Supporting Information for

"COMBINES-CID: An efficient method for de novo engineering of highly specific chemically induced protein dimerization systems"

Shoukai Kang, Kristian Davidsen, Luis Gomez-Castillo, Huayi Jiang, Xiaonan Fu, Zengpeng Li, Yu Liang, Molly Jahn, Mahmoud Moussa, Frank DiMaio and Liangcai Gu

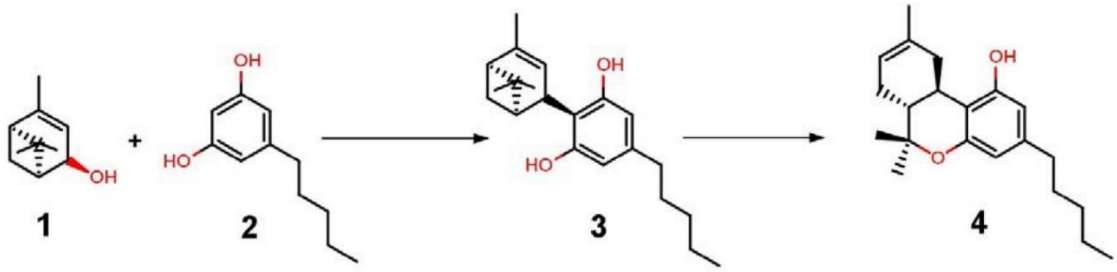
Table of Contents

SUPPLEMENTARY METHODS

1. Chemical synthesis

Biotinylated tetrahydrocannabinol (THC) (**8**) and cannabidiol (CBD) (**13**) were provided by the University of Washington Institute for Protein Design and the synthesis was performed by DermaXon LLC as described below.

All solvents were reagent grade and used without further purification. All procedures were carried out at room temperature (r.t.) unless otherwise stated. All glassware was oven-dried at 60˚C prior to use. Magnesium sulfate was used as the drying agent. Yields refer to chromatographically pure compounds as determined by TLC or HPLC analysis. The identities and purities of all final products were checked by HPLC-ESI-MS, ^{1}H, and ^{13}C (including APT and DEPT) experiments. NMR spectra were obtained on a Bruker Ascend TM-400 instrument. All spectra were baseline-corrected. All ^{13}C NMR spectra were obtained with complete proton decoupling. The signals in the ^{13}C NMR spectra were assigned by means of DEPT (Distortionless Enhancement by Polarization Transfer): these methods enable differentiation between the resonances of quaternary carbons (C) and the carbons of CH, CH_2 and CH_3 groups. NMR results of all target compounds were consistent with the assigned structures. High Resolution mass spectrometry (HRMS) analyses were performed on a Waters/Micromass LCT-TOF instrument.

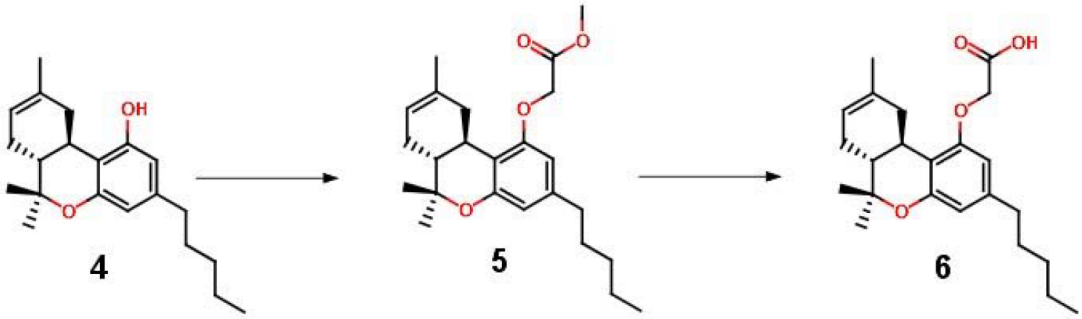


Protocols adapted for synthesis of compound **4** Δ^8-THC from a previous report.[1]

5-pentyl-2-[(1R,5S)-4,6,6-trimethylbicyclo[3.1.1]hept-3-en-2-yl]benzene-1,3-diol (3). To a solution of olivetol **2** (1.98 g, 11 mmol) and p-toluenesulfonic acid (191 mg, 1.1 mmol) in $CHCl_3$ (65 mL) at r.t. under N_2 was added (s)-cis-verbenol **1** (1.7 g, 11 mmol) in $CHCl_3$ (30 mL). After being stirred at r.t. for 2.5 h, the reaction mixture was poured into saturated $NaHCO_3$ and extracted two times with CH_2Cl_2. The combined organic extracts were dried over $MgSO_4$ and evaporated. The residue was purified by chromatography on silica gel, eluting with EtOAc/hexane (1:9) to afford **3** as a yellow oil (1.34 g, 4.26 mmol, 39%). ^{1}H NMR (400 MHz, $CDCl_3$): δ 6.20 (s, 1H), 5.70 (s, 1H), 3.92 (s, 1H), 2.48 – 2.39 (m, 1H), 2.29 (ddd, J = 16.9, 11.5, 5.8 Hz, 1H), 2.18 (t, J = 5.5 Hz, 1H), 1.85 (s, 2H), 1.61 – 1.52 (m, 3H), 1.43 (s, 5H), 1.32 (s, 5H), 0.96 (s, 2H), 0.89 (t, J = 6.9 Hz, 3H) . ^{13}C NMR (100 MHz, $CDCl_3$):

Δ^8-THC (4). To a solution of compound **3** (0.6 g, 1.9 mmol) in CH_2Cl_2 (35 mL) at 0˚C under N_2 was added BF_3/Et_2O (481 uL, 3.8 mmol), and the mixture was allowed to warm to r.t. After

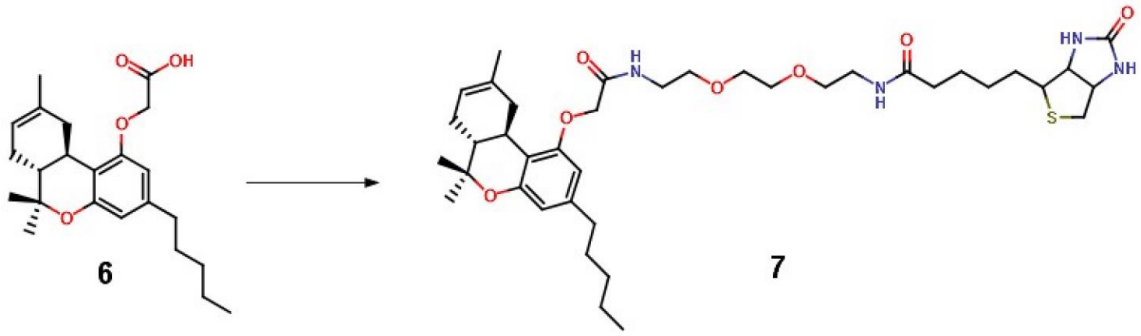
being stirred for 3 h, the mixture was poured into saturated $NaHCO_3$ and extracted two times with CH_2Cl_2. The combined organic extracts were dried over $MgSO_4$ and evaporated. The residue was purified by chromatography on silica gel, eluting with CH_2Cl_2/cyclohexane (5:5), to yield **4** as a yellow oil (280 mg, 0.89 mmol, 47%). ^{1}H NMR (400 MHz, $CDCl_3$): δ 6.28 (s, 1H), 6.08 (d, J = 1.4 Hz, 1H), 5.41 (d, J = 4.1 Hz, 1H), 5.18 (d, J = 4.9 Hz, 1H), 3.21 (dd, J = 16.4, 4.0 Hz, 1H), 2.78 – 2.61 (m, 1H), 2.40 (dd, J = 8.3, 5.8 Hz, 2H), 2.13 (dd, J = 10.9, 4.9 Hz, 1H), 1.93 – 1.73 (m, 3H), 1.68 (s, 3H), 1.61 – 1.45 (m, 2H), 1.39 – 1.34 (m, 3H), 1.33 – 1.19 (m, 4H), 1.09 (s, 3H), 0.87 (t, J = 6.9 Hz, 3H).



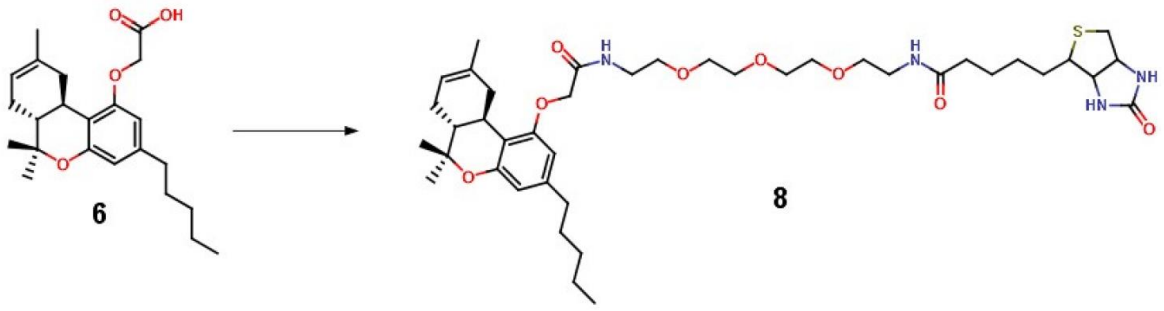
Methyl 2-{[(6aR,10aR)-6,6,9-trimethyl-3-pentyl-6H,6aH,7H,10H,10aH-benzo[c]isochromen-1-yl]oxy}acetate (5). To a solution of compound **4** (280 mg, 0.89 mmol) and potassium carbonate (123 mg, 8.9 mmol) in MEK (10 mL) at r.t. under N_2 was added methyl bromoacetate (163 mg, 1.07 mmol). After being stirred at 80˚C for 18 h, an additional 30 uL of methyl bromoacetate were added since the conversion of the starting material was only approximately 60%. The reaction was stirred at 80˚C for 8 h. 87 mg of cesium carbonate were added and the reaction was stirred at 80˚C overnight. The reaction mixture was poured into saturated NH_4Cl and extracted two times with AcOEt. The combined organic extracts were washed with water, dried over $MgSO_4$ and evaporated to afford **5** as a yellow oil (344 mg, 0.89 mmol, quantitative yield). ^{1}H NMR (400 MHz, $CDCl_3$): δ 6.35 (d, J = 1.1 Hz, 1H), 6.11 (d, J = 1.1 Hz, 1H), 5.42 (d, J = 4.3 Hz, 1H), 4.62 (s, 2H), 3.92 – 3.75 (m, 3H), 3.34 (dd, J = 16.4, 4.5 Hz, 1H), 2.74 (td, J = 11.0, 5.0 Hz, 1H), 2.48 (td, J = 7.3, 2.1 Hz, 2H), 2.23 – 2.07 (m, 1H), 1.92 – 1.75 (m, 3H), 1.71 (s, 3H), 1.64 – 1.49 (m, 2H), 1.37 (s, 3H), 1.34 – 1.23 (m, 2H), 1.10 (s, 3H), 0.88 (t, J = 6.9 Hz, 4H). ^{13}C NMR (100 MHz, $CDCl_3$): δ 169.46, 157.07, 154.66, 142.53, 135.09, 119.07, 112.53, 111.32, 103.78, 65.51, 52.13, 44.95, 36.08, 35.88, 31.75, 31.56, 30.70, 27.94, 27.57, 26.92, 23.59, 22.55, 18.47, 14.03.

2-{[(6aR,10aR)-6,6,9-trimethyl-3-pentyl-6H,6aH,7H,10H,10aH-benzo[c]isochromen-1-yl]oxy}acetic acid (6). A solution of compound **5** (269 mg, 0.73 mmol) and lithium hydroxide monohydrate (135 mg, 3.2mmol) in a mixture of THF (15 mL) and water (250 µL) under N_2 was stirred at 50˚C for 18 h. 250 uL of water were added since the conversion of the starting material was not achieved. The reaction was stirred at 65˚C for 8 h and at r.t. overnight. The reaction mixture was poured into 1N HCl, and extracted two times with AcOEt. The combined organic extracts were washed with water, dried over $MgSO_4$ and evaporated. The residue was purified by

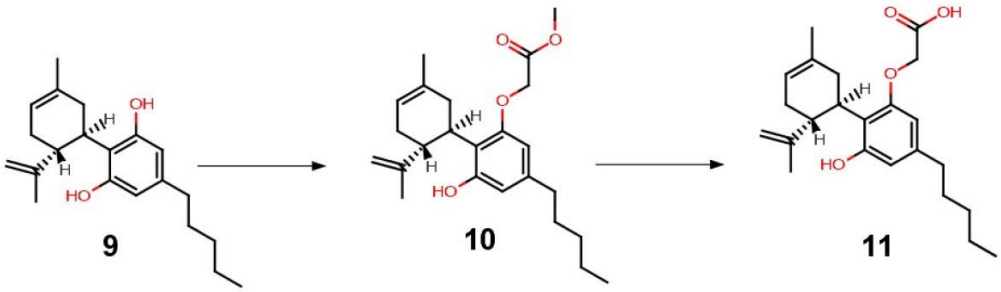
chromatography on silica gel, eluting with a gradient of CH_2Cl_2/MeOH (2% to 20% of MeOH), to yield **6** as a yellow oil (185 mg, 0.5 mmol, 68%). ^{1}H NMR (400 MHz, $CDCl_3$): δ 6.37 (d, J = 1.1 Hz, 1H), 6.14 (d, J = 1.1 Hz, 1H), 5.42 (d, J = 3.6 Hz, 1H), 4.67 (s, 2H), 3.28 (d, J = 16.6 Hz, 1H), 2.74 (td, J = 11.0, 4.9 Hz, 1H), 2.57 – 2.42 (m, 2H), 2.22 – 2.08 (m, 1H), 1.89 – 1.76 (m, 3H), 1.69 (s, 3H), 1.63 – 1.52 (m, 2H), 1.38 (s, 3H), 1.35 – 1.27 (m, 4H), 1.09 (s, 3H), 0.88 (t, J = 6.9 Hz, 3H). ^{13}C NMR (101 MHz, $CDCl_3$) δ 156.78, 154.76, 142.67, 135.02, 119.15, 112.58, 111.64, 103.88, 77.35, 77.24, 77.04, 76.72, 65.04, 44.95, 36.15, 35.89, 31.76, 31.58, 30.72, 27.95, 27.57, 23.55, 22.56, 18.48, 14.04.



