Electronic Supplementary Information

Triplex-mediated analysis of cytosine methylation at CpA sites in DNA

Marie W. Johannsen,[a,§] Simon R. Gerrard,[a] Tracy Melvin[b] and Tom Brown*[a,§]

[a] School of Chemistry, University of Southampton, Highfield, Southampton SO17 1BJ, UK.

[b] Optoelectronics Research Centre, University of Southampton, Highfield, Southampton, SO17 1BJ, UK

[§] Current address: Department of Chemistry, University of Oxford, Chemistry Research Laboratory, 12 Mansfield Rd, Oxford, OX1 3TA, UK. email:tom.brown@chem.ox.ac.uk. Tel: +44 (0)1865 275413.

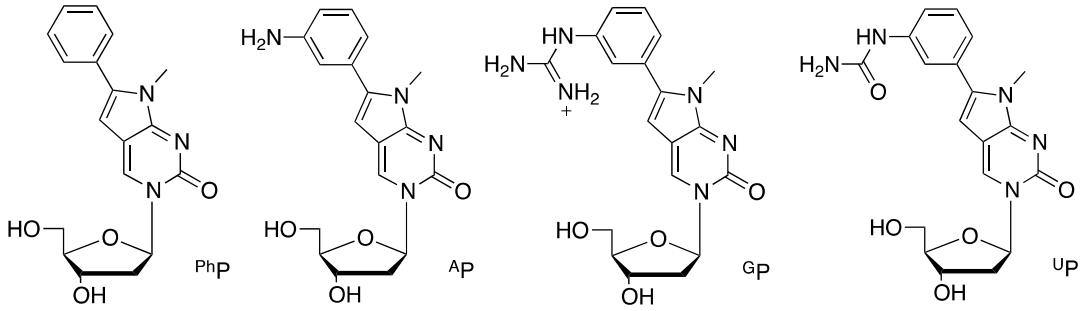

Figure S1 Structures of modified nucleosides[1]

General methods

Modified DNA phosphoramidites ^{Ph}P, ^{G}P, ^{A}P, and ^{U}P were synthesised as previously described.[1] Standard DNA phosphoramidites, solid supports, and additional reagents were purchased from Link Technologies and Applied Biosystems Ltd. The ^{hm}C (5-hydroxymethyl-dC II-CE phosphoramidite), ^{f}C (5-formyl-dC-CE phosphoramidite), and ^{ca}C (5-carboxy-dC-CE phosphoramidite) were purchased from Glen Research/Cambio and 5-methyl deoxycytidine phosphoramidite (^{Me}C) from Link Technologies. All oligonucleotides were synthesised on an Applied Biosystems 394 automated DNA/RNA synthesiser using a standard 0.2 or 1.0 μmol phosphoramidite cycle of acid-catalyzed detritylation, coupling, capping, and iodine oxidation. TFOs containing ^{G}P, ^{A}P, and ^{U}P were synthesised with no capping step. Stepwise coupling

[1] S. R. Gerrard, M. M. Edrees, I. Bouamaied, K. R. Fox, T. Brown, *Org. Biomol. Chem.*, 2010, **8**, 5087-5096

efficiencies were determined by the automated trityl cation conductivity monitoring facility and in all cases were >98.0%. All β-cyanoethyl phosphoramidite monomers were dissolved in anhydrous acetonitrile (or freshly distilled CH_2Cl_2 if poorly soluble in acetonitrile) to a concentration of 0.1 M immediately before use. The coupling time for normal A, G, C, and T monomers was 25 s (0.2 μmol) or 40 s (1 μmol), methyl cytosine monomers were coupled for 180 s, the ^{Ph}P, ^{G}P, ^{A}P, ^{U}P, and HEG (hexaethylene glycol) phosphoramidite monomers were coupled for 360 s, and for the ^{hm}C, ^{f}C and ^{ca}C phosphoramidite monomers, the coupling time was 600 s. Most oligonucleotides were deprotected and cleaved from the solid support using standard conditions (concentrated aqueous ammonia solution for 60 min at room temperature followed by heating in a sealed tube for 5 h at 55 °C). For oligonucleotides containing the ^{G}P, ^{A}P, and ^{U}P monomers, solid supports were washed after synthesis with 20% diethylamine in acetonitrile for 20 min (60 min for ^{G}P monomer) and dried, prior to cleavage of oligonucleotides from the solid-support and deprotection (conc. aq ammonia, rt for 24 h, followed by 55 °C for 1 h).

