ARTICLE

Triplex-forming peptide nucleic acids as emerging ligands to modulate structure and function of complex RNAs

Received 00th January 20xx,
Accepted 00th January 20xx

DOI: 10.1039/x0xx00000x

Martins Katkevics,[a,†] James A. MacKay,[b,†] and Eriks Rozners*[c,†]

Over the last three decades, our view of RNA has changed from a simple intermediate supporting protein synthesis to a major regulator of biological processes. In the expanding area of RNA research, peptide nucleic acid (PNA) is emerging as a promising ligand for triple-helical recognition of complex RNAs. As discussed in this feature article, the key advantages of PNAs are high sequence specificity and affinity for RNA (>10 fold higher than for DNA) that are difficult to achieve with small molecule ligands. Emerging studies demonstrate that triple-helical binding of PNAs can modulate biological function and control dynamic conformational equilibria of complex folded RNAs. These results suggest that PNA has a unique potential as a research tool and therapeutic compound targeting RNA. The remaining problems hampering advances in these directions are limitations of sequences that can be recognized by Hoogsteen triplexes (typically purine rich tracts), poor cellular uptake and bioavailability of PNA, and potential off-target effects in biological systems. Recent exciting studies are discussed that illustrate how synthetic nucleic acid chemistry provides innovative solutions for these problems.

Introduction

RNA has a host of diverse functions despite being most renowned as a mediator between DNA and proteins in the central dogma of molecular biology and new functions of RNA continue to be discovered.[1-4] Unlike proteins that are made of twenty-one standard monomeric units with great structural diversity, RNA's primary structure lacks diversity with only four similar nucleoside monomers and a homogeneous polyanionic phosphate backbone that historically discouraged interest in RNA as a target for molecular recognition. Still, a wide array of tertiary motifs that lead to RNA structural diversity are known[5] and there is a growing body of recent research toward the recognition of RNA using both small molecules[6, 7] and modified oligonucleotides.[8, 9] Given that double helical regions are quite common in non-coding RNA,[2, 10] duplex RNA is primed for recognition by a third oligonucleotide strand. In fact, nature has revealed the triple helix motif as a unique feature capable of regulating function by controlling folding patterns in RNA through winding and unwinding from duplex RNA.[11-13] Inspired by nature's design, triplex forming oligonucleotides (TFOs) have emerged as a well explored class of molecular probes for recognition of double-stranded DNA (dsDNA).[14]

Peptide nucleic acid (PNA, Fig. 1) was originally reported in 1991 as a nucleic acid analogue for triple-helical binding to dsDNA.[15] The guiding principles for the design of PNA's neutral pseudopeptide backbone were structural simplicity (no chiral centers), easy synthesis (amide bond formation), and lack of electrostatic repulsion with the negatively charged dsDNA. The latter was expected to be especially advantageous for formation of strong PNA-dsDNA triplexes.

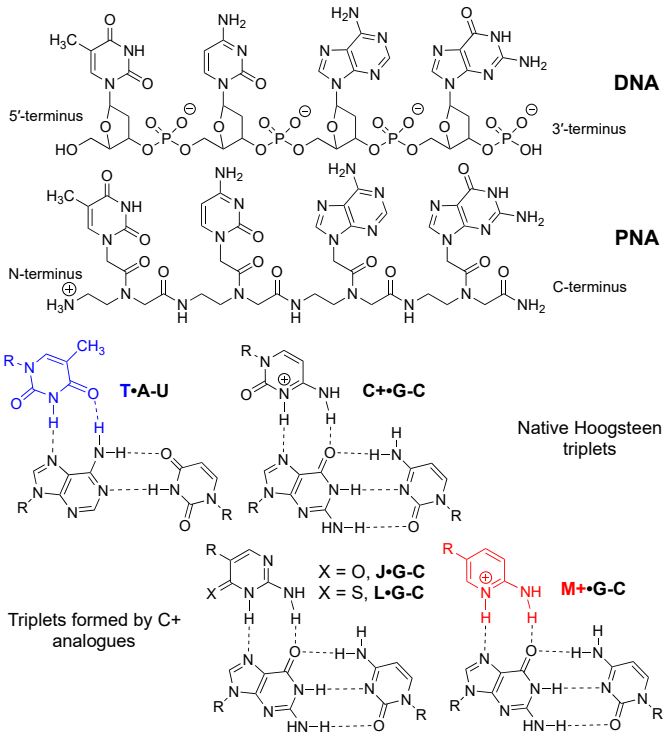

Fig. 1. Structures of DNA, PNA, and native and modified Hoogsteen triplets.

However, the first study on triple helical binding of PNA gave a surprising discovery of strand invasion in dsDNA.[15] In this

[a] Latvian Institute of Organic Synthesis, Aizkraukles 21, Riga, LV-1006, Latvia.
[b] Department of Chemistry and Biochemistry, Elizabethtown College, Elizabethtown, PA 17022, USA.
[c] Department of Chemistry, Binghamton University, Binghamton, NY 13902, USA. E-mail: erozners@binghamton.edu.
† The authors are listed alphabetically.

unexpected binding mode, the pyrimidine rich DNA strand was displaced by a second PNA molecule forming a PNA-DNA-PNA strand-invasion triplex instead of the expected PNA-dsDNA triplex. The exciting discovery that PNA could unwind and invade dsDNA was unprecedented in 1991 and captured the majority of the resulting PNA research for the next several decades.[16] Interestingly, relatively few studies continued exploring the original intent, the formation of PNA-dsDNA triplex.[17]

RNA triple helices have been studied even less than DNA triplexes.[11] Most surprisingly, triple helical binding of PNA to dsRNA was not studied (for an early report mentioning a PNA-dsRNA complex see ref.[18]) until 2010 when Rozners and co-workers reported that even short PNA 6-mers formed unusually strong and sequence specific triple helices with dsRNA at pH 5.5.[19] Fast forward to 2023, nucleobase-modified PNAs (Figs. 1 and 2) are emerging as promising high affinity ligands for sequence specific recognition and functional control of complex folded biologically relevant RNA molecules.[16, 20] In the present feature article, we review recent exciting studies in our and other laboratories to develop triplex-forming PNAs for recognition of functional biologically important RNAs. A distinct feature of our approach has been the use attractive electrostatic interactions engineered in modified nucleobases to improve binding affinity without compromising sequence selectivity. We are relying extensively on structural modeling and synthetic organic chemistry to design novel nucleobases that use Hoogsteen hydrogen bonding to achieve the ultimate but so far elusive goal of recognizing any sequence of double-stranded RNA.

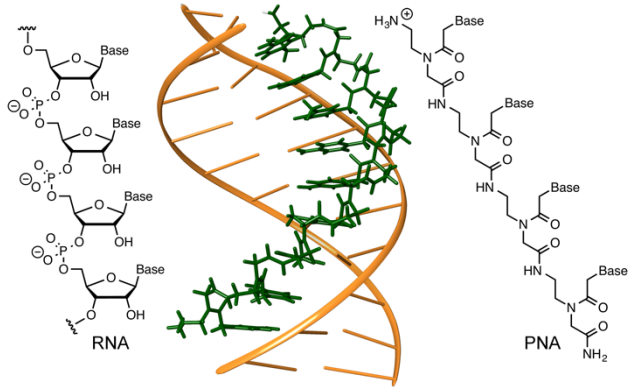

Fig. 2. Cartoon structure of double-stranded RNA complexed with triplex-forming PNA.