N-(2-{2-[2-(2-{[(6aR,10aR)-6,6,9-trimethyl-3-pentyl-6H,6aH,7H,10H,10aH-benzo[c]isochromen-1-yl]oxy}acetamido)ethoxy]ethoxy}ethyl)-5-{2-oxo-hexahydro-1H-thieno[3,4-d]imidazolidin-4-yl}pentanamide (7). A mixture of O-(7-Azabenzotriazol-1-yl)-N,N,N',N'-tetramethyluronium hexafluorophosphate HATU (53 mg, 0.14 mmol), DIPEA (28 mg, 0.215 mmol, 0.31 mL) and compound **6** (40 mg, 0.107 mmol) in DCM (2 mL) and DMF (2 mL) was stirred under N_2 at r.t. for 1 h. Biotin-$(PEO)_4$ amine (50 mg, 0.118 mmol) diluted in a mixture of DCM (2 mL) and DMF (2 mL) was added to the reaction mixture. The reaction mixture was stirred at r.t. for 46 h. The reaction medium was concentrated. Column chromatography (silica gel, gradient MeOH/DCM 2 to 30%) afforded the title compound as a white solid (12 mg, 16%). ^{1}H NMR (400 MHz, $CDCl_3$) δ 6.86 (br s, 1H), 6.41 (t, J = 5.4 Hz, 1H), 6.37 (s, 1H), 6.25 (br s 1H), 6.19 (s, 1H), 5.44 (d, J = 4.1 Hz, 1H), 5.34 (s, 1H), 4.58 – 4.44 (m, 3H), 4.34 – 4.27 (m, 1H), 3.64 – 3.49 (m, 10H), 3.45 – 3.37 (m, 2H), 3.18 – 3.10 (m, 1H), 3.07 – 2.98 (m, 1H), 2.90 (dd, J = 12.8, 4.90 Hz, 1H), 2.77 – 2.66 (m, 2H), 2.47 (dd, J = 8.7, 6.4 Hz, 2H), 2.26 – 2.12 (m, 3H), 1.92 – 1.60 (m, 10H), 1.62 – 1.51 (m, 2H), 1.44 (dd, J = 15.4, 7.8 Hz, 2H), 1.38 (s, 3H), 1.36 – 1.24 (m, 4H), 1.10 (s, 3H), 0.88 (t, J = 6.84 Hz, 3H). ^{13}C NMR (100 MHz, $CDCl_3$): δ 14.03, 18.44, 22.54, 23.64, 25.53, 27.53, 27.94, 28.09, 28.15, 30.74, 31.55, 31.83, 35.85, 35.92, 36.70, 38.84, 39.12, 40.51, 45.09, 53.43, 55.46, 60.17, 61.81, 67.89, 69.68, 69.97, 70.17, 70.20, 104.59, 111.75, 112.17, 119.70, 134.36, 143.03, 154.68, 156.63, 163.70, 168.92, 173.19, 192.65. HRMS (positive mode) for $C39H61N407S^+$ $[M+1]^+$ calc 729.4261, found 729.4268.



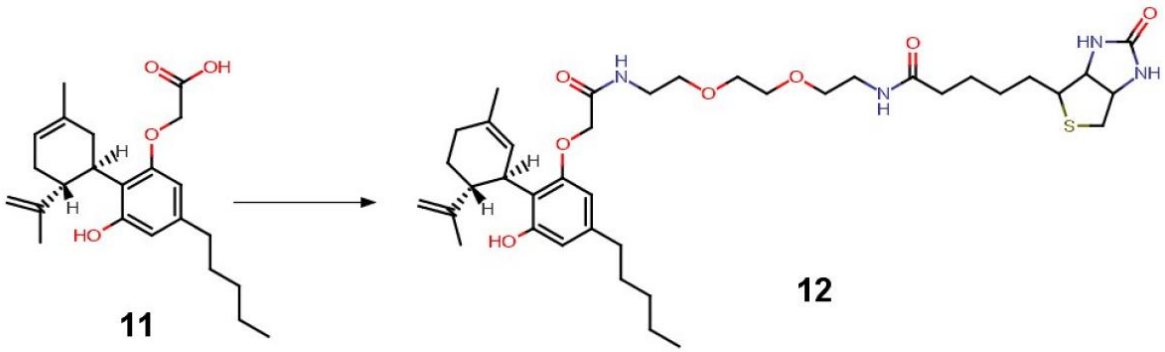
N-[2-(2-{2-[2-(2-{[(6aR,10aR)-6,6,9-trimethyl-3-pentyl-6H,6aH,7H,10H,10aH-benzo[c]isochromen-1-yl]oxy}acetamido)ethoxy]ethoxy}ethoxy)ethyl]-5-{2-oxo-hexahydro-1H-thieno[3,4-d]imidazolidin-4-yl}pentanamide (8). A mixture of O-(7-Azabenzotriazol-1-yl)-N,N,N',N'-tetramethyluronium hexafluorophosphate HATU (53 mg, 0.14 mmol), DIPEA (28 mg, 0.215 mmol, 0.31 mL) and compound **6** (40 mg, 0.107 mmol) in DCM (4 mL) and DMF (2 mL) was stirred under N_2 at r.t. for 1 h. Biotin-$(PEO)_4$ amine (50 mg, 0.118 mmol) was added. The reaction mixture was stirred at r.t. for 46 h. The reaction medium was concentrated. Column chromatography (silica gel, gradient MeOH/DCM 2 to 30 %) afforded the title compound as a white solid (12 mg, 16%). ^{1}H NMR (400 MHz, $CDCl_3$) δ 6.89 (br s, 1H), 6.51 (t, J = 5.4 Hz, 1H), 6.37 (d, J = 1.3 Hz, 1H), 6.19 (d, J = 1.3 Hz, 1H), 6.09 (s, 1H), 5.44 (d, J = 4.4 Hz, 1H), 5.16 (s, 1H), 4.65 – 4.41 (m, 3H), 4.42 – 4.24 (m, 1H), 3.69 – 3.49 (m, 13H), 3.42 (dd, J = 9.9, 5.1 Hz, 2H), 3.15 (dt, J = 11.9, 6.0 Hz, 1H), 3.09 – 2.95 (m, 1H), 2.90 (dd, J = 12.8, 5.0 Hz, 1H), 2.71 (dd, J = 17.8, 8.5 Hz, 2H), 2.47 (dd, J = 8.7, 6.4 Hz, 2H), 2.34 – 2.06 (m, 3H), 1.96 – 1.60 (m, 11H), 1.60 – 1.49 (m, 2H), 1.45 (dd, J = 15.4, 7.8 Hz, 2H), 1.38 (s, 2H), 1.34 – 1.25 (m, 4H), 1.10 (s, 3H), 0.89 (t, J = 6.8 Hz, 3H). ^{13}C NMR (101 MHz, CDCl 3) δ 173.10, 168.84, 163.48, 156.66, 154.66, 143.01, 134.44, 119.64, 112.18, 111.71, 104.60, 70.48, 70.44, 70.32, 70.09, 69.95, 69.71, 67.90, 61.76, 60.13, 55.37, 45.09, 40.52, 39.15, 38.85, 36.69, 35.85, 31.82, 31.55, 30.75, 28.09, 27.95, 27.54, 25.49, 23.64, 22.54, 18.45, 14.04. HRMS (positive mode) for $C41H65N408S^+$ $[M+1]^+$ calc 773.4523 found 773.4554.



Methyl 2-{3-hydroxy-2-[(1R,6R)-3-methyl-6-(prop-1-en-2-yl)cyclohex-3-en-1-yl]-5-pentylphen oxy}acetate (10). A solution of CBD, compound **9** (50 mg, 0.16 mmol) and cesium carbonate (52 mg, 0.16 mmol) in MEK (2 mL) under N_2 was heated at 80˚C. After being stirred at 80˚C for 1 h, methyl bromoacetate (15 µL, 0.16 mmol) was added. The reaction was stirred at

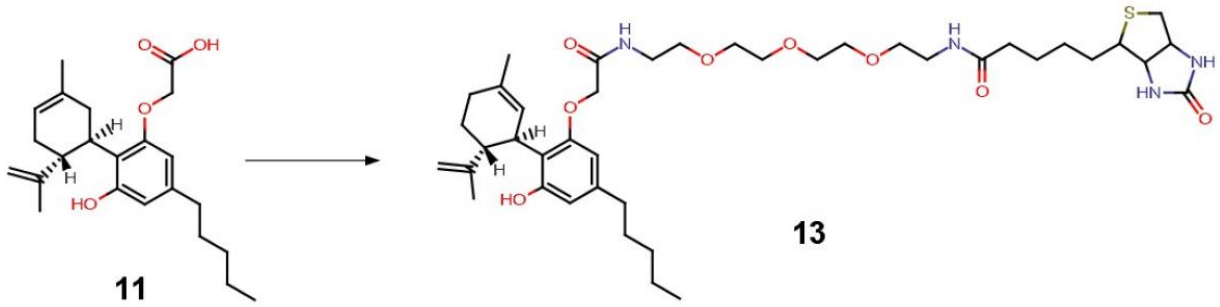
80˚C for 8 h. The reaction mixture was concentrated. Two batches of the reaction were combined and were purified by chromatography on silica gel, eluting with a gradient of cyclohexane/ethyl acetate (95/5) to (65/35). A mixture of CBD and the expected product (50/50, according to HPLC quantification) was obtained (140 mg overall).

2-{3-hydroxy-2-[(1R,6R)-3-methyl-6-(prop-1-en-2-yl)cyclohex-3-en-1-yl]-5-pentylphenoxy} acetic acid (11). A solution of mixture **10** previously obtained (140 mg) and lithium hydroxide monohydrate (65 mg, 1.5 mmol) in a mixture of THF (5 mL) and water (125 µL) under N_2 was stirred at r.t. overnight. The residue was purified by chromatography on silica gel, eluting with a gradient of CH_2Cl_2/MeOH (100/0 to 90/10), to yield **11** as a yellow oil (45 mg, 25% over the two steps). ^{1}H NMR (400 MHz, $CDCl_3$): δ ppm 6.38 (s, 1 H), 6.13 (br. s., 2 H), 5.57 (br. s., 1 H), 4.62 - 4.48 (m, 3 H), 4.36 (br. s., 1 H), 4.08 - 3.98 (m, 1 H), 2.53 - 2.33 (m, 3 H), 2.30 - 2.16 (m, 1 H), 2.14 - 2.06 (m, 1 H), 1.85 - 1.72 (m, 5 H), 1.64 (s, 3 H), 1.56 (dt, J=15.00, 7.56 Hz, 2 H), 1.37 - 1.22 (m, 4 H), 0.88 (t, J=6.96 Hz, 3 H).



N-(2-{2-[2-(2-{3-hydroxy-2-[(1R,6R)-3-methyl-6-(prop-1-en-2-yl)cyclohex-2-en-1-yl]-5-pentylphenoxy}acetamido)ethoxy]ethoxy}ethyl)-5-{2-oxo-hexahydro-1H-thieno[3,4-d]imidazolidin-4-yl}pentanamide (12). A mixture of O-(7-Azabenzotriazol-1-yl)-N,N,N’,N’-tetramethyluronium hexafluorophosphate HATU (32 mg, 0.084 mmol), DIPEA (20 µL, 0.128 mmol) and compound **11** (24 mg, 0.064 mmol) in DCM (1 mL) and DMF (1 mL) was stirred under N_2 at r.t. for 1h. Biotin-$(PEO)_3$ amine (13 mg, 0.035 mmol) diluted in DMF (1 mL) was added to the reaction mixture. The reaction mixture was stirred at r.t. for 46 h and concentrated. Column chromatography (silica gel, gradient MeOH/DCM 0 to 15 %) afforded the title compound as a white solid (9 mg, 35%). ^{1}H NMR (400 MHz, CDCl 3) δ 6.83 (br s, 1H), 6.45 - 6.34 (m, 1H), 6.25 (m, 1H), 6.13 (br s, 1H), 6.05 (br s, 1H), 5.57 (br s, 1H), 5.31 (br s, 1H), 4.64 (br s, 1H), 4.50 - 4.56 (m, 1 H), 4.50 – 4.37 (m, 3H), 4.36 – 4.29 (m, 1H), 3.98 – 3.89 (m, 1H), 3.64 – 3.46 (m, 10H), 3.46 – 3.37 (m, 2H), 3.19 - 3.12 (m, 1 H), 2.92 (dd, J=12.92, 4.89 Hz, 1 H), 2.73 (d, J= 12.80 Hz, 2H), 2.56 – 2.35 (m, 3H), 2.22 (t, J=7.22 Hz, 3 H), 2.13 – 2.03 (m, 1H), 1.87 – 1.60 (m, 11H), 1.54 (dt, J=14.87, 7.50 Hz, 3 H), 1.49 – 1.41 (m, 2H), 1.35 – 1.21 (m, 4H), 0.87 (t, J= 6.90 Hz, 3H). ^{13}C NMR (101 MHz, $CDCl_3$) δ ppm 14.06, 20.00, 22.54, 25.47, 28.02, 29.87, 30.84, 31.48, 35.89, 38.84, 39.18, 40.40, 40.51, 53.58, 55.35, 60.28, 61.92, 69.89, 70.17,

104.13, 110.85, 111.13, 120.30, 123.55, 129.39, 163.51, 169.35, 173.40. HRMS (positive mode) for $C_{39}H_{61}N_4O_7S^+$ $[M+1]^+$ calc 729.4261, found 729.4260.



