Nucleic Acids Research, 2017 **1**
doi: 10.1093/nar/gkx1211

Structural basis of DNA duplex distortion induced by thiazole-containing hairpin polyamides

Giacomo Padroni[1], John A. Parkinson[1], Keith R. Fox[2,*] and Glenn A. Burley[1,*]

[1]Department of Pure and Applied Chemistry, University of Strathclyde, Thomas Graham Building, 295 Cathedral Street, Glasgow G1 1XL, UK and [2]Centre for Biological Sciences, University of Southampton, Life Sciences Building 85, Southampton SO17 1BJ, UK

Received September 14, 2017; Revised November 14, 2017; Editorial Decision November 15, 2017; Accepted November 28, 2017

ABSTRACT

This manuscript reports the molecular basis for double-stranded DNA (dsDNA) binding of hairpin polyamides incorporating a 5-alkyl thiazole (Nt) unit. Hairpin polyamides containing an *N*-terminal Nt unit induce higher melting stabilisation of target dsDNA sequences relative to an archetypical hairpin polyamide incorporating an N-terminal imidazole (Im) unit. However, modification of the *N*-terminus from Im to Nt-building blocks results in an increase in dsDNA binding affinity but lower G-selectivity. A general G-selectivity trend is observed for Nt-containing polyamide analogues. G-selectivity increases as the steric bulk in the Nt 5-position increases. Solution-based NMR structural studies reveal differences in the modulation of the target DNA duplex of Nt-containing hairpin polyamides relative to the Im-containing archetype. A structural hallmark of an Nt polyamide•dsDNA complex is a more significant degree of major groove compression of the target dsDNA sequence relative to the Im-containing hairpin polyamide.

INTRODUCTION

Sequence-selective recognition of double-stranded, B-type DNA (dsDNA) by small to medium size molecules holds considerable promise as a general strategy for the exogenous modulation of gene expression (1–12). Hairpin pyrrole-imidazole polyamides (PIPs) are particularly powerful exemplars of this approach as their sequence-selective dsDNA binding is inherently programmable (13–17). Using a generalized set of pairing rules (Figure 1A), (13) the primary sequence of pyrrole (Py) and imidazole (Im) amino acid heterocycles defines the sequence selectivity of the target DNA duplex up to 24 base-pairs in length (18–33). One of the most important recognition profiles in hairpin PIP designs is the G-selective hydrogen-bonding recognition signature of the endocyclic N3 of the Im building block (ImN3) with the exocyclic N2 amine of G (34–36). Repulsive interactions between ImN3 and the electron-rich C, T and A nucleotides provide an auxiliary level of sequence discrimination (11). One distinct limitation of current hairpin PIPs for targeting G-rich sequences is their cell permeability, which generally decreases with increasing Im content (37,38). Although a variety of strategies exist to mitigate the enhanced polarity of Im-rich PIPs—such as modification of the C-terminus and the turn motifs (39–41)—these approaches tend to have an unpredictable impact on their overall physicochemical properties and toxicity (39,42,43).

Expanding the heterocyclic repertoire of current hairpin polyamide designs offers a powerful alternative for broadening the dsDNA binding lexicon currently available and provides a means to optimize cell uptake by tuning their physicochemical parameters (44–51). Thiazole-4-carboxylic acids (Nt) (46) are attractive analogues of Im (52) building blocks in this regard as they contain endocyclic heteroatoms (S1 and N3), which are potential hydrogen-bond acceptors for the exocyclic N2 amine of G (50,51,53–55). Moreover, facile synthetic routes are available to prepare Nt analogues containing alkyl substituents in the 5-position, thereby allowing the hydrophobicity of these building blocks to be tuned accordingly (56–58). Recent high-resolution X-ray crystal structures of cyclic 8-ring PIP•dsDNA complexes have revealed significant widening of the minor groove, and concomitantly, major groove compression, which results in an overall bending of the DNA duplex (17,59) relative to that observed when a hairpin polyamide binds to nucleosomal dsDNA (18,60). To date, solution-based structural characterizations of hairpin PIPs have been confined to six-ring hairpin and cyclic PIP•dsDNA complexes (14,15,61,62).

In this manuscript, we expand the heterocyclic repertoire of hairpin PIPs by replacing the *N*-terminal Im unit of **PA1** with Nt-building blocks (**PA2–4**, Figure 1B) (63). We show that Nt-building blocks incorporated in **PA2–4** differ in their relative binding preferences for a target G nucleotide in the minor groove with G-selectivity increasing

*To whom correspondence should be addressed. Tel: +44 141 548 2792; Email: glenn.burley@strath.ac.uk
Correspondence may also be addressed to Keith R. Fox. Tel: +44 23 8059 4374; Email: k.r.fox@soton.ac.uk

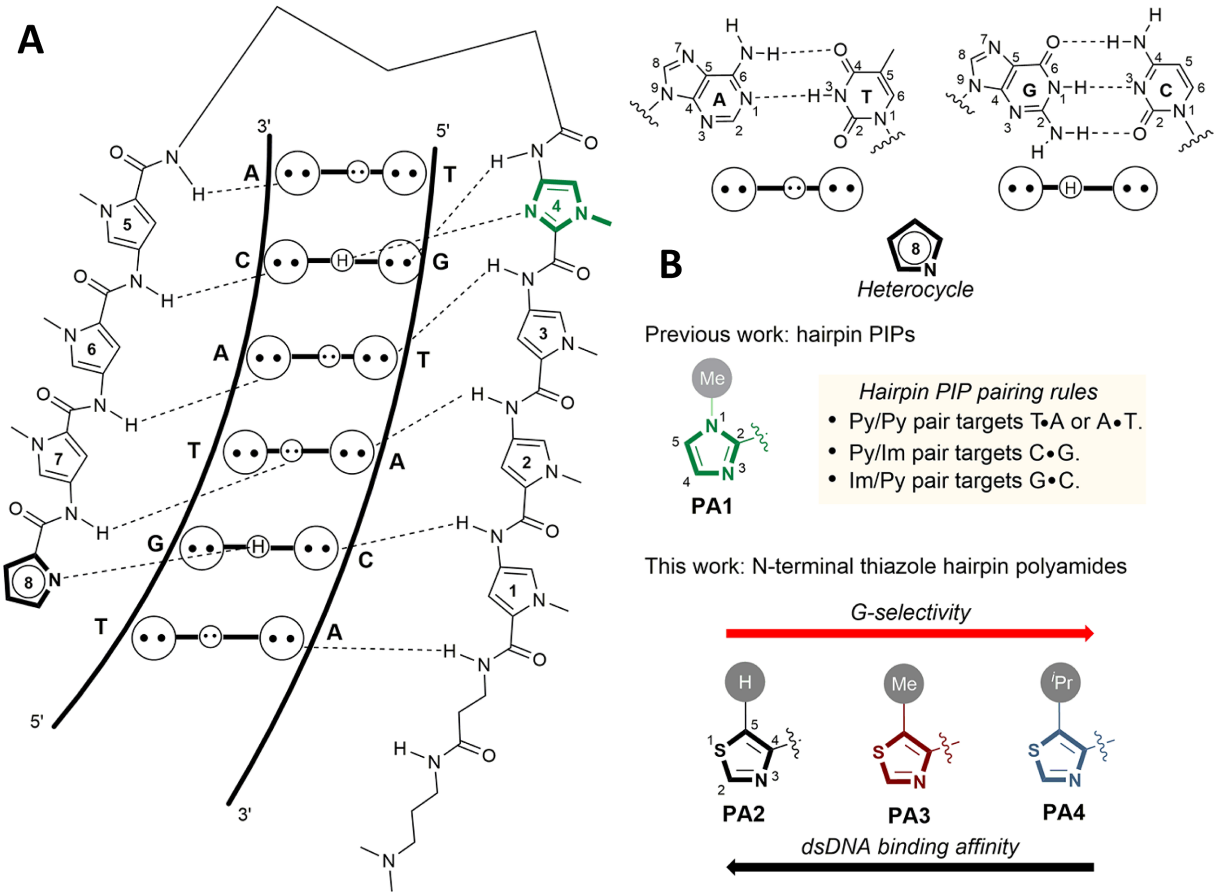


Figure 1. (**A**) General binding mode of hairpin polyamides. (**B**) Structures of polyamides used in this study.

with increasing steric bulk in the 5-position. Finally, NMR-derived models of **PA3**•dsDNA and **PA4**•dsDNA complexes show that Nt-containing polyamides (**PA3–4**) bind to dsDNA in a hairpin conformation and induce a more pronounced degree of major groove compression relative to an Im-containing **PA1**•dsDNA complex.

MATERIALS AND METHODS

Synthesis of polyamides PA1-4

PA1–4 were prepared by manual solid phase synthesis using Boc-β-Ala-PAM resin (0.5 mmol/eq, Sigma-Aldrich). HATU, HOAt and TFA were purchased from Fluorochem. DCC and Dp were purchased from Alfa Aesar.

UV melting analysis

Melting temperature analyses were performed on a Shimadzu UV spectrometer UV 1800 equipped with a thermo-controlled cell holder possessing a cell path length of 1 cm. An aqueous solution consisting of 10 mM monobasic phosphate, 10 mM dibasic phosphate, 100 mM NaCl, at pH 7.0 was used as the analysis buffer. The corresponding duplex DNA and the PA ligands were mixed at a stoichiometry of 1:1.75 providing a final dsDNA concentration of 1 μM for each experiment. Denaturating profiles were recorded at λ = 260 nm from 20 to 90°C, with a heating rate of 1°C/min. The reported melting temperatures were defined as averages of the maximum of first derivative plots of three independent denaturing profiles.

DNase I Footprinting

DNA fragments for footprinting were prepared by ligating synthetic oligonucleotides, which had been designed to contain six heptanucleotide potential ligand binding sites, each separated by GCGC, into the BamHI site of pUC19 (64,65). Successful clones were sequenced by Eurofins MWG Operon, Ebersberg, Germany. The sequence of the inserted fragment is shown in Supplementary Figure S1.

