PROTEIN DESIGN

Computational design of a modular protein sense-response system

Anum A. Glasgow[1]*, Yao-Ming Huang[1]*†, Daniel J. Mandell[1,2]*‡, Michael Thompson[1], Ryan Ritterson[1]§, Amanda L. Loshbaugh[1,3], Jenna Pellegrino[1,3], Cody Krivacic[1,4], Roland A. Pache[1]‖, Kyle A. Barlow[1,2]¶, Noah Ollikainen[1,2], Deborah Jeon[1], Mark J. S. Kelly[5], James S. Fraser[1,3,6], Tanja Kortemme[1,2,3,4,6,7]#

Sensing and responding to signals is a fundamental ability of living systems, but despite substantial progress in the computational design of new protein structures, there is no general approach for engineering arbitrary new protein sensors. Here, we describe a generalizable computational strategy for designing sensor-actuator proteins by building binding sites de novo into heterodimeric protein-protein interfaces and coupling ligand sensing to modular actuation through split reporters. Using this approach, we designed protein sensors that respond to farnesyl pyrophosphate, a metabolic intermediate in the production of valuable compounds. The sensors are functional in vitro and in cells, and the crystal structure of the engineered binding site closely matches the design model. Our computational design strategy opens broad avenues to link biological outputs to new signals.

In the past two decades, computational protein design has created diverse new protein structures spanning helical (*1*–*5*), alpha-beta (*6*–*8*), and beta-sheet (*9*, *10*) folds. By contrast, our ability to computationally design arbitrary protein functions de novo lags far behind, with relatively few examples that often require screening of many variants (*11*, *12*). One unsolved challenge is the de novo design of small-molecule sensor-actuators in which ligand binding by a protein directly controls changes in downstream functions, a key aspect of cellular signal transduction (*13*).

Sensing and responding to a small-molecule signal requires both recognition of the target and linking target recognition to an output response. Exciting progress has been made with the design of proteins recognizing new ligands (*10*, *11*, *14*–*16*). A general solution to the second problem, coupling ligand recognition to diverse output responses, has remained challenging. Existing approaches have used a ligand that fluoresces upon binding (*10*), engineered the sensor components to be unstable and hence inactive in the absence of the ligand (*14*, *17*), or repurposed an allosteric transcription factor (*18*). Each of these strategies places constraints on the input signals or output responses that can be used.

Here, we describe a computational strategy to engineer protein complexes that can sense a small molecule and respond directly using different biological outputs, creating modular sensor-actuator systems. Unlike previous work (*10*, *11*, *14*, *15*) that reengineered existing binding sites or placed ligands into preformed cavities, we build small-molecule binding sites de novo into heterodimeric protein-protein interfaces to create new and programmable chemically induced dimerization systems (CIDs). This strategy is inspired by naturally occurring and reengineered CID systems (*19*) that have been widely used but are limited to a small number of existing or similar input molecules. We aimed to design synthetic CIDs that could similarly link binding of a small molecule to modular cellular responses through genetically encoded fusions of each sensor monomer to a split reporter (Fig. 1A) but respond to new, user-defined inputs.

To demonstrate this strategy, we chose farnesyl pyrophosphate (FPP) as the target ligand. FPP is an attractive target because it is a toxic intermediate in a commonly engineered biosynthesis pathway for the production of valuable terpenoid compounds (*20*). Sensors for FPP could be used, for example, to optimize pathway enzymes or, when linked to appropriate outputs, to regulate pathway gene expression in response to changes in metabolite concentrations (*21*). Our computational strategy (Fig. 1B and supplementary materials and methods) proceeds in four main steps: (i) defining the geometries of minimal FPP-binding sites composed of three to four side chains (termed "motif residues") that form key interactions with the target ligand; (ii) modeling these geometries into a dataset of heterodimeric protein-protein interfaces (termed "scaffolds") and computationally screening for coarsely compatible scaffolds (*22*); (iii) optimizing the binding sites in these scaffolds using flexible backbone design methods previously used to predict ligand-binding specificities (*23*–*25*) but not tested in the de novo design of binding sites ("reshaping"); and (iv) ranking individual designs for experimental testing according to several design metrics, including ligand-binding energy predicted using the Rosetta force field (*26*) and ligand burial.

Starting with five FPP-binding-site geometries and up to 3462 heterodimeric scaffolds, we selected the highest-ranked designs across three engineered scaffolds for experimental testing (Fig. 1B and supplementary materials and methods): (i) the FKBP-FRB complex originally responsive to rapamycin (*27*) (one design), (ii) a complex of the bacterial proteins RapF and ComA (*28*) (four designs), and (iii) an engineered complex of maltose-binding protein (MBP) and an ankyrin repeat (AR) protein (*29*) (four designs) (Fig. 2A, table S1, and fig. S1). The ligand was placed into the rapamycin site in FKBP-FRB, but the binding sites in the other two complexes were modeled de novo.

To test these computationally designed FPP sensors, we genetically fused the engineered sensor proteins to a well-studied split reporter, the enzyme murine dihydrofolate reductase [mDHFR (*30*); Fig. 2B and supplementary materials, appendix 1] and expressed the fusion constructs in *Escherichia coli*. We reasoned that functional sensors should exhibit increased growth through FPP-driven dimerization of the sensor proteins and resulting complementation of functional mDHFR under conditions in which endogenous *E. coli* DHFR was specifically inhibited by trimethoprim. Because FPP does not efficiently enter *E. coli*, we added its metabolic precursor, mevalonate, to the growth medium and coexpressed an engineered pathway of five enzymes (*20*) (Fig. 2B, pMBIS) to produce FPP from mevalonate in the cells. We then monitored sensor function as change in growth in the presence or absence of mevalonate under otherwise identical conditions (Fig. 2B and supplementary materials and methods). In the following, we denote designs by their scaffold (S1, S2, and S3), design generation (1, 2, and 3), and successive letter (A, B, C, etc.) (table S1 and fig. S1).

[1]Department of Bioengineering and Therapeutic Sciences, University of California San Francisco, San Francisco, CA, USA. [2]Bioinformatics Graduate Program, University of California San Francisco, San Francisco, CA, USA. [3]Biophysics Graduate Program, University of California San Francisco, San Francisco, CA, USA. [4]UC Berkeley–UCSF Graduate Program in Bioengineering, University of California San Francisco, San Francisco, CA, USA. [5]Department of Pharmaceutical Chemistry, University of California San Francisco, San Francisco, CA, USA. [6]Quantitative Biosciences Institute, University of California San Francisco, San Francisco, CA, USA. [7]Chan Zuckerberg Biohub, San Francisco, CA, USA.
*These authors contributed equally to this work. †Present address: 23andMe, Inc., Therapeutics, South San Francisco, CA, USA. ‡Present address: GRO Biosciences, Boston, MA, USA. §Present address: Gryphon Scientific, LLC, Takoma Park, MD, USA. ‖Present address: Novozymes A/S, Biologiens Vej 2, 2800 Kgs. Lyngby, Denmark. ¶Present address: Adimab LLC, Palo Alto, CA, USA.
#Corresponding author. Email: tanjakortemme@gmail.com

Although seven of the nine selected designs showed only a small signal (S2-1A, B, C, D, S3-1A, B) or no signal (S1-1A), two designs (S3-1C, D) displayed a robust signal response to FPP (Fig. 2C and fig. S2). Both designs resulted from the AR-MBP scaffold (Fig. 2A, S3). For this scaffold, we also generated two libraries: library 1 based on our ensemble design predictions (Fig. 2A and table S2) and library 2 using error-prone polymerase chain reaction (epPCR) starting from design S3-1C. After an initial growth-based selection and subsequent plate-based screens in the presence and absence of FPP (supplementary materials and methods, fig. S3), we identified 36 hits from which we confirmed 27 FPP-responsive sequences by individual growth assays (figs. S4 and S5). One of the most active designs identified across both libraries (S3-2A) was a variant of design S3-1C with two additional mutations distal from the designed FPP-binding site introduced by epPCR. This variant displayed essentially equal activity as the original S3-1C design when tested under identical conditions (Fig. 2C, table S1, and fig. S2). These results show that library screening or epPCR was not necessary to identify functional sensors; instead, we obtained functional sensors directly by computational design. However, library 1 provided additional active sequences from the sequence tolerance predicted in the ensemble design simulations (Fig. 2A, table S2, and fig. S4).

To further characterize the identified best design, S3-2A (table S1), we performed single-site saturation mutagenesis at (SSM) 11 positions (table S3). We tested the resulting mutants with the growth-based split mDHFR reporter in the presence and absence of FPP under more stringent conditions by increasing the trimethoprim concentration (fig. S6). Whereas at most positions the originally designed amino acid (Fig. 2A, design S3-1C) appeared to be optimal under these conditions, we saw considerable improvements for mutations at two positions, R194A (Fig. 2A, design S3-2B) and R194A/L85G (Fig. 2A, design S3-2C). These two designs displayed increasing responses to mevalonate at higher trimethoprim concentrations (Fig. 2D). For the most active design, S3-2C, we confirmed that the sensor signal was dependent on the expression of the sensor proteins [Fig. 2E, -IPTG (isopropyl-β-D-thiogalactopyranoside)] and the metabolic pathway that converts added mevalonate to FPP (Fig. 2E, pMBIS). To test for specificity for FPP, we confirmed that the sensor signal was absent when preventing the accumulation of FPP either by inactivating the fifth enzyme in the pathway by a single point mutation (Fig. 2, B and E, *ispA* R116A) or by adding a sixth enzyme that converts FPP to amorphadiene (Fig. 2, B and E, pB5K). To test whether the sensor signal was dependent on the original four motif side chains, we mutated each individually to alanine and observed decreased sensitivity to the presence of mevalonate for three of the four motif side chains (Fig. 2F, L89, F133, and R145, but not W114). Finally, we tested whether the sensor signal of design

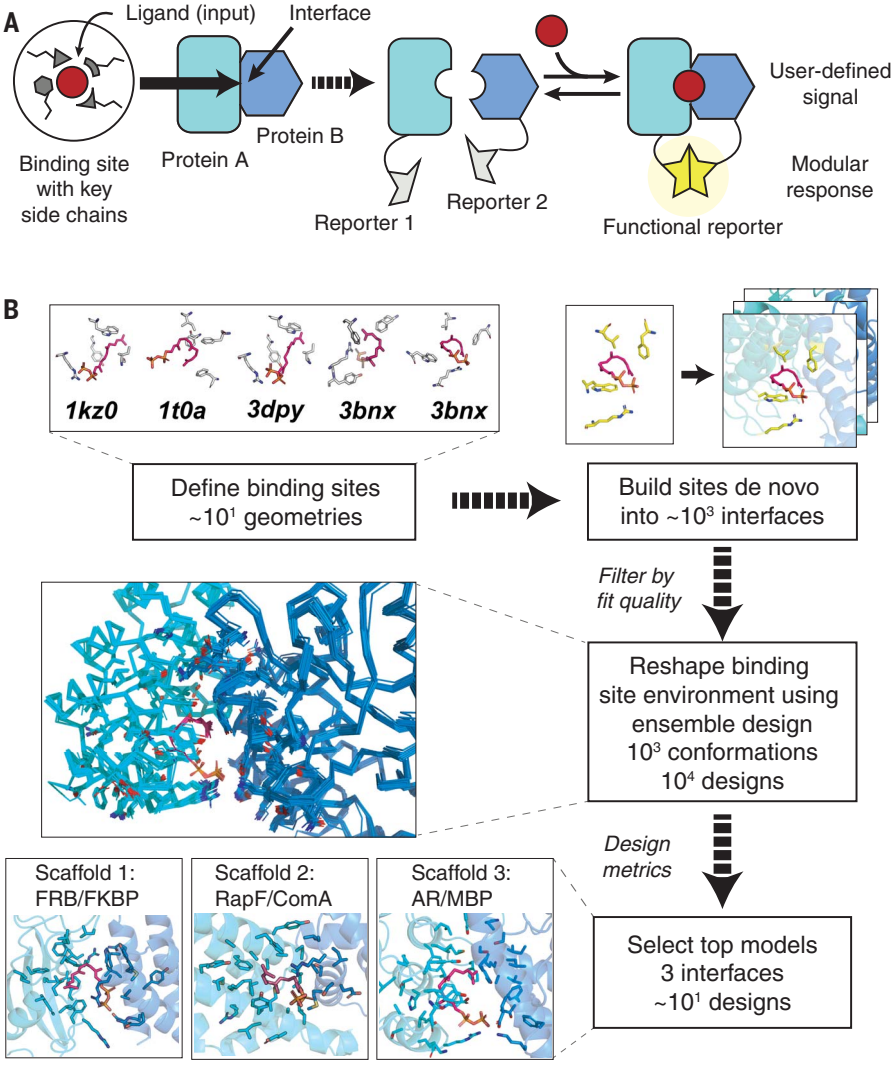


Fig. 1. Computational design. (**A**) Diagram of the design strategy. A small-molecule binding site is built de novo into protein-protein interfaces (left) to create synthetic CIDs (right). Linking the designed sensor proteins to split reporters yields modular CID systems in which different reporter outputs can be coupled to user-defined small-molecule input signals. (**B**) Steps in the design of a synthetic CID system sensing FPP. Top: Binding-site geometries with key interacting side chains selected from FPP-binding proteins (PDB codes indicated) are computationally modeled into a large number of protein-protein interfaces. Middle: Binding sites with feasible geometries are reshaped and optimized by flexible backbone design; shown is a conformational ensemble for a single sequence. Bottom: Top designs from three different scaffolds selected for experimental tests (Fig. 2).

S3-2C was dependent on the concentration of FPP by titrating the extracellular concentration of the mevalonate precursor (Fig. 2G). Although the sensor signal initially increased with increasing mevalonate concentrations, as expected, the signal decreased at the highest mevalonate concentration tested. This behavior likely arises from FPP-mediated toxicity previously observed at this mevalonate concentration using the same FPP-biosynthesis pathway (*20*). We confirmed a consistent dependency of the sensor signal both on sensor expression (by varying the concentration of the inducer, IPTG) and on mevalonate concentration in the growth medium for seven of our designs (fig. S7, S3-1A, B, C, D, S3-2A, B, C). These results confirm that sensor function in *E. coli* is specific to FPP produced by an engineered pathway, dependent on key residues in the de novo designed binding site, dose dependent in *E. coli*, and sensitive to FPP concentrations in a relevant range (i.e., below the toxicity level).

To confirm biochemically that FPP increases the binding affinity of the AR-MBP complex as designed, we purified the designed AR and MBP proteins without attached reporters [see the supplementary materials and methods; these constructs contained several previously published mutations to stabilize AR (*31*), which when tested in the split mDHFR reporter assay led to active sensor S3-2D (table S1, fig. S8, and supplementary materials, appendix 2)]. We determined the apparent binding affinity of the designed AR and MBP proteins comprising the S3-2D sensor (Fig. 3A, table S1, and fig. S1) in the absence and presence of 200 μM FPP using biolayer interferometry (Fig. 3B, fig. S9, and supplementary materials and methods). The presence of FPP led to a >100-fold stabilization of the interaction between the AR and MBP proteins comprising sensor S3-2D [dissociation constant (K_D) from >200 to 2.1 μM, Fig. 3C; for comparison, the original AR-MBP scaffold had a K_D of 4.4 nM (*29*)]. Binding of FPP to the designed AR component of S3-2D alone was weak, and binding of FPP to the designed MBP component of S3-2D alone was not detectable (Fig. 3D). These results confirm in vitro with purified components that design S3-2D functions as a CID system responding to FPP.

To determine whether FPP is recognized in the de novo engineered binding site as predicted by the design model, we determined a 2.2-Å resolution crystal structure of the ternary complex of FPP bound in the engineered AR-MBP interface (supplementary materials and methods; table S4). The crystal structure of the bound complex is in excellent overall agreement with the design model (Fig. 4, A to C). Despite twinning in the crystals, examining unbiased omit maps allowed modeling of unexplained density in the engineered binding site as FPP (Fig. 4B and fig. S10) and confirmed the side-chain conformations in the designed binding pocket (Fig. 4, C and D). Overall, in a 10-Å shell around FPP in the binding pocket, the Cα root mean square

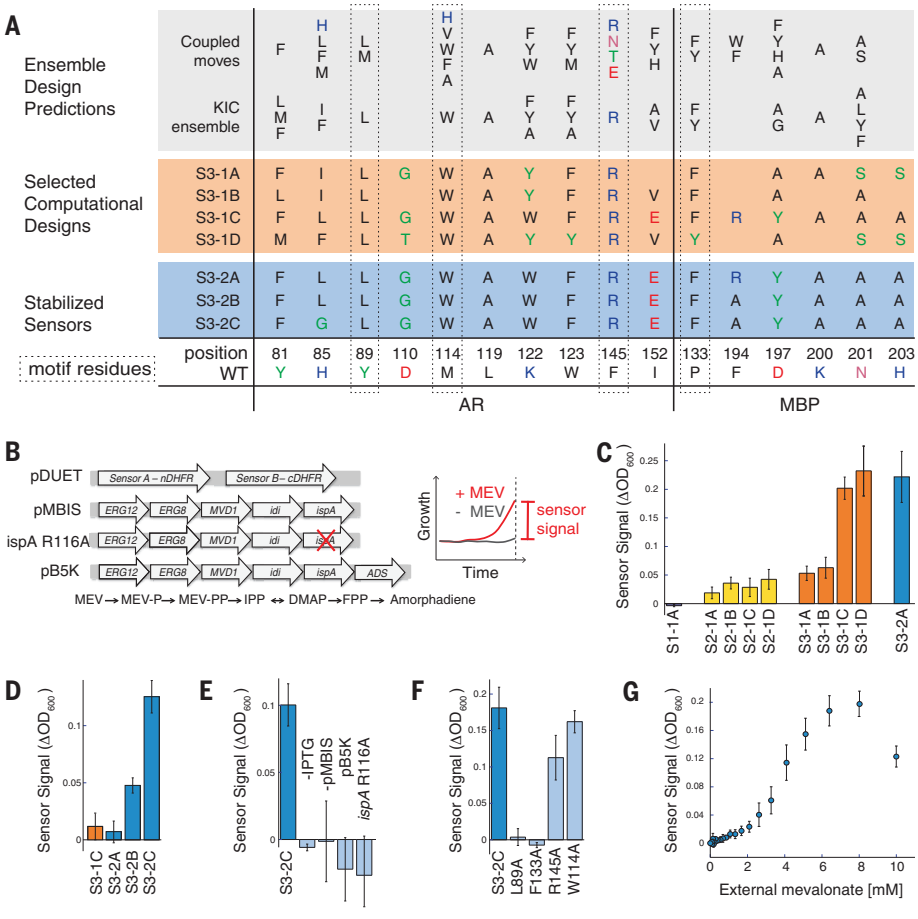


Fig. 2. Sensor function in bacteria. (**A**) Designed sequences at key positions for scaffold 3. Gray shading indicates preferred residues from flexible backbone reshaping by kinematic closure [KIC (*23*, *24*)] or coupled moves (*25*). Orange shading indicates individual computational designs selected based on ligand burial (S3-1A), consensus (S3-1B), optimized ligand packing (S3-1C), and predicted ligand-binding score (S3-1D). Blue shading indicates sensors stabilized by two additional mutations from SSM (S3-2B and S3-2C also contained two mutations from epPCR that were not in the designed FPP-binding site; fig. S1). (**B**) Constructs (left; for details, see supplementary materials, appendix 1) used in the split mDHFR reporter assay (right). pDUET, sensor proteins linked to the split mDHFR reporter; pMBIS, engineered pathway of five enzymes to convert mevalonate (MEV) into FPP (*20*); *ispA* R116A, pMBIS containing R116A mutation in *ispA* that reduces catalytic activity 13-fold (*37*); pB5K, pMBIS with amorphadiene synthase (ADS) (*20*). Sensor signal is quantified as the change in optical density at wavelength 600 nm (OD_{600}) in the presence and absence of mevalonate. (**C**) Sensor signal in the split mDHFR assay for computational designs based on scaffold 1 (FKBP-FRB12, purple bar), scaffold 2 (RapF-ComA, yellow bars), and scaffold 3 (AR-MBP, orange bars). Sensor S3-2A (identified from library 2 with two mutations distal from the designed FPP-binding site; table S1), is shown for comparison (blue bar). (**D**) Improvement of sensor signal by stability-enhancing mutations in S3-2B and S3-2C at increased stringency [trimethoprim concentration 6 μM versus 1 μM in (C)]. (**E**) Dependence of S3-2C sensor signal on sensor expression (-IPTG) and FPP production (-pMBIS, pB5K, *ispA* R116A). (**F**) Dependence of the S3-2C sensor signal on motif residues. (**G**) Dependence of the S3-2C sensor signal on the concentration of the FPP precursor mevalonate added extracellularly. Error bars indicate standard deviation from at least four biological replicates and eight technical replicates for each biological replicate.

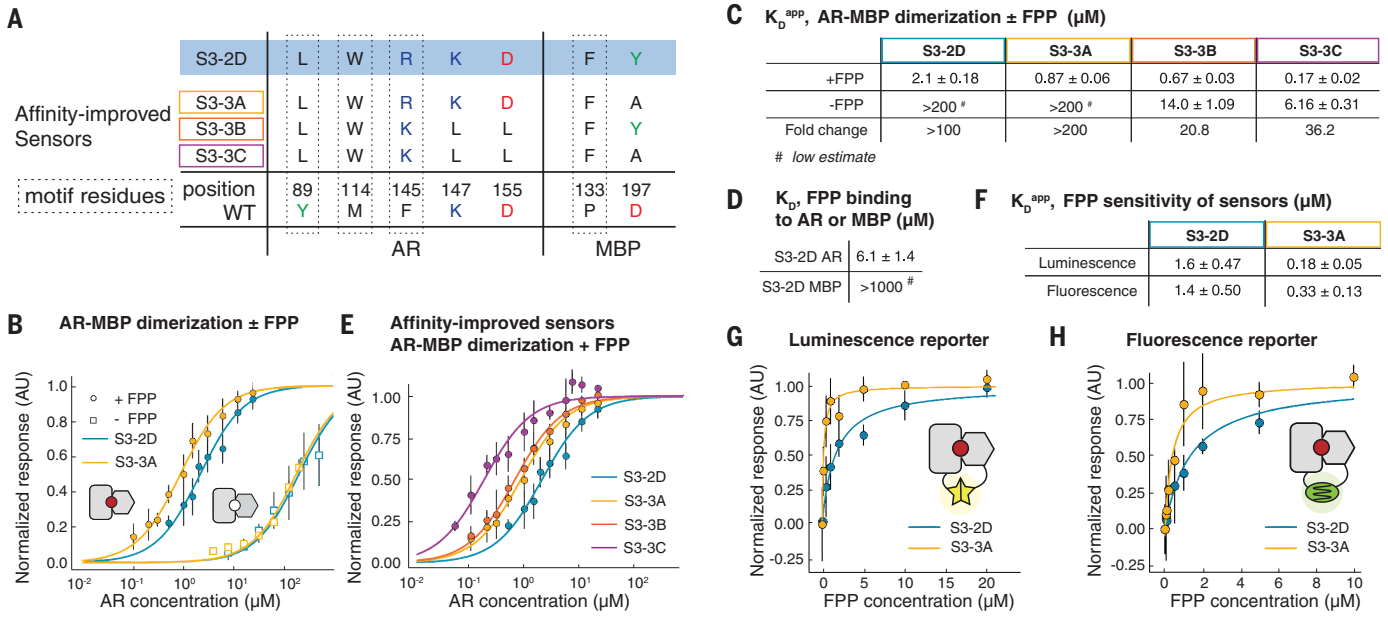


Fig. 3. In vitro sensor characterization and output modularity. (**A**) Sequence changes in sensor constructs tested in vitro. Motif residues are also shown. The starting construct, S3-2D (blue), is identical to S3-2C in the engineered FPP-binding site but contains additional previously published stabilizing mutations in AR (*31*) (shown in table S1). (**B** to **H**) In vitro binding measurements from biolayer interferometry (BLI) using purified protein (B to E) or FPP titrations with sensors expressed using in vitro TxTl (F to H). (B) Apparent AR interaction with immobilized MBP in the presence (closed circles) or absence (open squares) of 200 μM FPP, comparing designs S3-2D (blue) and S3-3A containing the Y197A mutation (orange). (C) Summary of BLI results for apparent AR-MBP dimerization with and without FPP. (D) Summary of BLI results for FPP binding to the individual designed AR and MBP proteins comprising design S3-2D (table S1). (E) Apparent AR interaction with immobilized MBP for a computationally designed variant using the S3-2D crystal structure as the input with (purple, S3-3C) or without (red, S3-3B) the Y197A mutation. (F) Apparent affinity of the S3-2D and S3-3A sensors for FPP using luminescent or fluorescent reporters in TxTl experiments. (G and H) FPP titrations in TxTl using the luminescent reporter (G) or the fluorescent reporter (H). Error bars indicate standard deviations for $n \geq 3$.

deviation (rmsd) between the model and the structure was 0.53 Å and the all-heavy-atom rmsd was 1.13 Å. Although crystals formed only in the presence of FPP, only one of the two complexes in the asymmetric unit contained FPP in the binding site (fig. S11). This behavior allowed us to compare *apo* and *holo* states of the complex. Most of the designed side chains are in identical conformations in the FPP-bound *holo* and FPP-minus *apo* states (Fig. 4E), suggesting favorable preorganization of the designed binding site. An exception is W114 on AR, which is partly disordered in the *apo* state (fig. S11), providing a potential explanation for why a W114A mutation is less detrimental for sensor activity (Fig. 2F) than expected based on the observed packing interactions between W114 and FPP in the *holo* state. A second slight deviation between the model and the crystal structure appeared to be caused by potential steric clashes of the engineered Y197 on MBP with the modeled FPP conformer, which led to rearrangements in the FPP structure and a rotamer change in designed residue F133 on MBP (Fig. 4D). Many of the original models from computational design favored a smaller alanine side chain at this position (Fig. 2A). These observations led to the prediction that a Y197A mutation might stabilize the ternary complex and, indeed, design S3-3A containing the Y197A mutation showed an increased (>200 fold) stabilization of the complex with FPP, with an apparent dissociation constant of 870 ± 60 nM for the designed AR and MBP proteins comprising sensor S3-3A in the presence of 200 μM FPP (Fig. 3, B and C). We also confirmed that design S3-3A (table S1) is active in *E. coli* (fig. S12). To further improve the design based on the crystal structure of design S3-2D, we used an additional round of flexible backbone design using the Rosetta coupled-moves method (*25*) starting from the FPP-bound crystal structure. These simulations suggested three additional mutations leading to design S3-3B: R145K, K147L, and D155L (Fig. 3A). These mutations, when combined with the Y197A mutation (design S3-3C), enhanced the apparent binding affinity of the designed AR and MBP proteins comprising sensor S3-3C in the presence of 200 μM FPP to 170 ± 20 nM (Fig. 3, C and E), which is within 40-fold of the original scaffold AR-MBP interaction affinity (*29*), but also strengthened the binding affinity of the protein–protein dimer in the absence of FPP to 6.2 ± 0.3 μM (fig. S13). The design simulations optimized sequences for stability of the ternary complex without also destabilizing the dimer in the absence of the small molecule. Methods integrating negative design (*32*) could be incorporated to improve the dynamic range of the system (supplementary text).

A key advantage of our CID design strategy is the ability to link an engineered sensor, the input of which is specific to a user-defined small-molecule signal, to a modular output that can in principle be chosen from many available split reporters (Fig. 1A). To test this concept, we linked the engineered CID sensors S3-2D and S3-3A to two additional outputs: a dimerization-dependent fluorescent protein (*33*) and a split luciferase (*34*) (Fig. 3, G and H, and supplementary materials, appendix 3). We tested input-output responses using an in vitro transcription-translation system (TxTl) (*35*) in which FPP can be added at defined concentrations to the assay extract, in contrast to the cell-based split mDHFR assay. The TxTl assay revealed a nanomolar FPP sensitivity (K_D^{app}) for our best sensor, S3-3A (Fig. 3F), that was essentially identical for both reporters (180 ± 50 and 330 ± 130 nM by luminescence and fluorescence detection, respectively; Fig. 3, G and H, and fig. S14), and additionally confirms the improvements in design S3-3A containing the Y197A mutation over design S3-2D (the K_D^{app} for S3-2D was 1.6 ± 0.5 and 1.4 ± 0.5 μM for the luminescence and fluorescence reporters, respectively; Fig. 3, F to H). These results show that our CID

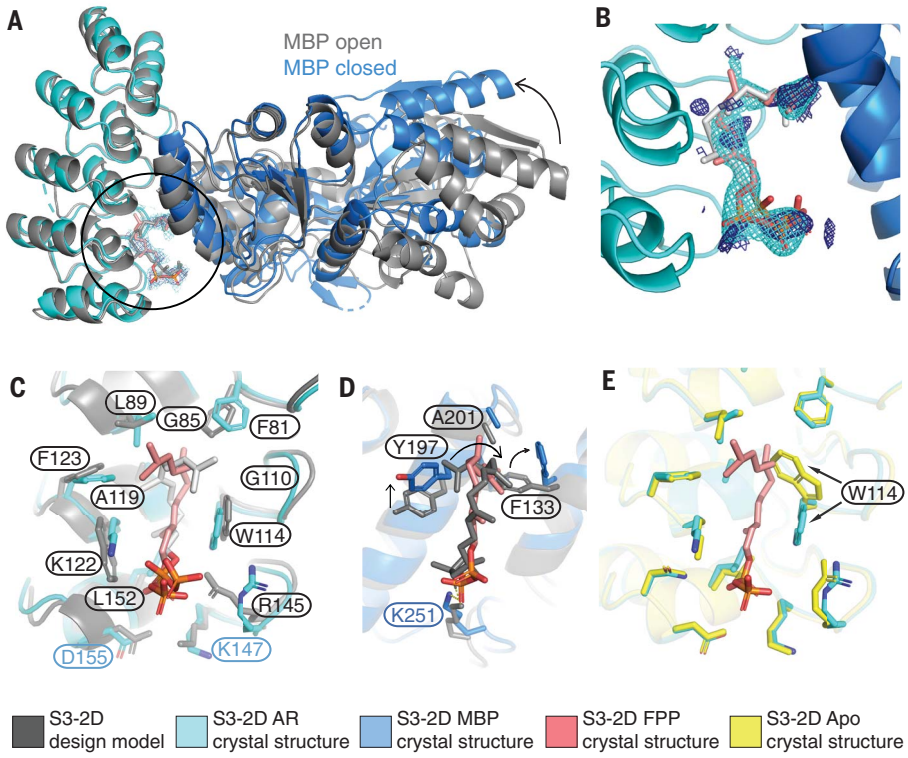


Fig. 4. The S3-2D crystal structure closely matches the computational design model. (**A**) Overlay of the design model (gray) with the crystal structure (designed AR, cyan; designed MBP, blue; FPP, pink) showing FPP binding in the computationally designed binding site at the AR-MBP interface (circle). The design crystallized in the closed MBP conformation, whereas MBP was in the open conformation in the original scaffold on which the model was based, leading to a difference in rigid-body orientation (arrow) of one lobe of MBP distal to the FPP-binding site. (**B**) FPP overlaid with $2mF_o - DF_c$ electron density map (1.2σ, cyan) and ligand $2mF_o - DF_c$ omit map (1.0σ, dark blue). Strong density peaks were present in both maps for the phosphates and several anchoring hydrophobic groups. (**C**) Open-book representation of the FPP-binding site on AR showing close match of designed side-chain conformations to the crystal structure. (**D**) Open-book representation of the FPP-binding site on MBP indicating a clash between the position of MBP Y197 in the crystal structure (blue) and the designed FPP orientation in the model (gray), causing slight rearrangements of FPP and F133 (arrows). (**E**) Alignment of the *holo* (cyan) and *apo* (yellow) structures of S3-2D showing overall agreement with the exception of the side chain of W114 (arrows). In (C) to (E), residues are labeled black when designed and cyan or blue when present in the original scaffold complex.

sensor design strategy is compatible with modular outputs.

The most critical feature of our approach is the ability to computationally design small-molecule binding sites de novo into protein–protein interfaces. A previous computational analysis suggested that the appearance of pockets around artificially generated protein–protein interfaces may be an intrinsic geometric feature of protein structure (*36*), lending support to the idea that our approach is extensible to many other small molecules and interfaces. The design method presented here thus introduces a generalizable way to create synthetic sensing systems with different outputs that can be used in diverse biological contexts to respond to user-specified molecular signals.

REFERENCES AND NOTES

1. P. S. Huang *et al.*, *Science* **346**, 481–485 (2014).
2. T. M. Jacobs *et al.*, *Science* **352**, 687–690 (2016).
3. A. R. Thomson *et al.*, *Science* **346**, 485–488 (2014).
4. R. B. Hill, D. P. Raleigh, A. Lombardi, W. F. DeGrado, *Acc. Chem. Res.* **33**, 745–754 (2000).
5. P. B. Harbury, J. J. Plecs, B. Tidor, T. Alber, P. S. Kim, *Science* **282**, 1462–1467 (1998).
6. G. J. Rocklin *et al.*, *Science* **357**, 168–175 (2017).
7. N. Koga *et al.*, *Nature* **491**, 222–227 (2012).
8. P. S. Huang *et al.*, *Nat. Chem. Biol.* **12**, 29–34 (2016).
9. E. Marcos *et al.*, *Nat. Struct. Mol. Biol.* **25**, 1028–1034 (2018).
10. J. Dou *et al.*, *Nature* **561**, 485–491 (2018).
11. H. Lechner, N. Ferruz, B. Höcker, *Curr. Opin. Chem. Biol.* **47**, 67–76 (2018).
12. J. Dou *et al.*, *Protein Sci.* **26**, 2426–2437 (2017).
13. R. M. Gordley, L. J. Bugaj, W. A. Lim, *Curr. Opin. Struct. Biol.* **39**, 106–114 (2016).
14. M. J. Bick *et al.*, *eLife* **6**, e28909 (2017).
15. C. E. Tinberg *et al.*, *Nature* **501**, 212–216 (2013).
16. N. F. Polizzi *et al.*, *Nat. Chem.* **9**, 1157–1164 (2017).
17. J. Feng *et al.*, *eLife* **4**, e10606 (2015).
18. N. D. Taylor *et al.*, *Nat. Methods* **13**, 177–183 (2016).
19. D. M. Spencer, T. J. Wandless, S. L. Schreiber, G. R. Crabtree, *Science* **262**, 1019–1024 (1993).
20. V. J. Martin, D. J. Pitera, S. T. Withers, J. D. Newman, J. D. Keasling, *Nat. Biotechnol.* **21**, 796–802 (2003).
21. F. Zhang, J. Keasling, *Trends Microbiol.* **19**, 323–329 (2011).
22. A. Zanghellini *et al.*, *Protein Sci.* **15**, 2785–2794 (2006).
23. M. Babor, D. J. Mandell, T. Kortemme, *Protein Sci.* **20**, 1082–1089 (2011).
24. D. J. Mandell, E. A. Coutsias, T. Kortemme, *Nat. Methods* **6**, 551–552 (2009).
25. N. Ollikainen, R. M. de Jong, T. Kortemme, *PLOS Comput. Biol.* **11**, e1004335 (2015).
26. R. F. Alford *et al.*, *J. Chem. Theory Comput.* **13**, 3031–3048 (2017).
27. J. Choi, J. Chen, S. L. Schreiber, J. Clardy, *Science* **273**, 239–242 (1996).
28. M. D. Baker, M. B. Neiditch, *PLOS Biol.* **9**, e1001226 (2011).
29. H. K. Binz *et al.*, *Nat. Biotechnol.* **22**, 575–582 (2004).
30. J. N. Pelletier, F. X. Campbell-Valois, S. W. Michnick, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 12141–12146 (1998).
31. M. A. Kramer, S. K. Wetzel, A. Plückthun, P. R. Mittl, M. G. Grütter, *J. Mol. Biol.* **404**, 381–391 (2010).
32. J. A. Davey, R. A. Chica, *Protein Sci.* **21**, 1241–1252 (2012).
33. S. C. Alford, Y. Ding, T. Simmen, R. E. Campbell, *ACS Synth. Biol.* **1**, 569–575 (2012).
34. A. S. Dixon *et al.*, *ACS Chem. Biol.* **11**, 400–408 (2016).
35. R. Marshall, V. Noireaux, *Methods Mol. Biol.* **1772**, 61–93 (2018).
36. M. Gao, J. Skolnick, *Proc. Natl. Acad. Sci. U.S.A.* **109**, 3784–3789 (2012).
37. L. Song, C. D. Poulter, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 3044–3048 (1994).

