

Self-Priming Enzymatic Fabrication of Multiply Modified DNA

Colette J. Whitfield,^[a] Rachel, C. Little,^[a] Kasid Khan,^[a] Kuniharu Ijiro,^[b] Bernard A. Connolly,^[c] Eimer M. Tuite,^{*[a]} and Andrew R. Pike^{*[a]}

Abstract: The self-priming synthesis of long multiply modified DNA by extension of repeating unit duplex “oligoseeds” provides a new robust and versatile approach to the construction of multi-functional nanoarchitectures. Sterically-demanding nucleotides 7-deaza-7-I-dATP, 5-Br-dUTP, 6-S-dGTP, 5-I-dCTP as well as 5-C₈-alkyne-dCTP are multiply incorporated into the extending sequences on one or both strands of DNA demonstrating their recognition by polymerase as both substrate and template. 5-acetyl-Hg-dCTP, however, was recognized only as a relatively poor substrate. Nucleobase modifications that lie in the major groove are reliably read and written by the polymerase during the extension reaction, even when bulky or in contiguous sequences. Repeat sequence DNA over 500 bp long bearing four different modified units are readily produced by this method. The number, position, type of modification, and the overall length of the DNA can be controlled to yield designer DNA that offers sequence-determined sites for further chemical adaptations, targeted small molecule binding studies, or sensing and sequencing applications.

DNA designed to bear multiple modifications has opened new avenues in synthetic biology,^[1] nanomaterial fabrication,^[2] bioanalytical^[3] and sequencing applications.^[4] DNA can be chemically modified at each of its three component parts – the phosphate backbone,^[5] the sugar ring^[6] or the nucleobase.^[7] Of the three, modified nucleobases offer the widest choice of functionality and have been commercially developed for insertion into DNA either enzymatically as deoxynucleotide triphosphates (dNTPs)^[8] or through automated DNA synthesis as phosphoramidite derivatives.^[9] Automated synthesis allows routine production of unmodified single-strand oligonucleotides up to 100 bases, but multiply modified oligonucleotides are typically limited to about 50 bases before low yields and high costs become prohibitive. To obtain high molecular weight DNA, oligonucleotides can be ligated in post-synthetic procedures,^[10]

either enzymatically^[11] or chemically.^[12] However, not all sequences are amenable to efficient phosphoramidite synthesis, particularly those with hairpins, highly repetitive sequences or high guanine content, each of which are of interest for nanomaterials and sensing applications. Additionally, some functional groups, such as those prone to oxidation or reactive with nucleophiles (e.g. azide) have limited compatibility with the phosphoramidite method. Enzymatic routes to modified DNA are flexible, can produce single- or double-stranded products, and have fewer restrictions on product length. Polymerase methods to copy DNA comprise isothermal amplification^[13] including loop-mediated amplification (LAMP),^[14] strand-displacement,^[15] slippage,^[16] rolling circle amplification (RCA),^[17] and primer-template extension,^[18] as well as thermal cycling amplification such as polymerase chain reaction (PCR).^[19] Methods such as primer-extension can introduce modified bases on the growing strand,^[20] whilst PCR can introduce different modifications on both strands as long as the modified nucleotides can be both incorporated and read by polymerases.^[18] DNA, RNA, threose nucleic acid (TNA), peptide nucleic acid (PNA) and other nucleic acid analogues can be fabricated chemically or enzymatically by incorporation of modified bases, bases with functional substituents, and artificial bases that possess non-standard absorption, emissive, redox or recognition properties. These include chromophoric,^[21] fluorescent,^[22] and electroactive^[23] base analogues; non-canonical bases;^[24] and highly specific artificial base-pairs.^[25] Alternatively, bases with reactive moieties such as azide,^[26] alkyne,^[27] and amino^[28] groups attached by linkers can be incorporated in DNA to allow post-synthetic functionalization and to reduce the synthetic burden at the monomer level. The wide choice of modified dNTPs and the plethora of commercially available polymerases are two powerful tools in the biochemists workshop. In this work, our goal is to exploit enzymatic synthesis to introduce multiple modifications into long repeat sequence DNA.

[a] Dr. C. J. Whitfield, Miss R. C. Little, Mr. K. Khan, Dr. E. M. Tuite, Dr. A. R. Pike

Chemistry – School of Natural and Environmental Sciences
Newcastle University
Newcastle upon Tyne, NE1 7RU, U.K.

E-mail: Andrew.Pike@newcastle.ac.uk and
Eimer.Tuite@newcastle.ac.uk

[b] Prof. Dr. K. Ijiro

Research Institute for Electronic Science
Hokkaido University
Sapporo, 001-0020, Japan

[c] Prof. Dr. B. A. Connolly

Institute for Cell and Molecular Biosciences
Newcastle University
Newcastle upon Tyne, NE2 4HH, U.K.

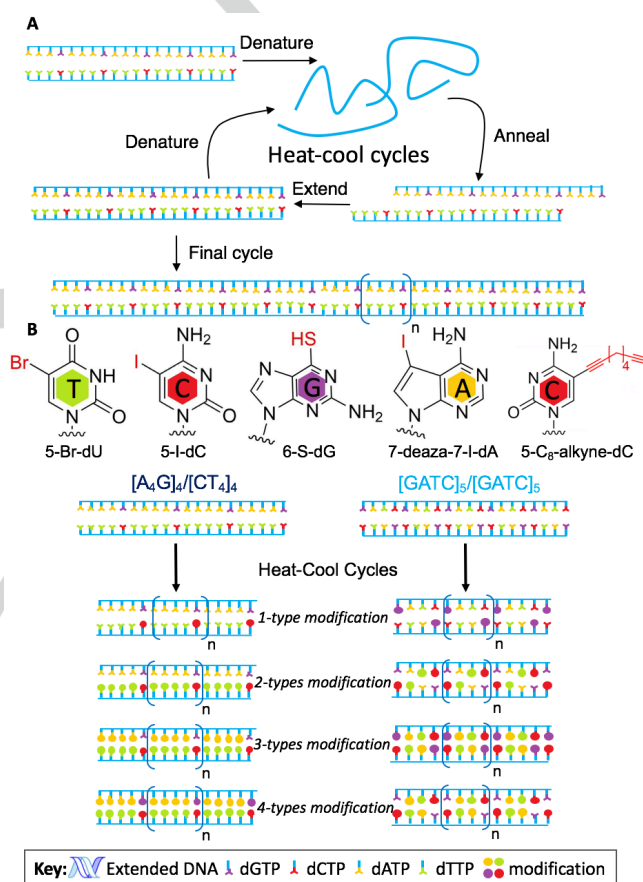
Supporting information for this article is given via a link at the end of the document.

