Nucleic Acids Research, 2026, **54**, gkag249
https://doi.org/10.1093/nar/gkag249
Chemical Biology and Nucleic Acid Chemistry

5-Triazolyl uracil improves nucleobase stacking, pyrimidine recognition, biological activity, and specificity of RNA-binding triplex-forming peptide nucleic acids

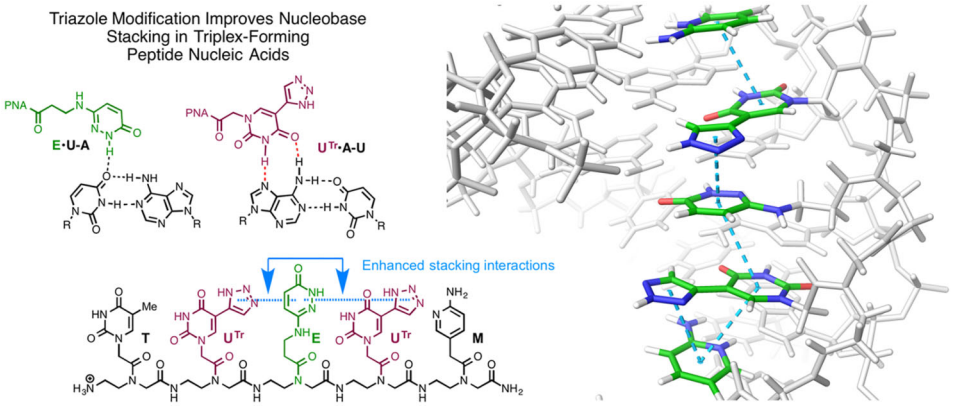
Sara F. Saei[1], Tamaki Endoh[2,3], Vladislavs Baskevics[4], Ilze Kumpina[1], Martins Katkevics[4], Naoki Sugimoto[2], Eriks Rozners[1,*]

[1]Department of Chemistry, Binghamton University, The State University of New York, Binghamton, NY 13902, United States
[2]Frontier Institute for Biomolecular Engineering Research (FIBER), Konan University, 7-1-20 Minatojima-minamimachi, Chuo-ku, Kobe 650-0047, Japan
[3]Graduate School of Frontiers of Innovative Research in Science and Technology (FIRST), Konan University, 7-1-20 Minatojima-minamimachi, Chuo-ku, Kobe 650-0047, Japan
[4]Biologically Active Compound Synthesis Laboratory, Latvian Institute of Organic Synthesis, Aizkraukles 21, Riga LV-1006, Latvia
*To whom correspondence should be addressed. Email erozners@binghamton.edu

Abstract

Cellular RNA folds into complex helical structures that are attractive targets for controlling biological activity and treating diseases using RNA-binding ligands. Peptide nucleic acid (PNA) is a neutral DNA analog uniquely suited for triple-helical recognition of RNA. However, the applications of triplex-forming PNAs have been limited by sequence restrictions, as stable base triples are formed only with purines. Hoogsteen recognition of pyrimidines has been a long-standing challenge. The present study explores a new approach to improve pyrimidine recognition by enhancing nucleobase stacking in the PNA strand. Placing 5-triazolyl uracil adjacent to the modified nucleobases designed for pyrimidine recognition increased the binding affinity while maintaining sequence specificity. Molecular dynamics simulations confirmed the beneficial effect of synergistic stacking interactions between the triazole and neighboring nucleobases. The increased binding affinity improved the biological activity of triplex-forming PNAs targeting microRNA precursors in a cell-based dual fluorescent protein assay. RNA-seq analysis showed that a PNA recognizing a mixed sequence of three nucleobases (G, A, and U) had higher specificity than a similar PNA featuring a uniform purine recognition tract. These results demonstrate a novel strategy for addressing the issue of pyrimidine recognition, a critical bottleneck in the triple helical targeting of nucleic acids.

Graphical abstract



Introduction

RNA plays a wide variety of regulatory roles in cell biology [1–4]. Recent approvals of small interfering RNAs, mRNA vaccines, and RNA-based gene editors as new therapeutic modalities have further increased the interest in these fascinating biopolymers [5, 6]. Regulatory RNAs fold into complex secondary and tertiary structures that govern their biological activity [7]. Sequence-specific recognition of folded

Received: September 6, 2025. **Revised:** February 12, 2026. **Accepted:** March 3, 2026

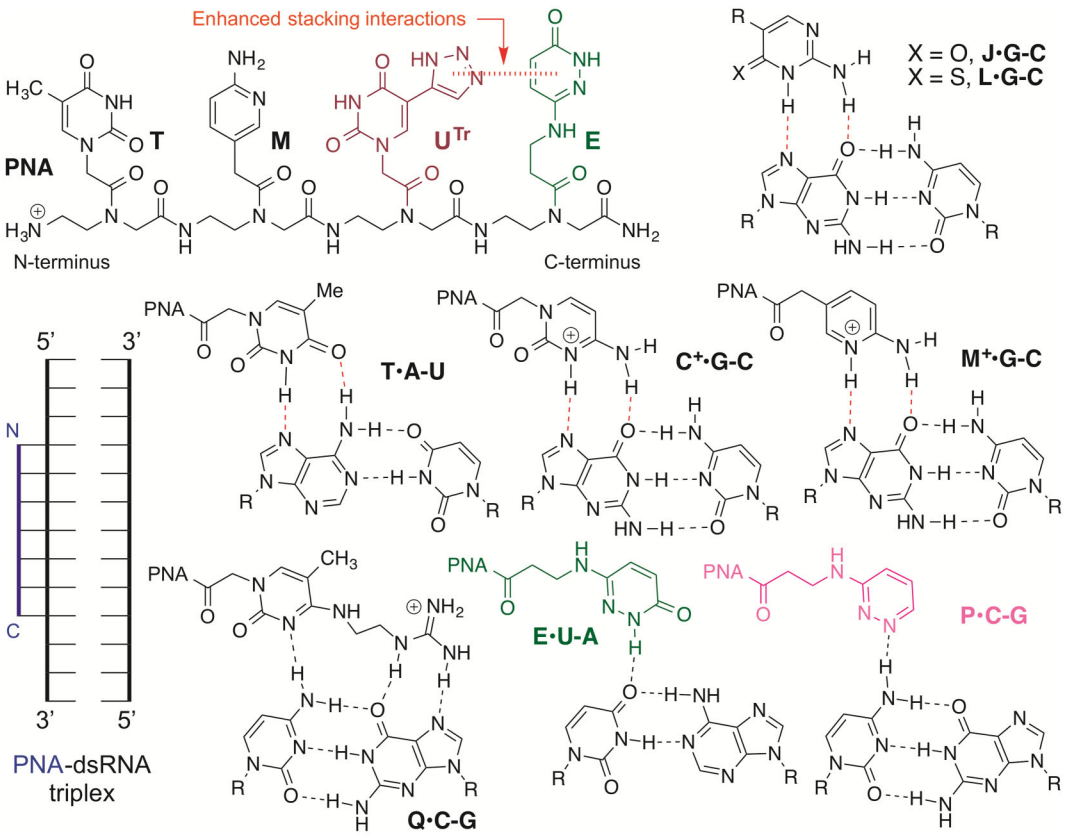


Figure 1. Structures of triplex-forming PNA and Hoogsteen hydrogen-bonded base triples.

RNA has the potential to modulate biological processes in both the laboratory and the clinic. However, the discovery of ligands that bind structured RNA with high affinity and specificity has been a long-standing and formidable problem [8–11]. Small-molecule ligands are being actively explored, with notable recent successes [12–15]. In the present study, we report advancement of an alternative approach to RNA recognition and functional modulation using triplex-forming peptide nucleic acids (PNAs).

PNAs (Fig. 1), first developed for triple helical recognition of DNA in 1991 [16], have recently emerged as promising ligands for binding to double-stranded RNA (dsRNA) [17–19]. Triple helical recognition of dsRNA is challenging because of electrostatic repulsion caused by phosphates aligning the deep and narrow major groove of the RNA helix. For example, homopyrimidine DNA does not form a triplex with the homopurine strand of dsRNA [20]. Therefore, triplex-forming PNAs that sequence-specifically recognize folded RNA could find broad applications in biotechnology and medicine [17–19]. For example, microRNAs are well-recognized regulators of gene expression implicated in many diseases. MicroRNAs are transcribed and processed as short double-helical structures well suited for recognition by triplex-forming PNAs [4]. Several research groups, including ours, have developed modified nucleobases to optimize the stability and sequence specificity of PNA-dsRNA triplexes [19]. Significant progress was achieved by the replacement of natural cytosine, which requires unfavorable protonation to form C^+•G-C triple, with thio-pseudocytosine [21] (L, Fig. 1), an analog of J base originally reported by Nielsen and co-workers for PNA-dsDNA triplexes [22], and 2-aminopyridine [23, 24] (M, Fig. 1). The L and M nucleobases formed highly stable L•G-C and M^+•G-C triples, enabling PNA-dsRNA triplex formation under physiological conditions [25]. Moreover, the M-modified PNAs were binding at least 10-fold stronger to the target dsRNA than to dsDNA of the same sequence [23, 24]. Structural studies showed that the RNA preference was primarily due to the hydrogen bonding between backbone amides in PNA and phosphates in RNA, made possible by matching distances in the PNA-dsRNA triplex [26].

