

Multiplexed Small Molecule Ligand Binding Assays by Affinity

Labeling and DNA Sequence Analysis

Bo Cai and Casey J. Krusemark*

Department of Medicinal Chemistry and Molecular Pharmacology, Purdue Center for Cancer Research, Purdue University, West Lafayette, Indiana 47907, United States

Abstract: Small molecule binding assays to target proteins are a core component of drug discovery and development. While a number of assay formats are available, significant drawbacks still remain in cost, sensitivity, and throughput. To improve assays by capitalizing on the power of DNA sequence analysis, we have developed an assay method that combines DNA encoding with split-and-pool sample handling. The approach involves affinity labeling of DNA-linked ligands to a protein target. Critically, the labeling event assesses ligand binding and enables subsequent pooling of several samples. Application of a purifying selection on the pool for protein-labeled DNAs allows detection of ligand binding by quantification of DNA barcodes. We demonstrate the approach in both ligand displacement and direct binding formats and demonstrate its utility in determination of relative ligand affinity, profiling ligand specificity, and high-throughput small molecule screening.

The identification and characterization of small molecule ligands to proteins are central to the development of drugs and chemical probes.¹⁻⁴ While a variety of assay approaches are available for detecting protein-ligand interactions,⁵⁻⁷ significant limitations exist, particularly in the areas of throughput, cost, and sensitivity.^{8,9} Labeled-ligand binding assays are commonplace and have a particularly long history in the quantitative determination of ligand receptor affinity. Typical labels include radioactive atoms,¹⁰ fluorophores for detection by fluorescence polarization (FP),¹¹ fluorescence energy transfer (FRET),¹² bioluminescence energy transfer (BRET),¹³ or solid phase beads (Alphascreen, e.g.).¹⁴ The use of DNA as a quantitative label in such assays has not been explored and stands to benefit from the several advantages of DNA detection (Figure 1).

The *in vitro* selection of DNA-encoded chemical libraries (DELs), as well as phage and mRNA display libraries, can be considered as a parallel ligand binding assay that uses DNA tags as detection labels. While there are a few examples of effective affinity ranking of ligands using selection approaches,¹⁵ enrichment values from selections often correlate modestly to affinity. A

number of complicating factors, such as the ratio of the ligand affinity constant to target protein concentration, ligand purity, and sequencing depth can affect the enrichment-affinity relationship.¹⁶ As these approaches are most often used for discovery, a strong correlation is typically not required.

The many benefits of DNA detection-based assays are well illustrated in DEL-based ligand discovery, which is a dramatically more accessible and lower cost alternative to high throughput screening (HTS).¹⁷ DNA sequence analysis as a detection platform allows for remarkable sensitivity, low sample requirements, and use of widely available instrumentation. Using DNA as a label readily allows for multiplexing by using unique labels as barcodes. Combining DNA encoding with split-and-pool approaches not only facilitates the synthesis of a large number of unique molecules in a DEL but also allows for a collective assay of molecules by *in vitro* selection. The ability to assay compounds within a pool, rather than individually (as in traditional HTS), is a significant benefit that dramatically decreases assay complexity, variability, and cost.¹⁸

We recently reported a DNA-based assay approach, which we call selection-based sensing, that uses DNA-linked molecules as sample probes and also exploits split-and-pool techniques with DNA-encoding. We have used this method for assaying enzymatic activity with DNA-linked substrates or active site-reactive probes.¹⁹ In this report, we extend the use of selection-based sensing as an assay platform with a DNA-based ligand binding assay for the characterization of protein-ligand interactions. We implement the approach in both a ligand displacement assay (Figure 3 and Figure S1a) suitable for small molecule screening and in a direct binding format to determine relative affinity of DNA-linked ligands (Figure 5 and Figure S1b) through protein titration. Both applications rely upon use of DNA-encoding to record the sample history of a molecule and covalent crosslinking of a DNA-linked ligand to a target protein using a dual display format. An approach that has previously shown to be effective for protein affinity labeling with DNA-linked ligands (Figure 2).²⁰

Our approach to a ligand displacement assay by selection-based sensing involves incubation of a protein target with a known ligand appended to DNA along with an electrophilic crosslinker, which serves as a binding probe (Figures 2, S1a). Co-incubation with a competing free ligand (not linked to DNA) will displace the DNA-linked probe, resulting in decreased affinity labeling and lower DNA recovery in a subsequent purification of DNA's linked to the protein. As an initial demonstration of the technique, we determined half maximal displacement (IC_{50}) values for three

compounds of known affinity to the Chromobox Homolog 7 Chromodomain (CBX7-ChD). DNA-linked probes were prepared by conjugating a peptidic ligand (ligand **1** (BrBA) (Figure S2), $K_d \approx 26 \text{ nM}$)²¹ to different encoding DNAs. In this application, the DNA barcodes of the probes serve to encode the sample identity. Barcodes are simply assigned to particular samples and encode both the identity and the concentration of competing free ligands in a titration series (Figures 3a, S1a). Following incubation, affinity labeling was effectively quenched with excess free ligand (1000x relative to DNA-linked probe) as a precaution, components from all samples were pooled, and the protein-labeled probes were purified via a HaloTag (HT)²² on the target protein using chloroalkane beads. The recovery of each probe was quantitated collectively by next generation DNA sequencing. A dose-dependent reduction in crosslinking was observed with increasing concentration of free ligand, enabling the calculation of IC_{50} values (Figure 3a). To verify these results, a fluorescence polarization (FP) assay was performed using identical concentrations of both protein and a fluorescein-labeled probe ligand, which provided comparable IC_{50} values (Figure S3).

In drug discovery and development, biochemical assays are generally considered low content in terms of the information provided.²³ The ability to assay compounds concurrently against a pool of multiple protein targets would improve this significantly by providing information-rich biochemical profiles while keeping costs low, as a large portion of HTS costs arise from the compounds themselves and from dispensing them. While most assay platforms are limited to a single protein target, this approach allows for multiple proteins to be assayed concurrently (provided they contain orthogonal purification tags). In addition, it is adequately robust and specific to allow assays within crude protein mixtures. To demonstrate these features, we determined the IC_{50} of free ligand **1** (Figure S2) to 5 protein targets labeled with 20-mer oligonucleotides as orthogonal purification tags within *E. coli* cell lysates (Figure 3b). Protein targets were recombinantly expressed in *E. coli*, and SDS-PAGE analysis of lysates confirmed nearly equal expression of the 5 targets (Figure S4a). To label the target proteins, a particular chloroalkane-labeled oligonucleotide was simply added to each lysate in excess (10 μM) and incubated for 30 minutes. Labeled proteins within crude lysates were then pooled and subsequently split for displacement assays, as in Figure 3a, without any protein purification steps. After incubation with free ligand, probes crosslinked to target proteins were selected by purification using complementary oligonucleotides on magnetic beads. The relative recoveries of

the DNA probes were determined by qPCR and gave IC_{50} values of the ligand to each of these five targets, which are consistent with reported values.²⁴

To evaluate robustness for high throughput applications, we prepared 96 probes composed of the DNA-linked ligand **1** with unique barcodes. Half of the probes were incubated with protein and crosslinker, while the other half were additionally incubated with excess free ligand. Enrichment of the probes from the two groups were well distinguished (Z' factor = 0.77, Figure S5). This indicates little variation in DNA barcode detection and a robustness suitable for HTS. Separate incubation of individual samples in a well plate gave a similarly robust Z' factor (0.54, Figure 4a).

Next, we developed a binding assay for *E. coli* dihydrofolate reductase (eDHFR) using a DNA-linked trimethoprim (TMP) probe, a commonly used model ligand receptor pair.²⁵ We screened a small molecule library of 1280 compounds (LOPAC@1280, Sigma-Aldrich) (Figure 4b). The three known eDHFR inhibitors in the library, as well as a novel compound, GW1929²⁶ (Figure S6), showed low recovery of DNA-linked probe. Inhibition assays of eDHFR with GW1929 and TMP validated the results of the screen (Figures 4c, S7, S8, K_i of GW1929 = 410 nM, K_i of TMP = 15 nM). Similarly, GW1929 was able to displace a fluorescein-labeled methotrexate (MTX) from eDHFR in an FP assay (Figure S9, IC_{50} = 7.9 μ M). In addition, GW1929 showed modest inhibition of human DHFR (Figure 4d), which may have implications for its use as a chemical probe for activation of peroxisome proliferator-activated receptor γ .²⁷

This technique can also be used in a direct binding, rather than competitive, format to determine relative affinity of numerous DNA-linked ligands concurrently (Figures 5a, S1b). In this approach, a ligand to be tested is conjugated to DNA in a way that allows tethering of a crosslinking moiety on the opposite strand, as with the probe ligand for the displacement assay. The barcodes within the DNA-linked molecules serve to encode the identity of the ligand and also the protein concentration of a given sample within a titration series. After crosslinking, samples are pooled, and DNAs linked to protein are purified. DNA sequencing then determines the relative crosslinking yield of each DNA to the target protein, which is dependent upon the fraction of the DNA-ligand bound to the protein. Because the labeling rate using the sulfonyl fluoride crosslinker is slow relative to ligand binding ($t_{1/2} \approx 1$ hour, Figure S10), dose response curves do not represent true binding curves. Thus, we have expressed affinities as effective protein concentrations that

gave half-maximal enrichment (EC_{50} values). The use of fast-reacting photocrosslinking groups, as previously demonstrated,²⁰ are expected to give EC_{50} values equal to dissociation constants.

As an initial test, four compounds were prepared on encoding DNA scaffolds (Figure 5a) with a range of affinities to CBX7-ChD. After incubation with various protein concentrations, affinity labeling was quenched with excess ligand, all samples were pooled, and the ligands linked to the target protein were purified. DNA sequencing of the pool gave enrichment curves, and the four ligands showed differential affinity to CBX7-ChD as expected.²¹ To evaluate assay performance in a complex mixture, we used a previously reported collection of 96 DNA-encoded crude compounds containing single monomer substitutions of the BrBA ligand **1** (Table S5-S8).²¹ In this case, DNA barcodes were appended to the compounds to indicate both the compound identity and the concentration of HaloTag-CBX7-Chd (0.2 nM – 50 μ M) used in the crosslinking reaction. Using this approach, 53 of the 96 compounds gave curves sufficient to determine EC_{50} 's (Figure S11-S14, Table S4), and 8 exemplary curves are shown in Figure 5b. The remaining compounds gave incomplete curves and are presumed to have low affinity.

This approach enables the affinity of DNA-linked ligands to be measured as crude compounds (provided contaminants are not ligands), which is a key benefit to improve throughput. Most ligand binding assays are conducted with the small molecule ligand in molar excess over the protein target. Thus, a high purity of a tested ligand is required, as this purity would affect the perceived potency. Because this approach uses the protein in excess, EC_{50} values are not dependent on the concentration of the DNA-linked ligand. To demonstrate this, we compared the values obtained for the DNA-conjugated compound **7** that was intentionally “contaminated” with a non-ligand DNA. As expected, the recovery yield of impure compound **7** was lower than the pure compound as assessed by qPCR, yet the EC_{50} values were within error (Figure 5c, 730 ± 200 nM vs 920 ± 90 nM). To additionally verify the EC_{50} values from the set of 96 crude compounds, we synthesized 12 compounds off-DNA and tested their K_d values (Figure S15). A good correlation ($R^2 = 0.85$, Figure 5d) was observed in the comparison between off-DNA K_d values of the pure compounds and EC_{50} values of the crude on-DNA compounds.

In summary, the affinity labeling of DNA-linked ligands enables high-level multiplexing capability in labeled-ligand binding assays. These assays capitalize on DNA encoding to record the sample history of a DNA-linked molecule to enable pooled manipulation and downstream analysis. As DNA sequence analysis continues to increase in accessibility and decrease in cost,

we expect this method to accelerate numerous aspects of the small molecule ligand discovery and development process.

Acknowledgements

The authors thank Drs. C. Park and E. C. Dykhuizen for critical review of the manuscript. We thank C. Park for assistance with protein expression and purification. This work was supported by NIH 1R35GM128894-01 to C.J.K. The Purdue University Mass Spectrometry and Genome Sequencing Shared Resources are supported by P30 CA023168 from the National Institutes of Health.

Keywords: Protein-ligand interactions, ligand binding assays, affinity labelling.

Figures and legends

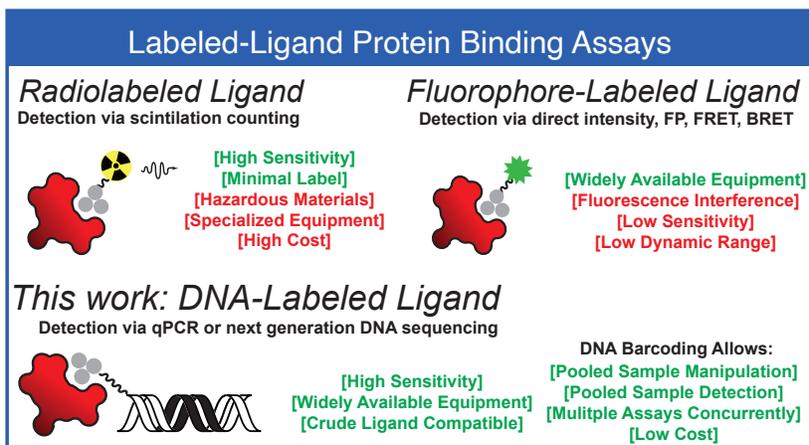


Figure 1. Use of DNA as a detection label in ligand binding assays offers advantages over conventional radioactive and fluorophore labels.



Figure 2. Selection-based sensing approach to ligand binding assays by covalent crosslinking of DNA-linked ligand to target proteins.

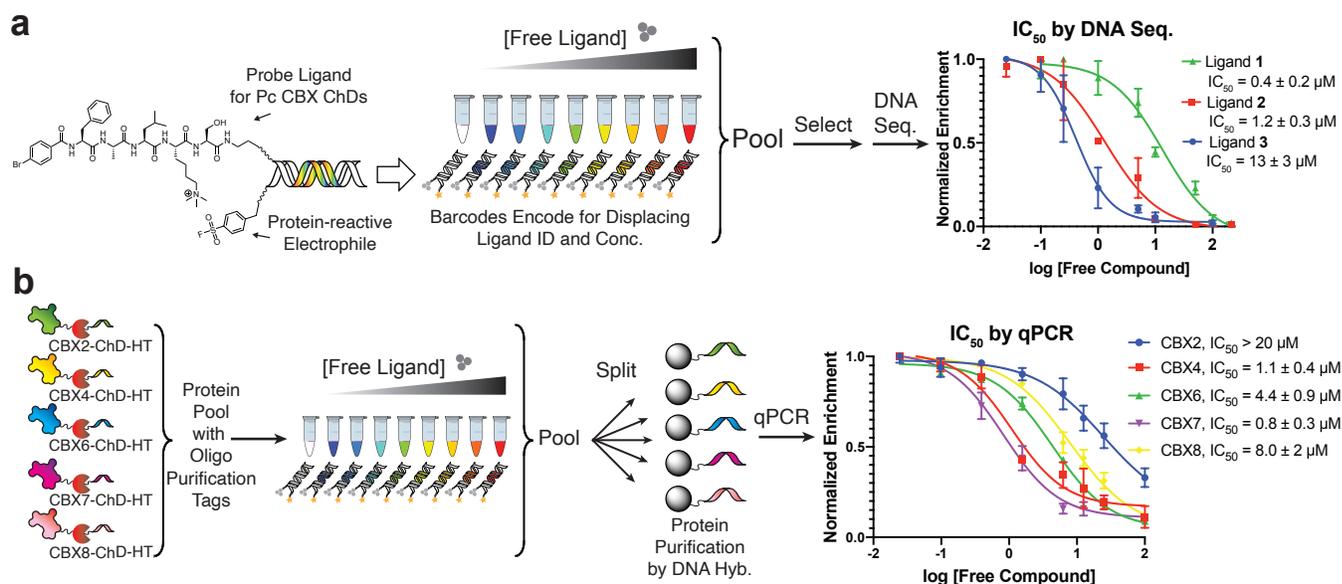


Figure 3. A DNA-based ligand displacement assay for quantifying affinity of non-labeled ligands by DNA sequence analysis. a) Schematic illustration of multiplexed ligand binding assays by displacement of affinity labeling of a DNA-linked probe ligand with free ligands and subsequent DNA sequence analysis. The assay was applied with a broad specificity ligand for the Polycomb (Pc) CBX Chromodomains (ChDs)²⁸ and 3 competing ligands (see Figure S2 for structures) to the CBX7-ChD. b) Concurrent determination of IC₅₀ values for ligand 1 to five protein targets within crude *E. coli* cell lysates. Labeling HaloTag (HT) fusion proteins with 20-mer oligos enabled orthogonal purification using complementary oligo beads. In all panels, error bars indicate one standard deviation of the signal mean for three unique DNA constructs at each free ligand concentration.