Thermodynamic stability of PNA-dsRNA triplexes

Native triple helices form when the third DNA strand binds in the major groove of a double helix using Hoogsteen hydrogen bonding to purine nucleobases (Fig. 1).[14] Native RNA triple helices are also well known to regulate a variety of biological functions.[12, 13] In the most common native triple helices, the third DNA (or RNA) strand binds parallel to the so-called polypurine tract where one strand of the double helix consists of mainly purines. Despite being a promising approach to targeting genomic DNA, triplex-forming oligonucleotides (TFOs) have not found practical applications either as probes or as antigene therapeutics. The main roadblock has been low thermodynamic stability of the Hoogsteen triple helices compared to the Watson-Crick double helices, especially at physiologically relevant pH. The origin of the low stability is electrostatic repulsion between the negatively charged phosphate backbones of the TFO and target DNA or RNA double helix. Using a neutral amide backbone, removes the electrostatic repulsion and improves the overall stability of triple helices formed by PNA.[17] However, the requirement for slightly acidic conditions (e.g., pH ~5) originates from the low pK_a of cytosine (~4.5) that needs to be protonated to form the C+•G-C triplet (Fig. 1). The problem of cytosine protonation is generally approached either by designing neutral analogues that have the required hydrogen bond donors and acceptors or by modifying cytosine to increase the pK_a.[14]

Several studies of kinetics and thermodynamics of binding of a triplex-forming PNA to dsRNA have been reported. Sugimoto disclosed that the entropic contribution toward the protonation of basic residues plays a key role in destabilizing triplexes.[21] Nishizawa and co-workers used a stopped-flow technique along with isothermal titration calorimetry (ITC) to determine that the association rate constant was dominated by the charge of the PNA molecule. They concluded that the triplex formation proceeds through a nucleation-zipping mechanism.[22] Collectively, previous studies using modified nucleobases have solved the problem of cytosine protonation and removed the requirement for acidic pH; however, optimization of PNA binding affinity and improvements of sequence scope that can be recognized using triplex-forming PNAs remain active areas of research.[16, 23]

Neutral cytosine analogues for recognition of G-C base pair

As early as 1995, Nielsen and co-workers[24] introduced pseudoisocytosine (J, Fig. 1) a neutral PNA nucleobase originally developed for TFOs by Kan and co-workers.[25] J mimics the hydrogen bonding ability of protonated C and using J instead of C is currently the most common approach to enable triple helical binding of PNAs at physiological pH.[26, 27] In PNA-dsRNA triplexes, the stability of the J•G-C triplet is similar to or slightly weaker than the stability of the T•A-U triplet.[28] Consequently, several research groups have explored other modified nucleobases to improve on J for recognition of biologically relevant RNAs.[16, 23] Most notably, Chen and co-workers[29] showed that 4-thio-pseudoisocytosine (L, Fig. 1) having a C=O substituted with C=S improved the affinity of triplex-forming PNAs for complementary dsRNA. L was originally developed for DNA triplexes by Sekine and co-workers,[30] but like J did not find wide applications in TFOs. Chen and co-workers used L-modified PNAs to stimulate ribosomal frameshifting[31] and inhibit replication of influenza A virus[32] demonstrating that L enabled functional control of biologically important RNAs.

Modulation of cytosine pK_a for recognition of G-C base pair

The benefits of increasing cytosine pK_a to enhance protonation and stabilize C+•G-C triplets at physiological pH was recognized

early in DNA triplexes. Povsic and Dervan reported that the slightly higher pK_a of 5-methylcytosine resulted in enhanced stability of MeC+•G-C triplets in TFOs.[33] Later, several groups explored 2-aminopyridine (M, Fig. 1) as a more basic (pK_a~6.7) cytosine analogue but the success of M in DNA triplexes was modest.[34, 35] Early studies by Rozners and co-workers suggested that M as a PNA nucleobase formed significantly stronger triplets than J at physiological salt and pH.[36] In a related approach, Chen and co-workers used a guanidinium group to mimic the hydrogen bonding scheme of protonated cytosine.[37]

To evaluate the binding affinity and sequence specificity of nucleobase-modified PNAs we have been using a model system of four dsRNA hairpins (HRP1-HRP4, Fig. 3) having a variable base pair in the middle of a polypurine tract.[28, 36, 38] HRP1-HRP4 were designed following the original publication by Roberts and Crothers on mixed DNA and RNA triplexes.[39] All new nucleobases (such as, P_9, E, and V shown in Figure 2 and discussed below) were incorporated in PNAs at the position facing the variable base pair and the binding affinity was measured using ITC and UV thermal melting.

The UV thermal melting uses hyperchromicity, the decrease in absorbance of nucleobases upon formation of stacked helical structures, to monitor formation and dissolution of nucleic acid complexes. Usually, hyperchromicity is measured at 260 nm where the native nucleobases have unique absorbance maxima. An important observation made in our group was that the M-modified PNAs had unique absorbance at 300 nm where native DNA and RNA do not appreciably absorb the UV light.[40] Therefore, UV melting at 300 nm allowed observation of only the triplex dissociation without interference of the hyperchromicity signal from melting of dsRNA.

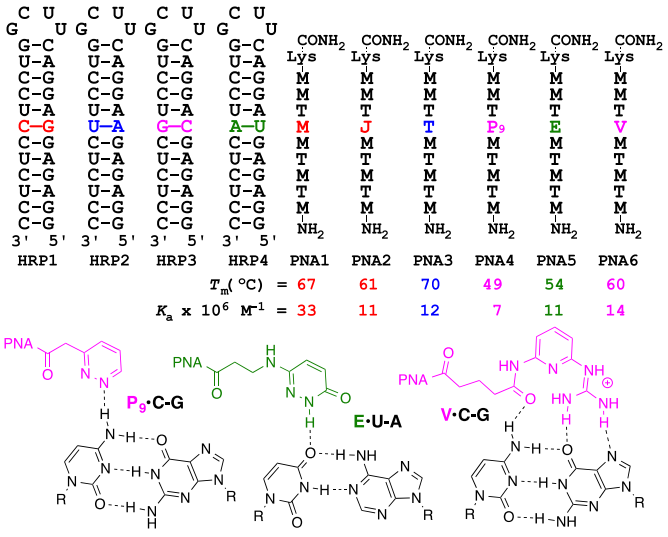

Fig. 3. Structures of model RNA hairpins and PNAs to study the affinity, specificity, and sequence scope of triplex formation (top) and Hoogsteen triplets recognizing pyrimidines. The matched triplets are color coded in HRP and PNA sequences.

Additionally, it is common to place a lysine residue at the C-terminus to aid in binding affinity, improve solubility, and deter aggregation of PNA. However, in a systematic study Ryan and Rozners reported that neither the location (N- or C-terminus) nor chirality of the Lys residue imparts significant differences in PNA affinity for RNA.[41]

Recent detailed studies using the model RNA hairpins showed that M+•G-C triplets were about three times stronger than J•G-C triplets.[28] While PNA1 (Fig. 3) had high affinity for the matched HRP1 (K_a = 33 × 10^6 M^{-1} and T_m = 67 °C), PNA2 containing the J base, had notably weaker binding (K_a = 11 × 10^6 M^{-1} and T_m = 61 °C). Replacement of all Ms with Js lowered the affinity by ~100-fold.[28] PNA1 showed high sequence selectivity for HRP1 because the mismatched complex with HRP2-HRP4 had ~25-fold lower K_a and ~30 °C lower T_m than the matched PNA1-HRP1 triplex. The sequence specificity of PNA1 was slightly better than that of PNA2 having J at the variable position.[28] Collectively, these results showed that M was a superb modified nucleobase enabling strong and sequence specific formation of PNA-dsRNA triplexes at physiological conditions.

Triple-helical recognition of A-U base pair

Natural RNAs make use of the U•A-U, while thymidine (T) is used most commonly in PNA to form a T•A-U triplet, as described above (Fig. 1). Given the stability of this natural triple, less attention has been given toward recognition of A-U compared to G-C (vide supra). Most successful modifications involved changing the functionality at the uracil 5-position. Chen and co-workers systematically replaced T's in triplex-forming PNA with 5-halouracils (^{F}U, ^{Cl}U, ^{Br}U, and ^{I}U) and demonstrated significant improvements in RNA binding affinity.[42] They attributed this to a lower pK_a of the N-H of the 5-halouracils compared to T. In most cases, the best binding was observed for BrU suggesting that pK_a may not be the only factor and that stacking interactions may be in play. Noteworthy, these modified PNA's showed no appreciable binding to dsDNA, however 5-halouracils enhanced binding to single-stranded RNAs through Watson-Crick base paring.[42] Similarly, our research groups found that a 5-triazolyl uridine derivative showed enhanced binding to A-U base pairs suggesting that substitutions at the 5-position are both sterically well tolerated and have potential to enhance stability of T•A-U triplets.

