Supplementary Materials for

Computational Design of Ligand Binding Proteins with High Affinity and Selectivity

Christine E. Tinberg[1], Sagar D. Khare[1], Jiayi Dou, Lindsey Doyle, Jorgen W. Nelson, Alberto Schena, Wojciech Jankowski, Charalampos G. Kalodimos, Kai Johnsson, Barry L. Stoddard & David Baker*

*To whom correspondence should be addressed. E-mail: dabaker@uw.edu

[1]These authors contributed equally to this work.

This PDF file includes:

- Supplementary Methods
- Supplementary Tables 1-22
- Supplementary Figures 1-21
- Supplementary Data

Supplementary Methods

Computational Methods. Digoxigenin binders were designed using an updated version[1] of RosettaMatch[2] to search for PDB scaffold backbones that can accommodate pre-defined interactions to the ligand followed by RosettaDesign[3] to optimize the binding site amino acid sequences of the matches for ligand binding affinity.

Generation of ligand and ligand conformer library. The 3-dimensional structure of digoxigenin (DIG) was obtained from PDB ID 1LKE[4]. Because our experimental validation and selection methods rely on the presence of a linker that connects the O5 hydroxyl of the DIG molecule to either biotin or carrier protein, we included this linker in our ligand model (Supplementary Fig. 1b). Linker atoms were added to DIG using the Build functionality of MacPyMOL (Schrödinger, LLC).

A ligand conformer library was generated by sampling conformations around the C3–O5 and N1-C26 bonds (Supplementary Fig. 1b) at -60° ± 30°, 60° ± 30°, and 180° ± 30°. Conformers were rejected if there were significant clashes within the molecule by using an intra_fa_rep cutoff value of 0.25 Rosetta energy units (Reu). Although the lactone–cardenoline bond (C17–C20) of the steroid is freely rotatable in solution, we restricted this torsion angle to that found in PDB ID 1LKE and PDB ID 1IGJ for simplicity.

Scaffold selection. A set of 401 scaffolds was generated for use as input structures for matching. This set contained 57 scaffold proteins previously used for enzyme design projects within our lab that span a variety of structural and functional classes including periplasmic binding proteins and lipid-binding proteins,[5-7] as well as 344 structural homologs[8] of a subset of these scaffolds (PDB codes 1m4w, 1oho, 1a53, 1thf, 1dl3, and 1e1a) having a DALI Z-score cutoff value of 8 from the input search model. These six scaffolds were chosen because of previous enzyme-design successes in these fold classes[5-7] and/or because of their thermostability, as directed evolution experiments have shown that more stable scaffolds can acquire new functions more easily than their less stable counterparts[9,10]. The homolog subset comprised 8 Concanavalin A-like lectins/glucanases (homologs of 1m4w from *Nonomurea flexuosa*), 91 cystatin-like proteins (homologs of 1oho from *Pseusomonas putida*), 208 TIM β/α-barrels (28 homologs of 1a53 from *Sulfobus solfataricus*, 46 unique homologs of 1thf from *Thermotoga maritima*, and 134 unique homologs of 1dl3 from *Thermotoga maritima*), and 37 6-bladed β-propeller proteins (homologs

of 1e1a from *Loligo vulgaris*). All of these proteins are enzymes that bind small molecule substrates. All 401 scaffolds comprise <350 amino acids, have been expressed previously in *E. coli*, and were stripped of their cognate bound small molecules and water molecules before use. The PDB IDs for all 401 scaffolds are listed in Supplementary Table 8. To identify residue positions to be used for matching in the homolog scaffolds, each homolog crystal structure was superimposed on that of its parent scaffold using the CEAlign plug-in of the PyMOL molecular visualization program, and then homolog residue positions within 5.0 Å of any ligand heavy atom present in the parent scaffold were identified. For PDBs 1a53, 1dl3 and 1oho, ligands present in the crystal structures were used in this search. For 1m4w, 1e1a and 1thf, ligand positions from the computational design models of a retroaldolase (RA60)[5], a Diels-Alderase (DA_20)[7], and a Kemp Eliminase (KE_007)[6] were used, respectively.

Geometric placement of ligand using a set of pre-selected interactions (matching). Geometric criteria for enforcing binding site interactions were determined by inspecting structures of digoxin bound to the anti-digoxigenin antibody 26-10, PDB ID 1IGJ[11], and of digoxigenin bound to the engineered lipocalin DigA16, PDB ID 1LKE[4]. From these structures we defined five interface criteria: (1) hydrogen bond between the lactone carbonyl oxygen O1 and a Tyr side chain, (2) hydrogen bond between the O2 hydroxyl and a histidine or Tyr side chain, (3) hydrogen bond between the O3 hydroxyl and a His or Tyr side chain, (4) hydrophobic packing interaction on the top face of the ligand, and (5) hydrophobic packing interaction on the bottom face of the ligand. Two active site configurations were specified: one having Tyr, Tyr, His, Phe/Tyr, and Phe/Tyr/Trp satisfying design criteria 1-5 (DIG_yyhff), and one having Tyr, His, His, Phe/Tyr/Trp, and Tyr/Trp satisfying design criteria 1-5 (DIG_yhhff).

Geometric criteria were defined using six degrees of freedom between the ligand and the desired interacting side chain using a matching constraints file[1]. Extra rotamer sampling (two half step standard deviations) was performed around all side chain torsion angles. To enforce burial of the lactone head group within a binding pocket, we considered only those residue positions in the binding site that had a minimum of 14 neighboring residues during matching for constraint 1 (hydrogen bond to the lactone carbonyl oxygen). A neighbor was defined as a residue having Cα within 10 Å of the Cα of the binding site position under consideration. Secondary matching[1] was used for constraints 3, 4, and 5. To eliminate high-energy rotamer conformations, a maximum Dunbrack energy (fa_dun) cutoff of 4.5 Reu (unweighted) was used while building rotamers for all constraints. Using these matching criteria, 29,274 and 30,861 matches were found for DIG_yyhff and DIG_yhhff, respectively.

Rosetta sequence design. Active site amino acid sequences of each match were designed to maximize binding affinity to the ligand according to the Rosetta energy function using the enzdes weights set for the energy terms[1,12]. Explicit electrostatics were not used. Design moves were followed by steepest descent gradient minimization in which side chain degrees of freedom and the relative orientation of the ligand with respect to the protein were allowed to minimize freely[13] but backbone minimization was restricted such that Cα atoms were only allowed to move ≤0.05 Å from their pre-minimization positions. Internal torsions of the ligand were allowed to minimize but were constrained to be within 5 degrees of their initial values.

Two successive rounds of sequence design were used to generate designs. The purpose of the first round was to maximize binding affinity for the ligand[1]. To prevent destabilization of the apo-protein that can result from mutating potentially stabilizing residues having side chains important for core packing, aromatic residues in the scaffold were only allowed to mutate to other aromatics during this round of design. A RosettaScripts XML file entitled ligdes.xml for running the first round of sequence design is provided in Supplementary Data.

After the first round, a second round of binding site sequence design was performed on the output files of the first round. The goal of this round was to optimize protein stability while maintaining the binding interface designed during the first round as much as possible. Ligand-protein interactions were up-weighted by a factor of 1.5 relative to intra-protein interactions during sequence optimization in attempt to ensure that the interface binding affinity was maintained, and two different criteria were used to optimize protein stability: (1) native scaffold residues identities were favored by 1.5 Rosetta energy units (Reu), and (2) no more than five residues were allowed to change from identities observed in a multiple sequence alignment (MSA) if (a) these residues were present in the MSA with a frequency greater than 0.6 as specified by a position-specific sequence matrix (PSSM) and, (b) if the calculated ΔΔG for mutation of the scaffold residue to alanine was greater than 1.5 Reu in the context of the scaffold sequence. The ΔΔG for mutation to alanine was estimated as described[14] and PSSM files were generated using NCBI PSI-BLAST. For both the DIG_yhhff and the DIG_yyhff designs, a first method restricted the amino acid identities of the hydrogen bonding (Tyr/His) residues to their pre-selected (matched) identities during the design. For the DIG_yhhff designs, we used an alternative second method in which the matched residues were allowed to mutate to any amino acid according to the MSA and ΔΔG criteria described above. Designs generated using this latter protocol were filtered to ensure the presence

of at least three hydrogen bonds between the protein and the ligand. RosettaScripts XML files for running the two second-round design methods, ligdes_fix_cst.xml and ligdes_flex_hb.xml, respectively, are given in Supplementary Data.

Evaluation of designs. Designs passing the filters encoded in the XML files (see Supporting Data; for the DIG_yyhff but not the DIG_yhhff designs, interface energy was used as a filter as noted in Supporting Data) were subjected to several additional filtering criteria. High shape complementary was enforced using by rejecting designs having $S_c < 0.5$. Shape complimentary was computed using the CCP4 package v.6.0.2[15] using the S_c program[16] and the Rosetta radii library. A common feature of the engineered DIG-binding lipocalin DigA16 (PDB IDs ILKE and 1KXO)[4] and the anti-DIG 26-10 antibody (PDB IDs 1IGJ and 1IGI)[11] is that the binding site is largely pre-organized; there are very few structural changes between the bound and unbound forms of the proteins. We therefore attempted to enforce pre-organization of the binding-competent conformation of the apo-protein by two metrics: (1) introducing second-shell amino acids that hold the pre-selected residues in place via hydrogen bonding or sterics using Foldit[17], and (2) selecting designs having Boltzmann-weighted side chain probabilities[18] > 0.1 for at least one hydrogen bonding residue.

Compatibility of designed sequence with local backbone structure. We reasoned that binding site pre-organization would be compromised if substitution of amino acid side chains during (fixed backbone) design leads to a change in the backbone conformational preference in regions sequence-local to the sites of substitution. Therefore, we developed a metric to estimate the impact of design on local backbone structure and used this metric to discard designs that were predicted to lead to backbone structure changes. Using the structure prediction modules of Rosetta[19], we generated a set of 9-mer fragment structures for each designed and wild type scaffold sequence and compared the average RMSD of these fragments to those of the scaffold backbone structures. If the average RMSD of conformations predicted in these fragments (200 9-mers) near any designed position was greater (> 0.8 Å) for the designed sequence than the wild type scaffold sequence, we flagged that region of the designed protein as unlikely to adopt the local backbone conformation of the scaffold protein and rejected that designed protein.

Design Scoring. Following automated filtering, all designs were inspected manually using Foldit[17] and some ligand-proximal residues were manually reverted back to their native scaffold identity to increase

the likelihood of design stability. Finally, 17 designs in 14 unique scaffolds were chosen for experimental testing (Supplementary Table 2). For scoring, all design models were relaxed with backbone and side chain heavy atom constraints[20] using Rosetta relax[21].

Modeling directed evolution mutations. Mutations arising from directed evolution studies were modeled using RosettaScripts[12]. Mutations were introduced in the parent model, then residues having Cα within 10 Å of any ligand heavy atom and having Cα within 12 Å of any ligand heavy atom and Cβ closer to any heavy atom in the ligand than Cα were repacked using the soft_rep score function[22]. All side chains, the rigid body orientation of the ligand with respect to the protein, and internal ligand torsions were minimized using the Rosetta energy function with the enzdes weights set. Backbone minimization was restricted such that Cα atoms were only allowed to move ≤ 0.05 Å from their pre-minimization positions. Ten trajectories were run and the one having the lowest interface energy was selected. An example RosettaScripts XML file, make_mutation.xml, is given in Supplementary Data.

Materials. Digoxigenin, digoxin, digitoxigenin, progesterone, and β-estradiol were purchased from Sigma Aldrich (St. Louis, MO) and were used as received. DIG-BSA was purchased from CalBioreagents (San Mateo, CA, ~10 DIG molecules per BSA). EZ-link-sulfo-NHS-biotin was purchased from Thermo Fisher Scientific (Waltham, MA). Ribonuclease A (RNase A) and DIG-NHS were from Sigma Aldrich (St. Louis, MO). Reagents and solvents used for the synthesis of the digoxigenin derivatives were purchased from Sigma Aldrich and used without any further purification. Dimethylsulfoxide was stored over activated molecular sieves (Sigma-Aldrich, 4A, beads 8-12 mesh) for at least 24 hours before use. High-resolution mass spectra (HRMS) were collected with a LCQ Fleet Ion Trap Mass Spectrometer (Thermo Scientific). Reverse-phase analytical high-pressure liquid chromatography (RP-HPLC) was run on a Dionex system equipped with a P680 pump, an ASI 100 automatic sample injector and an UltiMate 3000 diode array detector for product visualization using a Waters symmetry C18 column (5μm, 3.9 x 150 mm). Reverse-phase preparative high-pressure liquid chromatography was performed on a Dionex system equipped with an UltiMate 3000 pump and an UVD 170U UV-Vis detector for product visualization on a Waters SunFire™ Prep C18 OBD™ 5 μm 19×150 mm Column. Proton and carbon nuclear magnetic resonance (NMR) spectra were recorded at room temperature on a Bruker Avance-III 400 or on a Bruker DRX-600 equipped with a cryoprobe. Chemical

shifts (δ) are reported in ppm relative to the solvent residual signals. Synthetic schemes are given in Supplementary Fig. 21.

Biotinylation of DIG-BSA. DIG-BSA was prepared by reacting 50 μL of a 58 μM solution of DIG-BSA (2.9 nmol) with 8 μL of a 1.8125 mM solution of EZ-link-sulfo-NHS-biotin (14.5 nmol, 5 eq) in PBS for 1 hr at RT. A 10 μL portion of 14.5 mM glycine was added to quench the reaction. After 30 min, the reaction mixture was centrifuged and soluble protein was purified from excess small molecules by repeated rounds of centrifugal concentration and dilution into PBS until the absorbance of the flow-through remained constant.

Synthesis of DIG-RNase-biotin. A 460 μL portion of a 365 μM solution of Ribonuclease A (168 nmol; RNase A) prepared in PBS was reacted with 30 μL of a 9.73 mM solution of EZ-link-sulfo-NHS-biotin (292 nmol, 1.7 eq) prepared in PBS and 10 μL of a 106.3 mM solution of DIG-NHS (1 μmol, 6 eq) prepared in DMSO for 1 hour at RT. A 20 μL portion of 385 mM glycine was added to quench the reaction. After 20 min, the reaction mixture was centrifuged and soluble protein was purified from excess small molecules by repeated rounds of centrifugal concentration and dilution into PBS until the absorbance of the flow-through remained constant.

Synthesis of Biotin-PEG_3-NH_2 (2) Biotin (**1**, 13.5 mg, 55.3 μmol, 1 eq) was dissolved in 100 μL of dimethylsulfoxide (DMSO) and diisopropylethylamine (DIEA) was added (19.3 μL, 2 eq). O-(N-succinimidyl)-N,N,N',N'-tetramethyl-uronium (TSTU, 15.0 mg, 0.9 eq) was added and the clear solution was stirred for 10 minutes at room temperature to form the biotin-NHS ester. 4,7,10-Trioxa-1,13-tridecanediamine (18 mg, 1.5 eq) was dissolved in 200 μL of dry DMSO and the biotin-NHS was added drop wise under vigorous stirring over 5 minutes. The mixture was stirred for a further 10 minutes at room temperature. 1.5 mL of diethyl ether was added to the clear solution and the resulting suspension was centrifuged. The supernatant ether phase was discarded and the remaining oil was purified by preparative RP-HPLC (5mL/min, 10-100% acetonitrile in 0.1% TFA in H_2O). The fractions containing the product were lyophilized to afford **2** as a yellowish liquid (15 mg, 67%). [HRMS (ESI): 447.42 m/z (447.7 m/z expected). ^{1}H NMR (400 MHz, DMSO) δ 7.78 (t, 1 H, J = 5.6 Hz), 7.70 (m, 2 H), 6.42 (d, 1 H, J = 0.2 Hz), 6.37 (m, 1 H), 4.31 (m, 1 H), 4.13 (dd, 1 H, J = 7.6, 4.5 Hz), 3.50 (m, 11 H), 3.39 (t, 2 H, J = 6.3 Hz), 3.08 (m, 3 H), 2.85 (m, 3 H), 2.05 (t, 2 H, J = 7.4 Hz), 1.78 (m, 2 H), 1.61 (m, 4 H), 1.49 (m,

3 H), 1.30 (m, 2 H). ^{13}C NMR (101 MHz, DMSO) δ 172.4, 163.2, 70.2, 70.1, 70.0, 70.0, 68.6, 67.8, 61.5, 59.7, 55.9, 40.3, 37.3, 36.2, 35.7, 29.9, 28.7, 28.5, 27.7, 25.8.]

Synthesis of Digoxigenin-PEG$_3$-biotin (3) Digoxigenin-NHS ester (1 mg, 1.5 μmol) was dissolved in 100 μL of DMSO and DIEA (0.4 mg, 3.0 μmol) was added, followed by **2** (1.3 mg, 3.0 μmol). The reaction was stirred for 10 minutes at room temperature and then purified by preparative HPLC (5mL/min, 10-100% acetonitrile in 0.1% TFA in H_2O). The fractions containing the product were lyophilized to afford **3** as a yellowish liquid (0.4 mg, 27%). [HRMS (ESI): 990.4 m/z (990.6 m/z expected) ^{1}H NMR (400 MHz, DMSO) δ 7.74 (m, 2 H), 7.52 (m, 1 H), 6.44 (s, 1 H), 6.34 (s, 1 H), 5.83 (s, 1 H), 4.88 (m, 3 H), 4.32 (m, 2 H), 4.15 (m, 2 H), 3.77 (m, 1 H), 3.60 (m, 1 H), 3.52 (m, 2 H), 3.47 (m, 2 H), 3.44-3.2 (30 H), 3.08 (m, 2 H), 2.84 (m, 1 H), 2.81 (m, 1 H), 2.60 (m, 1 H), 2.57 (m, 2 H), 2.45 (m, 1 H), 2.05 (m, 2 H), 1.74 (m, 2 H), 1.61 (m, 4 H), 1.44 (m, 3 H), 1.25 (m, 2 H), 0.87 (s, 2 H), 0.66 (s, 2 H)]

Synthesis of Alexa488-PEG$_3$-NH$_2$ (5) Alexa Fluor 488 (**4**, 4.74 mg, 8.9 μmol) was dissolved in 100 μL of DMSO and treated with DIEA (3.1 μL, 17.8 μmol), followed by TSTU (3.22 mg, 10.7 μmol). The reaction was stirred at room temperature for 10 minutes. 4,7,10-Trioxa-1,13-tridecanediamine (3.92 mg, 17.8 μmol) was dissolved in 100 μL of dry DMSO and the Alexa 488 reaction mixture was added drop wise under vigorous stirring over 5 minutes. The clear orange solution was stirred for 10 minutes at room temperature and then purified by preparative HPLC (5mL/min, 10-100% acetonitrile in 0.1% TFA in H_2O). The fractions containing the product were lyophilized to afford **5** as a deep red liquid (2.8 mg, 43%). [HRMS (ESI): 738.3 m/z (738.7 m/z expected). 1H NMR (400 MHz, DMSO) δ 8.74 (m, 1 H), 8.62 (m, 1 H), 8.26 (m, 1 H), 7.62 (m, 2 H), 7.26 (m, 1 H), 6.86 (m, 3 H), 3.54 (m, 4 H), 3.48 (m, 2 H), 3.3-3.4 (6H), 2.83 (m, 2 H), 2.08 (d, 1 H, J = 0.7 Hz), 1.84 (m, 2 H), 1.73 (m, 2 H), 1.25 (m, 1 H), 1.10 (t, 4 H, J = 7.0 Hz)]

Synthesis of Digoxigenin-PEG$_3$-Alexa488 (6) 5 (0.56 mg, 0.76 μmol) was dissolved in 200 μL of DMSO and treated with DIEA (0.20 mg, 1.52 μmol). Digoxigenin-NHS ester (0.5 mg, 0.8 μmol) was added at once and the reaction stirred for 10 minutes at room temperature and then purified by preparative HPLC (5mL/min, 10-100% acetonitrile in 0.1% TFA in H_2O). The fractions containing the product were lyophilized to afford **6** as a deep red liquid (0.59 μmol, 78%). [^{1}H NMR (400 MHz, DMSO)

δ 8.87 (s, 1 H), 8.69 (s, 3 H), 8.27 (dd, 1 H, J = 7.9, 1.3 Hz), 7.74 (m, 1 H), 7.54 (m, 2 H), 7.00 (dd, 4 H, J = 3.2, 1.6 Hz), 5.81 (s, 1 H), 4.86 (m, 3 H), 3.8-3.5 (31 H), 3.37 (m, 3 H), 3.22 (m, 2 H), 3.16 (s, 2 H), 3.06 (m, 2 H), 2.08 (s, 4 H), 2.02 (t, 2 H, J = 7.3 Hz), 1.81 (d, 2 H, J = 6.5 Hz), 1.73 (m, 2 H), 1.59 (m, 4 H), 1.43 (m, 7 H), 1.19 (d, 3 H, J = 6.2 Hz), 1.07 (m, 2 H), 0.86 (s, 2 H), 0.64 (s, 2 H).]

Gene synthesis. Designs DIG1-17, DigA16, and 3hk4 were ordered from Genscript (Piscataway, NJ) between the *NdeI* and *XhoI* restriction sites of a custom pET29-based vector having an N-terminal FLAG tag and a C-terminal His_6 tag (pET29FLAG). Codon usage was optimized for both *E. coli* and yeast with preference given to *E. coli*. DNA sequences are given in Supplementary Table 1.

Yeast surface display assays. Designed proteins were tested for binding using yeast-surface display[23]. Designs DIG1-17/pET29FLAG, DigA16/pET29FLAG, and 3hk4/pET29FLAG were subcloned into the *NdeI*/*XhoI* cloning sites of pETCON[24]. Designs and control proteins in pETCON were transformed into EBY100 cells using lithium acetate and polyethylene glycol[25] with dH_2O instead of single stranded carrier DNA and were plated on selective media (C *–ura –trp*). Freshly transformed cells were inoculated into 1 mL of SDCAA media[23] and grown at 30 °C, 200 rpm. After ~12 hrs, 1e7 cells were collected by centrifugation at 1,700 x *g* for 3 min and resuspended in 1 mL of SGCAA media to induce protein expression. Following induction for 24-48 hrs at 18 °C, 4e6 cells were collected by centrifugation, and washed twice by incubation with PBSF (PBS supplemented with 1 g/L of BSA) for 10 min at room temperature.

Yeast surface protein expression was monitored by binding of anti-cmyc FITC (Miltenyi Biotec GmbH, Germany) to the C-terminal myc epitope tag of the displayed protein. DIG binding was assessed by quantifying the phycoerythrin (PE) fluorescence of the displaying yeast population following incubation with DIG-BSA-biotin, DIG-RNase-biotin, or DIG-PEG_3-biotin, and streptavidin-phycoerythrin (SAPE; Invitrogen, Carlsbad, CA). In a typical experiment using DIG-BSA-biotin or DIG-RNase-biotin, 4e6 cells were resuspended in 50 μL of a premixed solution of PBSF containing a 1:100 dilution of anti-cmyc FITC, 2.66 μM DIG-BSA-biotin or DIG-RNase biotin, and 664 nM SAPE. Following a 2-4 hr incubation at 4 °C in the dark on a rotator, cells were collected by centrifugation at 1,700 x *g* for 3 min and washed with 200 μL of PBSF at 4 °C. Cell pellets were resuspended in 200 μL of ice-cold PBSF immediately before use. Cellular fluorescence was monitored on an Accuri C6 flow

cytometer using a 488 nm laser for excitation and a 575 nm band pass filter for emission. Phycoerythrin fluorescence was compensated to minimize bleed-over contributions from the FITC fluorescence channel.

Two positive controls having different affinities for digoxigenin were used to assess the binding assay: DigA16[26], and a commercially available anti-DIG monoclonal antibody 9H27L19 (Life Technologies). Experiments using DigA16 were conducted in an identical fashion to designs DIG1-17. For those employing the DIG antibody, two tandem Z domains of protein A (ZZ domain)[27,28], were displayed on the yeast cell surface. Washed cells were resuspended in 20 μL of PBSF with 2 μL of rabbit anti-DIG mAB 9H27L19 (Invitrogen, Carlsbad, CA). Following a 30-min incubation at 4 °C on a rotator, excess antibody was removed by washing the cells with 200 μL of PBSF. Labeling reactions were then performed as above. Negative controls for binding were the ZZ domain without mAB and an orthogonal gp120-based library available in the Baker lab (S2). FlowJo software version 7.6 was used to analyze all flow cytometry data presented here.

Competition assays with free digoxigenin were performed as above except that between 750 μM and 1.5 mM of digoxigenin (Sigma Aldrich, St. Louis, MO) prepared as a stock solution in MeOH was added to each labeling reaction mixture. Control experiments performed in a similar manner showed that the small amount of MeOH added does not affect the fluorescence or binding properties of SAPE (data not shown).

Knockout mutations. Knockout mutations were introduced into the appropriate DIG design in pETCON or pET29b(+) by the method of Kunkel[29]. These variants included the single point mutants V117R, Y101F, Y115F, and Y34F and the triple mutant Y101F/Y115F/Y34F for DIG10, the single point mutants W119R, H58A, Y84F, and Y97F and the double mutant Y115F/Y84F for DIG5, the single point mutants V86R, H101A and the triple mutant Y10F/H101A/Y103F for DIG8, and single point mutants Y34F, Y101F, Y115F, the double mutant Y99F/Y101F, the triple mutant Y34F/Y99F/Y101F, and the quadruple mutant Y34F/Y99F/Y101F/Y115F for DIG10.3. Oligos were ordered from Integrated DNA Technologies, Inc. (Coralville, IA) and are listed in Supplementary Table 13 with the mutagenized region(s) highlighted in red.

Recursive PCR assembly of 1z1s. The gene for 1z1s having additional pETCON overlap fragments at either end for yeast homologous recombination was assembled via recursive PCR. Oligo sequences were designed using DNAWorks[30] and are given in Supplementary Table 14. Oligos were ordered from

Integrated DNA Technologies, Inc. (Coralville, IA). A 2 μL portion of a 2.5 μM stock solution of each oligo was combined and the mixture was added to 8 μL of 1.25 mM dNTPs, 20 μL of 5X Phusion buffer HF, 3 μL of DMSO, and 1 μL of Phusion high-fidelity polymerase (NEB, Waltham, MA) in 100 μL. Full-length gene product was assembled by 30 cycles of PCR (98 °C 10 s, 61 °C 30 s, 72 °C 15 s).

Correctly assembled PCR product was amplified by a second round of PCR. Reaction product (5 μL) was combined with 2 μL of 10 μM pCTCON2f (Supplementary Table 15), 2 μL of 10 μM pCTCON2r (Supplementary Table 15), 8 μL of 1.25 mM dNTPs, 20 μL of 5X Phusion buffer HF, 3 μL of DMSO, and 1 μL of Phusion high-fidelity polymerase (NEB, Waltham, MA) in 100 μL. Product was obtained by 30 cycles of PCR (98 °C 10 s, 60 °C 30 s, 72 °C 15 s). Following confirmation of a single band at the correct molecular weight by 1% agarose gel electrophoresis, the PCR product was purified using a Qiagen PCR cleanup kit (Qiagen) and eluted in dH_2O.

Yeast EBY100 cells were transformed with 240 ng of 1z1s gene DNA and 400 ng of gel-purified pETCON digested with *Nde1* and *Xho1* using lithium acetate and polyethylene glycol[25] with dH_2O instead of single-stranded carrier DNA. The correct sequence was confirmed by colony PCR and gene sequencing, and plasmids from these colonies were harvested using a Zymoprep Yeast Miniprep II kit (Zymo Research Corporation, Irvine, CA).

DIG10 site-saturation mutagenesis library (directed evolution round 1a). A DIG10 single site-saturation mutagenesis (SSM) library was constructed by Kunkel mutagenesis[29] using degenerate NNK primers targeting the following 34 amino acids positions: S10, L11, L14, W22, L32, Y34, A37, P38, G40, H41, H54, M55, L57, F58, Y61, V62, V64, F66, F84, G86, G88, H90, V92, S93, L97, A99, Y101, S103, Y115, V117, F119, V124, A127, and L128. These positions were chosen from the model based on the following requirements: (1) they have Cα within 7 Å of any ligand heavy atom, and/or (2) they have Cα within 9 Å of any ligand heavy atom and Cβ closer to any heavy atom in the ligand than Cα. The theoretical library size was 1088 clones. Primers were ordered from Integrated DNA Technologies (Coralville, IA) and are listed in Supplementary Table 16 with the mutagenized region highlighted in red.

Kunkel mutagenesis of each position was carried out independently. DNA from each reaction was dialyzed into dH_2O using a 0.025 μm membrane filter (Millipore, Billerica, MA), and then the dialyzed reaction mixtures were pooled, concentrated to a volume of < 10 μL using a Savant SpeedVac centrifugal vacuum concentrator, and transformed into yeast strain EBY100 using the method of Benatuil[31], yielding 2.5e5 transformants. After transformation, cells were grown in 250 mL of SDCAA media for 36 hrs at 30

°C. Cells (5e8) were collected by centrifugation at 1,700 x *g* for 4 min, resuspended in 50 mL of SGCAA media, and induced at 18 °C for 24 hrs.

Cells were subjected to three rounds of permissive cell sorting (Supplementary Table 9). For each round of sorting, cells were washed and then labeled with a pre-incubated mixture of 2.66 μM DIG-BSA-biotin, 644 nM SAPE, and anti-cmyc-FITC as noted above for single clones. After each sort, cells were grown in SDCAA for 24 hrs and then induced in SGCAA for 24 hrs before the next sort. After the final sort, the mean compensated PE fluorescence of the expressing population of the sorted cells was considerably higher than that of DIG10, indicating the presence of a point mutant(s) with increased binding affinity.

After each sort, a portion of cells were plated and grown at 30 ºC. Plasmids from individual colonies were harvested using a Zymoprep Yeast Miniprep II kit (Zymo Research Corporation, Irvine, CA) and the gene was amplified by 30 cycles of PCR (98 °C 10 s, 61 °C 30 s, 72 °C 15 s) using Phusion high-fidelity polymerase (NEB, Waltham, MA) with the pCTCON2r and pCTCON2f primers (Supplementary Table 15). Sanger sequencing (Genewiz, Inc., South Plainfield, NJ) was used to sequence at least 10 colonies from each population.

DIG10 combinatorial mutagenesis library (directed evolution round 1b). Beneficial mutations identified in the DIG10 SSM library were combined by Kunkel mutagenesis[29] using degenerate primers. At each mutagenized position, the original DIG10 amino acid and chemically similar amino acids to those identified were also allowed, resulting in a combinatorial library. Amino acid substitutions included S, A, or M at position S10, L, H, or Q at position L11, A or P at position A37, I, L, V, F, or M at position V62, I, L, V, F, or M at position V64, H, T, or N at position H90, I, L, V, F, or M at position V117, and A or P at position A127. The theoretical library size was 1.35e4 clones. Primers were ordered from Integrated DNA Technologies, Inc. (Coralville, IA) and are listed in Supplementary Table 17 with the mutagenized region(s) highlighted in red.

Four independent Kunkel reactions using different oligo concentrations ranging from 36 nM to 291 nM during polymerization were performed to minimize sequence-dependent priming bias. For the same reason, oligos encoding native substitutions contained at least one codon base change. Library DNA was pooled, prepared as above, and transformed into electrocompetent *E. coli* strain BL21(DE3) cells (1800 V, 200 Ω, 25 μF), yielding 8e4 transformants. Library plasmid DNA was isolated from expanded cultures using a Qiagen miniprep kit. Gene insert was amplified from 10 ng of library DNA by 30 cycles of PCR

(98 °C 10 s, 61 °C 30 s, 72 °C 15 s) using Phusion high-fidelity polymerase (NEB, Waltham, MA) with the pCTCON2r and pCTCON2f primers (Supplementary Table 15).

Yeast EBY100 cells were transformed with 4.0 μg of PCR-purified DNA insert and 1.0 μg of gel-purified pETCON digested with *Nde1* and *Xho1* using the method of Benatuil[31], yielding 8e5 transformants. After transformation, cells were grown in 150 mL of low pH SDCAA media supplemented with Pen/Strep for 48 hrs at 30 °C. Cells (5e8) were collected by centrifugation at 1,700 x *g* for 4 min, resuspended in 50 mL of SGCAA media, and induced at 18 °C for 24 hrs.

