Supporting Information

Self-Priming Enzymatic Fabrication of Multiply Modified DNA

Colette J. Whitfield[1], Rachel C. Little[1], Kasid Khan[1], Kuniharu Ijiro[2], Bernard A. Connolly[3], Eimer M. Tuite[1] and Andrew R. Pike[1]

[1] School of Natural and Environmental Sciences, Newcastle University, Newcastle upon Tyne, NE1 7RU
[2] Research Institute for Electronic Science, Hokkaido University, Sapporo, Japan
[3] Institute of Cellular and Molecular Biosciences, Newcastle University, Newcastle upon Tyne, NE2 4HH

Table of contents

Supporting Methods

Deoxyoligonucleotides

Deoxyoligonucleotides used to form the oligoseeds are outlined in Table 1.

Oligos to be combined	Oligoseed	Oligoseed length / bp
$(GATC)_5$ and $(GATC)_5$	$[GATC]_5/[GATC]_5$	20
$(AAAAG)_4$ and $(CTTTT)_4$	$[AAAAG]_4/[CTTTT]_4$	20
$(GATTC)_4$ and $(GAATC)_4$	$[GATTC]_4/[GAATC]_4$	20
$(G)_{20}$ and $(C)_{20}$	$[G]_{20}/[C]_{20}$	20

Table 1 Summary of oligos required to form the oligoseeds used.

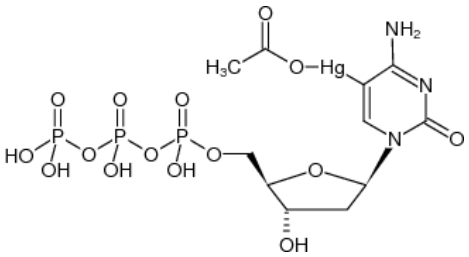
Deoxynucleotidetriphosphate combinations

Deoxynucleotidetriphosphate combinations required to synthesise each modified DNA extension product is outlines in Table 2.

Modified DNA	dNTPs present
I-dC-DNA	dGTP, 5-I-dCTP, dATP, dTTP
I-dA-DNA	dGTP, 7-deaza-7-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, 5-(octadiynyl)-dCTP, dATP, dTTP
S-dG-DNA	6-S-dGTP, dCTP, dATP, dTTP
I-dC, Br-dU-DNA	dGTP, 5-I-dCTP, dATP, 5-Br-dTTP
Alkyne-dC, Br-dU-DNA	dGTP, 5-(octadiynyl)-dCTP, dATP, 5-Br-dUTP
Alkyne-dC, Br-dU, I-dA-DNA	dGTP, 5-(octadiynyl)-dCTP, 7-deaza-7-I-dATP, 5-Br-dUTP
Alkyne-dC, Br-dU, S-dG, I-dA-DNA	6-S-dGTP, 5-(octadiynyl)-dCTP, 7-deaza-7-I-dATP, 5-Br-dUTP

Table 2 dNTPs present in each modified DNA reaction mixture.

Structure of 5-acetyl-Hg-dCTP



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 5 μg/ μL 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 provided by Thermo Scientific (Paisley, UK) for volumes less than 2 μL and a Varian Cary 100Bio UV‑Vis spectrophotometer with a Varian Cary temperature controller provided by Agilent Technologies (Stockport, UK) for larger volumes. Spectrometer was blanked using nanopure-H_2O or QIAGEN elusion buffer. Concentration and peak determination was performed using the Beer-Lambert law (Table S1).

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 dsDNA and incubated at 37 °C for 16‑18 hours. The dNs produced can be analysed by reverse phase HPLC on an APEX ODS C18 5um 250 mm column Waters (Herts, UK) 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 4.

Table 4 HPLC buffer system

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

Mass spectroscopy

A Waters (Herts, UK) 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 (Germany) except 5-(octadiynyl)-dCTP (5-(octa-1,7-diynyl)-cytosine) which was synthesised in house following the protocol by Chittepu *et al.*[3] 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.