N-[2-(2-{2-[2-(2-{3-hydroxy-2-[(1R,6R)-3-methyl-6-(prop-1-en-2-yl)cyclohex-2-en-1-yl]-5-pentylphenoxy}acetamido)ethoxy]ethoxy}ethoxy)ethyl]-5-{2-oxo-hexahydro-1H-thieno[3,4-d]imidazolidin-4-yl}pentanamide (13). A mixture of O-(7-Azabenzotriazol-1-yl)-N,N,N',N'-tetramethyluronium hexafluorophosphate HATU (30 mg, 0.077 mmol), DIPEA (20 µL, 0.118 mmol) and compound **11** (22 mg, 0.059 mmol) in DCM (1 mL) and DMF (1 mL) was stirred under N_2 at r.t. for 1 h. Biotin-$(PEO)_4$ amine (27 mg, 0.065 mmol) diluted in DCM (1 mL) was added. The reaction mixture was stirred at r.t. for 46 h. The reaction medium was concentrated. Column chromatography (silica gel, gradient MeOH/DCM 2 to 15 %) afforded the title compound as a white solid (20 mg, 45%). ^{1}H NMR (400 MHz, $CDCl_3$) δ 6.83 (br s, 1H), 6.66 - 6.47 (m, 1H), 6.37 (m, 1H), 6.31 – 6.17(m, 1H), 6.12 (br s, 1H), 5.58 (br s, 1H), 5.35 (br s, 1H), 4.63 (br s, 1H), 4.56 - 4.46 (m, 2 H), 4.46 – 4.38 (m, 2H), 4.38 – 4.26 (m, 1H), 3.98 – 3.89 (m, 1H), 3.66 – 3.46 (m, 14H), 3.46 – 3.37 (m, 2H), 3.19 - 3.11(m, 1 H), 2.91 (dd, J=12.92, 4.89 Hz, 1 H), 2.74 (d, J= 12.80 Hz, 2H), 2.56 – 2.34 (m, 3H), 2.15 - 2.31 (m, 3 H), 2.13 – 2.01 (m, 1H), 1.86 – 1.60 (m, 11H), 1.54 (dt, J=14.96, 7.51 Hz, 3 H), 1.48 – 1.39 (m, 2H), 1.35 – 1.21 (m, 4H), 0.88 (t, J= 6.96 Hz, 3H). ^{13}C NMR (101 MHz, $CDCl_3$) δ ppm 14.05, 19.92, 22.52, 23.73, 25.47, 27.56, 28.04, 28.06, 29.95, 30.83, 31.47, 35.83, 35.87, 36.40, 38.83, 39.18, 46.03, 53.43, 55.37, 60.23, 61.83, 69.90, 69.98, 70.06, 70.31, 70.36, 70.39, 104.15, 10.92, 110.94, 110.97, 111.04, 111.08, 120.22, 123.76, 123.77, 123.84, 129.16, 163.77 169.17, 173.32. HRMS (positive mode) for $C_{41}H_{65}N_4O_8S^+$ $[M+1]^+$ calc 773.4523, found 773.4562.

2. Nanobody library construction and validation

Library construction: The DNA library was designed with a universal nanobody scaffold and randomized CDR sequences similarly to a reported method.[2] The combinatorial DNA library was chemically synthesized by a trinucleotide mutagenesis technology (Thermo Fisher Scientific). To subclone the library into a phagemid vector, ~300 ng of the synthetic DNA library was used as templates to set up 30×50 µL PCR reactions, each with 1 µL of Platinum SuperFi DNA polymerase (Thermo Fisher Scientific) and 0.5 µM (each) primers (Table S3). The PCR protocol included an initial denaturation at 98˚C for 30 s followed by 20 cycles of 98˚C for 10 s, 70˚C for 10 s, and 72˚C for 10 s, and a final extension at 72˚C for 5 min. ~40 µg purified PCR products and ~60 µg of a pADL-23c phagemid vector (Antibody Design Labs) were digested for 1 h at 37˚C with *Bgl*I (Thermo Fisher Scientific) and purified before the ligation. ~8.5 µg digested vector and ~3.9 µg inserts were ligated with 260 units of T4 DNA ligase (Thermo Fisher Scientific) in a 2,600 µL reaction at 4˚C for overnight. Ligated products were purified by a Purelink PCR quick purification kit (Invitrogen) to obtain a final volume of 50 µL (~278.9 ng/µL) and then transformed to *E. coli* electrocompetent TG1 cells (Lucigen) in 30 electroporation cuvettes, each containing 1.4 µL ligated products and 50 µL cells, by following the manufacturer's instruction (1,800 V, 10 µF, 600 Ω). Electroporated cells were resuspended with a warm recovery medium (Lucigen) and incubated by shaking (250 rpm) at 37˚C for 1 h and then mixed with ~470 mL 2 × YT, 2% glucose. Cells were plated on 500 2×YT-ampicillin-glucose agar plates (140 mm), grown overnight at 30˚C, and scraped with 2×TY. The library was aliquoted and stored in 25% glycerol at -80˚C.

Next generation sequencing library validation: To sequence the phage-displayed nanobody library, phagemid DNAs were isolated from purified phage particles using a QIAprep Spin M13 kit (Qiagen). DNA concentration was measured by a Nanodrop 2000 spectrophotometer (Thermo Fisher Scientific). A two-step low-cycle-number PCR was performed to introduce the Illumina adaptors and 8-bp unique molecular identifiers (UMIs) to the 3' and 5' ends of amplicons with specific primers (Table S3). The library was sequenced by an Illumina NextSeq platform using a 2×150 bp high-output kit. The sequencing data were first processed to trim UMI sequences using Flexbar.[3] Low-quality reads were filtered with a minimum quality score of 20. To accommodate errors within UMIs while retrieving CDR sequences, undetermined bases in UMIs or shorter barcodes were allowed. Clean reads were then aligned to CDR-adjacent scaffold sequences to extract designed CDR sequences at a 0.1 mismatch rate. Three CDR sequences were obtained by merging paired-end reads. To retain correct sequences, CDR sequences were required to be in the same reading frame without any undetermined base or stop codon. A custom Perl script, shown below, was used to analyze translated protein sequences. CDR protein sequence logos were generated with WebLogo3.6.[4]

```
#!/usr/bin/perl -w
use strict;
my $fastqL = $ARGV[0];
my $fastqR = $ARGV[1];
```

```
my %processedReads;
my %phagereads;
open(FQL,"$fastqL") || die;
while(my $ln=<FQL>){
	if($.%4!=1){next;}
	chomp $ln;
	my @code = split(/\_/,$ln);
	my $coord = $code[0];
	my $tagcodeL = $code[1];
	my $cdr1_rna = $code[2];
	my $cdr2_rna = $code[3];
	#remove the peptides containing undertermined nul
	if($cdr1_rna=~/N/ or $cdr2_rna=~/N/){
		next;
	}
	#filter for fixed peptide length
	if(length($cdr1_rna)%3==0 and length($cdr2_rna)%3==0){
		$phagereads{$coord}=$tagcodeL.'|'.$cdr1_rna.'|'.$cdr2_rna;
	}
}
close(FQL);

open(FQR,"$fastqR") || die;
my $headertag;
while(my $ln=<FQR>){
	chomp $ln;
	if($.%4==1){
		$headertag = $ln;
		next;
	}
	if($.%4==2){
		my @code = split(/\_/,$headertag);
		my $coord = $code[0];
		$coord=~s/ 2\:N\:0\:/ 1\:N\:0\:/;
		my $tagcodeR = $code[1];
		my $cdr3_rna = $ln;
		#remove the peptides containing undertermined nul
		if($cdr3_rna=~/N/){
			next;
		}
		#filter for frame shift
		if(length($cdr3_rna)%3==0 and exists $phagereads{$coord}){

	$processedReads{$coord}=$phagereads{$coord}.'|'.$cdr3_rna.'|'.$tagcodeR;
		}
	}
}
close(FQR);

open(PRO,">$ARGV[2]") || die;
my %countPeptide;
foreach my $key(keys %processedReads){
	my @data = split(/\|/,$processedReads{$key});
	my $c1 = translate($data[1]);
	my $c2 = translate($data[2]);
	my $c3 = translate($data[3]);
	print PRO
$key,"\t",$data[0],"\t",$data[4],"\t",$data[1],"\t",$data[2],"\t",$data[3],"\t",$c1,"\
t",$c2,"\t",$c3,"\n";
	$countPeptide{$data[0].'_'.$data[4]}{$c1.'_'.$c2.'_'.$c3}++;
}
close(PRO);
```

```
foreach my $key(keys %countPeptide){
       foreach my $sky(keys %{$countPeptide{$key}}){
              print '>',$key,'-',$countPeptide{$key}{$sky},"\n";
              print $sky,"\n";
       }
}

sub reverse_complement {
        my $dna = shift;

        # reverse the DNA sequence
        my $revcomp = reverse($dna);

        # complement the reversed DNA sequence
        $revcomp =~ tr/ATCGN/TAGCN/;
        return $revcomp;
}
sub translate{
       my(%genetic_code) = (
         'TCA' => 'S', # Serine
         'TCC' => 'S', # Serine
         'TCG' => 'S', # Serine
         'TCT' => 'S', # Serine
         'TTC' => 'F', # Phenylalanine
         'TTT' => 'F', # Phenylalanine
         'TTA' => 'L', # Leucine
         'TTG' => 'L', # Leucine
         'TAC' => 'Y', # Tyrosine
         'TAT' => 'Y', # Tyrosine
         'TAA' => '4', # Stop
         'TAG' => '4', # Stop
         'TGC' => 'C', # Cysteine
         'TGT' => 'C', # Cysteine
         'TGA' => '4', # Stop
         'TGG' => 'W', # Tryptophan
         'CTA' => 'L', # Leucine
         'CTC' => 'L', # Leucine
         'CTG' => 'L', # Leucine
         'CTT' => 'L', # Leucine
         'CCA' => 'P', # Proline
         'CAT' => 'H', # Histidine
         'CAA' => 'Q', # Glutamine
         'CAG' => 'Q', # Glutamine
         'CGA' => 'R', # Arginine
         'CGC' => 'R', # Arginine
         'CGG' => 'R', # Arginine
         'CGT' => 'R', # Arginine
         'ATA' => 'I', # Isoleucine
         'ATC' => 'I', # Isoleucine
         'ATT' => 'I', # Isoleucine
         'ATG' => 'M', # Methionine
         'ACA' => 'T', # Threonine
         'ACC' => 'T', # Threonine
         'ACG' => 'T', # Threonine
         'ACT' => 'T', # Threonine
         'AAC' => 'N', # Asparagine
         'AAT' => 'N', # Asparagine
         'AAA' => 'K', # Lysine
         'AAG' => 'K', # Lysine
         'AGC' => 'S', # Serine
         'AGT' => 'S', # Serine
         'AGA' => 'R', # Arginine
         'AGG' => 'R', # Arginine
```

```
        'CCC' => 'P', # Proline
        'CCG' => 'P', # Proline
        'CCT' => 'P', # Proline
        'CAC' => 'H', # Histidine
        'GTA' => 'V', # Valine
        'GTC' => 'V', # Valine
        'GTG' => 'V', # Valine
        'GTT' => 'V', # Valine
        'GCA' => 'A', # Alanine
        'GCC' => 'A', # Alanine
        'GCG' => 'A', # Alanine
        'GCT' => 'A', # Alanine
        'GAC' => 'D', # Aspartic Acid
        'GAT' => 'D', # Aspartic Acid
        'GAA' => 'E', # Glutamic Acid
        'GAG' => 'E', # Glutamic Acid
        'GGA' => 'G', # Glycine
        'GGC' => 'G', # Glycine
        'GGG' => 'G', # Glycine
        'GGT' => 'G'  # Glycine
    );
    my $rna = shift;
    #if(length($rna)%3!=0){
    #      return(0);
    #}
    my $protein = "";
    for(my $i=0;$i<length($rna)-2;$i+=3){
           my $codon = substr($rna,$i,3);
           if(not exists $genetic_code{$codon}){
                  $protein .= 6;
           } else {
                  $protein .= $genetic_code{$codon};
           }
    }
    return ($protein);
}
```

3. Phage display selection

To prepare a phage library, the phagemid-containing TG1 bacterial stock from the "Library construction" was diluted with 2×YT media with 2% glucose and 100 μg/mL ampicillin to OD600 of ~0.1 and cultured at 37˚C to OD600 of ~0.4 to 0.5. Cells were superinfected by adding a helper phage CM13 at 5×10^9 pfu/mL for 1 h, pelleted to remove the glucose, resuspended with fresh 2×YT media containing 100 μg/mL ampicillin and 50 μg/mL kanamycin, and incubated at 25˚C for overnight. To purify phage particles, cells were removed by centrifugation at 5,000×g, 4˚C for 30 min and phage particles in the supernatant were precipitated with polyethylene glycol (PEG). The supernatant was added with 1/5 volume PEG/NaCl solution (20% (w/v) polyethylene glycol 6,000 and 2.5 M NaCl), placed on ice for 1 h, and centrifuged at 12,000×g, 4˚C for 30 min. Phage pellets were resuspended in 1×PBS and stored at 4˚C for short-term use or with 25% glycerol at -80˚C for long-term storage.

Anchor binder selection was performed using biotin- and biotinylated-CBD-bound streptavidin magnetic beads for negative and positive selections, respectively, in each selection round. Briefly, 300 μL 10 μM biotin or biotinylated CBD were captured by 300 μL streptavidin-coated

magnetic beads (Dynabeads M-280 Streptavidin, Thermo Fisher Scientific) and blocked with 1% casein and 1% BSA in 1×PBS pH 7.4 at r.t. for 1 h. In each round, the phage-displayed nanobody library was incubated with the biotin-bound beads for 1 h at r.t. to remove off-target binders and the supernatant was incubated with biotinylated CBD-bound beads for 1 h. Beads were washed with 10× 0.05 % PBST (1×PBS with 0.05% v/v Tween 20) and phage particles were competitively eluted with 50, 10 and 1 μM CBD in the first, second, and third selection rounds, respectively, and 100 nM CBD in the fourth to sixth rounds. After six rounds of biopanning, single colonies were picked and validated by phage ELISA followed by DNA sequencing.

Dimerization binders were selected using CBD-free and bound CA-14 for the negative and positive selections, respectively. Briefly, to prepare the bait for each selection round, 600 μL 1 μM biotinylated CA-14 were captured by 600 μL streptavidin beads and blocked with 1% casein and 1% BSA in 1×PBS pH 7.4 for 1 h at r.t. The beads were divided by a 2:1 ratio for the negative and positive selections, respectively. For the positive selection, the CA-14 bound beads were incubated with 50, 10, 1, and 1 μM CBD to generate the CBD–CA-14 complex as the bait from the rounds 1 to 4, respectively. Biopanning was performed as described above except that bound phages were eluted with 100 mM triethylamine. Four rounds of biopanning were performed before single colony picking and validation.

