Chemical Science

EDGE ARTICLE

View Article Online
View Journal | View Issue

Cite this: *Chem. Sci.*, 2015, **6**, 7097

Design, preparation, and selection of DNA-encoded dynamic libraries†

Gang Li,‡[a] Wenlu Zheng,‡[b] Zitian Chen,‡[c] Yu Zhou,[b] Yu Liu,[a] Junrui Yang,[c] Yanyi Huang[c] and Xiaoyu Li*[ab]

Received 8th July 2015
Accepted 9th September 2015

DOI: 10.1039/c5sc02467f

www.rsc.org/chemicalscience

We report a method for the preparation and selection of DNA-encoded dynamic libraries (DEDLs). The library is composed of two sets of DNA-linked small molecules that are under dynamic exchange through DNA hybridization. Addition of the protein target shifted the equilibrium, favouring the assembly of high affinity bivalent binders. Notably, we introduced a novel locking mechanism to stop the dynamic exchange and "freeze" the equilibrium, thereby enabling downstream hit isolation and decoding by PCR amplification and DNA sequencing. Our DEDL approach has circumvented the limitation of library size and realized the analysis and selection of large dynamic libraries. In addition, this method also eliminates the requirement for modified and immobilized target proteins.

Introduction

Dynamic combinatorial chemistry (DCC) employs reversible bond formation to create dynamic systems of continuous inter-exchanging chemical entities.[1–4] Built on the principle of DCC, dynamic combinatorial libraries (DCLs) have emerged as efficient tools for discovering novel ligands for biological targets.[5–8] Compared with a static library, a DCL has two advantages. First, a DCL allows for a spontaneous library synthesis based on the inter-conversion of compounds through reversible reactions among building blocks (BBs); the entire library can be synthesized by simply mixing the BBs without the need for spatial separation. Second, a DCL is adaptive: adding the target induces the selection pressure to redistribute the BBs, favouring the synthesis of target-binding compounds at the expense of non-binding ones.[9–12] Moreover, after reaching a new equilibrium in the presence of the target, the library can be "frozen" by stopping the dynamic exchange (*e.g.* by adding an additive or changing the pH to stop reversible reactions), so that the library population change is preserved and ready for subsequent hit identification.[1,6] DCLs have shown great potential in accelerating the discovery of lead compounds in drug discovery,[5,6,13,14] such as in fragment-based[15–18] and structure-based drug design.[5,19,20]

However, DCLs face a major limitation of low library diversity, mainly resulting from the lack of suitable analytical methods. Typically, chromatographic methods, such as HPLC, are used to resolve DCLs and to identify binders by comparing spectra with and without the target,[18,21–23] but HPLC does not have the capacity to resolve large libraries containing many different compounds.[16,24] Other methods, such as non-denaturing mass spectrometry,[25] NMR,[26] and spectroscopic methods (UV and fluorescence)[27–29] have been employed for DCLs, but the resolution and throughput of these methods are also not sufficient for large libraries. Otto, Miller, and their respective co-workers have developed several elegant approaches capable of analyzing and selecting large DCLs (~10 K compounds);[29–32] however, in most cases, DCLs only contain 10–100 compounds. Since the probability of discovering high affinity ligands increases with the library diversity, the limitation of the library size has presented a significant obstacle for DCLs.[23] New approaches capable of resolving and analyzing large DCLs are still highly desired.

A DNA-encoded library (DEL), in which each compound is linked with a unique DNA tag, is another combinatorial library approach employing mixed compounds in library processing.[33–42] In contrast to DCLs, due to DNA's high encoding capacity, DELs can contain millions of different compounds;[43–46] library selection can be feasibly decoded using PCR amplification and DNA sequencing.[47,48] Therefore, introducing DNA-encoding to DCLs could be an effective strategy to address the limitation of their library size. Previously, nucleic acids have been successfully used as programmable templates or scaffolds with spatial precision to display ligand

[a]*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, Beijing, China 100871. E-mail: xiaoyuli@pku.edu.cn*

[b]*Key Laboratory of Chemical Genomics, School of Chemical Biology and Biotechnology, Peking University Shenzhen Graduate School, Shenzhen, China 518055*

[c]*Biodynamic Optical Imaging Centre (BIOPIC) and College of Engineering, Peking University, Beijing, China 100871*

† Electronic supplementary information (ESI) available: Materials and general methods, experimental details, library selection and sequencing methods and fold enrichment calculations. See DOI: 10.1039/c5sc02467f

‡ These authors contributed equally to this work.

combinations interacting with various biological targets.[49–66] The Neri group developed a method named an Encoded Self-Assembling Combinatorial (ESAC) library, in which two sets of DNA-linked fragments form a static library by combinatorial duplex formation.[65,66] Hamilton and co-workers introduced dynamic exchange in DNA hybridization, so that the target can shift the equilibrium and enrich high affinity fragment combinations (Fig. 1a).[49,67] Very recently, Zhang and co-workers reported a similar system achieving target-induced enrichment of DNA duplexes.[68] These studies have nicely shown that the principle of dynamic exchange can be applied to DELs; however, more systematic methodology for the preparation and selection of DNA-encoded dynamic libraries (DEDLs) has yet to be developed. Moreover, previous studies require modified and immobilized targets in library selection, which is not compatible with proteins that are difficult to purify or modify, such as membrane proteins.[69,70] Aiming to address these issues, here we report the detailed study of a DEDL system, including library preparation, encoding, selection, hit deconvolution, and notably, a novel "locking" strategy to freeze the equilibrium shift for hit isolation and identification.

Results and discussion

Our strategy is shown in Fig. 1b. Libraries of BBs are conjugated to different DNA strands (ligand DNA), all having a common sequence that can form dynamically exchanging duplexes with an "anchor DNA", which is conjugated with an "anchor" molecule. Upon target addition, the equilibrium shifts to form more high affinity bivalent duplexes. Next, the photo-reactive group on the anchor DNA can crosslink the two DNA strands upon irradiation, thereby stopping the dynamic exchange and locking the shifted equilibrium. The distal region on the ligand DNA encodes the BB's chemical identity, and the crosslinked duplex can be isolated for hit identification with PCR amplification and DNA sequencing (Fig. 1b). By combining the features of DELs and DCLs, our design allows for the selection of high diversity DCLs to discover synergistic fragments for "affinity maturation" of the anchor molecule.[65,66,71,72]

We first verified that dynamic DNA duplex formation can be affected by the target protein.[49,68] As shown in Fig. 2a, a fluorescein (FAM) molecule and a quencher (DABCYL) were conjugated to two complementary DNA strands; the decrease of fluorescence therefore indicates DNA hybridization. The other end of the DNA was conjugated to a biotin, a desthiobiotin, or an iminobiotin molecule (Fig. 2b). These ligands are well known to bind to adjacent pockets on the tetrameric protein streptavidin (SA) with different affinities (K_d: 40 fM, 2.0 nM and 50 nM, respectively).[73] Moreover, we reason that, in order to establish

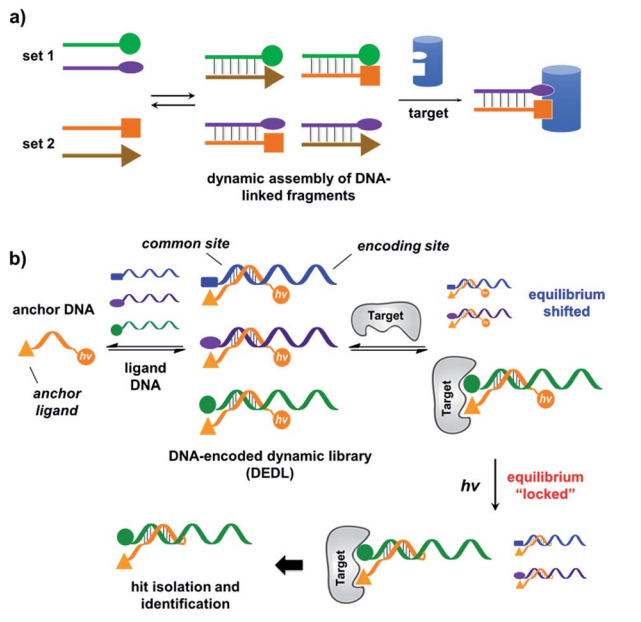


Fig. 1 (a) Previous work: two sets of DNA-linked fragments form dynamically exchanging duplexes; addition of the target enriches high affinity duplexes.[49,67,68] (b) DNA-encoded dynamic library (DEDL) (this work): an anchor DNA forms dynamic duplexes with multiple ligand DNA strands, forming the library. Adding the target shifts the equilibrium, favouring the formation of high affinity binders. A photo-crosslinker in the anchor DNA locks the shifted equilibrium under irradiation. Crosslinked binders can then be isolated for hit identification *via* PCR amplification and DNA sequencing.

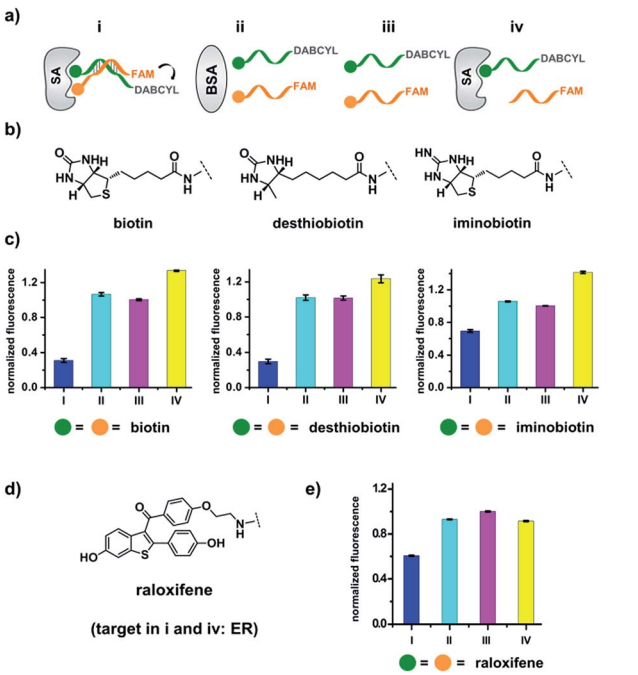


Fig. 2 (a) Two complementary DNA strands conjugated to biotin, desthiobiotin, or iminobiotin and FAM/DABCYL groups were mixed with the target SA (**i**), with BSA (**ii**), with no protein (**iii**), or with one ligand omitted (**iv**). (b) Structures of the small molecule ligands. (c) Fluorescence quenching results. FAM fluorescence values were measured and normalized to (**iii**). Left panel: with biotin; middle panel: with desthiobiotin; right panel: with iminobiotin. (d) Structure of the raloxifene ligand. (e) Fluorescence quenching results of the raloxifene–ER system. ER was used as the target in (**i**) and (**iv**). DNA: 200 nM each; protein: 400 nM. The DNA and protein were incubated at 30 °C for 1 h before measurement using a fluorophotometer. Excitation: 494 nm; emission: 522 nm. Error bars (standard deviation, SD) are based on three replicates of each experiment.

dynamic exchange, the DNA duplex should have a melting temperature (T_m) close to the experiment temperature, and it should also be sufficiently long to ensure hybridization specificity; therefore, either 6- or 7-base DNA duplexes were chosen in our study.

