Genetically Encoded Optical Biosensors for Drugs and Peptides Comprising De Novo Engineered Synthetic Mini-binders

Yuxin Pan

A dissertation

Submitted in partial fulfillment of the

Requirements for the degree of

Doctor of Philosophy

University of Washington

2024

Reading Committee:

Liangcai Gu, Chair

Andre Berndt

Richard Palmiter

Program Authorized to Offer Degree:

Biochemistry

University of Washington

Abstract

Genetically Encoded Optical Biosensors for Drugs and Peptides Comprising De Novo Engineered Synthetic Mini-binders

Yuxin Pan

Chair of the Supervisory Committee:

Liangcai Gu

Department of Biochemistry

Methods for *de novo* engineering of protein binders into chemically induced dimerization (CID) systems offer new opportunities for developing genetically encoded sensors for drugs and metabolites that lack suitable natural binders. However, integrating these binders into genetically encoded fluorescent sensors (GEFSs) remains largely unexplored. Here, we present a pipeline that efficiently selects synthetic CIDs and integrates them with a fluorescent domain to create single-protein GEFSs suitable for solution-based assays and mammalian cell applications.

As proof of concept, we used this pipeline to generate synthetic CID pairs from monobody and nanobody scaffolds and incorporated them into circularly permuted green fluorescent protein-based GEFSs. To optimize sensor performance, we created a library of 361 unique linker variants and conducted lysate-based screening to identify the most effective configurations. The resulting sensors exhibited a $\Delta F/F_0$ exceeding 100%. Furthermore, these sensors retained functionality in HEK293T cells when localized to the plasma membrane, cytoplasm, or ER lumen. We also adapted this

approach to additional optical domains, such as circularly permuted HaloTag, demonstrating its broad applicability.

Our results establish a foundation for expanding synthetic GEFSs, significantly enhancing the small molecule and peptide GEFS toolkit and reducing reliance on natural proteins and derivatives.

Contents

Acknowledgments

I thank Dr. Liangcai Gu for his mentorship and advise. You gave me much freedom to explore scientific problems in my own way and at my own pace, and had patience to allow me to learn through trial and error. You are a cheerleader that helps me overcome anxiety and frustration, and a great role model exemplifying resilience and optimism. You do not shy away from sharing with me your struggles and frustrations at work and seeing how you overcome those obstacles inspires me. My experience in your lab makes me a better scientist and a better person.

I also thank Dr. Shoukai Kang for his mentorship. As someone I work closely with, you are the epitome of an experimentalist for me. It is through you that I see my lack of thoroughness and commitment as a scientist, and I can never overstate how much you help shape both my experimental skills and scientific mindset. Comparing myself with you is always shameful, for I cannot imagine how I can be as good as you are, yet at the same time empowering, for I see a great example to chase after. Thinking of you easily reminds me of how much I still need to learn. I am indeed a very fortunate person to be trained by you.

I am blessed with wonderful lab mates who are delightful colleagues and friends. I thank Li Sun for being a very supportive lab manager, who not only works diligently to ensure the timely arrival of lab supplies, but also magically gets gadgets and instruments repaired with his mechanical engineering skills. I thank Dr. Xiaonan Fu for his help with my research and the fun we had fishing and camping together. I also thank Runze Dong, Dr. Runglawan Silakit, Dr. Cristen Molzahn, Dr. Mingzhu Lihu and Dr. Wen Xie for being cheerful companions that can help me. Outside of the lab, I need to say a big thank you to Lei Go for being a supportive friend and an inspiring role model. Our friendship is truly a bless.

I thank my committee members, Dr. Frank Dimaio, Dr. Richard Palmiter, Dr. James Hurley and Dr. Andre Berndt for their help and advice along the way. A big thank you to Erin Kirshner for keeping me on track and providing all the guidance and tips. Much of the work is funded by National Institute on Drug Abuse (5R21DA051555, 5R21DA051194 and 5R61DA051489).

Finally, I would like to thank my family, especially my parents, Yan Lei and Dr. Mingdao Pan, for their unconditional support. I will forever be grateful for being their child.

1. Introduction to Genetically Encoded Fluorescent Sensors

Optical biosensors translate molecular events, such as ion or molecule binding, enzymatic reactions, or changes in pH, into measurable optical signals. Genetically encoded fluorescent sensors (GEFSs) offer real-time tracking of biomolecule levels and various biological processes with exceptional spatiotemporal precision. Over the past two decades, these biosensors have become essential tools in biomedical research, substantially advancing bioimaging techniques across cellular, tissue, and whole-organism studies.

GEFSs can be generally categorized based on the principle of fluorescent readout they employ. The first category leverages Förster resonance energy transfer (FRET) technology, while the second uses a single fluorescent protein (FP). A solid understanding of the mechanisms behind both approaches enhances appreciation for the elegance of a single FP-GEFS design. In this section, I will briefly review these two categories, highlighting their underlying principles and discussing their respective strengths and limitations in various applications.

1.1. FRET-based GEFS

FRET-based GEFSs consist of a pair of fluorescent proteins (FPs) or fluorescent dyes, using FRET (Förster resonance energy transfer) as their optical readout. FRET is a non-radiative energy transfer process between two closely positioned chromophores, typically fluorophores in this context (1). The donor fluorophore releases energy, which is absorbed by the acceptor fluorophore. For FRET to occur, several key conditions must be met simultaneously. First, the emission spectrum of the donor must overlap with the excitation spectrum of the acceptor, usually by at least 30%, ensuring that the donor's energy is sufficient to excite the acceptor. Second, the distance between the donor and acceptor should be within 10 nm for optimal energy transfer efficiency. Third, the dipole alignment of the two fluorophores must be favorable. Finally, while not strictly required, a high quantum yield in the donor can enhance energy transfer, leading to more excited acceptor molecules, which is a desirable feature.

FRET is a powerful readout mechanism for GEFSs because biological processes, such as protein dimerization or conformational changes, can bring fluorophores into the necessary proximity for FRET to occur. When the fluorophores are too far apart for FRET, excitation of the donor leads to donor emission.

However, when biological events bring the fluorophores close enough, exciting the donor results in acceptor emission, indicating that FRET has taken place (Figure 1a). This shift from donor to acceptor emission serves as a clear signal of FRET activity. FRET-based sensors are commonly classified into two major types depending on the working mechanism: cleavage-based (Figure 1b) and conformational change/dimerization-based sensors (Figure 1c).

Cleavage-based sensors feature a donor and acceptor connected by a cleavable peptide linker and are frequently used to monitor protease or peptidase activity. In the absence of enzyme activation, the intact peptide keeps the donor and acceptor close enough for FRET to occur, and excitation of the donor leads to acceptor emission. Upon enzyme activation, the peptide linker is cleaved, separating the donor and acceptor, which disrupts FRET and shifts the emission from the acceptor back to the donor. One of the earliest examples of a cleavage-based FRET sensor monitors caspase-3 activity during apoptosis (2). In this system, a green fluorescent protein (GFP) is connected to a blue fluorescent protein (BFP) by a DEVD peptide, which is cleaved by caspase-3 during apoptosis, causing a shift from GFP (acceptor) emission to BFP (donor) emission. Another example is a sensor designed to monitor the proteolytic activity of the enteroviral protease $2A^{pro}$ (3). Here, enhanced cyan fluorescent protein (ECFP) and yellow fluorescent protein (EYFP) are linked by a peptide containing a $2A^{pro}$ cleavage site. When $2A^{pro}$ is active, a shift from EYFP emission to ECFP emission indicates viral infection.

Cleavage-based sensors are typically irreversible due to the permanent nature of peptide cleavage, limiting their application mainly to protease activity detection. In contrast, conformational change-based FRET sensors are generally reversible and can detect a wider range of biological processes and molecules.

Conformational change-based sensors feature a donor and acceptor fused to a protein that undergoes structural changes in response to specific biomolecules or biological events. These conformational shifts alter the distance between the donor and acceptor, ultimately regulating whether FRET can occur. In some cases, the conformational change is an intrinsic property of the protein itself. In other cases, this mechanism is engineered by linking a pair of heterodimers that dimerize in response to a particular ligand or biological phenomenon of interest.

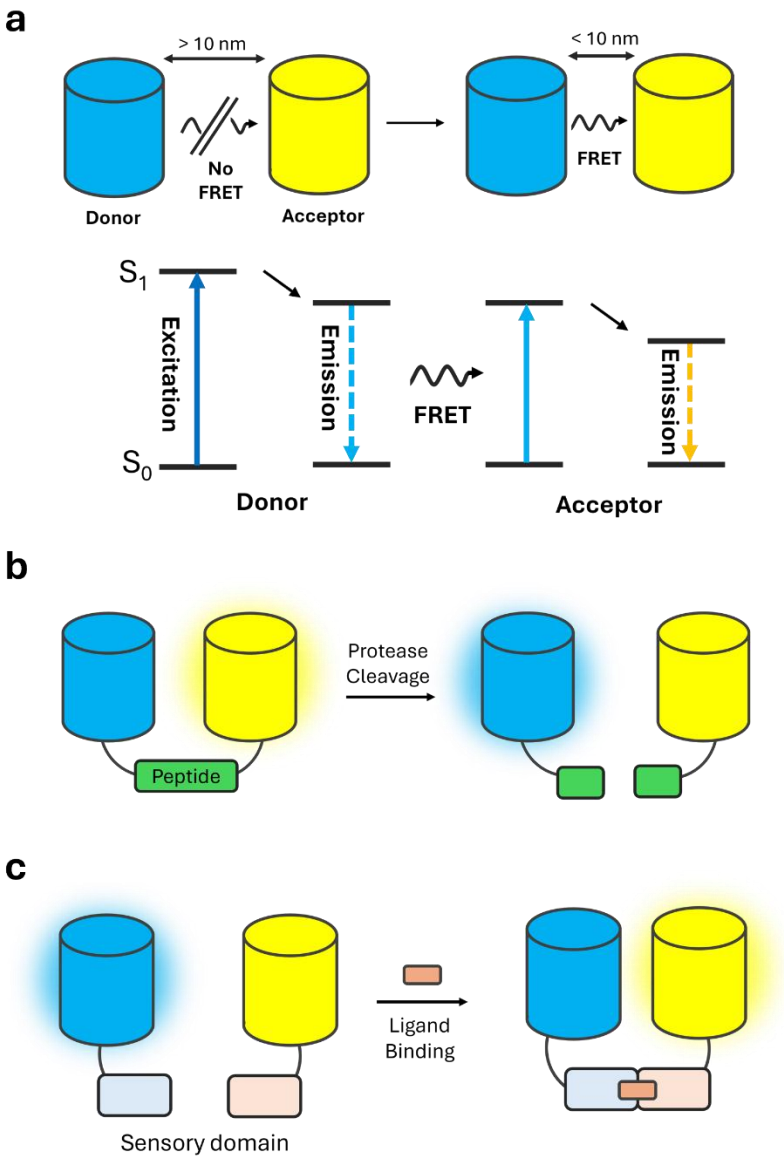


Figure 1 Schematic of FRET-based GEFSs. (a) Schematic of FRET principle. (b) FRET-GEFS based on enzymatic cleavage. (c) FRET-GEFS based on ligand binding and conformational changes.

The calcium sensor *Cameleon* is one of the earliest examples of conformational change-based sensors (4). In *Cameleon*, the calcium-binding protein calmodulin and the M13 peptide are linked by a short peptide, with each fused to one of the FRET pair proteins, CFP and YFP. Upon the presence of Ca^{2+}, calmodulin binds to M13, causing dimerization that brings the FRET pair into close proximity. This

interaction results in a shift from donor to acceptor emission. Since the development of *Cameleon*, many other sensors for detecting metal ions have been created (5).

Using a similar sensing mechanism, sensors for detecting small molecules have also been developed. One notable example is cAMP sensors based on Epac (exchange proteins directly activated by cAMP). In the earliest versions of these sensors, EGFP and citrine were fused to the termini of Epac (6); upon binding to cAMP, Epac undergoes a conformational change, resulting in an increase in FRET between the two fluorescent proteins. An early example of metabolite sensors is a maltose sensor in which a truncated maltose-binding protein (MBP) is flanked by ECFP and EYFP (7). Similarly, when MBP binds to maltose, it undergoes a conformational change that modulates FRET between the fluorescent proteins. Since these early developments, many sensors utilizing similar mechanisms have been developed (8).

In addition to ligand detection, conformational change-based sensors can also monitor biological events. A prominent example is the A-kinase activity reporter (AKAR) (9), which is designed to track the activity of protein kinase A (PKA). AKAR consists of a PKA-specific phosphorylatable peptide fused with a phosphoamino acid-binding domain, and this heterodimer pair is flanked by CFP and YFP. When PKA is active, it phosphorylates the specific peptide, allowing it to bind to the phosphoamino acid-binding domain. This dimerization event triggers an increase in FRET between the fluorescent proteins. Similarly, over the years, many sensors have been developed to monitor the activity of kinases and GTPases (10).

A key advantage of FRET sensors over single FP sensors is their ability to enable ratiometric measurements, making them better suited for quantitative studies. Another distinct feature of FRET sensors is their compatibility with fluorescence lifetime imaging microscopy (FLIM). Fluorescence lifetime refers to the delay between a fluorophore’s excitation and its emission. When FRET occurs, the donor fluorophore has an additional pathway for relaxation, which shortens its lifetime. FLIM offers excellent temporal resolution for live-cell imaging and is largely independent of fluorophore concentration. As a self-referenced method, FLIM is also less reliant on precise instrument calibration.

However, FRET sensors do have limitations. One major constraint is that the FRET pair requires a certain level of spectral overlap, which restricts the number of usable channels and hampers the application of FRET sensors in multiparameter imaging. Another issue is that fluorophores have varying photobleaching

rates, which can introduce bias in long-term experiments. Additionally, since acceptor fluorophores are excited by the energy emitted from donors rather than directly from the light source, the signal intensity of FRET sensors tends to be lower than that of single FP sensors. FRET sensors also frequently exhibit lower signal-to-noise ratios, and connecting two fluorescent proteins via a peptide linker inherently results in higher background noise compared to freely diffusing proteins in solution.

1.2. Single FP-based GEFS

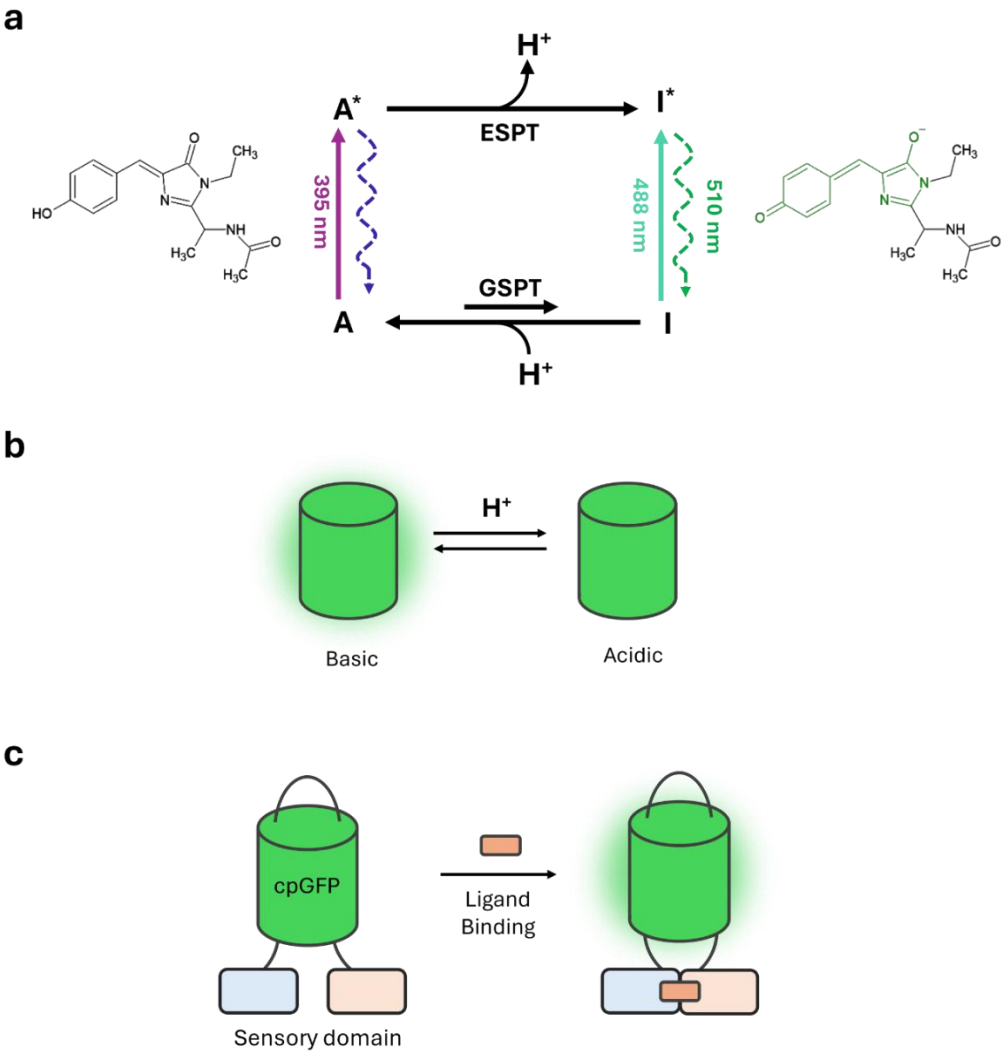


Figure 2 Schematic of single FP-based GEFSs. (a) Schematic of GFP fluorophore in protonation/deprotonation states. A: protonation state; I: deprotonation state. ESPT: excited-state proton transfer; GSPT: ground-state proton transfer. (b) Schematic of GFP pH sensor utilizing inherent proton sensitivity. (c) Schematic of cpGFP-based GEFS.

Unlike FRET-based sensors, single FP-based sensors contain only one fluorescent protein. Based on their design, these sensors can be categorized into two major types: intrinsic sensitivity-based sensors (Figure 2b) and exogenous sensory domain-based sensors (Figure 2c).

Intrinsic sensitivity-based sensors utilize FPs as both the reporter and the sensor, taking advantage of the fact that some FPs are inherently sensitive to specific chemical parameters. A prominent example is pH sensors. GFP, for instance, is highly sensitive to pH, with its fluorescence intensity significantly decreasing in acidic conditions due to increased protonation (Figure 2a). Although this sensitivity is often considered a disadvantage, it allows GFP to serve as the basis for pH indicators (11). One of the earliest and most widely used pH sensors for studying exocytosis and endocytosis is superecliptic pHluorin (SEP) (12). SEP is a mutant form of enhanced GFP (EGFP), which is bright at neutral pH but gradually loses fluorescence as the environment becomes more acidic, becoming nearly non-fluorescent at pH 5.5. This property makes SEP particularly useful for studying secretory vesicle formation and exocytosis, processes that involve acidification of microenvironments. More recently, a red variant with similar sensitivity, called pHuji, has been developed (13). Another example of intrinsic sensitivity-based sensors is redox-sensitive GFP. In these sensors, two amino acids on the surface of GFP are replaced by cysteines, allowing the formation of a disulfide bond in oxidizing environments (14). These redox-sensitive GFPs are valuable for monitoring redox potential in various cellular compartments, such as mitochondria.

Exogenous sensory domain-based sensors incorporate sensory domains into the structure of FPs. In these sensors, the FP is fused with a protein whose allosteric changes modulate the fluorescence intensity of the FP. To accomplish this, single-FP sensors are engineered so that inserting a protein at specific sites within the FP disrupts its structural integrity, leading to a loss of fluorescence. When the sensory domain undergoes a conformational change upon ligand binding, the FP’s structure is restored, resulting in an increase in fluorescence. Pioneering work by the Tsien lab (15, 16) demonstrated that the Y145 position in EYFP is particularly tolerant to protein insertion. By inserting calmodulin into EYFP at this position, they created a calcium indicator called Camgaroo, which shows a several-fold increase in fluorescence upon calcium binding.

Alternatively, FPs can also be inserted into sensory domains. To maximize allosteric modulation of fluorescence, these FPs are often circularly permuted (CP); the details of this engineering process will be discussed later. One of the earliest sensors using cpFP as the reporter domain is GCaMP (17,18,19,20,21,22). In GCaMP, cpEGFP is flanked by the Ca^{2+}-binding protein calmodulin and the M13 peptide. When Ca^{2+} is present, calmodulin binds to M13, restoring the structural integrity of the EGFP beta barrel and resulting in an increase in fluorescence.

Following the same design and engineering principles of GCaMP, numerous sensors targeting various ligands have been developed over the past two decades (23). Examples include the glutamate sensor iGluSnFR (24,25,26), dopamine sensors such as dLight (27,28) and GRABDA1m (29), the ATP sensor iATPSnFR (30,31), the peroxide sensor HyPer (32,33,34), and voltage sensors like ASAP (35,36,37,38,39,40,41) and rEstus (42), and many others.

One of the primary limitations of single-FP sensors is that their readout is intensiometric, making them less suitable for quantitative studies compared to ratiometric FRET sensors. However, single FP-based sensors offer several advantages. Since they are directly excited by the light source, they typically exhibit higher intensity and a better signal-to-noise ratio. Additionally, single-FP sensors do not suffer from issues related to differences in pH sensitivity or photobleaching kinetics between two fluorophores, which can lead to false positives in FRET systems. Another key benefit is that single FP-based sensors use only a single channel, making them more compatible with multiparameter imaging. Due to their versatility and ease of use, single-FP sensors have become invaluable tools in imaging and often play pivotal roles in biomedical research.

1.3. Components of GEFS

Within the scope of this work, I will now focus my discussion on single FP-based GEFSs with exogenous sensory domain-based sensors. Functionally, this type of GEFS can be viewed as a fusion of three key components (Figure 3): an optical domain that provides the fluorescent readout, a sensory domain that detects ligands of interest, and linkers that connect these two domains. In the following sections, I will provide an overview of the commonly used engineering strategies for each of these parts, starting with the sensory domains.

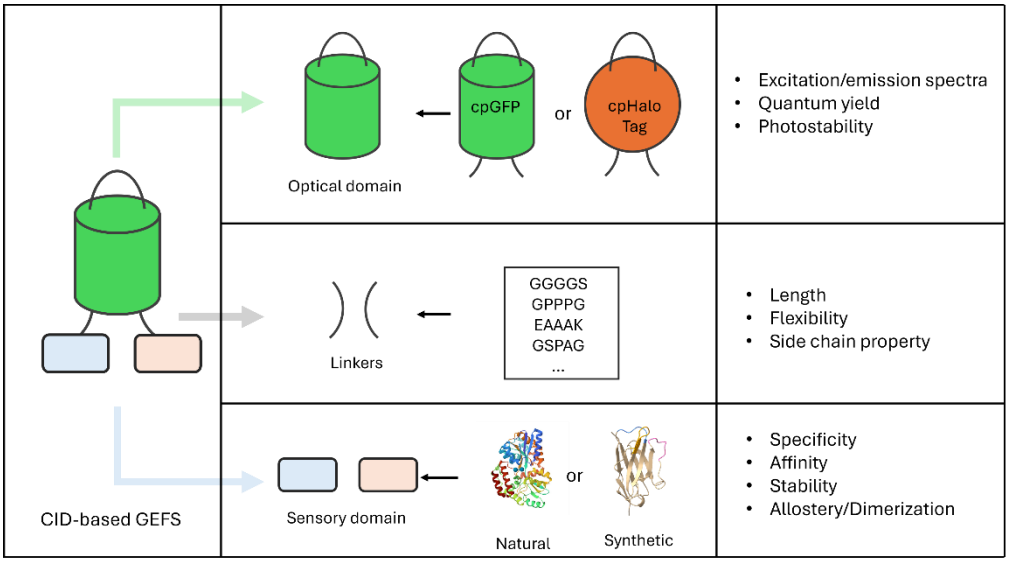


Figure 3 Schematic of single FP-based GEFS. Sensory domains can be either a single protein with insertion, or a pair of dimers.

1.3.1. GEFS components: sensory domains

The sensory domains of GEFSs serve two critical functions: first, they must bind to their target ligands with high specificity and appropriate affinity; second, they need to transmit the binding event to the optical domain through allosteric changes. Due to these requirements, sensory domains are typically protein binders found in nature that undergo conformational changes or dimerization upon ligand binding. Among these natural binders, periplasmic binding proteins (PBPs) and G-protein-coupled receptors (GPCRs) are particularly common choices for sensory domain engineering.

PBPs are a superfamily of bacterial receptors that mediate chemotaxis and transportation through ligand binding. PBPs can bind to a large variety of natural molecules, including ions, amino acids, peptides, carbohydrates, and so forth. PBPs share a common structural motif: two domains connected by a hinge region, with the ligand-binding site located at the interface between the two domains (Figure 4c, 4d). The protein shifts between two conformations through a bending motion around the hinge—a ligand-binding closed form and a ligand-free open form. Due to these structural characteristics, PBPs are often referred

to as "Venus flytrap" proteins. The inherent coupling of allosteric changes with ligand binding makes PBPs ideal sensory domains for biosensor engineering. Moreover, with over 300 PBP structures resolved to date (43), this protein family is particularly well-suited for structure-based engineering.

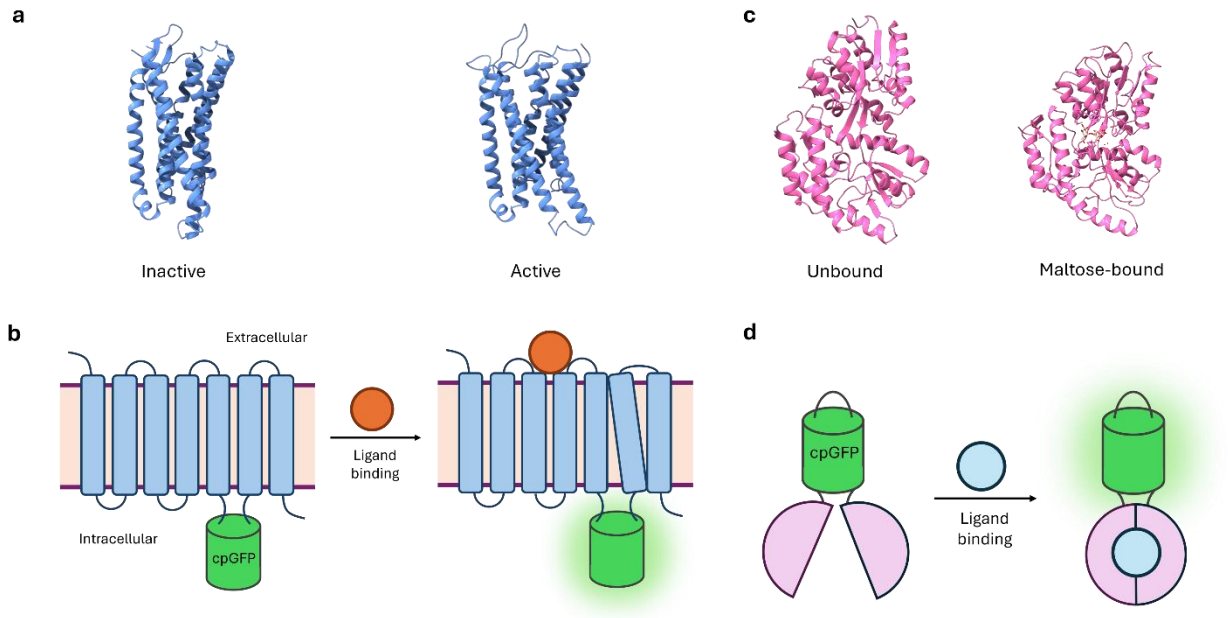


Figure 4. Schematics of GEFSs using GPCR and PBP as sensory domains. (a) Structures of mu-opioid receptor in inactive (7UL4) and active (6DDE) form. (b) Schematic of GPCR-based GEFSs. (c) Structure of maltose binding protein in unbound (1anf) and bound (1peb) form. (d) Schematic of PBP-based GEFSs.

