Supporting Information

Modular assembly and encoding strategies for dual-display DNA-encoded chemical libraries

Sebastian Oehler,[a] Louise Plais,[a] Gabriele Bassi,[a] Dario Neri,*[b] Jörg Scheuermann.*[a]

[a] Department of Chemistry and Applied Biosciences
ETH Zürich
Vladimir-Prelog-Weg 3, 8093 Zürich (Switzerland)
E-mail: jorg.scheuermann@pharma.ethz.ch, dario.neri@pharma.ethz.ch
[b] Philochem AG
Libernstrasse 3, 8112, Oteffingen (Switzerland)
[*] Corresponding authors

Author contributions
S.O., D.N. and J.S. wrote the manuscript. S.O. created the figures. S.O., D.N. and J.S. designed the experiments which were implemented by S.O. with support of L.P. and G.B. for DEL selections.
1. Supplementary Figures

**Figure S1.** Conversion of dimeric structures after hairpin library assembly into hairpin structures by a fast heat-cool cycle (95 °C for 2 min, on ice for 2 min).
Figure S2. LC-MS chromatogram including the MS trace (B) of the 1+1 hairpin format (A) after assembly and ligation. C) Deconvoluted masses of each peak: Expected masses = 7372 (Adapter), 14971 (Elib2) and 32101 (Hairpin construct); Observed masses = 7373 (Adapter), 14973 (Elib2) and 32110 (Hairpin construct).
Figure S3. LC-MS chromatogram including the MS trace (B) of the 1+1 circular format (A) after assembly, polymerase-mediated fill-in and ligation. C) Deconvoluted masses of each peak: Expected masses = 14971 (Elib2), 16949 (S1sh) and 32399 (Circular template); Observed masses = 14973 (Elib2), 16954 (S1sh) and 32403 (Circular template).
Figure S4. LC-MS chromatogram including the MS trace of assembled and encoded 1+1 circular constructs as crude (A) or after agarose gel purification (B). The mass-over-charge (m/z) ratios and deconvoluted masses of the circular template peaks are shown: Expected mass = 32399; Observed masses = 32402 (crude) and 32404 (gel-purified).
Figure S5. LC-MS chromatogram including the MS trace (B) of the 1+1 linear format (A) after assembly and encoding (fill-in and ligation.) C) Deconvoluted masses of each peak: Expected masses = 14971 (Elib2), 16949 (S1sh) and 31977 (Linear template); Observed masses = 14973 (Elib2), 16954 (S1sh), 31984 (Linear template) and 31680 (Linear template -T = thymidine monophosphate, -304 Da).
Figure S6. LC-MS chromatogram including the MS trace of assembled and encoded 1+1 linear constructs as crude (A) or after agarose gel purification (B). The mass-over-charge (m/z) ratios and deconvoluted masses of the linear template peaks are shown: Expected mass = 31977; Observed masses = 31984 (crude) and 31985 (gel-purified).

Figure S7. EcoRI digestion of the 1+1 circular (A) and 1+1 linear (B) construct with one cleavage site each. Digestion patterns were visualized by agarose gel.
Figure S8. LC-MS chromatogram including the MS trace of 1+1 linear constructs which were assembled and encoded at low (0.5 µM, A) and high (2 µM, B) concentrations. The deconvoluted masses of the linear template peaks are shown: Expected mass = 31977; Observed masses = 31984 (A) and 31983 (B). The side product was identified as extended relay primer: Expected mass = 24006, Observed mass = 24009 (A) and 24008 (B).
Figure S9. LC-MS chromatogram including the MS trace (B) of the 1+2 hairpin format (A) after assembly and ligation. C) Deconvoluted masses of each peak: Expected masses = 7437 (Adapter), 23660 (GB2) and 40798 (Hairpin construct); Observed masses = 7438 (Adapter), 23663 (Elib2) and 40808 (Hairpin construct).

Figure S10. Splint-ligation of the 3’-modified non-complementary sublibrary (S1sh; 1) to receive a single stranded sublibrary comprising two coding regions (S1lo; 4).
Figure S11. Assembly and encoding of a 3'-modified sublibrary with one coding region and a 5'-modified sublibrary with two coding regions to obtain the 1+2 circular construct. Construct formation was visualized by agarose gel.

Figure S12. Assembly and encoding of a 3'-modified sublibrary with one coding region and a 5'-modified sublibrary with two coding regions to obtain the 1+2 linear construct. Construct formation was visualized by agarose gel.

