UC Santa Cruz
UC Santa Cruz Previously Published Works

Title
De novo design of modular and tunable protein biosensors

Permalink

Journal

ISSN

Authors
Quijano-Rubio, Alfredo
Yeh, Hsien-Wei
Park, Jooyoung
et al.

Publication Date

DOI

Peer reviewed

De novo design of modular and tunable protein biosensors

Alfredo Quijano-Rubio[†,1,2], **Hsien-Wei Yeh**[†,1], **Jooyoung Park**[1], **Hansol Lee**[3], **Robert A. Langan**[1], **Scott E. Boyken**[1], **Marc J. Lajoie**[1], **Longxing Cao**[1], **Cameron M. Chow**[1], **Marcos C. Miranda**[1], **Jimin Wi**[4], **Hyo Jeong Hong**[4], **Lance Stewart**[1], **Byung-Ha Oh**[*,1,3], **David Baker**[*,1,5]

[1]Department of Biochemistry and Institute for Protein Design, University of Washington, Seattle, Washington 98195, USA

[2]Department of Bioengineering, University of Washington, Seattle, Washington 98195, USA

[3]Department of Biological Sciences, KAIST Institute for the Biocentury, Korea Advanced Institute of Science and Technology, Daejeon 34141, Republic of Korea

[4]Department of Systems Immunology, College of Biomedical Science, Kangwon National University, Chuncheon 200-701, Republic of Korea

[5]Howard Hughes Medical Institute, University of Washington, Seattle, Washington 98195, USA

Abstract

Naturally occurring protein switches have been repurposed for developing novel biosensors and reporters for cellular and clinical applications[1], but the number of such switches is limited, and engineering them is often challenging as each is different. Here, we show that a very general class of protein-based biosensors can be created by inverting the flow of information through *de novo* designed protein switches in which binding of a peptide key triggers biological outputs of interest[2]. The designed sensors are modular molecular devices with a closed dark state and an open luminescent state; binding of the analyte of interest drives switching from the closed to the open state. Because the sensor is based purely on thermodynamic coupling of analyte binding to sensor activation, only one target binding domain is required, which simplifies sensor design and allows direct readout in solution. We demonstrate the modularity of this platform by creating

[*]**Correspondence and requests for materials** should be addressed to D.B. or B.-H.O.
[†]These authors contributed equally to this work
Present Addresses:
J.P.: Sana Biotechnology, Inc, Seattle, United States
R.A.L., S.E.B., M.J.L.: Outpace Bio, Inc., Seattle, United States.

biosensors that, with little optimization, sensitively detect the anti-apoptosis protein Bcl-2, the IgG1 Fc domain, the Her2 receptor, and Botulinum neurotoxin B, as well as biosensors for cardiac Troponin I and an anti-Hepatitis B virus (HBV) antibody that achieve the sub-nanomolar sensitivity necessary to detect clinically relevant concentrations of these molecules. Given the current need for diagnostic tools for tracking COVID-19[3], we used the approach to design sensors of antibodies against SARS-CoV-2 protein epitopes and of the receptor-binding domain (RBD) of the SARS-CoV-2 Spike protein. The latter, which incorporates a *de novo* designed RBD binder[4], has a limit of detection of 15 pM and a signal over background of over 50-fold. The modularity and sensitivity of the platform should enable the rapid construction of sensors for a wide range of analytes and highlights the power of *de novo* protein design to create multi-state protein systems with new and useful functions.

Protein-based biosensors play important roles in synthetic biology and clinical applications, but thus far, biosensor design has been mostly limited to reengineering natural proteins[1]. However, finding analyte-binding domains that undergo sufficient conformational changes is challenging, and even when available, extensive protein engineering efforts are generally required to effectively couple them to a reporter domain[5,6]. Hence it is desirable to construct modular biosensor platforms that can be easily repurposed to detect different protein targets of interest. Modular systems have been developed for detecting antibodies[7–9] and small molecules[10,11], but systems for detecting proteins with very different structures, sizes and oligomerization states using semisynthetic protein platforms[12–14] or based on calmodulin switches[15,16], usually require considerable screening to find potential candidates due to limited predictability[17].

A protein biosensor can be constructed from a system with two nearly isoenergetic states, the equilibrium between which is modulated by the analyte being sensed. Desirable properties in such a sensor are (i) the analyte triggered conformational change should be independent of the details of the analyte, so the same overall system can be used to sense many different targets, (ii) the system should be tunable so that analytes with different binding energies and at relevant concentrations can be detected over a large dynamic range, and (iii) the conformational change should be coupled to a sensitive output. We hypothesized that these attributes could be attained by inverting the information flow in *de novo* designed protein switches in which binding to a target protein of interest is controlled by the presence of a peptide actuator[2]. We developed a system consisting of two protein components: (a) a 'lucCage' comprising a cage domain and a latch domain containing a target binding motif and a split luciferase fragment (SmBiT 114[18]), and (b) a "lucKey", containing a key peptide which binds to the open state of lucCage and the complementary split luciferase fragment (LgBit 11S[18], Fig. 1a). lucCage has two states: a closed state in which the cage domain binds the latch and sterically occludes the binding motif from binding target and SmBiT from combining with LgBit to reconstitute luciferase activity, and an open state in which these binding interactions are not blocked, and lucKey can bind the cage domain. Association of lucKey with lucCage results in the reconstitution of luciferase activity (Fig. 1a, right). The thermodynamics of the system are tuned such that the binding free energy of lucKey to lucCage (ΔG_{CK}) is insufficient to overcome the free energy cost of lucCage opening (ΔG_{open}) in the absence of target ($\Delta G_{open} - \Delta G_{CK} \gg 0$), but in the

presence of the target, the additional binding free energy of the latch to the target (ΔG_{LT}) drives latch opening and luciferase reconstitution ($\Delta G_{open} - \Delta G_{CK} - \Delta G_{LT} \ll 0$) (Fig. 1b,c). This system satisfies properties (i) and (ii) above, as a wide range of binding activities can be caged, and since the switch is thermodynamically controlled, the lucKey and target binding energies can be adjusted to achieve activation at the relevant target concentrations. Because lucKey and lucCage are always the same, the system is modular: the same molecular association can be coupled to the binding of many different targets. Bioluminescence provides a rapid and sensitive readout of analyte driven lucCage-lucKey association, satisfying property (iii).

The states of this biosensor system are in thermodynamic equilibrium, with the tunable parameters ΔG_{open} and ΔG_{CK} governing the populations of the possible species, along with the free energy of association of the analyte to the binding domain ΔG_{LT} (Fig. 1b). We simulated the dependence of the sensor system on ΔG_{open} (Extended Data Fig. 1a), ΔG_{LT} (Extended Data Fig. 1b), and the concentration of analyte and the sensor components (Extended Data Fig. 1c,d). The sensitivity of analyte detection is a function of ΔG_{LT}, with a lower limit of roughly one-tenth the K_D for analyte binding (Extended Data Fig. 1b). Above this lower limit, varying the concentration of lucCage and lucKey enables responding to different target concentration ranges (Extended Data Fig. 1c,d). Sensitivity can be further modulated by tuning the strength of the intramolecular cage-latch interaction and the intermolecular cage-key interaction (ΔG_{open} and ΔG_{CK}): for example too tight cage-latch interaction results in low signal in the presence of target, and too weak an interaction results in high background in the absence of target (Extended Data Fig. 1a,e). Our design strategy aims to find this balance through modulating ΔG_{open} and ΔG_{CK} by varying the length of the latch (and key) helix and by introducing either favorable hydrophobic or unfavorable buried polar interactions at the cage-latch/key interface[2] (Extended Data Fig. 1f,g).

Designing tunable lucCage sensors

To design sensors based on these principles, we developed a "GraftSwitchMover" Rosetta-based method to identify placements of target binding peptides within the latch such that the resulting protein is stable in the closed state and the interactions with the target are blocked (see Supplementary methods). As a first test, we grafted the SmBiT peptide and the Bim peptide in the closed state of the optimized asymmetric LOCKR switch described in Langan et al,[2] (Extended Data Fig. 2). SmBiT adopts a β-strand conformation within the luciferase holoenzyme, but we assumed that it could adopt a helical secondary structure in the context of the helical bundle scaffold, since secondary structure can be context dependent[19]. We sampled different threadings for the two peptide sequences across the latch, selected the lowest energy solutions (Extended Data Fig. 2a) and expressed twelve designs in *E. coli*. We mixed the designs with lucKey in a 1:1 ratio, then added Bcl-2, which binds with nanomolar affinity to Bim[20], and observed a rapid increase in luminescence (Extended Data Fig. 2b,f; we refer to the best of these as lucCageBim), showing that the LOCKR actuator[2] operated in reverse can function as a biosensor. The analyte detection range could be tuned by varying the concentration of the sensor (lucCage + lucKey) (Extended Data Fig. 2g) as anticipated in our model simulations (Extended Data Fig. 1c). lucCageBim has SmBiT at position 312 in

the latch (SmBiT312; Extended Data Fig. 2d). The Cage with this placement ("lucCage") was used as the base scaffold for the biosensors described below.

lucCage sensors with miniprotein sensing domains

We next investigated the incorporation of a range of binding modalities for analytes of interest within lucCage by developing methods for computationally caging target-binding proteins, rather than peptides, in the closed state (See supplementary methods). As a test case, we caged the *de novo* designed Influenza A H1 hemagglutinin (HA)[21] binding protein HB1.9549.2 into a shortened version of the LOCKR switch[22] (sCage), optimized to improve stability and facilitate crystallization efforts (Fig. 2a). Two of five designs were functional, and bound HA in the presence but not the absence of key (Extended Data Fig. 3b). The crystal structure of the best design, sCageHA_267–1S, determined to 2.0 Å resolution (Table S1, PDB ID: 7CBC), showed that all HA-binding interface residues except one (F273) interact with the cage domain (blocking binding of the latch to the target) as intended by design (Fig. 2a, Extended Data Fig. 3a–c).

With this structural validation of the design concept, we next sought to develop sensors for Botulinum neurotoxin B (BoNT/B), the immunoglobulin Fc domain, and the Her2 receptor. To do so, we grafted a *de novo* designed binder for Botulinum neurotoxin (Bot.0671.2)[21], the C domain of the generic antibody binding protein Protein A[23], and a Her2-binding affibody [24], into lucCage. After screening a few designs for each target (Extended Data Fig. 4–5), we obtained highly sensitive lucCages (lucCageBot, lucCageProA, and lucCageHer2) that can detect BoNT/B (Fig. 2b, Extended Data Fig. 4), human IgG Fc domain (Fig. 2c, Extended Data Fig. 5a–d), and Her2 receptor (Fig. 2d; Extended Data Fig. 5e–h) respectively, demonstrating the modularity of the platform. The designed sensors respond within minutes upon adding the target, and their sensitivity can be tuned by changing the concentration of lucCage and lucKey (Fig. 2). With further development, these sensors could enable rapid and low-cost detection of botulinum neurotoxins in the food industry[25], and detection of serological levels of soluble Her2 (>15 ng/mL; within the detection range of lucCageHer2) associated with metastatic breast cancer[26].

lucCage sensor for cardiac troponin

We next designed sensors for cardiac troponin I (cTnI), which is the standard early diagnostic biomarker for acute myocardial infarction (AMI)[27]. We took advantage of the high-affinity interaction between cTnT, cTnC, and cTnI (Fig. 3a) and designed eleven biosensor candidates by inserting 6 truncated cTnT sequences at different latch positions (Extended Data Fig. 6a). The best candidate, lucCageTrop627, was able to detect cTnI but not at sufficiently low levels for clinical use because the rule-in and rule-out levels of cTnI assay for diagnosis of AMI in patients are in the low pM range[27]. As noted above, the limit of detection (LOD) of our sensor platform is about 0.1 x Kd of the latch-target affinity (K_{LT}), we further increased the affinity of our sensor to cTnI by fusing cTnC to its terminus (Extended Data Fig. 6b–d). The resulting sensor, lucCageTrop, has a single-digit pM LOD suitable for quantification of clinical samples (Fig. 3b, Extended Data Fig. 6e,f).

lucCage sensors for anti-HBV and anti-SARS-CoV-2 antibodies

Detection of specific antibodies is important for monitoring the spread of a pathogen in a population[28], the success of vaccination[29], and levels of therapeutic antibodies[9]. To adapt our system for antibody serological analyses, we sought to incorporate linear epitopes recognized by the antibodies of interest into lucCage. We first developed a sensor for antibodies against the PreS1 domain of the hepatitis B surface protein L[30]. The best of 8 designs tested, lucCageHBV had a ~150% increase in luciferase activity upon addition of the anti-HBV antibody HzKR127–3.2[31] (Extended Data Fig. 7a–d). To further improve the dynamic range and LOD of lucCageHBV (~2 nM, Extended Data Fig. 7e), a second copy of the peptide was introduced at the end of the latch to increase latch affinity with the bivalent antibody (K_{LT}) (Fig. 3c,d). The resulting design, lucCageHBVα, had a LOD of 260 pM and a dynamic range of 225% (Fig. 3e; Extended Data Fig. 7g–i), with a luminescence intensity easily detectable with a camera (Extended Data Fig. 7j). Hence the platform is applicable to detecting specific antibodies with a LOD in the range for monitoring therapeutic antibodies[32].

We next sought to use the lucCageHBV sensor to detect hepatitis B surface antigen (HBsAg). Since our sensors are under thermodynamic control, we hypothesized that the pre-assembled sensor-antibody complex would re-equilibrate in the presence of the target HBsAg protein, PreS1, with antibody redistributing to bind free PreS1 instead of the epitope on lucCageHBV (Fig. 3f). Indeed, the luminescence of lucCageHBV plus HzKR127–3.2 mixture decreased shortly upon addition of the PreS1 domain (Fig. 3g); the sensitivity of this readout enabled quantification of PreS1 concentration in a clinically relevant range[33] (Fig. 3h, Extended Data Fig. 7f).

The COVID-19 pandemic has showcased the urgent need for diagnostics for both the SARS-CoV-2 virus and antiviral antibodies[3]. To design sensors for anti-SARS-CoV-2 antibodies, we first identified from the literature highly immunogenic linear epitopes in the SARS-CoV[34,35] and SARS-CoV-2 proteomes[36] that are not present in "common" strains of *coronaviridae*. Among these, we focused on two epitopes in the Membrane (M) and Nucleocapsid (N) proteins found to be recognized by SARS and COVID-19[35,36] patient sera for which cross-reactive animal-derived antibodies are commercially available (see Methods). We designed sensors for each epitope and identified designs that specifically responded to anti-M and anti-N antibodies (Extended Data Fig. 8a,b). These sensors reached full signal in 2–5 minutes and had a ~50–70% dynamic range in response to low nanomolar amounts of antibodies (Fig. 4a,b, Extended Data Fig. 8c,d). For robust serological analysis, generation of an expanded set of more sensitive sensors spanning multiple SARS-CoV-2 epitopes recognized will be necessary.

lucCage sensors for SARS-CoV-2 Spike protein

To create sensors capable of detecting SARS-CoV-2 viral particles directly, we integrated a *de novo* designed picomolar affinity binder to the receptor-binding domain (RBD) of the SARS-CoV-2 Spike protein named LCB1[4] into the lucCage format (Fig. 4c). Of 13 candidates tested, the best, which we refer to as lucCageRBD, was able to detect both

monomeric RBD and the full trimeric SARS-CoV2 spike protein[37] with 15 pM LOD and >1700% dynamic range for the RBD detection (Fig. 4c, Extended Data Fig. 9). We further increased the dynamic range of lucCageRBD to 5300% by using a short version of lucKey to tune K_{CK} (Extended data Fig. 10a–c).

