Artificial Base Pairs

International Edition: DOI: 10.1002/anie.201504731
German Edition: DOI: 10.1002/ange.201504731

A Crystal Structure of a Functional RNA Molecule Containing an Artificial Nucleobase Pair

Armando R. Hernandez, Yaming Shao, Shuichi Hoshika, Zunyi Yang, Sandip A. Shelke, Julien Herrou, Hyo-Joong Kim, Myong-Jung Kim, Joseph A. Piccirilli, and Steven A. Benner**

In memory of John M. (Jack) Templeton, Jr.

Abstract: *As one of its goals, synthetic biology seeks to increase the number of building blocks in nucleic acids. While efforts towards this goal are well advanced for DNA, they have hardly begun for RNA. Herein, we present a crystal structure for an RNA riboswitch where a stem C:G pair has been replaced by a pair between two components of an artificially expanded genetic-information system (AEGIS),* ***Z*** *and* ***P****, (6-amino-5-nitro-2(1H)-pyridone and 2-amino-imidazo[1,2-a]-1,3,5-triazin-4-(8H)-one). The structure shows that the* ***Z*****:****P** *pair does not greatly change the conformation of the RNA molecule nor the details of its interaction with a hypoxanthine ligand. This was confirmed in solution by in-line probing, which also measured a 3.7* nM *affinity of the riboswitch for guanine. These data show that the* ***Z*****:****P** *pair mimics the natural Watson–Crick geometry in RNA in the first example of a crystal structure of an RNA molecule that contains an orthogonal added nucleobase pair.*

The possibility of increasing the number of replicable nucleotides in DNA and RNA (collectively xNA) beyond the four standard nucleotides (GACT or GACU) found in natural genetic material was noted a quarter-century ago.[1,2] Since then, two contrasting strategies based on different "concepts" to expand the genetic alphabet have been explored.

In the first, additional nucleobase pairs were designed to fit the Watson–Crick pairing geometry, with pairing specificity determined by two kinds of complementarity: 1) size complementarity, where large purines pair with small pyrimidines, and 2) hydrogen bonding complementarity, where hydrogen-bond donors on one nucleobase match hydrogen-bond acceptors on the other. By shuffling hydrogen-bond donor and acceptor groups within this framework, up to 12 nucleobases forming six orthogonal nucleobase pairs are readily conceived (Figure 1). This has been called an "artificially expanded genetic information system" (AEGIS).[3]

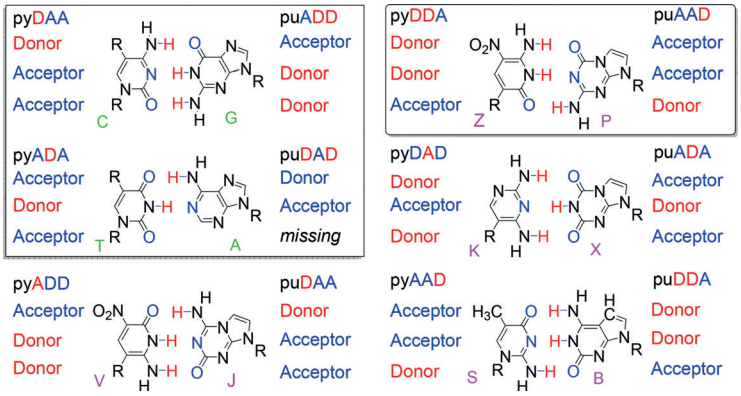


Figure 1. Shuffling hydrogen-bond donor and acceptor units in paired nucleobases adds eight new nucleotides that form four orthogonal nucleobase pairs following the Watson–Crick "concept". These are in addition to the two pairs formed from the four standard nucleobases (box upper left). The **Z**:**P** pair examined herein is in the upper right box.