Many repeat sequences have biological functions (e.g. telomeres) or are associated with disease states when expanded (e.g. trinucleotide repeats for polyQ disorders such as Huntington's disease). Detection of copy numbers for repeat-sequences such as variable number tandem repeats (VNTRs) provides insight into the progression of associated disorders, hence amplification of these sequences is of diagnostic interest.^[29] Production of high copy number repeat sequences is also important for nanomaterials applications of DNA, where the polynucleotide acts as a template for assembly of photonic or electronic elements. For example, isothermal enzymatic slippage extension has been used to synthesise tri-block DNA copolymers composed of natural and unnatural nucleobases for fabrication of hybrid 1D nanomaterials.^[30] The DNA strand itself can be used as a template to guide the deposition of metal ions to synthesise 1D nanomaterials,^[31] for example, Ag-DNA wires.^[32] Additionally, the DNA template can guide the polymerization of monomers to produce uniformly decorated materials to provide evenly spaced nucleation sites for metal ion deposition to improve DNA wire properties.^[33] Long DNA strands containing high copy numbers (up to 10 kbp) of short repeat sequences (2-3 nucleotides) are readily synthesised from 10-20 mer duplex oligonucleotides using isothermal slippage methods.^[16, 34] Although isothermal slippage has been shown to occur for 6-10 bp repeats such as telomeric sequences,^[35] we have found that intermediate sized repeat sequences (4-40 nucleotides) are most efficiently extended by thermal cycling methods.^[36] Moreover, it is difficult to produce high copy number repeat-sequence DNA containing modified nucleotides by isothermal extension, even for short repeat sequences and low density modifications. These constraints motivated us to extend our previously reported thermal cycling extension method to facilitate introduction of modified bases at specific positions on long repeat-sequence DNA chains, with the motivation of producing modified DNA for nanomaterials and diagnostic applications and potential uses in data/coding and security operations.

In this work, we present a versatile enzymatic method for the production of long modified DNA, which overcomes some limitations of phosphoramidite chemistry and other polymerase approaches. Two DNA polymerases were chosen to demonstrate the compatibility of the method with both in-house and commercially available DNA polymerases. These were, respectively, a *Thermococcus furiosus* Family B DNA polymerase (Tgo-Pol) exonuclease minus (exo-) variant, Z3, which has previously shown promise in handling modified dNTPs,^[37] and the *Pyrococcus* species GB-D DNA polymerase, Deep Vent exo-.^[38] By using selected repeat-sequences, DNA containing single or multiple bulky modifications at specific loci on either strand can be produced. Using an adapted PCR protocol,^[36] modified DNAs with designed repeating units that include handles for further functionalization are synthesized. The heat-cool cycle method extends DNA by forcing a slip in sequence of one or more repeating units within an initial unmodified oligoseed of around 20 bases (Scheme 1A). The resulting overhangs are then copied by a DNA polymerase, adding complementary canonical or modified nucleobases, thus

extending the DNA by the number of repeating units that have been displaced. This approach affords control over the spacing between modified sites and governs the total number and type of base modifications that are designed into final DNA products. Post-synthetic functionalization at the modified bases opens up possibilities for designer DNA-based materials, through a range of DNA-compatible chemistries depending on the type of modified dNTP inserted into the growing DNA strands.

The modified nucleotides used in this study were chosen to contain sterically-demanding groups attached to the ring system as well as anchor points for post-synthetic functionalization. They include single atom exchange modifications; viz. two C5-modified pyrimidines (5-I-dCTP and 5-Br-dUTP), a 7-modified purine (7-deaza-7-I-dATP) and a C6-modified purine (6-S-dGTP), and also a long carbon chain cytosine modification (5-C₈-alkyne-dCTP);



Scheme 1. (A) Schematic overview of the synthetic protocol to produce designer DNA sequences bearing single and multiple base modifications. (B) Molecular structures of the modified nucleobases are shown (top – 5-Br-dUTP, 5-I-dCTP, 6-S-dGTP, 7-deaza-7-I-dATP and 5-C₈-alkyne-dCTP, left to right), the two oligoseeds (middle) and examples of the possible extended modified DNA products (bottom).

structures of these modified nucleotides are shown in Scheme 1B. Some comparative studies were also performed with the heavy atom modified cytosine 5-acetyl-Hg-dCTP which, like the iodo- and bromo-derivatives, has potential applicability in next generation sequencing using electron microscopy.^[39] In previous

reports where several or all bases were replaced with modified nucleotides using PCR on a long template,^[12, 20] the modifications were displaced from the bases by alkene or alkyne linkers at the 5-pyrimidine and 7-deaza- or 8-purine positions. In this work, which complements previous studies, several bulky groups were directly attached to or included in the bases which introduces steric constraints. In the case of 6-S-dGTP, the bulky thiol is in the H-bonding region, which is likely to interfere with pairing and change the local duplex structure. 6-S-dGTP has previously been demonstrated to act as a relatively good substrate in a primer extension assay^[40] but there are no reports of its behaviour as a template. The thiol-modified guanosine affords a metal coordination site for potentially assembling of coordination polymers with metal ions,^[41] and also has photo-cross-linking potential.^[42] The halide substitutions on the bases are suitable chemical handles for the C-C linkage of additional functional groups through Pd-catalysed Sonogoshira, Suzuki, and Heck reactions, which have been adapted for aqueous media^[43] and recently exploited for labelling of DNA.^[44] Finally, the alkynyl-cytidine derivative is incorporated to exploit the established cycloaddition “click” reaction with azide groups, demonstrated here through fluorophore functionalization.