ACKNOWLEDGMENTS

We thank J. Keasling and F. Zhang for advice on FPP production in microbes and pathway constructs; E. de los Santos, Z. Sun, V. Noireux, and R. Murray for TxTl advice and reagents; S. Alford and R. Campbell for dimerization-dependent fluorescent protein constructs; A. Anand, V. Ruiz, B. Adler, and A. Maxwell for contributions to computational design and characterization; S. O'Connor for developing a database for design models; and members of the Kortemme lab for discussion. **Funding:** This work was supported by a grant from the National Institutes of Health (NIH) (R01-GM110089) and a W.M.F. Keck Foundation Medical Research Award to T.K. We additionally acknowledge the following fellowships: NIH IRACDA and UC Chancellor's Postdoctoral Fellowships (A.A.G.), PhRMA Foundation Predoctoral Fellowship in Informatics (D.J.M.), NIH F32 Postdoctoral Fellowship (M.T.), and National Science Foundation Graduate Research Fellowships (J.P. and N.O.). **Author contributions:** D.J.M. and T.K. conceived the idea for the project; D.J.M. developed and performed the majority of the computational design with contributions from A.A.G., R.A.P., K.A.B., N.O., J.P., and T.K.; A.A.G. and Y.-M.H. designed the experimental approach and performed the majority of the experimental characterization with contributions from R.R., A.L.L., C.K., D.J., and M.J.S.K.; M.T. and J.S.F. determined the crystal structure; M.J.S.K., J.S.F., and T.K. provided guidance, mentorship, and resources; and A.A.G. and T.K. wrote the manuscript with contributions from all authors. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** Coordinates and structure files have been deposited to the Protein Data Bank (PDB) with accession code 6OB5. All other relevant data are available in the main text or the supplementary materials. Rosetta source code is available from rosettacommons.org. Upon publication, constructs will be made available by Addgene.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/366/6468/1024/suppl/DC1
Materials and Methods
Supplementary Text
Tables S1 to S4
Figs. S1 to S14
References (*38*–*73*)

1 May 2019; accepted 7 October 2019
10.1126/science.aax8780

Computational design of a modular protein sense-response system

Anum A. Glasgow, Yao-Ming Huang, Daniel J. Mandell, Michael Thompson, Ryan Ritterson, Amanda L. Loshbaugh, Jenna Pellegrino, Cody Krivacic, Roland A. Pache, Kyle A. Barlow, Noah Ollikainen, Deborah Jeon, Mark J. S. Kelly, James S. Fraser and Tanja Kortemme

Science **366** (6468), 1024-1028.
DOI: 10.1126/science.aax8780

Sense and respond

Many signaling pathways start with cellular proteins sensing and responding to small molecules. Despite advances in protein design, creating a protein-based sense-and-respond system remains challenging. Glasgow *et al.* designed binding sites at the interface of protein heterodimers (see the Perspective by Chica). By fusing each monomer to one half of a split reporter, they linked ligand-driven dimerization to the reporter output. The computational design strategy provides a generalizable approach to create synthetic sensing systems with different outputs.

Science, this issue p. 1024; see also p. 952

Science (print ISSN 0036-8075; online ISSN 1095-9203) is published by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005. The title *Science* is a registered trademark of AAAS.

science.sciencemag.org/content/366/6468/1024/suppl/DC1

Supplementary Materials for

Computational design of a modular protein sense–response system

Anum A. Glasgow*, Yao-Ming Huang*, Daniel J. Mandell*, Michael Thompson, Ryan Ritterson, Amanda L. Loshbaugh, Jenna Pellegrino, Cody Krivacic, Roland A. Pache, Kyle A. Barlow, Noah Ollikainen, Deborah Jeon, Mark J. S. Kelly, James S. Fraser, Tanja Kortemme†

*These authors contributed equally to this work.
†Corresponding author. Email: tanjakortemme@gmail.com

Published 22 November 2019, *Science* **366**, 1024 (2019)
DOI: 10.1126/science.aax8780

This PDF file includes:

Materials and Methods

1. Computational design methods

Computational methods that resulted in the successful sensor design for scaffold 3 (AR-MBP) are as described in detail first(*38*). Differences in the design protocol that resulted in designs for scaffold 1 (FKBP12-FRB) and scaffold 2 (RapF-ComA) are described further below.

1.1 Selection of protein-protein interface scaffold set. To identify protein interface scaffolds suitable for accommodating small molecule binding sites, we searched the PDB for heterodimeric complexes with ≤95% sequence identity solved at ≤2.8 Å resolution between chains of 75 to 300 residues that were expressed in *E. coli*. This search resulted in 612 structures (an expanded search of the PDB resulted in up to 3462 scaffolds, see section 1.7) that were filtered to remove HETATM records and multiple densities (only the first densities listed were kept). Selenomethionines were converted to methionines.

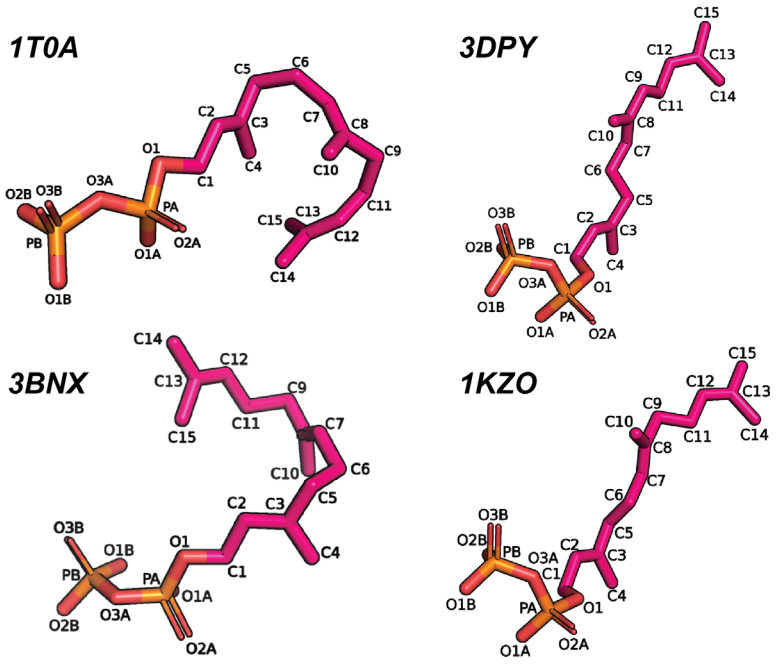
1.2 Definition of FPP binding site geometries. We identified 29 X-ray structures of FPP–protein complexes at ≤2.8 Å resolution in the PDB to serve as templates for binding site geometries. We visually inspected each protein-FPP complex to identify cases where 4 residues could define an encompassing portion of the FPP binding surface. 18 structures were discarded because FPP bound in complex with an inhibitor or other small molecule, forming a binding site that cannot easily be reproduced by amino acid side-chains. Other cases were discarded because the binding site was formed by small contributions from too many residues to define a suitable binding site geometry. Ultimately, 4 template binding site geometries (“motifs”) were selected for subsequent matching

and design (a fifth distinct binding site geometry derived from 3bnx is used in the runs leading to designs on scaffold 2, see section 1.7):

PDB	Protein	Motif residues
1kzo	Protein farnesyltransferase	chain B: R291, Y251, W303. chain C: I10.
1t0a	2C-Methyl-D-Erythritol-2,4-cyclodiphosphate Synthase	chain A: I101, F9. chain B: F9. chain C: F9.
3bnx	Aristolochene synthase	chain A: R314, W308, L184, F153.
3dpy	Protein farnesyltransferase	chain B: R291, Y251, W303. chain C: I2008.

Note that for PDB templates 1kzo and 3dpy, one of the motif residues comes from a co-associated peptide substrate, and the binding site geometry from 1t0a contains residues from a homotrimeric interface. In all cases non-polar hydrogens were added to FPP, and a single polar hydrogen was placed on the O5 oxygen.

The dihedral angles for the conformations of FPP used in the template binding site geometries, with PDB codes, are included below:



1KZO

Index	Atom 1	Atom 2	Atom 3	Atom 4	Angle
1	O1B	PB	O3A	PA	56.578
2	O2B	PB	O3A	PA	-66.439
3	O3B	PB	O3A	PA	179.393
4	PB	O3A	PA	O1A	-9.683
5	PB	O3A	PA	O2A	-132.538
6	O3A	PA	O1A	O2A	-119.396
7	O3A	PA	O1	C1	-82.579
8	PA	O1	C1	C2	129.629
9	O1	C1	C2	C3	108.359
10	C1	C2	C3	C4	-0.25
11	C2	C3	C5	C6	66.367
12	C4	C3	C5	C6	-113.68
13	C3	C5	C6	C7	173.344
14	C5	C6	C7	C8	141.175
15	C6	C7	C8	C10	0
16	C6	C7	C8	C9	179.895
17	C7	C8	C9	C11	-59.234
18	C8	C9	C11	C12	-174.743
19	C10	C8	C9	C11	120.679
20	C9	C11	C12	C13	177.13
21	C11	C12	C13	C14	-0.082
22	C11	C12	C13	C15	179.929

1T0A

Index	Atom 1	Atom 2	Atom 3	Atom 4	Angle
1	O1B	PB	O3A	PA	-0.78
2	O2B	PB	O3A	PA	-124.633
3	O3B	PB	O3A	PA	121.527
4	PB	O3A	PA	O2A	-69.47
5	PB	O3A	PA	O1A	60.816
6	O3A	PA	O1A	O2A	-130.509
7	O3A	PA	O1	C1	-170.861
8	PA	O1	C1	C2	-170.407
9	O1	C1	C2	C3	-152.506
10	C1	C2	C3	C4	1.032
11	C2	C3	C5	C6	84.311
12	C4	C3	C5	C6	-95.668
13	C3	C5	C6	C7	67.587
14	C5	C6	C7	C8	-102.897
15	C6	C7	C8	C10	1.709
16	C6	C7	C8	C9	-176.548
17	C7	C8	C9	C11	-102.331
18	C8	C9	C11	C12	-46.225
19	C10	C8	C9	C11	79.292

20	C9	C11	C12	C13	131.953
21	C11	C12	C13	C14	-179.556
22	C11	C12	C13	C15	0.225

3BNX

Index	Atom 1	Atom 2	Atom 3	Atom 4	Angle
1	O1B	PB	O3A	PA	3.718
2	O2B	PB	O3A	PA	123.804
3	O3B	PB	O3A	PA	-116.979
4	PB	O3A	PA	O2A	-142.881
5	PB	O3A	PA	O1A	-23.229
6	O3A	PA	O1A	O2A	-119.484
7	O3A	PA	O1	C1	179.743
8	PA	O1	C1	C2	-178.706
9	O1	C1	C2	C3	160.066
10	C1	C2	C3	C4	0.771
11	C2	C3	C5	C6	-121.614
12	C4	C3	C5	C6	58.834
13	C3	C5	C6	C7	102.037
14	C5	C6	C7	C8	-110.856
15	C6	C7	C8	C10	-0.052
16	C6	C7	C8	C9	179.905
17	C7	C8	C9	C11	-106.179
18	C8	C9	C11	C12	126.417
19	C10	C8	C9	C11	73.778
20	C9	C11	C12	C13	-132.226
21	C11	C12	C13	C14	-179.476
22	C11	C12	C13	C15	0.214

3DPY

Index	Atom 1	Atom 2	Atom 3	Atom 4	Angle
1	O1B	PB	O3A	PA	51.274
2	O2B	PB	O3A	PA	-68.156
3	O3B	PB	O3A	PA	171.897
4	PB	O3A	PA	O2A	-125.8
5	PB	O3A	PA	O1A	-1.447
6	O3A	PA	O1A	O2A	-127.941
7	O3A	PA	O1	C1	-70.847
8	PA	O1	C1	C2	142.79
9	O1	C1	C2	C3	112.354
10	C1	C2	C3	C4	-0.697
11	C2	C3	C5	C6	-36.839
12	C4	C3	C5	C6	143.835
13	C3	C5	C6	C7	-175.389
14	C5	C6	C7	C8	-137.806
15	C6	C7	C8	C10	0.636

16	C6	C7	C8	C9	-178.932
17	C7	C8	C9	C11	-8.651
18	C8	C9	C11	C12	176.477
19	C10	C8	C9	C11	171.775
20	C9	C11	C12	C13	132.37
21	C11	C12	C13	C14	-0.463
22	C11	C12	C13	C15	179.575

1.3 Building binding sites *de novo* into protein-protein interfaces. We scanned the interface scaffold set for backbones that may accommodate the small molecule target and binding site geometry using a geometric matching procedure(*22*). For each binding site geometry, the relationship between the motif side-chains and the target is uniquely defined by 6 geometric constraints. The matching algorithm scans the first motif residue constraints across a set of scaffold residue positions (here, all positions with Ca atoms within 15 Å of the other chain). At each position, the motif residue is placed into rotameric conformations from the Dunbrack backbone-dependent rotamer library(*39*). For each side-chain conformation, the small molecule target is placed relative to the motif residue using the geometric constraints defined from the template binding site geometry. Conformations that place the target without introducing steric clashes between the motif side-chain, the target, and the scaffold backbone are recorded. The process is iterated for the remaining motif residues, comparing the clash-free target positions to those from the previous motif residues using an efficient geometric hashing technique. Motif side-chain conformations that place the target within the same geometric 'bin' as positions from the previous motif side-chains are recorded as 'hits'. At the conclusion, cases where the target is placed into the same geometric bin for all motif residues are called 'matches'. Only matches where at least one motif side-chain is placed on a different chain than the remaining motif side-chains (i.e., across the scaffold interface) are considered further. Many matches may be found for a set of scaffold

interface residue positions, corresponding to highly similar conformations for the motif residue side-chains and the target that fall into the same geometric bin. One such match is randomly selected for design, which is termed a 'unique match'. The numbers of unique matches arising from each template binding site geometry for FPP are shown below:

Template PDB	Motif residues	Number of unique matches
1kzo	chain B: R291, Y251, W303. chain C: I10.	79
1t0a	chain A: I101, F9. chain B: F9. chain C: F9.	371
3bnx	chain A: R314, W308, L184, F153.	370
3dpy	chain B: R291, Y251, W303. chain C: I2008.	43

The quality and quantity of matching results are tuned by a number of parameters. To slightly relax the angle and torsion constraints, we sampled 5 degrees above and below the values computed from the template binding site geometry. There are also Euclidean and Euler parameters that determine the bin size for geometric hashing, which we set to 2.0 Å and 20.0 degrees, respectively. A bump tolerance parameter allowed for some steric overlap – to be resolved in the design stage – which we set to 0.6. We also allowed motif residues to be matched by other residue types with similar side-chain moieties. The following groups of residues were allowed to be matched by any residue in the group: "DE", "LVI", "FYW", and "ST". We tuned the matching parameters to produce a reasonable number of unique matches for design (on the order of several hundred). Evaluating matches directly is not necessarily informative, since a good match (one that faithfully reproduces the template binding site geometry) may yield poor designs due to an inability of the 'second shell' residues to accommodate the binding site geometry and target, while less precise matches may yield more promising designs after being subjected to rigid body optimization of the ligand and backbone relaxation of the scaffold. Thus, all unique matches are passed on to design.

A number of parameters control Rosetta-specific matching options, including the number of side-chain conformations sampled. The command line options used for Rosetta revision r35441 were as follows:

match.linuxgccrelease -database minirosetta_database -s 1SVX.pdb -match:lig_name LG1 -match:grid_boundary 1SVX.gridlig -match:scaffold_active_site_residues 1SVX.pos -match:geometric_constraint_file 3bnx.cst -extra_res_fa 3bnx_LG.fa.params -output_matches_per_group 10 -ex1 -ex2 -extrachi_cutoff 0 -euclid_bin_size 2.0 -euler_bin_size 20.0 -bump_tolerance 0.6 -match:output_format PDB -match:consolidate_matches -match:output_matchres_only

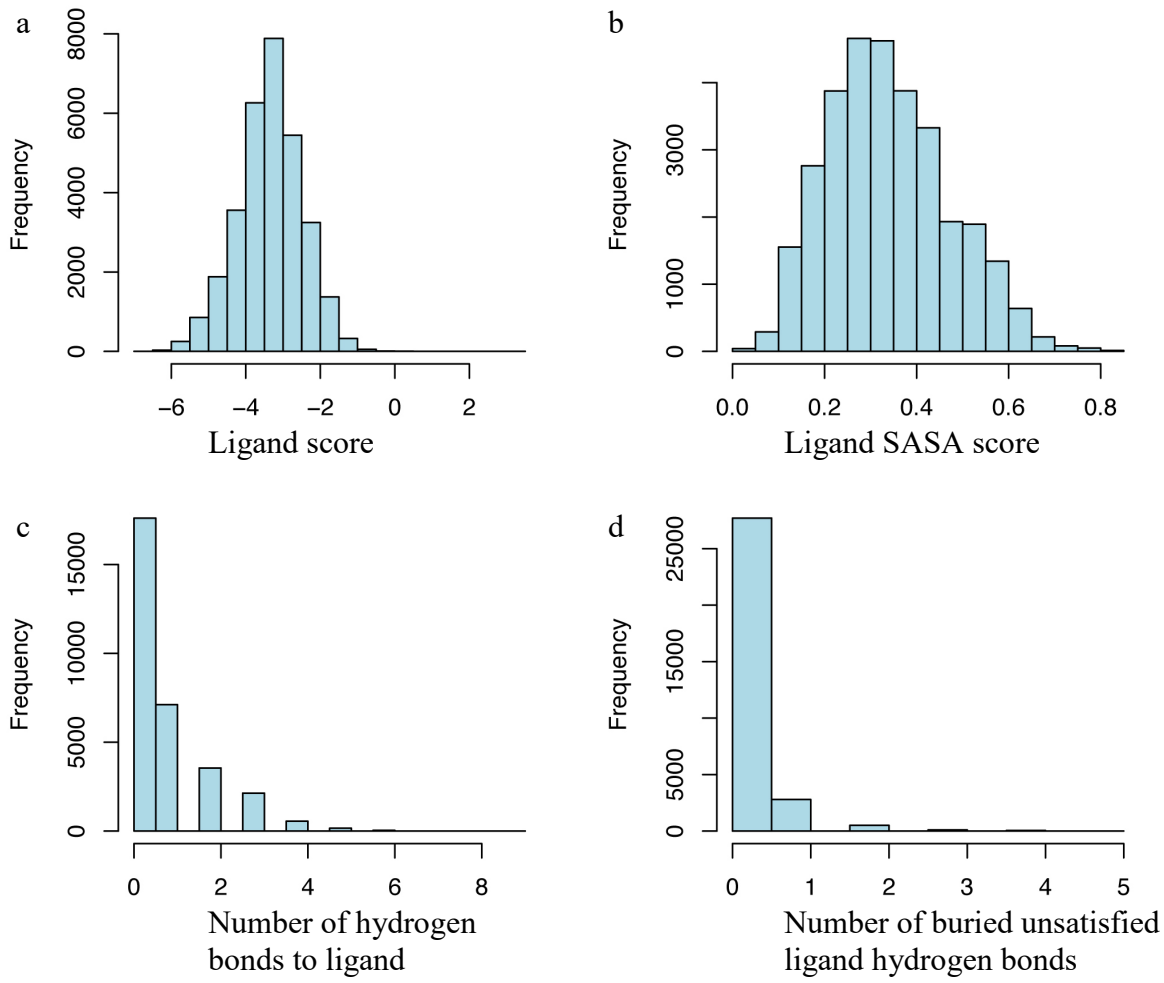
1.4 Initial designs and analysis. The matching algorithm places the motif residues and target into a scaffold interface while avoiding clashes with the backbone. However, the procedure will likely introduce unfavorable interactions with the residues surrounding the motif, or 'second shell' residues. In order to accommodate the target and binding motif, we applied a protocol that iterates between rigid body optimization of the ligand(*40*) and sequence design of the second shell residues(*41, 42*). Rigid body optimization of the ligand was performed by sampling from a Gaussian distribution with mean 2.0 degrees for rotation and from a Gaussian distribution with mean 0.1Å for translation of the ligand, which are the default values for -enzdes:trans_magnitude and -enzdes:rot_magnitude. In the design step, all residues with a Ca atom within 6.0 Å of any ligand heavy atom were designable (they could change residue type), as well as any residue with a Ca atom within 8.0 Å of any ligand heavy atom that also has a Cb atom closer to the ligand than the Ca atom. Additionally, all residues with a Ca atom within 10.0 Å of any ligand heavy atom are

subject to repacking (Metropolis Monte Carlo optimization of side-chain conformations) together with any residue with a Ca atom within 12.0 Å of the ligand with a Cb atom that is closer to the ligand than the Ca atom. The protocol iterates rigid body optimization and sequence design 3 times, producing interfaces with more favorable interactions between the motif residues, target, and second shell residues.

The command line options used for Rosetta revision r36129 were as follows:

EnzdesFixBB.linuxgccrelease -database minirosetta_database -s 1BH9_R33Y94L120F121.pdb -extra_res_fa 3bnx_LG.fa.params enzdes:detect_design_interface -enzdes:cut1 6.0 -enzdes:cut2 8.0 -enzdes:cut3 10.0 -enzdes:cut4 12.0 -enzdes:cst_opt -enzdes:cst_design -enzdes:cst_min -enzdes:cstfile 3bnx.cst -enzdes:bb_min -enzdes:chi_min -enzdes:design_min_cycles 3 -ex1 -ex2 -use_input_sc -nstruct 999 -enzdes:start_from_random_rb_conf

For each template binding site geometry we produced on the order of 10^4 designs and created distributions over computed physicochemical properties. Four of these distributions from the 3bnx template binding site geometry are shown below:



The ligand score (panel a) corresponds to the predicted binding energy between the ligand and scaffold interface. The ligand solvent accessible surface area (SASA, panel b) score measures the burial of the ligand from 0.0 (completely solvent exposed) to 1.0 (completely buried). The number of hydrogen bonds between the scaffold and the ligand (panel c) and the number of buried unsatisfied hydrogen bonds on the ligand (panel d) are also shown. For FPP, visual inspection of representative members of the distributions across all designs suggested the following filter for selecting designs for further refinement: ligand score < -6.0, ligand SASA > 0.6, ligand hydrogen bonds > 1, unsatisfied buried ligand hydrogen bonds = 0. The number of designs passing the filter for all FPP binding site geometries is given below:

Template PDB	Motif residues	Number of unique matches	Number of designs passing filter
1kzo	chain B: R291, Y251, W303. chain C: I10	79	31
1t0a	chain A: I101, F9. chain B: F9. chain C: F9.	371	0 (1 if SASA filter relaxed to 0.5)
3bnx	chain A: R314, W308, L184, F153.	370	81
3dpy	chain B: R291, Y251, W303. chain C: I2008.	43	0 (2 if SASA filter relaxed to 0.5)

Passing designs were then further filtered to remove interface scaffolds imposing additional challenges such as cases with small molecules crystallized at the predicted target binding site and complexes that were purified from inclusion bodies.

Ultimately, a design passing all filters with the best ligand score on an interface scaffolds, the complex (PDB 1svx) between an engineered ankyrin repeat protein (AR) and maltose binding protein (MBP) with a 4-residue binding site geometry from template 3bnx, was selected for refinement by flexible backbone ensemble design.

1.5 Flexible backbone ensemble design. To model the conformational adjustments that could occur in concert with sequence mutations, matched scaffold designs were subject to kinematic closure (KIC(*24*)) over their entire backbones producing conformational ensembles. In brief, KIC generates backbone conformations of segments in proteins by sampling backbone phi/psi torsion angles for *n*-6 degrees of freedom ("non-pivot" torsions) in a selected segment, and then solving the remaining 6 "pivot" degrees of freedom analytically to close the loop. To generate a protein backbone ensemble, different segment start and end points are sampled throughput the protein. Here, we generated near-native conformational ensembles with KIC (200 conformations with 0.9 Å average rmsd to the X-ray structure) using a modified protocol(*23*) compared to the published

de novo loop reconstruction method(*24*). The ensemble generation protocol skips the low-resolution centroid stage and fixes the temperature at 1.2 *k*T. Further, to focus sampling on near-native conformations, non-pivot torsions are sampled within a vicinity of 3 degrees of the input value before each kinematic move, instead of sampling from the allowable Ramachandran space.

A second round of Rosetta sequence design was then applied across the ensembles to the side-chains surrounding the *de novo* built binding site in order to accommodate the motif residues and the small molecule target. Designing across a conformational ensemble, rather than a single backbone, can improve agreement between the geometry of the binding site built into the scaffold interface and the original geometry in the binding site template, and generates a diversity of predicted low-energy sequences(*23, 43, 44*).

The command line used to generate the ensemble with Rosetta r36129 was as follows:

loopmodel.linuxgccrelease -database minirosetta_database -loops:refine refine_kic -loops:max_kic_build_attempts 10000 -loops:input_pdb 1SVX_R134W103L78Y286__DE_19.pdb -loops:loop_file 1SVX_R134W103L78Y286__DE_19.loop -extra_res_fa 1SVX_R134W103L78Y286__DE_19_LG.fa.params -in:file:extra_res_cen 1SVX_R134W103L78Y286__DE_19_LG.cen.params -in:file:native 1SVX_R134W103L78Y286__DE_19.pdb -loops:kic_max_seglen 12 -loops:outer_cycles 1 -loops:refine_init_temp 1.2 -loops:refine_final_temp 1.2 -loops:vicinity_sampling -loops:vicinity_degree 3 -ex1 -ex2

A similar protocol for small molecule rigid body optimization and scaffold sequence design from section 1. 4 was applied to refine every member of the KIC ensemble. Additional rotamers were included (*via* the -ex3 -ex4 Rosetta command line flags), and the selection of residues to be redesigned, fixed, or modeled as wild-type was performed manually rather than selected by distance cutoffs (redesigned positions for the 1svx scaffold are shown in **Figure 2A**). Designs resulting from this step were further optimized for small molecule dependence on complex formation by requiring a ligand score < -6.0, ligand SASA > 0.7, ligand hydrogen bonds > 1, and unsatisfied buried ligand hydrogen bonds = 0. In addition to selecting single designs (see below), designing across a conformational ensemble, rather than a single backbone, produced a sequence library(*45*) (**Table S2**) to be assayed in *E. coli* for biosensor activity. In KIC ensemble design, we performed 200 independent design simulations in parallel (one for each backbone) and then combined the results into a sequence profile computed from the lowest energy sequence and conformation from simulated annealing Metropolis Monte-Carlo simulations in the design step. Design calculations minimized the Rosetta full-atom energy (see command lines for version numbers and options). The most frequently designed amino acid residues at each position are shown in **Fig. 2A**.

In addition to KIC ensemble design, we also used a second method for flexible backbone design: coupled moves(*25*). In contrast to KIC ensemble design, which first pre-generates an ensemble and then performs design on each member of the ensemble, coupled moves simultaneously moves the backbone and side chain of an amino acid (or rigid-body and conformer of the ligand) during design. In these simulations, we started from the model of design S3-1C (**Fig. 2A**) with FPP placed into the AR-MBP scaffold through matching with 4 motif residues (AR: L89, W114, R145; MBP:

F133), followed by coupled moves design performed as described in (*25*), allowing the motif residues and the following additional residues to design: AR: Y81, H85, Y89, W112, M114, T115, H118, L119, K122, W123, F145, K147, I152, D155; MBP: E130, P133, F194, D197, K200, N201, K251. Ligand rigid-body rotations and translations were sampled using two Gaussian distributions with a 1 degree standard deviation for rotations and a 0.1Å standard deviation for translations. Coupled moves design simulations were performed using Rosetta rotamer flags "-ex1 -ex2 -extrachi_cutoff 0 -use_input_sc" and were run at a constant temperature of 0.6 kT. The backbone moves used in these simulations were 3-residue "backrub" moves(*46*) centered on the amino acid being designed. Amino acid residues preferred in these simulations at key positions are shown in **Figure 2A**.

KIC ensemble simulations took approximately 30 minutes on a single processor and were run over 200 nodes, for a total of approximately 100 CPU hours. In comparison, coupled moves simulations took approximately 4 CPU hours (12 minutes per simulation and 20 simulations total).

1.6 Design ranking and selection. We chose four computational designs for FPP sensors based on the AR/MBP (PDB 1svx) scaffold for testing:

S3-1A: Computational design ranked most highly by ligand burial

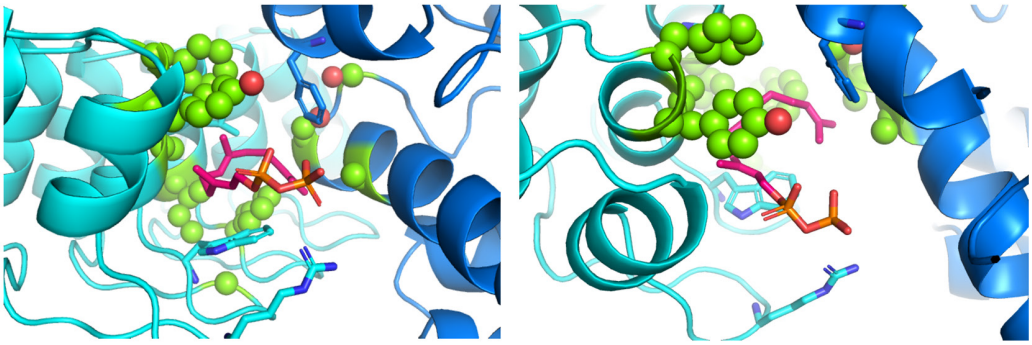
S3-1B: Computational design consensus sequence, based on most frequently selected amino acid residues in KIC ensemble design

S3-1C: Computational design with improved protein-ligand packing interactions

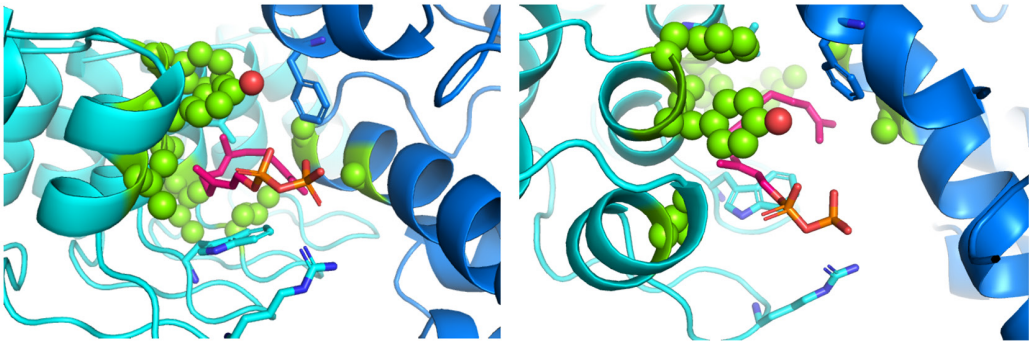
S3-1D: Computational design ranked most highly by predicted interaction affinity with the ligand

Models of these designs are shown below in two orientations. The orientation shown on the left was rotated 45° over the *x*-axis to attain the orientation shown on the right. AR: cyan, MBP: blue, FPP: magenta sticks, motif residues: sticks, designed residues: green spheres.

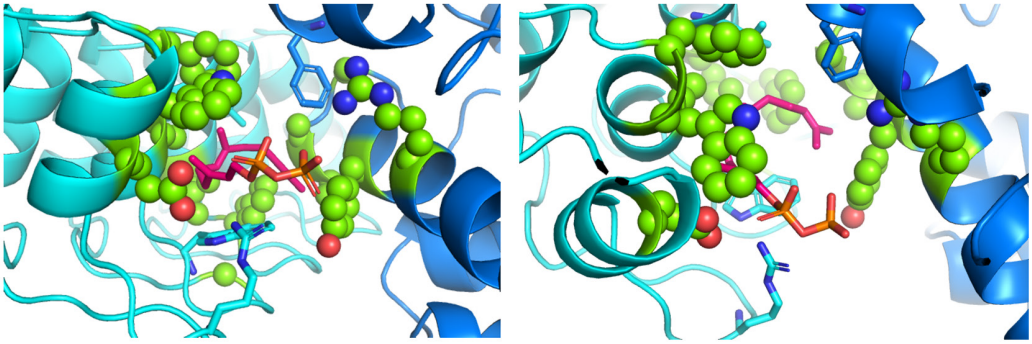
S3-1A



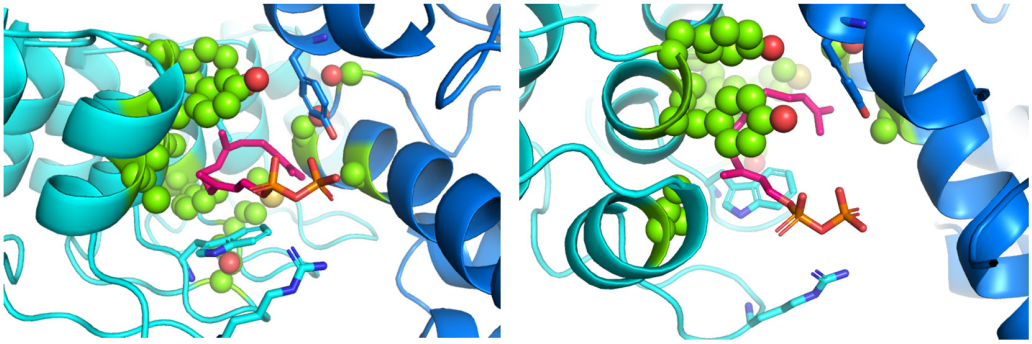
S3-1B



S3-1C



S3-1D



1.7 Computational methods that resulted in designs on scaffolds 1 and 2. Designs for scaffolds 1 and 2 resulted from the same overall protocol as described above, with the following modifications:

(i) We used expanded protein-protein interface scaffold sets consisting of 809 (for scaffold 1) or 3462 (for scaffold 2) heterodimers. Scaffolds were pre-relaxed using Rosetta FastRelax(*47*) with constraints to the starting coordinates for both backbone and side-chain atoms.

(ii) We used the following FPP binding site motifs:

Scaffold	**Template PDB**	**Motif residues**
Scaffold 1 FRB/FKBP (3FAP.pdb)	3bnx	F153, L184, W308, R314
Scaffold 2 RapF/ComA (3ULQ.pdb)	3bnx	F153, R314, Y315, F87

(iii) Optimization of the ligand orientation in the binding site used gradient-based minimization of the input structure before design; during this minimization, all designable residues that are not constrained are mutated to alanine and a reduced energy function that does not contain vdW_attractive or solvation terms (and is thus dominated by repulsive interactions) is used for minimization; during this minimization process, rigid body optimization of the ligand is performed (via setting jump edges between the ligand and protein residues to true when creating the enzdes movemap); the purpose of this stage is to optimize the motif-residue interactions with the ligand before design.

(iv) We did not use backbone ensembles to reshape the binding site environment around the motif residues built into the scaffold, but instead used Rosetta FastRelax to optimize both side chains and backbone in the binding site environment after the design step.

(v) We selected designs for testing based on the top predicted ligand binding score.

For scaffold 1, we selected one design (S1-1A) that contained the following wild-type reversions: A28G, T36F, W46F, H48F, M59W, L193Y determined by visual inspection.

For scaffold 2, we selected 4 designs:

S2-1A: computationally designed sequence.

S2-1B: computationally designed with wild-type reversions: V21I, M404L and mutations of designed alanine residues: A28G, A32S, A59R.

S2-1C: as S2-1B with 2 additional alanine mutations to destabilize the wild-type protein-protein interaction: D23A, N418A.

S2-1D: as S2-1A with 2 additional mutations: W12F, A411M.

(vi) Runs leading to the design on scaffold 1 used Rosetta version 44618 with the following command lines:

Building sites into scaffolds (matching):
```
match.linuxgccrelease
-database ROSETTA_DATABASE
-s INPUT_PDB
-match lig_name LG1
-match grid_boundary GRIDLIG_FILE
-match scaffold_active_site_residues POSFILE
-match geometric_constraint_file CST_FILE
-extra_res_fa FA_PARAMS
-output_matches_per_group 1
-ex1 -ex2 -extrachi_cutoff 0
-euclid_bin_size 1.75 -euler_bin_size 17.5 -bump_tolerance 0.525
-out path OUTDIR
-match output_format PDB
-match consolidate_matches
```

Design step:
```
enzyme_design.linuxgccrelease
-database ROSETTA_DATABASE
-enzdes:cstfile CONSTRAINTS_FILE
-extra_res_fa FA_PARAMS
-out:file:o design.score
-s INPUT_PDB
-enzdes:detect_design_interface
```

-enzdes:cst_design
-mute core.io.database
-mute core.scoring.hbonds
-overwrite
-out:pdb_gz
-enzdes:cut1 6.0
-enzdes:cut2 8.0
-enzdes:cut3 10.0
-enzdes:cut4 12.0
-enzdes:cst_opt
-enzdes:cst_min
-enzdes:bb_min
-enzdes:chi_min
-enzdes:design_min_cycles 3
-ex1
-ex2
-extrachi_cutoff 0
-use_input_sc
-enzdes:start_from_random_rb_conf
-enzdes:final_repack_without_ligand
-nstruct 10

Fastrelax of designs:
relax.linuxgccrelease
-database ROSETTA_DATABASE
-in:file:fullatom
-overwrite
-nstruct 1
-out:pdb_gz
-relax:fast
-constrain_relax_to_start_coords
-extra_res_fa FA_PARAMS
-in:file:s INPUT_PDB
-in:file:native INPUT_PDB

Rescoring relaxed designs:
enzyme_design.linuxgccrelease
-database ROSETTA_DATABASE
-s INPUT_PDB
-out:file:o design.rescore
-enzdes:detect_design_interface
-enzdes:cut1 0.0
-enzdes:cut2 0.0
-enzdes:cut3 10.0
-enzdes:cut4 12.0
-mute core.io.database

-mute core.scoring.hbonds
-out:pdb_gz
-ex1 -ex2 -extrachi_cutoff 0 -use_input_sc
-enzdes:no_unconstrained_repack
-enzdes:final_repack_without_ligand
-enzdes:enz_score
-extra_res_fa FA_PARAMS
-enzdes:cstfile CST_FILE

Runs leading to the designs on scaffold 2 used Rosetta version 49508 with the following

command lines:

Building sites into scaffolds (matching):
match.linuxgccrelease -database ROSETTA_DATABASE -s INPUT_PDB -match::lig_name LG1 -match::grid_boundary GRIDLIG_FILE -match::scaffold_active_site_residues POSFILE -match::geometric_constraint_file CST_FILE -extra_res_fa FA_PARAMS_FILE -output_matches_per_group 1 -match:consolidate_matches -out:file:scorefile -ex1 -ex2 -extrachi_cutoff 0 -use_input_sc -euclid_bin_size 1.5 -euler_bin_size 15 -bump_tolerance 0.5 -out::path OUTDIR -match:output_format PDB

Design step:
enzyme_design.linuxgccrelease -database ROSETTA_DATABASE -s INPUT_PDB -in:file:fullatom -out:file:o designs.score -extra_res_fa FA_PARAMS_FILE -enzdes:cstfile CONSTRAINTS_FILE -overwrite -out:pdb_gz -nstruct 10 -enzdes:cst_design -enzdes:detect_design_interface -enzdes:cut1 6.0 -enzdes:cut2 8.0 -enzdes:cut3 10.0 -enzdes:cut4 12.0 -enzdes:cst_opt -enzdes:cst_min -enzdes:bb_min -enzdes:chi_min -enzdes:design_min_cycles 3 -ex1 -ex2 -extrachi_cutoff 0 -use_input_sc -enzdes:start_from_random_rb_conf -enzdes:final_repack_without_ligand -score:weights enzdez.wts

Fastrelax of designs:
relax.linuxgccrelease -database ROSETTA_DATABASE -s INPUT_PDB -extra_res_fa FA_PARAMS_FILE -out:pdb_gz -out::path OUTDIR -ignore_zero_occupancy false -relax:fast -relax:constrain_relax_to_start_coords -ex1 -ex2 -extrachi_cutoff 0 -use_input_sc -score:weights enzdez.wts -preserve_header -nstruct 10

Rescoring relaxed designs:
enzyme_design.linuxgccrelease -database ROSETTA_DATABASE -s INPUT_PDB -in:file:fullatom -out:file:o designs.score -extra_res_fa FA_PARAMS_FILE -enzdes:cstfile CONSTRAINTS_FILE -overwrite -out:pdb_gz -enzdes:detect_design_interface -enzdes:cut1 0.0 -enzdes:cut2 0.0 -enzdes:cut3 10.0 -enzdes:cut4 12.0 -ex1 -ex2 -extrachi_cutoff 0 -use_input_sc -enzdes:no_unconstrained_repack -enzdes:lig_packer_weight 1.8 -enzdes:final_repack_without_ligand -score:weights enzdez.wts

1.8 Computational methods to further stabilize sensor S3-2D. Using the crystal structure for S3-2D as the input structure, we used the RosettaScripts platform to apply two successive cycles of coupled moves(*25*) to improve FPP-protein interactions and the binding pocket stability. The input structure was pre-relaxed using Rosetta FastRelax with constraints to the starting coordinates for both backbone and side-chain atoms. In each cycle of coupled moves, we chose a different group of neighboring residues in the binding pocket to be designed. The second cycle received the lowest-energy structure from the previous cycle as the input structure. Results were ranked according to total energy. Structures and sequences of top-ranked designs were manually inspected to choose which designs to experimentally test.

