DNA-Encoded Library Hot Paper

DOI: 10.1002/anie.201404830

Selection of DNA-Encoded Small Molecule Libraries Against Unmodified and Non-Immobilized Protein Targets**

Peng Zhao, Zitian Chen, Yizhou Li, Dawei Sun, Yuan Gao, Yanyi Huang, and Xiaoyu Li**

Abstract: *The selection of DNA-encoded libraries against biological targets has become an important discovery method in chemical biology and drug discovery, but the requirement of modified and immobilized targets remains a significant disadvantage. With a terminal protection strategy and ligand-induced photo-crosslinking, we show that iterated selections of DNA-encoded libraries can be realized with unmodified and non-immobilized protein targets.*

In 1992, Brenner and Lerner proposed a visionary concept of using DNA to encode combinatorial libraries.[1] During the past two decades, many synthesis, selection, and decoding strategies for DNA-encoded libraries (DELs) have been developed.[2] Today, DELs can be prepared with extremely large numbers of compounds,[2b,3] and hit decoding can be feasibly accomplished by reading the DNA tags.[3,4] Researchers have discovered many novel binders from DEL selections,[2g,h] and pharmaceutical companies have also adopted DELs in drug discovery.[3,5]

Typically, DELs are selected against immobilized proteins. However, immobilization is not compatible with proteins that require a native cellular environment, or that are difficult to purify or modify, such as protein complexes and membrane proteins. Previously, peptide nucleic acid (PNA) encoded small-molecule (SM) libraries were used to profile enzyme activities[6] and select against live cells.[7] Selections with targets of DNA–protein conjugates have also been reported.[8] Recently, Liu and co-workers developed a method called IDUP (interaction determination using unpurified proteins), which can directly select endogenous proteins in cell lysates.[9] However, the iterated selection of DELs against completely non-immobilized and unmodified proteins has yet to be realized, which would be essential to enrich low-abundance binders from a large population of nonbinders in a library.[10]

Recently, we reported a DNA-based protein-labeling method, the DNA-programmed affinity labeling (DPAL; Figure 1a),[11] in which a DNA-linked small molecule guides

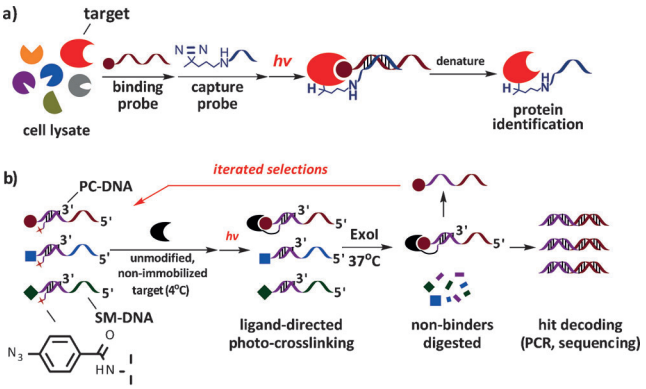


Figure 1. a) DNA-programmed affinity labeling (DPAL). b) Proposed selection method. A photo-reactive DNA (PC-DNA) hybridizes with library DNAs (SM-DNAs) at their 3′ end. PC-DNA is crosslinked to the target upon small-molecule binding under irradiation, protecting SM-DNAs from ExoI digestion. Surviving SM-DNAs can be directly decoded or subjected to iterated selections for further enrichments.

the capture and identification of its target through photo-crosslinking in cell lysates. We reason that this concept may be used conversely to select specific small-molecule binders from DELs against protein targets. In our design, DNA-encoded small molecules (SM-DNAs) hybridize at the selection temperature (4 °C) with a short 8-nt DNA strand that bears a 5′-azidophenyl group (PC-DNA; Figure 1b). After target binding and irradiation, a more stable "hairpin" structure is formed,[8a,9,12] protecting SM-DNAs from being digested by exonuclease I (ExoI). For nonbinders, the 8-nt duplex is denatured and readily degraded at digestion temperature (37 °C). Surviving SM-DNAs can be decoded or subjected to iterated selections for further enrichments. Selective nuclease digestion has been used in terminal protection assays for protein detection[13] and in selection decoding for PNA-