As shown in Fig. 2c, for all three ligands the fluorescence decreased significantly in the presence of SA, suggesting the formation of the ternary complex (**i**). In contrast, in control experiments with the non-specific protein BSA (bovine serum albumin) (**ii**), without SA (**iii**), or with one ligand omitted (**iv**), little or no fluorescence decrease was observed, indicating the quenching in (**i**) depends on specific bivalent binding to SA. Notably, ~40% quenching was observed for the weak binder iminobiotin (Fig. 2c, right panel). Furthermore, we performed similar fluorescence quenching experiments with raloxifene, an estrogen receptor (ER) modulator (Fig. 2d);[74] dimeric raloxifene ligands are able to bind to the two binding pockets on estrogen receptor dimers.[61,75,76] Similar to the biotin ligand series, a significant fluorescence decrease was observed in the presence of the specific target ER and the bivalent raloxifene duplex (Fig. 2e). In addition, as a thermodynamically-controlled system, an important feature of DCLs is that the same state of equilibrium can be reached from different starting points.[77,78] In order to verify this, we either altered the mixing order or incubated the mixture at 4 °C, 16 °C, 30 °C or 40 °C for 30 min before incubation at 30 °C for another hour (**QD** and **FD**; Fig. 3a). We observed that all experiments reached the same equilibrium based on fluorescence readings, proving the dynamic nature of our system (Fig. 3b).

Next, we investigated whether the target has shifted the equilibrium to promote the assembly of high affinity duplexes.

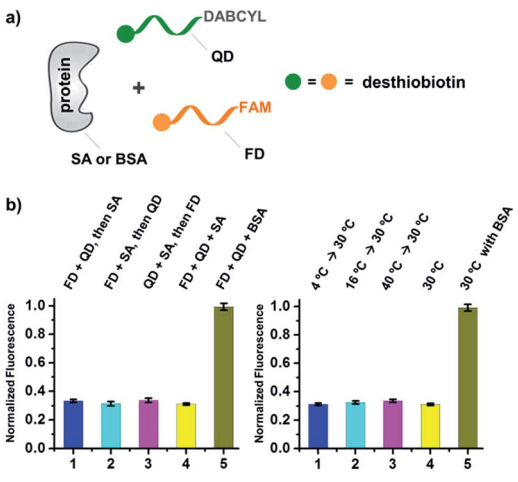


Fig. 3 (a) Two desthiobiotin-labelled DNA strands conjugated to FAM or DABCYL were mixed with SA or BSA in different orders or at different temperatures; fluorescence decreases were then measured. (b) Left panel: data from different mixing orders. Right panel: data from different temperatures; SA was used as the target except in 5 where BSA was used as a negative control. The experimental conditions were the same as those for Fig. 2 except for the mixing order and temperature. Error bars (standard deviation, SD) are based on three replicates of each experiment.

As shown in Fig. 4, we mixed a non-fluorescent background DNA (5′-GTCTGC-3′-NH_2; **BD-1**) with a fluorescent ligand DNA (5′-FAM-GTCTGC-3′-ligand; **LD-1**) at an 8 : 1 ratio. Both DNA strands dynamically compete for hybridization to **AD-1**, which is conjugated to an anchor ligand and a DABCYL quencher (5′-ligand-GCAGACT-3′-DABCYL). The bivalent **LD-1**/**AD-1** duplex is expected to have a higher affinity for SA than the monovalent **BD-1**/**AD-1** duplex. After mixing the DNA strands (**BD-1**/**LD-1**/**AD-1**: 8 : 1 : 1) with SA, we observed significant fluorescence quenching for all three ligands, indicating the equilibrium has been shifted to favour the formation of the (**LD-1**/**AD-1**)-SA ternary complex.

As a negative control, the non-binding BSA did not shift the equilibrium (cyan columns; Fig. 4). These results have demonstrated that the target can indeed promote the assembly of high affinity binders.

In the selection of DCLs, it is often necessary to stop the dynamic exchange and "freeze" the shifted equilibrium, so that the library population change, induced by the target, can be preserved for further characterization. For example, adding $NaBH_3CN$ to reduce imines to stable amines is a popular method to stop the dynamic imine formation,[21,22,79–81] and lowering the pH can effectively disable disulfide exchange and reversible Michael addition, which optimally occur at basic pH.[16,17,24,31,78] In this study, we designed a novel photo-crosslinking strategy to stop the dynamic DNA duplex exchange. Photo-crosslinking is kinetically fast and can be imposed/withdrawn conveniently with minimal perturbation to the system.[82] As shown in Fig. 5a, psoralen (PS), a photo-crosslinker widely used in nucleic acid crosslinking,[83–86] was conjugated to the 5′-end of a short 7-nt DNA bearing the anchor molecule (**AD-2**). **AD-2** is complementary to the 5′-end of a 24-nt DNA having a ligand and a FAM group (**LD-2**). Moreover, **LD-2** also contains a thymine group at the site opposite to PS, which is known to be able to improve the crosslinking efficiency.[87] After DNA incubation and target addition, irradiation triggers

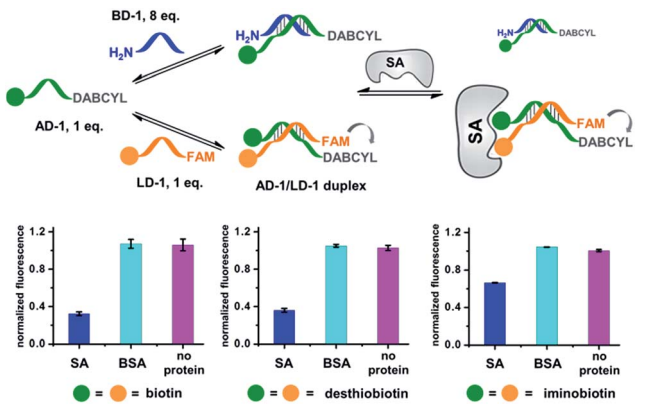


Fig. 4 Verification of the target-induced equilibrium shift, determined by a fluorescence decrease. Fluorescence values were normalized to the "no protein" experiment. AD-1 and LD-1: 200 nM; BD-1: 1.6 μM; proteins: 400 nM. The experimental procedures were the same as those for Fig. 2. Error bars (standard deviation, SD) are based on three replicates of each experiment.

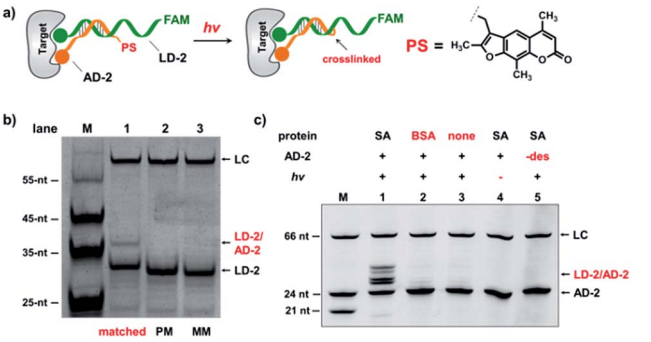
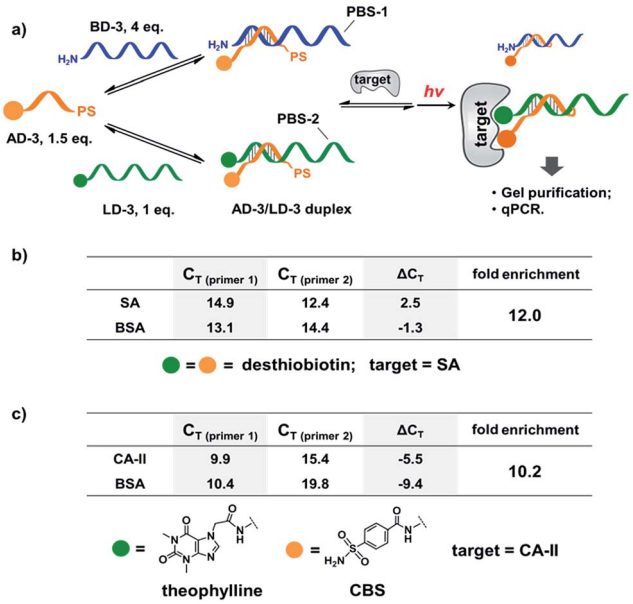


Fig. 5 (a) Psoralen (PS) was chosen to lock the equilibrium in the DEDL. (b) Three sets of **AD-2**/**LD-2** were mixed, irradiated, and analysed by denaturing electrophoresis. PM: partially mismatched (2-base mismatch); MM: fully mismatched; LC: a 66-nt DNA loading control. (c) 5′-Desthiobiotin-labeled **LD-2** and **AD-2** were mixed, irradiated under different conditions, and then analysed by denaturing electrophoresis (18% TBE–urea denaturing PAGE). Lane 1: with SA; lane 2: with BSA; lane 3: no protein added; lane 4: no irradiation; lane 5: no desthiobiotin on **AD-2**. **AD-2**: 300 nM; **LD-2**: 200 nM; irradiation: 365 nm for 30 s at 30 °C using a UV LED point light system; short irradiation reduces non-specific crosslinking in the background. M: marker; –des: no desthiobiotin.



	C_T (primer 1)	C_T (primer 2)	ΔC_T	fold enrichment
SA	14.9	12.4	2.5	12.0
BSA	13.1	14.4	-1.3	

	C_T (primer 1)	C_T (primer 2)	ΔC_T	fold enrichment
CA-II	9.9	15.4	-5.5	10.2
BSA	10.4	19.8	-9.4	

Fig. 6 (a) PS-based photo-crosslinking locks the shifted equilibrium for the subsequent hit isolation and qPCR analysis. After gel-purification of the crosslinked duplexes, qPCR was performed to determine the C_T values and to calculate the fold enrichment of the **AD-3**/**LD-3** duplex. (b) Results of the bivalent desthiobiotin–SA system. (c) Results of the theophylline/CBS–CA-II system. $\Delta C_T = C_{T\,(primer\,1)} - C_{T\,(primer\,2)}$. **AD-3**: 300 nM; **LD-3**: 200 nM; primers: 200 nM. The experimental procedure was the same as that for Fig. 5. See the ESI for details.†

crosslinking between **AD-2** and **LD-2**, thereby stopping strand exchange and locking the equilibrium. The crosslinked **AD-2**/**LD-2** duplex can then be isolated for PCR amplification and DNA sequencing to decode the ligand synergistically binding to the target with the anchor molecule. First, we prepared fully matched, partially mismatched, and fully mismatched **AD-2**/**LD-2** duplexes. These DNA duplexes were mixed, irradiated, and analysed by denaturing electrophoresis. The crosslinked product was only observed with the fully matched DNA duplexes (lane 1; Fig. 5b). Next, a set of desthiobiotin-labelled **AD-2** and **LD-2** strands was subjected to the same procedure; results show that only in the presence of SA was the crosslinking product detected (lane 1; Fig. 5c). Multiple bands appeared in lane 1 of Fig. 5c; mass analysis confirmed that all are crosslinked duplexes (see the ESI†). We hypothesize that the "T" shape of the crosslinked duplex may partially renature in the gel, a phenomenon that we have observed previously.[88] In all negative controls (with BSA, no protein, no irradiation and no desthiobiotin on **AD-2**; lanes 2–5, Fig. 5c), no or very little crosslinking was detected. The product bands were excised, extracted, and quantified. With SA, a 40% crosslinking yield was obtained. Collectively, these results have demonstrated the specificities of PS-based interstrand DNA crosslinking and its suitability for capturing target-induced duplex formation.

