Electronic Supplementary Material (ESI) for ChemComm.

Supporting information

Second Generation DNA-Encoded Dynamic Combinatorial Chemical Libraries

Francesco V. Reddavide[1,2#]*, Meiying Cui[1#], Weilin Lin[1], Naiqiang Fu[1], Stephan Heiden[2], Helena Andrade[1], Michael Thompson[2], Yixin Zhang[1]*

1. B CUBE Center for Molecular Bioengineering, Technische Universität Dresden, Germany.
2. DyNAbind GmbH, Dresden, Germany.

[#] These authors contributed equally to this work.

Table of Contents

1. List of abbreviations

AcOH: acetic acid
ACN: acetonitrile
AMA: ammonium hydroxide and aqueous methyl amine
ATP: adenosine triphosphate
CPG: controlled pore glass
CBS: 4-Carboxybenzenesulfonamide
DCM: Dichloromethane
DEAE: Diethylaminoethyl
DIPEA: *N,N*-diisopropylethylamine
DMF: N, N Dimethylformamide
DMSO: dimethylsulfoxid
EDC: 1-Ethyl-3-[3-dimethylaminopropyl]carbodiimide hydrochloride
HATU: 1-[Bis(dimethylamino)methylene]-1H-1,2,3-triazolo[4,5-b]pyridinium 3-oxid hexafluorophosphate
HBTU: 2-(1H-benzotriazol-1-yl)-1,1,3,3-tetramethyluronoium hexafluorophosphate
HEPES: 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HOAt: 1-hydroxy-7-azabenzotriazole
HPLC: High Performance Liquid Chromatography
Im: Iminobiotin
MMT: Methoxytrityl [(4-methoxyphenyl)diphenylmethyl]
NGS: Next Generation Sequencing
NHS: N-hydroxysuccinimide
PBS : phosphate buffer saline
qPCR: quantitative Polymerase Chain Reaction
Sulfo-NHS: N-hydroxysulfosuccinimide sodium salt
TCEP: tris(2-carboxyethyl) phosphine
TEA: triethylamine
TEAA: triethylammonium acetate
TEAB: triethylammonium bicarbonate
TFA: trifluoroacetic acid
THF: tetrahydrofuran
TIS: triisopropylsilane
TRIS: tris(hydroxymethyl)aminomethane
UPLC-ESI-MS: Ultra Performance Liquid Chromatography-Electrospray Ionization-Mass Spectrometry
Van: vancomycin

2. Materials

All chemicals were, unless otherwise noted, of reagent grade and purchased from Sigma-Aldrich (St. Louis, MO, USA), Merck Millipore (Darmstadt, Germany), Iris Biotec (Marktredwitz, Germany) or Thermo Fisher Scientific (Waltham, MA, USA). All oligonucleotides were purchased from IBA life sciences (Göttingen, Germany) and Metabion (Steinkirchen, Germany) in HPLC-purified grade, molecular biology grade, NGS grade or on the CPG solid support according to applications; other chemicals for oligonucleotide synthesis were obtained from Link Technologies (Bellshill, Scotland, UK). Streptavidin Sepharose High Performance and NHS-activated Sepharose were purchased from GE Healthcare (Little Chalfont, England, UK).

3. Coupling of oligonucleotides

3.1 Conjugation on the 5´ oligonucleotides

Unless otherwise noted, all the 5' modification of the oligonucleotides were based on amide bond formation between the amino group on the oligonucleotide and the carboxylic group of the organic molecule with the following protocol: 1 eq (0.01 mmol, 50 mM) carboxylic acid, 1 eq (0.01 mmol, 50 mM) HATU and 1 eq (0.1mmol, 50 mM) HOAt were dissolved in 200 µL DMSO and incubated 30 min on orbital shaker at 300 rpm at room temperature. The reaction mixture was added on top of the CPG beads and 3 eq. (150 mM) of DIPEA was added and incubated overnight at room temperature, then washed with DMSO followed by DCM. The DNA was cleaved from the CPG beads and deprotected with 1:1 mixture of AMA, while incubating on an orbital shaker at 420 rpm for 3 h at room temperature. The AMA solution of the crude conjugates was seperated from the CPG beads, shock frozen and vacuum concentrated at 50 °C. The crudes were re-suspended with 1 mL MilliQ water and purified via reverse-phase HPLC (Waters, USA) on a Clarity 3u Oligo-RT C18 reverse−phase HPLC column (Phenomenex, Torrance, CA, USA), applying a gradient from 5% ACN to 35% ACN over 30 min. The correct fraction was confirmed via UPLC-ESI-MS, Waters, USA) equipped with an analytical ACQUITY UPLC OST C18 column (Waters, USA). The fractions were dried using a speedvac (Christ RVC 2-25 CD plus/CT 02-50 SR, Osterode am Harz, Germany) equipped with a PC 3000 series vacuum controller (Vacuumbrand, Wertheim, Germany). According to the absorbance at 260 nm, the conjugates were quantified and equilibrated.

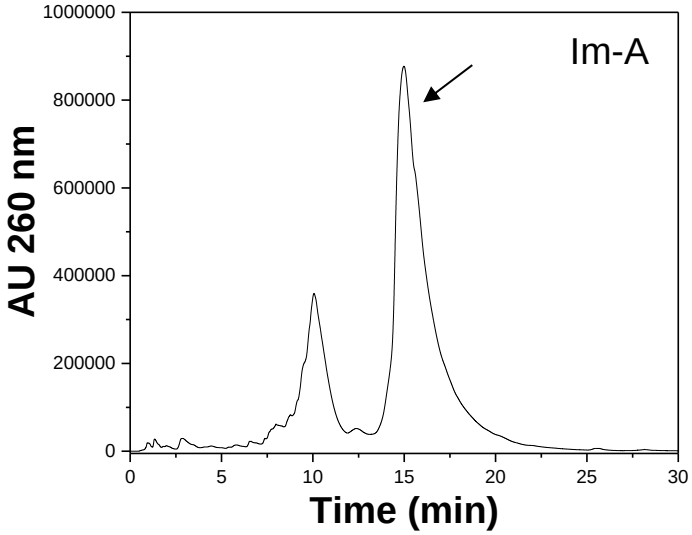


Figure S1. HPLC chromatogram of the 5' Iminobiotin modified oligonucleotide (Im-A). The arrow indicates the collected HPLC fractions.

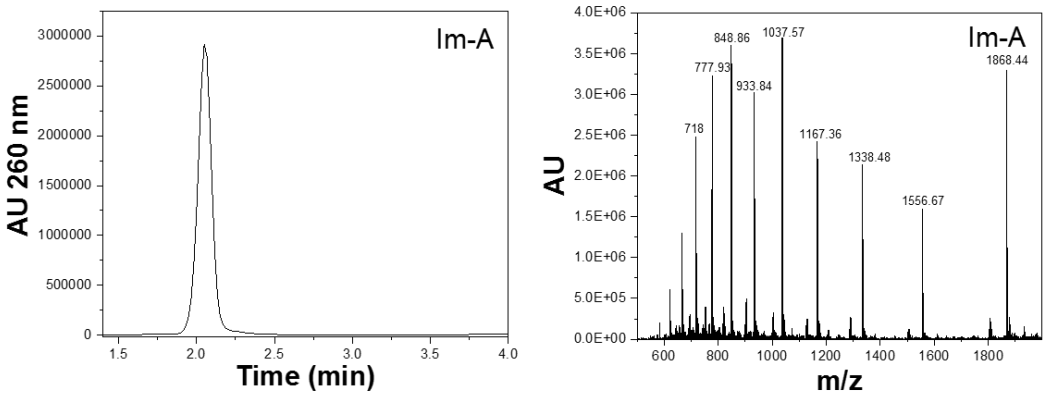


Figure S2. UPLC chromatogram and ESI-MS spectrum of the HPLC purified 5' Iminobiotin modified oligonucleotide (Im-A).

3.2 3'-Iminobiotin modified oligonucleotide

The 3' iminobiotin oligonucleotide was synthesized according to the adapted DEAE method described by Halpin and Halbury.[1]

The 3'-C6-amino-modified oligonucleotide (15 nmol) were immobilized on DEAE sepharose (0.1 mL of slurry). The resin was washed with 10 mM aq. AcOH / 0.005 % TritonX 100 (2 × 0.5 mL), MilliQ water supplemented with 0.005 % TritonX 100 (2 × 0.5 mL) and DMSO (2 × 0.5mL). 2-iminobiotin carboxylic acid was activated with EDC (HCl salt, 50 mM) and HOAt (5 mM) in DMSO and then added to oligonucleotides immobilized on the resin. The slurry was agitated for 3 h at room temperature. The solution was removed and the resin washed with DMSO (1 × 0.5 mL) and re-treated with freshly activated 2-iminobiotin carboxylic acid solution. These steps were repeated to reach three coupling steps of 1 h each. The reaction solution was removed and the resin was washed with DMSO (2 × 0.5 mL) and AcOH / 0.005 % TritonX 100 (2 × 0.5

mL). The DNA was eluted from the resin by incubating with 50 mM Tris-HCl, 1.5 M NaCl, 0.005 % TritonX 100, pH 8.0 for 10 min.
The eluted oligonucleotide was purified and characterized as described in section 3.1.

