JACS JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

Article

Cite This: J. Am. Chem. Soc. XXXX, XXX, XXX−XXX

pubs.acs.org/JACS

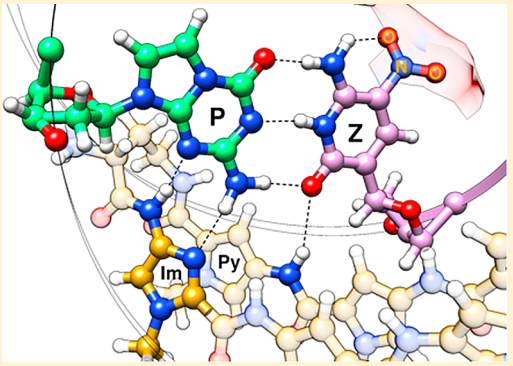
Sequence-Selective Minor Groove Recognition of a DNA Duplex Containing Synthetic Genetic Components

Giacomo Padroni, Jamie M. Withers,[†] Andrea Taladriz-Sender,[†] Linus F. Reichenbach, John A. Parkinson, and Glenn A. Burley*

Department of Pure and Applied Chemistry, University of Strathclyde, Thomas Graham Building, 295 Cathedral Street, Glasgow G1 1XL, United Kingdom

Supporting Information

ABSTRACT: The structural basis of minor groove recognition of a DNA duplex containing synthetic genetic information by hairpin pyrrole-imidazole polyamides is described. Hairpin polyamides induce a higher melting stabilization of a DNA duplex containing the unnatural **P**·**Z** base-pair when an imidazole unit is aligned with a **P** nucleotide. An NMR structural study showed that the incorporation of two isolated **P**·**Z** pairs enlarges the minor groove and slightly narrows the major groove at the site of this synthetic genetic information, relative to a DNA duplex consisting entirely of Watson−Crick base-pairs. Pyrrole-imidazole polyamides bind to a **P**·**Z**-containing DNA duplex to form a stable complex, effectively mimicking a G·C pair. A structural hallmark of minor groove recognition of a **P**·**Z** pair by a polyamide is the reduced level of allosteric distortion induced by binding of a polyamide to a DNA duplex. Understanding the molecular determinants that influence minor groove recognition of DNA containing synthetic genetic components provides the basis to further develop unnatural base-pairs for synthetic biology applications.



INTRODUCTION

Fundamental to all living organisms is the storage and transmission of genetic information in the form of a four-letter nucleic acid code. DNA is the predominant genetic repository used for this purpose, with information typically embedded in an antiparallel B-type duplex where A pairs with T, and G pairs with C. While nature uses two sets of Watson−Crick base-pairs, expanding this code by incorporating synthetic variants offers the opportunity to develop artificial genomes, which have the potential to perform novel functions not possible with natural systems.[1−3]

Rearranging the hydrogen-bond donor (D) and acceptor (A) groups is one design approach to expand the information content in DNA beyond nature's four letter alphabet.[4] The **P**·**Z** pair is the most prominent exemplar of these artificially expanded genetic information systems (AEGIS), where the mode of selective recognition is via a three hydrogen-bond arrangement not present in either a G·C or an A·T pair (Figure 1a).[5−8] This strategy is distinct from other artificial base-pairs (e.g., Romesberg's **dNAM**·**dTPT3**[9] pair and Hirao's **Ds**·**Pa**[10,11] pair) as **P**·**Z** pairing is more closely aligned with Watson−Crick-like hydrogen bonding rather than shape complementarity.[8,12−15]

From a topological perspective, the minor groove hydrogen-bond profile of a **P**·**Z** pair is equivalent to that of G·C (Figure 1a). However, the distinct structural differences of **P**·**Z** compared to G·C pairs could influence sequence-selective recognition. These structural differences include inversion of the central hydrogen bond, the absence of a major groove hydrogen-bond acceptor in the **P** nucleotide (i.e., **P**C7 versus GN7), and the presence of an electron-withdrawing nitro group (**Z**-NO_2) projecting into the major groove.[16,17] Thus, a key challenge in the further development of unnatural base-pairs for synthetic biology applications[6,8,11,12,18−28] is understanding how auxiliary molecular recognition interactions, such as major/minor groove recognition, are influenced by these structural differences.[29,30]

Pyrrole-imidazole polyamides (PIPs) are a class of programmable minor groove binders which bind to double-stranded DNA (dsDNA) in a sequence-selective fashion.[31−38] The well-established pairing rules of PIPs binding to naturally occurring dsDNA[39,40] enable this family of compounds to target sequences up to 24 base-pairs in length—a feature that can subsequently be used to modulate gene-selective transcription both in vitro and in vivo.[34−36]

Recent structural studies have revealed the incorporation of AEGIS results in a transition from conventional B-type to an A-type duplex when the density of **P**·**Z** in a duplex is increased from one to six consecutive pairs.[7,8] This suggests that AEGIS could impact auxiliary molecular recognition processes such as groove recognition, which is essential for transcriptional initiation. At present, the molecular basis for sequence-selective recognition of synthetic genetic components is not

Received: November 20, 2018
Published: May 22, 2019

DOI: 10.1021/jacs.8b12444
J. Am. Chem. Soc. XXXX, XXX, XXX−XXX

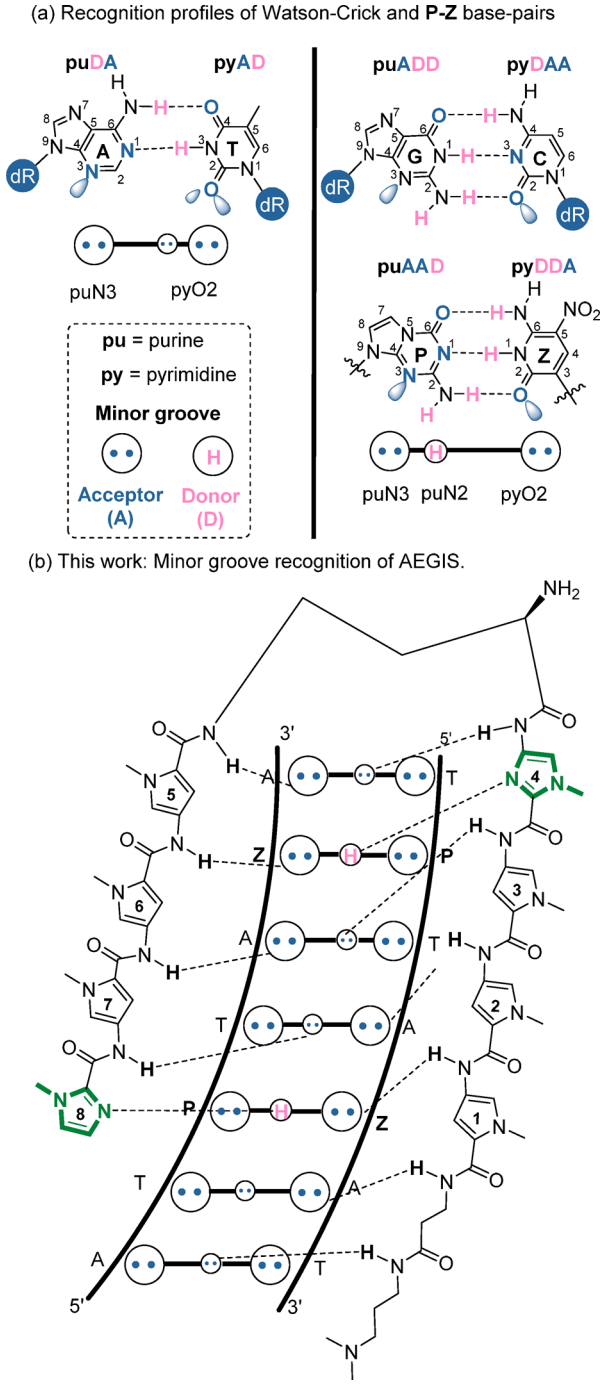


Figure 1. Schematic representation of (a) Watson–Crick and **P·Z** base-pairing, and (b) **PA1** in complex with a target DNA duplex containing two **P·Z** base-pairs. dR = deoxyribose.

known. Herein, we report the first NMR-based structural analysis of a DNA duplex containing AEGIS base-pairs and show how this synthetic genetic information is selectively targeted by a minor groove binding PIP (Figure 1b).[38,41,42]

