Supplementary Material (ESI) for Chemical Communications

Observation of a Dynamic Equilibrium Between DNA Hairpin and Duplex Forms of Covalent Adducts of a Minor Groove Binding Agent

Khondaker M. Rahman,[a] Vanelssa Mussa,[a] Mathangi Narayanaswamy,[a] Colin James,[a] Philip W. Howard,[a,b] and David E. Thurston,[*,a,b]

[a]*Gene Targeting Drug Design Research Group, The School of Pharmacy, University of London, 29/39 Brunswick Square, London WC1N 1AX and* [b]*Spirogen Ltd, The School of Pharmacy, University of London WC1N 1AX*.

Supplementary Information

Single-Stranded Oligonucleotides (*Seq-1* to *Seq-4*)

The single-stranded (SS) oligonucleotides were obtained in a lyophilised form from Eurogentec, U.K. (see **Table S1**).

Double-Stranded Oligonucleotide

Each single-stranded oligonucleotide was dissolved in 1M ammonium acetate (Sigma-Aldrich, U.K.) to form a stock solution of 1 mM. A solution of double-stranded DNA *Seq-1*/*2* (*Seq-1* + *Seq-2*) was prepared by heating the complementary SS DNA sequences (1 mM) in 1 M ammonium acetate to 90°C for 10 minutes in a heating/cooling block (Grant Bio, U.K.). The solution was then allowed to cool slowly to room temperature followed by storage at -20°C overnight to ensure completion of the annealing process. Working solutions of DS DNA duplex of 50 μM were prepared by diluting the stored solutions with 100 mM ammonium acetate. **Table S2** shows the structures of the annealed double-stranded oligonucleotide and its average mass.

PBD Conjugate 1 (GWL-78)

PBD conjugate **1** (GWL-78) was provided by Spirogen Ltd (Batch No. SG2274.005) and was dissolved in methanol to form a stock solution of 10 mM which was stored at -20°C for no longer than four months. Working solutions of the drug of 200 μM were prepared by diluting the above solution with 100 mM

ammonium acetate. These were stored at -20^{o}C for not more than one week and thawed to room temperature for use when required.

Preparation of Complexes of 1 and DNA

Ligand/ DNA complexes were prepared by incubating **1** with single stranded or duplex oligonucleotides at a 4:1 molar ratio at room temperature. Samples were withdrawn at various time intervals and subjected to ion-pair RPLC and mass spectrometry analysis as described below.

Ion-Pair Reversed Phase Liquid Chromatography (RPLC) Analysis

Chromatography was performed on a Thermo Electron HPLC system equipped with a 4.6 × 50 mm Xterra MS C18 column packed with 2.5 µM particles (Waters Ltd, U.K.), an UV 1000 detector, an AS3000 autosampler, a SCM1000 vacuum degasser and Chromquest software (Version 4.1). A gradient system of 100 mM triethyl ammonium bicarbonate (TEAB) as buffer A and 40% acetonitrile in water (HPLC grade, Fischer Scientific, U.K) as buffer B was used. For buffer A, a 1 M pre-formulated buffer of TEAB was purchased from Sigma-Aldrich, U.K and diluted to 100 mM with HPLC grade water (Fischer Scientific, U.K). The gradient was ramped from 90% A at 0 minutes to 50% A at 20 minutes and finally to 10% A at 35 minutes. UV absorbance was monitored at 254 nm and fractions containing separated components were collected manually, combined when appropriate, lyophilised and analysed using ESI-QTOF and MALDI TOF mass spectrometer as described below.