Coupled moves cycle	Designed protein	Designable residues
1	AR of S3-2D (AR-2.7)	80-82, 84-86, 88-90, 118-122, 141-145, 150-155
2	MBP of S3-2D (MBP-2.5)	132-134, 196-202, 250-252

Rosetta commands (Rosetta 3.8, commit f3a3f038d9419c0d79a6030ea4dc16f62668d69f):

FastRelax of crystal structure:
relax.linuxgccrelease -database ROSETTA_DATABASE -in:file:s INPUT_PDB -in:file:fullatom -relax:constrain_relax_to_start_coords -extra_res_fa FA_PARAMS_FILE

Design flags:
-in:file:s INPUT_PDB
-extra_res_fa FA_PARAMS_FILE
-packing
    -ex1
    -ex1aro
    -extrachi_cutoff 0
    -ex2
-number_ligands 1
-coupled_moves
    -ntrials 10000
    -initial_repack false
    -min_pack true
    -ligand_mode true
    -ligand_weight 2.0

-nstruct 20

RosettaScript to run coupled moves:

```
<ROSETTASCRIPTS>
  <SCOREFXNS>
    <ScoreFunction name="ref15"/>
  </SCOREFXNS>

  <TASKOPERATIONS>
    <ReadResfile name="AR_resfile" filename=RESFILE_NAME_1/>
    <ReadResfile name="MBP_resfile" filename= RESFILE_NAME_2/>
  </TASKOPERATIONS>

  <MOVERS>
    <CoupledMovesProtocol name="coupled_moves_AR" task_operations="AR_resfile"/>
    <CoupledMovesProtocol name="coupled_moves_MBP" task_operations="MBP_resfile"/>
    <DumpPdb name="dump" fname="dump" tag_time="1"/>
    <FastRelax name="fastrelax" repeats="5"/>
  </MOVERS>

  <PROTOCOLS>
    <Add mover_name="coupled_moves_AR"/>
    <Add mover_name="dump"/>
    <Add mover_name="coupled_moves_MBP"/>
  </PROTOCOLS>
</ROSETTASCRIPTS>
```

2. Complementation assay with split murine dihydrofolate reductase (mDHFR).

2.1 Constructs and strains. We tested sensor function in *E. coli* strain DH10B using complementation of split mDHFR(*30*). The constructs used in this assay are listed in **Appendix 1** and the gene sequences are listed in **Appendix 4**. For all MBP genes tested in the mDHFR assay, we deleted the sequence corresponding to the last alpha helix in the structure (residues 354-370). This modification was made to bring the two sensor protein termini closer together for facile complementation of the attached split mDHFR constructs upon small-molecule mediated dimerization. A detailed protocol of the assay is given below.

2.2 Liquid culture assay. Fresh preparations were made of filter-sterilized 0.2 M isopropyl β-D-1-thiogalactopyranoside (IPTG) in sterile deionized water, 20% L-arabinose in sterile deionized water, 5 mg/ml trimethoprim (TMP) (Sigma) solution in methanol, and autoclaved M9 medium (0.1 mM $CaCl_2$, 1X M9 salts, 2 mM $MgSO_4$, 0.4% glucose, 0.2% Casaminoacids, 10 mg/ml leucine, 10 mg/ml isoleucine, 5 mg/ml thiamine, all filter-sterilized). One 4 ml overnight culture was grown at 37 °C and 225 rpm for each strain from freshly transformed plates in M9 medium with 37 μg/ml chloramphenicol and 50 μg/ml spectinomycin. To prepare media for the assay, the following was added to 49 ml M9 medium: 37 μg/ml chloramphenicol and 50 μg/ml spectinomycin, 60 μM IPTG (unless otherwise noted), 0.4% L-arabinose, and 6 μg/ml TMP (unless otherwise noted). All reagents were thawed completely, vortexed, and centrifuged prior to addition to the master mix to avoid precipitants. All ingredients were gently mixed in a Falcon tube, and then 4 ml of the master mix was aliquoted into new culture tubes. Each overnight culture was gently vortexed, checked for complete resolubilization of any chunks, diluted 1 to 100 in fresh

medium, and poured into a reagent reservoir. Using a multichannel pipette, 200 µl of the diluted cultures were transferred to the wells of a transparent, flat-bottom 96-well plate (Costar), 8 wells at a time. A final concentration of 5 mM mevalonate (Sigma) (unless otherwise noted) was added to every other column via 1 to 1000 dilution of a 1 M stock solution in Millipore water. To prevent evaporation of medium, 50 µl mineral oil was gently pipetted on top of each well's contents. The plate was centrifuged for 30 seconds to pop bubbles and any remaining bubbles were popped manually with a pipette tip. The plate was placed in a shaker and grown at 35 °C and 200 rpm for 48 hours. After 48 hours of growth, cultures were mixed gently using a multichannel pipette set to 50 µl, avoiding disturbance of the mineral oil layer. Bubbles were removed using a Bunsen burner flame and/or pipette tips. Cell culture optical density (OD) was read over the whole plate, without a plate cap, at 600 nm using a Tecan Safire 2 microplate reader.

2.3 Library selection/screening and saturation mutagenesis

2.3.1 Library construction. The computationally designed library (library 1, **Table S2**) was constructed using gene assembly mutagenesis(*48*). Briefly, multiple oligonucleotides were synthesized to assemble the design regions of AR and MBP, and 14 and 16 oligonucleotides were generated to cover the protein-protein interface of AR and MBP respectively. Computationally predicted mutations within the regions were introduced into the designed oligonucleotides using degenerate codons that code for a list of desired amino acids (**Table S2**). These synthesized oligonucleotides were combined and assembled using a two-step PCR to form the library of AR and MBP fragments, and both sets of the fragments were cloned sequentially into pCDFDuet to

obtain the final library constructs. We also constructed an additional library using error-prone PCR starting from design S3-1C (library 2).

2.3.2 Library selection and screening. The library constructs were electroporated into DH10B *E. coli* and plated onto M9 minimal medium plates containing 100 μM IPTG, 0.4 % L-arabinose, 5 mM mevalonate and 6 μg/ml TMP. The plates were cultured for 10 days at room temperature. Clones that survived under TMP selection were potential active candidates that could form functional mDHFR and were enriched in this selection.

We performed selection on ~3 x 10^5 members of each library (limited by transformation efficiency), which led to the isolation of 1536 active clones from each library after enrichment for growth via split mDHFR complementation. The selected clones were then subjected to an array-based colony-printing assay (described below, Fig. S3). From this assay, we identified 36 hits from which we confirmed 27 unique sequences (7 from library 1 and 20 from library 2) by individual growth assays (Fig. S4, Fig. S5).

In the colony-printing assay, freshly grown overnight cultures of each selected clone were 100X diluted and spotted onto two sets of M9 minimal medium plates, one with 5 mM mevalonate and one without. Both sets of the plates contained 100 μM IPTG, 0.4 % L-arabinose and 6 μg/ml TMP. The plates were cultured for 10 days at room temperature. Clones that showed increased growth (as indicated by the size of the colonies) with the addition of mevalonate after 10 days were selected and individually tested using the liquid culture assay as described above. The same conditions were kept in the liquid culture screening.

2.3.3 Saturation mutagenesis. We performed two rounds of iterative saturation mutagenesis (ISM(*49*), **Table S3**). For each round, 11 positions (AR: position 85, 112, 118, 119, 122, 123, 152, 155; MBP: 194, 197, 251) around the designed binding site were allowed to mutate to any of 20 amino acids coded by NNK degenerate codon starting from S3-2A. Only one of 11 positions was allowed to mutate at a time (a.k.a. single-site saturation). The best clone from the first round of ISM (S3-2B) was used as the starting sequence for the second round of ISM. A total of 384 clones and 480 clones were screened in the colony printing assay in the presence and absence of FPP in the first and second round of ISM respectively. The top clones for each round were screened using the liquid culture assay as described above using 80 μM IPTG, 0.4 % L-arabinose, 5 mM mevalonate and 6 μg/ml TMP.

3. Protein purification

3.1 Constructs. We expressed and purified several S3 sensor proteins for further characterization. The expression plasmids are listed in **Appendix 2** and the gene sequences are listed in **Appendix 4**. For all MBP genes in expression plasmids, we included the sequence corresponding to the last alpha helix in the structure (residues 354-370).

3.2 Purification of ankyrin repeat (AR) proteins. Overnight cultures of BL21(DE3) pLysS cells (Novagen) harboring the pET47-6XHis-AR plasmid were grown with 37 μg/ml chloramphenicol and 50 μg/ml kanamycin for 14-18 hours at 37 °C and 225 rpm in lysogeny broth (LB) medium. Overnight cultures were diluted 1 to 50 in LB and grown at 37 °C and 225 rpm until the OD_{600} reached approximately 0.6 (about two hours). Cultures were induced with 0.5 mM IPTG and

further grown at 16 °C overnight. Cells were pelleted by centrifugation at 6,000 *g* for 20 minutes. Pellets from 1 L cell cultures were stored at -80 °C. Frozen cell pellets were thawed on ice for 20 minutes and resuspended by vortexing in 40 ml lysis buffer (25 mM Tris, 500 mM NaCl, pH 8.0). The following components were added to the cell resuspension: 5 mM $MgCl_2$, 1 mM $MnCl_2$, 100 µM $CaCl_2$, one protease inhibitor tablet (Thermo Fisher), and 5 µL 2000 units/ml DNaseI per ml lysis buffer. Cells were lysed by the addition of 10X BugBuster (Thermo Fisher) to a final concentration of 1X, mixed gently, and stored at room temperature for 5 minutes. Cell lysate was centrifuged at 27,000 *g* for 30 minutes at 4 °C. The supernatant was decanted into a fresh tube, and imidazole was added to 20 mM. The soluble lysate was further incubated on ice for 1 hour to allow any remaining precipitate to form, then centrifuged at 3,700 *g* for 10 minutes at 4 °C. The supernatant was transferred to a fresh tube, mixed with 3 ml Ni-NTA Slurry (Thermo Fisher) per 40 ml lysate, and mixed at 4 °C on a nutator for 1 hour. The lysate-slurry mix was centrifuged at 500 *g* for 5 minutes at 4 °C and the supernatant was immediately discarded. The pelleted resin was resuspended in two resin-volumes of cold wash buffer (25 mM Tris, 500 mM NaCl, 20 mM imidazole, pH 8.0) and applied to a Ni^{2+}-NTA gravity column, then the column was washed with two column volumes of cold wash buffer. Protein was eluted with 1.5 column volumes in 25 mM Tris, 500 mM NaCl, 250 mM imidazole, pH 8.0. The eluate was desalted into 25 mM Tris, 50 mM NaCl, pH 8.0 using an AKTA FPLC with a Hiprep 26/10 desalting column, then ion-exchanged using a Hitrap Q HP column into 25 mM Tris, 500 mM NaCl, pH 8.0 The resulting protein solution was concentrated down to 5 ml using 3 KDa MW cutoff spin concentrators (Sartorius), applied to a size-exclusion HiLoad 16/60 Superdex 75 column, and eluted in 50 mM Tris, 150 mM NaCl, pH 8.0. The eluate was concentrated to 1.5 ml and dialyzed three times against 1 L 25 mM Tris, 150 mM NaCl, pH 8.0, for two hours, overnight, and two hours, respectively. The final protein

concentration was determined in the denatured state in 6 M guanidinium hydrochloride, 20 mM potassium phosphate solution and the purity was confirmed by Coomassie-stained SDS-PAGE to be > 95%.

3.3 Purification of maltose binding proteins (MBP). Methods for the growth and lysis of BL21(DE3) pLysS cells with pET28-MBP plasmids were similar to those described above for AR proteins, with the following modifications: (1) overnight cultures were subcultured with 0.2% glucose, and (2) lysis buffer was composed of 25 mM Tris, 150 mM NaCl, pH 8.0. Isolation of the soluble lysate was followed by incubation on ice for 1 hour, then centrifugation at 3,700 *g* for 10 minutes at 4 °C. The supernatant was decanted into a fresh tube, applied to an MBP affinity column (MBPtrap HP column) using an AKTA FPLC, and eluted in 50 mM Tris, 150 mM NaCl, 30 mM maltose, pH 8.0. The volume of the protein solution was reduced to 5 ml using 10 KDa MW cutoff spin concentrators (Millipore) and applied to a size-exclusion HiLoad 16/60 Superdex 75 column, and eluted in 50 mM Tris, 150 mM NaCl, pH 8.0. The size exclusion eluate volume was concentrated down to 1.5 ml and dialyzed three times against 1 L 25 mM Tris, 150 mM NaCl, pH 8.0, for two hours, overnight, and two hours, respectively. The final protein concentration was determined in the denatured state in 6 M guanidinium hydrochloride, 20 mM potassium phosphate solution and the purity was confirmed by Coomassie-stained SDS-PAGE to be >98%.

4. *In vitro* binding assays using bio-layer interferometry (BLI)

4.1 Constructs. We purified several S3 sensor protein pairs to determine their *in vitro* binding affinity in the presence and absence of FPP by BLI. In our BLI experiments, one protein is tethered to an optically transparent biosensor tip by a biotin-streptavidin interaction, and the other protein is present as the analyte in solution in the microplate. BLI expression plasmids are listed in **Appendix 2** and the gene sequences are listed in **Appendix 4**. Constructs for BLI experiments included the sequence corresponding to the last alpha helix in MBP (residues 354-370).

4.2 Biotinylation of avi-MBP. Avi-tagged MBP (**Appendix 2**) was prepared at 66 μM in 10 mM Tris buffer, pH 8.0. The BirA-500 biotin-protein ligase kit from Avidity was used as per the manufacturer's instructions, with a final avi-MBP concentration of 40 μM and 2.5 μg BirA enzyme per 10 nmol protein. The reaction was performed at 30 °C for 45 minutes, and then buffer-exchanged by spin concentration and 50-fold dilution three times in fresh 25 mM Tris, 150 mM NaCl buffer, pH 8.0, to remove excess biotin and reaction components. Biotinylated avi-MBP was diluted to 40 μM and stored at 4 °C for use in BLI measurements.

4.3 BLI. Affinity measurements between avi-tagged MBP, AR and FPP were performed at room temperature using an Octet RED96 system and streptavidin (SA)-coated biosensor tips (Pall ForteBio). Avi-MBP was diluted to 400 nM in HBS-P buffer (0.01 M HEPES pH 7.4, 0.15 M NaCl, 0.005% v/v Surfactant P20) to be used as the antigen. Antigen-bound SA-tips were washed, exposed to the analyte solutions during an association period, and then allowed to unbind from the analyte during a dissociation period. For experiments including FPP, the analyte solution was

composed of 0-24 μM AR and 200 μM FPP. For experiments measuring the affinity of MBP for AR without FPP, the analyte solution was composed of 0-500 μM AR. The binding protocol was as follows: rinse tips in HBS-P buffer, 60 seconds; load tips with antigen, 300 seconds; establish baseline by rinsing tips in HBS-P buffer, 60 seconds; association with analyte, 60 seconds; dissociation in baseline wells, 600 seconds. Raw data was fit to 1:1 binding curves in Octet Data Analysis HT software using curve fitting kinetic analysis with local fitting. The theoretical equilibrium binding signal response data (R equilibrium) were normalized by the steady-state group maximum response (R_{max}) values, and the steady state affinity was determined using the Hill equation,

$$\theta = \frac{1}{\frac{K_A}{[L]}+1},$$

where θ is the fraction of MBP that is bound to AR, K_A is the AR concentration at which half the MBP are occupied; and $[L]$ is the concentration of unbound AR. Non-cooperative binding kinetics are assumed. All binding curves fit the BLI data with R^2>0.90.

5. Cell-free transcription-translation (TxTl) assay

5.1 Constructs. Plasmids for TxTl experiments were constructed using standard BsaI/BsmBI Golden Gate cloning methods as described in Engler *et al.*(*50*) using our group's adaptation of the MoClo Yeast Tool Kit strategy as described in Lee *et al.*(*51*). For all MBP genes in TxTl plasmids, we included the sequence corresponding to the last alpha helix in the structure (residues 354-370). A comprehensive list of plasmids is available in **Appendix 3**.

5.2 TxTl protocol. Separate 30 µl TxTl reactions were prepared for each plasmid (expressing MBP or AR variants) as described in Sun *et al.*(*52*) and incubated for 8-16 hours at 29 °C. Each reaction also included 0.2 nM of the TxTl11 plasmid for RNA polymerase expression (**Appendix 3**) and 1 mM IPTG. The cell extract was prepared from Rosetta 2 cells (EMD Millipore). All TxTl reactions were mixed 1 to 1 with phosphate buffered saline/1% bovine serum albumin (PBS/BSA solution) for a total protein solution volume of 60 µl. A master mix was prepared by mixing 1 volume TxTl extract expressing an MBP variant, 1 volume TxTl extract expressing an AR variant, and 2 volumes PBS/BSA solution. Master mix aliquots of 27 µl were mixed with 3 µl farnesyl pyrophosphate (FPP) (Sigma) stock solutions in water with concentrations ranging from 1 µM to 2 mM. These 30 µl solutions were distributed in 9.5 µl aliquots in a white opaque 384 well plate for luminescence measurements or a black opaque 384 well plate for fluorescence measurements, centrifuged for 30 seconds to remove bubbles, and stored at room temperature for 10 minutes. For luminescence measurements, 1 ml plus 10 µl per well of Promega NanoGlo buffer was thawed on ice and mixed with a 50-fold dilution of furimazine substrate to make NanoGlo reagent, as described in Dixon *et al.*(*34*). The reagent was distributed in 9.5 µl aliquots to all wells containing

9.5 µl protein/FPP sample. After ten minutes, a SpectraMax L luminometer (Molecular Devices) was used to take luminescence readings using analog detection and injection, following priming of the injection lines with 1 ml NanoGlo reagent. Data were collected with the following machine settings: integration time, 1 s; PMT setting, photon counting; automix setting, classic with 5 s mix duration and 30 mm/s mix speed; M-injection, with injector volume 10 µl, post-injection delay 1 s; injection speed 320 µl/s, and no shaking after injection; and dark adapt set to off. For fluorescence measurements, a Tecan Safire 2 microplate reader was used to read ddGFP complementation from the top of the plate at excitation/emission wavelengths of 380 nm and 508 nm, with the gain set to 100.

5.3 Data fitting for *in vitro* FPP titration measurements. Raw background-subtracted luminescence and fluorescence data from TxTl assay measurements were averaged for each FPP concentration and fit to a modified three-parameter logistic nonlinear regression model of the Hill equation,

$$\hat{Y} = a + \frac{b-a}{1+\frac{c}{x}},$$

where $\hat{Y}$ is the expected response at FPP concentration x, a is the minimum response when no FPP is present, b is the maximum response, and c is the FPP concentration at which half the protein sensors are bound, allowing for a fluorescent or luminescent output. Data and fit were normalized by subtracting the calculated minimum response and dividing by the difference of the calculated maximum and minimum responses.

6. Crystallography.

6.1 Constructs. The plasmids used to express S3-2D sensor proteins for crystallography are listed in **Appendix 2**. Their sequences are listed in **Appendix 4**. Methods for protein expression and purification are described in Section 3.

6.2 Crystallization. AR and MBP were each prepared at 340 μM and mixed 1:1 to obtain each protein at 170 μM in solution. FPP was prepared at 170 mM in 70% ethanol and diluted 100-fold in the protein solution for a final concentration of 1.7 mM FPP. We carried out initial crystallization trials in 15-well hanging drop format using EasyXtal crystallization plates (Qiagen) and a crystallization screen that was designed to explore the chemical space around the crystallization conditions reported by Binz *et al.* (*29*) for the co-crystal structure of their AR-MBP complex (PDBID:1SVX). Crystallization drops were prepared by mixing 1 μl of protein solution with 1 μl of the mother liquor, and sealing the drop inside a reservoir containing an additional 500 μl of the mother liquor solution. Thin, plate-like crystals were obtained in many of the conditions tested, with the largest single crystals growing from drops prepared using mother liquor containing 0.1 M Tris Buffer pH 8.7, 0.1 M sodium chloride, and 32% PEG-6000. The crystals could not be obtained without the addition of the FPP ligand, and we used SDS-PAGE to confirm that both proteins required to form the heterodimer were present in the crystals (after harvesting and washing in mother liquor without protein).

6.3 X-ray data collection and processing. Prior to X-ray data collection, crystals were dehydrated overnight in a solution containing 0.1 M Tris Buffer pH 8.7, 0.1 M sodium chloride, and 45%

PEG-6000. Next, the crystals were soaked in a solution containing 0.1 M Tris Buffer pH 8.7, 0.1 M sodium chloride, 42% PEG-6000, 7% ethanol, and 18 mM FPP for 30 minutes before they were harvested and flash-cooled in liquid nitrogen. The soaking steps were essential to improve the quality of the observed X-ray diffraction patterns, and to ensure adequate occupancy of the FPP ligand in the resulting electron density maps. No cryoprotectant was necessary due to the high concentration of PEG-6000 in the soaking solutions.

We collected single-crystal X-ray diffraction data on beamline 8.3.1 at the Advanced Light Source. The beamline was equipped with an ADSC Quantum 315r CCD detector, and the crystals were maintained at a cryogenic temperature (100 K) throughout the course of data collection.

We processed the X-ray data using the Xia2 system(*53*), which performed indexing, integration, and scaling with XDS and XSCALE(*54*), followed by merging with Pointless(*55*). A resolution cutoff (2.20 Å) was taken where the completeness of the data fell to a value of approximately 90%. Although other metrics of data quality (such as CC1/2 and $\langle I/\sigma I \rangle$) suggest that a more aggressive resolution cutoff would be acceptable, we were limited by physical constraints on the experimental geometry. Specifically, the plate-like morphology of the crystals resulted in the long axis of the unit cell always being perpendicular to the crystal rotation (phi) axis, which constrained the minimum sample-to-detector distance and the maximum Bragg angle that could be recorded on the detector without producing overlap between individual reflections. Further information regarding data collection and processing is presented in **Table S4**. The reduced diffraction data were analyzed with phenix.xtriage (http://www.ccp4.ac.uk/newsletters/newsletter43/articles/PHZ_RWGK_PDA.pdf) to check for

crystal pathologies , which revealed the presence of pseudomerohedral twinning based on the results of the L-test (<|L|>=0.399, <L2>=0.225)(*56*). Pseudomerohedral twinning is uncommon, however it has been reported previously in the literature(*57*) and is not unprecedented in macromolecular crystallography(*58-61*).

6.4 Structure determination. We obtained initial phase information for calculation of electron density maps by molecular replacement using the program Phaser(*62*), as implemented in the PHENIX suite(*63*). We searched for each component of the heterodimer independently, using separate models of AR (PDBID: 1SVX, chain A) and the "closed" (ligand-bound) form of MBP (PDBID: 1FDQ). We explicitly did not use the design model for molecular replacement to avoid the introduction of model bias. Two complete copies of the heterodimer were found in the unit cell, consistent with an analysis of Matthews probabilities for the observed unit cell and molecular weight of the heterodimer(*64, 65*).

Next, we attempted to rebuild the missing or incorrect parts of the molecular replacement solution using the electron-density maps calculated from model phases, however, the presence of twinning compromised the quality of the initial maps. To improve the interpretability of the electron density maps, we carried out a round of atomic refinement using phenix.refine(*66*) in which we also refined the twin fraction (twin operator = *h,-k,-l*; refined twin fraction, $\boldsymbol{\alpha}$ = 0.49). The twin refinement improved the map quality substantially, allowing us to rebuild the missing or incorrect parts of the structure. Additional iterative steps of manual model rebuilding and atomic refinement were performed, and during this process we found evidence for several small molecule ligands in both $2mF_o$-DF_c and mF_o-DF_c electron density maps. Specifically, we identified an FPP molecule

occupying one of the computationally-designed binding sites, as well as two maltose disaccharides (one bound to each copy of MBP). The presence of merohedral twinning can present challenges for map interpretation because the structure factor detwinning process can exacerbate the effects of model phase bias(*57*). Consequently, we took extra care with the placement of the FPP ligand into electron density features around the designed binding site. The initial placement of the ligand was motivated by the observation of several strong electron density peaks near a key lysine residue and in the hydrophobic binding cavity, which could be attributed to the pyrophosphate group and aliphatic tail of FPP, respectively. Refinement of the structure with FPP in the binding site resulted in a rearrangement of side chains in contact with the ligand, providing additional evidence that this binding site is occupied. Specifically, Trp114 of AR becomes ordered upon FPP binding, and Y197 in MBP rotates and displaces a small network of two neighboring water molecules (**Figure S11**). After adding the ligands, we performed additional refinement of atomic positions, and individual atomic displacement parameters (B-factors), and occupancies using non-crystallographic symmetry (NCS) and secondary structure restraints, a riding hydrogen model, and automatic weight optimization. The occupancy of the FPP ligand was set to 1.0 during the refinement, so that the heterogeneity of the binding site would be captured entirely by the B-factors. Following refinement, the atomic B-factors of the FPP ligand were similar to those of the surrounding protein atoms, suggesting that the binding site was relatively well occupied and that the ligand is modeled in a reasonable pose. Furthermore, there are no abrupt variations in the B-factors of atoms that are nearby in space, an additional indication that the ligand is appropriately placed. The average B factor of the ligand is 35.5, which is slightly, but not unusually higher than surrounding atoms: The average B factor of atoms on the AR (chain D) within 5 Å of any ligand atom is 29.3, within 10 Å is 29.1, within 15 Å is 29.1. The average B factor of atoms on the MBP (chain B) within 5 Å

is 23.4, within 10 Å is 20.1, within 15 Å is 19.1. After refining the model to convergence, we further verified the presence of the FPP ligand in the model by repeating our final refinement starting from atomic coordinates without the ligand atoms. This "omit refinement" produced $2mF_o$-DF_c and mF_o-DF_c omit maps and allowed us to assess the effect of including the ligand. The omit maps showed density that was highly similar to the original maps that permitted initial placement of FPP (**Figure S10**). The absence of FPP in the second ligand binding site could be the result of distortions to the heterodimeric interface caused by crystallization, or by the dehydration procedure that was required to improve the diffraction resolution. All model building was performed using Coot(*67*) and refinement steps were performed with phenix.refine (v1.13-2998) within the PHENIX suite(*63, 66*). Restraints for the maltose and FPP ligands were calculated using phenix.elbow(*68*). The final model coordinates were deposited in the Protein Data Bank (PDB(*69*)) under accession code 6OB5. Further information regarding model building and refinement is presented in **Table S4**.

Supplementary Text

There are several potential limitations that need to be considered when engineering and applying CID systems:

First, the difference in affinity of the two sensor proteins in the presence and absence of the small molecule should be large to maximize the achievable sensor amplitude (see also **Fig. S14**). It is hence advantageous for computational and experimental engineering approaches to select for both stabilization of the ternary complex in the presence of the small molecule and destabilization of the protein-protein heterodimer complex in the absence of the small molecule. In our case, we experimentally screened sensor candidates in both the presence and absence of the small molecule. Our initial computational design protocol selected for designs with large favorable ligand binding energies, and the design selected by most favorable ligand score (S3-1D) showed the largest sensor signal in the *in vivo* DHFR assay (**Fig. 2C**). *In vitro* affinity measurements on sensors S3-2D and S3-3A (**Fig. 3B,C**) showed a large destabilization of the protein-protein heterodimer complex in the absence of FPP (45,000-fold destabilized compared to the original AR-MBP scaffold). However, in our subsequent optimization of the design based on the S3-2D crystal structure, we only optimized for ternary complex stability but did not simultaneously optimize for dimer complex destabilization. The resulting mutations stabilized both the ternary and the dimer complexes, effectively reducing the dynamic range of the sensor. Negative design approaches(*32*) could be incorporated to destabilize the dimer complex in the simulation.

Second, there are several failure modes for computational designs, including (i) issues with properly reconstituting the split reporter, (ii) failure of the designs to fold or express at sufficient concentrations, or (iii) insufficient binding to the small molecule ligand across the complex interface (see also **Fig. S2**).

Third, there are considerations important to the application of CID systems. For example, we imagine two principal applications for an FPP sensor. The first would be to use FPP detection in directed evolution experiments to optimize expression or activity of enzymes in the pathway producing FPP. The second more complex application(*21*) would be to dynamically regulate the activity of enzymes in the FPP pathway to keep FPP accumulation below toxic levels. For both cases, but particularly the latter, a DHFR-based screen using growth as the output would not be optimal as growth conditions in an actual application would be vastly different; however, we show that different modular outputs that are not linked to cell survival can be coupled to our sensors (**Fig. 3G, H**). Moreover, in an application the sensitivity of a sensor would have to match FPP production levels; a family of FPP sensors with different sensitivities could be used. Finally, CID systems in general could be subject to combinatorial inhibition where the sensor signal is diminished at high ligand concentrations, because the ligand forms dimeric complexes with each of the sensor halves reducing the effective concentration of ternary complexes. In our case, however, binding of FPP to the designed MBP component of S3-2D alone was not detectable up to at least 200 μM.

Table S1. Summary of computational designs. The design ID lists scaffold first (S1, S2, S3), followed by design round (1, 2, 3), followed by a letter (A, B, etc). The nomenclature for the individual sensor proteins lists protein, design round, and a consecutive number denoting the protein's variant.

Scaffold (WT PDB)	Design ID	Sensor Protein A	Sensor Protein B	Design / Engineering Method
Scaffold 1: FKBP12 -FRB (3FAP.pdb)	S1-1A	FKBP12-1.1	FRB-1.1	Computational Design
Scaffold 2: RapF-ComA (3ULQ.pdb)	S2-1A	RapF-1.1	ComA-1.1	Computational Design
	S2-1B	RapF-1.2	ComA-1.2	Computational Design
	S2-1C	RapF-1.3	ComA-1.3	Computational Design
	S2-1D	RapF-1.4	ComA-1.4	Computational Design
Scaffold 3: AR-MBP (1SVX.pdb)	**Computational designs (group 1)**			
	S3-1A	AR-1.1	MBP-1.1	Computational Design, best ligand burial
	S3-1B	AR-1.2	MBP-1.2	Computational Design, consensus
	S3-1C	AR-1.3	MBP-1.3	Computational Design, optimized ligand packing
	S3-1D	AR-1.4	MBP-1.4	Computational Design, best ligand score
	Computational designs with stability enhancing mutations (group 2)			
	S3-2A	AR-2.5	MBP-1.3	Hit from error-prone PCR (identical to design S3-1C except for 2 mutations in AR: N102H, A104E)
	S3-2B	AR-2.5	MBP-2.5	S3-2A with stability enhancing mutation from saturation mutagenesis in MBP: R194A
	S3-2C	AR-2.6	MBP-2.5	S3-2A with stability enhancing mutations from saturation mutagenesis: MBP: R194A AR: L85G
	S3-2D	AR-2.7	MBP-2.5	S3-2C with published stabilizing mutations in AR*: W122K, E152L, A149P, S153A, L161I, I164V, L168A, N169A
	Computational designs with designed affinity enhancing mutations (group 3)			
	S3-3A	AR-2.7	MBP-3.6	S3-2D with affinity enhancing mutation in MBP from KIC ensemble design: Y197A
	S3-3B	AR-3.8	MBP-2.5	S3-2D with affinity enhancing mutations in AR from coupled moves: R145K, K147L, D155L
	S3-3C	AR-3.8	MBP-3.6	S3-2D with affinity enhancing mutations in AR and MBP: AR: R145K, K147L, D155L MBP: Y197A

* from reference (*31*)

Table S2. Computationally designed library. We designed a sequence library with 2.4×10^6 members using degenerate codons (top) based on the dominant residues predicted computationally (bottom). The library was tested using the plate-based split mDHFR assay. Hits were defined as sensor sequences that grew better in the presence of mevalonate. Residues are shown with motif positions in red and wild-type residues shaded grey. Residues with minor frequencies in the flexible design predictions are shown as lower-case letters.

Sequence position	81	85	89	114	119	122	123	145	152	133	194	197	200	201	203
Protein	AR									MBP					
Wild-type residue	Y	H	Y	M	L	K	W	F	I	P	F	D	K	N	H
Library codon	WTK	WTT	WTK	TGG	GCG	KHT	BHK	CGT	RYT	YWT	-	GVC	RMG	KHT	-
Library residues (plus F79A)	F I L M	F I	F I L M	W	A	A D F S V Y	F Y A S V D L P H Q E	R	A I T V	F H L Y	-	A D G	A E K T	A D F S V Y	-
Residues in library hits	I L M	F I	F M	W	A	D S V Y	Y A V L E	R	I T V	F Y H	-	D A G	K E A	N A D	-

Computational predictions

Sequence position	81	85	89	114	119	122	123	145	152	133	194	197	200	201	203
Protein	AR									MBP					
Computational Design (Top-ranked individual Sequences) (+99: D,G,T)	F I L M	F I L	L	W	A	Y W	F Y	R	I V E	F Y	F R	A Y	A K	A S	A H S
Computational Predictions (KIC Ensemble)	a F i L M	a F I m	L	W	A	A F l Y	A e F h q Y	R	A V	F Y	-	A G	A	A L F Y	-
Computational Predictions (coupled moves)	F	F H L M	L M	A F H V W	A	F W Y	F M Y	E R N T	H F Y	F Y	F W	A F H Y	A	A S	-

Table S3: Single-site saturation mutagenesis positions and results.

Sequence position	85	112	118	119	122	123	152	155	194	197	251
Protein	AR								MBP		
Wild-type residue	H	D	H	L	K	W	I	D	F	D	K
Residues in library hits	F I	Not varied in comp. library design	Not varied in comp. library design	A	D S V Y	Y A V L E	I T V	Not varied in comp. library design	Not varied in comp. library design	D A G	Not varied in comp. library design
Residue in design S3-1D	F	D	H	A	Y	Y	V	D	F	A	K
Residue in design S3-1C	L	D	H	A	W	F	E	D	R	Y	K
Residue in design S3-2A	L	D	H	A	W	F	E	D	R	Y	K
Saturation mutagenesis hits round 1, starting from S3-2A	-	-	L N P	-	V L	-	-	-	A G F	-	-
Best from combination of hits round 1 (design S3-2B)	L	D	H	A	W	F	E	D	A	Y	K
Saturation mutagenesis hits round 2, starting from S3-2B	G	-	T I	P R	-	-	-	-	-	-	-
Best from round 2 (design S3-2C)	G	D	H	A	W	F	E	D	A	Y	K

Table S4: X-ray data reduction and model refinement.