RESULTS

Experimental Approach. The objectives of this study were to understand how a **P·Z** base-pair incorporated into a DNA duplex influences (i) double-helix structure in solution, and (ii) minor groove recognition by PIPs (**PA1–3**, Figure 2). The well-established recognition rules of PIPs for Watson–Crick pairs[37,38,40,43–47] rendered PIPs excellent candidates to investigate whether the N3 atom of an *N*-methyl imidazole (Im) unit hydrogen bonds with the exocyclic amine (N2) of **P**, much akin to hydrogen bonding observed with the cognate N2 amine of G. Furthermore, 8-ring PIPs such as **PA1–3** predictably bind to their target dsDNA sequences in a "forward" orientation where the *N*-terminus of each PIP (in this study, all three PIPs contain an *N*-terminal Im8 unit) binds in a 5′ → 3′ direction with respect to the DNA backbone.[40,48] All three PIPs in the series (i.e., **PA1–3**) bind to 7 base-pair dsDNA sequences. The PIPs vary in their recognition of the base pair in position **X** (5′-WWG**X**WCW-3′, where W = A/T), which aligns a Py2/Py7 (**PA1**), Py2/Im7 (**PA2**), and a Im2/Py7 (**PA3**) pairing at position **X**.

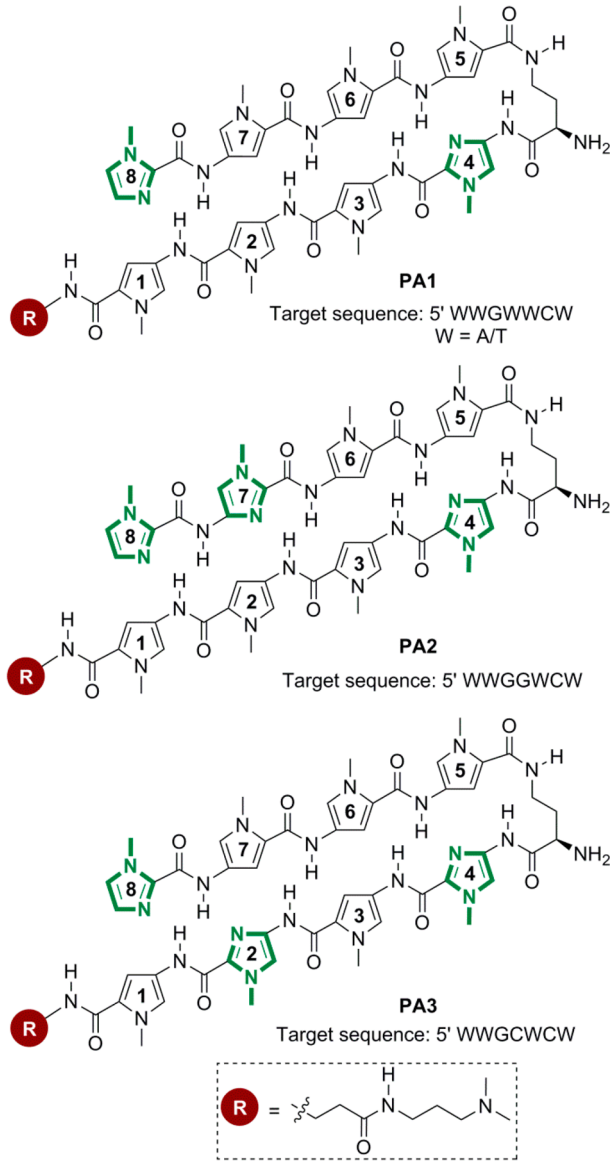


Figure 2. Structures of PIPs (**PA1–3**) used in this study.

PA1 was chosen for NMR-based structural studies as it has a well-demonstrated high-affinity binding profile for the palindromic sequence 5′-WWGWWCW,[44,49–51] whereas **PA2** and **PA3** preferentially bind to 5′-WWGGWCW and 5′-WWGCWCW, respectively.[52–55] Introduction of a **P·Z** pair into position **X·Y** of a target DNA sequence (Figure 3a) would

DOI: 10.1021/jacs.8b12444
J. Am. Chem. Soc. XXXX, XXX, XXX–XXX

allow a greater understanding of how each Py and Im PIP pairing combination influences duplex stabilization.

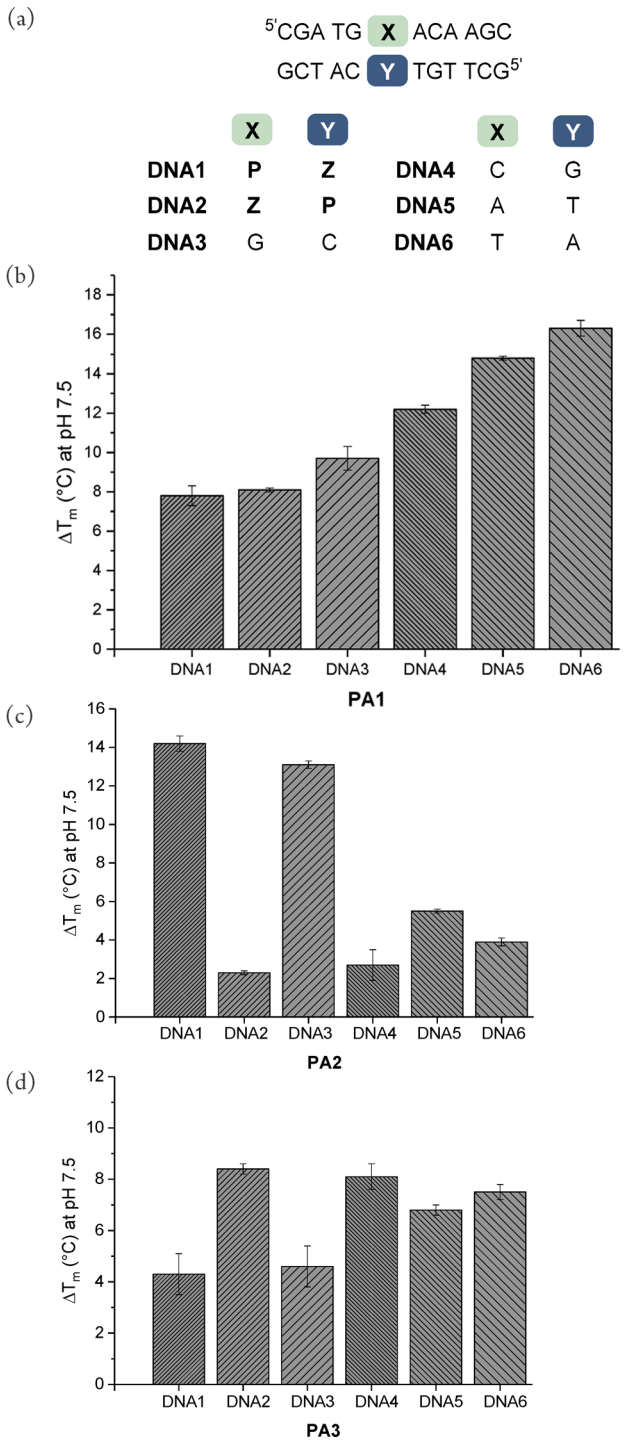


Figure 3. (a) DNA duplexes used in the UV–vis melting study. (b) UV–vis melt stabilization (pH 7.5) of **DNA1–6** (1 μM) using **PA1**, (c) **PA2**, and (d) **PA3** (1.5 μM).

Im Building Blocks Incorporated into Polyamides Preferentially Hydrogen-Bond with P Nucleotides in P·Z-Containing DNA Duplexes. To gain insight into the duplex stabilization and sequence preferences of PIPs (**PA1–3**) in the presence of **P·Z** pairs relative to naturally occurring Watson–Crick base-pairs, UV–vis melting experiments were conducted using duplexes **DNA1–6** (Figure 3a, Figures S1–S6, Table S1).[56–58] These duplexes were chosen in order to determine (i) if hairpin PIP pairings discriminate a **P** nucleotide over a **Z** in a **P·Z** pair, analogous to the preferential pairing of an ImN3 unit with G in a G·C base-pair (versus a C·G), and (ii) the relative differences in duplex stability of PIP binding to dsDNA when a C·G/G·C/A·T/T·A base-pair is in the same position (i.e., position 6). Each of the three pairing combinations binding to DNA at position **X·Y** is surveyed using PIPs where Py and Im are altered in position two and seven [i.e., a Py2/Py7 (**PA1**), Py2/Im7 (**PA2**), and Im2/Py7 (**PA3**)].

