

COMMUNICATION

2-Guanidyl Pyridine PNA Nucleobase for Triple-Helical Hoogsteen Recognition of Cytosine in Double-Stranded RNA

Christopher A. Ryan,^a Vladislavs Baskevics,^b Martins Katkevics,^b and Eriks Rozners ^{*a}Received 00th January 20xx,
Accepted 00th January 20xx

DOI: 10.1039/x0xx00000x

In triplex-forming peptide nucleic acid, a novel 2-guanidyl pyridine nucleobase (V) enables recognition of up to two cytosine interruptions in polypurine tracts of dsRNA by engaging the entire Hoogsteen face of C-G base pair. Ab initio and molecular dynamics simulations provided insights into H-bonding interactions that stabilized V•C-G triplets. Our results provided insights for future design of improved nucleobases, which is an important step towards the ultimate goal of recognition of any sequence of dsRNA.

Peptide nucleic acids (PNAs, Figure 1) are DNA mimics built of neutral amide backbone.¹ Because they bind complementary DNA and RNA with high affinity and specificity, PNAs have become highly useful research and diagnostic tools.² Recently, PNAs have also emerged as excellent ligands for triple helical recognition of double-stranded RNA (dsRNA).^{2, 3} Our research group showed that M-modified triplex-forming PNAs (Figure 1) bind to dsRNA at least ten-fold stronger than to the same sequence of dsDNA.^{4, 5} NMR structural studies showed that this unusually high affinity was driven by H-bonding of the PNA backbone amide N-H groups to RNA backbone phosphates.⁶ Studies by our group^{4, 5, 7} and others⁸⁻¹² have demonstrated that nucleobase-modified triplex-forming PNAs (Figure 1) are uniquely fit to bind dsRNA with high affinity and specificity over dsDNA or single-stranded nucleic acids. However, the requirement for polypurine tracts to form the Hoogsteen triple helices remains the main limitation of PNA-dsRNA recognition. In native triple-helical RNA, uridine forms Hoogsteen H-bonds with adenosine while protonated cytosine H-bonds with guanosine of the A-U and G-C base pairs, respectively.¹³ Nucleobase modifications such as 2-aminopyridine⁴ (M, Figure 1), pseudoisocytosine¹⁴ (J), and 2-thiopseudoisocytosine⁹ (L) have been developed to overcome the unfavorable protonation

of cytosine ($pK_a \sim 4.5$) in triplex-forming PNAs. Our recent comparative study showed that M••G-C triplets were significantly more stable than either J•G-C or T•A-U triplets likely due to M ($pK_a \sim 6.7$) being partially protonated at physiological conditions.⁷ However, similar attempts to overcome sequence limitations of triple helical recognition by developing modified nucleobases to form Hoogsteen-like triplets with pyrimidines of inverted T-A (or U-A) and C-G base pairs have so far given only modest results.^{15, 16}

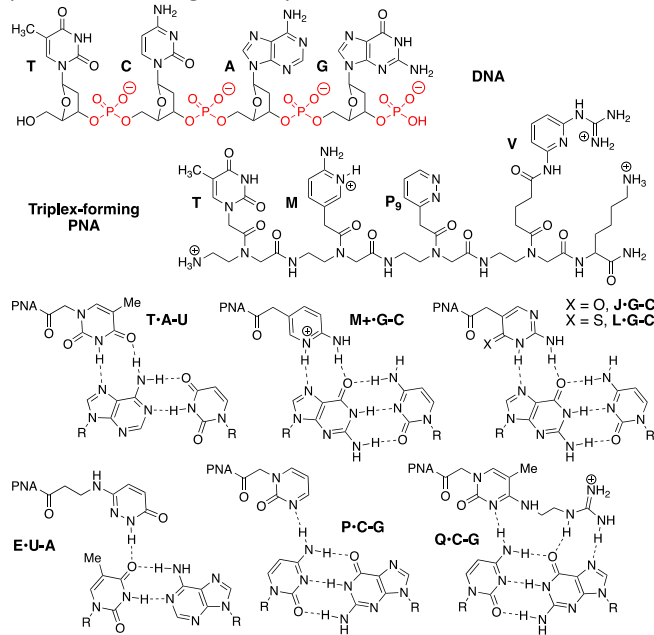


Figure 1. Structures of DNA, triplex-forming PNA, and Hoogsteen hydrogen-bonded base triplets. PNA denotes the amide backbone of peptide nucleic acid; R denotes sugar-phosphate backbone of DNA or RNA.

