Triple-Helical Recognition of RNA Using 2-Aminopyridine-Modified PNA at Physiologically Relevant Conditions**

Thomas Zengeya, Pankaj Gupta, and Eriks Rozners

Compared to DNA, molecular recognition of double-stranded RNA has received relatively little attention. Until the early 1990s, RNA was viewed as a passive messenger in the transfer of genetic information from DNA to proteins. However, since the discovery that RNA can catalyze chemical reactions, the number and variety of non-coding RNAs and the important roles they play in biology have been growing steadily.[1] Currently, the functional importance of most RNA transcripts is still unknown, and it is fairly safe to predict that we will discover many more regulatory RNAs in the near future. The ability to selectively recognize and control the function of such RNAs will be highly useful for both fundamental research and practical applications. However, recognition of double-helical RNA by sequence-selective ligand binding is a formidable challenge.[2,3]

Recently, we proposed that biologically relevant double-helical RNAs could be recognized by major-groove triple-helix formation using peptide nucleic acid (PNA).[4–6] We showed that PNAs as short as hexamers formed stable and sequence selective Hoogsteen triple helices with RNA duplexes ($K_a > 10^7$ M^{-1}) at pH 5.5.[4] A limitation of triple helical recognition was the requirement for long homopurine tracts, as only the Hoogsteen T(U)*A-T(U) and C+*G-C triplets could be used (Figure 1). We found that modification of PNA with 2-pyrimidinone[7] and 3-oxo-2,3-dihydropyrida-zine (**E**)[8] nucleobases allowed efficient and selective recognition of isolated C-G and U-A inversions, respectively, in polypurine tracts of double helical RNA at pH 6.25.[5,6] However, the high affinity of PNA at pH 5.5 was greatly reduced at pH 6.25, and no binding could be observed at physiologically relevant salt and pH 7.4.[5] The remaining problem was the unfavorable protonation of cytosine, which was required for formation of the Hoogsteen C+*G-C triplets (Figure 1). Because its $pK_a = 4.5$, cytosine is hardly protonated under physiological pH, which greatly decreases the stability of the triple helix. Herein we provide an efficient solution to this problem and demonstrate that sequence
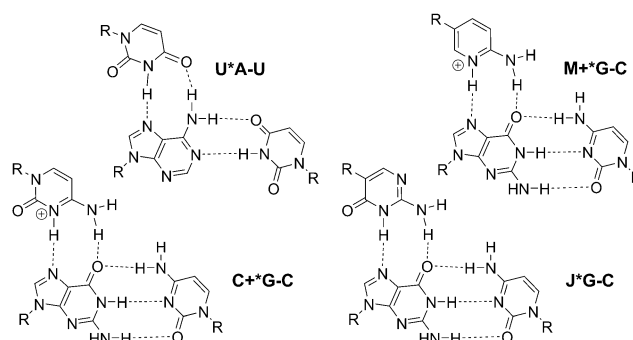

Figure 1. Standard and modified Hoogsteen triplets.

selective recognition of the RNA duplex can be achieved at physiologically relevant conditions by replacing cytosine with a more basic ($pK_a = 6.7$[9]) heterocycle, 2-aminopyridine (abbreviated as **M** in Figure 1).

Povsic and Dervan pioneered the chemical modulation of the cytosine pK_a by showing that triple helices containing 5-methylcytosine were more stable at higher pH than those of unmodified DNA.[10] More recently, derivatives of 2-amino-pyridine have been used to increase the stability of DNA triple helices at high pH.[11–14] Before our studies, this approach had not been used in PNA.

An alternative approach has used neutral nucleobases that mimic the hydrogen-bonding pattern of protonated cytosine. The most notable examples are pseudoisocytosine (abbreviated as **J** in Figure 1) by Kan and co-workers,[15] methyloxocytosine by McLaughlin and co-workers,[16,17] and a pyrazine derivative by von Krosigk and Benner.[18] The **J** base is widely used in PNA to alleviate the pH dependency of PNA-DNA triplexes.[19,20] The use of modified heterocycles to recognize double stranded RNA at physiologically relevant conditions had no precedent before our studies.