Given the range of modified nucleotides available, the possibilities to synthesize designer DNA of pre-determined sequence can be controlled by judicious choice of the repeating unit in the oligoseeds. Here, the duplexes [GATC]₅/[GATC]₅ and [A₄G]₄/[CT₄]₄ were chosen to demonstrate: i) the inclusion of one type of modification into one strand of a repeating nucleotide sequence ([A₄G]₄/[CT₄]₄), ii) the inclusion of one type of modification into both strands of a self-complementary sequence ([GATC]₅/[GATC]₅), and iii) the inclusion of two or more modifications within a nucleotide sequence, both [GATC]₅/[GATC]₅ and [A₄G]₄/[CT₄]₄.[‡] To synthesise modified DNA, the standard nucleotide is simply replaced with its modified derivative, *i.e.* swapping out dGTP for 6-S-dGTP during a set number of heat-cool cycles of the repeating polymerase reaction. Scheme 1B, summarises some of the possible DNA sequences that can be produced from two oligoseeds and five modified nucleotides which we exploit to produce exemplary designer DNAs with several sites for post-synthetic functionalization along the duplex.

Using a PCR-based approach that produces repeat-unit DNA, as outlined in Scheme 1,^[36] we explore the extension of two oligoseeds, [A₄G]₄/[CT₄]₄ and [GATC]₅/[GATC]₅ using combinations of five different modified nucleotides. Initially, we show that by replacing one of the four standard dNTPs with a modified equivalent, DNA bearing a single type of modified nucleobase on just one strand of the extended [A₄G]₄/[CT₄]₄ is produced. In that case, the modified nucleotide must act as a polymerase substrate. In the second case, we discuss incorporation of modified nucleobases in both DNA strands using the self-complementary oligoseed [GATC]₅/[GATC]₅. In this case, the modified nucleotide must be recognised by polymerase as both a substrate and in the template. Finally, using both types of oligoseed we demonstrate that the described

method can be used to synthesise fully modified DNA; *i.e.*, several modified bases within a user-defined sequence, on one or both strands to yield DNA products >500 bp in length. The results shown in the manuscript are those obtained using Tgo-Pol-Z3 exo- polymerase, but Deep Vent exo- works equally well (Figure S1).

The oligoseed [A₄G]₄/[CT₄]₄ was chosen to illustrate how a modified base, as either a modified G or C, could be multiply incorporated into a repeat sequence at a fixed base-separation, *i.e.* at every fifth site within one strand of the extending duplex. Conversely, using a modified A or T in the extension reaction demonstrated that multiple modifications can be located adjacent to each other over many repeating units of DNA, *i.e.* four contiguous modifications on one strand. First, we consider the incorporation of a single type of C5 or C7 modified dNTP. 10 heat-cool cycles of extension using [A₄G]₄/[CT₄]₄ and either 5-I-dCTP or 5-C₈-alkyne-dCTP yields highly extended DNA of 8 kb and 5 kb respectively (see gel electrophoresis images, Figure S2A – 1B).

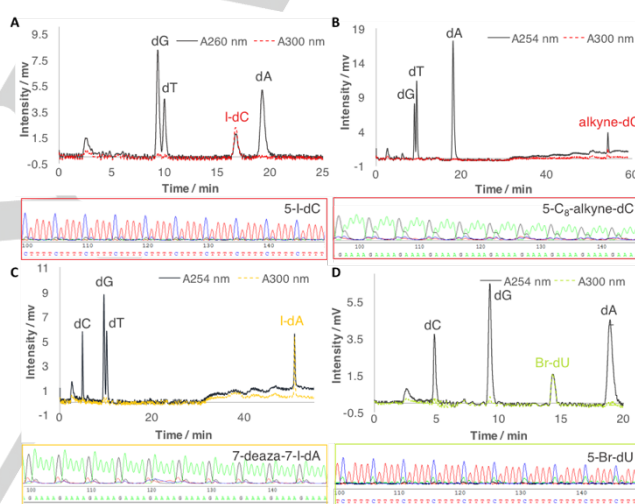


Figure 1. HPLC traces of the DNA components after digestion and the DNA Sanger sequencing data of the extended DNA products from oligoseed [A₄G]₄/[CT₄]₄ bearing: (A) 5-I-dC, (B) 5-C₈-alkyne-dC (C) 7-deaza-7-I-dA and (D) 5-Br-dU

The DNA products contain the modified cytidine at every fifth site on one strand of the extended duplex and, in both cases the DNA produced is similar in length to the control reaction using the unmodified dCTP.^[36] Nucleotides with more sterically demanding modifications, 6-S-dGTP and 5-acetyl-Hg-dCTP, were also incorporated into just one strand of the extending DNA but gave shorter products of ca. 500 bp which could not be further extended with additional cycling (Figure S2E and Figure S3). Employing the same starting oligoseed but inserting either a modified dATP or dUTP results in DNA bearing 4 consecutive modifications on the same strand. After 10 cycles, the method yields DNA products bearing 7-deaza-7-I-dA (500 bp) and 5-Br-dU (5 kb) (Figure S2C and S2D) although further cycling allows attainment of lengths of modified DNA comparable to that obtained with the native nucleotides.

[‡] oligoseeds are designated as [5'-3'] / [5'-3']

To confirm that the DNA products detected in the gel images contained the modified nucleobase sequences according to our design, base analysis by enzymatic digestion and HPLC was performed. Phosphodiesterase/phosphatase digestion of the synthesized DNA followed by HPLC analysis demonstrated that the modified nucleobases were reliably incorporated (Figure 1 and Figure 4B). UV-Vis (Figure S4C and Table S1) and HR-MS (Table S2) of the isolated fractions demonstrated that each modified base was included in the sequence as designed.