Cells were subjected to seven rounds of cell sorting (Supplementary Table 9). For the first four rounds, cells were washed and then labeled with a pre-incubated mixture of DIG-BSA-biotin, SAPE, and anti-cmyc-FITC as noted above for single clones. Label concentrations for rounds one through four were: (1) 1 μM DIG-BSA-biotin and 250 nM SAPE, (2) 750 nM DIG-BSA-biotin and 187.5 nM SAPE, (3) 50 nM DIG-BSA-biotin and 12.5 nM SAPE, and (4) 5 nM DIG-BSA-biotin and 1.25 nM SAPE. For rounds five through seven, DIG-RNase-biotin was used in a multistep labeling procedure to minimize selection for carrier protein (BSA) binding and because this procedure showed a larger dynamic range in several control experiments. In these experiments, cells were washed as before, labeled with DIG-RNase-biotin for 3-4 hrs at 4 °C, and then treated with a solution of PBSF containing a 1:100 dilution of SAPE and a 1:100 dilution of anti-cmyc-FITC (secondary label) for < 15 min at 4 °C before washing and sorting. DIG-RNase-biotin label concentrations were 10 pM, 5 pM, and 5 pM for rounds five through seven, respectively.

At least 10 clones from each round were sequenced as noted for the DIG10 SSM library. After seven rounds, the library converged to two sequences differing by a single point substitution: DIG10.1, harboring the S10A substitution, and DIG10.1b, containing S10M (Supplementary Figs 6 and 7). Mutations common to both DIG10.1 and DIG10.1b were V62M, V64I, V117L, and A127P. Analysis of both clones using the multistep labeling procedure with 5 pM DIG-RNase-biotin showed that DIG10.1 had a slightly higher signal for mean PE fluorescence of the expressing population than did DIG10.1b.

DIG10.1L library (directed evolution round 2). Library DNA was a mixture of DNA from DIG10.IL_f1, DIG10.IL_f2, and a third library (DIG10.IL_3) combining mutations from the two fragment libraries (see section on Next-Gen Library Construction). For DIG10.1L_3, the library was constructed using the oligos DIG10.1L_hr1, DIG10.1L_f1a_rc_variable, DIG10.1L_f1b_variable, DIG10.1L_f2_rc_variable, and DIG10.1L_hr2 and the procedures detailed below.

Yeast EBY100 cells were transformed with a mixture of DNA insert from DIG10.IL_f1 (3.0 μg), DIG10.1L_f2 (3.0 μg), and DIG10.IL_3 (24.0 μg) and 10.0 μg of gel-purified pETCON digested with *Nde1* and *Xho1* using the method of Benatuil[31], yielding 1.5e7 transformants. After transformation, cells were grown for 24 hrs in 250 mL of low-pH SDCAA media supplemented with Pen/Strep at 30 °C, passaged once, and grown for an additional 24 hrs under the same conditions. Cells (5e8) were collected by centrifugation, resuspended in 50 mL of SGCAA, and induced overnight at 18 °C.

Cells were subjected to five rounds of cell sorting using monovalent DIG-PEG_3-biotin and the multistep labeling procedure detailed for directed evolution round 1b sorts five through seven to increase stringency by avoiding avidity effects (Supplementary Fig. 11 and Supplementary Table 9). DIG-PEG_3-biotin label concentrations were 80 nM, 80 nM, 50 nM, 1 nM, and 1 nM for the five rounds. After the final sort, the mean compensated PE fluorescence of the expressing population of the sorted cells was considerably higher than that of DIG10.2, indicating the presence of a mutant(s) with increased binding affinity.

At least 10 clones from each round were sequenced as noted for the DIG10 SSM library. After four rounds, the library converged to one sequence (DIG10.2) having the two loop mutations A37P and H41Y (Supplementary Fig. 6), which were two of the most enriched single point mutations identified in the next-generation sequencing experiment.

DIG10.2 combinatorial library based on deep sequencing data (directed evolution round 3). Mutations having normalized next-generation sequencing enrichment values (ΔE_i^x) > ~3.5 were combined by Kunkel mutagenesis[29] using degenerate primers. DIG10.2 was used as the library template. At each mutagenized position, the original DIG10.2 amino acid and chemically similar amino acids to those identified were also allowed, resulting in a combinatorial library. Amino acid substitutions included C or S at position C23, F, H, or Y at position F45, M or F at position M62, H, I, L, F, or Y at position H90, V or A at position V92, A, V, I, T, F, or Y at position A99, S, A, or V at position S103, L, V, or W at position L105, I or F at position I112, V or F at position V124, and P, I, L, or V at position P127. The theoretical library size was 1.04e5 clones. Primers were ordered from Integrated DNA Technologies (Coralville, IA) and are listed in Supplementary Table 18 with the mutagenized region(s) highlighted in red.

Four Kunkel reactions using different oligo concentrations ranging from 36 nM to 291 nM during polymerization and two Kunkel reactions using reduced oligo concentrations for the M62M substitution

relative to the concentrations of the M62F substitution were performed to minimize sequence-dependent priming bias. For the same reason, oligos encoding native substitutions contained at least one codon base change. Library DNA was pooled, prepared as above, and transformed into *E. coli* strain ElectroMAX DH10B (Invitrogen, Carlsbad, CA) cells (2500 V, 200 Ω, 25 μF), yielding 1.6e7 transformants. Library plasmid DNA was isolated from expanded cultures using a Qiagen miniprep kit. Gene insert was amplified from 10 ng of library DNA by 30 cycles of PCR (98 °C 10 s, 61 °C 30 s, 72 °C 15 s) using Phusion high-fidelity polymerase (NEB, Waltham, MA) with the pCTCON2r and pCTCON2f primers (Supplementary Table 15).

Yeast EBY100 cells were transformed with 6.0 μg of PCR-purified DNA insert and 2.0 μg of gel-purified pETCON digested with *Nde1* and *Xho1* using the method of Benatuil[31], yielding 5e6 transformants. After transformation, cells were grown in 150 mL of low pH SDCAA media supplemented with Pen/Strep for 48 hrs at 30 °C. Cells (5e8) were collected by centrifugation at 1,700 x *g* for 4 min and resuspended in 50 mL of SGCAA media. Cells were induced at 18 °C for 24 hrs.

Cells were subjected to four rounds of cell sorting (Supplementary Fig. 12 and Supplementary Table 9). The first three sorts utilized monovalent DIG-PEG_3-biotin and the multistep labeling procedure detailed for directed evolution round 1b sorts five through seven to increase stringency by avoiding avidity effects. DIG-PEG_3-biotin label concentrations were 400 pM, 20 pM, and 20 pM for the first three rounds. For the fourth round, an off-rate selection[24] was used to better discriminate between high affinity binders. Cells (4e6) were washed and labeled with 20 pM DIG-PEG_3-biotin, as described above. Labeled cells were collected by centrifugation at 1,700 x *g* for 4 min and resuspended in 100 μL of 100 nM DIG in PBSF. Cells were incubated with free DIG for 20 min at room temperature (20 min was found to be the half-life of the DIG10.2–DIG-PEG_3-biotin complex in off-rate experiments) collected by centrifugation, labeled with secondary label as described above, washed, and sorted. After the final sort, the mean compensated PE fluorescence of the expressing population of the sorted cells was considerably higher than that of DIG10.2, indicating the presence of a mutant(s) with increased binding affinity.

At least 10 clones from each round were sequenced as noted for the DIG10 SSM library. After four rounds, the library converged to one sequence (DIG10.3) having the mutations C23S, H90L, V92A, A99Y, S103A, and L105W (Supplementary Fig. 6).

Next-generation DIG10.1 library construction and selections. Paired-end 151 Illumina sequencing was used to simultaneously assess the effects of mutation on binding of DIG10.1 to digoxigenin at 39

amino acid positions within the binding site pocket. Two libraries were constructed: an N-terminal library with mutations between residues S10 and F66 (fragment 1 library - DIG10.IL_f1) and a C-terminal library with mutations between residues F84 and L128 (fragment 2 library - DIG10.IL_f2). For each library, the full-length DIG10.1 gene having additional pETCON overlap fragments at either end for yeast homologous recombination was assembled via recursive PCR. To introduce mutations, we used degenerate PAGE-purified oligos in which selected positions within the binding site were doped with a small amount of each non-native base at a level expected to yield 1-2 mutations per gene (TriLink BioTechnologies, San Diego, CA). All other wild-type oligos were also PAGE-purified (Integrated DNA Technologies). Oligos and doping ratios are given in Supplementary Table 10. For DIG10.IL_f1, bases coding for the following 20 amino acid positions were allowed to vary: A10, L11, L14, W22, C23, F26, L32, Y34, A37, P38, G40, H41, F45, H54, M55, F58, Y61, M62, I64, and F66. For DIG10.1L_f2, bases coding for the following 19 amino acid positions were allowed to vary: F84, G86, G88, H90, V92, S93, G95, L97, A99, Y101, S103, L105, I112, Y115, L117, F119, V124, P127, and L128.

For assembly of DIG10.1L_f1, 2 μL of 2.5 μM DIG10.1L_hr1, 2 μL of 2.5 μM DIG10.1L_f1a_rc_variable, 2 μL of 2.5 μM DIG10.1L_f1b_variable, 2 μL of 2.5 μM DIG10.IL_f2_rc_WT, and 2 μL of 2.5 μM DIG10.1L_hr2 were combined with 8 μL of 1.25 mM dNTPs, 20 μL of 5X Phusion buffer HF, 3 μL of DMSO, and 1 μL of Phusion high-fidelity polymerase (NEB, Waltham, MA) in 100 μL. Reaction mixtures for assembly of DIG10.1l_f2 were the same, except that DIG10.1L_f1a_rc_variable, DIG10.1L_f1b_variable, and DIG10.1L_f2_rc_WT were substituted with DIG10.1L_f1a_rc_WT, DIG10.1L_f1b_WT, and DIG10.1L_f2_rc_variable, respectively. Full-length products were assembled by 30 cycles of PCR (98 °C 10 s, 61 °C 30 s, 72 °C 15 s).

Correctly assembled PCR products were amplified by a second round of PCR. Reaction products (5 μL) were combined with 2 μL of 10 μM DIG10.IL_assembly_fwd, 2 μL of 10 μM DIG10.IL_assembly_rev, 8 μL of 1.25 mM dNTPs, 20 μL of 5X Phusion buffer HF, 3 μL of DMSO, and 1 μL of Phusion high-fidelity polymerase (NEB, Waltham, MA) in 100 μL. Products were amplified by 30 cycles of PCR (98 °C 10 s, 60 °C 30 s, 72 °C 15 s). Following confirmation of a single band at the correct molecular weight by 1% agarose gel electrophoresis, PCR products were purified using a Qiagen PCR cleanup kit (Qiagen) and eluted in ddH_2O.

Yeast EBY100 cells were transformed with 5.4 μg of library DNA insert and 1.8 μg of gel-purified pETCON digested with *Nde1* and *Xho1* using the method of Benatuil[31], yielding 4e6 and 3e6 transformants for the DIG10.1L_f1 and DIG10.1L_f2 libraries, respectively. After transformation, cells

were grown for 24 hrs in 100 mL of low-pH SDCAA media supplemented with Pen/Strep at 30 °C, passaged once, and grown for an additional 24 hrs under the same conditions. Cells (5e8) were collected by centrifugation, resuspended in 50 mL of SGCAA, and induced overnight at 18 °C.

Induced cells (3e7) were labeled with 4 μL of anti-cymc-FITC (Miltenyi Biotec GmbH, Germany) in 200 μL of PBSF for 20 min (DIG10.1L_f1) or 60 min (DIG10.1L_f2) at 4 °C. Then, labeled cells were washed with PBSF and sorted. In this first round of sorting, all cells showing a positive signal for protein expression were collected (Supplementary Table 9 and Supplementary Fig. 9). Cells were recovered overnight in ~1 mL of low-pH SDCAA supplemented with Pen/strep at 30 °C, pelleted by centrifugation at 1,700 x *g* for 4 min, resupended in 5 mL of low-pH SDCAA supplemented with Pen/strep, and grown for an additional 24 hrs at 30 °C. Cells (2e7) were collected by centrifugation, resuspended in 2 mL of SGCAA, and induced overnight at 18 °C.

Induced cells from expression-sorted DIG10.1L_f1 (2e7 cells), expression-sorted DIG10.1L_f2 (2e7 cells), and two DIG10.1 reference samples (5e6 cells per sample) were washed with 600 μL of PBSF and then labeled with 100 nM of DIG-PEG_3-biotin in 400 μL of PBSF for the libraries or 200 μL of PBSF for the reference samples for > 3 hrs at 4 °C. Labeled cells were washed with 200 μL of PBSF, then incubated with a secondary label solution of 0.8 μL of SAPE (Invitrogen) and 4 μL of anti-cymc-FITC (Miltenyi Biotec GmbH, Germany) in 400 μL of PBSF for 8 min at 4 °C. Cells were washed with 200 μL PBSF, resuspended in either 800 μL μL of PBSF for the libraries or 400 μL of PBSF for the reference samples, and sorted (Supplementary Fig. 9 and Supplementary Table 9). Each library was sorted according to two different stringency conditions: (1) clones having binding signals higher than that of DIG10.1 (DIG10.1_f1_better and DIG10.1_f2_better), and (2) clones having binding signals equivalent to or higher than that of DIG10.1 (DIG10.1_f1_neutral and DIG10.1_f2_neutral). Collected cells were recovered overnight in ~1 mL of low-pH SDCAA supplemented with Pen/strep at 30 °C, pelleted by centrifugation at 1,700 x *g* for 4 min, resupended in 2 mL of low-pH SDCAA supplemented with Pen/strep, and grown for an additional 24 hrs at 30 °C. Cells (2e7) were resuspended in 2 mL of SGCAA and induced overnight at 18 °C.

To reduce noise from the first round of cell sorting, the sorted libraries were labeled and subjected to a second round of cell sorting using the same conditions and gates as in the first round (Supplementary Fig. 9 and Supplementary Table 9). Collected cells were recovered overnight in 800 μL of low-pH SDCAA supplemented with Pen/strep at 30 °C, pelleted by centrifugation at 1,700 x *g* for 4 min,

resupended in 2 mL of low-pH SDCAA supplemented with Pen/strep, and grown for an additional 24 hrs at 30 °C.

One hundred million cells from the expression-sorted DIG10.1L_f2 and DIG10.1L_f2 libraries and at least 2e7 cells from doubly-sorted DIG10.1_f1_better and DIG10.1_f2_better were pelleted by centrifugation at 1,700 x *g* for 4 min, resuspended in 1 mL of freezing solution (50% YPD, 2.5% glycerol), transferred to cryogenic vials, slow-frozen in an isopropanol bath, and stored at -80 °C until further use.

Next-generation library sequencing. Library DNA was prepared as detailed previously[32]. Illumina adapter sequences and unique library barcodes were appended to each library pool through PCR amplification using population-specific HPLC-purified primers (Integrated DNA Technologies, Coralville, IA) given in Supplementary Table 11. The library amplicons were verified on a 2% agarose gel stained with SYBR Gold (Invitrogen) and then purified using an Agencourt AMPure XP bead-based purification kit. (Beckman Coulter, Inc.) Each library amplicon was denatured using NaOH and then diluted to 6 pM. A sample of *PhiX* control DNA (Illumina, Inc., San Diego, CA) was prepared in the same manner as the library samples and added to the library DNA to create high enough sample diversity for the Illumina base-calling algorithm. The final DNA sample was prepared by pooling 300 μL of 6 pM *PhiX* control DNA (50%), 102 μL of 6 pM expression-sorted DIG10_1L_f1 (17.0%), 102 μL of 6 pM expression-sorted DIG10_1L_f2 (17.0%), 33 μL of 6 pM DIG10_1L_f1_neutral (5.5%), 33 μL of 6 pM DIG10_1L_f2_neutral (5.5%), 15 μL of 6 pM DIG10_1L_f1_better (2.5%), and 15 μL of 6 pM DIG10_1L_f2_better (2.5%). DNA was sequenced in paired-end mode on an Illumina MiSeq using a 300-cycle reagent kit and custom HPLC-purified primers (Integrated DNA Technologies, Inc., Coralville, IA) given in Supplementary Table 11.

Processing of sequencing results. Data from each next-generation sequencing library was demultiplexed using the unique library barcodes added during the amplification steps. Of a total 5,630,105 paired-end reads, 2,531,653 reads were mapped to library barcodes (Supplementary Table 12). For each library, paired end reads were fused and filtered for quality (Phred ≥ 30). The resulting full-length reads were aligned against the relevant segments of the DIG10.1 sequence using scripts from the software package Enrich[33]. The counts for each codon encoding a given amino acid substitution were summed to get the total count number for that mutation. For single mutations having ≥ 7 counts in the original input library

(see below), a relative enrichment ratio between the input library and each selected library was calculated[32,34,35]. A pseudocount value of 0.3 was added to the total reads for each selected library mutation to allow calculation of enrichment values for mutations that disappeared completely during selection. Sequencing counts for single point mutations observed in the input and selected libraries are given in Supplementary Tables 19 and 20, respectively.

To determine the minimum number of counts in the original input library necessary to be confident that the calculated enrichment values are statistically significant, we reasoned that there should be a negative correlation between initial read counts and enrichment values for mutations with a low number of initial reads (*e.g.* because there were ~4-times as many reads for the unselected library than the selected library and because of the applied psuedocount, a mutation with one count in the unselected library will always appear to have a positive enrichment value, even when it is not observed in the selected pool)[36]. A sliding window approach was applied in which the coefficient for the correlation between initial read count and enrichment value was calculated over contiguous 10-count windows (1-10, 2-11, 3-12, etc.). The cutoff was defined as the value at the bottom of the window (e.g. 1 for the 1-10 window) in which the correlation was no longer significant ($p > 0.1$), and was found to be 7 in this case (Supplementary Fig. 10c).

Protein expression and purification. Selected DIG designs and variants were expressed in *E. coli* in pET29FLAG or with a TEV protease-cleavable His_6 purification tag (pET29-TEV-His_6). For the latter, DIG genes were amplified from the appropriate pETCON-based plasmid using a forward primer and a reverse primer harboring a TEV-protease recognition insertion sequence (Supplementary Table 15). The PCR products were digested with *Nde1* and *XhoI* and ligated into similarly digested pET29b(+). Ligation products were transformed into Rosetta 2 (DE3) cells for expression. Rosetta 2 (DE3)/pET29b(+) cells were grown in 1L of LB or TB medium at 37 °C to an $O.D._{600}$ of ~0.7, and then protein expression was induced by the addition of 0.5 mM IPTG (isopropyl-β-D-thiogalactopyransoide). Cultures were incubated at 37 °C for 3-4 hrs or at 18 °C for 18 hrs and then harvested by centrifugation at 1,912 x *g* for 20 min. Cell pellets were stored at -20 °C until further use.

Proteins were purified by gravity flow chromatography over Ni-NTA resin (Qiagen, Hilden, Germany) columns. Frozen cell pellets were resuspended in 15 mL of wash buffer (PBS pH 7.4, 30 mM imidazole) supplemented with 300 μL of 100 mM phenylmethanesulfonyl fluoride (PBSF; Sigma Aldrich, St. Louis, MO) prepared in neat ethanol, 2 mg/mL of lysozyme, and 0.2 mg/mL of DNAse I.

Cells were lysed by sonication for a total of 4 min (30 s on, 20 s off) using a Branson sonifier at 75% power. Insoluble material was removed by centrifugation at 38,724 x *g* for 30 min, and particulate matter was further removed from the supernatant by filtration through a 0.45 μm syringe filter. Supernatant was then passed through gravity columns containing 3 mL of Ni-NTA resin (Qiagen, Hilden, Germany) equilibrated in wash buffer. Bound proteins were washed with 45 mL of wash buffer and then eluted in 20 mL of elution buffer (PBS pH 7.4, 200 mM imidazole). Proteins were concentrated to ~5-40 mg/mL using Vivaspin 5 kD MWCO centrifugal concentration devices (Sartorium Stedim Biotech GmbH, Goettingen, Germany) and imidazole was removed by dialysis (3 x 2L) into PBS pH 7.4 at 4 °C.

Yields for the DIG designs expressed in pET29FLAG are given in Supplementary Table 2. Typical yields for DIG10-TEV-his_6, DIG10.1-TEV-his_6, DIG10.2-TEV-his_6, DIG10.3-TEV-his_6, DIG10.3-TEV-his_6 variants, and 1z1s-TEV-his_6 range from 10 to 60 mg/L. For all solution experiments, protein concentrations were determined from absorbance at 280 nm measured on a NanoDrop spectrophotometer (Thermo Scientific) using extinction coefficients calculated from primary amino acid sequences (Supplementary Table 21).

Size-exclusion chromatography. Protein oligomerization states were assessed by size exclusion chromatography on an ÄKTA FPLC (GE Healthcare) using a Superdex 75 10/300 GL column equilibrated in running buffer (25 mM Tris-HCl pH 7.4, 250 mM NaCl). Proteins were run over the column at a flow rate of 0.5 mL/min. Horse heart cytochrome c (29 kDa), bovine erythrocytes carbonic anhydrase (12.4 kDa), and bovine aprotinin (6.5 kDa) molecular weight standards (Sigma Aldrich, St. Louis, MO) were analyzed in the same manner as the protein samples. Under these conditions, cytochrome c, carbonic anhydrase, DIG10-TEV-his_6 (expected MW of monomer: 17.9 kDa), DIG10.1b-TEV-his_6 (expected MW of monomer: 17.9 kDa), DIG10.2_t-his_6 (see below; expected MW of monomer: 15.9 kDa), DIG10.3-TEV (the his_6 tag was cleaved with TEV protease; expected MW of monomer: 16.9 kDa), and DIG5-TEV-his_6 (expected MW of monomer: 17.8 kDa) eluted at 12.05 mL, 13.65 mL, 11.88 mL, 11.81 mL, 11.40 mL, 11.35 mL, and 11.78 mL, respectively (Supplementary Fig. 15). For preparative runs, pure protein-containing fractions were identified by absorbance at 280 nm and by SDS gel electrophoresis. Analytical superdex 75 gel filtration analyses of 100 μM DIG10.3-TEV-his_6 and 100 μM DIG10.3-TEV-his_6 pre-incubated with 500 μM DIG for ~60 min at room temperature were also conducted using the above procedure. Under the conditions, DIG10.3 and the DIG10.3–DIG complex eluted at 11.29 mL and 10.71 mL, respectively (Supplementary Fig. 15).

Preparation of samples for crystallography. Crystallographic trials with the DIG10-based C-terminal TEV-his_6 constructs (cleaved with TEV protease or un-cleaved) failed to yield diffraction-quality crystals. All 1z1s-based designs contained a 12 residue C-terminal tail that was disordered in the structure of 1z1s but was maintained when we ordered the designs in case it was necessary for protein stability or folding. To reduce entropic effects from this disordered tail that might prevent crystal formation, we cloned the DIG10 designs into new pET29b(+)-based constructs in which all 12 residues of this tail were eliminated and a non-cleavable his_6 tag was placed immediately after the last ordered residue ($DIG10_t$-his_6, $DIG10.1_t$-his_6, $DIG10.2_t$-his_6, and $DIG10.3_t$-his_6). Sub-cloning primers are given in Supplementary Table 15. The amino acid sequences of these gene products are given in Supplementary Table 21.

Truncated samples were expressed and purified by gravity flow over Ni-NTA resin using the above procedure. Typical expression yields were comparable to their un-truncated, TEV-cleavable His_6-tagged counterparts (see above). Preparative size exclusion chromatography was used to further purify all proteins for crystallization attempts using the above procedure.

Crystallization. Purified DIG10 and its evolved variants were incubated at 4°C with 1 mM digoxigenin for 16-20 hours. The protein-ligand complex was then screened using several commercially available sparse matrix crystallization screens using a nanoliter drop volume crystallization robot (TTP LabTech 'Mosquito'). Potential hits were scaled up into vapor diffusion plates with reservoir solution to protein-ligand complex at a ratio of 1:1. Several diffraction quality crystals were obtained for $DIG10.2_t$-his_6 and $DIG10.3_t$-his_6. Crystals of $DIG10.2_t$-his_6 were grown at a concentration of 15 mg mL^{-1} in 0.1 M Acetate pH 5.5, 1.5% MPD, 2.5 M Sodium chloride and 12% PEG1500. Crystals of $DIG10.3_t$-his_6 were grown at a concentration of 13.5 mg mL^{-1} in 0.2 M Ammonium acetate, 0.1 M Bis-Tris pH 5.5 and 20% PEG3350. $DIG10.2_t$-his_6 and $DIG10.3_t$-his_6 crystals were transferred to artificial mother liquor containing 20% Sucrose or Glycerol, respectively, then individually removed in fiber loops and flash frozen in liquid nitrogen.

Crystallographic data collection and processing. Datasets from crystals of $DIG10.2_t$-his_6 and $DIG10.3_t$-his_6 were collected at the Advanced Light Source (ALS) synchrotron facility (Berkeley, CA) on beamline 5.0.2 using a CCD area detector. Data for $DIG10.2_t$-his_6 corresponded to 360° of 1° diffraction exposures collected at a distance of 180 mm and exposure times of 1 second per 1° oscillation. Data for

DIG10.3$_t$-his$_6$ corresponded to 360° of 1 ° diffraction exposures collected at a distance of 230 mm and exposure times of 1 second per 1° oscillation.

Data was processed using the HKL2000 software package[37]. Molecular replacement was performed using program PHASER[38] in the CCP4 software suite[39,40] using *Pseudomonas aeruginosa* hypothetical protein PA3332 (PDB 1Z1S) as the model[41]. Refinement and model building were carried out using Refmac5[42] and COOT (Crystallography Object-Oriented Toolkit)[43], respectively. DIG10.2$_t$-his$_6$ required extra refinement using simulated annealing in the PHENIX suite[44] to remove R-free reflections that overlapped with those used previously for validation of 1Z1S. The geometric quality of the final model was validated using ProCheck[45], SFCheck[46] and MolProbity[47], as well as the validation tools provided by the RCSB Protein Data Base[48].

The diffraction dataset collected from the DIG10.3$_t$-his$_6$ crystal collected could only be processed to 3.2Å resolution in space group C2. Significant disorder was displayed in several of the independent copies of protein-ligand complex in the asymmetric unit, which resulted in very high average B-factors. Data collection and refinement statistics for DIG10.2$_t$-his$_6$ and DIG10.3$_t$-his$_6$ are provided in Supplementary Table 22.

Fluorescence polarization equilibrium binding assays. Fluorescence polarization-based affinity measurements of designs and their evolved variants were performed as noted previously[49] using Alexa488-conjugated DIG (DIG-PEG$_3$-Alexa488). In a typical experiment, the concentration of DIG-PEG$_3$-Alexa488 was fixed below the K_d of the interaction being monitored and the effect of increasing concentrations of protein on the fluorescence anisotropy of Alexa488 was determined. Fluorescence anisotropy (r) was measured in 96-well plate format using Corning NBS black flat bottom 96 well plates on a SpectraMax M5e microplate reader (Molecular Devices) with λ_{ex} = 485 nM and λ_{em} = 538 nM using a 515 nm emission cutoff filter. In all experiments, PBS (pH 7.4) was used as the buffer system and the temperature was 25 °C. DIG-PEG$_3$-Alexa488 solutions were prepared from a 1 mM stock in DMSO. Equilibrium dissociation constants (K_d) were determined by fitting plots of the anisotropy averaged over a period of 20 to 40 min after reaction initiation versus protein concentration to eq 1, as described

$$A = A_f + (A_b - A_f) * \left(\frac{([L]_T + K_D + [R]_T) - \sqrt{(-[L]_T - K_D - [R]_T)^2 - 4[L]_T[R]_T}}{2[L_T]}\right) \qquad \text{eq 1}$$

previously[49], where A is the experimentally measured anisotropy, A_f is the anisotropy of the free ligand, A_b is the anisotropy of the fully bound ligand, $[L]_T$ is the total ligand concentration, and $[R]_T$ is the total receptor concentration. Reported K_d values represent the average of at least three independent measurements with at least two separate batches of purified protein.

Design-TEV-his$_6$ constructs were used for all measurements. The [DIG-PEG$_3$-Alexa488] used for sets of experiments on each protein are as follows: DIG5: 2 μM, DIG10: 2 μM, 1z1s: 2 μM, BSA: 2 μM, DIG10.1: 10 nM, DIG10.2: 1 nM, DIG10.3: 0.5 nM, DIG10.3 Y34F: 2 nM, DIG10.3 Y99F: 2 nM, DIG10.3 Y101F: 2 nM, DIG10.3 Y115: 2 nM, DIG10.3 Y99F/Y101F: 2 nM, DIG10.3 Y34F/Y99F/Y101F: 10 nM, and DIG10.3 Y34F/Y99F/Y101F/Y115F: 10 nM.

Fluorescence polarization equilibrium competition binding assays. Fluorescence polarization equilibrium competition binding assays were used to determine the binding affinities of DIG10.3 and its variants for unlabeled digoxigenin, digitoxigenin, progesterone, β-estradiol, and digoxin. In a typical experiment, the concentration of DIG-PEG$_3$-Alexa488 was kept near or below the K_d of the interaction being monitored, the concentration of protein was fixed at a saturating value such that >95% the DIG-PEG$_3$-Alexa488 in the system was bound to protein, and the effects of increasing concentrations of unlabeled ligand on the fluorescence anisotropy of Alexa488 were determined as noted above. Unlabeled stock solutions of digoxigenin, digitoxigenin, progesterone, and β-estradiol were prepared in methanol. Unlabeled stock solutions of digoxin were prepared in DMSO. Ligand stock solutions were 10 mM for DIG, digitoxigenin, and digoxin, and 1 mM for progesterone and β-estradiol. For each ligand concentration, a negative control sample containing only DIG-PEG$_3$-Alexa488 and the appropriate dilution of a corresponding methanol-only control solution in PBS was measured. At all concentrations employed, methanol did not affect fluorescence anisotropy (data not shown). Similarly, the highest concentration of DMSO employed also did not affect fluorescence anisotropy (data not shown).

Fluorescence anisotropy (r) was measured as noted above. In all experiments, PBS (pH 7.4) was used as the buffer system and the temperature was 25 °C. The concentration of total unlabeled ligand producing 50% binding signal inhibition (I_{50}) was determined by fitting a plot of the anisotropy averaged over a period of 30 min to 3 hr after reaction initiation versus unlabeled ligand concentration as described previously[49]. For some experiments, limiting steroid concentrations made it impossible to collect data in the regime of complete inhibition. In these cases, data were fit by fixing the anisotropy at infinite steroid concentration to a value measured for other steroids for which this value could be determined

experimentally. For cases in which K_d for steroid << K_d for DIG-PEG$_3$-Alexa488, the data could not be fit to the model and only qualitative conclusions could be reached (Fig. 4, dashed lines).

The inhibition constant for each protein-ligand interaction, K_i, was calculated from the measured IC_{50} and the K_d of the protein-label interaction according to a model accounting for receptor-depletion conditions[49]. IC_{50} values, the concentrations of free unlabeled ligand producing 50% binding signal inhibition, were calculated from the measured I_{50} values[49]. Reported I_{50} and subsequent K_i values represent the average of at least three independent measurements from at least two batches of purified protein and a fresh unlabeled inhibitor stock prepared for each. For DIG10.3, [DIG-PEG$_3$-Alexa488] = 1 nM and [DIG10.3-TEV-his$_6$] = 20 nM. For DIG10.3 Y34F, [DIG-PEG$_3$-Alexa488] = 10 nM and [DIG10.3 Y34F-TEV-his$_6$] = 200 nM. For DIG10.3 Y101F, [DIG-PEG$_3$-Alexa488] = 10 nM and [DIG10.3 Y101F-TEV-his$_6$] = 200 nM. For DIG10.3 Y34F/Y99F/Y101F, [DIG-PEG$_3$-Alexa488] = 500 nM and [DIG10.3 Y34F/Y99F/Y101F-TEV-his$_6$] = 5 μM.

Circular dichroism spectroscopy. Circular dichroism spectra were collected on an Aviv 62A DS spectrometer. Samples were prepared in PBS. Fixed-temperature scans were conducted at 25 °C. Guanadinium hydrochloride (GuHCl) denaturations were carried out at 25 °C by monitoring spectra of samples of 25 μM DIG10.3-TEV-his6 or 1z1s-TEV-his6 that were incubated with GuHCl in PBS, pH 7.4, for at least one hour prior to data collection. The concentration of the GuHCl stock solution was determine from the measured refractive indices of the solution and of the buffer[50]. Data are given in Supplementary Fig. 20.