A second concept, proposed originally by Kool and co-workers, dispenses with inter-nucleobase hydrogen bonding entirely.[4–6] Instead, specific pairing is sought through geometric complementarity alone. Expanding on this concept, Hirao and co-workers[7–9] and Romesberg and co-workers[10–12] more recently designed additional pairs that also lack inter-nucleobase hydrogen bonding, seeking specific and orthogonal pairing by way of simple steric complementarity. Romesberg and co-workers were even able to create an *E. coli* strain that maintained one example of his non-standard pair in a plasmid for a number of hours.[13]

Both strategies have been shown to increase the functional potential of DNA. For example, Hirao and co-workers recently showed that adding their fifth nucleotide to a DNA aptamer increased its affinity for its target.[7] More recently, we showed that oligonucleotides built from six-letter AEGIS

[*] Dr. A. R. Hernandez,[+] Dr. Y. Shao,[+] Dr. S. A. Shelke, Dr. J. Herrou, Prof. J. A. Piccirilli
Department of Biochemistry and Molecular Biology
Department of Chemistry
University of Chicago
Chicago, IL 60637 (USA)
E-mail: jpicciri@uchicago.edu
Dr. S. Hoshika,[+] Z. Yang, Dr. H.-J. Kim, Dr. M. J. Kim, Prof. S. A. Benner
Foundation for Applied Molecular Evolution
Firebird Biomolecular Sciences LLC
13709 Progress Boulevard, Alachua, FL 32615 (USA)
E-mail: sbenner@ffame.org
[+] These authors contributed equally

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/anie.201504731.

genetic alphabets can evolve in a DNA matrix under laboratory-created selective pressure to give molecules that bind to breast cancer[14] and liver cancer cells;[15] this success suggested a full search of the expanded "sequence space" created by the additional genetic "letters". Further, within DNA, as many as six AEGIS pairs can be placed in consecutive positions in a duplex without altering the overall structure of the DNA double helix.[16]

In contrast with DNA, RNA has seldom been examined as a matrix for an expanded genetic alphabet. Some early work focused on the **S**:**B** and **K**:**X** pairs (Figure 1), but was not supported by structural biology.[17–19] Only recently was this work extended to the enzymology of the pair between 6-amino-5-nitro-3-(1′-β-D-ribofuranosyl)-2(1*H*)-pyridone and 2-amino-8-(1′-β-D-ribo-furanosyl)-imidazo-[1,2-a]-1,3,5-triazin-4-(8*H*)-one, called **Z** and **P** (Figure 1, right box). Recent results showed that T7 RNA polymerase can generate RNA molecules containing **Z** and **P** AEGIS ribonucleosides from encoding GACT**ZP** DNA molecules, and reverse transcriptase can synthesize **Z** and **P** AEGIS DNA from complementary GACU**ZP** AEGIS RNA.[20]

For RNA, structural studies are (if anything) more important than with DNA. Because of its extra 2′-hydroxy group, RNA has many more conformational opportunities than DNA.[21] These are exploited in RNA molecules that perform functions beyond simple genetic encoding, including RNA receptors, RNA ligands, and RNA catalysts, including the ribosome. Crystallography is needed to learn whether a **Z**:**P** pair has a canonical Watson–Crick geometry in an RNA context, and does not disrupt the folding of a functional RNA molecule nor disturb mechanisms by which a functional RNA molecule act.

To take the first step into the structural biology of AEGIS RNA, we turned to one of those functional RNA molecules: a 67-nt xpt-pbuX riboswitch aptamer domain from *Bacillus subtilis* (Protein data bank code (PDB ID): 4FE5).[22] Riboswitches are highly structured *cis*-acting elements located in the 5′-untranslated regions of many messenger RNAs. There, they bind to small-molecule metabolites and, once bound, regulate the expression of the protein encoded by the message. In particular, the xpt-pbuX riboswitch binds to guanine and its structural analogue, hypoxanthine, with approximately 5 and 50 nM dissociation constants, respectively. The crystal structure of the natural riboswitch complexed to hypoxanthine was solved a decade ago.[23] That structure contains a P1 stem, two hairpins, and four nucleotides in the junction that contact the ligand (Figure 2).

Analyzing this structure, we selected a single C:G pair in the P1 stem for replacement by a **Z**:**P** pair. RNA fragments containing **Z** and **P** ribonucleotides were prepared by solid-phase synthesis using commercially available phosphoramidites (Firebird Biomolecular Sciences LLC, www.firebirdbio.com). These were then splint-ligated to the remaining RNA fragment using T4 RNA ligase 2 (Figure S1 in the Supporting Information). The product was purified by 8% PAGE. In parallel, a plasmid encoding the native riboswitch lacking the **Z**:**P** substitution was constructed to generate a wild type (wt) RNA transcript.