Inspired by their success with the L base for G-recognition, Chen and co-workers also demonstrated that 2-thiouracil (s^2U) modification enhanced binding affinity for A-U base pairs.[43] They proposed a lower dehydration energy imposed by the thiocarbonyl; however, the interplay between stacking, hydrogen-bonding and dehydration in any of these modified PNAs has yet to be explored fully.

Sequence scope of triple helical recognition

Native parallel RNA and DNA triple helices are inherently limited to polypurine tracts because stable triplets, C+•G-C and U•A-U (or T•A-T in DNA) are formed only by hydrogen bonding to the Hoogsteen faces of purines (Fig. 1).[14] Development of modified nucleobases to recognize the Hoogsteen faces of pyrimidines through formation of stable X•C-G and Y•U-A triplets is an active area of research that so far has given relatively modest

advances.[16, 23, 44] The main challenge is that pyrimidines present only one hydrogen bond donor (-NH of C) or acceptor (C=O of U) on their Hoogsteen face (see Fig. 3). The problem is further compounded by steric crowding because pyrimidines protrude further out in the major groove than purines leading to clashing with incoming PNA nucleobases. Taken together, these problems have greatly complicated design of nucleobases that could form highly stable and sequence specific X•C-G and Y•U-A triplets, which remains a key bottleneck for triple helical recognition of nucleic acids.[16, 23, 44]

Recognition of pyrimidines in X•C-G and Y•U-A triplets

Despite the inherent challenges, several research groups, including ours, have continued searching for nucleobase analogues that could recognize pyrimidines using a single hydrogen bond.[16, 23, 44] For triplex-forming PNAs, we started with pyrimidin-2-one (P_0, Fig. 4) that was first reported in DNA triplexes by Prevot-Halter and Leumann.[45] Consistent with their results in DNA triplexes,[45] in our PNA-dsRNA model triplexes, P_0 selectively recognized C-G over other RNA base pairs; however, with significantly decreased binding affinity.[46] In our model system (Fig. 3) PNA modified with P_0 at the variable position had K_a = 4 × 10^6 M^{-1} and T_m = 40 °C (Fig. 4).[47]

Since simple nitrogen heterocycles have been reported to recognize C-G in DNA triplexes,[48, 49] we screened all nine isomers of pyridine, pyrimidine, pyrazine, and pyridazine as PNA nucleobases.[47] Of these, 3-pyridazinyl nucleobase P_9 (Fig. 4) at the variable position formed the most stable triplet in PNAP$_N$-HRP3 (Fig. 4, K_a = 7 × 10^6 M^{-1} and T_m = 49 °C), which was an improvement over P_0 but still inferior to the high stability of M+•G-C and T•A-U triplets.

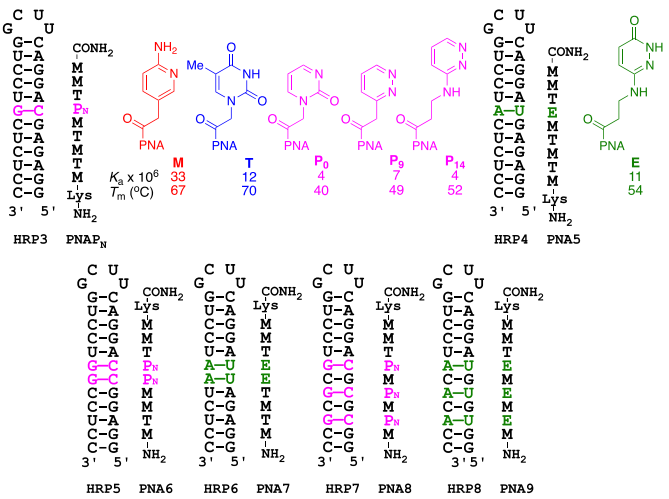

Fig. 4. Designer PNA nucleobases for recognition of C-G and U-A base pairs using a single Hoogsteen hydrogen bond.

For recognition of T-A base pairs in DNA, Nielsen and co-workers developed 3-oxo-2,3-dihydropyridazine (E, Fig. 4). The extended linker that connected E to PNA backbone was expected to circumvent steric hindrance from the 5-Me of T and allow E to form a single hydrogen bond with the C4 carbonyl of T.[50] We found that E also recognized U-A base pairs in RNA with good affinity (Fig. 4, K_a = 11 × 10^6 M^{-1} and T_m = 54 °C) albeit the sequence specificity was lower than for M and T.[46, 51] Molecular modeling suggested that in PNA-dsRNA triplexes E formed one hydrogen bond with U (Fig. 3) as originally proposed for the E•T-A triplet in DNA.[50] Interestingly, using the longer linker in P_{14} did not significantly change the binding properties of P_9.[51] Molecular dynamics simulations suggested that the longer linker pushed P_{14} out of optimal stacking alignment, which likely off-set any gains from optimized hydrogen bond alignment.[51] Most importantly, placing P_9 and E in PNA6-PNA9 designed to recognize several pyrimidine interruptions in HRP5-HRP8 (Fig. 4) resulted in notable losses of binding affinity. While the K_a for PNA6-HRP5 and PNA7-HRP6 was 1 and 3 × 10^6 M^{-1}, respectively, stability of PNA8-HRP7 and PNA9-HRP8 was too low to be measured by ITC or UV melting under our usual conditions.[51] Taken together, these results show that recognition of pyrimidines using simple heterocycles forming one hydrogen bond is a challenge and remains an unsolved problem for triplex-forming PNAs.

Extended nucleobases recognizing the entire Hoogsteen face

An alternative approach to triple helical recognition of pyrimidine interruptions in polypurine tracts is to design extended nucleobases that hydrogen bond to the entire Hoogsteen face of Watson-Crick base pairs. The advantages are that in theory one could form three hydrogen bonds and that extending the pi systems may enhance beneficial stacking interactions. While this approach was already explored in the 1990s, pioneered by Dervan and co-workers,[52] the initial designs suffered from low affinity and sequence selectivity.[53-55] Later it was discovered that, at least in some cases, the extended nucleobases intercalated between the base pairs instead of forming the desired hydrogen bonds to the entire Hoogsteen face of dsDNA.[56, 57] More recent studies have achieved some success with extended nucleobases in modified TFOs forming DNA triplexes and this approach has also been adopted in triplex-forming PNA for recognition of dsRNA.[16]

N^4-(2-Guanidoethyl)-5-methylcytosine[58] (Q, Fig. 5) and *N*-(4-(3-acetamidophenyl)-thiazol-2-yl)acetamide[59] (S, Fig. 5) have emerged as two especially successful extended nucleobases in DNA triplexes.[60, 61] Chen and co-workers introduced Q[62] and S[63] in triplex-forming PNAs to recognize C-G and U-A interruptions, respectively. Similar to P_N series nucleobases, Q formed triplets with C-G with good selectivity but decreased affinity. In contrast, S lacked sequence specificity and was binding to C-G and U-A base pairs with similar affinity. Though both Q and S have been explored only as single PNA modifications, it is conceivable that multiple substitutions may decrease the affinity and selectivity of PNA-dsRNA triplexes. However, as will be discussed later, single modifications Q and S have been used successfully in triplex forming PNAs to recognize isolated pyrimidine interruptions in complex biologically relevant dsRNAs.

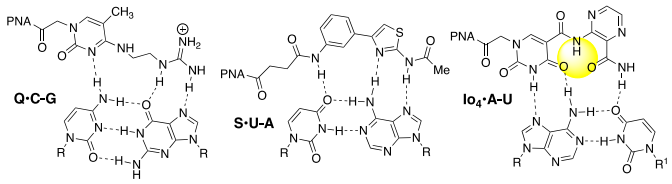

Fig. 5. Extended PNA nucleobases for recognition of the entire Hoogsteen face of Watson-Crick base pairs. The yellow sphere highlights intermolecular hydrogen bonding interactions.