To evaluate the ability of our sensor platform to function in complex biological matrices, we compared RBD detection by lucCageRBD in buffer, simulated nasal matrix[38], and human serum, and observed only a minor reduction in the latter two conditions (Fig. 4c). Following a suggestion by Maarten Merkx[39], we controlled for variation in absolute luminescence signal in spiked serum samples from four different donors and spiked simulated nasal matrix using a BRET internal reference[40] for internal calibration, and found that with such calibration the RBD could be accurately quantified without compromising sensor dynamic range (Extended Data Fig. 11). These results suggest that it should be feasible to implement the lucCage system for future point-of-care applications.

To test the specificity of the biosensors developed in this work, we measured the activation kinetics of each lucCage in response to all target proteins used in this work. Each sensor responded rapidly and sensitively to its cognate target, but not to any of the others (Fig. 4d). For the most part, the actual sensors (see Table S2 and Table S3) perform as predicted by the simple thermodynamic model; for example, experiments at varying key and sensor concentrations suggest little coupling between parameters. However, there is considerable variation between different sensors in the level of activation at saturating target concentrations or high lucKey concentrations, which for most is lower than that expected for the complete luciferase reconstitution predicted by the model (Extended Data Fig. 10d–g and Table S4). This may be a consequence of steric interference between target binding to the latch and luciferase reconstitution as the target binding motif and the luciferase SmBiT are adjacent to each other in the latch; such interference could be resolved by increasing the separation between the two in the switch. The potential of the lucCage system is illustrated by the high dynamic range (5300%) and picomolar sensitivity of the lucCageRBD sensor: the near optimal K_{open} value results in a very low background in the absence of target without compromising the extent of activation at low target concentrations.

Discussion

It is instructive to put our sensors in the context of the multiple protein-based biosensor platforms that have been developed over the years with considerable success (see Supplementary discussion, and Table S5). Our sensor platform is based on the thermodynamic coupling between defined closed and open states of the system, thus, its sensitivity depends on the free energy change upon the sensing domain binding to the target but not the specific geometry of the binding interaction (the semi-synthetic small molecule sensors[10,11] also have this property). This enables the incorporation of various binding modalities, including small peptides, globular miniproteins, antibody epitopes and *de novo* designed binders, to generate sensitive sensors for a wide range of protein targets with little or no optimization. For point of care (POC) applications, our system, like other luminescence based protein biosensor platforms[8], has the advantages of being homogeneous, no-wash, and a nearly instantaneous readout; the quantification of

luminescence can be carried out with inexpensive and accessible devices such as a cell phone camera[8]. In hospital settings, the ability to modularly design sensors with identical readouts for diverse targets could enable quick readout of large numbers of different compounds using an array of hundreds of different sensors.

Up until recently, the focus of *de novo* protein design was on the design of proteins with new structures corresponding to single deep free energy minima; our results highlight the progress in the field which now enables more complex multistate systems to be readily generated. Our sensors, like other *de novo* designed proteins, are expressed at high levels in cells and are very stable, which should considerably facilitate their manufacturing and distribution. As highlighted by the outstanding performance of the lucCageRBD sensor, there is a strong synergy between the general "molecular device" architecture of our platform and *de novo* designed high-affinity miniprotein binders[4,21] (these *de novo* miniproteins also synergize with other platforms[41]). As the power of computational design continues to increase, it should become possible to detect an ever wider range of targets with higher sensitivity using lucCage sensors. Beyond biosensors, our results highlight the potential of *de novo* protein design to create more general solutions for current day challenges than can be achieved by repurposing native proteins that have evolved to solve completely different challenges.

Methods

Design of the sensor system: lucCage and lucKey

The low affinity SmBiT 114 (VTGYRLFEEIL)[18] was grafted into the latch of the asymmetric LOCKR switch described in Langan et al, 2019[2] using GraftSwitchMover, a RosettaScripts-based protein design algorithm (See Supplementary methods for details). The grafting sampling range was assigned between residues 300–330. The resulting designs were energy-minimized, visually inspected and selected for subsequent gene synthesis, protein production and biochemical analyses. The best SmBit position on the latch was experimentally determined to be an insertion at residue 312, as described in Extended Data Fig. 2. This design was named lucCage. lucKey was assembled by genetically fusing the LgBit of NanoLuc[18] to the key peptide described in Langan et al, 2019. All protein sequences are listed in Table S6.

Computational grafting of sensing domains into lucCage

Peptides and epitopes: The amino acid sequence for each sensing domain was grafted using Rosettascripts[42] GraftSwitchMover into all α-helical registers between residues 325–359 of lucCage. In the cases where the desired sequence to be inserted exceeded the length of the lucCage latch, we made use of Rosetta Remodel [43] to model the C-terminus extension of lucCage (See Supplementary methods for details). The resulting lucCages were energy-minimized using Rosetta fast relax[44], visually inspected and typically less than ten designs were selected for subsequent protein production and biochemical characterization.

Protein domains: the main secondary structure element segment forming the interface of the binding protein domain with the target was identified. The amino acid sequence was

extracted and grafted into lucCage using the GraftSwitchMover or Rosetta Remodel as described above. Then, we used MergePDBMover and Pymol 2.0 to align, model and visualize the full-length binding domain in the context of the switch (See Supplementary methods for details). The designs were energy-minimized using Rosetta fast relax and visually inspected for selection.

Synthetic gene construction

The designed protein sequences were codon optimized for *E. coli* expression and ordered as synthetic genes in pET21b+ or pET29b+ *E. coli* expression vectors. The synthetic gene was inserted at the NdeI and XhoI sites of each vector, including an N-terminal hexahistidine tag followed by a TEV protease cleavage site and a stop codon was added at the C terminus.

General procedures for bacterial protein production and purification

The *E. coli* Lemo21(DE3) strain (NEB) was transformed with a pET21b+ or pET29b+ plasmid encoding the synthesized gene of interest. Cells were grown for 24 hours in LB media supplemented with carbenicillin or kanamycin. Cells were inoculated at a 1:50 mL ratio in the Studier TBM-5052 autoinduction media supplemented with carbenicillin or kanamycin, grown at 37 °C for 2–4 hours, and then grown at 18 °C for an additional 18 h. Cells were harvested by centrifugation at $4000g$ at 4 °C for 15 min and resuspended in 30 ml lysis buffer (20 mM Tris-HCl pH 8.0, 300 mM NaCl, 30 mM imidazole, 1 mM PMSF, 0.02 mg/mL DNAse). Cell resuspensions were lysed by sonication for 2.5 minutes (5 second cycles). Lysates were clarified by centrifugation at $24,000g$ at 4 °C for 20 min and passed through 2 ml of Ni-NTA nickel resin (Qiagen, 30250) pre-equilibrated with wash buffer, (20 mM Tris-HCl pH 8.0, 300 mM NaCl, 30 mM imidazole). The resin was washed twice with 10 column volumes (CV) of wash buffer, and then eluted with 3 CV of elution buffer (20 mM Tris-HCl pH 8.0, 300 mM NaCl, 300 mM imidazole). The eluted proteins were concentrated using Ultra-15 Centrifugal Filter Units (Amicon) and further purified by using a Superdex™ 75 Increase 10/300 GL (GE Healthcare) size exclusion column in Tris Buffered Saline (TBS; 25 mM Tris-HCl pH 8.0, 150 mM NaCl). Fractions containing monomeric protein were pooled, concentrated, and snap-frozen in liquid nitrogen and stored at −80 °C.

In vitro bioluminescence characterization

A Synergy Neo2 Microplate Reader (BioTek) was used for all in vitro bioluminescence measurements. Assays were performed in 1:1=DPBS (with calcium, Gibco):Nano-Glo (Promega) assay buffer for cTnI sensors while 1:1=HBS-EP (GE Healthcare Life Sciences) :Nano-Glo assay buffer was used for other sensors. 10X lucCage, 10X lucKey, and 10X target proteins of desired concentrations were first prepared from stock solutions. For each well of a white opaque 96-well plate, 10 μL of 10X lucCage, 10 μL of 10X lucKey, and 20 μL of buffer were mixed to reach the indicated concentration and ratio. The lucCage and lucKey components were incubated for 60 minutes at RT to enable pre-equilibration. The plate was centrifuged at $1000 \times g$ for 1 min and incubated at RT for additional 10 min. Then, 50 μL of 50X diluted furimazine (Nano-Glo luciferase assay reagent, Promega) was added to each well. For assays containing serum or simulated nasal matrix (110mM NaCl, 1% w/v mucin, 10μg/mL human genomic DNA[38]), buffer composition was replaced by the

biological matrix. Bioluminescence measurements in the absence of target were taken every 1 min post-injection (0.1 s integration and 10 s shaking during intervals). After ~15 min, 10 μL of serially diluted 10X target protein plus a blank was injected and bioluminescence kinetic acquisition continued for a total of 2 h. To derive EC_{50} values from the bioluminescence-to-analyte plot, the top three peak bioluminescence intensities at individual analyte concentrations were averaged, subtracted from blank, and used to fit the sigmoidal 4PL curve. To calculate the limit of detection (LOD), the linear region of bioluminescence responses of sensors to its analyte was extracted and a linear regression curve was obtained. It was used to derive the standard deviation (s.d.) of the response and the slope of the calibration curve (S). The LOD was determined as 3×(s.d./S).

Detection of spiked RBD in human serum specimens

Serum specimens were derived from excess plasma or sera from adults (>18 yo) of both genders kindly provided by the Director of the Clinical Chemistry Division, the hospital of University Washington. All anonymized donor specimens were provided de-identified. Since the donors consented to have their excess specimens be used for other experimental studies, they could be transferred to our study without additional consent. All samples were passed through 0.22 μm filters before use. 10 μL of 10X serial diluted monomeric RBD (167–0.69nM), 5 μL of 20X lucCage (20nM), 5 μL of 20X lucKey (20nM), 5 μL of 20X Antares2 (2nM), and 10, 20, 25, or 50 μL of human donor serum or simulated nasal matrix were mixed with 1:1=HBS:Nano-Glo assay buffer to reach a total volume of 75 μL. The plate was centrifuged at $1000 \times g$ for 1 min. Then, 25 μL of 25X diluted furimazine in buffer was added to each well. Bioluminescence signals were recorded from both 470/40 nm and 590/35 nm channels every 1 min for a total of 1h. Ratio at each time point was calculated by the equation described in Extended Data Figure 11b. Monomeric SARS-CoV-2 RBD was expressed and purified as described elsewhere[45].

Biolayer interferometry (BLI)

Protein-protein interactions were measured by using an Octet® RED96 System (ForteBio) using streptavidin-coated biosensors (ForteBio). Each well contained 200 μL of solution, and the assay buffer was HBS-EP+ Buffer (GE Healthcare Life Sciences, 10 mM HEPES pH 7.4, 150 mM NaCl, 3 mM EDTA, 0.05% v/v Surfactant P20) + 0.5% Non-fat dry milk blotting grade blocker (BioRad). The biosensor tips were loaded with analyte peptide/protein at 20 μg/mL for 300 s (threshold of 0.5 nm response), incubated in HBS-EP+ Buffer for 60 s to acquire the baseline measurement, dipped into the solution containing Cage and/or Key for 600 s (association step) and dipped into the HBS-EP+ Buffer for 600 s (dissociation steps). The binding data were analyzed with the ForteBio Data Analysis Software version 9.0.0.10.

Design and characterization of lucCageBim

The Bim peptide sequence (EIWIAQELRRIGDEFNAYYA) was threaded into the lucCage scaffold as described in the "Design of sensing domains into lucCage" section. The selected designs were expressed in *E. coli*, purified and characterized for luminescence activation. The bioluminescence detection signal was measured for each design lucCage at 20 nM

mixed with lucKey at 20 nM, in the presence or absence of target Bcl-2 protein at 200nM. Recombinant Bcl-2 was produced as described somewhere else[46].

Design and characterization of lucCageHer2, lucCageProA, lucCageBot and lucCageRBD

The main binding motifs of the Bot.0671.2 de novo binder, *S. aureus* Protein A domain C (SpaC), the Her2 affibody and the *de novo* RBD binder LCB1 were threaded into lucCage as described in the "Design of sensing domains into lucCage" section (See Table S3 and Table S6 for sequences). The selected designs were expressed in *E. coli*, purified and characterized for luminescence activation. The designs were screened by measuring bioluminescence signal for each design lucCage at 20 nM mixed with lucKey at 20 nM, in the presence or absence of 200 nM target protein. The target proteins used were: Botulinum Neurotoxin B HcB expressed as previously described[47], human IgG1 Fc-HisTag (AcroBiosystems, Cat. No. IG1-H5225) and human Her2-HisTag (AcroBiosystems, Cat. No. HE2-H5225). Monomeric SARS-CoV-2 RBD and the trimeric SARS-CoV-2 spike protein (Hexapro pre-stabilized version[37]) were expressed and purified as described previously[45].

Design and characterization of lucCageTrop

The cardiac Troponin T (cTnT) binding motif sequence was truncated into fragments of different length (see Extended Data Fig. 6) and threaded into the lucCage scaffold as described in the "Design of sensing domains into lucCage" section. The selected designs were expressed in *E. coli*, purified and characterized for luminescence activation. The designs were screened by measuring bioluminescence signal for each design lucCage at 20 nM mixed with lucKey at 20 nM in the presence or absence of 100 nM cardiac Troponin I (Genscript, Cat. No. Z03320–50). Subsequently, lucCageTrop, an improved version by fusion to cardiac Troponin C (cTnC), was created by genetically fusing the following sequence to the C terminus of lucCageTrop627.

Design and characterization of lucCageHBV and lucCageHBVα

The binding motif (GANSNNPDWDFN) of the PreS1 domain was threaded into the lucCage scaffold at every position after residues 336 using the Rosetta GraftSwitchMover. Following the Rosetta FastRelax protocol, eight designs were selected for protein production. The designs were screened by measuring bioluminescence signal for each design lucCage (20 nM) and lucKey (20 nM) in the presence or absence of the anti-HVB antibody HzKR127–3.2 (100 nM) to select lucCageHBV. Subsequently, lucCageHBVα was constructed by genetically fusing a sequence containing a second antigenic motif (GGSGGGSSGFGANSNNPDWDFNPN) to lucCageHBV.

Design and characterization of lucCageSARS2-M and lucCageSARS2-N

Antigenic epitopes of the SARS-CoV-2 membrane protein (a.a. 1–31, 1–17 and 8–24) and the nucleocapsid protein (a.a. 368–388 and 369–382) were computationally grafted into lucCage as described in the "Design of sensing domains into lucCage" section. The selected designs were expressed in *E. coli*, purified and characterized for luminescence activation. All designs at 50nM were mixed with 50nM lucKey and experimentally screened for an increase in luminescence in the presence of rabbit anti-SARS-CoV Membrane polyclonal

antibodies (ProSci, Cat. No.: 3527) at 100nM or mouse anti-SARS-CoV Nucleocapsid monoclonal antibody (clone 18F629.1, NovusBio Cat. No. NBP2–24745) at 100 nM.

Design and characterization of sCageHA variants

HB1.9549.2 was embedded into the parental six-helix bundle for sCage design at different positions along the latch helix of the scaffold. To promote more favorable intramolecular interactions, three consecutive residues on the latch were intentionally substituted with glycine to allow for conformational freedom. The five designs were produced in *E. coli*. Biolayer interferometry analysis was performed with purified Cages (1 μM) and biotinylated Influenza A H1 hemagglutinin (HA)[21] loaded onto streptavidin-coated biosensor tips (ForteBio) in the presence or absence of the key (2 μM) using an Octet instrument (ForteBio).

Production and purification of HzKR127–3.2

The synthetic V_H and V_L DNA fragments were subcloned into the pdCMV-dhfrC-cA10A3 plasmid containing the human Cγ1 and Cκ DNA sequences. The vector was introduced into HEK 293F cells using Lipofectamine (Invitrogen), and the cells were grown in FreeStyle 293 (GIBCO) in 5% CO_2 in a 37 °C humidified incubator. The culture supernatant was loaded onto a protein A-sepharose column (Millipore), and the bound antibody was eluted by the addition of 0.2 M glycine–HCl (pH 2.7), followed by immediate neutralization with 1 M Tris–HCl (pH 8.0). The solution was dialyzed against 10 mM HEPES-NaOH (pH 7.4), and the purity of the protein was analyzed by SDS-PAGE.