Analytical Ultracentrifugation. The sedimentation profile of DIG10.3-TEV-his$_6$ was collected in the absence and presence of DIG using a Beckman Proteome XL-I instrument. Samples of DIG10.3-TEV-his$_6$ purified by size exclusion chromatography (see above) were prepared at concentrations of 0.5 mg/mL (16.7 μM) and 0.25 mg/mL (8.4 μM) in PBS, pH 7.4. A four-fold molar excess of DIG diluted from a 10 mM stock solution prepared in MeOH was added to samples with ligand (66.8 μM DIG/0.66% MeOH for the 0.5 mg/mL sample; 33.4 μM DIG/0.33% MeOH for the 0.25 mg/mL sample). The same amount of MeOH was added to the corresponding control samples without DIG. Both samples were also analyzed in the absence of MeOH. Prior to data collection, all samples were incubated for 2 hrs at room temperature to ensure equilibrium. For each sample, a reference solution comprising buffer and the appropriate amount of MeOH was employed. A sedimentation velocity run was performed using six two-

sector charcoal-epon velocity cells. Absorbance at 280 nm was used to monitor the protein, and the cells were centrifuged at 50,000 rpm in an 8-hole rotor at 20 °C. The sample was observed to have fully sedimented to the bottom of the cell in 5 hrs.

Data were fit using direct Lamm equation modeling of the boundary data as implemented in the program SEDFIT[51]. A c(s) size distribution from 0 to 15 s or 0 to 20 s was determined using a resolution of 0.075 s. The program SEDENTERP[52] was used to calculate buffer density and viscosity. For the c(s) distribution analysis, the experimental boundary profile was fit using a Marquadt-Levenberg algorithm as mentioned previously[53]. All samples showed a single predominant peak comprising ~99% of the total signal. The molecular weight estimate for DIG10.3-TEV-his_6 averaged over the 0.5 mg/mL and 0.25 mg/mL samples is 36 ± 1.7 kDa, which agrees with the calculated molecular weight for the dimer (36,130 Da). For the DIG10.3-TEV-his_6–DIG complexes, the averaged molecular weight estimate is 37 ± 1.7 kDa. The aggregate level was measured to be 1 ± 0.2 kDa for the highest concentration sample and did not change upon addition of MeOH or DIG. Data are shown in Supplementary Fig. 15e.

Isothermal Titration Calorimetry. Calorimetric titrations of DIG10-TEV-his_6, DIG10.1-TEV-his_6, 10.2-TEV-his_6 and DIG10.3-TEV-his_6 with digoxigenin were performed on an iTC200 microcalorimeter (MicroCal) at 25 °C. Protein samples were exchanged against PBS buffer, pH 7.4. The ligand solution was prepared by diluting a stock solution of digoxigenin (100 mM in 100% DMSO) in the flow through of the last buffer exchange. The samples used for ITC measurements were prepared so that the final DMSO concentration in the PBS buffer was 1-3%. All solutions were filtered using membrane filters (pore size 0.45 μm) and thoroughly degassed for 20 min by stirring under vacuum. The 200 μL sample cell and the 40 μL injection syringe were typically filled with a 5–350 μM solution of protein and with 50-2000 μM of the titrating digoxigenin, respectively. Typical ITC experiments consisted of a preliminary 0.3 μL injection followed by 10-15 subsequent 3.0-3.5 μL injections. Data for the preliminary injections, which are affected by diffusion of the solution from and into the injection syringe during the initial equilibration period, were discarded. The ITC titration data were integrated and analyzed with Origin 7.0 (MicroCal).

Supplementary Tables

Supplementary Table 1. List of genes ordered in this study

Gene	Scaffold	Sequence
DIG1	1gy7	CATATGAGCCTGGACTTCAATACGCTGGCACAAAACTTCACGCAATTCTGGTACAATCAATTCGA CACCGACCGTTCGCAACTGGGCAATCTGTATCGTAATGAAAGTATGACCACGCATGAAACCTCCC AGCTGCAAGGCGCGAAAGATATTGTGGAAGGTCTGGTTAGCCTGCCGTTTCAGAAATATCAAGCA CGTATCACCACGCTGGATGCACAGCCGGCAAGTCCGTACGGCGACGTGCTGGTTATGGTCACCGG CGATGGTCTGACGGACGAAGAACAAAACCCGTGGCGTTATTCAGCGGTCTACCACCTGATTCCGG ACGGCAACTCGTATTACGTGTTCAATGCAATCTGGCGCTATAACTACAGCGCTGGTTCTCTCGAG
DIG2	2mve	CATATGGTCTCCGCAAAAGATTTCTCTGGTGCCAACCTGTACACGCTGGAAGAAGTGCAGTATGG TAAATTCGAAGCTCGTATGAAAATGGCGGCCGCATCAGGCACCGTGAGCTCTATGTATCTGCATC AGAACGGCTCGTGGATTGCCGATGGTCGTCCGTTTGTGGCAGTTTACATTATCGTCCTGGGCAAAA ACCCGGGTTCATTCCAATCGAATATCGTGACCGGCAAAGCTGGTGCGTATAAAACGAGTGCTAAA CATCACGCAGTTTCCCCGGCTGCAGATCAGGCATTTCACACCTATGGCCTGGAATGGACGCCGAA CTACGTTCGTTGGACCGTCGACGGTCAAGAAGTGCGCAAAACGGAAGGCGGTCAAGTTAGCAATC TGACCGGCACGCAAGGTCTGCGCTTTAGCCTGTACAGTTCCGAATCTGCCGCATGGGTGGGCCAG TTCGATGAATCCAAACTGCCGCTGTTTCAATTCATCAACTGGGTCAAAGTGTACAAATACACCCCG GGTCAGGGTGAAGGCGGTAGCGATTTTACCCTGGATTGGACGGACAATTTTGATACGTTCGACGG CTCTCGTTGGGGCAAGGGTGATTATACCTTTGACGGTAACCGCGTTGATTTCACGGACAAAAATA TTTACAGTCGTGATGGCATGCTGATCCTGGCTCTGACCCGCAAAGGCCAGGAATCTTTCAATGGTC AAGTCCCGCGTGATGACGAACCGGCACCGCAGTCATCGAGCTCTGCACCGGCAAGTTCCGGTAGC CTCGAG
DIG3	1pvx	CATATGGGCACGACCCCGAACTCGACCGGCTGGCACGACGGCTACTATTATCACTGGTGGAGCGA CGGCGGCGGCGACTCAACCTACACGAACAACAGTGGCGGTACCTATGAAATTACGTGGGGCAAT GGCGGTATTCTGATCGGCGGTAAAGGCTGGAACCCGGGTCTGAATGCGCGTGCCATCCATTTTAC CGGCGTCTATCAGCCGAACGGTACGAGTTTCCTGTCCGTTTACGGCTGGACCCGCAATCCGCTGGT CAGCTATTACATTGTGGAAAACTTCGGTAGCTCTAATCCGAGTTCCGGCTCCACCGATCTGGGTAC GGTGTCATGCGACGGCTCGACCTATACGCTGGGTCAGAGCACCTGGTATAACTACCCGTCTATCG ATGGTACCCAAACGTTTAATGCTTACTGGTCTGTGCGTCAGGACAAACGCTCATCGGGCACCGTTC AAACGGGTTGTCATTTCGATGCATGGGCTAGCGCGGGCCTGAACGTTACCGGTGACCACTATTAC CAAATTGTCGCAACGTTTGGCTGGTATAGCTCTGGTTACGCCCGTATCACCGTGGCAGATGTTGGC GGTAGCCTCGAG
DIG4	3b4o	CATATGTCTGACGTGGAATCACTGGAAAATACCTCCGAAAACCGTGCTCAAGTCGCCGCTCGCCA ACACAATCGCAAAATCGTGGAACAGTATATGCATACCCGCGGCGAAGCGGAACTGAAAATGCAC CTGCTGTTTACGGAAGATGGCGTCGGCGGTAGCTGGACCACGAGCTCTGGTCAGCCGATTGCCAT CCGTGGCCGCGAAAAACTGGGTGAACATGATGTGTTTCTGCTGCAAGTTTTCCCGGATTGGGTCTG GACCGACATTCAGATCTTTGAAACGCAAGACCCGAACTGGTTCTGGGTGGAATGCCGTGGCGAAG GTGCGATTGTTTTCCCGGGTTACCCGCGTGGTCAATATCGTGCACACTACCTGGCTAGTTTTCGTTT CGAAAACGGTCTGATTAAAGAAACCCGCTGGTTTTGGAATCCGTGTGAAGCGTTCCGTGCCCTGG GCATCGAAGGTTCCCTCGAG
DIG5	1z1s	CATATGAATGCGAAAGAAATCCTGGTCCATAGTCTGCGTCTGCTGGAAAACGGCGATGCTCGTGG TTGGTGTGACCTGTTCCACCCGGAAGGCGTCCTGGAATTTCCGTATGCCCCGCCGGGCTGGAAAA CCCGTTTCGAAGGTCGCGAAACGATTTGGGCACACATGCGTCTGCATCCGGAACACGTCACCGTG CGCTTTACGGATGTTCAGTTCTATGAAACCGCTGATCCGGACCTGGCGATCGGCGAATACCACGG TGACGGTGTGGTTACCGTGAGCGGCGGTAAATATGCGGCCGATTTTATTACCGTTCTGCGTACGCG CGACGGTCAGATCCTGCTGTACCGTGTGTTCTGGAACCCGCTGCGTGCACTGGAAGCAGCTGGCG GTGTTGAAGCGGCCGCAAAAATTGTCCAAGGCGCGGGTTCTCTCGAG
DIG6	3cu3	CATATGAACCTGCAAACCGACCAGACGACGACGACCGCCGACGAATCAGCAATCCGTGCCTTTAC CCGCCAGATGATCGACGCATGGAACCGTGGTAGCGGTGAAGGTTTTGCAGCACCGTTCTCTGAAA CCGCGGATTATATTACCTTTGACGGTACGCATCTGAAAGGCCGCAAAGAAATCGCAGCTTTTGCA CAGCAAGCTTTCGATACCGTTGCAAAGGGTACGCGTCACGAAGGCGAAGTCGATTTTGTGCGCTT CGTTAACAGTCAGCTGGCTCTGATGCTGACCGTCTGGCGTGTGATTCTGCCGGGTCAAACCGAAA

		CGAGCGCGTCTATGGATGCCCTGCCGCTGTATGTGGTTACGAAAGGTGACGAAGGCTGGCAGATC GAAGGTCTGCTGGCGACCTACAAACTGACGCTGGAACGTGGCTCCTTTCTGGATGACTTCGATAG TCTGTCCGCGGAAGCCCAGCGCCAAGTTACCGACCTGGTCGCCAGCCTGAAACAGTCACATGGCT CGCTCGAG
DIG7	3gwr	CATATGTCTGAACCGGTTTTCCCGACCCCGGAAGCGGCTGAAGATGCCTTTTATGCTGCCCTGGAA GCTGGCTCCCTGGATGACTATATGGCTGTTTGGGCGCGTGATGACCATGTCGCCTTTATTCACCCG CTGGCAGCACCGCTGAACGGTCGTGCAGCTGTGGCAGCAGGTTGGCGTAGTCATCTGGGTGCAGC TGGTCGTTTCCGCCTGCAGGTCAAAGCCGTGCATGAAATTCGTCAAGCAGATCACGTCATTCGCAT CACCGACATCTTTTTCACGGGCGGTGATGAAACCGCACCGCGTCCGGCAGCACTGGCAACGGCTG TTTATCGTCGCGAAGCTGACGGCTGGCGCATGGTGCTGTACCATGCAAGCCCGCTGCAGGTTGGT GCAAAAGCAGGTGCAGATACCCCGCCGGTGGTTTTTCACGGTTCTCTCGAG
DIG8	3hk4	CATATGACGATTGCGGAAATTGCCAAAGACTACACGGAACTGAACAAACAGGGCGACCAAGCGG GTGCCTATGAAAAATATGCTGCCGATGATATTGCGTATTACCAAGCCATGGAAGGCCCGATGGCA GTCAGCCATGGTAAAGAAGCATGGCGTCAGGCTCTGCAATGGTATCAGGAAAACGCCGAATTTCA CGGCGGTTCTGTCGAAGGCCCGTACGTGAATGGTGATCAGTTTGCGCTGCGTTTCAAATGGGACG TGACCCCGAAAGCAACGGGTGAACGTGTGACCGTTGACGGTGTTCATCTGTATACCGTCAAAAAC GGCAAAATCACGGAAGTTCGCTGGTATTACGGTAGTCTCGAG
DIG9	1i60	CATATGAAACTGTGCTTTAATGAATTTACCACCCTGGAAAACTCAAACCTGAAACTGGACCTGGA ACTGTGCGAAAAACACGGCTATGATTATATTCAGATCCGTACCATGGATAAACTGCCGGAATACC TGAAAGACCATAGTCTGGATGACCTGGCGGAATATTTTCAAACGCATCACATTAAACCGCTGGCC CTGCTGCATCTGCTGTTTTTCAACAATCGTGATGAAAAAGGCCACAATGAAATTATCACCGAATTC AAAGGCATGATGGAAACCTGCAAAACGCTGGGTGTGAAATACGTTCTGGCGTGGCCGCTGCGCAC CGAACAGAAAATCGTGAAAGAAGAAATTAAAAAATCTAGCGTGGATGTCCTGACGGAACTGTCA GACATTGCGGAACCGTATGGCGTTAAAATCGCCCTGAATTTTGGCGGTCTGCCGCAGTGTACCGT CAACACGTTCGAACAAGCCTACGAAATTGTCAACACCGTGAATCGTGATAACGTCGGTCTGTATC TGAATTCGTTTCATTTCCACGCAATGGGCTCAAACATCGAATCGCTGAAACAGGCTGATGGCAAG AAAATTTTTATCTACGGCATTGGTGATACCGAAGACTTTCCGATCGGCTTCCTGACGAGTGAAGAC ATGGTTTGGCCGGGCCAAGGTGCGATTGATCTGGACGCACACCTGTCCGCTCTGAAAGAAATCGG TTTTAGCGATGTGGTTTCTGTGGCACTGCTGCGCCCGGAATATTACAAACTGACCGCGGAAGAAG CCATTCAGACGGCGAAGAAAACCACGGTTGATGTCGTGTCTAAATATTTCAGTATGGGTTCCCTC GAG
DIG10	1z1s	CATATGAATGCTAAAGAAATTGTTGTCCACTCACTGCGTCTGCTGGAAAATGGCGATGCCCGTGG TTGGTGCGACCTGTTTCACCCGGAAGGCGTGCTGGAATATCCGTACGCCCCGCCGGGCCATAAAA CCCGTTTTGAAGGTCGCGAAACGATTTGGGCGCACATGCGTCTGTTCCCGGAATATGTGACCGTTC GCTTTACGGATGTCCAGTTCTACGAAACCGCCGATCCGGACCTGGCAATCGGCGAATTTCATGGT GACGGTGTCCACACCGTGAGCGGCGGTAAACTGGCGGCCGATTATATTTCTGTTCTGCGTACGCG CGACGGTCAGATCCTGCTGTACCGTGTGTTTTTCAACCCGCTGCGTGTCCTGGAAGCTCTGGGCGG TGTGGAAGCAGCTGCGAAAATTGTTCAAGGCGCGGGTAGTCTCGAG
DIG11	1zo2	CATATGGACCAATCAATCAATCTGAATCCGCAATTCGACCAAATCGGTAAACAGTTCGTGCAGCA TTATGTGCAAACGCATCAAACCAATCGTCCGGCACTGGGCGGTCTGTATGGTCCGCAGTCAATGC TGACCTTTGAAGATACGCAGTTCCAAGGCCAGGCGAACATCGTTAACAAATTCAACTCGTTCAAC TTCCAACGTGTGCAGGCCGAAATTACCCGCGTTGATTGCCAACCGAGCCCGAACAATGGTTCTAT CGTCTTTACCACGGGTGACGGTCGTGGTGATGACGGTCAGCCGTATAAATTCAGTCATGTGTACA ACCTGATGCCGTCCGGTAATGGCGGTTTTATGATTTTCAACGACCTGTGGCGCATCAATCTGGGCG GTTCTCTCGAG
DIG12	2owp	CATATGGGCATGGAAGTGAATCAGCCGGACATTGTGGCTCAAGTGCAAGCAGCATTTGTGGAATA CGAACGCGCCCTGGTGGAAAACGACATCGAAGCAATGAACGCTCTGTTTTGGCACACCCCGGAAA CGGTCTTCTATGGCGCGACCACGGTGCAGCATGGCGGTGAAGCGTGGCGTGCCCATGTTGAACGC AGTCAACCGCACCCGAAATCCCGTAAACTGCATCGCACGGTGGTTACCACGTTTGGTACCGATTT CGCCACCGTCAGCACGGAATTTACCTCTGACGGTACCCCGCTGCTGGGTCGTCAGATGCAAACCT GGGCACGCCTGTCTCCGGCTGATGGCTGGAAAATTGTGGCGGCCCACTTCTCACTGATCGCGATG CCGGGTTCGCTCGAG
DIG13	2ox1	CATATGAAACTGGTTGCGGGCCTGTCCTCGCCGGAAGAACTGGAACTGGCTGAAAAAGCGGATGT TGTGACCCTGCATATTGACCTGTTTGATTTTAGTGGCGCACGTGTTGACAAAGAAAAAGGTCTGAC CTGCATGCGCGTCTCCGATGGCGGTAAATTTGAAGGCGACGAACGTGAACGCATTGAAAAAATGA

		AACGTGCCTTCGATAGCCTGAACCCGGACTATGTTTACCTGGAAAGCGATCTGCCGGACTCTGCG TTTGATTTCAACTGTCGTATTGCCGAATTCTATGGCAATGTGATCCGCACCCCGGATTACAGCGAA CTGAAAGGTATTGTTGAAGGCCGTCGCGGTGATCTGGTGGTTATCGCAACGATGGGTAAATCTAA ACGTGATGTCGAAACCATTGTGCGCATCCTGACGAATTATGATGACGTCCTGGCTCATCTGATGG GCGAACGCTTTAGCTTCACGATGGTTCTGGCGGCCTATCTGGGTTCTCCGTGGATTTGGTGCTACG TGGGCTCACCGAAATTTCCGGGTGCAATCTCGCTGGATGACGCTCGTGAAATTATCAGTCGCCTG GGCGGTTCCCTCGAG
DIG14	3e5z	CATATGACCCTGCGTGCCGCCCGTCCGGAATTTCTGGACCTGTTTCCGGCTGGTGCTGAAGCTCGT CGCCTGGCTGATGGTTTTACCTGGACCTCCGGCCCGGTGTATGTTCCGGCACGTTCAGCTGTGATT TTTAGCGATTCTGCGCAGAACCGCACCTGGGCCTGGTCAGATGACGGTCAACTGTCGCCGGAAAT GCATCCGAGCCATCACCAGGGCGGTCACTGCCTGAATAAACAAGGCCATCTGATTGCATGTAGCC ACGGTCTGCGTCGCCTGGAACGTCAGCGTGAACCGGGCGGTGAATGGGAATCAATCGCTGATTCG TTCGAAGGCAAAAAACTGAACAGTCCGTCCGCAGTGTGCCTGGCACCGGATGGTAGCCTGTGGTT TTCTGATCCGACCTGGGGCATTGACCTGCCGGAATTCGGTTACGGCGGTGAAATGGAACTGCCGG GCCGTTGGGTTTTTCGTCTGGCCCCGGATGGTACGCTGAGCGCTCCGATCCGTGACCGTGTCAAAC CGACCGGTCTGGCCTTCCTGCCGTCTGGTAATCTGCTGGTGTCTGATGCAGGCGACAACGCTACGC ATCGTTATTGCCTGAATGCGCGTGGCGAAACCGAATACCAGGGTGTTCACTTTACGGTCGAACCG GGTGCGACCTATTACCTGCGCGTTGATGCCGGCGGTCTGATTTGGGCATCGGCAGGTGACGGTGT CCATGTGCTGACGCCGGATGGCGACGAACTGGGTCGTGTCCTGACCCCGCAAACCACGACCGGCC TGTGTTTTGGCGGTCCGGAAGGTCGCACGCTGTATATGACCGTGAGTACCGAATTCTGGTCCATCG AAACCAACGTTCGTGGCGGTTCCCTCGAG
DIG15	3fmz	CATATGGAACGTGACTGTCGTGTCTCCTCCTTCCGTGTGAAAGAAAACTTCGATAAAGCTCGCTTC TCGGGCACCTGGTATGCTATGGCGAAAAAAGACCCGACCGGCCTGTTTCTGCAGGATAACATTGT GGCAGAATACAGCGTTGATGAAACGGGCCAAATGTCTGCGACCGCCAAAGGTCGTGTCCGCCTGC TGAACAATTGGGACGTCTGCGCCGATTATGTGGGCACCTTTACGGACACCGAAGATCCGGCAAAA TTCAAAGGCAAATGGTGGGGTGTTGCTAGTTTCCTGCTGAAGGGTAATTCCGATCATTGGATCGTT GATACGGACTACGATACCTATGCAGTCATGTACGGCTGCACGCTGCTGAACCTGGACGGTACCTG TGCTGATAGCTATTCTTACGTGTATTCCCGTGACCCGAATGGTCTGCCGCCGGAAGCACAGAAAA TTGTTCGTCAGCGCCAAGAAGAACTGTGCCTGGCCCGTCAATACCGCCTGATCGTGCACAACGGC TATTGTGATGGTCGTTCAGAACGCAATCTGCTGGGTTCGCTCGAG
DIG16	3gwr	CATATGTCTGAACCGGTGTTCCCGACCCCGGAAGCCGCCGAAGATGCCTTTTACGCTGCCTTTGAA GCTCGCTCGCTGGATGATATGATGGCTGTGTGGGCGCGTGATGACCATGTTGCCTATATTCACTGG CTGGCAGCACCGCTGAACGGTCGTGCAGCTGTTGCAGCAGGTTGGCGCAGTATGTTTGGTGCAGC TGGTCGTTGGCGTGTCCAGGTGAAAGCTGTCCATGAAATTCGTCAAGCCGATCACGTTATTCGCAT CGTCGACCAGTTTTTCACCATCGGTGATGAAACGGCACCGGGTCCGGCAGGTCTGGCAACCAATG TGTACCGTCGCGAAGCGGACGGTTGGCGCATGGTCCTGACGCATCACAGCCCGCTGCAAGTGGGT GCAAAAGCAGGTGCAGATACCCCGCCGGTGGTTTTCCATGGCTCTCTCGAG
DIG17	3cu3	CATATGAACCTGCAAACCGACCAAACGACGACCACCGCTGATGAATCGGCTATCCGTGCCTTCCA CCGCCAACTGATTGATGCTTTTAACCGTGGTAGCGGTGAAGGTTTTGCAGCACCGTTCTCTGAAAC CGCAGATTTTATTACCGCTGAGGGTACGCATCTGAAAGGCCGCAAAGAAATCGCAGCTTATCACC AGCAAGCGTTCGATACCGTGGTTAAGGGTACGCGTCTGGAAGGCGAAGTGGACTTTGTTCGCTTC GTCAACTCCCAGCTGGCGCTGATGCTGGTCGTGAGCCGTATTATCCTGCCGGGTCAAACCGAAAC GAGCGCCTCTCGCGATTATCTGCCGCTGTACGTTGTCACCAAAGGTGACGAAGGCTGGCAGATTG AAGGCCTGCTGGCAACCCGTAAACTGACGCTGGAACGCCAATTTTTCCTGGATGACTTTGATAGT CTGTCCGCGGAAGCCCAGCGTCAAGTTACGGACCTGGTGGCGAGTCTGAAACAGTCACATGGTTC GCTCGAG
DigA16	-	CATATGGATGTGTATCACGATGGTGCTTGCCCGGAAGTCAAACCGGTGGACAACTTCGATTGGAG TCAGTATCATGGTAAATGGTGGCAGGTCGCGGCCTACCCGGATCATATTACCAAATATGGCAAAT GCGGTTGGGCAGAATACACGCCGGAAGGCAAAAGCGTCAAAGTGAGTCGTTATTCCGTGATCCAT GGCAAAGAATACTTTTCTGAAGGTACCGCCTATCCGGTTGGCGATAGTAAAATCGGTAAAATCTA CCACTCCTACACCATTGGCGGTGTGACGCAGGAAGGTGTTTTTAACGTTCTGTCAACGGATAACA AAAACTACATCATCGGCTACTTCTGCTCGTATGATGAAGACAAAAAAGGTCACATGGACCTGGTT TGGGTCCTGAGCCGCTCTATGGTCCTGACCGGCGAAGCGAAAACGGCCGTGGAAAACTACCTGAT CGGTAGCCCGGTGGTTGATTCTCAAAAACTGGTTTATTCAGACTTCTCGGAAGCAGCTTGTAAAGT CAATGGCTCACTCGAG
3hk4	-	CATATGACGATTGCGGAAATTGCGAAAGACTTTACGGAACTGCTGAAACAGGGCGATAATGCGG

		GTGCGGCGGAAAAATACAACGCTGATGATATTGCGAGTTATGAAGCGATGGAAGGCCCGATGGC AGTCAGCCACGGCAAAGAAGCTCTGCGTCAGAAATCTCAATGGTGGCAGGAAAACCATGAAGTG CACGGCGGTAGCGTTGAAGGCCCGTACGTCAATGGTGATCAATTTGCGCTGCGTTTTAAATTCGA CGTGACCCCGAAAGCAACGGGCGAACGCGTTACGATGGATGAAGTTGGTCTGTACACGGTCAAA AATGGCAAAATCACGGAAGAACGCTTCTACTACGGCAGTCTCGAG

Supplementary Table 2. Features of ordered designs

				Computational Characterization		Experimental Characterization		
Design	Scaffold	Fold Superfamily[1]	# Mutations from Scaffold	Interface Energy (Reu[2])[3]	S_c[3]	Expression Yield (mg/L)[4]	Binding Signal with DIG- BSA-biotin[5]	Binding Signal with DIG-RNase-biotin[5]
DIG1	1gy7	NTF2-like	16	-15.0	0.73	7.0	-	-
DIG2	1mve	Concanavalin A-like	15	-13.4	0.60	9.6	-	-
DIG3	1pvx	Concanavalin A-like	12	-12.8	0.69	1.1	-	-
DIG4	3b4o	NTF2-like	15	-11.1	0.46	6.3	-	-
DIG5	1z1s	NTF2-like	11	-14.8	0.68	48.5	+	++
DIG6	3cu3	NTF2-like	14	-11.1	0.58	1.4	-	-
DIG7	3gwr	NTF2-like	13	-13.5	0.67	6.1	-	-
DIG8	3hk4	NTF2-like	19	-15.2	0.67	28.5	-	+++
DIG9	1i60	Xylose isomerase-like	19	-9.3	0.50	0.7	-	-
DIG10	1z1s	NTF2-like	10	-15.4	0.60	31.1	+++	+++
DIG11	1zo2	NTF2-like	13	-14.7	0.64	insoluble	-	-
DIG12	2owp	NTF2-like	12	-8.7	0.67	2.2	-	-
DIG13	2ox1	Type I 3-dehydroquinase	18	-10.1	0.63	0.9	-	-
DIG14	3e5z	Gluconolaconase	16	-11.0	0.60	0.8	-	-
DIG15	3fmz	Lipocalin-like	13	-13.8	0.67	0.4	+++	-
DIG16	3gwr	NTF2-like	10	-11.4	0.70	insoluble	-	-
DIG17	3cu3	NTF2-like	7	-9.8	0.56	34.5	-	-

[1]SCOP or PFAM classification. [2]Rosetta Energy Units. [3]Designs were relaxed with heavy atom coordinate constraints and scored with score12 using the enzdes weights set, as detailed in Methods.

[4]Expression yields for pET29FLAG constructs. [5][label] =2.7 μM.

Supplementary Table 3. Thermodynamic parameters of DIG binding to DIG10 variants determined by isothermal titration calorimetry

Variant	K_d	ΔG	ΔH	-TΔS
DIG10	12.2 μM	-6.70	-0.46	-6.24
DIG10.1	196 nM	-9.14	-8.29	-0.85
DIG10.2	168 nM	-9.23	-9.92	0.69
DIG10.3	541 pM[1]	-12.60[1]	-10.83	-1.79[1]

[1]For DIG10.3, the isotherm was too steep to reliably determine ΔG; the value measured in FP experiments with DIG-PEG_3-Alexa488 is given instead and was used and compute –TΔS

Supplementary Table 4. DIG10.3 steroid binding affinities measured by fluorescence polarization equilibrium competition experiments

Variant	Steroid	IC_{50}	Inhibition Constant (K_i)	ΔΔG (DIG – steroid) (kcal/mol)[1]
DIG10.3	Digoxigenin (DIG)	71 ± 13 nM	653 ± 262 pM	-
	digitoxigenin	2.1 ± 0.1 μM	19 ± 7 nM	2.0
	progesterone	27 ± 3 μM	243 ± 91 nM	3.5
	β-estradiol	225 ± 32 μM	2.1 ± 0.8 μM	4.8
	digoxin	24 ± 7 nM	223 ± 105 pM	-0.7
Y101F	Digoxigenin (DIG)	1.6 ± 0.2 μM	39 ± 8 nM	-
	digitoxigenin	< 155 nM	< 3.8 nM	> 1.4
	progesterone	1.2 ± 0.3 μM	30 ± 9 nM	-0.2
	β-estradiol	68 ± 7 μM	1.7 ± 0.4 μM	2.2
Y34F	Digoxigenin (DIG)	1.36 ± 0.03 μM	59 ± 6 nM	-
	digitoxigenin	16.3 ± 0.7 μM	714 ± 79 nM	1.5
	progesterone	1.8 ± 0.3 μM	76 ± 14 nM	0.2
	β-estradiol	343 ± 40 μM	15 ± 2 μM	3.3
Y34F/Y99F/Y101F	Digoxigenin (DIG)	29 ± 4 μM	580 ± 229 nM	-
	digitoxigenin	< 786 nM	< 16 nM	> 2.1
	progesterone	< 857 nM	< 17 nM	> 2.1
	β-estradiol	81 ± 2 μM	1.6 ± 0.6 μM	0.6

[1] ΔΔG values were estimated using $\Delta\Delta G = -RT\ln(K_i(\text{DIG})/K_i(\text{steroid}))$

Supplementary Table 5. Binding affinities of DIG10.3 variants for DIG-PEG$_3$-Alexa488 measured by fluorescence polarization

Variant	Dissociation Constant (K_d)	ΔΔG (DIG10.3 – Variant) (kcal/mol)[1]
DIG10.3	541 ± 193 pM	-
Y34F	24.9 ± 2.4 nM	2.3
Y101F	14.3 ± 2.7 nM	1.9
Y115F	9.4 ± 1.4 nM	1.7
Y99F/Y101F	12.4 ± 2.0 nM	1.9
Y34F/Y99F/Y101F	293 ± 105 nM	3.7
Y34F/Y99F/Y101F/Y115F	166 ± 1.4 μM	7.5

[1]ΔΔG values were estimated using $\Delta\Delta G = -RT\ln(K_d(\text{DIG10.3})/K_d(\text{variant}))$

Supplementary Table 6. Boltmann-weighted probabilities of hydrogen bonding residue conformations

Design	Residue[1]	Boltzmann-weighted probability[2]	Average[3] probability
DIG1	Tyr17	0.089	0.085
	His39	0.029	
	Tyr62	0.156	
	Tyr103	0.129	
DIG2	Tyr40	0.215	0.174
	His42	0.093	
	Arg137	0.290	
	Tyr141	0.185	
	Tyr203	0.149	
DIG3	His16	0.119	0.152
	Tyr77	0.119	
	Tyr88	0.316	
	Tyr172	0.120	
DIG4	Tyr30	0.128	0.078
	His73	0.035	
	Tyr124	0.105	
DIG5	His58	0.043	0.118
	Tyr84	0.210	
	Tyr97	0.088	
	Tyr115	0.245	
DIG6	Tyr46	0.261	0.163
	His77	0.098	
	Tyr118	0.170	
DIG7	Tyr28	0.073	0.072
	His61	0.034	
	Tyr122	0.153	
DIG8	Tyr10	0.072	0.084
	Gln19	0.054	
	Tyr33	0.111	
	His101	0.044	
	Tyr103	0.220	
DIG9	His67	0.051	0.100
	Tyr172	0.259	
	Asn174	0.100	
	His177	0.285	
	Ser214	0.027	
DIG10	Tyr34	0.168	0.213
	Tyr101	0.173	
	Tyr115	0.333	
DIG11	His25	0.040	0.080
	His104	0.175	
	Tyr106	0.073	
DIG12	Tyr19	0.117	0.096
	Trp56	0.185	
	His59	0.062	
	His66	0.063	
DIG13	His25	0.062	0.097
	Tyr77	0.151	
	His149	0.097	
DIG14	Ser31	0.034	0.059
	Tyr222	0.102	
DIG15	Tyr45	0.203	0.112
	His104	0.076	
	Tyr135	0.101	
	Tyr137	0.102	
DIG16	Tyr40	0.123	0.072
	His42	0.019	
	Gln90	0.090	
	His124	0.129	
DIG17	His64	0.056	0.077
	Ser98	0.041	
	Tyr118	0.196	

[1]Boltzmann-weighted probabilities of all hydrogen-bonding residues, including those incorporated during matching and those designed during sequence optimization, are provided. [2]Metrics given are from designs that were relaxed with heavy atom constraints before Boltzmann-weighted probability calculations. During design filtering stages, calculations were performed on structures that were not relaxed. [3]Geometric mean.