DNA sequencing

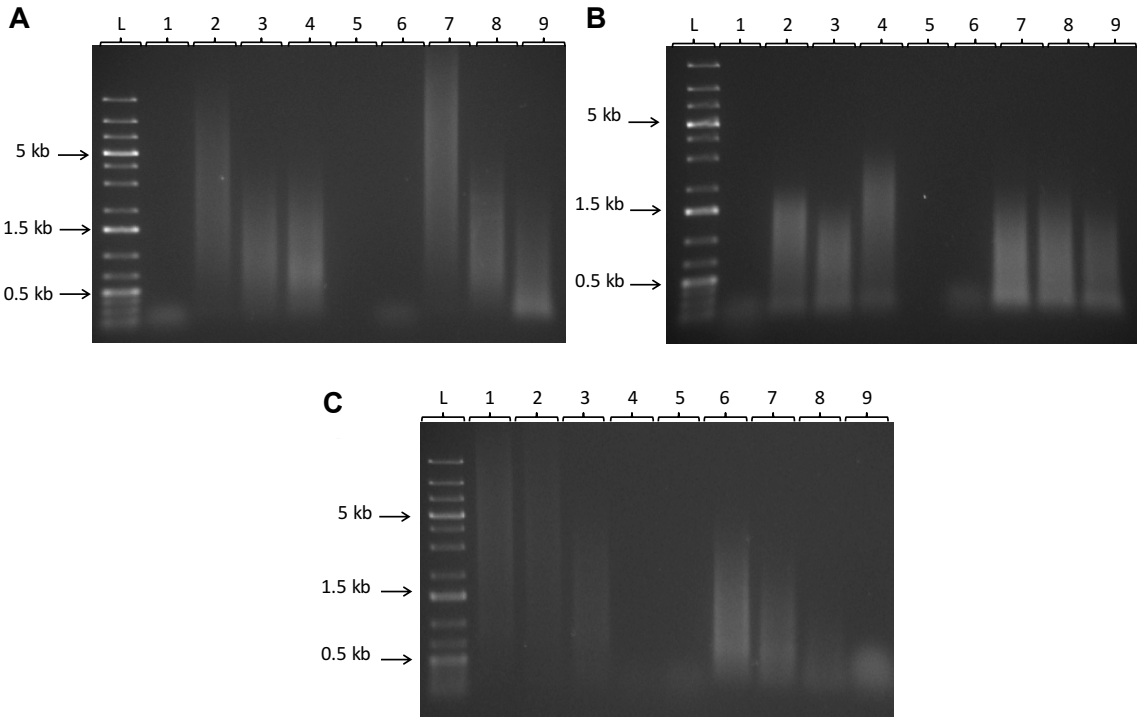
DNA Sanger sequencing was performed by GATC Biotech (Constance, Germany). In separate 1.5 mL microcentrifuge tubes, 20 μL of 10-50 ng/μL purified DNA extension product after 5 heat-cool cycles was sent along with 10 μM of one oligoseed, i.e. for $[A_4G]_n$/$[CT_4]_n$, $[A_4G]_4$ or $[CT_4]_4$ was provided as the sequencing primer and for $[GATC]_n$/$[GATC]_n$, $[GATC]_5$ was provided as the sequencing primer.

Lonza gel electrophoresis

Lonza Gel electrophoresis and extraction was performed using the pre-cast 1.2 % agarose gel kit, Lonza FlashGel™ DNA system (Lonza), and performed as stated in the manufacturers handbook. After exraction, sample was purified using a QIAquick PCR purification kit (25) (QIAGEN, Manchester, UK) following manufacturers protocol. Samle concentration was determined using a Nanodrop 1000 Spectrometer (Thermo Scientific).