Production and purification of the PreS1 domain

The DNA fragment encoding the PreS1 domain (residues 1–56) was cloned into the pGEX-2T (GE Healthcare) plasmid, and the protein was produced in the *E. coli* BL21(DE3) strain (NEB) at 18 ºC as a fusion protein with glutathion-S-transferase (GST) at the N-terminus. The cell lysates were prepared in a buffer solution (25 mM Tris-HCl pH 8.0, 300 mM NaCl), and clarified supernatant was loaded onto GSTBind™ Resin (Novagen). The GST-PreS1 domain was eluted with the same buffer containing additional 10 mM reduced glutathione, further purified using a Superdex™ 75 Increase 10/300 GL (GE Healthcare) size exclusion column, and concentrated to 34 μM.

Production of SCageHA_267–1S and its variants

sCageHA_267–1S and sCageHA_267–1S(E99Y/T144Y) were expressed at 18 °C in the *E. coli* LEMO21(DE3) strain (NEB) as a fusion protein containing a (His)$_{10}$-tagged cysteine protease domain (CPD) derived from *Vibrio cholerae* [48] at the C-terminus. The protein was purified using HisPur™ nickel resin (Thermo), a HiTrap Q anion exchange column (GE Healthcare) and a HiLoad 26/60 Superdex 75 gel filtration column (GE Healthcare). For Selenomethionine (SelMet)-labeling, an I30M mutation was introduced additionally to generate a sCageHA_267–1S(E99Y/T144Y/I30M) variant. This protein was expressed in the *E. coli* B834 (DE3) RIL strain (Novagen) in the minimal media containing SeMet, and purified according to the same procedure for purifying the other variants.

Crystallization and structure determination of sCageHA_267–1S

Two point mutations (Glu99Tyr and Thr144Tyr) were introduced in an attempt to induce favorable crystal packing interactions. Good-quality single crystals of sCageHA_267–1S(E99Y/T144Y/I30M) were obtained in a hanging-drop vapor-diffusion setting by micro-seeding in a solution containing 11% (v/v) ethanol, 0.25 M NaCl, 0.1 M TrisHCl (pH 8.5). The crystals required strict maintenance of the temperature at 25 °C. For cryoprotection, the crystals were soaked briefly in the crystallization solution supplemented with 15% *2,3-*butanediol and flash-cooled in the liquid nitrogen. A single-wavelength anomalous dispersion (SAD) data set was collected at the Se absorption peak and processed with *HKL2000*[49]. Se positions and initial electron density map were calculated using the AutoSol module in *PHENIX*[50]. The model building and structure refinement were performed by using *COOT*[51] and *PHENIX*.

Statistical analysis

No statistical methods were used to pre-determine the sample size. No sample was excluded from data analysis, and no blinding was employed. De-identified clinical serum samples were randomly used for spiking in target proteins. Results were successfully reproduced using different batches of pure proteins on different days. Unless otherwise indicated, data are shown as mean ± s.d., and error bars in figures represent s.d. of technical triplicate. BLI data was analyzed using ForteBio Data Analysis Software version 9.0.0.10. All data were analyzed and plotted using GraphPad Prism 8.

Data availability

The atomic coordinates of sCageHA_267–1S have been deposited in the Protein Data Bank (http://www.rcsb.org) under an accession code 7CBC. The original experimental data that supports the findings of this work are available from the corresponding authors upon request. Plasmids encoding the biosensor proteins described in this article are available from the corresponding authors upon request.

Code availability

The design models and RosettaScripts code used in the manuscript have been deposited to http://files.ipd.uw.edu/pub/de_novo_design_of_tunable_biosensors_2021/designcode_and_models.zip. The code for the numerical simulations shown in this manuscript are available at http://files.ipd.uw.edu/pub/de_novo_design_of_tunable_biosensors_2021/model_simulation.py
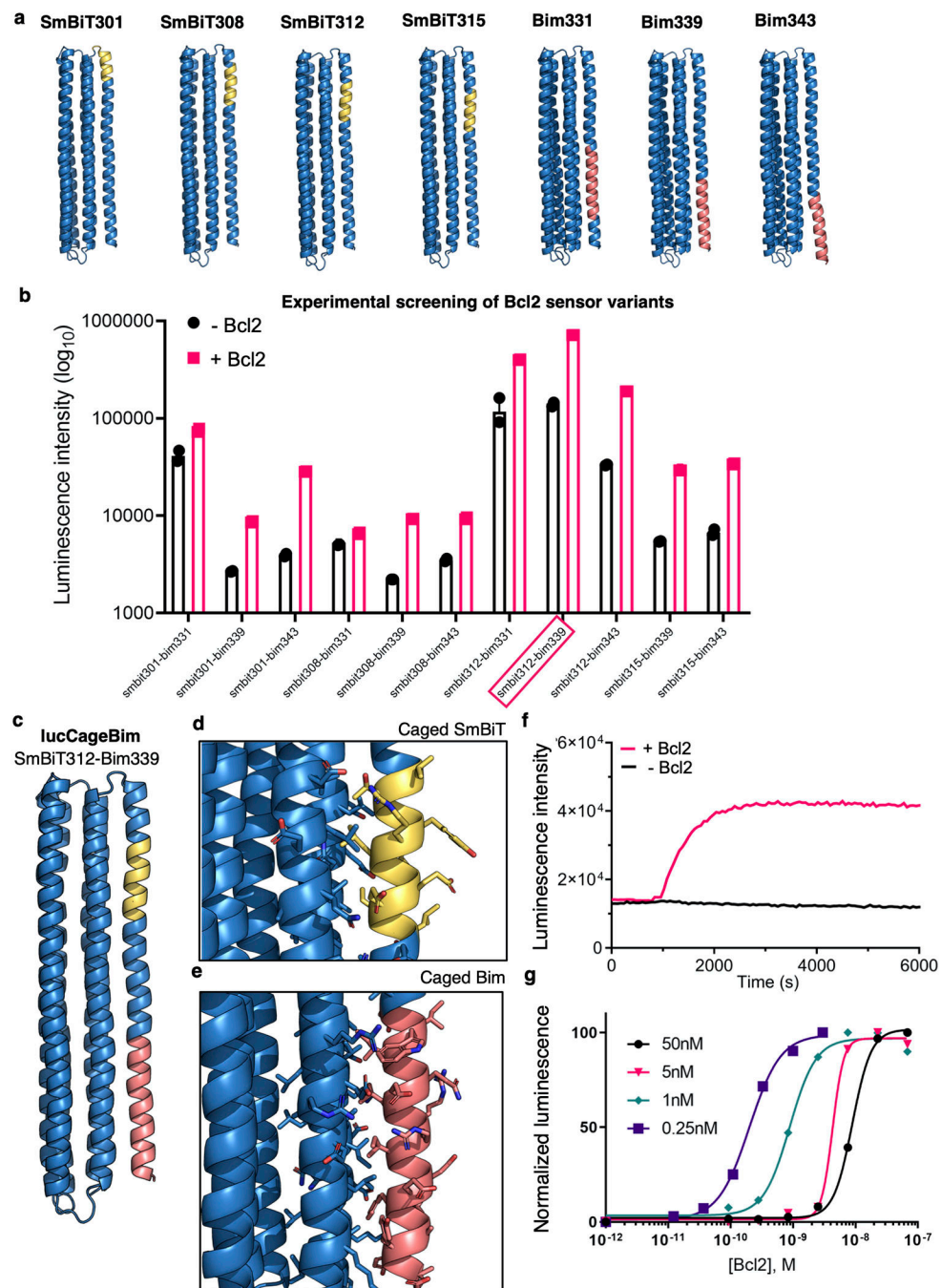
Extended Data
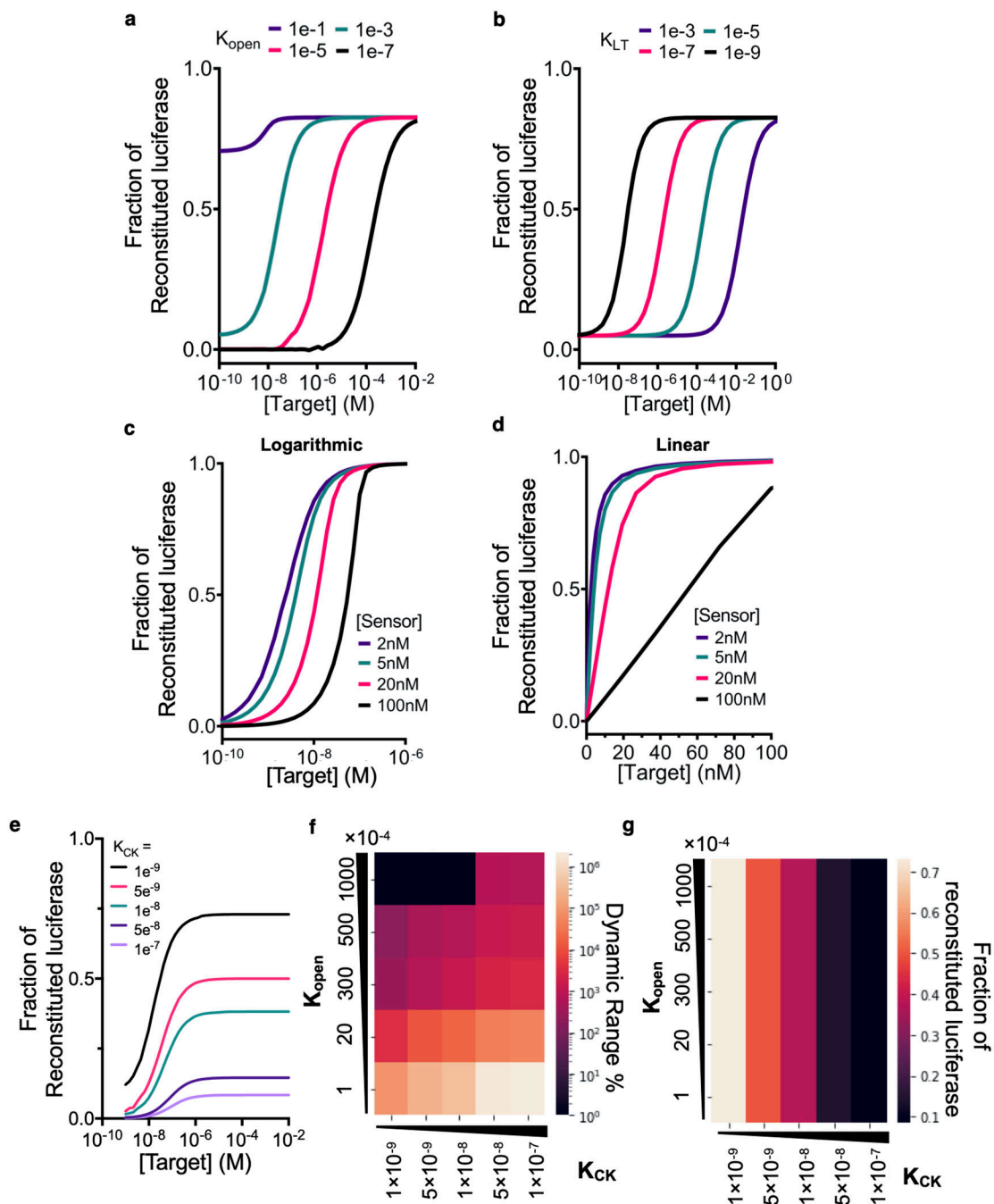
Extended Data Figure 1. Numerical simulations of the sensor thermodynamic equilibria showing the tunability of the lucCage platform to optimize sensitivity and dynamic range.
Numerical simulations of the coupled equilibria shown in Fig. 1b for different values of (**a**) K_{open}, (**b**) K_{LT}, (**c**)(**d**) [lucKey]$_{tot}$ and [lucCage]$_{tot}$ and (**e**) K_{open}. K_{LT} and K_{CK} were set to 1 $\times 10^{-3}$, 10^{-9} M, and 10^{-8} M respectively, and the concentration of the sensor components to 10:100 nM ([lucCage]$_{tot}$:[lucKey]$_{tot}$) except where explicitly indicated. **a**, Increasing G_{open} (smaller K_{open}) shifts the sensor response to higher analyte concentrations. **b**, The sensor

LOD is approximately $0.1 \times K_{LT}$; the driving force for opening the switch becomes too weak below this concentration. **c-d**, The effective target detection range can be tuned by changing the concentrations of the two sensor components. Simulation results shown in a logarithmic scale (**c**) or linear scale (**d**) for target concentration illustrate that the steepness of the response depends on the ratio of the sensor concentration to the K_D of the binding interaction (K_{LT}). **e**, K_{CK} values affect both species responsible for background and signal (species 6 and 7 in Fig. 1b, respectively), leading to different sensor dynamic ranges. **f-g**, Simulations with various K_{open} and K_{CK} values. Too large K_{open} value and strong lucCage-lucKey interaction (K_{CK}) increase the formation of the species 6 (in Fig. 1b). **f,** A heatmap representing the calculated sensor dynamic range according to the K_{open} and K_{CK} values. K_{open} exerts a predominant effect on the dynamic range, while K_{CK} provides an additional one-order of tunability. **g**, A heatmap showing the fraction of reconstituted luciferase (sensitivity) at saturating target concentration, indicating a trade-off of K_{CK} tuning.

Extended Data Figure 2. Determination of the optimal SmBiT position in lucCage and characterization of lucCageBim, a Bcl-2 biosensor.

a, Protein models showing the different threading positions of SmBiT (gold) and the Bim peptide (salmon) on the latch helix of the *de novo* LOCKR switch (blue). **b**, Experimental screening of 11 *de novo* Bcl-2 sensors. Eleven variants were generated by combining the SmBiT and Bim positions in (a) and characterized by activation of their luminescence upon addition of Bcl-2. Luminescence measurements were performed with each design (20 nM) and lucKey (20 nM) in the presence or absence of Bcl-2 (200 nM). SmBiT312-Bim339

