

## Improving Binding Affinity and Selectivity of Computationally Designed Ligand-Binding Proteins Using Experiments

Christine E. Tinberg and Sagar D. Khare

### Abstract

The ability to de novo design proteins that can bind small molecules has wide implications for synthetic biology and medicine. Combining computational protein design with the high-throughput screening of mutagenic libraries of computationally designed proteins is emerging as a general approach for creating binding proteins with programmable binding modes, affinities, and selectivities. The computational step enables the creation of a binding site in a protein that otherwise does not (measurably) bind the intended ligand, and targeted mutagenic screening allows for validation and refinement of the computational model as well as provides orders-of-magnitude increases in the binding affinity. Deep sequencing of mutagenic libraries can provide insights into the mutagenic binding landscape and enable further affinity improvements. Moreover, in such a combined computational–experimental approach where the binding mode is preprogrammed and iteratively refined, selectivity can be achieved (and modulated) by the placement of specified amino acid side chain groups around the ligand in defined orientations. Here, we describe the experimental aspects of a combined computational–experimental approach for designing—using the software suite Rosetta—proteins that bind a small molecule of choice and engineering, using fluorescence-activated cell sorting and high-throughput yeast surface display, high affinity and ligand selectivity. We illustrated the utility of this approach by performing the design of a selective digoxigenin (DIG)-binding protein that, after affinity maturation, binds DIG with picomolar affinity and high selectivity over structurally related steroids.

**Key words** Computational design, Rosetta macromolecular modeling, Affinity optimization, Binding selectivity, Steroid binding, Protein-small molecule interactions

---

### 1 Introduction

Computational de novo design of protein function has seen remarkable success in recent years, enabling, for example, the construction of enzymes for catalyzing reactions that are not natively catalyzed by natural enzymes [1, 2], protein binders against pathogenic proteins [3], and, more recently, the design of small-molecule binding proteins with high affinity and programmable selectivity [4]. In all cases, the initial hits obtained from the computational design

approach were weakly active, and the use of high-throughput experimental characterization to screen and improve designed proteins was critical for success. Many of the limitations of computational design methodology, including force field inaccuracies, lack of explicit modeling of solvent and properties such as protein solubility, and, more generally, our limited understanding of protein sequence–function relationships [5], were, at least in part, overcome by screening tens of computationally designed proteins using sensitive experimental assays, identifying weakly active hits and subsequently improving their efficacies using mutagenic screening or selection techniques [6]. Conversely, the directed evolution methods used to improve activities in these efforts could be made more efficient, compared to random mutagenesis approaches, by virtue of being guided by an atomic-resolution (but partially accurate) computational model of the bound state. This iterative, combined computational–experimental strategy builds upon the strengths of these complementary methods and will continue to be a key component of various protein design applications [7].

Here, we describe the experimental strategy and protocols used in our efforts to *de novo* design small-molecule binding sites in proteins—these computationally designed and subsequently laboratory-evolved proteins feature affinities and selectivities that rival those of natural small-molecule binding proteins. On the computational end, we developed and used a computational design approach, in the context of the Rosetta macromolecular modeling suite, to transplant idealized binding sites for a chosen ligand—the steroid digoxigenin (DIG)—into a set of protein scaffolds. The scaffolds were remodeled to accommodate predefined interactions to DIG, and then Rosetta Design [8] was used to optimize the binding site amino acid sequences for ligand-binding affinity. A more complete description of the computational strategy and protocols used to obtain the binders can be obtained elsewhere [4]. As mentioned above, the initial hits were weak affinity binders and could be detected only with a sensitive and relatively high-throughput yeast surface display assay that conveniently allowed testing tens of computationally designed proteins (referred to as designs hereafter) and their mutagenic libraries. We focus here on the experimental assays and methods for subsequent affinity maturation as well as selectivity modulation. Results from these experimental strategies (impact of point mutations on binding) were used to both validate (or invalidate) and refine initial designs, and models of mutagenized proteins were then used to guide further optimization, for instance, by the model-guided enumeration of ligand-proximal residue positions for which mutagenic libraries were constructed and tested. The experimental data-guided design model of one of our designs was subsequently validated by the observed atomic-resolution agreement with X-ray crystallographic

structures of a series of its variants [4]. Below, we describe our approach and offer some practical suggestions for the choices that are made while performing various steps.

---

## 2 Materials

Streptavidin–phycoerythrin (SAPE).

Yeast strain EBY100.

pCTCON2 or pETCON vector.

Highly avid ligand–biotin conjugate.

Monovalent ligand–biotin conjugate.

Monovalent ligand–fluorophore conjugate.