Supplmentary Table 7. Boltmann-weighted probabilities of Tyr34, Ty101, and Tyr115 confomations in DIG10-based designs

Residue	DIG10[1]	DIG10.2[2]	DIG10.3[2]
Tyr34	0.168	0.708	0.459
Tyr101	0.173	0.333	0.998
Tyr115	0.333	0.836	0.905
Geometric Mean	**0.213**	**0.582**	**0.746**

[1]The design model was relaxed with heavy atom coordinate constraints before Boltzmann-weighted probability calculations. [2]Boltzmann-weighted probability calculations were performed on the crystallographic copy with the least amount of missing density in the first and second coordination shells (<8 Å from ligand heavy atoms). For both DIG10.2 and DIG10.3, this was chain A. All chains were relaxed with heavy atom constraints before rotamer-Boltzmann calculations.

Supplementary Table 8. List of PDB entries used as scaffolds

1b54A	1d9eD	1e3vB	1aj0A	1f5zA	1aw1A	1g69B
1g7uA	1h1yA	1btmA	1h5yB	1dbtA	1i45A	1dl3A
1i60A	1i4nA	1eg9B	1dqwA	1j2wA	1eixC	1dvjA
1dxeA	1k32A	1k3uA	1eyeA	2a15A	2a4aA	1fwtA
1hg3A	1l6wA	1gpwA	1gqnA	1m1bA	1gthA	1gy7B
1ho1C	1gvfB	1m3uA	2b4wA	1hkxE	1m6jA	1gyhA
1n7kA	1jcjA	1o60A	1ka9F	1o66E	1o0yA	1o68B
1jn5B	1o1zA	1iubA	1jkgA	1izcA	1p0kB	1lbfA
2e7fA	1q40A	1q40B	2agkA	1kqyA	2f98C	2f6uA
1q7fB	2bdqA	2aqwA	2bhmA	2cc3A	1mo0A	2bmoB
2bngC	2bs5A	1s18A	2chcA	1s5aB	1of5A	1of5B
2h6rB	1mzhA	2h9lA	1nu3B	1ok6A	2cw6A	1nsjA
2cz5A	2dp3A	2i5iA	2i9eB	1oqfA	2ehhA	2dsoC
1oy0A	2duaA	2ekcB	1u83A	1pjxA	1qapA	2dyhA
2dzaB	1qdsA	2fjkC	1qo2A	1rd5A	1v93A	2fliC
1qmaA	1v5xA	1ri6A	1w37A	1qpnA	2gexA	2geyA
1w0mA	2ghsA	2gjlA	1qwgA	2gopA	2hesX	1sfjB
1rpxA	1rv8B	1sfsA	1sgjA	1x1oA	3b4oB	3b4uB
1sjwA	1rwiA	3b7cA	1x7iB	3b8iA	3b8lA	1y0eA
3c56A	1thfD	2htmC	3c2vA	1tp6A	1suxA	3c6cA
2imjA	2jbmA	1tqjC	1treA	2jgqA	1tqxA	1z1sA
1vc4A	1tuhA	1vd6A	1ujpA	3e02A	2p10C	1vcvA
3d9rB	3e49A	1vhcF	1vizA	1vkfB	3e99A	2p4oA
3e5zA	1wbhB	1vlwB	3f14A	1vs1D	3e9aA	3f40A
2p9wA	1vpxE	1vqtA	3bb9B	1xc4A	3f4nC	3f4wA
1wq5B	3bg5B	3g16A	3f8hA	3f7sA	3f7xA	1xbzB
1xi3A	3f8xB	3f9sA	2q8zA	3g0kA	1wqlB	1wx0A
1xm3B	3bo9A	2r91A	2r4iA	1yadB	3blzA	1ydnA
3g8zA	3ceuA	3h51A	1yfqA	3h2aB	3chvA	3h3hB

3cjpA	3daqA	3dasA	3i10A	1zccA	1zcoA	3cnxC
3ct7A	2oczA	3cpgA	3cu2B	3cu3A	3cu9A	2nv2I
3i0yA	3dm8A	1yx1B	3ec9B	2ojhA	1zo2A	2nuwA
3dmcA	3ef8A	3dr2A	1zlpB	3ecfA	3ebtA	1yyaA
3ebyA	2oogD	1yxyA	2pcqA	3fa5A	3ehcB	2ox1A
3dsmA	3en8A	2qc5A	3dukA	3ejvA	3ff0A	3ff2A
2qe8A	3ekzA	2owpA	2qf7A	3er7B	2oxnA	3dxoB
3fh1A	3k13C	2v30A	3fgcD	3fm0A	3k0zA	3fgyA
3ez4F	3fkaB	2v81A	2qjgA	2pz0B	2qiwA	2qiyA
3ewbX	3fljA	2v5jA	3fluA	3gayA	3exrA	3fs2A
2rcdA	3fokA	3gk0A	2rfgB	3l0gC	2rfrA	3fsdA
2rgqB	3frxB	3l2iA	2w2cH	2w6rA	3gnnA	3fyoB
3hk4A	3grdA	3hfqA	3gvgA	3gwrB	3ih1A	3gzbA
3iebB	3ii7A	3ii8B	3gzrA	3hv9A	3igsB	3hx8A
3hxjA	3inpA	3hzpA	2vc6A	2z2oC	3ke7B	3iwpA
2z6iB	3js3A	2vepA	2vfhA	2ux0A	3kkgA	3ks6A
2vpjA	3krsA	3kspA	3kstA	3ktsA	3kwsA	2vwsA
3kxqA	2wozA	2ze3A	2yr1A	2zbtA	2yw3E	2yswA
2yxgA	2yyuB	2yzrA	2zvrB	4stdA	1a18	1a40
1a53	1e1a	1abe	1ajk	1f5j	1anf	1h0b
1h1a	1h61	1btm	1dl3	1i4n	1ebg	1fhv
1eua	1gca	1ftx	2b3f	1m4w	1igs	1hsl
1n4a	1jcl	1ixh	1juk	1jul	1jvx	1lbf
1lbl	1p6o	1mac	1lst	2ayh	1mpd	1mve
1oho	2dry	1qo2	1pvx	2fnc	2gh9	2fwv
2ia4	1thf	1y3n	1tml	2jen	2o62	1tsn
1wdn	3f8a	3bdr	2r79	3cg1	3cfz	1xrl
3cij	2onr	3fa6	3emm	3elz	3fen	2qo4
3ge2	3fmz					

Supplementary Table 9. Flow cytometry statistics

Library	Library Size	Sort #	Label[1]	Labeling Procedure[2]	# of Cells Analyzed	# of Cells Collected	% Cells Collected
DIG10 SSM	1088	1	2.7 μM DIG-BSA-biotin	avidity	1,200,000	77,280	6.4%
(round 1a)		2	2.7 μM DIG-BSA-biotin	avidity	3,200,000	135,000	4.2%
		3	2.7 μM DIG-BSA-biotin	avidity	3,200,000	430,000	13.4%
DIG10 combinatorial	13,500	1	1 μM DIG-BSA-biotin	avidity	9,260,000	432,562	4.7%
(round 1b)		2	750 nM DIG-BSA-biotin	avidity	6,815,099	624,312	9.2%
		3	50 nM DIG-BSA-biotin	avidity	5,918,747	321,854	5.4%
		4	5 nM DIG-BSA-biotin	avidity	4,844,636	45,199	0.9%
		5	10 pM DIG-RNase-biotin	multistep	7,651,564	121,687	1.6%
		6	5 pM DIG-RNase-biotin	multistep	10,470,304	93,404	0.9%
		7	5 pM DIG-RNase-biotin	multistep	4,534,500	30,335	0.7%
DIG10.1 doped	-	1	80 nM DIG-PEG_3-biotin	multistep	12,074,870	217,664	1.8%
(round 2)		2	80 nM DIG-PEG_3-biotin	multistep	11,284,745	189,587	1.7%
		3	50 nM DIG-PEG_3-biotin	multistep	6,644,917	206,565	3.1%
		4	1 nM DIG-PEG_3-biotin	multistep	9,927,036	98,223	1.0%
		5	1 nM DIG-PEG_3-biotin	multistep	7,230,322	135,837	1.9%
DIG10.2 combinatorial	103,680	1	400 pM DIG-PEG_3-biotin	multistep	8,796,824	591,151	6.7%
(round 3)		2	20 pM DIG-PEG_3-biotin	multistep	3,335,115	108,604	3.3%
		3	20 pM DIG-PEG_3-biotin	multistep	1,426,445	40,501	2.8%
		4	20 pM DIG-PEG_3-biotin	off-rate	4,422,371	33,549	0.8%
DIG10.1L_f1 expression (deep sequencing)	-	1	anti-cmyc-FITC	-	9,301,475	1,000,760	10.8%
DIG10.1L_f2 expression (deep sequencing)	-	1	anti-cmyc-FITC	-	5,911,327	1,006,134	17.0%
DIG10.1L_f1 binding[3]	-	2	100 nM DIG-PEG_3-biotin	multistep	18,029,680	218,350	1.2%
(deep sequencing)		3	100 nM DIG-PEG_3-biotin	multistep	4,249,106	1,611,422	37.9%
DIG10.1L_f2 binding[4]	-	2	100 nM DIG-PEG_3-biotin	multistep	20,374,557	447,293	2.2%
(deep sequencing)		3	100 nM DIG-PEG_3-biotin	multistep	5,241,370	2,022,777	38.6%

[1]For DIG-conjugated protein labels, label concentrations refer to the concentration of protein, not DIG. [2]avidity: Cells were labeled with a pre-incubated mixture of the appropriate DIG-conjugated label and SAPE; multistep: Cells were labeled with the appropriate DIG-conjugated label, washed, and then labeled with SAPE; off-rate: Cells were labeled with the appropriate DIG-conjugated label, washed, resuspended in unlabeled DIG, washed, and then labeled with SAPE. Detailed protocols are given in Supporting Methods. [3]DIG10.1L_f1_better. [4]DIG10.1L_f2_better.

Supplementary Table 10. List of oligos used for construction of DIG10_IL libraries

Oligo Name	Sequence	Comments
DIG10.1L_hr1	GTTCCAGACTACGCTCTGCAGGCTAGTGGTGGAGGAGGCTCTGGTGGAGGCGGTAGCGGAGGCGGAGGGTCGGCTAGCCATATGAATGCTAAAGAAATTGTTGTCCAC	
DIG10.1L_f1a_rc_WT	GCCCAAATCGTTTCGCGACCTTCAAAACGGGTTTTATGGCCCGGCGGTGCGTACGGATATTCCAGCACGCCTTCCGGGTGAAACAGGTCGCACCAACCACGGGCATCGCCATTTTCCAGCAGACGGAGAGCGTGGACAACAATTTCTTTAGCATTCATATGGC	
DIG10.1L_f1a_rc_variable	GC CCA AAT CGT TTC GCG ACC TTC aaa ACG GGT TTT atg gcc CGG cgg tgc GTA CGG ata TTC cag CAC GCC TTC CGG GTG aaa CAG GTC gca cca ACC ACG GGC ATC GCC ATT TTC cag CAG ACG gag agc GTG GAC AAC AAT TTC TTT AGC ATT CAT ATG GC	Highlighted positions were varied by using a mix of 96.952% correct base and 1.016% of each of the other bases
DIG10.1L_f1b_WT	GAAGGTCGCGAAACGATTTGGGCGCACATGCGTCTGTTCCCGGAATATATGACCATCCGCTTTACGGATGTCCAGTTCTACGAAACCGCCGATCCGGACCTGGCAATCGG	
DIG10.1L_f1b_variable	GAA GGT CGC GAA ACG ATT TGG GCG cac atg CGT CTG ttc CCG GAA tat atg ACC atc CGC ttt ACG GAT GTC CAG TTC TAC GAA ACC GCC GAT CCG GAC CTG GCA ATC GG	Highlighted positions were varied by using a mix of 96.952% correct base and 1.016% of each of the other bases
DIG10.1L_f2_rc_WT	CAATTTTCGCAGCTGCTTCCACACCGCCCAGTGGTTCCAGGACACGCAGCGGGTTGAAAAAGAGACGGTACAGCAGGATCTGACCGTCGCGCGTACGCAGAACAGAAATATAATCGGCCGCCAGTTTACCGCCGCTCACGGTGTGGACACCGTCACCATGAAATTCGCCGATTGCCAGGTCCGGATCG	
DIG10.1L_f2_rc_variable	C AAT TTT CGC AGC TGC TTC CAC ACC GCC cag tgg TTC CAG gac ACG CAG CGG GTT gaa AAA gag ACG gta CAG CAG gat CTG ACC GTC GCG CGT ACG cag AAC aga AAT ata ATC ggc CGC cag TTT acc GCC gct cac GGT gtg GAC acc GTC acc ATG aaa TTC GCC GAT TGC CAG GTC CGG ATC G	Highlighted positions were varied by using a mix of 96.880% correct base and 1.040% of each of the other bases
DIG10.1L_hr2	GGTGTGGAAGCAGCTGCGAAAATTGTTCAAGGCGCGGGTAGTCTCGAGGGGGGCGGATCCGAACAAAAGCTTATTTCTGAAGAGGACTTGTAATAG	
DIG10.1L_assembly_fwd	TCCAGACTACGCTCTGCAGGCTAGTG	
DIG10.1L_assembly_rev	CTATTACAAGTCCTCTTCAGAAATAAGCTTTTGTTCGG	

Supplementary Table 11. List of oligos used for next-generation sample preparation and sequencing

Primer Name	Sequence	Use
DIG10.1L_illumina_amp_f1_fwd	AATGATACGGCGACCACCGAGATCTACAC -- GGCTAGCCATATGAATGCTAAAGAAATTGTTGTCCA	All DIG10.1L_f1 DNA fwd amplification
DIG10.1L_illumina_amp_f1_rev_0	CAAGCAGAAGACGGCATACGAGAT -- GAGCGATA -- TCGGCGGTTTCGTAGAACTGGACATC	Expression-sorted DIG10.1L_f1 (input) DNA rev amplification
DIG10.1L_illumina_amp_f1_rev_1	CAAGCAGAAGACGGCATACGAGAT -- CTCGAGAA -- TCGGCGGTTTCGTAGAACTGGACATC	DIG10.1L_f1_neutral DNA rev amplification
DIG10.1L_illumina_amp_f1_rev_2	CAAGCAGAAGACGGCATACGAGAT -- ATGACACC -- TCGGCGGTTTCGTAGAACTGGACATC	DIG10.1L_f2_better DNA rev amplification
DIG10.1L_illumina_amp_f2_fwd	AATGATACGGCGACCACCGAGATCTACAC -- GATCCGGACCTGGCAATCGGCGAA	All DIG10.1L_f2 fwd amplification
DIG10.1L_illumina_amp_f2_rev_0	CAAGCAGAAGACGGCATACGAGAT -- TCGTCGAT – CTTGAACAATTTTCGCAGCTGCTTCCACAC	Expression-sorted DIG10.1L_f2 (input) DNA rev amplification
DIG10.1L_illumina_amp_f2_rev_1	CAAGCAGAAGACGGCATACGAGAT -- TAAGAATC -- CTTGAACAATTTTCGCAGCTGCTTCCACAC	DIG10.1L_f2_neutral DNA rev amplification
DIG10.1L_illumina_amp_f2_rev_2	CAAGCAGAAGACGGCATACGAGAT -- AATGACAG -- CTTGAACAATTTTCGCAGCTGCTTCCACAC	DIG10.1L_f2_better DNA rev amplification
DIG10.1L_illumina_seq_f1_fwd	GGCTAGCCATATGAATGCTAAAGAAATTGTTGTCCA	Fragment 1 fwd sequencing
DIG10.1L_illumina_seq_f1_rev	TCGGCGGTTTCGTAGAACTGGACATC	Fragment 1 rev sequencing
DIG10.1L_illumina_seq_f1_barcode	GATGTCCAGTTCTACGAAACCGCCGA	Fragment 1 barcode sequencing
DIG10.1L_illumina_seq_f2_fwd	GATCCGGACCTGGCAATCGGCGAA	Fragment 2 fwd sequencing
DIG10.1L_illumina_seq_f2_rev	CTTGAACAATTTTCGCAGCTGCTTCCACAC	Fragment 2 rev sequencing
DIG10.1L_illumina_seq_f2_barcode	GTGTGGAAGCAGCTGCGAAAATTGTTCAAG	Fragment 2 barcode sequencing

Supplementary Table 12. Deep sequencing processing statistics

Library	# Pre-filtering reads	# Post-filtering reads	# Single mutation reads
DIG10.1L_f1 expression-sorted	1,067,069	624,607	208,828
DIG10.1L_f1 neutral	348,465	104,954	29,474
DIG10.1L_f1 better	190,272	89,946	48,112
DIG10.1L_f2 expression-sorted	831,137	660,512	236,617
DIG10.1L_f2 neutral	282,451	231,410	80,837
DIG10.1L_f2 better	137,804	108,349	59,282

Supplemetary Table 13. List of oligos used in this study for knockout mutation generation.

Primer Name	Sequence[1]	Use
DIG10_V117R_rc	GACACGCAGCGGGTTGAAAAA**GCG**ACGGTACAGCAGGATCTGAC	DIG10 V117R
DIG10_Y101F_rc	GCGTACGCAGAACAGAAAT**GAA**ATCGGCCGCCAGTTTACC	DIG10 Y101F
DIG10_Y115F_rc	CAGCGGGTTGAAAAACACACG**GAA**CAGCAGGATCTGACCGTCG	DIG10 Y115F
DIG10_Y34F_rc	GTTTTATGGCCCGGCGGGGCGTACGG**GAA**TTCCAGCACGCCTTC	DIG10 Y34F
DIG5_W119R_rc	CGCAGAACGGTAATAAAATCGGCCGC**GAA**TTTACCGCCGCTCACGG	DIG5 W119R
DIG5_H58A_rc	CCAGTGCACGCAGCGGGTT**ACG**GAACACACGGTACAGCAGGATCTG	DIG5 H58A
DIG5_Y84F_rc	CGTAAAGCGCACGGTGACGTGTTCCGG**AGC**CAGACGCATGTGTG	DIG5 Y84F
DIG5_Y97F_rc	GGTAACCACACCGTCACCGTG**GAA**TTCGCCGATCGCCAGGTCCGG	DIG5 Y97F
DIG8_V86R_rc	CCGTTGCTTTCGGGGT**CCT**GTCCCATTTGAAACGCA	DIG8 V86R
DIG8_Y10F_rc	TGTTTGTTCAGTTCCGT**GAA**GTCTTTGGCAATTTCCG	DIG8 Y10F/H101A/Y 103F
DIG8_H101A_rc	CCGTTTTTGACGGTATACAG**AGC**AACACCGTCAACGGTCACACG	DIG8 H101A
DIG8_H101A_Y103F_rc	CGTTTTTGACGGT**AAA**CAG**AGC**AACACCGTCAACGGTCACACG	DIG8 Y10F/H101A/Y 103F
DIG10.3_Y34F_rc	AGCCCGGCGGTGGGTACGG**AAA**TTCCAGCACGCCTTCCGGGTG	DIG10.3 Y34F
DIG10.3_Y99F_rc	GCGTACGCCAAACGGCAATATAATC**GAA**CGCCAGTTTACCGCCGCTTG	DIG10.3 Y99F
DIG10.3_Y101F_rc	CGCGTACGCCAAACGGCAAT**GAA**ATCGTACGCCAGTTTACCG	DIG10.3 Y101F
DIG10.3_Y115F_rc	CACGCAGCGGGTTGAAAAAGAGACG**GAA**CAGCAGAATCTGACCGTC	DIG10.3 Y115F
DIG10.3_Y99F_Y101F_rc	GCGTACGCCAAACGGCAAT**GAA**ATC**GAA**CGCCAGTTTACCGCCGCTTGC	DIG10.3 Y99F/Y101F

[1]Mutagenic region shown in red.

Supplementary Table 14. List of oligos used for gene assembly of 1z1s

Primer Name	Sequence
1z1s_assembly_1	GGTGGAGGAGGCTCTGGTGGAG
1z1s_assembly_2	ATGGCTAGCCGACCCTCCGCCTCCGCTACCGCCTCCACCAGAGCCTCCTC
1z1s_assembly_3	GAGGGTCGGCTAGCCATATGAATGCTAAAGAAATTCTGGTCCACTCACTG
1z1s_assembly_4	CACGGGCATCGCCATTTTCCAGCAGACGCAGTGAGTGGACCAGAATTTCT
1z1s_assembly_5	AAATGGCGATGCCCGTGGTTGGTGCGACCTGTTTCACCCGGAAGGCGTGC
1z1s_assembly_6	GGTTTTCCAGCCCGGCGGGGCGTACGGGAATTCCAGCACGCCTTCCGGGT
1z1s_assembly_7	GCCGGGCTGGAAAACCCGTTTTGAAGGTCGCGAAACGATTTGGGCGCACA
1z1s_assembly_8	AGCGAACGGTCAGGTGTTCCGGGAACAGACGCATGTGCGCCCAAATCGTT
1z1s_assembly_9	ACACCTGACCGTTCGCTTTACGGATGTCCAGTTCTACGAAACCGCCGATC
1z1s_assembly_10	TCACCATGAAATTCGCCGATTGCCAGGTCCGGATCGGCGGTTTCGTAGAA
1z1s_assembly_11	ATCGGCGAATTTCATGGTGACGGTGTCGCTACCGTGAGCGGCGGTAAACT
1z1s_assembly_12	GCGTACGCAGAACAGAAATATAATCCTGCGCCAGTTTACCGCCGCTCACG
1z1s_assembly_13	ATATTTCTGTTCTGCGTACGCGCGACGGTCAGATCCTGCTGTACCGTGAC
1z1s_assembly_14	GAGCTTCCAGGTGACGCAGCGGGTTCCAAAAGTCACGGTACAGCAGGATC
1z1s_assembly_15	TGCGTCACCTGGAAGCTCTGGGCGGTGTGGAAGCAGCTGCGAAAATTGTT
1z1s_assembly_16	CCGCCCCCCTCGAGACTACCCGCGCCTTGAACAATTTTCGCAGCTGCTTC
1z1s_assembly_17	TCTCGAGGGGGGCGGATCCGAACAAAAGCTTATTTCTGAAGAGGACTTGT
1z1s_assembly_18	CTATTACAAGTCCTCTTCAGAAATAAGC

Supplementary Table 15. List of cloning primers.

Cloning Purpose	Primer	Sequence
General (general PCR/sequencing)	pCTCON2f	CAACAACTAGCAAAGGCAGCCCCATAAACAC
	pCTCON2r	GGAACAAAGTCGATTTTGTTACATCTACACTG
DIG10-TEV-his_6 subcloning	DIG10_TEV_fwd	GCTAGTAACCATATGAATGCTAAAGAAATTGTTGTCCACTC
	DIG10_TEV_rev_rc	GTTACTCTCGAGGCCCTGGAAGTACAGGTTCTCACTACCCGCGCCTT GAACAATTTTCG
DIG5-TEV-his_6 subcloning	DIG5_TEV_fwd	GCTAGTAACCATATGAATGCGAAAGAAATCCTGGTCCATAG
	DIG5_TEV_rev_rc	GTTACTCTCGAGGCCCTGGAAGTACAGGTTCTCAGAACCCGCGCCTT GGACAATTTTTGC
1z1s-TEV-his_6 subcloning	1z1s_TEV_fwd	GCTAGTAACCATATGAATGCTAAAGAAATTCTGGTCCACTC
DIG10.1-TEV-his_6 subcloning	DIG10_S10A_TEV_ fwd	GCTAGTAACCATATGAATGCTAAAGAAATTGTTGTCCACGCTC
	DIG10_TEV_rev_rc	GTTACTCTCGAGGCCCTGGAAGTACAGGTTCTCACTACCCGCGCCTT GAACAATTTTCG
DIG10.2-TEV-his_6 subcloning	DIG10_S10A_TEV_ fwd	GCTAGTAACCATATGAATGCTAAAGAAATTGTTGTCCACGCTC
	DIG10_TEV_rev_rc	GTTACTCTCGAGGCCCTGGAAGTACAGGTTCTCACTACCCGCGCCTT GAACAATTTTCG
DIG10.3-TEV-his_6 subcloning	DIG10_S10A_TEV_ fwd	GCTAGTAACCATATGAATGCTAAAGAAATTGTTGTCCACGCTC
	DIG10_TEV_rev_rc	GTTACTCTCGAGGCCCTGGAAGTACAGGTTCTCACTACCCGCGCCTT GAACAATTTTCG
DIG10.2_t-his_6 subcloning	DIG10_S10A_TEV_ fwd	GCTAGTAACCATATGAATGCTAAAGAAATTGTTGTCCACGCTC
	DIG10_1a_trunc_his _rev_rc	GTGGTGCTCGAGGCCCAGTGGTTCCAGGACACG
DIG10.3_t-his_6 subcloning	DIG10_S10A_TEV_ fwd	GCTAGTAACCATATGAATGCTAAAGAAATTGTTGTCCACGCTC
	DIG10_1a_trunc_his _rev_rc	GTGGTGCTCGAGGCCCAGTGGTTCCAGGACACG

Supplementary Table 16. List of oligos used for DIG10 site-saturation mutagenesis library generation (affinity maturation round 1a).

Primer Name	Sequence[1]
DIG10_SSM.1_A127_rc	GCTGCTTCCACACCGCCCAG**MNN**TTCCAGGACACGCAGCGGGTTG
DIG10_SSM.1_A37_rc	GGTTTTATGGCCCGGCGG**MNN**GTACGGATATTCCAGCACGCCTT
DIG10_SSM.1_A99_rc	CGTACGCAGAACAGAAATATAATC**MNN**CGCCAGTTTACCGCCGCTCA
DIG10_SSM.1_F119_rc	CAGGACACGCAGCGGGTT**MNN**AAACACACGGTACAGCAGGATC
DIG10_SSM.1_F58_rc	CGAACGGTCACATATTCCGG**MNN**CAGACGCATGTGCGCCCAAATC
DIG10_SSM.1_F66_rc	CGTAGAACTGGACATCCGT**MNN**GCGAACGGTCACATATTCC
DIG10_SSM.1_F84_rc	GTGGACACCGTCACCATG**MNN**TTCGCCGATTGCCAGGTC
DIG10_SSM.1_G40_rc	CGACCTTCAAAACGGGTTTTATG**MNN**CGGCGGGGCGTACG
DIG10_SSM.1_G86_rc	CACGGTGTGGACACCGTC**MNN**ATGAAATTCGCCGATTGCCAGG
DIG10_SSM.1_G88_rc	GCTCACGGTGTGGAC**MNN**GTCACCATGAAATTCGCCGATTG
DIG10_SSM.1_H41_rc	GTTTCGCGACCTTCAAAACGGGTTTT**MNN**GCCCGGCGGGGCGTAC
DIG10_SSM.1_H54_rc	CATATTCCGGGAACAGACGCAT**MNN**CGCCCAAATCGTTTCGCGAC
DIG10_SSM.1_H90_rc	GTTTACCGCCGCTCACGGT**MNN**GACACCGTCACCATGAAATTCG
DIG10_SSM.1_L11_rc	CCATTTTCCAGCAGACG**MNN**TGAGTGGACAACAATTTC
DIG10_SSM.1_L128_rc	AGCTGCTTCCACACCGCC**MNN**AGCTTCCAGGACACGCAGC
DIG10_SSM.1_L14_rc	ACCACGGGCATCGCCATTTTC**MNN**CAGACGCAGTGAGTGGACAAC
DIG10_SSM.1_L32_rc	CGTACGGATATTC**MNN**CACGCCTTCCGGGTGA
DIG10_SSM.1_L57_rc	GAACGGTCACATATTCCGGGAA**MNN**ACGCATGTGCGCCCAAATCG
DIG10_SSM.1_L97_rc	AGAACAGAAATATAATCGGCCGC**MNN**TTTACCGCCGCTCACGGTG
DIG10_SSM.1_M55_rc	CACATATTCCGGGAACAGACG**MNN**GTGCGCCCAAATCGTTTCG
DIG10_SSM.1_P38_rc	CAAAACGGGTTTTATGGCCCGG**MNN**GGCGTACGGATATTCCAGC
DIG10_SSM.1_S103_rc	GTCGCGCGTACGCAGAAC**MNN**AATATAATCGGCCGCCAGTTTACC
DIG10_SSM.1_S10_rc	TCCAGCAGACGCAG**MNN**GTGGACAACAATTTCTTTAG
DIG10_SSM.1_S93_rc	ATCGGCCGCCAGTTTACCGCC**MNN**CACGGTGTGGACACCGTCACC
DIG10_SSM.1_V117_rc	GACACGCAGCGGGTTGAAAAA**MNN**ACGGTACAGCAGGATCTGAC
DIG10_SSM.1_V124_rc	CACCGCCCAGAGCTTCCAG**MNN**ACGCAGCGGGTTGAAAAACACAC
DIG10_SSM.1_V62_rc	CATCCGTAAAGCGAACGGT**MNN**ATATTCCGGGAACAGACGCATG
DIG10_SSM.1_V64_rc	GAACTGGACATCCGTAAAGCG**MNN**GGTCACATATTCCGGGAACAG
DIG10_SSM.1_V92_rc	GCCAGTTTACCGCCGCT**MNN**GGTGTGGACACCGTCACCA
DIG10_SSM.1_W22_rc	CTTCCGGGTGAAACAGGTCGCA**MNN**ACCACGGGCATCGCCAT
DIG10_SSM.1_Y101_rc	GCGTACGCAGAACAGAAAT**MNN**ATCGGCCGCCAGTTTACC
DIG10_SSM.1_Y115_rc	CAGCGGGTTGAAAAACACACG**MNN**CAGCAGGATCTGACCGTCG
DIG10_SSM.1_Y34_rc	GTTTTATGGCCCGGCGGGGCGTACGG**MNN**TTCCAGCACGCCTTC
DIG10_SSM.1_Y61_rc	CCGTAAAGCGAACGGTCAC**MNN**TTCCGGGAACAGACGCATG

[1]Mutagenic region shown in red.

Supplementary Table 17. List of oligos for DIG10 combinatorial library generation (affinity maturation round 1b).

Primer Name	Sequence[1]
DIG10_c.2_A127AP_rc	GCTGCTTCCACACCGCCCAG**TGS**TTCCAGGACACGCAGCGGGTTG
DIG10_c.2_A37AP_rc	GGTTTTATGGCCCGGCGG**TGS**GTACGGATATTCCAGCACGCCTT
DIG10_c.2_H90H_rc	GTTTACCGCCGCTCACGGT**ATG**GACACCGTCACCATGAAATTCG
DIG10_c.2_H90TN_rc	GTTTACCGCCGCTCACGGT**GKT**GACACCGTCACCATGAAATTCG
DIG10_c.2_S10AS-L11LHQ_rc	CATCGCCATTTTCCAGCAGACG**SWGAGM**GTGGACAACAATTTCTTTAGC
DIG10_c.2_S10M-L11LHQ_rc	CATCGCCATTTTCCAGCAGACG**SWGCAT**GTGGACAACAATTTCTTTAGC
DIG10_c.2_V117ILVF_rc	GACACGCAGCGGGTTGAAAAA**GAN**ACGGTACAGCAGGATCTGAC
DIG10_c.2_V117M_rc	GACACGCAGCGGGTTGAAAAA**CAT**ACGGTACAGCAGGATCTGAC
DIG10_c.2_V62ILVF-V64ILVF_rc	CTGGACATCCGTAAAGCG**GAN**GGT**GAN**ATATTCCGGGAACAGACGC
DIG10_c.2_V62ILVF-V64M_rc	CTGGACATCCGTAAAGCG**CAT**GGT**GAN**ATATTCCGGGAACAGACGC
DIG10_c.2_V62M-V64ILVF_rc	CTGGACATCCGTAAAGCG**GAN**GGT**CAT**ATATTCCGGGAACAGACGC

[1]Mutagenic region(s) shown in red.

