JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

Communication

pubs.acs.org/JACS

Multistep DNA-Templated Synthesis Using a Universal Template

Yizhou Li, Peng Zhao, Mingda Zhang, Xianyuan Zhao, and Xiaoyu Li*

Key Laboratory of Bioorganic Chemistry and Molecular Engineering of the Ministry of Education, Beijing National Laboratory of Molecular Sciences (BNLMS), College of Chemistry and Molecular Engineering, Peking University, 202 Chengfu Road, Beijing, China 100871

Supporting Information

ABSTRACT: We report a DNA-templated synthesis method that allows construction of the entire DNA-encoded library with a single DNA template. Taking advantage of deoxyinosine's indiscriminate base-pairing property, we designed a "universal template" that is capable of directing chemical reactions with multiple reactant DNAs with different sequences. In combination with other design features including photocleavable linkers and direct encoding by the reactant DNA, we demonstrated the capabilities of the universal template in library synthesis, target selection, and hit decoding. Our method can be generally and straightforwardly applied to prepare a variety of chemically diverse DNA-encoded libraries.

Originally in 1992, Brenner and Lerner proposed the concept of using DNA as tags to encode chemical reactions in combinatorial library synthesis.[1] During the past two decades, a variety of strategies of DNA-encoded libraries (DEL) have been developed.[2] Benefiting from the high encoding capacity of DNA molecules, today's DELs can be prepared with an extremely large number of compounds.[3] Selections of DELs against biological targets can be accomplished using highly sensitive PCR amplification and ultra-high-throughput DNA sequencing technology.[3,4] Selections of DELs have already generated many novel ligands against biological targets and have become a new discovery modality for researchers in both academia and pharmaceutical industry.[2e,h,i]

Despite technological variations, all DEL synthesis strategies share the common goal of creating compound collections in which each compound is conjugated to a unique encoding DNA tag. Initially developed by Liu and co-workers, DNA-templated synthesis (DTS) is an important method for DEL synthesis;[2h,5] it has also become a versatile approach in controlling chemical reactions and molecular interactions.[2f,6] In DTS-based library synthesis, each template has a unique sequence containing several "codons" (Figure 1a). An individual codon directs hybridization with a complementary "reagent DNA", enabling the DNA-templated reaction and delivering a building block from the specific reagent DNA encoded for by the template. A DNA-encoded library thus can be constructed with a large pool of templates of many different sequences through multistep DTS. For example, a pool of templates with $m \times n \times l$ different sequences (m, n, and l are the numbers of different codons in each reaction step) generates a library of $m \times n \times l$ compounds by a 3-step DTS

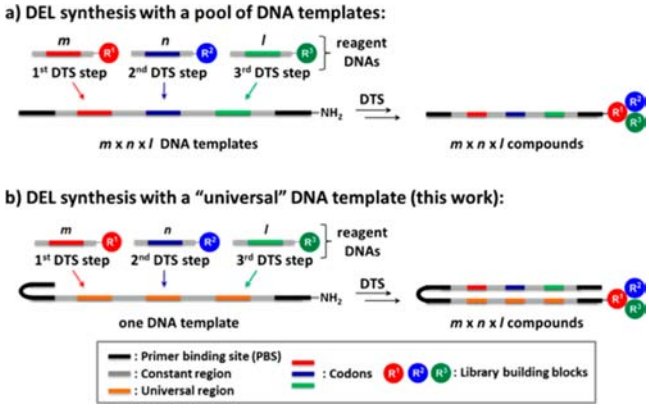

Figure 1. Synthesis of DNA-encoded library (DEL) by (a) regular DTS strategy with a pool of templates of many different sequences ($m \times n \times l$), and by (b) a single "universal" DNA template. A typical 3-step synthesis with 3 sets of "reagent DNAs" is shown. m, n, and l indicate the numbers of reagent DNAs with different codons encoding different building blocks (R^1, R^2, R^3) in each step.

