Supporting Information

Mannocci *et al.* 10.1073/pnas.0805130105

SI Text

Reagents. Unless otherwise denoted, chemical compounds and proteins were from Sigma–Aldrich–Fluka, resin for solid phase synthesis from Novabiochem, enzymes from New England Biolabs, and HPLC grade lyophilized oligonucleotides were from IBA. SpinX columns were purchased from Corning–Costar and ion-exchange cartridges for DNA purification from Qiagen, (PCR purification cat. no. 28104, nucleotides removal cat. No. 28306) and used according to the protocol described by the provider.

Synthesis of DEL4000 Compounds. The individual organic compounds to be coupled to the 5′ amino-modified 42-mer oligonucleotides were dissolved to a DMSO stock solution (100 mM), occasionally by further addition of water or diluted hydrochloric acid. All HPLC were performed on an XTerra Prep RP_{18} column (5 μm, 10 × 150 mm) using a linear gradient from 10% to 40% MeCN in 100 mM TEAA, pH 7. LC-ESI-MS were performed on an XTerra RP_{18} column (5 μm, 4.6 × 20 mm) using a linear gradient from 0% to 50% MeOH over 1 min in 400 mM HFIP/5 mM TEA. The mass spectra were measured from *m/z* 900 to 2000 by a Quattro Micro instrument (Waters).

Coupling Reactions of 20 Fmoc-Protected Amino Acids. To a reaction volume of 310 μL containing 70% (vol/vol) DMSO/water, compounds were added to the respective final concentrations: Fmoc-protected amino acids DMSO solution, 4 mM; *N*-hydroxysulfosuccinimide in DMSO, 10 mM; *N*-ethyl-*N*′-(3-dimethylaminopropyl)-carbodiimide in DMSO, 4 mM; aqueous triethylamine hydrochloride solution (pH 9.0), 80 mM; oligonucleotide aqueous solution, 50 μM, (5′-GGA GCT TGT GAA TTC TGG XXX XXX GGA CGT GTG TGA ATT GTC-3′, XXX XXX unambiguously identifies the individual Fmoc-protected amino acid compound). All coupling reactions were stirred overnight at 25 °C; residual activated species were then quenched and simultaneously Fmoc-deprotected by addition of piperidine (500 mM in DMSO). Before HPLC purification, 500 μL of 100 mM TEAA (pH 7.0) was added to the reaction mixture. The reactions were then purified by HPLC, and the desired fractions were dried under reduced pressure, redissolved in 100 μL of water, and analyzed by LC-ESI-MS. The samples showed the expected Fmoc-deprotected products. Typical coupling yields were >51% overall. Four nanomoles of each DNA–compound conjugate were pooled to generate a 20-member DNA-encoded sublibrary.

Coupling Reactions of 200 Carboxylic Acids. To a reaction volume of 310 μL, containing 70% (vol/vol) DMSO/water, compounds were added to the respective final concentrations: DMSO-dissolved carboxylic acid, 4 mM; *N*-hydroxysulfosuccinimide in DMSO, 10 mM; *N*-ethyl-*N*′-(3-dimethylaminopropyl)-carbodiimide in DMSO, 4 mM; aqueous triethylamine hydrochloride (pH9.0), 80 mM; DNA-oligonucleotide sublibrary pool, 1.5 μM. All coupling reactions were stirred overnight at 25 °C; residual activated species were then quenched by addition of 50 μL of Tris-Cl buffer, 500 mM (pH 9.0). The mixture was allowed to quantitatively precipitate by sequential addition of 25 μL of 1 M acetic acid, 12.5 μL of 3 M sodium acetate buffer (pH 4.7), and 500 μL of ethanol followed by 2-h incubation at −23 °C. The DNA was centrifuged, and the resulting oligonucleotide pellet was washed with ice-cold 90% (vol/vol) ethanol and then dissolved in 100 μL of water. Test coupling reactions were also performed with the reaction conditions described above; using model 42-mer 5′-Fmoc-deprotected amino acid oligonucleotide conjugates and model carboxylic acids. The reactions were analyzed by HPLC and the masses of the reacted oligonucleotides detected by LC-ESI-MS. Typical HPLC coupling yields on this step were >52% with purity >46%.

Polymerase Klenow Encoding of 200 Carboxylic Acid Reactions. To a reaction volume of 50 μL, reagents were added to the respective final concentrations: pool 20 conjugates 320 nM, 44-mer oligonucleotide (5′-GTA GTC GGA TCC GAC CAC XXXX XXXX GAC AAT TCA CAC ACG TCC-3′, XXXX XXXX unambiguously identifies the individual carboxylic acid compound, IBA) 600 nM, Klenow buffer (cat. no. B7002S; NEB), dNTPs (cat. no. 11969064001; Roche), 0.5 mM, Klenow Polymerase enzyme (cat. no. M0210L; NEB), 5 units. The Klenow polymerization reactions were incubated at 37 °C for 1 h and then purified on ion-exchange cartridges (cat. no. 28306; Qiagen). The 200 purified reactions were dissolved in 50 μL of water each and pooled to generate the 4,000 member library (DEL4000) to a final total oligonucleotide concentration of 300 nM.