Several follow-up studies have demonstrated the biological activity of L- and M-modified triplex-forming PNAs [27–30]; however, the limitation on what sequences can be targeted remains a major unsolved problem for triple helical recognition of nucleic acids. Triple helical binding to mixed purine-pyrimidine sequences in dsDNA or dsRNA remains a highly attractive but elusive goal because only purines form stable Hoogsteen hydrogen-bonded triples, such as T•A-U and M^+•G-C (Fig. 1) [31]. Designing modified nucleobases X and Y that would form stable X•C-G and Y•U-A triples is complicated due to the limited ability of pyrimidines to form Hoogsteen hydrogen bonds, i.e. cytosine donates only one, and uridine accepts only one hydrogen bond [17, 32]. Furthermore, the lack of symmetry of the major groove of RNA, where pyrimidines occupy more space than purines, aggravates the problem. Thus, a single isolated pyrimidine interruption of the polypurine tract can be recognized with some loss of binding affinity, but the presence of two or more pyrimidines strongly destabilizes the complex.

Despite the challenges, designing novel PNA nucleobases X and Y to improve pyrimidine recognition has been an active area of research [17–19]. For recognition of thymidine in dsDNA, Nielsen and co-workers introduced 3-oxo-2,3-dihydropyridazine (E, Fig. 1) [33]. The E base features an

extended four-atom (instead of the common two-atom) linker connecting the heterocycle to the PNA backbone. The longer linker was designed to circumvent the steric hindrance of the 5-methyl group of thymidine. The longer linker was also beneficial for optimal hydrogen bonding of E to U-A base pairs in RNA [34, 35]. For recognition of cytosine in dsRNA, Chen and co-workers [36] modified PNAs with guanidinylethyl-5-methylcytosine [37] (Q, Fig. 1) that formed Q•C-G triples with reduced stability but good sequence specificity compared to standard Hoogsteen triples. Our research group designed 2-guanidyl pyridine nucleobase (V) that recognized the entire Hoogsteen face of the C-G base pair using three hydrogen bonds (similar to Q) with good affinity [38]. However, we discovered that PNAs modified with three V bases engaged dsRNA in a sequence non-specific binding, apparently caused by the positive charges of guanidine groups. In a series of related studies [34, 35, 39], we screened a variety of simple nitrogen heterocycles and optimized their linkers to the PNA backbone. Of these studies, 3-aminopyridazine (P, Fig. 1) emerged as our currently most promising lead heterocycle for the recognition of cytosine [34]. Similar to E, P has an extended four-atom linker.

While evaluating the binding of E- and P-modified PNAs to dsRNA, we performed molecular dynamics (MD) studies on the PNA-dsRNA triplexes [34]. We observed that the longer linkers in E and P positioned the modified nucleobases favorably for hydrogen bonding; however, the favorable hydrogen bonding required a displacement of the nucleobases towards the major groove. This movement pushed E and P out of the helical stack, interrupting favorable pi-pi interactions and stacking, which decreased the stability of the E•U-A and P•C-G triples [34]. These observations led to a hypothesis that triple helical recognition of pyrimidines could be improved by restoring the stacking interactions lost by displacement of E and P deeper in the major groove.

In the present study, we demonstrated that the 5-triazolyl uracil (U^{Tr}) nucleobase adjacent to E or P improved the binding affinity using a synergetic enhancement of stacking interactions between the triazole ring and E or P heterocycles. MD simulations confirmed that the triazolyl group restored the stacking interactions lost due to the displacement of E and P by the longer linkers. A cell-based dual fluorescent protein reporter assay demonstrated that U^{Tr} modification enhanced the biological activity (inhibition of microRNA maturation) of triplex-forming PNAs. Focused RNA-seq analysis of mature microRNAs showed that an E- and U^{Tr}-modified PNA targeting an 8-nucleotide-long sequence of six purines interrupted by two uridines caused less perturbation in the microRNA expression levels and cellular cytotoxicity than a similar sequence targeting an uninterrupted polypurine tract. Our results demonstrate that modification of PNA nucleobases with extended heterocyclic systems is a new approach to improve pyrimidine recognition that has been a long-standing bottleneck for Hoogsteen triple helical targeting of nucleic acids. Moreover, targeting mixed pyrimidine-purine RNA sequences enhances the specificity of triplex-forming PNAs.

Materials and methods

Materials

Chemicals and reagents were obtained from commercial suppliers and utilized as received unless otherwise noted. Liquid chromatography-mass spectrometry (LC/MS) analysis was performed on a Shimadzu LCMS 2020 system equipped with an electrospray ionization (ESI) source. The RNA hairpins were acquired from Horizon Discovery Biosciences Ltd. (Dharmacon™), deprotected according to the manufacturer's instructions, and purified using reverse-phase high-performance liquid chromatography (HPLC) on a Shimadzu LC-40 system using an acetonitrile gradient in a 50 mM triethyl ammonium acetate buffer (pH 6.4). The purity of the RNA was verified by HPLC analysis. The RNA was quantified by measuring the absorbance at 260 nm. Finally, the RNA was lyophilized and reconstituted in nuclease-free water to a final concentration of 240 μM.

Synthesis of PNA

PNAs were synthesized on a 2 μmol scale using NovaSyn TG Sieber resin (Novabiochem) or TentaGel XV RAM (Rapp Polymere GmbH) on an Expedite 8909 synthesizer, following previously reported procedures [34, 40]. The commercial PNA-T monomer was sourced from Link Technologies, while the M [41], P [34], E [33], and U^{Tr} [40] monomers were synthesized following established literature methods. The PNAs were cleaved from the solid support with 0.6 mL of TFA over a 2-h period using a two-syringe pull-push technique. For PNAs having U^{Tr} in their sequences, the crude PNA in 0.1% aqueous formic acid solution was lyophilized and treated with 1 mL of ammonium hydroxide in a 1.5 mL Eppendorf tube for 3 h at room temperature. The crude PNA was distributed into three Eppendorf tubes (200 μL each), precipitated with chilled diethyl ether (∼1.0 mL), and centrifuged at 15 000 rpm for 20 min. After the diethyl ether was removed, the PNA was resuspended in 1.0 mL of diethyl ether and centrifuged again at 15 000 rpm for 10 min. Following the removal of the diethyl ether, the white solid was dissolved in 1.0 mL of 0.1% formic acid in water. The samples were analyzed by LC/MS to identify the target compound. The target PNA sequences were isolated from the crude mixtures using HPLC purification. Details of LC/MS and HPLC conditions, columns, and gradients are listed in Supplementary Information (Supplementary Table S1 and Figs S1–S23). After quantification using UV absorbance at 260 nm, the solutions were lyophilized and reconstituted in nuclease-free deionized water to a final concentration of 240 μM.

UV thermal melting

UV melting experiments were conducted using Shimadzu UV-2600 and UV-1800 spectrophotometers equipped with TMSPC-8 temperature controllers. The temperature was increased at 1.0°C per min. Absorbance measurements at 300 nm were recorded to analyze the formation of PNA-based triplex structures. The experiments were carried out in a buffer containing 50 mM potassium phosphate, 2 mM $MgCl_2$, 90 mM KCl, and 10 mM NaCl, pH 7.4. To prepare the sample, 17.3 μL of each PNA and RNA solution (240 μM stock concentration) were freeze-dried separately. The solid PNA was dissolved in 230 μL of buffer to reach 18 μM and subsequently mixed with the freeze-dried RNA. The resulting mixture was heated to 95°C for 10 min and then rapidly cooled to 4°C. The samples were measured through three melting cycles, ranging from 10°C to 95°C. Data from five replicates were used to determine the average melting temperature (T_m) and its standard deviation (Supplementary Tables S2–S11).

Isothermal titration calorimetry (ITC)

ITC experiments were performed using a MicroCal ITC200 calorimeter at 25°C. The experiments were carried out in a buffer containing 50 mM potassium phosphate, 2 mM $MgCl_2$, 90 mM KCl, and 10 mM NaCl, adjusted to pH 7.4. In each run, 2.45 μL of PNA solution (80 μM) was injected sequentially into 270 μL of RNA hairpin solution (10 μM) using a rotating syringe at 750 rpm. Data were analyzed using the MicroCal PEAQ-ITC software (Figs S24–S66 and Supplementary Table S12).

Molecular dynamics (MD)

The MD simulations were performed using Maestro Version 13.4.134, Release 2022–4 (Schrödinger, LLC). The initial PNA-dsRNA triple helix structure was derived from a previously validated model [34]. RNA sequences were modified using Maestro's built-in mutate residue tool, while PNA nucleobases were adjusted manually using default fragments provided by the software. Each modified triple helix was energy-minimized with MacroModel employing the OPLS4 force field and the Truncated Newton Conjugate Gradient (TNCG) method, achieving a convergence threshold of 0.05 kcal/mol/Å for the gradient. For the MD simulations, the OPLS4 force field was employed. The system was solvated using the TIP3P water model within a cubic box extending 10 Å beyond the solute in all directions. To simulate physiological conditions, the system was neutralized with Na^+ ions and supplemented with 0.15 M NaCl, mimicking a biologically relevant ionic environment. The relaxation protocol consisted of five stages, all conducted at 10 K:

1. Brownian Dynamics in the NVT ensemble for 100 ps,
2. NVT simulation for 12 ps,
3. NPT simulation for 12 ps,
4. NPT simulation with solute restraints for 12 ps,
5. NPT simulation without restraints for 24 ps.

The production MD simulation was performed in the NPT ensemble at 300 K and 1 atm for a total of 150 ns, using a 2 fs timestep. The final 50 ns, representing the equilibrated phase, were used for analysis. Structural images were generated from Desmond trajectory clustering, specifically using RMSD-based clustering of the equilibrated phase. The images presented depict the most abundant conformations. Relative binding energies (ΔG) were calculated using the Molecular Mechanics Generalized Born Surface Area (MM/GBSA) method, implemented via the *thermal_mmgbsa.py* script provided by Schrödinger Prime. The binding free energy is defined as:

$$\Delta G(\text{bind}) = \Delta G(\text{solv}) + \Delta E(\text{MM}) + \Delta G(\text{SA})$$

where:

- ΔG(solv) is the difference in GBSA solvation energy between the RNA-PNA complex and the sum of the solvation energies of the RNA and PNA.
- ΔE(MM) is the difference in minimized energies between the RNA-PNA complex and the sum of the energies of the RNA and PNA.
- ΔG(SA) is the difference in surface area energies between the complex and the sum of the surface area energies of the RNA and PNA.

