Enzymatic Protocols for the Synthesis of Designer DNA

Colette J. Whitfield

A thesis submitted for the degree of

Doctor of Philosophy in Chemistry

School of Chemistry

Newcastle University

December 2016

Abstract

The enzymatic synthesis of long DNA with a controllable sequence, length and functional content has been reported. This method, involves the heating and cooling of the reaction components, resulting in the extension of repeating units. The key components comprise of the oligo seed of interest, the deoxynucleotide triphosphates (dNTPs), and a DNA polymerase. Using a thermostable *Thermococcus gorgonarius* Family B DNA polymerase exonuclease minus variant, Z3, and 20 heat-cool cycles, long DNA up to 20,000 base pairs bearing repeating units between 1 to 40 bases was produced.

Incorporation of artificial nucleotides, with modifications ranging from single atom exchanges, 5-I-dCTP, 7-deaza-I-dATP, 5-Br-dUTP and 6-S-dGTP, to long chains, 5-C_8-alkyne-dCTP, was demonstrated. Modifications situated in the major groove have little effect on the DNA polymerase efficiency but reduced enzymatic processivity is observed if the modification lies in the hydrogen-bonding region. By tailoring the oligo seed, it is possible to synthesise long designer DNA to include modifications at user defined positions. The modified DNA product lengths are similar to the unmodified DNA products, except for 6-S-dGTP, which yielded DNA of 500 base pairs.

6-S-dGTP is renowned for strong metal interactions, and was exploited for the specific localisation of Au^{+}, Ni^{2+}, Cd^{2+} and Au^{3+} at repeating G positions. As the final 6-S-DNA product is limited in length, an alternative thiol modification was investigated. Using phosphorothioate dNTPs, sulfur bearing DNA products similar in length to the unmodified DNA were produced after 30 heat-cool cycles. This enabled the specific positioning of Au-nanoparticles through careful oligo seed design.

DNA bearing the 5-C_8-alkyne-dCTP provides alkyne anchors at sites sitting in the major groove. To demonstrate the ability to add a second layer of design, click chemistry with azide-fluor-545 was investigated. This opens up potential routes to more complex modifications via organic synthesis at precise sites within the designer DNA.

Acknowledgements

I would like to take this opportunity to thank everybody that has contributed to my PhD work. To my supervisors, Dr Andrew pike, Prof Bernard Connolly and Dr Eimer Tuite, I am thoroughly grateful for the guidance, advice and opportunities provided throughout my PhD. The opportunity to travel and extend my network has broadened my research interests and inspired me to continue in academia.

To all the Connolly and Chemical Nanoscience lab members, past and present, I am thankful for your inspiration and friendship and for making the workplace an enjoyable environment, especially the Pike group: Sam Lunn, Rachel Little and Tom Bamford. In particular, I'd like to thank Dr Scott Watson and Dr Chris McGurk for their ample AFM knowledge and skills, and Mrs Pauline Heslop and Dr Javier Abellon-Ruiz for their endless biochemical knowledge and advice.

I would also like to thank the group of Prof. Ijiro at Hokkaido University for their hospitality and use of facilities to conduct several experiments.

Daniel, for your incredible patience, laughter and late night hospital teas, thank you – I couldn't have done it without you. And his family, for their continuing interest and support.

Lastly to my family, thank you for your continuing enthusiasm for my research and endless encouragement and inspiration and for reminding me there is always a light at the end of the tunnel.

Thank you to the Biotechnology and Biological Sciences Research Council for providing this studentship.

Table of Contents

Abbreviations

α	alpha
ε	extinction coefficient
λ	lambda
μm	micrometer
μL	microliter
nm	nanometers
π	pi
1-D	one-dimension
2-D	two-dimensions
3-D	three-dimensions
AFM	atomic force microscopy
bp	base pairs
BSA	Bovine serum albumin
dA	deoxyriboadenine
dC	deoxyribocytosine
dd	dideoxy
dG	deoxyriboguanosine
dT	deoxyribothymine
Deep Vent	*Thermococcus litoralis*
DNA	deoxyribonucleic acid
dN	deoxynucleotide
dsDNA	double stranded DNA
E. coli	*Escherichia coli*
EDTA	ethylenediaminetetraacetic acid
exo-	exonuclease minus
exo+	exonuclease positive
FRET	fluorescence resonance energy transfer
HPLC	high performance liquid chromatography
HRMS	high resolution mass spectroscopy
Kb	kilo base pairs
mM	milli molar
MP	monophosphate

NMR	nuclear magnetic resonance
NP	nanoparticle
oligo	oligonucleotide
PCR	polymerase chain reaction
PEAR	polymerase-endonuclease amplification reaction
Pfu	*Pyrococcus furiosus*
Pol	polymerase
RCA	rolling circle amplification
RNA	ribonucleic acid
SDS	sodium dodecyl sulphate
ssDNA	single stranded DNA
Taq	*Thermus aquaticus*
TBE	tris, boric acid and EDTA
TBTA	tris(benzyltriazolylmethyl)amine
TEM	transmission electron microscopy
Tgo	*Thermococcus gorgonarius*
T_m	melting temperature
TOF	time of flight
TP	triphosphate
SERS	surface-enhanced Raman spectroscopy
UV-Vis	ultraviolet visible spectroscopy

Chapter 1.

Introduction: the Exploitation of DNA for Applications in Nanomaterials

Table of Contents

The elucidation of the DNA structure in 1962 heralded an era of intense research into its biological roles based on the antiparallel double helix published by Watson and Crick.[1] More recently, it has become an integral component in nanomaterial fabrication due to its favourable nanoscale dimensions and linear structure. However, fundamentally it is the programmable genetic code of DNA that has allowed DNA research to go beyond its use as the molecule of life. Many studies into its vital roles in both cellular function and the cellular mechanisms to maintain its integrity, have provided the foundation to explore DNA as a tool for drug delivery,[2,3] sensing,[4] conductive nanomaterials[5,6] and nanomotors.[7,8]

1.1 Key molecular insights of DNA

The initial discovery by Watson and Crick, outlined the complementary base pairing of; adenine (A) and thymine (T), and guanine (G) and cytosine (C),[1] Fig. 1. This discovery stimulated the many publications involving DNA as a tool in the synthesis of numerous nanoscale structures.[9,10] The complementary hydrogen bonding between strands, the $\pi-\pi$ interactions of the stacked nucleotides, Fig. 1,[11] and the increase in entropy on the dispersion of water molecules, results in the double helical DNA structure. The complementary base pairing system is highly conserved since faithful DNA copying is required during cell division and also protein synthesis to ensure the correct cellular function and protein coding sequence is maintained. In this section, the importance of correct base pairing will be discussed.

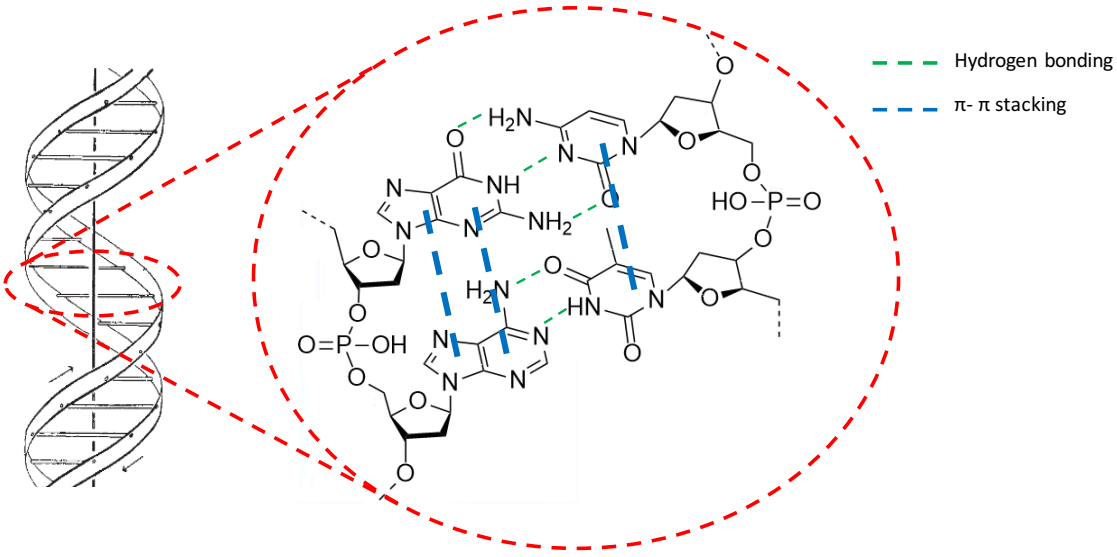


Fig. 1 A diagram of the DNA double helix.[1] The deoxyribonucleotides are joined together by a 3'- 5' phosphodiester bond. The blow-up images show the nucleotides in detail and how the stacking and hydrogen bonding interactions occur. A and T form a 2-hydrogen bond base pair where as G and C form a 3-hydrogen bond base pair and also have a better π orbital overlap, hence GC rich DNA duplexes are more thermostable.[12]

During replication high fidelity DNA polymerases faithfully recognize the Watson and Crick base pairs and maintain the genome.[13] However, one in every million nucleotide incorporations results in an error in the Watson and Crick base pairing rules.[13] Although there are many DNA repair proteins present in the cellular system, some mutations remain unfixed. The effects of these mutations have been studied due to their vital consequences in many diseases.[14] For example, DNA hypermethylation is a common marker for genetic disease and cancer progression.[15] DNA methylation at the 5 position of cytosine plays an important role in gene expression.[16] Consequentially, dysregulation of cytosine methylation can lead to genetic diseases, for example by deamination to become dT, therefore leading to mispairing.[17]

In addition, reactive oxidative stress is a common cause for genomic instability and disease.[18] Incorporation of 8-oxo-deoxyguanosinetriphosphate (8-oxo-dGTP) in the genetic code can lead to replication errors due to the dual coding potential to pair against either cytosine, via standard base pairing or adenine using the Hoogsteen base pairing mode, Fig. 2.[19] Once a lesion site has undergone mis-pairing, consequential mutations can occur – for example, the Hoogsteen mis-pair to A, would then replicate to incorporate dTTP, rather than dGTP. There are many DNA mismatch repair pathways in place – both incorporation against the lesion, and extension past the lesion determine the damage of the mutation. In the case of oxidative stress, once the mismatch replication has occurred, translation from messenger RNA my cause mis-coding to a different amino acid during protein synthesis - repercussions depend on the effect of the incorrect amino acid on protein structure and function. Replication mispairing can also occur from replication errors to produce base-base mismatches, for example, from T against T (purine-purine) to A against C (pyrimidine to purine).[20,21] This mispairing commonly occurs from an unfavourable tautomer to the imino or enol forms which are possible with each base.[21]

From studies such as those outlined above, the detailed understanding of how the DNA helix relies on a high-fidelity base-pairing system, and the repair of base mutations has enabled the exploitation of DNA as a precisely designed building block for nanoscale structures.

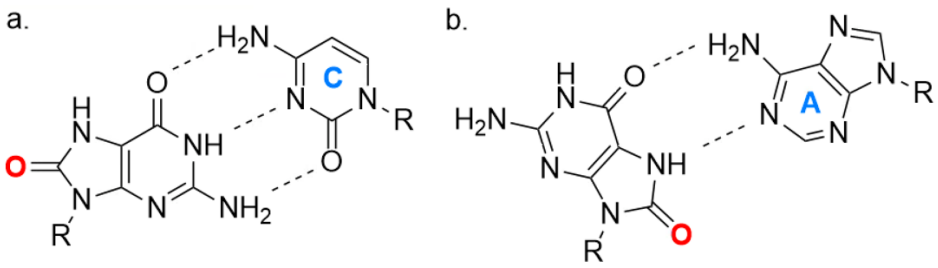


Fig. 2 Alternative base pairing with 8-oxo-dG, **(a)** conventional Watson and Crick base pairing rules, matching C to 8-oxo-G, **(b)** Hoogsteen base pair, matching A to 8-oxo-dG. R = DNA molecule.

For healthy cellular function DNA must remain intact. Therefore, mutations must be repaired, for example, by nucleotide excision repair mechanisms.[22] In nucleotide excision repair a cascade of protein interactions are in place, activated by the sensitive detection of local DNA instabilities,[23,24] to stimulate the specific repair proteins as molecular machines. The understanding of complex biological mechanisms has led to the development of synthetic biological molecules for the design of nanorobots and nanomotors.[25,26] Each nanosystem harnesses the simple building blocks of life, either nucleic acids or amino acids, to construct complex systems. For example, the ribosome can specifically detect the three base sequence that corresponds to an amino acid through the simple Watson and Crick base pairing rules. Likewise, the DNA polymerase is able to select and bind the correct complementary base when replicating DNA. Binding of the correct deoxyribonucleotide induces a conformational change in the DNA polymerase, which opens up the active site for the next deoxynucleotide selection.[27] These processes are possible by the intricate recognition of the nucleoside molecular structures. On detection of an incorrect structure, the polymerase will not reopen the active site, and will halt replication, Fig. 3,[28] – part of the quality control procedure of these molecular machines. The growing interest of DNA based biological machinery has developed into a research curiosity in how specifically the DNA enzymes can be manipulated.

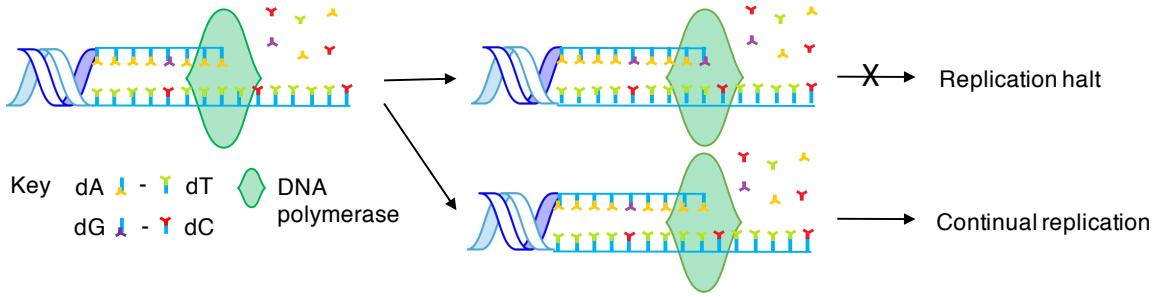


Fig. 3 DNA replication by a DNA polymerase, incorporating one deoxynucleotide at a time. The top pathway shows a mis-pair incorporation of dG-dT, not following the Watson and crick base pairing rules, therefore replication is stalled. The bottom pathway shows the faithful replication of DNA, allowing continuation of replication.

Many DNA mutations can be due to chemical alterations of the nucleobase structure. Chemical carginogens, for example polycyclic hydrocarbons, covalently attach to the nucleobase to form bulky DNA-adducts resulting in lesions.[29] Several groups have studied how the binding of bulky ligands can be accommodated by various DNA polymerases in order to bypass lesion formation.[30] In a similar study, the Kool group determined the consequence of steric effects on DNA polymerase efficiency by the replacement of both oxygen atoms on thymidine by either hydrogen or halogen atoms (F, Cl, Br and I), Fig. 4. The efficiency of pairing the thymidine derivatives against dA peaked with chlorine modifications suggesting steric constraints were not the only factor affecting the polymerase activity.[31] Strechenbach *et al.,* elaborated on the steric effects by studying the behaviour of a range of DNA polymerases.[32] The DNA polymerase handling of native, methyl and ethyl nucleobase derivatives was compared using a selection of DNA polymerases with different active sites.[32] The relative efficiencies varied - the *Sulfolobus solfataricus* P2 DNA polymerase IV demomstrated a lower fidelity than the *Escherichia coli* (*E. coli*) DNA polymerase I Klenow fragment. Additionally, Y-family DNA polymerases are more promiscuous than typical replicative DNA polymerase – displaying an ability for mutagenic bypass.[33] Structural studies with Y-family DNA polymerases revealed the lower fidelity is due to a weak DNA-DNA polymerase interaction as a result of a larger active site.[30] The importance of understanding the DNA polymerase active site structure, resulted in a boom of research into DNA polymerase structure-function relationships and how these can be exploited and manipulated.

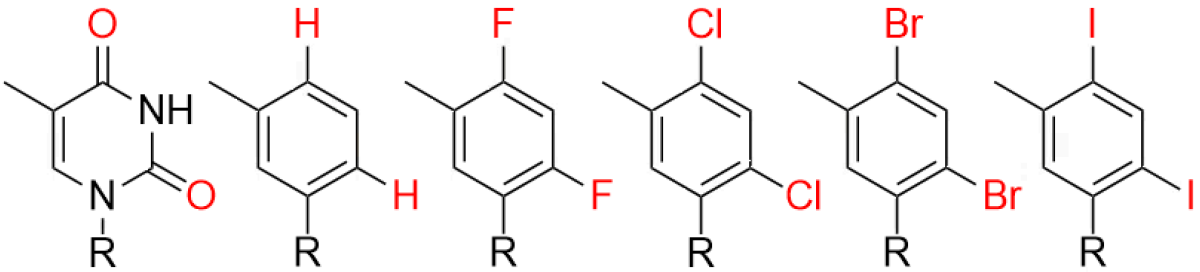


Fig. 4 Modified nucleotide structures of dT derivatives used for DNA polymerase efficiency studies.[31] The modifications range from an exchange of an H atom up to an I atom. R is the deoxyribose sugar. dT is shown on the left followed by the modifications with increasing bond length, left to right.

Discovering the structure-function relationship of DNA polymerases has sparked the development of artificial DNA polymerases with improved attributes for particular applications from high fidelity replication DNA polymerases for PCR to more promiscuous DNA polymerases for random mutagenesis assays.[34,35] Random

mutagenesis assays are performed with error-prone DNA polymerases to promote mis-pairing during PCR, to produce DNA which will code for a mutant protein – aiding the research of important structural motifs for protein interactions and active site function. The 3'-5' exonuclease domain accounts for the proof reading ability of the DNA polymerase.[36] The bacterial *Thermus aquaticus* (Taq) DNA polymerase (Pol) lacks the 3'-5' exonuclease domain (exo-) reducing its fidelity, however for error prone PCR it is still too accurate.[37] To provide an avenue for random mutagenesis, the proofreading 3'-5' exonuclease domain must be disabled in high fidelity DNA polymerases to allow mis-pairing to go unfixed.[34] Many artificial DNA polymerases are commercially available, and can be exploited for non-biological applications by the incorporation of modified dNTPs with potential for functional programming.

In addition, some DNA polymerases can cause errors by inserting a short extended sequence through the DNA slippage reaction,[38] a naturally occurring process that arises from loop formation from unstable tandem repeat sequences. The process occurs when one strand forms a stable internal loop, shifting the downstream sequence to accommodate the loop, revealing a sticky end. The sticky end can then be filled in by the DNA polymerase, leading to extension of the DNA sequence. Investigations into the tandem repeat instability initially began when notable deletion mutations occurred in association with the repeats.[39,40] The most commonly reported slippage sequence is the expanded CAG repeat that contributes to Huntington's disease.[41] The expansion of the CAG sequence can lead to an increase of 1,000 base pairs (bp) in the huntingtin protein gene – expansive repeat length has been show to correlate with disease severity.[42]

In another example, the telomere sequence consists of hexa-nucleotide tandem repeats, maintained by telomerase and DNA replication enzymes to reduce cell ageing.[43] The hexa-nucleotide telomere sequence has also been reported to extend by the slippage reaction.[44] These biological extension processes sparked studies into understanding the extension and slippage mechanisms of non-native DNA sequences, which led to the potential to enzymatically synthesise DNA for alternative applications.

DNA sequencing became possible once the base pairing system was elaborated, and suitably sensitive detection methods for nucleobases were developed based on the fundamental chemistry of DNA. The human genome project, launched in 1990,[45] fuelled the drive for second generation DNA sequencing methods – all DNA sequencing is based on the detection of the correct base pairing. DNA Sanger sequencing, developed in 1972,[46] originally sequenced DNA by extending one deoxynucleotide at a time and analyzing the extension using acrylamide gels. This was a heavily laborious technique, motivating the development of the more commonly used techniques today, such as, the dideoxynucleotide (ddN) terminating method, Fig. 6 (a). Sanger sequencing utilizes high fidelity DNA polymerases to incorporate the correct deoxynucleotide, in contrast, the SOLiD (sequencing by oligonucleotide (oligo) ligation and detection) method extends the DNA from 3'-5' using a DNA ligase. The SOLiD system – a four-colour DNA sequencing method, utilizes four short oligos, each conjugated to a fluorophore to detect base interactions, Fig. 6 (b).[47] Although there are many new DNA sequencing methods, Sanger sequencing is the most commonly used due to the long read lengths (300 - 900 bp) and short run time (0.3 - 3 hours) compared to read lengths of 50 - 75 bp and a run time 7 – 14 days for SOLiD. 3rd generation DNA sequencing utilises the electronic signatures of each individual base to detect their presence between the electron tunnelling junction.[48] However, optimisation of the 3rd generation techniques is required to remove background noise to accurately decipher between each base. The ability to sequence single DNA molecules would allow the sensitive analysis of mutations and cloning without the initial upscaling required for Sanger sequencing.

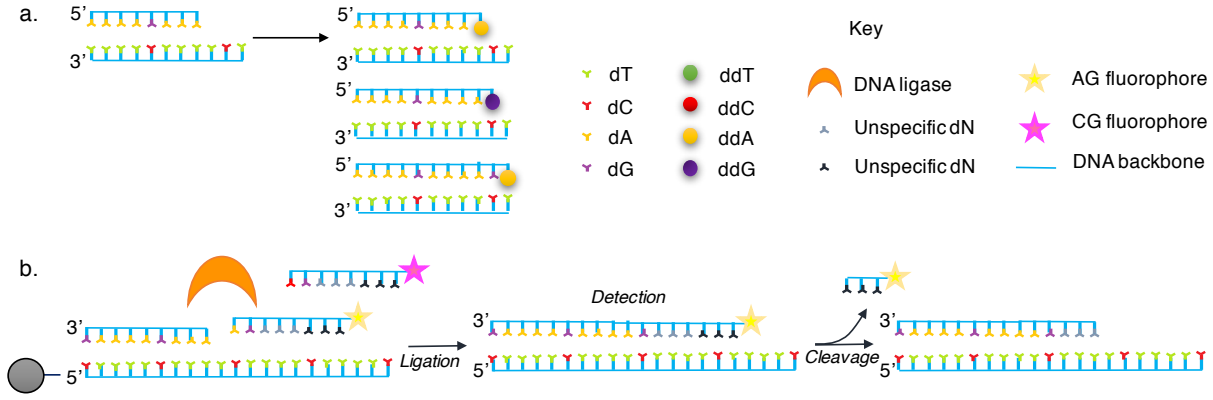


Fig. 6 DNA sequencing by **(a)** Sanger sequencing and **(b)** SOLiD sequencing. The example shown for Sanger sequencing uses the labelled ddN terminator method. The SOLiD method binds the template DNA to a bead before sequencing.[47]

The biomolecular study of DNA, has provided fundamental knowledge into DNA replication and DNA polymerase structure and function. DNA mutations and, in particular, information as to how modified or artificial bases may be incorporated into DNA whilst maintaining the overall double stranded (ds)DNA structure has been understood. The study of DNA polymerase structure-function relationships has aided the discovery of the 3'-5' exonuclease function and its importance in high-fidelity replication, or conversely, the importance of its removal to hinder detection of incorrect base pairing. As many artificial DNA polymerases are commercially available, DNA polymerases can be manipulated to work as adaptable nanomachines to synthesise DNA dependent on the biochemical study or function. The study of the intricate interactions between protein complexes and DNA has opened avenues to control DNA as a molecular tool.

1.2. DNA nanotechnology

The non-biological potential of DNA was first expressed in 1982 by Ned Seeman through his proposal "It is possible to generate sequences of oligomeric nucleic acids which will preferentially associate to form migrationally immobile junctions, rather than linear duplexes, as they usually do".[49] Building nanoscale structures is achievable by two approaches; bottom-up or top-down, that is, starting from singular atoms and building up by the addition of more atoms or starting on the large scale and carving out structures to leave nanoscale material. Bionanotechnology – the use of biological materials as building blocks is one bottom up approach, utilising attractive properties, optimized over billions of years. DNA has lent itself as a useful component in nanomaterials and DNA based bionanotechnology due to its versatility and ease of manipulation to control self-assembly by Watson and Crick base pairing. The DNA duplex has two dimensions on the nanoscale, *ca.* 2 nm high and wide, and up to many microns in length. This high aspect ratio makes it an attractive building block for nanomaterial assembly, particularly 1-dimensional (1-D) nanowires.[50,51]

The emergence of sub 100 nm material characterization techniques has allowed the visualization of intricate designs as higher level of detection is available through atomic force microscopy (AFM),[52] transmission electron microscopy (TEM)[53] and cryo electron microscopy.[54,55] DNA's predictable structure lends itself to be manipulated into higher ordered architectures, Fig. 6, and provides a key framework to build from, to

design nanomaterials. The emergence of 1-D DNA nanomaterials, 2-D DNA origami in 2006[56] followed by 3-D DNA origami in 2009,[54,57,58] have provided platforms to build nano robots,[25,26] nano capsules,[3] single molecule detection devices,[59] nano computers[60] and many more.

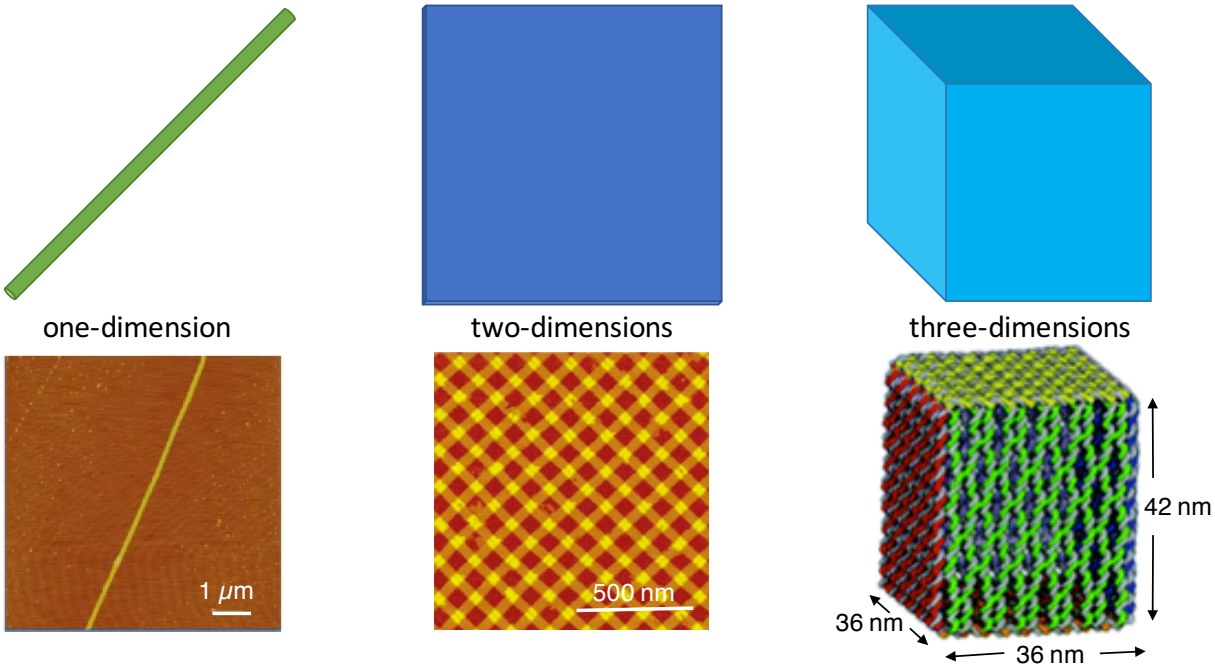


Fig. 6 An example of 1-D, 2-D and 3-D DNA nanostructures. 1-D DNA nanowires height average is 15 nm.[61] 2-D and 3-D structures are prepared by DNA origami self-assembly.[54,62]

The following sections describe recent progress into the use of DNA in nanomaterial synthesis from 1-D through to 3-D architectures and the incorporation of additional functionality through templating and scaffolding techniques.

1.2.1 *1-D DNA nanomaterials*

The high aspect ratio of DNA naturally lends itself to the use of DNA in 1-D nanomaterial. The obvious role for DNA to play would be to act as a nanowire. However, the conductivity of DNA has been a controversial topic for many years. In 1992, Murphy *et al.,*[63] suggested the naked DNA molecule possesses conductive character by the π–π staking of the nucleobases. Braun *et al*,[64] then contradicted this finding in 1998, expressing insulating characteristics. It is now understood that the conductive or insulating nature of DNA, is largely dependent on its individual components,[65] hence can express each feature under different circumstances. The Porath laboratory has demonstrated the conductive nature of native G-quadruplex DNA.[66] The G-quadruplex structure provides an additional charge transport avenue[67,68] and yields a rigid structure with limited defects as the electronic transport

doesn't rely on molecule-substrate interactions.[66] Therefore, there has been considerable work on the design of hybrid DNA systems, that exploit its 1-D nature but also add conductive material, through both templating or scaffolding approaches.

1.2.1.1 *DNA as a template*

Recent advancements in the production of a nanoscale conducting wire have been developed through the use of DNA as a template for conductive material.[64,69,70] Each deoxyribonucleotide component possesses a phosphate group, which is known to have a high affinity for cationic species, i.e. metal ions. Metal ion exchange between the stabilising counter ions (usually Na^+ or K^+) and the seed metal ion (Ag^{3+}, Cu^{2+}, Cd^{2+}, etc) is commonly used for metal ion deposition.[70-72] After electrostatic association of the metal cation with the phosphate backbone, a reduction step to the metal leads to the templating of DNA with a conductive metallic coating. Initial attempts to synthesise and characterise DNA templated silver wires were performed by Braun *et al.,*[64] by attaching specific primers to two electrodes, followed by the annealing of λ-DNA to the primer, thus resulting in a long length of DNA, positioned between two electrodes, Fig. 7. Today, many conductivity measurements exploit the same strong thiol – gold interaction by modifying the 3' and the 5' end of the DNA strands with a thiol group, which can interact with the gold electrode.

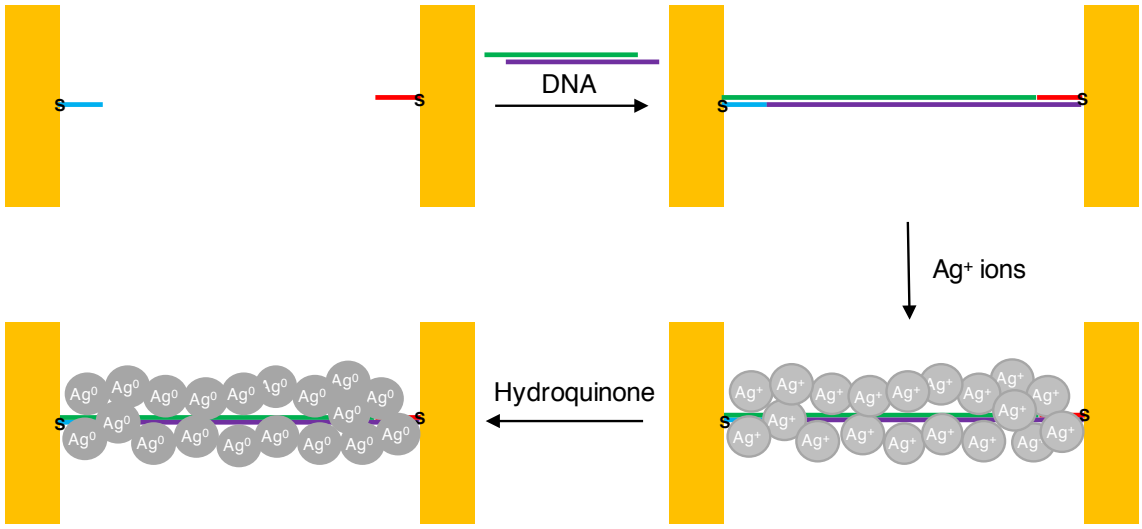


Fig. 7 Synthesis of a DNA templated Ag wire between a gold electrode gap by the attachment of specific primers via thiol-gold interactions, followed by the annealing of the DNA template and a succession of Ag^+ addition followed by reduction with hydroquinone addition.

Long range electronic transport along DNA has also been reported by analysis after deposition onto a surface bearing Ga nanoparticles, using platinum electrodes.[73] Since conductive analysis techniques emerged, several groups have successfully produced a conducting wire using DNA as a template. The most popular metals include, Ag,[64] Pd,[69] Cu[51] and Pt.[74] DNA templated with Cu, originally reported by Wooley *et al.,* is commonly the industrial metal of choice due to its low cost and high level of conductivity, however, deposition of Cu^{2+} alone exhibited a granular coating requiring refinement.[70] Seeding with heteroatoms has tackled the challenge to control nucleation to build a more homogeneous nanowire. Initial seeding with Pd^{2+}, followed by treatment with Cu^{2+} produced a nanowire structure of controllable height depending on the copper electroless plating time.[75] Due to the luminescent properties, CdS nanowires have been heavily studied by several groups.[72,76] Bigham *et al.,* studied the effect of base composition on nanoparticle size, expressing a 10 nm reduction in nanoparticle size on increasing poly[A] content, thus altering the extinction maximum of the nanowire.[77] This phenomenon provides an avenue to controlling nanowire size and thus nanowire properties. The formation of monodisperse CdS particles can then undergo Ostwald ripening to produce a smooth wire like singular structure, Fig. 8.[78] Other cadmium based nanocluster nucleation has been analysed on DNA, such as CdSe.[79] Artemyev *et al.,* successfully organised CdSe nanorods along λ-DNA, producing a strongly luminescent nanomaterial with high ordered organisation.[80]

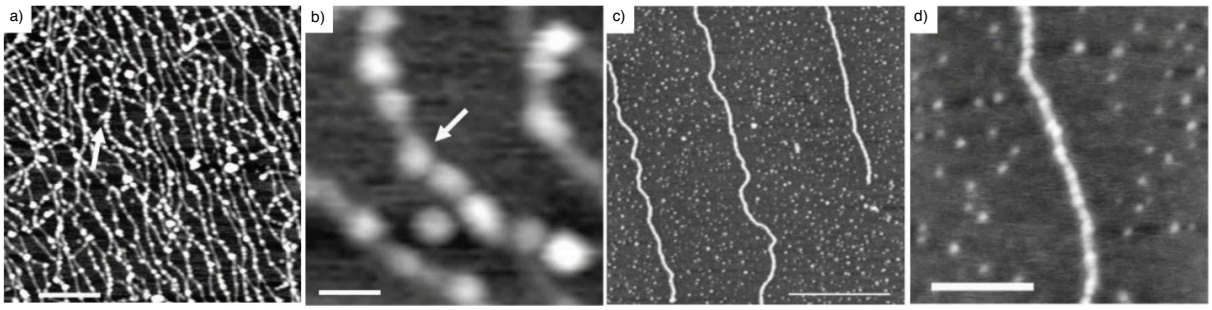


Fig. 8 AFM images of CdS templated on DNA, attached to a mica surface. Monodisperse particle formation is shown in **(a)** (scale bar 200 nm) and **(b)** (scale bar 25 nm). After Oswald Ripening, a smoother nanowire structure is shown **(c)** (scale bar 500 nm) and **(d)** (scale bar 100 nm).[78]

These results are positive for the production of a smaller dimension conducting wire as their width is on the sub 20 nm scale, however, by this method, the circumference and conductivity of the nanowire is not easily controlled, reducing the complexity of the structures possible. Ijiro's lab developed a method to synthesise triblock DNA (A-type, B-type, A-type), consisting of regions of poly(dG).poly(dC) (A-type) and

poly(7-deaza-dG).poly(dC) (B-type).[81] Cisplatin preferentially binds to regions of poly(dG),[82] providing an avenue for programmable deposition. The reduction of cisplatin covalently attached to the DNA, provides nucleation sites for the addition of free cisplatin and dimethylborane as a reducing agent to produce a nanowire.[81,83] The triblock nature of the 1-D DNA structure, presents a nano-gap specified by the length of the 7-deaza-dG region, which is controllable by the polymerisation time during synthesis of the triblock DNA.[81] This potentially leads to the fabrication of the next generation of sophisticated nanowires.

Pyrrole is groove binding and is readily polymerized by oxidation at neutral pH to produce nanowires with continuous coverage.[78] Similarly, polymerising aniline in acidic conditions in the presence of DNA, templates the polyaniline in 1-D.[84] Protonation of the DNA backbone is required to associate aniline and again, oxidation is required to produce the conductive wire. Both polyaniline and polypyrrole are chemical compounds, capable of chemical modification to present additional function on the DNA hybrid structure, for example, bearing an alkyne functional group provides an anchor for click chemistry with an azide.[72,85] Click chemistry is a commonly used reaction due to the mild conditions and specificity for the alkyne site. Alkyne bearing DNA can also be exploited for controlled Ag^+ binding.[64] Several articles discuss the binding of Ag^+ to alkyne groups[86,87] and thus, its presence along DNA provides evenly spaced nucleation sites to produce a uniformly distributed nanowire using Tollens reagent.[71]

The range of materials available by using DNA as a template is vast, however, for a wire to be useful in a circuit, it must not contain any defects, which is still posing a challenge. To reduce defects, a more designed approach, starting with a more intricately tailored DNA sequence, could overcome the challenge, producing a wholly homogenous DNA nanowire. Furthermore, the templating process covers the base-pairing ability of the DNA, removing the opportunity to self-assemble electronic components into a functioning circuit.

1.2.1.2 *DNA as a scaffold*

DNA is a complex structure and is capable of many interactions with conjugate substrates. Functionalisation of DNA as a scaffold is possible by the chemical modification of the native DNA structure. Modifications to the C5 position of purine deoxynucleotides by Sonogashira alkylations followed by clicking with pyrrole

derivatives provides a site specific anchor for the oxidative scaffolding and therefore, offers an additional element of control.[88] Further developments led to the use of polynucleotides as a solid support to bind porphyrin molecules, which chelate metal ions,[89,90] providing a promising method for the fabrication of nanowires. A significant advantage of these nanowires is that they are bio-compatible and therefore have potential within medical applications and additionally, the hydrogen bonding is retained. Furthermore, control of functionalised attachment is more precise than current DNA templating methods. This specific attachment chemistry can be used to produce self-assembling structures, conjugated with proteins or receptors by streptavidin and biotin linkers.[91] The advantage of this method of modification over DNA templating is the increased control over the location and density of the attachments.

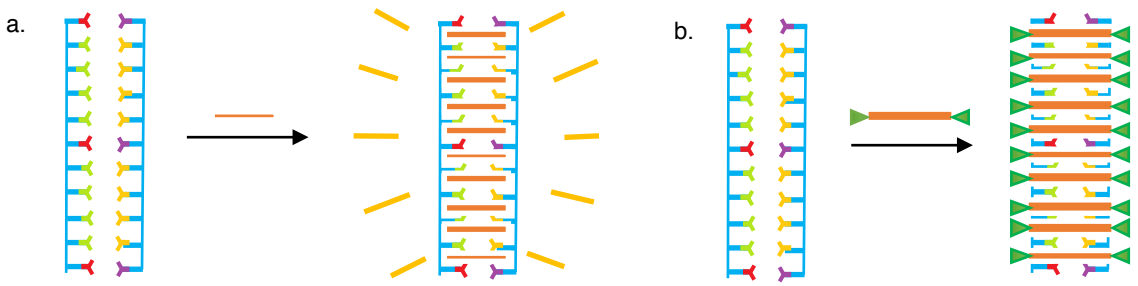


Fig. 9 DNA intercalation with **(a)** planar aromatic fluorescent compound and **(b)** planar aromatic compound with external functional groups for a second layer of DNA modification.

Intercalation of planar aromatic compounds has been a popular approach to DNA functionalisation since the biological application of ethidium bromide addition for gel visualisation, Fig. 9 (a). Developments have led to the synthesis of numerable fluorescent intercalators[92,93] with many commercially available, for example, from Thermo Fischer Scientific (Renfrew, UK), Biorad (Hempstead, UK) or Promega (Southampton, UK). DNA intercalators bearing external functional groups provide anchors for additional functionalisation, Fig. 9 (b), similar to the major and minor groove binders. Click chemistry can once again be exploited to react with redox active groups, such as ferrocene, to study the redox behaviour in close proximity to DNA – a reduction in the redox potential is observed upon reacting.[94] An ability to control the positioning of additional molecules along the DNA structure, opens avenues to study distance dependent interactions. Ferrocenyl-dT and ferrocenyl-dC phosphoramidite derivatives were prepared using the Pd-catalysed Sonogashira cross coupling reaction with I-deoxynucleotides and ethylene ferrocene and incorporated into oligos using

automated solid-phase synthesis.[95,96] By the incorporation of artificial dN into oligos, the position of modification can be controlled. However, the length of the sequence is limited by the phosphoramidite coupling method's reducing yield on increasing length. So far, this discussion has included DNA as a scaffold for external modifications. Modification within the DNA duplex is also possible, and commonly achieved by the inclusion of artificial bases. An artificial nucleobase can replace a natural one while maintaining the structural integrity of the DNA duplex due to the surrounding base pairs.[97-99] Artificial nucleobases are commonly planar and aromatic to maintain the base stacking interactions,[100-102] however, a large range of structures have been documented, Fig. 10. Fluorescent modifications, quenched by the stacking interactions, can reveal unstable regions of the DNA structure if fluorescence is detected.[103] Alternatively, the fluorescent DNA bases can be used to induce DNA melting by initiating a photodegrading process of the fluorophore, altering the structure and reducing stability.[104] Increasing the alphabet of DNA bases enhances the possibilities of synthetic DNA based materials and DNA probes.

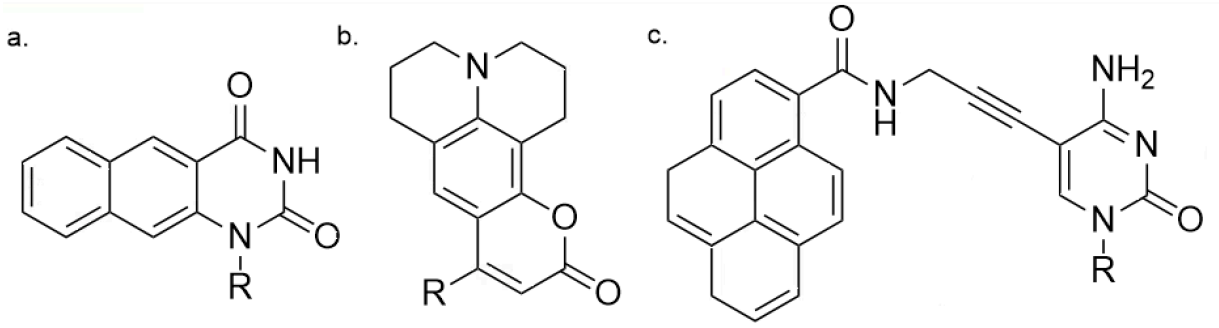


Fig. 10 Modified nucleobase analogues of thymidine and cytosine - **(a)** benzo(g)-quinazoline-2,4-(1H,3H)-dione T derivative[102] **(b)** coumarin β–C-riboside[105] and **(c)** 5-(1-pyrenecarboxamido)propyl-2'dexoy cytidine.[106] R = deoxyribose sugar

By using DNA as a scaffold, high cost organic synthesis can be used to produce artificial nucleosides, for example conducting polymer units, or redox reporting groups, which can be precisely located using the phosphoramidite method. However, only short oligos are possible unless enzymatic synthesis becomes feasible.

Despite the potential for DNA to assemble material on the nanoscale by either templating or scaffolding to give 1-D hybrids, more complex structures are desirable. There has been a long history of self-assembly of DNA oligos to fabricate intricate 2-D and 3-D structures and more recently to add functionality to these architectures.

1.2.2 DNA assembled nanostructures

dsDNA self-assembly can be precisely controlled by the synthesis of programmable building blocks which drive the fabrication of higher ordered complex structures – including DNA junctions, tiles and origami.[10] This has led to the production of DNA structures in 2-D[56,107,108] and 3-D.[54,109]

1.2.2.1 2-D DNA nanostructures

DNA nanotechnology began by the recognition of the Holliday junction in 1964.[110] Since then, many 2-D structures have been developed from 3- and 4- armed junctions to lattices and higher ordered motifs, Fig. 11.[62,111] Initial attempts by Seeman involved the repetitive restriction and ligation of new strands, covalently linking the structure.[112] In more recent examples, the DNA structures are formed by the planned self-assembly of oligos by designing specific complementary base pairing at the junctions.

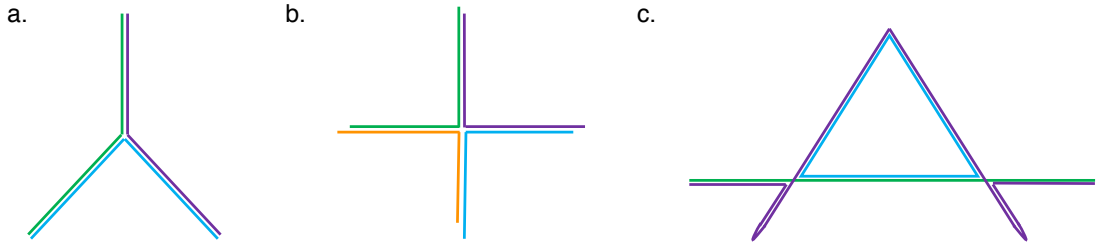


Fig. 11 2-D DNA structures: **(a)** 3-armed junction, **(b)** 4-armed junction,[49] **(c)** bulged 3-arm junction.[113]

Extended assemblies consisting of multiple junctions led to the synthesis and design of DNA cross over tiles, Fig. 12.[114] Cross over tiles are assembled using many Holliday junctions consisting of sticky ends to provide anchors for attachment to multiple tiles to form a higher level architecture.

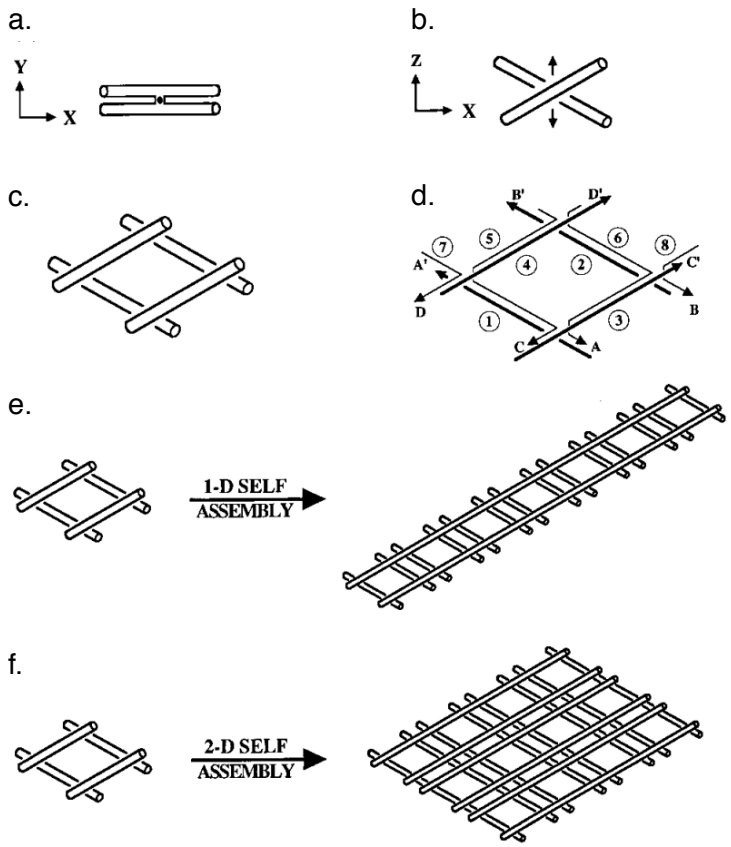


Fig. 12 The assembly of DNA tiles by the arrangement of Holliday junctions.[114] **(a)** a horizontal view down the central axis, **(b)** a top down view of the Holliday junction, **(c)** the combination of four Holliday junctions, **(d)** the strand positions of (c) – strands that are continuous are shown as a dark line where as strands involved in cross overs are shown as thin lines, **(e)** 1-D self-assembly of the Holliday junction motifs, **(f)** 2-D self-assembly of Holliday junction motifs.

A further breakthrough came when the assembly of short oligos by a single stand of ssDNA was achieved by Rothemund.[56] Here the design of several 'staple strands' complementary to precise positions on a longer ssDNA scaffold led to its controlled folding into a structure or shape pre-determined by the sequence, Fig. 13.[56]

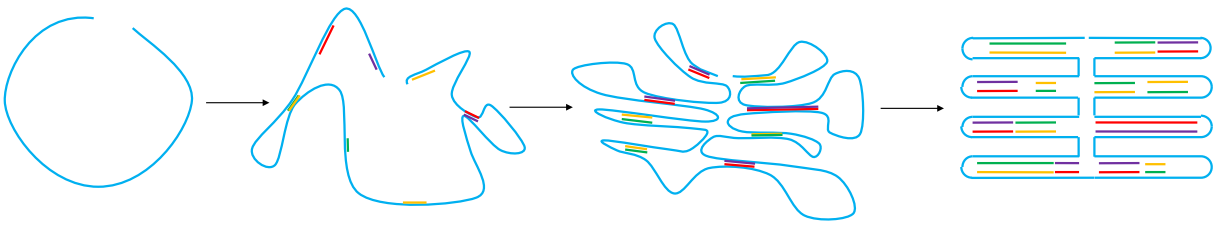


Fig. 13 DNA origami construction - the consecutive addition of programmed 'stable strands' to fold the long ssDNA scaffold in the DNA origami structure. The 'staple strands' are designed to bind only at specific regions of the ssDNA scaffold to direct the self-assembly of the designed structure, for example, in this case a rectangle is assembled.

DNA origami is now a widespread approach for nanomaterial fabrication and a wide range of shapes have been reported, Fig. 14.[56] Additionally, higher ordered structures

can be assembled by linking each origami shape or structure.[115] 2-D DNA origami nanostructures can be designed to bear regions of modification for binding sites to provide a route to position sensors, chemical content or biomolecules – much like a breadboard in microelectronics[116] - a useful surface for nanomedicine and detection.

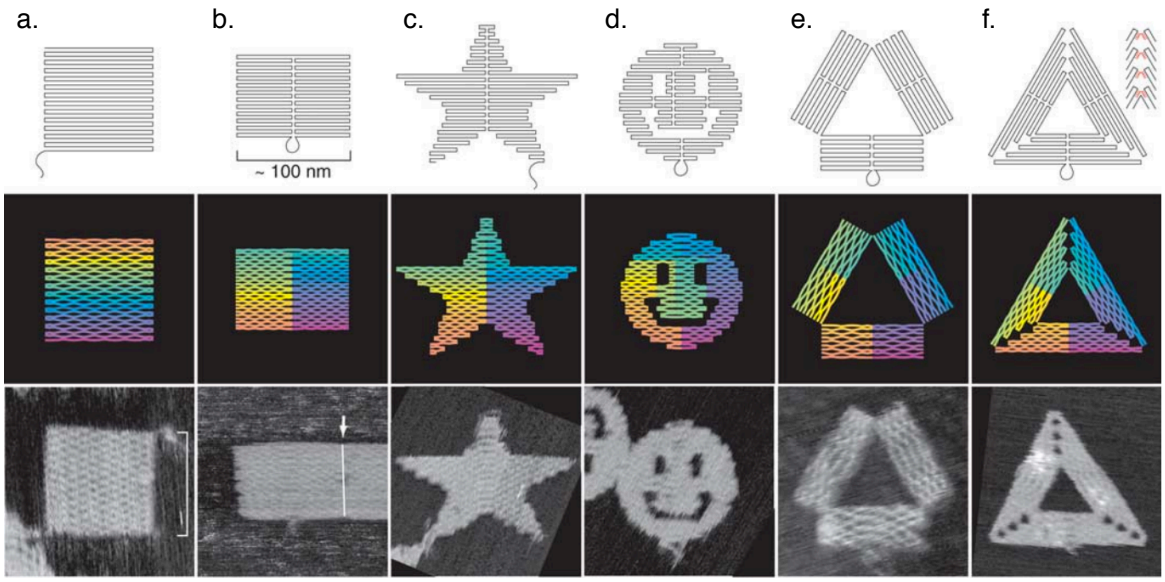


Fig. 14 DNA origami shapes **(a)** a square, **(b)** a rectangle, **(c)** a star, **(d)** a smiley face, **(e)** and **(f)** triangles. The top row outlines the folding of the ssDNA scaffold, the second row displays the helical structures and the bottom row shows the atomic force microscopy (AFM) images performed in a liquid cell.[56]

1.2.2.2 *3-D DNA nanostructures*

Since the development of the 2-D three armed junction in 1982,[49] significant advancements in folding more complex structures through both Holliday junctions and DNA origami have led to the production of 3-D DNA structures.[117]

Several designs have been successful so far, including octahedrons,[118] icosahedron encapsulating designs,[119] and nanotubes.[120] 3-D DNA wire frames, Fig. 15 (a), can be designed by the complementary pairing of the sticky ends from 3-armed junctions.[109] The tetrahedron is a more rigid structure than the similar 3-D cubes, exhibited by AFM studies.[121] Turberfield's laboratory has demonstrated the biological prospect for the DNA tetrahedron by encapsulating a protein,[122] exhibiting a chaperone role for potential in drug delivery systems. DNA nanotubes are similarly synthesised by the combination of many 2-D tiles or ribbons.[123] The tiles are designed to prefer a curved structure to direct a thermodynamically favourable smooth tube. An alternative method to DNA nanotubes synthesis was demonstrated by Seeman's laboratory by the combination of six parallel DNA helices, each with two crossover sites, Fig. 15 (b).[124]

The nanotube lengths can be adapted by designing either blunt or sticky ends at the terminals.

DNA origami has also been utilised to design 3-D DNA structures. The DNA tetrahedron was designed by the folding of one M13 ssDNA scaffold by 248 oligos to exhibit a single structure, Fig. 15 (d).[58] Likewise, Shih *et al.,* designed the assembly of a DNA octahedron from the folding of one ssDNA scaffold by five 40 base oligos.[125] The 3-D DNA box, Fig. 15 (c) and (f), is constructed by the assembly of six interconnected DNA origami sheets, which can subsequently be folded into a hollow box.[54] Anderson *et al.,* developed a dynamic box, able to open and release its' cargo on encountering certain the stimuli – a competitor DNA oligo, or key, to preferentially bind to the lock sequences.[54] An ability to release cargo at a specific stimulus would aid many medical applications, such as cancer therapeutics to reduce unspecific cell death.

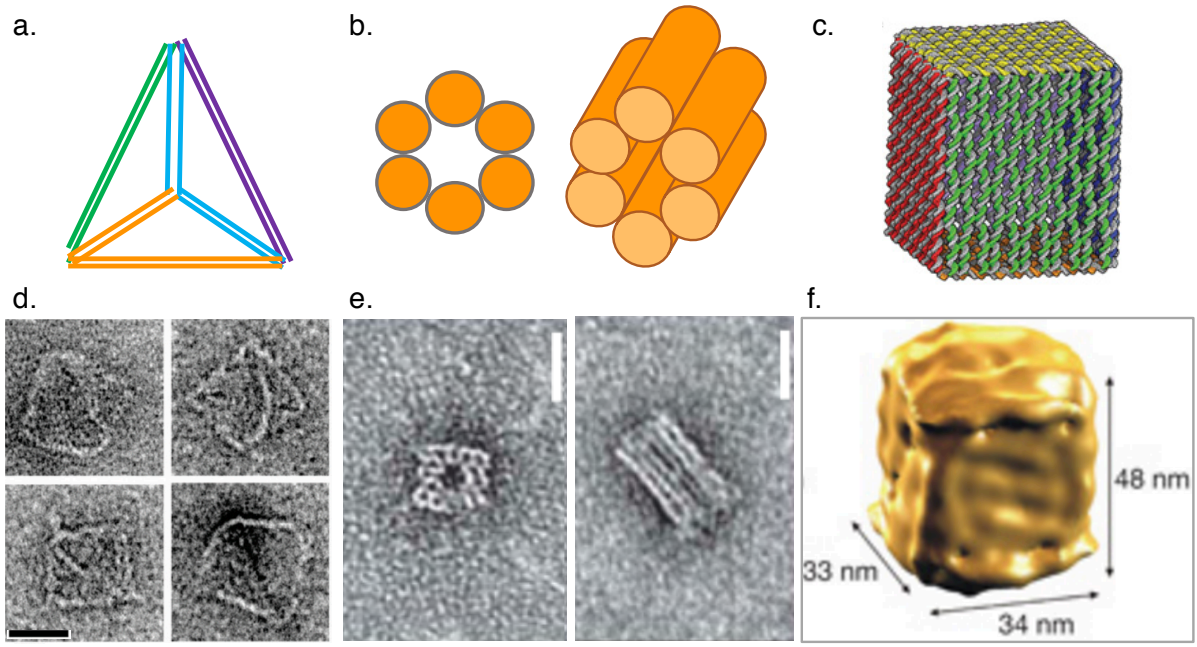


Fig. 15 3-D DNA origami structures: **(a)** DNA tetrahedron **(b)** DNA nanotubes[124] and **(c)** DNA origami 3D box - the different colours represent different DNA origami tiles, composing of different DNA sequences and joined together by 'staple strands'.[54] **(d)** TEM images of DNA tetrahedrons, **(e)** AFM images of DNA nanotubes[57] and **(f)** surface representation of cryo electron microscopy map.[54]

1.2.2.3 *Nanomaterial assembly on DNA origami structures*

DNA origami has presented DNA architectures with increased rigidity and a vast array of shapes, providing additional platforms to build nanomaterial assemblies. Geng *et al.,* designed a DNA origami structure, mimicking an electronic circuit, with improved metal ion deposition control.[126] However, height reproducibility was still lacking,

Fig. 16 (a), as compared to the 1-D DNA nanowires.[64] The site specific metallisation performed in the Wooley laboratory,[127] Fig. 16 (b), provides a method to position Au nanoparticles (NP) on certain regions of the DNA structure. The long DNA origami strand was folded into a T-shape by staple strands to bear single stranded (ss)DNA on the longest branch of the T-shape. Thiolated oligos, complementary to the ssDNA regions and previously bound to AuNPs could then site specifically attach to the longest branch of the T-shape. The advancement of 2-D DNA circuits presents the beginning of a possible next generation of electronic circuits.

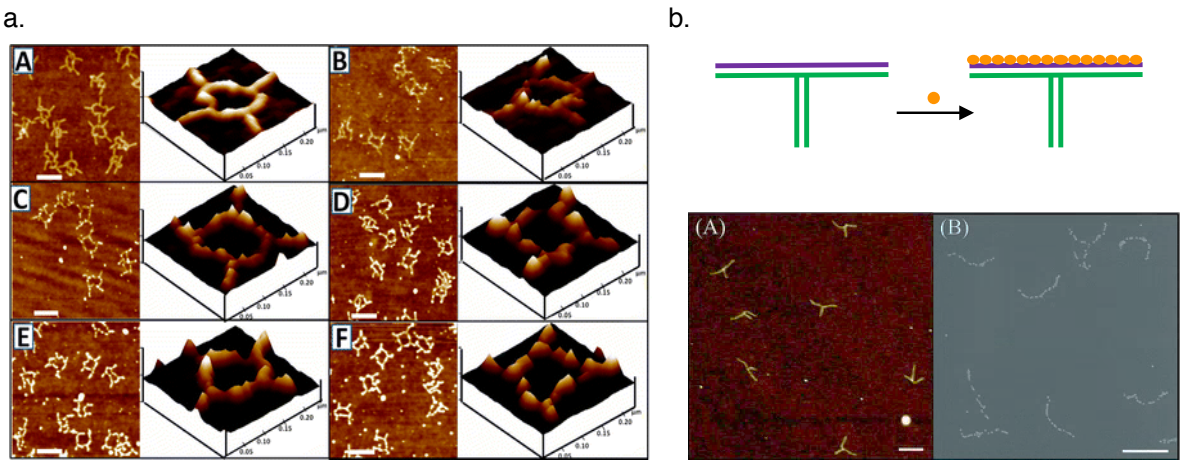


Fig. 16 2-D DNA origami, followed my metal templating **(a)** AFM height and 3D images of DNA origami structures on Si surface. (A) DNA structures. (B) 1, (C) 2, (D) 3, (E) 4, and (F) 5 palladium deposition. (left) Height images of substrate on surface (right).[126] **(b)** T-shape DNA structures bearing thiol groups (purple) on once region (A) to site specifically bind gold nanoparticles (orange) (B).[127]

One limitation for DNA origami is the DNA synthesis and sequence design of the start materials. To produce a DNA origami structure, many sequence specific oligos are required, as well as the long genetic ssDNA – a costly procedure. Therefore, the synthesis of both the DNA oligo staple strands and also the long ssDNA scaffold should receive more attention to diversify the possibilities the technique offers. It is conceivable that if a wider range of starting materials were available, in particular the choice of ssDNA scaffolds, then further complex nanoarchitectures could be envisaged. Additionally, by incorporating modifications into the synthesised DNA, controllable loading and an ability to precisely position functionality onto the origami structures is possible. This would open up avenues to molecular environment and spatial positioning studies as well as the positioning of NPs at precise nanoscale separations to enhance plasmon resonance. In light of the potential for DNA based functional nanomaterials outlined above the focus now turns to the synthesis of possible DNA components for these nanoarchitechtures.

1.3 DNA synthesis

As described above, DNA has played a vital role in the fabrication of nanomaterials. One limitation in the advancement of DNA based nanomaterials is the types of DNA available through the two main synthetic routes; automated synthesis and enzymatic routes.

The solid support method of DNA synthesis, the phosphoramidite method, involves the step by step build-up of each deoxynucleotide from 3'-5' to produce oligos. Nucleoside phosphoramidites were first developed by Beaucage and Carruthers in 1981[128] and since developed to synthesise oligos by the laborious coupling, washing, decoupling, and washing steps for each addition of the deoxyribonucleotides.[129,130] Key developments, such as the automation of the process, presented the phosphoramidite oligo synthesis as commercially viable. The oligo sequence can be designed by the user, with an added ability to synthesise oligos bearing modifications.[95,96] However, the length of the final oligo is still limited, a typical maximum sequence length is around 120 bases when ordered commercially (Eurofins Genomics). This is not surprising given the diminishing yield over many synthetic steps, thus reducing the overall yield even after each deoxynucleotide addition to 99.5 %, Fig. 17. A typical cost for a 20 base oligo is £20 for 0.2 μmole of purified oligo (Eurofins Genomics, UK Aug, 2016). For longer oligos, prices can reach up to £140 for 100 bases of 0.2 μmole purified unmodified oligo (Eurofins Genomics, UK, Aug, 2016). The inclusion of modifications into oligos is possible, however, the choice is limited by the commercial availability or it requires specifically designed and synthesised modified phosphoramidite nucleosides.

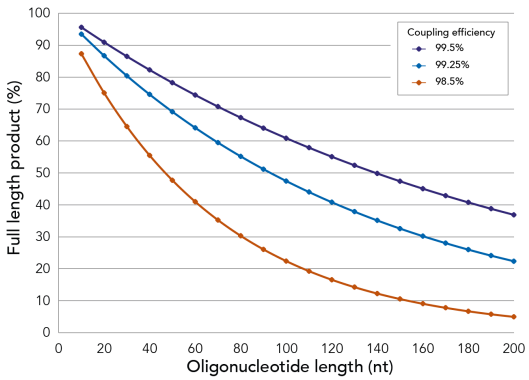


Fig. 17 The decreasing oligo product yield on increasing length by the phosphormaidite method. Data courtesy of Integrated DNA Technologies (www.idtdna.com).

Contrastingly, genomic DNA is readily available, with many sequence lengths ranging from around 500 bp for the histone gene up to 48,502 bp for Lambda DNA. However, DNA lengths that are not commercially available, would need to be designed precisely for each purpose. Similarly, within the genomic DNA, a specific sequence can be selected to encompass the desired properties (a gene code) or the reasonable frequency of the base required (GC rich regions), however, again the exact sequence cannot be easily designed. The availability of more user defined, designer DNA would allow further advances in biomolecular studies and nanomaterial assembly.

Therefore, enzymatic DNA synthesis methods could be manipulated to yield non genomic DNA. These enzymatic processes include primer extension reactions, the polymerase chain reaction (PCR) and rolling circle amplification (RCA).

Primer extensions synthesise DNA by isothermal extension of a short primer attached to a template to provide a single copy complementary to the template sequence.[131] Similarly, PCR synthesises DNA by replicating a double stranded template, commonly a genetic DNA sequence, by extending a short primer sequence through thermal cycling on both strands.[132] In this case, the amount of product is controlled by the number of cycles as each cycle duplicates the amount of DNA. RCA is similar to PCR in that it requires a template and primer, however, the template is circular ssDNA, commonly a plasmid and required isothermal extension.[133] Here the product is ssDNA consisting of a number of repetitions of the plasmid sequence depending on the length of incubation time.

Primer extensions, PCR and RCA are optimised techniques, however, they are all limited by the need for template DNA which is already commercially available, hence the sequence and length of the products cannot be easily designed outside the limitations of the sequences available. Additionally, each method permits the incorporation of enzymatically compatible modified dNTPs, however, the position of modifications cannot be absolutely controlled as the position or frequency of each base can again only be known due to DNA sequencing but not designed. A method, able to synthesise DNA where the length, sequence and functional content can be designed, would provide additional avenues for site specific nanomaterial assembly.

1.4 Aims of the project

DNA has proven itself to be a useful tool for nanomaterial assembly. The research conducted by biochemists, has provided interesting modified nucleotides with useful properties, and the DNA polymerases able to incorporate them. The DNA available on the commercial market is currently limited to genomic DNA or short oligos, therefore, DNA that is more decorated, with a more complex sequence and controllable length, would open up new avenues of DNA based nanomaterials.

Chapter 2 will discuss a new enzymatic DNA synthesis method, aiding sequence control and capable of repeat sequence extension and the selection of product length. Chapter 3 will enhance the designer nature of the enzymatic DNA synthesis method discussed, providing the user with an option to incorporate modifications into the DNA structure at precise positions. To add a second layer of design, several examples of using DNA as a scaffold will be discussed. In Chapter 4, DNA will be synthesized to encompass thiol modifications at precise positions, designed by the sequence chosen. The thiol modifications can be exploited by metal binding for nanomaterial assembly. Chapter 5 establishes the exploitation of chemical anchors, positioned on the DNA duplex by the repeating sequences. The second layer of design can be controlled by the initial loading of each modification, directing the nanomaterial assembly.

1.5 References

1 Watson, J. D. & Crick, F. H. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature* **171**, 737-738 (1953).

2 Li, J. *et al.* Self-Assembled Multivalent DNA Nanostructures for Noninvasive Intracellular Delivery of Immunostimulatory CpG Oligonucleotides. *Acs Nano* **5**, 8783-8789 (2011).

3 Lee, H. *et al.* Molecularly self-assembled nucleic acid nanoparticles for targeted in vivo siRNA delivery. *Nature Nanotechnology* **7**, 389-393 (2012).

4 Drummond, T. G., Hill, M. G. & Barton, J. K. Electrochemical DNA sensors. *Nat Biotechnol* **21**, 1192-1199 (2003).

5 Dong, L. *et al.* Synthesis, manipulation and conductivity of supramolecular polymer nanowires. *Chemistry a European Journal* **13**, 822-828 (2007).

6 Zhang, X. Y., Goux, W. J. & Manohar, S. K. Synthesis of polyaniline nanofibers by "nanofiber seeding". *Journal of the American Chemical Society* **126**, 4502-4503 (2004).

7 Venkataraman, S., Dirks, R. M., Rothemund, P. W. K., Winfree, E. & Pierce, N. A. An autonomous polymerization motor powered by DNA hybridization. *Nature Nanotechnology* **2**, 490-494 (2007).

8 Shin, J. S. & Pierce, N. A. A synthetic DNA walker for molecular transport. *Journal of the American Chemical Society* **126**, 10834-10835 (2004).

9 Seeman, N. C. An overview of structural DNA nanotechnology. *Mol Biotechnol* **37**, 246-257 (2007).

10 Feldkamp, U. & Niemeyer, C. M. Rational design of DNA nanoarchitectures. *Angew Chem Int Ed Engl* **45**, 1856-1876 (2006).

11 Cooper, V. R. *et al.* Stacking interactions and the twist of DNA. *J Am Chem Soc* **130**, 1304-1308 (2008).

12 Yakovchuk, P., Protozanova, E. & Frank-Kamenetskii, M. D. Base-stacking and base-pairing contributions into thermal stability of the DNA double helix. *Nucleic Acids Research* **34**, 564-574 (2006).

13 Kunkel, T. A. DNA replication fidelity. *Journal of Biological Chemistry* **279**, 16895-16898 (2004).

14 Loeb, K. R. & Loeb, L. A. Significance of multiple mutations in cancer. *Carcinogenesis* **21**, 379-385 (2000).

15 Robertson, K. D. & Wolffe, A. P. DNA methylation in health and disease. *Nat Rev Genet* **1**, 11-19 (2000).

16 Razin, A. & Cedar, H. DNA methylation and gene expression. *Microbiol Rev* **55**, 451-458 (1991).

17 Cooper, D. N., Mort, M., Stenson, P. D., Ball, E. V. & Chuzhanova, N. A. Methylation-mediated deamination of 5-methylcytosine appears to give rise to mutations causing human inherited disease in CpNpG trinucleotides, as well as in CpG dinucleotides. *Hum Genomics* **4**, 406-410 (2010).