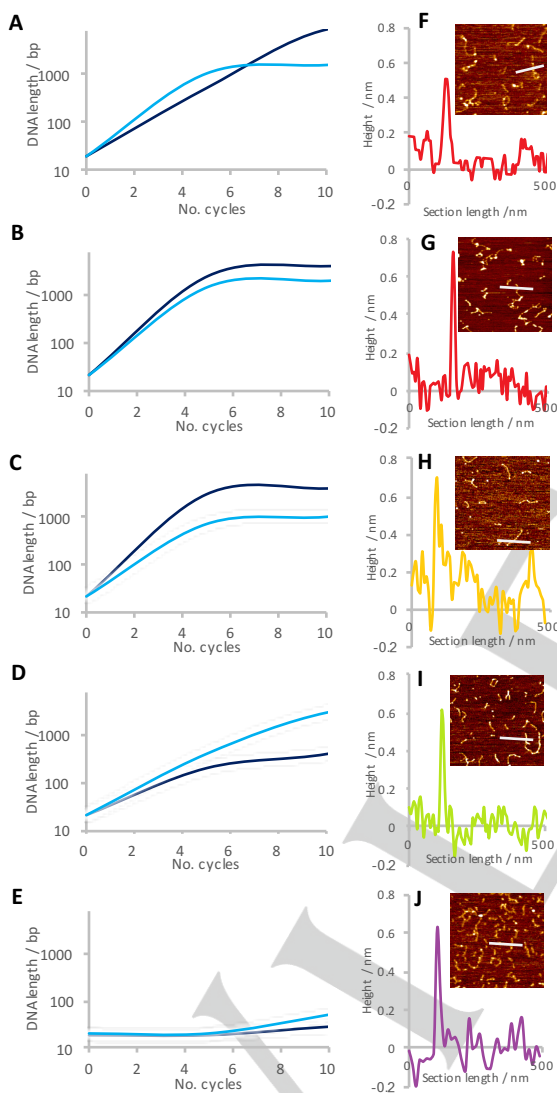


Figure 2. A-E compares the lengths of extension products by number of heat-cool cycles indicating the relative incorporation rates for the different modified bases with both oligoseeds, [A₄G]_n/[CT₄]_n (dark blue) and [GATC]₅/[GATC]₅ (light blue). Details of rate evaluation are provided in the Supplementary information, Figure 2. AFM images for 30 cycle DNA products from the extension of [A₄G]_n/[CT₄]_n with single nucleobase modifications as in Figure S1 are shown in F to J respectively. Height scale is 1.5 to 2.1 nm, scan size is 1 μm, and profile trace lines are 500 nm in each case.

Sanger sequencing of the above products bearing all the modified bases[#] confirmed the integrity of the repeating sequence (see boxes within Figure 1 and Figure S4B). The polymerase was able to select the modified dNTP correctly to form repeat sequences of single or multiple modifications. Equally, the polymerase was able to read the modified bases once incorporated and the oligoseed sequence was retained over multiple repetitions. Because of the difficulty of sequencing repeat sequences, only short segments of the product showed clear results, however, the oligoseed can bind at any position on the DNA product, therefore the correct sequence must be retained throughout the DNA product to conclude in accurate sequencing results. Additionally, analysis of synthesized DNA by enzymatic digestion with HPLC represents the length of the product up to 10 kbp. Therefore, polymerase fidelity is not significantly reduced as the "template" strand is increased.

To ensure that the polymerase retains its fidelity, *i.e.* it accurately writes only the correct modified base, extension reactions using both oligoseeds were performed where one dNTP was omitted from the reaction. In all cases no extension was observed, indicating that the polymerase did not mis-write into the sequence and continue the extension with an alternative base when the correct base was absent (see Figure S5). Therefore, base mis-incorporation does not occur when an artificial and potentially "hard-to-write" modified dNTP is used in the reaction.

Exemplary atomic force microscopy (AFM) images of extended [A₄G]_n/[CT₄]_n products (30 heat cool cycles) bearing single base modifications are shown in Figure 2 (F–J; further analysis is provided in Figure S6 and Table S3). In each instance the modified DNA exhibits a duplex height between 0.5 to 0.7 nm corresponding to previously reported literature values for dsDNA imaged on mica.^[45] It is noticeable that the expected linear morphology of DNA is recognizable, however some minor structural variations can be observed between the different extended products

Extension of the second oligoseed, the self-complementary [GATC]₅/[GATC]₅, using only a single type of modified dNTP in the reaction mixture yields a DNA product that has a repeating modification at every fourth position on each strand of the final duplex, which places the modifications in adjacent basepairs. After 10 heat-cool cycles, extension products of over 2 kb were produced for reactions involving any one of the four modified nucleobases, 5-I-dCTP, 5-Br-dUTP, 7-deaza-7-I-dATP or 5-C₈-alkyne-dCTP (Figure S2F-2I). Once again the extension using 6-S-dGTP yielded shorter DNA products of ca. 500 bp (Figure S2J).

Sanger sequencing from a similar oligoseed sequence, one that also bears all four bases on each strand but is not self-complementary, [GAATC]₄/[GATTC]₄, was used to confirm enzymatic fidelity, see Figure S7.

From analysis of the gel band intensities (Figure 2, and Figure S8), it is clear that in most cases there is minimal

[#] DNA containing 5-acetyl-Hg-dCTP could not be synthesised in the quantities required for sequencing

difference in the initial rate of incorporation of modified bases between the two types of oligoseed. However, in the case where multiple adjacent, 5-Br-dUTP units are incorporated into the $[A_4G]_4/[CT_4]_4$ oligoseed, a noticeable reduction in the length of the extension products up to 10 cycles was observed. This suggests that, since bromine is more electronegative than iodine, the higher charge density of several neighbouring bromo-modifications might slow the polymerase reading and/or writing processes even though after 30 cycles the final product lengths were similar. Likewise, we found little difference in the enzymatic processivity of different