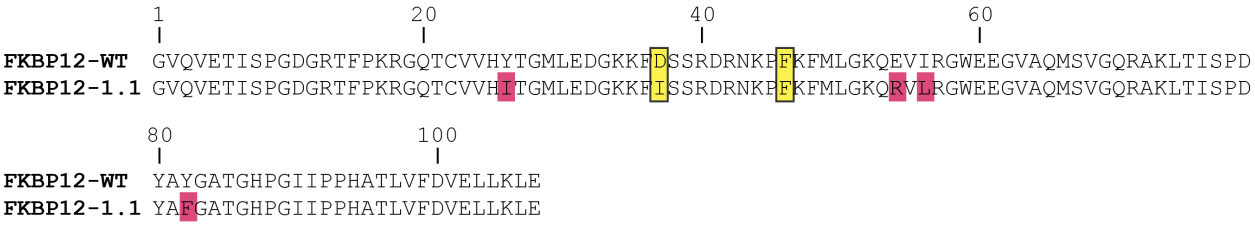
Wavelength	1.116Å
Resolution Range	95.47-2.20 (2.24-2.20)
Unit Cell	a=44.57Å , b=190.92Å , c=55.48Å **α**=**β**=***γ***=90°
Space Group	$P2_1$
Unique Reflections	45972 (7746)
Multiplicity	3.7 (3.6)
Completeness	98.5% (91.5%)
<I/σI>	17.5 (8.3)
$CC_{1/2}$[a]	0.997 (0.966)
R_{pim}[b]	0.033 (0.090)
R_{work}[c]	0.2071 (0.3448)
R_{free}[c]	0.2596 (0.4761)
Total Refined Atoms	8115
Protein Residues	1022
Solvent Molecules	141
Refined Ligand Atoms	70
Average B-factor	21.6Å^2
$RMSD_{bonds}$	0.003Å
$RMSD_{angles}$	0.56°
Rama. Plot:	
Favored	97.1%
Allowed	2.9%
Outliers	0.0%
Molprobity Clashscore[d]	3.97
PDB ID	6OB5

a. Reference (*70*)
b. Reference (*53*)
c. Reference (*71*)
d. Reference (*72*)

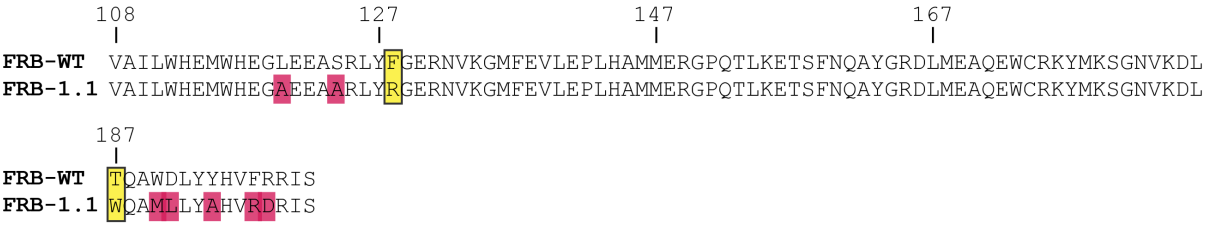
Figure S1

Scaffold 1:

FKBP-12

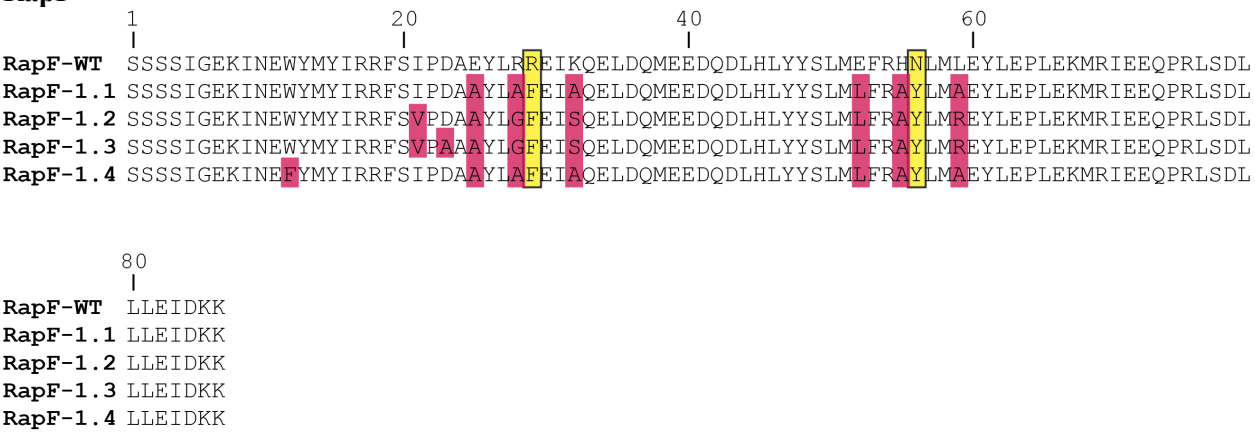


FRB

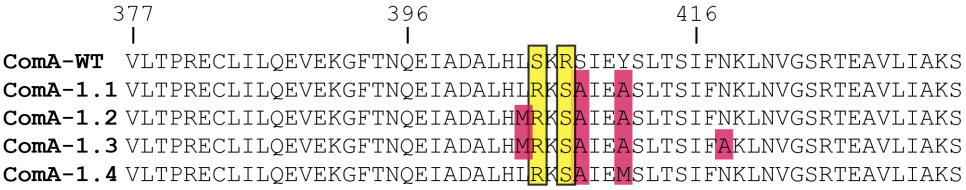


Scaffold 2:

RapF

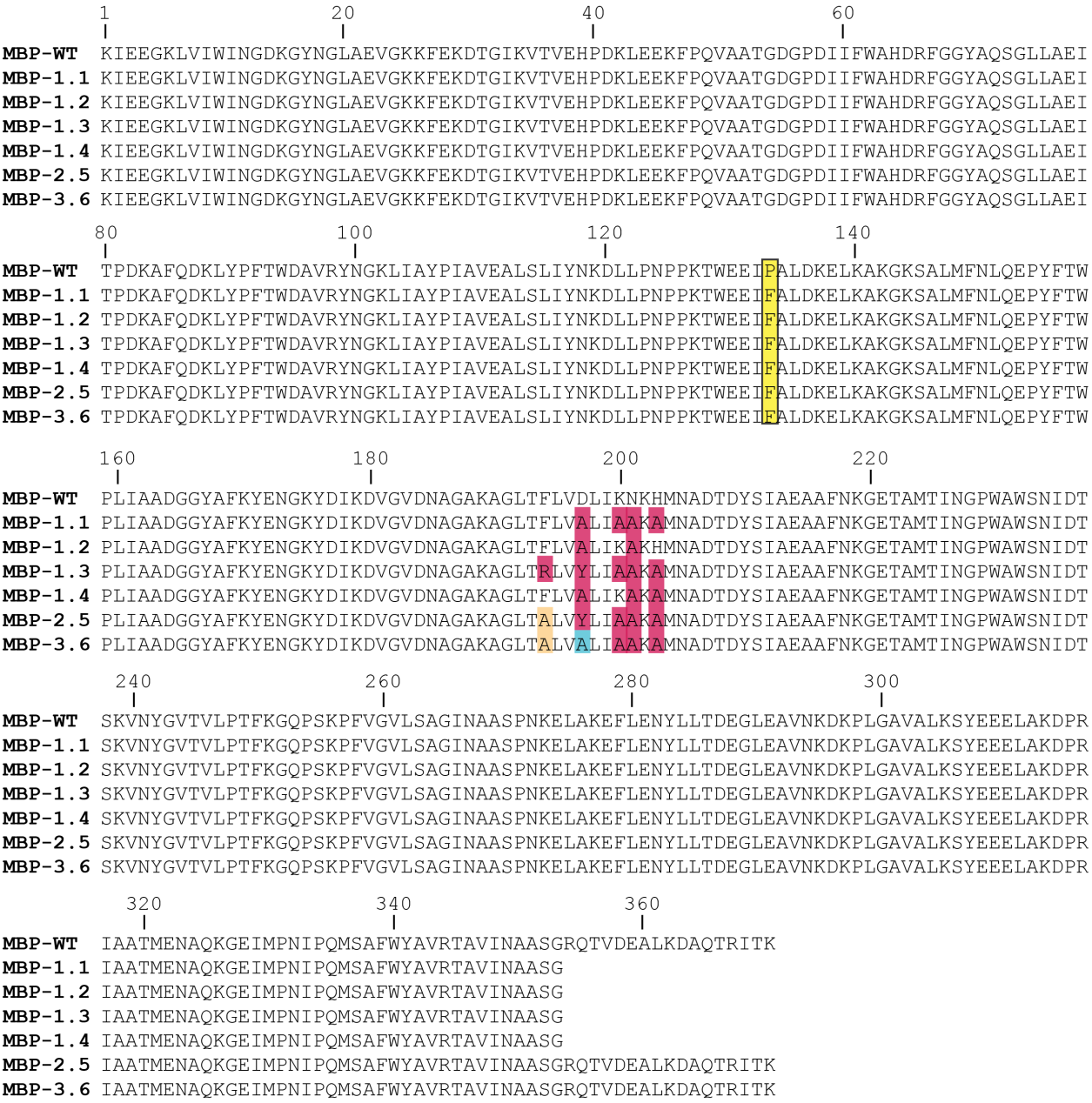


ComA



Scaffold 3:

MBP



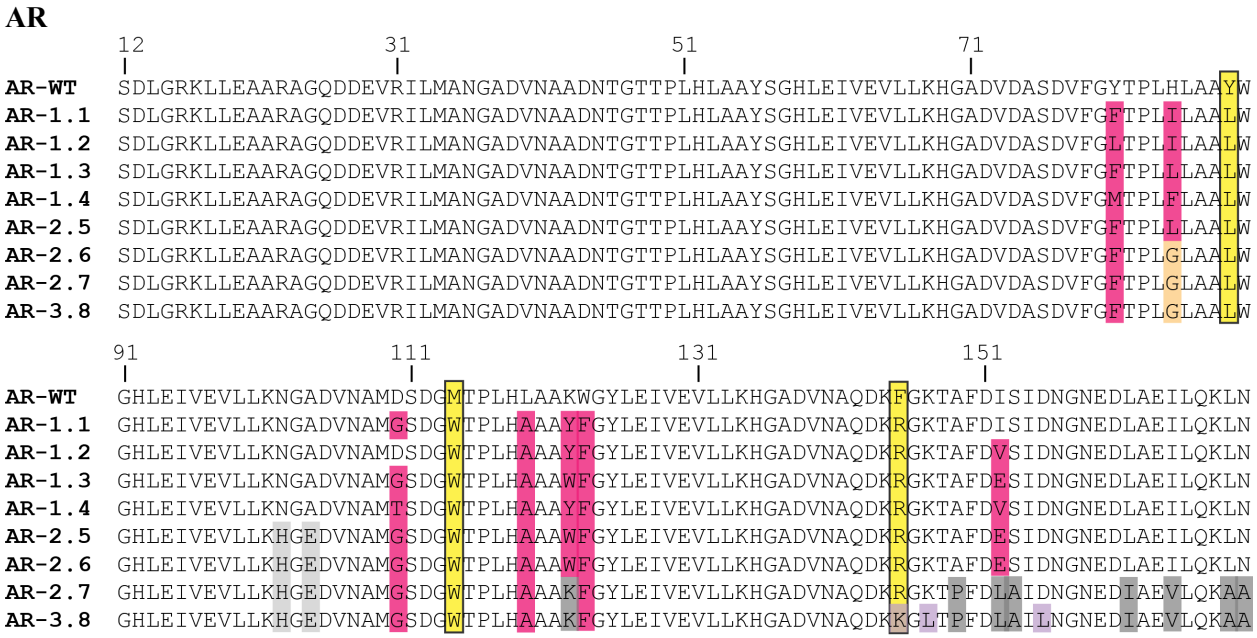


Fig. S1. Sequence alignments of designed proteins (Table S1). Residues that are different from the wild-type scaffold are highlighted, and motif residues are highlighted in yellow and boxed. Computationally predicted mutations for original designs are in magenta, stability-enhancing mutations from Kramer *et al.*(*31*) are in dark gray, mutations from error-prone PCR are in light grey, mutations from saturation mutagenesis are in orange, computationally predicted stabilizing mutations are in purple, and the computationally predicted reversion mutation is in blue.

Figure S2

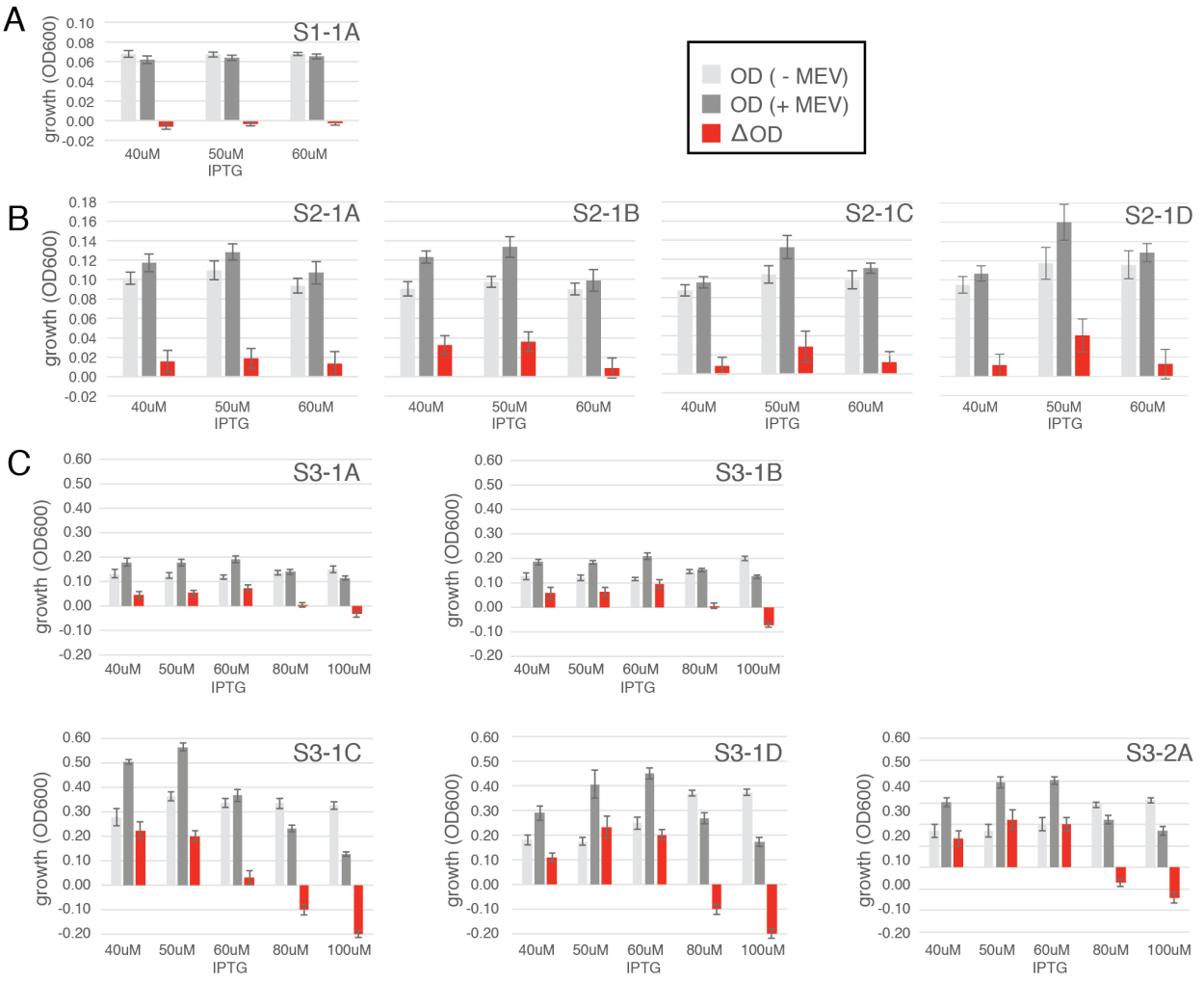


Fig. S2. Growth +/- 5 mM mevalonate and change in growth for data shown in Figure 2C. **(A)** Sensor design for scaffold 1 (FRB/FKBP). (**B**) Sensor designs for scaffold 2 (RapF/ComA). (**C**) Sensor designs for scaffold 3 (AR/MBP). Change in growth is dependent on IPTG concentration (panel C) as expected, where at high induction levels split mDHFR is complemented independent of mevalonate, and adding mevalonate leads to a growth disadvantage because of production of toxic metabolites (IPP, DMAP, and FPP). **Figure 2C** shows data at 50 μM IPTG. Experimental conditions: 0.4% L-arabinose, 1 μg/mL TMP, 35°C. Error bars are the standard deviation from at least 4 biological replicates and 8 replicates for each biological replicate.

Note that S3-2A, which contains two mutations introduced by error-prone PCR and was identified by library screening using the split mDHFR reporter, behaves similarly (except for a slight shift in

the effect of IPTG concentration) to the original computational design S3-1C, which has an identical sequence without the two error-prone PCR mutations (**Fig. 2A**, **Fig. S5**).

Designs for scaffolds 1 (panel A) and 2 (panel B) failed to give a robust signaling response. In general, designs could fail because (i) even though the designs dimerize, they do not properly reconstitute the split reporter (for example linkage between sensor and reporter components is problematic), (ii) the designs do not express at sufficient concentrations, are not properly folded, or a combination of both, or (iii) the designs express sufficiently and are folded but there is no or insufficient binding to the small molecule ligand across the complex interface. Scaffold 1 has been used extensively as rapamycin-induced signaling system linked to a variety of different outputs(*73*), so (ii) or (iii) are more likely reasons for failure. The designs for scaffold 2 show a weak but reproducible response to FPP, so the designs are either expressed at too low levels under our conditions, display only weak ligand binding, or both. A more detailed analysis would be required to dissect the failure modes more precisely.

Figure S3

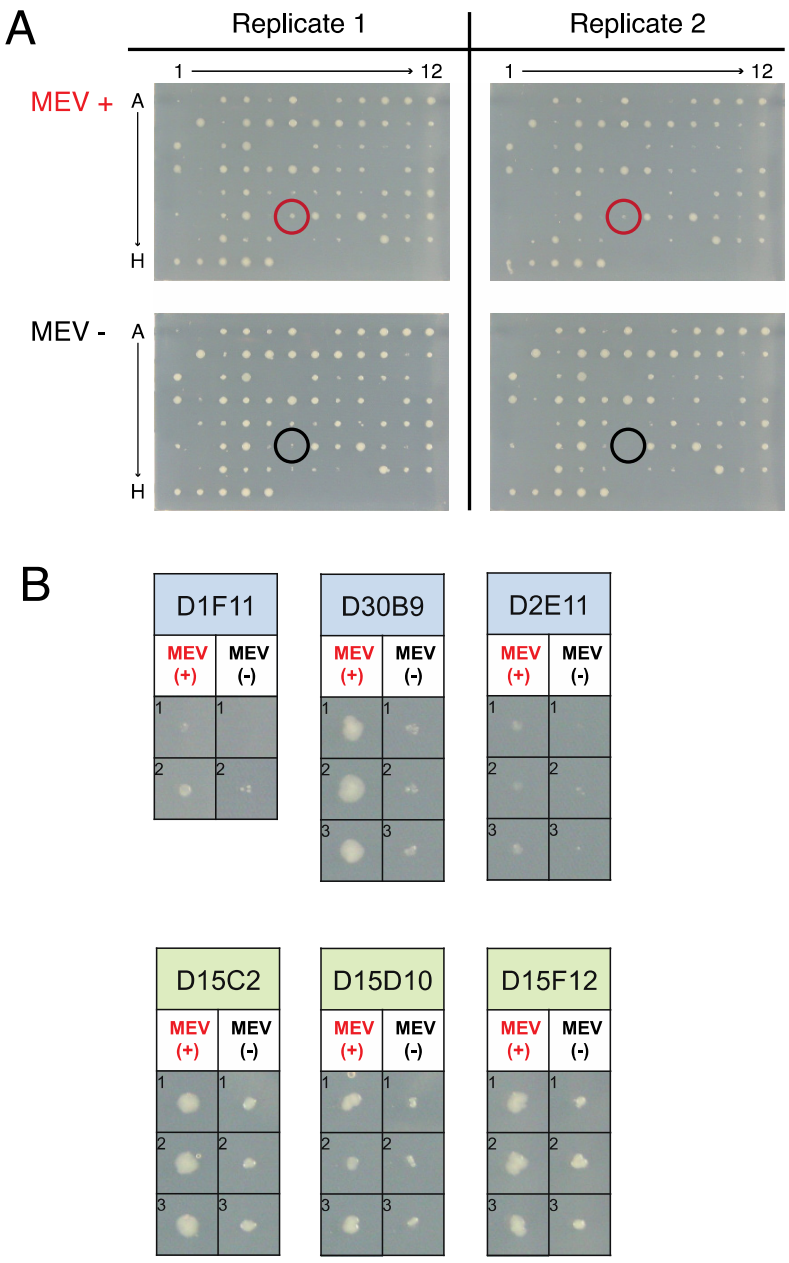
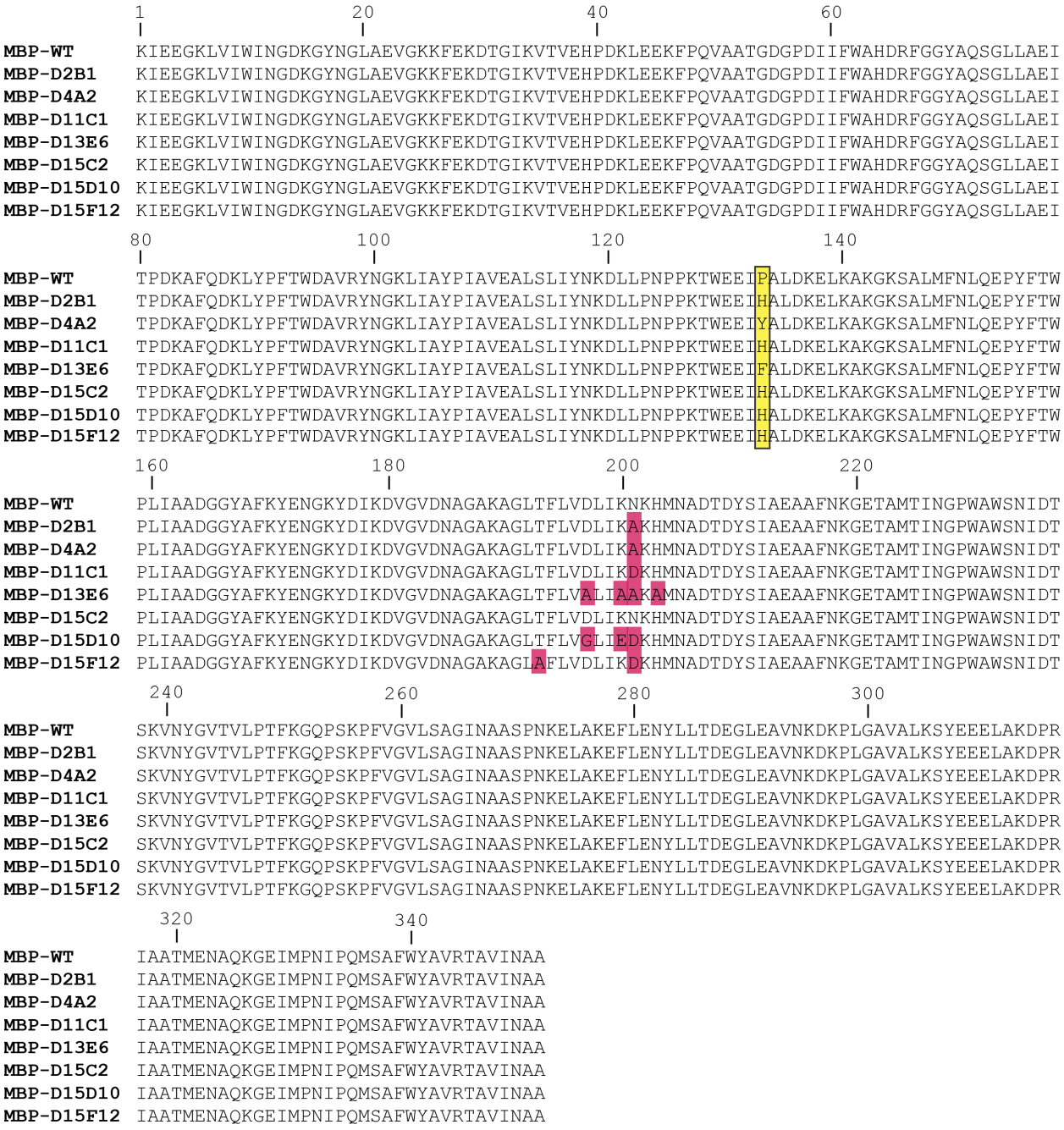


Fig. S3. Example data from library screening using a colony-printing assay. (A) Comparison of growth in the presence (top) and absence (bottom) of 5 mM mevalonate for two replicates. A colony growing better with mevalonate is circled in the replicates. (**B**) Examples of two to three replicates for library hits (concatenated from different plates). Experimental conditions: 0.4 % L-arabinose, 6 µg/mL TMP, 100 µM IPTG. Top row: hits from library 2; bottom row: hits from library 1.

Figure S4

MBP



AR

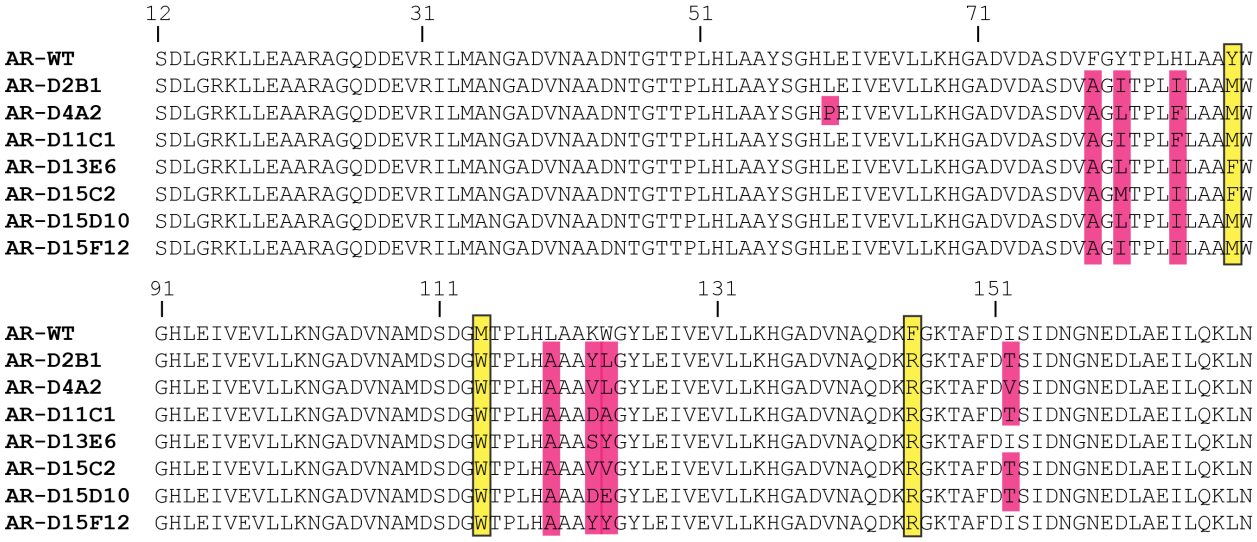
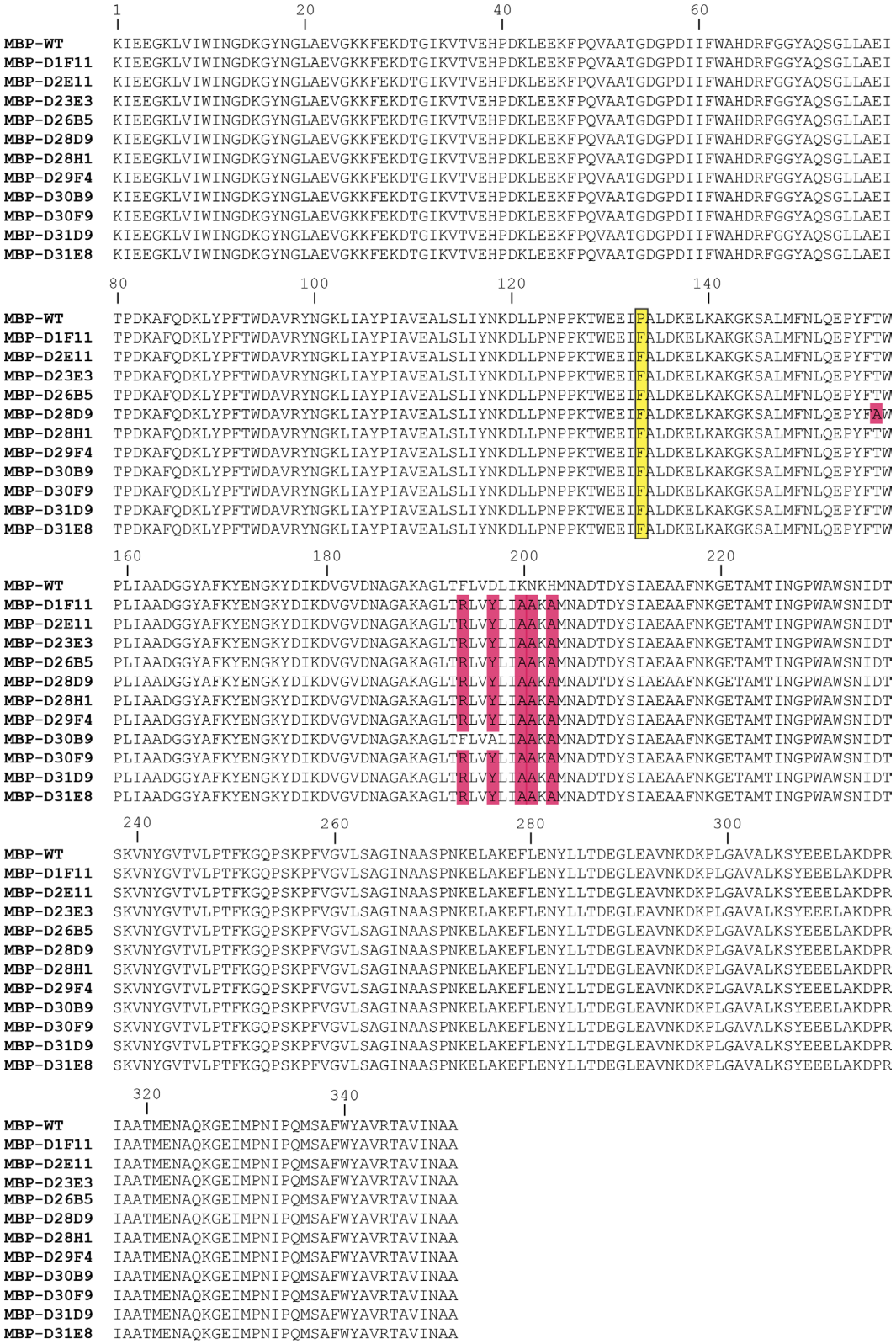


Fig. S4. Sequence alignments of hits from the computationally designed S3 library. These variants were identified using a plate-based split mDHFR assay (**Fig. S3**) and validated by individual solution growth assays (Methods). Residues different from the WT scaffold are highlighted in magenta, and motif residues are highlighted in yellow and boxed.

Figure S5

MBP



AR

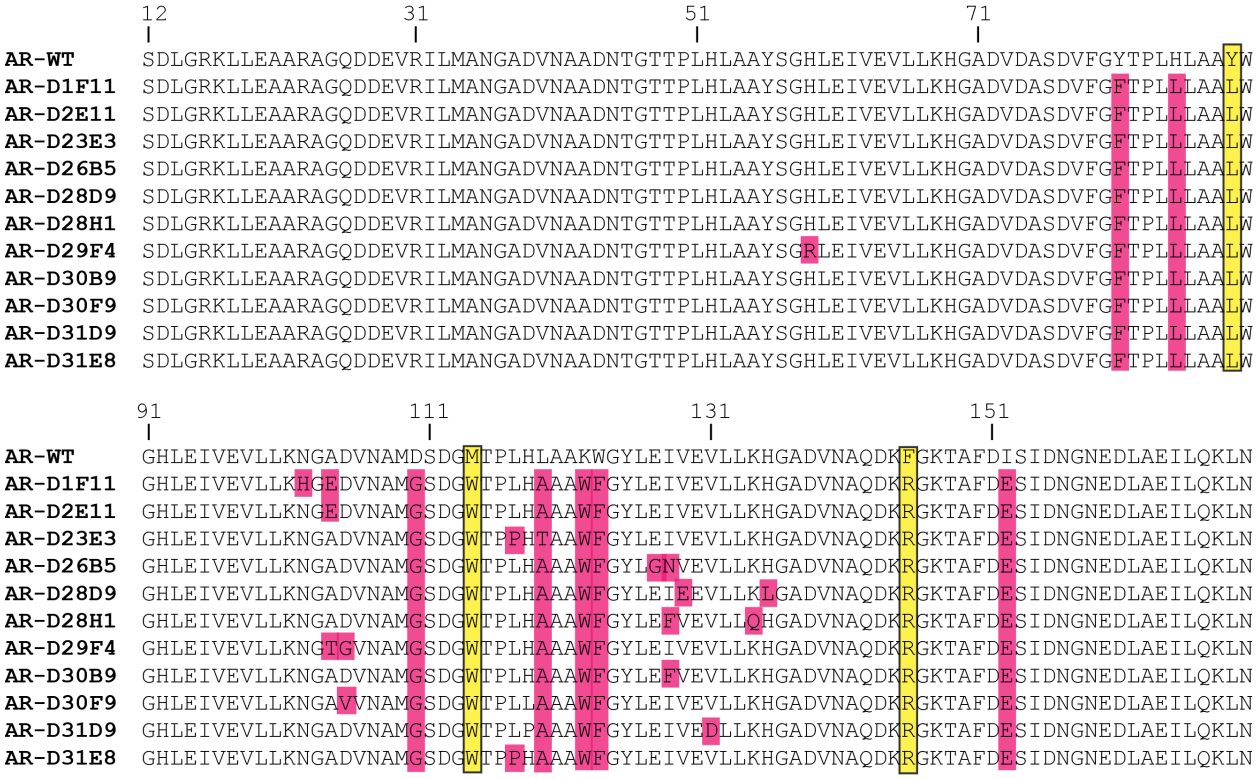
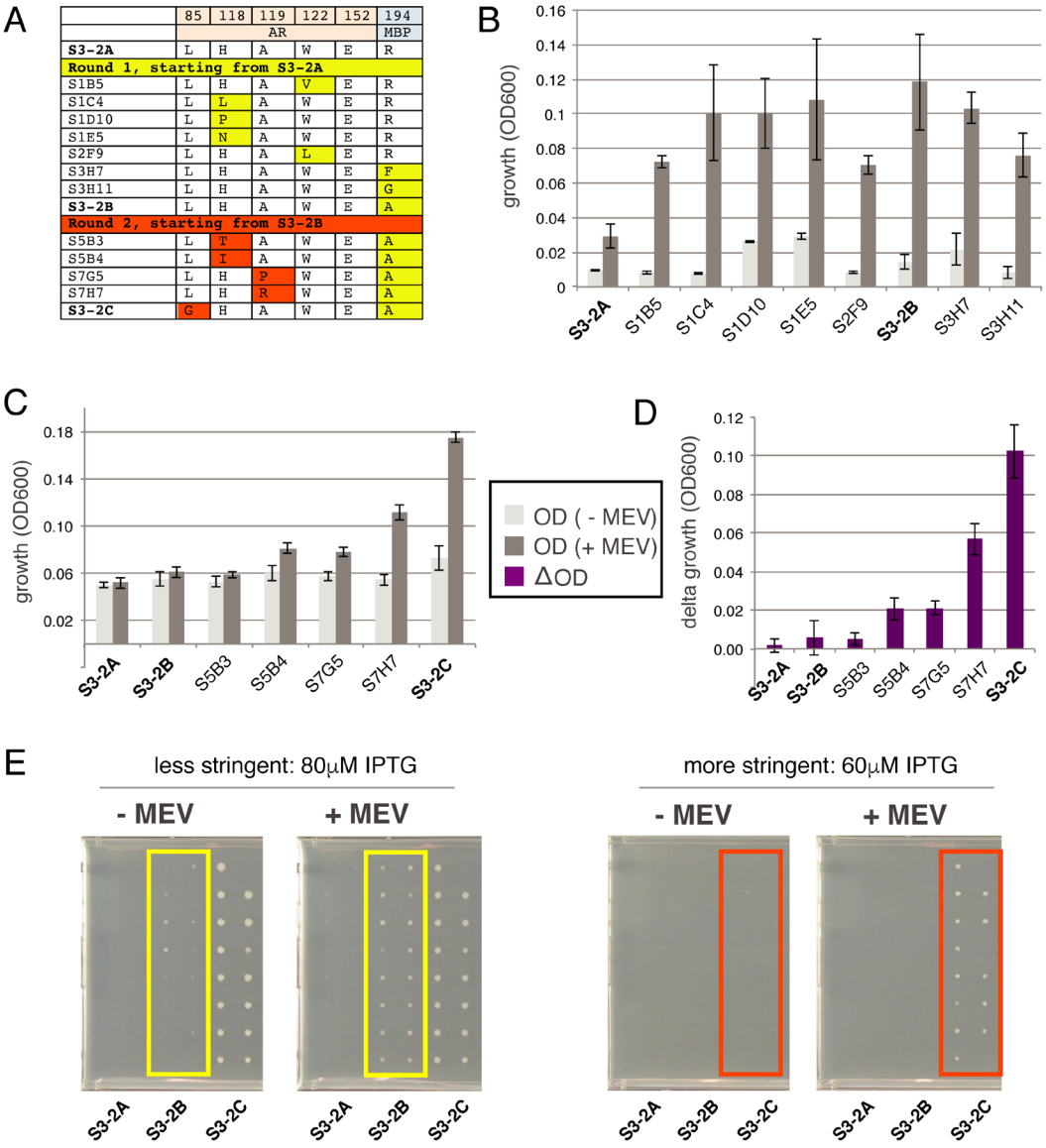


Fig. S5. Sequence alignments of hits from the error-prone PCR S3 library. These variants were identified using a plate-based split mDHFR assay (**Fig. S3**) and validated by individual solution growth assays (Methods). Residues different from the WT scaffold are highlighted in magenta, and motif residues are highlighted in yellow and boxed.

