LETTER

Computational design of ligand–binding proteins with high affinity and selectivity

Christine E. Tinberg[1]*, Sagar D. Khare[1]†*, Jiayi Dou[2,3], Lindsey Doyle[4], Jorgen W. Nelson[5], Alberto Schena[6], Wojciech Jankowski[7], Charalampos G. Kalodimos[7], Kai Johnsson[6], Barry L. Stoddard[4] & David Baker[1,8]

The ability to design proteins with high affinity and selectivity for any given small molecule is a rigorous test of our understanding of the physiochemical principles that govern molecular recognition. Attempts to rationally design ligand-binding proteins have met with little success, however, and the computational design of protein–small-molecule interfaces remains an unsolved problem[1]. Current approaches for designing ligand-binding proteins for medical[2] and biotechnological uses rely on raising antibodies against a target antigen in immunized animals[3,4] and/or performing laboratory-directed evolution of proteins with an existing low affinity for the desired ligand[5–7], neither of which allows complete control over the interactions involved in binding. Here we describe a general computational method for designing pre-organized and shape complementary small-molecule-binding sites, and use it to generate protein binders to the steroid digoxigenin (DIG). Of seventeen experimentally characterized designs, two bind DIG; the model of the higher affinity binder has the most energetically favourable and pre-organized interface in the design set. A comprehensive binding-fitness landscape of this design, generated by library selections and deep sequencing, was used to optimize its binding affinity to a picomolar level, and X-ray co-crystal structures of two variants show atomic-level agreement with the corresponding computational models. The optimized binder is selective for DIG over the related steroids digitoxigenin, progesterone and β-oestradiol, and this steroid binding preference can be reprogrammed by manipulation of explicitly designed hydrogen-bonding interactions. The computational design method presented here should enable the development of a new generation of biosensors, therapeutics and diagnostics.

Computational design could provide a general approach for creating new small molecule binding proteins with rationally programmed specificities. Structural and biophysical characterization of previous computationally designed ligand-binding proteins revealed numerous discrepancies with the design models, however, and it was concluded that protein–ligand interaction design is an unsolved problem[1,8]. The lack of accuracy in programming protein–small-molecule interactions also contributes to low catalytic efficiencies of computationally designed enzymes[9–14]. The development of robust computational methods for the design of small-molecule-binding proteins with high affinity and selectivity would have wide-ranging applications.

We developed a computational method for designing ligand-binding proteins with three properties characteristic of naturally occurring binding sites: (1) specific energetically favourable hydrogen-bonding and van der Waals interactions with the ligand; (2) high overall shape complementarity to the ligand; and (3) structural pre-organization in the unbound protein state, which minimizes entropy loss upon ligand binding[15,16]. To program in defined interactions with the small molecule,
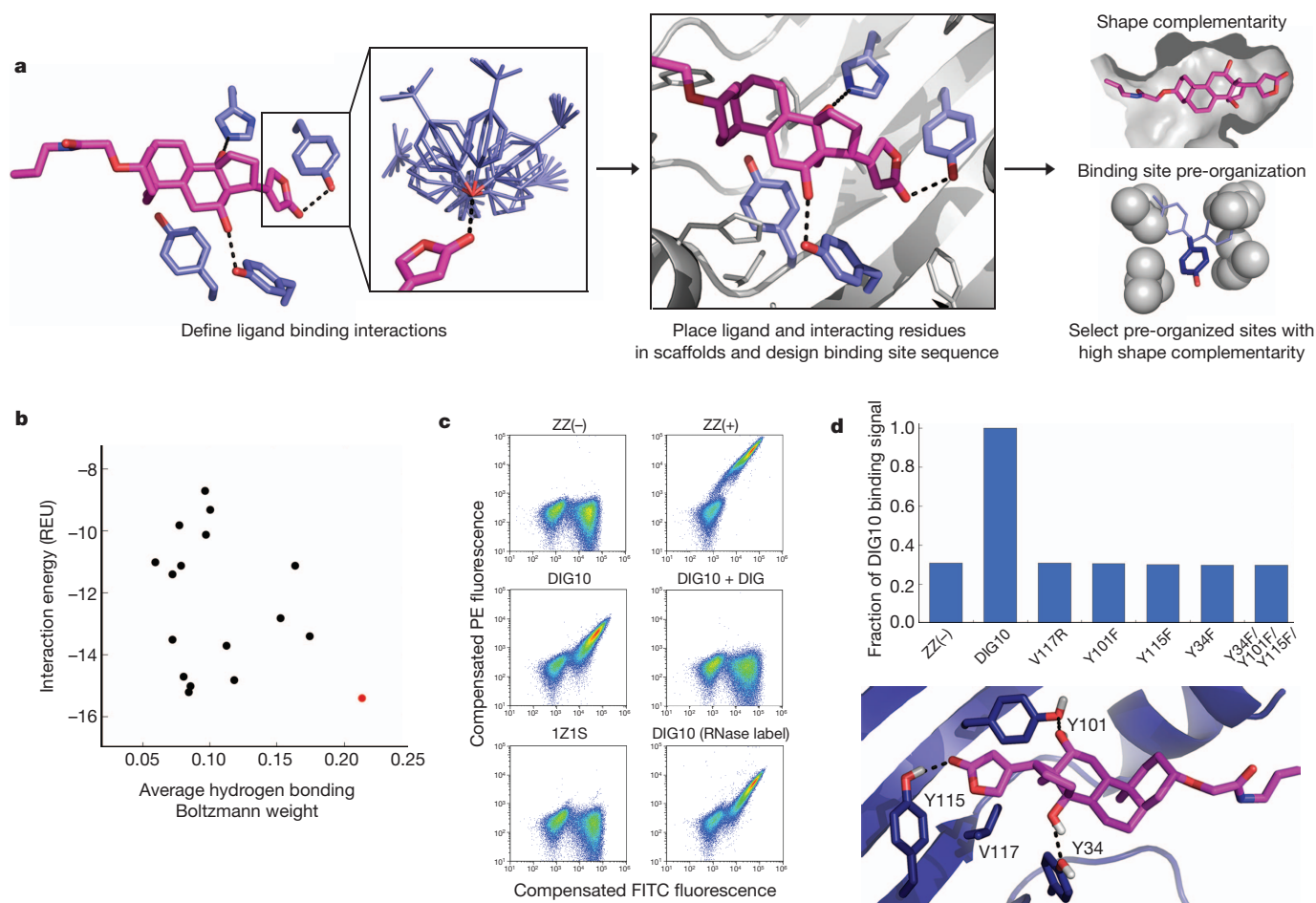
disembodied binding sites are created by positioning amino acid side chains around the ligand in optimal orientations and then placed at geometrically compatible sites in a set of scaffold protein structures[17]. The surrounding side chain identities and conformations are then optimized to generate additional protein–ligand and buttressing protein–protein interactions (Fig. 1a). Designs with protein–small-molecule shape complementarity below those typical of native complexes[18] or having interface side chain conformations with low Boltzmann-weighted probabilities in the unbound state[16] are then discarded.