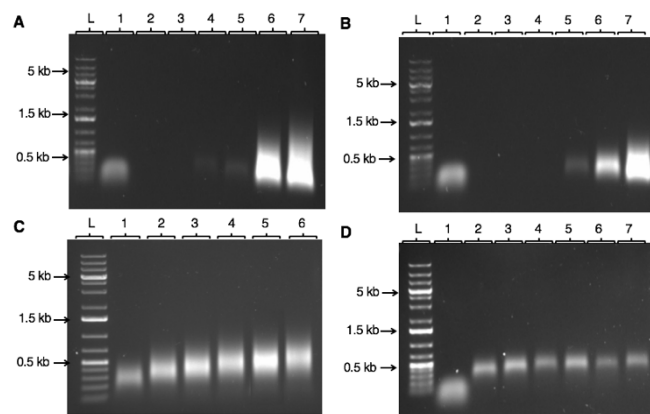


Figure 3. Agarose gels of DNA extension products with $[A_4G]_n/[CT_4]_n$ (A and B) and $[GATC]_5/[GATC]_5$ (C and D) with 6-S-dG, 5-Br-dU and 7-deaza-7-I-dA (A and C) and 6-S-dG, 5-Br-dU, 7-deaza-7-I-dA and 5-C₈-alkyne-dC (B and D).

types of base modification within the same oligoseed, namely between; i) a single atom or a long alkyl chain modification on the same base (cf. 5-I-dCTP and 5-C₈-alkyne-dCTP), ii) different nucleobases bearing the same single atom displacement (cf. 5-I-dCTP and 7-deaza-7-I-dATP). However, the same lag-effect for 5-Br-dUTP was observed when compared to a 5-I-dCTP modification. In these instances, the nucleobase modifications at positions C5 and C7 do not lie within the hydrogen bonding region but project into the major groove. Extension was also observed with 6-S-dGTP (Figure S2E and S2J) although with much lower extension rates for both oligoseeds, yielding 200 bp after 10 cycles, which can be rationalized by interference of the large sulphur atom in the base pairing region, thus inhibiting polymerase activity.^[40] Additionally, the enzymatic extension efficiency of $[A_4G]_n/[CT_4]_n$ was lowered considerably when dCTP was replaced with 5-Hg-acetyl-dCTP (Figure S3), comparable to previous studies where it was shown to act as a polymerase substrate.^[39] However, extension of $[GATC]_5/[GATC]_5$ to include Hg-modifications was not possible – suggesting that additional stabilization effects may be required to allow its incorporation into both strands of the dsDNA^[39] and that it is unlikely to be recognized in a template strand.

To explore whether polymerase could incorporate modified bases in longer stretches, the challenging substrate 6-S-dGTP was incorporated against cytosine in the extension of a $[G]_{20}/[C]_{20}$ seed (Figure S9). A poly⁶(S-dG).poly(dC) product of ca. 1 kbp was synthesized, the short length being consistent

with truncated lengths observed in extension of oligoseeds with 6-S-G incorporated at every fifth base (Figure S2E, S2J). We note that with 6-S-dGTP, the modification truncates the extension length as well as slowing incorporation. But for other modified dNTPs kinetics is the limiting factor, as seen with 5-Br-dUTP (Figure S2D, & S2I) where sequences as long as those with unmodified bases are produced after additional cycling.

Confident of the versatility of the method, *i.e.* that it can incorporate a single type of modification on one or both strands at either a fixed base separation or at multiple consecutive sites, the inclusion of two or more different modified dNTPs was investigated. Extension of either oligoseed, $[A_4G]_4/[CT_4]_4$ or $[GATC]_5/[GATC]_5$, with two halogen-modified pyrimidines, 5-I-dCTP and 5-Br-dUTP, in the same reaction showed high extension efficiency (Figure S10). Again, the polymerase maintained the correct repeat sequence (see Figure S10 for sequencing data). In a second extension reaction, the alkyl-modified nucleotide, 5-C₈-alkyne-dCTP, and the halogenated nucleotide, 5-Br-dUTP, were also both successfully incorporated into the extending DNA, (Figure S11), producing DNA products over 1 kb after 30 heat-cool cycles from both oligoseeds. The extension of both oligoseeds to include three modified nucleotides, either 5-C₈-alkyne-dCTP, 5-Br-dUTP and 7-I-dATP (Figure S1, Lane 8) or 6-S-dGTP, 5-Br-dUTP and 7-deaza-7-I-dATP (Figure 3A and 3C), demonstrated that three modifications can be written into the repeating sequence to yield shorter DNA products, < 1 kb. Final DNA lengths are notably shorter when 6-S-dGTP is incorporated. Finally, the possibility to design DNA bearing four different modifications was demonstrated using both oligoseed sequences. 6-S-dGTP, 5-Br-dUTP, 7-I-dATP and 5-C₈-alkyne-dCTP were incorporated starting with either type of oligoseed to extend and demonstrate the synthesis of DNA up to 500 bp comprising fully artificial bases (Figure 3B and 3D), which is considerably longer than is feasible through automated DNA synthesis and highlights the versatility of this repetitive self-priming enzymatic approach to the reading and writing of artificial nucleotides to yield multiply modified DNA.

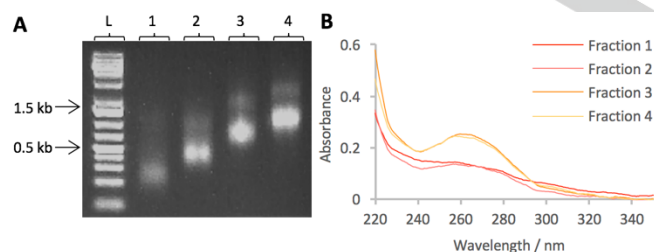
The synthetic method described in this work is versatile and complements our previous reports on the enzymatic synthesis of long repeat unit DNA.^[36] Through careful design of oligoseeds and functionally-modified nucleotides that are recognized by polymerase as both substrate and template, extended DNA of specified sequence can be synthesized to incorporate multiple sterically-demanding base modifications. Because this self-priming method does not rely on the availability of a template strand (either a natural sequence or one constructed by ligation of oligonucleotides), duplex DNA can be synthesized, the length depends on the type and number of modifications used. Since the sequence of the extended DNA is determined by the oligoseed, this method for producing modified DNA is very adaptable. Thus far, the method discussed synthesizes polydisperse DNA lengths. Undoubtedly, an ability to produce monodisperse DNA would be of great benefit.

Table 1. The incorporation efficiencies of single and multiple modifications as judged by DNA product lengths after 30 heat-cool cycles.