Mass Spectrometry Analysis (ESI-MS)

ESI-MS spectra were acquired on a Micromass Q-TOF Global Tandem Mass Spectrometer (Waters, Manchester, U.K.) fitted with a NanoSpray ion source. Negative mode was used for data acquisition and the instrument was calibrated with ions produced from a standard solution of taurocholic acid (10 pmole/µl) in acetonitrile. The HPLC fractions collected were lyophilised (Speedvac, Thermo Electron, U.K.) and mixed with a 1:1 v/v mixture of 40% acetonitrile/water and 20mM triethylamine/water (TEA, Fischer Scientific, U.K.) which was also used as the electrospray solvent. 3-5 µL of sample was loaded into a metal-coated borosilicate electrospray needle with an internal diameter of 0.7 mm and a spray orifice of 1-10 µm (NanoES spray capillaries, Proxeon Biosystems, U.K.) which was positioned at ~10 mm from the sample cone giving a flow rate of ~20 nl/min. Nitrogen was used as the API gas, and the capillary, cone and RF Lens 1 voltages were set to 1.8 - 2.0 kV, ~ 35 V and 50 V, respectively to ensure minimum fragmentation of

the ligand/DNA adducts. The collision voltage was 5 V and the MCP voltage was set at 2200 V. Spectra were acquired over the *m/z* range 1000-3000.

Mass Spectrometry Analysis (MALDI TOF)

Samples were prepared by diluting with matrix (37 mg THAP in 1 mL ACN, 45 mg ammonium citrate in 1 mL water – mix 1:1 for matrix) either 2:1, 1:1 or 1:5 (sample:matrix). 1 µl of sample was spotted onto the MALDI target plate and allowed to dry. Samples were analyzed on a Voyager DE-Pro with a nitrogen laser in positive linear mode using delayed extraction (500 nsec) and an accelerating voltage of 25000V. Acquisition was between 4000 – 15000 Da with 100 shots/spectrum.

Circular Dichroism and UV melting

CD spectra and UV melting curves were recorded on an Applied Photophysics Chirascan Circular Dichroism Spectrophotometer (Applied Photophysics Ltd, UK) using a quartz cell of 1-cm optical path length and an instrument scanning speed of 100 nm/min with a response time of 2 s, and over a wavelength range of 200-360 nm at 20ºC. The oligonucleotide concentrations used were 5 µM. All CD spectra are zero corrected and baseline-corrected for signal contributions due to the buffer. Ligand/oligonucleotide complexes were prepared by incubating **1** with oligonucleotides at a 4:1 molar ratio at room temperature in 100mM ammonium acetate buffer. *Seq-1* and *Seq-2* melting curves were recorded at 222 nm between 293 K to 368 K. The melting profile of 1:1 *Seq-1* and *Seq-2* adducts were recorded at 260 nm between 293 K to 368 K.

Molecular Modelling

General Method

In order to examine the structural feasibility of the DNA duplex and hairpin structures with and without **1** bound, molecular models were constructed. To test the integrity of the structures under energetic conditions, dynamics simulations were carried out at room temperature (300K). Initial models of DNA constructs were made using the 'nucgen' build module of the AMBER(v9) modeling software. The PBD adducts were constructed by means of MacroModel (v6.5). The minimized structure was exported in 'pdb' format and converted to the 'mol2' format with Gasteiger charges by means of the 'antechamber' routine before missing parameters were constructed with the 'parmchk' program. Covalent binding of **1** to the guanine residue of the DNA was performed graphically using AMBER 'Xleap' utilizing 'parm99' and the general Amber force field parameters (gaff). Care was taken to ensure that the (*S*)-configuration was maintained at the C11-

position of the central PBD ring at the point of attachment to the guanine N2-position. Constructs were saved for subsequent minimization and dynamics using the AMBER 'Sander' program. In Sander, the generalized Born/surface area (GB/SA) implicit solvent model was used with monovalent electrostatic ion screening simulated with SALTCON set to 0.2M. The dynamics integration time-step was 0.002ps while constraining all bonds to hydrogen atoms using the SHAKE algorithm. A temperature of 300K was maintained using the Langevin thermostat (NTT = 3, GAMMA_LN = 2.0), and a long range non-bonded cut-off of 100 (Å) was used.

