nature biotechnology

Supplementary information

https://doi.org/10.1038/s41587-026-03081-9

Artificial allosteric protein switches with machine-learning-designed receptors

In the format provided by the authors and unedited

Supplementary Information

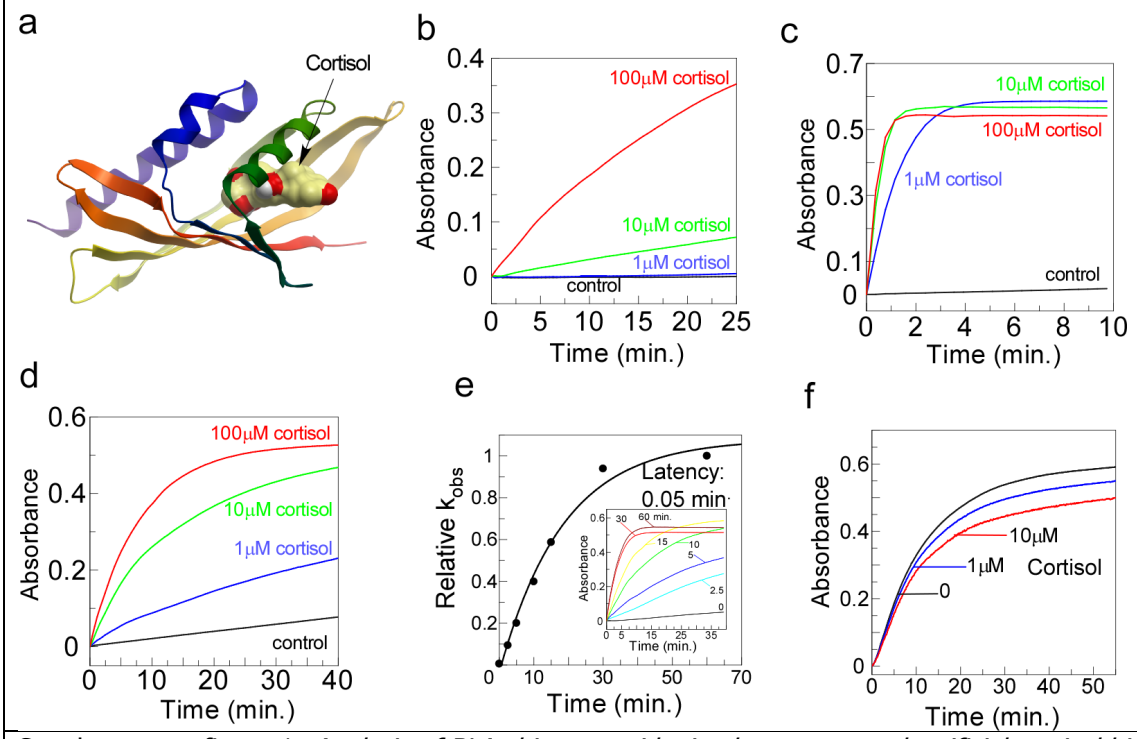
Artificial allosteric protein switches with machine learning-designed receptors

Table of contents:

Supplementary Figure 1
Supplementary Figure 2
Supplementary Figure 3
Supplementary Figure 4
Supplementary Figure 5
Supplementary Figure 6
Supplementary Figure 7
Supplementary Figure 8
Supplementary Figure 9
Supplementary Figure 10
Supplementary Figure 11
Supplementary Figure 12
Supplementary Figure 13
Supplementary Figure 14

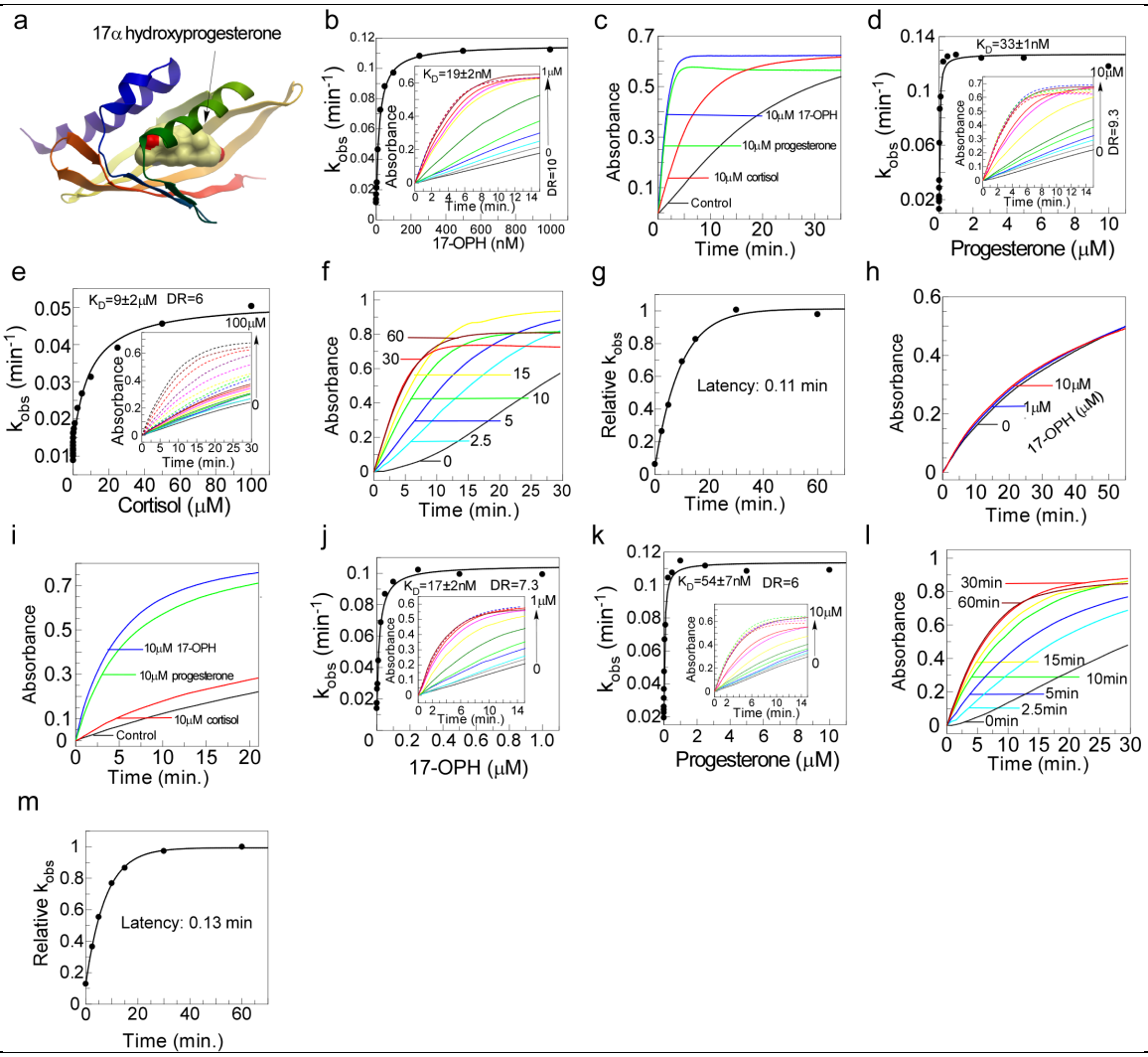
Supplementary text 1
Supplementary text 2
Supplementary text 3

Supplementary Table 1
Supplementary Table 2
Supplementary Table 3
Supplementary Table 4
Supplementary Table 5
Supplementary Table 6



Supplementary figure 1. *Analysis of BLA chimeras with circular permutated artificial cortisol binding domains.*

(a) Structure of the NTF2 cortisol binding domain HCY129.1 in complex with its ligand[2]. The protein is shown in ribbon representation and colored spectrally from the N- to C-terminus while cortisol is displayed as molecular surface colored by atom type.
(b) Time-resolved changes in absorption of a solution containing 100 nM cpHCY129.1-24-BLA-253 chimera, 50 μM β-lactamase chromogenic substrate and the indicated concentrations of cortisol.
(c) Same as (b) but using cpHCY129.1-35-BLA-253 chimera.
(d) Same as (b) but using cpHCY129.1-64-BLA-253 chimera.
(e) Latency analysis of 50 nM cortisol biosensor cpHCY129.1-35-BLA-253 incubated with 10 μM of cortisol for the indicated period of time and assayed for β-lactamase activity (inset). Fitting of the initial rates to a single exponential equation provided a numerical expression of latency.
(f) Activity analysis of 100 nM chimeric protein wtHCY129.1-BLA-253 and its substrate in the presence or absence of the indicated concentrations of cortisol.