We used the method to design proteins that bind the steroid DIG (Supplementary Fig. 1), the aglycone of digoxin, a cardiac glycoside used to treat heart disease[19] and a non-radioactive biomolecular labelling reagent[20]. Anti-digoxigenin antibodies are administered to treat overdoses of digoxin, which has a narrow therapeutic window[21], and are used to detect biomolecules in applications such as fluorescence *in situ* hybridization[20]. We created idealized DIG-binding sites featuring hydrogen bonds from Tyr or His to the polar groups of DIG and hydrophobic packing interactions between Tyr, Phe or Trp and the steroid ring system (Fig. 1a). These interactions were embedded in designed binding sites with high shape complementarity to DIG, and 17 designs were selected for experimental characterization based on computed binding affinity, shape complementarity, and the extent of binding site pre-organization in the unbound state (Fig. 1b and Supplementary Tables 1 and 2).

Binding of the designed proteins to DIG was probed by yeast surface display[22] and flow cytometry using DIG-functionalized bovine serum albumin (DIG-BSA) or RNase (DIG-RNase). Designed proteins DIG5 and DIG10 bound to both labels (Fig. 1c and Supplementary Fig. 2), and binding was reduced to background levels when unlabelled DIG was added as a competitor (Fig. 1c and Supplementary Fig. 3). Fluorescence polarization measurements with purified proteins and Alexa488-fluorophore-conjugated DIG (DIG-PEG$_3$-Alexa488) indicated affinities in the low-to-mid micromolar range, with DIG10 binding more tightly (Fig. 2a, b). Isothermal titration calorimetry (ITC) measurements confirmed that the affinity of DIG10 for DIG is identical to that for DIG-PEG$_3$-Alexa488 (Fig. 2b, Supplementary Fig. 4 and Supplementary Table 3). The scaffold from which both DIG5 and DIG10 derive, a protein of unknown function from *Pseudomonas aeruginosa* (Protein Data Bank (PDB) accession code 1Z1S), does not bind to either label (Fig. 1c and Supplementary Fig. 3a) when expressed on the yeast surface or to DIG-PEG$_3$-Alexa488 in solution (Fig. 2a), suggesting that the binding activities of both proteins are mediated by the computationally designed interfaces. Indeed, substitution of small nonpolar residues in the binding pockets of DIG5 and DIG10 with arginines resulted in complete loss of binding, and mutation of the designed hydrogen-bonding tyrosine and histidine residues to

Figure 1 | Computational design methodology and experimental binding validation. a, Overview of the design procedure. **b**, Ranking of experimentally characterized DIG designs by computed ligand interaction energy (Rosetta energy units, REU) and average (geometric mean) side-chain Boltzmann weight of residues designed to hydrogen bond to DIG. DIG10 (red) scores the best by both metrics. **c**, Flow cytometric analysis of yeast cells expressing designed proteins. Yeast surface expression and DIG binding were probed by labelling with anti-c-Myc-fluorescein (FITC) and a mixture of biotinylated DIG-functionalized BSA and phycoerythrin (PE)–streptavidin, respectively. ZZ(−), negative control; ZZ(+), positive control; 1Z1S, original scaffold. **d**, On-yeast substitutions of DIG10-designed interface residues reduce binding (phycoerythrin) signals to background negative control levels. See figure legends in Supplementary Information for details.

phenylalanine reduced (DIG5) or abolished (DIG10) binding (Fig. 1d and Supplementary Fig. 5). Optimization of DIG10 by site-saturation mutagenesis and selections using yeast surface display and fluorescence-activated cell sorting (FACS) identified several small-to-large hydrophobic amino acid changes that increase binding affinity 75-fold through enhanced binding enthalpy, yielding DIG10.1 (Fig. 2b, c, f, Supplementary Figs 4, 6–8 and Supplementary Table 3).

To provide feedback for improving the overall design methodology and to evaluate the contribution of each residue in the DIG10.1-binding site, we used next-generation sequencing to generate a comprehensive binding fitness map[23–25]. A library of variants with approximately 1–3 substitutions at 39 designed interface positions in DIG10.1 was generated using doped oligonucleotide mutagenesis, displayed on yeast, and subjected to selections using a monovalent DIG-PEG$_3$-biotin conjugate (Supplementary Fig. 9). Variants with increased affinity for DIG were selected by FACS, and next-generation sequencing was used to quantify the frequency of every single point mutation in the unselected and selected populations. A large majority of the interrogated variants were depleted in the selected population relative to the unselected input library, suggesting that most of the DIG10.1-binding site residues are optimal for binding (Fig. 2d, e and Supplementary Fig. 10). In particular, mutation of the three designed hydrogen-bonding residues, Tyr 34, Tyr 101 and Tyr 115, to any other amino acid was disfavoured. Several large hydrophobic residues that pack against the ligand in the computational model

are also functionally optimal (for example, Phe 66 and Phe 119). Besides Ala 99, which contacts DIG directly, most of the observed mutations that improve binding are located in the second coordination shell of the ligand and fall into two categories: (1) protein core substitutions tolerating mutation to chemically similar amino acids (for example, Leu 105 and Cys 23), and (2) solvent-exposed loop residues having high sequence entropy (for example, His 90 and Val 92). The best clone obtained from sorting the library to homogeneity, DIG10.2, contains two of the most highly enriched mutations, Ala37Pro and His41Tyr (Fig. 2b, c and Supplementary Figs 4, 6, 8 and 11).