4. Phage ELISA

To prepare single-phage ELISA, individual colonies were randomly picked, inoculated into 250 μL media (2×TY, 2% glucose, 100 μg/mL ampicillin) per well in sterile deep-well plates (Thermo Fisher Scientific), and grown at 37˚C for overnight. 10 μL cultures were inoculated into 500 μL fresh media and cells were grown to OD600 of ~0.5 and infected by CM13 helper phage with the multiplicity of infection (MOI) of ~18. The cultures were incubated at 37˚C for 1 h, added with kanamycin of 50 μg/mL final concentration, and grown at 25˚C for overnight. Plates were centrifuged for 10 min at 3,000×g and phage-containing supernatants were transferred to fresh plates for the ELISA assays.

For anchor-binder phage ELISA, ELISA plates (Nunc MaxiSorp, Thermo Fisher Scientific) were coated with 100 μL 5 μg/mL streptavidin in coating buffer (100 mM carbonate buffer, pH 8.6) at 4˚C for overnight. After washing with 3× 0.05 % PBST (1×PBS with 0.05% v/v Tween 20), each well was added with 100 μL 1 μM biotinylated CBD and incubated at r.t. for 1 h. Wells were washed with 5× 0.05 % PBST, blocked with 1% casein in 1×PBS, and then added with 100 μL phage supernatants. Phage particles were incubated with CBD in plates at r.t. for 1 h. Wells were washed with 10× 0.05% PBST, added with 100 μL HRP-M13 major coat protein antibody (RL-ph1, Santa Cruz Biotechnology; 1:10,000 dilution with 1×PBS with 1% casein), and incubated at r.t. for 1 h. A colorimetric detection was performed using a 1-Step Ultra TMB ELISA substrate

solution (Thermo Fisher Scientific) and OD450 was measured with a SpectraMax Plus 384 microplate reader (Molecular Devices).

The dimerization-binder phage ELISA used the plate-immobilized anchor binder as the bait. ~100 nM biotinylated anchor binder, CA-14, was bound to a streptavidin-coated plate prepared as above described. The immobilized CA-14 was bound to different concentrations of CBD in 1×PBS, pH 7.4, to form the anchor binder–CBD complex before binding to phage-displayed dimerization binders. Other steps were performed similarly to the anchor-binder phage ELISA.

5. Protein expression and purification

All nanobodies were expressed as a C-terminal AviTagged and His-tagged form in *E. coli* and purified by Ni-affinity and size-exclusion chromatography. In brief, *E. coli* strain WK6 was transformed with the expression constructs and grown in TB medium at 37˚C to an OD600 of ~0.7 before induction with 1 mM isopropyl-β-D-galactopyranoside (IPTG) at 28˚C for overnight. Harvested cell pellets from 1-liter cultures were resuspended in 15 mL ice-cold TES buffer (0.2 M Tris-HCL pH 8.0, 0.5 mM EDTA, 0.5 M sucrose) and incubated with gently shaking on ice for 1 h. To release proteins from the periplasm by osmotic shock, the resuspended pellets were added with 30 mL of TES/4 buffer (1:4 dilution of the TES buffer in ddH_2O) and gently shaken on ice for 45 min. Cell debris was removed by centrifugation at 15,000×g, 4˚C for 30 mins. The supernatant was loaded onto a 5 mL HisTrap column (GE Healthcare) pre-equilibrated with the lysis buffer (50 mM sodium phosphate, pH 8.0, 300 mM NaCl, 10 mM imidazole, 10% glycerol). The column was washed with a washing buffer (50 mM sodium phosphate, pH 8.0, 300 mM NaCl, 20 mM imidazole, 10% glycerol) and then His-tagged proteins were eluted with an elution buffer (50 mM sodium phosphate, pH 8.0, 300 mM NaCl, 250 mM imidazole, 10% glycerol). Eluates were concentrated with an Amicon Ultra-15 centrifugal filter unit (3 kDa cutoff, Millipore). Concentrated proteins were loaded onto a HiLoad 16/600 Superdex 200 pg column (GE Healthcare) pre-equilibrated with a storage buffer (1×PBS, 5% glycerol). Eluted proteins were concentrated, examined by SDS-PAGE, and quantified by a Bradford assay (BioRad), then flash frozen in 100 µL aliquots by liquid N_2 and stored at -80˚C.

6. Protein biotinylation

Nanobodies bearing AviTag were biotinylated by BirA using a BirA-500 kit (Avidity).[5] Typically, 90 µL BiomixA (10× concentration: 0.5 M bicine buffer, pH 8.3), 90 µL BiomixB (10× concentration: 100 mM ATP, 100 mM $Mg(OAc)_2$, 500 µM d-biotin), 4 µL 1 mg/mL BirA, and 216 µL ddH_2O were added to 500 µL ~1 mg/mL AviTagged nanobodies to a final volume of 900 µL. The biotinylation mixture was incubated at r.t. for 1 h and then loaded onto a HiPrep 26/10 desalting column (GE Healthcare) pre-equilibrated with a storage buffer (1×PBS, 5% glycerol). Eluted proteins were concentrated, examined by SDS-PAGE, and quantified by the Bradford assay, flash frozen in 100 µL aliquots by liquid N_2, and stored at -80˚C.

7. Bio-layer interferometry

The nanobody binding kinetics was analyzed using an Octet RED96 system (ForteBio) and Streptavidin (SA) or Super Streptavidin (SSA) biosensors. For the anchor-binder analysis, 200 nM biotinylated anchor binders were immobilized on SSA biosensors with a binding assay buffer (1×PBS, pH 7.4, 0.05% Tween 20, 0.2% BSA, 3% methanol; Note: because CBD and THC were dissolved in methanol to make stock solutions, 3% methanol was added to the buffer to fairly compare samples and controls). Serial dilutions of CBD and THC were used for the anchor binder assays. Dissociation constants (K_Ds) of anchor binder-ligand interactions were calculated based on a steady-state analysis using an Octet data analysis software 9.0. For the dimerization-binder analysis, 200 nM biotinylated dimerization binders were immobilized on SA biosensors with the binding assay buffer and then assayed with 1 μM CA-14 pre-equilibrated with serial dilutions of CBD or THC. K_Ds, k_{on}s, and k_{off}s of the interactions between anchor binder–ligand complexes and dimerization binders were calculated based on a heterogeneous ligand model global fit (2:1) of the data and simulated binding equilibria of the anchor binder and CBD described in Supplementary Note 1.

8. Analytical SEC

Anchor and dimerization nanobodies were analyzed by SEC with a Superdex 75 increase 10/300 GL column (GE Healthcare). For non-crosslinked samples, 500 μL ~30 μM nanobodies were injected into the column equilibrated with 1 × PBS and eluted at a flow rate of 0.75 mL/min at 4˚C. The column was calibrated with molecular weight standards (Sigma-Aldrich). Crosslinked samples were prepared by incubating 5 or 10 μM proteins in the presence or absence of CBD in 1×PBS with 0.1 mM $BS(PEG)_5$ at r.t. for 30 mins. 500 μL crosslinked samples were injected to the column.

9. Sandwich ELISA-like assay to determine CBD detection sensitivity

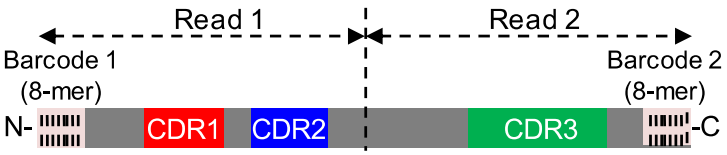
Urine and saliva samples were collected from three healthy volunteers (two males and one female). The urine and saliva samples were spun for 3 minutes at 14,000×g and the supernatants were diluted by 1/5 with 1×PBS, pH 7.4, and spiked with CBD. Serial dilutions (0 to 1,000 nM CBD) of samples were used for the sandwich ELISA-like assay. In each well, 100 μL biotinylated CA-14 (~100 nM) was immobilized to the plate and the bound CBD was detected with 100 μL phage-displayed DB-21 (~10^{10} phage particles; purified similarly as His-tagged nanobodies to remove empty phage particles) and 100 μL HRP-M13 antibody (1:10,000 dilution with 1×PBS with 1% casein). Each dilution was repeated 8 times and Limit of Detection (LoD) was calculated by $mean_{blank} + 3\times SD_{blank}$. Note: by comparing Ni-affinity purified phage particles and the DB-21 protein, we found that the phage-displayed DB-21 can be directly used as a relatively sensitive and stable affinity reagent. We typically purified ~10^{14} phage particles from 50 mL culture, which can be used for ~1,000 96-well microplate assays.

10. Split luciferase assay

A split luciferase assay with a NanoLuc complementation reporter was performed to analyze the CBD-dependent nanobody dimerization. A LgBIT (a ~18 kDa large subunit of an engineered split nanoluciferase, NanoBiT) and a SmBIT(~1.3 kDa small subunit of NanoBiT) tags were subcloned from pBiT1.1-N and pBiT2.1-C (Promega), fused to CA-14 and DB-21 C-termini, respectively, via a 15-amino-acid linker (GSSGGGGSGGGGSSG) to avoid steric hindrance. Protein fusion genes were cloned into pADL-23c using *Bgl*I restriction sites. Proteins were produced by periplasmic expression in *E. coli* WK6 with 0.1 mM IPTG induction at 18˚C for overnight and released by osmotic shock as described above. For luminescence measurement, 25 µL CA-14 supernatants were pre-equilibrated with serial dilutions of CBD, incubated with 25 µL DB-21 supernatants at r.t. for 30 mins, and then added with 10 µL/well furimazine (20× dilution, Promega). After 10 min incubation, luminescence intensities (OD450) were measured with a SpectraMax i3x (Molecular Devices).

SUPPLEMENTARY FIGURES

a



b

Unique clone count	Percentage (%)
1	95.5471
2	3.9760
3	0.4193
4	0.0434
5	0.0069
6	0.0023
7	0.0015
8	0.0008
9	0.0005
≥10	0.0004

c

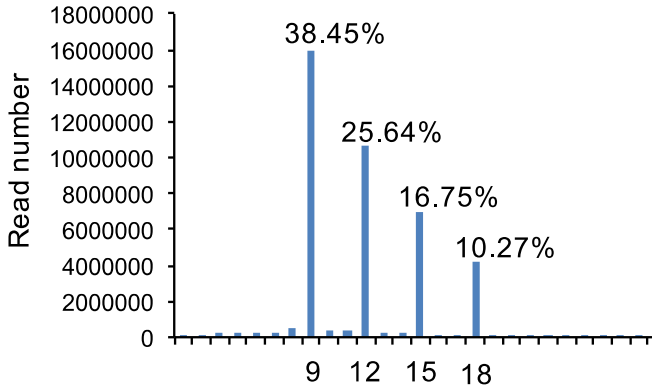


Figure S1. (a) Design of the nanobody library sequencing with an Illumina NextSeq 2×150 bp paired-end kit. Both ends of amplicons were barcoded by 8-bp unique molecular identifiers. (b) Distribution of unique clone counts in merged sequencing reads. (c) Length distribution of CDR3 in sequenced clones.

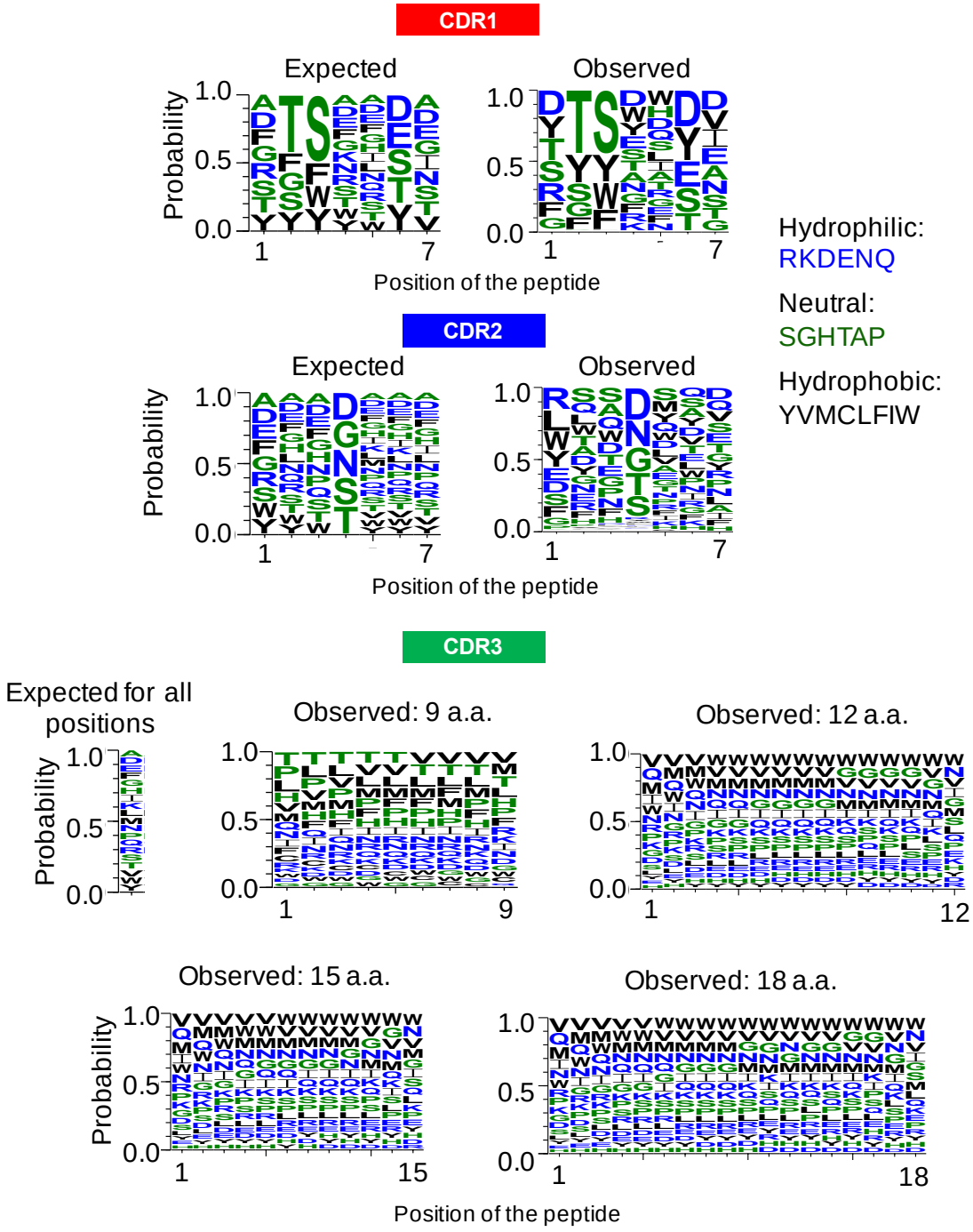


Figure S2. Designed and observed amino acid distributions in CDR randomization positions.