Supplementary figure 2. Analysis of BLA chimeras with circular permutated artificial 17 α-hydroxyprogesterone (17-OPH) binding domain.

(a) Structure of the NTF2 cortisol binding domain OHPFA_1952 in complex with 17-OPH[2]. (b) Enzymatic activity of 50 nM cpOHPFA1952-31-BLA-253 chimera at different concentrations of 17-OPH. The inset shows the absorbance changes of the solution containing 50 nM biosensor, 50 μM chromogenic BLA substrate UW154 and increasing concentrations of 17-OPH. The K_D value and dynamic range (DR) are denoted on the graph. (c) Catalytic activity of 100 nM cpOHPFA1952-31-BLA-253 chimera in the presence of 10 μM 17-OPH, progesterone or cortisol.

(d) Catalytic activity of 50 nM cpOHPFA1952-31-BLA-253 chimera in the presence of increasing concentrations of progesterone. The inset shows the time-resolved changes in absorption while the main plot shows the initial velocities of the reactions plotted versus the concentration of the titrant and fitted to a quadratic equation. The resulting K_D value and the dynamic range (DR) of the chimera are shown.

(e) Same as (d) but using cortisol as a titrant.

(f) Latency analysis of 50 nM 17-OPH biosensor cpOHPFA1952-31-BLA-253 incubated with 1 μM 17-OPH for indicated periods and monitored for β-lactamase activity.

(g) Fit of the initial rates from the graphs shown in (f) to a single exponential equation to obtain latency time constant.

(h) Activity analysis of 100 nM chimeric protein wtOHPFA1952-BLA-253 in the presence or absence of the indicated concentrations of 17-OPH.

(i) Catalytic activity of 100 nM cpOHPFA1952-20-BLA-253 chimera in the presence of 10 μM 17-OPH, progesterone and cortisol.

(j) Catalytic activity of 50 nM solution of cpOHPFA1952-20-BLA-253 chimera in the presence of increasing concentrations of 17-OPH. The inset shows the time resolved changes in absorption while

the main plot shows the initial velocities of the reactions plotted versus the concentration of the titrant and fitted to a quadratic equation.
(k) Same as (j) but using progesterone as the titrant.
(l) Latency analysis of 50 nM cortisol biosensor cpOHPFA1952-20-BLA-253 incubated with 1 μM 17-OPH for the periods indicated and monitored for β-lactamase activity. (m) Fit of the initial rates from the graphs shown in (l) to a single exponential equation to obtain a numerical expression of latency time constant.

Supplementary text 1.

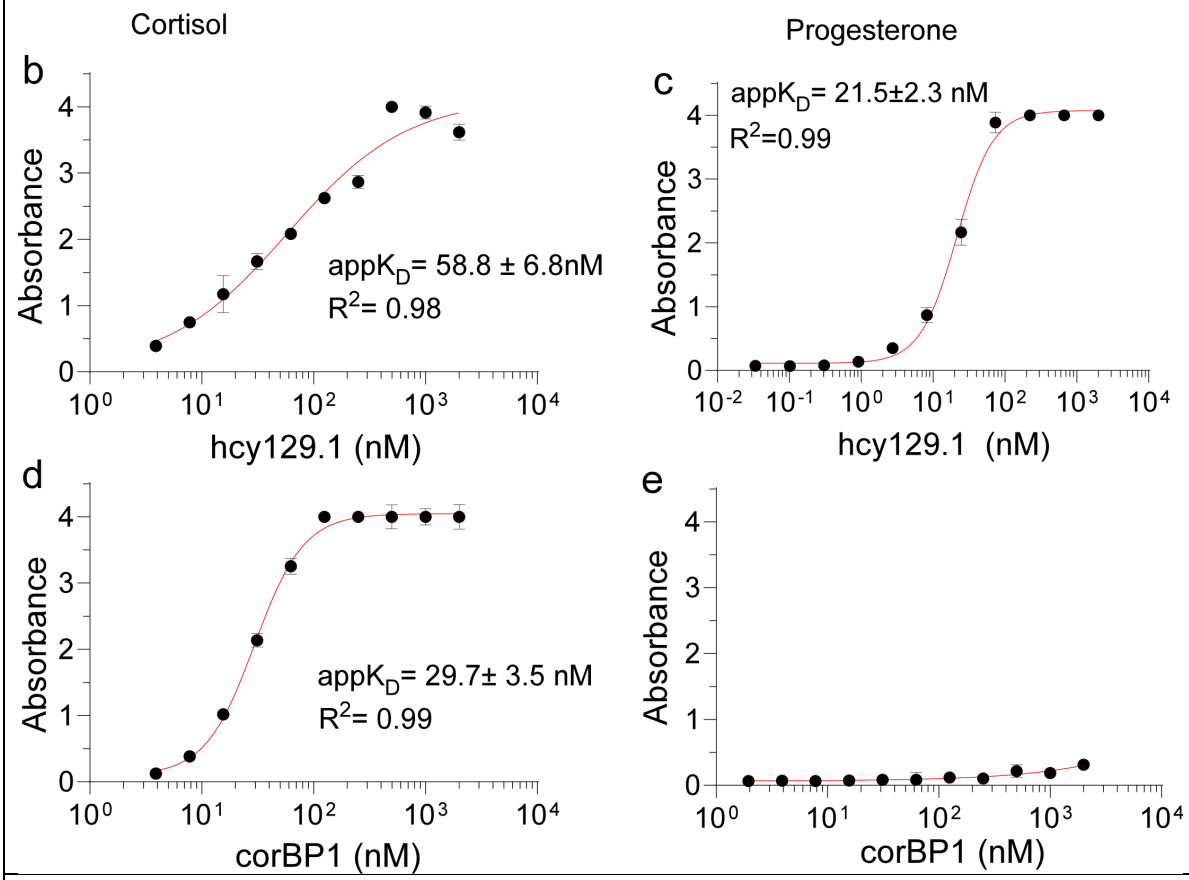
Improving selectivity of hcy121.1 cortisol binding domain

In order to obtain variants of hcy121.1 domain we designed a gene library by fully randomizing 12 amino acids forming the cortisol binding site (for sequence see Supplementary table 1). The corresponding gene library was commercially synthesized using trimer codons by Synbio Technologies (New Jersey). The resulting liner library was used as an input in mRNA display performed as described earlier [4].