---

## 3 Methods

### 3.1 Overview of Approach

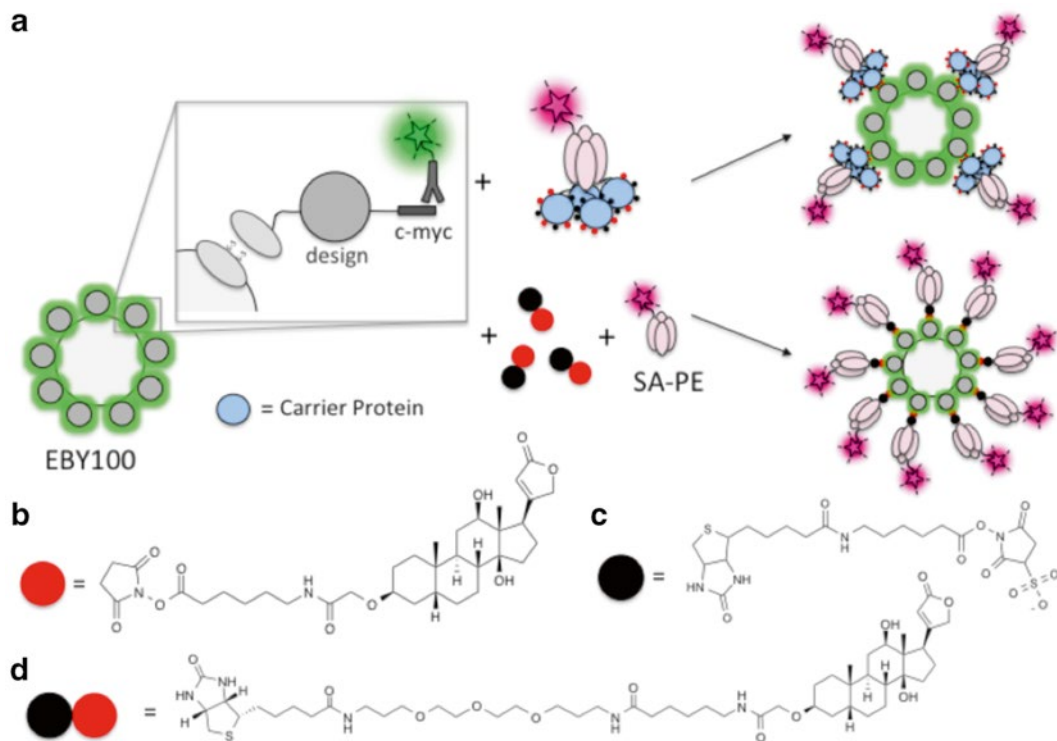
The overall goals of the approach are (1) to detect (initially weak) binding of the designed proteins and (2) to improve binding affinity and selectivity of the designed proteins. In the latter case, the choice of residue positions to mutate is based on the spatial proximity of these positions to the ligand in the computational model of the bound state. Typically, first-shell positions are chosen for site–saturation mutagenesis, beneficial mutations are combined (combinatorially), and these experimentally identified amino acid substitutions are used to refine or invalidate initial design model. For the optimized variant, a single-site mutagenic library at both first- and second-shell residue positions is generated, and high-throughput sequencing of screened libraries is used to guide further affinity improvements. The experimental data-guided computational model is then used to design mutations to predictively modulate the selectivity of designed proteins for the small molecule over a series of congeners.

### 3.2 Initial Screen of Computationally Designed Proteins

1. Designed proteins are tested for ligand binding using yeast surface display [9]. We used the vector pETCON and the *NdeI/XhoI* restriction sites in this vector to clone synthetic genes of the designs. Standard yeast surface display materials and protocols were used for growth and induction unless stated otherwise below.
2. For hydrophobic ligands (such as DIG) and designed proteins that are expected to have low affinities, it is important to guard against false-positives as exposed hydrophobic patches in proteins can nonspecifically bind the ligand with low affinity. To control for nonspecific binding, we used proteins that are both structurally and functionally unrelated to designed proteins as controls. Negative controls for binding were two tandem Z

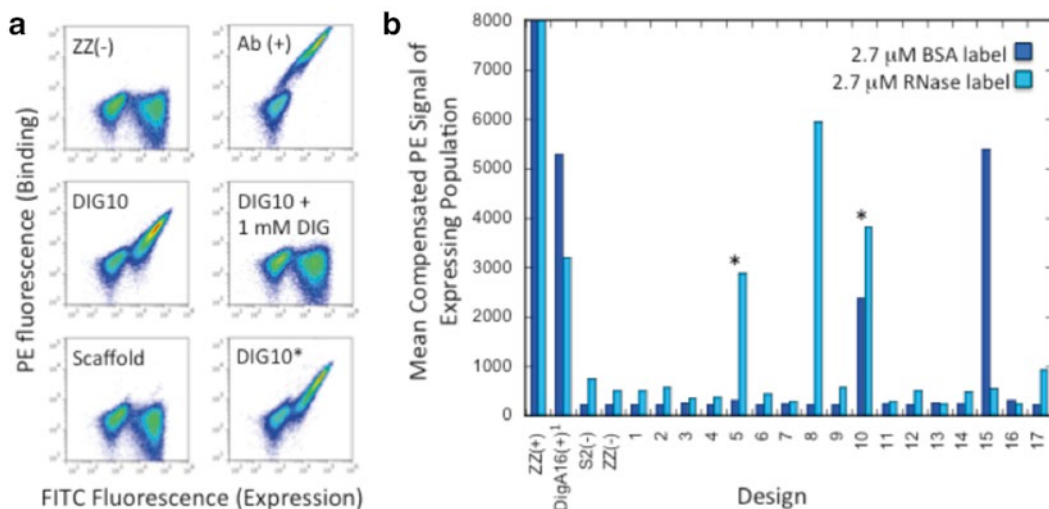
domains of protein A (ZZ domain) [10, 11] and a mutagenic library of HIV glycoprotein (gp120) variants developed for an unrelated project.

3. The genes for the “negative control” proteins as well as designs cloned into the pETCON vector are transformed into cells of the yeast strain EBY100 using lithium acetate and polyethylene glycol [12]. Transformants are plated on selective media (C – *ura –trp*) that select for both the strain and the vector.
4. Freshly transformed cells are inoculated into 1 mL of SDCAA media [9] and grown at 30 °C, 200 rpm. After ~12 h, 1e7 cells are collected by centrifugation at 1700 × *g* for 3 min and resuspended in 1 mL of SGCAA media to induce protein expression.
5. Following induction for 24–48 h at 18 °C, 4e6 cells are collected by centrifugation and washed twice by incubation with PBSF (PBS supplemented with 1 g/L of BSA) for 10 min at room temperature. Induction times and temperatures required to obtain the highest expression levels of displayed proteins can vary and need to be empirically determined. For our system, 24–48 h at 18 °C was optimal.
6. For proteins expressed from their gene in the pETCON vector, yeast surface protein expression can be monitored by the binding of anti-cmyc-FITC antibody to the C-terminal myc-epitope tag of the displayed protein (Fig. 1a).
7. Small-molecule (in our case, DIG) binding is assessed by quantifying the phycoerythrin (PE) fluorescence of the displaying yeast population following incubation with small-molecule-biotinylated protein conjugates: DIG-BSA-biotin, DIG-RNase-biotin (Fig. 1b, c), or DIG-PEG<sub>3</sub>-biotin (Fig. 1d) in our case, and streptavidin–phycoerythrin (SAPE). *See Note 1.*
8. Following a 2–4-h incubation at 4 °C in the dark on a rotator, cells are collected by centrifugation at 1700 × *g* for 3 min and washed with 200 μL of PBSF at 4 °C.
9. Cell pellets are resuspended in 200 μL of ice-cold PBSF immediately before use. For detecting weak affinity binders, it is important to keep the samples on ice until resuspension and resuspend immediately before use.
10. Cellular fluorescence is 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 is compensated to minimize bleed-over contributions from the FITC fluorescence channel.
11. While negative controls are important (*see* Subheading 3.2, **step 3**), positive controls of varying affinities, if available, should be used to validate, and tune the sensitivity of, the assay.