We started by comparing the binding of unmodified PNA1, **J**-modified PNA2 and **M**-modified PNA3 to HRPA (Figure 2), an A-rich RNA hairpin similar to the model systems used previously by us.[4,5] Fmoc-protected **J** monomer **1** was synthesized according to the literature procedure.[21] The **M** monomer **2** was synthesized from Fmoc-protected PNA backbone **3** and the known carboxylic acid **4** using DCC mediated coupling followed by deprotection of the allyl group as previously described by us (Scheme 1).[5] All of the PNAs were synthesized using a standard PNA procedure on an Expedite 8909 DNA synthesizer, purified by HPLC, and characterized by mass spectrometry as previously reported.[4–6]

Following the same approach as in our previous studies,[4–6] we used isothermal titration calorimetry (ITC) and UV

[*] T. Zengeya, Dr. P. Gupta, Dr. E. Rozners
Department of Chemistry, Binghamton University
The State University of New York
Binghamton, NY 13902 (USA)
E-mail: erozners@binghamton.edu
Homepage: http://www2.binghamton.edu/chemistry/people/roz-ners/rozners.html

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/anie.201207925.
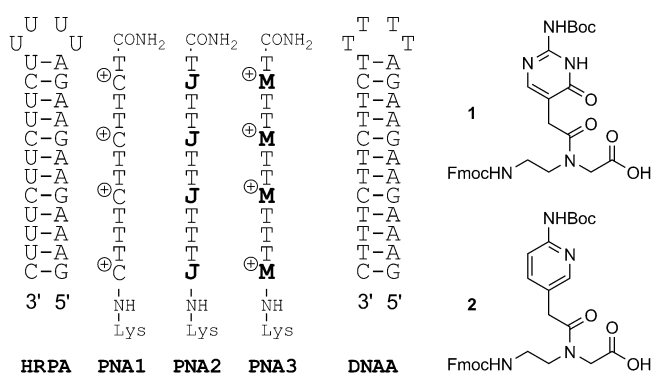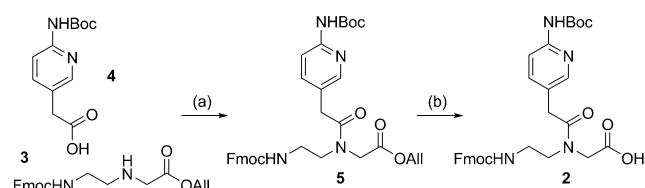
Figure 2. RNA, PNA, and DNA sequences used to compare C, J, and M nucleobases.

Scheme 1. Synthesis of **M** PNA monomer: a) DCC, 3-hydroxy-1,2,3-benzotriazin-4(3H)-one, DMF, RT, overnight, 57%; b) [Pd(PPh₃)₄], N-ethylaniline, THF, RT, 2 h, 79%.

thermal melting to characterize the binding of PNA to RNA hairpins. ITC directly measures the enthalpy of binding and, by fitting the binding data, provides binding affinity (association constant K_a in M^{-1}) and stoichiometry (the ratio of PNA to RNA in the final complex).[22] Owing to its operational simplicity, reliability, and rich thermodynamic data, ITC is one of the best methods to study ligand binding to RNA. The unmodified PNA1 formed a stable triplex with HRPA at pH 5.5 (Table 1) in sodium acetate buffer at 25 °C. As

Table 1: Binding of C-, J-, and M-containing PNA to RNA HRPA.[a]

PNA	Acetate pH 5.5[b]	Acetate pH 7.0[b]	Phosphate pH 7.4[c]
PNA1 (**C**)	0.76	0.06	NB
PNA2 (**J**)	–	0.41	0.17
PNA3 (**M**)	–	36.5	1.8

[a] Association constants $K_a \times 10^7$ M^{-1}, NB = no binding, $K_a < 10^3$.
[b] 100 mM Sodium acetate buffer at 25 °C. [c] 2 mM MgCl₂, 90 mM KCl, 10 mM NaCl, 50 mM potassium phosphate at 37 °C.

expected, because of the unfavorable protonation of cytosine at higher pH, the affinity decreased significantly when the pH of the buffer was increased to 7 and we could not observe any binding in phosphate buffer mimicking the physiological conditions at 37 °C.