18 Reuter, S., Gupta, S. C., Chaturvedi, M. M. & Aggarwal, B. B. Oxidative stress, inflammation, and cancer How are they linked? *Free Radical Bio Med* **49**, 1603-1616 (2010).

19 Oda, Y. *et al.* Nmr-Studies of a DNA Containing 8-Hydroxydeoxyguanosine. *Nucleic Acids Research* **19**, 1407-1412 (1991).

20 Kramer, B., Kramer, W. & Fritz, H. J. Different base/base mismatches are corrected with different efficiencies by the methyl-directed DNA mismatch-repair system of E. coli. *Cell* **38**, 879-887 (1984).

21 Topal, M. D. & Fresco, J. R. Complementary base pairing and the origin of substitution mutations. *Nature* **263**, 285-289 (1976).

22 Wood, R. D. Nucleotide excision repair in mammalian cells. *Journal of Biological Chemistry* **272**, 23465-23468 (1997).

23 Sugasawa, K. *et al.* A multistep damage recognition mechanism for global genomic nucleotide excision repair. *Gene Dev* **15**, 507-521 (2001).

24 Volker, M. *et al.* Sequential assembly of the nucleotide excision repair factors in vivo. *Mol Cell* **8**, 213-224 (2001).

25 Green, S. J., Bath, J. & Turberfield, A. J. Coordinated chemomechanical cycles: a mechanism for autonomous molecular motion. *Phys Rev Lett* **101**, 238101 (2008).

26 Sherman, W. B. & Seeman, N. C. A precisely controlled DNA biped walking device. *Nano Lett* **4**, 1203-1207 (2004).

27 Kiefer, J. R., Mao, C., Braman, J. C. & Beese, L. S. Visualizing DNA replication in a catalytically active Bacillus DNA polymerase crystal. *Nature* **391**, 304-307 (1998).
28 Erie, D. A., Hajiseyedjavadi, O., Young, M. C. & von Hippel, P. H. Multiple RNA polymerase conformations and GreA: control of the fidelity of transcription. *Science* **262**, 867-873 (1993).
29 Cheng, S. C., Hilton, B. D., Roman, J. M. & Dipple, A. DNA adducts from carcinogenic and noncarcinogenic enantiomers of benzo[a]pyrene dihydrodiol epoxide. *Chem Res Toxicol* **2**, 334-340 (1989).
30 Perlow, R. A. & Broyde, S. Extending the understanding of mutagenicity: Structural insights into primer-extension past a benzo[a]pyrene diol epoxide-DNA adduct. *J Mol Biol* **327**, 797-818 (2003).
31 Kim, T. W., Delaney, J. C., Essigmann, J. M. & Kool, E. T. Probing the active site tightness of DNA polymerase in subangstrom increments. *P Natl Acad Sci USA* **102**, 15803-15808 (2005).
32 Streckenbach, F., Rangam, G., Moller, H. M. & Marx, A. Steric Constraints Dependent on Nucleobase Pair Orientation Vary in Different DNA Polymerase Active Sites. *Chembiochem* **10**, 1630-1633 (2009).
33 Sale, J. E., Lehmann, A. R. & Woodgate, R. Y-family DNA polymerases and their role in tolerance of cellular DNA damage. *Nat Rev Mol Cell Biol* **13**, 141-152 (2012).
34 Biles, B. D. & Connolly, B. A. Low-fidelity Pyrococcus furiosus DNA polymerase mutants useful in error-prone PCR. *Nucleic Acids Res* **32**, e176 (2004).
35 Elshawadfy, A. M. *et al.* DNA polymerase hybrids derived from the family-B enzymes of Pyrococcus furiosus and Thermococcus kodakarensis: improving performance in the polymerase chain reaction. *Front Microbiol* **5**, 224 (2014).
36 Shevelev, I. V. & Hubscher, U. The 3' 5' exonucleases. *Nat Rev Mol Cell Biol* **3**, 364-376 (2002).
37 Neylon, C. Chemical and biochemical strategies for the randomization of protein encoding DNA sequences: library construction methods for directed evolution. *Nucleic Acids Res* **32**, 1448-1459 (2004).
38 Canceill, D., Viguera, E. & Ehrlich, S. D. Replication slippage of different DNA polymerases is inversely related to their strand displacement efficiency. *J Biol Chem* **274**, 27481-27490 (1999).
39 Albertini, A. M., Hofer, M., Calos, M. P. & Miller, J. H. On the Formation of Spontaneous Deletions - the Importance of Short Sequence Homologies in the Generation of Large Deletions. *Cell* **29**, 319-328 (1982).
40 Farabaugh, P. J., Schmeissner, U., Hofer, M. & Miller, J. H. Genetic Studies of the Lac Repressor .7. Molecular Nature of Spontaneous Hotspots in the Laci Gene of Escherichia-Coli. *J Mol Biol* **126**, 847-863 (1978).
41 Langbehn, D. R., Hayden, M. R., Paulsen, J. S. & Group, P.-H. I. o. t. H. S. CAG-repeat length and the age of onset in Huntington disease (HD): a review and validation study of statistical approaches. *Am J Med Genet B Neuropsychiatr Genet* **153B**, 397-408 (2010).
42 Kennedy, L. *et al.* Dramatic tissue-specific mutation length increases are an early molecular event in Huntington disease pathogenesis. *Hum Mol Genet* **12**, 3359-3367 (2003).
43 Hornsby, P. J. Telomerase and the aging process. *Exp Gerontol* **42**, 575-581 (2007).
44 Nozawa, K., Suzuki, M., Takemura, M. & Yoshida, S. In vitro expansion of mammalian telomere repeats by DNA polymerase alpha-primase. *Nucleic Acids Res* **28**, 3117-3124 (2000).
45 Collins, F. S., Lander, E. S., Rogers, J., Waterston, R. H. & Conso, I. H. G. S. Finishing the euchromatic sequence of the human genome. *Nature* **431**, 931-945 (2004).
46 Sanger, F., Nicklen, S. & Coulson, A. R. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A* **74**, 5463-5467 (1977).
47 McKernan, K. J. *et al.* Sequence and structural variation in a human genome uncovered by short-read, massively parallel ligation sequencing using two-base encoding. *Genome Res* **19**, 1527-1541 (2009).
48 Chang, S. *et al.* Electronic signatures of all four DNA nucleosides in a tunneling gap. *Nano Lett* **10**, 1070-1075 (2010).
49 Seeman, N. C. Nucleic acid junctions and lattices. *J Theor Biol* **99**, 237-247 (1982).
50 Stoltenberg, R. M. & Woolley, A. T. DNA-templated nanowire fabrication. *Biomed Microdevices* **6**, 105-111 (2004).
51 Watson, S. M., Wright, N. G., Horrocks, B. R. & Houlton, A. Preparation, characterization and scanned conductance microscopy studies of DNA-templated one-dimensional copper nanostructures. *Langmuir* **26**, 2068-2075 (2010).

52 Mouliere, F., El Messaoudi, S., Pang, D. L., Dritschilo, A. & Thierry, A. R. Multi-marker analysis of circulating cell-free DNA toward personalized medicine for colorectal cancer. *Mol Oncol* **8**, 927-941 (2014).
53 Marini, M. *et al.* The structure of DNA by direct imaging. *Sci Adv* **1**, e1500734 (2015).
54 Andersen, E. S. *et al.* Self-assembly of a nanoscale DNA box with a controllable lid. *Nature* **459**, 73-76 (2009).
55 Bednar, J. *et al.* Determination of DNA Persistence Length by Cryoelectron Microscopy - Separation of the Static and Dynamic Contributions to the Apparent Persistence Length of DNA. *J Mol Biol* **254**, 579-591 (1995).
56 Rothemund, P. W. K. Folding DNA to create nanoscale shapes and patterns. *Nature* **440**, 297-302, (2006).
57 Douglas, S. M. *et al.* Self-assembly of DNA into nanoscale three-dimensional shapes. *Nature* **459**, 414-418 (2009).
58 Ke, Y. G. *et al.* Scaffolded DNA Origami of a DNA Tetrahedron Molecular Container. *Nano Lett* **9**, 2445-2447 (2009).
59 Liu, J., Cao, Z. & Lu, Y. Functional nucleic acid sensors. *Chem Rev* **109**, 1948-1998 (2009).
60 Tabor, J. J. & Ellington, A. D. Playing to win at DNA computation. *Nat Biotechnol* **21**, 1013-1015 (2003).
61 Hassanien, R. *et al.* Preparation and characterization of conductive and photoluminescent DNA-templated polyindole nanowires. *Acs Nano* **4**, 2149-2159 (2010).
62 Liu, W., Zhong, H., Wang, R. & Seeman, N. C. Crystalline two-dimensional DNA-origami arrays. *Angew Chem Int Ed Engl* **50**, 264-267 (2011).
63 Murphy, C. J. *et al.* Long-range photoinduced electron transfer through a DNA helix. *Science* **262**, 1025-1029 (1993).
64 Braun, E., Eichen, Y., Sivan, U. & Ben-Yoseph, G. DNA-templated assembly and electrode attachment of a conducting silver wire. *Nature* **391**, 775-778 (1998).
65 Cohen, H., Nogues, C., Naaman, R. & Porath, D. Direct measurement of electrical transport through single DNA molecules of complex sequence. *Proc Natl Acad Sci U S A* **102**, 11589-11593 (2005).
66 Livshits, G. I. *et al.* Long-range charge transport in single G-quadruplex DNA molecules. *Nat Nanotechnol* **9**, 1040-1046 (2014).
67 Woiczikowski, P. B., Kubar, T., Gutierrez, R., Cuniberti, G. & Elstner, M. Structural stability versus conformational sampling in biomolecular systems: why is the charge transfer efficiency in G4-DNA better than in double-stranded DNA? *J Chem Phys* **133**, 035103 (2010).
68 Liu, S. P. *et al.* Direct measurement of electrical transport through G-quadruplex DNA with mechanically controllable break junction electrodes. *Angew Chem Int Ed Engl* **49**, 3313-3316 (2010).
69 Richter, J., Mertig, M., Pompe, W., Monch, I. & Schackert, H. K. Construction of highly conductive nanowires on a DNA template. *Appl Phys Lett* **78**, 536-538 (2001).
70 Monson, C. F. & Woolley, A. T. DNA-templated construction of copper nanowires. *Nano Lett* **3**, 359-363 (2003).
71 Al-Said, S. A. F. *et al.* Templating Ag on DNA/polymer hybrid nanowires: Control of the metal growth morphology using functional monomers. *Electrochem Commun* **11**, 550-553 (2009).
72 Houlton, A., Pike, A. R., Angel Galindo, M. & Horrocks, B. R. DNA-based routes to semiconducting nanomaterials. *Chem Commun (Camb)*, 1797-1806 (2009).
73 A.D. Chepelianskii *et al.* Long range electronic tranport in DNA molecules deposited across a disconnected array of metallic nanoparticles. *Comptes Rendus Physique* **13**, 967-992 (2012).
74 Ford, W. E., Harnack, O., Yasuda, A. & Wessels, J. M. Platinated DNA as precursors to templated chains of metal nanoparticles. *Adv Mater* **13**, 1793-1797 (2001).
75 Kudo, H. & Fujihira, M. DNA-Templated copper nanowire fabrication by a two-step process involving electroless metallization. *Ieee T Nanotechnol* **5**, 90-92 (2006).
76 Bigham, S. R. & Coffer, J. L. Deactivation of Q-Cds Photoluminescence through Polynucleotide Surface Binding. *J Phys Chem-Us* **96**, 10581-10584 (1992).
77 Bigham, S. R. & Coffer, J. L. The Influence of Adenine Content on the Properties of Q-Cds Clusters Stabilized by Polynucleotides. *Colloid Surface A* **95**, 211-219 (1995).
78 Dong, L. Q. *et al.* DNA-templated semiconductor nanoparticle chains and wires. *Adv Mater* **19**, 1748 (2007).
79 Sarangi, S. N., Rath, S., Goswami, K., Nozaki, S. & Sahu, S. N. DNA template driven CdSe nanowires and nanoparticles: Structure and optical properties. *Physica E* **42**, 1670-1674 (2010).

80 Artemyev, M. *et al.* Self-organized, highly luminescent CdSe nanorod-DNA complexes. *J Am Chem Soc* **126**, 10594-10597 (2004).
81 Mitomo, H., Watanabe, Y., Matsuo, Y., Niikura, K. & Ijiro, K. Enzymatic synthesis of a DNA triblock copolymer that is composed of natural and unnatural nucleotides. *Chem Asian J* **10**, 455-460 (2015).
82 Rajski, S. R. & Williams, R. M. DNA Cross-Linking Agents as Antitumor Drugs. *Chem Rev* **98**, 2723-2796 (1998).
83 Tanaka, A., Matsuo, Y., Hashimoto, Y. & Ijiro, K. Sequence-specifically platinum metal deposition on enzymatically synthesized DNA block copolymer. *Chem Commun (Camb)*, 4270-4272 (2008).
84 Nickels, P., Dittmer, W. U., Beyer, S., Kotthaus, J. P. & Simmel, F. C. Polyaniline nanowire synthesis templated by DNA. *Nanotechnology* **15**, 1524-1529 (2004).
85 Hannant, J. *et al.* Modification of DNA-templated conductive polymer nanowires via click chemistry. *Chem Commun (Camb)* **46**, 5870-5872 (2010).
86 Zang, S. Q. & Mak, T. C. Assembly of silver(I)-organic networks from flexible supramolecular synthons with pendant ethynide arms attached to a naphthyl skeleton. *Inorg Chem* **47**, 7094-7105 (2008).
87 Guo, G. C., Zhou, G. D. & Mak, T. C. W. Structural variation in novel double salts of silver acetylide with silver nitrate: Fully encapsulated acetylide dianion in different polyhedral silver cages. *Journal of the American Chemical Society* **121**, 3136-3141 (1999).
88 Galindo, M. A. *et al.* Pyrrolyl-, 2-(2-thienyl)pyrrolyl- and 2,5-bis(2-thienyl)pyrrolyl-nucleosides: synthesis, molecular and electronic structure, and redox behaviour of C5-thymidine derivatives. *Org Biomol Chem* **9**, 1555-1564 (2011).
89 Pasternack, R. F., Goldsmith, J. I., Szep, S. & Gibbs, E. J. A spectroscopic and thermodynamic study of porphyrin/DNA supramolecular assemblies. *Biophysical Journal* **75**, 1024-1031 (1998).
90 Haug, R. & Richert, C. A porphycene-DNA hybrid and its DNA-templated interactions with a porphyrin. *Journal of Porphyrins Phthalocyanines* **16**, 545 (2012).
91 Wan, Y. *et al.* A surface-initiated enzymatic polymerization strategy for electrochemical DNA sensors. *Biosens Bioelectron* **41**, 526-531 (2013).
92 Rye, H. S. *et al.* Stable fluorescent complexes of double-stranded DNA with bis-intercalating asymmetric cyanine dyes: properties and applications. *Nucleic Acids Res* **20**, 2803-2812 (1992).
93 Kanony, C., Akerman, B. & Tuite, E. Photobleaching of asymmetric cyanines used for fluorescence imaging of single DNA molecules. *J Am Chem Soc* **123**, 7985-7995 (2001).
94 Moradpour Hafshejani, S., Watson, S. M., Tuite, E. M. & Pike, A. R. Click modification of diazido acridine intercalators: a versatile route towards decorated DNA nanostructures. *Chemistry* **21**, 12611-12615 (2015).
95 Pike, A. R. *et al.* Metallocene-DNA: synthesis, molecular and electronic structure and DNA incorporation of C5-ferrocenylthymidine derivatives. *Chemistry* **8**, 2891-2899 (2002).
96 Pike, A. R. *et al.* Ferrocenyl-modified DNA: synthesis, characterization and integration with semiconductor electrodes. *Chemistry* **11**, 344-353 (2004).
97 Matray, T. J. & Kool, E. T. Selective and Stable DNA Base Pairing without Hydrogen Bonds. *J Am Chem Soc* **120**, 6191-6192 (1998).
98 Matsuda, S., Henry, A. A., Schultz, P. G. & Romesberg, F. E. The effect of minor-groove hydrogen-bond acceptors and donors on the stability and replication of four unnatural base pairs. *J Am Chem Soc* **125**, 6134-6139 (2003).
99 Yu, C., Henry, A. A., Romesberg, F. E. & Schultz, P. G. Polymerase recognition of unnatural base pairs. *Angew Chem Int Ed Engl* **41**, 3841-3844 (2002).
100 Brotschi, C., Haberli, A. & Leumann, C. J. A stable DNA duplex containing a non-hydrogen-bonding and non-shape-complementary base couple: interstrand stacking as the stability determining factor. *Angew Chem Int Ed Engl* **40**, 3012-3014 (2001).
101 Wilhelmsson, L. M., Holmen, A., Lincoln, P., Nielsen, P. E. & Norden, B. A highly fluorescent DNA base analogue that forms Watson-Crick base pairs with guanine. *J Am Chem Soc* **123**, 2434-2435 (2001).
102 Godde, F., Toulme, J. J. & Moreau, S. Benzoquinazoline derivatives as substitutes for thymine in nucleic acid complexes. Use of fluorescence emission of benzo[g]quinazoline-2,4-(1H,3H)-dione in probing duplex and triplex formation. *Biochemistry* **37**, 13765-13775 (1998).
103 Wilson, J. N. & Kool, E. T. Fluorescent DNA base replacements: Reporters and sensors for biological systems. *Org Biomol Chem* **4**, 4265-4274 (2006).

104 Preus, S. *et al.* The photoinduced transformation of fluorescent DNA base analogue tC triggers DNA melting. *Photochem Photobiol Sci* **12**, 1416-1422 (2013).
105 Coleman, R. S. & Madaras, M. L. Synthesis of a novel coumarin C-riboside as a photophysical probe of oligonucleotide dynamics. *J Org Chem* **63**, 5700-5703 (1998).
106 Saito, Y., Miyauchi, Y., Okamoto, A. & Saito, I. Base-discriminating fluorescent (BDF) nucleoside: distinction of thymine by fluorescence quenching. *Chem Commun (Camb)*, 1704-1705 (2004).
107 Wei, B., Dai, M. & Yin, P. Complex shapes self-assembled from single-stranded DNA tiles. *Nature* **485**, 623-626 (2012).
108 Seeman, N. C. & Kallenbach, N. R. Design of immobile nucleic acid junctions. *Biophys J* **44**, 201-209 (1983).
109 Goodman, R. P., Berry, R. M. & Turberfield, A. J. The single-step synthesis of a DNA tetrahedron. *Chem Commun (Camb)*, 1372-1373 (2004).
110 Holliday, R. A mechanism for gene conversion in fungi (Reprinted). *Genet Res* **89**, 285-307 (2007).
111 Chelyapov, N. *et al.* DNA triangles and self-assembled hexagonal tilings. *Journal of the American Chemical Society* **126**, 13924-13925 (2004).
112 Zhang, Y. W. & Seeman, N. C. A Solid-Support Methodology for the Construction of Geometrical Objects from DNA. *Journal of the American Chemical Society* **114**, 2656-2663 (1992).
113 Liu, D., Wang, M. S., Deng, Z. X., Walulu, R. & Mao, C. D. Tensegrity: Construction of rigid DNA triangles with flexible four-arm DNA junctions. *Journal of the American Chemical Society* **126**, 2324-2325 (2004).
114 Mao, C. D., Sun, W. Q. & Seeman, N. C. Designed two-dimensional DNA Holliday junction arrays visualized by atomic force microscopy. *Journal of the American Chemical Society* **121**, 5437-5443 (1999).
115 Woo, S. & Rothemund, P. W. K. Self-assembly of two-dimensional DNA origami lattices using cation-controlled surface diffusion (vol 5, 4889, 2014). *Nat Commun* **5** (2014).
116 Gopinath, A., Miyazono, E., Faraon, A. & Rothemund, P. W. Engineering and mapping nanocavity emission via precision placement of DNA origami. *Nature* **535**, 401-405 (2016).
117 Lin, C., Liu, Y. & Yan, H. Designer DNA Nanoarchitectures. *Biochemistry* **48**, 1663-1674 (2009).
118 Zhang, Y. W. & Seeman, N. C. Construction of a DNA-Truncated Octahedron. *Journal of the American Chemical Society* **116**, 1661-1669 (1994).
119 Banerjee, A. *et al.* Controlled release of encapsulated cargo from a DNA icosahedron using a chemical trigger. *Angew Chem Int Ed Engl* **52**, 6854-6857 (2013).
120 Jiang, Q. *et al.* DNA origami as a carrier for circumvention of drug resistance. *J Am Chem Soc* **134**, 13396-13403 (2012).
121 Goodman, R. P. *et al.* Rapid chiral assembly of rigid DNA building blocks for molecular nanofabrication. *Science* **310**, 1661-1665 (2005).
122 Erben, C. M., Goodman, R. P. & Turberfield, A. J. Single-molecule protein encapsulation in a rigid DNA cage. *Angew Chem Int Ed Engl* **45**, 7414-7417 (2006).
123 Zhang, D. Y., Hariadi, R. F., Choi, H. M. T. & Winfree, E. Integrating DNA strand-displacement circuitry with DNA tile self-assembly. *Nat Commun* **4** (2013).
124 Mathieu, F. *et al.* Six-helix bundles designed from DNA. *Nano Lett* **5**, 661-665 (2005).
125 Shih, W. M., Quispe, J. D. & Joyce, G. F. A 1.7-kilobase single-stranded DNA that folds into a nanoscale octahedron. *Nature* **427**, 618-621 (2004).
126 Geng, Y. *et al.* Electrically conductive gold- and copper-metallized DNA origami nanostructures. *Langmuir* **29**, 3482-3490 (2013).
127 Gates, E. P., Jensen, J. K., Harb, J. N. & Woolley, A. T. Optimizing gold nanoparticle seeding density on DNA origami. *Rsc Adv* **5**, 8134-8141 (2015).
128 Beaucage, S. L. & Caruthers, M. H. Deoxynucleoside Phosphoramidites - a New Class of Key Intermediates for Deoxypolynucleotide Synthesis. *Tetrahedron Lett* **22**, 1859-1862 (1981).
129 Caruthers, M. H. *et al.* Chemical Synthesis of Deoxyoligonucleotides by the Phosphoramidite Method. *Method Enzymol* **154**, 287-313 (1987).
130 Caruthers, M. H. Gene synthesis machines: DNA chemistry and its uses. *Science* **230**, 281-285 (1985).
131 Carey, M. F., Peterson, C. L. & Smale, S. T. The primer extension assay. *Cold Spring Harb Protoc* **2013**, 164-173 (2013).

132 Mullis, K. *et al.* Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. *Cold Spring Harb Symp Quant Biol* **51 Pt 1**, 263-273 (1986).
133 Ali, M. M. *et al.* Rolling circle amplification: a versatile tool for chemical biology, materials science and medicine. *Chem Soc Rev* **43**, 3324-3341 (2014).

Chapter 2.

Enzymatic Extension of Double Stranded Repeat Sequence DNA

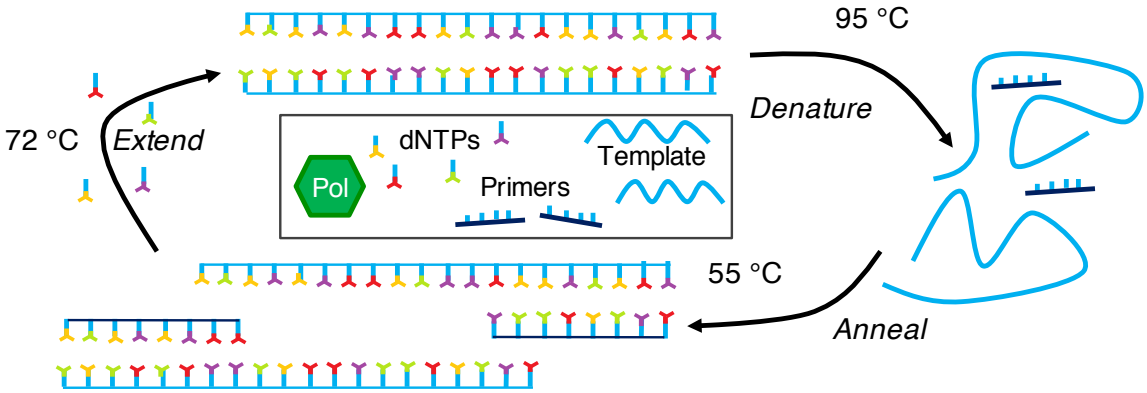
This chapter is based on the following article:

Colette J. Whitfield, Andrew T. Turley, Eimer M. Tuite, Bernard A. Connoly and Andrew R. Pike, Enzymatic Method for the Synthesis of Long DNA Sequences with Multiple Repeat Units *Angew. Chem. Int. Ed.* 2015, **54**, 8971 –8974

Table of Contents

2.1 Introduction

As introducted in Chapter 1, DNA is a highly versatile molecule and has been recognised as a suitable building block for hybrid nanomaterials.[1,2] The enzymatic synthesis of DNA has been studied since the 1950's and requires: a DNA template and primer, deoxynucleotide triphosphates (dNTPs) and a DNA polymerase.[3,4] Historically, radioactive analytical methods were used to determine DNA composition and reliability of enzymatic DNA synthesis outside *in vivo* conditions.[4] Initially, primer extension assays were commonly employed to investigate the enzymatic DNA synthesis.[5] Primer extension reactions involve the isothermal incorporation of complementary dNTPs against a single stranded template, the DNA polymerase extends a primer to produce one copied DNA strand per template strand. Further developments led to the PCR - originally established in 1983 to amplify DNA sequences[6] – mainly for protein preparation and purification by target gene amplification. PCR involves thermal cycling, consisting of repeated heating and cooling of dsDNA to allow denaturing and annealing, followed by extension of the overhangs by a DNA polymerase, Scheme 1 and Fig. 1 (a).



Scheme 1 PCR scheme depicting denaturing of the template DNA, followed by annealing of each primer to the complementary region on the template DNA and extension of the overhangs by a DNA polymerase (green) through insertion of dNTPs. Each cycle doubles the copy number of the DNA template.

In 1976, a thermostable DNA polymerase, from *Thermococus aquaticus* was isolated,[7] allowing continual replication without denaturing the DNA polymerase during PCR.[8] Currently, there are many commercially available thermostable DNA polymerases, each possessing a specific attribute for either high or low fidelity replication.[9] Across species, the DNA polymerase structure varies greatly,[10,11] placing each polymerase into a distinct family.[12] Nevertheless, all DNA polymerases consist of 3 common subdomains; the palm, thumb and fingers subdomains, Fig. 1 (b).[13]

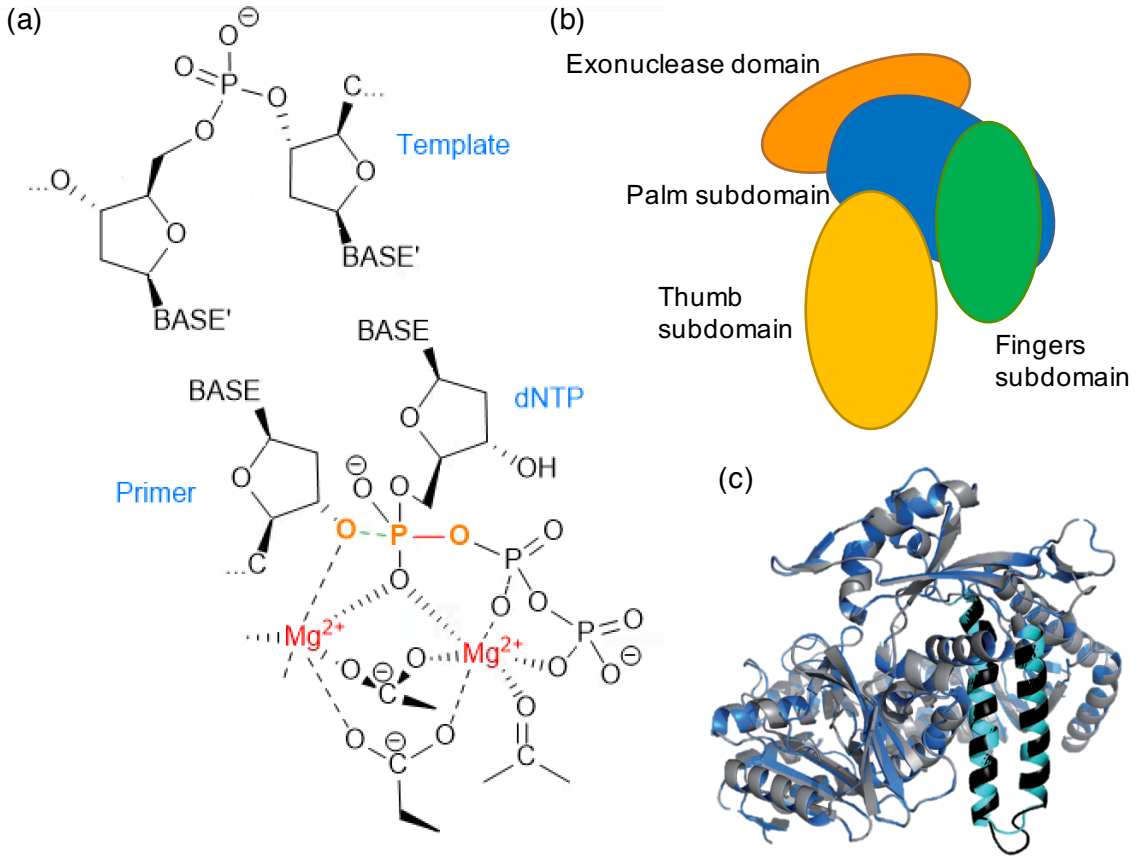
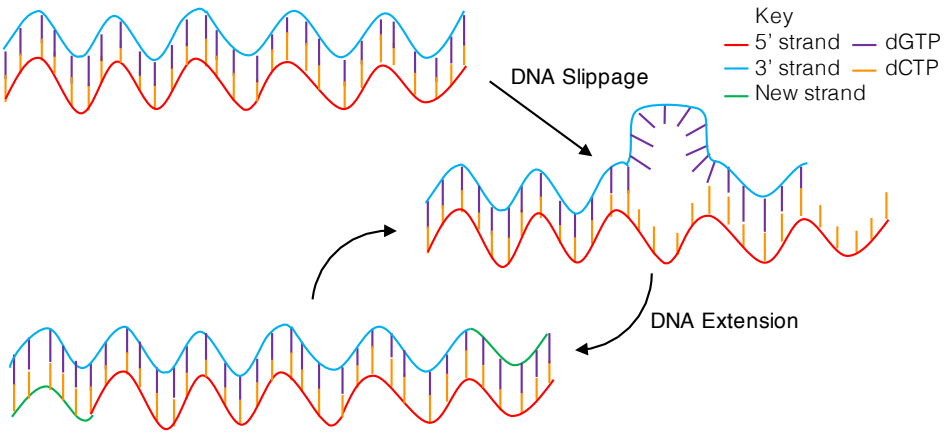


Fig. 1 (a) Transition state of DNA-dNTP-*E. coli* DNA polymerase 1 complex (based on a Fig. from Brautigam and Steitz, 1998).[14] The two *E.coli* DNA polymerase 1 conserved aspartates are shown. Metal ions are used to stabilise the reaction components; one to activate the attaching 3' oxygen from the primer and the other to stabilise the di-phosphate. The new bond formed is shown in green. The interaction broken is shown in red. The template and primer strands are already either polymers or short oligomers. **(b)** DNA polymerase diagram. Each domain is shown as a simple circular form in approximate positions and size of the known structure.[15] The palm, thumb and fingers make up the polymerase domain. **(c)** Overlay crystal structure of Tgo-Pol Z3 exo-.[16] The fingers domain is high-lighted in teal and black. The extension loop can be seen joining the two teal/black helices.

The two polymerase families relevant to this thesis are Family A and Family B. In viruses, Family A and B polymerases are used for genome replication, as opposed to only Family B polymerases for eukaryotes and archaea.[17] Many of the replicative polymerases encompass an exonuclease domain capable of 3'-5' proof-reading, which improves fidelity[10] – a desirable function *in vivo* to ensure correct copying of the genome. Archael Family B DNA polymerases are common research tools for PCR due to their thermostable and high fidelity nature.[18] At the high temperature habitat of Archaea, common mutations such as cytosine deamination[19,20] occur in a higher abundance. Therefore, Archael DNA polymerases have additional structural adaptations, such as a read-ahead function,[21] to counteract the increased mutations (error rate per base pair is 1-20 $x10^{-5}$ for the *Thermus aquaticus* DNA polymerase,

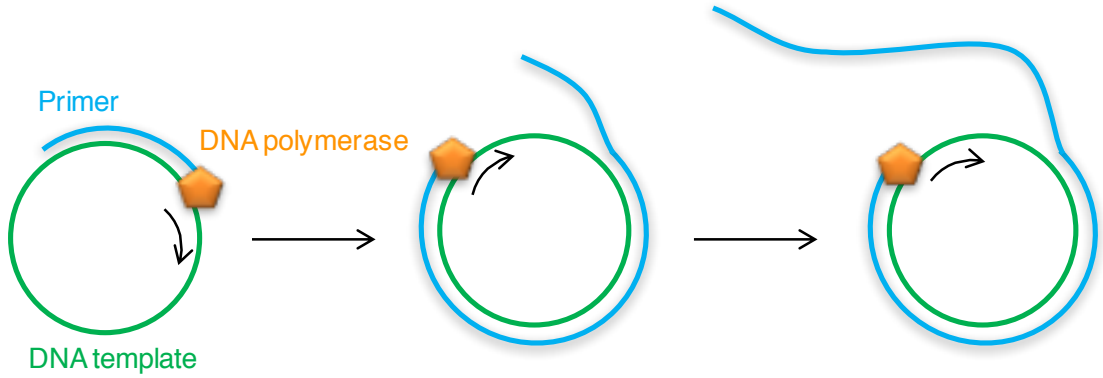
Taq, and 1-2 x 10^{-6} for the Archael *Pyrococcus furiosus* Family B DNA polymerase (Pfu-Pol)).[22,23] In contrast, many research questions necessitate the use of mutated DNA, hence require a more promiscuous DNA polymerase to incorporate a mis-pair and replicate against the mutation. DNA polymerases disabled in the the exonuclease domain have become a popular choice for conventional PCR. PCR protocols are now highly optimised for each bioscience application. However, conventional PCR only replicates the sequence present, and does not extend the DNA.

Until recently, DNA slippage has been a key synthetic approcach for the production of long DNA, Scheme 2. The DNA slippage reaction occurs *in vivo* due to the instability of tandem repeat sequences in the genome – commonly involved in neurodegenerative diseases.[24,25] Initial work focused on naturally occurring trinucleotide repeat sequences, showing extension up to 700 bp.[26] DNA slippage has subsequently been exploited as a technique for long DNA synthesis with a variety of nucleotide sequences. The final DNA sequence is an elongated product of the starting sequence. An examination of DNA polymerases showed that the *E. coli* DNA polymerase I Klenow exonuclease minus fragment was the most proficient in DNA extension, extending $[G]_n/[C]_n$ up to 2,000 bp at 37 °C within 60 minutes.[27,28] Though DNA slippage is highly efficient at producing long DNA of repeating sequences – the breadth of repeat sequences able to slip is limited. Hence, the need for an improvement in long repeat sequence DNA synthesis.



Scheme 2 dsDNA slippage. A hairpin forms on one strand, causing that strand to shift. This hairpin can then 'flatten' and extend the strand past the lower strand. This results in two 'sticky ends' which the DNA polymerase can fill in with new nucleotides.

Another example of how DNA polymerases can be expoited as biological tools is in the production of ssDNA. This has been demonstrated by RCA.[29,30] RCA permits the isothermal production of long ssDNA by the replication of circular or plasmid DNA, displacing the single strand once it has been extended, Scheme 3. RCA is highly efficient at producing long DNA of interest, however, the length of the repeating sequence is limited by the size of the plasmid and the sequence itself is limited to the commercially or biologically available DNA templates.



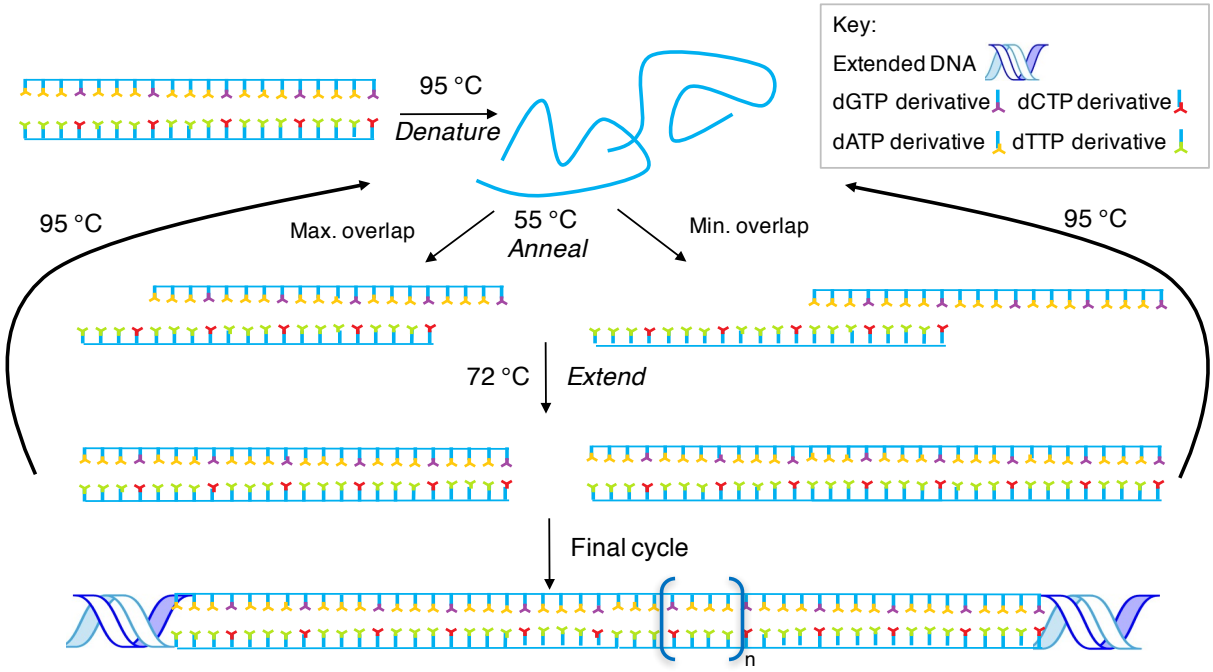
Scheme 3 RCA. Extension of the primer by the DNA polymerase is shown followed by the strand displacement and continued extension.

Several attempts to enzymatically replicate DNA on a large scale using heat-cool cycles have been published.[31–33] Two methods showed the amplification of trinucleotide repeat sequences, similar to slippage, to study neurodegenerative related sequences.[32,33] The polymerase endonuclease amplification reaction (PEAR), developed in 2010 by Xu *et al.,*[31] was established to synthesise antisense oligonucleotides. As the replication in each study involved an oligonucleotide consisting of two repeating sequences, a side reaction of DNA extension was observed.

These methods are applicable for the synthesis of biologically relevant dsDNA, however, DNA length and content is again limited by the source of templates. Long dsDNA of designer sequence and length would be a highly desirable addition to the nanomaterials toolbox. Here, we have developed a DNA extension method, able to extend many repeating units of DNA to a user defined length.

2.2 Results and discussion

In this chapter, a PCR-based method for the synthesis of designer DNA is described - the heat-cool cycles of PCR, dsDNA consisting of a repeating sequence and the standard dNTPs are required. The dsDNA, acts as both primer and template and is now refered to as an “oligo seed”. These oligo seeds encode the repeat element to be extended and must contain a minimum of two repeat units. Sufficient base overlap between units to form a stable sticky ended duplex is also a requirement. In principle, if the strands are cooled quickly, they will not attain the lowest energy structure – completely aligned double strands. This would result in directing the assembly of ‘sticky ends’, capable of extension by a DNA polymerase. In most instances, there are more than two possible conformations in which the strands can anneal, shown in Scheme 4. For each potential conformation, the length of the overhang will vary and therefore the DNA polymerase will extend the duplex by a different number of bases which yields a range of DNA products, Table 1. The difference in length between the smallest and largest DNA products can be considerable. The theoretical maximum and minimum lengths can be calculated based on the length of the repeat sequence, the length of the oligo seed, the number of cycles and is also limited by the behaviour of the DNA polymerase.



Scheme 4 A PCR-based heat-cool cycle model. DNA melting is performed at 95 °C to form two single strands. The annealing of complementary regions with the minimum (shown left) and maximum (shown right) extension is at 55 °C. The temperature is then increased to 72 °C at which the DNA polymerase binds complementary nucleotides to the ‘sticky ends’ of the stands. This process can be repeated for the desired number of cycles.

As the method of heat cool cycles demands a thermostable DNA polymerase – a *Thermococcus gorgonarius* family B DNA polymerase (Tgo-Pol) mutant, Z3, was used.[16] Tgo-Pol Z3 exo- is disabled in the 3'-5' proof-reading exonuclease region and has an insertion in the fingers domain, Fig. 1 (c) - mutations that confer the ability to read non-standard bases in the template strand[17] – a desirable attribute for future experiments involving modified dNTP incorporation.

2.2.1 *Heat-cool cycle extension development*

The most common repeat sequence used for trialling extension techniques is a tri-nucleotide repeat unit – regularly seen in biological examples of slippage.[24,25] With this in mind, the initial sequence chosen was a tetra-nucleotide repeat unit to broaden the length of repeat units currently explored. Other oligo seeds that were later explored included $[AG]_{10}/[TC]_{10}$ (b = 2), $[A_2G]_7/[T_2C]_7$ (b = 3), $[A_3G]_5/[T_3C]_5$ (b = 4). $[A_4G]_4/[T_4C]_4$ (b = 5) and $[A_9G]_2/[T_9C]_2$ where b = the number of bases in the repeating unit. The maximum and minimum lengths of the expected products from the above oligo seeds were calculated and are shown in Table 1.

No. of cycles	$[AG]_{10}/[TC]_{10}$ (b = 2)		$[A_2G]_7/[T_2C]_7$ (b = 3)		$[A_3G]_5/[T_3C]_5$ (b = 4)		$[A_4G]_4/[T_4C]_4$ (b = 5)		$[A_9G]_2/[T_9C]_2$ (b = 10)	
	Min[a](bp)	Max[b](bp)	Min[a](bp)	Max[b](bp)	Min[a](bp)	Max[b](bp)	Min[a](bp)	Max[b](bp)	Min[a](bp)	Max[b](bp)
0	20	20	21	21	20	20	20	20	20	20
5	30	392	36	393	40	392	45	330	70	330
10	40	5746	51	5713	60	5746	70	5650	120	5650
15	50	10746	66	10713	80	10746	95	10560	170	10560
20	60	15746	81	15713	100	15746	120	15560	220	15560

Table 1: Theoretical minimum and maximum DNA extension products by repeat unit length (b = 2-10). The algorithm used to determine the theoretical minimum and maximum product lengths assumes that 1) at least eight bp must exist to give a duplex stable enough for elongation, that is, the minimum stable overlap is eight for $[AG]_{10}/[TC]_{10}$ and $[A_3G]_5/[T_3C]_5$, nine for $[A_2G]_7/[T_2C]_7$, and ten for $[A_4G]_4/[T_4C]_4$ and $[A_9G]_2/[T_9C]_2$ and that 2) the polymerase can fill in up to 500 bases per 30 seconds extension time.[34] [a] Minimum theoretical length (bp) = original length+(y x b). [b] Maximum theoretical length (bp) = original length x 2^y-[(min. stable overlap x 2^y)-min. stable overlap] , where b = number of bp in the repeat unit and y = number of cycles.

The first extension assay was performed with $[A_3G]_n/[T_3C]_n$, Fig. 2, and clearly illustrates the increase in length of DNA as the number of heat-cool cycles increases. As described in Scheme 4 and Table 1, a distribution of lengths is produced. The lane absorbance intensity shows normal distribution curves for the range of lengths produced which was anticipated, Fig. 2 (b). After 10 cycles the duplex had extended to a modal length of 700 bp. This shows the method is capable of extending a 20 base oligo seed sequence. To determine if longer DNA lengths are achievable, additional heat-cool cycles were investigated.

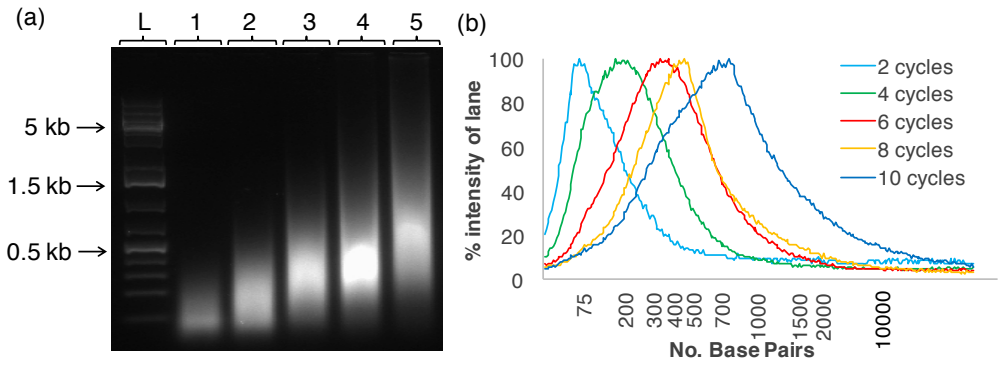


Fig. 2 Heat-cool cycle extension with Tgo-Pol Z3 exo- and $[A_3G]_5/[T_3C]_5$ for 2, 4, 6, 8, and 10 cycles. **(a)** Agarose gel: lanes 1-5: extension products after 2, 4, 6, 8 and 10 cycles, respectively. **(b)** Image J analysis of (a). Analysis shows the intensity of each band as a percentage of the highest intensity per lane, L = DNA ladder.

The enzymatic extension of $[A_3G]_5/[T_3C]_5$ from 5 to 20 cycles expressed the continual extension to a modal length of 1,500 bp, Fig. 3 (a). The distribution of lengths (from 400 to 10,000 bp) was consistent with Fig. 3 and within the extreme limits outlined in Table 1 (from 100 to 15,746 bp).

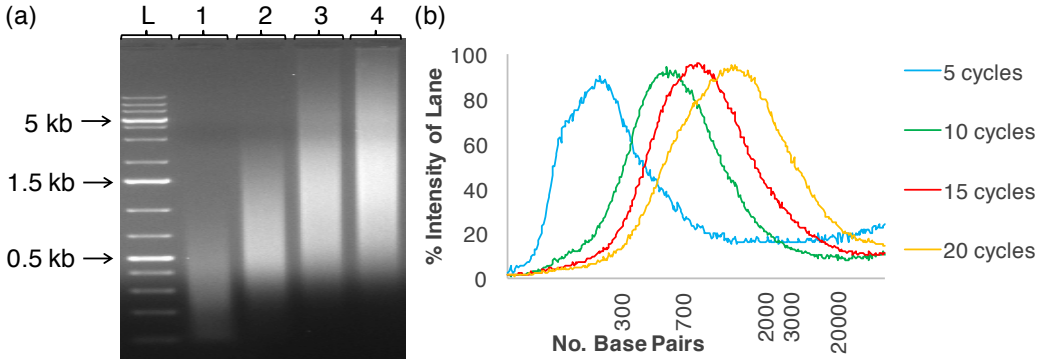


Fig. 3 Heat-cool cycle extension with Tgo-Pol Z3 exo- and $[A_3G]_5/[T_3C]_5$ for 5, 10, 15 and 20 cycles. **(a)** agarose gel: lanes 1-4: extension products after 5, 10, 15 and 20 cycles, respectively. **(b)** Image J analysis of (a). Analysis shows intensity of band as a % of the highest intensity per lane, L=DNA ladder.

To investigate whether continual DNA extension is possible after 20 cycles, the number of cycles was increased further. After 20 cycles, the rate of DNA extension tailors off, Fig. 4 and 5, suggesting a reduction in DNA polymerase activity. Previous unpublished results have shown Tgo-Pol Z3 exo- to denature after 30 cycles, which would explain the reduction in activity (C. J. Whitfield MRes Dissertation, 2013). To determine if the reduced extension is due to enzyme activity additional DNA polymerase was added during the heating and cooling cycles.

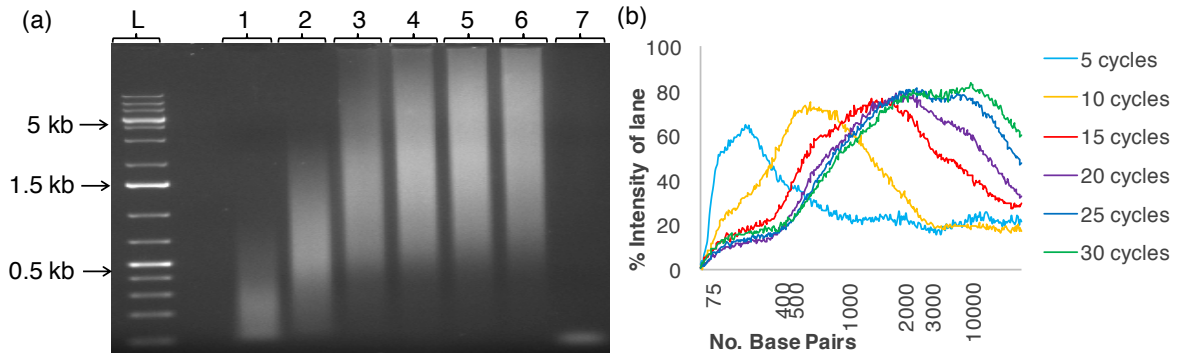


Fig. 4 Heat-cool cycle extension with Tgo-Pol Z3 exo- and $[A_3G]_5/[T_3C]_5$ for 5, 10, 15, 20, 25 and 30 cycles. **(a)** agarose gel: lanes 1-6: DNA extension product after 5, 10, 15, 20, 25 and 30 cycles, respectively, lane 7: no polymerase control. **(b)** Image J analysis of (a). Analysis shows intensity of band as a % of the highest intensity per lane, L=DNA ladder.

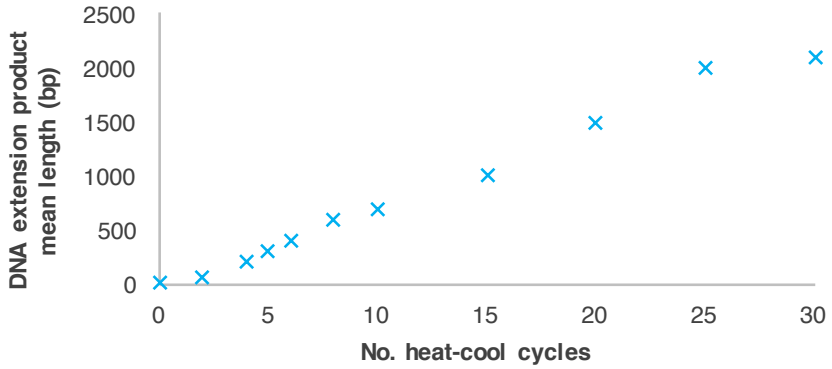


Fig. 5 The DNA extension trend of the heat-cool cycles. An extension rate of 72 bp per cycle is calculated for extension up to 25 cycles. After 25 cycles the extension tailors off.

To examine whether the stall in extension is due to Tgo-Pol Z3 exo- a comparison of the heat-cool cycles with and without the addition of fresh DNA polymerase was performed, Fig. 6. A difference can be seen between lanes 4 and 8, showing longer extension products after addition of fresh DNA polymerase. This coincides with the polymerase denaturation expected after 30 heat cool cycles, hence the continuous extension up to 40 cycles is possible if fresh DNA polymerase is added. If longer lengths of DNA are required, this modified extension method could be adopted.

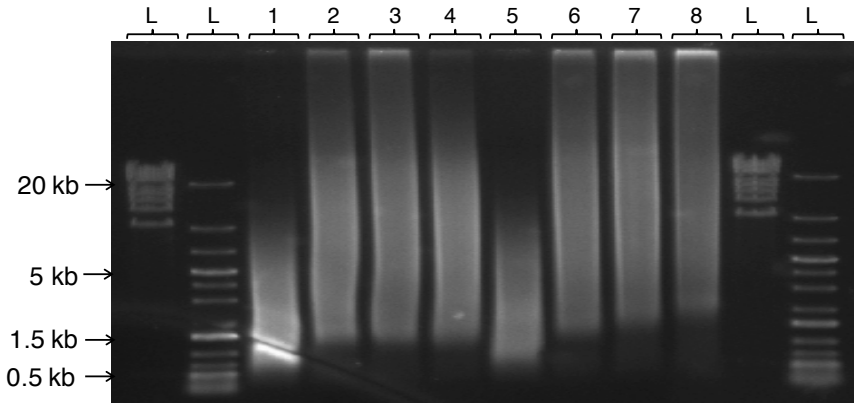


Fig. 6 Heat-cool cycle extension of $[A_3G]_n/[T_3C]_n$,with and without extra addition of 2 μL containing 1.0 μM dNTP and 2.0 μM Tgo-Pol Z3 exo- every 10 cycles (2 μL H_2O was added to no addition controls). Lanes 1-5: 10, 20, 30 and 40 cycles with no addition, lanes 5-8: 10, 20, 30 and 40 cycles with addition, L = DNA ladder.

Due to the nature of the repetitive sequence within the oligo seeds, it is conceivable that larger structures such as concatamers could assemble. Therefore, a DNA extension assay was performed in the absence of a DNA polymerase to investigate whether such larger structures were produced. As can be seen from the agarose gel data, after 10 heat-cool cycles, the oligo seed $[A_3G]_5/[T_3C]_5$ does not self assemble into a concatamer structure, Fig. 7.

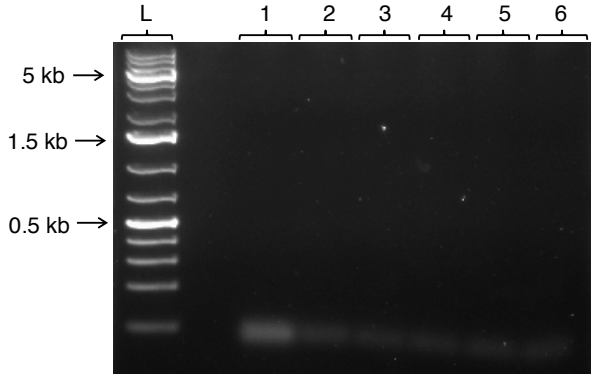


Fig. 7 Agarose gel of a no polymerase control extension assay with $[A_3G]_5/[T_3C]_5$. Lane 1: 0 cycles, lane 2: 2 cycles, lane 3: 4 cycles, lane 4: 6 cycles, lane 5: 8 cycles, lane 6: 10 cycles, L = DNA ladder.

2.2.2 *DNA Gel extraction*

Access to a narrower size distribution of DNA would be beneficial in the design of precisely controlled nanomaterials, therefore, agarose gel extraction was explored. The Lonza FlashGel™ DNA system allows for the simple removal of DNA from a second set of extraction wells, Fig. 8 (a). The extracted samples, A-C, were analysed on a second gel and show much tighter bands of reduced size distribution, Fig. 8 (b).

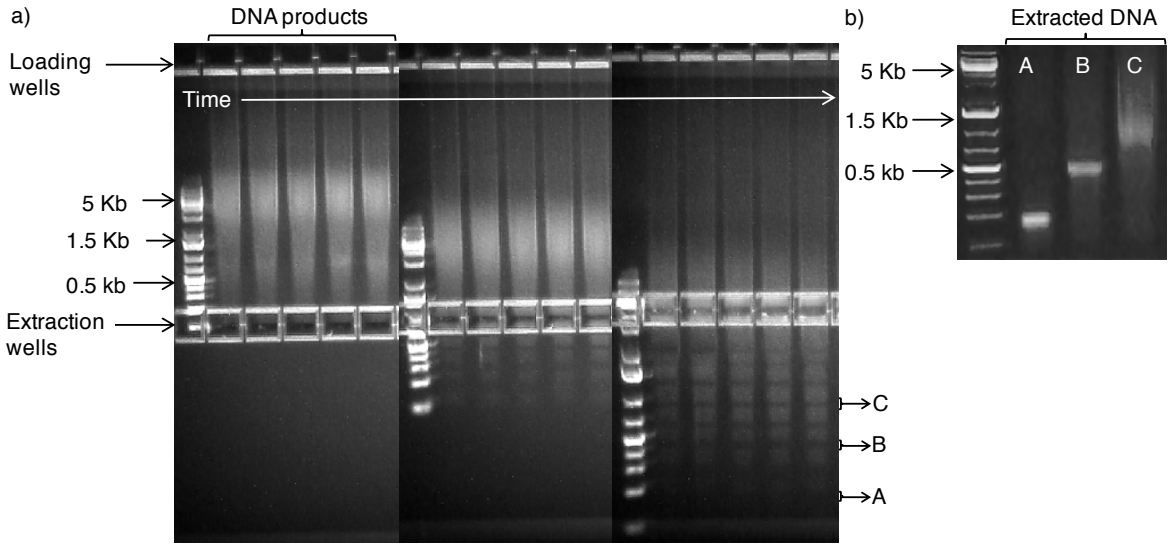


Fig. 8 (a) Lonza FlashGel™ DNA system of DNA product after 20 heat-cool cycles with $[A_3G]_n/[T_3C]_n$. Sample was removed from the extraction wells and length was determined by comparison to the DNA ladder. **(b)** Lonza FlashGel™ of extracted DNA products A-C.

The concentration of each extracted band is consistent with the Image J analysis intensity of the smeared band before size recovey. A comparison between the DNA concentration as ascertained by UV-Vis, Fig. 9 (a) and the Image J analysis data, Fig. 9 (b), shows good correlation. Due to low extraction yields from the double tier Lonza FlashGel™ system, conventional agarose gel extraction was adopted to give DNA samples at 500, 1,500 and 7,000 bp for further characterisation.

(a)

Sample	Conc (ng/μL)
500	9.3
700	24.6
1000	33.5
1500	33.6
2000-3000	11.1

(b)

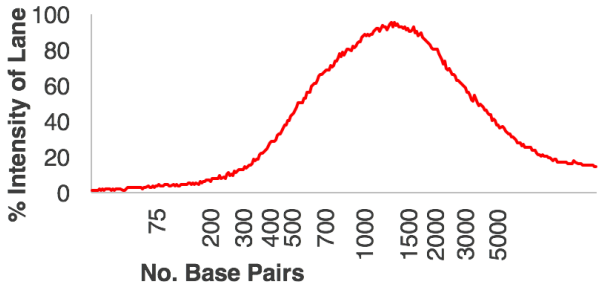


Fig. 9 (a) A table of DNA concentrations of samples extracted from the Lonza gel, Fig. 9. **(b)** Image J analysis of DNA extension products after 20 heat-cool cycles on an agarose gel.

AFM was performed with samples excised at 500 and 1,500 bp, Fig. 10 (a) and (b). The AFM images show the mean height of both DNA samples to be 0.8 ± 0.2 nm and 0.7 ± 0.2 nm, which is in agreement with previously reported DNA dimensions.[35] Also, in total, 10 length measurements were performed per AFM image of each size of DNA, (see Appendix A) which gave mean observed lengths of 180 ± 18 nm, (*ca.* 0.5 kb) and 420 ± 55 nm, (*ca.* 1.2 kb) for the 500 and 1,500 bp samples respectively.

To visualise the excised DNA by fluorescence microscopy, YOYO-1, a fluorescent intercalator, was added. The fluorescence micrographs of DNA excised at 7,000 bp, Fig. 10 (c), outlines average DNA lengths of 3.3 μm ± 0.5 μm, (*ca.* 9.7 kb ± 1.6 kb) – slightly longer than expected. However, in similar studies, other YOYO-1-labelled samples also appeared to be up to 10 % longer than expected due to limitations of the fluorescence microscope optics.[36] This extraction process outlines a simple method for user defined DNA production of a specific length.

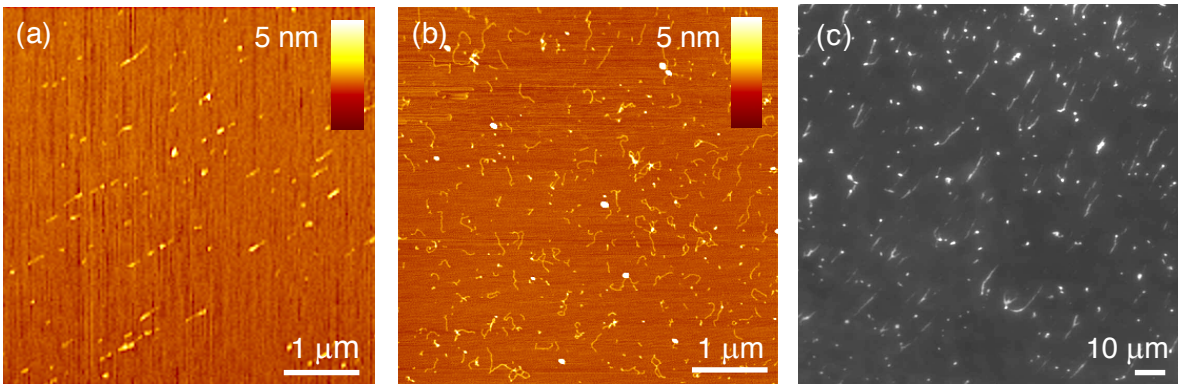


Fig. 10 Visual analysis of DNA extension products selected by length through gel excision **(a)** AFM of 500 bp repeat sequence DNA, **(b)** AFM of 1,500 bp repeat sequence DNA and **(c)** fluorescence micrograph of 7,000 bp repeat sequence DNA with YOYO-1.

2.2.3 *Heat-cool cycle extension with repeating units of length b, where b = 1-10*

To demonstrate the flexibility of this heat-cool cycle method to produce a wider range of DNA products, various repeating units were trialled in extension assays. It is worth nothing that in previous work (C. Whitfield MRes Dissertation, 2012), $[A]_n/[T]_n$ (b = 1), was extended and produced DNA with a modal length of 5 kb. Repeating units were selected to encompass units of length b, where b = 2, 3, 4, 5, and 10 bases. $[AG]_{10}/[TC]_{10}$ (b = 2), $[A_2G]_7/[T_2C]_7$ (b = 3), $[A_3G]_5/[T_3C]_5$ (b = 4) and $[A_4G]_4/[T_4C]_4$ (b = 5) all extended to a modal length of 5 kb after 20 cycles, Fig. 11 (a). As expected from the extension length model in Table 1, a distribution of product lengths is produced. The extension of the above sequences over 20 heat-cool cycles were replicated several times and in each case similar DNA product lengths were observed, indicating the reproducibility of this enzymatic DNA extension method, Fig. 11 (b).

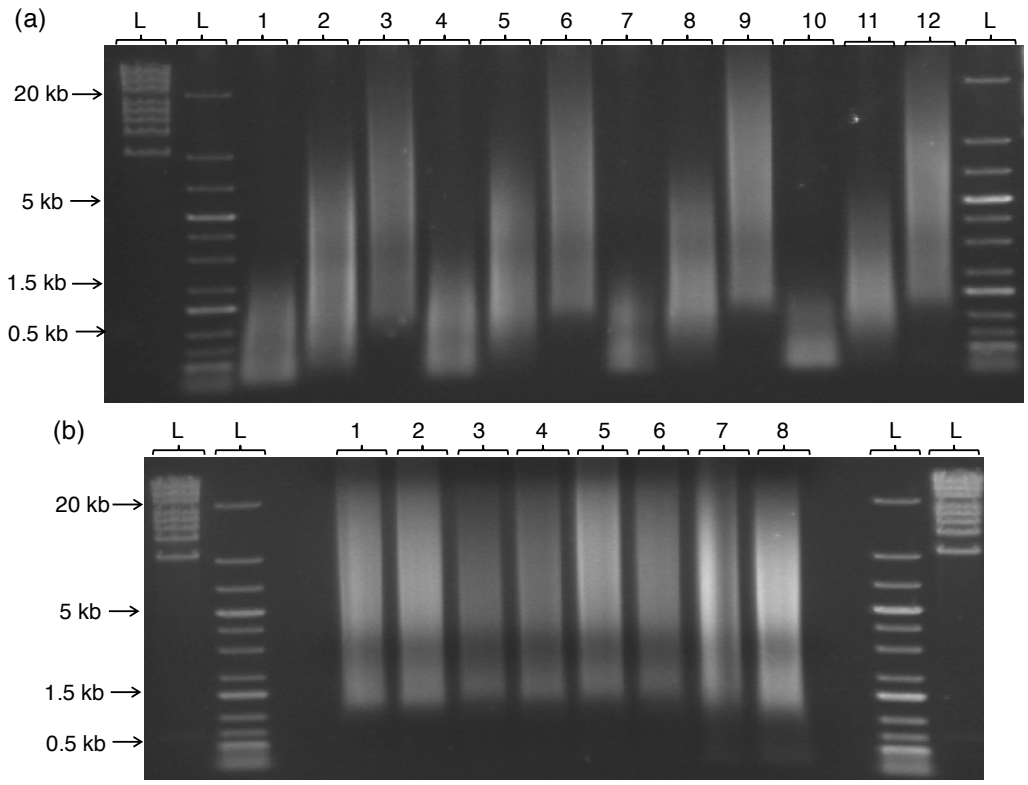


Fig. 11 Agarose gel of **(a)** extension products after 5, 10 and 20 cycles, lanes 1-3: $[AG]_n/[TC]_n$ extension for 5, 10 and 20 cycles respectively, lanes 4-6: $[A_2G]_n/[T_2C]_n$ extension for 5, 10 and 20 cycles respectively, lanes 7-9: $[A_3G]_n/[T_3C]_n$ extension for 5, 10 and 20 cycles respectively, lanes 10-12: $[A_4G]_n/[T_4C]_n$ extension for 5, 10 and 20 cycles respectively. **(b)** extension products after 20 cycles, lanes 1 and 2: $[AG]_n/[TC]_n$ extension, lanes 3 and 4: $[A_2G]_n/[T_2C]_n$ extension, lanes 5 and 6: $[A_3G]_n/[T_3C]_n$ extension, lanes 7 and 8: $[A_4G]_n/[T_4C]_n$ extension, L = DNA ladder.

The extension assay of the 10 base pair repeat sequence, using the oligo seed, $[A_9G]_2/[T_9C]_2$, (b = 10), Fig. 12, indicated that the extension was less efficient than with the smaller repeating units, Fig. 11. The initial probability of a shifted annealing when b = 10 is lower for the 20 bp oligo seed, (only 50 % - either 20 bp complete overlap or 10 bp shifted overlap) than with the smaller repeating units, (67 % when b = 5, 75 % when b = 4, 80 % when b = 3, 86 % when b = 2,). However, after the initial 10 cycles, the extension appears to follow the same trend as the other sequences considered, Fig. 12. The ability to extend repeats of the sequence type $[A_xG]_n/[T_xC]_n$, where x = 1 to 9, affords control over the G:C base pair density, that is, the ability to position individual G:C pairs at defined distances from each other.

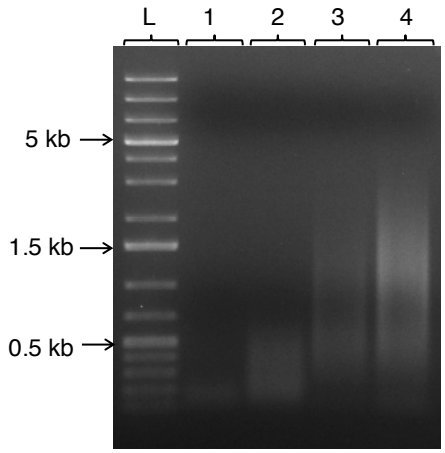


Fig. 12 Agarose gel of the DNA extension products consisting of $[A_9G]_n/[T_9C]_n$, lanes 1-4: extension products after, 5, 10, 15 and 20 cycles, respectively. L = DNA ladder.

Thermal cycling is also possible with sequences containing all four bases - $[GATC]_5/[CTAG]_5$, (b = 4) and $[ACTGATCAGC]_2$/ $[TGACTAGTCG]_2$ (b = 10), Fig. 13 (a) and (b). This validates the method as an efficient and reliable route to fabricate long DNA with designer sequences.

