ChemPubSoc Europe

DOI: 10.1002/chem.201801976

CHEMISTRY A European Journal Full Paper

Biochemistry

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 multiply modified DNA by the extension of repeating unit duplex "oligoseeds" provides a source of versatile DNA. Sterically-demanding nucleotides 5-Br-dUTP, 7-deaza-7-I-dATP, 6-S-dGTP, 5-I-dCTP as well as 5-(octadiynyl)-dCTP were incorporated into two extending oligoseeds; $[GATC]_5/[GATC]_5$ and $[A_4G]_4/[CT_4]_4$. The products contained modifications on one or both strands of DNA, demonstrating their recognition by the polymerase as both template (reading) and substrate (writing). Nucleobase modifications that lie in the major groove were 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 was produced by this method. The number, position and type of modification, as well as 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.

Introduction

DNA that bears 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 are reactive with nucleophiles 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, and some modified dNTPs are even better substrates for polymerases than the natural dNTPs.[13] Polymerase methods to copy DNA comprise isothermal amplification,[14] including loop-mediated amplification (LAMP),[15] strand-displacement,[16] slippage,[17] rolling circle amplification (RCA),[18] and primer-template extension,[19] as well as thermal cycling amplification such as the polymerase chain reaction (PCR).[20] Methods such as primer-extension can introduce modified bases on the growing strand,[21] whilst PCR can introduce different modifications on both strands as long as the modified nucleotides can be both recognized and inserted by polymerases.[19] 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,[22] fluorescent,[23] and electroactive[24] base analogues; non-canonical bases;[25] and highly specific artificial base-pairs.[26] Alternatively, bases with reactive moieties such as azide,[27] alkyne,[28] and amino[29] groups attached by linkers can be incorporated into 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

[a] Dr. C. J. Whitfield, R. C. Little, K. Khan, Dr. E. M. Tuite, Dr. A. R. Pike
Chemistry-School of Natural and Environmental Sciences
Newcastle University, Newcastle upon Tyne, NE1 7RU (UK)
E-mail: eimer.tuite@newcastle.ac.uk
andrew.pike@newcastle.ac.uk

[b] Prof. Dr. K. Ijiro
Research Institute for Electronic Science
Hokkaido University, Sapporo, 001-0021 (Japan)

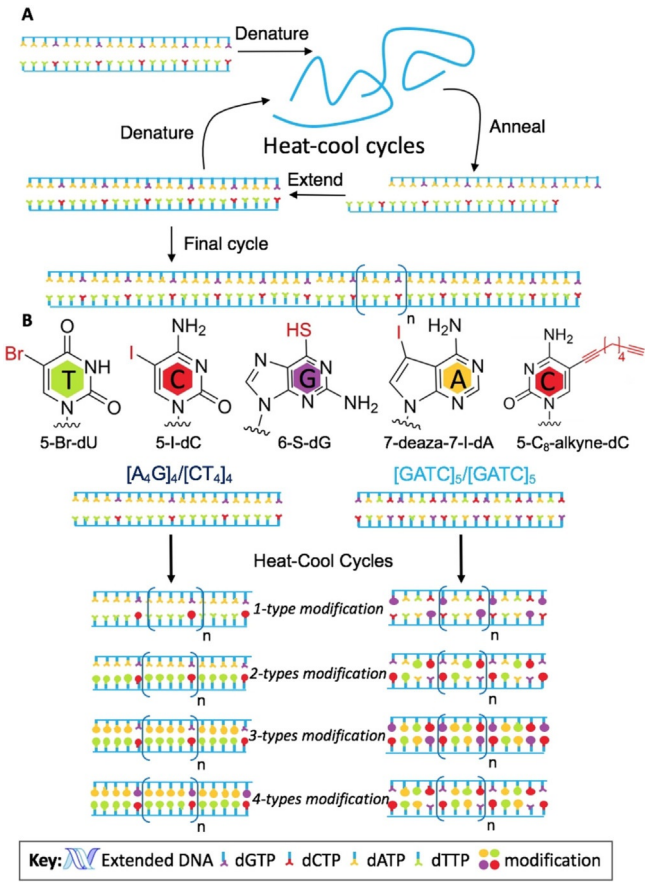
[c] Prof. Dr. B. A. Connolly
Institute for Cell and Molecular Biosciences
Newcastle University, Newcastle upon Tyne NE2 4HH (UK)