Modelling of Ligand/DNA Adducts

Modelling of 1/Matched Hairpin Structures:
5'-TATAAGAAAATCTTATA-3' (*Seq-1*) and 5'-TATAAGATTTTCTTATA-3' (*Seq-2*)

AMBER9 was used to construct DNA hairpin "duplexes" for the two oligonucleotides *Seq-1* and *Seq-2* with complementary base pairs at positions 1-7/11-17, and the hairpin loop positioned at the central three residues, AAA or TTT, respectively (see *Seq-1* and *Seq-2*, **Figure S9**). The resulting files were edited to give a continuous 5' to 3' sequence. After minimization, a dynamics simulation was performed over 300ps according to the general method described above. A single molecule of **1** was then added in each of the two possible orientations (**Figure S9: A** and **B** for *Seq-1*, **C** and **D** for *Seq-2*). The first orientation (**S9A** or **S9C**) had **1** positioned snugly within the minor groove with the *N*-methylpyrrole heterocycles pointing towards the 5'-end of the covalently-bound (at G6) arm of the hairpin. In the second orientation (**9B** or **9D**), **1** was still covalently bound to G6 but the *N*-methylpyrrole heterocycles were pointing in the direction of the hairpin loop (*i.e.*, towards the 3'-end of the DNA). Molecular dynamics simulations of these different constructs were performed over 300ps. With the residues re-numbered as a continuous 5'- to 3'-sequence, the atoms in each dynamics frame were designated as rigid bodies fitted to those in the initial frame for comparison using residues 1-7 and 11-17, and the RMS deviation was measured (excluding the loop residues 8-10). The models of the initially energy minimized structures are shown in **Figure S10**.

Modelling of 1/Duplex DNA Adducts:

5'-TATAAGAAAATCTTATA-3' (*Seq-1*)/5'-TATAAGATTTTCTTATA-3' (*Seq-2*)

Models of the four possible 2:1 **1**/DS DNA adducts were constructed for the duplex 5'-TATAAGAAAATCTTATA-3' (*Seq-1*)/5'-TATAAGATTTTCTTATA-3' (*Seq-2*) as illustrated in **Figure S11** (**A-D**). Subsequent minimization and dynamics were performed over 100ps. The models of the initially energy minimized structures are shown in **Figure S12**.

Table S1: Structures and average masses of the single stranded (SS) oligonucleotides used in the study and the average mass of the adducts formed from their covalent interaction with one molecule of **1** (as measured by MS).

Table S2: Structure and average mass of the double stranded (DS) oligonucleotide duplex used in the study and the average mass of the adduct formed from its covalent interaction with two molecules of **1** (as measured by MS).

Figure Legends:

Figure S1: Schematic diagram of the various possible species resulting from the combination of oligonucleotides *Seq-1* and *Seq-2* and their covalent reaction with PBD conjugate **1** - **A**, DS *Seq-1*/*Seq-2*; **B**, Adduct of two molecules of **1** and DS *Seq-1*/*Seq-2* (four possible configurations – see **Figures S11** and **S12**); **C**, SS *Seq-1* (hairpin form); **D**, SS *Seq-2* (hairpin form); **E**, 1:1 **1**/*Seq-1* hairpin adduct; **F**, 1:1 **1**/*Seq-2* hairpin adduct, **G**, Adduct of two molecules of **1** and DS *Seq-1*/*Seq-2* (same as **B)**, in equilibrium with species **E** and **F**

Figure S2: HPLC chromatograms - **A**, *Seq-1*, retention time 26.4 mins; **B**, *Seq-1* immediately after incubating with **1**, gradual appearance of 1:1 *Seq 1* adduct, retention time 27.9 mins; **C**, *Seq-1* after incubating with **1** for 1 hour, 1:1 *Seq-1* adduct formation is complete, retention time 27.9 mins. **D**, *Seq-2*, retention time 26.9 mins; **E**, *Seq-2* immediately after incubating with **1**, gradual appearance of 1:1 *Seq 2* adduct, retention time 28.7 mins; **F**, *Seq-2* after incubating with **1** for 3 hours, 1:1 *Seq-2* adduct formation is complete, retention time 28.7 mins .