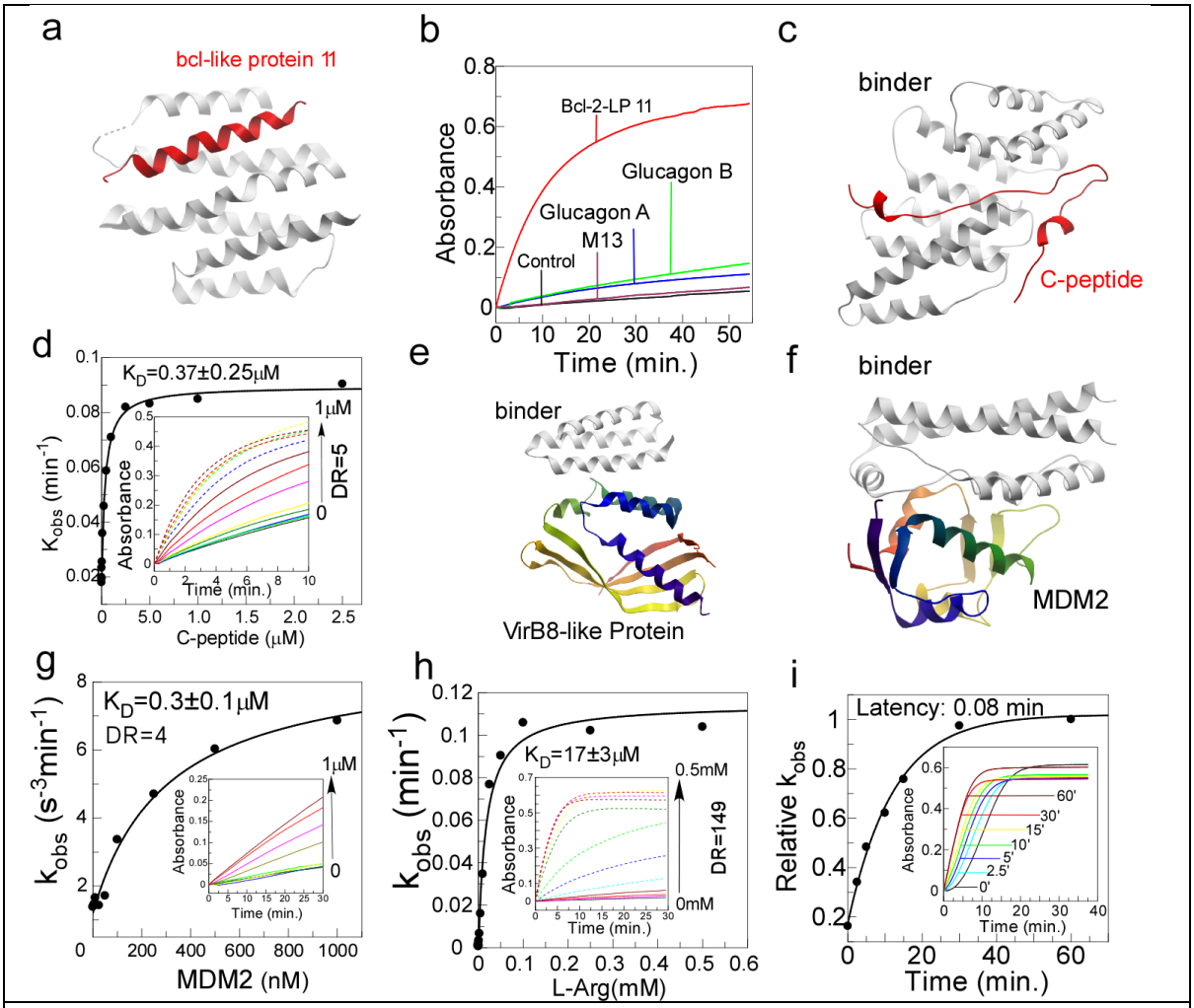
Following in vitro translation the mRNA-protein conjugate library was first subjected to pre-clearing on biotin-saturated magnetic streptavidin Dynabeads to remove nonspecific streptavidin and biotin binders. The unbound library fraction was then incubated with the same beads saturated with biotin-PEG-cortisol. After washing three times, the cDNA remained on the beads was amplified, transcribed and ligated to puromycin oligonucleotide for the next round of selection. Five iterative rounds were performed, with selection pressure increasing in each round by increasing the number of wash cycles, introducing high-salt washes, extending pre-clearing incubation, and shortening incubation time with target-loaded beads. Enrichment across rounds was assessed by polyclonal ELISA using *in vitro* translation products from recovered DNA pools. The DNA pool from the final round was cloned into pET XX expression vector and the individual clones were expressed in *E.coli* . The selectivity of the resulting binding domains was analysed by surface immobilized bioin-PEG-cortisol ELISA.

a

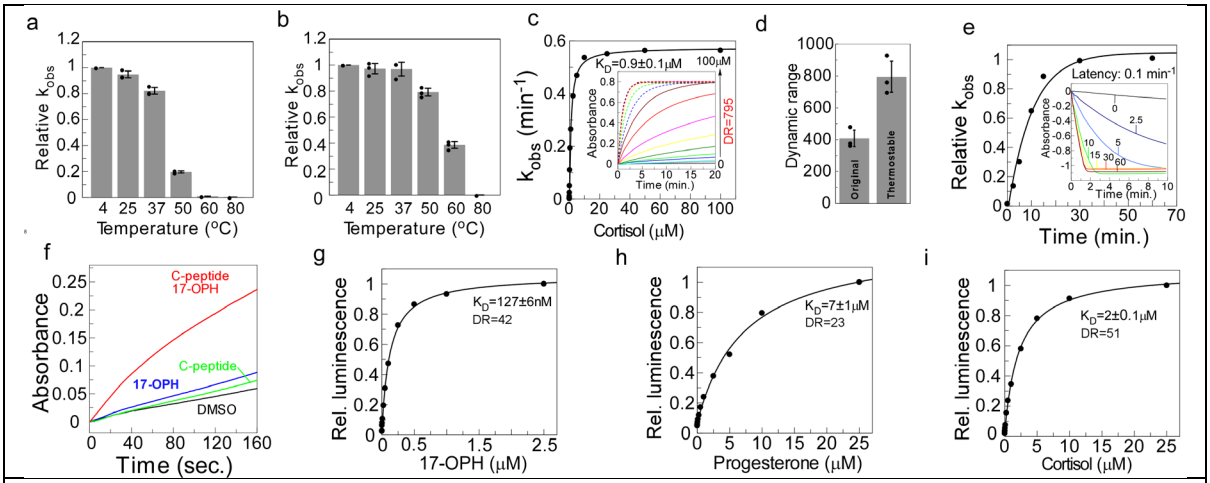
```
hcy129.1   1     TSSKEAEEAIRDMLRRWYEAINKGDMEKLKSLVDPDASFHFARTNQQYDKEQFLEMIKEA  60
                 TSSKEAEEAIRDMLRRWYEAINKGDMEKLKSLVDPDAS+HFARTNQQYDKEQ LE++KEA
corBP1     1     TSSKEAEEAIRDMLRRWYEAINKGDMEKLKSLVDPDASWHFARTNQQYDKEQHLELLKEA  60

hcy129.1   61    LKQDLKVEVKSIHIQQQPRGDHVTVTVHVEAHMNRNGQTHTFTVTDHYHFVRKGDSWKIT  120
                 LKQDLKVEVKSIHIQQQPRGDHVTVTVHVE HMNRNGQTHTFT+TDHYHFVRKGDSWKIT
corBP1     61    LKQDLKVEVKSIHIQQQPRGDHVTVTVHVEGHMNRNGQTHTFTLTDHYHFVRKGDSWKIT  120

hcy129.1   121   RTQWHIHQQ  129
                 R Q H HQQ
corBP1     121   RCQVHDHQQ  129
```



Supplementary figure 3. Ligand-interaction analysis of the hcy121.1 cortisol- binding domain and its variant corBP1 selected by mRNA display. (a) Sequence alignment of the parental cortisol binding domain hcy129.1 and its derivative corBP1 obtained by mRNA selection from the library of binding site mutants. (b) A plot of absorbance values of ELISA assay where surface immobilized biotin-PEG-cortisol was probed with the increasing concentrations of recombinant hcy129.1. (c) as in (b) but using biotin-PEG-progesterone as a ligand. (d) as in (b) but using corBP1 domain obtained by selection of hcy121.1-derived library on biotin-PEG-cortisol-loaded matrix. (e) as in (c) but using corBP1 domain as a probe. Data represent mean ± s.d. from duplicate measurements.