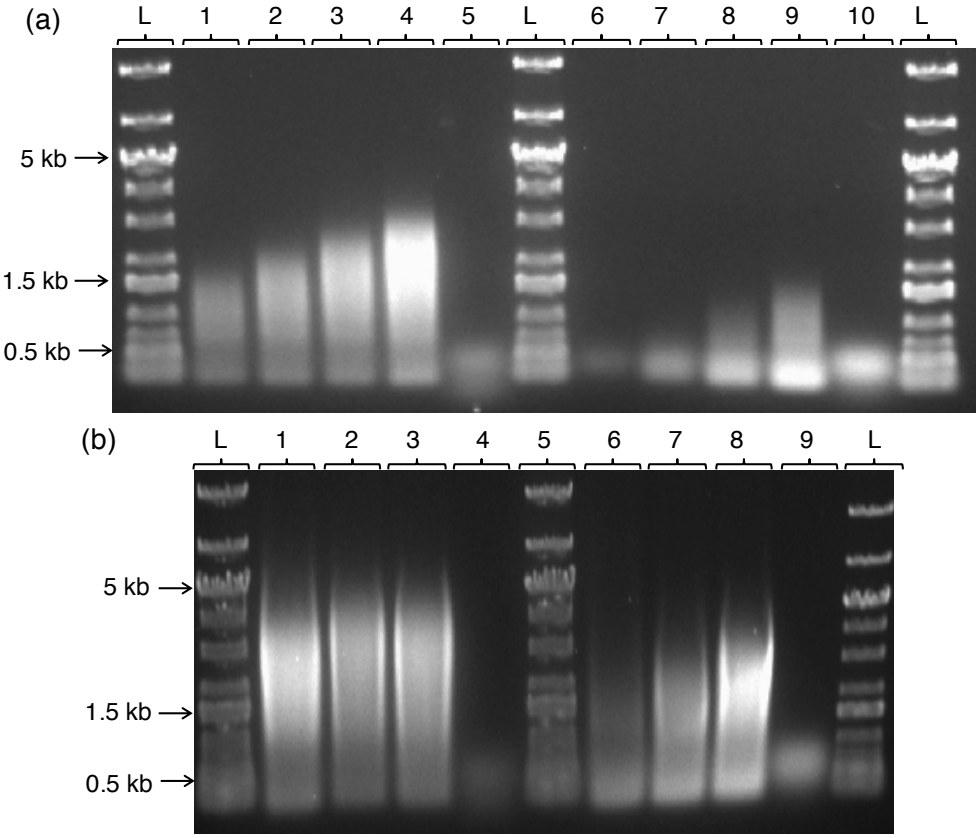


Fig. 13 Agarose gels of **(a)** lanes 1-5: DNA products from $[GATC]_5/[CATG]_5$ after 5, 10, 15, 20 and 0 cycles respectively, lanes 6-10: DNA extension products from $[ACTGATCAGC]_2/[TGACTAGTCG]_2$ after 5, 10, 15, 20 and 0 cycles respectively, **(b)** lanes 1-4: DNA extension products from $[GATC]_5/[CATG]_5$ after 20, 40, 60 and 0 cycles respectively, lanes 6-9: DNA extension products from $[ACTGATCAGC]_2/[TGACTAGTCG]_2$ after 20, 40, 60 and 0 cycles respectively, L = DNA ladder.

DNA Sanger sequencing was performed by GATC-biotech (Cologne, Germany), for both the $[A_4G]_n$ strand and the $[T_4C]_n$ strand, yielding a sequence length of 307 and 340 nucleotides respectively, Fig. 14 (a) and (b). The DNA sequencing shows correct reading and incorporation of each base by the DNA polymerase, thus providing evidence for high fidelity of the DNA polymerase to maintain the Watson and Crick base pairing rules.

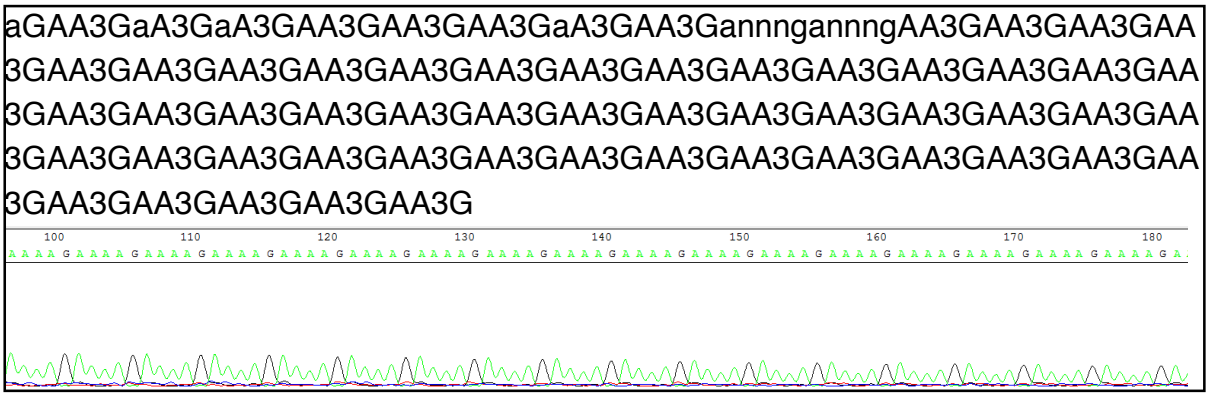


Fig. 14 (a) Sanger sequencing results of the $(A_4G)_n$ strand of the DNA extension product.

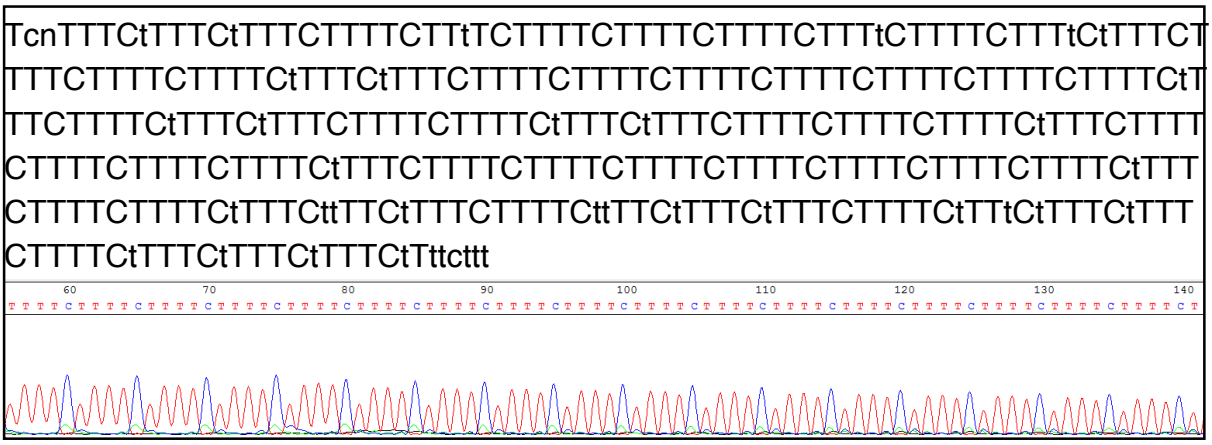


Fig. 14 (b) Sanger sequencing results of the $(T_4C)_n$ strand of the DNA extension product.

DNA melting was performed using pico-green (a DNA intercalator) with 2 μM dsDNA (both oligo seed and DNA extension products), monitoring the fluorescence emission while heating from 45 to 95 °C, Fig. 15. As expected, the DNA extension products after 30 heat-cool cycles have a higher melting temperature (T_m) than the original oligo seeds. A comparison between the control oligo duplex and the DNA samples shows the initial gradual decrease in fluorescence intensity is expected for the DNA melting profiles. All the samples follow a similar trend to the control sample, except the $[A_9G]_2/[T_9C]_2$ oligo seed which shows a small melting curve, suggesting the oligo seed may not form a completely annealed duplex under these conditions.

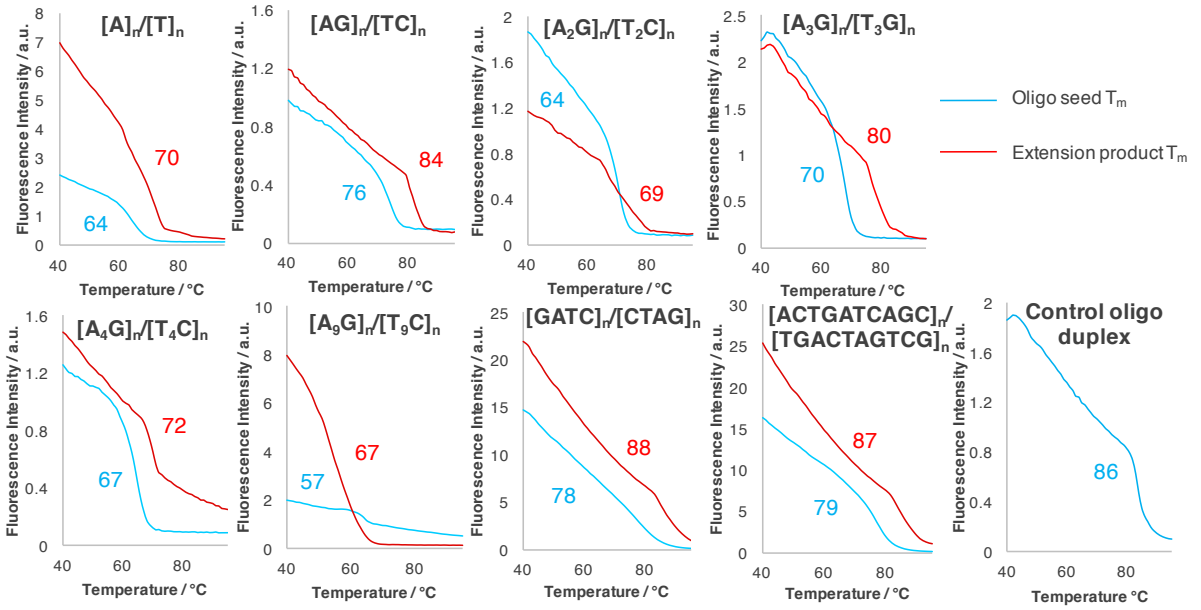


Fig. 15 T_m analysis of the oligo seeds, the extension products synthesised after 30 heat-cool cycles and a commercial DNA duplex of known melting point.

2.2.4 *Selection of DNA polymerase*

Tgo-Pol Z3 exo- was used throughout the experiments in this chapter. Tgo-Pol Z3 exo- is a thermostable DNA polymerase with reduced fidelity as it is disabled in the proof reading exonuclease region.[16] Additional mutations by the insertion of 3 amino acids into the fingers domain adds to Z3's ability to incorporate non-standard dNTPs.[16] However, for extension with the four standard bases, a standard exonuclease minus DNA polymerase variant is adequate. This was demonstrated by comparing the extension efficiency of the Pfu-Pol exo- and the variant Tgo-Pol Z3 exo-. There was a slight increase in extension efficiency of Pfu-Pol exo- in comparison to Tgo-Pol Z3 exo-, Fig. 16 (a). However, as Tgo-Pol Z3 exo- is more accommodating of modified nucleotides, it may be a more apt reagent for future applications in producing DNA as a customisable nanomaterial. The importance of the exonuclease activity in this method was investigated by a comparison with Pfu-Pol exo+. The exonuclease containing DNA polymerase is unable to extend DNA by this method of extension and appears to degrade the starting oligo seed, Fig. 16 (b). Therefore, exo- DNA polymerase variants are required to extend DNA by this method.

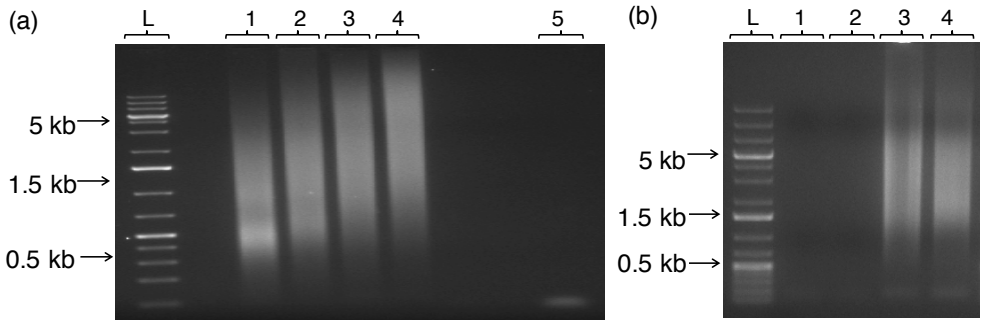


Fig. 16 (a) Agarose gel comparing extension efficiency between Tgo-Pol Z3 exo- and Pfu-Pol exo-. Lane 1: Tgo-Pol-Z3 exo- 7 cycles, lane 2: Tgo-Pol-Z3 exo- 10 cycles, lane 3: Pfu-Pol exo- 7 cycles, lane 4: Pfu-Pol exo- 10 cycles, lane 6: oligo seed. **(b)** Agarose gel comparing extension efficiency between Tgo-Pol Z3 (exo-) and Pfu-Pol exo+. Lane 1: Pfu-Pol (exo+) 5 cycles, lane 2: Pfu-Pol (exo+) 10 cycles, lane 3: Tgo-Pol Z3 (exo-) 5 cycles, lane 4: Tgo-Pol Z3 (exo-) 10 cycles, L=DNA ladder.

2.2.5 *Heat-cool extension vs DNA slippage*

A comparison in extension efficiency between the heat-cool cycling method demonstrated here, and the previously reported isothermal slippage reaction,[26,27] showed that for oligo seeds where b = 2 - 5, the products from heat-cool cycling are consistently longer, Fig. 17 (a)-(c). DNA slippage is a random process, whereas, thermal cycling directs extension. Therefore, heat-cool cycling is more suitable for the synthesis of long repeating sequences of DNA. Additionally, evidence is provided to support the proposed mechanism for the heat-cooling cycling method of extension as DNA slippage alone does not result in the same product lengths.

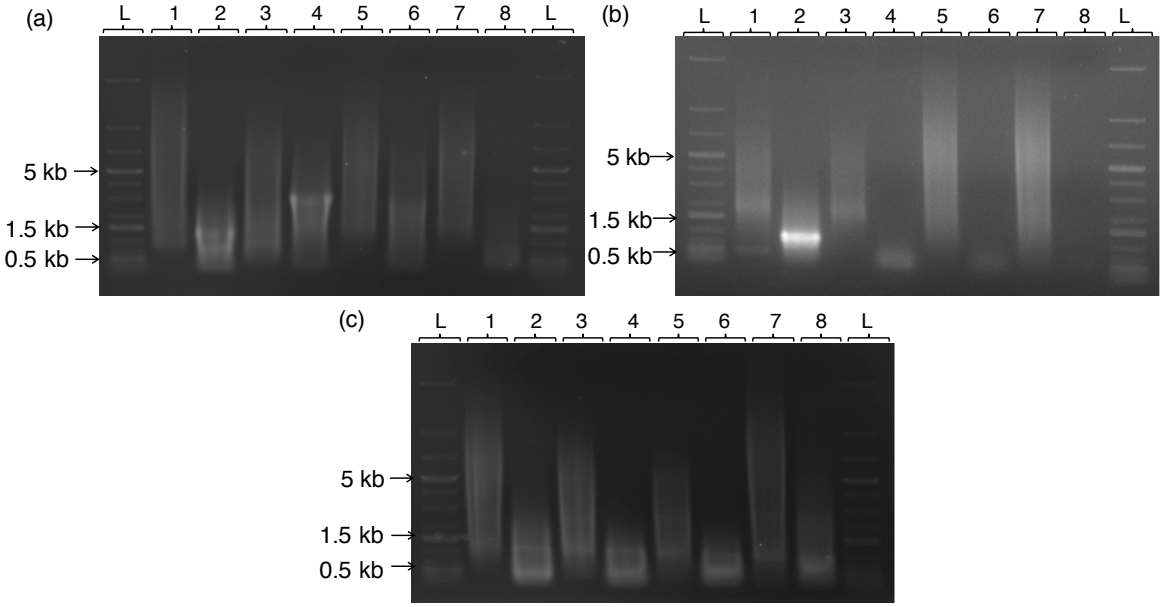


Fig. 17 Agarose gels of **(a)** Tgo-Pol Z3 exo-, **(b)** Pfu-Pol exo- and **(c)** Tgo-Pol exo- comparisons; lane 1: $[AG]_{10}/[TC]_{10}$ DNA extension products, lane 2: $[AG]_{10}/[TC]_{10}$ DNA slippage products, lane 3: $[A_2G]_7/[T_2C]_7$ DNA extension products, lane 4: $[A_2G]_7/[T_2C]_7$ DNA slippage products, lane 5: $[A_3G]_5/[T_3C]_5$ DNA extension products, lane 6: $[A_3G]_5/[T_3C]_5$ DNA slippage products, lane 7: $[A_4G]_4/[T_4C]_4$ DNA extension products, lane 8: $[A_4G]_4/[T_4C]_4$ DNA slippage products, L = DNA ladder.

2.3 Conclusions

DNA extension via heat-cool cycles provides a flexible method for the synthesis of long dsDNA of designed length and sequence from short oligo seeds. Repeating units of a number or bases, b, where b = 2, 3, 4, 5 and 10 bases can be extended to produce a range of lengths of long DNA. As shown with $[A_3G]_5/[T_3C]_5$, DNA lengths of smaller distribution can be extracted from agarose gels to yield uniform DNA samples. Oligo seeds containing repeating units of random sequences can also extend – showing the versatility of this method. Extension comparisons between two exonuclease minus and an exonuclease positive DNA polymerases illustrates the requirement for an exonuclease minus variant, however, additional mutations are not required. Finally, this method establishes a means for the introduction of multiple, uniformly spaced bases, controlled by user defined repeat units within the oligo seeds, to assemble DNA with potential for adaptable nanomaterials. If modified DNA synthesis was possible by this method of DNA extension, it would further expand the DNA nanotechnology tool-box and provide avenues to a variety of designer DNA. Chapter 3 will consider the addition of modified dNTPs into the heat-cool cyclic method of DNA synthesis reported here.

2.4 Experimental

Deoxynucleotide concentration

Deoxyoligonucleotides were purchased from Eurofins (Ebersberg, Germany). The concentration of the deoxynucleotides were determined using absorbance at 260nm. Absorbance was measured using a NanoDrop 1000 Spectrometer (Thermo Scientific, UK). Deoxyoligonucleotide concentrations were calculated using the Beer-Lambert Law (equation 1).

$$A=\varepsilon cl \qquad \text{(Equation 1)}$$

c= concentration (M)
A= Absorbance
ε = extinction coefficient of the deoxyoligonucleotide at 260 nm ($mM^{-1}cm^{-1}$) (Table 2).
l = path length (cm)

Name	Sequence	Extinction Coefficient $mM^{-1}cm^{-1}$
$(AG)_{10}$	AGAGAGAGAGAGAGAGAGAG	242.1
$(TC)_{10}$	TCTCTCTCTCTCTCTCTCTC	144.9
$(AAG)_7$	AAGAAGAAGAAGAAGAAGAAG	266.5
$(TTC)_7$	TTCTTCTTCTTCTTCTTCTTC	156.2
$(AAAG)_5$	AAAGAAAGAAAGAAAGAAAG	259.6
$(TTTC)_5$	TTTCTTTCTTTCTTTCTTTC	150.8
$(AAAAG)_4$	AAAAGAAAAGAAAAGAAAAG	263.16
$(TTTTC)_4$	TTTTCTTTTCTTTTCTTTTC	151.9
$(AAAAAAAAAG)_2$	AAAAAAAAAGAAAAAAAAAG	270.2
$(TTTTTTTTTC)_2$	TTTTTTTTTCTTTTTTTTTC	171.4
$(GATC)_5$	GATCGATCGATCGATCGATC	193.5
$(ACTGATCAGC)_2$	ACTGATCAGCACTGATCAGC	195.8
$(TGACTAGTCG)_2$	TGACTAGTCGTGACTAGTCG	191.2

Table 2 Oligos used in this study.

Each deoxynucleotide extinction coefficient ($mM^{-1}cm^{-1}$) was calculated using equation 2 and Table 3.

$$\varepsilon oligodeoxynucleotide = \sum \varepsilon bases \times 0.9^{51} \qquad \text{(Equation 2)}$$

Nucleotide	Extinction coefficient at 260 nm
dA	15400
dC	7400
dG	11500
dT	8700

Table 3 Nucleotide extinction coefficients taken from Sigma-Aldrich (http://www.sigmaaldrich.com/technical-documents/articles/biology/quantitation-of-oligos.html).

Primer template annealing

The primer duplexes, shown in Table 4, were prepared by addition of DNA annealing buffer (10 mM Hepes pH 7.5, 100 mM NaCl and 1 mM EDTA) to the deoxynucleotides and heating to 95 °C for 10 minutes and cooling slowly to room temperature, 25 °C. Annealed duplexes were stored at -20 °C.

Oligos to be combined	**Final oligo seed**	**DNA duplex length / bp**
$(AG)_{10}$ and $(TC)_{10}$	$[AG]_{10}/[TC]_{10}$	20
$(AAG)_7$ and $(TTC)_7$	$[AAG]_7/[TTC]_7$	21
$(AAAG)_5$ and $(TTTC)_5$	$[AAAG]_5/[TTTC]_5$	20
$(AAAAG)_4$ and $(TTTTC)_4$	$[AAAAG]_4/[TTTTC]_4$	20
$(AAAAAAAAAG)_2$ and $(TTTTTTTTTC)_2$	$[AAAAAAAAAG]_2/[TTTTTTTTTC]_2$	20
$(GATC)_5$	$[GATC]_5/[CTAG]_5$	20
$(ACTGATCAGC)_2$ and $(TGACTAGTCG)_2$	$[ACTGATCAGC]_2/[TGACTAGTCG]_2$	20

Table 4 Summary of oligos required to form the primer duplexes used.

DNA polymerases

The DNA polymerases used in this study are described in Table 5.

DNA Polymerase	**Properties**	**Source**	**Extinction coefficient $M^{-1}cm^{-1}$**	**DNA pol concentration for extension**
Pfu-Pol exo+	Archael family B polymerase	Agilent	121710	1 unit / 20 μl reaction
Pfu-Pol exo-	Archael family B polymerase with the 3´→ 5´ exonuclease activity removed	Prepared and purified in house[16,37]	121710	200 nM
Tgo-Pol exo-	Archael family B polymerase with the 3´→ 5´ exonuclease activity removed	Prepared and purified in house[16,37]	116020	200 nM
Tgo-Pol Z3 exo-	Archael family B polymerase low fidelity variant with the 3´→ 5´ exonuclease activity removed and alterations to the fingers domain	Prepared and purified in house[16,37]	116020	200 nM

Table 5 Description of DNA polymerases used.
Extinction coefficients of DNA polymerases used in this study were calculated at 280 nm. The values were determined using ExPASy ProtParam. The extinction coefficients of enzymes with single amino acid substitutions are assumed to be the same as wild type.

DNA duplex extension using heat-cool cycles

0.5 µM DNA duplex (oligo seed), DNA polymerase (Table 6), corresponding reaction buffer (Table 7), and 0.5 mM dNTPs (dCTP, dATP, dTTP and dGTP) were mixed. Thermocycling was carried out using an Applied Biosciences Veriti 96 well Thermal Cycler by the following method:

N (number of cycles stated)	x	30 seconds at 95 °C
		30 seconds at 55 °C
		30 seconds at 72 °C

The products were cooled to 4 °C after the reaction. The reactions were terminated by the addition of 40 mM EDTA.

DNA extension product was then purified using a QIAquick PCR purification kit (25) (QIAGEN, Manchester, UK) following the manufacturers protocol.

Where stated in the results and discussion, additional dNTPs (2 mM) and Tgo-Pol Z3 exo- (200 nM) were added every 10 cycles to the reaction mixture. As a control, an equal volume of nanopure-H_2O was added to additional reaction solutions to match dilution.

DNA Slippage reaction

0.5 µM DNA duplex (Table 4), 200 nM DNA polymerase, DNA polymerase reaction buffer (Table 6), and 0.5 mM deoxynucleotide triphosphates (dNTPs) (dCTP, dATP, dTTP and dGTP) were mixed. Reaction mixtures were incubated at 37 °C for 2 hours and quenched by the addition of 40 mM EDTA. Product was then purified using a QIAquick PCR purification kit (25) (QIAGEN, Manchester, UK) following manufacturers protocol.

DNA polymerase	**Buffer**	**Components**
Pfu-Pol exo-, Tgo-Pol exo- and Tgo-Pol Z3 exo-	10 x Pfu-Pol exo- and Pfu-Pol-Z3 exo- Reaction Buffer	200 mM Tris-HCl (pH 8.8, 25 °C), 100 mM $(NH_4)_2SO_4$, 100 mM KCl, 1 % Triton X-100, 1 mg/mL Bovine Serum Albumin (BSA) and 20 mM $MgSO_4$

Table 6 Composition of 10x polymerase reaction buffers.

Agarose Gel Electrophoresis

The DNA extension products were analysed by gel electrophoresis in either TAE (Tris, Acetic Acid and $Na_2EDTA.2H_2O$) or TBE (Tris, Boric Acid and $Na_2EDTA.2H_2O$)

buffer. Either 0.5 % or 1% Agarose (Melford, Ipswich, UK) or 0.3 % Seakem Gold Agarose (Lonza, Reading, UK) was added to the 1 % TBE or TAE buffer and heated to dissolve. The agarose solution was either supplemented with ethidium bromide (Sigma Aldrich) and poured once it had cooled to 50 °C or post-stained with a 5 µg/mL solution of ethidium bromide. The 100 bp+, 1 kb and 1 kb+ DNA ladders were purchased from Thermo Scientific and provided with a Loading dye (2.5 % Ficoll-400, 11 mM EDTA, 3.3 mM Tris-HCl (pH 8.0, 25 °C), 0.017 % SDS and 0.015 % bromophenol blue). DNA samples were supplemented with the gel loading dye. The gels were run at 100 V, 100 mA, 10 W for approximately 1 hour and then visualised using an ultra-violet transilluminator.

Lonza Gel Electrophoresis

Lonza Gel electrophoresis was performed using the pre-cast Lonza Gel kit (Lonza™) and performed as stated in the manufactures handbook. DNA was recovered by addition of nano pure H_2O to the wells. Concentration of extracted DNA was performed using a rotary evaporator (Savant Speed Vac).

Extraction of DNA Bands from Agarose Gel

DNA bands were manually excised from 1 % agarose gel using a surgical blade following resolution by electrophoresis. The excised gel was then weighed and product extracted using a QIAquick gel extraction kit (QIAGEN) and performed following the manufacturers protocol. Product was then purified using a QIAquick PCR purification kit (25) (QIAGEN) following manufacturers protocol.

AFM Imaging

The top layer of the mica surface was cleaved using sticky tape. 5 µL of the DNA sample was dropped onto the mica surface at 25 °C using a micro pipette. After 5 minutes, 5 µL nanopure-H_2O was dropped onto the mica surface, maintaining the 25° angle. Nitrogen gas was then passed over the surface and allowed drying for 1 hour under laminar flow. Light microscopy was used to locate the sample and cantilever position. AFM images were collected using a Dimension V with a nanoscope controller (Veeco Instruments Inc., Metrology Group, Santa Barbara, CA). The tapping mode was used with an etched silicon tip (Tap 300Al-G, 300 kHz,

40 N/m) on an isolation table (Veeco Inc., Metrology Group) to reduce interference. Nanoscope 7.00b19 software was used to acquire data.

Ultra-violet/ Visible Spectroscopy

UV-Vis spectroscopy was performed using a NanoDrop 1000 Spectrometer (Thermo Scientific) for volumes less than 2 µL and using a Varian Cary 100Bio UV-Vis spectrophotometer with a Varian Cary temperature controller for larger volumes. Spectrometer was blanked using nanopure-H_2O or QIAGEN elusion buffer. Concentration and peak determination was performed as shown on Table 4.

Fluorescent Microscope Imaging

DNA samples were prepared as stated above in 'Heat-cool cycle DNA extension', diluted to approximately 10 ng/µL and supplemented with 0.5 nM YOYO-Iodine (Life technologies, Paisley, UK). Glass slides were cleaned using a diener electronic (Femto, Korea) and 5 µL sample was deposited using a micro pipette and spread on the surface. Sample was allowed to dry and mounted onto an Axioshop 2 plus (Zeiss, Germany) image platform, set to filter 44, with a Plan-NEOFLUAR 40 x/ 0.75 objective lens (Zeiss). The sample was excited at 490 nm from a ebq100 isolated mercury lamp (LEJ, Germany) and captured using an axioCam HRm (Zeiss).

DNA Melting Temperature

DNA oligo seeds were pre-annealed and DNA products were synthesised following the method in "heat-cool cycles DNA extension" for 30 cycles. Control dsDNA annealed from

5'GGGGATCCTCTAGAGTCGACCTGCTGGAAAGCAAGCTTGTCTCATGGACCGATA

5'TATCGGTCCATGAGACAAGCTTGCTTGCCAGCAGGTCGACTCTAGAGGATCCCC

200 fold Quant-iT™ PicoGreen® was added in a 1:1 ratio to 2 µM dsDNA. DNA melting was monitored via fluorescence detection from 45 to 95 °C rising 1 °C every 10 seconds on a Corbett Research Rotor-Gene 6000.

2.5 References

1. Zahid, M., Kim, B., Hussain, R., Amin, R. & Park, S. H. DNA nanotechnology: a future perspective. *Nanoscale Res. Lett.* **8,** 119 (2013).
2. Pinheiro, A. V, Han, D., Shih, W. M. & Yan, H. Challenges and opportunities for structural DNA nanotechnology. *Nat Nano* **6,** 763–772 (2011).
3. Simms, E. S. & Kornberg, A. Enzymatic synthesis of deoxyribonucleic acid III. The incorporation of pyrimidine and purine analogues into deoxyribonucleic acid. *Proc. Natl. Acad. Sci.* **44,** 633–640 (1958).
4. Josse, J., Kaiser, A. D. & Kornberg, A. Enzymatic Synthesis of Deoxyribonucleic Acid. *J. Biol. Chem.* **236,** 864–875 (1961).
5. Sokolov, B. P. Primer extension technique for the detection of single nucleotide in genomic DNA. *Nucleic Acids Res.* **18,** 3671 (1989).
6. Mullis, K. B., Elrich, H. A., Horn, G. T., Saiki, R. K. & Scharf, S. J. Process for amplifying, detecting, and/or-cloning nucleic acid sequences. (1987).
7. Chien, A., Edgar, D. B. & Trela, J. M. Deoxyribonucleic Acid Polymerase from the Extreme Thermophile Thermus aquaticus. *J. Bacteriol.* **127,** 1550–1557 (1976).
8. Cerrito, E., Frances, C. & Saiki, R. K. Purified thermostable enzyme. (1989).
9. Eckert, K. A. & Kunkel, T. A. DNA Polymerase Fidelity and the Polymerase Chain Reaction. *Genome Res.* **1,** 17–24 (1991).
10. Bebenek, K. & Kunkel, T. A. Functions of DNA polymerase. *Adv. Protein Chem.* **69,** 137–165 (2004).
11. Wang, J. *et al.* Crystal Structure of a pol α Family Replication DNA Polymerase from Bacteriophage RB69. *Cell* **89,** 1087–1099 (1997).
12. Joyce, C. M. & Steitz, T. A. Function and Structure Relationship in DNA Polymerase. *Annu. Rev. Biochem.* **63,** 777–822 (1994).
13. Ollis, D. L., Brick, P., Hamlin, R., Xuong, N. G. & Steitz, T. A. Structure of large fragment of Escherichia coli DNA polymerase I complexed with dTMP. *Nature* **313,** 762–766 (1985).
14. Brautigam, C. A. & Steitz, T. A. Structural and functional insights provided by crystal structures of DNA polymerases and their substrate complexes. *Curr. Opin. Struct. Biol.* **8,** 54–63 (1998).
15. Beard, W. A. & Wilson, S. H. Structural insights into the origins of DNA polymerase fidelity. *Structure* **11,** 489–496 (2003).
16. Jozwiakowski, S. K. & Connolly, B. A. A modified family-B archaeal DNA polymerase with reverse transcriptase activity. *ChemBioChem* **12,** 35–37 (2011).
17. Braithwaite, D. K. & Ito, J. Compilation, alignment, and phylogenetic relationships of DNA polymerases. *Nucleic Acids Res.* **21,** 787–802 (1993).
18. Lundberg, K. S. *et al.* High-fidelity amplification using a thermostable DNA polymerase isolated from Pyrococcus furiosus. *Gene* **108,** 1–6 (1991).
19. Lindahl, T. Instability and decay of the primary structure of DNA. *Nature* **362,** 709–715 (1993).
20. Pearl, L. H. Structure and function in the uracil-DNA glycosylase superfamily. *Mutat. Res.* **460,** 165–181 (2000).
21. Greagg, M. A. *et al.* A read-ahead function in archaeal DNA polymerases detects promutagenic template-strand uracil. *Proc. Natl. Acad. Sci.* **96,** 9045–9050 (1999).
22. Mcinerney, P., Adams, P. & Hadi, M. Z. Error Rate Comparison during Polymerase Chain Reaction by DNA Polymerase. *Mol. Biol. Int.* **2014,** (2014).
23. Eckert, K. A. & Kunkel, T. A. High fidelity DNA synthesis by the Thermus aquaticus DNA polymerase. *Nucleic Acids Res.* **18,** 3739–3744 (1990).
24. Petruska, J., Hartenstine, M. J. & Goodman, M. F. Analysis of Strand Slippage in DNA Polymerase Expansions of CAG / CTG Triplet Repeats Associated with Neurodegenerative Disease *. *J. Biol. Chem.* **273,** 5204–5210 (1998).
25. Wells, R. D., Dere, R., Hebert, M. L., Napierala, M. & Son, L. S. Advances in mechanisms of genetic instability related to hereditary neurological diseases. *Nucleic Acids Res.* **33,** 3785–3798 (2005).
26. Schlötterer, C. & Tautz, D. Slippage synthesis of simple sequence DNA. *Nucleic Acids Res.* **20,** 211–215 (1992).
27. Kotlyar, A. *et al.* Synthesis of novel poly(dG)–poly(dG)–poly(dC) triplex structure by Klenow exo– fragment of DNA polymerase I. *Nucleic Acids Res.* **33,** 6515–6521 (2005).
28. Klenow, H. & Henningsen, I. Selective Elimination of the Exonuclease Activity of the Deoxyribonucleic Acid Polymerase from Escherichia coli B by Limited Proteolysis *. *Proc. Natl.*

Acad. Sci. **65,** 168–175 (1970).
29. Fire, a & Xu, S. Q. Rolling replication of short DNA circles. *Proc. Natl. Acad. Sci. U. S. A.* **92,** 4641–4645 (1995).
30. Gilbert, W. & Dressler, D. *DNA Replication: The Rolling Circle Model*. (Cold Spring Harb Symp Quant Biol, 1968).
31. Wang, X., Gou, D. & Xu, S. Y. Polymerase-endonuclease amplification reaction (PEAR) for large-scale enzymatic production of antisense oligonucleotides. *PLoS One* **5,** 1–7 (2010).
32. Wu, M., Chow, L. & Hsieh, M. Amplification of GAA / TTC triplet repeat in vitro : preferential expansion of (TTC) n strand. *Biochem. Biophys. ACTA* **1407,** 155–162 (1998).
33. Takahashi, N., Sasagawa, N., Suzuki, K. & Ishiura, S. Synthesis of long trinucleotide repeats in vitro. *Neurosci. Lett.* **262,** 45–8 (1999).
34. Nielson, K. B., Costa, G. L. & Braman, J. Optimization of PCR using Pfu DNA polymerase. *Strategies Mol. Biol.* **9** 24-25 (1996).
35. Watson, S. M. D., Pike, A. R., Pate, J., Houlton, A. & Horrocks, B. R. DNA templated nanowires: morphology and electrical conductivity. *Nanoscale* **6** 4027–4037 (2014).
36. Chan, T. F. *et al.* A simple DNA stretching method for fluorescence imaging of single DNA molecules. *Nucleic Acids Res.* **34,** 1–6 (2006).
37. Evans, S. J. *et al.* Improving dideoxynucleotide-triphosphate utilisation by the hyper-thermophilic DNA polymerase from the archaeon Pyrococcus furiosis. *Nucleic Acids Res* **28**, 1059-1066 (2000).

Chapter 3.

Designer DNA: Synthesis of Modified and Sequence Controlled Double Stranded DNA

Table of Contents

3.1 Introduction

Enzymatic incorporation of modified dNTPs has potential in both biomedical[1,2] and nanotechnology applications,[3-6] with probable aid in DNA sequencing,[7,8] conjugation[9,10] and sensing.[3,11] The insertion of modified dNTPs would reveal a new array of specialist DNA structures capable of further function and added control. Currently, the DNA used within these advanced applications consists of a known length and sequence, however, both these attributes are difficult to design beyond the capabilities of an automated DNA synthesiser. Along with a high fidelity controlled base pairing system,[12] DNA is a stable molecule constructed from three components, the base, ribose sugar and phosphate – each of these groups can be artificially modified, Fig. 1. The synthesis of non-natural dNTPs is now highly established[13,14] with many now commercially available (Jena Biosciences (Germany) and Trilink (USA)).

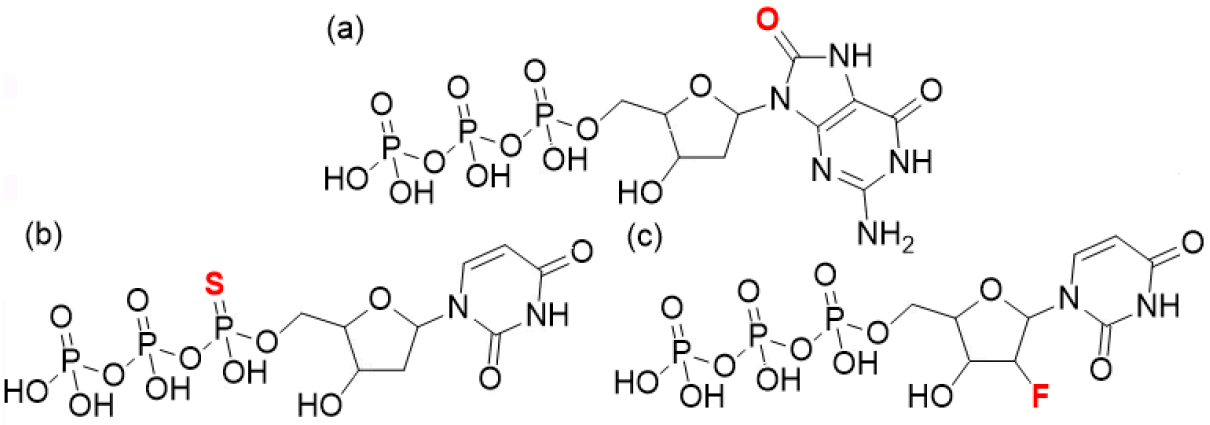


Fig. 1 dNTPs with modifications on the base **(a)** 8-oxo-dGTP,[15] the phosphate **(b)** dTTPαS[1] and the sugar **(c)** 2'-F-dUTP.[1,16]

The incorporation of modified nucleotides into dsDNA has been possible, by either phosphoramidite chemistry using automated DNA synthesisers[17,18] or enzymatically by primer extension assays[15] and template copying.[8] Shionoya's laboratory synthesised a range of unnatural bases, modified in the base pairing region, and incorporated these into short oligos by the phosphoramidite method.[5,17,19] These unnatural bases provide a specific anchor for certain metal ions, affording additional control over typical DNA templating.[20] However, this method of modified DNA synthesis is limited by the reducing yield of increasing oligo length during the phosphoramidite method. A PCR based method, PEAR, allows the enzymatic synthesis of modified antisense oligos.[1] PEAR is a more accommodating method for artificial DNA synthesis compared to the phosphoramidite method as the DNA

polymerase is more flexible, for example, azide functional groups cannot be incorporated into oligos by the phosphoramite method.[21]

The incorporation of oxidised nucleotides has been studied due to their interference with the faithful replication of chromosomal DNA,[2] resulting in cellular mutations.[22] Incorporations in this instance are shown using primer extension assays where the end length of the modified DNA strand is equal to the template used.[15,23] A detailed study by the Loeb laboratory[15] outlines incorporation efficiencies of 8-oxo-dGTP and 8-NH_2-dGTP versus dGTP in the reaction pot, revealing reduced efficiency in both Pol α and Pol β, however, some level of incorporation is present. This is an efficient method for modified DNA synthesis, however, the lengths of DNA are limited by the template provided.

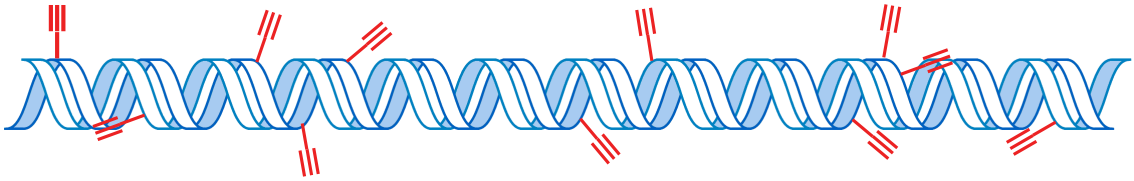


Fig. 2 Modified DNA, decorated with alkyne functional groups positioned on dT bases throughout the DNA molecule.[24]

Longer lengths of DNA are currently obtainable using DNA extracted from bacterial or mammalian cells, for example, lambda DNA and calf thymus DNA. Modifications cannot be built-in to the original DNA sequence, however, they can be incorporated into the new replica under PCR conditions. Carell's laboratory prepared modified DNA by this method to provide alkyne functional anchors for click chemistry on 300 and 900 bp lengths of DNA, Fig. 2.[24] The DNA sequence and length were known as the Pol η gene was used, however, neither of these attributes were designed specifically for this experiment. Although, 900 bp is an improvement on oligo lengths of DNA, functional nanomaterials necessitate user-definable lengths. The synthesis of longer lengths of modified DNA with designed sequence and length was performed by Ijiro's laboratory using the slippage reaction.[6] $[dG]_{15}/[dC]_{15}$ underwent isothermal incubation to allow extensions up to 2,000 bp in length where the sequence is controlled by the starting oligo duplex and the length is controlled by the time of incubation. This method in turn was used to synthesise DNA tri-block copolymers with blocks of unmodified versus modified regions of polynucleotides, Fig. 3 – additionally affording control over functional content location. Although Ijiro's results report a method with added control,

this control is still limited to the $[dX]_n/[dY]_n$ sequence and can only direct the site of modifications in large blocks.

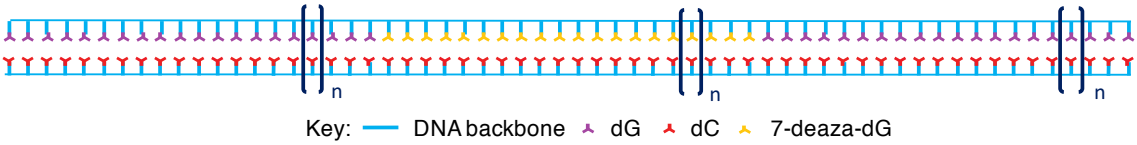


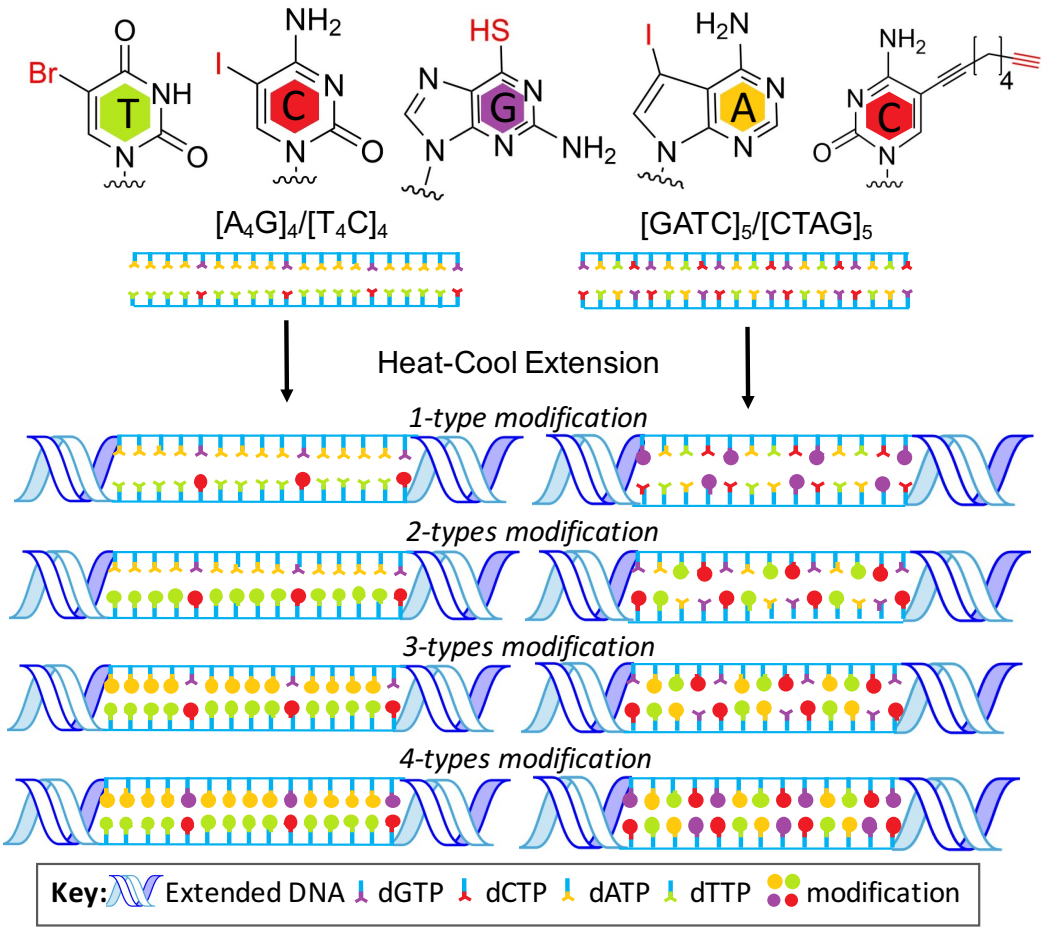
Fig. 3 DNA triblock copolymer consisting of an unmodified, a modified and an unmodified region of DNA of n bp in length, following three rounds of the slippage reaction.[6]

The integrity of the genome relies on the faithful replication of the DNA by a set of DNA polymerases. Many studies into the DNA polymerase ability to read mutations has led to several exo- derivatives with added abilities to handle artificial dNTPs. The *E. coli* Klenow exo- fragment is commonly used for primer extension assays and the slippage reaction due to its' ability at performing many biosciences applications, such as, DNA sequencing[25] and blunt end formation.[26] As the Klenow exo- fragment is a proteolytic product of *E. coli* DNA polymerase 1, it is not stable at high temperatures,[27] hence, thermostable DNA polymerases from the Archaea genus such as *Pyrococcus* and *Thermococcus* have been investigated to aid PCR based methods.[28-30] Many thermophilic Archaeal DNA polymerases are now commercially available; Vent exo-, Deep Vent exo- and Pfu-Pol exo- - all are absent in the 5'-3' proofreading ability, however, they are derived from a highly accurate DNA polymerase, providing support to enable correct reading and incorporation of artificial bases.

A method, able to incorporate artificial bases at not only known, but designed positions throughout the DNA duplex, would provide anchors for additional function – aiding the synthesis of designable nanomaterials.

3.2 Results and discussion

To build on the method reported in Chapter 2, adaptations to improve the control over the range of products was investigated to produce designer DNA. Scheme 1 outlines the chosen modifications to establish this designer DNA synthesis method. Single atom modifications of varying sizes and positions have been selected, 5-I-dCTP, 7-deaza-I-dATP, 5-Br-dUTP and 6-S-dGTP, along with a long chain modification, 5-alkyne-C_8-dCTP, to ascertain DNA polymerase capacity. Two DNA polymerases; the Tgo-Pol Z3 exo-[29] and *Thermococcus litoralis* DNA polymerase (Deep Vent exo-)[31] have been assessed during this chapter to ascertain requirements for modified base incorporation. A comparison with the commercially available Deep Vent exo- was performed to gauge the behaviour of Tgo-Pol Z3 exo-. Both DNA polymerases are disabled in the exonuclease domain and are therefore suitable for modified dNTP incorporation. Tgo-Pol Z3 exo- has an insertion of 3 amino acids in the loop subdomain – a subdomain involved in dNTP incorporation – allowing added flexibility for larger artificial base incorporation.[29]



Scheme 1 Designer DNA synthesis by heat-cool cycles from $[A_4G]_4/[T_4C]_4$ and $[GATC]_5/[CTAG]_5$ to control the production of DNA with specific sequence, length and chemical content.

3.2.1 DNA extension reactions incorporating 1-type of modification

3.2.1.1 DNA extensions with $[GATC]_5/[CTAG]_5$

$[GATC]_5/[CTAG]_5$ is a self-complementary sequence, containing an equal distribution of each base on both strands, hence will lead to no bias when analysing the DNA polymerase ability to incorporate modified bases using the heat-cool cycles described in Chapter 2.

3.2.1.1.1 5-I-dCTP incorporation with $[GATC]_5/[CTAG]_5$

Incorporation of the single atom modification, 5-I-dCTP, was initially investigated with the self-commentary oligo seed $[GATC]_5/[CTAG]_5$ and Tgo-Pol Z3 exo-. Extension using Tgo-Pol Z3 exo- is clearly highly proficient when incorporating this artificial base – lengths up to 3,000 bp are reached within 5 heat-cool cycles, Fig. 4 (a). The agarose gel, depicts a distribution of products lengths, consistent with the DNA extension results in Chapter 2. A maximum extension length is reached before 30 cycles - this is shown clearly on the Image J plot, Fig. 4 (b), illustrating a consistent maximum intensity (modal length) from 20-30 cycles.

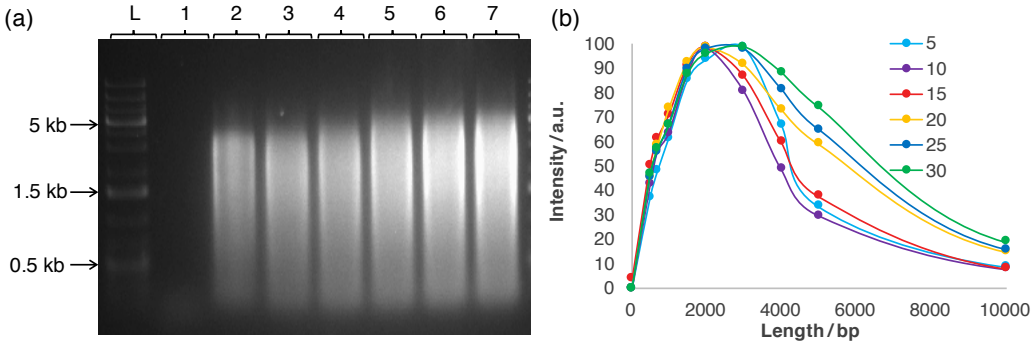


Fig. 4 (a) 5-I-dC-DNA agarose gel of $[GATC]_5/[CTAG]_5$ DNA extension products after 0, 5, 10, 15, 20, 25 and 30 cycles from lanes 1-7, respectively. **(b)** Image J analysis of (a) agarose gel, normalised to 100 for the maximum fluorescence of each lane.

3.2.1.1.2 5-alkyne-C_8-dCTP incorporation with $[GATC]_5/[CTAG]_5$

To establish whether the DNA polymerase can handle bulky chain derivatives, 5-alkyne-C_8-dCTP was added to the reaction pot in place of dCTP. Equally, this artificial base had little effect on the DNA polymerase capability, Fig. 5 (a). DNA extension performed with this artificial base also reached the maximum length before 30 cycles, exhibiting similar product lengths from 15-30 cycles as depicted in more detail by the Image J analysis, Fig. 5 (b). Once again, a large band is visible on the agarose gel, confirming a distribution of lengths were produced, spanning from 200 to 3,000 bp, with a peak maximum at 2,000 bp.

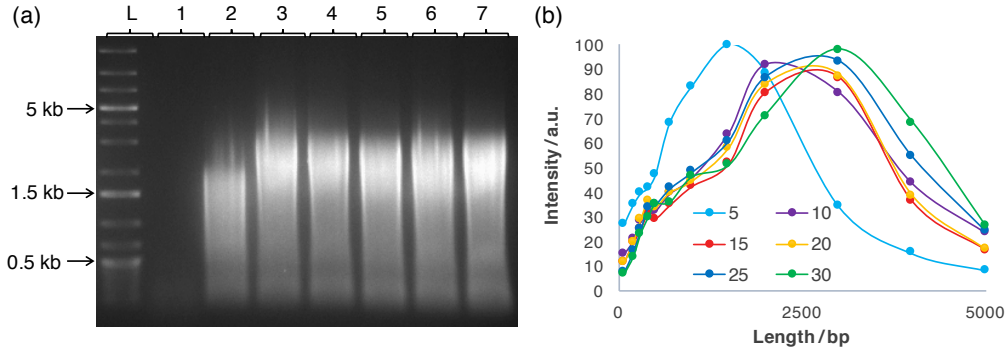


Fig. 5 (a) 5-alkyne-C_8-dC-DNA agarose gel of $[GATC]_5/[CTAG]_5$ DNA extension products after 0, 5, 10, 15, 20, 25 and 30 cycles from lanes 1-7, respectively. **(b)** Image J analysis of (a) agarose gel, normalised to 100 for the maximum fluorescence of each lane.

3.2.1.1.3 *7-deaza-I-dATP incorporation with $[GATC]_5/[CTAG]_5$*

To determine whether the position of the modification on the base affects the incorporation, dATP was replaced with 7-deaza-I-dATP. DNA extension was again monitored up to 30 cycles, Fig. 6 (a) and expressed a reduced extension proficiency in comparison to 5-I-dCTP. A closer analysis of the Image J plot, reveals the maximum extension length was not reached until 25 cycles suggesting incorporation is less favourable in the N7 position of dATP derivatives. This dATP derivative also involves the exchange of nitrogen for carbon, potentially altering the electron density distribution and thus, the DNA polymerase handling of the modification.

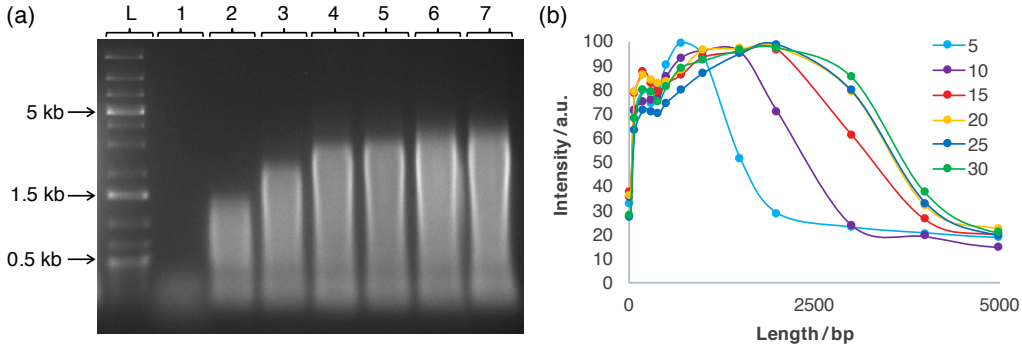


Fig. 6 (a) 7-deaza-I-dA-DNA agarose gel of $[GATC]_5/[CTAG]_5$ DNA extension products after 0, 5, 10, 15, 20, 25 and 30 cycles from lanes 1-7, respectively. **(b)** Image J analysis of (a) agarose gel, normalised to 100 for the maximum fluorescence of each lane.

3.2.1.1.4 *5-Br-dUTP incorporation with $[GATC]_5/[CTAG]_5$*

As iodine has a larger atomic radius than bromine, bromine containing derivatives were expected to incorporate at an equal rate, if not greater than the I-dNTPs. Analysis of extension products however, elucidates a reduced incorporation rate compared to the I-derivatives, Fig. 4 and 6 - a 2,000 bp extension product length was not achieved until 15 cycles – confirmed by the Image J analysis depicting near identical plot profiles from 15-30 cycles, Fig. 7. The agarose gel and Image J plot illustrates a lag phase in

enzymatic extension which is overcome after 5 cycles. The slight reduction in extension rate may be due to the larger electronegativity difference between hydrogen and bromine versus hydrogen and iodine. Bromine is more electronegative, potentially changing the electron density of the delocalised electrons in the base ring, altering the hydrogen bonding environment – a potentially unfavourable attribute for DNA stability and enzyme recognition.

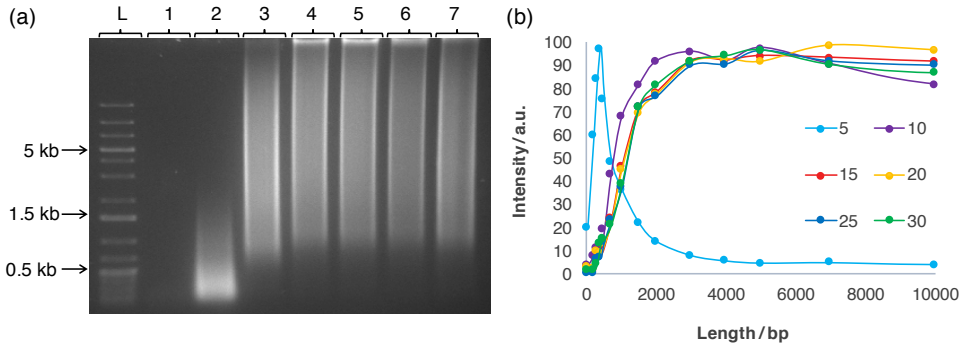


Fig. 7 (a) 5-Br-dU-DNA agarose gel of $[GATC]_5/[CTAG]_5$ DNA extension products after 0, 5, 10, 15, 20, 25 and 30 cycles from lanes 1-7, respectively. **(b)** Image J analysis of (a) agarose gel, normalised to 100 for the maximum fluorescence of each lane.

3.2.1.1.5 *6-S-dGTP incorporation with $[GATC]_5/[CTAG]_5$*

Thus far, the modifications discussed are located in the major groove of DNA once incorporated. As implied by the bromine derivative, alterations in the base pairing region may reduce polymerase ability. In 6-S-dGTP there is an exchange between the nucleobase 6-oxygen for a sulfur – a larger atom with lower electronegativity. As anticipated, extension rate is largely reduced when exchanging dGTP for 6-S-dGTP in the extension reaction, Fig. 8 (a). Additionally, the reaction also appears to have a lag time of 5 cycles before extension begins, Fig. 8 (b). Due to the increased atomic radius of sulfur, some distortion in the dsDNA helical structure and a reduced dsDNA stability is anticipated.[32] Therefore, re-annealing with a shifted structure may be less favourable – this could account for the lag time before the dsDNA is long enough to overcome this instability challenge.

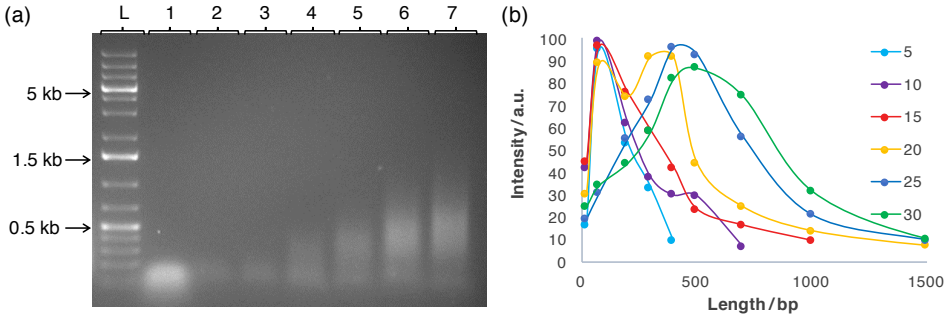


Fig. 8 (a) 6-S-dG-DNA agarose gel of $[GATC]_5/[CTAG]_5$ DNA extension products after 0, 5, 10, 15, 20, 25 and 30 cycles from lanes 1-7, respectively. **(b)** Image J analysis of (a) agarose gel, normalised to 100 for the maximum fluorescence of each lane.

3.2.1.2 *DNA extensions with $[A_4G]_4/[T_4C]_4$*

To understand the true potential of Tgo-Pol Z3 exo-, a sequence with repeated bases, $[A_4G]_4/[T_4C]_4$, was investigated. $[A_4G]_4/[T_4C]_4$ affords further control over base positions – dG and dC appear once every 5 bases and each nucleotide only appears on one strand. In future studies the number of dA and dT bases can be varied in the form, $[A_xG]_n/[T_xC]_n$ to study how the base separation between the dC/dG sites can be exploited for nanomaterial assembly.

3.2.1.2.1 *5-I-dCTP incorporation with $[A_4G]_4/[T_4C]_4$*

5-I-dC incorporation into $[A_4G]_n/[T_4C]_n$ showed similar extensions when compared to $[GATC]_n/[CTAG]_n$ (modal DNA product length of 3,000 bp for both oligo seeds), Fig. 9, confirming this modified dNTP has little effect on DNA polymerase efficiency. From the Image J analysis, it is clear that maximum extension is reached after 10 cycles, Fig. 9 (b), producing a distribution of lengths as exhibited by the long bands on the agarose gel, Fig. 9 (a).

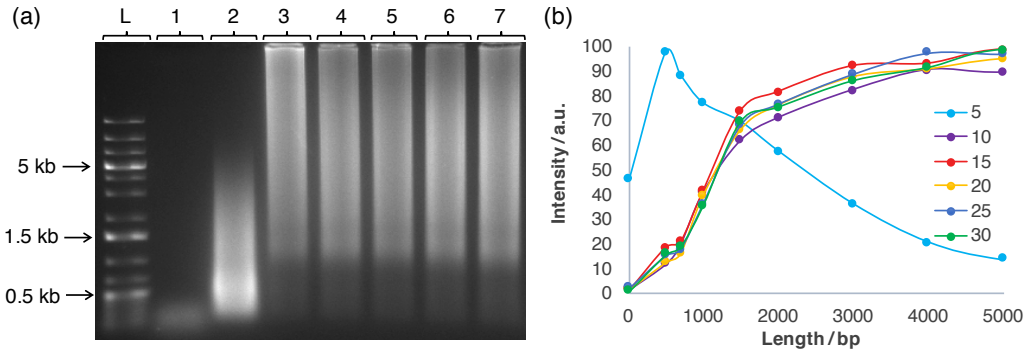


Fig. 9 (a) 5-I-dC-DNA agarose gel of $[A_4G]_4/[T_4C]_4$ DNA extension products after 0, 5, 10, 15, 20, 25 and 30 cycles from lanes 1-7, respectively. **(b)** Image J analysis of (a) agarose gel, normalised to 100 for the maximum fluorescence of each lane.

3.2.1.2.2 *5-alkyne-C_8-dCTP incorporation with $[A_4G]_4/[T_4C]_4$*

5-alkyne-C_8-dCTP incorporation into $[A_4G]_n/[T_4C]_n$ is almost identical to $[GATC]_n/[CTAG]_n$ incorporation, exhibited by the agarose gel images. Both gels display a similar distribution of product lengths, ranging from 200 to 4,000 bp and reach a maximum extension length after 10 heat-cool cycles. Equally, the modal length is approximately 2,000 bp for each sequence as expressed from the Image J analysis, Fig. 5 (a) and 10 (a).

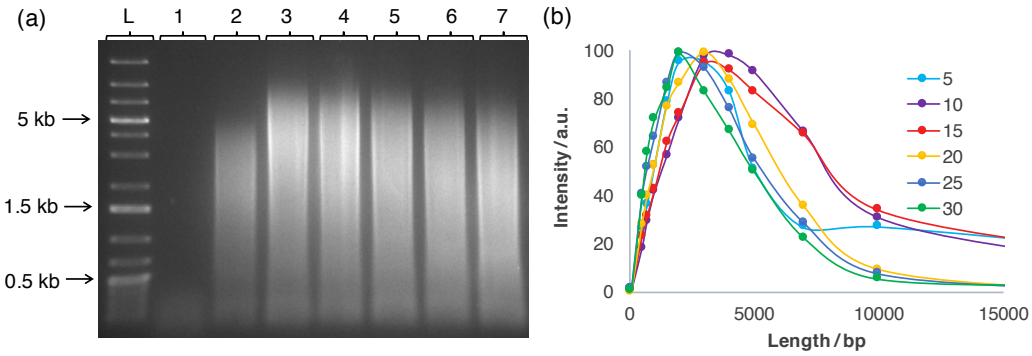


Fig. 10 (a) 5-alkyne-C_8-dC-DNA agarose gel of $[A_4G]_4/[T_4C]_4$ DNA extension products after 0, 5, 10, 15, 20, 25 and 30 cycles from lanes 1-7, respectively. **(b)** Image J analysis of (a) agarose gel, normalised to 100 for the maximum fluorescence of each lane.

3.2.1.2.3 *7-deaza-I-dATP incorporation with $[A_4G]_4/[T_4C]_4$*

Heat-cool cycles performed with $[A_4G]_4/[T_4C]_4$ and in the presence of 7-deaza-I-dATP in the reaction produced a maximum modal product length of 4,000 bp after 10 cycles, Fig. 11. Surprisingly, the maximum product length was reached earlier with the $[A_4G]_n/[T_4C]_n$ than with the $[GATC]_n/[CTAG]_n$ sequence. $[A_4G]_n/[T_4C]_n$ extension requires consecutive incorporation of the artificial dA derivative, hence a slower rate of incorporation was anticipated. However, these results show this modified dNTP has negligible effect on DNA extension as the overall rate of extension is not hindered by a required increase in the number of modifications to produce the same product length.

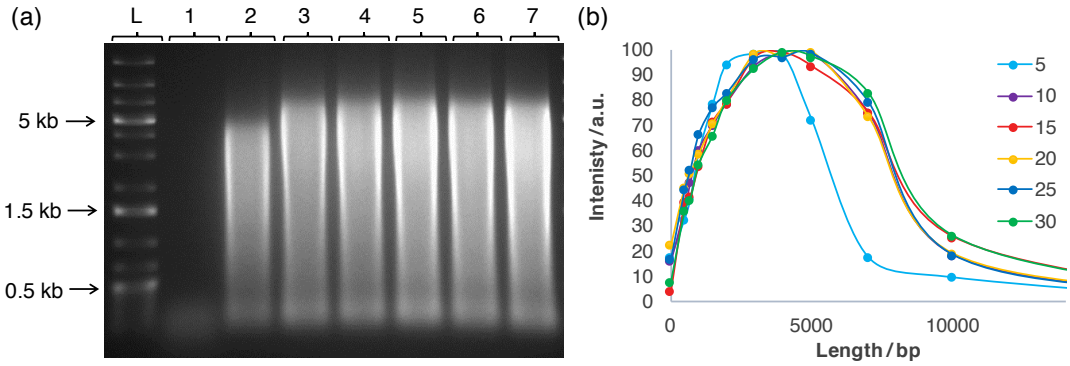


Fig. 11 (a) 7-deaza-I-dA-DNA agarose gel of $[A_4G]_4/[T_4C]_4$ DNA extension products after 0, 5, 10, 15, 20, 25 and 30 cycles from lanes 1-7, respectively. **(b)** Image J analysis of (a) agarose gel, normalised to 100 for the maximum fluorescence of each lane.

3.2.1.2.4 *5-Br-dUTP incorporation with $[A_4G]_4/[T_4C]_4$*

A comparison between 5-Br-dUTP incorporation into the $[A_4G]_n/[T_4C]_n$ sequence versus the $[GATC]_n/[CTAG]_n$ sequence, exhibited a slower rate, reaching only 200 bp versus 1,000 bp, after 5 cycles, Fig. 12. As recalled, 5-Br-dUTP has a negative effect on DNA extension due to the potential disruption of the hydrogen bonding region due to the increased electronegativity of the Br atom, hence, in a sequence where four consecutive incorporations are required, a larger impact on the rate of extension is observed. The adjustment in hydrogen bonding strength may cause a reduction in stability of the DNA duplex. On beginning extension, a shift is required – if there are 4 consecutive bases of unfavourable or less stable annealing, the overlapping duplex may not be completely double stranded and thus disable DNA polymerase binding and sequence copying. As the final maximum product length is similar to the dNTPs which the DNA polymerase can handle well, it is likely that the reduced initial extension rate is due to initial instabilities, rather than DNA polymerase handling.

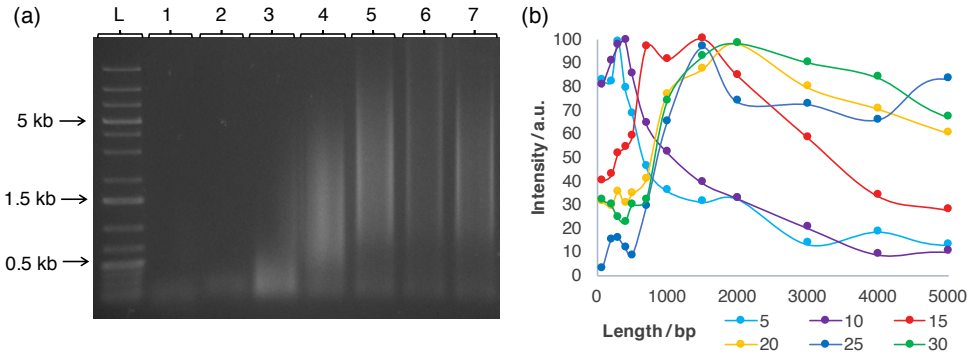


Fig. 12 (a) 5-Br-dU-DNA agarose gel of $[A_4G]_4/[T_4C]_4$ DNA extension products after 0, 5, 10, 15, 20, 25 and 30 cycles from lanes 1-7, respectively. **(b)** Image J analysis of (a) agarose gel, normalised to 100 for the maximum fluorescence of each lane.

3.2.1.2.5 *6-S-dGTP incorporation with $[A_4G]_4/[T_4C]_4$*

6-S-dGTP also has a reduced incorporation rate using $[A_4G]_4/[T_4C]_4$ as the oligo seed, Fig. 13 (a). Again, we expect the modification to induce some structural instabilities into the dsDNA. Conversely to $[GATC]_n/[CTAG]_n$, the modification only appears on one strand in $[A_4G]_n/[T_4C]_n$, hence it is proposed that as any deformations are unequal between the two strands the 6-S-dG sites may result in reduced stability, stalling dNTP incorporation. Slow extension is observed up to 30 heat-cool cycles, and only reaching modal product lengths of 300 bp.