We used the affinity of PNA1 at pH 5.5 as a benchmark to gauge the effect of **J** and **M** modifications on PNA affinity at higher pH. The affinity of **J**-modified PNA2 for HRPA in acetate buffer at pH 7 was lower than the affinity of PNA1 at pH 5.5 and decreased even more under the more demanding physiological conditions (Table 1). Nielsen and co-workers[20]

reported that the affinity of an unmodified PNA 15 mer (having 5 isolated cytosines) for a DNA duplex dropped by three orders of magnitude (K_d changed from 2 nM to 2.2 mM) when changing the pH from 5.5 to 7.2. Substitution of all five cytosines by **J** base increased the affinity only about tenfold ($K_d = 0.15$ mM).[20] Thus, our result was qualitatively consistent with that reported by Nielsen, only smaller in magnitude, and suggested that the positive charge on cytosine contributed significantly to stability of the Hoogsteen triplet, presumably by electrostatic attraction to the negatively charged nucleic acid. Consequently, an ideal design for recognition of G-C pairs would include both a correct hydrogen bonding pattern and a positive charge on the heterocycle. Because unmodified PNA containing cytosine (p$K_a = 4.5$) forms a stable triple helix at pH 5.5, we hypothesized that PNA modified with 2-aminopyridine **M** (p$K_a = 6.7$) would form at least equally strong triple helices at physiological pH 7.4 (owing to a similar pH/pK_a difference).

Confirming our hypothesis, ITC showed that **M** modification strongly enhanced the binding affinity of PNA3. In acetate buffer at pH 7, **M**-modified PNA3 had about two orders of magnitude higher affinity ($K_a = 3.7 \times 10^8$) for HRPA than the **J**-modified PNA2 (Figure 2 and Table 1). Under physiologically relevant conditions, PNA3 bound to HRPA with $K_a = 1.8 \times 10^7$, which was an order of magnitude higher than the affinity of PNA2 at the same conditions and twice that of unmodified PNA1 at pH 5.5. The larger drop in affinity for PNA3 compared to PNA2 going from acetate to phosphate buffer is most likely due to screening of the electrostatic interactions (that are more important for the charged **M**) by higher salt concentration and the presence of MgCl₂ in the physiologically relevant buffer. Most remarkably, binding of PNA3 to the matched DNA hairpin (DNAA) in physiological phosphate buffer was about two orders of magnitude weaker ($K_a = 3 \times 10^5$) than binding to RNA HRPA. This result suggested that the **M**-modified PNA might have unique selectivity for triple helical recognition of RNA over DNA. For all experiments at physiologically relevant conditions, fitting of ITC titration curves gave a 1:1 PNA/RNA stoichiometry (Supporting Information, Table S1), which was consistent with triple-helix formation.

UV thermal melting experiments (Supporting Information, Figure S15) confirmed the ITC results. Consistent with our previous observations,[4] the complexes of HRPA and high affinity PNAs melted in one transition of triple helix to single strands without an intermediate duplex. In phosphate buffer at pH 7.4 adding PNA2 had little effect on the stability of HRPA: $t_m = 75$ °C for HPRA alone and 74 °C for a 1:1 complex of HRPA-PNA2. Consistent with the higher K_a observed in the ITC experiments, the thermal stability of a 1:1 complex of HRPA-PNA3 was significantly higher at 80 °C. Taken together, the results confirmed our hypothesis that the charged **M** would have an advantage over the neutral **J** for triple-helical recognition of RNA.