Supporting information and the ORCID identification number(s) for the author(s) of this article can be found under: https://doi.org/10.1002/chem.201801976.

repeat sequence DNA unobtainable by automated solid-phase synthesis.

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.[30] Production of high copy number repeat sequences is also important in the realm of nanomaterials applications of DNA, where the polynucleotide acts as a template for the 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.[31] The DNA strand itself can then be used as a template to guide the deposition of metal ions to synthesise 1D nanomaterials,[32] for example, Ag-DNA wires.[33] 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.[34] 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.[17,35] Although isothermal slippage has been shown to occur for 6–10 bp repeats such as telomeric sequences,[36] we have found that intermediate sized repeat sequences (4–40 nucleotides) are most efficiently extended by thermal cycling methods.[37] 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 the 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,[38a] and the *Pyrococcus* species GB-D DNA polymerase, Deep Vent *exo*-.[38b] 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,[37] modified DNA with designed repeating units that include handles for further functionalization are synthesized. The heat–cool cycle method extends DNA by relying on imperfect hybridisation within the repetitive sequence of the initial unmodified oligoseed of around 20 bases (Scheme 1 A). 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 (through design of the initial sequence of the oligoseeds) and governs both the type of base modifications (choice of NTPs included in the reaction) and the total number of bases (through the number of heat–cool cycles performed) that are designed into the final DNA products. This kind of control over the type and number of modified bases in synthetic DNA would create possibilities for designer DNA-based materials, and is therefore a valuable synthetic challenge. It is envisaged that the multiply modified DNA presented here, through a range of DNA-compatible chemistries depending on the type of modified dNTP inserted into the growing DNA strands is one route to realising this goal.



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-(octadiynyl)-dCTP, left to right), the two starting oligoseeds (middle) and examples of the possible extended modified designer DNA products (bottom).

Results and Discussion

The modified nucleotides used in this study were chosen to contain sterically demanding groups attached directly to the

ring system as well as anchor points for postsynthetic 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 modified cytosine (5-(octadiynyl)-dCTP); structures of these modified nucleotides are shown in Scheme 1 B. Some comparative studies were also performed with the heavy atom modified cytosine, 5-acetyl-Hg-dCTP,[39] which, like the iodo- and bromo-derivatives, has potential in next generation sequencing using electron microscopy. In previous reports, several or all bases were replaced with modified nucleotides using PCR on a long template.[12,21] In these instances, the modifications were located away from the nucleobase through either alkene or alkyne linkers attached to the 5-pyrimidine, 7-deaza- or 8-purine positions and so were not intimately involved in the base recognition sites.[12,21] In this work, several bulky groups were directly attached to or were included in the core nucleobase structure, which introduces steric constraints. In the case of 6-S-dGTP, an oxygen atom has been replaced by the bulkier thiol-group, which lies in the H-bonding region, and is therefore likely to interfere with base-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 coordination polymers with metal ions,[41,42] and also has photo-cross-linking potential.[43] The halide substitutions on the bases are suitable chemical handles for the C-C linkage of additional functional groups through Pd-catalysed Sonogashira, Suzuki, and Heck reactions, which have been adapted for aqueous media[44] and recently exploited for the labelling of DNA.[45]

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 $[A_4G]_4/[CT_4]_4$[1] were chosen to demonstrate: i) the inclusion of one type of modification into one strand of a repeating nucleotide sequence ($[A_4G]_4/[CT_4]_4$), 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 $[A_4G]_4/[CT_4]_4$. To synthesise modified DNA, the standard nucleotide is simply replaced with its modified derivative, that is, swapping out dGTP for 6-S-dGTP during a set number of heat–cool cycles of the polymerase reaction. Scheme 1 B, summarises some of the possible DNA sequences that can be produced from the two oligoseeds and five modified nucleotides, that is, designer DNA with sites for post-synthetic functionalization decorated along the duplex.