Radiolabelled DNA fragments were obtained by cutting the plasmid with HindIII and SacI and labelling the 3′-end of the HindII site with α-^{32}P[dATP] using exo-Klenow fragment of DNA polymerase I (New England Biolabs). The labelled DNA was separated from the remainder of the plasmid on a 6% polyacrylamide gel, eluted and dissolved in 10 mM Tris–HCl pH 7.5, containing 0.1 mM EDTA at a concentration of at least 10 cps/μl as determined on a hand-held Geiger counter (∼10 nM DNA).

DNase I footprinting was performed by mixing 1.5 μl radiolabelled DNA with 1.5 μl ligand (diluted in 10 mM Tris–HCl containing 10 mM NaCl, from a 10 mM stock solution in DMSO). Samples were equilibrated for at least 30 min

before adding 2 μl DNase I (typically 0.01 units/ml) which had been diluted in 20 mM NaCl, 2 mM $MgCl_2$, 2 mM $MnCl_2$. The reaction was stopped after 1 min by adding 4.5 μl 80% formamide containing 10 mM EDTA, 1 mM NaOH and bromophenol blue. Samples were denatured by heating at 100°C for 3 min before cooling on ice and loading onto a 40 cm, 8% denaturing polyacrylamide gel containing 8 M urea, which was run at 1500 V for ∼90 min. The gel was then fixed in 10% (v/v) acetic acid, dried onto 3MM filter paper and exposed overnight to a phosphor screen. The screen was scanned using a Typhoon phosphorimager.

NMR sample preparation

Two separated batches of HPLC purified self-complementary $d(CGATGTACATCG)_2$ (**ODN1**) were purchased from Eurogentec. Each sample was dissolved in 1.2 ml of H_2O buffered with 100 mM phosphate at pH 7.4 and containing 0.1 mM deuterated trimethylsilylpropionic acid sodium salt-d_4, divided in two aliquots (600 μl each) and freeze-dried to dryness. The samples were re-dissolved in 600 μl of 90% H_2O with 10% D_2O and in 99% D_2O. The solutions in two separate 5 mm high-precision NMR tubes were used for **ODN1** data acquisition. After acquisition, the two 99% D_2O samples were freeze-dried and re-dissolved in 10% D_2O/90% H_2O. A concentrated solution of the **PA1–3-4** (16 mM in up to 10% d_6-DMSO in water) was added in small aliquots. After each addition of polyamide to the buffered solution containing **ODN1**, 1D ^{1}H NMR spectra were acquired until the end point of the titration was reached as determined by inspection of the imino proton resonance region of the data. After acquisition of the full set of data in 10% D_2O/90% H_2O, the samples were freeze-dried and re-dissolved in D_2O to complete the acquisition of all data sets.

NMR data acquisition

NMR data were typically acquired on a Bruker AVANCE-II$^+$ 600 NMR spectrometer operating at a magnetic field strength of 14.1 T (600.13 MHz for proton resonance) under TopSpin (version 3.5 patch level 5, Bruker, Reinstetten, Germany) running on a Hewlett Packard Z420 workstation under Windows Professional version 7 (Microsoft Inc.). A triple resonance probehead [TBI-z] equipped with actively shielded z-gradient coil for delivery of pulsed field gradients was manually tuned for a [^{1}H, ^{13}C, ^{31}P] frequency configuration on each sample studied to allow uninterrupted data acquisition to progress using all channels in their turn. The probe temperature was maintained at 298 K in all instances.

One-dimensional (1D) ^{1}H NMR data were acquired using either presaturation (pulse program zggppr) or double pulsed field gradient spin-echo (pulse program zgesgp) to eliminate the residual solvent resonance. Data were typically acquired for 90% H_2O samples with between 64 and 256 transients into 32K data points (acquisition time: 1.09 s) over a frequency width equivalent to 20.0276 ppm centred at $\delta^1H = 4.692$ ppm. Hard proton 90° pulses were typically calibrated at 9.6 W with an average value of 9.7 μs across the samples studied. Data acquired for D_2O-dissolved samples were acquired over a frequency width equivalent to 12.0166 ppm (acquisition time: 2.27 s).

One-dimensional (1D) ^{31}P–{^{1}H} (66) NMR spectra were acquired over a frequency width of 3 ppm into 512 data points (acquisition time: 0.35 s) centred at $\delta^{31}P = -1.185$ ppm with the decoupler offset set to $\delta^1H = 4.00$ ppm. Between 512 and 4096 transients were acquired with a relaxation delay of 0.5 s.

Proton inversion recovery profiles were measured on one sample to gauge the requirement for relaxation delays during the acquisition of 2D NOESY data sets used in structure calculations. The largest T_1 values were measured as 5.65 s (for adenine H2 protons). Relaxation delays were typically set to 2.0 s for most data acquisitions with the exception of NOESY data sets whose relaxation delays were typically set to 7 s to allow at least one T_1 period of magnetization recovery without excessively extending 2D NOESY data acquisitions.

Two-dimensional (2D) NMR data sets were acquired as follows for samples dissolved in D_2O. Proton-only data were acquired over frequency widths equivalent to 12 ppm (7194 Hz) in both F2 and F1 centred at an offset of $\delta^1H = 4.702$ ppm. 2D DQF-COSY data (pulse program: cosydfphpr) were acquired with 16 transients for each of 1024 States-TPPI t_1 increments for acquisition times (aq) of 285 ms (ω_2) and $aq_{max} = 71$ ms (ω_1) and a total data accumulation time of 10 hrs. 2D TOCSY data (pulse program: dipsi2phpr) were acquired into 2048 data points (ω_2 acquisition time = 142 ms) with eight transients for each of 512 States-TPPI t_1 increments (aq_{max} $\omega_1 = 35.4$ ms) using a spin-lock time of 70 ms. The same conditions were used to acquire 2D NOESY data (pulse program: noesyphpr) with mixing times ranging from 100 ms to 250 ms. 2D [^{1}H, ^{31}P] correlation data were acquired for assignment purposes only according to the constant-time procedure (pulse program: nahpctco) (67). Data were acquired over frequency widths equivalent to 7 ppm (ω_2) and 2 ppm (ω_1) centred at $\delta^1H = 4.701$ ppm and $\delta^{31}P = -1.185$ ppm respectively with 128 transients for each of 128 State-TPPI t_1 increments and a relaxation delay of 1.5 s to give a total data accumulation period of 2.75 h. 2D [^{1}H, ^{13}C] correlation data were acquired using echo/anti-echo-TPPI data acquisition schemes with sensitivity improvement, with (pulse program: hsqcedetgpsisp.2) and without (pulse program: hsqcetgpsisp.2) DEPT-editing and without/with non-uniform sampling (NUS) respectively. NUS data were acquired with 25% sampling of 512 t_1 increments into 2048 data points for frequency widths equivalent to 10 ppm (F2) and 170 ppm (F1) with 128 transients per t_1 increment for a total data accumulation time of 10 h. DEPT-edited HSQC data were acquired with a traditional acquisition mode using the same conditions and 64 transients per t_1 increment for a total data accumulation time of 20 hrs. For samples dissolved in 90% H_2O, 2D NOESY data were acquired with excitation sculpting for solvent suppression (pulse program: noesyesgpph) using State-TPPI for 1024 t_1 increments and 4096 data points over F2 and F1 frequency widths equivalent to 25 ppm centred at the solvent frequency resonance. Data were acquired with 16 transients per t_1 increment with a relaxation delay of 7 s between transients for a total data accumulation time of 33 h. Mixing times were defined as 180 and 250 ms. All data were processed according to established NMR data process-

ing protocols and transferred into SPARKY for data assignment and reduction in preparation for molecular structure calculations.

Preliminary molecular dynamics

Preliminary molecular dynamics for distance restraint calculations of free **ODN1** and the complexes were run in explicit solvent using the SANDER module of AmberTools16 (68). The starting model of **ODN1** was generated using the UCSF CHIMERA program (69). The structures of each polyamide•**ODN1** complex were generated using an X-ray-derived structure of a cyclic polyamide•dsDNA complex as the starting structure (PDB ID 3OMJ) (59). Input files (prmtop and inpcrd) were generated using the nucleic acid force field (LEAP module) ff99bsc1 and GAFF for **ODN1** and ligands respectively. 22 and 21 Na^+ ions for the free **ODN1** and the complexes were added to neutralize the negative charge of the resultant structures. An octahedral box with outer edges of ∼10.0 Å of TIP3P water was placed around the structures. A 10 Å cut-off for non-bonded interaction were applied with the particle mesh Ewald (PME) method to account for long range electrostatic interactions. Initial minimizations of 1000 steps (500 steps of steepest descent and 500 steps of conjugate gradient minimization) were performed on each complex keeping the solute fixed with a 500 kcal/mol·Å^2 force constant. The entire system was then minimized over 2500 steps. A first MD run of 10 000 steps was performed (20 ps) using the SHAKE algorithm. The initial temperature of 0 K was brought gradually up to 300 K and kept constant using Langevin dynamics and keeping the solute fixed with weak restraints (10 kcal/mol·Å^2). The RST file was then used for a second MD run of 50 000 steps (100 ps) removing the restraints. A snapshot was recorded every 100 steps to generate the trajectory file. RST files were checked for general consistency with the NOE data and used for distance calculations

NMR distance restraints

In all cases, the distance restraints were obtained from 2D [^{1}H–^{1}H] NOESY (250 ms). NMR-FAM SPARKY was used for the assignment process (70). Non-overlapping peaks were integrated using the Gaussian fit to generate the intensity file for MARDIGRAS (71). The RST files obtained from preliminary molecular dynamics were converted to pdb file format to be used in MARDIGRAS through CORMA.IN using an isotropic model and relative error of 10%. The correlation times used were roughly established in each case checking MARDIGRAS calculation outputs from 1 to 8 ns. Absolute un-normalized noise was defined as a fraction of the smallest peak that could be integrated in the spectrum. The methyl jump 3 model was used for all models. MARDIGRAS was run for 50 cycles. The resulting .dst files were associated with RANDRESTR to generate an average distance file. These coordinates were converted to restraints to be used in AMBER using M2AHOMO or MARDI2AMBER and filtering out selected out-of-range distances. The force constants applied were 10 kcal/mol·Å for lower and upper bounds. The width of the parabola was set to be 2 Å in all the cases. Calculations were repeated at the beginning of each molecular dynamics cycle as reported below.