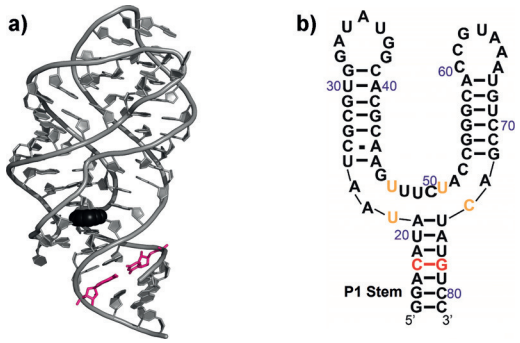


Figure 2. The guanine riboswitch in a) a three dimensional representation (PDB ID: 4FE5) and b) in its secondary structure representation. Nucleotides in yellow contact a guanine or hypoxanthine ligand. In this case, the C:G pair (red) joining positions 18 and 78 in the P1 stem was replaced by a **Z**:**P** pair.

The enzymatic ligation of the **Z**:**P** riboswitch from synthetic oligonucleotides produced a product that differed from the product produced by T7 RNA polymerase transcription; it has a 5′-triphosphate, while that from T4 RNA ligase 2 had a 5′-phosphate only. This difference notwithstanding, initial studies showed that both the 5′-monophosphorylated native riboswitch and its analogue containing the **Z**:**P** pair crystallized well (see Experimental Section in the Supporting Information). Typical growth times were four days at room temperature.

Crystals of both the native and AEGIS riboswitches were then analyzed by X-ray diffraction with the synchrotron source at Argonne National Laboratory. Both crystals had the same space groups and unit cell dimensions. The structures of the native riboswitch and the riboswitch containing the **Z**:**P** pair were then solved at 3.05 Å and 3.22 Å resolution, respectively. Table 1 shows the data collection and refinement statistics, after building and refinement of both structures using COOT24 and PHENIX25 crystallography software packages.[24,25]

Our resolution with the wild-type (wt) riboswitch was lower than that reported for the native riboswitch.[23] Never-

