Contents lists available at ScienceDirect



Bioorganic & Medicinal Chemistry Letters

journal homepage: www.elsevier.com/locate/bmcl

PNA containing isocytidine nucleobase: Synthesis and recognition of double helical RNA

Thomas Zengeya, Ming Li, Eriks Rozners*

Department of Chemistry, Binghamton University, The State University of New York, Binghamton, NY 13902, USA

ARTICLE INFO

Article history: Received 27 December 2010 Revised 25 January 2011 Accepted 28 January 2011 Available online 1 February 2011

Keywords: RNA recognition Peptide nucleic acid Triple helix Modified nucleobases Isothermal titration calorimetry

ABSTRACT

Peptide nucleic acid (**PNA1**) containing a 5-methylisocytidine (iC) nucleobase has been synthesized. Triple helix formation between **PNA1** and RNA hairpins having variable base pairs interacting with iC was studied using isothermal titration calorimetry. The iC nucleobase recognized the proposed target, C-G inversion in polypurine tract of RNA, with slightly higher affinity than the natural nucleobases, though the sequence selectivity of recognition was low. Compared to non-modified PNA, **PNA1** had lower affinity for its RNA target.

© 2011 Elsevier Ltd. All rights reserved.

Compared to DNA, molecular recognition of double helical RNA has received little attention. Until relatively recently, RNA was viewed mostly as a passive messenger in the transfer of genetic information from DNA to proteins. However, recent discoveries of the central role that various non-coding RNAs play in gene expression have revitalized interest in molecular recognition of double helical RNA. Although RNA is a well-established target of current antibiotics, designing of new compounds that selectively recognize RNA has been a challenging task.¹

Hydrogen bond mediated base pairing is the key feature of helical nucleic acids and would be inherently the most effective way of sequence selective recognition of RNA. Recognition of double helical DNA via triple helix forming oligonucleotides² and polyamides³ that form extensive hydrogen bond contacts in the major and minor grooves of DNA, respectively, is well established. Compared to DNA, the wide and shallow minor groove of RNA double helix is less suited for molecular recognition. Surprisingly and in contrast to DNA, sequence selective recognition of double helical RNA using major groove triple helix is little studied.⁴ The major groove of RNA is deep and narrow, which may hinder the formation of triple helix. While DNA does not form triple helix with RNA, modestly stable RNA triple helices are formed via parallel binding of a pyrimidine rich RNA third strand to the purine rich strand of a double helix.⁴ The sequence selectivity derives from recognition of adenosineuridine base pairs by uridine (U*A-U triplet) and guanosine-cytidine base pairs by protonated cytidine (C*G-C triplet) via the Hoogsteen hydrogen bonding scheme (Fig. 1). Recently, we discovered that peptide nucleic acid (PNA),⁵ a neutral amide-linked analogue of DNA formed highly stable and sequence selective triple helices with double stranded RNA.⁶

A major limitation of triple helical recognition is the requirement for a polypurine tract, because only A and G can be recognized by the Hoogsteen hydrogen bonds (Fig. 1). While double stranded RNA typically does not have long polypurine tracts, it is common to find stretches of eight and more purine bases interrupted by a couple of pyrimidines in micro RNAs (for structures, see www.mirbase.org) and other non-coding RNAs. Thus, if the sequence range of triple helical recognition could be expanded to recognize isolated pyrimidine inversions, the approach could be rendered useful for fundamental biochemical studies and practical applications.

While extensive work has been done on modified nucleobases for recognition of C–G and T–A inversion in polypurine tracts of DNA,² similar recognition of pyrimidines in RNA is unprecedented. Herein, we tested a hypothesis that 5-methylisocytosine (iC) would form a triplet with a C–G base pair in RNA (Fig. 1, iC*C–G). iC was designed by Benner and co-workers⁷ to engineer unnatural base pairing schemes as alternatives to Watson–Crick pairing and, before our study, had not been tried in triple helix context. We envisioned that iC would recognize cytosine by forming a O4– N4H hydrogen bond and, potentially, also engaging in an attractive N3–C5H interaction (Fig. 1). This is similar to the N3–N4H and O2– C5H hydrogen bonding scheme used for recognition of C–G in DNA by 5-methyl-2-pyrimidinone⁸ and derivatives.⁹ Other designs of C– G recognition in DNA have included interactions that reach across

^{*} Corresponding author. Tel.: +1 607 777 2441; fax: +1 607 777 4478. *E-mail address:* erozners@binghamton.edu (E. Rozners).