Prime MM-GBSA calculates the energy of optimized free receptors, free ligand, and a complex of the ligand with a receptor. It also calculates the ligand strain energy by placing the ligand in a solution autogenerated by the VSGB 2.0 suite.

Construction of reporter plasmid

A plasmid vector encoding AcGFP with a complementary sequence of active miR-155 (miR-155–5p) or miR-27a (miR-27a-3p) in the 5′-untranslated region (UTR) under control of a cytomegalovirus (CMV) promoter was constructed based on the pAcGFP-N3 vector (Clontech). The DNA fragments to be targeted by the miR-155–5p and miR-27a-3p were prepared by annealing the miR-155-sense (5′-CTAGCATCCAACCCCTATCACGATTAGCATTAA AGAACTAGTGG-3′) and miR-155-antisense (5′-GATCCCACTAGTTCTTTAATGCTAATCGTGATA GGGGTTGGATG -3′), and miR-27a-sense (5′-CTAGCAT CCGCGGAACTTAGCCACTGTGAAAGAACTAGTGG-3′) and miR-27a-antisense (5′-GATCCCACTAGTTCTTTCA CAGTGGCTAAGTTCCGCGGATG-3′) oligonucleotides, respectively. The DNA fragments were cloned into the NheI and BamHI sites of pAcGFP-N3 to make pCMV-UTR-miR155-5p-AcGFP and pCMV-UTR-miR-27a-3p-AcGFP.

Preparation of pre-miRNAs

Pre-miR-155 and pre-miR-27a were prepared by *in vitro* transcription. The template DNA fragments were prepared by primer extension using the sense (5′-CTTAATACGACTCACTATAGGCTGTTAATGCTA ATCGTGATAGGGGTTTTTGCCTCCAACTG-3′) and antisense (5′-CTGTTAATGCTAATATGTAGGAGTCAGTTGG AGGCAAAAACCCC-3′) primers for pre-miR-155, and sense (5′-CTTAATACGACTCACTATAGGCTGAGGAGCAGG GCTTAGCTGCTTGTGAGCAGGGTCCACACCAAGTCG TG-3′) and antisense (5′-CTGGGGGGCGGAACTTAGCCA CTGTGAACACGACTTGGTGTGGACCCTG-3′) primers for pre-miR-27a, respectively, and were purified using the QIAquick PCR Purification Kit (Qiagen). Pre-miRNAs were transcribed from the template DNA using the ScriptMAX Thermo T7 Transcription Kit (Toyobo) and purified using a denaturing polyacrylamide gel. The RNA concentration was determined by measuring the absorbance at 260 nm at 90°C using a Shimadzu UV-1700 spectrophotometer connected to a thermo-programmer.

Electroporation of PNAs into cells

Human neuroblastoma (SH-SY5Y) and breast cancer (MCF7) cell lines were cultured in Dulbecco's modified Eagle's medium (DMEM) and Eagle's minimum essential medium (EMEM) with non-essential amino acids (NEAA), respectively, supplemented with 10% fetal bovine serum and antibiotics (100 U/mL penicillin and 100 μg/mL streptomycin), and maintained at 37°C under 5% CO_2. Cells were trypsinized and counted by Countess II (Invitrogen). PNA (400 pmol) was mixed with SH-SY5Y (5×10^5 cells) or MCF7 (4×10^5 cells) in a solution of 4D-Nucleofector X Kit S (Lonza) and introduced into the cells using 4D-nucleofector (Lonza) with the optimal program for each cell line provided by the manufacturer. After the electroporation, cells were replated on a collagen-coated 96-well plate (Iwaki, Thermo Scientific, or Corning). Each well contained 1/20 cells used for the electroporation. To purify total RNA, 1/2 of the cells used for the electroporation were replated in each of the two wells of a collagen-coated 24-well plate (Iwaki).

Dual fluorescent protein reporter assay

After electroporation of PNAs, cells were incubated for 18 h at 37°C under 5% CO_2. pCMV-UTR-miR155-5p-AcGFP or pCMV-UTR-miR27a-3p-AcGFP reporter plasmid (7.5 ng) and a pDsRed-Express2-N1 (Clontech) control plasmid (7.5 ng) were co-transfected with or without purified pre-miRNA (0.25 pmol) using 0.0625 μL of Lipofectamine 2000 (Invitrogen). After continuous incubation at 37°C under 5% CO_2 for 48 h, the cellular media were exchanged to a buffer containing 20 mM HEPES–KOH (pH 7.4), 115 mM NaCl, 5.4 mM KCl, 1.8 mM $CaCl_2$, 0.8 mM $MgCl_2$, and 13.8 mM glucose. Expression levels of AcGFP and DsRed-Express2 were evaluated by measuring fluorescence intensities of cells using a microwell plate reader (Infinite 200 Pro; Tecan). Excitation and emission wavelengths for measuring the fluorescence intensities were 488 and 530 nm for AcGFP and 560 and 600 nm for DsRed-Express2, respectively, and fluorescence intensities at five different positions were measured in one well. Averaged fluorescence intensities obtained from three wells, after subtraction of intensities from wells without cells, were used to analyze the expression levels of reporter proteins. Reproducibility of the experiment was evaluated by repeating a series of processes starting with electroporation. After the measurement, AcGFP and DsRed-Express2 fluorescence in cells were imaged using a spinning disk field scanning confocal microscopy (CSU-W1, Yokogawa). Wavelengths at 488 and 561 nm lasers and 530 and 600 nm emission filters were used for imaging AcGFP and DsRed-Express2, respectively. Cellular fluorescence was imaged using a 20 × objective lens, and merged images of the reporter proteins were obtained by NIS-Elements software (Nikon).

RNA-seq analysis targeting mature miRNAs

After electroporation of PNAs, cells in a 24-well plate were continuously cultured for 66 h at 37°C under 5% CO_2. Cellular total RNA was purified using the miRNeasy Mini Kit (Qiagen) and a robotic workstation for automated purification (QIAcube, Qiagen) according to the standard program without optional DNase treatment. Concentration of the purified RNA was measured by NanoDrop 1000 (Thermo Fisher Scientific). A library for RNA-seq analysis was prepared from 100 ng total RNA using QIAseq miRNA Library Kit for Thermo Fisher Scientific NGS Systems (Qiagen) according to the manufacturer's protocol. QIAseq miRNA 12 Index TF (Qiagen) was used to distinguish samples. Quality of the libraries and distributions of the library size were confirmed by using a microchip electrophoresis system (MultiNA, Shimadzu) (Supplementary Fig. S70). Concentrations of the libraries were quantified using a Qubit 4 Fluorometer (Thermo Fisher Scientific) (Supplementary Table S13). NGS analysis of the mixed library was performed using the Ion GeneStudio S5 system with the Ion 540 chip (Thermo Fisher Scientific) according to the manufacturer's protocol. The Ion Chef System (Thermo Fisher Scientific) was used for automated template preparation and chip loading. From NGS data, expression levels of miRNAs were analyzed using CLC Genomics Workbench (Qiagen) according to workflows provided by the manufacturer using mirBase_v22 as reference miRNA data. At least 2 million reads were annotated to the reference data in all samples. A total of 421 mature miRNAs (listed in Supplementary Table S14), in which expression levels counted as counts per million reads (CPM) were more than 4 in control cells without PNA treatment, were analyzed to quantify the Log [2] fold change caused by the PNA treatment. Statistical analyses were performed using NGS data obtained from triplicate experiments.

Cell proliferation assay

After 18 h of incubation from electroporation of PNAs, the cellular media in the 96-well plate were replaced with fresh media and further incubated for 48 h at 37°C under 5% CO_2. The number of cells in the 96-well plate was evaluated by Cell Counting Kit-8 (Dojindo). Briefly, the cellular media was replaced with that containing 10 μL/well substrate and incubated at 37°C. The 450 nm absorbance of each well was measured after 60-min incubation using Varioskan Flash (Thermo Fisher Scientific), and the values were normalized by subtracting those of the wells without cells. Cellular proliferation was evaluated by comparing the values with electroporated cells without PNA, for which the value was set as 100%.

Results

To test our hypothesis that restoring the stacking interactions will improve the triple helical recognition of pyrimidines, we used a dsRNA model system of HRP1-HRP4 (Fig. 2A). We previously used these hairpins, having a variable base pair in the middle of the stem (color-coded in Fig. 2A), to study the binding affinity and sequence specificity of nucleobase-modified PNAs [24, 34, 38, 39]. The U^{Tr} nucleobase (Fig. 1) was originally synthesized and incorporated into a PNA as a control to gauge the performance of extended nucleobases recognizing the entire Hoogsteen face of RNA base pairs [40]. The binding affinity and sequence specificity of modified PNAs were studied using UV thermal melting and isothermal titration calorimetry (ITC), as in our previous studies (Supplementary Tables S2–S12, and Supplementary Figs S24–S66) [24, 34, 38, 39]. We also used MD simulations to study changes in binding energy and stacking interactions upon U^{Tr} modification.