Next, we mixed a background DNA (5′-NH_2, 28-nt; **BD-3**) with a ligand DNA (5′-desthiobiotin, 28-nt; **LD-3**) at a 4 : 1 ratio. **BD-3** and **LD-3** have orthogonal primer binding sites (PBS-1 and PBS-2; Fig. 6a). Both **BD-3** and **LD-3** have a 7-base region complementary to a short anchor DNA (3′-desthiobiotin, 7-nt; **AD-3**). These DNA strands were mixed at a 4 : 1 : 1.5 ratio to form the dynamic library. After adding SA, the mixture was irradiated and the crosslinked duplexes were gel-purified for qPCR (quantitative PCR) analysis. The qPCR threshold cycle values (C_T's) were determined to calculate the initial copy numbers of **LD-3**/**AD-3** and **BD-3**/**AD-3** duplexes with their respective primers.[89,90] In order to offset possible biases from experimental factors, the library was also subjected to the same procedure (irradiation, gel purification, and qPCR) with the control protein BSA. Fold enrichments were then calculated by comparing the results from these two selections (see the ESI for the calculation method; Fig. S2–S4†). As a result, a 12.0-fold enrichment of the high affinity **LD-3**/**AD-3** duplex was achieved (Fig. 6b), which is comparable to typical DCL-based selections.[16,18,19,91,92] Gel analysis also directly confirmed the enrichment of the crosslinked **LD-3**/**AD-3** duplex (Fig. S5†). Moreover, in order to test the generality of our method, we conjugated another pair of ligands, theophylline and CBS to **LD-3** and **AD-3** DNA strands, respectively. Theophylline and CBS were found to synergistically bind the target of carbonic anhydrase-II (CA-II) in an ESAC library selection.[66] After mixing with the background DNA **BD-3**, the formed dynamic library was subjected to the selection against the target CA-II and the negative control BSA with the same procedure. The results show that a 10.2-fold enrichment of the **LD-3**/**AD-3** duplex was achieved (Fig. 6c). Collectively, these results have demonstrated that the PS-based crosslinking mechanism is suitable for locking and analysing the equilibrium shift in DEDL selections.

Encouraged by these results, we further prepared several model DCLs (Fig. 7a). These libraries contain a desthiobiotin-labelled ligand DNA (**LD-4**) and 4 background (**BD-4**) DNA strands, all dynamically competing for an anchor DNA (with

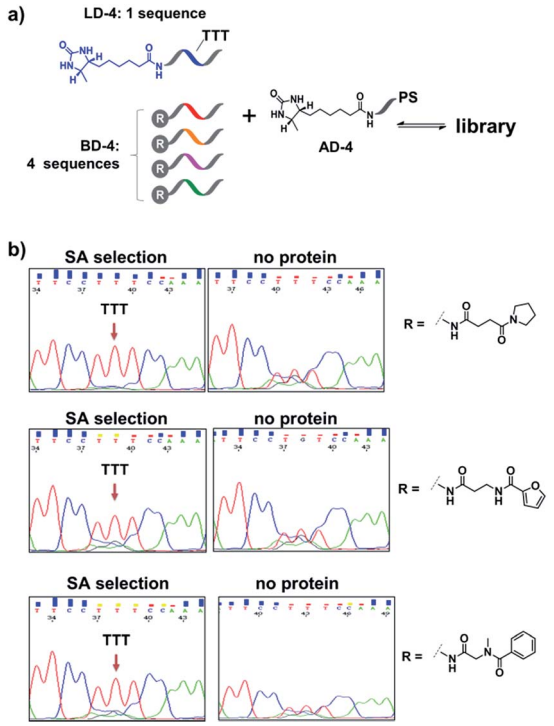


Fig. 7 (a) Components of the model libraries. Libraries were selected against SA with the same procedure as that for Fig. 6, except Sanger sequencing was used. (b) Sequencing results; left panels: after SA selection; right panels: control selection without SA. **LD-4**: 200 nM; **BD-4**: 800 nM (total); **AD-4**: 300 nM; SA: 400 nM. See the ESI for details.†

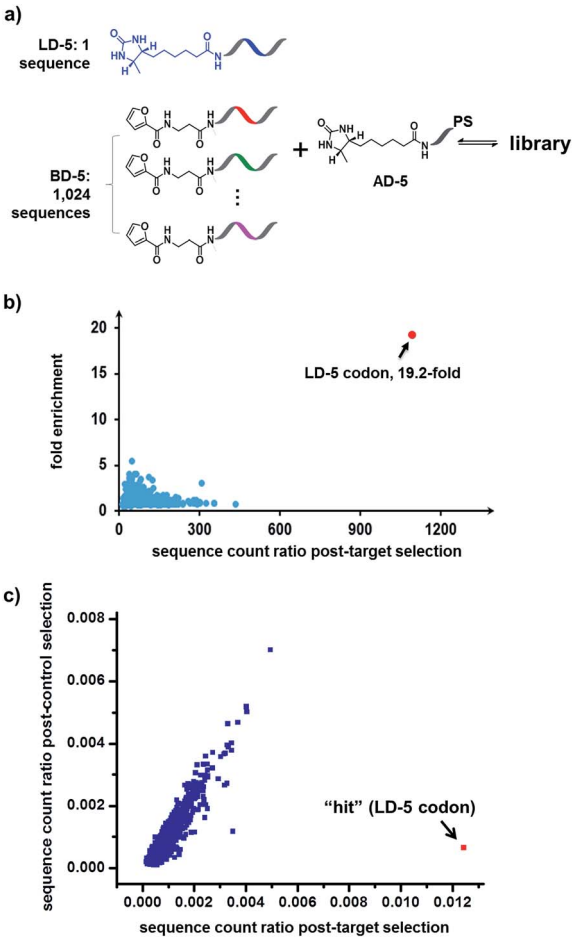


Fig. 8 (a) Components of the 1025-member library. The library was selected against SA with the same procedure as that for Fig. 6, except Illumina® sequencing was used. (b) Plot of the fold enrichment *versus* sequence count after the target selection. Fold enrichment = (post-target selection fraction)/(post-control selection fraction). (c) Plot of sequence count ratios after the control selection (no protein added) *versus* count ratios after the target selection (with SA). Sequence ratio = (sequence count)/(total sequence count of the library). Each dot represents the DNA sequence corresponding to a library member. The "hit" containing the desired **LD-5** codon is highlighted in red. **LD-5**: 0.19 nM; **BD-5**: 200 nM (total); **AD-5**: 300 nM; SA: 400 nM. The fold enrichments for the low-count library members vary widely due to statistical under-sampling. See the ESI† for more details on the experimental procedure, data analysis and further discussion of the sequencing results.

desthiobiotin, **AD-4**). The **BD-4** strands were also conjugated with several small molecules that are not known to bind SA, but represent typical fragment structures in a library. The ligand of desthiobiotin in **LD-4** is encoded by a "TTT" codon, while the **BD-4** strands contain varied sequences at the encoding site ("AAG", "GCA", "ACA" and "CGC"). These DNA strands were mixed at an equal ratio to form the library and then selected against SA. After irradiation, "hit compounds" were isolated and decoded with the same procedure as that for Fig. 6, except Sanger sequencing was used. As shown in Fig. 7b, in all cases, the "TTT" codon encoding the desthiobiotin in **LD-4** has been enriched markedly by SA due to the high affinity of the **LD-4**/**AD-4** duplex (left panels), whereas negative selections (no protein) only generated scrambled sequences at the encoding site (Fig. 7b, right panels).

Finally, in order to mimic library diversity, we prepared a model DCL containing 1024 background (**BD-5**) DNA strands, a ligand DNA (with desthiobiotin, **LD-5**), and an anchor DNA (with desthiobiotin, **AD-5**) (Fig. 8). The **LD-5** and all **BD-5** strands were mixed at an equal ratio, realizing a 1024-fold excess of background DNA strands relative to **LD-5**. This library was selected against the target SA and also subjected to a "no-protein" control selection, similar to that for Fig. 6, to control for biases from the selection procedures (irradiation, gel purification, PCR, sequencing, *etc.*). The selection results were decoded by high throughput DNA sequencing (Illumina®). The fold enrichments of selected sequences were plotted against the sequence counts to identify "hit compounds" (Fig. 8b). Again, due to the high affinity of the **LD-5**/**AD-5** duplex, the sequence that encodes **LD-5** was distinctly enriched (19.2-fold). In addition, the expected "hit", **LD-5**, shows a high sequence count ratio after the target selection, while having an average count ratio in the control selection, further confirming its target specificity (Fig. 8c and S6†). It is worth noting that the wide distribution of sequence counts in both the target and control selections indicates that sufficient sequencing depth and high library synthesis quality (even distribution of library members)[93] are both important in library selections. Although

this model library only has a limited chemical diversity, these results have demonstrated our approach's suitability for the selection of large dynamic libraries.

Conclusions

In conclusion, we have developed a DNA-encoded dynamic library (DEDL) approach for the preparation and selection of large dynamic libraries. Notably, we introduced a novel locking mechanism, which is able to take a "snapshot photo" of the library equilibrium altered by the target protein, thereby enabling the downstream hit isolation and identification. Second, our method eliminated the requirement of target immobilization and physical washing; therefore, target-induced perturbation of the library equilibrium is better preserved, and unmodified, non-immobilized proteins can be used as targets.[69,70,90]

However, the present method only encodes one fragment and thus is limited to the "affinity maturation" of known ligands (the "anchor"),[66,71,72] rendering it unsuitable for the *de novo* discovery of synergistic fragment combinations.[38] In contrast, nucleic acids have previously been successfully used as templates to pair DNA/PNA-linked small molecule ligands, therefore enabling the selection of synergistic fragment pairs for biological targets,[50,52,53,57–64] and the strategy of interstrand code-transfer also realized the dual-pharmacophore ESAC libraries.[65] These elegant studies highlight the importance of further development of dual-display DNA-encoded dynamic library,[8,94] which indeed is currently being pursued in our laboratory using an alternative DNA architecture, more efficient crosslinker,[95] and different decoding scheme.[96] We will report the results in due course.

Acknowledgements

This work was supported by Ministry of Science and Technology Basic Research Program (2011CB809100), NSFC (21272016, 21002003, 91013003) and the Doctoral Fund of Ministry of Education of China (20120001110083). We thank the mass spectrometry facility of National Centre for Protein Sciences at Peking University and Dr Rong Meng for the kind help with mass spectroscopy analysis.