⁰⁹⁶⁰⁻⁸⁹⁴X/\$ - see front matter \odot 2011 Elsevier Ltd. All rights reserved. doi:10.1016/j.bmcl.2011.01.130



Figure 1. Structures of Hoogsteen U*A-U and C*G-C and the proposed iC*C-G base triplets and PNA (PNA1).

the groove to form hydrogen bonds with G, cationic base substituents and sugar modifications to stabilize the non-Hoogsteen triplets.¹⁰

Inspired by our own recent results⁶ on triple helical recognition of RNA using PNA, we set out to test if iC installed on a PNA backbone (Fig. 1, **PNA1**) could recognize the C–G inversion in polypurine tract of RNA. While Nielsen and co-workers¹¹ have designed PNA containing pyridazinone nucleobase for recognition of T–A and U–A inversions in DNA, to the best of our knowledge, recognition of C–G inversion by modified PNA had no precedent at the outset of our studies.¹² We found that the iC nucleobase 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 relatively low.

For synthesis of iC PNA monomer we first tried to adopt the standard PNA chemistry (Scheme 1).¹³ 5-Methylisocytosine¹⁴ was protected with Boc group and alkylated to produce ester **2**. The correct N1 position of alkylation was confirmed by observation of NOE between H6 and CH₂ of the acetate linkage (Scheme 1). The ethyl ester was cleaved under basic conditions to produce carboxylic acid **3**.

To our surprise, coupling of carboxylic acid **3** with amine **4**, following standard procedures,¹³ resulted in a transfer of the Boc group on the central amine and gave compound **5** instead of the expected PNA monomer. A potential explanation for this unexpected result could be that the activation of carboxylic acid **3** led to an intramolecular cyclization and formation of imide **6**, which then could transfer the Boc group to amine **4**. Because our goal was to prepare the iC PNA monomer, we redirected our efforts to alternative syntheses and the intermediacy of imide **6** remained a plausible but unconfirmed hypothesis.

We envisioned that late stage alkylation of 5-methylisocytidine with chloroamide **7** (Scheme 2), following a strategy similar to that used by Meltzer and co-workers,¹⁵ might provide an alternative route to the desired iC monomer. Chloroamide **7** was obtained by coupling of amine **4** with chloroacetic acid in the presence of dicy-clohexylcarbodiimide (DCC).

Although alkylation of **1** with chloroamide **7** gave the expected product in a relatively low yield, we were able to produce enough material to proceed with the PNA synthesis and binding studies. As previously reported,¹³ cleavage of benzyl ester proceeded uneventfully and gave the required Fmoc-PNA-iC(Boc)-OH **9**, which was used in a standard PNA synthesis protocol on an Expedite 8909 DNA synthesizer to prepare **PNA1** containing the 5-methylisocytidine nucleobase (see Figs. 1 and 2).

The particular sequence of PNA1 allowed convenient testing of the effect of iC via direct comparison with our previous work that tested the binding of PNA3-PNA6 (containing a variable nucleobase highlighted in bold in Fig. 2) to HRP1-HRP4 (containing a variable base pair).⁶ The binding affinity and sequence selectivity of PNA1 was studied using isothermal titration calorimetry (ITC), as previously described by us.⁶ ITC directly measures enthalpy of binding and is one of the best methods to study RNA-ligand interactions.¹⁶ ITC has been used to characterize the thermodynamics of PNA binding to DNA and formation of modified DNA triple helices.¹⁷ Our recent study confirmed that ITC was an excellent method to study binding of PNA to double helical RNA.⁶ In our experimental system, iC in PNA1 was expected to give a matched triplet with the C–G base pair of HRP3 (see Fig. 1). ITC experiments (Fig. 3) showed that the affinity of PNA1 for HRP3 was slightly higher than affinity for HRP1, HRP2 and HRP4, which were expected to form mismatched triplets with iC (Table 1).



Scheme 1. Attempted synthesis of the iC PNA monomer using standard strategy.



Scheme 2. Synthesis of the iC PNA monomer using late stage alkylation strategy.