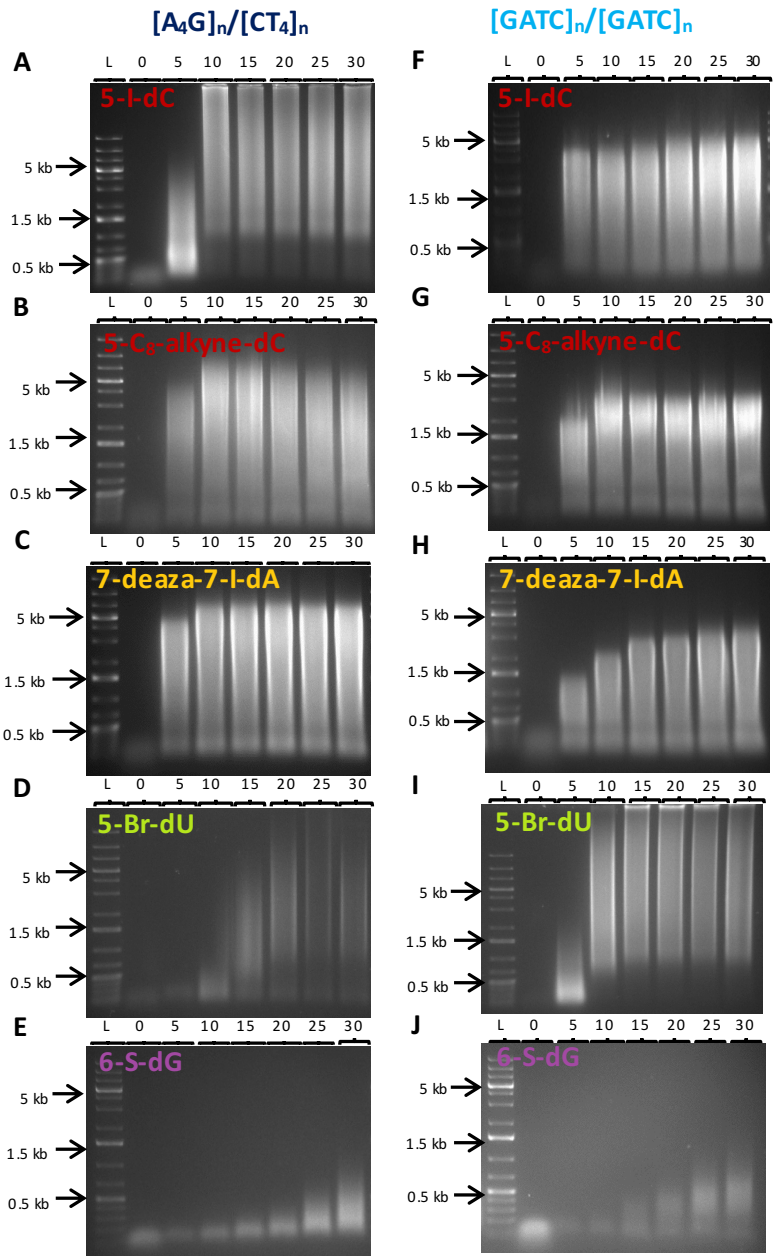
Supporting Data

Figure S1: Heat-cool cycles performed with Deep Vent exo-



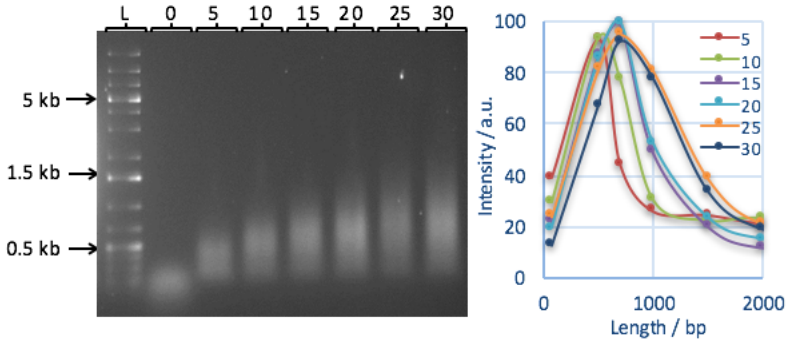
Agarose gels of DNA extension products with Deep Vent exo- and (**A**) $[A_4G]_n/[CT_4]_n$, (**B**) $[GATC]_n/[GATC]_n$ and (**C**) $[GATTC]_n/[GAATC]_n$ after 30 cycles. Lane 1 (A, B) and lane 9 (C) depict the oligoseed used for each extension assay. Lanes 2-9 (A, B) and lanes 1-8 (C), depict extensions with 5‑I‑dCTP, 5-Br-dUTP, 7-deaza-I-dATP, 5-acetly-Hg-dCTP, 6-S-dGTP, 5-(octadiynyl)-dCTP, and 5‑Br-dUTP, and 7-deaza-I-dATP, respectively.

Figure S2: Heat-cool cycles with one modified base per reaction



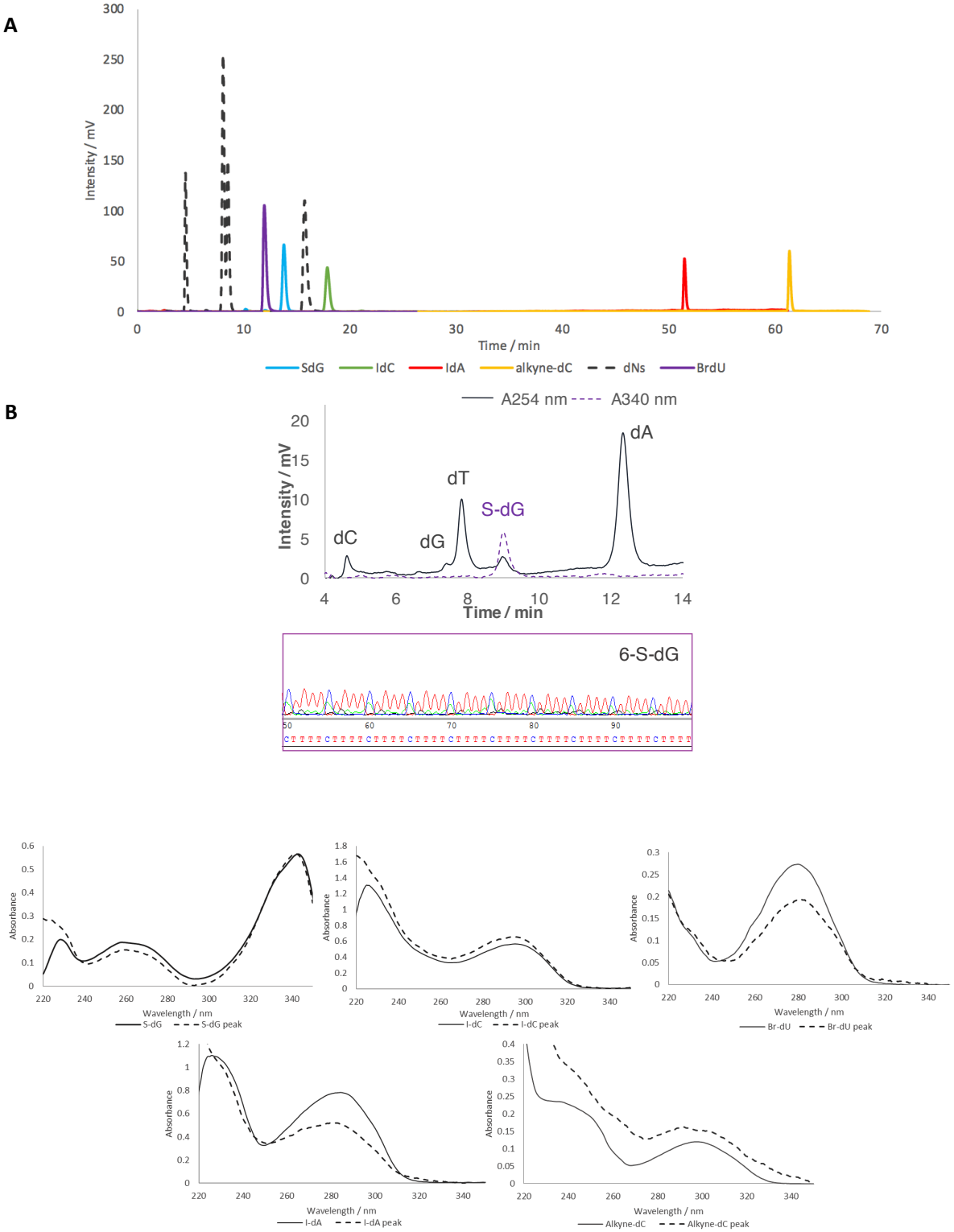
Agarose gels of DNA extension products with $[A_4G]_n/[CT_4]_n$ (A-E) and $[GATC]_n/[GATC]_n$ (F-J) and the specified modified dNTP with Tgo-Pol Z3 exo-.

