Pyridazine Nucleobase in Triplex-Forming PNA Improves Recognition of Cytosine Interruptions of Polypurine Tracts in RNA

Nikita Brodyagin[1], Ilze Kumpina[2,†], Justin Applegate[1], Martins Katkevics[2] and Eriks Rozners[1,*]

[1] Department of Chemistry, Binghamton University, The State University of New York, Binghamton, New York 13902, United States
[2] Latvian Institute of Organic Synthesis, Aizkraukles 21, Riga, LV-1006, Latvia

ABSTRACT

Sequence specific recognition of regulatory non-coding RNAs would open new possibilities for fundamental science and medicine. However, molecular recognition of such complex double-stranded RNA (dsRNA) structures remains a formidable problem. Recently, we discovered that peptide nucleic acids (PNAs) form an unusually stable and sequence specific triple helix with dsRNA. Triplex-forming PNAs could become universal tools for recognition of non-coding dsRNAs but are limited by the requirement of polypurine tracts in target RNAs as only purines form stable Hoogsteen hydrogen bonded base triplets. Herein, we systematically surveyed simple nitrogen heterocycles P_N as modified nucleobases for recognition of cytosine in P_N*C-G triplets. We found that 3-pyridazinyl nucleobase formed significantly more stable P_N*C-G triplets than other heterocycles including pyrimidin-2-one previously used by us and others for recognition of cytosine interruptions in polypurine tracts of PNA-dsRNA triplexes. Our results improve triple helical recognition of dsRNA and provide insights for future development of new nucleobases to expand the sequence scope of non-coding dsRNAs that can be targeted by triplex-forming PNAs.

INTRODUCTION

Non-coding RNA (ncRNA) plays important but not well-understood roles in biological processes and disease development.[1-3] Research tools for sequence specific recognition, detection and functional inhibition of such RNAs would be highly useful for fundamental biology and practical applications in biotechnology and medicine. Many ncRNAs adopt complex tertiary structures having double helical regions that can be targeted using triple-helical recognition.[4, 5] However, molecular recognition of such helical RNA structures has been little studied.

Double-helical DNA and RNA can be recognized sequence-specifically by triplex-forming oligonucleotides (TFOs) or their chemically modified derivatives.[6] However, a major limitation for triple helix formation is that only the purine nucleobases can be recognized through two Hoogsteen hydrogen bonds in the natural T*A-T (or U*A-U) and C+*G-C triplets (Figure 1).[6] Recognition of pyrimidine nucleobases in inverted U-A and C-G base pairs is a formidable problem because they present only one hydrogen bond donor or acceptor in the major groove of duplex. Additional problems are low binding affinity and slow kinetics of TFOs, due to the electrostatic repulsion between the negatively charged phosphates of a TFO and the target nucleic acid duplex, and the need for cytosine (pK_a~4.5) protonation to form the C+*G-C triplet, which is unfavorable under physiological conditions.

The latter two problems have been solved in the recent developments of cationic triplex-forming peptide nucleic acid (PNA, Figure 1).[7-12]

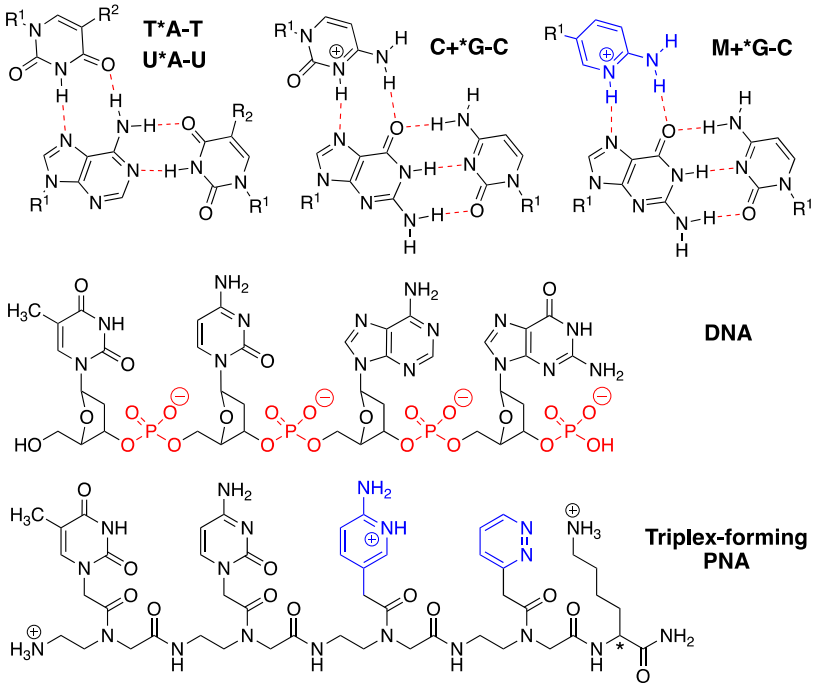


Figure 1. Structures of Hoogsteen hydrogen-bonded base triplets, DNA and triplex-forming PNA. R^1 denotes DNA, RNA, or modified nucleic acid backbones (such as, PNA); R^2 = H or -CH_3.

PNA is a neutral DNA analogue where the nucleobases are connected through a pseudopeptide backbone (Figure 1).[13] PNA was originally developed as a triplex-forming ligand to recognize double-stranded DNA (dsDNA),[14] but the discovery that PNA can invade dsDNA by displacing a pyrimidine rich DNA strand and forming a PNA-DNA-PNA triplex, shifted the PNA development to this new and exciting mode of DNA recognition.[13] While triple-helical binding of PNA to dsDNA has also been explored,[7, 8] binding of PNA to dsRNA was first reported only in 2010 by Rozners and co-workers.[9] Follow up studies by us[10-12] and others[15-19] showed that nucleobase-modified PNA formed triple helices with dsRNA with high affinity and sequence specificity. Replacement of cytosine with the more basic 2-aminopyridine (M) nucleobase (pK_a~6.7, hence, partially protonated at physiological conditions) enabled PNA-dsRNA triplex formation at physiological pH and salt concentration.[10] Following the precedents by Corey[20, 21] and Gait[22, 23], conjugation with cationic lysine residues further increased the binding affinity and facilitated cellular uptake of the cationic triplex-forming PNA (Figure 1) without compromising the binding specificity.[11, 12] Most remarkably, PNA had about ten-fold higher affinity for complementary dsRNA than for the same sequence of dsDNA.[10-12] However, formation of highly stable PNA-dsRNA triplexes was still restricted to sequences where one strand of dsRNA consisted of mostly purine nucleobases (polypurine tracts).[24]

The need for polypurine tracts is arguably the most significant limitation of triple-helical recognition of nucleic acids.[25] Several research groups have designed heterocyclic nucleobases to

form a single hydrogen bond with the exocyclic -NH_2 of cytosine (Figure 2). Leumann showed that 5-methylpyrimidin-2-one (^{4H}T, R^2 = CH_3) modified TFOs selectively recognized the C-G inversion over other base pairs, but the binding affinity was lower than that of standard Hoogsteen triplets.[26] We made similar observations when pyrimidin-2-one (P, R^2 = H) was incorporated in triplex-forming PNA targeting dsRNA.[24] However, despite the lowered affinity, P-modified PNA formed a triplex with a hairpin structure in mRNA and suppressed its translation in vitro and in cells.[27] The latter study was the first demonstration of biological effect of PNA-dsRNA triplex formation in live cells.

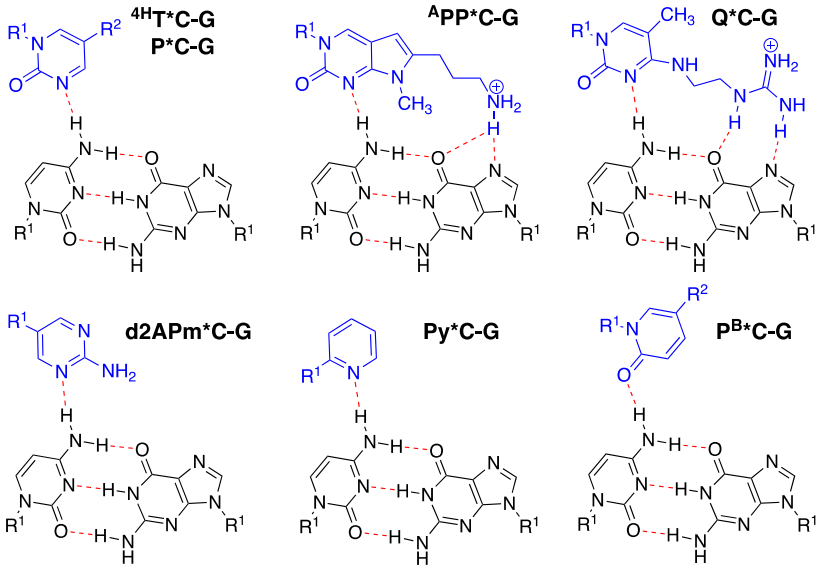


Figure 2. Structures of modified heterocycles for recognition of C-G inversions in polypurine tracts and the proposed hydrogen-bonding schemes. R^1 denotes DNA, RNA, or modified nucleic acid backbones (such as, PNA); R^2 = H or -CH_3.