**Fig. 1** Outline of assay used for detection and evolution of binding affinity of designed proteins. **(a)** Designs are expressed on the surface of yeast using the plasmid pETCON as described by Wittrup and co-workers. A c-myc tag is attached at the C-terminus of the protein to enable detection using an anti-c-myc antibody that is conjugated to a fluorophore (e.g., FITC, green). Binding can be detected in a high-avidity format to identify initial hits (*top*) or low-avidity format to enable more sensitive detection of affinity increase during affinity maturation (*bottom*). **(b, c)** NHS esters of DIG and biotin that are used for conjugation to a carrier protein (e.g., BSA or RNase) in the high-avidity format. **(d)** The DIG-biotin conjugate that was used in the low-avidity format

In our case, two positive controls having different affinities for digoxigenin were used in the binding assay: a previously engineered steroid binding protein DigA16 [13] and a commercially available anti-DIG monoclonal antibody 9H27L19 (Fig. 2). Experiments using DigA16 were conducted in an identical fashion to design DIG1-17. For those employing the anti-DIG antibody, an F<sub>c</sub>-region-binding protein, the ZZ domain (*see* Subheading 3.2, step 3), was displayed on the yeast cell surface, and washed cells were resuspended in 20  $\mu$ L of PBSF with 2  $\mu$ L of rabbit anti-DIG mAb 9H27L19. Following a 30-min incubation at 4  $^{\circ}$ C on a rotator, excess antibody was removed by washing the cells with 200  $\mu$ L of PBSF. Labeling reactions were then performed as above.

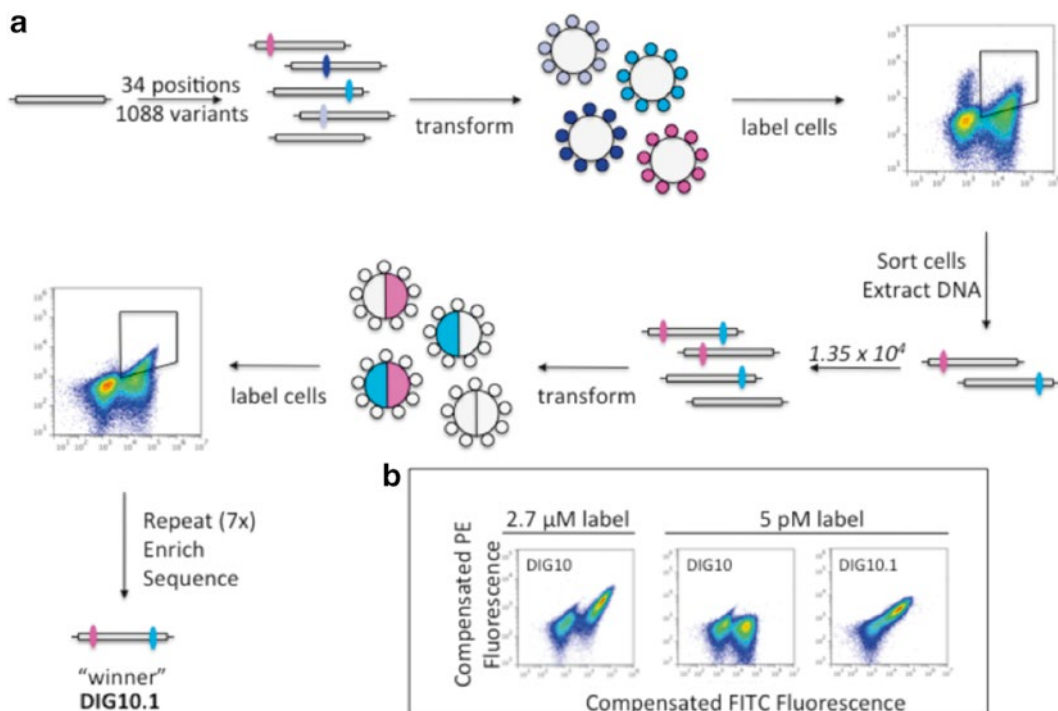


**Fig. 2** Typical assay results for hits obtained in a set of computationally designed proteins. **(a)** Example results and validation experiments carried out for a hit identified from the binding assay showing no binding signal for negative control (ZZ (-)), high binding signal for positive control (Ab (+)), binding signal for the design (DIG10), no binding signal for design incubated with excess unlabeled DIG (DIG10 + 1 mM DIG), no binding signal for the wild-type scaffold protein on which the design DIG10 is based (scaffold), and similar binding signal (as DIG10) when an alternative carrier protein, RNase, is used (DIG10\*). **(b)** Binding signals for controls and all 17 tested designs. Designs DIG10 showed reproducible binding signals with both carrier proteins, DIG5 and DIG8 showed high signals with RNase carrier protein but not BSA, and DIG15 showed high signals with BSA but not RNase. Tests described in **(a)** identified DIG10 and DIG5 as being specific binders to DIG. These were used for further affinity maturation

12. To test if the hits identified above are not false-positives on account of binding to other assay components (such as SAPE), it is important to perform competition experiments with the free ligand (Fig. 2a). See **Note 2**.
13. To further ensure specific binding to the small molecule, knockout mutagenesis of key interacting residues is performed. Residues that interact with the ligand in the computational model are mutated to amino acids that disfavor binding. This step serves to confirm that the ligand and not other assay components are binding the design as well as confirm the design model.