[*] Prof. X. Li
Key Laboratory of Bioorganic Chemistry and Molecular Engineering
College of Chemistry and Molecular Engineering, Peking University
Beijing 100871 (China)
and
Key Laboratory of Chemical Genomics, School of Chemical Biology and Biotechnology, Peking University Shenzhen Graduate School
Shenzhen 518055 (China)
E-mail: xiaoyuli@pku.edu.cn

Prof. Y. Huang
Biodynamic Optical Imaging Center (BIOPIC) and
College of Engineering, Peking University
Beijing, 100871 (China)
E-mail: yanyi@pku.edu.cn

P. Zhao, Y. Li, D. Sun, Y. Gao
College of Chemistry and Molecular Engineering, Peking University
Beijing, 100871 (China)

Z. Chen
College of Engineering, Peking University
Beijing, 100871 (China)

[**] This work was supported by the Ministry of Science and Technology Basic Research Program (2011CB809100), the NSFC (21272016, 91013003, 21222501, and 91313302), and a Doctoral Fund of the Ministry of Education (20120001110083).

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/anie.201404830.

encoded libraries.[7a] We reasoned that the "hairpin" formation would strengthen SM–protein interactions so that moderate and low-affinity binders may also be protected.

We first tested two SM–protein pairs: GLCBS–CA-II ($K_i = 9$ nM; CA-II = carbonic anhydrase II) and CBS–CA-II ($K_d = 3.2$ μM). The small molecule (i.e., GLCBS or CBS) was conjugated to the 3′ end of a 21-nt DNA strand with a fluorescein (FAM) group at the 5′ end.[11] A 27-nt DNA strand that was conjugated with CBM, a nonbinding small molecule (CBM-DNA; Figure S2) was used as the negative control. After incubating SM-DNA, CBM-DNA, PC-DNA, and CA-II at 4 °C, the mixture was irradiated and subjected to ExoI digestion. Figure 2 shows that GLCBS- and CBS-DNA can be protected in the presence of CA-II after irradiation (lane 2). No or little protection was observed for CBM-DNA and in negative control experiments (lane 3–7), proving that the protection requires both specific SM–protein interaction and photo-crosslinking.

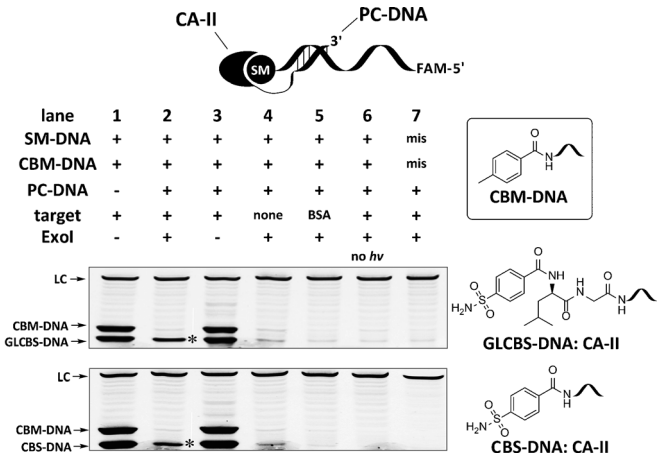
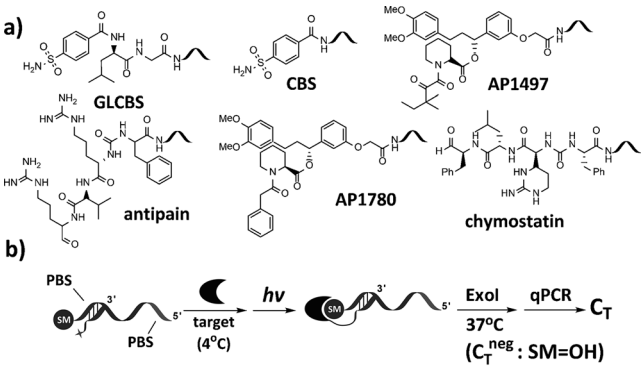