Figure S13. EcoRI digestion of the 1+2 circular (A) and 1+2 linear (B) construct with one cleavage site each. Digestion patterns were visualized by agarose gel (the cleavage products of the linear construct had similar sizes, i.e. 64 and 68 bp, and thus appeared as one band).
Figure S14. LC-MS chromatogram including the MS trace (B) of the 1+2 circular format (A) after assembly, polymerase-mediated fill-in and ligation. C) Deconvoluted masses of each peak: Expected masses = 16949 (S1sh), 23660 (GB2) and 41013 (Circular template); Observed masses = 16953 (S1sh), 23662 (GB2) and 41020 (Circular template).
Figure S15. LC-MS chromatogram including the MS trace (B) of the 1+2 linear format (A) after assembly, polymerase-mediated fill-in and ligation. C) Deconvoluted masses of each peak: Expected masses = 16949 (S1sh), 24006 (side product = extended relay primer), 23660 (GB2) and 40591 (Linear template); Observed masses = 16953 (S1sh), 24009 (side product), 23663 (GB2) and 40605 (Linear template).
Figure S16. Assembly and encoding of a 3’-modified sublibrary with two coding regions and a 5’-modified sublibrary with one coding region to obtain the 2+1 circular construct. Construct formation was visualized by agarose gel.

Figure S17. Assembly and encoding of a 3’-modified sublibrary with two coding regions and a 5’-modified sublibrary with one coding region to obtain the 2+1 linear construct. Construct formation was visualized by agarose gel.

Figure S18. EcoRI digestion of the 2+1 circular (A) and 2+1 linear (B) construct with one cleavage site each. Digestion patterns were visualized by agarose gel.
Figure S19. LC-MS chromatogram including the MS trace (B) of the 2+1 circular format (A) after assembly, polymerase-mediated fill-in and ligation. C) Deconvoluted masses of each peak: Expected masses = 14971 (Elib2), 29880 (S1lo) and 45341 (Circular template); Observed masses = 14973 (Elib2), 29886 (S1lo) and 45350 (Circular template).
Figure S20. LC-MS chromatogram including the MS trace (B) of the 2+1 linear format (A) after assembly, polymerase-mediated fill-in and ligation. C) Deconvoluted masses of each peak: Expected masses = 14971 (Elib2), 29880 (S1lo) and 44919 (Linear template); Observed masses = 14974 (Elib2), 29887 (S1lo) and 44928 (Linear template).
**Figure S21.** Assembly and encoding of a 3’-modified sublibrary with two coding regions and a 5’-modified sublibrary with two coding regions to obtain the 2+2 circular construct. Construct formation was visualized by agarose gel.

**Figure S22.** Assembly and encoding of a 3’-modified sublibrary with two coding regions and a 5’-modified sublibrary with two coding regions to obtain the 2+2 linear construct. Construct formation was visualized by agarose gel.

**Figure S23.** EcoRI digestion of the 2+2 circular (A) and 2+2 linear (B) construct with one cleavage site each. Digestion patterns were visualized by agarose gel.
Figure S24. LC-MS chromatogram including the MS trace (B) of the 2+2 circular format (A) after assembly, polymerase-mediated fill-in and ligation. C) Deconvoluted masses of each peak: Expected masses = 23660 (GB2), 29880 (S1lo) and 53956 (Circular template); Observed masses = 23663 (GB2), 29886 (S1lo) and 53995 (potentially circular template with Na⁺).
Figure S25. LC-MS chromatogram including the MS trace (B) of the 2+2 linear format (A) after assembly, polymerase-mediated fill-in and ligation. C) Deconvoluted masses of each peak: Expected masses = 23660 (GB2), 29880 (S1lo) and 53534 (Linear template). Observed masses = 23665 (GB2), 29888 (S1lo) and 53554 (potentially linear template).
Figure S26. PCR amplifiability of 1+2 hairpin (A), circular (B) and linear (C) library formats. Dilution series of the constructs ranging from $10^{10}$ to 0 DNA molecules as input allowed to visualize the input threshold for productive PCR amplification (indicated by the dotted lines) via agarose gel.
Figure S27. PCR amplifiability of 2+1 circular (A) and linear (B) library formats. Dilution series of the constructs ranging from $10^{10}$ to 0 DNA molecules as input allowed to visualize the input threshold for productive PCR amplification (dotted lines) via agarose gel.

Figure S28. PCR amplifiability of 2+2 circular (A) and linear (B) library formats. Dilution series of the constructs ranging from $10^{10}$ to 0 DNA molecules as input allowed to visualize the input threshold for productive PCR amplification (as indicated by the dotted lines) via agarose gel.
Figure S29. Chemical structures with the respective code identities of building blocks that were coupled to the 3'- (left) and 5'- (right) modified sublibraries.

Figure S30. Fingerprints of the naive ESAC library (A) and of the library after selection against “empty” beads on which no protein has been immobilized (B). Each library member is represented by a dot for which size and color shows the normalized sequence count. CodeA:1 was spiked-in 100-fold. Sequencing and selections were performed in duplicates. Total counts (TC) and average counts (AC): A (TC=153’191, AC=2.66; TC=139’224, AC=2.42) and B (TC=61’474, AC=1.07; TC=79’849, AC=1.38).
Figure S31. Selection fingerprints of the ESAC library after screening against carbonic anhydrase IX (CAIX) immobilized on beads. Each individual dot represents one library member for which size and color shows the normalized sequence count. CodeA:1 was spiked-in 100-fold. The two-dimensional planes show all library members with acetazolamide as constant 3’-coupled building block (CodeA:2) while CodeB:493 codes for the known binder fragment[1] which led to AAZ+. Selections were performed in duplicates. Total counts (TC) and average counts (AC): TC=152'212, AC=2.65; TC=145'157, AC=2.52.

