RESEARCH ARTICLE

Improved Triplex-Forming Isoorotamide PNA Nucleobases for A-U Recognition of RNA Duplexes**

John M. Talbott[+],[a] Brandon R. Tessier[+],[b] Emily E. Harding,[a] Grant D. Walby,[a] Kyle J. Hess,[a] Vladislavs Baskevics,[c] Martins Katkevics,[c] Eriks Rozners,[b] and James A. MacKay*[a]

[a] J. M. Talbott, E. E. Harding, Dr. G. D. Walby, K. J. Hess, Prof. Dr. J. A. MacKay
Department of Chemistry and Biochemistry
Elizabethtown College
Elizabethtown, Pennsylvania 17022 (USA)
E-mail: mackayj@etown.edu
https://www.etown.edu/depts/chemistry-biochemistry/research

[b] B. R. Tessier, Prof. Dr. E. Rozners
Department of Chemistry
Binghamton University
Binghamton, New York 13902 (USA)

[c] V. Baskevics, Dr. M. Katkevics
Latvian Institute of Organic Synthesis
Aizkraukles 21, Riga, LV-1006 (Latvia)

[+] These authors contributed equally to this work

[**] PNA = peptide nucleic acid

Supporting information for this article is given via a link at the end of the document.

Abstract: Four new isoorotamide (***Io***) containing PNA nucleobases were designed for A-U recognition of double helical RNA. New PNA monomers were prepared efficiently and incorporated into PNA 9-mers for binding A-U in a PNA:RNA_2 triplex. Isothermal titration calorimetry and UV thermal melting experiments revealed slightly improved binding affinity for singly-modified PNA compared to known A-binding nucleobases. Molecular dynamics simulations provided further insights into binding of ***Io*** bases in the triple helix. Together, data revealed interesting insights into binding modes including the notion that three Hoogsteen-hydrogen bonds are unnecessary for strong selective binding of an extended nucleobase. Cationic monomer ***Io8*** additionally gave the highest affinity observed for an A-binding nucleobase to date. These results will help inform future nucleobase design toward the goal of recognizing any sequence of double helical RNA.

Introduction

RNA has grown in its relevance and prevalence despite its historical role of simply bridging the gap between DNA and proteins. In 2012, the Encyclopedia of DNA Elements (ENCODE) project revealed that only a small percentage (~2%) of the human genome is transcribed into coding RNA while most RNA is noncoding.[1] The 21st century has since seen an outburst of investigation into the structure and function of noncoding RNAs (ncRNAs) with biological functions including regulating gene expression, catalyzing chemical reactions, and post transcriptional modification.[2] The ability to selectively recognize, bind, and modulate function of regulatory RNAs represents an important goal for fundamental applications in biology as demonstrated by techniques including CRISPR-Cas9 gene editing[3] and therapeutic oligonucleotides.[4]

Many ncRNAs adopt double helical motifs or conformations through tertiary structures. Furthermore, the triple helix or triplex motif is well known in natural RNA transcripts.[5] Drawing inspiration from nature, double helical regions of RNA afford the opportunity to utilize triplex forming oligonucleotides (TFOs) for sequence recognition (Figure 1a). However, formation of the TFO:RNA_2 triple helix presents several limitations. First, formation of the triplex relies on insertion of the TFO into the deep and narrow major groove of the A-form RNA helix which presents an electrostatic conundrum due to charge repulsion of three strands. Further, selectivity results from affinity of the TFO base to the Hoogsteen face of the Watson-Crick base pairs through the U*A-U and C^{+}*G-C base triples (Figure 1b) - both of which form two hydrogen bonds with the purine residue. Binding such as this requires polypurine tracks of RNA and a pH below that of many biological systems due to the requirement for protonation of cytosine (pK_a ~4.5).

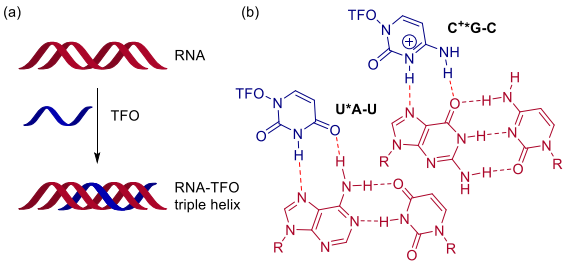


Figure 1. (a) TFO strategy for RNA recognition (b) Natural Hoogsteen base triples U*A-U and C+*G-C

Peptide nucleic acid (PNA, Figure 2) is a DNA analogue that was developed by Nielsen and coworkers in 1991[6] as a TFO and has emerged as an important tool for nucleic acid recognition in biotechnology.[7] In its original form, PNA replaces the phosphodiester backbone in DNA or RNA with an *N*-(aminoethyl)glycine backbone, the monomer of which can be

prepared easily and modified with synthetic nucleobases.[7a, 8] From protected PNA monomers, oligomeric PNA can then be prepared using standard peptide coupling chemistry on solid phase. Owing to its ease of synthesis, stability to proteases, and its uncharged backbone, PNA provides an invaluable molecular tool that binds to both DNA and RNA with high affinity and specificity through complementary hydrogen bonding patterns including the triple helix. Recent results demonstrate that PNA is an excellent ligand for the sequence specific recognition of double helical RNA.[7]

In 2010, Rozners and co-workers were the first to bind PNA in the major groove of double helical RNA.[9] Improvements in PNA binding were disclosed in 2012 using 2-aminopyridine, (M in Figure 2).[10] The greater basicity of M (pK_a ~ 6.7) compared to the natural cytosine increases the likelihood for protonation at physiological pH and improves the binding affinity of M for G-C. Interestingly, PNAs demonstrated greater affinity for dsRNA than dsDNA in triple helix formation.[11] A follow-up report by our group using NMR structural studies demonstrated that PNA:RNA_2 affinity owes to a favorable hydrogen-bonding geometry between the PNA backbone amide N-H and RNA backbone phosphates.[12] These findings along with developments from others[13] have resulted in the use of nucleobase-modified PNA for the recognition of regulatory RNA targets.[14] Still, the requirement for polypurine tracks remains, and the ability to bind RNA sequences with multiple pyrimidine interruptions is an unsolved problem.[15] Accordingly, there exists a need for new modified nucleobases that can address this limitation and ultimately allow PNA to recognize any sequence of double helical RNA.

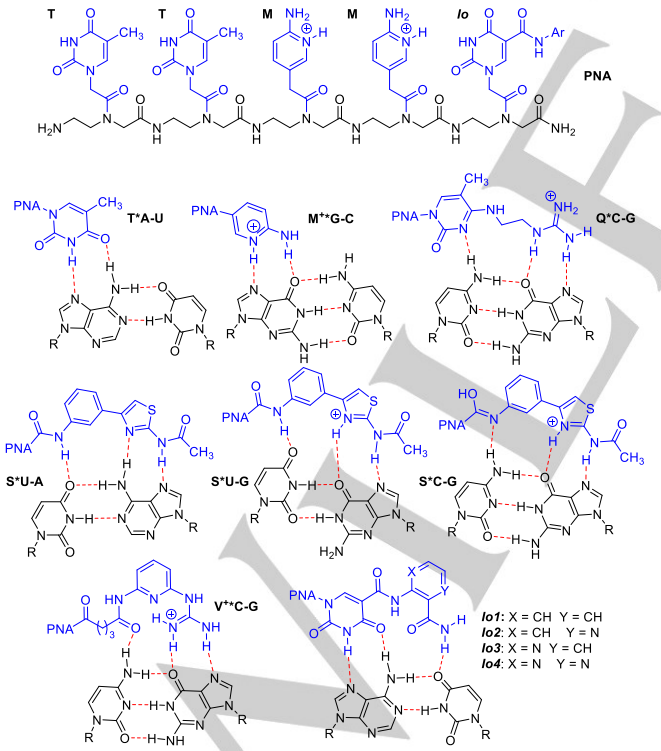


Figure 2. Structures of triplex-forming PNA and Hoogsteen hydrogen-bonded base triplets. PNA nucleobases are denoted in blue and R denotes the sugar-phosphate backbone of RNA.