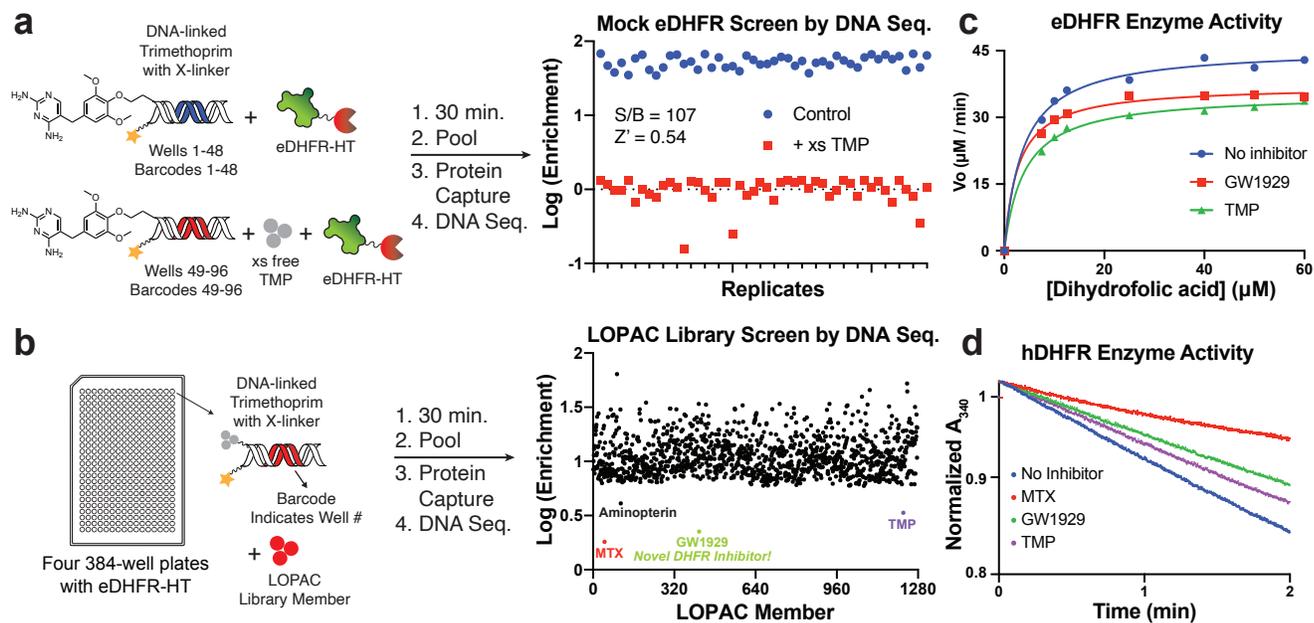


Figure 4. Application of DNA-based ligand displacement assay for screening. a) Assessment of assay robustness with a trimethoprim (TMP) probe and *E. coli* DHFR (eDHFR) in a 96-well plate. b) Screening of the LOPAC library against eDHFR. Samples exhibiting read numbers 1 or less (40 of 1280) were excluded as failed wells due to low sampling. c) Inhibition of GW1929 and TMP of eDHFR. d) Real time assay of inhibition of human DHFR enzyme activity.

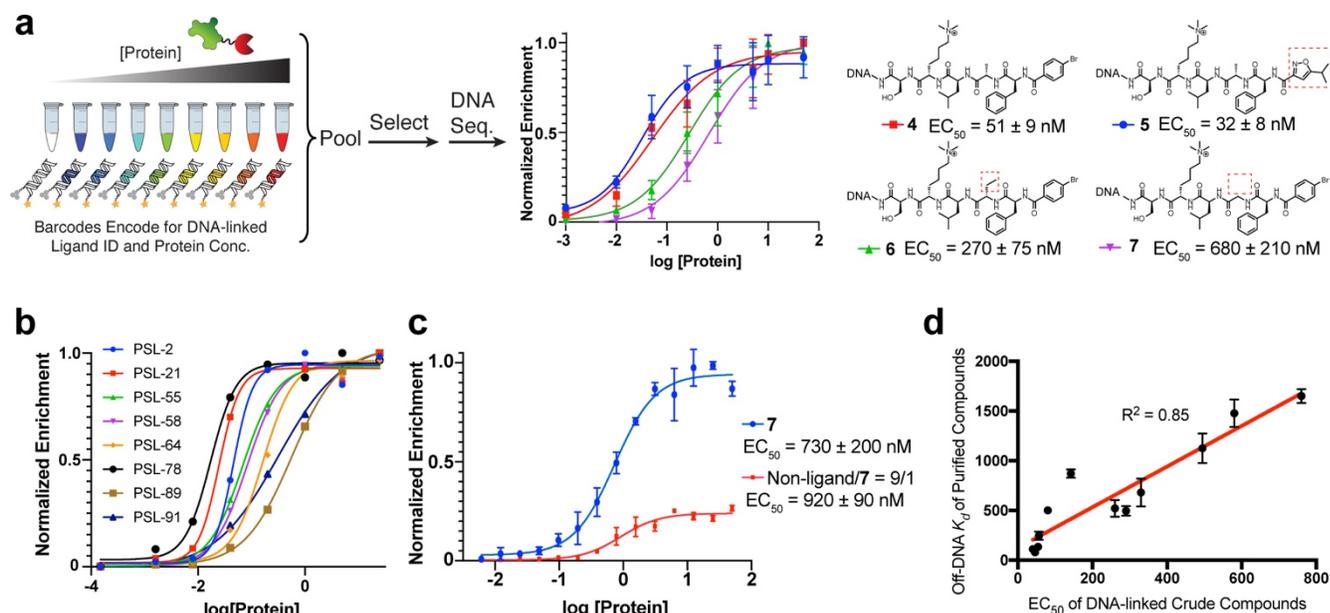


Figure 5. DNA-based direct binding assay for the determination of relative binding affinity of DNA-linked CBX7-ChD ligands. a) Schematic illustration of multiplexed ligand binding assays by affinity labeling of DNA-linked ligands and subsequent DNA sequence analysis. Application of the assay with ligands **4-7** to the CBX7-ChD. Red boxes indicate structural changes from ligand **4**. b) Direct binding assays to determine relative affinity of 96 DNA-linked ligands concurrently were conducted, and binding curves of representative 8 compounds are shown. See Figs. S11-14 for additional curves. c) Measurement of the EC_{50} values of pure and “contaminated” compound **7** to CBX7-ChD. d) Correlation plot between off-DNA K_d of purified compounds and on-DNA EC_{50} values of crude library members. In panels a and c, error bars indicate one standard deviation of the signal mean for three unique DNA constructs at each protein concentration. In panel d, error bars indicate one standard deviation of the signal mean for two off-DNA K_d values.

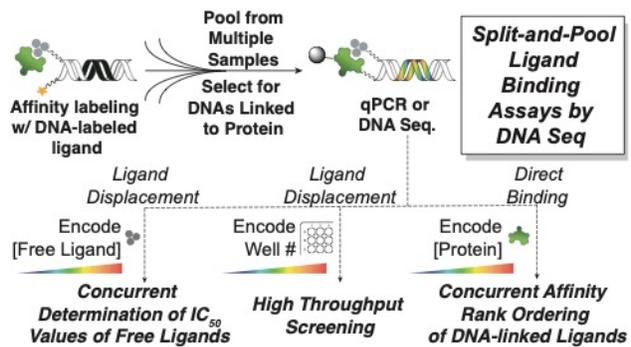
References:

- [1] Geysen, H. M., Schoenen, F., Wagner, D. and Wagner, R. *Nature Reviews Drug Discovery*. 2003, 2, 222-230.
- [2] Schreiber, S.L., Kotz, J.D., Li, M., Aubé, J., Austin, C.P., Reed, J.C., Rosen, H., White, E.L., Sklar, L.A., Lindsley, C.W. and Alexander, B.R. *Cell*. 2015, 161, 1252-1265.
- [3] Annis, D. A., Nazef, N., Chuang, C. C., Scott, M. P. and Nash, H. M. *Journal of the American Chemical Society*. 2004, 126, 15495-15503.
- [4] Gesmundo, N. J., Sauvagnat, B., Curran, P. J., Richards, M. P., Andrews, C. L., Dandliker, P. J. and Cernak, T. *Nature*. 2018, 557, 228-232.
- [5] Weber, G. *Biochemical Journal*. 1952, 51, 145-155.

- [6] Wigle, T.J., Martin Herold, J., Senisterra, G.A., Vedadi, M., Kireev, D.B., Arrowsmith, C.H., Frye, S.V. and Janzen, W.P. *Journal of Biomolecular Screening*. 2010, 15, 62-71.
- [7] Glickman, J. F., Wu, X., Mercuri, R., Illy, C., Bowen, B. R., He, Y. and Sills, M. *Journal of Biomolecular Screening*. 2002, 7, 3-10.
- [8] Huang, X. *Journal of Biomolecular Screening*. 2003, 8, 34-38.
- [9] Lea, W. A. and Simeonov, A. *Expert Opinion on Drug Discovery*. 2011, 6, 17-32.
- [10] Wingler, L.M., Skiba, M.A., McMahon, C., Staus, D.P., Kleinhenz, A.L., Suomivuori, C.M., Latorraca, N.R., Dror, R.O., Lefkowitz, R.J. and Kruse, A.C. *Science*. 2020, 367, 888-892.
- [11] Rossi, A. M. and Taylor, C. W. *Nature protocols*. 2011, 6, 365-387.
- [12] Lohse, M. J., Nuber, S. and Hoffmann, C. *Pharmacological reviews*. 2012, 64, 299-336.
- [13] Robers, M.B., Dart, M.L., Woodroffe, C.C., Zimprich, C.A., Kirkland, T.A., Machleidt, T., Kupcho, K.R., Levin, S., Hartnett, J.R., Zimmerman, K. and Niles, A.L. *Nature communications*. 2015, 6, 1-10.
- [14] Eglen, R. M., Reisine, T., Roby, P., Rouleau, N., Illy, C., Bossé, R. and Bielefeld, M. *Current chemical genomics*. 2008, 1, 2-10.
- [15] (a). Jenson, J. M., Xue, V., Stretz, L., Mandal, T. and Keating, A. E. *Proceedings of the National Academy of Sciences*. 2018, 115, 10342-10351. (b). Rogers, J. M., Passioura, T. and Suga, H. *Proceedings of the National Academy of Sciences*. 2018, 115, 10959-10964. (c). Wang, S., Denton, K.E., Hobbs, K.F., Weaver, T., McFarlane, J.M., Connelly, K.E., Gignac, M.C., Milosevich, N., Hof, F., Paci, I., Musselman, C.A., Dykhuizen, E. C., Krusemark, C. J. *ACS Chemical Biology*. 2019, 15, 112-131. (d). Franzini, R.M., Ekblad, T., Zhong, N., Wichert, M., Decurtins, W., Nauer, A., Zimmermann, M., Samain, F., Scheuermann, J., Brown, P.J., Hall, J., Gräslund, S., Schüler, H., Neri, D. Identification of Structure–Activity Relationships from Screening a Structurally Compact DNA-Encoded Chemical Library. *Angew. Chem.* 2015, 127, 3999-4003; *Angew. Chem. Int. Ed.* 2015, 54, 3927–3931.
- [16] (a). Kuai, L., O’Keeffe, T. and Arico-Muendel, C. *SLAS DISCOVERY: Advancing Life Sciences R&D*. 2018, 23, 405-416. (b). Satz, A. L. *ACS Combinatorial Science*. 2016, 18, 415-424. (c). Hall, J., Foley, T.L., Chen, Q., Israel, D.I., Xu, Y., Ford, K.F., Xie, P., Fan, J. and Wan, J. *Biochemical and Biophysical Research Communications*. 2020, 527, 250-256.
- [17] (a). Neri, D. and Lerner, R. A. *Annual Review of Biochemistry*. 2018, 87, 479-502. (b). Flood, D. T., Kingston, C., Vantourout, J. C., Dawson, P. E. and Baran, P. S. *Israel Journal of Chemistry*. 2020, 60, 268-280. (c). Clark, M. A., Acharya, R. A., Arico-Muendel, C. C., Belyanskaya, S. L., Benjamin, D. R., Carlson, N. R., ... and Morgan, B. A. *Nature Chemical*

- Biology. 2009, 5, 647-654. (d). Brenner, S. and Lerner, R. A. Proceedings of the National Academy of Sciences. 1992, 89, 5381-5383.
- [18] Zhao, G., Huang, Y., Zhou, Y., Li, Y. and Li, X. Expert Opinion on Drug Discovery. 2019, 14, 735-753.
- [19] (a). Jetson, R. R. and Krusemark, C. J. Angew. Chem. 2016, 128, 9714-9718; Angew. Chem. Int. Ed. 2016, 55, 9562–9566. (b). Kim, D., Jetson, R. R. and Krusemark, C. J. Chemical Communications. 2017, 53, 9474-9477.
- [20] (a). Li, G., Liu, Y., Liu, Y., Chen, L., Wu, S., Liu, Y. and Li, X. Angew. Chem. 2013, 125, 9723 –9728; Angew. Chem. Int. Ed. 2013, 52, 9544-9549. (b). Denton, K. E. and Krusemark, C. J. MedChemComm. 2016, 7, 2020-2027.
- [21] Cai, B., Kim, D., Akhand, S., Sun, Y., Cassell, R.J., Alpsy, A., Dykhuizen, E.C., Van Rijn, R.M., Wendt, M.K. and Krusemark, C.J. Journal of the American Chemical Society. 2019, 141, 17057-17061.
- [22] Los, G. V., Encell, L. P., McDougall, M. G., Hartzell, D. D., Karassina, N., Zimprich, C., ... and Wood, K.V. ACS Chemical Biology. 2008, 3, 373-382.
- [23] Blay, V., Tolani, B., Ho, S. P. and Arkin, M. R. Drug Discovery Today. 2020, 25, 1807-1821.
- [24] Stuckey, J. I., Dickson, B. M., Cheng, N., Liu, Y., Norris, J. L., Cholensky, S. H., ... and Frye, S. V. Nature Chemical Biology. 2016, 12, 180-187.
- [25] Baccanari, D. P., Daluge, S. and King, R. W. Biochemistry. 1982, 21, 5068.
- [26] Brown, K.K., Henke, B.R., Blanchard, S.G., Cobb, J.E., Mook, R., Kaldor, I., Kliewer, S.A., Lehmann, J.M., Lenhard, J.M., Harrington, W.W. and Novak, P.J. Diabetes. 1999, 48, 1415-1424.
- [27] Henke, B. R., Blanchard, S. G., Brackeen, M. F., Brown, K. K., Cobb, J. E., Collins, J. L., Harrington, W. W., Hashim, M. A., Hull-Ryde, E. A., Kaldor, I. and Kliewer, S. A. Journal of Medicinal Chemistry. 1998, 41, 5020-5036.
- [28] Simhadri, C., Daze, K.D., Douglas, S.F., Quon, T.T., Dev, A., Gignac, M.C., Peng, F., Heller, M., Boulanger, M.J., Wulff, J.E. and Hof, F. 2014. Journal of medicinal chemistry. 2014, 57, 2874-2883.

Entry for the Table of Contents



The affinity labeling of DNA-barcoded ligands to target proteins allows ligand binding assays using DNA as a label. Pooled sample manipulation and DNA sequence detection enabled by DNA encoding yields quantitative binding assays amenable to both ligand displacement and direct binding formats.