NMR-restrained molecular dynamics

Initial molecular dynamics were run using the generalized Born implicit-solvent model. After initial minimization (500 total steps, 250 steepest descent and 250 conjugated gradient), 100 ps MD were calculated keeping the temperature constant. NMR restraints were gradually increased over the first 20 ps and kept constant during the remaining 80 ps. After error checking, the RST file was used for a second cycle of MARDIGRAS distance calculation. The second round of molecular dynamics was carried out in explicit solvent applying the same condition reported for preliminary molecular dynamics. Newly generated NMR restraints were gradually applied over the first 20 ps of MD simulation and kept constant over the remaining 80 ps. The RST file was then used for the last cycle of MARDIGRAS calculation and the obtained restraints used for the third MD simulation. Production runs of 500 ps were obtained. The production runs were analyzed using the cpptraj module of AMBERTools16. Clustering of the last 400 ps was carried out using UCSF Chimera to produce ten representative structures, which were minimized and deposited in the ProteinDataBank (PDB accession codes 5OCZ, 5OE1, 5ODF and 5ODM). An average minimized structure obtained from this ensemble was obtained using the CPPTRAJ module of AMBERTools16 and used to calculate the DNA parameters taking a complementary approach using Curves+ and 3DNA software (72,73).

RESULTS AND DISCUSSION

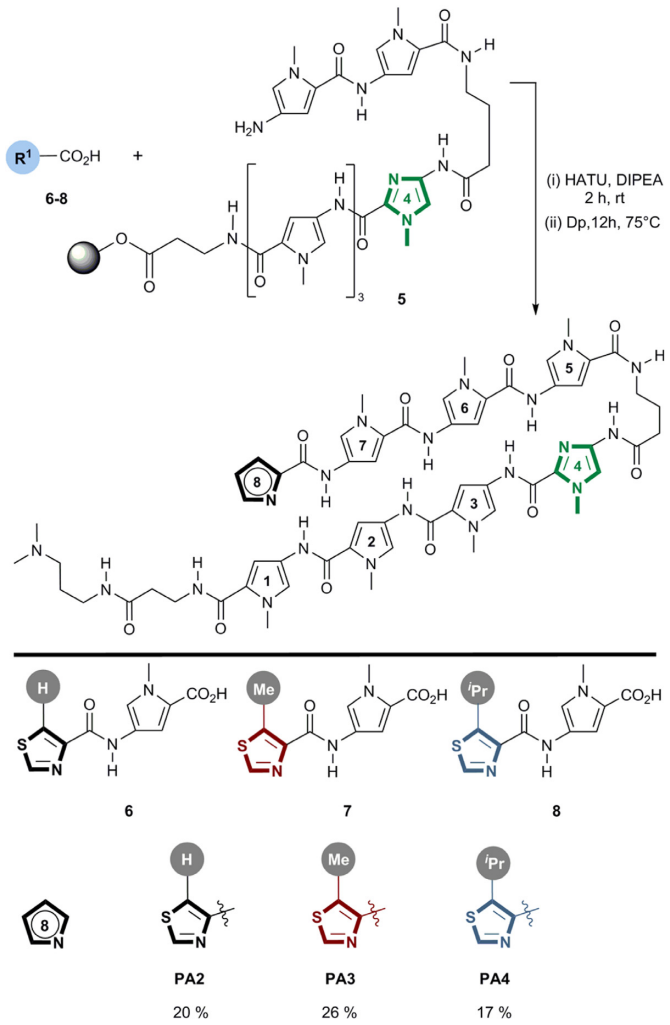
The heterocyclic core of the anti-tumour hairpin PIP (25,63) **PA1** was chosen as our benchmark design to explore the impact of replacing the N-terminal Im unit (Im8) with Nt analogues. **PA1** has an established high affinity binding profile for 5'-WGWWCW (where W = A/T), which is recognized by the Androgen Receptor (59,63).

Synthesis of hairpin polyamides PA1-4

PA1 was prepared *via* Boc-based solid phase synthesis on a β-Ala PAM resin (0.5 mmol/eq) (74–78). For the synthesis of **PA2–4**, the resin-bound hexamer **5** was used (Scheme 1). Amide coupling of dimer carboxylic acids **6–8** (Supplementary Scheme S1) (79) with **5** using HATU as the condensing agent, followed by aminolytic cleavage with (*N,N*-dimethylamino)propylamine (Dp) afforded **PA2–4** after purification by RP-HPLC.

G-Selectivity of PA2-4 binding to dsDNA sequences increases with Nt steric bulk

The dsDNA binding affinity and sequence selectivity of **PA1–4** were determined using a combination of DNase I footprinting and UV–vis melting experiments (65). Footprinting of **PA1** [3–0.01 μM] on the radiolabelled DNA fragment containing the 75 base-pair insert with variants of the putative binding sites revealed a bias towards binding



Scheme 1. Synthesis of **PA1–4**.

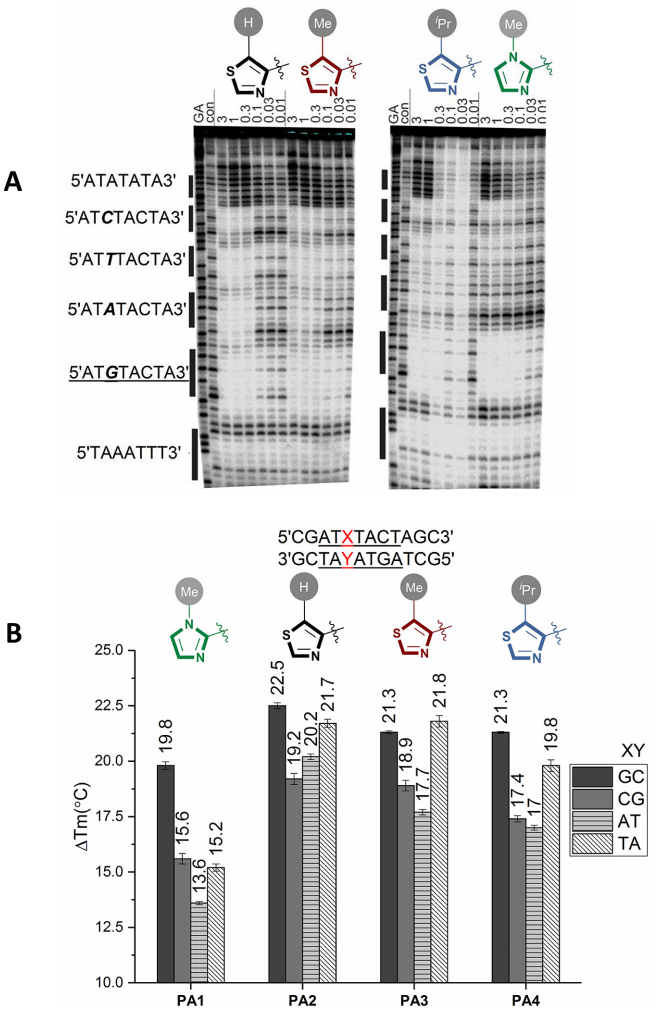


Figure 2. (**A**) DNase I footprint titration using an engineered DNA 75 bp fragment in the presence of **PA1–4** at concentrations [3–0.01 μM]. (**B**) UV melt stabilization of **PA1–4**.

to 5'-AT***G***TACT (Figure 2A, Supplementary Figure S1). For **PA2**, a more promiscuous DNA binding profile was observed, characterized by footprints at 5'-AT***T***TACT, 5'-AT***A***TACT and 5'-AT***G***TACT. In contrast, a bias towards G-selective binding was observed for **PA4**, with **PA4** exhibiting a clear preference for 5'-AT***G***TACT; i.e. a similar trend observed for **PA1**. To gain further insight into the relative sequence preferences of **PA1–4**, UV–vis melting experiments were conducted using dsDNA sequences covering the four possible nucleotides (5'-AT***X***TACT, where **X** = G, C, A or T; Figure 2B; Supplementary Tables S1 and S2). A 19.8°C stabilization was observed for **PA1** in the presence of the target sequence 5′-AT***G***TACT (Figure 2B), with a clear binding preference for G at this position (Supplementary Tables S1 and S2). Enhanced thermal stabilization for all three Nt-containing polyamides (**PA2–4**) relative to **PA1** was observed with **PA2** exhibiting the highest degree of thermal stabilization of the polyamide series. However, enhanced dsDNA binding affinity results in sequence selectivity for G nucleotides being somewhat compromised. In agreement with the sequence selectivity observed in the DNase I footprinting experiments, sequence discrimination of **PA4** was higher than **PA2–3**, with a $\Delta\Delta T_m$ of 1.5°C stabilization for 5′-AT***G***TACT relative to 5′-AT***T***TACT.