We have recently been interested in base recognition through an extended nucleobase design.[16] This strategy capitalizes on hydrogen bonding from the PNA nucleobase across the entire Hoogsteen-face of the Watson-Crick base pair to afford multiple Hoogsteen H-bonds capable of increasing binding affinity and selectivity. Our group[16] and others[13d, 17] have utilized this strategy for recognition of each of the four natural nucleobases with the ultimate goal of finding a PNA "genetic code" capable of recognition of any sequence of RNA. Given the strong and selective binding through two Hoogsteen hydrogen bonds, much of the preliminary work in this field has focused on purine recognition. Notable advancements toward pyrimidine recognition have included the extended cationic bases Q[17] and V[16c] for C-recognition (Figure 2). Chen and coworkers also reported the use of the S base for promiscuous recognition of U-A, C-G and wobble U-G base pairs in RNA duplexes.[13d] However, Q and S were only tested on RNA with a single pyrimidine interruption and PNA with more than one V modification quickly lost affinity and specificity.

In 2020, our group disclosed a series of extended PNA nucleobases based on isoorotic acid aimed at improved affinity for the recognition of A-U base pairs in dsRNA (Figure 2, ***Io1-Io4***).[16b] These bases used a planar isoorotamide system with a pendant primary amide-containing aryl substituent. In the design, we hypothesized that the primary amide formed a third Hoogsteen hydrogen-bond with the uracil carbonyl of the A-U base pair and expected this to aid in the affinity and selectivity for A-U. We found that ***Io1-Io3*** had similar but slightly stronger binding to A, compared to T-control and that ***Io4*** showed higher affinity. Additionally, we showed for the first time in PNA:RNA_2 recognition that four consecutive extended nucleobases were tolerated with cooperativity effects improving binding.[16b] But despite improved binding, selectivity of the ***Io4*** nucleobase was lower than that of T or M for their matched base pairs with off target affinity for G being nearly 6 times that for the other mismatches and around half that of the matched A residue. Given the noted limitations of extended nucleobases including Q, S, V, and ***Io***, we aimed to further our study with the intent of gaining insight into nucleobase design and binding chemistry with the ultimate goal of using an isoorotamide scaffold for U-recognition.

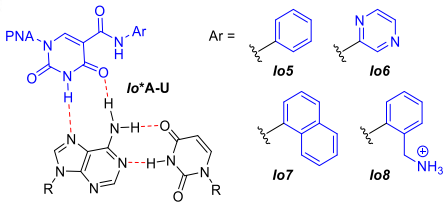


Figure 3. Next generation ***Io*** nucleobases

The increased binding affinity of sequential ***Io4*** bases led to the hypothesis that improved binding could be due to the π-π overlap of the extended nucleobases in the triplex. We also hypothesized that an internal H-bond between the primary amide in ***Io1-Io4*** and the isoorotamide N-H preorganized the extended nucleobase for ideal binding geometries. Still, given the similar binding of ***Io1-Io3*** bases compared to the T control, it is plausible that the third hydrogen bond is not as essential as we had proposed. To investigate these ideas and pave the way for better RNA recognizing nucleobases, a new series of ***Io*** derivatives was envisioned (Figure 3). ***Io5*** and ***Io6*** were conceived to determine the role of the primary amide in ***Io1-Io4*** and we expected to find decreased affinity for A using these bases if the third hydrogen bond was essential for binding. ***Io7*** was designed to see if a more extended π system might improve stacking interactions. Finally,

Io8 was envisioned as a cationic analogue that might be aided by a third Hoogsteen hydrogen-bonding interaction with U. Given the success of cationic residues M and V, we conjectured that a cationic A-binding residue could increase binding affinity and add to our arsenal of triplex forming nucleobases.

Results and Discussion

Prior to embarking on the synthesis of these new compounds, we used computational methods to study if ***Io5*** and ***Io6*** exhibited similar conformational preferences to ***Io1*** and ***Io4*** respectively. In our previous study, HF-631G(d) calculations of ***Io1*** and ***Io4*** demonstrated good hydrogen-bonding distances, strong planarity, and the presence of the 'third H-bond' between the primary amide of the ***Io*** derivative and the U-base of the Watson-Crick base pair.[16b] We recognized that ***Io5*** and ***Io6***, lacking amides, could possibly distort from planarity and performed ab initio calculations to determine their ground state geometries.

Following our previously established protocols,[16a, 16b] ab initio calculations [HF-631G(d)] afforded ground state geometries of ***Io5*** and ***Io6*** (with a methyl in the place of the PNA backbone). In the case of ***Io5***, the geometry of the conjugated molecule was fully planar throughout the entire system and nearly superimposable with the optimized ***Io1*** ground state calculation. In contrast, ***Io6*** differed in that several local minima were found where there seemed an apparent repulsion of lone pairs between the isoorotamide carbonyl and the ortho nitrogen on the pyrazine. Two energy minima involved a 45-degree twist of the pyrazine aryl outside of planarity. However, when the dihedral angle about the N-aryl bond of the pyrazine was adjusted so the pyrazine ortho nitrogen was 180 degrees with respect to the amide carbonyl and the geometry optimization was reinitiated, a planar conformation was found that represented a global energy minimum (Supporting Information, Figure S30a).

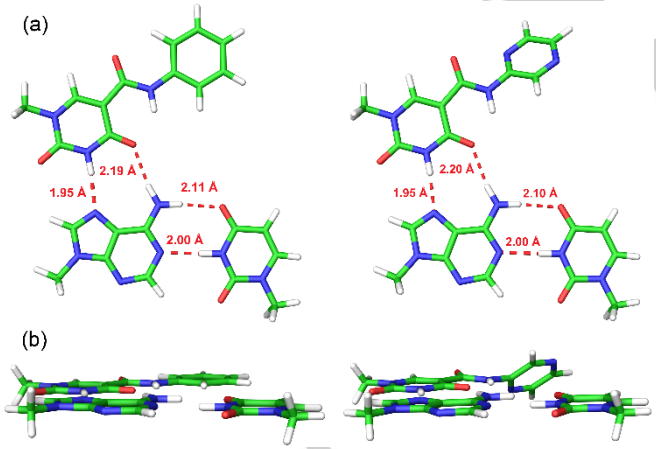
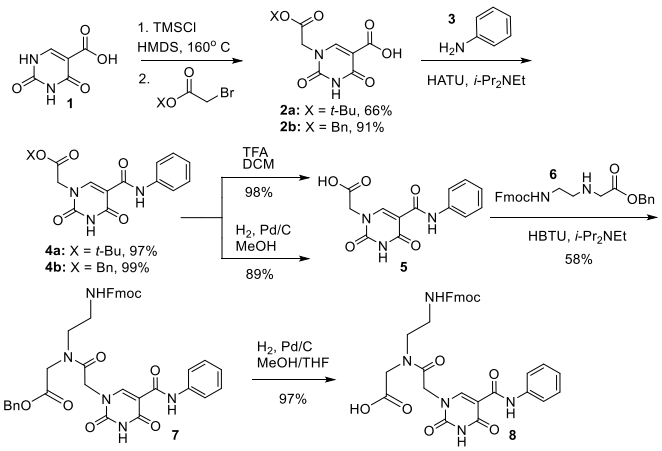


Figure 4. Geometry optimizations of ***Io5****A-U (left) and ***Io6****A-U (right) using HF-631G(d). (a) top view showing H-bond distances; (b) side view showing planarity of extended system.