Figure S3: HPLC chromatograms - **A**, 1:1 *Seq-1* adduct after incubating Seq-1 with **1** for 24 hours, retention time 27.9 mins; **B**, 1:1 *Seq-2* adduct after incubating Seq-2 with **1** for 24 hours, retention time 28.7 mins; **C**, chromatogram obtained after mixing 1:1 *Seq-1* adduct and 1:1 *Seq-2* adduct, incubation time 24 hours, 1:1 adduct peaks are co-eluting at 29.0 mins, 2:1 adduct peaks are eluting at 31.1 mins & 32.0 mins; **D**, chromatogram obtained after re-injection of 2:1 adduct peak with retention time 31.1 mins, essentially similar to **S3C** ; **E**, chromatogram obtained after re-injection of 2:1 adduct peak with retention time 32.0 mins, essentially similar to **S3C**.

Figure S4: Time course HPLC chromatograms after mixing the 1:1 *Seq-1* and *Seq-2* adducts. The chromatograms obtained after 1, 2, 3 and 24 hours were essentially identical to S3C showing equilibrium has been established within 1 hour. **A**, chromatogram obtained after 1 hour of mixing; **B**, chromatogram obtained after 2 hours of mixing; **C**, chromatogram obtained after 3 hours of mixing; D, chromatogram obtained after 4 hours of mixing and **E**, chromatogram obtained after 24 hours of mixing.

Figure S5: MALDI TOF MS spectra of HPLC peaks in **Figure S3A**, **S3B** and **S3C** (co-eluting 1:1 adducts). **A**, MALDI TOF MS spectrum of HPLC peak at RT 27.9 mins (**S3A)**, $[M+H]^+$ ion at 5790.4 is consistent with 1:1 **1**/*Seq-1* adduct ; **B**, MALDI TOF MS spectrum of HPLC peak at RT 28.7 mins (**S3B)**, $[M+H]^+$ ion at 5766.1 is consistent with 1:1 **1**/*Seq-2* adduct; **C,** MALDI TOF MS spectrum of HPLC peak at RT 29.0 mins (**S3C)**, $[M+H]^+$ ions at 5789.7 and 5764.4 are consistent with both the **1**/*Seq-1* and **1**/*Seq-2* adducts, confirming that they are co-eluting.

Figure S6: Primary negative-ion ESI-MS spectra of HPLC peaks in **Figure S3C**, **A**, 29.0 min peak. Ions at m/z 1929.4 and 1920.3 ($[M-3H]^{3-}$), 1446.8 ($[M-4H]^{4-}$) and 1723.4 ($[M-3H]^{3-}$) are consistent with both the **1**/*Seq-1* and **1**/*Seq-2* adducts; **B**, 31.1 min peak. Ions at m/z 1924.6 ($[M-6H]^{6-}$) and 1649.6 ($[M-7H]^{7-}$) are consistent with a 2:1 **1**/(*Seq-1*/*Seq-2*) adduct; **C**, 32.0 min peak. Ions at m/z 1924.8 ($[M-6H]^{6-}$), 1649.7 ($[M-7H]^{7-}$) and 1732.5 ($[M-3H]^{3-}$) are consistent with a 2:1 **1**/(*Seq-1*/*Seq-2*) adduct.

Figure S7: A, CD spectra of **1** alone (———) and oligonucleotides in 100mM ammonium acetate buffer: *Seq-1* (———), *Seq-2* (— — —) and *Seq-4* (—●●—); **B**, CD spectra of **1** alone (———) and oligonucleotide: *Seq-1*/**1** (———), *Seq-2*:/**1** (— — —), *Seq-3*/**1** (—●●—).