Notes and references

1 P. T. Corbett, J. Leclaire, L. Vial, K. R. West, J.-L. Wietor, J. K. Sanders and S. Otto, *Chem. Rev.*, 2006, **106**, 3652–3711.
2 F. B. Cougnon and J. K. Sanders, *Acc. Chem. Res.*, 2011, **45**, 2211–2221.
3 Y. Jin, C. Yu, R. J. Denman and W. Zhang, *Chem. Soc. Rev.*, 2013, **42**, 6634–6654.
4 C. D. Meyer, C. S. Joiner and J. F. Stoddart, *Chem. Soc. Rev.*, 2007, **36**, 1705–1723.
5 M. Mondal and A. K. Hirsch, *Chem. Soc. Rev.*, 2015, **44**, 2455–2488.
6 O. Ramstrom and J. M. Lehn, *Nat. Rev. Drug Discovery*, 2002, **1**, 26–36.
7 S. Otto, R. L. Furlan and J. K. Sanders, *Curr. Opin. Chem. Biol.*, 2002, **6**, 321–327.
8 S. Otto, R. L. Furlan and J. K. Sanders, *Drug Discovery Today*, 2002, **7**, 117–125.
9 J.-M. Lehn, *Angew. Chem., Int. Ed.*, 2015, **54**, 3276–3289.
10 J. M. Lehn, *Chem. Soc. Rev.*, 2007, **36**, 151–160.
11 O. Ramstrom, T. Bunyapaiboonsri, S. Lohmann and J. M. Lehn, *Biochim. Biophys. Acta*, 2002, **1572**, 178–186.
12 B. L. Miller, *Dynamic combinatorial chemistry: in drug discovery, bioorganic chemistry, and materials science*, John Wiley & Sons, 2009.
13 K. R. West, K. D. Bake and S. Otto, *Org. Lett.*, 2005, **7**, 2615–2618.
14 K. R. West and S. Otto, *Curr. Drug Discovery Technol.*, 2005, **2**, 123–160.
15 M. F. Schmidt, A. Isidro-Llobet, M. Lisurek, A. El-Dahshan, J. Tan, R. Hilgenfeld and J. Rademann, *Angew. Chem., Int. Ed.*, 2008, **47**, 3275–3278.
16 B. Shi, R. Stevenson, D. J. Campopiano and M. F. Greaney, *J. Am. Chem. Soc.*, 2006, **128**, 8459–8467.
17 D. E. Scott, G. J. Dawes, M. Ando, C. Abell and A. Ciulli, *ChemBioChem*, 2009, **10**, 2772–2779.
18 V. T. Bhat, A. M. Caniard, T. Luksch, R. Brenk, D. J. Campopiano and M. F. Greaney, *Nat. Chem.*, 2010, **2**, 490–497.
19 M. Mondal, N. Radeva, H. Koster, A. Park, C. Potamitis, M. Zervou, G. Klebe and A. K. Hirsch, *Angew. Chem., Int. Ed.*, 2014, **53**, 3259–3263.
20 G. Klebe, *Nat. Rev. Drug Discovery*, 2015, **14**, 95–110.
21 I. Huc and J.-M. Lehn, *Proc. Nat. Acad. Sci. U. S. A.*, 1997, **94**, 2106–2110.
22 S. Zameo, B. Vauzeilles and J.-M. Beau, *Angew. Chem., Int. Ed.*, 2005, **44**, 965–969.
23 S. Ladame, *Org. Biomol. Chem.*, 2008, **6**, 219–226.
24 O. Ramström and J. M. Lehn, *ChemBioChem*, 2000, **1**, 41–48.
25 M. Demetriades, I. K. Leung, R. Chowdhury, M. C. Chan, M. A. McDonough, K. K. Yeoh, Y. M. Tian, T. D. Claridge, P. J. Ratcliffe and E. C. Woon, *Angew. Chem., Int. Ed.*, 2012, **51**, 6672–6675.
26 R. Larsson, Z. Pei and O. Ramstrom, *Angew. Chem., Int. Ed.*, 2004, **43**, 3716–3718.
27 S. M. Voshell, S. J. Lee and M. R. Gagne, *J. Am. Chem. Soc.*, 2006, **128**, 12422–12423.
28 A. Buryak and K. Severin, *Angew. Chem., Int. Ed.*, 2005, **44**, 7935–7938.
29 B. R. McNaughton and B. L. Miller, *Org. Lett.*, 2006, **8**, 1803–1806.
30 R. F. Ludlow and S. Otto, *J. Am. Chem. Soc.*, 2008, **130**, 12218–12219.
31 P. C. Gareiss, K. Sobczak, B. R. McNaughton, P. B. Palde, C. A. Thornton and B. L. Miller, *J. Am. Chem. Soc.*, 2008, **130**, 16254–16261.
32 B. R. McNaughton, P. C. Gareiss and B. L. Miller, *J. Am. Chem. Soc.*, 2007, **129**, 11306–11307.
33 R. M. Franzini, D. Neri and J. Scheuermann, *Acc. Chem. Res.*, 2014, **47**, 1247–1255.

34 R. E. Kleiner, C. E. Dumelin and D. R. Liu, *Chem. Soc. Rev.*, 2011, **40**, 5707–5717.
35 K. Gorska and N. Winssinger, *Angew. Chem., Int. Ed.*, 2013, **52**, 6820–6843.
36 L. Mannocci, M. Leimbacher, M. Wichert, J. Scheuermann and D. Neri, *Chem. Commun.*, 2011, **47**, 12747–12753.
37 K. Yea, J. Xie, H. Zhang, W. Zhang and R. A. Lerner, *Curr. Opin. Chem. Biol.*, 2015, **26**, 1–7.
38 J. Scheuermann and D. Neri, *Curr. Opin. Chem. Biol.*, 2015, **26**, 99–103.
39 A. I. Chan, L. M. McGregor and D. R. Liu, *Curr. Opin. Chem. Biol.*, 2015, **26**, 55–61.
40 P. Blakskjaer, T. Heitner and N. J. V. Hansen, *Curr. Opin. Chem. Biol.*, 2015, **26**, 62–71.
41 A. D. Keefe, M. A. Clark, C. D. Hupp, A. Litovchick and Y. Zhang, *Curr. Opin. Chem. Biol.*, 2015, **26**, 80–88.
42 G. Li, W. Zheng, Y. Liu and X. Li, *Curr. Opin. Chem. Biol.*, 2015, **26**, 25–33.
43 M. A. Clark, R. A. Acharya, C. C. Arico-Muendel, S. L. Belyanskaya, D. R. Benjamin, N. R. Carlson, P. A. Centrella, C. H. Chiu, S. P. Creaser, J. W. Cuozzo, C. P. Davie, Y. Ding, G. J. Franklin, K. D. Franzen, M. L. Gefter, S. P. Hale, N. J. V. Hansen, D. I. Israel, J. W. Jiang, M. J. Kavarana, M. S. Kelley, C. S. Kollmann, F. Li, K. Lind, S. Mataruse, P. F. Medeiros, J. A. Messer, P. Myers, H. O'Keefe, M. C. Oliff, C. E. Rise, A. L. Satz, S. R. Skinner, J. L. Svendsen, L. J. Tang, K. van Vloten, R. W. Wagner, G. Yao, B. G. Zhao and B. A. Morgan, *Nat. Chem. Biol.*, 2009, **5**, 647–654.
44 H. Deng, H. O'Keefe, C. P. Davie, K. E. Lind, R. A. Acharya, G. J. Franklin, J. Larkin, R. Matico, M. Neeb, M. M. Thompson, T. Lohr, J. W. Gross, P. A. Centrella, G. K. O'Donovan, K. L. Bedard, K. van Vloten, S. Mataruse, S. R. Skinner, S. L. Belyanskaya, T. Y. Carpenter, T. W. Shearer, M. A. Clark, J. W. Cuozzo, C. C. Arico-Muendel and B. A. Morgan, *J. Med. Chem.*, 2012, **55**, 7061–7079.
45 J. S. Disch, G. Evindar, C. H. Chiu, C. A. Blum, H. Dai, L. Jin, E. Schuman, K. E. Lind, S. L. Belyanskaya, J. Deng, F. Coppo, L. Aquilani, T. L. Graybill, J. W. Cuozzo, S. Lavu, C. Mao, G. P. Vlasuk and R. B. Perni, *J. Med. Chem.*, 2013, **56**, 3666–3679.
46 C. S. Kollmann, X. Bai, C.-H. Tsai, H. Yang, K. E. Lind, S. R. Skinner, Z. Zhu, D. I. Israel, J. W. Cuozzo, B. A. Morgan, K. Yuki, C. Xie, T. A. Springer, M. Shimaoka and G. Evindar, *Bioorg. Med. Chem.*, 2014, **22**, 2353–2365.
47 L. Mannocci, Y. X. Zhang, J. Scheuermann, M. Leimbacher, G. De Bellis, E. Rizzi, C. Dumelin, S. Melkko and D. Neri, *Proc. Nat. Acad. Sci. U. S. A.*, 2008, **105**, 17670–17675.
48 R. E. Kleiner, C. E. Dumelin, G. C. Tiu, K. Sakurai and D. R. Liu, *J. Am. Chem. Soc.*, 2010, **132**, 11779–11791.
49 K. I. Sprinz, D. M. Tagore and A. D. Hamilton, *Bioorg. Med. Chem. Lett.*, 2005, **15**, 3908–3911.
50 D. Sil, J. B. Lee, D. Luo, D. Holowka and B. Baird, *ACS Chem. Biol.*, 2007, **2**, 674–684.
51 D. M. Tagore, K. I. Sprinz, S. Fletcher, J. Jayawickramarajah and A. D. Hamilton, *Angew. Chem., Int. Ed.*, 2007, **46**, 223–225.
52 K. Gorska, K. T. Huang, O. Chaloin and N. Winssinger, *Angew. Chem., Int. Ed.*, 2009, **48**, 7695–7700.
53 B. A. Williams, C. W. Diehnelt, P. Belcher, M. Greving, N. W. Woodbury, S. A. Johnston and J. C. Chaput, *J. Am. Chem. Soc.*, 2009, **131**, 17233–17241.
54 D. Margulies and A. D. Hamilton, *Angew. Chem., Int. Ed.*, 2009, **48**, 1771–1774.
55 D. Margulies and A. D. Hamilton, *J. Am. Chem. Soc.*, 2009, **131**, 9142–9143.
56 K. Gorska, J. Beyrath, S. Fournel, G. Guichard and N. Winssinger, *Chem. Commun.*, 2010, **46**, 7742–7744.
57 C. Scheibe, A. Bujotzek, J. Dernedde, M. Weberb and O. Seitz, *Chem. Sci.*, 2011, **2**, 770–775.
58 M. Ciobanu, K. T. Huang, J. P. Daguer, S. Barluenga, O. Chaloin, E. Schaeffer, C. G. Mueller, D. A. Mitchell and N. Winssinger, *Chem. Commun.*, 2011, **47**, 9321–9323.
59 H. Eberhard, F. Diezmann and O. Seitz, *Angew. Chem., Int. Ed.*, 2011, **50**, 4146–4150.
60 J. P. Daguer, M. Ciobanu, S. Alvarez, S. Barluenga and N. Winssinger, *Chem. Sci.*, 2011, **2**, 625–632.
61 F. Abendroth, A. Bujotzek, M. Shan, R. Haag, M. Weber and O. Seitz, *Angew. Chem., Int. Ed.*, 2011, **50**, 8592–8596.
62 K. T. Huang, K. Gorska, S. Alvarez, S. Barluenga and N. Winssinger, *ChemBioChem*, 2011, **12**, 56–60.
63 E. A. Englund, D. Wang, H. Fujigaki, H. Sakai, C. M. Micklitsch, R. Ghirlando, G. Martin-Manso, M. L. Pendrak, D. D. Roberts, S. R. Durell and D. H. Appella, *Nat. Commun.*, 2012, **3**, 614.
64 J. P. Daguer, C. Zambaldo, M. Ciobanu, P. Morieux, S. Barluenga and N. Winssinger, *Chem. Sci.*, 2015, **6**, 739–744.
65 M. Wichert, N. Krall, W. Decurtins, R. M. Franzini, F. Pretto, P. Schneider, D. Neri and J. Scheuermann, *Nat. Chem.*, 2015, **7**, 241–249.
66 S. Melkko, J. Scheuermann, C. E. Dumelin and D. Neri, *Nat. Biotechnol.*, 2004, **22**, 568–574.
67 D. M. Tagore, K. I. Sprinz and A. D. Hamilton, *Supramol. Chem.*, 2007, **19**, 129–136.
68 F. V. Reddavide, W. Lin, S. Lehnert and Y. Zhang, *Angew. Chem., Int. Ed.*, 2015, **54**, 7924–7928.
69 L. M. McGregor, T. Jain and D. R. Liu, *J. Am. Chem. Soc.*, 2014, **136**, 3264–3270.
70 P. Zhao, Z. Chen, Y. Li, D. Sun, Y. Gao, Y. Huang and X. Li, *Angew. Chem., Int. Ed. Engl.*, 2014, **53**, 10056–10059.
71 S. Melkko, Y. Zhang, C. E. Dumelin, J. Scheuermann and D. Neri, *Angew. Chem., Int. Ed.*, 2007, **46**, 4671–4674.
72 J. Scheuermann, C. E. Dumelin, S. Melkko, Y. Zhang, L. Mannocci, M. Jaggi, J. Sobek and D. Neri, *Bioconjugate Chem.*, 2008, **19**, 778–785.
73 N. Green, *Biochem. J.*, 1966, **101**, 774–780.
74 C. D. Jones, M. G. Jevnikar, A. J. Pike, M. K. Peters, L. J. Black, A. R. Thompson, J. F. Falcone and J. A. Clemens, *J. Med. Chem.*, 1984, **27**, 1057–1066.