Supplementary Table 18. List of oligos used for DIG10.2 combinatorial library generation based on deep sequencing data (directed evolution round 3).

Primer Name	Sequence
DIG10_2a_C23SC_rc	GCCTTCCGGGTGAAACAGGTC**ASA**CCAACCACGGGCATCGCCATTTTC
DIG10_2a_F45FY_rc	CCCAAATCGTTTCGCGACCTTC**GWA**ACGGGTTTTATAGCCC
DIG10_2a_F45H_rc	CCCAAATCGTTTCGCGACCTTC**ATG**ACGGGTTTTATAGCCC
DIG10_2a_M62F_rc	CTGGACATCCGTAAAGCGGATGGT**GAA**ATATTCCGGGAACAGACGCATGTG
DIG10_2a_M62M_rc	CTGGACATCCGTAAAGCGGATGGT**CAT**ATATTCCGGGAACAGACGCATGTG
DIG10_2a_H90HY_V92VA_rc	ACCGCCGCT**TRC**GGT**ATR**AACACCGTCACCATGAAATTCGCCGATTG
DIG10_2a_H90ILF_V92VA_rc	ACCGCCGCT**TRC**GGT**GAD**AACACCGTCACCATGAAATTCGCCGATTG
DIG10_2a_A99AVIT_S103SA_L105LV_I112IF_rc	GAGACGGTACAGCAG**AAW**CTGACCGTCGCGCGTACG**GAS**AAC**GGM**AATATAATC**ARY**CGCCAGT
DIG10_2a_A99AVIT_S103SA_L105W_I112IF_rc	GAGACGGTACAGCAG**AAW**CTGACCGTCGCGCGTACG**CCA**AAC**GGM**AATATAATC**ARY**CGCCAGT
DIG10_2a_A99AVIT_S103V_L105LV_I112IF_rc	GAGACGGTACAGCAG**AAW**CTGACCGTCGCGCGTACG**GAS**AAC**AAC**AATATAATC**ARY**CGCCAGT
DIG10_2a_A99AVIT_S103V_L105W_I112IF_rc	GAGACGGTACAGCAG**AAW**CTGACCGTCGCGCGTACG**CCA**AAC**AAC**AATATAATC**ARY**CGCCAGT
DIG10_2a_A99FY_S103SA_L105LV_I112IF_rc	GAGACGGTACAGCAG**AAW**CTGACCGTCGCGCGTACG**GAS**AAC**GGM**AATATAATC**GWA**CGCCAGT
DIG10_2a_A99FY_S103SA_L105W_I112IF_rc	GAGACGGTACAGCAG**AAW**CTGACCGTCGCGCGTACG**CCA**AAC**GGM**AATATAATC**GWA**CGCCAGT
DIG10_2a_A99FY_S103V_L105LV_I112IF_rc	GAGACGGTACAGCAG**AAW**CTGACCGTCGCGCGTACG**GAS**AAC**AAC**AATATAATC**GWA**CGCCAGT
DIG10_2a_A99FY_S103V_L105W_I112IF_rc	GAGACGGTACAGCAG**AAW**CTGACCGTCGCGCGTACG**CCA**AAC**AAC**AATATAATCGWACGCCAGT
DIG10_2a_V124VF_P127ILV_rc	CAATTTTCGCAGCTGCTTCCACACCGCCCAG**GAB**TTCCAG**AAM**ACGCAGCGGGTTG
DIG10_2a_V124VF_P127P_rc	CAATTTTCGCAGCTGCTTCCACACCGCCCAG**CGG**TTCCAG**AAM**ACGCAGCGGGTTG

[1]Mutagenic region(s) shown in red.

Supplementary Table 19. Input DIG10.1L library deep sequencing counts

Position	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y
10	430743	21	595	24	34	1215	5	91	0	29	4284	15	232	0	9	1405	1756	2515	0	20
11	9	21	13	2	2117	18	1477	1568	0	430743	24	21	652	33	1001	33	20	1382	0	29
14	9	1	0	5	46	6	27	36	13	430743	1396	0	325	1423	1214	43	19	1240	31	0
22	25	2965	0	5	37	1367	0	0	1	1539	20	0	18	25	521	1571	32	26	430743	21
23	9	430743	25	0	1177	1079	16	15	0	27	3	15	28	1	1338	3109	16	9	949	1577
26	3	615	6	0	430743	1	6	1600	0	3987	18	4	5	0	2	371	8	1030	11	1432
32	6	0	0	0	63	3	11	49	3	430743	1284	0	388	1292	422	18	25	1208	6	0
34	13	1499	1187	24	1235	4	1634	17	26	38	2	1406	17	28	9	1192	17	10	4	430743
37	430743	0	6	965	0	942	0	14	7	62	0	0	1695	9	7	1220	1391	1625	0	0
38	832	0	0	8	0	15	34	0	17	1738	32	0	430743	1182	985	1732	1225	17	24	1
40	1744	1788	2092	30	23	430743	27	31	0	9	0	33	29	0	1740	43	16	1592	11	9
41	8	18	1098	19	9	7	430743	3	22	1150	0	1459	1310	2517	1078	31	31	20	1	1411
45	21	1160	9	0	430743	7	20	1465	0	3922	5	6	1	0	15	1366	16	1266	8	1435
54	10	14	1639	43	25	13	430743	26	15	2219	0	1454	782	3068	1217	34	13	31	0	1716
55	7	0	0	9	22	15	1	3548	805	2393	430743	21	5	3	1229	29	877	1452	10	0
58	19	1750	16	0	430743	37	8	1368	0	4093	20	19	10	1	13	920	13	2225	19	993
61	11	1746	1310	15	1226	18	1012	9	20	37	0	1007	9	13	7	956	6	11	8	430743
62	12	0	0	5	27	17	0	3581	986	2312	430743	13	12	6	1039	27	806	1445	7	0
64	14	1	10	0	1506	15	6	430743	19	1207	1992	1068	29	1	24	833	960	1486	0	23
66	8	1297	4	0	430743	5	4	1197	0	3408	10	12	9	0	7	1047	15	1403	2	1187
84	11	2433	9	5	461352	25	22	2422	0	7953	38	23	37	1	20	2246	42	2070	37	2685
86	2123	1520	1987	26	10	461352	24	11	0	13	0	25	24	1	1553	1604	8	1901	17	27
88	2244	2021	2006	38	26	461352	19	26	1	16	0	32	21	1	1963	1972	31	1737	17	32
90	25	21	1800	38	35	14	461352	15	43	1696	5	1762	1777	3307	1915	46	20	30	0	3091
92	1424	0	52	1464	32	1310	0	66	7	3209	1727	0	12	18	11	15	24	461352	36	0
93	22	1822	27	0	18	1541	10	1846	61	14	19	2129	19	0	5929	461352	2061	23	16	29
95	2037	1995	2213	32	22	461352	38	42	3	23	1	41	21	0	2789	2243	27	1871	25	31
97	15	1	0	21	63	12	47	67	25	461352	1670	0	1653	1682	1485	33	13	1542	44	9
99	461352	27	2091	55	41	63	27	25	0	17	0	33	1620	0	34	1930	2075	2163	0	36
101	39	1844	1971	15	1455	13	1661	13	35	37	0	1610	15	31	36	1651	9	20	16	461352
103	2018	1454	18	0	1340	6	14	18	3	65	0	8	1873	0	11	461352	1841	10	19	1323
105	18	1	0	5	31	7	54	72	15	461352	1522	0	801	1621	1349	10	28	1563	12	0
112	13	9	3	0	1509	5	16	461352	9	1514	1668	776	15	0	21	958	1692	1643	0	1
115	14	1678	1548	9	1746	5	1671	31	33	65	0	1937	6	37	15	1750	26	10	15	461352
117	11	19	12	0	1594	7	1865	1663	1	461352	14	11	1667	46	1266	55	19	1876	0	19
119	39	1972	15	0	461352	9	41	1848	0	4887	13	23	27	0	38	2125	20	2218	10	1825
124	1792	41	2026	33	1524	1744	13	1598	4	1659	17	11	23	0	13	37	22	461352	10	30
127	11711	0	0	5	0	23	36	19	21	1192	0	0	461352	1424	1370	1421	1460	14	1	0
128	34	0	0	25	58	11	35	31	21	461352	1650	0	1953	1809	1692	14	32	1598	17	0

Supplementary Table 20. Selected DIG10.1L library deep sequencing counts

	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y
10	24648	0	14	0	0	22	0	0	0	0	79	0	2	0	0	15	34	33	0	0
11	0	0	0	0	37	0	23	21	0	24648	0	0	11	0	23	0	0	25	0	0
14	0	0	0	0	1	1	0	2	0	24648	17	0	7	20	18	0	0	42	1	0
22	0	36	0	0	1	20	0	0	0	23	0	0	0	0	7	20	1	0	24648	2
23	9	24648	0	0	8	43	0	0	0	0	0	0	0	0	26	6971	1	0	24	26
26	0	13	0	0	24648	0	1	31	0	69	0	0	0	0	0	6	1	15	0	426
32	0	0	0	0	0	0	0	0	0	24648	19	0	6	21	8	1	0	25	0	0
34	0	19	28	0	20	2	25	0	0	0	0	26	2	0	0	17	0	0	0	24648
37	24648	0	18	4286	1	50	0	0	44	2	0	0	8192	5	8	38	564	35	0	1
38	9	0	0	0	0	0	0	0	0	17	0	0	24648	33	10	40	15	2	0	0
40	59	38	38	0	1	24648	1	0	0	5	0	1	1	0	51	0	0	28	0	0
41	0	0	11	0	45	0	24648	49	2	5730	0	27	38	46	8	1	6	149	3	8023
45	0	19	0	0	24648	0	23	35	0	64	7	2	0	1	0	27	10	24	0	4967
54	0	4	26	0	1	0	24648	2	0	43	0	35	21	63	27	0	1	1	0	14
55	0	0	0	0	2	0	0	1293	14	47	24648	0	0	1	20	0	23	32	0	0
58	1	25	0	0	24648	0	0	97	0	63	0	0	1	0	0	18	0	2005	2	18
61	0	47	11	0	59	0	22	0	0	0	0	16	0	0	0	11	0	0	5	24648
62	0	0	0	0	0	0	0	66	22	79	24648	0	0	0	14	0	18	18	0	0
64	0	0	0	0	85	1	0	24648	1	28	27	33	2	0	0	14	14	27	0	0
66	0	21	0	0	24648	0	0	16	0	55	0	0	0	0	0	11	0	18	0	10
84	0	14	1	0	9447	0	0	23	0	43	0	0	0	0	0	19	0	22	1	36
86	18	10	18	0	0	9447	0	0	0	0	0	0	0	0	6	12	0	14	0	0
88	89	28	7	0	1	9447	0	19	0	16	0	2	0	0	17	35	0	18	0	0
90	33	5	12	0	88	0	9447	68	0	5678	22	17	33	182	88	3	44	79	0	10024
92	3577	0	5	1099	1	89	0	2	0	724	262	0	3	11	7	7	2	9447	2	0
93	0	14	0	0	0	18	0	16	1	0	0	15	0	0	79	9447	27	0	0	0
95	23	10	43	1	0	9447	0	0	0	0	0	0	0	0	31	24	0	15	2	2
97	0	0	0	0	4	0	0	3	0	9447	255	0	11	13	10	1	1	10	3	0
99	9447	1	19	0	140	6	0	5	0	3	0	3	23	0	0	122	7935	8584	2	133
101	0	12	17	0	4	0	14	0	0	0	0	19	0	0	1	9	0	1	0	9447
103	1926	955	3	0	24	6	0	0	0	6	0	0	19	0	0	9447	49	12	1	15
105	2	0	0	0	13	0	0	85	1	9447	46	0	14	15	7	1	1	3325	8	0
112	0	0	0	0	2587	0	0	9447	0	111	14	9	1	0	1	11	21	42	0	0
115	1	10	14	0	13	0	23	0	0	0	0	31	0	0	0	30	0	0	0	9447
117	0	0	0	0	14	0	6	25	0	9447	2	0	15	0	8	3	0	15	0	1
119	0	19	0	0	9447	0	0	15	0	35	0	0	0	0	1	14	0	21	0	10
124	12	0	10	1	2955	14	0	64	0	27	0	0	1	0	0	1	0	9447	0	0
127	110	0	0	0	0	0	3	30	0	3378	0	0	9447	27	22	86	24	33	3	0
128	0	0	0	2	0	1	1	20	1	9447	34	0	64	171	189	0	7	826	0	0

Supplementary Table 21. Amino acid sequences of relevant DIG-binding protein constructs

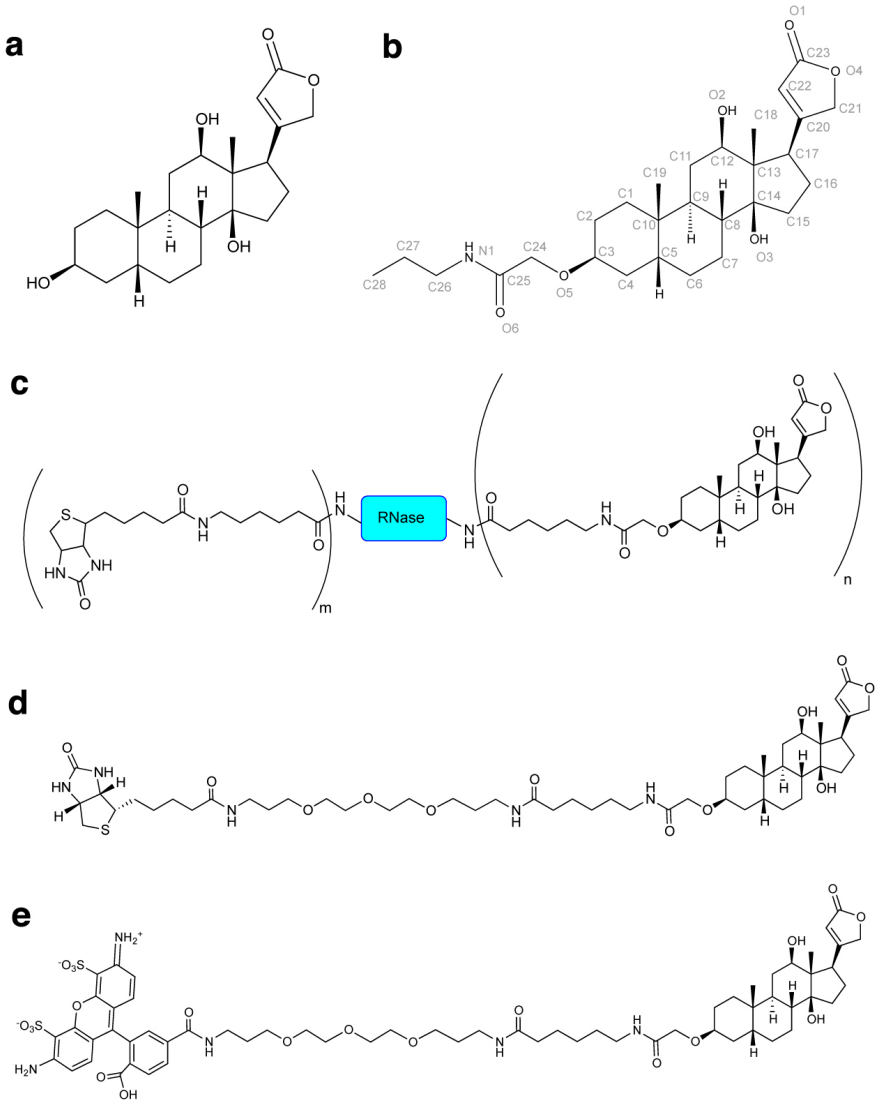
Variant	Sequence	Calculated ε (M^{-1} cm^{-1})
DIG5-TEV-his_6	MNAKEILVHSLRLLENGDARGWCDLFHPEGVLEFPYAPPGWKTRFEGRETIWAHMRL HPEHVTVRFTDVQFYETADPDLAIGEYHGDGVVTVSGGKYAADFITVLRTRDGQILL YRVFWNPLRALEAAGGVEAAAKIVQGAGSENLYFQGLEHHHHHH	30,940
DIG10-TEV-his_6	MNAKEIVVHSLRLLENGDARGWCDLFHPEGVLEYPYAPPGHKTRFEGRETIWAHMR LFPEYVTVRFTDVQFYETADPDLAIGEFHGDGVHTVSGGKLAADYISVLRTRDGQILL YRVFFNPLRVLEALGGVEAAAKIVQGAGSENLYFQGLEHHHHHH	21,430
DIG10.1-TEV-his_6	MNAKEIVVHALRLLENGDARGWCDLFHPEGVLEYPYAPPGHKTRFEGRETIWAHMR LFPEYMTIRFTDVQFYETADPDLAIGEFHGDGVHTVSGGKLAADYISVLRTRDGQILL YRLFFNPLRVLEPLGGVEAAAKIVQGAGSENLYFQGLEHHHHHH	21,430
DIG10.2-TEV-his_6	MNAKEIVVHALRLLENGDARGWCDLFHPEGVLEYPYPPPGYKTRFEGRETIWAHMR LFPEYMTIRFTDVQFYETADPDLAIGEFHGDGVHTVSGGKLAADYISVLRTRDGQILL YRLFFNPLRVLEPLGGVEAAAKIVQGAGSENLYFQGLEHHHHHH	22,920
DIG10.3-TEV-his_6	MNAKEIVVHALRLLENGDARGWSDLFHPEGVLEYPYPPPGYKTRFEGRETIWAHMRL FPEYMTIRFTDVQFYETADPDLAIGEFHGDGVLTASGGKLAYDYIAVWRTRDGQILLY RLFFNPLRVLEPLGGVEAAAKIVQGAGSENLYFQGLEHHHHHH	29,910
DIG10.3 Y34F-TEV-his_6	MNAKEIVVHALRLLENGDARGWSDLFHPEGVLEFPYPPPGYKTRFEGRETIWAHMRL FPEYMTIRFTDVQFYETADPDLAIGEFHGDGVLTASGGKLAYDYIAVWRTRDGQILLY RLFFNPLRVLEPLGGVEAAAKIVQGAGSENLYFQGLEHHHHHH	28,420
DIG10.3 Y99F-TEV-his_6	MNAKEIVVHALRLLENGDARGWSDLFHPEGVLEYPYPPPGYKTRFEGRETIWAHMRL FPEYMTIRFTDVQFYETADPDLAIGEFHGDGVLTASGGKLAFDYIAVWRTRDGQILLY RLFFNPLRVLEPLGGVEAAAKIVQGAGSENLYFQGLEHHHHHH	28,420
DIG10.3 Y101F-TEV-his_6	MNAKEIVVHALRLLENGDARGWSDLFHPEGVLEYPYPPPGYKTRFEGRETIWAHMRL FPEYMTIRFTDVQFYETADPDLAIGEFHGDGVLTASGGKLAYDFIAVWRTRDGQILLY RLFFNPLRVLEPLGGVEAAAKIVQGAGSENLYFQGLEHHHHHH	28,420
DIG10.3 Y115F-TEV-his_6	MNAKEIVVHALRLLENGDARGWSDLFHPEGVLEYPYPPPGYKTRFEGRETIWAHMRL FPEYMTIRFTDVQFYETADPDLAIGEFHGDGVLTASGGKLAYDYIAVWRTRDGQILLF RLFFNPLRVLEPLGGVEAAAKIVQGAGSENLYFQGLEHHHHHH	28,420
DIG10.3 Y99F/Y101F-TEV-his_6	MNAKEIVVHALRLLENGDARGWSDLFHPEGVLEYPYPPPGYKTRFEGRETIWAHMRL FPEYMTIRFTDVQFYETADPDLAIGEFHGDGVLTASGGKLAFDFIAVWRTRDGQILLY RLFFNPLRVLEPLGGVEAAAKIVQGAGSENLYFQGLEHHHHHH	26,930
DIG10.3 Y34F/Y99F/Y101F-TEV-his_6	MNAKEIVVHALRLLENGDARGWSDLFHPEGVLEFPYPPPGYKTRFEGRETIWAHMRL FPEYMTIRFTDVQFYETADPDLAIGEFHGDGVLTASGGKLAFDFIAVWRTRDGQILLY RLFFNPLRVLEPLGGVEAAAKIVQGAGSENLYFQGLEHHHHHH	25,440
DIG10.3 Y34F/Y99F/Y101F/Y115F-TEV-his_6	MNAKEIVVHALRLLENGDARGWSDLFHPEGVLEFPYPPPGYKTRFEGRETIWAHMRL FPEYMTIRFTDVQFYETADPDLAIGEFHGDGVLTASGGKLAFDFIAVWRTRDGQILLF RLFFNPLRVLEPLGGVEAAAKIVQGAGSENLYFQGLEHHHHHH	23,950
1z1s-TEV-his_6	MNAKEILVHSLRLLENGDARGWCDLFHPEGVLEFPYAPPGWKTRFEGRETIWAHMRL FPEHLTVRFTDVQFYETADPDLAIGEFHGDGVATVSGGKLAQDYISVLRTRDGQILLY RDFWNPLRHLEALGGVEAAAKIVQGAGSENLYFQGLEHHHHHH	29,450
DIG10.2_t-his_6	MNAKEIVVHALRLLENGDARGWCDLFHPEGVLEYPYPPPGYKTRFEGRETIWAHMR LFPEYMTIRFTDVQFYETADPDLAIGEFHGDGVHTVSGGKLAADYISVLRTRDGQILL YRLFFNPLRVLEPLGLEHHHHHH	21,430
DIG10.3_t-his_6	MNAKEIVVHALRLLENGDARGWSDLFHPEGVLEYPYPPPGYKTRFEGRETIWAHMRL FPEYMTIRFTDVQFYETADPDLAIGEFHGDGVLTASGGKLAYDYIAVWRTRDGQILLY RLFFNPLRVLEPLGLEHHHHHH	28,420

Supplementary Table 22. Crystallization data collection and refinement statistics

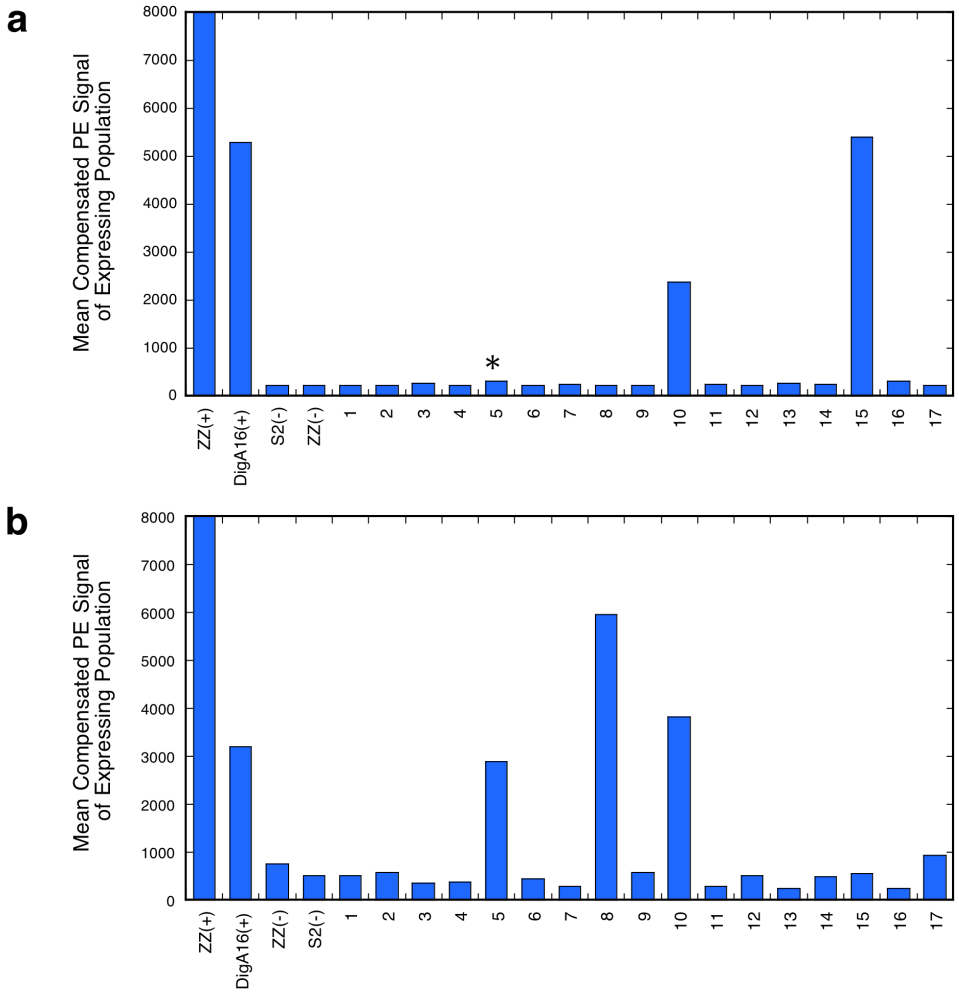
	DIG10.2$_t$-his$_6$	DIG10.3$_t$-his$_6$
Data collection		
Space group	$P6_5$	C2
Cell dimensions		
a, *b*, *c* (Å)	74.4, 74.4, 161.1	132.8, 91.0, 110.1
α, β, γ (°)	90,90,120	90, 92.7, 90
Resolution (Å)	50.0-2.05(2.12-2.05)	45.5-3.2 (3.31-3.20)
R_{sym} or R_{merge}	0.101(0.336)	0.109(0.515)
$I/\sigma I$	23.7(2.70)	17.8(3.86)
Completeness (%)	96.3	97.4
Redundancy	15.8(4.8)	7.3(7.1)
Refinement		
Resolution (Å)	2.05	3.2
No. reflections	30440	21172
$R_{work/}$ R_{free}	21.1/24.6	26.2/32.0
No. atoms		
Protein	4036	7050
Ligand/ion	152	252
Water	86	0
B-factors		
Protein	29.24	73.73
Ligand/ion	29.18	60.70
Water	27.34	N/A
R.m.s deviations		
Bond lengths (Å)	0.0083	0.0106
Bond angles (°)	1.153	1.736

*Highest resolution shell is shown in parenthesis.

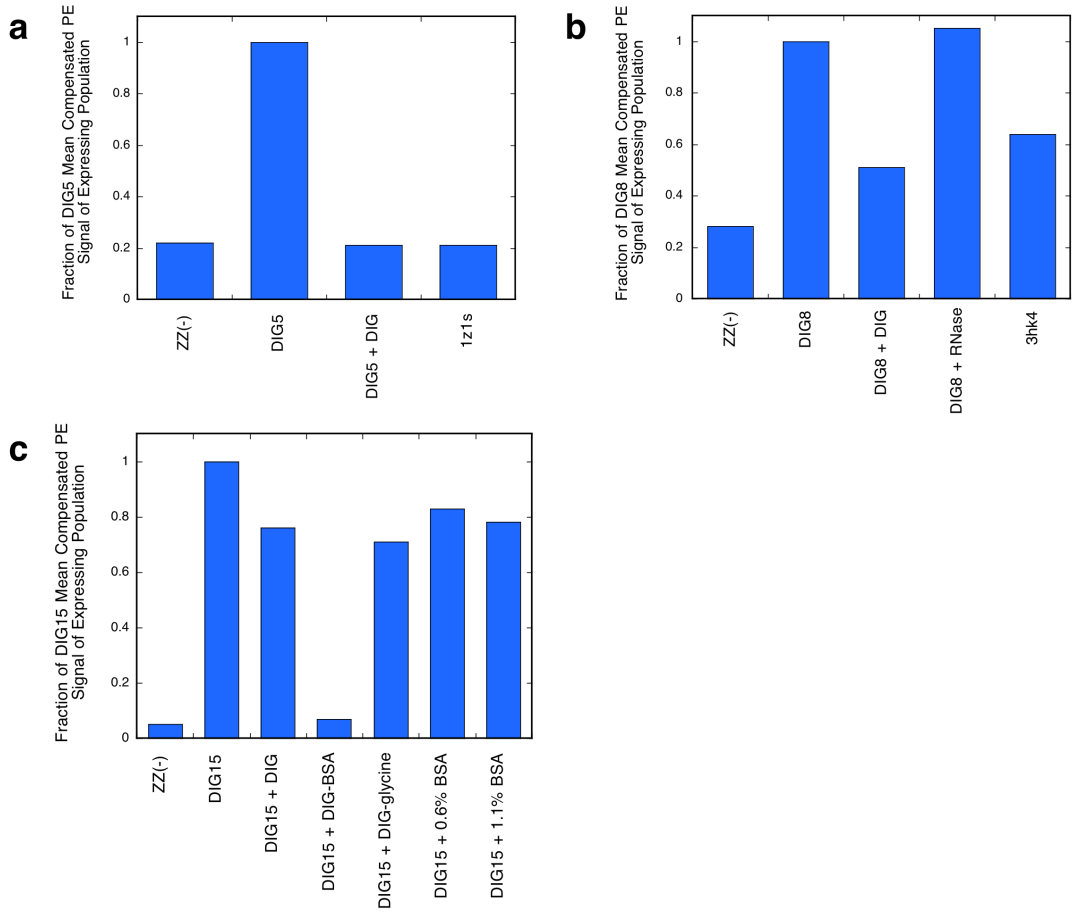
Supplementary Figures



Supplementary Figure 1 | DIG–based materials used in this study. a, Digoxigenin. **b,** A linker-modified digoxigenin used in Rosetta design calculations. Atom names are given in gray. **c,** Schematic of DIG-RNase-biotin. **d,** DIG-PEG_3-biotin. **e,** DIG-PEG_3-Alexa488.

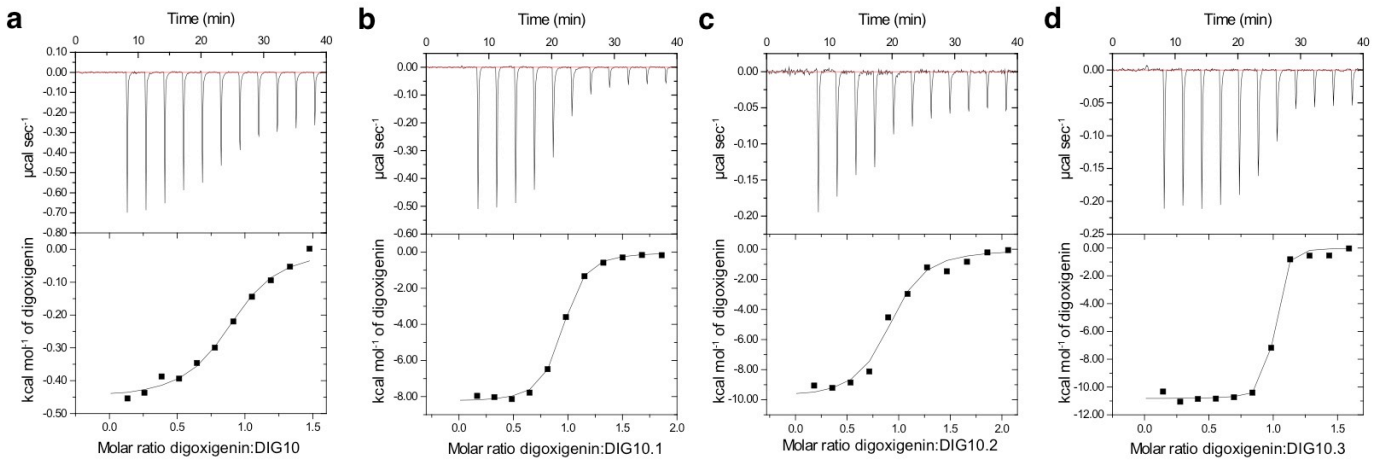


Supplementary Figure 2 | Experimental characterization of computationally designed DIG binders by yeast surface display. a, Compensated mean binding (PE) signals of the expressing populations of yeast cells displaying computationally designed proteins on their cell surfaces. Binding was interrogated by labeling cells with a pre-incubated mixture of 2.7 μM biotinylated DIG-functionalized BSA (~10 DIG/BSA) and phycoerythrin (PE)-conjugated streptavidin. Cell populations shown are an anti-DIG antibody serving as a positive control for binding (ZZ(+)), an engineered DIG-binding lipocalin (DigA16(+)), two negative controls for binding (ZZ(-) and S2(-)), and designed proteins DIG1 through DIG17. DIG10 and DIG15 show strong binding signals. DIG5 shows a reproducible signal that is slightly above background levels (starred). **b,** Binding was interrogated by labeling cells with a pre-incubated mixture of 2.7 μM biotinylated DIG-functionalized RNase (~6 DIG/RNase) and phycoerythrin (PE)-conjugated streptavidin. DIG5, DIG8, and DIG10 show strong binding signals. DIG10 and DIG5 bind to both labels.