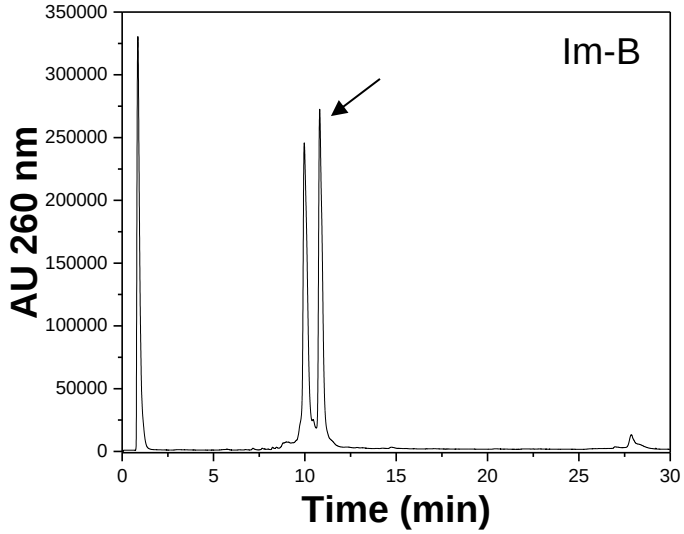


Figure S3. HPLC chromatogram of the 3' Iminobiotin modified oligonucleotide (Im-B). The arrow indicates the collected HPLC fractions.

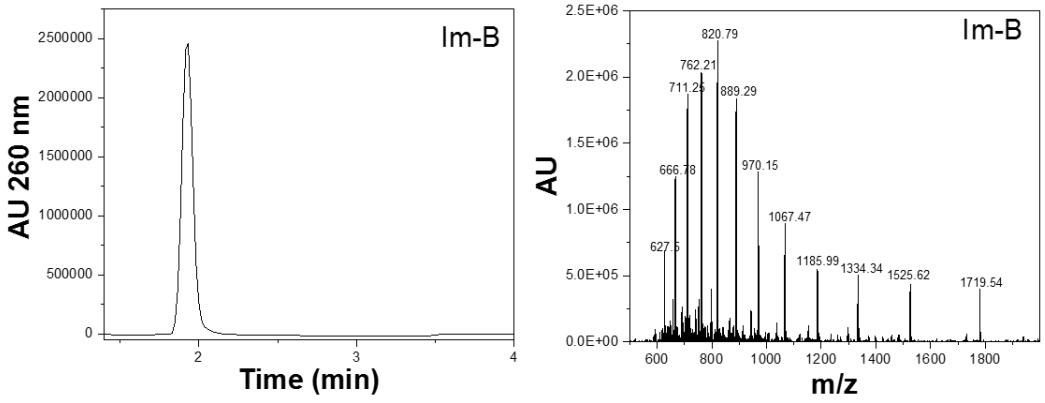


Figure S4. UPLC chromatogram and ESI-MS spectrum of the HPLC purified 3' Iminobiotin modified oligonucleotide (Im-B).

3.3 5'- and 3'-Vancomycin modified oligonucleotides

Coupling of vancomycin was accomplished through maleimide reaction chemistry.
The same protocol was used both for 5' and 3' modified oligonucleotides. Amino modified oligonucleotides (20 nmol, in 100 mM KH_2PO_4, pH 7.2) were mixed with 10 mM Sulfo-SMCC (in DMSO, 20 % of total volume). The reaction mixture was incubated at room temperature for 2 h. Subsequently the oligonucleotides were precipitated via ethanol precipitation.
The ethanol precipitation protocol used was the following: firstly, 1/10 volume of sodium acetate (3 M, pH 5.2) and 2.5 volume of absolute EtOH were used to precipitate DNA.

Then the solution was incubated at -20 ºC for 2h. The precipitated DNA was centrifuged at 13,000 rpm for 30 min at 4 ºC. The DNA pellet was then washed with 70% EtOH and air dried.

The sulfhydryl modifying vancomycin (Van-SH) was obtained from the reduction of 10 mM dimeric vancomycin (Van-S-S-Van) using 50 mM TCEP, pH 7 in water for 30 min. The maleimide terminal oligonucleotide pellets were mixed with freshly reduced vancomycin in 100 mM KH_2PO_4, pH 7.2 for 2 h.

The synthesis of bis(van) was performed according to the following protocol adapted from Ren et al., 2014.[2] Cystamine dihydrochloride (0.03 mmol) was dissolved in 1 mL dry DMF firstly, then the resulted solution was added to the solution of vancomycin–HCl (0.07 mmol) in 1 mL DMSO. The mixture was cooled to 0 °C, HBTU (0.09 mmol) and DIPEA (0.36 mmol) were added following. The solution was allowed to warm to room temperature and stirred overnight. Reverse-phase preparative HPLC was used to purify the product and the product was lyophilized. The product was confirmed with ESI-MS (MW 3015).

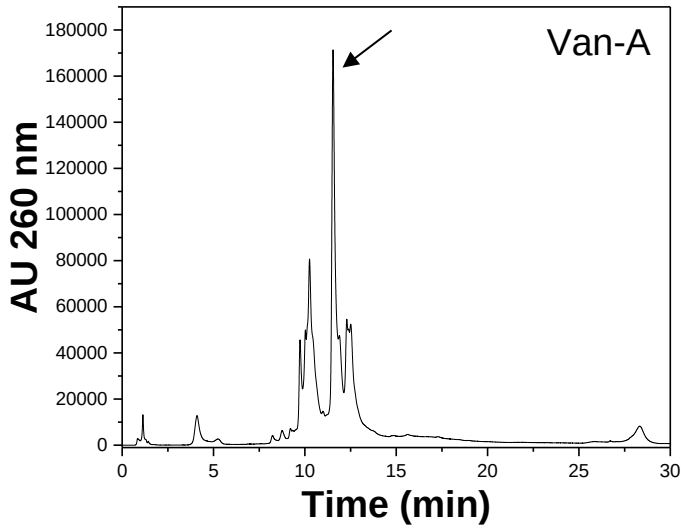


Figure S5. HPLC chromatogram of the 5' vancomycin modified oligonucleotide (Van-A). The arrow indicates the collected HPLC fractions.

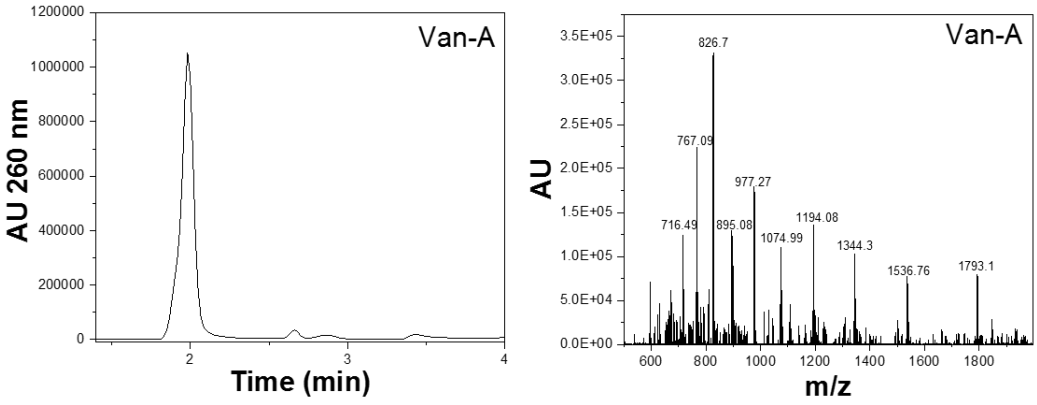


Figure S6. UPLC chromatogram and ESI-MS spectrum of the HPLC purified 5' vancomycin modified oligonucleotide (Van-A).

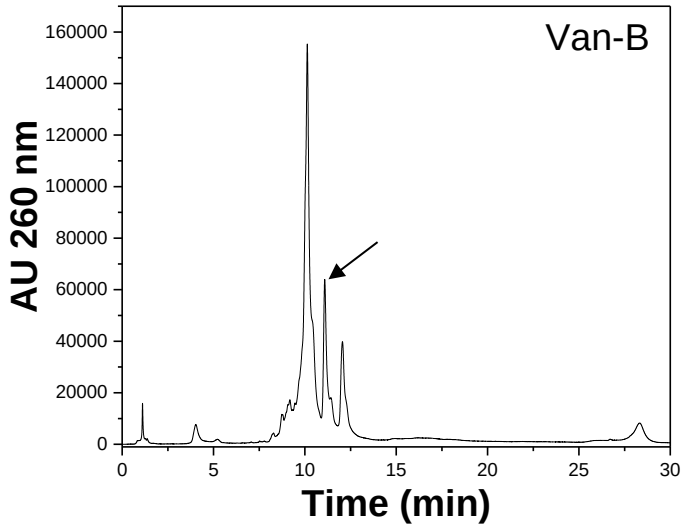


Figure S7. HPLC chromatogram of the 3' vancomycin modified oligonucleotide (Van-B). The arrow indicates the collected HPLC fractions.

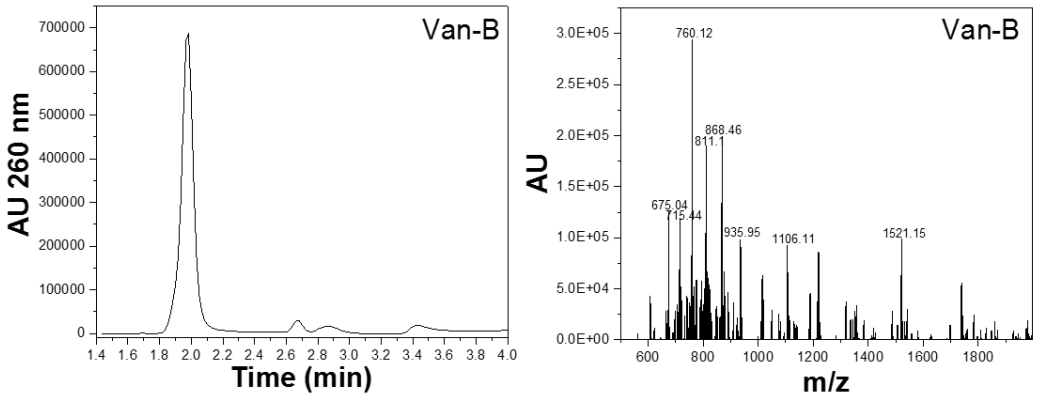


Figure S8. UPLC chromatogram and ESI-MS spectrum of the HPLC purified 3' vancomycin modified oligonucleotide (Van-B).