NMR structural characterization of dsDNA●polyamide complexes

Data arising from DNase I footprinting and UV–vis melts revealed a G-selective binding profile for Nt-containing polyamides as steric bulk is increased in the 5-position. NMR studies were undertaken to gain structural insight into the binding profile of **PA3–4** compared to **PA1** when in complex with the dodecamer oligodeoxyribonucleotide (**ODN1**) sequence d(CG**AT*G*TACA**TCG)$_2$ (Figure 3A and B). A 1:1 stoichiometric complex for each polyamide was observed by 1D ^{1}H NMR and characterized by a doubling of the number of DNA imino proton resonances and a single set of polyamide resonances (Figure 3C–F). Assignment of selected NMR data for **ODN1** and the polyamide complexes **PA1●ODN1**, **PA3●ODN1** and **PA4●ODN1** was achieved by a combination of 2D [^{1}H,^{1}H] NOESY (150 ms and 250 ms mixing times), DQF-COSY, TOCSY, 2D [^{1}H, ^{13}C] HSQC and 2D [^{1}H, ^{31}P] COSY for samples in 99% D_2O

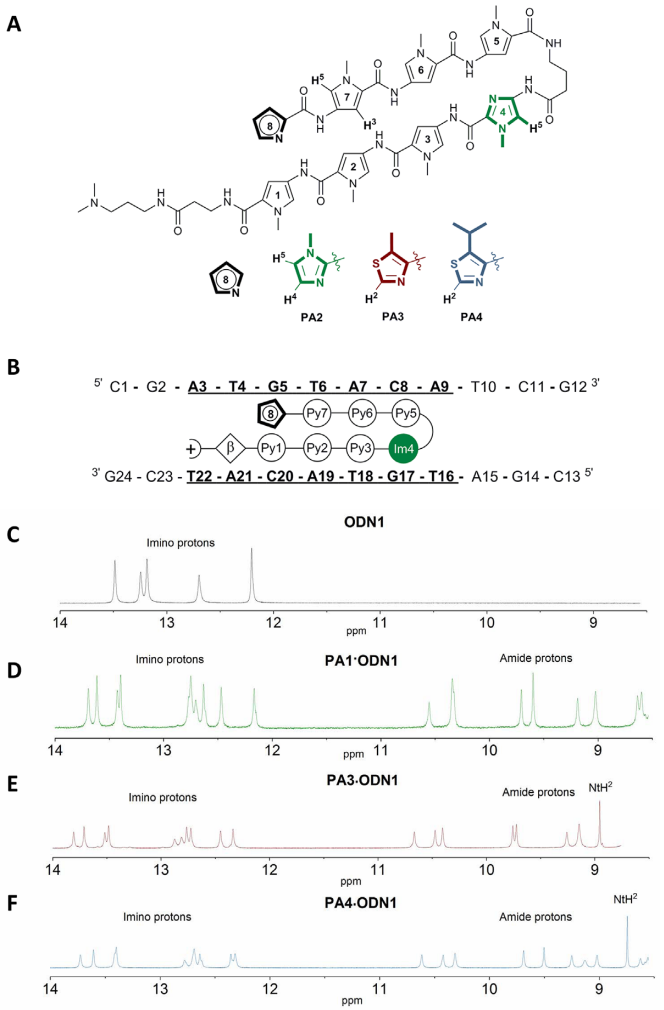


Figure 3. (**A**) Schematic representations and nomenclature of **PA1**, **PA3** and **PA4** in complex with **ODN1**. (**B**) Schematic depiction of a hairpin polyamide in complex with the target DNA duplex. (**C**) Selected region of 1D ^{1}H NMR of free **ODN1**, (**D**) **PA1•ODN1**, (**E**) **PA3•ODN1**, (**F**) **PA4•ODN1**.

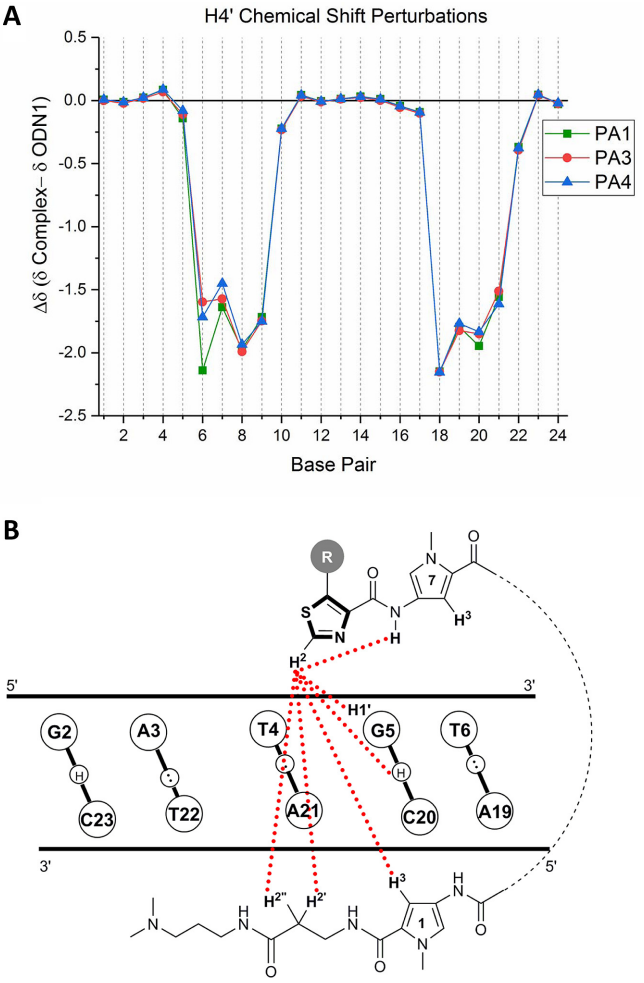


Figure 4. (**A**) Chemical shift perturbations plots of H4′ protons of **ODN1**. (**B**) Schematic representations of strong inter and intramolecular NOE cross-correlations of NtH^2 (**PA3** and **PA4**) in complex with **ODN1**.

and 2D [^{1}H, ^{1}H] NOESY for samples in 10% D_2O/90% H_2O. For **PA3•ODN1** and **PA4•ODN1**, the NtH^2 proton was diagnostic (8.97 ppm for **PA3**; 8.87 ppm for **PA4**) and used as a starting point for the structural assignment of these polyamides. A strong NOE cross-correlation from NtH^2 to the adjacent Py7NH and Py7H^3 was observed in conjunction with TOCSY correlations to identify Py7H^5 (Supplementary Figure S2A). NOE correlations between Py7H^5/Py7H^3 and (Me)Py7 completed the assignment of Py7 resonances. This general assignment pattern was found for all three polyamide complexes. Finally, 2D [^{1}H, ^{1}H] TOCSY and COSY NMR data were used to assign the alkyl protons of γ-turn, β-alanine and Dp moieties.

Nt polyamides bind to target dsDNA sequences in a hairpin conformation. Upfield perturbation of the H4' chemical shift was observed for nucleotides located in the central 5'-AT***G***TACT sequence, which is a diagnostic marker for minor groove binding (Figure 4A) (14–16,34). This was corroborated by strong NOE correlations between amide NH, PyH^3 and H1' sugar protons of the adjacent base (Supplementary Figure S2B). NOE cross-peaks from NtH^2 of both Nt building blocks to β-alanine and Py1 protons in **PA3•ODN1** and **PA4•ODN1** suggest that both Nt-containing polyamides bind to their target sequence in a hairpin conformation (Figure 4B). NOE data also provide insight into the nature of Nt-binding to G5 in the minor groove. NOE correlations from NtH^2 of **PA3** and **PA4** to G5H1′ and the G5N2 exocyclic amine suggest both Nt building blocks have the endocyclic N3 directed towards the floor of the minor groove. The presence of a hydrogen bond between NtN3 and the exocyclic G5N2 amine is implied by slower exchange rate of the exocyclic N–H of G5N2 not involved in Watson–Crick base-pairing (G5H22) and NOE cross-correlations from G5H22 to G5H1 and T6H3, all of which are absent in the NMR data for the free **ODN1** duplex (Figure 5A–D). Relevant differences in the chemical shift of G5H22 in **PA1•ODN1** (7.87 ppm) and Nt polyamides (**PA3•ODN1** 6.76 ppm and **PA4•ODN1** 6.59 ppm) indicate that the presence of the endocyclic sulphur in Nt building blocks influences Nt•G5N2 pairing. Taken col-

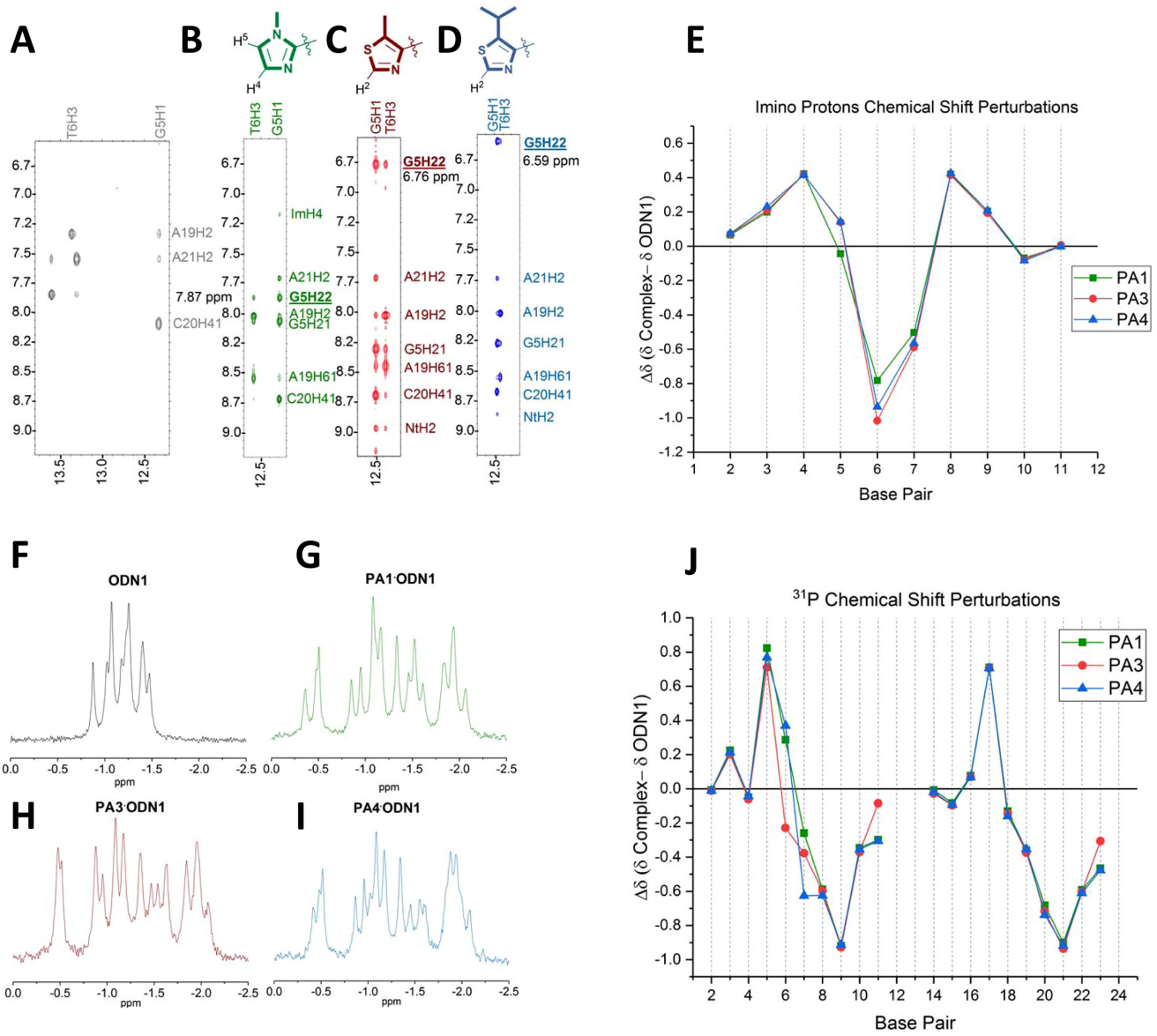


Figure 5. (**A**–**D**) Strip plots from 2D [^{1}H, ^{1}H] NOESY NMR data of free **ODN1** and **PA1●ODN1**, **PA3●ODN1** and **PA4●ODN1**. (**E**) Chemical shift perturbations plots of imino proton resonances of **ODN1**. (**F**–**I**) ^{31}P-{^{1}H} NMR spectra of free **ODN1**, **PA1●ODN1**, **PA3●ODN1** and **PA4●ODN1** complexes. (**J**) Chemical shift perturbations plots of phosphorous atom NMR resonances of **ODN1**.

lectively, the NMR data of each polyamide•dsDNA complex correlates with the implied formation of a hydrogen bond between the exocyclic amine of G5N2 (i.e. G5H22) and the endocyclic N3 of Im8 and both Nt building blocks.

