THE PROTEIN SOCIETY

Sampling and energy evaluation challenges in ligand binding protein design

Jiayi Dou,[1,2,3] Lindsey Doyle,[4] Per Jr. Greisen,[5] Alberto Schena,[6] Hahnbeom Park,[3,7] Kai Johnsson,[6] Barry L. Stoddard,[4] and David Baker[3,7,8]*

[1]Department of Bioengineering, University of Washington, Seattle, Washington
[2]Graduate Program in Biological Physics, Structure, and Design, University of Washington, Seattle, Washington
[3]Institute for Protein Design, University of Washington, Seattle, Washington
[4]Division of Basic Sciences, Fred Hutchinson Cancer Research Center, Seattle, Washington
[5]Global Research, Novo Nordisk A/S, DK-2760, Måløv, Denmark
[6]Ecole Polytechnique Federale de Lausanne, Institute of Chemical Sciences and Engineering, Institute of Bioengineering, National Centre of Competence in Research (NCCR) in Chemical Biology, Lausanne, Switzerland
[7]Department of Biochemistry, University of Washington, Seattle, Washington
[8]Howard Hughes Medical Institute, University of Washington, Seattle, Washington

Received 5 July 2017; Accepted 2 October 2017
DOI: 10.1002/pro.3317
Published online 4 October 2017 proteinscience.org

Abstract: The steroid hormone 17α-hydroxylprogesterone (17-OHP) is a biomarker for congenital adrenal hyperplasia and hence there is considerable interest in development of sensors for this compound. We used computational protein design to generate protein models with binding sites for 17-OHP containing an extended, nonpolar, shape-complementary binding pocket for the four-ring core of the compound, and hydrogen bonding residues at the base of the pocket to interact with carbonyl and hydroxyl groups at the more polar end of the ligand. Eight of 16 designed proteins experimentally tested bind 17-OHP with micromolar affinity. A co-crystal structure of one of the designs revealed that 17-OHP is rotated 180° around a pseudo-two-fold axis in the compound and displays multiple binding modes within the pocket, while still interacting with all of the designed residues in the engineered site. Subsequent rounds of mutagenesis and binding selection improved the ligand affinity to nanomolar range, while appearing to constrain the ligand to a single bound conformation that maintains the same "flipped" orientation relative to the original design. We trace the discrepancy in the design calculations to two sources: first, a failure to model subtle backbone changes which alter the distribution of sidechain rotameric states and second, an underestimation of the energetic cost of desolvating the carbonyl and hydroxyl groups of the ligand. The difference between design model and crystal structure thus arises from both sampling limitations and energy function inaccuracies that are exacerbated by the near two-fold symmetry of the molecule.

Keywords: Computational protein design; ligand binding design; hydrophobic small molecules; 17-hydroxylprogesterone

Introduction

The hormone 17α-hydroxylprogesterone(17-OHP) (Fig. S1.a**1**) is a biomarker for a group of autosomal recessive disorders called congenital adrenal hyperplasia(-CAH).[1] Quick and accurate monitoring of blood concentration of 17-OHP is critical for disease diagnosis and treatment. Newborn screening based on elevated levels of 17-OHP is performed in many countries for early diagnosis of CAH. The most widely

Additional Supporting Information may be found in the online version of this article.

*Correspondence to: David Baker, University of Washington, Molecular Engineering and Sciences, Box 351655, Seattle, WA. E-mail: dabaker@u.washington.edu

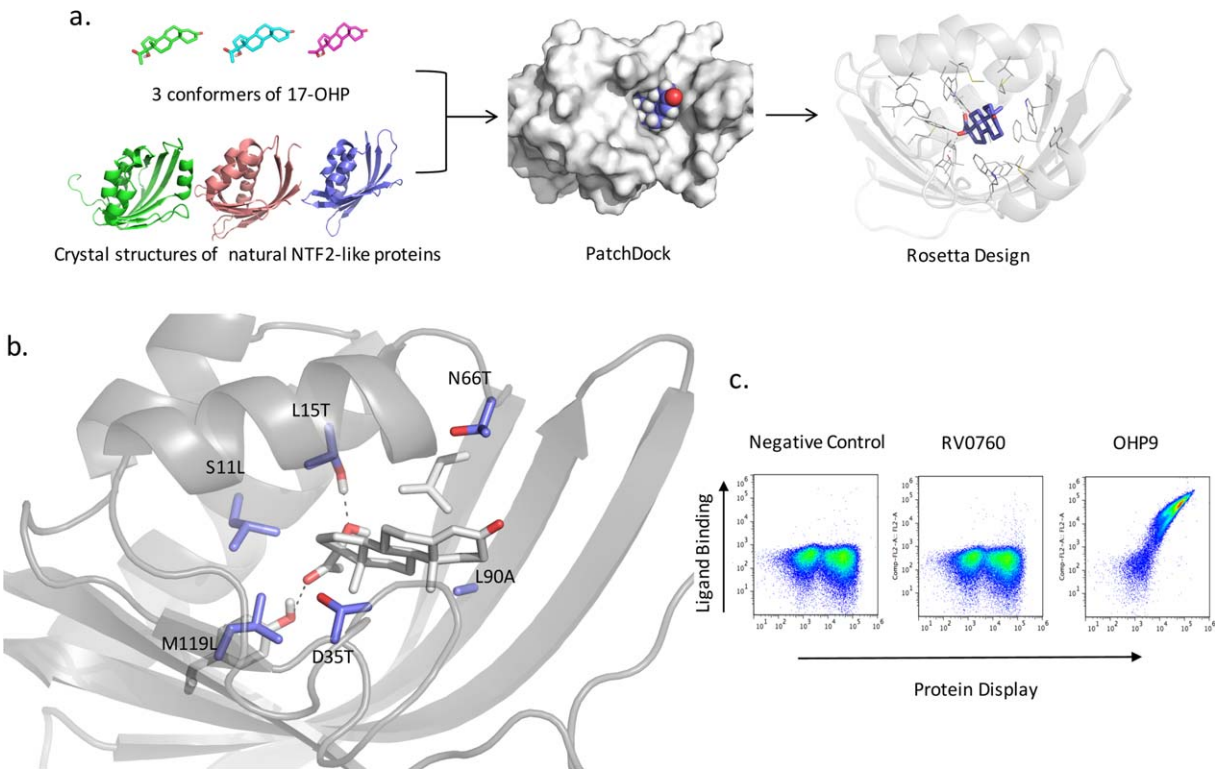

Figure 1. Computational design method and OHP9 design model. (**a**) Computational design method for binding 17-OHP. PatchDock[9] was used to place three conformers of 17-OHP into all 257 crystal structures of NTF2-like proteins from PDB based on shape complementarity; pocket residues were redesigned by Rosetta to fulfill the ligand hydrogen bonding and hydrophobic packing interactions. (**b**) OHP9 design model. Six mutations introduced by Rosetta design are highlighted in purple. 17-OHP is shown as gray sticks with the more polar end inside the protein pocket. (**c**) Yeast display flow cytometry results. Under the same labeling condition, scaffold protein RV0760 does not bind 17-OHP while design protein OHP9 shows a clear binding signal. Negative control is OHP9 cells labeled with streptavidin-PE and FITC-conjugated anti-cMyc antibody without biotinylated 17-OHP compound.

used 17-OHP-based CAH diagnosis methods are radioimmunoassays, enzyme-linked fluoroimmunoassays, gas chromatography mass spectrometry(GC-MS), and liquid chromatography linked with tandem mass spectrometry(LC-MS/MS).[2] There is a need for highly specific 17-OHP screening methods with lower cost and improved sensitivity.