Figure 2. PAGE analysis of SM-DNA protection by CA-II. SM-DNA: 1 μM each; PC-DNA: 2 μM; proteins: 2 μM. ExoI: 20 units, 37 °C, 60 min. Lane 1: SM-DNAs and CA-II; lane 2: digestion of SM-DNA, CBM-DNA, and PC-DNA with CA-II after *hν*; lane 3–7: same as lane 2, but no ExoI, no CA-II, with BSA, no *hν*, and with mismatched PC-DNA. LC: a 66-nt 5′-FAM-labeled control DNA. Asterisks mark surviving SM-DNAs.

Next, several SM–protein pairs of different affinities were tested and then analyzed by quantitative PCR (qPCR),[8a,10] as target-bound SM-DNA that survived digestion will be amplified more rapidly to give smaller threshold cycle value (C_T).[8a,12] As shown in Figure 3c, the ΔC_T values clearly indicate that SM-DNAs were amplified more rapidly in the presence of their targets. Even for the weak interaction between chymostatin and papain ($K_i = 14$ μM), a ΔC_T value of 4.7 was observed. Cell lysates spiked with CA-II can also be used as the "target" to afford large ΔC_T values (Figure 3c, entry 8 and 9, and Figure S4), indicating the method may be used for targets that require a more native environment. ExoI digestions of individual SM-DNAs were tested to show that ExoI can efficiently digest SM-DNAs with different chemical structures (Figure S5). Control qPCR experiments again confirmed the specificity (Figure S6). These results have demonstrated the generality of the method for various SM–protein interactions.

The primary goal of a selection is to identify small numbers of specific binders from a large library population, which requires multiple rounds of enrichments.[10] To demonstrate this, we mixed SM-DNA and CBM-DNA with orthogonal primer binding sites (PBS) at various ratios (Figure 4a). After the first round of selection with CA-II, the selected SM-DNAs were isolated and mixed with a fresh batch of CA-II and PC-DNA for the second round. The percentage of SM-DNA after each round of selection was calculated based on the C_T values. Results show that both GLCBS- and CBS-DNA were significantly enriched after the selections (Figure 4b). At an initial GLCBS-DNA/CBM-DNA ratio of 1:100, GLCBS-DNA became the major species in the mixture (70%) after two rounds of selection. At an initial GLCBS-DNA/CBM-DNA ratio of 1:1000, GLCBS-DNA was only enriched by a factor of 7.4 after the first round of selection, but became 12% (120-fold increase) of the selected population after the second round, showing the importance of iterated selections. The weaker binder CBS-DNA can also be enriched from 1% to 31% after two rounds at an initial ratio of 1:100 of CBS-DNA/CBM-DNA. In comparison, CBS-DNA was barely detectable in a selection with the immobilized CA-II target (Figure S8).

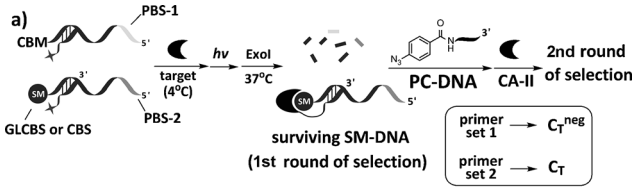
Next, we prepared a model DEL composed of GLCBS, CBS, and CBM (each encoded by a 6-base codon) and a pool of background DNAs of 1024 sequences at equal ratio (for simplicity, no small molecule was conjugated). After the addition of PC-DNA, the library was subjected to two rounds



c)

entry	SM : protein	$K_d/K_i/IC_{50}$ (nM)	ΔC_T
1	GLCBS : CA-II	9.0	7.7
2	AP1497 : FKBP12	25.7	5.7
3	antipain : trypsin	100	3.6
4	chymostatin : chymotrypsin	290	3.6
5	CBS : CA-II	3,200	4.7
6	AP1780 : FKBP12	5,170	3.0
7	chymostatin : papain	14,000	4.7
8	GLCBS : CA-II/HeLa	9,0	6.5
9	GLCBS : CA-II/293T	9,0	7.6

