





ARTICLE

DOI: 10.1038/s42004-018-0080-5

OPEN

# Shape selective bifacial recognition of double helical DNA

Shivaji A. Thadke <sup>1</sup>, V.M. Hridya<sup>2</sup>, J. Dinithi R. Perera<sup>1</sup>, Roberto R. Gil <sup>1</sup>, Arnab Mukherjee <sup>2</sup> & Danith H. Ly <sup>1</sup>

An impressive array of antigene approaches has been developed for recognition of double helical DNA over the past three decades; however, few have exploited the ‘Watson-Crick’ base-pairing rules for establishing sequence-specific recognition. One approach employs peptide nucleic acid as a molecular reagent and strand invasion as a binding mode. However, even with integration of the latest conformationally-preorganized backbone design, such an approach is generally confined to sub-physiological conditions due to the lack of binding energy. Here we report the use of a class of shape-selective, bifacial nucleic acid recognition elements, namely Janus bases, for targeting double helical DNA or RNA. Binding occurs in a highly sequence-specific manner under physiologically relevant conditions. The work may provide a foundation for the design of oligonucleotides for targeting the secondary and tertiary structures of nucleic acid biopolymers.

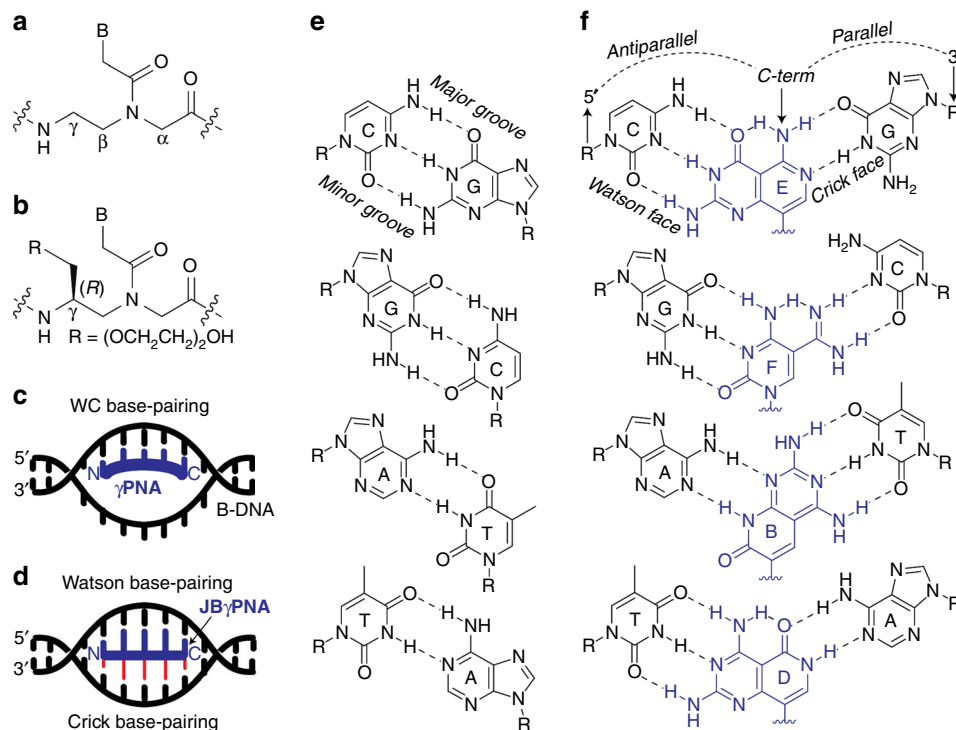
<sup>1</sup>Department of Chemistry and Institute for Biomolecular Design and Discovery (IBD), Carnegie Mellon University, 4400 Fifth Avenue, Pittsburgh, PA 15213, USA. <sup>2</sup>Department of Chemistry, Indian Institute of Science Education and Research (IISER), Pune, Maharashtra 411008, India. Correspondence and requests for materials should be addressed to S.A.T. (email: [sathadke@andrew.cmu.edu](mailto:sathadke@andrew.cmu.edu)) or to D.H.L. (email: [dly@andrew.cmu.edu](mailto:dly@andrew.cmu.edu))

It has been over six decades since the structure of double helical DNA was first unveiled<sup>1</sup>. To chemists and biologists working in the field of molecular recognition at the time, the structure provided more than just insight into the mechanism for the storage and transmission of genetic information. It provided a paradigm for recognition of genetic materials, in a way in which the two opposing strands are held together: through hydrogen-bonding interaction between the adenine (A) nucleobase and thymine (T), and between cytosine (C) and guanine (G). These so-called ‘Watson-Crick’ base-pairing rules are routinely employed in the design of oligonucleotide molecules for targeting single-stranded or perturbed regions of double-stranded DNA or RNA; however, they are rarely applied in the recognition of double helical B-form DNA (B-DNA)—the most stable form of DNA double helix. The reason is that most oligonucleotide molecules developed to date do not have sufficient binding energy to invade B-DNA. To circumvent this challenge, past research efforts have generally been focused on the minor and major grooves, in large part, because of the ease of accessibility of the chemical groups and the precedence for their recognition in Nature<sup>2,3</sup>. This pursuit has led to the development of several major classes of antigene molecules including polyamides<sup>4</sup>, triplex-forming oligonucleotides<sup>5</sup>, zinc-finger peptides<sup>6,7</sup>, and transcription activator-like effectors (TALEs)<sup>8,9</sup>. Notwithstanding considerable advances in these areas, the issues of sequence selection and/or specificity, to a certain extent, still remain with many of these classes of molecules. Recently, a CRISPR-Cas9 system has been harnessed from bacteria for the manipulation of genomes and has achieved remarkable success<sup>10,11</sup>. Even with this state-of-the-art technology, though, an improvement in sequence-specificity is warranted for many of the anticipated biomedical applications due to the concern for off-target genome modifications<sup>12,13</sup>.

In concert with these efforts, Nielsen and coworkers<sup>14,15</sup> have shown over the past two decades that peptide nucleic acid (PNA), a nucleic acid mimic comprising a pseudopeptide backbone (Fig. 1a), can invade B-DNA. Strand invasion of DNA by PNA

occurs predominantly through Watson-Crick<sup>16</sup>, or in combination with Hoogsteen base-pairing<sup>17</sup>, depending on the sequence context. The simplicity and the generality in sequence design, along with its exquisite recognition-specificity, make PNA an attractive antigene agent. However, with the original (achiral) backbone design, PNA recognition is restricted to mostly homopyrimine and homopyrimidine targets. Mixed-sequence PNA generally does not have sufficient binding energy to invade B-DNA. Progress has been made to relax the sequence constraint with the application of tail-clamp<sup>18,19</sup> and double-duplex invasion strategies<sup>20</sup>, but they are not without limitations<sup>21</sup>. More recently we showed that PNA, when preorganized into a right-handed helical motif by installing an (*R*)-stereogenic center at the  $\gamma$ -backbone (Fig. 1b), can invade B-DNA without sequence limitation (Fig. 1c)<sup>22</sup>. This newly endowed property of PNA has been exploited in a number of biological applications, including diagnostics<sup>23–25</sup> and gene editing<sup>26</sup>. Despite this recent advance, it remains a challenge to invade B-DNA at elevated, physiologically relevant ionic strengths. With increasing ionic strengths, particularly in the presence of divalent metal cations, such as  $Mg^{2+}$ , the DNA double helix becomes significantly thermodynamically more stable. As such, additional binding energy would be required to invade B-DNA, a major hurdle facing the design of oligonucleotide molecules for targeting mixed-sequence double helical B-DNA (or RNA). It should be noted that our work is focused on targeting double stranded DNA *in vitro* and not in its native state in a cell, which is epigenetically modified, wrapped around histones, and bound to other proteins.

The required binding energy, in principle, could be attained by replacing natural nucleobases with synthetic analogs<sup>27</sup>, such as the replacements of adenine with 2,6-diaminopurine<sup>28</sup> and cytosine with G-clamp<sup>29</sup>, with improved hydrogen-bonding and base-stacking interactions. However, such an approach could compromise recognition-specificity due to the propensity of the resultant ultra-high affinity probes to bind to DNA (or RNA) with closely related sequences, in addition to their intended



**Fig. 1** Chemical structures and binding modes of  $\gamma$ PNAs. Structure of (a) PNA and (b)  $\gamma$ PNA, (c) binding mode of  $\gamma$ PNA containing natural nucleobases, and (d) that containing JBs. e Hydrogen-bonding interactions of natural bases, and (f) that of JBs with the canonical base-pairs

binding sites. The challenge is in how to design oligonucleotide molecules that can invade B-DNA under physiologically relevant conditions without compromising sequence-specificity.