We started with testing the impact of triazole addition on PNA binding affinity and specificity. Replacement of a single T with U^{Tr} caused little change in the on-target affinity of PNA2-PNA4 (Table 1). Within the range of error, both K_a and T_m values for the triazole-modified PNAs were similar or slightly higher than those for PNA1. The triazole modification caused a small increase in the stability of the mismatched triplexes, most notable for the off-target binding to HRP1. Thus, isolated triazole modifications had a relatively small effect on binding affinity and specificity. Next, we studied the effect of placing U^{Tr} adjacent to E•U-A and P•C-G triples. The sequence of HRP1-HRP4 allows placement of U^{Tr} at the -$CONH_2$ side to the variable base (shown as "above" in Fig. 2A) of the PNAs. MD studies showed that displacement of E and P towards the major groove resulted in a loss of stacking (lack of blue line above P in Fig. 2B) and positioned the modified nucleobase directly "below," referring to the perspective shown in Fig. 2B, the 5-methyl group of thymidine.

This observation led to our hypothesis that replacing the methyl group with triazole would restore the lost stacking interactions. Biophysical studies in our model hairpin system confirmed the hypothesis. U^{Tr} modification increased the binding affinity of PNA6 and PNA8 (Table 1). While the increase in T_m was relatively modest (~2 and 5°C, respectively), ITC

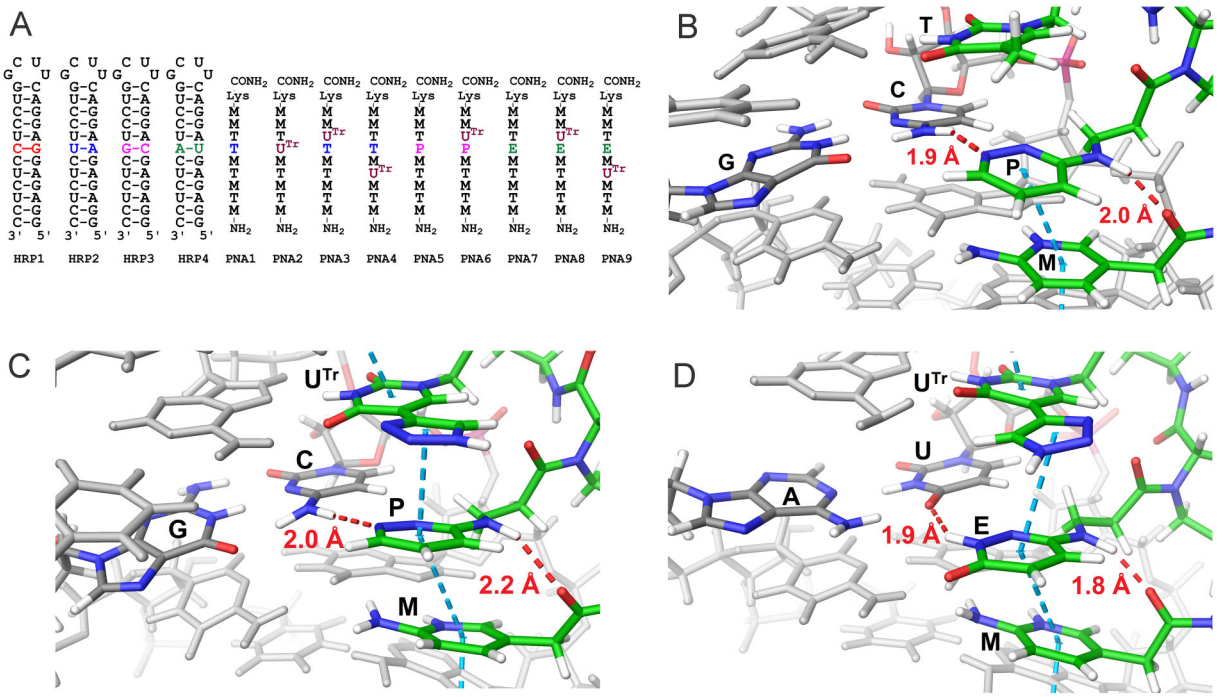


Figure 2. (**A**) Sequences of model RNA hairpins (HRP1-HRP4) and triplex-forming PNAs. (**B**-**D**) Snapshots of MD of triplexes formed by (**B**) PNA5-HRP3, focusing on the P•C-G triple. (**C**) PNA6-HRP3, focusing on the P•C-G triple. (**D**) PNA8-HRP4, focusing on the E•U-A triple. Dashed red lines show hydrogen bonding, dashed blue lines show pi-pi stacking.

Table 1. Affinity and sequence specificity of nucleobase-modified PNAs binding to HRP1-HRP4

PNA	Sequence (NH_2- to -$CONH_2$)	Affinity [a]	HRP1(C-G)	HRP2(U-A)	HRP3(G-C)	HRP4(A-U)	ΔG [b] kcal/mol	pi-pi contacts [c]
PNA1	MTMTMTTMM	$K_a \times 10^6$ M^{-1}	1.9 ± 0.1	**12 ± 1**	0.9 ± 0.1	0.4 ± 0.1	-198	**1.49(5.96)**
		T_m°C	46.4 ± 0.5	**65.9 ± 0.3**	35.4 ± 0.4	34.6 ± 0.2		
PNA2	MTMTMU^{Tr}TMM	$K_a \times 10^6$ M^{-1}	2.9 ± 0.4	**14 ± 1**	1.2 ± 0.1	1.0 ± 0.1	-191	**1.95(6.65)**
		T_m°C	56.7 ± 0.4	**65.0 ± 0.6**	39.7 ± 0.3	37.4 ± 0.3		
PNA3	MTMTMTU^{Tr}MM	$K_a \times 10^6$ M^{-1}	3.8 ± 0.5	**14 ± 1**	2.1 ± 0.1	1.5 ± 0.1	-209	**1.69(6.24)**
		T_m°C	55.1 ± 0.5	**68.3 ± 0.6**	38.7 ± 0.4	36.2 ± 0.3		
PNA4	MTMU^{Tr}MTTMM	$K_a \times 10^6$ M^{-1}	2.5 ± 0.3	**11 ± 1**	0.4 ± 0.3	0.5 ± 0.3	-169	**1.45(6.07)**
		T_m°C	48.6 ± 0.5	**65.8 ± 0.4**	34.9 ± 0.4	32.8 ± 0.5		
PNA5	MTMTMPTMM	$K_a \times 10^6$ M^{-1}	0.9 ± 0.2	0.5 ± 0.1	**5.4 ± 0.1**	0.4 ± 0.2	-198	**0.95 (5.37)**
		T_m°C	32.4 ± 0.4	41.5 ± 0.4	**51.7 ± 0.7**	36.4 ± 0.2		
PNA6	MTMTMPU^{Tr}MM	$K_a \times 10^6$ M^{-1}	4.2 ± 0.1	4.2 ± 0.1	**23 ± 1**	3.2 ± 0.1	-218	**1.73 (6.04)**
		T_m°C	43.0 ± 0.4	43.0 ± 1.0	**54.1 ± 0.8**	42.4 ± 0.6		
PNA7	MTMTMETMM	$K_a \times 10^6$ M^{-1}	2.2 ± 0.2	4.2 ± 0.7	6.6 ± 0.2	**11 ± 1**	-194	**0.95 (5.75)**
		T_m°C	33.7 ± 0.5	49.6 ± 0.3	49.0 ± 0.4	**51.9 ± 0.5**		
PNA8	MTMTMEU^{Tr}MM	$K_a \times 10^6$ M^{-1}	2.2 ± 0.1	5.1 ± 0.1	5.2 ± 0.4	**49 ± 2**	-219	**1.82 (6.60)**
		T_m°C	43.6 ± 0.5	48.1 ± 0.5	52.0 ± 0.6	**56.6 ± 0.5**		
PNA9	MTMU^{Tr}METMM	$K_a \times 10^6$ M^{-1}	1.9 ± 0.1	4.3 ± 0.6	2.5 ± 0.6	**8.4 ± 0.4**	ND [d]	ND [d]
		T_m°C	32.1 ± 0.7	44.8 ± 1.0	44.2 ± 0.5	**50.0 ± 1.0**		

[a] Association constants, $K_a \times 10^6$ M^{-1}, average of three experiments ± S.D., for binding of PNAs to the respective hairpins in 50 mM potassium phosphate buffer (pH 7.4) containing 2 mM $MgCl_2$, 90 mM KCl, 10 mM NaCl at 25°C. UV thermal melting temperatures, T_m°C, an average of five experiments ± S.D. measured at 300 nm and 18 μM of each dsRNA and PNA in the ITC buffer as above. [b] MM-GBSA free energies of binding (ΔG, kcal/mol) of the matched triplexes calculated from 150 ns molecular dynamics. [c] Average number of pi-pi contacts around the variable base and, in parentheses, around the entire PNA strand. [d] ND – not determined.

showed a remarkable more than four-fold increase in K_a. The improvement of binding affinity was due to synergistic interactions between E or P and U^{Tr}, because PNA9, having U^{Tr} and E separated by an M nucleobase, did not have enhanced binding affinity compared to PNA7. As with PNA1-PNA4, triazole substitution led to a small decrease in sequence specificity of P-modified PNA6, most notable for binding to HRP1. However, PNA6 still maintained good specificity, with all off-target triplexes having more than 10°C and five-fold lower T_m and K_a, respectively. In contrast, the E-modified PNA7 had a relatively modest specificity that was somewhat improved by triazole modification in PNA8.

MD simulations showed that in the U^{Tr}-modified triplexes, P•C-G and E•U-A triples maintained the expected hydrogen bonding network (Fig. 2C and D) observed in our previous study [34]. Both P and E formed one hydrogen bond to C and U, respectively, and the N-H of the linker hydrogen bonded to the C = O of the linker connecting adjacent M to the PNA backbone. Consistent with the improved affinity, MD simulations showed the expected stacking interactions between triazole and P and E nucleobases (blue lines in Fig. 2C and D), which resulted in an overall increase in the number of average pi-pi contacts from 0.95 to 1.73 and 1.82 (Table 1 rightmost column) for the P and E nucleobases, respectively. Calculated free energies of binding (ΔG, Table 1) agreed well with the improved stacking and affinity. Collectively, biophysical and MD results validated our hypothesis and encouraged further studies on the synergy between adjacent U^{Tr} and E or P.