The most extensive duplex stabilization induced by **PA1** was observed when position **X·Y** was either an A·T (**PA1·DNA6**, ΔT_m 16.3 °C) or T·A (**PA1·DNA5**, ΔT_m 14.8 °C, Figure 3b). Reduced levels of stabilization were observed for **PA1·DNA1** (ΔT_m 7.8 °C), **PA1·DNA2** (ΔT_m 8.1 °C) and **PA1·DNA3** (i.e., ΔT_m 9.7 °C and **PA1·DNA4**, ΔT_m 12.2 °C). The Py/Py arrangement present in **PA1** exhibits a binding preference for a T·A pair (**PA1·DNA6**) over other sequences, consistent with previous studies.[47,50,52]

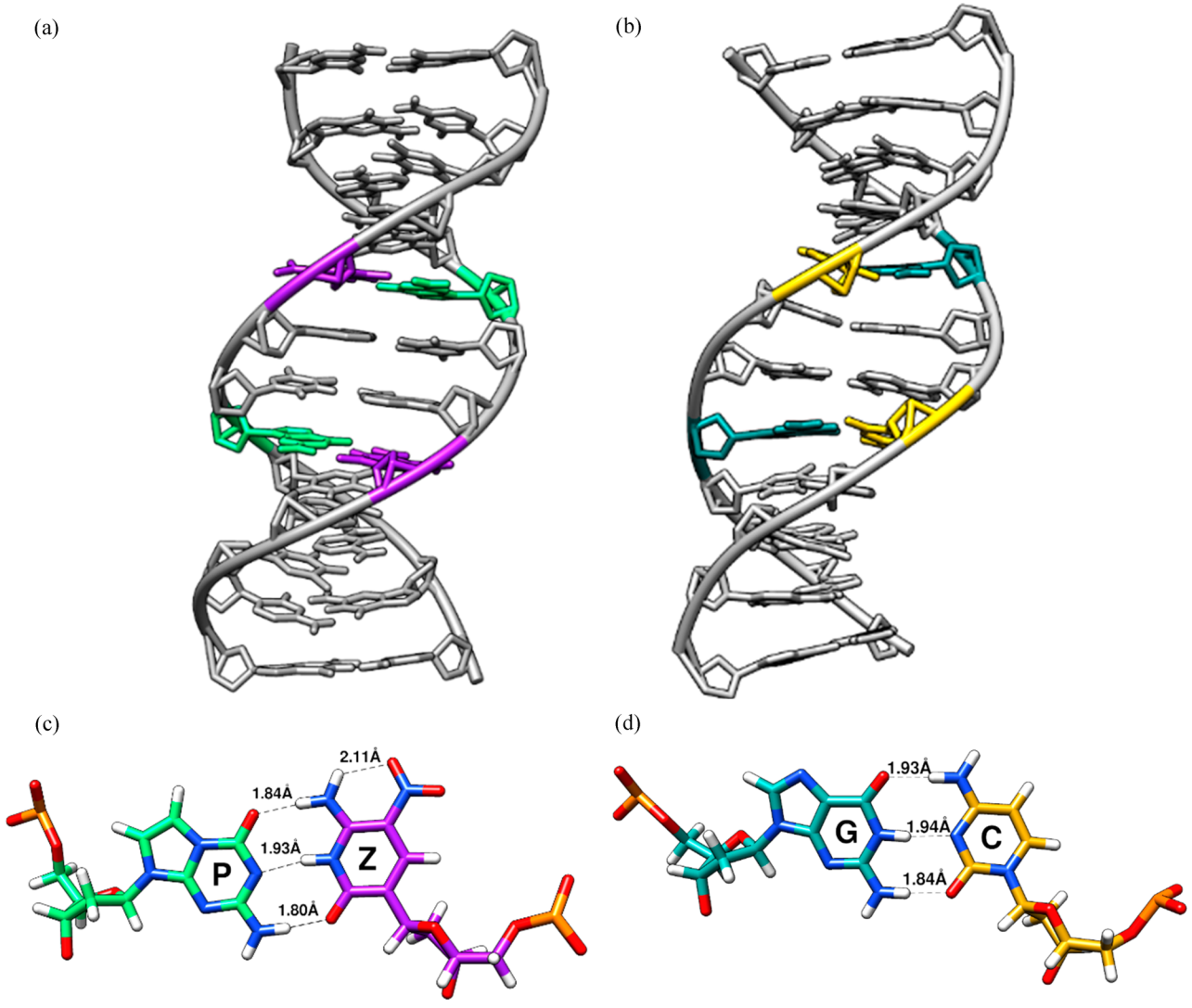
In contrast to the reduced level of duplex stabilization observed when a Py/Py pairing is aligned with a **P·Z**, **PA2** exhibits a larger level of duplex stabilization for **P·Z** in **DNA1** (ΔT_m 14.2 °C) relative to **Z·P** (**DNA2**; ΔT_m 2.3 °C). In this context, this would preferentially align the Im7 unit of **PA2** with a **P** nucleotide in **DNA1**. Duplex stabilization of **PA2·DNA1** was 1.1 °C higher than in the **PA2·DNA3** complex (ΔT_m 13.1 °C), where Im7 is paired with a G (Figure 3c).[52] A reduction in duplex stabilization was observed when this base-pair is inverted to a C·G (**PA2·DNA4**; ΔT_m 2.7 °C) or replaced by A·T/T·A (**PA2·DNA5** ΔT_m 5.5 °C; **PA2·DNA6** ΔT_m 3.9 °C).

Enhanced duplex stabilization was also observed when Im2 in **PA3** is aligned with either a **Z·P** (**PA3·DNA2**, ΔT_m 8.4 °C) or a C·G (**PA3·DNA4**, ΔT_m 8.1 °C) relative to the inverted base-pairing in **PA3·DNA1** (ΔT_m 4.3 °C), **PA3·DNA3** (ΔT_m 4.6 °C) complexes (Figure 3d). The trend of preferential binding of the Im2 unit for **P**/G was also consistent with lower duplex stabilization in the presence of an A·T (**PA3·DNA5**, ΔT_m 6.8 °C) or T·A (**PA3·DNA6** ΔT_m 7.5 °C).

Consistent with a lower binding determined for **PA3** relative to **PA2**,[52] the extent of duplex stabilization is lower. Taken collectively, these studies reveal a preferential recognition mode of a PIP binding to a target dsDNA sequence when an Im unit is aligned with a **P**/G nucleotide relative to **Z**/C. Furthermore, Im units show a higher duplex stabilization with a **P·Z** pair relative to a G·C in the same position.

NMR Structural Characterization of a DNA Duplex Containing Unnatural P·Z Base-Pairs. Solution-based NMR studies were undertaken to explore how the incorporation of two **P·Z** base-pairs in a self-complementary dodecamer d(CGAT**P**TA**Z**ATCG)$_2$ (**DNA7**) impacts duplex structure relative to a duplex incorporating two G·C base pairs in the equivalent position [i.e., d(CGAT**G**TACATCG)$_2$, **DNA8**]. Characteristic NOEs between H6/H8 of the natural nucleobases to the H1′ (Figure S7–S8) and H2′/H2″ sugar protons correlate with the formation of B-type DNA in both duplexes.[8] Further evidence of a B-type duplex being maintained in **DNA7** at the site of the **P·Z** base-pair was the presence of NOEs between **P**H8 and **Z**H4 with sugar H1′ and H2′/H2″ resonances of the flanking bases.[59]

DOI: 10.1021/jacs.8b12444
J. Am. Chem. Soc. XXXX, XXX, XXX–XXX



(e)

$^{5'}$C1 - G2 - A3 - T4 - **X5** - T6 - A7 - **Y8** - A9 - T10 - C11 - G12$^{3'}$

$^{3'}$G24 - C23 - T22 - A21 - **Y20** - A19 - T18 - **X17** - T16 - A15 - G14 - C13$^{5'}$

	X•Y	Shear (Å)	Stretch (Å)	Propeller (°)
DNA7	**P5•Z20**	-0.50 ± 0.12	-0.16 ± 0.04	-3.85 ± 4.13
	P17•Z8	0.47 ± 0.09	-0.16 ± 0.04	-5.28 ± 4.57
DNA8	G5•C20	-0.13 ± 0.07	-0.03 ± 0.02	-17.79 ± 0.87
	G17•C8	0.23 ± 0.10	-0.04 ± 0.03	-20.95 ± 2.67

Figure 4. Average structures of (a) **DNA7** and (b) **DNA8**[49] (PDB 5OCZ) obtained from the ten most representative conformations of the NMR-restrained MD production run. Comparative hydrogen-bonding profile of (c) **P**·**Z** (**DNA7**) and (d) G·C (**DNA8**). Green = **P**; Purple = **Z**; Cyan = G; Gold = C. (e) Comparative analysis of shear, stretch and propeller twist parameters for **DNA7** and **DNA8** (average values ± standard deviation obtained from structure ensembles).

