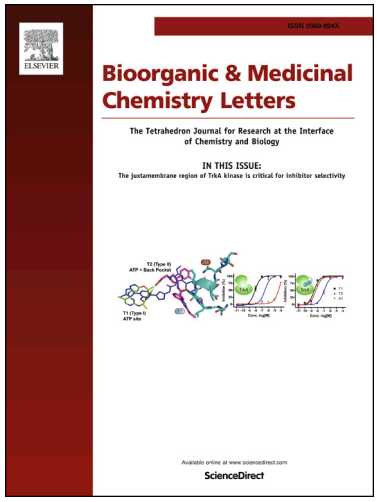
Accepted Manuscript



Assessment of new triplet forming artificial nucleobases as RNA ligands directed towards HCV IRES IIId loop

Mauro Safir Filho, Anthony R. Martin, Rachid Benhida

PII: S0960-894X(17)30204-4
DOI: http://dx.doi.org/10.1016/j.bmcl.2017.02.061
Reference: BMCL 24733

To appear in: *Bioorganic & Medicinal Chemistry Letters*

Received Date: 16 December 2016
Revised Date: 21 February 2017
Accepted Date: 23 February 2017

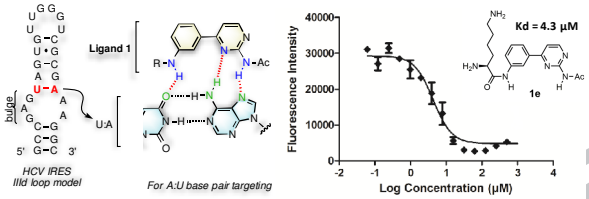
Please cite this article as: Filho, M.S., Martin, A.R., Benhida, R., Assessment of new triplet forming artificial nucleobases as RNA ligands directed towards HCV IRES IIId loop, *Bioorganic & Medicinal Chemistry Letters* (2017), doi: http://dx.doi.org/10.1016/j.bmcl.2017.02.061

Graphical Abstract

Assessment of new triplet forming artificial nucleobases as RNA ligands directed towards HCV IRES IIId loop.

Mauro Safir Filho, Anthony R. Martin and Rachid Benhida*

Leave this area blank for abstract info.



Assessment of new triplet forming artificial nucleobases as RNA ligands directed towards HCV IRES IIId loop.

Mauro Safir Filho[ab], Anthony R. Martin[a], and Rachid Benhida[a,*]

[a]*Université Côte d'Azur, CNRS, Institut de Chimie de Nice UMR7272 – 06108 NICE, France;* [b]*CAPES foundation, Ministry of Education of Brazil, Brasília – DF 70040-020, Brazil.*

Abstract: We report the synthesis of two new artificial nucleobase scaffolds, **1** and **2**, featuring adequate hydrogen bonding donors and acceptors for the molecular recognition of U:A and C:G base pairs, respectively. The tethering of these structures to various amino acids and the assessment of these artificial nucleobase-amino acid conjugates as RNA ligands against a model of HCV IRES IIId domain are also reported. Compound **1e** displayed the highest affinity (K_d twice lower than neomycin – control). Moreover, it appears that this interaction is enthalpically and entropically favored.

Keywords: Hepatitis virus C; IRES IIId loop; RNA ligands; Inhibitors; Binding assay

Main Text:

Hepatitis C virus (HCV) affects about 170-million people worldwide hence representing a major health problem. It is a main cause of chronic hepatitis and may lead to hepatocellular carcinoma and liver cirrhosis.[1] HCV is a single-stranded positive-sense RNA virus belonging to the *flaviviridae* family. Its genome is about 9,600 bases length and the several strains can be divided into 7 genotypes and 67 subtypes.[2] Until recently, only few treatments against HCV were available, ribavirin and (pegylated)-interferon, while exhibiting limited efficacy. In 2013, the FDA approval of sofosbuvir, a nucleotide inhibitor of the NS5B viral polymerase, revolutionized HCV treatment. It yields impressive sustained virological response (SVR) and limited adverse effects.[3] Nevertheless, this treatment is contraindicated for patients suffering severe renal impairments and its efficacy on HCV genotype 3 is not satisfactory.[4] Thus, new complementary approaches to circumvent HCV infection are still needed. One of them focuses on the inhibition of HCV translation by targeting its Internal Ribosomal Entry Site (IRES).[5,6]