Preparation of D-Desthiobiotin Oligonucleotide-Conjugate (Positive Control). D-desthiobiotin-oligonucleotide conjugate was synthesized and unambiguously encoded as described above.

Library DEL 4000 Streptavidin Selections. The resulting library DEL 4000 (total oligonucleotide conjugate concentration 300 nM) was diluted 1:15 in PBS (20 nM final concentration), spiked with D-desthiobiotin oligonucleotide-conjugate (final concentration 1 pM). Fifty microliters of the library, 20 nM, was either added to 50 μL of streptavidin–Sepharose slurry (cat. no. 17-5113-01; GE Healthcare) or to 50 μL of Sepharose slurry without streptavidin. Both resins were preincubated with PBS and 0.3 mg/ml herring sperm DNA. After incubation for 1 h at 25 °C, the mixture was transferred to a SpinX column, the supernatant was removed, and the resin washed 4 times with 400 μL of PBS. After washing, the resin was resuspended in 100 μL of water.

Identification of Binding Molecules. The codes of the oligonucleotide–compound conjugates were amplified by PCR (total volume 50 μL, 25 cycles of 1 min at 94 °C, 1 min at 55 °C, 40 s at 72 °C) with either 5 μL of 100 fM DEL4000 library before selection as template, or 5 μL of each resuspended resin after selection as template. The PCR primers DEL_P1_A (5′-*GCC TCC CTC GCG CCA TCA G*GG AGC TTG TGA ATT CTG G-3′) and DEL_P2_B (5′-*GCC TTG CCA GCC CGC TCA G*GT AGT CGG ATC CGA CCA C-3′) additionally contain at one extremity a 19-bp domain (italicized) required for high-throughput sequencing with the 454 genome sequencer system. The PCR products were purified on ion-exchange cartridges. Subsequent high-throughput sequencing was performed on a 454 Life Sciences–Roche GS 20 sequencer platform (sequencing service by Eurofins MWG). Analyses of the codes from high-throughput sequencing were performed by an in-house program written in C++. The frequency of each code has been assigned to each individual pharmacophore.

Synthesis of the Binding Molecules as Fluorescein Conjugates. In a polypropylene syringe, 50 mg (46 μmol) of *O*-bis-(aminoethyl) ethylene glycol trityl resin (cat. no. 01-64-0235; Novabiochem) was suspended in a mixture of the appropriate Fmoc-protected

amino acid (100 μmol, 1 mL), HBTU (200 μmol, 1 mL; Aldrich), and DIEA (400 μmol, 0.5 mlL; Fluka) in dry DMF. After overnight incubation at 25 °C, the resin was washed 6 times with 2 mL of dry DMF. and the Fmoc moiety was removed by addition of 1 mL of piperidine (50% in dry DMF) for 1 h at 25 °C. After washing 6 times with 2 mL of dry DMF, the corresponding carboxylic acid (100 μmol, 1 mL of DMF) was added and a further amide bond formation reaction was performed as described above. The resulting product was cleaved by treating the resin 10 times with 2 mL of TFA (1% in CH_2Cl_2). The methylenchloride fractions were quenched in 5 mL of $NaHCO_3$ aq_{sat} and the water phase was back extracted 2 times with 5 mL of CH_2Cl_2. The pooled organic phases were washed 3 times with water, dried on Na_2SO_4 and concentrated *in vacuo*. The crude product was reacted with 2 equivalents of fluorescein isothiocyanate (800 μL of DMF) and 200 μL of$NaHCO_3$ aq_{sat} in the dark overnight at 25 °C. After HPLC purification on an XTerra Prep RP_{18} column (5 μM, 10 × 150 mm) using a linear gradient from 10% to 100% MeCN 0.1% TFA, the desired fractions were collected and lyophilized. ESI-MS analysis confirmed the mass of the expected FITC conjugated products: **02-78** (C45H49BrN4O10S2) expected: 949.93 measured: 951.31 [M+H$^+$]; **07-78** (C50H59N5O12S2) expected: 985.36 measured: 986.37 [M+H$^+$]; **15-78** (C44H49N5O12S3) expected: 935.25 measured: 936.25 [M+H$^+$]; **02-107** (C47H47BrN4O9S) expected: 923.87 measured: 925.12 [M+H$^+$]; **13-40** (C46H49N5O14S) expected: 927.30 measured: 928.42 [M+H$^+$]; **11-78** (C49H58N4O13S2) expected: 974.34 measured: 975.41; **17-49** (C45H49IN4O11S) expected: 980.22 measured: 981.29 [M+H$^+$]; **17-78** (C45H49IN4O10S2) expected: 996.19 measured: 997.26; **16-78** (C43H48N4O10S3) expected: 876.25 measured: 877.33; **15-117** (C45H45N5O12S2) expected: 911.25 measured: 912.33; **02-49** (C45H49BrN4O11S) expected: 932.23; measured: 933.32 [M+H$^+$].