Initially, we show that by replacing just 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 $[A_4G]_4/[CT_4]_4$ oligoseed is produced. In this case, the modified nucleotide must act as an efficient polymerase substrate, that is, be incorporated correctly against its canonical complementary base, but it also must be read correctly for incorporation of its unmodified complement to allow extension to continue along both strands. In the second case, we discuss incorporation of a single type of modified nucleobase into both DNA strands using the self-complementary oligoseed $[GATC]_5/[GATC]_5$.

In this case again, the modified nucleotide must be recognised correctly by the polymerase as a substrate for insertion and also be read whilst in the growing template. Finally, using both types of oligoseed we demonstrate that the described method can be used to synthesise fully modified DNA; that is, several modified bases within a user-defined sequence, on one or both strands to yield DNA products >500 bp in length. The results shown are those obtained using Tgo-Pol-Z3 *exo-* polymerase, however Deep Vent *exo-* works equally well (Figure S1).

Firstly, the oligoseed $[A_4G]_4/[CT_4]_4$ was chosen to illustrate how either a modified G or C, could be multiply incorporated into a repeat sequence at a fixed base-separation, that is, at every fifth site within just one strand of the extending duplex. Alternatively, by using a modified A or T in the extension reaction, multiple modifications can be located adjacent to each other over many repeating units of DNA, that is, four contiguous modifications on one strand. First, we consider the incorporation of a single type of C5 modified dCTP. 10 heat–cool cycles of extension using $[A_4G]_4/[CT_4]_4$ and either 5-I-dCTP or 5-(octadiynyl)-dCTP yields highly extended DNA of 8 kb and 5 kb, respectively (see gel electrophoresis images, Figure S2 A and S2 B). The DNA products contain the modified cytidine at every fifth site on just 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.[37] 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 around 500 bp which could not be further extended with additional cycling (Figure S2 E and Figure S3). Employing the same starting oligoseed but inserting either a modified dATP or dTTP 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 S2 C and S2 D) and further cycling allows attainment of lengths of modified DNA comparable to that obtained with the native nucleotides.

To confirm that the DNA products 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 S4 B). UV/Vis (Figure S4 C and Table S1) and HR-MS (Table S2) of the isolated fractions demonstrated that each modified base was included in the sequence as designed. Sanger sequencing of the above products confirmed the integrity of the repeating sequence (see boxes within Figure 1 and Figure S4 B). The

[1] *Oligoseeds are designated as [5′-3′]/[5′-3′].*

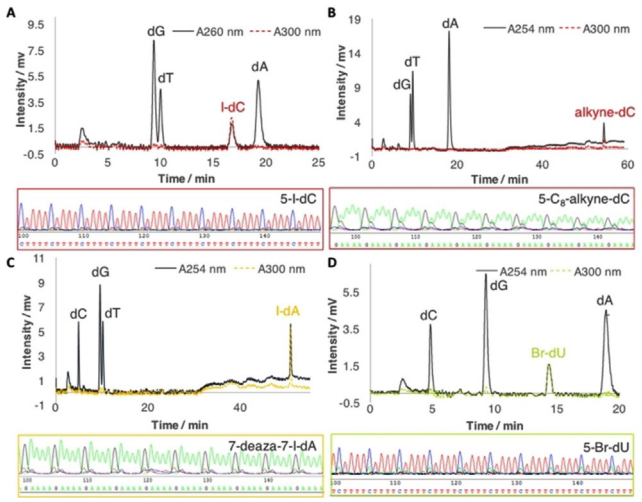


Figure 1. HPLC traces of the DNA components after digestion and the Sanger sequencing data of the extended products from oligoseed $[A_4G]_4$/$[CT_4]_4$ bearing: (A) 5-I-dC, (B) 5-C_8-alkyne-dC (C) 7-deaza-7-I-dA and (D) 5-Br-dU.