HCV IRES is a highly conserved sequence located in the 5'-untranslated region (5'-UTR) of HCV genome and composed of three folding domains (II-IV). It allows the recruitment of the ribosomal machinery together with a restricted number of eukaryotic initiation factors (eIFs) to promote the translation of the viral RNA in a 5'-cap-independent fashion.[7] Thus, many efforts have been devoted to the development of new drugs able to inhibit the HCV translation by hampering its IRES functions.[8] In particular, the discovery of ligands targeting the IRES IIId loop is highly relevant as this domain is of utmost importance for the interaction with the 40S ribosomal

* Corresponding author. Tel.: +33-4-9207-6143; fax: +33-4-9207-6189; e-mail: benhida@unice.fr

subunit.[9] Known ligands for this IRES IIId loop encompass antisense oligonucleotides[10-12] and RNA aptamers[13] as well as small-sized organic molecules whose structures remain undisclosed.[14]

While non-coding RNA molecules are even more recognized as key players in numerous biological processes, including gene regulation, tumorigenesis, viral translation, etc. they also appear as important druggable targets. For this purpose, several research group, including ours, devoted much effort to develop small-sized organic molecule targeting RNAs.[15-18] Our ligand design consisted in the combination of molecular recognition elements to enhance site specificity with electrostatic interactions to strengthen the complex stability. The preparation of these multimodal ligands was accomplished by assembling the nucleobase analog "**S**", able to form triplets through Hoogsteen interactions with U:A base pair, with basic amino acid residues (Figure 1). This strategy, based on an unprecedented rational design of RNA ligands, was successfully applied to the discovery of new specific binders of the stem-bulge of HIV1 TAR RNA.[19,20]

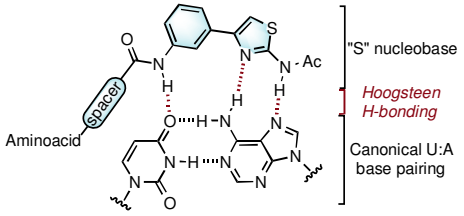


Figure 1. Amino acid tethered "**S**" nucleobase in a "**S**"-U:A triplet.

In continuation of our efforts to produce innovative RNA ligands using our triplex-based rational design, we present herein the synthesis of two new artificial nucleobase scaffolds, **1** and **2**, featuring adequate hydrogen bonding donors and acceptors for the molecular recognition of U:A and C:G base pairs, respectively. The tethering of these structures to various amino acids is also reported (Figure 2). It is worth noting that while nucleobases **1** and **2** may involve three hydrogen bonds to increase the specificity (recognition of U:A *versus* C:G), the amino acid counterparts are critical for high affinity binding (H-bonding and electrostatic ammonium-phosphate interaction). Finally, we disclose the binding affinity of our novel structures to a model of HCV IRES IIId domain.

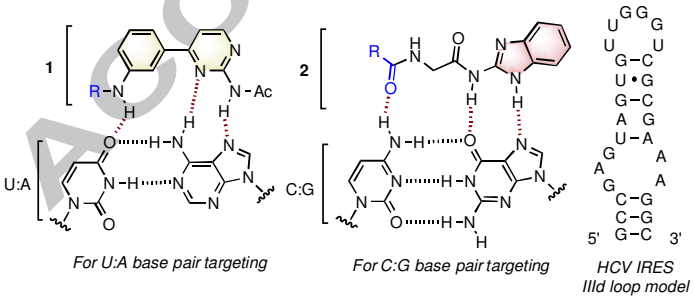
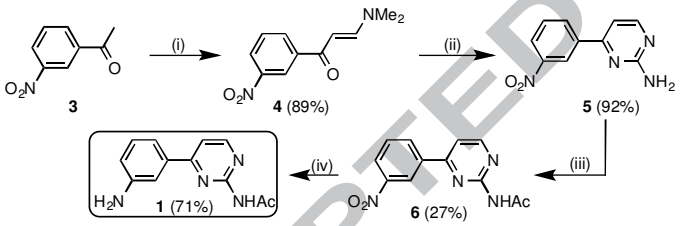


Figure 2. Structures of **1**, **2** and HCV IRES IIId loop model.