Next, we tested the synergistic binding affinity improvements in a sequence context that allows placing U^{Tr} on both sides of P and E. We used HRP5 and HRP6, variants of our model hairpins, where the variable base pair was shifted down the stem and placed between two U-A base pairs (Fig. 3A). Consistent with results observed with model hairpins HRP1-HRP4, placing U^{Tr} either on the amino ("below" in Fig. 3A) or carboxy ("above") side of P and E notably increased the affinity of PNAs binding to HRP5 and HRP6 (Table 2). The affinity increase was similar for both positions and correlated well with the calculated ΔG values. Placing U^{Tr} at both sides of P and E further improved the affinity. MD simulations showed significant enhancement of pi–pi stacking interactions (blue lines in Fig. 3B and C). The number of pi-pi contacts (Table 2) increased from ~0.9 to ~2.4 for both P and E nucleobases. However, in the case of double modifications, the calculated ΔG values did not correlate with the observed affinity changes. At this point, we do not have a good explanation for the mismatched trend in ΔG of doubly modified PNAs. Overall, the biophysical and MD results with HRP5 and HRP6 supported our hypothesis of increased stacking interactions by U^{Tr} modification.

Next, we tested whether the increased stacking interactions may improve the biological activity of PNAs. Because triplex-forming PNAs bind about 10-fold stronger to dsRNA than to dsDNA [23, 24], they are excellent ligands for targeting biological RNAs, such as microRNAs. Their unique RNA selectivity reduces potential off-target binding to genomic DNA. Previously, we have demonstrated that the binding of triplex-forming PNAs to microRNA precursor (pre-miR) hairpins inhibited maturation of miR-197 in human neuroblastoma (SH-SY5Y) cells [29]. In the present study, we selected miR-155 as the target to demonstrate the advantages of U^{Tr}-modified PNAs. MiR-155 is a multifunctional microRNA involved in the regulation of genes involved in immune response, inflammation, and cancer development, such as *SHIP-1*, *VHL*, *SOCS-1*, *BCL6*, and *TAB2* [42, 43]. The biological role of miR-155 varies across different cell and tissue types, developmental stages, and diseases.

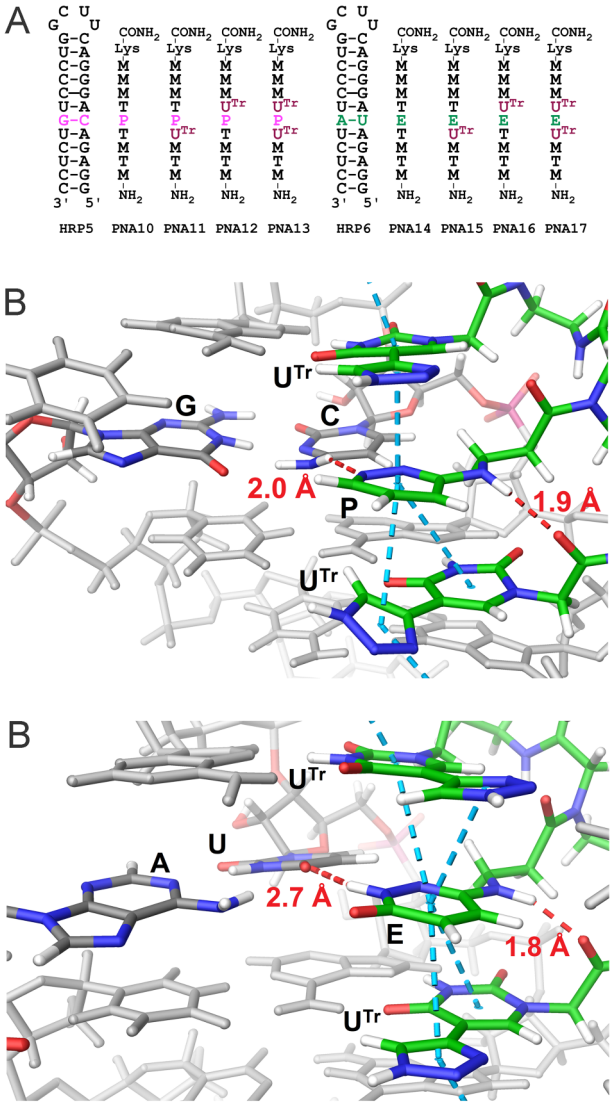


Figure 3. (**A**) Sequences of HRP5 and HRP6 and the matched triplex-forming PNAs allowing placement of U^{Tr} on both sides of P and E. (**B** and **C**) Snapshots of MD of triplexes formed by (**A**) PNA13-HRP5 and (**B**) PNA17-HRP6, focusing on the P•C-G and E•U-A triples.

As the target site of PNA, we chose a purine-rich sequence interrupted by two uridines (bold in Fig. 4A), having a relatively balanced distribution of three nucleotides (four Gs, two As, and two Us). The miR-155 target sequence featured a bulged A. Our recent study showed that the bulged A could be efficiently recognized by a T in PNA, presumably through a Hoogsteen A-T base pair [44]. This PNA sequence allowed the placement of U^{Tr} on both sides of the central E (PNA19-21, in Fig. 4A). PNA22 was a scrambled control sequence. We expected that the unmodified PNA18 would have modest activity because the two uridines interrupting the purine stretch represent a relatively poor binding site for Hoogsteen triplex formation. We hypothesized that the U^{Tr} modifications would enhance the activity of PNA19-21.

Table 2. Affinity of PNAs binding to HRP5 and HRP6.

PNA [a]	$K_a \times 10^6$ M^{-1}	T_m °C	ΔG kcal/mol	pi-pi contacts [b]
PNA10	5.0 ± 0.2	42.6 ± 0.2	-167	0.91 (4.09)
PNA11	18 ± 1	50.4 ± 1.0	-178	1.52 (6.04)
PNA12	13 ± 1	49.9 ± 0.4	-182	1.74 (5.42)
PNA13	29 ± 3	53.2 ± 0.7	-157	2.36 (6.08)
PNA14	5.9 ± 0.3	51.7 ± 0.6	-177	0.94 (4.49)
PNA15	7.5 ± 0.4	54.9 ± 0.6	-186	1.50 (6.25)
PNA16	11 ± 1	54.2 ± 0.8	-199	1.72 (6.48)
PNA17	18 ± 1	57.9 ± 0.6	-179	2.45 (6.58)

[a]For experimental details, see footnotes in Table 1.
[b]Average number of pi-pi contacts around the variable base and, in parentheses, around the entire PNA strand.