(Figure 1a).[5b,c] However, this "one-template, one-compound" strategy presents certain challenges and limitations, especially for DELs containing large numbers of compounds. First, a template pool corresponding to every single compound has to be prepared as the starting material (e.g., a 1-million compound library requires 1 million unique DNA templates). Even with automated DNA synthesis and combinatorial "split-and-pool" strategy,[7] it is still a tedious and laborious process to synthesize and purify large numbers of DNAs. More importantly, template sequences need to be carefully designed and experimentally validated in order to avoid mismatched reactions and to ensure encoding fidelity. This "codon design" is not a trivial task and often requires sophisticated computation and iterated optimization.[5c,8] Recently, two strategies that do not require a template pool have been reported: the ESAC library by the Neri group[9] and the YoctoReactor system by Vipergen.[10] ESAC is specific for fragment-based libraries, whereas the YoctoReactor is a proximity-based approach that assembles compounds within the tight spaces of multiway DNA junctions. It is still highly desired to develop a more general approach that is compatible with typical DTS but that also does not require a large pool of templates.

Here we report such a method that replaces the complex template pool with just a single DNA template. We designed a

Received: September 26, 2013
Published: November 14, 2013

   dx.doi.org/10.1021/ja409936r | J. Am. Chem. Soc. 2013, 135, 17727−17730

novel "universal template" that can hybridize with not a specific one but multiple reagent DNAs with different sequences (Figure 1b). With this approach, library synthesis is significantly simplified as only one DNA template is necessary. More importantly, sophisticated codon design is no longer necessary as the universal template is designed to hybridize with multiple DNAs.

In the "universal template" (**UT**, Figure 2a), we incorporated three deoxyinosines as "anticodons" ("**III**"). Deoxyinosine can

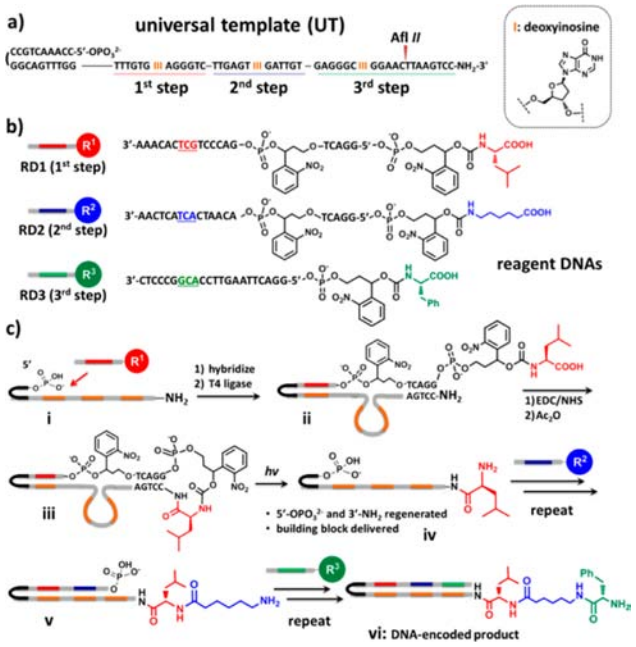

Figure 2. (a) Sequence and structure of the universal template (**UT**). Reagent DNA hybridization sites are underlined. **III** indicates 3 deoxyinosines. (b) Structures and sequences of the reagent DNAs (**RD1**, **RD2**, **RD3**). Building blocks are encoded by 3-base codons (underlined). c) Reaction scheme of a 3-step templated synthesis with the **UT**. After hybridization, T4 DNA ligase ligates the reagent DNAs to the **UT** before amidation reaction. EDC, 1-ethyl-3-(3-dimethyl-aminopro-pyl) carbodiimide hydrochloride, 200 mM; NHS, *N*-hydroxy succinimide, 30 mM. After capping unreacted amines with Ac_2O, photocleavage regenerates the 5′-phosphate and 3′-amine for the next cycle. **UT**, 1 μM; each reagent DNA, 1.2 μM; T4 DNA ligase, 350 units at 16 °C for 15 h. Irradiation: 365 nm at 0 °C for 10 min.