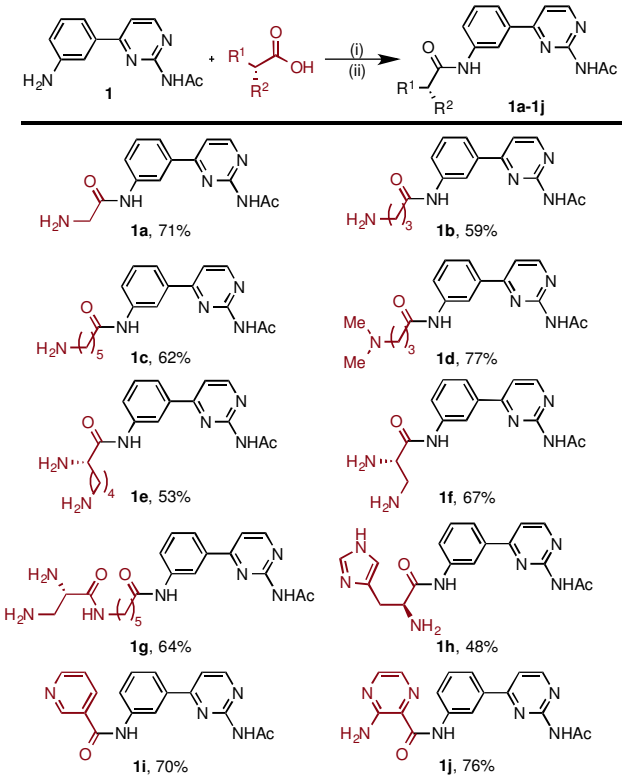
Design of the platforms **1** and **2** was achieved by analogy with the existing structures of artificial nucleobases known to establish Hoogsteen base pairing in triplex-forming oligonucleotides.[21] Thus, the molecular structure of **1** is an analog of the artificial nucleobase "**S**", already reported by our group, where the thiazole ring is replaced by a pyrimidine moiety. It is worth noting that the aminopyrimidine system should account for more H-bonding ability and π-stacking interactions along the RNA helix, hence modulating the binding properties. Moreover, this replacement will also affect the possible stacking interactions. Similarly to **1**, the structure of **2** is analogous to a benzimidazolylglycine compound (**BIG**) that was reported by Sasaki and Maeda[22] and further studied by our group in aminoglycoside conjugates.[23] Notably, the **BIG** moiety have not been linked to amino acids. Thus, we herein complement this study and will use these data for comparative purpose. This should allow us to identify which base pair is best suited for IRES IIId loop targeting using our multimodal ligands.

The first step of our study consists of the synthesis of platforms **1** and **2** and their derivatization with various amino acids. Thus, we first undertook the preparation of **1**, following a linear 4-step strategy (Scheme 1, steps *i-iv*) starting from *m*-nitroacetophenone (**3**). Firstly, **3** was homologated into the corresponding enaminone **4** using the *N,N*-dimethylamine dimethylacetal reagent. Next, the construction of the 2-aminopyrimidine ring was performed by cyclization of **4** with guanidine in *n*BuOH to afford the biaryl compound **5**. Then, the free amino function of **5** was acylated using acetyl chloride in anhydrous pyridine. This acylation provided **6** in moderate yields (27%) probably owing to the low nucleophilicity of the 2-aminopyrimidine moiety. Finally, compound **1** was obtained after reduction of the nitro function into an amino group that will serve as anchoring point for various amino acids tethering (Scheme 2).