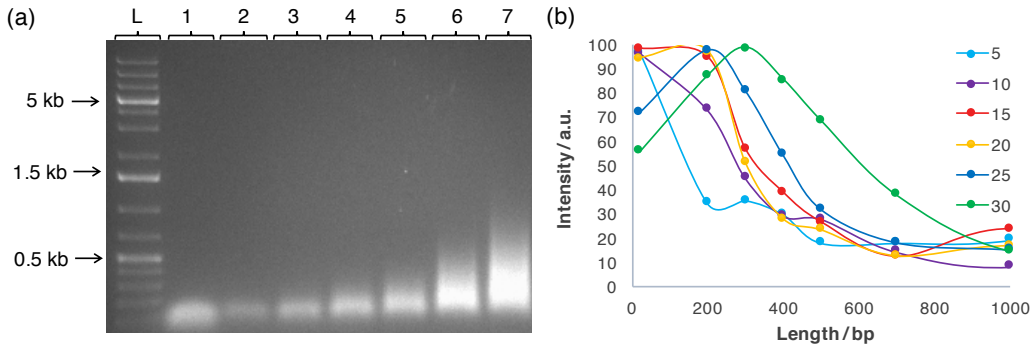


Fig. 13 (a) 6-S-dG-DNA agarose gel of $[A_4G]_4/[T_4C]_4$ DNA extension products after 0, 5, 10, 15, 20, 25 and 30 cycles from lanes 1-7, respectively. **(b)** Image J analysis of (a) agarose gel, normalised to 100 for the maximum fluorescence of each lane.

3.2.1.3 Summary of 1-type of modification per extension reaction

The 5 artificial bases discussed above are each within the Tgo-Pol Z3 exo- handling capacity to perform replication incorporations. The DNA polymerase activity varies depending on the position and type of modification, within each modified nucleotide, as shown by the variation in modal DNA product lengths after 5 and 10 cycles, Fig. 14. Extensions with $[GATC]_5/[CTAG]_5$ and 5-I-dCTP, 5-C_8-alkyne-dCTP and 7-deaza-I-dATP clearly show maximum extension is rapidly reached by 5 heat-cool cycles. The same trend is seen with $[A_4G]_4/[T_4C]_4$ and 5-C_8-alkyne-dCTP and 7-deaza-I-dATP, however maximum extension is not observed with 5-I-dCTP until after 10 heat-cool cycles. Due to the less favourable incorporation of 5-Br-dUTP and 6-S-dGTP, a lag phase is observed. 5-Br-dUTP and 6-S-dGTP interfere with the base pairing regions, therefore, the reduced stability affects enzyme processivity. However, 5-Br-dUTP does reach similar maximum product lengths to 5-I-dCTP, 5-C_8-alkyne-dCTP and 7-deaza-I-dATP once more favourable duplex stability has been established. 6-S-dGTP, however, continues to extend up to 30 cycles with a heavily reduced DNA product length profile.

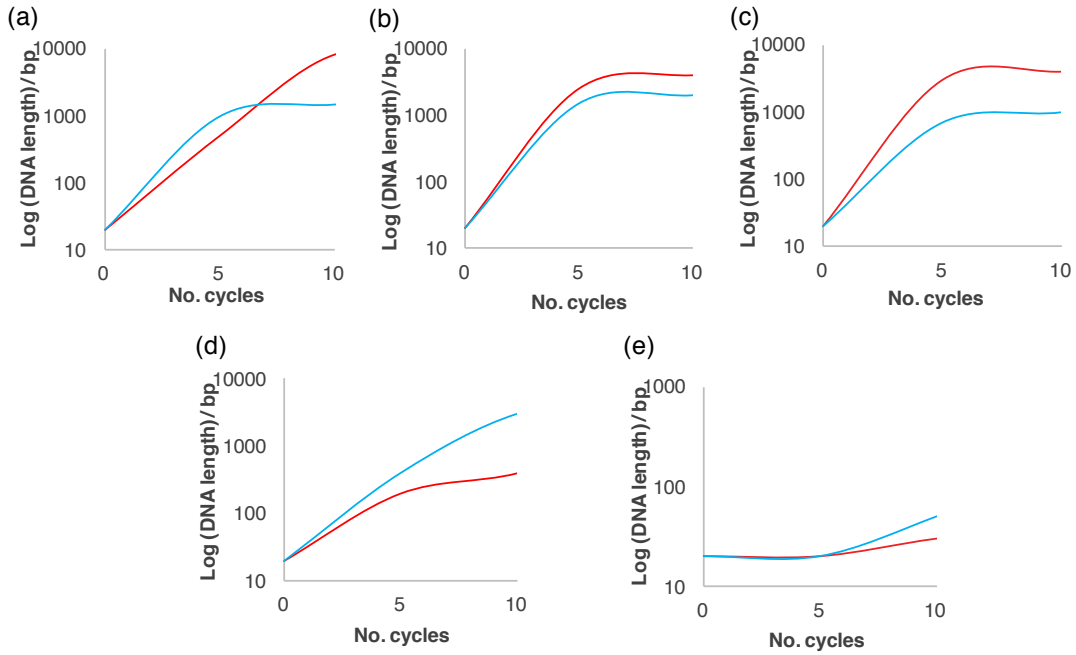


Fig. 14 DNA modal lengths determined from maximum fluorescence from the Image J analysis (Fig. 4-13) versus the number of heat-cool cycles, $[GATC]_5/[CTAG]_5$ (blue) and $[A_4G]_4/[T_4C]_4$ (red), **(a)** 5-I-dC-DNA, **(b)** 5-alkyne-C_8-dC-DNA, **(c)** 7-deaza-I-dA-DNA, **(d)** 5-Br-dU-DNA and **(e)** 6-S-dG-DNA.

3.2.1.4 *DNA extensions with 5-acetyl-Hg-dCTP*

Having established a method to incorporate artificial dNTPs bearing common organic functionalities, it would be interesting to assess dNTPs modified with heavy metals to gain an avenue towards controllable DNA labelling and sequencing.[33] Extensions with 5-acetyl-Hg-dCTP were performed to study the DNA polymerase handling of heavy metal dNTP derivatives. However, only low levels of incorporation were observed for the extension of the $[GATC]_5/[CTAG]_5$ oligo seed, Fig. 15. 5-acetly-Hg-dCTP incorporation has been trialled previously using the slippage reaction and shows no extension is possible under standard isothermal polymerase conditions.[33] However, on addition of 2-mercaptoethanol, incorporation was observed.[33] If future DNA synthesis bearing mercury functionalised nucleobases was to be explored, an updated reaction solution containing 2-mercaptoethanol would be implemented.

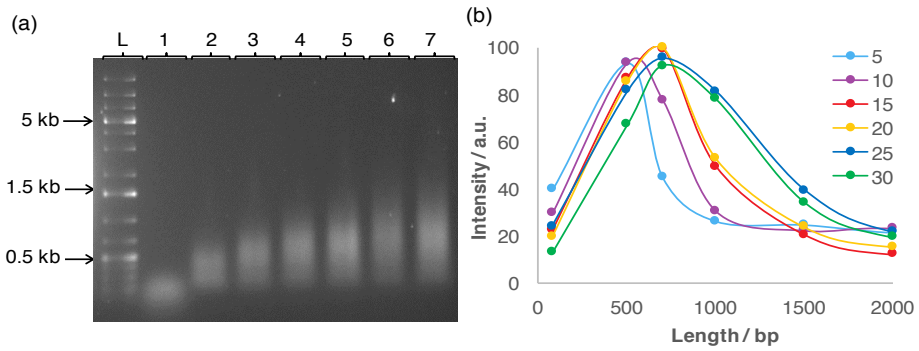


Fig. 15 (a) Agarose gel of 5-acetyl-Hg-dC-DNA extension products from $[GATC]_5/[CTAG]_5$ after 0, 5, 10, 15, 20, 25 and 30 cycles, lanes 1-7, respectively. **(b)** Image J analysis of (a) agarose gel, normalised to 100 for the maximum fluorescence of each lane.

3.2.1.5 DNA extension controls

It is conceivable, that the gel electrophoresis results discussed in section 3.2.1 above, could in fact be due to incorrect base incorporation by the DNA polymerase. An unmodified base could be inserted preferentially over the modified base or alternatively, it is possible that the dNTP reagents were impure. Control extensions were performed in the absence of each individual dNTP per reaction to identify any misincorporations or impurities. For each sequence and DNA polymerase, no extension was observed when the reaction was performed in the absence of any one of the four standard dNTPs, Fig. 16 (a-c). This confirms that no contamination was present in any of the dNTP samples and that the DNA polymerase does not perform incorrect base pairing incorporation.

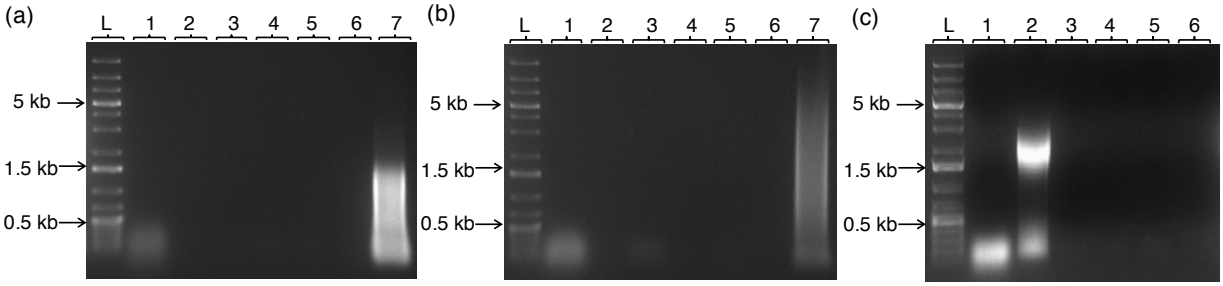


Fig. 16 DNA extenion controls with **(a)** $[GATC]_n/[CTAG]_n$ and **(b)** $[A_4G]_n/[T_4C]$ with Deep Vent exo- and **(c)** $[A_4G]_n/[T_4C]$ with Tgo-Pol Z3 exo-. Lane 1(a-c): starting oligo seed, lanes 3-6(a-c): negative controls performed in the absence of dTTP, dATP, dGTP and dCTP, respectively, lane 7(a-b) and lane 2(c): positive control performed in the presence of all four unmodified dNTPs and lane 2(a-b) blank.

3.2.2 Atomic force microscopy

The DNA products described in section 3.2.1 were only characterised by agarose gel electrophoresis, which, does not distinguish between the expected linear DNA products and other potential DNA structures, for example, concatamers between shorter DNA sequences or branched extension products. Ensuring this enzymatic DNA

synthesis method produces linear DNA is important, hence, AFM was performed for each of the $[A_4G]_n/[T_4C]_n$ samples after 30 cycles of extension. Each AFM image shows the expected topographical image of dsDNA aligned on a mica surface, Fig. 17.[34] Additionally, length statistics of the DNA products are in good agreement with those from agarose gel electrophoresis, Table 1. DNA heights are likewise in agreement with literature guidelines[34] – profile scans were taken perpendicular to the linear DNA (Appendix B). Although each AFM image depicts singular linear molecules of DNA, a variety of different curvatures are notable. Standard dsDNA has a persistence length of less than 43 nm in the presence of 100 mM Na^+ counterions.[35] 1 mM Mg^{2+} were used for this study, therefore, a shorter DNA persistence length is expected which could not be observed using the AFM parameters operated here. However, the 7-deaza-I-dA-DNA has a prominently more rigid structure than the other modified DNA molecules produced, Fig. 17 (c). Further studies could expose distortions in the dsDNA structure and rigidity on inclusion of multiple modified bases to the DNA sequence. On account of the AFM images, we can confirm the synthesis of linear 1-D-dsDNA.

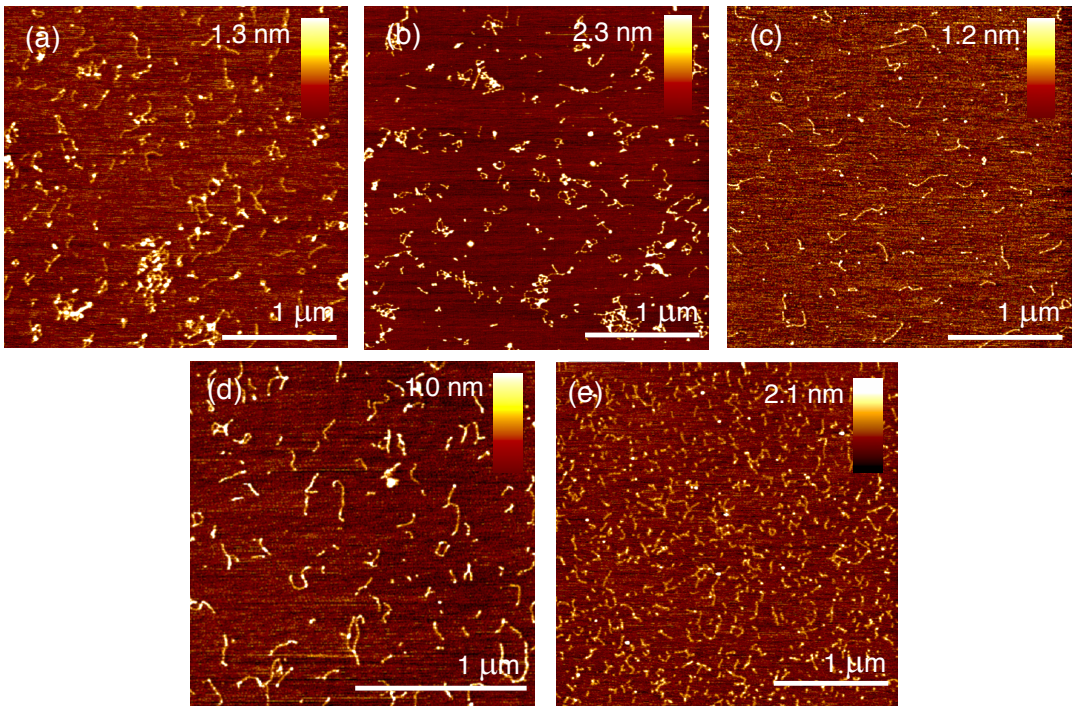


Fig. 17 AFM of modified $[A_4G]_n/[T_4C]_n$ DNA extension products after 30 cycles, **(a)** 5-I-dC-DNA, **(b)** 5-C_8-alkyne-dC-DNA, **(c)** 7-deaza-I-dA-DNA, **(d)** 5-Br-dU-DNA, **(e)** 6-S-dG-DNA. Additional examples of each AFM image are shown in Appendix B.

	I-dC	I-dA	Br-dU	Alkyne-dC	S-dG
Average (nm)	907	798	958	871	326
SD (nm)	128	191	231	180	46

Table 1 DNA average length summary calculated from AFM images in Fig. 17. A sample of DNA lengths are shown in Appendix B.

3.2.3 DNA sequencing

Enzymatic synthesis of long linear DNA with artificial dNTPs has been shown by agarose gel electrophoresis and AFM, however, the Tgo-Pol Z3 exo- DNA polymerase chosen for this project has attributes that induce an increased error rate. To ensure the elevated DNA polymerase promiscuity did not result in incorrect base pairing, DNA sequencing was performed (GATC Biotech). Sequencing of $[A_4G]_n/[T_4C]_n$ DNA extension products yielded results in perfect agreement with the oligo seed sequence, Fig. 18 (a) – (e). Submission of $[GATC]_n/[CTAG]_n$ extended DNA samples did not yield any sequencing data – sequencing of repetitive sequences is known to be challenging and as $[GATC]_n/[CTAG]_n$ is self-complementary, additional difficulty is expected. In order to obtain sequencing data, representative of the $[GATC]_n/[CTAG]_n$ sequence, $[GAATC]_n/[GATTC]_n$ was adopted and produced sequencing data in equally good agreement with the oligo seed sequence, Fig. 18 (f) – (j). Therefore, the correct incorporation of dNTPs according to the Watson and Crick base pairing rules was maintained throughout the above oligo seed extensions.

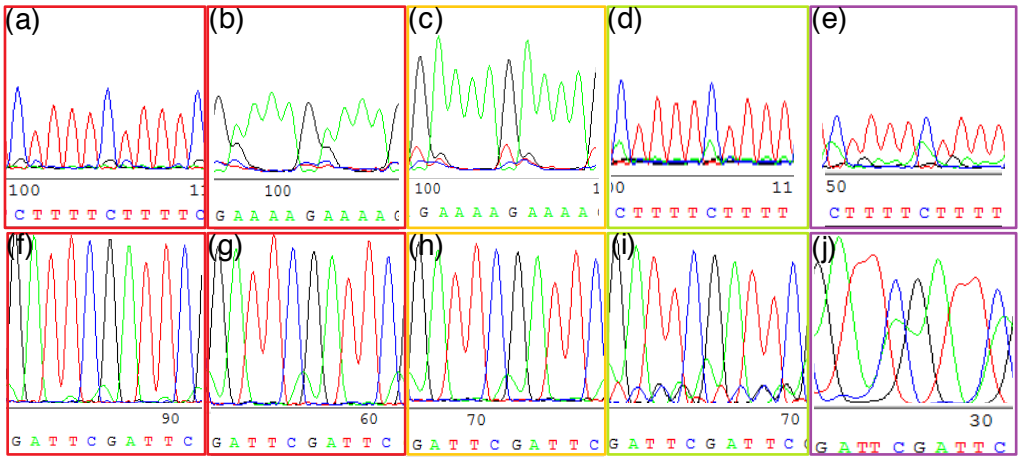


Fig. 18 DNA sanger sequencing of modified $[A_4G]_n/[T_4C]_n$; **(a)** 5-I-dC-DNA, **(b)** 5-C_8-alkyne-dC-DNA, **(c)** 7-deaza-I-dA-DNA, **(d)** 5-Br-dU-DNA, **(e)** 6-S-dG-DNA. and modified $[GAATC]_n/[CTTAG]_n$; **(f)** 5-I-dC-DNA, **(g)** 5-C_8-alkyne-dC-DNA, **(h)** 7-deaza-I-dA-DNA, **(i)** 5-Br-dU-DNA, **(j)** 6-S-dG-DNA. Longer regions of sequencing data are found in Appendix B.

3.2.4 DNA digestions and characterisations

Thus far, the data has shown promise for a method capable of designing and synthesising DNA by sequence and functional content. For absolute confirmation of modified dNTP incorporation, DNA digestions were performed to analyse the constituent nucleosides. Analysis of the DNA digested products was performed by high performance liquid chromatography (HPLC) and each peak was comparable to dN standards (Appendix B). The HPLC traces of each digested extension product confirm

the presence of the modified nucleoside and absence of the natural derivative, Fig. 19 (a-e). A slight shoulder is noted in the 6-S-dG-DNA digested products, which compares to the dG present in the digested oligo seed, Fig. 19 (f). As the 6-S-dG-DNA product is considerably shorter than the other modified-DNA products (300 bp versus 2,000 bp), the unmodified dG in the starting oligo seed will be present in a larger ratio and so can be seen in the chromatogram.

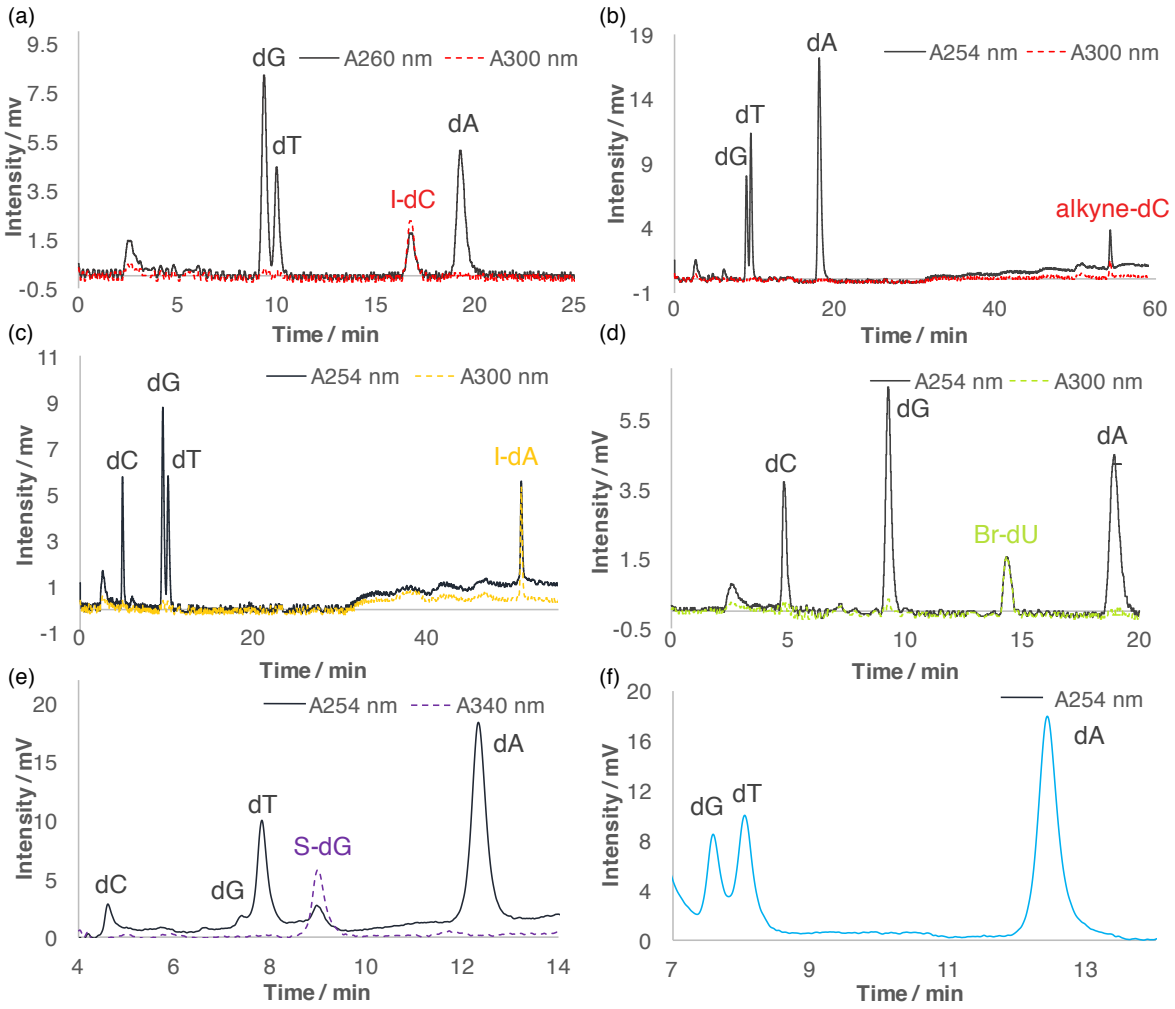


Fig. 19 HPLC traces of DNA digestion products from **(a)** 5-I-dC-$[A_4G]_n/[T_4C]_n$, **(b)** 5-alkyne-C_8-dC-$[A_4G]_n/[T_4C]_n$, **(c)** 7-deaza-I-dA-$[A_4G]_n/[T_4C]_n$, **(d)** 5-Br-dU-$[A_4G]_n/[T_4C]_n$, **(e)** 6-S-dG-$[A_4G]_n/[T_4C]_n$ and **(f)** $[A_4G]_4/[T_4C]_4$ oligo seed.

To further confirm the presence of each artificial base, each peak at the corresponding retention time of the modified nucleobase standards was collected for additional analysis. Each collected fraction showed comparable UV-Vis data, Fig. 20 and the corresponding mass in high resolution mass spectroscopy (HRMS) to the relevant standard, Table 2. Therefore, confidence is given to the correct incorporation of each artificial dNTP into the dsDNA whilst maintaining the correct repeating sequence.

Collected dN	HRMS
5-I-dC	$[M-H]^+$ calculated for $[C_9H_{13}IN_3O_4]^+$ 353.9951 found 353.9951
7-deaza-I-dA	$[M-H]^+$ calculated for $[C_{10}H_{13}IN_5O_3]^+$ 376.9985 found 377.0103
6-S-dG	$[M-H]^+$ calculated for $[C_{10}H_{14}N_5O_3S]^+$ 284.0817 found 284.0815
5-alkyne-C_8-dC	$[M-H]^+$ calculated for $[C_{17}H_{22}N_3O_4]^+$ 332.1610 found 332.1572
5-Br-dU	$[M-H]^+$ calculated for $[C_9H_{12}BrN_2O_5]^+$ 306.9930 found 306.9862.

Table 2 HRMS product peaks of collected modified dNs from the HPLC traces. Each dN found mass is compared to the calculated mass.

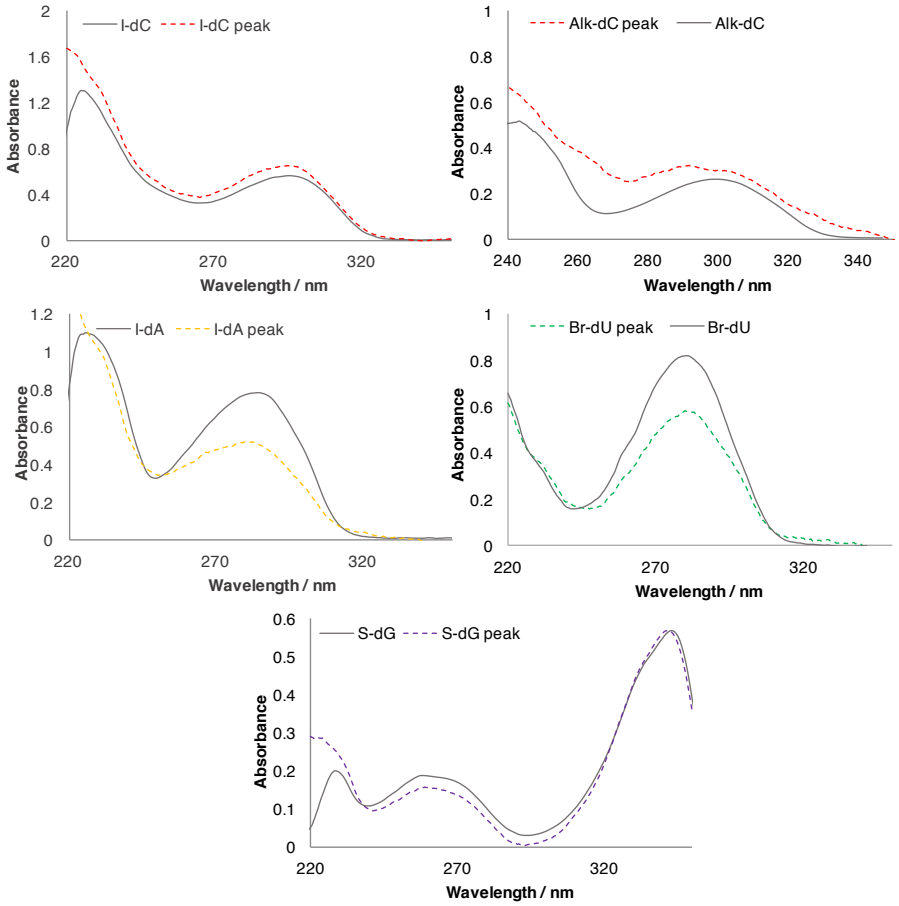


Fig. 20 UV-Vis spectrums of collected modified dN peaks (dashed line) compared to the corresponding standard (solid line).

Having proven incorporation of a single type of artificial dNTP per reaction to produce multiply modified DNA, extensions involving more than one type of modification were investigated to enhance the scope for the synthesis of designer DNA.

3.2.5 DNA extension reactions incorporating multiple types of modification

Since 1-type of artificial base per extension reaction had been established, to add to the designer DNA tool-box, 2- to 4-types of modification in one extension pot was assessed. The 2-type incorporation of both 5-I-dCTP and 5-Br-dUTP into a single DNA duplex is observed by agarose gel electrophoresis, Fig. 21, showing a similar extension pattern to the 1-type extension reactions. Likewise, with both 5-alkyne-C_8-dCTP and 5-Br-dUTP – final DNA products after 30 heat-cool cycles are equivalent in length as the 1-type extension product of the limiting artificial base, 5-Br-dUTP, Fig. 23. $[A_4G]_n/[T_4C]_n$, shows reduced extension due to the limiting repetitive incorporation of 5-Br-dUTP, however, the $[GATC]_n/[CTAG]_n$ does not contain the repetitive dT sequence, hence has equal product length to both 1-type extensions. DNA sequencing verifies the correct base incorporation and AFM confirms linear DNA synthesis is maintained as the number of modifications per sequence is increased, Fig. 21 (c) – (e).

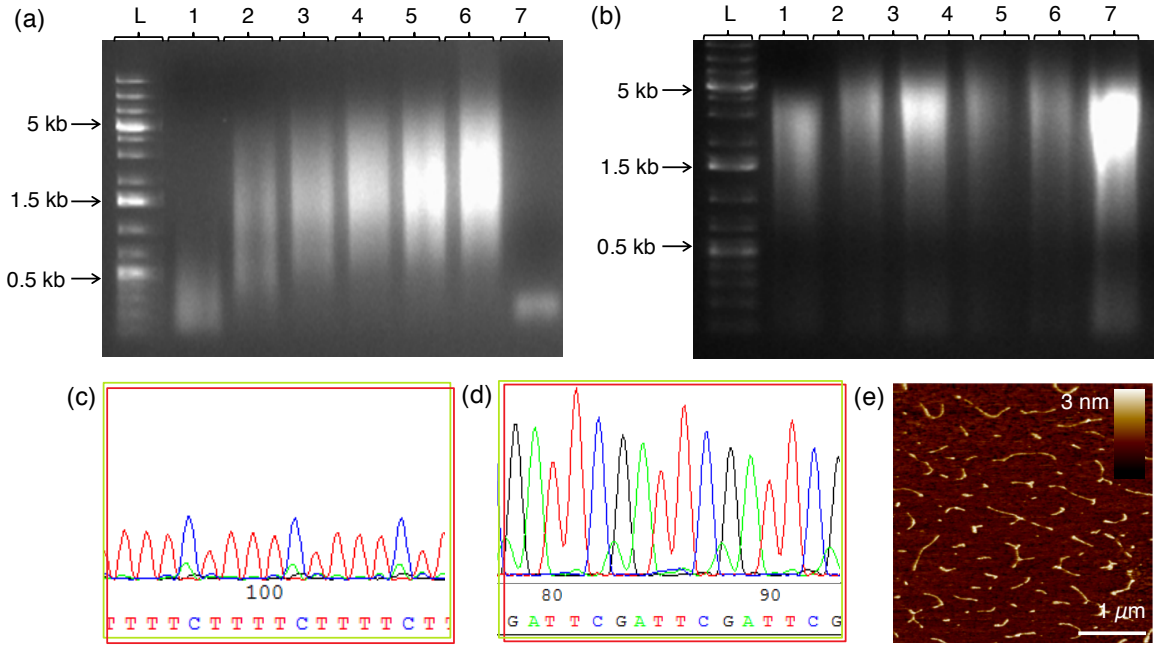


Fig. 21 DNA extension analysis of products after extension in the presence of 5-I-dCTP, 5-Br-dUTP, dGTP and dATP. **(a)** DNA extension products after 5, 10, 15, 20, 25 and 30 cycles with Tgo-Pol Z3 exo- and $[A4G]_n/[T4C]_n$, **(b)** DNA extension products after 5, 10, 15, 20, 25 and 30 cycles with Tgo-Pol Z3 exo- and $[GATC]_n/[CTAG]_n$, **(c)** DNA sequencing of the 5-I-dC, 5-Br-dU-products after 30 cycles with $[A_4G]_n/[T_4C]_n$. **(d)** DNA Sanger sequencing of DNA extension products of 5-alkyne-C_8-dC, 5-Br-dU-product after 30 cycles, **(e)** AFM of 5-I-dC, 5-Br-dU-products after 30 cycles.

3- and 4- types of modifications per extension reaction are also possible and again, show final DNA product lengths similar to the limiting artificial base, Fig. 22 (a) – (d). A similar lag phase is visualised on the agarose gels of any the extension products from

$[A_4G]_4/[T_4C]_4$ where 6-S-dGTP has been incorporated. This delay in extension corresponds closely to the 1-type modification reaction for the incorporation of 6-S-dGTP. DNA sequencing of each 3- and 4-type modification extension products was possible to confirm that the expected fidelity is maintained by the DNA polymerase, Fig. 22 (e) – (h).

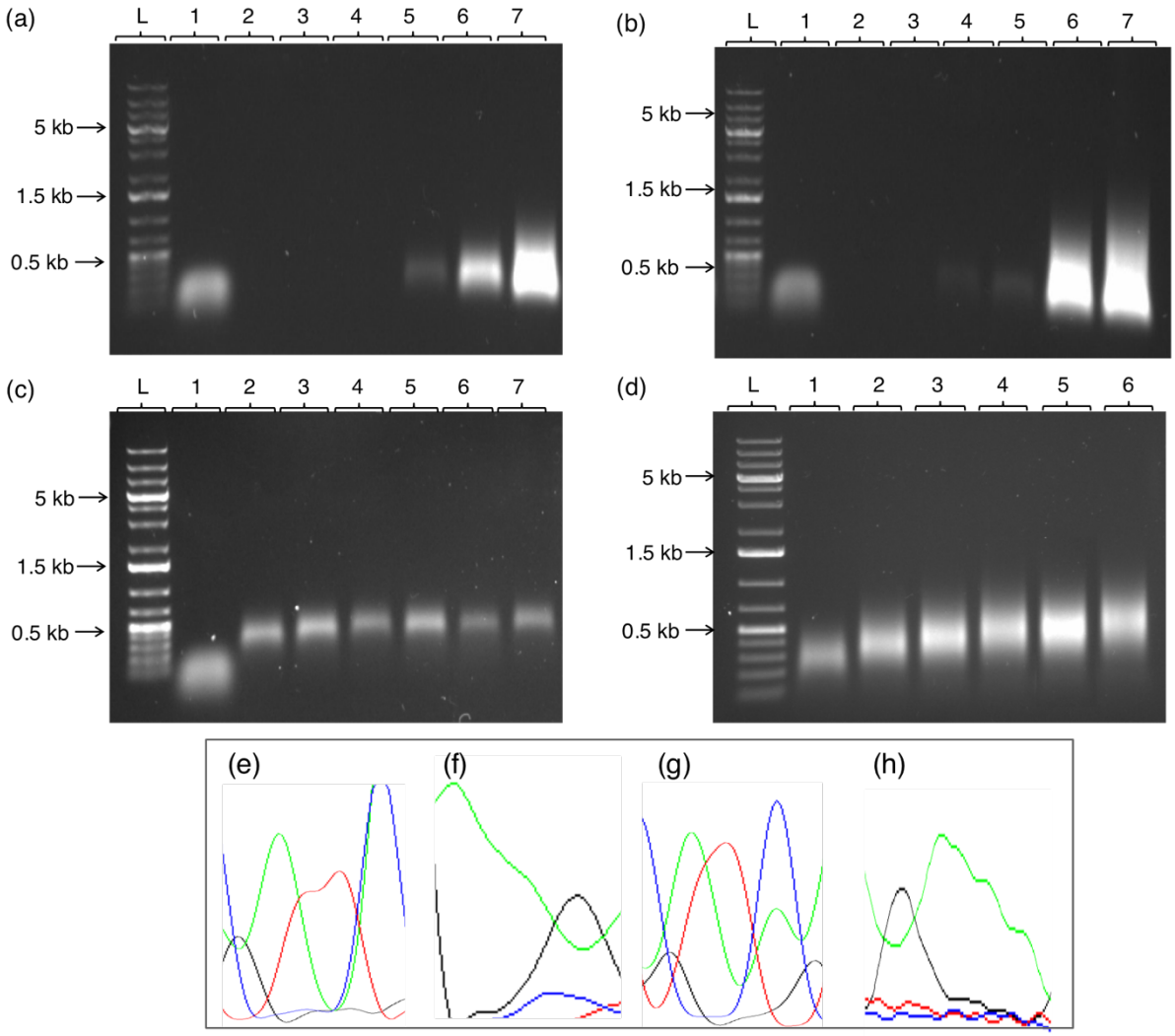


Fig. 22 Agarose gels of DNA extension products of **(a)** and **(c)** 6-S-dG, 5-Br-dU, 7-deaza-I-dA and 5-C_8-alkyne-dC modified $[A_4G]_n/[T_4C]_n$ and $[GATC]_n/[CTAG]_n$, respectively **(b)** and **(d)** 6-S-dG, 5-Br-dU and 7-deaza-I-dA modified $[A_4G]_n/[T_4C]_n$ and $[GATC]_n/[CTAG]_n$, respectively. DNA extension products were anayalyed after 5, 10, 15, 20, 25 and 30 cycles. Extension was performed using Tgo-Pol Z3 exo-. DNA sequencing of 6-S-dG, 7-I-dA and 5-Br-dU modified **(e)** $[GAATC]_n[GATTC]_n$ and **(f)** $[A_4G]_n/[T_4C]_n$. 6-S-dG, 7-I-dA, 5-C_8-alkyne-dC and 5-Br-dU modified **(g)** $[GAATC]_n[GATTC]_n$ and **(h)** $[A_4G]_n/[T_4C]_n$.

The modified DNA product lengths after 30 heat-cool cycles using Deep Vent exo- instead of Tgo-Pol Z3 exo- as the DNA polymerase depicted the same trend of extension efficiency for all oligo seeds and modified nucleotides, Fig. 23. Therefore, this method of extension is also feasible with a commercially available enzyme. Additionally, extensions with $[GATTC]_4/[CTAAG]_4$ as the oligo seed, Fig. 23 (c), yielded

DNA product lengths similar to $[GATC]_4/[CTAG]_4$, for example, extension with 7-deaza-I-dATP produced modal lengths of 1,500 bp. $[GATTC]_n/[CTAAG]_n$ was used as the representative sequence for $[GATC]_n/[CTAG]_n$ to perform DNA sequencing, therefore similar extension profiles confim its acceptable use as a replacement sequence, Fig. 23 (c).

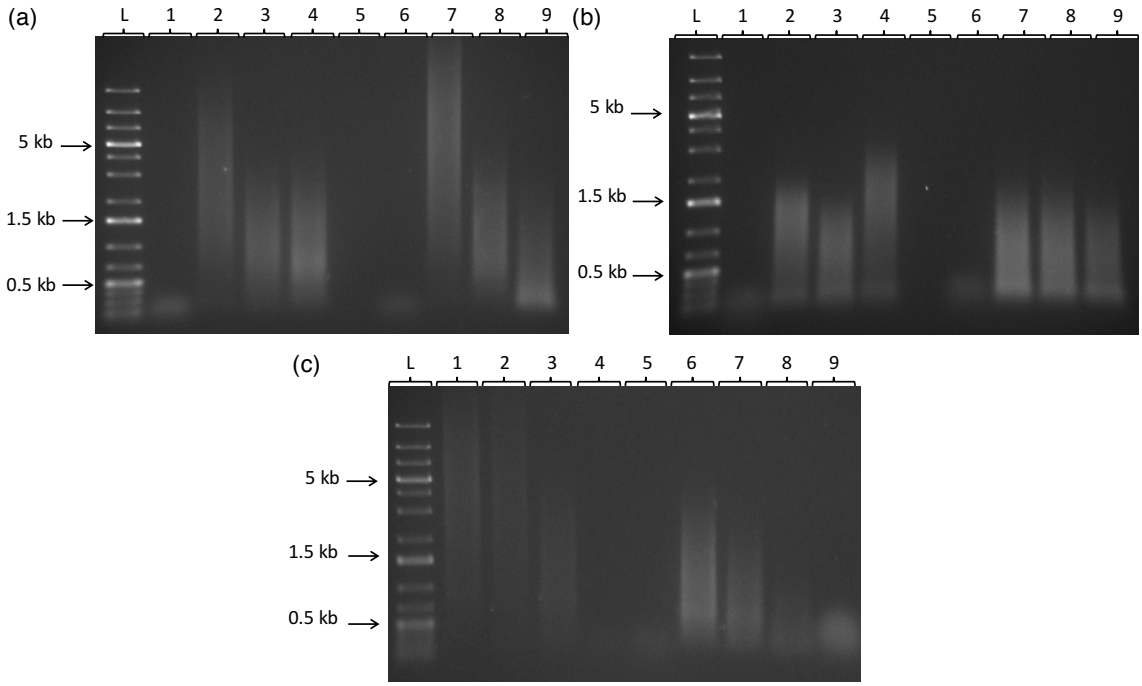


Fig. 23 Agarose gels of DNA extension products with Deep Vent exo- and **(a)** $[A_4G]_n/[T_4C]_n$, **(b)** $[GATC]_n/[CTAG]_n$ and **(c)** $[GATTC]_n/[CTAAG]_n$ after 30 cycles. Lane 1 (a,b) and lane 9 (c) depict the oligo seed used for each extension assay. Lanes 2-9 (a, b) and lanes 1-8 (c), depict extensions with 5-I-dCTP, 5-Br-dUTP, 7-deaza-I-dATP, 5-acetly-Hg-dCTP, 6-S-dGTP, 5-C_8-alkyne-dCTP, 5-C_8-alkyne-dCTP and 5-Br-dUTP, 5-C_8-alkyne-dCTP, 5-Br-dUTP and 7-deaza-I-dATP, respectively.

Table 3 summarises the incorporation efficiencies of each modified nucleotide against the rate of extension using standard dNTPs, as considered in Chapter 2. As discussed, each 1-type modification reaction exhibited a similar extension profile to unmodified dNTP incorporation, except in the case of 6-S-dGTP. Modifications situated in the major groove appear to have little effect on DNA polymerase handling for correct incorporation to produce dsDNA. Modifications that interfere with the base pairing hydrogen bonds lead to a decrease in DNA stability and thus a reduction in extension lengths. This trend is consequently seen in 2-, 3- and 4-type modification reactions – where length is limited by the least efficiently incorporated dNTP. However, DNA extension is still observed with multiple modified dNTPs in the reaction. Therefore, here the enzymatic synthesis of modified DNA has been successfully demonstrated and the

potential for variations in the sequence, length and now the number of modifications is attractive for the realisation of truly designer DNA.

Modified dNs	$[GATC]_n/[CATG]_n$	$[A_4G]_n/[T_4C]_n$
	Grade	Grade
I-dC	◎	◎
≡-dC	◎	◎
I-dA	◎	◎
Br-dU	◎	○
S-dG	○	△
I-dC, Br-dU	◎	◎
Br-dU, ≡-dC	○	○
≡-dC, Br-dU, I-dA	○	○
S-dG, Br-dU, I-dA	○	△
S-dG, Br-dU, I-dA, ≡-dC	○	△
No modification	◎	◎

Table 3 Extension efficiency summary, comparing each 1-, 2-, 3-, and 4-type modified dNTP extension pot to 4 unmodified dNTPs. DNA product length after 30 heat-cool cycles, ◎ represents DNA products > 2 kb, ○ > 500 bp and △ > 20 bp.

3.3 Conclusions

An enzymatic DNA synthesis method has been reported with the added amenability and control over the sequence and functional content. Single atom modifications, 5-I-dCTP, 7-deaza-I-dATP, 5-Br-dUTP and 6-S-dGTP and a chain modification 5-C_8-alkyne-dCTP were examined to determine DNA polymerase ability to incorporate these dNTPs efficiently into dsDNA using the heat-cool cycle method. Two oligo seeds, $[A_4G]_n/[T_4C]_n$ and $[GATC]_n/[CTAG]_n$ were selected to establish the extension profile and limits of this desinger DNA synthetic method. Extensions with the four dNTPs, where modifications are positioned in the major groove, showed little deviation in extension profile when compared to unmodified dNTP incorporation. 5-Br-dUTP, however, did show a slight reduction when consecutively incorporated into the $[A_4G]_n/[T_4C]_n$ sequence. Conversly, 6-S-dGTP, a modification altering the base pairing region, had a largely reduced extension profile. On expanding to 2-, 3- and 4-type modification extensions, this trend continued. No added reduction in extension rate was observed on increasing the number of artifical dNTPs, however, product length was limited by the least efficiently incorporated dNTP.

In summary, an enzymatic DNA synthetic method has been established which provides freedom to design the repeating sequence, functional content, and length. To demonstrate that this highly designed DNA has potential as a tool in the fabrication of nanomaterials, metal binding and organic chemistry at the modified dN sites is described in Chapter 4 and Chapter 5 respectively.

3.4 Experimental details

Preparation and Purification of DNA polymerases

Tgo-Pol Z3 exo- was prepared and purified as described previously.[29,36] Deep Vent exo- was purchased from New England Biolabs (Herts, UK).

Deoxynucleotide preparation

Deoxyoligonucleotides were purchased from Eurofins (Ebersberg, Germany) and complementary strands were annealed as shown in Table 5.

Oligos to be combined	Oligo seed	Oligo seed length / bp
$(GATC)_5$ and $(CATG)_5$	$[GATC]_5/[CATG]_5$	20
$(AAAAG)_4$ and $(TTTTC)_4$	$[AAAAG]_4/[TTTTC]_4$	20
$(GATTC)_4$ and $(GAATC)_4$	$[GATTC]_4/[GAATC]_4$	20

Table 5 Summary of oligos required to form the oligo seeds used.

Heat-cool cycle DNA extension

0.5 μM oligo seed (Table 5), 200 nM DNA polymerase or 1 unit, DNA polymerase reaction buffer (200 mM Tris-HCl (pH 8.8, 25 °C), 100 mM $(NH_4)_2SO_4$, 100 mM KCl, 1 % Triton X-100, 1 mg/mL Bovine Serum Albumin and 20 mM $MgSO_4$) or 10 x ThermoPol buffer, and 0.5 mM dNTPs (Table 6) were mixed. Thermocycling was carried out using an Applied Biosciences Veriti 96 well Thermal Cycler by the following method:

N (number of cycles stated) x 30 seconds at 95 °C, 30 seconds at 55 °C and 120 seconds at 72 °C.

The products were cooled to 4 °C after the reaction and the product was then purified using a QIAquick PCR purification kit (25) (QIAGEN, Manchester, UK) following manufacturers protocol.

Modified DNA	dNTPs present
I-dC-DNA	dGTP, 5-I-dCTP, dATP, dTTP
I-dA-DNA	dGTP, 7-deaza-I-dATP, dATP, dTTP
Br-dU-DNA	dGTP, dCTP, dATP, 5-Br-dUTP
Hg-dC-DNA	dGTP, 5-acetyl-Hg-dCTP, dATP, dTTP
Alkyne-dC-DNA	dGTP, C8-alkyne-dCTP, dATP, dTTP
S-dG-DNA	6-S-dGTP, dCTP, dATP, dTTP
I-dC, Br-dU-DNA	dGTP, dCTP, dATP, dTTP
Alkyne-dC, Br-dU-DNA	dGTP, 5-C_8-alkyne-dCTP, dATP, 5-Br-dUTP
Alkyne-dC, Br-dU, I-dA-DNA	dGTP, 5-C_8-alkyne-dCTP, 7-deaza-I-dATP, 5-Br-dUTP

Table 6 dNTPs present in each modified DNA reaction mixture.

Agarose Gel Electrophoresis

The extension products were analysed by gel electrophoresis in TBE (Tris, boric Acid and $Na_2EDTA.2H_2O$) buffer. 0.75 % agarose (Melford, Ipswich, UK) was added to the 1 x TBE buffer and heated to dissolve. The 0.75 % agarose solution was supplemented with ethidium bromide (Sigma Aldrich) and poured once it had cooled to 50 °C. The 1 kb+ ladder was purchased from Thermo Scientific and provided with a loading dye (2.5 % Ficoll-400, 11 mM EDTA, 3.3 mM Tris-HCl (pH 8.0, 25 °C), 0.017 % SDS and 0.015 % BPB). DNA samples were supplemented with the gel loading dye. The gels were run at 100 V, 100 mA, 10 W for approximately 1 hour and then visualised using an ultra-violet transilluminator.

Atomic force microscopy imaging

The top layer of the mica surface was cleaved using sticky tape. The mica surface was placed at 30° to the bench and 5 μL of the DNA sample (1.5 ng / μL) in 1 mM $MgCl_2$ was dropped onto the surface using a micro pipette After 5 minutes, 5 μL of H_2O was then dropped onto the surface whilst maintaining the angle. Nitrogen gas was passed over the surface to straighten the DNA and allowed drying for 1 hour. A light microscope was used to locate the sample and cantilever position. AFM images were collected using a Dimension V with a nanoscope controller (Veeco Instruments Inc., Metrology Group, Santa Barbara, CA). The tapping mode was used with an etched silicon tip (Tap 300 Al-G, 300 kHz, 40 N/m) on an isolation table (Veeco Inc., Metrology Group) to reduce interference. Nanoscope 7.00b19 software was used to acquire data.

Ultraviolet-Visible spectroscopy

UV-Vis spectroscopy was performed using a NanoDrop 1000 Spectrometer (Thermo Scientific for volumes less than 2 μL and a Varian Cary 100Bio UV-Vis spectrophotometer with a Varian Cary temperature controller for larger volumes. Spectrometer was blanked using nanopure-H_2O or QIAGEN elusion buffer. Concentration and peak determination was performed as shown on Table 4.

DNA digestion

0.2 mg of snake venom phosphodiesterase, 100 units bacterial alkaline phosphotase in 10 mM potassium phosphate buffer, pH 7 and 10 mM magnesium chloride was added to 0.5 absobance units at 260 nm of dsDNA and incubated at 37 °C for 16-18 hours. The dNMPs produced can be analysed by reverse phase HPLC on an

APEX ODS C18 5um 250 mm column with 0.1 M triethylammonium acetate, pH 6.5 containing 5 % ACN (buffer A) and 0.1 M triethylammonium acetate, pH 6.5 containing 65 % ACN (buffer B) operated at 1 $mLmin^{-1}$ at room temperature. 100 μL of digestion product was injected into a Waters 2487 Dual Wavelength Absorbance detector with a Waters 600 controller under the buffer system described in Table 7.

Time / min	Buffer A %	Buffer B %
0	100	0
25	5	95
30	10	90
40	15	85
45	20	80
50	25	75
55	30	70
60	35	65
70	100	0

Table 7 HPLC buffer system

Mass spectroscopy

A Waters Micromass LCT Premier TOF system was used in positive mode, with a desolvation temperature of 250 °C and data collected using masslynx v. 4.1. Samples were injected directly in nanopure-H_2O.

dN standards

All standards were purchased from Jena Biosciences (Jena, Germany) except 5-C_8-alkyne-dC (5-(octa-1,7-diynyl)-cytosine) which was synthesised in house following the protocol by Chittepu *et al.*[37] 72 %, TLC (CH_2Cl_2/MeOH 9:1): R_f 0.41. UV-Vis λ_{max} (MeOH)/nm 299. ^{1}H NMR ((D_6) DMSO): 1.52-1.67 (m, 4H $2CH_2$), 1.96-2.17 (m, 2H CH_2(2')), 2.19-2.23 (m, 2H CH_2), 2.42-2.45 (t, *J*=2.44 CH_2), 2.78 (t, 1H C≡CH), 3.54-3.65 (m, 2H CH_2(5')), 3.78-3.81 (m, HC(4')), 4.20-4.24 (m, HC(3')), 5.06-5.09 (t, *J*=5.08 OH-C(5')), 5.23-5.24 (d, *J*=5.23 OH-C(3')), 6.11-6.14 (t, *J*=6.13 H-C(1')), 6.74 (s, NH_a), 7.69 (s, NH_b), 8.08 (s, H-C(6)). ^{13}C NMR ((D_6) DMSO): 17.7 (CH_2-C≡H), 19.0 (CH_2-C≡H), 27.6 (CH_2CH_2-C≡H), 27.7 (CH_2CH_2-C≡H), 39.9 (C-2'), 61.5 (C-5'), 70.6 (C-3'), 71.8 (C≡CH), 72.6 (C≡C), 84.8 (C-1'), 85.7 (C-4'), 87.9 (C≡C), 90.8 (C≡C), 95.8 (C-5), 144.0 (C-6), 154.0 (C-2), 164.8 (C-4). HRMS (ESI +ve) calcd. for $C_{17}O_4N_3H_{21}$ $[M-H]^+$ 332.1610, found 332.1589.

3.5 References

1 Wang, X., Zhang, J., Li, Y., Chen, G. & Wang, X. Enzymatic synthesis of modified oligonucleotides by PEAR using Phusion and KOD DNA polymerases. *Nucleic Acid Ther* **25**, 27-34 (2015).

2 Macpherson, P. *et al.* 8-oxoguanine incorporation into DNA repeats in vitro and mismatch recognition by MutSalpha. *Nucleic Acids Res* **33**, 5094-5105 (2005).

3 Yuen, L. H., Franzini, R. M., Tan, S. S. & Kool, E. T. Large-scale detection of metals with a small set of fluorescent DNA-like chemosensors. *J Am Chem Soc* **136**, 14576-14582 (2014).

4 Wan, Y. *et al.* A surface-initiated enzymatic polymerization strategy for electrochemical DNA sensors. *Biosens Bioelectron* **41**, 526-531 (2013).

5 Tanaka, K. *et al.* Programmable self-assembly of metal ions inside artificial DNA duplexes. *Nat Nanotechnol* **1**, 190-194 (2006).

6 Mitomo, H., Watanabe, Y., Matsuo, Y., Niikura, K. & Ijiro, K. Enzymatic synthesis of a DNA triblock copolymer that is composed of natural and unnatural nucleotides. *Chem Asian J* **10**, 455-460 (2015).

7 Bentley, D. R. *et al.* Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* **456**, 53-59 (2008).

8 Bell, D. C. *et al.* DNA base identification by electron microscopy. *Microsc Microanal* **18**, 1049-1053 (2012).

9 Lercher, L., McGouran, J. F., Kessler, B. M., Schofield, C. J. & Davis, B. G. DNA modification under mild conditions by Suzuki-Miyaura cross-coupling for the generation of functional probes. *Angew Chem Int Ed Engl* **52**, 10553-10558 (2013).

10 Shigdel, U. K., Zhang, J. & He, C. Diazirine-based DNA photo-cross-linking probes for the study of protein-DNA interactions. *Angew Chem Int Ed Engl* **47**, 90-93 (2008).

11 Nakayama, S., Yan, L. & Sintim, H. O. Junction probes - sequence specific detection of nucleic acids via template enhanced hybridization processes. *J Am Chem Soc* **130**, 12560-12561 (2008).

12 Watson, J. D. & Crick, F. H. The structure of DNA. *Cold Spring Harb Symp Quant Biol* **18**, 123-131 (1953).

13 Shapiro, R. *et al.* Conformation of amine-modified DNA: 2-aminofluorene- and 2-(acetylamino)fluorene-modified deoxydinucleoside monophosphates with all possible nearest neighbors. A comparison of search and optimization methods. *Chem Res Toxicol* **7**, 239-253 (1994).

14 Froehler, B. C., Ng, P. G. & Matteucci, M. D. Synthesis of DNA via deoxynucleoside H-phosphonate intermediates. *Nucleic Acids Res* **14**, 5399-5407 (1986).

15 Kamath-Loeb, A. S., Hizi, A., Kasai, H. & Loeb, L. A. Incorporation of the guanosine triphosphate analogs 8-oxo-dGTP and 8-NH2-dGTP by reverse transcriptases and mammalian DNA polymerases. *J Biol Chem* **272**, 5892-5898 (1997).

16 Lauridsen, L. H., Rothnagel, J. A. & Veedu, R. N. Enzymatic recognition of 2'-modified ribonucleoside 5'-triphosphates: towards the evolution of versatile aptamers. *Chembiochem* **13**, 19-25 (2012).

17 Tanaka, K., Yamada, Y. & Shionoya, M. Formation of silver(I)-mediated DNA duplex and triplex through an alternative base pair of pyridine nucleobases. *J Am Chem Soc* **124**, 8802-8803 (2002).

18 Clever, G. H., Polborn, K. & Carell, T. A highly DNA-duplex-stabilizing metal-salen base pair. *Angew Chem Int Ed Engl* **44**, 7204-7208 (2005).

19 Takezawa, Y., Nishiyama, K., Mashima, T., Katahira, M. & Shionoya, M. Bifacial Base-Pairing Behaviors of 5-Hydroxyuracil DNA Bases through Hydrogen Bonding and Metal Coordination. *Chemistry* **21**, 14713-14716 (2015).

20 Watson, S. M. D., Pike, A. R., Pate, J., Houlton, A. & Horrocks, B. R. DNA-templated nanowires: morphology and electrical conductivity. *Nanoscale* **6**, 4027-4037 (2014).

21 Jawalekar, A. M. *et al.* Conjugation of nucleosides and oligonucleotides by [3+2] cycloaddition. *J Org Chem* **73**, 287-290 (2008).

22 Nakabeppu, Y. Cellular levels of 8-oxoguanine in either DNA or the nucleotide pool play pivotal roles in carcinogenesis and survival of cancer cells. *Int J Mol Sci* **15**, 12543-12557 (2014).

23 Burak, M. J., Guja, K. E. & Garcia-Diaz, M. Nucleotide binding interactions modulate dNTP selectivity and facilitate 8-oxo-dGTP incorporation by DNA polymerase lambda. *Nucleic Acids Res* **43**, 8089-8099 (2015).

24 Gutsmiedl, K., Fazio, D. & Carell, T. High-density DNA functionalization by a combination of Cu-catalyzed and cu-free click chemistry. *Chemistry* **16**, 6877-6883 (2010).
25 Sanger, F., Nicklen, S. & Coulson, A. R. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A* **74**, 5463-5467 (1977).
26 Sambrook, J., Fritsch, E. F. & Maniatis, T. *Molecular Cloning: A Laboratory Manual*. (Cold Spring Harbor Laboratory Press, 1977).
27 Datta, K., Wowor, A. J., Richard, A. J. & LiCata, V. J. Temperature dependence and thermodynamics of Klenow polymerase binding to primed-template DNA. *Biophys J* **90**, 1739-1751 (2006).
28 Staiger, N. & Marx, A. A DNA polymerase with increased reactivity for ribonucleotides and C5-modified deoxyribonucleotides. *Chembiochem* **11**, 1963-1966 (2010).
29 Jozwiakowski, S. K. & Connolly, B. A. A modified family-B archaeal DNA polymerase with reverse transcriptase activity. *Chembiochem* **12**, 35-37 (2011).
30 Biles, B. D. & Connolly, B. A. Low-fidelity Pyrococcus furiosus DNA polymerase mutants useful in error-prone PCR. *Nucleic Acids Res* **32**, e176 (2004).
31 Kong, H., Kucera, R. B. & Jack, W. E. Characterization of a DNA polymerase from the hyperthermophile archaea Thermococcus litoralis. Vent DNA polymerase, steady state kinetics, thermal stability, processivity, strand displacement, and exonuclease activities. *J Biol Chem* **268**, 1965-1975 (1993).
32 Somerville, L. *et al.* Structure and dynamics of thioguanine-modified duplex DNA. *Journal of Biological Chemistry* **278**, 1005-1011 (2003).
33 Dale, R. M., Livingston, D. C. & Ward, D. C. The synthesis and enzymatic polymerization of nucleotides containing mercury: potential tools for nucleic acid sequencing and structural analysis. *Proc Natl Acad Sci U S A* **70**, 2238-2242 (1973).
34 Thundat, T., Allison, D. P. & Warmack, R. J. Stretched DNA structures observed with atomic force microscopy. *Nucleic Acids Res* **22**, 4224-4228 (1994).
35 Brunet, A. *et al.* Dependence of DNA Persistence Length on Ionic Strength of Solutions with Monovalent and Divalent Salts: A Joint Theory-Experiment Study. *Macromolecules* **48**, 3641-3652 (2015).
36 Evans, S. J. *et al.* Improving dideoxynucleotide-triphosphate utilisation by the hyper-thermophilic DNA polymerase from the archaeon Pyrococcus furiosus. *Nucleic Acids Res* **28**, 1059-1066 (2000).
37 Seela, F., Sirivolu, V. R. & Chittepu, P. Modification of DNA with octadiynyl side chains: synthesis, base pairing, and formation of fluorescent coumarin dye conjugates of four nucleobases by the alkyne--azide "click" reaction. *Bioconjug Chem* **19**, 211-224 (2008).

Chapter 4.

Synthesis of Thiolated DNA and the Site Specific Coordination of Metal Species

Table of Contents

4.1 Introduction

The bioinspired bottom-up assembly of nanomaterials is an attractive approach to achieve a wide range of properties.[1,2,3] In particular, controlled metal ion deposition on the nanoscale has been highlighted for nanoelectronics.[4] For nanomaterial purposes, 1-D, 2-D and 3-D structures could provide additional function and properties, but require carefully designed synthetic strategies. 0-D materials (NPs) are fundamentally easier to fabricate and have a comprehensive research background.[5] For example, ferredoxin enzymes provide a route for metal ion mineralisation due to the biologically established pathway to sequester and store iron.[6] However, this only forms 0-D metal spheres, approximately 6 nm in diameter.[7]

The biological interactions between metal ions and DNA have been studied for many years due to their vital roles in biological systems.[8,9] Such studies have highlighted how certain metal ions destabilise the double helix, for example copper, and some stabilise, for example magnesium.[8] Nuclear Magnetic Resonance (NMR) demonstrated the sequence specific binding of several metal ions to dsDNA.[10] Nickel expressed a preference to bind between G-G and G-C sequences.[10] Manganese, nickel and cobalt showed binding interactions at the G-H8 and G-H1 positions.[10] This provided the motivation for multiple studies into the use of DNA as a template for metal deposition in 1-D, 2-D and 3-D nanomaterial synthesis.[11,12]

1-D DNA templating has been performed by many groups with several metal ions including Cu^{2+}, Pd^{2+} and Pt^{2+}.[13-15] Such DNA based nanowires have shown promise as semiconductors and conductors.[13-15] Ijiro's laboratory demonstrated a method for the controlled metallisation of DNA, dependant on the nucleobase composition.[16] The incorporation of 7-deaza-dGTP, alternating with the generic dGTP into dsDNA by the slippage reaction, provided specific binding regions for cisplatin – binding does not occur to 7-deaza-dG.[16] This advancement in DNA templated patterning provides an avenue to circuit gate synthesis. However, control over metallisation on the nanoscale is still limited as non-uniform deposition can often yield defects.[17] Therefore, nucleation control and a method to regulate deposition sites is required.

Many artificial nucleotides have been developed to extend the 4 letter alphabet of DNA and have been incorporated into oligos by phosphoramidite chemistry.[18-20] Some artificial nucleotides replace the base pairing region with metal binding ligands. The ligand affinity for specific metal ions can be exploited to tailor the magnetic and

electronic properties of the DNA hybrids.[21,22] Tanaka and Shionoya prepared a series of artificial ligands able to bind Pd^{2+},[19,23] Cu^{2+}[24] and B^{3+}.[25] Several other metal ions have since been specifically bound within the base pairing region of oligonucleotides, including Hg^{2+}[26] and Ag^{+}.[27,28] Carell *et al.,* designed an oligo sequence that can bind two metals, Cu^{2+} and Hg^{2+},[26] Fig. 1 (b). These results expand the capability to design functional content into the DNA structure through artificial bases. The metal-mediated base pairs can also induce an increased thermal stability in comparison to the Watson and Crick base pairing, Fig 1 (a).[27,29] By employing standard automated DNA synthesis procedures, the position of the metal ion within the duplex can be controlled. In instances where artificial nucleotides are situated at 'sticky ends', higher ordered structures can be assembled, such as DNA tiles.[30] However, the incorporation of the artificial bases discussed here is still restricted to short oligos. Therefore, a method able to incorporate site specific metallisation into long DNA would provide an avenue for the templating of metal ions onto DNA to design nanowires.

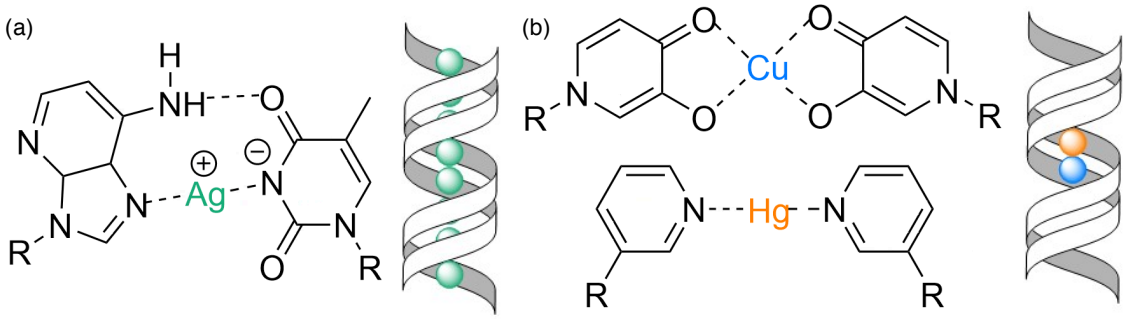


Fig. 1 Artificial base pairs of **(a)** 1-deaza-adenine and thymine[27] and **(b)** hydroxypyridone with Cu^{2+} and pyridine with Hg^{2+},[26] R = deoxyribose of a DNA backbone.

Alternatively, modified nucleosides have also been explored for metal ion interactions with DNA. The strong thio-gold interaction has been exploited for many years to attach DNA to gold surfaces through terminal DNA-thio-modification.[31] 2-deoxy-6-thio-guanosine (6-S-dG), can form highly ordered coordination complexes with metal ions, for example Co^{2+}.[32] The coordination structure depends on the metal ion chosen and whether the ribose sugar is with or without the 2' OH group, Fig. 2.[32] X-ray crystallography indicates two binding forms of thio-guanosine to Co^{2+}, the thiol and thione group are both present in the coordination structure, Fig. 2 (c) and (d). 1-D polymers have also been reported by the addition of Ni^{2+} ions to 6-S-G,[33] revealing a versatile interaction of the modified nucleoside with a variety of metal ions.

Modifications at the phosphate group have also been explored, for example, alpha-phosphorothioate-(α-S-)dUMP, Fig 2 (f).[34,35] Metal ions were added to

α-S-dUMP to study their binding strength compared to the phosphate derivative (dUMP). It was found that Zn^{2+} and Cd^{2+} formed more stable complexes with α-S-dUMP, however, the binding geometries were unknown.[34] Again, although the metal ion interaction with the nucleotide was observed, no evidence of metal coordination to long DNA bearing artificial groups was reported.

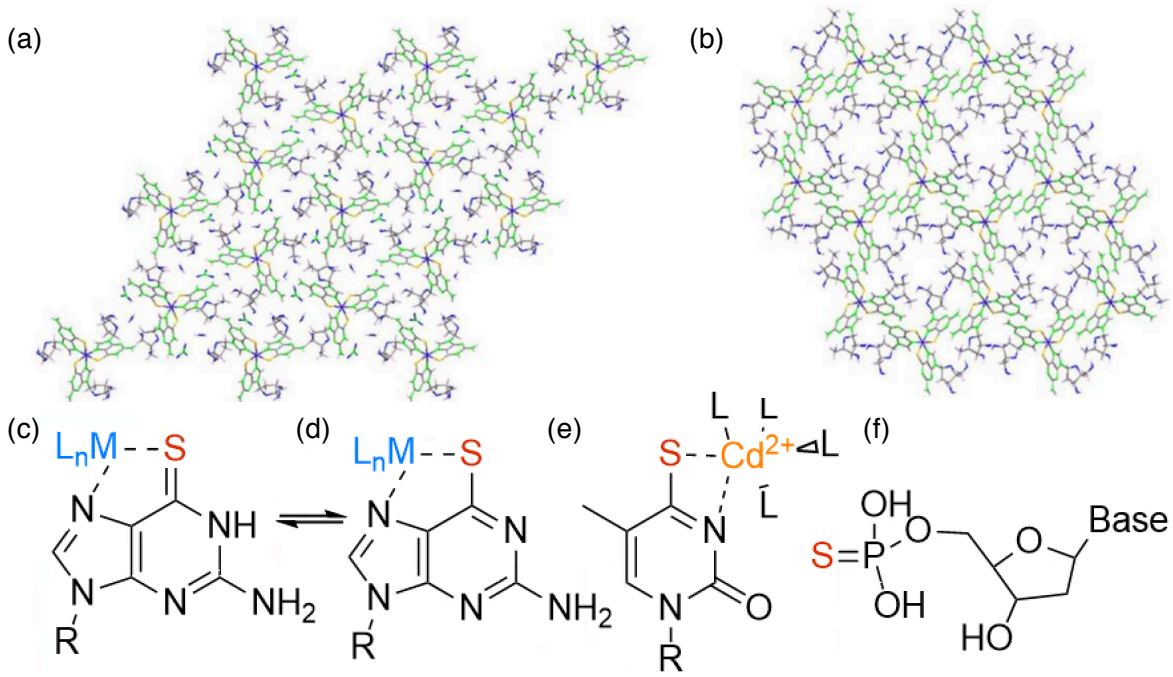


Fig. 2 (a) $[Co(2'\text{-d-6-thio-GH})^3]^{2+}$, **(b)** $[Co(6\text{-thio-G})^3]$, **(c)** 6-thione-guanosine, **(d)** 6-thiol-guanosine, **(e)** 4-thio-thymine, **(f)** α-S-dN, L = ligand, R = DNA.

An alternative route to metal DNA templating exploits DNA origami, folded using functionalised oligo staple strands.[36-39] In one example, the origami structure was designed to present regions of ssDNA, complementary to thio-functionalised oligos bound to metal NPs.[38] However, this approach only afforded control over relatively large regions (*ca.* 500 nm) of DNA.

Multiple α-S-dNTPs have been incorporated into long oligos (60 to 80 bp) to provide thio-anchors at specific separations via a bifunctional fastener.[40] This approach is attractive for the assembly of Au-NPs at specific separations and makes it possible to specifically functionalise DNA nanostructures with various nanoscale materials.

A method able to introduce multiple metal binding functional groups at controllable positions on long DNA (> 500 bp) would provide a synthetic tool towards advances of 1-D electronic and photonic nanostructures. The following section describes some progress in this respect.

4.2 Results and discussion

In Chapter 2, a method to extend repeat sequence DNA where base position and DNA length is controlled was described. Chapter 3 outlined an adaptation to the heat-cool extension method to allow for the precise incorporation of multiple non-standard nucleotides into DNA. Here, the enzymatic incorporation of thiolated dNTPs was investigated with the aim to control metal deposition along the DNA duplex. The range of oligo seeds, the type of thiolated dNTPs and the number of incorporations were all examined as variables in the design of DNA hybrid molecules.