The **Z** nucleotide in **DNA7** exhibited a number of unique spectroscopic features relative to C nucleotides. First, an upfield shift in the H1′ (δ 4.44) and C1′ (δ 72.2) resonances relative to C (H1′ δ 5.51 and C1′ δ 83.2) was attributed to the presence of a *C*-glycosidic linkage. Second, the **Z**-NO_2 group strongly induces a downfield shift (δ 8.47) in the exocyclic N6 resonance (H62) not involved in base pairing, which suggests the presence of an intramolecular hydrogen bond[60] as previously purported in earlier crystal structures.[7,8]

DOI: 10.1021/jacs.8b12444
J. Am. Chem. Soc. XXXX, XXX, XXX−XXX

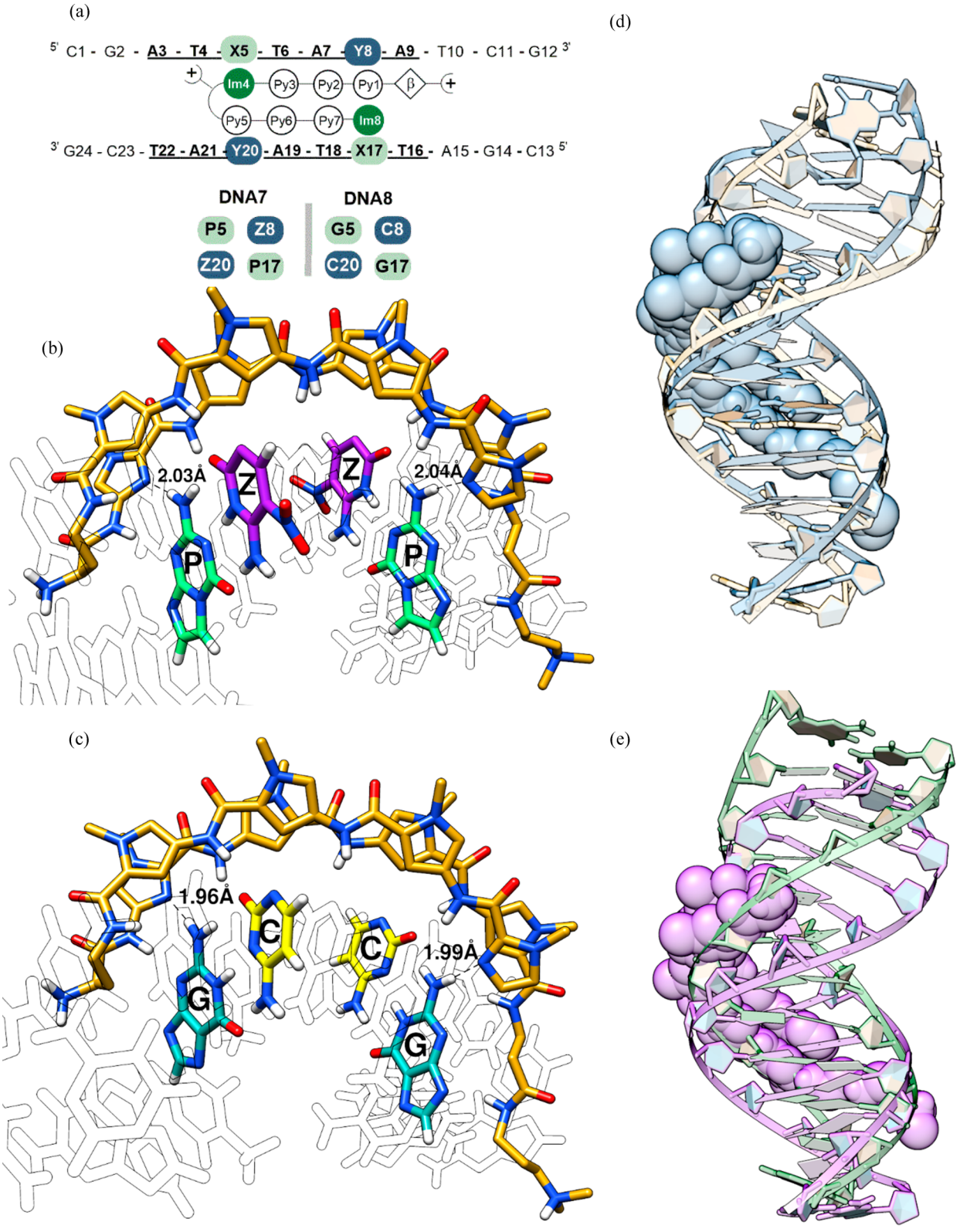


Figure 5. (a) Schematic representation of the **PA1**·**DNA7** complex. Representation of the hydrogen bonding between (b) **PA1** and **P** in the **PA1**·**DNA7** complex (black dashed lines), and (c) **PA1** and G in the **PA1**·**DNA8** complex (black dashed lines). Overlays of (d) **DNA7** (ghost-yellow) and the **PA1**·**DNA7** complex (blue), (e) **DNA8** (ghost-green) and the **PA1**·**DNA8** complex (magenta).

Third, a downfield shift in the exocyclic **Z**N6 proton (H61, δ 9.38) which pairs with **P**O4 is indicative of the formation of a strong hydrogen bond (Figure S9). Further spectroscopic evidence of strong base pairing was the reduced level of solvent exchange of the exocyclic **P**N2 resonances (**P**H21 δ 7.31 and **P**H22 δ 6.10) in 10% D_2O/90% H_2O relative to G·C resonances in the control duplex **DNA8** (Figure S9). Stronger NOE intensities were also observed for all correlations

DOI: 10.1021/jacs.8b12444
J. Am. Chem. Soc. XXXX, XXX, XXX−XXX

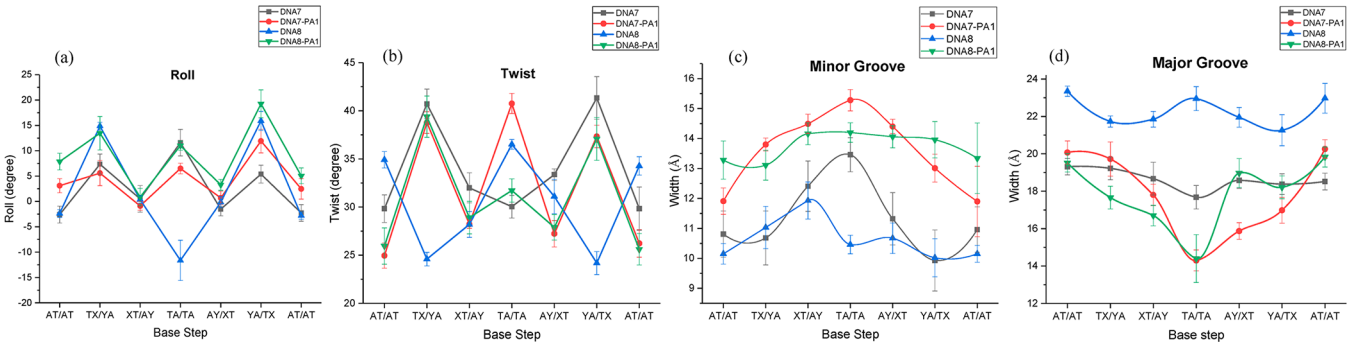


Figure 6. Base step parameters of (a) roll and (b) twist derived from 10 representative conformations obtained from the production runs of **DNA7**, **DNA8**, **PA1·DNA7**, and **PA1·DNA8**. (c) Minor and (d) major groove width (P−P distances) of **DNA7**, **DNA8**, **PA1·DNA7**, and **PA1·DNA8** (X = G/**P**, Y = C/**Z**).

between the nucleobase protons (**P**H8 and **Z**H4) and H2′ compared with H3′ (Figure S14).[59] In addition, correlations between H1′ and H2′/H2′′ were observed in COSY and TOCSY spectra, which is consistent with all the deoxyribose sugars present in **DNA7** and **DNA8** adopting the C2′ *endo* conformation.[61] Finally, the structural impact of incorporating a **P·Z** base-pair into a DNA duplex is evident by comparative analysis of the ^{31}P NMR spectrum of **DNA7** relative to **DNA8**. Perturbations of the sugar-phosphodiester backbone in **DNA7** is evident at the site of the **P·Z** pair (i.e., **P**5/**Z**8) and the flanking base pairs T6·A19 and T4·A21. (Figure S10). While the data confirms that **DNA7** adopts an overall B-type duplex, the **P·Z** pair imparts local distortions to the phosphodiester backbone relative to **DNA8**, which contains a G·C pair in the same position.