Affinity Measurements. Fluorescein–compound conjugates (500 nM) were incubated with increasing amounts of streptavidin (cat. no. S002-6; BIOSPA) in PBS, 5% DMSO, for 1 h at 25 °C. The fluorescence polarization was determined with a TECAN Polarion instrument by excitation at 485 nm and measuring emission at 535 nm (ε = 72,000 $M^{-1}cm^{-1}$). [All of the curves were fitted by applying the following formula: $y = a \times 0.5 \times (((x + \text{const} + b) - \sqrt{((x + \text{const} + b)\hat{2} - 4 \times x \times \text{const})}))$; const = 500 nM.]

Polyclonal Human IgG Coating of Sepharose Beads. One hundred micrograms of CNBr-activated Sepharose (GE Healthcare) was swollen in 1 mM HCl, washed, and mixed in separate tubes with 2.5 mg/ml PBS solution of polyclonal human IgG (Sigma–Aldrich–Fluka. After 4-h incubation at 4 °C, the slurry was repeatedly and alternatively washed with 0.1 M $NaHCO_{3aq}$ 0.1 M Tris-Cl, 0.5 M NaCl (pH 8.3), and 0.1 M NaOAc, 0.5 M NaCl (pH 4) and then stored in PBS at 4 °C.

Library DEL 4000 Polyclonal Human IgG Selection. The library DEL4000 (total oligonucleotide conjugate concentration 300 nM) was diluted 1:15 in PBS (20 nM final concentration). Fifty microliters of the library 20 nM was added to 50 μL of IgG–Sepharose slurry. The resin was preincubated with PBS, 0.3 mg/ml herring sperm DNA (Sigma). After incubation for 1 h at 25 °C, the mixture was transferred to a SpinX column (Corning–Costar), the supernatant was removed, and the resin washed 4 times with 400 μL of PBS. After washing, the resin was resuspended in 100 μL of water.

Identification of Human IgG-Binding Molecules. The codes of the oligonucleotide–compound conjugates were amplified by PCR (total volume 50 μL, 25 cycles of 1 min at 94 °C, 1 min at 5 5°C, 40 s at 72 °C) with either 5 μL of 100 fM DEL4000 library before selection as template, or 5 μL of each resuspended resin after selection as template. The PCR primers DEL_P1_A (5′-*GCC TCC CTCGCG CCA TCA G*GG AGC TTG TGA ATT CTG G-3′) and DEL_P2_B (5′-GCC TTG CCA GCC CGC TCA *GGT AGT CGG ATC CGA CCA C*-3′) additionally contain, at one extremity, a 19-bp domain (italicized) required for high-throughput sequencing with the 454 genome sequencer system. The PCR products were purified on ion-exchange cartridges. Subsequent high-throughput sequencing was performed on a 454 Life Sciences–Roche GS 20 sequencer platform. Analyses of the codes from high-throughput sequencing were performed by an in-house program written in C++. The frequency of each code has been assigned to each individual pharmacophore.

Synthesis of Affinity Chromatography Resin Containing the Compounds 02-40 or 16-40. In a polypropylene syringe, 129 mg (120 μmol) of *O*-bis-(aminoethyl)ethylene glycol trityl resin (cat. no. 01-64-0235; Novabiochem) was suspended in a mixture of the appropriate Fmoc-protected amino acid (180 μmol, 1 mL), HBTU (Aldrich, 360 μmol, 1 mL), and DIEA (Fluka, 720 μmol, 0.5 mL) in dry DMF. After overnight incubation at 25 °C, the resin was washed 6 times with 2 mL of dry DMF and the Fmoc moiety was removed by addition of 3 mL of piperidine (50% in dry DMF) for 1 h at 25 °C. After washing 6 times with 2 mL of dry DMF, 4-(4-(1-hydroxyethyl)-2-methoxy-5-nitrophenoxy)butanoic acid (**40**, 54 mg, 180 μmol, 1 mL of DMF) was added, and a further amide bond formation reaction was performed as described above. The resulting product was cleaved by treating the resin 10 times with 2 mL of TFA (1% in CH_2Cl_2). The dichloromethylene fractions were quenched in 5 mL of $NaHCO_3$ aq_{sat} and the water phase was back extracted 2 times with 5 mL of CH_2Cl_2. The pooled organic phases were washed 3 times with water, dried on Na_2SO_4, and concentrated *in vacuo*. After HPLC purification on an XTerra Prep RP_{18} column (5 μM, 10 × 150 mm) using a linear gradient from 10% to 100% MeCN 0.1% TFA, the desired fractions were collected and lyophilized. ESI-MS analysis confirmed the mass of the expected products: [M+H$^+$]; **02-40** (C29H41BrN4O9) expected: 668.21 measured: 669.37 [M+H$^+$]; **16-40** (C27H40N4O9S) expected: 596.25 measured: 597.12 [M+H$^+$]. 200 mg CNBr-activated Sepharose (GE Healthcare) was swollen in 1 mM HCl, washed, and mixed in separate tubes with 15 μmol of the compounds dissolved in 2 mL of 0.1 M $NaHCO_{3aq}$, 10% DMF. After 4-h incubation at 25 °C, the slurry was repeatedly and alternatively washed with 0.1 M $NaHCO_{3aq}$, 0.1 M Tris-Cl, 0.5 M NaCl (pH 8.3), and 0.1 M NaOAc, 0.5 M NaCl (pH 4) and then stored in PBS at 4 °C.