In the context of single FP-based sensors, converting PBPs into sensors is typically achieved through domain insertion, where cpFPs are inserted into PBPs. Since the conformational change in PBPs occurs near the hinge region, cpFPs are usually inserted close to this area to maximize the coupling between the allosteric shifts in PBPs and the resulting changes in fluorescence intensity of the cpFPs.

Numerous single FP-based GEFSs using PBPs as sensory domains have been developed. One of the earliest examples of a PBP-single FP fusion is a maltose sensor that uses the maltose-binding protein (MBP) as its sensory domain (44). Following similar design principles, sensors for glucose (45,46), phosphonate (47), histidine (48), glutamate (24), GABA (49), acetylcholine (50), and other ligands have been created over the past decade.

A major limitation of PBPs is that they primarily bind to metabolites, reflecting their natural biological roles. This constraint can make it difficult to find a PBP that binds to specific targets, such as synthetic molecules. Consequently, developing sensors using PBPs often requires extensive searches of protein databases and genomes across various species, and in some cases, it may not be feasible. On the other hand, the fact that PBPs rarely bind synthetic compounds makes them less prone to nonspecific responses induced by pharmaceutical compounds, making them especially suitable for studies involving drug administration.

GPCRs are a family of membrane receptors that bind to extracellular ligands and transduce these binding events into intracellular signals by coupling with G proteins and arrestins. Each GPCR contains seven transmembrane domains connected by extracellular and intracellular loops. Unlike simple "on-off" switches, GPCRs are highly dynamic systems capable of adopting multiple conformations. These conformational states are selected by ligand binding, which can influence downstream signaling or induce signaling bias (51). The most significant conformational change in GPCRs occurs between transmembrane helices 5 and 6 (Figure 4a). To optimize the coupling between allosteric changes and fluorescence intensity shifts, cpFPs are typically inserted into the third intracellular loop that connects these two helices (Figure 4b).

One of the earliest examples of GPCR-based sensors is an acetylcholine sensor with cpGFP inserted into the third intracellular loop of a human muscarinic acetylcholine receptor (M_3R) (52). Since then, following similar design principles, sensors sensing various biomolecules have been developed, including dopamine (53, 27), norepinephrine (54), ATP (55), serotonin (56), GABA (57), acetylcholine (58), oxytocin (59, 60) and orexin (61). It is particularly noteworthy that recently Yulong lab has developed a generalizable method to generate GPCR-sensors by grafting the same peptide linkers and cpGFP onto various GPCRs at the third intracellular loop. Using the method, Yulong lab has developed an array of neuropeptide sensors with high specificity and affinity (62).

GPCRs offer unique advantages as sensory domains. Since GPCRs primarily bind to neurotransmitters and neuromodulators in nature, they are naturally suited for sensors designed to monitor these molecules. Unlike PBPs, which are derived from prokaryotic systems, GPCRs originate from eukaryotic

systems where their ligands function naturally. This makes their *in vivo* specificity and affinity for ligands more biologically relevant. Additionally, even for the same ligand, GPCR subtypes often exhibit distinct affinities, specificities, and signaling patterns (63), enabling the creation of sensors with tailored performance characteristics.

However, GPCRs are also more sensitive to pharmaceutical interventions, which can limit their use in studies involving drug administration. Furthermore, as membrane proteins, GPCR-based sensors are difficult to direct to subcellular locations. Their membrane-bound nature also renders them incompatible with conventional high-throughput screening systems such as those used in yeast and bacteria. As a result, screening and characterizing GPCR-based sensors must be conducted on the plasma membrane of mammalian cells, which can be time-consuming and labor-intensive.

In addition to PBPs and GPCRs, several other natural protein binders are used in sensory domain engineering. A prominent example is the plant hormone abscisic acid (ABA) receptor PYR1, along with its coreceptor HAB1 (64). Upon ABA binding, PYR1 undergoes a conformational change that enables it to interact with HAB1, forming a heterodimer. By mutating the amino acids within the PYR1 binding pocket, this ABA-sensitive heterodimerization system can be re-engineered to detect small molecules other than ABA (65). Zinc finger proteins, such as ZF1 and ZF2, provide another example. In the presence of zinc ions, ZF1 and ZF2 dimerize, leading to increased cpGFP brightness (66). While these natural protein binders are less common than PBPs and GPCRs, they offer valuable alternatives for ligand detection.

1.3.2. GEFS components: optical domains

The optical domain is another crucial component of a GEFS. It typically consists of a fluorescent protein, whose fluorescence intensity can be modulated through allosteric changes or dimerization of the sensory domain. One of the most used optical domains with this capability is circularly permuted green fluorescent protein (cpGFP).

Circular permutation refers to the rearrangement of a protein’s sequence that alters the connectivity of its termini without significantly affecting its overall structure. The phenomenon of naturally occurring circular permutation was first reported in 1974 (67), and years later, artificial circular permutations were employed

to study protein folding (68, 69). Circularly permuted GFP (cpGFP) was subsequently identified as a GFP variant with increased sensitivity to its surrounding microenvironment (15).

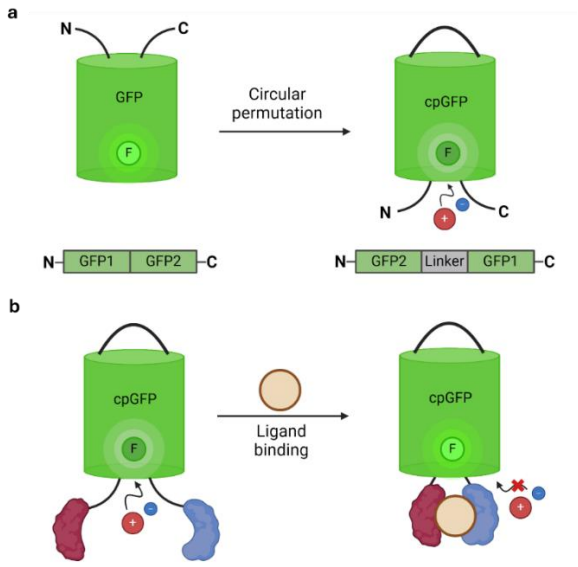


Figure 5. Schematic of circularly permuted GFP. (a) Circular permutation repositions protein termini into proximity of the fluorophore. (b) Repositioned protein termini expose fluorophore to solvent, making it sensitive to surrounding microenvironment

Artificial circular permutation mimics the natural process by connecting the original N- and C-termini of a protein with a peptide linker and introducing new termini at a different, strategically selected position. Like its natural counterpart, artificially circularly permuted proteins retain their structure despite the altered connectivity (70).

In the case of cpGFP, the protein remains fluorescent, but the newly introduced termini are positioned adjacent to the fluorophore, which was previously well-shielded within the beta-barrel structure (Figure 5a). As a result, the fluorophore is exposed to the solvent, making it sensitive to the surrounding microenvironment. This structural feature underpins cpGFP's ability to modulate fluorescence intensity through allostery and dimerization (Figure 5b).

cpGFP has become a widely used optical domain in many GEFS designs, with examples provided in the previous sections. However, one limitation of cpGFP is its inherent sensitivity to pH (11): under acidic conditions, cpGFP becomes protonated, causing a significant loss of fluorescence intensity and allosteric functionality. While this pH sensitivity allows cpGFP to be engineered into pH sensors, it also restricts its use in subcellular environments with low pH, such as lysosomes. Furthermore, cpGFP's excitation by shorter wavelengths makes it less suitable for deep tissue imaging due to light scattering. In multi-parameter imaging, having optical domains with diverse color spectra is also advantageous. For these reasons, although cpGFP remains a popular optical domain for GEFS engineering, there is ongoing research to develop optical domains with different colors and properties.

To develop optical domains beyond cpGFP, other fluorescent proteins (FPs), such as red-shifted FPs have been studied. However, previous studies have shown that optimizing red-shifted cpFPs presents more challenges (71, 72), often resulting in a lower dynamic range and brightness. As an alternative to cpFPs, cpHaloTag labeled with the rhodamine dye JF635 has been developed in recent years (73). HaloTag is a self-labeling protein (74) that can covalently bind various fluorescent dyes via nucleophilic attack on a chloroalkane group to acquire fluorescence (Figure 6a). JF635 is highly fluorogenic (75), meaning its brightness significantly increases upon conjugation with another molecule. The fluorogenic property of JF635 arises from its equilibrium between a nonfluorescent lactone form and a fluorescent zwitterion form, with conjugation shifting the balance toward the fluorescent state.

In the case of cpHaloTag, JF635 is exposed to the solvent, making its equilibrium sensitive to the surrounding microenvironment and thus modulable by the allostery or dimerization of sensory domains (Figure 6b). In addition to replacing the optical domain in GCaMP (73), cpHaloTag-JF635 has been used to develop kinase sensors (76) and peroxide sensors (77). Although the red-shifted spectra of cpHaloTag-JF635 offer advantages for deep-tissue imaging and JF635 is less pH-sensitive than GFP, these sensors require an external supply of JF635. For use in an animal, the cost of JF635 can be substantial. Moreover, dye delivery in complex systems poses another challenge, as dye can quickly diffuse and be actively cleared from the system. These limitations may explain the lack of widespread *in vivo* applications for cpHaloTag-based GEFSs.

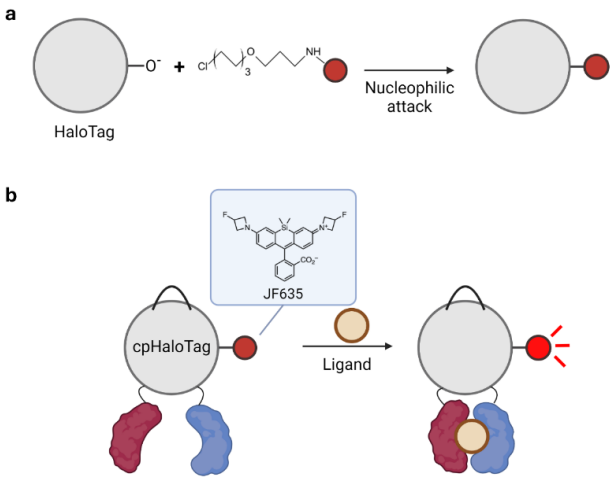


Figure 6. Schematic of HaloTag and cpHaloTag-based GEFS. (a) HaloTag can be conjugated to chloroalkane substrates through nucleophilic attack (b) J635-labeled cpHaloTag is modulable by sensory domain allostery.

1.3.3. GEFS components: linkers

Another crucial component of GEFS is the linker that connects the sensory domain to the optical domain. The functionality of a GEFS relies on the coupling between the allostery or dimerization of the sensory domain and the optical domain, with linkers playing a pivotal role in transmitting mechanical shifts between them. Peptide linkers are typically evaluated based on several parameters: length, hydrophobicity, secondary structure, and rigidity (78). Length is determined by the number of amino acids in the linker, while the other parameters are influenced by the physical and chemical properties of the amino acid side chains. Generally, smaller and more hydrophobic side chains, like glycine and serine, provide greater flexibility. In contrast, proline significantly increases the rigidity of the linker. Several studies have demonstrated how these linker properties impact the function of fusion proteins (79, 80).

In the context of cpGFP-based GEFS, an additional key consideration is how the linker influences the microenvironment surrounding the fluorophore. Specifically, the two amino acids located at the N- and C-termini of cpGFP, often referred to as the "gatepost" residues, are particularly important (81). These amino acids are positioned at the opening of the beta-barrel, which shields the fluorophore, and their side chains can significantly affect the coupling between cpGFP and the sensory domain. Optimizing these

residues is typically case-specific and requires mutagenesis and variant screening to identify the ideal configuration for each sensor.

1.3.4. GEFS component engineering

Identifying suitable proteins for each component of a GEFS is just the first step; engineering and fine-tuning these components into a functional fusion protein with optimal synergy presents a more complex challenge. Of the three components, the optical domain generally requires the least engineering effort. This is because cpGFP is a well-established and highly adaptable optical domain that can be easily paired with a variety of sensory domains. However, the sensory domain and linkers are often more case-specific and require extensive optimization to effectively modulate cpGFP, achieving both a good dynamic range and the desired ligand affinity. In this section, I will review common approaches to engineering sensory domains and linkers.

GEFS component engineering: sensory domain

In most cases, a critical step in engineering GEFSs is determining the insertion position of the sensory domain, which is connected to the optical domain via peptide linkers. This insertion position must enable conformational changes in the sensory domain to be efficiently transmitted to the optical domain, typically cpGFP. The combination of an optimal insertion position and well-optimized peptide linkers is key to achieving a broad dynamic range in GEFSs.

Since many sensory domains used in GEFS engineering are naturally occurring protein receptors or binders with known structures, a common strategy for identifying optimal insertion sites is structure-guided screening. Potential insertion positions are selected by examining protein structures for solvent-exposed loops that are likely to undergo conformational changes. Prototype sensors are then screened to identify the variant with the largest ligand-induced response. When only one conformation has available structures, further inference is needed to predict the regions undergoing conformational changes. For example, in a study developing a lactate sensor, 70 different positions were screened (82). In contrast, when both the unbound and ligand-bound structures are available, it becomes easier to identify regions undergoing conformational shifts. In a study using maltose-binding protein to create a maltose sensor, only four positions were screened (44).

Due to the complexity of allosteric coupling between sensory and optical domains, rational design using structural information often requires screening of a narrowed candidate pool, rather than directly pinpointing the optimal insertion site. As a result, determining the insertion site remains a labor-intensive and rate-limiting step. To improve the efficiency of this process, methods for rapid screening of insertion positions have been developed. One such approach, called domain insertion profiling, uses transposons and transposase to randomly insert cpGFP into sensory domains. Combined with cell sorting, this method allowed for efficient screening of 210 out of 370 productive insertions and successfully identified a new insertion site that improved the sensor's maximum response by several folds (83).

GEFS component engineering: linker optimization

Compared with optimizing the insertion position of domains, linker optimization involves exploring a significantly larger space of countless combinations of peptide lengths and amino acid compositions. Structural information can facilitate this process; for example, the performance of GCaMP2 was improved by mutating its linker region with structural guidance (84). Similarly, the development of GCaMP5 involved rational design (85). However, due to the vast number of possible linkers, screening a library of variants is usually necessary. Even in cases where linkers are short, around two amino acids, to maximize allosteric coupling between domains (81), it is still common to screen hundreds, if not thousands, of variants to find one with optimal performance.

Given the large number of variants to be screened, methods have been developed to increase efficiency. One such method is bacterial lysate screening (82), where each bacterial clone (a unique sensor variant) is grown individually in a 96-deep-well plate. After induction and protein expression, bacteria are lysed to release sensor variants, and the lysates are evaluated using a fluorescent plate reader to identify clones with desirable performance. This method works for sensors compatible with bacterial expression systems; however, some sensors, such as those utilizing GPCRs as sensory domains, cannot be expressed in bacteria. To address this, a mammalian cell-based screening platform called the optogenetic microwell array screening system was developed (86). This system isolates individual mammalian clones in microwell arrays, allowing them to be individually analyzed under a microscope. This approach

successfully screened over 13,000 dopamine sensor variants and identified an improved sensor with more than a six-fold signal increase compared to the original.

Another method, called sort-seq (or FACS-seq), combines fluorescence-activated cell sorting (FACS) with deep sequencing to screen constructs based on their fluorescence levels (87-91). This method has been used in recent years to search for single-FP-based GEFS with improved dynamic range (92, 93). Cells are sorted into bins based on fluorescence intensity, and high-throughput sequencing is used to infer the dynamic range of each variant. Using this method, researchers improved the maximum response of a pyruvate sensor by threefold. In another study, more than 15,000 variants were screened, resulting in the discovery of a variant with a dynamic range exceeding 1000% (76).

GEFS component engineering: other considerations

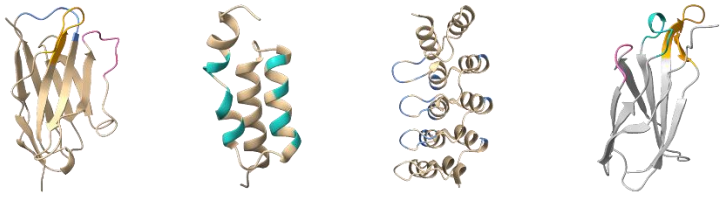
An optimal insertion position in the sensory domain, combined with optimized linkers, can significantly enhance GEFS performance. However, other factors, such as the interface between sensory and optical domains, also play a role. Amino acids at the domain interface can affect the microenvironment surrounding the fluorophore, and site-directed mutagenesis at this interface can lead to enhanced sensor performance. For example, in the development of GCaMP6, 16 positions at the interface between cpGFP and calmodulin were mutagenized, resulting in the discovery of three ultrasensitive GCaMP6 variants (21).

It is also common to improve sensor performance through directed evolution, using whole-gene mutagenesis. For instance, in the development of the extracellular lactate sensor iLACCO1, 11 rounds of directed evolution led to a variant with a more than tenfold improvement in dynamic range (94). Naturally, the screening methods used for linker optimization can also be applied to variants created through site-directed or whole-gene mutagenesis.

1.4. Synthetic protein binders as GEFS sensory domains

In the current research landscape, the development of GEFSs largely centers on inserting cpGFP into naturally occurring protein receptors. This process typically begins with identifying potential natural protein receptors, followed by determining the optimal insertion site through structural analysis and screening.

Multiple rounds of directed evolution are often required to optimize the linkers and the entire gene. This well-established engineering approach has yielded many successes, such as the widely recognized GCaMP-series sensors. However, despite its effectiveness, this approach has limitations. While naturally occurring protein receptors offer several advantages as sensory domains, they are often unsuitable for ligands of artificial origin, such as drugs, compared to natural biomolecules. Even when receptors for specific drugs are identified, these interactions may lack the necessary specificity, as seen with opioid receptors and opioid drugs.



Scaffold name	Nanobody	Affibody	DARPin	Monobody
Molecular Weight	~14 kDa	~6.5 kDa	~14-21 kDa	~10 kDa
Residues	~120	58	67+33*n	~100
Immunoglobulin or not	Yes	No	No	No
Parent protein origin	Camelid	Bacteria	Human	Human

Table 1. Overview of compact synthetic binding scaffolds and some of their characteristics.

Synthetic protein binders present a potential solution to the limitations of naturally occurring proteins (Table 1). However, most synthetic protein binders do not exhibit allostery, making them unsuitable for direct coupling with cpGFP as sensory domains for GEFSs. Nevertheless, engineering methods, although rarely practiced, have been developed to introduce allostery into some synthetic binder scaffolds, which will be briefly discussed in later sections. More straightforwardly, synthetic binders can be coupled with cpGFP through an alternative mechanism to allostery: dimerization, specifically chemically induced dimerization (CID).

Chemically induced dimerization (CID) refers to the process in which two molecules are brought together by chemical ligands, such as small molecules or peptides. CIDs play a critical role in biological processes like transcription regulation, signal transduction, and protein degradation (95, 96). Since the first demonstration of controlling signaling pathways using FK1012 and FK binding proteins (FKBPs) (97), CIDs have become powerful tools in research and therapeutic contexts for controlling biological events at the cellular level (98-102). Just as conformational changes in a sensory domain can alter the microenvironment around the cpGFP fluorophore to modulate its brightness, dimerization can achieve similar results. In this way, synthetic binders forming CIDs can serve as sensory domains for GEFSs.

Despite the theoretical potential, few GEFSs using CIDs as sensory domains have been reported. This is partly due to the limited availability of naturally occurring CIDs. For example, PYR1-HAB is one of the few natural CIDs used for GEFS development, but redesigning its binding pocket to accommodate different small molecules is challenging. Computational analysis of the PYR1-HAB binding pocket has been conducted (103) and may provide guidance for future engineering projects, but the number of available natural CIDs remains limited. However, recent years have seen the emergence of methods to generate synthetic CIDs from combinatorial libraries (103), opening exciting new possibilities for GEFS engineering. Unlike naturally occurring protein receptors, synthetic CIDs can be engineered specifically for a wide range of ligands, including drugs and biosynthetic intermediates, not just natural biomolecules such as metabolites or signal transmitters. This could significantly expand the detection range of GEFSs.

Another advantage of using synthetic CIDs as sensory domains is that it eliminates the need to determine an optimal insertion site for cpGFP or other optical domains, which can be a challenging and time-consuming process. The first obstacle is that many proteins lack well-resolved structures, often requiring the use of homologous structures as references. Additionally, identifying positions that undergo the most significant conformational changes requires structures of different conformations, which are even rarer. Even with structural information, determining the optimal insertion site is often case-specific and usually requires some level of screening. In contrast, due to the inherent nature of dimers, connecting an optical domain to CIDs is straightforward and does not require insertion site optimization. This simplifies the experimental process and accelerates the engineering of GEFSs.

In summary, synthetic binders/CIDs offer a sensory domain that complements naturally occurring proteins, both in terms of application and engineering. In this section, we will explore commonly used synthetic binders, methods to generate synthetic CIDs, and strategies for integrating them into GEFSs. Here, I will briefly review some of the synthetic binder scaffolds that are compact (single domain, no more than 20 kDa) and thus more suitable for GEFS fusion (Table 1). Understandably, most of these compact synthetic binder scaffolds are non-immunoglobulin, apart from nanobody.

1.4.1. Synthetic protein binders

Nanobodies (Nbs), also known as single-domain antibodies or VHHs, consist of a single N-terminal variable domain of the heavy chain. Nanobodies were developed following the discovery that camelids produce functional antibodies devoid of light chains (114). In addition to being derived from camelid antibodies, Nbs can also be selected from synthetic libraries using phage, ribosome, yeast, and bacterial display (115-118). Nanobodies contain four framework regions and three complementarity-determining regions (CDRs). The framework regions form the core scaffold, while the CDRs are responsible for antigen binding. Therefore, synthetic nanobody libraries typically maintain a fixed scaffold while introducing high amino acid diversity into the CDRs. Due to their compact size, nanobodies exhibit faster tissue penetration and clearance, and their single-domain structure enables easy genetic manipulation.

Nanobodies have been engineered into various types of biosensors, such as fluoro-immunosensors called Quenchbody (Q-body) (119). This design leverages photoinduced electron transfer (PET), in which the indole side chain of tryptophan (Trp) acts as an electron donor to certain fluorescent dye molecules. This means tryptophan in the protein can quench fluorescent dyes. In Q-body, a nanobody is conjugated to 5-carboxytetramethylrhodamine (TAMRA). In the presence of methotrexate (MTX), nanobody binding to MTX alters the distance between its tryptophan residues and TAMRA, resulting in a change in TAMRA's fluorescent intensity, thereby functioning as a sensor. Additionally, nanobodies can be engineered into FRET-based sensors by fusing them with fluorescent proteins. For instance, ochratoxin-specific nanobodies have been fused with sfGFP to detect quantum dot-labeled ochratoxin (120). Upon binding ochratoxin, the nanobody-sfGFP complex is brought into proximity with red quantum dots, enabling FRET to occur.

Nanobodies have also been engineered to bind specific conformations of allosteric proteins, allowing them to serve as sensors for detecting conformational changes in these proteins. For example, nanobody Nb80 binds specifically to the agonist-bound form of the β2-adrenoceptor (121), while nanobody Nb39 binds to the agonist-occupied opioid receptor (122). Nb80 was directly fused with GFP and used as a probe to detect the conformation of the β2-adrenoceptor (123). In a more sophisticated design, Nb39 was fused with cpGFP and connected to the opioid receptor (124). In this system, unbound Nb39 interferes with cpGFP maturation, rendering the sensor nonfluorescent. Upon binding the activated opioid receptor, Nb39 is released from cpGFP, allowing cpGFP to mature, which results in increased fluorescence.

Nanobodies have also been engineered to become allosteric, though the methods are rarely practiced. For instance, a single, split nanobody was fused with cpGFP to create a GEFS capable of sensing seven amino acids from the C-terminus of bone gla protein (125). More recently, a new technique has emerged to circularly permute nanobodies, introducing allostery and allowing direct modulation of cpGFP brightness by a single nanobody protein (126).

An affibody is a small protein binder (6.5 kDa) consisting of a three-helix bundle, originally derived from the Z-domain of staphylococcal protein A (104). It contains 58 amino acid residues, with 13 positions in helices one and two randomized to create variable binding surfaces. Due to their compact size and rapid folding, affibodies have been increasingly applied not only in therapeutic settings but also in sensor development (105). Notably, compared with nanobodies and other binding scaffolds, affibodies can be chemically synthesized due to their small size, allowing for straightforward chemical labeling. This property enables convenient radiolabeling or fluorescent labeling of affibodies for imaging applications. For instance, affibodies have been used as tracers for human epidermal growth factor receptor 2 (HER2). In one study, the affibody ZHER2:342 was synthesized as a peptide, coupled with DOTA, and labeled with 111In. The DOTA-affibody demonstrated potential for *in vivo* imaging of HER2-expressing tumors, showing rapid labeling at low temperatures (within 30 minutes) and efficient tumor uptake (106). Additionally, affibodies have been utilized in FRET-based sensor development. In one example, affibodies were labeled with fluorophores in a site-specific manner, where a donor fluorophore was attached to a cysteine in the B domain (introduced via mutagenesis) and an acceptor fluorophore was attached to the N

terminus (107). Upon binding to target ligands, the fluorophores were brought into close proximity, enabling intramolecular FRET to report ligand presence.

DARPins (designed ankyrin repeat proteins) are small (14 to 21 kDa), single-domain binding proteins. As the name suggests, DARPins are derived from natural ankyrin repeat proteins. Repeat proteins feature an architecture in which a structural motif repeats multiple times to form the repeat protein domain that can bind to target molecules (108). The original DARPin libraries were consensus designed (109). Consensus design is based on hypothesis that the consensus amino acid in a given sequence contributes more than average to the desirable traits. Each DARPin library element consists of three parts: an N-cap, a C-cap and a library module. The N- and C-caps are unchanged for all elements, whereas library modules contain variable number of repeats with various amino acid composition. Each library module contains 33 amino acids among which 7 are variable, and each DARPin molecule typically contains 2-4 library modules (110). Alternatively, DARPin libraries can also be engineered with combinatorial randomization (111), in which variable positions in library modules are synthesized with degenerate codons.

Like other synthetic binding scaffolds, DARPins have been increasingly used for therapeutic applications as well as biosensor development (112). In one study, DARPins that specifically bind to the mitogen-activated protein kinase (MAPK) extracellular signal-regulated kinase (ERK) in either inactive (unphosphorylated) or active (phosphorylated) form were labeled with a solvatochromic merocyanine dye to become a fluorescent biosensor (113). This merocyanine dye has stronger fluorescence when exposed to a more hydrophobic environment. Therefore, upon DARPins binding to ERKs in active/inactive forms, the dye emits stronger fluorescence and thus report the presence of ERKs.in respective forms.