4.2.1 *6-S-dG-DNA*

As discussed in Chapter 3, the Tgo-Pol exo- mutant, Z3, is able to incorporate 6-S-dGTP into dsDNA. Therefore, the potential to design DNA bearing specific sites of thiolation is possible by careful tailoring of the oligo seed sequences.

4.2.1.1 *Incorporation of 6-S-dGTP into repeat unit DNA*

To assess the flexibility of the heat-cool cycle method with 6-S-dGTP, a larger range of repeat unit sequences were analysed to determine possible spacing and control of thio-positioning. Several oligo seeds were designed to assess the dependence of extension on the separation between nearest neighbour G-bases; $[G]_{20}/[C]_{20}$, $[AG]_{10}/[TC]_{10}$, $[A_2G]_7/[T_2C]_7$, $[A_3G]_4/[T_3G]_5$, $[A_4G]_4/[T_4G]_4$ and $[A_9G]_2/[T_9C]_2$. These oligo seeds allow the positioning of the thio-modification every base and every 2^{nd}, 3^{rd}, 4^{th}, 5^{th} and 10^{th} base. Extension with $[GATC]_5/[CTAG]_5$ was also performed to show that a mixed base repeat sequence is compatible with this method. Extension efficiency increased on decreasing 6-S-dG frequency, peaking at $[A_3G]_4/[T_3G]_5$, as seen by gel electrophoresis data, Fig. 3 (a). In Chapter 3, this was attributed to the reduced stability of the DNA duplex upon 6-S-dGTP incorporation through less favourable hydrogen bonding to dC. Therefore, oligo seed sequences bearing > 25 % thio-groups are expected to have reduced extension. The reduced extension trend for oligo seeds bearing < 25 % thio-groups, is due to the diminishing extension of the longer repeat sequence rather than kinetics of 6-S-dGTP incorporation. As each repeat sequence starts as a 20 bp oligo seed, sequences with longer repeats have a lower probability of forming a shifted structure on annealing, therefore they may take more cycles to begin extension. Even though the $[A_3G]_5/[T_3G]_5$ oligo seed yielded the longest products, $[A_4G]_4/[T_4C]_4$, was selected as the thio-modification appears at regular

intervals on alternating sides of the DNA duplex, Fig. 4 (a). A negative control extension, performed in the absence of a dGTP derivative results in no extension of the initial oligo seed, Fig 3 (b). An extension performed in the presence of all four unmodified dNTPs in the same conditions as used for the $[A_4{}^{6S}G]_n/[T_4C]_n$, produces DNA lengths comparable to Chapter 2.

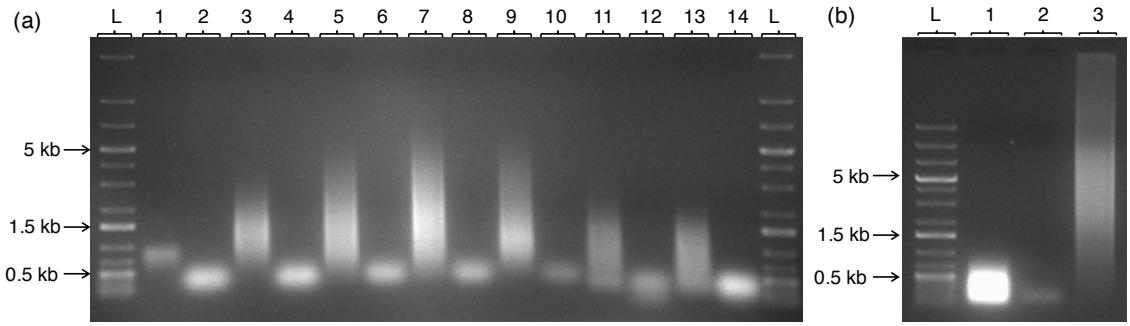


Fig. 3 Agarose gels of **(a)** DNA extension products versus oligo seeds and **(b)** DNA extension controls. L = DNA ladder, **(a)** lane 1: $[^{6S}G]_n/[C]_n$, lane 2: $[G]_{20}/[C]_{20}$, lane 3: $[A^{6S}G]_n/[TC]_n$, lane 4: $[AG]_{10}/[TC]_{10}$, lane 5: $[A_2{}^{6S}G]_n/[T_2C]_n$, lane 6: $[A_2G]_7/[T_2C]_7$, lane 7: $[A_3{}^{6S}G]_n/[T_3C]_n$, lane 8: $[A_3G]_4/[T_3C]_5$, lane 9: $[A_4{}^{6S}G]_n/[T_4C]_n$, lane 10: $[A_4G]_4/[T_4C]_4$, lane 11: $[A_9{}^{6S}G]_n/[T_9C]_n$, lane 12: $[A_9G]_2/[T_9C]_2$, lane 13: $[^{6S}GATC]_n/[CTA^{6S}G]_n$, lane 14: $[GATC]_5/[CTAG]_5$. **(b)** lane 1: $[A_4{}^{6S}G]_n/[T_4C]_n$, lane 2: $[A_4G]_4/[T_4C]_4$, lane 3: $[A_4G]_n/[T_4C]_n$.

UV-Vis of the thiolated nucleotides is a useful tool to identify the thio-functionality.[41] An absorbance at 340 nm for the 6-S-dG in the purified thiolated-DNA extension products is evidence of the successful incorporation, Fig. 4 (b). Additional characterisations are described in Chapter 3.

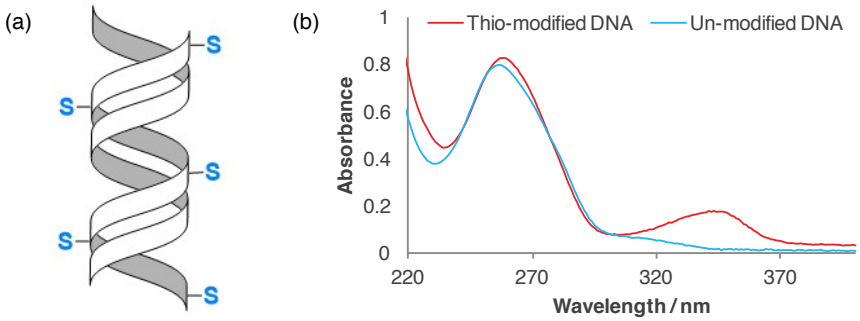


Fig. 4 (a) Schematic of $[A_4{}^{6S}G]_n/[T_4C]_n$, showing how thio-modifications at every 5th base results in potential binding sites on alternate sides of the DNA helix **(b)** UV-Vis spectrum of unmodified and thio-modified DNA extension products.

The above extensions were performed with an unmodified oligo seed, resulting in products containing a single unmodified oligo seed sequence at the ends of the DNA product. As control over the spacing of the modification throughout the dsDNA is desirable, a modified oligo seed is required to ensure completely thiolated products. Thiolation of the $[A_4G]_4$ oligo seed was performed as described in the materials and methods.[42] Heat-cool extension reactions were then carried out and the product

lengths were assessed by gel electrophoresis. Unfortunately, no extension was observed using the thiolated oligo seed. As discussed in Chapter 3, the incorporation of 6-S-dGTP into DNA, reduces the stability of the duplex, therefore, the initial oligo seed may not anneal at the 55 °C employed in the heat-cool cycle. The melting temperature (T_m) of $[A_4G]_4/[T_4G]_4$ is 67 °C (Chapter 2), however, a 6 °C reduction has been documented on the inclusion of one 6-S-dG into a 13 bp duplex.[43] Here, $[A_4{}^{6S}G]_4/[T_4C]_4$ contains four 6-S-dG units, therefore a total T_m reduction of up to 24 °C is conceivable, resulting in an estimated T_m of 43 °C. Therefore, the $[A_4{}^{6S}G]_4/[T_4C]_4$will most likely remain in its single strand form at 55 °C and thus not provide a binding site for the DNA polymerase to begin extension.

4.2.1.2 *Metal ion coordination with 6-S-dG-DNA*

Metal : 6-S-dG coordination polymers have been studied due to their conductive properties.[32] The extended structure of the coordination polymer depends on the presence of the ribose sugar or phosphate group on the nucleobase. In these previous studies, the coordination polymer was formed from modified nucleobases, nucleosides and nucleotides. Here the focus is to exploit the 6-S-dG binding sites already embedded into long dsDNA. Thus, the elucidation of stable metal ion coordination modes with dsDNA could present an avenue to realise DNA scaffolding through site specific metal ion deposition. The interactions between $[A_4{}^{6S}G]_n/[T_4C]_n$ (where n > *ca.*100) and several metal ions were investigated; Au^+, Ni^{2+}, Cd^{2+} and Au^{3+}. The concentration of 6-S-dG was determined using the extinction coefficient, 9485 $M^{-1}cm^{-1}$. To establish if metal binding is possible when the thio-group is distributed throughout dsDNA, UV-Vis binding titrations were performed. On coordination, the thio-absorbance shifts to a lower wavelength (from c.a. 348 nm to 320 nm), corresponding to an increased proportion of the thiol form, which is known to absorb between 320 and 330 nm.[44] By monitoring the change in absorbance, a binding plot can be produced to understand the approximate stoichiometry and dissociation energy. The binding curves for each metal ion show interactions with the thio-functional group, Fig. 5.

Analysis of the binding titration plots revealed the dissociation constant, K_d, to be 91 ± 16 μM for Au^+ binding, 102 ± 12 μM for Ni^{2+} binding, 23 ± 4.6 μM for Cd^{2+} binding and 24 ± 3.3 μM for Au^{3+} binding (inverse percentage plot and titration tables are found in Appendix C). Au^+ has low solubility in 100 % aqueous solution, and hence may be

unfavourable in the DNA environment resulting in the higher dissociation constant. On the other hand, Cd^{2+} has the lowest dissociation constant as expected due to the strong interaction between Cd ions and thio-groups.[45]

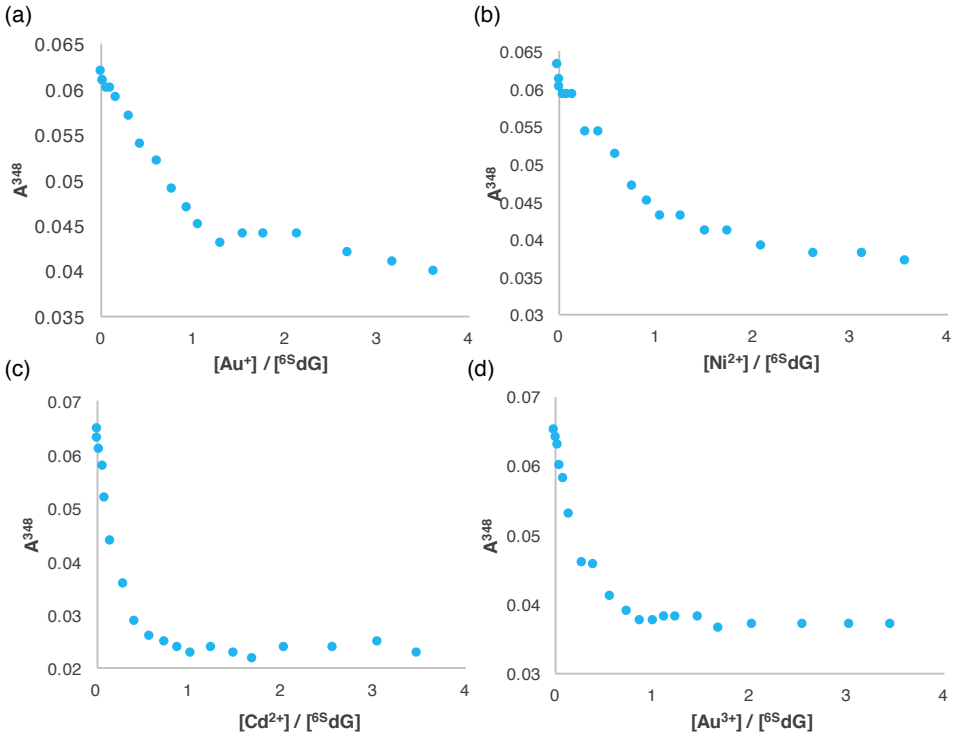


Fig. 5 Static $[A_4{}^{6S}G]_n/[T_4C]_n$ binding titration plots with **(a)** Au^+, **(b)** Ni^{2+}, **(c)** Cd^{2+} and **(d)** Au^{3+}.

Binding stoichiometry can be determined by performing a two variable titration and plotting a Job plot. Au^+ appears to form two binding modes with 6-S-dG in dsDNA, Fig. 6 (a). The thio-group appears throughout the DNA on one strand and every 5th base, so the observed 2:1 binding ratio must result from interstrand interactions of two DNA duplexes. As more Au^+ is added during the titration, a 1:1 binding mode is also apparent. Interstrand interactions may be unfavourable due to the steric bulk of the DNA duplex, therefore, in the presence of a higher concentration of Au^+, 1:1 binding is adopted to remove the crosslinking requirement. Ni^{2+} binding to 6-S-dG modified DNA expresses a 3:1 and a 1:1 binding ratio, Fig. 6 (b). A 3:1 binding ratio was reported previously for Ni^{2+} with the 6-S-dG monomer (S. Hribesh PhD thesis, 2013), and similar coordination could exist within the long DNA here. However, due to the stoichiometric challenge to assemble three DNA duplexes together, again a 1:1 binding ratio could be envisaged to relieve some of the unfavourable thermodynamic architectures. Cd^{2+} binding to 6-S-dG also expresses 3:1 binding, as seen from the Job plot analysis, Fig. 6 (c). In this instance 2:1 binding is also observed. Cd^{2+} binding is stronger than the other metal ions analysed here, hence there is a potential to overcome the steric

hindrance possible from the surrounding DNA duplex. Binding studies with Au^{3+} exhibit similar characteristics to binding with Ni^{2+} as a 3:1 and 1:1 binding ratio is evident from the Job plot, Fig. 6 (d). The binding stoichiometry for each metal ion determined by the titration plots are similar but not identical to the Job plots due to the variation in the methods used (see section 4.4 Experimental Details).

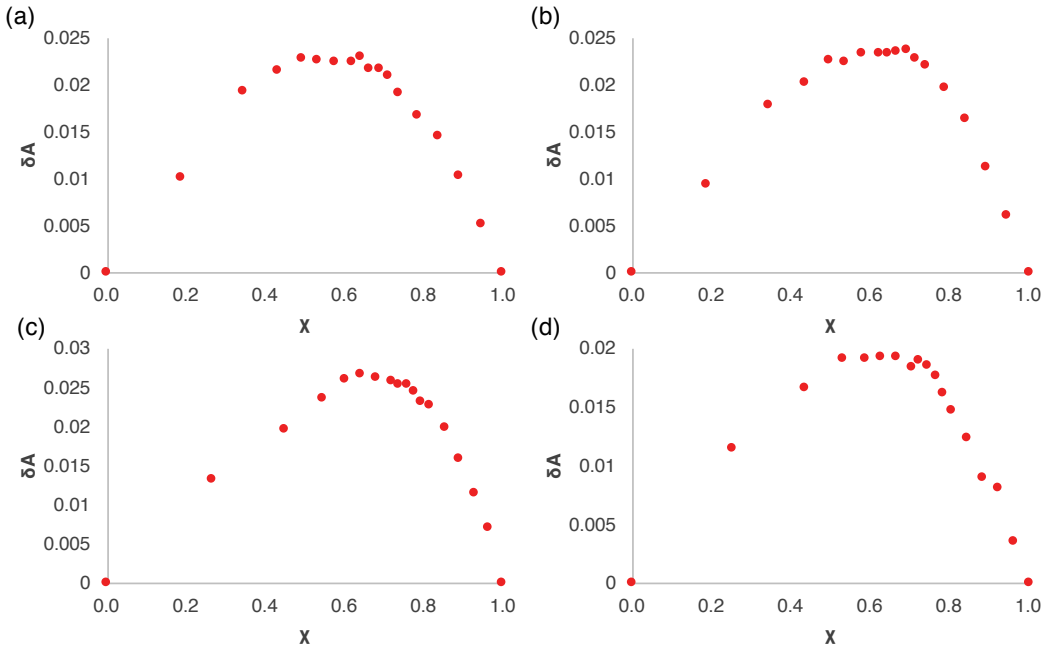


Fig. 6 Job Plots of the two variables titration of 6-S-dG-DNA with **(a)** Au^{+}, **(b)** Ni^{2+}, **(c)** Cd^{2+} and **(d)** Au^{3+}. X = mole fraction = 6-S-dG / (6-S-dG + metal)

To further elucidate the different binding modes observed for each metal ion with 6-S-dG-DNA, Raman spectroscopy was performed. Standard DNA, shown as the yellow plot in Fig. 7, has a well-known fingerprint; each peak corresponding to certain bases and bonds or rings within those bases.[9] On incorporation of the thio-modification, there is a shift in two peaks, one corresponding to the purine ring stretch, shifted from 1544 cm^{-1} to 1568 cm^{-1} and the other the C=O/C=S stretch, shifted from 1643 cm^{-1} to 1671 cm^{-1}. The shift of these peaks suggests the thio-function has been incorporated and has led to the higher wavenumber peak due to the lower electronegativity of S compared to O. Additionally, on coordination with metal ions, the two shifted peaks maintain the same wavenumber, except for Au^{3+}, however, peak intensities vary with each metal ion – Cd^{2+} peaks increase, due to the strong binding affinity to both the thio-group and N-7.[47] The other three metal ions have similar, if not, reduced intensities suggesting a different binding mode as elucidated from the Job plots. The N-H functional group contributes to the broad peaks between 3200 and 3500 cm^{-1}. A reduction in intensity is observed for Cd^{2+} : 6-S-dG-DNA hybrids in

comparison to the other M^{y+} : 6-S-dG-DNA, which again suggests a potential variation in binding mode.

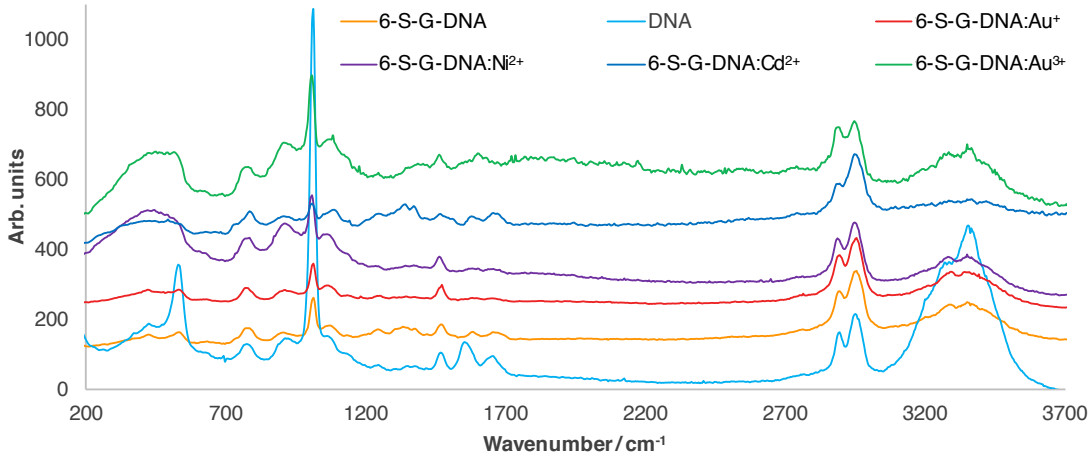
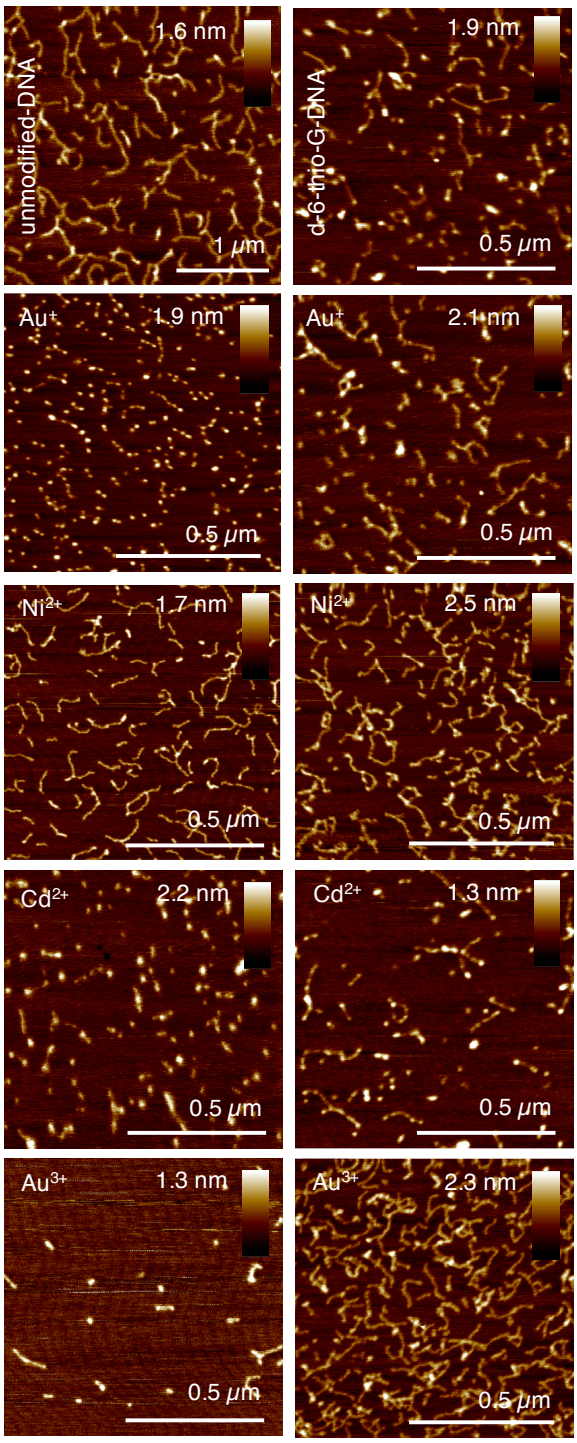


Fig. 7 Raman spectrum of $[A_4G]_n/[T_4C]_n$ DNA extension products, with and without the thio-modification, and after the addition of metal ions to 6-S-DNA, with each intensity at 2950 cm^{-1} normalised to 150 units.

As the binding ratios are now known, it would be interesting to investigate the morphology of the DNA-metal coordination complex. AFM elucidated the hybrid height profiles to reveal any aggregation occurring, Fig. 8. On addition of metal ions (relative concentration to 6-S-dG-DNA was chosen by the highest binding ratio observed), a height increase is noted in each instance. However, by comparing the heights of 6-S-bearing DNA compared with unmodified DNA, an increase is observed for both types of DNA with Au^{+}, Cd^{2+} and Au^{3+}. Unspecific binding in these instances is likely to occur with the negatively charged phosphate backbone. Nickel DNA nanowires have been reported previously,[48] therefore, binding to the phosphate backbone is also expected with Ni^{2+}. However, in this instance no aggregation is detected by AFM with unmodified DNA. DNA solutions were prepared in the presence of 1 mM $MgCl_2$ – Mg^{2+} has a higher affinity than Ni^{2+} for phosphate binding, hence may displace Ni^{2+}.[8] Therefore, changes in morphology on Ni^{2+} addition are specific to the thio-modification. Bare DNA heights for 6-S-DNA and DNA are 0.6 ± 0.1 nm. On addition of Ni^{2+}, a height increase to 1.2 ± 0.2 nm is noted for 6-S-DNA, suggesting the presence of multiple DNA duplexes per linear structure. In order to achieve the 3:1 binding ratio observed for 6-S-dG : Ni^{2+}, DNA aggregation is necessary to reach this stoichiometry. The double height structures apparent by AFM analysis, provide evidence to suggest the self-assembly of linearly aggregated DNA structures through Ni coordination between 2 or more duplexes.



	DNA height / nm		6-S-dG-DNA height / nm	
	Average	StDev	Average	StDev
Bare	0.6	0.1	0.6	0.1
Au^{+}	0.9	0.3	1.0	0.3
Ni^{2+}	0.6	0.1	1.2	0.2
Cd^{2+}	1.0	0.3	0.8	0.2
Au^{3+}	0.8	0.2	0.9	0.2

Fig. 8 AFM images of unmodified and 6-S-dG-DNA with and without the addition of Au^{+}, Ni^{2+}, Cd^{2+} and Au^{3+}. The average heights and standard deviations were calculated from the AFM images in Appendix C.

By selecting the metal ion, the binding ratio, and consequently the self-assembly of higher ordered DNA structures bearing 6-S-dG can be directed. Further investigations into how the DNA sequence could be used to control the structure assembled by the metal ions would open further avenues for the exploitation of the output DNA in molecular electronics.

4.2.2 alpha-phosphorothioate-DNA

Although 6-thio-dG derivatives are useful for metal ion coordination,[32] the reduced stability on incorporation into DNA apparent by the lack of extension during heat-cool cycles, means only limited DNA lengths are available. To overcome the length challenge, other thio-containing derivatives were investigated. Previously, α-S-dNs have mainly been studied as a mechanism to stall nuclease degradation.[49] Metal ion coordination with this thio-ligand has also been considered to assess the affinity of *in vivo* metal ions.[34] Using the heat-cool cycle enzymatic DNA extension method discussed in Chapter 3, incorporation of α-S-dNTPs were investigated to produce DNA bearing thio-groups on the backbone of the duplex from a $[GATC]_5/[CTAG]_5$ oligo seed. Each dNTP derivative was incorporated individually, and an extension with all four dNTP derivatives in one reaction was performed, Fig. 9. Each α-S-dNTP can be incorporated into the $[GATC]_n/[GATC]_n$ sequence to produce DNA lengths greater than 1000 bp after 30 heat-cool cycles using Tgo-Pol Z3 exo-. The extension proficiency varied with each dNTP derivative - α-S-dTTP extension is the most efficient followed by α-S-dCTP, α-S-dGTP and α-S-dATP being the least efficient. Extension with all four α-S-dNTPs is also possible, to produce DNA bearing a phosphorothioate between every base on both sides of the duplex. DNA, bearing thio-groups readily accessible for metal binding, provides a route to the fabrication of 1-D nanomaterials with site specific deposition along the helix.

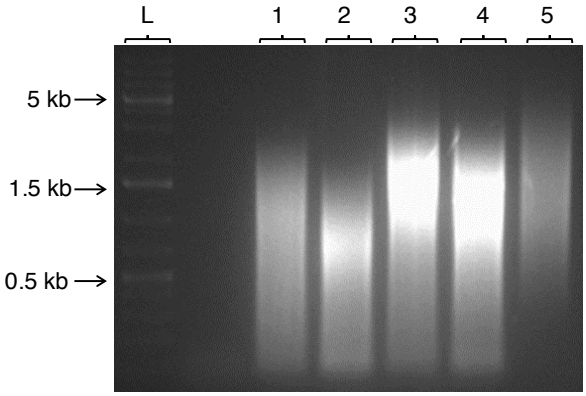


Fig. 9 Agarose gel of DNA extension products from a $[GATC]_5$ oligo seed with α-S-dNTPs, lane 1: α-S-dGTP, lane 2: α-S-dATP, lane 3: α-S-dTTP, lane 4: α-S-dCTP, lane 5: all α-S-dNTPs, L = DNA ladder.

As reported previously, modifications may reduce the DNA duplex stability. To investigate the effect of α-S-incorporation, T_m analysis was performed on short 20 bp oligos with and without the α-S-modification, Fig. 10. A slight decrease in stability is noted on addition of α-S to the duplex – 55 °C for $[A_4{}^{\alpha S}G]_4/[T_4C]_4$ and 56 °C for

$[A_4G]_4/[T_4C]_4$, however, as only a slight decrease is notable, DNA stability is deemed acceptable for continuing study for nanomaterials.

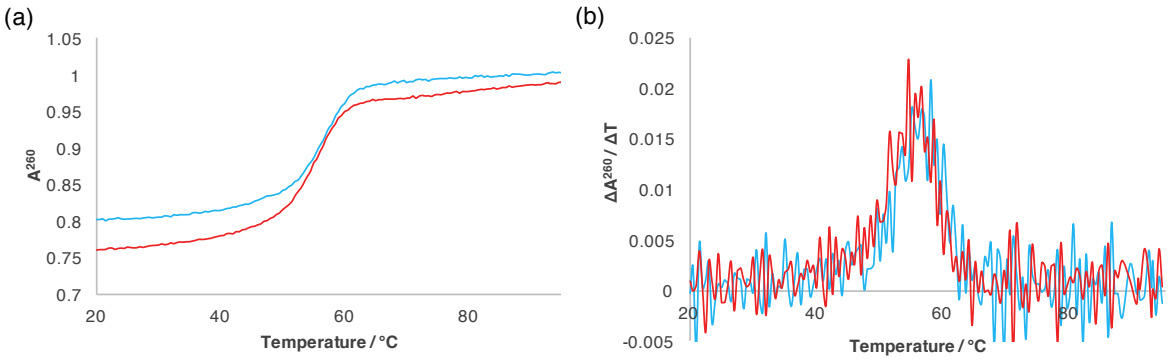


Fig. 10 (a) UV-Vis melting curves and **(b)** the first derivatives of $[A_4G]_4/[T_4C]_4$ (blue) and $[A_4{}^{\alpha\text{-S}}G]_4/[T_4C]_4$ (red).

As described in Chapter 3, DNA digestions can be used to analyse the constituent base components of a DNA duplex. However, as the enzymes used to digest the DNA produce the nucleoside constituents - Snake venom phosphodiesterase breaks the 3' phosphodiester bond and alkaline phosphatase cleaves the 5' phosphodiester bond – the α-S-modification is removed during the digestion process.[50] Hence, DNA digestions cannot be used to determine whether the DNA products contain the α-S-group. However, control extensions performed in Chapters 2 and 3 indicate that the effective promiscuity of Tgo-Pol Z3 exo- should allow for the high fidelity incorporation of the α-S-dCTP. Additionally, DNA sequencing (GATC Biotech) showed that the oligo seed sequence was maintained, Fig. 11, confirming correct base pairing occurred.

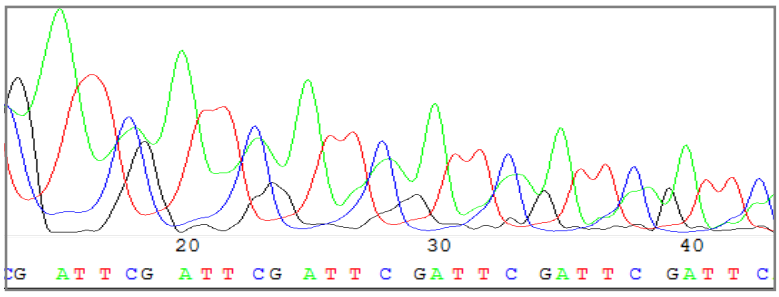
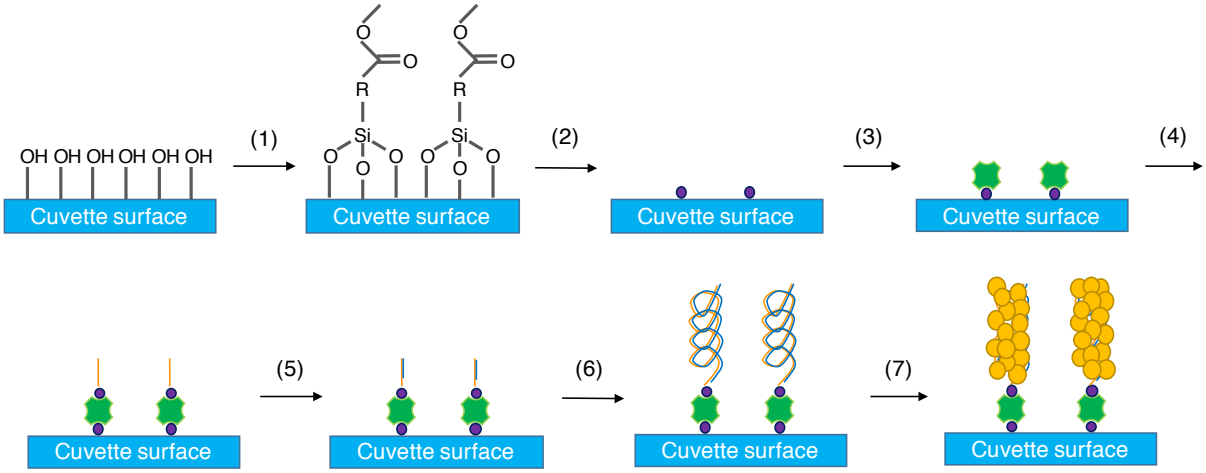


Fig. 11 DNA sequencing of $[^{\alpha\text{-S}}GATTC]_n/[CAAT^{\alpha\text{-S}}G]_n$ extension product.

To utilise the α-S-DNA for nanomaterial purposes, an investigation into AuNP deposition was conducted. To determine the binding specificity of the phosphorothioate backbone linkage for AuNPs, UV-Vis was performed to quantify the binding event. Firstly, a DNA brush was immobilised onto the inner surface of a cuvette, Scheme 1. Surface immobilisation ensures that thorough washing could be performed

to determine that the bound AuNP was in fact attached to the DNA. A biotinylated oligonucleotide was first bound to a streptavidin modified cuvette, followed by polymerisation using the *E. coli* DNA polymerase I Klenow exo- fragment. As DNA extension using Tgo-Pol Z3 exo- is performed at 72 °C, the Klenow exo- fragment was used as it is active at 37 °C,[51] which reduces streptavidin and DNA denaturation.[52,53] An initial duplex of $(dG)_{20}/(dC)_{20}$ was extended to produce a single layer DNA brush (c.a. 4,500 bp) on the cuvette inner surface. α-S-dGTP and dCTP were incorporated into the sticky ends during the slippage reaction, resulting in long linear dsDNA bearing thio groups on one strand.



Scheme 1 DNA brush preparation on the inner surface of a glass cuvette. **(1)** 2-carbomethoxyethyltrichlorosilane in super dehydrated toluene, **(2)** 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide, HCl and amine-PEG_2-biotin in HEPES, pH 7.6 **(3)** streptavidin in Tris, pH 7.9. **(4)** C_{15}-biotin in Tris-HCl, pH 7.9 and NaCl, **(5)** G_{15} in Tris-HCl, pH 7.9 and NaCl. **(6)** polymerisation solution –dNTPs, Klenow exo- in $MgCl_2$, Tris-HCl, pH 7.9 and NaCl, **(7)** 5 nm AuNP citrate stabilised. Cuvettes were washed thoroughly between each step.

The consequential incubation of the cuvette surface with 5 nm AuNPs stabilised with citrate, was monitored by UV-Vis, Fig. 12. Coating the nanoparticles with citrate groups, provides a negative charge, and hence a repulsion to the DNA backbone. Binding of the NPs to the α-S-DNA is evident by the absorbance peak at 525 nm and only a small background peak is present in the control, unmodified DNA, Fig. 12. Therefore, the specific binding of the AuNPs to α-S-DNA can be concluded.

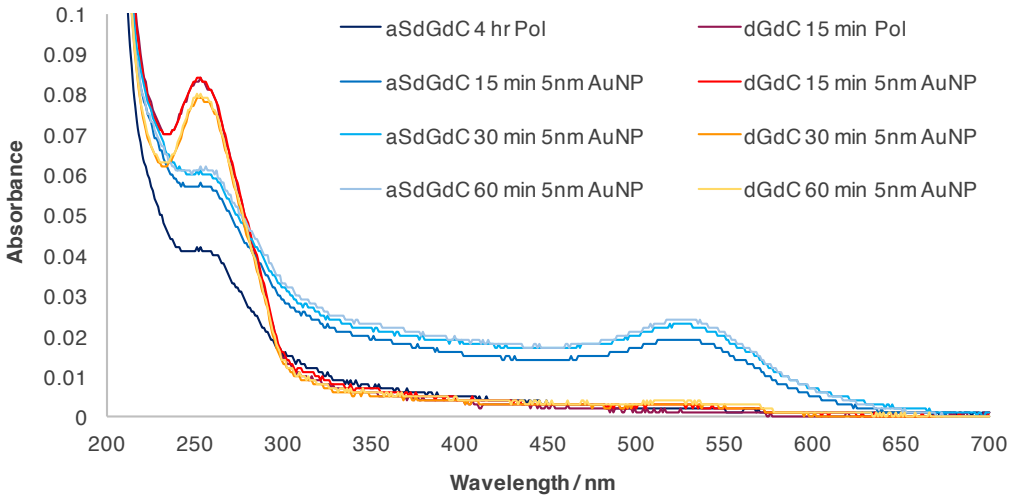


Fig. 12 UV-Vis of DNA brush modified cuvette cells, followed by incubation with citrate stabilised 5 nm AuNP over specified times.

An initial AFM experiment to visualise the effect of AuNP binding to α-S-DNA ($[GATC]_n/[GATC]_n$) where each phosphate linkage on both strands bears a α-S-group was performed. AFM images of the bare α-S-DNA and unmodified DNA of the same sequence indicated no significant differences, Fig. 13 (ai) and (bi). On addition of citrate stabilised 3 nm AuNPs, binding to α-S-DNA is apparent, Fig. 13 (aii), visible by the positioning of the NPs on the linear DNA structure. No specific binding is observed to the unmodified DNA, Fig. 13 (bii). Therefore, specific binding of AuNPs to α-S-DNA brushes as seen by UV-Vis, is also observed on linear DNA deposited onto a mica surface by AFM. In both these studies, the DNA bears α-S functional groups throughout the DNA structure, i.e. NPs can bind to any position. 3 nm AuNPs at maximum loading (intimate contact between nearest neighbours) could in theory position themselves approximately once every helical turn (c.a. 3.4 nm), to produce a uniformly templated DNA hybrid material. However, this is not observed as the initial deposition and incubation of AuNPs was performed within 15 minutes which may have limitations on the assembly of a fully coated uniform structure. To direct the assembly of a uniformly distributed DNA-NP hybrid material with regular spatial control, the repeating sequence can be designed to position the α-S-binding sites at the desired spacing for NP deposition.

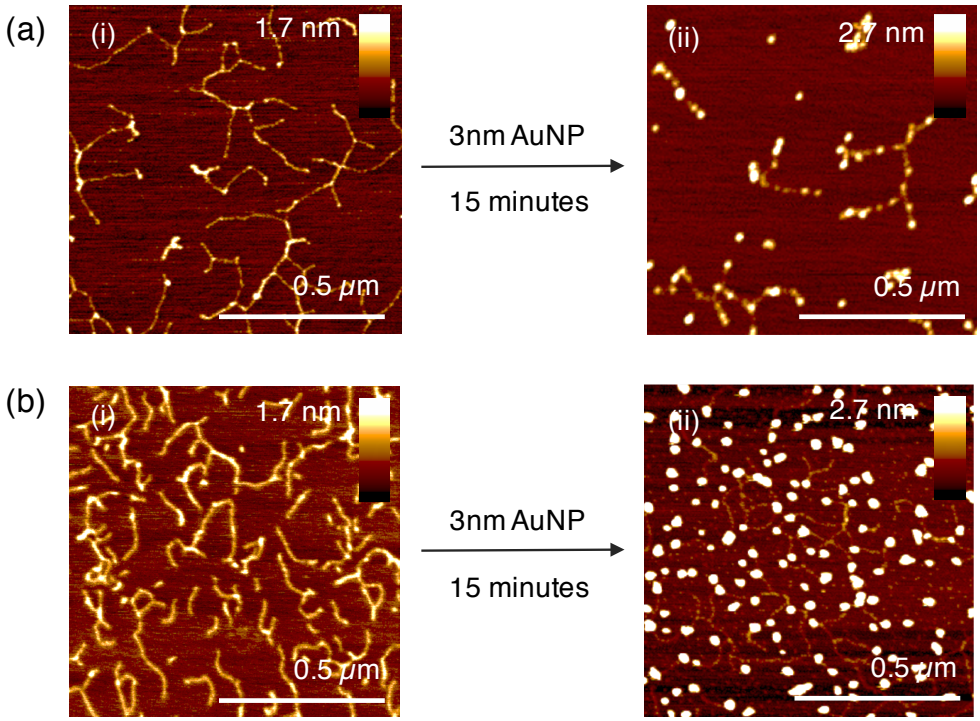


Fig. 13 AFM of **(a)** α-S-DNA and **(b)** DNA on mica, before (i) and after incubation with 3 nm AuNP for 15 mins (ii).

To direct the even deposition of AuNPs to once every helical turn, the repeating sequence $[(AT)_4G_2]_n/[(TA)_4C_2]_n$ was extended with α-S-dCTP to produce DNA consisting of two adjacent α-S-groups per turn to enhance NP immobilisation. After 30 heat-cool cycles, 1,000 bp lengths of DNA were observed in the gels for both the control, using all four unmodified dNTPs and the α-S-DNA, replacing dCTP with α-S-dCTP, Fig. 14 (a). α-S-DNA was analysed by AFM and exhibited DNA lengths as expected from the agarose gel and heights in correspondence with literature values,[54] Fig. 14 (bi). The α-S-DNA immobilised on mica was incubated with 3 nm AuNPs for 10 minutes before washing with 1 mL water, Fig. 14 step (1). Deposition of the 3 nm AuNPs onto the DNA is apparent, however, on scanning the length of a sample of DNA molecules, it is clear the deposition is not uniformly distributed, Fig. 14 (bii). A fresh AuNP incubation was repeated for another 10 minutes with gentle shaking, Fig. 14 step (2). After washing, AFM analysis revealed DNA molecules with additional AuNP deposition suggesting sequential incubation steps allows additional binding. This process was repeated a further two times, Fig 14 steps (3) and (4), to yield DNA with near uniform deposition of AuNPs as exhibited by the height traces shown below the AFM images, Fig. 14 (bv). Therefore, the incorporation of α-S linkages into DNA provides an additional approach to the DNA scaffolding of metal NPs. Using the

enzymatic extension of oligo seeds additional control over the deposition sites through designer sequences could have implications for complex nanomaterial synthesis.

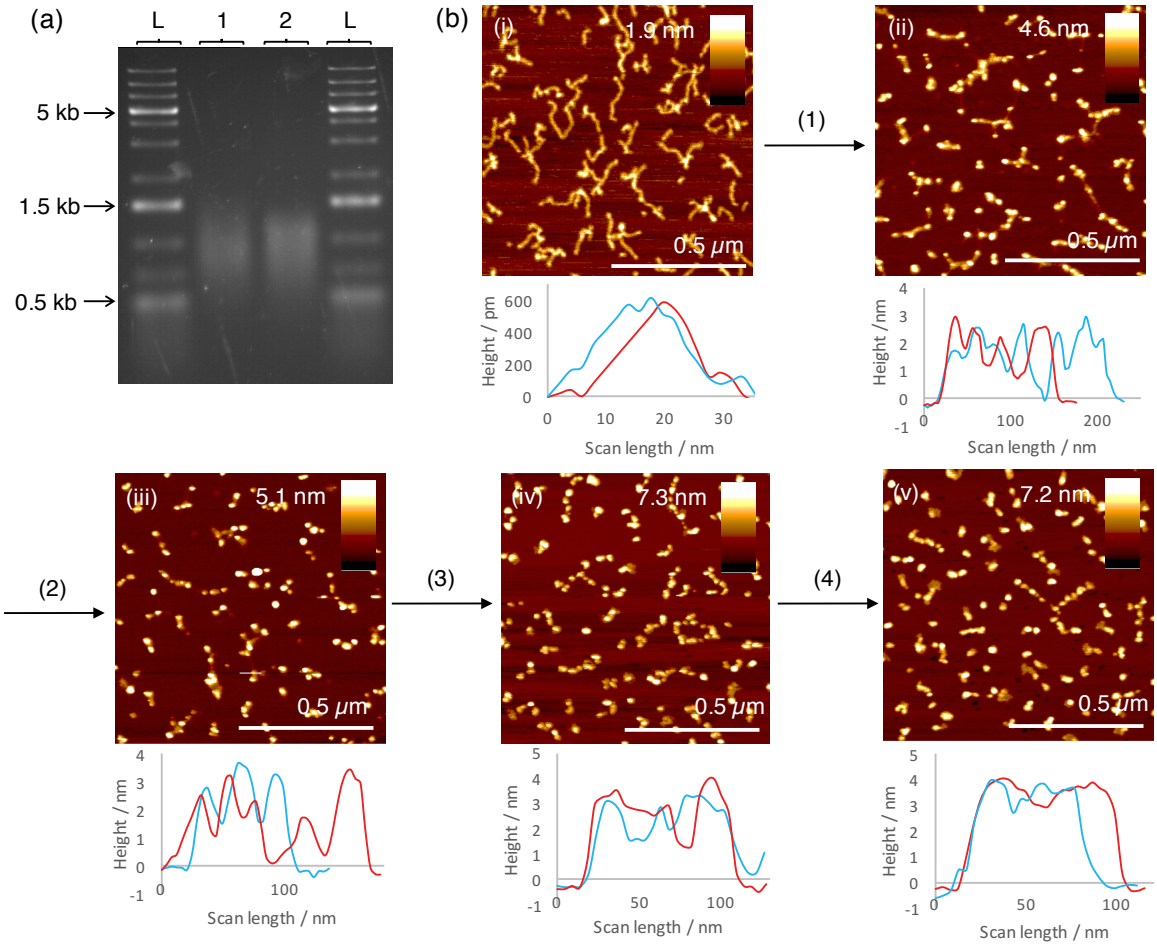


Fig. 14 (a) Agarose gel of $[(AT)_4G_2]_n/[(TA)_4{}^{\alpha S}C_2]_n$ (lane 1) and $[(AT)_4G_2]_n/[(TA)_4C_2]_n$ (lane 2) DNA extension products, L = DNA ladder. **(b)** AFM of $[(AT)_4G_2]_n/[(TA)_4{}^{\alpha S}C_2]_n$ on mica, followed by the consecutive deposition and washing of citrate stabilised 3 nm AuNP, (1) 10 mins incubation, (2) and (3) 10 mins incubation with gentle shaking, (4) 20 mins incubation with gentle shaking.

Therefore, this method for the controlled spacing of NPs may provide a route to optimise surface enhanced plasmon resonance (SERS), in particular the ability to create small gaps (< 30 nm) between adjacent NPs to generate SERS "hot-spots".[55,56] The sequence positioning of the α-S-group once every 4 turns (c.a. 13.6 nm separation) within the DNA helix provides approximately a 10 nm gap between each 3 nm AuNP. Therefore, the 40-base repeat oligo seed $[(AT)_{19}{}^{\alpha S}G_2]_2/[(TA)_{19}C_2]_2$ was extended in the presence of α-S-dGTP and produced DNA modal lengths of 700 bp after 30 heat-cool cycles, Fig. 15 (a). The same extension lengths were achieved using standard dNTPs and an unmodified oligo seed, Fig. 15 (a). Initial analysis by AFM depicted short strands of DNA, which were difficult to image bare and became "invisible" after NP deposition. To be able to analyse by AFM the deposition of 3 nm

AuNPs onto this 40 base repeat DNA, longer lengths of DNA are required to prevent aggregation and allow for more facile molecular combing. Therefore, to establish whether further extension beyond 700 bp is possible, 40 and 50 cycle extensions were performed. Analysis of the DNA extension products by agarose gel electrophoresis suggests however that additional extension is not possible using the conditions employed here, Fig. 15 (b). On further analysis of the agarose gel by Image J, Fig. 15 (c), it is clear that no further extension occurred beyond 30 heat-cool cycles. Therefore, to analyse the deposition of AuNPs on DNA of this length, (c.a. 250 nm) a more sensitive visualisation technique is required - liquid AFM or AFM with an ultra-sharp tip could provide a possible route to making this observation.[57]

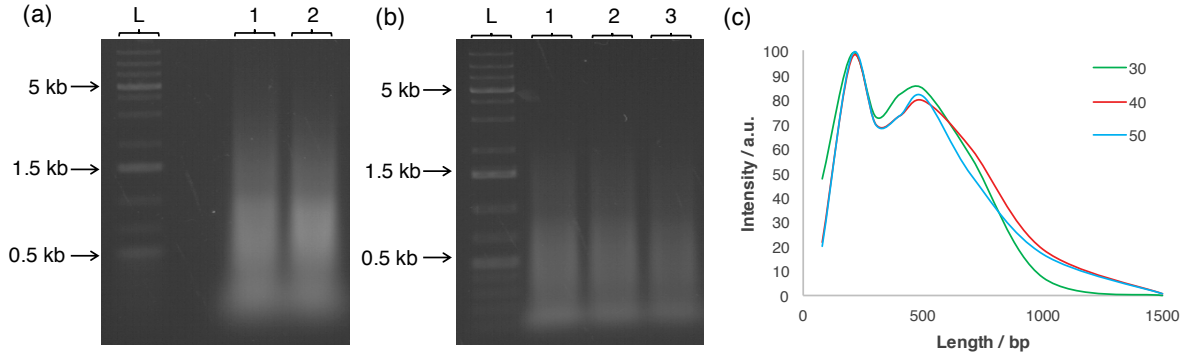


Fig. 15 Analysis of DNA heat-cool enzymatic extension products **(a)** agarose gel of DNA extension products after 30 cycles from lane 1: $[(AT)_{19}G_2]_2/[(TA)_{19}C_2]_2$ with dGTP and lane 2: $[(AT)_{19}{}^{\alpha S}G_2]_2/[(TA)_{19}C_2]_2$ with α-S-dGTP, **(b)** agarose gel of $[(AT)_{19}{}^{\alpha S}G_2]_2/[(TA)_{19}C_2]_2$ DNA extension products after lane 1: 30 cycles, lane 2: 40 cycles, and lane 3: 50 cycles, L = DNA ladder. **(c)** Image J analysis of (b), normalising the maximum intensity of each lane to 100.

4.3 Conclusions

A potential route towards the directed assembly of DNA-metal hybrid nanomaterials has been described here through the designed incorporation of thio-modifications into the DNA sequence. Several oligo seeds were extended with 6-S-dGTP in exchange of dGTP in the reaction pot to yield DNA with a thio-group every base, every other base and so on up to every 10^{th} base. DNA extension product length varied depending on the spatial frequency of the 6-S-dG base – extension efficiency increased on decreasing 6-S-dG frequency up to a maximum length for $[A_3G]_n/[T_3C]_n$. Metal ion titrations were performed with the 6-S-dG-DNA to determine binding stoichiometry. Titrations were performed with Au^+, Ni^{2+}, Cd^{2+} and Au^{3+} to yield primary binding ratios of 2:1 for Au^+ and 3:1 for Ni^{2+}, Cd^{2+} and Au^{3+}. For these binding ratios to occur, higher ordered structures consisting of multiple DNA duplexes would be required. On closer analysis of the Job plots, a 1:1 binding ratio is observed which is explained by the unfavourable steric hindrance to bring DNA duplexes together. On AFM analysis of unmodified and 6-S-modified DNA, Ni^{2+} binding to 6-S-DNA appears specific and leads to DNA- nanostructures double in height compared to the unmodified DNA control, suggesting the ability to direct the assembly of a higher ordered DNA structure through metal ion coordination. As 6-S-DNA extension was limited by the reduced compatibility of 6-S-dGTP during enzymatic incorporation, DNA extension only reached 500 bp, therefore, an alternative thio-containing dNTP was investigated - α-S-dNTP. Extension of $[GATC]_5/[CTAG]_5$ with each α-S-dNTP was performed and yielded DNA of similar length to the unmodified derivative, approximately 1,500 bp. UV-Vis and AFM analysis confirmed the specific binding of citrate stabilised AuNPs to α-S-containing DNA. To direct an even distribution of 3 nm AuNPs along the DNA structure. $[(AT)_4G_2]_2/[(TA)_4C_2]_2$ was extended and deposited onto a mica surface. Four consecutive incubations with 3 nm AuNPs followed by washing with H_2O, produced a uniformly decorated hybrid DNA molecule. To establish whether the spacing of the NPs can be determined by the DNA sequence, $[(AT)_{19}{}^{\alpha S}G_2]_2/[(TA)_{19}C_2]_2$ was extended to yield designer DNA of lengths up to 700 bp. On addition on 3 nm AuNPs, it was difficult to ascertain whether the NPs were positioned on the DNA molecules. To conclude whether the NPs were spaced along the DNA backbone, a more sensitive technique is required. However, the data provided illustrates a method to specifically scaffold either metal ions or NPs on the designer DNA structure at precise positions, determined by the initial oligo seed.

4.4 Experimental details

Preparation and Purification of DNA polymerases

Tgo-Pol Z3 exo- was prepared and purified as described previously.[58,59]

Deoxynucleotide preparation

Oligos were purchased from Eurofins (Ebersberg, Germany) and complementary strands were annealed as shown in Table 1.

Oligos to be combined	Oligo seed	Duplex length / bp
$(A)_{20}$ and $(T)_{20}$	$[A]_{20}/[T]_{20}$	20
$(AG)_{10}$ and $(TC)_{10}$	$[AG]_{10}/[TC]_{10}$	20
$(AAG)_7$ and $(TTC)_7$	$[AAG]_7/[TTC]_7$	21
$(AAAG)_5$ and $(TTTC)_5$	$[AAAG]_5/[TTTC]_5$	20
$(AAAAG)_4$ and $(TTTTC)_4$	$[AAAAG]_4/[TTTTC]_4$	20
$(AAAAAAAAAG)_2$ and $(TTTTTTTTTC)_2$	$[AAAAAAAAAG]_2/[TTTTTTTTTC]_2$	20
$(GC)_{10}$	$[GC]_{10}/[CG]_{10}$	20

Table 1 Summary of oligos required to form the oligo seeds used.

Heat-cool cycle DNA extension

0.5 µM DNA duplex (Table 1), 200 nM DNA polymerase, DNA polymerase reaction buffer (200 mM Tris-HCl (pH 8.8, 25 °C), 100 mM $(NH_4)_2SO_4$, 100 mM KCl, 1 % Triton X-100, 1 mg/mL Bovine Serum Albumin (BSA) and 20 mM MgSO), and 0.5 mM deoxynucleotide triphosphates (dNTPs) (dCTP, dATP, dTTP and dGTP or 6-S-dGTP) were mixed. Thermocycling was carried out using an Applied Biosciences Veriti 96 well Thermal Cycler by the following method:

> N (number of cycles stated) x 30 seconds at 95 °C, 30 seconds at 55 °C and 120 seconds at 72 °C.

The products were cooled to 4 °C after the reaction and then purified using a QIAquick PCR purification kit (25) (QIAGEN, Manchester, UK) following manufacturers protocol.

Agarose Gel Electrophoresis

The extension products were analysed by gel electrophoresis in TBE (Tris, boric acid and NaEDTA.$2H_2O$) buffer. 0.3 % Seakem agarose (Lonza, UK) was added to the 1 % TBE buffer and heated to dissolve. The 0.3 % Seakem agarose solution was supplemented with ethidium bromide (Sigma Aldrich) and poured once it had cooled to 50 °C. The 1 kb+ ladder were purchased from Thermo Scientific and provided with a loading dye (2.5 % Ficoll-400, 11 mM EDTA, 3.3 mM Tris-HCl (pH 8.0, 25 °C), 0.017 % SDS and 0.015 % BPB). DNA samples were supplemented with the gel

loading dye. The gels were run at 100 V, 100 mA, 10 W for approximately 1 hour and then visualised using an ultra-violet transilluminator.

Thiolation of on-column oligo

Oligo 5'-AAAAGAAAAGAAAAGAAAAG- 3'

2 µmol oligo ($[A_4G]_4$) column (MWG Operon) was crushed with a spatula followed by addition of 10 mL pyrimidine. 1 mL trifluoroacetic anhydride was added drop wise followed by 40 minutes of stirring. A slurry of 1 g sodium sulphide and 40 mL N,N-dimethylformamide was added to the stirring solution and left for 24 hours. Solution was filtered using a Buchner funnel and washed with 3 x H_2O and 3 x ethanol. Solid beads were placed in a sample vial and filled with methylamine overnight. Vial was returned to atmosphere followed by addition of 1 mL H_2O to the vial. H_2O – bead suspension was run through a DNA purification column (QIAGEN PCR Purification kit), separating the free oligos from the beads.

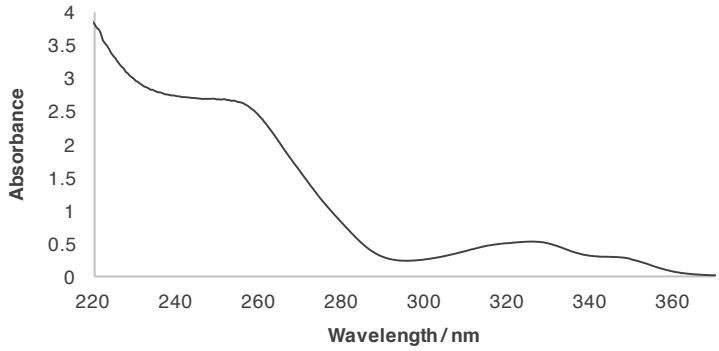


Fig. 16 UV-Vis of thioalted oligo after column purification.

HPLC purification of thio-oligo

A buffer system consisting of A: 0.1 mM triethyl ammonium acetate, pH 6.5 in 5 % acetonitrile and B: 0.1 mM triethyl ammonium acetate, pH 6.5 in 65 % acetonitrile was used to purify un-thiolated oligos on the gradient from 0 to 50 % buffer B over 30 minutes and on an APEX ODS 25 cm 4 µm column at 25 °C. Oligos eluted at the times stated in Table 2 and shown Fig. 17. The thio-oligo was collected for further use.

Analyte	Time / min
Un-modified oligo	3.2
Thio-modified oligo	6.2

Table 2 Elution times for the un-modified and thio-modified oligo.

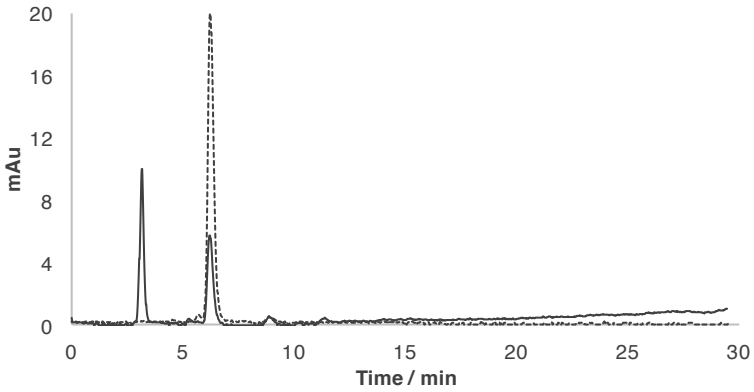


Fig. 17 HPLC purification of thiolatied oligo. Absorbance is detected at 254 (solid line) and 340 nm (dashed line) to determine the peaks that contain the thiol-function group.

Binding curves

Solutions of 6-thio-dG-DNA were prepared at the stated concentrations, found in Appendix C, in nanopure-H_2O. Metal ion solutions (AuCl, $Ni(NO_3)_2.6H_2O$, $Cd(NO_3)_2.4H_2O$) and $AuCl_3$) were prepared by dilution in nanopure-H_2O from a 100 mM stock solution. A 2 x concentration of 6-thio-dG-DNA was prepared for addition during titrations to maintain a constant 6-thio-dG concentration. Concentration of M^{y+} was altered by the removal of 10 µL coordination polymer mixture, followed by the addition of 5 µL M^{y+} and 5 µL 2 x concentration 6-thio-dG-DNA. Binding curves were performed by detecting the change in A^{348} as the M^+ concentration increases, see Tables 2-5 for more details. The absorbance was detected using a Nanodrop 1000. Binding constants were determined using a nonlinear fit of one site specific binding on Prism 7 (GraphPad, California, USA).

Job plots

Job plots were performed to determine binding stoichiometry by the gradual reduction of the 6-thio-dG mole fraction in a 100 µL mixture. Concentration was altered by the removal of binding solution, followed by the addition of M^{n+} of the same volume.

Raman spectroscopy

Raman spectroscopy was performed with a Melles Griot HeNe laser (California, USA) and Witec focus innovations system (Ulm, Germany). A 488 nm laser was used with a 600 g / nm grating and a spectral offset of 1800 cm^{-1}. Samples were dropped onto a clean glass cover slip using a micro pipette and allowed drying before analysis. Raman spectra was taken over a 5 x 5 µm square with a 5 second integration time per

spectrum and 15 points per line. The data shown in Fig. 7 is an average of each sample image.

Atomic Force Microscopy

The top layer of the mica surface was cleaved using sticky tape. 5 µL of the DNA sample at 4 ng / µL in 1 mM MgCl was spread onto the mica surface using a micro pipette. Nitrogen gas was passed over the surface to straighten the DNA and allowed drying for 1 hour. A light microscope was used to locate the sample and cantilever position. AFM images were collected using a Dimension V with a nanoscope controller (Veeco Instruments Inc., Metrology Group, Santa Barbara, CA). The tapping mode was used with an etched silicon tip (Tap 300 Al-G, 300 kHz, 40 N/m) on an isolation table (Veeco Inc., Metrology Group) to reduce interference. Nanoscope 7.00b19 software was used to acquire data.

High resolution mass spectroscopy

A Waters Micromass LCT Premier TOF system was used in positive mode, with a desolvation temperature of 250 °C and data collected using masslynx v. 4.1. Samples were injected directly in nanopure-H_2O.

DNA brush preparation

Cuvettes were cleaned using piranha solution followed by washing with H_2O and drying. A solution of 2-carbomethoxyethyltrichlorosilane was added in dry toluene in < 25 % humidity. After 1 hour, cells were washed, filled with HCl and left over night. The cells were washed thoroughly and filled with 50 mM 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide and 1 mM amine-PEG_2-biotin in 10 mM Hepes. After 1 hour, cells were washed and filled with 0.1 mg/mL streptavidin in 10 mM Tris, pH 7.9. After 1 hour, cells were washed and filled with 1 µM C_{15}-biotin in 10 mM Tris, pH 7.9 and 200 mM NaCl. After 1 hour, cells were washed and filled with 1 µM G_{15} 10 mM Tris-HCl, pH 7.9 and 200 mM NaCl. After 1 hour, cells were washed and filled with polymerisation solution – 0.5 mM dNTPs, 0.08 mg/mL Klenow exo- in 2 mM $MgCl_2$, 10 mM Tris-HCl, pH 7.9 and 200 mM NaCl. Reactions were performed as stated in the results and quenched by the removal of reaction solution and washing with 10 mM Tris, pH 7.9 and 200 mM NaCl.

4.5 References

1 Willner, I. & Willner, B. Biomolecule-Based Nanomaterials and Nanostructures. *Nano Lett* **10**, 3805-3815 (2010).

2 Huang, J. L. *et al.* Bio-inspired synthesis of metal nanomaterials and applications. *Chemical Society Reviews* **44**, 6330-6374 (2015).

3 Bhattacharya, P., Du, D. & Lin, Y. H. Bioinspired nanoscale materials for biomedical and energy applications. *J R Soc Interface* **11** (2014).

4 Weber, J., Singhal, R., Zekri, S. & Kumar, A. One-dimensional nanostructures: fabrication, characterisation and applications. *Int Mater Rev* **53**, 235-255 (2008).

5 Hasan, S. A Review on Nanoparticles: Their Synthesis and Types. *Research Journal of Recent Sciences* **4**, 12-15 (2015).

6 Chasteen, N. D. & Harrison, P. M. Mineralization in ferritin: An efficient means of iron storage. *J Struct Biol* **126**, 182-194 (1999).

7 Yoshimura, H. Protein-assisted nanoparticle synthesis. *Colloid Surface A* **282**, 464-470 (2006).

8 Eichhorn, G. L. & Shin, Y. A. Interaction of metal ions with polynucleotides and related compounds. XII. The relative effect of various metal ions on DNA helicity. *J Am Chem Soc* **90**, 7323-7328 (1968).

9 Duguid, J., Bloomfield, V. A., Benevides, J. & Thomas, G. J., Jr. Raman spectroscopy of DNA-metal complexes. I. Interactions and conformational effects of the divalent cations: Mg, Ca, Sr, Ba, Mn, Co, Ni, Cu, Pd, and Cd. *Biophys J* **65**, 1916-1928 (1993).

10 Moldrheim, E., Andersen, B., Froystein, N. A. & Sletten, E. Interaction of manganese(II), cobalt(II) and nickel(II) with DNA oligomers studied by (1)H NMR spectroscopy. *Inorg Chim Acta* **273**, 41-46 (1998).

11 Ito, Y. & Fukusaki, E. DNA as a 'nanomaterial'. *J Mol Catal B-Enzym* **28**, 155-166 (2004).

12 Pinheiro, A. V., Han, D. R., Shih, W. M. & Yan, H. Challenges and opportunities for structural DNA nanotechnology. *Nature Nanotechnology* **6**, 763-772 (2011).

13 Monson, C. F. & Woolley, A. T. DNA-templated construction of copper nanowires. *Nano Lett* **3**, 359-363 (2003).

14 Stoltenberg, R. M. & Woolley, A. T. DNA-templated nanowire fabrication. *Biomed Microdevices* **6**, 105-111 (2004).

15 Watson, S. M., Wright, N. G., Horrocks, B. R. & Houlton, A. Preparation, characterization and scanned conductance microscopy studies of DNA-templated one-dimensional copper nanostructures. *Langmuir* **26**, 2068-2075 (2010).

16 Mitomo, H., Watanabe, Y., Matsuo, Y., Niikura, K. & Ijiro, K. Enzymatic synthesis of a DNA triblock copolymer that is composed of natural and unnatural nucleotides. *Chem Asian J* **10**, 455-460 (2015).

17 Mao, V., Dwyer, C. & Chakrabarty, K. Fabrication Defects and Fault Models for DNA Self-Assembled Nanoelectronics. *Int Test Conf P*, 796-805 (2008).

18 Weizman, H. & Tor, Y. 2,2 '-bipyridine ligandoside: A novel building block for modifying DNA with intra-duplex metal complexes. *Journal of the American Chemical Society* **123**, 3375-3376 (2001).

19 Tanaka, K. & Shionoya, M. Synthesis of a novel nucleoside for alternative DNA base pairing through metal complexation. *J Org Chem* **64**, 5002-5003 (1999).

20 Takezawa, Y., Nishiyama, K., Mashima, T., Katahira, M. & Shionoya, M. Bifacial Base-Pairing Behaviors of 5-Hydroxyuracil DNA Bases through Hydrogen Bonding and Metal Coordination. *Chem-Eur J* **21**, 14713-14716 (2015).

21 Clever, G. H., Kaul, C. & Carell, T. DNA-metal base pairs. *Angew Chem Int Edit* **46**, 6226-6236 (2007).

22 Muller, J. Metal-ion-mediated base pairs in nucleic acids. *Eur J Inorg Chem*, 3749-3763 (2008).

23 Tasaka, M., Tanaka, K., Shiro, M. & Shionoya, M. A palladium-mediated DNA base pair of a beta-C-nucleoside possessing a 2-aminophenol as the nucleobase. *Supramol Chem* **13**, 671-675 (2001).

24 Tanaka, K., Tasaka, M., Cao, H. H. & Shionoya, M. An approach to metal-assisted DNA base pairing: novel beta-C-nucleosides with a 2-aminophenol or a catechol as the nucleobase. *Eur J Pharm Sci* **13**, 77-83 (2001).

25 Cao, H., Tanaka, K. & Shionoya, M. An alternative base-pairing of catechol-bearing nucleosides by borate formation. *Chem Pharm Bull* **48**, 1745-1748 (2000).

