

Nobel Lectures 2013: Multiscale Modeling

Reviews by M. Karplus, M. Levitt, and A. Warshel ACIEFS 53 (38) 9963–10252 (2014) · ISSN 1433–7851 · Vol. 53 · No. 38

WILEY-VCH

DNA-Encoded Library Hot Paper

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.

n 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

[*] 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) 7 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.

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 1 a),^[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 Exol digestion. Surviving SM-DNAs can be directly decoded or subjected to iterated selections for further enrichments.

the capture and identification of its target through photocrosslinking 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- 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 $(K_i = 9 \text{ nM}; \text{CA-II} = \text{carbonic anhydrase II})$ and CBS–CA-II ($K_d = 3.2 \,\mu\text{M}$). 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.

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 ($K_i = 14 \mu M$), 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 SMprotein 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-



Figure 2. PAGE analysis of SM-DNA protection by CA-II. SM-DNA: 1 μ M each; PC-DNA: 2 μ M; proteins: 2 μ M. 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 $h\nu$; lane 3–7: same as lane 2, but no ExoI, no CA-II, with BSA, no $h\nu$, and with mismatched PC-DNA. LC: a 66-nt 5'-FAM-labeled control DNA. Asterisks mark surviving SM-DNAs.

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



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.





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 immobilizationbased 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: zoomin of the lower left portion of the top panel. Enrichment fold = (postselection fraction)/(preselection fraction).^[2g] See the Supporting Information for details.

10058 www.angewandte.org

© 2014 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

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

- 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.



Supporting Information © Wiley-VCH 2014

69451 Weinheim, Germany

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*

anie_201404830_sm_miscellaneous_information.pdf

Table of Contents

1.	Abbreviations	S1
2.	Materials and General Methods.	S2
3.	Oligonucleotide Structures, Synthesis, Purification, and Characterization	S3
4.	DNA Oligonucleotide Sequences	S6
5.	Oligonucleotide Characterization.	S8
6.	Selection Conditions and Procedures	S9
7.	CA-II Binding Assay of CBM-DNA.	S10
8.	qPCR Analysis and Calculation Methods	S12
9.	High-Throughput Sequencing (Illumina [®]) Procedures and Conditions	S14
10.	Control Selection Experiments in Cell Lysates Only	S15
11.	Effects of Different Chemical Modifications at the 3'- of SM-DNA on ExoI's Digestion Reactivity	S17
12.	Control qPCR Experiments of the Protection of GLCBS-DNA and CBM-DNA from ExoI Digestion	S18
13.	Enrichment Fold Calculation Method Based on qPCR Analysis.	S19
14.	Selection Experiments with Immobilized CA-II.	S22
15.	Selection of the 1,207-Member Model Library.	S23
16.	Selection of the 4,801-MemberDNA-Encoded Library.	S24
17.	References	S29

1. Abbreviations.

AMA: 1:1 (v:v) aqueous methylamine (40% wt.): aqueous ammonium hydroxide (30% wt.) CA-II: carbonic anhydrase II CBS: 4-sulfamoylbenzoic acid CPG: controlled-pore glass DCC: N, N'-dicyclohexylcarbodiimide DCM: dichloromethane DIPEA: N, N'-diisopropylethylamine DMF: N, N'-dimethylformamide DMSO: dimethyl sulfoxide DMT: di-(4-methoxyphenyl)phenylmethyl DMT-MM: 4-(4, 6-dimethoxy-1, 3, 5-triazin-2-yl)-4-methylmorpholinium chloride DSC: N, N'-disuccinimidyl carbonate DTS: DNA-templated synthesis EDC: 1-ethyl-3- (3-dimethylaminopropyl) carbodiimide hydrochloride ExoI: Exonuclease I ETT: 5-ethylthio-1H-tetrazole FAM: carboxyfluorescein Fmoc: 9-fluorenylmethyloxycarbonyl GLCBS: Gly-Leu-4-sulfamoylbenzoic acid HBTU: O-benzotriazol-1-yl-N, N, N', N'-tetramethyluronium hexafluorophosphate HOBt: N-hydroxylbenzotriazole 3-HPA: 3-hydroxypicolinic acid MMT: (4-methoxyphenyl) diphenylmethyl NHS: N-hydroxysuccinimide PAGE: polyacrylamide gel electrophoresis PBS: phosphate-buffered saline SE: succinimidyl ester SM: small molecule sNHS: N-hydroxysulfosuccinimide sodium salt TBE: tris-borate-EDTA TCA: trichloroacetic acid TEAA: triethylammonium acetate TEA: triethylamine THAP: 2', 4', 6'-trihydroxyacetophenone monohydrate

TPCK: tosyl phenylalanyl chloromethyl ketone

2. Materials and General Methods.

Unless otherwise noted, all reagents and solvents were purchased from commercial sources and used as received. All chemical reagents were purchased from J&K Scientific Ltd. and Beijing Ouhe Technology Co. Ltd. Avidin and all PCR primers were purchased from Sangon Biotech. NHS-activated SepharoseTM 4 Fast Flow beads were purchased from GE Healthcare Life Science. Exonuclease I was purchased from New England Biolabs (NEB). CA-II and TPCK-treated trypsin

were purchased from Sigma-Aldrich. FKBP12 were purchased from Sino Biological Inc. qPCR reagents were purchased from Transgen Biotech. Water was purified with a Thermo Scientific Barnstead Nanopure system. Oligonucleotides were synthesized on standard CPG (Controlled Pore Glass, 1000 Å) beads by an automated Applied Biosystems 394 synthesizer following the machine's built-in synthesis programs. Standard phosphoramidites, other synthesis reagents and solvents were purchased from Hai Phoenix Technology and Glen Research. Anhydrous acetonitrile was freshly distilled over P_2O_5 prior to use. DNA oligonucleotides were purified by reverse-phase HPLC (Agilent 1200) using a gradient of acetonitrile (5-80%) in 100 mM TEAA (pH 7.0). DNA oligonucleotides were analyzed and purified by denaturing TBE-Urea PAGE containing 25% formamide, stained with ethidium bromide. All gel images were captured by a Bio-Rad Chemidoc system or a Tanon-1600 gel image system. Photo-crosslinking experiments were conducted by a UVP CL–1000L Ultraviolet crosslinker at 365 nm wavelengths with an intensity of approximately 100 μJ /cm². Column chromatography was performed on silica gel (200-300 mesh).

3. Oligonucleotide Structures, Synthesis, Purification, and Characterization.

(a) Chemical structures of SM-DNA (including CBM-DNA) and PC-DNA.



Figure S1: Representative structures of modified DNAs used in this study.

(b) DNA oligonucleotide synthesis and purification.

