ChemComm

Cite this: *Chem. Commun.*, 2011, **47**, 11125–11127

www.rsc.org/chemcomm

COMMUNICATION

Triple helical recognition of pyrimidine inversions in polypurine tracts of RNA by nucleobase-modified PNA†

Pankaj Gupta,‡ Thomas Zengeya‡ and Eriks Rozners*

Received 31st July 2011, Accepted 24th August 2011
DOI: 10.1039/c1cc14706d

Peptide nucleic acids containing 2-pyrimidinone (P) and 3-oxo-2,3-dihydropyridazine (E) heterocycles recognized C–G and U–A inversions in a polypurine tract of double helical RNA with high affinity and sequence selectivity at pH 6.25. E-modified PNA bound strongly to bacterial A-site RNA, while no binding was observed to the human A-site RNA.

Recent discoveries of important roles that non-coding RNAs play in regulating gene expression make them attractive targets for molecular recognition.[1] However, most non-coding RNAs are in a *double helical* conformation and molecular recognition of such structures is a formidable problem. Designing small molecules that selectively recognize RNA using hydrophobic or electrostatic interactions has been a challenging and involved process.[2] Herein, we show that RNA can be recognized with high affinity and sequence selectivity using nucleobase-modified PNA that forms stable triple helix in the major groove of RNA duplex.

Hydrogen bonding is an inherently effective means of selective recognition of nucleic acids. Dervan and co-workers pioneered this concept in DNA recognition using minor groove-binding polyamides[3] and major groove-binding oligonucleotides.[4] Compared to DNA, the minor groove of RNA is wide and shallow and less suited for molecular recognition. The major groove of RNA is deep and narrow, which complicates triple helical binding. RNA triple helices are little studied. Relatively stable RNA triple helices are formed *via* parallel binding of a pyrimidine oligoribonucleotide to a purine tract of the double helical RNA.[5] The sequence selectivity is achieved through Hoogsteen hydrogen bonding, which allows formation of U*A–U and C^+*G–C triplets (Fig. 1).

Practical applications of triple helical recognition of nucleic acids are limited by (1) low affinity of the third strand oligonucleotide for the double helix caused, at least in part, by electrostatic repulsion between the negatively charged phosphate backbones and (2) the *requirement for long homopurine tracts*, as only U*A–U and C^+*G–C triplets are used in the common triple helical recognition. Moreover, the requirement for cytosine (pK_a = 4.5) protonation greatly decreases the stability of C^+*G–C at physiological pH. We recently reported[6] that the problem of low affinity could be overcome by using peptide nucleic acid (PNA), a neutral oligonucleotide analogue that does not suffer from the electrostatic repulsion. Although binding of PNA to double helical DNA has been extensively explored,[7] triple helix formation between PNA and double helical RNA was unknown before our study.[6] This discovery inspired a hypothesis that PNA may serve as a general ligand for sequence selective recognition of biologically relevant RNA helices.

Despite significant research activity,[8] the requirement for long purine tracts remains a major limitation of triple helical recognition. Biologically relevant double-helical RNAs typically do not contain long polypurine stretches. However, it is common to find seven or more contiguous purines interrupted by one or two pyrimidines in ribosomal RNAs[9] and microRNAs.[10] Thus, if the sequence range of triple helical recognition could be expanded to recognize isolated pyrimidines, the approach could be rendered useful for fundamental studies in RNA biology as well as practical biomedical applications.

We recently found that 5-methylisocytidine when installed in PNA recognized the C–G inversion in polypurine tract of RNA with slightly higher affinity than the natural nucleobases, though the sequence selectivity of recognition was low.[11] Herein we show that 2-pyrimidinone (**P**) and 3-oxo-2,3-dihydropyridazine (**E**) recognized C–G and U–A inversions (Fig. 1) with high affinity and sequence selectivity at pH 6.25. The use of **P** base in PNA had no precedent before our study, although derivatives of pyrimidinone have been used as oligonucleotide modifications for recognition of C–G inversions in polypurine