Two independent NMR-restrained molecular dynamic (MD) production runs in explicit solvent were obtained after a simulated annealing protocol for **DNA7**. The trajectories were analyzed separately and an ensemble of ten structures was obtained by combining the five most relevant geometrical conformations for each run (Figure S15), whereas the ensemble for **DNA8** was previously reported (PDB 5OCZ).[49] NMR-restrained MD structures of **DNA7** and **DNA8** provided further insight into the structural impact of a **P·Z** pair relative to a G·C in the same position (Figure 4a,b). Subtle but distinct differences in the geometry of the **P·Z** pairing (**DNA7**) relative to equivalent G·C pairs (**DNA8**) are present. In particular, the **P**O6-**Z**N6 hydrogen bond (1.84 Å, Figure 4c) is slightly shorter than the equivalent GO6-CN4 (1.93 Å, Figure 4d). Our MD structures also suggest an average hydrogen-bond distance of 2.11 Å for the exocyclic **Z**N6 and the **Z**-NO_2 group, is consistent with previously observed distances for other 2-nitroanilines.[60]

While our MD calculations show that **DNA7** adopts an overall B-type duplex (Figure 4a,b), there is a change in the shearing, stretching, and propeller twist at the site of the **P·Z** pair in **DNA7** relative to the cognate G·C pair in **DNA8** (Figure 4e and S20). Second, the twist of TP/ZA and ZA/TP steps of **DNA7** is larger (approximately 40°) than the corresponding TG/CA and CA/TG steps of **DNA8** (approximately 25°). As a result of this, the local inclination of **DNA7** is reduced relative to **DNA8** (Figure S21). Lastly, roll and inclination of the central base pair step AT/TA parameters in **DNA7** versus **DNA8** have similar magnitude but opposite direction. These local differences could be due to the stacking of the **Z**-NO_2 group with the adjacent adenine nucleobase (Figure S23). Thus, the overall result of these structural changes is a slightly enlarged minor groove but narrower major groove at the central base step AT/TA for **DNA7** compared to **DNA8**.

In summary, although the incorporation of isolated **P·Z** pairs maintains a B-type structure, these unnatural base-pairs induce structural perturbations, which may arise from a stronger hydrogen-bond network between the **P** and **Z** nucleotides relative to a G·C pair in the same position.[62] The presence of the electron-withdrawing **Z**-NO_2 group is contributing to the stronger **P·Z** pairing by a combination of unique molecular features such as the presence of an intramolecular hydrogen bond with the exocyclic amine of **Z**N6,[60] and influencing base-stacking of the **Z**-NO_2 group with A7/A21. This is also consistent with previous crystal structures of **Z·P**-containing duplexes.[7,8]

Minor Groove Recognition of P·Z Base-Pairs by PA1. NMR structural analyses of the free duplexes **DNA7** and **DNA8** revealed a widening of the minor groove at the site of incorporation of a **P·Z** pair. The unique structural changes induced by these non-natural base-pairs may influence PIP binding. To test this hypothesis, NMR structural analyses of two PIP·dsDNA complexes (i.e., **PA1·DNA7** and **PA1·DNA8**) were undertaken (Figure 5a).

A 1:1 complex for **PA1·DNA7** and **PA1·DNA8** was formed by ^{1}H NMR titration, followed by a full 2D NMR analysis (Figure S11−S12). An NOE connectivity "walk" and shift perturbations of the H1′ and H4′ resonances were consistent with minor groove binding of **PA1** in a hairpin conformation in both complexes (Figure S11).[63] Furthermore, downfield perturbations of **P**H22 resonances compared to the free duplex **DNA7** ($\Delta\delta$ 2.62 and $\Delta\delta$ 2.22) suggest strong hydrogen bonding between Im8N3/Im4N3 of **PA1** and **P**H22 in **DNA7** (Figure S13).

Comparative Structural Analyses of Polyamide·DNA complexes Containing P·Z Base-Pairs (PA1·DNA7) Relative to a Naturally Occurring Duplex (PA1·DNA8). In both **PA1·DNA7** and **PA1·DNA8** complexes, a distinct hydrogen bond with similar length between the N3 atoms of Im4/Im8 with the N2 exocyclic amine of **P**/G was observed (Figure 5b,c). These structural studies confirm that the **P** nucleotide acts as a minor groove "G mimic" via the formation of a hydrogen bond with the endocyclic N3 atom of Im4/Im8 in **PA1**. An unexpected observation when inspecting the overlays of both complexes with their corresponding free duplexes was a reduced level of helical bending in **PA1·DNA7** complex (Figure 5d) relative to **PA1·DNA8** (Figure 5e). We attribute this to the enlarged minor groove width of **DNA7**

DOI: 10.1021/jacs.8b12444
J. Am. Chem. Soc. XXXX, XXX, XXX−XXX

which more closely resembles the bound state (Figure 4a) compared to the natural duplex (**DNA8**, Figure 4b). This is further suggested by comparative analysis of the change in base step parameters upon binding of **PA1** to **DNA7** and **DNA8**. In particular, **PA1** binding to **DNA8** induces extensive changes in the twist of T4G5/C20A21 and C8A9/T16G17 steps of **DNA8** which align with comparable values for T4**P**5/**Z**20A21 and **Z**8A9/T16**P**17 of **DNA7** and **PA1**·**DNA7**. The same trend is observed for roll and inclination of the central T6A7/T18A19 base step (Figure 6a,b). Thus, a more extensive conformational rearrangement occurs for **DNA8** in order to accommodate **PA1** in the **PA1**·**DNA8** complex relative to **PA1**·**DNA7**.

DISCUSSION

This work was designed to gain insight as to how the incorporation of **P**·**Z** within a DNA duplex influences duplex structure and minor groove molecular recognition in solution. We highlight here several key observations which have emerged from our results.

Incorporation of P·Z Base-Pairs in a DNA Duplex Induces Local Structural Perturbations. Previous crystallographic analysis of a DNA duplex containing two isolated **P**·**Z** pairs highlighted a widening of both grooves[8] at the site of the synthetic nucleotides. In addition, increasing the density of **P**·**Z** up to six consecutive base-pairs induces the formation of an A-type duplex where the **Z**-NO_2 group prefers to stack on top of the adjacent nucleobase.[7,16] Thus, the steric and/or electronic properties of the **Z**-NO_2 group plays an influential role in altering dsDNA structure, particularly in duplexes containing consecutive **P**·**Z** pairs.

Our NMR-derived structure of **DNA7** also shows a widening of the minor groove at the site of the **P**·**Z** pair (Figure 4a and 6c). The **P**·**Z** pairing geometry differs quite markedly from a G·C pair in this position. In our structure, a slight narrowing of the major groove at the site of the **P**·**Z** base-pair was observed (Figure 6d). We attribute this to a combination of a shorter hydrogen bond between **P**O6-**Z**N6 (Figure 4c,d) most likely facilitated by the presence of the **Z**-NO_2 group forming an intramolecular hydrogen bond and stacking of the **Z**-NO_2 group with the adjacent adenine base. Concerning the latter point, we speculate that the **Z**-NO_2 group projecting into the major groove could be playing a role in constraining conformational freedom both at the site of a **P**·**Z** pair and on adjacent pairs.[7,8] These localized perturbations are particularly evident on the adjacent sugar atoms (A7C2′/H2′-H2″), which are positioned ∼1 Å further away from **Z**8-NO_2 compared to the C8H5 of **DNA8** (Figure S22). These localized changes are also observed by an upfield shift at A7C2′ ($\Delta\delta$ −0.74) and a downfield shift on the A7H2′-H2″ resonances ($\Delta\delta$ 0.054 and 0.237) relative to the equivalent atoms present in **DNA8** (Table S2−S7).