pair indiscriminately with each of the four standard bases;[11] therefore, the "**III**" anticodon can hybridize with multiple DNAs composed of different codon sequences. In regular DTS, chemical synthesis is pre-encoded within individual templates and reagent DNAs are removed after delivering the building block. In our system, we ligate reagent DNAs to the template to directly encode the synthesis (Figure 1b). To achieve this, we designed a 5′-hairpin structure with a 5′-phosphate group on the **UT**; thus, the template can be enzymatically ligated to the 3′-hydroxyl group of reagent DNAs to record the synthesis. Ligation forms a native DNA backbone fully compatible with the postselection PCR amplification and DNA sequencing.

Figure 2b shows the architecture of the reagent DNAs. Each building block in the reagent DNA is encoded by a 3-base codon. Bases flanking the codon are unique and complementary to the ones flanking the respective "**III**" site on the **UT**. This important feature ensures the correct positioning of reagent DNAs at each step. In **RD1** and **RD2**, a constant 5-base sequence is included to form an "omega" structure upon hybridization, a feature known to improve DTS reaction yields.[12] We incorporated two photocleavable linkers[13] which serve a dual purpose. Unlike other linkers used in DTS,[14] photocleavage by irradiation is more convenient and milder because no cleavage reagent is necessary (e.g., strong base or strong oxidant).[13c] More importantly, linker cleavage regenerates both the 5′-phosphate group and the 3′-amino group on the **UT** (Figure 2c), which are necessary for further ligation and chemical reaction. Enabled by these features, after three cycles of DNA hybridization, ligation, templated reaction, and photocleavage, the final product can be synthesized at the 3′-terminus of the template, directly encoded by the original reagent DNA strands delivering the building blocks (Figure 2c).

First, we prepared template **UT** and reagent DNAs. **UT** can be straightforwardly synthesized via automated DNA synthesis using commercially available modifiers. We confirmed that **I**-containing duplexes are sufficiently stable for DTS reactions (Figure S1). Photocleavable linkers in reagent DNAs were incorporated with two special phosphoramidites based on a report by Taylor and co-workers (Figure S2).[13a,b] We confirmed that the **UT** can direct reactions with multiple reagent DNAs made of different sequences (Figure S3). These reagent DNAs could only hybridize at the site with complementary flanking bases on the template (Figure S4). These results demonstrated the specificity in positioning reagent DNAs by the universal template. In addition, T4-mediated ligation is highly sensitive to mismatches at the ligation site and it provides high encoding fidelity in library synthesis (Figure S5). Finally, we verified the regeneration of the 5′-phosphate and 3′-amino groups by photocleavage (Figure S6 and S7).

Next, we performed a multistep synthesis with the template **UT** and **RD1**−**RD3** (Figure 2c). The reactions can be followed with denaturing electrophoresis (Figure 3a). With the linear template used in DTS, the template/reagent DNA duplex can be denatured and analyzed based on DNA lengths; however, we observed more severe band smearing in our reactions than for regular DTS. We reason it may be due to the hairpin structure of the **UT** and the cyclic nature of the reaction intermediates (**iii** in Figure 2c; **iv-3**, and **v-3** in Figure S8), which may not fully denature and can form various secondary structures. We gel-purified the cyclic intermediates (**iii**, **iv-3** and **v-3**) before irradiation in order to isolate them from unligated and/or unreacted template/reagents. After irradiation, band smearing was significantly reduced, suggesting a cyclic to linear structural change. Intermediates (**iv** and **v**) and the final product (**vi**) were also gel-purified. After *Afl*II digestion to remove the bulk of the DNA, we were able to characterize the intermediates and the product by MALDI-MS (Figure 3b). This data has demonstrated the capability of the universal template in directing multistep DTS. It is important to note while a very low DNA concentration is required to minimize mismatched reactions (<120 nM/codon) in standard DTS,[2j,5c,10] with the universal template, much higher DNA concentrations may be employed (~1 μM; Figure S9), a highly advantageous feature for large-scale DEL synthesis.