Here we report the use of a class of shape-selective, bifacial nucleic acid recognition elements for targeting double helical DNA with improvements in binding energy and sequence-specificity, enabling an effective strand invasion of double helical B-DNA, as well as RNA, under a physiologically simulated condition.

## Results

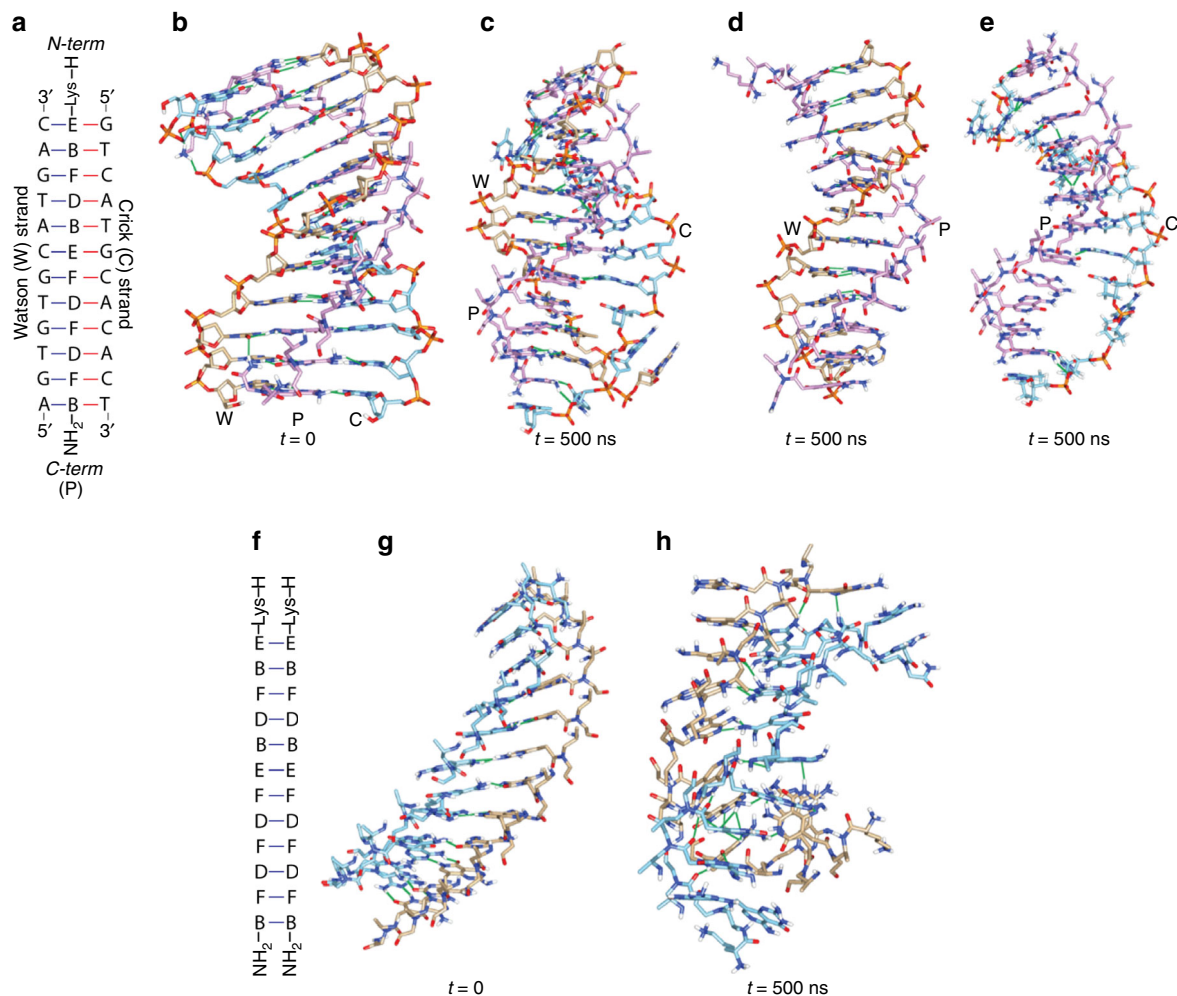
**Design rationale of bifacial recognition.** To augment the binding energy of  $\gamma$ PNA, while simultaneously improving its recognition specificity, we envisioned the application of Janus bases (or JB) as recognition elements since they have the potential to form hydrogen-bonds with nucleobases in both strands of the DNA double helix (Fig. 1d). We expected their interactions with double helical DNA to be more favorable than that of natural nucleobases (Fig. 1e), as a result of a significant increase in the number of hydrogen-bonding and in the degree of base-stacking interactions of the resulting triplex as compared to that of a duplex (Fig. 1f). An improvement in base-stacking interactions was anticipated due to the expanded aromatic ring-systems of JB and as a result of the formation of base-triads. For every two hydrogen-bonds in an A–T or T–A pair, and for every three in a C–G or G–C that are broken, five new ones are formed upon the invasion of DNA by JB $\gamma$ PNA. Moreover, we expected their binding to be more sequence-specific because a single base-mismatch that would normally occur on one face of a natural nucleobase would occur on both faces of a JB. The concept of bifacial nucleic acid recognition, however, is not new. It was conceived by Lehn<sup>30</sup> more than two decades ago and, subsequently, expounded upon by others in the development of Janus-wedges<sup>31–38</sup>. Despite concerted efforts from several research groups, only a small set of Janus-wedges has been developed, and they vary considerably in shapes and sizes, such that they cannot be effectively combined in a modular format for recognition of a non-homogenous nucleic acid sequence. Examples include those developed by Chen and McLaughlin<sup>31,32</sup>, which contained single-aromatic ring systems. As such, not only are they degenerate in recognition, unable to distinguish a C–G from G–C pair or an A–T from T–A, they neither possess the necessary binding energy to invade B-DNA nor a means to suppress self-hybridization. The latter is an intrinsic property of JB, one that presents a major challenge in the design of bifacial nucleic acid recognition elements for targeting (canonical) double helical DNA or RNA. In contrast, the JB under development are novel in structures and enabling in recognition capabilities. They were strategically designed and optimized so that they contain the appropriate shapes, sizes, chemical functionalities, and tautomeric structures for proper recognition of the respective DNA base-pairs (Fig. 1f), in addition to the asymmetry in the helically-folded  $\gamma$ PNA backbone that enables them to hybridize to the designated base-pairs in double helical DNA or RNA, but not to each other.

**Molecular dynamics simulations.** Molecular dynamics (MD) simulations were initially performed to assess the feasibility of JB recognition (Supplementary Table 1). A dodecameric  $\gamma$ PNA containing a mixture of all four JB, H-Lys-EBFDBEFDFB-NH<sub>2</sub>, was chosen as a model system for computational modeling (Fig. 2a). This extended sequence was selected to ensure its proper binding with the Watson strand, as well as the Crick strand of DNA; and should it fail to hybridize to each other, this occurs as a result of steric clash in the backbone and not due to the short sequence. The four C–E–G, G–F–C, A–B–T, and T–D–A base-triads

were built and optimized by the HF/6-31G\* basis set and grafted onto the respective DNA and  $\gamma$ PNA backbone (Supplementary Fig. 1)<sup>39</sup>. The structure of DNA–JB $\gamma$ PNA–DNA was created using the NAB module of AmberTools<sup>40</sup> (Fig. 2b and Supplementary Data 1) and that of JB $\gamma$ PNA–JB $\gamma$ PNA was adopted from an existing NMR structure (Fig. 2f, g, and Supplementary Data 2)<sup>41</sup>. The result showed that the W–P portion of the triplex was stable after 500 ns (Fig. 2c and Supplementary Movie 1), remained intact throughout the simulations (Fig. 2d), while that of the P–C segment displayed a significant structural distortion (Fig. 2e). This is reflected in the number of hydrogen-bonds and in the inter-strand interaction energy (Supplementary Figs. 2–4). The weaker interaction of P–C, as compared to that of W–P, is attributed to the number of hydrogen-bonds being fewer and to the fact that hybridization occurs in a less favorable parallel orientation<sup>42</sup>. In contrast, as the result of steric clash in the backbone, the structure of JB $\gamma$ PNA–JB $\gamma$ PNA unraveled upon restraint release (Fig. 2h, Supplementary Fig. 5 and Supplementary Movie 2). Self-hybridization was a major concern in the design of JB $\gamma$ PNA, or any bifacial nucleic acid recognition system for targeting the canonical base-pairs of DNA or RNA for that matter, due to the complementarity of the two faces of JB. However, we conjectured that such an event is less likely to occur with an asymmetrical, right-handed helically-induced chiral  $\gamma$ PNA than with the achiral counterpart<sup>39,43</sup>. Indeed, the simulations revealed that not only can JB $\gamma$ PNA hybridize to both strands of the DNA double helix, but that it is unable to hybridize to each other—a prerequisite for a successful design of the bifacial nucleic acid system.