4. DNA-encoded Dynamic library selection experiments

The resin with the immobilized protein was added to a 1.5 mL test tube (DNA LoBind Eppendorf and tips were used for all the experiments). The resin was washed three times using 1 mL binding buffer with gentle shaking and the supernatant was discarded after centrifuging the sepharose resin at 2,000 rpm for 5 min. After removal of supernatant the resin was ready for the binding assay.

4.1 Library preparation for streptavidin and bacterial cell wall model

The different model libraries were prepared mixing together the two sublibraries (Y-5 and Y-3) composing the Y-shaped construct (Fig. S9).
The sublibrary Y-5 was assembled by mixing 1 eq. of the Y-5-I construct (1 nM or 0.1 nM) carrying the positive control for the binding fragment and the excess of the non-binding competitor Y-5-N at different molar ratios. In the same way, the sublibrary Y-3

was generated using the construct Y-3-I as carrier of the positive binding fragment and Y-3-N as the non-binding competitor.

The sublibrary constructs were assembled via an annealing protocol in a PCR machine where the samples were heated up at 85 ˚C and slowly cooled down to 20 ˚C over 90 min. Each construct, Y-5-I, Y-5-N, Y-3-I, Y-3-N, was annealed separately at a concentration of 5 µM in annealing buffer (1mM $MgCl_2$).

The library size of around one hundred thousand was simulated by placing the different sublibrary constructs in the following ratio: 1 eq. (1 nM, 200 fmol) of Y-5-I and Y-3-I, and 320 eq (320 nM, 64 pmol) of Y-5-N and Y-3-N. The library sizes of 1 million and 100 million members were simulated by placing the different sublibrary constructs in the following ratio: 1 eq (0.1 nM, 20 fmol) of Y-5-I and Y-3-I and 1,000- or 10,000- eq (0.1 µM, 20 pmol or 1 µM, 200 pmol) of Y-5-N and Y-3-N.

To avoid a preferential assemble of the positive binding fragments the library was prepared via adding the constructs in the following order: Y-5-N, Y-3-I, Y-3-N, Y-5-I.

The 100,000 static library was prepared with a slightly different protocol. Four annealing tubes were prepared. Each tube contained the annealed construct of each pair of fragments on a full length template. The occurrence of each combination of high- low- or non- affinity binding member of the library was calculated and the different pairs added accordingly before being added to the resin. In particular, the concentration of each pair was the following: Y-5-I/Y-3-I (64 pmol ÷ 102400 = 0.625 fmol or also 200 fmol ÷ 320 = 0.625 fmol), Y-5-I/Y-3-N and Y-5-N/Y-3-I (200 fmol – 0.625 = 199.38 fmol), Y-5-N/Y-3-N (64 pmol – 0.625 fmol – 199.38*2 fmol = 63.6 pmol). The choice to not to generate a self-assembly static library was done to exclude any effect of the ligation process on the enrichment. The use of the Y-shaped construct for static library has no advantages over the already known linear contracts where the enzymatic reactions work quantitatively. Therefore, it was more of interest to compare the dynamic Y-shaped construct performance with the static library maximizing the possible enrichment of the static library avoiding any final ligation step.

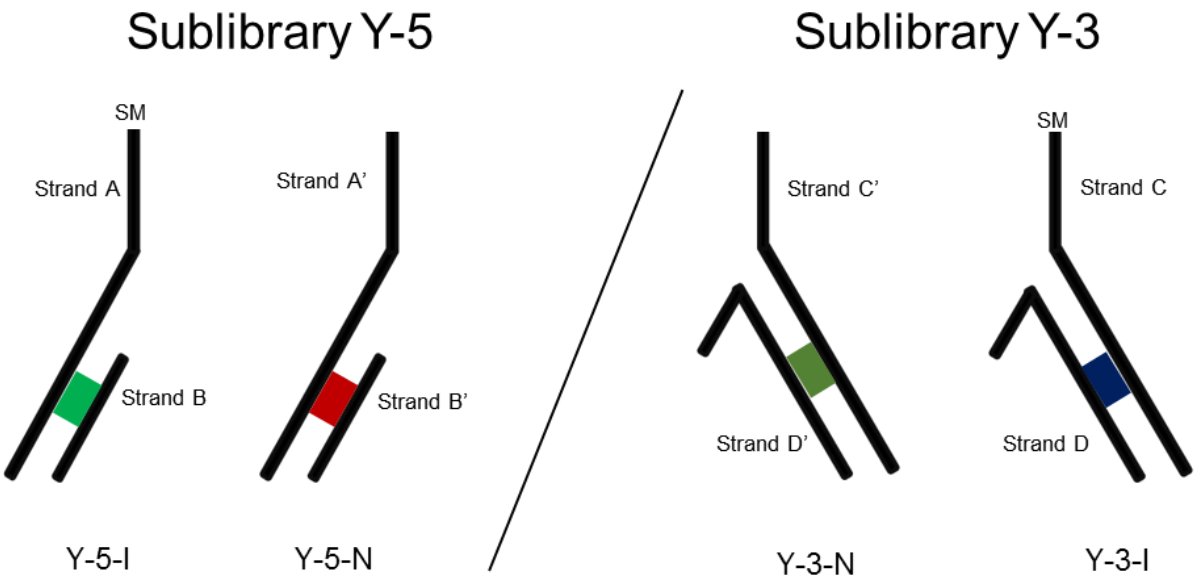


Figure S9. Constructs of the y-shape DNA codes. The oligonucleotide strands A, B and A', B' were annealed forming the partial double stranded constructs 5-Y-I and 5-Y-N, respectively. They were used to generate the sublibrary Y-5. The oligonucleotide strands C, D and C', D' were annealed forming the partial double stranded constructs 3-Y-I and 3-Y-N, respectively.
In the different model selections, the strand A carried the positive binding fragment (SM, small molecule) for the Y-5 sublibrary (iminobiotin, vancomycin), while the strand B carried a unique primer region for the positive binding control quantification. The strand C carried the positive binding fragment for the Y-3 sublibrary (iminobiotin, vancomycin), while the strand D carried a unique primer region for the positive binding control quantification. The excess of non-binding elements simulating the library is represented by the partial double stranded constructs Y-5-N and Y-3-N formed via the separate annealing of the strand A' and B' and C' and D', respectively. The strand B' and D' carried two unique primer regions used to quantify the occurrence of the non-binding member of the library.

4.2 Library preparation for Carbonic Anhydrase II affinity maturation

The dynamic behavior of 6+5 and 6+7 formats were explored. Selection against streptavidin was performed in the same way as in section 4.1, using 6+5 format and 6+7 format libraries separately. The enrichment of dual iminobiotin from two different format of libraries were measured by QPCR and were calculated by setting dual blank as reference. Fold of enrichment=2^(-ΔCt)* library size.

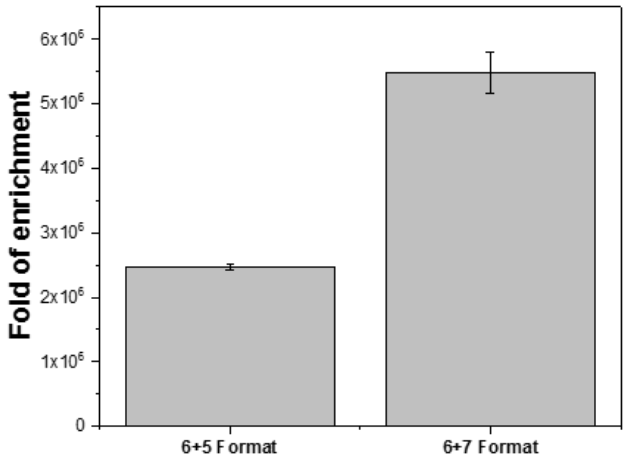


Figure S10. Enrichment of dual iminobiotin from selections against streptavidin using 6+5 and 6+7 format libraries.

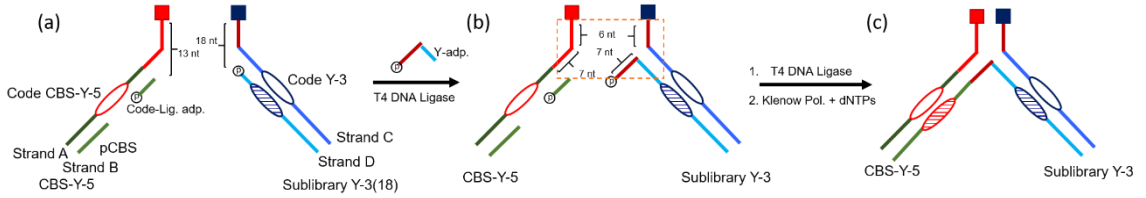


Figure S11. 285-member library for CA II affinity maturation. (a) The Y-EDCCL for the CA II affinity maturation was prepared using a CBS-Y-5 construct displaying a CBS chemical moiety on the 5' end of strand A. This construct has a 13 ssDNA overhang on the 5' and a phosphorylated adapter used to attach the code for the CBS moiety. The 3' end of strand A is complementary with the primer pCBS used later as a PCR primer. (b) The 285-member sublibrary Y-3 was prepared ligating (via T4 DNA ligase) the Y-adapter (Y-adp.) to the sublibrary Y-3(18) having 18 nt overhang on the 3' end. The Y-EDCCL is then generated by mixing the CBS-Y-5 construct with the 285-member sublibrary Y-3. (c) After the library selection on the solid support, the ligation mixture is added (1) allowing the enriched sublibrary Y-3 members to be ligated with the CBS-Y-5 construct while still binding on the target protein. Afterwards, the addition of klenow polymerase and dNTPs (2), with the preexisting T4 DNA ligase, allows the formation of the joined code pair template. The construct is finally eluted from the solid support for further analysis. The template is represented by the joined strand B and D and the code flanking regions are used as primers for PCR and sequencing. The phosphorylated position are indicate by a (p) at the 5 end of the oligonucleotides.