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 to conclude in accurate sequencing results. Additionally, the analysis of the synthesized DNA by enzymatic digestion and HPLC was performed on the 10 kb product, therefore we are confident that polymerase fidelity is not diminished as the "template" strand is increased.

Exemplary atomic force microscopy (AFM) images of extended $[A_4G]_n$/$[CT_4]_n$ products (30 heat cool cycles) bearing single base modifications are shown in Figure 2 (F–J; further analysis is provided in Figure S5 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.[46] It is noticeable that the expected linear morphology of DNA is recognizable, however some minor structural variations between the different extended products can be observed.

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 both strands of the final duplex, placing the modifications at adjacent base-pair sites. After 10 heat–cool cycles, extension products of over 2 kb were produced for reactions involving any one of the following four modified nucleobases, 5-I-dCTP, 5-Br-dUTP, 7-deaza-7-I-dATP or 5-(octadiynyl)-dCTP (Figure S2 F–I). 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

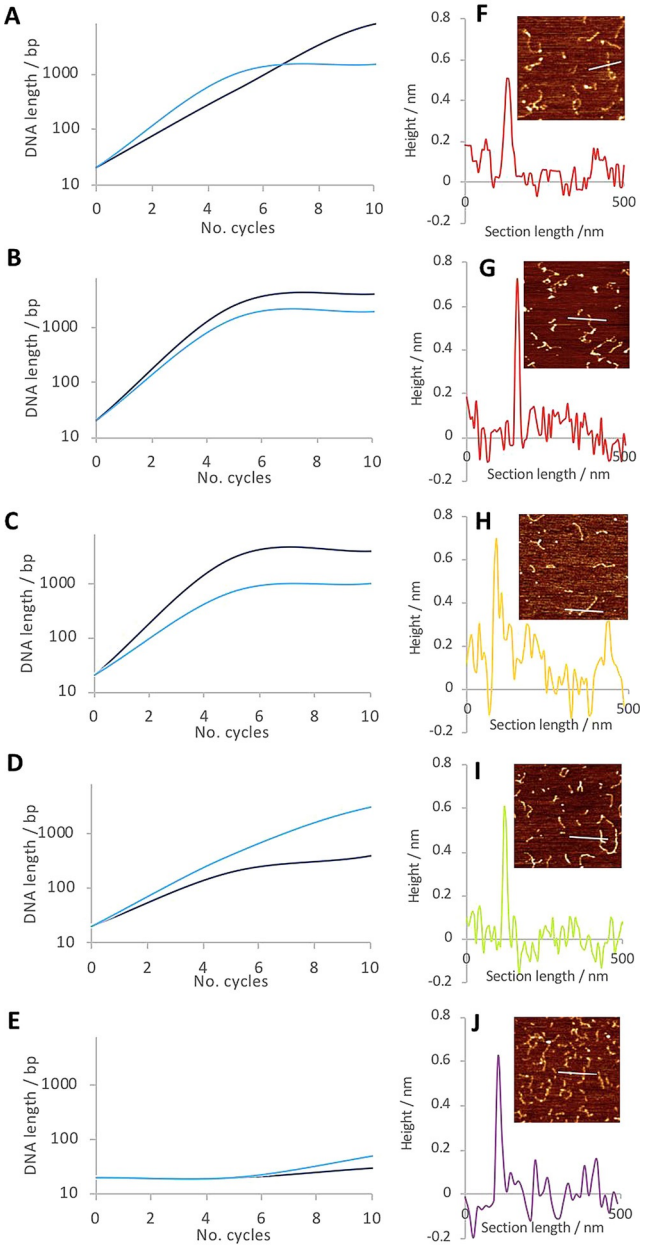


Figure 2. Plots A-E compare the lengths of extension products by number of heat–cool cycles, indicating the relative incorporation rates for the different modified bases with both oligoseeds, $[A_4G]_4$/$[CT_4]_4$ (dark blue line) and $[GATC]_5$/$[GATC]_5$ (light blue line), where A = 5-iodo-dC, B = 5-C8-alkyne-dC, C = 7-deaza-7-I-dA, D = 5-Br-dU and E = 6-S-dG. Details of rate evaluation are provided in the Supplementary Information. F-J shows the AFM images on mica of the corresponding DNA products (A–E) from the extension of $[A_4G]_4$/$[CT_4]_4$ after 30 cycles. Height scale is 2.1 nm, scan size is 1 μm, and profile trace lines are 500 nm in each case.

each strand but is not self-complementary, $[GAATC]_4$/$[GATTC]_4$, was used to confirm enzymatic fidelity, see Figure S6.

The gel-electrophoresis data from both extension reactions, (Figure S2) indicated that the final DNA products consisted of a range of lengths, from 200 bp to 5000 bp, as expected from our previous report on this method using unmodified NTPs.[37] Further analysis of the gel band intensities, (Figure 2 A–E, and Figure S7) indicated that in most cases there is minimal differ-

ence in DNA product lengths after 5 cycles upon incorporation of the modified bases by the two types of oligoseed. However, the most marked difference occurred in the case where multiple adjacent, 5-Br-dUTP units are incorporated into the $[A_4G]_4$/$[CT_4]_4$ oligoseed, resulting in a noticeable reduction in the length of the extension products after 10 cycles, Figure 2D. This suggests that, the size and electronegativity of bromine may have steric and/or electronic impacts on the polymerase and the higher charge density and bulk of several neighbouring bromo-modifications might affect the polymerase reading