

Evolution of Functional Six-Nucleotide DNA

Liqin Zhang,^{†,○} Zunyi Yang,^{‡,○} Kwame Sefah,[†] Kevin M. Bradley,[‡] Shuichi Hoshika,[‡] Myong-Jung Kim,[‡] Hyo-Joong Kim,[§] Guizhi Zhu,^{†,||} Elizabeth Jiménez,[†] Sena Cansiz,[†] I-Ting Teng,[†] Carole Champanhac,[†] Christopher McLendon,[‡] Chen Liu,[⊥] Wen Zhang,^{#,∇} Dietlind L. Gerloff,[‡] Zhen Huang,^{*,#,∇} Weihong Tan,^{*,†,||} and Steven A. Benner^{*,‡,§}

[†]Department of Chemistry, Department of Physiology and Functional Genomics, UF Health Cancer Center, UF Genetics Institute, University of Florida, Gainesville, Florida 32611, United States

[‡]Foundation for Applied Molecular Evolution, Gainesville, Florida 32601, United States

[§]Firebird Biomolecular Sciences LLC, Alachua, Florida 32615, United States

^{||}Molecular Science and Biomedicine Laboratory, State Key Laboratory of Chemo/Biosensing and Chemometrics, College of Chemistry and Chemical Engineering, College of Biology, Collaborative Innovation Center for Chemistry and Molecular Medicine, Hunan University, Changsha 410082, China

[⊥]Department of Pathology, Immunology, and Laboratory Medicine, University of Florida College of Medicine, Gainesville, Florida 32610, United States

[#]Department of Chemistry, Georgia State University, Atlanta, Georgia 30303, United States

[∇]SeNA Research, Inc., Atlanta, Georgia 30303, United States

Supporting Information

ABSTRACT: Axiomatically, the density of information stored in DNA, with just four nucleotides (GACT), is higher than in a binary code, but less than it might be if synthetic biologists succeed in adding independently replicating nucleotides to genetic systems. Such addition could also add functional groups not found in natural DNA, but useful for molecular performance. Here, we consider two new nucleotides (Z and P, 6-amino-5-nitro-3-(1'-β-D-2'-deoxyribo-furanosyl)-2(1H)-pyridone and 2-amino-8-(1'-β-D-2'-deoxyribofuranosyl)-imidazo[1,2-a]-1,3,5-triazin-4(8H)-one). These are designed to pair via complete Watson–Crick geometry. These were added to a library of oligonucleotides used in a laboratory *in vitro* evolution (LIVE) experiment; the GACTZP library was challenged to deliver molecules that bind selectively to liver cancer cells, but not to untransformed liver cells. Unlike in classical *in vitro* selection, low levels of mutation allow this system to evolve to create binding molecules not necessarily present in the original library. Over a dozen binding species were recovered. The best had Z and/or P in their sequences. Several had multiple, nearby, and adjacent Zs and Ps. Only the weaker binders contained no Z or P at all. This suggests that this system explored much of the sequence space available to this genetic system and that GACTZP libraries are richer reservoirs of functionality than standard libraries.

The possibility of increasing the number of replicable nucleotides in DNA and RNA (collectively xNA) above the standard four found in natural terran xNA was noted a quarter-century ago.^{1,2} However, only recently has the broader scientific community come to recognize that expanded genetic

“alphabets” might also expand the functional potential of nucleic acids. Key contributions include the observation by Hiraó and his co-workers that adding a fifth nucleotide to a DNA aptamer increased its affinity for its target,³ the use of expanded genetic alphabets to increase the amino acid “lexicon” of proteins in ribosome-based translation,⁴ the development of full *in vitro* selection protocols that exploit 6-nucleotide DNA libraries,⁵ and the improved performance of DNA that does not add nucleotides to the DNA “alphabet”, but rather appends functional groups to the four standard nucleobases.^{6,7} More recently, Romesberg created a strain of *E. coli* that maintains in a plasmid (for 15 h) one exemplar of a nonstandard nucleobase pair *not* joined by hydrogen bonds.⁸