a

Round	Input phage	Output phage
1	2.4×10^{13}	3.5×10^{5}
2	1.3×10^{12}	4.7×10^{5}
3	1.4×10^{12}	2.3×10^{6}
4	1.5×10^{12}	1.2×10^{7}
5	1.8×10^{12}	1.8×10^{9}
6	2.2×10^{12}	5.2×10^{9}

b

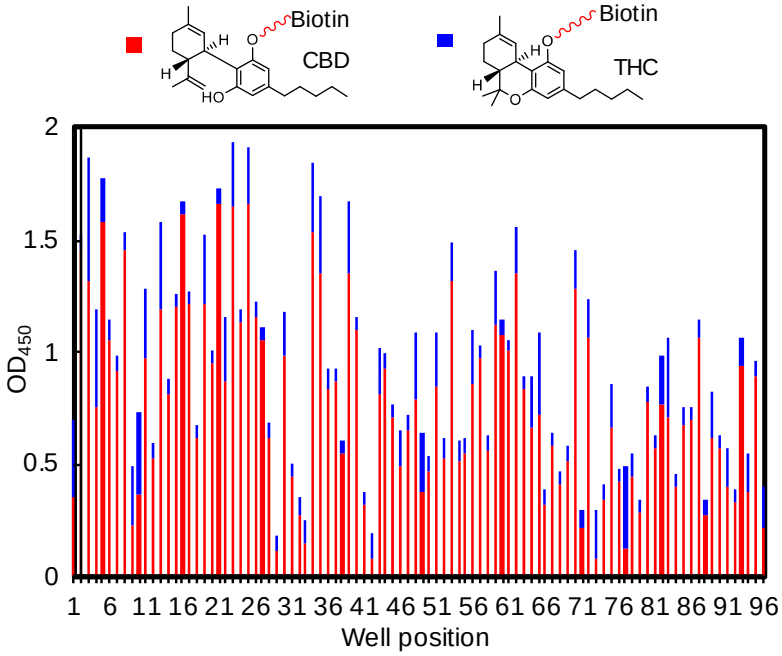


Figure S3. (a) Enrichment of phage titers following each round of biopanning for the anchor binder selection. (b) ELISA of 96 randomly picked clones after the six rounds of the anchor binder selection.

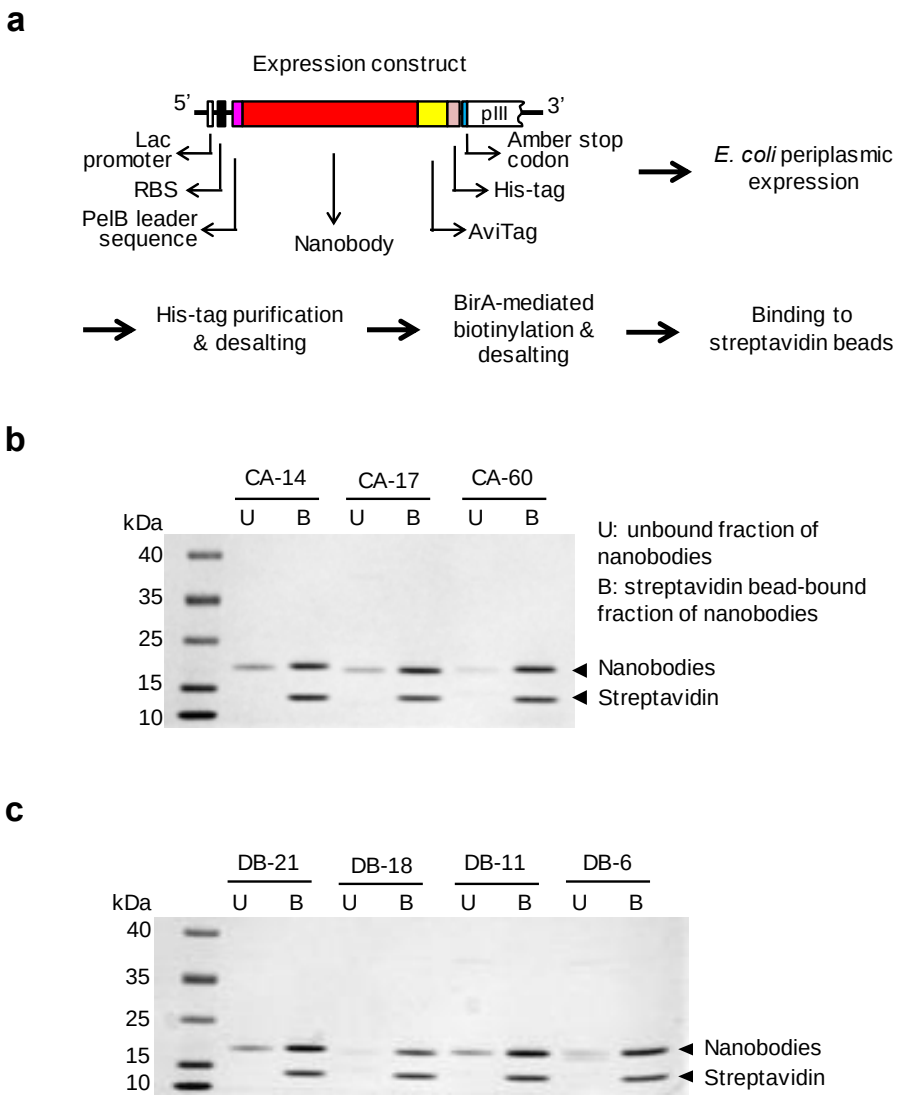


Figure S4. (a) Schematic of nanobody expression, purification, and biotinylation. SDS-PAGE analysis of free and streptavidin bead-bound anchor (b) and dimerization (c) binders after BirA-mediated biotinylation. The biotinylation efficiency was analyzed by incubating biotinylated nanobodies with streptavidin beads and then comparing the ratio of free to bead-bound nanobodies. Bead-bound nanobodies were eluted by boiling beads in a SDS sample loading buffer.

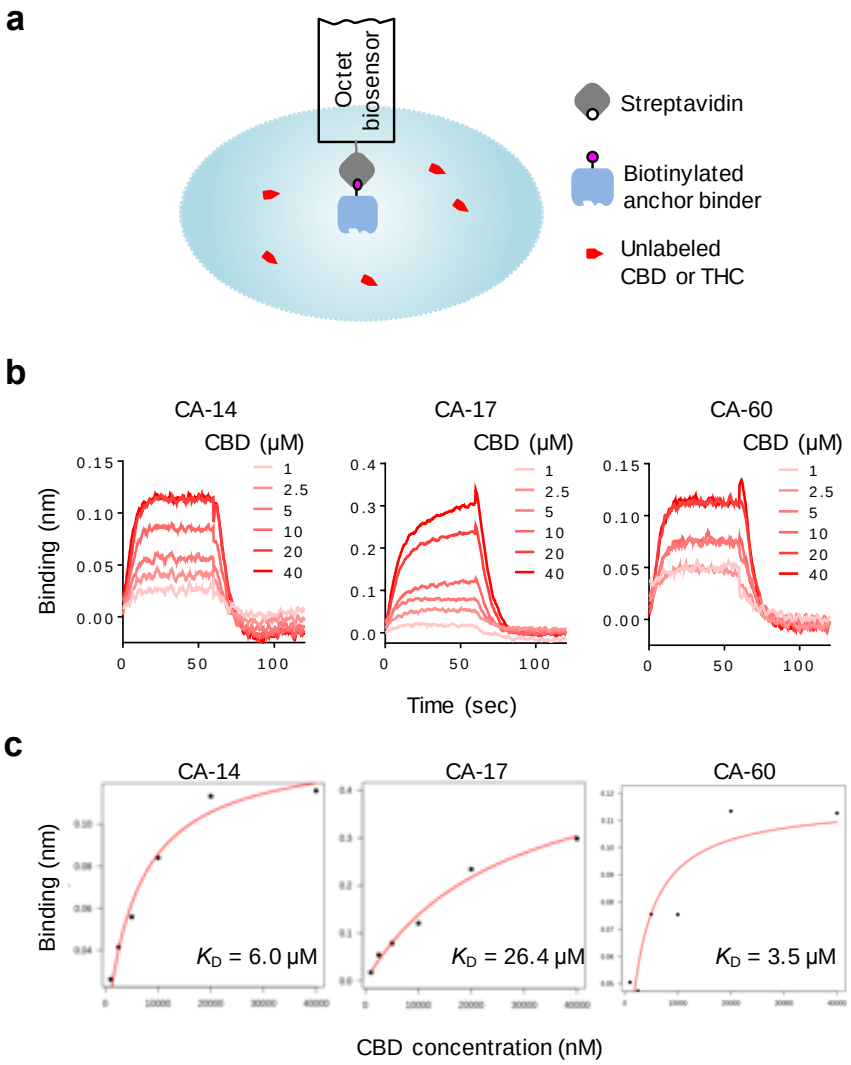


Figure S5. (a) BLI setup for the anchor binder–ligand interaction analysis. (b) BLI sensorgrams showing CBD binding to a panel of anchor binders. CA-14 was analyzed with one sensor per condition, while others were analyzed with a single sensor for all conditions. (c) Steady-state binding curve fitting.

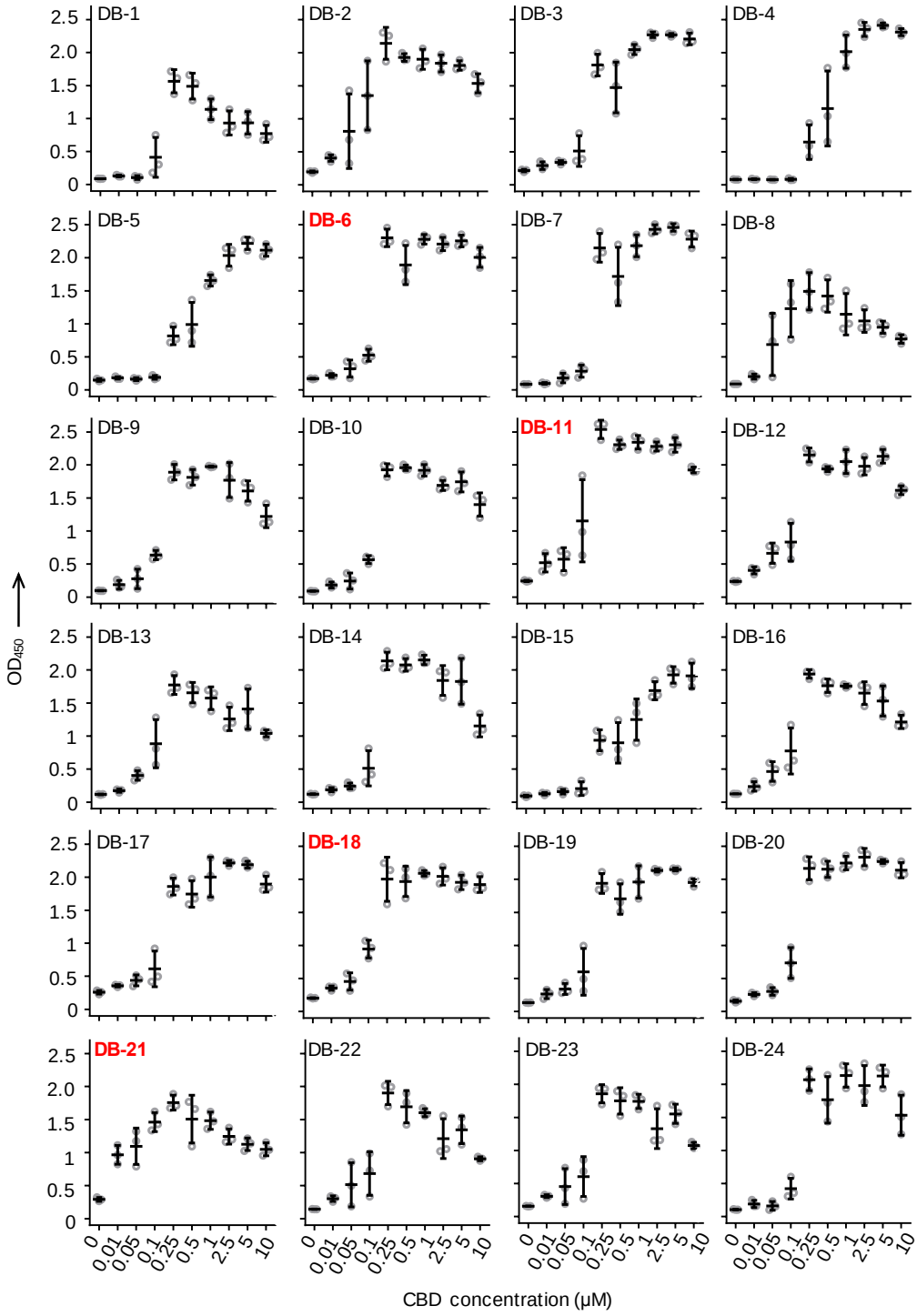


Figure S6. Titration ELISA of the anchor binder, CA-14, interacted with dimerization binder hits and different concentrations of CBD. Purified CA-14 was biotinylated and immobilized on the plate and bound to phage-displayed dimerization binders. Data represent mean values of 6 measurements; error bars, standard deviation.