Monobodies are synthetic binders derived from the fibronectin type III domain (127). In the original library, the three variable loops of monobodies are located at the N-terminus of the protein, resembling the structure of nanobodies (128). Alternatively, amino acids on the beta-sheet surface can be diversified for binding, resulting in what is called a "side and loop" library (129). Like nanobodies, monobodies are compact in size, but unlike nanobodies, monobody scaffolds lack disulfide bonds. This makes them less

rigid than nanobodies and more compatible with oxidizing environments, such as those found within cytoplasm.

Monobodies, like nanobodies, can be engineered into FRET-based fluorescent biosensors. For example, a monobody targeting R-phycoerythrin was fused with CFP and an enzyme-specific peptide to create a biosensor that probes membrane-type-1 matrix metalloproteinase activity (130). Monobodies can also be engineered to target specific protein conformations and serve as biosensors for monitoring these conformational changes. For instance, a monobody engineered to bind the activated conformation of Src-family kinases was fused with merocyanine fluorescent dye and a fluorescent protein (131). The solvent-sensitive merocyanine dye was optimized to emit increased fluorescence when in proximity to the SH3 domain of the target protein. As a result, when the monobody binds to Src-family kinases, the conjugated merocyanine dye's fluorescence intensity increases.

1.4.2. Synthetic protein binder combinatorial library

Although synthetic binders can be generated by computational design, a more common, accessible, and generalizable method for producing synthetic binders is through combinatorial library screening powered by recombinant DNA technologies.

Combinatorial libraries were originally developed as large collections of compounds with high structural diversity, first applied to generate peptide libraries (132). Shortly thereafter, combinatorial libraries of antibodies were created, leading to the formal introduction of the "combinatorial library" concept (133-135). Since then, countless combinatorial libraries have been developed for antibodies and other protein binders, including ScFv (136), monobodies (127), and nanobodies (137-139).

Combinatorial libraries are powerful tools for generating protein binders, as they theoretically provide binders with unlimited diversity, significantly increasing the likelihood of finding a protein binder for a specific ligand (140). This high diversity is achieved by varying the amino acid compositions of the variable regions within a protein binder scaffold. For example, in single-domain antibodies (nanobodies), the scaffold typically contains three complementarity-determining regions (CDRs) with variable amino acid sequences, potentially generating around 20^{20} to 20^{30} unique nanobody sequences. However, such

astronomical numbers exceed the experimental capabilities of current technology, so protein binder combinatorial libraries are usually refined through rational design. This means that amino acid distributions at each position are carefully controlled, often mimicking natural diversity while tuning properties such as hydrophobicity to prevent aggregation.

A common method for synthesizing combinatorial libraries or introducing randomized amino acids into scaffold genes is saturation mutagenesis (141-144). Saturation mutagenesis incorporates degenerate codons, such as NNK, into regions of interest to diversify the amino acids encoded. NNK codons are trinucleotide cassettes (145) that encode all 20 amino acids and the stop codon, though they reduce overall codon diversity by representing only 32 of the 64 possible codons. Synthetic DNAs are typically produced via PCR amplification of sequences containing degenerate codons at selected positions, followed by assembly of the PCR fragments into expression vectors. Once the library is synthesized, it undergoes high-throughput sequencing to assess diversity and quality, and then proceeds to display and screening.

1.4.3. Combinatorial library display technologies

The purpose of library display technologies is to establish a direct link between the genotypes in combinatorial libraries and the phenotypes that can be screened. Not only must the protein of interest be expressed and presented in a manner that allows for screening under relevant conditions, but the DNA sequence encoding the protein must also be physically connected to the protein, enabling researchers to trace back to the corresponding genetic information. To achieve this, various display technologies have been developed using different expression systems (Figure 7), like bacteriophage-based (146), cell-based including bacteria (147) and yeast (148), and cell-free ones such as ribosome display (149). Each expression system has its own advantages and disadvantages.

In phage display, exogenous proteins (or peptides) are expressed on the surface of phage particles and screened based on various physical parameters, such as binding affinity to target ligands. To display the protein on the surface, the exogenous protein is typically fused to one of the phage coat proteins. As the phage assembles, the fusion protein becomes part of the coat, and thus, is displayed on the surface of the phage particle.

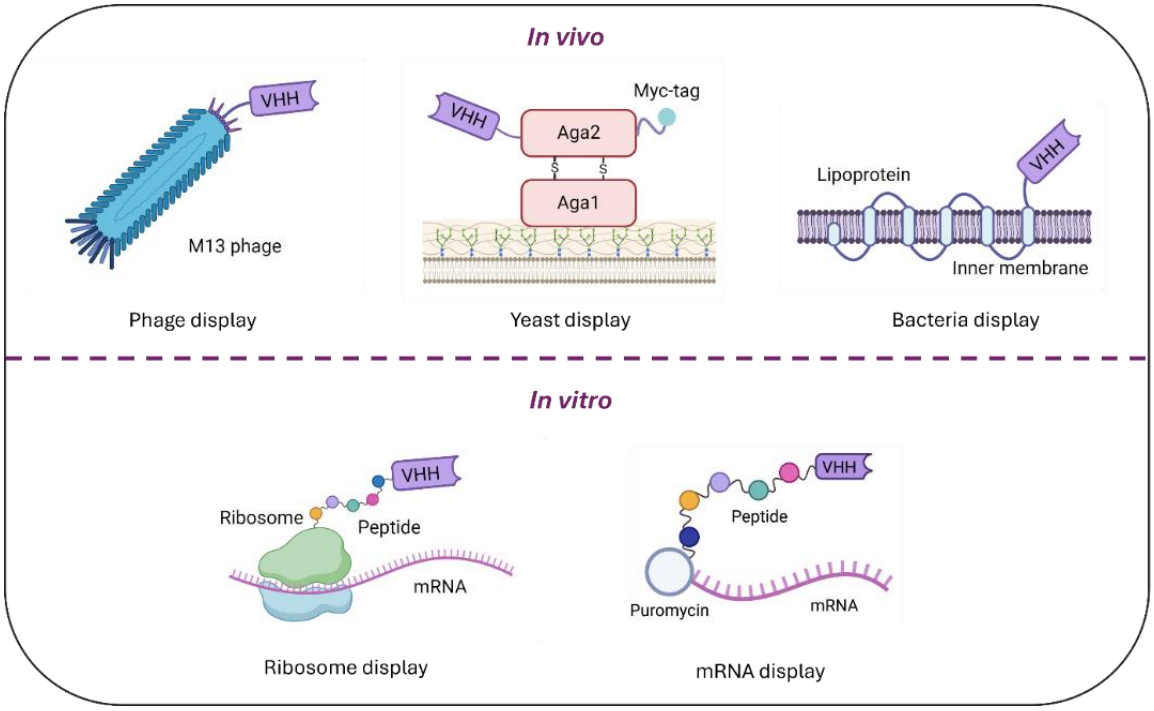


Figure 7. Schematics of in-vivo and in-vitro display technologies.

Filamentous phages offer several advantages as hosts for phage display. They tolerate gene insertions in non-essential regions, remain infective even with engineered coat proteins, and are stable across a wide range of temperatures and pH levels (150). Among filamentous phages, M13 is the most used. Several of its coat proteins have been utilized for fusion purposes, with pIII and pVIII being the most common choices. pIII, approximately 400 amino acids long, plays a role in host infection, while pVIII, at around 50 amino acids, forms the body of the phage. Since pIII is more tolerant to larger insertions and has fewer copies (only five), it is the preferred choice for protein fusions.

Even though filamentous phages are tolerant, the displayed protein can still interfere with coat protein function, especially when overcrowding occurs. To prevent polyvalency (the display of too many copies of the protein), phagemids are often used. A phagemid is a viral plasmid that encodes only the coat protein fusion and contains a packaging signal, while the other components needed for phage assembly, including the wild-type coat protein, are provided by helper phages. Infecting host bacteria containing

phagemids with helper phages produces phage particles with only one or a few copies of the fusion coat protein, thereby preventing polyvalency.

Due to high titer of filamentous phages and high transformation efficiency of bacterial hosts, phage libraries can achieve very high diversity in relatively small volumes, allowing screening of more than 10^{10} variants at a time. However, because the coat protein folds in the periplasm of bacterial hosts, the displayed proteins may not fold or function in the same way as they would in other environments. This is particularly true for proteins with disulfide bonds, as the reducing environment of the cytoplasm can disrupt these bonds, while the periplasm is more favorable for proper disulfide bond formation.

In yeast display, exogenous proteins are displayed on the cell surface by fusing them with glycosylphosphatidylinositol (GPI)-anchored proteins, which contain hydrophobic peptides at their C-termini that embed in the lipid bilayer of the cell membrane.

The most used GPI-anchored protein for yeast display is the *a*-agglutinin mating protein Aga2p subunit. The exogenous protein is fused to Aga2p at either the N- or C-terminus, and Aga2p is linked to the GPI-anchored Aga1p subunit via disulfide bonds. This arrangement displays the exogenous protein on the surface of the yeast cell. To create a yeast display library, plasmids expressing the Aga2p fusion with the proteins of interest are prepared and then transformed into yeast hosts lacking Aga2p. As a result, the fusion protein is expressed and displayed on the yeast surface.

Unlike phage display, where the protein of interest is displayed in only a few copies per phage particle, yeast display can present up to 10^5 copies of the protein per cell (151). This high copy number can introduce variability in protein expression levels, which may interfere with library screening. To address this, yeast display systems often include epitope tags that allow for quantification and normalization of protein expression levels per cell.

Although yeast libraries cannot achieve the same level of diversity and screening throughput as phage libraries due to the lower transformation efficiency of yeast cells, yeast display offers distinct advantages. As a eukaryotic system, it provides post-translational modifications that are unavailable in phage display and is capable of displaying larger, more complex proteins, including those containing disulfide bonds.

Similar to yeast display, bacterial display is a cell-surface display method in which proteins are fused to surface-anchoring motifs that act as carriers. The most used host for bacterial display is *E. coli*, a Gram-negative bacterium known for its high transformation efficiency and availability of extensive genetic manipulation tools. A variety of outer membrane proteins serve as carriers in this system.

For antibody display, the outer membrane presents a significant challenge. One common solution is anchored periplasmic expression, where single-chain antibodies are anchored to the inner membrane via lipoproteins (153). Additionally, various other display systems and anchor proteins have been developed, such as OmpA, FhuA, and CPX, to facilitate bacterial surface display (152). Despite the higher transformation efficiency of *E. coli*, bacterial display systems, like yeast display, are limited by the throughput of available screening methods.

Compared with yeast display, bacterial display has some disadvantages. Yeast display is generally considered safer for pharmaceutical and food applications. Moreover, as eukaryotic cells, yeast cells more closely resemble mammalian cells in terms of protein expression and secretion machinery, making them more suitable for displaying mammalian proteins or proteins intended for mammalian cellular environments.

Ribosome display is a cell-free display system. Unlike cell-based systems, where cells expressing proteins naturally contain corresponding DNA sequences, ribosome display links phenotypes and genotypes through *in vitro* translation of mRNAs lacking stop codons. Normally, release factors bind to stop codons at the end of translation, catalyzing the release of polypeptide chains from ribosomes. However, in the absence of stop codons, polypeptides remain trapped in the ribosome. Additionally, without ribosome recycling factors, the mRNA stays attached to the ribosome (154). This setup establishes a connection between the protein and its encoding gene.

While the concept of ribosome display is straightforward, it requires extra precautions to avoid certain issues. For example, an unstructured peptide spacer must be added to the C-terminus of the protein of interest to allow the polypeptide to fold properly while still attached to the ribosome. Furthermore, folding-assisting proteins, such as chaperones, are often included in the *in vitro* translation mixture. If the protein

of interest contains disulfide bonds, enzymes such as protein disulfide isomerase should be added to catalyze bond formation.

A key advantage of ribosome display over other cell-based display technologies is that it bypasses the limitations of library size due to host cell transformation efficiency, a main challenge faced in yeast display. Ribosome display can screen libraries with over 10^{10} variants, comparable to the capacity of phage display. Additionally, ribosome display eliminates the need for time-consuming cell culture preparation, which can take several days. Another unexpected benefit of ribosome display is that proteins in ternary complexes appear to be less aggregation prone. However, to ensure display efficiency, the amount of active ribosomes must be adjusted according to the desired library size.

One major issue with ribosome display is the low stability of the mRNA-ribosome-protein complex, making library screening more challenging. Moreover, mRNA is more prone to nuclease degradation and less stable than DNA. For these reasons, ribosome display is not as robust or widely used as phage display.

Like ribosome display, mRNA display is also a cell-free display system (173). A major difference between mRNA display and ribosome display is that in mRNA display proteins are covalently linked to mRNA molecules. This is usually achieved by modifying the 3' end of mRNA molecules with puromycin. Puromycin is an antibiotic structurally like an amino-acylated tRNA; when puromycin enters the A site of ribosome, it forms a peptide bond with the peptide chain, thus covalently linking mRNAs to proteins. The resulting mRNA-ribosome complexes can be screened, and mRNAs can be readily amplified for library enrichment or sequence analysis. Like ribosome display, mRNA display benefit from advantages of *in vitro* expression systems, such as high library diversity, bypassing transformation efficiency limits and surface display restrictions. On the other hand, *in vitro* translation frequently results in a certain fraction of translated molecules to be incomplete, which limits enrichment factor, and mRNA as a naturally unstable molecule is prone to degradation. Additionally, it has been suggested that applying mRNA-display to robust folding proteins might encounter issues with protein-mRNA complex purification (149). Strategies have been adopted to overcome certain limitations. For example, in some studies, a branched DNA nucleotide is ligated to mRNA and acts as a primer for the reverse transcription of cDNA (174).

1.4.4. Combinatorial library screening methods

The choice of library screening methods depends on which display technology is used. Biopanning is used for phage display, and cell sorting is used for cell surface display technologies. Biopanning is an affinity-based screening method used to screen phage display libraries (Figure 8a). The process is generally divided into four major steps: immobilization, binding, washing, and elution (155).

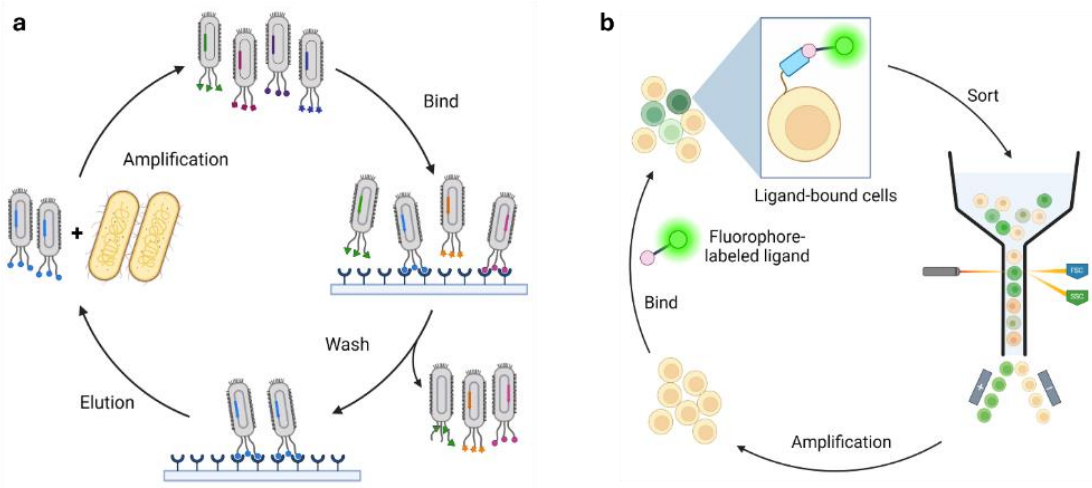


Figure 7. Schematic of common library screening methods. (a) Schematic of biopanning; (b) Schematic of fluorescence activated cell sorting.

In the immobilization step, the ligand is fixed onto a solid surface, often by incubating biotin-conjugated ligands with streptavidin-coated plastic wells or magnetic beads. This solid surface is referred to as the solid phase. During the binding step, the phage library is mixed with the solid phase, where phages displaying strong binders tightly adhere to the surface, while those displaying weak binders do not. In the wash step, unbound or weakly bound phages are flushed off the solid phase, leaving only the strongly bound phages. In the elution step, the strong, non-covalent interactions between the phage (or more precisely, the protein it displays) and the solid phase are disrupted, typically using acidic or basic solutions, or alternatively by enzymatic cleavage. Once eluted, phages are pooled together and used to infect bacteria to generate a new library. In theory, this new library contains more strong binders than the previous one, with the population of desired protein binders enriched. The biopanning process can be

repeated over several rounds to further enrich the library, until its diversity is reduced enough and satisfactory binders are identified.

To assess the binding affinity of individual binders, bacteria clones can be isolated, phage particles produced, and enzyme-linked immunosorbent assay (ELISA) performed. If strong binders are identified, corresponding DNA from bacteria can be amplified by PCR and sequenced.

Each step of biopanning can be adjusted to meet specific screening goals. For example, to eliminate non-specific binders, one can perform negative selection by using unwanted targets during the immobilization step. In this case, phages that remain tightly bound to the solid phase are likely displaying non-specific binders, while those easily washed off can be screened against the real ligand of interest. To increase the stringency of screening and improve the likelihood of finding strong binders, one can apply harsher wash conditions, such as increasing detergent concentration in the wash buffer.

Biopanning is a high-throughput screening method that, when combined with phage display, enables efficient screening of libraries with over 10^{10} variants in relatively short time. Unless overnight incubation is required, one round of screening can typically be completed in a single day. However, biopanning has certain limitations. First, the ligand must be conjugated to a molecule that can attach to the solid surface, such as biotin or bovine serum albumin (BSA). Conjugation relies on specific chemical groups in the ligand, which can sometimes be difficult to implement. Additionally, the presence of a conjugation molecule increases the risk of non-specific or false-positive binders. This issue is particularly pronounced when using BSA as the conjugation molecule, as it is much larger than most short peptides or small molecules and can become a target for off-target binding. Non-specific binders with low affinity are difficult to remove and can create a noisy background, especially when strong binders are scarce. Moreover, steric hindrance from the solid surface can restrict the approach of binders to immobilized ligands, potentially limiting their affinity and binding orientation.

Cell sorting methods can be used to isolate rare cell populations for research and to screen cell-based display libraries. For yeast display, it is common to combine two cell sorting techniques: magnetic-activated cell sorting (MACS) and fluorescence-activated cell sorting (FACS). MACS is similar to the biopanning, where ligands are immobilized onto magnetic beads. Cells displaying protein binders are

incubated with magnetic beads, washed, and eluted based on how tightly they bind to the beads (156). However, because the number of protein copies on each yeast cell's surface is highly heterogeneous, cells with more protein copies may bind more tightly due to avidity. As a result, MACS can only isolate cells with binders that attach to the ligand but does not necessarily identify strong binders. It is more commonly used as an initial screen to reduce the library's diversity.

FACS is often used to further separate cells based on binding affinity (Figure 8b). It is an automated cell sorting method that sorts cells into bins based on fluorescence intensity. Yeast display libraries typically use epitope tags attached to the displayed protein, which can be labeled with antibodies conjugated to fluorescent dyes, and the fluorescence intensity of cells will reflect the amount of protein displayed on each cell. In FACS, cells are labeled with these antibodies, then incubated with fluorescently labeled ligands and washed. The instrument processes each cell, measuring the fluorescence intensity of both the ligand and the antibody. The ratio of ligand to antibody fluorescence provides an indication of binding affinity: the more ligand bound and the less protein displayed, the higher the ratio, indicating stronger binding. By setting an appropriate threshold, cells displaying strong binders can be isolated, used to grow a more enriched library, and the screening process can be repeated over multiple rounds as needed. Binding affinity can be conveniently assessed by flow cytometry (157), and corresponding DNA stored in yeast cells can be readily amplified by PCR and sequenced.

One advantage of yeast display combined with cell sorting is the ability to perform very stringent negative selection. Due to high copy number of displayed proteins, the binding affinity of binders is amplified through avidity. This means that non-specific binders, even those with low affinity, can bind tightly to non-specific targets and be efficiently eliminated from the library. However, a major limitation of cell sorting and yeast display is the lower throughput compared to biopanning and phage display. Modern FACS instruments process approximately 10^7 to 10^8 events per hour, and yeast libraries are generally several orders of magnitude smaller than phage libraries.

1.5. Synthetic chemically induced dimers as GEFS sensory domains

The library preparation, display, and screening methods described for generating a single synthetic binder can also be applied to the creation of synthetic chemically induced dimerization (CID) pairs, as CIDs

consist of two synthetic binders. However, since a CID pair must only associate in the presence of a target ligand, its generation requires a specially designed methodology. COMBINES-CID is one such method that generates synthetic CIDs through combinatorial library screening in a sequential manner (158) (Figure 9).

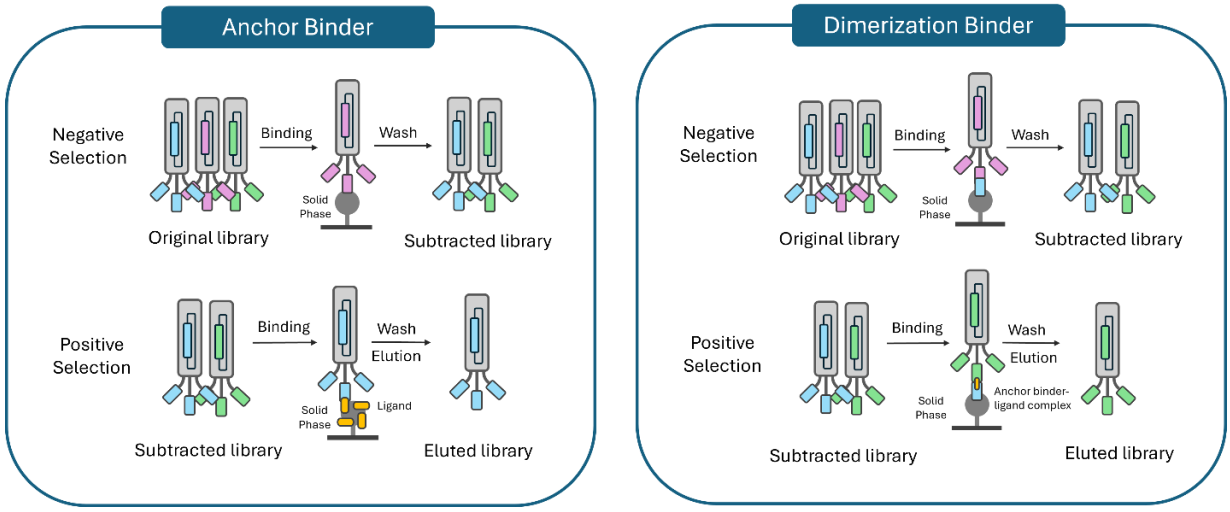


Figure 8. Schematic of COMBINES-CID. Combinatorial libraries of synthetic binders are displayed by phage. Biopanning involves a combination of positive and negative control. Selection of CID pairs is carried out in a sequential manner.

In COMBINES-CID, synthetic DNA libraries of nanobodies and monobodies are created with an estimated diversity of over 10^9. *E. coli* are then transformed with these DNA libraries to produce phage libraries, which are then screened using biopanning. To generate CIDs, COMBINES-CID selects two sets of synthetic binders in a sequential process. First, it selects synthetic binders that bind the target ligand with high affinity and specificity, known as anchor binders. These serve as the anchor for initiating dimerization. Second, it selects synthetic binders, called dimerization binders, which bind specifically to the anchor binder-ligand complex but not to the anchor binder alone. These complete the dimerization event.

For anchor binder selection, target ligands are immobilized onto either magnetic beads or plastic well plates, typically using biotin-streptavidin interactions or passive BSA absorption. A negative-selection step

is performed before positive selection. Depending on the positive-selection method, either magnetic beads or empty plastic well plates (without ligands) are used as the "fake" target. Phage libraries are incubated with the fake target, and the wash is collected to create a subtracted library. This library is then incubated with the real target, followed by washing and elution. The screened library is used to infect bacteria, generating an enriched library, which can undergo further rounds of screening as needed.

To evaluate the level of enrichment, a pool ELISA is performed. In this step, the phage library is incubated with ligand-coated or empty wells, and anti-phage antibodies are used to quantify the number of phage particles bound to the ligand compared to the background. If enrichment is sufficient, individual bacteria clones are isolated, and phage particles with single, unique DNA sequences are produced. These phage particles are analyzed for binding affinity using titration ELISA. Those with strong signals have their DNA amplified and sequenced. The DNA sequences of promising candidates are then expressed in bacteria and purified for binding affinity quantification. Binders with desirable traits are retained for the dimerization binder selection process.

The selection of dimerization binders follows a similar process as that of anchor binders, with the key difference occurring in the immobilization step. Instead of immobilizing target ligands, the anchor binder-ligand complex is used as bait. If multiple anchor binders are available, they are mixed in equal molar concentrations. For negative selection, unbound anchor binders are used to remove binders that can bind to anchor binders in the absence of ligands. Pool ELISA and titration ELISA are performed as with the anchor binders, using the anchor binder-ligand complex as bait. Promising dimerization binders are then identified through their isolated DNA sequences and expressed for further characterization. The ideal dimerization binder should bind the anchor binder-ligand complex with high affinity and specificity, without binding to the unoccupied anchor binder in the absence of the ligand.

Through this process, COMBINES-CID enables the selection of synthetic CID pairs from combinatorial libraries, ensuring that dimerization is tightly controlled by the ligand of choice.

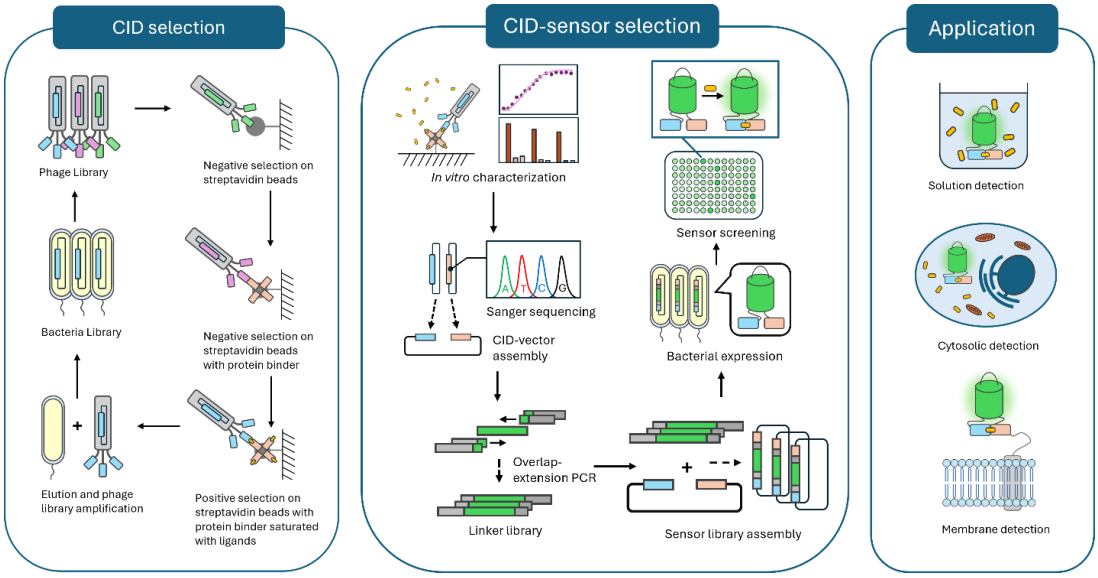


Figure 9. Overview of CID-based genetically encoded fluorescent sensor engineering. CID pairs are selected by phage display and phage biopanning. After sequencing and in vitro characterization, a linker library of cpGFP flanked by various length of GS linkers is inserted between CID pairs to generate a sensor library. This library is further screened to identify sensors variants with desirable performance. CID-based sensors engineered by this method can be applied to solution, subcellular locations and cell surface.