Axiomatically, adding two nucleotides to the four found in standard xNA increases the “sequence space” of the system. However, if the added pairs deviate too greatly from canonical Watson–Crick geometry, standard polymerases will be unable to explore that space, especially that part where nonstandard nucleotides are nearby or adjacent in the sequence. For example, nonstandard nucleotides from the Hiraó,^{3,9,10} Romesberg,^{8,11–13} and Kool^{14–16} groups (Figure S1) all lack internucleobase hydrogen bonding, designed to pair edge-on by steric complementarity alone. In addition to deviating substantially from the Watson–Crick pairing “concept”, some do not pair as designed. For example, the Romesberg nucleobases intercalate^{11,12} rather than lying coplanar in a DNA double helix; an edge-on geometry is enforced only by interaction with a polymerase.¹³ This creates challenges in building DNA duplexes with adjacent and nearby nonstandard pairs.

Received: March 3, 2015

Published: May 12, 2015

An artificially expanded genetic information system (AEGIS) can, however, be designed to retain interstrand hydrogen bonding as well as steric complementarity within a complete Watson–Crick pairing geometry.¹⁷ Here, hydrogen bond donor and acceptor groups are rearranged within that geometry to create up to 12 independently replicating nucleotides forming six orthogonal base pairs.¹⁷ Various AEGIS pairs support “six letter” PCR amplification,¹⁸ transcription into RNA, reverse transcription back to DNA,¹⁹ sequencing,²⁰ and other processes known in natural molecular biology.

As newly reported in a separate manuscript,²¹ a new pair (Z and P, Figure 1) joined by an orthogonal hydrogen bonding

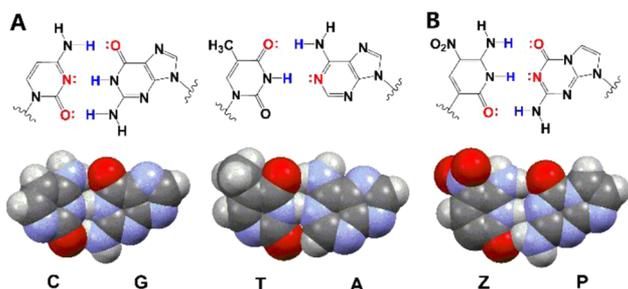


Figure 1. Crystal structures of C:G, T:A, and Z:P pairs (top as structures, bottom space filling) showing their similarity (PDB ID: 4RHD). (A) C:G and T:A pairs. (B) Z:P pair retaining hydrogen bonding. The structure for this Z:P pair was obtained by cocrystallization of 5'-G^{5-MeSe}dUGT-Z-ACAC-3' and 5'-G^{5-MeSe}dUGT-P-ACAC-3', where selenium derivatization was used to speed crystallization.^{22,23} For structures of unmodified DNA containing Z:P pairs, see ref 21.

pattern was found to adopt a standard Watson–Crick geometry. Indeed, the geometry was sufficiently “natural” that *six consecutive Z:P pairs* (of 16) were compatible with the double helix. This result was confirmed in a separate structure, reported here (Figures 1 and S1).

This encouraged us to ask whether the considerably larger sequence space created by GACTZP DNA, which also carries the nitro group, could be explored by polymerases in laboratory

in vitro evolution (LIVE) experiments to create useful DNA molecules. LIVE is analogous to *in vitro* selection or SELEX.^{23–26} In these, however, the high fidelity of polymerases does not allow substantial sequence evolution; sequences that emerge must, in general, already be present in the original library. However, in LIVE experiments, Z:P pairs can be gained or lost, allowing the system to explore sequences outside of those already present in the initial library.