### 3.3 Affinity Improvement Using Yeast Surface Display Selections and Fluorescence-Activated Cell Sorting of Mutagenic Libraries

1. Based on the identified hits in Subheading 3.2, affinity maturation is performed using single site-saturation mutagenesis (SSM) library constructed by Kunkel mutagenesis [14] using degenerate NNK primers (Fig. 3).
2. Positions for mutagenesis are chosen based on the computational design model. Positions are chosen from the model based on the following requirements: (1) they have C $\alpha$  within



**Fig. 3** Directed evolution of computational designs. **(a)** Outline of scheme used for site-directed mutagenesis of designs for affinity improvement. Several rounds of single site-saturation mutagenesis followed by combinatorial mutagenesis using identified beneficial single mutations are performed to obtain affinity improvements. **(b)** Comparison of the binding properties of the initial hit (DIG10) with the affinity matured variant (DIG10.1). High binding signals are detectable at ~6 orders-of-magnitude lower labeled ligand concentrations after affinity maturation

7 Å of any ligand heavy atom, and/or (2) they have C $\alpha$  within 9 Å of any ligand heavy atom and C $\beta$  closer to any heavy atom in the ligand than C $\alpha$ . The theoretical library size can be calculated (in our case, we chose 34 positions for design DIG10 yielding a size of 1088 clones).

3. Kunkel mutagenesis of each position using mutagenic oligonucleotides is carried out independently. DNA from each reaction is dialyzed into dH<sub>2</sub>O using a 0.025  $\mu$ m membrane filter, and then the dialyzed reaction mixtures are pooled, concentrated to a volume of <10  $\mu$ L using a Savant SpeedVac centrifugal vacuum concentrator, and transformed into yeast strain EBY100 using the method of Benatuil [15]. Typical yields are 1E7–1E8. See **Note 3**.
4. After transformation, cells are grown in 250 mL of SDCAA media for 36 h at 30 °C. Cells (5e8) are collected by centrifugation at 1700  $\times g$  for 4 min, resuspended in 50 mL of SGCAA media, and induced at 18 °C for 24 h.

5. Cells are subjected to multiple (we used three) rounds of permissive cell sorting to enrich for improved variants. During each round of sorting, cells are washed and then labeled with a preincubated mixture of 2.66  $\mu\text{M}$  DIG-BSA-biotin, 644 nM SAPE, and anti-cmyc-FITC as noted above for single clones. During each round, the top  $\sim 10\%$  of cells in the PE channel are collected. It is important to sort 10–100 times the library transformation efficiency to ensure that each clone in the library is sampled during the sort. *See Note 4.*
6. After each round of sorting, cells are grown in SDCAA for 24 h and then induced in SGCAA for 24 h before the next sort. It is important to recover the cells in this way so that low representation clones are allowed to amplify.
7. After the final sort, an increase in the mean compensated PE fluorescence of the expressing population of the sorted cells compared to that of the original design indicates the presence of a point mutant(s) with increased binding affinity.
8. After each sort, a portion of cells are plated and grown at 30  $^{\circ}\text{C}$ . Plasmids from individual colonies are harvested and the gene is amplified by PCR. Sanger sequencing is used to sequence at least ten colonies from each population to identify mutations that increase affinity.

### **3.4 Combinatorial Mutagenesis Using Identified Beneficial Single-Point Mutations**

1. Beneficial mutations identified in the SSM library (Subheading 3.3) are combined by Kunkel mutagenesis [14] using degenerate primers. At each mutagenized position, the original DIG10 amino acid and chemically similar amino acids to those identified in the first round of directed evolution are also allowed, resulting in a combinatorial library.
2. Four independent Kunkel reactions using different mutagenic oligonucleotide concentrations ranging from 36 to 291 nM during polymerization are performed to minimize sequence-dependent priming bias. For the same reason, oligonucleotides encoding native substitutions contain at least one codon base change.
3. Library DNA is pooled, prepared as above, and transformed into electrocompetent *E. coli* strain BL21(DE3) cells (1800 V, 200  $\Omega$ , 25  $\mu\text{F}$ ). Library plasmid DNA is isolated from expanded cultures. Gene insert is amplified from 10 ng of library DNA by 30 cycles of PCR (98  $^{\circ}\text{C}$  10 s, 61  $^{\circ}\text{C}$  30 s, 72  $^{\circ}\text{C}$  15 s) using Phusion high-fidelity polymerase with the pCTCON2r and pCTCON2f primers. *See Note 5.*
4. Yeast EBY100 cells are transformed with 4.0  $\mu\text{g}$  of PCR-purified DNA insert generated in the previous step and 1.0  $\mu\text{g}$  of gel-purified pETCON digested with *NdeI* and *XhoI* using the method of Benatuil [15], yielding  $1\text{E}7$ – $1\text{E}8$  transformants.