With the ground states determined, each minimized nucleobase was docked with a previously optimized A-U base pair and another geometry optimization was performed on the triplet affording the computed base triples shown in Figure 4. It is noteworthy that in both cases, there was little to no change in the original nucleobase geometry upon completion of the base-triple calculation. As seen in Figure 4, ***Io5*** and ***Io6*** provided typical hydrogen-bonding distances (approximately 2 Å in length) between the ***Io*** pyrimidine and the A-base while maintaining a planar geometry between the 3 molecules in the triplex. For the ***Io6*** base triple, a second local minimum with similar computed binding energy was found that contains a fully planar pyrazine ring. Unsurprisingly, planarity may not be required for the hydrogen bonding interactions, but we still expected that planarity may help lead to more efficient π-stacking in the triple helix.

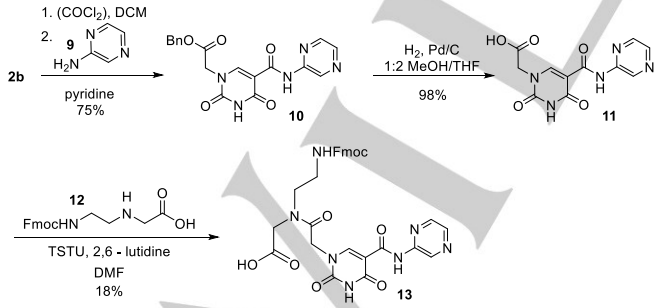
A third H-bond in ***Io1-Io4*** could be essential in promoting triplex formation but would also require an additional level of preorganization. Further, modeling of ***Io8****A-U determined that the third hydrogen bond was feasible in this system provided the ammonium moiety was forced into a conformation where it was not binding either of the ***Io*** carbonyl carbons (Supporting Information, Figure S31). Thus, we saw value in the synthesis of these novel compounds to test their affinity in PNA:RNA_2 triplexes.

The synthetic route to ***Io5*** monomer **8** (Scheme 1) began with alkylation of isoorotic acid (**1**). We explored two protecting groups for the acetic ester moiety in compound **2**. The *tert*-butyl group was previously used to make **2a** in our synthesis of ***Io1-Io4***.[16b] However, alkylation with benzylbromoacetate, afforded **2b** in better yield after precipitation from methanol without the need for chromatographic purification. Next, **2b** was coupled to aniline (**3**) using 1-[bis(dimethylamino)methylene]-1H-1,1,2,3-triazole[4,5-b]pyridinium-3-oxide hexafluorophosphate (HATU), yielding a yellow solid, **4b**, in 99% yield consistently on a multi-gram scale. This reaction was also successfully completed using closely related (2-(1*H*-benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (HBTU) with similar yields. However, while deprotection of the benzyl group with Pd/C and H_2 formed the carboxylic acid **5** in good yield, poor reproducibility likely due to the inadequate solubility of **4b** in methanol resulted. Thus, compound **5** was derived from **2a** through the same coupling reaction with **3** and then the deprotection of **4a** using TFA, giving noticeably better yields (98%) of the carboxylic acid. Standard conditions[16a, 16b] were employed to couple **5** to the benzyl protected PNA backbone (**6**). Finally, while there are reported examples of Fmoc cleavage under hydrogenolysis conditions,[18] selective debenzylation is known to be effective using short reaction times and monitoring by TLC.[19] This strategy allows for selective deprotection of the benzyl ester in **7** and afforded the desired carboxylic acid monomer **8** in high yields.



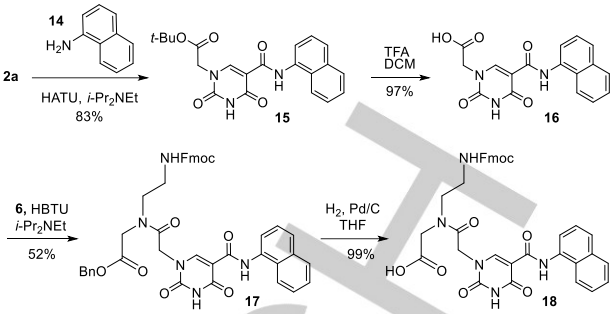
Scheme 1. Synthesis of the ***Io5*** monomer

We then attempted to prepare the ***Io6*** monomer (**13**, Scheme 2) using the same chemistry as for ***Io4*** and/or ***Io5***. However, pyrazine containing compounds (**10**, **11**, and **13**) proved to be quite difficult to prepare due to unexpected insolubility of most intermediates in common solvents. We attributed this to the nitrogen atoms in the pyrazine moiety severely altering the polarity and aggregation states, thus decreasing solubility in organic solvents. Standard coupling reagents (HBTU, HATU, and EDCI) afforded no product **10** when **2a** or **2b** were treated with 2-aminopyrazine (**9**). We had faced similar problems in the synthesis of ***Io4***,[16b] remedying the problem using Ghosez's reagent.[20] While Ghosez's reagent successfully produced product in the reaction of **9** to **10** in about 15% conversion, we ultimately discovered that the reaction was much more efficient when the acyl chloride was produced using oxalyl chloride in dichloromethane followed by a solvent swap with pyridine, then the addition of **9**. Nitrogen in the pyrazine greatly affected the acidity of the amide resulting in unwanted byproducts in basic conditions[21] and initial attempts at purification proved challenging. It was eventually determined that addition of water directly to the reaction precipitated pure **10** without the need for column purification (75% yield). Hydrogenolysis of the benzyl ester in **10** afforded **11** once the ideal solvent for reactant solubility was determined. A 1:2 ratio of methanol/THF was employed with more or less THF hampering solubility. Additionally, increasing catalyst loading from 15 mol% to 20 mol% allowed for efficient conversion of **10** to **11** as an extended reaction time drastically decreased yields due to degradation of **11**. With the carboxylic acid in hand, **11** and **6** were subjected to previously employed peptide coupling conditions (HBTU, HATU, EDCI, etc.) however isolation of pure benzyl protected product was elusive. Acyl chloride formation, like in the formation of **10**, proved problematic as the desired product was unstable to the reaction conditions. Ultimately, direct coupling of the free carboxylic acid PNA backbone (**12**) with carboxylic acid **11** using TSTU afforded **13** that could be sufficiently purified by column chromatography albeit in poor yield due to undesired side products and incomplete conversion. It is noteworthy that TSTU has been used with unprotected PNA backbone in the past[16a, 16b, 22] and while yields typically suffer from these conditions, it offers an attractive alternative to protected backbone reactants when problems arise in either the coupling or the deprotection step.



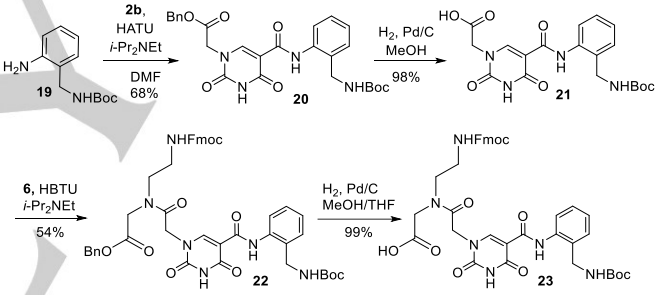
Scheme 2. Synthesis of the ***Io6*** monomer

We prepared the ***Io7*** PNA monomer **18** (Scheme 3) using chemistry fully analogous to the preparation of **8** (Scheme 1) and with similar yields in each respective step.