Computational methods for designing small molecule binding proteins seek to generate proteins with a binding pocket with shape and chemical complementarity to the ligand of interest. Rosetta protein design calculations have previously been used to generate high affinity binders of the steroid digoxigenin (DIG).[3] The designed DIG binding domain has subsequently been incorporated into sensors with fluorescence and transcriptional readouts.[4,5] Here, we describe the computational design and experimental characterization of 17α-hydroxylprogesterone binding domains. Unlike the DIG case, the crystal structure of the ligand-protein complex differs considerably from the design model. Investigation of the origin of this discrepancy reveals challenges to designing specific binding pockets for pseudosymmetric ligands with few hydrogen bond donors or acceptors.

Results

17-OHP and OHP9 design

We used the Rosetta computational design program[6] to design proteins to bind 17-OHP using crystal structures of NTF2-like proteins from the RCSB Protein Data Bank (PDB) as starting scaffolds (Table S3). NTF2-like proteins share a common fold that was first observed in the structure of the rat Nuclear Transport Factor 2(NTF2) protein.[7] This fold is well suited for binding small molecules, with a cone-like shape formed by four curved antiparallel beta strands and three or four alpha helices. Our previous work of designing Digoxigenin binding proteins demonstrated that NTF2-like proteins can tolerate multiple mutations introduced by computational design.[3] In our previous work with DIG binding protein design, we used RosettaMatch[8] to place the ligand relative to the scaffold backbone such that amino acids could be placed to make hydrogen bonds with each of the DIG polar groups.[3] As 17-OHP has fewer polar groups, we experimented with approaches in which ligand placement is primarily determined by shape complementarity with the scaffold. We developed protocols that first, identify shape complementary placements of the ligand in

Table I. *Data Collection and Refinement of Crystal Structures of OHP9 and OHP9_1C*

	OHP9	OHP9_1c
PDB ID	5IER	5IF6
Data collection		
Space group	P 3_1 2 1	P 2_1 2_1 2_1
Unit cell		
a, b, c (A°)	78.8, 78.8, 184.5	80.0, 100.8, 118.2
α, β, γ (°)	90, 90, 120	90, 90, 90
Wavelength (Å)	1.0	1.0
Resolution range (Å)	36.2–2.0 (2.1–2.0)	36.7–2.5 (2.6–2.5)
Unique reflections	45362 (4441)	33404 (3054)
R-merge	0.067 (0.369)	0.072 (0.608)
R-meas	0.071 (0.397)	0.077 (0.664)
$CC^1/_2$	(0.933)	(0.988)
I/σ(I)	22.25 (5.35)	26.55 (2.81)
Chî2	1.093 (0.783)	1.162 (0.991)
Multiplicity	10.2 (7.4)	7.1 (6.4)
Completeness (%)	99.95 (99.48)	99.6 (99.5)
Wilson B-factor	29.97	52.5
Refinement		
R-work	0.1777	0.2129
R-free	0.2098	0.2437
Number of non-hydrogen atoms	4292	6029
Macromolecules	3877	5834
Ligands	125	147
Water	290	48
Protein residues	519	800
RMS(bonds)	0.003	0.008
RMS(angles)	0.69	0.86
Ramachandran favored (%)	98	97
Ramachandran allowed (%)	0	2.75
Ramachandran outliers (%)	0	0.25
Clashscore	1.02	1.12
Average B-factor	36.2	60
Macromolecules	35.3	60.2
Ligands	50.3	56.3
Solvent	41.5	51.6

the pocket and second, search for hydrogen bonds with sidechains from neighboring backbone segments [Fig. 1(A)]. For the first step, we used PatchDock,[9] and for the second, either HBnet[10] or Rosetta design allowing small perturbations of the ligand rigid body orientation (Materials and Methods). Sixteen designs with favorable predicted binding energy and high shape complementarity (in comparison with existing steroid binding proteins) were selected for experimental characterization (Tables S1 and S7).

Synthetic genes encoding the 16 designs were obtained and the proteins were displayed on the yeast surface[11] (Methods, Table S2). We synthesized biotin-conjugated and an Alexa488-conjugated probes of 17-OHP for fluorescent labeling. Both probes replace the ketone atom O3 in 17-OHP with a nitrogen atom tethered to an oxime ether linker (Methods, Fig. S1). Eight of the designs showed binding signal in the flow cytometry assay using this biotinylated probe and streptavidin–phycoerythrin (PE); the protein display level in the assay was measured by a fluorescein isothiocyanate (FITC)-conjugated chicken anti-c-Myc antibody (Methods, Fig. S2). Fluorescence polarization assay using the Alexa488-conjugated probe and purified proteins showed the binding affinity was in the low to high micro-molar range (Methods, Fig. S1).

We focused on the OHP9 design, as the crystal structure of this design in complex with 17-OHP provides considerable insight into current design challenges (see below). OHP9 is based on the *Mycobacterium Tuberculosis* protein RV0760, which currently has no annotated biological function.[12] Six substitutions were introduced by Rosetta into this protein scaffold in the design calculations [Fig. 1(B)]. RV0760 does not bind 17-OHP in the yeast display assay [Fig. 1(C)]. OHP9 was expressed and purified from *E. coli*, and its binding dissociation constant (K_D) for 17-OHP was estimated by fluorescence polarization to be approximately 15 μM (Fig. S3).

OHP9 crystal structure

The crystal structure of OHP9 in complex with 17-OHP was solved at 2.0Å resolution (Materials and Methods, Table I). Ligand density is clear and unambiguous in all four copies in the crystallographic asymmetric unit [Fig. 2(D)]. While the protein backbone in the crystal structure is similar to that in the design model (Cα root-mean-square