To increase binding affinity further, we constructed a library in which the residues at 11 positions that acquired beneficial substitutions in the deep sequencing experiment were varied in combination to allow for non-additive effects. Selections led to DIG10.3 (Supplementary Figs 4, 6, 8 and 12), which binds DIG and its cardiac glycoside derivative digoxin with picomolar affinity (Fig. 2b, Supplementary Fig. 13 and Supplementary Table 4), rivalling the affinities of anti-digoxin antibody therapeutics[21] and an evolved single-chain variable anti-digoxin antibody fragment[7]. Fluorescence-polarization-based affinity measurements of DIG10.3 and Tyr knockouts suggest that the designed hydrogen bonds each contribute ~2 kcal mol^{-1} to binding energy (Supplementary Table 5 and Supplementary Fig. 8).

The crystal structures of DIG10.2 and DIG10.3 in complex with DIG were solved to 2.05 Å and 3.2 Å resolution, respectively (Fig. 3a, b and
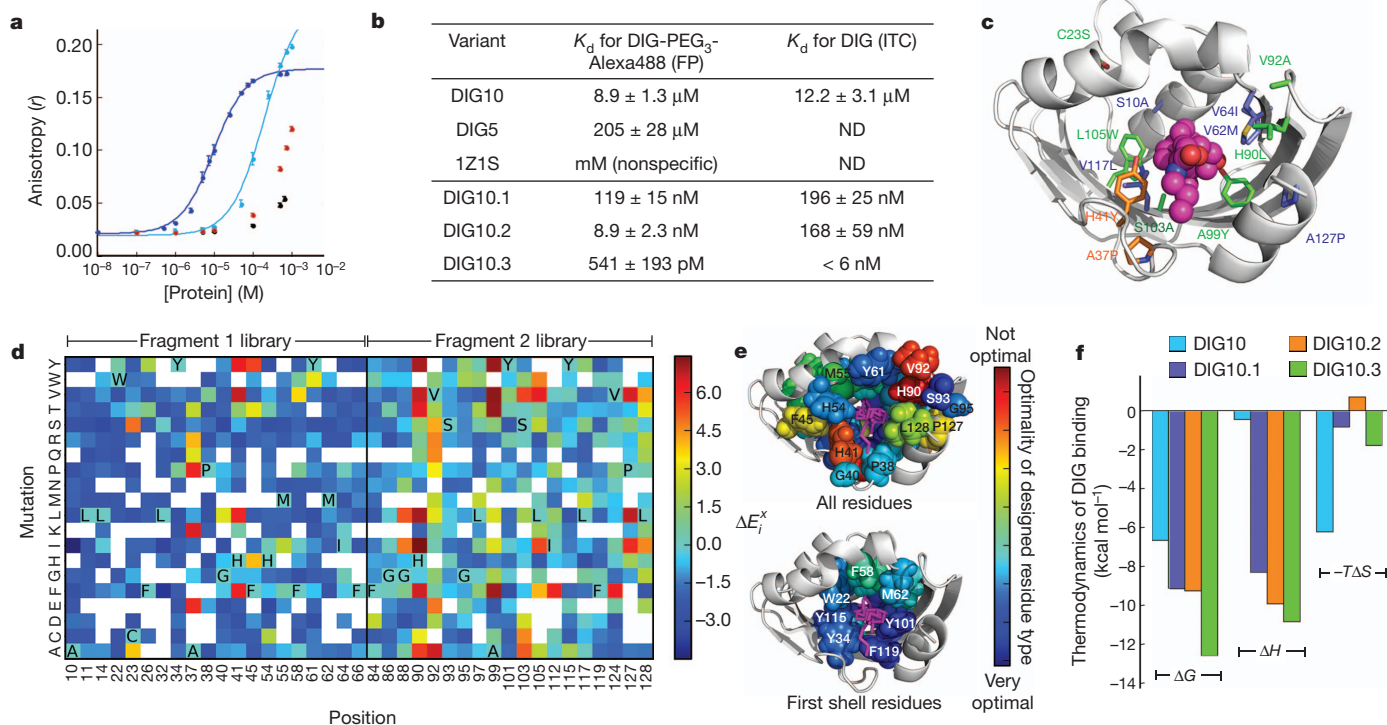
Figure 2 | Binding characterization and affinity maturation. a, Equilibrium fluorescence anisotropy of DIG-PEG$_3$-Alexa488 mixed with purified DIG10 (blue), DIG5 (cyan), 1Z1S scaffold (black) and BSA (red). Error bars represent s.d. for at least three independent measurements. **b**, Dissociation constants (K_d values) of designs. Differences in fluorescence polarization (FP)- and isothermal titration calorimetry (ITC)-derived K_d values probably result from enhanced interactions with the linker of DIG-PEG$_3$-Alexa488 used in the fluorescence polarization experiments. ND, not determined. **c**, Mutations identified during affinity maturation to generate DIG10.1 (blue), DIG10.2 (orange) and DIG10.3 (green) mapped onto the computational model of DIG10.3. **d**, Fitness landscape of DIG10.1 showing the effects of single amino acid substitutions on binding (ΔE_i^x; see equation (1) in Methods). Red and blue indicate enrichment and depletion, respectively. The original DIG10.1 amino acid at each position is indicated in bold. White indicates mutations for which there were not enough sequences in the unselected library to make a statistically significant conclusion about function. **e**, The optimality of each initial DIG10.1 residue type mapped onto the computational model of DIG10.1. **f**, DIG binding thermodynamic parameters determined by ITC. ΔG, free binding energy; ΔH, binding enthalpy; $-T\Delta S$, binding entropy. See figure legends in Supplementary Information for details.