Scheme 3. Synthesis of the ***Io7*** monomer

In the preparation of the ***Io8*** monomer (**23**), we opted to protect the benzylamine group with Boc, knowing that **19** was readily available,[23] that Boc was stable to standard Fmoc chemistry in PNA synthesis, and that the amine protecting group would be easily removed upon acidic cleavage of the synthetic PNA from solid support. For this synthesis (Scheme 4), the benzyl containing **2b** was chosen in lieu of **2a** to avoid acidic deprotection conditions that may prematurely remove the requisite Boc group. Following typical amide coupling, hydrogenolysis, and PNA backbone coupling procedures, **22** was prepared in good yields. Final deprotection of the benzyl moiety led to **23** in quantitative yield.



Scheme 4. Synthesis of the ***Io8*** monomer

PNA synthesis using new ***Io*** monomers was carried out on solid support using Fmoc chemistry on an Expedite 8909 DNA/RNA/PNA synthesizer and following established protocols.[16b, 24] Three PNA sequences were employed for binding studies with the aim of exploring the effects of a single synthetic modification (Figure 5, **PNA1**) and multiple modifications (**PNA2** and **PNA3**). To determine the affinity of ***Io*** PNA nucleobases for their matched RNA base pairs, RNA hairpins containing a purine rich 5' arm with variable base pairs (**HRP 1-4**, in red), were chosen.

During PNA oligomerization, an important observation was made that yields were low with particularly hydrophobic aromatic residues such as ***Io5*** and ***Io7***. For example, even a single hydrophobic naphthyl group (***Io7***) limited PNA solubility and made multiple modifications with ***Io7*** an impossibility. Further, when four ***Io5*** bases were employed in **PNA3** the yield was exceptionally low presumably owing to the same reasons. Additionally, given the challenges encountered in the synthesis and purification of ***Io6*** monomer (**13**), we were delighted to obtain purified PNA containing ***Io6*** from the solid phase synthesis and PNA purification conditions.

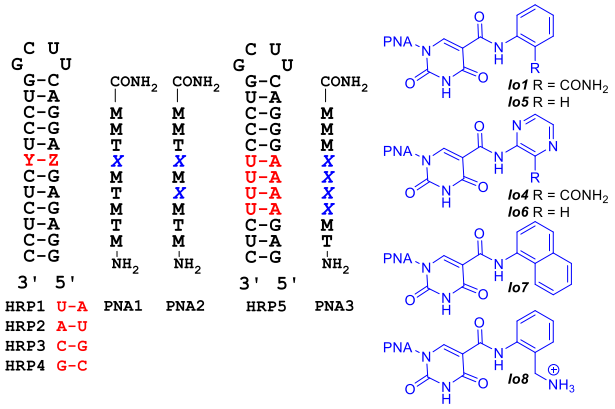


Figure 5. Sequences for RNA hairpins and PNA ligands for binding studies.

Table 1. Binding affinity for ***Io*** containing PNA with single modifications using UV melting (˚C, top)[a] and ITC (10^6 M^{-1}, bottom)[b]

Entry	PNA1	HRP1 (A)	HRP2 (U)	HRP3 (G)	HRP4 (C)
1	X = T	65.0 ± 0.5[c] 8.6 ± 0.7[c]	-	-	-
2	X = ***Io1***	69.1 ± 0.1[c] 8.3 ± 0.2[c]	-	-	-
3	X = ***Io4***	68.4 ± 0.7[c] 15.8 ± 0.1[c]	35.2 ± 7[c] 0.9 ± 0.1[c]	56.7 ± 0.4[c] 6.1 ± 0.6[c]	31.6 ± 0.4[c] 0.7 ± 0.0[c]
4	X = ***Io5***	73 ± 2 14.3 ± 0.2	35.1 ± 0.6 3.1 ± 0.1	51.4 ± 0.4 7.1 ± 0.3	33.1 ± 0.2 3.3 ± 0.2
5	X = ***Io6***	74 .7± 0.3 15.9 ± 0.1[d]	38.6 ± 0.4 ND[e]	57.1 ± 0.5 ND[e]	38.8 ± 0.4 ND[e]
6	X = ***Io7***	71.5 ± 0.5 ND[e]	50.7 ± 0.7 ND[e]	51.2 ± 0.6 ND[e]	49.5 ± 0.6 ND[e]
7	X = ***Io8***	73 ± 1 28.5 ± 0.4	34.4 ± 0.6 3.4 ± 0.0	55.0 ± 0.8 16.3 ± 0.1	35.0 ± 0.9 3.1 ± 0.1

[a] UV melting temperatures, T_m (average of five experiments ± stand. dev.). [b] Association constants K_a x 10^6 M^{-1} ± stand. dev. are averages of three ITC measurements using a Malvern MicroCal iTC200 in 2 mM $MgCl_2$, 90 mM KCl, 10 mM NaCl, 50 mM potassium phosphate buffer pH 7.4 at 25 ˚C. [c] Data from ref 16b. [d] average of two ITC measurements [e] ND – Not determined

With the PNA oligonucleotides in hand, the first binding studies involved the comparison of **PNA1** containing single modifications using each ***Io*** base in concert with matched **HRP1** (Table 1). We were immediately surprised to find that ***Io5*** (entry 4), lacking the pendant primary amide moiety showed improved binding relative to T control (entry 1) and ***Io1*** (entry 2). In fact, ***Io5*** had similar K_a values to the previously best A-binding base ***Io4*** (entry 3) with slightly elevated T_m values. ***Io6*** (entry 5) displayed similar UV melting compared to ***Io5*** and ITC experiments did not show significant enough improvement over ***Io4*** to warrant further exploration of this difficult to prepare monomer. ***Io7*** (entry 6) showed a similar UV melting profile to ***Io5*** and ***Io6*** using **HRP1**. Finally, when cationic ***Io8*** was employed, K_a improved to over triple that of T control (compare entry 7 and entry 1), while UV melting was similar to ***Io4***-***Io7***.

Selectivity for ***Io*** nucleobases was also examined by comparing each of the singly-mismatched hairpins (**HRP2-4**, last three columns in Table 1) using both UV melting and ITC. With **PNA1** and ***Io5*** (entry 4), mismatched hairpins **HRP2-4** all showed weaker binding compared to the expected matched pair **HRP1**. However, for **HRP3** (C-G variable base pair) the triple helix was more stable than the other mismatched triplexes. Similarly, ***Io6*** and ***Io8*** showed weak off-target binding with A-U (**HRP2**) and G-C (**HRP4**) but moderate binding to C-G (**HRP3**). These new data profile similar to ***Io4*** (entry 3) which also showed moderate binding to the **HRP3** mismatch.[16b] Finally, it was noted that ***Io7*** showed the poorest selectivity from UV melting.

We next opted to explore the effects of multiple ***Io*** nucleobase modifications in **PNA2** and **PNA3** (Table 2). When **PNA2** with two modified bases using ***Io5*** and ***Io8*** were employed in UV melting experiments, an improved affinity was noted compared to T-control (Table 2, compare entries 1-3). However, similar affinities were observed compared to singly modified **PNA1**, found in Table 1. For example, two ***Io5*** modifications (Table 2, entry 2) reveal a comparable UV melting temperature to the single ***Io5*** modification (Table 1, entry 4). A similar result is observed for ***Io8*** (Table 2, entry 3 compared to Table 1, entry 7) where in this case, K_a was nearly the same as well. Notably, selectivity was similar for the matched versus mismatched hairpins with **PNA2** (Supporting Information, Tables S11 and S12). When **PNA3** was used to test the effects of four consecutive modified nucleobases, UV melting revealed an enhancement in binding compared to a single modification, albeit with a lower magnitude enhancement compared with ***Io4*** (entries 4-7). **PNA3**/**HRP5** binding with X = ***Io5*** and ***Io8*** both exhibited UV melting temperatures values higher than T-control. However, as noted earlier, when hydrophobic ***Io5*** residues were employed (entries 2 & 6), solubility limited the ability to gather ITC binding data and the T_m value observed may be attributed to hydrophobic effects in solvation rather than melting due to disruption of the hydrogen bonded triplex. For **PNA3**, containing ***Io8*** (entry 7), we observed the expected solubility and were able to perform ITC. But in this case, to obtain saturation in the ITC curve, it required an increased concentration of PNA while keeping the RNA concentration constant. The result revealed a stoichiometry of >3:1 (PNA:RNA) and may be attributed to nonspecific aggregation (Supporting Information, Figure S19).