**Chemical synthesis and probe design.** Encouraged by the results of MD simulations, we synthesized nucleobases E and F<sup>44</sup>, along with the corresponding  $\gamma$ PNA monomers (Supplementary Fig. 6) and oligomer, for the initial testing (Fig. 3a and Supplementary Figs. 7 and 8). They were chosen based on the presumption that, if JB $\gamma$ PNA containing such recognition elements is able to invade B-DNA with the corresponding sequence, then those that contain B and D, along with their mixture with E and F, should be able to accomplish the same task but more effectively, since G–C and C–G pairs are thermodynamically more stable and, thus, more difficult to invade than A–T and T–A pairs. We selected a relatively short JB $\gamma$ PNA (P1), six units in length, for the initial proof-of-concept study because MD simulations suggest that this length is sufficient for binding DNA (or RNA) (Fig. 3a). P2, a homolog of P1 containing natural nucleobases, was also prepared for comparison. Binding studies were conducted with model DNA targets containing matching sequence and binding-orientation (W1 and C1, Fig. 3b, c), as well as those with mismatched sequence (W2 and C2) and mismatched binding-orientation (W3 and C3). Figure 3d shows the preferential binding orientation of the triplex. Double-stranded DNA and RNA were also prepared (Fig. 3e, f), and their bindings with P1 and P2 were assessed using electrophoretic mobility-shift assays (EMSA). The difference between the two perfectly-matched DNA targets, HP1 (Fig. 3e, i) and HP1D (Fig. 3f), is that the former is an intramolecular hairpin and the latter is an intermolecular duplex—designed to assess the recognition generality of P1.

**Conformational analysis.** Circular dichroism (CD) was employed to assess the conformation of P1, along with those of the bound complexes. Consistent with the previous finding<sup>43</sup>, P1 adopted a right-handed helical motif, as evident by the biphasic exciton coupling pattern with maxima at 235, 255, and 340 nm and minima at 210 and 275 nm (Fig. 3g). We ruled out the possibility for self-hybridization on the basis of



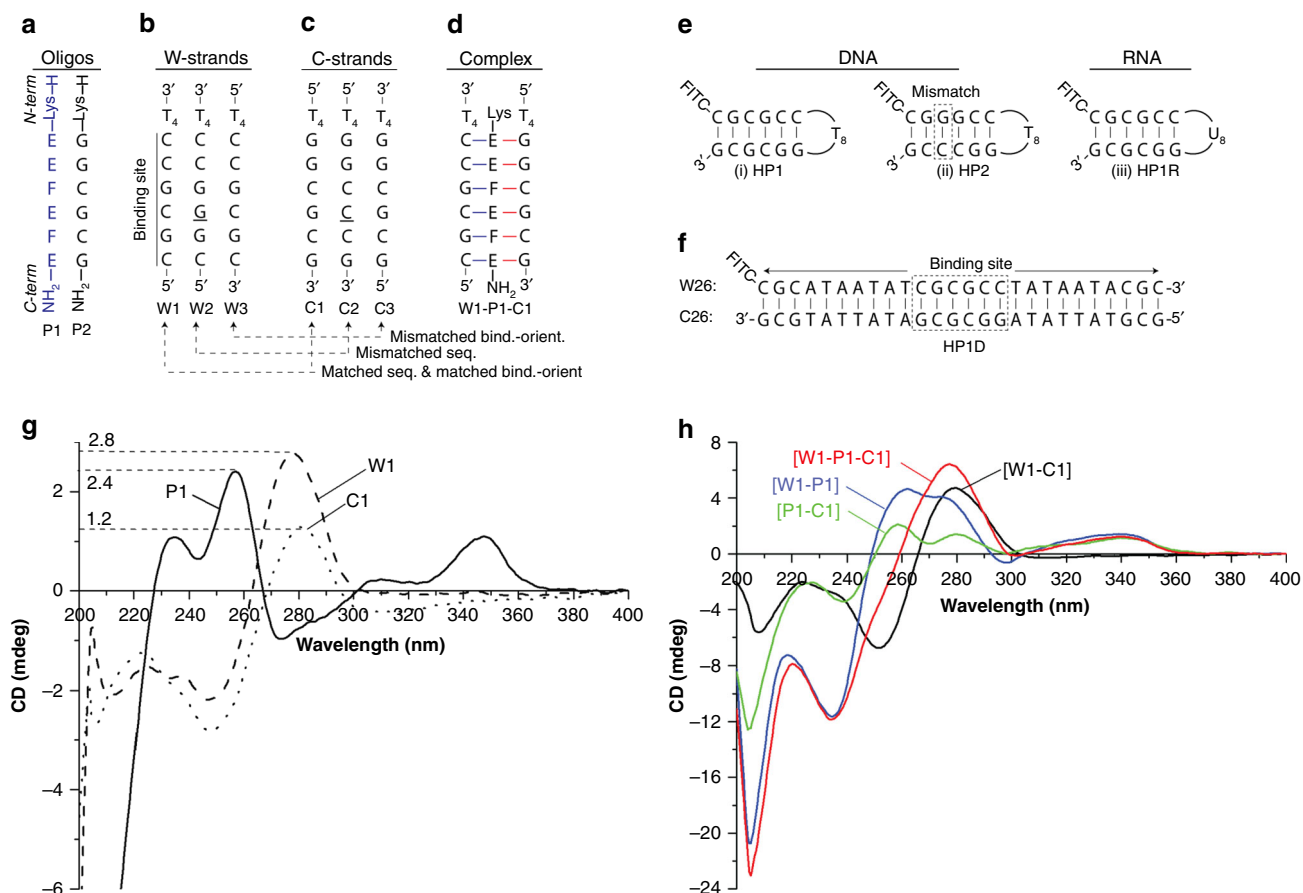
**Fig. 2** MD simulations of the triplex and duplex structures. **a–e** DNA-JByPNA-DNA triplex, and **(f–h)** JByPNA-JByPNA duplex. **a** Sequence of triplex, **(b)** initial structure of triplex, **(c)** simulated structure of triplex after 500 ns, and **(d)** and **(e)** are the same as **(c)** but with the respective C and W strands removed for clarification. **f** Sequence of parallel duplex, **(g)** initial structure of duplex, and **(h)** simulated structure of duplex after 500 ns

concentration-dependent measurements (Supplementary Fig. 9), along with other supporting evidence as discussed in the sections below. The W1-C1 mixture showed modest differential in CD signals. On the other hand, notable differences in the signal strengths and spectral patterns were observed with W1-P1, P1-C1, and W1-P1-C1, in comparison to that of the sum of individual strands (Fig. 3h and Supplementary Fig. 10)—suggesting that their interactions are more favorable than that of W1-C1. Such drastic changes in the CD signals, however, were not observed with the mismatched sequence (W2 and C2) or the mismatched binding-orientation (W3 and C3) (Supplementary Fig. 11). Collectively, these results provide supporting evidence for the binding of P1 with the Watson strand, as well as the Crick strand of DNA double helix in a sequence- and orientation-specific manner.