To answer this question, we adapted the cell-LIVE procedure of Sefah et al.⁵ to create GACTZP aptamers that bind to a line of HepG2 liver cancer cells (Figure 2), adding a negative selection. A six-nucleotide single stranded DNA library (GACTZP DNA library) was synthesized with a randomized region (25 nucleotides) flanked by two primer binding sites. This was incubated with target liver cancer cells. Unbound species were then removed by washing; species having affinity for the cancer cells were collected. “Survivors” after positive selection were then negatively selected, removing those that bound to untransformed liver cells (Hu1545V), the counter-cells. These were then amplified by GACTZP PCR using a fluorescein-labeled forward primer and a biotinylated reverse primer. Enriched fluorescein-conjugated ssDNA libraries from the PCR was used for the next round of selection.

The binding affinities of survivors in rounds 8 through 13 were monitored by flow cytometer. Binding was observed in the bulk pools after 12 rounds of affirmative selection (Figure S2) into which had been embedded (in rounds 3–6) four rounds of negative selection. The entire process included approximately 200 cycles of PCR.

After round 13, the LIVE was stopped, and the selected pool was subjected to deep sequencing (see Tables S3–S7 for the procedure, experimental details, and additional data). Then, 17 motifs (contributing from 0.14 to 26% of the total surviving population, Figure 3A) were resynthesized with a 5'-biotin tag. Their affinities toward HepG2 cells were then measured by measuring fluorescence per cell as a function of increasing concentration of aptamers (Figure S3), with the fluorescence was created by a phycoerythrin–streptavidin tag. Data are collected in Figure 3A.

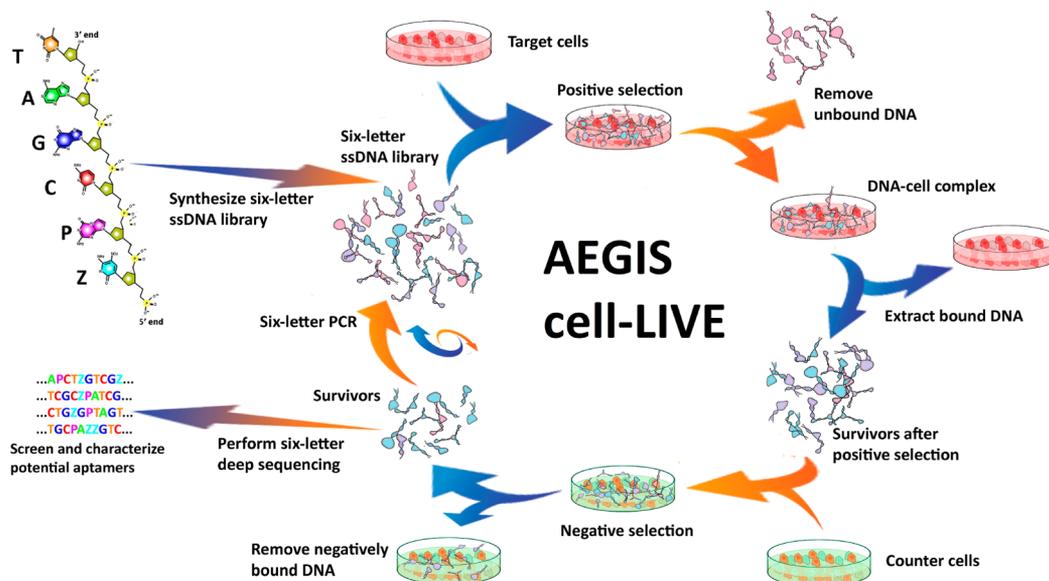


Figure 2. Schematic of the AEGIS cell-LIVE with both positive and counter selections.