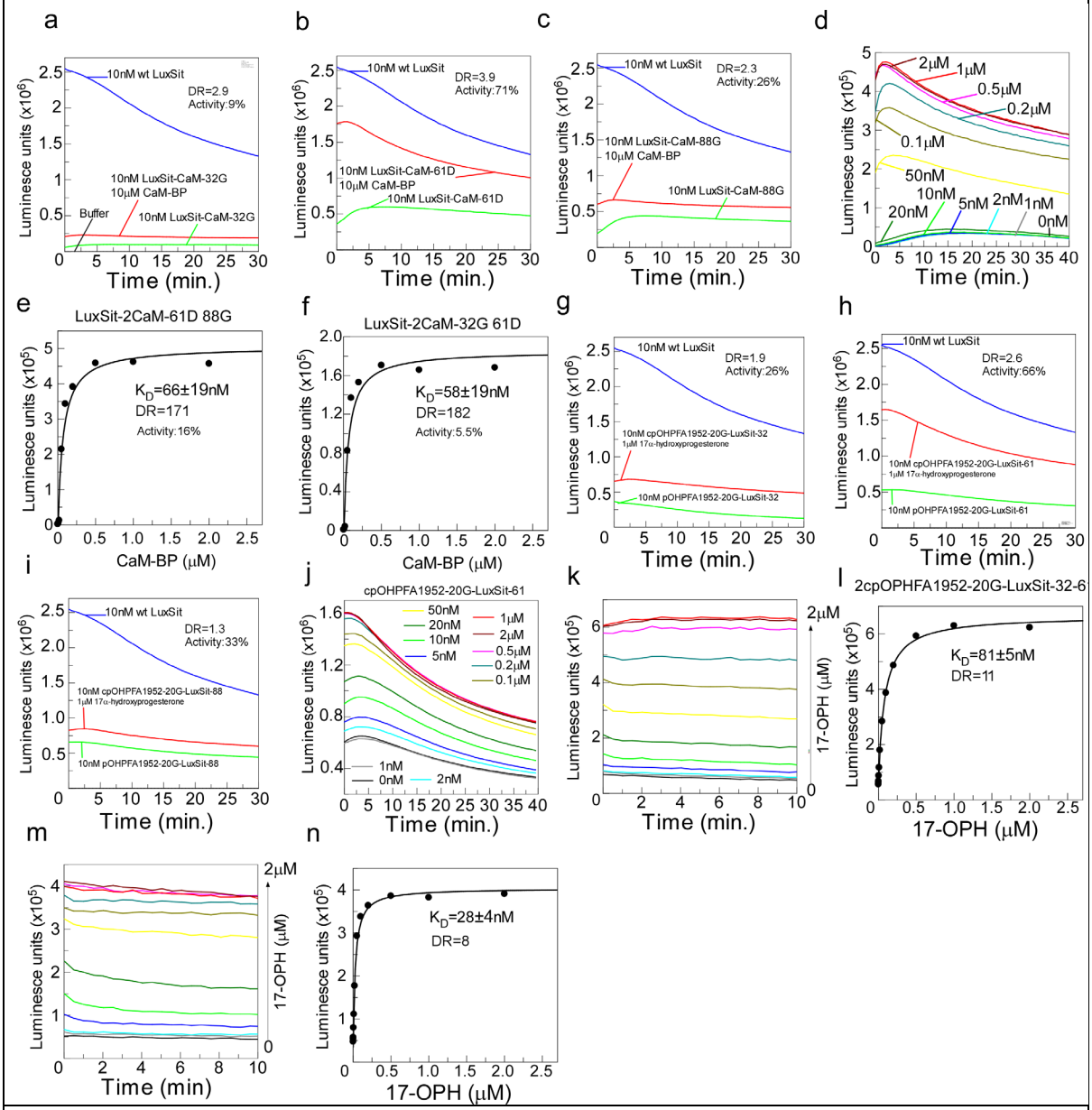
Supplementary Figure 4. Analysis of BLA biosensors with artificial peptide and protein receptors.
(a) Model of the bcl-like protein 11 peptide binding protein (PDB: 8T5E) displayed as grey ribbon with its ligand peptide (red).
(b) Time-resolved absorption traces of 100 nM cp8T5E-118-BLA-253 chimera supplemented with 10 μM cognate ligand peptide Bcl-2-LP11 or the same concentration of the control peptides.
(c) Ribbon representation of artificial binding domain CPH02 (grey) with a bound C-peptide ligand (red).
(d) Analysis of the titration experiment where 100 nM cpCPH02-52-BLA-253 chimera was titrated with increasing concentrations of C-peptide. The outer plot displays k_{obs} values plotted against the peptide concentrations while the inset displays the activity traces. The K_D and dynamic range are indicated.
(e) Structure of the VirB8-like protein (VirB8-LP) shown as a ribbon spectrally colored from N- to C-terminus in complex with an artificial binding domain shown as grey ribbon (PDB: 7SH3).
(f) Ribbon representation of a MDM2 fragment spectrally colored N- to C-terminus bound to an artificial binding domain which is colored grey. The AlphaFold2 model of the complex, the corresponding sequences and the experimental binding data were extracted from [3]. (g) Titration data of 100 nM cp-MDM2-27-BLA-253 with increasing concentrations of recombinant MDM2. The kinetic data were processed and fitted as in (d).
(h) Titration of 100 nM 2cpArgSBP-206-BLA-41-197 chimera containing two cpArgSBP receptor domains with increasing concentrations of L-Arg (inset). The kinetic data were processed as in (g).
(i) Fit of the k_{obs} values of 100 nM 2cpArgSBP-206-BLA-41-197 in the presence of 0.25 mM L-Arg where the samples were collected at the indicated point of time and analyzed for activity. The inset shows a fit to a single exponential to obtain the latency time constant.



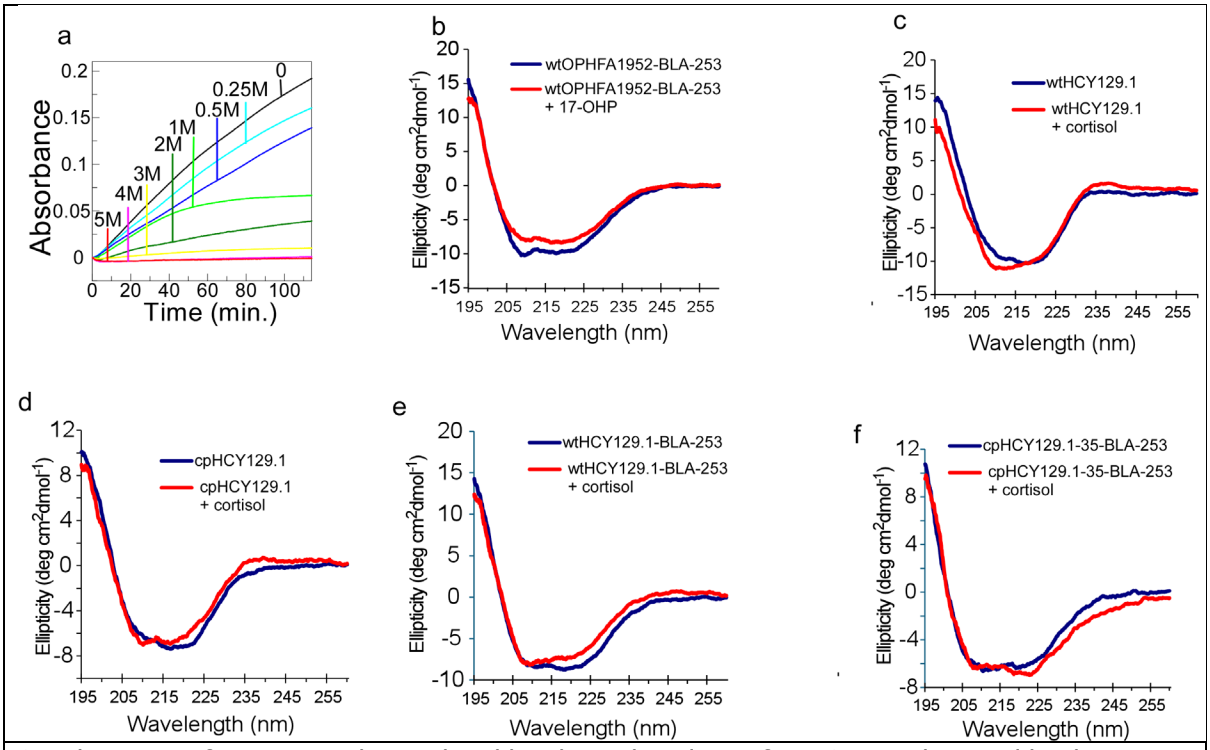
Supplementary figure 5. Engineering of chimera stability and reporter chemistry.
(a) Thermostability of cpHCY129.1-35-BLA-253 chimera analyzed by incubating 100 nM samples at the indicated temperatures and assessing their activity by supplementing the reactions with 10 μM of cortisol together with the chromogenic substrate and measuring the k_{obs} rate constant.
(b) Same as (a) but using thermostable cpHCY129.1-35-BLA-253 chimera (Supplementary table 1).
(c) Activity analysis of the thermostabilized cpHCY129.1-35-BLA-253 chimera upon titration with increasing concentrations of cortisol. The inset shows absorbance changes of solutions containing 50 nM biosensor and 50 μM chromogenic BLA substrate. The K_D value was extracted by plotting the initial reaction velocities against cortisol concentration and fitting them to a quadratic equation. DR represents the dynamic range of the biosensor.
(d) Comparison of dynamic ranges of cpHCY129.1-35-BLA-253 chimera and its thermostabilized variant obtained by analyzing time traces of reactions containing 100 nM solutions of the respective chimera and 10 μM cortisol.
(e) Latency analysis of 10 nM cpOHPFA1952-20-GDH-404 in the presence of 10 μM 17–OPH. After mixing the assay components, samples were collected at the indicated time points and activity was measured. The inset shows the time traces of samples analyzed after the various incubation times.
(f) Activity of 100 nM cpOHPFA1952-20-cpCP-H02-52-BLA-41-197 in the presence or absence of 2 μM 17-OPH and 10 μM C-peptide.
(g) Titration of 20 nM cpHCY129.1-35-NanoLuc-157 with increasing concentrations of 17–OPH. K_D values were calculated by fitting the data to the quadratic equation.
(h) Same as (g) but using progesterone as the titrant.
(i) Same as (g) but using cortisol as the titrant.