Supplementary Figs 14–19). The structure of DIG10.2 bound to DIG shows atomic-level agreement with the design model (all-atom root mean squared deviation (r.m.s.d.) = 0.54 Å; Fig. 3c). The ligand–protein interface has high shape complementarity (S_c = 0.66) and no water molecules are observed in the binding pocket. The DIG binding mode is nearly identical in the X-ray structure and the computational model, with an average r.m.s.d. of 0.99 Å for all ligand heavy atoms (Fig. 3d). As designed, Tyr 34, Tyr 101 and Tyr 115 hydrogen bond with O3, O2 and O1 of DIG, respectively. Tyr 41, a residue identified during affinity maturation, forms a weak hydrogen bond with the terminal hydroxyl group of DIG (O5) (Supplementary Fig. 16). Of 27 non-glycine/alanine residues within ~10 Å of the ligand, 21 adopt the computationally designed conformations (Supplementary Fig. 17), including Tyr 101 and Tyr 115 (in chain B) as well as the first-shell packing residues Trp 22, Phe 58 and Phe 119. The structure of DIG10.3 bound to DIG (Supplementary Fig. 18) also agrees closely with the design model (r.m.s.d. = 0.68 Å).

We assessed the binding specificity of DIG10.3 by determining affinities for a series of related steroids by equilibrium competition fluorescence polarization. Digitoxigenin, progesterone and β-oestradiol bind less tightly to DIG10.3 than DIG (Fig. 4a, b and Supplementary Table 4). The magnitudes of the affinity decreases are consistent with the loss of one, two and three hydrogen bonds, respectively (assuming ~1.8 kcal mol^{-1} per hydrogen bond[26]), suggesting that these compounds bind in the same orientation as DIG. We next investigated whether the observed steroid selectivity could be reprogrammed by mutagenesis of the key hydrogen-bonding tyrosines. The variants Tyr101Phe, Tyr34Phe and Tyr34Phe/Tyr99Phe/Tyr101Phe show clear preferences for more hydrophobic

steroids in a predictable manner that depends on the hydrogen-bonding capabilities of both the protein and the steroid. Tyr101Phe eliminates the DIG-specific hydrogen bond with DIG O2 and provides a more hydrophobic environment that favours the other three steroids (Fig. 4c). Tyr34Phe removes a hydrogen bond common to DIG and digitoxigenin, thus enhancing the preference for progesterone (Fig. 4d). Tyr34Phe/Tyr99Phe/Tyr101Phe has decreased affinity for DIG and increased affinity for the more hydrophobic steroids (Fig. 4e). These results confirm that the selectivity of DIG10.3 for DIG is conferred through the designed hydrogen-bonding interactions and demonstrate how this feature can be programmed using positive design alone through the explicit placement of designed polar and hydrophobic interactions.

Comparison of the properties of successful and unsuccessful designs provides a test of the hypotheses underlying the design methodology. Although all 17 designed proteins had high computed shape complementarity to DIG by construction, the DIG10 design, which had the highest affinity for DIG, had the most favourable computed protein–ligand interaction energy and was predicted to have the most pre-organized binding site (Fig. 1b and Supplementary Table 6), suggesting that these attributes should continue to be the focus of future design methodology development. One potential avenue for obtaining more favourable interaction energies would be incorporating backbone flexibility during design to achieve more tightly packed binding sites: the fact that substitution of small hydrophobic interface residues to larger ones increased binding affinity indicates that the original DIG10 design was under-packed.

The binding fitness landscape in combination with the X-ray co-crystal structures highlight the importance of second shell interactions
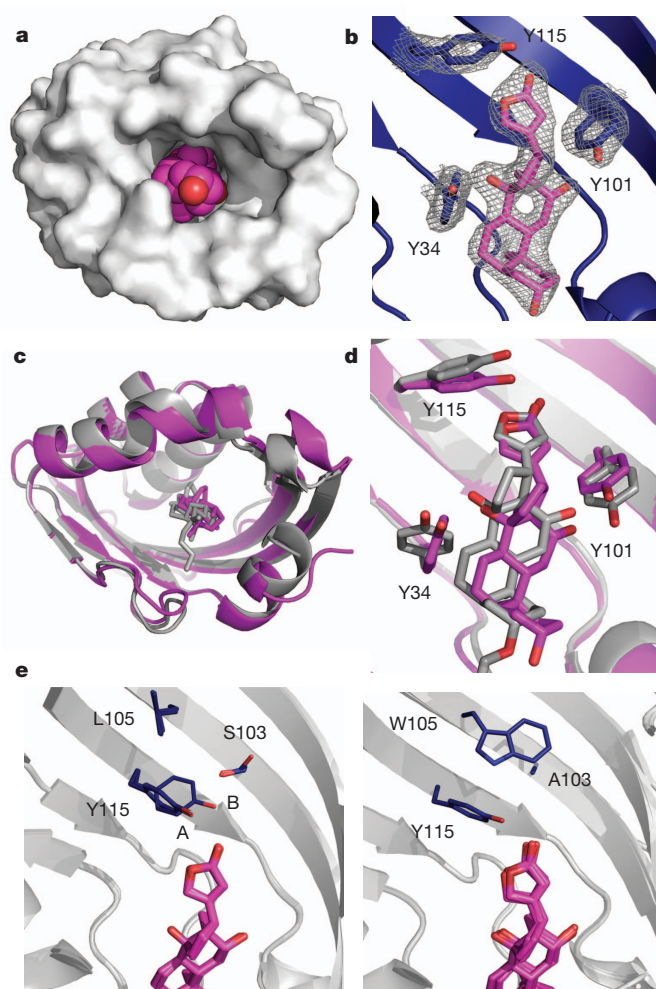
Figure 3 | Crystal structures of DIG10.2–DIG and DIG10.3–DIG. a, Surface representation of the DIG10.2−DIG complex. DIG is shown in magenta. DIG10.2 is a dimer and crystallized with four copies in the asymmetric unit. **b**, DIG10.2−DIG $2F_o − F_c$ omit map electron density showing DIG and interacting tyrosines contoured at 1.0σ. **c**, Backbone superposition of the crystal structure of DIG10.2−DIG (magenta) with the computational model (grey). **d**, DIG10.2−DIG-binding site backbone superposition. **e**, Comparison of an overlay of the four DIG10.2–DIG crystallographic copies (left) with that of DIG10.3–DIG chains A, B, C, H and I (right; the density was poorly resolved for the other four chains). DIG10.2 Tyr 115 conformation A has a more canonical hydrogen-bonding geometry than conformation B, and in DIG10.3 Tyr 115 is locked into conformation A by Trp 105.