Figure S6



	85	118	119	122	152	194
	AR					MBP
S3-2A	L	H	A	W	E	R
Round 1, starting from S3-2A						
S1B5	L	H	A	V	E	R
S1C4	L	L	A	W	E	R
S1D10	L	P	A	W	E	R
S1E5	L	N	A	W	E	R
S2F9	L	H	A	L	E	R
S3H7	L	H	A	W	E	F
S3H11	L	H	A	W	E	G
S3-2B	L	H	A	W	E	A
Round 2, starting from S3-2B						
S5B3	L	T	A	W	E	A
S5B4	L	I	A	W	E	A
S7G5	L	H	P	W	E	A
S7H7	L	H	R	W	E	A
S3-2C	G	H	A	W	E	A

Fig. S6. Data from the single-site saturation mutagenesis screen. (A) Sequences of hits at screened positions. (**B**) Validation of hits from round 1 (starting from S3-2A) with the split mDHFR assay in liquid culture. (**C**) Validation of hits from round 2 (starting from S3-2B) with split mDHFR assay in liquid culture. (**D**) Sensor signal (change in growth as measured by OD_{600}) from round 2. (**E**) Split mDHFR plate assay showing that sensors S3-2B and S3-2C function under more stringent conditions (lower IPTG inducer concentration). Experimental conditions: Round 1 (panel B): cultures grown with 80 μM IPTG at 35 °C, *n*=3. Round 2: (panels C, D): cultures grown with 60 μM IPTG at 35 °C, *n*=4. Error bars reflect standard deviation.

Figure S7

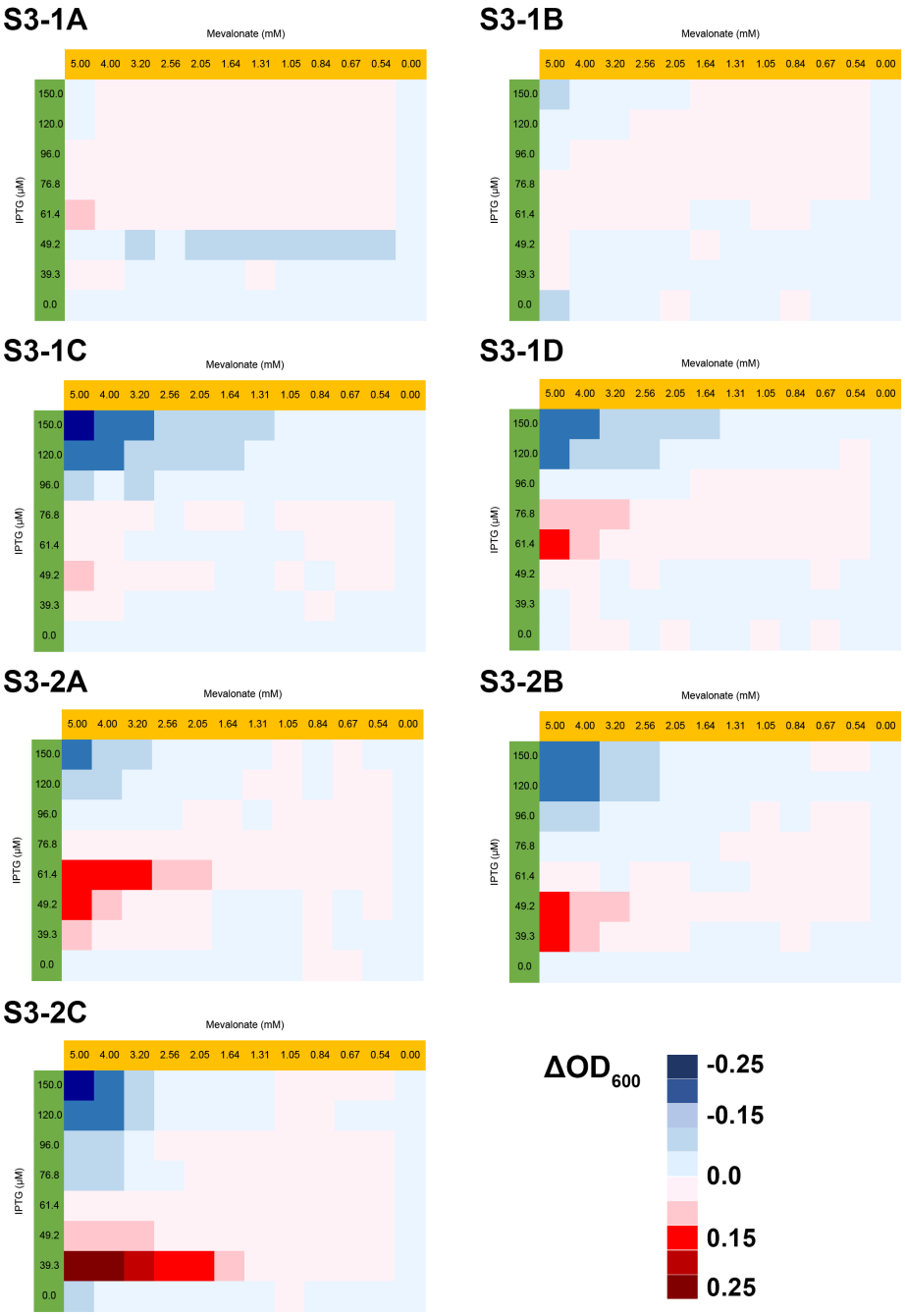


Fig. S7. Dependency of the sensor signal on IPTG and mevalonate concentrations. Data are from a split mDHFR assay in liquid culture. Experimental conditions: varied concentrations of IPTG and mevalonate, 0.4% L-arabinose, 1 µg/ml TMP, cultured at 35 °C. Data are averaged over 6 plates per design.

Figure S8

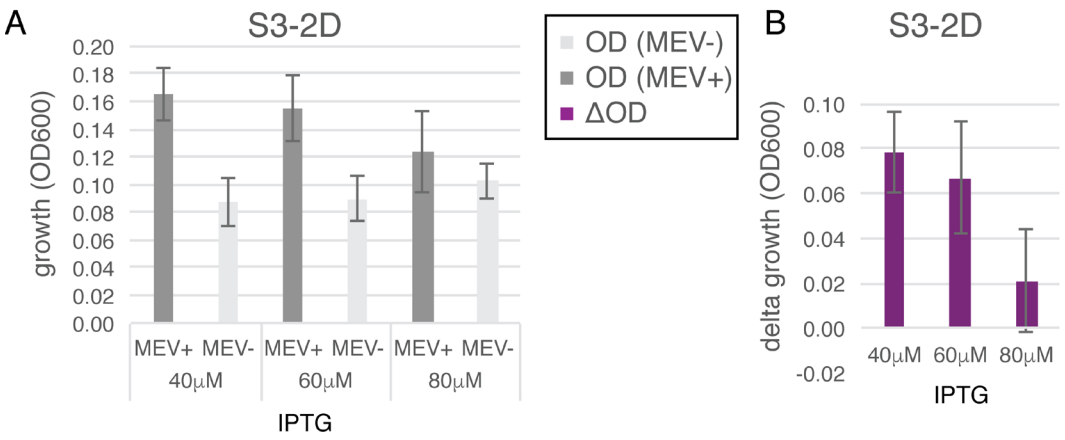


Fig. S8. (A) Growth +/- mevalonate and (B) change in growth for design S3-2D. Experimental conditions: 35 °C, 1 µg/ml TMP, 5 mM mevalonate. Error bars are standard deviation from at least 4 biological replicates and 8 replicates for each biological replicate.

Figure S9

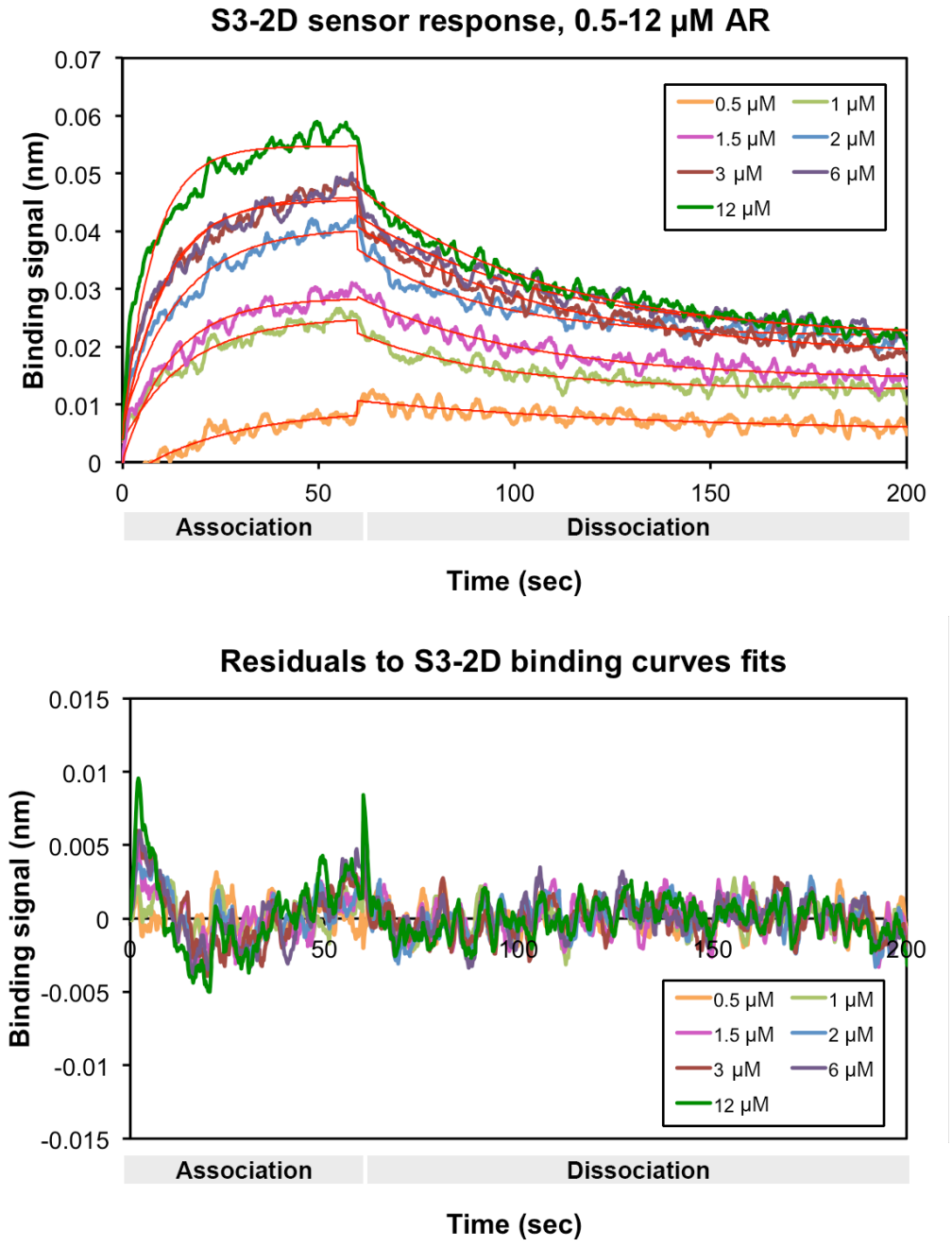


Fig. S9. BLI binding assay representative data and fits. The apparent K_D of design S3-2D was measured using immobilized avi-MBP-2.5, titrated AR-2.7 at indicated concentrations (top), in the presence of 200 µM FPP. Red lines in the top plot indicate fit to the data and residuals are shown in the bottom plot.

Figure S10

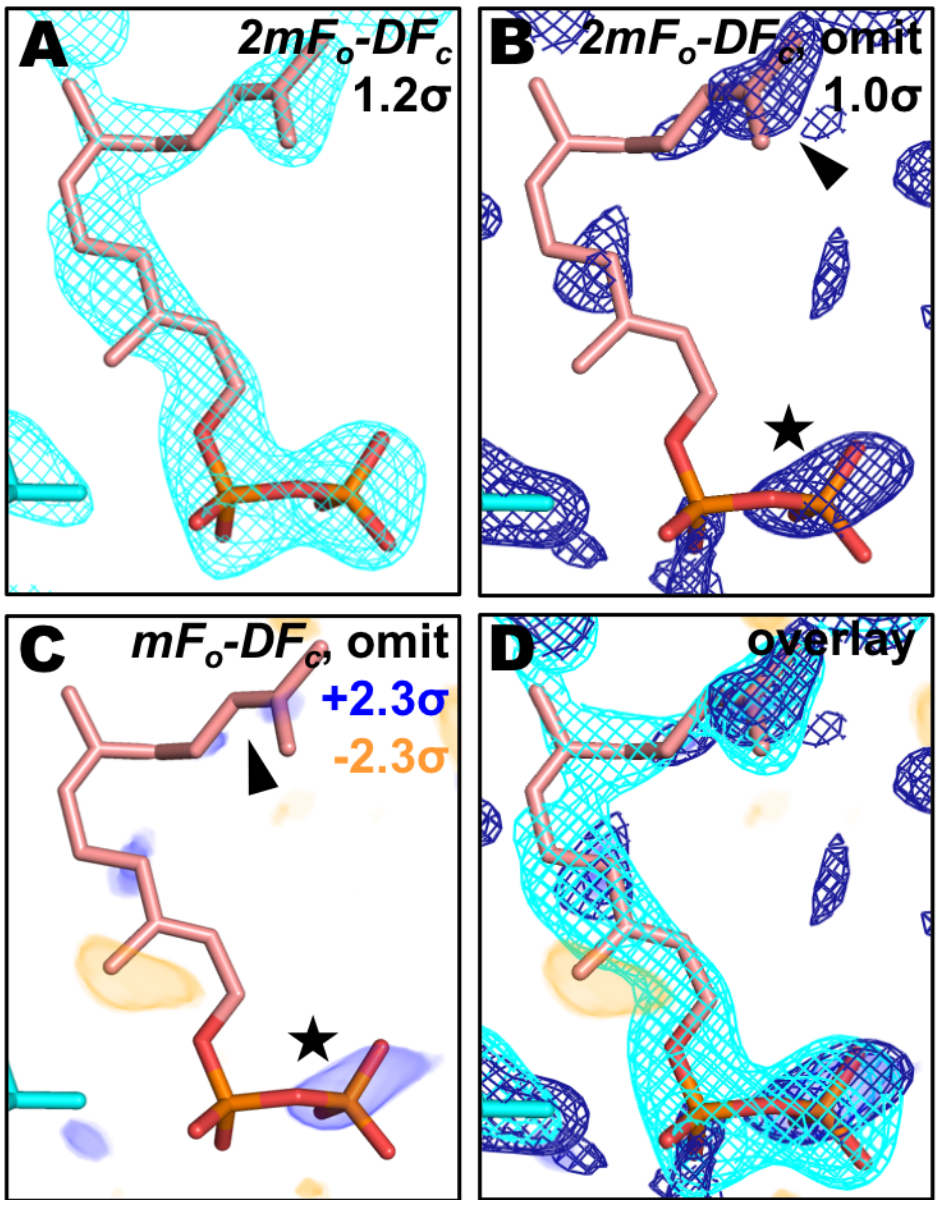


Fig. S10. Electron density maps supporting placement of the FPP ligand. The $2mF_o$-DF_c electron density map created using the final model phases (**A**, cyan mesh) shows good agreement with the modeled ligand (RSCC=0.89). The $2mF_o$-DF_c (**B**, dark blue mesh) and mF_o-DF_c (**C**, blue/orange volumes) omit maps contain notable peaks corresponding to the pyrophosphate group (marked with a star) and the aliphatic tail (marked with an arrow), which resembled peaks in the $2mF_o$-DF_c and mF_o-DF_c maps that were used to initially identify the ligand. An overlay of the different electron density maps (**D**) shows the co-localization of these features in real space.

Figure S11

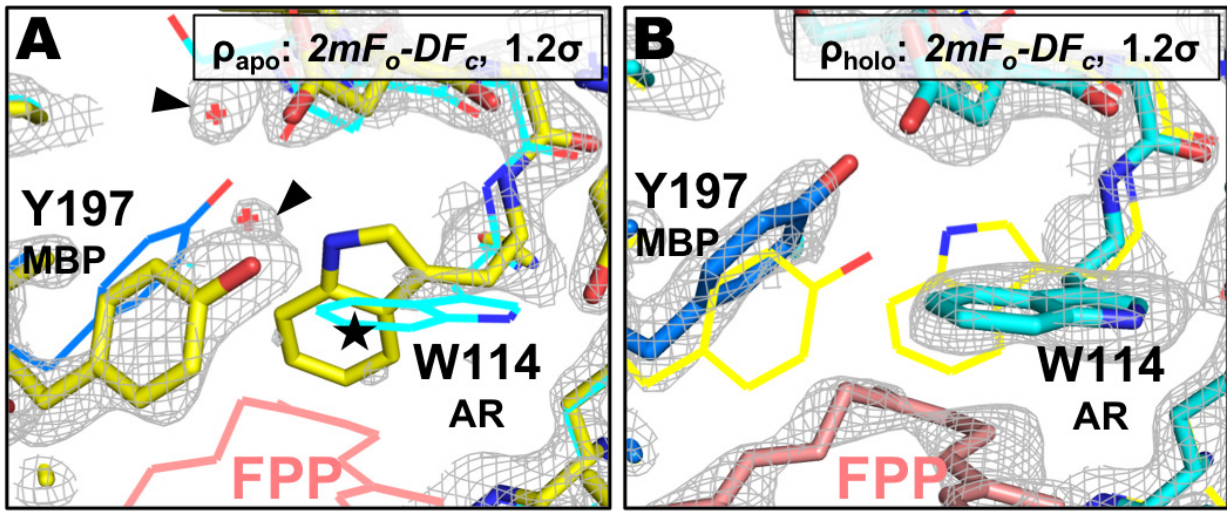


Fig. S11. Electron density supporting ligand-induced rearrangement of the FPP binding site. In the unoccupied FPP binding site (yellow models, density shown in panel (**A**), Y197 in the MBP monomer is rotated down into the binding pocket, and W114 of the AR monomer appears flexible, likely occupying multiple rotameric states (denoted by a star). When the binding site is occupied (MBP models in blue, AR models in cyan, density shown in panel (**B**), the Y197 side chain rotates away from the FPP molecule, displacing a small network of neighboring water molecules (indicated by the black arrows in panel A), and W114 becomes well-ordered with its indole group packed against the bound FPP molecule (pink).

Figure S12

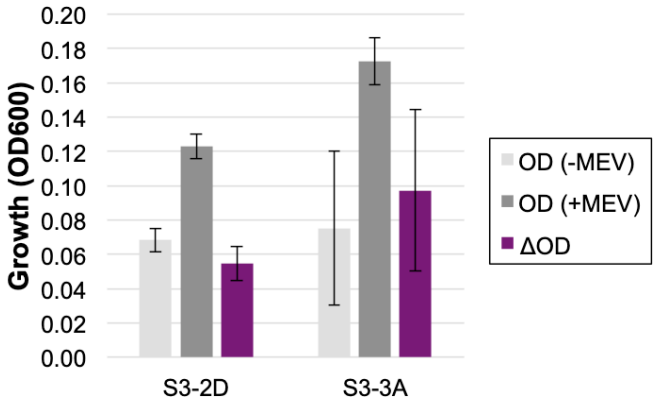


Fig. S12. Comparison of S3-2D and S3-3A (Y197A mutation) in the split mDHFR assay. Design S3-3A is an active sensor in *E. coli*. Experimental conditions: 35 °C, 60 µM IPTG, 6 µg/ml TMP, 0.4% L-arabinose. Error bars are standard deviations, *n*=8 for each design and +/- mevalonate condition.

Figure S13

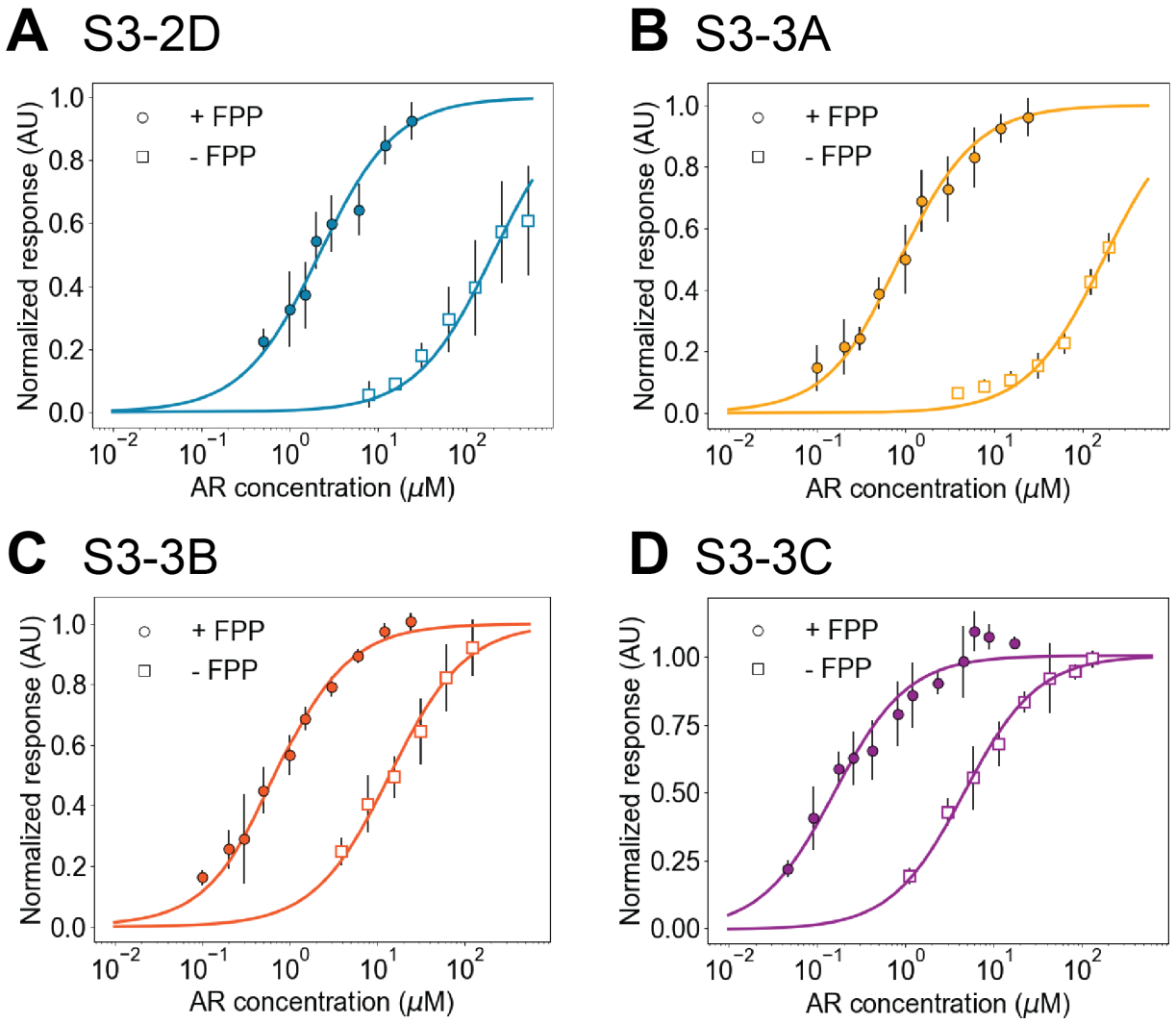


Fig. S13. BLI binding data +/- FPP for designs S3-2D, S3-3A, S3-3B, S3-3C. Experiments use immobilized MBP variants titrating in AR variants at the indicated concentrations. Data for the +FPP condition are also shown in **Figure 3E** and depicted here for comparison. Apparent K_D values are given in **Figure 3C**. Error bars are standard deviations, $n \geq 3$ for each of +/- FPP conditions.

Figure S14

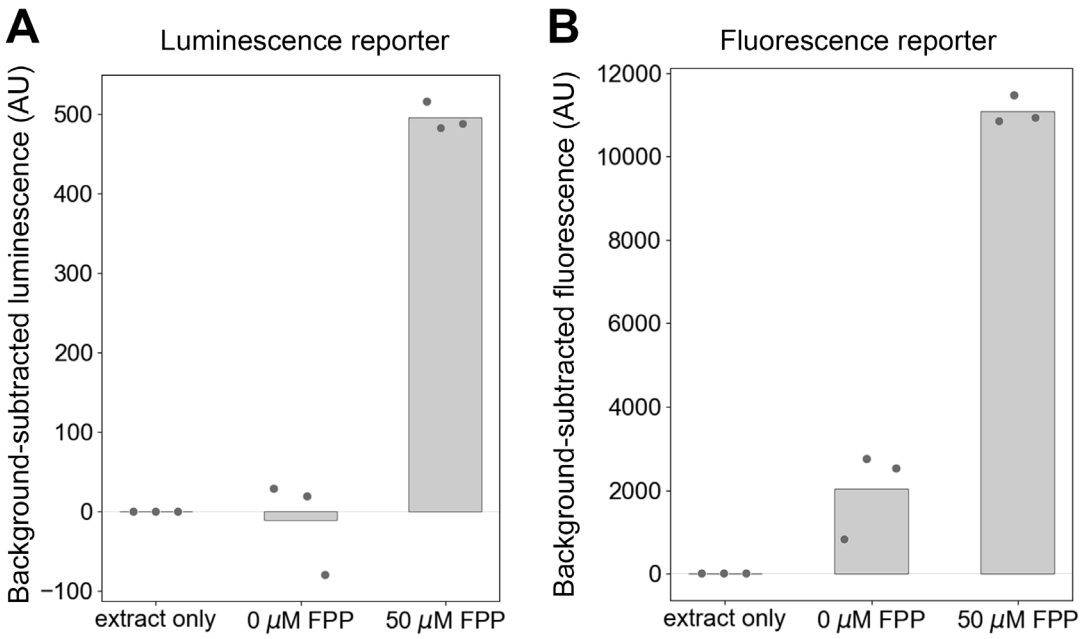


Fig. S14. TxTl data for sensor S3-2D. Background-subtracted sensor signal for *in vitro* TxTl experiments using (**A**) split luciferase and (**B**) dimerization-dependent fluorescent protein reporters. The background is the fluorescent or luminescent signal measured in the assay for the cell extract only control (no AR or MBP plasmids). The bar height shows the average, with individual data points from technical triplicates overlaid. In (**B**), the observed signal in the absence of FPP for sensors fused to the fluorescence reporter is about 20% of the total signal at high (saturating) concentrations of FPP, which could be a result of the higher affinity of the reporter components for each other as compared to the split luciferase components.

In general in CID systems, the maximum amplitude of the sensor signal is intrinsically limited by the difference in sensor protein heterodimerization affinity in the presence and absence of the small molecule (as shown in **Fig. 3B**; the sensor protein affinity difference can be additionally influenced by residual affinity of the reporter halves for each other at the relevant concentrations), the concentration of ternary complex at which this difference is maximized, and the reporter signal at this concentration. Compared to the non-split reporter alone at the same concentration, the signal could be reduced by imperfect reconstitution of the split reporter and local differences in structure, stability, or dynamics.

Appendix 1: Plasmids / constructs for split mDHFR reporter assays. See **Appendix 4** for full DNA sequences for each gene and vector listed below.

Designs with C-terminal split mDHFR

Construct	Vector background	Description
DHFR1	pCDFDuet*	FKBP12/FRB design S1-1A
DHFR2	pCDFDuet	RapF/ComA design S2-1A
DHFR3	pCDFDuet	RapF/ComA design S2-1B
DHFR4	pCDFDuet	RapF/ComA design S2-1C
DHFR5	pCDFDuet	RapF/ComA design S2-1D
DHFR6	pCDFDuet	AR/MBP design S3-1A
DHFR7	pCDFDuet	AR/MBP design S3-1B
DHFR8	pCDFDuet	AR/MBP design S3-1C
DHFR9	pCDFDuet	AR/MBP design S3-1D
DHFR10	pCDFDuet	AR/MBP Design S3-2A
DHFR11	pCDFDuet	AR/MBP Design S3-2B
DHFR12	pCDFDuet	AR/MBP Design S3-2C
DHFR13	pCDFDuet	AR/MBP Design S3-2D
DHFR14	pCDFDuet	AR/MBP Design S3-3A

***** pCDFDuet (Novagen) is spectromycin/spectinomycin-resistant and has the CDF origin of replication. Each protein is fused to one of the subunits of split mDHFR. See **Appendix 3** for sequences of split mDHFR subunits.

Mevalonate pathway constructs

Construct	Vector background	Description
pMBIS	pB8a**	*ERG12-ERG8-MVD1-idi-ispA*
ispA R116A	pB8a	pMBIS + loss of function in *ispA*
pB5K	pB8a	pMBIS + *ADS*

****** pB8a is **chloramphenicol-resistant** and has the **pBBR1 origin of replication**.

ERG12: mevalonate kinase from S. cerevisiae;
ERG8: phosphomevalonate kinase from *S. cerevisiae*;
MVD: mevalonate pyrophosphate decarboxylase from *S. cerevisiae*;
idi: IPP isomerase from E. coli;
ispA: FPP synthase from E. coli.
Plasmid ispA R116A contains a single point mutation in ispA that reduces catalytic activity 13-fold(36). Plasmid pB5K adds amorphadiene synthase (ADS) that converts FPP to amorphadiene(20).

Motif residue alanine reversions with C-terminal split mDHFR

Construct	Vector background	Description
DHFR15	pCDFDuet	Design S3-2C AR L89A
DHFR16	pCDFDuet	Design S3-2C AR W114A
DHFR17	pCDFDuet	Design S3-2C AR R145A
DHFR18	pCDFDuet	Design S3-2C MBP F133A

Appendix 2: Plasmids / constructs for *in vitro* binding experiments and crystallography. These constructs included the C-terminal alpha helix in MBP (residues 354-370). See **Appendix 4** for full DNA sequences for each gene and vector listed below.

Constructs for bio-layer interferometry (Designs S3-2D, S3-3A, S3-3B and S3-3C)

Construct	Vector background	Description*
BLI1	pET28b-GG	AR-2.7
BLI2	pET28b-GG	Avi-MBP-2.5
BLI3	pET28b-GG	AR-3.8
BLI4	pET28b-GG	Avi-MBP-3.6

***Avi tag sequence**
DNA: GGTCTGAACGACATCTTCGAGGCTCAGAAAATCGAATGGCACGAA
Protein: GLNDIFEAQKIEWHE

Constructs for crystallography (Design S3-2D)

Construct	Vector background	Description
CR1	pET28a	MBP-2.5
CR2	pET47b(+)	AR-2.7

Appendix 3: Plasmids / constructs for TxTl reporter assays. See **Appendix 4** for full DNA sequences for each gene, vector, and reporter listed below.

Designs S3-2D and S3-3A constructs for TxTl cell-free reporter assay

Construct	Vector background*	Description**
TxTl1	TxTl T7 A1	LgBIT-AR-2.7
TxTl2	TxTl T7 A2	MBP-2.5-SmBIT
TxTl3	TxTl T7 A2	ddRFPb-AR-2.7
TxTl4	TxTl T7 A2	MBP-2.5-ddGFPa
TxTl5	TxTl T7 A2	MBP-3.6-SmBIT
TxTl6	TxTl T7 A2	MBP-3.6-ddGFPa

* The TxTl T7 A1 vector has a kanamycin resistance cassette and a p15a origin of replication.
The TxTl T7 A2 vector has an ampicillin resistance cassette and a ColE1 origin of replication.

**LgBIT: 18 kDa large subunit of the engineered split nanoluciferase, NanoBiT
SmBIT: 1.3 kDa (11 amino acid peptide) small subunit of NanoBiT

ddRFPb: 26 kDa dimerization-dependent protein that fluoresces in complex with ddGFPa.
ddGFPa: 26 kDa dimerization-dependent protein that fluoresces in complex with ddRFPb.

ddRFPb + ddGFPa complex is excited at 493 and 380 nm, with emission peaks at 448 and 508 nm.

Controls for TxTl cell-free reporter assay

Construct	Vector background	Description
TxTl7	TxTl T7 A1	LgBIT-WT AR
TxTl8	TxTl T7 A2	WT MBP-SmBIT
TxTl9	TxTl T7 A2	ddRFPb-WT AR
TxTl10	TxTl T7 A2	WT MBP-ddGFPa

Additional construct

Construct	Vector background***	Description
TxTl11	TxTl T500 A3	T7 RNA polymerase expression

***The TxTl T500 A3 vector has an ampicillin resistance cassette and a p15a origin of replication. The target gene is expressed from the Pr promoter from lambda phage.