Figure S8: A, Melting profile of *Seq-1* (black) & *Seq-2* (red) at 222 nm; **B,** Melting profile of **1**/*Seq-1* (black) and **1**/*Seq-2* (red) adducts at 260 nm; **C,** Melting profile of *Seq-1* and *Seq-2* annealed together, 20→95^{o}C (black), 20←95^{o}C (red); **D,** Melting profile of **1**/Seq-1 and Seq-2 annealed together, 20→95^{o}C (black), 20←95^{o}C (red)

Figure S9: Schematic representation of the *Seq-1* and *Seq-2* hairpin structures showing the alignment of base pairs (left), and the four possible adducts resulting from covalent binding of **1** to guanine-6 with the PBD A-ring oriented either to the 5′-end of the covalently modified guanine (centre, **A** or **C**) or the 3′-end (right, **B** or **D**).

Figure S10: Molecular models of the energy minimized structures of the four possible adducts of **1** covalently bound to guanine-6 of *Seq-1* and *Seq-2* with the A-ring of the PBDs oriented to either the 5′-end of the covalently bound guanine (**A** and **C** for *Seq-1* and *Seq-2*, respectively) or the 3′-end (**B** and **D** for *Seq-1* and *Seq-2*, respectively). The DNA is shown in stick model form (purple) and the ligand **1** in Van der Waals representation with atom type colouring.

Figure S11: Schematic representation of the four possible configurations of duplex DNA adducts formed from the covalent binding of **1** to both 5′-guanine-6 positions of annealed *Seq-1*/*Seq-2* DNA, with the PBD A-rings oriented either both to the 3′- (**A**) or 5′- (**C**) ends, or to opposing 3′/5′ (**B**) or 5′/3′ (**D**) ends.

Figure S12: Molecular models of the energy minimized structures of the four possible adducts of **1** covalently bound to 5′-guanine-6 of both strands of duplex *Seq-1*/*Seq-2* DNA according to **Figure S9**, with the PBD A-rings oriented to the 3′- or 5′-ends of the covalently bound guanine (**A** and **C**, respectively), or

the 3′-/5′- (**B**) or 5′-/3′-ends (**D**). The DNA is shown in stick model form (purple) and the ligand **1** in Van der Waals representation with atom type colouring.

Table S1: Structures and average masses of the single stranded (SS) oligonucleotides used in the study and the average mass of the adducts formed from their covalent interaction with **1** (as measured by MS)**.**

Label	SS DNA Sequence	Average Mass of SS DNA	Average Mass of 1:1 Adduct
Seq-1	5′-TATAAGAAAATCTTATA-3′	5200.49	5791.10
Seq-2	5′-TATAAGATTTTCTTATA-3′	5173.45	5764.06
Seq-3	5′-GCGCAGATTTTCTTATA-3′	5175.45	n/a
Seq-4	5′-TATAAIAAAATCTTATA-3′	5185.43	n/a

Table S2: Structure and average mass of the double stranded (DS) oligonucleotide duplex used in the study and the average mass of the adduct formed from its covalent interaction with **1** (as measured by MS).

Label	DS DNA Sequence	Average Mass of DS DNA	Average Mass of 2:1 Adduct
Seq-1/2	5′-TATAAGAAAATCTTATA-3′ 3′-ATATTCTTTTAGAATAT-5′	10373.94	11555.16

Table S3 : UV melting temperatures of *Seq-1, Seq-2, Seq-1 & Seq-2* annealed, 1/*Seq-1*, 1/*Seq-2* and 1/*Seq-1* & *Seq-2* annealed

Label	Melting Temperature (Tm)	
	Transition 1	Transition 2
Seq-1	311 K	337 K
Seq-2	314 K	321 K
1/*Seq-1*	337 K	352 K
1/*Seq-2*	343 K	355 K
Seq-1 and *Seq-2* annealed	314 K	331 K
1/*Seq-1* and *Seq-2* annealed	349 K	356 K