Figure 3. a) Structures of SM-DNAs. b) SM-DNA/PC-DNA was subjected to the same selection procedure with respective target as in Figure 2. c) Table of ΔC_T values from qPCR analysis. $\Delta C_T = C_T^{neg} - C_T$. ($C_T^{neg}$: from a nonbinding control DNA). All values are averages of three experiments. PBS = primer-binding site.



b)

ratio →	GLCBS-DNA 1:100	GLCBS-DNA 1:1000	CBS-DNA 1:10	CBS-DNA 1:100
before selection	1%	0.1%	10%	1%
1st round	13%	0.74%	51%	3.8%
2nd round	70%	12%	80%	31%

Figure 4. a) GLCBS- or CBS-DNA were mixed with CBM-DNA at various ratios. The mixture was subjected to the same selection procedure as in Figure 2. After the first round, selected SM-DNAs were isolated for the second round. b) GLCBS- and CBS-DNA percentages in the mixture before and after each round of selection, deduced from qPCR. See the Supporting Information for details.

of selections against CA-II. Selected members were amplified and decoded by high-throughput sequencing. After the first round, GLCBS and CBS were enriched by factors of 38.4 and 7.2, respectively, while much higher enrichments were observed (factors of 179.3 and 213.7, respectively) after the second round (see Figure 5b), again demonstrating the importance of selection iteration.

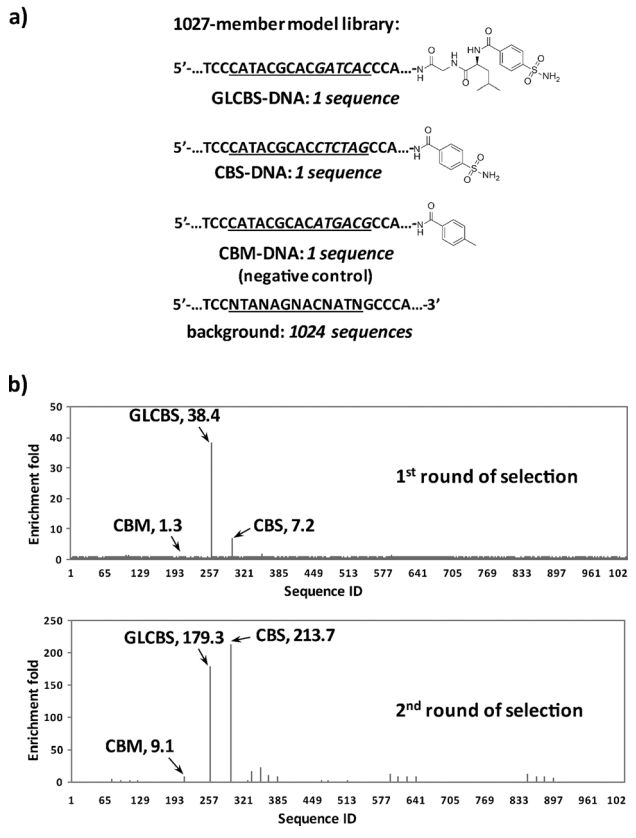


Figure 5. a) A model library was selected against CA-II employing the same procedure as in Figure 2. An equal amount of PC-DNA was added before the selection. N denotes either A, T, C, or G. b) Selection results. Enrichment folds of SM-DNAs are marked. See the Supporting Information for details. Enrichment fold = (postselection fraction)/(preselection fraction).[2g]