All DNA oligonucleotides were synthesized on an Applied Biosystems 394 DNA synthesizer using standard phosphoramidite protocols and purified by C18 reverse-phase HPLC with aqueous 0.1 M triethylammonium acetate (TEAA)/CH₃CN gradient on Agilent 1200 HPLC systems (Eclipse-XDB C18, 5 µM, 9.4 x 250 mm or 4.6 x 150 mm). After solid-phase synthesis, oligonucleotides were cleaved by an AMA cleavage solution (50 : 50; 40% aqueous ammonium hydroxide : methylamine; v: v) (55 °C, 55 min) over a dry bath. The cleaved oligonucleotide were then concentrated and neutralized (2.0 M TEAA) before HPLC purification (using a gradient of acetonitrile (5-80%) from 5 min to 25 min). For all phosphoramidites other than the four standard ones (dA-, dT-, dC-, and dG-phosphoramidites), coupling time was modified to 999 seconds. Non-standard phosphoramidites were either purchased or prepared in our laboratory as described below. Oligonucleotides with a 3'-amino group were synthesized using 3'-amino-modifier C7 CPG. Oligonucleotides were

quantitated by a BioTek Epoch UV-Vis spectrometer based on their calculated extinction coefficients at 260 nm. Oligonucleotides were characterized by either a Bruker APEX IV (for ESI–MS) or a Kratos PC Axima CFR plus V2.4.1 [for MALDI–MS, matrix: 8:1 (50 mg/mL 3-HPA or THAP in 1:1 water: acetonitrile) : (50 mg/mL ammonium citrate in water)] mass spectrometer. All DNA sequences are written in 5'- to 3'- orientation unless otherwise noted.

(c) Synthesis of SM-DNAs and PC-DNAs.

i) Dual-labeled SM-DNA with 5'-FAM and GLCBS:



5'-FAM labeling: Solid phase DNA synthesis was performed with the 3'-amino-modifier CPG. 5'-MMT was deprotected with 3% TCA in CH_2Cl_2 ; the beads were then washed with CH_2Cl_2 and then dried under vacuum. 5, 6-FAM (15.0 mg, 40 mmol) was dissolved in 200 µL anhydrous DMF along with 1 equiv. HBTU (15.2 mg, 40 mmol). After vortexing for 1 hour at room temperature, the activation mixture was added to the CPG along with 2.3 equiv. DIPEA (15.2 µL, 92 mmol). The suspension was then incubated at 37 °C overnight with agitation. The CPG was washed with DMF (3x 600 µL), CH₃CN (3x 600 µL), and then dried with gentle airstream. The 5'-FAM-labeled DNA was cleaved and purified by HPLC as described in the general method.

GLCBS-SE synthesis: GLCBS carboxylic acid (175 mg, 0.47 mmol) and NHS (54 mg, 0.47 mmol) were dissolved in 1.5 mL anhydrous DMF. *N*, *N*'-dicyclohexylcarbodiimide (115 mg, 0.56 mmol) was added to the solution at room temperature. After the mixture was stirred for 8 hours, precipitated urea was removed by filtration. The filtrate was dried by a rotary evaporator. After washing briefly with diethyl ether, the obtained white powder (desthiobiotin SE) was used directly without further characterization.

GLCBS labeling: The 3'-amine, 5'-FAM-modified DNA was dissolved in 40 μ L high purity water and 40 μ L phosphate buffer (pH = 7.2). GLCBS-SE (5.6 mg, 12 μ mol) was dissolved in 40 μ L DMSO (final concentration: 0.3 M). The two solutions were mixed and the reaction was incubated with sonication at 55 °C for 1 hour. After brief centrifugation, the reaction mixture was desalted by a NAP-5 column (GE Pharmacia) before the product was purified by HPLC as described in the general method.

ii) Dual-labeled SM-DNA with 5'-FAM and CBS:



The DNA synthesis and modification procedure was the same as described above in **i**) except CBS was used instead of the desthiobiotin.

iii) SM-DNA, SM = GLCBS.



GLCBS (4.5 mg, 12 µmol), DCC (2.5 mg, 12 µmol) and NHS (1.4 mg, 12 µmol) were dissolved in 40 µL anhydrous DMSO and then the solution was vortexed for 1 hour at room temperature. The formed urea precipitate was removed by filtration. The filtrate was added to a solution containing 40 µL 3'-amine-modified DNA (20 nmol) and 40 µL phosphate buffer (pH 7.5). The mixture was the maintained with sonication at 37 °C for 1 hour. After brief centrifugation, the reaction mixture was desalted by a NAP-5 column (GE Pharmacia) before the product was purified by HPLC.

iv) SM-DNA, SM = CBS.



The synthesis procedure was the same as described in iv) except CBS was used instead of GLCBS.

v) SM-DNA, SM = antipain.



Antipain (0.85 mg, 1.25 μ mol) in 12.5 μ L DMSO was added to 215 μ L DMSO, before sNHS (0.72 mg, 3.33 μ mol) in 10 μ L 2:1 DMSO : H₂O and EDC (0.24 mg, 1.2 μ mol) in 12 μ L DMSO were added. After the reaction mixture (final volume: 249.5 μ L) was vortexed at room temperature for 30 minutes, 3'-amine-modified DNA in triethylamine-HCl buffer (pH= 10.0, 50 μ L of a 500 mM stock solution) was added. The reaction was then vortexed at room temperature for another 12 hours. Tris-HCl (pH = 8.0, 20 μ L of a 500 mM stock solution) was added to quench the reaction (1 hour at room temperature). The products were purified by ethanol precipitation and reverse-phase HPLC.

vi) SM-DNA, SM = AP1497.



AP1497 was prepared following a report by Holt and co-workers in the carboxylic acid form,^{S1} which was further activated and conjugated to the 3'-amine-modified DNA with the same procedure as described in **vi**).

vii) SM-DNA, SM = AP1480.



AP1480 was prepared following a report by Holt and co-workers in the carboxylic acid form,¹ which was further activated and conjugated to the 3'-amine-modified DNA with the same procedure as described in **vi**).

viii) PC-DNA.



The synthesis and purification procedure is the same as described in **iv**) except that 4-azidobenzoic acid was activated to be conjugated to a 5'-amine-modified DNA.

4. DNA Oligonucleotide Sequences.

(a) SM-DNAs, CBM-DNA, and NH₂-DNA.

- Sequence of SM-DNAs in Figure 2: 5'-FAM-GGT TGA GCC ATT CTT AGT TTC-SM-3'
- Sequence of SM-DNAs for qPCR analysis in Figure 3 and 4: 5'-TGC CCA ATG TCT AGC TTG GTT GAG-SM-3'
- Sequence of negative control DNA for qPCR analysis in Figure 3: 5'-TGC CCA ATG TCT AGC TTG GTT GAG-3'
- CBM-DNA sequence in Figure 2: 5'-FAM-AAG GAG GGT TGA GCC ATT CTT AGT TTC-CBM-3'
- CBM-DNA sequence for qPCR analysis in Figure 4: 5'-TCG CCT ATC TTC CTC TTG GTT GAG-CBM-3'
- NH₂-DNA sequence in Figure 2: 5'-FAM-AAG GAG GGT TGA GCC ATT CTT AGT TTC-NH₂- 3'

(b) PC-DNA.