26 Tanaka, K. *et al.* Programmable self-assembly of metal ions inside artificial DNA duplexes. *Nat Nanotechnol* **1**, 190-194 (2006).
27 Polonius, F. A. & Muller, J. An artificial base pair, mediated by hydrogen bonding and metal-ion binding. *Angew Chem Int Edit* **46**, 5602-5604 (2007).
28 Megger, D. A. *et al.* Contiguous Metal-Mediated Base Pairs Comprising Two Ag-I Ions. *Chem-Eur J* **17**, 6533-6544 (2011).
29 Takezawa, Y., Nishiyama, K., Mashima, T., Katahira, M. & Shionoya, M. Bifacial Base-Pairing Behaviors of 5-Hydroxyuracil DNA Bases through Hydrogen Bonding and Metal Coordination. *Chemistry* **21**, 14713-14716 (2015).
30 Dupre, N., Welte, L., Gomez-Herrero, J., Zamora, F. & Muller, J. Bipyridine-modified oligonucleotides: Aggregation in the presence of metal ions. *Inorg Chim Acta* **362**, 985-992 (2009).
31 Pensa, E. *et al.* The chemistry of the sulfur-gold interface: in search of a unified model. *Acc Chem Res* **45**, 1183-1192 (2012).
32 Amo-Ochoa, P. *et al.* Coordination Chemistry of 6-Thioguanine Derivatives with Cobalt: Toward Formation of Electrical Conductive One-Dimensional Coordination Polymers (vol 52, pg 5290, 2013). *Inorg Chem* **52**, 7306-7306 (2013).
33 Amo-Ochoa, P. *et al.* Synthesis of Designed Conductive One-Dimensional Coordination Polymers of Ni(II) with 6-Mercaptopurine and 6-Thioguanine. *Inorg Chem* **48**, 7931-7936 (2009).
34 Da Costa, C. P., Okruszek, A. & Sigel, H. Complex formation of divalent metal ions with uridine 5 '-O-thiomonophosphate or methyl thiophosphate: Comparison of complex stabilities with those of the parent phosphate ligands. *Chembiochem* **4**, 593-602 (2003).
35 Da Costa, C. P., Krajewska, D., Okruszek, A., Stec, W. J. & Sigel, H. Stabilities of lead(II) complexes formed in aqueous solution with methyl thiophosphate (MeOPS2-), uridine 5 '-O-thiomonophosphate (UMPS2-) or adenosine 5 '-O-thiomonophosphate (AMPS(2-)). *J Biol Inorg Chem* **7**, 405-415 (2002).
36 Hung, A. M. *et al.* Large-area spatially ordered arrays of gold nanoparticles directed by lithographically confined DNA origami. *Nature Nanotechnology* **5**, 121-126 (2010).
37 Ding, B. Q. *et al.* Gold Nanoparticle Self-Similar Chain Structure Organized by DNA Origami. *Journal of the American Chemical Society* **132**, 3248-+ (2010).
38 Gates, E. P., Jensen, J. K., Harb, J. N. & Woolley, A. T. Optimizing gold nanoparticle seeding density on DNA origami. *Rsc Adv* **5**, 8134-8141 (2015).
39 Pearson, A. C. *et al.* DNA origami metallized site specifically to form electrically conductive nanowires. *J Phys Chem B* **116**, 10551-10560 (2012).
40 Lee, J. H. *et al.* Site-specific control of distances between gold nanoparticles using phosphorothioate anchors on DNA and a short bifunctional molecular fastener. *Angew Chem Int Edit* **46**, 9006-9010 (2007).
41 Zheng, Q. G., Wang, Y. & Lattmann, E. Synthesis of S-6-(2,4-dinitrophenyl)-6-thioguanosine phosphoramidite and its incorporation into oligoribonucleotides. *Bioorg Med Chem Lett* **13**, 3141-3144 (2003).
42 Kung, P. P. & Jones, R. A. One-Flask Syntheses of 6-Thioguanosine and 2'-Deoxy-6-Thioguanosine. *Tetrahedron Lett* **32**, 3919-3922 (1991).
43 Somerville, L. *et al.* Structure and dynamics of thioguanine-modified duplex DNA. *Journal of Biological Chemistry* **278**, 1005-1011 (2003).
44 Santhosh, C. & Mishra, P. C. Electronic-Structures and Spectra of 6-Mercaptopurine and 6-Thioguanine. *Spectrochim Acta A* **49**, 985-993 (1993).
45 Vairavamurthy, M. A., Goldenberg, W. S., Shi, O. Y. & Khalid, S. The interaction of hydrophilic thiols with cadmium: investigation with a simple model, 3-mercaptopropionic acid. *Mar Chem* **70**, 181-189 (2000).
46 Perrin, D. D. & Watt, A. E. Complex Formation of Zinc and Cadmium with Glutathione. *Biochim Biophys Acta* **230**, 96-& (1971).
47 Griffith, E. A. H. & Amma, E. L. Crystal-Structure and Cd-113(48) N M R Spectrum of Di-Mu-Chloro-Dichlorobis-(6-Mercaptopurine)Diaquodicadmium(Ii). *J Chem Soc Chem Comm*, 1013-1014 (1979).
48 Becerril, H. A., Ludtke, P., Willardson, B. M. & Woolley, A. T. DNA-templated nickel nanostructures and protein assemblies. *Langmuir* **22**, 10140-10144 (2006).
49 Connolly, B. A., Potter, B. V. L., Eckstein, F., Pingoud, A. & Grotjahn, L. Synthesis and Characterization of an Octanucleotide Containing the Ecori Recognition Sequence with a Phosphorothioate Group at the Cleavage Site. *Biochemistry* **23**, 3443-3453 (1984).

50 Wojcik, M., Cieslak, M., Stec, W. J., Goding, J. W. & Koziolkiewicz, M. Nucleotide pyrophosphatase/phosphodiesterase 1 is responsible for degradation of antisense phosphorothioate oligonucleotides. *Oligonucleotides* **17**, 134-145 (2007).
51 Datta, K., Wowor, A. J., Richard, A. J. & LiCata, V. J. Temperature dependence and thermodynamics of Klenow polymerase binding to primed-template DNA. *Biophys J* **90**, 1739-1751 (2006).
52 Gonzalez, M., Argarana, C. E. & Fidelio, G. D. Extremely high thermal stability of streptavidin and avidin upon biotin binding. *Biomol Eng* **16**, 67-72 (1999).
53 Holmberg, A. *et al.* The biotin-streptavidin interaction can be reversibly broken using water at elevated temperatures. *Electrophoresis* **26**, 501-510 (2005).
54 Watson, S. M. D., Pike, A. R., Pate, J., Houlton, A. & Horrocks, B. R. DNA-templated nanowires: morphology and electrical conductivity. *Nanoscale* **6**, 4027-4037 (2014).
55 Xu, H. X., Bjerneld, E. J., Kall, M. & Borjesson, L. Spectroscopy of single hemoglobin molecules by surface enhanced Raman scattering. *Physical Review Letters* **83**, 4357-4360 (1999).
56 Qin, L. D. *et al.* Designing, fabricating, and imaging Raman hot spots. *P Natl Acad Sci USA* **103**, 13300-13303 (2006).
57 Pang, D., Thierry, A. R. & Dritschilo, A. DNA studies using atomic force microscopy: capabilities for measurement of short DNA fragments. *Front Mol Biosci* **2**, 1 (2015).
58 Evans, S. J. *et al.* Improving dideoxynucleotide-triphosphate utilisation by the hyper-thermophilic DNA polymerase from the archaeon Pyrococcus furiosus. *Nucleic Acids Res* **28**, 1059-1066 (2000).
59 Jozwiakowski, S. K. & Connolly, B. A. A modified family-B archaeal DNA polymerase with reverse transcriptase activity. *Chembiochem* **12**, 35-37 (2011).

Chapter 5.

Exploitation of the Click Reaction for the Designed Assembly of Functional DNA

Table of Contents

5.1 Introduction

The controlled attachment of biological or functional molecules to DNA provides yet another route to DNA based materials or devices.[1,2] The functional modification of DNA is required to utilise DNA for detection, for example, as a molecular beacon.[2] Fluorescent DNA molecules are developing into highly powerful systems to target and monitor molecular properties and pathways.[2,3] Intercalation of planar, aromatic fluorescent molecules has been a popular method to track DNA for many years – expanded from the ethidium bromide staining used in DNA electrophoresis.[4] Intercalation of fluorescent molecules for DNA labelling is popular due to real time detection of, and specific binding to DNA. Many fluorescent intercalators have been developed with improved extinction coefficients and emission quantum yields,[5,6] aiding the progress of more sensitive detection methods.

Intercalation of planar aromatic compounds bearing external reactive sites, provides a route for chemical functionalisation of any length of DNA through in situ reactions such as the click reaction, Fig. 1.[7] However, for nanomaterial purposes, it is difficult to control the intercalation loading, and this approach is therefore not suitable for site specific modifications.

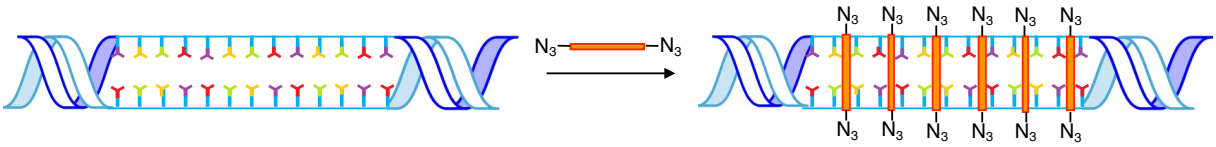


Fig. 1 The intercalation of proflavin diazide, a planar aromatic compound, into dsDNA. The intercalation in this instance provides DNA expressing azide-functional groups, available for a second layer of modification through the click reaction.[7,8]

Direct and specific incorporation of fluorescent and biomarker labels into oligos is made possible by the chemical alteration of phosphoramidites followed by inclusion through solid phase oligo synthesis.[9,10] Fluorescent modifications are commonly cited using this method to establish routes to study genetic variations, for example DNA mutations.[11] Although the solid phase oligo synthesis method permits the design of the DNA sequence, lengths are limited due to the reduction in yield after each dN addition. Additionally, chemical modifications must be compatible with the deprotection, protection and coupling steps.

To enhance the range of chemical modifications incorporated into DNA, the synthesis of functionalised dNTPs followed by their inclusion into DNA during PCR has been performed.[12] A common approach to modified dNTP synthesis is the Pd(0) or Pd(II)

catalysed Sonogashira coupling reaction: the coupling of a terminal alkyne with an aryl or vinyl halogen.[13-15] Performing the Sonogashira reaction on DNA components requires adaptations to the reaction conditions to make it compatible with an aqueous environment – commonly a water soluble Pd(0) catalyst was used.[16] Thus far, Sonogashira reactions have been performed on nucleosides (> 50 % yield[12]) to produce modified dNTPs for incorporation into DNA strands, however, aqueous-based Sonogashira reactions have not been performed on modified DNA strands.

The incorporation of reactive-site bearing dNTPs for post-labelling is an attractive approach for DNA chemical modification as it removes the limitations of DNA polymerase handling and the complex multi-step modified dNTP synthesis. One example of post-labelling was performed by an invasive acid depurination technique to reveal aldehyde groups for reaction with hydrazine containing fluorophores.[17] Although this is an efficient method for DNA labelling, the product lacks the ability to undergo Watson and Crick base pairing, which may result in unknown effects on the DNA structure and function.

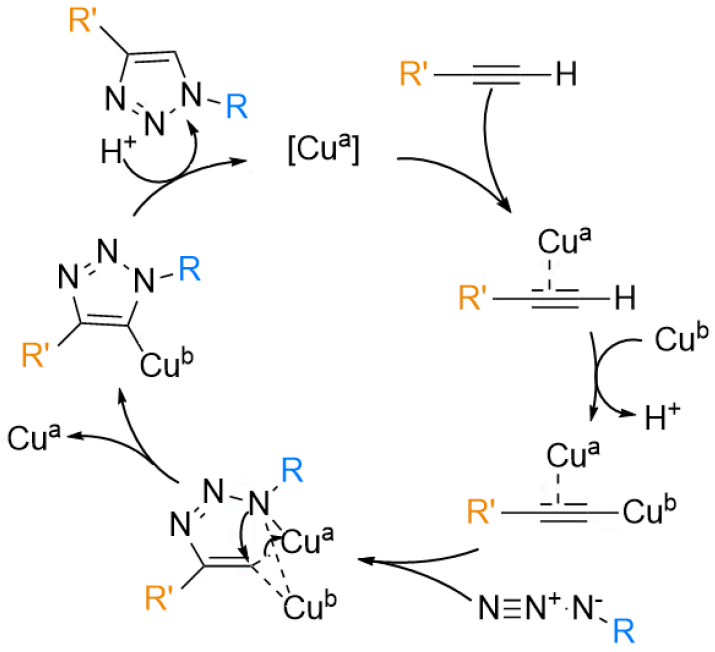


Fig. 2 Cu(I) catalysed click reaction based on the scheme by Fokin and coworkers.[18] A copper acetylene interaction in formed which increases the acidity of the acetyl-H. Deprotonation can then occur. An intermediate is formed with the azide-group, followed by a rearrangement to the five-membered 1,2,3-triazole ring. Protonation releases the copper complex to begin the next catalytic cycle.

Click chemistry is a term that was first introduced in 2001 by Sharpless and co-workers and is renowned for its bio-orthogonality, requirement of mild conditions and reliable yields.[19] The Cu(I) catalysed azide-alkyne cycloaddition is the most commonly used

reaction, Fig. 2.[20] This cycloaddition forms a high yielding triazole ring, linking the two original compounds. Click chemistry on DNA was initially performed on oligos bearing alkyne groups for DNA labelling, incorporated through the phosphoramidite method, Fig. 3 (a).[21,22] An increase in duplex stability is noted for the inclusion at C7 of deoxyguanine of alkyl-groups which lie in the major groove.[23] Sequential chemical additions of three different azide building blocks can yield triply modified oligonucleotides.[24] The click reactions were performed sequentially by protecting the second and third alkyne with a trimethylsilyl and triisopropylsilyl ether group, followed by step-wise deprotection after completion of the first and then second click reaction was performed.[24] By labelling DNA with three different reporter groups, it may be possible to report multiple genetic instabilities simultaneously. However, oligo lengths are again limited, hence enzymatic routes have also been explored to produce "clickable" DNA of more than 100 bp. PCR performed in the presence of alkyne-dNTPs, provides an avenue to loading DNA with reporter groups for click chemistry. Carell's group performed a series of click reactions on DNA bearing a range of alkyne densities, demonstrating a method for the high-density labelling of DNA.[25,26] One application of the covalent attachment of fluorescent tags is reduced unspecific leaching which commonly causes background fluorescence as seen with intercalating dyes.[27] The occurrence of the alkyne group is determined by the genetic sequence and which artificial alkyne dNTP is incorporated. Therefore, flexibility to design the position of functionalisation is limited to the sequences commercially available and can only be chosen by the user. A method able to design the repetition and specific location of the alkyne and consequently the azido-marker, would provide an additional avenue to nanoscale assembly with specific location attachment.

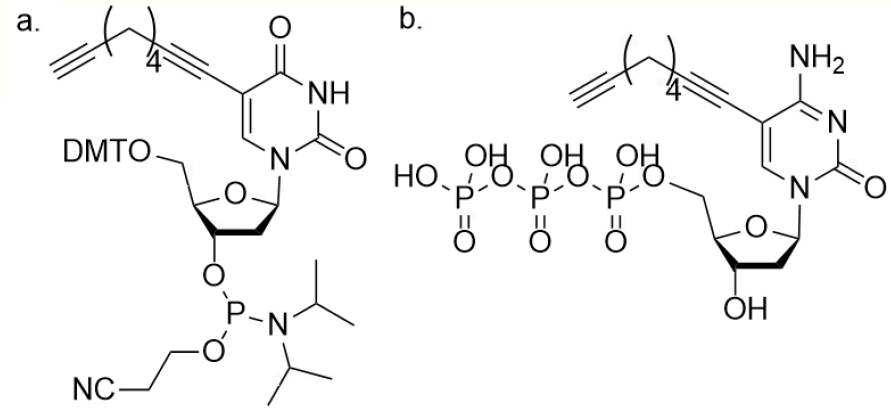


Fig. 3 (a) 5'-4,4'-dimethoxytriphenyl-methyl-O-5-octa-1,7-diynyl)-2'-dU, 3'-[2-(nitrile)-(N,N-diisopropyl)]-phosphoramidite. **(b)** 5-(Octa-1,7-diynyl)-2'-deoxycytidine 5'-triphosphate.

5.2 Results and Discussion

In this chapter, a method for the fluorescent tagging of DNA at sequence specific positions to control fluorescence loading is reported. In Chapter 2, the controlled elongation of a DNA repeat sequence DNA, of up to 10 bases was reported.[28] Through considered design of the repeat sequence, the arrangement and frequency of certain base pairs can be controlled. In Chapter 3, a range of DNA polymerases were investigated to determine relative incorporation efficiencies of modified nucleotides, including 5-C_8-alkyne-dCTP which was found to have a similar behaviour to the unmodified dCTP. DNA synthesised with multiple 5-C_8-alkyne-dCTP anchors allows for organic chemistry, specifically click modification, to be carried out at precise positions on the DNA backbone.

5.2.1 *5-C_8-alkyne-dC-DNA synthesis*

DNA extensions were performed with 5-C_8-alkyne-dCTP and the oligo seeds $[AG]_{10}$/$[TC]_{10}$, $[A_4G]_4$/$[T_4C]_4$ and $[A_9G]_2$/$[T_9C]_2$ to yield DNA with three different loadings (1 in 2, 1 in 5 and 1 in 10) of alkyne groups. Extension with each sequence yielded DNA of 2,000 bp for $[AG]_n$/$[TC]_n$, and 5,000 bp for both $[A_4G]_n$/$[T_4C]_n$ and $[A_9G]_n$/$[T_9C]_n$, Fig. 4 (a). DNA extension in the absence of either the dCTP or 5-C_8-alkyne-dCTP derivatives was not observed with any sequence, demonstrating that the DNA polymerase did not mis-incorporate a dTTP, dATP or dGTP. Once again, the extension reaction performed in the presence of only dCTP and unmodified dNTPs, yielded similar DNA product lengths as observed for the 5-C_8-alkyne-dC-DNA products for all three oligo seeds.

A test of two DNA polymerases for extension of the oligo seed $[A_4G]_4$/$[T_4C]_4$ was performed to establish the optimal enzyme for the incorporation of the 5-C_8-alkyne-dCTP: Deep Vent exo-,[29] a commercially available *Pyrococcus* species GB-D DNA polymerase and Tgo-Pol mutant, Z3 exo-, an in-house synthesised DNA polymerase were compared, Fig. 4 (b). Both DNA polymerases are exo- mutants, hence exhibit favourable properties for incorporation of modified dNTP. Both DNA polymerases produced extended DNA of similar modal length, *ca.* 5,000 bp. However, Deep Vent exo- appears to yield more DNA from the extension reactions as exhibited by the more intense bands in lanes 2 and 3 versus lanes 5 and 6. As Deep Vent exo- produced a higher yield of modified DNA, it was the chosen DNA polymerase for the enzymatic extensions performed in this Chapter.

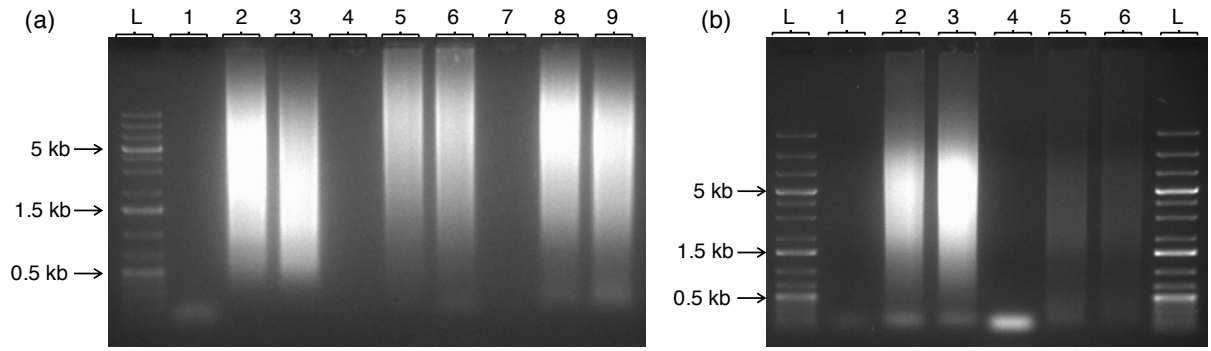


Fig. 4 Agarose gel of DNA extension products after 30 heat-cool cycles, **(a)** lanes 1-3: $[AG]_n/[TC]_n$, in the absence of dCTP, in the presence of dCTP and in exchange with 5-C_8-alkyne-dCTP, respectively, lanes 4-6: $[A_4G]_n/[T_4C]_n$ in the absence of dCTP, in the presence of dCTP and in exchange with 5-C_8-alkyne-dCTP, respectively, lanes 7-9 $[A_9G]_n/[T_9C]_n$ in the absence of dCTP, in the presence of dCTP and in exchange with 5-C_8-alkyne-dCTP, respectively. **(b)** $[A_4G]_n/[T_4C]_n$ extensions performed with Deep Vent exo-, lanes 1-3, and Tgo-Pol Z3 exo-, lanes 4-6, in the absence of dCTP, in the presence of dCTP and in exchange with 5-C_8-alkyne-dCTP, respectively. L = DNA ladder.

Fig. 5 (a) represents schematically the range of loading of the alkyne functional groups along the DNA duplex. By incorporating 5-C_8-alkyne-dCTP into $[AG]_n/[TC]_n$, the functionality will be evenly distributed throughout the DNA at every other base on one strand. It is known that C5 modifications of the pyrimidines extend into the major groove and do not interfere with hydrogen bonding between base pairs. As there are ten base pairs per turn in B-DNA, $[A_4G]_4/[T_4C]_4$ provides a method for presenting alkyne groups on alternate faces of the duplex and $[A_9G]_2/[T_9C]_2$ bears the modification on only one face of the duplex, Fig. 5 (a).

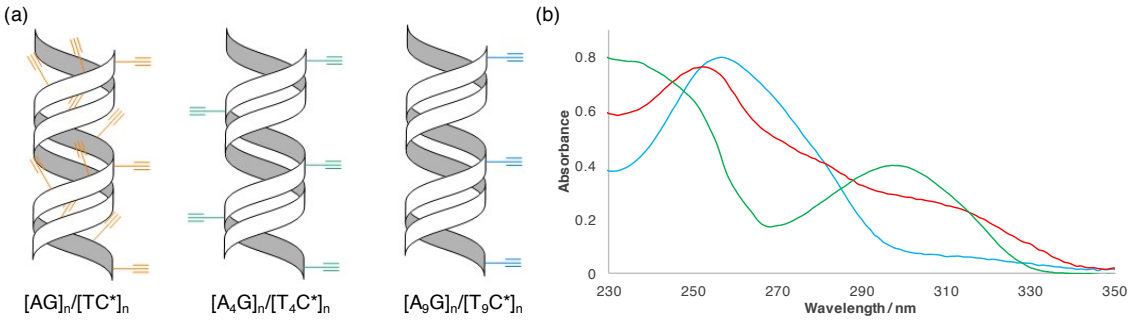


Fig. 5 (a) position of alkyne functional groups depending on the DNA sequence chosen. **(b)** UV-Vis of unmodified (blue) and alkyne-modified (red) DNA and 5-C_8-alkyne-dC (green). C* = 5-C_8-alkyne-dC.

The alkyne functionality of 5-C_8-alkyne-dCTP causes a shift in the λ_{max} of the dC-nucleotide from 271 nm to 299 nm (5-C_8-alkyne-dC spectrum shown as the green line in Fig. 5 (b)).[30,31] Therefore detection of a shoulder in the UV-Vis spectrum of the DNA extension products indicates successful incorporation of the 5-C_8-alkyne-dCTP, red line in Fig. 5 (b). Additionally, further characterisation and confirmation of 5-C_8-alkyne-dCTP incorporation was reported in Chapter 3.

5.2.2 *Click chemistry with 5-C_8-alkyne-dC-DNA*

DNA presenting alkyne groups can now undergo click chemistry for secondary modification with any desired azido-group. In this investigation, click functionalisation with fluorophores was chosen due to ease of analysis. Two different azide-modified fluorophores were selected to determine the ability to control the fluorescent loading of DNA using the three designed repeat sequences. Azide-fluor 545 is commercially available from Sigma Aldrich and the bodipy-azide was synthesised in-house and provided from the group of Prof. A. Benniston,[32] Fig. 6. Click reactions of the two azide-fluorophores were performed with both 5-C_8-alkyne-dC-DNA and unmodified DNA, Fig. 6. TBTA is required to stabilise the oxidation state of Cu(I) and produce high yields.[33] The long reaction time is necessary due to the low concentrations used (50 μM), the possible steric hindrance of the DNA duplex and the restricted diffusion of reactive alkyne sites throughout the solution.

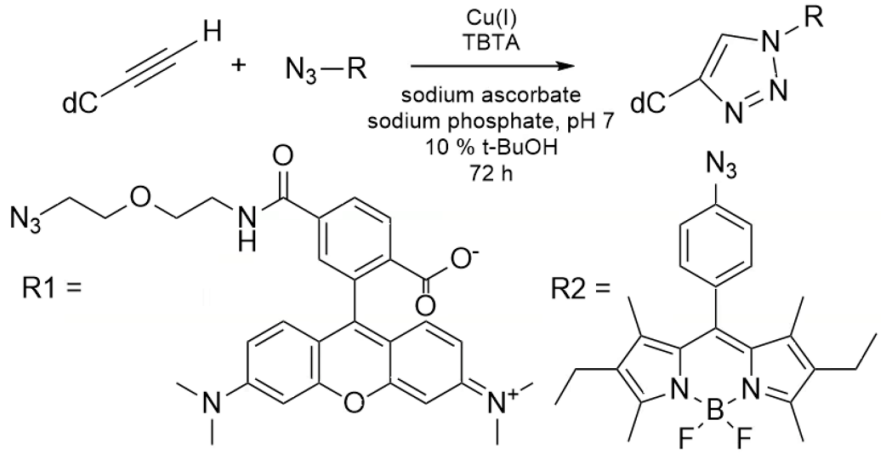


Fig. 6 The click reaction of 5-C_8-alkyne-dC with either Azide-fluor 545 (left) or bodipy-azide (right) to form the 1,2,3-triazole ring.

Purification after the click reaction was achieved using QIAQuick PCR Purification columns following the standard protocol. After DNA binding to the column and two washing steps, the pink or colourless column demonstrated the presence or absence respectively, of fluorophores in the products, Fig. 7. The DNA-click product was then eluted to release purified products from the columns.

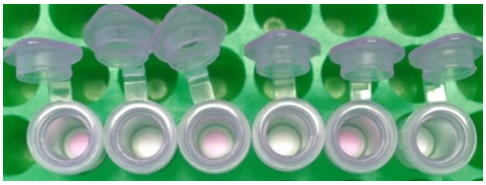


Fig. 7 QIAquick PCR Purification columns after binding of fluor 545-DNA click samples and 2 x washing with PE buffer, 1-6 (left to right). Sample 1: $[AG]_n/[TC^*]_n$, sample 2: $[AG]_n/[TC]_n$, sample 3: $[A_4G]_n/[T_4C^*]_n$ sample 4: $[A_4G]_n/[T_4C]_n$ sample 5: $[A_9G]_n/[T_9C^*]_n$, sample 6: $[A_9G]_n/[T_9C]_n$. C* = 5-C_8-alkyne-dC.

UV-Vis spectroscopic analysis of the products from the click reaction with unmodified DNA (control) and the alkyne-bearing DNA was performed. In the case of the control reaction between Azide-fluor 545 and unmodified DNA, no absorbance peak at 545 nm after 72 hours was observed, as expected, Fig. 8 (a), (c) and (e). This confirms the complete removal of any free fluorophore through efficient column purification and also indicates that intercalation of the dye into the DNA did not occur. Alkyne-bearing DNA, however, produced an absorbance peak at 545 nm after click with Azide-fluor 545, suggesting association of the fluorophore with the DNA, Fig. 8. Similar observations for the click reaction with the bodipy dye were noted as in Fig. 8 (b), (d) and (f).

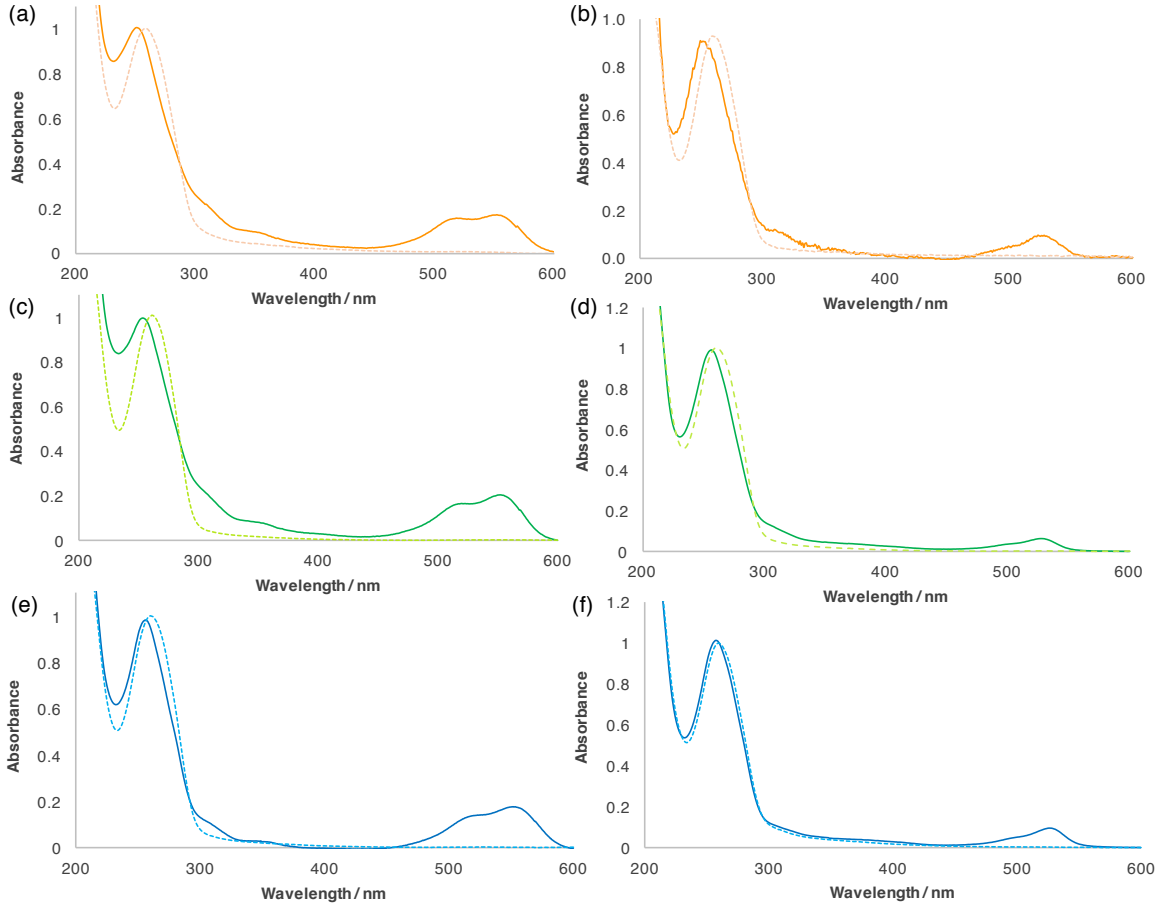


Fig. 8 UV-Vis of DNA click products from alkyne-bearing-C DNA (smooth line) and unmodified-C-DNA (dashed line). **(a)** $[AG]_n/[TC]_n$, **(c)** $[A_4G]_n/[T_4C]_n$, and **(e)** $[A_9G]_n/[T_9C]_n$ click with azide-fluor 545, **(b)** $[AG]_n/[TC]_n$ **(d)** $[A_4G]_n/[T_4C]_n$, and **(f)** $[A_9G]_n/[T_9C]_n$ click with bodipy-azide. All spectra are normalised at 260 nm.

To further characterise the click products, fluorescence spectroscopy was performed to ensure that the fluorophore remained photo-active after attachment, Fig. 9 (a). The characteristic emission band at 580 nm was observed for all three fluor 545-DNA products, however, the intensity did not decrease as alkyne loading was reduced. On normalising the emission intensity to DNA absorbance, a ratio of 2:1:4 for $[AG]_n/[TC]_n$, $[A_4G]_n/[T_4C]_n$ and $[A_9G]_n/[T_9C]_n$ was observed rather than the expected 5:2:1 ratio as

by loading. This difference in ratios between loading and emission intensity suggests additional interactions with either the DNA duplex or more likely, neighbouring fluorophores. On closer analysis of the UV-Vis spectra, Fig. 9 (b), the shoulder at 520 nm is apparent at different ratios for each clicked DNA product. At high concentrations, rhodamine is known to aggregate,[34,35] resulting in a reduced quantum yield. A reduced quantum yield is also noted on conjugation to DNA.[36] A lower emission intensity is observed for the $[AG]_n/[TC]_n$, repeating sequence, suggesting the fluorophores are in close enough proximity to interact and reduce the quantum yield. Alternatively, in the DNA sequences where reactive sites are in close proximity, it is possible that not all alkyne sites reacted.

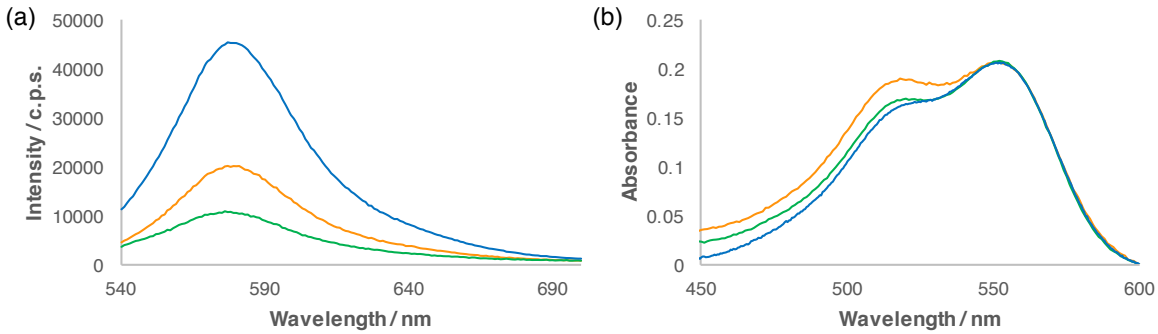


Fig. 9 (a) Fluorescence spectroscopy and **(b)** UV-Vis of each click product from $[AG]_n/[TC^*]_n$ (orange), $[A_4G]_n/[T_4C^*]_n$ (green) and $[A_9G]_n/[T_9C^*]_n$ (blue). Samples were excited at 520 nm. C* = 5-C_8-alkyne-dC.

Thus far, UV-Vis and fluorescence spectroscopic studies have shown that after the click reaction the fluorophore is present in the product, however, additional characterisation to visualise the fluorophore associated with the DNA would be beneficial. Fluorescence microscopy confirms the fluor 545 is located only on the DNA molecules, as linear bright structures, *ca.* 1,500 nm in length which corresponds to 4,500 bp of DNA, Fig. 10. Such linear features are not observed after click addition of Azide-fluor 545 to the control DNA. Further fluorescence micrographs of the click products of each sequence, with and without the alkyne group modification are shown in Appendix D.

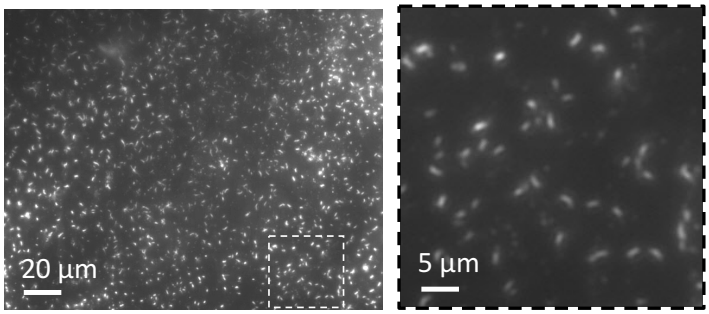


Fig. 10 Fluorescence microscopy of $[A_4G]_n/[T_4C^*]_n$ click products on glass. C* = 5-C_8-alkyne-dC.

Fluorescence microscopy expresses the association of fluorophores with DNA, however it does not necessarily confirm covalent attachment through the expected 1,2,3-triazole ring linkage. Enzymatic digestion of the DNA-click products leads to the break-down of DNA into its individual nucleosides. Therefore, in this case, the digestion of the 5-C_8-alkyne-dC-DNA-click product should give the dC-fluor 545 clicked monomer as one of the individual digestion products. HPLC of the digested click products for both the unmodified and the 5-C_8-alkyne-modified DNA showed the presence of dT, dG and dA in both samples but only dC in the unmodified DNA sample, as expected. In addition, there is no peak at 54.5 minutes corresponding to the unreacted 5-C_8-alkyne-dC peak (all standard retention times are shown in Appendix B, Table 1 and Fig. 9). Therefore, high yields are considered for the $[A_9G]_n/[T_9C]_n$ sequence due to the disappearance of the 5-C_8-alkyne-dC peak as expected from Carell's studies.[37] However, two new peaks at longer retention times of 61 and 62.5 minutes, assigned to the successfully clicked fluor-545-dC product were observed. The presence of a split peak is expected of the fluor-545 rhodamine derivative due to the equilibrium of the charged and uncharged forms.[38] The linear detection range of the HPLC is not known for the click product, hence the small product peaks may be due to limited detection at lower concentrations.

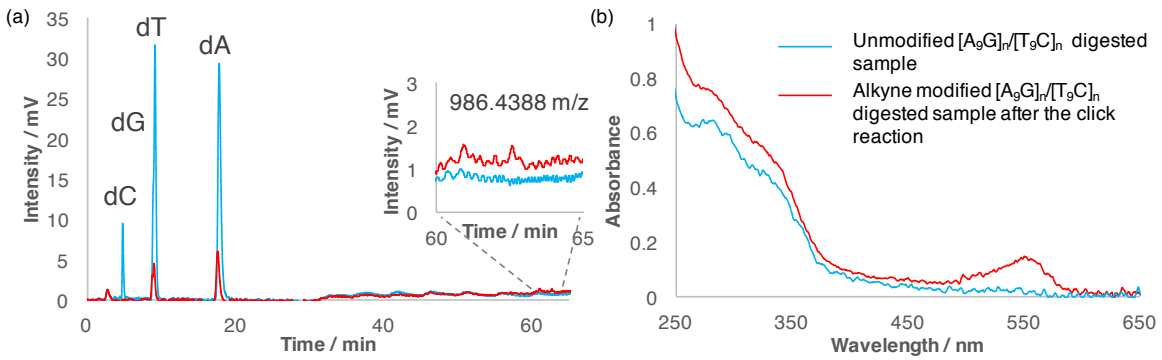


Fig. 11 (a) HPLC trace of digestion products at A^{254} and **(b)** UV-Vis spectrum of the eluent collected from 61-63 minutes.

The UV-Vis spectrum of the eluent collected from 61-63 minutes illustrated the presence of the clicked dC-rhodamine derivative, Fig. 11. To ensure the peak collected was the click product, HRMS was performed and the mass ion peak for $C_{50}H_{60}N_9O_{11}Na$, expected at 986.4388 m/z, was found at 986.4384 m/z. The mass peak without sodium was not found, suggesting the sodium counter ion is required for stabilisation. The presence of the new longer retention time peak in HPLC, which absorbs at 545 nm and has the correct molecular mass ion confirms that the click reaction was successful on long DNA bearing multiple repeating alkyne groups.

5.3 Conclusions

The flexible incorporation of 5-C_8-alkyne-dCTP into three different repeating sequences of DNA provides a route to loading DNA with chemical anchors at pre-designed positions. The enzymatic heat-cool method led to DNA extension products with modal lengths of 2,000 bp for $[AG]_{10}/[TC]_{10}$, and 5,000 bp for $[A_4G]_4/[T_4C]_4$ and $[A_9G]_2/[T_9C]_2$ after 30 cycles. The DNA extension products with the incorporation of 5-C_8-alkyne-dCTP provide alkyne groups situated in the DNA major groove available for the further functionalisation via the click reaction. The click reaction was performed with alkyne-dC derivatives of $[AG]_n/[TC]_n$, $[A_4G]_n/[T_4C]_n$ and $[A_9G]_n/[T_9C]_n$ with azide-fluor-545 and bodipy-azide. Characterisation by UV-Vis and fluorescence spectroscopy revealed the association of each fluorophore with the modified DNA. To ensure the fluorophore was covalently attached, DNA digestion followed by HPLC analysis revealed the HRMS product peak for the 5-C_8-alkyne-dC-fluorophore click product.

In this study, the ability to further functionalise designer modified DNA through the 1,3-dipolar cycloaddition click reaction has been demonstrated. The location of alkyne reactive sites along the helix can be controlled by the design of the starting oligo seed. Each alkyne site provides an anchor for further chemistry and could be a promising route to exploit DNA as an advanced tool in diagnostics and in delivery systems.

5.4 Experimental details

Preparation and Purification of DNA polymerases

Deep Vent exo- was purchased from New England Biolabs (Ipswich) and Tgo-Pol Z3 exo- was prepared and purified as described previously.[39,40]

Deoxynucleotide preparation

Oligos were purchased from Eurofins (Ebersberg, Germany) and complementary strands were annealed as shown in Table 1.

Oligos to be combined	Final oligo seed	Duplex length / bp
$(AG)_{10}$ and $(TC)_{10}$	$[AG]_{10}/[TC]_{10}$	20
(AAAAG) and (TTTTC)	[AAAAG]/[TTTTC]	20
(AAAAAAAAAG) and (TTTTTTTTTC)	[AAAAAAAAAG]/[TTTTTTTTTC]	20

Table 1 Summary of the oligos required to form the oligo seeds used.
Deoxyguanosine (G), deoxythymidine (T), deoxycytidine (C), deoxyadenine (A) and base pair (bp)

Heat-cool cycle DNA extension

0.5 μM oligo seed (Table 1), 200 nM DNA polymerase, DNA polymerase reaction buffer (200 mM Tris-HCl (pH 8.8, 25 °C), 100 mM (NH)SO, 100 mM KCl, 1 % Triton X-100, 1 mg/mL Bovine serum albumin (BSA) and 20 mM MgSO), and 0.5 mM deoxynucleotide triphosphates (dNTPs (Thermo Scientific, UK)) (dCTP, dATP, dTTP and dGTP or 5-C_8-alkyne-dCTP (Jena Biosciences, Germany)) were mixed. Thermocycling was carried out using an Applied Biosciences Veriti 96 well Thermal Cycler by the following method:

N (number of cycles stated) x 30 seconds at 95 °C, 30 seconds at 55 °C and 120 seconds at 72 °C.

The products were cooled to 4 °C after the reaction and purified using a QIAquick PCR Purification Kit (25) (QIAGEN, Manchester, UK) following manufacturers protocol.

Click reaction

50 μM 5-C_8-alkyne-dC in DNA was reacted with 500 μM azide-fluoro-545 (Sigma Aldrich, UK) or bodpy-azide (a kind gift from Prof. A. Benniston), 500 μM CuI, 1 mM TBTA, 5 mM sodium ascorbate in 10 % isopropanol and 10 mM sodium phosphate, pH 7 in a total volume of 200 μL and gently shaken for 72 hours. Samples were purified using QIAquick PCR Purification Kits. On binding the sample to the column, reaction outcome is expressed by colouring the matrix, Fig. 7.

Agarose Gel Electrophoresis

The extension products were analysed by gel electrophoresis in TBE (Tris, boric Acid and NaEDTA.2H_2O) buffer. 1 % agarose (Melford, Ipswich, UK) was added to the 1 % TBE buffer and heated to dissolve. The agarose solution was supplemented with ethidium bromide (Sigma Aldrich) and poured once it had cooled to 50 °C. The 1 kb+ ladder Thermo Scientific) was provided with a loading dye (2.5 % Ficoll-400, 11 mM EDTA, 3.3 mM Tris-HCl (pH 8.0, 25 °C), 0.017 % sodium dodecyl sulphate and 0.015 % bromophenol blue). DNA samples were supplemented with the gel loading dye. The gels were run at 100 V, 100 mA, 10 W for approximately 1 hour and then visualised using an ultra-violet transilluminator.

Ultraviolet-Visible spectroscopy

UV-Vis spectroscopy was performed using a NanoDrop 1000 Spectrometer (Thermo Scientific) for volumes less than 2 μL and a Varian Cary 100Bio UV-Vis spectrophotometer with a Varian Cary temperature controller for larger volumes. Spectrometer was blanked using nanopure-H_2O or QIAGEN elusion buffer.

DNA digestion

0.2 mg of snake venom phosphodiesterase, 100 units bacterial alkaline phosphotase in 10 mM potassium phosphate buffer, pH 7 and 10 mM magnesium chloride was added to 0.5 absorbance units at 260 nm of ds DNA and incubated at 37 °C for 16-18 hours. The dNMPs produced can be analysed by reverse phase HPLC on an APEX ODS C18 5μm 250 mm column with 0.1 M triethylammonium acetate, pH 6.5 containing 5 % ACN (buffer A) and 0.1 M triethylammonium acetate, pH 6.5 containing 65 % ACN (buffer B) operated at 1 $mLmin^{-1}$ at room temperature. 100 μL of digestion product was injected into a Waters 2487 Dual Wavelength Absorbance detector with a Waters 600 controller under the buffer system described in Table 2.

Time / min	Buffer A %	Buffer B %
0	100	0
25	5	95
30	10	90
40	15	85
45	20	80
50	25	75
55	30	70
60	35	65
70	100	0

Table 2 HPLC buffer system

Mass spectroscopy

A Waters Micromass LCT Premier TOF system (Hertfordshire, UK) was used in positive mode, with a desolvation temperature of 250 °C and data collected using masslynx v. 4.1. Samples were injected directly in nanopure-H_2O.

dN standards

All standards were purchased from Jena Biosciences (Jena, Germany) except 5-C_8-alkyne-dC (5-(octa-1,7-diynyl)-cytosine) which was synthesised in house following the protocol by Chittepu *et al.*[30] 72 %, TLC (CH_2Cl_2/MeOH 9:1): R_f 0.41. UV λ_{max} (MeOH)/nm 299. ^{1}H NMR ((D_6) DMSO): 1.52-1.67 (m, 4H $2CH_2$), 1.96-2.17 (m, 2H CH_2(2')), 2.19-2.23 (m, 2H CH_2), 2.42-2.45 (t, *J*=2.44 CH_2), 2.78 (t, 1H C≡CH), 3.54-3.65 (m, 2H CH_2(5')), 3.78-3.81 (m, HC(4')), 4.20-4.24 (m, HC(3')), 5.06-5.09 (t, *J*=5.08 OH-C(5')), 5.23-5.24 (d, *J*=5.23 OH-C(3')), 6.11-6.14 (t, *J*=6.13 H-C(1')), 6.74 (s, NH_a), 7.69 (s, NH_b), 8.08 (s, H-C(6)). ^{13}C NMR ((D_6) DMSO): 17.7 (CH_2-C≡H), 19.0 (CH_2-C≡H), 27.6 (CH_2CH_2-C≡H), 27.7 (CH_2CH_2-C≡H), 39.9 (C-2'), 61.5 (C-5'), 70.6 (C-3'), 71.8 (C≡CH), 72.6 (C≡C), 84.8 (C-1'), 85.7 (C-4'), 87.9 (C≡C), 90.8 (C≡C), 95.8 (C-5), 144.0 (C-6), 154.0 (C-2), 164.8 (C-4). HRMS (ESI +ve) calcd. for $C_{17}O_4N_3H_{21}$ $[M-H]^+$ 332.1610, found 332.1589.

Fluorescent microscope imaging

DNA samples were diluted to approximately 10 ng/μL. Glass slides were cleaned using a diener electronic plasma cleaner (Femto, Korea) and 5 μL sample was deposited using a micro pipette and spread on the surface. Sample was allowed to dry and mounted onto an Axioshop 2 plus (Zeiss, Germany) image platform, set to filter 44, with a Plan-NEOFLUAR 40 x/ 0.75 objective lens (Zeiss). The sample was excited at 490 nm from a ebq100 isolated mercury lamp (LEJ, Germany) and captured using an axioCam HRm (Zeiss).

5.5 References

1 Dai, N. & Kool, E. T. Fluorescent DNA-based enzyme sensors. *Chem Soc Rev* **40**, 5756-5770 (2011).

2 Wang, K. *et al.* Molecular engineering of DNA: molecular beacons. *Angew Chem Int Ed Engl* **48**, 856-870 (2009).

3 Didenko, V. V. DNA probes using fluorescence resonance energy transfer (FRET): Designs and applications. *Biotechniques* **31**, 1106 (2001).

4 Waring, M. J. Complex formation between ethidium bromide and nucleic acids. *J Mol Biol* **13**, 269-282 (1965).

5 Fujimoto, K. Highly Emissive Pyrene-based Fluorophores and Highly Sensitive Fluorescent Sensors Using Pyrene Emission Switching. *Yakugaku Zasshi* **130**, 1283-1287 (2010).

6 Huang, Q. & Fu, W. L. Comparative analysis of the DNA staining efficiencies of different fluorescent dyes in preparative agarose gel electrophoresis. *Clin Chem Lab Med* **43**, 841-842 (2005).

7 Moradpour Hafshejani, S., Watson, S. M., Tuite, E. M. & Pike, A. R. Click modification of diazido acridine intercalators: a versatile route towards decorated DNA nanostructures. *Chemistry Eur. J.* **21**, 12611-12615 (2015).

8 MoradpourHafshejani, S., Hedley, J. H., Haigh, A. O., Pike, A. R. & Tuite, E. M. Synthesis and binding of proflavine diazides as functional intercalators for directed assembly on DNA. *Rsc Adv* **3**, 18164-18172 (2013).

9 Roget, A., Bazin, H. & Teoule, R. Synthesis and use of labelled nucleoside phosphoramidite building blocks bearing a reporter group: biotinyl, dinitrophenyl, pyrenyl and dansyl. *Nucleic Acids Res* **17**, 7643-7651 (1989).

10 Theisen, P., McCollum, C., Upadhya, K., Jacobson, K., Vu, K. & Andrus, A. Fluorescent dye phosphoramidite labelling of oligonucleotides. *Tetrahedron Lett* **33** 5033-5036 (1992).

11 Hall, L. M., Gerowska, M. & Brown, T. A highly fluorescent DNA toolkit: synthesis and properties of oligonucleotides containing new Cy3, Cy5 and Cy3B monomers. *Nucleic Acids Res* **40**, e108 (2012).

12 Thoresen, L. H., Jiao, G. S., Haaland, W. C., Metzker, M. L. & Burgess, K. Rigid, conjugated, fluoresceinated thymidine triphosphates: syntheses and polymerase mediated incorporation into DNA analogues. *Chemistry* **9**, 4603-4610 (2003).

13 Vrabel, M., Pohl, R., Klepetarova, B., Votruba, I. & Hocek, M. Synthesis of 2'-deoxyadenosine nucleosides bearing bipyridine-type ligands and their Ru-complexes in position 8 through cross-coupling reactions. *Org Biomol Chem* **5**, 2849-2857 (2007).

14 Brazdilova, P. *et al.* Ferrocenylethynyl derivatives of nucleoside triphosphates: synthesis, incorporation, electrochemistry, and bioanalytical applications. *Chemistry* **13**, 9527-9533 (2007).

15 Capek, P. *et al.* An efficient method for the construction of functionalized DNA bearing amino acid groups through cross-coupling reactions of nucleoside triphosphates followed by primer extension or PCR. *Chemistry* **13**, 6196-6203 (2007).

16 Casalnuovo, A. L. & Calabrese, J. C. Palladium-Catalyzed Alkylations in Aqueous-Media. *Journal of the American Chemical Society* **112**, 4324-4330 (1990).

17 Proudnikov, D. & Mirzabekov, A. Chemical methods of DNA and RNA fluorescent labeling. *Nucleic Acids Res* **24**, 4535-4542 (1996).

18 Worrell, B. T., Malik, J. A. & Fokin, V. V. Direct Evidence of a Dinuclear Copper Intermediate in Cu(I)-Catalyzed Azide-Alkyne Cycloadditions. *Science* **340**, 457-460 (2013).

19 Kolb, H. C., Finn, M. G. & Sharpless, K. B. Click Chemistry: Diverse Chemical Function from a Few Good Reactions. *Angew Chem Int Ed* **40**, 2004-2021 (2001).

20 Fliege, W., Grashey, R. & Huisgen, R. 1,3-Dipolar Cyclo-Additions .91. The Chemistry of N-Methyl-C-Phenylnitrilimine. *Chem Ber-Recl* **117**, 1194-1214 (1984).

21 Gramlich, P. M., Wirges, C. T., Manetto, A. & Carell, T. Postsynthetic DNA modification through the copper-catalyzed azide-alkyne cycloaddition reaction. *Angew Chem Int Ed* **47**, 8350-8358 (2008).

22 Seela, F. & Sirivolu, V. R. DNA containing side chains with terminal triple bonds: Base-pair stability and functionalization of alkynylated pyrimidines and 7-deazapurines. *Chem Biodivers* **3**, 509-514 (2006).

23 Ramzaeva, N., Mittelbach, C. & Seela, F. 7-deazaguanine DNA: Oligonucleotides with hydrophobic or cationic side chains. *Helv Chim Acta* **80**, 1809-1822 (1997).

24 Gramlich, P. M., Warncke, S., Gierlich, J. & Carell, T. Click-click-click: single to triple modification of DNA. *Angew Chem Int Ed Engl* **47**, 3442-3444 (2008).
25 Gierlich, J., Burley, G. A., Gramlich, P. M., Hammond, D. M. & Carell, T. Click chemistry as a reliable method for the high-density postsynthetic functionalization of alkyne-modified DNA. *Org Lett* **8**, 3639-3642 (2006).
26 Gutsmiedl, K., Fazio, D. & Carell, T. High-density DNA functionalization by a combination of Cu-catalyzed and cu-free click chemistry. *Chemistry* **16**, 6877-6883 (2010).
27 Stadler, A. L. *et al.* Fluorescent DNA nanotags featuring covalently attached intercalating dyes: synthesis, antibody conjugation, and intracellular imaging. *Bioconjug Chem* **22**, 1491-1502 (2011).
28 Whitfield, C. J., Turley, A. T., Tuite, E. M., Connolly, B. A. & Pike, A. R. Enzymatic Method for the Synthesis of Long DNA Sequences with Multiple Repeat Units. *Angew Chem Int Ed* **54**, 8971-8974 (2015).
29 Jannasch, H. W., Wirsen, C. O., Molyneaux, S. J. & Langworthy, T. A. Comparative Physiological-Studies on Hyperthermophilic Archaea Isolated from Deep-Sea Hot Vents with Emphasis on Pyrococcus Strain Gb-D. *Appl Environ Microb* **58**, 3472-3481 (1992).
30 Seela, F., Sirivolu, V. R. & Chittepu, P. Modification of DNA with octadiynyl side chains: synthesis, base pairing, and formation of fluorescent coumarin dye conjugates of four nucleobases by the alkyne--azide "click" reaction. *Bioconjug Chem* **19**, 211-224 (2008).
31 Cavaluzzi, M. J. & Borer, P. N. Revised UV extinction coefficients for nucleoside-5'-monophosphates and unpaired DNA and RNA. *Nucleic Acids Res* **32**, e13 (2004).
32 Bai, D., Benniston, A. C., Whittle, V. L., Lemmetyinen, H. & Tkachenko, N. V. ROFRET: A Molecular-Scale Fluorescent Probe Displaying Viscosity-Enhanced Intramolecular Forster Energy Transfer. *Chemphyschem* **15**, 3089-3096 (2014).
33 Chan, T. R., Hilgraf, R., Sharpless, K. B. & Fokin, V. V. Polytriazoles as copper(I)-stabilizing ligands in catalysis. *Organic Letters* **6**, 2853-2855 (2004).
34 Ogawa, M., Kosaka, N., Choyke, P. L. & Kobayashi, H. H-Type Dimer Formation of Fluorophores: A Mechanism for Activatable, in Vivo Optical Molecular Imaging. *Acs Chem Biol* **4**, 535-546 (2009).
35 Setiawan, D., Kazaryan, A., Martoprawiro, M. A. & Filatov, M. A first principles study of fluorescence quenching in rhodamine B dimers: how can quenching occur in dimeric species? *Phys Chem Chem Phys* **12**, 11238-11244 (2010).
36 Vamosi, G., Gohlke, C. & Clegg, R. M. Fluorescence characteristics of 5-carboxytetramethylrhodamine linked covalently to the 5' end of oligonucleotides: Multiple conformers of single-stranded and double-stranded dye-DNA complexes. *Biophysical Journal* **71**, 972-994 (1996).
37 Gierlich, J. *et al.* Synthesis of highly modified DNA by a combination of PCR with alkyne-bearing triphosphates and click chemistry. *Chem-Eur J* **13**, 9486-9494 (2007).
38 Hell, S. *et al.* Novel fluorinated rhodamines as photostable fluorescent dyes for labelling and imaging techniques Germany patent (2014).
39 Evans, S. J. *et al.* Improving dideoxynucleotide-triphosphate utilisation by the hyper-thermophilic DNA polymerase from the archaeon Pyrococcus furiosus. *Nucleic Acids Res* **28**, 1059-1066 (2000).
40 Jozwiakowski, S. K. & Connolly, B. A. A modified family-B archaeal DNA polymerase with reverse transcriptase activity. *Chembiochem* **12**, 35-37 (2011).

Outlook

The enzymatic synthesis techniques outlined in this thesis provide a platform to build DNA based nanomaterials with an advanced level of precision and control. The initial protocol involves the enzymatic synthesis of DNA consisting of repeating sequences up to 10 bases. This extension method utilises the heating and cooling parameters of the polymerase chain reaction which results in incomplete annealing by one or more repeat sequences, providing 'sticky ends' for the DNA polymerase to extend. Several repeat sequences of varying G-C percentages ($[A_xG]_n/[T_xC]_n$, where x = 1 to 9 and n = number of repeat units) were extended up to 20,000 bp. Mixed repeat sequences were also shown to extend, for example, $[GATC]_n/[CTAG]_n$. Specific DNA lengths (500, 1,500 and 7,000 bp) were selected by the simple excision and extraction from agarose gels. This method requires a DNA polymerase compatible with thermal cycling, therefore the activity of several thermophilic DNA polymerases was investigated. The Archaeal Family B Tgo-Pol exo- mutant, Z3, the Pfu-Pol exo- and the commercially available Pfu-Pol exo+ and Deep Vent exo- all exhibited similar extension efficiency except the Pfu-Pol exo+, suggesting an exo- variant is required. As Tgo-Pol Z3 exo- is known to allow for some base mis-incorporation, it was selected to investigate the synthesis of modified DNA using non-standard dNTPs.

Single atom modifications, 5-I-dCTP, 5-Br-dUTP, 6-S-dGTP and 7-deaza-I-dATP and a long chain modification 5-C_8-alkyne-dCTP were incorporated into two oligo seeds, $[A_4G]_4/[T_4C]_4$ and $[GATC]_5/[CTAG]_5$. These oligo seed sequences allow for the study of single and multiple modified dNTP incorporations per repeating unit. Each of the five modified dNTPs were able to incorporate into the dsDNA efficiently, except 6-S-dGTP which may interfere with base pairing and duplex stability. Two or more different modified dNTPs can be incorporated simultaneously to yield DNA bearing up to four different functional groups within a designer repeat sequence determined by the initial oligo seed.

The incorporation of 6-S-dGTP, provides thio-anchors for site directed metal ion binding within the duplex. Au^{+}, Ni^{2+}, Cd^{2+} and Au^{3+} were titrated with 6-S-dG-DNA. Cd^{2+} expressed the strongest binding followed by Au^{3+}, Au^{+} and Ni^{2+}. Job plots revealed the preferred binding ratios of 2:1 for Au^{+} and 3:1 for Ni^{2+}, Cd^{2+} and Au^{3+}. However, a 1:1 binding mode was also expressed for Au^{+}, Ni^{2+} and Au^{3+} and a 2:1 binding ratio for Cd^{2+}, indicating mixed binding geometries in all cases. A binding ratio higher than 1:1 for $[A_4{}^{6S}G]_n/[T_4C]_n : M^{y+}$ (y = metal charge) requires the close proximity of two or more DNA duplexes. This may be unfavourable due to steric constraints and repulsive

charges between the assembled duplexes resulting in some 1:1 binding modes for metal ions with a higher K_d - Ni^{2+} and Au^{+}. AFM analysis of the DNA hybrid structure revealed a double height increase in comparison to bare DNA for $[A_4{}^{6S}G]_n/[T_4C]_n$: Ni^{2+} structures, suggesting the metal ion directed assembly of multiple helices.

Oligo seed extension using 6-S-dGTP produced DNA modal lengths up to a maximum of 500 bp (170 nm) which makes visualisation and analysis by AFM challenging. Ideally, DNA extension products over 700 bp would provide a more accessible building block for the directed assembly of nanomaterials. Therefore, the incorporation of α-S-dNTPs provides an alternative avenue to the synthesis of long functional DNA (up to 2,000 bp (680 nm)). The α-S-DNA can be exploited for NP deposition, through the interaction between gold and the thioate-group. The oligo seed sequence can be designed to control the α-S-dNTP position and hence the anchor location for NP binding. Uniform deposition of 3 nm AuNPs along α-S-DNA was observed by AFM after repetitive incubation with NPs, which was not evident with unmodified DNA of the same sequence. This approach could potentially be used to space NPs at user defined distances determined by the location of the α-S-group within the repeating sequence.

Similarly, DNA extended using 5-C_8-alkyne-dCTP provides a synthetic route to DNA bearing chemical anchors for click chemistry. Again, through appropriate design of the oligo seed, the spacing between chemical anchors can be controlled. The click reaction was performed successfully with azide-modified fluorophores to yield fluorescent DNA. DNA loaded with a clicked fluorophore at every other, every 5^{th} and every 10^{th} base exhibited complex emission spectra potentially due to intra-strand interactions.

The work presented in this thesis provides a flexible method to design and synthesise DNA with control over length, composition and functionality. Initial investigations into exploiting the designer DNA for the fabrication of nanomaterials has been demonstrated.

Some future considerations to further optimise the DNA product from the heat-cool cycle extension method, include amplification of the extended DNA due to the relatively low yields of the agarose gel recovery system. A standard PCR amplification method could be adopted, however, due to the repeating sequence, unspecific primer binding could lead to shorter DNA lengths. Therefore, the ligation of a short non-repeating

sequence to each end of the product DNA, would provide a specific attachment point for the complementary primer for standard PCR amplification to occur.

Additionally, variation of the composition and length of the repeating sequences could provide near infinite possibilities for designer DNA synthesis. In each instance, the optimisation of the extension parameters would need careful consideration. Longer repeat lengths would allow the spacing of biologically important sequences throughout a user defined DNA product which may have applications in biomedical research.

Here, six different modified nucleotides were incorporated into DNA by the heat-cool cycle method. Analysis of the extension products indicated that the exo- DNA polymerases are capable of handling a range of dNTP substrates. Therefore, the incorporation of a wider range of modified dNTPs would provide additional chemical anchors for secondary functionalisation.

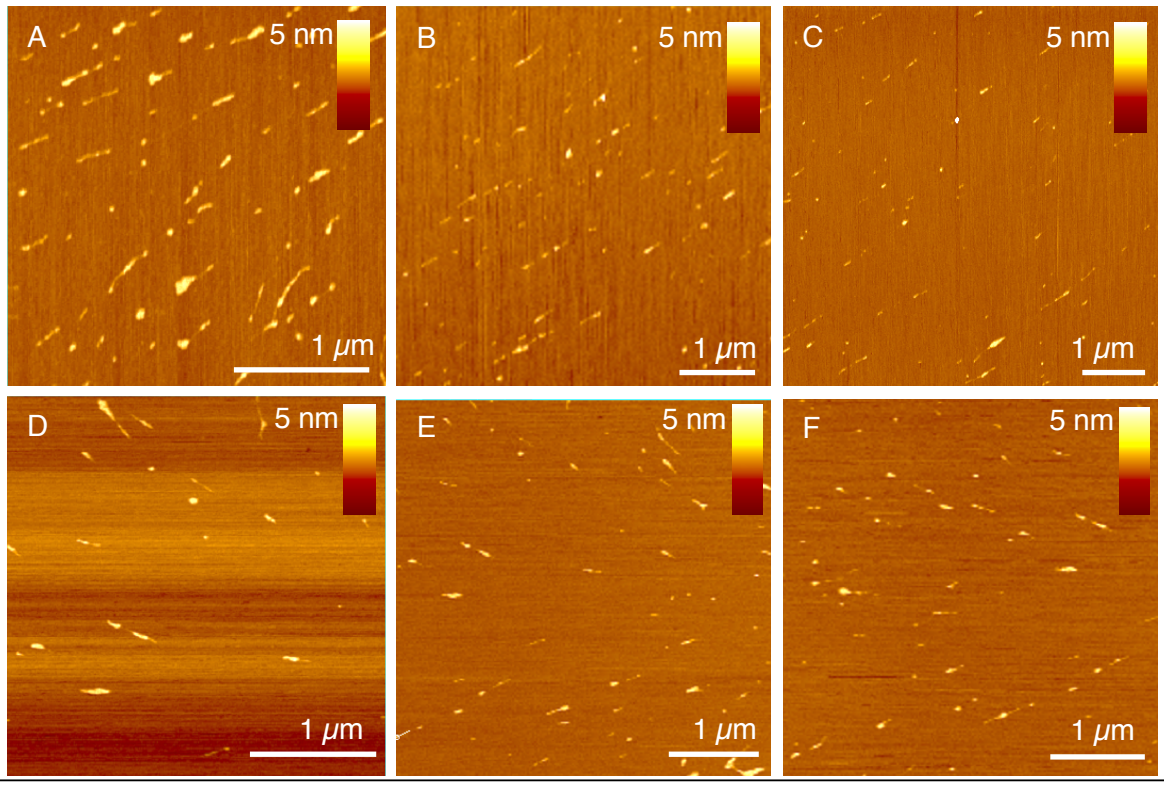
The successful site specific binding of Au-NPs to α-S-DNA, suggests the NPs can be spatially positioned along the DNA backbone. Further optimisation to extend the 40-base repeat sequence would allow for larger separations between particle binding sites. High resolution TEM analysis could provide an understanding into the role of spacing between NPs on DNA-molecule interactions through vibrational spectroscopy. Similarly, the controlled spatial positioning of chemical anchors along the DNA allows for the immobilisation of functional molecules at pre-designed positions. The example presented here was the alkyne-azide click reaction, however, the incorporation of an iodo-modified dNTP could provide a route to perform the Sonogashira reaction at pre-determined sites on DNA. The ability to perform two different chemical reactions on DNA would provide an avenue to attach two different molecules. The spatial separation between chemically active sites could be controlled to study small molecule interactions between one another and DNA.

In conclusion, this method of enzymatic DNA extension provides a protocol for the design and synthesis of DNA based nanomaterials, tailored by length, sequence and functional content.

Appendix

Appendix A

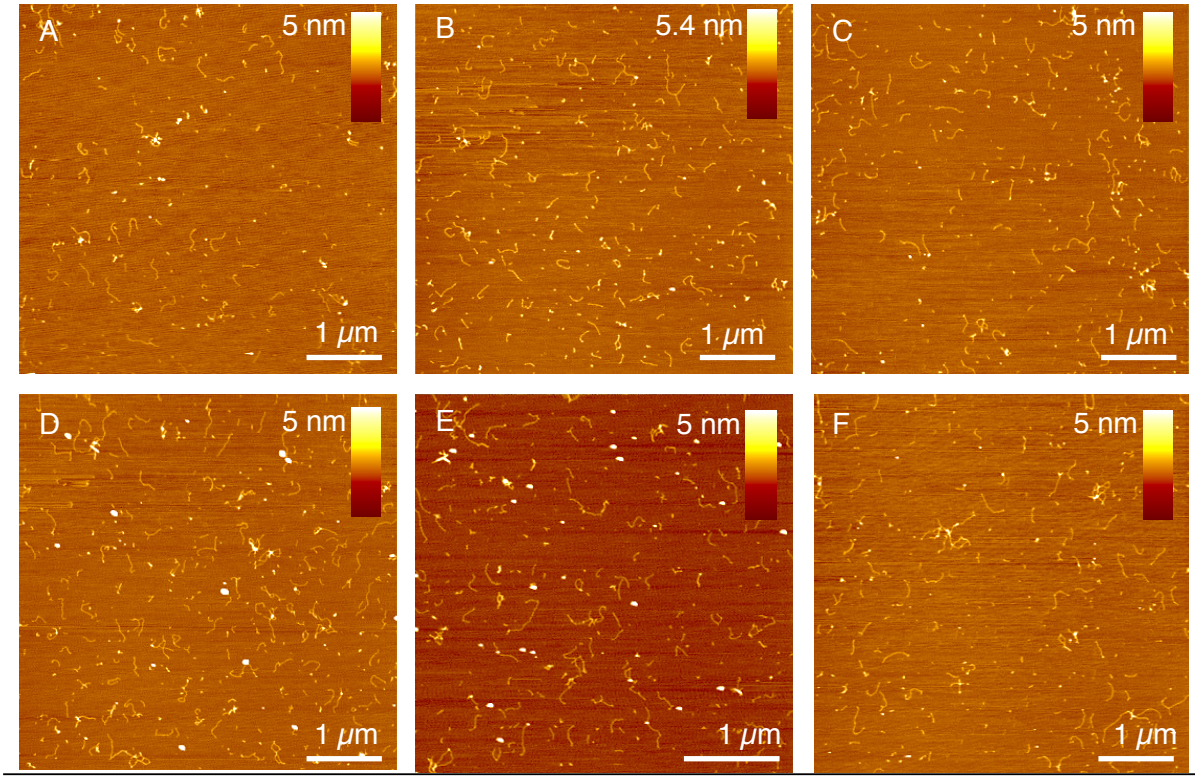
AFM and fluorescent microscopy images of DNA extension products excised from an agarose gel at 500, 1,500 and 7,000 bp, Fig. 1-3.