The library was prepared by coupling each building block (as a carboxylic acid) to a C6 amino-3'-modified DNA (AtdBio, Southampton, UK) on a CPG solid support. The DNA was reverse-synthesized on the solid support, therefore, it would expose the 3' amino-modifier as last monomer.

The coupling protocol and the sample characterization was the same as in section 3.1. Successfully coupled DNA-building blocks were then phosphorylated using T4 PNK (10 U, New England Biolabs, USA) in 1X NEB2 buffer supplemented with 1 mM ATP, in a total volume of 50 µL and then encoded with respective code DNA sequences using T4 DNA ligase (4U, New England Biolabs, USA) in 1X NEB2 buffer supplemented with 1 mM ATP in 100 µL volume. The oligonucleotides were added in the following concentrations: 5 µM building block oligo-conjugate, 1.25 eq. of the respective code and ligation adapter. First, the oligonucleotides were heated at 60 ºC for 5 minutes and slowly cooled to 25 ºC over a time of 2 h to form dsDNA. Next, 4 U of T4 ligase and 1 µL of 100 mM ATP was added and incubated O/N at 18 ºC. The T4 DNA ligase was inactivated incubating at 65 ºC for 20 min and then slowly cooling to room temperature. Each ligation was tested via qPCR (Thermo Fisher Scientific) using the code flanking primers.

For 3' library, 2 µL of 2mM dNTP and 2 µL, 10U of klenow exo (NEB, USA) were added to extend from 3' of Adapter and form double strand at code region. Finally, the ligase and klenow was inactivated for 20 min at 75 °C in a water bath and the temperature decreased slowly till room temperature to allow re-annealing of the construct.

Coupling and encoding process of Y-5-CBS were essentially the same as 3' library, except that extension with klenow was done after selection and ligation of two sublibraries. Schematics of the sublibraries can be seen in Fig. S11.

The 285-member- 3L library was prepared by mixing equimolar each member. The Y-3 library exposes 13 bases complementary to Y-5-CBS.

This Y-3 library was then further diluted in 1X PBST and mixed with Y-5-CBS. Selection was performed with 1 nM (100 fmol) Y-5-CBS and 1 nM (100 fmol) each member of Y-3 library in 100 µL volume.

4.3 General binding assay

In general, the library was exposed to protein immobilized on resin (Fig. S12) and incubated for 2 h on a rotating mixer. Then the beads were washed 5 times with 1 mL of binding buffer before proceeding with the ligation and the final elution.

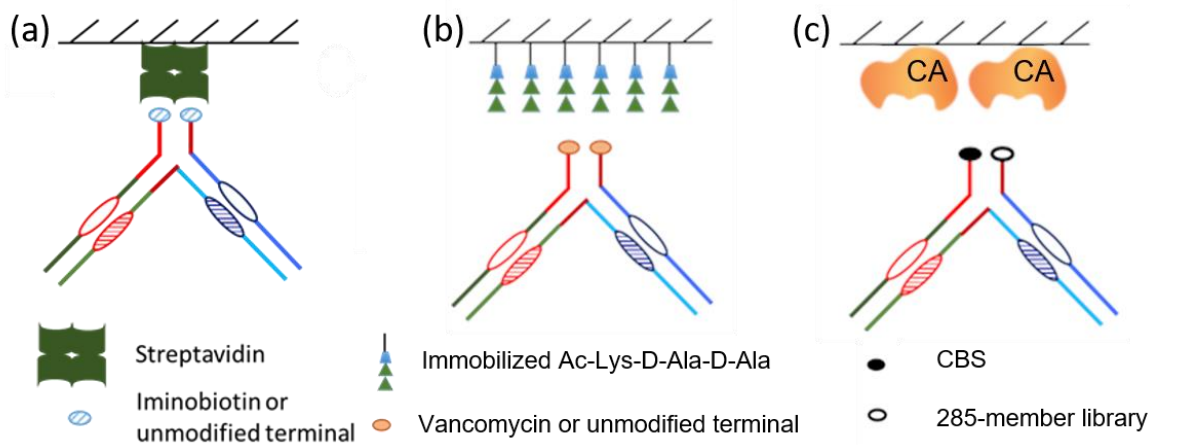


Figure S12. Y-EDCCL selection experiments. (a) Streptavidin model system. The chelate effect of the bidentate iminobiotin Y-shaped construct against streptavidin was used as positive control for binding. A model library was prepared by adding excess blank DNA constructs. (b) Bacterial cell wall model system, with monomeric vancomycin coupled to each side of the Y-shaped construct (Y-5 and Y-3). Assembly of the dimeric vancomycin construct on DNA and binding on the D-Ala-D-Ala peptide modified solid support. (c) Carbonic anhydrase II Y-EDCCL affinity maturation using CBS as known binder (Y-5) and a library of 285 fragments (Y-3).

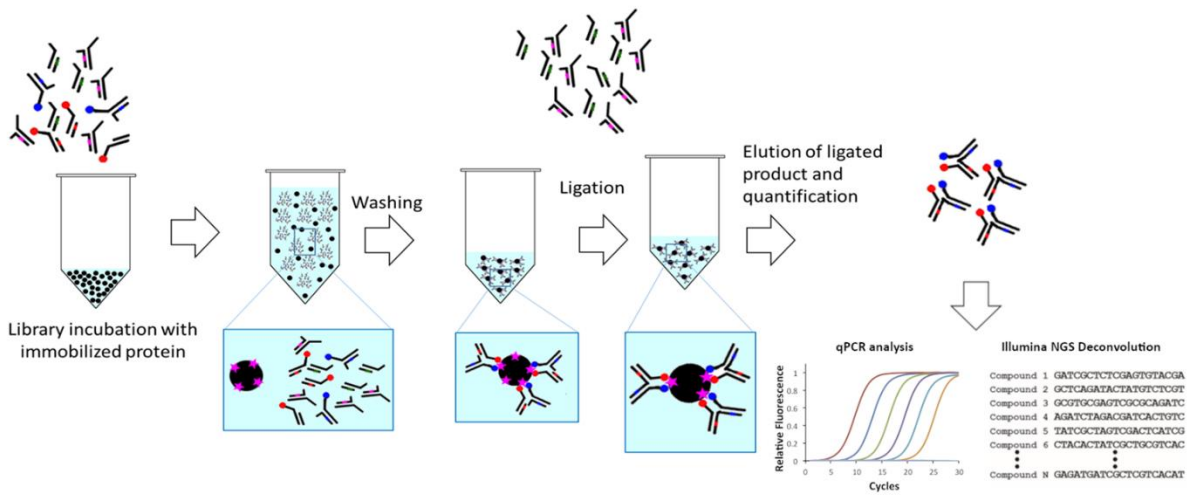


Figure S13. Y-EDCCL experimental workflow. Moving left to right in the diagram, the library is incubated with the solid support-immobilized protein. The unbound part is removed and the solid support is washed several times. The bound part is ligated using DNA Ligase, after the bound part is eluted via protein denaturation and the different binding population analyzed via qPCR or Illumina New Generation Sequencing.

4.4 Streptavidin model binding assay

The library was added to 10 µL slurry of High Performance Streptavidin Sepharose resin. The binding buffer used had the following composition: 150 mM NaCl, 25 mM $NaHCO_3$, pH 9.2, and 0.05 % Tween20.

4.5 Bacterial cell-wall model binding assay

The peptidoglycan peptide of gram positive bacteria (Ac-Lys-D-Ala-D-Ala) was synthesized by standard Fmoc chemistry on a solid-phase (d-ALA Wang Resin) with HBTU activation on an automated solid-phase peptide synthesizer (ResPep SL, Intavis, Cologne, Germany). Each amino acid was coupled twice with each 5 times excess and all non-reacted amino groups were capped with acetic anhydride (5% in DMF). Then the peptide was cleaved form the resin with TFA/TIS/water (92.5 (v/v):5 (v/v):2.5 (v/v)) for 1.5 h. The product was precipitated and washed with ice-cold diethyl ether. The peptide was purified by HPLC.

The peptide was first coupled on NHS-activated Sepharose 4 Fast Flow (GE Healthcare). The resin (500 µL) was transferred into a 1.5 mL LoBind tube and washed 5 times with 1 mL of cold 1 mM HCl immediate. Then 3.2 mg peptide dissolved in 250 µL binding buffer (0.2 M $NaHCO_3$, 0.5 M NaCl, pH 8.3) was mixed with the washed resin. After 2 h incubation on a rotator at RT the resin was centrifuged at 2,000 rpm and the supernatant was discarded, and then the resin was kept in 0.5 M ethanolamine, 0.5 M NaCl, pH 8.3 or 0.1 M Tris-HCl, pH 8.5 for 1 h to block any non-reacted groups. Finally, the resin was washed with 0.1 M Tris-HCl buffer pH 8 followed by a washing with 0.1 M acetate buffer, 0.5 M NaCl, pH 4, this step was repeated for 3 times. The peptide coupled resin was ready for use and stored at 4 ˚C.