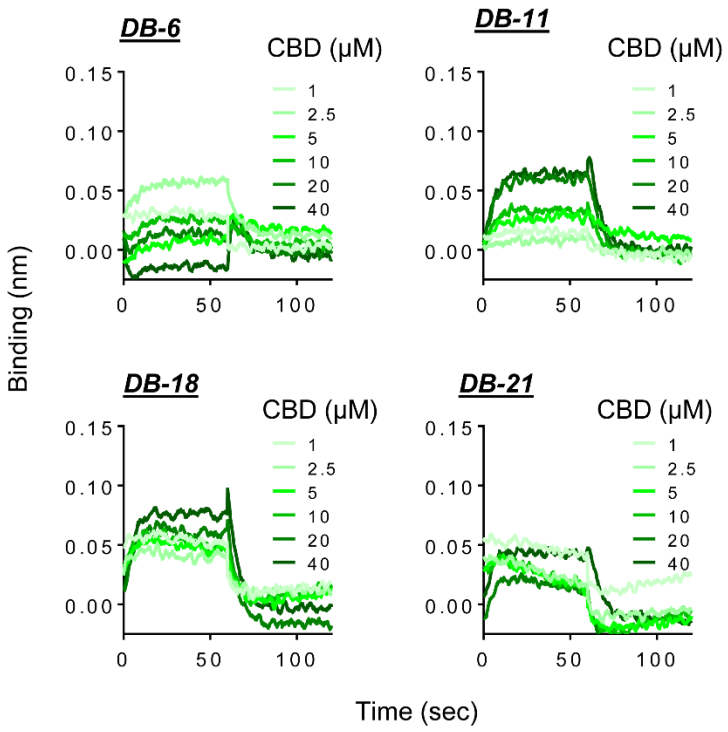


Figure S7. BLI sensorgrams showing CBD binding to dimerization binders, DB-6, 11, 18, and 21. Dimerization binders were immobilized on Super Streptavidin biosensors and interacted with unlabelled CBD.

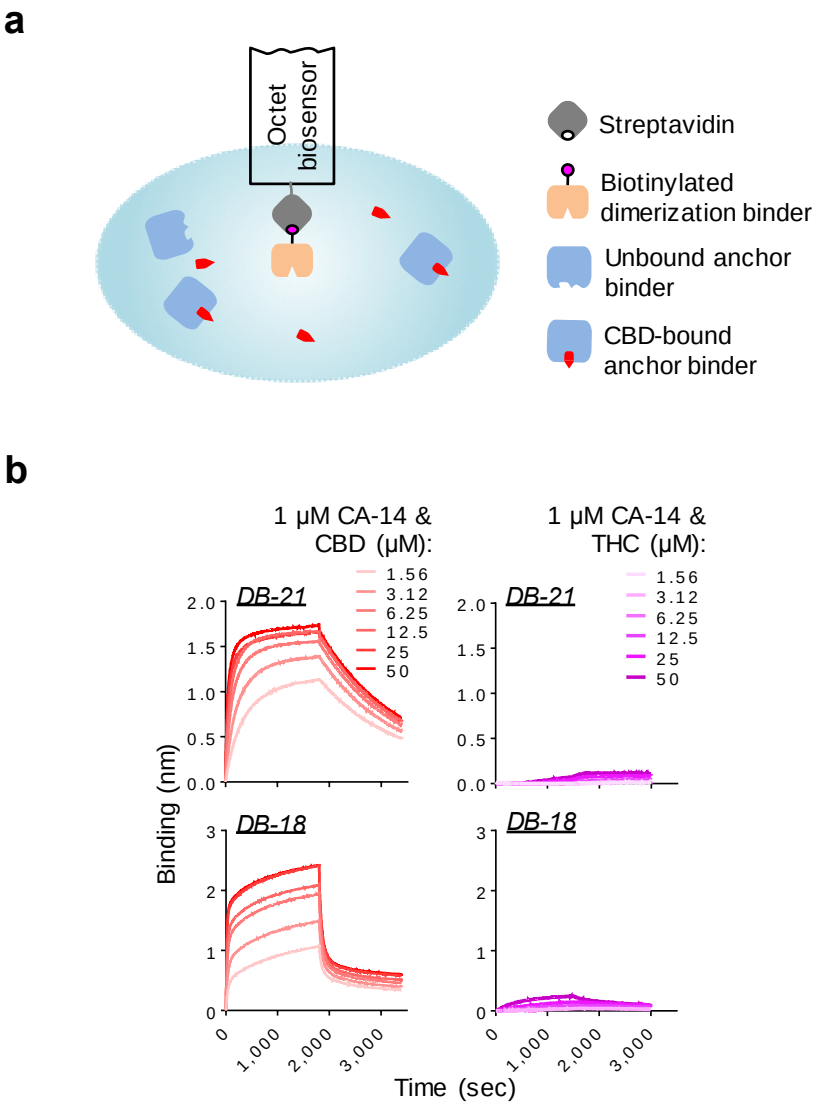
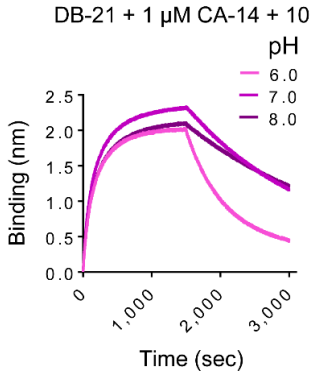


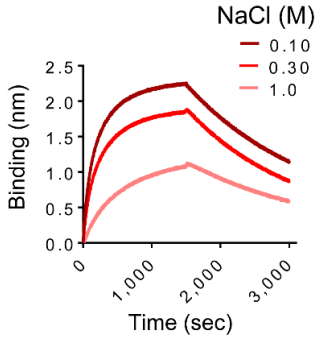
Figure S8. (a) BLI setup for the anchor binder-dimerization binder interaction analysis. (b) BLI sensorgrams showing the anchor binder-dimerization binder interactions in the presence of CBD or THC.

a



pH	k_{on}^{app} ($10^3 M^{-1}s^{-1}$)	k_{off}^{app} ($10^{-4}s^{-1}$)
6.0	1.50 ± 0.006	6.3 ± 0.03
7.0	0.35 ± 0.002	3.8 ± 0.06
8.0	0.10 ± 0.002	2.3 ± 0.02

b



NaCl (M)	k_{on}^{app} ($10^3 M^{-1}s^{-1}$)	k_{off}^{app} ($10^{-4}s^{-1}$)
0.10	0.24 ± 0.001	2.9 ± 0.02
0.30	0.19 ± 0.001	3.6 ± 0.02
1.0	0.02 ± 0.009	3.5 ± 0.16

Figure S9. The pH (a) and ion strength (b) effects on the DB-21 dimerization. For the pH effect analysis, the sensor-immobilized DB-21 was interacted with 1 μM CA-14 and 10 μM CBD in a 0.1 M phosphate buffers with different pHs. For the ion strength effect analysis, the assay was performed in the phosphate buffers, pH 7.0, with different concentrations of NaCl. For the three-component CID binding, apparent k_{on}s (k_{on}^{app}) and k_{off}s (k_{off}^{app}) were calculated using the Octet data analysis software 9.0.

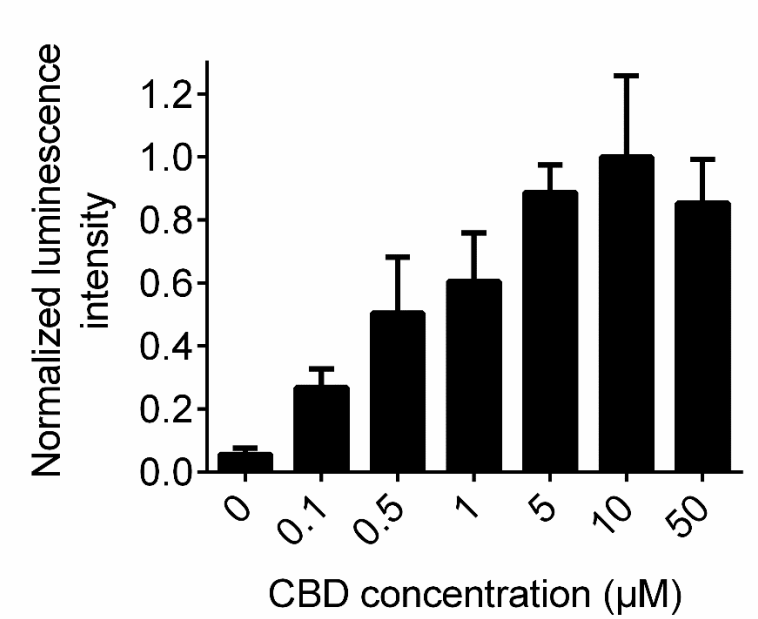


Figure S10. The NanoLuc complementation reporter-based luciferase assay. CA-14 and DB-21 bearing C-terminal LgBIT and SmBIT tags, respectively, were expressed in *E. coli* periplasm and released by osmotic shock. Released proteins (25 μL each) were mixed with CBD at indicated concentrations for 30 min at r.t. before measuring luminescence intensities (OD450). Data represent mean values of 3 measurements; error bars, standard deviation.

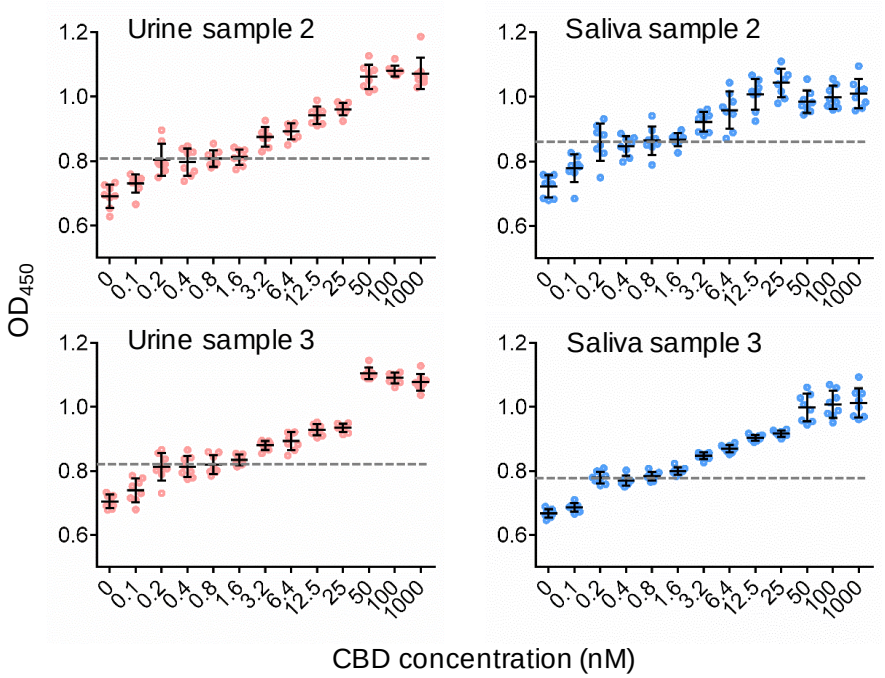


Figure S11. The sandwich ELISA-like assay of CBD-spiked urine and saliva samples in addition to the samples shown in Figure 4b. LoD ($mean_{blank}$ + 3 × (standard deviation, SD_{blank}) for the urine and saliva samples are ~0.8 nM.

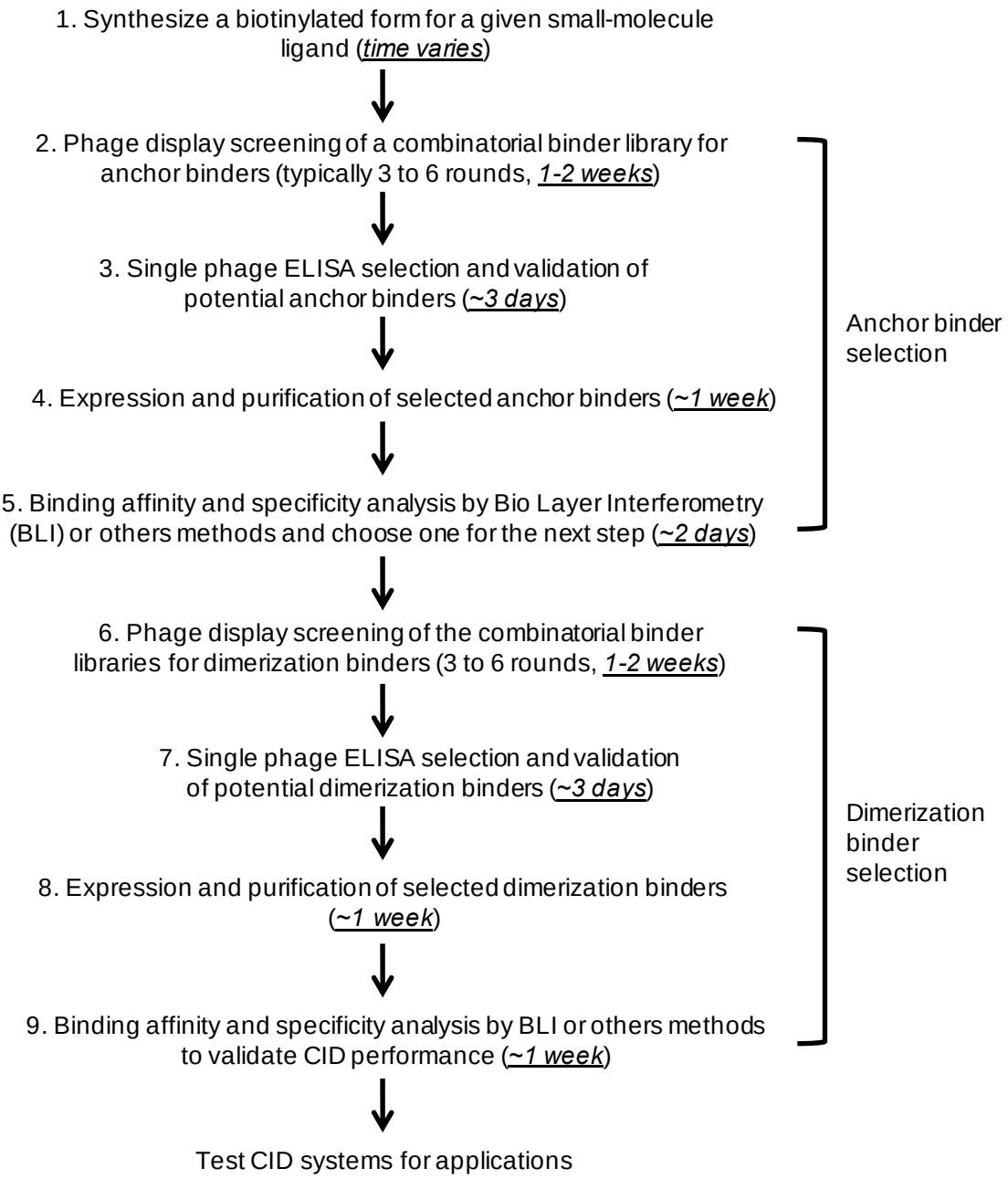


Figure S12. Flowchart and timeline of COMBINES-CID.