Modified dNs	$[A_4G]_n/[CT_4]_n$	$[GATC]_n/[GATC]_n$
I-dC	⊙	⊙
≡-dC	⊙	⊙
I-dA	⊙	⊙
Br-dU	⊙	⊙
S-dG	△	○
I-dC, Br-dU (Lane 1)	⊙	⊙
≡-dC, Br-dU (Lane 2)	○	○
≡-dC, Br-dU, I-dA (Lane 3)	○	○
S-dG, Br-dU, I-dA (Lane 4)	△	○
≡-dC, Br-dU, I-dA, S-dG (Lane 5)	△	△
No modification	⊙	⊙

⊙ represents DNA products > 2 kb, ○ > 500 bp and △ < 500 bp.

Size selection can be achieved by electrophoretic separation and sample extraction.^[36] A LonzaGel™ system can be adopted to select DNA lengths, for example, 250, 450, 800 and 1200 bp (Figure 4A), with reduced length distribution within each extracted sample. The yield of the size separated DNA products can then be quantified by UV-Vis (Figure 4B); 110, 109, 206 and 199 % for each fraction, respectively, compared to the starting oligo seed. Nonetheless, in several applications a fixed length is not required, e.g. the conductivity of a fixed length of extended DNA can be determined by careful placement of the electrodes or by laying DNA between electrodes.^[46]

**Figure 4.** Size separation of I-dC-modified DNA extension products. A) Agarose gel of size separated fractions, 1-4. B) UV-Vis of the fractions 1-4, corresponding to Lanes 1-4 in A.

In conclusion, we have shown that DNA bearing single and multiple modifications can be enzymatically synthesised from short oligoseeds using an efficient and facile PCR-based extension method. 5-C₈-alkyne-dCTP, 5-I-dCTP, 5-Br-dUTP, 7-deaza-7-I-dATP and 6-S-dGTP can all be incorporated into repeat sequence DNA derived from $[A_4G]_4/[CT_4]_4$ and $[GATC]_5/[GATC]_5$ oligoseeds. Single and multiple incorporations of one, two, three and four types of modified nucleobase were shown to yield designer DNA products from 200 bp to over 8 kb. This method affords DNA sequences previously unobtainable by automated DNA synthesis, being longer and with multiple sites

for further decoration. We anticipate that this approach will provide adaptable and emergent materials for DNA nanotechnology and find new applications in biomedicine, diagnostics, 1-D nanomaterials, DNA sequencing and coding.

Experimental Section

DNA oligoseeds

Deoxyoligonucleotides were purchased lyophilized from Eurofins and complementary strands were annealed by heating to 95 °C for 10 minutes in anneal buffer (10 mM HEPES, 100 mM NaCl and 1 mM EDTA) and allowing to cool at 1 °C per minutes to room temperature. The DNA sequences can be found in the Supplementary Methods.

DNA polymerases

Tgo-Pol-Z3 exo- was prepared and purified as described previously (30,32). Deep Vent exo- was purchased from NEB.

DNA extension

0.5 μM oligo seed (see the Supplementary Methods for more details), 200 nM or 1 unit DNA polymerase, DNA polymerase reaction buffer (200 mM Tris-HCl (pH 8.8, 25 °C), 100 mM (NH₄)₂SO₄, 100 mM KCl, 1 % Triton X-100, 1 mg/mL Bovine Serum Albumin and 20 mM MgSO₄) or 10 x ThermoPol buffer, and 0.5 mM dNTPs (see supplementary data for combinations) 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 following manufacturer's protocol.

DNA analysis

Details of the analytical techniques, (AFM, gel electrophoresis, DNA digestions and HPLC analysis, HR-MS, and Sanger sequencing) can be found in the supporting information.

Acknowledgements

This work was supported by the BBSRC Doctoral Training Partnership in Biosciences, Liverpool, Durham and Newcastle [BB/J014516/1 to C. J. W], the School of Chemistry, Newcastle University [R. C. L and K. K] and the Cooperative Research Program of "Network Joint Research Center for Materials and Devices". This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 778001.

Keywords: designer DNA • enzymatic extension • modified dNTPs • nanomaterials • PCR

- [1] A. Czogalla, H. G. Franquelim, P. Schwillie, *Biophys. J.* **2016**, 110, 1698-1707.
 [2] Y. Ito, E. Fukusaki, *J. Mol. Catal. B: Enzym.* **2004**, 28, 155-166.