Early attempts by our collaborative team illustrated the challenges in designing of improved extended nucleobases.[64] Our approach initially involved designing an extended nucleobase scaffold aimed at the more straightforward purine recognition, with the aspiration toward using the same scaffold design to attach PNA to the opposite side of the extended nucleobase for pyrimidine recognition. While molecular modeling suggested several viable hydrogen bonding schemes, synthetic difficulties limited the designs that could be tested. The early designs had low binding affinities most likely due to less-than-ideal hydrogen-bonding and π-stacking of the extended nucleobases that was compounded with high entropic cost of rearranging the scaffolds with multiple conformations.[64] More recent studies confirmed the need for conformational preorganization of extended nucleobases. MacKay and co-workers[65] reported that extended isoorotamide containing nucleobases (e.g., Io_4 in Fig. 5) had improved affinity for U-A base pairs that was maintained in PNAs with multiple Io_4 modifications. The success of Io_4 was at least in part due to a favorable planar preorganization of the scaffold by intermolecular hydrogen bonds involving the isoorotamide N-H (yellow highlight in Fig. 5).

While 2D structures like those in Figure 4 are instructive toward understanding proposed Hoogsteen-base triples, the hydrogen-bonding schemes of modified nucleobases have rarely been rigorously established by structural studies. In many studies, molecular modeling and molecular dynamics simulations of triple-helical structures including the modified nucleobases have been used to support the suggested Hoogsteen hydrogen-bonding (e.g., Io_4•A-U in Fig. 5 and V•C-G in Fig. 6); however, the models need to be continuously refined based on new experimental data. For example, a most recent study[66] suggested that the third hydrogen bond (amide to C=O of U) in the Io_4•A-U triplet may play a less important role than initially assumed given that analogues lacking the primary amide afforded equal or better binding affinities.

Inspired by naphthyridine based extended nucleobases designed by Ohkubo, Sekine and co-workers,[61] Ryan et al. designed a cationic 2-guanidyl pyridine PNA nucleobase (V in Figs. 2 and 5).[67] As a single modification, V formed a slightly less stable V•C-G triplet (Fig. 3, K_a = 14 × 10^6 M^{-1} and T_m = 60 °C) compared to M+•G-C triplets. This affinity was similar to slightly better than that of T•A-U; however, the K_a was approximately double that of P_9, which had been the best performing base for the C-G base pair. Similar to Io_4, the heterocyclic system of V was preorganized into a planar scaffold by intermolecular hydrogen bonding (Fig. 6A). In addition, the long linker connecting V to the PNA backbone engaged in hydrogen bonding not only to the -NH_2 of C, but also to the C=O of the linker connecting an adjacent M base to PNA backbone (Fig. 6B). Despite the favorable hydrogen bonding network in Fig. 6, PNAs with two V modifications formed triplexes with dsRNA with reduced affinity (T_m < 50 °C) while PNA with three V modifications showed a non-sequence specific binding to RNA.[67]

These results revealed that cationic base modifications (such as, Q and V) will help maintaining binding affinity, but sequence specificity might be compromised when several modifications are used. While V remains our best extended nucleobase for recognition of isolated C-G interruptions, it is not suitable for sequences having multiple V•C-G triplets. A similar phenomenon was observed when a cationic Io derivative was prepared and tested.[66]

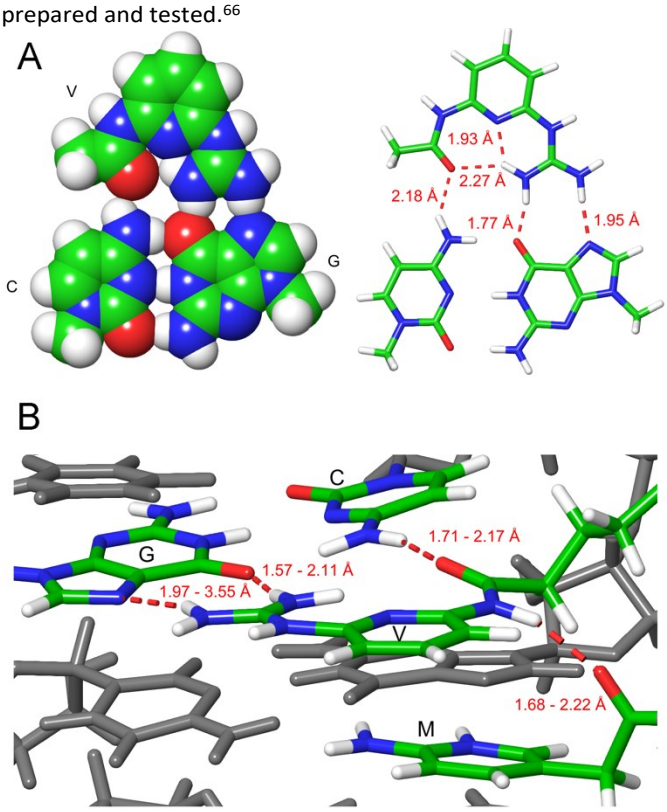

Fig. 6. Geometry optimization and molecular modeling of V•C-G triplet using (A) B3LYP 6-31G*(d, p) and (B) molecular dynamics simulations. Reproduced with permission from ref. 67. Copyright (2022) RSC.

Altogether, these studies have continued to advance our understanding of PNA/RNA triplex formation and continued improvements towards the goal of recognition of any sequence of dsRNA, but limitations still remain. These studies have revealed the difficulties of designing large heterocyclic systems that have precise hydrogen bonding arrangements required to recognize the entire Hoogsteen face of Watson-Crick base pairs. Another important conclusion was the need for conformational rigidity, provided by additional intra- and inter-molecular hydrogen bonds, to diminish the entropic cost of free rotation in multiple bonds of extended heterocyclic systems and long linkers connecting nucleobases to PNA's backbone. These insights should help future designs of better modified nucleobases. Currently, we do not have a general solution to the

problem of pyrimidine recognition, especially in sequences with multiple pyrimidine interruptions. The search for modified nucleobases to recognize any sequence of dsRNA remains an ongoing and formidable challenge.

Structure and RNA preference of PNA triplexes

A surprising discovery from our biophysical studies was a unique preference of M-modified PNAs for binding to dsRNA over dsDNA.[28, 36, 38] PNA1 and PNA3 (Fig. 3) formed at least 10-fold stronger triplexes with HRP1 and HRP2, respectively, than with dsDNA hairpins having the same sequence.[28] The reason for such unusual difference in binding affinity was not immediately obvious, but had to be related to the different structures of RNA and DNA helices, A- and B-type, respectively. Interestingly, in a previously published crystal structure of PNA-DNA-PNA triplex, the DNA strand adopted a conformation of P-type helix (~16 base pairs per turn), more similar to the A-type structure of RNA than to the B-type structure of DNA, suggesting that the PNA binding forced DNA to adopt an RNA-like conformation.[68]

These intriguing results prompted us to perform an NMR structural study of PNA-dsRNA triplex similar to those in Figure 3.[69] Scott Kennedy built a model of a PNA-dsRNA triplex, based on the published crystal structure,[68] and optimized the geometry using distance restraints obtained from experimental NOESY data on a PNA-dsRNA complex. In the resulting triplex, the RNA assumed the expected conformation, similar to the crystal structure of the PNA-DNA-PNA triplex, and the PNA aligned along the purine strand of RNA forming the expected Hoogsteen hydrogen bonded M+•G-C and T•A-U triplets (Fig. 7).

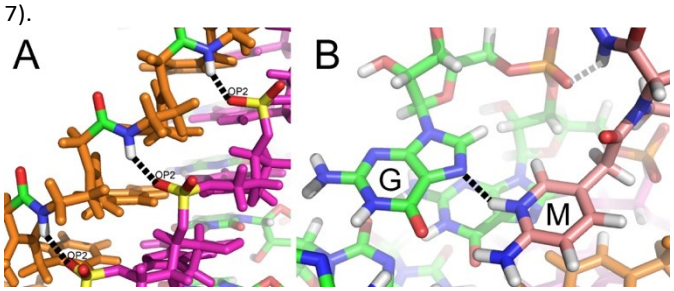

Fig. 7. Hydrogen-bonding interactions stabilizing PNA-dsRNA triplex: (A) PNA amide to RNA phosphate backbone interactions and (B) M to G hydrogen bonding. Reproduced with permission from ref. 69. Copyright (2019) Wiley-VCH.