Our synthetic CID-based GEFS engineering pipeline comprises two key phases: CID selection and CID-sensor selection (Figure 10). In the first phase, CID selection, we employed an in-house technique, COMBINES-CID, to generate synthetic CIDs for a specific ligand. This method uses combinatorial libraries—with more than 10^9 unique protein binders—and leverages sequential selection via phage display and biopanning (172). The process begins by selecting anchor binders from the library, using the target ligand as bait. These anchor binders demonstrate both high affinity and specificity for the ligand. Subsequently, dimerization binders are selected, targeting the anchor binder-ligand complex as bait. These binders interact exclusively with the anchor binder-ligand complex, without binding to the unoccupied anchor binders. To ensure precision at each stage, we implemented negative selection protocols to remove non-specific binders. For anchor binder selection, streptavidin beads were used as the negative bait, while for dimerization binder selection, the unoccupied anchor binders were the negative control.

After successful CID selection, we moved on to the CID-sensor selection phase. We created a CID-sensor library by introducing a linker library between the selected synthetic CIDs. This library was

subsequently screened in 96-deep well plates to identify sensors with optimal performance. The top-performing sensor variants were then characterized *in vitro* and localized within HEK293T cells to validate functionality. This streamlined pipeline provides a reliable strategy for integrating synthetic CIDs into GEFS, facilitating applications both *in vitro* and within cellular environments.

2. CID Generation using COMBINES-CID

2.1. Introduction

GEFSs using synthetic CIDs as sensory domains offer several advantages. First, synthetic CIDs are not constrained by naturally available proteins, allowing for a broader range of target interactions. Furthermore, because CIDs rely on dimerization, integrating them into GEFSs does not require extensive screening to optimize the insertion site, reducing the need for multiple resolved structures. Additionally, synthetic CIDs generated from combinatorial libraries can provide various binding affinities and behaviors, easily adaptable for specific applications.

The initial step in engineering CID-based GEFSs is generating synthetic CIDs, which can be achieved through a specialized phage biopanning method known as COMBINES-CID. This approach leverages robust combinatorial libraries for effective CID generation. Building on our earlier work with a nanobody-based library, we now present a synthetic library based on the monobody scaffold. While monobodies, like nanobodies, contain three variable loops, they lack disulfide bonds, potentially enhancing folding robustness, though at the cost of some structural rigidity. This characteristic makes monobodies a valuable complement to our existing nanobody library. To improve the sequence correct rate and folding efficiency of the library, we adopted His-purification and CAT screening to optimize the library diversity.

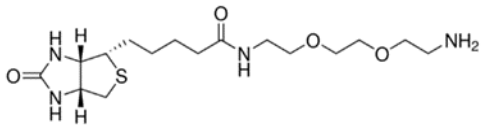
To validate the monobody-scaffold library and develop sensory domains for GEFSs, we used COMBINES-CID to select synthetic CID pairs specifically targeting metorphamide. We characterized the binding affinity of these CID pairs *in vitro*, assessing association affinity both in the presence and absence of the ligand to confirm that ligand binding modulates CID interactions.

Additionally, we stabilized a nanobody-based CID pair by introducing mutations at specific scaffold positions while preserving disulfide bonds. We then tested the expression of these stabilized variants when fused with cpGFP.

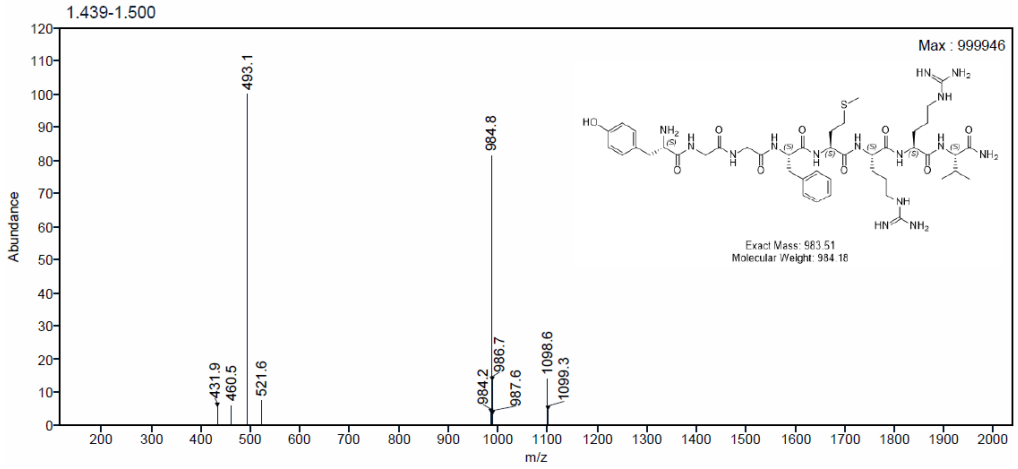
2.2. Material & Methods

2.2.1. Chemical Synthesis of Ligand

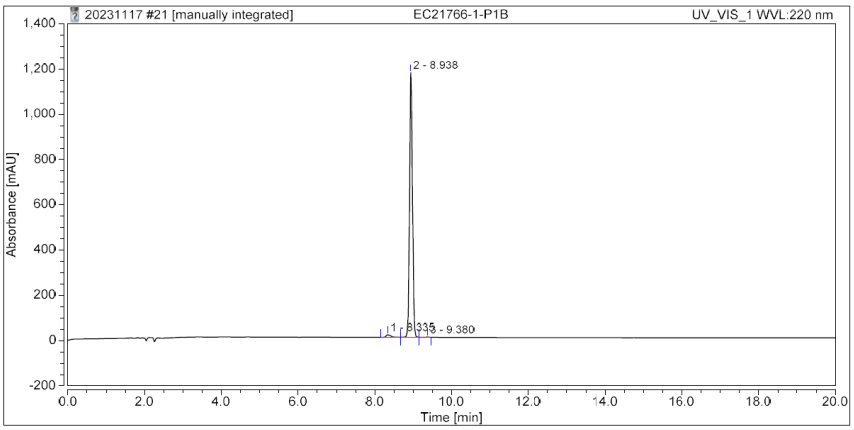
Biotinylated MP was synthesized by Wuxi AppTec. The sequence is YGGFMRRV-linker-biotin. The linker-biotin has the following structure:



Metorphamide MS profile:



Metorphamide-biotin HPLC purification profile:



2.2.2. Pre-library DNA synthesis

The preliminary DNA library was designed using a universal monobody scaffold with randomized sequences in three variable loops: BC (6 amino acids), DE (4 amino acids), and FG (7, 10, or 13 amino acids). NNK codons were incorporated to randomize the sequences in these loops. The monobody gene was split into three fragments: fragment 1 (loop BC), fragment 2 (loop DE, using MNN codons), and fragment 3 (loop FG, with NNK codons). All fragments were ordered from Integrated DNA Technologies and assembled into a complete monobody gene via overlap-extension PCR.

For assembly, the fragments were mixed and incubated with T4 DNA polymerase at room temperature for 30 minutes, followed by PCR amplification with Q5 polymerase and primers (0.5 μM). The PCR conditions were: initial denaturation at 98˚C for 30 seconds, 12 cycles of 98˚C for 10 seconds, 70˚C for 10 seconds, 72˚C for 10 seconds, and a final extension at 72˚C for 5 minutes. Approximately 3 μg of purified PCR products and 10 μg of pADL-23c phagemid vector were digested with *Bgl*I for 1 hour at 37˚C, followed by T4 DNA ligation overnight. The ligated products were purified with the Purelink PCR Quick Purification Kit and transformed into *E. coli* TG1 cells by electroporation (1,800 V, 10 μF, 600 Ω). Cells were recovered, incubated in 2×YT with 2% glucose, and plated on 2×YT-ampicillin-glucose agar plates. Colonies were grown overnight at 30˚C, scraped into 2×TY, and the library was subjected to three rounds of His-tag purification and two rounds of CAT screening. Next-generation sequencing was then used to analyze amino acid distribution in the variable regions.

2.2.3. Pre-library enrichment by His-purification

Purified phage particles were incubated with Ni-NTA agarose for 1 hour at room temperature, then packed into a column. The column was washed with sodium phosphate buffer (50 mM Na_2HPO_4, 300 mM NaCl, 50 mM imidazole, 5% glycerol, pH 7.4) and eluted using a gradient of elution buffer (50 mM Na_2HPO_4, 300 mM NaCl, 500 mM imidazole, 5% glycerol, pH 7.4) mixed with varying percentages of wash buffer. The eluted phage particles were then used to transfect TG1 cells, generating a new bacterial library.

2.2.4. Pre-library screening with CAT assay

Preliminary library-transformed TG1 cells were inoculated into 2xYT medium with 100 μg/mL ampicillin and 0.2% glucose, then grown at 37°C until OD_{600} reached ~0.6. Monobody-CAT expression was induced

with 0.3 mM IPTG for 4 hours, after which the bacteria were plated on 2xYT-agar plates containing 0.1 mM IPTG and increasing chloramphenicol concentrations (34, 68, 136, 272, and 500 µg/mL). Plates were incubated overnight at 30°C. The following day, clones were scraped from the plates, pooled, and prepared for codon distribution analysis with next-generation sequencing.

2.2.5. Pre-library diversity analysis with next-generation sequencing

Phagemid DNAs were isolated from purified monobody library phage particles using the QIAprep Spin M13 kit (Qiagen) for sequencing. DNA concentrations were measured with a Nanodrop 2000 spectrophotometer (Thermo Fisher Scientific). A two-step PCR (10 cycles) was used to introduce Illumina adaptors and 8-bp unique molecular identifiers (UMIs) to both ends of the amplicons. The library was sequenced on an Illumina NextSeq platform using a 2×150 bp high-output kit.

Sequencing data were processed with Flexbar (170) to trim UMI sequences and reads with a quality score below 20 were filtered out. Errors within UMIs were tolerated when retrieving variable loop sequences, allowing for undetermined bases or shorter barcodes. Clean reads were aligned to the scaffold sequences adjacent to the variable loops, extracting loop sequences with a 0.1 mismatch rate. Paired-end reads were merged to produce the three variable loop sequences. Only in-frame sequences without undetermined bases or stop codons were retained. A custom Perl script was used for protein sequence analysis, and sequence logos were generated using WebLogo3.6 (171).

2.2.6. Library construction

After determining the optimal amino acid distribution, the final combinatorial DNA library was synthesized using trinucleotide mutagenesis (Thermo Fisher Scientific). Fragments were amplified by 10 cycles of 100 PCR reactions with Q5 DNA polymerase, 0.5 µM primers, 30 ng template. The PCR product was purified with the Purelink PCR Quick Purification Kit. Digested vector (24 µg) and inserts (7 µg) were ligated with T4 DNA ligase at 4˚C overnight. The ligated products were purified with the Purelink PCR Quick Purification Kit and transformed into *E. coli* TG1 cells by electroporation (1,800 V, 10 µF, 600 Ω). Cells were recovered, incubated in 2×YT with 2% glucose, and plated on 2×YT-ampicillin-glucose agar plates. The library (17 µg) was transformed into TG1 cells via 110 electroporation reactions, generating approximately 8.45×10^{10} clones, plated on 618 plates. The cells were scraped into 2×TY and stored in

25% glycerol at -80˚C. Phage library was produced from bacteria library, with estimated concentration of $0.5x10^{14}$/mL.

2.2.7. Synthetic CID selection

To prepare the phage library, the bacterial stock of the monobody library was diluted in 2×YT medium with 2% glucose and 100 μg/mL ampicillin to an OD_{600} of ~0.1. The culture was grown at 37°C until OD_{600} reached ~0.5, then superinfected with CM13 helper phage at $5×10^9$ pfu/mL for 1 hour. Afterward, the cells were pelleted, resuspended in fresh 2×YT with 100 μg/mL ampicillin and 50 μg/mL kanamycin, and incubated overnight at 25°C. Phage particles were purified by centrifugation (5,000×g, 4°C, 30 minutes) to remove cells, followed by precipitation of the supernatant with PEG/NaCl solution (20% w/v PEG 6,000 and 2.5 M NaCl). The mixture was incubated on ice for 1 hour, centrifuged (12,000×g, 4°C, 30 minutes), and the phage pellets were resuspended in 1×PBS for short-term storage at 4°C or with 25% glycerol at -80°C for long-term storage.

Anchor binders were selected using streptavidin magnetic beads bound to biotin and biotinylated metorphamide (MP) for negative and positive selections, respectively. Briefly, 300 μL of 10 μM biotin or biotinylated MP was captured by 300 μL of streptavidin-coated magnetic beads (Dynabeads M-280, Thermo Fisher) and blocked with 1% casein in 1×PBS (pH 7.4) for 1 hour. During each round, the phage-displayed nanobody library was incubated with biotin-bound beads for 1 hour to remove off-target binders, followed by incubation with MP-bound beads. After washing the beads 10 times with 0.05% PBST, phage particles were eluted using 100, 50, and 10 μM MP in rounds one through three, and 1 μM MP in rounds four through six. After six rounds of biopanning, colonies were picked and validated by phage ELISA, then sequenced.

Dimerization binders were selected using unbound oAS28/oAS14 for negative selection and MP-bound oAS28/oAS14 for positive selection. For the bait, 600 μL of 1 μM biotinylated oAS28/oAS14 was incubated with 600 μL streptavidin beads and blocked with 1% casein in 1×PBS (pH 7.4) for 1 hour. The beads were split 2:1 for negative and positive selections. For positive selection, oAS28/oAS14-loaded beads were incubated with 50, 10, 2, and 1 μM MP in rounds one through four. The biopanning method

was similar to anchor binder selection, except phage was eluted with 100 mM triethylamine. After four rounds, single colonies were picked and evaluated by phage ELISA.

2.2.8. Phage ELISA

ELISA plates were prepared by coating 96-well plates (Nunc MaxiSorp, Thermo Fisher Scientific) with 100 µL of 5 µg/mL streptavidin in 100 mM carbonate buffer (pH 8.6) and incubating overnight at 4˚C. Before use, plates were washed three times with 0.05% PBST (1×PBS with 0.05% Tween 20) to remove residual streptavidin.

For single-phage ELISA, individual clones were inoculated into 250 µL of 2×YT medium containing 2% glucose and 100 µg/mL ampicillin in 96 deep-well plates (Thermo Fisher Scientific) and incubated overnight at 37˚C. A 10 µL aliquot from each culture was transferred into 1 mL of fresh medium and grown until OD_{600} reached ~0.5. The cultures were then infected with CM13 helper phage at a multiplicity of infection (MOI) of ~18 and incubated at 37˚C for 1 hour. After infection, 50 µg/mL kanamycin was added, and cultures were incubated overnight at 25˚C. The next day, the plates were centrifuged at 3,000×g for 10 minutes, and the supernatants were transferred to fresh plates for ELISA.

For anchor-binder phage ELISA, each well was coated with 100 µL of 1 µM biotinylated MP and incubated at room temperature for 1 hour. After five washes with 0.05% PBST, wells were blocked with 1% casein in PBS. Then, 100 µL of phage supernatant was added to each well and incubated for 1 hour. Afterward, the wells were washed 10 to 15 times with 0.05% PBST and incubated with 100 µL of HRP-conjugated anti-M13 major coat protein antibody (RLph1, Santa Cruz Biotechnology; 1:10,000 in 1×PBS with 1% casein) for 1 hour. To develop the signal, 100 µL of 1-Step Ultra TMB ELISA substrate (Thermo Fisher Scientific) was added and incubated at 37˚C for 5 minutes, followed by 100 µL of Tris-HCl (pH 8.0) to stop the reaction. OD_{450} was measured using a SpectraMax Plus 384 microplate reader (Molecular Devices).

For dimerization-binder phage ELISA, the procedure was similar to the anchor-binder ELISA, except that the anchor binder-MP complex was used as bait. In brief, ~100 nM biotinylated anchor binder was added to streptavidin-coated plates, followed by incubation with varying MP concentrations in 1×PBS (pH 7.4).

For titration ELISA, ELISA plates were prepared as described in the “Phage ELISA” section. Wells were coated with biotinylated oAS28, washed three times with 0.05% PBST, and blocked with 1% casein in PBS for 1 hour. For each dimerization binder, 1 μM of phage particles displaying the binder were mixed with 0.1, 0.3, 1, 3, 10, 30, and 90 μM MP and added to the wells. After a 1-hour incubation at room temperature, the plates were washed 10 times with 0.05% PBST. Color development followed the same procedure as described in the “Phage ELISA” section.

2.2.9. Protein expression and purification

All monobodies were expressed with C-terminal Avi-tags and His-tags in *E. coli* and purified using Ni-NTA chromatography. Monobodies were cloned into the pET24b vector and transformed into *E. coli* C41(DE3). A single colony was inoculated into 1 L of 2×YT medium and grown at 37°C until OD_{600} reached ~0.6. Protein expression was induced with 0.3 μM IPTG, and the culture was incubated overnight at 16°C.

Cells were harvested by centrifugation at 8,000×g for 15 minutes, resuspended in 30 mL of binding buffer (10 mM sodium phosphate, 0.5 M NaCl, 20 mM imidazole, pH 7.4), and lysed by sonication (25% amplitude, 5-second on/off cycles for 5 minutes). The lysate was centrifuged at 20,000×g for 30 minutes at 4°C.

The supernatant was loaded onto a 1 mL HisTrap column (GE Healthcare) pre-equilibrated with washing buffer (50 mM sodium phosphate, pH 8.0, 300 mM NaCl, 50 mM imidazole, 10% glycerol). After washing, His-tagged proteins were eluted with a gradient of elution buffer (50 mM sodium phosphate, pH 8.0, 300 mM NaCl, 250 mM imidazole, 10% glycerol). Eluted proteins were analyzed by SDS-PAGE, quantified using a Nanodrop spectrophotometer (Thermo Fisher Scientific), flash-frozen in 500 μL aliquots, and stored at -80°C.

2.2.10. Protein biotinylation

Monobodies were biotinylated using two methods: during bacterial culture or after purification, both requiring an Avi-tag included in the modified pET24b vector. For in-culture biotinylation, monobody-expressing plasmids were transformed into a customized *E. coli* C41(DE3) strain co-expressing the BirA

enzyme. During protein induction, 50 μM biotin and IPTG were added to the media, allowing BirA to biotinylate the monobodies in the bacterial cytosol.

For post-purification biotinylation, monobodies were biotinylated *in vitro* using the BirA-500 kit (Avidity). A typical reaction included 200 μL of 10× BiomixA (0.5 M bicine buffer, pH 8.3), 200 μL of 10× BiomixB (100 mM ATP, 100 mM $Mg(OAc)_2$, 500 μM D-biotin), and 8 μL of 1 mg/mL BirA, added to 1.6 mL of ~1 mg/mL Avi-tagged monobodies (final volume of 900 μL). The reaction was incubated either overnight at 4°C or for 1 hour at room temperature. Post-biotinylation, proteins were purified using a HiPrep 26/10 desalting column (GE Healthcare) pre-equilibrated with storage buffer (1×PBS, 5% glycerol).

Biotinylation efficiency was assessed by mixing purified proteins with streptavidin and analyzing via SDS-PAGE. Proteins were quantified with a Nanodrop spectrophotometer (Thermo Fisher Scientific), flash-frozen in 500 μL aliquots, and stored at -80°C.

2.2.11. Bio-layer interferometry

Binding kinetics was analyzed using an Octet R8 system (ForteBio) and Streptavidin (SA) biosensors. 200 nM biotinylated dimerization binder were immobilized on SA biosensors with the binding assay buffer (1×PBS, pH 7.4, 0.05% Tween 20, 0.2% BSA) and then assayed with 1 μM anchor binder pre-equilibrated with serial dilutions of CBD. For CA14-DB21, the resulting binding curves were compared with the ones collected in previous research to get an estimated binding affinity. For oAS28-6 and oAS14-89, 1:1 global fitting was used to calculate binding affinity.

2.2.12. Isothermal calorimetry

Binding affinities and thermodynamics of oAS28 and oAS14 to MP were measured using a MicroCal PEAQ-ITC (Malvern) at 25°C. Approximately 210 μL of 20 μM oAS28/oAS14 was loaded into the sample cell, and 38 μL of 200 μM MP was titrated in 19 injections (2 μL each). Background heat transfer from dilution was measured by titrating PBS into the sample. Raw ITC data were analyzed using NITPIC version 1.2.7 to obtain the Kd.

2.2.13. Size-exclusion chromatography

Monobody CIDs were analyzed by size-exclusion chromatography (SEC) using a Superdex 75 Increase 10/300 GL column (GE Healthcare). For non-crosslinked samples, 500 µL of ~10 µM monobodies were injected into the column, pre-equilibrated with 1×PBS, and eluted at 0.5 mL/min at 4°C. The column was calibrated with molecular weight standards (Sigma-Aldrich).

Crosslinked samples were prepared by incubating 10 µM proteins, with or without MP at saturation, in 1×PBS containing 0.1 mM BS(PEG)5 for 30 minutes at room temperature. Then, 500 µL of the crosslinked samples were injected into the column for analysis.

2.3. Results

2.3.1. Combinatorial Library Design and Preparation

We aimed to create a monobody library complementary to the pre-existing nanobody library, choosing the monobody scaffold for its structural similarity to nanobodies but greater robustness in bacterial expression due to the absence of disulfide bonds. Developed by the Koide lab, the monobody scaffold is derived from the tenth domain of human fibronectin type III (127), and two main monobody library subtypes were created based on different variable region configurations. The first, the “loop only” library (Figure 1a), contains three variable regions confined to flexible loops at the N-terminal. The second, the “side and loop” library (Figure 1b), includes two loops and positions in the beta-sheet, forming a palm-like binding surface (129). Since our synthetic CIDs target small molecules and short peptides, we chose the “loop only” library, as the broader binding surface of the “side and loop” library would be less advantageous.

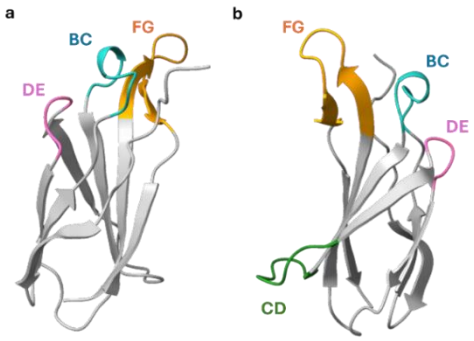


Figure 1. Schematics of monobody libraries. (a) "Loop" only library involves three variable loops, BC, DE, FG; (b) "Side and Loop" library involves four variable loops and beta sheet in-between.

To design our monobody library, we prioritized two factors: achieving high sequence accuracy without unexpected mutations and ensuring strong expression in *E. coli*. We began by creating a preliminary library (pre-library) with unbiased amino acid diversity in the variable loops (Figure 2a). Optimal amino acid composition was identified through His-tag purification and a CAT screening assay previously developed (139), guided by insights from other monobody library designs. His-purification removed clones lacking a His-tag due to folding inefficiencies or sequence errors. In the CAT screening, an HA-tagged chloramphenicol acetyltransferase (CAT) was fused to the C-terminus of the monobody. Only bacteria expressing correctly folded monobodies could produce functional CAT and survive in chloramphenicol-supplemented media, naturally enriching favorable amino acid distributions in the loops. After these enrichment steps, we analyzed the enriched pre-library diversity using next-generation sequencing (NGS) and synthesized an enhanced library reflecting this optimized diversity. We expected this approach to yield a library with high sequence accuracy and robust folding.

To construct the preliminary library, we introduced synthetic diversity into loops BC, DE, and FG using NNK degenerate codons. This library was intended to explore optimal diversity without controlling amino acid distribution in each loop: loop BC contained 7 NNKs, DE had 4, and FG had 7, 10, or 13 NNKs (Table 1). The synthetic DNA library was assembled via overlap extension PCR, creating an N-terminal fragment with loop BC and C-terminal fragments with loops DE or FG. These fragments were joined using T4 DNA polymerase and amplified by PCR. The assembled library was then inserted into the pADL-Avi-CAT phagemid vector, which contained His- and Avi-tags.

We performed three rounds of His-purification using NiNTA chromatography, followed by one round of CAT screening. In each round of His-purification, phage libraries generated from bacteria were purified with His-columns. Starting with 10^{13} phages in the first round, we recovered approximately 10^{11} phages (1% recovery rate). The second round, starting with 10^{11} phages, recovered about 0.6×10^{10} (27% recovery rate), which slightly increased to 30% in the third round. This pattern indicated that the first two rounds of His-purification substantially enriched the library, with slower enrichment in the third round.

During CAT screening, only plates with 34 µg/mL and 68 µg/mL chloramphenicol showed clone growth, suggesting further enrichment. We then proceeded to diversity analysis.

At each stage of His-purification and CAT screening, we analyzed the amino acid composition of the variable loops using NGS (Figure 3). Diversity did not shift significantly across the three rounds of His-purification (Figures 3a–3c), while CAT screening with 34 and 68 µg/mL chloramphenicol resulted in more substantial shifts in diversity (Figures 3d, 3e). We combined the enriched pre-library diversity with diversity insights from previous studies (129) to generate a customized diversity profile.

The resulting profile disfavored proline and cysteine, which is logical as proline is bulky and cysteine can form disulfide bonds. Loop BC contained 10% glycine, 15 % serine and tyrosine was strongly favored at 25%; this was a combined diversity based on our observation in CAT68 library (where glycine and serine were highly favored) and the design in previous study, where tyrosine had high distribution. Similarly, in loop DE, a limited set of amino acids, including glycine, serine, and tyrosine, was favored at specific positions; such bias was again supported by both enrichment in our screening and practice from previous studies. Loop FG showed a more even distribution, with tryptophan generally excluded except at position X12; this is because we did not observe any significant diversity bias in this loop, except for tryptophan had a low distribution. For the similar reason, we also excluded methionine at position X12 and X13.

We synthesized a new synthetic DNA library, termed the enhanced library, incorporating this optimized diversity. The DNA fragments were synthesized by trinucleotide mutagenesis, followed by 10 cycles of amplification. The resulting fragments were cloned into the pADL-23c phagemid vector and transformed into *E. coli* strain TG1 via 110 electroporation reactions, yielding approximately 8.45×10^{10} clones across 618 agar plates. Analysis of 96 clones revealed all to be unique sequences, with 77 containing correct sequences, resulting in an accuracy rate of 80%. We then proceeded to test the library by selecting synthetic CIDs from it.