Polyclonal Human IgG Cy5 Labeling. Polyclonal human IgG (Sigma–Aldrich–Fluka) was labeled with Cy5 Mono-reactive kit (cat. no. PA25001; Amersham) according to the protocol of the provider and purified over a PD10 column (cat. no. 17-0851-01; GE Healthcare) as described by the supplier.

Biotinylated Polyclonal Human IgG. Polyclonal human IgG (Sigma–Aldrich–Fluka) was labeled with NHS-LC-Biotin reagent (cat. no. 21336; Pierce) according to the protocol of the provider and purified over a PD10 column (cat. no. 17-0851-01; GE Healthcare) as described by the supplier.

Affinity Chromatography of CHO Cells Supernatant Spiked with Human IgG Cy5-Labeled or Biotinylated Human IgG on IgG-Binding Resin. Seventy milligrams of the resin containing compound **02-40** resin were loaded on a chromatography cartridge (cat. no. 20-0030-00; Glen Research) and washed 3 times with 1 mL of PBS before loading a CHO cell supernatant (60 μL) spiked with human IgG Cy5-labeled (40 μL, 9.68 μM) or with biotinilated human IgG (30 μL, 17.2 μM). The flow-through, the washing fractions (washing 1 time with 10 mL of PBS; 1 time with 10 mL of 500

mM NaCl, 0.5 mM EDTA; 1 time with 10 mL of 100 mM NaCl, 0.1% Tween 20, 0.5 mM EDTA) and the elutate (elution 3 times with 200 μL of aqueous triethylamine, 100 mM) were collected and concentrated back to a final volume of 100 μL by centrifugation in a Vivaspin 500 tube (cat. no. VS0101, cut-off 10.000 MW; Vivascience). The samples were then analyzed by gel electrophoresis on a NuPAGE 4–12% Bis-Tris Gel (cat. no. NP0321; Invitrogen) by using Mops SDS as running buffer and stained with Coomassie blue. Cy5 activity was detected by a Diana III Chemiluminescence Detection System (Raytest) by excitation at 675 nm and measuring emission at 694 nm (ε = 250,000 $M^{-1}cm^{-1}$).

Western blot analysis has been performed by transferring the proteins to NC membrane (Millipore) with the Xcell II blot module (Invitrogen) by using standard procedures. The membrane was quickly rinsed with water before soaking twice in methanol. The membrane was dried at room temperature for 15 min and incubated for 1 h with 1:500 dilutions in 4% defatted milk-containing PBS of the following protein: Streptavidin–horseradish peroxidase conjugate (HRP–Streptavidin cat. no. RPN1231V; Amersham Biosciences). For detection of immunoreactive bands, the membrane was washed 3 times for 5 min with PBS and soaked in chemiluminescent reagent (ECL1plus Western Blotting Detection System from Amersham Biosciences) for 5 sec and exposed to BioMax films (Kodak) in an autoradiographic cassette.

Human MMP3 Coating of Sepharose Beads. CNBr-activated Sepharose (100 mg; GE Healthcare) was swollen in 1 mM HCl, washed, and mixed in separate tubes with 1 mg/mL polyclonal human IgG (Sigma–Aldrich–Fluka) dissolved in. After 4-h incubation at 4 °C, the slurry was repeatedly and alternatively washed with 0.1 M $NaHCO_{3aq}$, 0.1 M Tris-Cl, 0.5 M NaCl (pH 8.3), and 0.1 M NaOAc, 0.5 M NaCl (pH 4) and then stored in PBS at 4 °C.

Library DEL4000 human MMP3 Selection. The library DEL4000 (total oligonucleotide conjugate concentration 300 nM) was diluted 1:15 in PBS (20 nM final concentration). Fifty microliters of the library 20 nM was added to 50 μL of MMP3–Sepharose slurry. The resin was preincubated with PBS, 0.3 mg/ml herring sperm DNA (Sigma). After incubation for 1 h at 25 °C, the mixture was transferred to a SpinX column (Corning–Costar), the supernatant was removed, and the resin washed 4 times with 400 μL of PBS. After washing, the resin was resuspended in 100 μL of water.

Identification of Human MMP3-Binding Molecules. The codes of the oligonucleotide–compound conjugates were amplified by PCR (total volume 50 μL, 25 cycles of 1 min at 94 °C, 1 min at 55 °C, 40 s at 72 °C) with either 5 μL of 100 fM DEL4000 library before selection as template or 5 μLof each resuspended resin after selection as template. The PCR primers DEL_P1_A (5′-*GCC TCC CTC GCG CCA TCA G*GG AGC TTG TGA ATT CTG G-3′) and DEL_P2_B (5′-GCC TTG CCA GCC CGC TCA *GGT AGT CGG ATC CGA CCA C*-3′) additionally contain at one extremity a 19-bp domain (italics) required for high-throughput sequencing with the 454 genome sequencer system. The PCR products were purified on ion-exchange cartridges. Subsequent high-throughput sequencing was performed on a 454 Life Sciences–Roche GS 20 sequencer platform. Analyses of the codes from high-throughput sequencing were performed by an in-house program written in C++. The frequency of each code has been assigned to each individual pharmacophore.