Lastly, we prepared a model "64 × 28 × 64" library with a universal template (**UT1**) and 3 sets of reagent DNAs (**RD4**, **RD5**, **RD6**; Figure 4). There were 64 different 3-base codons in **RD4** and **RD6**. For simplicity, they only encoded one amino acid (Leu in **RD4** and Phe in **RD6**). In **RD5**, a single 3-base sequence ("**CCC**") encoded a special biotinyl-labeled lysine

dx.doi.org/10.1021/ja409936r | J. Am. Chem. Soc. 2013, 135, 17727−17730

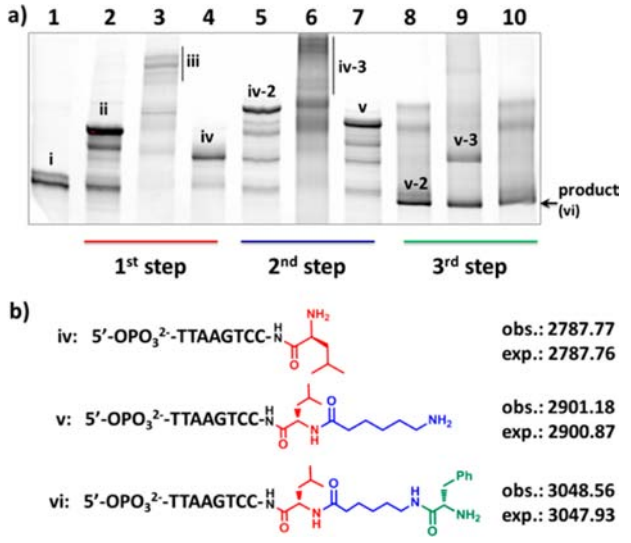

Figure 3. (a) Reactions in Figure 2c were analyzed by denaturing electrophoresis. Lane 1, **UT**; lane 2, **UT**/**RD1** after ligation; lane 3, lane 2 after amidation; lane 4, lane 3 after irradiation; lanes 5−7 and 8−10, 2nd and 3rd reaction cycles with **RD2** and **RD3**. Presumed structures of **i**, **ii**, **iii**, and **iv** are shown in Figure 2, and structures of **iv-2**, **iv-3**, **v-2**, and **v-3** are shown in Figure S8. (b) MALDI-MS characterizations of **iv**, **v**, and **vi** after *Afl*II digestion. See Figures S10−S12 for full spectra. Ethanol precipitation was performed after T4 ligation and amidation. Cyclic intermediates in lanes 3, 6, and 9 were gel-purified before irradiation and the next reaction cycle.

(bio-Lys, **RD5-1**), while the 27 other codons ("**DDD**", D = A, G, or T; **RD5-2**) encoded ε-Ahx (6-aminohexanoic acid). We intentionally mixed the bio-Lys-containing **RD5-1** with a large excess of **RD5-2** (1:100) as the "**RD5**". After library synthesis with **UT1** and reagent DNAs following the procedure shown in Figure 2c, we performed a selection against immobilized streptavidin as the target, which should only pull-down library compounds containing bio-Lys encoded by the "**CCC**" codon. The selected compounds were eluted, digested by *Hind*III (for hairpin removal; Figure S13), PCR-amplified and analyzed by DNA sequencing (Figure S14). As shown in Figure 4c, before selection, the sequences detected at all three codon positions were scrambled. After selection, the encoding sequence **CCC** for bio-Lys at codon 2 was significantly enriched (85.6-fold, see the Supporting Information) and distinctly identified (as "**GGG**" of the complementary strand). This result is corroborated by the sequencing result of the opposite strand (Figure S15). As expected, sequences at the other two codon sites remained scrambled after the selection. To further demonstrate the generality and performance of our method, we prepared a similar "64 × 64 × 28" library containing a phenyl sulfonamide, a known specific binder for carbonic anhydrase II (CA-II; K_i = 9.0 nM),[15] at the codon 3 position. This library was selected against immobilized CA-II and 94.7-fold enrichment of the sulfonamide was obtained (Figure S16). Again, no sequence enrichment was observed at the other two codon positions. Collectively, these data have demonstrated the encoding fidelity of our method and also its capability in DEL library synthesis, selection and hit identification.