Table 2. Thermal melting and ITC binding affinity for ***Io*** containing PNA with 2 and 4 modifications.

Entry	PNA (X)	HRP	T_m[b]	K_a[a]
1	PNA2 (X = T)	HRP1	65.0 ± 0.5[c]	8.6 ± 0.7[c]
2	PNA2 (X = ***Io5***)	HRP1	73.3 ± 0.8	ND[d]
3	PNA2 (X = ***Io8***)	HRP1	70.8 ± 0.8	27.0 ± 0.2
4	PNA3 (X = T)	HRP5	71.2 ± 0.7[c]	18.1 ± 1[c]
5	PNA3 (X = ***Io4***)	HRP5	80.4 ± 0.4[c]	45 ± 3[c]
6	PNA3 (X = ***Io5***)	HRP5	74.7 ± 0.5	ND[d]
7	PNA3 (X = ***Io8***)	HRP5	77.5 ± 0.6	NS[d]

[a] Association constants K_a x 10^6 M^{-1} ± stand. dev. are averages of three ITC measurements using a Malvern MicroCal iTC200 in 2 mM $MgCl_2$, 90 mM KCl, 10 mM NaCl, 50 mM potassium phosphate buffer pH 7.4 at 25 ˚C. [b] UV melting temperatures T_m (average of five experiments ± stand. dev.). [c] Data from ref 16b. [d] ND – Not determined; NS – Non-specific binding.

To gain better insight into molecular interactions in the triplex, we turned to molecular dynamics. Using model **PNA1-HRP1** based off a template of the PNA-dsRNA triplex provided by previous NMR structural studies,[12] (for details, see Supporting Information) the ***Io*** nucleobases were separately inserted into the PNA model strand and subjected to 150 ns unrestricted Desmond molecular dynamics. From each simulation, data was extracted to account for geometry of the ***Io****A-U base triples. Table 3 displays X-H----X hydrogen bond probability, average bond distance, and average bond angle for each of the key Hoogsteen hydrogen bonds in the ***Io****A-U base triples. In all the observed entries, the formation of hydrogen bonds between the uracil (U) moiety in ***Io*** and adenine (A) in the ***Io****A-U base triple was found to be permanent (95-99%) with the expected bond distances (1.8 – 2.2 Å) and exhibited favorable bond angles (159-166°). These findings indicate the optimal positioning of the pyrimidine moiety within the ***Io*** base triple.

Table 3. Molecular dynamics data for PNA1-HRP1 interactions in last 50ns of simulation. Hydrogen bonding is reported as probabilities (%), average bond lengths (Å), and average bond angles.

Entry	Base	NH(***Io***)-N(A)	CO(***Io***)-NH_2(A)	NH_2(***Io***)-CO(U)
1	***Io1***	99%, 1.87 Å, 166°	99%, 1.95 Å, 164°	94%, 2.08 Å, 159°
2	***Io4***	99%, 1.87 Å, 166°	99%, 2.00 Å, 163°	91%, 2.17 Å, 157°
3	***Io5***	99%, 1.92 Å, 165°	95%, 2.19 Å, 159°	-
4	***Io6***	99%, 1.94 Å, 165°	98%, 1.86 Å, 160°	-
5	***Io7***	99%, 1.86 Å, 166°	96%, 2.20 Å, 159°	-
8	***Io8***	99%, 1.97 Å, 160°	99%, 1.87 Å, 163°	1%

Figure 6a shows a snapshot of ***Io1*** binding, where it is observed that both aromatic rings in the ***Io*** nucleobase retain a fairly planar conformation with only the amide NH_2 twisting out of plane to make the third hydrogen bond to U. This is consistent with ab initio calculations performed on this triplex in our previous studies.[16b] An additional H-bond with 96% probability, average length 1.98 Å, and angle 160° was observed between the amide carbonyl and the NH_2 group of cytosine (C) in the adjacent **M***G-C triplet in the cases of ***Io1*** and ***Io4***. This interaction formally can improve affinity, but extra amide functionality also could potentially introduce nonselective interactions which may impact nucleobase arrangement within the triplex. In contrast, dynamics for ***Io5*** (entry 3 and Figure 6b) suggest that the distal aromatic ring has a propensity to twist and is more dynamic. This is unsurprising given that there is not an additional H-bond to lock the ring in place, but it does contrast with the binding data that displays moderately improved binding of ***Io5*** relative to ***Io1***, a result that was unexpected if planarity would be disrupted significantly. However, modeling also suggests that there is enough space in the major groove to accommodate twisted bases and that allows them to retain their affinity. With this propensity for twisting, rotation about the N-aryl bond in ***Io7*** and ***Io8*** positions the naphthyl and the cationic benzylammonium moieties outside of the helix and towards the anionic RNA backbone. This may lead to the unfavorable hydrophobic interactions with solvent in the case of the naphthyl group in ***Io7***, which help explain solubility challenges. In fact, dynamic simulations with ***Io8*** led to twisting of the benzylammonium group away from the Hoogsteen face (resulting in only 1% probability of H-bonding between the ammonium moiety and the uracil C-4 carbonyl) and into a conformation where the ammonium moiety formed an internal hydrogen bond with isoorotamide carbonyl and with solvent (Figure 6c). This helps explain non-specific binding when multiple ***Io8*** residues are employed due to the cationic nature of the substituents being exposed.