Synthesis of the MMP3-Binding Molecules as Fluorescein Conjugates. In a polypropylene syringe, 50 mg (46 μmol) of *O*-bis-(aminoethyl)ethylene glycol trityl resin (cat. no. 01-64-0235; Novabiochem) was suspended in a mixture of the appropriate Fmoc-protected amino acid (100 μmol, 1 mL), HBTU (200 μmol, 1 mL; Aldrich), and DIEA (400 μmol, 0.5 mL; Fluka) in dry DMF. After overnight incubation at 25 °C, the resin was washed 6 times with 2 mL of dry DMF, and the Fmoc moiety was removed by addition of 1 mL of piperidine (50% in dry DMF) for 1 h at 25 °C. After washing 6 times with 2 mL of dry DMF, the corresponding carboxylic acid (100 μmol, 1 mL of DMF) was added, and a further amide bond formation reaction was performed as described above. The resulting product was cleaved by treating the resin 10 times with 2 mL of TFA (1% in CH_2Cl_2). The methylenchloride fractions were quenched in 5 mL of $NaHCO_3$ aq_{sat}, and the water phase was back extracted 2 times with 5 mL of CH_2Cl_2. The pooled organic phases were washed 3 times with water, dried on Na_2SO_4 and concentrated *in vacuo*. The crude product was reacted with 2 equivalents of fluorescein isothiocyanate (800 μL of DMF) and 200 μ of $NaHCO_3$ aq_{sat} in the dark overnight at 25 °C. After HPLC purification on an XTerra Prep RP_{18} column (5 μM, 10 × 150 mm) using a linear gradient from 10% to 100% MeCN 0.1% TFA, the desired fractions were collected and lyophilized. ESI-MS analysis confirmed the mass of the expected FITC conjugated products: **02-118** ($C_{55}H_{63}BrN_4O_{10}S$) expected: 1052.08 measured: 1053.30 [M+H$^+$]; **13-17** ($C_{49}H_{55}N_5O_{11}S$) expected: 921.36 measured: 922.42 [M+H$^+$]; **15-117** ($C_{45}H_{45}N_5O_{12}S_2$) expected: 911.25 measured: 912.33 [M+H$^+$]; **17-104** ($C_{46}H_{43}I_2N_5O_{10}S$) expected: 1111.08 measured: 1112.19 [M+H$^+$]; **18-96** ($C_{49}H_{45}BrN_4O_9S$) expected: 945.87 measured: 947.10 [M+H$^+$].

Stepwise Encoding. Gel electrophoresis was performed either using 15% Tris-Borate-EDTA-Urea denaturing polyacrylamide gels (TBE-Urea, cat. no. EC68852; Invitrogen) or 20% Tris-Borate-EDTA native polyacrylamide gels (TBE, cat. no. EC63152; Invitrogen) and stained with SYBR green II. DNA ethanol precipitation of DNA was performed by adding 1/10 vol of 3 M AcOH/AcONa buffer (pH 4.7) and 3 vol of ethanol relative to the volume of the DNA sample. After 2-h incubation at −23 °C, the mixture was centrifuged in a table-top centrifuge for 40min (16,000 × *g*) at 4 °C, the supernatant removed, and the pellet washed with 300 μL of ice-cold ethanol 90%. After a further 20-min centrifugation (16.000 × *g*) at 4/°C, the pellet was dried and redissolved in water.

Stepwise Encoding by Ligation (Method A Italicized Sequences Represent Coding Sequences). Hybridization of 3 pairs (A, B, C) of oligonucleotides (A: 5′-CAT GGA ATT CGC TCA CTC C*GA CTA G*AG G-3′ and 5′-(Phosphate)-CGT ACC T*CT AGT C*GG AGT GAG CGA ATT CCA TG-3′; B: 5′-(Phosphate)-TAC GTG AGC *TTG ACC* TGG TGA G-3′ and 5′- (Phosphate)-GCT TCT CAC CA*G GTC AA*G CTC A-3′; C: 5′-(Phosphate)-AAG C*AC GTT C*GC TGG ATC CTC AAC TGT G-3′ and 5′-CAC AGT TGA GGA TCC AGC *GAA CGT*-3′) was carried out by mixing the oligonucleotides at a concentration of 1.25 μM per oligonucleotide in 1× ligase buffer [40 mM Tris·HCl, 10 mM MgCl2, 10 mM DTT, 0.5 mM ATP (pH 7.8)] and incubating the mixtures for 10 min at 50 °C. Subsequently, the ligations were performed by mixing 10 μL of hybridized oligonucleotide pairs A and B with 10 μL of 1× ligase buffer and 1 μL of T4 ligase (Roche Applied Science), and incubated at 25 °C for 2 h. The ligation product was purified by using a Qiagen Nucleotide Removal kit and eluted with 50 μL of 10 mM Tris·HCl (pH 8.0). Eighteen microliters of the eluate was mixed with 10 μL of hybridized oligonucleotide pair C (which was present in excess), 2 μL of 10× ligase buffer, and 1 μL of T4 ligase and incubated for 2 h at 25 °C. Aliquots of the two starting oligonucleotides and the different ligation products were subjected to electrophoresis on a 20% TBE gel.