Figure S32. Selection fingerprints of the ESAC library after screening against alpha-glycoprotein 1 (AGP) immobilized on beads. Each individual dot represents one library member for which size and color shows the normalized sequence count. CodeA:1 was spiked-in 100-fold. The two-dimensional planes show all library members with a furan derivative (AGP ligand fragment) as constant 3’-coupled building block (CodeA:3) and CodeB:117 coding for the known AGP binder fragment[1]. Selections were performed in duplicates. Total counts (TC) and average counts (AC): TC=74'819, AC=1.30; TC=119'506, AC=2.08.
Figure S33. Fingerprints of the naive hairpin library (A) and of the library after selection against “empty” beads (B, no protein immobilized). Each library member is represented by a dot for which dot size and color shows the normalized sequence count. CodeA:1 was spiked-in 100-fold. Sequencing and selections were performed in duplicates. Total counts (TC) and average counts (AC): A (TC=158’236, AC=2.75; TC=163’849, AC=2.85) and B (TC=143’485, AC=2.49; TC=145’146, AC=2.52).

Figure S34. Selection fingerprints of the hairpin library after screening against carbonic anhydrase IX (CAIX) immobilized on beads. Each individual dot represents one library member for which size and color shows the normalized sequence count. CodeA:1 was spiked-in 100-fold. The two-dimensional planes show all library members with acetazolamide as constant 3’-coupled building block (CodeA:2) while CodeB:493 encodes the known binder fragment[1] which led to AAZ+∗. Selections were performed in duplicates. Total counts (TC) and average counts (AC): TC=170’905, AC=2.97; TC=159’875, AC=2.78.
Figure S35. Selection fingerprints of the hairpin library after screening against alpha-glycoprotein 1 (AGP) immobilized on beads. Each individual dot represents one library member for which size and color shows the normalized sequence count. CodeA:1 was spiked-in 100-fold. The two-dimensional planes show all library members with a furan derivative (AGP binder fragment) as constant 3’-coupled building block (CodeA:3). CodeB:117 codes for the counterpart of the known AGP ligand[1]. Selections were performed in duplicates. Total counts (TC) and average counts (AC): TC=123’509, AC=2.15; TC=140’708, AC=2.45.

A) Circular construct - Naive library

B) No protein selections

Figure S36. Fingerprints of the naive circular library (A) and of the library after selection against “empty” beads (B, no protein immobilized). Each library member is represented by a dot for which dot size and color shows the normalized sequence count. CodeA:1 was spiked-in 100-fold. Sequencing and selections were performed in duplicates. Total counts (TC) and average counts (AC): A (TC=161’827, AC=2.81; TC=157’268, AC=2.73) and B (TC=139’175, AC=2.42; TC=151’821, AC=2.64).

24
Figure S37. Selection fingerprints of the circular library after screening against carbonic anhydrase IX (CAIX) immobilized on beads. Each individual dot represents one library member for which size and color shows the normalized sequence count. CodeA:1 was spiked-in 100-fold. The two-dimensional planes show all library members with acetazolamide as constant 3' coupled building block (CodeA:2). Selections were performed in duplicates. Total counts (TC) and average counts (AC): TC=137'153, AC=2.38; TC=140'261, AC=2.44.

Figure S38. Selection fingerprints of the circular library after screening against alpha-glycoprotein 1 (AGP) immobilized on beads. Each individual dot represents one library member for which size and color shows the normalized sequence count. CodeA:1 was spiked-in 100-fold. The two-dimensional planes show all library members with a furan derivative (AGP binder fragment) as constant 3' coupled building block (CodeA:3). Selections were performed in duplicates. Total counts (TC) and average counts (AC): TC=121'242, AC=2.10; TC=125'101, AC=2.18.
**Figure S39.** Fingerprints of the naive linear library (A) and of the library after selection against "empty" beads (B, no protein immobilized). Each library member is represented by a dot for which dot size and color shows the normalized sequence count. CodeA:1 was spiked-in 100-fold. Sequencing and selections were performed in duplicates. Total counts (TC) and average counts (AC): A (TC=104'721, AC=1.82; TC=101'559, AC=1.77) and B (TC=91'358, AC=1.59; TC=98'906, AC=1.72).