As DELs with chemical diversity and sequence complexity would truly test the performance of our method, following the well-developed strategy for DNA-encoded macrocycle libraries by Liu and co-workers,[4d,14] a DEL of 4800 macrocycles was prepared (Figure S9). GLCBS-DNA was added as a positive control (Figure 6a). After adding PC-DNA, the library was selected against CA-II. Results show that GLCBS was distinctly enriched by a factor of 98.2 (Figure 6b). No noticeable enrichment of any macrocycles was observed, possibly as a result of the moderate diversity of the library and the lack of a privileged CA-II-binding structure.[15] When we selected the same library with immobilized CA-II, we also did not observe a significant enrichment of macrocycles (Figure S10). To control for false positives from ExoI resistance, we also performed a selection without CA-II and observed no enrichment (Figure S11); this type of control is analogous to the "beads-only" selection frequently used in immobilization-based selection.[3,4d] Finally, a "blank library" without small molecules was selected to control for DNA–protein interactions, and again no enrichment was observed (Figure S12). Overall, these results validated the capability of the method to select low-abundance-specific binders from chemically diverse DELs with complex sequences.

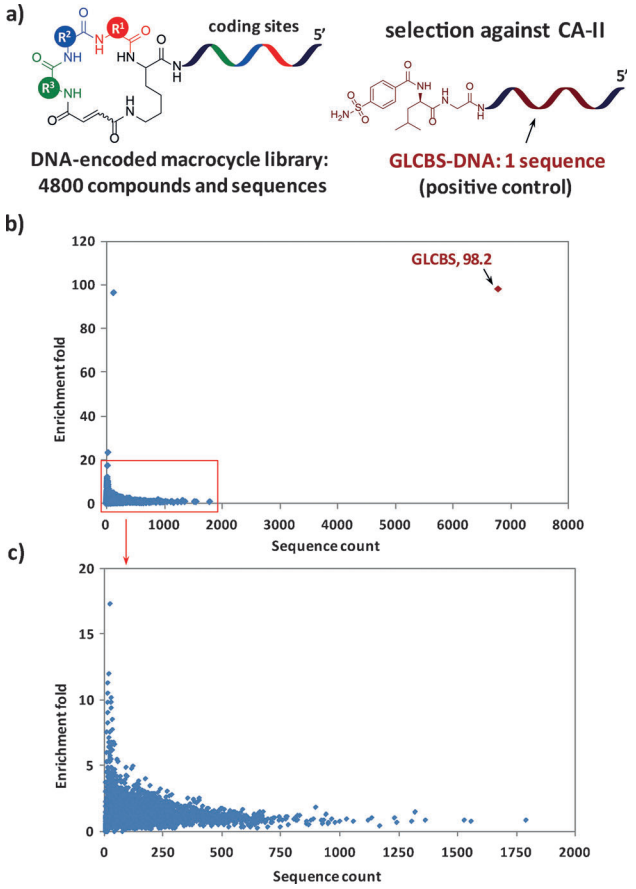


Figure 6. a) Composition of the macrocycle library. b) Plot of enrichment fold versus post-selection sequence count. Bottom panel: zoom-in of the lower left portion of the top panel. Enrichment fold = (postselection fraction)/(preselection fraction).[2g] See the Supporting Information for details.

In summary, we have developed a novel DEL selection strategy and realized iterated selections against unmodified and non-immobilized protein targets. The selective removal of nonspecific background by ExoI digestion enables enrichment and iterated selections. 5′-Specific nucleases[16] or ExoIII may be used for DELs that bear small molecules at the 5′ end or are encoded by double-stranded DNA. Our strategy requires little library redesign or resynthesis, as most DELs already have PBS available for PC-DNA hybridization. This advantageous feature may enable a rapid utilization of existing DELs to interrogate previously intractable targets.

Received: April 30, 2014
Published online: July 7, 2014

Keywords: DNA · DNA-encoded library · DNA-templated chemistry · drug discovery · selection

[1] S. Brenner, R. A. Lerner, *Proc. Natl. Acad. Sci. USA* **1992**, *89*, 5381 – 5383.