Oligonucleotides containing ^{hm}C, ^{f}C and ^{ca}C were cleaved from solid support and largely deprotected using 0.4M NaOH in MeOH/H_2O (4:1 v/v) at room temperature for 17-26 hr. The solutions were neutralised with 2M TEAA, desalted by gel filtration then freeze-dried. The deprotection for the ^{hm}C and ^{ca}C containing oligonucleotides was completed (once the ^{hm}C/^{ca}C nucleotide was determined to be fully deprotected by HPLC-MS), using concentrated aqueous ammonia solution, heating at 55 °C for up to 2 hr.

For oligonucleotides containing ^{f}C, the NaOH-mediated deprotection was repeated (5 hr deprotection time) where necessary to ensure complete deprotection of masked ^{f}C. The product 1,2-diols were oxidised to the corresponding formyl groups *via* oxidation with $NaIO_4$. Each product was suspended in water (250 µL) and chilled to 4 °C. A solution of 50 mM $NaIO_4$ (250 µL) was added, and the mixture was vortexed and chilled at 4 °C for 30 min. After gel filtration, the products were RP-HPLC purified using triethylammonium acetate buffer, to prevent nucleophilic addition of ammonia to the formyl group.

Purification and analysis of oligonucleotides

The deprotected oligonucleotides were purified by reversed-phase HPLC on a Gilson system using a Luna 10u C8 100Å 10 × 250 mm column (Phenomenex) with a gradient of acetonitrile in triethylammonium acetate (TEAA), (0 – 35% buffer B over 15 min, flow rate 4 mL/min. Buffer A: 0.1 M TEAA, buffer B: 0.1 M TEAA with 50% acetonitrile), or ammonium acetate (0 – 25% buffer B over 15 min, flow rate 4 mL/min. Buffer A: 0.1 M ammonium acetate, buffer B: 0.1 M ammonium acetate with 50% acetonitrile). Hairpin oligonucleotides were purified on a ResourceQ 6 mL column (GE Healthcare) using 0 – 75% buffer B over 23 min (flow rate 6 mL/min, Buffer A: 0.05 M NaCl, 0.01 M NaOH, pH 12, buffer B: 1 M NaCl, 0.01M NaOH, pH 12). Elution of oligonucleotides was monitored by ultraviolet absorption at 280-300 nm. After HPLC purification, oligonucleotides were desalted using NAP gel filtration columns (GE Healthcare) according to the manufacturer's instructions.

The identity and purity of oligonucleotides was confirmed by mass spectrometry and HPLC analysis. Negative ion electrospray mass spectra (350-3500 *m/z*) were recorded using a MicrOTOF ESI-TOF mass spectrometer (Bruker Daltonics, Bremen, Germany). All data were processed using Maximum Entropy *via* DataAnalysis (Bruker Daltonics, Bremen, Germany). Chromatographic separations were performed using a Dionex 3000 Ultimate 3000 UHPLC system (Thermo Scientific, Hemel Hempstead, UK). 2 µL samples in water were injected onto an Acquity UPLC BEH C18 column 1.7 µm 1.0 x 100 mm (Waters, Manchester, UK) heated to 40° C. Mobile phases used were (A) 10 mM TEAA, 100 mM HFIP in water and (B) 20 mM TEAA in acetonitrile, and separation was achieved using a gradient of 5% B to 40% B in 14 minutes at a flow rate of 0.1 mL/min. UV data were recorded at 290 nm.
HFIP = 1,1,1,3,3,3-hexafluoro-2-propanol.