Stepwise Encoding by "Hybrid Klenow-Ligation" (Method B Italicized Sequences Represent Coding Sequences). To a reaction volume of 50 μL, reagents were added to the respective final concentrations: a 42-mer 5′-amino-C12-DNA-oligonucleotide (5′-GGA GCT TGT GAA TTC TGG *ATC TTA* GGA CGT GTG TGA ATT GTC-3′), 2 μM, a 42-mer 5′-C6-biotinylated-oligonucleotide containing the nonpalindromic BssSI restriction site (in bold type) (5′-GTA GTC GGA **CAC GAG** TAC *TGG TAA TC*G ACA ATT CAC ACA CGT CC-3′), 3 μM, Klenow buffer, dNTPs (cat. no. 11969064001; Roche), 0.5 mM, and Klenow Polymerase enzyme, 5 units. After incubation at 37 °C for 1 h, the reaction mixture was purified on ion-exchange cartridge and eluted in 25 μL of water. Eight units of BssSI enzyme were added to the purified Klenow product in 50 μL of BssSI restriction buffer. The restriction cutting reaction was carried out at 37 °C for 1.5 h. Streptavidin–Sepharose slurry (50 μL, cat. no. 17-5113-01; GE Healthcare) was added and the slurry was incubated for 30 min at 4 °C. After SpinX column centrifugation, the supernatant was collected and purified on ion-exchange cartridge and eluted in 25 μL of water. Subsequently, the following reagents were added to the final volume of 50 μL: preincubated mixture 1:1 of hybridized oligonucleotides (27-mer 5′-phosphate-TCG TGA AAT TTG *CT*A GGA TCC ATA TTG-3′ and 23-mer 5′-CAA TAT GGA TCC T*AG CAA A*TT TC-3′), 3 μM, T4 ligase buffer (Roche Applied Science) and T4 ligase (Roche Applied Science), 4 units. The ligation was performed overnight at 16 °C and then purified on ion-exchange cartridge. Aliquots of the starting oligonucleotides, and the different Klenow, restriction, and ligation products were analyzed on a 15% TBE-Urea gel. Sequencing of the excised band after 3 stepwise encoding confirmed the identity of the expected product.

Stepwise Encoding by Klenow Polymerase (Method C Italicized Sequences Represent Coding Sequences). To a reaction volume of 50 μL, reagents were added to the respective final concentrations: a 42-mer 5′-amino-C12-DNA-oligonucleotide (5′-GGA GCT TGT GAA TTC TGG *ATC TTA* GGA CGT GTG TGA ATT GTC-3′), 2 μM, 42-mer 3′-C6-biotinylated-oligonucleotide (5′-GTA GTC GGA TCC GAC CAC *GTT CCT* GAC AAT TCA CAC ACG TCC-3′), 3 μM, Klenow buffer, dNTPs (cat. no. 11969064001; Roche), 0.5 mM, Klenow Polymerase enzyme, 5 units. The Klenow polymerization reaction was incubated at 37 °C for 1 h, purified on ion-exchange cartridge, and eluted in 100 μL of 4 M urea. After incubating at 94 °C for 2 min, 50 μL of streptavidin–Sepharose slurry (cat. no. 17-5113-01; GE Healthcare) were added, and the slurry was incubated for 1 h at 4 °C. The streptavidin Sepharose resin and the supernatant were separated by centrifugation in a SpinX column. The DNA in the supernatant was ethanol precipitated as described above. The resulting single-stranded oligonucleotide was mixed with a 42-mer unmodified DNA oligonucleotide (5′-GTC GTA TCG CCA TGG TCC *AAC ATC* GTA GTC GGA GAG GAC CAC-3′), and a Klenow polymerization reaction was performed as described above. Aliquots of the 3 starting oligonucleotides and the different Klenow products were applied on a 15% TBE-Urea gel.