- Figure 2: 5'-azidophenyl-<u>GAA ACT AA</u>G CTT-3'
- Figure 3 and 4: 5'-azidophenyl-<u>CTC AAC CA</u>T CGA-3'

Note: only underlined bases are complementary to the SM-DNA.

(c) DNA sequences of the model library (Figure 5 of the main text).

• Positive controls:

- 5'-CCTGAATTCC<u>CATACGCACGATCAC</u>CCAAACTGCC-GLCBS-3'
- 5'-CCTGAATTCC<u>CATACGCACCTCTAG</u>CCAAACTGCC-CBS-3'

• Negative control:

5'-CCTGAATTCC<u>CATACGCACATGACG</u>CCAAACTGCC-CBM-3'

Background DNAs:

5'-CCTGAATTCC<u>NTANAGNACNATNGC</u>CCAAACTGCC-OH-3' (N=A/T/C/G)

• PC-DNAs:

5'-azidophenyl-GGCAGTTT-3'

Note: Colored sequences are barcodes. N represents variable base positions. There are five variable bases in the background DNAs (therefore $4^5 = 1,024$ different sequences).

(d) DNA sequences of the macrocycle library.

- Positive control GLCBS-DNA (encoding bases are underlined): 5'-CCTGAATTCC AAAGCC CTCACAATCCCCAC<u>TCACTC</u>ACTCCAAACTGCC-GLCBS-3'
- **PC-DNAs:** 5'-azidophenyl-GGCAGTTT-3'

• Macrocycle-encoding sequences: see Section 16.

(e) Other oligonucleotides.

• Loading control in Figure 2: 5'-FAM-CTT TCA GAC ATT CTT GAA CCT TCT CAC ATC TTG CTA TTC ACT TAC TGC TTA TCG TTA CTC ATT GTT-3'

5. Oligonucleotide Characterization.

Oligonucleotide	Expected mass (Da)	Observed mass (Da)	Note
5'-FAM-DNA-SM (GLCBS, 1; Figure 2)	7502.2	7502.1	MALDI
5'-FAM-DNA-SM (CBS, 2; Figure 2)	7333.1	7334.5	MALDI
5'-FAM-DNA-SM (CBM, 3; Figure 2)	9192.4	9193.7	MALDI
5'-DNA-SM (GLCBS; Figure 3 and 4)	7934.3	7934.9	MALDI
5'-DNA-SM (CBS; Figure 3 and 4)	7764.2	7764.4	MALDI
5'-DNA-SM (AP1497; Figure 3)	8146.6	8146.3	MALDI
5'-DNA-SM (antipain; Figure 3)	8167.7	8168.8	MALDI
5'-DNA-SM (AP1780; Figure 3)	8137.4	8138.6	MALDI
5'-DNA-SM (OH; Figure 3)	7371.0	7372.9	MALDI
5'-DNA-SM (CBM; Figure 4)	7601.2	7601.0	MALDI
5'-azidophenyl-DNA (Figure 2)	3963.7	3964.5	MALDI
5'-azidophenyl-DNA (Figure 3 and 4)	3884.6	3885.6	MALDI
5'-DNA-SM (GLCBS; Figure 5)	11107.0	11105.4	MALDI
5'-DNA-SM (CBS; Figure 5)	10927.9	10926.9	MALDI
5'-DNA-SM (CBM; Figure 5)	10911.9	10911.8	MALDI
5'-azidophenyl-DNA (Figure 5)	2751.5	2749.9	ESI

6. Selection Conditions and Procedures.

- (a) Fluorescence-based electrophoresis analysis (Figure 2 in the main text).
 - Selection sample preparation and photo-crosslinking procedure: In a typical experiment, 25 pmol of SM-DNA and 25 pmol of NH₂-DNA or CBM-DNA and 50 pmol PC-DNA were added to a 200 μL centrifuge tube, before 50 pmol of the corresponding protein target and 1.5 μL 10x PBS buffer were added. The mixture was supplemented with high purity water to a final volume of 15 μL, incubated at 4 °C for 1 hour, and then was subjected to irradiation under 365 nm at 0 °C (over ice) for 30 minutes.
 - ExoI-mediated digestion: After photo-crosslinking, the reaction mixture was transferred to a 600 μL centrifuge tube; then 2.4 μL 10x Exonuclease I buffer, 1 μL ExoI (20 units, 20 units/μL), and 5.6 μL high purity water were added to reach a final volume of 24 μL. The reaction mixture was incubated at 37 °C for 1 hour. After the digestion, the proteins were heat-denatured at 95 °C for 20 minutes on a dry bath.
 - Sample preparation for denaturing PAGE analysis: After heat-denaturation, the mixture was supplemented with 76 μL water, 1.5 μL glycogen (20 mg/mL), 10 μL NaOAc buffer (3M, pH 5.0), and 250 μL ethanol, along with 8 pmol internal DNA standard (a 66-nt, 5'-FAM-labeled DNA), for ethanol precipitation. The pellet obtained was dissolved in loading buffer directly for denaturing PAGE analysis (15%).
- (b) qPCR-based analysis (Figure 3 and 4 in the main text).
 - Sample preparation and photo-crosslinking procedure: The procedure and experiment conditions are the same as described above in (a).
 - *ExoI*-mediated digestion: The procedure and conditions are the same as described above in (a).
 - Sample preparation for PCR analysis: After heat-denaturation, the mixture was supplemented with 476 μL water. 100 μL of the solution was then removed and further diluted to 1000 μL. 2 μL of the final solution (1/2,500 of the original sample) was added with 1.8 μL PCR primers (5 μM) and 10 μL 2x Trans Top qPCR Mix. The solution was then again diluted to 20 μL for qPCR analysis.
- (c) Illumina[®] sequencing for the model library selection (Figure 5 in the main text).
 - Sample preparation and photo-crosslinking: In a typical experiment, 0.22 pmol of GLCBS-DNA/CBS-DNA/CBM-DNA, and 225.28 pmol of NH₂-DNA and 225.94 pmol PC-DNA were added to a 200 μL centrifuge tube, before 225.94 pmol of corresponding protein target and 6.78 μL 10x PBS buffer were added. The mixture was supplemented with high purity water to a final volume of 67.8 μL, incubated at 4 °C for 1 hour, and then subjected to irradiation under 365 nm at 0 °C (over ice) for 30 minutes.
 - *ExoI*-mediated digestion: After photo-crosslinking, 30 μL of the reaction mixture was transferred to a 600 μL centrifuge tube; then 4.8 μL 10x Exonuclease I buffer, 1 μL *ExoI* (20 units, 20 units/μL), and 12.2 μL high purity water

were added to reach a final volume of 48 μ L. The reaction mixture was incubated at 37 °C for 1 hour. After the digestion, the proteins were heat-denatured at 95 °C for 20 minutes on a dry bath.