Fox, Brown and co-workers[28, 29] showed that TFOs modified with ^{A}PP, having the pyrimidin-2-one moiety extended with a cationic amino group, recognized C-G inversion with good affinity and sequence selectivity, comparable to other modified triplets. Seidman and co-workers[30] showed that Q*C-G triplet (Q = guanidinylethyl-5-methylcytosine, Figure 2) had similar stability as T*A-T triplet in DNA, though the sequence selectivity of Q-modified TFOs was somewhat lower. Chen and co-workers studied Q in short triplex-forming PNAs targeting dsRNA and found that Q had good sequence selectivity, but the stability of Q*C-G triplet was reduced compared to T*A-U (~8-fold) or L*G-C (~24-fold, L = thiopseudoisocytosine).[31] Recent studies have demonstrated the potential of L- and Q-modified triplex-forming PNA in modulation of biological activity of a pathogenic RNAs. Chen and co-workers demonstrated that triplex formation by PNA targeting a hairpin structure present in HIV-1 virus stimulated ribosomal frameshifting of viral mRNA.[32] Most recently, Kierzek, Chen, Prabhakaran and co-workers reported that a triplex-forming PNA-neamine conjugate targeting the dsRNA panhandle structure of influenza virus significantly reduced viral replication.[33]

Simple pyridine and pyrimidine derivatives have also shown promise as modified nucleobases for recognition of C-G inversions. Chen and McLaughlin[34] tested 2-aminopyrimidine (d2APm) and

Obika and co-workers[35] tested pyridine (Py) as novel nucleobases in TFOs. Both recognized C-G inversions better than other base pairs, but the overall affinity of the modified TFOs was reduced. The hydrogen-bonding schemes shown in Figure 2 are proposed as conceivable interactions but are not experimentally proven. For example, Imanishi and co-workers[36, 37] showed that 2-pyridone (P^B, Figure 2) had similar binding properties as ^{4H}T and P, despite lacking the endocyclic N, and proposed a different hydrogen-bonding interaction. Therefore, the exact nature of stabilizing interactions, hydrogen-bonding, stacking, charge-charge attraction, etc., is not well understood, leaving space for empirical improvement and optimization.

Despite the significant effort reviewed above, a general and effective solution for recognition of C-G interruptions of polypurine tracts is still lacking. Herein, we report a systematic study of simple nitrogen heterocycles (P_N) as nucleobases in triplex-forming PNA for formation of P_N*C-G triplets. We found that pyridazine-modified PNAs had useful binding affinity and sequence specificity for recognition of C-G inversions in polypurine tracts of dsRNA. Our results show that 3-pyridazinyl nucleobase outperforms other heterocycles, including pyrimidin-2-one previously used by us[24, 27] and others,[19] and forms the most stable P_N*C-G triplets in PNA-dsRNA triplex.

RESULTS

Synthesis and initial screening of modified PNA nucleobases for recognition of cytosine

We started our study by screening all possible variants of six-membered heterocyclic nucleobases containing one (P_1-P_3, pyridine) and two (P_4-P_9, pyrimidine, pyrazine, and pyridazine) nitrogen atoms using the model system of 9-mer PNA_N and RNA hairpin HRP_C (Figure 3), as in our previous studies.[10-12, 38]

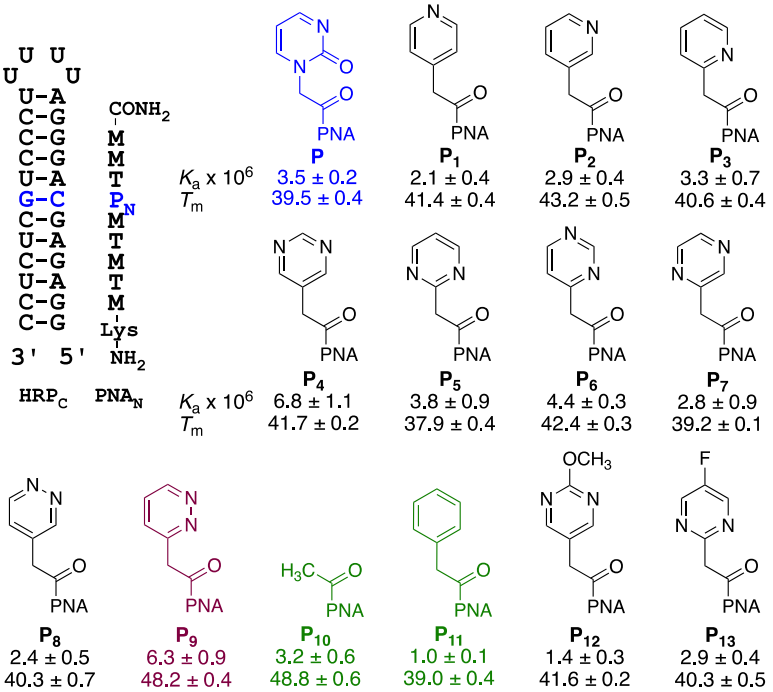
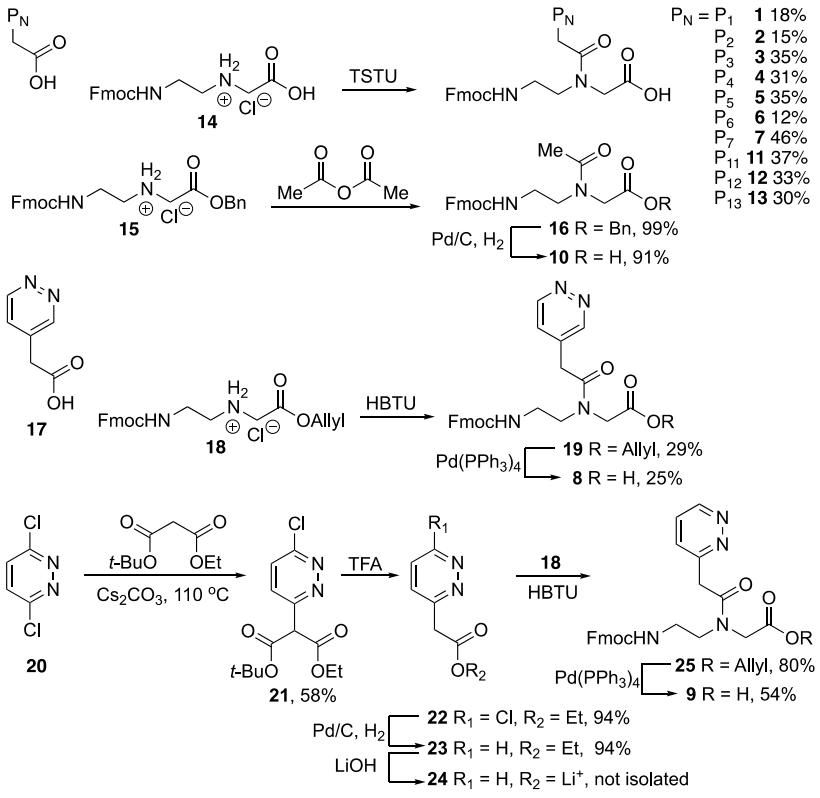


Figure 3. Structures of the heterocyclic nucleobases screened for recognition of the C-G inversion (blue) in the polypurine tract of HRP_C. The numbers following the nucleobase codes (P_1-P_{13}) are association constants ($K_a \times 10^6$ M^{-1}, average of three experiments ± stand. dev., except P_2, which is average of two experiments) for binding of the respective P_N-modified PNA_N to HRP_C and UV thermal melting temperatures (T_m °C, average of five experiments ± stand. dev.) of the corresponding triplexes. Structure of M is shown in Figure 1.