Table S1. CDR sequences of anchor and dimerization binder clones identified in the work. Clones selected to the detailed CID characterization are highlighted in bold.

Anchor Binders			
IDs	**CDR1**	**CDR2**	**CDR3**
CA-14	*STSRQYD*	*SNQDQPP*	*FKQHHANGA*
CA-17	DTSEDYD	FTSSNHT	FKKHASFPP
CA-60	HTSNAYD	SFPDAAV	YKNHPYDPP
Dimerization Binders			
IDs	**CDR1**	**CDR2**	**CDR3**
DB-1	DTYRLDT	YRTDQDH	GHSWWDLDE
DB-2	TGWEIES	FRANRFE	STFDSPSRR
DB-3	YTSFQYV	WLNGQVH	SMVFDHPQSGGGVET
DB-4	YGSDLDS	YAQDDWV	MSIWPEQHH
DB-5	DSSWWDG	WAFDNWR	YTNIDFQAYQSWFQNPPE
DB-6	*RFSWGEE*	*WAATPWQ*	*DEWHIGHVS*
DB-7	YTSDQDA	SSQSEIA	YRQSVHPQIASM
DB-8	FTFSQEE	FEDGMKG	WWYESHPQFQHQ
DB-9	DTFDLSA	WRDNPFR	MLQLHHHDG
DB-10	DTYNWDV	YEPSMYT	MMSSLHTFWANFQSD
DB-11	*TTSDNDT*	*WNGGRDE*	*YQDNRSWQE*
DB-12	GSYSWDA	YFGHNAY	VHFWKLLNE
DB-13	STYEWYS	WDEDNWN	EPQDGWTGV
DB-14	YTSAGEI	WWDGFAL	AHPSSTKMS
DB-15	RFSWGEE	WATAPWQ	YEWHIGHVS
DB-16	DFSSWDA	EGHSMTA	DIEFDLSMNHMYLVQ
DB-17	TTSDWYD	WWPTRAV	DWSFGMMQQ
DB-18	*GYSRADD*	*FGETDSF*	*YHNYTNMFE*
DB-19	DFYKLYS	WEAGMSH	LQDWMREWE
DB-20	RFSWGEK	WAAAPWQ	DEWRIDHVS
DB-21	*TTYGQTN*	*GLQGRDL*	*FHDFLRMWE*
DB-22	DTSNAST	WSSSPGN	MDAFHPQAW
DB-23	YGSFLDS	YAKDDGV	MSIWAEQHH
DB-24	DYSSTEI	AQPGVQQ	NVAFRHNHD

Table S2. Dissociation constants of anchor and dimerization binders in the presence or absence of CBD.

Dimerization binder ID	CBD	K_D (10^{-9} M)	k_{on} ($10^5 M^{-1}s^{-1}$)	k_{off} ($10^{-4}s^{-1}$)
DB-21	+	56.4	0.11	6.17
	-	N.D.	N.D.	N.D.
DB-18	+	553	0.41	227
	-	N.D.	N.D.	N.D.
DB-11	+	1,380	0.61	840
	-	N.D.	N.D.	N.D.
DB-6	+	19,000	0.096	1,820
	-	N.D.	N.D.	N.D.

Table S3. Synthetic oligos used for the nanobody library construction and deep sequencing in this study.

	Name	**Sequences (5' to 3')**	**Notes**
TG1 library	TG1F	TATAAT*GCCCAGCCGGC*CATGGCAGAAGTTCAGC TGCAGGCAAGC	*Bgl*I sites are underlined.
	TG1R	AATTAA*GCCTCCCGGGC*TGCTGCTAACGGTAACC TGGGT	
NGS library	UMI (F)	GTCGGAGCGTCAGATGTGTATAAGAGACAG*NNNN NNNN*TCTGAGCTGTGCACCCAGCGGT	UMIs are underlined; N=A/T/C/G
	UMI (R)	GGCTCGGAGATGTGTATAAGAGACAG*NNNNNNNN* TAACCTGGGTGCCCTGACCCCAATT	
	P5	AATGATACGGCGACCACCGAGATCTACACTCGTC GGCAGCGTCAGATGTGTATAAGAG	Index is underlined.
	P7 with an index	CAAGCAGAAGACGGCATACGAGAT*GCGTAGTA*GT CTCGTGGGCTCGGAGATGTGTATAAG	

SUPPLEMENTARY NOTES

1. Binding model expectations based on previous CID systems and ligand docking

Two mechanisms can explain the currently existing CID systems[6] (***Figure S13***): a) the chemical induces a large structural change upon binding one of the proteins, this structural change then allows for dimerization to the other protein, and b) the chemical inducer works as a bridge between the two proteins dimerizing.

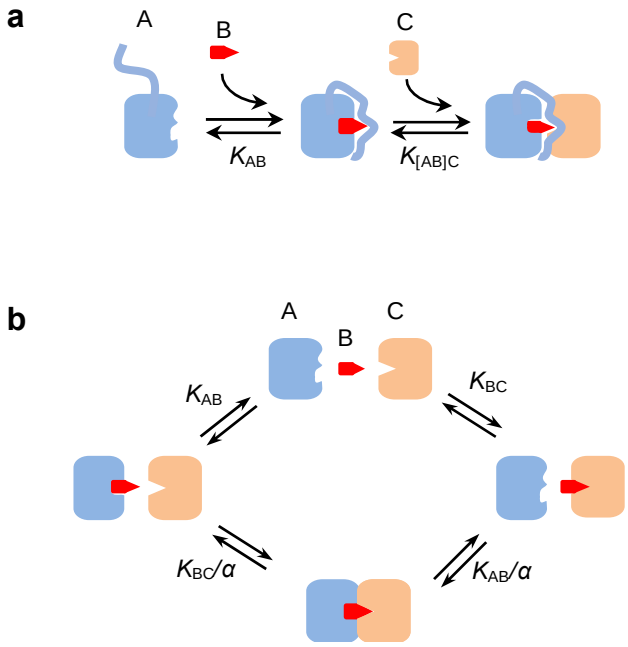


Figure S13. Two mechanisms explaining CID. (a) Conformational change induced binding typical of the gibberellin CID system. (b) Three-body binding equilibrium typical of the rapamycin CID system.

The gibberellin CID system is an example of protein dimerization which depends on a necessary conformational change in one of the dimerizing proteins. In the gibberellin CID system, the GID1A domain contains an N-terminal extension that, upon binding to gibberellin, forms a stable interaction surface for the DELLA domain.[7] In this system gibberellin is completely buried in a deep binding pocket and only the conformationally changed GID1A domain is free to bind the DELLA domain. Such system is kinetically simple with only two dissociation constants defining the system (***Figure S13a***).

The rapamycin CID system is an example of the "bridge molecule" mechanism. In this system the protein FKBP binds to one part of the rapamycin molecule, while the protein FRB binds to the other end.[8-9] In such systems, because both proteins can bind independently to the small molecule, the kinetics are more complex defined by two dissociation constants and a parameter, α, quantifying "binding cooperation" (***Figure S13b***). When $\alpha > 0$, this is termed positive cooperation and represents a binding enhancement upon ternary complex formation. When $\alpha < 0$, this is termed negative cooperation and represents a binding diminishment upon ternary complex

formation. Others have studied the three-body binding equilibrium in great detail.[10] Using established affinities,[8] the rapamycin CID system has positive corporation of ~2,000.

We hypothesize that our CBD CID system works by either of the two above described binding mechanisms. We primarily tested CBD anchor 14 (CA-14) which has relatively short CDR sequences (***Table S1***). With short CDRs it is unlikely that CBD binding induces a gibberellin-like conformational change through CDR loop stabilization. This also makes it unlikely that CBD is completely buried with no exposed parts, although, we cannot rule out this possibility. Indeed, our docking study suggests that CBD is still exposed after binding to CA-14 (***Figure S16***). This leads us to believe that CBD induced dimerization is governed by the "bridge molecule" mechanism similar to the rapamycin CID. The distinction between these binding models is important because binding through a bridge molecule leads to a system with possible auto-inhibition,[10] also known as the hook effect or the prozone effect. Auto-inhibition negatively affects dimerization at high concentrations of the bridge molecule because it competes against dimerization. The effect is described in detail in a mathematic model.[10]

Bio-layer interferometry to determine binding kinetics

In order to determine the binding kinetics of our nanobody CID system we performed a series of bio-layer interferometry (BLI) experiments. In all experiments we used an Octet RED96 system (ForteBio) and Streptavidin Biosensors (ForteBio). 200 nM biotinylated protein (***Figure S4b & 4C***) were immobilized on biosensors (Super Streptavidin, SSA, biosensors used for anchor binder-ligand tests; Streptavidin, SA, biosensors used for dimerization binder tests) with a binding assay buffer (1×PBS, pH 7.4, 0.05% Tween 20, 0.2% BSA, 3% methanol; Note: because CBD and THC were dissolved in methanol to make stock solutions, 3% methanol was added to the buffer to fairly compare samples and controls). Ligand loading was followed by a quenching step using free biotin. All experiments were performed using four measurements to control for sensor drift and unspecific binding:

a. No ligand, no analyte
b. With ligand, no analyte
c. No ligand, with analyte
d. With ligand, with analyte

The ligand is defined as the protein loaded to the SSA biosensor and the analyte is either protein or small molecule in solution. The final sensorgrams were derived from: (d - c) - (b - a).

We first determined the binding affinity between the CBD anchor 14 (CA-14) and CBD via steady-state analysis (***Figure S5***). For this experiment, BLI sensorgrams were collected using one SSA sensor per CBD concentration over six concentrations. Data analysis were performed in ForteBio Data Analysis v9 using a global 1:1 model (standard Langmuir isotherm). The global model requires binding parameters for all six concentrations to be the same, except the ligand

loading amount, R_{max}, which was unlinked to allow for differences in ligand loading caused by sensor variance. Steady-state analysis was then performed using the steady-state values estimated by this global model. We performed a similar experiment, using a single sensor for multiple concentrations, on the dimerization binders (DB). We observed minimal binding between DBs and CBD, but the K_Ds are too weak to be determined (***Figure S7***).

Next, we wanted to determine the binding affinity between CA-14 bound to CBD (referred to as CA-14*) and a panel of DBs. We observed no detectable binding between CA-14 and any of the DBs (***Figure 3a***). We also rely on the observation that the DBs have weak binding to CBD, and therefore we can assume that only CA-14* binds to the DB loaded to the sensor. We control the CA-14* concentration by using a fixed concentration of 1 μM CA-14 preincubated to equilibrium with six different concentrations of CBD. These CA-14* concentrations were then allowed to bind to DB loaded SA sensors and binding was measured (***Figure S8***). One sensor was used per CBD concentration. The sensorgrams initially showed a fast binding phase which was followed by a phase with slow binding. Such biphasic binding is an indication of a heterogeneous surface.[11] Attempts to optimize the binding behavior was unsuccessful so we turned towards fitting methods that take into account multiple binding phases. One such method is implemented as a publicly available software called EVILFIT.[12] EVILFIT integrates over a grid of possible affinities and assigns probabilities to each. Using CA-14's affinity to CBD, we calculated the CA-14* concentration at equilibrium for the different CBD concentrations and used this as the analyte concentrations to EVILFIT. At the two largest CBD concentrations we observed CBD based auto-inhibition of dimerization (discussed below) and therefore these were removed. When restricting our analysis to the four smallest CBD concentrations, EVILFIT showed good agreement between the data and the fit (***Figure S14***). Using these fits, we located the affinity distribution for the first binding phase and integrated over this to derive a point estimate of the affinities between CA-14 and the DBs (***Table S2***).

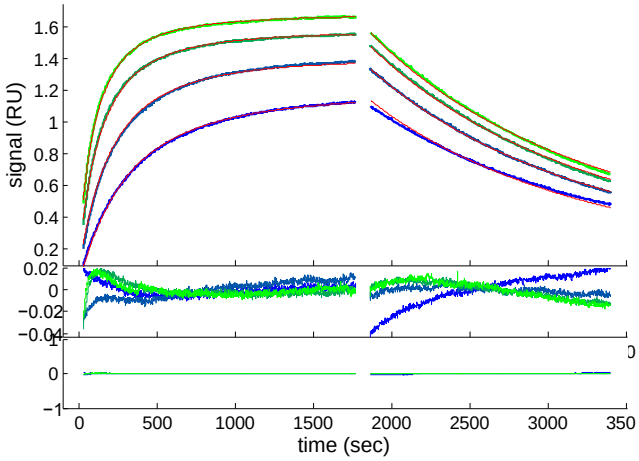


Figure S14. EVILFIT analysis for CBD inducible dimerization of CA-14 and DB-21. Upper panel showing BLI sensorgrams in green and blue shades and model fit in red. Lower panels showing the residues between data and fit.

Suggested binding model for CA-14 dimerizing DB-21

CA-14 and DB-21 is our most well characterized binding pair. BLI analysis established the binding affinity between CA-14 and CBD to be 6.0 μM (***Figure S5***). Docking studies suggested that CBD bound CA-14 in a way that exposes a part of CBD (***Figure S16***) potentially acting as a bridge to DB-21. This hypothesis is supported by the characteristic auto-inhibition we observed in our ELISA titration data (***Figure S6***), indicating that CBD at high concentrations is competing with dimerization. We observe a similar kind of auto-inhibition at the two highest CBD concentrations in our BLI assay (***Figure S8***).