(hence referred to as lucCageBim) was selected for posterior characterization due to its higher brightness, dynamic range and stability. **c-g**, Characterization of lucCageBim. **c**, Structural design model in ribbon representation. **d**, close up view showing the predicted interface of SmBiT (gold) and Cage (blue). **e**, close up view showing the predicted interface of Bim (salmon) and Cage (blue). **f**, Kinetic luminescence measurements upon addition of Bcl-2 (200 nM) to a mix of lucCageBim (20 nM) and lucKey (20 nM). **g**, Tunable sensitivity of lucCageBim to Bcl-2 by changing the concentrations of sensor (lucCageBim and lucKey) components (colored curves).
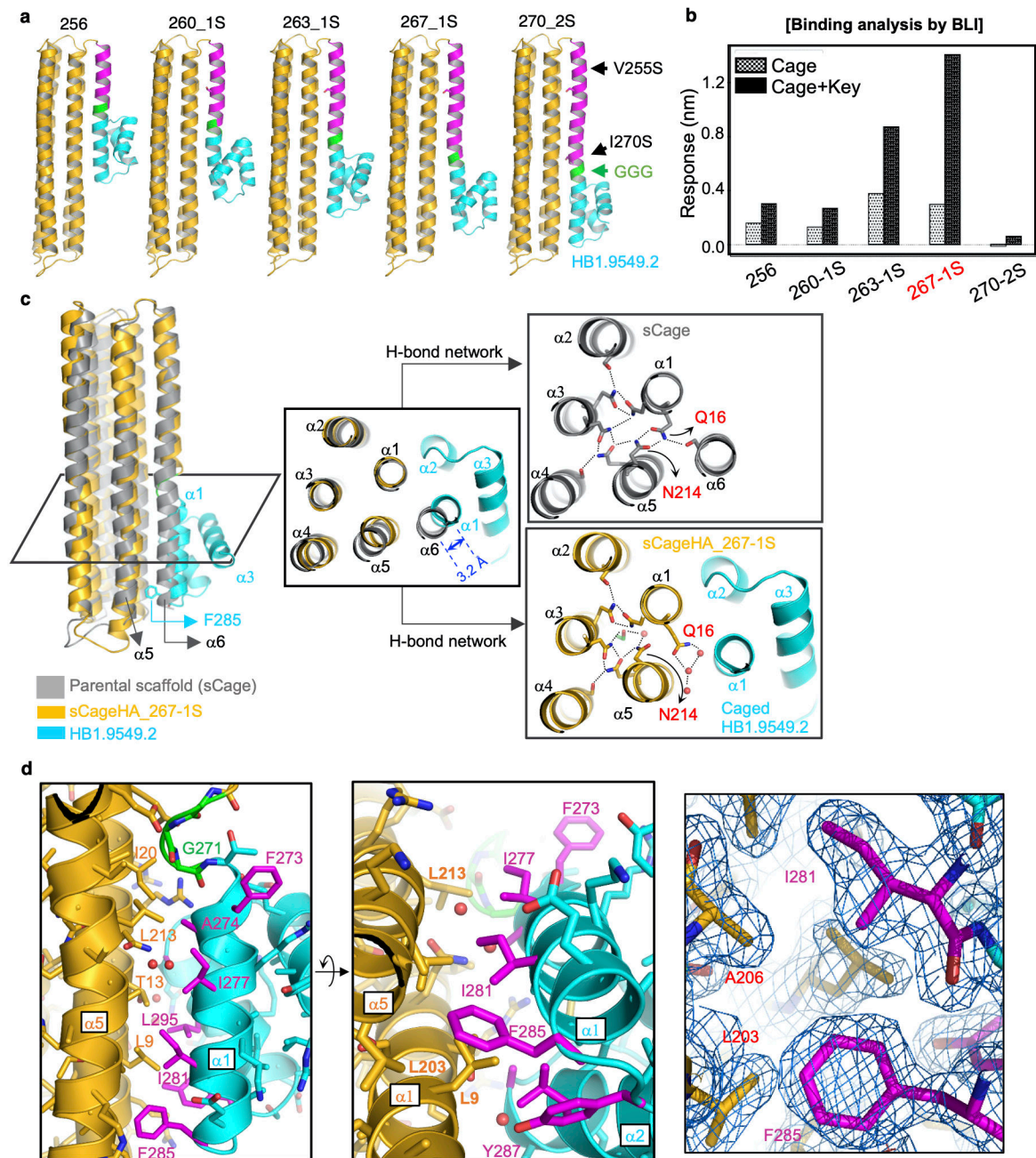
Extended Data Figure 3. Functional screening of sCageHA designs and crystal structure of sCageHA_267–1S.

a, Structural models of sCageHA designs with the embedded *de novo* binder HB1.9549.2. The HB1.9549.2 protein (cyan) was grafted into a parental six-helix bundle (sCage, yellow) at different positions along the latch helix (magenta) including three consecutive glycine residues (green). The black arrows indicate the additionally introduced single V255S (1S) or double V255S/I270S (2S) mutation(s) on the latch. **b**, Experimental validation of five sCageHA designs binding to HA in the presence or absence of the key by biolayer interferometry. The concentration of the sCages and the key were 1 μM and 2 μM, respectively. Each experiment was performed once. sCageHA_267–1S exhibited the highest fold of activation. **c**, Structural comparison showing the flexible nature of sCage to enable caging of HB1.9549.2. The structural model of sCage (grey) and the crystal structure of sCageHA_267–1S (gold) are superposed, and a narrow section (black box) is shown in an orthogonal view for detail. The N-terminal helix of HB1.9549.2 is displaced from the latch helix (α6) by 3.2 Å (middle panel) with a concomitant displacement of α5 and partial disruption of a hydrogen-bond network involving Q16 and N214 of sCage (right panels). **d**, A close up view of the intramolecular interactions of sCageHA_267–1S. The HA-binding residues are highlighted in magenta. Both the N-terminal helix (cyan α1) and the following helix (cyan α2) of HB1.9549.2 interact with the cage. The intramolecular interactions are all hydrophobic. The bulky hydrophobic side chain of F285 tightly abuts against the backbone atoms of α5 of sCage, which is unlikely to happen without a bending of α5. Unfavorable interactions are also found: F273 is solvent-exposed, and the Y287 hydroxyl group is buried in the apolar environment. The rightmost panel shows the quality of the electron density map.
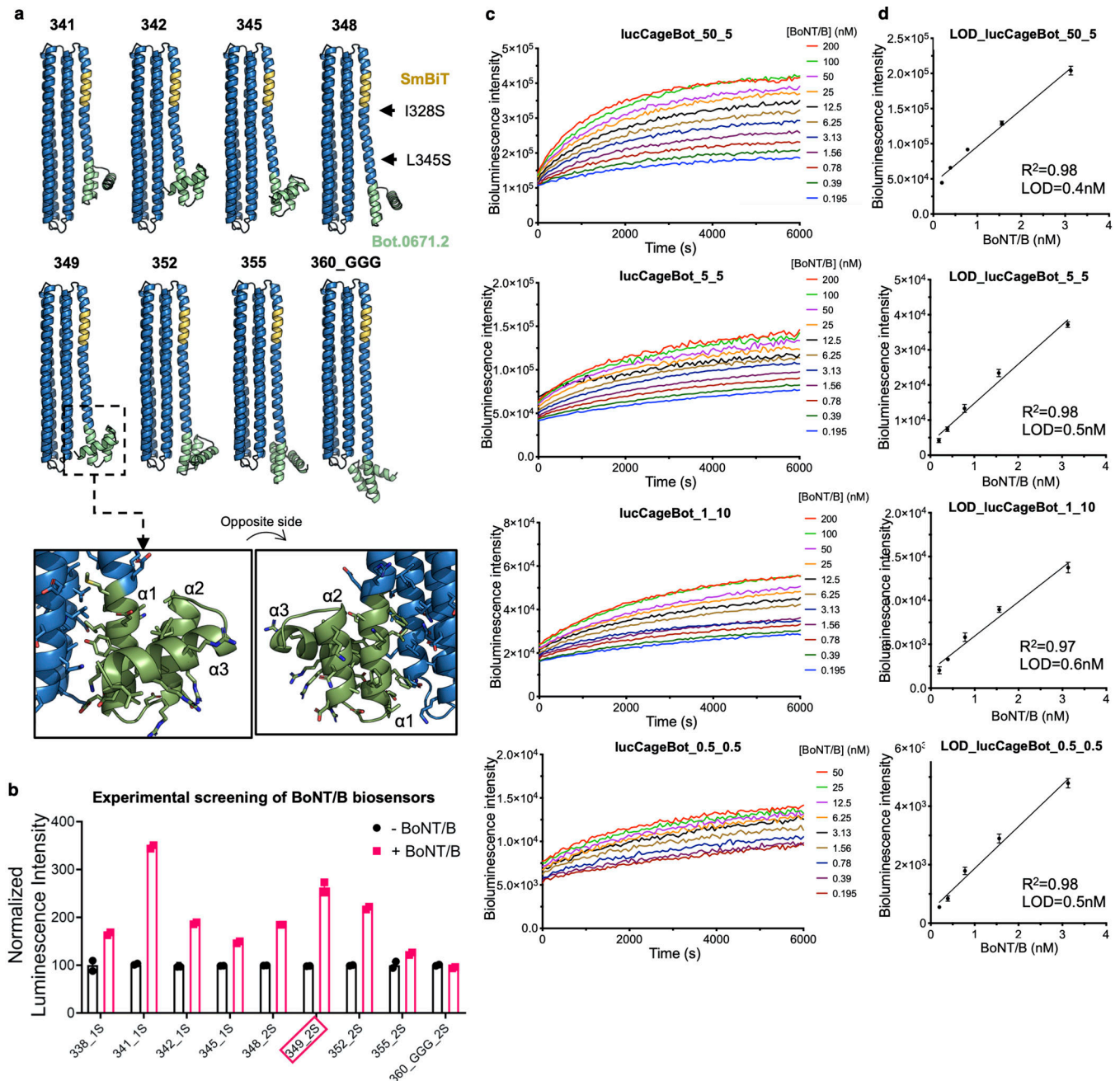
a, Structural models of the botulinum neurotoxin B (BoNT/B) sensor designs showing the different threading positions of Bot.0671.2 (green, PDB ID: 5VID) on the latch of lucCage (blue). The SmBiT peptide is shown in gold ribbon representation. I328S and L345S indicate mutations introduced to tune the latch-cage interface (1S=I328S, 2S=I328S/L345S) [2], and "GGG" indicates the presence of three consecutive glycine residues between the latch and the grafted protein. The black box shows a close-up view of the interface of Cage (blue) and Bot.0671.2 (green) in the 349_2S design. **b**, Experimental screening of 9 *de novo* BoNT/B sensors. Luminescence measurements were performed for each design (20 nM) and

lucKey (20 nM) in the presence or absence of the BoNT/B protein (200 nM). The luminescence values for each design were normalized to 100 in the absence of BoNT/B. Design 349_2S was selected as the best candidate due to high sensitivity and stability, and was named lucCageBot. **c**, Determination of lucCagerBot sensitivity. Bioluminescence was measured over 6000 s in the presence of serially diluted BoNT/B protein. From top to bottom - lucCageBot:lucKey concentration (nM) = 50:5, 5:5, 1:10, 0.5:0.5. **d**, Limit of detection (LOD) calculations for the sensor at different concentrations. From top to bottom - lucCageBot:lucKey concentration (nM) = 50:5, 5:5, 1:10, 0.5:0.5. All experiments were performed in triplicate, representative data are shown, and data are presented as mean values +/− s.d.
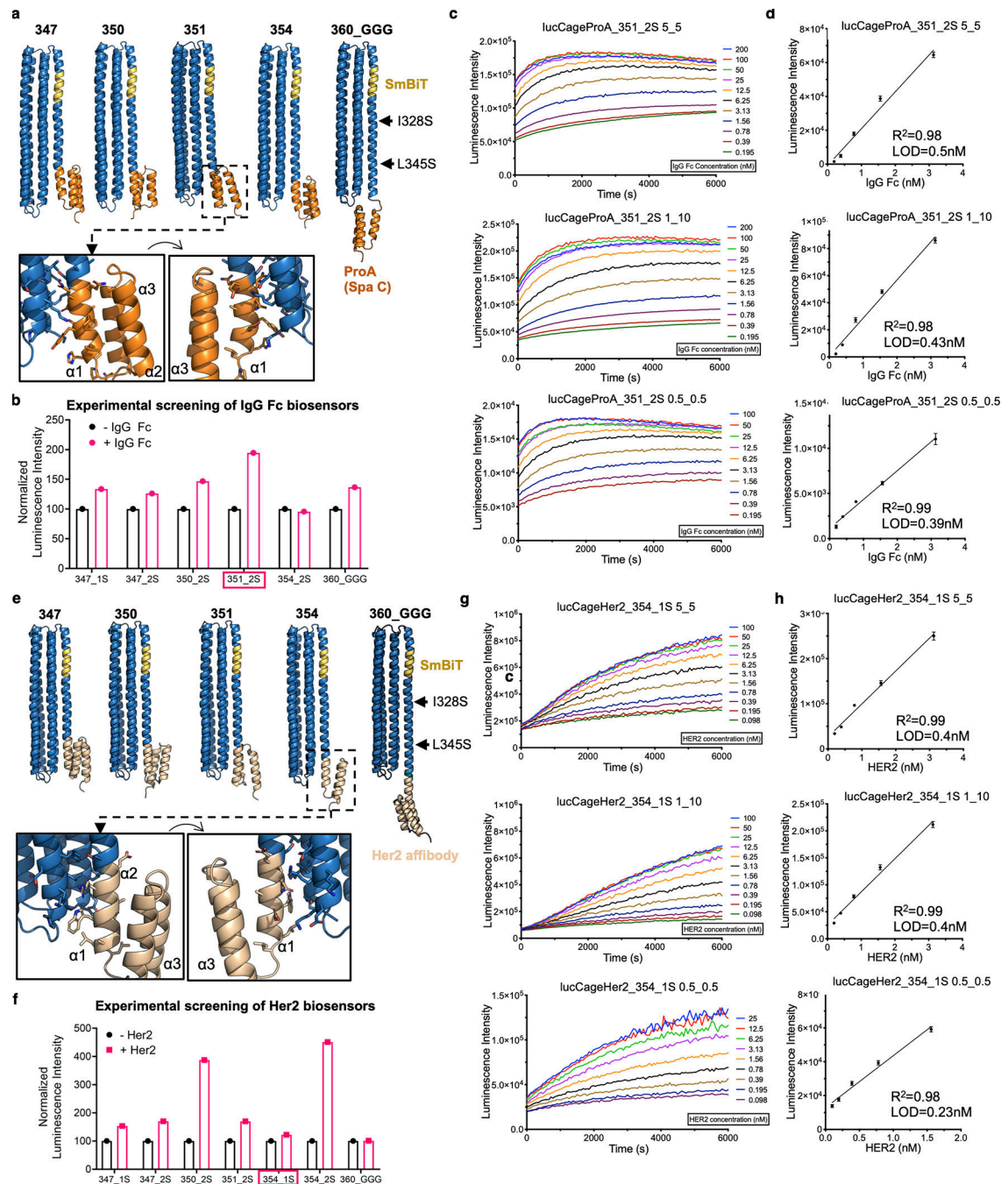
Extended Data Figure 5. Design and characterization of an Fc domain sensor (lucCageProA) and a Her2 sensor (lucCageHer2).

a, Structural models of the Fc sensor designs showing the different threading positions of the *S. aureus* Protein A domain C (orange, PDB ID: 4WWI) on the latch of lucCage (blue). The SmBit peptide is shown in gold ribbon representation. I328S and L345S indicate mutations introduced to tune the latch-cage interface, (1S=I328S, 2S=I328S/L345S) [2], and "GGG" indicates the presence of three consecutive glycine residues between the latch and the grafted protein. **b**, Experimental screening of 6 *de novo* Fc domain sensors. Luminescence