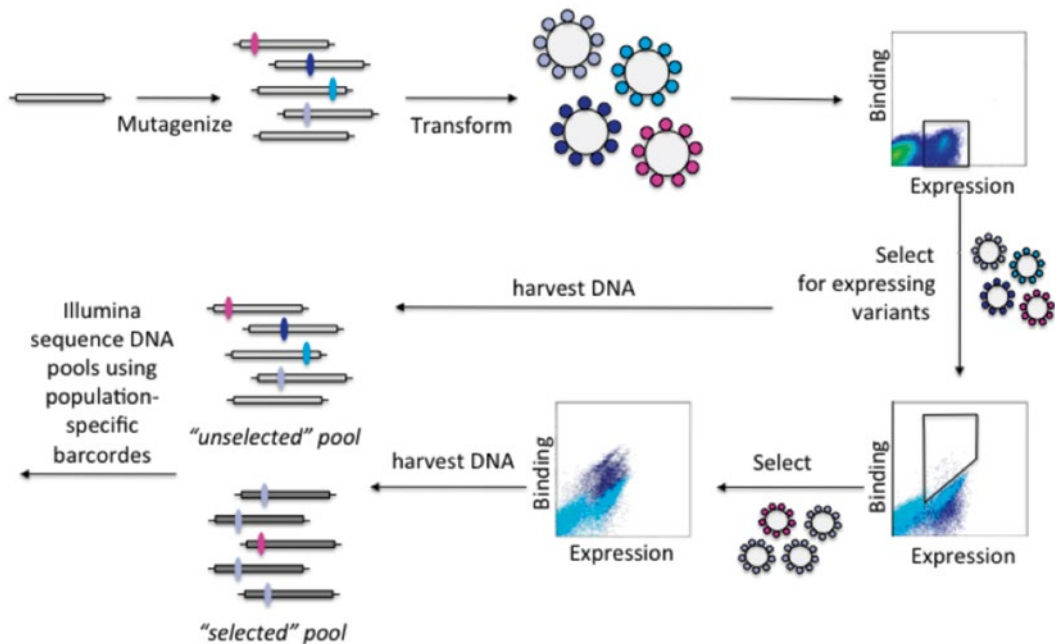


After transformation, cells are grown in 150 mL of low-pH SDCAA media supplemented with Pen/Strep for 48 h at 30 °C. Cells ( $\sim 5 \times 10^8$ ) are collected by centrifugation at  $1700 \times g$  for 4 min, resuspended in 50 mL of SGCAA media, and induced at 18 °C for 24 h.

5. Cells are subjected to several rounds of cell sorting (we performed seven rounds). For the first four rounds, cells are washed and then labeled with a preincubated mixture of small-molecule BSA-biotin, SAPE, and anti-cmyc-FITC as noted above for single clones. Small-molecule-label concentrations can be decreased progressively in every round to increase the selection stringency. It is important to maintain a 4:1 (biotin/SAPE) ratio. For example, our concentrations for rounds one through four were (1) 1  $\mu$ M 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. Selection stringency is increased in each round by dropping the label concentration or decreasing the avidity of the label. Note that these concentrations in this example refer to the concentration of carrier protein molecules, not DIG molecules.
6. To ensure that the identified mutations do not select for binding to the carrier protein (e.g., BSA in our case) or a specific linkage between small molecule and carrier protein or other assay components (e.g., SAPE), it is important to use an unrelated protein for labeling with small molecule (Fig. 3b). For rounds five through seven, we used DIG-RNase-biotin in a multistep labeling procedure to minimize selection for carrier protein (BSA) binding. The use of RNase also allowed a larger dynamic range in several control experiments. DIG-RNase-biotin label concentrations were 10, 5, and 5 pM (concentrations referenced to RNase) for rounds five through seven, respectively.
7. At least ten clones from each round are sequenced as noted for the SSM library. After several rounds, the library typically converges to a small number of sequences differing by a single or a few point substitutions.

### **3.5 Mutagenic Libraries and Deep Sequencing**

1. Paired-end 151 Illumina sequencing is used to simultaneously assess the effects of mutation on binding.
2. A number of mutagenic libraries are designed, based on the distribution of mutagenized positions in and length of the gene under consideration and the optimal read length of the deep-sequencing approach being used (Fig. 4). In our case, two libraries were constructed to allow optimal probing of the mutagenic landscape using 151-bp paired-end sequencing on an Illumina MiSeq.



**Fig. 4** Preparation for the deep sequencing-based illumination of the mutagenic landscape of binding. A mutagenic library is synthesized (see main text) and is screened first for expression and then binding. Harvested DNA at both stages is deep sequenced, and the relative frequency of individual mutations in the selected and unselected pools is used to compute the landscape

3. For each library, the full-length protein gene having additional pETCON overlap fragments at either end for yeast homologous recombination is assembled via recursive PCR. To introduce mutations, degenerate PAGE-purified oligos are used in which selected positions within the binding site are doped with a small amount of each nonnative base at a level expected to yield 1–2 mutations per gene. For this study, we ordered custom-doped oligos. See **Note 6**.
4. For each library assembly, overlapping oligonucleotides, including overlapping regions with the ends of the pETCON plasmid, are combined with dNTPs, 5× Phusion buffer HF, DMSO, and Phusion high-fidelity polymerase. Full-length products are assembled by PCR, and correctly assembled PCR products are amplified by a second round of PCR using oligonucleotides that overlap with the pETCON plasmid. Correct length PCR products are isolated using agarose gel electrophoresis and are purified using a Qiagen PCR cleanup kit and eluted in ddH<sub>2</sub>O.
5. Yeast EBY100 cells are transformed with 5.4 μg of library DNA insert and 1.8 μg of gel-purified pETCON digested with *NdeI* and *XhoI* using the method of Benatuil [15], yielding ~1e6 transformants.