Figure 2. Sequences of RNA hairpins and PNA ligands. Numbering of HRP1-HRP4 and PNA3-PNA6 is retained from Ref. 6.

Analysis of the thermodynamic data in Table 1 revealed that the binding was driven by favorable entropy that more than compensated for the less favorable enthalpy contribution. The binding order (stoichiometry) indicated that **PNA1** formed the expected 1:1 complex with **HRP3**, a result that is consistent with the proposed triple helical recognition. We do not have an explanation for why the binding order of the other mismatched combinations was 0.5

or less. However, we have observed similar phenomenon with mismatched complexes in our previous study.⁶

The binding of **PNA1** to **HRP3** was confirmed using UV spectroscopy. The melting curve (Fig. 4) showed two well-resolved transitions. The lower temperature (around 30 °C) transition was characteristic for melting of triplex to hairpin and single strand PNA, while the higher temperature (around 90 °C) transition was due to hairpin melting.⁶ Analogous experiments using the mismatched complexes (**HRP1** and **HRP2**) did not produce clear melting curves. The complex formation was further confirmed by circular dichroism spectroscopy (Fig. S1). Overall, the results of UV melting and circular dichroism experiments were consistent with our previous observations⁶ and indicated triple helix formation.

Comparison of **PNA1** with PNAs featuring natural nucleobases at the variable position (**PNA3-PNA6**) is shown in **Table 2**. Affinity of **PNA1** for the matched **HRP3** was decreased approximately 50 times compared to **PNA3** and **PNA4** binding to their complementary hairpins (**HRP1** and **HRP2**) that featured the matched Hoogsteen triplets (highlighted in bold in Table 2). Meanwhile, the affinities of mismatched combinations involving **PNA1** were similar to that of other mismatches previously studied.⁶ Thus, the



Figure 3. ITC titration curve of PNA1 binding to HRP3.

Table 1								
Thermodynamic	data	for	binding	of PNA1	to	RNA	hair	pins

Entry	RNA	Ka	$-\Delta H$ (kcal/mol)	$-\Delta S$ (cal/mol K)	$-\Delta G$ (kcal/mol)	order
1	HRP1	0.51	42.3	116	7.8	0.5
2	HRP2	0.82	27.2	64	8.1	0.4
3	HRP3	1.01	15.5	25	8.2	1.1
4	HRP4	0.77	28.7	69	8.0	0.3

^a Association constants $K_a \times 10^6 \text{ M}^{-1}$ in 100 mM sodium acetate, 1.0 mM EDTA, pH 5.5.



Figure 4. UV thermal melting curve of HRP3 (5.25 $\mu M)$ and PNA1 (18 $\mu M)$ in 100 mM sodium acetate, 1.0 mM EDTA, pH 5.5.

Table 2

Sequence specificity of PNA binding to RNA hairpins^a

PNA (variable base)	HRP1 ^a (G –C)	HRP2 ^a (A –U)	HRP3 ^a (C -G)	HRP4 ^a (U-A)
PNA1 (iC)	0.5	0.8	1.0	0.8
PNA3 $(C)^{b}$	84.0	0.4	0.5	0.2
PNA4 $(T)^b$	2.7	47.0	0.5	0.02
PNA5 (G) ^b	1.5	0.4	0.2	0.09
PNA6 (A) ^b	6.0	1.6	0.7	0.05

 a Association constants $K_a \times 10^6 \, M^{-1}$ in sodium acetate buffer, pH 5.5. b Data from our previous study, Ref. 6.

selectivity of **PNA1** for **HRP3** was marginal, which is most likely result of a less than ideal alignment of iC in the iC*C–G triplet. However, among all nucleobases studied for the recognition of the C–G inversions in RNA, iC gave the highest binding affinity (see the **HRP3** column in Table 2), though the improvement is small.

In summary, we have synthesized the first PNA analogue bearing 5-methylisocytidine nucleobase. While iC is currently the best nucleobase to recognize C–G inversions in the polypurine tract of an RNA double helix, the binding affinity and sequence selectivity remain modest. In fact, iC is also the best base for recognition of U– A inversions (Table 2, last column) and may serve as universal base to be used against pyrimidine inversions instead of the natural nucleobases. Decreased binding affinity and sequence selectivity, similar to our results, has also been observed in recognition of C– G and T–A inversions in polypurine tracts of DNA, and illustrates the challenge of Hoogsteen recognition of pyrimidine nucleobases having only one hydrogen bonding site.^{2,8–11} To the best of our knowledge, our study is the first attempt at overcoming the requirement for a polypurine tract in RNA triple helices using modified PNA. Work is in progress in our lab to test other modified nucleobases that are expected to lead to further improvements in recognition of C–G inversions in RNA double helices.