**Figure S40.** Selection fingerprints of the linear library after screening against carbonic anhydrase IX (CAIX) immobilized on beads. Each individual dot represents one library member for which size and color shows the normalized sequence count. CodeA:1 was spiked-in 100-fold. The two-dimensional planes show all library members with acetazolamide as constant 3'-coupled building block (CodeA:2). Selections were performed in duplicates. Total counts (TC) and average counts (AC): TC=1'010'446, AC=17.6; TC=873'566, AC=15.2.
Figure S41. Selection fingerprints of the linear library after screening against alpha-glycoprotein 1 (AGP) immobilized on beads. Each individual dot represents one library member for which size and color shows the normalized sequence count. CodeA:1 was spiked-in 100-fold. The two-dimensional planes show all library members with a furan derivative (AGP binder fragment) as constant 3’-coupled building block (CodeA:3). Selections were performed in duplicates. Total counts (TC) and average counts (AC): TC=79’330, AC=1.38; TC=87’308, AC=1.52.
**Figure S42.** Electrophoretic mobility shift assay to measure the affinity of the AGP ligand pair (A2/B117) presented on different DNA structures. AGP was titrated against the AGP ligand pair (200 nM) for which a band shift could only be observed for the ESAC and hairpin format as highlighted by the black arrows. 20 µM AGP without DNA (-) as well as the respective on-DNA AGP-ligand pair pre-mixed with 20 µM bovine serum albumin (BSA) served as negative control.
2. Materials and Methods

Chemicals and solvents used in this study were purchased from Merck, VWR, TCI, Sigma-Aldrich, Fisher chemicals, or abcr GmbH in analytical or HPLC grade. Custom-made oligonucleotides were purchased from LGC Biosearch Technologies, ethanol precipitated, characterized via LC-MS analysis and quantified with a Nanodrop 2000c Spectrophotometer. Ligation buffer (#B0202S), T4 DNA Polymerase (#M0203L), T4 DNA Ligase (#M0202S) and high-fidelity Phusion DNA polymerase (#M0530L) were purchased from New England Biolabs. Kits for PCR purification and gel extraction (#28506) were purchased from Qiagen.

Oligomer sequences

| Oligomer | Sequence (5’ → 3’) |
|----------|-------------------|
| 5’-modified sublibrary with 1 code (Elib2)[3] | NH₂-(CH₂)₆-GGAGCTTCTGAATTTCTGTGCTGXXXXXCGAGTCCCATGGCGGCAGC (Codes: e.g. AGACTC) |
| 5’-modified sublibrary with 2 codes (GB2)[3] | NH₂-(CH₂)₆-GGAGCTTCTGAATTTCTGTGCTGXXXXXCGAGGTCAGGCGACCTGCTGXXXXXCGAGTCCCATGGCGGCAGC (codes: e.g. TCAGAT (code1) and TTCACG (code2)) |

Circular and linear constructs

| Oligomer | Sequence |
|----------|----------|
| 3’-modified sublibrary with 1 code (S1sh) | Phos-AAGGAAGGCTAGACGTATCTCACTXXXXXGTCGCTGTTCTTAAATGTCTTCCAGGG - (CH₂)₆-NH₂ |
| 3’-modified sublibrary with 2 codes (S1lo) | TGACTCCCCAATCGATGTGTCAGXXXXXGAAAGGTCAGAACAAAGGAAGGCTAGACGTTCTCACTXXXXXGTCGCTGTTCTTAAATGTCTTCCAGGGG-(CH₂)₆-NH₂ (codes: e.g. TTCACG (code2) and GAACGT (code1)) |

Relay primer (near modified library ends)

| Oligomer | Sequence |
|----------|----------|
| 3’-modified sublibrary with 1 code (S1sh) | Phos-GCACACAGAATTCAGAAGCTCCATCTCTCAAGACTTAAGACACAG |
| Distal relay primer (circular) | Phos-TGAGATACGTCTAGCCTCTCCTTddGCTGCGCCATGGGGACTCG (for S1sh+Elib2) Phos-TGAGATACGCTCTAGCCTCCTTddGCGCCATGGGGACTCG (for S1sh+GB2) Phos-GACCACATCGATTTGGGAGTCAAddGCTCGCCATGGGGACTCG (for S1lo+Elib2) Phos-GACCACATCGATTTGGGAGTCAddGCGCCATGGGGACTCG (for S1lo+GB2) |
| Terminal relay primer (linear) | GCTGCGCCATGGGGACTCG (for Elib2) GCGCCATGGGGACTCG (for GB2) |

Hairpin construct

| Oligomer | Sequence |
|----------|----------|
| 3’-modified sublibrary (S4) | Phos-CCTGTCTTTACTGXXXXXCTTCGAGGCTTAAGTCTTddAGAATTCAGAAGCTCC-(CH₂)₆-NH₂ |
| Adapter | CAGTAAGACACAGGGCTGCGCCAT/dd (for Elib2)  
          | CAGTAAGACACAGGGCGCCATGGG (for GB2) |

### ESAC library

3'-modified sublibrary with 1 code (Elib4)

| CAGTAAGACACAGGGCTGCGCCAT/dd (for Elib2) |
|-----------------------------------------|
| CAGTAAGACACAGGGCGCCATGGG (for GB2)     |