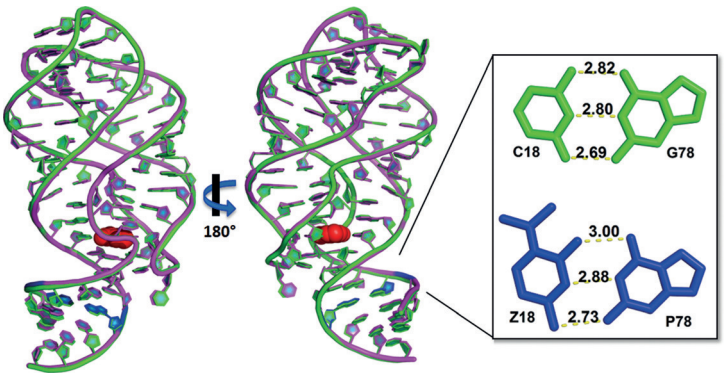
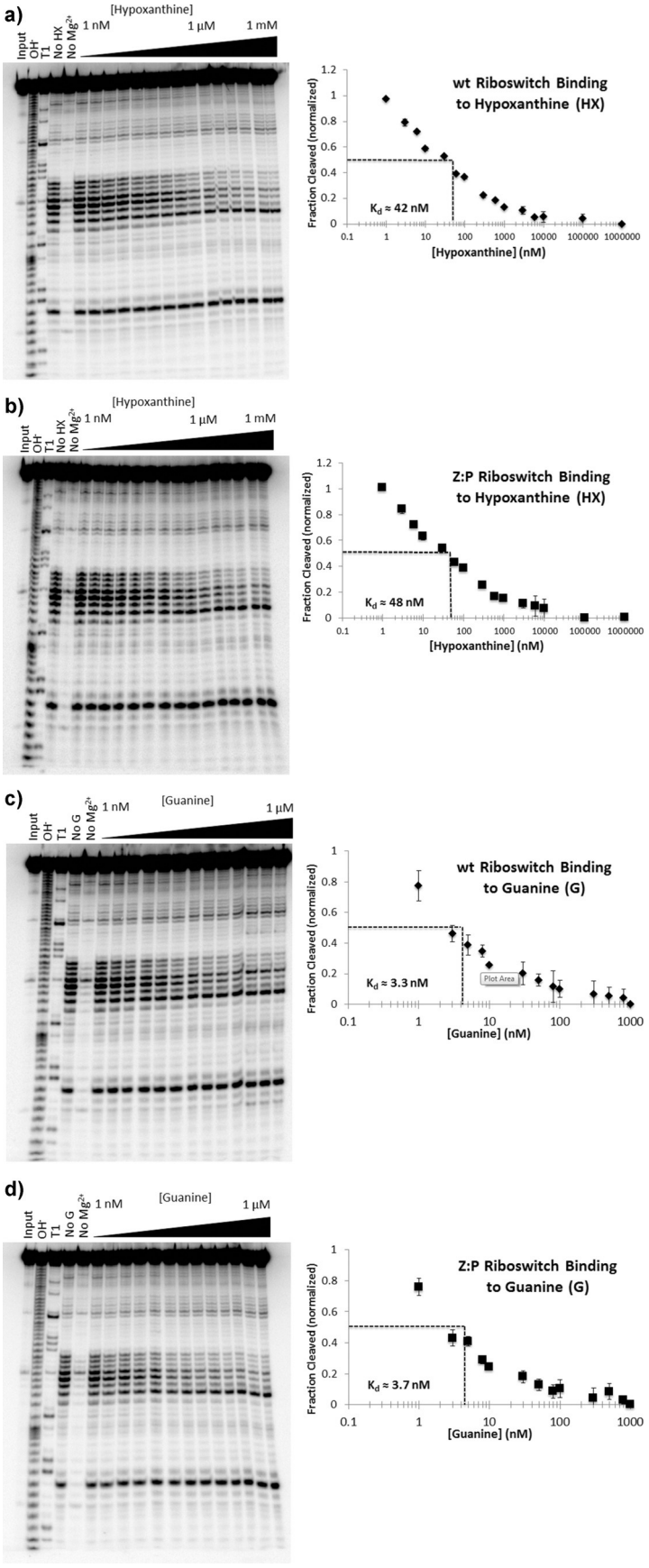


Figure 3. Superimposed structures of the 5′-monophosphate wt (green) and **Z**:**P** (magenta) guanine riboswitch bound to hypoxanthine (red). The **Z**:**P** base pair in the P1 stem is highlighted in blue. Close-up views of corresponding C:G and **Z**:**P** base pairs from the structure models are also presented. A difference map is found in the Supporting Information, Figure S3.



theless, our structure matched well the previously determined structure (RMSD = 0.803, all atoms aligned) (Figure S2). Figure 3 shows the overlapped model for the wt and the **Z**:**P** riboswitches in two orientations. In both cases, the hypoxanthine ligand was present. In the AEGIS riboswitch, electron density was clearly seen at the position expected for the nitro group on **Z** (Figure S3).

Overall, the wild type riboswitch bound to hypoxanthine superimposed rather well on its analogue containing the **Z**:**P** pair (Figure 3). Only in the P1 stem, which holds the **Z**:**P** substitution, did the structures deviate noticeably (Figure S4). Even here, the structural divergence was modest. The inferred inter-nucleobase distances between **Z** and **P** were similar to those seen in various C:G pairs, the closest structural analogues of the **Z**:**P** pair (Figure 3). The interactions between the riboswitch and the bound hypoxanthine were also quite superimposable (Figure S5).