Figure S3: Heat-cool cycle extension with 5-acetyl-Hg-dCTP and $[A_4G]_4/[CT_4]_4$



Agarose gel of Hg-dC-modified DNA extension products after 5, 10, 15, 20, 25 and 30 cycles is shown. Extension was performed with $[A_4G]_4/[CT_4]_4$ and shows succesful DNA extension.

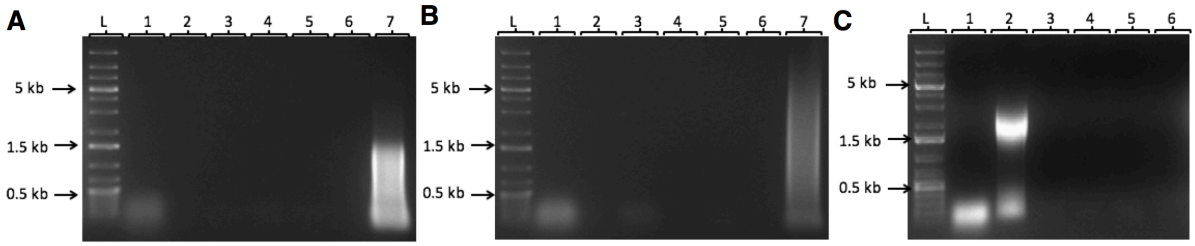
Figure S4: DNA digestions and characterisations



Standards of each deoxynucleotide (dN) were used to determine HPLC elusion times, shown as a HPLC trace (**A**) using the method described, see Supplementary Table 1. To determine extinction coefficients, a known mass of each dN was measured and prepared to a known concentration. Each dN was then analysed by

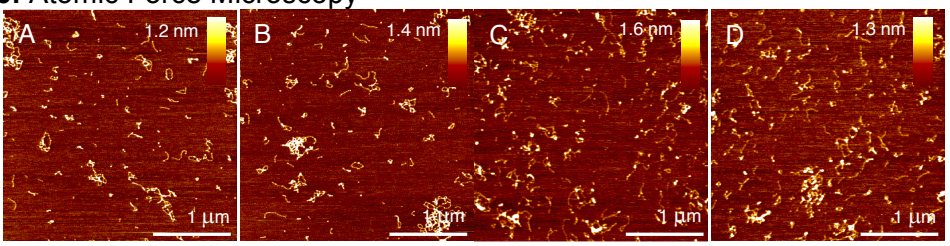
UV-Vis to determine extinction coefficients using Beer's Law. The HPLC trace of the 6-S-dG-$[A_4G]_n/[CT_4]_n$ digestion products (**B**) were compared with each standard dN to determine the peaks. DNA sequencing is also shown (**B**) for the 6-S-dG-$[A_4G]_n/[CT_4]_n$ modified DNA. Modified dNs were collected for the whole peak and UV-Vis of standard dNs compared to the collected dN peak from the HPLC run was performed (**C**). The collected modified dN peak was also analysed by high resolution mass spectroscopy, see Supplementary Table 2.

Figure S5: Heat-cool extensions performed in the absence of each dNTP individually

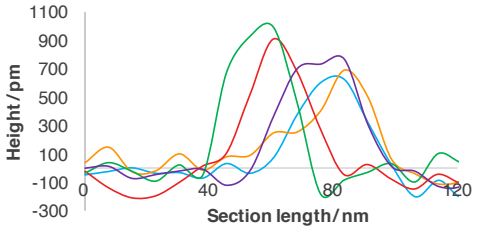


Negative controls of the DNA extension were also performed to confirm no contamination in the other un-modified nucleotides which could lead to incorrect incorporation. The controls were performed in the absence of each dNTP in turn and reinforce the absence of any contamination within the single dNTP stock solutions purchased from Roche. DNA extension controls with (**A**) $[GATC]_n/[GATC]_n$ and (**B**) $[A_4G]_n/[CT_4]$ with DVE- and (**C**) $[A_4G]_n/[CT_4]$ 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.

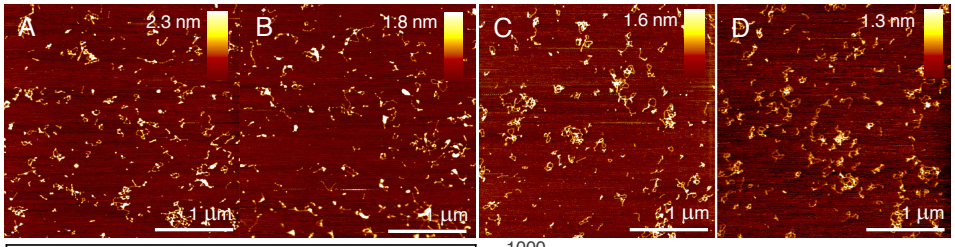
Figure S6: Atomic Force Microscopy



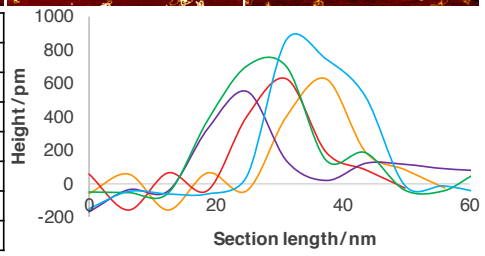
	A	B	C	D
Length (nm)	838	1188	1032	853
	976	897	732	735
	706	897	1029	885
	1135	653	1003	900
	1124	794	935	829
Average (nm)	**956**	**886**	**946**	**841**
SD (nm)	166	176	113	58