In summary, we have developed a novel DNA-templated synthesis strategy. Taking advantage of deoxyinosine's indiscriminate base-pairing property, we have shown a single DNA template is capable of directing DNA-templated reactions with multiple reagent DNAs composed of different sequences. Along

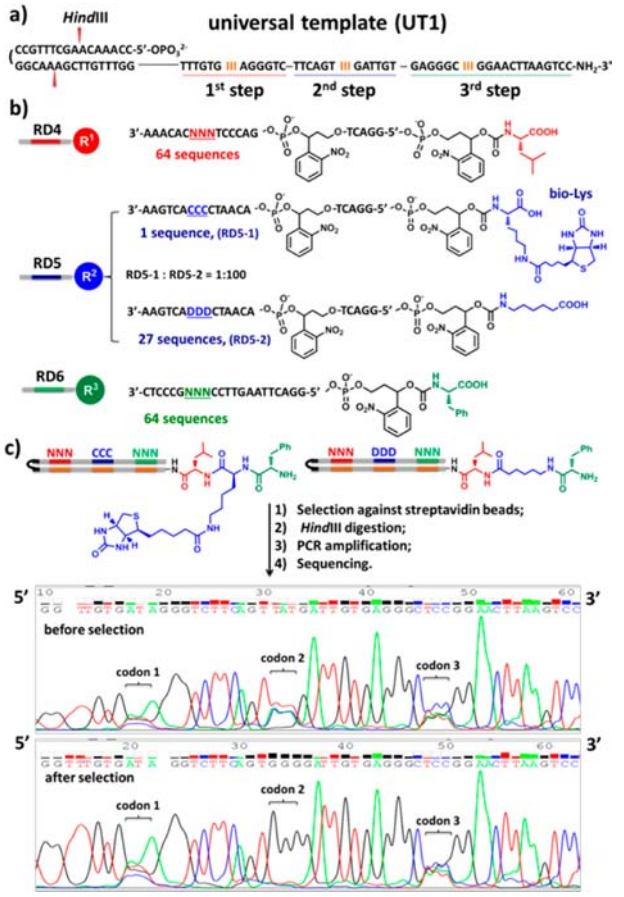

Figure 4. (a) The universal template **UT1**. (b) Reagent DNAs (**RD4**, **RD5**, and **RD6**) used for the model library. Amino acids in **RD4** and **RD6** are encoded by a scrambled 3-base codon. Bio-Lys is encoded by "**CCC**" (**RD5-1**) and ε-Ahx is encoded by "**DDD**" (**RD5-2**). (c) The model library was selected against immobilized streptavidin. Bound molecules were eluted, digested by *Hind*III, amplified, and sequenced. Sequencing results at 3 codon sites before and after the selection were compared as marked in the figure.

with other design features, such as the photocleavable linkers and ligation-based encoding, an entire DNA-encoded library can be prepared with a single universal template regardless of the size of the library. Our approach significantly simplifies template preparation and codon design. This method directs the synthesis of library compounds at the DNA terminus; therefore, most previously reported DTS reactions,[6h] linker designs,[14] selection strategies and decoding methods[2h,4a] can be straightforwardly utilized in our method to prepare a variety of chemically diverse DNA-encoded libraries. Certainly, more deoxyinosines can be used in anticodons to increase the encoding capacity. Currently, our laboratory is actively implementing this method in DEL synthesis and selections against biological targets.

■ ASSOCIATED CONTENT

Ⓢ Supporting Information

Detailed DNA preparation, structures, sequences, characterization data, library synthesis and selection procedures and conditions, PCR protocols, DNA sequencing methods and data analysis, and other experimental details. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author
xiaoyuli@pku.edu.cn

Notes
The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was supported by Ministry of Science and Technology Basic Research Program (2011CB809100), National Natural Science Foundation of China (21002003, 91013003, 21272016), the Beijing Nova Program (2010B002), and the Doctoral Fund of Ministry of Education of China (20120001110083).