- Sample preparation for PCR: After heat-denaturation, the mixture was supplemented with 452 μL water. 50 μL of the solution was then removed and further diluted to 1000 μL. 2 μL of the final solution (1/5000 of the original sample) was added with 5 μL PCR primers (5 μM), 10 μL 5X Phusion HF Reaction Buffer, 1 μL 1 mM Deoxynucleotide Solution Mix and 0.5 μL Phusion HF DNA Polymerase. The solution was then again diluted to 50 μL before PCR amplification. The PCR products were recovered by gel extraction and submitted for high throughput sequencing (see Section 9).
- **Pre-selection Library:** Pre-selection library was directly PCR amplified. PCR products were gel-purified and submitted for sequencing (see Section 9).

(d) Illumina[®] sequencing for macrocycle library selections (Figure 6 in the main text).

- Sample preparation and photo-crosslinking: The procedure and experiment conditions are the same as described above in (c).
- ExoI-mediated digestion: The procedure and experiment conditions are the same as described above in (c).
- Sample preparation for PCR: The procedure and experiment conditions are the same as described above in (c).
- Control selection in the absence of CA-II (Figure S11): The procedure is the same as described above except no CA-II was added to the selection mixture.
- Control selection with the blank library (Figure S12): The procedure is the same as described above except the macrocycle library was replaced with the same quantity of blank library (without conjugated small molecules).

7. CA-II Binding Assay of CBM-DNA.

GLCBS and CBS are well known inhibitors of CA-II.^{S2} Previously we have shown that DNA conjugation to these small molecules does not significantly alter their binding affinities against CA-II.^{S3} In this study, we used CBM, a structurally similar analog lacking the Zn^{2+} -binding sulfonamide motif, as the negative control for CA-II. Following a previously reported procedure,^{S4} we performed IC₅₀ assays with small molecule CBM and CBM-DNA against CA-II. The results show that both CBM and CBM-DNA do not inhibit CA-II. IC₅₀ values are known to well correlate with K_d values when competitive inhibition applies as employed in these inhibition assays; therefore the results show that CBM-DNA can be used as a non-binding negative control for target CA-II.

The esterase activity of CA-II is measured by spectrometric monitoring of the hydrolysis of substrate 4-nitrophenyl acetate. Experimentally, to a 100 μ L HEPES (50 mM, pH 7.4) buffer solution containing 1 μ M CA-II, we added 5 μ L inhibitor solutions with varied concentrations (for small molecule CBM: 0.01 μ M, 0.1 μ M, 1.0 μ M, 100 μ M, 1 mM, 10 mM; for CBM-DNA: 0.01 μ M, 0.1 μ M, 1.0 μ M, 10 μ M, 10



Figure S2: CA-II inhibition assays with small molecule CBM and CBM-DNA conjugate. a) IC_{50} measurements of CBM; b) IC_{50} measurements of CBM-DNA. Error bars (SD) are based on three replicates for each experiment.

8. qPCR Analysis and Calculation Methods.

(a) General qPCR conditions and materials.

All PCR reactions were carried out with 2x Trans Start Top Green qPCR Super Mix (9 µL and 20 µL reaction volume, Transgen Biotech) in the buffer provided. PCR reactions include Mg²⁺ (2 mM), dNTPs (200 mM), and primers (200 nM each). Templates were amplified from a standard initial concentration of 1nM unless otherwise specified. The thermal cycling sequence was set as follows: 95 °C for 2 minutes, then iterated cycles of 95 °C for 30 seconds, 50 °C for 30 seconds, and 68 °C for 15 seconds. The thermal cycling lasts 40 times unless otherwise noted. Quantitative PCR experiments were always performed in triplicates on a Bio-Rad CFX96 Real-Time PCR Detection System with SYBR Green as the detection dye.

(b) Quantitative PCR analysis and calculation methods.

We utilized the quantitative feature of qPCR (amplification cycle threshold value, C_T) to determine the initial copy number of amplifiable DNA templates in each PCR experiment. We installed orthogonal PCR primers on SM-DNAs and negative control DNAs respectively, so that one reaction mixture can be subjected to two parallel PCR amplifications with two sets of primers. Two sets of C_T values were measured and used to calculate initial copy numbers of surviving SM-DNA and negative control DNA. A comparison of C_T values obtained with different primer sets generates the ΔC_T (C_T^{neg} - C_T), which is used as a quantitative measure to reflect the protection of SM-DNAs in the presence of their specific protein targets. This type parallel C_T measurement was used for the experiments in Figure 4 of the main text.

First, in order to verify the analytical method performs well-behaved PCR amplification over a range of concentrations, we generated a standard plot by qPCR. Five different concentrations of the standard template (5 nM, 500 pM, 50 pM, 5pM, 500 fM, and 50 fM) were subjected to qPCR under standard conditions described above. The log of initial template concentration was plotted *vs*. the threshold cycle and a linear function was fitted to the data. A representative plot (Figure S3) and a C_T table (Table S1) are shown below. This type of validation and calibration has been employed throughout all qPCR experiments in this study.



Figure S3: Standard curve plot for validation of qPCR performance over a range of initial template concentrations.

From the above plot:

Equation: Y = -3.65X + 25.46 (r²=0.997; linear function);

 $Y = C_T$, X = log ([DNA]/[5 fM]) (in a 20 µL PCR system);

Therefore: X = (Y-25.46)/-3.65, then the DNA concentration can be calculated as:

 $[DNA] = [5 \text{ fM}] \times 10^{X} = [5 \text{ fM}] \times 10^{(Y-25.46)/-3.65} = [5 \text{ fM}] \times 10^{(C} \text{ }^{-25.46)/-3.65}$

DNA Template Concentration	С _{т1}	C _{T2}	C _{T3}	C _T (average)
5nM	3.7	4.4	4.0	4.0
500pM	7.7	7.5	6.0	7.1
50pM	10.1	10.6	10.3	10.3
5pM	14.6	14.6	14.8	14.7
500fM	18.6	17.9	17.5	18.0
50fM	22.9	21.7	21.8	22.1

Table S1: C_T measurements for the validation of qPCR performance over a range of template DNA concentrations.

Template sequence:

5'-TGCCCAATGTCTAGCTTGGTTGAG-3'

Primer sequences:

P1: 5'-TAGTCTGCCCAATGT-3' P2:5'-AATCGCTCAACCAAG-3'

(c) Orthogonality validation of PCR primer sets.