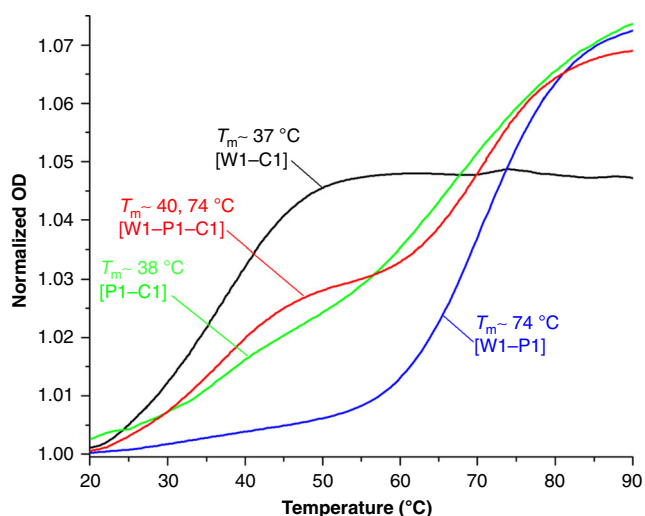
**Thermal and thermodynamic analyses.** UV-melting experiment confirmed the formation of a W1-C1 duplex, although relatively weak, with a melting transition ( $T_m$ ) of  $\sim 37^\circ\text{C}$  (Fig. 4). However, under identical conditions, no discernible  $T_m$  was observed for P1-C1, suggesting that either the duplex did not form or that it did, but its melting profile merged with that of P1, which exhibited a broad melting pattern with  $T_m$  in the  $35\text{--}80^\circ\text{C}$  range (Supplementary Fig. 12). The former scenario is unlikely based on the CD finding. The unusually high and broad  $T_m$  of P1 could be

attributed to the increase in base-stacking interactions of E and F, in comparison to that of the natural nucleobases. A mixture of W1, P1, and C1 at an equimolar ratio displayed two well-defined melting transitions at  $40$  and  $74^\circ\text{C}$  (Fig. 4). We assigned the first transition to the melting of P1-C1 and the second to that of W1-P1. W1-P1 and W1-P2 showed similar  $T_{ms}$ ,  $74^\circ\text{C}$  and  $72^\circ\text{C}$ , respectively (Supplementary Fig. 13). This result was expected since they both possess the same number of hydrogen-bonds. The thermodynamic parameters for W1-C1, W1-P1, and W1-P2 duplexes, as determined by van't Hoff analyses, are shown in Table 1. In contrast to the observations made with W1 and C1, samples containing P1 and mismatched sequence (W2 and C2, Supplementary Fig. 14) or mismatched binding-orientation DNA (W3 and C3, Supplementary Fig. 15) did not yield well-defined melting patterns. These results are consistent with the CD findings, indicating that P1 has strong binding sequence and orientation preferences.

**Confirmation of binding by NMR.** To further substantiate the binding of P1 with W1 and C1, we carried out 1D and 2D NMR experiments at variable temperature and concentration. Figure 5a shows  $^1\text{H-NMR}$  spectra of the various combinations, prepared in a PBS buffer containing  $\text{H}_2\text{O}:\text{D}_2\text{O}$  at a 9:1 ratio; and those of the individual strands are shown in Supplementary Figs. 16–18. Several key observations were noted. First, the NMR spectrum of



**Fig. 3** Probes and nucleic acid targets employed in this study along with their CD spectra. **a** Sequence of P1 (blue) and P2 (black), **(b)** Watson strands (W1, W2, W3), **(c)** Crick strands (C1, C2, C3), **(d)** depiction of a bound W1-P1-C1 triplex, **(e)** intramolecular DNA (HP1 and HP2) and RNA (HP1R) targets, **(f)** intermolecular DNA target (HP1D), **(g)** CD spectra of the individual strands (P1: solid line, W1: dashed line, C1: dotted line), and **(h)** CD spectra of W1-P1-C1 (red line), W1-P1 (blue line), W1-C1 (black line), and P1-C1 (green line) complexes. In **g** and **h** the concentration of each strand was 2.5  $\mu\text{M}$ , prepared in a 1 $\times$  PBS buffer (137 mM NaCl, 2.7 mM KCl, 10 mM NaPi, pH 7.4) and recorded at 25  $^{\circ}\text{C}$



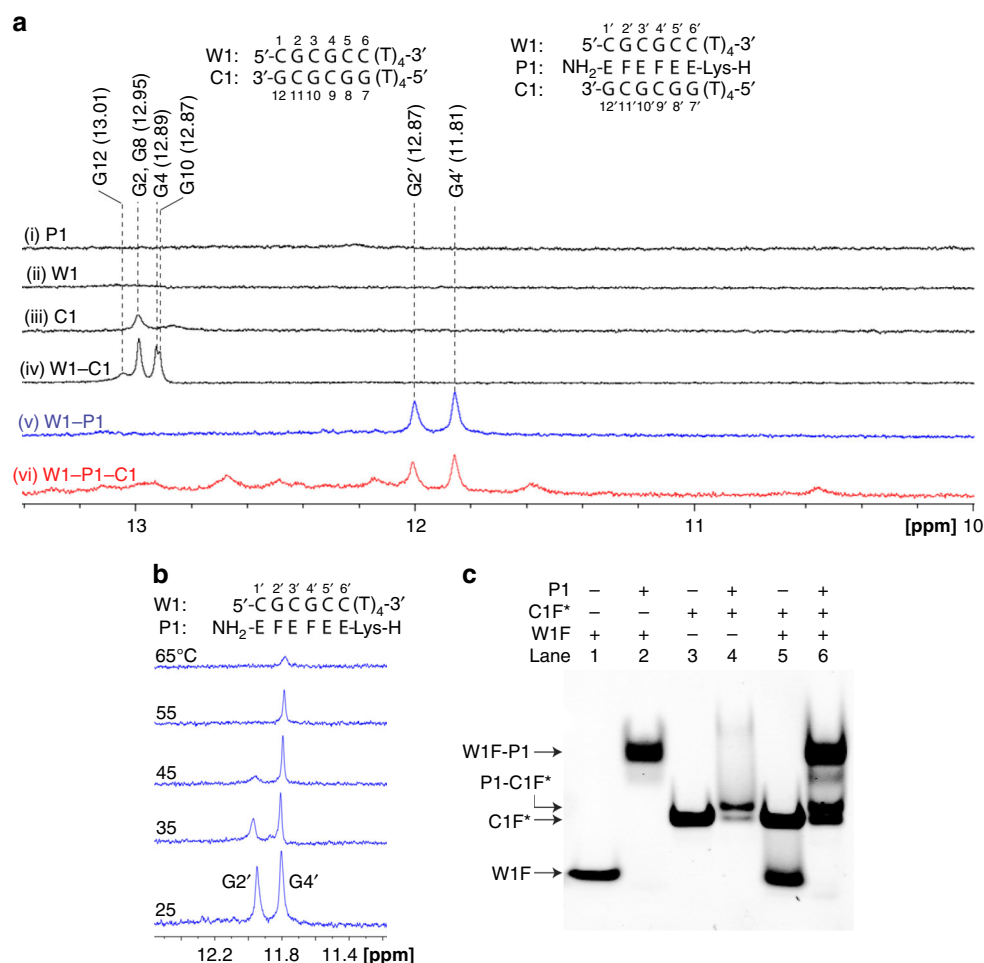
**Fig. 4** Thermal stability of the bound complexes. UV-melting profiles of W1-P1-C1 (red line), W1-P1 (blue line), W1-C1 (black line), and P1-C1 (green line). The concentration of each strand was 2.5  $\mu\text{M}$ , prepared in a 1 $\times$  PBS buffer

**Table 1** Thermodynamic parameters

Complex	$T_m$ ( $^{\circ}\text{C}$ )	$-\Delta H^{\circ}$ (kJ/mol)	$-T\Delta S^{\circ}$ (kJ/mol)	$-\Delta G^{\circ}$ (kJ/mol)
W1-C1	37	244 $\pm$ 22	202 $\pm$ 24	42 $\pm$ 3
W1-P1	74	288 $\pm$ 8	213 $\pm$ 6	75 $\pm$ 1
P1-C1	40	ND	ND	ND
W1-P2	72	325 $\pm$ 17	250 $\pm$ 13	75 $\pm$ 3