These results show the compatibility of the **Z**:**P** pair with the Watson–Crick and non-Watson–Crick geometries that allow this particular RNA molecule to perform its non-coding function. We therefore sought to confirm this structural conclusion by studies in solution by applying "in-line probing".[21] In-line probing analyzes patterns of fragmentation created through chemical degradation of the RNA segments that are not in A-form helices. Since the conformation of the RNA changes upon binding of the ligand, an apparent dissociation constant can be estimated by analyzing the changing fragmentation patterns while varying the concentrations of guanine or hypoxanthine ligands over a 1 nM to 1 mM range. Using the Origin software package, this analysis generated a dissociation constant of 42 nM for hypoxanthine (Figure 4) and 3.3 nM for guanine for the wild type riboswitch (Figure S6 and S8), and 48 nM and 3.7 nM for the **Z**:**P** riboswitch (Figure S7 and S9). For the native riboswitch, values of 50 and 5 nM are reported in the literature.[26]

These are the first crystallographic analyses and solution biophysical measurements for any RNA molecule that contains an added orthogonal nucleobase pair. From this combination of structural and biophysical measurements, we might expect that the **Z**:**P** pair will fit well within the constraints of an RNA structure, consistent with its enzymological behavior.

Accession codes**:** The atom coordinates and structural factors have been deposited in the Protein Data Bank with accession code 5C7U for the wild type native guanine riboswitch and 5C7W for the Z:P guanine riboswitch.

Figure 4. In-line probing data for native and **Z**:**P** riboswitches as a function of ligand concentrations. Gels show gradual protection of regions within RNA at increasing concentrations of ligand. Plots generated by the Origin software give apparent K_d values for ligand binding. a) Wild type riboswitch and hypoxanthine ligand; $K_d \approx 42$ nM. b) **Z**:**P** riboswitch and hypoxanthine ligand, $K_d \approx 48$ nM. c) Wild type riboswitch and guanine ligand; $K_d \approx 3.3$ nM. d) Z:P riboswitch and guanine ligand; $K_d \approx 3.7$ nM.

Table 1: Data collection and refinement statistics.

RNA	wt riboswitch[a]	**Z**:**P** riboswitch[a]
Data collection:		
Wavelength [Å]	0.97900	0.97918
Space group	$P6_522$	$P6_522$
Resolution [Å]	46.36–3.05 (3.16–3.05)	44.59–3.22 (3.34–3.22)
Cell dimensions		
a, *b*, *c* [Å]	52.61, 52.61, 278.17	52.18, 52.18, 274.5
R_{merge} [%]	6.5 (50.2)	8.3 (53.7)
$I/\sigma I$	33 (5.1)	24.9 (4.1)
Completeness [%]	100 (100)	99.8 (100)
Redundancy	14 (14.82)	8.54 (8.95)
Refinement:		
No. reflections	4923	4107
R_{work}/R_{free}	0.2011/0.2411	0.2277/0.2797
R.M.S deviations		
Bond angles [°]	1.7	1.8
Bond length [Å]	0.008	0.007
B-factors	95	104

[a] Values in the parentheses are for the highest resolution shell.

Acknowledgements

We are indebted to the Defense Threat Reduction Agency (HDTRA1-13-1-0004), NASA (NNX10AT28G), and the Templeton World Charitable Fund for support of this research.

Keywords: crystal structure · expanded DNA · in-line probing · riboswitch · synthetic biology

How to cite: *Angew. Chem. Int. Ed.* **2015**, *54*, 9853–9856
Angew. Chem. **2015**, *127*, 9991–9994