Supplementary text 2

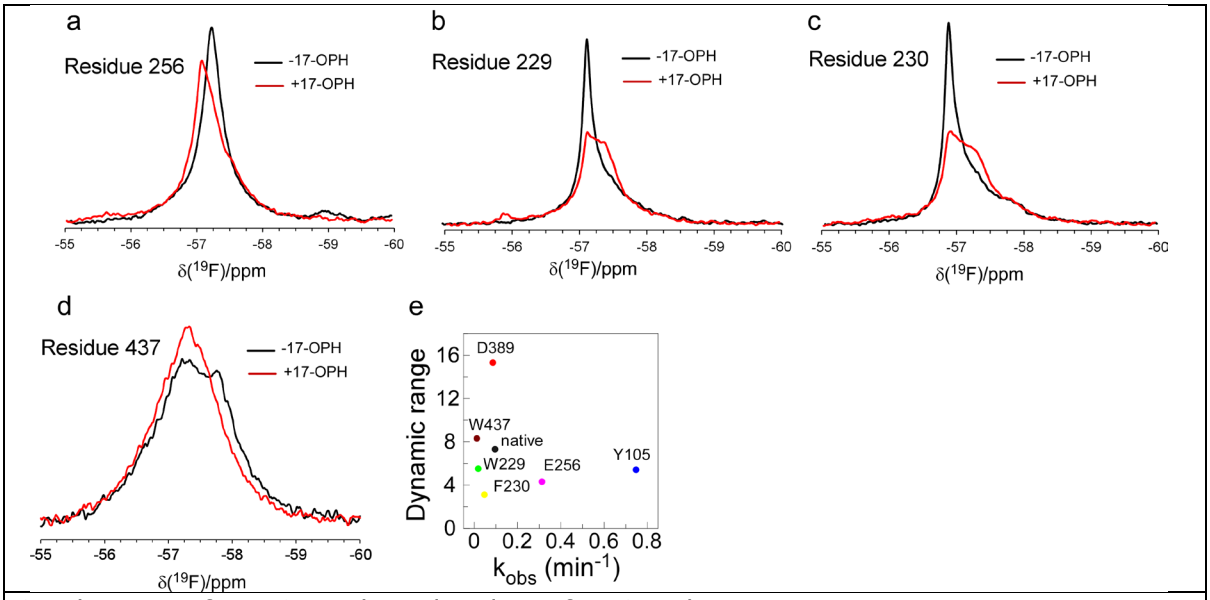
To further explore the fitness landscape of the developed switches we examined the thermostability of the developed cortisol biosensor and found that it was low (Supplementary figure 7a). We hypothesized that the introduction of known thermostabilizing mutations (K55Q, S82A, G92D, T140K, H153R, V184A) into the β-lactamase reporter could improve the overall stability of the chimera[22]. The resulting protein demonstrated improved thermostability, 2-fold increase in the dynamic range and 8-fold increase in k_{cat}. These results indicate that the performance parameters of chimeras with synthetic receptors are amenable to optimization by varying the amino acid sequence of the individual domains (Supplementary figure 5a-d, Supplementary table 1).



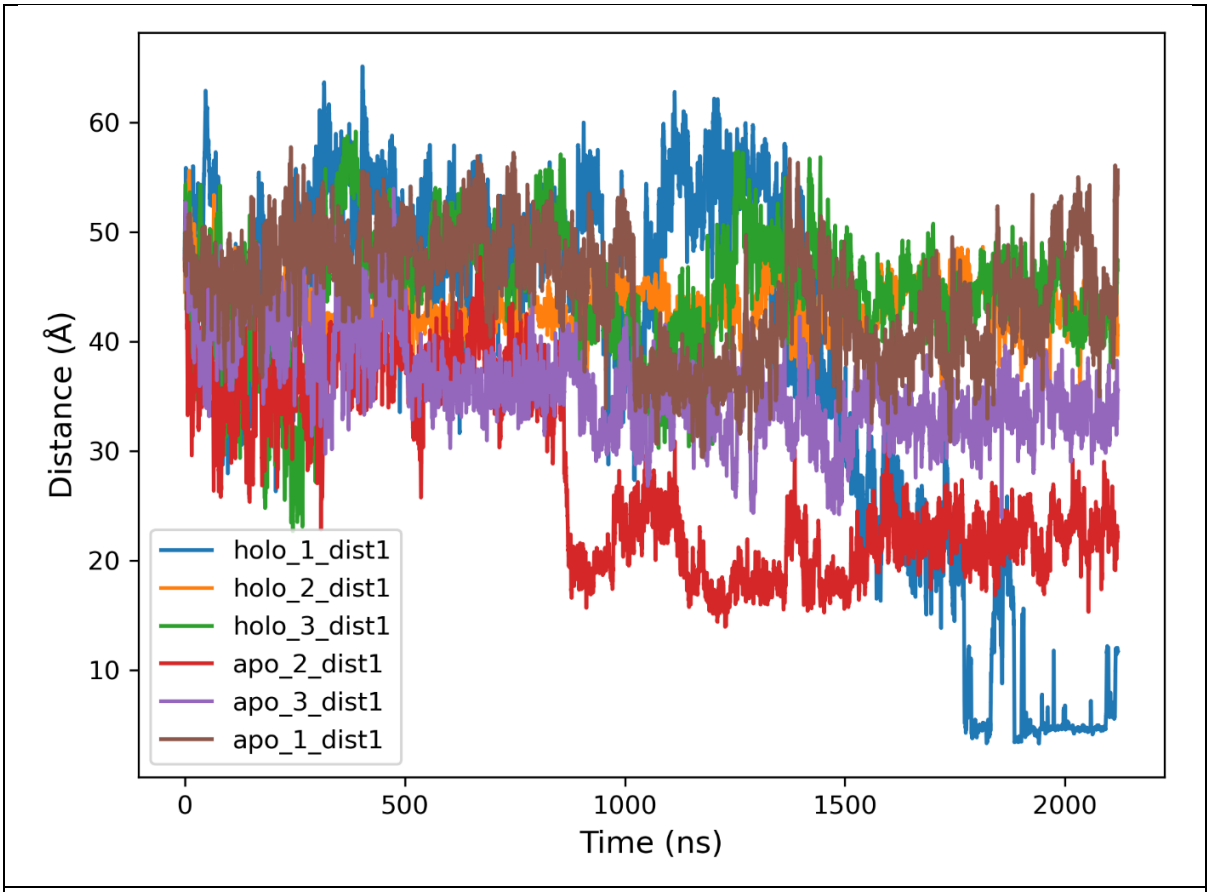
Supplementary figure 6. Activity analysis of LuxSit Pro chimeras.
(a) Time-resolved luminescence of 10 nM wild-type LuxSit Pro or CaM-LuxSit Pro-32 chimera in the absence or presence of 10 μM calmodulin binding peptide (Cam-BP).
(b) Same as (a) but using CaM-LuxSit Pro -61 chimera.
(c) Same as (a) but using CaM-LuxSit Pro-88 chimera.
(d) Titration of 10 nM 2CaM-LuxSit Pro-61-88 chimera with increasing concentration of Cam-BP.
(e) Fit of the data from (d). The dynamic range (DR), K_D and residual catalytic activity are indicated.
(f) Same as (e) but fitting the data from the titration of 2CaM-LuxSit Pro-32-61 chimera. (g) Plot of time-resolved luminescence of cpOHPFA1952-20-LuxSit Pro-32 chimera in the presence or absence of saturating concentrations of 17-OPH. 10 nM LuxSit was used as reference.
(h) Same as (g) but for cpOHPFA1952-20-LuxSit Pro-61 chimera.
(i) Same as (g) but using cpOHPFA1952-20-LuxSit Pro-88 chimera.
(j) Time-resolved luminescence of 10 nM cpOHPFA1952-20-LuxSit Pro-61 chimera in the absence or presence of increasing concentrations of 17-OPH.
(k) Same as (j) but using 25 nM 2cpOPHFA1952-20G-LuxSit Pro-32-61.
(i) Plot of luminescence values extracted from (k) and fitted to a quadratic equation.
(m) Time-resolved luminescence of 25 nM 2cpOPHFA1952-20G-LuxSit Pro-61-88.
(n) Plot of luminescence values extracted from (m) fitted to a quadratic equation.



Supplementary figure 7. Biochemical and biophysical analysis of 17-OPH and cortisol binding domains and thereon based switchable chimeras.
(a) Activity analysis of cp35-HCY129.1-253-BLA chimera in the presence of 20 μM of cortisol and indicated concentrations of urea.
(b) Far-UV CD spectra of 6.7 μM wild-type OPHFA1952-253-BLA chimera in the absence (blue line) and presence (red line) of 20 μM 17-OPH.
(c) Same as (b) but for 6.4 μM cortisol binding domain HCY129.1 in the absence (blue line) and presence (red line) of 20 μM cortisol.
(d) Same as (c) but using 6.4 μM circularly permutated cortisol binding domain cp35-HCY129.1.
(e) Same as (c) but using 6.7 μM chimera of the wild-type cortisol binder HCY129.1 and BLA.
(f) Same as (c) but using 6.7 μM cortisol biosensor cp35HCY129.1-253-BLA.