in stabilizing binding competent conformations. For example, the enriched substitution Leu105Trp (Fig. 2d) causes the adjacent Tyr 115 side chain, which shows conformational variability in DIG10.2, to adopt a single conformation in DIG10.3 that makes a more canonical hydrogen bond to DIG than that of the DIG10.2 alternative state (Fig. 3e and Supplementary Fig. 19). The calculated side chain pre-organization of all three hydrogen-bonding tyrosine residues increases from DIG10.2 to DIG10.3 (Supplementary Table 7), suggesting that the increased affinity may arise in part from a higher proportion of the binding competent conformation of apo-DIG10.3 (refs 15, 27). Indeed, ITC studies confirm that entropic as well as enthalpic factors contribute to the enhanced binding affinity of DIG10.3 (Fig. 2f and Supplementary Table 3). Similarly, reduced backbone conformational entropy is probably responsible for the increased fitness of substitutions increasing β-sheet propensity at inter-strand loop positions 90 and 92 (Fig. 2d). That flexibility is selected against during affinity maturation suggests that maximizing the free-energy gap between binding-competent and alternative states of the binding site[28] by explicitly designing second shell interactions to buttress side
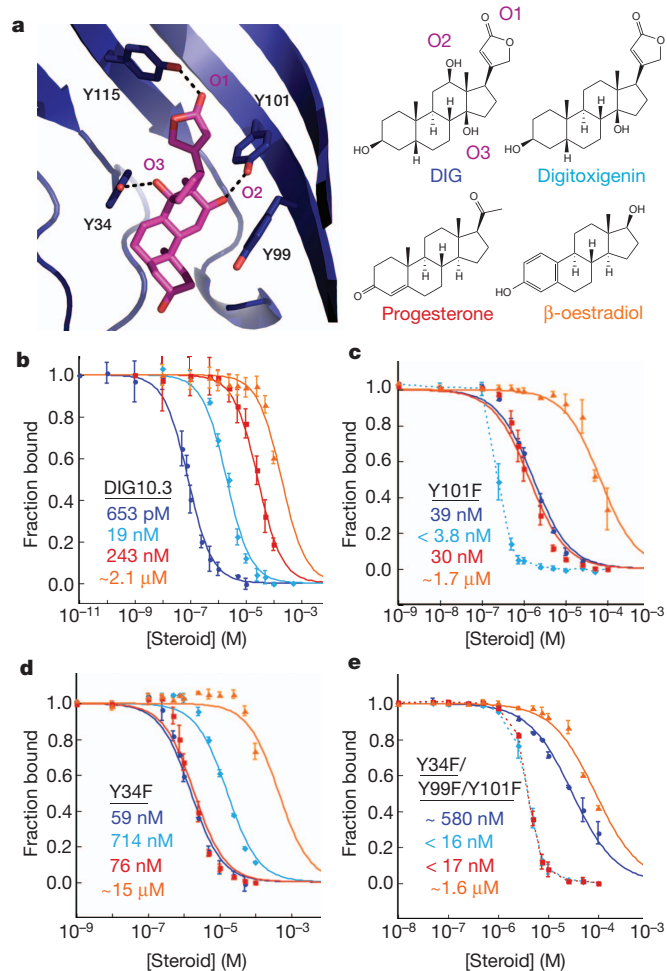
Figure 4 | Steroid binding selectivity. a, X-ray crystal structure of the DIG10.3−DIG complex (left) and chemical structures of related steroids (right). **b**, Steroid selectivity profile of DIG10.3 determined by measuring the equilibrium effects of unlabelled steroids on the anisotropy of the DIG-PEG$_3$-Alexa488−DIG10.3 complex. Solid lines represent fits of the data to a competitive binding model and error bars indicate s.d. for at least three independent measurements. Inhibition constants (K_i values, coloured numbers) were estimated from inhibitory half-maximal concentrations (IC$_{50}$ values) obtained from the fits. **c**, Steroid selectivity profile of DIG10.3(Tyr101Phe). Dashed lines show qualitative assessments of the inhibitory effects for cases in which the data could not be fit due to overly strong inhibition (see Supplementary Methods). **d**, Steroid selectivity profile of DIG10.3(Tyr34Phe). **e**, Steroid selectivity profile of DIG10.3(Tyr34Phe/Tyr99Phe/Tyr101Phe).

chains making key ligand contacts should help to achieve high affinity in the next generation of computationally designed ligand-binding proteins.

The binding affinity of DIG10.3 is similar to those of anti-digoxin antibodies[21], and because it is stable for extended periods (>3 months) at ambient temperatures (Supplementary Fig. 20) and can be expressed at high levels in bacteria, it could provide a more cost-effective alternative for biotechnological and for therapeutic purposes if it can be made compatible with the human immune response. With continued improvement in the methodology and feedback from experimental results, computational protein design should provide an increasingly powerful approach to creating small molecule receptors for synthetic biology, therapeutic scavengers for toxic compounds, and robust binding domains for diagnostic devices.

METHODS SUMMARY

Design calculations were performed using RosettaMatch[17] to incorporate five pre-defined interactions to DIG into a set of 401 scaffolds. RosettaDesign[29] was then

used to optimize each binding site sequence for maximal ligand-binding affinity. Designs having low interface energy, high shape complementarity, and high binding site pre-organization were selected for experimental characterization. Example command lines and full design protocols are given in the Supplementary Data.

Designs were displayed on the surface of yeast strain EBY100 and examined for binding to a mixture of 2.7 μM biotinylated DIG-conjugated BSA or DIG-conjugated RNase and streptavidin–phycoerythrin on an Accuri C6 flow cytometer. Binding clones from yeast-surface displayed libraries based on DIG10 were selected using highly avid DIG-BSA or DIG-RNase or monovalent DIG-conjugated biotin on a Cytopeia inFlux cell sorter. DIG10.1-derived library DNA was sequenced in paired-end mode on an Illumina MiSeq. For single mutations having ≥ 7 counts in the original input library, a relative enrichment ratio between the input library and each selected library was calculated[23–25]. The effect of each amino acid substitution on binding, ΔE_i^x, was computed with equation (1),

$$\Delta E_i^x = \log_2\left(\frac{f_i^{x,\text{sel}}}{f_i^{x,\text{unsel}}}\right) - \log_2\left(\frac{f_i^{\text{orig,sel}}}{f_i^{\text{orig,unsel}}}\right) \qquad (1)$$

in which $f_i^{x,\text{sel}}$ is the frequency of mutation x at position i in the selected population, $f_i^{x,\text{unsel}}$ is the frequency in the unselected population, $f_i^{\text{orig,sel}}$ is the frequency of the original amino acid at position i in the selected population, and $f_i^{\text{orig,unsel}}$ is the frequency of the original amino acid in the unselected population.