measurements were performed for each design (20 nM) and lucKey (20 nM) in the presence or absence of recombinant human IgG1 Fc (200 nM). The luminescence values were normalized to 100 in the absence of Fc. Design 351_2S was selected as the best candidate due to high sensitivity and stability, and was named lucCageProA. This experiment was performed using single replicates in two independent instances, representative data are shown. **c**, Determination of lucCageProA's sensitivity. Bioluminescence was measured over 6000 s in the presence of serially diluted Fc protein. From top to bottom - lucCageBot:lucKey concentration (nM) = 5:5, 1:10, 0.5:0.5. **d**, LOD calculations for the sensor at different concentrations. From top to bottom - lucCageBot:lucKey concentration (nM) = 5:5, 1:10, 0.5:0.5. **e**, Structural models of the Her2 sensor designs showing the different threading positions of the Her2 affibody protein (PDB ID: 3MZW, beige) on the latch of lucCage (blue). The SmBiT peptide is shown in gold ribbon representation. I328S and L345S indicate mutations introduced to tune the latch-cage interface, (1S=I328S, 2S=I328S/L345S) [2], and "GGG'' indicates the presence of three consecutive glycine residues between the latch and the grafted protein. The black boxes show a close-up view of the interface of Cage (blue) and the Her2 affibody (beige) in the 354_2S design. **f**, Experimental screening of 7 *de novo* Her2 sensors. Luminescence measurements were taken for each design (20 nM) and lucKey (20 nM) in the presence or absence of the ectodomain of Her2 (200 nM). The luminescence values were normalized to 100 in the absence of Her2 ectodomain. This experiment was performed using single replicates in two independent instances, representative data are shown. Design 354_2S was selected as the best candidate due to high sensitivity and stability, and was named lucCageHer2. **g**, Determination of lucCagerHer2's sensitivity. Bioluminescence was measured over 6000 s in the presence of serially diluted Her2 ectodomain protein. From top to bottom - lucCageBot:lucKey concentration (nM) = 5:5, 1:10, 0.5:0.5. **h**, Limit of detection (LOD) calculations for the sensor at different concentrations. From top to bottom - lucCageBot:lucKey concentration (nM) = 5:5, 1:10, 0.5:0.5. All experiments were performed in triplicate unless specifically indicated, representative data are shown, and data are presented as mean values +/− s.d.
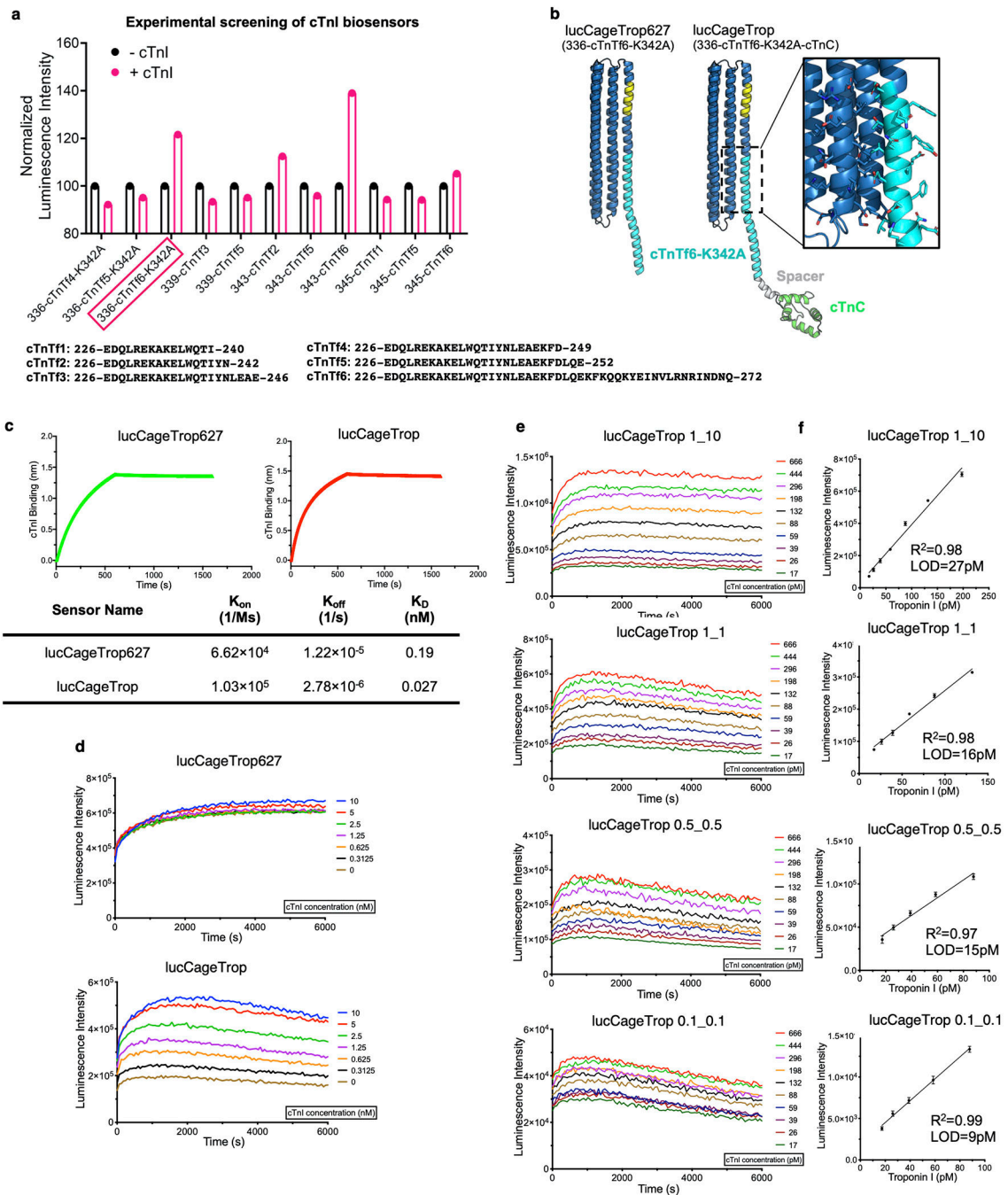
Extended Data Figure 6. Design, selection, and engineering of lucCageTrop for cardiac Troponin I detection.

a, Experimental screening of designed sensors for cardiac Troponin I (cTnI). Fragments of cardiac Troponin T, namely cTnTf1-f6, were computationally grafted into lucCage at different positions of the latch. All designs were produced in *E. coli* and experimentally screened at 20 nM and 20 nM lucKey for an increase in luminescence in the presence of cTnI (100 nM). The luminescence values were normalized to 100 in the absence of cTnI. This experiment was performed using single replicates in two independent instances,

representative data are shown. Design 336-cTnTf6-K342A was selected as the best candidate (named lucCageTrop627) based on its sensitivity, activation fold-change, and stability. **b**, Models of lucCageTrop627 and lucCageTrop, an improved version by fusion of cardiac Troponin C (cTnC) at the C-terminus of lucCageTrop627. The models are shown in ribbon representation comprising SmBit (gold) a fragment of cTnT (cyan, PDB ID: 4Y99), and cTnC (green, PDB ID: 4Y99). The black box shows a close-up view of the interface of Cage (blue) and cTnT (cyan) in the lucCageTrop design. **c**, The binding affinity of lucCageTrop627 and lucCageTrop to cTnI was measured by biolayer interferometry. lucCageTrop showed 7-fold higher affinity to cTnI than lucCageTrop627. **d**, Comparison of bioluminescence kinetics between lucCageTrop627 (top) and lucCageTrop (bottom) in the presence of serially diluted cTnI. Higher binding affinity leads to improved dynamic range and sensitivity of the sensor. **e**, Determination of lucCageTrop's sensitivity. Bioluminescence was measured over 6000 s in the presence of serially diluted cTnI. From top to bottom - lucCageTrop:lucKey concentration (nM) = 1:10, 1:1, 0.5:0.5, 0.1:0.1. **f**, LOD calculations for the sensor at different concentrations. From top to bottom - lucCageTrop:lucKey concentration (nM) = 1:10, 1:1, 0.5:0.5, 0.1:0.1. All experiments were performed in triplicate unless otherwise indicated, representative data are shown, and data are presented as mean values +/− s.d.
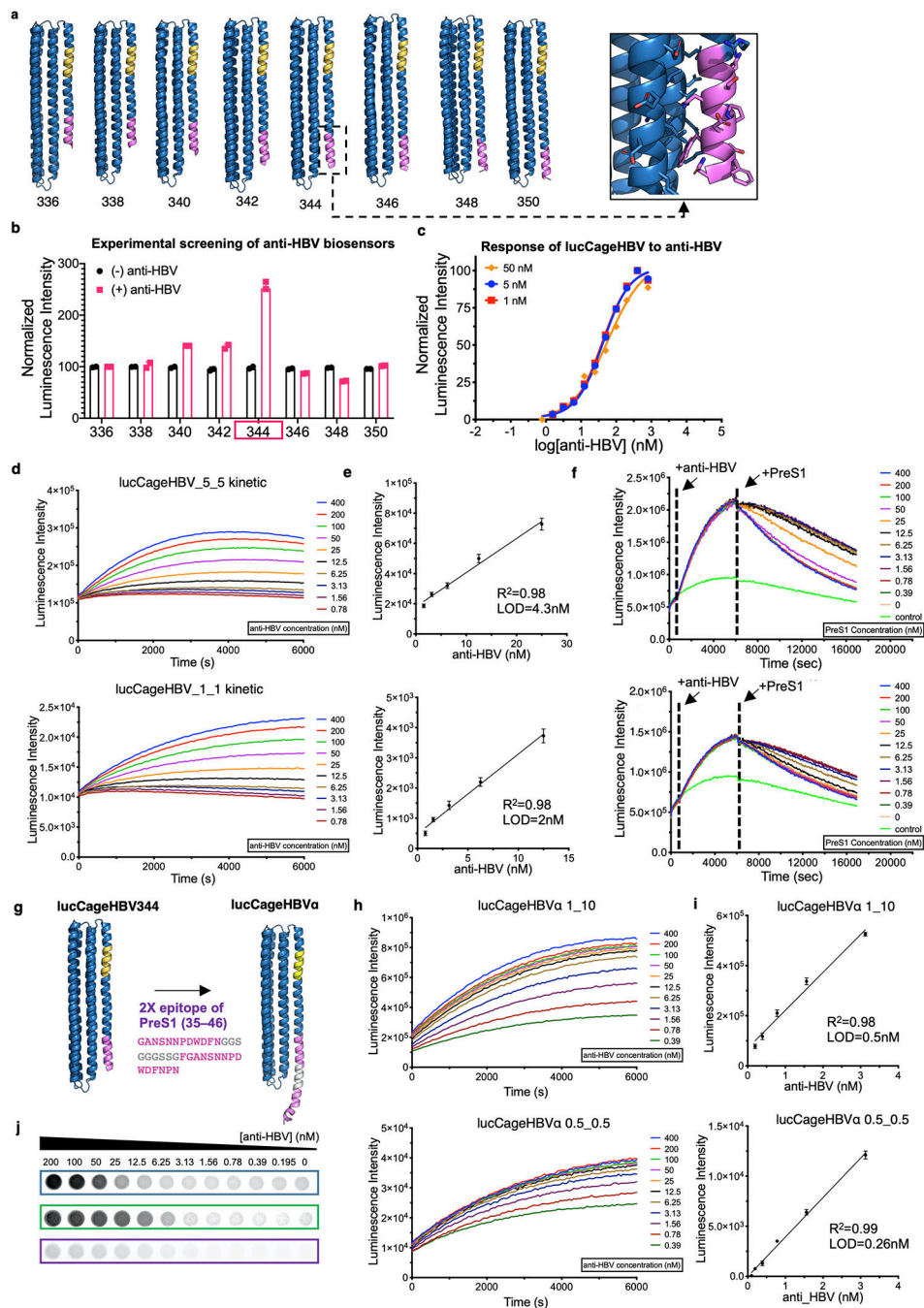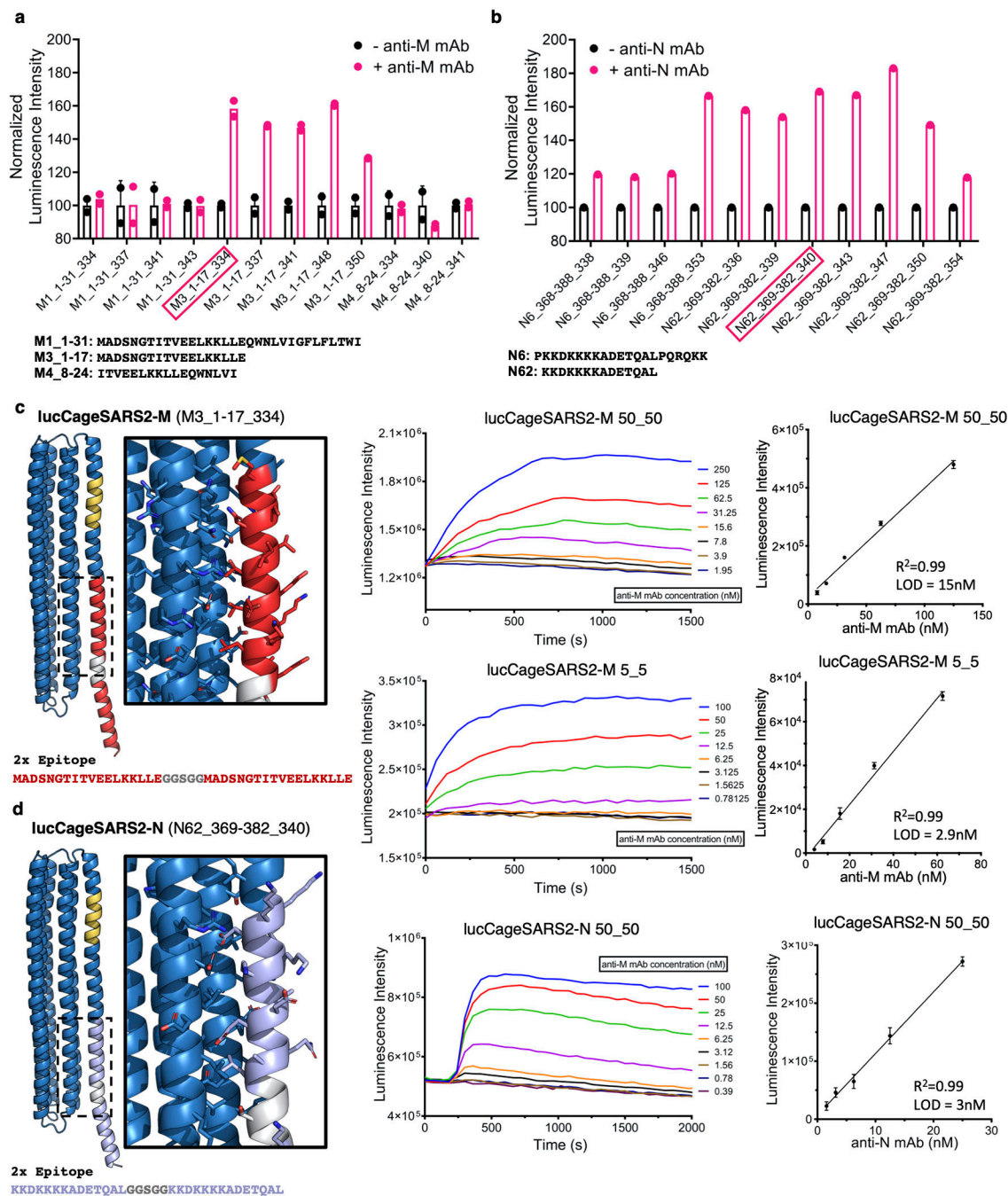
Extended Data Figure 7. Design and characterization of an anti-HBV antibody sensor.
a, The energy-minimized models of lucCage designs are shown with the threaded segments of SmBiT (gold) and the antigenic motif of PreS1 (magenta). The right box shows a close-up view of the cage-motif interface of the HBV344 design. **b**, Experimental screening of all designs performed by monitoring the luminescence of each lucCage (20 nM) and lucKey (20 nM) in the presence or absence of the anti-HBV antibody HzKR127–3.2 (100 nM). The luminescence values were normalized to 100 in the absence of anti-HBV. This experiment was performed in duplicate in two independent instances, representative data are shown. The

design HBV344 was selected due to its better performance and was named lucCageHBV. **c,d,** Determination of lucCageHBV sensitivity. Bioluminescence was measured over 6000 s in the presence of serially diluted HzKR127–3.2. From top to bottom - lucCageHBV:lucKey concentration (nM) = 5:5, 1:1. The maximum values of the curves in **d**, are used to obtain the curves in **c**. **e**, LOD calculations for the sensor at different concentrations. From top to bottom - lucCageHBV:lucKey concentration (nM) = 5:5, 1:1. **f**, Detection of PreS1 by competition of lucCageHBV344 and HzKR127–3.2 shown in Fig. 3f. Luminescence kinetics after the addition of the antibody (anti-HBV, first arrow). From top to bottom - anti-HBV antibody concentrations = 50, 12.5 nM. At 6000 s, different concentrations of the PreS1 domain were injected into the wells, and the decreased luminescence signals were used to detect PreS1. **g,** Design of lucCageHBVα for improved detection of an anti-HBV antibody. The structural model of lucCageHBVα is shown with a close-up detail of the predicted interface between the PreS1 epitope (magenta) and lucCage (blue). The design comprises two copies of the epitope PreS1 (a.a. 35–46), spaced by a flexible linker (grey) to enable bivalent interaction with the antibody. The SmBit peptide is shown in gold. **h**, Determination of lucCageHBVα detection sensitivity to the presence of the antibody HzKR127–3.2 (anti-HBV). Bioluminescence was measured over 6000 s in the presence of serially diluted HzKR127–3.2. From top to bottom - lucCageHBVα:lucKey concentration (nM) = 1:10, 0.5:0.5. **i,** The linear region of a calibration curve was used to determine the LOD and the dynamic range of antibody detection. **j**, Bioluminescence images acquired with a BioRad ChemiDoc imaging system. From top to bottom, lucCageHBVα:lucKey concentration (nM) = 50:5, 5:5, 1:10. Changes in bioluminescence intensity levels were detected as a function of the concentration of HzKR127–3.2. All experiments were performed in triplicate unless specifically indicated, and representative data are presented as mean values +/− s.d.