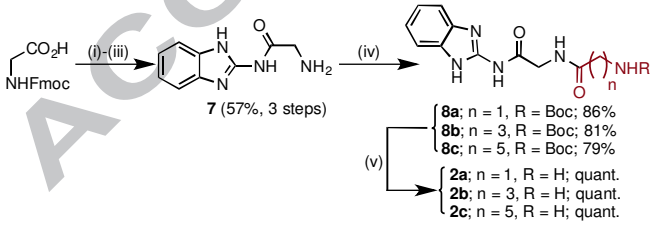
Scheme 1. Synthetic pathway for the preparation of **1**. Reagents and conditions (i) *N,N*-dimethylformamide dimethylacetal, xylene, 140°C, 12 h; (ii) guanidine hydrochloride, nBuOH, 100°C, 24 h; (iii) AcCl, pyridine, 0-50°C, 12 h; (iv) $NaBH_4$, Pd/C, MeOH, t.a., 3 h.

With compound **1** in hands, we proceeded to its connection and unmasking to various amino acids through their carboxylic acid function. These structures, **1a-1j,** were obtained in decent to good yields (48-77%) and are summarized on Scheme 2. The variety of connected amino acids was selected to evaluate the structure-binding relationship in terms of (1) chain length between the carboxylic acid and the amine, (2) number and pK_a of amino functions, (3) presence of an α-amino group or an aromatic ring.



Scheme 2. Synthesis and structures of ligands **1a-1j**. Reagents and conditions (i) *N*-Boc protected amino acid, HBTU, HOBt, DIEA, DMF, 40°C, 24 h; (ii) TFA, CH_2Cl_2, r.t., overnight. Yields are isolated overall yields over the 2 steps.

Following the preparation of the first family of ligands, we proceeded to the synthesis of the second series (**2a-2c**) designed for C:G base pair targeting (**2**). As aforementioned, this second set of compounds was prepared for comparative purpose with the first family (**1a-1j**). The synthesis of **2a-2c**, was performed in a 4-step sequence. Starting from the *N*-Fmoc-glycine, after the formation of its corresponding acid chloride (Scheme 3, step *i*), the amide bond was established with the 2-aminobenzimidazole and the *N*-Fmoc was subsequently unmasked, thus affording **7** in a 57% overall yield (3 steps). Next, the free amine was engaged in a peptide coupling with three different *N*-Boc protected amino acids to yield **8a-8c**. Finally, acidic treatment prompted the *N*-Boc deprotection and furnished **2a-2c** in 79-86% yields over two steps (Scheme 3).



Scheme 3. Synthetic pathway for the preparation of **1a-1j**. Reagents and conditions (i) thionyl chloride, THF/CH_2Cl_2, 60°C, 0.5 h; (ii) 2-aminobenzimidazole, DMAP, DIEA, DMF, 100°C, 24 h; (iii) piperidine, DMF, r.t., 2 h; (iv) *N*-Boc protected amino acid, HBTU, HOBt, DIEA, DMF, 40°C, 24 h; (v) TFA, CH_2Cl_2, r.t., overnight.

With these fully characterized new compounds in hand (**1a-1j** and **2a-2c**),[24] we proceeded to their evaluation as ligands of the RNA model of HCV IRES IIId loop presented on figure 2. We used for this study a classical fluorescence-based assay for quantification of RNA-ligand interactions (IRES fragment labelled at its 5'-end with a fluorescent reporter Alexa488®). Indeed, the dissociation constant (K_d) of each ligand was determined by plotting the fluorescence intensity as a function of the ligand concentrations. Neomycin, a well-known non-specific RNA binder[20] was used as a positive control. Based on these results concerning compounds **1a-1j** (Table 1), some structure-binding relationship (SBR) could be drawn. We observed that the nature of the linked amino acid dramatically influences the $^{20°C}K_d$ with values ranging from 4.3 to 863.4 µM; neomycin exhibiting a K_d of 8.5 µM. Interestingly, the nucleobase **1** with no appended amino acids, or the lysine amino acid alone are unable to bind IRES RNA (Table 1, entry 15 & 16), demonstrating therefore that both domains (nucleoabse and amino acid) cooperate in binding to RNA. The first trend of this SBR concerns the length of the hydrocarbon chain between carbonyl and terminal amine of the amino acid moiety. Its lengthening from one to five methylene units correlates with a 73-fold decrease in K_d (Table 1, entries 2-4). Additionally, switching from a primary terminal amine to a more basic tertiary dimethylamino function is also beneficial in terms of binding (Table 1, entry 3 *vs* entry 5). The number of amino group and H-bond donor/acceptor is also highly important as their increase favors the binding (entries 6-8). Nevertheless, this last point must be in balance with an appropriate chain length for an optimal effect. The use of histidine or other aromatic amino acids (entries 9-11) displayed limited efficacy. Finally, concerning the three compounds belonging to the 2-aminobenzimidazole series (**2a-2c**), their binding were too weak to be measured. In the present study, the ligand exhibiting the highest affinity for IRES IIId loop is **1e** with a K_d of 4.3 µM, two-fold more potent than neomycin. While it obviously does not compare favorably with oligonucleotide ligands,[10-12] this K_d value falls in the same range of other small molecule ligands of IRES IIa loop.[8]