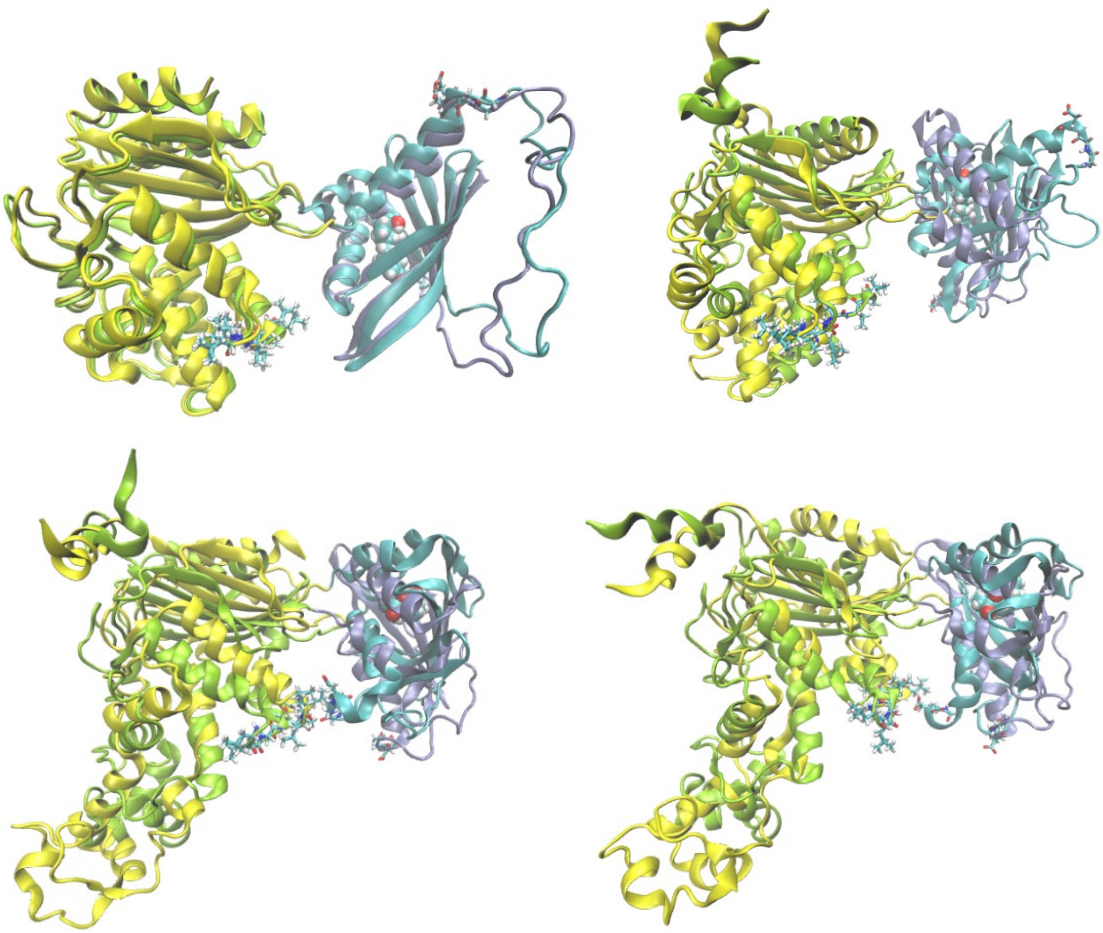
Supplementary figure 8. Biophysical analysis of 17-OPH biosensor
(a) ^{19}F-NMR spectra of 0.6 mM cpOHPFA_1952-20-BLA-253 labeled with *p*-(O)CF_3-tyrosine at position 256 in the absence (black) and presence (red) of 1.5 mM 17-OPH.
(b) Same as (a) but using cpOHPFA_1952-20-BLA-253 labeled at position 229.
(c) Same as (a) but with the label installed at position 230.
(d) Same as (a) but with the label in position 430.
(e) Activity analysis of *p*-(O)CF_3-tyrosine labeled probes. In the experiment 50 nM cpOHPFA_1952-20-BLA-253 labeled at the indicated position was assayed for activity in the presence and absence of 1 μM 17-OPH and the dynamic range was plotted versus the maximal k_{obs} values. Unlabeled chimera (marked as 'native') was used a reference.



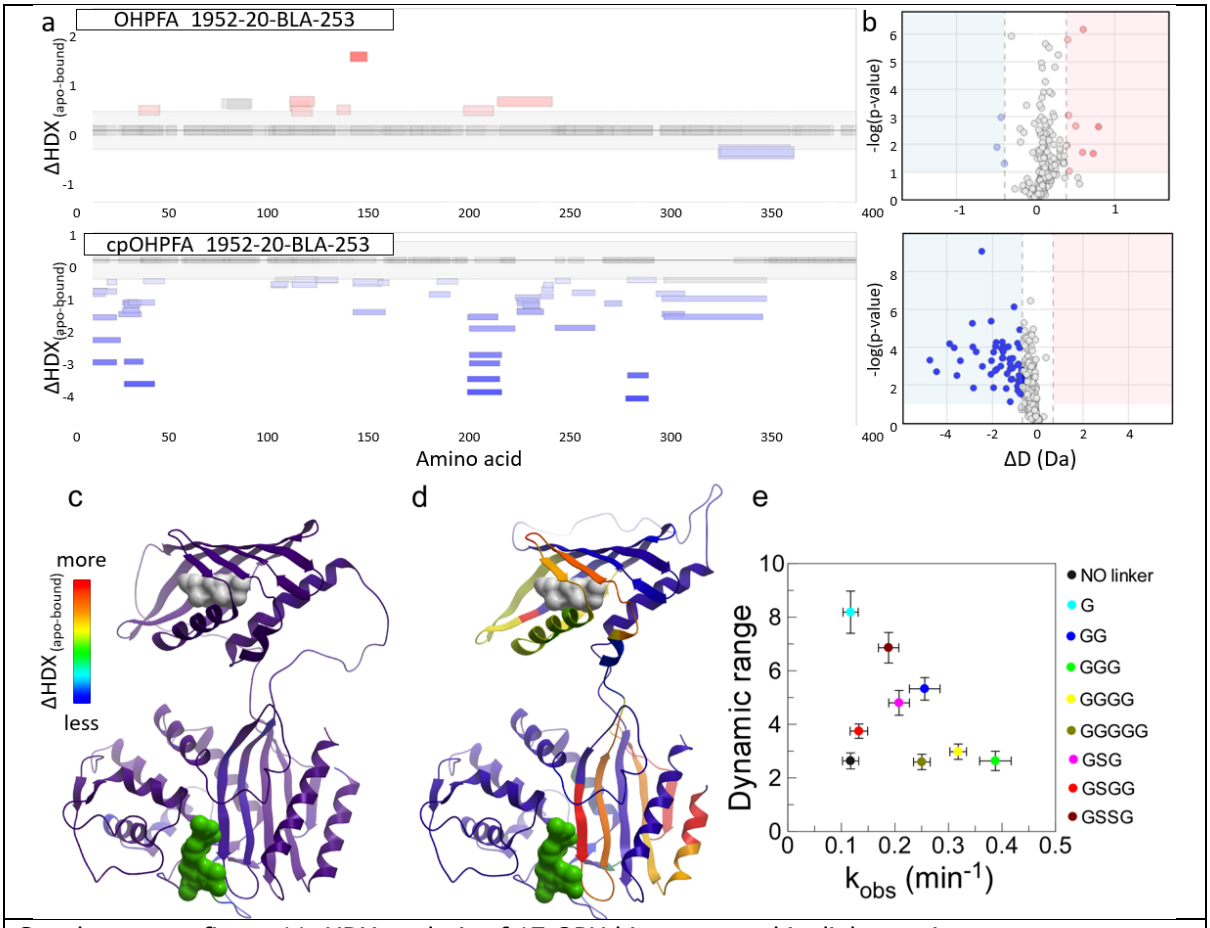
Supplementary figure 9. Distance between the speculative catalytic and allosteric domain contacts for the three holo and three apo replicates over the course of the production MD trajectory. Of a particular note is the stable interaction that formed between the catalytic and allosteric binding domains in the holo_1 replicate at ~1.77 µs.