A)

Name	Sequences	K_d (nM)	Percentage in pool
LZH1	~CCAATAATCTPGCZGPGGATCGG~	14±3	0.40%
LZH2	~GGAAGTGACGGTAGCPTTTTGGAGG~	24±3	0.28%
LZH3	~CGACCZGACTTTTAGCPTCGAATAG~	24±5	7.84%
LZH4	~GGATAAGTCTPACZGPGGTATCATG~	36±5	0.50%
LZH5	~GCTACPTGGGCCCTGGTPTCTGTGC~	41±7	0.66%
LZH6	~TATCAGCCGATTAACTCPZATGG~	47±5	2.15%
LZH7	~CAATAATCTPGCZGCGGTAITGGG~	55±8	23.04%
LZH8	~TATTAGTACGGCTAAACCCPCATGG~	96±10	0.65%
LZH9	~TATCPGTTGCCCTAAAGGCTATGG~	214±26	25.67%
LZH10	~CGCCACGGAGAGTCTCTGCGGCC~	234±53	1.35%
LZH11	~CGCCCGCZGAGCAGGPPCCCCCGCC~	247±25	0.27%
LZH12	~CGGCTTGACAGACPGCATZGATCAG~	298±37	0.14%
LZH13	~GTGCGGCCACCATACCTCTCTGGCC~	326±53	0.91%
LZH14	~CCAACCTGCGACCCACAACCTATG~	346±50	1.27%
LZH15	~TTGGCATGCCACCACCTACCAGGC~	727±84	1.25%
LZH16	~TCCCTACATGCGAGTACCACCTCTG~	>1000	0.74%
LZH17	~CCACCTAAGCTCTGTTTCCCGTGG~	>1000	0.54%

B)

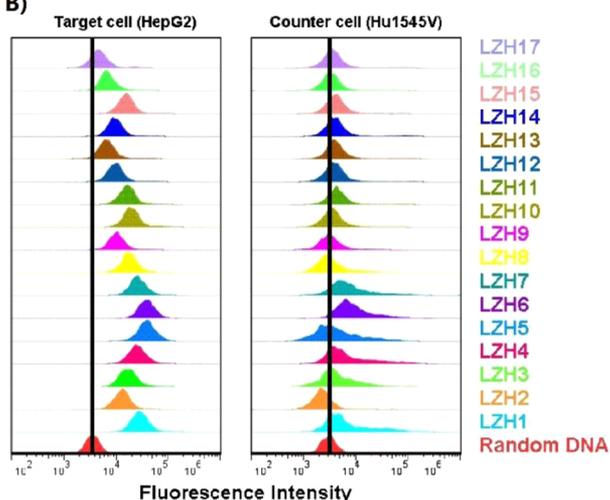


Figure 3. DNA aptamers recovered from cell-LIVE. (A) Sequences (only showing randomized region), dissociation constants (K_d), and percent in pool of binders (the AEGIS Z and P are shown in red). Sequences are arranged in order of increasing dissociation constant. (B) Binding and specificity of selected DNA aptamers. Aptamers are arranged from the least tightly binding (top) to the most tightly binding (bottom). (left) Binding to transformed liver cells, the “positive” in the selection. (right) Binding to untransformed liver cells, the counter-selection cells. The red distributions at the bottom of each panel is the signal generated from DNA that has the same length as the aptamers but random sequence.

The GACTZP binders also have good specificity (Figure 3B), presumably a result of the negative selection. This was shown by incubating both transformed and untransformed cells (30 min, 4 °C) with biotinylated aptamers at high concentrations (250 nM) that, for aptamers LZH1 through LZH8, are well above their dissociation constants; for these, the extent of fluorescence (the x axis) indicates the effective number of sites on each cell. Then, after the biotin captured fluorescent phycoerythrin–streptavidin conjugate, the total labeling of each cell type was determined by flow cytometry. For the best eight aptamers (LZH1 through LZH8), the 250 nM concentration used was considerably greater than their respective dissociation constants. Therefore, the total fluorescent labeling of each cell offers an estimate of the number of binding sites per cell.