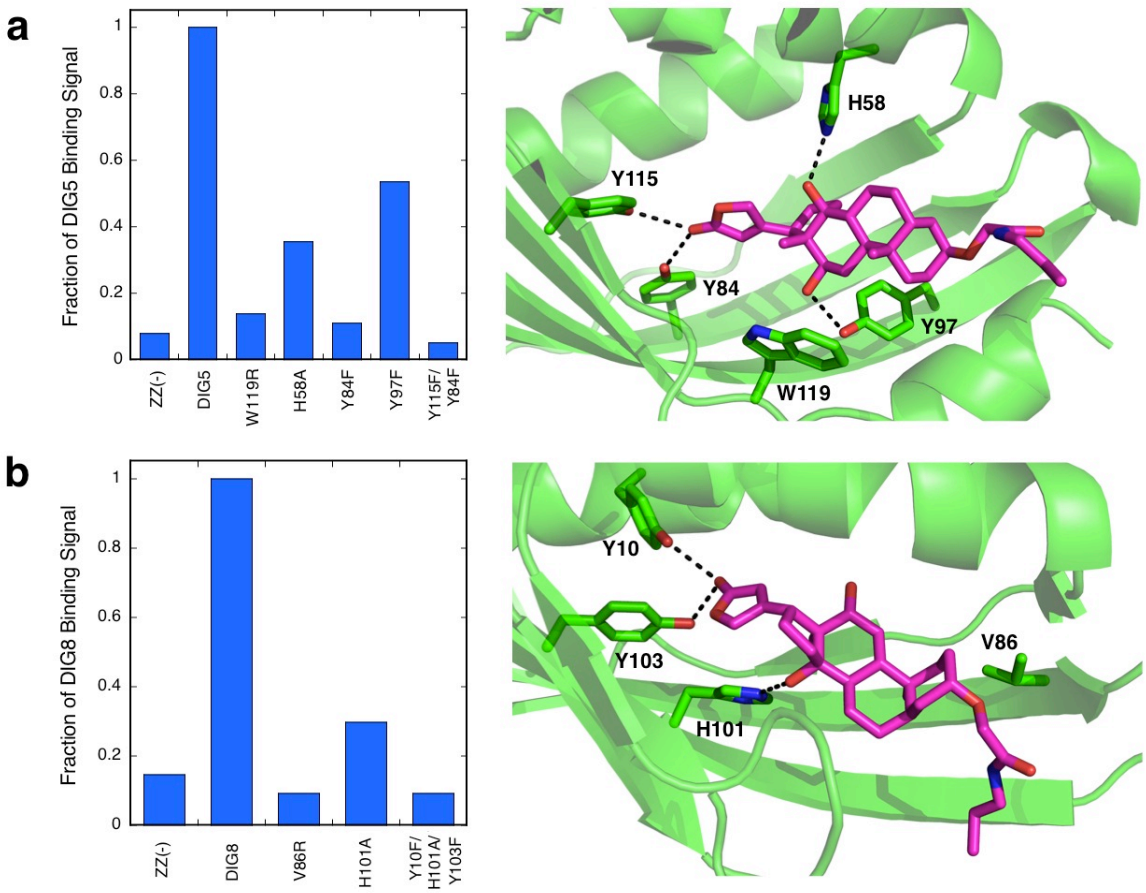


Supplementary Figure 3 | Experimental yeast surface display competition assays of DIG5, DIG8, and DIG15. a, Compensated mean binding (PE) signals of the expressing populations of yeast cells displaying DIG5 and DIG5 scaffold 1z1s. 1z1s (structural genomics target PA3332) is a protein of unknown function from *Pseudomonas aeruginosa* and has no functionally characterized homologs with > 20% sequence identity. 1z1s belongs to the nuclear transport 2 (NTF2)-like structural superfamily, a functionally diverse fold class of which the steroid-metabolizing enzyme ketosteroid isomerase is also a member. Cells were labeled with 2.7 μM DIG-RNase-biotin and SAPE. In the presence of 790 μM unlabeled DIG, the signal is reduced to that of the negative control ZZ(-), revealing that binding is specific for DIG and not other assay components. Scaffold 1z1s does not show a binding signal, suggesting that binding is mediated by the designed interface. **b,** Compensated mean binding (PE) signals of the expressing populations of yeast cells displaying DIG8 and DIG8 scaffold 3hk4. Cells were labeled with 2.7 μM DIG-RNase-biotin and SAPE. In the presence of 790 μM unlabeled DIG, the DIG8 signal is reduced by half, suggesting that binding is likely specific for DIG. The binding signal is not reduced to background levels probably because binding is weak and the unlabeled DIG is not present in high enough concentration to overcome the avidity affects of the DIG-RNase-biotin label. This explanation is corroborated by the observation that unlabeled 130 μM RNase does not affect the DIG8 signal. Scaffold 3hk4, NTF2-like superfamily member and structural genomics target MLR7391 (PDB ID 3hk4) from *Mesorhizobium loti*, does bind to DIG-RNase-biotin with a weaker (and more avidity-based) signal than DIG8. 3hk4 has not been functionally characterized. **c,** Compensated mean binding (PE) signals of the

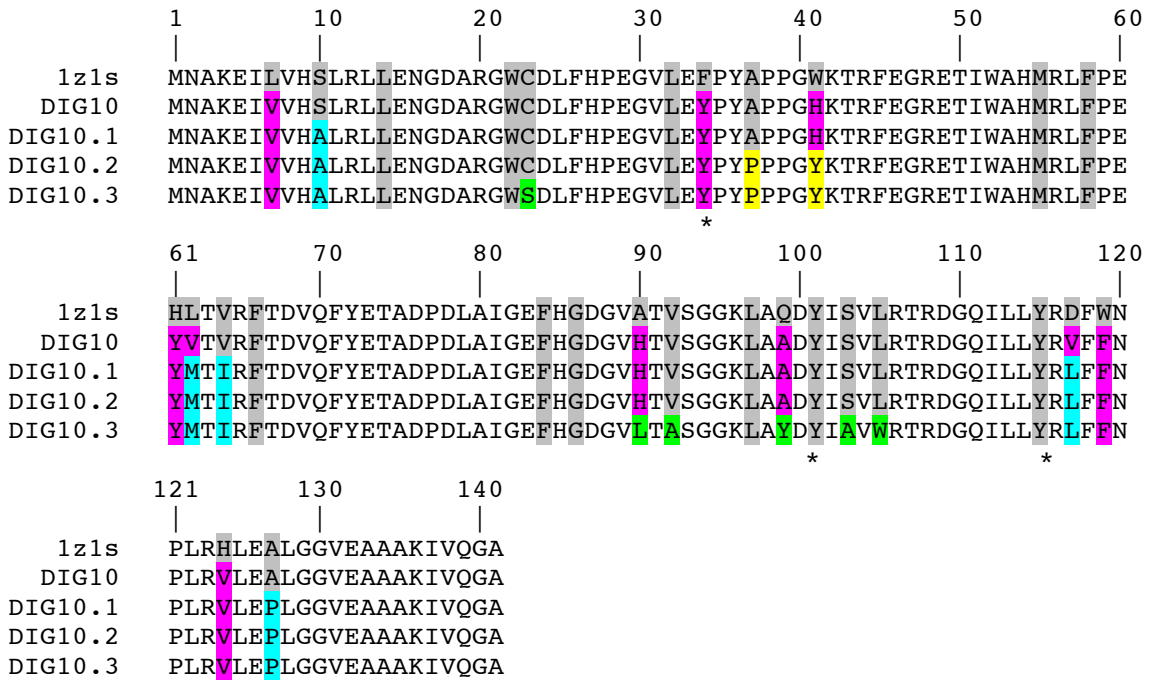
expressing populations of yeast cells displaying DIG15. Cells were labeled with 2.7 μM DIG-BSA-biotin and SAPE. In the presence of 1.6 mM unlabeled DIG, the binding signal is only reduced slightly. Similar effects are seen with ~1.6 mM DIG-linker conjugate (DIG-NHS ester reacted with glycine) and BSA. However, the signal was completely reduced upon incubation with 18 μM DIG-BSA. The signal is slightly reduced in the presence of additional BSA (0.6% or 1.1% BSA). Either the DIG15 binding signal is not reduced to background levels in the presence of unlabeled DIG because binding is weak and the amount of competitor in the assay is not enough to overcome the avidity affects of the DIG-RNase-biotin label or because the design recognizes both DIG and BSA non-specifically. Due to these complications, DIG15 was not characterized further.



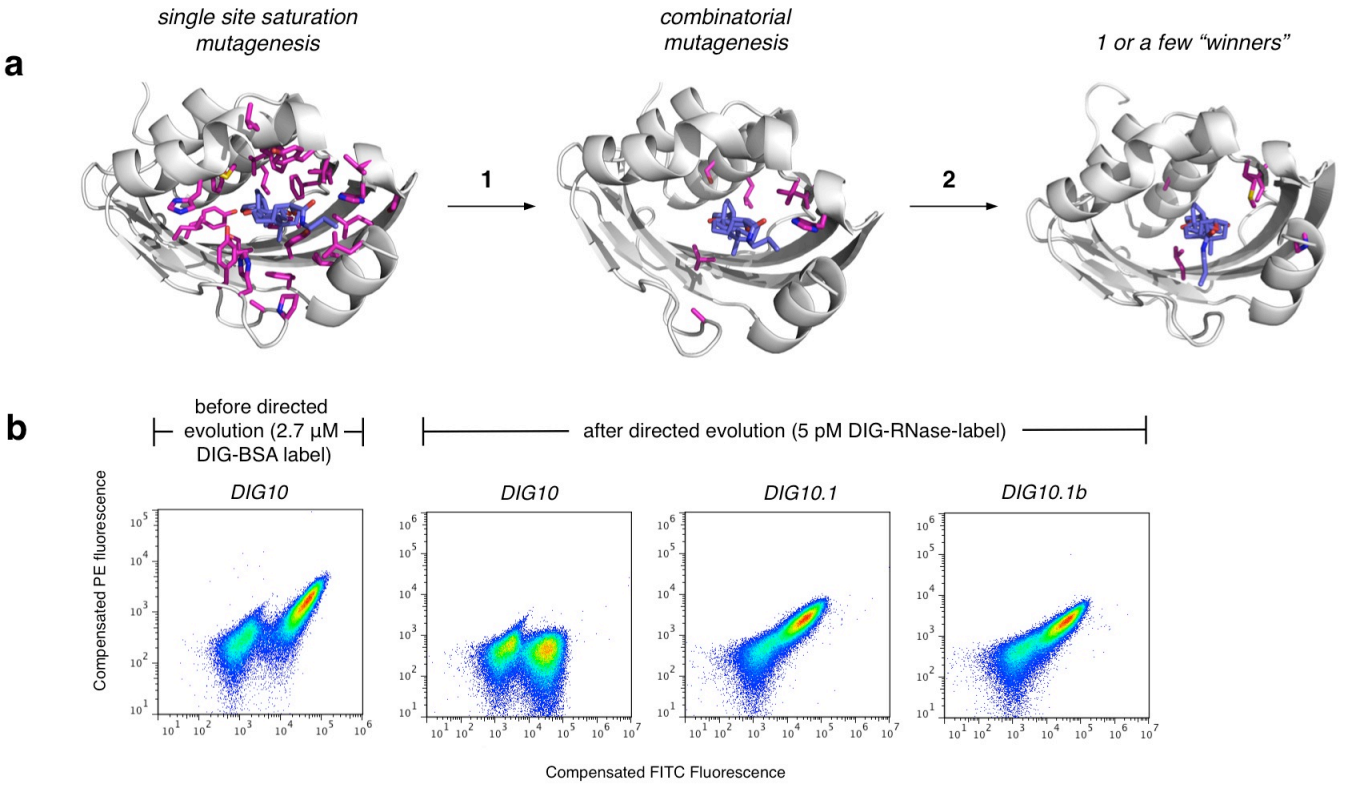
Supplementary Figure 4 | Representative DIG-Binding isotherms for DIG10 variants. DIG10 (a), DIG10.1 (b), DIG10.2 (c), and DIG10.3 (d) are shown. For DIG10.3, the determined ΔG (and K_d) is an upper limit of this value, because the binding isotherm is too steep to be fit reliably.



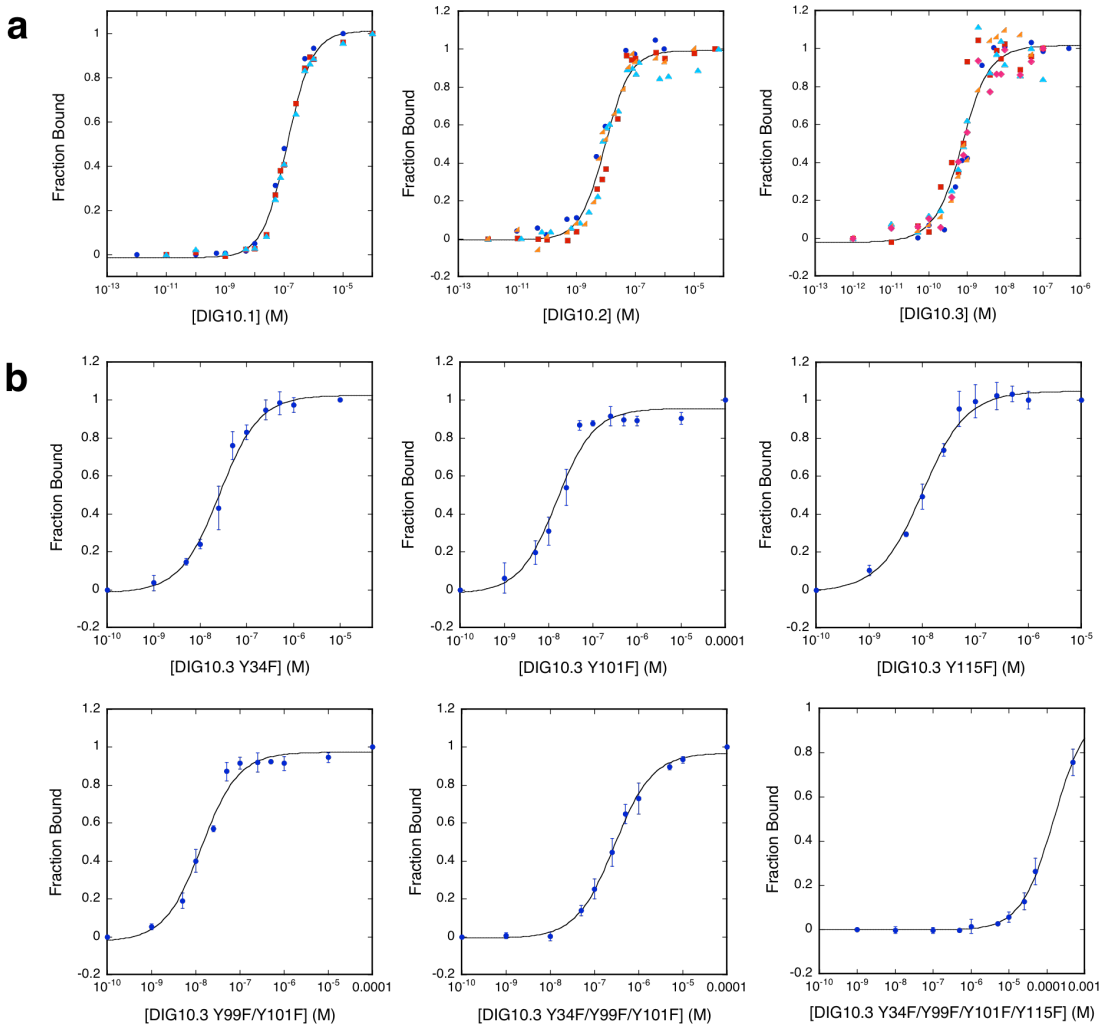
Supplementary Figure 5 | Yeast surface display knockout mutagenesis studies of DIG5 and DIG8. a, Functional substitutions of DIG5 key modeled interacting residues leads to expressing-population compensated mean binding (PE) signals that are reduced relative to DIG5 (left panel). Cells were labeled with 1.5 μM DIG-RNase-biotin and SAPE. Mutation of binding site residue Trp119 to larger Arg indicates that DIG binds in the intended pocket of the computational model (right panel). Mutation of hydrogen bonding residue Tyr84 to Phe alone and in combination with the Tyr115Phe substitution leads to a binding signal that is reduced to background negative control (ZZ(-)) levels, confirming that this residue is necessary for binding. Mutation of His58 and Tyr97, which make hydrogen bonds in the computational model, to Ala and Phe, respectively, also lead to reduced binding signals. **c,** Functional substitutions of DIG8 key modeled interacting residues leads to expressing-population compensated mean binding (PE) signals that are reduced relative to DIG8 (left panel). Cells were labeled with 2.7 μM DIG-RNase-biotin and SAPE. Mutation of binding site residue Val86 to larger Arg indicates that DIG binds in the intended pocket of the computational model (right panel). Simultaneous mutation of hydrogen bonding residues Tyr10, His101, and Tyr103 to Phe, Ala, and Phe, respectively, leads to a binding signal that is reduced to background negative control (ZZ(-)) levels, confirming that the combination of these three residues is critical for binding. Mutation of His101 to Ala also leads to a reduced binding signal. Despite the observation that the DIG8 scaffold PDB ID 3hk4 does bind to DIG-RNase (Supplementary Fig. 3), these results indicate that the designed interface could contribute to binding of the design.



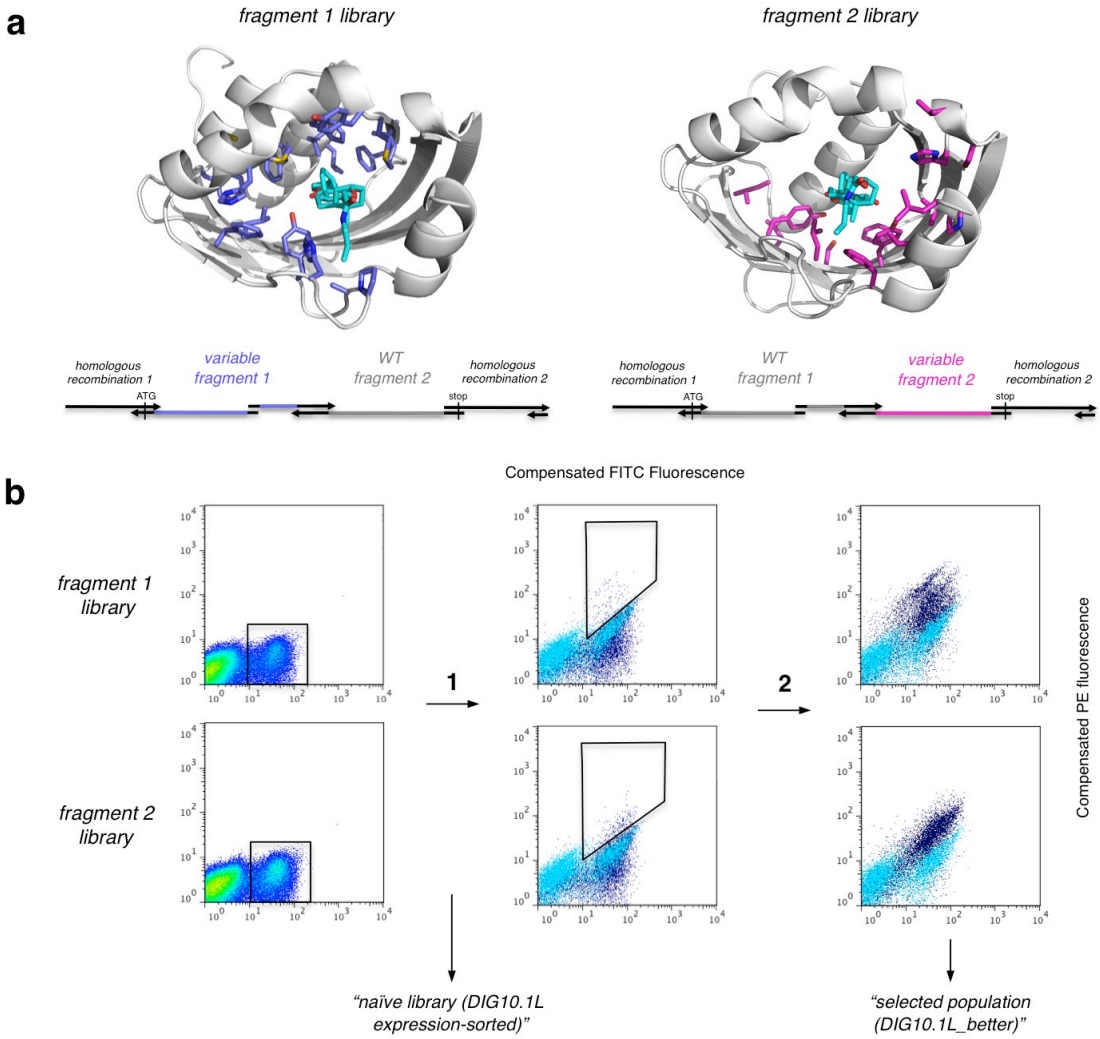
Supplementary Figure 6 | Sequence Alignment of DIG10-based Designs. Binding site residues are highlighted. Residues in magenta represent designed amino acids that differ from the scaffold 1z1s. Cyan, yellow, and green indicate residues that evolved during affinity maturation to yield DIG10.1, DIG10.2, and DIG10.3, respectively. The three hydrogen-bonding tyrosines are marked with a star.



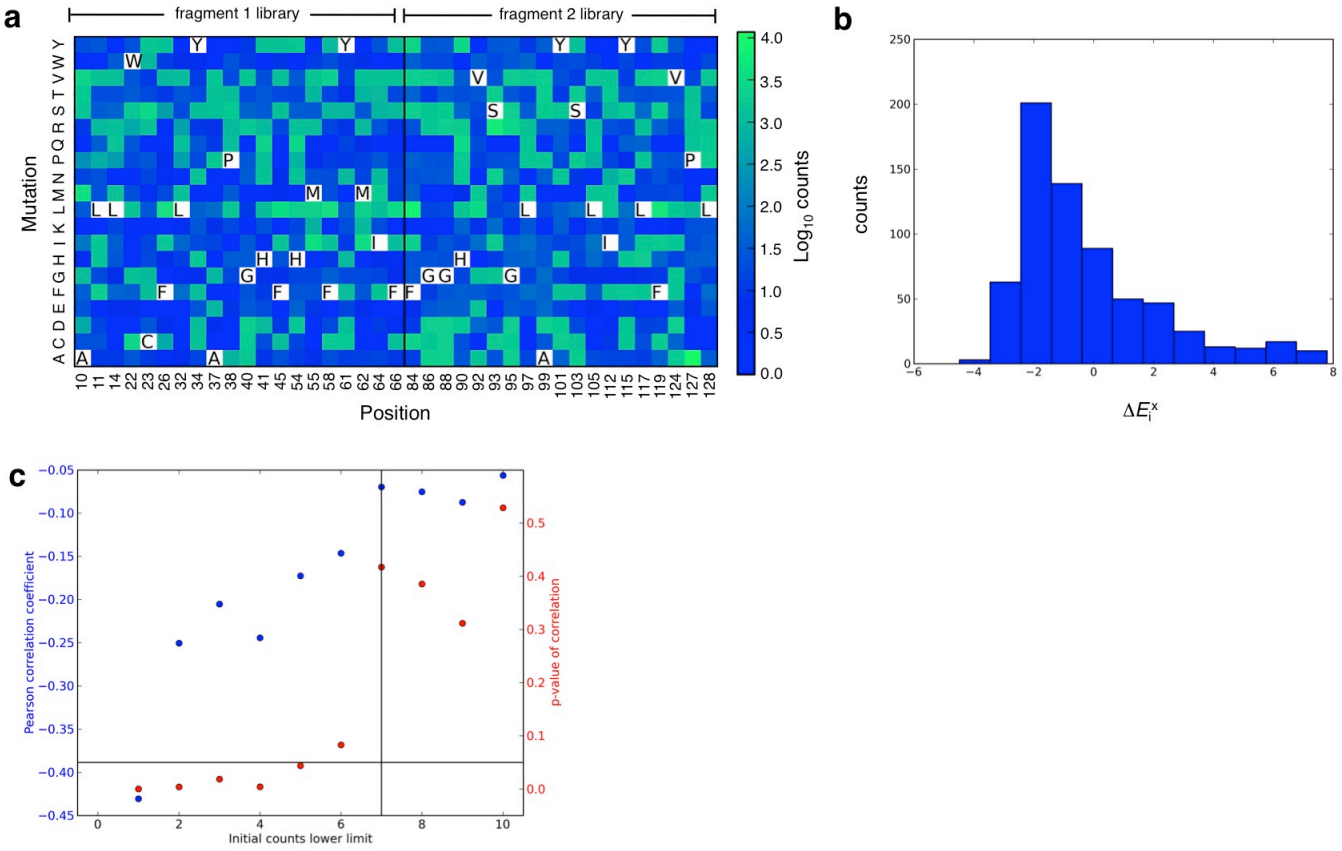
Supplementary Figure 7 | Affinity maturation of DIG10: Round 1. a, Strategy for round 1 affinity maturation of DIG10. A single site-saturation mutagenesis library was constructed by mutating each of 34 positions in the binding pocket (magenta sticks, left panel) to all other amino acids by Kunkel mutagenesis with degenerate NNK primers. After several rounds of selections with highly avid DIG-BSA-biotin and SAPE (step 1), eight positions were identified for which mutations lead to improved binding. Beneficial mutations, chemically similar residue types, and the DIG10 "wild type" amino acids at these positions (magenta sticks, middle panel) were combined combinatorially by Kunkel mutagenesis with degenerate primers. Following several rounds of increasingly stringent selections with DIG-BSA-biotin or DIG-RNase and SAPE (step 2), two variants, DIG10.1, and DIG10.1b, were identified, each having five mutations from DIG10 (magenta sticks, right panel). **b,** Flow cytometric analysis of cells expressing DIG10, DIG10.1, and DIG10.1b. DIG10 shows a strong binding signal when labeled with a pre-incubated mixture of 2.7 μM DIG-BSA-biotin and SAPE but not when subjected to a more stringent multistep labeling procedure in which cells were first labeled with 5 pM DIG-RNase and then with SAPE. DIG10.1 and DIG10.1b show strong binding signals from the latter labeling procedure, however, demonstrating that these variants have improved binding affinities.



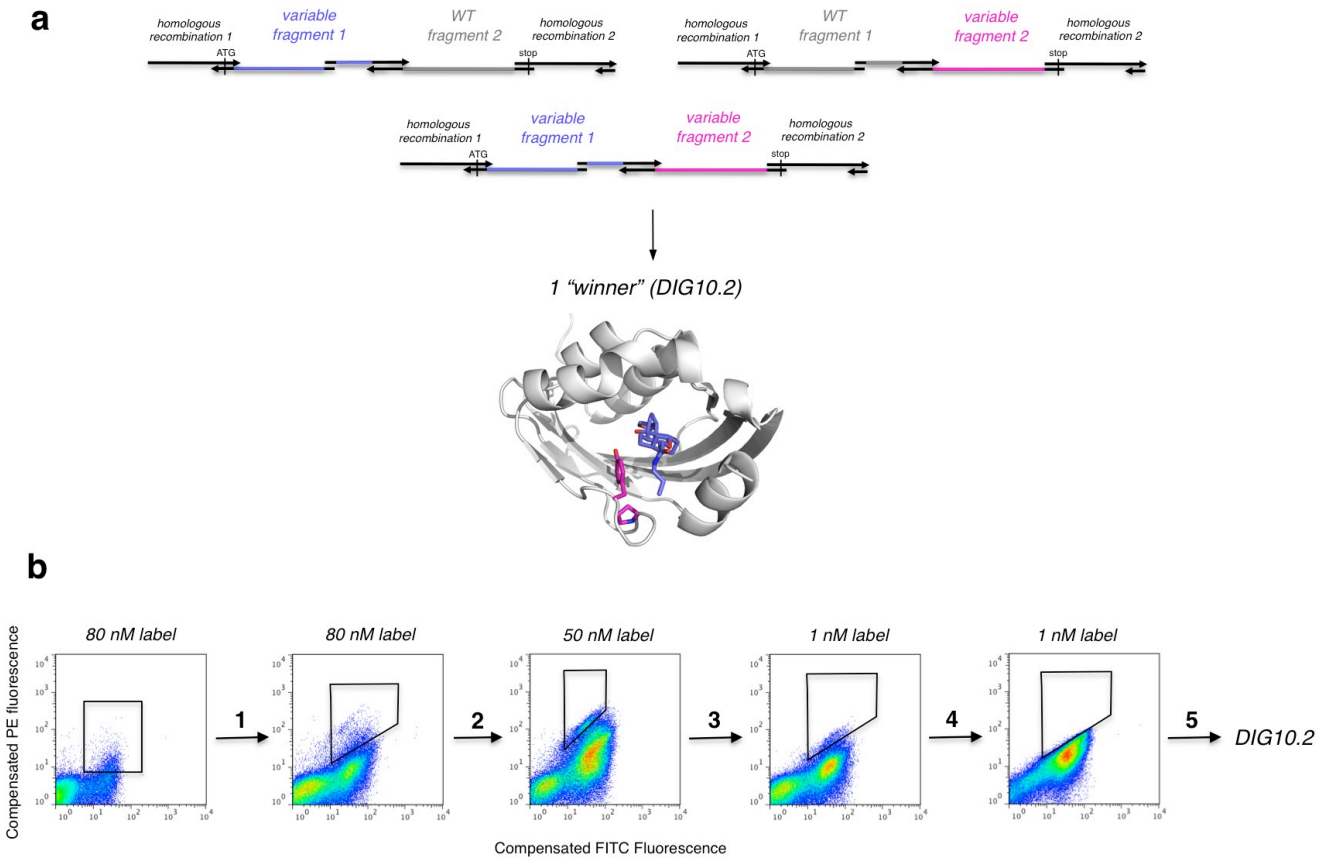
Supplementary Figure 8 | Fluorescence polarization affinity measurements of DIG10 evolved variants. a, Equilibrium fluorescence polarization measurements of DIG-PEG_3-Alexa488 treated with increasing amounts of DIG10.1 (left panel), DIG10.2 (middle panel), and DIG10.3 (right panel). Solid lines represent fits to the data to obtain dissociation constants (K_d values). Different colored symbols represent different independent measurements collected with at least two separate batches of purified protein. For DIG10.1, [DIG-PEG_3-Alexa488] = 10 nM. For DIG10.2, [DIG-PEG_3-Alexa488] = 1 nM. For DIG10.3, [DIG-PEG_3-Alexa488] = 0.5 nM. **b,** Equilibrium fluorescence polarization measurements of DIG-PEG_3-Alexa488 treated with increasing amounts of DIG10.3 variants Tyr34Phe (panel 1), Tyr101Phe (panel 2), Tyr115Phe (panel 3), Tyr99Phe/Tyr101Phe (panel 4), Tyr34Phe/Tyr99Phe/Tyr101Phe (panel 5), and Tyr34Phe/Tyr99Phe/Tyr101Phe/Tyr115Phe (panel 6). Solid lines represent fits to the data to obtain dissociation constants (K_d values). Error bars represent standard deviations for at least three independent measurements collected using at least two different batches of purified protein. For Tyr34Phe, Tyr101Phe, Tyr115Phe, and Tyr99Phe/Tyr101Phe, [DIG-PEG_3-Alexa488] = 2 nM. For Tyr34Phe/Tyr99Phe/Tyr101Phe and Tyr34Phe/Tyr99Phe/Tyr101Phe/Tyr115Phe, [DIG-PEG_3-Alexa488] = 10 nM.