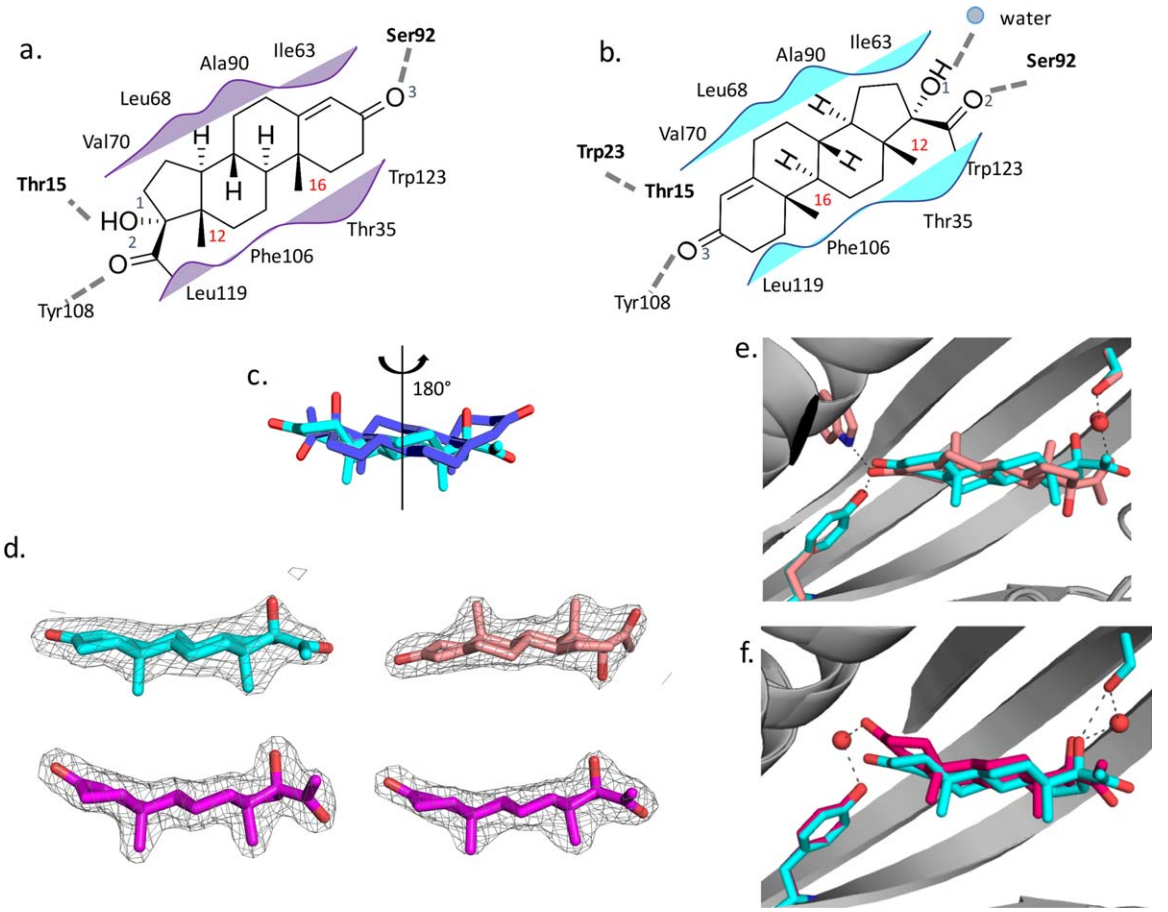

Figure 2. Distinct binding conformations revealed in the crystal structure of OHP9. (**a**) 2D representation of designed interactions around 17-OHP. Hydrogen bonding interactions are highlighted as dashed gray lines between protein sidechains and 17-OHP. Hydrophobic packing interactions are represented as purple shades. (**b**) 2D representation of interactions in OHP9 crystal chain A in the same fashion as in **a**. In addition, a water molecule is represented as a gray dot. Observed intraprotein hydrogen bond between Thr15 and Trp23 is highlighted. Atom O1, O2, O3, C12, and C16 of 17-OHP are labeled explicitly for direct comparison. (**c**) Superimposed ligands upon aligned protein backbones. 17-OHP in design model is shown in purple. The same ligand in crystal chain A is in cyan and its orientation deviates by a 180° rotation. (**d**) 2Fo-Fc density maps of four ligand copies in OHP9 crystal: chain A in cyan; chain B and D in magenta; chain C in salmon. (**e** and **f**) Superimposed ligands upon protein alignment. In comparison with the bound ligand in crystal chain A(cyan), chain C ligand(salmon) reveals the second ligand binding configuration where 17-OHP flips long the longer axis with two methyl groups pointing up; (**f**) Chain B and D(magenta) reveal the third binding configuration with an additional water molecule between Tyr108 and 17-OHP.

deviation (RMSD) < 0.52Å, Table II), there are considerable differences in both the rotameric state of binding site sidechains and the ligand placement [Table S4 and Fig. 2(A–C)]. The four individual copies of the ligand found in the crystallographic asymmetric unit display three distinct configurations that each interact with different sidechain rotameric states and make different water-mediated hydrogen bonds (Fig. 2). The observed presence of several similar, but distinct binding modes within the crystal implies that sufficient energetic and kinetic barriers (and structural differences) separate each state to drive consistent positioning of the distinct modes in the crystallographic asymmetric unit.

In the first configuration (crystal chain A), 17-OHP is flipped 180° in comparison with the design model [Fig. 2(C)]. Thr15, originally designed to form a hydrogen bond to the ligand hydroxyl group (atom name O1) instead turns to form an inter-residue hydrogen bond with Trp23, leaving its methyl group facing the pocket and preferring a hydrophobic interaction [Fig. 2(A)]. Without the Thr15 hydrogen bonding donor to distinguish its two polar ends, 17-OHP's ketone oxygen atom O3 on the other end recapitulates the designed hydrogen bond with Tyr108 [Fig. 2(B)]. Two methyl groups (C12 and C16) switch their positions in the flipped orientation making hydrophobic interactions with Leu119, Phe106, Thr95, and Trp123 [Fig. 2(A, B)]. For all of these residues, except Phe106, the rotameric states are modeled incorrectly in the design model (Table S4). In the second configuration (crystal chain C), 17-OHP is rotated another 180° along the longer axis while maintaining the flipped orientation seen in the first configuration [Fig. 2(E)]. The protein pocket stays almost exactly the same in the first and second configurations except slight

Table II. *Backbone RMSD Between Four Chains of OHP9 Crystal Structure and OHP9 Design Model*

	Chain B (Å)	Chain C (Å)	Chain D (Å)	Design model (Å)
Chain A	0.194	0.165	0.223	0.417
Chain B		0.248	0.225	0.520
Chain C			0.286	0.492
Chain D				0.443

backbone movements on Leu68 and Val102. In the third configuration (crystal chain B and chain D), the indole nitrogen proton of Trp12 faces solvent and forms water-mediated hydrogen bonds with the unpaired beta-strand backbone, leaving an open space in the pocket. Water-mediated hydrogen bonds are formed between 17-OHP and Tyr108. 17-OHP is tilted along the longer axis in comparison with the first configuration [Fig. 2(F)]. Since the probes we used in the binding assays have a PEG linker (Fig. S1), the open space in crystal chain B and chain D could allow the linker to exit (this is speculative since the compound used in the crystal structure determination does not have the linker). In addition to the Trp12 conformation change, Leu68 and Ser92 adopt different rotamers in the third configuration compared with the first and second configuration.

Competition between intra protein and protein–ligand hydrogen bonding

As noted above, a central flaw in the original computational design calculations was the failure to recognize that the side chain of Thr15, initially designed to form a hydrogen bond with the hydroxyl group in 17-OHP, instead can assume an alternate rotameric conformation forming an inter-residue hydrogen bond with Trp23 [Fig. 2(A, B)]. This failure could be due to either a backbone or sidechain sampling problem or to energy function inaccuracy. To address this question, we systematically evaluated the energies of each rotameric state of Trp23 in the absence of 17-OHP, allowing all neighboring sidechains to reconfigure into their lowest energy states for each rotamer choice (Materials and Methods). When the backbone of the crystal structure is used in the calculations, the Trp23-Thr15 interaction is the lowest energy state, as in the crystal structure. However, when the design model backbone was used, the Trp23-Thr15 hydrogen bond is not observed in the low energy ensemble [Fig. 3(A)]. The shortcoming thus arises from the fixed-backbone approximation used in our design protocol: local refinement by short molecular dynamics (MD) simulation was able to sample a subtle change in the N-Cα-Cβ bond angle at position 23 which allows Trp23 to hydrogen bond with Thr15 [Materials and Methods, Fig. 3(B)].

Energetic balance in ligand docking

To investigate whether the unanticipated intra-protein hydrogen bond between Trp23 and Thr15 can explain the lack of observation of the designed ligand orientation in the crystal structure, we performed ligand docking simulations starting with both design model and the crystal structure using Rosetta,[13] Glide,[14] and Vina[15] (we used Glide and Vina in addition to Rosetta to reduce bias arising from the use of the same energy function for design and docking) (Materials and Methods). The docking solutions generated by Glide and Vina with the design model closely matched the designed ligand orientation, suggesting that the designed binding configuration is indeed a deep local minimum given the design model backbone [Fig. 3(C)]. Starting from the crystal structure, Glide correctly generates the flipped orientation with ligand position moved 2.2 Å away from the crystalized configuration; the best model generated by Vina from the crystal structure is still in the (incorrect) designed configuration [Fig. 3(D)]. Rosetta ligand docking with flexible sidechains (Glide and Vina were run with fixed sidechains) samples all three of the ligand configurations observed in the crystal structure, but the lowest energy configuration is that of the design model for both the model backbone and crystal structure backbone [Fig. 3(E)]. In the crystal structure backbone, the lower energy of the incorrect designed ligand configuration comes from both electrostatic and hydrogen bonding interactions (Table III).