Dissociation constants were also sought for aptamers against untransformed liver cells. In most cases, no binding was detected. In cases where binding was detected, the affinity was (with experimental error) the same as with cancer cells. This implies the aptamer target on the cancer cells is also present on the untransformed cells, but just in much smaller amounts. (Figures 3B and S3, Tables S8 and S9). Fifteen other types of cells were also used to confirm the specificity (Table S9).

To show that the AEGIS nucleotides were essential for binding in the two most abundant aptamers containing both Z and P (LZH3 and LZH7), their analogues were synthesized with P replaced by G or A, and Z replaced by C or T. In all cases, observed binding was diminished in the analogue lacking the Z and/or P relative to the parent aptamer. Binding was not restored when the Z and P were replaced by the complementary C and G (Figure 4 and Table S10).

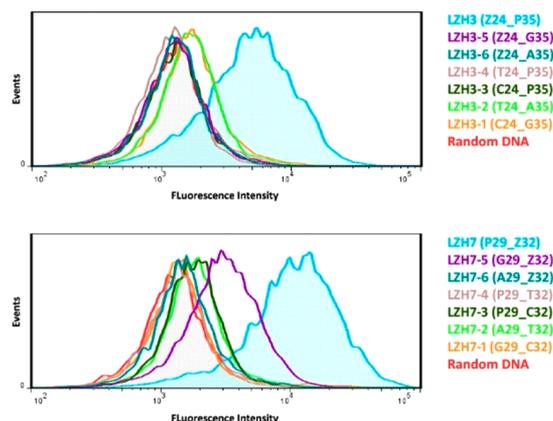


Figure 4. Binding of analogues of aptamer LZH3 and LZH7 with Z and P replaced by standard nucleotides. The indicated aptamers and their analogues (50 nM) were incubated with target cell (4 °C for 30 min) and then analyzed using flow cytometry as described for Figure 3.

These data provide direct evidence, the first for any artificially expanded genetic information system, that a LIVE experiment can explore substantial fractions of sequence space in a six letter genetic system.

It is significant that LIVE differs from classical *in vitro* selection, which generally finds only species pre-existing in a library. In the GACTZP system, slow gain or loss of Z and P is possible, with loss rate slightly larger than the rate of gain.¹⁸ Thus, the system at evolutionary equilibrium will have (and, in these experiments, did have) fewer than the original amounts of Z and P. The generation of binders with more Zs and Ps than generated by Sefah et al.⁵ may reflect the improvement in polymerase fidelity that retains Z:P pairs.

Interestingly, this also suggests that while they are modestly disfavored during PCR, Z and P are favored by the selection itself, with selection to retain Z/P balancing the tendency to lose Z/P during PCR. This also suggests that GACTZP libraries may be richer reservoirs of selective binders than GACT libraries. If this were not so, and if Z and/or P does *not* contribute to the overall “fitness potential” of the reservoir, the LIVE experiment could have delivered binders with competing affinity lacking Z and/or P through Z/P loss. These would come either from (quite scarce) original library members that lacked the Z and P (the initial library contained on average 3 of each) or through the net loss of Z and P during the PCR amplification.

Indeed, binders lacking both Z and P *do* emerge in this selection. However, their affinities were generally weaker, often much weaker, than binders with Z and/or P (Figure 3B). This may, of course, reflect the scarcity of GACT sequences in the original library; the best specific GACT binders may have been undersampled. Alternatively, the evolution that removed Z and/or P may not have delivered the best GACT binders. Subject to these caveats, these experimental outcomes suggest that GACTZP libraries are richer reservoirs of binding function than standard GACT libraries.

We can only speculate as to why GACTZP libraries might be richer reservoirs of binding molecules. For example, Z has a nitro moiety that may be a “universal weak binder” (note the affinity of many proteins to nitrocellulose). Alternatively, the added nucleotides, by increasing the information density of the sequences, may have removed folding ambiguity, an ambiguity that has been shown in other systems to diminish the performance of functional xNA species.²⁷

This notwithstanding, these LIVE experiments have shown that this particular six-letter artificially expanded genetic information system can explore regions of sequence space that require pairing to be sufficiently “Watson–Crick-like” for polymerases to accept. Thus, they lay the ground for the development of a broad laboratory evolution program in AEGIS-LIVE.