Synthesis of the PNA monomers **1**-**13** bearing the various heterocyclic nucleobases (P_1-P_{13}) was done following our previously reported procedures[39] by coupling of nucleobase acetic acid derivatives with Fmoc-protected PNA backbone **14**, or its benzyl or allyl esters **15** and **18**, respectively, followed by deprotection of the carboxy group (Scheme 1). Nucleobase acetic acid derivatives were either commercially available (P_1, P_2, P_7, P_8, P_{10}, P_{11}), or synthesized (P_3-P_6, P_9, P_{12}, P_{13}) as described in Experimental Section and in Supporting Information. As an example, synthesis of monomer **9** (Scheme 1, P_9) started with nucleophilic aromatic substitution of 3,6-dichloropyridazine with *tert*-butyl ethyl malonate.[40] Cleavage of the *tert*-butyl group and decarboxylation was followed by removal of aromatic chlorine by hydrogenation over Pd/C in the presence of 2 equivalents of triethylamine, which prevented reduction of the pyridazine ring and gave ethyl ester **23** in excellent yield. The ester was cleaved with LiOH and the resulting salt was coupled directly with allyl protected PNA backbone **18** without further purification. The final monomer **9** was obtained after deprotection of the allyl group. Using of allyl protected backbone **18** gave better yields of **9** than direct coupling to unprotected carboxylic acid **14** or benzyl protected backbone **15**.



Scheme 1. Synthesis of PNA monomers.

Binding affinity of P_N-modified PNAs was measured using isothermal titration calorimetry (ITC, as in our previous studies[10-12, 38]) and UV thermal melting at 300 nm. The latter reported specifically on the triple helix formation by measuring changes in absorbance at a wavelength (300 nm) unique to the M nucleobases (for structure of M, see Figure 1). While the correlation between K_a (ITC) and T_m (UV melting) data was not always consistent, overall, both methods showed that most of the modified nucleobases formed triplets with stability similar to that of the original P. The K_a range for most nucleobases with one and two nitrogen atoms was 2.1 to 4.4 × 10^6 M^{-1} (3.5 for the original P) and T_m range was 37.9 to 43.2 °C (39.5 °C for the original P). Notable exceptions were P_4 that had higher K_a, but not T_m, and P_9 (highlighted in maroon in Figure 3) that stood out as forming the most stable triplet as judged by both ITC and UV melting. Two nucleobase surrogates, P_{10} and P_{11} were used as controls to provide insights in the impact of hydrogen bonding, stacking, and steric hindrance on the stability of P_N*C-G triplets. Surprisingly, omission of the aromatic ring in P_{10} did not lead to significant loss of binding affinity. On contrary, the T_m of this modification was similar to that of P_9. Two additional pyrimidine derivatives, P_{12} and P_{13} were tested, but did not show promising binding properties.

Comparison of P_N*C-G triplets in different sequence contexts

The initial screening did not reveal a clear correlation between nitrogen position and the stability of P_N*C-G triplets, which led us to hypothesize that the stability might be driven by stacking of the nucleobase analogues in the triple helical structures. To explore this hypothesis, we selected the original P, and P_4, P_5, P_7 and P_9 (representing various arrangements of nitrogens ortho and meta to the linker connecting to PNA backbone) to further study the stability of P_N*C-G triplets in various sequence contexts (Figure 4). The RNA hairpins in Figure 4 were designed by replacing the A(UUUU)U hairpin loop of HRP_C with the unusually stable C(UUCG)G tetraloop,[41] which improved the quality of ITC data, most likely, by rigidifying the RNA hairpins. Furthermore, we varied the location and neighboring base pairs of the C-G interruption to create all four possible sequence contexts (MP_NT, TP_NM, MP_NM, and TP_NT) while maintaining the same number of T and M nucleobases in the triplex-forming PNAs. If the stacking would play a major role, we expected to observe significant differences in stability of P_N*C-G triplets depending on the immediate neighbors of P_N, T or M in the modified PNAs.

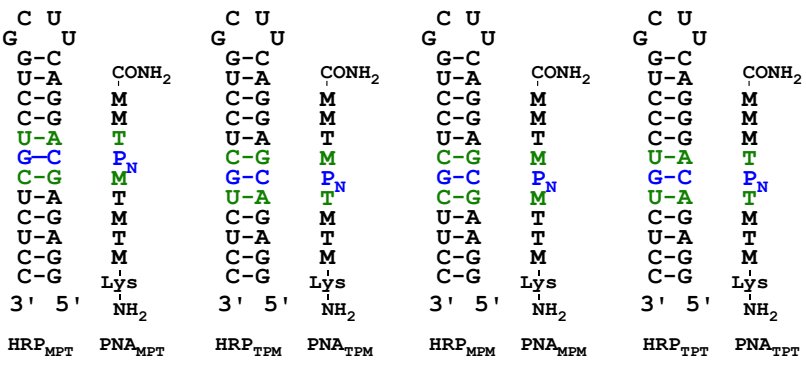
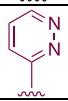


Figure 4. Structures of the RNA hairpins and P_N-modified PNAs used to study the effect of sequence context (neighboring nucleobases) on the stability of P_N*C-G triplets.

The results of ITC and UV thermal melting experiments are shown in Table 1. As in the initial screening, we observed some variance in correlation of K_a and T_m values for the selected nucleobases in different sequence contexts. Nevertheless, the data reveled some common trends. Both ITC and UV melting showed that overall the stabilities of triplexes formed with the MP_NT and TP_NT hairpins were lower than the stability of triplexes formed with the TP_NM and MP_NM hairpins. This suggested that P_N nucleobases were stacking better with M on the carboxy side. ITC showed that the MP_NT sequence context gave the least stable triplexes, while UV melting suggested that the TP_NT triplexes were the least stable.

Table 1. Stability of P_N*C-G triplets in different sequence contexts.

Modified Nucleobase P_N		**HRP_{MPT}** [a]	**HRP_{TPM}** [a]	**HRP_{MPM}** [a]	**HRP_{TPT}** [a]
Abbrev.	Structure	U-A*T G-C*P_N C-G*M	C-G*M G-C*P_N U-A*T	C-G*M G-C*P_N C-G*M	U-A*T G-C*P_N U-A*T
P		1.3 ± 0.1 42.6 ± 0.2	10 ± 1 46.1 ± 0.5	9.7 ± 0.3 48.6 ± 0.6	3.5 ± 0.2 38.0 ± 0.3
P_4		3.9 ± 0.4 44.5 ± 0.4	18 ± 1 49.1 ± 0.3	14 ± 1 50.5 ± 0.6	14 ± 0.5 42.8 ± 0.5
P_5		2.8 ± 0.3 41.4 ± 0.4	12 ± 0.5 43.0 ± 0.5	16 ± 0.5 47.6 ± 0.6	3.3 ± 0.4 35.0 ± 0.4
P_7		1.7 ± 0.1 42.8 ± 0.6	13 ± 1 44.1 ± 0.7	12 ± 1 49.0 ± 0.1	3.8 ± 0.5 37.1 ± 0.5
P_9		**6.7 ± 0.4** [b] **48.5 ± 0.2**	**18 ± 2** **51.0 ± 0.4**	**17 ± 0.5** **49.8 ± 0.5**	**18 ± 1** **48.4 ± 0.3**
T [c]	T*A-U control	12 ± 0.5 69.6 ± 0.8	25 ± 1 75.5 ± 0.7	21 ± 2 76.7 ± 0.5	28 ± 2 70.5 ± 0.5

[a] Association constants, $K_a \times 10^6$ M^{-1}, average of three experiments ± stand. dev. for binding of complementary PNA to the respective RNA hairpin and UV thermal melting temperatures, T_m °C, average of five experiments ± stand. dev. of the corresponding triplexes; [b] Average of two experiments; [c] Control experiments were P_N*C-G triplet in each respective triplex is replaced by the canonical T*A-U triplet.

In the MP_NT and TP_NT sequence contexts, pyridazine P_9 formed the most stable P_N*C-G triplet by both ITC and UV melting (highlighted bold maroon in Table 1). In the TP_NM and MP_NM sequence contexts, all nucleobases tested showed similar stability, with P_9 being only slightly better than P_4, P_5, or P_7. Compared to control triplexes that had the canonical T*A-U triplet in place of the P_N*C-G, ITC results suggested that triplexes formed by P_9-modified PNAs were approaching the stability of the canonical controls. In contrast, UV melting results suggested that the canonical

triplexes were always significantly more stable. We used Van't Hoff analysis to extrapolate ΔG from UV melting curves of P_N*C-G and canonical T*A-U controls to 25 °C (Table S23). Consistent with T_m analysis in Table 1, comparison of ΔG_{25} obtained by ITC and UV melting results (Table S23) showed that the canonical triplexes were more stable. Despite these discrepancies, both ITC and UV melting results clearly showed that pyridazine P_9 was the modified nucleobase of choice that performed consistently well across all sequence contexts.