OHP9 single-mutation binding landscape

To probe the sequence-structure-function relationships underlying the three different ligand conformations, we performed a single-mutation scanning analysis spanning the entire sequence of OHP9. Every residue in OHP9 was mutated in parallel to each of the 20 amino acids (Methods, Table S5). Yeast cells displaying the mutant pool were incubated with the biotinylated probe, labeled with streptavidin-PE and anti-cMyc FITC conjugated antibody for fluorescence-activated cell sorting (FACS) (Materials and Methods). Naïve and selected pools were deep sequenced and mutation counts were summarized as enrichment values (Methods, Figs. S4 and S5). The hydrogen bonding residues Tyr108 and Thr15 are highly conserved after three rounds of binding selection; major packing residues introduced by computational design also were conserved [Fig. 4(A)]. Overall, the mutational data are compatible with both our design model and the crystal structure. Although the ligand configurations are

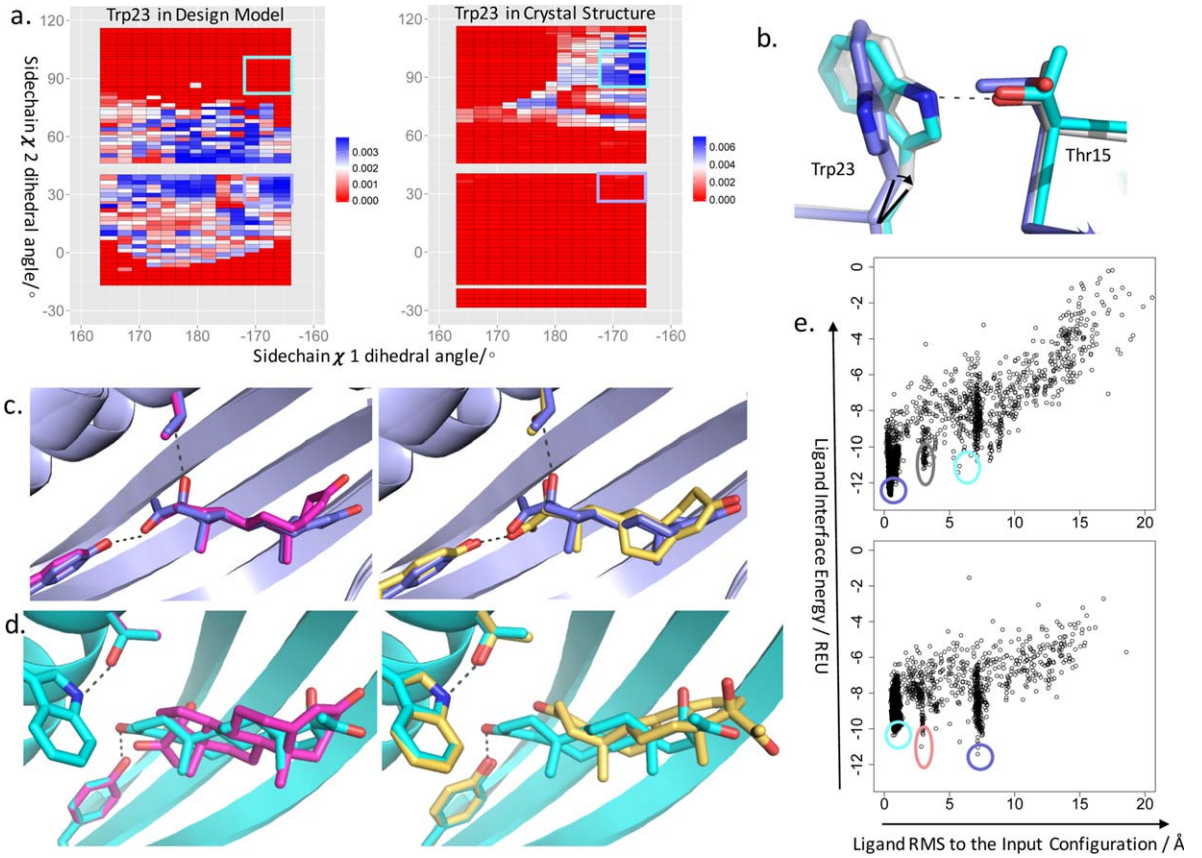

Figure 3. Computational analysis of observed discrepancies in OHP9 design model. (**a**) Computed sidechain rotamer distribution for Trp23 in OHP9 design model backbone and OHP9 crystal backbone. Each colored grid represents one conformational state of Trp23 with sidechain χ1 and χ2 angles indicated by X and Y axes, respectively. Colorimetric scale is based on Boltzmann probability[23] calculated from Rosetta energy term, where blue representing high-probability(low-energy) states and red representing low-probability(high-energy) states. The designed Trp23 rotamer is indicated by a purple window (χ1 ∼ −170°, χ2 ∼ 30°), and crystalized Trp23 rotamer by a cyan window (χ1 ∼ −170°, χ2 ∼ 90°). (**b**) Change of Cα-Cβ vector in molecular dynamics simulation that captures the Trp23-Thr15 hydrogen bond. Design model shown in purple serves as the starting point for simulation; representative MD model after a short simulation shown in gray closely matches the conformations in crystal structure(cyan). (**c** and **d**) Docked ligand conformations using Vina[15] and Glide[14]. OHP9 design model in purple on the upper panel where Vina ligand (pink) and Glide ligand (yellow) are superimposed with the design ligand(purple) for comparison; (**d**) OHP9 crystal chain A in cyan on the lower panel where Vina ligand (pink) and Glide ligand(yellow) overlay with the chain A ligand(cyan). Ligand hydrogen bonds are highlighted by dashed gray lines. (**e**) Ligand energy landscapes generated by Rosetta ligand docking. OHP9 design model was used as the input conformation for the docking simulation summarized on the upper panel, where the purple color circles the design ligand configuration; cyan circle is close to the crystal chain A ligand configuration; gray color circles the ligand configuration that is 180° rotated from chain C ligand (with two polar groups inside the protein pocket); For the lower-panel docking landscape, crystal chain A was used as the input docking conformation where salmon circle represents the crystal chain C ligand configuration. The same colors are used for indicating design(purple) and chain A(cyan) ligand configurations.

quite different, the designed pocket residues play similar structural and chemical roles conferring binding function. Several peripheral mutations are better understood based on the crystal structure. Small amino acids (Gly and Ala) are highly preferred at position 12, where the linker of the biotinylated probe must exit from the binding pocket. The preference of position 13 for Tyr, Trp, and Phe and

Table III. *17-OHP Energy Terms in Designed and Crystalized Configurations (Rosetta Energy Function: Talaris2014; Unit: Rosetta Energy Unit)*

	fa_atr	fa_rep	fa_sol	fa_intra_rep	fa_elec	hbond_sc	Total
Design model	−7.36709	0.55663	2.4725	0.0916	−0.27093	−1.02311	−5.54041
Crystal chain A	−7.28305	0.62063	2.2467	0.06463	−0.13581	−0.54221	−5.02912

fa_atr and fa_rep describe the attractive and repulsive portions of Van der Waals energy, respectively. fa_sol is the implicit solvation energy term based on Lazaridis–Karplus approximation. fa_intra_rep describes the internal repulsive Van der Waals energy of the small molecule ligand 17-OHP. fa_elec is the Coulumb electrostatic energy term. hbond_sc describes the hydrogen bonding energy between protein sidechain and the small molecule ligand.