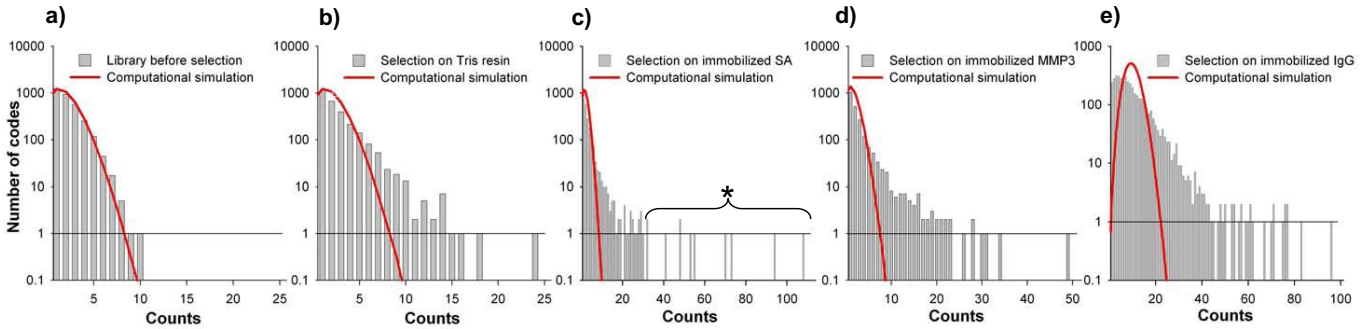

Fig. S1. Simulated and experimental distribution of sequence counts observed for members of the library before selection (*A*) and after selection on Tris-quenched resin (*B*), and streptavidin-resin (*C*) as well as resin coated with human matrix metalloproteinase 3 (*D*) and with polyclonal human IgG (*E*). The plots display the number of codes (i.e., DNA-encoded compounds in the library) that were observed with a given number of counts (i.e., number of sequences) either in the experimental 454-assisted sequence of PCR (performed before or after selection) or in a computer-assisted simulation. Although in the library before selection, experimental findings and simulation are in excellent agreement, in selection experiments, certain compounds are enriched much more compared with what would be predicted from the statistical distribution of sequence counts in an equimolar mixture of compounds. The sequences of compounds in plot *C* identified with an asterisk were found more than 30 times; these compounds were then chosen for the experimental affinity determination (Fig. 3). The individual plots exhibit a different maximum for the simulated curve of number of codes observed with a certain number of counts, due to differences in the overall number of experimental sequences (e.g., 7,336 overall sequence counts for the library before selection; 39,032 overall sequence counts for IgG selections). The simulated distribution of number of codes represented by individual counts, which are related to the probability that certain counts are experimentally found in a nonbiased mixture of equimolar compounds, was computed by using a home-written software program. The basic principle used in the simulation relies on the computer-assisted random generation of numbers corresponding to any of the 4,000 compounds in the library. The repetition of the simulation more than once allows the computation of fractional values for the number of codes associated to a given ''count'' value. For example, a number of codes value of 0.1 corresponds to the observation of a given ''Counts'' value in only 1 of 10 simulations, each performed with a total number of counts equal to the total number of experimental sequences in a given experiment.