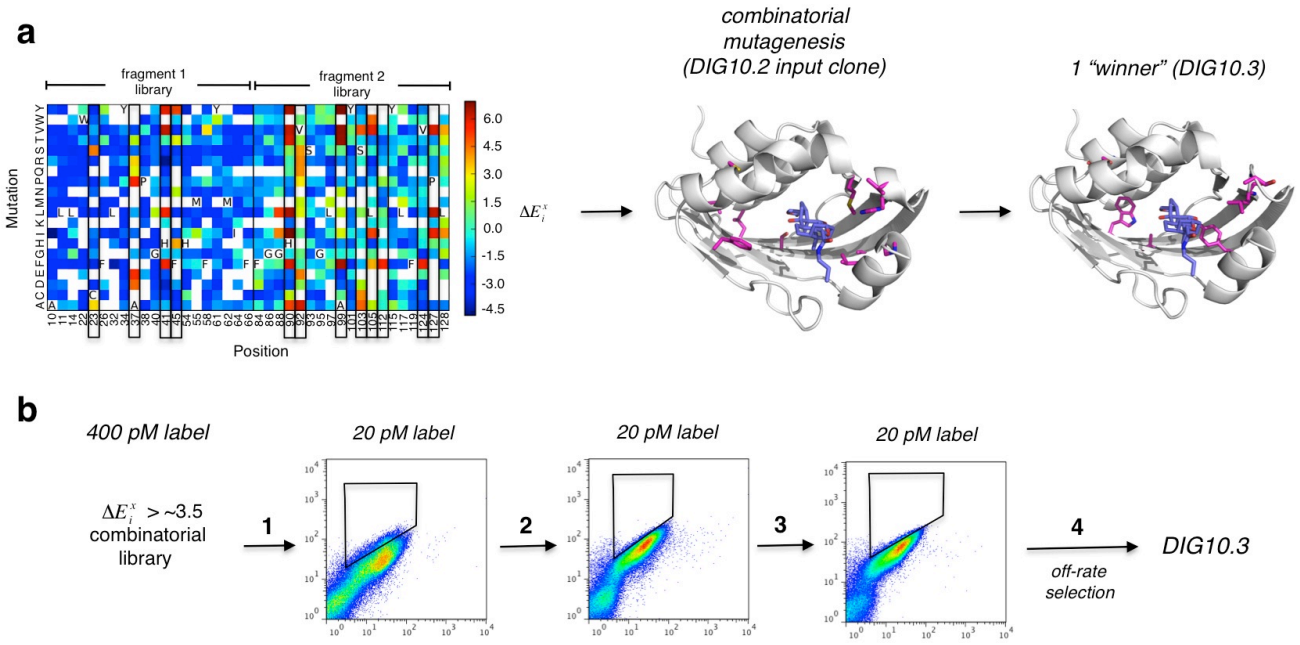
Supplementary Figure 9 | DIG10.1 deep sequencing library construction and selections. a, DIG10.1-based deep sequencing fragment libraries. Residues included in the fragment 1 (left) and fragment 2 (right) libraries are depicted in blue and magenta, respectively (upper panels). Libraries were constructed by recursive PCR using a combination of mutagenized (colored) and wild-type (gray) oligos (lower panels). **b,** Flow cytometry plots of yeast surface display selections for deep sequencing experiments. Fragment libraries 1 and 2 were first labeled with anti-cmyc-FITC and the expressing populations were collected using fluorescent gates (black squares, left panel step 1). Expressing cells were recovered, labeled with 100 nM monovalent DIG-PEG_3-biotin and then SAPE, and then library clones (navy) having higher PE binding signals than DIG10.1 (aqua) were collected using fluorescent gates (black quadrilaterals, center panel, step 2). To reduce noise, this procedure was repeated once more using the same conditions. Following these two rounds of binding selections, the selected cells showed higher binding signals than DIG10.1 (right panel). DNA from the expression-sorted naïve libraries and the selected libraries were subjected to deep sequencing.



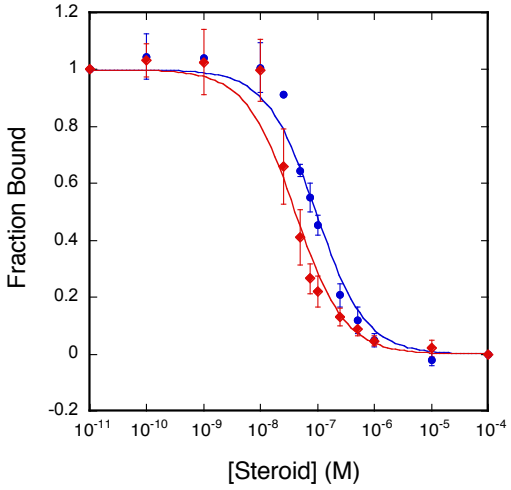
Supplementary Figure 10 | DIG10.1 deep sequencing library statistics. **a,** A data matrix showing the number of counts for each single mutation in the unselected N-terminal (fragment 1) and C-terminal (fragment 2) deep sequencing libraries. Blue and green indicate a low and a high number of counts, respectively. The DIG10.1 amino acid at each position is indicated in bold using its one-letter amino acid code. **b,** A histogram of the deep sequencing data in Fig. 2e indicates that most mutations are deleterious for binding. **c,** Determination of the significance of deep sequencing data for mutations with low numbers of counts in the unselected library. At low initial read counts, there is a negative correlation between the number of initial reads and enrichment value. The Pearson coefficient (blue) and the p-value (red) describing the correlation between the number of counts in the unselected library and the enrichment value were calculated for windows of unselected library counts (window size 10) and are plotted versus the count value at the bottom of each window. The count cutoff in which the correlation is no longer significant ($p > 0.05$, horizontal line) was found to be 7 (vertical line).



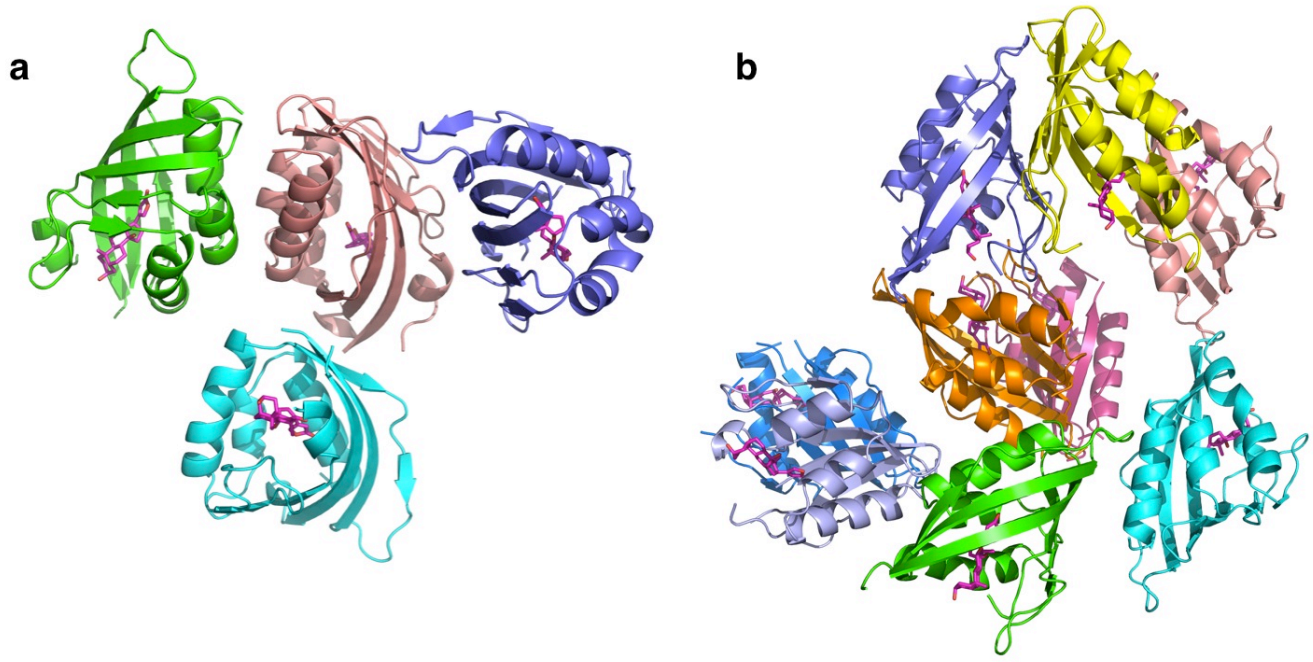
Supplementary Figure 11 | Affinity maturation of DIG10.1: Round 2. a, The round 2 DIG10.1 affinity maturation library was constructed by pooling the products of three recursive PCR reactions using different combinations of mutagenized (colored) and wild-type (gray) oligos (upper panel). After several rounds of selections of the library with monovalent DIG-BSA-biotin and then SAPE, a single best variant, DIG10.2, having two mutations from DIG10.1, was identified. **b,** Flow cytometry plots of yeast surface display selections for affinity maturation round 2. Yeast cells were subjected to five increasingly stringent rounds of selections with DIG-PEG_3-biotin and then SAPE using fluorescent gates (black quadrilaterals).



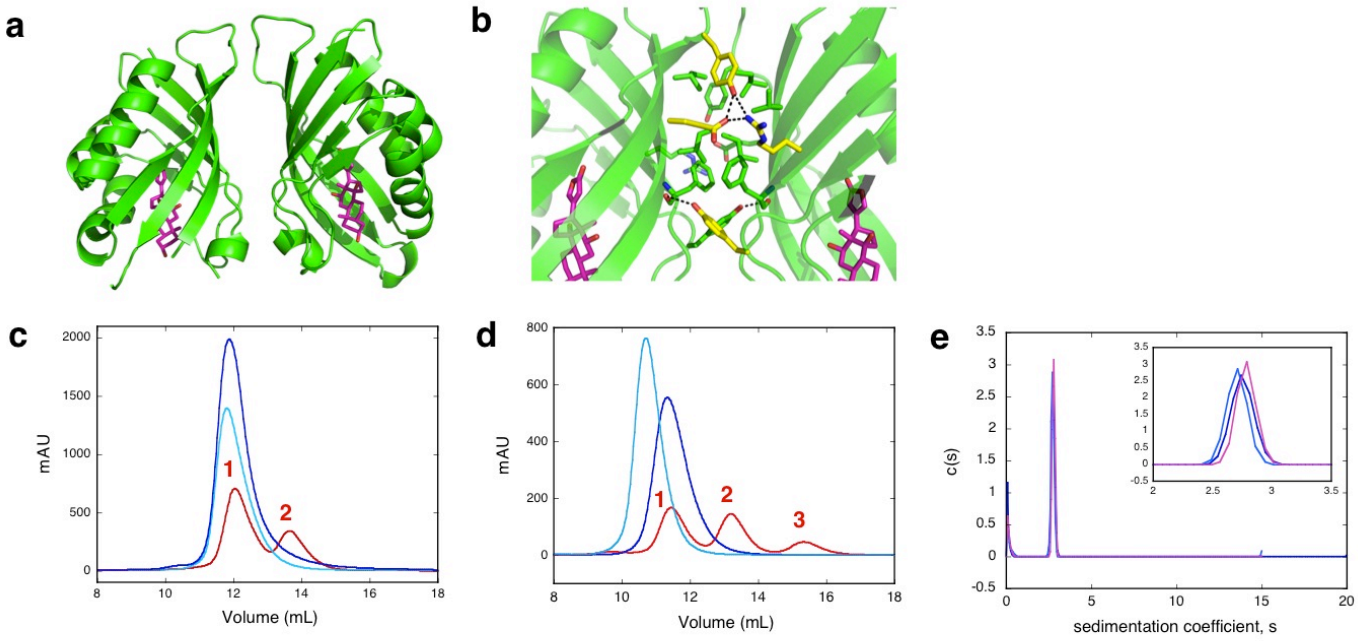
Supplementary Figure 12 | Affinity maturation of DIG10.2: Round 3. a, Strategy for round 3 affinity maturation of DIG10.2. Mutations having ΔE_i^x >~3.5 in the deep sequencing experiment (left panel) and the DIG10.2 amino acids at these positions (magenta sticks, middle panel) were combined combinatorially by Kunkel mutagenesis with degenerate primers. Selections converged to a single variant, DIG10.3, having six mutations from DIG10.2 (right panel). **b,** Flow cytometry plots of yeast surface display selections for affinity maturation round 3. The library was subjected to four increasingly stringent rounds of selections with DIG-PEG_3-biotin and then SAPE using fluorescent gates (black quadrilaterals). An off-rate selection was used in the last round.



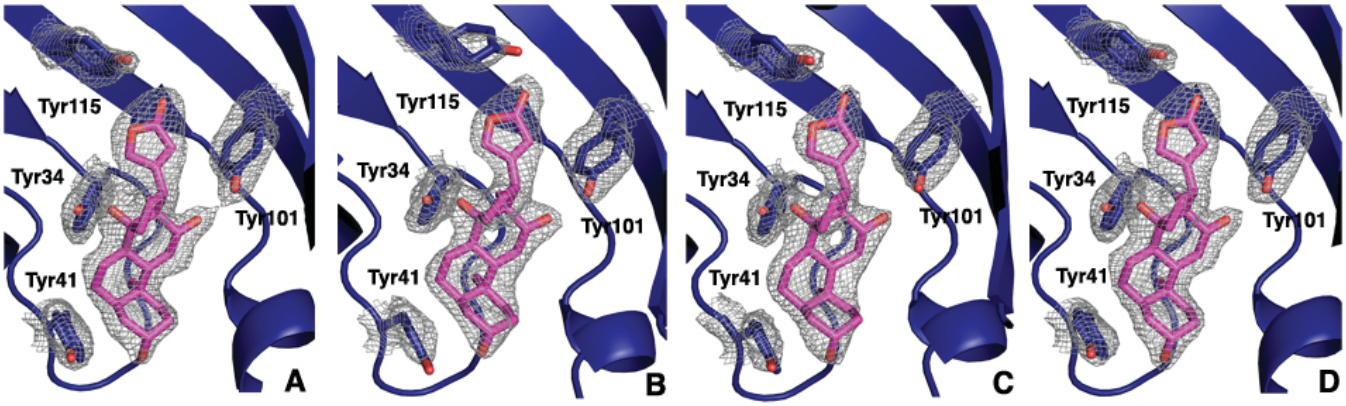
Supplementary Figure 13 | Equilibrium competition fluorescence polarization assays of DIG10.3 with digoxin. Unlabeled digoxin (red) or digoxigenin (blue) was added to a solution of DIG10.3 and DIG-PEG_3-Alexa488 in increasing amounts. Solid lines represent fits to the data to obtain the half-maximal inhibitory concentrations (IC_{50} values). Error bars represent standard deviations for at least three independent measurements collected using at least two different batches of purified protein. The affinity of DIG10.3 for digoxin is slightly higher than that for DIG (see Supplementary Table 4).



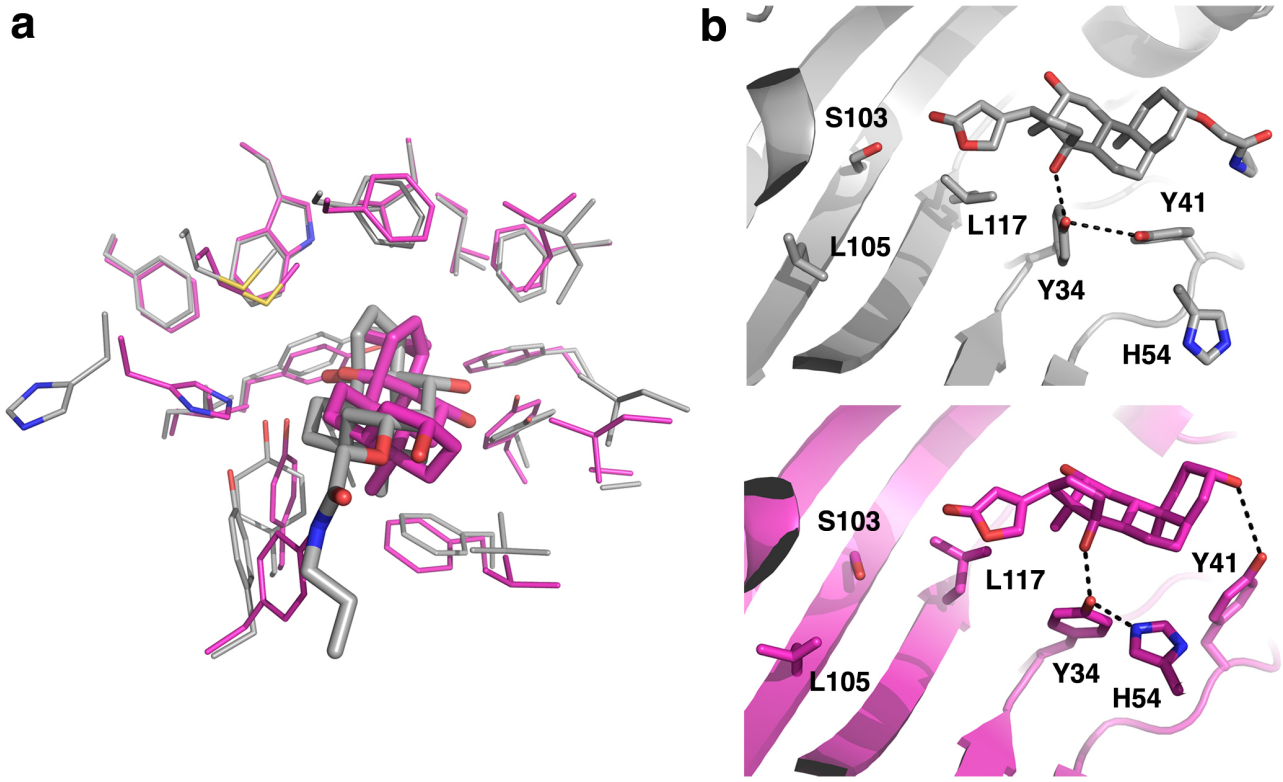
Supplementary Figure 14 | Crystal packing in the DIG10.2–DIG and DIG10.3–DIG Complexes. a, The asymmetric unit of the DIG10.2 crystal structure contains four copies, each of which bind a molecule of DIG. **b,** The asymmetric unit of the DIG10.3 crystal structure contains nine copies, each of which bind a molecule of DIG.



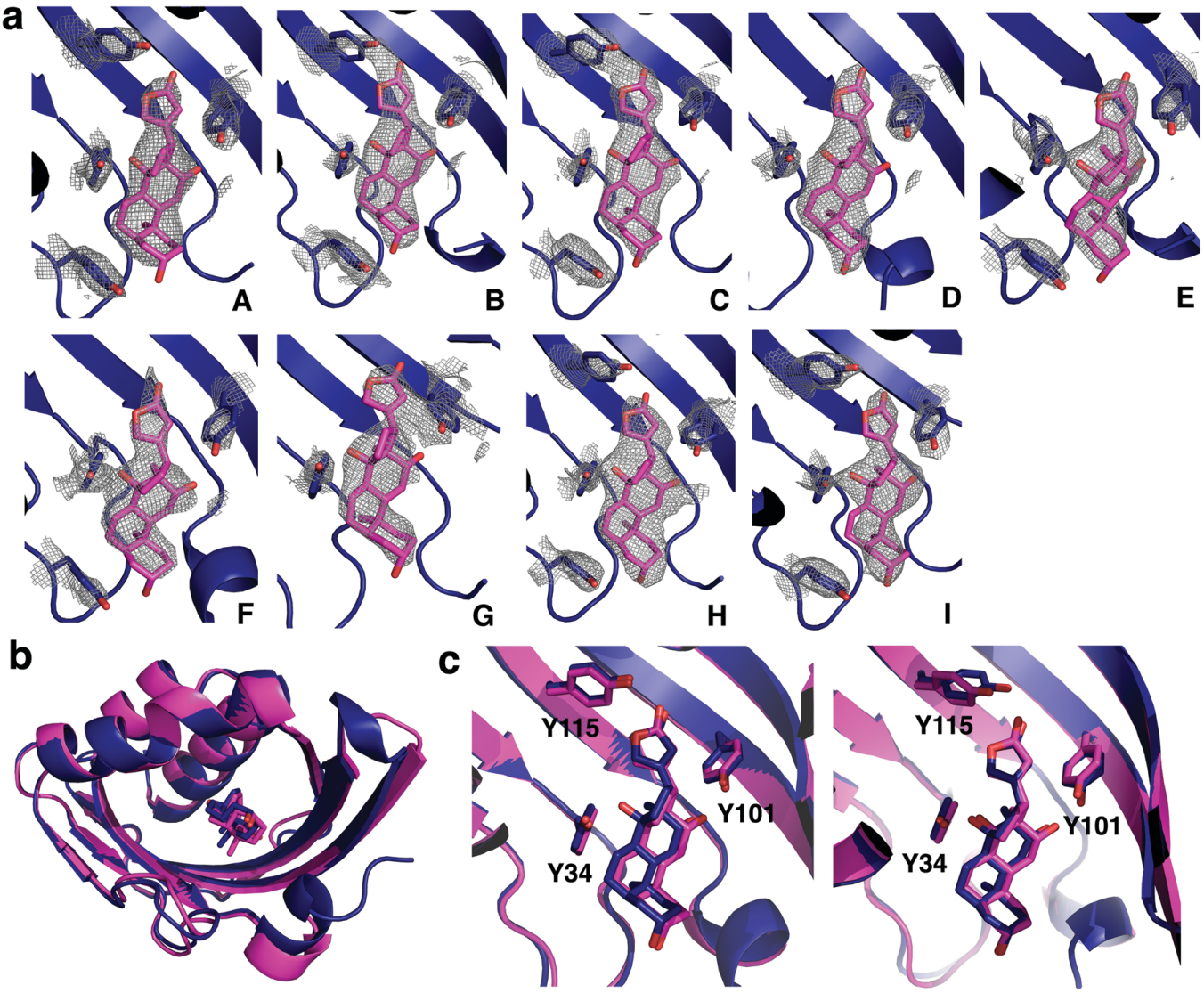
Supplementary Figure 15 | DIG10-based designs are dimers. a, The dimeric unit of DIG10.2 observed in the crystal structure. The protomers are related by a pseudosymmetric or symmetric C2 axis. DIG is shown in magenta sticks. **b,** The dimer interface is formed by specific intermolecular salt-bridges, packing interactions, and hydrogen bonds between the solvent-facing sides of the curved β-sheets. **c,** Preparative Superdex 75 gel filtration traces of DIG10 (cyan), DIG10.1b (navy), molecular weight standard horse heart cytochrome c (M_r = 29 kDa, red **1**), and molecular weight standard bovine erythrocytes carbonic anhydrase (M_r = 12.4 kDa, red **2**). Both DIG10 and DIG10.1b elute near their expected dimeric molecular weights (36 kDa). Both proteins are well-behaved in solution, and the traces show no evidence for sample heterogeneity or higher-order aggregate species. **d,** Analytical Superdex 75 gel filtration traces of DIG10.3 (navy), pre-formed DIG10.3–DIG complex (cyan), molecular weight standard horse heart cytochrome c (M_r = 29 kDa, red **1**), molecular weight standard bovine erythrocytes carbonic anhydrase (M_r = 12.4 kDa, red **2**) and bovine aprotinin (6.5 kDa, red **3**). DIG10.3 elutes near its expected dimeric molecular weight (36 kDa). The DIG10.3–DIG complex elutes at a slightly shorter retention volume. DIG10.3 and the DIG10.3–DIG complex are both well behaved in solution, and the traces show no evidence for sample heterogeneity or higher-order aggregate species. **e,** Analytical ultracentrifugation c(s) size distributions of DIG10.3-TEV-his_6 in PBS (blue), DIG10.3-TEV-his_6 in PBS, 0.66% MeOH (cyan), and the DIG10.3-TEV-his6–DIG complex in PBS, 0.66% MeOH (pink), at 0.5 mg/mL and 20 °C. The distributions show a single peak for each sample that corresponds to protein dimer. Similar data were observed for the 0.25 mg/mL samples.



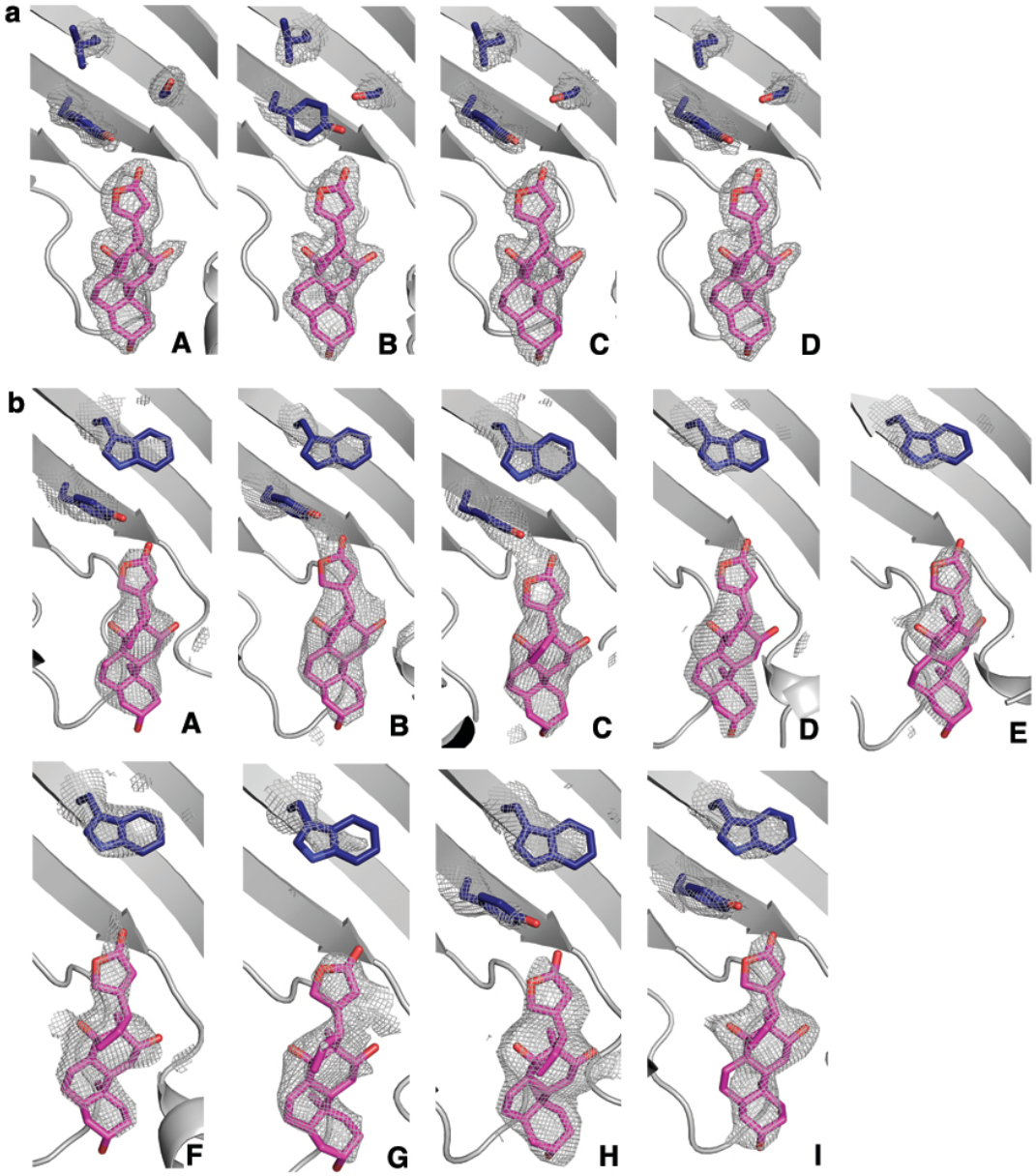
Supplementary Figure 16 | Ligand binding site $2F_o$-F_c maps of the DIG10.2–DIG complex. $2F_o - F_c$ omit map electron density of DIG interacting with Tyr34, Tyr41, Tyr101, and Tyr115 in chains A, B, C, and D contoured at 1.0 sigma.



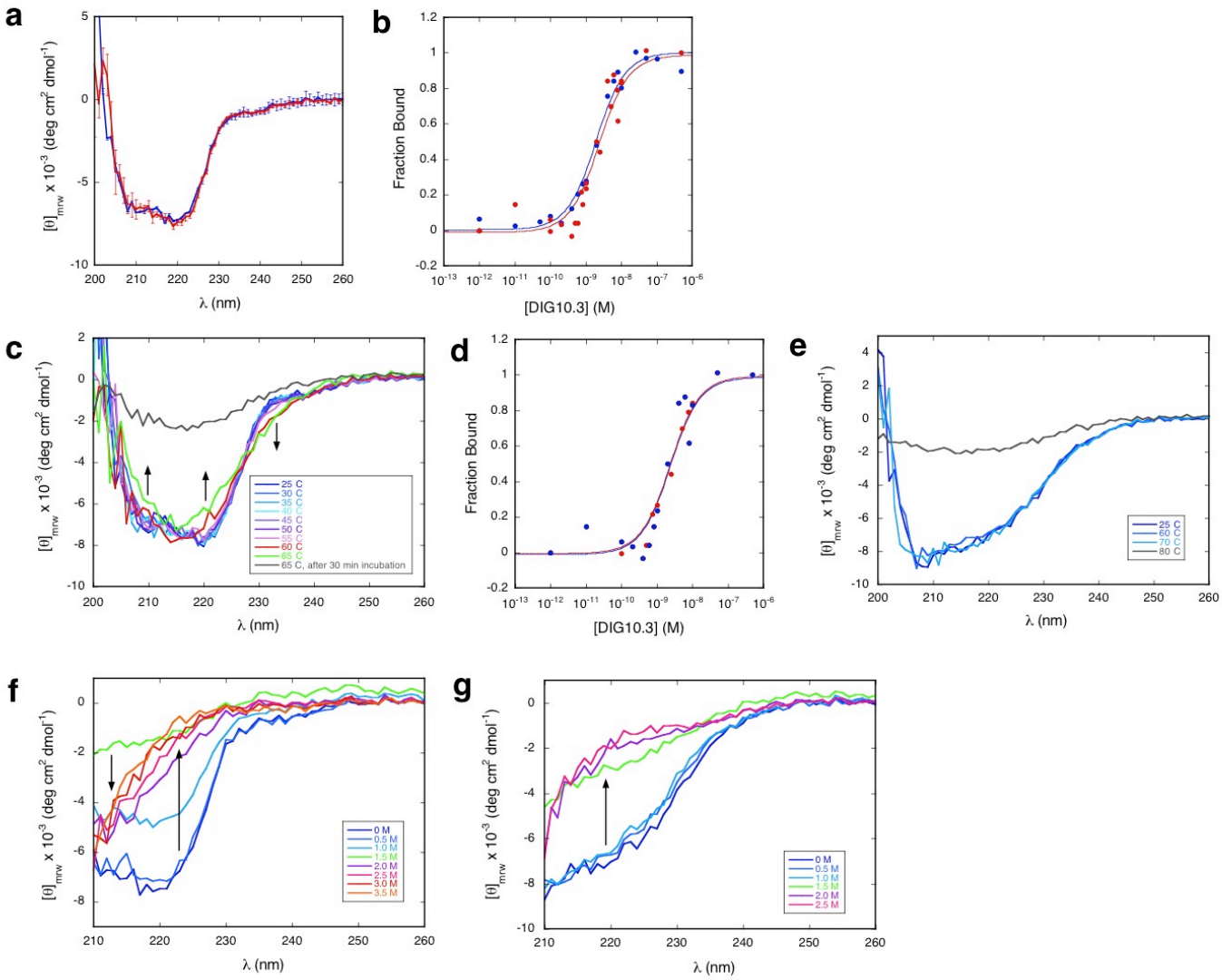
Supplementary Figure 17 | Comparison of the side chain rotamers in the DIG10.2–DIG crystal structure with the computational model. a, A backbone superposition of the computational model (gray) and the x-ray crystal structure (magenta) of DIG10.2 shows that the majority of the amino acid side chains in the binding cavity adopt their modeled conformations. **b,** A side-by-side comparison of the computational model (gray, top panel) and the x-ray crystal structure (magenta, bottom panel) of DIG10.2 highlighting the conformations of the six incorrectly modeled amino acids. Tyr34 adopts a statistically less probable rotamer (chi2 = 153°) in the crystal structure than the computational model (chi2 = 80°), which may result from an unanticipated hydrogen bond with His54: a subtle shift in the backbone position of this histidine allows it to face inwards towards the binding cavity and interact with Tyr34 instead of being fully solvent-exposed as predicted by the model. Perhaps to relieve hydrophobic clashes with the crystallographic Tyr34 rotamer, Leu117 also has a different side conformation. Finally, Try41, which engages in a second shell hydrogen-bonding interaction with Tyr34 in the computational model, adopts a different chi1 rotamer, and instead participates in a long (3.7 – 4.0 Å) hydrogen bond with the A-ring hydroxyl group of DIG. Ser103 and Leu105 also show different conformations in the structure and the computational model, although Ser103 is characterized by high conformational heterogeneity between the four protomers in the crystallographic asymmetric unit (see Fig. 3e).