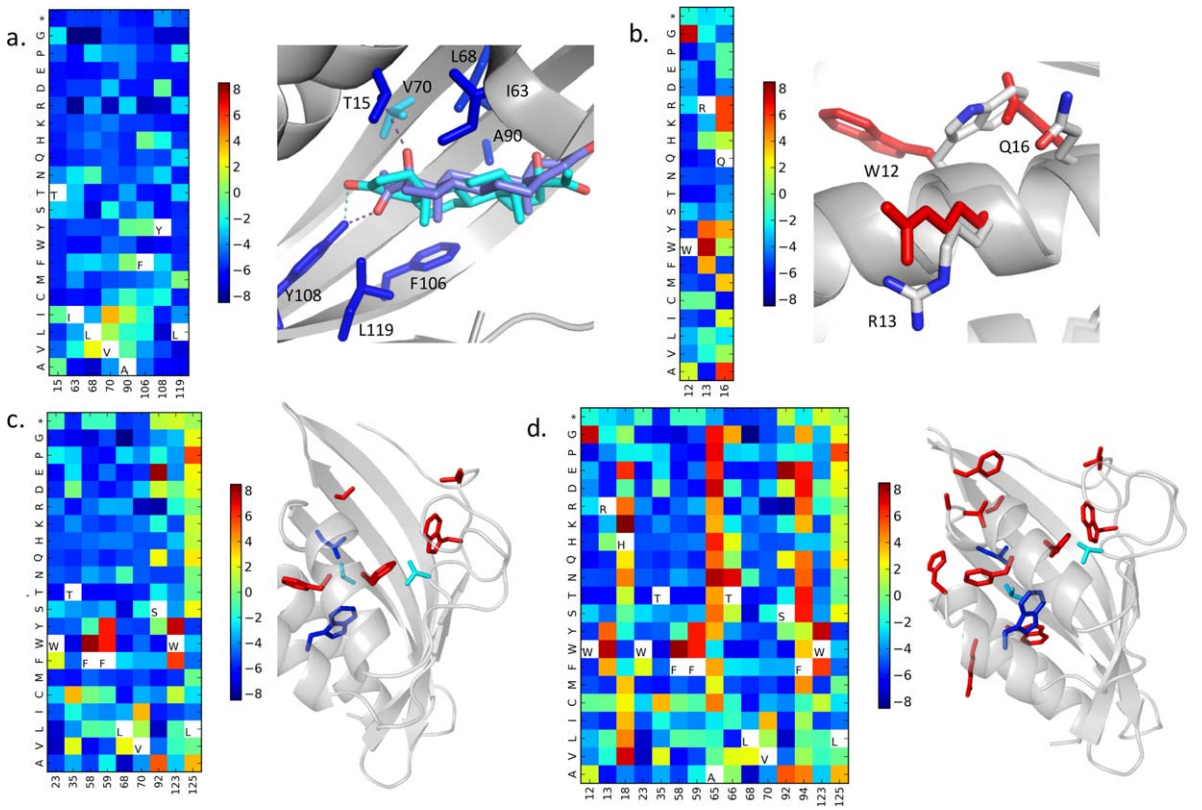

Figure 4. Binding fitness landscape of OHP9. The effect of each amino acid substitution(Y axis) at selected protein positions(X axis) is assessed by enrichment in the binding population (ΔE^x: enrichment value Supplementary Figure S4). Colored grids represent single mutant substitutes, where red and blue indicate high enrichment and depletion, respectively, after three rounds of selection for better binding (Supplementary Figure S4 and S5). The initial OHP9 amino acid at each position is indicated by its one-letter amino acid code in the white box. (**a**) Designed interacting residues in OHP9 are highly conserved during the affinity selection. Few or no substitutions are enriched shown in the colored data matrix. Residue positions are mapped on to the OHP9 structure, where design ligand(purple) and crystal chain A ligand(cyan) are superimposed, and their hydrogen bonds are indicated by dashed lines in the same color. (**b**) Periphery beneficial mutations that seemingly conflict with the design model (red sidechains) can be partially explained by the crystal chain B and D conformation (gray sidechains). (**c**) Beneficial mutations in close vicinity to the ligand were included for constructing a pocket-only combinatorial library of OHP9. The nine positions are mapped onto the OHP9 structure. (**d**) Top enriched substitutions, mostly in the periphery region of the protein, were included for making a general combinatorial library. In total, 15 positions were mutated all around the protein.

favorable mutations seen at position 16 might affect stability of the exposed conformation of Trp12 in chain B and chain D of the crystal structure [Fig. 4(B)].

Using this comprehensive mutational map as a guide, we constructed two combinatorial libraries to explore the synergy between the beneficial mutations found at either the periphery or pocket positions (Methods, Table S6). The libraries were sorted to convergence to identify the highest affinity binders (Fig. S6). OHP9_1A, isolated from the pocket-mutant-only library [Fig. 4(C)], carries five mutations. Of particular note, Trp23 (which appears to prevent Thr15 from forming a hydrogen bond to the ligand) is mutated to Phe in OHP9_1A [Figs. 5(A) and S8]. Its dissociation constant (K_D) for binding 17-OHP was determined (via fluorescent polarization) to be 5.1nM, an estimated 3000-fold affinity increase from OHP9 [Figs. 5(B) and S7]. Attempts to obtain high-resolution diffracting crystals of OHP9_1A were unsuccessful.

OHP9_1C, which displayed a K_D value of 59 nM, was identified from the library designed to combine all of the top beneficial point mutants [Fig. 4(D)]. It carries seven mutations from OHP9, none of which is seen in OHP9_1A [Fig. 5(A)]. The co-crystal structure of OHP9_1C in complex with 17-OHP was solved at 2.5 Å (Materials and Methods, Table I), and the ligand was found in a configuration similar to that in the chain B and D of OHP9 crystal structure, with a water-bridged hydrogen bond between Tyr108 and 17-OHP [Fig. 5(D)]. All six copies of ligand in the crystal structure adopt the same configuration [Fig. 5(C)]. The average backbone Cα RMS change from OHP9 to OHP9_1C is 0.305 Å (Table V). Mutation of Trp12 to Gly likely opens an exit for the probe linker (a sodium ion is present in the empty space in the crystal structure) [Fig. 5(D)]. The Ser92-to-Ala mutation releases a structured water in OHP9. The Trp123-to-Tyr mutation enables another water-bridged hydrogen bond between Tyr123