Differences in DNA duplex distortion imparted by Nt-containing polyamides. Analysis of chemical shift perturbations of **PA3●ODN1** and **PA4●ODN1** reveal subtle, but distinct differences in DNA duplex distortion relative to **PA1●ODN1**. For example, a significant upfield shift of H4' (T6 and A7) and imino (G5H1) protons in the **PA1●ODN1** complex is observed relative to **PA3●ODN1** and **PA4●ODN1** (Figures 4A and 5E). In the case of T6H3 and T18H3 imino protons, a more significant upfield shift is evident for the Nt-containing polyamides **PA3●ODN1** and **PA4●ODN1** relative to **PA1●ODN1** (Figure 5E). ^{31}P NMR spectra were assigned by scalar cross-correlations from the sugar H3' to the O3′-P. ^{31}P NMR spectra of the polyamide•dsDNA complexes revealed broadening of the ^{31}P resonances with downfield shifts of G5P and G17P, and upfield shifts of A9P and A21P, standout features relative to free **ODN1** (Figure 5F–J). Further distinct chemical shift changes for G5P, T6P and A7P were observed across all three complexes indicative of different levels of phosphodiester backbone distortion imparted by each polyamide (Figure 5J, Supplementary Table S22).

NMR retrained molecular dynamics of polyamide•dsDNA complexes

NMR-derived structures of free **ODN1** and **PA1–3-4** complexes were obtained by NMR restrained molecular dy-

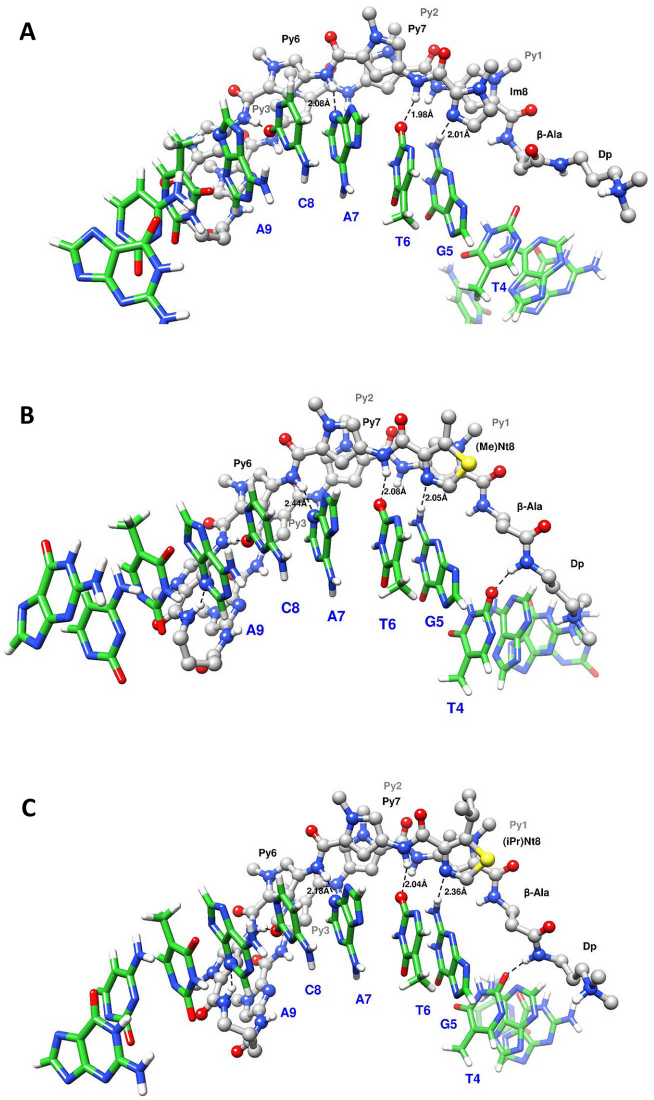


Figure 6. View of the curvature and anticipated hydrogen-bonding pattern between (**A**) **PA1●ODN1**, (**B**) **PA3●ODN1** and (**C**) **PA4●ODN1**.

namics (57). The ten most representative geometric conformational states of each model obtained from molecular dynamics simulations were then minimized and chosen as representative ensembles (Supplementary Figure S3). Although all three polyamides retain a ring-over-ring conformation in each complex, (14–17,59) the Nt-containing polyamides (**PA3–4**) adopt a distinct dsDNA binding geometry relative to **PA1**. In the case of **PA1●ODN1**, the smaller Im8 building block induces a more pronounced curvature within the minor groove relative to **PA3●ODN1** (Figure 6A and B). In the **PA4●ODN1** complex, the steric clash between isopropyl and (Me)Py1 prevents close contact between NtN3 and the exocyclic G5N2 amine (Figure 6C). In addition, more pronounced base inclination and axis bending is seen in the **PA1●ODN1** complex compared with **PA3●ODN1** and **PA4●ODN1** (Supplementary Figure S4, Supplementary Tables S24, S28, S32 and S36).

Analysis of the minor and major groove geometries in each complex reveals a number of distinct structural features. **PA1●ODN1** and **PA3●ODN1** complexes are characterized by an enlarged minor groove relative to **PA4●ODN1** (17,59). This is most likely due to deeper minor groove penetration of the less hindered Im8 and (Me)Nt8 (Figure 7A, Supplementary Table S23). An unexpected observation is the structural impact on the major groove width when **PA1●ODN1** is compared with **PA4●ODN1**. A more pronounced narrowing of the major groove is observed at the site of (iPr)Nt binding to G (i.e. NtN3●G5N2), changing from 22.4 Å in the free duplex (i.e. **ODN1**) to 12.2 Å in **PA4●ODN1** (Figure 7B, Supplementary Table S23). In contrast, only a slight narrowing (21.0 Å) of the major groove is observed for the **PA1●ODN1** complex, which is consistent with crystal structural studies of a cyclic PIP●dsDNA complex (59). For the **PA3●ODN1** complex, an intermediate change (16 Å) in the major groove width relative to **PA4●ODN1** suggests that the increased steric bulk of the isopropyl substituent in **PA4●ODN1** plays a significant role in inducing major groove compression of the DNA duplex.

Structural changes to the Watson-Crick pairing profile in each polyamide●DNA complex also differ. In **PA1●ODN1**, the larger propeller twist in T4●A21 and increased buckle of T6●A19 contribute in inducing a widening of the major groove, which is also typified by the larger shift changes of the imino and H4' protons (Supplementary Tables S25, S29, S33, S37). In the case of the **PA4●ODN1** complex, the sterically more encumbered (iPr)Nt induces significant changes in the χ torsional angle (O4'–C1'–N9–C4) of the G5 (–87.9°) nucleosidic bond relative to **PA3●ODN1** (–103.2°) and **PA1●ODN1** (–117.1°; Supplementary Figure S3 and Tables S27, S31, S35, S39). Finally, the significant deviation in the ^{31}P resonances of G5P, T6P and A7P in **PA4●ODN1** relative to **PA1●ODN1** is evidence of significant changes in the phosphodiester bond angles ζ (C3'–O3'–P-O5') and ϵ (C4'–C3'–O3'-P) located in close proximity to the site of Nt-binding (Supplementary Figure S5 and Tables S27, S31, S35, S39).

Rationalisation of the differences in DNA selectivity and duplex distortion of Nt-containing polyamides

These experiments were designed to probe how structural changes to the *N*-terminal heterocycle of a hairpin PA affects overall dsDNA binding affinity and sequence selectivity. Our data reveal that the replacement of an *N*-terminal Im unit for an Nt heterocycle enhances overall dsDNA binding, but at the expense of sequence selectivity for a target G nucleotide. We hypothesise that the overall differences in hydrophobicity, hydrogen bond strength and steric bulk between Nt and Im heterocycles are contributing factors. Underpinning all of these is the replacement of the nitrogen N1 of Im with a sulfur (S1) atom, which in Nt is also oriented differently, occupying the same position of the smaller carbon (C5) atom in the Im unit. This key structural change appears to contribute to a general loss of G-selectivity for **PA2–3** when there is minimal steric bulk in the 5-position relative to the archetypical hairpin **PA1**. This has implications in, for example, weaker hydrogen bonding between the exocyclic N-H of G5N2 (i.e.