75 M. Shan, A. Bujotzek, F. Abendroth, A. Wellner, R. Gust, O. Seitz, M. Weber and R. Haag, *ChemBioChem*, 2011, **12**, 2587–2598.
76 K. K. Sadhu, T. Eierhoff, W. Romer and N. Winssinger, *J. Am. Chem. Soc.*, 2012, **134**, 20013–20016.
77 B. Shi and M. F. Greaney, *Chem. Commun.*, 2005, **7**, 886–888.
78 S. Ladame, A. M. Whitney and S. Balasubramanian, *Angew. Chem., Int. Ed.*, 2005, **44**, 5736–5739.
79 M. Hochgürtel, R. Biesinger, H. Kroth, D. Piecha, M. W. Hofmann, S. Krause, O. Schaaf, C. Nicolau and A. V. Eliseev, *J. Med. Chem.*, 2003, **46**, 356–358.
80 J. C. Leitzel and D. G. Lynn, *Chem. Rec.*, 2001, **1**, 53–62.
81 P. Luo, J. C. Leitzel, Z.-Y. J. Zhan and D. G. Lynn, *J. Am. Chem. Soc.*, 1998, **120**, 3019–3031.
82 G. W. Preston and A. J. Wilson, *Chem. Soc. Rev.*, 2013, **42**, 3289–3301.
83 T. Stafforst and D. Hilvert, *Angew. Chem., Int. Ed.*, 2011, **50**, 9483–9486.
84 H. Baigude, Z. Li, Y. Zhou and T. M. Rana, *Angew. Chem., Int. Ed.*, 2012, **51**, 5880–5883.
85 J. E. Hearst, *Annu. Rev. Biophys. Bioeng.*, 1981, **10**, 69–86.
86 U. Pieles and U. Englisch, *Nucleic Acids Res.*, 1989, **17**, 285–299.
87 H. Baigude, Ahsanullah, Z. Li, Y. Zhou and T. M. Rana, *Angew. Chem., Int. Ed.*, 2012, **51**, 5880–5883.
88 G. Li, Y. Liu, L. Chen, S. Wu and X. Li, *Angew. Chem., Int. Ed.*, 2013, **52**, 9544–9549.
89 D. J. Gorin, A. S. Kamlet and D. R. Liu, *J. Am. Chem. Soc.*, 2009, **131**, 9189–9191.
90 L. M. McGregor, D. J. Gorin, C. E. Dumelin and D. R. Liu, *J. Am. Chem. Soc.*, 2010, **132**, 15522–15524.
91 M. Demetriades, I. K. H. Leung, R. Chowdhury, M. C. Chan, M. A. McDonough, K. K. Yeoh, Y.-M. Tian, T. D. W. Claridge, P. J. Ratcliffe, E. C. Y. Woon and C. J. Schofield, *Angew. Chem., Int. Ed.*, 2012, **51**, 6672–6675.
92 D. E. Scott, G. J. Dawes, M. Ando, C. Abell and A. Ciulli, *ChemBioChem*, 2009, **10**, 2772–2779.
93 A. L. Satz, *ACS Chem. Biol.*, 2015, DOI: 10.1021/acschem bio.5 b00378.
94 V. Goral, M. I. Nelen, A. V. Eliseev and J. M. Lehn, *Proc. Nat. Acad. Sci. U. S. A.*, 2001, **98**, 1347–1352.
95 H. Kashida, T. Doi, T. Sakakibara, T. Hayashi and H. Asanuma, *J. Am. Chem. Soc.*, 2013, **135**, 7960–7966.
96 A. Litovchick, M. A. Clark and A. D. Keefe, *Artif. DNA PNA. XNA*, 2014, **5**, e27896.

Supporting Information

Design, Preparation, and Selection of DNA-encoded Dynamic Libraries

Gang Li,[a,‡] Wenlu Zheng,[b,‡] Zitian Chen,[c,‡] Yu Zhou,[b] Yu Liu,[a] Junrui Yang,[c] Yanyi Huang,[c]

Xiaoyu Li*[,a,b]

[a] *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*
[b] *Key Laboratory of Chemical Genomics, School of Chemical Biology and Biotechnology, Peking University Shenzhen Graduate School, Shenzhen, China 518055*
[c] *Biodynamic Optical Imaging Centre (BIOPIC) and College of Engineering, Peking University, Beijing, China 100871*
[‡] *These authors contributed equally to this work.*

E-mail: xiaoyuli@pku.edu.cn

Table of Contents

1. Abbreviations.

AMA: 1:1 (v:v) aqueous methylamine (40% wt.) : aqueous ammonium hydroxide (30% wt.)

BSA: bovine serum albumin

CA-II: carbonic anhydrase-II

CBS: 4-sulfamoylbenzoic acid

CPG: controlled-pore glass

DABCYL: (E)-4-((4-(dimethylamino)phenyl)diazenyl)benzoic acid

DB: desthiobiotin

DCC: *N*, *N*'-dicyclohexyl carbodiimide

DCM: dichloromethane

DIPEA: *N*, *N*'-diisopropylethylamine

DMF: *N*, *N*'-dimethyl formamide

DMSO: dimethyl sulfoxide

DMT: di-(4-methoxyphenyl)phenylmethyl

EDC: 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide hydrochloride

FAM: carboxyfluorescein

Fmoc: 9-fluorenylmethyloxycarbonyl

HBTU: *O*-benzotriazol-1-yl-*N*, *N*, *N*', *N*'-tetramethyluroniumhexafluorophosphate

HOBt: *N*-hydroxylbenzotriazole

3-HPA: 3-hydroxybenzotriazole

MMT: 4-methoxytrityl

NHS: *N*-hydroxysuccinimide

PAGE: polyacrylamide gel electrophoresis

PBS: phosphate buffered saline

PC: photo-crosslinker

PS: psoralen

SA: streptavidin

SE: succinimidyl ester

SM: small molecule

TBE: tris-borate-EDTA

TCA: trichloroacetic acid

TEA: triethylamine

TEAA: triethylammonium acetate

2. Materials and general methods.

Unless otherwise noted, all reagents and solvents were purchased from commercial sources and used as received. Streptavidin and BSA were purchased from Sangon Biotech; CA-II was purchased from Sigma-Aldrich; Estrogen receptor (ER) alpha was purchased from ThermoFisher Scientific. Biotin, desthiobiotin and iminobiotin were purchased from Sigma-Aldrich. Theophylline-7-acetic acid was purchased from Alfa Aesar. 4-sulfamoylbenzoic acid (CBS) was purchased from J&K Scientific Ltd. Fluorescent dyes were purchased from Tianjin Heowns Biochemical Technology Co. Ltd. (DABCYL) and Beijing Fanbo Biochemicals (5, 6-FAM and 6-FAM-SE). Other common chemicals and reagents were purchased from Beijing Ouhe Technology Co. Ltd. and J&K Scientific Ltd. qPCR Supermix was purchased from Beijing Transgen Biotech. Other PCR reagents and agarose gel loading markers were purchased from Beijing Liuhe Tong Trade Co. Ltd. Water was purified with a Thermo Scientific Barnstead Nanopure system. Oligonucleotides were synthesized on standard CPG (1000 Å) beads by an automated Applied Biosystems 394 synthesizer following instrument's built-in 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 gel or native TBE PAGE gel (18% or 20%). Samples for Illumina sequencing were analyzed by agarose gel electrophoresis, stained with ethidium bromide. All gel images were captured by a Bio-Rad Chemidoc system. Photo-crosslinking experiments were conducted by a UVP CL-1000L Ultraviolet crosslinker at 365 nm wavelength with an intensity of approximately 100 μJ /cm^2 or Uvata UV LED point light system. Concentrations of oligonucleotides were determined by the absorbance and extinction coefficient at 260 nm. Concentrations of proteins were determined by the absorbance and extinction coefficient at 280 nm.

3. Experiment conditions for the assembly and analysis of DNA-encoded dynamic library.

a) FRET experiments:

- In Figure 2 and Figure 3, fluorescent DNA (5'-FAM-DNA-ligand-3', 200 nM) or the control DNA (5'-FAM-DNA-NH_2-3', 200 nM), and quenching DNA (5'-ligand-DNA-DABCYL-3', 200 nM) were mixed with respective protein (SA, ER or BSA; 400 nM) in a buffer of 0.1M TAPS/0.1 M NaCl (pH = 9.0). The total solution volume was 200 μL. Samples were then incubated at 30 ºC for 1 h before fluorescence measurement with temperature control. Excitation wavelength: 494 nm; detection wavelength: 522 nm.
- In Figure 4, **LD-1** (200 nM), **AD-**1 (200 nM), and **BD-1** (1.6 μM) were mixed with SA (400 nM), BSA (400 nM), or water (no protein control) in a buffer of 0.1M TAPS/0.1M NaCl (pH = 9.0) to reach a total volume of 200 μL. Samples were incubated at 30 ºC for 1 h before fluorescence measurement with temperature control. Excitation wavelength: 494 nm; detection wavelength: 522 nm.

b) Photo-crosslinking experiments:

- In Figure 5, 6 and 7: DNA strands were mixed in a buffer of 0.1M TAPS/0.1M NaCl (pH = 9.0) at the concentrations as specified in the figure. Samples were heated at 50 ºC briefly before slowing cooling to room temperature before protein addition. The total volume of incubation mixture was 200 μL. Samples were incubated at 30 ºC for 1 h before light irradiation at 365 nm for 10 min by a UVP CL-1000L Ultraviolet crosslinker or 30s by a Uvata UV LED point light source. 20 μL 1% SDS was added and protein was denatured at 90 ºC for 10 min over a dry bath. A 66-nt fluorescent DNA was added to all solutions as the internal fluorescence control. After ethanol precipitation, recovered samples were resolved by 18% TBE-Urea denature PAGE gel (Figure 5) or 20% TBE native PAGE gel (Figure 6, 7). Product bands were excised, extracted by TE buffer, and purified by ethanol precipitation. Purified samples were directly subjected to MS characterization, PCR and/or sequencing.
- In Figure 8: experiment conditions and procedures are the same as described above except that samples were incubated at 16 ºC and 22% TBE native PAGE gel was used. The isolated crosslinking products were subjected to PCR with long Illumina primers; PCR products were resolved by 3% agarose gel and recovered by Axygen plasmid recovery kit. The products

were quantified by Nanodrop and then submitted to BIOPIC's high-throughput sequencing platform at Peking University for Illumina sequencing.