and/or writing processes. Otherwise we found little difference in the enzymatic processivity of different types of base modification within the same oligoseed, namely between; i) a single atom or a long alkyl chain modification on the same base, Figure 2A,B, (cf. 5-I-dCTP and 5-(octadiynyl)-dCTP) or ii) different nucleobases bearing the same single atom displacement, Figure 2A,C (cf. 5-I-dCTP and 7-deaza-7-I-dATP). In these instances, the nucleobase modifications are at positions C5 and C7 and do not lie within the hydrogen bonding region but project into the major groove. Extension was also observed with 6-S-dGTP (see Figure S2E and S2J for gel electrophoretic data) although with much shorter extension products for both oligoseeds (Figure 2D) yielding only 200 bp lengths after 10 cycles. This observation can be rationalized by considering the 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.

To ensure that the polymerase retains its fidelity, that is, it accurately writes-in 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 miswrite into the sequence an alternative base when the correct base was absent (see Figure S8). Therefore, we conclude that base misincorporation does not occur when an artificial and potentially "hard-to-write" modified dNTP is used in the reaction.

To explore whether polymerases could incorporate modified bases into longer stretches of DNA, the challenging substrate 6-S-dGTP was incorporated against cytosine in the extension of a $[G]_{20}$/$[C]_{20}$ seed (Figure S9). A poly(6-S-dG).poly(dC) product of ca. 1 kb was synthesized, consistent with the truncated lengths observed in the extension of oligoseeds with 6-S-G incorporated at every fourth or fifth base, Figure S2E and Figure S2J. We note that with 6-S-dGTP, the thiol-modification limits the final extension length as well as slowing incorporation. But for other modified dNTPs kinetics alone is limited, as seen with 5-Br-dUTP (Figure S2D, & S2I) where sequences of equal length to products using unmodified bases are produced but only after additional cycling. Confident of the versatility of the method, that is, 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 and the polymerase maintained the correct repeat sequence giving products over 2000 bp after 30 cycles (see Figure S10 for sequencing data). In a second extension reaction, the alkyl-modified nucleotide, 5-C_8-alkyne-dCTP, and the halogenated nucleotide, 5-Br-dUTP, were both successfully incorporated into the extending DNA, (Figure S11), producing DNA products of over 1 kb after 30 heat–cool cycles from both types of oligoseed, (Figure S1, Lane 7). The extension of both oligoseeds to include three modified nucleotides, either 5-(octadiynyl)-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,C, respectively), demonstrated that three modifications can be written into the repeating sequence to yield triply-modified DNA products, <1 kb. Final DNA lengths are notably shorter when 6-S-dGTP is incorporated, as previously noted.