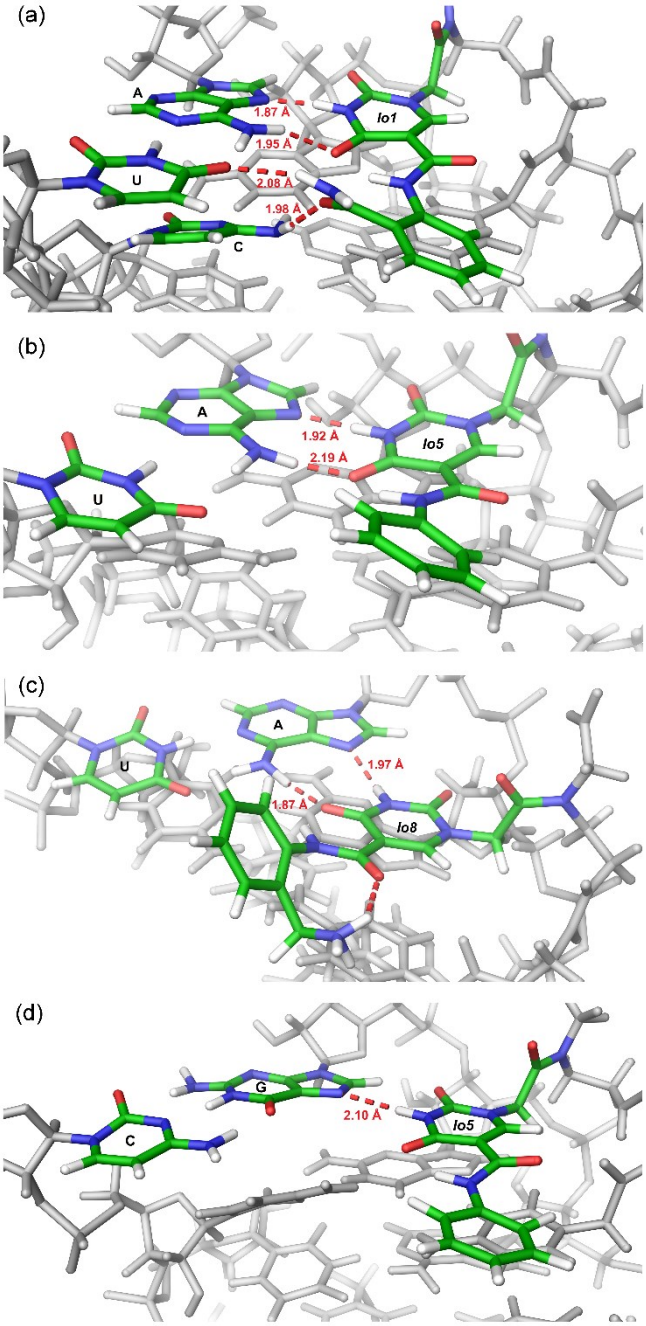


Figure 6. Major groove view of hydrogen-bonding interactions in the (a) ***Io1****A-U triplet (b) ***Io5****A-U triplet (c) ***Io8****A-U triplet and (d) ***Io5****G-C triplet from molecular dynamics simulations of the **PNA1-HRP1** triplex model. The average hydrogen bonding distance observed during molecular dynamics simulations are highlighted in red. Carbon, hydrogen, oxygen, and nitrogen atoms are labeled in green, white, red, and blue respectively.

Additionally, molecular dynamics simulation of **PNA 1**(X = ***Io5***)-**HRP3**, shined the light on ***Io*** interactions with guanine (G) that result in loss of selectivity. It was found, that the isoorotamide N3 forms a reasonable hydrogen bond with N7 of guanine on the Hoogsteen face (Figure 6d). This hydrogen bond is apparently responsible for the moderate off-target affinity of ***Io*** bases to guanine nucleobases. Dynamics simulations reveal that the carbonyl oxygens on the pyrimidine ring of ***Io*** do not clash with G. This is further supported by the observation that canonical T also has some affinity to G[11c, 16c]

Finally, in the context of **HRP5**-**PNA3** triplexes, it was observed through computations that simultaneous to the planar Hoogsteen-bonding interactions with A-U, four consecutive ***Io4*** bases formed additional stabilizing interactions through hydrogen bonding. The secondary amide NH group in ***Io4*** formed hydrogen bonds with uracil carbonyl and the primary amide NH_2 was stabilized by the adjacent primary amide carbonyl group of the adjacent ***Io4*** base (Figure 7). The later interactions introduced three extra hydrogen bonds to the system, potentially contributing to its overall stabilization. This supports the data in Table 2, showing the highest triplex stability for **PNA3** containing four ***Io4*** residues. Such stabilizing interactions may help explain the cooperative effects observed when consecutive ***Io4*** residues exist and why triplex stability appears diminished with ***Io5*** compared to ***Io4*** in the **HRP5** model.

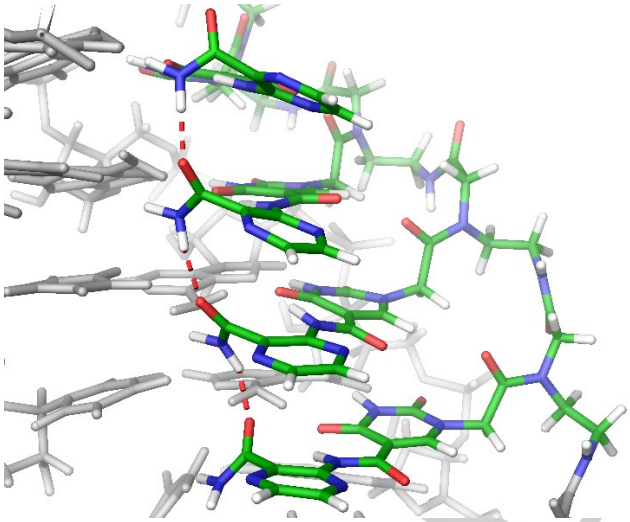


Figure 7. Representation of hydrogen-bonding interactions in the **HRP5-PNA3** triplex with ***Io4*** nucleobases. Stabilization by sequential ***Io4*** primary amides would not be present with other ***Io*** nucleobases lacking the amide moiety.

Conclusion

This study reveals improvements to A-binding using new isoorotic acid derivatives ***Io5***-***Io8*** that afford affinity and selectivity in recognition of dsRNA using PNA as a TFO. Moreover, while these new bases were primarily intended to help inform design of future nucleobases, several interesting conclusions emerged from this study that provide insight instructing design of new U-recognizing monomers. Most obvious is that results for ***Io5*** and ***Io6*** support the notion that the 'third hydrogen-bond' in ***Io1***-***Io4*** is less significant than initially expected. Computations suggest that such a hydrogen-bond is forming, though binding data reveal that the energetic contribution of that hydrogen bond is compensated by other subtle interactions. ***Io5*** is symmetrical about the aryl-nitrogen bond and would require less preorganization than any of the other ***Io*** derivatives. This entropic advantage could lead to improvements in binding affinity. This is especially notable given the different local minima found in the ***Io6*** ground state computations (*vide supra*).

While heterocycle extensions of U-type bases have previously exhibited non-planarity in modeling,[25] ***Io*** bases are unique in that the isoorotamide carbonyl remains conjugated with the pyrimidine heterocycle in all our modeled examples. This effect would not only provide extended conjugation for stacking but may aid in decreasing the pK_a of the imino nitrogen on the pyrimidine. It is reported that lowering the pK_a of pyrimidines using electron withdrawing substituents at the 5-position may improve Hoogsteen hydrogen bonding interactions with A,[26] and all ***Io*** derivatives show improved binding relative to T. It is plausible that the conjugated carbonyl improves affinity through this type of pK_a lowering, but the lack of correlation between binding strength and aryl ring electronics in the ***Io*** series is a result of multiple additional factors at play.

As discussed earlier, another important finding from the studies of ***Io5*** and ***Io7*** is that naphthyl and even too many phenyl groups can become limiting in the synthesis of PNA and disrupt binding affinity with RNA due to hydrophobic effects. This may require the incorporation of polar functional groups in the aromatic rings that can aid in solubility of PNA while interacting with external solvent upon triplex formation. Still, the seemingly simple differences between ***Io5*** and ***Io6*** and the synthetic challenges associated with the latter underscore the subtle complexity of this conclusion.

Finally, though the success of cationic PNA using M residues has a made a significant impact on the field of RNA recognition using synthetically modified PNA,[11c] the results from PNA incorporating cationic ***Io8***, or V[16c] residues, inform that there is likely a limit to the cationic character PNA may exhibit before non-specific binding becomes problematic. For example, cationic ***Io8*** showed impressive binding affinity when used as a single modification in PNA. However, when four ***Io8*** residues were incorporated into **PNA3** along with four adjacent M residues, it is possible that an abundance of basic residues and the potentially high number of resulting cationic charges disrupted the necessary protonation state of M required for strong and selective binding. In fact, a study on DNA triplex stability by Lane and Brown reveals that internal cytosine bases have higher pK values than terminal bases and protonation state effects triplex stability.[27] Given that T_m values looked encouraging in our system, but ITC data revealed a large heat release upon titration for both the matched and mismatched hairpins, it is reasonable to conclude that PNA-RNA aggregation takes place but unlikely through the expected triple helix. Future PNA nucleobases will be designed with the idea that interspersed cationic and neutral heterobases is most likely ideal for PNA sequences with fewer cationic bases, and ***Io8*** may provide a suitable or even improved alternative to T in some circumstances.