4. Modified DNA structures, sequences, and characterization.

a) General.

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 TEAA/CH_3CN gradient on Agilent 1200 HPLC systems (Eclipse-XDB C18, 5 μM, 9.4 x 200 mm or 4.6 x 150 mm). After solid-phase synthesis, oligonucleotides on CPG were cleaved by AMA (55 ºC, 55 min) over a dry bath, then concentrated, neutralized (2.0 M TEAA) before HPLC purification. For all non-standard phosphoramidites (other than standard dA-, dT-, dC- and dG-phosphoramidites), coupling time was modified to 999 seconds. Non-standard phosphoramidites were purchased from Glen Research and Wuhu Huaren. Oligonucleotides were quantitated by a BioTek Epoch UV-Vis spectrometer based on extinction coefficient at 260 nm. Oligonucleotides were characterized by either a Bruker APEX IV (for ESI-MS) or a Bruker ultrafleXtreme [for MALDI-MS, matrix: 9:1 (50 mg/mL 3-HPA 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.

b) Chemical structures of DNAs.

We used the standard 3'-amino-modifier C7 CPG and the 5'-amino-modifier 5 in DNA synthesis to introduce amino functional groups at the ends of DNA and various modifications (FAM, DABCYL, small molecule ligand, etc.) were coupled to DNA via these amine groups, either in solution or directly on solid phase. The PS group was introduced by using the commercially available psoralen-C2-phosphoramidite. Detailed preparations are described in Section 5. Representative chemical structures of modified DNAs are shown below:

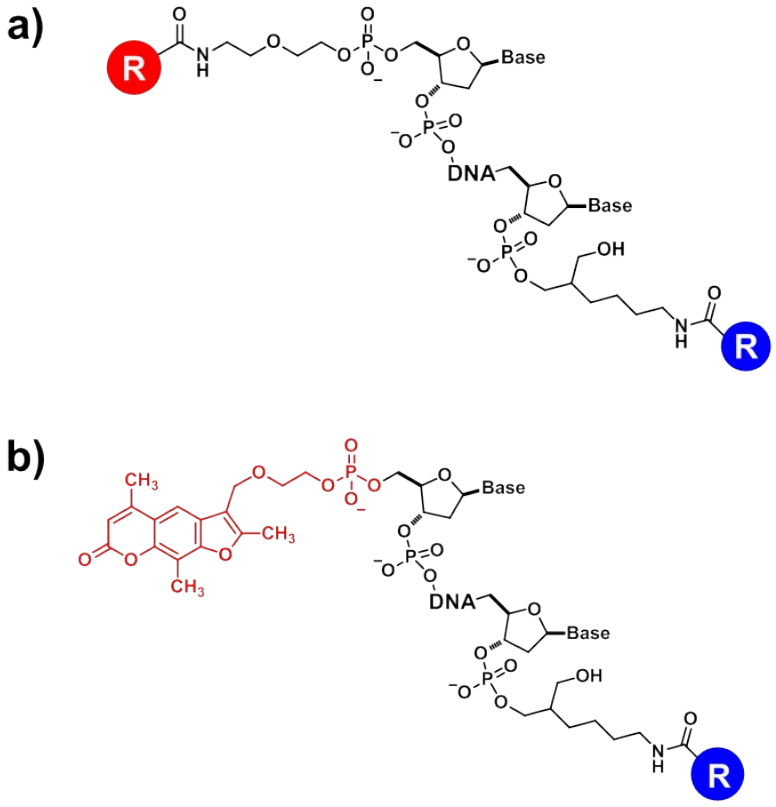


Figure S1: Representative structures of modified DNAs: a) modification via 5'- and 3'- amino modifiers; b) 5'-psoralen-modified DNAs.

c) DNA sequences.

- **Sequences in Figure 2, 3, 4:**

  5'-FAM-GTC TGC-3'-ligand (**FD**, **LD-1**)

  5'-ligand-GCA GAC T-3'-DABCYL (**QD**, **AD-1**)

  5'-GTC TGC-NH_2-3' **(BD-1**)

- **Sequences in Figure 5:**

  AD-2: 5'-PS-TAT CAA C-ligand-3'

  Fully matched **LD-2** (Figure 5b): 5'-GTT GAT ATT TTC CCG AAA GTA GAT AGG CGA-3'

Partially mismatched **LD-2** (Figure 5b): 5'-GTA GAC ATT TTC CCG AAA GTA GAT AGG CGA-3'

Fully mismatched **LD-2** (Figure 5b): 5'-CAA ATG TTT TTC CCG AAA GTA GAT AGG CGA-3'

LD-2 (Figure 5c): 5'-ligand-GTT GAT ATC CGC AGA TAG GCG ACA-FAM-3'

66-nt DNA loading control: 5'-CTT TCA GAC ATT CTT GAA CCT TCT CAC ATC TTG CTA TTC ACT TAC TGC TTA TCG TTA CTC ATT GTT-3'

- **Sequences in Figure 6:**

AD-3: 5'-PS-TAT CAA C-ligand-3'

LD-3: 5'-ligand-GTT GAT ATT CCC GAA AGT AGA TAG GCG A-3' (for desthiobiotin)

LD-3: 5'-ligand-GTT GAT ATT CCC GAA AGT AGA TAG GCG T-3' (for CBS)

BD-3: 5'-NH_2-GTT GAT ATT CCC GAA AGT ACA TTG GGC A-3'

Forward Primer: 5'-TAG AGT CCC GAA AGT-3'

Reverse Primer for **LD-3**: 5'-TTC TGT CGC CTA TCT-3'

Reverse Primer for **BD-3**: 5'-TAG TCT GCC CAA TGT-3'

- **Sequences in figure 7:**

AD-4: 5'-PS-TAT CAA C-ligand-3'

LD-4: 5'-GTT GAT ATC CTG AAT TCC <u>TTT</u> CCA AAC TGC C-3'

BD-4-1: 5'-GTT GAT ATC CTG AAT TCC <u>AAG</u> CCA AAC TGC C-3'

BD-4-2: 5'-GTT GAT ATC CTG AAT TCC <u>GCA</u> CCA AAC TGC C-3'

BD-4-3: 5'-GTT GAT ATC CTG AAT TCC <u>ACA</u> CCA AAC TGC C-3'

BD-4-4: 5'-GTT GAT ATC CTG AAT TCC <u>CGC</u> CCA AAC TGC C-3'

Forward Primer: 5'-AAT GAT ACG GCG ACC ACC GAG ATC TAC ACT CTT TCC CTA CAC GAC GCT CTT CCG ATC TCC TGA ATT CC-3'

Reverse Primer:5'-CAA GCA GAA GAC GGC ATA CGA GAT GCT GTA GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC TGG CAG TTT GG-3'

- **Sequences in Figure 8:**

AD-5: 5'-PS-TAT CAA C-DB-3'

LD-5: 5'-ligand-GTT GAT ATC CTG AAT TCC TGA CGG GTC CAA ACT GCC-3'

BD-5: 5'-GTT GAT ATC CTG AAT TCC NCN ANC NNC CAA ACT GCC-3'

(N = A, C, G, C)

Forward Primer: 5'-AAT GAT ACG GCG ACC ACC GAG ATC TAC ACT CTT TCC CTA CAC GAC GCT CTT CCG ATC TCC TGA ATT CC-3'

Reverse Primer for **LD-5**: 5'-CAA GCA GAA GAC GGC ATA CGA GAT TCG AAG GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC TGG CAG TTT GG-3'

Reverse Primer for **BD-5**: 5'-CAA GCA GAA GAC GGC ATA CGA GAT TCG GGA GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC TGG CAG TTT GG-3'

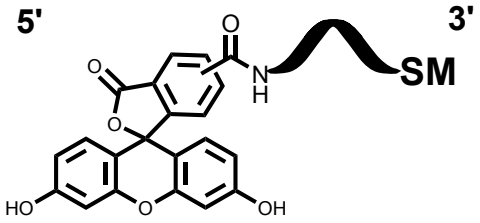
d) DNA characterization.

oligonucleotide	expected mass (Da)	observed mass (Da)
FAM-DNA-biotin in Figure 2, **LD-1**	2742.53	2742.12
FAM-DNA-desthiobiotin in Figure 2, **FD**, **LD-1**	2712.57	2712.23
FAM-DNA-iminobiotin in Figure 2, **LD-1**	2741.55	2741.18
FAM-DNA-raloxifene in Figure 2, **LD-1**	3017.57	3019.11
DABCYL-DNA-biotin in Figure 2, **AD-1**	2957.67	2957.39
DABCYL-DNA-desthiobiotin in Figure 2, **QD**, **AD-1**	2927.71	2927.23
DABCYL-DNA-iminobiotin in Figure 2, **AD-1**	2956.68	2956.16
DABCYL-DNA-raloxifene in Figure 2, **AD-1**	3232.72	3234.56
FAM-DNA-NH_2 in Figure 2	2731.60	2731.94
AD-2, **AD-3**, **AD-4**, **AD-5** (desthiobiotin)	2832.65	2833.17
AD-3 (CBS)	2819.53	2860.62
LD-2 in Figure 5c	8328.52	8329.00
LD-2, matched	9285.06	9286.50
LD-2, partially mismatched	9279.07	9280.10

LD-2, fully mismatched	9254.05	9255.20
LD-3, theophylline	9050.55	9058.90
LD-5	11368.98	11369.43

5. DNA synthesis.

a) 5'-FAM-DNA-3'-SM.



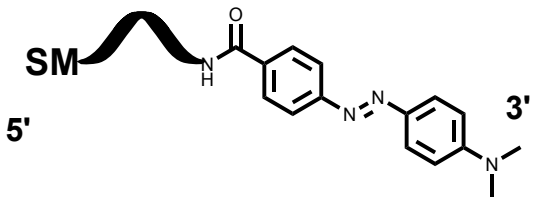
After solid phase DNA synthesis with the 3'-amino-modifier C7 CPG and 5'-amino-modifier 5, the 5'-MMT group was removed to expose the 5'-amine with 3% TCA. In a separate vial, 5, 6-FAM (15.0 mg, 40 μmol) was dissolved in 200 μL anhydrous DMF with 1.0 equiv. HBTU (15.2 mg, 40 μmol), 1.0 equiv. HOBT (5.4 mg, 40 μmol) and 2.3 equiv. DIPEA (15.2 μL, 92 μmol). After 1 hour incubation at room temperature, the activation mixture was added to the CPG. The suspension was incubated at 37 ºC overnight with agitation. The CPG was washed with DMF (3x 600 μL), CH_3CN (3x 600 μL), H_2O (3x 600 μL), CH_3CN (3x 600 μL) and then dried by gentle air stream. The 5'-FAM-labeled DNA was cleaved from CPG by AMA, desalted by a NAP-10 column (GE Pharmacia), and then purified by HPLC.

Biotin-SE (4.1 mg, 12 μmol) or desthiobiotin-SE (3.7 mg, 12μmol) was dissolved in 40 μL DMSO and then added to the 5'-FAM-labeled DNA (in 80 μL phosphate buffer, 0.25 M, pH 7.2) with sonication at 37 ºC for 2 hours. The mixture was desalted by a NAP-5 column (GE Pharmacia) and the product was purified by HPLC. For the iminobiotin, *in situ* activation was employed. Iminobiotin (12.2 mg, 50 μmol), EDC (9.6 mg, 50 μmol) and NHS (5.8 mg, 50 μmol) were dissolved in 150 μL DMSO. The mixture was incubated at 4 ºC overnight with agitation. The activated product was then added to the 5'-FAM-labeled DNA solution with sonication at 37 ºC

for 1 hour. After brief centrifugation, the mixture was separated by a NAP-5 column (GE Pharmacia) and the product was purified by HPLC.