Next we probed the sequence specificity of **M**-modified PNA using a model system from our previous studies (Figure 3).[4,5] Table 2 shows that PNA5 (four **M** modifications) had high affinity for the matched HRP1 at physiologically relevant conditions while maintaining excellent
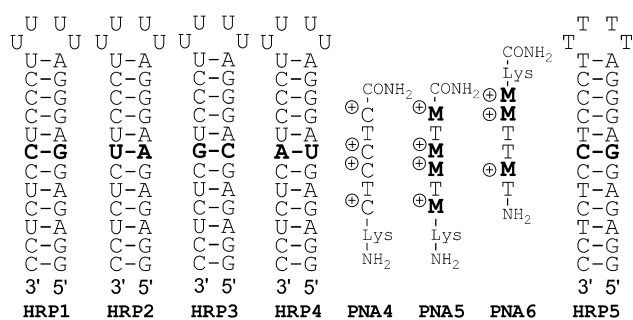
Figure 3. PNA, RNA, and DNA used in the sequence specificity study.

Table 2: Binding of M-modified PNA to RNA hairpins.[a]

PNA	HRP1 (G-C)	HRP2 (A-U)	HRP3 (C-G)	HRP4 (U-A)
PNA4[b]	**8.4**	0.04	0.05	0.02
PNA5[c]	**2.0**	<0.001[d]	NB[e]	NB[e]
PNA6[c]	NB[e]	**0.4**	NB[e]	NB[e]

[a] Association constants $K_a \times 10^7$ M^{-1}. [b] From Ref. [4], in 100 mM sodium acetate buffer, pH 5.5 at 25°C. [c] In phosphate buffer, pH 7.4 at 37°C. [d] Highest estimate; the low binding prevented more accurate curve fit. [e] NB = no binding, $K_a < 10^3$.

sequence specificity. The binding affinity of **M**-modified PNA5 at pH 7.4 was somewhat lower than that of unmodified PNA4 at pH 5.5.[4] This was in contrast to a similar comparison of **M**-modified PNA3 and unmodified PNA1 in Table 1. The discrepancy might be related to electrostatic repulsion of adjacent charged nucleobases, which has been reported to have a negative effect on affinity,[23] and may affect PNA5 more than PNA4. Nevertheless, the strong and highly selective RNA binding by PNA5 at physiologically relevant conditions was extremely encouraging. UV thermal melting (Supporting Information, Figure S25) of the matched PNA5-HRP1 complex showed a broad and relatively weak transition at about 55°C that might be assigned to triplex melting preceding the melting of the HRP1 hairpin at about 100°C. Similar transitions unique to the matched PNA5-HRP1 complex were also observed in CD melting plots (Supporting Information, Figure S26). Consistent with the high sequence selectivity, we could not observe any transitions above 30°C that could be assigned to triplex melting of the mismatched complexes. Confirming the unique RNA selectivity observed for PNA3, PNA5 showed little, if any binding to its matched DNA hairpin HRP5 (Supporting Information, Figure S24).

PNA6 (three **M** modifications) had five times lower affinity for the matched HRP2 than PNA5 for HRP1, which was consistent with a higher stability of triplets involving G–C base pairs and the notion that the positive charges are important for high binding affinity. As expected, PNA6 showed excellent sequence specificity. The PNA-RNA stoichiometry was 1:1 in all experiments shown in Table 2 (see also the Supporting Information, Table S1).