For biochemical assays, proteins were expressed in *E. coli* Rosetta 2 (DE3) cells with a carboxy-terminal tobacco etch virus (TEV) protease-cleavable His6 tag. For crystallographic analysis of DIG10 variants, a 12-amino-acid structurally disordered C terminus deriving from the scaffold protein 1Z1S was replaced directly with a His6 tag. Binding affinities were determined by equilibrium fluorescence polarization[30] on a SpectraMax M5e microplate reader by monitoring the anisotropy of DIG-conjugated Alexa488 as a function of protein concentration. Equilibrium fluorescence polarization competition assays were performed by examining the effect of increasing concentrations of unlabelled DIG, digitoxigenin, progesterone and β-oestradiol on the anisotropy of designed protein−DIG-conjugated Alexa488 complex. ITC studies were performed on an iTC200 microcalorimeter.

Full Methods and any associated references are available in the online version of the paper.

1. Schreier, B., Stumpp, C., Wiesner, S. & Höcker, B. Computational design of ligand binding is not a solved problem. *Proc. Natl Acad. Sci. USA* **106,** 18491–18496 (2009).
2. de Wolf, F. A. & Brett, G. M. Ligand-binding proteins: their potential for application in systems for controlled delivery and uptake of ligands. *Pharmacol. Rev.* **52,** 207–236 (2000).
3. Hunter, M. M., Margolies, M. N., Ju, A. & Haber, E. High-affinity monoclonal antibodies to the cardiac glycoside, digoxin. *J. Immunol.* **129,** 1165–1172 (1982).
4. Shen, X. Y., Orson, F. M. & Kosten, T. R. Vaccines against drug abuse. *Clin. Pharmacol. Ther.* **91,** 60–70 (2012).
5. Bradbury, A. R. M., Sidhu, S., Dübel, S. & McCafferty, J. Beyond natural antibodies: the power of *in vitro* display technologies. *Nature Biotechnol.* **29,** 245–254 (2011).
6. Brustad, E. M. & Arnold, F. H. Optimizing non-natural protein function with directed evolution. *Curr. Opin. Chem. Biol.* **15,** 201–210 (2011).
7. Chen, G. *et al.* Isolation of high-affinity ligand-binding proteins by periplasmic expression with cytometric screening (PECS). *Nature Biotechnol.* **19,** 537–542 (2001).
8. Telmer, P. G. & Shilton, B. H. Structural studies of an engineered zinc biosensor reveal an unanticipated mode of zinc binding. *J. Mol. Biol.* **354,** 829–840 (2005).
9. Baker, D. An exciting but challenging road ahead for computational enzyme design. *Protein Sci.* **19,** 1817–1819 (2010).
10. Jiang, L. *et al. De novo* computational design of retro-Aldol enzymes. *Science* **319,** 1387–1391 (2008).
11. Khare, S. D. & Fleishman, S. J. Emerging themes in the computational design of novel enzymes and protein–protein interfaces. *FEBS Lett.* **587,** 1147–1154 (2013).
12. Khersonsky, O. *et al.* Bridging the gaps in design methodologies by evolutionary optimization of the stability and proficiency of designed Kemp eliminase KE59. *Proc. Natl Acad. Sci. USA* **109,** 10358–10363 (2012).
13. Röthlisberger, D. *et al.* Kemp elimination catalysts by computational enzyme design. *Nature* **453,** 190–195 (2008).
14. Wang, L. *et al.* Structural analyses of covalent enzyme–substrate analog complexes reveal strengths and limitations of *de novo* enzyme design. *J. Mol. Biol.* **415,** 615–625 (2012).
15. Boehr, D. D., Nussinov, R. & Wright, P. E. The role of dynamic conformational ensembles in biomolecular recognition. *Nature Chem. Biol.* **5,** 789–796 (2009).
16. Fleishman, S. J., Khare, S. D., Koga, N. & Baker, D. Restricted sidechain plasticity in the structures of native proteins and complexes. *Protein Sci.* **20,** 753–757 (2011).
17. Zanghellini, A. *et al.* New algorithms and an in silico benchmark for computational enzyme design. *Protein Sci.* **15,** 2785–2794 (2006).
18. Lawrence, M. C. & Colman, P. M. Shape complementarity at protein/protein interfaces. *J. Mol. Biol.* **234,** 946–950 (1993).
19. The Digitalis Investigation Group. The effect of digoxin on mortality and morbidity in patients with heart failure. *N. Engl. J. Med.* **336,** 525–533 (1997).
20. Eisel, D., Seth, O., Grünewald-Janho, S. & Kruchen, B. *DIG Application Manual for Nonradioactive in situ Hybridization* 4th edn (Roche Diagnostics, 2008).
21. Flanagan, R. J. & Jones, A. L. Fab antibody fragments: some applications in clinical toxicology. *Drug Saf.* **27,** 1115–1133 (2004).
22. Chao, G. *et al.* Isolating and engineering human antibodies using yeast surface display. *Nature Protocols* **1,** 755–768 (2006).
23. Fowler, D. M. *et al.* High-resolution mapping of protein sequence-function relationships. *Nature Methods* **7,** 741–746 (2010).
24. McLaughlin, R. N. Jr, Poelwijk, F. J., Raman, A., Gosal, W. S. & Ranganathan, R. The spatial architecture of protein function and adaptation. *Nature* **491,** 138–142 (2012).
25. Whitehead, T. A. *et al.* Optimization of affinity, specificity and function of designed influenza inhibitors using deep sequencing. *Nature Biotechnol.* **30,** 543–548 (2012).
26. Fersht, A. R. *et al.* Hydrogen bonding and biological specificity analysed by protein engineering. *Nature* **314,** 235–238 (1985).
27. Frederick, K. K., Marlow, M. S., Valentine, K. G. & Wand, A. J. Conformational entropy in molecular recognition by proteins. *Nature* **448,** 325–329 (2007).
28. Fleishman, S. J. & Baker, D. Role of the biomolecular energy gap in protein design, structure, and evolution. *Cell* **149,** 262–273 (2012).
29. Kuhlman, B. & Baker, D. Native protein sequences are close to optimal for their structures. *Proc. Natl Acad. Sci. USA* **97,** 10383–10388 (2000).
30. Rossi, A. M. & Taylor, C. W. Analysis of protein-ligand interactions by fluorescence polarization. *Nature Protocols* **6,** 365–387 (2011).