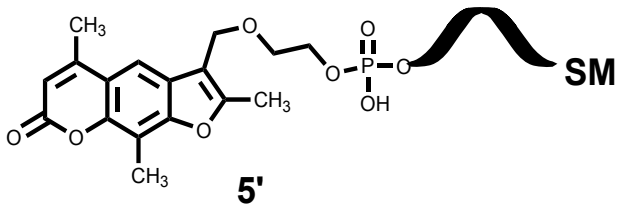
Raloxifene was synthesized according to a published protocol (*Chem. Res. Toxicol.* **2005**, *18*, 1485-1496). The "uncleavable linker", bis(2,5-dioxopyrrolidin-1-yl)glutarate (0.7 mg, 2.1 μmol), and raloxifene (2.3 mg, 5.6 μmol) were dissolved in 20 μL DMSO, respectively. The 5,6-FAM-labeled DNA was dissolved in 40 μL water and 40 μL phosphate buffer (~ 500 μM). The DNA solution was placed in a sonicator at 40 ºC before the uncleavable linker was added. After 2.5 minutes, the raloxifene solution was added to the mixture. The reaction was then maintained at 40 ºC for 1 hour. The mixture was desalted by NAP-5 (GE Pharmacia) and the final product was purified by HPLC.

b) 5'-SM -DNA-3'- DABCYL.



The Fmoc protecting group on the 3'-amino-modifier C7 CPG was removed by 20% piperidine in DMF. DABCYL (10.8 mg, 40 μmol) was dissolved in 400 μL anhydrous DMSO with 1.0 equiv. HBTU (15.2 mg, 40 μmol), 1.0 equiv. HOBT (5.4 mg, 40 μmol) and 2.3 equiv. DIPEA (15.2 μL, 92 μmol). After 1 hour incubation at room temperature, the activation mixture was added to the CPG. The suspension was incubated at 37 ºC overnight with agitation. The CPG was washed with DMF (3x 600 μL), CH_3CN (3x 600 μL), H_2O (3x 600 μL), CH_3CN (3x 600 μL) and then dried by gentle air stream. The DABCYL-labeled CPG was used for automated DNA synthesis. After removing the MMT protecting group by TCA on solid phase, DNA was cleaved, desalted by a NAP-10 column (GE Pharmacia), and purified by HPLC. The following labeling of biotin, desthiobiotin, iminobiotin, and raloxifeneat the 5'-end was the same as described above in a).

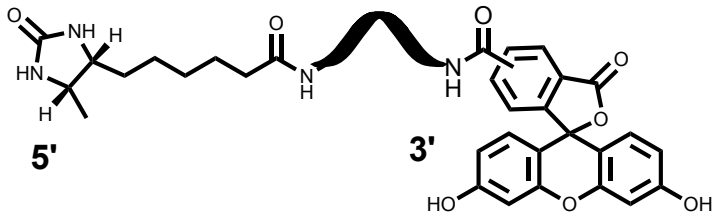
c) 5'-PS-DNA-3'-small molecule.



Automated DNA synthesis was performed with the 3'-amino-modified C7 CPG. Psoralen-C2-phosphoramidate (35.0 mg, 0.07 M) was dissolved in 1 mL redistilled anhydrous acetonitrile and then used as the last base. The coupling time was modified to 999 seconds and repeated twice. After cleavage, desalting, and HPLC purification, the 5'-PS-modified DNA was obtained.

Desthiobiotin was labeded at the 3'-end of the 5'-PS-modified DNA with the same method as described in a). As for the CBS modification, 5'-PS-modifed DNA was dissolved in 40 μL phosphate buffer (pH = 7.2) and CBS-NHS ester (3.6 mg, 12 μmol) was dissolved in 40 μL anhydrous DMSO. The solutions were mixed and the reaction was incubated with sonication at 37 ºC 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. The product was lyophilized and quantitated by a UV spectrometer and characterized by MALDI-TOF mass spectrometry.

d) LD-2 (5'- desthiobiotin-DNA-3'- FAM).

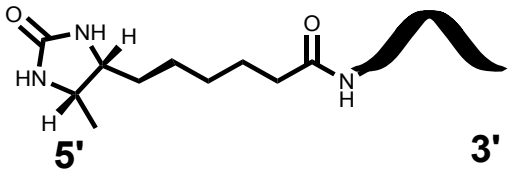


Automated DNA synthesis was performed with the 3'-amino-modified CPG. The 5'-MMT was deprotected with 3% TCA. Desthiobiotin (10.7 mg, 50 μmol) was dissolved in 200 μL anhydrous DMF with 1.0 equiv. HBTU (18.9 mg, 50 μmol), 1.0 equiv. HOBT (6.1 mg, 45 μmol) and 2.3 equiv. DIPEA (20.0 μL, 115 μmol). After 1 hour incubation at room temperature, the

activation mixture was added to CPG. The suspension was incubated at 37 ºC overnight with agitation. The CPG was washed with DMF (3x 600 μL), CH_3CN (3x 600 μL), H_2O (3x 600 μL), CH_3CN (3x 600 μL) and then dried by gentle air stream. The desthiobiotin-labeled DNA was cleaved from CPG, desalted by a NAP-10 column (GE Pharmacia), and then purified by HPLC.

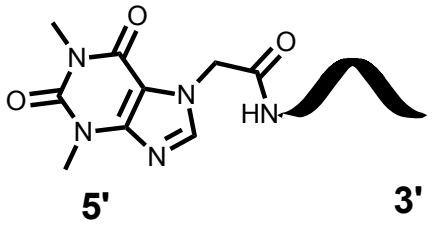
FAM-SE (2.4 mg, 5.0 μmol) was dissolved in 50 μL DMF and then added to the 5'-desthiobiotin-labeled DNA (in 60 μL phosphate buffer, 0.33 M, pH 7.2) with sonication at 37 ºC for 1 hours. The mixture was desalted by a NAP-5 column (GE Pharmacia) and the product was purified by HPLC.

e) LD-3, LD-4, LD-5 (5'- desthiobiotin-DNA-3').



Automated DNA synthesis was performed with the 5'-amino-modifier 5. The DNA was cleaved and purified by HPLC. After MMT removal by 80% acetic acid was employed to remove the MMT protecting group, the DNA was purified again by HPLC. The labeling of desthiobiotin at the 5'-amine is the same as described above in a).

f) LD-3 (5'-theophylline-DNA-3')



The theophylline (23.7 mg, 0.10 mmol) and NHS (13.8 mg, 0.12 mmol) were dissolved in 1.0 mL DMF. EDCI (28.7 mg, 0.15 mmol) was added to the system at room temperature. After the

mixture was stirred for 12 h, the solvent was removed by a rotary evaporator. After dissolving the residue with 10 mL chloroform, the organic phase was washed with 5 mL 2.5% $NaHSO_4$ (three times). After drying with anhydrous sodium sulfate, the crude product was obtained after the solvent was removed.

The 5'-amino modified DNA was dissolved in 12 μL sodium carbonate buffer (pH = 9.0, 0.5 M). Theophylline-NHS (0.67 mg, 2 μmol) dissolved in 100 μL DMSO was added to the system and the reaction was incubated for 12 hours. After ethanol precipitation, the pellet was purified by HPLC to obtain the product.

g) BD-4 (5'-SM-DNA-3').

The synthesis procedure is the same as in e), except either the small molecule was activated *in situ* by EDC/NHS or the crude activated ester (SM-SE) was used in coupling with DNA.

h) 1,024 background DNA (BD-5's).

Based on the relative reactivity of different phosphoramidites (dA/dC/dG/dT = 3: 3: 2: 2.4), we mixed appropriate quantity of reagents to final concentration 0.1 M in anhydrous CH_3CN. The mixed phosphoramidite solution was used at the five "N" positions (shown above in Section 4c). The labeling of small molecule was the same as described in f).

6. MS characterization of the crosslinking products (Figure 5c):

- **DNA sequence:**

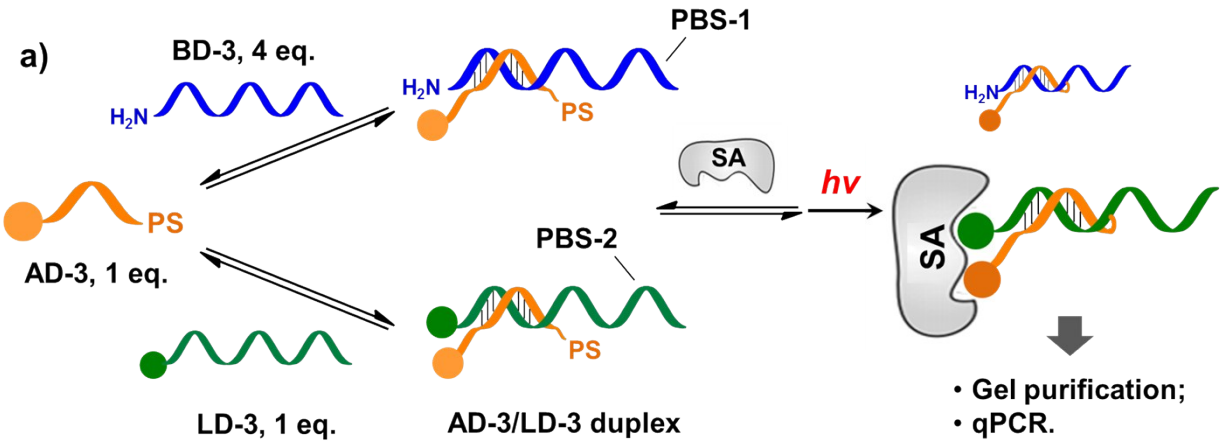
  5'-NH_2-GTT GAT ATT CCC GAA AGT ACA TTG GGC A-3'

  5'-PS-TAT CAA C-NH_2-3'

- **MS results:**

DNA	Expected MS	Observed MS
Upper bands	11217.97	11211.54
Lower bands	11217.97	11217.78

7. qPCR and enrichment fold calculation:



b)

	$C_{T\ (primer\ 1)}$	[AD-3/LD-3]/μM	$C_{T\ (primer\ 2)}$	[AD-3/BD-3] /μM	enrichment fold
SA	14.9	11.0	12.4	50.4	12.0
BSA	13.1	39.2	14.4	15.0	

Figure S2: a) PS-based photo-crosslinking locks the shifted equilibrium for the subsequent hit isolation and qPCR analysis. b) After gel-purification of crosslinked duplexes, qPCR was performed to determine C_T values and to calculate enrichment fold of the AD-3/LD-3 and AD-3/BD-3 duplexes.

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 experiment. We installed orthogonal PCR primers on ligand DNA and background DNA, respectively (Figure S2a).

First, for each set of primers, we conducted a series of PCR experiments with different concentrations of standard template (10 nM, 1 nM, 100 pM, 10 pM and 1 pM). The log values of the initial template concentrations were plotted against the threshold cycle and a linear function was fitted to the data (Figure S3 and S4).