Most remarkably, in both the crystal structure and our PNA-dsRNA triplex model, the PNA formed additional hydrogen bonds from the N-H of PNA backbone amides to the OP2 of RNA backbone phosphates (Fig. 7A).[69] Our results suggested that this hydrogen bonding "zipper" was in large part responsible for the higher stability of PNA-dsRNA triplexes compared to PNA-dsDNA triplexes of the same sequence. The hydrogen bonding was possible because of the matching distance (~5.5 Å) between the backbone N-H and OP2 in RNA-like P-type and A-type helices, while this distance was significantly longer (~7 Å) in DNA-like B-type helices, which prevents a continuous hydrogen bonding "zipper". It is conceivable that compared to RNA, DNA needs to undergo significantly larger reorganization to adopt the P-type conformation for PNA binding, as is observed in the crystal structure of the PNA-DNA-PNA triplex.[68] The energetic cost of this preorganization may explain the lower stability of PNA-dsDNA triplexes. This analysis is consistent with biophysical studies by Nishizawa and co-workers showing that triplex formation requires conformational changes that are larger for dsDNA than for dsRNA targets.[70] Collectively, these studies showed that PNA is naturally a better ligand for triple-helical recognition of dsRNA than for recognition of dsDNA.

Interestingly, the NMR data also suggested that M formed only one strong hydrogen bond with G of G-C base pair: N1-H of M+ to N7 of G (Fig. 7B). The exocyclic amino group of M appeared to freely rotate and engage in only weak interactions with C=O of G.[69] These intriguing observations underscore the need for rigorous structural studies to fully understand the molecular interactions that drive formation of Hoogsteen triplets and, ultimately, the RNA recognition. Without such studies, the hydrogen bonding schemes proposed in Fig. 5 and elsewhere should be considered only reasonable possibilities, providing that molecular modeling studies support them.

Towards applications of triplex-forming PNAs

Triple-helical binding of PNA controls biological activity of RNA

Historically, poor cellular uptake and bioavailability have hampered biological and medicinal applications of PNAs.[71] The triplex-forming PNAs discussed above are no exception and improving the cellular delivery of PNAs remains an active area of research.[16] Despite these issues, several recent studies have demonstrated intriguing biological activity of triplex-forming PNAs targeting dsRNA.

Chen and co-workers reported that L- and Q-modified PNAs targeting model mRNA hairpins stimulated ribosomal frameshifting in a cell-free in vitro assay.[31] In another study, similar PNAs inhibited replication of influenza A virus in MDCK cells by forming a triplex with the conserved panhandle duplex region of viral genomic RNA.[32] In the latter study, PNA was conjugated with neamine[72] to stimulate the cellular uptake.

Recent studies by Rozners and co-workers showed that M-modification improved the cellular uptake of PNAs, most likely by mimicking the structure of arginine.[38, 73] M-modified PNAs conjugated to short cell-penetrating peptides were taken up efficiently in cells and apparently escaped endosomes but remained trapped in unknown cellular ompartments.[73] On the other hand, delivering PNAs using electroporation enabled their biological activity. Studies in collaboration with Profs. Sugimoto and Endo at FIBER (Konan University, Japan) demonstrated that M-modified PNAs suppressed mRNA translation[74] and microRNA maturation[75] by forming sequence specific triple helices with target RNAs in cells. However, the latter study also revealed that a scrambled control PNA slightly increased the levels of microRNAs suggesting that PNAs may have off-target effects in complex biological systems. While biological RNAs contain many purine rich regions, the two letter recognition code of M+•G-C and T•A-U triplets becomes redundant for longer sequences leading to many partially matched potential

off-targets. Going forward, the development of novel nucleobases that recognize any sequence of dsRNA should help with improving the biological specificity of triplex-forming PNAs by decreasing the cationic character (fewer M modifications) and increasing the diversity of recognition sequences (four letter recognition code). However, the biological specificity and off-target effects remain concerns that need to be studied and controlled for any application of PNAs ether as tools or therapeutics.

Triple-helical binding of PNA controls dynamic alternative conformations of RNA

Native nucleic acid triple helices are usually less stable than the corresponding double helices.[39] However, the first study on triple helical binding of PNA to dsRNA by Rozners and co-workers revealed that the PNA-dsRNA triplex was notably more stable than the dsRNA duplex.[19] This observation promoted a hypothesis that that binding of triplex-forming PNAs to dynamic RNA structures may be used to drive the equilibria of alternative conformations towards one specific structure. PNAs that lock the RNA structure in one of the alternative conformations could serve as enabling tools for studying the biological role of dynamic RNA switches and evaluating their potential as novel drug targets.

In contrast to the uniform double helix of DNA, RNA folds in complex structures where single-stranded loops, junctions, and bulges interrupt double-helical portions of RNA. Bulges formed when one or several nucleotides on one RNA strand do not have base pairing partners on the other strand are most common structural motifs in RNA.[76] The hypothesis that PNAs could control the conformation of RNA bulges was tested using modified model hairpins HRP9N (Fig. 8, N was A, U, G, or C) where a single nucleotide bulge was added to our original HRP1.[77]

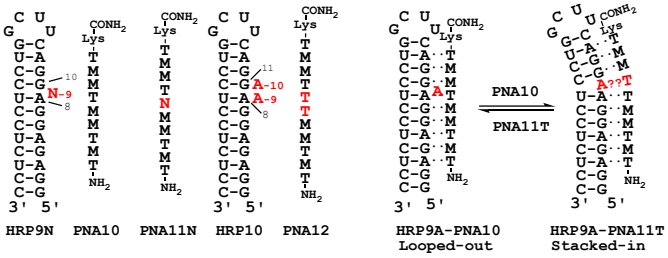

Fig. 8. Structures of model RNA hairpins featuring single nucleotide and double A bulges and PNAs to study the recognition of RNA bulges.

The unpaired nucleotides in HRP9N may adopt either looped-out or stacked-in conformations and serve as an excellent model system to test the ability of triplex forming PNAs to control these alternative conformations. Consistent with the hypothesis, PNA10 (Fig. 8) was binding to HRP9N with affinity similar to that for HRP1 (K_a = 30-55 × 10^6 M^{-1} and T_m = 67-77 °C).[77] This result suggested that in PNA10-HRP9N complexes the unpaired nucleotides were fixed in looped-out conformations. When an additional nucleotide N was inserted in PNA11N (Fig. 8, N was A, T, G, C, or M), PNA11T showed unusually high affinity for HRP9A (K_a = 220 × 10^6 M^{-1} and T_m = 80 °C) while other PNA-dsRNA combinations were notably less stable.[77] Similar, unusually high stability was observed for PNA11C and PNA11M complexes with HRP9G. These results suggested that PNA11T engaged the bulged A in HRP9A in an extraordinarily stable stacked-in conformation (Fig. 8 right), possibly through a Hoogsteen-like hydrogen bonding, which was consistent with similar high stability of PNA11C-HRP9G and PNA11M-HRP9G complexes (though currently we cannot exclude alternative structures). Interestingly, PNA12 also formed highly stable triplex with HRP10 featuring a two adenosine bulge.[77] While the exact structures and reasons behind the unusually high stability of triplexes involving RNA bulges are awaiting detailed structural studies, chemical probing of all triplexes was consistent with the proposed looped-out and stacked-in conformations.

Collectively, these results suggested that, depending on the sequence, triplex-forming PNAs had a unique ability to shift dynamic structures of single (and possibly also double) nucleotide bulges from looped-out to stacked-in conformations. Bulges of unpaired nucleotides are dynamic RNA structures that play important roles in driving RNA interactions with proteins, small molecules and other RNAs.[76] The ability of triplex-forming PNAs to control the conformation of RNA bulges will be useful for fundamental studies in RNA biology and may find practical biomedical applications.