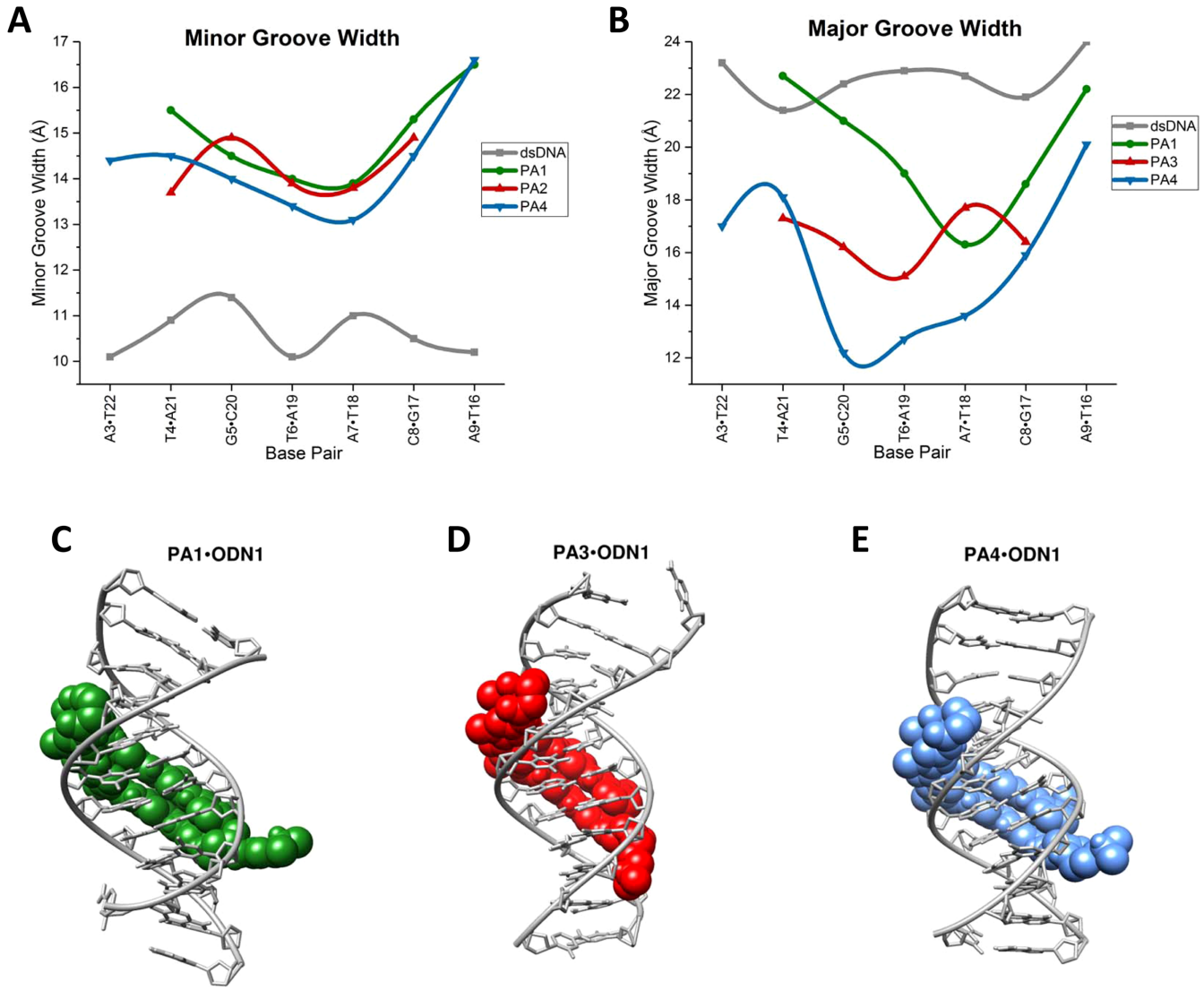


Figure 7. (**A**) Minor Groove and (**B**) Major Groove width of **ODN1** and **PA1●ODN1**, **PA3●ODN1**, **PA4●ODN1**. Major groove view of (**C**) **PA1●ODN1** (D) **PA3●ODN1** (E) **PA4●ODN1**.

G5H22) and the N3 of Nt compared to the more basic ImN3, which is implied by a downfield shift of G5H22 for the Im-containing polyamide (**PA1●ODN1**, 7.87 ppm) compared to Nt polyamide (**PA3●ODN1** 6.76 ppm, and **PA4●ODN1** 6.59 ppm). The higher dsDNA melting stabilisation observed for Nt polyamides (**PA2–4**) compared to the archetypical Im polyamide (**PA1**) is attributed to a stronger overall hydrophobic interaction for the Nt-containing **PA2–4**, which appears to compensate for the weaker hydrogen bonding interaction with G5H22, and is most likely responsible for an increase in non-specific dsDNA binding for **PA2** and **PA3**. However, the installation of an isopropyl group (i.e. **PA4**) increases G selectivity *with respect to PA2–3*. Whilst these changes in G selectivity cannot be rationalised by this study alone, we hypothesise that the steric effect of the isopropyl group in **PA4** biases the binding towards wider minor groove G●C regions.

A structural rationalization for the observed dsDNA binding differences of **PA3–4** relative to **PA1** would likely involve an interplay between the geometrical differences of the Nt and Im heterocycles, and their inherent differences in their respective non-covalent interactions with dsDNA in the minor groove. For example, the stronger hydrogen-bonding as well as the smaller overall dimension of the Im8 unit in **PA1** results in deeper penetration into the dsDNA minor groove. From a topological perspective, this induces an overall increase in dsDNA bending in the **PA1●ODN1** complex (Supplementary Figure S4), which concomitantly compresses the major groove around the central T6●A7●C8 region of the DNA duplex. In the case of **PA3●ODN1**, the weaker hydrogen-bonding interaction as well as the steric effect of the sulfur in (Me)Nt reduces the overall bending (Supplementary Figure S4), which in turn compresses the major groove over a larger DNA region (T4●G5●T6●A7●C8). The 5-position substituent of the (iPr)Nt building block of **PA4●ODN1** only partially affects the overall DNA bending (Supplementary Figure S4**).** A unique characteristic is the larger steric bulk of the

(iPr)Nt unit induces a more extensive compression of the major groove (in particular over G5), which is characterised by an increased hydrogen bond distance between the N3 of (iPr)Nt and G5H22 arising from reduced minor groove penetration. Taken collectively, the steric effect of the isopropyl Nt substituent in **PA4** enhances the selective recognition of Nt-containing hairpin polyamides by compensating for the weaker hydrogen bonding interaction with the exocyclic N2 of G with an increase in the overall hydrophobicity of this *N*-terminal building block.

CONCLUSION

In conclusion, our data shows that hairpin polyamides which incorporate an *N*-terminal Nt unit (**PA3–4**) induce more pronounced compression of the major groove of target DNA duplexes compared to archetypical hairpin PIPs and untethered minor groove binders (e.g. **PA1**) (51). The enhanced level of major groove compression in **PA3●ODN1** and **PA4●ODN1** complexes is akin to Nt-containing polyamides 'pinching' the target dsDNA sequence at the polyamide extremities. These differences in dsDNA distortion could offer new opportunities for modulating gene expression either *via* disruption of transcription-dsDNA binding or by blocking transcriptional elongation (80–82). We envisage that further tuning of dsDNA binding affinity and sequence selectivity is indeed possible at other sites within the polyamide scaffold, which could be used as a design proxy for their continuing development as gene modulatory agents across a range of cell types.

AVAILABILITY

Atomic coordinates for the reported structures have been deposited with the Protein Data bank under accession numbers 5OCZ, 5OE1, 5ODF and 5ODM.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank Dr Nikolai Ulyanov for his assistance in using the MARDIGRAS software.

FUNDING

This work was supported by the University of Strathclyde studentship scheme [G.P.] and by the STFC "A Dynamic Perspective on DNA-binding" [ST/M000125/1]. Funding for open access charge: university funds.

Conflict of interest statement. K.R.F. is Senior Executive Editor of *Nucleic Acids Research*. The other authors declare no competing financial or academic conflict of interest.