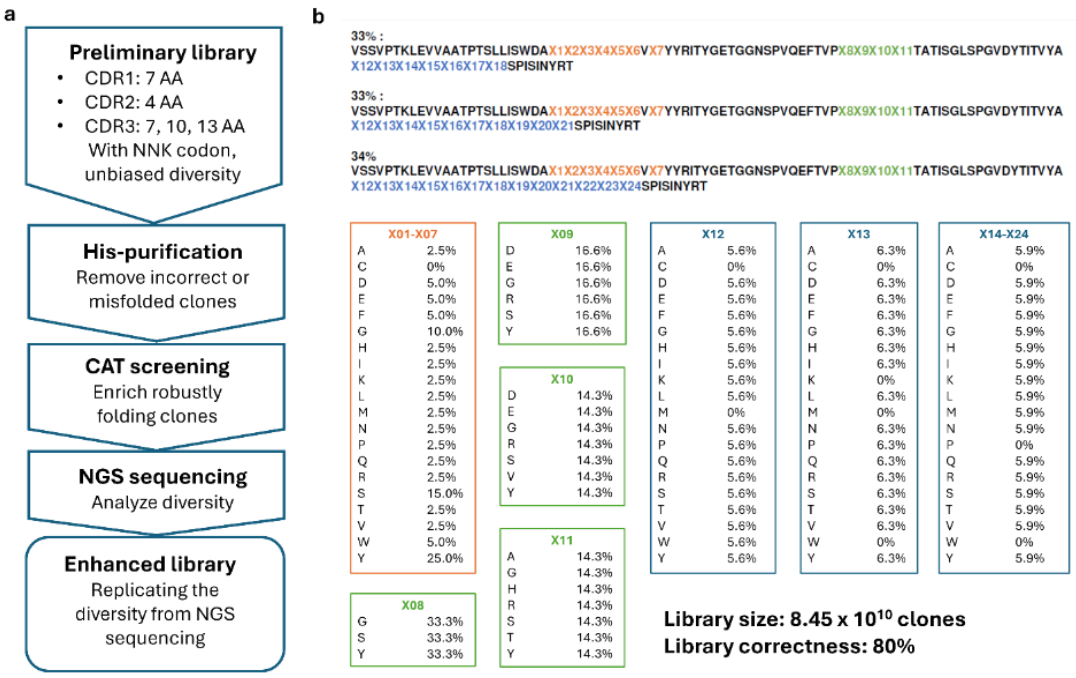


Figure 2. Overview of monobody library construction. (a) Workflow to synthesize a library with improved sequence correctness and folding robustness. A preliminary library was initially generated without any diversity bias. This library was subject to His-purification and CAT screening to enrich correct and robustly folding clones. The diversity of the enriched library was analyzed by NGS sequencing. A finalized, enhanced library was synthesized replicating the diversity found in the enriched pre-library. (b) Final diversity at each variable position in the enhanced library.

Fragments	DNA sequence
P1	ctcgcggcccagccggccGTATCTAGCGTACCGACTAAACTGGAAGTTGTTGCTGCTACC CCGACCTCTCTGCTGATCTCCTGGGACGCAnnknnknnknnknnknnkGTCnnkTATTA CCGTATCACCTACGGTGAAACT
P2 (5' phosphorylation)	TCAGACCGGAGATAGTAGCAGTmnnmnnmnnmnnTGGCACAGTGAACTCCTGAA CCGGAGAGTTACCACCAGTTTCACCGTAGGTGATACGGTAATA
P3 (5' phosphorylation)	ACTGCTACTATCTCCGGTCTGAGCCCGGGTGTCGACTACACCATCACCGTCTACG CAnnknnknnknnknnknnknnkAGCCCGATTAGCATCAACTATCGTACC
	ACTGCTACTATCTCCGGTCTGAGCCCGGGTGTCGACTACACCATCACCGTCTACG CAnnknnknnknnknnknnknnknnknnknnkAGCCCGATTAGCATCAACTATCGTACC
	ACTGCTACTATCTCCGGTCTGAGCCCGGGTGTCGACTACACCATCACCGTCTACG CAnnknnknnknnknnknnknnknnknnknnknnknnknnkAGCCCGATTAGCATCAACTA TCGTACC
P4	gctttggcctcccgggctGGTACGATAGTTGATGCTAATCGGG

Table 1. Fragments used in overlap-extension PCR to generate monobody. P1, P2 and P3 contain loop BC, DE, FG respectively. There are three variants of P3, containing 7, 10, 13 NNKs, respectively. The first 45 bp in P1 is the forward primer, and P4 is the reverse primer.

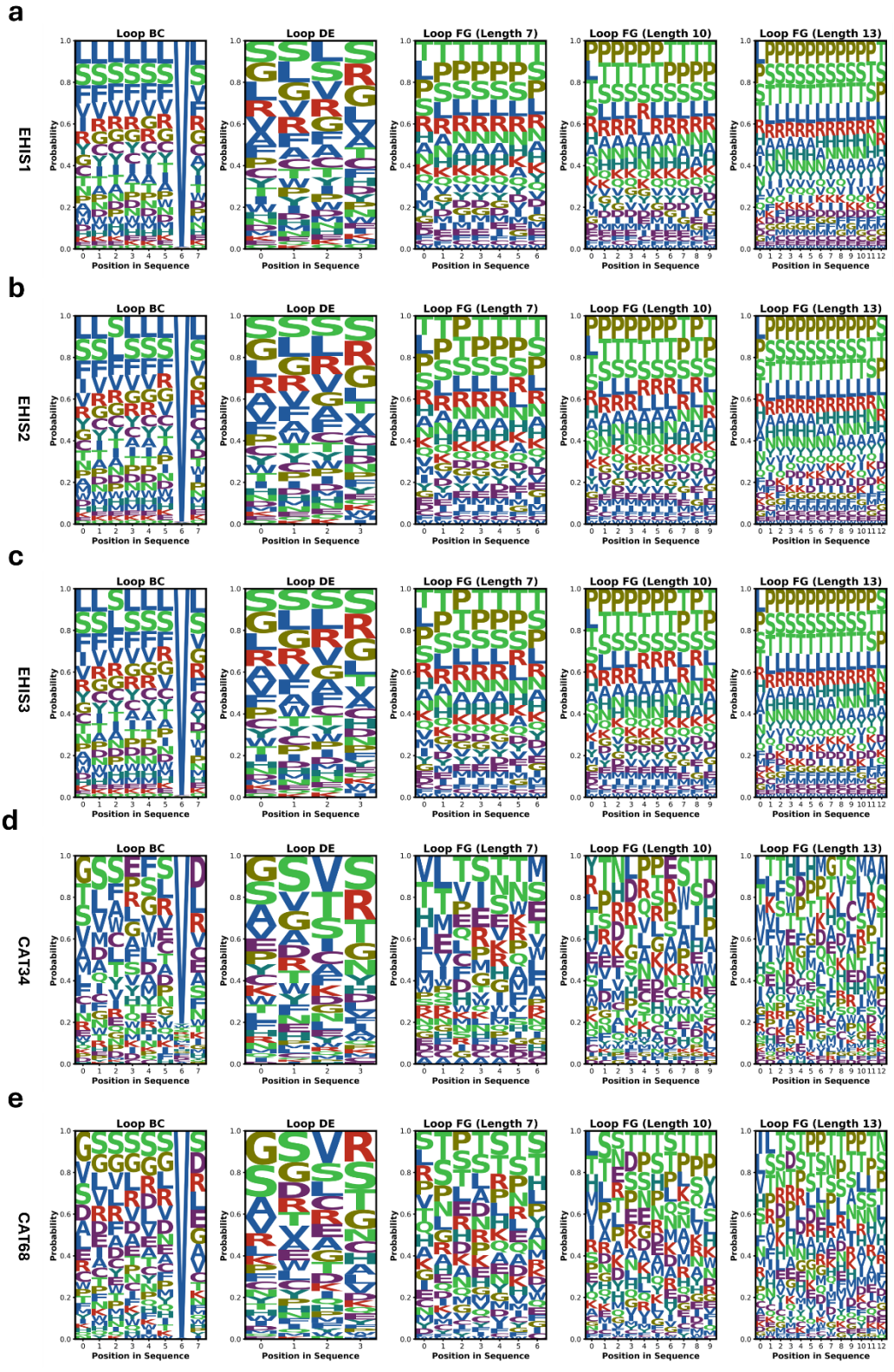

a
b
c
d
e
EHIS1
EHIS2
EHIS3
CAT34
CAT68
Loop BC
Loop DE
Loop FG (Length 7)
Loop FG (Length 10)
Loop FG (Length 13)
Probability
Position in Sequence

Figure 3. Sequence diversity diagram in each variable loop for preliminary libraries from each round of screening. EHIS1, EHIS2 and EHIS3 refer to resulting libraries of His-purification round 1, 2 and 3, respectively. CAT34 and CAT68 refer to resulting libraries of CAT assay with 34 and 68 μg/mL of chloramphenicol, respectively.

2.3.2. Anchor binder selection

Our lab previously generated nanobody-scaffold CIDs targeting cannabinoid (CBD) using the COMBINES-CID method. To evaluate the versatility of this approach and simultaneously validated our newly synthesized monobody library, we applied COMBINES-CID to create monobody-scaffold CIDs. While both nanobodies and monobodies possess three variable loops (159, 127), monobodies lack the disulfide bonds found in nanobodies, which makes nanobodies more structurally rigid. Metorphamide (MP), a neuropeptide, differs from CBD in both size and chemical properties (160). Given these differences in scaffold structures and ligands, we hypothesized that selecting monobody-scaffold CIDs targeting metorphamide would effectively demonstrate the broad applicability of our monobody library and CID selection method.

To select anchor binders, we screened the monobody library using phage biopanning with MP as the target. We began with 10^{13} phages in the first selection round, progressively increasing selection stringency by raising the number of washes and detergent concentration. After six biopanning rounds, individual clones were isolated and evaluated using phage ELISA. Results showed that most binders specifically recognized MP-coated streptavidin, with no binding to streptavidin alone. Of the 76 candidates tested, 18 exhibited significant binding affinity to MP, with oAS28 and oAS14 emerging as the strongest binders (Figure 4a).

We next assessed the expression levels of these binders by expressing them in *E. coli* C41(DE3) cells using the pET24b vector, finding both oAS28 and oAS14 displayed robust expression (Figure 4b). We further quantified their binding affinities using isothermal calorimetry (ITC), finding that oAS28 had a Kd of ~6 μM toward MP, while oAS14 had a Kd of ~22 μM (Figure 4c). The stronger binding affinity of oAS28 compared to oAS14 also correlates with the results of the phage ELISA. It is worth noting that while we previously identified nanobody anchor binders with Kd values in the low nanomolar range, the strongest anchor binders from the monobody library in this selection exhibited micromolar Kd values. Although

further selections are needed to draw more definitive conclusions, we speculate that this difference may arise from the rigidity of the nanobody scaffold, which contains disulfide bonds. A more rigid scaffold could theoretically form tighter, more specific binding pockets compared to the more flexible monobody scaffold.

Despite oAS14's relatively weaker binding, we proceeded to screen for its dimerization binders for two reasons. First, selecting synthetic CIDs with varying affinities could provide more options for sensor sensitivity in GEFS engineering. Second, our hypothesis suggests that a dimerization binder could stabilize the anchor-ligand complex, even with a weaker anchor binder. Screening for dimerization binders with a weak anchor binder allowed us to test this hypothesis effectively.

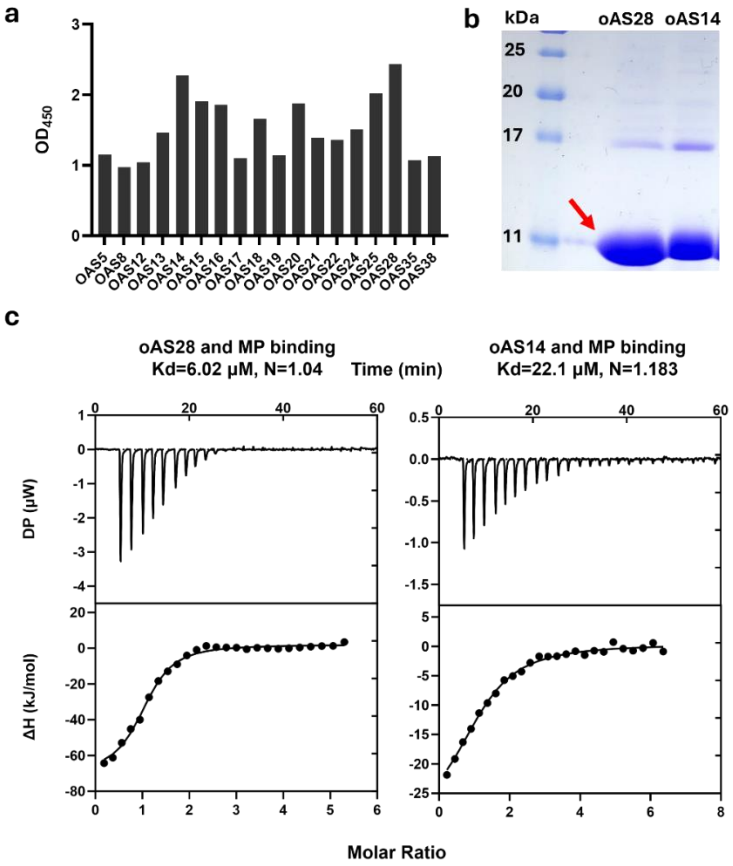


Figure 4. Anchor binder selection results. (a) Phage ELISA of anchor binder candidates. (b) SDS-PAGE of His-purified oAS28 and oAS14 expressed in E.coli. The red arrow points at the band of interest, around 11 kDa. (c) ITC binding curves of oAS28 and oAS14 binding to MP.

2.3.3. Dimerization binder selection

To obtain dimerization binders, we screened the monobody library using the anchor-ligand complex as bait. Both anchor binder candidates were expressed, purified, and mixed in equal molar concentrations, with MP added at a saturating level. To reduce false positives—binders that might interact with anchor binders in the absence of ligands—we incorporated negative selection steps in which only the anchor binder mixture was used as bait in each screening round. After four rounds of biopanning, individual clones were isolated and tested for binding to each anchor binder.

Binding was assessed both with and without 100 μM MP. Binders that showed a signal increase of more than threefold in the presence of MP were considered potential dimerization candidates. Phage ELISA results confirmed the presence of dimerization binders for both oAS28 and oAS14 (Figure 5a). The background binding for oAS14 was notably stronger, consistent with oAS14's weaker affinity as an anchor binder, which may have led to more frequent dissociation from the ligand during selection, thus attracting false positives that bound only to the unoccupied anchor binder.

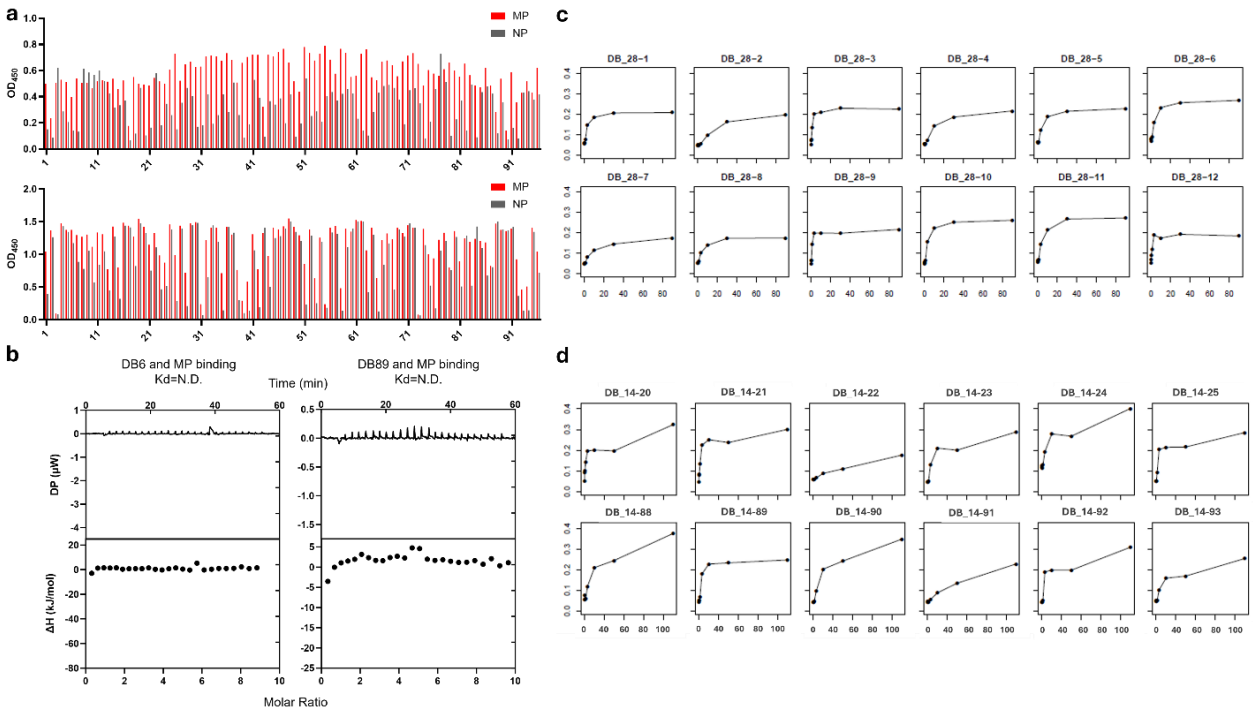


Figure 5. Phage ELISA results of dimerization binder selection for oAS14 and oAS28. (a) Phage ELISA of oAS28 (top) and oAS14 (bottom) dimerization binders; (b) ITC binding curves of DB6 and DB89. (c) Titration ELISA of oAS28 binding to dimerization binder candidates. (d) Titration ELISA of oAS14 binding to dimerization binder candidates.

One goal of using the COMBINES-CID method was to discover CIDs with diverse binding behaviors for a given ligand from a single screening experiment. To evaluate whether CID pairs selected from the monobody library displayed such diversity, we titrated each dimerization binder candidate for oAS28 and oAS14, generating binding curves that reflected the distinct binding behaviors of each CID pair (Figure 5c, 5d). ELISA results indicated that the COMBINES-CID method successfully produced synthetic CIDs with varied binding profiles for the chosen ligand. Among the dimerization binder candidates, DB6 and DB89 demonstrated both favorable binding behavior and good expression (data not shown).

We measured the binding affinity of DB6 and DB89 to the ligand using isothermal calorimetry (ITC). Neither candidate showed detectable affinity for the ligand alone (Figure 5b), consistent with our selection strategy, which favors dimerization binders that primarily interact with the anchor-ligand complex rather than the ligand by itself. Given the COMBINES-CID design, the anchor binders were selected specifically for their ability to bind the ligand, whereas the dimerization binders were chosen based on their affinity for the anchor binder-ligand complex. Due to the size differences between the anchor binder and the ligand, it is not surprising that many dimerization binders cannot bind the ligand on their own, requiring the more extensive binding surface provided by the anchor binder.

To streamline further testing, we focused on two CID pairs: oAS28-DB6 (abbreviated as oAS28-6) and oAS14-DB89 (abbreviated as oAS14-89). We proceeded to characterize the binding behavior of these CID pairs.

2.3.4. CID *in vitro* Characterization

We began with quantifying the binding affinity of the CIDs. CID dimerization involves a three-body interaction (161), which includes association and dissociation between each of the two components. Even when dimerization binders show minimal affinity toward the ligand, as seen with the oAS28-6 and oAS14-89 pairs, the overall dissociation constant (Kd) for CID dimerization consists of two steps: the binding of

the anchor binder to the ligand and the binding of the dimerization binder to the anchor-ligand complex. To capture this complexity, we measured the affinity of the CIDs for the ligand using two parameters.

In the first biolayer interferometry (BLI) assay, we immobilized the dimerization binders onto sensors and titrated them with varying concentrations of the ligand in the presence of a constant, low concentration of anchor binders. This assay provided an empirical estimate of the tendency of CIDs to form in the presence of the ligand . In this setup, the oAS28-6 pair (Figure 6a) exhibited a Kd of 1.5 μM, while the oAS14-89 pair (Figure 6b) had a Kd of 11 μM. These results confirmed that the synthetic CIDs could indeed dimerize in the presence of ligands. The difference in Kd between the two CID pairs correlates with the binding affinities of their respective anchor binders. It is logical that the dimerization affinity of CIDs would correspond to the anchor binder's affinity for the ligand: the stronger the anchor binder's interaction with the ligand, the more anchor-ligand complexes are available for the dimerization binder, leading to an overall higher apparent affinity of the CID for the ligand.

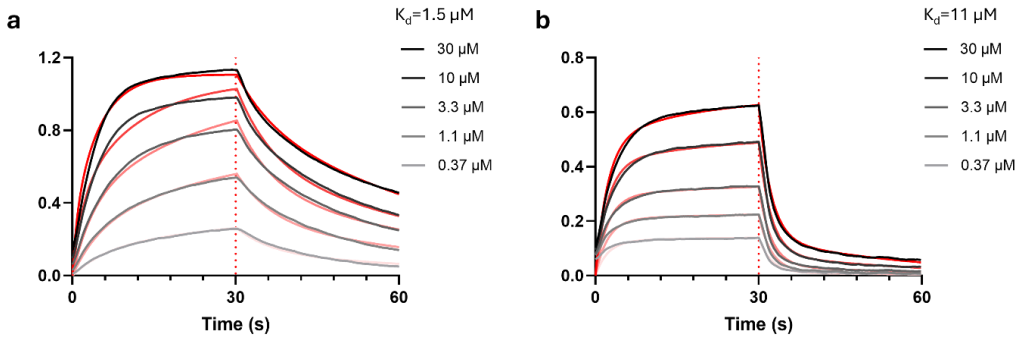


Figure 6. BLI data of CIDs binding to various concentrations of MP. (a) oAS28-6 binding to MP; (b) oAS14-89 binding to MP.

In the second BLI assay, the dimerization binders were again immobilized on sensors, but this time they were titrated with varying concentrations of the anchor binder in the presence or absence of an excess of ligand. This assay demonstrated how much the ligand enhances the affinity between the CID components (Figure 7). Without the ligand, the Kd between oAS28 and DB6 was 4.9 μM (Figure 7a), but in the presence of the ligand, the Kd improved to 127 nM—nearly a 40-fold enhancement (Figure 7b). Similarly,

the addition of ligand increased the Kd between oAS14 and DB89 from 25 μM to 832 nM (Figure 7c, 7d), representing more than a 30-fold improvement.

Together, these data demonstrate that COMBINES-CID effectively generates CIDs, and the binding behavior of these CIDs reflects the sequential nature of the method. Moreover, the method can produce multiple CIDs with diverse binding behaviors for a single ligand, facilitating the fine-tuning of GEFS sensitivity.

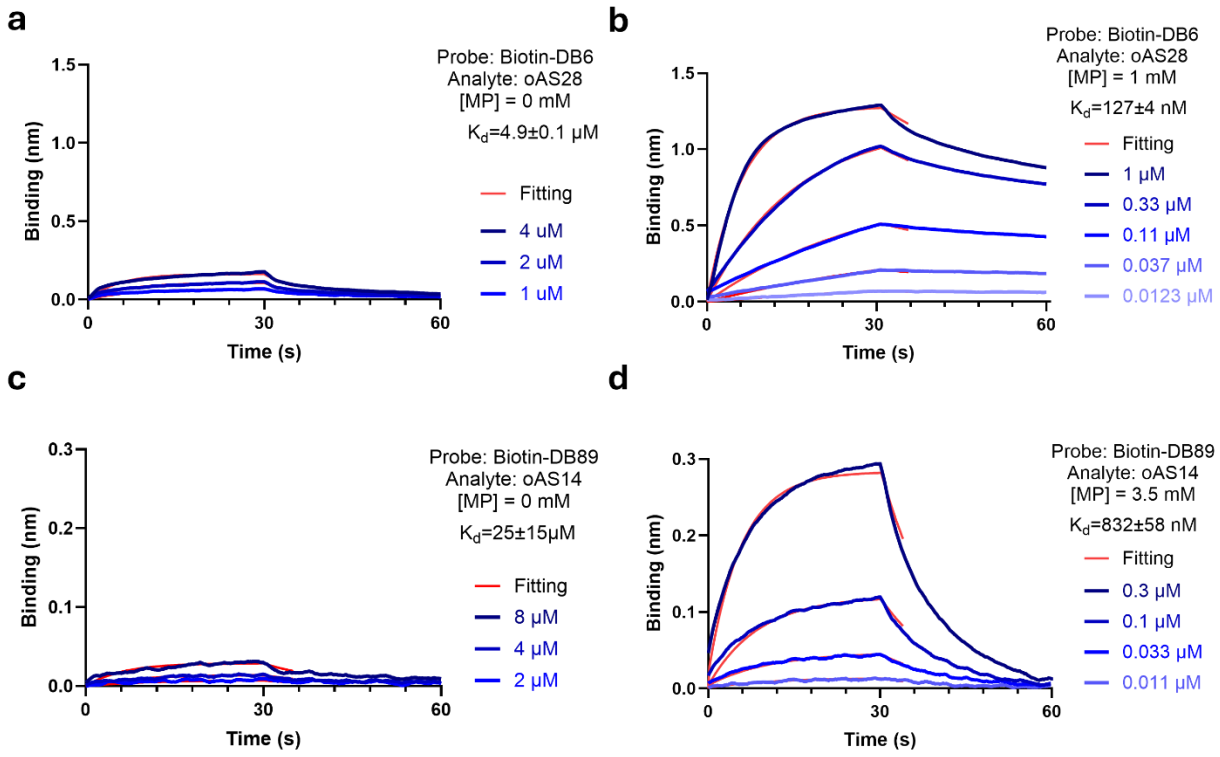


Figure 7. BLI data of CID formation in the presence or absence of MP. (a) DB6 binding to oAS28 in the presence of MP; (b) DB6 binding to oAS28 in the absence of MP; (c) DB89 binding to oAS14 in the presence of MP; (d) DB89 binding to oAS14 in the absence of MP.

2.3.5. Stabilization of Nanobody-scaffold CIDs

Our lab previously discovered a nanobody-scaffold CID pair, CA14-DB21, which binds to cannabinoid (CBD), using the COMBINES-CID method (158). However, when we fused this CID pair with cpGFP, the resulting protein failed to express in bacteria. This was expected, as nanobody binders in phage display

are expressed and folded in the periplasm of *E. coli*, where the formation of disulfide bonds is facilitated. In contrast, the cytoplasm of *E. coli* is a reducing environment that hinders disulfide bond formation, preventing proper folding of nanobody binders. Indeed, nanobody stability in cellular systems is known to be inconsistent, with aggregation and degradation being common issues.

Fortunately, previous research has addressed this problem (162). In that study, the authors demonstrated that nanobody stability could be systematically improved by introducing mutations at key positions within the scaffold. Inspired by these findings, we sought to stabilize our nanobody CIDs using a similar strategy. After analyzing the amino acid (AA) sequences of CA14-DB21 and consulting the methods described in the research, we introduced rational mutations in the protein scaffold while preserving the AAs near or within the binding loops (Table 2). Our goal was to enhance the stability of the CID without significantly compromising binding affinity.

CA14	
Stable Scaffold:	QVQLVESGGGLVQAGGSLRLSCAAS MGWFRQAPGKEREFVAA TYYADSVKGRFTISRDNAKNTVYLQMNSLKPEDTAVYYC WGQGTQVTVSS
Original Sequence:	EVQLQASGGGFVQPGGSLRLSCAASG STSRQYD MGWFRQAPGKEREFVSAIS SNQDQPP YYADSVKGRFTISRDNSKNTVYLQMNSLRAEDTATYYCA FKQHHANGA YWGQGTQVTVSS
Stable Sequence:	**Q**VQLQ**E**SGGG**L**VQPGGSLRLSCAASG STSRQYD MGWFRQAPGKEREFV**S**AIS SNQDQPP YYADSVKGRFTISRDN**A**KNTVYLQMNSL**KP**EDTATYYCA FKQHHANGA YWGQGTQVTVSS
DB21	
Stable Scaffold:	QVQLVESGGGLVQAGGSLRLSCAAS MGWFRQAPGKEREFVAA TYYADSVKGRFTISRDNAKNTVYLQMNSLKPEDTAVYYC WGQGTQVTVSS
Original Sequence:	EVQLQASGGGFVQPGGSLRLSCAASG TTYGQTN MGWFRQAPGKEREFVSAIS GLQGRDL YYADSVKGRFTISRDNSKNTVYLQMNSLRAEDTATYYCA FHDFLRMWE YWGQGTQVTVSS
Stable Sequence:	**Q**VQLQ**E**SGGG**L**VQPGGSLRLSCAASG TTYGQTN MGWFRQAPGKEREFV**S**AIS GLQGRDL YYADSVKGRFTISRDN**A**KNTVYLQMNSL**KP**EDTATYYCA FHDFLRMWE YWGQGTQVTVSS

Table 2. Stabilized nanobody amino acid sequences compared with the original. Bold amino acids in the stable scaffold are not conserved. Colored amino acids in the nanobody sequences contact with either CBD or the other protein binder. Red: CBD-contact; blue: protein-contact; purple: both.