Triplex-forming PNAs as fluorescent probes for RNA recognition

In contrast to their slow clinical development, duplex-forming PNAs have become powerful research tools, probes, and diagnostics.[16, 78] Fluorogenic PNAs offer an attractive strategy for nucleic acid detection,[79] so it is natural that several fluorescent PNA nucleobases have been explored for detecting triplex formation with dsRNA. Nishizawa, Sato, and co-workers were the first to develop a fluorescent probe for dsRNA detection using thiazole orange (TO, Fig. 9) as a universal base surrogate that exhibited a fluorescence light-up response upon triplex formation through intercalation.[80] The triplex-forming forced intercalation (tFIT) PNA probes were developed based on earlier work of Seitz and co-workers on TO-modified duplex-forming PNAs as forced intercalation (FIT) probes.[81, 82] The fluorescence light-up signal resulted from a rigidification of the TO fluorophore upon binding to single or double stranded nucleic acid target. Nishizawa, Sato, and co-workers later determined that a longer linker between the nucleobase and the PNA backbone enhanced binding ~10-fold while maintaining the function of TO as a fluorescent universal base.[83] Second generation red-emitting bases (QB, TR, and BIQ, Fig. 9) retained binding affinity and showed unique photophysical properties.[84, 85] In a related approach, Chen and co-workers[86] showed that L-modified PNAs containing 5-benzothiophene uracil (btU, Fig. 9), a fluorogenic nucleobase originally reported by Sabale and Srivatsan[87] for duplex-forming PNAs, exhibited a light up response upon binding dsRNA.

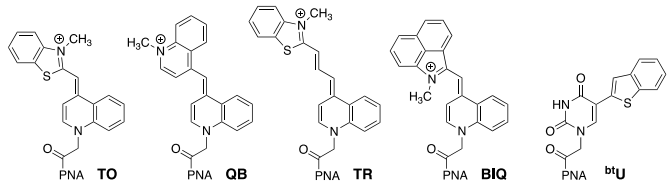

Fig. 9. Fluorescent PNA nucleobases used for triple-helical recognition and detection of dsRNA.

More recently Sato, Nishizawa and co-workers have used the TO tFIT probes conjugated with small molecules to target biologically relevant dsRNAs. A TO-modified PNA conjugate with fluorescent trimethylated naphthyridine derivative showed high affinity and selectivity for bacterial A-site RNA, and exhibited strong enhancement of fluorescence upon A-site binding.[88] Another PNA conjugate with 6,7-dimethoxy-2-(1-piperazinyl)-4-quinazolinamine (DPQ) was targeted to the panhandle region of influenza A virus (IAV) promoter containing a unique (A•A)-U internal loop.[89, 90] Impressively, the PNA conjugate containing DPQ and TO modifications gave a dramatic light up response alongside improved and selective binding to the IAV RNA promoter region. A related application used tFIT probes to simultaneously bind the 3'-overhang and double-stranded region of siRNAs.[91, 92] Winssinger and co-workers described a dsRNA-templated reaction using triplex-forming PNA-reagent conjugates that generated signal by unmasking a coumarin fluorophore.[93] Collectively, these studies demonstrate rich potential applications of fluorogenic triplex-forming PNAs as probes and diagnostics to detect biologically relevant dsRNA species. Further development of fluorescently labeled PNAs should enable structural studies using fluorescence resonance energy transfer (FRET) and other related advanced spectroscopic techniques. To this end, we have developed an efficient protocol for solid phase terminal fluorescent labeling of PNAs that is complementary to using fluorogenic nucleobases discussed above.[94]

Synthesis of nucleobase-modified PNAs

The original design of PNA was in part motivated by a straightforward synthesis at both monomer and oligomer level. PNA is typically synthesized following the well-established peptide synthesis methods using either Fmoc or Boc as the temporary protecting groups for the growing PNA chain and other orthogonal protecting groups for the heterocyclic bases (Fig. 10).[95] Assembling of PNA oligomers consists of standard solid phase synthesis steps: attaching of monomer to the growing chain of PNA using carboxyl acid activating reagents such as HATU, optional capping of unreacted amino groups by acetylation, and removal of the Fmoc protection. PNA monomers with the canonical nucleobases and unmodified aminoethylglycine (AEG) backbone are commercially available, but monomers carrying modified nucleobases for triplex-forming PNAs (as shown throughout the article) need to be custom synthesized from AEG backbone and the corresponding carboxylic acid derivatives (as exemplified for M in Fig. 10). Preferably, the carboxyl function of the AEG backbone should be protected as benzyl or allyl esters; however, a free AEG acid (R = H) can also be used if the required nucleobase is not compatible with benzyl or allyl cleavage conditions.[64-66, 96, 97] Synthesis procedures for AEG backbone are well developed;[98] therefore, the main challenge for the generation of new PNA monomers is the synthesis of carboxylic acids derivatives. We recently developed a straightforward and efficient synthesis of M monomer starting from 5-bromo-2-nitropyridine (Fig. 10) that makes M-modified PNAs readily available to a broader community of laboratories requiring only basic organic synthesis capabilities.[99] We and others have also published detailed protocols for preparation of other modified PNA monomers (for specific monomers, see references cited throughout the article) and nucleobase-modified PNA oligomers.[100-103]

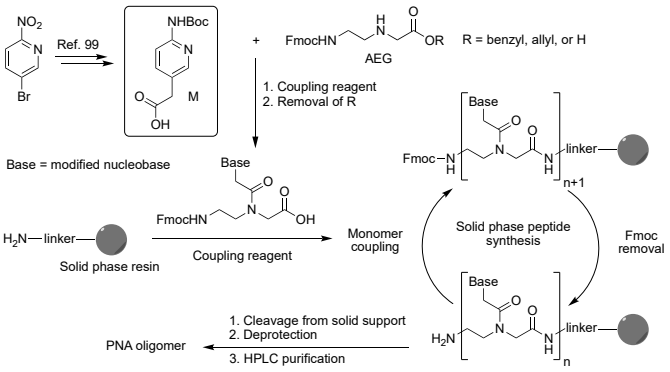

Fig. 10. General workflow for Fmoc solid phase synthesis of nucleobase-modified PNAs.

Conclusions

Molecular recognition of folded complex dsRNA has attracted less attention than recognition of dsDNA. However, since the discoveries of the various and intriguing regulatory functions that non-coding RNAs play in cell biology, the interest in RNA recognition is rapidly growing. In this area, triplex-forming PNAs offer unique advantages because of the unusually high affinity, sequence specificity, and programmable nature (e.g., the sequence of dsRNA determines the sequence of the PNA). As discussed above, several academic laboratories are exploring PNA-dsRNA triplexes and developing new modified PNAs to address remaining problems. Preliminary results suggest that PNAs are promising ligands to modulate biological function and control dynamic conformational equilibria of complex folded RNAs. The most significant bottleneck in this field remains the requirement for long polypurine tracts as effective solutions for stable X•C-G and Y•U-A triplets are still lacking. Additional concerns that generally apply for the use of PNAs are poor cellular uptake and bioavailability, and potential off-target effects in biological systems. Despite the challenges, the rapid progress and growing interest in PNA-dsRNA triplexes, since they were first reported in 2010, inspire confidence that these

problems can be addressed with focused collaborative efforts of nucleic acid chemists and biologists.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

The authors thank all students and postdoctoral associates who contributed to the PNA projects in their laboratories. Their names appear in the references cited throughout the article. The authors thank Vladislavs Baskevics for help with RNA and PNA images in Figure 2 and TOC. This work was supported by the Latvian Institute of Organic Synthesis internal research fund (IG-2019-04 and IG-2023-08 to M.K.), the National Science Foundation (CHE-1406433, CHE-1708761, and CHE-2107900 to E.R., and CHE-1708699 and CHE-2107911 to J.A.M.) and the National Institutes of Health (R01 GM71461 and R35 GM130207 to E.R.).