Extended Data Figure 8. Design and characterization of sensors for anti-SARS-CoV-2 antibodies.

a-b, Experimental screening of *de novo* sensors for antibodies against the SARS-CoV-2 membrane protein (**a**), and the nucleocapsid protein (**b**). Selected epitopes of the membrane protein (M1, M3 and M4) and the nucleocapsid protein (N6 and N62) were computationally grafted into lucCage at different positions of the latch. Each design comprised two tandem copies of each epitope, separated by a flexible linker, to take advantage of the bivalent binding of antibodies. All designs were experimentally screened for increase in

luminescence at 20nM of each lucCage design and 20nM of lucKey in the presence of anti-M rabbit polyclonal antibodies (ProSci, 3527) (**a**) or anti-N mouse monoclonal antibody at 100nM (clone 18F629.1) (**b**). These experiments were performed in duplicate (**a**) or single replicate (**b**) in two independent instances, representative data are shown. The luminescence values were normalized to 100 in the absence of antibodies. Designs M3_1–17_334 and N62_369–382_340 were selected as the best candidates due to high sensitivity and stability, and were named lucCageSARS2-M and ucCageSARS2-N respectively. **c**, Left panel: structural model of lucCageSARS2-M, showing a close-up view of the predicted interface between the M3 epitope (red) and lucCage (blue). Middle panel: determination of lucCageSARS2-M sensitivity to anti-M pAb. Bioluminescence was measured over 4000 s in the presence of serially diluted anti-M pAb. From top to bottom - lucCageSARS2-M:lucKey concentration (nM) = 50:50, 5:5. Right panel: LOD calculations for the sensor at different concentrations. **d**, Left panel: structural model of lucCageSARS2-N, showing a close-up view of the predicted interface between the N62 epitope (purple) and lucCage (blue). Middle panel: determination of lucCageSARS2-N sensitivity to anti-N mAb. Bioluminescence was measured over 4000 s for lucCageSARS2-N + lucKey at 50 nM in the presence of serially diluted anti-N antibody. Right panel: LOD calculations for the sensor. All experiments were performed in triplicate unless specifically indicated, representative data are shown, and data are presented as mean values +/− s.d.
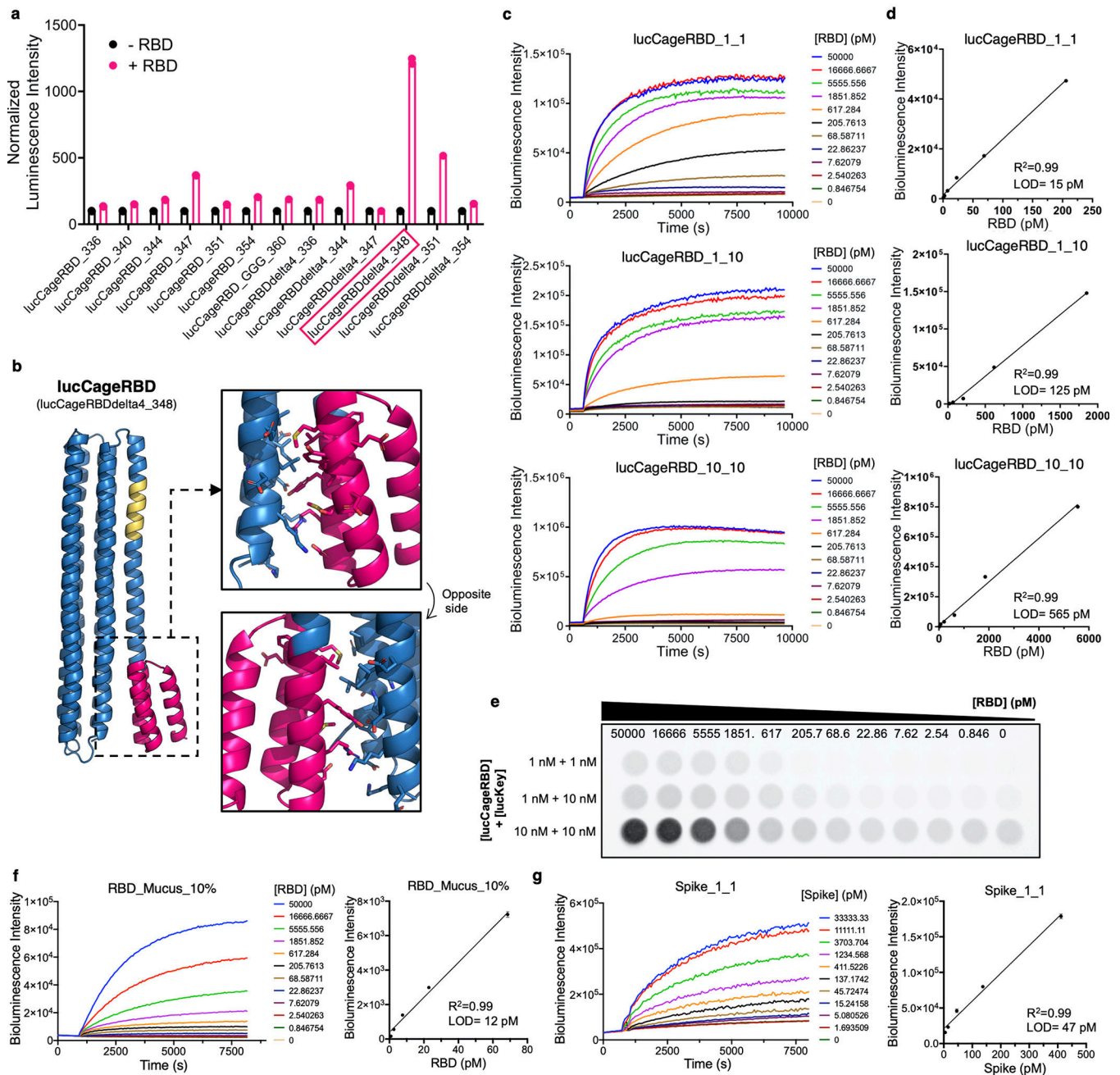
**Extended Data Figure 9. Design and characterization of SARS-CoV-2 RBD sensors.**
a, Experimental screening of *de novo* sensors for the receptor-binding domain (RBD) of the
SARS-CoV-2 Spike protein. All designs were experimentally screened for increase in
luminescence at 20 nM of each lucCage design and 20 nM of lucKey in the presence of 200
nM RBD. The luminescence values were normalized to 100 in the absence of RBD. This
experiment was performed in duplicate in two independent instances, representative data are
shown. Design lucCageRBDdelta4_348 was selected as the best candidate due to high
sensitivity and stability, and was named lucCageRBD. **b**, Structural model of lucCageRBD
composed of the LCB1 binder (magenta) grafted into lucCage (blue) comprising a caged

SmBiT fragment (gold). The black boxes show a close-up view of the interface of Cage (blue) and LCB1 binder (magenta) in the lucCageRBD design. **c**, Determination of lucCagerRBD's sensitivity. Bioluminescence was measured over 10000 s in the presence of serially diluted RBD protein. From top to bottom - lucCageRBD:lucKey concentration (nM) = 1:1, 1:10, 10:10. **d**, LOD calculations for the sensor at different concentrations. From top to bottom - lucCageRBD:lucKey concentration (nM) = 1:1, 1:10, 10:10. **e**, Bioluminescence images acquired with a BioRad ChemiDoc imaging system. Changes in bioluminescence intensity levels were detected as a function of the concentration of RBD with lucCageRBD at 1 nM and lucKey at 10 nM. **f**, Detection of RBD in 10% simulated nasal matrix. Left: Bioluminescence was measured overtime in the presence of serially diluted RBD protein. Right: LOD was calculated to be 12 pM. **g**, Detection of spike protein in a 20% diluted pooled serum. Left: Bioluminescence was measured overtime in the presence of serially diluted HexaPro spike protein. Right: LOD was calculated to be 47 pM. All experiments were performed in triplicate unless otherwise indicated, representative data are shown, and data are presented as mean values +/− s.d.
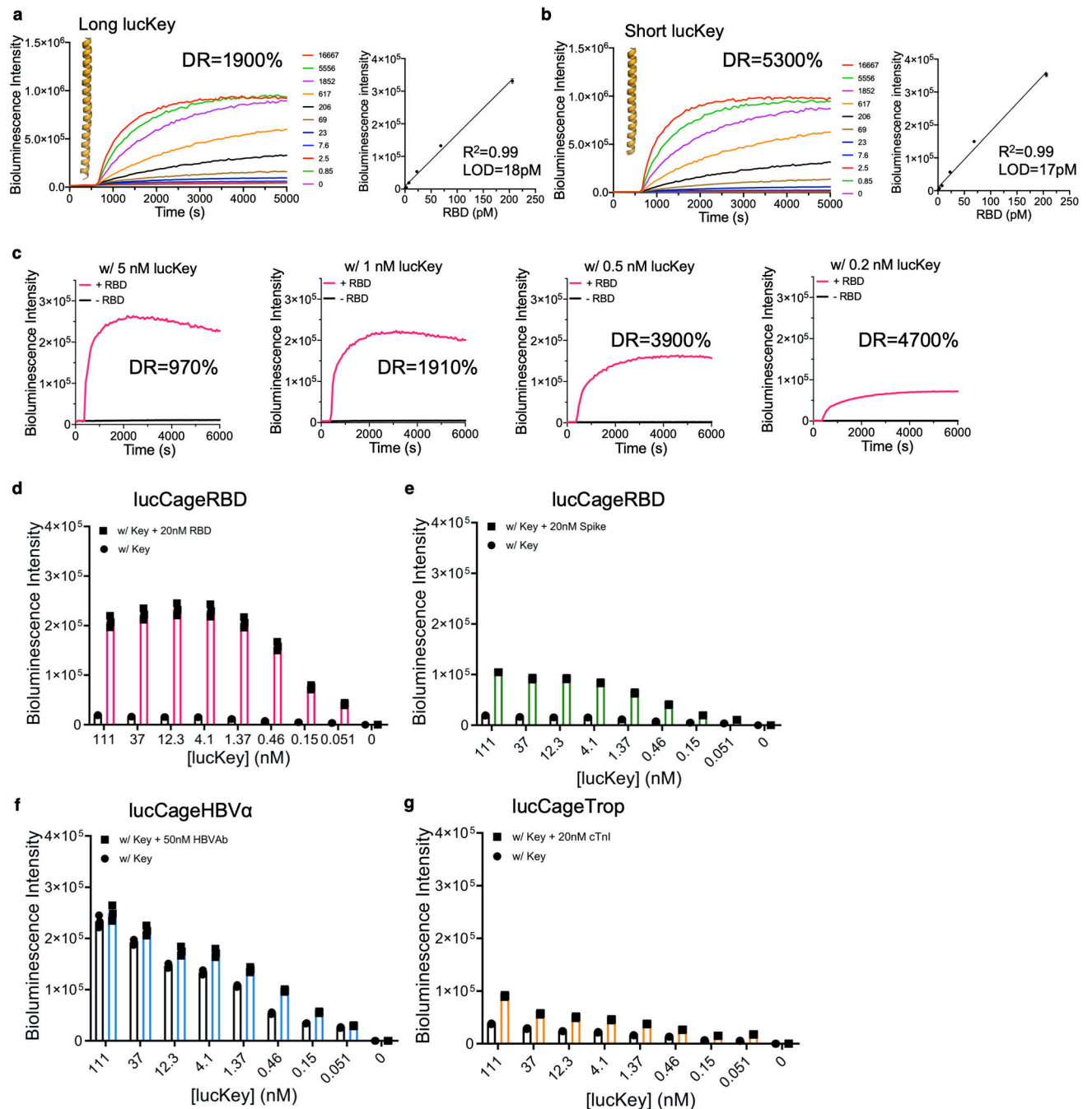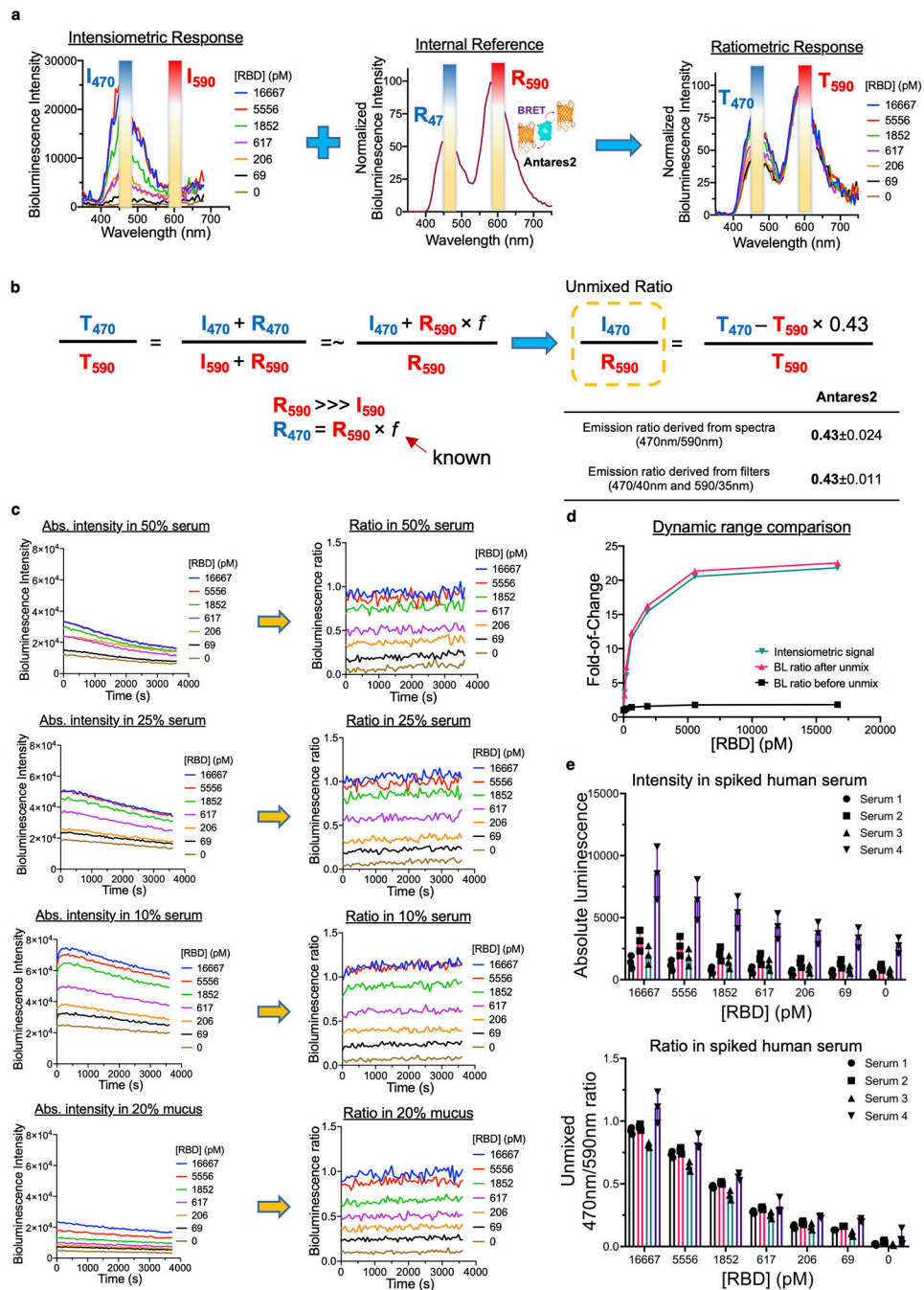
Extended Data Figure 10. lucCageRBD tunability by varying the lucKey length (K_{CK}) and lucKey concentration (a-c) and the comparison of bioluminescent signals over a range of lucKey concentrations in the presence of target at saturating concentration (d-g).

a-b, Experimental evaluation of the effect of K_{CK} on the dynamic range (DR) of lucCageRBD to detect monomeric SARS-CoV-2 RBD. A truncated lucKey (short lucKey), 14 residue shorter than the full-length key at its C-terminus (**b**), provides better dynamic range than the full-length lucKey (**a**) owing to reduced background signal, as predicted by the simulation in Extended Data Fig. 1f while the LOD remains the same. **c**, The effect of lucKey concentration on the dynamic range. Decreasing lucKey concentration increases the

dynamic range of lucCageRBD due to reduced background signal, but with accompanying reduced maximum bioluminescence signal. **d-e,** lucCageRBD (1 nM) was incubated with RBD (20 nM, **d**) or spike protein (20 nM, **e**), which are expected to result in full reconstitution of the luciferase activity. In the presence of spike protein, the same sensor was unable to yield the maximal bioluminescent signal, suggesting the effect of factors not captured by the simulations such as steric hindrance against complete luciferase reconstitution. **f**, lucCageHBVα (1 nM) incubated with 50 nM of the HBV antibody HzKR127–3.2 shows almost complete activation, but suffers from high background signal. **g**, lucCageTrop (1 nM) shows non-ideal background signal and moderate target-driven activation in the presence of 20 nM cTnI. All experiments were performed in triplicate, representative data are shown, and data are presented as mean values +/− s.d.