We first validated the performance as well as the orthogonality of these primer sets. As shown in Table S2 below, primer set #1 and #2 can efficiently amplify a control DNA (Template 1 with the primer binding site for set #1, entry 1) and SM-DNA (Template 2 with the primer binding site for set #2, entry 3) respectively. However, they cannot effectively cross-amplify the other DNA template having non-complementary primer binding site (entry 5 and entry 6). These primer sets were used for experiments in Figure 3 and 4 in the main text. (Note: the two primer sets share a same primer hybridizing at the 3'- of the template).

Entry	Template 1	Template 2	Primer set #1 1A & 2	Primer set #2 1B & 2	CT
1	+	-	+	-	7.8
2	-	-	+	-	23.6
3	-	+	-	+	7.5
4	-	-	-	+	32.5
5	+	-	-	+	29.5
6	-	+	+	-	21.8

Table S2: C_T measurements to validate the performance and orthogonality of the primer sets.

Template 1 sequences:

5'-TCGCCTATCTTCCTCTTGGTTGAG-3'

Primer set 1 sequence: P1A: 5'-TTCTGTCGCCTATCT-3' P2:5'-AATCGCTCAACCAAG-3'

Template 2 sequences:

5'-TGCCCAATGTCTAGCTTGGTTGAG-3'

Primer set 2 sequences: P1B: 5'-TAGTCTGCCCAATGT-3' P2:5'-AATCGCTCAACCAAG-3'

9. High-Throughput Sequencing (Illumina[®]) Procedures and Conditions.

- (a) Sample preparation, photo-crosslinking, *Exo*I-mediated digestion, and sample preparation for PCR: Same as described in Section 6c.
- (b) Gel extraction and product quantitation procedure: After PCR amplification, the products were recovered by gel extraction. The procedure was performed with the Gel Extraction Kit (CWBIO) according to the manufacturer's protocol.
- (c) Primer sequences for sequencing library construction: Forward Primer:

5'-AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTCCTGAATTCC-3'

Reverse Primers:

5'-CAAGCAGAAGACGGCATACGAGAT<u>GCTGTA</u>GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGGCAGTTTGG-3' 5'-CAAGCAGAAGACGGCATACGAGAT<u>TCGGGA</u>GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGGCAGTTTGG-3' 5'-CAAGCAGAAGACGGCATACGAGAT<u>GAATGA</u>GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGGCAGTTTGG-3' 5'-CAAGCAGAAGACGGCATACGAGAT<u>CTTCGA</u>GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGGCAGTTTGG-3'

Note: Italian fonts indicate the Illumina's sequencing adaptors, while the red fonts represent the primer binding sites for PCR amplification.

(d) High-throughput sequencing:

High-throughput sequencing was performed on an Illumina HiSeq 2500 sequencer using the standard 2x100bp paired-end sequencing procedure. For each sample we obtained at least 10^6 paired-end reads for accurate assessment of the enrichment fold changes.

10. Control Selection Experiments in Cell Lysates Only.



Figure S4: Selection of GLCBS-DNA in cell lysate without added protein target CA-II. a) Selection scheme. The procedure and conditions are the same as described in Figure 3 of the main text, except only HeLa or 293T cell lysates were used as the background (no CA-II added). b) Selection results analyzed by qPCR. $\Delta C_T = C_T^{neg} - C_T$. C_T values are averages from three replicate experiments. HeLa and 293T cell lysates: 1 mg/mL.

We performed control selection in lysate only (without added CA-II) and confirmed that the observed effect in the main text (Figure 3) is predominantly from the added CA-II, not from other endogenous cellular proteins (Figure S4).

Specific selection procedure and conditions:

- Sample preparation and photo-crosslinking: In a typical experiment, 25 pmol of GLCBS-DNA, 25 pmol CBM-DNA, and 50 pmol CP-DNA were added to a 200 μL centrifuge tube, before 4.83 μL 3 mg/mL cell lysate and 1.5 μL 10x PBS buffer were added. The mixture was supplemented with water to a final volume of 15 μL, incubated at 4 °C for 1 hour and then was subjected to irradiation under 365 nm at 0 °C (over ice) for 30 minutes.
- *Exo*I-mediated digestion: the procedure and conditions are the same as described in 6(b).
- Sample preparation for PCR: the procedure and conditions are the same as described in 6(b).

11. Effects of Different Chemical Modifications at the 3'- of SM-DNA on ExoI's Digestion Reactivity.



Figure S5: *Exo*I's digestion reactivities on various SM-DNA substrates were tested. SM-DNAs were directly digested by *Exo*I and analyzed by qPCR to determine C_T values. a) Reaction scheme; b) structures of SM-DNAs; c) qPCR results. Entry 1: native DNA with a 3'-OH group, which was used as the reference of the standard *Exo*I reactivity. Digestion and qPCR conditions are the same as described in Figure 3 of the main text. C_T values are averages from three replicate experiments.

We simply mixed individual SM-DNAs with *Exo*I for digestion at 37 °C and then used qPCR to evaluate the digestion efficiency (Figure S5). An unmodified native DNA was used as the reference (entry 1). The results are shown in Figure S5c. First, indeed there are variations from different 3'-chemical structures; however, C_T values of all SM-DNAs we tested are *close to or even larger* than the native DNA substrate ($C_T = 18.4$; entry 1), indicating they can be efficiently digested by *Exo*I despite having very different 3'- structures. Interestingly, for all the SM-DNA tested, only the 3'-NH₂-DNA appeared to be slightly resistant to *Exo*I digestion with a C_T smaller than native DNA ($C_T = 17.5$; entry 2).

12. Control qPCR Experiments of the Protection of GLCBS-DNA and CBM-DNA from *ExoI* Digestion under Different Conditions.



Figure S6: Analysis of the protection of GLCBS-DNA and CBM-DNA by qPCR. GLCBS-DNA or CBM-DNA was subjected to different experimental conditions as shown in the table above. Reaction conditions are the same as described in Figure 3. BSA: CA-II was replaced by non-specific protein BSA. C_T values are averages from three replicate experiments.

As shown in Figure S6 above, effective protection, indicated by small C_T values, took place only in the presence of specific SM-target interaction and PC-DNA-mediated target crosslinking. Other negative control experiments all showed C_T values close to the background.

13. Enrichment Fold Calculation Method Based on qPCR Analysis.

a) Validation of qPCR performance with orthogonal primer pairs over a series of dilutions.

First, we used standard samples to validate the performance of PCR amplification over a wide range of dilutions with the orthogonal primer pairs. As shown below in Table S3 and Figure S7, DNA-1 and DNA-2 were installed with primer set #1 and #2 respectively. These DNAs were mixed at various ratios and subjected to qPCR analysis. C_T 's are listed below in Table S3 (All values are the average of three replicate experiments). ΔC_T values ($\Delta C_T = C_{T2}-C_{T1}$) were calculated and plotted against dilution ratio (Figure S7). Excellent linear relationship is obtained, indicating qPCR performs well-behaved amplification with these primer pairs over a wide range of dilutions.