In all cases, the melting transitions of triplexes where well separated from the melting transitions of hairpins; the latter were difficult to measure precisely because of T_m > 95 °C (Figure S44). To test if the UV thermal meltings measured at 0.5 °C per minute gave true equilibrium T_ms, we recorded the melting curves of PNA TP9T and PNA TP9M complexed with their matched hairpins under both heating and cooling conditions. The resulting melting curves (Figure S46) were almost identical and showed minimal hysteresis of ~ 1 °C. These results were consistent with fast binding of M-modified PNAs to dsRNAs observed in our previous studies[10-12] and confirmed that the recorded T_ms represented true equilibrium transitions of the PNA-dsRNA triplexes.

Stability of PNA-dsRNA triplexes having several P_N*C-G triplets

Next, we tested P_4, P_5, and P_9 in model triplexes featuring two consecutive (PNA2$_P$-HRP2) and three dispersed (PNA3$_P$-HPR3) P_N*C-G triplets (Figure 5). Adding of the second modified nucleobase significantly decreased the stability of PNA2$_P$-HRP2 triplex. The stability of PNA3$_P$-HRP3 triplex was too low to be measured by ITC at pH 7.4; however, decreasing the pH to 6.5 increased the overall binding affinity of M-modified PNAs and enabled comparison of the three nucleobases. In both cases, P_9-modified PNAs were forming significantly stronger triplexes than P_4- or P_5-modified PNAs.

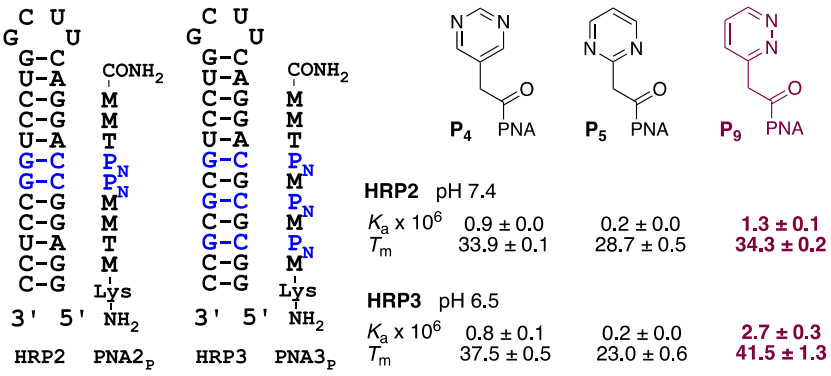
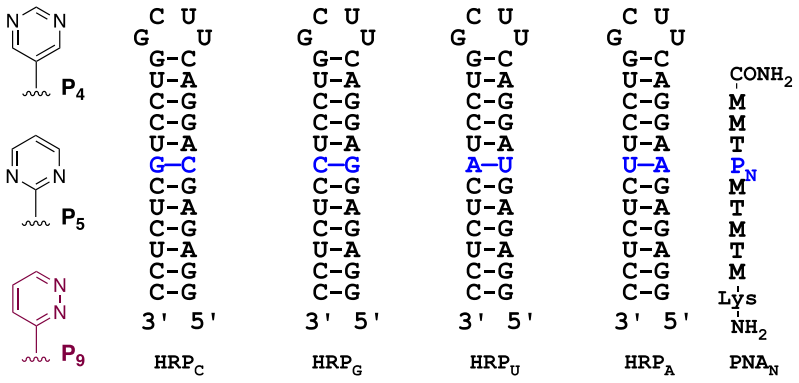


Figure 5. Structures of the RNA hairpins and PNAs having several P_N*C-G triplets. The numbers in the right panel are association constants (K_a × 10^6 M^{-1}, average of three experiments ± stand. dev.) and UV thermal melting temperatures (T_m °C, average of five experiments ± stand. dev.) of the corresponding triplexes.

Sequence specificity of P-modified PNAs

Finally, we tested the sequence specificity of P_4, P_5, and P_9 in model triplexes featuring all four Watson-Crick base pairs at the position of the modified nucleobase (Figure 6). All three

nucleobases had higher affinity for the C-G over other base pairs. The specificity of P_4 and P_5 was somewhat lower as they also formed P_N*A-U triplets with comparable stability. While the selectivity of P_9 was good with affinity consistently about five-fold higher for the matched C-G than for any other base pair, it was somewhat lower than the at least eight-fold higher selectivity of M for its matched G-C base pair observed in our previous study.[12]



	HRP$_C$	HRP$_G$	HRP$_U$	HRP$_A$
P$_4$				
K_a x 10^6	**3.9 ± 0.4**	0.3 ± 0.0	0.8 ± 0.1	2.0 ± 0.1
T_m	**44.5 ± 0.4**	34.7 ± 0.3	37.9 ± 0.4	47.7 ± 0.3
P$_5$				
K_a x 10^6	**2.8 ± 0.4**	0.4 ± 0.0	0.8 ± 0.2	1.1 ± 0.1
T_m	**41.4 ± 0.4**	31.6 ± 0.4	35.6 ± 0.1	38.5 ± 0.6
P$_9$				
K_a x 10^6	**6.7 ± 0.4**	1.5 ± 0.1	1.3 ± 0.2	1.4 ± 0.1
T_m	**48.5 ± 0.2**	36.2 ± 0.3	36.4 ± 0.2	36.5 ± 0.3

Figure 6. Structures of the RNA hairpins used to study sequence specificity of modified nucleobases (P_4, P_5, and P_9). The numbers in the lower panel are association constants (K_a × 10^6 M^{-1}, average of three experiments ± stand. dev.) and UV thermal melting temperatures (T_m °C, average of five experiments ± stand. dev.) of the corresponding triplexes.

DISCUSSION

While only ~2% of DNA encodes for proteins, >75% is transcribed into RNA.[42, 43] Recent discoveries of a variety of non-coding RNAs and the important roles that some of them (e.g., microRNAs) play in cell biology have changed the traditional view of RNA as a passive messenger in the transfer of genetic information from DNA to proteins.[1-3] The functional importance of most RNA transcripts is still unknown. Therefore, the ability to selectively recognize, detect and inhibit the function of complex regulatory RNAs will be highly useful for fundamental biology and practical applications in biotechnology and medicine. However, currently we do not have a general method for sequence selective recognition of dsRNA.

Recently, we discovered that PNA forms an unusually stable and sequence specific triple helix with dsRNA.[9-12] Triplex-forming PNAs could become general tools for molecular recognition of complex regulatory RNA molecules. However, practical applications of triplex-forming PNAs are limited by the requirement of a polypurine tract at the recognition site in dsRNA because only purine nucleobases can be effectively recognized using two Hoogsteen hydrogen bonds (Figure 1). Despite considerable effort in development of modified nucleobases,[6, 25] recognition of pyrimidine interruptions in purine tracts remains a formidable problem. In the present study, we systematically surveyed all nine isomeric pyridine, pyrimidine, pyrazine, and pyridazine heterocycles (Figure 3) as modified nucleobases for recognition of cytosine interruptions in polypurine tracts of dsRNA. We found that 3-pyridazinyl nucleobase P_9 formed more stable P_N*C-G triplets than other isomeric heterocycles. Most importantly, pyridazine P_9 performed significantly better than pyrimidin-2-one P that has been previously used by us[24, 27] and others[19] as the PNA nucleobase of choice for recognition of C-G base pairs in dsRNA triplexes.

In the present study, we measured the triplex stabilities using ITC and UV thermal melting. The UV melting studies were done at 300 nm, a wavelength where we could observe triplex dissociation with little interference from signal due to the hairpin melting (Figures S44-S46). This was possible because the natural nucleobases do not absorb at 300 nm where the M nucleobase still has relatively strong absorbency. Overall, the K_a obtained by ITC and T_m obtained by UV melting confirmed the same trends, but the specific correlation was not always entirely consistent. As discussed in literature,[44] the binding affinities measured at different temperatures by UV thermal melting and isothermal methods (e.g., ITC) may not correlate strongly because of variable dependency of thermodynamic parameters (ΔG, ΔH, etc.) on temperature. Nevertheless, both ITC and UV melting confirmed that P_9 was the best nucleobase for recognition of cytosine in P_N*C-G triplets.