	A		B		C		D		E		F	
	Length (nm)	Height (nm)	Length (nm)	Height (nm)	Length (nm)	Height (nm)	Length (nm)	Height (nm)	Length (nm)	Height (nm)	Length (nm)	Height (nm)
	205	0.9	202	0.6	165	0.5	145	1.0	178	1.1	190	0.7
	162	1.2	205	1.2	199	0.9	140	0.9	145	0.7	157	0.4
	173	0.6	203	0.5	179	0.8	181	1.0	195	0.6	173	1.1
	163	0.8	179	0.9	152	0.4	163	0.5	178	1.2	172	1.2
	178	1.1	160	0.5	216	0.6	166	1.3	200	1.1	194	0.7
	179	1.0	191	0.5	176	0.5	199	1.4	165	0.5	156	1.0
	143	1.2	187	0.9	148	0.6	143	1.2	198	0.8	190	0.5
	192	1.0	179	0.6	209	0.7	161	1.1	195	1.2	153	1.2
	145	1.1	176	0.5	140	0.6	180	1.2	163	0.4	160	0.8
	182	0.6	204	0.7	133	0.6	173	1.0	169	0.7	168	0.6
Average	172	1.0	189	0.7	172	0.6	165	1.1	179	0.8	171	0.8
SD	19	0.2	14	0.2	28	0.1	18	0.2	17	0.3	15	0.3

Summary	Length (nm)	Length (bp)	Height (nm)
Average	175	513	0.8
SD	18	54	0.2

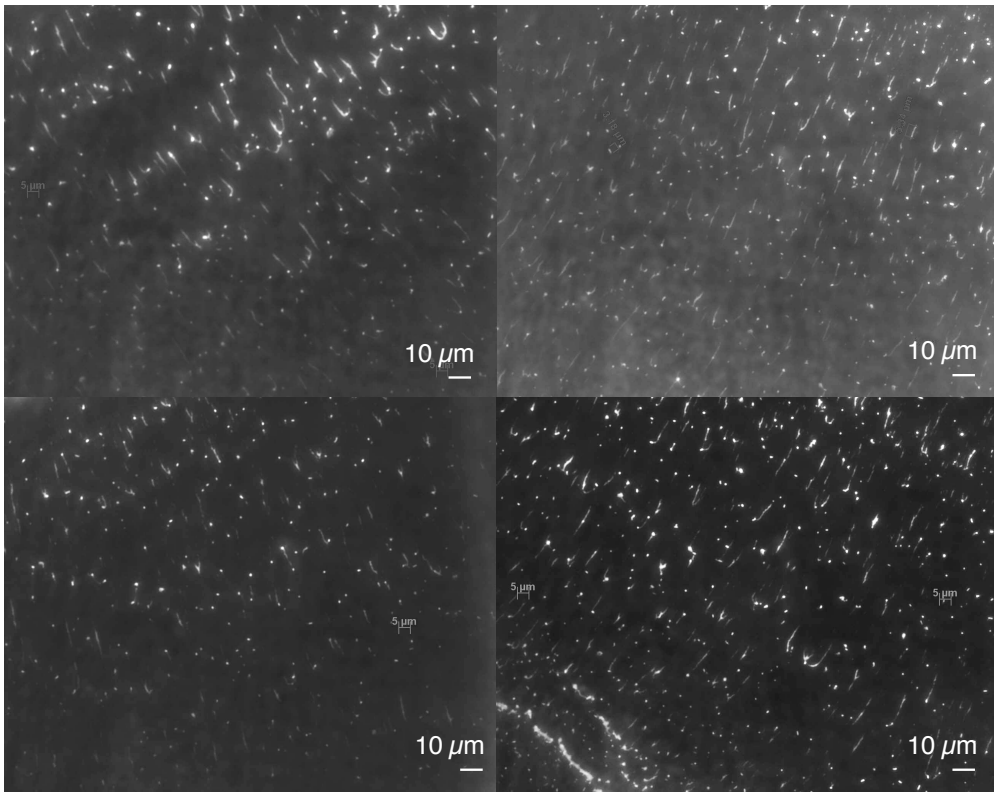
Fig. 1 AFM images of DNA extension products on mica. $[A_3G]_5/[T_3C]_5$ DNA extension products after 20 cycles and size recovered at 0.5 kb.



	A		B		C		D		E		F	
	Length (nm)	Height (nm)	Length (nm)	Height (nm)	Length (nm)	Height (nm)	Length (nm)	Height (nm)	Length (nm)	Height (nm)	Length (nm)	Height (nm)
	293	0.8	447	0.8	416	0.6	491	0.7	501	0.5	426	1.0
	361	0.7	323	1.2	403	0.8	379	0.5	371	0.6	560	0.8
	395	1.0	304	0.7	386	0.5	441	0.8	524	1.1	428	0.9
	414	0.7	442	0.6	410	0.6	305	0.7	525	0.5	412	0.5
	402	0.7	422	0.8	379	0.5	482	0.8	426	0.8	454	1.1
	398	0.8	430	1.0	376	1.1	373	0.5	507	0.8	470	0.5
	451	1.1	451	0.5	444	0.8	428	0.6	430	0.6	364	0.6
	349	0.6	384	0.7	436	0.9	335	0.9	511	0.9	407	0.9
	406	0.7	419	0.6	393	0.7	454	0.8	412	0.5	382	0.7
	457	0.6	513	0.8	453	0.7	405	0.6	469	0.7	416	0.5
Average	404	0.8	414	0.8	410	0.7	409	0.7	468	0.7	432	0.8
SD	36	0.2	62	0.2	27	0.2	61	0.1	54	0.2	55	0.2

Summary	Length (nm)	Length (bp)	Height (nm)
Average	421	1238	0.7
SD	56	165	0.2

Fig. 2 AFM images of DNA extension products on mica. $[A_3G]_5/[T_3C]_5$ DNA extension products after 20 cycles and size recovered at 1.5 kb.



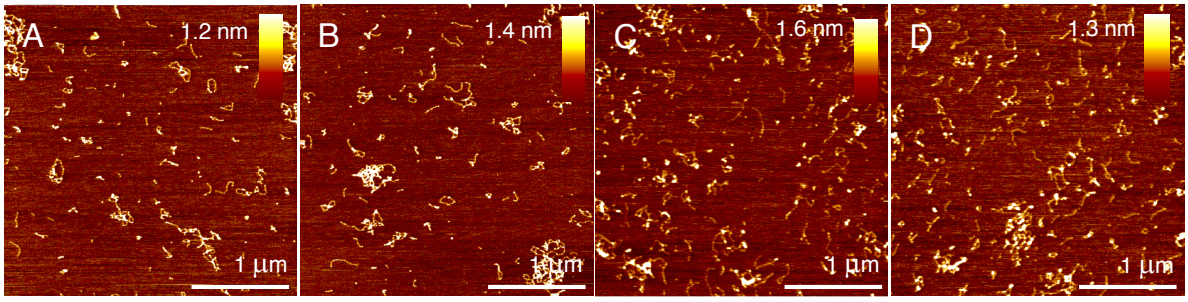
	A length (μm)	B length (μm)	C length (μm)	D length (μm)
	3.05	3.53	4.32	2.55
	2.61	3.32	2.76	3.36
	3.11	2.27	3.94	3.94
	2.50	2.62	2.76	3.37
	3.37	3.63	2.35	4.38
	4.04	4.11	2.55	3.32
	3.67	3.43	3.10	3.87
	3.35	2.63	2.87	3.83
	3.09	2.78	3.11	4.32
	3.23	4.36	2.96	3.52
Average	3.20	3.27	3.07	3.65
SD	0.43	0.65	0.58	0.51

Summary	Length (μm)	Length (bp)
Average	3.30	9700
SD	0.54	1600

Fig. 3 Fluorescent micrographs of 7 kb DNA synthesised using $[A_3G]_5/[T_3C]_5$ as the oligo seed and extending for 20 cycles followed by size recovery from an agarose gel.

Appendix B

AFM of modified DNA extension products after 30 heat-cool cycles, Fig. 1-5. Five DNA length scans were performed for each image, see Tables and several height scans are plotted, bottom right of Fig. 1-5.



	A	B	C	D
Length (nm)	838	1188	1032	853
	976	897	732	735
	706	897	1029	885
	1135	653	1003	900
	1124	794	935	829
Average (nm)	**956**	**886**	**946**	**841**
SD (nm)	166	176	113	58

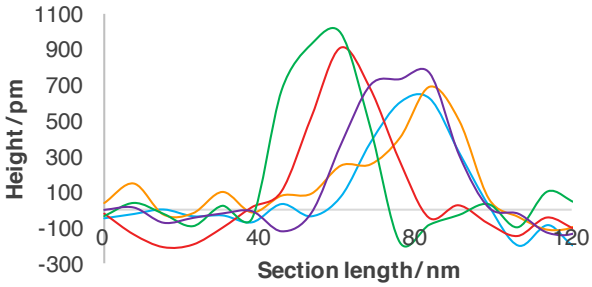
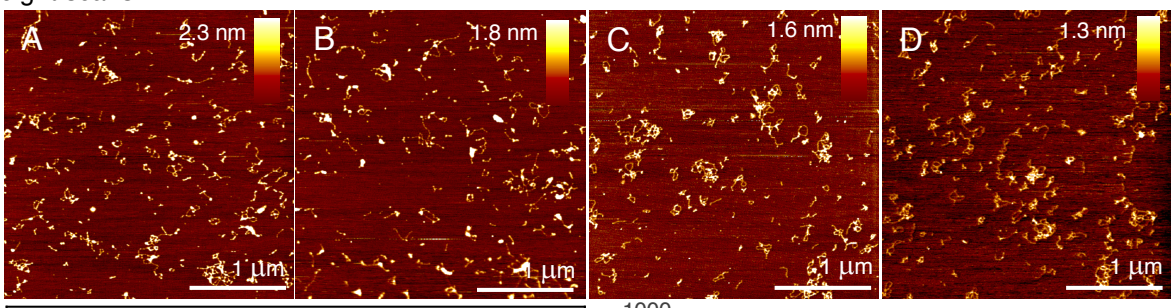


Fig. 1 AFM images (A-D) of 5-I-dC-DNA, a table of the analysed lengths and a selection of perpendicular height scans.



	A	B	C	D
Length (nm)	715	1050	571	609
	471	1026	768	512
	547	1206	1015	991
	824	865	1165	1165
	835	1100	956	1041
Average (nm)	**678**	**1049**	**895**	**864**
SD (nm)	147	111	206	256

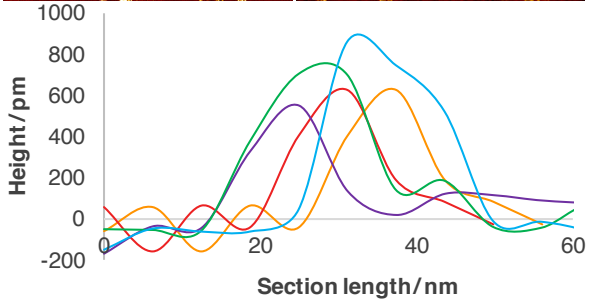
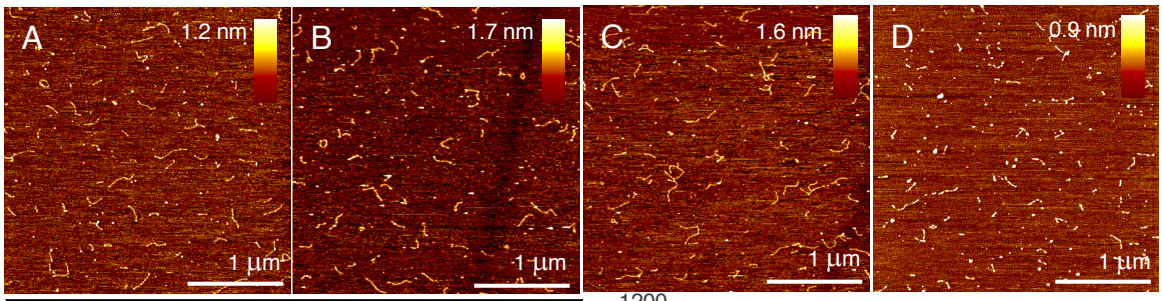


Fig. 2 AFM images (A-D) of 5-C_8-alkyne-dC-DNA, a table of the analysed lengths and a selection of perpendicular height scans.



	A	B	C	D
Length (nm)	644	671	818	691
	1238	485	938	953
	621	818	791	929
	1056	621	897	1041
	647	656	647	797
Average (nm)	**841**	**650**	**818**	**882**
SD (nm)	256	106	101	123

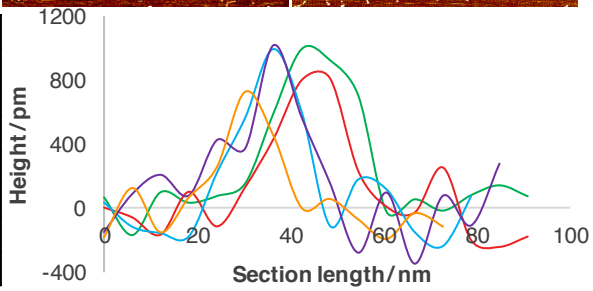
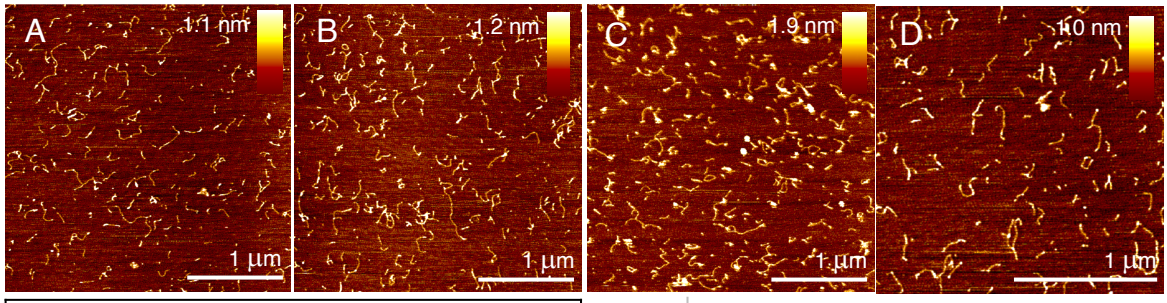


Fig. 3 AFM images (A-D) of 7-deaza-I-dA-DNA, a table of the analysed lengths and a selection of perpendicular height scans.



	A	B	C	D
Length (nm)	1441	1156	982	968
	1112	785	1679	641
	1271	918	965	965
	1121	815	1000	600
	794	624	718	606
Average (nm)	**1148**	**859**	**1069**	**756**
SD (nm)	214	176	322	172

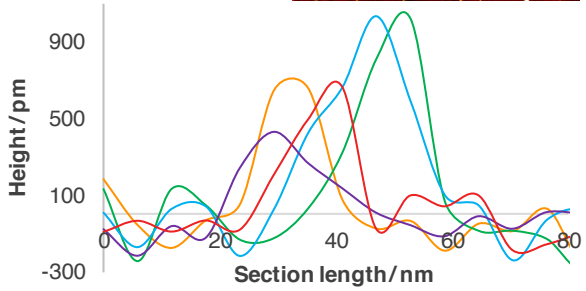
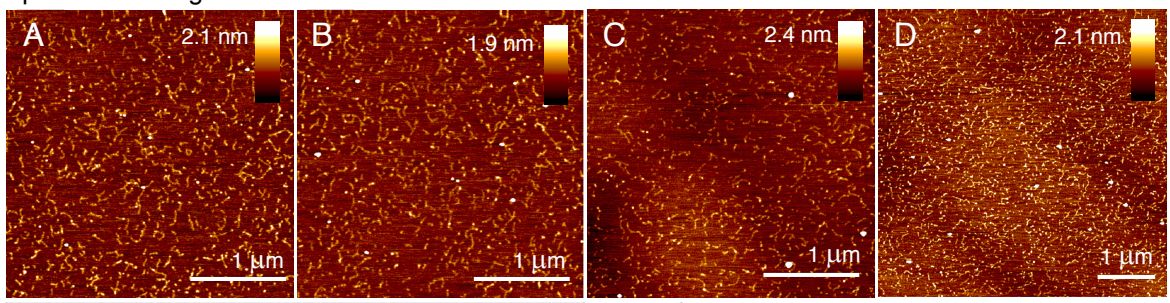


Fig. 4 AFM images (A-D) of 5-Br-dU-DNA, a table of the analysed lengths and a selection of perpendicular height scans.



	A	B	C	D
Length (nm)	521	259	426	285
	353	324	297	326
	362	297	312	315
	306	265	332	294
	274	309	300	362
Average (nm)	**363**	**291**	**334**	**316**
SD (nm)	85	25	48	27

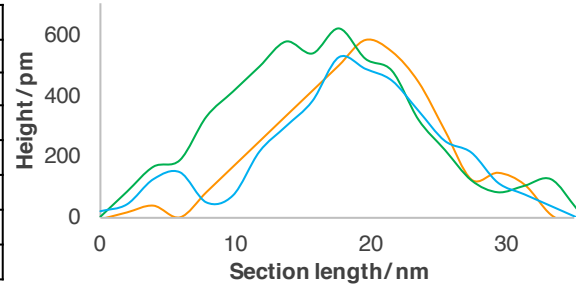


Fig. 5 AFM images (A-D) of 6-S-dG-DNA, a table of the analysed lengths and a selection of perpendicular height scans.

DNA sanger sequencing was performed (GATC Biotech) for each modified DNA product, Fig. 6-8.

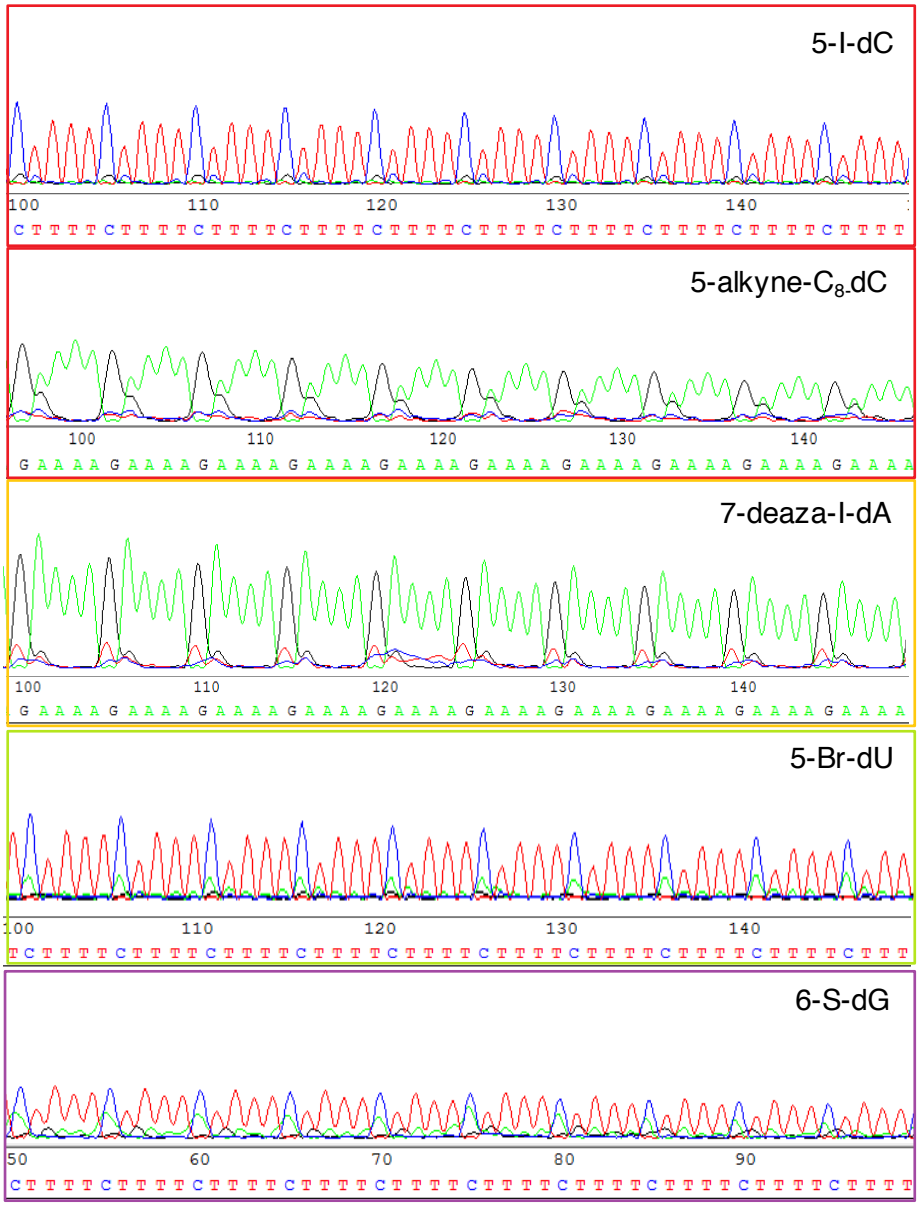


Fig. 6 DNA sanger sequencing of $[A_4G]_n/[T_4C]_n$ extended with a single type of modification per DNA extension pot.

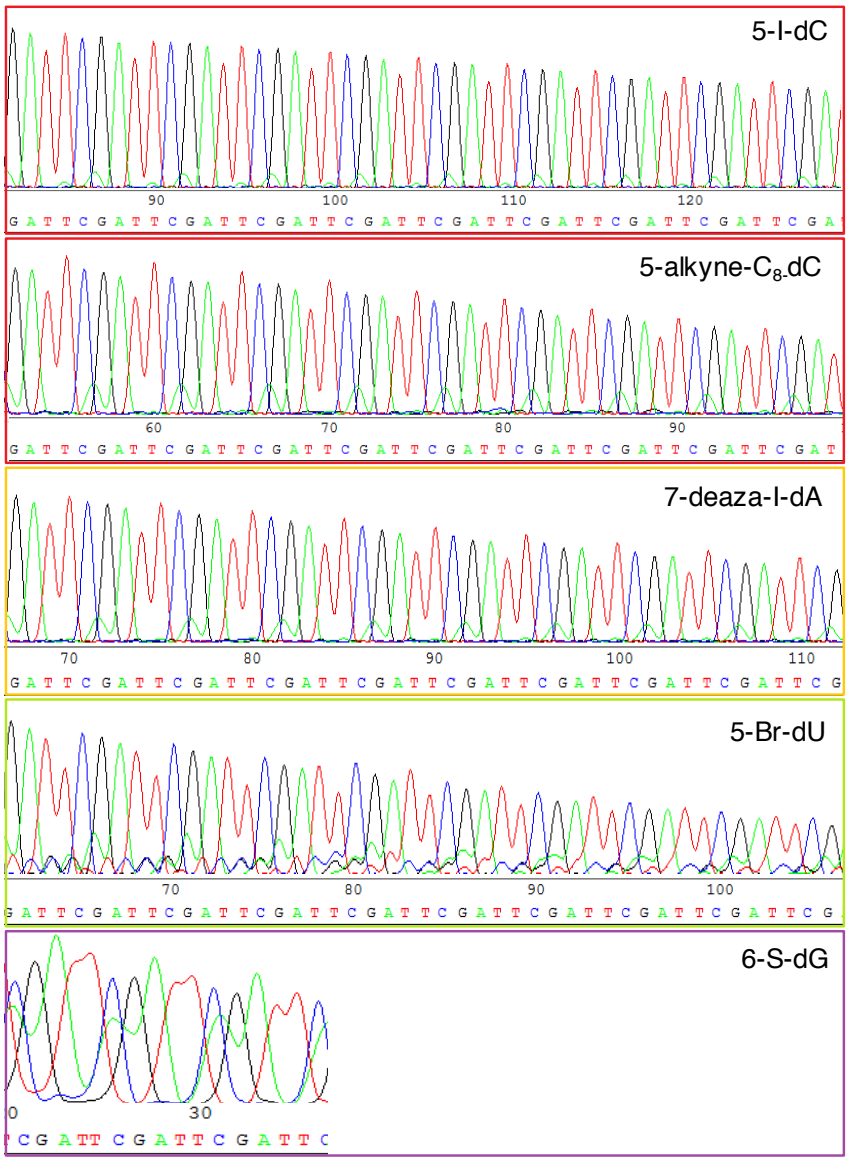


Fig. 7 DNA sanger sequencing of $[GAATC]_n/[GATTC]_n$ extended with a single type of modification per DNA extension pot.

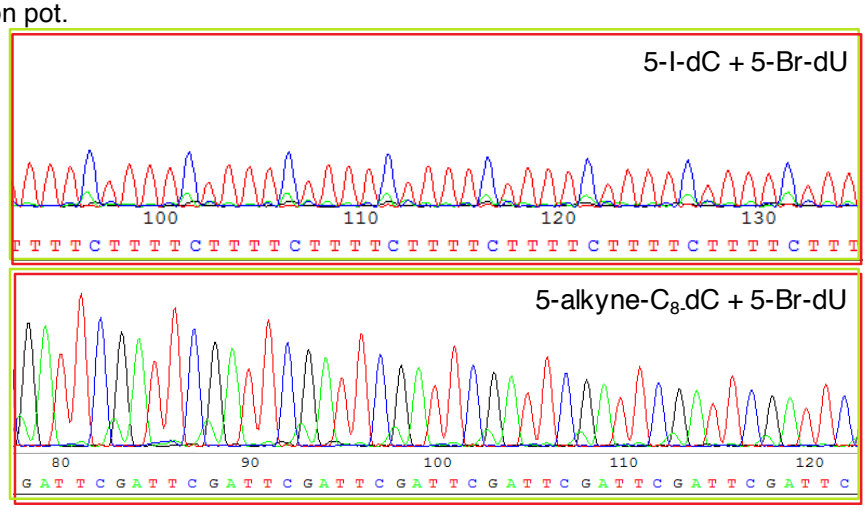


Fig. 8 DNA sanger sequencing of the two-type of modified DNA products.

The HPLC elution time and extinction coefficients of the dN standard for each unmodified and modified dNTP incorporated into dsDNA in Chapter 3 are shown in Table 1. The elusion times were determined from the HPLC traces in Fig. 9 and the extinction coefficients were calculated from the UV-Vis analysis of a known concentration.

dN	HPLC elution time / min	Extinction coefficient / $M^{-1}cm^{-1}$
dC	5.0	5180
dG	10.3	14190
T	10.8	7790
dA	21.3	9850
6-S-dG	13.8	3820
5-I-dC	17.9	4250
5-Br-dU	14.6	2700
7-I-dA	52.1	5630
5-alkyne-C_8-dC	54.5	7200*

Table 1 dN elution times and extinction coefficients. Extinction coefficients are shown for an absorbance at 254 nm. *Extinction coefficient at λmax(299 nm).

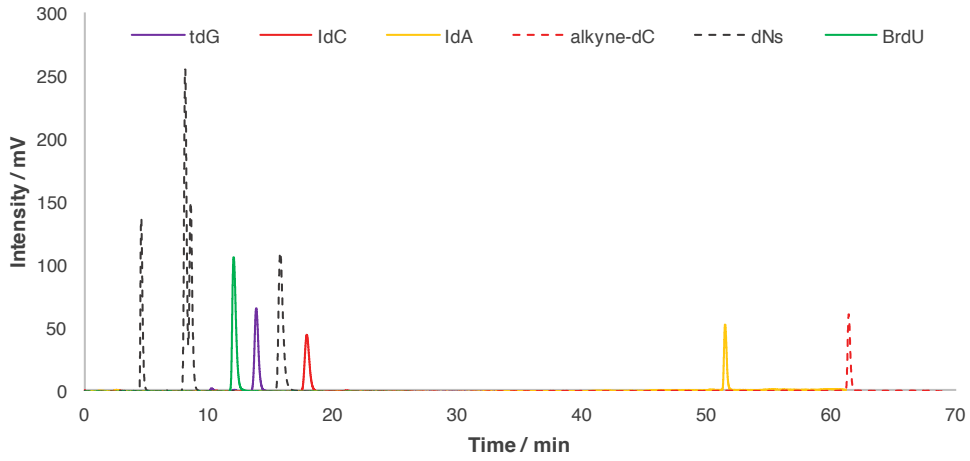


Fig. 9 HPLC traces of each dN standard.

References

1 Seela, F., Sirivolu, V. R. & Chittepu, P. Modification of DNA with octadiynyl side chains: synthesis, base pairing, and formation of fluorescent coumarin dye conjugates of four nucleobases by the alkyne--azide "click" reaction. *Bioconjug Chem* **19**, 211-224 (2008).

Appendix C

Static 6-S-dG-DNA titrations with Au^{+}, Table 1, Ni^{2+}, Table 2, Cd^{2+}, Table 3, Au^{3+}, Table 4 and the percentage binding plot used to determine the K_d for each metal ion, Fig. 1.

[S-dG] / μM	[Au^{+}] / μM	A^{348}	$1/A^{348}$	% Binding	[Au^{+}] / [S-dG]
65.4	0.00	0.062	16.1	0.00	0.000
65.4	1.00	0.061	16.4	2.98	0.015
65.4	1.90	0.061	16.4	2.98	0.029
65.4	4.21	0.060	16.7	6.06	0.064
65.4	6.29	0.060	16.7	6.06	0.096
65.4	10.7	0.059	16.9	9.24	0.16
65.4	19.6	0.057	17.5	15.9	0.30
65.4	27.6	0.054	18.5	26.9	0.42
65.4	39.9	0.052	19.2	35.0	0.61
65.4	50.9	0.049	20.4	48.2	0.78
65.4	60.8	0.047	21.3	58.0	0.93
65.4	69.7	0.045	22.2	68.7	1.1
65.4	77.7	0.046	21.7	63.2	1.2
65.4	85.0	0.043	23.3	80.3	1.3
65.4	101	0.044	22.7	74.4	1.6
65.4	116	0.044	22.7	74.4	1.8
65.4	140	0.044	22.7	74.4	2.1
65.4	176	0.042	23.8	86.6	2.7
65.4	208	0.041	24.4	93.1	3.2
65.4	237	0.040	25.0	100	3.6

Table 1 Static 6-S-dG-DNA binding to Au^{+}.

[S-dG] / μM	[Ni^{2+}] / μM	A^{348}	$1/A^{348}$	% Binding	[Ni^{2+}] / [S-dG]
66.4	0.00	0.063	15.9	0.00	0.000
66.4	1.00	0.061	16.4	4.67	0.015
66.4	1.90	0.060	16.7	7.12	0.029
66.4	4.21	0.059	16.9	9.65	0.063
66.4	6.29	0.059	16.9	9.65	0.095
66.4	10.7	0.059	16.9	9.65	0.16
66.4	19.6	0.054	18.5	23.7	0.30
66.4	27.6	0.054	18.5	23.7	0.42
66.4	39.9	0.051	19.6	33.5	0.60
66.4	50.9	0.047	21.3	48.4	0.77
66.4	60.8	0.045	22.2	56.9	0.92
66.4	69.7	0.043	23.3	66.2	1.0
66.4	77.7	0.044	22.7	61.5	1.2
66.4	85.0	0.043	23.3	66.2	1.3
66.4	101	0.041	24.4	76.4	1.5
66.4	116	0.041	24.4	76.4	1.8
66.4	140	0.039	25.6	87.6	2.1
66.4	176	0.038	26.3	93.6	2.6
66.4	208	0.038	26.3	93.6	3.1
66.4	237	0.037	27.0	100	3.6

Table 2 Static 6-S-dG-DNA binding to Ni^{2+}.

[S-dG] / μM	[Cd^{2+}] / μM	A^{348}	$1/A^{348}$	% Binding	[Cd^{2+}] / [S-dG]
68.5	0.00	0.065	15.4	0.00	0.000
68.5	1.00	0.063	15.9	1.62	0.015
68.5	1.90	0.061	16.4	3.35	0.028
68.5	4.21	0.058	17.2	6.17	0.061
68.5	6.29	0.052	19.2	12.8	0.092
68.5	10.7	0.044	22.7	24.4	0.16
68.5	19.6	0.036	27.8	41.2	0.29
68.5	27.6	0.029	34.5	63.5	0.40
68.5	39.9	0.026	38.5	76.7	0.58
68.5	50.9	0.025	40.0	81.9	0.74
68.5	60.8	0.024	41.7	87.4	0.89
68.5	69.7	0.023	43.5	93.4	1.0
68.5	77.7	0.25	40.0	81.9	1.1
68.5	85.0	0.024	41.7	87.4	1.2
68.5	101	0.023	43.5	93.4	1.5
68.5	116	0.022	45.5	100	1.7
68.5	140	0.024	41.7	87.4	2.0
68.5	176	0.024	41.7	87.4	2.6
68.5	208	0.025	40.0	81.9	3.0
68.5	237	0.023	43.5	93.4	3.5

Table 3 Static 6-S-dG-DNA binding to Cd^{2+}.

[S-dG] / μM	[Au^{3+}] / μM	A^{348}	$1/A^{348}$	% Binding	[Au^{3+}] / [S-dG]
68.5	0.00	0.065	15.4	0.00	0.000
68.5	1.00	0.064	15.6	2.00	0.015
68.5	1.90	0.063	15.9	4.07	0.028
68.5	4.21	0.060	16.7	10.7	0.061
68.5	6.29	0.058	17.2	15.5	0.092
68.5	10.7	0.053	18.9	29.0	0.16
68.5	19.6	0.046	21.7	52.9	0.29
68.5	27.6	0.046	22.0	54.9	0.40
68.5	39.9	0.041	24.4	75.0	0.58
68.5	50.9	0.039	25.6	85.4	0.74
68.5	60.8	0.038	26.7	93.9	0.89
68.5	69.7	0.038	26.7	93.9	1.0
68.5	77.7	0.038	26.3	91.0	1.1
68.5	85.0	0.038	26.3	91.0	1.2
68.5	101	0.038	26.3	91.0	1.5
68.5	116	0.037	27.4	100	1.7
68.5	140	0.037	27.0	96.9	2.0
68.5	176	0.037	27.0	96.9	2.6
68.5	208	0.037	27.0	96.9	3.0
68.5	237	0.037	27.0	96.9	3.5

Table 4 Static 6-S-dG-DNA binding to Au^{3+}.

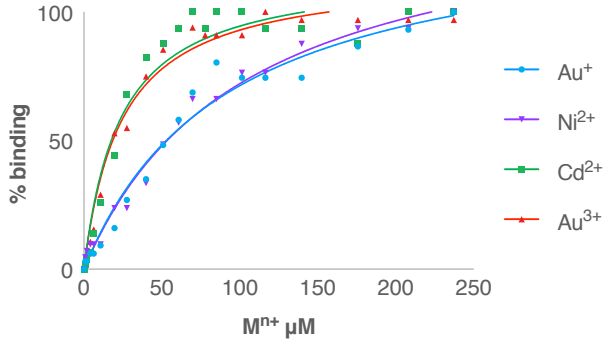


Fig. 1 Percentage binding plots of each metal ion with 6-S-dG-DNA from Tables 1-4.

AFM of DNA and 6-S-dG-DNA samples with and without the addition of Au^{+}, Ni^{2+}, Cd^{2+} and Au^{3+} on a mica substrate, Fig. 2-11 and Tables 5-14.

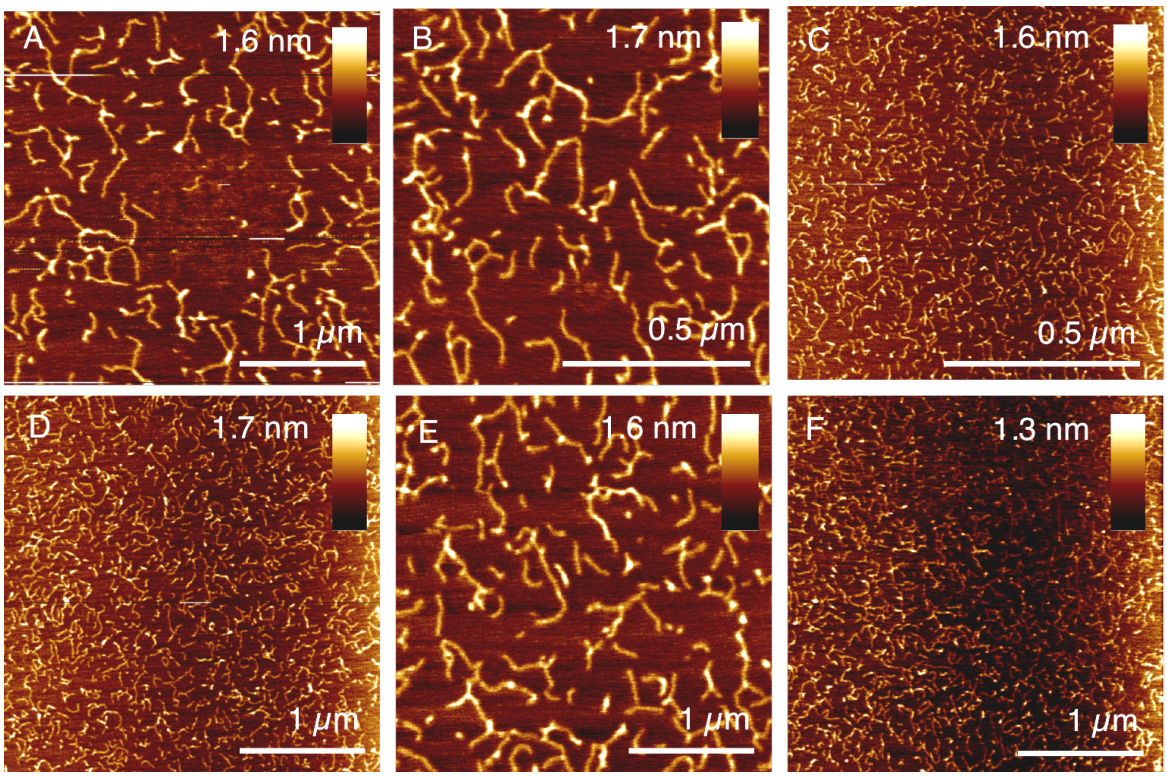


Fig. 2 AFM of $[A_4G]_n/[T_4C]_n$.

	A	B	C	D	E	F	
		0.5	0.6	0.6	0.6	0.8	0.6
		0.7	0.5	0.5	0.7	0.5	0.7
		0.6	0.5	0.6	1.1	0.5	0.6
		0.5	0.8	0.5	0.6	0.6	0.5
DNA		0.6	0.8	0.7	0.5	0.5	0.6
		0.8	0.8	0.8	0.7	0.7	0.6
		0.8	0.6	0.7	0.6	0.8	0.5
		0.6	0.5	0.6	0.7	0.6	0.9
		0.5	0.6	0.7	0.7	0.7	0.5
		0.7	0.5	0.6	0.5	0.6	0.6
Average		0.6	0.6	0.6	0.7	0.6	0.6
StDev		0.11	0.12	0.09	0.16	0.11	0.11

Table 5 Sample of DNA height scans from Fig. 2.

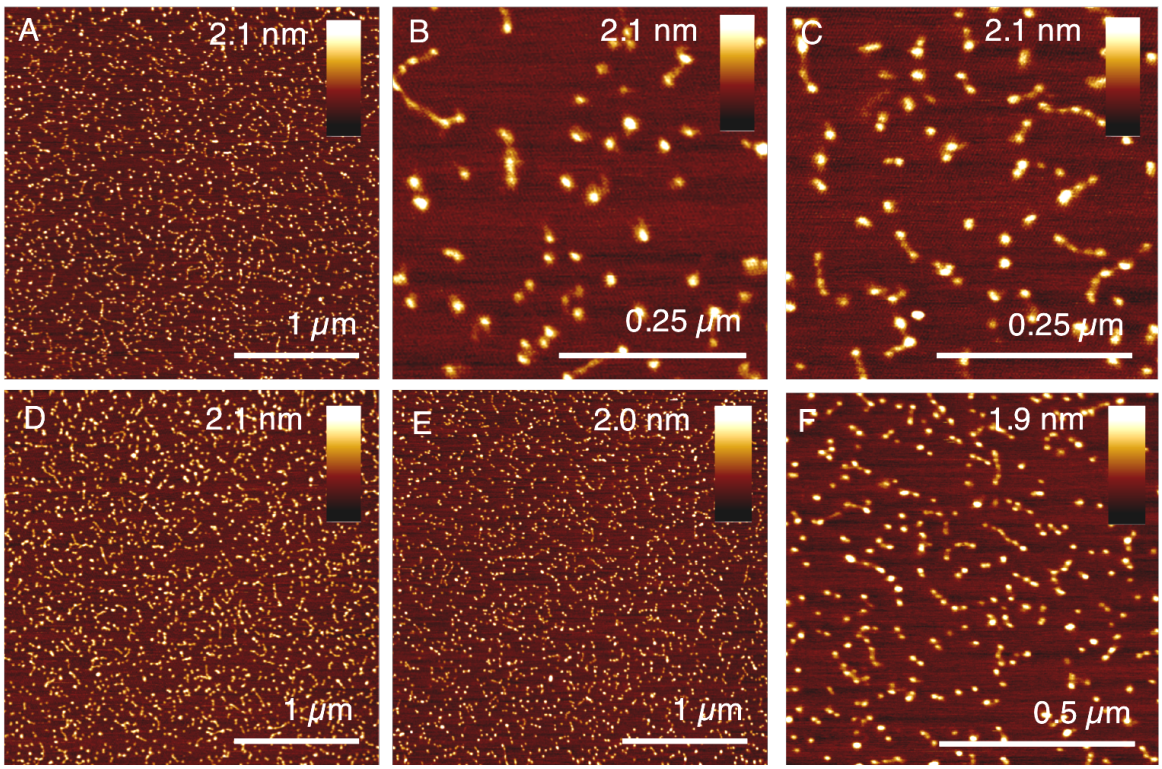


Fig. 3 AFM of $[A_4G]_n/[T_4C]_n$ with Au^+.

	A	B	C	D	E	F	
		0.6	0.9	0.5	0.6	0.8	0.7
		0.6	1.1	0.9	0.1	0.8	0.5
		0.8	1.1	0.8	0.9	1.1	0.9
		0.9	1.5	0.6	1.3	0.5	1.5
DNA Au^+		1.1	0.6	0.7	0.5	0.9	1.1
		0.9	1.1	1.2	0.7	0.9	1.0
		0.7	1.0	0.5	1.1	0.8	1.1
		1.0	0.8	0.8	1.4	0.5	1.1
		1.1	0.7	1.1	0.8	1.3	1.3
		1.0	1.3	0.6	0.8	1.8	1.0
Average		0.9	1.0	0.8	0.8	0.9	1.0
StDev		0.18	0.26	0.23	0.37	0.37	0.27

Table 6 Sample of DNA height scans from Fig. 3.

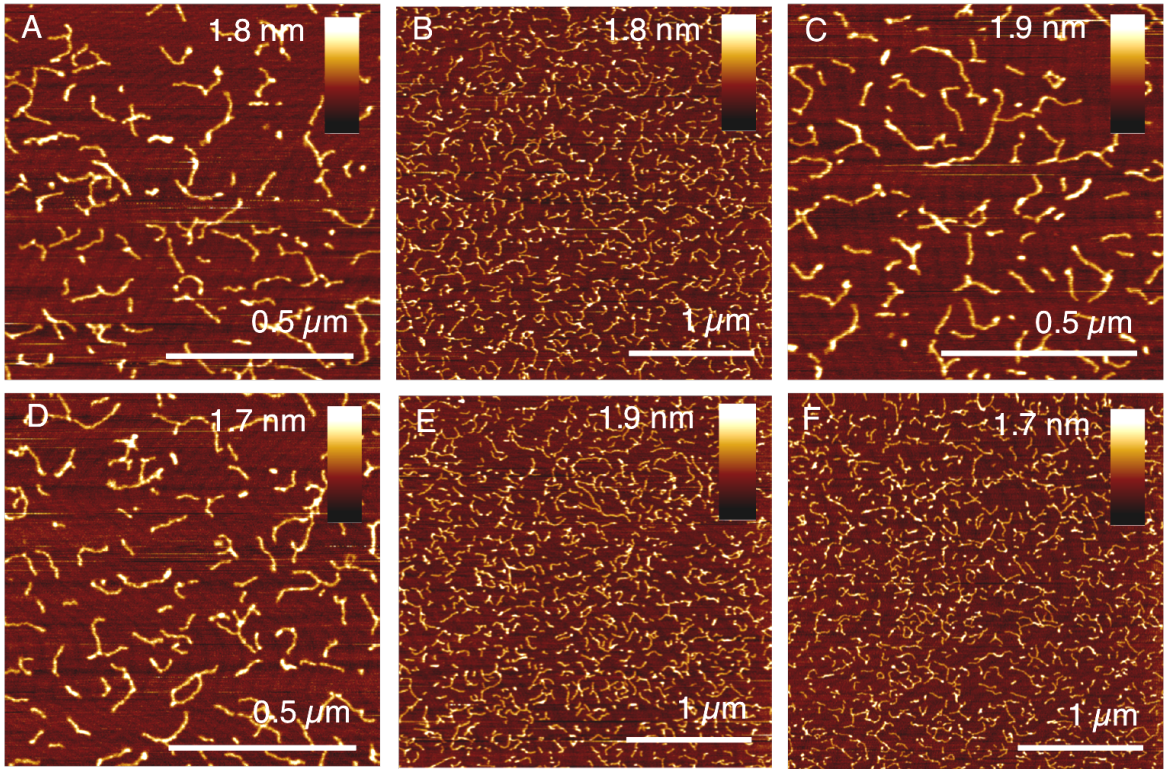


Fig. 4 AFM of $[A_4G]_n/[T_4C]_n$ with Ni^{2+}.

	A	B	C	D	E	F	
		0.6	0.6	0.6	0.5	0.6	0.5
		0.6	0.6	0.5	0.6	0.8	0.6
		0.5	1.1	0.7	0.8	0.6	0.7
		0.5	0.7	0.6	0.9	0.5	0.5
DNA Ni^{2+}		0.7	0.4	0.9	0.6	0.6	1.0
		0.7	0.8	0.7	0.7	0.6	0.6
		0.5	0.7	0.6	0.6	0.7	0.7
		0.8	0.5	0.8	0.7	0.5	0.5
		0.7	0.5	0.7	0.7	0.7	0.6
		0.7	0.6	0.6	0.6	0.6	0.8
Average		0.6	0.7	0.7	0.7	0.6	0.7
StDev		0.10	0.19	0.11	0.11	0.09	0.15

Table 7 Sample of DNA height scans from Fig. 4.

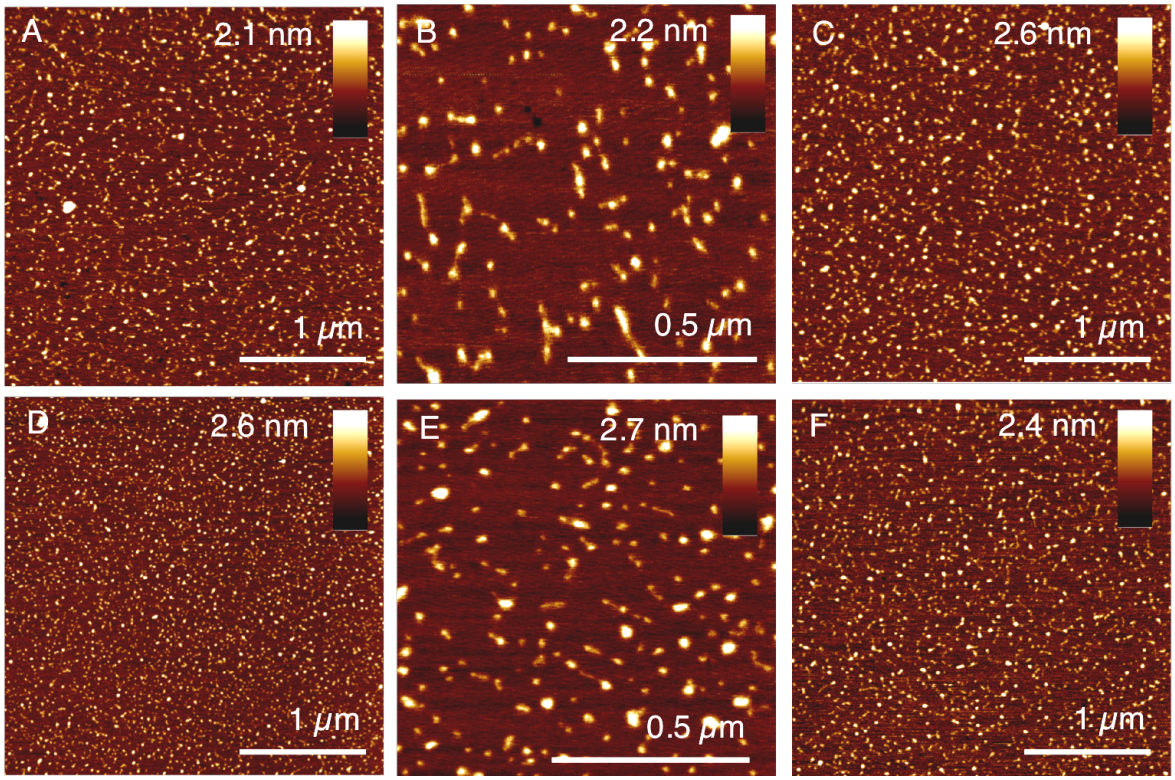


Fig. 5 AFM of $[A_4G]_n/[T_4C]_n$ with Cd^{2+}.

	A	B	C	D	E	F	
		0.5	1.0	1.5	0.8	0.6	0.7
		1.3	0.8	1.3	0.7	1.3	1.0
		1.5	1.1	1.0	0.9	1.6	1.3
		0.5	1.0	0.7	0.8	1.3	0.9
DNA Cd^{2+}		0.6	0.7	0.8	1.5	0.7	0.8
		0.8	0.6	0.5	1.3	1.5	1.6
		1.0	0.5	0.8	1.7	1.0	0.7
		0.7	1.3	1.0	0.8	0.5	0.6
		1.1	0.8	1.0	1.3	0.7	1.5
		1.3	1.3	1.6	0.5	1.1	1.0
Average		0.9	0.9	1.0	1.0	1.0	1.0
StDev		0.34	0.26	0.33	0.37	0.37	0.33

Table 8 Sample of DNA height scans from Fig. 5.

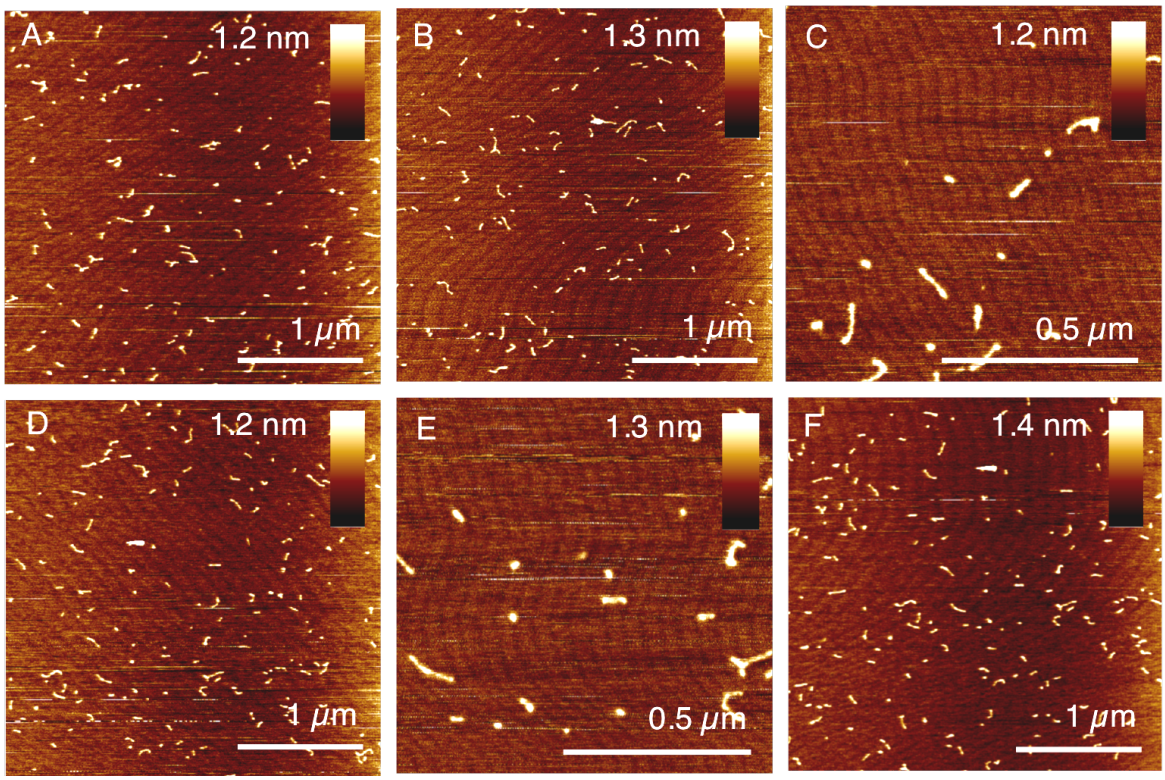


Fig. 6 AFM of $[A_4G]_n/[T_4C]_n$ with Au^{3+}.

	A	B	C	D	E	F	
		1.1	0.6	0.7	0.7	1.1	1.0
		0.9	0.6	1.2	0.8	1.0	0.6
		0.9	0.8	0.7	0.7	0.9	0.7
		0.7	0.7	0.8	0.7	1.2	0.9
DNA Au^{3+}		0.8	0.8	1.0	0.8	1.0	0.8
		0.5	0.6	0.8	0.7	1.1	0.6
		1.4	0.7	1.0	0.6	1.3	0.1
		0.9	0.6	0.8	0.8	1.2	0.8
		0.7	0.9	0.9	0.6	0.9	0.9
		0.9	0.7	1.0	0.6	0.7	0.8
Average		0.9	0.7	0.9	0.7	1.0	0.7
StDev		0.23	0.10	0.15	0.08	0.17	0.25

Table 9 Sample of DNA height scans from Fig. 6.

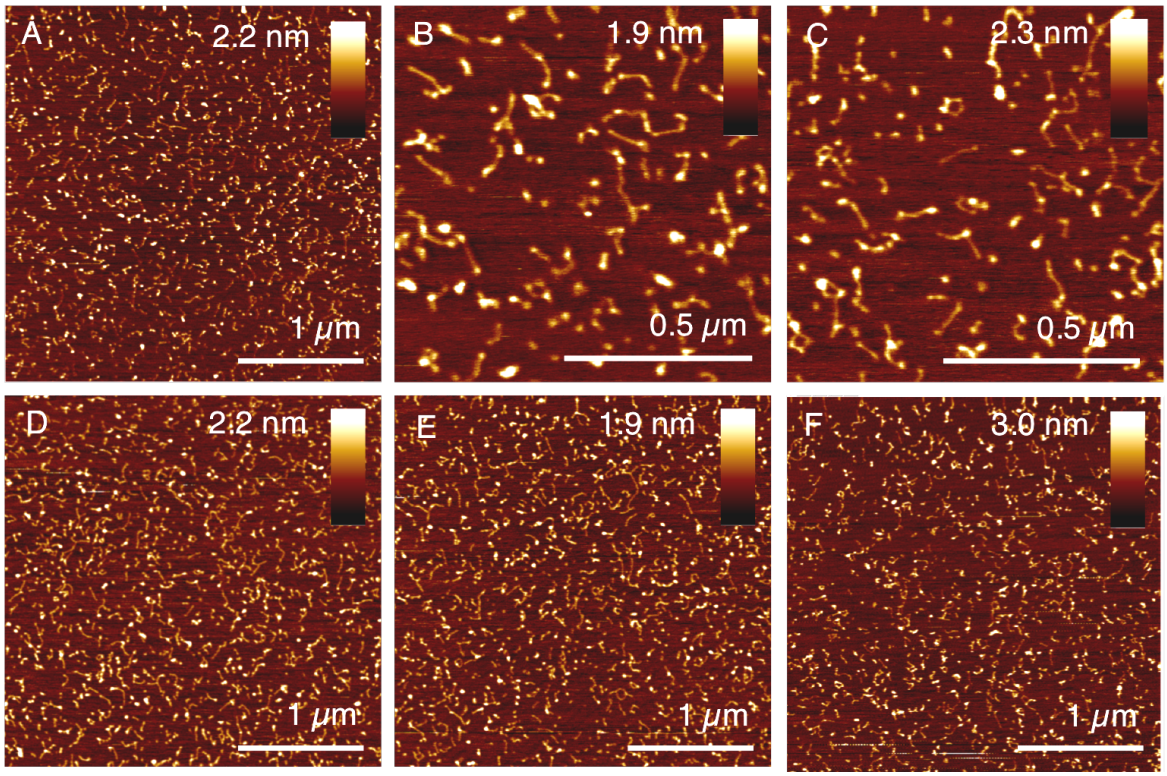


Fig. 7 AFM of $[A_4{}^{6S}G]_n/[T_4C]_n$.

	A	B	C	D	E	F	
		0.6	0.5	0.5	0.7	0.5	0.6
		0.5	0.6	0.6	0.6	0.7	0.5
		0.8	0.8	0.8	1.0	0.5	0.6
		0.5	0.7	0.7	0.9	0.6	0.7
6-S-dG-DNA		0.6	0.6	0.5	0.5	0.9	0.6
		0.8	0.7	0.7	0.7	0.6	0.6
		0.5	0.8	0.6	0.5	1.0	0.6
		0.5	1.0	0.5	0.5	0.7	0.7
		0.9	0.7	0.7	0.6	0.6	0.8
		0.7	0.5	0.8	0.7	0.5	0.6
Average		0.6	0.7	0.6	0.7	0.7	0.6
StDev		0.14	0.14	0.11	0.16	0.16	0.08

Table 10 Sample of DNA height scans from Fig. 7.

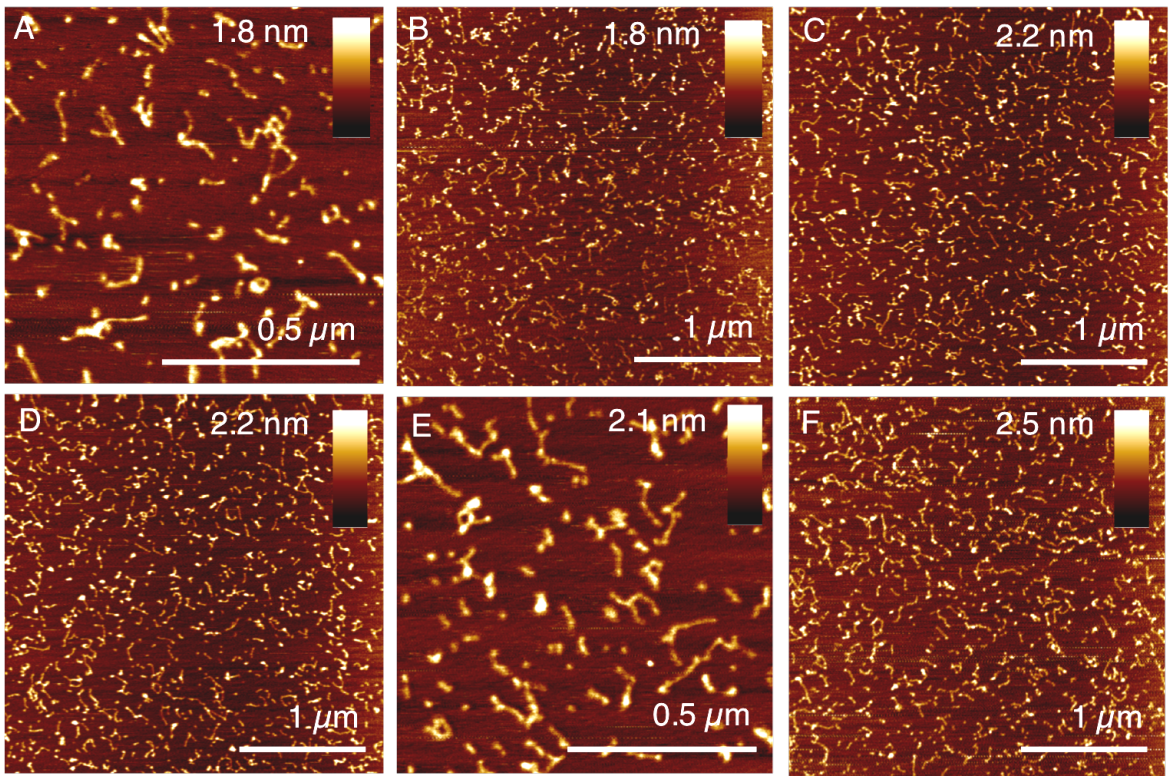


Fig. 8 AFM of $[A_4{}^{6S}G]_n/[T_4C]_n$ with Au^+.

	A	B	C	D	E	F	
		1.1	0.8	1.3	0.7	1.6	1.1
		0.8	0.8	0.5	1.1	1.5	1.0
		0.8	0.9	0.8	0.9	1.0	0.9
6-S-dG-DNA Au⁺		1.2	1.1	1.0	0.9	1.1	0.5
		1.1	1.2	1.5	1.5	1.3	1.5
		1.5	1.6	1.2	1.1	0.8	0.6
		1.6	1.1	1.5	0.6	1.0	0.5
		0.7	1.0	1.6	1.0	1.1	1.0
		1.0	0.9	1.0	0.8	0.9	0.8
		1.4	0.6	1.0	1.6	0.8	1.1
Average		1.1	1.0	1.1	1.0	1.1	0.9
StDev		0.29	0.26	0.33	0.31	0.26	0.30

Table 11 Sample of DNA height scans from Fig. 8.

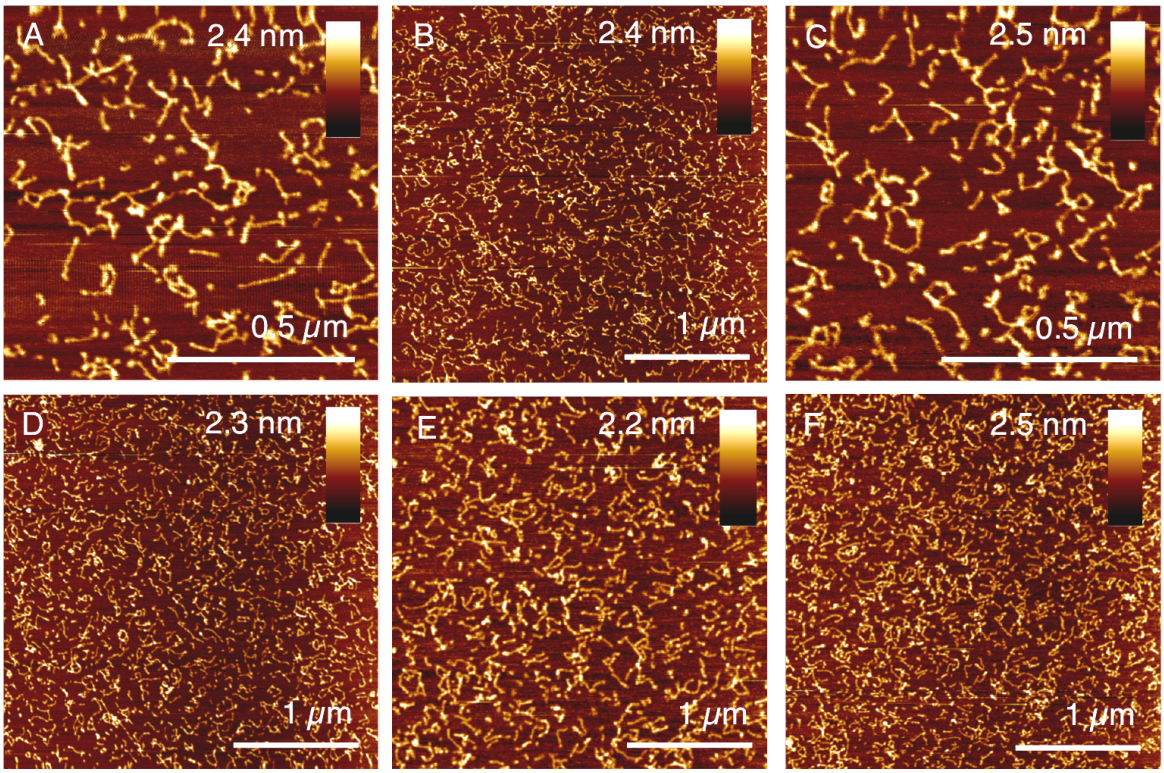


Fig. 9 AFM of $[A_4{}^{6S}G]_n/[T_4C]_n$ with Ni^{2+}.

	A	B	C	D	E	F	
		1.0	1.2	1.0	1.3	1.2	1.2
		1.1	1.3	0.9	1.1	1.5	1.0
		1.0	1.1	1.3	1.0	1.3	0.9
6-S-dG-DNA Ni^{2+}		1.8	1.0	1.4	0.9	1.3	1.0
		1.3	1.3	1.3	1.6	0.9	1.3
		1.1	0.9	1.1	1.2	1.4	1.0
		1.3	0.7	1.0	0.9	1.5	0.9
		1.1	1.3	1.0	1.5	1.0	1.1
		1.2	0.9	1.5	1.4	1.0	0.8
		1.8	1.0	1.4	1.0	1.2	0.9
Average		1.3	1.1	1.2	1.2	1.2	1.0
StDev		0.28	0.20	0.20	0.24	0.20	0.14

Table 12 Sample of DNA height scans from Fig. 9.

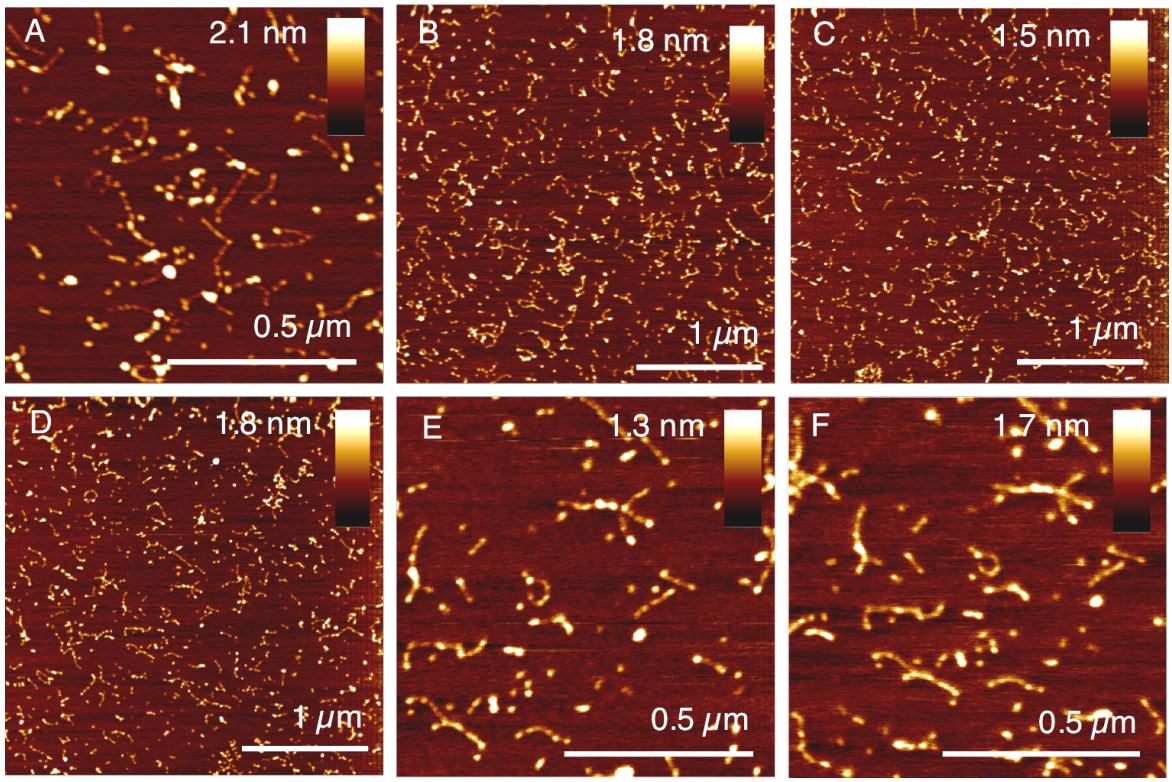


Fig. 10 AFM of $[A_4{}^{6S}G]_n/[T_4C]_n$ with Cd^{2+}.

	A	B	C	D	E	F	
		0.7	1.0	0.8	0.6	0.6	0.7
		0.7	0.6	0.6	1.5	0.7	1.0
		0.6	1.5	0.8	0.9	0.6	0.8
6-S-dG-DNA Cd^{2+}		0.5	1.4	0.9	0.5	0.7	0.5
		1.3	1.5	0.8	0.9	0.8	0.9
		0.5	0.9	0.7	0.5	0.5	1.0
		0.7	0.6	0.9	1.0	0.5	0.8
		0.6	0.8	1.1	0.8	0.6	0.6
		0.8	0.9	0.7	0.7	0.8	0.8
		0.7	0.7	1.2	0.7	0.7	0.9
Average		0.7	1.0	0.9	0.8	0.7	0.8
StDev		0.22	0.34	0.17	0.28	0.10	0.15

Table 13 Sample of DNA height scans from Fig. 10.

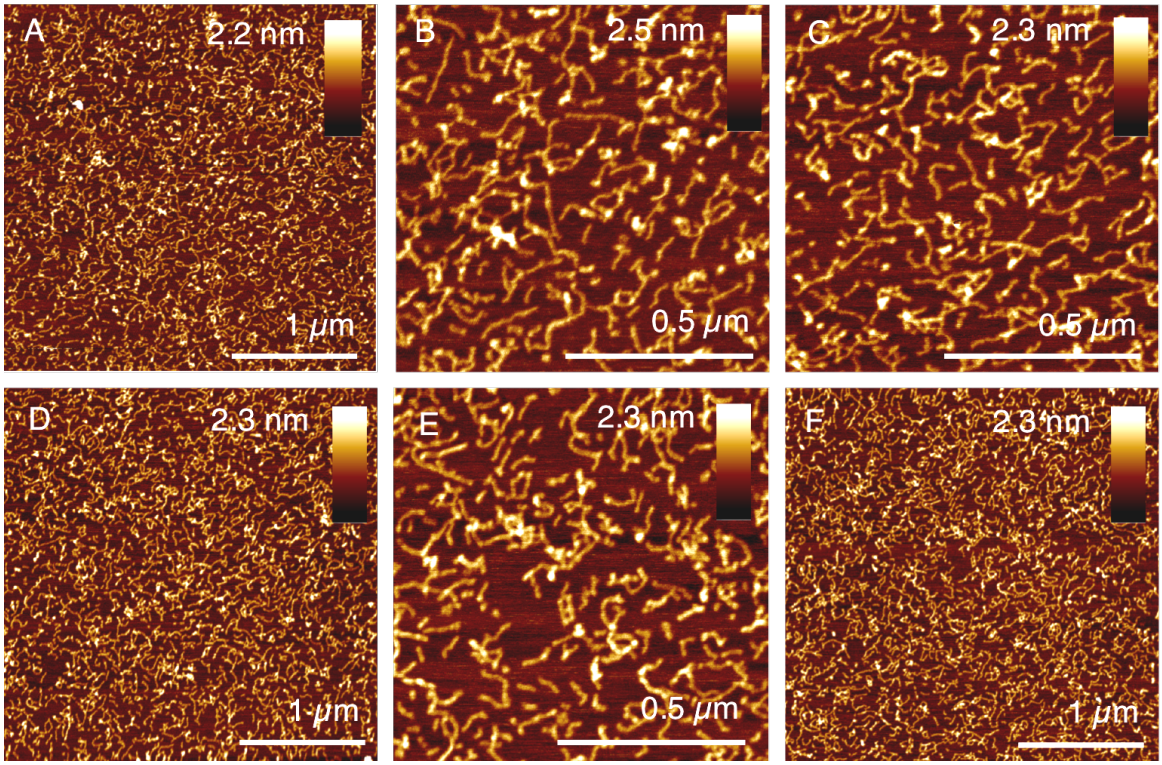


Fig. 11 AFM of $[A_4{}^{6S}G]_n/[T_4C]_n$ with Au^{3+}.