DNA-1/DNA-2	C _{T1} (primer set 1)	C _{T2} (primer set 2)	ΔCτ
1000:1	5.6	15.6	10.0
100:1	5.6	13.3	7.7
10:1	7.8	11.8	4.0
1:1	8.7	9.5	0.9
1:10	11.3	9.3	-2.0
1:100	12.8	6.9	-5.9
1:1000	15.6	6.7	-8.9

Table S3: Validation of qPCR performance over a range of mixing ratios of DNA-1 and DNA-2.



Figure S7: Plot of ΔC_T values against DNA-1/DNA-2 mixing ratios.

Calculation method:

From the above plot:

Equation: Y = 3.21X+0.83 (r^2 =0.997; linear function), Y = $\Delta C_T = C_{T2} - C_{T1}$, X = log ([DNA-1]/[DNA-2]) (in a 20 µL PCR reaction);

Therefore: X = (Y-0.83)/3.21, [DNA-1]/[DNA-2] = $10^{X} = 10^{(Y-0.83)/3.21} = 10^{(\Delta C_{T}^{-0.83)/3.21}}$; then the percentage of DNA-1 can be calculated as: [DNA-1]% = [DNA-1]/([DNA-1] + [DNA-2]) x 100%

b) Enrichment calculation.

The quantitative feature of qPCR enable us to use C_T values to calculate initial DNA copy numbers; and subsequently, the enrichment fold after each round of selection can be calculated by comparing the fraction percentage of SM-DNA before and after the selection. The results are shown in Figure 4 of the main text; detailed data and enrichment calculations are shown below.

First, we measured the C_T values with respective primer sets complementary to GLCBS- or CBS-DNA and CBM-DNA for the experiments with different mixing ratios. The ΔC_T 's ($C_T^{CBM} - C_T^{GLCBS}$ or $C_T^{CBM} - C_T^{CBS}$) were calculated to derive the percentages of GLCBS-DNA, CBS-DNA, and CBM-DNA in each experiment. (Note: in below, DNA-1 is GLCBS-DNA or CBS-DNA; DNA-2 is CBM-DNA)

	ratio	ΔC _T (1 st round)	ΔC _T (2 nd round)	GLCBS or CBS% (1 st round)	GLCBS or CBS% (2 nd round)
CLCDS.CDM	1:100	-1.8	1.2	13%	70%
GLCDS:CDIVI	1:1000	-6.0	-2.0	0.74%	12%
CDC+CDM	1:10	0.9	2.8	51%	80%
CD3:CBIVI	1:100	-3.7	-0.3	3.8%	31%

Table S4: Measurement of ΔC_T 's and corresponding percentages of SM-DNAs in the dilution experiments.

- For the GLCBS 1:100 experiment, after 1st round selection: $\Delta C_T = -1.8$ Therefore, [DNA-1]/[DNA-2] = $10^X = 10^{(Y-0.83)/3.21} = 10^{(\Delta C_T - 0.83)/3.21} = 10^{(-1.8 - 0.83)/3.21} = 1/6.6$ [DNA-1]% = [DNA-1]/([DNA-1] + [DNA-2]) x 100% = 1/(1+6.6) = 13%.
- For the GLCBS 1:100 experiment, after 2^{nd} round selection: $\Delta C_T = 1.2$ Therefore, [DNA-1]/[DNA-2] = $10^X = 10^{(Y-0.83)/3.21} = 10^{(\Delta C_T - 0.83)/3.21} = 10^{(1,2-0.83)/3.21} = 2.3$ [DNA-1]% = [DNA-1]/([DNA-1] + [DNA-2]) x 100% = 2.3/(2.3+1) = 70%
- For the GLCBS 1:1000 experiment, after 1st round selection: $\Delta C_T = -6.0$ Therefore, [DNA-1]/[DNA-2] = $10^X = 10^{(Y-0.83)/3.21} = 10^{(\Delta C_T - 0.83)/3.21} = 10^{(-6.0 - 0.83)/3.21} = 1/134.2$ [DNA-1]% = [DNA-1]/([DNA-1]+ [DNA-2]) x 100% = 1/(1+134.2) = 0.74%
- For the GLCBS 1:1000 experiment, after 2^{nd} round selection: $\Delta C_T = -2.0$ Therefore, [DNA-1]/[DNA-2] = $10^X = 10^{(Y-0.83)/3.21} = 10^{(\Delta C_T - 0.83)/3.21} = 10^{(-2.0 - 0.83)/3.21} = 1/7.6$ [DNA-1]% = [DNA-1]/([DNA-1] + [DNA-2]) x 100% = 1/(1+7.6) = 12%
- For the CBS 1:10 experiment, after 1st round selection: $\Delta C_T = 0.9$ Therefore, [DNA-1]/[DNA-2] = $10^X = 10^{(Y-0.83)/3.21} = 10^{(\Delta C_T - 0.83)/3.21} = 10^{(0.9 - 0.83)/3.21} = 1.05$

 $[DNA-1]\% = [DNA-1]/([DNA-1] + [DNA-2]) \times 100\% = 1.05/(1.05+1) = 51\%$

- For the CBS 1:10 experiment, after 2^{nd} round selection: $\Delta C_T = 2.8$ Therefore, [DNA-1]/[DNA-2] = $10^X = 10^{(Y-0.83)/3.21} = 10^{(\Delta C_T^{-0.83)/3.21}} = 10^{(2.8 \cdot 0.83)/3.21} = 4.1$ [DNA-1]% = [DNA-1]/([DNA-1] + [DNA-2]) x 100% = 4.1/(4.1+1) = 80%
- For the CBS 1:100 experiment, after 1st round selection: $\Delta C_T = -3.7$ Therefore, [DNA-1]/[DNA-2] = $10^X = 10^{(Y-0.83)/3.21} = 10^{(\Delta C_T - 0.83)/3.21} = 10^{(-3.7 - 0.83)/3.21} = 1/25.1$ [DNA-1]% = [DNA-1]/([DNA-1] + [DNA-2]) x 100% = 1/(1+25.1) = 3.8%
- For the CBS 1:100 experiment, after 2nd round selection: $\Delta C_T = -0.3$ Therefore, [DNA-1]/[DNA-2] = $10^X = 10^{(Y-0.83)/3.21} = 10^{(\Delta C_T - 0.83)/3.21} = 10^{(-0.3 - 0.83)/3.21} = 1/2.2$ [DNA-1]% = [DNA-1]/([DNA-1] + [DNA-2]) x 100% = 1/(1+2.2) = 31%



14. Selection Experiments with Immobilized CA-II.

Figure S8: qPCR analysis results of the selection of CBS-DNA/CBM-DNA (1:100) against immobilized CA-II. a) Selection scheme. qPCR with orthogonal primer pairs were used to analyze the selected SM-DNAs. b) qPCR analysis results. The two "no template" experiments were performed with primers only as the background. C_T values are averages from three replicate experiments.