Following stabilization, the CA14-cpGFP-DB21 fusion protein was successfully expressed at a satisfactory concentration and purity, indicating that not only were the individual nanobody binders stabilized, but their folding was also robust enough to support a multi-domain protein fusion. Further biolayer interferometry (BLI) analysis showed that the stabilized CA14-DB21 could still bind to CBD, although the affinity decreased to approximately one-fifth of the original value (Figure 8). These findings suggest that a more stable nanobody scaffold exists compared to the one used in our current libraries, and developing a new nanobody library based on this improved scaffold could be a promising direction for future research.

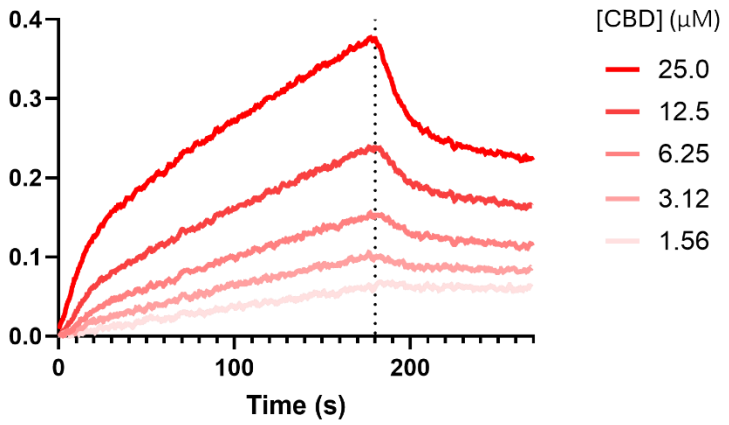


Figure 8. BLI data of stabilized CID pair binding to CBD.

2.4. Conclusion

We have developed and synthesized a combinatorial library based on the monobody scaffold, with a diversity of approximately 10^{10}. To ensure robust folding efficiency, we employed a CAT assay during library preparation to enrich for clones with high folding stability.

To validate this library, we used our COMBINES-CID method to select CID pairs binding specifically to metorphamide (MP). This approach yielded multiple anchor binders, and, notably, we identified several dimerization binders for oAS28 with distinct binding profiles. Our focus then shifted to two CID pairs, oAS28-6 and oAS14-89. Using ITC and BLI assays, we characterized the binding affinity of each anchor binder and CID pair, finding that ligand presence enhanced the dimerization of these pairs by more than 10-fold. These findings confirm both the validity of our monobody library and the broad applicability of the COMBINES-CID method.

We also explored improving the folding efficiency and stability of nanobody CID pairs by introducing mutations to the nanobody scaffold. The mutated CA14-DB21 pair exhibited robust expression in bacteria, though with a moderate reduction in binding affinity (estimated at around one-third to one-fifth of the original). Importantly, these results demonstrate the feasibility of stabilizing nanobodies without

removing disulfide bonds, supporting the potential for redesigning a nanobody library with enhanced stability in the future.

In summary, we have enhanced our CID generation pipeline by designing a combinatorial library with improved expression in bacterial systems and by validating a strategy for stabilizing the nanobody scaffold. This optimized pipeline allows us to efficiently generate synthetic CIDs that can serve as sensory domains in genetically encoded fluorescent sensors (GEFSs).

.

3. Linker Optimization with Library Screening

3.1. Introduction

With synthetic CIDs in hand, our next objective was to develop a method for efficiently integrating these CIDs into GEFSs. The critical factor in achieving this goal is optimizing the inter-domain linkers between the CID pair and cpGFP. These linkers play a crucial role in the functionality of cpGFP-based sensors (79, 81) by transmitting the conformational rearrangements of the sensory domain (CIDs) to alter the chemical environment surrounding the optical domain. Linker length is especially important, as it must be short enough to maintain tight coupling between domains, yet long enough to provide the mobility needed for the sensory domain to undergo conformational changes or dimerization.

In GEFSs where naturally occurring proteins serve as sensory domains, the focus of linker optimization typically revolves around amino acid composition, while linker length is kept short (only a few amino acids) because cpGFP is generally inserted at sites experiencing the most dramatic domain rearrangements. Short linkers work well in this configuration. However, in CID-based GEFSs, linker length requires careful optimization (Figure 1a). Dimerization can involve significant angular rotation, so the linkers need to be long enough to allow for this motion, while still short enough to preserve the allosteric coupling between the dimers and cpGFP (Figure 1b).

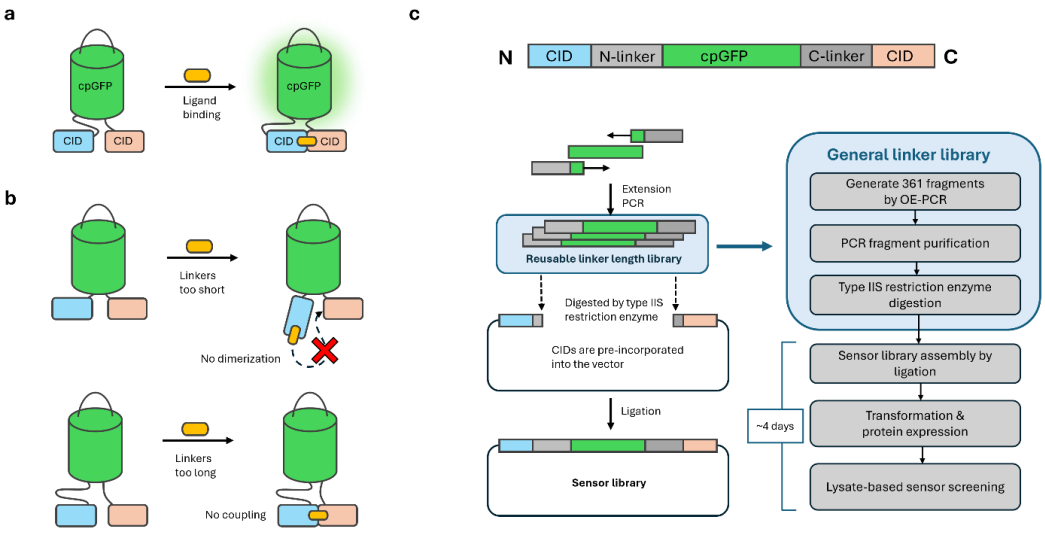


Figure 1. Schematic of CID-cpGFP sensor. In the presence of the ligand, the CID pair dimerizes, which makes cpGFP brighter through allosteric coupling. (b) The importance of linker length. If linkers are too short, the CID pair cannot dimerize properly; if linkers are too long, the dimerization of CID pair has no impact on cpGFP brightness. (c) Schematic of linker library and sensor library preparation and screening workflow.

To efficiently screen the optimal linker length, we adopted a linker library screening method. We generated a linker library containing flexible GS-linkers of various lengths (2-20 amino acids), and systematically screened all the 361 length combinations to find ones resulting in the highest sensor dynamic range (Figure 1c). We designed and made our linker library as a generalizable DNA fragment library. This library can be inserted into any synthetic CID pair generated by our pipeline, and the resulting sensor library can be rapidly screened with deep well plates and a plate reader.

To validate the linker library, we inserted it into several synthetic CID pairs of monobody and nanobody scaffolds and screened the resulting sensor libraries. We analyzed the correlation between linker lengths and sensor performance in terms of both the basal fluorescence and dynamic range. We also isolated the variants with high dynamic range and characterized their functionality, both *in vitro* and in mammalian cells.

3.2. Material and Methods

3.2.1. Linker and sensor library construction

N-terminal and C-terminal linkers with ~250 bp overlap with cpGFP and a cpGFP middle fragment of 300 bp were ordered as E-blocks from IDT. All starting DNA fragments used to generate the linker library are listed in Supplementary. To perform overlap-extension PCR, each combination of N-terminal and C-terminal linkers was mixed with cpGFP fragments in 96-well PCR plates. PCR reaction volume was 20 µL, including 10 µL Q5 Master Mix (New England Biolabs), and 1 µL of each DNA fragments (10 ng/µL each). PCR was conducted using a single-step protocol with 5-minute annealing/extension at 72°C for 30 cycles. The PCR product was purified using magnetic beads (Axygen) and subsequently digested with BbsI (NEB) at 37°C for 30 min and deactivated at 75 °C for 20 min.

To generate the receiving vector, CID sequences were amplified by PCR and inserted into the pBAD vector via Gibson assembly, following the manufacturer's instructions. A connector containing *Bbs*I digestion sites and two glycine residues (DNA sequence: ggtggcttgtcttcatgttctctcccttggcagtaatagtgag aagacggcggt) was inserted between the CID pairs during the Gibson assembly. Note that overlaps with CID pairs are excluded, as it depends on the CID scaffold being used. The complete vector was then digested by *Bbs*I.

To generate the sensor library, linker library fragments were ligated into the receiving vector in 96-well PCR plates using T4 DNA ligase (NEB) at a 3:1 insert-to-vector molar ratio, with incubation at RT for 15 min.

3.2.2. Sensor library screening

The sensor library was individually transformed into DH10B bacteria using heat shock. Each clone was inoculated into 1 mL 2xYT media containing 100 µg/mL ampicillin in a 96-deep-well plate. Cultures were grown at 37˚C until reaching an OD600 of ~0.6, then induced with 0.02% arabinose and incubated overnight at 18˚C. The cells were pelleted by centrifugation at 4,000 g for 5 minutes at room temperature and resuspended in 40 µL of B-PER bacterial protein extraction reagent (Thermo Fisher Scientific). After incubating on a benchtop plate shaker at 1,500 rpm for 15 minutes at room temperature, 320 µL PBS buffer was added to each well and thoroughly mixed. The mixture was centrifuged at 4,000 g for 5 minutes at room temperature, and the supernatant was carefully transferred to a 96-well plate, avoiding disturbance of the pellet. The fluorescence intensity and response of each clone were measured using a fluorescent plate reader (BioTek Synergy Neo2) at excitation/emission wavelengths of 488 nm/530 nm. Appropriate controls were included for baseline fluorescence and calibration. Triplicates were measured and averaged; plots were generated using MatPlotLib.

3.2.3. Sensor protein expression and purification

All protein constructs were expressed in *E. coli* using a pBAD vector, with a C-terminal Avi-tag and His-tag and were subsequently purified by Ni-affinity chromatography. In brief, E. coli strain DH10B was transformed with the expression constructs using heat shock. Transformed cells were grown in 2xYT

media containing 100 μg/mL ampicillin at 37˚C until reaching an OD_{600} of ~0.6. Protein expression was then induced with 0.02% arabinose, and cultures were incubated overnight at 18˚C. Cell pellets from 1-liter cultures were harvested by centrifugation at 8000 g at 4 ˚C, resuspended in 15 mL ice-cold His-wash buffer (50 mM sodium phosphate, pH 8.0, 300 mM NaCl, 20 mM imidazole, 10% glycerol), and kept on ice. Bacterial cells were lysed using a Branson cell sonicator at 25% amplitude for 5 minutes with 5 seconds on and 5 seconds off, in an ice bath. The lysate was then centrifuged at 10,000 g, 4˚C for 20 minutes to remove cell debris, followed by a second centrifugation of the supernatant at 20,000 g, 4˚C for 30 minutes. The clarified supernatant was filtered through a 0.45 μm syringe filter (Millipore) at room temperature to remove any remaining particulates The filtered supernatant was loaded onto a 1 mL HisTrap column (GE Healthcare) pre-equilibrated with His-wash buffer at a flow rate of 1 mL/min. The column was washed with 30 mL of His-wash buffer at 1 mL/min, and His-tagged proteins were eluted in a gradient manner, at ~ 50% of elution buffer (50 mM sodium phosphate, pH 8.0, 300 mM NaCl, 250 mM imidazole, 10% glycerol). Eluted proteins were desalted using a 50 mL Hi-Trap desalting column (GE Healthcare), equilibrated with PBS buffer with 5% PBS, at a flow rate of 5 mL/min. Desalted proteins were examined by SDS-PAGE.

3.2.4. Sensor *in vitro* characterization

Unless otherwise stated, all measurements were taken in triplets. Monobody sensors and nanobody-scaffold sensors were characterized by similar methods, and nanobody sensor N14C2 is used here as an example.

To obtain excitation/emission spectra,100 μL sensor protein solution of 1 μM was transferred to a black 96-well microplate (Greiner) and scanned by a fluorescent plate reader (BioTek Synergy Neo2). The wavelength range for excitation/emission spectra is 400-520 nm/492-590 nm, with 2 nm increment. To obtain spectra in the presence of the ligand, 30 μM CBD was added to protein solution.

To get dose-response curves, various concentrations of ligands were added to 100 μL sensor protein solution of 1 μM in a black 96-well microplate, mixed at 1000 rpm on a benchtop plate shaker for 30 seconds, and measured by a fluorescent plate reader. For CBD sensors, the CBD concentrations are 30,

3, 0.3, 0.03, 0.003, $3x10^{-4}$, $3x10^{-5}$ μM; for MP sensors, the MP concentrations are 300, 100, 33.3,11.1, 3.7,1.2, 0.4, 0 μM.

To get basal fluorescence at various pH values, PBS buffers of pH 6, 6.4, 6.8, 7.2, 7.6, 8 were prepared. Sensor protein solution was diluted using these buffers to a final concentration of 1 μM, 100 μL of which was transferred to a black 96-well microplate and measured by a fluorescent plate reader. To get fluorescence response at various pH values, 30 μM CBD was added to the same wells and mixed at 1000 rpm on a benchtop plate shaker for 30 seconds before measurement. Readings before/after addition of CBD were used to calculate $\Delta F/F_0$, which is defined as F/F_0-1. Both basal fluorescence and $\Delta F/F_0$ collected at different pH values were normalized against values at pH8. Data was analyzed using excel. Briefly, the average and standard deviation for each triplicate were calculated; for dose response curves, nonlinear regression was performed using a four-parameter model (d + (a - d) / (1 + (x / c) ^b)). All plots were generated using MatPlotLib.

3.2.5. Mammalian expression plasmid construction

To target cytoplasmic, ER lumen and cell surface expression, sensor variants were amplified by PCR and inserted into the pCMV vector (38), pCMV-PRLSS vector (38) and pAEMXT- vector (27) using Gibson assembly. pCMV vector does not add any tags to the protein construct. pCMV-PRLSS vector adds the PRLSS leader sequence and KDEL ER-retention sequence to the protein. The pAEMXT vector here refers to a modified pAEMXT used in literature where a CD59 leader sequence and a CD59 anchor sequence are added to the protein. pCMV vector was digested with *Eco*RI and *Not*I (NEB); pCMV-PRLSS vector was digested with *Age*I (NEB) and *Not*I; pAEMXT vector was digested with *Bgl*II (NEB) and *Eco*RI. PCR was conducted using a two-step protocol with 30-second annealing at 60°C and 2-minute extension at 72°C for 30 cycles. All digestion reactions were carried out at 37˚C for 30 minutes and deactivated at 75 ˚C for 20 minutes.

3.2.6. Sensor validation in mammalian cells

HEK293T cells were cultured in high-glucose DMEM (Gibco) with 10% FBS and 1% penicillin-streptomycin (Gibco) at 37°C with 5% CO_2. Cells were seeded at 4×10^4 per well in 8-well poly-D-lysine-coated channel slides (Ibidi) and transiently transfected with 0.3 μg of sensor plasmids using PEI Prime

(Serochem) at a 3:1 PEI ratio (μL:μg). After 48 hours, cells were washed with HBSS and imaged on a confocal microscope (Nikon).

For imaging over 300 seconds, HBSS with 30 μM CBD or THC was added, and images were taken at 0, 30, 90, 150, and 300 seconds. For dose-response imaging, HBSS with various CBD concentrations (0, 0.015, 0.15, 1.5 nM, 0.015, 0.15, 1.5, 15 μM) was added to each well, replacing the existing HBSS. Cells were incubated for 3 minutes at 37°C before imaging.

Images were acquired using a Nikon Ti-E automated inverted microscope equipped with a Perfect Focus System, Nikon Plan Apo Lambda 40×/1.0 objective, motorized stage (Nikon Ti-S-ER), and an Andor iXon Ultra 888 EMCCD camera. Illumination was provided by a SPECTRA X LED (Lumencor) with an excitation filter (448 ± 19 nm), dichroic mirror (506 nm), and emission filter (510 ± 20 nm) (Chroma). Images were analyzed with ImageJ, using line plots for membrane-displayed sensors and the freehand tool for cytosolic and ER-lumen sensors. Each cell was measured three times, and intensity data was processed as described in the "Sensor *in vitro* characterization" section.

3.2.7. AlphaFold prediction

Sensor structure was predicted by feeding the amino acid sequence to AlphaFold 2. The sequence used for prediction is as follows:

QVQLQESGGGLVQPGGSLRLSCAASGSTSRQYDMGWFRQAPGKEREFVSAISSNQDQPPYYADSVKG
RFTISRDNAKNTVYLQMNSLKPEDTATYYCAFKQHHANGAYWGQGTQVTVSSGGSGGSGGGGSYNVFI
MADKQKNGIKANFKIRHNIEDGGVQLAYHYQQNTPIGDGPVLLPDNHYLSVQSKLSKDPNEKRDHMVLLE
FVTAAGITLGMDELYKGGTGGSMVSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFIC
TTGKLPVPWPTLVTTLTYGVQCFSRYPDHMKQHDFFKSAMPEGYIQERTIFFKDDGNYKTRAEVKFEGDT
LVNRIELKGIDFKEDGNILGHKLEYNFNGGSGQVQLQESGGGLVQPGGSLRLSCAASGTTYGQTNMGWF
RQAPGKEREFVSAISGLQGRDLYYADSVKGRFTISRDNAKNTVYLQMNSLKPEDTATYYCAFHDFLRMW
EYWGQGTQVTVSS

3.3. Results

3.3.1. Linker Library Design and Generation

Since we aimed to optimize linker lengths without relying on structural guidance—given that we were not resolving structures for our synthetic CIDs—and wanted a method applicable across different synthetic CID scaffolds, our approach had to be generalizable to avoid overly specific optimizations.

To tackle this, we designed a linker library composed of glycine-serine (GS) linkers ranging from 2 to 20 amino acids in length. By restricting the amino acid composition to G and S, we maximize flexibility and eliminate the confounding effects of differing side-chain properties. This design produces 361 unique combinations of N- and C-terminal linker lengths (19 for each terminus). The central portion of each library construct contains the same cpGFP fragment from previous research (44, 83), flanked by variable-length N- and C-terminal linkers (Figure 1c), with all linker elements detailed in Table 1. To ensure compatibility with any CID pairs across different scaffolds, we added a two-glycine (GG) adapter between the linker and the CIDs.

The linker library was generated using overlap extension PCR (108), with each PCR reaction verified by agarose gel electrophoresis to ensure quality (Figure 2). The purified PCR products were then stored at -20°C for long-term use in future experiments.

Linker Index	Amino Acid	Length	Linker Index	Amino Acid	Length
N2	GG	2	C2	GG	2
N3	GGS	3	C3	SGG	3
N4	GGGS	4	C4	SGGG	4
N5	(GGGGS)	5	C5	SGGGG	5
N6	GGSGGS	6	C6	SGGSGG	6
N7	GGSGGGS	7	C7	SGGGSGG	7
N8	GGS(GGGGS)	8	C8	(SGGGG)SGG	8
N9	GGGS(GGGGS)	9	C9	(SGGGG)SGGG	9
N10	$(GGGGS)_2$	10	C10	(SGGGG)SGGGG	10
N11	GGSGGS(GGGGS)	11	C11	(SGGGG)SGGSGG	11
N12	GGSGGGS(GGGGS)	12	C12	$(SGGGG)_2GG$	12
N13	$GGS(GGGGS)_2$	13	C13	$(SGGGG)_2SGG$	13
N14	$GGGS(GGGGS)_2$	14	C14	$(SGGGG)_2SGGG$	14
N15	$(GGGGS)_3$	15	C15	$(SGGGG)_3$	15
N16	$GGSGGS(GGGGS)_2$	16	C16	$(SGGGG)_2SGGSGG$	16
N17	$GGSGGGS(GGGGS)_2$	17	C17	$(SGGGG)_2SGGGSGG$	17
N18	$GGS(GGGGS)_3$	18	C18	$(SGGGG)_3SGG$	18
N19	$GGGS(GGGGS)_3$	19	C19	$(SGGGG)_3SGGG$	19
N20	$(GGGGS)_4$	20	C20	$(SGGGG)_4$	20

Table 1. All linker elements in the linker library. N means linkers at the N-terminus of cpGFP; C means linkers at the C-terminus of cpGFP.

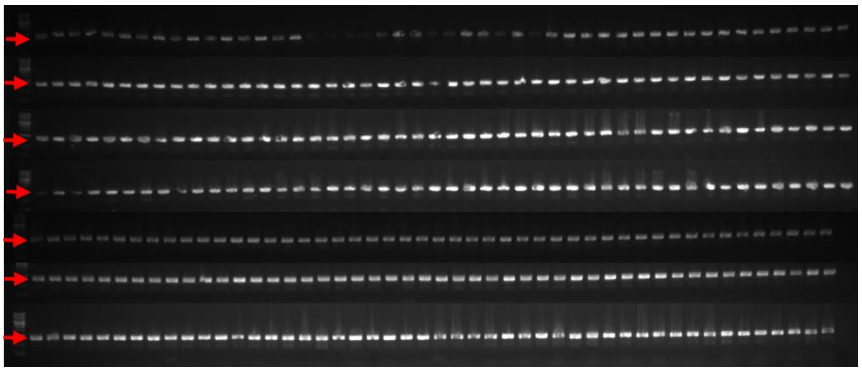


Figure 2. Agarose gel image of all 361 linker library fragments generated by overlap-extension PCR. Red arrows point to the band of interest (800-1000 bp).

3.3.2. Sensor Library Generation

We aimed to test the linker DNA library on three CID pairs: the stabilized nanobody-scaffold CA14-DB21, as well as the two monobody-scaffold CID pairs generated in this study, oAS28-6 and oAS14-89. The assembly of the sensor library was designed to be straightforward, as the linker DNA library allows for insertion between CID pairs with just a single ligation step. However, isolating clones, extracting plasmid DNA, and performing sequencing for all 361 unique constructs would be a significant burden—particularly since the library is intended to be generalizable and reused for future GEFS engineering projects.

We reasoned that if the ligation efficiency was high enough for correctly ligated plasmids to dominate the bacterial culture, these clones would be naturally enriched during overnight growth. With a population enriched for correctly ligated constructs, even if not individually isolated, the bacterial culture should still produce proteins with the correct reading frame and amino acid sequences upon induction. Based on this rationale, we decided to insert the linker library between the CIDs, transform *E. coli* with the plasmids, and inoculate the transformants directly, without isolating individual clones.

To ensure that most cells carried correctly ligated plasmids, we optimized the ligation protocol to achieve >96% efficiency (Figure 3). To further enrich the culture, we passaged the cells through an additional round of growth after the initial overnight incubation. It's worth mentioning that we initially failed to express the sensors in a pET vector and had to switch to the pBAD vector. We hypothesized that the slow, tightly regulated induction of the pBAD system would better support sensor folding. This switch was

important, as cpGFP does not fold as robustly as GFP, and cpGFP fusions often face performance challenges in mammalian cells (163).

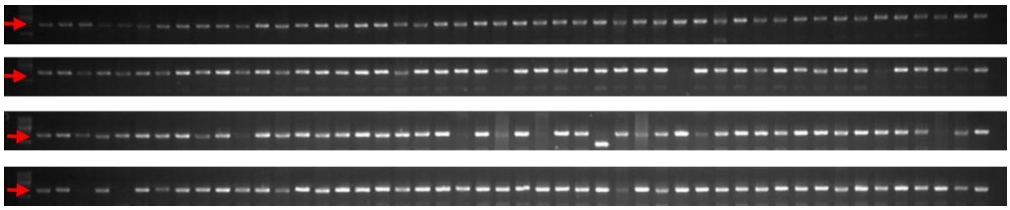


Figure 3. Agarose gel image of colony PCR of randomly picked 196 clones after ligation. Red arrows point to the band of interest (~2000 bp).

3.3.3. Screening and Characterization of Sensor Libraries

Once individual cell cultures with enriched correct clones were prepared, each unique clone was inoculated and induced in 96-deep well plates. The bacteria were lysed, releasing the sensor proteins into the solution (82). For each sensor library, we first measured the basal fluorescence intensity of each well to verify that our sensor library preparation method was working as intended. We ensured that all wells had the same culture volume, induction, and incubation conditions. We used the lysate from untransformed bacteria as a blank control, subtracting its fluorescence intensity from each well's measurement. These data were organized into a heat map.

For all three CID pairs, the sensor variants produced consistent fluorescence signals above the blank (Figure 4), confirming that even without isolating individual clones, sufficiently enriched cell cultures could still produce correctly folded proteins. The heatmaps for the three distinct CID pairs also revealed interesting patterns. First, the overall fluorescence intensity of the CID-based sensors varied depending on the specific CIDs used. For example, the fluorescence range for oAS28-6-based sensors (Figure 4a) was 5,000 to 25,000; for oAS14-89 (Figure 4b), it was 4,000 to 18,000; and for CA14-DB21 (Figure 4c), the range was 10,000 to 60,000. While differences between scaffolds are expected, the fact that oAS28-6 and oAS14-89, despite being from the same scaffold, produced sensors with distinct basal brightness suggests that the specific amino acid composition of each binder also significantly impacts fluorescence.

An even more striking observation was that linker lengths dramatically influenced the basal brightness of the CID-based sensors, even for the same CID pair. Interestingly, C-terminal linkers had a greater effect on brightness than N-terminal linkers. For example, a 5-amino acid C-terminal linker increased the basal brightness of oAS28-6 sensors from ~5,000 to ~20,000, a several-fold enhancement. Similar trends were observed for oAS14-89 and CA14-DB21, indicating this property might be common across CID-based sensors with this configuration. Although C-terminal linkers were dominant in all three CID pairs, N-terminal linkers also had some impact on brightness in monobody scaffolds: a 3-amino acid N-terminal linker consistently improved the brightness of sensors for both oAS28-6 and oAS14-89.

While inter-domain linkers clearly influence protein folding, the brightness of GEFSs cannot be attributed solely to folding or expression levels, as the optical domain is sensitive to its surrounding chemical environment. It is possible that varying linker lengths not only affect the folding efficiency of CID sensors but also alter the spatial relationship between the CIDs and the optical domain, creating a combined effect on sensor brightness. Quantifying protein concentration may provide further insights into this phenomenon. Fluorophore-labeled anti-GFP antibodies of different colors could be a useful tool for this, enabling simultaneous measurement of sensor brightness and expression levels, if improperly folded proteins lack a properly folded cpGFP domain.