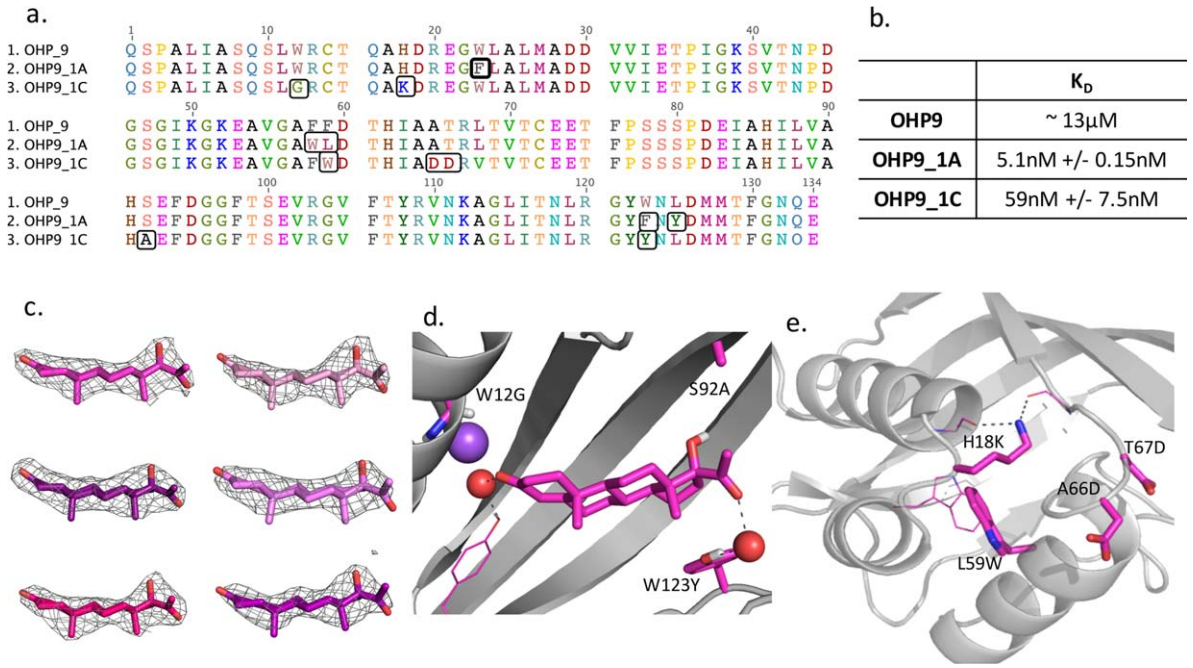

Figure 5. Evolved variants of OHP9: OHP9_1A and OHP9_1C. (**a**) Sequence alignment of OHP9, OHP9_1A and OHP9_1C. Black windows mark the positions mutated in OHP9_1A and OHP9_1C. (**b**) Equilibrium dissociation constants of OHP9, OHP9_1A, and OHP9_1C determined by fluorescence polarization assays. (**c**) 2Fo–Fc density maps of six ligand copies in OHP9_1C crystal. (**d**) The converged ligand binding configuration in OHP9_1C. Mutations inside the binding site are labeled and highlighted by magenta sticks. Water molecules are shown as red spheres and the sodium ion as a purple sphere. (**e**) Periphery mutations in OHP9_1C in magenta sticks mapped onto the crystal structure. Hydrogen bonds are represented by gray dashed lines.

and 17-OHP and helps lock the ligand in the same configuration in all six copies of crystallographic asymmetric unit [Fig. 5(D)]. The other four mutations are located at least 10 Å away from the binding pocket: Lys18 caps the C-terminus of the first helix and connects the helix to the unpaired beta-strand backbone; Asp66 and Asp67 forms water-bridged hydrogen bond to cap the third helix [Fig. 5(E)]. The convergence on a single binding mode and the increase in affinity thus likely result from a combination of backbone changes favoring the selected binding configuration and small adjustments to the ligand binding site.

Discussion

The errors in modeling OHP9 that result in differences in ligand conformation are subtle and highlight the challenges of ligand binding protein design

The discrepancy between the position of the 17-OHP in the design model of OHP9 and the actual ligand positions observed in the crystal structure highlights several of the critical challenges to computational design of small molecule binding proteins. Binding a small molecule with high affinity and specificity depends on the formation of a set of weak non-covalent interactions including hydrogen bonds, electrostatic attractions, and van der Waals attractions. The first challenge is accurate energy evaluation. In the OHP9 case, in the design model the more polar end of 17-OHP is buried such that the polar groups make hydrogen bonds with designed side chain residues. In the crystal structure, the ligand configurations are flipped with the more polar portion sticking out into solvent, suggesting that the cost of desolvating these groups outweighs the energy gain from hydrogen bond formation. The energy function used in design clearly underestimates the desolvation penalty; this is also the case for Vina.

The second challenge is to properly model intra protein sidechain and backbone conformational changes. As noted above, a slight change of the N-Cα-

Table IV. *Backbone RMSD Between Six Chains of OHP9_1C Crystal Structure[TQ1]*

	Chain B (Å)	Chain C (Å)	Chain D (Å)	Chain E (Å)	Chain F (Å)
Chain A	0.174	0.160	0.152	0.149	0.154
Chain B		0.190	0.159	0.196	0.200
Chain C			0.200	0.186	0.194
Chain D				0.177	0.184
Chain E					0.159

Table V. *Backbone RMSD Between OHP9_1C Crystal Structure and OHP9 Crystal Structure*

OHP9_1C	c.A (Å)	c.B (Å)	c.C (Å)	c.D (Å)	c.E (Å)	c.F (Å)	Average (Å)
OHP9_c.A	0.285	0.294	0.248	0.278	0.266	0.245	0.269
OHP9_c.B	0.367	0.334	0.317	0.338	0.312	0.291	0.327
OHP9_c.C	0.331	0.344	0.299	0.294	0.308	0.291	0.311
OHP9_c.D	0.320	0.365	0.277	0.308	0.304	0.302	0.313

Cβ bond angle of residue Trp23 allows it to hydrogen bond to Thr15 instead of interacting with 17-OHP. In addition, several hydrophobic sidechains adopt different rotamers to accommodate the configurations of the ligand observed in the crystal structure. These intra-protein conformational rearrangements disfavor the design target binding mode and favor the alternative observed modes. These challenges are particularly acute when designing binders for hydrophobic small molecules, where there are only one or two polar moieties to make directional interactions and provide "handles" for residues in the binding pocket to grab on to. It is instructive to compare 17-OHP to DIG: the latter has two additional hydrogen bonding groups which the high affinity DIG binding exploits to achieve both specificity and precision. The general problem of specifically orienting non-polar ligand is exacerbated for 17-OHP by its near two-fold symmetry—a flip around the pseudo two-fold axis does not change the steric and non-polar interactions significantly and hence it is challenging for designs to strongly favor one of the two orientations.

There are several clear routes forward. First, the error in energy evaluation of ligand–protein interactions results mainly from the improper balance between the unfavorable cost of desolvating ligand polar groups and the favorable hydrogen bonding interaction. Future work on improving the solvation model will help correct those errors. Second, improved sampling methods coupled with more computing resources and an improved energy function should be able to recapitulate the subtle backbone changes that enable multiple sidechain rearrangement.

Materials and Methods

Computational methods

Small molecule preparation. 3D coordinates of 17α-hydroxylprogesterone (17-OHP) were downloaded from PubChem open chemistry database with CID 6238. The default method in Avogadro[16] was used to assign Gasteiger-Marsili empirical atomic partial charges[17] (Table S8). Three low-energy conformers with all-atom RMSD above 0.2 Å were generated for 17-OHP by OMEGA from OpenEye toolkits.[18]

Scaffold Collection

Crystal structures of NTF2-like proteins from RCSB PDB were collected based on TM alignment to a known NTF2-like protein (PDB ID: 1Z1S) with TM score cutoff of 0.5.[19] Structures with multiple domains were removed afterwards by manual inspection. The bound ligands, water molecules and other crystalized non-protein molecules were removed. 257 crystal structures of NTF2-like proteins were used for design calculation (Table S3). The crystal structures were further prepared using an energy minimization protocol with pre-generated heavy-atom coordinate constraints as previously described.[20]

PatchDock. A PatchDock constraints file that defines the receptor binding pocket was generated for each scaffold. All three 17-OHP conformers were docked into the defined pocket in parallel for all the scaffolds. PatchDock scores were used to rank the docking solutions and top 100 docked configuration from each scaffold were selected for following design calculations. Ligand orientation was calculated based on the distances from the center of mass of the complex to the two ends of 17-OHP (See Supplementary Material). We chose the docked configurations that put the more polar end of 17-OHP buried inside pocket for designing directional hydrogen bonding interactions.

Rosetta design. 7727 docks were continued for Rosetta design. Interface residues were designed using RosettaScripts application.[21] Two RosettaScripts protocols were developed to design hydrogen bonds in the docked configurations. The first protocol used HBnet method to explicitly search for hydrogen bonds. Once a set of hydrogen bonds that satisfy the buried polar atoms of the ligand was found, Rosetta geometric constraint files were used to keep the hydrogen bonding patterns.[22] The other pocket residues were optimized based on the calculated interface energy. The second protocol relied on full-atom Rosetta energy function to capture possible hydrogen bonds while optimizing the binding energy. It did a fine grid searching of ligand configuration and generated 1000 starting ligand positions for each dock. Each ligand configuration went through the pocket design process and filtered by ligand hydrogen bond satisfaction (see Supplementary Material for detailed design protocols).