Author Contributions C.E.T., S.D.K. and D.B. designed the research. S.D.K., C.E.T. and J.D. performed computations. S.D.K. wrote program code. C.E.T. experimentally characterized the designs, constructed libraries, performed affinity maturation and deep sequencing selections, and conducted binding and biochemical studies of DIG10. J.D. characterized DIG5. J.W.N. analysed deep sequencing data. C.E.T. and J.D. prepared protein samples for crystallographic analysis. L.D. and B.S. crystallized the protein samples and analysed crystallographic data. A.S. and K.J. synthesized DIG-PEG3-biotin and DIG-PEG3-Alexa488. C.E.T. prepared protein samples for ITC studies, and W.J. and C.G.K. performed ITC experiments and analysed ITC data. C.E.T., S.D.K. and D.B. analysed data and wrote the manuscript. All authors discussed the results and commented on the manuscript.

METHODS

General methods. Full details for all computational and experimental methods are given in Supplementary Methods. Design calculations were performed using RosettaDesign[29]. Dissociation constants (K_d values) of designs were determined by equilibrium fluorescence polarization[30] and by ITC. Example command lines and RosettaScripts[31] design protocols are provided in Supplementary Data. Source code is freely available to academic users through the Rosetta Commons agreement (http://www.rosettacommons.org/). Design models, the scaffold library, and scripts for running design calculations are provided on the Baker laboratory website.

Matching. RosettaMatch[17] was used to identify backbone constellations in 401 protein scaffold structures where a DIG molecule and side chain conformations interacting with DIG in a pre-defined geometry could be accommodated. This set contained scaffolds previously used for design projects within our laboratory[10,13,32], as well as structural homologues of a subset of these scaffolds that are known to tolerate mutations. Full details are given in Supplementary Methods.

Rosetta sequence design. Two successive rounds of sequence design were used. The purpose of the first was to maximize binding affinity for the ligand[33]. The goal of the second was to minimize protein destabilization due to aggressive scaffold mutagenesis while maintaining the binding interface designed during the first round. During the latter round, ligand–protein interactions were up-weighted by a factor of 1.5 relative to intra-protein interactions to ensure that binding energy was preserved. Two different criteria were used to minimize protein destabilization: (1) native scaffold residues identities were favoured by 1.5 REU, and (2) no more than five residues were allowed to change from residue types observed in a multiple sequence alignment (MSA) of the scaffold if (a) these residues were present in the MSA with a frequency greater than 0.6, or (b) if the calculated $\Delta\Delta G$ for mutation of the scaffold residue to alanine[34] was greater than 1.5 REU in the context of the scaffold sequence. In some design calculations, identities of the matched hydrogen-bonding residues were allowed to vary according to the MSA and $\Delta\Delta G$ criteria described above. In these cases, designs having fewer than three hydrogen bonds between the protein and the ligand were discarded.

Design evaluation. Designs were evaluated on interface energy, ligand solvent exposed surface area, ligand orientation, shape complementarity, and apo-protein binding site pre-organization. The latter was enforced by explicitly introducing second-shell amino acids that bolster the programmed interacting residues using Foldit[35] and by selecting designs having rotamer Boltzmann probabilities[16] > 0.1 for at least one hydrogen-bonding residue (Supplementary Table 6). High shape complementarity was enforced by rejecting designs having $S_c < 0.5$. Shape complimentary was computed using the *CCP*4 package v.6.0.2 (ref. 36) using the S_c program[18] and the Rosetta radii library. All designs were evaluated for local sequence secondary structure compatibility, and those predicted to have backbone conformations that varied by > 0.8 Å from their native scaffold were rejected (see Supplementary Methods).

General experimental methods. Detailed procedures for the syntheses of DIG-BSA-biotin, DIG-RNase-biotin, DIG-PEG₃-biotin, and DIG-PEG₃-Alexa488, as well as protein expression, purification, crystallization, cloning and mutagenesis methods are given in Supplementary Methods. Details about fluorescence polarization binding assays, ITC, gel filtration analysis, analytical ultracentrifugation experiments, and circular dichroism protein stability measurements are also provided in Supplementary Methods.

Yeast surface display. Designed proteins were tested for binding using yeast-surface display[22]. Yeast surface protein expression was monitored by binding of anti-c-Myc-FITC to the C-terminal Myc epitope tag of the displayed protein. DIG binding was assessed by quantifying the phycoerythrin (PE) fluorescence of the displaying yeast population following incubation with DIG-BSA-biotin, DIG-RNase-biotin, or DIG-PEG₃-biotin, and streptavidin-phycoerythrin (SAPE). In a typical experiment using DIG-BSA-biotin or DIG-RNase-biotin, cells were resuspended in a premixed solution of PBSF (PBS plus 1 g L⁻¹ of BSA) containing a 1:100 dilution of anti-c-Myc-FITC, 2.66 µM DIG-BSA-biotin or DIG-RNase-biotin, and 664 nM SAPE for 2–4 h at 4 °C. Cellular fluorescence was monitored on an Accuri C6 flow cytometer using a 488-nm laser for excitation and a 575-nm bandpass filter for emission. Phycoerythrin fluorescence was compensated to minimize bleed-over contributions from the FITC channel. Competition assays with free DIG were performed as above except that between 750 µM and 1.5 mM DIG was added to each labelling reaction mixture. Full details are given in Supplementary Methods.