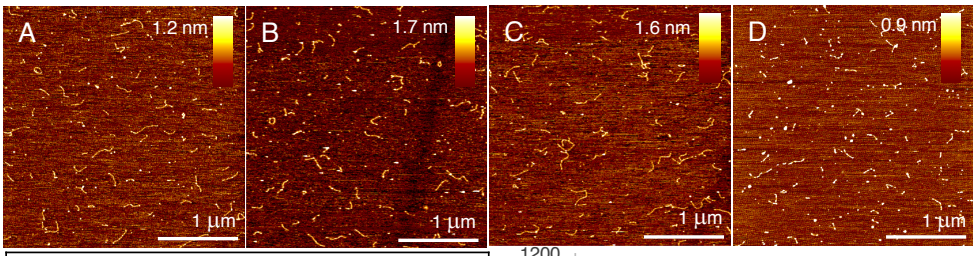
AFM images (A-D) of 5-I-dC-DNA, a table of the analysed lengths and a selection of perpendicular height scans.



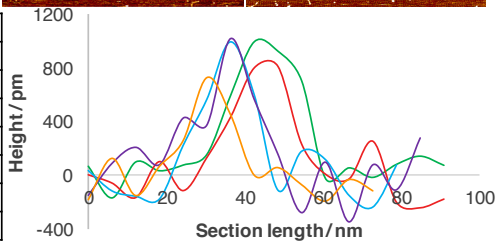
	A	B	C	D
Length (nm)	715	1050	571	609
	471	1026	768	512
	547	1206	1015	991
	824	865	1165	1165
	835	1100	956	1041
Average (nm)	**678**	**1049**	**895**	**864**
SD (nm)	147	111	206	256



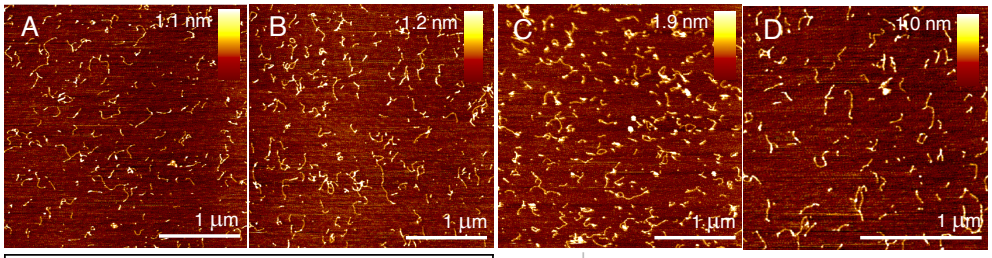
AFM images (A-D) of 5-(octadiynyl)-dCTP, a table of the analysed lengths and a selection of perpendicular height scans.



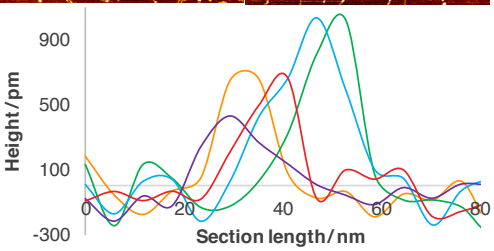
	A	B	C	D
Length (nm)	644	671	818	691
	1238	485	938	953
	621	818	791	929
	1056	621	897	1041
	647	656	647	797
Average (nm)	**841**	**650**	**818**	**882**
SD (nm)	256	106	101	123



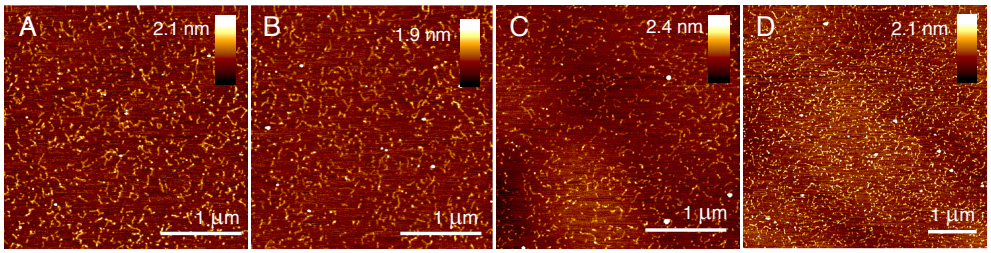
AFM images (A-D) of 7-deaza-7-I-dA-DNA, a table of the analysed lengths and a selection of perpendicular height scans.