If CBD is acting as a bridging molecule it must also bind DB-21. However, another BLI analysis showed that DB-21 only binds CBD very weakly (***Figure S7***), and therefore bridging cannot by itself explain the strong dimerization we observe (***Figure S8***). The observations can, however, be explained by positive cooperation, which enhances dimerization, and has before been shown to cause ~2,000 enhancement in rapamycin inducible dimerization.[8]

Douglass *et al.*[10] derives the equilibrium solution to such three-body binding systems with cooperation. Unfortunately, we cannot reliably measure DB-21's affinity to CBD (K_{BC} in ***Figure S13b***) due to its weak affinity and the limited solubility of CBD. However, we have been able to determine CBD's affinity to CA-14 and CA-14*'s affinity to DB-21 (K_{AB} and K_{BC}/α in ***Figure S13b***). As for the cooperation parameter, α, it must be so large that K_{BC} turns form undetectable by BLI to 56 nM (K_{BC}/α). α = 1,000 is a reasonable guess, also within the limit of what has been observed in the rapamycin CID, making K_{BC} equal to 56 μM. Plugging this into the three-body binding model from Douglass *et al.* (***Figure S15***), we get an explanation for the auto-inhibition we observe in ELISA titrations, as well as a reason why the two highest CBD concentrations in our BLI sensorgrams of dimerization (***Figure S8***) does not follow a standard steady-state curve.

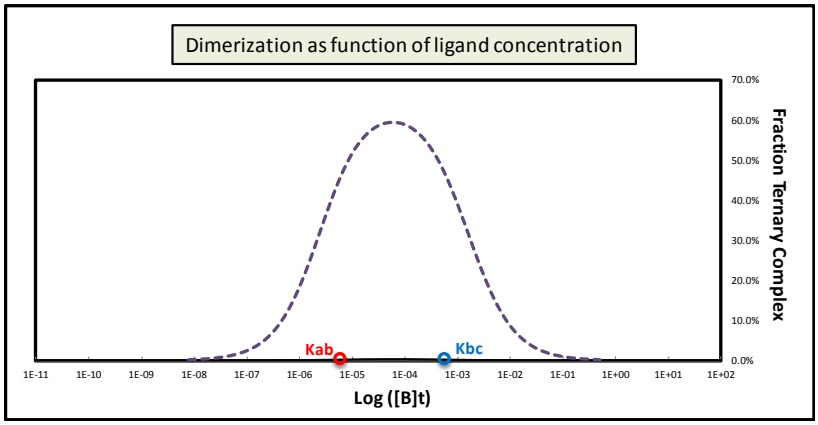


Figure S15. Example of a cooperative binding model.[10] Approximating the binding behavior of a BLI experiment where the concentration of immobilized ligand (in this case DB-21) is much

smaller than the analyte (in this case CA-14): [L] << [A]. α was set to 1,000 and K_{AB} and K_{BC} was set using BLI data from the CA-14, DB-21 and the dimerization.

2. Docking studies of CBD binding CA-14

Rosetta-based protein modelling and protein/ ligand docking was carried out with the anchor binder CA-14 and CBD. First, models of the nanobody CA-14 were constructed with RosettaCM.[13] Modelling used four template structures (pdb ids: 1mqk, 1t2j, 1uac, and 6cnw). The command used to run RosettaCM follows:

```
$ROSETTA/source/bin/rosetta_scripts.default.linuxgccrelease \
       -in:file:fasta CA14.fasta \
       -parser::protocol hybrid.xml \
       -relax::dualspace \
       -relax:default_repeats 5 \
       -default_max_cycles 200 \
       -score:weights beta_cart \
       -beta
```

The command above was provided the following XML script:

```
<ROSETTASCRIPTS>
    <SCOREFXNS>
        <ScoreFunction name="stage1" weights="score3">
            <Reweight scoretype="atom_pair_constraint" weight="0.1"/>
        </ScoreFunction>
        <ScoreFunction name="stage2" weights="score4_smooth_cart">
            <Reweight scoretype="atom_pair_constraint" weight="0.1"/>
        </ScoreFunction>
        <ScoreFunction name="fullatom" weights="beta_cart">
            <Reweight scoretype="atom_pair_constraint" weight="0.1"/>
        </ScoreFunction>
    </SCOREFXNS>
    <MOVERS>
        <Hybridize name="hybridize" stage1_scorefxn="stage1"
                   stage2_scorefxn="stage2" fa_scorefxn="fullatom" batch="1">
            <Template pdb="1mqkH_201.pdb" weight="1.0" cst_file="AUTO"/>
            <Template pdb="1t2jA_202.pdb" weight="1.0" cst_file="AUTO"/>
            <Template pdb="6cnwA_203.pdb" weight="1.0" cst_file="AUTO"/>
            <Template pdb="1uacH_204.pdb" weight="1.0" cst_file="AUTO"/>
       </Hybridize>
    </MOVERS>
    <PROTOCOLS>
        <Add mover="hybridize"/>
    </PROTOCOLS>
    <OUTPUT scorefxn="fullatom"/>
</ROSETTASCRIPTS>
```

RosettaCM was run in parallel, generating a total of 400 initial models. The 16 lowest energy models (***Figure S16***) were then used as receptor structures in ligand docking.

Ligand docking used an unpublished docking protocol in Rosetta, with a recently developed generalized energy function using prior methodology[14] for energy function fitting. A CBD ligand conformation was generated using OpenBabel.[15] This model was then minimized using AM1 and

AM1-BCC partial charges were generated with *antechamber*.[16] Finally, the refined small molecule was converted to a Rosetta parameter file with the following command:

```
python $ROSETTA/source/scripts/python/public/generic_potential/
       mol2genparams.py --amide_chi -s LG1.mol2
```

Docking in Rosetta was carried out running the following command:

```
$ROSETTA/source/bin/rosetta_scripts.linuxgccrelease \
     -database ~/Rosetta_efunc/database \
     -s input.pdb \
     -parser:protocol dock.xml \
     -beta_cart \
     -extra_res_fa CBD.params  \
     -no_autogen_cart_improper
```

The input XML, *dock.xml*:

```
<ROSETTASCRIPTS>
  <SCOREFXNS>
      <ScoreFunction name="dockscore" weights="beta">
        <Reweight scoretype="fa_rep" weight="0.2"/>
        <Reweight scoretype="coordinate_constraint" weight="0.1"/>
      </ScoreFunction>
      <ScoreFunction name="relaxscore" weights="beta_cart"/>
  </SCOREFXNS>
  <MOVERS>
    <GALigandDock name="dock" runmode="dockflex" scorefxn="dockscore"
                  scorefxn_relax="relaxscore" sidechains="auto">
      <Stage repeats="10" npool="100" pmut="0.2" smoothing="0.375"
             rmsdthreshold="2.0" maxiter="50" pack_cycles="100"
             ramp_schedule="0.1,1.0"/>
    </GALigandDock>
  </MOVERS>
  <PROTOCOLS>
    <Add mover="dock"/>
  </PROTOCOLS>
  <OUTPUT scorefxn="relaxscore"/>
</ROSETTASCRIPTS>
```

The input files from this run were the ten lowest-energy receptor structures from stage 1. Initially, input ligands were placed in a random orientation such that the ligand center of mass was coincident with the C-beta of residue 33, the residue closest to the center of mass of the variable loops in the nanobody.

The previous XML samples an ensemble of 100 structures in a single trajectory. We ran a total of 160 such trajectories, ten trajectories for each of the 16 initial models, generating a total of 16,000 models. Visual inspection of the lowest-energy 20 docked conformations yielded two unique configurations, illustrated in ***Figure S16***. Both suggest that the solvent-exposed phenolic hydroxyl group of anchor-bound CBD, where biotin was attached in the anchor binder selection, might be involved in the dimerization recognition.

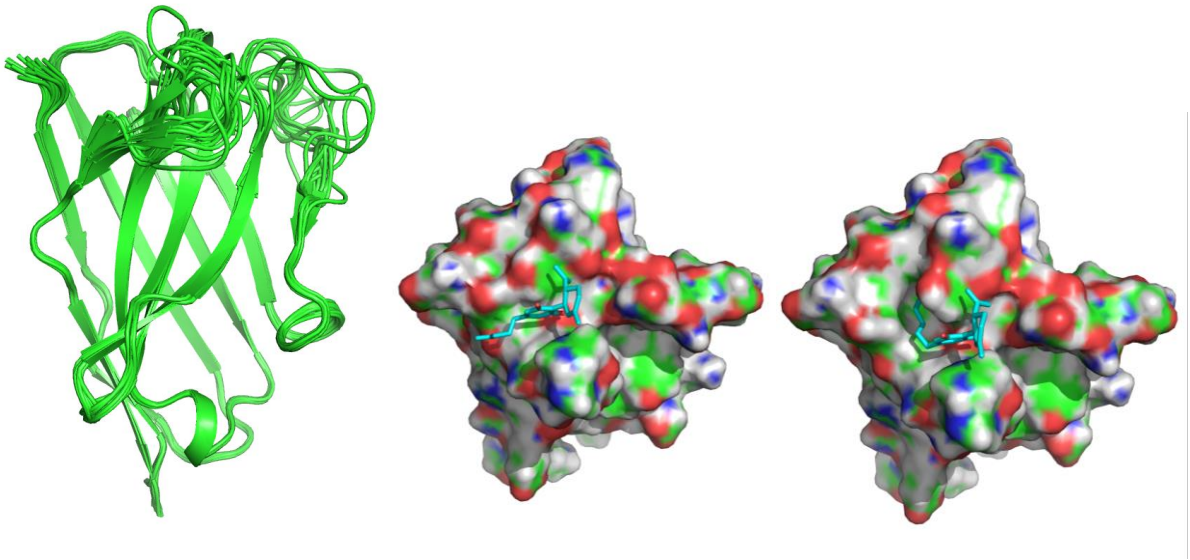


Figure S16. Docking results. (left) The 16 low-energy models of the CA-14 construct. (middle, right) The two putative binding models of CBD suggested from the docking simulation.

References

(1) Qi, L. W.; Yamamoto, N.; Meijler, M. M.; Altobell, L. J.; Koob, G. F.; Wirsching, P.; Janda, K. D. Delta(9)-tetrahydrocannabinol immunochemical studies: Haptens, monoclonal antibodies, and a convenient synthesis of radiolabeled Delta(9)-tetrahydrocannabinol. *J. Med. Chem.* **2005,** *48*, 7389-7399.
(2) Moutel, S.; Bery, N.; Bernard, V.; Keller, L.; Lemesre, E.; de Marco, A.; Ligat, L.; Rain, J. C.; Favre, G.; Olichon, A.; Perez, F. NaLi-H1: A universal synthetic library of humanized nanobodies providing highly functional antibodies and intrabodies. *eLife* **2016,** *5,* e16228.
(3) Roehr, J. T.; Dieterich, C.; Reinert, K. Flexbar 3.0-SIMD and multicore parallelization. *Bioinformatics* **2017,** *33*, 2941-2942.
(4) Crooks, G. E.; Hon, G.; Chandonia, J. M.; Brenner, S. E. WebLogo: A sequence logo generator. *Genome Res.* **2004,** *14*, 1188-1190.
(5) Beckett, D.; Kovaleva, E.; Schatz, P. J. A minimal peptide substrate in biotin holoenzyme synthetase-catalyzed biotinylation. *Protein Sci.* **1999,** *8*, 921-929.
(6) Stanton, B. Z.; Chory, E. J.; Crabtree, G. R. Chemically induced proximity in biology and medicine. *Science* **2018,** *359*, eaao5902.
(7) Murase, K.; Hirano, Y.; Sun, T. P.; Hakoshima, T. Gibberellin-induced DELLA recognition by the gibberellin receptor GID1. *Nature* **2008,** *456*, 459-463.
(8) Banaszynski, L. A.; Liu, C. W.; Wandless, T. J. Characterization of the FKBP-Rapamycin-FRB ternary complex. *J. Am. Chem. Soc.* **2005,** *127*, 4715-4721.
(9) Liang, J.; Choi, J.; Clardy, J. Refined structure of the FKBP12-rapamycin-FRB ternary complex at 2.2 angstrom resolution. *Acta Crystallogr. D Biol. Crystallogr.* **1999,** *55*, 736-744.
(10) Douglass, E. F.; Miller, C. J.; Sparer, G.; Shapiro, H.; Spiegel, D. A. A comprehensive mathematical model for three-body binding equilibria. *J. Am. Chem. Soc.* **2013,** *135*, 6092-6099.
(11) Schuck, P.; H., Z. The role of mass transport limitation and surface heterogeneity in the biophysical characterization of macromolecular binding processes by SPR biosensing. *Methods. Mol. Biol.* **2010,** *627*, 15-54.
(12) Zhao, H. Y.; Gorshkova, II; Fu, G. L.; Schuck, P. A comparison of binding surfaces for SPR biosensing using an antibody-antigen system and affinity distribution analysis. *Methods* **2013,** *59*, 328-335.
(13) Song, Y. F.; DiMaio, F.; Wang, R. Y. R.; Kim, D.; Miles, C.; Brunette, T. J.; Thompson, J.; Baker, D. High-resolution comparative modeling with RosettaCM. *Structure* **2013,** *21*, 1735-1742.
(14) Park, H.; Bradley, P.; Greisen, P.; Liu, Y.; Mulligan, V. K.; Kim, D. E.; Baker, D.; DiMaio, F. Simultaneous Optimization of Biomolecular Energy Functions on Features from Small Molecules and Macromolecules. *J. Chem. Theory. Comput.* **2016,** *12*, 6201-6212.
(15) O'Boyle, N. M.; Banck, M.; James, C. A.; Morley, C.; Vandermeersch, T.; Hutchison, G. R. Open Babel: An open chemical toolbox. *J. Cheminform.* **2011,** *3, 33*.
(16) Wang, J. M.; Wang, W.; Kollman, P. A.; Case, D. A. Automatic atom type and bond type perception in molecular mechanical calculations. *J. Mol. Graph. Model.* **2006,** *25*, 247-260.