Allosteric Perturbations Induced by Polyamide Binding to Naturally Occurring DNA8 relative to Z·P-Containing DNA7. A structural hallmark of previous PIP·dsDNA complexes is that PIP binding induces extensive helical bending, widening of the minor groove, and concomitantly, compression of the major groove directly opposite to the site of binding.[43,44,49,63−65] In contrast to this being observed for **PA1**·**DNA8**, this is less evident in the **PA1**·**DNA7** complex where two **P**·**Z** pairs are incorporated within the PIP-binding sequence (Figure 5d−e). In fact, **PA1** binding to **DNA7** results in only minor structural perturbations to the overall duplex. Our thermal UV studies reveal the existence of high stability complexes formed between **P**·**Z**-containing dsDNA and PIPs (Figure 3b,c), which demonstrates that the reduced level of allosteric perturbation observed upon binding of the **PA1** to **DNA7** does not negatively impact minor groove recognition. We surmise that **PA1** binding profile to **DNA7** is more akin to a lock and key model rather than induced fit as previously observed for other minor groove binders such as Hoechst33258 in complex with A-tract dsDNA sequences.[66,67]

Taken collectively, the unique structural features of a **P**·**Z** pair not only influences groove width but also reduces allosteric modulation in a DNA duplex. This is manifested in the reduced level of allosteric perturbation of minor groove recognition by a PIP which, we speculate, is preorganized for more optimal binding. Since sequence-selective recognition of DNA duplexes involves an interplay between direct base contacts and shape complementarity of the duplex,[29,30] our work could be used as a basis to design next-generation molecules which could preferentially bind to synthetic genetic information with enhanced selectivity.

CONCLUSION

In summary, this work demonstrates sequence-selective minor groove recognition of synthetic genetic information incorporated in dsDNA by PIPs. Although the unnatural **P**·**Z** base-pair mimics the hydrogen-bond profile of a G·C projected into the minor groove, the distinct differences of the **P**·**Z** base-pairing geometry plays an influential role in modulating the local helical structure of the free duplex (**DNA7**) and when in complex with a PIP (**PA1**·**DNA7**). We envisage that the unique structural signatures of DNA duplexes containing synthetic genetic information could offer new opportunities in the field of synthetic biology, including the development of new strategies to regulate gene expression[68] or as orthogonal pathways for transcriptional initiation and elongation.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/jacs.8b12444.

PIPs characterization, NMR chemical shifts, and molecular dynamics statistics (PDF)

AUTHOR INFORMATION

Corresponding Author

*glenn.burley@strath.ac.uk

ORCID

John A. Parkinson: 0000-0003-4270-6135

Glenn A. Burley: 0000-0002-4896-113X

Author Contributions

†JMW and ATS contributed equally.

Notes

The authors declare no competing financial interest.

Distance restraints and structure ensembles are deposited in the PDB databank with access codes 6I4O (**DNA7**), 6I4N (**PA1**·**DNA7**), 5OCZ (**DNA8**), and 6RIO (**PA1**·**DNA8**).

ACKNOWLEDGMENTS

This work was supported by the University of Strathclyde studentship scheme [G.P.], the STFC [ST/M000125/1],

DOI: 10.1021/jacs.8b12444
J. Am. Chem. Soc. XXXX, XXX, XXX−XXX

BBSRC [BB/R006857/1], and Leverhulme Trust [RPG-2018-149].

REFERENCES

(1) Benner, S. A.; Karalkar, N. B.; Hoshika, S.; Laos, R.; Shaw, R. W.; Matsuura, M.; Fajardo, D.; Moussatche, P. Alternative Watson-Crick Synthetic Genetic Systems. *Cold Spring Harbor Perspect. Biol.* **2016**, *8*, a023770.

(2) Feldman, A. W.; Romesberg, F. E. Expansion of the Genetic Alphabet: A Chemist's Approach to Synthetic Biology. *Acc. Chem. Res.* **2018**, *51*, 394−403.

(3) Hirao, I.; Kimoto, M.; Yamashige, R. Natural versus Artificial Creation of Base Pairs in DNA: Origin of Nucleobases from the Perspectives of Unnatural Base Pair Studies. *Acc. Chem. Res.* **2012**, *45*, 2055−2065.

(4) Hoshika, S.; Leal, N. A.; Kim, M.-J.; Kim, M.-S.; Karalkar, N. B.; Kim, H.-J.; Bates, A. M.; Watkins, N. E.; SantaLucia, H. A.; Meyer, A. J.; DasGupta, S.; Piccirilli, J. A.; Ellington, A. D.; SantaLucia, J.; Georgiadis, M. M.; Benner, S. A. Hachimoji DNA and RNA: A genetic system with eight building blocks. *Science* **2019**, *363*, 884−887.

(5) Yang, Z.; Chen, F.; Chamberlin, S. G.; Benner, S. A. Expanded Genetic Alphabets in the Polymerase Chain Reaction. *Angew. Chem., Int. Ed.* **2010**, *49*, 177−180.

(6) Yang, Z.; Chen, F.; Alvarado, J. B.; Benner, S. A. Amplification, Mutation, and Sequencing of a Six-Letter Synthetic Genetic System. *J. Am. Chem. Soc.* **2011**, *133*, 15105−15112.

(7) Georgiadis, M. M.; Singh, I.; Kellett, W. F.; Hoshika, S.; Benner, S. A.; Richards, N. G. J. Structural Basis for a Six Nucleotide Genetic Alphabet. *J. Am. Chem. Soc.* **2015**, *137*, 6947−6955.

(8) Reichenbach, L. F.; Sobri, A. A.; Zaccai, N. R.; Agnew, C.; Burton, N.; Eperon, L. P.; de Ornellas, S.; Eperon, I. C.; Brady, R. L.; Burley, G. A. Structural Basis of the Mispairing of an Artificially Expanded Genetic Information System. *Chem.* **2016**, *1*, 946−958.

(9) Feldman, A. W.; Romesberg, F. E. Expansion of the Genetic Alphabet: A Chemist's Approach to Synthetic Biology. *Acc. Chem. Res.* **2018**, *51*, 394−403.

(10) Kimoto, M.; Mitsui, T.; Yamashige, R.; Sato, A.; Yokoyama, S.; Hirao, I. A New Unnatural Base Pair System between Fluorophore and Quencher Base Analogues for Nucleic Acid-Based Imaging Technology. *J. Am. Chem. Soc.* **2010**, *132*, 15418−15426.

(11) Hirao, I.; Kimoto, M.; Mitsui, T.; Fujiwara, T.; Kawai, R.; Sato, A.; Harada, Y.; Yokoyama, S. An unnatural hydrophobic base pair system: site-specific incorporation of nucleotide analogs into DNA and RNA. *Nat. Methods* **2006**, *3*, 729−735.

(12) Wang, X. Y.; Hoshika, S.; Peterson, R. J.; Kim, M. J.; Benner, S. A.; Kahn, J. D. Biophysics of Artificially Expanded Genetic Information Systems. Thermodynamics of DNA Duplexes Containing Matches and Mismatches Involving 2-Amino-3-nitropyridin-6-one (Z) and Imidazo 1,2-a −1,3,5-triazin-4(8H)one (P). *ACS Synth. Biol.* **2017**, *6*, 782−792.

(13) Laos, R.; Shaw, R.; Leal, N. A.; Gaucher, E.; Benner, S. Directed Evolution of Polymerases To Accept Nucleotides with Nonstandard Hydrogen Bond Patterns. *Biochemistry* **2013**, *52*, 5288−5294.

(14) Yang, Z. Y.; Hutter, D.; Sheng, P. P.; Sismour, A. M.; Benner, S. A. Artificially expanded genetic information system: a new base pair with an alternative hydrogen bonding pattern. *Nucleic Acids Res.* **2006**, *34*, 6095−6101.