Appendix 4: Gene and vector DNA sequences.

pCDFDuet

CGCTGACGTCGGTACCCTCGAGTCTGGTAAAGAAACCGCTGCTGCGAAATTTGAACGCCAGCACATGGAC
TCGTCTACTAGCGCAGCTTAATTAACCTAGGCTGCTGCCACCGCTGAGCAATAACTAGCATAACCCCTTG
GGGCCTCTAAACGGGTCTTGAGGGGTTTTTTGCTGAAACCTCAGGCATTTGAGAAGCACACGGTCACACT
GCTTCCGGTAGTCAATAAACCGGTAAACCAGCAATAGACATAAGCGGCTATTTAACGACCCTGCCCTGAA
CCGACGACCGGGTCATCGTGGCCGGATCTTGCGGCCCCTCGGCTTGAACGAATTGTTAGACATTATTTGC
CGACTACCTTGGTGATCTCGCCTTTCACGTAGTGGACAAATTCTTCCAACTGATCTGCGCGCGAGGCCAA
GCGATCTTCTTCTTGTCCAAGATAAGCCTGTCTAGCTTCAAGTATGACGGGCTGATACTGGGCCGGCAGG
CGCTCCATTGCCCAGTCGGCAGCGACATCCTTCGGCGCGATTTTGCCGGTTACTGCGCTGTACCAAATGC
GGGACAACGTAAGCACTACATTTCGCTCATCGCCAGCCCAGTCGGGCGGCGAGTTCCATAGCGTTAAGGT
TTCATTTAGCGCCTCAAATAGATCCTGTTCAGGAACCGGATCAAAGAGTTCCTCCGCCGCTGGACCTACC
AAGGCAACGCTATGTTCTCTTGCTTTTGTCAGCAAGATAGCCAGATCAATGTCGATCGTGGCTGGCTCGA
AGATACCTGCAAGAATGTCATTGCGCTGCCATTCTCCAAATTGCAGTTCGCGCTTAGCTGGATAACGCCA
CGGAATGATGTCGTCGTGCACAACAATGGTGACTTCTACAGCGCGGAGAATCTCGCTCTCTCCAGGGGAA
GCCGAAGTTTCCAAAAGGTCGTTGATCAAAGCTCGCCGCGTTGTTTCATCAAGCCTTACGGTCACCGTAA
CCAGCAAATCAATATCACTGTGTGGCTTCAGGCCGCCATCCACTGCGGAGCCGTACAAATGTACGGCCAG
CAACGTCGGTTCGAGATGGCGCTCGATGACGCCAACTACCTCTGATAGTTGAGTCGATACTTCGGCGATC
ACCGCTTCCCTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCG
GATACATATTTGAATGTATTTAGAAAAATAAACAAATAGCTAGCTCACTCGGTCGCTACGCTCCGGGCGT
GAGACTGCGGCGGGCGCTGCGGACACATACAAAGTTACCCACAGATTCCGTGGATAAGCAGGGGACTAAC
ATGTGAGGCAAAACAGCAGGGCCGCGCCGGTGGCGTTTTTCCATAGGCTCCGCCCCTCCTGCCAGAGTTCA
CATAAACAGACGCTTTTCCGGTGCATCTGTGGGAGCCGTGAGGCTCAACCATGAATCTGACAGTACGGGC
GAAACCCGACAGGACTTAAAGATCCCCACCGTTTCCGGCGGGTCGCTCCCTCTTGCGCTCTCCTGTTCCG
ACCCTGCCGTTTACCGGATACCTGTTCCGCCTTTCTCCCTTACGGGAAGTGTGGCGCTTTCTCATAGCTC
ACACACTGGTATCTCGGCTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTAAGCAAGAACTCCCCGTTC
AGCCCGACTGCTGCGCCTTATCCGGTAACTGTTCACTTGAGTCCAACCCGGAAAAGCACGGTAAAACGCC
ACTGGCAGCAGCCATTGGTAACTGGGAGTTCGCAGAGGATTTGTTTAGCTAAACACGCGGTTGCTCTTGA
AGTGTGCGCCAAAGTCCGGCTACACTGGAAGGACAGATTTGGTTGCTGTGCTCTGCGAAAGCCAGTTACC
ACGGTTAAGCAGTTCCCCAACTGACTTAACCTTCGATCAAACCACCTCCCCAGGTGGTTTTTTCGTTTAC
AGGGCAAAAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACTGAACCGCTC
TAGATTTCAGTGCAATTTATCTCTTCAAATGTAGCACCTGAAGTCAGCCCCATACGATATAAGTTGTAAT
TCTCATGTTAGTCATGCCCCGCGCCCACCGGAAGGAGCTGACTGGGTTGAAGGCTCTCAAGGGCATCGGT
CGAGATCCCGGTGCCTAATGAGTGAGCTAACTTACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGT
CGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGG
GCGCCAGGGTGGTTTTTCTTTTCACCAGTGAGACGGGCAACAGCTGATTGCCCTTCACCGCCTGGCCCTG
AGAGAGTTGCAGCAAGCGGTCCACGCTGGTTTGCCCCAGCAGGCGAAAATCCTGTTTGATGGTGGTTAAC
GGCGGGATATAACATGAGCTGTCTTCGGTATCGTCGTATCCCACTACCGAGATGTCCGCACCAACGCGCA
GCCCGGACTCGGTAATGGCGCGCATTGCGCCCAGCGCCATCTGATCGTTGGCAACCAGCATCGCAGTGGG
AACGATGCCCTCATTCAGCATTTGCATGGTTTGTTGAAAACCGGACATGGCACTCCAGTCGCCTTCCCGT
TCCGCTATCGGCTGAATTTGATTGCGAGTGAGATATTTATGCCAGCCAGCCAGACGCAGACGCGCCGAGA
CAGAACTTAATGGGCCCGCTAACAGCGCGATTTGCTGGTGACCCAATGCGACCAGATGCTCCACGCCCAG
TCGCGTACCGTCTTCATGGGAGAAAATAATACTGTTGATGGGTGTCTGGTCAGAGACATCAAGAAATAAC
GCCGGAACATTAGTGCAGGCAGCTTCCACAGCAATGGCATCCTGGTCATCCAGCGGATAGTTAATGATCA
GCCCACTGACGCGTTGCGCGAGAAGATTGTGCACCGCCGCTTTACAGGCTTCGACGCCGCTTCGTTCTAC
CATCGACACCACCACGCTGGCACCCAGTTGATCGGCGCGAGATTTAATCGCCGCGACAATTTGCGACGGC
GCGTGCAGGGCCAGACTGGAGGTGGCAACGCCAATCAGCAACGACTGTTTGCCCGCCAGTTGTTGTGCCA
CGCGGTTGGGAATGTAATTCAGCTCCGCCATCGCCGCTTCCACTTTTTCCCGCGTTTTCGCAGAAACGTG

GCTGGCCTGGTTCACCACGCGGGAAACGGTCTGATAAGAGACACCGGCATACTCTGCGACATCGTATAAC
GTTACTGGTTTCACATTCACCACCCTGAATTGACTCTCTTCCGGGCGCTATCATGCCATACCGCGAAAGG
TTTTGCGCCATTCGATGGTGTCCGGGATCTCGACGCTCTCCCTTATGCGACTCCTGCATTAGGAAATTAA
TACGACTCACTATAGGGGAATTGTGAGCGGATAACAATTCCCCTGTAGAAATAATTTTGTTTAACTTTAA
TAAGGAGATATACC

pB8a

GGATCCTAACTCGAGTAAGGATCTCCAGGCATCAAATAAAACGAAAGGCTCAGTCGAAAGACTGGGCCTT
TCGTTTTATCTGTTGTTTGTCGGTGAACGCTCTCTACTAGAGTCACACTGGCTCACCTTCGGGTGGGCCT
TTTGCGTTTATACCTAGGCTACAGCCGATAGTCTGGAACAGCGCACTTACGGGTTGCTGCGCAACCCAAG
TGCTACCGGCGCGGCAGCGTGACCCGTGTCGGCGGCTCCAACGGCTCGCCATCGTCCAGAAAACACGGCT
CATCGGGCATCGGCAGGCGCTGCTGCCCGCGCCGTTCCCATTCCTCCGTTTCGGTCAAGGCTGGCAGGTC
TGGTTCCATGCCCGGAATGCCGGGCTGGCTGGGCGGCTCCTCGCCGGGGCCGGTCGGTAGTTGCTGCTCG
CCCGGATACAGGGTCGGGATGCGGCGCAGGTCGCCATGCCCCAACAGCGATTCGTCCTGGTCGTCGTGAT
CAACCACCACGGCGGCACTGAACACCGACAGGCGCAACTGGTCGCGGGGCTGGCCCCACGCCACGCGGTC
ATTGACCACGTAGGCCAACACGGTGCCGGGGCCGTTGAGCTTCACGACGGAGATCCAGCGCTCGGCCACC
AAGTCCTTGACTGCGTATTGGACCGTCCGCAAAGAACGTCCGATGAGCTTGGAAAGTGTCTTCTGGCTGA
CCACCACGGCGTTCTGGTGGCCCATCTGCGCCACGAGGTGATGCAGCAGCATTGCCGCCGTGGGTTTCCT
CGCAATAAGCCCGGCCCACGCCTCATGCGCTTTGCGTTCCGTTTGCACCCAGTGACCGGGCTTGTTCTTG
GCTTGAATGCCGATTTCTCTGGACTGCGTGGCCATGCTTATCTCCATGCGGTAGGGGTGCCGCACGGTTG
CGGCACCATGCGCAATCAGCTGCAACTTTTCGGCAGCGCGACAACAATTATGCGTTGCGTAAAAGTGGCA
GTCAATTACAGATTTTCTTTAACCTACGCAATGAGCTATTGCGGGGGGTGCCGCAATGAGCTGTTGCGTA
CCCCCCTTTTTTAAGTTGTTGATTTTTAAGTCTTTCGCATTTCGCCCTATATCTAGTTCTTTGGTGCCCA
AAGAAGGGCACCCCTGCGGGGTTCCCCCACGCCTTCGGCGCGGCTCCCCCTCCGGCAAAAAGTGGCCCCT
CCGGGGCTTGTTGATCGACTGCGCGGCCTTCGGCCTTGCCCAAGGTGGCGCTGCCCCCTTGGAACCCCCG
CACTCGCCGCCGTGAGGCTCGGGGGGCAGGCGGGCGGGCTTCGCCCTTCGACTGCCCCCACTCGCATAGG
CTTGGGTCGTTCCAGGCGCGTCAAGGCCAAGCCGCTGCGCGGTCGCTGCGCGAGCCTTGACCCGCCTTCC
ACTTGGTGTCCAACCGGCAAGCGAAGCGCGCAGGCCGCAGGCCGGAGGCACTAGTGCTTGGATTCTCACC
AATAAAAAACGCCCGGCGGCAACCGAGCGTTCTGAACAAATCCAGATGGAGTTCTGAGGTCATTACTGGA
TCTATCAACAGGAGTCCAAGCGAGCTCGTAAACTTGGTCTGACAGTTACGCCCCGCCCTGCCACTCATCG
CAGTACTGTTGTAATTCATTAAGCATTCTGCCGACATGGAAGCCATCACAAACGGCATGATGAACCTGAA
TCGCCAGCGGCATCAGCACCTTGTCGCCTTGCGTATAATATTTGCCCATGGTGAAAACGGGGGCGAAGAA
GTTGTCCATATTGGCCACGTTTAAATCAAAACTGGTGAAACTCACCCAGGGATTGGCTGAGACGAAAAAC
ATATTCTCAATAAACCCTTTAGGGAAATAGGCCAGGTTTTCACCGTAACACGCCACATCTTGCGAATATA
TGTGTAGAAACTGCCGGAAATCGTCGTGGTATTCACTCCAGAGCGATGAAAACGTTTCAGTTTGCTCATG
GAAAACGGTGTAACAAGGGTGAACACTATCCCATATCACCAGCTCACCGTCTTTCATTGCCATACGGAAT
TCCGGATGAGCATTCATCAGGCGGGCAAGAATGTGAATAAAGGCCGGATAAAACTTGTGCTTATTTTTCT
TTACGGTCTTTAAAAAGGCCGTAATATCCAGCTGAACGGTCTGGTTATAGGTACATTGAGCAACTGACTG
AAATGCCTCAAAATGTTCTTTACGATGCCATTGGGATATATCAACGGTGGTATATCCAGTGATTTTTTTC
TCCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATAT
TTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGACGT
C

pET28b-GG (with GFP dropout)

TGGCGAATGGGACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTGACC
GCTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTCGCCG
GCTTTCCCCGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTCCGATTTAGTGCTTTACGGCACCTCGA
CCCCAAAAAACTTGATTAGGGTGATGGTTCACGTAGTGGGCCATCGCCCTGATAGACGGTTTTTCGCCCT
TTGACGTTGGAGTCCACGTTCTTTAATAGTGGACTCTTGTTCCAAACTGGAACAACACTCAACCCTATCT
CGGTCTATTCTTTTGATTTATAAGGGATTTTGCCGATTTCGGCCTATTGGTTAAAAAATGAGCTGATTTA

ACAAAAATTTAACGCGAATTTTAACAAAATATTAACGTTTACAATTTCAGGTGGCACTTTTCGGGGAAAT
GTGCGCGGAACCCCTATTTGTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCATGAATTAATTCT
TAGAAAAACTCATCGAGCATCAAATGAAACTGCAATTTATTCATATCAGGATTATCAATACCATATTTTT
GAAAAAGCCGTTTCTGTAATGAAGGAGAAAACTCACCGAGGCAGTTCCATAGGATGGCAAGATCCTGGTA
TCGGTCTGCGATTCCGACTCGTCCAACATCAATACAACCTATTAATTTCCCCTCGTCAAAAATAAGGTTA
TCAAGTGAGAAATCACCATGAGTGACGACTGAATCCGGTGAGAATGGCAAAAGTTTATGCATTTCTTTCC
AGACTTGTTCAACAGGCCAGCCATTACGCTCGTCATCAAAATCACTCGCATCAACCAAACCGTTATTCAT
TCGTGATTGCGCCTGAGCGAGACGAAATACGCGATCGCTGTTAAAAGGACAATTACAAACAGGAATCGAA
TGCAACCGGCGCAGGAACACTGCCAGCGCATCAACAATATTTTCACCTGAATCAGGATATTCTTCTAATA
CCTGGAATGCTGTTTTCCCGGGGATCGCAGTGGTGAGTAACCATGCATCATCAGGAGTACGGATAAAATG
CTTGATGGTCGGAAGAGGCATAAATTCCGTCAGCCAGTTTAGTCTGACCATCTCATCTGTAACATCATTG
GCAACGCTACCTTTGCCATGTTTCAGAAACAACTCTGGCGCATCGGGCTTCCCATACAATCGATAGATTG
TCGCACCTGATTGCCCGACATTATCGCGAGCCCATTTATACCCATATAAATCAGCATCCATGTTGGAATT
TAATCGCGGCCTAGAGCAAGACGTTTCCCGTTGAATATGGCTCATAACACCCCTTGTATTACTGTTTATG
TAAGCAGACAGTTTTATTGTTCATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCACTGAGCGTCAGAC
CCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAA
AAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAAC
TGGCTTCAGCAGAGCGCAGATACCAAATACTGTCCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAG
AACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATA
AGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGG
GGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTGAGCTA
TGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAG
GAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCT
CTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCG
GCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCCTGCGTTATCCCCTGATT
CTGTGGATAACCGTATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGCAGCCGAACGACCGAGCGCAG
CGAGTCAGTGAGCGAGGAAGCGGAAGAGCGCCTGATGCGGTATTTTCTCCTTACGCATCTGTGCGGTATT
TCACACCGCATATATGGTGCACTCTCAGTACAATCTGCTCTGATGCCGCATAGTTAAGCCAGTATACACT
CCGCTATCGCTACGTGACTGGGTCATGGCTGCGCCCCGACACCCGCCAACACCCGCTGACGCGCCCTGAC
GGGCTTGTCTGCTCCCGGCATCCGCTTACAGACAAGCTGTGACCGTCTCCGGGAGCTGCATGTGTCAGAG
GTTTTCACCGTCATCACCGAAACGCGCGAGGCAGCTGCGGTAAAGCTCATCAGCGTGGTCGTGAAGCGAT
TCACAGATGTCTGCCTGTTCATCCGCGTCCAGCTCGTTGAGTTTCTCCAGAAGCGTTAATGTCTGGCTTC
TGATAAAGCGGGCCATGTTAAGGGCGGTTTTTTCCTGTTTGGTCACTGATGCCTCCGTGTAAGGGGGATT
TCTGTTCATGGGGGTAATGATACCGATGAAACGAGAGAGGATGCTCACGATACGGGTTACTGATGATGAA
CATGCCCGGTTACTGGAACGTTGTGAGGGTAAACAACTGGCGGTATGGATGCGGCGGGACCAGAGAAAAA
TCACTCAGGGTCAATGCCAGCGCTTCGTTAATACAGATGTAGGTGTTCCACAGGGTAGCCAGCAGCATCC
TGCGATGCAGATCCGGAACATAATGGTGCAGGGCGCTGACTTCCGCGTTTCCAGACTTTACGAAACACGG
AAACCGAAGACCATTCATGTTGTTGCTCAGGTCGCAGACGTTTTGCAGCAGCAGTCGCTTCACGTTCGCT
CGCGTATCGGTGATTCATTCTGCTAACCAGTAAGGCAACCCCGCCAGCCTAGCCGGGTCCTCAACGACAG
GAGCACGATCATGCGCACCCGTGGGGCCGCCATGCCGGCGATAATGGCCTGCTTCTCGCCGAAACGTTTG
GTGGCGGGACCAGTGACGAAGGCTTGAGCGAGGGCGTGCAAGATTCCGAATACCGCAAGCGACAGGCCGA
TCATCGTCGCGCTCCAGCGAAAGCGGTCCTCGCCGAAAATGACCCAGAGCGCTGCCGGCACCTGTCCTAC
GAGTTGCATGATAAAGAAGACAGTCATAAGTGCGGCGACGATAGTCATGCCCCGCGCCCACCGGAAGGAG
CTGACTGGGTTGAAGGCTCTCAAGGGCATCGGTCGAGATCCCGGTGCCTAATGAGTGAGCTAACTTACAT
TAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGG
CCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCCAGGGTGGTTTTTCTTTTCACCAGTGAGACGGG
CAACAGCTGATTGCCCTTCACCGCCTGGCCCTGAGAGAGTTGCAGCAAGCGGTCCACGCTGGTTTGCCCC
AGCAGGCGAAAATCCTGTTTGATGGTGGTTAACGGCGGGATATAACATGAGCTGTCTTCGGTATCGTCGT
ATCCCACTACCGAGATATCCGCACCAACGCGCAGCCCGGACTCGGTAATGGCGCGCATTGCGCCCAGCGC
CATCTGATCGTTGGCAACCAGCATCGCAGTGGGAACGATGCCCTCATTCAGCATTTGCATGGTTTGTTGA
AAACCGGACATGGCACTCCAGTCGCCTTCCCGTTCCGCTATCGGCTGAATTTGATTGCGAGTGAGATATT
TATGCCAGCCAGCCAGACGCAGACGCGCCGAGACAGAACTTAATGGGCCCGCTAACAGCGCGATTTGCTG

GTGACCCAATGCGACCAGATGCTCCACGCCCAGTCGCGTACCGTCTTCATGGGAGAAAATAATACTGTTG
ATGGGTGTCTGGTCAGAGACATCAAGAAATAACGCCGGAACATTAGTGCAGGCAGCTTCCACAGCAATGG
CATCCTGGTCATCCAGCGGATAGTTAATGATCAGCCCACTGACGCGTTGCGCGAGAAGATTGTGCACCGC
CGCTTTACAGGCTTCGACGCCGCTTCGTTCTACCATCGACACCACCACGCTGGCACCCAGTTGATCGGCG
CGAGATTTAATCGCCGCGACAATTTGCGACGGCGCGTGCAGGGCCAGACTGGAGGTGGCAACGCCAATCA
GCAACGACTGTTTGCCCGCCAGTTGTTGTGCCACGCGGTTGGGAATGTAATTCAGCTCCGCCATCGCCGC
TTCCACTTTTTCCCGCGTTTTCGCAGAAACGTGGCTGGCCTGGTTCACCACGCGGGAAACGGTCTGATAA
GAGACACCGGCATACTCTGCGACATCGTATAACGTTACTGGTTTCACATTCACCACCCTGAATTGACTCT
CTTCCGGGCGCTATCATGCCATACCGCGAAAGGTTTTGCGCCATTCGATGGTGTCCGGGATCTCGACGCT
CTCCCTTATGCGACTCCTGCATTAGGAAGCAGCCCAGTAGTAGGTTGAGGCCGTTGAGCACCGCCGCCGC
AAGGAATGGTGCATGCAAGGAGATGGCGCCCAACAGTCCCCCGGCCACGGGGCCTGCCACCATACCCACG
CCGAAACAAGCGCTCATGAGCCCGAAGTGGCGAGCCCGATCTTCCCCATCGGTGATGTCGGCGATATAGG
CGCCAGCAACCGCACCTGTGGCGCCGGTGATGCCGGCCACGATGCGTCCGGCGTAGAGGATCGAGATCTC
GATCCCGCGAAATTAATACGACTCACTATAGGGGAATTGTGAGCGGATAACAATTCCCCTCTAGAAATAA
TTTTGTTTAACTTTAAGAAGGAGATATACCATGGGAGACCTCCCCTATCAGTGATAGAGATTGACATCCC
TATCAGTGATAGAGATACTGAGCACGGATCTTAGCTACTAGAGAAAGAGGAGAAATACTAGATGCGTAAA
GGCGAAGAGCTGTTCACTGGTGTCGTCCCTATTCTGGTGGAACTGGATGGTGATGTCAACGGTCATAAGT
TTTCCGTGCGTGGCGAGGGTGAAGGTGACGCAACTAATGGTAAACTGACGCTGAAGTTCATCTGTACTAC
TGGTAAACTGCCGGTACCTTGGCCGACTCTGGTAACGACGCTGACTTATGGTGTTCAGTGCTTTGCTCGT
TATCCGGACCATATGAAGCAGCATGACTTCTTCAAGTCCGCCATGCCGGAAGGCTATGTGCAGGAACGCA
CGATTTCCTTTAAGGATGACGGCACGTACAAAACGCGTGCGGAAGTGAAATTTGAAGGCGATACCCTGGT
AAACCGCATTGAGCTGAAAGGCATTGACTTTAAAGAAGACGGCAATATCCTGGGCCATAAGCTGGAATAC
AATTTTAACAGCCACAATGTTTACATCACCGCCGATAAACAAAAAAATGGCATTAAAGCGAATTTTAAAA
TTCGCCACAACGTGGAGGATGGATCTGTGCAGCTGGCTGATCACTACCAGCAAAACACTCCAATCGGTGA
TGGTCCTGTTCTGCTGCCAGACAATCACTATCTGAGCACGCAAAGCGTTCTGTCTAAAGATCCGAACGAG
AAACGCGATCATATGGTTCTGCTGGAGTTCGTAACCGCAGCGGGCATCACGCATGGTATGGATGAACTGT
ACAAATGAGGTCTCTTAAGAGCTCCGTCGACAAGCTTGCGGCCGCACTCGAGCACCACCACCACCACCAC
TGAGATCCGGCTGCTAACAAAGCCCGAAAGGAAGCTGAGTTGGCTGCTGCCACCGCTGAGCAATAACTAG
CATAACCCCTTGGGGCCTCTAAACGGGTCTTGAGGGGTTTTTTGCTGAAAGGAGGAACTATATCCGGAT

pET28a

TGGCGAATGGGACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTGACC
GCTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTCGCCG
GCTTTCCCCGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTCCGATTTAGTGCTTTACGGCACCTCGA
CCCCAAAAAACTTGATTAGGGTGATGGTTCACGTAGTGGGCCATCGCCCTGATAGACGGTTTTTCGCCCT
TTGACGTTGGAGTCCACGTTCTTTAATAGTGGACTCTTGTTCCAAACTGGAACAACACTCAACCCTATCT
CGGTCTATTCTTTTGATTTATAAGGGATTTTGCCGATTTCGGCCTATTGGTTAAAAAATGAGCTGATTTA
ACAAAAATTTAACGCGAATTTTAACAAAATATTAACGCTTACAATTTAGGTGGCACTTTTCGGGGAAATG
TGCGCGGAACCCCTATTTGTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCATGAATTAATTCTT
AGAAAAACTCATCGAGCATCAAATGAAACTGCAATTTATTCATATCAGGATTATCAATACCATATTTTTG
AAAAAGCCGTTTCTGTAATGAAGGAGAAAACTCACCGAGGCAGTTCCATAGGATGGCAAGATCCTGGTAT
CGGTCTGCGATTCCGACTCGTCCAACATCAATACAACCTATTAATTTCCCCTCGTCAAAAATAAGGTTAT
CAAGTGAGAAATCACCATGAGTGACGACTGAATCCGGTGAGAATGGCAAAAGTTTATGCATTTCTTTCCA
GACTTGTTCAACAGGCCAGCCATTACGCTCGTCATCAAAATCACTCGCATCAACCAAACCGTTATTCATT
CGTGATTGCGCCTGAGCGAGACGAAATACGCGATCGCTGTTAAAAGGACAATTACAAACAGGAATCGAAT
GCAACCGGCGCAGGAACACTGCCAGCGCATCAACAATATTTTCACCTGAATCAGGATATTCTTCTAATAC
CTGGAATGCTGTTTTCCCGGGGATCGCAGTGGTGAGTAACCATGCATCATCAGGAGTACGGATAAAATGC
TTGATGGTCGGAAGAGGCATAAATTCCGTCAGCCAGTTTAGTCTGACCATCTCATCTGTAACATCATTGG
CAACGCTACCTTTGCCATGTTTCAGAAACAACTCTGGCGCATCGGGCTTCCCATACAATCGATAGATTGT
CGCACCTGATTGCCCGACATTATCGCGAGCCCATTTATACCCATATAAATCAGCATCCATGTTGGAATTT
AATCGCGGCCTAGAGCAAGACGTTTCCCGTTGAATATGGCTCATAACACCCCTTGTATTACTGTTTATGT

AAGCAGACAGTTTTATTGTTCATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCACTGAGCGTCAGACC
CCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAA
AAAACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACT
GGCTTCAGCAGAGCGCAGATACCAAATACTGTCCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGA
ACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAA
GTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGG
GGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTGAGCTAT
GAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGG
AGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTC
TGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGG
CCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCCTGCGTTATCCCCTGATTC
TGTGGATAACCGTATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGCAGCCGAACGACCGAGCGCAGC
GAGTCAGTGAGCGAGGAAGCGGAAGAGCGCCTGATGCGGTATTTTCTCCTTACGCATCTGTGCGGTATTT
CACACCGCATATATGGTGCACTCTCAGTACAATCTGCTCTGATGCCGCATAGTTAAGCCAGTATACACTC
CGCTATCGCTACGTGACTGGGTCATGGCTGCGCCCCGACACCCGCCAACACCCGCTGACGCGCCCTGACG
GGCTTGTCTGCTCCCGGCATCCGCTTACAGACAAGCTGTGACCGTCTCCGGGAGCTGCATGTGTCAGAGG
TTTTCACCGTCATCACCGAAACGCGCGAGGCAGCTGCGGTAAAGCTCATCAGCGTGGTCGTGAAGCGATT
CACAGATGTCTGCCTGTTCATCCGCGTCCAGCTCGTTGAGTTTCTCCAGAAGCGTTAATGTCTGGCTTCT
GATAAAGCGGGCCATGTTAAGGGCGGTTTTTTCCTGTTTGGTCACTGATGCCTCCGTGTAAGGGGGATTT
CTGTTCATGGGGGTAATGATACCGATGAAACGAGAGAGGATGCTCACGATACGGGTTACTGATGATGAAC
ATGCCCGGTTACTGGAACGTTGTGAGGGTAAACAACTGGCGGTATGGATGCGGCGGGACCAGAGAAAAAT
CACTCAGGGTCAATGCCAGCGCTTCGTTAATACAGATGTAGGTGTTCCACAGGGTAGCCAGCAGCATCCT
GCGATGCAGATCCGGAACATAATGGTGCAGGGCGCTGACTTCCGCGTTTCCAGACTTTACGAAACACGGA
AACCGAAGACCATTCATGTTGTTGCTCAGGTCGCAGACGTTTTGCAGCAGCAGTCGCTTCACGTTCGCTC
GCGTATCGGTGATTCATTCTGCTAACCAGTAAGGCAACCCCGCCAGCCTAGCCGGGTCCTCAACGACAGG
AGCACGATCATGCGCACCCGTGGGGCCGCCATGCCGGCGATAATGGCCTGCTTCTCGCCGAAACGTTTGG
TGGCGGGACCAGTGACGAAGGCTTGAGCGAGGGCGTGCAAGATTCCGAATACCGCAAGCGACAGGCCGAT
CATCGTCGCGCTCCAGCGAAAGCGGTCCTCGCCGAAAATGACCCAGAGCGCTGCCGGCACCTGTCCTACG
AGTTGCATGATAAAGAAGACAGTCATAAGTGCGGCGACGATAGTCATGCCCCGCGCCCACCGGAAGGAGC
TGACTGGGTTGAAGGCTCTCAAGGGCATCGGTCGAGATCCCGGTGCCTAATGAGTGAGCTAACTTACATT
AATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGC
CAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCCAGGGTGGTTTTTCTTTTCACCAGTGAGACGGGC
AACAGCTGATTGCCCTTCACCGCCTGGCCCTGAGAGAGTTGCAGCAAGCGGTCCACGCTGGTTTGCCCCA
GCAGGCGAAAATCCTGTTTGATGGTGGTTAACGGCGGGATATAACATGAGCTGTCTTCGGTATCGTCGTA
TCCCACTACCGAGATATCCGCACCAACGCGCAGCCCGGACTCGGTAATGGCGCGCATTGCGCCCAGCGCC
ATCTGATCGTTGGCAACCAGCATCGCAGTGGGAACGATGCCCTCATTCAGCATTTGCATGGTTTGTTGAA
AACCGGACATGGCACTCCAGTCGCCTTCCCGTTCCGCTATCGGCTGAATTTGATTGCGAGTGAGATATTT
ATGCCAGCCAGCCAGACGCAGACGCGCCGAGACAGAACTTAATGGGCCCGCTAACAGCGCGATTTGCTGG
TGACCCAATGCGACCAGATGCTCCACGCCCAGTCGCGTACCGTCTTCATGGGAGAAAATAATACTGTTGA
TGGGTGTCTGGTCAGAGACATCAAGAAATAACGCCGGAACATTAGTGCAGGCAGCTTCCACAGCAATGGC
ATCCTGGTCATCCAGCGGATAGTTAATGATCAGCCCACTGACGCGTTGCGCGAGAAGATTGTGCACCGCC
GCTTTACAGGCTTCGACGCCGCTTCGTTCTACCATCGACACCACCACGCTGGCACCCAGTTGATCGGCGC
GAGATTTAATCGCCGCGACAATTTGCGACGGCGCGTGCAGGGCCAGACTGGAGGTGGCAACGCCAATCAG
CAACGACTGTTTGCCCGCCAGTTGTTGTGCCACGCGGTTGGGAATGTAATTCAGCTCCGCCATCGCCGCT
TCCACTTTTTCCCGCGTTTTCGCAGAAACGTGGCTGGCCTGGTTCACCACGCGGGAAACGGTCTGATAAG
AGACACCGGCATACTCTGCGACATCGTATAACGTTACTGGTTTCACATTCACCACCCTGAATTGACTCTC
TTCCGGGCGCTATCATGCCATACCGCGAAAGGTTTTGCGCCATTCGATGGTGTCCGGGATCTCGACGCTC
TCCCTTATGCGACTCCTGCATTAGGAAGCAGCCCAGTAGTAGGTTGAGGCCGTTGAGCACCGCCGCCGCA
AGGAATGGTGCATGCAAGGAGATGGCGCCCAACAGTCCCCCGGCCACGGGGCCTGCCACCATACCCACGC
CGAAACAAGCGCTCATGAGCCCGAAGTGGCGAGCCCGATCTTCCCCATCGGTGATGTCGGCGATATAGGC
GCCAGCAACCGCACCTGTGGCGCCGGTGATGCCGGCCACGATGCGTCCGGCGTAGAGGATCGAGATCTCG
ATCCCGCGAAATTAATACGACTCACTATAGGGGAATTGTGAGCGGATAACAATTCCCCTCTAGAAATAAT

TTTGTTTAACTTTAAGAAGGAGATATACCATGGGCAGCAGCCATCATCATCATCATCACAGCAGCGGCCT
GGTGCCGCGCGGCAGCCATATGGCTAGCATGACTGGTGGACAGCAAATGGGTCGCGGATCCGAATTCGAG
CTCCGTCGACAAGCTTGCGGCCGCACTCGAGCACCACCACCACCACCACTGAGATCCGGCTGCTAACAAA
GCCCGAAAGGAAGCTGAGTTGGCTGCTGCCACCGCTGAGCAATAACTAGCATAACCCCTTGGGGCCTCTA
AACGGGTCTTGAGGGGTTTTTTGCTGAAAGGAGGAACTATATCCGGAT

pET47b(+)

ATCCGGATATAGTTCCTCCTTTCAGCAAAAAACCCCTCAAGACCCGTTTAGAGGCCCCAAGGGGTTATGC
TAGTTATTGCTCAGCGGTGGCAGCAGCCAACTCAGCTTCCTTTCGGGCTTTGTTTAGCAGCCTAGGTTAA
TTAAGCCTCGAGAGCAGCAGAAGTAGAGCTGTCCATGTGCTGGCGTTCGAATTTAGCAGCAGCGGTTTCT
TTACTACCGCGTGGCACCAGAGCGAGCTCTGCGGCCGCAAGCTTGTCGACGGACGTCGGGCGCGCCAAGG
CCTGTACAGAATTCGGATCCTGGTACCCGGGTCCCTGAAAGAGGACTTCAAGAGCCGCGGAGTGATGGTG
GTGGTGATGTGCCATATGTATATCTCCTTCTTAAAGTTAAACAAAATTATTTCTAGAGGGGAATTGTTAT
CCGCTCACAATTCCCCTATAGTGAGTCGTATTAATTTCGCGGGATCGAGATCGATCTCGATCCTCTACGC
CGGACGCATCGTGGCCGGCATCACCGGCGCCACAGGTGCGGTTGCTGGCGCCTATATCGCCGACATCACC
GATGGGGAAGATCGGGCTCGCCACTTCGGGCTCATGAGCGCTTGTTTCGGCGTGGGTATGGTGGCAGGCC
CCGTGGCCGGGGGACTGTTGGGCGCCATCTCCTTGCATGCACCATTCCTTGCGGCGGCGGTGCTCAACGG
CCTCAACCTACTACTGGGCTGCTTCCTAATGCAGGAGTCGCATAAGGGAGAGCGTCGAGATCCCGGACAC
CATCGAATGGCGCAAAACCTTTCGCGGTATGGCATGATAGCGCCCGGAAGAGAGTCAATTCAGGGTGGTG
AATGTGAAACCAGTAACGTTATACGATGTCGCAGAGTATGCCGGTGTCTCTTATCAGACCGTTTCCCGCG
TGGTGAACCAGGCCAGCCACGTTTCTGCGAAAACGCGGGAAAAAGTGGAAGCGGCGATGGCGGAGCTGAA
TTACATTCCCAACCGCGTGGCACAACAACTGGCGGGCAAACAGTCGTTGCTGATTGGCGTTGCCACCTCC
AGTCTGGCCCTGCACGCGCCGTCGCAAATTGTCGCGGCGATTAAATCTCGCGCCGATCAACTGGGTGCCA
GCGTGGTGGTGTCGATGGTAGAACGAAGCGGCGTCGAAGCCTGTAAAGCGGCGGTGCACAATCTTCTCGC
GCAACGCGTCAGTGGGCTGATCATTAACTATCCGCTGGATGACCAGGATGCCATTGCTGTGGAAGCTGCC
TGCACTAATGTTCCGGCGTTATTTCTTGATGTCTCTGACCAGACACCCATCAACAGTATTATTTTCTCCC
ATGAAGACGGTACGCGACTGGGCGTGGAGCATCTGGTCGCATTGGGTCACCAGCAAATCGCGCTGTTAGC
GGGCCCATTAAGTTCTGTCTCGGCGCGTCTGCGTCTGGCTGGCTGGCATAAATATCTCACTCGCAATCAA
ATTCAGCCGATAGCGGAACGGGAAGGCGACTGGAGTGCCATGTCCGGTTTTCAACAAACCATGCAAATGC
TGAATGAGGGCATCGTTCCCACTGCGATGCTGGTTGCCAACGATCAGATGGCGCTGGGCGCAATGCGCGC
CATTACCGAGTCCGGGCTGCGCGTTGGTGCGGACATCTCGGTAGTGGGATACGACGATACCGAAGACAGC
TCATGTTATATCCCGCCGTTAACCACCATCAAACAGGATTTTCGCCTGCTGGGGCAAACCAGCGTGGACC
GCTTGCTGCAACTCTCTCAGGGCCAGGCGGTGAAGGGCAATCAGCTGTTGCCCGTCTCACTGGTGAAAAG
AAAAACCACCCTGGCGCCCAATACGCAAACCGCCTCTCCCCGCGCGTTGGCCGATTCATTAATGCAGCTG
GCACGACAGGTTTCCCGACTGGAAAGCGGGCAGTGAGCGCAACGCAATTAATGTAAGTTAGCTCACTCAT
TAGGCACCGGGATCTCGACCGATGCCCTTGAGAGCCTTCAACCCAGTCAGCTCCTTCCGGTGGGCGCGGG
GCATGACTAGCATGATCGTGCTCCTGTCGTTGAGGACCCGGCTAGGCTGGCGGGGTTGCCTTACTGGTTA
GCAGAATGAATCACCGATACGCGAGCGAACGTGAAGCGACTGCTGCTGCAAAACGTCTGCGACCTGAGCA
ACAACATGAATGGTCTTCGGTTTCCGTGTTTCGTAAAGTCTGGAAACGCGGAAGTCAGCGCCCTGCACCA
TTATGTTCCGGATCTGCATCGCAGGATGCTGCTGGCTACCCTGTGGAACACCTACATCTGTATTAACGAA
GCGCTGGCATTGACCCTGAGTGATTTTTCTCTGGTCCCGCCGCATCCATACCGCCAGTTGTTTACCCTCA
CAACGTTCCAGTAACCGGGCATGTTCATCATCAGTAACCCGTATCGTGAGCATCCTCTCTCGTTTCATCG
GTATCATTACCCCCATGAACAGAAATCCCCCTTACACGGAGGCATCAGTGACCAAACAGGAAAAAACCGC
CCTTAACATGGCCCGCTTTATCAGAAGCCAGACATTAACGCTTCTGGAGAAACTCAACGAGCTGGACGCG
GATGAACAGGCAGACATCTGTGAATCGCTTCACGACCACGCTGATGAGCTTTACCGCAGCTGCCTCGCGC
GTTTCGGTGATGACGGTGAAAACCTCTGACACATGCAGCTCCCGGAGACGGTCACAGCTTGTCTGTAAGC
GGATGCCGGGAGCAGACAAGCCCGTCAGGGCGCGTCAGCGGGTGTTGGCGGGTGTCGGGGCGCAGCCATG
ACCCAGTCACGTAGCGATAGCGGAGTGTATACTGGCTTAACTATGCGGCATCAGAGCAGATTGTACTGAG
AGTGCACCATATATGCGGTGTGAAATACCGCACAGATGCGTAAGGAGAAAATACCGCATCAGGCGCTCTT
CCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAA
GGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAA

AAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATC
ACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCC
TGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCT
TCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCA
AGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGA
GTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGG
TATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGGACAGTATTTG
GTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAAC
CACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAA
GATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCA
TGAACAATAAAACTGTCTGCTTACATAAACAGTAATACAAGGGGTGTTATGAGCCATATTCAACGGGAAA
CGTCTTGCTCTAGGCCGCGATTAAATTCCAACATGGATGCTGATTTATATGGGTATAAATGGGCTCGCGA
TAATGTCGGGCAATCAGGTGCGACAATCTATCGATTGTATGGGAAGCCCGATGCGCCAGAGTTGTTTCTG
AAACATGGCAAAGGTAGCGTTGCCAATGATGTTACAGATGAGATGGTCAGACTAAACTGGCTGACGGAAT
TTATGCCTCTTCCGACCATCAAGCATTTTATCCGTACTCCTGATGATGCATGGTTACTCACCACTGCGAT
CCCCGGCAAAACAGCATTCCAGGTATTAGAAGAATATCCTGATTCAGGTGAAAATATTGTTGATGCGCTG
GCAGTGTTCCTGCGCCGGTTGCATTCGATTCCTGTTTGTAATTGTCCTTTTAACAGTGATCGCGTATTTC
GTCTCGCTCAGGCGCAATCACGAATGAATAACGGTTTGGTTGATGCGAGTGATTTTGATGACGAGCGTAA
TGGCTGGCCTGTTGAACAAGTCTGGAAAGAAATGCATAAACTTTTGCCATTCTCACCGGATTCAGTCGTC
ACTCATGGTGATTTCTCACTTGATAACCTTATTTTTGACGAGGGGAAATTAATAGGTTGTATTGATGTTG
GACGAGTCGGAATCGCAGACCGATACCAGGATCTTGCCATCCTATGGAACTGCCTCGGTGAGTTTTCTCC
TTCATTACAGAAACGGCTTTTTCAAAAATATGGTATTGATAATCCTGATATGAATAAATTGCAGTTTCAT
TTGATGCTCGATGAGTTTTTCTAAGAATTAATTCATGAGCGGATACATATTTGAATGTATTTAGAAAAAT
AAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGAAATTGTAAACGTTAATATTTTGT
TAAAATTCGCGTTAAATTTTTGTTAAATCAGCTCATTTTTTAACCAATAGGCCGAAATCGGCAAAATCCC
TTATAAATCAAAAGAATAGACCGAGATAGGGTTGAGTGTTGTTCCAGTTTGGAACAAGAGTCCACTATTA
AAGAACGTGGACTCCAACGTCAAAGGGCGAAAAACCGTCTATCAGGGCGATGGCCCACTACGTGAACCAT
CACCCTAATCAAGTTTTTTGGGGTCGAGGTGCCGTAAAGCACTAAATCGGAACCCTAAAGGGAGCCCCCG
ATTTAGAGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAAGGAAGGGAAGAAAGCGAAAGGAGCGGGC
GCTAGGGCGCTGGCAAGTGTAGCGGTCACGCTGCGCGTAACCACCACACCCGCCGCGCTTAATGCGCCGC
TACAGGGCGCGTCCCATTCGCCA