■ ASSOCIATED CONTENT

Supporting Information

Experimental procedures and supplementary data. The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/jacs.5b02251.

■ AUTHOR INFORMATION

Corresponding Authors

*huang@gsu.edu

*tan@chem.ufl.edu

*sbenner@ffame.org

Author Contributions

○These authors contributed equally to this work.

Notes

The authors declare the following competing financial interest(s). S.A.B. owns intellectual property in this area.

■ ACKNOWLEDGMENTS

We are indebted to the Defense Threat Reduction Agency (HDTRA1-13-1-0004) and the National Aeronautics and Space Administration (NNX10AT28G) for long-standing support of this and other basic research in the S.A.B. laboratory in the area of nucleic acid chemistry. Part of this work was supported by the Templeton World Charitable Fund. The W.T. and S.A.B. laboratories are indebted to the National Institutes of Health (R21CA122648, GM079359, CA133086, and R01GM111386), the National Key Scientific Program of China (2011CB911000), NSFC grants (NSFC 21221003 and NSFC 21327009), and the China National Instrumentation Program 2011YQ03012412 for support of this work. The Z.H. laboratory is indebted to the National Science Foundation Innovation Corps Program (1340153) and National Institutes of Health (R01GM095881) for support of this work.

■ REFERENCES

- (1) Switzer, C.; Moroney, S. E.; Benner, S. A. *J. Am. Chem. Soc.* **1989**, *111*, 8322.
- (2) Piccirilli, J. A.; Krauch, T.; Moroney, S. E.; Benner, S. A. *Nature* **1990**, *343*, 33.
- (3) Kimoto, M.; Yamashige, R.; Matsunaga, K.; Yokoyama, S.; Hirao, I. *Nat. Biotechnol.* **2013**, *31*, 453.
- (4) Bain, J. D.; Switzer, C.; Chamberlin, A. R.; Benner, S. A. *Nature* **1992**, *356*, 537.
- (5) Sefah, K.; Yang, Z.; Bradley, K. M.; Hoshika, S.; Jimenez, E.; Zhang, L.; Zhu, G.; Shanker, S.; Yu, F.; Turek, D.; Tan, W.; Benner, S. A. *Proc. Natl. Acad. Sci. U.S.A.* **2014**, *111*, 1449.
- (6) Hollenstein, M.; Hipolito, C. J.; Lam, C. H.; Perrin, D. M. *Nucleic Acids Res.* **2009**, *37*, 1638.
- (7) Gold, L.; Ayers, D.; Bertino, J.; Bock, C.; Bock, A.; Brody, E. N.; Carter, J.; Dalby, A. B.; Eaton, B. E.; Fitzwater, T.; Flather, D.; Forbes, A.; Foreman, T.; Fowler, C.; Gawande, B.; Goss, M.; Gunn, M.; Gupta, S.; Halladay, D.; Heil, J.; Heilig, J.; Hicke, B.; Husar, G.; Janjic, J.; Jarvis, T.; Jennings, S.; Katilius, E.; Keeney, T. R.; Kim, N.; Koch, T. H.; Kraemer, S.; Kroiss, L.; Le, N.; Levine, D.; Lindsey, W.; Lollo, B.; Mayfield, W.; Mehan, M.; Mehler, R.; Nelson, S. K.; Nelson, M.; Nieuwlandt, D.; Nikrad, M.; Ochsner, U.; Ostroff, R. M.; Otis, M.; Parker, T.; Pietrasiewicz, S.; Resnicow, D. I.; Rohloff, J.; Sanders, G.; Sattin, S.; Schneider, D.; Singer, B.; Stanton, M.; Sterkel, A.; Stewart, A.; Stratford, S.; Vaught, J. D.; Vrkljan, M.; Walker, J. J.; Watrobka, M.; Waugh, S.; Weiss, A.; Wilcox, S. K.; Wolfson, A.; Wolk, S. K.; Zhang, C.; Zichi, D. *PLoS One* **2010**, *5*, e15004.