[1] C. Switzer, S. E. Moroney, S. A. Benner, *J. Am. Chem. Soc.* **1989**, *111*, 8322–8323.
[2] J. A. Piccirilli, T. Krauch, S. E. Moroney, S. A. Benner, *Nature* **1990**, *343*, 33–37.
[3] S. A. Benner, Z. Yang, F. Chen, *C. R. Chim.* **2011**, *14*, 372–387.
[4] K. M. Guckian, J. C. Morales, E. T. Kool, *J. Org. Chem.* **1998**, *63*, 9652–9656.
[5] K. M. Guckian, T. R. Krugh, E. T. Kool, *J. Am. Chem. Soc.* **2000**, *122*, 6841–6847.
[6] J. C. Delaney, P. T. Henderson, S. A. Helquist, J. C. Morales, J. M. Essigmann, E. T. Kool, *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 4469–4473.
[7] M. Kimoto, R. Yamashige, K. Matsunaga, S. Yokoyama, I. Hirao, *Nat. Biotechnol.* **2013**, *31*, 453–457.
[8] M. Kimoto, R. S. Cox 3rd. , I. Hirao, *Expert Rev. Mol. Diagn.* **2011**, *11*, 321–331.
[9] R. Yamashige, M. Kimoto, Y. Takezawa, A. Sato, T. Mitsui, S. Yokoyama, I. Hirao, *Nucleic Acids Res.* **2012**, *40*, 2793–2806.
[10] K. Betz, D. A. Malyshev, T. Lavergne, W. Welte, K. Diederichs, T. J. Dwyer, P. Ordoukhanian, F. E. Romesberg, A. Marx, *Nat. Chem. Biol.* **2012**, *8*, 612–614.
[11] K. Betz, D. A. Malyshev, T. Lavergne, W. Welte, K. Diederichs, F. E. Romesberg, A. Marx, *J. Am. Chem. Soc.* **2013**, *135*, 18637–18643.
[12] D. A. Malyshev, K. Dhami, H. T. Quach, T. Lavergne, P. Ordoukhanian, A. Torkamani, F. E. Romesberg, *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 12005–12010.
[13] D. A. Malyshev, K. Dhami, T. Lavergne, T. Chen, N. Dai, J. M. Foster, I. R. Correa, Jr., F. E. Romesberg, *Nature* **2014**, *509*, 385–388.
[14] K. Sefah, Z. Yang, K. M. Bradley, S. Hoshika, E. Jimenez, G. Zhu, S. Shanker, F. Yu, W. Tan, W. , S. A. Benner, *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 1449–1456.
[15] L. Zhang, Z. Yang, K. Sefah, K. M. Bradley, S. Hoshika, M.-J. Kim, H.-J. Kim, G. Zhu, E. Jimenez, S. Cansiz, I.-T. Teng, C. Champanhac, C. McLendon, C. Liu, W. Zhang, D. L. Gerloff, Z. Huang, W. H. Tan, S. A. Benner, *J. Am. Chem. Soc.* **2015**, *137*, 6734–6737.
[16] M. M. Georgiadis, I. Singh, W. F. Kellett, S. Hoshika, S. A. Benner, N. G. J. Richards, *J. Am. Chem. Soc.* **2015**, 137, 6947–6955.
[17] J. D. Bain, C. Switzer, A. R. Chamberlin, S. A. Benner, *Nature* **1992**, *356*, 537–539.
[18] J. A. Piccirilli, S. E. Moroney, S. A. Benner, *Biochemistry* **1991**, *30*, 10350–10356.
[19] J. J. Voegel, S. A. Benner, *Helv. Chim. Acta* **1996**, *79*, 1881–1898.
[20] H.-J. Kim, N. A. Leal, S. Hoshika, S. A. Benner, *J. Org. Chem.* **2014**, *79*, 3194–3199.
[21] K. T. Schroeder, S. A. McPhee, J. Ouellet, D. M. J. Lilley, *RNA* **2010**, *16*, 1463–1468.
[22] M. Mandal, B. Boese, J. E. Barrick, W. C. Winkler, R. R. Breaker, *Cell* **2003**, *113*, 577–586.
[23] R. T. Batey, S. D. Gilbert, R. K. Montange, *Nature* **1996**, *379*, 411–415.
[24] P. Emsley, K. Cowtan, *Acta Crystallogr. Sect. D* **1996**, *60*, 2126–2132.
[25] P. D. Adams, P. V. Afonine, G. Bunkóczi, V. B. Chen, I. W. Davis, N. Echols, J. J. Headd, L.-W. Hung, G. J. Kapral, R. W. Grosse-Kunstleve, A. J. McCoy, N. W. Moriarty, R. Oeffner, R. J. Read, D. C. Richardson, J. S. Richardson, T. C. Terwilliger, P. H. Zwart, et al., *Acta Crystallogr. Sect. D* **1996**, *66*, 213–221.
[26] A. Nahvi, N. Sudarsan, M. S. Ebert, X. Zou, K. L. Brown, R. R. Breaker, *Chem. Biol.* **1996**, *3*, 1043–1049.

Received: May 25, 2015
Published online: July 16, 2015