Finally, we chose microRNA-215, which is implicated in cancer development and drug resistance,[24,25] as an initial target to check if **M**-modified PNA could bind to biologically

relevant double helical RNA. MicroRNAs (miRNAs) are transcribed as long hairpin structures, pri-miRNAs, which are processed into mature miRNA duplexes (ca. 22 nt) by Drosha and Dicer endonucleases. It is common to find stretches of eight and more contiguous purines interrupted by one or two pyrimidines in pri-miRNA hairpins.[26] Triple-helical binding to such sites could be used to detect miRNAs and interfere with their function, which would find broad applications in fundamental science, medicine, and biotechnology. We chose HRP6 as a model that contains the purine-rich recognition site present in pri-miRNA-215.[26] HRP6 has a stretch of nine purines interrupted by a uridine and features several non-canonical base pairs, which are hallmarks of pri-miRNA hairpins. For recognition of the uridine interruption, we used nucleobase **E** (Figure 4) that was originally designed for thymidine recognition in DNA[8] and later adopted by us for uridine recognition in RNA.[5] PNA7 having three **M** and one **E** modification was prepared using monomers **2** and the previously reported[5] **6** (Figure 4).
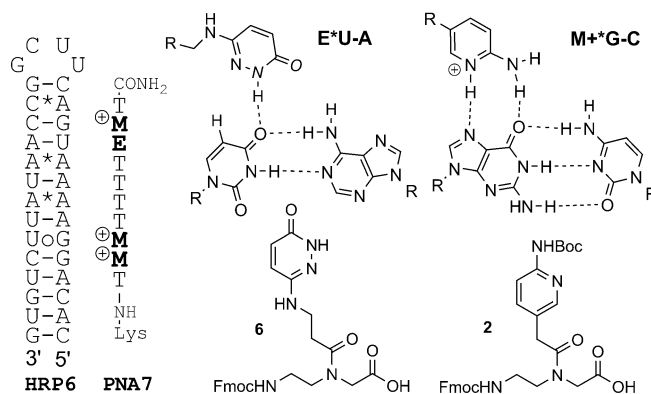
Figure 4. Binding of **E**- and **M**-modified PNA7 to HRP6 to model the pri-miRNA-215 hairpin structure.

Consistent with results obtained with other **M**-modified PNAs, PNA7 recognized HRP6 with high affinity ($K_a = 1.2 \times 10^7$) and 1:1 stoichiometry (Supporting Information, Table S1) under physiologically relevant conditions. Remarkably, the non-canonical C*A and A*A and the wobble UoG base pairs did not prevent formation of the PNA-RNA complex.

In summary, we have demonstrated that modification of PNA with 2-aminopyridine (**M**) nucleobases allows formation of stable and sequence selective triple helices with double-stranded RNA at physiologically relevant conditions. For triple-helical RNA recognition, modulation of nucleobase basicity (cf. $pK_a = 6.7$ for **M** with 4.5 for C) was a more efficient approach than using the neutral **J** base. The **M**-modified PNAs exhibited unique RNA selectivity and had two orders of magnitude higher affinity for the double stranded RNAs than for the same DNA sequences. It is conceivable that the deep and narrow major groove of RNA presented a better steric fit for the PNA ligands than the wider major groove of DNA. In preliminary experiments, nucleobase-modified PNA recognized a purine-rich model

sequence of a double-helical miRNA precursor with high affinity at physiologically relevant conditions. While this is a relatively new area of research, Beal and co-workers[27] have already demonstrated the potential of targeting pri-miRNAs using helix-threading peptides. Taken together, our results suggest that PNA may have unique and previously under-appreciated potential for triple-helical recognition of biologically relevant RNA. Low stability at pH 7.4 has been a long-standing problem for practical applications of triple helices. The excellent performance of **M**-modified PNAs at pH 7.4 observed herein provide efficient solution to this problem that should open the door for new approaches to detection and interference with the function of double-stranded RNA molecules.

[1] P. A. Sharp, *Cell* **2009**, *136*, 577–580.

[2] J. R. Thomas, P. J. Hergenrother, *Chem. Rev.* **2008**, *108*, 1171–1224.

[3] L. Guan, M. D. Disney, *ACS Chem. Biol.* **2012**, *7*, 73–86.

[4] M. Li, T. Zengeya, E. Rozners, *J. Am. Chem. Soc.* **2010**, *132*, 8676–8681.

[5] P. Gupta, T. Zengeya, E. Rozners, *Chem. Commun.* **2011**, *47*, 11125–11127.