For recognition of T-A inversions, Nielsen and co-workers developed 3-oxo-2,3-dihydropyridazine (E, Figure 1) having a two atoms longer β -amino acid linker to PNA backbone and forming a single H-bond with T.¹⁷ For recognition of C-G inversions, pyrimidin-2-one (P)¹⁸ and guanidinyethyl-5-

^a Department of Chemistry, Binghamton University, Binghamton, NY 13902, USA.

^b Latvian Institute of Organic Synthesis, Aizkraukles 21, Riga, LV-1006, Latvia.

Electronic Supplementary Information (ESI) available: Synthesis, purification, and LC-MS characterization of PNA monomers and oligomers; UV melting and ITC results; details of computational studies; copies of ¹H and ¹³C NMR spectra of new compounds. See DOI: 10.1039/x0xx00000x

methylcytosine (Q)¹⁹ have been used in PNA. In a recent study, we showed that the 3-pyridazinyl nucleobase (P₉, Figure 1) formed stronger triplets than P providing a notable improvement in recognition of a single C-G inversion.²⁰ While Winssinger and co-workers have reported¹² formation of 13-nucleotides long PNA-dsRNA triplex having as many as six pyrimidine inversions (recognized with three E•U-A and three P•C-G triplets) under specific experimental conditions, overall, the current performance of modified nucleobases limit the stability of PNA-dsRNA triplexes under physiological conditions to a single pyrimidine interruption. In this Communication, we report that a new cationic 2-guanidyl pyridine nucleobase (V, Figure 1) recognizes up to two cytosine interruptions in a nine-nucleotide polypurine tract with affinity and sequence specificity sufficient for practical applications at physiological conditions. The design of V base originated from our study on the 2,7-diamino-1,8-naphthyridine ^{DA}N (Figure 2) nucleobase, previously reported by Ohkubo, Sekine and co-workers to recognize pyrimidine interruptions in polypurine tracts of DNA triplexes.²¹ The PNA monomers of ^{DA}N and derivatives thereof were synthesized using well-established procedures (for details, see Supplementary Information) and incorporated in PNA oligomers (Figure 2) using our previously reported methods.²² We measured the stability and sequence specificity of PNA-dsRNA triplexes using UV thermal melting at 300 nm and the model hairpins (HRP1-HRP4, Figure 2) used in our previous studies.^{7, 20}

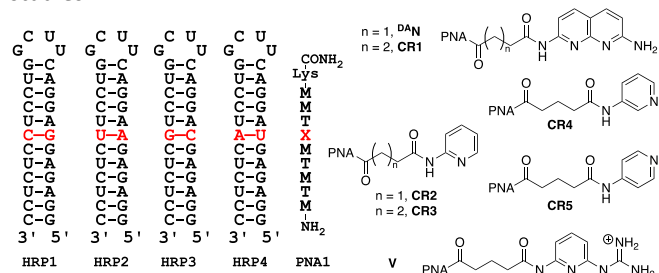


Figure 2. Structures of the RNA hairpins, PNA1, and heterocyclic nucleobases screened for recognition of the cytosine inversion in polypurine tract of HRP3.

Table 1. Binding Affinities and Sequence Selectivities of PNA1 by UV Thermal Melting.