Extended Data Figure 11. Integration of Antares2 as the internal reference for calibration of lucCageRBD in different biological matrices.

a, The bioluminescent emission spectra of lucCageRBD (Left) in response to varying concentrations of RBD. Antares2 is an efficient CyOFP1-teLuc-CyOFP1 BRET system[40] with a peak emission at 590 nm (Middle). The emission spectra were recorded from a mixture of lucCageRBD and lucKey (both at 1 nM), Antares2 (0.1 nM) and RBD at varying concentrations (Right). By acquiring the individual signal from 470/40 nm and 590/35 nm channels, the intensiometric responses from lucCageRBD were converted into ratiometric

readouts. **b,** Equations to calculate the spectrally unmixed ratio. The total signal from the 470/40 nm channel (T_{470}) is the sum of the signals from the lucCageRBD sensor (I_{470}) and the Antares2 reference (R_{470}), while the total signal from the 590/35 nm channel (T_{590}) is equal to the sensor signal (I_{590}) plus reference signal (R_{590}). Since lucCageRBD gives negligible emission at 590/35 nm channel, T_{590} is approximately equal to R_{590} ($R_{590} \gg I_{590}$). R_{470} is $R_{590} \times f$, a predetermined constant for Antares2, and therefore the unmixed ratio (I_{470}/R_{590}) could be calculated in real time during signal acquisition. The constant f for Antares2 was consistently determined to be 0.43 by either recording the full spectra or from the filter set. **c,** RBD at varying concentrations were spiked in 50%, 25%, 10% pooled serum or in 20% simulated nasal fluid. Absolute bioluminescence intensities and the emission kinetics were different across the matrices due to matrix inhibition effect and substrate turnover[52]. In contrast, calibration with Antares2 resulted in stable ratiometric signals (I_{470}/R_{590}). **d,** The bioluminescence intensity of lucCageRBD at saturating RBD concentration (green curve) is ~20 folds higher than the background level. Reporting the raw ratio (T_{470}/T_{590}) as a function of the RBD concentration compromises the sensor dynamic range (black curve) due to a significant emission at 470/40 nm channel (R_{470}) from Antares2. After calculation and conversion of the unmixed ratio, the dynamic range becomes ~20 folds over the background with ratiometric readouts (magenta curve). **e,** Detection of spiked RBD in four different anonymized human sera (50%) shows that calibration using spectrally resolved Antares2 as an internal reference can minimize the variations of the intensiometric bioluminescence in these matrices. Bioluminescent signals and s.d. were measured in triplicate, and a representative one is shown for emission spectra and emission kinetics, respectively. Data are presented as mean values +/− s.d.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

1. Stein V & Alexandrov K Synthetic protein switches: design principles and applications. Trends Biotechnol. 33, 101–110 (2015). [PubMed: 25535088]

2. Langan RA et al. De novo design of bioactive protein switches. Nature 572, 205–210 (2019). [PubMed: 31341284]

3. Udugama B et al. Diagnosing COVID-19: The Disease and Tools for Detection. ACS Nano vol. 14 3822–3835 (2020). [PubMed: 32223179]

4. Cao L et al. De novo design of picomolar SARS-CoV-2 miniprotein inhibitors. Science 370, 426–431 (2020). [PubMed: 32907861]

5. Yeh H-W & Ai H-W Development and Applications of Bioluminescent and Chemiluminescent Reporters and Biosensors. Annu. Rev. Anal. Chem 12, 129–150 (2019).

6. Greenwald EC, Mehta S & Zhang J Genetically Encoded Fluorescent Biosensors Illuminate the Spatiotemporal Regulation of Signaling Networks. Chem. Rev 118, 11707–11794 (2018). [PubMed: 30550275]

7. Banala S, Aper SJA, Schalk W & Merkx M Switchable reporter enzymes based on mutually exclusive domain interactions allow antibody detection directly in solution. ACS Chem. Biol. 8, 2127–2132 (2013). [PubMed: 23941162]

8. Arts R et al. Detection of Antibodies in Blood Plasma Using Bioluminescent Sensor Proteins and a Smartphone. Anal. Chem 88, 4525–4532 (2016). [PubMed: 27018236]

9. van Rosmalen M et al. Dual-Color Bioluminescent Sensor Proteins for Therapeutic Drug Monitoring of Antitumor Antibodies. Anal. Chem 90, 3592–3599 (2018). [PubMed: 29443503]

10. Yu Q et al. Semisynthetic sensor proteins enable metabolic assays at the point of care. Science 361, 1122–1126 (2018). [PubMed: 30213915]

11. Yu Q et al. A biosensor for measuring NAD levels at the point of care. Nat Metab 1, 1219–1225 (2019). [PubMed: 32694678]

12. Schena A, Griss R & Johnsson K Modulating protein activity using tethered ligands with mutually exclusive binding sites. Nat. Commun 6, 7830 (2015). [PubMed: 26198003]

13. Arts R et al. Semisynthetic Bioluminescent Sensor Proteins for Direct Detection of Antibodies and Small Molecules in Solution. ACS Sens 2, 1730–1736 (2017). [PubMed: 29037030]

14. Xue L, Prifti E & Johnsson K A General Strategy for the Semisynthesis of Ratiometric Fluorescent Sensor Proteins with Increased Dynamic Range. J. Am. Chem. Soc 138, 5258–5261 (2016). [PubMed: 27071001]

15. Guo Z et al. Generalizable Protein Biosensors Based on Synthetic Switch Modules. J. Am. Chem. Soc 141, 8128–8135 (2019). [PubMed: 31074995]

16. Edwardraja S et al. Caged activators of artificial allosteric protein biosensors. ACS Synth. Biol. 9, 1306–1314 (2020). [PubMed: 32339455]

17. Ribeiro LF, Warren TD & Ostermeier M Construction of Protein Switches by Domain Insertion and Directed Evolution. Methods Mol. Biol. 1596, 43–55 (2017). [PubMed: 28293879]

18. Dixon AS et al. NanoLuc Complementation Reporter Optimized for Accurate Measurement of Protein Interactions in Cells. ACS Chem. Biol. 11, 400–408 (2016). [PubMed: 26569370]

19. Minor DL Jr & Kim PS Context-dependent secondary structure formation of a designed protein sequence. Nature 380, 730–734 (1996). [PubMed: 8614471]

20. Kale J, Osterlund EJ & Andrews DW BCL-2 family proteins: changing partners in the dance towards death. Cell Death Differ. 25, 65–80 (2018). [PubMed: 29149100]

21. Chevalier A et al. Massively parallel de novo protein design for targeted therapeutics. Nature 550, 74–79 (2017). [PubMed: 28953867]

22. Lajoie MJ et al. Designed protein logic to target cells with precise combinations of surface antigens. Science 369, 1637–1643 (2020). [PubMed: 32820060]

23. Deis LN et al. Suppression of conformational heterogeneity at a protein-protein interface. Proc. Natl. Acad. Sci. U. S. A 112, 9028–9033 (2015). [PubMed: 26157136]

24. Eigenbrot C, Ultsch M, Dubnovitsky A, Abrahmsén L & Härd T Structural basis for high-affinity HER2 receptor binding by an engineered protein. Proc. Natl. Acad. Sci. U. S. A 107, 15039–15044 (2010). [PubMed: 20696930]

25. Hobbs RJ, Thomas CA, Halliwell J & Gwenin CD Rapid Detection of Botulinum Neurotoxins-A Review. Toxins 11, (2019).

26. Perrier A, Gligorov J, Lefèvre G & Boissan M The extracellular domain of Her2 in serum as a biomarker of breast cancer. Lab. Invest 98, 696–707 (2018). [PubMed: 29491426]

27. Rubini Gimenez M et al. One-hour rule-in and rule-out of acute myocardial infarction using high-sensitivity cardiac troponin I. Am. J. Med 128, 861–870.e4 (2015). [PubMed: 25840034]

28. Collins MH Serologic Tools and Strategies to Support Intervention Trials to Combat Zika Virus Infection and Disease. Trop Med Infect Dis 4, (2019).

29. Pondé RA de A. Expression and detection of anti-HBs antibodies after hepatitis B virus infection or vaccination in the context of protective immunity. Arch. Virol 164, 2645–2658 (2019). [PubMed: 31399876]

30. Chi S-W et al. Broadly neutralizing anti-hepatitis B virus antibody reveals a complementarity determining region H3 lid-opening mechanism. Proc. Natl. Acad. Sci. U. S. A 104, 9230–9235 (2007). [PubMed: 17517649]

31. Kim JH et al. Enhanced humanization and affinity maturation of neutralizing anti-hepatitis B virus preS1 antibody based on antigen-antibody complex structure. FEBS Lett. 589, 193–200 (2015). [PubMed: 25481411]

32. Ovacik M & Lin K Tutorial on Monoclonal Antibody Pharmacokinetics and Its Considerations in Early Development. Clin. Transl. Sci 11, 540–552 (2018). [PubMed: 29877608]

33. Locarnini S & Bowden S Hepatitis B surface antigen quantification: Not what it seems on the surface. Hepatology 56, 411–414 (2012). [PubMed: 22454331]

34. Chow SCS et al. Specific epitopes of the structural and hypothetical proteins elicit variable humoral responses in SARS patients. J. Clin. Pathol. 59, 468–476 (2006). [PubMed: 16461566]

35. He Y, Zhou Y, Siddiqui P, Niu J & Jiang S Identification of immunodominant epitopes on the membrane protein of the severe acute respiratory syndrome-associated coronavirus. J. Clin. Microbiol 43, 3718–3726 (2005). [PubMed: 16081901]

36. Wang H et al. SARS-CoV-2 proteome microarray for mapping COVID-19 antibody interactions at amino acid resolution. ACS Central Science 6, 2238–2249 (2020). [PubMed: 33372199]

37. Hsieh C-L et al. Structure-based design of prefusion-stabilized SARS-CoV-2 spikes. Science 369, 1501–1505 (2020). [PubMed: 32703906]

38. Panpradist N et al. Swab sample transfer for point-of-care diagnostics: characterization of swab types and manual agitation methods. PLoS One 9, e105786 (2014). [PubMed: 25181250]

39. Ni Y et al. RAPPID: a platform of ratiometric bioluminescent sensors for homogeneous immunoassays. doi:10.1101/2020.10.31.363044.

40. Yeh H-W et al. Red-shifted luciferase-luciferin pairs for enhanced bioluminescence imaging. Nat. Methods 14, 971–974 (2017). [PubMed: 28869756]

41. Yang C et al. Bottom-up de novo design of functional proteins with complex structural features. Nat. Chem. Biol (2021) doi:10.1038/s41589-020-00699-x.

Methods References

42. Fleishman SJ et al. RosettaScripts: a scripting language interface to the Rosetta macromolecular modeling suite. PLoS One 6, e20161 (2011). [PubMed: 21731610]

43. Huang P-S et al. RosettaRemodel: a generalized framework for flexible backbone protein design. PLoS One 6, e24109 (2011). [PubMed: 21909381]

44. Khatib F et al. Algorithm discovery by protein folding game players. Proc. Natl. Acad. Sci. U. S. A 108, 18949–18953 (2011). [PubMed: 22065763]

45. Walls AC et al. Elicitation of potent neutralizing antibody responses by designed protein nanoparticle vaccines for SARS-CoV-2. Cell 183, P1367–1382.E17 (2020).

46. Berger S et al. Computationally designed high specificity inhibitors delineate the roles of BCL2 family proteins in cancer. Elife 5, (2016).

47. Jin R, Rummel A, Binz T & Brunger AT Botulinum neurotoxin B recognizes its protein receptor with high affinity and specificity. Nature 444, 1092–1095 (2006). [PubMed: 17167421]

48. Shen A et al. Mechanistic and structural insights into the proteolytic activation of Vibrio cholerae MARTX toxin. Nat. Chem. Biol 5, 469–478 (2009). [PubMed: 19465933]

49. Otwinowski Z & Minor W [20] Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol. 276, 307–326 (1997).

50. Liebschner D et al. Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. Acta Crystallogr D Struct Biol 75, 861–877 (2019). [PubMed: 31588918]

51. Potterton L et al. Developments in the CCP4 molecular-graphics project. Acta Crystallogr. D Biol. Crystallogr. 60, 2288–2294 (2004). [PubMed: 15572783]