To obtain insights into hydrogen bonding and stacking interactions that govern the stability of P_N*C-G triplets, we also studied two nucleobase surrogates, acetate P_{10} and phenylacetate P_{11} (highlighted in green in Figure 3). Absence of hydrogen bond acceptors in phenylacetate P_{11} decreased the K_a (but not T_m) indicating that the heterocyclic nitrogens may have been involved in stabilizing electrostatic or stacking interactions. Surprisingly, acetate P_{10} formed more stable triplets

(higher K_a and T_m) than phenylacetate P_{11} indicating that a six membered ring may experience some negative steric interactions when forming the P_N*C-G triplets. This notion is supported by our recent structural analysis of PNA-dsRNA triplex[45] showing that pyrimidine rings of Watson-Crick base pairs protrude further in the major groove than purine rings, which may cause some steric clash with the third nucleobase of the triplex. Others have previously shown that abasic sites are tolerated in the third strand of triplexes. Early studies on TFO showed that abasic site formed modestly stable triplets with C-G and T-A interruptions in purine tracts, but was binding poorly to G-C and A-T base pairs in DNA.[46] Nielsen and co-workers reported[47] that P_{10} formed more stable P_{10}*T-A triplets than guanosine, which forms fairly stable G*T-A triplets in DNA.[48] Taken together, these results suggest that future development of novel nucleobases for recognition of pyrimidine interruptions in polypurine tracts should focus on smaller heterocycles, such as five membered rings that may reduce the negative steric interactions.

Our survey of all isomeric one and two nitrogen heterocycles does not reveal a clear correlation between the position of nitrogen and stability of P_N*C-G triplets (Figure 3 and Table 1). Most likely, this is because the binding affinity is driven by subtle interplay of hydrogen bonding and stacking interactions. The variation of binding affinity across different sequence contexts in Table 1 suggests that stacking plays a significant role, with P_N nucleobases stacking better with M on the carboxy side. On the other hand, the sequence specificity of P_4, P_5, and P_9 in model triplexes in Figure 6 suggests that the nitrogen atoms are involved in some hydrogen bonding like electrostatic interactions that discriminate among various Hoogsteen partners. We hypothesize that these interactions are not ideal in either P_4 or P_5, and that P_9 delivers the best compromise by conceivably involving both nitrogen atoms in a bifurcated electrostatic interaction with the exocyclic -NH_2 of cytosine. Bifurcated hydrogen bonds are well-established in non-canonical base pairs of RNA.[49, 50] Testing of this hypothesis may be one of the goals of future structural studies.

CONCLUSIONS

In the present study, we discovered that 3-pyridazinyl nucleobase P_9 formed the most stable P_N*C-G triplet among all possible isomeric pyridine, pyrimidine, pyrazine, and pyridazine heterocycles. Our results suggest that P_9 can be used effectively to recognize single C-G inversion in polypurine tracts of RNA and may have useful affinity for targets having two C-G inversions. However, the latter case and potential targets with three C-G inversions will most likely require longer PNAs than the 9-mers used in this study. Comparison with T*A-T triplets shows that there is still plenty of room for improvement of P_N*C-G triplets. In this context, our survey of 14 different modified nucleobases, and studies of five of them in different sequence contexts provide useful insights for future development of new nucleobases to expand the sequence scope of non-coding RNAs that can be targeted by triplex-forming PNAs. Until then, we propose that 3-pyridazinyl P_9 is currently the PNA nucleobase of choice for triple helical recognition of cytosine interruptions in polypurine tracts of dsRNA.

METHODS

For general procedures and synthesis of PNA monomers, see Supporting Information, pages S1-S17; analytical data (copies of ^{1}H and ^{13}C NMR) are given in Supporting Information, pages S78-S107.

1-(*tert*-Butyl)-3-ethyl-2-(6-chloropyridazin-3-yl)malonate (21). *tert*-Butyl ethyl malonate (2.8 mL, 15 mmol) was added to a solution of 3,6-dichloropyridazine (**20**, 1.5 g, 10 mmol) in DMSO (3 mL). Then Cs_2CO_3 (6.5 g, 20 mmol) was added and the reaction mixture was kept at 110 ˚C for 1 h. The mixture was diluted with EtOAc (100 mL) and water (50 mL) and the aqueous phase was washed with EtOAc (2 × 100 mL). The organic phases were combined, dried over Na_2SO_4, and concentrated under reduced pressure. The crude product was purified by silica gel chromatography using a linear gradient (0-30%) of EtOAc in hexanes to afford the title compound as a yellow oil (1.7 g, 58% yield). R_f = 0.43 (25% EtOAc in hexanes). HRMS (ESI/TOF) *m/z*: $[M + H]^+$ calcd. for $C_{13}H_{17}N_2O_4ClNa$, 323.0775; found 323.0771.^{1}H NMR (400 MHz, $CDCl_3$, ppm) δ: 7.76 (1H, d, *J* = 8.9 Hz), 7.51 (1H, d, *J* = 8.9 Hz), 5.10 (1H, s), 4.31 – 4.05 (2H, m), 1.40 (9H, s), 1.23 (3H, t, *J* = 7.1 Hz). ^{13}C NMR (101 MHz, $CDCl_3$, ppm) δ: 166.76, 165.41, 156.74, 155.94, 130.06, 128.28, 83.72, 62.38, 59.00, 27.76, 13.98.

Ethyl 2-(6-chloropyridazin-3-yl)acetate (22). TFA (6 mL) was added to a solution of 1-(tert-butyl) 3-ethyl 2-(6-chloropyridazin-3-yl)malonate (**21**, 1.6 g, 5.4 mmol) in CH_2Cl_2 (6 mL). The reaction was completed in 30 min at room temperature. The reaction mixture was concentrated under reduced pressure and purified by silica gel chromatography using a linear gradient (0-50%) of EtOAc in hexanes to afford the title compound as a pale-yellow oil (1.0 g, 94% yield). R_f = 0.68 (5% MeOH in CH_2Cl_2). HRMS (ESI/TOF) *m/z*: $[M + H]^+$ calcd. for $C_8H_{10}N_2O_2Cl$, 201.0431; found 201.0433. ^{1}H NMR (400 MHz, $CDCl_3$, ppm) δ: 7.55 (1H, d, *J* = 8.8 Hz), 7.48 (1H, d, *J* = 8.8 Hz), 4.14 (2H, q, *J* = 7.1 Hz), 4.00 (2H, s), 1.22 (3H, t, *J* = 7.2 Hz).^{13}C NMR (101 MHz, $CDCl_3$, ppm) δ: 169.39, 156.40, 155.97, 130.03, 128.31, 61.52, 40.84, 14.07.

Ethyl 2-(pyridazin-3-yl)acetate (23). NEt_3 (1.26 mL, 9.01 mmol) was added to a solution of ethyl 2-(6-chloropyridazin-3-yl)acetate (**22**, 909 mg, 4.53 mmol) in EtOH (12 mL). The mixture was purged with nitrogen followed by addition 10% Pd/C (182 mg). Hydrogen gas (1 atm) was bubbled through the mixture for 4 h at room temperature. Then reaction mixture was filtered through a pad of celite, the filtrate was concentrated, and the residue was purified by silica gel chromatography using a linear gradient (0-4%) of MeOH in CH_2Cl_2 to afford the title compound as a pale-yellow oil (708 mg, 94% yield). R_f = 0.38 (5% MeOH in CH_2Cl_2). HRMS (ESI/TOF) *m/z*: $[M + H]^+$ calcd. for $C_8H_{11}N_2O_2$, 167.0821; found 167.0826. ^{1}H NMR (400 MHz, $CDCl_3$, ppm) δ: 9.07 (1H, dd, *J* = 4.9, 1.8 Hz), 7.53 (1H, dd, *J* = 8.5, 1.7 Hz), 7.43 (1H, dd, *J* = 8.5, 4.9 Hz), 4.15 (2H, q, *J* = 7.2 Hz), 4.02 (2H, s), 1.22 (3H, t, *J* = 7.2 Hz).^{13}C NMR (101 MHz, $CDCl_3$, ppm) δ: 169.78, 157.16, 150.35, 127.45, 126.49, 61.35, 41.73, 14.08.