ND not determined

P1 was devoid of any imino proton signals in the 10.0–20.0 ppm region (Fig. 5a, i), indicating that self-hybridization did not take place even at such a high concentration (500  $\mu\text{M}$  strand concentration). Second, C1 showed two imino proton signals at 12.95 and 12.80 ppm (Fig. 5a, iii, and Supplementary Fig. 19), as a result of the formation of a partial C1-C1 duplex stabilized by a terminal GG-diad. Self-hybridization did not occur with W1 because it lacked the terminal purine-cap (Fig. 5a, ii). Third, the titration of P1 with W1-C1 resulted in a gradual disappearance of the imino proton signals of G12, G2/G8, G10, and G4 (Fig. 5a, compare *vi* to *iii* and *iv*), concomitant with the formation of that of G2' and G4' of W1-P1 and other imino protons of P1-C1

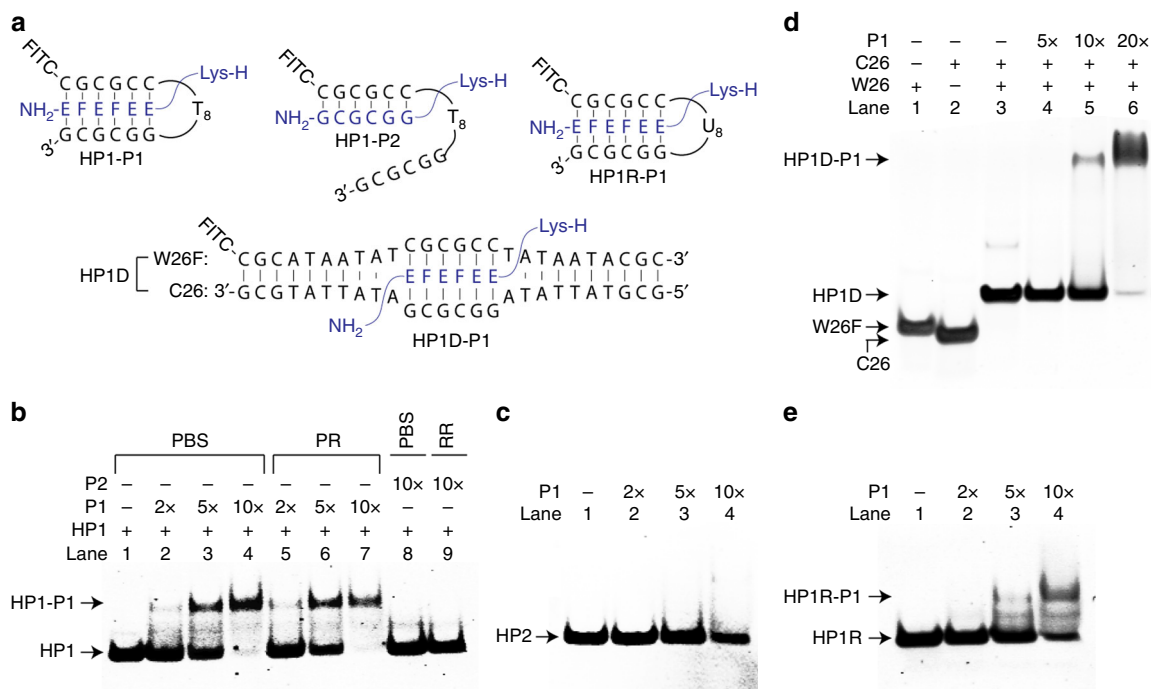


**Fig. 5** Imino proton signals of the bound complexes. **a**  $^1H$ -NMR spectra of the indicated samples in the 10.0–13.4 ppm range (no signals were observed beyond 13.4 ppm). Note P1, W1, C1, W1-C1: black line; W1-P1: blue line; W1-C1-P1: red line. **b** Imino proton signals of a W1-P1 duplex as a function of temperature. **c** Result of an EMSA showing the binding of various partners. Experimental conditions: pre-annealed DNA (W1F or C1F\*) was mixed with P1 in a PR buffer (137 mM NaCl, 2.7 mM KCl, 10 mM NaPi, 2 mM MgCl<sub>2</sub>, pH 7.4) and incubated at 37 °C for 4 h; the resulting mixtures were separated by 10% non-denaturing PAGE. The sequences of FITC-attached DNA are as followed: W1F, 5'-CGCGCC-3'; and C1F\*, 5'-(T)<sub>12</sub>GGCGCG-3'. The poly-(T)<sub>12</sub> tail in C1F\* was employed to provide separation between the two target strands. The strand concentrations of W1F, C1F\*, and P1 were 1 μM each

(Fig. 5a, compare *vi* to *v*, and Supplementary Figs. 20 and 21; the other controls are shown in Supplementary Figs. 22–27). COSY and NOESY experiments enabled the assignment of imino protons of W1-C1 (Supplementary Figs. 28–33 and Supplementary Table 2) and W1-P1 (Supplementary Figs. 34–39). Fourth, W1-P1 complex appeared to be highly stable, as judging from the line sharpness and the persistence of imino proton signals at a temperature as high as 65 °C (Fig. 5b and Supplementary Figs. 40 and 41). While it is not currently feasible to assign the imino proton signals of E and F due to significant spectral overlaps with that of DNA nucleobases, a cursory inspection of W1-P1-C1 complex (Supplementary Fig. 42) confirms the binding of P1 with W1 and C1, as noted from the emergence of the aromatic W1-P1 proton signals (asterisk) and from the disappearance of that of C1 (filled circle). This suggestion was further corroborated by findings from EMSA (Fig. 5c) and MALDI-TOF MS (Supplementary Fig. 43) experiments, showing W1-P1 duplex being the most stable (compare lane 2 to lane 1), followed by P1-C1 (compare lane 4 to lane 3) and W1-C1 (compare lane 5 to lanes 1 and 3).

**Assessment of DNA strand invasion.** To determine whether JByPNA can invade a double helical B-DNA, we carried out EMSA comparing the binding of P1 and P2 with model DNA

targets containing a perfectly-matched HP1 (Fig. 3e, *i*) and a base-pair inversion HP2 (Fig. 3e, *ii*) at a moderate ionic strength (1xPBS buffer), as well as at a physiologically relevant (PR) condition<sup>45</sup>. Figure 6a depicts compositions of the various binding complexes. The result demonstrated that not only can P1 invade a highly stable HP1 in PBS (Fig. 6b, lanes 2–4), but also in a PR buffer (lanes 5–7). Formation of the HP1-P1 invasion complex was confirmed by MALDI-TOF MS (Supplementary Fig. 44). The UV-melting profiles of the individual duplexes are shown in Supplementary Fig. 45. Binding occurred in a concentration-dependent manner, resulting in a complex that remained intact throughout the electrophoresis—as inferred from the sharpness of the shifted bands. Binding was relatively fast, complete within 10 min at 37 °C (Supplementary Fig. 46). In contrast, no evidence of binding was observed with P2, not even in a PBS buffer (compare lanes 8 and 9 to lanes 4 and 7, respectively). The fact that P1 and P2 have similar binding energy with the Watson strand (Table 1), but that only P1 was able to invade HP1 in PBS, as well as in a PR buffer, highlights the significant contribution of hydrogen-bonding interactions and, perhaps, steric clashes with the Crick strand of DNA double helix in stabilization of the invasion complex. No binding was observed with the mismatched HP2 (Fig. 6c). In addition to the hairpin



**Fig. 6** Invasion of double-stranded DNA and RNA by P1 and P2. **a** Compositions of the various complexes (HP1-P1, HP1-P2, HP1R-P1, HP1D-P1). **b** Comparison of binding of P1 (lanes 2–7) and P2 (lanes 8 and 9) with the perfectly-matched DNA hairpin (HP1) at a moderate (1× PBS) (lanes 1–4, and 8) and at a physiologically relevant (PR) ionic strength (lanes 5–7, and 9). The concentration of HP1 was 1 μM and that of P1 and P2 are as shown. The samples were prepared in the indicated buffers and incubated at 37 °C for 4 h prior to separation by non-denaturing PAGE. **c** Result of EMSA of P1 with mismatched HP2, **(d)** with HP1D, and **(e)** with HP1R RNA. In **c–e**, the samples were prepared in a PR buffer and incubated at the same temperature and duration, and separated under identical conditions as that in **b**. PR buffer: 137 mM NaCl, 2.7 mM KCl, 10 mM NaPi, 2 mM MgCl<sub>2</sub>, pH 7.4

structure, P1 was also able to target an internal binding site embedded within a highly stable double helix (Supplementary Fig. 45), as demonstrated with HP1D (Fig. 3f and Fig. 6d). The result showed that being able to simultaneously target both strands of the DNA double helix is essential for the invasion of B-DNA at a physiologically relevant ionic strength. A similar observation was made by Nielsen<sup>20</sup> in the invasion of DNA by pseudo-complementary PNAs, *albeit* strand invasion requires simultaneous binding of two separate strands of PNA. Such a design, however, has practical limitations due to the slow hybridization kinetics and low invasion efficiency<sup>21</sup>.