Entry	PNA1 ^[a]	HRP1 (G)	HRP2 (A)	HRP3 (C)	HRP4 (U)
1	X = T ^[b]	46.4 ± 0.5	69.6 ± 0.8	35.4 ± 0.4	34.6 ± 0.2
2	X = P ₉ ^[c]	36.2 ± 0.3	36.5 ± 0.3	48.5 ± 0.2	36.4 ± 0.2
3	X = DAN	40.8 ± 0.4	50.6 ± 0.6	39.1 ± 0.5	35.2 ± 0.3
4	X = CR1	32.2 ± 0.4	40.6 ± 0.2	51.1 ± 0.3	35.6 ± 0.2
5	X = CR2	NB ^[d]	55.5 ± 0.3	40.3 ± 0.3	34.0 ± 0.4
6	X = CR3	30.3 ± 0.5	39.1 ± 0.3	51.5 ± 0.3	34.5 ± 0.2
7	X = CR4	31.8 ± 0.3	37.1 ± 0.5	51.2 ± 0.3	34.6 ± 0.3
8	X = CR5	32.6 ± 0.4	37.8 ± 0.1	52.5 ± 0.3	35.7 ± 0.3
9	X = V	36.4 ± 0.3	42.3 ± 0.5	60.3 ± 0.4	37.0 ± 0.3

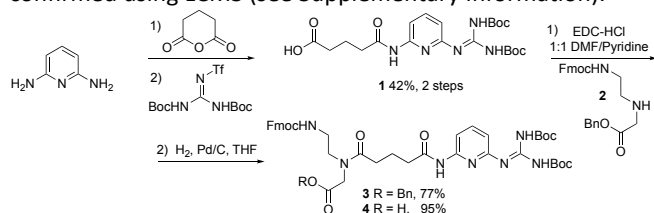
[a] UV thermal melting temperatures (T_m , °C) are averages of five experiments ± the standard deviation measured at 300 nm and 18 μM of each dsRNA and PNA in 50 mM potassium phosphate buffer (pH 7.4) containing 2 mM MgCl₂, 90 mM KCl, and 10 mM NaCl. The results for matched target dsRNA are highlighted in bold. [b] Benchmark data for all purine triplexes are from ref.⁷ [c] Benchmark data for previous best nucleobase (P₉) from ref.²⁰ [d] No melting curve observed.

The T_m = 69.6 °C of the triplex between PNA1 X = T and HRP2 (uninterrupted polypurine tract) and T_m = 48.5 °C of the triplex between PNA1 X = P₉ (currently our best nucleobase for recognition of cytosine interruption) and HRP3 served as the benchmarks for desired stability and current state of the art, respectively (Table 1).

In contrast to results reported for DNA triplexes,²¹ ^{DA}N-modified PNA1 showed increased affinity for recognition of A in HRP2 (Table 1, entry 3). Extension of the linker between nucleobase and PNA backbone by one carbon in CR1 shifted the recognition in favor of C in HRP3 (Table 1, entry 4). Next, we explored the importance of H-bond donors and acceptors of ^{DA}N simplifying the ^{DA}N heterocycle to 2-amidopyridine in CR2 and CR3. Somewhat surprisingly, this had relatively little effect on the binding affinity; PNAs modified with CR2 and CR3 showed similar binding properties as PNAs modified with ^{DA}N and CR1 nucleobases, respectively (Table 1, entries 5 and 6). Furthermore, moving the ring nitrogen to meta and para positions in CR4 and CR5 did not change the binding affinity or specificity (Table 1, entries 7 and 8).

Ab initio calculations at the B3LYP/6-31G+(d,p) level of theory showed that in CR1•C-G triplet, the CR1 nucleobase formed two H-bonds, one between the C=O of linker and -NH₂ of cytosine and the second between -NH₂ of CR1 and N7 of guanosine (Figure S22A). In contrast, CR3 formed only one H-bond between the C=O of linker and -NH₂ of cytosine (Figure S22B). Collectively, the experimental and computational results suggested that in CR2-CR5 series, only the C=O of the amide group connecting the heterocycle and PNA linker was participating in H-bonding interactions with RNA base pairs.