Mass spectrometry data for oligonucleotides

Set name	Sequence (5′ to 3′)	Expected mass	Observed mass
Set 1 TFO	TTT TTM T^{Ph}PT MTM TMT	4596	4595.2
Set 1 TFO T	TTT TTM TTT MTM TMT	4497	4496.6
Set 1 pur. C	GCT AAA AAG ACA GAG AGA TCG	6521	6521.1
Set 1 pur ^{Me}C	GCT AAA AAG AMA GAG AGA TCG	6535	6534.9
Set 1 pur ^{hm}C	GCT AAA AAG AQA GAG AGA TCG	6551	6550.9
Set 1 pur ^{f}C	GCT AAA AAG AOA GAG AGA TCG	6549	6547.8
Set 1 pur ^{ca}C	GCT AAA AAG APA GAG AGA TCG	6565	6564.9
Set 1 pyr.	CGA TCT CTC TGT CTT TTT AGC	6329	6327.8
Set 2 TFO	CTM CC^{Ph}P MCG MCC TMT MTC	5428	5426.3
Set 2 pur. C	GGA GAG GGC GGT GGG AGA GAG GGC	7654	7653.9
Set 2 pur ^{Me}C	GGA GAG GGM GGT GGG AGA GAG GGC	7668	7667.0
Set 2 pur ^{hm}C	GGA GAG GGQ GGT GGG AGA GAG GGC	7684	7683.0
Set 2 pur ^{f}C	GGA GAG GGO GGT GGG AGA GAG GGC	7682	7680.9
Set 2 pur ^{ca}C	GGA GAG GGP GGT GGG AGA GAG GGC	7698	7697.0
Set 2 pyr. C	GCC CTC TCT CCC ACC GCC CTC TCC	7057	7056.8
Set 2 TFO T	CTM CCT MCG MCC TMT MTC	5328	5327.8
Set 2 TFO ^{G}P	CTM CC^{G}P MCG MCC TMT MTC	5485	5483.9
Set 3 TFO	CTM CT^{Ph}P TMT MCC TMT MTC	5433	5431.9
Set 3 pur. C	GGA GAG GAC AGA GGG AGA GAG GGC	7631	7630.0
Set 3 pur. ^{Me}C	GGA GAG GAM AGA GGG AGA GAG GGC	7645	7644.1
Set 3 pyr. C	GCC CTC TCT CCC TCT GTC CTC TCC	7078	7077.9
Set 3 hp C	GGA GAG GAC AGA GGG AGA GAG GGC-H-GCC CTC TCT CCC TCT GTC CTC TCC	15116	15115.0
Set 3 hp ^{Me}C	GGA GAG GAM AGA GGG AGA GAG GGC-H-GCC CTC TCT CCC TCT GTC CTC TCC	15130	15130.1
Set 4 TFO	TTT TTM C^{Ph}PT MTM TMT	4581	4580.9
Set 4 pur. C	GCT AAA AAG GCA GAG AGA TCG	6537	6536.2
Set 4 pur ^{Me}C	GCT AAA AAG GMA GAG AGA TCG	6551	6551.2
Set 4 pyr. C	CGA TCT CTC TGC CTT TTT AGC	6314	6314.0
Set 5 TFO	CTM CT^{Ph}P MTT MCC TMT MTC	5433	5431.9
Set 5 hp C	GGA GAG GAC GAA GGG AGA GAG GGC H GCC CTC TCT CCC TTC GTC CTC TCC	15116	15113.6
Set 5 hp ^{Me}C	GGA GAG GAM GAA GGG AGA GAG GGC H GCC CTC TCT CCC TTC GTC CTC TCC	15130	15128.7
Set 5 TFO ^{G}P	CTM CT^{G}P MTT MCC TMT MTC	5490	5488.9

Set name	Sequence (5′ to 3′)	Expected mass	Observed mass
Set 6 TFO	CTM CC^{Ph}P CMT MCC TMT MTC	5403	5400.8
Set 6/7 pur. C	GGA GAG GGC GGA GGG AGA GAG GGC	7663	7662.0
Set 6/7 pur ^{Me}C	GGA GAG GGM GGA GGG AGA GAG GGC	7677	7676.0
Set 6/7 pyr. C	GCC CTC TCT CCC TCC GCC CTC TCC	7048	7047.8
Set 7 TFO	CTM CM^{Ph}P MCT MCC TMT MTC	5417	5415.8
M = ^{Me}C, Q = ^{hm}C, O = ^{f}C, P = ^{ca}C, H = hexaethylene glycol			
Additional TFOs with ^{G}P, ^{A}P, ^{U}P as previously described[1]			

For HPLC traces and MS spectra of modified oligonucleotides, see end of Supplementary Information

Conditions for melting temperature (T_m) measurements

UV melting experiments were performed on a Cary 4000 UV-Visible Spectrophotometer (Varian) in Hellma® SUPRASIL synthetic quartz cuvettes (10 mm pathlength, 1.5 mL sample volume), monitoring at 260 nm, using Cary WinUV Thermal application software. The TFO and duplex strands were suspended in 1.0 mL of the correct buffer (see table captions) to afford a 5 μM: 1 μM concentration of TFO:duplex. When measuring duplexes alone, the samples contained 1 μM duplex. When measuring a (parallel) duplex composed of the purine strand and the TFO strand, samples were made up to either 5 μM:1 μM or 1 μM:1 μM concentration of TFO:purine as noted in tables S.2, S.3). The samples were transferred to cuvettes and subjected to the DNA melting programme alongside a matched cell reference blank. The samples were equilibrated by initial denaturation by heating to 85 °C at 10 °C/min. After holding for 2 min, the samples were annealed by cooling to 12-15 °C at 0.5 °C/min and held at that temperature for 20 min. Melting cycles in the range between 12 and 85 °C at 0.5 °C/min with a hold time at 85 °C of 2 min and 20 min at 12-15 °C were carried out. Average T_m values were derived from the first derivatives of the melting curves for each experiment using OriginPro 8.6.