6. After transformation, cells are grown for 24 h in 100 mL of low-pH SDCAA media supplemented with Pen/Strep at 30 °C, passaged once, and grown for an additional 24 h under the same conditions. Cells ( $\sim 5 \times 10^8$ ) are collected by centrifugation, resuspended in 50 mL of SGCAA, and induced overnight at 18 °C.
7. Induced cells ( $3 \times 10^7$ ) were labeled with 4  $\mu\text{L}$  of anti-cymc-FITC in 200  $\mu\text{L}$  of PBSF for 20 min at 4 °C to label cells expressing full-length protein variants. Then, labeled cells are washed with PBSF and sorted. In this first round of sorting, all cells showing a positive signal for protein expression are collected.
8. Cells were recovered overnight in  $\sim 1$  mL of low-pH SDCAA supplemented with Pen/Strep at 30 °C, pelleted by centrifugation at  $1700 \times g$  for 4 min, resuspended in 5 mL of low-pH SDCAA supplemented with Pen/Strep, and grown for an additional 24 h at 30 °C.
9. Cells ( $\sim 2 \times 10^7$ ) are collected by centrifugation, resuspended in 2 mL of SGCAA, and induced overnight at 18 °C.
10. Induced cells from expression-sorted libraries and two reference samples of the template protein ( $5 \times 10^6$  cells per sample) prepared similarly are washed with 600  $\mu\text{L}$  of PBSF and then labeled with a chosen concentration of the small-molecule-biotin complex (100 nM of DIG-PEG<sub>3</sub>-biotin in our case) in 400  $\mu\text{L}$  of PBSF for the libraries or 200  $\mu\text{L}$  of PBSF for the reference samples for  $>3$  h at 4 °C. The concentration of the label should be sufficient to observe a binding signal with the parent clone. Labeled cells are washed with 200  $\mu\text{L}$  of PBSF and then incubated with a secondary label solution of 0.8  $\mu\text{L}$  of SAPE (Invitrogen) and 4  $\mu\text{L}$  of anti-cymc-FITC in 400  $\mu\text{L}$  of PBSF for 8 min at 4 °C. Cells are washed with 200  $\mu\text{L}$  PBSF, resuspended in either 800  $\mu\text{L}$  of PBSF for the libraries or 400  $\mu\text{L}$  of PBSF for the reference samples, and sorted.
11. Clones having binding signals higher than that of the parent reference sample are collected using FACS. Collected cells are recovered overnight in  $\sim 1$  mL of low-pH SDCAA supplemented with Pen/Strep at 30 °C, pelleted by centrifugation at  $1700 \times g$  for 4 min, resuspended in 2 mL of low-pH SDCAA supplemented with Pen/Strep, and grown for an additional 24 h at 30 °C. Cells ( $2 \times 10^7$ ) are resuspended in 2 mL of SGCAA and induced overnight at 18 °C.
12. To reduce noise from the first round of cell sorting, the sorted libraries are labeled and subjected to a second round of cell sorting using the same conditions and gates as in the first round. Collected cells are recovered and grown as described above.
13. One hundred million cells from the expression-sorted libraries and at least  $2 \times 10^7$  cells from doubly sorted library are pelleted by

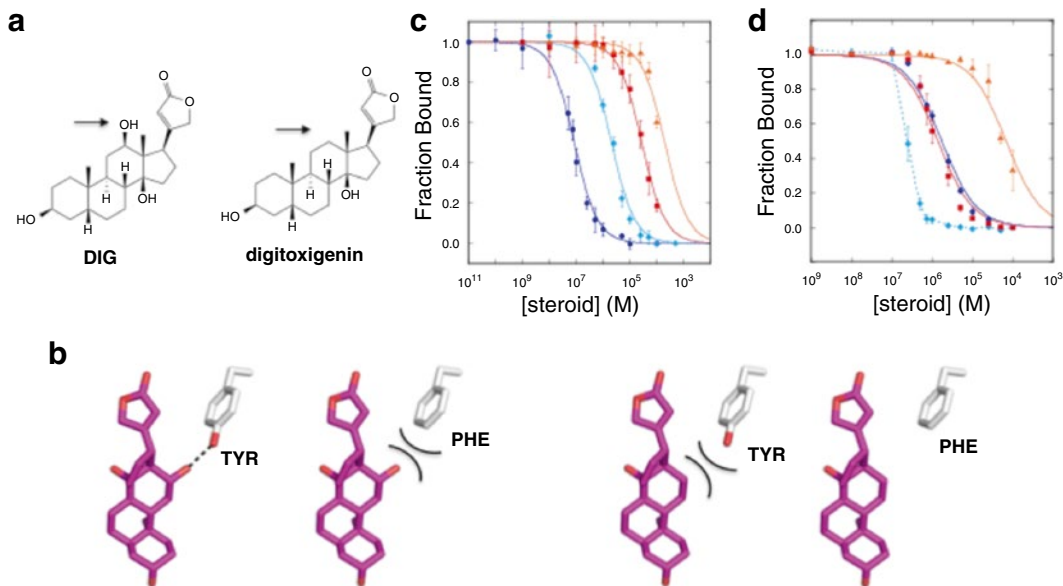
centrifugation at  $1700 \times g$  for 4 min, resuspended in 1 mL of freezing solution (50 % YPD, 2.5 % glycerol), transferred to cryogenic vials, slow-frozen in an isopropanol bath, and stored at  $-80^\circ\text{C}$  until further use.

### **3.6 Next-Generation Library Sequencing**

1. Library DNA is prepared as detailed previously [16]. Illumina adapter sequences and unique library barcodes are appended to each library pool through PCR amplification using population-specific HPLC-purified primers.
2. The library amplicons are verified on a 2 % agarose gel stained with SYBR Gold and then purified using an Agencourt AMPure XP bead-based purification kit. Each library amplicon is denatured using NaOH and then diluted to 6 pM. A sample of *PbiX* control DNA is 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 is prepared by pooling 300  $\mu\text{L}$  of 6 pM *PbiX* control DNA (50 %), 102  $\mu\text{L}$  of 6 pM expression-sorted library, and 33  $\mu\text{L}$  of 6 pM sorted libraries each.
3. DNA is sequenced in paired-end mode on an Illumina MiSeq using a 300-cycle reagent kit and custom HPLC-purified primers.
4. Data from each next-generation sequencing library is demultiplexed using the unique library barcodes added during the amplification steps. For example, in our experiment, of a total 5,630,105 paired-end reads, 2,531,653 reads were mapped to library barcodes. For each library, paired-end reads are fused and filtered for quality ( $\text{Phred} \geq 30$ ).
5. The resulting full-length reads are aligned against the relevant segments of the template gene sequence using scripts from the software package Enrich [17].
6. For single mutations having  $\geq 7$  counts in the original input library, a relative enrichment ratio between the input library and each selected library is calculated [16, 18, 19]. This cutoff value is used to establish statistical significance in the final data set.
7. A pseudocount value (0.3 in our case) is added to the total reads for each selected library mutation, to allow calculation of enrichment values for mutations that disappeared completely during selection.