- [3] a) L. H. Yuen, R. M. Franzini, S. S. Tan, E. T. Kool, *J. Am. Chem. Soc.* **2014**, *136*, 14576-14582; b) M. Hollenstein, *Molecules* **2015**, *20*, 20777-20804.
- [4] D. C. Bell, W. K. Thomas, K. M. Murtagh, C. A. Dionne, A. C. Graham, J. E. Anderson, W. R. Glover, *Microsc. Microanal.* **2012**, *18*, 1049-1053.
- [5] X. Wang, J. Zhang, Y. Li, G. Chen, X. Wang, *Nucleic Acid Ther.* **2015**, *25*, 27-34.
- [6] L. H. Lauridsen, J. A. Rothnagel, R. N. Veedu, *Chembiochem* **2012**, *13*, 19-25.
- [7] A. S. Kamath-Loeb, A. Hizi, H. Kasai, L. A. Loeb, *J. Biol. Chem.* **1997**, *272*, 5892-5898.
- [8] A. Hottin, A. Marx, *Acc. Chem. Res.* **2016**, *49*, 418-427.
- [9] a) G. H. Clever, K. Polborn, T. A. Carell, *Angew. Chem. Int. Ed.* **2005**, *44*, 7204-7208; b) K. Tanaka, Y. Yamada, M. Shionoya, *J. Am. Chem. Soc.* **2002**, *124*, 8802-8803.
- [10] A. Shivalingam, T. Brown, *Biochem. Soc. Trans.* **2016**, *44*, 709-715.
- [11] K. Jayaraman, S. A. Fingar, J. Shah, J. Fyles, *Proc. Natl. Acad. Sci. USA*, **1991**, *88*, 4084-4088.
- [12] A. H. El-Sagheer, A. P. Sanzone, R. Gao, A. Tavassoli, T. Brown, *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 11338-11343.
- [13] Y. Zhao, F. Chen, Q. Li, L. Wang, C. Fan, *Chemical Reviews* **2015**, *115*, 12491-12545.
- [14] T. Notomi, H. Okayama, H. Masubuchi, T. Yonekawa, K. Watanabe, N. Amino, T. Hase, *Nucleic Acids Res.* **2000**, *28*, e63.
- [15] G. T. Walker, M. S. Fraiser, J. L. Schram, M. C. Little, J. G. Nadeau, D. P. Malinowski, *Nucleic Acids Res.* **1992**, *20*, 1691-1696.
- [16] A. B. Kotlyar, N. Borovok, T. Molotsky, L. Fadeev, M. Gozin, *Nucleic Acids Res.* **2005**, *33*, 525-535.
- [17] a) M. M. Ali, F. Li, Z. Zhang, K. Zhang, D.-K. Kang, J. A. Ankrum, X. C. Le, W. Zhao, *Chem. Soc. Rev.* **2014**, *43*, 3324-3341; b) A. Fire, S.-Q. Xu, *Proc. Natl. Acad. Sci. USA* **1995**, *92*, 4641-4645; c) D. Liu, S. L. Daubendiek, M. A. Zillman, K. Ryan, E. T. Kool, *J. Am. Chem. Soc.* **1996**, *118*, 1587-1594.
- [18] a) M. Hocek, *J. Org. Chem.* **2014**, *79*, 9914-9921; b) E. Fahy, R. Nazarbaghi, M. Zomorodi, C. Hernstadt, W. D. Parker, R. E. Davis, S. S. Ghosh, *Nucleic Acids Res.* **1997**, *25*, 3102-3109.
- [19] R. D. Saiki, D. H. Gelfand, S. Stoffel, S. S. Scharf, R. Higuchi, G. T. Horn, K. B. Mullis, H. A. Erlich, *Science* **1988**, *239*, 487-491.
- [20] K. Gutmiedl, D. Fazio, T. Carell, *Chem. Eur. J.* **2010**, *16*, 6877-6883.
- [21] Y. N. Teo, E. T. Kool, *Chem. Rev.* **2012**, *112*, 4221-4245.
- [22] a) D. Shin, R. W. Sinkeldam, Y. Tor, *J. Am. Chem. Soc.* **2011**, *133*, 14912-14915; b) M. Sholokh, R. Improta, M. Mori, R. Sharma, C. Kenfack, A. Shin, K. Voltz, R. H. Stote, O. A. Zaporozhets, M. Botta, Y. Tor, Y. Mely, *Angew. Chem. Int. Ed.* **2016**, *55*, 7974-7978; c) M. Wilhelmsson, *Quarterly Review of Biophysics* **2010**, *43*, 159-183; d) M. Suchy, R. H. E. Hudson, *J. Org. Chem.* **2014**, *79*, 3326-3347; e) H. – A. Wagenknecht, *Ann. N.Y. Acad. Sci.* **2008**, *1130*, 122-130; f) D. Jiang, F. Seela, *J. Am. Chem. Soc.* **2010**, *132*, 4016-4024; g) S. K. Jarchow-Choy, A. T. Krueger, H. Liu, J. Gao, E. T. Kool, *Nucleic Acids Res.* **2011**, *39*, 1586-1594; h) Y. Saito, A. Suzuki, Y. Okada, Y. Yamasaka, N. Nemoto, I. Saito, *Chem. Commun.* **2013**, *49*, 5684-5686; i) A. A. Tanpure, S. G. Srivatsan, *Nucleic Acids Res.* **2015**, *43*, e149; j) A. Matarazzo, R. H. E. Hudson, *Tetrahedron* **2015**, *71*, 1627-1657; k) G. Mata, O. P. Schmidt, N. W. Luedtke, *Chem. Commun.* **2016**, *52*, 4718-4721; l) H. Mei, C. Shi, R. M. Jimenez, Y. Wang, M. Kardouh, C. Chaput, *Nucleic Acids Res.* **2017**, *45*, 5629-5638; m) D. Dziuba, P. Jurkiewicz, M. Cebecauer, M. Hof, M. Hocek, *Angew. Chem. Int. Ed.* **2015**, *1*, 174-178.
- [23] a) M. Vrabel, P. Horakova, H. Pivonkova, L. Kalachova, H. Cernocka, H. Cahova, R. Pohl, P. Sebest, L. Havran, M. Hocek, M. Fojta, *Chem. Eur. J.* **2009**, *15*, 1144-1154; b) M. Hocek, M. Fojta, *Chem. Soc. Rev.* **2011**, *40*, 5802-5814.
- [24] E. Eremeeva, M. Abramov, L. Margamuljana, J. Rozenski, V. Pezo, P. Marlière, P. Herdewijn, *Angew. Chem. Int. Ed.* **2016**, *55*, 7515-7519.
- [25] a) H. Liu, J. Gao, S. R. Lynch, Y. D. Saito, L. Maynard, E. T. Kool, *Science* **2003**, *302*, 868-871; b) R. Yamashige, M. Kimoto, Y. Takezawa, A. Sato, T. Mitsui, S. Yokoyama, I. Hirao, *Nucleic Acids Res.* **2011**, *40*, 2793-2806; c) L. Zhang, Z. Yang, K. Sefah, K. M. Bradley, S. Hoshika, M.-J. Kim, H.-J. Kim, G. Zhu, E. Jiménez, S. Cansiz, I.-T. Teng, C. Champanhac, C. McLendon, C. Liu, W. Zhang, D. L. Gerloff, Z. Huang, W. Tan, S. A. Benner, *J. Am. Chem. Soc.* **2015**, *137*, 6734-6737; d) D. A. Malyshev, F. E. Romesberg, *Angew. Chem. Int. Ed.* **2015**, *54*, 11930-11944; e) N. Tarashima, Y. Komatsu, K. Furukawa, N. Minakawa, *Chemistry* **2015**, *21*, 10688-10695; f) Y. Zhang, B. M. Lamb, A. W. Feldman, A. X. Zhou, T. Lavergne, L. Li, F. E. Romesberg, *Proc. Natl. Acad. Sci. USA* **2017**, *114*, 1317-1322.
- [26] X. Ren, A. H. El-Sagheer, T. Brown, *Nucl. Acids Res.* **2016**, *44*, e79.
- [27] J. Gierlich, K. Gutmiedl, P. M. Gramlich, A. Schmidt, G. A. Burley, T. Carell, *Chem. Eur. J.* **2007**, *13*, 9486-9494.
- [28] a) S. Jäger, G. Rasched, H. Kornreich-Leshem, M. Engeser, O. Thum, M. Famulok, *J. Am. Chem. Soc.* **2005**, *127*, 15071-15082; b) P. Capek, H. Cahová, R. Pohl, M. Hocek, C. Gloeckner, A. Marx, *Chem. Eur. J.* **2007**, *13*, 6196-6203.
- [29] M. Jama, A. Millson, C. E. Miller, E. Lyon, *J. Mol. Diagnostics* **2013**, *15*, 255-262.
- [30] H. Mitomo, Y. Watanabe, Y. Matsuo, K. Niikura, K. Ijiro, *Chem. Asian J.* **2015**, *10*, 455-460.
- [31] S. M. D. Watson, A. R. Pike, J. Pate, A. Houlton, B. R. Horrocks, *Nanoscale* **2014**, *6*, 4027-4037.
- [32] S. A. F. Al-Said, R. Hassanien, J. Hannant, M. A. Galindo, S. Pruneanu, A. R. Pike, A. Houlton, B. R. Horrocks, *Electrochem. Commun.* **2009**, *11*, 550-552.
- [33] J. Hannant, J. H. Hedley, J. Pate, A. Walli, S. A. F. Al-Said, M. A. Galindo, B. A. Connolly, B. R. Horrocks, A. Houlton, A. R. Pike, *Chem. Comm.* **2010**, *46*, 5870-5872.
- [34] a) A. R. Morgan, M. B. Coulter, W. F. Flintoff, V. H. Paetkau, *Biochemistry* **1974**, *13*, 1596-603; b) H. K. Schachman, J. Adler, C. M. Radding, I. R. Lehman, A. Kornberg, *J. Biol. Chem.* **1960**, *235*, 3242-3249.
- [35] J. S. Hartig, E. T. Kool, *Nucleic Acids Res.* **2005**, *33*, 4922-4927.
- [36] C. J. Whitfield, A. T. Turley, E. M. Tuite, B. A. Connolly, A. R. Pike, *Angew. Chem. Int. Ed.* **2015**, *54*, 8971-8974.
- [37] a) S. K. Jozwiakowski, B. A. Connolly, *ChemBioChem* **2011**, *12*, 35-37; b) B. D. Biles, B. A. Connolly, *Nucleic Acids Res.* **2004**, *32*, e176.
- [38] H. W. Jannasch, C. O. Wirsen, S. J. Molyneaux, T. A. Langworthy, *Appl. Environ. Microbiol.* **1992**, *58*, 3472-3481.
- [39] R. M. Dale, D. C. Livingston, D. C. Ward, *Proc. Natl. Acad. Sci. U.S.A.* **1973**, *70*, 2238-2242.