For each binding experiment, 20 µL slurry were used and PBS supplemented with 0.05 % Tween20 was used as binding buffer.

4.6 Carbonic Anhydrase II binding assay

2 X 350 µL NHS- activated Sepharose 4 Fast Flow (GE Healthcare, UK) were prepared in SpinX columns (Corning Inc.) The beads were washed 15 times with fresh, cold 1mM HCl and three times with cold 1X PBS with 0.05% Tween 20, spinning 1min at 2000 rpm. Then, 1.25 mg of CA II was dissolved in 500 µL of coupling buffer (0.1 M $NaHCO_3$, 0.5 M NaCl at pH 8.3). Meanwhile, Tris beads were prepared as a negative control by adding 500 µl of 0.1 M Tris-HCl, pH 8.0 instead of protein solution. The beads were incubated with either CA II or Tris buffer for 5 h at 4 °C. After incubation, the supernatants were discarded and 500 µL of Tris buffer was added on top to quench and left on ice for 30 min. Then, the beads were washed by repeating three times the following cycle: three times 400 µL 0.1 M Tris-HCl at pH 8, and three times 400 µL 0.1

M sodium acetate, 0.5 M NaCl at pH 4. Finally, a three times wash with 400 µL 1X PBST was done. The beads were resuspended in 400 µL1X PBST and stored at 4 °C.

For the binding assay, 20 µL of Tris beads and CA II coated beads were washed and incubated with 10 µg/ml herring sperm DNA and the library supplemented in 100 µL 1X PBST.

4.7 Code-joining via T4 DNA ligation

The washed resin was mixed with 1 µL T4 DNA ligase, 2 µL NEB2 buffer, 2 µL of ATP (10 mM) and 15 µL MilliQ water. The mixture was gently vortexed and incubated O/N at 16 ˚C.

For CA II binding assay, 1 µL (2 U) klenow and 2 µL dNTP (2 mM) were added into ligation mixture and stirred on an orbital shaker for 3 h, to finally form the complete Y-shape template joining two codes.

4.8 Elution of the bound components

The on-resin ligated bound components were eluted for the following qPCR analysis. For streptavidin coated beads, 1 µL of biotin in DMSO (0.2 mg/µL) was added to the O/N incubated ligation mixture and incubated for 15 min on a rotating mixer. Then the mixture was heated to 95 ˚C for 5 min. Finally, the resin was centrifuged and supernatant was collected.

For the bacterial cell wall model and the CA II coupled on the NHS-activated Sepharose, the elution were carried in 100 µL and 50 µL elution buffer (10 mM Tris, 0.1% Tween 20, pH 8.0), respectively. Then the mixture was heated to 95 ˚C for 5 min. Finally, the resin was centrifuged and supernatant was collected.

To prove the formation of Y-shape DNA construct in the selection output of CA II, the eluted sample was amplified and subjected to Sanger sequencing before submission to Illumina high-throughput sequencing.

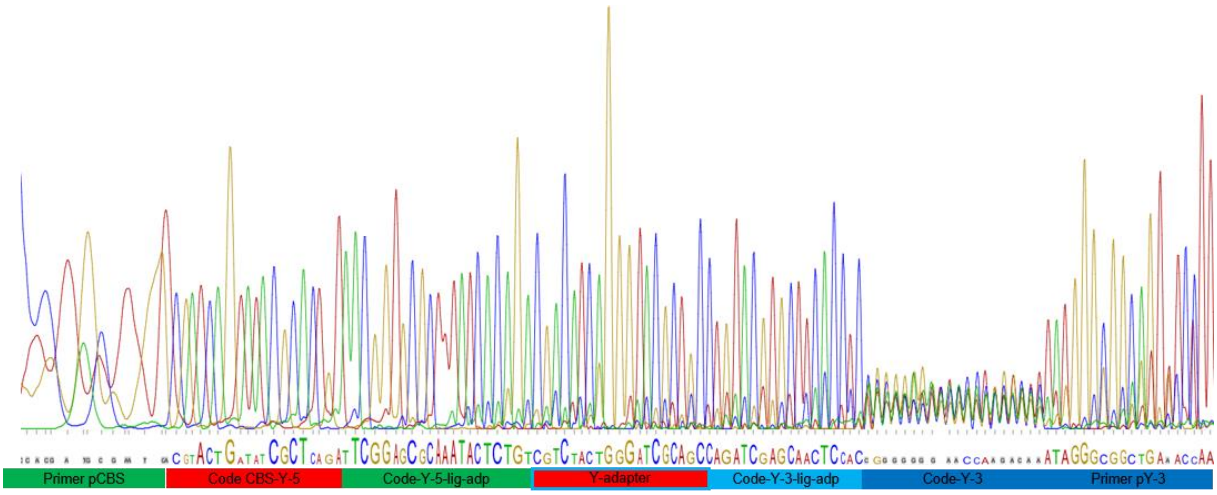


Figure S14. The Sanger sequencing chromatogram of eluted Y-shape construct from selection against CA II. The sequenced DNA contained all the parts of sequence used for construction of Y-shape library. The signal of pCBS was ambiguous, because the region was close to the sequencing primer. Other parts showed clear sharp peaks, except for code-Y-3 part (20 nt) due to high sequence diversity of the library.

4.9 qPCR quantification and data analysis

The enrichment of each combination eluted after the ligation of the y-shaped construct for DNA-encoded dynamic combinatorial chemical library was analyzed via qPCR. Four orthogonal sets of primers permit to distinguish the high affinity binding from the low affinity and non-binding compounds by qPCR amplification (Fig. S15).

The qPCR reaction mixture (10 µL) contained: Maxima Sybr Green qPCR 2X Master Mix, primers (500 nM), and 1 µL of elution product. The samples were loaded in a 96 well plate for Piko Real-Time PCR system with the following amplification method: 10 min at 95 °C, then 35 cycles of: 15 s at 95 °C, 30 s at 60 °C, and 30 s at 72 °C (Sybr Green signal collected at this step). The inability to perform qPCR amplification with the wrong template showed that the primers are orthogonal between each other.

To quantify the enrichment of each binding combination, a standard curve for the different templates were performed.

As templates were used oligonucleotides having the same sequences as the ligated ones (Y-5-I/Y-3-I, Y-5-I/Y-3-N, Y-5-N/Y-3-I, Y-5-N/Y-3-N).

The four plots and equations are shown in Fig. S16. A 10-fold dilution series was used (from 1 nM to 0.1 pM), 1 µL was used as template and the qPCR was performed in triplicate.

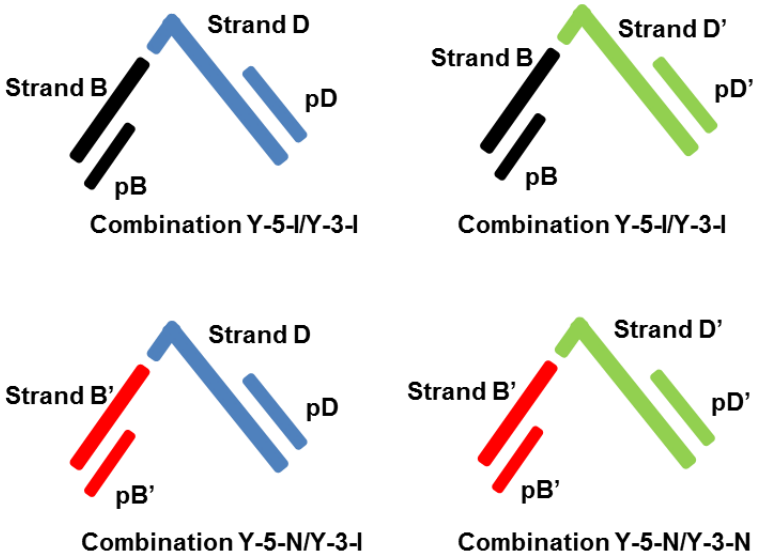


Figure S15. The four sets of primers and templates. Four orthogonal sets of primers were used to amplify the four different combinations. The combination Y-5-I/Y-3-I represents the high affinity binders and was specifically amplified by the primer set pB and pD. The combinations Y-5-I/Y-3-N and Y-5-N/Y-3-I represent the low affinity binders and were amplified using the primer sets pB - pD' and pB' - pD, respectively. The combination Y-5-N/Y-3-N represents the non-binding members of the library and was amplified using the primer set pB' and pD'.

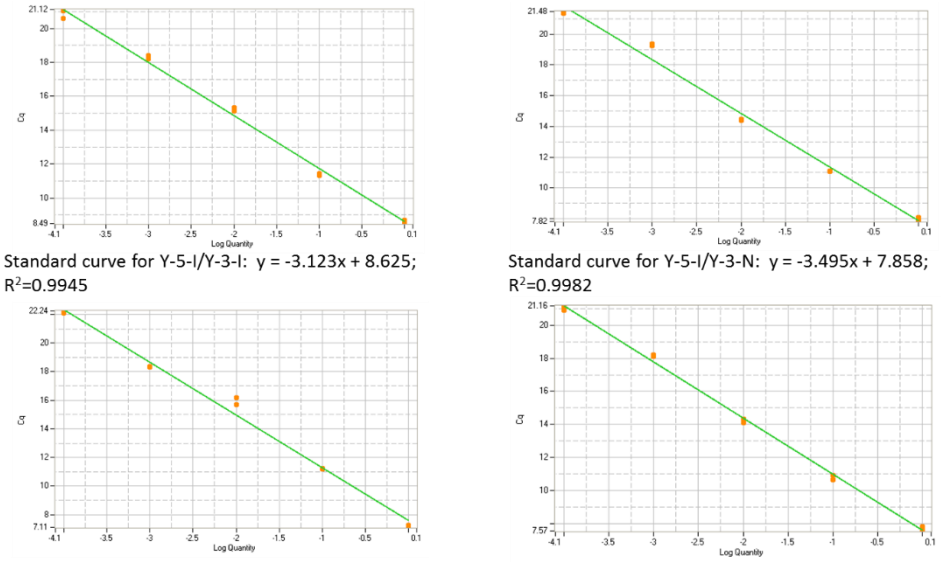


Figure S16. Standard curves for the four different template used. A 10 fold dilution series was prepared (from 1 nM to 0.1 pM) and 1 µL was used as template.