### **3.7 Selectivity Assays by Equilibrium Fluorescence Polarization Competition Assays**

1. To verify binding and to measure binding dissociation constants, fluorescence polarization assays are using purified protein and fluorescent ligand (Fig. 5). Fluorescence polarization-based affinity measurements of designs and their evolved variants are performed as noted previously [20] using a small-molecule-fluorescent dye conjugate (in our case Alexa488-conjugated DIG; DIG-PEG<sub>3</sub>-Alexa488).



**Fig. 5** Measuring and modulating selectivity of designed proteins guided by the computational model of binding. **(a)** The specificity of the designed binding protein can be modulated for congeneric ligands that differ in their chemical structure by as little as a hydroxyl group, as is the case with DIG and digitoxigenin. **(b)** Guided by the computational model of DIG10.3, in which tyrosine side chain groups were positioned to make hydrogen bonds with the DIG hydroxyl, a Tyr to Phe substitution was chosen, and **(c and d)** the selectivity of DIG10.3 and DIG10.3\_Y110F was measured as described in the text. Robust specificity switching was observed (compare **c** and **d**), demonstrating the programmability of computationally designed ligand-binding proteins

- In a typical experiment, the concentration of the conjugate is fixed near the  $K_d$  of the interaction being monitored, and the effect of the increasing concentrations of protein on the fluorescence anisotropy of the fluorescent dye is determined.
- Fluorescence anisotropy ( $r$ ) is measured in 96-well plate format at appropriate excitation and emission wavelengths ( $\lambda_{\text{ex}} = 485 \text{ nM}$  and  $\lambda_{\text{em}} = 538 \text{ nM}$  using a 515 nm emission cut-off filter, in our case). In all experiments, PBS (pH 7.4) is used as the buffer system and the temperature is 25 °C. For high-affinity complexes, it is important to use NBS-coated plates to improve the signal-to-noise aspect.
- Equilibrium dissociation constants ( $K_d$ ) are determined by fitting plots of the anisotropy averaged over a period of 20–40 min (equilibrium) after reaction initiation versus protein concentration to Eq. 1:

$$A = A_f + (A_b - A_f) \times \left( \frac{([L]_T + K_D + [R]_T) - \sqrt{(-[L]_T - K_D - [R]_T)^2 - 4[L]_T[R]_T}}{2[L]_T} \right) \quad (1)$$

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.

5. For ensuring assay robustness, reported  $K_d$  values should represent the average of at least three independent measurements with at least two separate batches of purified protein.

### **3.8 Fluorescence Polarization Equilibrium Competition Binding Assays**

1. Fluorescence polarization equilibrium competition binding assays are used to determine the binding affinities of designed proteins and their variants for unlabeled ligands and congeneric compounds (for which selectivity measurements and modulation is desired; in our case, these were digoxigenin, digitoxigenin, progesterone,  $\beta$ -estradiol, and digoxin; Fig. 5a). During the computational design procedure, careful placement of interacting amino acid side chains allows for explicit design of selectivity (Fig. 5b). Selectivity can be switched by manipulation of these residues. In our case, we considered Tyr to Phe mutations as candidates to switch the specificity toward more hydrophobic steroids (Fig. 5b). The labeled small molecule (Subheading 3.5) is used, and the ability of different ligands to inhibit its binding to the designed protein variant is used to calculate their affinities for the protein.
2. In a typical experiment, the concentration of labeled small molecule is kept near or below the  $K_d$  of the interaction being monitored, the concentration of protein is fixed at a saturating value such that >95 % the labeled small molecule in the system is bound to protein, and the effects of increasing concentrations of unlabeled ligand on the fluorescence anisotropy of the fluorescent dye are determined as described above in Subheading 3.5.
3. If the ligands being considered are insoluble or sparingly soluble in aqueous buffers, stock solutions are typically made in organic solvents such as DMSO or methanol. For each ligand concentration, a negative control sample containing only the appropriate dilution of the corresponding organic solvent-only control solution (in aqueous assay buffer, PBS in our case) is measured. While we found that at all concentrations employed, methanol or DMSO solvents did not affect fluorescence anisotropy with our binding assay. However, correction for this effect must be made.
4. The concentration of total unlabeled ligand producing 50 % binding signal inhibition ( $I_{50}$ ) is determined by fitting a plot of the anisotropy averaged over a period of 30 min to 3 h after reaction initiation versus unlabeled ligand concentration [20]. See Note 7.

5. For cases in which  $K_d$  for competitor is much smaller than  $K_d$  for the labeled small molecule, the data cannot be fit to the model and only qualitative conclusions can be reached (Fig. 5c, d).
6. The inhibition constant for each protein–ligand interaction,  $K_i$ , is calculated from the measured  $IC_{50}$  and the  $K_d$  of the protein–label interaction according to a model accounting for receptor-depletion conditions [20].
7.  $IC_{50}$  values, the concentrations of free unlabeled ligand producing 50 % binding signal inhibition, are calculated from the measured  $I_{50}$  values [20].
8. For assay robustness, reported  $I_{50}$  and subsequent  $K_i$  values should 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 experiment.

---

## 4 Notes

1. In a typical experiment using DIG-BSA-biotin or DIG-RNase-biotin, 4e6 cells are resuspended in 50  $\mu$ L of a premixed solution of PBSF containing a 1:100 dilution of anti-cmyc-FITC, 2.66  $\mu$ M DIG-BSA-biotin or DIG-RNase-biotin, and 664 nM SAPE (to achieve a 1:4 streptavidin/biotin ratio). The use of carrier protein–ligand molecules offers a highly avid label for detection of weak binders. The avidity of the system (i.e., number of copies of the ligand on the carrier protein) can be tailored by changing the concentration of reagents in the carrier protein–ligand conjugation reaction.
2. In our case, competition assays with free digoxigenin were performed: between 750  $\mu$ M 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, and binding of the resultant samples was determined as above. For “true” hits, the addition of excess free ligand should abolish the binding signal. 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.
3. It is best to restrict the library size such that each clone in the library can be oversampled by 10–100 in the transformed pool.
4. The stringency of the sort can be increased from round to round in order to hone in on one or a few binding clones by lowering the label concentration. However, it is important for the first round to be permissive to ensure that clones with low representation in the library pool are able to enrich if they have desirable binding properties.