Allyl *N*-(2-((((9*H*-fluoren-9-yl)methoxy)carbonyl) amino)ethyl)-*N*-(2-(pyridazin-3-yl)acetyl)glycinate (24). A solution of LiOH (50 mg, 2.1 mmol) water (2.2 mL) was added to a solution of ethyl 2-(pyridazin-3-yl)acetate (**23**, 315 mg, 1.90 mmol) in EtOH (5.4 mL). After 1 h at room

temperature, the mixture was concentrated under reduced pressure. The crude product was dried on high vacuum for 6 h and used directly in the next step without further purification. *i*-Pr_2NEt (430 µL, 2.47 mmol, 1.3 equiv.) was added under nitrogen to a solution of the crude lithium 2-(pyridazin-3-yl)acetate (1.90 mmol), HBTU (722 mg, 1.90 mmol), N-Fmoc-Aeg-O-Allyl HCl salt (**18**, 632 mg, 1.52 mmol) in anhydrous DMF (13 mL). The reaction mixture was stirred for 12 h at room temperature, and then concentrated under reduced pressure. The residue was dissolved in CH_2Cl_2 (50 mL) and washed with aqueous 5% $NaHCO_3$ (30 mL). The aqueous layer was back-extracted with CH_2Cl_2 (50 mL), the organic layers were combined, dried over Na_2SO_4 and concentrated under reduced pressure. The crude product was purified by silica gel chromatography using a linear gradient (0-5%) of MeOH in CH_2Cl_2 to afford the title compound as a yellow oil (757 mg, 80% yield). R_f = 0.27 (5% MeOH in CH_2Cl_2). HRMS (ESI/TOF) *m/z*: $[M + H]^+$ calcd. for $C_{28}H_{29}N_4O_5$, 501.2138; found 501.2133. ^{1}H NMR (400 MHz, $CDCl_3$, ppm) (mixture of rotamers) δ: 9.30 – 8.80 (1H, m), 7.74 (2H, dd, *J* = 7.6, 3.8 Hz), 7.64 – 7.49 (3H, m), 7.37 (3H, tt, *J* = 6.3, 3.2 Hz), 7.29 (2H, dt, *J* = 7.4, 1.7 Hz), 5.88 (1H, tq, *J* = 11.7, 5.8 Hz), 5.39 – 5.19 (2H, m), 4.62 (2H, dt, *J* = 6.0, 1.3 Hz), 4.47 – 4.29 (2H, m), 4.26 – 3.96 (4H, m), 3.76 – 3.50 (2H, m), 3.38 (2H, tt, *J* = 12.1, 5.8 Hz). ^{13}C NMR (101 MHz, $CDCl_3$, ppm) (mixture of rotamers) δ: 170.02, 169.63, 158.14, 156.61, 150.43, 150.35, 143.97, 143.86, 141.31, 131.43, 131.16, 127.88, 127.73, 127.71, 127.11, 127.07, 126.65, 126.55, 125.09, 119.99, 119.97, 119.63, 119.00, 66.82, 66.71, 66.58, 66.10, 51.15, 49.79, 49.02, 47.23, 41.41, 40.46, 39.44, 39.21.

***N*-(2-((((9*H*-Fluoren-9-yl)methoxy)carbonyl) amino)ethyl)-*N*-(2-(pyridazin-3-yl)acetyl)glycine (9).** $Pd(PPh_3)_4$ (37 mg, 0.03 mmol) and *N*-ethylaniline (173 µL, 1.38 mmol) were added sequentially to a solution of allyl ester of pyridazin-3-yl monomer (**24**, 407 mg, 0.81 mmol) in anhydrous THF (30 mL). The solution was stirred under nitrogen for 3 h at room temperature. After the reaction was complete, the mixture was acidified to pH 5-6 with 1 M aqueous HCl and solvent was removed under reduced pressure. The crude product was purified by C18 reverse phase flash chromatography using a linear gradient (0-80%) of MeCN in water to afford the title compound as a pale-yellow foam (202 mg, 54% yield). R_f = 0.06 (20% MeOH in CH_2Cl_2). HRMS (ESI/TOF) *m/z*: $[M + H]^+$ calcd. for $C_{25}H_{25}N_4O_5$, 461.1825; found 461.1827. ^{1}H NMR (400 MHz, DMSO-d_6, ppm) (mixture of rotamers) δ: 9.07 (1H, dd, *J* = 4.4, 2.2 Hz), 7.88 (3H, d, *J* = 7.5 Hz), 7.68 (2H, d, *J* = 7.5 Hz), 7.62 – 7.49 (2H, m), 7.40 (2H, t, *J* = 7.4 Hz), 7.31 (2H, t, *J* = 7.4 Hz), 4.34 – 4.16 (3H, m), 4.04 (1H, d, *J* = 4.6 Hz), 3.97 (2H, d, *J* = 3.4 Hz), 3.81 (2H, d, *J* = 8.1 Hz), 3.56 – 3.47 (1H, m), 3.39 (2H, d, *J* = 6.2 Hz), 3.26 (1H, d, *J* = 6.3 Hz), 3.18 (2H, q, *J* = 6.2 Hz). ^{13}C NMR (101 MHz, DMSO-d_6, ppm) (mixture of rotamers) δ: 170.59, 159.70, 156.57, 150.59, 144.42, 141.14, 128.72, 128.07, 127.63, 126.80, 125.83, 125.66, 120.59, 120.54, 66.04, 47.87, 47.18, 40.75, 39.07, 38.66.

The PNAs were synthesized on an Expedite 8909 synthesizer at 2 µmol scale on NovaSyn TG Sieber resin (Novabiochem) using methods previously developed in our group.[10-12, 38] Commercial PNA-T-monomer was purchased from Link Technologies. M monomer was synthesized using the synthetic route reported by our group.[10] Crude PNAs were analyzed by LC-MS and purified using a semi-preparative Supelco Discovery Wide Pore C18 column (4.6 x 150 mm) and a linear gradient of acetonitrile in water containing 0.1% formic acid. The purity and identity of the PNA sequences were

confirmed by LC-MS (ESI) analysis (Figures S1-S39). PNA was quantified as previously reported [38]. RNA hairpins were purchased crude from Dharmacon and purified prior to use on reverse phase HPLC using a gradient of acetonitrile in 50 mM aqueous triethylammonium acetate buffer as previously reported.[38]

Isothermal titration calorimetry experiments were done on a MicroCal iTC200 instrument at 25 °C in 50 mM potassium phosphate buffer (pH 7.4) containing 2 mM $MgCl_2$, 90 mM KCl, 10 mM NaCl, as previously reported.[38] In a typical ITC experiment, 2.45 μL aliquots of 90 μM PNA solution were sequentially injected from a 40 μL rotating syringe (750 rmp) into 200 μL of 10 μM RNA hairpin solution. All ITC experiments were run in triplicate (Tables S2-S11). Representative ITC titration traces are given in Supporting Information, Figures S40-S43.

UV thermal melting experiments were done on a Shimadzu UV-2600 spectrophotometer equipped with a TMSPC-8 temperature controller. Experiments were done with 18 μM PNA and dsRNA in phosphate buffer (2 mM $MgCl_2$, 90 mM KCl, 10 mM NaCl, 50 mM potassium phosphate at pH 7.4). Absorbance vs. temperature profiles were measured at 300 nm (Figures S44-S46). The temperature was increased at a rate of 0.5 °C per minute. The melting temperatures (T_m, °C) were obtained using Shimadzu LabSolutions Tm Analysis software version 1.31 (Tables S12-S23).

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website.

Synthesis of PNA monomers, LC-MS characterization of PNA oligomers; ITC results and representative titration images; UV melting results; copies of ^1H and ^{13}C NMR spectra of synthetic intermediates and products (PDF).

AUTHOR INFORMATION

Corresponding Author

* Eriks Rozners − Department of Chemistry, Binghamton University, The State University of New York, Binghamton, New York 13902, United States; orcid.org/0000-0001-7649-0040; Phone: (1) 607-777-2441; Email: erozners@binghamton.edu.

Present Addresses

† Current address: Department of Chemistry, Binghamton University, The State University of New York, Binghamton, New York 13902, United States.

Funding Sources

This work was supported by National Institutes of Health (R35 GM130207), National Science Foundation (CHE-1708761), and Latvian Institute of Organic Synthesis internal research fund (IG-2019-04).

ACKNOWLEDGMENT

We thank C. A. Ryan and V. Kumar for help with the ITC data analysis, and V. Kumar for design of UV melting experiments at 300 nm. This work was supported by National Institutes of Health (R35 GM130207 to E.R.), National Science Foundation (CHE-1708761 to E.R.), and Latvian Institute of Organic Synthesis internal research fund (IG-2019-04 to I.K.).