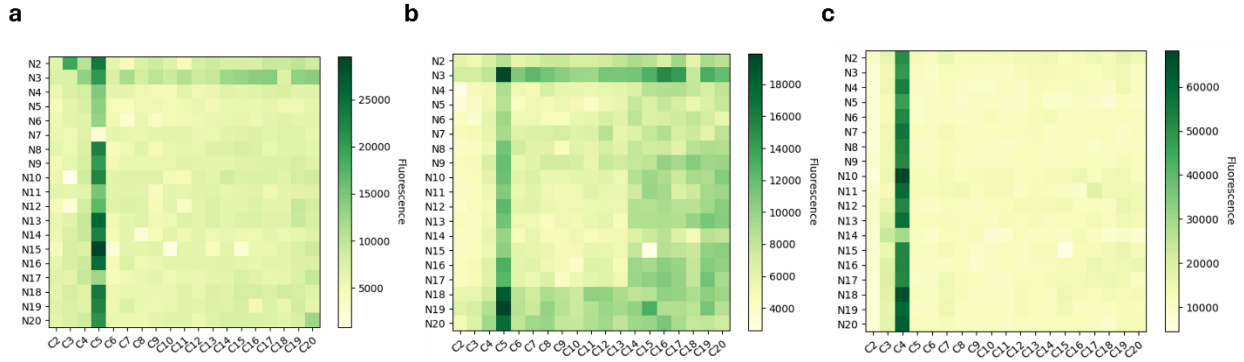


Figure 4. Heatmaps of each sensor library's basal fluorescence intensity. (a) Heatmap of oAS28-6 sensor library; (b) heatmap of oAS14-89 sensor library; (c) heatmap of CA14-DB21 library.

After identifying interesting correlations between linker length and basal brightness, we next evaluated the dynamic range of each variant in the sensor libraries. To do this, we added a saturation level of ligand

(more than 10 times the binding Kd) to each well and measured the fold change in fluorescence, defined as F/F_0.

Sensor variants exhibiting F/F_0 > 200% were identified in all three sensor libraries (Figure 12), demonstrating that we successfully generated GEFSs with a respectable dynamic range using our linker library and screening method. A closer examination of the data reveals that linker lengths on both the N- and C-terminal ends influenced the sensor's dynamic range, with the C-terminal linker exerting a more pronounced effect. These findings align with our observations on basal brightness, further supporting the notion that linker lengths affect the microenvironment around the optical domains, thereby impacting both the basal brightness and dynamic range of the sensors.

Interestingly, for all three CID pairs, the C-terminal linker lengths that maximized basal fluorescence and dynamic range were not the same. In fact, we consistently observed that the linkers yielding the highest dynamic range resulted in lower basal fluorescence. This suggests that the influence of linker length on basal fluorescence likely involves alterations in the chemical environment around cpGFP rather than merely affecting protein expression levels. The high dynamic range appears to be achieved through disruption and subsequent restoration of cpGFP fluorescence.

Another intriguing observation was that sensor dynamic range depends on the lengths of linkers on both ends independently, rather than the total combined length. Even with the same sum of linker lengths, sensor dynamic range varied significantly across different length combinations, indicating that the dynamic range of cpGFP-based sensors is shaped by a complex interplay of factors.

Unlike basal brightness, which varied considerably depending on the CID pair used, the dynamic range of the sensors was more consistent across different pairs. The oAS14-89 pair produced the most responsive sensor, with F/F_0 exceeding 250%, while the highest F/F_0 for the other two CID pairs was ~200%. Notably, each CID pair had a different optimal linker combination for maximizing dynamic range. For oAS28-6, the optimal combination was N20C3 (Figure 5a); for oAS14-89, it was N11C3 (Figure 5b); and for CA14-DB21, it was N13C2 (Figure 5c). This suggests that the dimerization orientation of each CID pair is distinct. The relatively similar dynamic range across different CID pairs, despite their varying

dimerization orientations, suggests that the specific binding interface of each CID has limited influence on cpGFP. This could be advantageous or disadvantageous depending on the goals of GEFS engineering: on one hand, it indicates that achieving a decent dynamic range with CID-based GEFSs is relatively straightforward; on the other hand, further improvement through mutagenesis might be challenging, as the CIDs' binding interfaces have minimal direct interaction with cpGFP, leaving little room for enhancement. Directed evolution of linker amino acid composition and whole-gene mutagenesis could provide further insights.

Another important observation from both the basal brightness and dynamic range data is that the optimal C-terminal linker length appears to be scaffold-dependent. For oAS28-6 and oAS14-89, the C-terminal linker length that maximized basal fluorescence was 5 amino acids, while for CA14-DB21, it was 4. Similarly, the linker lengths that maximized dynamic range were 3 amino acids for oAS28-6 and oAS14-89, and 2 amino acids for CA14-DB21. Notably, the difference between the optimal C-terminal linker lengths for basal fluorescence and dynamic range was consistently 2 amino acids. This suggests that the C-terminal linker plays a crucial role in CID-sensor function and operates in a generalizable manner.

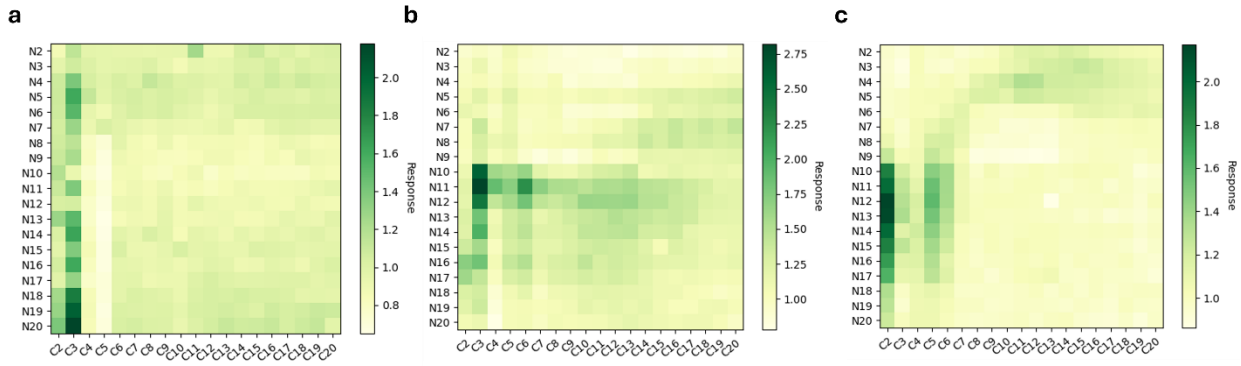


Figure 5. Heatmaps of each sensor library's F/F_0. (a) Heatmap for oAS28-14 sensor library; (b) heatmap for oAS14-89 sensor library; (c) heatmap for CA14-DB21 sensor library.

To explore the role of the C-terminal linker further, we used AlphaFold to predict the structure of a CBD sensor, N14C2 (164, 165) (Figure 6). Examining the predicted structure, we hypothesized that since the binding loops of the C-terminal binder are positioned close to the beta-barrel opening of cpGFP, a

sufficiently short C-terminal linker would position the CID pair near the cpGFP fluorophore. The presence of the CID pairs, whether in monomeric or dimeric form, could significantly affect the microenvironment of cpGFP. The scaffold-specific nature of the optimal C-terminal linker length, as well as the consistent 2-amino-acid difference between the lengths that maximize basal fluorescence and dynamic range, remain unclear. Resolving the structures of CID sensors in both ligand-bound and unbound states may provide insights into these mechanisms.

Regardless of the open questions and speculations, we successfully identified sensor variants with F/F_0 > 200% for each CID pair. These results demonstrate that our linker library is generalizable across different synthetic CID scaffolds, allowing for the rapid identification of sensor variants with robust dynamic ranges, all without requiring structural guidance or extensive directed evolution.

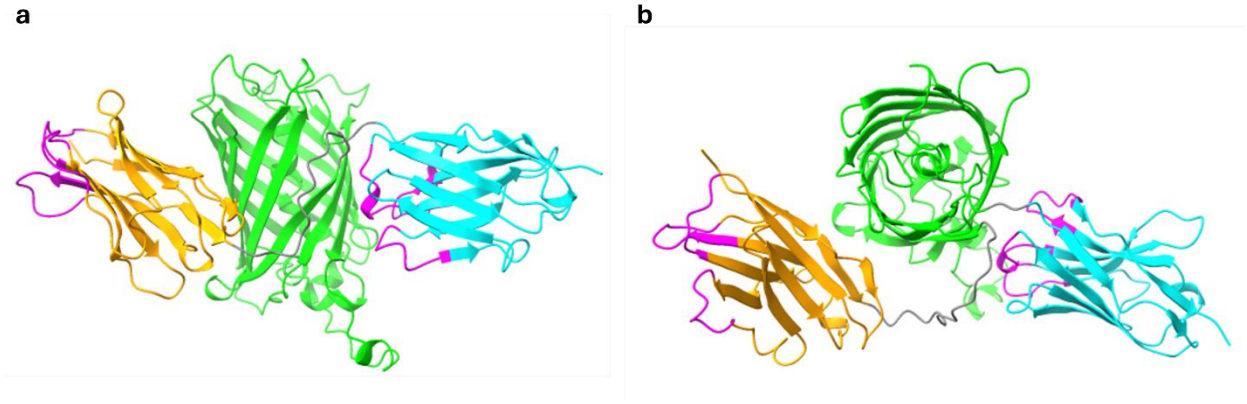


Figure 6. AlphaFold-predicted structure of CA14-DB21N14C2 sensor variant. (a) Side view; (b) top-down view. Green: cpGFP; magenta: variable loops; orange: CA14; cyan: DB21; grey: inter-domain linkers.

3.3.4. *In vitro* Characterization of Sensor Variants

Having established the generalizability of our linker library across different scaffolds, we proceeded to conduct a detailed characterization of the binding behavior and spectra of the sensor variants. We began by analyzing two sensor variants utilizing monobody CIDs: 28N20C3 (using the oAS28-6 pair) and 89N11C3 (using the oAS14-89 pair). Both sensors were maintained at a concentration of 1 μM, and we introduced varying concentrations of MP to measure the fold change in fluorescence.

Surprisingly, both sensor variants exhibited a significantly higher dynamic range compared to the results observed during library screening (Figure 7). At approximately 30 μM of MP, the 28N20C3 sensor reached around 400% F/F_0 (Figure 7a, 7c), while the 89N11C3 sensor exceeded 1000% F/F_0 at around 100 μM MP (Figure 7b, 7d). Notably, instead of leveling off at a plateau, the binding curves displayed a bell-shaped profile. Bell-shaped curves are a hallmark of CID-based sensors (166), indicating that the dimerization binders have weak affinity for the ligand. At higher ligand concentrations, the dimerization binders start to interact with the ligand, leading to simultaneous occupation of both the anchor and dimerization binders. This simultaneous binding prevents CID formation, causing a drop in the fluorescence signal. To investigate why the purified sensors exhibited a much higher dynamic range compared to those in lysate, we hypothesized that the increased sensor concentration in the purified form could be a contributing factor. To test this, we repeated the measurements with sensors at a concentration of 0.5 μM, which is half of the original concentration. Surprisingly, at the lower concentration, the binding curves deviated noticeably from the original data, with the effect being more pronounced for 89N11C3 than for 28N20C3.

We further titrated 89N11C3 at various concentrations and confirmed that the sensor's binding curve was indeed concentration dependent. Similar trends were observed with other sensor variants (Figure 7a), including those using the CA14-DB21 pair (Figure 7b). Based on these findings, we speculated that the sensing mechanism of our CID sensors may partially rely on oligomerization, meaning intermolecular interactions play a role (Figure 8). This is not unusual, as oligomerization in GEFSs has been previously reported, such as in the case of GCaMP2 (167).

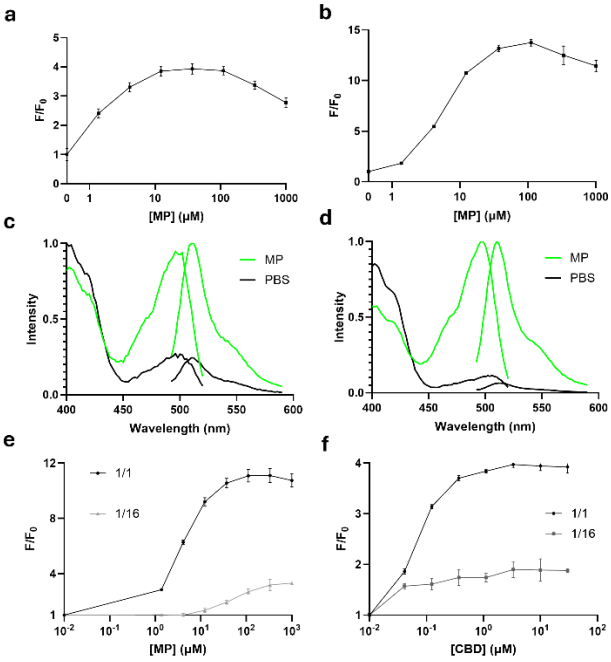


Figure 7. In vitro characterization of sensor variants. (a) Titration curve of 28N20C3 against MP; (b) titration curve of 89N11C3 against MP; (c) Spectra of 28N20C3 in the presence/absence of MP; (d) Spectra of 89N11C3 in the presence/absence of MP; (e) titration curves of two concentrations of 89N11C3 against MP; (f) titration curves of two concentrations of 14N14C2 against CBD. "1/1" and "1/16" Refers to the concentration of sensor relative to 1 μM.

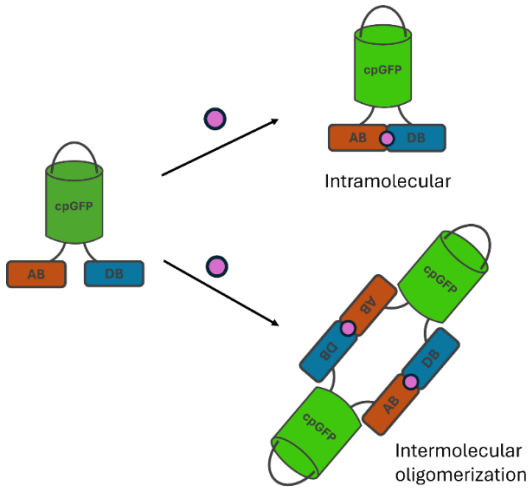


Figure 8. Schematics of CID-sensor mechanisms. CID-sensors can undergo intramolecular interactions where CID pair from the same molecule dimerizes; alternatively, CID-sensors can undergo intermolecular interactions where CID pairs from different molecules dimerize, leading to dimerization even oligomerization of sensor molecules.

To test whether CID sensors undergo oligomerization as speculated, we performed size exclusion chromatography (SEC) with and without ligands. To stabilize sensors in their ligand-bound forms, we pre-incubated the sensors with ligands and added cross-linking reagents. The SEC data showed that at saturating ligand levels, the chromatographic profiles of both 28N20C3 and 89N11C3 shifted towards the formation of larger oligomers, though not completely (Figure 9a, 9b). This suggests that the sensing mechanism of CID sensors involves a combination of intermolecular and intramolecular interactions. At higher sensor concentrations, sensors tend to favor oligomerization over intramolecular interaction, and the reverse occurs at lower concentrations.

It is intuitive that sensor oligomerization alters the chemical environment around cpGFP differently from intramolecular dimerization. However, it's important to note that oligomerization does not necessarily result in a higher dynamic range than intramolecular dimerization. We also found sensor variants that are more responsive at lower sensor concentrations, indicating that intramolecular interactions contribute more significantly to sensor dynamic range in certain cases. For example, in the case of 89N19C2, the F/F_0 increased by more than 100% when the sensor concentration was diluted 16-fold (Figure 9c, 9d). Notably, the saturation point of the sensor also increased by at least two orders of magnitude, suggesting that oligomerization enhances CID affinity for ligands, and cooperativity may be involved in the process.

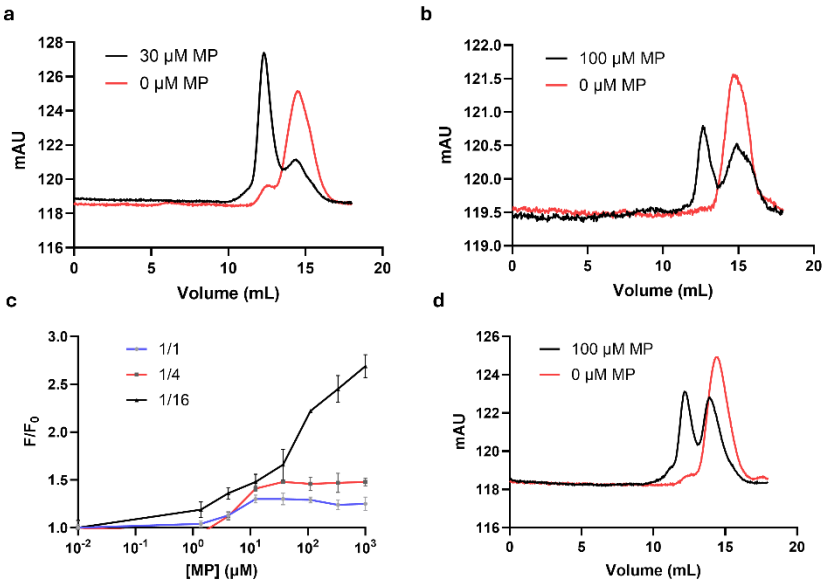


Figure 9. (a) SEC profile of 28N20C3 with/without MP; (b) SEC profile of 89N11C3 with/without MP; (c) Titration curves of different concentrations of 89N19C2 against MP; (d) SEC profile of 89N19C2.

We moved on to complete characterization of CA14-DB21-based sensor variant. Among the variants N11C2, N12C2, N13C2, and N14C2, which exhibited similar responses, we selected 14N14C2 for further experiments due to its slightly higher *in vitro* expression level. We first analyzed the excitation and emission spectra of N14C2, finding them slightly red-shifted compared to EGFP (Ex λ=498 nm, Em λ=514 nm; Figure 10a). The significantly higher excitation peak at 400 nm suggests a larger protonated population, likely due to the unique microenvironment of the fluorophore when fused with synthetic CIDs. Titration with CBD revealed a maximum $\Delta F/F_0$ of ~110% and an EC50 of 44.37 nM (Figure 10b). In contrast, titration with THC produced an EC50 of 2154.5 nM—about 50 times higher than CBD—with a maximum $\Delta F/F_0$ of ~20%, demonstrating good specificity. N14C2 exhibited a dynamic range from 10 nM to 1 μM, indicating a broad sensing range. As with most cpGFP sensors, N14C2 was pH-sensitive, with basal fluorescence and response to CBD decreasing at lower pH levels; at pH 6.0, it retained only 50% of its maximum $\Delta F/F_0$ observed at pH 8 (Figure 10d). These results suggest that synthetic CID-based sensors largely retain the biochemical properties of cpGFP and CIDs, with minor impacts on cpGFP spectra.

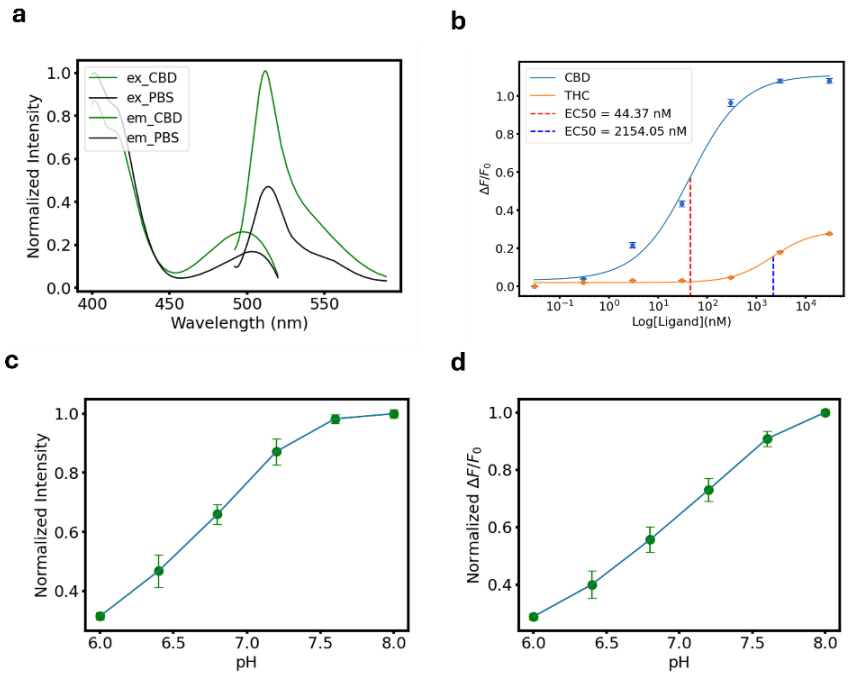


Figure 10. Excitation and emission spectra of CBD sensor N14C2. (b) Dose response curves for N14C2 sensing CBD and THC; $\Delta F/F_0=F/F_0-1$. (c) N14C2 basal fluorescence at various pH. Fluorescence intensity is normalized against the value measured at pH8. (d) N14C2 $\Delta F/F_0$ to 30 μM CBD at various pH. $\Delta F/F_0$ is normalized against the value measured at pH8.

Overall, the results demonstrated that our method successfully generated CID-based GEFSs with strong *in vitro* performance. Additionally, CID-based sensors exhibited both intra- and intermolecular dimerization, which can offer distinct advantages depending on the application. For instance, when applied at high concentrations, CID sensors like 89N11C3 can achieve very high dynamic ranges. Conversely, even at low sensor concentrations, it is possible to select sensors where the dynamic range primarily depends on intramolecular dimerization, as observed with 89N19C2.

However, a more in-depth exploration of the sensing mechanisms is warranted, particularly to better understand the relationship between sensing mechanisms and linker lengths. Gaining such insights would significantly enhance the screening process for CID-GEFSs, enabling more precise adaptation to specific application requirements.

3.3.5. Validation of Sensor Function in Cellular Environments

The strength of GEFSs lies in their ability to monitor cellular systems non-invasively. To demonstrate the potential of CID-GEFSs as valuable research tools, it is crucial to validate their functionality in cellular environments. As discussed earlier, the nanobody scaffold showed stability issues when expressed in the bacterial cytoplasm, and although mutations significantly improved expression, we have not yet tested the stabilized CID pair in mammalian cells. Moreover, given the challenges we faced expressing CID sensors in the pET vector, it appears that the folding of these multi-domain fusions may not be particularly robust.

As a first step, we tested whether CID-GEFSs can be successfully expressed in mammalian cells. We transfected four variants of the oAS28-6 and oAS14-89 sensors into HEK293T cells, since the monobody-scaffold CIDs may have a better chance of proper expression compared to the nanobody-scaffold ones. Fluorescent microscopy images confirmed successful expression in HEK293T cells, and the brightness of the sensors was consistent with the screening results (Figure 11). As expected, variants with 5-amino-acid C-terminal linkers exhibited the brightest fluorescence, and overall, the sensors based on oAS14-89 (Figure 11b) were dimmer than those based on oAS28-6 (Figure 11a).

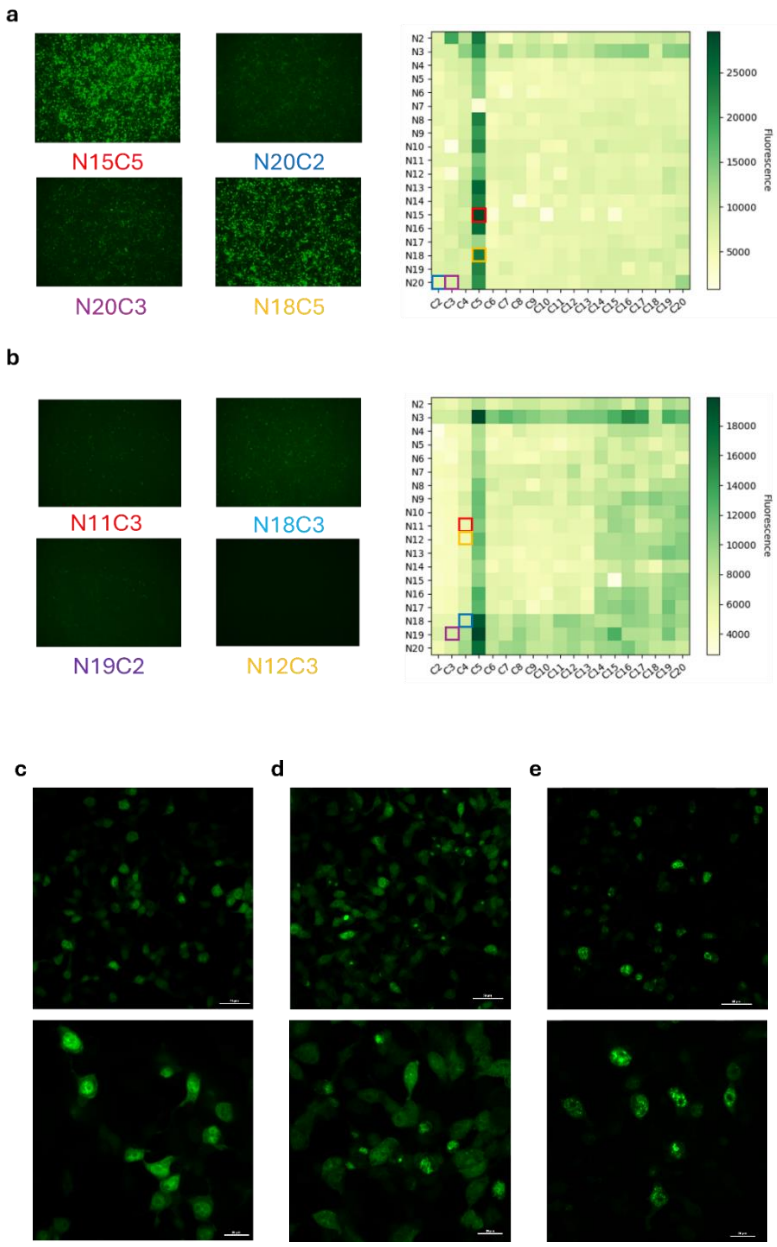


Figure 11. HEK293T cellular expression of CID-sensor variants. (a) Expression of selected oAS28-6 sensor variants and their basal brightness in screening; (b) expression of selected oAS14-89 sensor variants and their basal brightness in screening; (c) Zoom-in view of cellular expression of 28N18C5; (d) zoom-in view of cellular expression of 28N20C3; (e) zoom-in view of cellular expression of 89N11C3. Scale bar: 20 μM.

Zoomed-in images of 28N18C5 (Figure 11c), 28N20C3 (Figure 1d), and 89N11C3 (Figure 11e) provided further insights into the cellular distribution of the sensors. Due to the size of the sensors (~55 kDa), they

were able to diffuse into the nucleus, as seen in all three variants. Both 28N18C5 and 28N20C3 were evenly distributed in the cytoplasm, while 89N11C3 was primarily concentrated in the nuclei and tended to form aggregates, appearing as small bright dots in the images. These results suggest that oAS28-6 folds more robustly than oAS14-89 in HEK293T cells. Additionally, we successfully targeted the sensor to the ER lumen (Figure 12).