Inspired by the design of the Q base (Figure 1)¹⁹ and following up on our previous study to recognize the entire Hoogsteen face with extended nucleobases,¹⁶ we added a guanidine group to CR3 creating the V nucleobase (Figure 2). Synthesis of the PNA monomer started with ring opening of glutaric anhydride with 2,6-diaminopyridine followed by installation of Boc-protected guanidine (Scheme 1). Coupling of carboxylic acid **1** to PNA backbone **2** gave, after debenzoylation, the final monomer **4**. Because the NMR spectra PNA monomers were complicated by the presence of rotamers, the purity and identity were also confirmed using LCMS (see Supplementary Information).



Scheme 1. Synthesis of Fmoc/Boc-protected monomer to incorporate V base in PNA.

Addition of the guanidine group increased the on-target affinity of V-modified PNA1 for HRP3 (Table 1, entry 9) while maintaining good sequence specificity. This result was confirmed using isothermal titration calorimetry (Table 2). Compared with our benchmarks in entries 1 and 2 in Tables 1 and 2, V base significantly improved C-G recognition over P₉ and CR3. The UV melting showed that V base was binding somewhat

weaker to C-G ($T_m = 60.3^\circ\text{C}$) than T to A-T ($T_m = 69.6^\circ\text{C}$), while ITC results suggested that the affinities were comparable.

Table 2. Binding Affinities and Sequence Selectivities ($K_a \times 10^6 \text{ M}^{-1}$) of PNAs by ITC.

Entry	PNA1 ^[a]	HRP1 (G)	HRP2 (A)	HRP3 (C)	HRP4 (U)
1	X = T ^[b]	1.9 ± 0.1	12 ± 1	0.9 ± 0.1	0.4 ± 0.1
2	X = P ₉ ^[c]	1.5 ± 0.1	1.4 ± 0.1	6.7 ± 0.4	1.3 ± 0.2
3	X = CR3	ND ^[d]	ND ^[d]	7.9 ± 0.3	ND ^[d]
4	X = V	1.1 ± 0.2	1.7 ± 0.1	14 ± 0.3	1.0 ± 0.2

[a] Association constants ($K_a \times 10^6 \text{ M}^{-1}$) are averages of three experiments $\pm$ the standard deviation, for binding of PNAs to the respective hairpins in 50 mM potassium phosphate buffer (pH 7.4) containing 2 mM MgCl₂, 90 mM KCl, and 10 mM NaCl at 25°C . The results for matched target dsRNA are highlighted in bold. [b] Benchmark data for all purine triplexes are from ref.⁷ [c] Benchmark data for previous best nucleobase (P₉) from ref.²⁰ [d] ND – not determined.

Next, we evaluated the ability of V to recognize several C-G interruptions in the polypurine tracts of dsRNA. The affinity of V-modified PNA2 for HRP5 having two consecutive C-G base pairs (Figure 3) decreased compared to that of PNA1 for HRP3 (Tables 1 and 2) by both UV melting and ITC. However, the V-modified PNA2 bound to HRP5 significantly stronger than P₉-modified PNA2 in our previous study.²⁰ Binding affinity was further decreased for PNA3 having the two V bases moved apart in the sequence. Binding of PNA2 and PNA3 was sequence selective for their matched HRP5 and HRP6 as demonstrated by significantly lower stability of mismatched triplexes PNA2-HRP6 ($T_m = 40.1 \pm 0.8^\circ\text{C}$), and PNA3-HRP5 ($T_m = 30.0 \pm 0.4^\circ\text{C}$); neither PNA2, nor PNA3 had detectable affinity for HRP7 (Figures S17). In contrast, incorporation of three V bases in PNA4 led to non-specific binding to all three hairpins (HRP5-HRP7) with stoichiometry of more than two PNA molecules for each dsRNA target (Figures S20 and S21). Such a binding mode ($n > 2$) is most likely caused by non-specific binding of guanidine to RNA backbone phosphates, and may be driven by the highly cationic nature of PNA4 having up to eight positively charged nucleobases interrupted by a single T.