ABBREVIATIONS

Aeg, aminoethyl glycine; dsDNA, double-stranded DNA; dsRNA, double-stranded RNA; Fmoc, fluorenylmethyloxycarbonyl; HBTU, 2-(1H-benzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate; HRP, hairpin; ITC, isothermal titration calorimetry; K_a, association constant; M, 2-aminopyridine; ncRNA, non-coding RNA; PNA, peptide nucleic acid; TFA, trifluoroacetic acid, TFOs, triplex-forming oligonucleotides; T_m, melting temperature; TSTU, N,N,N′,N′-Tetramethyl-O-(N-succinimidyl)uronium tetrafluoroborate; UV, ultraviolet.

REFERENCES

[1] Smith, K. N., Miller, S. C., Varani, G., Calabrese, J. M., and Magnuson, T. (2019) Multimodal Long Noncoding RNA Interaction Networks: Control Panels for Cell Fate Specification, *Genetics 213*, 1093.

[2] Yao, R.-W., Wang, Y., and Chen, L.-L. (2019) Cellular functions of long noncoding RNAs, *Nat. Cell Biol. 21*, 542-551.

[3] Treiber, T., Treiber, N., and Meister, G. (2019) Regulation of microRNA biogenesis and its crosstalk with other cellular pathways, *Nat. Rev. Mol. Cell Biol. 20*, 5-20.

[4] Mercer, T. R., and Mattick, J. S. (2013) Structure and function of long noncoding RNAs in epigenetic regulation, *Nat. Struct. Mol. Biol. 20*, 300-307.

[5] Novikova, I. V., Hennelly, S. P., Tung, C.-S., and Sanbonmatsu, K. Y. (2013) Rise of the RNA machines: Exploring the structure of long non-coding RNAs, *J. Mol. Biol. 425*, 3731-3746.

[6] Fox, K. R., Brown, T., and Rusling, D. A. DNA Recognition by Parallel Triplex Formation, In *Chemical Biology No. 7: DNA-targeting Molecules as Therapeutic Agents* (Waring, M. J., Ed.), Royal Society of Chemistry, 2018, 1-32.

[7] Hansen, M. E., Bentin, T., and Nielsen, P. E. (2009) High-affinity triplex targeting of double stranded DNA using chemically modified peptide nucleic acid oligomers, *Nucleic Acids Res. 37*, 4498-4507.

[8] Wittung, P., Nielsen, P., and Norden, B. (1997) Extended DNA-recognition repertoire of peptide nucleic acid (PNA): PNA-dsDNA triplex formed with cytosine-rich homopyrimidine PNA, *Biochemistry 36*, 7973-7979.

[9] Li, M., Zengeya, T., and Rozners, E. (2010) Short Peptide Nucleic Acids Bind Strongly to Homopurine Tract of Double Helical RNA at pH 5.5, *J. Am. Chem. Soc. 132*, 8676-8681.

[10] Zengeya, T., Gupta, P., and Rozners, E. (2012) Triple Helical Recognition of RNA Using 2-Aminopyridine-Modified PNA at Physiologically Relevant Conditions, *Angew. Chem., Int. Ed. 51*, 12593-12596.

[11] Muse, O., Zengeya, T., Mwaura, J., Hnedzko, D., McGee, D. W., Grewer, C. T., and Rozners, E. (2013) Sequence Selective Recognition of Double-Stranded RNA at Physiologically Relevant Conditions Using PNA-Peptide Conjugates, *ACS Chem. Biol. 8*, 1683-1686.

[12] Hnedzko, D., McGee, D. W., Karamitas, Y. A., and Rozners, E. (2017) Sequence-selective recognition of double-stranded RNA and enhanced cellular uptake of cationic nucleobase and backbone-modified peptide nucleic acids, *RNA 23*, 58-69.

[13] Nielsen, P. E., Egholm, M., Berg, R. H., and Buchardt, O. (1991) Sequence-selective recognition of DNA by strand displacement with a thymine-substituted polyamide, *Science 254*, 1497-1500.

[14] Buchardt, O., Egholm, M., Berg, R. H., and Nielsen, P. E. (1993) Peptide nucleic acids and their potential applications in biotechnology, *Trends Biotechnol. 11*, 384-386.

[15] Zhou, Y., Kierzek, E., Loo, Z. P., Antonio, M., Yau, Y. H., Chuah, Y. W., Geifman-Shochat, S., Kierzek, R., and Chen, G. (2013) Recognition of RNA duplexes by chemically modified triplex-forming oligonucleotides, *Nucleic Acids Res. 41*, 6664-6673.

[16] Devi, G., Yuan, Z., Lu, Y., Zhao, Y., and Chen, G. (2014) Incorporation of thio-pseudoisocytosine into triplex-forming peptide nucleic acids for enhanced recognition of RNA duplexes, *Nucleic Acids Res. 42*, 4008-4018.

[17] Sato, T., Sakamoto, N., and Nishizawa, S. (2018) Kinetic and thermodynamic analysis of triplex formation between peptide nucleic acid and double-stranded RNA, *Org. Biomol. Chem. 16*, 1178-1187.

[18] Taehtinen, V., Granqvist, L., Murtola, M., Stroemberg, R., and Virta, P. (2017) 19F NMR Spectroscopic Analysis of the Binding Modes in Triple-Helical Peptide Nucleic Acid (PNA)/MicroRNA Complexes, *Chem. Eur. J. 23*, 7113-7124.

[19] Kim, K. T., Chang, D., and Winssinger, N. (2018) Double-Stranded RNA-Specific Templated Reaction with Triplex Forming PNA, *Helv. Chim. Acta 101*, e1700295.

[20] Hu, J., Matsui, M., Gagnon, K. T., Schwartz, J. C., Gabillet, S., Arar, K., Wu, J., Bezprozvanny, I., and Corey, D. R. (2009) Allele-specific silencing of mutant huntingtin and ataxin-3 genes by targeting expanded CAG repeats in mRNAs, *Nat. Biotechnol. 27*, 478-484.

[21] Hu, J., and Corey, D. R. (2007) Inhibiting Gene Expression with Peptide Nucleic Acid (PNA)-Peptide Conjugates That Target Chromosomal DNA, *Biochemistry 46*, 7581-7589.

[22] Fabani, M. M., and Gait, M. J. (2008) miR-122 targeting with LNA/2'-O-methyl oligonucleotide mixmers, peptide nucleic acids (PNA), and PNA-peptide conjugates, *RNA 14*, 336-346.

[23] Turner, J. J., Ivanova, G. D., Verbeure, B., Williams, D., Arzumanov, A. A., Abes, S., Lebleu, B., and Gait, M. J. (2005) Cell-penetrating peptide conjugates of peptide nucleic acids (PNA) as inhibitors of HIV-1 Tat-dependent trans-activation in cells, *Nucleic Acids Res. 33*, 6837-6849.

[24] Gupta, P., Zengeya, T., and Rozners, E. (2011) Triple helical recognition of pyrimidine inversions in polypurine tracts of RNA by nucleobase-modified PNA, *Chem. Commun. 47*, 11125-11127.

[25] Hari, Y., Obika, S., and Imanishi, T. (2012) Towards the sequence-selective recognition of double-stranded DNA containing pyrimidine-purine interruptions by triplex-forming oligonucleotides, *Eur. J. Org. Chem. 2012*, 2875-2887.

[26] Prevot-Halter, I., and Leumann, C. J. (1999) Selective recognition of a C-G base-pair in the parallel DNA triple-helical binding motif, *Bioorg. Med. Chem. Lett. 9*, 2657-2660.

[27] Endoh, T., Hnedzko, D., Rozners, E., and Sugimoto, N. (2016) Nucleobase-Modified PNA Suppresses Translation by Forming a Triple Helix with a Hairpin Structure in mRNA In Vitro and in Cells, *Angew. Chem., Int. Ed. 55*, 899-903.

[28] Ranasinghe, R. T., Rusling, D. A., Powers, V. E. C., Fox, K. R., and Brown, T. (2005) Recognition of CG inversions in DNA triple helices by methylated 3H-pyrrolo[2,3-d]pyrimidin-2(7H)-one nucleoside analogues, *Chem. Commun.*, 2555-2557.

[29] Rusling, D. A., Powers, V. E. C., Ranasinghe, R. T., Wang, Y., Osborne, S. D., Brown, T., and Fox, K. R. (2005) Four base recognition by triplex-forming oligonucleotides at physiological pH, *Nucleic Acids Res. 33*, 3025-3032.