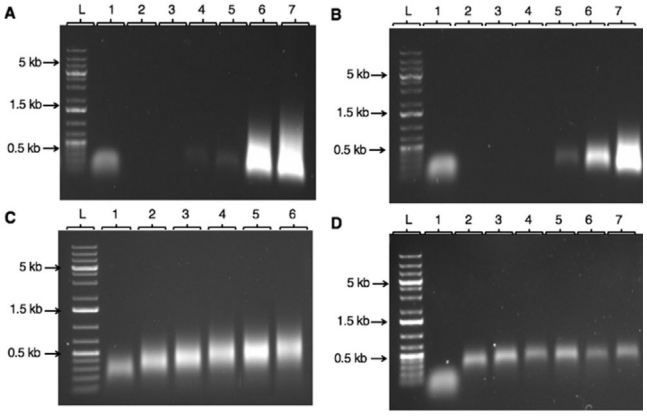


Figure 3. Image of agarose gels of the DNA extension products from oligoseed $[A_4G]_4$/$[CT_4]_4$ (A and B) and oligoseed $[GATC]_5$/$[GATC]_5$ (C and D). Triply modified DNA was fabricated using 6-S-dG, 5-Br-dU and 7-deaza-7-I-dA (A and C) and quadruply modified DNA from 6-S-dG, 5-Br-dU, 7-deaza-7-I-dA and 5-C_8-alkyne-dC (B and D). Lanes 1 (A, B, D) represent products after 0 cycles, lanes 2–7 (A, B, D) and lanes 1–6 (C) represent products after 5, 10, 15, 20, 25 and 30 cycles, respectively, lane L is a 1 kb + DNA ladder.

Finally, the possibility to design DNA bearing four different modifications was demonstrated, again using both oligoseed sequences. 6-S-dGTP, 5-Br-dUTP, 7-I-dATP and 5-(octadiynyl)-dCTP were incorporated over 30 heat–cool cycles starting with either type of oligoseed (Table 1). In both cases DNA products of up to 500 bp were achieved, demonstrating the synthesis of fully-modified DNA through this approach (Figure 3B,D). These products are considerably longer than what is routinely feasible through automated DNA synthesis and highlights the importance of this repetitive self-priming enzymatic approach to the reading and writing of artificial nucleotides to yield multiply modified DNA.

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

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

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

The synthetic method described in this work is versatile and through careful design of the oligoseeds and the choice of functionally-modified nucleotides, extended DNA of a specified repeating sequence bearing multiple sterically-demanding base modifications can be synthesised. Because this self-priming method does not rely on the availability of a long template strand (either a natural sequence or one constructed by ligation of oligonucleotides), extended DNA sequences can be synthesized, from readily available short oligoseeds. 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, varying in size from 200 bp to 5000 bp, this is due to the range of overlap lengths possible during each heat–cool cycle. Undoubtedly, an ability to produce monodisperse DNA would be of great benefit. Size selection can be simply achieved by electrophoretic separation and sample extraction using a LonzaGel double-well system to select the desired DNA lengths.[37] For example, 250, 450, 800 and 1200 bp samples were extracted from the dispersion of 5-I-dC bearing products, after the extension of the $[A_4G]_4/[CT_4]_4$ oligoseed, see gel image in Figure 4A. The yield of the size-separated DNA products in Lanes 1–4 was quantified by UV/Vis spectroscopy, see Figure 4B.