Table 1. K_d values and thermodynamic parameters for ligand/IRES IIId interaction.

Entry	Ligand	Number of amines[a]	$^{20°C}K_d$ (µM)[b]	ΔG° (kJ/mol)	ΔH° (kJ/mol)	TΔS° (kJ/mol)
1	Neomycin	6	8.5	-28.1	-39.8 ± 4.4	-11.7 ± 4.4
2	**1a**	1	863.4	n.a.[c]	n.a.[c]	n.a.[c]
3	**1b**	1	220.5	n.a.[c]	n.a.[c]	n.a.[c]
4	**1c**	1	11.8	-27.7	-3.1 ± 0.4	+24.6 ±0.4
5	**1d**	1	36.9	-25.2	-8.0 ± 1.1	+17.2 ± 1.1
6	**1e**	2	4.3	-30.0	-25.7 ± 2.6	+4.3 ± 2.7
7	**1f**	2	10.1	-27.8	-40.3 ± 3.1	-12.5 ± 3.2
8	**1g**	2	16.8	-26.5	-33.8 ± 2.4	-7.3 ± 2.4
9	**1h**	2	19.7	-25.9	-48.5 ± 5.1	-22.6 ± 5.2

10	**1i**	1	659.3	n.a.[c]	n.a.[c]	n.a.[c]
11	**1j**	1	232.3	n.a.[c]	n.a.[c]	n.a.[c]
12	**2a**	1	n.b.[d]	n.a.[c]	n.a.[c]	n.a.[c]
13	**2b**	1	n.b.[d]	n.a.[c]	n.a.[c]	n.a.[c]
14	**2c**	1	n.b.[d]	n.a.[c]	n.a.[c]	n.a.[c]
15	**1**	0	n.b.[d]	n.a.[c]	n.a.[c]	n.a.[c]
16	**Lysine**	2	n.b.[d]	n.a.[c]	n.a.[c]	n.a.[c]

[a]Number of amino functions, *i.e.* sites of protonation on the amino acid side chain; [b]Kd measurements were performed at 20°C in a 20 mM HEPES buffer (pH = 7.4) containing 20 mM of NaCl, 140 mM of KCl and 3 mM of $MgCl_2$ – K_d values are reported with an uncertainty of ± 10%.[20] [c]n.a.: not accessible; [b]n.b.: no binding at >30 °C.

Following the determination of the K_d of each ligand for IRES IIId at 20°C, we complemented these datasets by measuring the equilibrium constants at various temperature ranging from 5 to 36°C; thus we could access the corresponding thermodynamic parameters (Table 1).[24] It appears that the Gibbs free energy variations ($\Delta G°$) fall in the same range for all the ligands ($\Delta\Delta G°_{max}$ = 4.1 kJ/mol); although the enthalpic and entropic contributions may vary drastically. In fact, compounds exhibiting the longest flexible alkylamino side chains tend to bind their RNA target in an entropically driven fashion (*i.e.* **1c-1e**). This could be due to some degrees of residual motility of the formed complex and/or desolvation contribution.[25] Counterintuitively, **1g** binding seems to be enthalpically driven. This could be explained by the presence of two amines and one amide function at the end of **1g** side chain, which might lock the conformation of the ligand at its binding site, hence minimizing the entropic contribution. Finally, ligands displaying short side chain bearing H-bond donors/acceptors (**1f** & **1h**) feature the largest enthalpic contribution.