[30] Semenyuk, A., Darian, E., Liu, J., Majumdar, A., Cuenoud, B., Miller, P. S., MacKerell, A. D., Jr., and Seidman, M. M. (2010) Targeting of an Interrupted Polypurine:Polypyrimidine Sequence in Mammalian Cells by a Triplex-Forming Oligonucleotide Containing a Novel Base Analogue, *Biochemistry 49*, 7867-7878.

[31] Toh, D.-F. K., Devi, G., Patil, K. M., Qu, Q., Maraswami, M., Xiao, Y., Loh, T. P., Zhao, Y., and Chen, G. (2016) Incorporating a guanidine-modified cytosine base into triplex-forming PNAs for the recognition of a C-G pyrimidine-purine inversion site of an RNA duplex, *Nucleic Acids Res. 44*, 9071-9082.

[32] Puah, R. Y., Jia, H., Maraswami, M., Kaixin Toh, D.-F., Ero, R., Yang, L., Patil, K. M., Lerk Ong, A. A., Krishna, M. S., Sun, R., Tong, C., Huang, M., Chen, X., Loh, T. P., Gao, Y.-G., Liu, D. X., and Chen, G. (2018) Selective Binding to mRNA Duplex Regions by Chemically Modified Peptide Nucleic Acids Stimulates Ribosomal Frameshifting, *Biochemistry 57*, 149-159.

[33] Kesy, J., Patil, K. M., Kumar, S. R., Shu, Z., Yong, H. Y., Zimmermann, L., Ong, A. A. L., Toh, D.-F. K., Krishna, M. S., Yang, L., Decout, J.-L., Luo, D., Prabakaran, M., Chen, G., and Kierzek, E. (2019) A Short Chemically Modified dsRNA-Binding PNA (dbPNA) Inhibits Influenza Viral Replication by Targeting Viral RNA Panhandle Structure, *Bioconjugate Chem. 30*, 931-943.

[34] Chen, D. L., and McLaughlin, L. W. (2000) Use of pKa Differences To Enhance the Formation of Base Triplets Involving C-G and G-C Base Pairs, *J. Org. Chem. 65*, 7468-7474.

[35] Hari, Y., Matsugu, S., Inohara, H., Hatanaka, Y., Akabane, M., Imanishi, T., and Obika, S. (2010) 2′,4′-BNA bearing a 2-pyridine nucleobase for CG base pair recognition in the parallel motif triplex DNA, *Org. Biomol. Chem. 8*, 4176-4180.

[36] Obika, S., Hari, Y., Sekiguchi, M., and Imanishi, T. (2001) A 2',4'-bridged nucleic acid containing 2-pyridone as a nucleobase: efficient recognition of a C.bul.G interruption by triplex formation with a pyrimidine motif, *Angew. Chem., Int. Ed. 40*, 2079-2081.

[37] Obika, S., Hari, Y., Sekiguchi, M., and Imanishi, T. (2002) Stable oligonucleotide-directed triplex formation at target sites with CG interruptions: strong sequence-specific recognition by

2',4'-bridged nucleic-acid-containing 2-pyridones under physiological conditions, *Chem. Eur. J. 8*, 4796-4802.

[38] Brodyagin, N., Hnedzko, D., MacKay, J. A., and Rozners, E. Nucleobase-Modified Triplex-Forming Peptide Nucleic Acids for Sequence-Specific Recognition of Double-Stranded RNA, In *In Peptide Nucleic Acids. From Chemistry to Animals (Methods in Molecular Biology)* (Nielsen, P., Ed.), Springer Nature, 2020, 157-172.

[39] Kumpina, I., Brodyagin, N., MacKay, J. A., Kennedy, S. D., Katkevics, M., and Rozners, E. (2019) Synthesis and RNA-binding properties of extended nucleobases for triplex-forming peptide nucleic acids, *J. Org. Chem. 84*, 13276-13298.

[40] Wu, H.; Long, C.; Lin, J.; Chen, X.; Li, Y.; Liu, Z.; Wei, C.; Chen, L.; Chen, S. Quinoline derivatives as SMO inhibitors. US20170174703. June 22, 2017.

[41] Antao, V. P., Lai, S. Y., and Tinoco, I., Jr. (1991) A thermodynamic study of unusually stable RNA and DNA hairpins, *Nucleic Acids Res. 19*, 5901-5905.

[42] Consortium, I. H. G. S. (2004) Finishing the euchromatic sequence of the human genome, *Nature 431*, 931-945.

[43] Djebali, S., Davis, C. A., Merkel, A., Dobin, A., Lassmann, T., Mortazavi, A., Tanzer, A., Lagarde, J., Lin, W., Schlesinger, F., Xue, C., Marinov, G. K., Khatun, J., Williams, B. A., Zaleski, C., Rozowsky, J., Roeder, M., Kokocinski, F., Abdelhamid, R. F., Alioto, T., Antoshechkin, I., Baer, M. T., Bar, N. S., Batut, P., Bell, K., Bell, I., Chakrabortty, S., Chen, X., Chrast, J., Curado, J., Derrien, T., Drenkow, J., Dumais, E., Dumais, J., Duttagupta, R., Falconnet, E., Fastuca, M., Fejes-Toth, K., Ferreira, P., Foissac, S., Fullwood, M. J., Gao, H., Gonzalez, D., Gordon, A., Gunawardena, H., Howald, C., Jha, S., Johnson, R., Kapranov, P., King, B., Kingswood, C., Luo, O. J., Park, E., Persaud, K., Preall, J. B., Ribeca, P., Risk, B., Robyr, D., Sammeth, M., Schaffer, L., See, L.-H., Shahab, A., Skancke, J., Suzuki, A. M., Takahashi, H., Tilgner, H., Trout, D., Walters, N., Wang, H., Wrobel, J., Yu, Y., Ruan, X., Hayashizaki, Y., Harrow, J., Gerstein, M., Hubbard, T., Reymond, A., Antonarakis, S. E., Hannon, G., Giddings, M. C., Ruan, Y., Wold, B., Carninci, P., Guigo, R., and Gingeras, T. R. (2012) Landscape of transcription in human cells, *Nature 489*, 101-108.

[44] Mergny, J.-L., and Lacroix, L. (2003) Analysis of thermal melting curves, *Oligonucleotides 13*, 515-537.

[45] Kotikam, V., Kennedy, S. D., MacKay, J. A., and Rozners, E. (2019) Synthetic, Structural, and RNA Binding Studies on 2-Aminopyridine-Modified Triplex-Forming Peptide Nucleic Acids, *Chem. Eur. J. 25*, 4367-4372.

[46] Horne, D. A., and Dervan, P. B. (1991) Effects of an abasic site on triple helix formation characterized by affinity cleaving, *Nucleic Acids Res. 19*, 4963-4965.

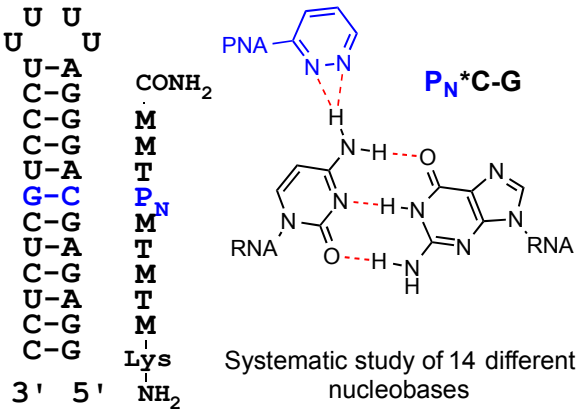
[47] Eldrup, A. B., Dahl, O., and Nielsen, P. E. (1997) A Novel Peptide Nucleic Acid Monomer for Recognition of Thymine in Triple-Helix Structures, *J. Am. Chem. Soc. 119*, 11116-11117.

[48] Greenberg, W. A., and Dervan, P. B. (1995) Energetics of Formation of Sixteen Triple Helical Complexes Which Vary at a Single Position Within a Purine Motif, *J. Am. Chem. Soc. 117*, 5016-5022.

[49] Sponer, J., Leszczynski, J., and Hobza, P. (2001) Electronic properties, hydrogen bonding, stacking, and cation binding of DNA and RNA bases, *Biopolymers 61*, 3-31.

[50] Berger, K. D., Kennedy, S. D., and Turner, D. H. (2019) Nuclear Magnetic Resonance Reveals That GU Base Pairs Flanking Internal Loops Can Adopt Diverse Structures, *Biochemistry 58*, 1094-1108.

Table of Contents (TOC)



Systematic study of 14 different nucleobases