All PCR experiments were performed in triplicates on a Bio-Rad CFX96 Real-Time PCR Detection System with SYBR Green as the detection dye. Thermal cycling lasts 40 times unless otherwise noted. The thermal cycling sequence was set as following: 95 ºC for 2 minutes, then iterated cycles of 95 ºC for 30 seconds, 50 ºC for 30 seconds, and 68 ºC for 15 seconds. All qPCR reactions were carried out with 2X Trans Start Top Green qPCR Super Mix in the buffer provided with a total volume of 20 μL containing Mg^{2+} (2 mM), dNTPs (200 mM), and primers (200 nM each). PCR templates were the psoralen-based photo-crosslinking products extracted from the PAGE gel. Typically, 1% of the extracted products were sufficient for the PCR amplification.

- For primer 1:

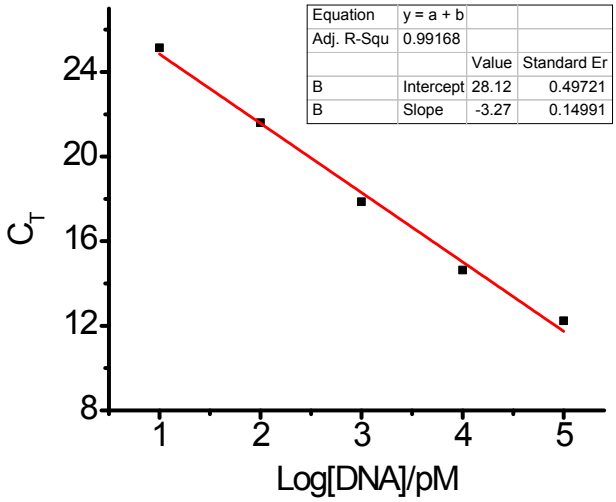


Figure S3: Standard plot of primer set #1.

From the above plot, we obtained the equation for primer set #1:

$Y = -3.27\ X + 28.12$ ($r^2 = 0.992$; linear function)

- For primer 2:

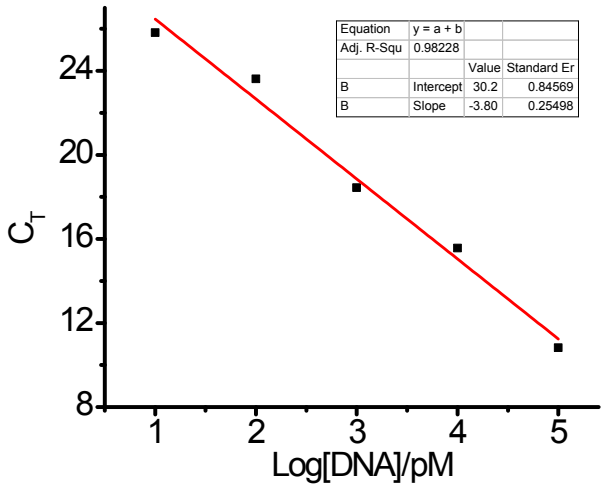


Figure S4: Standard plot of primer set #2.

From the above plot, we obtained the equation for primer set #1:

$Y = -3.80\ X + 30.27$ ($r^2 = 0.982$; linear function)

After each selection, two parallel PCR amplifications with two sets of primers were performed independently; two sets of C_T values were determined and used to calculate initial copy numbers of respective DNA duplexes. Since the process of irradiation, gel purification, and qPCR may also generate biases, a control selection against BSA was performed and C_T values were also measured.

As shown in Figure S2b, the initial copy numbers of **AD-3/LD-3** and **AD-3/BD-3** duplexes can be calculated based on C_T values and their respective standard equations. By comparing SA to BSA, an enrichment fold of 12.0 for the higher affinity **AD-3/LD-3** duplex was obtained. The enrichment fold for the theophylline/CBS-CA-II system (Figure 6c) was calculated with the same method.

8. Direct gel validation of the equilibrium shift by PS-based photo-crosslinking.

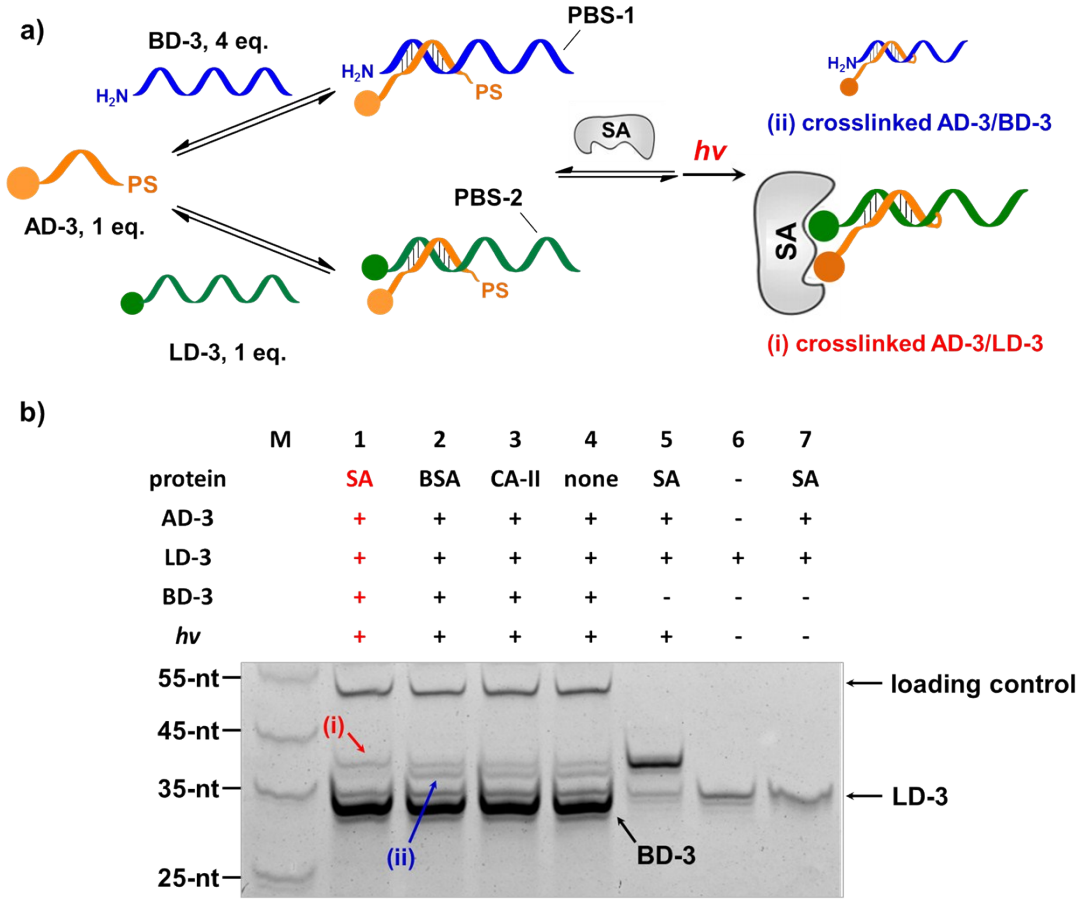


Figure S5: a) PS-based photo-crosslinking locks the shifted equilibrium and the crosslinked duplexes; (i) and (ii) were analysed by gel electrophoresis. b) Gel images of crosslinking reactions (20% TBE native PAGE). M: DNA marker; lane 1: AD-3/LD-3/BD-3 with specific target SA after irradiation; lane 2-4: same as lane 1 but with BSA, CA-II, or no protein; lane 5: standard sample of the crosslinked AD-3/LD-3 (i) duplex; lane 6: standard sample of LD-3; lane 7: AD-3/LD-3 with SA but no irradiation. LD-3: 200 nM; AD-3: 300 nM; BD-3; 800 nM; loading control: a 50-nt DNA standard. Reaction conditions and procedures are the same as Figure 6 in the main text.

Besides the qPCR analysis shown in Figure 6, we also analysed the equilibrium "freeze" and the enrichment of high affinity duplex by direct gel electrophoresis. As shown above in Figure S5, after mixing **AD-3**, **LD-3**, **BD-3** with BSA or CA-II, or with no protein added, photo-crosslinking generated two upper bands (lane 2-4). The top band is the crosslinked **AD-3**/**LD-3** duplex, as confirmed by the standard sample in lane 5; the bottom band is presumably the crosslinked **AD-3**/**BD-3** duplex. In contrast, in the presence of the specific target SA, the high affinity crosslinked

AD-3/**LD-3** duplex (**i**) was significantly enriched at the expense of the **AD-3**/**BD-3** duplex (band significantly weakened), despite the large excess of **BD-3** DNA in the mixture. This result has corroborated our qPCR analysis on enrichment shown in Figure 6 of the main text.

9. Analysis and discussion of the Illumina sequencing data in Figure 8.

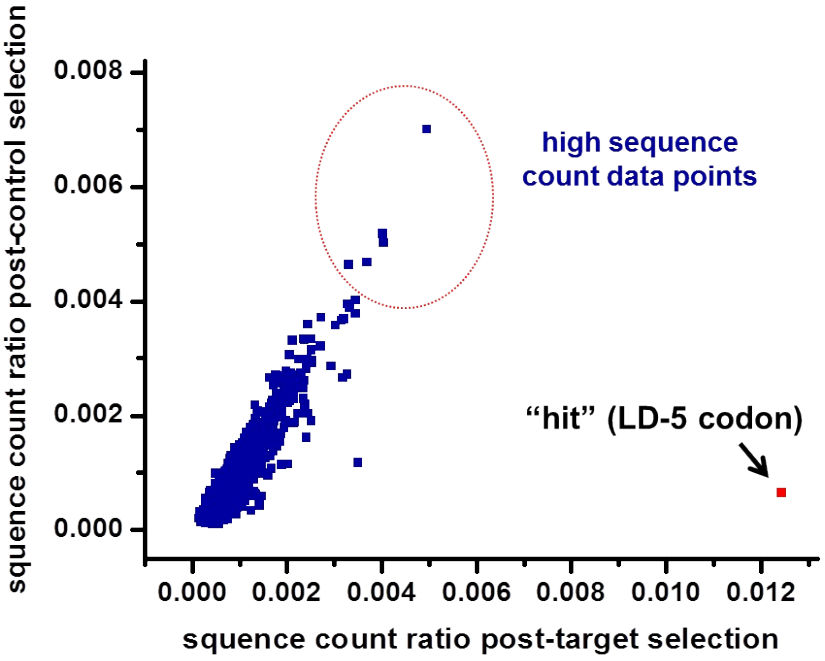


Figure S6: Plot of sequence count ratios after the control selection (no SA) versus count ratios after the target selection (with SA). The sequencing data is the same as shown in Figure 8 of the main text. "Hit" containing the desired LD-5 codon is highlighted in red. Red oval highlights several high sequence count data points. Sequence ratio = (sequence count)/(total sequence count of the library)

The control selection (no target added) can also be used to control for sequence count variations resulting from factors not related to target binding, such as non-ideal oligonucleotide synthesis (where mixed phosphoramidites were used to prepare mixed sequences and phosphoramidites may have different reactivities) and biases in PCR amplification and sequencing due to sequence difference.

In addition, a recent report by Satz has shown that library synthesis quality (even distribution of library compounds) is also important in sequencing (A. L. Satz, *ACS Chem. Biol.* 2015, **DOI:** 10.1021/acschem bio.5b00378; ref. 93 in the main text). Collectively, these observations

indicate that both sequencing depth and library quality are essential to ensure the reliability of the selection data.

The plot shown above compares the sequence count ratios of the target selection and the control selection. A linear correlation was observed, indicating those high count data points in target selection also have high counts in the control selection. It is important to always include such a data analysis in any selection to control for false results.