Having evaluated the affinity of these new ligands for IRES IIId RNA, we selected the best ligand (**1e**) and its related congener (**1c**) to study their specificity towards the unique U:A base pair located at the stem-bulge junction. For this purpose, we used a mutated sequence of the IRES IIId (Mut-IIId) in which the U:A base pair is replaced by a C:G (Table 2). Importantly, we found that the binding of **1e** to Mut-IIId was drastically decreased with more than 100-fold compared with the wild-type sequence (K_d values 4.3 *vs* 554.4 µM, Table 2, entry 1), proving therefore the high selectivity for U:A over C:G. The same behavior was observed using the ligand **1c** since the corresponding K_d was too high to be determined (Table 2, entry 2). These pieces of evidence confirmed again that the ligands derived from **1** effectively target the U:A base pair of IRES IIId. To further support this result, we assessed the ligand **2c** against Mut-IIId. While **2c** does not bind at all to the wild-type sequence, it displayed a K_d value of 41.4 µM against Mut-IIId (Table 2, entry 3). Since several C:G base pairs are present in Mut-IIId and the wild-type IRES IIId sequences, this last observation suggests that our ligands feature some degree of site specificity and are tailored to target the stem-bulge junction where the U:A base pair is located. Finally, we evaluated the ability of our ligands **1e** and **1c** to bind the IRES IIId loop sequence in the presence of natural tRNA competitors (Table 2). To do so, we measured the K_d value of **1e** and **1c** againt IRES IIId in the presence of a 50-fold excess of a

mixture of natural tRNA. This competition experiment revealed that **1e** and **1c** are able to selectively target the IIId loop of IRES since only a marginal increase of the corresponding K_d values was observed (2.4- and 1.4-fold, respectively).

Table 2. K_d determination to probe the site-selectivity and the specificity of the ligands.

Entry	Ligand	$^{20°C}K_d$ (μM)[a]		
		IRES IIId	Mut-IIId[b]	IRES IIId + tRNA[c]
1	**1e**	4.3	554.4	10.5
2	**1c**	11.8	n.b.[d]	17.1
3	**2c**	n.b.[d]	41.4	—

[a]Kd measurements were performed at 20°C in a 20 mM HEPES buffer (pH = 7.4) containing 20 mM of NaCl, 140 mM of KCl and 3 mM of $MgCl_2$ – K_d values are reported with an uncertainty of ± 10%.[20] [b]Sequence of Mut-IIId: 5'-GCCGAGGAGUGUUGGGUCGCGCAAGGC-3'. [c]Measured in the presence of a 50-fold excess of a mixture of natural tRNA [d]n.b.: no binding.

In conclusion, we reported herein two new series of rationally design multimodal RNA ligands featuring new artificial nucleobases. These compounds proved easily accessible in only 4 to 6 steps. We presented their ability to bind to a IIId loop model of HCV IRES and compared their potency with neomycin. It is noteworthy that this study discloses the first structures of small-molecule ligands of IRES IIId loop. Low micromolar dissociation constant could be obtained with a two-fold lower value for our best ligand (**1e**) compared with neomycin. Moreover, it seems that its binding is both enthalpically and entropically favored ($T\Delta S° = +4.3$ kJ/mol). This study validates our rational design strategy and could now be used to design new series of multimodal RNA ligands and paves the way towards more potent compounds.

Acknowledgments

This work was supported by CNRS, University of Nice Sophia Antipolis and CAPES (fellowship to MSF, process number 99999.001495/2015-01).