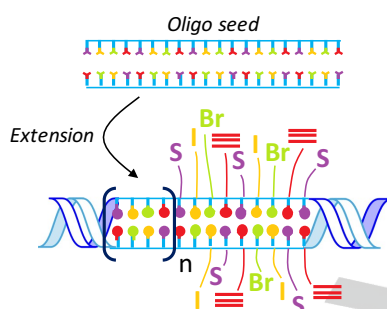
- [40] Y. Ling, J. A. Nelson, Y. C. Cheng, R. S. Anderson, K. L. Beattie, *Mol. Pharmacol.* **1991**, *40*, 508-514.
- [41] E. Dubler, *Metal Ions in Biological Systems, Vol. 32*, (Eds.: A. Sigel, H. Sigel), Marcel Dekker, NY, **1996**, pp. 301-338.
- [42] Z. Wang, T. M. Rana, *Biochemistry* **1998** *37*, 4235-4243.
- [43] a) K. H. Shaughnessy, *Molecules* **2015**, *20*, 9419-9454; b) B. H. Lipshutz, D. W. Chung, B. Rich, *Org. Lett.* **2008**, *10*, 3793-3796; c) L. Lercher, J. F. McGouran, B. M. Kessler, C. J. Schofield, B. G. Davis, *Angew. Chem. Int. Ed.* **2013**, *52*, 10553-10558; d) J. Dadová, P. Vidláková, R. Pohl, L. Havran, M. Fojta, M. Hocek, *J. Org. Chem.* **2013**, *78*, 9627-9637; e) G. Hervé, C. Len, *RSC Adv.* **2014**, *4*, 46926-46929;
- f) G. Hervé, G. Sartori, G. Enderlin, G. Mackenzie, C. Len, *RSC Adv.* **2014**, *4*, 18558-18594.
- [44] E. Defrancq, S. Messaoudi, *ChemBioChem* **2017**, *18*, 426-431.
- [45] H. G. Hansma, I. Revenko, K. Kim, D. E. Laney, *Nucleic Acids Res.* **1996**, *24*, 713-720.
- [46] G. I. Livshits, A. Stern, D. Rotem, N. Borovok, G. Eidelstein, A. Migliore, E. Penzo, S. J. Wind, R. Di Felice, S. S. Skourtis, J. C. Cuevas, L. Gurevich, A. B. Kotlyar, D. Porath, *Nature Nanotech.* **2014**, *9*, 1040-1046.

Entry for the Table of Contents (Please choose one layout)

Layout 1:

FULL PAPER**Designer DNA**

An enzymatic synthesis method able to produce long DNA decorated with multiple functional groups at controllable positions along the structure. DNA sequences of any repeating unit can be extended and incorporate modified dNTPs to produce DNA products from 200 to 5000 bp with potential in 1D nanomaterial applications.



Colette J. Whitfield, Rachel, C. Little,
Kasid Khan, Kuniharu Ijiro, Bernard A.
Connolly, Eimer M. Tuite,* and Andrew
R. Pike*

Page No. – Page No.

Self-Priming Enzymatic Fabrication of
Multiply Modified DNA