(15) Li, L. J.; Degardin, M.; Lavergne, T.; Malyshev, D. A.; Dhami, K.; Ordoukhanian, P.; Romesberg, F. E. Natural-like Replication of an Unnatural Base Pair for the Expansion of the Genetic Alphabet and Biotechnology Applications. *J. Am. Chem. Soc.* **2014**, *136*, 826−829.

(16) Molt, J. R. W.; Georgiadis, M. M.; Richards, N. G. J. Consecutive non-natural PZ nucleobase pairs in DNA impact helical structure as seen in 50 μs molecular dynamics simulations. *Nucleic Acids Res.* **2017**, *45*, 3643−3653.

(17) Chawla, M.; Credendino, R.; Chermak, E.; Oliva, R.; Cavallo, L. Theoretical Characterization of the H-Bonding and Stacking Potential of Two Nonstandard Nucleobases Expanding the Genetic Alphabet. *J. Phys. Chem. B* **2016**, *120*, 2216−2224.

(18) Winiger, C. B.; Kim, M. J.; Hoshika, S.; Shaw, R. W.; Moses, J. D.; Matsuura, M. F.; Gerloff, D. L.; Benner, S. A. Polymerase Interactions with Wobble Mismatches in Synthetic Genetic Systems and Their Evolutionary Implications. *Biochemistry* **2016**, *55*, 3847−3850.

(19) Zhang, L. Q.; Yang, Z. Y.; Sefah, K.; Bradley, K. M.; Hoshika, S.; Kim, M. J.; Kim, H. J.; Zhu, G. Z.; Jimenez, E.; Cansiz, S.; Teng, I. T.; Champanhac, C.; McLendon, C.; Liu, C.; Zhang, W.; Gerloff, D. L.; Huang, Z.; Tan, W. H.; Benner, S. A. Evolution of Functional Six-Nucleotide DNA. *J. Am. Chem. Soc.* **2015**, *137*, 6734−6737.

(20) Leal, N. A.; Kim, H. J.; Hoshika, S.; Kim, M. J.; Carrigan, M. A.; Benner, S. A. Transcription, Reverse Transcription, and Analysis of RNA Containing Artificial Genetic Components. *ACS Synth. Biol.* **2015**, *4*, 407−413.

(21) Feldman, A. W.; Romesberg, F. E. In Vivo Structure-Activity Relationships and Optimization of an Unnatural Base Pair for Replication in a Semi-Synthetic Organism. *J. Am. Chem. Soc.* **2017**, *139*, 11427−11433.

(22) Sefah, K.; Yang, Z. Y.; Bradley, K. M.; Hoshika, S.; Jimenez, E.; Zhang, L. Q.; Zhu, G. Z.; Shanker, S.; Yu, F. H.; Turek, D.; Tan, W. H.; Benner, S. A. In vitro selection with artificial expanded genetic information systems. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 1449−1454.

(23) Biondi, E.; Lane, J. D.; Das, D.; Dasgupta, S.; Piccirilli, J. A.; Hoshika, S.; Bradley, K. M.; Krantz, B. A.; Benner, S. A. Laboratory evolution of artificially expanded DNA gives redesignable aptamers that target the toxic form of anthrax protective antigen. *Nucleic Acids Res.* **2016**, *44*, 9565−9577.

(24) Seo, Y. J.; Malyshev, D. A.; Lavergne, T.; Ordoukhanian, P.; Romesberg, F. E. Site-Specific Labeling of DNA and RNA Using an Efficiently Replicated and Transcribed Class of Unnatural Base Pairs. *J. Am. Chem. Soc.* **2011**, *133*, 19878−19888.

(25) Lavergne, T.; Lamichhane, R.; Malyshey, D. A.; Li, Z. T.; Li, L. J.; Sperling, E.; Williamson, J. R.; Millar, D. P.; Romesberg, F. E. FRET Characterization of Complex Conformational Changes in a Large 165 Ribosomal RNA Fragment Site-Specifically Labeled Using Unnatural Base Pairs. *ACS Chem. Biol.* **2016**, *11*, 1347−1353.

(26) Kimoto, M.; Yamashige, R.; Matsunaga, K.; Yokoyama, S.; Hirao, I. Generation of high-affinity DNA aptamers using an expanded genetic alphabet. *Nat. Biotechnol.* **2013**, *31*, 453−457.

(27) Ishizuka, T.; Kimoto, M.; Sato, A.; Hirao, I. Site-specific functionalization of RNA molecules by an unnatural base pair transcription system via click chemistry. *Chem. Commun.* **2012**, *48*, 10835−10837.

(28) Someya, T.; Ando, A.; Kimoto, M.; Hirao, I. Site-specific labeling of RNA by combining genetic alphabet expansion transcription and copper-free click chemistry. *Nucleic Acids Res.* **2015**, *43*, 6665−6676.

(29) Slattery, M.; Zhou, T. Y.; Yang, L.; Machado, A. C. D.; Gordan, R.; Rohs, R. Absence of a simple code: how transcription factors read the genome. *Trends Biochem. Sci.* **2014**, *39*, 381−399.

(30) Rohs, R.; Jin, X.; West, S. M.; Joshi, R.; Honig, B.; Mann, R. S. Origins of Specificity in Protein-DNA Recognition. *Annu. Rev. Biochem.* **2010**, *79*, 233−269.

(31) Trauger, J. W.; Baird, E. E.; Dervan, P. B. Recognition of DNA by designed ligands at subnanomolar concentrations. *Nature* **1996**, *382*, 559−561.

(32) Geierstanger, B. H.; Mrksich, M.; Dervan, P. B.; Wemmer, D. E. Design of a G.C-specific DNA minor groove-binding peptide. *Science* **1994**, *266*, 646−650.

(33) Sharma, S. K.; Reddy, B. S. N.; Lown, J. W. Approaches to develop DNA sequence-specific agents. *Drugs Future* **2001**, *26*, 39−49.

(34) Yang, F.; Nickols, N. G.; Li, B. C.; Marinov, G. K.; Said, J. W.; Dervan, P. B. Antitumor activity of a pyrrole-imidazole polyamide. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 1863−1868.