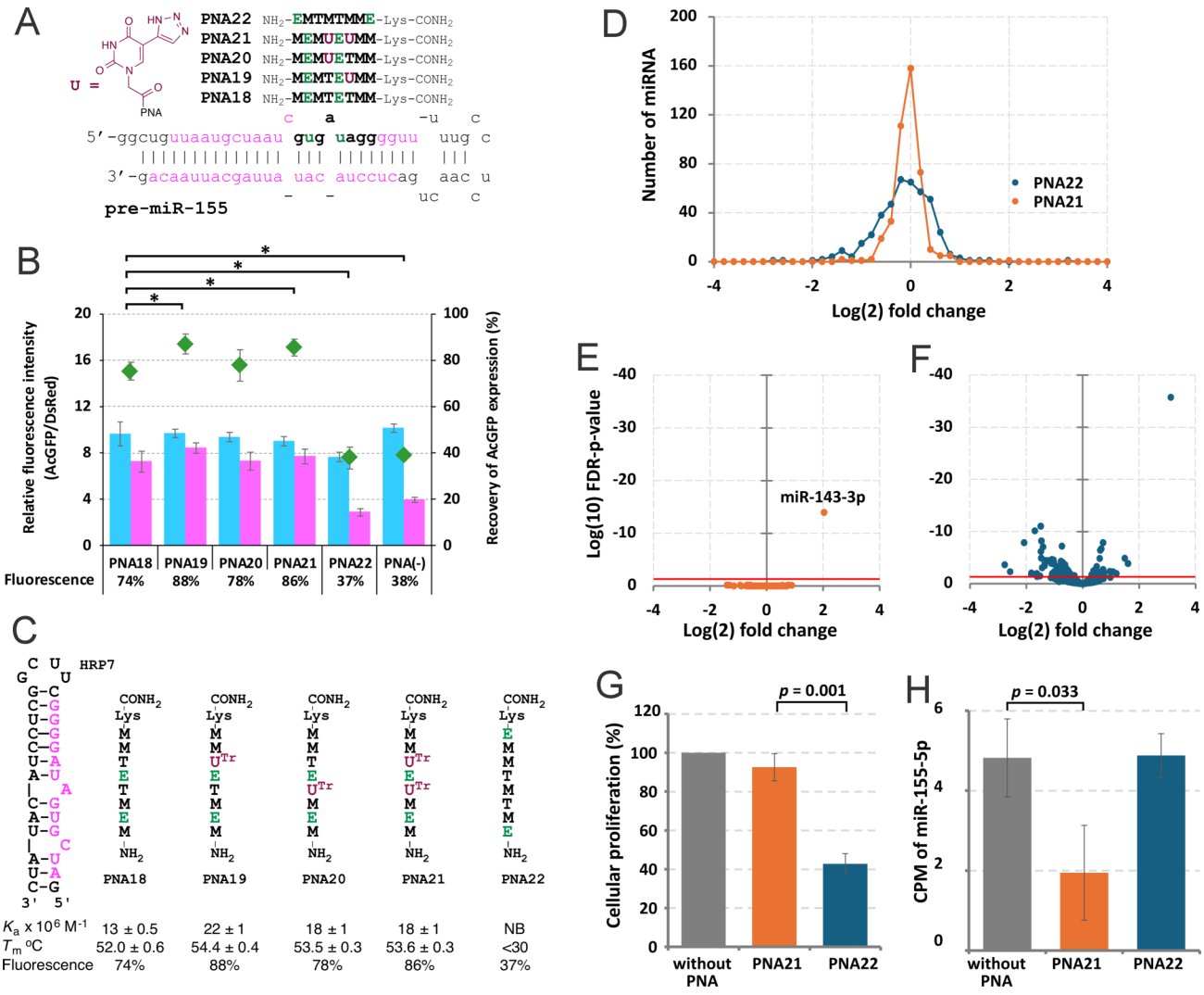


Figure 4. (**A**) Sequences of triplex-forming PNAs and pre-miR-155 hairpin. The mature microRNA is indicated in pink, and the PNA binding site is highlighted in bold. **(B)** Fluorescence intensity of AcGFP relative to DsRed-Express2 (AcGFP/DsRed, left axis). Reporter plasmids were co-transfected into cells without (blue) or with (pink) pre-miR-155. Recovery of fluorescence (AcGFP expression) in the presence of triplex-forming PNA18-PNA22 (green diamond, right axis) after normalization of the AcGFP/DsRed values by the cells co-transfected without pre-miR-155. ∗: $p < 0.02$. (**C**)Binding affinity of PNA18-22 to a short model RNA hairpin featuring the sequence of miR-155 (highlighted in pink) targeted by the PNAs. (**D**) Histogram of Log[2] fold changes in microRNA expression caused by PNA21 and PNA22. (**E** and **F**) Volcano plots comparing Log[2] fold change and Log[10] FDR-*p*-values of each microRNA in cells treated with PNA21 (**E**) and PNA22 (**F**). The red line indicates FDR-*p*-value = 0.05. (**G**) Changes in normalized cellular proliferation. (**H**) Averaged CPM values of miR-155–5p. In **G** and **H**, the values of average ± standard deviation obtained from triplicate experiments are shown. Values of Student's t-test are shown above the graphs.

To test our hypothesis, we used a dual fluorescent protein assay based on a reporter plasmid encoding AcGFP mRNA, which contains a complementary sequence of miR-155 (miR-155–5p) in its 5′ untranslated region (5′-UTR) under the control of a cytomegalovirus promoter (Supplementary Fig. S67). In this assay, suppression of exogenously transfected pre-miR-155 maturation through triplex formation rescued the down-regulation of reporter gene expression (Supplementary Fig. S67). In normal SH-SY5Y cells, miR-155 is expressed at low levels but is upregulated during ischemia, inflammation, or neurodegeneration [45]. The low expression level of miR-155 in normal SH-SY5Y cells, also confirmed by our previous study [29], enabled quantitative evaluation of the rescue activity by co-transfection of the reporter constructs and pre-miR-155 hairpin prepared *in vitro*. We used the dual fluorescent protein assay to test the hypothesis that U^{Tr}-modified PNAs would exhibit higher rescue activity than the non-modified PNA.

PNAs were first electroporated into SH-SY5Y cells, and the cells were incubated for 18 h to allow for full recovery of cell viability and minimize perturbations in the reporter system responses due to stress caused by the electroporation. The AcGFP reporter plasmid and a pDsRed-Express2-N1 (Clontech) control plasmid were co-transfected with or without synthetic RNA hairpin mimicking the native pre-miR-155 (Fig. 4A) using Lipofectamine 2000. After 48 h incubation, the levels of AcGFP expression relative to DsRed-Express2 expression (AcGFP/DsRed) were measured using a microwell plate reader (blue and pink bars in Fig. 4B) and imaged by a spinning disk field scanning confocal microscopy (Supplementary Fig. S68). Recovery of AcGFP expression (green diamonds in Fig. 4B) was calculated by dividing the AcGFP/DsRed values obtained from cells, in which reporter plasmids were co-transfected with pre-miR-155, by those from cells co-transfected without pre-miR-155.

Transfection of reporter plasmids and pre-miR-155 in the absence of PNA or in the presence of the scrambled PNA22 suppressed the AcGFP fluorescence to 38 and 37%, respectively (green diamonds in the two rightmost columns in Fig. 4B). In this assay, the transfected pre-miR-155, in the absence of complementary PNA, is processed by the cellular microRNA machinery into mature miR-155 that suppresses AcGFP expression from the reporter plasmid. Merged images of cellular fluorescence also showed red color (Supplementary Fig. S68). Electroporation of the unmodified PNA18 recovered the AcGFP fluorescence to 74%. This result is consistent with the triplex-forming PNA18 binding to the pre-miR-155 hairpin and inhibiting its maturation. Confirming our hypothesis, placing a single U^{Tr} on the carboxy side of the central E in PNA19 increased the fluorescence recovery to 88%. Interestingly, placing a single U^{Tr} on the amino side of the central E had relatively little effect on the activity of PNA20 (fluorescence recovery to 78%) and did not further increase the activity of the doubly modified PNA21 (fluorescence recovery to 86%). The U^{Tr} on the amino side recognized the single adenosine bulge, and it is conceivable that the U^{Tr}-A base pair had different conformation and stacking interactions than the triples in our model hairpins. Cellular fluorescence also showed recovery of AcGFP fluorescence in these samples (Supplementary Fig. S68). The dual fluorescent protein reporter assay results were consistent with improvements in binding affinity by the U^{Tr} modification in triplexes formed by PNA18-22 and HRP7 having the target sequence of miR-155 (Fig. 4C). To demonstrate the generality of the approach and ability of U^{Tr} modification to achieve similar activity enhancement when placed adjacent to P, we chose an octamer PNA sequence targeting the dicer cleavage site in pre-miR-27a (Supplementary Fig. S69A). In this system, a single U^{Tr} modification placed on the carboxy side of the central P increased the fluorescence recovery from 67% in unmodified PNA23 to 77% in the U^{Tr}-modified PNA24 (Supplementary Fig. S69B). This increase was again consistent with the improved binding affinity of the U^{Tr}-modified PNA24 (Supplementary Fig. S69C).

In our previous study on the inhibition of miR-197 maturation [29], we observed off-target activity of a scrambled PNA that was partially complementary to other microRNAs, particularly pre-miR-27a. We hypothesized that the off-target activity was caused by the low specificity of the "two-letter code" that uses only two Hoogsteen T•A-U and M^{+}•G-C triples. MicroRNA hairpins frequently have short sequences of eight or more consecutive purines interrupted by one or two isolated pyrimidines, as illustrated for pre-miR-155 in Fig. 4A. Because these sequences consist mostly of repeating G and A nucleosides, they are highly redundant. In other words, a triplex-forming PNA built of T and M nucleobases might inherently find many partially complementary off-targets. Consequently, we hypothesized that PNAs targeting mixed pyrimidine-purine sequences would have significantly higher biological specificity.

To test this hypothesis, we performed next-generation sequencing (NGS) to measure changes in mature microRNA levels upon electroporating cells with PNA21 and PNA22. While both PNAs had the same nucleobase content, the two E nucleobases were placed at the ends of PNA22, creating a six-nucleobase-long T/M tract. As discussed above, normal SH-SY5Y cells used in our previous study [29] expressed a low level of miR-155–5p that was not suitable for analysis of endogenous miR-155–5p. To analyze the expression levels of endogenous microRNAs, we chose breast carcinoma (MCF7) cells because the activity of miR-155 has been demonstrated in these cells [46]. To focus on the expression levels of mature microRNAs, a microRNA library kit (Qiagen) was used (Supplementary Fig. S70 and Supplementary Table S13). From the NGS data, changes in expression levels (values of Log[2] fold change) for 421 microRNAs (including miR-155–5p) that had an average intrinsic expression level of counts per million (CPM) more than 4 were analyzed (Supplementary Fig. S71 and Supplementary Table S14).

Consistent with our hypothesis, treatment of MCF7 cells with E- and U^{Tr}-modified PNA21 caused notably less perturbation of microRNA levels than PNA22, as the histogram of Log[2] fold change is sharper for PNA21 than PNA22 (Fig. 4D). When volcano plots of Log[10] FDR-*p*-values of changes in expression level of microRNAs were analyzed, PNA21 did not show significant changes in mature microRNA levels except for miR-143–3p (Fig. 4E). In contrast, PNA22 significantly perturbed the levels of many microRNAs, suggesting off-target activity of PNA22 toward a wide range of microRNAs (Fig. 4F). Treatment with PNA22 also causes a significant decrease, compared to PNA21, of cellular proliferation (Fig. 4G), which was likely the effect of the off-target activity. PNA21 reduced the miR-155–5p level with statistical significance using the standard *p*-value of Student's t-test (Fig. 4H), although the FDR-*p*-value was higher than 0.05, likely due to the relatively low expression level of miR-155. This

result confirms the ability of PNA21 to inhibit the maturation of endogenous miR-155. In addition, a novel miR-155/miR-143 cascade was recently discovered in breast cancer cells, where upregulation of miR-155 downregulates the expression of miR-143 via targeting C/EBPβ (a transcriptional activator for miR-143) [47]. Thus, the significant increase of the miR-143–3p observed in Fig. 4E was consistent with the reduction of miR-155 by PNA21. In contrast, there was no change in the expression level of miR-155 in cells treated with PNA22.