### PCR primers

| Primer Type | Sequence Details |
|-------------|------------------|
| PCR1_fw_Circ&Lin (with Elib2) | CAGACGTGTGCTCTTCCGATCXXXXXXGCTGCGCCATGGGACTC  
                       | (codes: ACGTTG, ACTGAT, AGACTA, ACTCTG) |
| PCR1_fw_Circ&Lin (with GB2) | CAGACGTGTGCTCTTCCGATCXXXXXXGCSCCACTGGGACTC  
                         | (codes: TATCGA, GATCT) |
| PCR1_rv_Circ&Lin (with S1sh) | TACACGACGCTCTTCCGATCTXXXXXAAAGGAAGGCTAGACGTATCTCTC |
| PCR1_rv_Circ&Lin (with S1lo) | TACACGACGCTCTTCCGATCTXXXXXGCTTCTGAATTTCTGTGCTG |
| PCR1_fw_hairpin | TACACGACGCTCTTCCGATCTXXXXXGCTTCTGAATTTCTGTGCT |
| PCR1_rv_hairpin (with Elib2 and GB2) | TACACGACGCTCTTCCGATCTXXXXXGCTTCTGAATTTCTGTGCT |
| PCR1_fw_ESAC | TACACGACGCTCTTCCGATCTXXXXXXGCTGCGCCATGGGACTC  
                         | (codes: ACGTTG, ACTGAT, AGACTA, ACTCTG) |
| PCR1_rv_ESAC | CAGACGTGTGCTCTTCCGATCXXXXXXGCTGCGCCATGGGACTC  
                      | (codes: CGATAT, CTGACA) |
| PCR2_a | AATGATACGGCGACCACCGAGATCTACACTCTTCTTCTTCTCCGATCT |
| PCR2_b | CAACGACTGGAACGGCATTACGAGATXXXXXGCTGACTGGAAGTTCAGACGTCTCTTCTCCGATC  
                      | TCCGATC |
|               | (codes: ACATCG, CAGAAT, TCTCCGGA, AATGAGCG, GTACAT) |

\[ d = \text{abasic spacer, XXXXX = coding region} \]

### Construct sequences and forward/reverse PCR1 primer annealing sites

Circular and linear (S1sh+Elib2 = 1+1)

\[ \text{CAGACGTGTGCTCTTCCGATCXXXXXXGCTGCGCCATGGGACTC} \]
\[ \text{TACACGACGCTCTTCCGATCTXXXXXAAAGGAAGGCTAGACGTATCTCTC} \]

Circular and linear (S1lo+Elib2 = 2+1)

\[ \text{CAGACGTGTGCTCTTCCGATCXXXXXXGCTGCGCCATGGGACTC} \]
\[ \text{TACACGACGCTCTTCCGATCTXXXXXAAAGGAAGGCTAGACGTATCTCTC} \]
Circular and linear (S1sh+GB2 = 1+2)

Circular and linear (S1lo+GB2 = 2+2)

Hairpin (S4+Elib2 = 1+1)

Hairpin (S4+GB2 = 1+2)

ESAC

Encoding of 3’-modified sublibraries for model selections (5’-modified as published\[2\])

| CodeA | SMILES of building block | Circular&Linear (S1sh) | Hairpin (S4) | ESAC (Elib4) |
|-------|--------------------------|-----------------------|--------------|--------------|
| 1     | CC(O)=O                 | GAACTC                | TTCACG       | TGCTTCAA     |
| 2     | NS(C(S1)=NN=C1NC(CCC(O)=O)=O)(=O)=O | CAGAAC       | CAGAAC       | ATCTCCAA     |
| 3     | OC(CCC1=CC=C(C2=CC=CC=C2(F)(F)F)O1)=O | CTGTTC       | CTGTTC       | TGGAGGAA     |
| 4     | N=C(N)C1=CC=C(CNC(CCC(O)=O)=O)C=C1 | TAGCAG       | TAGCAG       | GATAGGAA     |
| 5     | OC(C1=CC=C(C2=CC=NC[2@H](J)C)CC(N2)=O)=O | AGTCGA       | ACTGCA       | TTATCTAA     |

On-DNA synthesis

Ethanol precipitation

DNA, typically dissolved in aqueous solutions, was precipitated by the addition of 10 v/v% 5 M NaCl\[aq\] or 2.5 M acetic acid buffer pH 5 (not before coupling reactions), followed by 3.5x volumes absolute ethanol (EtOH). The mixture was thoroughly vortexed and left at -20 °C overnight to receive the DNA pellet by centrifugation (16’100 rcf at 4 °C for 1 hour, discard supernatant) which was dried in a SpeedVac vacuum concentrator.
EDC-sNHS method (for carboxylic acid building blocks: acetic acid, acetazolamide derivative (4-oxo-4-((5-sulfamoyl-1,3,4-thiadiazol-2-yl)amino)butanoic acid) and furan derivative (3-[5-(2-(Trifluoromethyl)phenyl)furan-2-yl]propionic acid))

The carboxylic acid was dissolved in DMSO at a concentration of 100 mM of which 42 µL were mixed with 90 µL DMSO, 20 µL 100 mM sNHS (in DMSO:water 2:1) and 40 µL 100 mM EDC (in DMSO). The acid was pre-activated for 10 min at 37 °C before adding the mixture to 5 nmol DNA in 50 mM TEA HCl pH 10. The reaction proceeded at 37 °C for 3-5 h for which the conversion could be followed by LC-MS analysis. DNA was EtOH precipitated and purified via HPLC.