[2] a) J. Nielsen, S. Brenner, K. D. Janda, *J. Am. Chem. Soc.* **1993**, *115*, 9812; b) M. C. Needels, D. G. Jones, E. H. Tate, G. L. Heinkel, L. M. Kochersperger, W. J. Dower, R. W. Barrett, M. A. Gallop, *Proc. Natl. Acad. Sci. USA* **1993**, *90*, 10700 – 10704; c) D. R. Halpin, P. B. Harbury, *Plos Biol.* **2004**, *2*, 1015 – 1021; d) J. Scheuermann, C. E. Dumelin, S. Melkko, D. Neri, *J. Biotechnol.* **2006**, *126*, 568 – 581; e) S. K. Silverman, *Angew. Chem.* **2010**, *122*, 7336 – 7359; *Angew. Chem. Int. Ed.* **2010**, *49*, 7180 – 7201; f) M. A. Clark, *Curr. Opin. Chem. Biol.* **2010**, *14*, 396 – 403; g) R. E. Kleiner, C. E. Dumelin, D. R. Liu, *Chem. Soc. Rev.* **2011**, *40*, 5707 – 5717; h) L. Mannocci, M. Leimbacher, M. Wichert, J. Scheuermann, D. Neri, *Chem. Commun.* **2011**, *47*, 12747 – 12753; i) J. P. Daguer, S. Alvarez, S. Barluenga, N. Winssinger, *Chem. Sci.* **2011**, *2*, 9; j) S. J. Wrenn, R. M. Weisinger, D. R. Halpin, P. B. Harbury, *J. Am. Chem. Soc.* **2007**, *129*, 13137 – 13143; k) N. Winssinger, K. Gorska, M. Ciobanu, J. P. Daguer, S. Barluenga, *Methods Mol. Biol.* **2014**, *1050*, 95 – 110.

[3] M. A. Clark, et al., *Nat. Chem. Biol.* **2009**, *5*, 647 – 654.

[4] a) L. Mannocci, Y. X. Zhang, J. Scheuermann, M. Leimbacher, G. De Bellis, E. Rizzi, C. Dumelin, S. Melkko, D. Neri, *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 17670 – 17675; b) F. Buller, L. Mannocci, Y. Zhang, C. E. Dumelin, J. Scheuermann, D. Neri, *Bioorg. Med. Chem. Lett.* **2008**, *18*, 5926 – 5931; c) F. Buller, Y. Zhang, J. Scheuermann, J. Schafer, P. Buhlmann, D. Neri, *Chem. Biol.* **2009**, *16*, 1075 – 1086; d) R. E. Kleiner, C. E. Dumelin, G. C. Tiu, K. Sakurai, D. R. Liu, *J. Am. Chem. Soc.* **2010**, *132*, 11779 – 11791; e) F. Buller, M. Steiner, K. Frey, D. Mircsof, J. Scheuermann, M. Kalisch, P. Buhlmann, C. T. Supuran, D. Neri, *ACS Chem. Biol.* **2011**, *6*, 336 – 344.

[5] a) M. H. Hansen, P. Blakskjaer, L. K. Petersen, T. H. Hansen, J. W. Hojfeldt, K. V. Gothelf, N. J. V. Hansen, *J. Am. Chem. Soc.* **2009**, *131*, 1322 – 1327; b) J. S. Disch, et al., *J. Med. Chem.* **2013**, *56*, 3666 – 3679; c) H. Encinas, et al., *J. Med. Chem.* **2014**, *57*, 1276 – 1288.