	A	B	C	D	E	F	
		0.5	1.0	1.1	0.9	0.6	0.7
		0.6	0.6	0.9	0.8	1.1	0.6
		0.6	0.6	1.1	1.1	0.7	0.9
6-S-dG-DNA Au^{3+}		0.9	0.7	0.8	1.3	0.7	1.0
		1.0	0.6	0.6	0.6	0.6	1.1
		0.9	1.1	0.7	0.7	1.1	0.8
		0.6	0.9	1.2	1.1	0.9	0.7
		1.1	0.8	0.9	0.8	1.0	0.8
		1.0	1.3	1.6	1.0	0.7	1.1
		1.3	0.8	0.6	0.8	0.6	0.8
Average		0.9	0.8	1.0	0.9	0.8	0.9
StDev		0.25	0.22	0.29	0.20	0.19	0.16

Table 14 Sample of DNA height scans from Fig. 11.

AFM of α-S-[GATC]$_n$/[CTAG]$_n$ DNA where each phosphorous bears a thio-group, Fig. 12, followed by the templating of 3 nm AuNP, Fig. 13. A comparison of unmodified DNA templating with 3 nm AuNP is shown Fig. 14.

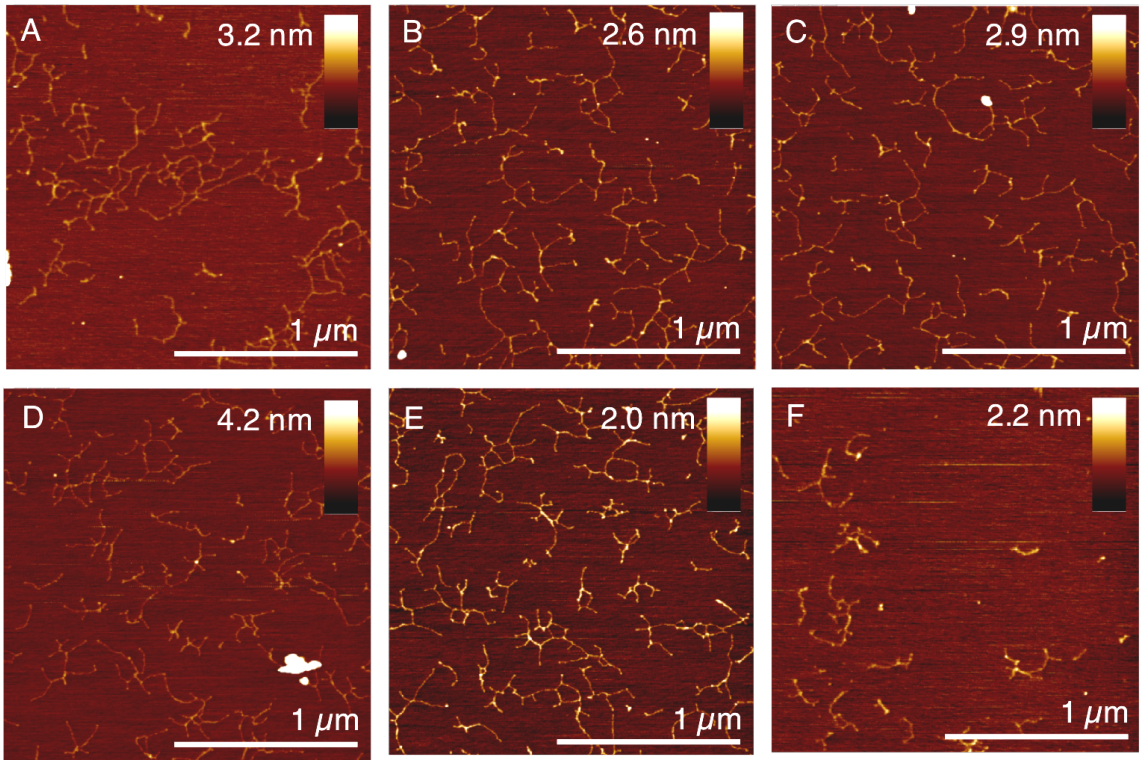


Fig. 12 AFM of α-S-[GATC]$_n$/[CTAG]$_n$.

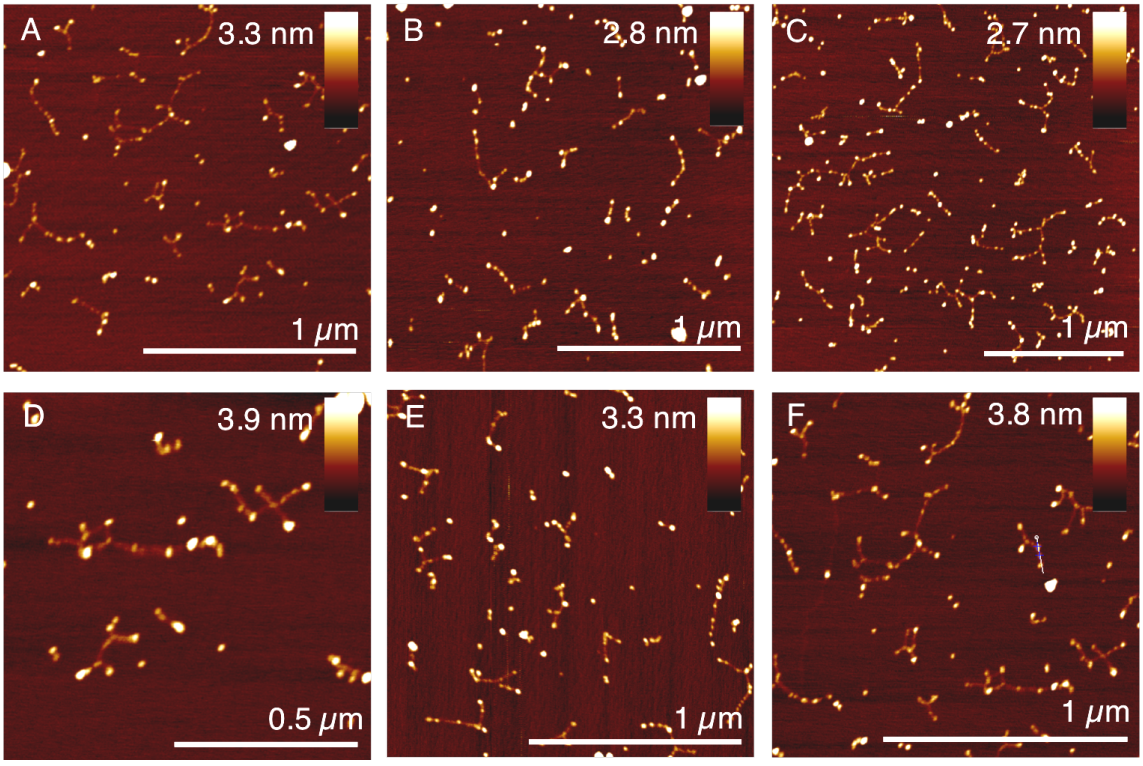


Fig. 13 AFM of α-S-[GATC]$_n$/[CTAG]$_n$ after templating with 3 nm AuNPs.

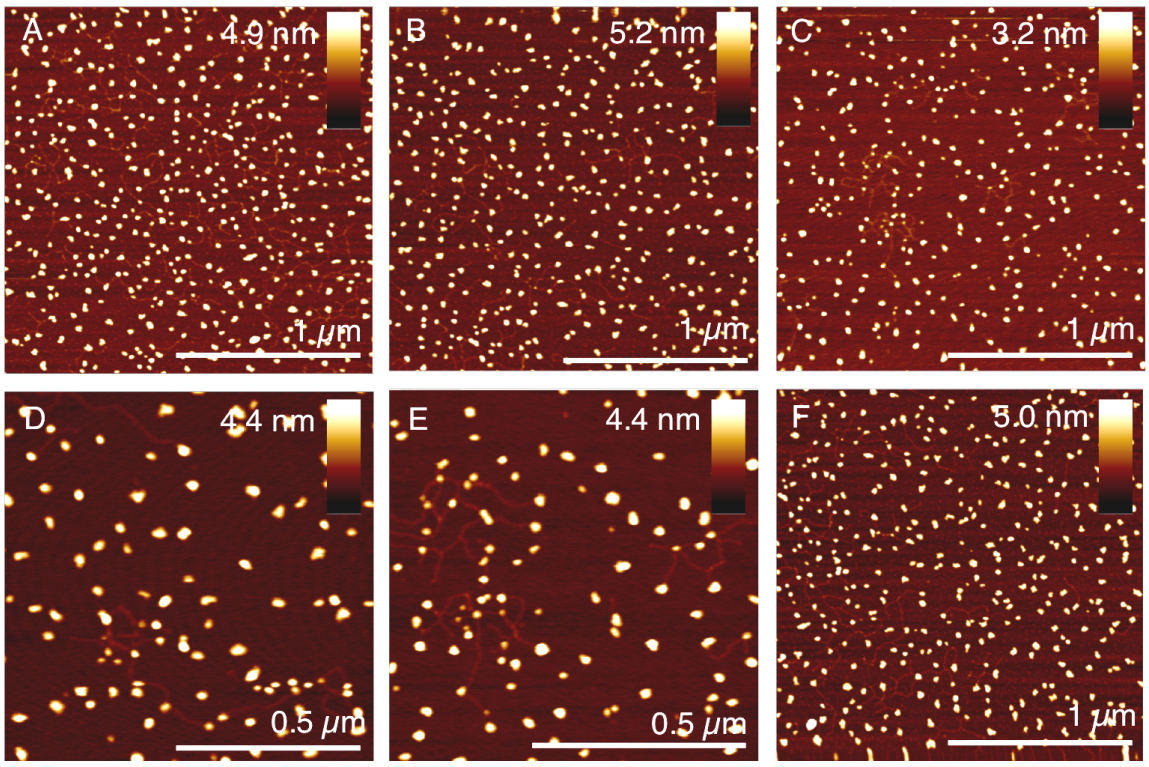


Fig. 14 AFM of $[GATC]_n/[CTAG]_n$ after templating with 3 nm AuNPs.

DNA heat-cool extension was performed with $[(AT)_4G_2]_n/[(TA)_4C_2]_n$ in the presence of α-S-dCTP and analysed by AFM, Fig. 15. Consecutive incubations were performed with the deposited sample and 10 mL of 3 nm AuNPs and analysed between each wash step, Fig. 16-19.

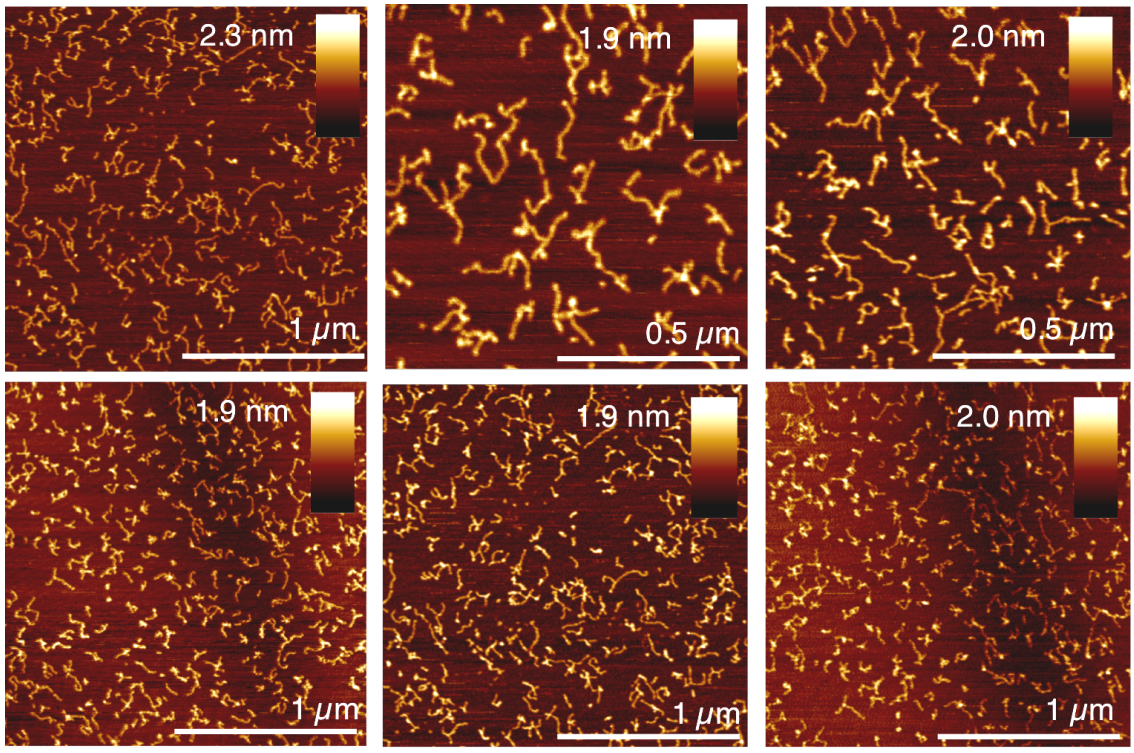


Fig. 15 AFM of $[(AT)_4G_2]_n/[(TA)_4{}^{\alpha S}C_2]_n$.

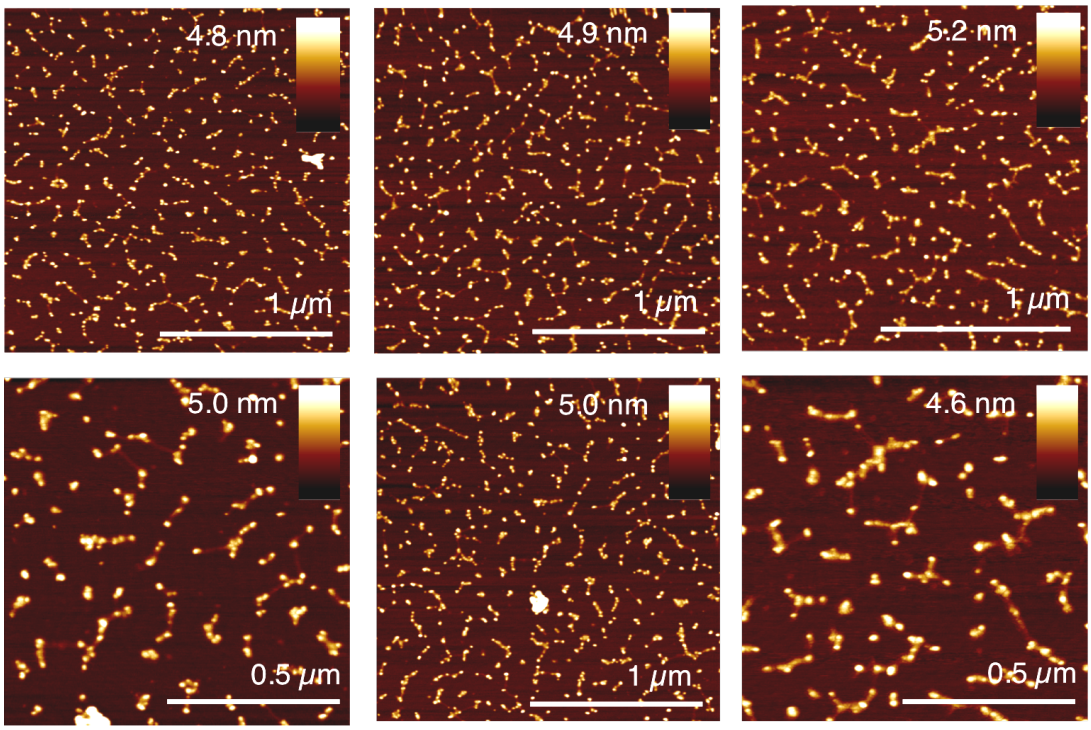


Fig. 16 AFM of $[(AT)_4G_2]_n/[(TA)_4{}^{\alpha S}C_2]_n$ after 10 minutes of incubation with 3 nm AuNPs.

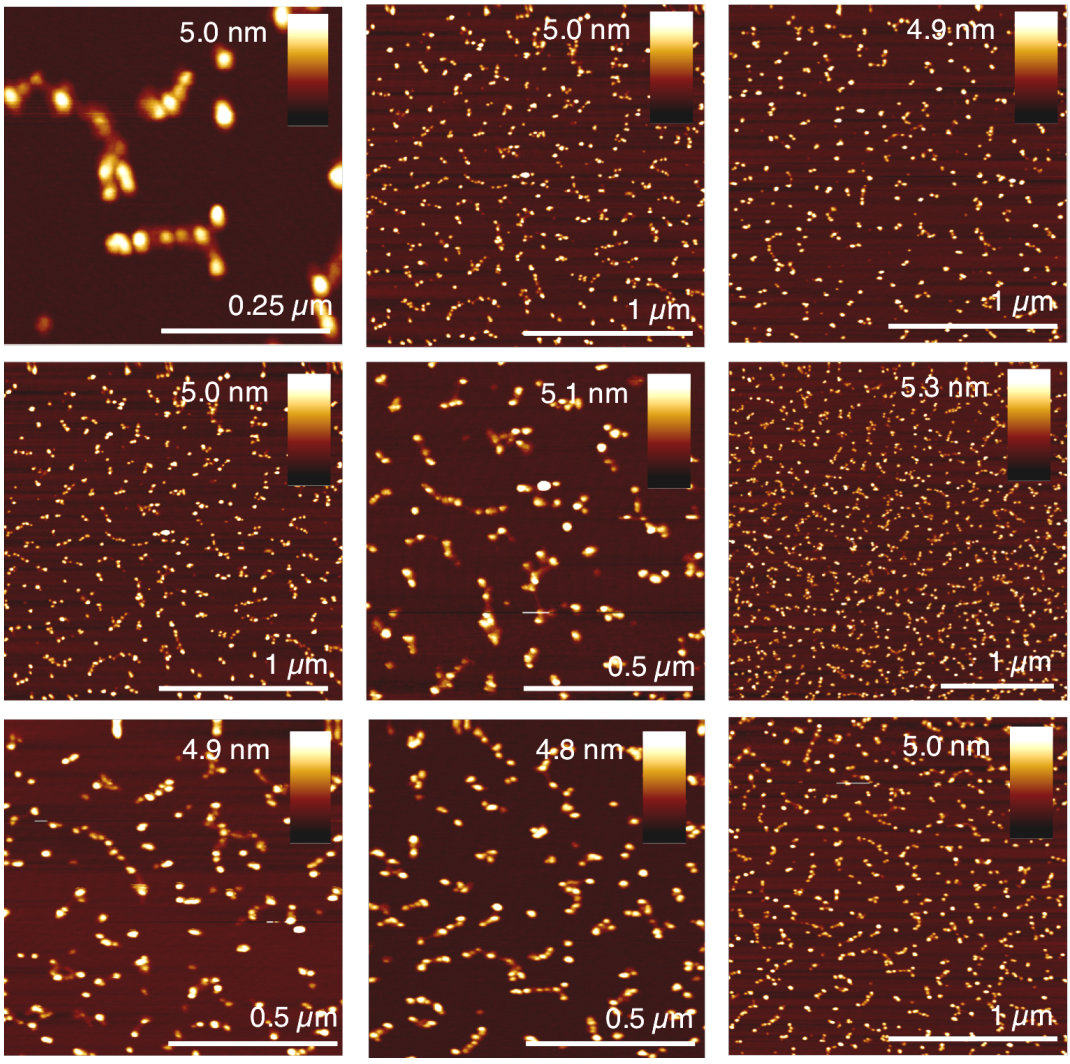


Fig. 17 AFM of $[(AT)_4G_2]_n/[(TA)_4{}^{\alpha S}C_2]_n$ after an additional 10 minutes of incubation with 3 nm AuNPs.

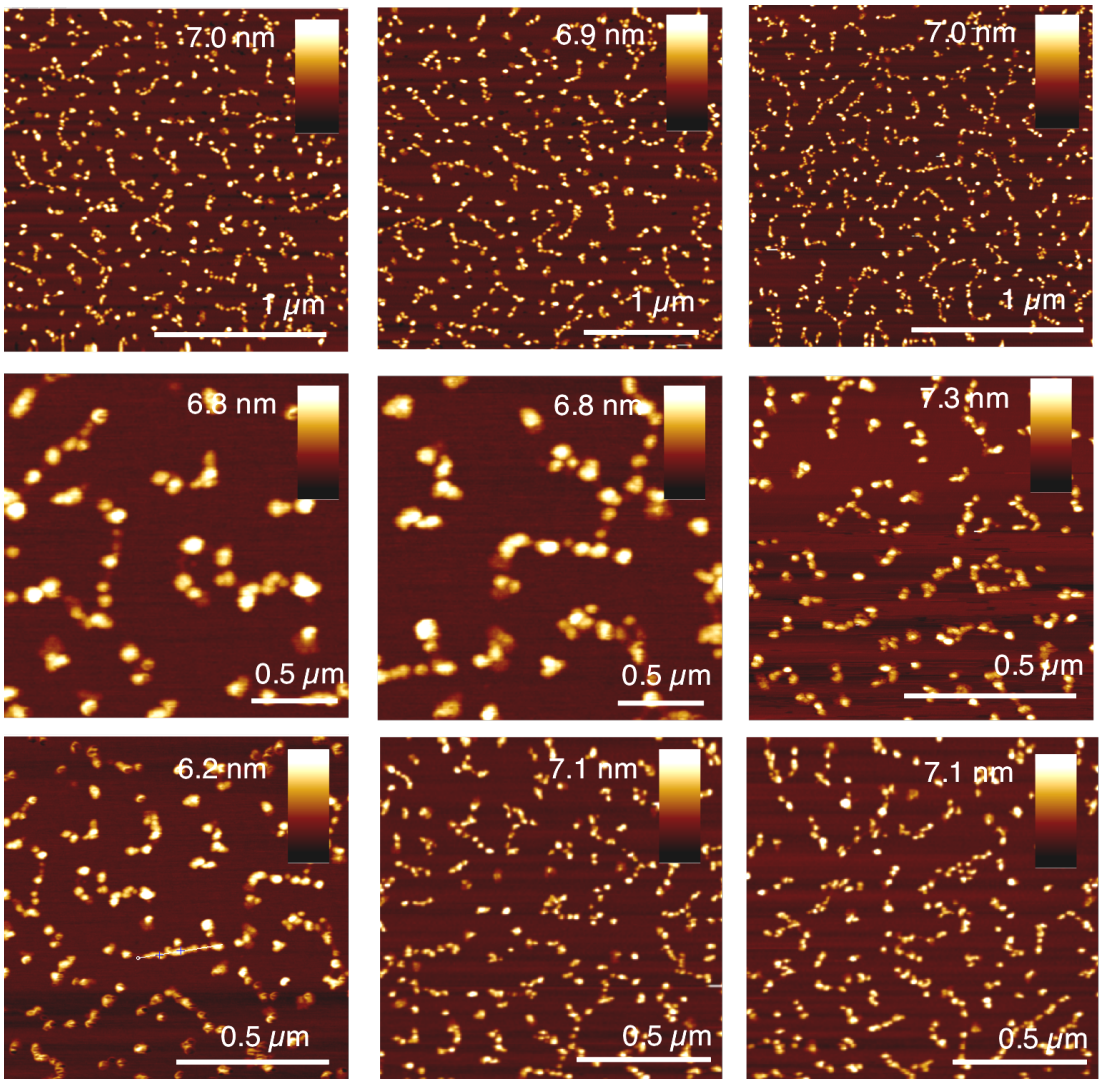


Fig. 18 AFM of $[(AT)_4G_2]_n/[(TA)_4{}^{\alpha S}C_2]_n$ after an additional 10 minutes of incubation with 3 nm AuNPs.

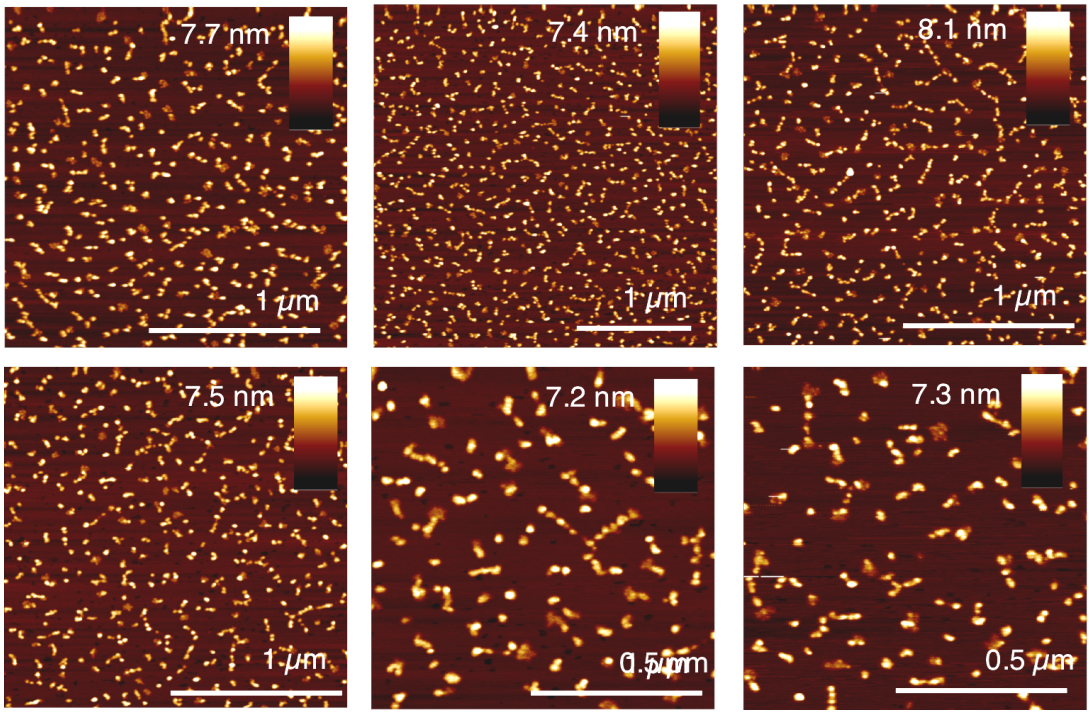


Fig. 19 AFM of $[(AT)_4G_2]_n/[(TA)_4{}^{\alpha S}C_2]_n$ after an additional 10 minutes of incubation with 3 nm AuNPs.

Appendix D

Fluorescence microscopy images of alkyne modified (C* = 5-C_8-alkyne-dC) and unmodified (C) DNA after the click reaction with Azide-fluor 545, Fig. 1-6.

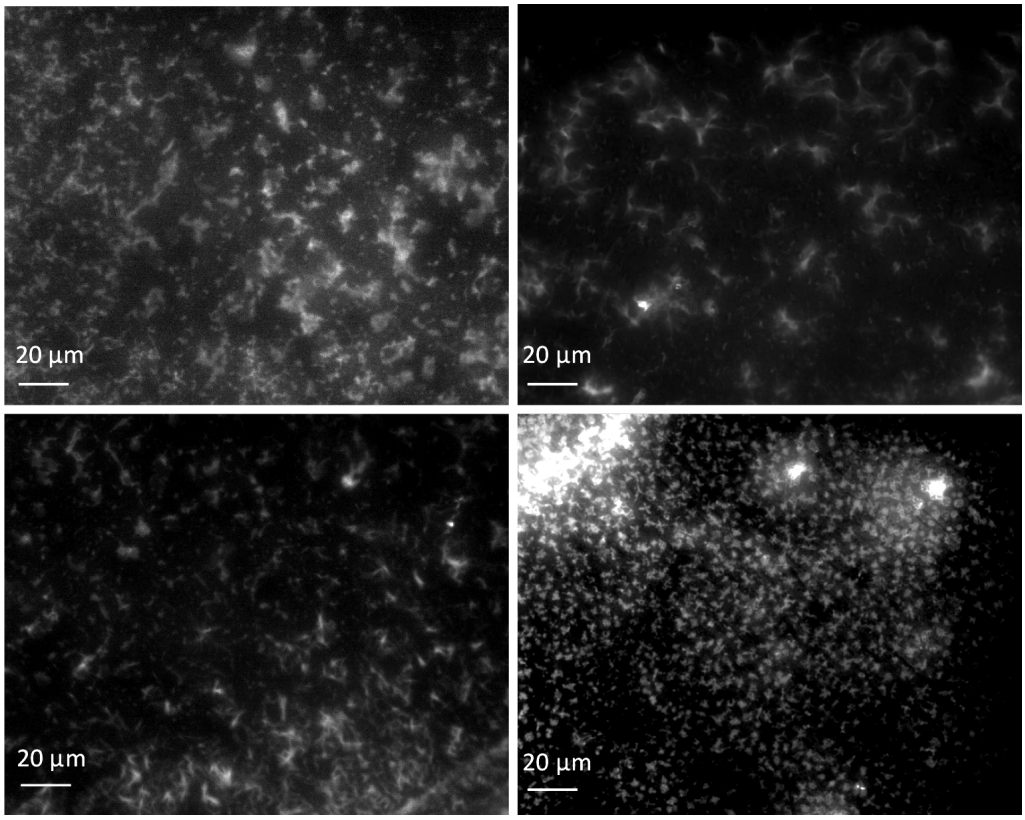


Fig. 1 Fluorescence microscopy of $[AG]_n/[TC^*]_n$ after the click reaction with Azide-fluor 545.

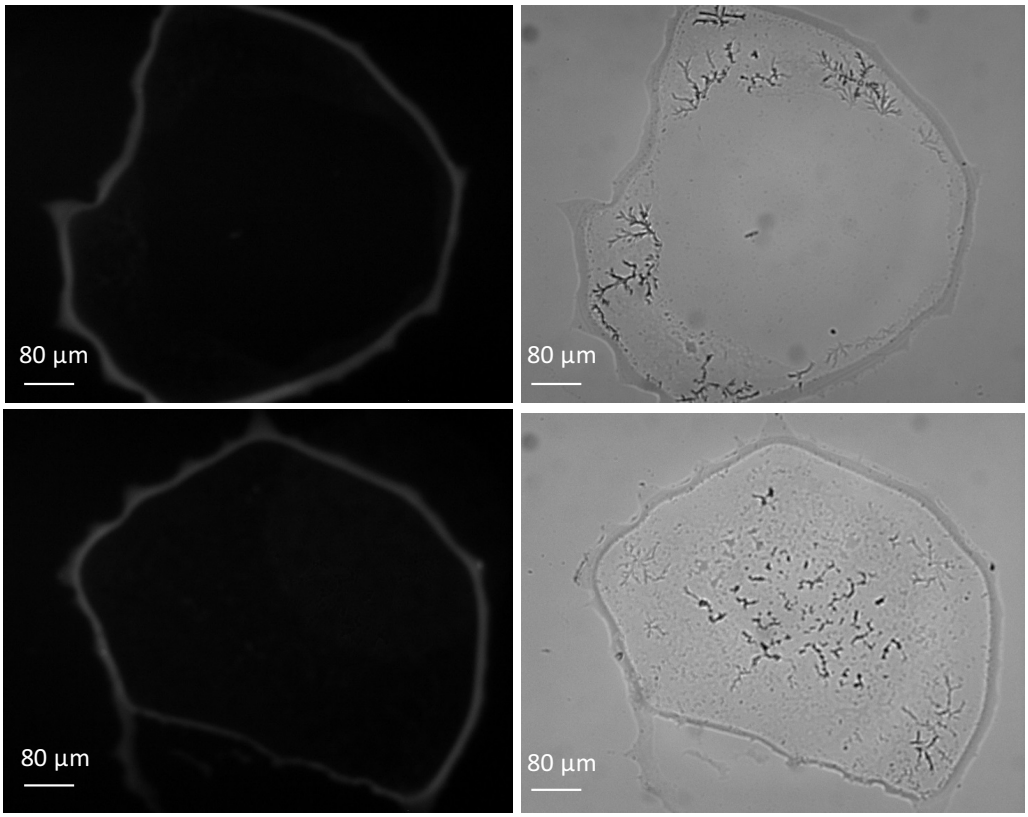


Fig. 2 Fluorescence microscopy of $[AG]_n/[TC]_n$ after the click reaction with Azide-fluor 545.

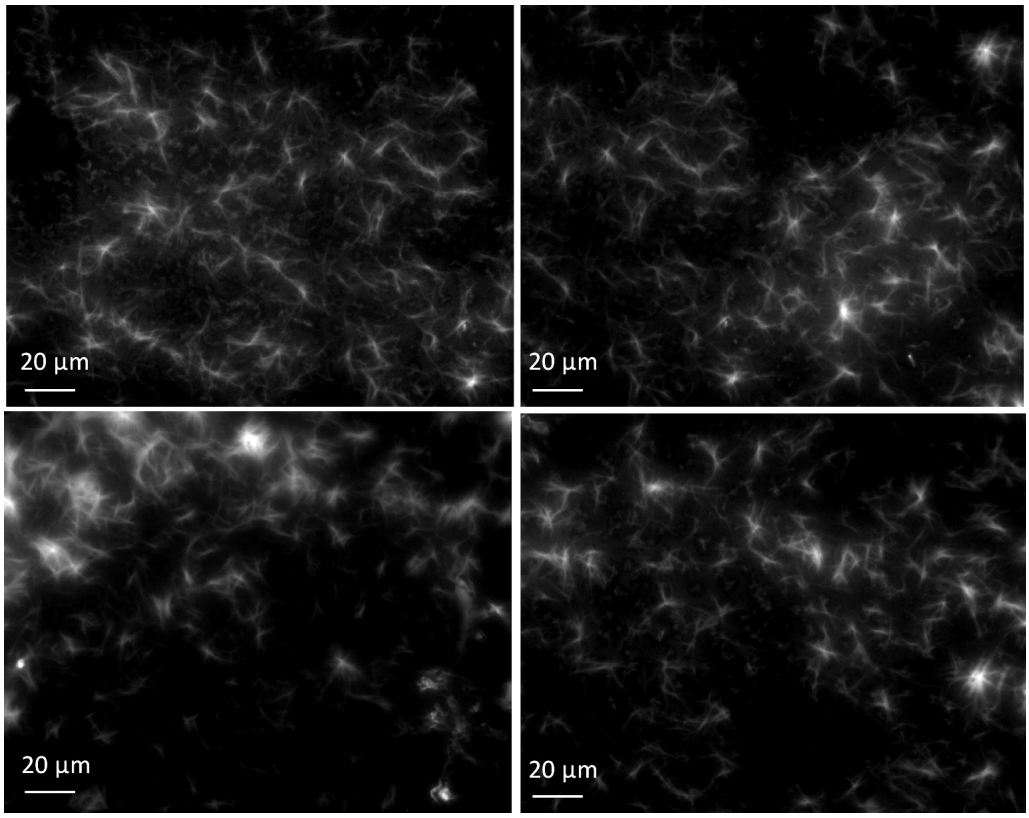


Fig. 3 Fluorescence microscopy of $[A_4G]_n/[T_4C^*]_n$ after the click reaction with Azide-fluor 545.

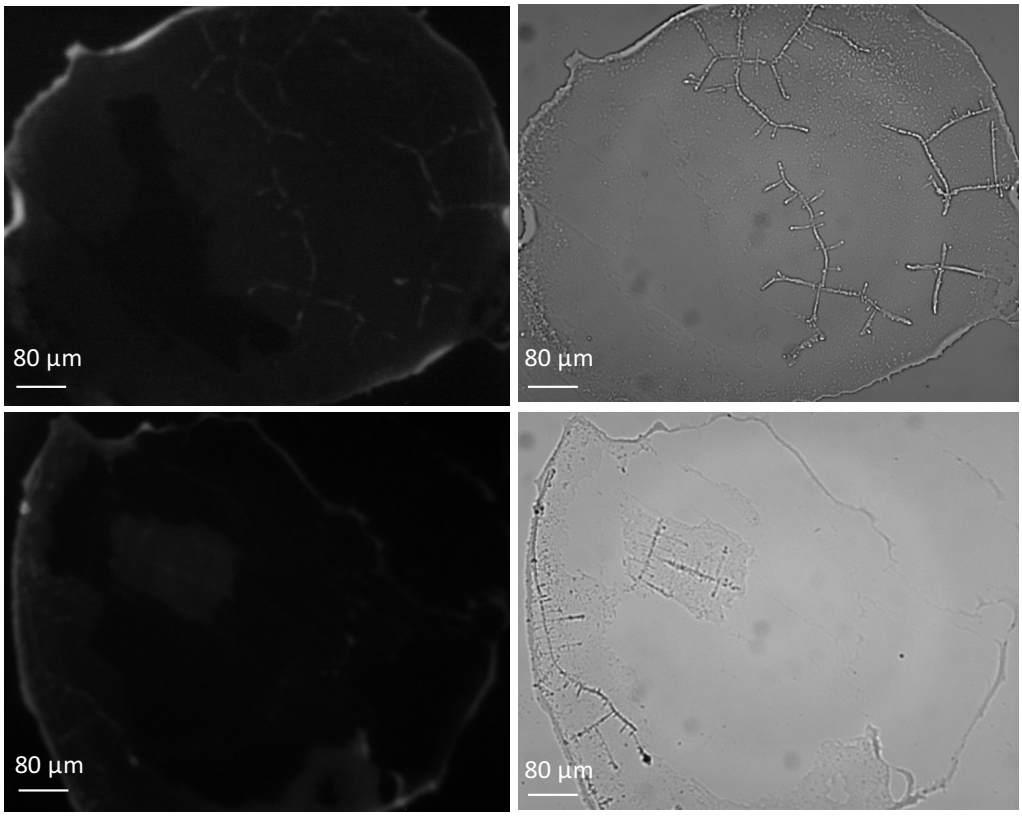


Fig. 4 Fluorescence microscopy of $[A_4G]_n/[T_4C]_n$ after the click reaction with Azide-fluor 545.

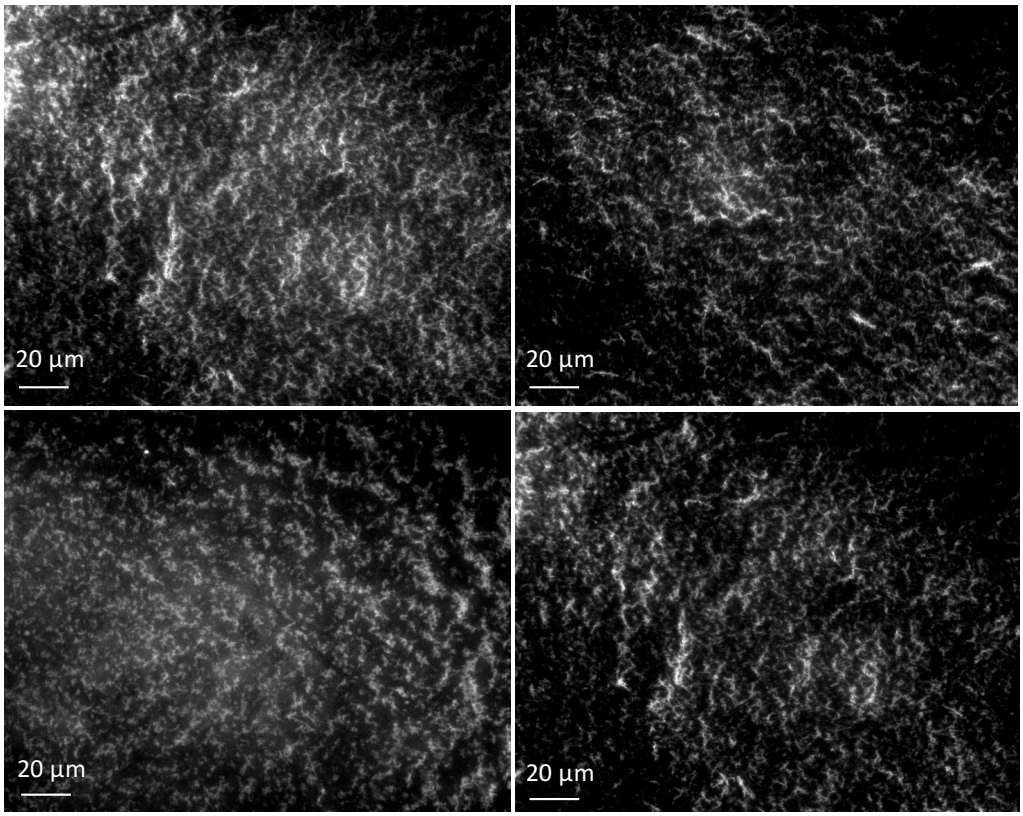


Fig. 5 Fluorescence microscopy of $[A_9G]_n/[T_9C^*]_n$ after the click reaction with Azide-fluor 545.

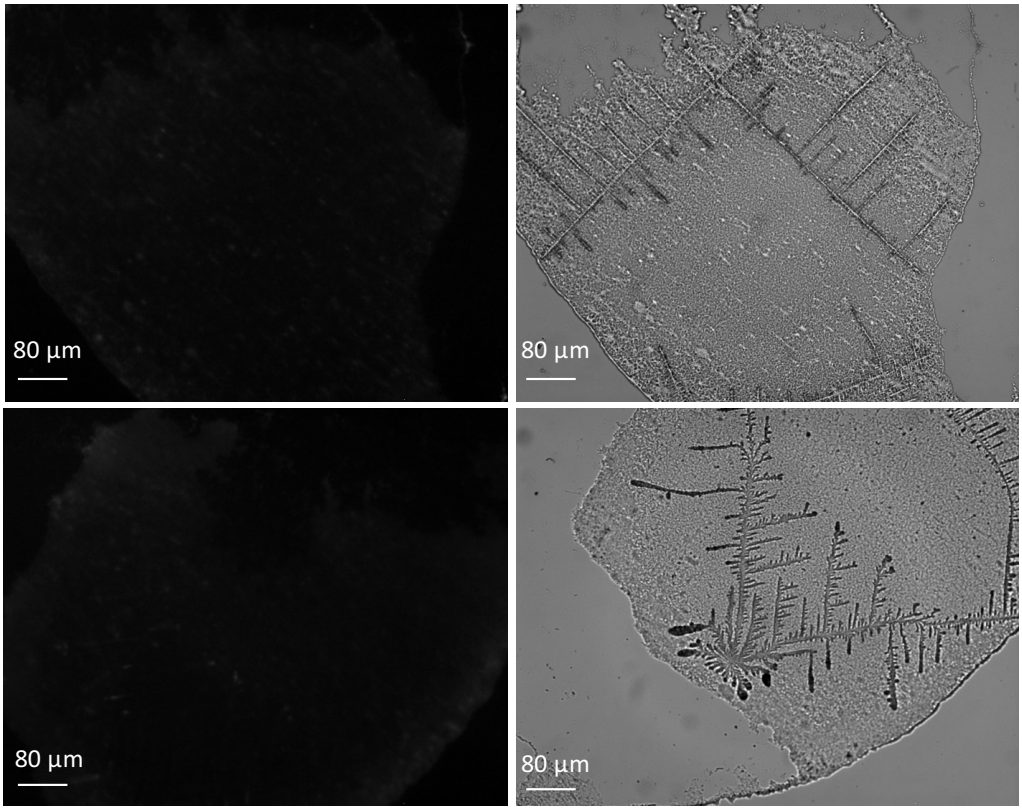


Fig. 6 Fluorescence microscopy of $[A_9G]_n/[T_9C]_n$ after the click reaction with Azide-fluor 545.

DNA Nanotechnology Hot Paper

International Edition: DOI: 10.1002/anie.201502971
German Edition: DOI: 10.1002/ange.201502971

Enzymatic Method for the Synthesis of Long DNA Sequences with Multiple Repeat Units**

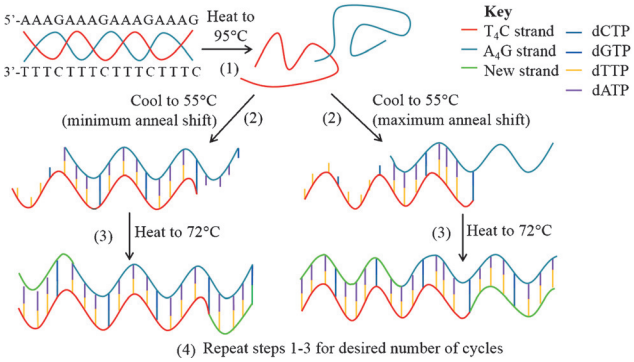
*Colette J. Whitfield, Andrew T. Turley, Eimer M. Tuite, Bernard A. Connolly, and Andrew R. Pike**

Abstract: *A polymerase chain reaction (PCR) derived method for preparing long DNA, consisting of multiple repeat units of one to ten base pairs, is described. Two seeding oligodeoxynucleotides, so-called oligoseeds, which encode the repeat unit and produce a duplex with 5′-overhangs, are extended using a thermostable archaeal DNA polymerase. Multiple rounds of heat–cool extension cycles, akin to PCR, rapidly elongate the oligoseed. Twenty cycles produced long DNA with uniformly repeating sequences to over 20 kilobases (kb) in length. The polynucleotides prepared include* $[A]_n/[T]_n$, $[AG]_n/[TC]_n$, $[A_2G]_n/[T_2C]_n$, $[A_3G]_n/[T_3C]_n$, $[A_4G]_n/[T_4C]_n$, $[A_9G]_n/[T_9C]_n$, $[GATC]_n/[CTAG]_n$, *and* $[ACTGATCAGC]_n/[TGACTAGTCG]_n$, *indicating that the method is extremely flexible with regard to the repeat length and base sequence of the initial oligoseeds. DNA of this length (20 kb ≈ 7 μm) with strictly defined base reiterations should find use in nanomaterial applications.*

The double-helical structure of DNA and the simple Watson–Crick base-pairing rules that determine helix formation[1] have led to its increasing use as a nanomaterial. Multiple applications have been reported, including nanoscale conducting wires,[2,3] in drug transport and target-driven medicine,[4,5] and as highly ordered one- and two-dimensional nanomaterials with functionalities at specific sites.[6] Developments in DNA origami[7,8] provide the ability to control the self-assembly of DNA into specific and intricate shapes with useful modifications in precise locations.[9]

Therefore, one key to using DNA in nanoscale engineering is the ability to prepare long sequences that consist of multiple repeat units, which is currently best achieved using the slippage reaction.[10–12] This provides a route towards the fabrication of long DNA that has a uniform repeat sequence covering many nanometers. Initial work demonstrated the extension of several two- and three-base repeats to a maximum of 700 base pairs (235 nm) by incubation with *Escherichia coli* DNA polymerase I.[10] A further advance showed that poly(dG).poly(dC) could be slipped to 10000 base pairs (3.4 μm) using the *E. coli* DNA polymerase I Klenow fragment.[13,14] More recent work investigated the range of sequences that could be extended by the slippage mechanism, but did not report on the lengths of the final products.[15] Commonly, the repeat unit must be able to form an internal loop for slippage to occur, limiting the number of reiterated polymers that can be prepared.[16] The level of control over the length of the DNA produced by the slippage approach relies on the quenching of the polymerase enzyme at a chosen time. These limitations prompted our investigations into an alternative method that affords DNA that consists of more complex repeat units, with lengths that can be controlled by the cyclical approach, common to the polymerase chain reaction (PCR).

PCR is routinely used in the biosciences to amplify long tracts of DNA.[17] However, products are usually limited to copies of naturally occurring DNA, and therefore, it is difficult to prepare polymers with repeating units. A general method that is capable of accurately extending short starting oligodeoxynucleotide repeat elements would be a valuable addition to the toolbox of DNA nanoengineers; such a protocol is outlined in Scheme 1. The starting materials are two short fully complementary oligodeoxynucleotides (ca. 20 bases). These oligoseeds encode the repeat element to be extended; in Scheme 1, the initial duplex has the sequence



Scheme 1. PCR-derived heat–cool cycles to extend a short oligoseed containing a repetitive element. Step 1: Melting of the duplex at 95 °C to form two single strands. Step 2: Annealing of complementary regions at 55 °C to give "shifted" duplexes with 5′ single-stranded extensions. Both a small (left) and a large (right) shift are illustrated; for the large shift, eight base pairs are considered as the minimum for stable duplex formation. Step 3: Incubation at 72 °C for the polymerase-catalyzed extension of the 5′ overhangs. Step 4: Repetition of steps 1 to 3 until the desired length of DNA is obtained.

[*] C. J. Whitfield, Prof. B. A. Connolly
Institute for Cell and Molecular Biosciences
University of Newcastle upon Tyne
Newcastle upon Tyne, NE2 4HH (UK)
A. T. Turley, Dr. E. M. Tuite, Dr. A. R. Pike
School of Chemistry, University of Newcastle upon Tyne
Newcastle upon Tyne, NE1 7RU (UK)
E-mail: andrew.pike@ncl.ac.uk

[**] This research was supported by the Biotechnology and Biological Sciences Research Council (BBSRC).

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/anie.201502971.

Table 1: Theoretical minimum and maximum DNA extension products by repeat unit length ($n = 2$–10).

No. of cycles	$[AG]_{10}/[TC]_{10}$ ($n = 2$)		$[A_2G]_7/[T_2C]_7$ ($n = 3$)		$[A_3G]_5/[T_3C]_5$ ($n = 4$)		$[A_4G]_4/[T_4C]_4$ ($n = 5$)		$[A_9G]_2/[T_9C]_2$ ($n = 10$)	
	min.[a]	max.[b]	min.[a]	max.[b]	min.[a]	max.[b]	min.[a]	max.[b]	min.[a]	max.[b]
0	20	20	21	21	20	20	20	20	20	20
5	30	392	36	393	40	392	45	330	70	330
10	40	11152	51	11151	60	11152	70	10130	120	10130
15	50	36152	66	36141	80	36152	95	35130	170	35130
20	60	61152	81	61131	100	61152	120	60130	220	60130

The algorithm used to determine the theoretical minimum and maximum product lengths assumes that 1) at least eight base pairs must exist to give a duplex stable enough for elongation, that is, the minimal stable overlap is eight for $[AG]_{10}/[TC]_{10}$, nine for $[A_2G]_7/[T_2C]_7$, eight for $[A_3G]_5/[T_3C]_5$, and ten for $[A_4G]_4/[T_4C]_4$, and that 2) the polymerase completely fills any single-stranded regions. [a] Minimum theoretical length (base pairs) = original length+($\gamma \times n$). [b] Maximum theoretical length (base pairs) = original length $\times 2^{\gamma}$−[(min. stable overlap $\times 2^{\gamma}$)−min. stable overlap], where n = number of base pairs in the repeat unit and γ = number of cycles.

$[A_3G]_5/[T_3C]_5$, that is, a repeat of four base pairs ($n = 4$). Heating denatures the duplex, and cooling allows for re-hybridization. Although the most stable duplex will be the starting 20-mer, structures displaced by an integral value of the repeat length are possible. For the example given in Scheme 1, the minimum possible shift is four bases (1 × repeat length). The maximum displacement cannot be rigorously stated; however, assuming that at least eight base pairs are needed for stable duplex formation, a shift of eight bases (2 × repeat length) is a possibility. All shifted structures possess 5′ single-strand overhangs and, as such, are DNA polymerase substrates. Filling in by the polymerase extends the original seeds while exactly maintaining the repeat unit. Repeating the heat–cool cycles, akin to PCR, rapidly elongates the seeds. The theoretical lengths attainable are shown in Table 1, assuming that an overlap of at least eight base pairs is required to generate an elongation-competent duplex and that the polymerase is fully capable of filling the overlaps produced. If only a minimal shift takes place, the seeds grow in a slow linear fashion; however, with maximal displacement, rapid expansion is possible. The greatest theoretical length is unlikely to be achieved as maximal displacement at each stage is improbable and the polymerase may lack the capacity to fill the large gaps present in later cycles. Furthermore, selecting more than eight base pairs for stable duplex formation also reduces the lengths that may be realizable. Nevertheless, Table 1 clearly illustrates the potential of this method to generate long DNA segments with multiple repeat units. The heat–cool cycles that are necessary for oligoseed extension demand a thermostable DNA polymerase, and a *Thermococcus gorgonarius* family B polymerase (Tgo-Pol) mutant, Z3, was used.[18,19] Z3 is disabled in the 3′→5′ proof-reading exonuclease (exo−) and has an insertion in the fingers domain—mutations that confer the ability to read non-standard bases in the template strand.

To gain an idea of the maximum DNA lengths that could be achieved, extension was initially investigated with Tgo-Pol Z3 (exo−) and $[A_3G]_5/[T_3C]_5$ as the starting oligoseed. DNA of escalating size builds up as the number of heat–cool cycles increases, with products of over 20000 bases (20 kb) visible after 15 cycles. The maximum size range was generated after around 20–25 cycles, and further rounds of up to 40 cycles did not lead to longer products (Figure 1; see also the Supporting Information, Figures S1 and S2). The addition

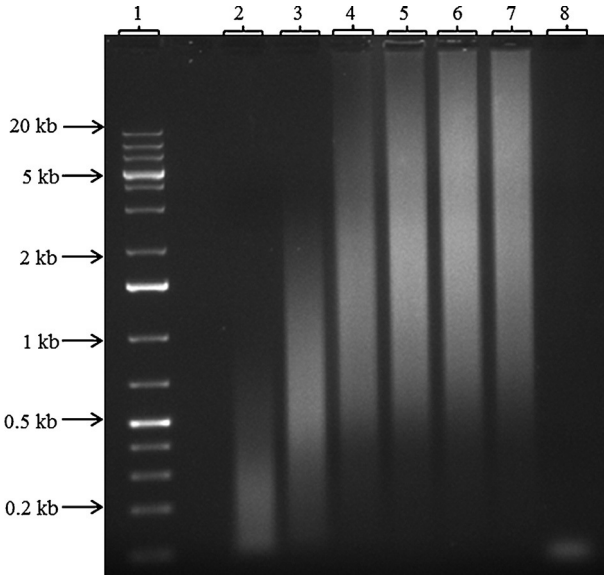


Figure 1. The extension products produced from an $[A_3G]_5/[T_3C]_5$ seed duplex during heat–cool cycling were analyzed on 0.3% SeaKem Gold Agarose gel. Lane 1: DNA ladder; lane 2: 5 cycles; lane 3: 10 cycles; lane 4: 15 cycles; lane 5: 20 cycles; lane 6: 25 cycles; lane 7: 30 cycles; lane 8: no polymerase (control).

of a fresh aliquot of 2′-deoxyribonucleoside triphosphates (dNTPs) also failed to increase product size, showing that the depletion or degradation of the dNTPs is not limiting the attainable length (Figure S3). An Image J analysis plot shows that the mean DNA product size increases between cycles 2 and 20 (Figures S4 and S5). Therefore, the ability to stop the reaction after any number of rounds affords some control over the length of the DNA produced. In all instances, the extension products appeared as an elongated smeared band covering a size range of approximately 0.5 to 20 kb. This dispersion undoubtedly arises from the multiple modes in which intermediate fragments can anneal, giving differently sized single-stranded extensions and, consequently, products of various lengths. The smallest size after 15 cycles, approximately 0.5 kb, is greater than the theoretical minimum, and as expected, the longest DNA segment (20 kb) did not reach the theoretical limit of 36 kb, but it is still of a suitable length for nanomaterial applications. However, access to DNA products with a narrower size distribution is required, and so a further gel-purification step was employed.

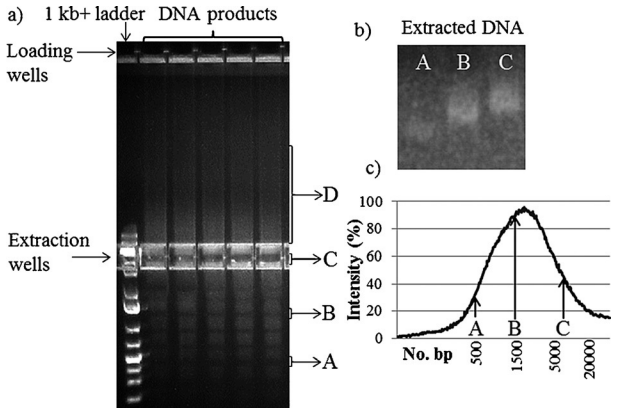


Figure 2. a) Lonza FlashGel DNA system depicting the DNA extraction process of samples A, B, and C and the unextracted DNA (D). b) Lonza gel analysis of the extracted DNA products A–C. c) Image J analysis of lane 4, Figure 1, showing the product length distribution as a percentage of the highest level of intensity in the lane.

The Lonza FlashGel DNA system enables the simple purification of DNA according to length[20] and was used to enrich products to a more limited size range. DNA with a size distribution of 0.5–20 kb was applied to multiple lanes of the Lonza gel, and selected aliquots of DNA were removed from the extraction wells (Figure 2). As an illustration, regions corresponding to DNA of 0.5 kb (sample A), 1.5 kb (sample B), and 7 kb (sample C; based on comparison with the standard ladder) were extracted from the gel. The successful removal of samples A and B is confirmed by the gaps in the elongated smeared band running beyond the extraction wells (Figure 2a). Sample C has just entered the extraction wells, and DNA at position D (>10 kb) is pre-purification, still to reach the wells. The three samples were subsequently analyzed on a second Lonza gel, showing much tighter bands (Figure 2b). Before purification, the total amount of all extended DNA products from a typical starting oligoseed concentration of 6 ng μL^{-1} was estimated from the electronic absorption at 260 nm[21] and was typically 200 ng μL^{-1}. The amounts in the purified samples varied in line with the Image J analysis plot (0.5 kb, sample A: 9 ng μL^{-1}; 1.5 kb, sample B: 34 ng μL^{-1}; and 7 kb, sample C: 11 ng μL^{-1}; see Table S2). In each instance, the DNA products A, B, and C were characterized by atomic force microscopy (AFM) and fluorescence microscopy.

AFM confirmed the presence of ds DNA in samples A and B. A 5 μL drop of each sample was spotted onto a freshly cleaved mica surface, and the DNA was immobilized by molecular combing. Typical AFM images are shown in Figure 3a and b. The mean height of the DNA across all observed images was 0.7 ± 0.2 nm, which is in agreement with previous AFM reports of DNA duplex dimensions.[22,23] A length analysis was performed on each sample and indicated that the mean observed lengths, 175 ± 18 nm (0.5 kb) and 420 ± 55 nm (1.2 kb), were close to the DNA lengths expected for samples A and B, respectively (for details, see the Supporting Information).

Labelling of sample C with YOYO-1, a fluorescent intercalator,[24] allowed for spectroscopic characterization of the

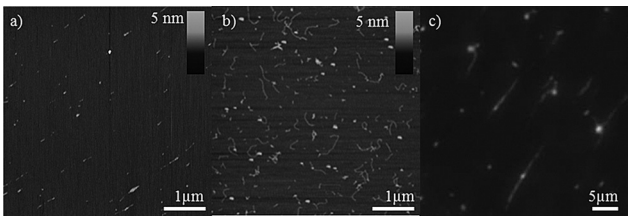


Figure 3. a) AFM image of an enriched sample A (500 bp) on mica. b) AFM image of an enriched sample B (1500 bp) on mica. c) Fluorescence microscopy image of an enriched sample C (7000 bp) on a glass slide.

enriched 7 kb band. The fluorescence micrograph in Figure 3c shows DNA strands that are 2.5–4 μm long, which corresponds to 7–12 kb. In, similar studies, other YOYO-1-labelled samples also appeared to be longer than expected owing to limitations of the optics of the fluorescence microscope (see Figure S9 and Table S5),[25] but in each case, the lengths of the observed DNA strands were in good agreement (± 1.6 μm) with the targeted extraction lengths.

Having established that the heat–cool cycling can extend the duplex $[A_3G]_5/[T_3C]_5$ (repeat length $n=4$), alternative seeds were explored. The extension products observed with $[AG]_{10}/[TC]_{10}$ ($n=2$), $[A_2G]_7/[T_2C]_7$ ($n=3$), and $[A_4G]_4/[T_4C]_4$ ($n=5$) as the primer templates are compared in Figures 4 and S10. These two-, three-, and five-base repeat units all gave extension products of comparable lengths to the four-base oligoseed and again showed efficient extension to more than 20 kb after 20 cycles. The products from the extension of the $[A_4G]_4/[T_4C]_4$ seed after five cycles were sequenced (Figure 4b, for more details see Figures S11 and S12), and 305 bases of AAAAG repeat units were identified, which correlates well with the 330 bp product expected after five cycles (see Table 1).

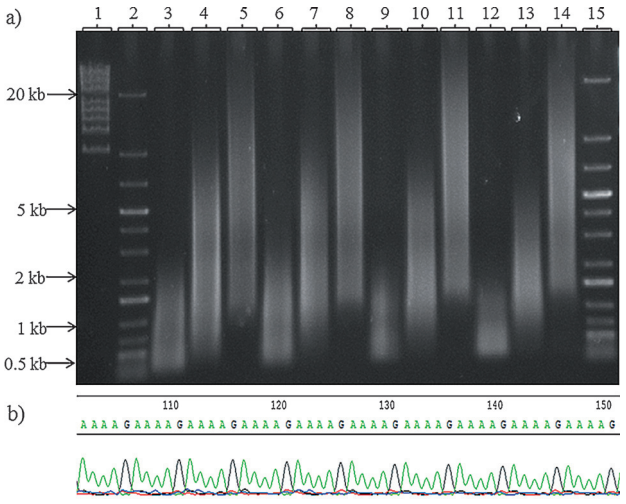


Figure 4. a) DNA extension products from $[AG]_{10}/[TC]_{10}$ (lanes 3–5), $[A_2G]_7/[T_2C]_7$ (lanes 6–8), $[A_3G]_5/[T_3G]_5$ (lanes 9–11), and $[A_4G]_4/[T_4C]_4$ (lanes 12–14) after 5, 10, and 20 cycles, respectively, were analyzed on 0.3% SeaKem Gold Agarose gel. Lanes 1, 2, and 15 show DNA ladders. b) A segment of the DNA sequencing data of the DNA extension product of $[A_4G]_4/[T_4C]_4$ after 5 cycles (Figure 4a, lane 12).

To increase potential applications, the diversity in the DNA repeat sequences was expanded. The simplest seed, the repeating polymer of $[A]_{20}/[T]_{20}$ ($n=1$), could be extended to give very long polymers (Figure S13) comparable to the equivalent product obtained by the slippage reaction.[13] The 20 bp oligoseed $[A_9G]_2/[T_9C]_2$ ($n=10$) could also be elongated, but yielded shorter products of up to approximately 2 kb. The ability to extend repeats of the sequence type $[A_xG]/[T_xC]$, where $x=1$ to 10, affords control over the base-pair density, that is, the ability to locate individual G:C pairs at defined distances from each other. Thermal cycling also succeeds with sequences that contain all four bases as shown with $[GATC]_5/[CTAG]_5$, ($n=4$) and $[ACTGATCAGC]_2/[TGACTAGTCG]_2$ ($n=10$; see Figure S14). Analysis of the DNA melting temperature outlined the expected increase in melting temperature from the oligoseed to the product (Figure S15).

The experiments described thus far have been carried out with Tgo-Pol Z3 (exo−), a mutated version with improved ability to copy non-standard bases.[18] Therefore, this polymerase is potentially useful for a longer-term aim, namely the incorporation of modified bases into repeating oligodeoxynucleotides for the preparation of future hybrid DNA nanomaterials. Nevertheless, all extensions described in this publication, which involve only the four standard bases, can be efficiently executed with a standard, commercially available archaeal DNA polymerase, and the method is not dependent on an exotic enzyme. This was demonstrated by comparing the extension efficiencies of the wild-type Tgo-Pol Z3 (exo−) DNA polymerase and the *Pyrococcus furiosus* DNA polymerase (Pfu-Pol exo−). After 10 cycles, the wild-type Pfu-Pol exo− polymerase was slightly more efficient than the Z3 mutant (Figure S16). However, the difference in efficiency between the DNA polymerases is minimal, and as Z3 is more accepting of modified bases, it shows promise for further applications. In a second comparison between the exo− and exo+ variants of Pfu-Pol, the exo+ variant was unable to extend the oligoseeds by the heat–cool method described here (Figure S17). Finally, this heat–cool cycling method was compared with the previously described isothermal slippage approach.[13] For oligoseeds of $n=2$ to 5, the length of the product produced was consistently higher when our cycling procedure was applied (Figure S18–S20), suggesting that this method, at least with archaeal exo− polymerases, is more suitable for the assembly of long repeating DNA sequences.

In summary, the elongation procedure described here provides a flexible method for preparing long DNA sequences with multiple repeating units, and is compatible with very many oligoseed sequences. This offers an alternative technique to the established isothermal slippage route towards long DNA. With both the cycling and slippage methods, the DNA product size distribution necessitates further gel enrichment. However, especially with the Tgo-Pol Z3 (exo−) polymerase, our method affords a route towards the introduction of multiple, uniformly spaced modified bases to give DNA with properties useful in nanomaterial fabrication, that could be used as metal-binding sites for the assembly of conducting nanowires.

Keywords: DNA · DNA structures · nanotechnology · polymerase chain reaction

How to cite: *Angew. Chem. Int. Ed.* **2015**, *54*, 8971 – 8974
Angew. Chem. **2015**, *127*, 9099 – 9102

[1] J. D. Watson, F. H. C. Crick, *Nature* **1953**, *171*, 737 – 738.
[2] S. A. Farha Al-Said, R. Hassanien, J. Hannant, M. A. Galindo, S. Pruneanu, A. R. Pike, A. Houlton, B. R. Horrocks, *Electrochem. Commun.* **2009**, *11*, 550 – 553.
[3] G. I. Livshits et al., *Nat. Nanotechnol.* **2014**, *9*, 1040 – 1046.
[4] V. Bagalkot, O. C. Farokhzad, R. Langer, S. Jon, *Angew. Chem. Int. Ed.* **2006**, *45*, 8149 – 8152; *Angew. Chem.* **2006**, *118*, 8329 – 8332.
[5] E. S. Andersen et al., *Nature* **2009**, *459*, 73 – 76.
[6] T. Wang, D. Schiffels, S. Martinez Cuesta, D. K. Fygenson, N. C. Seeman, *J. Am. Chem. Soc.* **2012**, *134*, 1606 – 1616.
[7] P. Rothemund, *Nature* **2006**, *440*, 297 – 302.
[8] B. Wei, M. Dai, P. Yin, *Nature* **2012**, *485*, 623 – 626.
[9] A. C. Pearson, J. Liu, E. Pound, B. Uprety, A. T. Woolley, R. C. Davis, J. N. Harb, *J. Phys. Chem. B* **2012**, *116*, 10551 – 10560.
[10] C. Schlötterer, D. Tautz, *Nucleic Acids Res.* **1992**, *20*, 211 – 215.
[11] M. J. Hartenstine, M. F. Goodman, J. Petruska, *J. Biol. Chem.* **2000**, *275*, 18382 – 18390.
[12] J. Petruska, M. J. Hartenstine, M. F. Goodman, *J. Biol. Chem.* **1998**, *273*, 5204 – 5210.
[13] A. B. Kotlyar, N. Borovok, T. Molotsky, L. Fadeev, M. Gozin, *Nucleic Acids Res.* **2005**, *33*, 525 – 535.
[14] H. Klenow, I. Henningsen, *Proc. Natl. Acad. Sci. USA* **1970**, *65*, 168 – 175.
[15] L. M. Chi, S. L. Lam, *J. Phys. Chem. B* **2012**, *116*, 1999 – 2007.
[16] T. Kato, X. Liang, H. Asanuma, *Biochemistry* **2012**, *51*, 7846 – 7853.
[17] L. Raeymaekers, *Genome Res.* **1995**, *5*, 91 – 94.
[18] S. K. Jozwiakowski, B. A. Connolly, *ChemBioChem* **2011**, *12*, 35 – 37.
[19] B. D. Biles, B. A. Connolly, *Nucleic Acids Res.* **2004**, *32*, e176.
[20] M. Riley, H. White, FlashGel System for DNA Recovery, Resource Notes (App Note) **2009**, *7*, 17 – 20.
[21] A. V. Tataurov, Y. You, R. Owczarzy, *Biophys. Chem.* **2008**, *133*, 66 – 70.
[22] S. M. D. Watson, N. G. Wright, B. R. Horrocks, A. Houlton, *Langmuir* **2010**, *26*, 2068 – 2075.
[23] S. J. T. Van Noort, K. O. Van der Werf, B. G. De Grooth, N. F. Van Hulst, J. Greve, *Ultamicroscopy* **1997**, *69*, 17 – 27.
[24] M. L. Bennink, O. D. Schärer, R. Kanaar, K. Sakata-Sogawa, J. M. Schins, J. S. Kanger, B. G. de Grooth, J. Greve, *Cytometry Part A* **1999**, *36*, 200 – 208.
[25] T. F. Chan, C. Ha, A. Phong, D. Cai, E. Wan, L. Leung, P. Y. Kwok, M. Xiao, *Nucleic Acids Res.* **2006**, *34*, e113.

Received: March 31, 2015
Published online: June 10, 2015