Supplementary Information

Multiplexed Small Molecule Ligand Binding Assays by Affinity Labeling and DNA Sequence Analysis

Bo Cai and Casey J. Krusemark*

Department of Medicinal Chemistry and Molecular Pharmacology, Purdue Center for Cancer Research, Purdue University, West Lafayette, IN 47907, USA

*correspondence to: cjk@purdue.edu

Table of Contents

1. Supplementary Methods	14
2. Supplementary Schemes	26
3. Supplementary Protein Sequences	29
4. Supplementary Figures	34
5. Supplementary Tables	44
6. Supplementary References	52

1. Supplementary Methods

General materials and instrumentation

Unless otherwise specified, all chemical reagents were purchased from commercial sources and used without further purification. Common chemical reagents were purchased from Chem-Impex and Sigma-Aldrich. DNA oligonucleotides were purchased from Integrated DNA Technologies. All modified oligonucleotides were synthesized on DEAE Sepharose[®] Fast Flow (GE Healthcare) and then purified by reverse-phase high-pressure liquid chromatography (HPLC) on an Agilent 1100 Series HPLC system. After HPLC purification, DNA was precipitated, and the DNA concentrations were measured using an IMPLEN NanoPhotometer[™] P-Class. All reagents used in PCR were purchased from Thermo Fisher Scientific. PCR reactions were performed using a BIO RAD C1000 Touch Thermo Cycler, and qPCR tests were conducted using an Applied Biosystems[™] ViiA[™] 7 Real Time PCR system. Purified water in all reactions was obtained using a Millipore MilliQ water filtration system. SDS-PAGE gel imaging was performed on GE Typhoon FLA 9500 Fluorescent Image Analyzer Scanner. The MALDI TOF/TOF MS analysis of all purified DNA constructs were performed using Applied Biosystems Voyager DE Pro in the Center for Drug Discovery at Purdue University. Mass spectroscopy experiments were conducted on the Waters Acquity UPLC with SQD2 mass spectrometer at the Purdue University Mass Spectrometry Center. Next-generation Illumina sequencing was performed at the Purdue Genomic Core Facility. Sequencing reads were parsed from FASTQ file to text file format. Reads of the variable sequence region were matched to barcodes using Matlab, which were collated and filtered to count total read numbers.

General procedure for the synthesis of peptides

The ligand **1-3** in Fig. S2 were synthesized on Rink Amide MBHA resin (Novabiochem[®]). All couplings and deprotections were monitored by ninhydrin tests. Briefly, 50 mg of Rink Amide

MBHA resin was swelled for 15 min in N, N'- dimethylformamide (DMF) and 15 min in dichloromethane (DCM). The resin was resuspended in 20 % piperidine in DMF for 30 min to deprotect 9-fluorenylmethyloxycarbonyl (Fmoc). Each coupling reaction contained 5 equivalents of Fmoc-amino acid, 5 equivalents of 1-Hydroxy-7- azabenzotriazole (HOAt), and 5 equivalents of 0.1 M N, N'-diisopropylcarbodiimide (DIC) and was allowed to proceed for 90 min. Following each coupling, the Fmoc group was removed by treating the resin with 20 % piperidine in DMF for 30 min. Peptides were finally cleaved with 95 % trifluoroacetic acid (TFA), 2.5 % triisopropylsilane (TIPS), and 2.5 % H₂O for 3 h. The crude peptide was precipitated out of cold ethyl ether and purified on a reversed-phase HPLC using a H₂O/MeCN + 0.1% TFA gradient with detection at 215 nm. The purity of all peptides was > 95% (confirmed by LC-MS). The mass of all peptides was confirmed by Waters Acquity UPLC with SQD2 mass spectrometer (Scheme S1-S3, Table S3).

General procedure for the synthesis of ligands on-DNA

The benzoic acid (non-ligand control) and the ligands of CBX7-ChD and eDHFR were synthesized on-DNA as previously described¹. All peptides were synthesized on a 5'-amine-modified 40-mer (ssDNAa-linker-ssDNAb-5'-NH₂) on cartridges. Briefly, 200 µL of diethylaminoethanol (DEAE) slurry (GE Healthcare) in 50 % aq. ethanol was loaded on a cartridge. The cartridge was washed with 1 mL of DEAE binding buffer (10 mM HOAc and 0.005 % Triton X-100) for 3 times. 20 nmol of the 40-mer DNA was then diluted with 1 mL of DEAE binding buffer and went through the cartridge. The DEAE slurry was washed with 3 x 1 mL MeOH. Each coupling reaction contained 50 mM carboxylic acids, 5 mM HOAt and 50 mM N-(3-dimethylaminopropyl)-N'-ethylcarbodiimide hydrochloride (EDC-HCl) in 40 % DMF/60 % MeOH and was allowed to proceed for 30 min. Each coupling reaction was repeated twice at room temperature. Following the second coupling, the Fmoc group was removed by treating the DEAE slurry with 20% piperidine in DMF for 30 min.

The final DNA-peptide conjugate was eluted using DEAE Elution Buffer (1.5 M NaCl and 0.005 % Triton X-100). The trimethoprim (TMP) carboxylic acid derivative was synthesized as previously described^{1,3} and was conjugated to the 5'-amine-modified 40-mer using the same procedure as mentioned above. Eluted conjugates were HPLC purified, ethanol precipitated and MALDI confirmed by Applied Biosystems Voyager DE Pro (Scheme S4-S8, Table S2).

Synthesis of sulfonyl fluoride reactive group on DNA

The 4-[2-(hex-5-ynoylamino)ethyl]benzenesulfonyl fluoride was synthesized and conjugated to ssDNA α' -3'-CapN₃-5'-FAM using click chemistry as previously described². Desired DNA product was HPLC purified, ethanol precipitated and MALDI confirmed by Applied Biosystems Voyager DE Pro (Scheme S9, Table S2).

Synthesis of DNA-linked chloroalkane tag

The chloroalkane tag (1-(2-(2-azidoethoxy)ethoxy)-6-chlorohexane) was synthesized as previously described¹. The ssDNA-5'-chloroalkane was synthesized by conjugating chloroalkane tag to hexynyl DNA using click chemistry (Scheme S10). The final product was HPLC purified, precipitated and re-suspended in water. The MALDI of ssDNA-5'-chloroalkane was confirmed by Applied Biosystems Voyager DE Pro (Table S2).

PCR amplification of DNA-ligands onto dsDNA constructs

Each of the ligand-40-mer conjugates were attached to 60-mer dsDNA constructs by PCR using the following conditions: 1X DreamTaq buffer (Thermo Fisher Scientific), 0.2 mM dNTPs, 0.5 μ M ligand-40 mer as forward primer, 0.5 μ M ssDNA α' as reverse primer, 0.01 ng/ μ L 60-mer dsDNA

template, and 0.025 U/ μ L DreamTaq DNA polymerase for 20 cycles (94 °C for 3 min, 94 °C for 15 s, annealing at 58 °C for 15s, extension at 72 °C for 30s, and a final extension of 72 °C for 5 min). The PCR product was purified by homemade Solid-phase Reversible Immobilization (SPRI) beads and quantified by UV absorbance at 260 nm using an IMPLEN NanoPhotometerTM P-Class.

Quantitative PCR experiments

All qPCR reactions were performed in 384-well plates. The total volume of each qPCR reaction was 10 μ L containing 2.5 μ L of DNA template sample, 2.5 μ L of 1 μ M forward and reverse primers and 5 μ L of Applied BiosystemsTM PowerUpTM SYBRTM Green Master Mix (Applied BiosystemsTM). The thermal cycling procedure of the qPCR machine was set as follows: Hold stage: The wells were brought to 50°C at a rate of 1.6°C/s and held at 50°C for 2 min. Following hold stage, the wells were brought to 95°C at a rate of 1.6°C/s and held at 95°C for 10 min. The following 35 PCR cycles consisted of 95°C for 15 s, 60°C for 20 s at a rate of 1.6°C/s. The melt curve stage consisted of 95°C for 15 s, 60°C for 1 min and 95°C for 15 s (the speed of temperature change was set to 1.6°C/s, 1.6°C/s and 0.05°C/s, respectively).

DNA sequencing data analysis

Sequencing reads were parsed from the forward read FASTQ file to text file format, giving the variable 20-mer barcode sequences only. Reads of the variable sequence region were matched to barcodes as individual 20-mers using Matlab scripts. Reads with any of the 20-mer sequences matching less than 18 bases were discarded. Only sequences with at least 18 out of 20 correct base matches were assigned to barcodes. Reads were collated and filtered to count total read numbers. Enrichments were calculated relative to the non-ligand control and initial preselection sample mixture.

Cloning information

The Halotag fusion of CBX constructs were obtained from Addgene (Dr. Xiaojun Ren lab plasmids). Addgene Plasmids ID for these constructs are: # 82510 (HT-CBX2), # 82513 (HT-CBX4), # 82514 (HT-CBX6), # 82520 (HT-CBX7) and # 82516 (HT-CBX8). PCR amplification was performed with the five constructs to obtain the Halotag fusion of CBX chromodomains. The Halotag-CBX2/4/6/7/8-ChD was cloned into pET28b (Novagen®). Halotag-eDHFR was cloned into pET28b using the procedure as previously described¹. The cloned plasmids were confirmed by sequencing at Purdue sequencing facility. Cloned plasmids in pET28b vector were used for expression and purification of proteins.

Protein expression

Each of the expression plasmid was transformed into *E. coli* BL21 (DE3) Rosetta cells (Novagen®). 50 µl of each transformation was plated on a LB agar plate containing kanamycin (50 µg/mL) and chloramphenicol (34 µg/mL) and was incubated overnight at 37°C. A single colony was selected and resuspended in 3 mL of LB broth (with antibiotics) for 12 h at 37 °C. The starter culture was then added to a flask with 300 mL of LB broth containing kanamycin (50 µg/mL) and chloramphenicol (34 µg/mL). The cells were grown at 37 °C until the OD₆₀₀ reached 0.5 to 0.8. The induction was performed by adding IPTG (final concentration = 0.5 mM) to the culture and incubated at 37 °C for 4 h. After induction, the cells were spun down at 3,000 xg for 10 min. Protein expression was checked by Coomassie stained protein gel (Fig. S4).

LDA for the determination of the IC₅₀ values of free ligands 1-3 in Fig. S2

Each crosslinking reaction contains 250 nM Halotag-CBX7-ChD, 1 μ M BSA, 1 mg/mL Salmon Sperm DNA (Thermo Fisher Scientific) in 20 μ L of buffer (100 mM sodium phosphate, pH 8.0, 250 mM NaCl, 0.02% (v/v) Tween-20). To each solution was added varying concentration of the free ligand (100 nM – 200 μ M for ligand **1** and ligand **2**; 25 nM – 100 μ M for ligand **3**, eight different concentrations of each ligand were tested). Following 30 minutes' incubation, to each of the eight tubes was added three unique DNA-linked BrBA probes (final concentration = 50 nM) and ssDNAa'-linked sulfonyl fluoride reactive group (final concentration = 100 nM). The mixture was then allowed to incubate for another 30 min. To each tube was then added 1000x free BrBA (relative to DNA-linked BrBA probes) to quench the crosslinking reaction. All eight tubes for each ligand were pooled and captured by 2 μ L of Magne[®] HaloTag[®] Beads (Promega) at 4°C for 1 h. The beads were washed with PBS containing 1 μ M BSA and 1 mg/mL Salmon Sperm DNA (Thermo Fisher Scientific) with 5% SDS for 3 times. For the fourth wash, beads were washed with PBS. Beads were resuspended in 20 μ L water and boiled under 95°C for 10 min. The final elution was PCR amplified followed by the addition of specific sequencing adapters. The recovery of each DNA-linked BrBA was determined by DNA sequencing. The DNA-linked benzoic acid was doped into each tube to serve as the non-ligand control. The enrichment of DNA-linked BrBA was given as the fold change in the recovery of DNA-BrBA relative to the recovery of DNA-linked non-ligand control. The data were normalized using the highest enrichment as 100% signal and 0 as 0% signal. The IC₅₀ of each ligand was derived with GraphPad Prism 8 using the following model: log (inhibitor) vs. response – variable slope (four parameters).

LDA for the determination of the IC₅₀ values of BrBA to Halo-CBX2/4/6/7/8

Each of the five proteins (in crude cell lysates) was pre-incubated with a unique 10 μ M 20-mer DNA-linked chloroalkane in 100 μ L of buffer (100 mM sodium phosphate, pH 8.0, 250 mM NaCl,

0.02% (v/v) Tween-20) for 30 min. The labeling was quenched by the addition of 1000x free chloroalkane (relative to DNA-linked chloroalkane). All five proteins were pooled and was evenly split to eight tubes containing varying concentration of the free ligand **1** (BrBA) (20 nM – 100 μ M). The solution was allowed to incubate for 30 min, followed by the addition of eight unique DNA-linked BrBA (final concentration = 20 nM) and ssDNAa'-linked sulfonyl fluoride reactive group (final concentration = 40 nM) to each tube. After 30 min, 1000x free BrBA (relative to DNA-linked BrBA probes) was added to each tube to quench the crosslinking reaction. All eight tubes were pooled and evenly separated to 5 new tubes. To each tube was added corresponded DNA-linked biotin and 4 μ L of NanoLink Streptavidin Magnetic Beads (Vector Laboratories). Following 1 h of incubation, the beads were washed with PBS containing 1% SDS, 1 μ M BSA and 1 mg/mL Salmon Sperm DNA (Thermo Fisher Scientific) for 3 times. For the fourth wash, beads were washed with PBS. Beads were resuspended in 20 μ L water and boiled under 95°C for 10 min. The recovery of the eight DNA-linked probes for each protein was quantified by qPCR. The IC₅₀ values of BrBA to five proteins were derived with GraphPad Prism 8 using the following model: log (inhibitor) vs. response – variable slope (four parameters).

LOPAC library screening

The LOPAC library screening was performed in Chemical Genomics Facility at Purdue University. To generate the probe for LOPAC library screening, the trimethoprim-40-mer conjugate (Scheme S8) was attached to unique 60-mer dsDNA constructs by PCR. The screening was performed in four 384-well plates. Each well contained approximately 50 nM Halotag-eDHFR, 2 nM DNA-linked trimethoprim and 10 nM crosslinker in 22 μ L of buffer (100 mM sodium phosphate, pH 8.0, 250 mM NaCl, 0.02% (v/v) Tween-20). First, 220 nL of the compounds in the LOPAC library were added to each well using Echo Acoustic Liquid Handler (final concentration of each compound =

10 μM). Then, 15 μL of the buffer containing 75 nM protein was dispensed into each well using a multidrop dispenser (Thermo Fisher Scientific) and was allowed to incubate with compounds in the LOPAC library for 30 min. Following incubation, 2 μL of 20 nM DNA-linked trimethoprim was transferred from storage 384-well plate to the assay plate using the Fluent[®] Automation Workstation (Tecan). The four 384-well assay plates were allowed to shake for 30 min. After 30 min-incubation, 5 μL of 40 nM crosslinker was added to each well using the Mantis Liquid Handling Robots and was allowed to incubate on shaker for another 30 min. All the fractions in each of the four 384-well assay plates were combined together by centrifuging the components from the entire plate into a V-bottomed polypropylene reservoir VBLOK200 (Clickbio). Each of the four elution samples was captured separately by 5 μL of Magne[®] HaloTag[®] Beads (Promega) at 4°C for 1 h. The beads were washed with PBS containing 1 μM BSA and 1 mg/mL Salmon Sperm DNA (Thermo Fisher Scientific) with 5% SDS for 3 times. For the fourth wash, beads were washed with PBS. Beads were resuspended in 20 μL water and boiled under 95°C for 10 min. The final elution was PCR amplified followed by the addition of specific sequencing adapters. The recovery of each DNA-linked trimethoprim was determined by DNA sequencing. DNA samples with read numbers below 2 in sequencing considered as failed wells and were excluded from analysis due to low sampling.