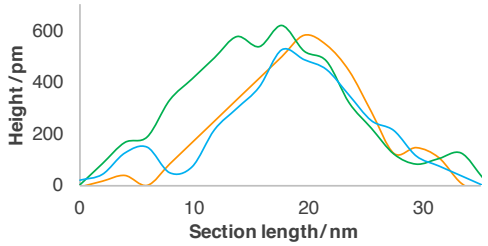
	A	B	C	D
Length (nm)	1441	1156	982	968
	1112	785	1679	641
	1271	918	965	965
	1121	815	1000	600
	794	624	718	606
Average (nm)	**1148**	**859**	**1069**	**756**
SD (nm)	214	176	322	172



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



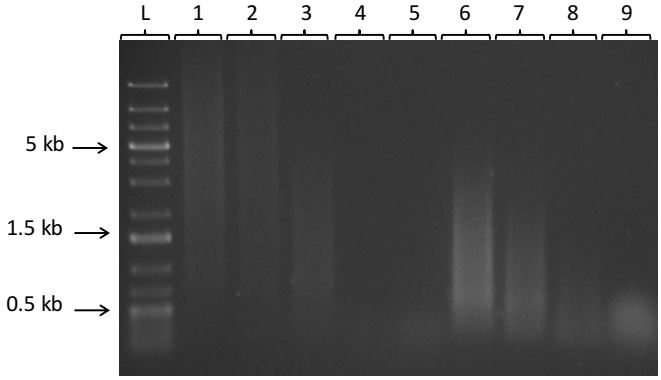
	A	B	C	D
Length (nm)	521	259	426	285
	353	324	297	326
	362	297	312	315
	306	265	332	294
	274	309	300	362
Average (nm)	**363**	**291**	**334**	**316**
SD (nm)	85	25	48	27



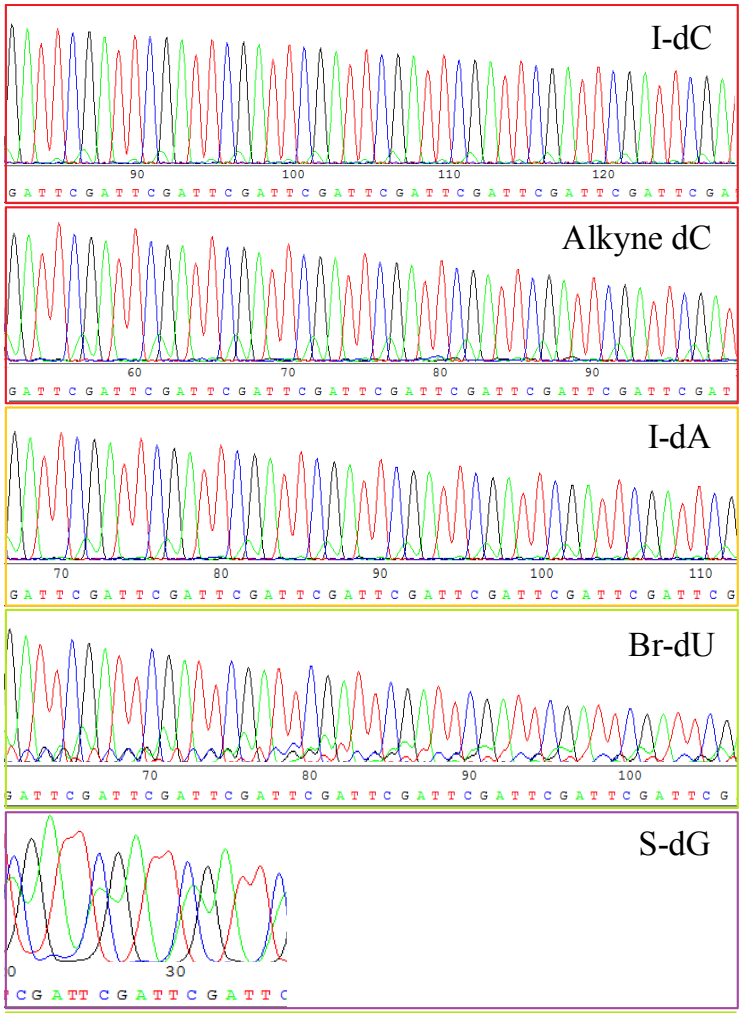
AFM images (A-D) of 6-S-dG-DNA, a table of the analysed lengths and a selection of perpendicular height scans.

Atomic Force Microscopy (AFM) was performed as described in methods for each sample with the $[A_4G]_n/[CT_4]_n$ sequence. A summary of this data is shown in Supplementary Table 3, depicting DNA lengths similar to the lengths depicted on the agarose gels.

Figure S7: DNA extensions with $[GAATC]_4/[GATTC]_4$



Agarose gel of DNA extension products from $[GAATC]_4[GATTC]_4$ after 30 cycles is shown. Extensions were performed with, lane 1: 5-I-dCTP, lane 2: 5-Br-dUTP, lane 3: 7-deaza-7-I-dATP, lane 4: 5-acetly-Hg-dCTP, lane 5: 6-S-dGTP, lane 6: 5-C_8-alkyne-dCTP, lane 7: 5-(octadiynyl)-dCTP and 5-Br-dUTP, lane 8: 5-C_8-alkyne-dCTP, 5-Br-dUTP and 7-deaza-7-I-dATP, lane 9: $[GAATC]_4/[GATTC]_4$ oligoseed.



DNA sequencing of one-type modification DNA extension products after 30 cycles with $[GAATC]_n[GATTC]_n$. DNA Sanger sequencing with $[GATC]_n/[GATC]_n$ sequences appeared to be challenging – possibly due to its ability to form self-complementary structures. An attempt to emulate this was made by including extensions with $[GATTC]_n/[GAATC]_n$ which yielded similar extension efficiencies with each modified base derivative and showed positive DNA sequencing results.

Figure S8: DNA average length determination.