C U	C U	C U	C U	C U	C U
G-C	CONH ₂	G-C	CONH ₂	G-C	CONH ₂
U-A	Lys	U-A	Lys	U-A	Lys
C-G	M	C-G	M	C-G	M
C-G	M	C-G	M	C-G	M
U-A	T	C-G	M	U-A	T
C-G	X	C-G	M	C-G	X
C-G	X	C-G	M	C-G	X
C-G	M	C-G	M	C-G	X
C-G	M	C-G	M	C-G	X
U-A	T	U-A	T	C-G	X
C-G	M	C-G	M	C-G	M
C-G	M	C-G	M	C-G	M
3' 5'	NH ₂	3' 5'	NH ₂	3' 5'	NH ₂
HRP5	PNA2	HRP6	PNA3	HRP7	PNA4

Figure 3. Structures of the RNA hairpins and PNAs for recognition of several cytosine inversions in polypurine tracts of dsRNA. ND – not determined.

Ab initio geometry optimization (Figure 4) showed that the V•C-G triplet was planar, and the V base adopted a “closed” conformation rigidified by intermolecular H-bond between the -NH₂ of guanidine and the pyridine nitrogen, and a longer H-bond between the -NH₂ of guanidine and the C=O of linker. V base used the C=O of the linker and both -NH₂ groups of guanidine to engage the entire Hoogsteen face of C-G base pair.

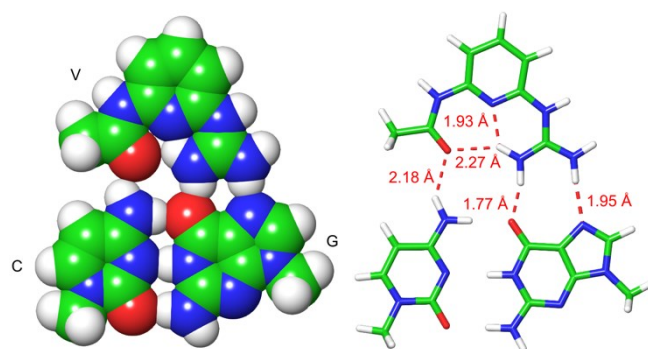


Figure 4. Geometry optimization of V•C-G triplet using B3LYP 6-31G*(d, p). Carbon, hydrogen, oxygen, and nitrogen atoms are labelled in green, white, red, and blue, respectively.

To study if these H-bonding geometries are maintained in the context of PNA-dsRNA triple helix, we constructed a model of PNA1-HRP3 using a template of PNA-dsRNA triplex constructed in our previous NMR structural studies⁶ (for details, see Supplementary Information). The conformation of the V•C-G triplet was simulated by running 50 ns unrestricted Desmond molecular dynamics and analyzing the last 10 ns of simulation, when the system had stabilized. The simulations showed that the V base maintained the conformation observed in the *ab initio* optimization (Figure 5); however, the H-bonding pattern to C-G base pair slightly changed. In the context of a triple helix, molecular dynamics simulations showed a dynamic V•C-G structure where V base formed two stable H-bonds to the Hoogsteen face of C-G: the C=O of V-base linker H-bonded to the cytosine -NH₂ (1.71–2.17 Å) and one of the guanidine -NH₂ H-bonded to the guanosine C=O (1.57–2.11 Å). The other guanidine -NH₂ only occasionally engaged in H-bonding with the guanosine N7 as illustrated by a longer distance and wider range of 1.97 to 3.55 Å. The intramolecular H-bond between the guanidine -NH₂ and the C=O of linker observed in *ab initio* triplet (Figure 4) was not detected in molecular dynamics simulations, while the H-bond between the pyridine nitrogen and the guanidine -NH₂ remained. Interestingly, a third strong H-bond (1.68–2.22 Å) was formed between the N-H of V base linker to the C=O of linker connecting the adjacent M base to PNA backbone (Figure 5).