HATU method (for carboxylic acid building blocks: 4-(4-aminoanilino)-4-oxobutanoic acid and (R)-4,5-dihydrobenzodiazepinone (THBD[4])

The pre-activation was performed by mixing 40 µL 100 mM carboxylic acid (in DMSO) with 13.3 µL 300 mM DIPEA (in DMSO) and 8.2 µL 500 mM HATU (in NMP, prepared fresh) for subsequent incubation at rt for 10 min (for (R)-THBD added 260 µL DMSO and direct addition of mixture to DNA was necessary to avoid multiple small molecule additions). The mixture was added to 5 nmol DNA in 100 mM MOPS 1M NaCl buffer pH 8. After 3-6 h the DNA was EtOH precipitated and purified by HPLC.

HPLC purifications of oligomers

A preparative reverse-phase high-pressure liquid chromatography (HPLC) device (Waters Alliance HT RP-HPLC) equipped with an X-Terra Shield 5 µm, RP 150 × 10 mm C18 column was used for purifications of oligomers. All purifications were performed with the following gradient of solvent C (0.1 M triethylammonium acetate (TEAA) pH 7) and solvent D (0.1 M TEAA, 80% acetonitrile): 0-1 min 90% C, 1-16 min 90-50% C, 16-17 min 50-0% C, 17-20 min 0% C, 20-21 min 0-90% C and 21-25 min 90% C. The flow was set to 4 mL/min at 30 °C column temperature. Fractions from the HPLC were dried in a SpeedVac vacuum concentrator and the remaining residue re-dissolved in 100 µL MilliQ for subsequent EtOH precipitation.

Liquid chromatography-mass spectrometry (LC-MS) analysis

Oligomers were characterized on a Waters Acquity UPLC H-Class system coupled to an ESI-ToF-MS (Waters Xevo G2XS Qtof). The UPLC instrument was equipped with a Waters Acquity BEH C18 column (2.1 x 50 mm, 130 Å, 1.7 µm). The following gradient of solvent C (5 mM triethylamine and 400 mM hexafluoroisopropanol in MilliQ water) and solvent D (methanol) was applied at 60 °C column temperature and a flow rate of 0.5 mL/min: 0-0.5 min 100-95% C, 0.5-7 min 95-50% C, 7-7.1 min 50-100% C, 7.1-8 min 100% C.
For protein measurements, the same system was used, however, with a Waters Acquity BEH C4 column (2.1 x 50 mm 300 Å, 1.7 µm column) and a different solvent system (0.1% formic acid in MQ and 0.1% formic acid in acetonitrile).

**Agarose gel**

To produce 2% agarose gels, 1 g agarose (#A9539, Sigma-Aldrich) was dissolved in 50 mL TBE buffer (89 mM Tris base, 89 mM boric acid and 0.4 mM EDTA) by heating in the microwave. Subsequent to the addition of 5 µL 10'000x GelRed nucleic acid stain (#41003, Biotium) the gel was cast and solidified in the chamber (passively cool to rt for 30 min). The gel ran in TBE buffer at 110 V and 80 mA for 30-45 min, using GeneRuler ultra low range DNA ladder (#SM1213, Thermo Fisher Scientific) as size reference. DNA could be visualized under UV (Diana III gel imager, Isotopenmessgeräte GmbH).

**Electrophoretic mobility shift assay**

Electrophoretic mobility shift assays were performed based on the general procedure described for agarose gels (see above) with following adjustments: TBE buffer was acidified with glacial acetic acid to pH 6 and the gel ran at 110 V and 100 mA for 30 min. Samples were prepared by mixing 6 µL protein (e.g., AGP in PBS) with 2 µL 1 µM modified oligonucleotides (i.e., on-DNA AGP binder pair assembled in different formats, 200 nM final) and 2 µL 6x loading dye (#R1161, Thermo Fisher Scientific). Samples were incubated for 15 min at room temperature previous to loading.