Determination of the EC₅₀ values of compound 4 - 7 in Fig. 5a

The compound to be tested was conjugated to ssDNAa-linker-ssDNA_b-5'-NH₂ to allow covalent crosslinking (Scheme S4-S7). Each of the conjugates was attached to 24 unique 60-mer dsDNA constructs by PCR. The DNA barcodes serve to encode both the identity of the DNA-linked compound and also the protein concentration of a given sample within a titration series. For each crosslinking reaction, a DNA-linked nonligand control and three of the DNA-linked compound

(same compound, three encoding DNA to get triplicate results, final concentration = 20 nM) were incubated with eight varying concentration of Halotag-CBX7-ChD (from 1 nM to 50 μ M), 1 μ M BSA, 1 mg/mL Salmon Sperm DNA (Thermo Fisher Scientific) in 20 μ L of buffer (100 mM sodium phosphate, pH 8.0, 250 mM NaCl, 0.02% (v/v) Tween-20) for 30 min. Then, to each of the eight tubes was added 40 nM ssDNAa'-linked sulfonyl fluoride reactive group. The mixture was allowed to incubate for another 30 min. Following crosslinking, to each tube was then added 1000x free BrBA (relative to DNA-linked compounds) to quench the crosslinking reaction. All eight fractions of each DNA-linked compound were pooled and captured by 2 μ L of Magne[®] HaloTag[®] Beads (Promega) at 4°C for 1 h. The beads were washed with PBS containing 1 μ M BSA and 1 mg/mL Salmon Sperm DNA (Thermo Fisher Scientific) with 5% SDS for 3 times. For the fourth wash, beads were washed with PBS. Beads were resuspended in 20 μ L water and boiled under 95°C for 10 min. The final elution was PCR amplified followed by the addition of specific sequencing adapters. The recovery of each DNA-linked compound was determined by DNA sequencing. The EC₅₀ of each compound was derived with GraphPad Prism 8 using the following model: Sigmoidal dose-response (variable slope).

Determination of the EC₅₀ values of 96 compounds in the positional scanning library

The positional scanning library used was prepared as previously described¹. The 96 compounds (Table S5-S8) to be tested were conjugated to ssDNAa-linker-ssDNAb-5'-NH₂ to allow covalent crosslinking. Each of the DNA-linked compounds was attached to unique 60-mer dsDNA constructs by PCR. Following PCR amplification, all of the 96 DNA-linked compounds were pooled. For each crosslinking reaction, a DNA-linked nonligand control and the 96 DNA-linked compounds (final total concentration = 1 μ M) were incubated with eight varying concentration of Halotag-CBX7-ChD (from 0.15 nM to 50 μ M), 1 μ M BSA, 1 mg/mL Salmon Sperm DNA (Thermo

Fisher Scientific) in 20 μL of buffer (100 mM sodium phosphate, pH 8.0, 250 mM NaCl, 0.02% (v/v) Tween-20) for 30 min. Then, to each of the eight tubes was added 2 μM ssDNAa'-linked sulfonyl fluoride reactive group. The mixture was allowed to incubate for another 30 min. Following crosslinking, to each tube was added 1000x free BrBA (relative to DNA-linked compounds) to quench the crosslinking reaction. The eight fractions were separately captured by 2 μL of Magne[®] HaloTag[®] Beads (Promega) at 4°C for 1 h. The beads were washed with PBS containing 1 μM BSA and 1 mg/mL Salmon Sperm DNA (Thermo Fisher Scientific) with 5% SDS for 3 times. For the fourth wash, beads were washed with PBS. Beads were resuspended in 20 μL water and boiled under 95°C for 10 min. The final elution was PCR amplified followed by the addition of specific sequencing adapters to each of the eight elution. The recovery of each DNA-linked compound was determined by DNA sequencing. The EC_{50} of each DNA-linked compound was derived with GraphPad Prism 8 using the following model: Sigmoidal dose-response (variable slope).

Study on the time-dependency of crosslinking

Each crosslinking reaction contains 1 μM DNA-linked BrBA, 1 μM Halotag-CBX7-ChD, 1 μM BSA, 1 mg/mL Salmon Sperm DNA (Thermo Fisher Scientific) in 20 μL of buffer (100 mM sodium phosphate, pH 8.0, 250 mM NaCl, 0.02% (v/v) Tween-20). The ssDNAa'-linked BrBA was incubated with protein for 30 min prior to the addition of the ssDNAa'-linked sulfonyl fluoride (DNA-SF, final concentration = 1 μM). The crosslinking reaction was allowed to proceed for different time periods, quenched by the addition of 6x SDS-loading buffer, directly boiled under 95 °C for 2 min, and analyzed by SDS-PAGE. To study the effect of hydrolysis of DNA-SF, the DNA-SF was incubated in buffer for 1 h prior to the addition of DNA-linked BrBA and Halotag-CBX7-ChD; the crosslinking was allowed to proceed for 1 h, quenched by SDS-loading buffer and analyzed

by SDS-PAGE. The intensity of crosslinked bands was normalized using 0 as 0% signal and the intensity of crosslinked band after 1 h – crosslinking as 100% signal (Fig. S10).

Fluorescence polarization assay

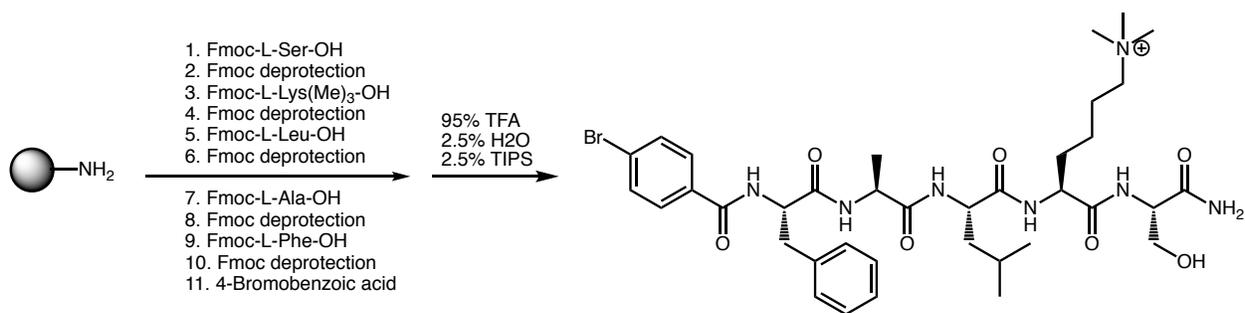
The fluorescence polarization assay was performed in black 384-well plates. The assay buffer contains 20 mM Tris-HCl, pH 8.0, 250 mM NaCl, 1 mM DTT, 1 mM PMSF and 0.02% (v/v) Tween-20. For eDHFR, the probe fluorescein-methotrexate (Biotium) and inhibitors were dissolved in DMSO. The concentration of inhibitors was varied, with constant concentration of fluorescein-methotrexate and Halotag-eDHFR at 2 nM and 5 nM, respectively. For CBX7-ChD, the FAM-probe and protein concentration was set to 50 nM and 250 nM, respectively. Plates were incubated for 15 min in darkness prior to reading with a Synergy™ 4 plate reader (BioTek) with $\lambda_{\text{exc}} = 485 \text{ nm}$, $\lambda_{\text{emi}} = 530 \text{ nm}$. The IC_{50} of each compound was derived with GraphPad Prism 8 using the following model: log (inhibitor) vs. response – variable slope (four parameters).

Dihydrofolate reductase assay

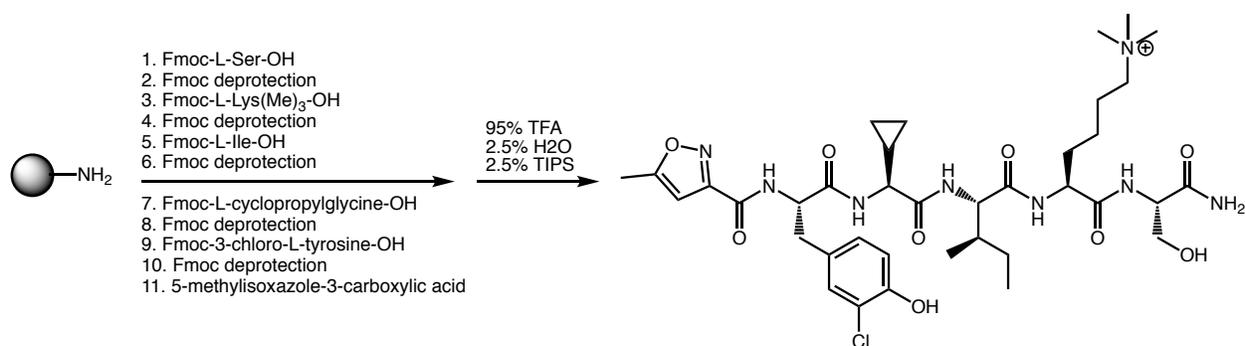
The dihydrofolate reductase assay was performed using the Dihydrofolate Reductase Assay Kit (Sigma-Aldrich, Catalog Number CS0340) with a Varian Cary® 50 UV-Vis Spectrophotometer (Agilent). The kinetic program was set to read every 0.1 second for 3 minutes at 340 nm and 22 °C. 1 mL of 1x assay buffer was added to a quartz cuvette as the blank sample. For the inhibition assay performed with human DHFR, each quartz cuvette contains 20 μL of supplied human DHFR, 6 μL of 10 mM NADPH, 5 μL of 10 mM dihydrofolic acid, 1 μL of 100 μM inhibitor and 968 μL of 1x supplied assay buffer. For the inhibition assay performed with eDHFR, the enzyme concentration was set to 20 nM. The absorbance at 340 nm (A_{340}) was recorded with the spectrophotometer, which was normalized using the highest value within each test as 100% signal

and 0 as 0% signal. The A_{340} – Time plot was generated using GraphPad Prism 8. The inhibition kinetics tests of GW1929 and TMP were performed in quartz cuvettes. Each cuvette contains 990 μL of 1x supplied assay buffer with 20 nM eDHFR in the presence of 100 nM GW1929 (or 5 nM TMP), 6 μL of 10 mM NADPH and varying concentration of dihydrofolic acid (0 to 60 μM). The Lineweaver-Burk plot was generated with GraphPad Prism 8 using the reciprocal of the initial velocity as y axis and the reciprocal of the substrate concentration as x axis ($1/V$ vs. $1/[S]$).

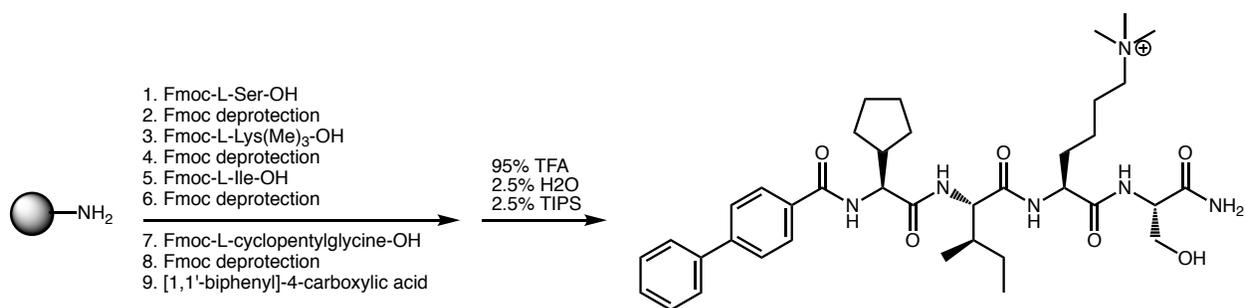
2. Supplementary Schemes



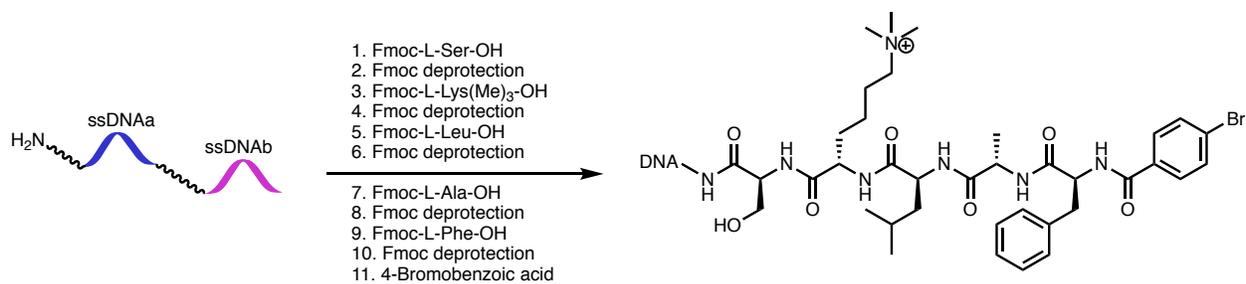
Scheme S1. Synthesis of compound **1** in Figure S2



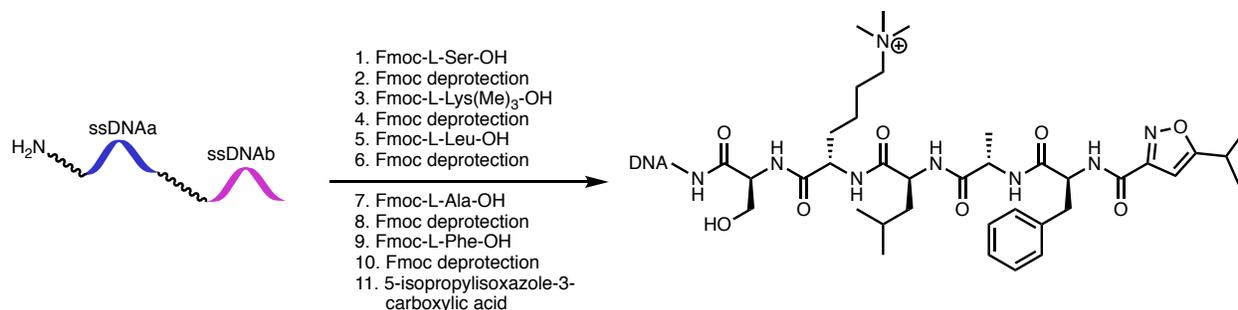
Scheme S2. Synthesis of compound **2** in Figure S2



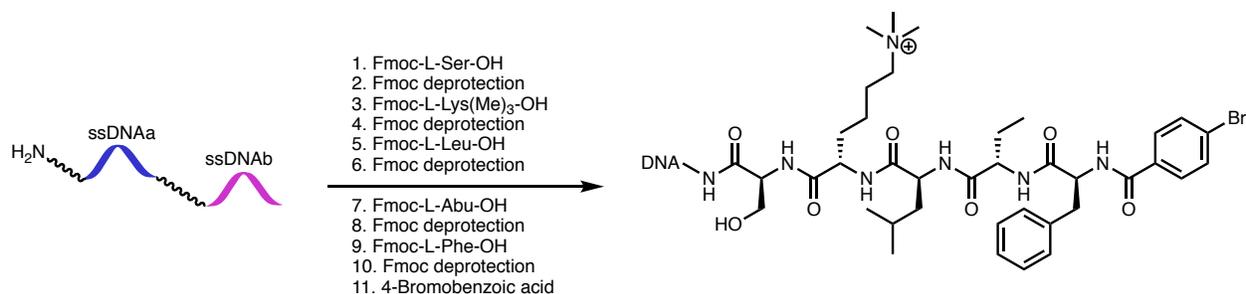
Scheme S3. Synthesis of compound **3** in Figure S2



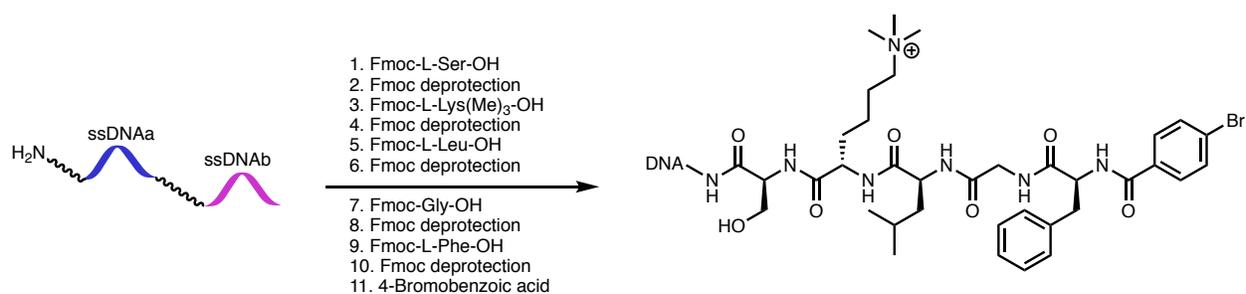
Scheme S4. Synthesis of compound **4** in Figure 5a