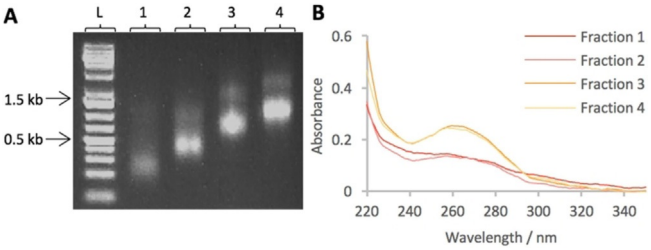


Figure 4. Size separation of the dispersion of I-dC-modified DNA extension products after 20 heat–cool cycles from the starting oligoseed $[A_4G]_4/[CT_4]_4$. A) Agarose gel of size separated fractions, 1–4. B) UV/Vis of the fractions 1–4, corresponding to Lanes 1–4 in A.

Yields of 110, 109, 206 and 199% for each 250, 450, 800 and 1200 bp fraction, respectively were obtained (yields are given as a percentage of the starting oligoseed concentration of 0.5 μM, hence yields over 100% are expected). Therefore, this approach both fabricates and amplifies modified DNA.

Conclusions

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-(octadiynyl)-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 functional 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, 1D 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 annealing buffer (10 mM HEPES, 100 mM NaCl and 1 mM EDTA) and allowing to cool at 1 °C per minutes to room temperature. Full DNA sequences can be found in the Supplementary Methods.

DNA polymerases

Tgo-Pol-Z3 *exo-* was prepared and purified as described previously.[38a] Deep Vent *exo-* was purchased from NEB.

DNA extension

Oligoseed (0.5 μM) (see the Supporting Information 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 $(NH_4)_2SO_4$, 100 mM

KCl, 1% Triton X-100, 1 mg mL^{-1} Bovine Serum Albumin and 20 mM $MgSO_4$) or 10×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 using the following programmed method:

Number of cycles stated, where one cycle = 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 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".

Conflict of interest

The authors declare no conflict of interest.