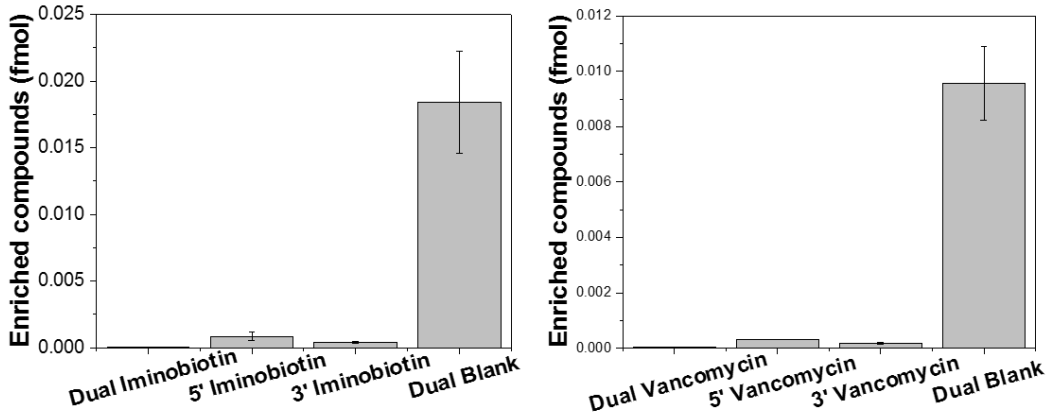


Figure S17. Control on the solid support for the Model selections with streptavidin and bacterial cell wall model. Binding profiles of the four possible combinations from the 10^5-member Y-EDCCL model library on the solid support without the intended target (Tris base modified sepharose beads).

4.10 Preparation for High-Throughput Sequencing

The eluted library members were subjected to two rounds of PCR to attach selection specific barcodes and adapter sequences for Illumina sequencing.
The length and purity of first PCR product was analyzed by 4% agarose electrophoresis. PCR product was then purified from the gel with Nucleospin Gel and PCR clean-up kit (Macherey-Nagel, Berlin, Germany). Next, PCR product was diluted to 10 nM with elution buffer in the kit to serve as template for second PCR. The second PCR product was analyzed by 4% agarose gel and purified from the gel with Nucleospin Gel and PCR clean-up kit (Macherey-Nagel, Berlin, Germany). The concentration of PCR purification product was measured by Nanodrop. Sample was diluted and submitted to Illumina high-throughput sequencing (HiSeq 2000, DKMS Life Science Lab, Dresden, Germany).

PCR were performed at 98 °C for 45 s, and 2 cycles of 98 °C for 30 s, 60 °C for 1 min, 72 °C for 30 s, then 13-18 cycles of 98 °C for 30 s, 65 °C for 1 min, 72 °C for 30 s, and finally 72 °C for 10 min.

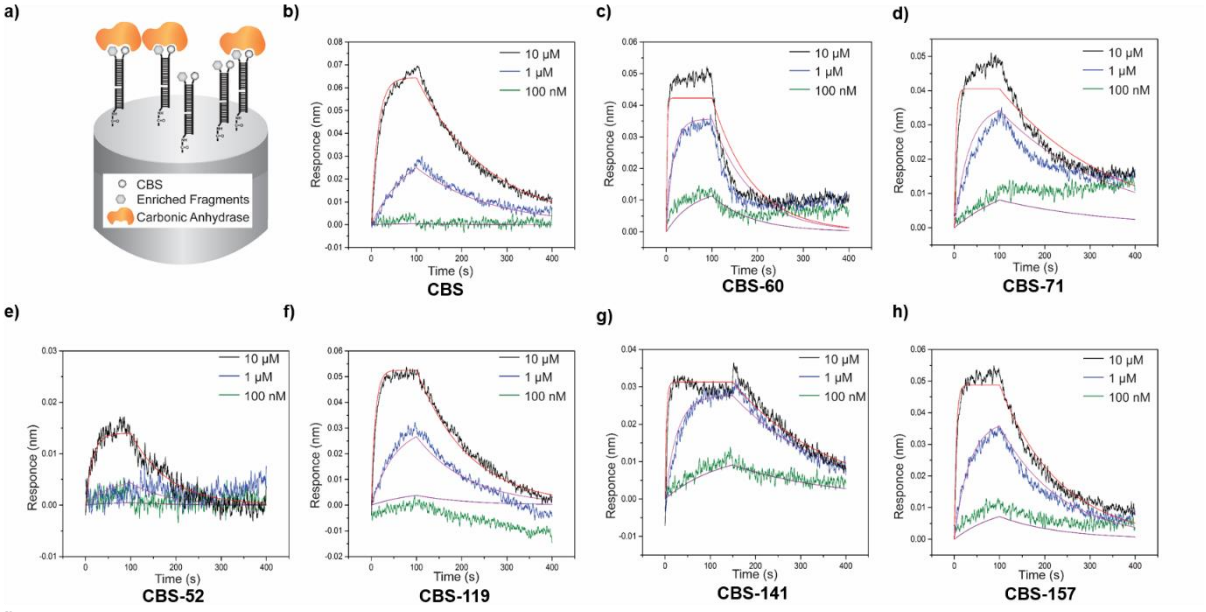
5. Hit validation with interferometry biosensor chip

Biosensors were functionalized by first soaking in dd H_2O for 10 minutes. Sensors were then functionalized with three cycles of soaking in 200 mM EDC, 50 mM NHS in DMF for 5 minutes, followed by soaking in a solution of Binding Profiler handle DNA (100 µM in dd H_2O) for 10 minutes. Following the third cycle, sensors were quenched by dipping into 100 mM ethanolamine. Functionalized sensors were stored in 1X PBS buffer.
DNA-conjugate compounds were mixed to a final concentration of 0.1 µM in 0.1M KH_2PO_4, pH 7.5. Functionalized sensors were loaded by soaking in this solution for 10

minutes. Loaded sensors were dipped in 1% SDS solution to remove any non-annealed DNA-conjugates.

Carbonic Anhydrase II was prepared as solutions of 10 µM, 1 µM, and 100 nM in Binding Profiler measurement buffer (10 mM HEPES, 1M NaCl, 0.05% Tween-20, pH 7.2). Measurements were performed on an Octet Red 386 platform. Affinity was measured by taking a 60 second baseline in measurement buffer, followed by a 150-second association measurement in protein solution, then a 300-second disassociation period in the same well used for the baseline. Remaining bound protein was removed by dipping in a 1% SDS solution. This process was repeated for each concentration of protein.

The measurements were blank-subtracted and set to align to the dissociation steps. Fitting was performed globally on each sensor.



i)

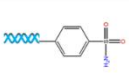
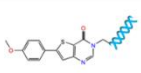
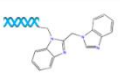
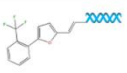
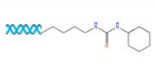
Fragment ID	CBS	CBS-60	CBS-71	CBS-119	CBS-141	CBS-157	CBS-52 (unenriched)
Structure							
K_d (nM)	1020 ± 13.6	211 ± 10.7	149 ± 5.92	819 ± 27.7	140 ± 2.56	337 ± 6.80	2100 ± 320
k_{on} (1/Ms)	5.21e+5 ± 7.60e+4	5.87e+4 ± 2.70e+3	3.00e+4 ± 1.03e+3	1.05e+4 ± 2.77e+2	3.36e+4 ± 5.30e+3	2.27E+4 ± 4.15E+2	7.80e+3 ±1.07 e+3
k_{off} (1/s)	5.22e-4 ± 2.47e-4	1.11e-2 ± 2.42e-4	3.94e-3 ± 8.29e-5	8.62e-3 ± 1.31e-4	4.72e-3 ± 4.29e-5	7.50e-3 ± 7.27e-5	1.64e-2 ±1.18e-3

Figure S18. (a) Scheme for Biolayer Interferometry kinetic binding assay where the fragments are paired via DNA-hybridization and immobilized on the biosensor surface. The biosensor is then exposed to the target protein at different concentration for the kinetic measurement. Carbonic Anhydrase II kinetic measurement. (b-h) Binding kinetics of CA II to immobilized CBS-fragment pairs. (i) Kinetic characterization (K_d, k_{on}, k_{off}) of different fragment pairs with the known binder CBS that were enriched in the Y-EDCCL selection.

6. Synthesis of CBS-hit conjugates

The selected two Compound 60 (5-hexylthiophene-2-carboxylic acid) and Compound 157 (3-{5-[3-(trifluoromethyl) phenyl]-2-furyl} acrylic acid) were linked to CBS via three different linkers (L1, L2, and L3).