- (8) Malyshev, D. A.; Dhami, K.; Lavergne, T.; Chen, T.; Dai, N.; Foster, J. M.; Correa, I. R., Jr.; Romesberg, F. E. *Nature* **2014**, *509*, 385.
- (9) Kimoto, M.; Cox, R. S., 3rd; Hirao, I. *Expert Rev. Mol. Diagn.* **2011**, *11*, 321.
- (10) Yamashige, R.; Kimoto, M.; Takezawa, Y.; Sato, A.; Mitsui, T.; Yokoyama, S.; Hirao, I. *Nucleic Acids Res.* **2012**, *40*, 2793.
- (11) Betz, K.; Malyshev, D. A.; Lavergne, T.; Welte, W.; Diederichs, K.; Dwyer, T. J.; Ordoukhanian, P.; Romesberg, F. E.; Marx, A. *Nat. Chem. Biol.* **2012**, *8*, 612.
- (12) Betz, K.; Malyshev, D. A.; Lavergne, T.; Welte, W.; Diederichs, K.; Romesberg, F. E.; Marx, A. *J. Am. Chem. Soc.* **2013**, *135*, 18637.
- (13) Malyshev, D. A.; Dhami, K.; Quach, H. T.; Lavergne, T.; Ordoukhanian, P.; Torkamani, A.; Romesberg, F. E. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109*, 12005.
- (14) Guckian, K. M.; Morales, J. C.; Kool, E. T. *J. Org. Chem.* **1998**, *63*, 9652.
- (15) Delaney, J. C.; Henderson, P. T.; Helquist, S. A.; Morales, J. C.; Essigmann, J. M.; Kool, E. T. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 4469.
- (16) Guckian, K. M.; Krugh, T. R.; Kool, E. T. *J. Am. Chem. Soc.* **2000**, *122*, 6841.
- (17) Benner, S. A.; Yang, Z. Y.; Chen, F. C. R. C. R. *Chim.* **2011**, *14*, 372.
- (18) Yang, Z.; Chen, F.; Chamberlin, S. G.; Benner, S. A. *Angew. Chem., Int. Ed.* **2010**, *49*, 177.
- (19) Leal, N. A.; Kim, H. J.; Hoshika, S.; Kim, M. J.; Carrigan, M. A.; Benner, S. A. *ACS Synth. Biol.* **2015**, No. 4, 407.
- (20) Yang, Z.; Chen, F.; Alvarado, J. B.; Benner, S. A. *J. Am. Chem. Soc.* **2011**, *133*, 15105.
- (21) Georgiadis, M. M.; Singh, I.; Kellett, W. F.; Hoshika, S.; Benner, S. A.; Richards, N. G. J. *J. Am. Chem. Soc.* **2015**, DOI: 10.1021/jacs.5b03482.
- (22) Lin, L.; Sheng, J.; Huang, Z. *Chem. Soc. Rev.* **2011**, *40*, 4591.
- (23) Hassan, A. E.; Sheng, J.; Jiang, J.; Zhang, W.; Huang, Z. *Org. Lett.* **2009**, *11*, 2503.
- (24) Tuerk, C.; Gold, L. *Science* **1990**, *249*, 505.
- (25) Ellington, A. D.; Szostak, J. W. *Nature* **1990**, *346*, 818.
- (26) Breaker, R. R.; Joyce, G. F. *Chem. Biol.* **1994**, *1*, 223.
- (27) Carrigan, M.; Ricardo, A.; Ang, D. N.; Benner, S. A. *Biochemistry* **2004**, *43*, 11446.