Supplementary text 3: Molecular dynamics simulations

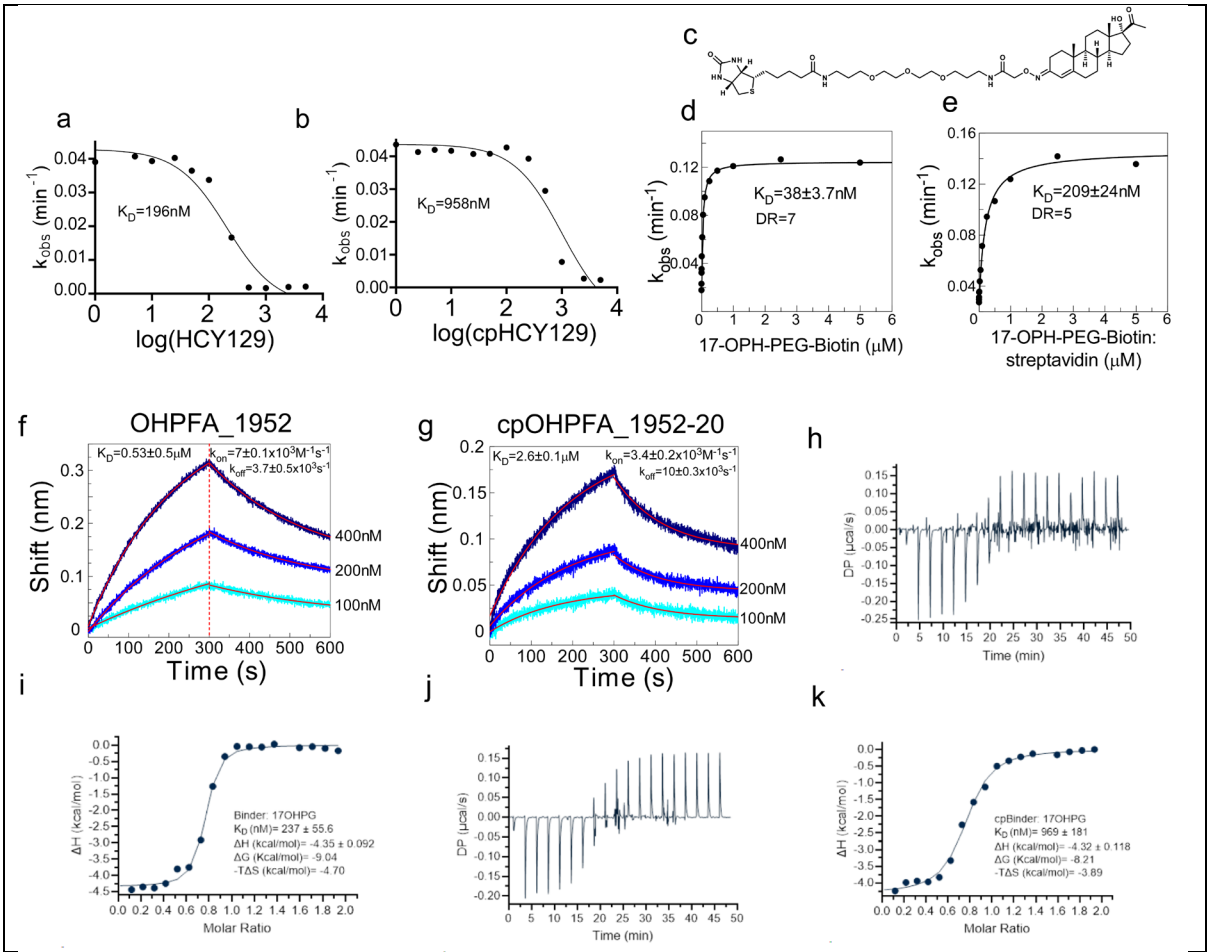
As can be seen in Supplementary figure 11 there was no discernible consistency or correlation within the replicate trajectories that allowed us to distinguish between the apo and holo chimeric proteins. While there appeared to be a potential stable contact between the allosteric binding domain and the catalytic domain in one of the holo trajectories, it was not observed in replicate trajectories, although the closest in behavior was a trajectory in the apo replicates. Interestingly, this interaction was quite stable once established (Supplementary Figure 11). However, the extreme variability between replicates suggests strongly that a vastly more extensive MD campaign would be required for us to be able to formulate any hypotheses worth testing in the laboratory.



Supplementary figure 10. Trajectory snapshots of holo_1 and apo_2 (see Figure 1) at 0 ns production (after 100 ns of equilibration, top left), at 600 ns (top right), 1.77 ns where first firm contact between the catalytic and allosteric domains was observed in the holo trajectory. The contact residues are shown in stick representation, 17-hydroxyprogesterone is shown in VDW representation, the catalytic domains are shown in yellow and green ribbons for the holo and apo respectively, and the allosteric domains are shown in blue and purple ribbons for the holo and apo, respectively.



Supplementary figure 11. HDX analysis of 17-OPH biosensor and its linker variants.
(a) Difference in HDX labeling for the wtOHPFA_1952-20-BLA-253 (top) and the circular permutant cpOHPFA_1952-20-BLA-253 (bottom) biosensor proteins when incubated with 17-hydroxyprogesterone (17-OPH). Woods plot bars indicate uniquely identified peptide slices. Bars positioned on y-axis and colored according to the size of the HDX difference upon ligand binding (red – deprotection; grey – no significant difference; blue – protection). Shaded region indicates limit of statistical significance ($p<0.001$) confidence limit from global significance testing. Note y-axes are not on same scale
(b) Volcano plot of HDX data filtered by global significance and Welch's t-test ($p<0.01$). Each data point represents a unique peptide slice from the OHPFA_1952-20-BLA-253 (top) and the cpOHPFA_1952-20-BLA-253 (bottom) biosensor proteins.
(c-d) Ribbon representation of the OHPFA_1952-20-BLA-253 (left) and cpOHPFA_1952-20-BLA-253 structure colored according to the difference in D-labeling with saturating 17-OPH. 17-OPH is displayed as a grey molecular surface and the active site of β-lactamase is marked by a green inhibitor molecule imported from PDB:6c79.
(e) A plot of dynamic ranges and maximal catalytic activities of 50 nM solution of cpOHPFA_1952-20-BLA-253 chimera with different linker sequences.



Supplementary figure 12. Biophysical analysis of native and circular permutated ML-generated steroid binding domains.

(a) Plot of k_{obs} values of reactions containing 50 nM thermostable cpHCY129.1-35-BLA-253 mixed with 500 nM cortisol and 50 μM of chromogenic BLA substrate UW154 and titrated with HCY129.1. The titration data were fitted to a competitive equation to determine the affinity of both cortisol-binding entities.

(b) Same as (a) but titrating cpHCY129.1. The resulting K_D values are indicated on the plot.

(c) Structure of 17-OPH modified with PEG linker and a biotin group that was used as a bait in bilayer interferometry analysis.

(d) A plot of k_{obs} values obtained by titrating 100nM of cpOHPFA_1952-20-BLA-253 chimera with increasing concentrations of derivatized 17-OPH shown in (c). The data was fitted to a quadratic equation to obtain the K_D values.

(e) As in (c) but using an equimolar mixture of derivatized 17-OPH and streptavidin as titrant.

(f) Association and dissociation kinetics of 17-OPH derivative shown in (c) immobilized on streptavidin chip against wt OHPFA_1952 17-OPH binding domain titrated from 100 to 400 nM. The rates obtained from the fit of the data and the calculated K_D values are displayed on the plot.

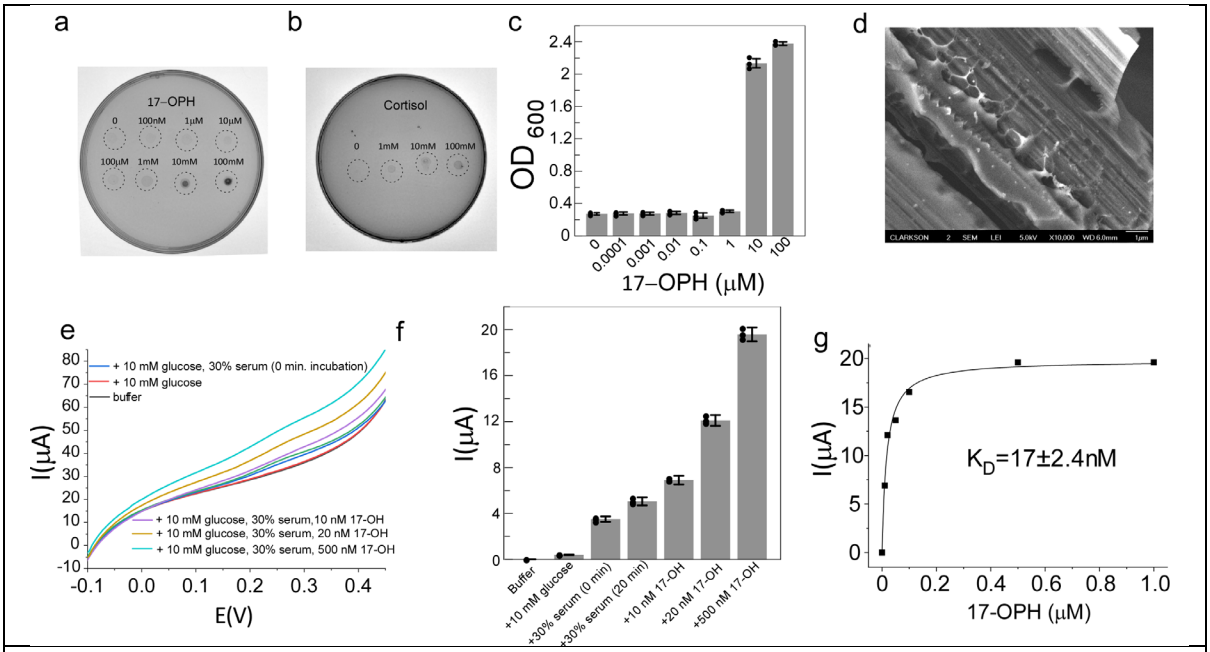
(g) As in (f) but using circular permutated cpOHPFA_1952-20 variant as titrant.