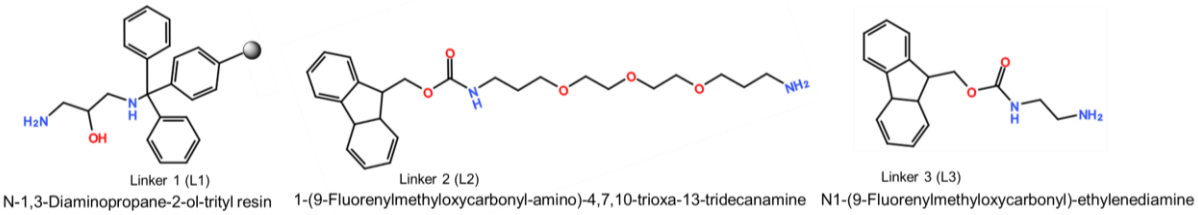


Figure S19. Structure of L1, L2, and L3 utilized to link hit compounds with CBS

CBS-L1 was synthesized by two steps. First, coupling of CBS on N-1, 3-Diaminopropane-2-ol-trityl resin (L1). 1 g of the resin was swollen in 5 mL DMF for 5 min three times. In the meanwhile CBS (2 eq, 0.7 mmol) was activated with HATU (2 eq) and HOAt (2 eq) in 5 mL DMF. 6 eq of DIPEA was added and reaction was stirred for 30 min at RT. The activated CBS was then added to the resin and incubated on an orbital shaker O/N. The resin was washed with DMF and DCM three times each. Second, CBS-L1 was cleaved from the resin by incubating in cleavage solution (10 % TFA, 5% water, and 85% DCM) for 1h at RT. The elution was collected left under fume hood for 2 h to evaporate DCM. The product was purified by HPLC and the best fraction was identified by LC-MS.

CBS-L2 and CBS-L3 synthesis were carried out in solution. 2 eq of preactivated CBS as above mentioned method were added to 1 eq (0.3 mmol) of L2 and L3 in DMF, accordingly. The mixtures were stirred at RT for 2 h. After identification of reaction product by LC-MS, the crudes were flash frozen in liquid nitrogen and solvent was evaporated by lyophilization. The condensed crude was then purified by HPLC and the best fraction was identified by LC-MS. To expose the second amino group on each linker, CBS-L2 and CBS-L3 were mixed with 20% piperidine in 5 mL ACN for Fmoc removal. After 1 h incubation, the crude was freeze-dried again to remove the solvent. The Fmoc-deprotected CBS-L2 and CBS-L3 were dissolved in ACN and purified by HPLC and the best fractions were identified by LC-MS.

CBS-L1, CBS-L2, and CBS-L3 were subjected to coupling of Compound 60 and 157. The compounds were activated in the same way as CBS activation using 2 eq relative to CBS-linker conjugates. To the solution of CBS-linker conjugates in DMF was added activated compounds and reaction last for 2 h at RT. The different products were purified by HPLC and the best fractions were identified by LC-MS, freeze-dried in speedvac O/N. The dried products were stored in -20℃.

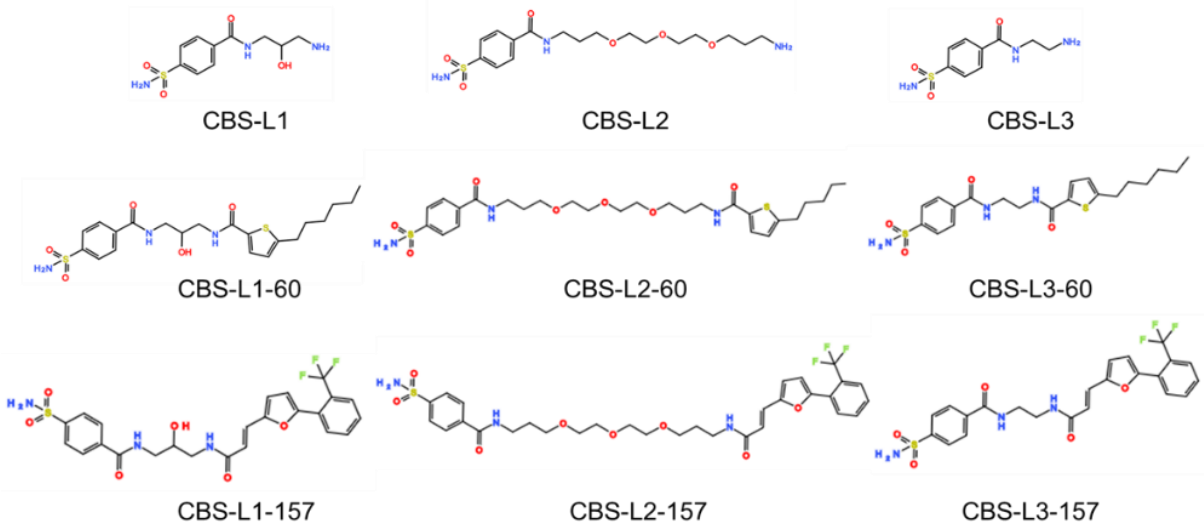


Figure S20. Structure of synthesized CBS-hit compound conjugates

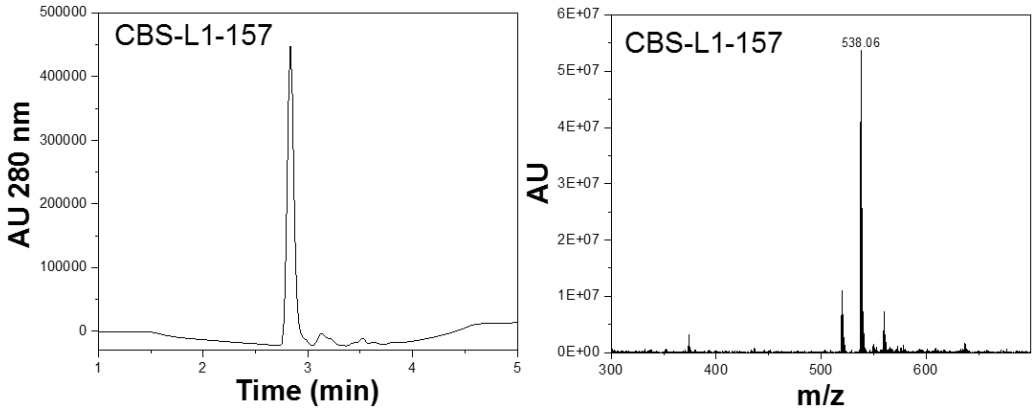


Figure S21. UPLC chromatogram and ESI-MS spectrum of the HPLC purified CBS-L1-157 compound.

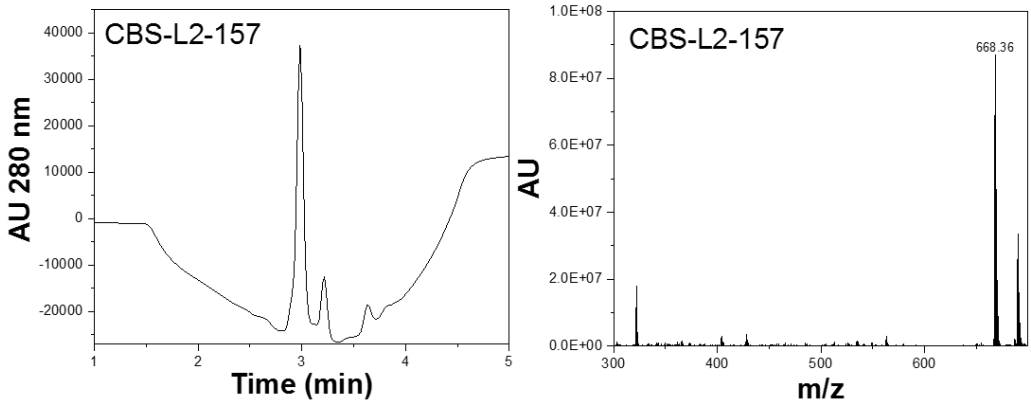


Figure S22. UPLC chromatogram and ESI-MS spectrum of the HPLC purified CBS-L2-157 compound.

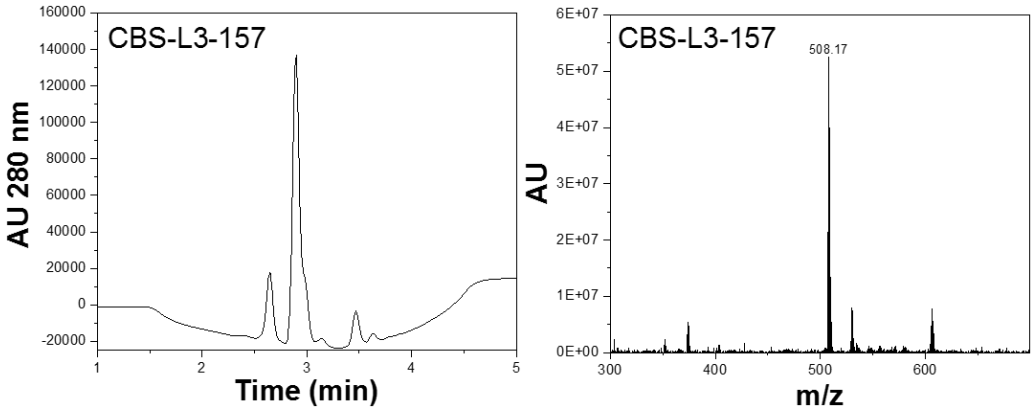


Figure S23. UPLC chromatogram and ESI-MS spectrum of the HPLC purified CBS-L3-157 compound.