TxTl T7 A1

ATCCCCAGGCATCAAATAAAACGAAAGGCTCAGTCGAAAGACTGGGCCTTTCGTTTTATCTGTTGTTTGT
CGGTGAACGCTCTCTACTAGAGTCACACTGGCTCACCTTCGGGTGGGCCTTTCTGCGTTTATAGCTGAGC
ATGAGACGGAAATCTGCTCGTCAGTGGTGCTCACACTGACGAATCATGTACAGATCATACCGATGACTGC
CTGGCGACTCACAACTAAGCAAGACAGCCGGAACCAGCGCCGGCGAACACCACTGCATATATGGCATATC
ACAACAGTCCAACTAGTGCACTGCAGTACATTATTCTTAGAAAAACTCATCGAGCATCAAATGAAACTGC
AATTTATTCATATCAGGATTATCAATACCATATTTTTGAAAAAGCCGTTTCTGTAATGAAGGAGAAAACT
CACCGAGGCAGTTCCATAGGATGGCAAGATCCTGGTATCGGTCTGCGATTCCGACTCGTCCAACATCAAT
ACAACCTATTAATTTCCCCTCGTCAAAAATAAGGTTATCAAGTGAGAAATCACCATGAGTGACGACTGAA
TCCGGTGAGAATGGCAAAAGTTTATGCATTTCTTTCCAGACTTGTTCAACAGGCCAGCCATTACGCTCGT
CATCAAAATCACTCGCATCAACCAAACCGTTATTCATTCGTGATTGCGCCTGAGCGAGACGAAATACGCG
ATCGCTGTTAAAAGGACAATTACAAACAGGAATCGAATGCAACCGGCGCAGGAACACTGCCAGCGCATCA
ACAATATTTTCACCTGAATCAGGATATTCTTCTAATACCTGGAATGCTGTTTTCCCGGGGATCGCAGTGG
TGAGTAACCATGCATCATCAGGAGTACGGATAAAATGCTTGATGGTCGGAAGAGGCATAAATTCCGTCAG
CCAGTTTAGTCTGACCATCTCATCTGTAACATCATTGGCAACGCTACCTTTGCCATGTTTCAGAAACAAC
TCTGGCGCATCGGGCTTCCCATACAATCGATAGATTGTCGCACCTGATTGCCCGACATTATCGCGAGCCC
ATTTATACCCATATAAATCAGCATCCATGTTGGAATTTAATCGCGGCCTAGAGCAAGACGTTTCCCGTTG
AATATGGCTCATAACACCCCTTGTATTACTGTTTATGTAAGCAGACAGTTTTATTGGAGTTTTTTCCATA
GGCTCCGCCCCCCTGACAAGCATCACGAAATCTGACGCTCAAATCAGTGGTGGCGAAACCCGACAGGACT

ATAAAGATACCAGGCGTTTCCCCCTGGCGGCTCCCTCGTGCGCTCTCCTGTTCCTGCCTTTCGGTTTACC
GGTGTCATTCCGCTGTTATGGCCGCGTTTGTCTCATTCCACGCCTGACACTCAGTTCCGGGTAGGCAGTT
CGCTCCAAGCTGGACTGTATGCACGAACCCCCCGTTCAGTCCGACCGCTGCGCCTTATCCGGTAACTATC
GTCTTGAGTCCAACCCGGAAAGACATGCAAAAGCACCACTGGCAGCAGCCACTGGTAATTGATTTAGAGG
AGTTAGTCTTGAAGTCATGCGCCGGTTAAGGCTAAACTGAAAGGACAAGTTTTGGTGACTGCGCTCCTCC
AAGCCAGTTACCTCGGTTCAAAGAGTTGGTAGCTCAGAGAACCTTCGAAAAACCGCCCTGCAAGGCGGTT
TTTTCGTTTTCAGAGCAAGAGATTACGCGCAGACCAAAACGATCTCAAGAAGATCATCTTATTCCCTGAA
TTCGCATCTAGACTGATGAGACGTGGTAGAGCCACAAACAGCCGGTACAAGCAACGATCTCCAGGACCAT
CTGAATCATGCGCGGATGACACGAACTCACGACGGCGATCACAGACATTAACCCACAGTACAGACACTGC
GACAACGTGGCAATTCGTCGCAATACAACGGAGGCCGTTGAGCACCGCCGCCGCAAGGAATGGTGCATGC
AAGGAGATGGCGCCCAACAGTCCCCCGGCCACGGGGCCTGCCACCATACCCACGCCGAAACAAGCGCTCA
TGAGCCCGAAGTGGCGAGCCCGATCTTCCCCATCGGTGATGTCGGCGATATAGGCGCCAGCAACCGCACC
TGTGGCGCCGGTGATGCCGGCCACGATGCGTCCGGCGTAGAGGATCGAGATCTCGATCCCGCGAAATTAA
TACGACTCACTATAGGGGAATTGTGAGCGGATAACAATTCCCCTCTAGAAATAATTTTGTTTAACTTTAA
GAAGGAGATATAT

TxTl T7 A2

ATCCCCAGGCATCAAATAAAACGAAAGGCTCAGTCGAAAGACTGGGCCTTTCGTTTTATCTGTTGTTTGT
CGGTGAACGCTCTCTACTAGAGTCACACTGGCTCACCTTCGGGTGGGCCTTTCTGCGTTTATAGCTGCCA
ATGAGACGACGGGGTCATCACGGCTCATCATGCGCCCAACAAATGTGTGCCATACACGCTCGGATGACTG
CCTGATGACCGCACTGACTGGGGACAGCCGATCCACCTAAGCCTGTGAGAGAAGCAGACACCCGACAGAT
CAAGGCAGTTAACTAGTGCACTGCAGTACATTATTCTTAGAAAAACTCATCGAGCATCAAATGAAACTGC
AATTTATTCATATCAGGATTATCAATACCATATTTTTGAAAAAGCCGTTTCTGTAATGAAGGAGAAAACT
CACCGAGGCAGTTCCATAGGATGGCAAGATCCTGGTATCGGTCTGCGATTCCGACTCGTCCAACATCAAT
ACAACCTATTAATTTCCCCTCGTCAAAAATAAGGTTATCAAGTGAGAAATCACCATGAGTGACGACTGAA
TCCGGTGAGAATGGCAAAAGTTTATGCATTTCTTTCCAGACTTGTTCAACAGGCCAGCCATTACGCTCGT
CATCAAAATCACTCGCATCAACCAAACCGTTATTCATTCGTGATTGCGCCTGAGCGAGACGAAATACGCG
ATCGCTGTTAAAAGGACAATTACAAACAGGAATCGAATGCAACCGGCGCAGGAACACTGCCAGCGCATCA
ACAATATTTTCACCTGAATCAGGATATTCTTCTAATACCTGGAATGCTGTTTTCCCGGGGATCGCAGTGG
TGAGTAACCATGCATCATCAGGAGTACGGATAAAATGCTTGATGGTCGGAAGAGGCATAAATTCCGTCAG
CCAGTTTAGTCTGACCATCTCATCTGTAACATCATTGGCAACGCTACCTTTGCCATGTTTCAGAAACAAC
TCTGGCGCATCGGGCTTCCCATACAATCGATAGATTGTCGCACCTGATTGCCCGACATTATCGCGAGCCC
ATTTATACCCATATAAATCAGCATCCATGTTGGAATTTAATCGCGGCCTAGAGCAAGACGTTTCCCGTTG
AATATGGCTCATAACACCCCTTGTATTACTGTTTATGTAAGCAGACAGTTTTATTGGAGTTTTTTCCATA
GGCTCCGCCCCCCTGACAAGCATCACGAAATCTGACGCTCAAATCAGTGGTGGCGAAACCCGACAGGACT
ATAAAGATACCAGGCGTTTCCCCCTGGCGGCTCCCTCGTGCGCTCTCCTGTTCCTGCCTTTCGGTTTACC
GGTGTCATTCCGCTGTTATGGCCGCGTTTGTCTCATTCCACGCCTGACACTCAGTTCCGGGTAGGCAGTT
CGCTCCAAGCTGGACTGTATGCACGAACCCCCCGTTCAGTCCGACCGCTGCGCCTTATCCGGTAACTATC
GTCTTGAGTCCAACCCGGAAAGACATGCAAAAGCACCACTGGCAGCAGCCACTGGTAATTGATTTAGAGG
AGTTAGTCTTGAAGTCATGCGCCGGTTAAGGCTAAACTGAAAGGACAAGTTTTGGTGACTGCGCTCCTCC
AAGCCAGTTACCTCGGTTCAAAGAGTTGGTAGCTCAGAGAACCTTCGAAAAACCGCCCTGCAAGGCGGTT
TTTTCGTTTTCAGAGCAAGAGATTACGCGCAGACCAAAACGATCTCAAGAAGATCATCTTATTCCCTGAA
TTCGCATCTAGACTGATGAGACGTGGTAGAGCCACAAACAGCCGGTACAAGCAACGATCTCCAGGACCAT
CTGAATCATGCGCGGATGACACGAACTCACGACGGCGATCACAGACATTAACCCACAGTACAGACACTGC
GACAACGTGGCAATTCGTCGCAATACAACGGAGGCCGTTGAGCACCGCCGCCGCAAGGAATGGTGCATGC
AAGGAGATGGCGCCCAACAGTCCCCCGGCCACGGGGCCTGCCACCATACCCACGCCGAAACAAGCGCTCA
TGAGCCCGAAGTGGCGAGCCCGATCTTCCCCATCGGTGATGTCGGCGATATAGGCGCCAGCAACCGCACC
TGTGGCGCCGGTGATGCCGGCCACGATGCGTCCGGCGTAGAGGATCGAGATCTCGATCCCGCGAAATTAA
TACGACTCACTATAGGGGAATTGTGAGCGGATAACAATTCCCCTCTAGAAATAATTTTGTTTAACTTTAA
GAAGGAGATATAT

TxTl T500 A3

CTCGAGGAATTCGACTCAATTAGTTCAGTCAGTTTCAGGATATTAGTCATCTCTACATTGATTATGAGTA
TTCAGAAATTCCTTAAATATTCTGACAAATGCTCTTTCCCTAAACTCCCCCCATAAAAAAACCCGCCGAA
GCGGGTTTTTACGTTATTTGCGGATTAACGATTACTCGTTATCAGAACCGCCCAGACCTGCGTTCAGCAG
TTCTGCCAGGCTGGCAGATGCGTCTTCCGAATTGATCCGTCGACCAAAGCCCGCCGAAAGGCGGGCTTTT
CTGTGCCGGCATGATAAGCTGTCAAACATGAGAATTACAACTTATATCGTATGGGGCTGACTTCAGGTGC
TACATTTGAAGAGATAAATTGCACTGAAATCTAGAAATATTTTATCTGATTAATAAGATGATCTTCTTGA
GATCGTTTTGGTCTGCGCGTAATCTCTTGCTCTGAAAACGAAAAAACCGCCTTGCAGGGCGGTTTTTCGA
AGGTTCTCTGAGCTACCAACTCTTTGAACCGAGGTAACTGGCTTGGAGGAGCGCAGTCACCAAAACTTGT
CCTTTCAGTTTAGCCTTAACCGGCGCATGACTTCAAGACTAACTCCTCTAAATCAATTACCAGTGGCTGC
TGCCAGTGGTGCTTTTGCATGTCTTTCCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGG
TCGGACTGAACGGGGGGTTCGTGCATACAGTCCAGCTTGGAGCGAACTGCCTACCCGGAACTGAGTGTCA
GGCGTGGAATGAGACAAACGCGGCCATAACAGCGGAATGACACCGGTAAACCGAAAGGCAGGAACAGGAG
AGCGCACGAGGGAGCCGCCAGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCACTGA
TTTGAGCGTCAGATTTCGTGATGCTTGTCAGGGGGGCGGAGCCTATGGAAAAACGGCTTTGCCGCGGCCC
TCTCACTTCCCTGTTAAGTATCTTCCTGGCATCTTCCAGGAAATCTCCGCCCCGTTCGTAAGCCATTTCC
GCTCGCCGCAGTCGAACGACCGAGCGTAGCGAGTCAGTGAGCGAGGAAGCGGAATATATCCTGTATCACA
TATTCTGCTGACGCACCGGTGCAGCCTTTTTTCTCCTGCCACATGAAGCACTTCACTGACACCCTCATCA
GTGCCAACATAGTAAGCCAGTATACACTCCGCTAGGGTCATGAGATTATCAAAAAGGATCTTCACCTAGA
TCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTA
CCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGACTC
CCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAG
ACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGG
TCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCA
GTTAATAGTTTGCGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGG
CTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGT
TAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCA
GCACTGCATAATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCA
AGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGC
GCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATC
TTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTT
TCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACG
GAAATGTTGAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATG
AGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAG
TGCCACCTGACGTCTAAGAAACCATTATTATCATGACATTAACCTATAAAAATAGGCGTATCACGAGGCC
CTTTCGTCTTCAAGAATTCTGGCGAATCCTCTGACCAGCCAGAAAACGACCTTTCTGTGGTGAAACCGGA
TGCTGCAATTCAGAGCGGCAGCAAGTGGGGGACAGCAGAAGACCTGACCGCCGCAGAGTGGATGTTTGAC
ATGGTGAAGACTATCGCACCATCAGCCAGAAAACCGAATTTTGCTGGGTGGGCTAACGATATCCGCCTGA
TGCGTGAACGTGACGGACGTAACCACCGCGACATGTGTGTGCTGTTCCGCTGGGCATGCCAGGACAACTT
CTGGTCCGGTAACGTGCTGAGCTAACACCGTGCGTGTTGACAATTTTACCTCTGGCGGTGATAATGGTTG
CAGCTAGCAATAATTTTGTTTAACTTTAAGAAGGAGGATCCAA

FKBP12-1.1

ATGGGCGGAGTGCAGGTGGAAACCATCTCCCCAGGAGACGGGCGCACCTTCCCCAAGCGCGGCCAGACCT
GCGTGGTGCACATTACCGGGATGCTTGAAGATGGAAAGAAATTTATTTCCTCCCGGGACAGAAACAAGCC
CTTTAAGTTTATGCTAGGCAAGCAGCGTGTGCTGCGAGGCTGGGAAGAAGGGGTTGCCCAGATGAGTGTG
GGTCAGAGAGCCAAACTGACTATATCTCCAGATTATGCCTTTGGTGCCACTGGGCACCCAGGCATCATCC
CACCACATGCCACTCTCGTCTTCGATGTGGAGCTTCTAAAACTGGAA

FRB-1.1

ATGGGTGTGGCCATCCTCTGGCATGAGATGTGGCATGAAGGCGCGGAAGAGGCAGCGCGTTTGTACCGTG
GGGAAAGGAACGTGAAAGGCATGTTTGAGGTGCTGGAGCCCTTGCATGCTATGATGGAACGGGGCCCCCA
GACTCTGAAGGAAACATCCTTTAATCAGGCCTATGGTCGAGATTTAATGGAGGCCCAAGAGTGGTGCAGG
AAGTACATGAAATCAGGGAATGTCAAGGACCTCTGGCAAGCCATGCTGCTCTATGCGCATGTGCGTGATC
GAATCTCA

RapF-1.1

ATGGGCAGCAGCAGCAGCATTGGCGAAAAGATTAACGAATGGTATATGTACATACGCCGATTCAGCATAC
CCGATGCAGCGTATTTGGCGTTTGAAATCGCGCAAGAGCTGGATCAAATGGAAGAAGATCAAGACCTTCA
TTTGTACTATTCACTGATGCTGTTTCGGGCGTATCTAATGGCCGAGTACCTTGAACCGTTAGAAAAAATG
AGGATTGAGGAACAGCCGAGACTGTCTGATCTGCTGCTTGAGATTGATAAAAAA

RapF-1.2

ATGGGCAGCAGCAGCAGCATTGGCGAAAAGATTAACGAATGGTATATGTACATACGCCGATTCAGCGTGC
CCGATGCAGCGTATTTGGGCTTTGAAATCAGCCAAGAGCTGGATCAAATGGAAGAAGATCAAGACCTTCA
TTTGTACTATTCACTGATGCTGTTTCGGGCGTATCTAATGCGTGAGTACCTTGAACCGTTAGAAAAAATG
AGGATTGAGGAACAGCCGAGACTGTCTGATCTGCTGCTTGAGATTGATAAAAAA

RapF-1.3

ATGGGCAGCAGCAGCAGCATTGGCGAAAAGATTAACGAATGGTATATGTACATACGCCGATTCAGCGTGC
CCGCGGCAGCGTATTTGGGCTTTGAAATCAGCCAAGAGCTGGATCAAATGGAAGAAGATCAAGACCTTCA
TTTGTACTATTCACTGATGCTGTTTCGGGCGTATCTAATGCGTGAGTACCTTGAACCGTTAGAAAAAATG
AGGATTGAGGAACAGCCGAGACTGTCTGATCTGCTGCTTGAGATTGATAAAAAA

RapF-1.4

ATGGGCAGCAGCAGCAGCATTGGCGAAAAGATTAACGAATTTTATATGTACATACGCCGATTCAGCATAC
CCGATGCAGCGTATTTGGCGTTTGAAATCGCGCAAGAGCTGGATCAAATGGAAGAAGATCAAGACCTTCA
TTTGTACTATTCACTGATGCTGTTTCGGGCGTATCTAATGGCCGAGTACCTTGAACCGTTAGAAAAAATG
AGGATTGAGGAACAGCCGAGACTGTCTGATCTGCTGCTTGAGATTGATAAAAAA

ComA-1.1

ATGGGTTCCTCTCAAAAAGAACAAGATGTGCTCACACCTAGAGAATGCCTGATTCTTCAAGAAGTTGAAA
AGGGATTTACAAACCAAGAAATCGCAGATGCCCTTCATTTACGTAAGAGCGCGATTGAAGCGAGCTTGAC
ATCGATTTTCAATAAGCTGAATGTCGGTTCACGGACGGAAGCGGTTTTGATTGCGAAATCAGACGGTGTA
CTT

ComA-1.2

ATGGGTTCCTCTCAAAAAGAACAAGATGTGCTCACACCTAGAGAATGCCTGATTCTTCAAGAAGTTGAAA
AGGGATTTACAAACCAAGAAATCGCAGATGCCCTTCACATGCGTAAGAGCGCGATTGAAGCGAGCTTGAC
ATCGATTTTCAATAAGCTGAATGTCGGTTCACGGACGGAAGCGGTTTTGATTGCGAAATCAGACGGTGTA
CTT

ComA-1.3

ATGGGTTCCTCTCAAAAAGAACAAGATGTGCTCACACCTAGAGAATGCCTGATTCTTCAAGAAGTTGAAA
AGGGATTTACAAACCAAGAAATCGCAGATGCCCTTCACATGCGTAAGAGCGCGATTGAAGCGAGCTTGAC
ATCGATTTTCGCGAAGCTGAATGTCGGTTCACGGACGGAAGCGGTTTTGATTGCGAAATCAGACGGTGTA
CTT

ComA-1.4

ATGGGTTCCTCTCAAAAAGAACAAGATGTGCTCACACCTAGAGAATGCCTGATTCTTCAAGAAGTTGAAA
AGGGATTTACAAACCAAGAAATCGCAGATGCCCTTCATTTACGTAAGAGCGCGATTGAAATGAGCTTGAC
ATCGATTTTCAATAAGCTGAATGTCGGTTCACGGACGGAAGCGGTTTTGATTGCGAAATCAGACGGTGTA
CTT

AR-1.1

SDLGRKLLEAARAGQDDEVRILMANGADVNAADNTGTTPLHLAAYSGHLEIVEVLLKHGADVDASDVFGF
TPLILAALWGHLEIVEVLLKNGADVNAMGSDGWTPLHAAAYFGYLEIVEVLLKHGADVNAQDKRGKTAFD
ISIDNGNEDLAEILQKLN

AR-1.2

SDLGRKLLEAARAGQDDEVRILMANGADVNAADNTGTTPLHLAAYSGHLEIVEVLLKHGADVDASDVFGL
TPLILAALWGHLEIVEVLLKNGADVNAMDSDGWTPLHAAAYFGYLEIVEVLLKHGADVNAQDKRGKTAFD
VSIDNGNEDLAEILQKLN

AR-1.3

SDLGRKLLEAARAGQDDEVRILMANGADVNAADNTGTTPLHLAAYSGHLEIVEVLLKHGADVDASDVFGF
TPLLLAALWGHLEIVEVLLKNGADVNAMGSDGWTPLHAAAWFGYLEIVEVLLKHGADVNAQDKRGKTAFD
ESIDNGNEDLAEILQKLN

AR-1.4

SDLGRKLLEAARAGQDDEVRILMANGADVNAADNTGTTPLHLAAYSGHLEIVEVLLKHGADVDASDVFGM
TPLFLAALWGHLEIVEVLLKNGADVNAMTSDGWTPLHAAAYFGYLEIVEVLLKHGADVNAQDKRGKTAFD
VSIDNGNEDLAEILQKLN

AR-2.5

SDLGRKLLEAARAGQDDEVRILMANGADVNAADNTGTTPLHLAAYSGHLEIVEVLLKHGADVDASDVFGF
TPLLLAALWGHLEIVEVLLKHGEDVNAMGSDGWTPLHAAAWFGYLEIVEVLLKHGADVNAQDKRGKTAFD
ESIDNGNEDLAEILQKLN

AR-2.6

SDLGRKLLEAARAGQDDEVRILMANGADVNAADNTGTTPLHLAAYSGHLEIVEVLLKHGADVDASDVFGF
TPLGLAALWGHLEIVEVLLKHGEDVNAMGSDGWTPLHAAAWFGYLEIVEVLLKHGADVNAQDKRGKTAFD
ESIDNGNEDLAEILQKLN

AR-2.7

SDLGRKLLEAARAGQDDEVRILMANGADVNAADNTGTTPLHLAAYSGHLEIVEVLLKHGADVDASDVFGF
TPLGLAALWGHLEIVEVLLKHGEDVNAMGSDGWTPLHAAAKFGYLEIVEVLLKHGADVNAQDKRGKTPFD
LAIDNGNEDIAEVLQKAA

AR-3.8

SDLGRKLLEAARAGQDDEVRILMANGADVNAADNTGTTPLHLAAYSGHLEIVEVLLKHGADVDASDVFGF
TPLGLAALWGHLEIVEVLLKHGEDVNAMGSDGWTPLHAAAKFGYLEIVEVLLKHGADVNAQDKKGLTPFD
LAILNGNEDIAEVLQKAA

MBP-1.1

KIEEGKLVIWINGDKGYNGLAEVGKKFEKDTGIKVTVEHPDKLEEKFPQVAATGDGPDIIFWAHDRFGGY
AQSGLLAEITPDKAFQDKLYPFTWDAVRYNGKLIAYPIAVEALSLIYNKDLLPNPPKTWEEIFALDKELK
AKGKSALMFNLQEPYFTWPLIAADGGYAFKYENGKYDIKDVGVDNAGAKAGLTFLVALIAAKAMNADTDY
SIAEAAFNKGETAMTINGPWAWSNIDTSKVNYGVTVLPTFKGQPSKPFVGVLSAGINAASPNKELAKEFL
ENYLLTDEGLEAVNKDKPLGAVALKSYEEELAKDPRIAATMENAQKGEIMPNIPQMSAFWYAVRTAVINA
ASG

MBP-1.2

KIEEGKLVIWINGDKGYNGLAEVGKKFEKDTGIKVTVEHPDKLEEKFPQVAATGDGPDIIFWAHDRFGGY
AQSGLLAEITPDKAFQDKLYPFTWDAVRYNGKLIAYPIAVEALSLIYNKDLLPNPPKTWEEIFALDKELK
AKGKSALMFNLQEPYFTWPLIAADGGYAFKYENGKYDIKDVGVDNAGAKAGLTFLVALIKAKHMNADTDY
SIAEAAFNKGETAMTINGPWAWSNIDTSKVNYGVTVLPTFKGQPSKPFVGVLSAGINAASPNKELAKEFL
ENYLLTDEGLEAVNKDKPLGAVALKSYEEELAKDPRIAATMENAQKGEIMPNIPQMSAFWYAVRTAVINA
ASG

MBP-1.3

KIEEGKLVIWINGDKGYNGLAEVGKKFEKDTGIKVTVEHPDKLEEKFPQVAATGDGPDIIFWAHDRFGGY
AQSGLLAEITPDKAFQDKLYPFTWDAVRYNGKLIAYPIAVEALSLIYNKDLLPNPPKTWEEIFALDKELK
AKGKSALMFNLQEPYFTWPLIAADGGYAFKYENGKYDIKDVGVDNAGAKAGLTRLVYLIAAKAMNADTDY
SIAEAAFNKGETAMTINGPWAWSNIDTSKVNYGVTVLPTFKGQPSKPFVGVLSAGINAASPNKELAKEFL
ENYLLTDEGLEAVNKDKPLGAVALKSYEEELAKDPRIAATMENAQKGEIMPNIPQMSAFWYAVRTAVINA
ASG

MBP-1.4

KIEEGKLVIWINGDKGYNGLAEVGKKFEKDTGIKVTVEHPDKLEEKFPQVAATGDGPDIIFWAHDRFGGY
AQSGLLAEITPDKAFQDKLYPFTWDAVRYNGKLIAYPIAVEALSLIYNKDLLPNPPKTWEEIFALDKELK
AKGKSALMFNLQEPYFTWPLIAADGGYAFKYENGKYDIKDVGVDNAGAKAGLTFLVALIKAKAMNADTDY
SIAEAAFNKGETAMTINGPWAWSNIDTSKVNYGVTVLPTFKGQPSKPFVGVLSAGINAASPNKELAKEFL
ENYLLTDEGLEAVNKDKPLGAVALKSYEEELAKDPRIAATMENAQKGEIMPNIPQMSAFWYAVRTAVINA
ASG

MBP-2.5

KIEEGKLVIWINGDKGYNGLAEVGKKFEKDTGIKVTVEHPDKLEEKFPQVAATGDGPDIIFWAHDRFGGY
AQSGLLAEITPDKAFQDKLYPFTWDAVRYNGKLIAYPIAVEALSLIYNKDLLPNPPKTWEEIFALDKELK
AKGKSALMFNLQEPYFTWPLIAADGGYAFKYENGKYDIKDVGVDNAGAKAGLTALVYLIAAKAMNADTDY
SIAEAAFNKGETAMTINGPWAWSNIDTSKVNYGVTVLPTFKGQPSKPFVGVLSAGINAASPNKELAKEFL
ENYLLTDEGLEAVNKDKPLGAVALKSYEEELAKDPRIAATMENAQKGEIMPNIPQMSAFWYAVRTAVINA
ASGRQTVDEALKDAQTRITK

MBP-3.6

KIEEGKLVIWINGDKGYNGLAEVGKKFEKDTGIKVTVEHPDKLEEKFPQVAATGDGPDIIFWAHDRFGGY
AQSGLLAEITPDKAFQDKLYPFTWDAVRYNGKLIAYPIAVEALSLIYNKDLLPNPPKTWEEIFALDKELK
AKGKSALMFNLQEPYFTWPLIAADGGYAFKYENGKYDIKDVGVDNAGAKAGLTALVALIAAKAMNADTDY
SIAEAAFNKGETAMTINGPWAWSNIDTSKVNYGVTVLPTFKGQPSKPFVGVLSAGINAASPNKELAKEFL
ENYLLTDEGLEAVNKDKPLGAVALKSYEEELAKDPRIAATMENAQKGEIMPNIPQMSAFWYAVRTAVINA
ASGRQTVDEALKDAQTRITK

ERG12

ATGTCATTACCGTTCTTAACTTCTGCACCGGGAAAGGTTATTATTTTTGGTGAACACTCTGCTGTGTACA
ACAAGCCTGCCGTCGCTGCTAGTGTGTCTGCGTTGAGAACCTACCTGCTAATAAGCGAGTCATCTGCACC
AGATACTATTGAATTGGACTTCCCGGACATTAGCTTTAATCATAAGTGGTCCATCAATGATTTCAATGCC
ATCACCGAGGATCAAGTAAACTCCCAAAAATTGGCCAAGGCTCAACAAGCCACCGATGGCTTGTCTCAGG
AACTCGTTAGTCTTTTGGACCCGTTGTTAGCTCAACTATCCGAATCCGCCCACTACCATGCAGCGTTTTG
TTTCCTGTATATGTTTGTTTGCCTATGCCCCCATGCCAAGAATATTAAGTTTTCTTTAAAGTCTACTTTA
CCCATCGGTGCTGGGTTGGGCTCAAGCGCCTCTATTTCTGTATCACTGGCCTTAGCTATGGCCTACTTGG
GGGGGTTAATAGGATCTAATGACTTGGAAAAGCTGTCAGAAAACGATAAGCATATAGTGAATCAATGGGC
CTTCATAGGTGAAAAGTGTGCTCACGGTACCCCTTCAGGAATAGATAACGCTGTGGCCACTTATGGTAAT
GCCCTGCTATTTGAAAAAGACTCACATAATGGAACAATAAACACAAACAATTTTAAGTTCTTAGATGATT
TCCCAGCCATTCCAATGATCCTAACCTATACTAGAATCCCAAGGTCTACAAAAGACCTTGTTGCTCGCGT
TCGTGTGTTGGTCACCGAGAAATTTCCTGAAGTTATGAAGCCAATTCTAGATGCCATGGGTGAATGTGCC
CTACAAGGCTTAGAGATCATGACTAAGTTAAGTAAATGTAAAGGCACCGATGACGAGGCTGTAGAAACTA
ATAATGAACTGTATGAACAACTATTGGAATTGATAAGAATAAATCATGGACTGCTTGTCTCAATCGGTGT
TTCTCATCCTGGATTAGAACTTATTAAAAATCTGAGCGATGATTTGAGAATTGGCTCCACAAAACTTACC
GGTGCTGGTGGCGGCGGTTGCTCTTTGACTTTGTTACGAAGAGACATTACTCAAGAGCAAATTGACAGCT
TCAAAAAGAAATTGCAAGATGATTTTAGTTACGAGACATTTGAAACAGACTTGGGTGGGACTGGCTGCTG
TTTGTTAAGCGCAAAAAATTTGAATAAAGACCTTAAAATCAAATCCCTAGTATTCCAATTATTTGAAAAT
AAAACTACCACAAAGCAACAAATTGACGATCTATTATTGCCAGGAAACACGAATTTACCATGGACTTCA

ERG8

ATGTCAGAGTTGAGAGCCTTCAGTGCCCCAGGGAAAGCGTTACTAGCTGGTGGATATTTAGTTTTAGATA
CAAAATATGAAGCATTTGTAGTCGGATTATCGGCAAGAATGCATGCTGTAGCCCATCCTTACGGTTCATT
GCAAGGGTCTGATAAGTTTGAAGTGCGTGTGAAAAGTAAACAATTTAAAGATGGGGAGTGGCTGTACCAT
ATAAGTCCTAAAAGTGGCTTCATTCCTGTTTCGATAGGCGGATCTAAGAACCCTTTCATTGAAAAAGTTA
TCGCTAACGTATTTAGCTACTTTAAACCTAACATGGACGACTACTGCAATAGAAACTTGTTCGTTATTGA
TATTTTCTCTGATGATGCCTACCATTCTCAGGAGGATAGCGTTACCGAACATCGTGGCAACAGAAGATTG
AGTTTTCATTCGCACAGAATTGAAGAAGTTCCCAAAACAGGGCTGGGCTCCTCGGCAGGTTTAGTCACAG
TTTTAACTACAGCTTTGGCCTCCTTTTTTGTATCGGACCTGGAAAATAATGTAGACAAATATAGAGAAGT
TATTCATAATTTAGCACAAGTTGCTCATTGTCAAGCTCAGGGTAAAATTGGAAGCGGGTTTGATGTAGCG
GCGGCAGCATATGGATCTATCAGATATAGAAGATTCCCACCCGCATTAATCTCTAATTTGCCAGATATTG
GAAGTGCTACTTACGGCAGTAAACTGGCGCATTTGGTTGATGAAGAAGACTGGAATATTACGATTAAAAG

TAACCATTTACCTTCGGGATTAACTTTATGGATGGGCGATATTAAGAATGGTTCAGAAACAGTAAAACTG
GTCCAGAAGGTAAAAAATTGGTATGATTCGCATATGCCAGAAAGCTTGAAAATATATACAGAACTCGATC
ATGCAAATTCTAGATTTATGGATGGACTATCTAAACTAGATCGCTTACACGAGACTCATGACGATTACAG
CGATCAGATATTTGAGTCTCTTGAGAGGAATGACTGTACCTGTCAAAAGTATCCTGAAATCACAGAAGTT
AGAGATGCAGTTGCCACAATTAGACGTTCCTTTAGAAAAATAACTAAAGAATCTGGTGCCGATATCGAAC
CTCCCGTACAAACTAGCTTATTGGATGATTGCCAGACCTTAAAAGGAGTTCTTACTTGCTTAATACCTGG
TGCTGGTGGTTATGACGCCATTGCAGTGATTACTAAGCAAGATGTTGATCTTAGGGCTCAAACCGCTAAT
GACAAAAGATTTTCTAAGGTTCAATGGCTGGATGTAACTCAGGCTGACTGGGGTGTTAGGAAAGAAAAAG
ATCCGGAAACTTATCTTGATAAA

MVD1

ATGACCGTTTACACAGCATCCGTTACCGCACCCGTCAACATCGCAACCCTTAAGTATTGGGGGAAAAGGG
ACACGAAGTTGAATCTGCCCACCAATTCGTCCATATCAGTGACTTTATCGCAAGATGACCTCAGAACGTT
GACCTCTGCGGCTACTGCACCTGAGTTTGAACGCGACACTTTGTGGTTAAATGGAGAACCACACAGCATC
GACAATGAAAGAACTCAAAATTGTCTGCGCGACCTACGCCAATTAAGAAAGGAAATGGAATCGAAGGACG
CCTCATTGCCCACATTATCTCAATGGAAACTCCACATTGTCTCCGAAAATAACTTTCCTACAGCAGCTGG
TTTAGCTTCCTCCGCTGCTGGCTTTGCTGCATTGGTCTCTGCAATTGCTAAGTTATACCAATTACCACAG
TCAACTTCAGAAATATCTAGAATAGCAAGAAAGGGGTCTGGTTCAGCTTGTAGATCGTTGTTTGGCGGAT
ACGTGGCCTGGGAAATGGGAAAAGCTGAAGATGGTCATGATTCCATGGCAGTACAAATCGCAGACAGCTC
TGACTGGCCTCAGATGAAAGCTTGTGTCCTAGTTGTCAGCGATATTAAAAAGGATGTGAGTTCCACTCAG
GGTATGCAATTGACCGTGGCAACCTCCGAACTATTTAAAGAAAGAATTGAACATGTCGTACCAAAGAGAT
TTGAAGTCATGCGTAAAGCCATTGTTGAAAAAGATTTCGCCACCTTTGCAAAGGAAACAATGATGGATTC
CAACTCTTTCCATGCCACATGTTTGGACTCTTTCCCTCCAATATTCTACATGAATGACACTTCCAAGCGT
ATCATCAGTTGGTGCCACACCATTAATCAGTTTTACGGAGAAACAATCGTTGCATACACGTTTGATGCAG
GTCCAAATGCTGTGTTGTACTACTTAGCTGAAAATGAGTCGAAACTCTTTGCATTTATCTATAAATTGTT
TGGCTCTGTTCCTGGATGGGACAAGAAATTTACTACTGAGCAGCTTGAGGCTTTCAACCATCAATTTGAA
TCATCTAACTTTACTGCACGTGAATTGGATCTTGAGTTGCAAAAGGATGTTGCCAGAGTGATTTTAACTC
AAGTCGGTTCAGGCCCACAAGAAACAAACGAATCTTTGATTGACGCAAAGACTGGTCTACCAAAGGAA

idi

ATGATAATGCAAACGGAACACGTCATTTTATTGAATGCACAGGGAGTTCCCACGGGTACGCTGGAAAAGT
ATGCCGCACACACGGCAGACACCCGCTTACATCTCGCGTTCTCCAGTTGGCTGTTTAATGCCAAAGGACA
ATTATTAGTTACCCGCCGCGCACTGAGCAAAAAAGCATGGCCTGGCGTGTGGACTAACTCGGTTTGTGGG
CACCCACAACTGGGAGAAAGCAACGAAGACGCAGTGATCCGCCGTTGCCGTTATGAGCTTGGCGTGGAAA
TTACGCCTCCTGAATCTATCTATCCTGACTTTCGCTACCGCGCCACCGATCCGAGTGGCATTGTGGAAAA
TGAAGTGTGTCCGGTATTTGCCGCACGCACCACTAGTGCGTTACAGATCAATGATGATGAAGTGATGGAT
TATCAATGGTGTGATTTAGCAGATGTATTACACGGTATTGATGCCACGCCGTGGGCGTTCAGTCCGTGGA
TGGTGATGCAGGCGACAAATCGCGAAGCCAGAAAACGATTATCTGCATTTACCCAGCTTAAA

ispA

ATGGACTTTCCGCAGCAACTCGAAGCCTGCGTTAAGCAGGCCAACCAGGCGCTGAGCCGTTTTATCGCCC
CACTGCCCTTTCAGAACACTCCCGTGGTCGAAACCATGCAGTATGGCGCATTATTAGGTGGTAAGCGCCT
GCGACCTTTCCTGGTTTATGCCACCGGTCATATGTTCGGCGTTAGCACAAACACGCTGGACGCACCCGCT
GCCGCCGTTGAGTGTATCCACGCTTACTCATTAATTCATGATGATTTACCGGCAATGGATGATGACGATC
TGCGTCGCGGTTTGCCAACCTGCCATGTGAAGTTTGGCGAAGCAAACGCGATTCTCGCTGGCGACGCTTT
ACAAACGCTGGCGTTCTCGATTTTAAGCGATGCCGATATGCCGGAAGTGTCGGACCGCGACAGAATTTCG
ATGATTTCTGAACTGGCGAGCGCCAGTGGTATTGCCGGAATGTGCGGTGGTCAGGCATTAGATTTAGACG
CGGAAGGCAAACACGTACCTCTGGACGCGCTTGAGCGTATTCATCGTCATAAAACCGGCGCATTGATTCG
CGCCGCCGTTCGCCTTGGTGCATTAAGCGCCGGAGATAAAGGACGTCGTGCTCTGCCGGTACTCGACAAG