[6] P. Gupta, O. Muse, E. Rozners, *Biochemistry* **2012**, *51*, 63–73.

[7] S. Buchini, C. J. Leumann, *Angew. Chem.* **2004**, *116*, 4015–4018; *Angew. Chem. Int. Ed.* **2004**, *43*, 3925–3928.

[8] A. B. Eldrup, O. Dahl, P. E. Nielsen, *J. Am. Chem. Soc.* **1997**, *119*, 11116–11117.

[9] L. R. Stewart, M. G. Harris, *J. Org. Chem.* **1978**, *43*, 3123–3126.

[10] T. J. Povsic, P. B. Dervan, *J. Am. Chem. Soc.* **1989**, *111*, 3059–3061.

[11] S. Hildbrand, A. Blaser, S. P. Parel, C. J. Leumann, *J. Am. Chem. Soc.* **1997**, *119*, 5499–5511.

[12] S. A. Cassidy, P. Slickers, J. O. Trent, D. C. Capaldi, P. D. Roselt, C. B. Reese, S. Neidle, K. R. Fox, *Nucleic Acids Res.* **1997**, *25*, 4891–4898.

[13] D. A. Rusling, V. E. C. Powers, R. T. Ranasinghe, Y. Wang, S. D. Osborne, T. Brown, K. R. Fox, *Nucleic Acids Res.* **2005**, *33*, 3025–3032.

[14] D. L. Chen, L. W. McLaughlin, *J. Org. Chem.* **2000**, *65*, 7468–7474.

[15] A. Ono, P. O. P. Ts'o, L. S. Kan, *J. Am. Chem. Soc.* **1991**, *113*, 4032–4033.

[16] G. Xiang, W. Soussou, L. W. McLaughlin, *J. Am. Chem. Soc.* **1994**, *116*, 11155–11156.

[17] G. Xiang, R. Bogacki, L. W. McLaughlin, *Nucleic Acids Res.* **1996**, *24*, 1963–1970.

[18] U. von Krosigk, S. A. Benner, *J. Am. Chem. Soc.* **1995**, *117*, 5361–5362.

[19] M. Egholm, L. Christensen, K. L. Dueholm, O. Buchardt, J. Coull, P. E. Nielsen, *Nucleic Acids Res.* **1995**, *23*, 217–222.

[20] M. E. Hansen, T. Bentin, P. E. Nielsen, *Nucleic Acids Res.* **2009**, *37*, 4498–4507.

[21] R. H. E. Hudson, F. Wojciechowski, *Can. J. Chem.* **2008**, *86*, 1026–1029.

[22] A. L. Feig, *Methods Enzymol.* **2009**, *468*, 409–422.

[23] E. S. Priestley, P. B. Dervan, *J. Am. Chem. Soc.* **1995**, *117*, 4761–4765.

[24] B. Song, Y. Wang, M. A. Titmus, G. Botchkina, A. Formentini, M. Kornmann, J. Ju, *Mol. Cancer* **2010**, *9*, 96.

[25] Z. Jin, F. M. Selaru, Y. Cheng, T. Kan, R. Agarwal, Y. Mori, A. V. Olaru, J. Yang, S. David, J. P. Hamilton, J. M. Abraham, J. Harmon, M. Duncan, E. A. Montgomery, S. J. Meltzer, *Oncogene* **2011**, *30*, 1577–1585.

[26] S. Griffiths-Jones, H. K. Saini, S. van Dongen, A. J. Enright, *Nucleic Acids Res.* **2008**, *36*, D154–D158.

[27] M. Krishnamurthy, K. Simon, A. M. Orendt, P. A. Beal, *Angew. Chem.* **2007**, *119*, 7174–7177; *Angew. Chem. Int. Ed.* **2007**, *46*, 7044–7047; B. D. Gooch, P. A. Beal, *J. Am. Chem. Soc.* **2004**, *126*, 10603–10610.