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

[1] A. Czogalla, H. G. Franquelim, P. Schwille, *Biophys. J.* **2016**, *110*, 1698 – 1707.
[2] Y. Ito, E. Fukusaki, *J. Mol. Catal. B* **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; *Angew. Chem.* **2005**, *117*, 7370 – 7374; 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] a) P. Kielkowski, J. Fanfrlík, M. Hocek, *Angew. Chem. Int. Ed.* **2014**, *53*, 7552 – 7555; *Angew. Chem.* **2014**, *126*, 7682 – 7685; b) H. Cahová, A. Panattoni, P. Kielkowski, J. Fanfrlík, M. Hocek, *ACS Chem. Biol.* **2016**, *11*, 3165 – 3171.
[14] Y. Zhao, F. Chen, Q. Li, L. Wang, C. Fan, *Chem. Rev.* **2015**, *115*, 12491 – 12545.
[15] T. Notomi, H. Okayama, H. Masubuchi, T. Yonekawa, K. Watanabe, N. Amino, T. Hase, *Nucleic Acids Res.* **2000**, *28*, e63.
[16] 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.
[17] A. B. Kotlyar, N. Borovok, T. Molotsky, L. Fadeev, M. Gozin, *Nucleic Acids Res.* **2005**, *33*, 525 – 535.
[18] 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.
[19] a) M. Hocek, *J. Org. Chem.* **2014**, *79*, 9914 – 9921; b) E. Fahy, R. Nazarbaghi, M. Zomorrodi, C. Herrnstadt, W. D. Parker, R. E. Davis, S. S. Ghosh, *Nucleic Acids Res.* **1997**, *25*, 3102 – 3109.
[20] 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.
[21] K. Gutsmiedl, D. Fazio, T. Carell, *Chem. Eur. J.* **2010**, *16*, 6877 – 6883.
[22] Y. N. Teo, E. T. Kool, *Chem. Rev.* **2012**, *112*, 4221 – 4245.
[23] 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; *Angew. Chem.* **2016**, *128*, 8106 – 8110; c) M. Wilhelmsson, *Q. Rev. Biophys.* **2010**, *43*, 159 – 183; d) M. Suchý, 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. Jiminez, 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.* **2016**, *55*, 174 – 178; *Angew. Chem.* **2016**, *128*, 182 – 186.
[24] a) M. Vrábel, P. Horáková, H. Pivonkova, L. Kalachova, H. Cernocka, H. Cahová, 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.
[25] E. Eremeeva, M. Abramov, L. Margamuljana, J. Rozenski, V. Pezo, P. Marlière, P. Herdewijn, *Angew. Chem. Int. Ed.* **2016**, *55*, 7515 – 7519; *Angew. Chem.* **2016**, *128*, 7641 – 7645.
[26] 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.* **2012**, *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; *Angew. Chem.* **2015**, *127*, 12098 – 12113; e) N. Tarashima, Y. Komatsu, K. Furukawa, N. Minakawa, *Chem. Eur. J.* **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.
[27] X. Ren, A. H. El-Sagheer, T. Brown, *Nucleic Acids Res.* **2016**, *44*, e79.
[28] J. Gierlich, K. Gutsmiedl, P. M. Gramlich, A. Schmidt, G. A. Burley, T. Carell, *Chem. Eur. J.* **2007**, *13*, 9486 – 9494.
[29] 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.
[30] M. Jama, A. Millson, C. E. Miller, E. Lyon, *J. Mol. Diagn.* **2013**, *15*, 255 – 262.
[31] H. Mitomo, Y. Watanabe, Y. Matsuo, K. Niikura, K. Ijiro, *Chem. Asian J.* **2015**, *10*, 455 – 460.
[32] S. M. D. Watson, A. R. Pike, J. Pate, A. Houlton, B. R. Horrocks, *Nanoscale* **2014**, *6*, 4027 – 4037.
[33] 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.
[34] 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. Commun.* **2010**, *46*, 5870 – 5872.

[35] a) A. R. Morgan, M. B. Coulter, W. F. Flintoff, V. H. Paetkau, *Biochemistry* **1974**, *13*, 1596 – 1603; b) H. K. Schachman, J. Adler, C. M. Radding, I. R. Lehman, A. Kornberg, *J. Biol. Chem.* **1960**, *235*, 3242 – 3249.
[36] J. S. Hartig, E. T. Kool, *Nucleic Acids Res.* **2005**, *33*, 4922 – 4927.
[37] a) C. J. Whitfield, A. T. Turley, E. M. Tuite, B. A. Connolly, A. R. Pike, *Angew. Chem. Int. Ed.* **2015**, *54*, 8971 – 8974; *Angew. Chem.* **2015**, *127*, 9099 – 9102.
[38] a) B. D. Biles, B. A. Connolly, *Nucleic Acids Res.* **2004**, *32*, e176; b) 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. USA* **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, New York, **1996**, pp. 301 – 338.
[42] S. M. L. Lunn, S. Hribesh, C. J. Whitfield, M. J. Hall, A. Houlton, A. K. Bronowska, E. M. Tuite, A. R. Pike, *ChemBioChem* **2018**, *19*, 1115 – 1118.
[43] Z. Wang, T. M. Rana, *Biochemistry* **1998**, *37*, 4235 – 4243.
[44] 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; *Angew. Chem.* **2013**, *125*, 10747 – 10752; 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.
[45] E. Defrancq, S. Messaoudi, *ChemBioChem* **2017**, *18*, 426 – 431.
[46] H. G. Hansma, I. Revenko, K. Kim, D. E. Laney, *Nucleic Acids Res.* **1996**, *24*, 713 – 720.

Manuscript received: April 20, 2018
Revised manuscript received: June 19, 2018
Accepted manuscript online: June 21, 2018
Version of record online: September 14, 2018