Scheme S5. Synthesis of compound **5** in Figure 5a



Scheme S6. Synthesis of compound **6** in Figure 5a



Scheme S7. Synthesis of compound **7** in Figure 5a

3. Supplementary Protein Sequences

Sequences of Halotag-CBX2/4/6/7/8-ChD, Halotag-eDHFR

Halotag-CBX2-ChD:

ATGGCAGAAATCGGTA CTGGCTTTCCATTTCGACCCCAATTATGTGGAAGTCCTGGGCGAG
CGCATGCACTACGTGATGTTGGTCCGCGCGATGGCACCCCTGTGCTGTTCTGCACGGT
AACCCGACCTCCTCCTACGTGTGGCGCAACATCATCCC GCATGTTGCACCGACCCATCGC
TGCATTGCTCCAGACCTGATCGGTATGGGCAAATCCGACAAACCAGACCTGGGTTATTTCT
TCGACGACCACGTCCGCTTCATGGATGCCTTCATCGAAGCCCTGGGTCTGGAAGAGGTCTG
TCCTGGTCATTACGACTGGGGCTCCGCTCTGGGTTTCCACTGGGCCAAGCGCAATCCAG
AGCGCGTCAAAGGTATTGCATTTATGGAGTTCATCCGCCCTATCCCGACCTGGGACGAATG
GCCAGAATTTGCCCGCGAGACCTTCCAGGCCTTCCGCACCACCGACGTCCGGCCGCAAGC
TGATCATCGATCAGAACGTTTTTATCGAGGGTACGCTGCCGATGGGTGTCTGTCGGCCCGC
TGACTGAAGTCGAGATGGACCATTACCGCGAGCCGTTCTGAATCCTGTTGACCGCGAGC
CACTGTGGCGCTTCCCAAACGAGCTGCCAATCGCCGGTGAGCCAGCGAACATCGTCGCG
CTGGTCAAGAATACATGGACTGGCTGCACCAAGTCCCCTGTCCC GAAGCTGCTGTTCTGG
GGCACCCCAAGCGTTCTGATCCCACCGGCCGAAGCCGCTCGCCTGGCCAAAAGCCTGCC
TAACTGCAAGGCTGTGGACATCGGCCCGGGTCTGAATCTGCTGCAAGAAGACAACCCGGA
CCTGATCGGCAGCGAGATCGCGCGCTGGCTGTGACGCTCGAGATTTCCGGCGCGAATT
CATCGATAGATCTGATATCGGTACCAGTCTGACTCTAGAGCAGTCTCTATGGAGGAGCTGAG
CAGCGTGGGCGAGCAGGTCTTCGCCGCGGAGTGCATCCTGAGCAAGCGGCTCCGCAAGG
GCAAGCTGGAGTACCTGGTCAAGTGGCGCGGCTGGTCCTCCAACATAACAGCTGGGAG
CCGGAGGAGAACATCCTGGACCCGAGGCTGCTCCTGGCCTTCCAGAAGAAGGAAGAGCT
CCGTGACAAGCTTGCGGCCGCACTCGAGCACCACCACCACCACCTGA

Halotag-CBX4-ChD:

ATGGCAGAAATCGGTA CTGGCTTTCCATTTCGACCCCAATTATGTGGAAGTCCTGGGCGAG
CGCATGCACTACGTGATGTTGGTCCGCGCGATGGCACCCCTGTGCTGTTCTGCACGGT
AACCCGACCTCCTCCTACGTGTGGCGCAACATCATCCC GCATGTTGCACCGACCCATCGC
TGCATTGCTCCAGACCTGATCGGTATGGGCAAATCCGACAAACCAGACCTGGGTTATTTCT
TCGACGACCACGTCCGCTTCATGGATGCCTTCATCGAAGCCCTGGGTCTGGAAGAGGTCTG
TCCTGGTCATTACGACTGGGGCTCCGCTCTGGGTTTCCACTGGGCCAAGCGCAATCCAG
AGCGCGTCAAAGGTATTGCATTTATGGAGTTCATCCGCCCTATCCCGACCTGGGACGAATG
GCCAGAATTTGCCCGCGAGACCTTCCAGGCCTTCCGCACCACCGACGTCCGGCCGCAAGC
TGATCATCGATCAGAACGTTTTTATCGAGGGTACGCTGCCGATGGGTGTCTGTCGGCCCGC
TGACTGAAGTCGAGATGGACCATTACCGCGAGCCGTTCTGAATCCTGTTGACCGCGAGC
CACTGTGGCGCTTCCCAAACGAGCTGCCAATCGCCGGTGAGCCAGCGAACATCGTCGCG
CTGGTCAAGAATACATGGACTGGCTGCACCAAGTCCCCTGTCCC GAAGCTGCTGTTCTGG
GGCACCCCAAGCGTTCTGATCCCACCGGCCGAAGCCGCTCGCCTGGCCAAAAGCCTGCC
TAACTGCAAGGCTGTGGACATCGGCCCGGGTCTGAATCTGCTGCAAGAAGACAACCCGGA
CCTGATCGGCAGCGAGATCGCGCGCTGGCTGTGACGCTCGAGATTTCCGGCGCGAATT
CACTTGCTATGGAGCTGCCAGCTGTTGGCGAGCACGTCTTCGCGGTGGAAAGCTTCGAGA
AGAAGCGGATCCGCAAGGGCAGAGTGGAGTATCTGGTCAAATGGAGAGGCTGGTCCGCC
AAATATAACACGTGGGAACCGGAGGAGAACATCCTGGACCCCAAGGCTGCTGATCGCCTTC
CAGAACAGGGAACGGCAGGAGCAGGAGCTCCGTGACAAGCTTGCGGCCGCACTCGAGC
ACCACCACCACCACCACCTGA

Halotag-CBX6-ChD:

ATGGCAGAAATCGGTA CTGGCTTTCCATTGACCCCCATTATGTGGAAGTCCTGGGCGAG
CGCATGCACTACGTGATGTTGGTCCGCGCGATGGCACCCCTGTGCTGTTCCCTGCACGGT
AACCCGACCTCCTCCTACGTGTGGCGCAACATCATCCCGCATGTTGCACCGACCCATCGC
TGCATTGCTCCAGACCTGATCGGTATGGGCAAATCCGACAAACCAGACCTGGGTTATTTCT
TCGACGACCACGTCCGCTTCATGGATGCCTTCATCGAAGCCCTGGGTCTGGAAGAGGTCCG
TCCTGGTCATTACGACTGGGGCTCCGCTCTGGGTTTCCACTGGGCCAAGCGCAATCCAG
AGCGCGTCAAAGGTATTGCATTTATGGAGTTCATCCGCCCTATCCCGACCTGGGACGAATG
GCCAGAATTTGCCCGCGAGACCTTCCAGGCCTTCCGCACCACCGACGTCCGGCCGCAAGC
TGATCATCGATCAGAACGTTTTTATCGAGGGTACGCTGCCGATGGGTGTCGTCCGCCCGC
TGACTGAAGTCGAGATGGACCATTACCGCGAGCCGTTCCCTGAATCCTGTTGACCGCGAGC
CACTGTGGCGCTTCCCAAACGAGCTGCCAATCGCCGGTGAGCCAGCGAACATCGTCGCG
CTGGTCAAGAATACATGGACTGGCTGCACCAGTCCCCTGTCCCGAAGCTGCTGTTCTGG
GGCACCCAGGCGTTCTGATCCACCGGCCGAAGCCGCTCGCCTGGCCAAAAGCCTGCC
TAACTGCAAGGCTGTGGACATCGGCCCGGGTCTGAATCTGCTGCAAGAAGACAACCCGGA
CCTGATCGGCAGCGAGATCGCGCGCTGGCTGTGACGCTCGAGATTTCCGGCGCGAATT
CCTGCAGGATGGAGCTGTCTGCAGTGGGCGAGCGGGTCTTCGCGGCCGAATCCATCATC
AAACGGCGGATCCGAAAGGGACGCATCGAGTACCTGGTGAATGGAAGGGGTGGGCGAT
CAAGTACAGCACTTGGGAGCCCGAGGAGAACATCCTGGACTCGCGGCTCATTGCAGCCTT
CGAACAAAAGGAGAGGGAGCGTGAGGAGCTCCGTGACAAAGCTTGGCGCCGCACTCGAG
CACCACCACCACCCTGA

Halotag-CBX7-ChD:

CGAGGGCCACCATGGCAGAAATCGGTA CTGGCTTTCCATTGACCCCCATTATGTGGAAG
TCCTGGGCGAGCGCATGCACTACGTGATGTTGGTCCGCGCGATGGCACCCCTGTGCTGT
TCCTGCACGGTAACCCGACCTCCTCCTACGTGTGGCGCAACATCATCCCGCATGTTGCAC
CGACCCATCGCTGCATTGCTCCAGACCTGATCGGTATGGGCAAATCCGACAAACCAGACC
TGGGTTATTTCTTCGACGACCACGTCCGCTTCATGGATGCCTTCATCGAAGCCCTGGGTCT
GGAAGAGGTGCTCCTGGTCATTACGACTGGGGCTCCGCTCTGGGTTTCCACTGGGCCAA
GCGCAATCCAGAGCGCGTCAAAGGTATTGCATTTATGGAGTTCATCCGCCCTATCCCGACC
TGGGACGAATGGCCAGAATTTGCCCGCGAGACCTTCCAGGCCTTCCGCACCACCGACGTC
GGCCGCAAGCTGATCATCGATCAGAACGTTTTTATCGAGGGTACGCTGCCGATGGGTGTC
GTCCGCCCGCTGACTGAAGTCGAGATGGACCATTACCGCGAGCCGTTCCCTGAATCCTGTT
GACCGCGAGCCACTGTGGCGCTTCCCAAACGAGCTGCCAATCGCCGGTGAGCCAGCGAA
CATCGTCGCGCTGGTCAAGAATACATGGACTGGCTGCACCAGTCCCCTGTCCCGAAGCT
GCTGTTCTGGGGCACCCAGGCGTTCTGATCCCACCGGCCGAAGCCGCTCGCCTGGCCA
AAAGCCTGCCTAACTGCAAGGCTGTGGACATCGGCCCGGGTCTGAATCTGCTGCAAGAAG
ACAACCCGGACCTGATCGGCAGCGAGATCGCGCGCTGGCTGTGACGCTCGAGATTTCC
GGCGCGAATTCTGGCGGTTCCGGGAGGCGGGAGTGGAGGTGAGCAGGTGTTCCCGTGG
AGAGCATCCGGAAGAAGCGCGTGCGGAAGGGTAAAGTCGAGTATCTGGTGAAGTGGAAA
GGATGGCCCCAAAGTACAGCACGTGGGAGCCAGAAGAGCACATCTTGGACCCCGCCT
CGTCATGGCCTACGAGGAGAAGGAGGAGTGA

Halotag-CBX8-ChD:

ATGGCAGAAATCGGTA CTGGCTTTCCATTGACCCCCATTATGTGGAAGTCCTGGGCGAG
CGCATGCACTACGTGATGTTGGTCCGCGCGATGGCACCCCTGTGCTGTTCCCTGCACGGT
AACCCGACCTCCTCCTACGTGTGGCGCAACATCATCCCGCATGTTGCACCGACCCATCGC

TGCATTGCTCCAGACCTGATCGGTATGGGCAAATCCGACAAACCAGACCTGGGTTATTTCT
TCGACGACCACGTCCGCTTCATGGATGCCTTCATCGAAGCCCTGGGTCTGGAAGAGGTGCG
TCCTGGTCAATCACGACTGGGGCTCCGCTCTGGGTTTCCACTGGGCCAAGCGCAATCCAG
AGCGCGTCAAAGGTATTGCATTTATGGAGTTCATCCGCCCTATCCCGACCTGGGACGAATG
GCCAGAATTTGCCCGGAGACCTTCCAGGCCTTCCGCACCACCGACGTGGGCCGCAAGC
TGATCATCGATCAGAACGTTTTTATCGAGGGTACGCTGCCGATGGGTGTCGTCCGCCCGC
TGAATGAAGTTCGAGATGGACCATTACCGCGAGCCGTTCTGAATCCTGTTGACCGCGAGC
CACTGTGGCGCTTCCCAAACGAGCTGCCAATCGCCGGTGAGCCAGCGAACATCGTCGCG
CTGGTCAAGAATACATGGACTGGCTGCACCAGTCCCCTGTCCCGAAGCTGCTGTTCTGG
GGCACCCCAGGCGTTCTGATCCCACCGGCCGAAGCCGCTCGCCTGGCCAAAAGCCTGCC
TAACTGCAAGGCTGTGGACATCGGCCCGGGTCTGAATCTGCTGCAAGAAGACAACCCGGA
CCTGATCGGCAGCGAGATCGCGCGCTGGCTGTCGACGCTCGAGATTTCCGGCGCGAATT
CACTTGCTATGGAGCTTTCAGCGGTGGGGGAGCGGGTGTTCGCGGCCGAAGCCCTCCTG
AAGCGGCGCATACGGAAAGGACGCATGGAATACCTCGTGAAATGGAAGGGATGGTCGCA
GAAGTACAGCACATGGGAACCGGAGGAAAACATCCTGGATGCTCGCTTGTCTCGCAGCCTT
TGAGGAAAGGGAAGAGCTCCGTCGACAAGCTTGCGGCCGCACTCGAGCACCACCACCAC
CACCCTGA

Halotag-eDHFR:

CGAGGGCCACCATGGCAGAAATCGGTACTGGCTTTCATTCGACCCCCATTATGTGGAAG
TCCTGGGCGAGCGCATGCACTACGTGCATGTTGGTCCGCGCGATGGCACCCCTGTGCTGT
TCCTGCACGGTAACCCGACCTCCTCCTACGTGTGGCGCAACATCATCCCGCATGTTGCAC
CGACCCATCGCTGCATTGCTCCAGACCTGATCGGTATGGGCAAATCCGACAAACCAGACC
TGGGTTATTTCTTCGACGACCACGTCCGCTTCATGGATGCCTTCATCGAAGCCCTGGGTCT
GGAAGAGGTCGTCTGGTCATTCACGACTGGGGCTCCGCTCTGGGTTTCCACTGGGCCAA
GCGCAATCCAGAGCGCGTCAAAGGTATTGCATTTATGGAGTTCATCCGCCCTATCCCGACC
TGGGACGAATGGCCAGAATTTGCCCGGAGACCTTCCAGGCCTTCCGCACCACCGACGTC
GGCCGCAAGCTGATCATCGATCAGAACGTTTTTATCGAGGGTACGCTGCCGATGGGTGTC
GTCCGCCCGCTGACTGAAGTCGAGATGGACCATTACCGCGAGCCGTTCTGAATCCTGTT
GACCGCGAGCCACTGTGGCGCTTCCCAAACGAGCTGCCAATCGCCGGTGAGCCAGCGAA
CATCGTCGCGCTGGTCAAGAATACATGGACTGGCTGCACCAGTCCCCTGTCCCGAAGCT
GCTGTTCTGGGGCACCCCAGGCGTTCTGATCCCACCGGCCGAAGCCGCTCGCCTGGCCA
AAAGCCTGCCTAACTGCAAGGCTGTGGACATCGGCCCGGGTCTGAATCTGCTGCAAGAAG
ACAACCCGGACCTGATCGGCAGCGAGATCGCGCGCTGGCTGTCGACGCTCGAGATTTCC
GGCGCGAATTCTGGCGGTTCCGGGAGGCGGGAGTGGAGGTATGATCAGTCTGATTGCGGC
GTTAGCGGTAGATCGCGTTATCGGCATGGAAAACGCCATGCCGTGGAACCTGCCTGCCGA
TCTCGCCTGGTTTAAACGCAACACCTTAAATAAACCCGTGATTATGGGCCCGCCATACCTGG
GAATCAATCGGTGCTCCGTTGCCAGGACGCAAAAATATTATCCTCAGCAGTCAACCGGGTA
CGGACGATCGCGTAACGTGGGTGAAGTCGGTGGATGAAGCCATCGCGGCGTGTGGTGAC
GTACCAGAAATCATGGTGATTGGCGGCGGTGCGGTTTATGAACAGTTCTTGCCAAAAGCG
CAAAAATGTATCTGACGCATATCGACGCAGAAGTGAAGGGCGACCCCATTTCCCGGATT
ACGAGCCGGATGACTGGGAATCGGTATTCAGCGAATTCACGATGCTGATGCGCAGAACT
CTCACAGCTATTGCTTTGAGATTCTGGAGCGGCGGTAA

