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

J. Am. Chem. Soc., Just Accepted Manuscript • Publication Date (Web): 22 May 2019

Downloaded from http://pubs.acs.org on May 24, 2019

Just Accepted

"Just Accepted" manuscripts have been peer-reviewed and accepted for publication. They are posted online prior to technical editing, formatting for publication and author proofing. The American Chemical Society provides "Just Accepted" as a service to the research community to expedite the dissemination of scientific material as soon as possible after acceptance. "Just Accepted" manuscripts appear in full in PDF format accompanied by an HTML abstract. "Just Accepted" manuscripts have been fully peer reviewed, but should not be considered the official version of record. They are citable by the Digital Object Identifier (DOI®). "Just Accepted" is an optional service offered to authors. Therefore, the "Just Accepted" Web site may not include all articles that will be published in the journal. After a manuscript is technically edited and formatted, it will be removed from the "Just Accepted" Web site and published as an ASAP article. Note that technical editing may introduce minor changes to the manuscript text and/or graphics which could affect content, and all legal disclaimers and ethical guidelines that apply to the journal pertain. ACS cannot be held responsible for errors or consequences arising from the use of information contained in these "Just Accepted" manuscripts.
**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, Glenn A. Burley*

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

**DNA • minor groove recognition • unnatural base-pairs • synthetic biology • NMR**

**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 \( PZ \) base-pair when an imidazole unit is aligned with a \( P \) nucleotide. An NMR structural study showed that the incorporation of two isolated \( PZ \) 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 \( PZ \)-containing DNA duplex to form a stable complex, effectively mimicking a \( GC \) pair. A structural hallmark of minor groove recognition of a \( PZ \) 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 which 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 anti-parallel B-type duplex where A pairs with T, and G pairs with C. Whilst 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.

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. The \( PZ \) pair is the most prominent exemplar of these artificially-expanded genetic information systems (AEGIS), where the mode of selective recognition is via a three-base-pair motif not present in either a \( GC \) or \( AT \) pair (Figure 1a). This strategy is distinct from other artificial base-pairs (e.g., Romesberg’s \( dNAM\cdot dTPT3 \) pair and Hirao’s \( Ds\cdot Pa \) pair) as \( PZ \) pairing is more closely aligned with Watson-Crick-like hydrogen-bonding rather than shape complementarity.

From a topological perspective, the minor groove hydrogen-bond profile of a \( PZ \) pair is equivalent to that of \( GC \) (Figure 1a). However, the distinct structural differences of \( PZ \) compared to \( GC \) 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., PC7 versus GN7), and the presence of an electron-withdrawing nitro group (\( Z\)-NO\(_2\)) projecting into the major groove. Thus, a key challenge in the further development of unnatural base-pairs for synthetic biology applications is understanding how auxiliary molecular recognition interactions, such as major/minor groove recognition, are influenced by these structural differences.

Pyrrole-Imidazole polyamides (PIPs) are a class of programmable minor groove binders which bind to double-stranded DNA (dsDNA) in a sequence-selective fashion. The well-established pairing rules of PIPs binding to naturally-occurring dsDNA 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-expression both in vitro and in vivo.

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. 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 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).

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 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. 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’-WWGXWCW-3’, where W = A/T), which aligns a Py2/Py7 (PA1), Py2/Im7 (PA2) and a Im2/Py7 (PA3) pairing at position X.

PA1 was chosen for NMR-based structural studies as it has a well-demonstrated high-affinity binding profile for the palindromic sequence 5’-WWGWGCW-3’, whereas PA2 and PA3 preferentially bind to 5’-WWGWCW and 5’-WWGCWCW, respectively. Introduction of a P•Z pair into position X•Y of a target DNA sequence (Figure 3a) would allow a greater understanding of how each Py and Im PIP pairing combinations influences duplex stabilization.

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 S1-6, Table S1). 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 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, ΔTm 16.3°C) or T•A (PA1•DNA5, ΔTm 14.8°C, Figure 3b). Reduced levels of stabilization were observed for PA1•DNA1 (ΔTm 7.8°C), PA1•DNA2 (ΔTm 8.1°C) and PA1•DNA3 (ΔTm 9.7°C and PA1•DNA4, ΔTm 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.