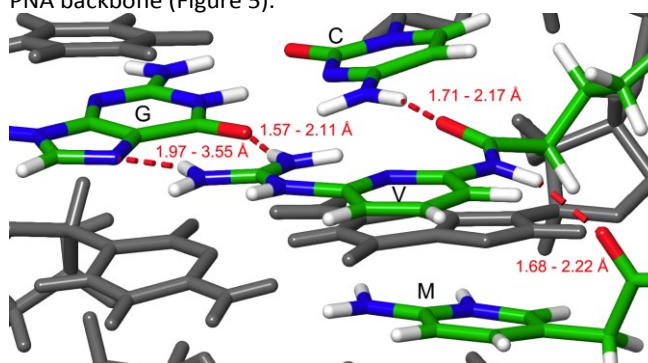


Figure 5. Major groove view of hydrogen-bonding interactions in V•C-G triplet from molecular dynamics simulations of the PNA1-HRP3 triplex model. The hydrogen-bonding interactions and distance ranges observed during molecular dynamics simulations are highlighted in red. Carbon, hydrogen, oxygen, and nitrogen atoms are labelled in green, white, red, and blue, respectively.

Collectively, the *ab-initio* calculations (Figure 4) and molecular dynamics simulations (Figure 5) revealed that V recognized the entire Hoogsteen face of the C-G base pair using two strong H-bonds and one weaker electrostatic interaction. The overall conformation was further stabilized by an H-bond between the amides of nucleobase linkers: the N-H of V base to the C=O of adjacent M base.

The present study significantly improved recognition of C-G base pairs that form cytosine interruptions in polypurine tracts of dsRNA. Compared to our previous state of the art nucleobase P₉ (Figure 3), the affinity of V-modified PNA 9-mers recognizing two consecutive cytosine interruptions increased ~3-fold (K_a by ITC) and ~15 °C (T_m by UV melting) allowing triplex formation under physiological salt and pH with T_m > 10 °C higher than 37 °C. While this is significant improvement, recognition of two and more cytosine interruptions is sequence dependent as shown by lower binding affinity of PNA3 having two separated V bases and loss of sequence specificity for PNA4 modified with three V bases. Most likely, the highly cationic nature of PNA4 and high affinity of guanidine for RNA phosphates caused the non-specific binding, which suggested that future improvement of the V base should focus on reducing the positive charge while maintaining the favorable H-bonding scheme. Meanwhile, sequences of triplex-forming PNAs containing several V bases will need to be carefully optimized to balance affinity and specificity.

Ab-initio calculations and molecular dynamics simulations provided important insights into design principles for future optimization of PNA nucleobases. The intermolecular H-bonding that stabilized the “closed” conformation of V base was similar to what we previously observed for an isosrotic acid derived extended nucleobase with high affinity for A-U base pairs.²³ This common pattern suggested that minimizing rotational freedom (hence, reducing unfavorable entropy) may be a general feature of successful designs for extended nucleobases recognizing the entire Hoogsteen face of Watson-Crick base pairs. Another notable feature was the H-bonding between the amides of nucleobase linkers (the N-H of V base to the C=O of adjacent M base) that positioned the V base favorably for C-G recognition and likely helped reducing unfavorable entropy of the relatively long linker connecting V base to PNA's backbone. Collectively, our experimental and computational results validate the 2-guanidyl pyridine V as a novel and improved nucleobase for recognition of cytosine interruptions in polypurine tracts of dsRNA and provide insights for future optimization of next generation designs of PNA nucleobases.

We thank Ilze Kumpina and Brandon R Tessier for help with the ITC data analysis and Dr. Nikita Brodyagin and Dr. James A. MacKay for their conceptual contribution to this work. This work was supported by National Science Foundation (CHE-2107900 to E.R.) and Latvian Institute of Organic Synthesis internal research fund (B-06-K to V.B.).

Conflicts of interest

There are no conflicts to declare.