Primary sequences of Halotag-CBX2/4/6/7/8-ChD, Halotag-eDHFR

Halo-CBX2-ChD:

MAEIGTGFPFDPHYVEVLGERMHYVDVGPRDGTPLVFLHGNPTSSYVWRNIIPHVAPTHRCIAP
DLIGMGKSDKPDLDGYFFDDHVRFMDFIEALGLEEVVLIHDWGSALGFHWAKRNPERVKGIA
FMFIRPIPTWDEWPEFARETFQAFRTTDVGRKLIIDQNVFIEGTLPMGVVRPLTEVEMDHYRE
PFLNPVDREPLWRFPNELPIAGEPANIVALVEEYMDWLHQSPVPKLLFWGTPGVLIPPAEAARL
AKSLPNCKAVDIGPGLNLLQEDNPDIGSEIARWLSTLEISGANSSIDLISVPVDSRAVSMEELSS
VGEQVFAAECILSKRLRKGKLEYLVKWRGWSSKHNSWEPEENILDPRLLLAFQKKEELRRQAC
GRTRAPPPPL

Halo-CBX4-ChD:

MAEIGTGFPFDPHYVEVLGERMHYVDVGPRDGTPLVFLHGNPTSSYVWRNIIPHVAPTHRCIAP
DLIGMGKSDKPDLDGYFFDDHVRFMDFIEALGLEEVVLIHDWGSALGFHWAKRNPERVKGIA
FMFIRPIPTWDEWPEFARETFQAFRTTDVGRKLIIDQNVFIEGTLPMGVVRPLTEVEMDHYRE
PFLNPVDREPLWRFPNELPIAGEPANIVALVEEYMDWLHQSPVPKLLFWGTPGVLIPPAEAARL
AKSLPNCKAVDIGPGLNLLQEDNPDIGSEIARWLSTLEISGANSLAMELPAVGEHVFAVESFEK
KRIRKGRVEYLVKWRGWSPKYNTWEPEENILDPRLLIQNRERQEELRRQACGRTRAPP
PPL

Halo-CBX6-ChD:

MAEIGTGFPFDPHYVEVLGERMHYVDVGPRDGTPLVFLHGNPTSSYVWRNIIPHVAPTHRCIAP
DLIGMGKSDKPDLDGYFFDDHVRFMDFIEALGLEEVVLIHDWGSALGFHWAKRNPERVKGIA
FMFIRPIPTWDEWPEFARETFQAFRTTDVGRKLIIDQNVFIEGTLPMGVVRPLTEVEMDHYRE
PFLNPVDREPLWRFPNELPIAGEPANIVALVEEYMDWLHQSPVPKLLFWGTPGVLIPPAEAARL
AKSLPNCKAVDIGPGLNLLQEDNPDIGSEIARWLSTLEISGANSCRMELSAVGERVFAAESIIKR
RIRKGRIEYLVKWKGWAIKYSTWEPEENILDSRLIAAFEQKEREREELRRQACGRTRAPPPPL

Halotag-CBX7-ChD:

MAEIGTGFPFDPHYVEVLGERMHYVDVGPRDGTPLVFLHGNPTSSYVWRNIIPHVAPTHRCIAP
DLIGMGKSDKPDLDGYFFDDHVRFMDFIEALGLEEVVLIHDWGSALGFHWAKRNPERVKGIA
FMFIRPIPTWDEWPEFARETFQAFRTTDVGRKLIIDQNVFIEGTLPMGVVRPLTEVEMDHYRE
PFLNPVDREPLWRFPNELPIAGEPANIVALVEEYMDWLHQSPVPKLLFWGTPGVLIPPAEAARL
AKSLPNCKAVDIGPGLNLLQEDNPDIGSEIARWLSTLEISGANSSGGSGGGSGGEQVFAVESIR
KKRVRKKGVEYLVKWKGWPPKYSTWEPEEHILDPRLVMAYEEKEE

Halotag-CBX8-ChD:

MAEIGTGFPFDPHYVEVLGERMHYVDVGPRDGTPLVFLHGNPTSSYVWRNIIPHVAPTHRCIAP
DLIGMGKSDKPDLDGYFFDDHVRFMDFIEALGLEEVVLIHDWGSALGFHWAKRNPERVKGIA
FMFIRPIPTWDEWPEFARETFQAFRTTDVGRKLIIDQNVFIEGTLPMGVVRPLTEVEMDHYRE
PFLNPVDREPLWRFPNELPIAGEPANIVALVEEYMDWLHQSPVPKLLFWGTPGVLIPPAEAARL
AKSLPNCKAVDIGPGLNLLQEDNPDIGSEIARWLSTLEISGANSLAMELSAVGERVFAAEALLK
RRIRKGRMEYLVKWKGWSQKYSTWEPEENILDARLLAAFEEREELRRQACGRTRAPPPPL

Halotag-eDHFR:

MAEIGTGFPFDPHYVEVLGERMHYVDVGPRDGTPLVFLHGNPTSSYVWRNIIPHVAPTHRCIAP
DLIGMGKSDKPDLDGYFFDDHVRFMDFIEALGLEEVVLIHDWGSALGFHWAKRNPERVKGIA

FMEFIRPIPTWDEWPEFARETFFQAFRTTDVGRKLIIDQNVFIEGTLPMGVWRPLTEVEMDHYRE
PFLNPVDREPLWRFPNELPIAGEPANIVALVEEYMDWLHQSPVPKLLFWGTPGVLIPPAEAARL
AKSLPNCKAVDIGPGLNLLQEDNPDIGSEIARWLSTLEISGANSSGGSSGGSSGGMISLIAALAVD
RVIGMENAMPWNLPADLAWFKRNTLNKPVIMGRHTWESIGRPLPGRKNIILSSQPGTDDRVTW
VKSVDIAIACGDVPEIMVIGGGRVYEQFLPKAQKLYLTHIDAEVEGDTHFPDYEPDDWESVFS
EFHDADAQNSHSYCFEILERR

4. Supplementary Figures

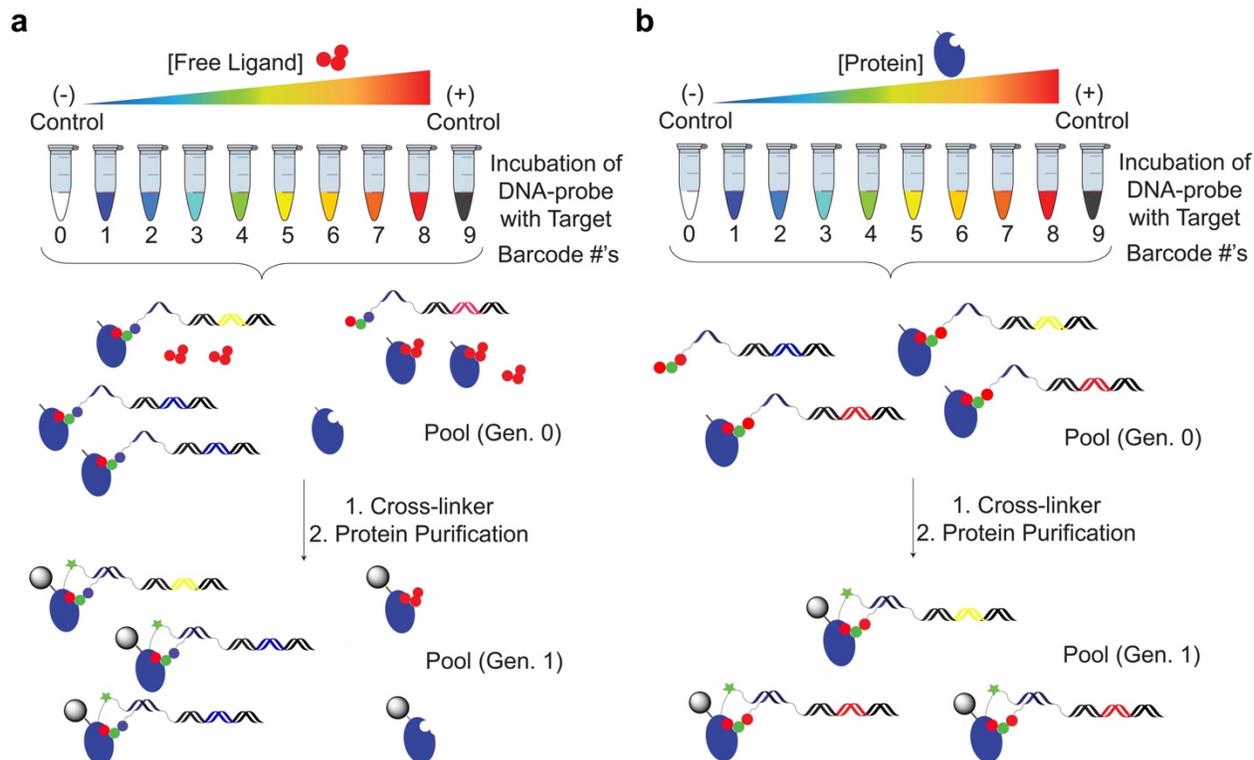


Figure S1. Affinity Labeling of DNA-linked Ligands for High Throughput Ligand Binding Assays. (a) Ligand displacement assay for compounds' IC_{50} determination. (b) DNA-based crosslinking assay for the determination of ligands' EC_{50} values.

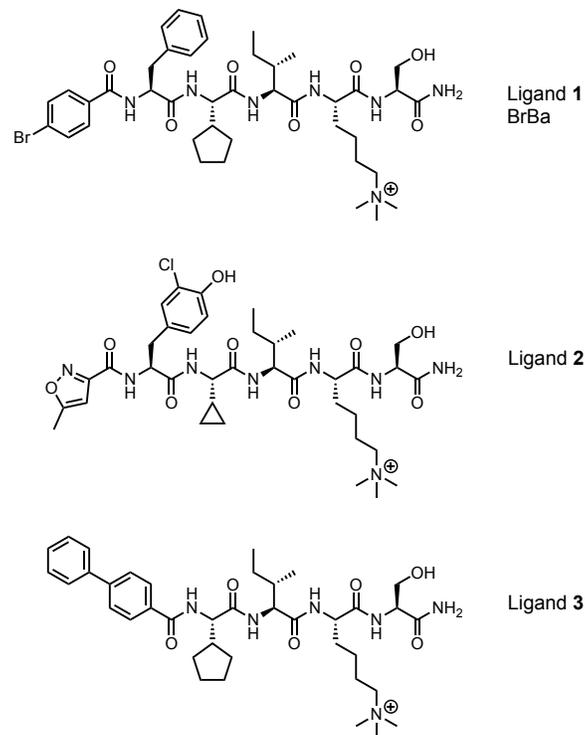


Figure S2. Structures of tested free ligands 1-3 to the CBX7-chromodomain used in ligand displacement assays in Figure 3a.

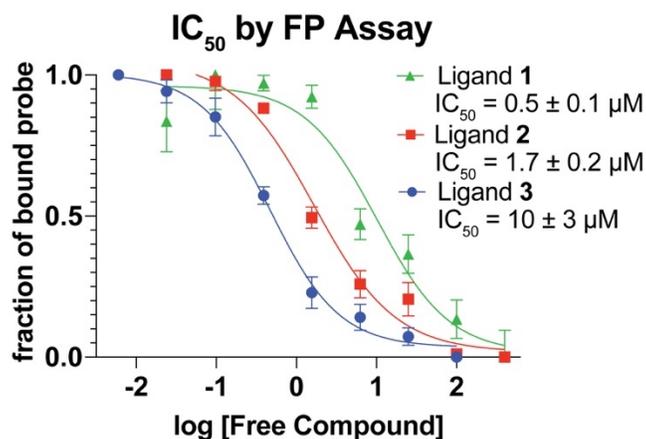


Figure S3. IC₅₀ of ligands 1-3 determined by a traditional fluorescence polarization (FP) assay.

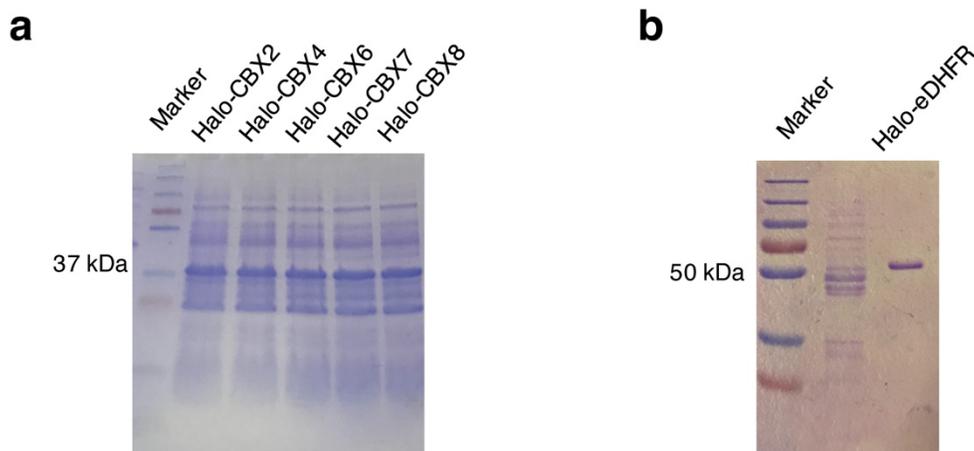


Figure S4. SDS-PAGE analysis of the protein expression. (a). Expression of HaloTag fusion of the CBX family Chromodomains (CBX2, CBX4, CBX6, CBX7, CBX8) used within crude lysates for ligand displacement assays in Fig. 3b. (b). Expression analysis in crude *E. coli* lysate and analysis of purified Halo-eDHFR.

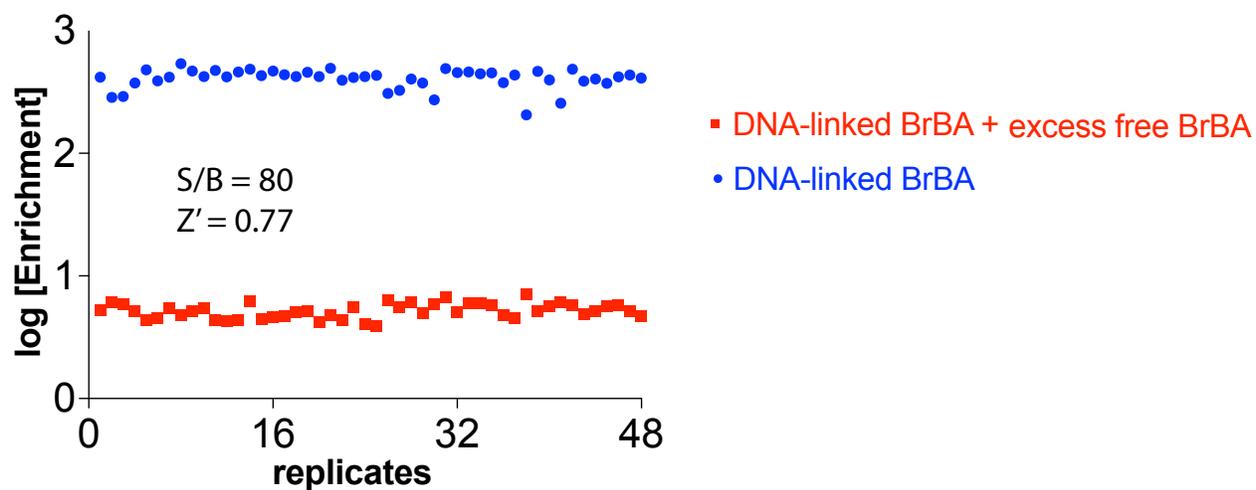


Figure S5. Assessment of the DNA-barcode robustness of the ligand displacement assay with CBX7 chromodomain in 2 wells.