Supplementary Figure 18 | Crystal structure of the DIG10.3–DIG complex. a, $2F_o - F_c$ omit map electron density of DIG and hydrogen-bonding residues Tyr34, Tyr41, Tyr101, and Tyr115 in chains A though I of the DIG10.3–DIG crystal structure contoured at 1.0 sigma. At this contour level and poor resolution (3.2 Å), density is not observed for all hydrogen bonding residues in all crystallographic copies of the protein. **b,** Backbone binding site superposition of the crystal structures of DIG10.3 (magenta) and DIG10.2 (navy). **c,** Backbone binding site superposition of the crystal structures of DIG10.3 (magenta) and DIG10.2 (navy) chains A (left) and B (right). The DIG10.3 Tyr115 rotamer is similar to that observed in DIG10.2 chains A, C, and D but different from that observed in chain B. In DIG10.3 and chains A, C, and D of DIG10.2, the hydroxyl group of Tyr115 is plane with the lactone ring (~5° torsion), but in DIG10.2 chain B, it is out of plane (-70° torsion) and therefore expected to make a weaker interaction.

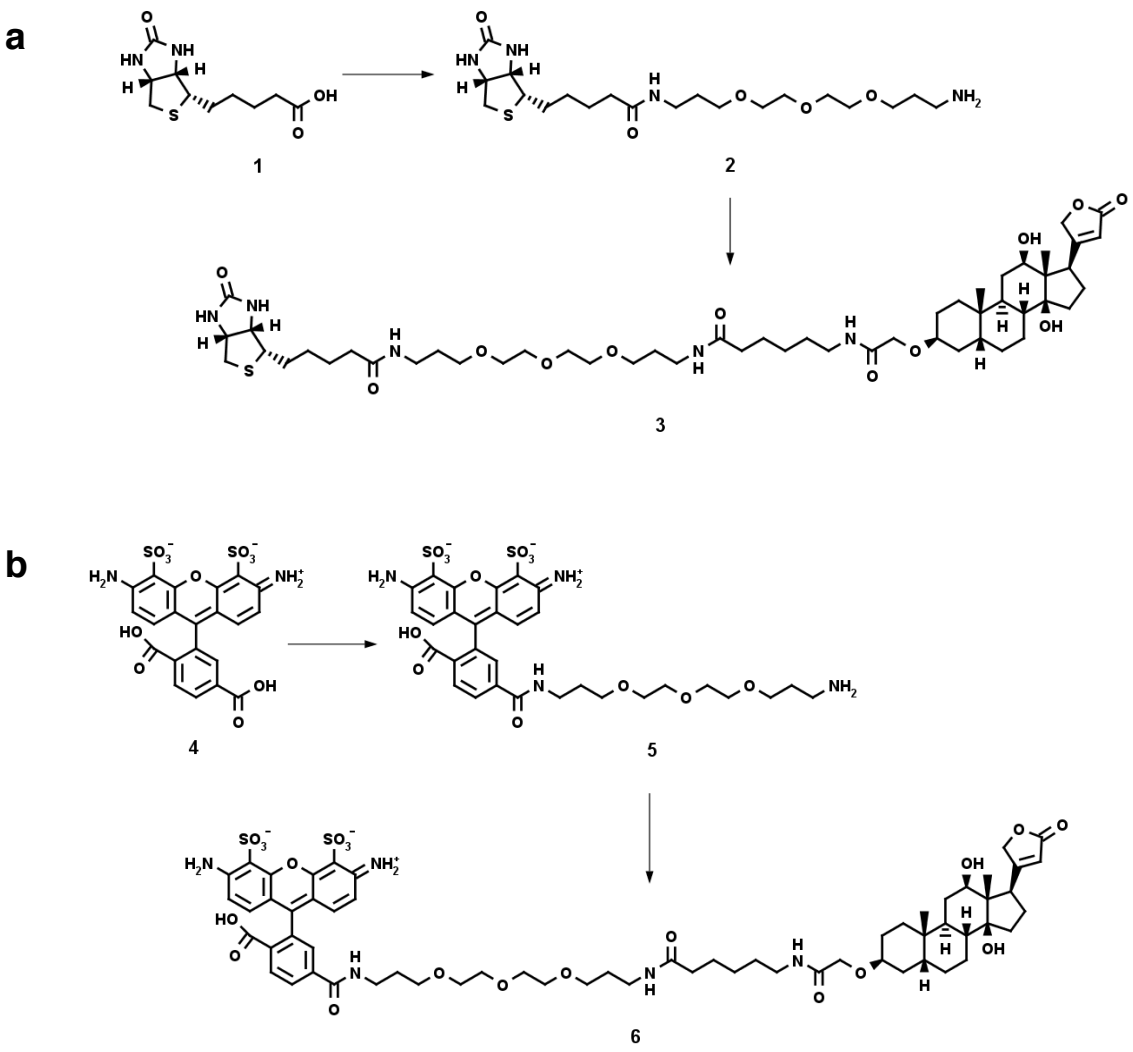


Supplementary Figure 19 | Side chain conformational heterogeneity in the crystal structures of DIG10.2–DIG and DIG10.3–DIG. a, $2F_o - F_c$ omit map electron density of DIG, Tyr115 and S103 in chains A through D of the DIG10.2–DIG crystal structure contoured at 1.0 sigma. Tyr115 and S103 explore more than one rotameric conformation. L105 is also depicted **b,** $2F_o - F_c$ omit map electron density of DIG, Tyr115, and Trp105 in chains A though I of the DIG10.3–DIG crystal structure contoured at 1.0 sigma. At this contour level and poor resolution (3.2 Å), density for Try115 is only observed in five of the nine copies; however, for copies in which density is observed, this amino acid is clearly in the same conformation. The position of Trp105, which is the same in all nine crystallographic copies, is inconsistent with the alternative conformation of Tyr115 observed in chain B of DIG10.2–DIG, which has a less canonical hydrogen-bonding geometry than the Tyr115 conformations observed in chains A, C, and D of DIG10.2-DIG and in DIG10.3-DIG.



Supplementary Figure 20 | Stability of DIG10.3. a, Temporal stability of DIG10.3. CD spectra of DIG10.3-TEV-his$_6$ samples that were either incubated at room temperature for 3 months (blue) or freshly purified (red) at 25 °C. Proteins were prepared at 25 μM in PBS, pH 7.4. The wavelength scans of the two samples are identical, indicating that DIG10.3 is folded even after long incubations at room temperature. **b,** Equilibrium fluorescence polarization binding measurements of DIG10.3-TEV-his$_6$ samples that were either incubated at room temperature for 3 months (blue) or freshly purified (red). After this extended period at room temperature, DIG10.3 retains binding activity. **c,** Temperature stability of DIG10.3. The CD spectrum of 25 μM DIG10.3-TEV-his$_6$ in PBS, pH 7.4, was measured at different temperatures in increments of 5 °C. The sample was allowed to equilibrate at each temperature for 10 min prior to recording each spectrum. Below 60 °C, the protein exhibits a signal indicative of stable mixed α/β character. At 60 °C, the spectrum shows subtle differences that suggest formation of an alternative protein state (arrows). Upon incubation of the sample at 65 °C for 30 min longer, the protein

irreversibly precipitates, suggesting that this alternative state could be a kinetic precursor to protein precipitation. **d,** Equilibrium fluorescence polarization binding measurements of DIG10.3-TEV-his_6 samples that were either incubated at 50 °C (blue) or room temperature (red) for 30 min and then equilibrated at room temperature for 60 min prior to data collection. After a 30-min incubation at 50 °C, DIG10.3 retains binding activity. **e,** Temperature stability of 1z1s. The CD spectrum of 25 μM 1z1s-TEV-his_6 in PBS, pH 7.4, was measured at different temperatures. The sample was allowed to equilibrate at each temperature for 10 min prior to recording the spectrum. Above 70 °C, the protein irreversibly precipitates. **f,** Stability of DIG10.3 in the presence of guanadium hydrochloride (GuHCl). Samples of DIG10.3-TEV-his_6 were equilibrated with GuHCl in PBS, pH 7.4 at room temperature for at least 60 min prior to data collection. DIG10.3 exhibits multi-state behavior indicated by arrows – the protein precipitates at 1.5 M GuHCl but exists in solution as a random coil at concentrations above 1.5 M GuHCl. **g,** Stability of 1z1s in the presence of guanadium hydrochloride (GuHCl). Samples of 1z1s-TEV-his_6 were equilibrated with GuHCl in PBS, pH 7.4 at room temperature for at least 60 min prior to data collection. 1z1s forms a soluble random coil at concentrations at or above 1.5 M GuHCl and does not exhibit the multi-state behavior observed for DIG10.3 on the same experimental timescale.



Supplementary Figure 21 | Synthetic schemes. a, Synthetic scheme for the preparation of DIG-PEG$_3$-biotin. **b,** Synthetic scheme for the preparation of DIG-PEG$_3$-Alexa488.

Supplementary Data

Example Command Line for Generating Ligand Conformers:

```
~/rosetta/bin/generate_ligens.linuxiccrelease -database <rosetta_database_path>
-in:file::s <ligand_pdb_model.pdb> -in:file::exra_res_fa <ligand.params> @flags

@flags:
-packing -use_input_sc -enzdes -rot_ensemble_ecutoff 0.25 -cst_design -
no_unconstrained_repack
```

Example Command Line for Matching:

```
~/rosetta/bin/match.static.linuxiccrelease -database <rosetta_database_path>
-extra_res_fa <ligand.params> -s <scaffold.pdb> -match:scaffold_active_site_residues
<scaffold.pos> -match:geometric_constraint_file <constraint.cst> @match.flags


@match.flags:
-match:lig_name:DIG -match:filter_colliding_upstream_residues
-match:filter_upstream_downstream_collisions -match:updown_collision_tolerance 0.3
-match::bump_tolerance 0.3 -match_grouper SameSequenceAndDSPositionGrouper
-match:euclid_bin_size 0.9 -match:euler_bin_size 9.0 -packing -extrachi_cutoff 0
-use_input_sc -in:ignore_unrecognized_res -output_format CloudPDB
-enumerate_ligand_rotamers -only_enumerate_non_match_redudant_ligand_rotamers
-out::file::output_virtual
```

Example Command Line for Design:

```
~/rosetta/bin/rosetta_scripts.static.linuxiccrelease -nstruct 1 -jd2:ntrials 1
-parser:protocol <RosettaScripts_protocol.xml> -database <rosetta_database_path>
-out::overwrite -s <input.pdb> @ligdes.flags

@ligdes.flags:
-run::preserve_header -enzdes::minimize_ligand_torsions 5.0
-enzdes::detect_design_interface -enzdes::cut1 6.0 -enzdes::cut2 8.0
-enzdes::cut3 10.0 -enzdes::cut4 12.0 -enzdes::bb_min_allowed_dev 0.05
-score:weights ~/rosetta_database/scoring/weights/enzdes.wts -packing::use_input_sc
-packing::extrachi_cutoff 1 -packing::ex1 -packing::ex2 -linmem_ig 10
-no_optH false -in:file::pssm scaffold.fasta.pssm
-extra_res_fa <DIG.params>
```

RosettaScripts: ligdes.XML:

```
<dock_design>
      <SCOREFXNS>
            <myscore weights=enzdes.wts/>
      </SCOREFXNS>

      <TASKOPERATIONS>
            <DetectProteinLigandInterface name=edto_aro design=1 cut1=6.0 cut2=8.0
             cut3=10.0 cut4=12.0 preserve_aromatics=1/>
            <DetectProteinLigandInterface name=edto_repack design=0 cut1=6.0
             cut2=8.0 cut3=10.0 cut4=12.0/>
            <LimitAromaChi2 name=limchi2/>
            <SetCatalyticResPackBehavior name=catres fix_catalytic_aa=0/>
            <SetCatalyticResPackBehavior name=fixcat fix_catalytic_aa=1/>
      </TASKOPERATIONS>

      <FILTERS>
            <EnzScore name="allcst" score_type=cstE scorefxn=myscore whole_pose=1
             energy_cutoff=10/>
            <LigInterfaceEnergy name="interfE" scorefxn=myscore energy_cutoff=-7.5/>
            <DiffAtomBurial name="pointingO1" res1_res_num=0 atomname1="C7"
             res2_res_num=0 atomname2="O5" sample_type="less"/>
            <CompoundStatement name="myfilter">
                  <AND filter_name="allcst"/>
                  <AND filter_name="interfE"/>
                  <AND filter_name="pointingO1"/>
            </CompoundStatement>
      </FILTERS>

      <MOVERS>
            <AddOrRemoveMatchCsts name=cstadd cst_instruction=add_new/>
            <EnzRepackMinimize name=cstopt cst_opt=1 minimize_rb=1 minimize_sc=1
             minimize_bb=0 cycles=1 min_in_stages=0 minimize_lig=1/>
            <GenericMonteCarlo name=multicstopt mover_name=cstopt filter_name=allcst
             trials=10 sample_type=low temperature=0.6 drift=0/>
            <EnzRepackMinimize name=desmin_fixcat design=1 repack_only=0
             scorefxn_minimize=myscore scorefxn_repack=soft_rep minimize_rb=1
             minimize_sc=1 minimize_bb=0 cycles=2 minimize_lig=1 min_in_stages=0
             backrub=0 task_operations=edto_aro,limchi2,fixcat/>
            <EnzRepackMinimize name=desmin design=1 repack_only=0
             scorefxn_minimize=myscore scorefxn_repack=soft_rep minimize_rb=1
             minimize_sc=1 minimize_bb=0 cycles=2 minimize_lig=1 min_in_stages=0
             backrub=0 task_operations=edto_aro,limchi2,catres/>
            <EnzRepackMinimize name=fin_min repack_only=0 design=0
             scorefxn_minimize=myscore scorefxn_repack=myscore minimize_rb=1
             minimize_sc=1 minimize_bb=1 cycles=1
             task_operations=edto_aro,limchi2,catres/>
            <EnzRepackMinimize name=fin_rpkmin repack_only=1 design=0
             scorefxn_minimize=myscore scorefxn_repack=myscore minimize_rb=1
             minimize_sc=1 minimize_bb=0 cycles=1
             task_operations=edto_repack,limchi2,catres/>
            <AddOrRemoveMatchCsts name=cstrem cst_instruction=remove/>
            <AddOrRemoveMatchCsts name=fincstadd cst_instruction=add_pregenerated/>
            <FavorNativeResidue name=fnr bonus=1.25/>
```

```
      </MOVERS>

      <PROTOCOLS>
            <Add mover_name=fnr/>
            <Add mover_name=cstadd/>
            <Add mover_name=multicstopt filter_name=allcst/>
            <Add filter_name=pointingO1/>
            <Add mover_name=desmin_fixcat/>
            <Add mover_name=desmin/>
            <Add mover_name=cstrem/>
            <Add mover_name=fin_min/>
            <Add mover_name=fin_rpkmin filter_name=interfE/>
            <Add mover_name=fincstadd filter_name=myfilter/>
      </PROTOCOLS>
</dock_design>
```

RosettaScripts: ligdes_fix_cst.xml:

```
<dock_design>
     <SCOREFXNS>
            <myscore weights=enzdes.wts/>
     </SCOREFXNS>

     <TASKOPERATIONS>
            <DetectProteinLigandInterface name=edto design=1 cut1=6.0 cut2=8.0
             cut3=10.0 cut4=12.0/>
            <ProteinLigandInterfaceUpweighter name=up interface_weight=1.5/>
            <LimitAromaChi2 name=limchi2/>
            <InitializeFromCommandline name=init/>
            <SetCatalyticResPackBehavior name=catres fix_catalytic_aa=0/>
            <RestrictConservedLowDdg name=ddg ddG_filename="ddg_predictions.out"
             conservation_cutoff=0.6 ddG_cutoff=1.5 verbose=1 force_similar=1/>
     </TASKOPERATIONS>

     <FILTERS>
            <LigInterfaceEnergy name="interfE" scorefxn=myscore energy_cutoff=-6.5/>
            <DSasa name="lsasa" lower_threshold=0.3 upper_threshold=0.95/>
            <ConservedPosMutationFilter name="conserve_mut"
             ddG_filename="ddg_predictions.out" conservation_cutoff=0.6
             ddG_cutoff=1.5 max_conserved_pos_mutations=5/>
     </FILTERS>

    <MOVERS>
            <AddOrRemoveMatchCsts name=cstadd cst_instruction=add_new
             cstfile=DIG_yyhff.cst/>
            <EnzRepackMinimize name=desmin design=1 repack_only=0
             scorefxn_minimize=myscore scorefxn_repack=soft_rep minimize_rb=1
             minimize_sc=1 minimize_bb=0 cycles=3 minimize_lig=1 min_in_stages=0
             backrub=0 task_operations=edto,limchi2,ddg,up,init,catres/>
            <FavorNativeResidue name=fnr bonus=1.5/>
     </MOVERS>

     <PROTOCOLS>
            <Add mover_name=fnr/>
            <Add mover_name=cstadd/>
            <Add mover_name=desmin/>
            <Add filter_name=interfE/>
            <Add filter_name=conserve_mut/>
            <Add filter_name=lsasa/>
     </PROTOCOLS>
</dock_design>
```

RosettaScripts: ligdes_flex_hb.xml:

```
<dock_design>
     <SCOREFXNS>
            <myscore weights=enzdes.wts/>
     </SCOREFXNS>

     <TASKOPERATIONS>
            <DetectProteinLigandInterface name=edto design=1 cut1=6.0 cut2=8.0
             cut3=10.0 cut4=12.0/>
            <ProteinLigandInterfaceUpweighter name=up interface_weight=1.5/>
            <LimitAromaChi2 name=limchi2/>
            <InitializeFromCommandline name=init/>
            <RestrictConservedLowDdg name=ddg ddG_filename="ddg_predictions.out"
             conservation_cutoff=0.6 ddG_cutoff=1.5 verbose=1 force_similar=1/>
     </TASKOPERATIONS>

     <FILTERS>
            <LigInterfaceEnergy name="interfE" scorefxn=myscore energy_cutoff=-7.5/>
            <DSasa name="lsasa" lower_threshold=0.3 upper_threshold=0.95/>
            <HbondsToResidue name=hbond partners=3 energy_cutoff=-0.4 pdb_num=1X/>
    </FILTERS>

    <MOVERS>
            <EnzRepackMinimize name=desmin design=1 repack_only=0
             scorefxn_minimize=myscore scorefxn_repack=soft_rep minimize_rb=1
             minimize_sc=1 minimize_bb=0 cycles=3 minimize_lig=1 min_in_stages=0
             backrub=0 task_operations=edto,limchi2,ddg,up,init/>
            <FavorNativeResidue name=fnr bonus=1.5/>
    </MOVERS>

    <PROTOCOLS>
            <Add mover_name=fnr/>
            <Add mover_name=desmin filter_name=interfE/>
            <Add filter_name=lsasa/>
            <Add filter_name=hbond/>
     </PROTOCOLS>
</dock_design>
```

Procedure for Estimating Sequence-Structure Compatibility

```
/path/to/script/getFastaFromCoords.pl -p <design.pdb> -chain A > design.fasta
sed -i -e '/DIG/d' $2.pdb
/path/to/script/make_fragments.9and3.pl -verbose -nocleanup design.fasta -id
<scaffold_pdb_code> -xx aa -nohoms
ls aa* > list.txt
/path/to/rosetta/rosetta_source/bin/r_frag_quality2.static.linuxgccrelease -database
/path/to/rosetta_database/ -in:file:native <scaffold.pdb> -s <design.pdb> -frags
list.txt
/path/to/script/getfragrmsave.pl aa$109_05.200_v1_3.rms > frags09.stats
```

Example Command Line and RosettaScripts Protocol for Modeling Directed Evolution Variants Using make_mutation.xml:

```
~/rosetta/bin/rosetta_scripts.linuxiccrelease -database <rosetta_database_path>
-nstruct 10 -jd2:ntrials 1 -s <input.pdb> -parser:protocol <make_mutation.xml>
-extra_res_fa <ligand.params> @make_mutation.flags

@make_mutation.flags:
-run::preserve_header -enzdes::minimize_ligand_torsions 5.0
-enzdes::detect_design_interface -enzdes::bb_min_allowed_dev 0.05
-score:weights ~/rosetta_database/scoring/weights/enzdes.wts
-packing::use_input_sc -packing::extrachi_cutoff 1 -packing::ex1 -packing::ex2
-linmem_ig 10 -ignore_unrecognized_res -no_optH false -correct -no_his_his_pairE
-score::hbond_params correct_params -nblist_autoupdate
-lj_hbond_hdis 1.75 -lj_hbond_OH_donor_dis 2.6
```

make_mutation.xml

```
<dock_design>
      <SCOREFXNS>
            <myscore weights=enzdes.wts/>
      </SCOREFXNS>

      <TASKOPERATIONS>
            <DetectProteinLigandInterface name=edto cut1=6.0 cut2=8.0 cut3=10.0
             cut4=12.0 design=0/>
      </TASKOPERATIONS>

      <MOVERS>
            <EnzRepackMinimize name=rpkmin repack_only=1 design=0
             scorefxn_minimize=myscore scorefxn_repack=soft_rep minimize_rb=1
             minimize_lig=1 minimize_sc=1 minimize_bb=1 cycles=1
             task_operations=edto/>
            <MutateResidue name=mutate target=1A new_res=MET/>
      </MOVERS>

      <PROTOCOLS>
              <Add mover_name=mutate/>
              <Add mover_name=rpkmin/>
      </PROTOCOLS>
</dock_design>
```

Command line and RosettaScripts XML for calculating Rotamer Boltzmann scores:

```
~/rosetta/bin/rosetta_scripts.linuxiccrelease -nstruct 1 -jd2:ntrials 1 -s $1
-parser:protocol <rb.xml> @rb.flags –database <rosetta_database_path>
-out::overwrite -in:file::extra_res_fa <ligand.params>

@rb.flags:
-run::preserve_header -enzdes::detect_design_interface -enzdes::cut1 12.0
-enzdes::cut2 12.0 -enzdes::cut3 14.0 -enzdes::cut4 14.0
-enzdes::bb_min_allowed_dev 0.05
-score:weights /rosetta_database/scoring/weights/enzdes.wts
-packing::use_input_sc -packing::extrachi_cutoff 1 -packing::ex1 -packing::ex2
-linmem_ig 10 -no_optH false –correct -no_his_his_pairE
-score::hbond_params correct_params -lj_hbond_hdis 1.75 -lj_hbond_OH_donor_dis 2.6
-linmem_ig 10 -nblist_autoupdate true -flip_HNQ
```

rb.xml

```
<dock_design>
       <TASKOPERATIONS>
            <DetectProteinLigandInterface name=enzto cut1=14 cut2=14 cut3=14 cut4=14
             design=1/>
       </TASKOPERATIONS>
       <SCOREFXNS>
            <enz weights=enzdes.wts/>
       </SCOREFXNS>
       <FILTERS>
            <Ddg name=ddg confidence=0 repeats=3/>
            <RotamerBoltzmannWeight name=boltz scorefxn=enz ddG_threshold=0
             task_operations=enzto jump=1 unbound=1 temperature=0.8
             skip_ala_scan=1/>
       </FILTERS>
       <MOVERS>
            <MinMover name=min chi=1 bb=0 jump=1/>
       </MOVERS>
       <PROTOCOLS>
              <Add mover_name=min/>
              <Add filter_name=boltz/>
       </PROTOCOLS>
</dock_design>
```

Supplementary References

1. Richter, F., Leaver-Fay, A., Khare, S.D., Bjelic, S., & Baker, D. De novo enzyme design using Rosetta3. *PLOS ONE* **6**, e19230 (2011).
2. Zanghellini, A. *et al.* New algorithms and an in silico benchmark for computational enzyme design. *Protein Sci.* **15**, 2785-2794 (2006).
3. Kuhlman, B. & Baker, D. Native protein sequences are close to optimal for their structures. *Proc. Natl. Acad. Sci. USA* **97**, 10383-10388 (2000).
4. Korndörfer, I.P., Schlehuber, S., & Skerra, A. Structural mechanism of specific ligand recognition by a lipocalin tailored for the complexation of digoxigenin. *J. Mol. Biol.* **330**, 385-396 (2003).
5. Jiang, L. *et al.* De novo computational design of retro-Aldol enzymes. *Science* **319**, 1387-1391 (2008).
6. Röthlisberger, D. *et al.* Kemp elimination catalysts by computational enzyme design. *Nature* **453**, 190-195 (2008).
7. Siegel, J.B. *et al.* Computational design of an enzyme catalyst for a stereoselective bimolecular Diels-Alder reaction. *Science* **329**, 309-313 (2010).
8. Holm, L. & Rosenström, P. Dali server: conservation mapping in 3D. *Nucleic Acids Res.* **38**, W545-W549 (2010).
9. Bloom, J.D., Labthavikul, S.T., Otey, C.R., & Arnold, F.H. Protein stability promotes evolvability. *Proc. Natl. Acad. Sci. USA* **103**, 5869-5874 (2006).
10. Tokuriki, N. & Tawfik, D.S. Chaperonin overexpression promotes genetic variation and enzyme evolution. *Nature* **459**, 668-673 (2009).
11. Jeffrey, P.D. *et al.* 26-10 Fab–digoxin complex: affinity and specificity due to surface complementarity. *Proc. Natl. Acad. Sci. USA* **90**, 10310-10314 (1993).
12. Fleishman, S.J. *et al.* RosettaScripts: A scripting language interface to the Rosetta macromolecular modeling suite. *PLOS ONE* **6**, e20161 (2011).
13. Davis, I.W. & Baker, D. RosettaLigand docking with full ligand and receptor flexibility. *J. Mol. Biol.* **385**, 381-392 (2009).
14. Kellogg, E.H., Leaver-Fay, A., & Baker, D. Role of conformational sampling in computing mutation-induced changes in protein structure and stability. *Proteins* **79**, 830-838 (2010).
15. Collaborative Computational Project, N. The *CCP4* suite: Programs for protein crystallography. *Acta Crystallogr. Sect. D* **50**, 760-763 (1994).
16. Lawrence, M.C. & Colman, P.M. Shape complementarity at protein/protein interfaces. *J. Mol. Biol.* **234**, 946-950 (1993).
17. Cooper, S. *et al.* Predicting protein structures with a multiplayer online game. *Nature* **466**, 756-760 (2010).
18. Fleishman, S.J., Khare, S.D., Koga, N., & Baker, D. Restricted sidechain plasticity in the structures of native proteins and complexes. *Protein Sci.* **20**, 753-757 (2011).
19. Raman, S. *et al.* Structure prediction for CASP8 with all-atom refinement using Rosetta. *Proteins* **77**, 89-99 (2009).
20. Nivón, L.G., Moretti, R., & Baker, D. A pareto-optimal refinement method for protein design scaffolds. *PLOS ONE* **8**, e59004 (2013).
21. Tyka, M.D. *et al.* Alternate states of proteins revealed by detailed energy landscape mapping. *J. Mol. Biol.* **405**, 607-618 (2011).
22. Kuhlman, B. *et al.* Design of a novel globular protein fold with atomic-level accuracy. *Science* **302**, 1364-1368 (2003).

23.Chao, G. *et al.* Isolating and engineering human antibodies using yeast surface display. *Nat. Protoc.* **1**, 755-768 (2006).
24. Fleishman, S.J. *et al.* Computational design of proteins targeting the conserved stem region of influenza hemagglutinin. *Science* **332**, 816-821 (2011).
25. Gietz, R.D. & Schiestl, R.H. High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. *Nat. Protoc.* **2**, 31-34 (2007).
26. Schlehuber, S., Beste, G., & Skerra, A. A novel type of receptor protein, based on the lipocalin scaffold, with specificity for digoxigenin. *J. Mol. Biol.* **297**, 1105-1120 (2000).
27. Mazor, Y., Blarcom, T.V., Mabry, R., Iverson, B.L., & Georgiou, G. Isolation of engineered, full-length antibodies from libraries expressed in *Escherichia coli*. *Nat. Biotechnol.* **25**, 563-565 (2007).
28. Nilsson, B. *et al.* A synthetic IgG-binding domain based on staphylococcal protein A. *Protein Eng.* **1**, 107-113 (1987).
29. Kunkel, T.A. Rapid and efficient site-specific mutagenesis without phenotypic selection. *Proc. Natl. Acad. Sci. USA* **82**, 488-492 (1985).
30. Hoover, D.M. & Lubkowski, J. DNAWorks: an automated method for designing oligonucleotides for PCR-based gene synthesis. *Nucleic Acids Res.* **30**, e43 (2002).
31. Benatuil, L., Perez, J.M., Belk, J., & Hsieh, C.-M. An improved yeast transformation method for the generation of very large human antibody libraries. *Protein Eng. Des. Sel.* **23**, 155-159 (2010).
32. Whitehead, T.A. *et al.* Optimization of affinity, specificity and function of designed influenza inhibitors using deep sequencing. *Nat. Biotechnol.* **30**, 543-548 (2012).
33. Fowler, D.M., Araya, C.L., Gerard, W., & Fields, S. Enrich: software for analysis of protein function by enrichment and depletion of variants. *Bioinformatics* **27**, 3430-3431 (2011).
34. Fowler, D.M. *et al.* High-resolution mapping of protein sequence-function relationships. *Nat. Methods* **7**, 741-746 (2010).
35. McLaughlin Jr, R.N., Poelwijk, F.J., Raman, A., Gosal, W.S., & Ranganathan, R. The spatial architecture of protein function and adaptation. *Nature* **491**, 138-142 (2012).
36. Fowler, D.M. Personal Communication.
37. Otwinowski, Z., Minor, W., & Carter, C.W., Jr. Processing of X-ray diffraction data collected in oscillation mode in *Methods in Enzymology* (Academic Press, 1997), Vol. 276, pp. 307-326.
38. McCoy, A.J. *et al. Phaser* crystallographic software. *J. Appl. Crystallogr.* **40**, 658-674 (2007).
39. Potterton, E., Briggs, P., Turkenburg, M., & Dodson, E. A graphical user interface to the *CCP4* program suite. *Acta Crystallogr. Sect. D* **59**, 1131-1137 (2003).
40. Winn, M.D. *et al.* Overview of the *CCP4* suite and current developments. *Acta Crystallogr. Sect. D* **67**, 235-242 (2011).
41. Kim, Y. *et al.* Crystal structure of the conserved hypothetical protein PA3332 from *Pseudomonas aeruginosa*. *To Be Published*.
42. Murshudov, G.N., Vagin, A.A., & Dodson, E.J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. Sect. D* **53**, 240-255 (1997).
43. Emsley, P. & Cowtan, K. *Coot*: model-building tools for molecular graphics. *Acta Crystallogr. Sect. D* **60**, 2126-2132 (2004).
44. Adams, P.D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. Sect. D* **66**, 213-221 (2010).
45. Laskowski, R.A., MacArthur, M.W., Moss, D.S., & Thornton, J.M. *PROCHECK*: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **26**, 283-291 (1993).

46. Vaguine, A.A., Richelle, J., & Wodak, S.J. *SFCHECK*: a unified set of procedures for evaluating the quality of macromolecular structure-factor data and their agreement with the atomic model. *Acta Crystallogr. Sect. D* **55**, 191-205 (1999).
47. Chen, V.B. *et al. MolProbity*: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. Sect. D* **66**, 12-21 (2010).
48. Rose, P.W. *et al.* The RCSB Protein Data Bank: redesigned web site and web services. *Nucleic Acids Res.* **39**, D392-D401 (2011).
49. Rossi, A.M. & Taylor, C.W. Analysis of protein-ligand interactions by fluorescence polarization. *Nat. Protoc.* **6**, 365-387 (2011).
50. Nozaki, Y. The preparation of guanidine hydrochloride in *Methods in Enzymology* (Academic Press, 1972), Vol. Volume 26, pp. 43-50.
51. Schuck, P. Size-Distribution Analysis of Macromolecules by Sedimentation Velocity Ultracentrifugation and Lamm Equation Modeling. *Biophys. J* **78**, 1606-1619 (2000).
52. Laue, T.M., Shah, B.D., Ridgeway, T.M., & Pelletier, S.L. Computer-aided interpretation of analytical sedimentation data for proteins in *Analytical Ultracentrifugation in Biochemistry and Polymer Science*, edited by S. E. Harding & J. C. Horton (The Royal Society of Chemistry, Cambridge, UK, 1992), pp. 90-125.
53. King, N.P. *et al.* Computational Design of Self-Assembling Protein Nanomaterials with Atomic Level Accuracy. *Science* **336**, 1171-1174 (2012).