REFERENCES

(1) Brenner, S.; Lerner, R. A. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 5381−5383.
(2) (a) Nielsen, J.; Brenner, S.; Janda, K. D. . *Am. Chem. Soc.* **1993**, *115*, 9812−9813. (b) Needels, M. C.; Jones, D. G.; Tate, E. H.; Heinkel, G. L.; Kochersperger, L. M.; Dower, W. J.; Barrett, R. W.; Gallop, M. A. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 10700−10704. (c) Halpin, D. R.; Harbury, P. B. *PLoS Biol.* **2004**, *2*, 1015−1021. (d) Scheuermann, J.; Dumelin, C. E.; Melkko, S.; Neri, D. *J. Biotechnol.* **2006**, *126*, 568−581. (e) Melkko, S.; Dumelin, C. E.; Scheuermann, J.; Neri, D. *Drug Discovery Today* **2007**, *12*, 465−471. (f) Silverman, S. K. *Angew. Chem., Int. Ed.* **2010**, *49*, 7180−7201. (g) Clark, M. A. *Curr. Opin. Chem. Biol.* **2010**, *14*, 396−403. (h) Kleiner, R. E.; Dumelin, C. E.; Liu, D. R. *Chem. Soc. Rev.* **2011**, *40*, 5707−5717. (i) Mannocci, L.; Leimbacher, M.; Wichert, M.; Scheuermann, J.; Neri, D. *Chem. Commun.* **2011**, *47*, 12747−12753. (j) Daguer, J. P.; Ciobanu, C.; Alvarez, S.; Barluenga, S.; Winssinger, N. *Chem. Sci.* **2011**, *2*, 625−632.
(3) Clark, M. A.; Acharya, R. A.; Arico-Muendel, C. C.; Belyanskaya, S. L.; Benjamin, D. R.; Carlson, N. R.; Centrella, P. A.; Chiu, C. H.; Creaser, S. P.; Cuozzo, J. W.; Davie, C. P.; Ding, Y.; Franklin, G. J.; Franzen, K. D.; Gefter, M. L.; Hale, S. P.; Hansen, N. J. V.; Israel, D. I.; Jiang, J. W.; Kavarana, M. J.; Kelley, M. S.; Kollmann, C. S.; Li, F.; Lind, K.; Mataruse, S.; Medeiros, P. F.; Messer, J. A.; Myers, P.; O'Keefe, H.; Oliff, M. C.; Rise, C. E.; Satz, A. L.; Skinner, S. R.; Svendsen, J. L.; Tang, L. J.; van Vloten, K.; Wagner, R. W.; Yao, G.; Zhao, B. G.; Morgan, B. A. *Nat. Chem. Biol.* **2009**, *5*, 647−654.
(4) (a) Mannocci, L.; Zhang, Y. X.; Scheuermann, J.; Leimbacher, M.; De Bellis, G.; Rizzi, E.; Dumelin, C.; Melkko, S.; Neri, D. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 17670−17675. (b) Buller, F.; Mannocci, L.; Zhang, Y.; Dumelin, C. E.; Scheuermann, J.; Neri, D. *Bioorg. Med. Chem. Lett.* **2008**, *18*, 5926−5931. (c) Buller, F.; Zhang, Y.; Scheuermann, J.; Schafer, J.; Buhlmann, P.; Neri, D. *Chem. Biol.* **2009**, *16*, 1075−1086. (d) Kleiner, R. E.; Dumelin, C. E.; Tiu, G. C.; Sakurai, K.; Liu, D. R. *J. Am. Chem. Soc.* **2010**, *132*, 11779−11791. (e) Buller, F.; Steiner, M.; Frey, K.; Mircsof, D.; Scheuermann, J.; Kalisch, M.; Buhlmann, P.; Supuran, C. T.; Neri, D. *ACS Chem. Biol.* **2011**, *6*, 336−344.