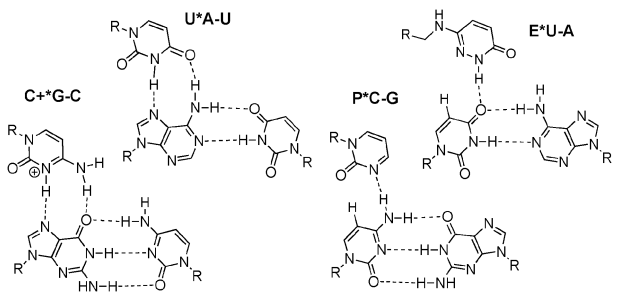
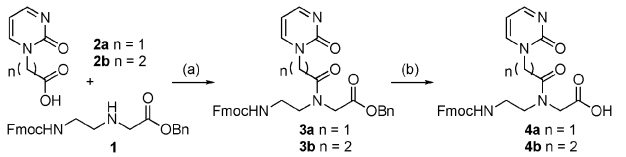


Fig. 1 Hoogsteen triplets recognition of purines and pyrimidines.

Department of Chemistry, Binghamton University, The State University of New York, Binghamton, New York 13902, USA. E-mail: erozners@binghamton.edu; Tel: (+) 1-607-777-2441

† Electronic supplementary information (ESI) available: Details of PNA synthesis, ITC experiments and data, UV melting data and copies of NMR spectra. See DOI: 10.1039/c1cc14706d

‡ These authors contributed equally to this work.



Scheme 1 Synthesis of **P** and $\mathbf{P_{ex}}$ PNA monomers. Steps: (a) EDC, DhbtOH, DMF, 40 °C, 12 h, **3a** 67%, **3b** 65%; (b) Pd/C, H_2, MeOH or MeOH:CH_2Cl_2 (1 : 1), rt, 3h, **4a** 91%, **4b** 84%.

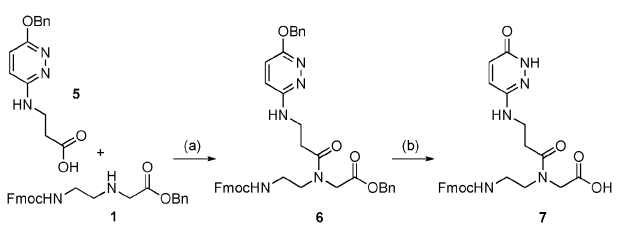
tracts of DNA.[12] The **E** base was originally designed as a PNA modification by Nielsen and co-workers,[13] who used it to recognize T–A inversions in DNA. Apart from our preliminary study using 5-methylisocytidine,[11] recognition of pyrimidine inversions in polypurine tracts of *double helical RNA* had no previous precedents.

Our study started with synthesis of Fmoc-protected **P**, $\mathbf{P_{ex}}$ and **E** PNA monomers **4a**, **4b** and **7**, respectively (Scheme 1 and 2). *N*-(3-dimethylaminopropyl)-*N*′-ethylcarbodiimide (EDC) and 3-hydroxy-1,2,3-benzotriazine-4(3H)-one (DhbtOH) mediated coupling of the commercially available (2-oxo-1(2H)-pyrimidinyl)-acetic acid **2a** and PNA backbone **1** followed by hydrogenation gave **P** PNA monomer **4a**.

Analogous *N*,*N*′-dicyclohexylcarbodiimide (DCC) mediated coupling of **1** with the known carboxylic acid **5**[13] followed by hydrogenation gave **E** PNA monomer **7** (Scheme 2). Inspired by the design of the **E** base, which features longer linker between the backbone of PNA and the heterocycle, we also prepared an extended variant of **P**, $\mathbf{P_{ex}}$ PNA monomer **4b** (Scheme 1). Our hypothesis was that the one carbon longer linker in **4b** would relax the conformation of PNA's backbone allowing better positioning of the **P** base for recognition of the C-G inversion.