Results from this study ultimately reinforce that ***Io*** nucleobases provide an excellent scaffold for binding to A. Further, this work supports the notion that the third hydrogen bond to U in the extended nucleobase does not seem to be critical for base-triple formation. Taken together, these results will inform future nucleobase design towards the ultimate goal of 'U-recognition' using bases that span the entire major groove, bypass the pyrimidine, and take advantage of excellent recognition for A.

Experimental Section

General Synthetic Details: Solvents and reagents were purchased from commercial suppliers and used without further purification except for diisopropylethylamine (DIPEA) which was distilled with CaH_2, and aniline which was distilled under vacuum with CaH_2. DIPEA and DMF were stored under an N_2 atmosphere using Schlenk techniques. DCM and toluene were obtained from a dry solvent system by passing over molecular sieves. THF was dried using a sodium metal and benzophenone still. All reactions were performed with oven dried glassware under an inert atmosphere (N_2 gas). When reactions required inert atmosphere, evacuation of the gaseous headspace was performed under vacuum and refilled with dry N_2. This cycle was commonly repeated 3 times. Analytical thin layer chromatography (TLC) was carried out using silica gel 60 F_{254} glass. TLCs were viewed through use of a UV lamp (254 and/or 365 nm). Flash chromatography was carried out using SiliaFlash® F60 silica (40-63 µm, 230-400 mesh) or automated using a *Teledyne ISCO* CombiFlash® EZ Prep flash chromatography system with RediSep Gold HP silica columns. IR spectra were obtained using a Nicolet iS50 FT-IR with diamond ATR accessory. ^{1}H and ^{13}C NMR spectra were obtained with a Varian 400-MR spectrometer and samples prepared in 5 mm OD tubes with deuterated solvent. Chemical shifts (δ) were reported relative to the solvent peak ($CDCl_3$ or DMSO-d_6) as a reference. The following abbreviations (or combinations thereof) were used to describe multiplicities: s = singlet, d = doublet, t = triplet, q = quartet, m = multiplet, br = broad. Coupling constants (J) were reported in Hertz (Hz). ^{13}C NMR spectra were obtained with proton decoupling. Chemical shifts for peaks overlapping with solvent were determined by HSQC. High Resolution Mass Spectrometry (HRMS) analyses used positive electrospray ionization (ESI^+) on an Orbitrap Q-Exactive instrument.

PNA Synthesis: All PNA sequences were synthesized using our standard procedures[24] on an automated PNA Expedite 8909 synthesizer at 2 µmol scale on NovaSyn TG Sieber resin. Commercial PNA-T-monomer was purchased from Link Technologies. M monomer was prepared using the reported synthetic route from our group.[10] PNA sequences were cleaved from the solid support using standard cleavage conditions using 0.6 mL of 5% m-cresol in TFA for 2 hours using the two-syringe pull-push method. The cocktail was collected in an Eppendorf tube and the resin was washed again with an additional 0.2 mL of the fresh cleaving cocktail. Crude PNA (200 µL in each Eppendorf) was precipitated by the addition of chilled diethyl ether (~1.0 mL) followed by centrifugation. Crude PNA was analyzed and purified by LCMS and then quantified (see Supporting Information).

Isothermal Titration Calorimetry experiments were performed on a MicroCal iTC200 instrument at 25 ˚C in phosphate buffer (pH 7.4) containing 2 mM $MgCl_2$, 90 mM KCl, 10 mM NaCl, 50 mM potassium phosphate. Aliquots (2.45 µL) of 75 µM PNA solution were sequentially injected from a 40 µL rotating syringe (750 rmp) into 270 µL of 10 µM RNA hairpin solution. Complete data tables and representative ITC titration traces are shown in the Supporting Information.

UV Thermal Melting Experiments were performed on a Shimadzu UV-2600 spectrophotometer equipped with a TMSPC-8 temperature controller. Each PNA-dsRNA complex was the average of five replicates at 18 µM in phosphate buffer (2 mM $MgCl_2$, 90 mM KCl, 10 mM NaCl, and 50 mM potassium phosphate at pH 7.4). Absorbance versus temperature profiles were measured at 300 nm. The temperature was increased at a rate of 1°C per minute. The melting temperatures were obtained using Shimadzu LabSolutions T_m Analysis software version 1.31. Complete data tables are shown in the Supporting Information.

Supporting Information

The authors have cited additional references within the Supporting Information.[28]

Acknowledgements

This work was supported by the U.S. National Science Foundation grants CHE-2107911 to J.A.M. and CHE-2107900 to E.R and the Latvian Institute of Organic Synthesis internal research fund (IG-2023-08) to V.B. We thank Ilze Kumpina and Christopher Ryan for help with the ITC data analysis.

Keywords: Hoogsteen hydrogen bonding • nucleobases • peptide nucleic acids • RNA recognition • triple-helix

[1] The Encode Project Consortium, *Nature* **2012**, *489*, 57-74.

[2] a) T. R. Mercer, J. S. Mattick, *Nat. Struct. Mol. Biol.* **2013**, *20*, 300-307; b) K. N. Smith, S. C. Miller, G. Varani, J. M. Calabrese, T. Magnuson, *Genetics* **2019**, *213*, 1093-1110; c) R.-W. Yao, Y. Wang, L.-L. Chen, *Nat. Cell Biol.* **2019**, *21*, 542-551; d) X.-W. Wang, C.-X. Liu, L.-L. Chen, Q. C. Zhang, *Nat. Chem. Biol.* **2021**, *17*, 755-766.

[3] a) J. Y. Wang, P. Pausch, J. A. Doudna, *Nat. Rev. Microbiol.* **2022**, *20*, 641-656; b) E. Rozners, *J. Am. Chem. Soc.* **2022**, *144*, 12584-12594.

[4] a) X. Shen, D. R. Corey, *Nucleic Acids Res.* **2018**, *46*, 1584-1600; b) M. Egli, M. Manoharan, *Nucleic Acids Res.* **2023**, *51*, 2529-2573.

[5] a) N. K. Conrad, *Wiley Interdiscip. Rev.: RNA* **2014**, *5*, 15-29; b) J. A. Brown, *Wiley Interdiscip. Rev.: RNA* **2020**, *11*, e1598.

[6] P. E. Nielsen, M. Egholm, R. H. Berg, O. Buchardt, *Science* **1991**, *254*, 1497-1500.

[7] a) N. Brodyagin, M. Katkevics, V. Kotikam, C. A. Ryan, E. Rozners, *Beilstein J. Org. Chem.* **2021**, *17*, 1641-1688; b) X. Zhan, L. Deng, G. Chen, *Biopolymers* **2022**, *113*, e23476.

[8] a) F. Wojciechowski, H. R. E. Hudson, *Curr. Top. Med. Chem.* **2007**, *7*, 667-679; b) A. Y. Shaikh, F. Björkling, P. E. Nielsen, H. Franzyk, *Eur. J. Org. Chem.* **2021**, 2792-2801.

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

[10] T. Zengeya, P. Gupta, E. Rozners, *Angew. Chem. Int. Ed.* **2012**, *51*, 12593-12596.

[11] a) O. Muse, T. Zengeya, J. Mwaura, D. Hnedzko, D. W. McGee, C. T. Grewer, E. Rozners, *ACS Chem. Biol.* **2013**, *8*, 1683-1686; b) D. Hnedzko, D. W. McGee, Y. A. Karamitas, E. Rozners, *RNA* **2017**, *23*, 58-69; c) C. A. Ryan, N. Brodyagin, J. Lok, E. Rozners, *Biochemistry* **2021**, *60*, 1919-1925.