Acknowledgment

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

Supplementary data

Supplementary data (experimental procedures, copies of CD and NMR data, details PNA synthesis and ITC experiments and data) associated with this article can be found, in the online version, at doi:10.1016/j.bmcl.2011.01.130.

References and notes

- (a) Thomas, J. R.; Hergenrother, P. J. Chem. Rev. 2008, 108, 1171; (b) Sucheck, S. J.; Wong, C. H. Curr. Opin. Chem. Biol. 2000, 4, 678; (c) Chow, C. S.; Bogdan, F. M. Chem. Rev. 1997, 97, 1489.
- 2. Fox, K. R.; Brown, T. Q. Rev. Biophys. 2005, 38, 311.
- 3. Dervan, P. B.; Edelson, B. S. Curr. Opin. Struct. Biol. 2003, 13, 284.
- (a) Roberts, R. W.; Crothers, D. M. Science **1992**, 258, 1463; (b) Han, H.; Dervan, P. B. Proc. Natl. Acad. Sci. U.S.A. **1993**, 90, 3806; (c) Escude, C.; Francois, J. C.; Sun, J. S.; Ott, G.; Sprinzl, M.; Garestier, T.; Helene, C. Nucleic Acids Res. **1993**, 21, 5547.
- 5. Nielsen, P. E.; Egholm, M.; Berg, R. H.; Buchardt, O. Science 1991, 254, 1497.
- 6. Li, M.; Zengeya, T.; Rozners, E. J. Am. Chem. Soc. 2010, 132, 8676.
- 7. Benner, S. A. Acc. Chem. Res. 2004, 37, 784.
- (a) Buchini, S.; Leumann, C. J. Angew. Chem., Int. Ed. 2004, 43, 3925; (b) Prevot-Halter, I.; Leumann, C. J. Bioorg. Med. Chem. Lett. 1999, 9, 2657.
- (a) Ranasinghe, R. T.; Rusling, D. A.; Powers, V. E. C.; Fox, K. R.; Brown, T. Chem. Commun. 2005, 2555; (b) Rusling, D. A.; Powers, V. E. C.; Ranasinghe, R. T.; Wang, Y.; Osborne, S. D.; Brown, T.; Fox, K. R. Nucleic Acids Res. 2005, 33, 3025.
- (a) Semenyuk, A.; Darian, E.; Liu, J.; Majumdar, A.; Cuenoud, B.; Miller, P. S.; MacKerell, A. D., Jr.; Seidman, M. M. *Biochemistry* **2010**, *49*, 7867; (b) Wachowius, F.; Rettig, M.; Palm, G.; Weisz, K. *Tetrahedron Lett.* **2008**, *49*, 7264.
- 11. Eldrup, A. B.; Dahl, O.; Nielsen, P. E. J. Am. Chem. Soc. 1997, 119, 11116.
- 12. Wojciechowski, F.; Hudson, R. H. E. Curr. Top. Med. Chem. 2007, 7, 667.
- 13. Wojciechowski, F.; Hudson, R. H. E. J. Org. Chem. 2008, 73, 3807.
- 14. Stoss, P.; Kaes, E.; Eibel, G.; Thewalt, U. J. Heterocycl. Chem. 1991, 28, 231.
- 15. Meltzer, P. C.; Liang, A. Y.; Matsudaira, P. J. Org. Chem. 1995, 60, 4305.
- Salim, N. N.; Feig, A. L. Methods 2009, 47, 198; Feig, A. L. Methods Enzymol. 2009, 468, 409.
- Ratilainen, T.; Holmen, A.; Tuite, E.; Nielsen, P. E.; Norden, B. *Biochemistry* 2000, 39, 7781; Chin, T.-M.; Tseng, M.-H.; Chung, K.-Y.; Hung, F.-S.; Lin, S.-B.; Kan, L.-S. *J. Biomol. Struct. Dyn.* 2001, 19, 543.