Melting temperature (T_m) data for triplexes with other modified nucleosides

Table S1 Data from additional melting studies

Set name	pH	*Target trinucleotide*	Cytosine analogue	T	^{Ph}P	^{A}P	^{U}P	^{G}P
Set1	6.2	AYA	C	26.7*	32.0	31.5	30.9*	29.0*
	6.2		^{Me}C	26.7*	18.1	19.2	21.8*	20.2*
Set2	6.2	GYG	C	16.9	16.5			
	6.2		^{Me}C	17.0	16.7			
Set 2	5.8	GYG	C	25.0	23.2			25.3
	5.8		^{Me}C	25.5	23.0			26.0
Set 5 hp	5.8	AYG	C		40.4			44.6
	5.8		^{Me}C		37.7			41.9

Triplex melting data, temperatures in °C, Y = C or ^{Me}C, 10 mM phosphate buffer, 200 mM NaCl, 1 mM EDTA. pH as written. Final T_m values are averages of at least three measurements, except where marked*, which are average of at least two measurements.

T_m data of additional transitions (not due to triplex denaturation)

These are data relating either to duplex denaturation, for example, denaturation of the Set 2 duplex:

```
5′- GGA GAG GGY GGT GGG AGA GAG GGC
3′- CCT CTC CCG CCA CCC TCT CTC CCG
```

or to denaturation of any duplex structure formed by the purine and TFO strand alone, for example (Set 2 parallel purine:TFO duplex):

```
    5′- CTM CCX MCG MCC TMT MTC
5′- GGA GAG GGY GGT GGG AGA GAG GGC
```

Table S2 Data measured at pH 6.2

Set name	*Target trinucleotide*	Cytosine analogue	*Duplex denaturation*	*Other duplex transitions*	*TFO:purine 1 μM:1 μM*	*TFO :purine 5 μM:1 μM*
Set1	AYA	C	63.0			
		^{Me}C	63.3			
Set2	GYG	C	77.4			
		^{Me}C	78.0			
Set 3	AYA	C	75.0	18.8	45.3	
		^{Me}C	75.2		46.7	32.7, 49.9
Set 3 hp	AYA	C	> 85	No	N/A	N/A
		^{Me}C	> 85	No	N/A	N/A
Set 5 hp	AYG	C	> 85	No	N/A	N/A
		^{Me}C	> 85	No	N/A	N/A
Set 4	GYA	C	65.5	No	27.6	
		^{Me}C	66.3	No	26.3	

Triplex melting data, temperatures in °C, Y = C or ^{Me}C, 10 mM phosphate buffer, 200 mM NaCl, 1 mM EDTA, pH 6.2. Final T_m values are averages of at least two measurements. N/A = not applicable (cannot be tested).

Table S3 Data measured at pH 5.8

Set name	*Target trinucleotide*	Cytosine analogue	*Duplex denaturation*	*Other duplex transitions*	*TFO:purine 1 μM:1 μM*	*TFO :purine 5 μM:1 μM*
Set1	AYA	C	61.6	No	33.3	
		^{Me}C	63.0	No	33.9	
Set2	GYG	C	77.8	No		25.8, 67.5
		^{Me}C	78.3	No		26.3, 68.6
Set 6	GYG	C	77.3	No	39.9, 52.7	
		^{Me}C	77.4	No	52.4	
Set 7	GYG	C	77.3	No	39.6, 53.8	
		^{Me}C	77.2	No	53.9	
Set 3	AYA	C	74.4	28.5		
		^{Me}C	74.3	No	46.9	
Set 3 hp	AYA	C	> 85	No	N/A	N/A
		^{Me}C	> 85	No	N/A	N/A
Set 5 hp	AYG	C	> 85	No	N/A	N/A
		^{Me}C	> 85	No	N/A	N/A
Set 4	GYA	C	64.6	No	33.4	
		^{Me}C	65.2		32.8	

Triplex melting data, temperatures in °C, Y = C or ^{Me}C, 10 mM phosphate buffer, 200 mM NaCl, 1 mM EDTA, pH 5.8. Final T_m values are averages of at least two measurements. N/A = not applicable (cannot be tested).