**Strand invasion of RNA.** Interestingly, we noticed that P1 was also able to invade RNA double helix, HP1R (Fig. 3e, *iii*), in a PR buffer (Fig. 6e), where P2 was not. This result is intriguing, suggesting the possibility for targeting the secondary and tertiary structures of RNA. Such RNA structural motifs are highly abundant in cells and are known to play key roles in many biological processes<sup>46</sup>. Generally, they are challenging to target with conventional antisense oligonucleotides<sup>47,48</sup> or with small-molecule ligands with high selectivity<sup>49,50</sup>, although some progress has been made with the development and approval of nusinersen<sup>51</sup> and patisiran drugs<sup>52</sup>, and in the application of oligonucleotides for targeting RNA-repeated expansions<sup>53</sup>.

## Discussion

The A–T (or A–U) and C–G base-pairing interactions are employed by Nature in the storage and transmission of genetic information, and by researchers in the recognition of genetic materials because of their exquisite recognition sequence-specificity. However, such principles have rarely been applied in the

recognition of double helical DNA or RNA because most oligonucleotide molecules developed to date do not have sufficient binding energy to invade such a canonical duplex structure. The present study is focused on determining whether a bifacial JByPNA could be designed, and, if so, whether it could invade a double helical DNA (or RNA) at a physiologically relevant ionic strength. A critical requirement for the successful development of such a dual-recognition nucleic acid system is its preferential binding with DNA (or RNA) but not with itself. We showed that not only can JByPNA be prepared, but that it can form hydrogen-bonds with nucleobases in both strands of the DNA double helix without any evidence for self-hybridization. The helically-induced chirality in the backbone prevents JByPNA from approaching and properly hybridizing to each other. In addition to its tight binding with the Watson strand, and to a lesser extent with the Crick strand, JByPNA can invade double helical B-DNA, as well as RNA, under a simulated physiological condition.

A significant structural perturbation in the P-C segment of triplex, as revealed by MD simulations, was surprising but not unexpected. This is because the number of H-bonds formed in P-C (12 H-bonds) is fewer than that in W-P (18 H-bonds), and binding occurs in an unfavorable parallel orientation. This result was substantiated by experimental findings, as demonstrated by CD, UV-melting, NMR, and EMSA measurements. The fact that both P1 and P2 form the same number of H-bond with the Watson strand and that the resulting duplexes exhibited virtually identical thermodynamic stability, but P1 was able to invade B-DNA whereas P2 was not, indicates the importance of being able to engage the Crick strand in stabilization of the invasion complex. It is well established that DNA double helix is relatively dynamic, with the deoxyribose phosphate backbones and the nucleobases in constant flux with continual bending and twisting

motions<sup>54</sup>. The reason that most oligonucleotide molecules are unable to productively invade B-DNA is not due to the lack of base-pair accessibility, but rather due to their inability to compete with the complementary DNA strand, especially in the context of a relatively long double helix<sup>27</sup>. In addition to the binding energy gained in the formation of P-C, although relatively small compared to the overall contribution, their interactions may provide a physical barrier in preventing the complementary DNA strand from re-hybridizing with its partner. Such a phenomenon has been illustrated by Nielsen and coworkers<sup>20</sup>.

While this proof-of-concept study is focused on the synthesis and evaluation of the binding property of a relatively short probe, the binding energy that would be required to invade a biologically relevant genomic DNA target, estimated to be ~18-nt in length, could be attained by expanding the probe length accordingly. This is because a  $\gamma$ PNA-DNA duplex is thermodynamically more stable than DNA-DNA on a per-unit basis, as demonstrated in a previous study.<sup>55</sup> We expected the thermodynamic stability of JByPNA-DNA to be even greater than that of  $\gamma$ PNA-DNA. If necessary, additional binding energy could be attained by employing second-generation JBs, such as those shown in Supplementary Fig. 47a, with improved hydrogen-bonding and base-stacking interactions. For instance, E\* and F\*\* are capable of forming six hydrogen-bonds each with the respective C-G and G-C base-pairs, as compared to five each for E and F. Although F\*\* is not the most stable isomer, previous studies showed that tautomers interconvert on a relatively fast time scale, and that rare occurrences can occur if they aid in the folding and stabilization of nucleic acid structures<sup>56</sup>.

Another potential benefit of a bifacial nucleic system is that it could be designed to bind more favorably with a double-stranded over a single-stranded DNA (or RNA) by furnishing it with a protective companion (Supplementary Fig. 47b). The application of toehold and strand displacement reaction to tune the binding kinetics and thermodynamics of nucleic acids—for the purpose of controlling their binding selectivity—has already been demonstrated<sup>57</sup>. Since binding occurs on both sides, a relatively short companion strand would be suffice to provide protection against single-stranded DNA (or RNA) binding, and maintain the required kinetics and thermodynamics for a successful invasion of double helical DNA (or RNA)<sup>58</sup>. Besides its protective role, the companion strand could be functionalized with a specific chemical group, such as guanidine, for improving cellular uptake<sup>59</sup>, and possibly for enhancing the rate of DNA (or RNA) strand invasion<sup>60</sup>. A distinctive advantage of such a molecular feature is that once it is displaced, upon a successful invasion of DNA (or RNA) by the designer probe, the companion strand is quenched through self-hybridization. Such a probe design may provide a safer and more effective means for targeting double-stranded DNA (or RNA) than the conventional antisense or antigene approach for biomedical applications.

In addition to the four JBs shown in Fig. 1f, which are designed to bind to the canonical base-pairs, the remaining twelve JBs within this class that are designed to bind to non-canonical base-pairs<sup>44</sup> could be employed in combination to target the secondary and tertiary structures of RNA, in attempts to elucidate their physiological functions and to develop therapeutic interventions for treating genetic disorders. However, this does not imply that conventional antisense molecules, those containing natural nucleobases, cannot be effectively used to manipulate the structures and functions of RNA. In fact, there are a number of diseases that are considered ‘undrugable’ by small molecules that can be corrected by targeting single-stranded RNA<sup>61</sup>—case in point are the recent development and approval of nusinersen (Spinraza) by Ionic Pharmaceuticals and patisiran (Onpattro) by Alnylam Pharmaceuticals for treatment of the respective spinal muscular

atrophy and hereditary transthyretin-mediated amyloidosis. The present work provides another dimension in the design of ‘millimolecular’ oligonucleotide molecules for targeting nucleic acid biopolymers, especially those structured regions that may be difficult to access by conventional reagents.

In summary, we have shown that  $\gamma$ PNA containing a selected set of JBs could be developed and could hybridize to both strands of the DNA double helix without undergoing self-hybridization. JByPNA is able to invade a highly stable double helical B-DNA at a physiologically relevant ionic strength, where a homolog containing natural nucleobases is not. Overall, the bifacial recognition mode is general, applicable to targeting not only double helical DNA, but also RNA. Due to their tight binding, significantly shorter probes could potentially be used to target the secondary and tertiary structures of RNA in our efforts to decipher their physiological roles and to develop molecular therapies for treating genetic disorders.

## Methods

**Molecular dynamics simulations.** See Supplementary Methods.

**UV-melting analysis.** All UV-melting samples were prepared by mixing probes with DNA targets at the indicated concentrations and in the appropriate buffers, and annealed by incubation at 90 °C for 5 min followed by gradual cooling to room temperature. UV melting curves were collected using Agilent Cary UV-Vis 300 spectrometer equipped with a thermoelectrically controlled multi-cell holder. UV-melting spectra were collected by monitoring UV-absorption at 260 nm from 20 to 95 °C in the heating runs, and from 95 to 20 °C in the cooling runs, both at the rate of 1 °C per min. The heating and cooling curves were nearly identical, indicating that the hybridization process is reversible. The recorded spectra were smoothed using a 10-point adjacent averaging algorithm. The first-order derivatives of the melting curves were taken to determine the melting temperatures of the duplex and single strand.