We mixed CBS-DNA with non-binding CBM-DNA at the 1:100 ratio and selected this mixture against *immobilized* CA-II. After a selection against *immobilized* CA-II with a typical washing procedure,^{S5} either the specific binder CBS-DNA and the non-specific CBM-DNA was barely detectable with qPCR analysis as their C_T values are very close to the primer only background (Figure S8). We reason that, due to the relatively weak affinity of CBS-DNA to CA-II (K_d: 3.2 μ M),^{2b} target-bound CBS-DNA was nearly completely washed away. Although milder condition may be attempted, this comparison has shown the disadvantage of immobilization-based selections, especially for moderate and low affinity binders.

• Specific procedure and condition for selection against the immobilized CA-II:

NHS-activated SepharoseTM 4 Fast Flow (GE Healthcare) was prepared according to the manufacturer's protocol. CA-II

 C_{T} of primer set 2 only (no template): 30.2

from bovine erythrocytes was purchased from Sigma. Protein was dissolved in 0.2 M NaHCO₃ buffer (pH 8.3, 0.5 M NaCl), at a concentration of 200 μ M. Protein concentration was determined using UV-Vis spectrometry by measuring CA-II's extinction coefficient at 280 nM.

First, the protein solution (50 μ L, 10 nmol) was incubated with 10 μ L of the NHS-beads for 16 hours at 4 °C. Beads were capped for 4 hours with 0.1 M Tris-HCl buffer (1x 100 μ L; pH 8.5). After coupling, the beads were then washed with 0.1 M Tris-HCl buffer (3x 100 μ L; pH 8.5) and 0.1 M sodium acetate buffer in 0.5 M NaCl (3x 100 μ L; pH 4.9). This wash cycle was repeated 4 times. The beads was then stored in 20% ethanol. Before use, the beads were washed with 1x PBS twice before being incubated with the 1:100 CBS-DNA/CBM-DNA at 4 °C for 1hour. Then the beads were washed with the 1x PBS for 10 times.⁵ Finally the beads were heat-denatured in water at 95 °C for 20 minutes over a dry bath to elute any SM-DNAs bound to CA-II. The eluted solution was directly subjected to qPCR as described above.

15. Selection of the 1,207-Member Model Library.

The library selection procedure is described in Section 6c and Section 9. After Illumina sequencing, data were exported for processing. Sequence counts for each library member before and after the selection were tallied to calculate the enrichment fold for each library compound, following the method of previous reports.^{S6} First, the number of encoding DNA sequence reads for each library compound was counted. Second, the number was divided by the total number of interpretable sequence reads to give the abundance of that library member. Finally, enrichment folds for each library member were obtained by dividing the post-selection abundance by the pre-selection abundance. Enrichment folds were then plotted against the sequence number (see Figure 5 of the main text).

As expected, results show positive controls GLCBS and CBS were highly enriched after two rounds of selections; negative control CBM was only enriched 1.3-fold (after 1st round) and 9.1-fold (after 2nd round). There are some other background DNA sequences enriched to a very small degree (Figure 5 of the main text), possibly due to either weak DNA-protein binding. Nevertheless, the enrichment folds of these background DNAs are much smaller than GLCBS and CBS, and it also can be addressed by control selections in the absence of target CA-II.

16. Selection of the 4,801-MemberDNA-Encoded Library.

This library was prepared following the same procedure as reported by Liu and co-workers in synthesizing a DNAencoded macrocycle library except the small molecules are at the 3'-terminus of the DNA templates.^{S7} Amino acid building blocks used in our library are shown in Figure S9. DNA sequences of the macrocycle library are shown in Figure S10.

The library selection procedure is described in Section 6d and Section 9. Results are shown in Figure 6 of the main text. Similar to the model library selection, after the sequencing experiment, data were exported for processing. Sequence counts for each library member before and after the selection were tallied to calculate the enrichment fold for each library compound with the same method shown above in Section 15.

For the macrocycle library selection results, enrichment folds were plotted against post-selection sequence counts (see Figure 6 of the main text). Previous reports have shown that enrichment folds varied widely for low-count DNA sequences and only sequences with both high count post-selection and high enrichment fold should be considered as potential hits due to statistical under-sampling of low-count library compounds.^{S6b, c, 8}





DNA sequences of the macrocycle library:

• Macrocycle encoding sequences:

PBS-1	codon 1		codon 2		codon 3		PBS-2
	GTTCCC		GACCAC				
	TTGCAC		CTGCAC				CCAAACTGCC-3'-SM
	ACGAAC		CAGTCC				
	CATCGC		CTAGTC				
	CGTTCC		CAATGC		AGACCC		
	GTCAAC		ATACGC		ACTGTC		
	TACCAC		TTACCC		TCTAGC		
	ATGTCC		CCAAGC		TAGACC		
	GCCTAC		CGATCC		ACATCC		
5'-CCTGAATTCC	CGATTC	стслс	TACAGC	CAC	ATCACC	СТС	
J-CETGAATTEE	TCCATC		GTAACC	CAC	AACCTC	cic	
	GCAATC		AACGAC		CTGTAC		
	CCTAAC		ATCCCC		CCAGAC		
	CCGTTC		ACCTTC		TTCGAC		
	AGTCTC		CCTTTC		GCATAC		
	AAGCCC		ACGTAC		ATGCTC		
	AACAGC		CATGCC				
	TACGTC		CATCAC				
	ATCCGC		GCTAAC				
	CTGATC		TTCGCC				

> Sequences in purple are constant regions. Sequences in red, blue and green are encoding regions.



Figure S10: Control selection of the 4,801-member library against *immobilized* target CA-II. The same macrocycle library (with spiked-in GLCBS-DNA positive control) was subjected to the same incubation, wash, and elution procedures as described in Section 14 with immobilized CA-II. The eluted compounds were amplified and decoded as described above. a) Structure and composition of the library. b) Selection results are shown in a scattered plot (enrichment fold *vs.* sequence count). c) Zoom-in of the lower left portion of b). Positive control GLCBS-DNA is labeled in brown color along with its enrichment fold obtained. Note: no PC-DNA was added to the library.

The selection was performed with the same 4,801-member macrocycle library against the *immobilized CA-II* as a comparison with the selection against soluble CA-II shown in Figure 6 of the main text. As shown above, except the positive control GLCBS-DNA, no high-count sequences were significantly enriched.