References

1. S. R. Eddy, *Nat. Rev. Genet.*, 2001, **2**, 919-929.
2. T. R. Mercer and J. S. Mattick, *Nat. Struct. Mol. Biol.*, 2013, **20**, 300-307.
3. K. N. Smith, S. C. Miller, G. Varani, J. M. Calabrese and T. Magnuson, *Genetics*, 2019, **213**, 1093.
4. R.-W. Yao, Y. Wang and L.-L. Chen, *Nat. Cell Biol.*, 2019, **21**, 542-551.
5. L. R. Ganser, M. L. Kelly, D. Herschlag and H. M. Al-Hashimi, *Nat. Rev. Mol. Cell Biol.*, 2019, **20**, 474-489.
6. K. D. Warner, C. E. Hajdin and K. M. Weeks, *Nat. Rev. Drug Discovery*, 2018, **17**, 547-558.
7. J. L. Childs-Disney, X. Yang, Q. M. R. Gibaut, Y. Tong, R. T. Batey and M. D. Disney, *Nat. Rev. Drug Discovery*, 2022, **21**, 736-762.
8. S. T. Crooke, B. F. Baker, R. M. Crooke and X.-h. Liang, *Nat. Rev. Drug Discovery*, 2021, **20**, 427-453.
9. S. T. Crooke, X.-H. Liang, B. F. Baker and R. M. Crooke, *J. Biol. Chem.*, 2021, **296**.
10. I. V. Novikova, S. P. Hennelly, C.-S. Tung and K. Y. Sanbonmatsu, *J. Mol. Biol.*, 2013, **425**, 3731-3746.
11. G. Devi, Y. Zhou, Z. Zhong, D.-F. K. Toh and G. Chen, *Wiley Interdisciplinary Reviews: RNA*, 2015, **6**, 111-128.
12. J. A. Brown, *WIREs RNA*, 2020, **11**, e1598.
13. N. K. Conrad, *WIREs RNA*, 2014, **5**, 15-29.
14. K. R. Fox, T. Brown and D. A. Rusling, in *Chemical Biology No. 7: DNA-targeting Molecules as Therapeutic Agents*, ed. M. J. Waring, Royal Society of Chemistry, 2018, vol. 7, ch. 1, pp. 1-32.
15. P. E. Nielsen, M. Egholm, R. H. Berg and O. Buchardt, *Science*, 1991, **254**, 1497-1500.
16. N. Brodyagin, M. Katkevics, V. Kotikam, C. A. Ryan and E. Rozners, *Beilstein J. Org. Chem.*, 2021, **17**, 1641-1688.
17. M. E. Hansen, T. Bentin and P. E. Nielsen, *Nucleic Acids Res.*, 2009, **37**, 4498-4507.
18. K. Aupeix, R. Le Tinevez and J. J. Toulme, *FEBS Lett.*, 1999, **449**, 169-174.
19. M. Li, T. Zengeya and E. Rozners, *J. Am. Chem. Soc.*, 2010, **132**, 8676-8681.
20. X. Zhan, L. Deng and G. Chen, *Biopolymers*, 2022, **113**, e23476.
21. T. Endoh, C. Annoni, D. Hnedzko, E. Rozners and N. Sugimoto, *Phys. Chem. Chem. Phys.*, 2016, **18**, 32002-32006.
22. T. Sato, N. Sakamoto and S. Nishizawa, *Org. Biomol. Chem.*, 2018, **16**, 1178-1187.
23. M. López-Tena, S.-K. Chen and N. Winssinger, *Bioconjugate Chem.*, 2023, **34**, 111-123.
24. M. Egholm, L. Christensen, K. L. Dueholm, O. Buchardt, J. Coull and P. E. Nielsen, *Nucleic Acids Res.*, 1995, **23**, 217-222.
25. A. Ono, P. O. P. Ts'o and L. S. Kan, *J. Am. Chem. Soc.*, 1991, **113**, 4032-4033.
26. R. Bahal, N. Ali McNeer, E. Quijano, Y. Liu, P. Sulkowski, A. Turchick, Y.-C. Lu, D. C. Bhunia, A. Manna, D. L. Greiner, M. A. Brehm, C. J. Cheng, F. Lopez-Giraldez, A. Ricciardi, J. Beloor, D. S. Krause, P. Kumar, P. G. Gallagher, D. T. Braddock, W. Mark Saltzman, D. H. Ly and P. M. Glazer, *Nat. Commun.*, 2016, **7**, 13304.
27. W. Li, H. Shi, B. Dong, K. Nie, Z. Liu and N. He, *Curr. Med. Chem.*, 2016, **23**, 4681-4705.
28. C. A. Ryan, N. Brodyagin, J. Lok and E. Rozners, *Biochemistry*, 2021, **60**, 1919-1925.
29. G. Devi, Z. Yuan, Y. Lu, Y. Zhao and G. Chen, *Nucleic Acids Res.*, 2014, **42**, 4008-4018.
30. S.-Q. Cao, I. Okamoto, H. Tsunoda, A. Ohkubo, K. Seio and M. Sekine, *Tetrahedron Lett.*, 2011, **52**, 407-410.
31. 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 and G. Chen, *Biochemistry*, 2018, **57**, 149-159.
32. 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 and E. Kierzek, *Bioconjugate Chem.*, 2019, **30**, 931-943.
33. T. J. Povsic and P. B. Dervan, *J. Am. Chem. Soc.*, 1989, **111**, 3059-3061.
34. S. Hildbrand, A. Blaser, S. P. Parel and C. J. Leumann, *J. Am. Chem. Soc.*, 1997, **119**, 5499-5511.
35. S. A. Cassidy, P. Slickers, J. O. Trent, D. C. Capaldi, P. D. Roselt, C. B. Reese, S. Neidle and K. R. Fox, *Nucleic Acids Res.*, 1997, **25**, 4891-4898.
36. T. Zengeya, P. Gupta and E. Rozners, *Angew. Chem., Int. Ed.*, 2012, **51**, 12593-12596.
37. M. S. Krishna, Z. Wang, L. Zheng, J. Bowry, A. A. L. Ong, Y. Mu, M. Prabakaran and G. Chen, *Biochemistry*, 2019, **58**, 3777-3788.
38. D. Hnedzko, D. W. McGee, Y. A. Karamitas and E. Rozners, *RNA*, 2017, **23**, 58-69.
39. R. W. Roberts and D. M. Crothers, *Science*, 1992, **258**, 1463-1466.
40. V. Kumar, N. Brodyagin and E. Rozners, *ChemBioChem*, 2020, **21**, 3410-3416.
41. C. A. Ryan and E. Rozners, *ACS Omega*, 2020, **5**, 28722–28729.
42. M. Patil Kiran, K. Toh Desiree-Faye, Z. Yuan, Z. Meng, Z. Shu, L. Ong Alan Ann, S. Krishna Manchugondanahalli, Y. Lu, G. Chen, H. Zhang and L. Lu, *Nucleic Acids Res.*, 2018, **46**, 7506-7521.