Monomers **4a**, **4b** and **7** were used in a standard PNA synthesis protocol on an Expedite 8909 DNA synthesizer to prepare modified PNA nonamers **PNA3**, **PNA4** and **PNA5**, respectively (Fig. 2). The target RNA hairpins **HRP1-HRP4** (containing a variable base pair) were chosen from our previous studies.[6,11] We expected that the modified PNAs would form the following matched triplexes: the **P**-modified **PNA3** and **PNA4** with **HRP3** and the **E**-modified **PNA5** with **HRP4**. The binding affinity and sequence selectivity were studied using isothermal titration calorimetry (ITC), as previously described by us.[6]

ITC results (Table 1, Fig. S1–19†) showed that at pH 5.5 the unmodified **PNA1** and **PNA2** had excellent binding affinity ($K_a > 8 \times 10^8$) and good selectivity for the matched hairpins **HRP1** and **HRP2**, respectively. Replacement of the variable base in PNA with the **P** base (monomer **4a**) led to an overall decrease in binding affinity for **PNA3**, while good binding selectivity was maintained (Table 1, entry 3). In support of our



Scheme 2 Synthesis of **E** PNA monomer. Steps: (a) DhbtOH, DCC, DMF, 0 °C, 1h then rt, overnight, 89%; (b) Pd/C, H_2, ethanol, rt, 3h, 73%.

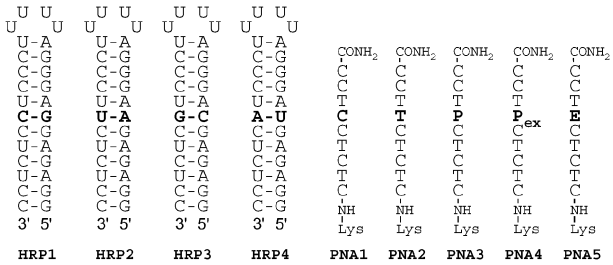


Fig. 2 Sequences of RNA hairpins and PNA ligands.

hypothesis that an extended linker would improve binding affinity, **PNA4** featuring the $\mathbf{P_{ex}}$ base (monomer **4b**) had significantly higher affinity than **PNA3** for the matched target **HRP3** (*cf.*, Table 1, entries 3 and 4). However, the higher affinity was accompanied by somewhat lower sequence selectivity of **PNA4**. Surprisingly, **PNA5** featuring the **E** base had high binding affinity to all hairpins tested without any sequence selectivity. This result was in contrast to the original report[13] that **E** base demonstrated high selectivity for T and U over other nucleobases in DNA at pH 7. The discrepancy prompted us to investigate the binding of modified PNAs at higher and more physiologically relevant pH.

Initial ITC experiments did not detect binding of the unmodified **PNA2** to **HRP2** at pH 7 ($K_a < 10^3$). However, good affinity was observed at pH 6.25 (Table 2, Fig. S20–35†), which has biological relevance being in the range of cytoplasmic pH of acidophilic bacteria.[14] Consistent with unfavourable cytosine ($pK_a = 4.5$) protonation, changing the pH from 5.5 to 6.25 reduced the affinity of PNAs, while the sequence selectivity was significantly improved (*cf.*, Tables 1 and 2).

The increase of pH restored the expected binding profile for **PNA5**, which now had the highest affinity for the matched **HRP4** (bold in entry 4, Table 2). It is conceivable that lower pH may have favored protonation or alternative tautomeric forms of **E**, which may explain the high affinity and low selectivity of **PNA5** at pH 5.5. Most importantly, at pH 6.25 the affinity of PNAs targeting single isolated pyrimidine interruption was similar to the affinity of unmodified PNAs targeting all-purine strands (see bold numbers in Table 2). As observed in our previous study,[6] the binding stoichiometry (Table S1) was consistent with a 1 : 1 PNA-RNA triple helix. The **P** and **E** bases showed excellent sequence selectivity for the target C–G and U–A base pairs, respectively.

The binding of PNAs to RNA hairpins was further confirmed by the expected biphasic UV melting curves typical for triple helices (Fig. 3, for full set of data, see Fig. S37–40†). The low temperature transition was assigned to the dissociation of

Table 1 Binding of PNA to RNA hairpins at pH 5.5[a]

Entry	PNA (variable base)	HRP1 (**G**–C)	HRP2 (**A**–U)	HRP3 (**C**–G)	HRP4 (**U**–A)
1	**PNA1** (C)	**890**	96	74	3
2	**PNA2** (T)	72	**820**	62	8
3	**PNA3** (P)	9	3	**46**	3
4	**PNA4** (P_{ex})	47	172	**293**	6
5	**PNA5** (E)	270	210	ND[b]	**38**

[a] Association constants $K_a \times 10^6$ M^{-1} in sodium acetate buffer. [b] Not determined.