**Circular dichroism analysis.** The samples were prepared in a 1× PBS buffer. All circular dichroism (CD) spectra represent an average of at least 15 scans collected at a rate of 100 nm/min between 200–400 nm, in a 1-cm path-length cuvette at 25 °C. The CD spectrum from buffer solution was subtracted from the sample spectra, which were then smoothed via a 10-point adjacent averaging algorithm.

**Nuclear magnetic resonance analysis.** All 1D nuclear magnetic resonance (NMR) spectra for DNA and DNA:PNA complex were recorded on a 500 MHz spectrometer using Watergate p3919gp in 90:10% H<sub>2</sub>O:D<sub>2</sub>O and in D<sub>2</sub>O solvent. All 2D NOESY spectra for DNA and DNA:PNA complex were recorded on a 500 MHz spectrometer using Watergate NOESY or noesygpph19 as a pulse program with 100, 200, and 300 ms time scale. All COSY spectra were recorded using cosygpprqf as the pulse program.

**Mass spectrometric analysis.** For PNA, a solution of  $\alpha$ -cyano-4-hydroxycinnamic acid (10 mg of  $\alpha$ -cyano-4-hydroxycinnamic acid in 500  $\mu$ L of water with 0.1% TFA and 500  $\mu$ L of acetonitrile with 0.1% TFA) was used as the matrix for MALDI-TOF analysis. The PNA samples were prepared by mixing with 2  $\mu$ L of matrix and 1  $\mu$ L PNA (1–5  $\mu$ M) at 37 °C. MALDI-TOF analysis was performed about 10 min after spotting. For DNA and DNA:PNA complex, a solution of 6-aza-2-thiothymine (ATT, 10 mg/mL) in 1:1 (v/v) CH<sub>3</sub>CN/ammonium citrate [20 mM] were prepared. The DNA and DNA:PNA complex were prepared by mixing 2  $\mu$ L of matrix and 1  $\mu$ L of DNA or DNA:PNA complex (either in RNase-free water or in a 1× PBS buffer). The MALDI-TOF plate was dried in vacuum desiccator for 10 min after spotting the sample and analyzed by MALDI-TOF MS.

**Electrophoretic mobility-shift assay.** All the DNA samples were prepared in the indicated buffers, either in 1× PBS (137 mM NaCl, 2.7 mM KCl, 10 mM NaPi, pH 7.4) or in a PR buffer (137 mM NaCl, 2.7 mM KCl, 10 mM NaPi, 2 mM MgCl<sub>2</sub>, pH 7.4), and annealed by heating to 90 °C for 5 min followed by gradual cooling to room temperature. PNA and DNA targets were mixed at the indicated concentrations and incubated at 37 °C for 4 h. The samples were then loaded onto 10% non-denaturing PAGE with 1× Tris-borate buffer and electrophoretically separated at 120 V for 60 min. The fluorescent probe attached to DNA was visualized by UV-Transilluminator. The gels without fluorescent probe were stained with SYBR-Gold and visualized by UV-Transilluminator.

**Preparation and characterization of PNA P1 and P2 and biophysical studies with DNA.** See Supplementary Methods.

**Data availability**

All data are available from the corresponding authors upon reasonable request.