Figure S11: Control selection of the 4,801-member library. The same macrocycle library (with spiked-in GLCBS-DNA positive control) was subjected to the same selection and decoding procedure as described in Figure 6 of the main text, except *no target protein CA-II* was added. a) Structure and composition of the library. b) Selection results are shown in a scattered plot (enrichment fold *vs.* sequence count). c) Zoom-in of the plot in b). Positive control GLCBS-DNA is labeled in brown color along with its enrichment fold obtained from the selection. Equal molar PC-DNA was added to the library before selection.

This selection was performed with the same macrocycle library but *without CA-II* in order to control for false positives arising from *Exo*I resistance due to 5'-conjugated small molecules or secondary DNA structures. As shown in Figure S11, no high-count sequences were significantly enriched. Most sequences with high counts were enriched less than 3 folds. As expected, in the absence of target CA-II, GLCBS-DNA was only enriched by 4.8-fold and the corresponding sequence count is very low, so that it cannot be considered a hit.



Figure S12: Control selection of a blank library. a) A blank library was prepared with the exact DNA sequence composition as the 4,801member macrocycle library but without the macrocycles. GLCBS-DNA was again added as the internal positive control. The blank library was subjected to the same selection and decoding procedure as described in Figure 6 of the main text. b) Selection results are shown in a scattered plot (enrichment fold *vs.* sequence count). c) Zoom-in of the plot in b). Positive control GLCBS-DNA is labeled in brown color along with its enrichment fold obtained from the selection. Equal molar PC-DNA was added to the library before selection.

This selection was performed against CA-II but with a blank library with the exact DNA sequence composition as the macrocycle library, in order to control for false positives arising from specific DNA-protein interactions. As shown in Figure S12, again no DNA sequence with high counts was enriched significantly.

17. References.

(S1) (a) Keenan, T.; Yaeger, D. R.; Courage, N. L.; Rollins, C. T.; Pavone, M. E.; Rivera, V. M.; Yang, W.; Guo, T.;
Amara, J. F.; Clackson, T.; Gilman, M.; Holt, D. A. *Bioorg. Med.Chem.* 1998, *6*, 1309; (b) Ong, S. E.; Schenone, M.; Margolin,
A. A.; Li, X.; Do, K.; Doud, M. K.; Mani, D. R.; Kuai, L.; Wang, X.; Wood, J. L.; Tolliday, N. J.; Koehler, A. N.; Marcaurelle,
L. A.; Golub, T. R.; Gould, R. J.; Schreiber, S. L.; Carr, S. A. *Proc. Natl. Acad. Sci. USA* 2009, *106*, 4617.

(S2) (a) Mincione, F.; Starnotti, M.; Menabuoni, L.; Scozzafava, A.; Casini, A.; Supuran, C. T. Bioorg. Med. Chem. Lett.

2001, 11, 1787; (b) West, G. M.; Tang, L.; Fitzgerald, M. C. Anal. Chem. 2008, 80, 4175; (c) McGregor, L. M.; Gorin, D. J.;

Dumelin, C. E.; Liu, D. R. J. Am. Chem. Soc. 2010, 132, 15522.

(S3) Li, G.; Liu, Y.; Chen, L.; Wu, S.; Li, X. Angew. Chem. Intl. Ed. 2013, 52, 9544.

(S4) Takaoka, Y.; Tsutsumi, H.; Kasagi, N.; Nakata, E.; Hamachi, I. J. Am. Chem. Soc. 2006, 128, 3273.

(S5) (a) Wei, X.; Lin, W.; Ma, N.; Luo, F.; Lin, Z.; Guo, L.; Qiu, B.; Chen, G. *Chem. Commun.* 2012, *48*, 6184; (b) Wu,
Z.; Wang, H.; Guo, M.; Tang, L. J.; Yu, R. Q.; Jiang, J. H. *Anal. Chem.* 2011, *83*, 3104; (c) Wu, Z.; Zhen, Z.; Jiang, J. H.; Shen,
G. L.; Yu, R. Q. *J. Am. Chem. Soc.* 2009, *131*, 12325; (d) Wei, X.; Zheng, L.; Luo, F.; Lin, Z.; Guo, L.; Qiu, B.; Chen, G. *J. Mater. Chem. B* 2013, *1*, 1812.

(S6) Li, Y.; Zhao, P.; Zhang, M.; Zhao, X.; Li, X. J. Am. Chem. Soc. 2013, 135, 17727.

(S7) (a) 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; (b) Kleiner, R. E.; Dumelin, C. E.; Tiu, G. C.; Sakurai, K.; Liu, D. R. J. Am. Chem. Soc. 2010, 132, 11779; (c) Mannocci,

L.; Zhang, Y. X.; Scheuermann, J.; Leimbacher, M.; De Bellis, G.; Rizzi, E.; Dumelin, C.; Melkko, S.; Neri, D. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 17670.

(S8) Tse, B. N.; Snyder, T. M.; Shen, Y. H.; Liu, D. R. J. Am. Chem. Soc. 2008, 130, 15611.

Full citations of the abbreviated references in the main text:

(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. Cheml Biol.* 2009, *5*, 647.
(5b) Disch, J. S.; Evindar, G.; Chiu, C. H.; Blum, C. A.; Dai, H.; Jin, L.; Schuman, E.; Lind, K. E.; Belyanskaya, S. L.;
Deng, J.; Coppo, F.; Aquilani, L.; Graybill, T. L.; Cuozzo, J. W.; Lavu, S.; Mao, C.; Vlasuk, G. P.; Perni, R. B. *J. Med. Chem.*

2013, 56, 3666.

(5c) Encinas, L.; O'Keefe, H.; Neu, M.; Remuiñán, M. J.; Patel, A. M.; Guardia, A.; Davie, C. P.; Pérez-Macías, N.; Yang
H.; Convery, M. A.; Messer, J. A.; Pérez-Herrán, E.; Centrella, P. A.; Alvarez-Gómez, D.; Clark, M. A.; Huss, S.; O'Donovan,

G. K.; Ortega-Muro, F.; McDowell, W.; Castañeda, P.; Arico-Muendel, C. C.; Pajk, S.; Rullás, J.; Angulo-Barturen, I.;
Alvarez-Ruíz, E.; Mendoza-Losana, A.; Ballell Pages, L.; Castro-Pichel, J.; Evindar, G. *J. Med. Chem.* 2014, *57*, 1276-1288.
(6c) J. Harris, D. E. Mason, J. Li, K. W. Burdick, B. J. Backes, T. Chen, A. Shipway, G. Van Heeke, L. Gough, A.
Ghaemmaghami, F. Shakib, F. Debaene, N. Winssinger, *Chem. Biol.* 2004, *11*, 1361-1372.
(8b) P. Blakskjaer; A. B. Christensen; N. J. V. Hansen; T. H. Hansen; J. Holmkvist; L. K. Larsen; L. K. Petersen; J.

Rasmussen-Dietvorst; F. A. Slok, Patent number: WO2012041633 A1, 2012.4.5.