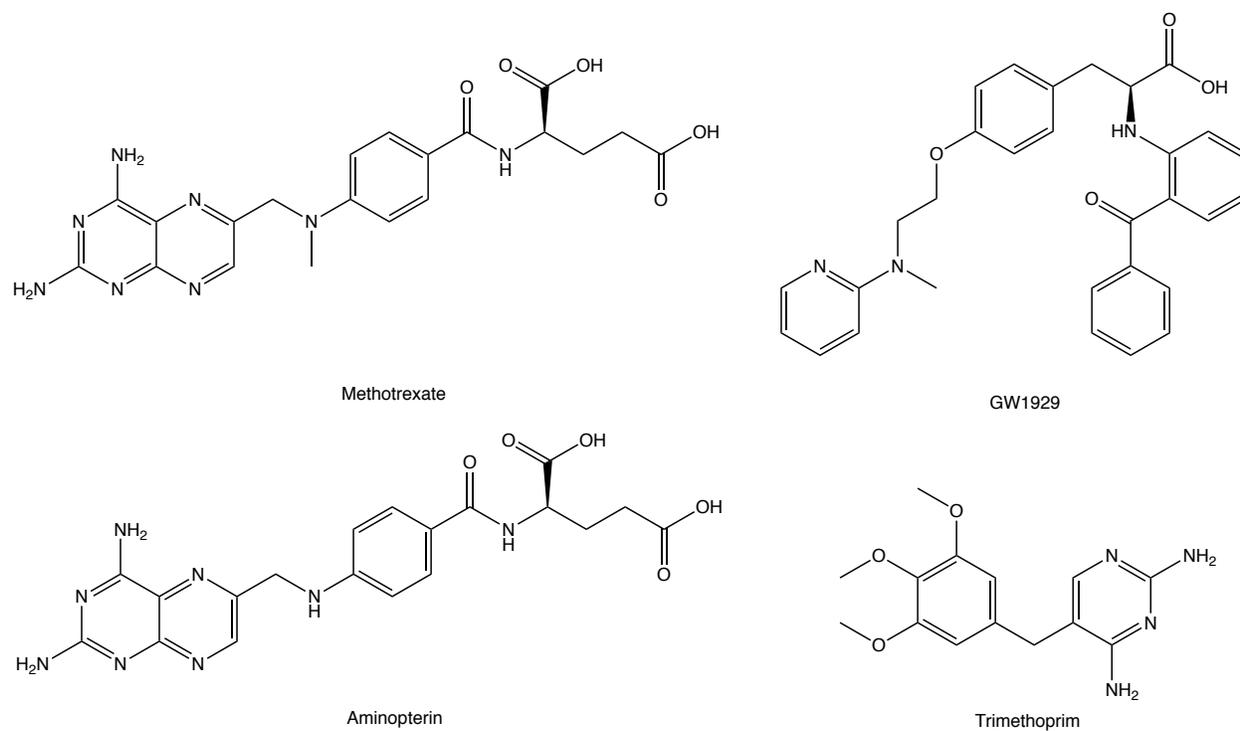


Figure S6. Structures of hit compounds from the LOPAC library screening

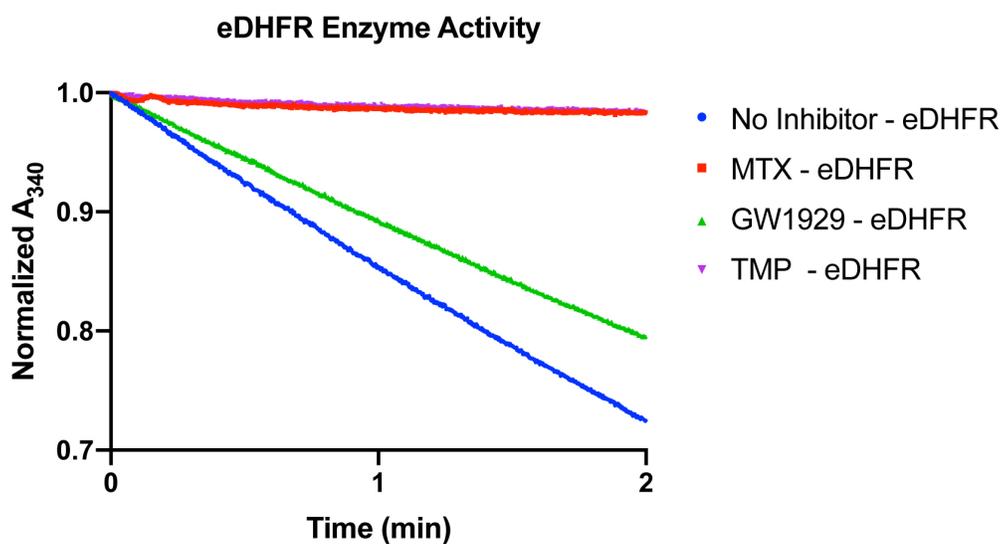


Figure S7. Real time assay of the inhibition of enzyme activity of eDHFR using 0.1 μM of the indicated inhibitors.

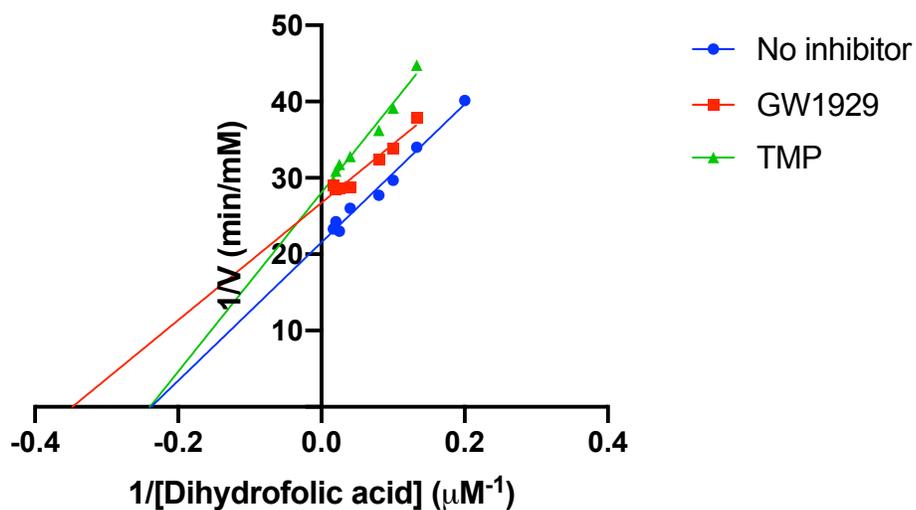


Figure S8. Lineweaver-Burk plot of inhibition of eDHFR with GW1929 and TMP using data presented in Fig. 4c.

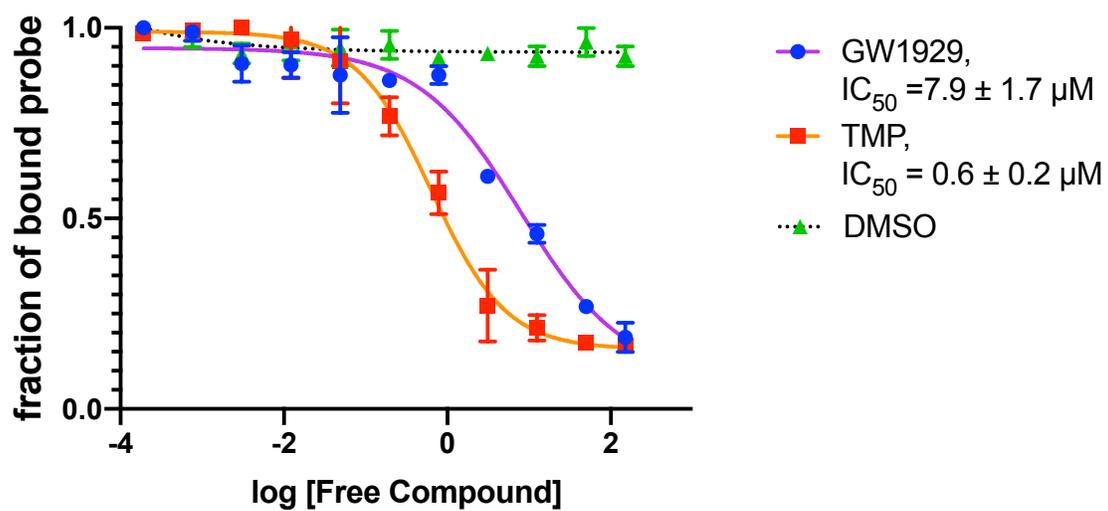


Figure S9. Fluorescence polarization assay for the determination of the IC_{50} of GW1929 using a fluorescein-conjugated methotrexate as a probe.

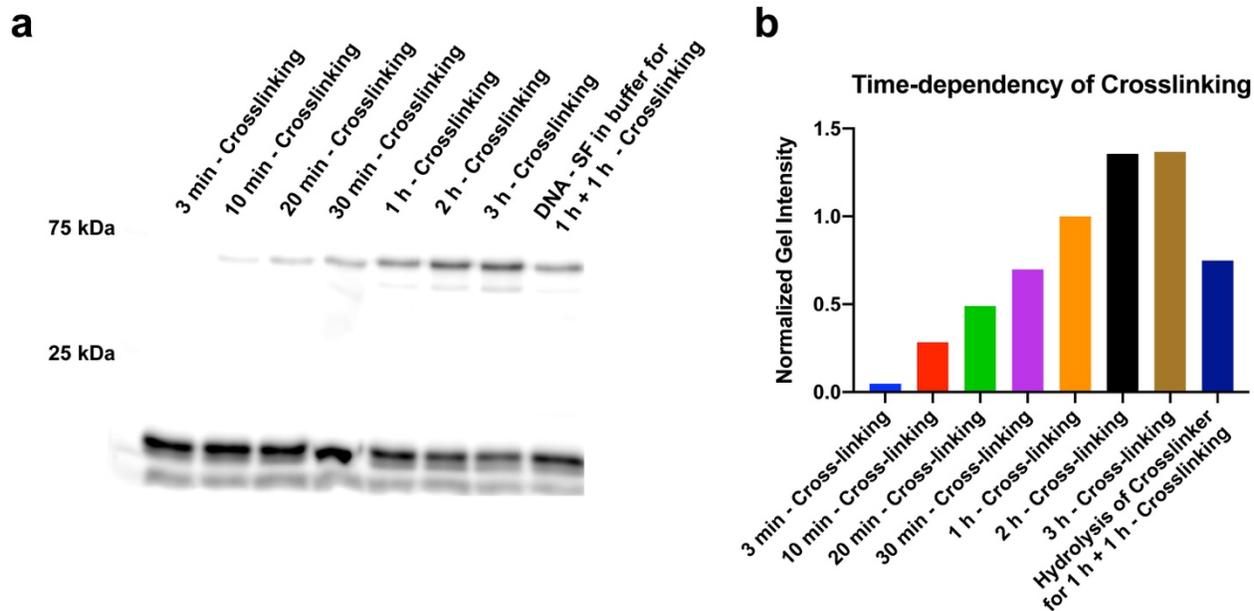


Figure S10. Time-dependency of crosslinking. (a) Crosslinking of DNA-linked BrBA (Ligand 1, Fig. S2) to Halo-CBX7-ChD using DNA-linked sulfonyl fluoride as the crosslinker. The crosslinking reaction was quenched at different time points (from 3 min to 3 h). A hydrolysis test in the final lane involved preincubation of the sulfonyl fluoride oligo in labeling buffer alone for 1 hour prior to addition to the labeling reaction. (b) Normalized intensity of crosslinked bands using background signal from a protein-free lane as 0% signal and the intensity of crosslinked band after 1 h – crosslinking as 100% signal.

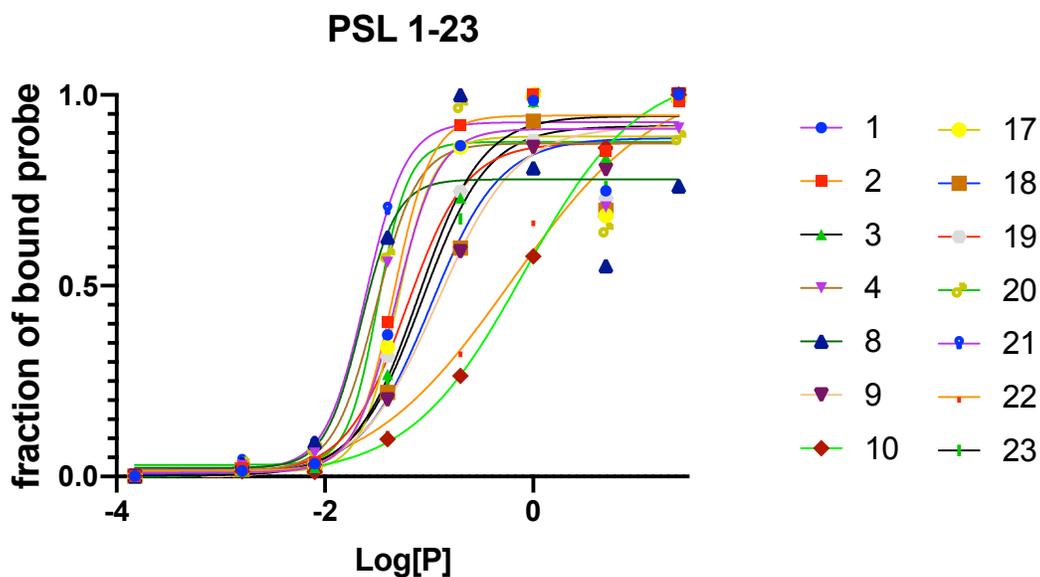


Figure S11. On-DNA EC_{50} of compound 1-23 in the PSL determined by DNA-based cross-linking assay

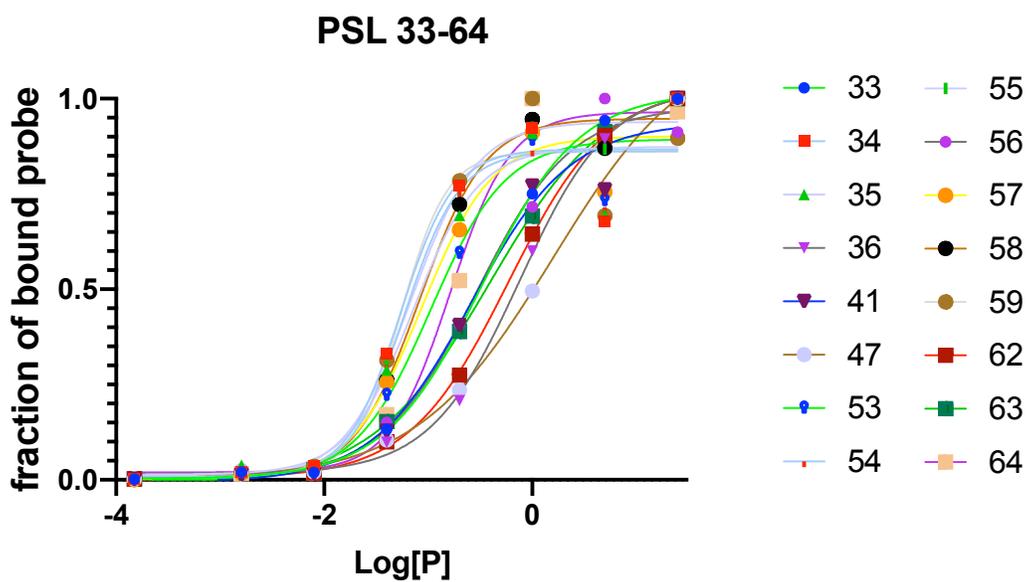


Figure S12. On-DNA EC_{50} of compound 33-64 in the PSL determined by DNA-based cross-linking assay