DOI: 10.1021/jacs.8b12444
J. Am. Chem. Soc. XXXX, XXX, XXX−XXX

(35) Erwin, G. S.; Grieshop, M. P.; Ali, A.; Qi, J.; Lawlor, M.; Kumar, D.; Ahmad, I.; McNally, A.; Teider, N.; Worringer, K.; Sivasankaran, R.; Syed, D. N.; Eguchi, A.; Ashraf, M.; Jeffery, J.; Xu, M.; Park, P. M. C.; Mukhtar, H.; Srivastava, A. K.; Faruq, M.; Bradner, J. E.; Ansari, A. Z. Synthetic transcription elongation factors license transcription across repressive chromatin. *Science* **2017**, *358*, 1617−1622.
(36) Hiraoka, K.; Inoue, T.; Taylor, R. D.; Watanabe, T.; Koshikawa, N.; Yoda, H.; Shinohara, K.; Takatori, A.; Sugimoto, H.; Maru, Y.; Denda, T.; Fujiwara, K.; Balmain, A.; Ozaki, T.; Bando, T.; Sugiyama, H.; Nagase, H. Inhibition of KRAS codon 12 mutants using a novel DNA-alkylating pyrrole-imidazole polyamide conjugate. *Nat. Commun.* **2015**, *6*, 8.
(37) Kawamoto, Y.; Bando, T.; Sugiyama, H. Sequence-specific DNA binding Pyrrole-imidazole polyamides and their applications. *Bioorg. Med. Chem.* **2018**, *26*, 1393−1411.
(38) Withers, J. M.; Padroni, G.; Pauff, S. M.; Clark, A. W.; Mackay, S. P.; Burley, G. A. DNA Minor Groove Binders as Therapeutic Agents. In *Reference Module in Chemistry, Molecular Sciences and Chemical Engineering*; Elsevier, 2017; pp 149−178.
(39) Dervan, P. B.; Edelson, B. S. Recognition of the DNA minor groove by pyrrole-imidazole polyamides. *Curr. Opin. Struct. Biol.* **2003**, *13*, 284−299.
(40) Meier, J. L.; Yu, A. S.; Korf, I.; Segal, D. J.; Dervan, P. B. Guiding the Design of Synthetic DNA-Binding Molecules with Massively Parallel Sequencing. *J. Am. Chem. Soc.* **2012**, *134*, 17814−17822.
(41) Dervan, P. B.; Doss, R. M.; Marques, M. A. Programmable DNA Binding Oligomers for Control of Transcription. *Curr. Med. Chem.: Anti-Cancer Agents* **2005**, *5*, 373−387.
(42) Bando, T.; Sugiyama, H. Synthesis and Biological Properties of Sequence-Specific DNA-Alkylating Pyrrole−Imidazole Polyamides. *Acc. Chem. Res.* **2006**, *39*, 935−944.
(43) Chenoweth, D. M.; Dervan, P. B. Structural Basis for Cyclic Py-Im Polyamide Allosteric Inhibition of Nuclear Receptor Binding. *J. Am. Chem. Soc.* **2010**, *132*, 14521−14529.
(44) Chenoweth, D. M.; Dervan, P. B. Allosteric modulation of DNA by small molecules. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 13175−13179.
(45) Dose, C.; Farkas, M. E.; Chenoweth, D. M.; Dervan, P. B. Next Generation Hairpin Polyamides with (R)-3,4-Diaminobutyric Acid Turn Unit. *J. Am. Chem. Soc.* **2008**, *130*, 6859−6866.
(46) Puckett, J. W.; Muzikar, K. A.; Tietjen, J.; Warren, C. L.; Ansari, A. Z.; Dervan, P. B. Quantitative Microarray Profiling of DNA-Binding Molecules. *J. Am. Chem. Soc.* **2007**, *129*, 12310−12319.
(47) Warren, C. L.; Kratochvil, N. C. S.; Hauschild, K. E.; Foister, S.; Brezinski, M. L.; Dervan, P. B.; Phillips, G. N.; Ansari, A. Z. Defining the sequence-recognition profile of DNA-binding molecules. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 867−872.
(48) White, S.; Baird, E. E.; Dervan, P. B. Orientation preferences of pyrrole-imidazole polyamides in the minor groove of DNA. *J. Am. Chem. Soc.* **1997**, *119*, 8756−8765.
(49) Padroni, G.; Parkinson, J. A.; Fox, K. R.; Burley, G. A. Structural basis of DNA duplex distortion induced by thiazole-containing hairpin polyamides. *Nucleic Acids Res.* **2018**, *46*, 42−53.
(50) Nickols, N. G.; Dervan, P. B. Suppression of androgen receptor-mediated gene expression by a sequence-specific DNA-binding polyamide. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 10418−10423.
(51) Chenoweth, D. M.; Harki, D. A.; Phillips, J. W.; Dose, C.; Dervan, P. B. Cyclic Pyrrole-Imidazole Polyamides Targeted to the Androgen Response Element. *J. Am. Chem. Soc.* **2009**, *131*, 7182−7188.
(52) Hsu, C. F.; Phillips, J. W.; Trauger, J. W.; Farkas, M. E.; Belitsky, J. M.; Heckel, A.; Olenyuk, B. Z.; Puckett, J. W.; Wang, C. C. C.; Dervan, P. B. Completion of a programmable DNA-binding small molecule library. *Tetrahedron* **2007**, *63*, 6146−6151.
(53) Krpetic, Z.; Singh, I.; Su, W.; Guerrini, L.; Faulds, K.; Burley, G. A.; Graham, D. Directed Assembly of DNA-Functionalized Gold Nanoparticles Using Pyrrole-Imidazole Polyamides. *J. Am. Chem. Soc.* **2012**, *134*, 8356−8359.
(54) Su, W.; Schuster, M.; Bagshaw, C. R.; Rant, U.; Burley, G. A. Site-Specific Assembly of DNA-Based Photonic Wires by Using Programmable Polyamides. *Angew. Chem., Int. Ed.* **2011**, *50*, 2712−2715.
(55) Su, W.; Bagshaw, C. R.; Burley, G. A. Addressable and unidirectional energy transfer along a DNA three-way junction programmed by pyrrole-imidazole polyamides. *Sci. Rep.* **2013**, *3*, 1883.
(56) Wang, S.; Nanjunda, R.; Aston, K.; Bashkin, J. K.; Wilson, W. D. Correlation of Local Effects of DNA Sequence and Position of *β*-Alanine Inserts with Polyamide−DNA Complex Binding Affinities and Kinetics. *Biochemistry* **2012**, *51*, 9796−9806.
(57) Pilch, D. S.; Poklar, N.; Baird, E. E.; Dervan, P. B.; Breslauer, K. J. The Thermodynamics of Polyamide−DNA Recognition: Hairpin Polyamide Binding in the Minor Groove of Duplex DNA. *Biochemistry* **1999**, *38*, 2143−2151.
(58) James, P. L.; Le Strat, L.; Ellervik, U.; Bratwall, C.; Nordén, B.; Brown, T.; Fox, K. R. Effects of a hairpin polyamide on DNA melting: comparison with distamycin and Hoechst 33258. *Biophys. Chem.* **2004**, *111*, 205−212.
(59) Pandav, K.; Pandya, P.; Barthwal, R.; Kumar, S. Structure Determination of DNA Duplexes by NMR. In *Chemistry of Phytopotentials: Health, Energy and Environmental Perspectives*; Khemani, L. D., Srivastava, M. M., Srivastava, S., Eds.; Springer: Berlin, Heidelberg, 2012; pp 155−158.
(60) Panunto, T. W.; Urbanczyk-Lipkowska, Z.; Johnson, R.; Etter, M. C. Hydrogen-bond formation in nitroanilines: the first step in designing acentric materials. *J. Am. Chem. Soc.* **1987**, *109*, 7786−7797.
(61) Wijmenga, S. S.; van Buuren, B. N. M. The use of NMR methods for conformational studies of nucleic acids. *Prog. Nucl. Magn. Reson. Spectrosc.* **1998**, *32*, 287−387.
(62) Wang, X. Y.; Hoshika, S.; Peterson, R. J.; Kim, M. J.; Benner, S. A.; Kahn, J. D. Biophysics of Artificially Expanded Genetic Information Systems. Thermodynamics of DNA Duplexes Containing Matches and Mismatches Involving 2-Amino-3-nitropyridin-6-one (Z) and Imidazo 1,2-a −1,3,5-triazin-4(8H)one (P). *ACS Synth. Biol.* **2017**, *6*, 782−792.
(63) de Clairac, R. P. L.; Geierstanger, B. H.; Mrksich, M.; Dervan, P. B.; Wemmer, D. E. NMR Characterization of Hairpin Polyamide Complexes with the Minor Groove of DNA. *J. Am. Chem. Soc.* **1997**, *119*, 7909−7916.
(64) Zhang, Q.; Dwyer, T. J.; Tsui, V.; Case, D. A.; Cho, J.; Dervan, P. B.; Wemmer, D. E. NMR Structure of a Cyclic Polyamide−DNA Complex. *J. Am. Chem. Soc.* **2004**, *126*, 7958−7966.
(65) de Clairac, R. P. L.; Seel, C. J.; Geierstanger, B. H.; Mrksich, M.; Baird, E. E.; Dervan, P. B.; Wemmer, D. E. 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.* **1999**, *121*, 2956−2964.
(66) Ramakers, L. A. I.; Hithell, G.; May, J. J.; Greetham, G. M.; Donaldson, P. M.; Towrie, M.; Parker, A. W.; Burley, G. A.; Hunt, N. T. 2D-IR Spectroscopy Shows that Optimized DNA Minor Groove Binding of Hoechst33258 Follows an Induced Fit Model. *J. Phys. Chem. B* **2017**, *121*, 1295−1303.
(67) Bostock-Smith, C. E.; Harris, S. A.; Laughton, C. A.; Searle, M. S. Induced fit DNA recognition by a minor groove binding analogue of Hoechst 33258: fluctuations in DNA A tract structure investigated by NMR and molecular dynamics simulations. *Nucleic Acids Res.* **2001**, *29*, 693−702.
(68) Hoshika, S.; Singh, I.; Switzer, C.; Molt, R. W.; Leal, N. A.; Kim, M.-J.; Kim, M.-S.; Kim, H.-J.; Georgiadis, M. M.; Benner, S. A. Skinny" and "Fat" DNA: Two New Double Helices. *J. Am. Chem. Soc.* **2018**, *140*, 11655−11660.

DOI: 10.1021/jacs.8b12444
J. Am. Chem. Soc. XXXX, XXX, XXX−XXX