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 (ΔTm 14.2 °C) relative to Z•P (DNA2, ΔTm 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 (ΔTm 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; ΔTm 2.7 °C) or replaced by A•T/T•A (PA2•DNA5 ΔTm 5.5 °C; PA2•DNA6 ΔTm 3.9 °C).

Enhanced duplex stabilization was also observed when Im2 in PA3 is aligned with either a Z•P (PA3•DNA2, ΔTm 8.4 °C) or a C•G (PA3•DNA4, ΔTm 8.1 °C) relative to the inverted base-pairing in PA3•DNA1 (ΔTm 4.3 °C), PA3•DNA3 (ΔTm 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, ΔTm 6.8 °C) or T•A (PA3•DNA6 ΔTm 7.5 °C).

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

(a)
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(CGATPTIAZATCG)₂ (DNA7) impacts duplex structure relative to a duplex incorporating two G•C base pairs in the equivalent position [i.e., d(CGATPTIAZATCG)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. Further evidence of a B-type duplex being maintained in DNA7 at the site of the P•Z base-pair was the presence of NOE between PH8 and ZH4 with sugar H1' and H2'/H2'' resonances of the flanking bases.

The Z nucleotide in DNA7 exhibited a number of unique spectroscopic features relative to C nucleotides. Firstly, an upfield shift in the H1' (6.44) and C1' (7.22) resonances relative to C (H1' δ 5.51 and C1' δ 8.32) was attributed to the presence of a C-glycosidic linkage. Secondly, the Z•NO₂ 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 as previously purported in earlier crystal structures.

Thirdly, a downfield shift in the exocyclic ZN6 proton (H61, δ 9.38) which pairs with PO4 is indicative of the formation of a stronger hydrogen-bond (Figure S9). Further spectroscopic evidence of strong base pairing was the reduced level of solvent exchange of the exocyclic PN2 resonances (PH21 δ 7.31 and PH22 δ 6.10) in 10% D₂O/90% H₂O relative to G•C resonances in the control duplex DNA8 (Figure S9). Stronger NOE intensities were also observed for all correlations between the nucleobase protons (PH8 and ZH4) and H2' compared with H3' (Figure S14). 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. Finally, the structural impact of incorporating a P•Z base-pair into a DNA duplex is evident by comparative analysis of the 3¹P 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., P5/Z6) and the flanking base pairs T6•A19 and T4•A21. (Figure S10). Whilst 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 5QCZ). 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 P60-ZN6 hydrogen-bond (1.84 Å, Figure 4c) is slightly shorter than the equivalent G06-CN4 (1.93 Å, Figure 4d). Our MD structures also suggest an average hydrogen-bond distance of 2.11 Å for the exocyclic ZN6 and the Z-NO₂ group, is consistent with previously observed distances for other 2-nitroanilines.

Whilst 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 & S20). Secondly, the twist of TP/ZA and ZA/TP steps of DNA7 is larger (approx. 40°) than the corresponding TG/CA and CA/TG steps of DNA8 (approx. 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-NO₂ 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. The presence of the electron-withdrawing Z-NO₂ 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 ZN6 and influencing base-stacking of the Z-NO₂ group with A7/A21. This is also consistent with previous crystal structures of Z-P-containing duplexes and DNA8.

**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 ¹H 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). Furthermore, downfield perturbations of PH22 resonances compared to the free duplex DNA7 (δ 2.62 and δ 2.22) suggest strong hydrogen-bonding between Im8N3/Im4N3 of PA1 and PH22 in DNA7 (Figure S13).
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\(\cdot\)Z (DNA7) and (d) G\(\cdot\)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).