Energy evaluation of sidechain rotameric states. Rotamer scanning calculation was performed by running *RotamerBoltzmannWeightFilter* within RosettaScripts application.[23] The source code

was modified to report both the $\chi1$ and $\chi2$ dihedral angles of each trial rotamer (See supplementary information).

Structure Refinement Using Molecular Dynamics (MD) Simulation

We used the forcefield and tools from AMBER12 package[24] and simulation parameters that were similar to the method described before.[25] Prepared systems were minimized and equilibrated for 50ps with increasing temperature from 50 to 300 K. Five independent 10-ns trajectories were initialized from the equilibrated state with harmonic restraints of 0.05 kcal/mol/Å^2 applied to the protein Cα coordinates. Langevin dynamics at 300 K was carried out with integration step of 2fs. Single representative model was extracted by structural averaging on all five trajectories using *ptraj* tool. The naïve averaged model was then regularized using Rosetta *Fastrelax* in dual space.[26]

Rosetta ligand docking. Rosetta ligand docking simulation was performed using a previously published protocol.[13] The same ligand parameters and conformers from design calculation were used for Rosetta ligand docking.

Glide and Vina Docking. We submitted both the design model and crystal structure of OHP9 to the weekly Continuous Evaluation for Ligand Pose Prediction (CELPP) for docking prediction (CELPP: www.drugdesigndata.org/about/CELPP Github page: https://github.com/drugdata/d3r/wiki). Ligands for both Vina[15] and Glide[14] protocols were prepared with Schrodingers LigPrep (Schrödinger, LLC, New York, NY, 2017). The best docking models from Vina and Glide were returned to us after the automated simulations were finished.

Experimental materials and methods

Probe synthesis. Reagents and solvents used for the synthesis of the 17-hydroxyprogesterone 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 h 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 (5 μm, 3.9 × 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 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 (Fig. S1).

Yeast display. Synthetic genes with 5′ and 3′ vector-overlapping sequences were synthesized (Gen9) with codon usage optimized for *E.coli* expression. They were first cloned between the NdeI and XhoI sites of pETCON for yeast surface display as an Aga2p-fusion protein.[27] EBY100 yeast cells were treated and induced for protein display according to the published protocol.[11]

Library construction. OHP9 single site-saturation mutagenesis(SSM) library was generated using overlapping PCR method (Table S5).[28] Two combinatorial libraries were generated using oligo assembly methed with synthetic DNA oligos containing degenerate codons (Integrated DNA Technologies) (Table S6). Libraries were transformed as linear PCR product together with linear cut pETCON(digested with NdeI and XhoI) in to EBY100 yeast cells by electroporation.[29]

Flow cytometry binding assay. Yeast cells displaying the designed protein were pre-coated by 0.01% BSA in 100 μL PBSF buffer.[11] In a 1.5 mL Eppendorf tube, pre-coated cells were incubated with 10 μg/mL (0.33 μM) streptavidin–phycoerythrin(PE) (Invitrogen), 1.65 μM biotinylated probes and and 5 μg/mL fluorescein isothiocyanate(FITC)-conjugated chicken anti-c-Myc (Immunology Consultants Laboratory) in 50 μL PBSF for 1 h on a benchtop rotator at room temperature. Cells were spun down and washed twice by 50 μL cold PBSF before running through a Accuri C6 flow cytometer (BD Biosciences) or sorted with a BD Influx cell sorter (BD Biosciences).

Deep sequencing and data process. Yeast cells of OHP9 naive SSM library and sorted libraries were lysed and plasmid DNA was extracted in the same way as described before.[28] Genes were PCR amplified using primers that annealed to the plasmid, followed by a secondary PCR to add 6-bp barcodes for distinguishing different libraries and flanking sequences for annealing to the sequencing flow cell. DNA was sequenced with a Miseq sequencer (Illumina, San Diego, CA) using Illunima Miseq Reagent kit v3 (Catalog#:MS-102–3003). Sequencing cluster density was 862K/mm^2 and 88.9% clusters passing quality filter. Paired-end sequences were aligned by PEAR[30] and analyzed with scripts adapted from Enrich.[31]

Protein samples for binding assay. Functional designs were cloned between the NdeI and XhoI sites of pET29b (Novagen), placing a 6His-tag on the protein's C-terminus. Plasmids were transformed into *E. coli* BL21(DE3) cells for protein expression. Cells were grown in Terrific Broth at 37°C to OD600 ~0.6–0.9 and induced with 0.5 mM IPTG overnight at 18°C. Cells were lysed in phosphate-buffered saline (PBS) (140 mM NaCl, 1 mM KCl, 12 mM Na_2HPO_4, and 1.2 mM KH_2PO_4, pH 7.4) containing 0.5 mM phenylmethylsulfonyl fluoride (PMSF) and 0.05 mg/mL DNase by sonication. Cleared lysate was loaded on NiNTA resin (Qiagen) and washed with 30 column volumes of wash buffer (PBS, 20mM imidazole). Protein were eluted with elution buffer (PBS, 200mM imidazole) and concentrated by centrifugal ultrafiltration before dialyzing overnight at 4°C against PBS. Protein concentration was determined by absorbance at 280 nm using calculated extinction coefficients.

Protein samples for crystallography. An 8His-tag for protein purification and the SUMO protein Smt3 from *Saccharomyces cerevisiae* for tag cleavage were fused to the N terminus of the genes encoding OHP9 and it variants using Gibson Assembly.[32] Purified fusion proteins were cleaved by homemade SUMO Protease (Ubl-specific protease 1) followed by a secondary Ni-NTA affinity resin (Qiagen). Tag-free proteins were further purified by size-exclusion chromatography on an AKTA Pure using Superdex75 column (GE Healthcare).

Crystal structure determination and refinement. Purified proteins were initially tested for crystallization via sparse matrix screens in 96-well sitting drops using a mosquito (TTP LabTech). Crystallization conditions were then optimized with constructs that proved capable of crystallizing in larger 24-well hanging drops. OHP9 crystallized in 100 mM HEPES sodium pH7.5, 1% (w/v) Polyethylene Glycol 200 and 2.0 M Ammonium sulfate at a concentration of 24 mg/mL. The crystal was transferred to a solution containing 75% mother liquor plus 25% ethylene glycol and flash frozen in liquid nitrogen. Data was collected at ALS on BL5.0.2 at 1.0 Å wavelength and processed on HKL2000. OHP9_1c crystallized in 100 mM sodium acetate pH5.5 and 15% (w/v) Polyethylene Glycol 3000 at a concentration of 14.5 mg/mL. The crystal was transferred to a solution containing 75% mother liquor plus 25% ethylene glycol and flash frozen in liquid nitrogen. Data was collected at ALS on BL5.0.2 at 1.0 Å wavelength and processed on HKL2000.[33] OHP9 and OHP9_1c were solved by Molecular Replacement with Phaser via phenix[34,35] using the original scaffold 2A15. The structure was then built and refined using Coot[36] and phenix, [35] respectively, until finished.

Acknowledgments

We thank Dr. Shuai Liu, Jeffrey R. Wagner, and Dr. Vitoria Feher from *The Drug Design Data Resource* (D3R) at University of California, San Diego for generating the docked files using Vina and Glide. We thank Dr. Christine E. Tinberg for her comments on the manuscript. This work was supported by the Defense Threat Reduction Agency (DTRA). The computational work was facilitated though the use of advanced computational, storage, and networking infrastructure provided by the Hyak supercomputer system at the University of Washington. Some of the computation was done using resources provided by the Open Science Grid, which is supported by the National Science Foundation and the U.S. Department of Energy's Office of Science. P.J.G was supported by Carlsbergfondet and an EMBO Long-term postdoctoral fellowship.