(h) ITC thermogram showing differential power during sequential injections of 500 μM solution of wt OHPFA_1952 17-OPH binding into the cell containing 50 μM 17-OPH.

(i) Integrated binding isotherm obtained by plotting the normalized heat per injection (ΔH, kcal/mol) against the molar ratio of OHPFA_1952 to 17-OPH. The data were fitted to a one-site binding model and the thermodynamic parameters, including K_D, ΔH, ΔG, and –TΔS, are shown in the figure.

(j) As in (h) but with circular permutated cpOHPFA_1952-20 variant.

(k) As in (i) but using circular permutated cpOHPFA_1952-20 variant.



Supplementary figure 13. Applications of the developed protein switches

(a) Agar plate containing ampicillin and chloramphenicol antibiotics seeded with *E. coli* strain DH5α expressing 2cpOHPFA1952-20-BLA-41-197 BLA chimera. After bacterial plating 2 μL drops of 17-OPH at the indicated concentration were spotted at the circled positions and bacterial growth was allowed to develop overnight.

(b) Same as (a) but spotting the indicated concentrations of cortisol.

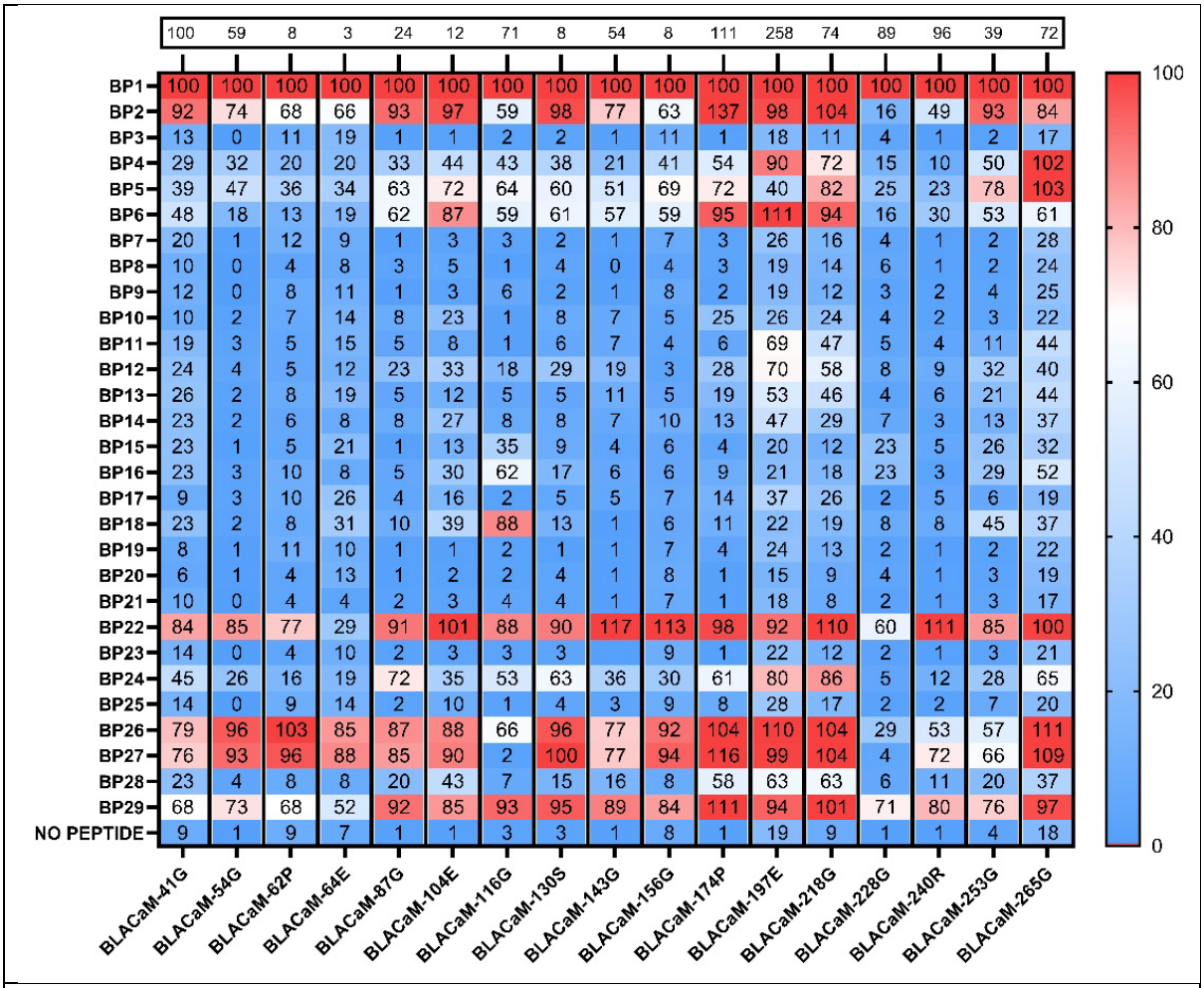
(c) Density plot of suspension cultures of *E. coli* DH5α transformed with 2cpohpFA1952-20-BLA-41-197-BLA in LB liquid medium in the presence of the indicated concentrations of 17–OPH. Error bars were calculated using values obtained from three independent bacterial cultures derived from different clones of a single transformation. Bars present mean values +/- SD.

(d) SEM microphotograph of working electrodes formed by carbon fibers decorated with graphene nanosheets. Multiple micrographs were taken and the representative one was chosen for display.

(e) A voltammogram GDH-17-OPH-based bioelectrode scanned at 5 mV/s rate *vs.* Ag/AgCl/3MKCl reference electrode at room temperature under the conditions indicated on the plot. Only oxidation output is shown on the voltammogram to simplify its visual inspection. The initial scan was performed in buffer containing 25 mM HEPES, pH 7.8, 100 mM Na_2SO_4, and 3 mM calcium acetate.

(f) A bar plot of the background adjusted values shown in figure (e). Error bars were calculated using values obtained from three independent scans of the same electrode under the corresponding conditions and present mean values +/- SD.

(g) A fit of the electric current values recorded on 17-OPH-GDH bioelectrode in the presence of 25 mM HEPES, pH 7.8, 100 mM Na_2SO_4 and 3 mM calcium acetate, 10mM glucose, 30% human serum and the increasing concentrations of 17-OPH. The K_D value obtained from the data fitting is displayed on the graph.



Supplementary figure 14. Heat map of the relative activity of the BLA-CaM mutant library exposed to 29 different CaM-BPs. The relative activity (%) of the BLA-CaM mutants was determined comparing their reaction rate (k_{obs}) with that measured with the calmodulin binding peptide BP1 (M13 peptide). Each column represents the activity of a chimera with the identity indicated by the insertion number. The last row reports the activity of the chimeras in the absence of ligand peptide. The heat map scale bar is shown on the right.

References:

1. Choi, J. H., Xiong, T. & Ostermeier, M. The interplay between effector binding and allostery in an engineered protein switch. *Protein Sci.* **25**, 1605–16 (2016).

2. Lee, G. R. *et al.* Small-molecule binding and sensing with a designed protein family. *bioRxiv* https://doi.org/10.1101/2023.11.01.565201 (2023) doi:10.1101/2023.11.01.565201.

3. Watson, J. L. *et al.* De novo design of protein structure and function with RFdiffusion. *Nature* **620**, 1089–1100 (2023).