**Sodium dodecyl sulfate - polyacrylamide gel electrophoresis (SDS-PAGE)**

TruPAGE Precast gels 4-12% (#PCG2007-10EA, Sigma-Aldrich) were used according to the supplier’s procedure with 1 h run time at 180 V and 100 mA. PageRuler Plus Prestained Protein Ladder (Thermo Scientific, #26620) was loaded (4 µL) as size reference and proteins were visualized by Coomassie staining.
Construction of circular, linear and hairpin DELs

Circular and Linear construct:
The constructs were produced in 80 pmol batches with the following composition and procedure:

| Component                                      | Concentration | Volume |
|------------------------------------------------|---------------|--------|
| 5’-modified sublibrary (e.g. Elib2)            | 16 µM         | 5 µL   |
| 3’-modified sublibrary (e.g. S1sh)             | 16 µM         | 5 µL   |
| T4 DNA Ligase Buffer                          | 10x           | 5 µL   |
| Relay primer (near modified library ends)      | 20 µM         | 4 µL   |
| Distal relay primer (circular) or Terminal relay primer (linear) | 20 µM         | 4 µL   |

**Annealing:** 95 °C 2 min, 15 min at rt, 1 h 45 min at 16 °C

| Component           | Concentration | Volume |
|---------------------|---------------|--------|
| dNTPs               | 5 mM          | 10 µL  |
| T4 DNA Polymerase   | 3 U/µL        | 3.3 µL |

**Polymerization:** 30 min at 16 °C

| Component           | Concentration | Volume |
|---------------------|---------------|--------|
| T4 DNA Ligase       | 400 U/µL      | 1 µL   |
| ATP                 | 10 mM         | 3 µL   |

**Ligation:** 2 h at 16 °C

EtOH precipitation with acetic acid buffer, after precipitation washed pellet with 500 µL ice-cold 80% EtOH

The resulting library pellets were dried and re-dissolved in 80 µL MQ to receive a 1 µM solution.
The hairpin construct was generated as following:

| Component                        | Concentration | Volume |
|----------------------------------|---------------|--------|
| 5’-modified sublibrary (e.g. Elib2) | 16 µM         | 5 µL   |
| 3’-modified sublibrary (S4)       | 16 µM         | 5 µL   |
| T4 DNA Ligase Buffer             | 10x           | 5 µL   |
| Adapter                          | 20 µM         | 4 µL   |

**Annealing:** 95 °C 2 min, 15 min at rt, 1 h 45 min at 16 °C

| Component     | Concentration | Volume |
|---------------|---------------|--------|
| T4 DNA Ligase | 400 U/µL      | 1 µL   |
| ATP           | 10 mM         | 3 µL   |

**Ligation:** 2 h at 16 °C

**EtOH precipitation with acetic acid buffer, after precipitation washed pellet with 500 µL ice-cold 80% EtOH, re-dissolved pellet in 80 µL MQ**

**Short heat-cool cycle:** 95 °C 2 min to directly transfer onto ice

A short heat-cooling cycle was applied to revert dimeric structures to hairpin structures (see **Figure S1**).

**Klenow polymerization**

The 3’-modified sublibrary (Elib4), consisting of oligomers with five differently encoded building blocks and a d-spacer region to allow assembly with the 5’-modified sublibrary in ESAC format was premixed with 100-fold input of acetylated DNA. The 5'-modified sublibrary with 553 members (Elib2) was used as reported previously.[2] The sublibraries were mixed in equimolar amounts (200 nM final) together with 10x NEBuffer™ 2 (1x final, #B7002S, New England Biolabs) to adjust the volume with MQ to 124 µL. To allow annealing, the mixture was heated to 90 °C for 2 min to passively cool down to 25 °C for 30 min. Klenow fill-in was initiated by subsequent addition of 5 mM dNTPs (200 µM final) and 1.2 µL Klenow fragment polymerase (#M0210S, New England Biolabs) and proceeded for 1 h at 25 °C. The resulting ESAC library was directly used for affinity selections or could be stored in the fridge for up to 1 month.
Elib4 code addition

The 3’-modified sublibrary strands used for the ESAC format had to be encoded after addition of the building blocks (BB) to the general sequence:

5’-Phos-CATGGGACTCGddddddCAGCACACAGAATTCAGAAGCTCC-(CH₂)₆-NH-BBs-3’

For ligation to the code, 25 µL 4 µM of building block carrying oligomer was mixed with 13 µL 10 µM code (5’-CCTGCATCGAATGGATCCGTGXXXXXXXXXGAGCTGCAGC-3’, X for coding region, 1.3 eq), 6.7 µL 30 µM of a chimeric DNA/RNA adaptor (5’-CGA GTC CCA TGG CGC AGC TGC-3’, RNA nucleotides in bold, 2 eq), 10 µL 10x T4 DNA ligase buffer and 43 µL MQ. The mixture was heated for 2 min at 90 °C, passively cooled to rt before addition of 1 µL T4 DNA ligase. Ligation was left at 16 °C for 5 h before heat inactivation of the enzyme (10 min at 65 °C). Subsequent enzymatic adapter degradation was performed by addition of 11 µL ThermoPol® Reaction Buffer (1x final, # B9004S, New England Biolabs) and 2 µL RNase HII (#M0288S, New England Biolabs) and incubation for 6 h or overnight at 37 °C. The oligomers were cartridge purified (#SK-PCPU-100, Smartpure PCR Kit, Eurogentec) following the manufacturer’s instructions.