References

1. Speiser PW, White PC (2003) Congenital adrenal hyperplasia. N Engl J Med 349:776–788.
2. White PC (2009) Neonatal screening for congenital adrenal hyperplasia. Nat Rev Endocrinol 5:490–498.
3. 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 501:212–216.
4. 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 10:598–603.
5. Feng J, Jester BW, Tinberg CE, Mandell DJ, Antunes MS, Chari R, Morey KJ, Rios X, Medford JI, Church GM, Fields S, Baker D, (2015) A general strategy to construct small molecule biosensors in eukaryotes. eLife 4:e10606.
6. Leaver-Fay A, Tyka M, Lewis SM, Lange F, Thompson J, Jacak R, Kaufman K, Renfrew PD, Smith CA, Sheffler W, Davis IW, Cooper S, Treuille A, Mandell DJ, Richter F, Ban Y-EA, Fleishman SJ, Corn JE, Kim DE, Lyskov S, Berrondo M, Mentzer S, Popovic Z, Havranek JJ, Karanicolas J, Das R, Meiler J, Kortemme T, Gray JJ, Kuhlman B, Baker D, Bradley P (2011) ROSETTA3: An object-oriented software suite for the simulation and design of macromolecules. Methods Enzymol 487:545–574.
7. Bullock TL, Clarkson WD, Kent HM, Stewart M (1996) The 1.6 angstroms resolution crystal structure of nuclear transport factor 2 (NTF2). J Mol Biol 260:422–431.
8. Zanghellini A, Jiang L, Wollacott A (2009) New algorithms and an in silico benchmark for computational enzyme design. Protein Sci 15:2785–2794.
9. Schneidman-Duhovny D, Inbar Y, Nussinov R, Wolfson HJ (2002) PatchDock and SymmDock: servers for rigid and symmetric docking. Nucl. Acids. Res 33:363–367, 2005.
10. Boyken SE, Chen Z, Groves B, Langan RA, Oberdorfer G, Ford A, Gilmore JM, Xu C, Dimaio F, Pereira JH, Sankaran B, Seelig G, Zwart PH, Baker D (2016) De novo design of protein homo-oligomers with modular hydrogen-bond network-mediated specificity. Science 352:680–687.

11. 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: 755–768.
12. Cherney MM, Garen CR, James MNG (2008) Crystal structure of *Mycobacterium tuberculosis* Rv0760c at 1.50A resolution, a structural homolog of delta5-3-ketosteroid isomerase. Biochim Biophys Acta 1784: 1625–1632.
13. Combs SA, Deluca SL, Deluca SH, Lemmon GH, Nannemann DP, Nguyen ED, Willis JR, Sheehan JH, Meiler J (2013) Small-molecule ligand docking into comparative models with Rosetta. Nat Protoc 8:1277–1298.
14. Friesner RA, Murphy RB, Repasky MP, Frye LL, Greenwood JR, Halgren TA, Sanschagrin PC, Mainz DT (2006) Extra precision glide: Docking and scoring incorporating a model of hydrophobic enclosure for protein-ligand complexes. J Med Chem 49:6177–6196.
15. Trott O, Olson A (2010) AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading. J Comput Chem 31:455–461.
16. Hanwell MD, Curtis DE, Lonie DC, Vandermeersch T, Zurek E, Hutchison GR (2012) Avogadro: an advanced semantic chemical editor, visualization, and analysis platform. J Cheminform 4:17.
17. Gasteiger J, Marsili M (1980) Iterative partial equalization of orbital electronegativity: A rapid access to atomic charges. Tetrahedron 36:3219–3228.
18. Hawkins PCD, Skillman AG, Warren GL, Ellingson BA, Stahl MT (2010) Conformer generation with OMEGA: Algorithm and validation using high quality structures from the Protein Databank and Cambridge Structural Database. J Chem Inf Model 50:572–584.
19. Zhang Y, Skolnick J (2005) TM-align: A protein structure alignment algorithm based on the TM-score. Nucleic Acids Res 33:2302–2309.
20. Nivón LG, Moretti R, Baker D (2013) A pareto-optimal refinement method for protein design scaffolds. PLoS One 8:1–5.
21. Fleishman SJ, Leaver-Fay A, Corn JE, Strauch E-M, Khare SD, Koga N, Ashworth J, Murphy P, Richter F, Lemmon G, Meiler J, Baker D (2011) RosettaScripts: a scripting language interface to the Rosetta macromolecular modeling suite. PLoS One 6:e20161.
22. Richter F, Leaver-Fay A, Khare SD, Bjelic S, Baker D (2011) De novo enzyme design using Rosetta3. PLoS One 6:e19230.
23. Fleishman SJ, Khare SD, Koga N, Baker D (2011) Restricted sidechain plasticity in the structures of native proteins and complexes. Protein Sci 20:753–757.
24. Case DA, Darden TA, Cheatham TEI, Simmerling CL, Wang J, Duke RE, Luo R, Walker RC, Zhang W, Merz KM, Roberts B, Hayik S, Roitberg A, Seabra G, Swails J, Gotz AW, Kolossvary I, Wong KF, Paesani F, Vanicek J, Wolf RM, Liu J, Wu X, Brozell SR, Steinbrecher T, Gohlke H, Cai Q, Ye X, Wang J, Hsieh W-J, Cui G, Roe DR, Mathews DH, Seetin MG, Salomon-Ferrer R, Sagui C, Babin V, Luchko T, Gusarov S, Kovalenko A, Kollman PA (2012) AMBER 12, University of California, San Francisco.
25. Mirjalili V, Feig M (2013) Protein structure refinement through structure selection and averaging from molecular dynamics ensembles. J Chem Theory Comput 9: 1294–1303.
26. Conway P, Tyka MD, DiMaio F, Konerding DE, Baker D (2014) Relaxation of backbone bond geometry improves protein energy landscape modeling. Protein Sci 23:47–55.
27. Fleishman SJ, Whitehead TA, Ekiert DC, Dreyfus C, Corn JE, Strauch E-M, Wilson IA, Baker D (2011) Computational design of proteins targeting the conserved stem region of influenza hemagglutinin. Science 332:816–821.
28. Procko E, Hedman R, Hamilton K, Seetharaman J, Fleishman SJ, Su M, Aramini J, Kornhaber G, Hunt JF, Tong L, Montelione GT, Baker D (2013) Computational design of a protein-based enzyme inhibitor. J Mol Biol 425:3563–2575.
29. 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 23:155–159.
30. Zhang J, Kobert K, Flouri T, Stamatakis A (2014) PEAR: a fast and accurate Illumina Paired-End reAd mergeR. Bioinformatics 30:614–620.
31. Fowler DM, Araya CL, Gerard W, Fields S (2011) Enrich: software for analysis of protein function by enrichment and depletion of variants. Bioinformatics 27:3430–3431.
32. Gibson DG, Young L, Chuang R-Y, Venter JC, Hutchison C. a, Smith HO (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat Methods 6:343–345.
33. Otwinowski Z, Minor W (1997) Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol 276:307–326.
34. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ (2007) Phaser crystallographic software. J Appl Cryst 40:658–674.
35. Adams PD, Afonine PV, Bunkóczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung L-W, Kapral GJ, Grosse-Kunstleve RW, McCoy AJ, Moriarty NW, Oeffner R, Read RJ, Richardson DC, Richardson JS, Terwilliger TC, Zwart PH (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Cryst D66:213–221.
36. Emsley P, Lohkamp B, Scott WG, Cowtan K (2010) Features and development of Coot. Acta Cryst D66: 486–501.