Triplex T_m data recorded at pH 5.8

Table S4 Triplex sequences investigated at pH 5.8

Set name	Sequence
Set 1	5′- TTT TTM T**X**T MTM TMT 5′- GCT AAA AAG A**Y**A GAG AGA TCG 3′- CGA TTT TTC T**G**T CTC TCT AGC
Set 2	5′- CTM CC**X** MCG MCC TMT MTC 5′- GGA GAG GG**Y** GGT GGG AGA GAG GGC 3′- CCT CTC CC**G** CCA CCC TCT CTC CCG
Set 3	5′- CTM CT**X** TMT MCC TMT MTC 5′- GGA GAG GA**Y** AGA GGG AGA GAG GGC 3′- CCT CTC CT**G** TCT CCC TCT CTC CCG
Set 3 hp	5′- CTM CT**X** TMT MCC TMT MTC 5′- GGA GAG GA**Y** AGA GGG AGA GAG GGC **H** 3′- CCT CTC CT**G** TCT CCC TCT CTC CCG
Set 4	5′- TTT TTM C**X**T MTM TMT 5′- GCT AAA AAG G**Y**A GAG AGA TCG 3′- CGA TTT TTC C**G**T CTC TCT AGC
Set 5 hp	5′- CTM CT**X** MTT MCC TMT MTC 5′- GGA GAG GA**Y** GAA GGG AGA GAG GGC **H** 3′- CCT CTC CT**G** CTT CCC TCT CTC CCG
Set 6	5′- CTM CC**X** CMT MCC TMT MTC 5′- GGA GAG GG**Y** GGA GGG AGA GAG GGC 3′- CCT CTC CC**G** CCT CCC TCT CTC CCG
Set 7	5′- CTM CM**X** MCT MCC TMT MTC 5′- GGA GAG GG**Y** GGA GGG AGA GAG GGC 3′- CCT CTC CC**G** CCT CCC TCT CTC CCG

M = ^{Me}C, X = ^{Ph}P, Y = C or ^{Me}C, H = hexaethylene glycol.

Table S5 Melting temperatures of triplex to duplex transitions at pH 5.8

Set name	Target trinucleotide	C	^{Me}C	ΔT_m
Set 1	AYA	40.0	25.1	14.9
Set 2	GYG	23.2	23.0	0.2
Set 3	AYA	46.8[a]	37.0[a]	9.8
Set 3 hp	AYA	46.4	36.9	9.5
Set 4	GYA	35.4	26.1	9.4
Set 5 hp	AYG	40.4	37.7	2.7
Set 6	GYG	37.9[b]	37.1[b]	0.8
Set 7	GYG	37.3[b]	36.7[b]	0.6

T_m data measured in °C. Y = C or ^{Me}C, 10 mM phosphate buffer, 200 mM NaCl, 1 mM EDTA, pH 5.8. [a]A transition is also seen at lower temperature, most likely due to weak duplex secondary structure. It is also observed when the T_m is measured for the duplex only, and is not seen in the hairpin structure (Tables S.2, S.3. [b] The triplex transition appears to consist of two partially overlapping transitions, which give additional peaks in the first derivative of the triplex denaturation curves. For Set 6 at 25.4°C (C) and 25.6°C (^{Me}C) and for Set 7 at 31°C (C) and 32°C (^{Me}C). No low temperature transitions appear for the duplex alone and when the TFO and purine strand are measured together, two transitions are seen for C but not for ^{Me}C (Tables S.2, S.3). As above, this is likely due to weak secondary structure.

Some additional triplexes were investigated at this pH (Sets 6 and 7), which are closely related to Set 2. Both have a TFO sequence where a G.TA triplet is replaced with T.AT in order to stabilise the triplex. While this does indeed stabilise the triplex relative to Set 2, the methylation status of the purine strand does not affect the melting temperature for either of sets 2, 6, or 7. Again, at this lower pH, Set 1 and Set 3, which have the central trinucleotide sequence AYA, show a large methylation-dependent difference in melting temperature (ΔT_m = 14.9 °C and 9.8 °C respectively), discrimination similar to that seen at pH 6.2. Importantly, Set 3, Set 3 hp, and Set 4 share nearly identical ΔT_m values.

Additional melting curves.

The following conditions apply for the thermal denaturation curves on the next pages:

A) UV absorption at 260 nm recorded as a function of temperature from 15 °C to 80 °C. **B**) Smoothed first derivative of the thermal denaturation curves shown in **A**. The first transition represents triplex denaturation, the second is denaturation of the duplex. 10 mM phosphate buffer, 200 mM NaCl, 1 mM EDTA, pH 5.8.

^{Ph}P.CG (——), ^{Ph}P.^{Me}CG (- - -), except where otherwise indicated.