TATGCAGAGAGCATCGGCCTTGCCTTCCAGGTTCAGGATGACATCCTGGATGTGGTGGGAGATACTGCAA
CGTTGGGAAAACGCCAGGGTGCCGACCAGCAACTTGGTAAAAGTACCTACCCTGCACTTCTGGGCCTTGA
GCAAGCCCGGAAGAAAGCCCGGGATCTGATCGACGATGCCCGTCAGTCGCTGAAACAACTGGCTGAACAG
TCACTCGATACCTCGGCACTGGAAGCGCTAGCGGACTACATCATCCAGCGTAATAAA

ADS

ATGGCCCTGACCGAAGAGAAACCGATCCGCCCGATCGCTAACTTCCCGCCGTCTATCTGGGGTGACCAGT
TCCTGATCTACGAAAAGCAGGTTGAGCAGGGTGTTGAACAGATCGTAAACGACCTGAAGAAAGAAGTTCG
TCAGCTGCTGAAAGAAGCTCTGGACATCCCGATGAAACACGCTAACCTGCTGAAACTGATCGACGAGATC
CAGCGTCTGGGTATCCCGTACCACTTCGAACGCGAAATCGACCACGCACTGCAGTGCATCTACGAAACCT
ACGGCGACAACTGGAACGGCGACCGTTCTTCTCTGTGGTTTCGTCTGATGCGTAAACAGGGCTACTACGT
TACCTGTGACGTTTTTAACAACTACAAGGACAAGAACGGTGCTTTCAAACAGTCTCTGGCTAACGACGTT
GAAGGCCTGCTGGAACTGTACGAAGCGACCTCCATGCGTGTACCGGGTGAAATCATCCTGGAGGACGCGC
TGGGTTTCACCCGTTCTCGTCTGTCCATTATGACTAAAGACGCTTTCTCTACTAACCCGGCTCTGTTCAC
CGAAATCCAGCGTGCTCTGAAACAGCCGCTGTGGAAACGTCTGCCGCGTATCGAAGCAGCACAGTACATT
CCGTTTTACCAGCAGCAGGACTCTCACAACAAGACCCTGCTGAAACTGGCTAAGCTGGAGTTCAACCTGC
TGCAGTCTCTGCACAAAGAAGAACTGTCTCACGTTTGTAAGTGGTGGAAGGCATTTGACATCAAGAAAAA
CGCGCCGTGCCTGCGTGACCGTATCGTTGAATGTTACTTCTGGGGTCTGGGTTCTGGTTATGAACCACAG
TACTCCCGTGCACGTGTGTTCTTCACTAAAGCTGTAGCTGTTATCACCCTGATCGATGACACTTACGATG
CTTACGGCACCTACGAAGAACTGAAGATTTTTACTGAAGCTGTAGAACGCTGGTCTATCACTTGCCTGGA
CACTCTGCCGGAGTACATGAAACCGATCTACAAACTGTTCATGGATACCTACACCGAAATGGAGGAGTTC
CTGGCAAAAGAAGGCCGTACCGACCTGTTCAACTGCGGTAAAGAGTTTGTTAAAGAGTTCGTACGTAACC
TGATGGTTGAAGCTAAATGGGCTAACGAAGGCCATATCCCGACTACCGAAGAACATGACCCGGTTGTTAT
CATCACCGGCGGTGCAAACCTGCTGACCACCACTTGCTATCTGGGTATGTCCGACATCTTTACCAAGGAA
TCTGTTGAATGGGCTGTTTCTGCACCGCCGCTGTTCCGTTACTCCGGTATTCTGGGTCGTCGTCTGAACG
ACCTGATGACCCACAAAGCAGAGCAGGAACGTAAACACTCTTCCTCCTCTCTGGAATCCTACATGAAGGA
ATATAACGTTAACGAGGAGTACGCACAGACTCTGATCTATAAAGAAGTTGAAGACGTATGGAAAGACATC
AACCGTGAATACCTGACTACTAAAAACATCCCGCGCCCGCTGCTGATGGCAGTAATCTACCTGTGCCAGT
TCCTGGAAGTACAGTACGCTGGTAAAGATAACTTCACTCGCATGGGCGACGAATACAAACACCTGATCAA
ATCCCTGCTGGTTTACCCGATGTCCATC

nDHFR

ATGGTTCGACCATTGAACTGCATCGTCGCCGTGTCCCAAAATATGGGGATTGGCAAGAACGGAGACCTAC
CCTGGCCTCCGCTCAGGAACGAGTTCAAGTACTTCCAAAGAATGACCACAACCTCTTCAGTGGAAGGTAA
ACAGAATCTGGTGATTATGGGTAGGAAAACCTGGTTCTCCATTCCTGAGAAGAATCGACCTTTAAAGGAC
AGAATTAATATAGTTCTCAGTAGAGAACTCAAAGAACCACCACGAGGAGCTCATTTTCTTGCCAAAAGTT
TGGATGATGCCTTAAGACTTATTGAACAACCGGAATTGGGTACC

cDHFR

ATGAGTAAAGTAGACATGGTTTGGATAGTCGGAGGCAGTTCTGTTTACCAGGAAGCCATGAATCAACCAG
GCCACCTCAGACTCTTTGTGACAAGGATCATGCAGGAATTTGAAAGTGACACGTTTTTCCCAGAAATTGA
TTTGGGGAAATATAAACTTCTCCCAGAATACCCAGGCGTCCTCTCTGAGGTCCAGGAGGAAAAAGGCATC
AAGTATAAGTTTGAAGTCTACGAGAAGAAAGAC

nDHFR fusion linker

GGGAGCAGTGCTAGCGGAACTTCTAGCACTAGTTCCGGAATT

cDHFR fusion linker

GGTGGCTCTGGCAGTGGAGCTAGCACT

LgBIT

ATGGTCTTCACACTCGAAGATTTCGTTGGGGACTGGGAACAGACAGCCGCCTACAACCTGGACCAAGTCC
TTGAACAGGGAGGTGTGTCCAGTTTGCTGCAGAATCTCGCCGTGTCCGTAACTCCGATCCAAAGGATTGT
CCGGAGCGGTGAAAATGCCCTGAAGATCGACATCCATGTCATCATCCCGTATGAAGGTCTGAGCGCCGAC
CAAATGGCCCAGATCGAAGAGGTGTTTAAGGTGGTGTACCCTGTGGATGATCATCACTTTAAGGTGATCC
TGCCCTATGGCACACTGGTAATCGACGGGGTTACGCCGAACATGCTGAACTATTTCGGACGGCCGTATGA
AGGCATCGCCGTGTTCGACGGCAAAAAGATCACTGTAACAGGGACCCTGTGGAACGGCAACAAAATTATC
GACGAGCGCCTGATCACCCCCGACGGCTCCATGCTGTTCCGAGTAACCATCAACAGC

SmBIT

ATGGTGACCGGCTACCGGCTGTTCGAGGAGATTCTG

LgBIT and ddFP fusion linker

GGTAGCGGCAGCGGCAG

SmBIT fusion linker

GGTAGCGGCAGCGGCAGGGTAGCGGCTTCT

ddRFPb

ATGGTGAGCAAGGGCGAGGAGACCATCAAAGAGTTCATGCGCTTCAAGGTGCGCATGGAGGGCTCCATGA
ACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGCACCCAGACCGCCAAGCT
GAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATCCTGTCCCCCCAGTTCATGTACGGCTCC
GAGGCGTACGTGAGGCACCCCGCCGACATCCCCGATTACAAGAAGCTGCCCTTCCCCGAGGGCTTCAAGT
GGGAGCGCGTGATGAACTTCGAAGACGGCGGTCTGGTGACCGTTACCCAGGACTCCTCCCTGCAGGACGG
CACGCTGATCTGCAAGGTGAAGATGCGCGGCACCAACTTCCCCCCCGACGGCCCCGTAATGCAGAAGAAG
ACCATGGGCTGGGAGGCCTCCACCGAGATGCTGTACCCCGAAGACGGCGTGCTGAAGGGCCATAGCTATC
AGGCCCTGAAGCTGAAGGACGGCGGCCACTACCTGGTGGAGTTCGAGACCATCTACATGGCCAAGAAGCC
CGTGCAACTGCCCGGCGATTACTGTGTGGACACCAAGCTGGACATCACCTCCCACAACGAGGACTACACC
ATCGTGGAACAGTACGGGCGCTCCGAGGGCCGCCACCGTCTGGGCATGGACGAGCGGTACAAG

ddGFPa

ATGGCGAGCAAGAGCGAGGAGGTCATCAAAGAGTTCATGCGCTTCAAGGTGCGCTTGGAGGGCTCCATGA
ACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCTTACGAGGGCACCCAGACCGCCAAGCT
GAAGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATCCTGTCCCCCCTGATCATGTACGGCTCC
AAGATGTACGTGAAGCACCCCGCCGACGTCCCCGATTACATGAAGCTGTCCTTCCCCGAGGGCTTCAAGT
GGGAGCGCGTGATGCACTTCGAGGACGGCGGTCTGGTGACCGCGACACAGGACTCCTCCCTGCAGGACGG
CACGCTGATCTACAAGGTGAAGATGCGCGGCACCAACTTCCCCCCCGACGGCCCCGTAATGCAGAAGAAG
ACCTTGGGCTGGGATTATGCCACCGAGCGCCTGTACCCCGAAGAAGGCGTGCTGAAGGGCGAGCTCCTGG
GGCGCCTGAAGCTGAAGGACGGCGGCCTCAACCTGGTGGAGTCCAAGACCATCTACATGGCCAAGAAGCC
CGTGCAACTGCCCGGCTACTACTTCGTGGACACCAAGCTGGACATCACCTCCCACAACGAGGACTACACC
ATCGTGGAACAGTACGAGCGCTCCGAGGGCCGCCACCACCTGGGGCATGGTACAGGTAGCACAGGCAGC

T7 RNA polymerase

ATGAACACGATTAACATCGCTAAGAACGACTTCTCTGACATCGAACTGGCTGCTATCCCGTTCAACACTC
TGGCTGACCATTACGGTGAGCGTTTAGCTCGCGAACAGTTGGCCCTTGAGCATGAGTCTTACGAGATGGG
TGAAGCACGCTTCCGCAAGATGTTTGAGCGTCAACTTAAAGCTGGTGAGGTTGCGGATAACGCTGCCGCC
AAGCCTCTCATCACTACCCTACTCCCTAAGATGATTGCACGCATCAACGACTGGTTTGAGGAAGTGAAAG
CTAAGCGCGGCAAGCGCCCGACAGCCTTCCAGTTCCTGCAAGAAATCAAGCCGGAAGCCGTAGCGTACAT
CACCATTAAGACCACTCTGGCTTGCCTAACCAGTGCTGACAATACAACCGTTCAGGCTGTAGCAAGCGCA
ATCGGTCGGGCCATTGAGGACGAGGCTCGCTTCGGTCGTATCCGTGACCTTGAAGCTAAGCACTTCAAGA
AAAACGTTGAGGAACAACTCAACAAGCGCGTAGGGCACGTCTACAAGAAAGCATTTATGCAAGTTGTCGA
GGCTGACATGCTCTCTAAGGGTCTACTCGGTGGCGAGGCGTGGTCTTCGTGGCATAAGGAAGACTCTATT
CATGTAGGAGTACGCTGCATCGAGATGCTCATTGAGTCAACCGGAATGGTTAGCTTACACCGCCAAAATG
CTGGCGTAGTAGGTCAAGACTCTGAGACTATCGAACTCGCACCTGAATACGCTGAGGCTATCGCAACCCG
TGCAGGTGCGCTGGCTGGCATCTCTCCGATGTTCCAACCTTGCGTAGTTCCTCCTAAGCCGTGGACTGGC
ATTACTGGTGGTGGCTATTGGGCTAACGGTCGTCGTCCTCTGGCGCTGGTGCGTACTCACAGTAAGAAAG
CACTGATGCGCTACGAAGACGTTTACATGCCTGAGGTGTACAAAGCGATTAACATTGCGCAAAACACCGC
ATGGAAAATCAACAAGAAAGTCCTAGCGGTCGCCAACGTAATCACCAAGTGGAAGCATTGTCCGGTCGAG
GACATCCCTGCGATTGAGCGTGAAGAACTCCCGATGAAACCGGAAGACATCGACATGAATCCTGAGGCTC
TCACCGCGTGGAAACGTGCTGCCGCTGCTGTGTACCGCAAGGACAAGGCTCGCAAGTCTCGCCGTATCAG
CCTTGAGTTCATGCTTGAGCAAGCCAATAAGTTTGCTAACCATAAGGCCATCTGGTTCCCTTACAACATG
GACTGGCGCGGTCGTGTTTACGCTGTGTCAATGTTCAACCCGCAAGGTAACGATATGACCAAAGGACTGC
TTACGCTGGCGAAAGGTAAACCAATCGGTAAGGAAGGTTACTACTGGCTGAAAATCCACGGTGCAAACTG
TGCGGGTGTCGATAAGGTTCCGTTCCCTGAGCGCATCAAGTTCATTGAGGAAAACCACGAGAACATCATG
GCTTGCGCTAAGTCTCCACTGGAGAACACTTGGTGGGCTGAGCAAGATTCTCCGTTCTGCTTCCTTGCGT
TCTGCTTTGAGTACGCTGGGGTACAGCACCACGGCCTGAGCTATAACTGCTCCCTTCCGCTGGCGTTTGA
CGGGTCTTGCTCTGGCATCCAGCACTTCTCCGCGATGCTCCGAGATGAGGTAGGTGGTCGCGCGGTTAAC
TTGCTTCCTAGTGAAACCGTTCAGGACATCTACGGGATTGTTGCTAAGAAAGTCAACGAGATTCTACAAG
CAGACGCAATCAATGGGACCGATAACGAAGTAGTTACCGTGACCGATGAGAACACTGGTGAAATCTCTGA
GAAAGTCAAGCTGGGCACTAAGGCACTGGCTGGTCAATGGCTGGCTTACGGTGTTACTCGCAGTGTGACT
AAGCGTTCAGTCATGACGCTGGCTTACGGGTCCAAAGAGTTCGGCTTCCGTCAACAAGTGCTGGAAGATA
CCATTCAGCCAGCTATTGATTCCGGCAAGGGTCTGATGTTCACTCAGCCGAATCAGGCTGCTGGATACAT
GGCTAAGCTGATTTGGGAATCTGTGAGCGTGACGGTGGTAGCTGCGGTTGAAGCAATGAACTGGCTTAAG
TCTGCTGCTAAGCTGCTGGCTGCTGAGGTCAAAGATAAGAAGACTGGAGAGATTCTTCGCAAGCGTTGCG
CTGTGCATTGGGTAACTCCTGATGGTTTCCCTGTGTGGCAGGAATACAAGAAGCCTATTCAGACGCGCTT
GAACCTGATGTTCCTCGGTCAGTTCCGCTTACAGCCTACCATTAACACCAACAAAGATAGCGAGATTGAT
GCACACAAACAGGAGTCTGGTATCGCTCCTAACTTTGTACACAGCCAAGACGGTAGCCACCTTCGTAAGA
CTGTAGTGTGGGCACACGAGAAGTACGGAATCGAATCTTTTGCACTGATTCACGACTCCTTCGGTACCAT
TCCGGCTGACGCTGCGAACCTGTTCAAAGCAGTGCGCGAAACTATGGTTGACACATATGAGTCTTGTGAT
GTACTGGCTGATTTCTACGACCAGTTCGCTGACCAGTTGCACGAGTCTCAATTGGACAAAATGCCAGCAC
TTCCGGCTAAAGGTAACTTGAACCTCCGTGACATCTTAGAGTCGGACTTCGCGTTCGCG

References and Notes

1. P. S. Huang, G. Oberdorfer, C. Xu, X. Y. Pei, B. L. Nannenga, J. M. Rogers, F. DiMaio, T. Gonen, B. Luisi, D. Baker, High thermodynamic stability of parametrically designed helical bundles. *Science* **346**, 481–485 (2014). doi:10.1126/science.1257481 Medline

2. T. M. Jacobs, B. Williams, T. Williams, X. Xu, A. Eletsky, J. F. Federizon, T. Szyperski, B. Kuhlman, Design of structurally distinct proteins using strategies inspired by evolution. *Science* **352**, 687–690 (2016). doi:10.1126/science.aad8036 Medline

3. A. R. Thomson, C. W. Wood, A. J. Burton, G. J. Bartlett, R. B. Sessions, R. L. Brady, D. N. Woolfson, Computational design of water-soluble α-helical barrels. *Science* **346**, 485–488 (2014). doi:10.1126/science.1257452 Medline

4. R. B. Hill, D. P. Raleigh, A. Lombardi, W. F. DeGrado, De novo design of helical bundles as models for understanding protein folding and function. *Acc. Chem. Res.* **33**, 745–754 (2000). doi:10.1021/ar970004h Medline

5. P. B. Harbury, J. J. Plecs, B. Tidor, T. Alber, P. S. Kim, High-resolution protein design with backbone freedom. *Science* **282**, 1462–1467 (1998). doi:10.1126/science.282.5393.1462 Medline

6. G. J. Rocklin, T. M. Chidyausiku, I. Goreshnik, A. Ford, S. Houliston, A. Lemak, L. Carter, R. Ravichandran, V. K. Mulligan, A. Chevalier, C. H. Arrowsmith, D. Baker, Global analysis of protein folding using massively parallel design, synthesis, and testing. *Science* **357**, 168–175 (2017). doi:10.1126/science.aan0693 Medline

7. N. Koga, R. Tatsumi-Koga, G. Liu, R. Xiao, T. B. Acton, G. T. Montelione, D. Baker, Principles for designing ideal protein structures. *Nature* **491**, 222–227 (2012). doi:10.1038/nature11600 Medline

8. P. S. Huang, K. Feldmeier, F. Parmeggiani, D. A. F. Velasco, B. Höcker, D. Baker, De novo design of a four-fold symmetric TIM-barrel protein with atomic-level accuracy. *Nat. Chem. Biol.* **12**, 29–34 (2016). doi:10.1038/nchembio.1966 Medline

9. E. Marcos, T. M. Chidyausiku, A. C. McShan, T. Evangelidis, S. Nerli, L. Carter, L. G. Nivón, A. Davis, G. Oberdorfer, K. Tripsianes, N. G. Sgourakis, D. Baker, De novo design of a non-local β-sheet protein with high stability and accuracy. *Nat. Struct. Mol. Biol.* **25**, 1028–1034 (2018). doi:10.1038/s41594-018-0141-6 Medline

10. J. Dou, A. A. Vorobieva, W. Sheffler, L. A. Doyle, H. Park, M. J. Bick, B. Mao, G. W. Foight, M. Y. Lee, L. A. Gagnon, L. Carter, B. Sankaran, S. Ovchinnikov, E. Marcos, P.-S. Huang, J. C. Vaughan, B. L. Stoddard, D. Baker, De novo design of a fluorescence-activating β-barrel. *Nature* **561**, 485–491 (2018). doi:10.1038/s41586-018-0509-0 Medline

11. H. Lechner, N. Ferruz, B. Höcker, Strategies for designing non-natural enzymes and binders. *Curr. Opin. Chem. Biol.* **47**, 67–76 (2018). doi:10.1016/j.cbpa.2018.07.022 Medline

12. J. Dou, L. Doyle, P. Greisen Jr., A. Schena, H. Park, K. Johnsson, B. L. Stoddard, D. Baker, Sampling and energy evaluation challenges in ligand binding protein design. *Protein Sci.* **26**, 2426–2437 (2017). doi:10.1002/pro.3317 Medline

13. R. M. Gordley, L. J. Bugaj, W. A. Lim, Modular engineering of cellular signaling proteins and networks. *Curr. Opin. Struct. Biol.* **39**, 106–114 (2016). doi:10.1016/j.sbi.2016.06.012 Medline

14. M. J. Bick, P. J. Greisen, K. J. Morey, M. S. Antunes, D. La, B. Sankaran, L. Reymond, K. Johnsson, J. I. Medford, D. Baker, Computational design of environmental sensors for the potent opioid fentanyl. *eLife* **6**, e28909 (2017). doi:10.7554/eLife.28909 Medline

15. C. E. Tinberg, S. D. Khare, J. Dou, L. Doyle, J. W. Nelson, A. Schena, W. Jankowski, C. G. Kalodimos, K. Johnsson, B. L. Stoddard, D. Baker, Computational design of ligand-binding proteins with high affinity and selectivity. *Nature* **501**, 212–216 (2013). doi:10.1038/nature12443 Medline

16. N. F. Polizzi, Y. Wu, T. Lemmin, A. M. Maxwell, S.-Q. Zhang, J. Rawson, D. N. Beratan, M. J. Therien, W. F. DeGrado, De novo design of a hyperstable non-natural protein-ligand complex with sub-Å accuracy. *Nat. Chem.* **9**, 1157–1164 (2017). doi:10.1038/nchem.2846 Medline

17. J. Feng, B. W. Jester, C. E. Tinberg, D. J. Mandell, M. S. Antunes, R. Chari, K. J. Morey, X. Rios, J. I. Medford, G. M. Church, S. Fields, D. Baker, A general strategy to construct small molecule biosensors in eukaryotes. *eLife* **4**, e10606 (2015). doi:10.7554/eLife.10606 Medline

18. N. D. Taylor, A. S. Garruss, R. Moretti, S. Chan, M. A. Arbing, D. Cascio, J. K. Rogers, F. J. Isaacs, S. Kosuri, D. Baker, S. Fields, G. M. Church, S. Raman, Engineering an allosteric transcription factor to respond to new ligands. *Nat. Methods* **13**, 177–183 (2016). doi:10.1038/nmeth.3696 Medline

19. D. M. Spencer, T. J. Wandless, S. L. Schreiber, G. R. Crabtree, Controlling signal transduction with synthetic ligands. *Science* **262**, 1019–1024 (1993). doi:10.1126/science.7694365 Medline

20. V. J. Martin, D. J. Pitera, S. T. Withers, J. D. Newman, J. D. Keasling, Engineering a mevalonate pathway in Escherichia coli for production of terpenoids. *Nat. Biotechnol.* **21**, 796–802 (2003). doi:10.1038/nbt833 Medline

21. F. Zhang, J. Keasling, Biosensors and their applications in microbial metabolic engineering. *Trends Microbiol.* **19**, 323–329 (2011). doi:10.1016/j.tim.2011.05.003 Medline

22. A. Zanghellini, L. Jiang, A. M. Wollacott, G. Cheng, J. Meiler, E. A. Althoff, D. Röthlisberger, D. Baker, New algorithms and an in silico benchmark for computational enzyme design. *Protein Sci.* **15**, 2785–2794 (2006). doi:10.1110/ps.062353106 Medline

23. M. Babor, D. J. Mandell, T. Kortemme, Assessment of flexible backbone protein design methods for sequence library prediction in the therapeutic antibody Herceptin-HER2 interface. *Protein Sci.* **20**, 1082–1089 (2011). doi:10.1002/pro.632 Medline

24. D. J. Mandell, E. A. Coutsias, T. Kortemme, Sub-angstrom accuracy in protein loop reconstruction by robotics-inspired conformational sampling. *Nat. Methods* **6**, 551–552 (2009). doi:10.1038/nmeth0809-551 Medline

25. N. Ollikainen, R. M. de Jong, T. Kortemme, Coupling protein side-chain and backbone flexibility improves the re-design of protein-ligand specificity. *PLOS Comput. Biol.* **11**, e1004335 (2015). doi:10.1371/journal.pcbi.1004335 Medline

26. R. F. Alford, A. Leaver-Fay, J. R. Jeliazkov, M. J. O'Meara, F. P. DiMaio, H. Park, M. V. Shapovalov, P. D. Renfrew, V. K. Mulligan, K. Kappel, J. W. Labonte, M. S. Pacella, R. Bonneau, P. Bradley, R. L. Dunbrack Jr., R. Das, D. Baker, B. Kuhlman, T. Kortemme, J. J. Gray, The Rosetta all-atom energy function for macromolecular modeling and design. *J. Chem. Theory Comput.* **13**, 3031–3048 (2017). doi:10.1021/acs.jctc.7b00125 Medline

27. J. Choi, J. Chen, S. L. Schreiber, J. Clardy, Structure of the FKBP12-rapamycin complex interacting with the binding domain of human FRAP. *Science* **273**, 239–242 (1996). doi:10.1126/science.273.5272.239 Medline

28. M. D. Baker, M. B. Neiditch, Structural basis of response regulator inhibition by a bacterial anti-activator protein. *PLOS Biol.* **9**, e1001226 (2011). doi:10.1371/journal.pbio.1001226 Medline

29. H. K. Binz, P. Amstutz, A. Kohl, M. T. Stumpp, C. Briand, P. Forrer, M. G. Grütter, A. Plückthun, High-affinity binders selected from designed ankyrin repeat protein libraries. *Nat. Biotechnol.* **22**, 575–582 (2004). doi:10.1038/nbt962 Medline

30. J. N. Pelletier, F. X. Campbell-Valois, S. W. Michnick, Oligomerization domain-directed reassembly of active dihydrofolate reductase from rationally designed fragments. *Proc. Natl. Acad. Sci. U.S.A.* **95**, 12141–12146 (1998). doi:10.1073/pnas.95.21.12141 Medline

31. M. A. Kramer, S. K. Wetzel, A. Plückthun, P. R. Mittl, M. G. Grütter, Structural determinants for improved stability of designed ankyrin repeat proteins with a redesigned C-capping module. *J. Mol. Biol.* **404**, 381–391 (2010). doi:10.1016/j.jmb.2010.09.023 Medline

32. J. A. Davey, R. A. Chica, Multistate approaches in computational protein design. *Protein Sci.* **21**, 1241–1252 (2012). doi:10.1002/pro.2128 Medline

33. S. C. Alford, Y. Ding, T. Simmen, R. E. Campbell, Dimerization-dependent green and yellow fluorescent proteins. *ACS Synth. Biol.* **1**, 569–575 (2012). doi:10.1021/sb300050j Medline

34. A. S. Dixon, M. K. Schwinn, M. P. Hall, K. Zimmerman, P. Otto, T. H. Lubben, B. L. Butler, B. F. Binkowski, T. Machleidt, T. A. Kirkland, M. G. Wood, C. T. Eggers, L. P. Encell, K. V. Wood, NanoLuc complementation reporter optimized for accurate measurement of protein interactions in cells. *ACS Chem. Biol.* **11**, 400–408 (2016). doi:10.1021/acschembio.5b00753 Medline

35. R. Marshall, V. Noireaux, Synthetic biology with an all *E. coli* TXTL system: Quantitative characterization of regulatory elements and gene circuits. *Methods Mol. Biol.* **1772**, 61–93 (2018). doi:10.1007/978-1-4939-7795-6_4 Medline

36. M. Gao, J. Skolnick, The distribution of ligand-binding pockets around protein-protein interfaces suggests a general mechanism for pocket formation. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 3784–3789 (2012). doi:10.1073/pnas.1117768109 Medline

37. L. Song, C. D. Poulter, Yeast farnesyl-diphosphate synthase: Site-directed mutagenesis of residues in highly conserved prenyltransferase domains I and II. *Proc. Natl. Acad. Sci. U.S.A.* **91**, 3044–3048 (1994). doi:10.1073/pnas.91.8.3044 Medline

38. D. J. Mandell, "Backbone flexibility in computational design," thesis, University of California, San Diego (2010).

39. R. L. Dunbrack Jr., F. E. Cohen, Bayesian statistical analysis of protein side-chain rotamer preferences. *Protein Sci.* **6**, 1661–1681 (1997). doi:10.1002/pro.5560060807 Medline

40. I. W. Davis, D. Baker, RosettaLigand docking with full ligand and receptor flexibility. *J. Mol. Biol.* **385**, 381–392 (2009). doi:10.1016/j.jmb.2008.11.010 Medline

41. D. Röthlisberger, O. Khersonsky, A. M. Wollacott, L. Jiang, J. DeChancie, J. Betker, J. L. Gallaher, E. A. Althoff, A. Zanghellini, O. Dym, S. Albeck, K. N. Houk, D. S. Tawfik, D. Baker, Kemp elimination catalysts by computational enzyme design. *Nature* **453**, 190–195 (2008). doi:10.1038/nature06879 Medline

42. L. Jiang, E. A. Althoff, F. R. Clemente, L. Doyle, D. Röthlisberger, A. Zanghellini, J. L. Gallaher, J. L. Betker, F. Tanaka, C. F. Barbas 3rd, D. Hilvert, K. N. Houk, B. L. Stoddard, D. Baker, De novo computational design of retro-aldol enzymes. *Science* **319**, 1387–1391 (2008). doi:10.1126/science.1152692 Medline

43. J. A. Davey, R. A. Chica, Improving the accuracy of protein stability predictions with multistate design using a variety of backbone ensembles. *Proteins* **82**, 771–784 (2014). doi:10.1002/prot.24457 Medline

44. J. A. Davey, R. A. Chica, Multistate computational protein design with backbone ensembles. *Methods Mol. Biol.* **1529**, 161–179 (2017). doi:10.1007/978-1-4939-6637-0_7 Medline

45. E. L. Humphris, T. Kortemme, Prediction of protein-protein interface sequence diversity using flexible backbone computational protein design. *Structure* **16**, 1777–1788 (2008). doi:10.1016/j.str.2008.09.012 Medline

46. C. A. Smith, T. Kortemme, Backrub-like backbone simulation recapitulates natural protein conformational variability and improves mutant side-chain prediction. *J. Mol. Biol.* **380**, 742–756 (2008). doi:10.1016/j.jmb.2008.05.023 Medline

47. P. Conway, M. D. Tyka, F. DiMaio, D. E. Konerding, D. Baker, Relaxation of backbone bond geometry improves protein energy landscape modeling. *Protein Sci.* **23**, 47–55 (2014). doi:10.1002/pro.2389 Medline

48. D. M. Hoover, J. Lubkowski, DNAWorks: An automated method for designing oligonucleotides for PCR-based gene synthesis. *Nucleic Acids Res.* **30**, e43 (2002). doi:10.1093/nar/30.10.e43 Medline

49. M. T. Reetz, J. D. Carballeira, Iterative saturation mutagenesis (ISM) for rapid directed evolution of functional enzymes. *Nat. Protoc.* **2**, 891–903 (2007). doi:10.1038/nprot.2007.72 Medline

50. C. Engler, R. Kandzia, S. Marillonnet, A one pot, one step, precision cloning method with high throughput capability. *PLOS ONE* **3**, e3647 (2008). doi:10.1371/journal.pone.0003647 Medline

51. M. E. Lee, W. C. DeLoache, B. Cervantes, J. E. Dueber, A Highly Characterized Yeast Toolkit for Modular, Multipart Assembly. *ACS Synth. Biol.* **4**, 975–986 (2015). doi:10.1021/sb500366v Medline

52. Z. Z. Sun, C. A. Hayes, J. Shin, F. Caschera, R. M. Murray, V. Noireaux, Protocols for implementing an *Escherichia coli* based TX-TL cell-free expression system for synthetic biology. *J. Vis. Exp.* **79**, e50762 (2013). doi:10.3791/50762 Medline

53. G. Winter, xia2: An expert system for macromolecular crystallography data reduction. *J. Appl. Cryst.* **43**, 186–190 (2010). doi:10.1107/S0021889809045701

54. W. Kabsch, XDS. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 125–132 (2010). doi:10.1107/S0907444909047337 Medline

55. P. Evans, Scaling and assessment of data quality. *Acta Crystallogr. D Biol. Crystallogr.* **62**, 72–82 (2006). doi:10.1107/S0907444905036693 Medline

56. J. E. Padilla, T. O. Yeates, A statistic for local intensity differences: Robustness to anisotropy and pseudo-centering and utility for detecting twinning. *Acta Crystallogr. D Biol. Crystallogr.* **59**, 1124–1130 (2003). doi:10.1107/S0907444903007947 Medline

57. M. C. Thompson, Identifying and overcoming crystal pathologies: Disorder and twinning. *Methods Mol. Biol.* **1607**, 185–217 (2017). doi:10.1007/978-1-4939-7000-1_8 Medline

58. W. M. Kim, A. B. Sigalov, L. J. Stern, Pseudo-merohedral twinning and noncrystallographic symmetry in orthorhombic crystals of SIVmac239 Nef core domain bound to different-length TCRzeta fragments. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 163–175 (2010). doi:10.1107/S090744490904880X Medline

59. S. Jenni, N. Ban, Imperfect pseudo-merohedral twinning in crystals of fungal fatty acid synthase. *Acta Crystallogr. D Biol. Crystallogr.* **65**, 101–111 (2009). doi:10.1107/S0907444909000778 Medline

60. A. Sultana, I. Alexeev, I. Kursula, P. Mäntsälä, J. Niemi, G. Schneider, Structure determination by multiwavelength anomalous diffraction of aclacinomycin oxidoreductase: Indications of multidomain pseudomerohedral twinning. *Acta Crystallogr. D Biol. Crystallogr.* **63**, 149–159 (2007). doi:10.1107/S0907444906044271 Medline

61. M. Gilski, P. Drozdzal, R. Kierzek, M. Jaskolski, Atomic resolution structure of a chimeric DNA-RNA Z-type duplex in complex with Ba (2+) ions: A case of complicated multi-domain twinning. *Acta Crystallogr. D Struct. Biol.* **72**, 211–223 (2016). doi:10.1107/S2059798315024365 Medline

62. A. J. McCoy, R. W. Grosse-Kunstleve, P. D. Adams, M. D. Winn, L. C. Storoni, R. J. Read, Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007). doi:10.1107/S0021889807021206 Medline

63. P. D. Adams, P. V. Afonine, G. Bunkóczi, V. B. Chen, I. W. Davis, N. Echols, J. J. Headd, L.-W. Hung, G. J. Kapral, R. W. Grosse-Kunstleve, A. J. McCoy, N. W. Moriarty, R. Oeffner, R. J. Read, D. C. Richardson, J. S. Richardson, T. C. Terwilliger, P. H. Zwart, PHENIX: A comprehensive Python-based system for macromolecular structure solution.

Acta Crystallogr. D Biol. Crystallogr. **66**, 213–221 (2010). doi:10.1107/S0907444909052925 Medline

64. K. A. Kantardjieff, B. Rupp, Matthews coefficient probabilities: Improved estimates for unit cell contents of proteins, DNA, and protein-nucleic acid complex crystals. *Protein Sci.* **12**, 1865–1871 (2003). doi:10.1110/ps.0350503 Medline

65. C. X. Weichenberger, B. Rupp, Ten years of probabilistic estimates of biocrystal solvent content: New insights via nonparametric kernel density estimate. *Acta Crystallogr. D Biol. Crystallogr.* **70**, 1579–1588 (2014). doi:10.1107/S1399004714005550 Medline

66. P. V. Afonine, R. W. Grosse-Kunstleve, N. Echols, J. J. Headd, N. W. Moriarty, M. Mustyakimov, T. C. Terwilliger, A. Urzhumtsev, P. H. Zwart, P. D. Adams, Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr. D Biol. Crystallogr.* **68**, 352–367 (2012). doi:10.1107/S0907444912001308 Medline

67. P. Emsley, B. Lohkamp, W. G. Scott, K. Cowtan, Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501 (2010). doi:10.1107/S0907444910007493 Medline

68. N. W. Moriarty, R. W. Grosse-Kunstleve, P. D. Adams, Electronic Ligand Builder and Optimization Workbench (eLBOW): A tool for ligand coordinate and restraint generation. *Acta Crystallogr. D Biol. Crystallogr.* **65**, 1074–1080 (2009). doi:10.1107/S0907444909029436 Medline

69. F. C. Bernstein, T. F. Koetzle, G. J. B. Williams, E. F. Meyer Jr., M. D. Brice, J. R. Rodgers, O. Kennard, T. Shimanouchi, M. Tasumi, The Protein Data Bank: A computer-based archival file for macromolecular structures. *J. Mol. Biol.* **112**, 535–542 (1977). doi:10.1016/S0022-2836 (77)80200-3 Medline

70. P. A. Karplus, K. Diederichs, Linking crystallographic model and data quality. *Science* **336**, 1030–1033 (2012). doi:10.1126/science.1218231 Medline

71. A. T. Brünger, Free R value: Cross-validation in crystallography. *Methods Enzymol.* **277**, 366–396 (1997). doi:10.1016/S0076-6879 (97)77021-6 Medline

72. V. B. Chen, W. B. Arendall 3rd, J. J. Headd, D. A. Keedy, R. M. Immormino, G. J. Kapral, L. W. Murray, J. S. Richardson, D. C. Richardson, MolProbity: All-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 12–21 (2010). doi:10.1107/S0907444909042073 Medline

73. T. Inoue, W. D. Heo, J. S. Grimley, T. J. Wandless, T. Meyer, An inducible translocation strategy to rapidly activate and inhibit small GTPase signaling pathways. *Nat. Methods* **2**, 415–418 (2005). doi:10.1038/nmeth763 Medline