Table 2 Binding of PNA to RNA hairpins at pH 6.25[a]

Entry	PNA (variable base)	HRP1 (**G**–C)	HRP2 (**A**–U)	HRP3 (**C**–G)	HRP4 (**U**–A)
1	**PNA1** (C)	**8.1**	LB[b]	LB[b]	LB[b]
2	**PNA2** (T)	7.5	**20.0**	LB[b]	0.9
3	**PNA4** (P_{ex})	LB[b]	LB[b]	**4.4**	LB[b]
4	**PNA5** (E)	3.0	0.5	LB[b]	**28.0**

[a] Association constants $K_a \times 10^6$ M^{-1} in sodium acetate buffer. [b] K_a estimated < 0.01; the low binding prevented accurate curve fit.

the PNA from the triple helical structure. The high temperature transition was assigned to the melting of the hairpin to single strand. The triple helix melting for the matched **PNA4-HRP3** (t_m = 43 °C, solid line in Fig. 3) was shifted 14 °C higher than the transitions of the mismatched complexes (t_m = 19 to 29 °C). Overall, the melting data were consistent with triple helix formation and confirmed our ITC results.

The encouraging results obtained in our model system (Fig. 2) prompted us to check if nucleobase-modified PNAs could recognize purine-rich strands in biologically important RNAs. Intriguingly, the sequence of ribosomal A-site conserved among several pathogenic bacteria, such as *E. coli*, *P. aeruginosa* and *S. aureus*, features a stretch of eight purines (Fig. 4, bold in **HRP5**) interrupted by single uridine.[9] ITC experiments (Fig. S36†) showed that the **E**-modified **PNA6** recognized the bacterial A-site with affinity similar to that observed in our model hairpins. The binding stoichiometry was close to 1 : 1, as expected for the triple helix. In contrast, we observed no binding of **PNA6** to **HRP6**, which features the sequence of the human ribosomal A-site.

Remarkably, the non-canonical A*G and A*A base pairs and the looped out adenosine did not significantly lower the stability of the PNA-RNA complex. We expected that of **PNA6** would show excellent selectivity for **HRP5** over **HRP6** because the purine-rich strands of human and bacterial A-site sequences had only four out of nine nucleosides common (bold in **HRP6**). This is in contrast to the A-rich loop, the target of aminoglycoside antibiotics, which is remarkably similar for different organisms.

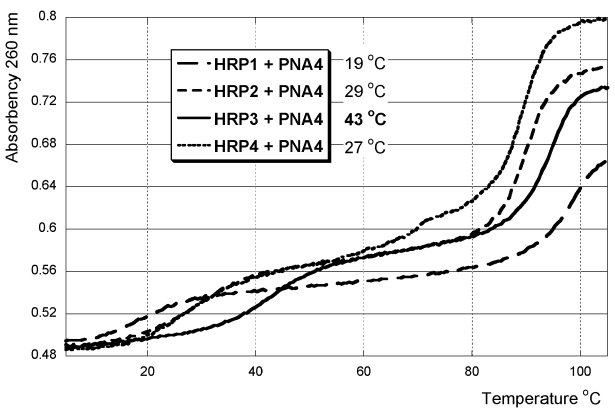


Fig. 3 UV thermal melting curves of $\mathbf{P_{ex}}$-modified **PNA4** bound to **HRP1-HRP4**. Solid line is the matched **PNA4-HRP3** complex.