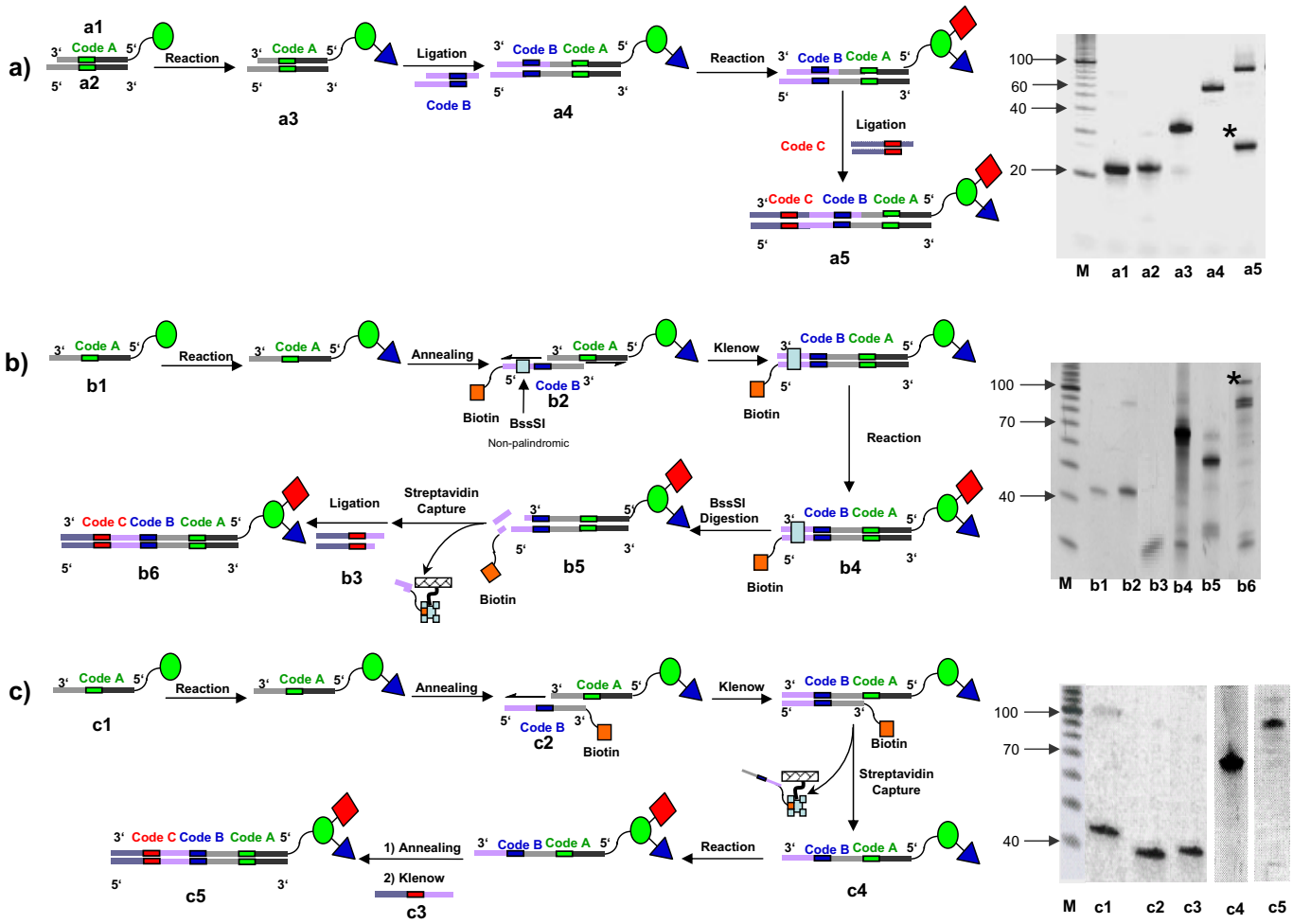

Fig. S2. Schematic representation of three different alternatives for the stepwise encoding of chemical libraries, consisting of three sets of building blocks. The first scheme (*A*) features the stepwise addition of groups of chemical moieties onto an initial scaffold, by using suitable orthogonal chemical reactions and/or protecting strategies, followed by the sequential addition of the corresponding DNA codes by an iterative ligation procedure. This scheme is conceptually simple and can be implemented experimentally but requires 2 double-stranded DNA fragments with sticky ends for each encoding event (i.e., 200 + 200 + 200 oligonucleotides for a library containing 100 × 100 × 100 chemical groups). The second synthetic and encoding strategy (*B*) represents a natural extension of the encoding strategy depicted in Fig. 1. Here, the double-stranded DNA fragment generated by Klenow fill-in is followed by a digestion with a nonpalindromic cutter and by a subsequent ligation step with a complementary double-stranded DNA fragment. Finally, one could envisage a synthetic strategy featuring only the use of Klenow polymerization (*C*), which would require the lowest number of oligonucleotides for library encoding (100 + 100 + 100 oligonucleotides for a library containing 100 × 100 × 100 chemical groups). The feasibility of the experimental procedures (*A*–*C*) was demonstrated by gel- electrophoretic analysis, which monitored the stepwise assembly of DNA-fragments of suitable size, and by DNA sequencing. (*A*) Encoding strategy based on the sequential ligation of double-stranded DNA fragments. Native PAGE analysis with a 20% TBE gel revealed the identity and purity of the DNA fragments used in the encoding procedure. M: marker; a1, single-strand 28-mer DNA fragment; a2, single-stranded 32-mer DNA fragment; a3, hybridization of 28-mer (a1) with the 32-mer (a2) DNA fragments; a4, double-stranded DNA 50-mer first ligation step product; a5, double-stranded 78-mer second ligation step product. *, The band is the hybridized oligonucleotide (of a 28-mer and 24-mer) carrying the Code C which was used in excess. (*B*) Encoding strategy based on the formation of a double-stranded DNA fragment by a Klenow-assisted polymerization step, followed by the ligation of a DNA-fragment carrying the third code. Denatured PAGE analysis using a 15% TBE-Urea gel revealed the purity and identity of the DNA fragments generated in the encoding steps. M, marker; b1, single-strand 42-mer DNA fragment; b2, 44mer partially complementary 5′-biotinylated single-stranded DNA fragment; b3, 27-mer and 23-mer hybridized DNA ligation fragments; b4, Klenow-assisted polymerization 68-mer product; b5, BssSI digestion product (54-mer); b6, full-length (81-mer) DNA fragment. *, The band is an artifact resulting from incomplete denaturation. If excised, extracted, and loaded on a gel, this band migrates at the expected height of a double-stranded 81-base DNA fragment.(*C*) Encoding strategy based on the formation of a double-stranded DNA fragment by the sequential use of 2 Klenow-assisted polymerization steps, starting from partially complementary oligonucleotides. Denatured PAGE analysis performed by using a 15% TBE-Urea gel reveled the purity and identity of the DNA fragments generated in the 3 Klenow-mediated encoding steps. M, marker; c1, single-stranded 5′-aminomodified 42-mer DNA fragment; c2, partially complementary 3′-biotinylated single stranded DNA fragment; c3, 42-mer single-strand DNA fragment partially complementary to first Klenow step product; c4, single-strand 66-mer DNA product after first Klenow step polymerization and purification; c5, full-length (90-mer) DNA fragment, after purification.

Table S1. Dissociation constants toward the different targets

Fluorescent compound	Counts after DEL4000 selection	Streptavidin K_d, μM	Lysozyme K_d, μM	Carbonic anhydrase K_d, μM
13-40	2	54	384	703
11-78	48	3.5	753	781
17-49	32	0.35	1.9e3	225
17-78	70	0.37	834	264
16-78	55	1.1	5.2e3	1e5
15-117	0	99	1.3e7	1.58e8
02-49	41	0.80	1.9e4	452
02-78	108	0.38	3.9e3	5.4e7
07-78	73	11	9.8e8	448
15-78	7	79	1.9e7	6.9e6
02-107	0	50	694	1.4e8

Other Supporting Information Files

Dataset S1 (PDF)