REFERENCES

1. Dervan,P.B., Doss,R.M. and Marques,M.A. (2005) Programmable DNA binding oligomers for control of transcription. *Curr. Med. Chem. Anticancer Agents*, **5**, 373–387.
2. Han,Y.-W., Sugiyama,H. and Harada,Y. (2016) The application of fluorescence-conjugated pyrrole/imidazole polyamides in the characterization of protein-DNA complex formation. *Biomater. Sci.*, **4**, 391–399.
3. Baraldi,P.G., Bovero,A., Fruttarolo,F., Preti,D., Tabrizi,M.A., Pavani,M.G. and Romagnoli,R. (2004) DNA minor groove binders as potential antitumor and antimicrobial agents. *Med. Res. Rev.*, **24**, 475–528.
4. Nanjunda,R. and Wilson,W.D. (2001) Binding to the DNA minor groove by heterocyclic dications: from AT-specific monomers to GC recognition with dimers. *Curr. Protoc. Nucleic Acid Chem.*, doi:10.1002/0471142700.nc0808s51.
5. Withers,J.M., Padroni,G., Pauff,S.M., Clark,A.W., Mackay,S.P. and Burley,G.A. (2016) Minor groove binders as therapeutic agents. *Reference Module in Chemistry, Molecular Sciences and Chemical Engineering, Comprehensive Supramolecular Chemistry II*, 149–178.
6. Wemmer,D.E. (2000) Designed sequence-specific minor groove ligands. *Annu. Rev. Biophys. Biomol. Struct.*, **29**, 439–461.
7. Barrett,M.P., Gemmell,C.G. and Suckling,C.J. (2013) Minor groove binders as anti-infective agents. *Pharmacol. Ther.*, **139**, 12–23.
8. Pett,L., Hartley,J.A. and Kiakos,K. (2015) Therapeutic agents based on DNA sequence specific binding. *Curr. Top. Med. Chem.*, **15**, 1293–1322.
9. Wilson,W.D., Tanious,F.A., Mathis,A., Tevis,D., Hall,J.E. and Boykin,D.W. (2008) Antiparasitic compounds that target DNA. *Biochimie*, **90**, 999–1014.
10. Vaijayanthi,T., Bando,T., Pandian,G.N. and Sugiyama,H. (2012) Progress and prospects of pyrrole-imidazole polyamide-fluorophore conjugates as sequence-selective DNA probes. *ChemBioChem*, **13**, 2170–2185.
11. Reddy,B.S.P., Sondhi,S.M. and Lown,J.W. (1999) Synthetic DNA minor groove-binding drugs. *Pharmacol. Ther.*, **84**, 1–111.
12. Pandian,G.N., Sato,S., Anandhakumar,C., Taniguchi,J., Takashima,K., Syed,J., Han,L., Saha,A., Bando,T., Nagase,H. *et al.* (2014) Identification of a small molecule that turns on the pluripotency gene circuitry in human fibroblasts. *ACS Chem. Biol.*, **9**, 2729–2736.
13. Dervan,P.B. and Edelson,B.S. (2003) Recognition of the DNA minor groove by pyrrole-imidazole polyamides. *Curr. Opin. Struct. Biol.*, **13**, 284–299.
14. Hawkins,C.A., Baird,E.E., Dervan,P.B. and Wemmer,D.E. (2002) Analysis of hairpin polyamide complexes having DNA binding sites in close proximity. *J. Am. Chem. Soc.*, **124**, 12689–12696.
15. de Clairac,R.P.L., Geierstanger,B.H., Mrksich,M., Dervan,P.B. and Wemmer,D.E. (1997) NMR characterization of hairpin polyamide complexes with the minor groove of DNA. *J. Am. Chem. Soc.*, **119**, 7909–7916.
16. Zhang,Q., Dwyer,T.J., Tsui,V., Case,D.A., Cho,J., Dervan,P.B. and Wemmer,D.E. (2004) NMR structure of a cyclic polyamide–DNA complex. *J. Am. Chem. Soc.*, **126**, 7958–7966.
17. Chenoweth,D.M. and Dervan,P.B. (2009) Allosteric modulation of DNA by small molecules. *Proc. Natl. Acad. Sci. U.S.A.*, **106**, 13175–13179.
18. Edayathumangalam,R.S., Weyermann,P., Gottesfeld,J.M., Dervan,P.B. and Luger,K. (2004) Molecular recognition of the nucleosomal "supergroove". *Proc. Natl. Acad. Sci. U.S.A.*, **101**, 6864–6869.
19. Hirata,A., Nokihara,K., Kawamoto,Y., Bando,T., Sasaki,A., Ide,S., Maeshima,K., Kasama,T. and Sugiyama,H. (2014) Structural evaluation of tandem hairpin pyrrole-imidazole polyamides recognizing human telomeres. *J. Am. Chem. Soc.*, **136**, 11546–11554.
20. Kawamoto,Y., Bando,T., Kamada,F., Li,Y., Hashiya,K., Maeshima,K. and Sugiyama,H. (2013) Development of a new method for synthesis of tandem hairpin pyrrole-imidazole polyamide probes targeting human telomeres. *J. Am. Chem. Soc.*, **135**, 16468–16477.
21. Erwin,G.S., Grieshop,M.P., Bhimsaria,D., Do,T.J., Rodriguez-Martinez,J.A., Mehta,C., Khanna,K., Swanson,S.A., Stewart,R., Thomson,J.A. *et al.* (2016) Synthetic genome readers target clustered binding sites across diverse chromatin states. *Proc. Natl. Acad. Sci. U.S.A.*, **113**, E7418–E7427.
22. Erwin,G.S., Bhimsaria,D., Eguchi,A. and Ansari,A.Z. (2014) Mapping polyamide-DNA interactions in human cells reveals a new

design strategy for effective targeting of genomic sites. *Angew. Chem. Int. Ed.*, **53**, 10124–10128.
23. Kurmis,A.A., Yang,F., Welch,T.R., Nickols,N.G. and Dervan,P.B. (2017) A pyrrole-imidazole polyamide is active against enzalutamide-resistant prostate cancer. *Cancer Res.*, **77**, 2207–2212.
24. Hargrove,A.E., Martinez,T.F., Hare,A.A., Kurmis,A.A., Phillips,J.W., Sud,S., Pienta,K.J. and Dervan,P.B. (2015) Tumor repression of VCaP xenografts by a pyrrole-imidazole polyamide. *PLoS One*, **10**, e014316.
25. Yang,F., Nickols,N.G., Li,B.C., Marinov,G.K., Said,J.W. and Dervan,P.B. (2013) Antitumor activity of a pyrrole-imidazole polyamide. *Proc. Natl. Acad. Sci. U.S.A.*, **110**, 1863–1868.
26. Hiraoka,K., Inoue,T., Taylor,R.D., Watanabe,T., Koshikawa,N., Yoda,H., Shinohara,K., Takatori,A., Sugimoto,H., Maru,Y. *et al.* (2015) Inhibition of KRAS codon 12 mutants using a novel DNA-alkylating pyrrole-imidazole polyamide conjugate. *Nat. Commun.*, **6**, 6706.
27. Wang,X., Nagase,H., Watanabe,T., Nobusue,H., Suzuki,T., Asami,Y., Shinojima,Y., Kawashima,H., Takagi,K., Mishra,R. *et al.* (2010) Inhibition of MMP-9 transcription and suppression of tumor metastasis by pyrrole-imidazole polyamide. *Cancer Sci.*, **101**, 759–766.
28. Ueno,T., Fukuda,N., Tsunemi,A., Yao,E.H., Matsuda,H., Tahira,K., Matsumoto,T., Matsumoto,K., Matsumoto,Y., Nagase,H. *et al.* (2009) A novel gene silencer, pyrrole-imidazole polyamide targeting human lectin-like oxidized low-density lipoprotein receptor-1 gene improves endothelial cell function. *J. Hypertens.*, **27**, 508–516.
29. Edwards,T.G., Vidmar,T.J., Koeller,K., Bashkin,J.K. and Fisher,C. (2013) DNA damage repair genes controlling human papillomavirus (HPV) episome levels under conditions of stability and extreme instability. *PLoS One*, **8**, e75406.
30. Edwards,T.G., Helmus,M.J., Koeller,K., Bashkin,J.K. and Fisher,C. (2013) Human papillomavirus episome stability is reduced by aphidicolin and controlled by DNA damage response pathways. *J. Virol.*, **87**, 3979–3989.
31. Edwards,T.G., Koeller,K.J., Slomczynska,U., Fok,K., Helmus,M., Bashkin,J.K. and Fisher,C. (2011) HPV episome levels are potently decreased by pyrrole–imidazole polyamides. *Antiviral Res.*, **91**, 177–186.
32. Kawamoto,Y., Sasaki,A., Chandran,A., Hashiya,K., Ide,S., Bando,T., Maeshima,K. and Sugiyama,H. (2016) Targeting 24 bp within telomere repeat sequences with tandem tetramer pyrrole-imidazole polyamide probes. *J. Am. Chem. Soc.*, **138**, 14100–14107.
33. Kawamoto,Y., Sasaki,A., Hashiya,K., Ide,S., Bando,T., Maeshima,K. and Sugiyama,H. (2015) Tandem trimer pyrrole-imidazole polyamide probes targeting 18 base pairs in human telomere sequences. *Chem. Sci.*, **6**, 2307–2312.
34. Geierstanger,B.H., Mrksich,M., Dervan,P.B. and Wemmer,D.E. (1994) Design of a G.C-specific DNA minor groove-binding peptide. *Science*, **266**, 646–650.
35. Trauger,J.W., Baird,E.E. and Dervan,P.B. (1996) Recognition of DNA by designed ligands at subnanomolar concentrations. *Nature*, **382**, 559–561.
36. Kissinger,K., Krowicki,K., Dabrowiak,J.C. and Lown,J.W. (1987) Molecular recognition between oligopeptides and nucleic acids. Monocationic imidazole lexitropsins that display enhanced GC sequence dependent DNA binding. *Biochemistry*, **26**, 5590–5595.
37. Nishijima,S., Shinohara,K., Bando,T., Minoshima,M., Kashiwazaki,G. and Sugiyama,H. (2010) Cell permeability of Py-Im-polyamide-fluorescein conjugates: influence of molecular size and Py/Im content. *Bioorg. Med. Chem.*, **18**, 978–983.
38. Liu,B. and Kodadek,T. (2009) Investigation of the relative cellular permeability of DNA-binding pyrrole-imidazole polyamides. *J. Med. Chem.*, **52**, 4604–4612.
39. Jacobs,C.S. and Dervan,P.B. (2009) Modifications at the C-terminus to improve pyrrole−imidazole polyamide activity in cell culture. *J. Med. Chem.*, **52**, 7380–7388.
40. Edelson,B.S., Best,T.P., Olenyuk,B., Nickols,N.G., Doss,R.M., Foister,S., Heckel,A. and Dervan,P.B. (2004) Influence of structural variation on nuclear localization of DNA-binding polyamide-fluorophore conjugates. *Nucleic Acids Res.*, **32**, 2802–2818.
41. Best,T.P., Edelson,B.S., Nickols,N.G. and Dervan,P.B. (2003) Nuclear localization of pyrrole-imidazole polyamide-fluorescein conjugates in cell culture. *Proc. Natl. Acad. Sci. U.S.A.*, **100**, 12063–12068.
42. Hargrove,A.E., Raskatov,J.A., Meier,J.L., Montgomery,D.C. and Dervan,P.B. (2012) Characterization and solubilization of pyrrole–imidazole polyamide aggregates. *J. Med. Chem.*, **55**, 5425–5432.
43. Yang,F., Nickols,N.G., Li,B.C., Szablowski,J.O., Hamilton,S.R., Meier,J.L., Wang,C.M. and Dervan,P.B. (2013) Animal toxicity of hairpin pyrrole-imidazole polyamides varies with the turn unit. *J. Med. Chem.*, **56**, 7449–7457.
44. Marques,M.A., Doss,R.M., Foister,S. and Dervan,P.B. (2004) Expanding the repertoire of heterocycle ring pairs for programmable minor groove DNA recognition. *J. Am. Chem. Soc.*, **126**, 10339–10349.
45. Renneberg,D. and Dervan,P.B. (2003) Imidazopyridine/pyrrole and hydroxybenzimidazole/pyrrole pairs for DNA minor groove recognition. *J. Am. Chem. Soc.*, **125**, 5707–5716.
46. Marques,M.A., Doss,R.M., Urbach,A.R. and Dervan,P.B. (2002) Toward an understanding of the chemical etiology for DNA minor-groove recognition by polyamides. *Helv. Chim. Acta*, **85**, 4485–4517.
47. Satam,V., Babu,B., Patil,P., Brien,K.A., Olson,K., Savagian,M., Lee,M., Mepham,A., Jobe,L.B., Bingham,J.P. *et al.* (2015) AzaHx, a novel fluorescent, DNA minor groove and G.C recognition element: synthesis and DNA binding properties of a p-anisyl-4-aza-benzimidazole-pyrrole-imidazole (azaHx-PI) polyamide. *Bioorg. Med. Chem. Lett.*, **25**, 3681–3685.
48. Foister,S., Marques,M.A., Doss,R.M. and Dervan,P.B. (2003) Shape selective recognition of T·A base pairs by hairpin polyamides containing N-Terminal 3-Methoxy (and 3-Chloro) thiophene residues. *Bioorg. Med. Chem.*, **11**, 4333–4340.
49. Khalaf,A.I., Bourdin,C., Breen,D., Donoghue,G., Scott,F.J., Suckling,C.J., MacMillan,D., Clements,C., Fox,K. and Sekibo,D.A.T. (2012) Design, synthesis and antibacterial activity of minor groove binders: the role of non-cationic tail groups. *Eur. J. Med. Chem.*, **56**, 39–47.
50. Alniss,H.Y., Anthony,N.G., Khalaf,A.I., Mackay,S.P., Suckling,C.J., Waigh,R.D., Wheate,N.J. and Parkinson,J.A. (2012) Rationalising sequence selection by ligand assemblies in the DNA minor groove: the case for thiazotropsin A. *Chem. Sci.*, **3**, 711–722.
51. Anthony,N.G., Johnston,B.F., Khalaf,A.I., MacKay,S.P., Parkinson,J.A., Suckling,C.J. and Waigh,R.D. (2004) Short lexitropsin that recognizes the DNA minor groove at 5 '-ACTAGT-3 ': understanding the role of isopropyl-thiazole. *J. Am. Chem. Soc.*, **126**, 11338–11349.
52. Sharma,S.K., Tandon,M. and Lown,J.W. (2000) Design and synthesis of novel thiazole-containing cross-linked polyamides related to the antiviral antibiotic distamycin. *J. Org. Chem.*, **65**, 1102–1107.
53. Khalaf,A.I., Waigh,R.D., Drummond,A.J., Pringle,B., McGroarty,I., Skellern,G.G. and Suckling,C.J. (2004) Distamycin analogues with enhanced lipophilicity: synthesis and antimicrobial activity. *J. Med. Chem.*, **47**, 2133–2156.
54. Plouvier,B., Houssin,R., Hecquet,B., Colson,P., Houssier,C., Waring,M.J., Henichart,J.-P. and Bailly,C. (1994) Antitumor combilexin. A thiazole-containing analog of netropsin linked to an acridine chromophore. *Bioconjugate Chem.*, **5**, 475–481.
55. Bailly,C., Colson,P., Houssier,C., Houssin,R., Mrani,D., Gosselin,G., Imbach,J.L., Waring,M.J., Lown,J.W. and Henichart,J.P. (1992) Binding properties and DNA sequence-specific recognition of two bithiazole-linked netropsin hybrid molecules. *Biochemistry*, **31**, 8349–8362.
56. James,P.L., Merkina,E.E., Khalaf,A.I., Suckling,C.J., Waigh,R.D., Brown,T. and Fox,K.R. (2004) DNA sequence recognition by an isopropyl substituted thiazole polyamide. *Nucleic Acids Res.*, **32**, 3410–3417.
57. Alniss,H.Y., Salvia,M.V., Sadikov,M., Golovchenko,I., Anthony,N.G., Khalaf,A.I., MacKay,S.P., Suckling,C.J. and Parkinson,J.A. (2014) Recognition of the DNA minor groove by thiazotropsin analogues. *ChemBioChem*, **15**, 1978–1990.
58. Suckling,C. (2012) From multiply active natural product to candidate drug? Antibacterial (and other) minor groove binders for DNA. *Future Med. Chem.*, **4**, 971–989.