Affinity maturation. Detailed procedures for constructing and selecting all libraries, including those for deep sequencing, are provided in Supplementary Methods. Yeast surface display library selections were conducted on a Cytopeia inFlux cell sorter using increasingly stringent conditions. In all labelling reactions for selections, care was taken to maintain at least a tenfold molar excess of label to cell surface protein. Cell surface protein molarity was estimated by assuming that

an attenuance at 600 nm ($D_{600\,nm}$) of $1.0 = 1 \times 10^7$ cells ml⁻¹, and that each cell displays 50,000 copies of protein[22]. For each round of sorting, we sorted at least 10 times the theoretical library size. FlowJo software v. 7.6 was used to analyse all data. Cell sorting parameters and statistics for all selections are given in Supplementary Table 9.

Next-generation sequencing. Two sequencing libraries based on DIG10.1 were assembled by recursive PCR: an amino-terminal library (fragment 1 library) and a carboxy-terminal library (fragment 2 library). To introduce mutations, we used degenerate PAGE-purified oligonucleotides in which the bases coding for 39 selected binding site amino acid residues were doped with a small amount of each non-native base at a level expected to yield 1–2 mutations per gene (TriLink BioTechnologies) (Supplementary Table 10). Yeast cells were transformed with DNA insert and restriction-digested pETCON[37]. Surface protein expression was induced[22] and cells were labelled with anti-c-Myc-FITC and sorted for protein expression. Expressing cells were recovered, induced, labelled with 100 nM of DIG-PEG₃-biotin for >3 h at 4 °C and then SAPE and anti-c-Myc-FITC for 8 min at 4 °C, and then sorted. For each library, clones having binding signals higher than that of DIG10.1 were collected (Supplementary Fig. 9). To reduce noise from the first round of cell sorting, the sorted libraries were recovered, induced, and subjected to a second round of sorting using the same conditions (see Supplementary Methods).

Library DNA was prepared as described[25]. Illumina adaptor sequences and unique library barcodes were appended to each library pool by PCR amplification using population-specific primers (Supplementary Table 11). DNA was sequenced in paired-end mode on an Illumina MiSeq using a 300-cycle reagent kit and custom primers (see Supplementary Methods). Of a total of 5,630,105 paired-end reads, 2,531,653 reads were mapped to library barcodes (Supplementary Table 12). For each library, paired-end reads were fused and filtered for quality (Phred ≥ 30). The resulting full-length reads were aligned against DIG10.1 using Enrich[38]. For single mutations having ≥7 counts in the original input library, a relative enrichment ratio between the input library and each selected library was calculated[23–25]. The effect of each amino acid substitution on binding, ΔE_i^x, was computed with equation (1),

$$\Delta E_i^x = \log_2 \left(\frac{f_i^{x,\text{sel}}}{f_i^{x,\text{unsel}}} \right) - \log_2 \left(\frac{f_i^{\text{orig,sel}}}{f_i^{\text{orig,unsel}}} \right) \qquad (1)$$

in which $f_i^{x,\text{sel}}$ is the frequency of mutation x at position i in the selected population, $f_i^{x,\text{unsel}}$ is the frequency in the unselected population, $f_i^{\text{orig,sel}}$ is the frequency of the original amino acid at position i in the selected population, and $f_i^{\text{orig,unsel}}$ is the frequency of the original amino acid in the unselected population.

Fluorescence polarization binding assays. Fluorescence-polarization-based affinity measurements of designs and their evolved variants were performed as described[30] using Alexa488-conjugated DIG (DIG-PEG₃-Alexa488). Fluorescence anisotropy (r) was measured in 96-well plate format on a SpectraMax M5e microplate reader (Molecular Devices) with $\lambda_{ex} = 485$ nM and $\lambda_{em} = 538$ nM using a 515-nm emission cut-off filter. Fluorescence polarization equilibrium competition binding assays were used to determine the binding affinities of DIG10.3 and its variants for unlabelled DIG, digitoxigenin, progesterone, β-oestradiol and digoxin. The inhibition constant (K_i) for each protein–ligand interaction was calculated from the measured total unlabelled ligand producing 50% binding signal inhibition (I_{50}; see Supplementary Methods) and the K_d of the protein-label interaction according to a model accounting for receptor-depletion conditions[30]. Full details are provided in Supplementary Methods.

ITC. ITC studies were performed on an iTC200 microcalorimeter (MicroCal) at 25 °C in PBS, pH 7.4. Ligand solutions were prepared by diluting a stock solution of DIG (100 mM in 100% dimethylsulphoxide (DMSO)) into the flow-through of the last buffer aliquot used to exchange the protein (final DMSO concentrations were 1–3%). ITC titration data were integrated and analysed with Origin 7.0 (MicroCal). Full details are provided in Supplementary Methods.

31. Fleishman, S. J. *et al.* RosettaScripts: A scripting language interface to the Rosetta macromolecular modeling suite. *PLoS ONE* **6,** e20161 (2011).
32. Siegel, J. B. *et al.* Computational design of an enzyme catalyst for a stereoselective bimolecular Diels–Alder reaction. *Science* **329,** 309–313 (2010).
33. Richter, F., Leaver-Fay, A., Khare, S. D., Bjelic, S. & Baker, D. *De novo* enzyme design using Rosetta3. *PLoS ONE* **6,** e19230 (2011).
34. Kellogg, E. H., Leaver-Fay, A. & Baker, D. Role of conformational sampling in computing mutation-induced changes in protein structure and stability. *Proteins* **79,** 830–838 (2011).
35. Cooper, S. *et al.* Predicting protein structures with a multiplayer online game. *Nature* **466,** 756–760 (2010).
36. Collaborative Computational Project, Number 4. The *CCP*4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50,** 760–763 (1994).

37. Benatuil, L., Perez, J. M., Belk, J. & Hsieh, C.-M. An improved yeast transformation method for the generation of very large human antibody libraries. *Protein Eng. Des. Sel.* **23,** 155–159 (2010).

38. Fowler, D. M., Araya, C. L., Gerard, W. & Fields, S. Enrich: software for analysis of protein function by enrichment and depletion of variants. *Bioinformatics* **27,** 3430–3431 (2011).