References

1. P. E. Nielsen, M. Egholm, R. H. Berg and O. Buchardt, *Science*, 1991, **254**, 1497-1500.
2. N. Brodyagin, M. Katkevics, V. Kotikam, C. A. Ryan and E. Rozners, *Beilstein J. Org. Chem.*, 2021, **17**, 1641-1688.
3. X. Zhan, L. Deng and G. Chen, *Biopolymers*, 2022, **113**, e23476.
4. T. Zengeya, P. Gupta and E. Rozners, *Angew. Chem., Int. Ed.*, 2012, **51**, 12593-12596.
5. D. Hnedzko, D. W. McGee, Y. A. Karamitas and E. Rozners, *RNA*, 2017, **23**, 58-69.
6. V. Kotikam, S. D. Kennedy, J. A. MacKay and E. Rozners, *Chem. Eur. J.*, 2019, **25**, 4367-4372.
7. C. A. Ryan, N. Brodyagin, J. Lok and E. Rozners, *Biochemistry*, 2021, **60**, 1919-1925.
8. Y. Zhou, E. Kierzek, Z. P. Loo, M. Antonio, Y. H. Yau, Y. W. Chuah, S. Geifman-Shochat, R. Kierzek and G. Chen, *Nucleic Acids Res.*, 2013, **41**, 6664-6673.
9. G. Devi, Z. Yuan, Y. Lu, Y. Zhao and G. Chen, *Nucleic Acids Res.*, 2014, **42**, 4008-4018.
10. T. Sato, N. Sakamoto and S. Nishizawa, *Org. Biomol. Chem.*, 2018, **16**, 1178-1187.
11. V. Tahtinen, L. Granqvist, M. Murtola, R. Stroemberg and P. Virta, *Chem. Eur. J.*, 2017, **23**, 7113-7124.
12. K. T. Kim, D. Chang and N. Winssinger, *Helv. Chim. Acta*, 2018, **101**, e1700295.
13. G. Devi, Y. Zhou, Z. Zhong, D.-F. K. Toh and G. Chen, *Wiley Interdisciplinary Reviews: RNA*, 2015, **6**, 111-128.
14. M. Egholm, L. Christensen, K. L. Dueholm, O. Buchardt, J. Coull and P. E. Nielsen, *Nucleic Acids Res.*, 1995, **23**, 217-222.
15. Y. Hari, S. Obika and T. Imanishi, *Eur. J. Org. Chem.*, 2012, **2012**, 2875-2887.
16. I. Kumpina, N. Brodyagin, J. A. MacKay, S. D. Kennedy, M. Katkevics and E. Rozners, *J. Org. Chem.*, 2019, **84**, 13276-13298.
17. A. B. Eldrup, O. Dahl and P. E. Nielsen, *J. Am. Chem. Soc.*, 1997, **119**, 11116-11117.
18. P. Gupta, T. Zengeya and E. Rozners, *Chem. Commun.*, 2011, **47**, 11125-11127.
19. D.-F. K. Toh, G. Devi, K. M. Patil, Q. Qu, M. Maraswami, Y. Xiao, T. P. Loh, Y. Zhao and G. Chen, *Nucleic Acids Res.*, 2016, **44**, 9071-9082.
20. N. Brodyagin, I. Kumpina, J. Applegate, M. Katkevics and E. Rozners, *ACS Chem. Biol.*, 2021, **16**, 872-881.
21. A. Ohkubo, K. Yamada, Y. Ito, K. Yoshimura, K. Miyauchi, T. Kanamori, Y. Masaki, K. Seio, H. Yuasa and M. Sekine, *Nucleic Acids Res.*, 2015, **43**, 5675-5686.
22. N. Brodyagin, D. Hnedzko, J. A. MacKay and E. Rozners, *Methods Mol. Biol.*, 2020, **2105**, 157-172.
23. N. Brodyagin, A. L. Maryniak, I. Kumpina, J. M. Talbott, M. Katkevics, E. Rozners and J. A. MacKay, *Chem. Eur. J.*, 2021, **27**, 4332-4335.