59. Chenoweth,D.M. and Dervan,P.B. (2010) Structural basis for cyclic Py-Im polyamide allosteric inhibition of nuclear receptor binding. *J. Am. Chem. Soc.*, **132**, 14521–14529.
60. Gottesfeld,J.M., Melander,C., Suto,R.K., Raviol,H., Luger,K. and Dervan,P.B. (2001) Sequence-specific recognition of DNA in the nucleosome by pyrrole-imidazole polyamides. *J. Mol. Biol.*, **309**, 615–629.
61. Hawkins,C.A., de Clairac,R.P., Dominey,R.N., Baird,E.E., White,S., Dervan,P.B. and Wemmer,D.E. (2000) Controlling binding orientation in hairpin polyamide DNA complexes. *J. Am. Chem. Soc.*, **122**, 5235–5243.
62. de Clairac,R.P.L., Seel,C.J., Geierstanger,B.H., Mrksich,M., Baird,E.E., Dervan,P.B. and Wemmer,D.E. (1999) NMR characterization of the aliphatic β/β pairing for recognition of A·T/T·A base pairs in the minor groove of DNA. *J. Am. Chem. Soc.*, **121**, 2956–2964.
63. Nickols,N.G. and Dervan,P.B. (2007) Suppression of androgen receptor-mediated gene expression by a sequence-specific DNA-binding polyamide. *Proc. Natl. Acad. Sci. U.S.A.*, **104**, 10418–10423.
64. Cardew,A.S. and Fox,K.R. (2010) In: Fox,KR (ed). *Drug-DNA Interaction Protocols*. Humana Press, Totowa, pp. 153–172.
65. Hampshire,A.J., Rusling,D.A., Broughton-Head,V.J. and Fox,K.R. (2007) Footprinting: a method for determining the sequence selectivity, affinity and kinetics of DNA-binding ligands. *Methods*, **42**, 128–140.
66. Bunkenborg,J., Behrens,C. and Jacobsen,J.P. (2002) NMR characterization of the DNA binding properties of a Novel Hoechst 33258 analogue peptide building block. *Bioconjugate Chem.*, **13**, 927–936.
67. Carlomagno,T., Hennig,M. and Williamson,J.R. (2002) A novel PH-CT-COSY methodology for measuring JPH coupling constants in unlabeled nucleic acids. Application to HIV-2 TAR RNA. *J. Biomol. NMR*, **22**, 65–81.
68. Case,D.A., Cerutti,D.S., Cheatham,I.T.E., Darden,T.A., Duke,R.E., Giese,T.J., Gohlke,H., Goetz,A.W., Greene,D., Homeyer,N. *et al.* (2017) *AMBER 2017*.
69. Pettersen,E.F., Goddard,T.D., Huang,C.C., Couch,G.S., Greenblatt,D.M., Meng,E.C. and Ferrin,T.E. (2004) UCSF chimera - a visualization system for exploratory research and analysis. *J. Comput. Chem.*, **25**, 1605–1612.
70. Lee,W., Tonelli,M. and Markley,J.L. (2015) NMRFAM-SPARKY: enhanced software for biomolecular NMR spectroscopy. *Bioinformatics*, **31**, 1325–1327.
71. Borgias,B.A. and James,T.L. (1990) MARDIGRAS-A procedure for matrix analysis of relaxation for discerning geometry of an aqueous structure. *J. Magn. Reson.*, **87**, 475–487.
72. Blanchet,C., Pasi,M., Zakrzewska,K. and Lavery,R. (2011) CURVES+ web server for analyzing and visualizing the helical, backbone and groove parameters of nucleic acid structures. *Nucleic Acids Res.*, **39**, W68–W73.
73. Lu,X.J. and Olson,W.K. (2003) 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucleic Acids Res.*, **31**, 5108–5121.
74. Baird,E.E. and Dervan,P.B. (1996) Solid phase synthesis of polyamides containing imidazole and pyrrole amino acids. *J. Am. Chem. Soc.*, **118**, 6141–6146.
75. Su,W., Gray,S.J., Dondi,R. and Burley,G.A. (2009) Highly efficient synthesis of DNA-binding hairpin polyamides via the use of a new triphosgene coupling strategy. *Org. Lett.*, **11**, 3910–3913.
76. Fallows,A.J., Singh,I., Dondi,R., Cullis,P.M. and Burley,G.A. (2014) Highly efficient synthesis of DNA-binding polyamides using a convergent fragment-based approach. *Org. Lett.*, **16**, 4654–4657.
77. Fang,L., Pan,Z., Cullis,P.M., Burley,G.A. and Su,W. (2015) Pyrrole-imidazole polyamides: automated solid-phase synthesis. *Curr. Protoc. Nucleic Acid Chem.*, doi:10.1002/0471142700.nc0811s63.
78. Pauff,S.M., Fallows,A.J., Mackay,S.P., Su,W., Cullis,P.M. and Burley,G.A. (2015) Pyrrole-imidazole polyamides: manual solid-phase synthesis. *Curr. Protoc. Nucleic Acid Chem.*, doi:10.1002/0471142700.nc0810s63.
79. Wetzler,M. and Wemmer,D.E. (2010) Facile dimer synthesis for DNA-binding polyamide ligands. *Org. Lett.*, **12**, 3488–3490.
80. Hidaka,T., Pandian,G.N., Taniguchi,J., Nobeyama,T., Hashiya,K., Bando,T. and Sugiyama,H. (2017) Creation of a synthetic ligand for mitochondrial DNA sequence recognition and promoter-specific transcription suppression. *J. Am. Chem. Soc.*, **139**, 8444–8447.
81. Xu,L., Wang,W., Gotte,D., Yang,F., Hare,A.A., Welch,T.R., Li,B.C., Shin,J.H., Chong,J., Strathern,J.N. *et al.* (2016) RNA polymerase II senses obstruction in the DNA minor groove via a conserved sensor motif. *Proc. Natl. Acad. Sci. U.S.A.*, **113**, 12426–12431.
82. Raskatov,J.A., Nickols,N.G., Hargrove,A.E., Marinov,G.K., Wold,B. and Dervan,P.B. (2012) Gene expression changes in a tumor xenograft by a pyrrole-imidazole polyamide. *Proc. Natl. Acad. Sci. U.S.A.*, **109**, 16041–16045.