References and notes

1. Webster, D. P.; Klenerman, P.; Dusheiko, G. M. *Lancet* **2015**, *385*, 1124–1135.
2. Smith, D. B.; Bukh, J.; Kuiken, C.; Muerhoff, A. S.; Rice, C. M.; Stapleton, J. T.; Simmonds, P. *Hepatology* **2014**, *59*, 318–327.
3. Gentile, I.; Maraolo, A. E.; Buonomo, A. R.; Zappulo, E.; Borgia, G. *Expert Opin. Drug Discovery* **2015**, *10*, 1363–1377.
4. Nakamura, M.; Kanda, T.; Haga, Y.; Sasaki, R.; Wu, S.; Nakamoto, S.; Yasui, S.; Arai, M.; Imazeki, F.; Yokosuka, O. *World J. Hepatol.* **2016**, *8*, 183–190.
5. Gallego, J.; Varani, G. *Biochem. Soc. Trans.* **2002**, *30*, 140–145.
6. Davis, D. R.; Seth, P. P. *Antiviral Chem. Chemother.* **2011**, *21*, 117–128.
7. Lukavsky, P. J. *Virus Res.* **2009**, *139*, 166–171.
8. Dibrov, S. M.; Parsons, J.; Carnevali, M.; Zhou, S.; Rynearson, K. D.; Ding, K.; Garcia Sega, E.; Brunn, N. D.; Boerneke, M. A.; Castaldi, M. P.; Hermann, T. *J. Med. Chem.* **2014**, *57*, 1694–1707.
9. Angulo, J.; Ulryck, N.; Deforges, J.; Chamond, N.; Lopez-Lastra, M.; Masquida, B.; Sargueil, B. *Nucleic Acids Res.* **2015**, gkv1325–17.
10. Jubin, R.; Vantuno, N. E.; Kieft, J. S.; Murray, M. G.; Doudna, J. A.; Lau, J. Y.; Baroudy, B. M. *J. Virol.* **2000**, *74*, 10430–10437.

11. Hanecak, R.; Brown-Driver, V.; Fox, M. C.; Azad, R. F.; Furusako, S.; Nozaki, C.; Ford, C.; Sasmor, H.; Anderson, K. P. *J. Virol.* **1996**, *70*, 5203–5212.
12. Tallet-Lopez, B.; Aldaz-Carroll, L.; Chabas, S.; Dausse, E.; Staedel, C.; Toulmé, J.-J. *Nucleic Acids Res.* **2003**, *31*, 734–742.
13. Kikuchi, K.; Umehara, T.; Fukuda, K.; Kuno, A.; Hasegawa, T.; Nishikawa, S. *Nucleic Acids Res.* **2005**, *33*, 683–692.
14. Baugh, C.; Shaohui Wang; Bin Li; Appleman, J. R.; Thompson, P. A. *J. Biomol. Screening* **2009**, *14*, 219–229.
15. Thomas, J. R.; Hergenrother, P. J. *Chem. Rev.* **2008**, *108*, 1171–1224.
16. Guan, L.; Disney, M. D. *ACS Chem. Biol.* **2012**, *7*, 73–86.
17. Hermann, T. *WIREs RNA* **2016**, *7*, 726–743.
18. Di Giorgio, A.; Tran, T. P. A.; Duca, M. *Future Med. Chem.* **2016**, *8*, 803–816.
19. Duca, M.; Malnuit, V.; Barbault, F.; Benhida, R. *Chem. Commun.* **2010**, *46*, 6162.
20. Joly, J.-P.; Mata, G.; Eldin, P.; Briant, L.; Fontaine-Vive, F.; Duca, M.; Benhida, R. *Chem. Eur. J.* **2014**, *20*, 2071–2079.
21. Malnuit, V.; Duca, M.; Benhida, R. *Org. Biomol. Chem.* **2011**, *9*, 326–336.
22. Sasaki, S.; Nakashima, S.; Nagatsugi, F.; Tanaka, Y.; Hisatome, M.; Maeda, M. *Tetrahedron Lett.* **1995**, *36*, 9521–9524.
23. Vo, D. D.; Staedel, C.; Zehnacker, L.; Benhida, R.; Darfeuille, F.; Duca, M. *ACS Chem. Biol.* **2014**, *9*, 711–721.
24. *See Electronic Supplementary Information (ESI) for more details.*
25. Klebe, G. *Nat Rev Drug Discov* **2015**, *14*, 95–110.

Supplementary Material

Supplementary material includes the synthetic procedures for the preparation of all reported compounds, their characterization, copies of ^{1}H, ^{13}C and HRMS spectra; experimental procedures for the determination of K_d and thermodynamic parameters.