S1sh code addition

The 3’-modified oligomer S1sh was mixed with adapter (1.2 eq) and code2 (1.1 eq) to add T4 DNA ligase buffer (1x final) and heat the solution to 65 °C for 5 min. After passively cooling to rt, ligase was added (typically 1 µL for 10 nmol batch) so that the ligation could proceed for 2 h at 20 °C. The enzyme was quenched by heating to 65 °C for 10 min. The extended oligomer was EtOH precipitated and purified via HPLC (see “HPLC purifications of oligomers” but at 60 °C column temperature and with a slow gradient: 0-1 min 95% C, 1-15 min 95-82% C, 15-20 min 82-20% C, 20-22 min 20-0% C, 22-29 min 0% C, 29-29.1 min 0-90% C, 29.1-30 min 90% C).

The sequences were the following:

**S1sh** (X for coding region):

5’-Phos-AAGGAAAGCGTACAATCTCTCXXGCTGGTCTGTCTTTAAGTCTCTGG-(CH₂)₆-NH-BB-3’

**Adapter:** 5’-CTAGCGCTTCTCTGCTGACCTT-3’

**Code2:** 5’-TGACTCCCAATCGATGTCGTCAGXXXXXXGAAGGTCAAGCA-3’.

EcoRI digestion

To 0.1 µg oligomer 10 U of HF EcoRI (#R3101S, New England Biolabs) was added together with 1 µL 20 mg/mL BSA (#B9000S, New England Biolabs), 5 µL 10x CutSmart® buffer (1x
final, #B7204S, New England Biolabs) to adjust with MQ to a final volume of 50 µL. Incubation at 37 °C for 1 h was followed by heat inactivation of the enzyme (65 °C for 10 min).

**Affinity selections of DELs and sequencing**

Automated affinity selections were performed with Dynabeads™ MyOne™ Streptavidin C1 (Thermo Fisher Scientific, # 65001), following the established protocol. All selections were performed at an input of 10⁷ copies per library member. For sequencing, selection eluates were amplified via two sequential PCR steps, following the general procedure with following adjustments for PCR1:

| PCR1 mixture composition (per reaction) | Volume |
|----------------------------------------|--------|
| Selection eluate                       | 30 µL  |
| Phusion HF buffer (5x)                  | 10 µL  |
| Phusion MgCl₂ (50 mM)                   | 2 µL   |
| dNTPs (5 mM)                           | 2.5 µL |
| PCR1 primer (10 µM)                    | 2 µL each (reverse and forward) |
| Phusion enzyme (2 U/µL)                 | 0.5 µL |

Furthermore, the annealing temperature was adjusted to 70 °C for the linear and circular constructs and to 55 °C for the hairpin construct.

**Protein biotinylation**

Alpha-glycoprotein 1 (AGP) and carbonic anhydrase IX (CAIX) were kindly provided by Florent Samain and Jacopo Millul (Philochem AG). CAIX (21 µM in PBS pH 7.4) was chemically biotinylated by addition of 20 equivalents NHS-LC-Biotin (10 mM in DMSO, #21336, ThermoFisher Scientific) and incubation at rt for 2 h. Biotinylation was monitored by LC-MS analysis (see Figure S4 A→B, +339) and upon completion the mixture was directly loaded onto a PD-10 column (#17085101, Cytiva) equilibrated in PBS to collect protein-containing fractions. AGP was dissolved in 1 mL PBS (0.56 mg/mL) to add 5 µL 22 mM NHS-LC-Biotin (8.9 equivalents). After 45 min of gentle mixing at rt, the solution was loaded onto a pre-equilibrated (PBS pH 7.4) PD-10 column to combine the protein-containing fractions. Due to heavy glycosylation, AGP biotinylation was checked by SDS-PAGE, loading 3 µg biotinylated
AGP, 3.5 µg avidin and a mixture of b-AGP (3 µg) and avidin (3.5 µg) in non-reducing conditions (see Figure S43 C).

**Figure S43.** Quality control of chemical protein biotinylations. MS spectra of carbonic anhydrase IX before (A) and after biotinylation (B) revealed the expected mass shift of +339 Da. Alpha-glycoprotein 1 (AGP) biotinylation was visualized by SDS-PAGE (complex formation with avidin, C).

### 3. References

[1] M. Wichert, N. Krall, W. Decurtins, et al., *Nat. Chem.* 2015, 7, 241-249.
[2] C. E. Dumelin, J. Scheuermann, S. Melkko, et al., *Bioconjugate Chem.* 2006, 17, 366-370.
[3] G. Bassi, N. Favalli, M. Vuk, et al., *Adv. Sci. (Weinheim, Ger.)* 2020, 7, 2001970.
[4] M. Catalano, M. Moroglu, P. Balbi, et al., *ChemMedChem* 2020, 15, 1752-1756.
[5] W. Decurtins, M. Wichert, R. M. Franzini, et al., *Nat. Protoc.* 2016, 11, 764-780.