Intrigued by our discovery that long T/M tracts caused higher off-target activity of PNAs, we reevaluated the activities of PNA26-PNA28 (Supplementary Fig. S72A), which were used in our previous study in SH-SY5Y cells [29], to target a different RNA sequence in pre-miR-197 (Supplementary Fig. S72A). Similar to PNAs targeting miR-155, PNA26 had an eight-nucleotide-long T/M tract with an E nucleobase in the middle. We also tested PNA27, which had three additional lysine residues designed to serve as a short cell-penetrating peptide in our earlier studies [48]. The scrambled sequence, PNA28, also had the E nucleobase at the terminal (-NH_2) position. Confirming our observations with PNA21 and PNA22, RNA-seq showed that among PNAs targeting pre-miR-197, PNA26 and PNA27 having an E nucleobase in the middle caused less perturbation of microRNA levels in MCF7 cells than PNA28, having an uninterrupted T/M tract (Supplementary Fig. S72B and C). Both PNA26 and PNA27 showed a significant reduction of mature miR-197 level (Supplementary Fig. S72C and D), consistent with our previous study using SH-SY5Y cells and quantitative RT-PCR analysis [29]. Interestingly, the presence of three additional lysine residues in PNA27 caused little effect on either activity or specificity of PNAs targeting pre-miR-197 (c.f., PNA26 and PNA27 in Supplementary Fig. S72C and D). Collectively, these results confirm our hypothesis that PNAs targeting mixed pyrimidine-purine sequences are inherently more specific than PNAs targeting purine-only sequences.

Discussion

Sequence-specific recognition of complex folded RNAs is an important area of biomedical research with potential broad applications in medicine and biotechnology. Small-molecule RNA-binding ligands are being actively developed but face formidable challenges due to the nonspecific electrostatic interactions and inherent flexibility of the dynamic RNA structure [8–11]. Triple-helical recognition of folded RNA using oligonucleotide analogues, such as PNA, has emerged as an attractive alternative [17–19]. However, the requirement for long polypurine tracts has been a long-standing problem for advancing practical applications of Hoogsteen triple helical recognition. Moreover, our earlier [29] and present results suggest that the "two-letter code" that uses only two Hoogsteen triples lacks specificity in biological systems. Hence, the development of new nucleobases and approaches to stable X•C-G and Y•U-A triples that would allow recognition of mixed purine-pyrimidine sequences becomes a high priority.

In the present study, we demonstrated that the 5-triazolyl uracil (U^{Tr}) nucleobase adjacent to E or P improved the binding affinity of triplex-forming PNAs and enabled targeting purine-rich sequences having a higher number of pyrimidine interruptions than previously reported [17–19]. Biophysical studies and MD simulations showed that the beneficial effect of the U^{Tr} nucleobase was due to a synergetic enhancement of stacking interactions between the triazole ring and E or P heterocycles. These results establish a new approach to pyrimidine recognition in Hoogsteen triple helices that optimizes not only the hydrogen bonding but also the stacking interactions of the PNA ligand.

We also demonstrated that improving the pyrimidine recognition enhanced the biological activity and specificity of triplex-forming PNAs. Our previous studies showed that PNA 9-mers having one E [29] or one P [30] nucleobase (11% pyrimidines in the target RNA sequence) inhibited microRNA maturation and protein synthesis, respectively. Chen and coworkers reported the biological activity of PNA 8-mer and PNA 10-mer having one Q nucleobase (10–13% pyrimidines in the target RNA sequence). The PNA 8-mer stimulated ribosomal frameshifting in a cell-free assay [27], while the 10-mer inhibited influenza viral replication in MDCK cells [28]. In all previous reports on the biological activity of triplex-forming PNAs, the purine-rich RNA target had only one pyrimidine interruption. In the present study, for the first time, we demonstrated the biological activity and specificity of triplex-forming PNA targeting a binding site with two pyrimidine interruptions. As 8-mers having two E nucleobases (25% pyrimidines in the target RNA sequence), PNA18-PNA21 represented a significant advance toward recognizing mixed purine-pyrimidine RNA sequences, and ultimately, any sequence of folded RNA.

RNA-seq results showed that targeting mixed pyrimidine-purine sequences enhanced the biological on-target specificity of triplex-forming PNAs. Biologically significant regulatory RNAs are rich in purine nucleosides that form a variety of non-canonical base pairs and secondary structures. We recently showed that triplex-forming PNAs recognize non-canonical base pairs with high affinity, similar to that for the Watson-Crick base pairs [49]. However, the abundance of purine-rich sequences creates many potential off-target sites for triplex-forming PNAs using only two nucleobases (T and M) for Hoogsteen hydrogen bonding. For example, we previously found that a PNA, used as a negative control (scrambled) sequence to study the inhibition of miR-197, showed an unanticipated inhibition of miR-27a, most likely because of a partial sequence match of seven nucleobases in the pri-miR-27a sequence [29]. In other words, the redundancy of the "two-letter code" leads to many potential off-target sequences for PNAs using only T and M nucleobases. Conversely, in our present study, PNA18-PNA21, having a relatively balanced distribution of four Ms, two Ts, and two Es nucleobases, showed significantly higher biological specificity (Figs. 4D–4F) than PNA22, having a continuous tract of four Ms and two Ts. These results clearly show that PNAs recognizing mixed pyrimidine-purine sequences are inherently more specific in biological systems and further emphasize the importance of developing X•C-G and Y•U-A triples to remove the requirement for polypurine sequences in triple-helical recognition of RNA.

Collectively, our binding affinity and biological activity results demonstrate the potential for improving Hoogsteen recognition of pyrimidines using extended heterocyclic PNA nucleobases. The present study demonstrates that new nucleobases that recognize pyrimidines in PNA-dsRNA triplexes have a promising potential to develop PNA molecules with high affinity and biological specificity. Removing the requirement for polypurine tracts will enable recognition of any RNA

sequence and unlock the full potential of triple helical recognition of biological RNAs.

Acknowledgements

We thank B.R. Tessier for help with the ITC data analysis and Bapeks (www.bapeks.com) for a generous gift of a PNA backbone starting material.

Author contributions: S.F.S. synthesized PNA monomers (M, E, and P), synthesized the modified PNAs, performed UV melting and ITC experiments, and performed data analysis, T.E. conceived and performed the dual fluorescent protein reporter and RNA-seq experiments and data analysis, V.B. conceived and performed MD simulations and data analysis, I.K. synthesized PNA monomers (U^{Tr}), synthesized the modified PNAs, and assisted with data analysis, M.K. conceived the MD experiments and supervised the work, N.S. conceived the study and supervised the work, and E.R. conceived the overall study, supervised the work, and wrote the manuscript with input from the other authors.

Supplementary data

Supplementary data is available at NAR online.

Conflict of interest

None declared.

Funding

This work was supported by the National Science Foundation (CHE-2107900 and CHE-2427911 to E.R.), the National Institutes of Health (R35 GM130207 to E.R.), and the Recovery and Resilience Facility (RRF) grants no. 51/OSI/PA and 33/OSI/DG (RRF project No. 5.2.1.1.i.0/2/24/I/CFLA/001) to M.K. and V.B. We also thank the support from the Ministry of Education, Culture, Sports, Science and Technology (MEXT), Japan Society for the Promotion of Science (KAKENHI grant no. 23K26780, 25K21727, and Grant-in-Aid for Scientific Research (S) (22H04975), JSPS Core-to-Core Program (JPJSCCA20220005), Konan New Century Strategic Research Project, and Chubei Itoh Foundation. Funding to pay for the Open Access publication charges for this article was provided by Binghamton University through the Lyrasis (Fully OA) Open Access Agreement.

Data availability

The data are available in the article and in its online supplementary material. The RNAseq data are deposited in GEO (GSE307169).