|        | Shear (Å) | Stretch (Å) | Propeller (°) |
|--------|-----------|-------------|---------------|
| DNA7   |           |             |               |
| P5\(\cdot\)Z20 | -0.50 ± 0.12 | -0.16 ± 0.04 | -3.85 ± 4.13  |
| P17\(\cdot\)Z8  | 0.47 ± 0.09  | -0.16 ± 0.04 | -5.28 ± 4.57  |
| DNA8   |           |             |               |
| G5\(\cdot\)C20 | -0.13 ± 0.07 | -0.03 ± 0.02 | -17.79 ± 0.87 |
| G17\(\cdot\)C8  | 0.23 ± 0.10  | -0.04 ± 0.03 | -20.95 ± 2.67 |
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)
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 overlaps of both complexes with their corresponding free duplexes was a reduced level of helical bending the in PA1•DNA7 complex (Figure 5d) relative to PA1•DNA8 (Figure 5e). We attribute this to the enlarged minor groove width of DNA7 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 T4P5/Z20A21 and Z8A9/T16P17 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 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-NO2 group prefers to stack on top of the adjacent nucleobase. Thus, the steric and/or electronic properties of the Z-NO2 group play 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 & 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 PO6-ZN6 (Figure 4c-d) most likely facilitated by the presence of the Z-NO2 group forming an intramolecular hydrogen-bond and stacking of the Z-NO2 group with the adjacent adenine base. Concerning the latter point, we speculate that the Z-NO2 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. These localized perturbations are particularly evident on the adjacent sugar atoms (A7C2'/H2'-H2"'), which are positioned ~1 Å further away from Z8-NO2 compared to the C8H5 of DNA8 (Figure S2). These localized changes are also observed by an upfield shift at A7C2' (Δδ ~0.74) and a downfield shift on the A7H2'-H2" resonances (Δδ 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 P binding induces extensive helical bending, widening of the minor groove, and concomitantly, compression of the major groove directly opposite to the site of binding. 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 5b). 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 Atract dsDNA sequences.

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 pre-organized for more optimal binding. Since sequence-selective recognition of DNA duplexes involves an interplay between direct base contacts and shape complementarity of the duplex, 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 or as orthogonal pathways for transcriptional initiation and elongation.
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).

ASSOCIATED CONTENT
Distance restraints and structures ensembles are deposited in the PDB databank with access code 6I4O (DNA7), 6I4N (PA1•DNA7), 5OCZ (DNA8) and 6RIO (PA1•DNA8). PIPs characterization, NMR chemical shifts and molecular dynamics statistics can be found in the supporting information. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION
Corresponding Author
* glenn.burley@strath.ac.uk

Author Contributions
The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

§ These authors contributed equally.

Funding Sources
This work was supported by the University of Strathclyde studentship scheme [G.P.], the STFC [ST/M000125/1], BBSRC [BB/R006857/1] and Leverhulme Trust [RPG-2018-149].

ABBREVIATIONS
PIP, pyrrole-imidazole polyamide.

REFERENCES
1. Benner, S. A.; Karalkar, N. B.; Hoshika, S.; Laos, R.; Shaw, R. W.; Matsura, 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. Chemie 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.; Bradley, 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. Watanabe, M.; Mitsui, T.; Yamashige, R.; Sato, A.; Yokoyama, S.; Hirao, L., 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. Hiroa, I.; Kimoto, M.; Mitsuji, 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.; Hoshioka, 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-nitrophenyl-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. F.; Sismon, 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.; Malsyhev, D. A.; Dhami, K.; Orduokhuanian, 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.; Cavallero, 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.; Hoshioka, 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.; Sehaf, K.; Bradley, K. M.; Hoshioka, S.; Kim, M. J.; Kim, H. J.; Zhu, G. Z.; Jimenez, E.; Cansiz, S.; Teng, I. T.; Champanac, 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*, 6734-6739.

20. Leal, N. A.; Alam, H. H.; Hoshioka, 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. Sehaf, K.; Yang, Z. Y.; Bradley, K. M.; Hoshioka, S.; Jimenez, E.; Zhang, L. Q.; Zhu, G. Z.; Shanker, S.; Yu, H. F.; Turek, D.; Tan, W. H.; Benner, S. A., Invitro 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.; Hoshioka, 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.; Orduokhuanian, 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.; Malsyhev, 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. 2011, 6*, 1347-1353.
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. Chernoweth, 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.; Faulks, 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. Chemie 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. Pandey, 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: Berlin, Heidelberg, 2012; pp 155-158.

60. Panunto, T. W.; Urbanczyk-Lipowska, 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. I.; Tsai, 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.; Towne, 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.
TABLE OF CONTENTS FIGURE