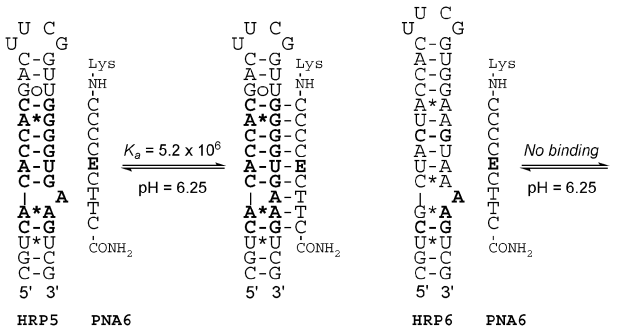


Fig. 4 Binding of **E**-modified **PNA6** to models of bacterial (**HRP5**) and human (**HRP6**) A-site RNAs.

In summary, PNA nonamers modified with **P** and **E** nucleobases recognized single isolated pyrimidine interruptions in polypurine tracts of double helical RNA with similar affinity and sequence selectivity than unmodified PNAs binding to all-purine strands of RNA at pH 6.25. Preliminary results suggested that the approach could be further developed to recognize complex biologically relevant RNAs featuring bulges and non-canonical base pairs, such as pre-microRNAs and ribosomal RNAs.[9,10] Further development of more basic cytosine analogues should increase the binding affinity at physiological conditions. Although recognition of pre-microRNA is a relatively new area of research, promising results have already been reported that binding of helix-threading peptides inhibit maturation of pre-microRNA.[15]

We thank Binghamton University and NIH (R01 GM071461) for financial support of this research.

Notes and references

1 G. Meister and T. Tuschl, *Nature*, 2004, **431**, 343; R. H. A. Plasterk, *Cell*, 2006, **124**, 877.
2 J. R. Thomas and P. J. Hergenrother, *Chem. Rev.*, 2008, **108**, 1171.
3 P. B. Dervan and B. S. Edelson, *Curr. Opin. Struct. Biol.*, 2003, **13**, 284.
4 H. E. Moser and P. B. Dervan, *Science*, 1987, **238**, 645.
5 R. W. Roberts and D. M. Crothers, *Science*, 1992, **258**, 1463; H. Han and P. B. Dervan, *Proc. Natl. Acad. Sci. U. S. A.*, 1993, **90**, 3806; C. Escude, J. C. Francois, J. S. Sun, G. Ott, M. Sprinzl, T. Garestier and C. Helene, *Nucleic Acids Res.*, 1993, **21**, 5547.
6 M. Li, T. Zengeya and E. Rozners, *J. Am. Chem. Soc.*, 2010, **132**, 8676.
7 M. E. Hansen, T. Bentin and P. E. Nielsen, *Nucleic Acids Res.*, 2009, **37**, 4498; P. E. Nielsen, *Curr. Opin. Mol. Ther.*, 2010, **12**, 184.
8 K. R. Fox and T. Brown, *Q. Rev. Biophys.*, 2006, **38**, 311.
9 http://www.rna.ccbb.utexas.edu/.
10 http:// www.mirbase.org/; S. Griffiths-Jones, H. K. Saini, S. van Dongen and A. J. Enright, *Nucleic Acids Res.*, 2007, **36**, D154.
11 T. Zengeya, M. Li and E. Rozners, *Bioorg. Med. Chem. Lett.*, 2011, **21**, 2121.
12 D. A. Gianolio and L. W. McLaughlin, *J. Am. Chem. Soc.*, 1999, **121**, 6334; S. Buchini and C. J. Leumann, *Angew. Chem., Int. Ed.*, 2004, **43**, 3925; R. T. Ranasinghe, D. A. Rusling, V. E. C. Powers, K. R. Fox and T. Brown, *Chem. Commun.*, 2005, 2555.
13 A. B. Eldrup, O. Dahl and P. E. Nielsen, *J. Am. Chem. Soc.*, 1997, **119**, 11116.
14 J. L. Slonczewki, M. Fujisawa, M. Dopson and T. A. Krulwich, *Adv. Microb. Physiol.*, 2009, **55**, 1.
15 M. Krishnamurthy, K. Simon, A. M. Orendt and P. A. Beal, *Angew. Chem., Int. Ed.*, 2007, **46**, 7044; B. D. Gooch and P. A. Beal, *J. Am. Chem. Soc.*, 2004, **126**, 10603.