References

1. Rinn JL, Chang HY. Long noncoding RNAs: molecular modalities to organismal functions. *Annu Rev Biochem* 2020;89:283–308. https://doi.org/10.1146/annurev-biochem-062917-012708
2. Chen L-L, VN K. Small and long non-coding RNAs: past, present, and future. *Cell* 2024;187:6451–85. https://doi.org/10.1016/j.cell.2024.10.024
3. Nemeth K, Bayraktar R, Ferracin M *et al.* Non-coding RNAs in disease: from mechanisms to therapeutics. *Nat Rev Genet* 2024;25:211–32. https://doi.org/10.1038/s41576-023-00662-1
4. Bartel DP. Metazoan microRNAs. *Cell* 2018;173:20–51. https://doi.org/10.1016/j.cell.2018.03.006
5. Egli M and Manoharan M. Chemistry, structure and function of approved oligonucleotide therapeutics. *Nucleic Acids Res* 2023;51:2529–73. https://doi.org/10.1093/nar/gkad067
6. Tang Q, Khvorova A. RNAi-based drug design: considerations and future directions. *Nat Rev Drug Discov* 2024;23:341–64. https://doi.org/10.1038/s41573-024-00912-9
7. Butcher SE, Pyle AM. The molecular interactions that stabilize RNA tertiary structure: RNA motifs, patterns, and networks. *Acc Chem Res* 2011;44:1302–11. https://doi.org/10.1021/ar200098t
8. Ursu A, Childs-Disney JL, Andrews RJ *et al.* Design of small molecules targeting RNA structure from sequence. *Chem Soc Rev* 2020;49:7252–70. https://doi.org/10.1039/D0CS00455C
9. Childs-Disney JL, Yang X, Gibaut QMR *et al.* Targeting RNA structures with small molecules. *Nat Rev Drug Discov* 2022;21:736–62. https://doi.org/10.1038/s41573-022-00521-4
10. Warner KD, Hajdin CE, Weeks KM. Principles for targeting RNA with drug-like small molecules. *Nat Rev Drug Discov* 2018;17:547–58. https://doi.org/10.1038/nrd.2018.93
11. Zafferani M, Hargrove AE. Small molecule targeting of biologically relevant RNA tertiary and quaternary structures. *Cell Chem Biol* 2021;28:594–609. https://doi.org/10.1016/j.chembiol.2021.03.003
12. Velagapudi SP, Gallo SM, Disney MD. Sequence-based design of bioactive small molecules that target precursor microRNAs. *Nat Chem Biol* 2014;10:291–7. https://doi.org/10.1038/nchembio.1452
13. Velagapudi SP, Cameron MD, Haga CL *et al.* Design of a small molecule against an oncogenic noncoding RNA. *Proc Natl Acad Sci USA* 2016;113:5898–903. https://doi.org/10.1073/pnas.1523975113
14. Rzuczek SG, Colgan LA, Nakai Y *et al.* Precise small-molecule recognition of a toxic CUG RNA repeat expansion. *Nat Chem Biol* 2017;13:188–93. https://doi.org/10.1038/nchembio.2251
15. Gibaut QMR, Bush JA, Tong Y *et al.* Transcriptome-wide studies of RNA-targeted small molecules provide a simple and selective r(CUG)exp degrader in myotonic dystrophy. *ACS Cent Sci* 2023;9:1342–53. https://doi.org/10.1021/acscentsci.2c01223
16. Nielsen PE, Egholm M, Berg RH *et al.* Sequence-selective recognition of DNA by strand displacement with a thymine-substituted polyamide. *Science* 1991;254:1497–500. https://doi.org/10.1126/science.1962210
17. Brodyagin N, Katkevics M, Kotikam V *et al.* Chemical approaches to discover the full potential of peptide nucleic acids in biomedical applications. *Beilstein J Org Chem* 2021;17:1641–88. https://doi.org/10.3762/bjoc.17.116
18. Zhan X, Deng L, Chen G. Mechanisms and applications of peptide nucleic acids selectively binding to double-stranded RNA. *Biopolymers* 2022;113:e23476. https://doi.org/10.1002/bip.23476
19. Katkevics M, MacKay JA, Rozners E. Triplex-forming peptide nucleic acids as emerging ligands to modulate structure and function of complex RNAs. *Chem Commun* 2024;60:1999–2008. https://doi.org/10.1039/D3CC05409H
20. Roberts RW, Crothers DM. Stability and properties of double and triple helices: dramatic effects of RNA or DNA backbone composition. *Science* 1992;258:1463–6. https://doi.org/10.1126/science.1279808
21. Devi G, Yuan Z, Lu Y *et al.* Incorporation of thio-pseudoisocytosine into triplex-forming peptide nucleic acids for enhanced recognition of RNA duplexes. *Nucleic Acids Res* 2014;42:4008–18. https://doi.org/10.1093/nar/gkt1367
22. Egholm M, Christensen L, Dueholm KL *et al.* Efficient pH-independent sequence-specific DNA binding by pseudoisocytosine-containing bis-PNA. *Nucl Acids Res* 1995;23:217–22. https://doi.org/10.1093/nar/23.2.217

23. Zengeya T, Gupta P, Rozners E. Triple helical recognition of RNA using 2-aminopyridine-modified PNA at physiologically relevant conditions. *Angew Chem Int Ed* 2012;51:12593–6. https://doi.org/10.1002/anie.201207925
24. Ryan CA, Brodyagin N, Lok J *et al.* 2-aminopyridine nucleobase improves triple helical recognition of RNA and DNA when used instead of pseudoisocytosine in peptide nucleic acids. *Biochemistry* 2021;60:1919–25. https://doi.org/10.1021/acs.biochem.1c00275
25. Kumpina I, Loubidi M, Rozners E. Comparison of 2-aminopyridine and 4-thiopseudisocytosine PNA nucleobases for hoogsteen recognition of guanosine in RNA. *ACS Omega* 2024;9:7249–54. https://doi.org/10.1021/acsomega.3c09775
26. Kotikam V, Kennedy SD, MacKay JA *et al.* Synthetic, structural, and RNA-binding studies on 2-aminopyridine-modified triplex-forming peptide nucleic acids. *Chem A Eur J* 2019;25:4242. https://doi.org/10.1002/chem.201900719
27. Puah RY, Jia H, Maraswami M *et al.* Selective binding to mRNA duplex regions by chemically modified peptide nucleic acids stimulates ribosomal frameshifting. *Biochemistry* 2018;57:149–59. https://doi.org/10.1021/acs.biochem.7b00744
28. Kesy J, Patil KM, Kumar SR *et al.* A short chemically modified dsRNA-binding PNA (dbPNA) inhibits influenza viral replication by targeting viral RNA panhandle structure. *Bioconjug Chem* 2019;30:931–43. https://doi.org/10.1021/acs.bioconjchem.9b00039
29. Endoh T, Brodyagin N, Hnedzko D *et al.* Triple-helical binding of peptide nucleic acid inhibits maturation of endogenous microRNA-197. *ACS Chem Biol* 2021;16:1147–51. https://doi.org/10.1021/acschembio.1c00133
30. Endoh T, Hnedzko D, Rozners E *et al.* 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* 2016;55:899–903. https://doi.org/10.1002/anie.201505938
31. Fox KR, Brown T, DA R. In: Waring MJ (ed.), *Chemical Biology No. 7: DNA-targeting Molecules as Therapeutic Agents: DNA Recognition by Parallel Triplex Formation*. Cambridge, UK: Royal Society of Chemistry, 2018, 1–32.
32. Hari Y, Obika S, Imanishi T. Towards the sequence-selective recognition of double-stranded DNA containing pyrimidine-purine interruptions by triplex-forming oligonucleotides. *Eur J Org Chem* 2012;2012:2875–87. https://doi.org/10.1002/ejoc.201101821
33. Eldrup AB, Dahl O, Nielsen PE. A novel peptide nucleic acid monomer for recognition of thymine in triple-helix structures. *J Am Chem Soc* 1997;119:11116–7. https://doi.org/10.1021/ja9717424
34. Kumpina I, Baskevics V, Nguyen KD *et al.* Nucleobase and linker modification for triple-helical recognition of pyrimidines in RNA using peptide nucleic acids. *ChemBioChem* 2023;24:e202300291.
35. Gupta P, Zengeya T, Rozners E. Triple helical recognition of pyrimidine inversions in polypurine tracts of RNA by nucleobase-modified PNA. *Chem Commun* 2011;47:11125. https://doi.org/10.1039/c1cc14706d
36. Toh D-FK, Devi G, Patil KM *et al.* 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* 2016;44:9071–82.
37. Semenyuk A, Darian E, Liu J *et al.* Targeting of an interrupted polypurine:polypyrimidine sequence in mammalian cells by a triplex-forming oligonucleotide containing a novel base analogue. *Biochemistry* 2010;49:7867–78. https://doi.org/10.1021/bi100797z
38. Ryan CA, Baskevics V, Katkevics M *et al.* 2-Guanidyl pyridine PNA nucleobase for triple-helical Hoogsteen recognition of cytosine in double-stranded RNA. *Chem Commun* 2022;58:7148–51. https://doi.org/10.1039/D2CC02615E
39. Brodyagin N, Kumpina I, Applegate J *et al.* Pyridazine nucleobase in triplex-forming PNA improves recognition of cytosine interruptions of polypurine tracts in RNA. *ACS Chem Biol* 2021;16:872–81. https://doi.org/10.1021/acschembio.1c00044
40. Kumpina I, Brodyagin N, MacKay JA *et al.* Synthesis and RNA-binding properties of extended nucleobases for triplex-forming peptide nucleic acids. *J Org Chem* 2019;84:13276–98. https://doi.org/10.1021/acs.joc.9b01133
41. Kumpina I, Baskevics V, Walby GD *et al.* Synthesis of 2-aminopyridine-modified peptide nucleic acids. *Synlett* 2023;35:649–53.
42. Mahesh G, Biswas R. MicroRNA-155: a master regulator of inflammation. *J Interferon Cytokine Res* 2019;39:321–30. https://doi.org/10.1089/jir.2018.0155
43. Alivernini S, Gremese E, McSharry C *et al.* MicroRNA-155—at the critical interface of innate and adaptive immunity in arthritis. *Front Immunol* 2018;8:1932. https://doi.org/10.3389/fimmu.2017.01932
44. Ryan CA, Rahman M, Md KV *et al.* Triplex-forming peptide nucleic acid controls dynamic conformations of RNA bulges. *J Am Chem Soc* 2023;145:10497–504. https://doi.org/10.1021/jacs.2c12488
45. Meng L, Lin J, Huang Q *et al.* Panax notoginseng saponins attenuate oxygen–glucose deprivation/reoxygenation-induced injury in human SH-SY5Y cells by regulating the expression of inflammatory factors through miR-155. *Biol Pharm Bull* 2019;42:462–7. https://doi.org/10.1248/bpb.b18-00799
46. Zhang C-M, Zhao J, Deng H-Y. MiR-155 promotes proliferation of human breast cancer MCF-7 cells through targeting tumor protein 53-induced nuclear protein 1. *J Biomed Sci* 2013;20:79. https://doi.org/10.1186/1423-0127-20-79
47. Jiang S, Zhang LF, Zhang HW *et al.* A novel miR-155/miR-143 cascade controls glycolysis by regulating hexokinase 2 in breast cancer cells. *EMBO J* 2012;31:1985–1998. https://doi.org/10.1038/emboj.2012.45
48. Brodyagin N, Kataoka Y, Kumpina I *et al.* Cellular uptake of 2-aminopyridine-modified peptide nucleic acids conjugated with cell-penetrating peptides. *Biopolymers* 2022;113:e23484. https://doi.org/10.1002/bip.23484
49. Farshineh Saei S, Baskevics V, Katkevics M *et al.* Recognition of noncanonical RNA base pairs using triplex-forming peptide nucleic acids. *ACS Chem Biol* 2025;20:179–85.

Received: September 6, 2025. **Revised:** February 12, 2026. **Accepted:** March 3, 2026