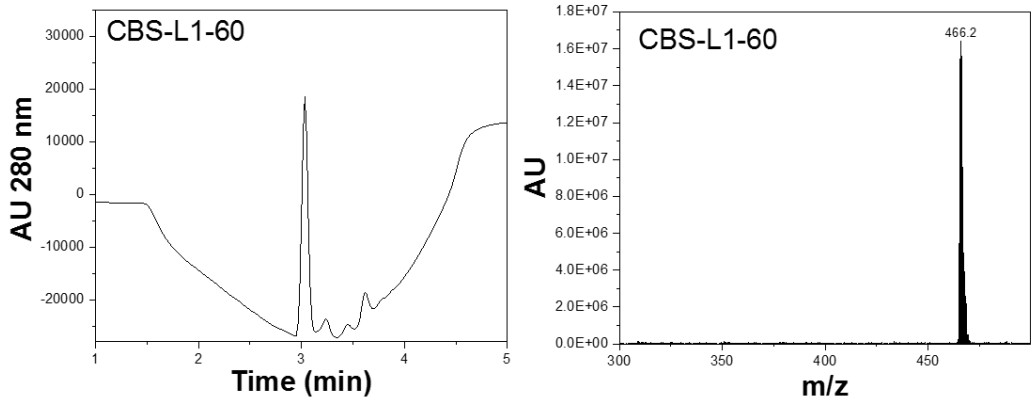


Figure S24. UPLC chromatogram and ESI-MS spectrum of the HPLC purified CBS-L1-60 compound.

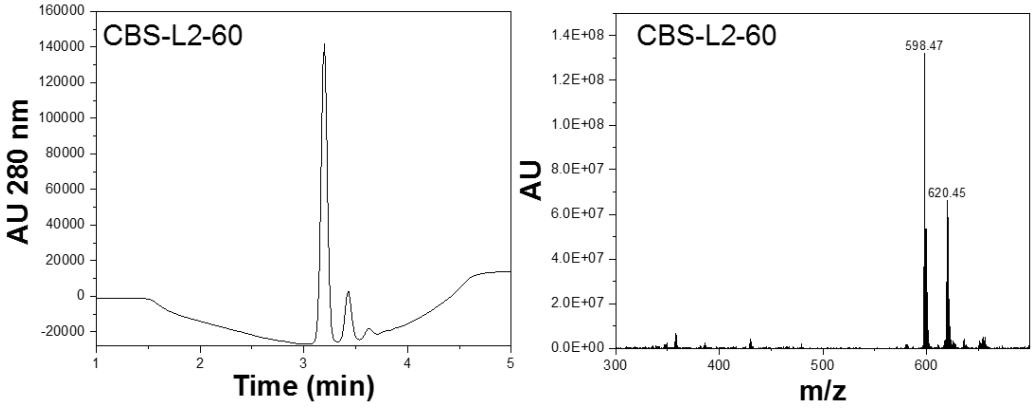


Figure S25. UPLC chromatogram and ESI-MS spectrum of the HPLC purified CBS-L2-60 compound.

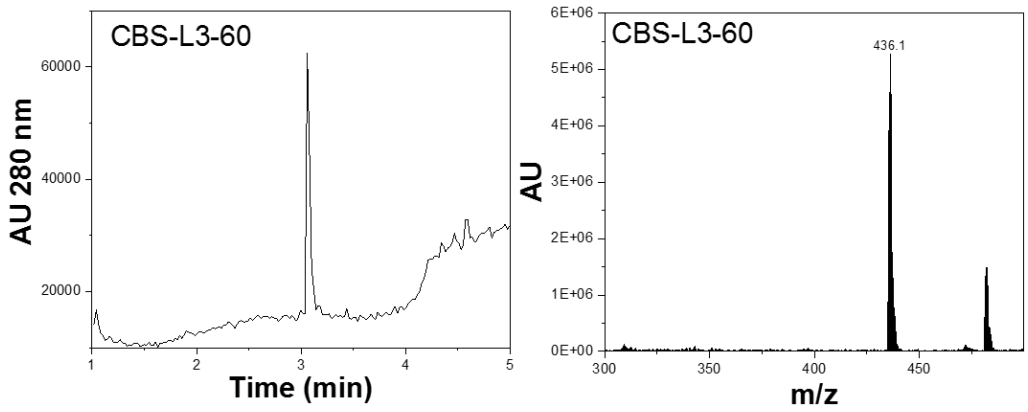


Figure S26. UPLC chromatogram and ESI-MS spectrum of the HPLC purified CBS-L3-60 compound.

7. Carbonic Anhydrase II inhibition assay

CAII inhibition assay was performed to measure the inhibitory effect of resynthesized CBS-conjugates. The assay was performed by using esterase activity of CA II with a chromogenic substrate p-nitrophenyl acetate (pNPA). The rates of hydrolysis were determined by increase of absorption at 410 nm after incubating different compounds with CAII. All possible compounds (CBS, CBS-L1, CBS-L2, CBS-L3, CBS-L1-60, CBS-L2-60, CBS-L3-60, CBS-L1-157, CBS-L2-157, and CBS-L3-157) were serial diluted in DMSO to reach final concentration from µM to nM via two-fold dilution with 10 data points. To 94 µL 100 mM HEPES, pH 7.4 buffer containing 650 nM CAII in each well was added 1 µL of compounds of each concentration in 384-well plate and the plate was incubated for 30 min prior to the measurement. 5 µL of pNPA (50 mM in 20 % DMSO) was added to plate right before the measurement. The absorption was measured at each 10 sec for 15 min using a Beckman Coulter's Paradigm Detection Platform (Brea, CA, USA). The initial Vmax was obtained from increase of absorption and the Vmax was plotted against inhibitor concentrations and logistic fitting was performed to obtain IC_{50} values using OriginPro 9.0G software.

8. Tables

Table S1: Uncoupled and standard oligonucleotides.

Name	Sequence	Type
Strand A'	GGAGGTTCAGACGACGCTCTTCCGATCTC	unmodified
Strand A	GGAGGTTCAGAGCTCGCTCTTCCGATCTC	5'-NH2
Strand C'	TGGTCTCAGCCGCCCTATTCTGAAGCCTACCTCC	unmodified
Strand C	TGGTCTCAGCCGCCCATCTCTGAAGCCTACCTCC	3'-NH2
Strand B'	GAGATCGGAAGAGCGTCG	unmodified
Strand B	GAGATCGGAAGAGCGAGC	unmodified
Strand D'	TCTGAAGGCTTCAGAATAGGGCGGCTGAGACCA	5'-PO4
Strand D	TCTGAAGGCTTCAGAGATGGGCGGCTGAGACCA	5'-PO4
Code-CBS-Y-5	CGCTCCGAATCTGAGCGATATCAGTACGTTCCTTCTGCATTCGGTCGG	5'-PO4
Code-Y-5-lig-adp	TTCGGAGCGCAAATACTCTGT	5'-PO4
Y-adapter	CGTCTACTGGGATCGCAGC	5'-PO4
Code-Y-3-lig-adp	CAGATCGAGCAACTCCAC	5'-PO4
Y-5-I/Y-3-I	GAGATCGGAAGAGCGAGCTCTGAAGGCTTCAGAGATGGGCGGCTGAGACC	unmodified
Y-5-I/Y-3-N	GAGATCGGAAGAGCGAGCTCTGAAGGCTTCAGAATAGGGCGGCTGAGACC	unmodified
Y-5-N/Y-3-I	GAGATCGGAAGAGCGTCGTCTGAAGGCTTCAGAGATGGGCGGCTGAGACC	unmodified
Y-5-N/Y-3-N	GAGATCGGAAGAGCGTCGTCTGAAGGCTTCAGAATAGGGCGGCTGAGACC	unmodified

Table S2: 5'-modified oligonucleotides:

Name	Sequence	Type	Expected mass	Measured mass
Im-A	<C12-Im>GGAGGTTCAGAGCTCGCTCTTCCGATCTC	5'-Iminobiotin	9349.1	9347.3
Van-A	<C6-Van>GGAGGTTCAGAGCTCGCTCTTCCGATCTC	5'-Vancomycin	10766.8	10762.1

Table S3: 3'-modified oligonucleotides

Name	Sequence	Type	Expected mass	Measured mass
Im-B	TGGTCTCAGCCGCCCATCTCTGAAGCCTACCTCC<C7-Im>	3'-Iminobiotin	10685.0	10681.9
Van-B	TGGTCTCAGCCGCCCATCTCTGAAGCCTACCTCC<C7-Van>	3'-Vancomycin	12187.7	12184.9

Table S4: CBS-hit conjugates

Name	Expected mass	Measured mass
CBS-L1-157	537.52	537.06
CBS-L2-157	667.88	667.36
CBS-L3-157	507.77	507.17
CBS-L1-60	467.6	467.2
CBS-L2-60	597.97	597.47
CBS-L3-60	437.86	437.1

Table S5: Primers for qPCR, PCR for Sanger sequencing and Illumina NGS

Name	Sequence
pB'	GAGATCGGAAGAGCGTCG
pB	GAGATCGGAAGAGCGAGC
pD'	TGGTCTCAGCCGCCCTAT
pD	TGGTCTCAGCCGCCCATC
pY-3	TGGTCTCAGCCGCCCTAT
pCBS	CCGACCGAATCCAGAAGG
Sanger F	ACACTCTTTCCCTACCCGACAACCTACTGCCTTACCTAACCTGAACCGACCGAATCCAGAAGG
Sanger R	CAAGCAGAAGACGGCATACGAGATTGGTCTCAGCCGCCCTAT
Sanger sequencing Primer	ACACTCTTTCCCTACCCGACAACCTACTG
Illumina F1	Same as Sanger F
Illumina R1	Same as Sanger R
Illumina F2	AATGATACGGCGACCACCGAGATCTACACTCTGAGCGATATACACTCTTTCCCTACCCGA
Illumina R2	CAAGCAGAAGACGGCATACGAGAT

References

[1] D. R. Halpin, J. A. Lee, S. J. Wrenn, P. B. Harbury, *Plos Biol* **2004**, *2*, 1031-1038.

[2] C. Ren, H. Wang, X. Zhang, D. Ding, L. Wang, Z. Yang, *Chem Commun* **2014**, *50*, 3473-3475.