5. Transformation of Kunkel libraries is typically not as efficient as is transformation of other library formats, so we found that preparing the library DNA in more efficient *E. coli* prior to transformation into yeast led to higher overall transformation efficiencies and a better chance of having complete clone coverage in the transformed library.
6. It is best to restrict the total library size so that each clone can be oversampled at 10–100 in both the transformed library and in the sequencing run (Illumina MiSeq runs currently yield up to  $10^7$  reads/run).
7. Note that for some experiments, due to the lack of solubility, limiting competitor ligand concentrations can make it impossible to collect data in the regime of complete inhibition. In these cases, data are fit by fixing the anisotropy at infinite steroid concentration to a value measured for other ligands for which this value could be determined experimentally.

---

## Acknowledgment

SDK acknowledges support from the NSF (grant MCB1330760).

## References

1. Rothlisberger D, Khersonsky O, Wollacott AM, Jiang L, DeChancie J, Betker J, Gallaher JL, Althoff EA, Zanghellini A, Dym O, Albeck S, Houk KN, Tawfik DS, Baker D (2008) Kemp elimination catalysts by computational enzyme design. *Nature* 453(7192):190–195. doi:[10.1038/nature06879](https://doi.org/10.1038/nature06879)
2. Jiang L, Althoff EA, Clemente FR, Doyle L, Rothlisberger D, Zanghellini A, Gallaher JL, Betker JL, Tanaka F, Barbas CF 3rd, Hilvert D, Houk KN, Stoddard BL, Baker D (2008) De novo computational design of retro-aldol enzymes. *Science* 319(5868):1387–1391. doi:[10.1126/science.1152692](https://doi.org/10.1126/science.1152692)
3. Fleishman SJ, Whitehead TA, Ekiert DC, Dreyfus C, Corn JE, Strauch EM, Wilson IA, Baker D (2011) Computational design of proteins targeting the conserved stem region of influenza hemagglutinin. *Science* 332(6031):816–821. doi:[10.1126/science.1202617](https://doi.org/10.1126/science.1202617)
4. Tinberg CE, Khare SD, Dou J, Doyle L, Nelson JW, Schena A, Jankowski W, Kalodimos CG, Johnsson K, Stoddard BL, Baker D (2013) Computational design of ligand-binding proteins with high affinity and selectivity. *Nature* 501(7466):212–216. doi:[10.1038/nature12443](https://doi.org/10.1038/nature12443)
5. Khare SD, Fleishman SJ (2013) Emerging themes in the computational design of novel enzymes and protein-protein interfaces. *FEBS Lett* 587(8):1147–1154. doi:[10.1016/j.febslet.2012.12.009](https://doi.org/10.1016/j.febslet.2012.12.009)
6. Fleishman SJ, Baker D (2012) Role of the biomolecular energy gap in protein design, structure, and evolution. *Cell* 149(2):262–273. doi:[10.1016/j.cell.2012.03.016](https://doi.org/10.1016/j.cell.2012.03.016)
7. Griss R, Schena A, Reymond L, Patiny L, Werner D, Tinberg CE, Baker D, Johnsson K (2014) Bioluminescent sensor proteins for point-of-care therapeutic drug monitoring. *Nat Chem Biol* 10(7):598–603. doi:[10.1038/nchembio.1554](https://doi.org/10.1038/nchembio.1554)
8. Kuhlman B, Baker D (2000) Native protein sequences are close to optimal for their structures. *Proc Natl Acad Sci U S A* 97(19):10383–10388
9. Chao G, Lau WL, Hackel BJ, Sazinsky SL, Lippow SM, Wittrup KD (2006) Isolating and engineering human antibodies using yeast surface display. *Nat Protoc* 1(2):755–768
10. Mazor Y, Blarcom TV, Mabry R, Iverson BL, Georgiou G (2007) Isolation of engineered, full-length antibodies from libraries expressed in *Escherichia coli*. *Nat Biotechnol* 25(5):563–565
11. Nilsson B, Moks T, Jansson B, Abrahmsén L, Elmblad A, Holmgren E, Henrichson C, Jones TA, Uhlén M (1987) A synthetic IgG-binding



- domain based on staphylococcal protein A. *Protein Eng* 1(2):107–113
12. Gietz RD, Schiestl RH (2007) High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. *Nat Protoc* 2(1):31–34
  13. Schlehuber S, Beste G, Skerra A (2000) A novel type of receptor protein, based on the lipocalin scaffold, with specificity for digoxigenin. *J Mol Biol* 297(5):1105–1120
  14. Kunkel TA (1985) Rapid and efficient site-specific mutagenesis without phenotypic selection. *Proc Natl Acad Sci U S A* 82(2):488–492
  15. Benatuil L, Perez JM, Belk J, Hsieh C-M (2010) An improved yeast transformation method for the generation of very large human antibody libraries. *Protein Eng Des Sel* 23(4):155–159
  16. Whitehead TA, Chevalier A, Song Y, Dreyfus C, Fleishman SJ, De Mattos C, Myers CA, Kamisetty H, Blair P, Wilson IA, Baker D (2012) Optimization of affinity, specificity and function of designed influenza inhibitors using deep sequencing. *Nat Biotechnol* 30(6):543–548
  17. Fowler DM, Araya CL, Gerard W, Fields S (2011) Enrich: software for analysis of protein function by enrichment and depletion of variants. *Bioinformatics* 27(24):3430–3431
  18. Fowler DM, Araya CL, Fleishman SJ, Kellogg EH, Stephany JJ, Baker D, Fields S (2010) High-resolution mapping of protein sequence-function relationships. *Nat Methods* 7(9):741–746
  19. McLaughlin RN Jr, Poelwijk FJ, Raman A, Gosal WS, Ranganathan R (2012) The spatial architecture of protein function and adaptation. *Nature* 491(7422):138–142
  20. Rossi AM, Taylor CW (2011) Analysis of protein-ligand interactions by fluorescence polarization. *Nat Protoc* 6(3):365–387