43. A. A. L. Ong, D.-F. K. Toh, M. S. Krishna, K. M. Patil, K. Okamura and G. Chen, *Biochemistry*, 2019, **58**, 3444-3453.
44. Y. Hari, S. Obika and T. Imanishi, *Eur. J. Org. Chem.*, 2012, **2012**, 2875-2887.
45. I. Prevot-Halter and C. J. Leumann, *Bioorg. Med. Chem. Lett.*, 1999, **9**, 2657-2660.
46. P. Gupta, T. Zengeya and E. Rozners, *Chem. Commun.*, 2011, **47**, 11125-11127.
47. N. Brodyagin, I. Kumpina, J. Applegate, M. Katkevics and E. Rozners, *ACS Chem. Biol.*, 2021, **16**, 872-881.
48. D. L. Chen and L. W. McLaughlin, *J. Org. Chem.*, 2000, **65**, 7468-7474.
49. Y. Hari, S. Matsugu, H. Inohara, Y. Hatanaka, M. Akabane, T. Imanishi and S. Obika, *Org. Biomol. Chem.*, 2010, **8**, 4176-4180.
50. A. B. Eldrup, O. Dahl and P. E. Nielsen, *J. Am. Chem. Soc.*, 1997, **119**, 11116-11117.
51. I. Kumpina, V. Baskevics, K. D. Nguyen, M. Katkevics and E. Rozners, *ChemBioChem*, 2023, **24**, e202300291.
52. L. C. Griffin, L. L. Kiessling, P. A. Beal, P. Gillespie and P. B. Dervan, *J. Am. Chem. Soc.*, 1992, **114**, 7976-7982.
53. C. Y. Huang and P. S. Miller, *J. Am. Chem. Soc.*, 1993, **115**, 10456-10457.
54. C.-Y. Huang, G. Bi and P. S. Miller, *Nucleic Acids Res.*, 1996, **24**, 2606-2613.
55. S. C. Zimmerman and P. Schmitt, *J. Am. Chem. Soc.*, 1995, **117**, 10769-10770.
56. T. E. Lehmann, W. A. Greenberg, D. A. Liberles, C. K. Wada and P. B. Dervan, *Helv. Chim. Acta*, 1997, **80**, 2002-2022.
57. H. Han and P. B. Dervan, *Proc. Natl. Acad. Sci. U.S.A.*, 1993, **90**, 3806-3810.
58. A. Semenyuk, E. Darian, J. Liu, A. Majumdar, B. Cuenoud, P. S. Miller, A. D. MacKerell, Jr. and M. M. Seidman, *Biochemistry*, 2010, **49**, 7867-7878.
59. D. Guianvarc'h, R. Benhida, J.-L. Fourrey, R. Maurisse and J.-S. Sun, *Chem. Commun.*, 2001, 1814-1815.
60. D. A. Rusling, V. E. C. Powers, R. T. Ranasinghe, Y. Wang, S. D. Osborne, T. Brown and K. R. Fox, *Nucleic Acids Res.*, 2005, **33**, 3025-3032.
61. 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.
62. 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.
63. 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 and G. Chen, *Biochemistry*, 2019, **58**, 1319-1331.
64. I. Kumpina, N. Brodyagin, J. A. MacKay, S. D. Kennedy, M. Katkevics and E. Rozners, *J. Org. Chem.*, 2019, **84**, 13276-13298.
65. 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.
66. J. M. Talbott, B. R. Tessier, E. E. Harding, G. D. Walby, K. J. Hess, V. Baskevics, M. Katkevics, E. Rozners and J. A. MacKay, *Chem. Eur. J.*, 2023, **29**, e202302390.
67. C. A. Ryan, V. Baskevics, M. Katkevics and E. Rozners, *Chem. Commun.*, 2022, **58**, 7148-7151.
68. L. Betts, J. A. Josey, J. M. Veal and S. R. Jordan, *Science*, 1995, **270**, 1838-1841.
69. V. Kotikam, S. D. Kennedy, J. A. MacKay and E. Rozners, *Chem. Eur. J.*, 2019, **25**, 4242.
70. T. Sato, Y. Sato and S. Nishizawa, *Biopolymers*, 2022, **113**, e23474.
71. U. Koppelhus and P. E. Nielsen, *Advanced Drug Delivery Reviews*, 2003, **55**, 267-280.
72. E. Riguet, S. Tripathi, B. Chaubey, J. Desire, V. N. Pandey and J.-L. Decout, *J. Med. Chem.*, 2004, **47**, 4806-4809.
73. N. Brodyagin, Y. Kataoka, I. Kumpina, D. W. McGee and E. Rozners, *Biopolymers*, 2022, **113**, e23484.
74. T. Endoh, D. Hnedzko, E. Rozners and N. Sugimoto, *Angew. Chem., Int. Ed.*, 2016, **55**, 899-903.
75. T. Endoh, N. Brodyagin, D. Hnedzko, N. Sugimoto and E. Rozners, *ACS Chem. Biol.*, 2021, **16**, 1147-1151.
76. T. Hermann and D. J. Patel, *Structure*, 2000, **8**, R47-R54.
77. C. A. Ryan, M. Rahman, Md and E. Rozners, *J. Am. Chem. Soc.*, 2023, **145**, 10497–10504.
78. J. Saarbach, P. M. Sabale and N. Winssinger, *Curr. Opin. Chem. Biol.*, 2019, **52**, 112-124.
79. T. Vilaivan, *Beilstein J. Org. Chem.*, 2018, **14**, 253-281.
80. T. Sato, Y. Sato and S. Nishizawa, *J. Am. Chem. Soc.*, 2016, **138**, 9397-9400.
81. O. Koehler and O. Seitz, *Chem. Commun.*, 2003, DOI: 10.1039/b308299g, 2938-2939.
82. F. Hoevelmann and O. Seitz, *Acc. Chem. Res.*, 2016, **49**, 714-723.
83. T. Sato, Y. Sato and S. Nishizawa, *Chem. Eur. J.*, 2017, **23**, 4079-4088.
84. Y. Yoshino, Y. Sato and S. Nishizawa, *Anal. Chem.*, 2019, **91**, 14254-14260.
85. T. Chiba, T. Sato, Y. Sato and S. Nishizawa, *Org. Biomol. Chem.*, 2017, **15**, 7765-7769.
86. M. S. Krishna, D.-F. K. Toh, Z. Meng, A. A. L. Ong, Z. Wang, Y. Lu, K. Xia, M. Prabakaran and G. Chen, *Anal. Chem.*, 2019, **91**, 5331-5338.
87. P. M. Sabale and S. G. Srivatsan, *ChemBioChem*, 2016, **17**, 1665-1673.
88. E. T. T. Lee, Y. Sato and S. Nishizawa, *Chem. Commun.*, 2020, **56**, 14976-14979.
89. Y. Sato, H. Miura, T. Tanabe, C. U. Okeke, A. Kikuchi and S. Nishizawa, *Anal. Chem.*, 2022, **94**, 7814-7822.
90. C. U. Okeke, H. Miura, Y. Sato and S. Nishizawa, *Org. Biomol. Chem.*, 2023, **21**, 3402-3410.
91. T. Tanabe, T. Sato, Y. Sato and S. Nishizawa, *RSC Advances*, 2018, **8**, 42095-42099.
92. Y. Sato, Y. Takahashi, T. Tanabe and S. Nishizawa, *Org. Biomol. Chem.*, 2020, **18**, 4009-4013.
93. K. T. Kim, D. Chang and N. Winssinger, *Helv. Chim. Acta*, 2018, **101**, e1700295.
94. D. Hnedzko, D. W. McGee and E. Rozners, *Bioorg. Med. Chem.*, 2016, **24**, 4199.
95. K. P. Nandhini, D. A. Shaer, F. Albericio and B. G. de la Torre, *Chem. Soc. Rev.*, 2023, **52**, 2764-2789.
96. F. Debaene, J. A. Da Silva, Z. Pianowski, F. J. Duran and N. Winssinger, *Tetrahedron*, 2007, **63**, 6577-6586.
97. S. Pothukanuri, Z. Pianowski and N. Winssinger, *Eur. J. Org. Chem.*, 2008, **2008**, 3141-3148.
98. T. A. Feagin, N. I. Shah and J. M. Heemstra, *J. Nucleic Acids*, 2012, **2012**, 354549.
99. I. Kumpina, V. Baskevics, G. D. Walby, B. R. Tessier, S. Farshineh Saei, C. A. Ryan, J. A. MacKay, M. Katkevics and

E. Rozners, *Synlett*, 2024, **published online**, DOI: 10.1055/a-2191-5774.

100. N. Brodyagin, D. Hnedzko, J. A. MacKay and E. Rozners, in *Peptide Nucleic Acids: Methods and Protocols*, ed. P. E. Nielsen, Springer US, New York, NY, 2020, pp. 157-172.
101. A. M. Hansen, G. Bonke, W. F. J. Hogendorf, F. Björkling, J. Nielsen, K. T. Kongstad, D. Zabicka, M. Tomczak, M. Urbas, P. E. Nielsen and H. Franzyk, *Eur. J. Med. Chem.*, 2019, **168**, 134-145.
102. A. Y. Shaikh, F. Björkling, P. E. Nielsen and H. Franzyk, *Eur. J. Org. Chem.*, 2021, **2021**, 2792-2801.
103. A. Y. Shaikh, A. M. Hansen and H. Franzyk, in *Peptide Nucleic Acids: Methods and Protocols*, ed. P. E. Nielsen, Springer US, New York, NY, 2020, pp. 1-16.