DNA average product lengths were determined using Image J. Plots were derived by scanning over each lane on the agarose gels from Figure 1(**A**) to 1(**J**) and peak maximum chosen from the highest intensity value. The peak maximum was correlated to the ladder to determine DNA length at peak maximum.

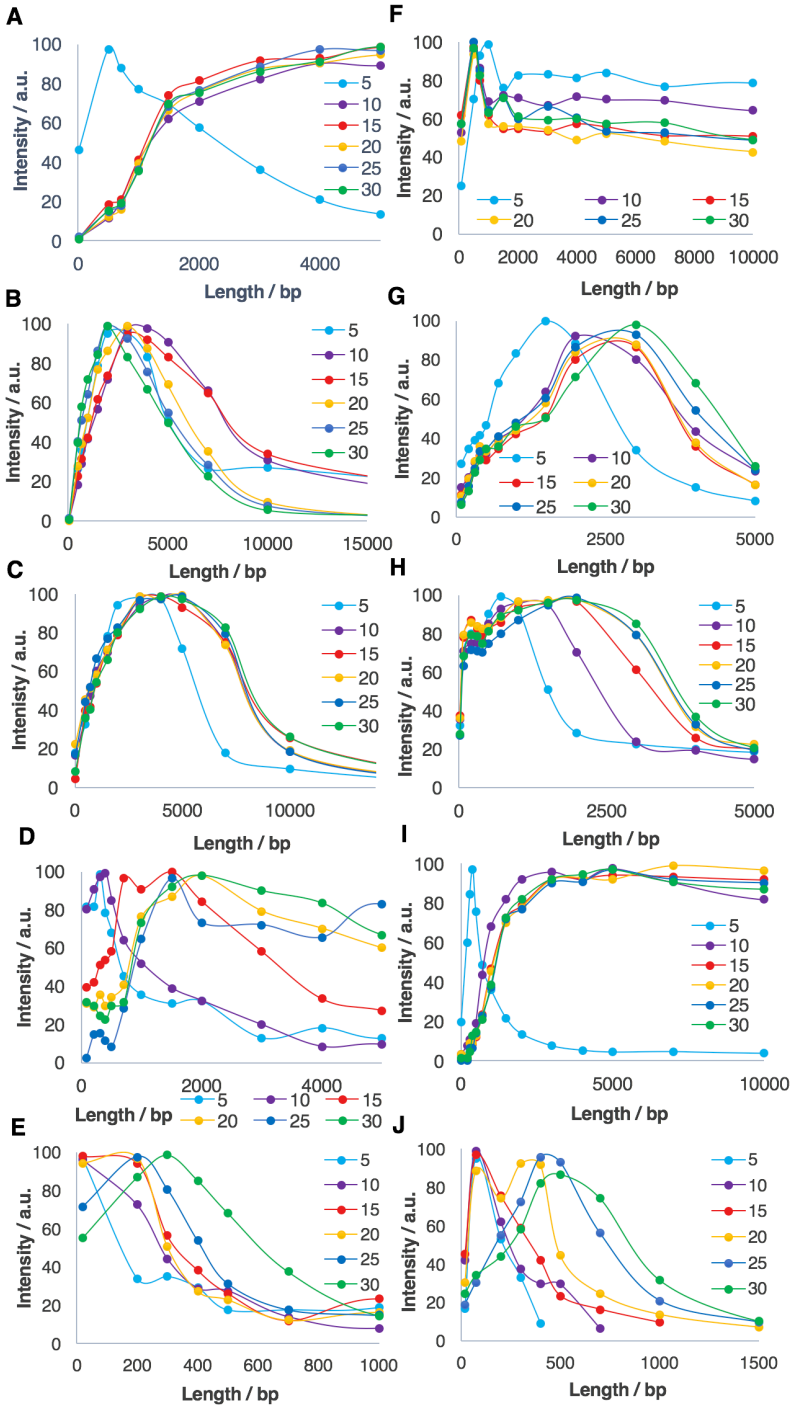
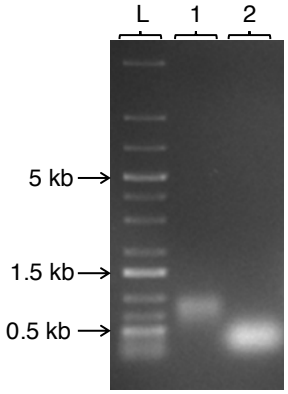
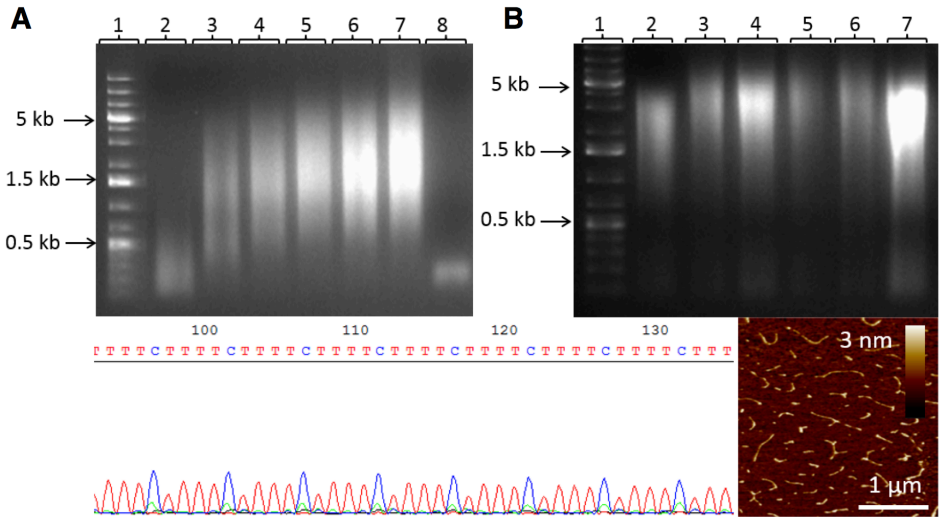