Extended Data figure legends-only references

52. Yeh H-W et al. ATP-Independent Bioluminescent Reporter Variants To Improve in Vivo Imaging. ACS Chemical Biology 14, 959–965 (2019). [PubMed: 30969754]
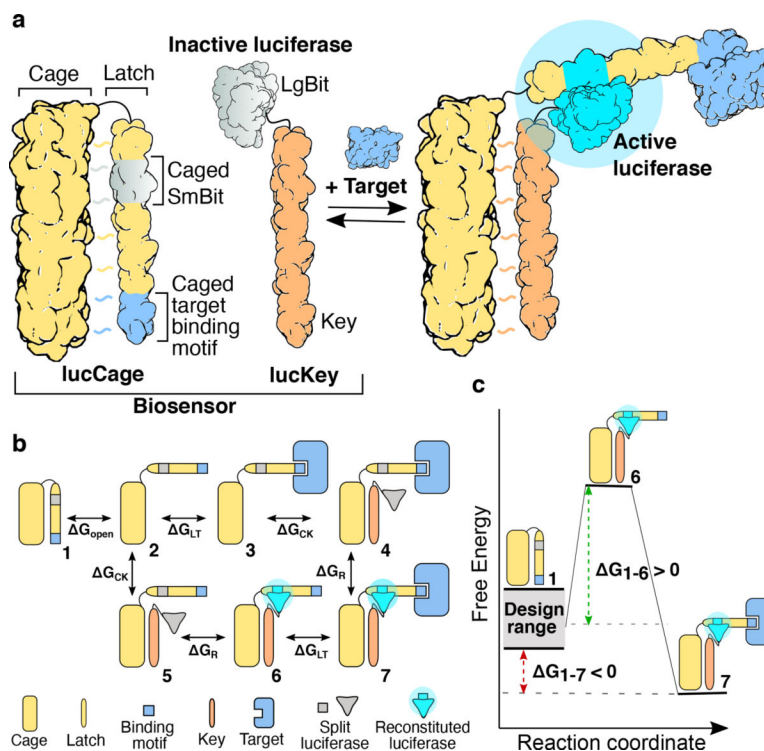
Fig. 1. *De novo* design of multi-state biosensors.
a, Sensor schematic mechanism. The closed form of lucCage (left) can not bind to lucKey, thus preventing the split luciferase SmBiT fragment from interacting with LgBit. The open form (right) can bind both target and key, allowing the reconstitution of SmBiT and LgBiT for luciferase activity. **b,** Thermodynamics of biosensor activation. The free energy cost ΔG_{open} of the transition from closed cage (species 1) to open cage (species 2) disfavors association of key (species 5) and reconstitution of luciferase activity (species 6) in the absence of target. In the presence of the target, the combined free energies of target binding (2→3; ΔG_{LT}), key binding (3→4; ΔG_{CK}), and SmBiT-LgBiT association (4→7; ΔG_R) overcome the unfavorable ΔG_{open}, driving opening of the lucCage and reconstitution of luciferase activity. **c**, Thermodynamics of biosensor design. The designable parameters are ΔG_{open} and ΔG_{CK}; ΔG_R is the same for all targets, and ΔG_{LT} is pre-specified for each target. For sensitive but low background analyte detection, ΔG_{open} and ΔG_{CK} must be tuned such that the closed state (species 1) is substantially lower in free energy than the open state (species 6) in the absence of target, but higher in free energy than the open state in the presence of target (species 7).
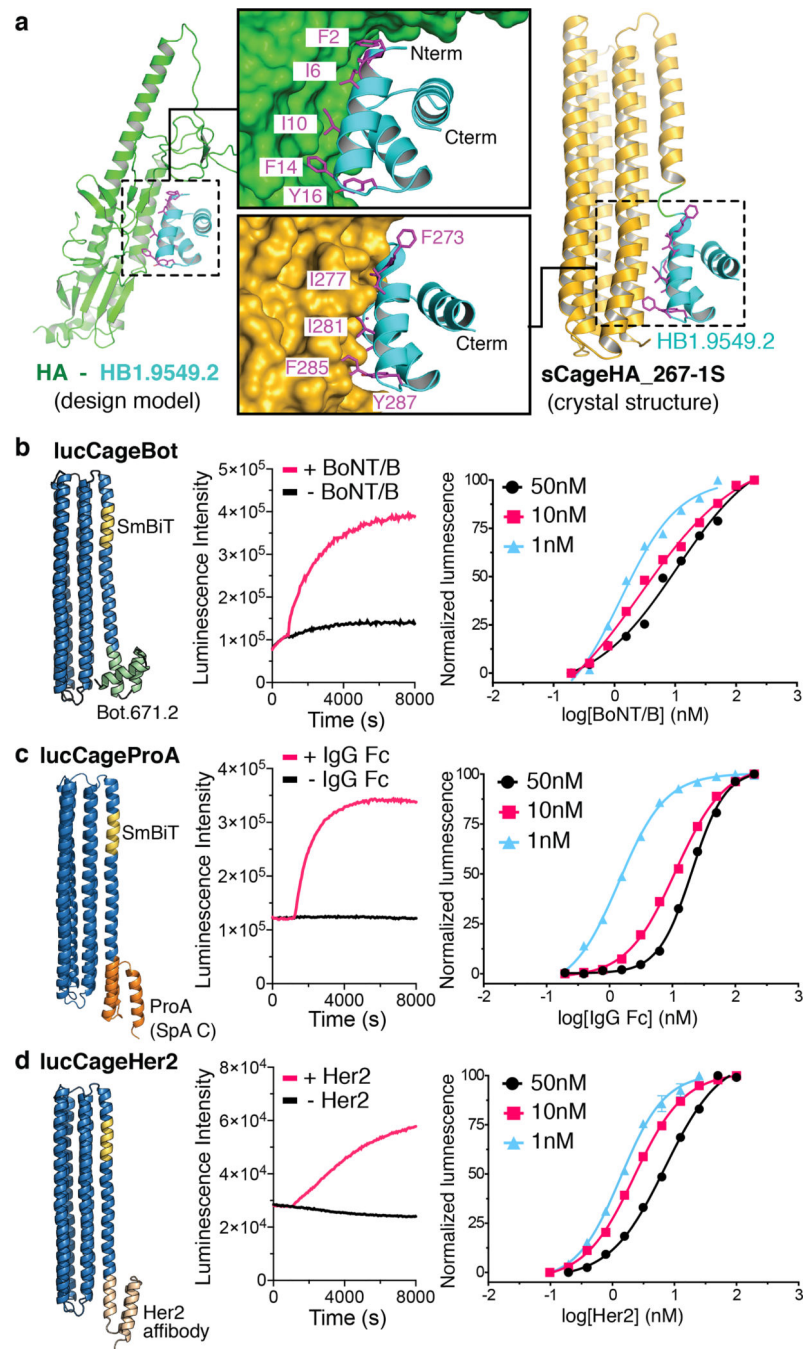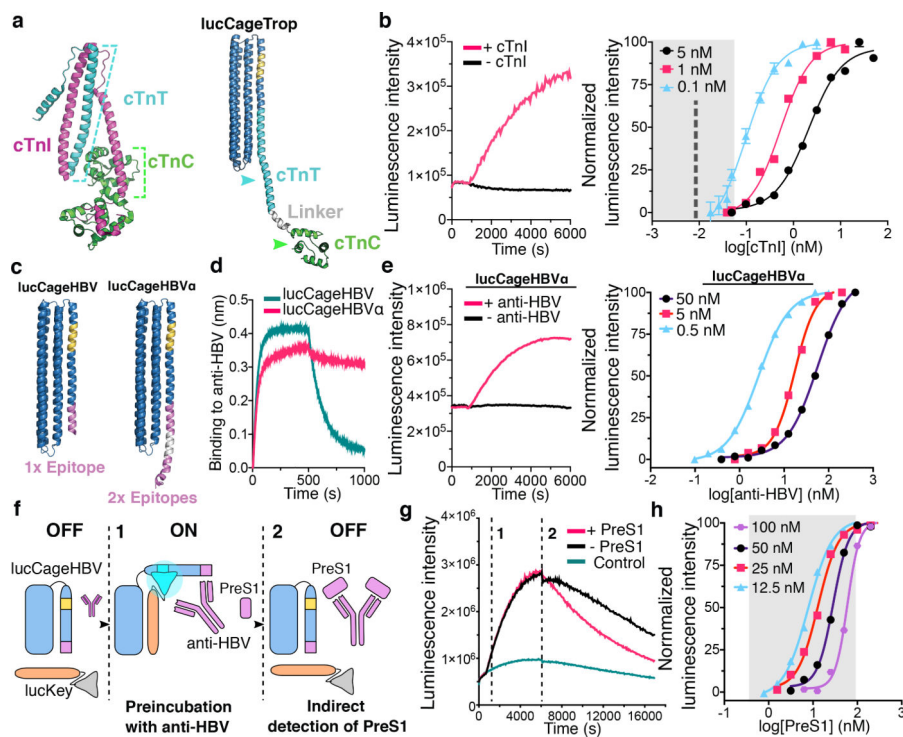
Fig. 2. Design and characterization of *de novo* biosensors incorporating small proteins as sensing domains.

a, Structural validation of sCageHA_267–1S, caging small protein domains into a LOCKR switch. Left: design model of the *de novo* binder HB1.9549.2 (cyan ribbon) bound to the stem region of influenza hemagglutinin (HA, green ribbon)[21]. Right: crystal structure (PDB ID: 7CBC) of sCageHA_267_1S, comprising HB1.9549.2 (cyan) grafted into a shortened and stabilized version of the LOCKR switch[22] (sCage, yellow ribbon). Middle: All residues of HB1.9549.2 involved in binding to HA (magenta, top) except for F273 are buried in the

closed state of the switch (bottom). The labels in magenta indicate the same set of amino acids in the two panels (e.g., F2 in the top panel corresponds to F273 in the lower panel). **b-d**, Functional characterization of lucCageBot, lucCageProA, and lucCageHer2. Left: structural models incorporate a *de novo* designed binder for BoNT/B (Bot.671.2) [21], the C domain of Protein A (SpaC)[23] or a Her2-binding affibody[24], respectively, into lucCage (blue ribbon) with caged SmBiT fragment (gold ribbon). Middle: Measurement of luminescence intensity upon addition of 50 nM of analyte (BoNT/B, IgG Fc, or Her2) to a mixture of 10 nM of each lucCage and 10 nM of lucKey. Right: detection over a wide range of analyte concentrations by changing the biosensor (lucCage + lucKey) concentration (colored lines). All experiments were performed in triplicate, representative data are shown, and data are presented as mean values +/− s.d.

Fig. 3. Design and characterization of biosensors for cardiac troponin I and an anti-HBV antibody.

a, Design of cardiac Troponin I sensor. Left: Structure of cardiac troponin (PDB ID: 4Y99); Troponin T, C and I (cTnT, cTnC and cTnI) are shown in cyan, green, and magenta, respectively. Right: Design model of lucCageTrop. **b**, Left: Kinetics of luminescence increase upon addition of 1 nM cTnI to 0.1 nM lucCageTrop+lucKey. Right: Wide detection range accessible by changing the concentration of the sensor components (colored lines). Grey area covers the cTnI concentration range relevant to the diagnosis of acute myocardial infarction (AMI) [27]; the dotted line indicates clinical AMI cut-off defined by W.H.O. (0.6 ng/mL, 25 pM). **c**, HBV sensor design models (gold, SmBiT; grey, linker; magenta, HBV PreS1 epitope). **d**, lucCageHBVα with two epitope copies shows higher affinity by biolayer interferometry for the anti-HBV antibody HzKR127–3.2 (Kd=0.68 nM) than lucCageHBV (Kd=20 nM). **e**, Left: Kinetics of bioluminescence increase upon addition of 50 nM anti-HBV antibody to 1 nM lucCageHBVα+lucKey. Right: Sensitive anti-HBV antibody detection over a wide concentration range. **f**, Mechanism for PreS1 detection using lucCageHBV. **g**, Kinetics of bioluminescence following addition of the anti-HBV antibody ("1") and subsequently PreS1 ("2"), which decreases bioluminescence by competing with the sensor for the antibody. **h**, Detection of PreS1 can be achieved over the relevant post-HBV infection concentration levels (grey area) by varying the concentration of antibody (indicated by colored labels). All experiments were performed in triplicate, representative data are shown, and data are presented as mean values +/− s.d.

Page 41
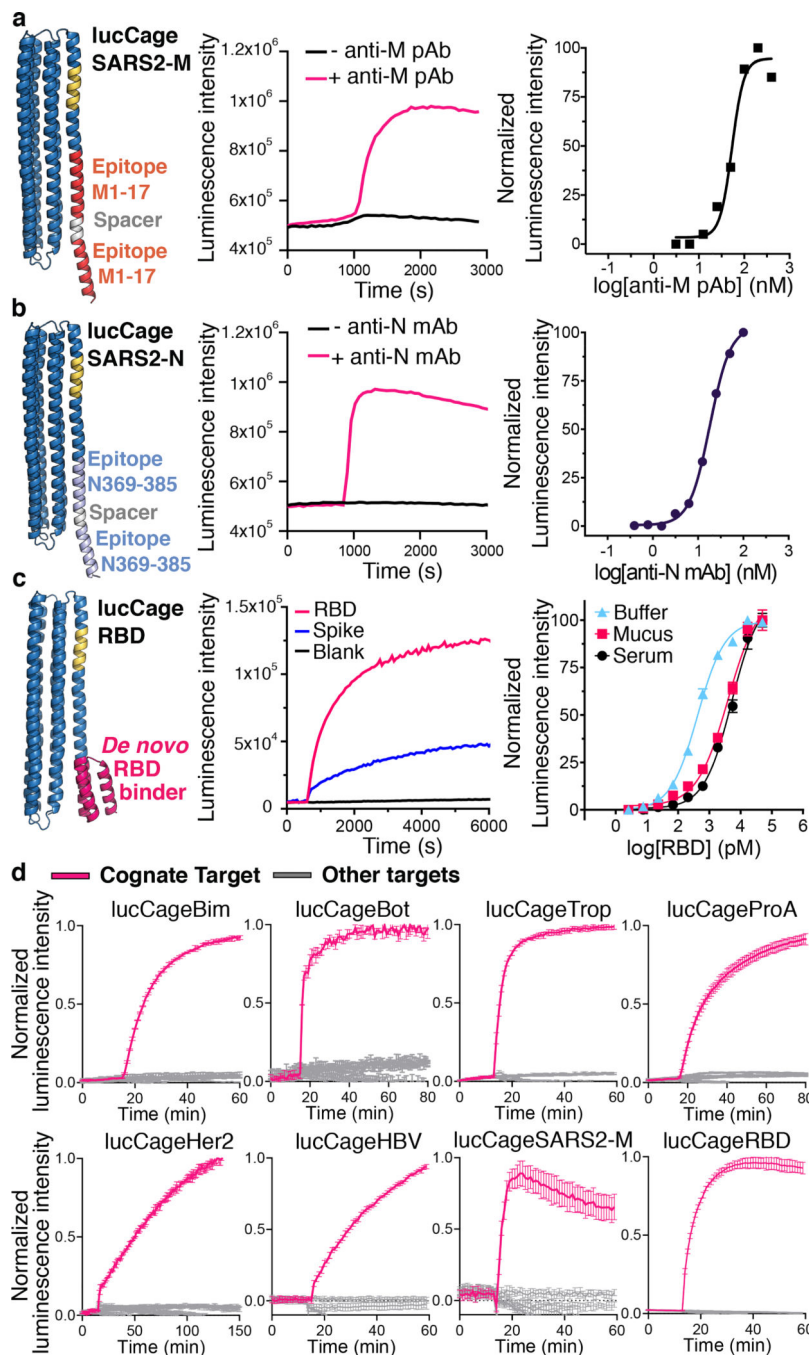
Fig. 4. Design of highly specific biosensors for the detection of anti-SARS-CoV-2 antibodies and SARS-CoV-2 viral proteins.

a, Left panel: lucCageSARS2-M sensor incorporates two copies of the SARS-CoV-2 membrane protein 1–17 epitope (red) connected with a flexible spacer. Middle panel: kinetics of luminescence activation of 50 nM lucCageSARS2-M+lucKey upon addition of 100 nM anti-SARS-CoV-1-M rabbit polyclonal antibodies (ProSci, 3527) that cross-react with residues 1–17 of the SARS-CoV-2-M. Right panel: response of 5 nM lucCageSARS2-M+lucKey to varying concentrations of target anti-M pAb. **b,** Left panel: lucCageSARS2-N

incorporates two copies of the SARS-CoV-2 nucleocapsid protein 369–382 epitope (light blue). Middle panel: kinetics of luminescence activation of 50 nM lucCageSARS2-N +lucKey upon addition of 100 nM anti-SARS-CoV-1-N mouse monoclonal antibody (clone 18F629.1), that recognizes the epitope. Right panel: response of 50 nM lucCageSARS2-N +lucKey to varying concentration of anti-N mAb. **c,** Left panel: lucCageRBD incorporates a *de novo* SARS-CoV-2 RBD binder[4] (LCB1, magenta). Middle panel: luminescence intensity upon addition of 16.7 nM SARS-Cov-2 RBD or trimeric spike protein to a mixture of 1 nM lucCageRBD+lucKey. Right panel: detection over a range of analyte concentrations in buffer, 10% synthetic nasal matrix[38] or 10% serum. **d,** Biosensor specificity. Each sensor at 1 nM was incubated with 50 nM of its cognate target (magenta lines) and the targets for the other biosensors (grey lines). Targets are Bcl-2, BoNT/B, human IgG Fc, Her2, cardiac Troponin I, anti-HBV antibody (HzKR127–3.2), anti-SARS-CoV-1-M polyclonal antibody and SARS-CoV-2 RBD. All experiments were performed in triplicate, representative data are shown, and data are presented as mean values +/− s.d.