Published online: 07 November 2018

**References**

- Watson, J. D. & Crick, F. H. C. A structure for deoxyribose nucleic acid. *Nature* **171**, 737–738 (1953).
- Kopka, M. L., Yoon, C., Goodsell, D., Pjura, P. & Dickerson, R. E. The molecular origin of DNA-drug specificity in netropsin and distamycin. *Proc. Natl Acad. Sci. USA* **82**, 1376–1380 (1985).
- Brown, D. D. The role of stable complexes that repress and activate eucaryotic genes. *Cell* **37**, 359–365 (1984).
- Kielkopf, C. L. et al. A structural basis for recognition of A-T and T-A base pairs in the minor groove of B-DNA. *Science* **282**, 111–115 (1998).
- Sun, J. S. & Helene, C. Oligonucleotide-directed triple-helix formation. *Curr. Opin. Struct. Biol.* **3**, 345–356 (1993).
- Greisman, H. A. & Pabo, C. O. A general strategy for selecting high-affinity zinc finger proteins for diverse DNA target sites. *Science* **275**, 657–661 (1997).
- Jantz, D., Amann, B. T., Gatto, G. J. & Berg, J. M. The design of functional DNA-binding proteins based on zinc finger domains. *Chem. Rev.* **104**, 789–800 (2004).
- Moscou, M. J. & Bogdanove, A. J. A simple cipher governs DNA recognition by TAL effectors. *Science* **326**, 1501 (2009).
- Mak, A. N.-S., Bradley, P., Cernadas, R. A., Bogdanove, A. J. & Stoddard, B. L. The crystal structure of TAL effector PthXo1 bound to its DNA target. *Science* **335**, 716–719 (2012).
- Jinek, M. et al. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816–821 (2012).
- Lander, E. S. The heroes of CRISPR. *Cell* **164**, 18–28 (2016).
- Tsai, S. Q. et al. GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. *Nat. Biotech.* **33**, 187–197 (2015).
- Wang, X. et al. Unbiased detection of off-target cleavage by CRISPR-Cas9 and TALENs using integrase-defective lentiviral vectors. *Nat. Biotech.* **33**, 175–178 (2015).
- Nielsen, P. E., Egholm, M., Berg, R. H. & Buchardt, O. Sequence-selective recognition of DNA by strand displacement with a thymine-substituted polyamide. *Science* **254**, 1497–1500 (1991).
- Nielsen, P. E. Peptide nucleic acid. A molecule with two identities. *Acc. Chem. Res.* **32**, 624–630 (1999).
- Nielsen, P. E. & Christensen, L. Strand displacement binding of a duplex-forming homopurine PNA to a homopyrimidine duplex DNA target. *J. Am. Chem. Soc.* **118**, 2287–2288 (1996).
- Egholm, M. et al. PNA hybridizes to complementary oligonucleotides obeying the Watson-Crick hydrogen-bonding rules. *Nature* **365**, 566–568 (1993).
- Kaihatsu, K., Shah, R. H., Zhao, X. & Corey, D. R. Extending recognition by peptide nucleic acids (PNAs): binding to duplex DNA and inhibition of transcription by tail-clamp PNA-peptide conjugate. *Biochemistry* **42**, 13996–14003 (2003).
- Bentin, T., Larsen, H. J. & Nielsen, P. E. Combined triplex/duplex invasion of double-stranded DNA by "tail-clamp" peptide nucleic acid. *Biochemistry* **42**, 13987–13995 (2003).
- Lohse, J., Dahl, O. & Nielsen, P. E. Double duplex invasion by peptide nucleic acid: A general principle for sequence-specific targeting of double-stranded DNA. *Proc. Natl Acad. Sci. USA* **96**, 11804–11808 (1999).
- Demidov, V. V. et al. Kinetics and mechanism of the DNA double helix invasion by pseudocomplementary peptide nucleic acids. *Proc. Natl Acad. Sci. USA* **99**, 5953–5958 (2002).
- Bahal, R., Sahu, B., Rapireddy, S., Lee, C.-M. & Ly, D. H. Sequence-unrestricted, Watson-Crick recognition of double helical B-DNA by (R)-MiniPEG-gPNAs. *ChemBiochem* **13**, 56–60 (2012).
- Singer, A., Rapireddy, S., Ly, D. H. & Meller, A. Electronic barcoding of a viral gene at the single-molecule level. *Nano. Lett.* **12**, 1722–1728 (2012).
- Nolling, J. et al. Duplex DNA-invading  $\gamma$ -modified peptide nucleic acids enable rapid identification of bloodstream infections in whole blood. *mBio* **7**, e00345–00316 (2016).
- Morin, T. J. et al. Nanopore-based target sequence detection. *PLoS One* **11**, e0154426 (2016).
- Bahal, R. et al. In vivo correction of anemia in  $\beta$ -thalassemic mice by yPNA-mediated gene editing with nanoparticle delivery. *Nat. Commun.* **7**, 13304 (2016).
- Rapireddy, S., Bahal, R. & Ly, D. H. Strand invasion of mixed-sequence, double-helical B-DNA by  $\gamma$ -peptide nucleic acids containing G-clamp nucleobases under physiological conditions. *Biochemistry* **50**, 3913–3918 (2011).
- Haaime, G., Hansen, H. F., Christensen, L., Dahl, O. & Nielsen, P. E. Increased DNA binding and sequence discrimination of PNA oligomers containing 2,6-diaminopurine. *Nucleic Acids Res.* **25**, 4639–4643 (1997).
- Chenna, V. et al. A simple cytosine-to-G-clamp nucleobase substitution enables chiral gamma-PNAs to invade mixed-sequence double helical B-form DNA. *ChemBiochem* **9**, 2388–2391 (2008).
- Branda, N. & Lehn, J.-M. JANUS WEDGES: A new approach towards nucleobase-pair recognition. *Chem. Commun.* **0**, 2443–2444 (1996).
- Chen, D., Meena, Sharma, S. K. & McLaughlin, L. W. Formation and stability of a janus-wedge type of DNA triplex. *J. Am. Chem. Soc.* **126**, 70–71 (2004).
- Chen, H., McLaughlin, M. & McLaughlin, L. M. A janus-wedge DNA triplex with A-W1-T and G-W2-C base triplets. *J. Am. Chem. Soc.* **130**, 13190–13191 (2008).
- Arambula, J. F., Ramisetty, S. R., Baranger, A. M. & Zimmerman, S. C. A simple ligand that selectively targets CUG trinucleotide repeats and inhibits MBNL protein binding. *Proc. Natl Acad. Sci. USA* **106**, 16068–16073 (2009).
- Zhao, H. et al. Synthesis of a complete Janus-type guanosine-cytosine base and its 2'-deoxyribonucleoside. *Chem. Lett.* **40**, 684–686 (2011).
- Shin, D. & Tor, Y. Bifacial nucleoside as a surrogate for both T and A in duplex DNA. *J. Am. Chem. Soc.* **133**, 6926–6929 (2011).
- Zeng, Y., Pratumyot, Y., Piao, X. & Bong, D. Discrete assembly of synthetic peptide-DNA triplex structures from polyvalent melamine-thymine bifacial recognition. *J. Am. Chem. Soc.* **134**, 832–835 (2012).
- Artigas, G. & Marchan, V. Synthesis of janus compounds for the recognition of G-U mismatched nucleobase pairs. *J. Org. Chem.* **78**, 10666–10677 (2013).
- Robinson, C. J. et al. Modular riboswitch toolsets for synthetic genetic control in diverse bacterial species. *J. Am. Chem. Soc.* **136**, 10615–10624 (2014).
- Yeh, J. I. et al. Crystal structure of chiral gammaPNA with complementary DNA strand: Insights into the stability and specificity of recognition and conformational preorganization. *J. Am. Chem. Soc.* **132**, 10717–10727 (2010).
- Macke, T. J. & Case, D. A. Modeling unusual nucleic acid structures. *Molecular Modeling of Nucleic Acids*. 379–393 (1997).
- He, W. et al. The structure of a gamma-modified peptide nucleic acid duplex. *Mol. Biosyst.* **6**, 1619–1629 (2010).
- Uhlmann, E., Peyman, A., Breipohl, G. & Will, D. W. PNA: synthetic polyamide nucleic acids with unusual binding properties. *Angew. Chem., Int. Ed.* **37**, 2796–2823 (1998).
- Dragulescu-Andrasi, A. et al. A simple gamma-backbone modification preorganizes peptide nucleic acid into a helical structure. *J. Am. Chem. Soc.* **128**, 10258–10267 (2006).
- Thadke, S. A. et al. Design of bivalent nucleic acid ligands for recognition of RNA-repeated expansion associated with Huntington's disease. *Biochemistry* **57**, 2094–2108 (2018).
- Santoro, S. W. & Joyce, G. F. A general purpose RNA-cleaving DNA enzyme. *Proc. Natl Acad. Sci. USA* **94**, 4262–4266 (1997).
- Mortimer, S. A., Kidwell, M. A. & Doudna, J. A. Insights into RNA structure and function from genome-wide studies. *Nat. Rev. Gen.* **15**, 469–479 (2014).
- Smith, L. M., Andersen, K. B., Hovgaard, L. & Jaroszewski, J. W. Rational selection of antisense oligonucleotide sequences. *Eur. J. Pharm. Sci.* **11**, 191–198 (2000).
- Tu, G.-c., Cao, Q.-n., Zhou, F. & Israel, Y. Tetranucleotide GGGG motif in primary RNA transcripts. *J. Biol. Chem.* **273**, 25125–25131 (1998).
- Thomas, J. R. & Hergenrother, P. J. Targeting RNA with small molecules. *Chem. Rev.* **108**, 1171–1224 (2008).
- Guan, L. & Disney, M. D. Recent advances in developing small molecules targeting RNA. *Acc. Chem. Biol.* **7**, 73–86 (2012).
- Corey, D. R. Nusinersen, an antisense oligonucleotide drug for spinal muscular atrophy. *Nat. Neurosci.* **20**, 497–499 (2017).
- Garber, K. Alnylam launches era of RNAi drugs. *Nat. Biotech.* **36**, 777–778 (2018).
- Wheeler, T. M. et al. Targeting nuclear RNA for in vivo correction of myotonic dystrophy. *Nature* **488**, 111–115 (2012).
- Brauns, E. B., Murphy, C. J. & Berg, M. A. Local dynamics in DNA by temperature-dependent Stokes Shift of an intercalated dye. *J. Am. Chem. Soc.* **120**, 2449–2456 (1998).
- He, G., Rapireddy, S., Bahal, R., Sahu, B. & Ly, D. H. Strand invasion of extended, mixed sequence B-DNA by gammaPNAs. *J. Am. Chem. Soc.* **131**, 12088–12090 (2009).
- Singh, V., Fedeles, B. I. & Essignmann, J. M. Role of tautomerism in RNA biochemistry. *RNA* **21**, 1–13 (2015).
- Zhang, D. Y. & Seelig, G. Dynamic DNA nanotechnology using strand-displacement reactions. *Nat. Chem.* **3**, 103–113 (2011).
- Zhang, D. Y. & Winfree, E. Control of DNA strand displacement kinetics using toehold exchange. *J. Am. Chem. Soc.* **131**, 17303–17314 (2009).

59. Stanzl, E. G., Trantow, B. M., Vargas, J. R. & Wender, P. A. Fifteen years of cell-penetrating, guanidinium-rich molecular transporters: Basic science, research tools, and clinical applications. *Acc. Chem. Res.* **46**, 2944–2954 (2013).
60. Corey, D. R. 48,000-fold acceleration of hybridization of chemically modified oligonucleotides. *J. Am. Chem. Soc.* **117**, 9373–9374 (1995).
61. Rinaldi, C. & Wood, M. J. A. Antisense oligonucleotides: the next frontier for treatment of neurological disorders. *Nat. Rev. Neurol.* **14**, 9–21 (2018).

### Acknowledgements

A.M. would like to thank SERB (Grant no. EMR/2016/001069), Department of Science and Technology, Government of India. Nuclear Magnetic Resonance and MALDI-TOF MS instrumentations at Carnegie Mellon University were partially supported by the National Science Foundation (CHE-9808188, CHE0130903, and CHE-1039870). D.H.L. and S.A.T. would like to thank the National Institutes of Health (R21NS098102), the National Science Foundation (CHE-1609159), and the DSF Charitable Foundation for the financial supports.

### Author contributions

S.A.T. and D.H.L. designed the experiments. S.A.T. conducted all the experiments except MD simulations. J.D.R.P. performed binding studies. V.M.H. and A.M. carried out MD simulations. S.A.T., R.R.G. and D.H.L. designed NMR experiments and interpreted the NMR data. All authors wrote the manuscript and supporting information. All authors have given approval to the final version of the manuscript.

### Additional information

**Supplementary information** accompanies this paper at <https://doi.org/10.1038/s42004-018-0080-5>.

**Competing interests:** The authors declares no competing interests.

**Reprints and permission** information is available online at <http://npg.nature.com/reprintsandpermissions/>

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018