Figure S9: Heat-cool extensions performed with 6-S-dGTP



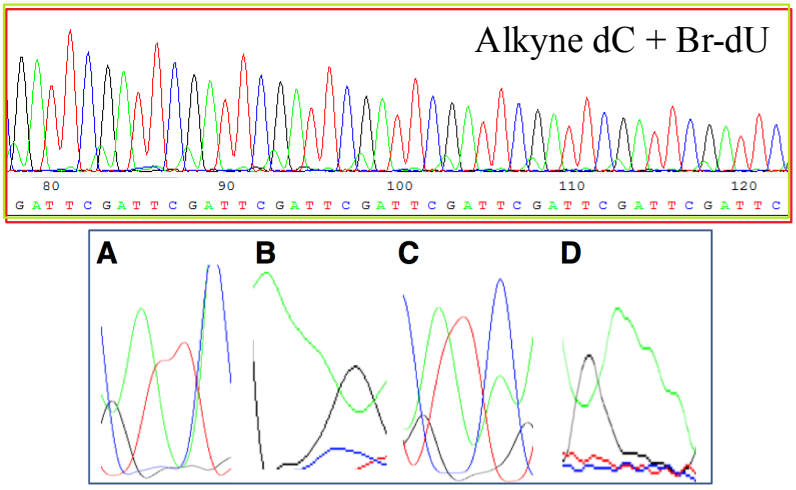
Agarose gel of heat-cool cycles extension products of $(G)_n.(C)_n$ with 6-S-dGTP, Lane 1. Lane 2 depicts the starting oligoseed, $(G)_{20}.(C)_{20}$. L = DNA ladder.

Figure S10: Heat-cool extension incorporation of multiple modified dNTPs per DNA pot.



Analysis of DNA products after extension in the presence of 5-I-dCTP, 5-Br-dUTP, dGTP and dATP. Gel (**A**) shows the DNA extension products after 5, 10, 15, 20, 25 and 30 cycles with Tgo-Pol-Z3 exo-, along with DNA sequencing and AFM of the products after 30 cyclces with $[A_4G]_4/[CT_4]_4$. Gel (**B**) showns the DNA extension products after 5, 10, 15, 20, 25 and 30 cycles with Tgo-Pol-Z3 exo- and $[GATC]_5/[GATC]_5$.

Figure S11: DNA sequencing of 2-, 3-, and 4-modifications per DNA extension reaction pot



DNA sequencing was performed with 5-C_8-alkyne-dC and 5-Br-dU modified DNA and yielded the correct resulting sequence. DNA sequencing of (**A**) and (**B**) 6-S-dG, 7-deaza-7-I-dA and 5-Br-dU modified (**A**) $[GAATC]_n[GATTC]_n$ and (**B**) $[A_4G]_n/[CT_4]_n$ and (**C**) and (**D**) 6-S-dG, 7-deaza-7-I-dA, 5-(octadiynyl)-dCTP and 5-Br-dU modified (**C**) $[GAATC]_n[GATTC]_n$ and (**D**) $[A_4G]_n/[CT_4]_n$. These confirm the correct incorporation of each modified nucleotide following the Watson Crick base pairing rules.[4]

Supporting Tables

***Table S1**: deoxynucleotide elution times and extinction coefficients*

dN	HPLC elution time / min	Extinction coefficient at A254 / $M^{-1}cm^{-1}$
dC	4.6	5180
dG	8.0	14190
T	8.5	7790
dA	15.7	9850
6-S-dG	13.8	3820
5-I-dC	17.9	4250
5-Br-dU	14.6	2700
7-deaza-7-I-dA	50.8	5630
5-(octadiynyl)-dCTP	54.5	7200*[3]

*Extinction coefficient at λmax(299 nm).

***Table S2:** High resolution mass spectrum (HRMS) product peaks of collected modified dNs from the HPLC traces*

Collected dN	HRMS
5-I-dC	$[M-H]^+$ calculated for $[C_9H_{13}IN_3O_4]^+$ 353.9951 found 353.9951
7-deaza-7-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-(octadiynyl)-dCTP	$[M-H]^+$ 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 S3**: Average DNA lengths determined from AFM*

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

References

1. Jozwiakowski, S. K. and Connolly, B. A. (2011) A Modified Family-B Archaeal DNA Polymerase with Reverse Transcriptase Activity. *Chembiochem,* **12**, e35.

2. Evans, S. J., Fogg, M. J., Mamon, A., Davis, M., Pearl, L. H. and Connolly, B. A. (2000) Improving dideoxynucleotide-triphosphate utilisation by the hyper-thermophilic DNA polymerase from the archaeon *Pyrococcus furiosus. Nucleic Acids Res.,* **28**, e1059.

3. Seela, F., Sirivolu, V. R. and Chittepu, P. (2008) 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**, e211.

4. Watson, J. D., and Crick, F. H. C. (1953) A Structure for Deoxyribose Nucleic Acid. *Nature* **171**, e737