(5) (a) Gartner, Z. J.; Liu, D. R. *J. Am. Chem. Soc.* **2001**, *123*, 6961−6963. (b) Gartner, Z. J.; Tse, B. N.; Grubina, R.; Doyon, J. B.; Snyder, T. M.; Liu, D. R. *Science* **2004**, *305*, 1601−1605. (c) Tse, B. N.; Snyder, T. M.; Shen, Y. H.; Liu, D. R. *J. Am. Chem. Soc.* **2008**, *130*, 15611−15626.
(6) (a) Summerer, D.; Marx, A. *Angew. Chem., Int. Ed.* **2002**, *41*, 89−90. (b) Silverman, A. P.; Kool, E. T. *Chem. Rev.* **2006**, *106*, 3775−3789. (c) Gorska, K.; Huang, K. T.; Chaloin, O.; Winssinger, N. *Angew. Chem., Int. Ed.* **2009**, *48*, 7695−7700. (d) Boersma, A. J.; Megens, R. P.; Feringa, B. L.; Roelfes, G. *Chem. Soc. Rev.* **2010**, *39*, 2083−2092. (e) Xu, L.; Liu, D. *Chem. Soc. Rev.* **2010**, *39*, 150−155. (f) Diezmann, F.; Seitz, O. *Chem. Soc. Rev.* **2011**, *40*, 5789−5801. (g) Englund, E. A.; Wang, D.; Fujigaki, H.; Sakai, H.; Micklitsch, C. M.; Ghirlando, R.; Martin-Manso, G.; Pendrak, M. L.; Roberts, D. D.; Durell, S. R.; Appella, D. H. *Nat. Commun.* **2012**, *3*, 614. (h) Gorska, K.; Winssinger, N. *Angew. Chem., Int. Ed.* **2013**, *52*, 6820−6843.
(7) Furka, A.; Sebestyen, F.; Asgedom, M.; Dibo, G. *Int. J. Pept. Protein Res.* **1991**, *37*, 487−493.
(8) Snyder, T. M.; Tse, B. N.; Liu, D. R. *J. Am. Chem. Soc.* **2008**, *130*, 1392−1401.
(9) Melkko, S.; Scheuermann, J.; Dumelin, C. E.; Neri, D. *Nat. Biotechnol.* **2004**, *22*, 568−574.
(10) Hansen, M. H.; Blakskjaer, P.; Petersen, L. K.; Hansen, T. H.; Hojfeldt, J. W.; Gothelf, K. V.; Hansen, N. J. V. *J. Am. Chem. Soc.* **2009**, *131*, 1322−1327.
(11) (a) Martin, F. H.; Castro, M. M.; Aboul-ela, F.; Tinoco, I., Jr. *Nucleic Acids Res.* **1985**, *13*, 8927−8938. (b) Loakes, D. *Nucleic Acids Res.* **2001**, *29*, 2437−2447.
(12) Gartner, Z. J.; Grubina, R.; Calderone, C. T.; Liu, D. R. *Angew. Chem., Int. Ed.* **2003**, *42*, 1370−1375.
(13) (a) Ordoukhanian, P.; Taylor, J. S. *Bioconjugate Chem.* **2000**, *11*, 94−103. (b) Ordoukhanian, P.; Taylor, J. S. *J. Am. Chem. Soc.* **1995**, *117*, 9570−9571. (c) Li, Y.; Zhang, M.; Zhang, C.; Li, X. *Chem. Commun.* **2012**, *48*, 9513−9515.
(14) (a) Gartner, Z. J.; Kanan, M. W.; Liu, D. R. *J. Am. Chem. Soc.* **2002**, *124*, 10304−10306. (b) Hojfeldt, J. W.; Blakskjaer, P.; Gothelf, K. V. *J. Org. Chem.* **2006**, *71*, 9556−9559.
(15) Mincione, F.; Starnotti, M.; Menabuoni, L.; Scozzafava, A.; Casini, A.; Supuran, C. T. *Bioorg. Med. Chem. Lett.* **2001**, *11*, 1787−1791.