[6] a) N. Winssinger, J. L. Harris, B. J. Backes, P. G. Schultz, *Angew. Chem.* **2001**, *113*, 3254 – 3258; *Angew. Chem. Int. Ed.* **2001**, *40*, 3152 – 3155; b) N. Winssinger, S. Ficarro, P. G. Schultz, J. L. Harris, *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 11139 – 11144; c) J. Harris, et al., *Chem. Biol.* **2004**, *11*, 1361 – 1372; d) H. D. Urbina, F. Debaene, B. Jost, C. Bole-Feysot, D. E. Mason, P. Kuzmic, J. L. Harris, N. Winssinger, *ChemBioChem* **2006**, *7*, 1790 – 1797.

[7] a) N. Svensen, J. J. Diaz-Mochon, M. Bradley, *Chem. Biol.* **2011**, *18*, 1284 – 1289; b) N. Svensen, J. J. Diaz-Mochon, K. Dhaliwal, S. Planonth, M. Dewar, J. D. Armstrong, M. Bradley, *Angew. Chem.* **2011**, *123*, 6257 – 6260; *Angew. Chem. Int. Ed.* **2011**, *50*, 6133 – 6136.

[8] a) L. M. McGregor, D. J. Gorin, C. E. Dumelin, D. R. Liu, *J. Am. Chem. Soc.* **2010**, *132*, 15522 – 15524; b) P. Blakskjaer, et al., Patent WO2012041633A1, **2012**.4.5.

[9] L. M. McGregor, T. Jain, D. R. Liu, *J. Am. Chem. Soc.* **2014**, *136*, 3264 – 3270.

[10] J. B. Doyon, T. M. Snyder, D. R. Liu, *J. Am. Chem. Soc.* **2003**, *125*, 12372 – 12373.

[11] G. Li, Y. Liu, L. Chen, S. Wu, X. Li, *Angew. Chem.* **2013**, *125*, 9723 – 9728; *Angew. Chem. Int. Ed.* **2013**, *52*, 9544 – 9549.

[12] D. J. Gorin, A. S. Kamlet, D. R. Liu, *J. Am. Chem. Soc.* **2009**, *131*, 9189.

[13] a) Z. Wu, Z. Zhen, J. H. Jiang, G. L. Shen, R. Q. Yu, *J. Am. Chem. Soc.* **2009**, *131*, 12325 – 12332; b) Z. Wu, H. Wang, M. Guo, L. J. Tang, R. Q. Yu, J. H. Jiang, *Anal. Chem.* **2011**, *83*, 3104 – 3111; c) X. Wei, W. Lin, N. Ma, F. Luo, Z. Lin, L. Guo, B. Qiu, G. Chen, *Chem. Commun.* **2012**, *48*, 6184 – 6186; d) Y. Cao, S. Zhu, J. Yu, X. Zhu, Y. Yin, G. Li, *Anal. Chem.* **2012**, *84*, 4314 – 4320; e) Z. Zhen, L. J. Tang, J. Lin, J. H. Jiang, R. Q. Yu, X. Xiong, W. Tan, *Anal. Chem.* **2012**, *84*, 5708 – 5715.

[14] a) Z. J. Gartner, B. N. Tse, R. Grubina, J. B. Doyon, T. M. Snyder, D. R. Liu, *Science* **2004**, *305*, 1601 – 1605; b) B. N. Tse, T. M. Snyder, Y. H. Shen, D. R. Liu, *J. Am. Chem. Soc.* **2008**, *130*, 15611 – 15626.

[15] C. Kerr, P. D. Sadowski, *J. Biol. Chem.* **1972**, *247*, 305 – 310.

[16] a) J. Wang, T. Li, X. Guo, Z. Lu, *Nucleic Acids Res.* **2005**, *33*, e23; b) X. Xu, M. S. Han, C. A. Mirkin, *Angew. Chem.* **2007**, *119*, 3538 – 3540; *Angew. Chem. Int. Ed.* **2007**, *46*, 3468 – 3470; c) L. J. Ou, P. Y. Jin, X. Chu, J. H. Jiang, R. Q. Yu, *Anal. Chem.* **2010**, *82*, 6015 – 6024; d) A. Cao, C. Y. Zhang, *Anal. Chem.* **2013**, *85*, 2543 – 2547.