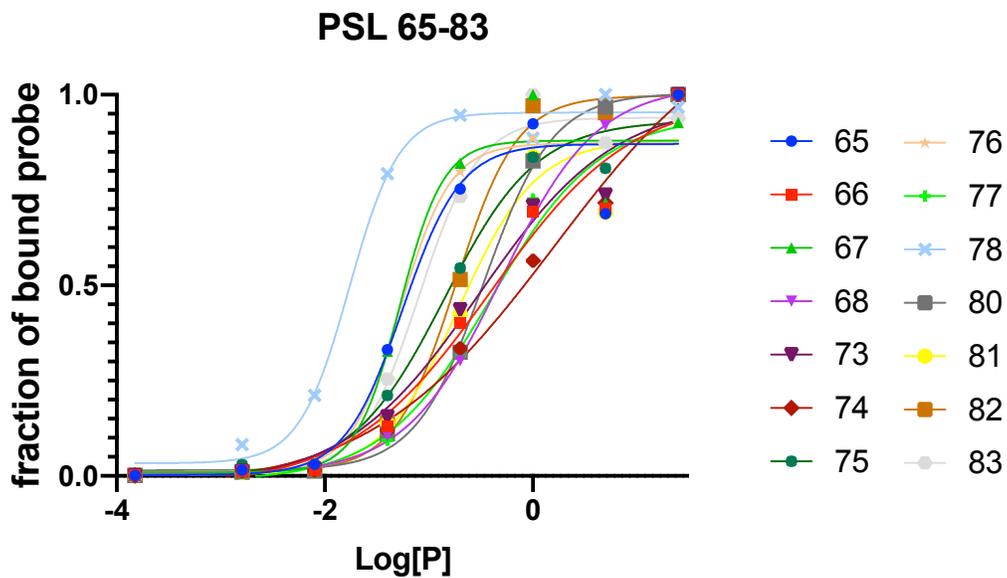


Figure S13. On-DNA EC_{50} of compound 65-83 in the PSL determined by DNA-based cross-linking assay

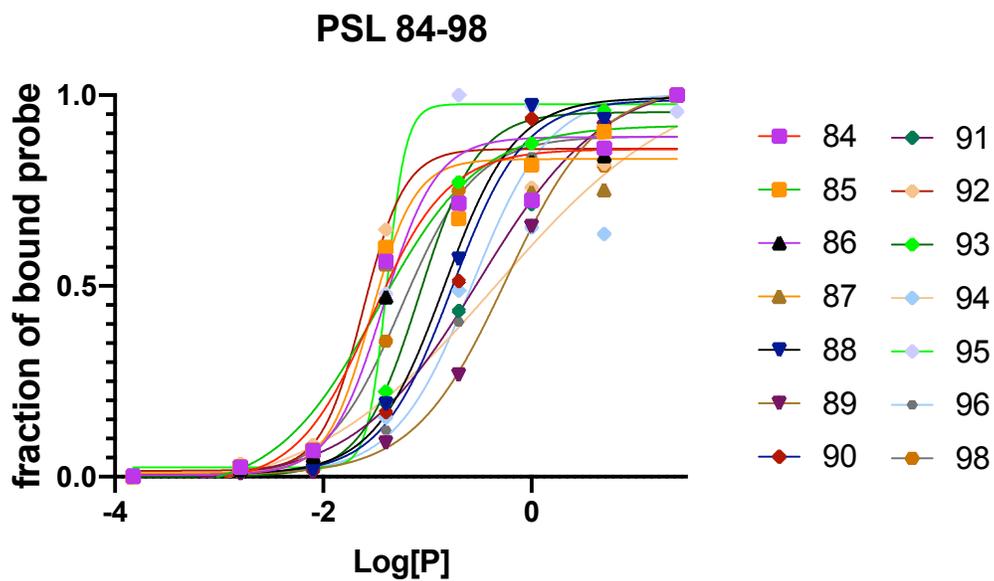


Figure S14. On-DNA EC_{50} of compound 84-98 in the PSL determined by DNA-based cross-linking assay. Compound 98 is the doped positive control (DNA-linked BrBA).

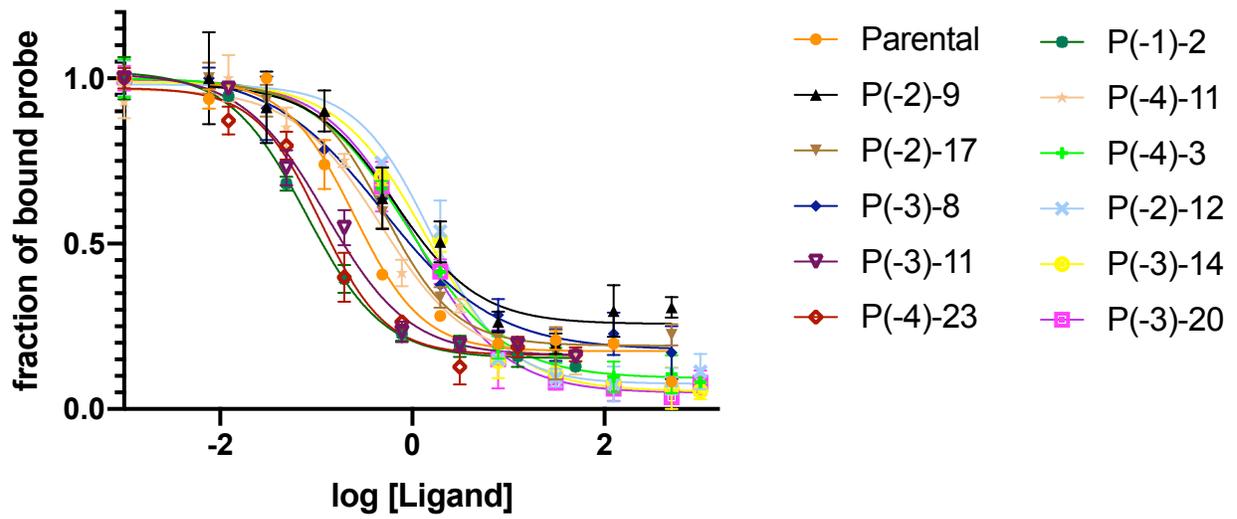


Figure S15. IC₅₀ values of off-DNA hits in binding assay to Halotag-CBX7-ChD. The IC₅₀ values were converted to K_d values using Cheng Prusoff equation.

5. Supplementary Tables

Table S1. Oligonucleotide sequences and modifications

Name	Sequences and modifications
ssDNAa-5'-C ₁₂ -NH ₂	/5AmMC12/ATGGTATCAAGCTTGCCACA
ssDNAa'-5'-FAM	/56-FAM/TGTGGCAAGCTTGATACCAT
ssDNAa'-3'-C ₁₂ -NH ₂ -5'-FAM	/56-FAM/TGTGGCAAGCTTGATACCATXXXXX
ssDNAa-linker-ssDNAb-5'-C ₁₂ -NH ₂	/5AmMC12/ATGGTATCAAGCTTGCCACA/iSp18/GTCGAGCTCTCTACTGCATA
ssDNAc'	TGACACCTTGCCCCGGTTC
ssDNAf1-5'-alkyne	/5Hexynyl/GCCTGATCCCCGCTGATTAT
ssDNAf2-5'-alkyne	/5Hexynyl/GATCTCTGTGAAGTTAGTGC
ssDNAf3-5'-alkyne	/5Hexynyl/GATCTCGATTATGCTCAAGG
ssDNAf4-5'-alkyne	/5Hexynyl/GAGACACTTATGGCTCATGT
ssDNAf5-5'-alkyne	/5Hexynyl/CGTCGTGCTGCGTGACTATA
ssDNAf1'-3'-biotin	ATAATCAGCGGAATCAGGC/3Bio/
ssDNAf2'-3'-biotin	GCACTAACTTCACAGAGATC/3Bio/
ssDNAf3'-3'-biotin	CCTTGAGCATAATCGAGATC/3Bio/
ssDNAf4'-3'-biotin	ACATGAGCCATAAGTGTCTC/3Bio/
ssDNAf5'-3'-biotin	TATAGTCACGCAGCACGACG/3Bio/

Table S2. MALDI analysis of chemically-modified oligonucleotides

Oligonucleotide	Calculated [M-H] ⁻	Observed [M-H] ⁻
ssDNAf1-5'-chloroalkane	6467.2	6468.8
ssDNAf2-5'-chloroalkane	6571.3	6571.6
ssDNAf3-5'-chloroalkane	6540.3	6542.0
ssDNAf4-5'-chloroalkane	6540.3	6541.7
ssDNAf5-5'-chloroalkane	6532.3	6534.2
ssDNAa-linker-ssDNAb-5'-non-ligand	12941.7	12941.5
ssDNAa-linker-ssDNAb-5'-Compound 4	13610.4	13611.3
ssDNAa-linker-ssDNAb-5'-Compound 5	13564.5	13567.1
ssDNAa-linker-ssDNAb-5'-Compound 6	13624.4	13626.3
ssDNAa-linker-ssDNAb-5'-Compound 7	13596.3	13599.0
ssDNAa-linker-ssDNAb-5'-TMP	13197.0	13197.6

Table S3. Observed masses of small molecules

Small Molecules	Calculated [M+H] ⁻	Observed [M+H] ⁺
Compound 1 (Figure S2)	789.33	788.74
Compound 2 (Figure S2)	792.39	792.06
Compound 3 (Figure S2)	694.43	694.12
Trimethoprim carboxylate-derivative (Scheme S8)	377.17	377.40

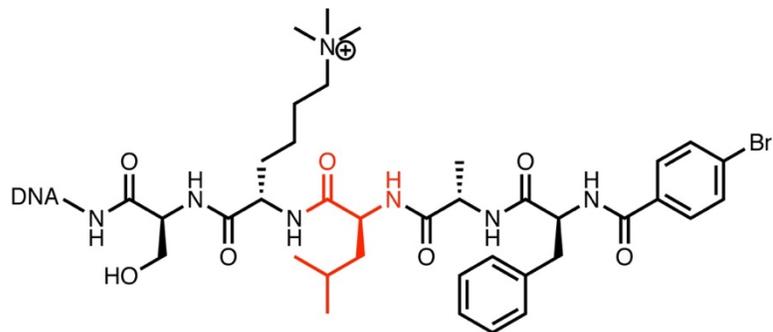
Table S4. On-DNA EC₅₀ values and enrichment of compounds within the Positional Scanning Library (PSL)

Compound ID in Figure S11-S14	Compound ID in Table S5-S8	On-DNA EC ₅₀ determined by cross-linking (nM)	Enrichment from selection against 250 nM protein
78	P(-4)-6	17	640
8	P(-1)-8	22	533
21	P(-1)-21	25	312
92	P(-4)-20	26	781
84	P(-4)-12	28	486
85	P(-4)-13	29	665
87	P(-4)-15	30	711
4	P(-1)-4	31	351
20	P(-1)-20	32	275
86	P(-4)-14	39	518
95	P(-4)-23	40	931
2	P(-1)-2	46	327
1	P(-1)-1	48	542
17	P(-1)-17	49	392
67	P(-3)-19	52	493
76	P(-4)-4	53	246
34	P(-2)-10	54	255
59	P(-3)-11	54	457
98	P(-3)-17	56	473
65	Parental	56	290
54	P(-3)-6	59	419

19	P(-1)-19	60	149
55	P(-3)-7	68	289
35	P(-2)-11	71	120
3	P(-1)-3	80	531
83	P(-4)-11	80	506
58	P(-3)-10	83	379
93	P(-4)-21	85	761
23	P(-1)-23	86	326
57	P(-3)-9	86	312
18	P(-1)-18	103	267
53	P(-3)-5	105	480
9	P(-1)-9	120	430
75	P(-4)-3	142	133
88	P(-4)-16	145	434
64	P(-3)-16	164	369
90	P(-4)-18	173	621
82	P(-4)-10	182	536
81	P(-4)-9	195	545
74	P(-4)-2	212	413
41	P(-2)-17	260	286
96	P(-4)-24	270	430
56	P(-3)-8	290	272
73	P(-4)-1	310	457
91	P(-4)-19	315	361
33	P(-2)-9	330	210
80	P(-4)-8	340	469
77	P(-3)-15	390	146
63	P(-4)-5	390	472
66	P(-3)-18	394	329
94	P(-4)-22	451	479
68	P(-3)-20	495	275
89	P(-4)-17	570	332
62	P(-3)-14	580	169

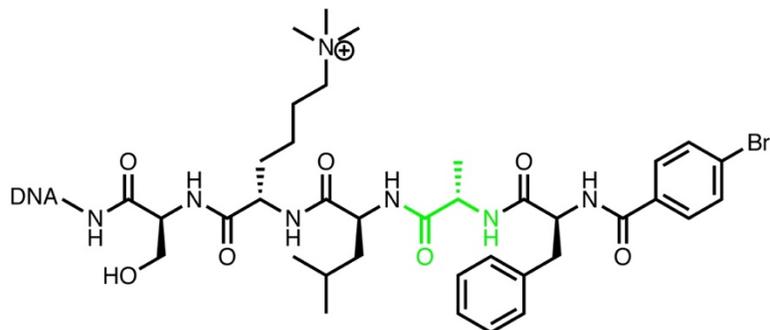
22	P(-1)-22	615	185
36	P(-2)-12	760	163
10	P(-1)-10	810	257
47	P(-2)-23	1590	143

Table S5. Building blocks of the P (-1) Position of the PSL.



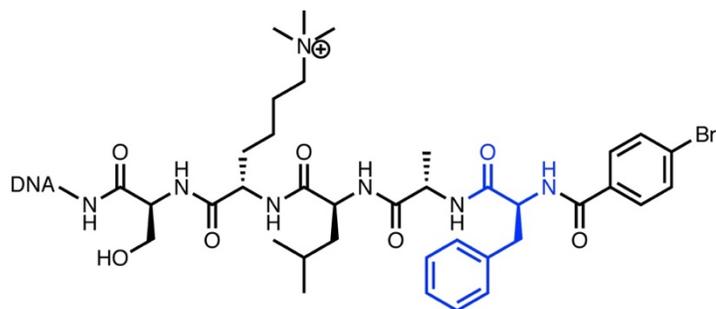
1		2		3		4	
5		6		7		8	
9		10		11		12	
13		14		15		16	
17		18		19		20	
21		22		23		24	

Table S6. Building blocks of the P (-2) Position of the PSL



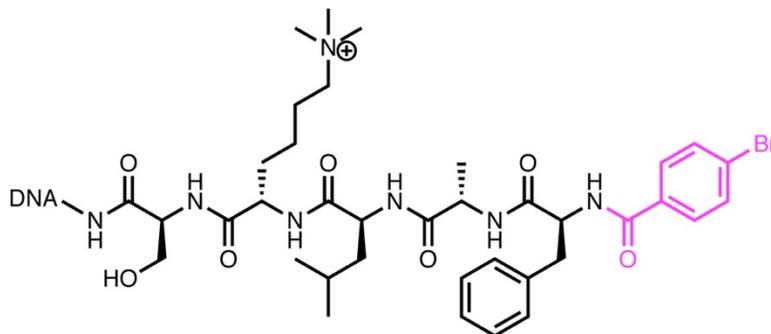
1		2		3		4	
5		6		7		8	
9		10		11		12	
13		14		15		16	
17		18		19		20	
21		22		23		24	

Table S7. Building blocks of the P (-3) Position of the PSL



1		2		3		4	
5		6		7		8	
9		10		11		12	
13		14		15		16	
17		18		19		20	
21		22		23		24	

Table S8. Building blocks of the P (-4) Position of the PSL



1		2		3		4	
5		6		7		8	
9		10		11		12	
13		14		15		16	
17		18		19		20	
21		22		23		24	

6. Supplementary References

1. Cai, B., Kim, D., Akhand, S., Sun, Y., Cassell, R. J., Alpsoy, A., ... & Krusemark, C. J. (2019). Selection of DNA-encoded libraries to protein targets within and on living cells. *Journal of the American Chemical Society*, 141, 17057-17061.
2. Hong, V., Presolski, S. I., Ma, C., & Finn, M. A. G. (2009). Analysis and optimization of copper-catalyzed azide–alkyne cycloaddition for bioconjugation. *Angewandte Chemie*, 121, 10063-10067.
3. Miller, L. W., Cai, Y., Sheetz, M. P., & Cornish, V. W. (2005). In vivo protein labeling with trimethoprim conjugates: a flexible chemical tag. *Nature methods*, 2, 255-257.