[12] V. Kotikam, S. D. Kennedy, J. A. MacKay, E. Rozners, *Chem. - Eur. J.* **2019**, *25*, 4367-4372.

[13] a) Y. Zhou, E. Kierzek, Z. P. Loo, M. Antonio, Y. H. Yau, Y. W. Chuah, S. Geifman-Shochat, R. Kierzek, G. Chen, *Nucleic Acids Res.* **2013**, *41*, 6664-6673; b) G. Devi, Z. Yuan, Y. Lu, Y. Zhao, G. Chen, *Nucleic Acids Res.* **2014**, *42*, 4008-4018; c) A. A. L. Ong, D.-F. K. Toh, M. S. Krishna, K. M. Patil, K. Okamura, G. Chen, *Biochemistry* **2019**, *58*, 3444-3453; d) A. A. L. Ong, D.-F. K. Toh, K. M. Patil, Z. Meng, Z. Yuan, M. S. Krishna, G. Devi, P. Haruehanroengra, Y. Lu, K. Xia, K. Okamura, J. Sheng, G. Chen, *Biochemistry* **2019**, *58*, 1319-1331; e) T. Sato, N. Sakamoto, S. Nishizawa, *Org. Biomol. Chem.* **2018**, *16*, 1178-1187; f) T. Sato, Y. Sato, S. Nishizawa, *Biopolymers* **2022**, *113*, e23474; g) K. T. Kim, D. Chang, N. Winssinger, *Helv. Chim. Acta* **2018**, *101*, e1700295; h) V.

Tähtinen, L. Granqvist, M. Murtola, R. Strömberg, P. Virta, *Chem. - Eur. J.* **2017**, *23*, 7113-7124.

[14] a) T. Endoh, D. Hnedzko, E. Rozners, N. Sugimoto, *Angew. Chem. Int. Ed.* **2016**, *55*, 899-903; b) R. Y. Puah, H. Jia, M. Maraswami, D. F. Kaixin Toh, R. Ero, L. Yang, K. M. Patil, A. A. Lerk Ong, M. S. Krishna, R. Sun, C. Tong, M. Huang, X. Chen, T. P. Loh, Y. G. Gao, D. X. Liu, G. Chen, *Biochemistry* **2018**, *57*, 149-159; c) J. Kesy, K. M. Patil, S. R. Kumar, Z. Shu, H. Y. Yong, L. Zimmermann, A. A. L. Ong, D.-F. K. Toh, M. S. Krishna, L. Yang, J.-L. Decout, D. Luo, M. Prabakaran, G. Chen, E. Kierzek, *Bioconjugate Chem.* **2019**, *30*, 931-943; d) T. Endoh, N. Brodyagin, D. Hnedzko, N. Sugimoto, E. Rozners, *ACS Chem. Biol.* **2021**, *16*, 1147-1151.

[15] I. Kumpina, V. Baskevics, K. D. Nguyen, M. Katkevics, E. Rozners, *Chembiochem* **2023**, e202300291.

[16] a) I. Kumpina, N. Brodyagin, J. A. MacKay, S. D. Kennedy, M. Katkevics, E. Rozners, *J. Org. Chem.* **2019**, *84*, 13276-13298; b) N. Brodyagin, A. L. Maryniak, I. Kumpina, J. M. Talbott, M. Katkevics, E. Rozners, J. A. MacKay, *Chem. - Eur. J.* **2021**, 4332-4335; c) C. A. Ryan, V. Baskevics, M. Katkevics, E. Rozners, *Chem. Commun.* **2022**, *58*, 7148-7151.

[17] D.-F. K. Toh, G. Devi, K. M. Patil, Q. Qu, M. Maraswami, Y. Xiao, T. P. Loh, Y. Zhao, G. Chen, *Nucleic Acids Res.* **2016**, *44*, 9071-9082.

[18] T. Maegawa, Y. Fujiwara, T. Ikawa, H. Hisashi, Y. Monguchi, H. Sajiki, *Amino Acids* **2009**, *36*, 493-499.

[19] M. Kurfiřt, Č. Š. a. Lucie, P. Cuřínová, V. Hamala, J. Karban, *J. Org. Chem.* **2021**, *86*, 5073-5090.

[20] A. Devos, J. Remion, A.-M. Frisque-Hesbain, A. Colens, L. Ghosez, *Chem. Commun.* **1979**, 1180-1181.

[21] N. A. Magnus, T. M. Braden, J. Y. Buser, A. C. DeBaillie, P. C. Heath, C. P. Ley, J. R. Remacle, D. L. Varie, T. M. Wilson, *Org. Process Res. Dev.* **2012**, *16*, 830-835.

[22] F. Debaene, J. A. Da Silva, Z. Pianowski, F. J. Duran, N. Winssinger, *Tetrahedron* **2007**, *63*, 6577-6586.

[23] S. Langolf, U. Machon, M. Ehlers, W. Sicking, T. Schirmeister, C. Büchhold, C. Gelhaus, P. J. Rosenthal, C. Schmuck, *ChemMedChem* **2011**, *6*, 1581-1586.

[24] N. Brodyagin, D. Hnedzko, J. A. MacKay, E. Rozners, in *Peptide Nucleic Acids. Methods and Protocols* (Ed.: P. E. Nielsen), Springer US, New York, **2020**, pp. 157-172.

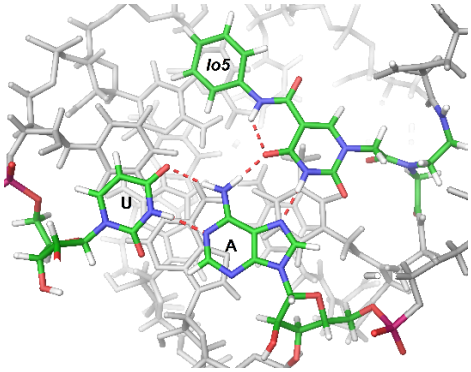
[25] M. S. Krishna, D.-F. K. Toh, Z. Meng, A. A. L. Ong, Z. Wang, Y. Lu, K. Xia, M. Prabakaran, G. Chen, *Anal. Chem.* **2019**, *91*, 5331-5338.

[26] a) T. J. Povsic, P. B. Dervan, *J. Am. Chem. Soc.* **1989**, *111*, 3059-3061; b) K. M. Patil, D.-F. K. Toh, Z. Yuan, Z. Meng, Z. Shu, H. Zhang, Alan Ann L. Ong, M. S. Krishna, L. Lu, Y. Lu, G. Chen, *Nucleic Acids Res.* **2018**, *46*, 7506-7521.

[27] J. L. Asensio, A. N. Lane, J. Dhesi, S. Bergqvist, T. Brown, *J. Mol. Biol.* **1998**, *275*, 811-822.

[28] a) C. Russo, G. Graziani, R. Cannalire, G. C. Tron, M. Giustiniano, *Green Chem.* **2022**, *24*, 3993-4003; b) T. A. Feagin, N. I. Shah, J. M. Heemstra, *J. Nucleic Acids* **2012**, *2012*, e354549.

Entry for the Table of Contents



Four isoorotamide nucleobases (***Io5-Io8***) are reported for incorporation into peptide nucleic acids aimed at RNA triple-helix formation. These extended nucleobases require only two hydrogen bonds across the Hoogsteen face of the A-U base pair for strong and selective binding, unlike previously reported extended nucleobases. Molecular dynamics simulations provide further insight into triple helical binding affinity and selectivity.

Institute and/or researcher Twitter usernames: @etowncollege; @osi.lv