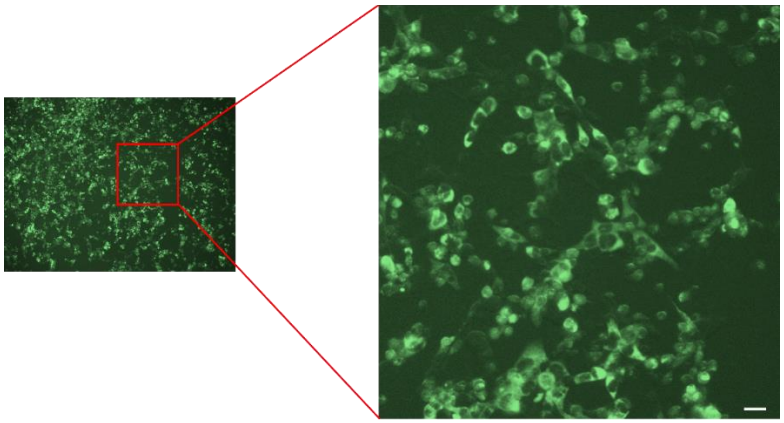


Figure 12. Expression of 28N18C5 in HEK393T ER lumen. Scale bar: 20 µm.

Next, we tested whether our sensors could be displayed on the surface of HEK293T cells. We incorporated 28N20C3 and 89N11C3 into a pDisplay vector and transfected HEK293T cells. To assess display efficiency, we labeled the surface-displayed sensors with anti-GFP-Alexa647 antibodies. Sensors expressed via the pDisplay vector exhibited relatively low display efficiency, with most of the protein trapped inside the cells, as seen in the microscope images and line profiles across the cells (Figure 13a, 13c). This issue has been encountered in previous studies, where researchers have addressed it by using alternative display anchors and leader sequences.

To improve display efficiency, we switched to a previously published display vector, pAEMXT, which replaces the Igκ leader sequence and PDGFRβ anchor protein with the CD59 leader sequence and CD59 anchor. While some cytosolic expression and aggregation were still observed, the new vector significantly improved display efficiency for both 28N20C3 and 89N11C3 (Figure 13b, 13d). In the future, it may be possible to further enhance display efficiency by screening through a selection of leader sequence-anchor combinations, as suggested in previous studies (168).

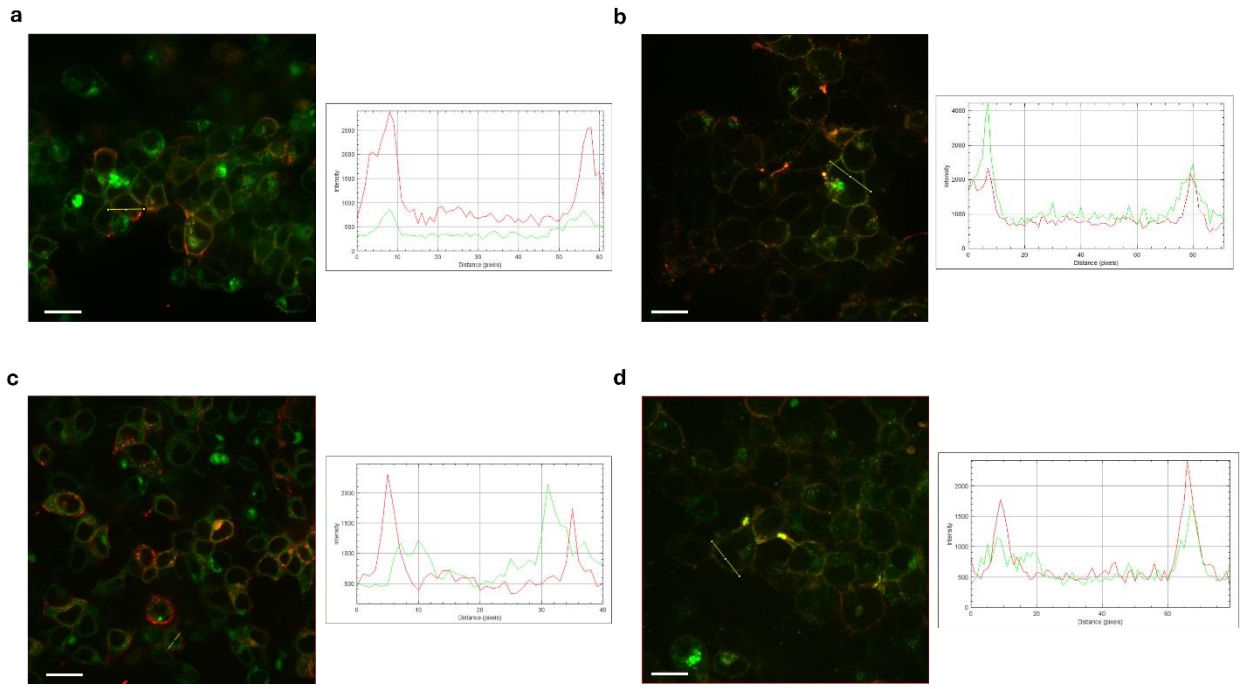


Figure 13. Images of displayed CID-sensors on the plasma membrane of HEK293T and line profiles. (a) 28N20C3 displayed with pDisplay vector; (b) 28N20C3 displayed with pAEMXT vector; (c) 89N11C3 displayed with pDisplay vector; (d) 89N11C3 displayed with pAEMXT vector.

With the sensors successfully displayed on the cell surface, we next tested whether they remained functional after passing through the expression and secretion machinery of HEK293T cells. Since 89N11C3 showed a higher tendency to aggregate, we focused on 28N20C3. To evaluate sensor functionality, we displayed 28N20C3 on the surface of HEK293T cells (Figure 14b) and titrated it with MP concentrations ranging from 1 to 1000 μM. The maximum $\Delta F/F_0$ was 40% ($\Delta F/F_0 = F/F_0 - 1$), and the EC_{50} was 2.75 μM (Figure 14a). The EC_{50} of 28N20C3 was close to the Kd of oAS28-6 as estimated by BLI, but the dynamic range was lower than what was observed during library screening.

Several factors could explain this reduced dynamic range. Firstly, it is common for GEFSs to exhibit different performance when transferred between expression systems due to inherent differences (163). Secondly, as shown in previous studies, the sensing mechanism of our CID-based GEFSs relies on a combination of inter- and intramolecular interactions, and the dynamic range is sensor concentration-dependent. Thus, it is not surprising that the sensor underperformed when displayed on the HEK293T cell surface, likely due to lower sensor concentration. Additionally, surface display anchors the sensor protein

to the cell membrane, which greatly limits its mobility and ability to undergo intermolecular interactions and oligomerization, potentially contributing to the lower dynamic range.

Despite the reduced dynamic range, these results demonstrate that our monobody-scaffold CID-GEFSs are compatible with mammalian expression systems. They can be expressed and correctly localized to various cellular locations, though further optimization of conditions may be necessary.

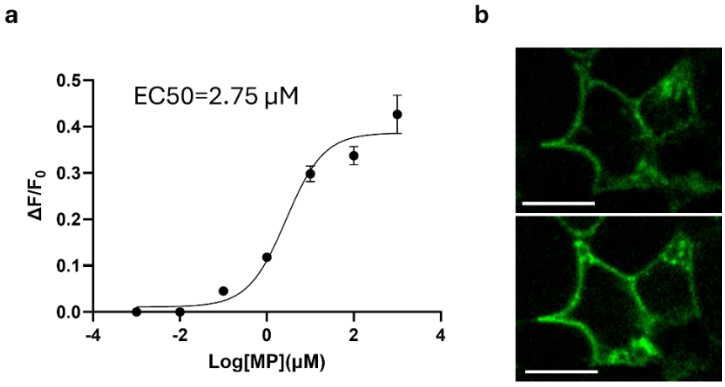


Figure 14. Response of 28N20C3 to MP while displayed on the plasma membrane of HEK293T. (a) Titration curve of 28N20C3 against MP; (b) images of cell surface-displayed 28N20C3 before and after the addition of 100 μM MP. Scale bar: 10 μm.

Building on the initial success with monobody-scaffold sensors, we then investigated whether nanobody-scaffold sensors are also compatible with mammalian expression systems. Based on the screening results of CA14-DB21-based sensors, several variants exhibited F/F_0 > 200%, or equivalently $\Delta F/F_0$ > 100%. Given that the bacterial expression of N14C2 produced slightly more protein than other variants, we focused on N14C2 for further testing.

We displayed N14C2 on the plasma membrane of HEK293T cells using the pAEMXT vector (82), as described earlier, and monitored its response to CBD. To assess whether the sensor maintained its specificity in mammalian systems, we also titrated N14C2 with THC. In the presence of 30 μM CBD, the fluorescence intensity gradually increased over 150 seconds (Figure 15a, c), while no significant change was observed with 30 μM THC over 300 seconds (Figure 15b, c), indicating no response to THC. This lack of response to THC aligns with the weak *in vitro* response and may reflect subtle differences in protein folding in mammalian systems. In a separate experiment, additional titrations of N14C2 with CBD

revealed a maximum $\Delta F/F_0$ of approximately 100% and an EC_{50} of ~213 nM (Figure 15d), confirming that the sensor remained functional and specific in mammalian cells.

It is noteworthy that N14C2 requires over 100 seconds to reach its maximum $\Delta F/F_0$. The slow kinetics may be due to the restricted mobility of the surface-anchored sensors, which could slow down intermolecular dimerization and oligomerization. Varying the length of the sensor-anchor linkers could provide insight into this phenomenon. Moreover, since MP is cell-impermeable and CBD diffusion takes time, the HEK293T system is not ideal for measuring intracellular sensor kinetics, an issue we hope to address in the future.

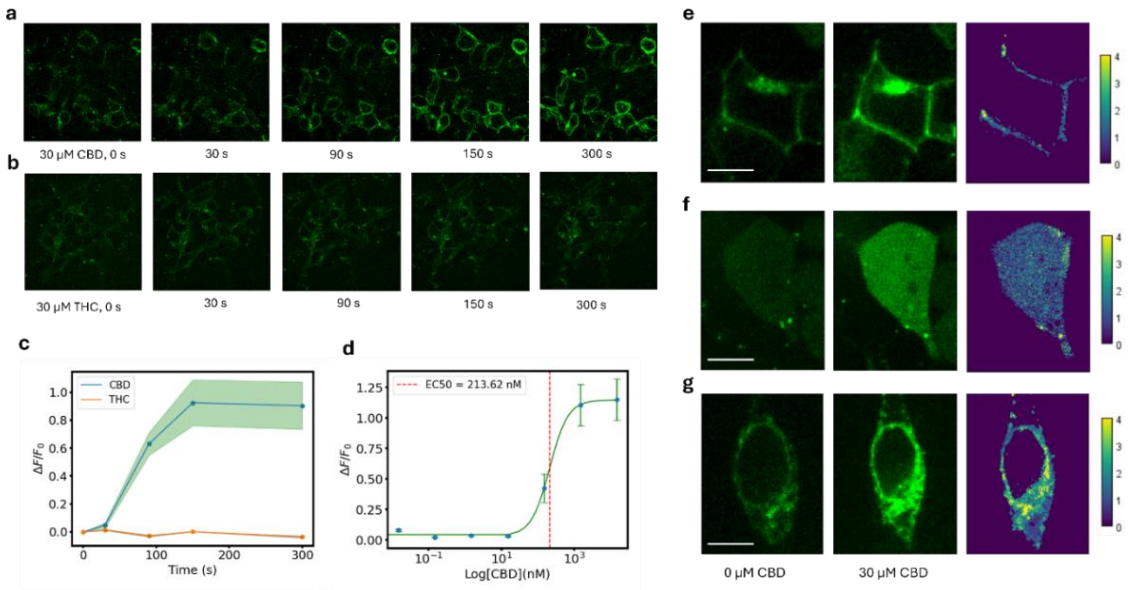


Figure 15. Express and characterize N14C2 in cellular environments. (a) N14C2 displayed on the surface of HEK293T responding to 30 μM CBD over 300 s. (b) N14C2 displayed on the surface of HEK293T responding to 30 μM THC over 300 sec. (c) $\Delta F/F_0$ of N14C2 in the presence of 30 μM CBD vs. THC over 300 s, n=10. (d) Titration of N14C2 on the surface of HEK293T against various concentrations of CBD, n=8. (e), (f), (g) N14C2 expressed on the surface (e), in the cytoplasm (f) and in the ER lumen (g) of HEK293T responding to 30 μM CBD. Scale bar: 20 μM. Calibration bar: $\Delta F/F_0$.

To test the flexibility of nanobody-scaffold sensors in different subcellular locations, we also targeted N14C2 to the cytoplasm and ER lumen of HEK293T cells. The sensor was successfully expressed on the plasma membrane (Figure 15e), in the cytosol (Figure 15f), and in the ER lumen (Figure 15g), and 30 μM CBD led to increased fluorescence in all cases, confirming sensor functionality in each location.

These findings demonstrate that our nanobody-scaffold sensors are compatible with HEK293T cells and retain functionality across different subcellular compartments. Overall, our experiments provide compelling evidence that synthetic CID-based GEFSs hold significant potential for applications in mammalian cell systems.

3.4. Conclusion

To integrate synthetic CIDs into GEFSs, we designed and constructed a flexible linker library. The aim was to optimize linker length to provide CID pairs sufficient mobility for dimerization. Our linker library comprised GS linkers ranging from 2 to 20 amino acids, resulting in 361 unique length combinations. This design allowed us to explore the relationship between linker length and sensor dynamic range while maintaining flexibility.

We tested the linker library by inserting it into three CID pairs: CA14-DB21, oAS28-6, and oAS14-89, generating corresponding sensor libraries. These libraries were transformed into *E. coli* and assayed in 96-deep-well plates. All three sensor libraries exhibited consistent and detectable fluorescence signals, indicating robust protein expression and demonstrating that our linker library is generalizable across different synthetic scaffolds.

Under saturating ligand conditions, sensor variants displayed varying levels of response, showing a clear correlation between linker length and dynamic range. Notably, the C-terminal linker length significantly affected both basal fluorescence and dynamic range, underscoring the importance of C-terminal linkers in modulating the cpGFP microenvironment. Across all three CID pairs, the C-linkers yielding the highest basal fluorescence were consistently two amino acids longer than those producing the highest dynamic range. This effect appeared to be scaffold-independent, as it was observed in both the nanobody- and monobody-based scaffolds. Structural predictions from AlphaFold suggest that this might be due to the proximity of the C-terminal binder's binding pocket to cpGFP.

We identified sensor variants with a ~100% $\Delta F/F_0$, validating the effectiveness of our linker library and screening approach. *In vitro* characterization of high-dynamic-range sensor variants revealed that their

dynamic range was concentration-dependent, which was confirmed by titrating sensors at varying concentrations. Size-exclusion chromatography (SEC) profiles indicated that the sensing mechanism involved both inter- and intramolecular interactions, though more detailed mechanisms will require further study.

We then validated the sensors in HEK293T cells, expressing them in the cytoplasm, ER lumen, and on the plasma membrane. The results showed that our sensors were compatible with mammalian expression systems. When displayed on the plasma membrane, we were able to titrate sensors, achieving EC_{50} values comparable to those observed *in vitro*, confirming that the sensors remained functional in mammalian cells.

Overall, these results demonstrate the generalizability of our linker library and the compatibility of synthetic CID-based sensors with mammalian expression systems.

4. Method Generalization with cpHaloTag as the Optical Domain

4.1. Introduction

While generating monobody- and nanobody-scaffold sensors has demonstrated our linker library's adaptability across different CID scaffolds, we have yet to determine if this versatility extends to other optical domains. For example, although less commonly used than cpGFP, JF635-labeled cpHaloTag operates in the red- light spectrum, which generally penetrates deeper into tissues than green light. If our methodology can be applied to cpHaloTag-based GEFS engineering, it would expand our toolkit for various applications. Despite cpGFP and cpHaloTag sharing similar principles, the specific interface between the sensory domain and optical domain can subtly impact sensor functionality. Therefore, we decided to replace cpGFP with cpHaloTag in our sensor design to see if the same method could yield sensor variants with approximately 100% $\Delta F/F_0$.

Additionally, given the critical role of the C-terminal linker in CID-based sensors, we investigated whether we could streamline the linker screening process. Our approach involved initially screening a small set of short C-terminal linkers, followed by testing a small set of N-terminal linkers with incremental length variations. This approach was informed by previous results showing that optimal C-terminal linker lengths were either 2 or 3 amino acids, while N-terminal linker lengths displayed broader variability.

4.2. Material and Methods

4.2.1. Protein expression and purification

All protein constructs were expressed in E. coli using a pBAD vector, with a C-terminal Avi-tag and His-tag and were subsequently purified by Ni-affinity chromatography. In brief, E. coli strain DH10B was transformed with the expression constructs using heat shock. Transformed cells were grown in 2xYT media containing 100 μg/mL ampicillin at 37˚C until reaching an OD600 of ~0.6. Protein expression was then induced with 0.02% arabinose, and cultures were incubated overnight at 18˚C. Cell pellets from 1-liter cultures were harvested by centrifugation at 8000 g at 4 ˚C, resuspended in 15 mL ice-cold His-wash buffer (50 mM sodium phosphate, pH 8.0, 300 mM NaCl, 20 mM imidazole, 10% glycerol), and kept on ice. Bacterial cells were lysed using a Branson cell sonicator at 25% amplitude for 5 minutes with 5

seconds on and 5 seconds off, in an ice bath. The lysate was then centrifuged at 10,000 g, 4˚C for 20 minutes to remove cell debris, followed by a second centrifugation of the supernatant at 20,000 g, 4˚C for 30 minutes. The clarified supernatant was filtered through a 0.45 µm syringe filter (Millipore) at room temperature to remove any remaining particulates The filtered supernatant was loaded onto a 1 mL HisTrap column (GE Healthcare) pre-equilibrated with His-wash buffer at a flow rate of 1 mL/min. The column was washed with 30 mL of His-wash buffer at 1 mL/min, and His-tagged proteins were eluted in a gradient manner, at ~ 50% of elution buffer (50 mM sodium phosphate, pH 8.0, 300 mM NaCl, 250 mM imidazole, 10% glycerol). Eluted proteins were desalted using a 50 mL Hi-Trap desalting column (GE Healthcare), equilibrated with PBS buffer with 5% PBS, at a flow rate of 5 mL/min. Desalted proteins were examined by SDS-PAGE.

4.2.2. Sensor labeling with fluorogenic dye *in vitro*

To maximize labeling efficiency and minimize background, a 2:1 protein-to-dye molar ratio was used. Typically, 1 µM JF635 dye was added to 2 µM purified protein in labeling buffer (50 mM HEPES, 50 mM NaCl, pH 7.4) and incubated for 30 minutes at room temperature. Given JF635's low background, the labeled sensor can be directly characterized without further purification. If JF635i was used, the labeled protein requires desalting with a HiTrap desalting column (Cytiva).

4.2.3. Sensor characterization *in vitro*

To get dose-response curves, various concentrations of MP were added to 100 µL sensor protein solution of 1 µM in a black 96-well microplate, mixed at 1000 rpm on a benchtop plate shaker for 30 seconds, and measured by a fluorescent plate reader. The MP concentrations are 300, 100, 33.3,11.1, 3.7,1.2, 0.4, 0 µM. In the case where only maximum response is concerned, 30 and 100 µM of MP were added to sensor variants using oAS28-6 and oAS14-89 CID pair, respectively. Data was analyzed using excel. Briefly, the average and standard deviation for each triplicate were calculated; for dose response curves, nonlinear regression was performed using a four-parameter model (d + (a - d) / (1 + (x / c) ^b)). All plots were generated using MatPlotLib.

4.3. Results

To validate the compatibility of our linker library with different optical domains, we aimed to replace cpGFP in our sensor design with circularly permuted HaloTag (cpHaloTag). HaloTag is a self-labeling protein (74) that becomes fluorescent when labeled with a dye. In cpHaloTag, the termini are repositioned near the fluorophore binding site; when labeled with the highly fluorogenic dye JF635, cpHaloTag becomes fluorescent and is sensitive to its chemical environment. Like cpGFP, cpHaloTag's fluorescence intensity can be modulated by the sensory domain through allosteric coupling (73, 75, 169). To test this, we fused an MP-binding CID with cpHaloTag, following protocols described in the literature (76). We opted for the MP-binding CID over the CBD-binding CID to minimize potential folding issues.

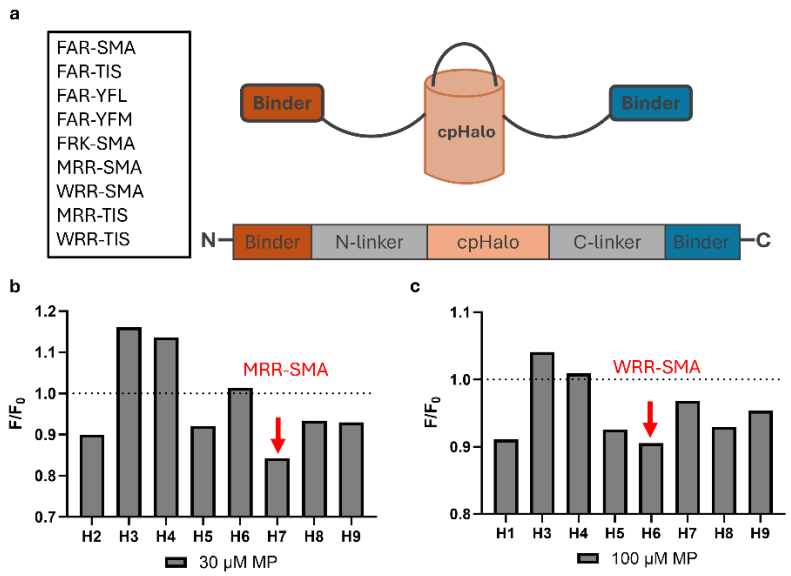


Figure 1. Screening linkers for cpHaloTag-based sensors. (a) Schematics of cpHaloTag-based sensor and linker list. (b) Screening results for oAS28-6 CID pair. (c) Screening results for oAS14-89 CID pair.

Linkers in cpHaloTag-sensors are crucial to sensor dynamic range, just like in cpGFP-sensors. Optimization of linkers, especially amino acid composition, can take considerable amount of time. To simplify the problem, we decided to first screen a small set of short linkers reported in literature (76) (Figure 1a). We reasoned that even if sensory domains are drastically different, linkers work for one sensor might as well retain partial functionality in the other, due to shared cpHaloTag domain. Screening

these linkers for oAS28-6 and oAS14-89 revealed two linker pairs, MRR-SMA and WRR-SMA, that led to a 10-15% decrease in fluorescent intensity (Figure 1b, 1c). Although the signal shift was small, we reasoned that it was sufficient as a start point, because we have seen in previous experiments that linker lengths can drastically change the dynamic range of CID-based sensors. Therefore, we proceeded to optimize linker lengths with these linkers.

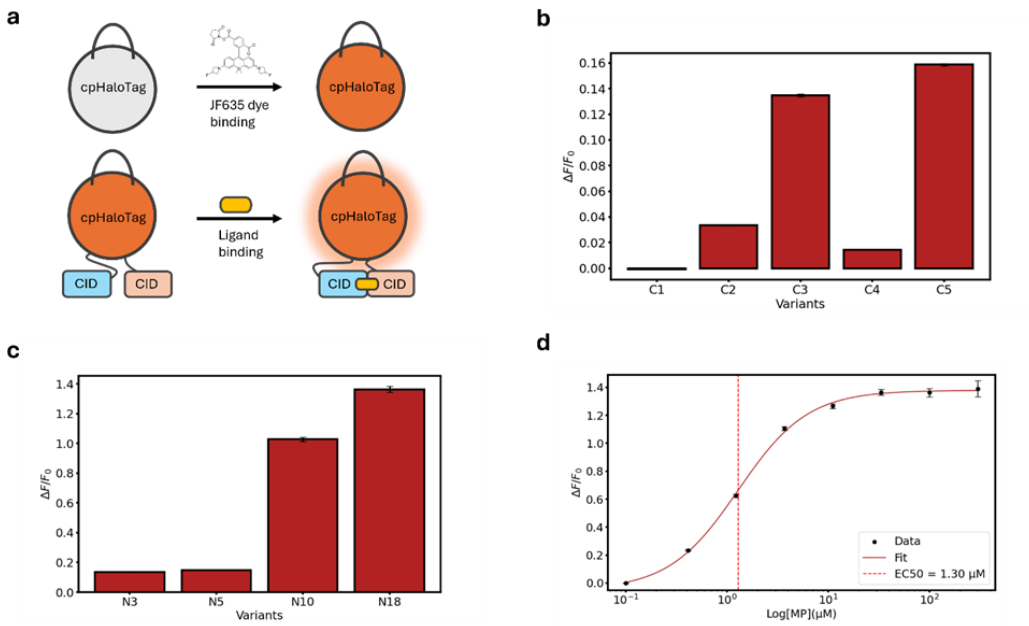


Figure 2. (a) Schematic of cpHaloTag labeling and CID-cpHaloTag sensors. (b) $\Delta F/F_0$ at 30 μM MP of MP sensor variants with various length of C-terminal linkers. (c) $\Delta F/F_0$ at 30 μM MP of MP sensor variants with various length of N-terminal linkers. (d) Dose response curve of MP sensor N18C3 sensing MP.

Having decided linkers adjacent to cpHaloTag, we proceeded to add various length of GS-extensions to linkers on both ends of the optical domain. To streamline the linker screening workflow, we began by screening C-terminal linkers ranging from 1 to 5 AAs, while keeping the N-terminal linker fixed at 3 AAs. We found that C3 and C5 variants had similar performance (Figure 2b). We then fixed the C-terminal linker, and screened N-terminal linkers of 5, 10 and 18 AAs, and found that the N18 variant had significantly improved $\Delta F/F_0$ compared to the N3 variant, increasing from less than 20% to more than 100% (Figure 2c). We further titrated the newly found N18C3 variant with various concentrations of ligand

to further characterize its performance and confirmed the maximum $\Delta F/F_0$ ~130% (Figure 2d). To get this improved sensor, in total we simply tested 9 variants.

These results indicate that our linker library method is generalizable across various synthetic CID scaffolds and optical domains. Moreover, insights from the linker library screening enable rapid enhancement of sensor performance in future, where testing a small set of variants instead of screening the whole library is sufficient to improve $\Delta F/F_0$ above 100%.

4.4. Conclusion

Using the same method we employed to engineer cpGFP-based GEFSs, we developed cpHaloTag-based GEFSs. The primary difference in this workflow was that, instead of screening a large library with hundreds of linker length combinations, we tested only a small, sequential set of linkers on both termini. Using the oAS28-6 CID pair as a proof of concept, we screened just nine variants and identified the N18C3 variant, which achieved a $\Delta F/F_0$ of approximately 120%. These results suggest that our method can be successfully extended to other optical domains. By applying insights from previous screenings, we can rapidly enhance sensor performance by screening only a limited set of variants.

Discussion

Despite being powerful biomedical research tools, GEFSs remain challenging to engineer due to lack of Genetically encoded fluorescent sensors (GEFSs) are powerful tools in biomedical research, but they remain challenging to engineer due to the lack of generalizable platforms and methodologies. Synthetic protein binders offer a potential solution, and in this work, we have developed a versatile and efficient method for engineering synthetic CIDs into GEFSs using a variety of protein scaffolds and optical domains. This approach has proven effective in creating sensors that can be targeted not only to the plasma membrane or cytoplasm but also to specific subcellular locations, such as the ER lumen. By integrating our previously established synthetic CID selection method, we have created a comprehensive pipeline for the de novo engineering of CID-cpGFP-based fluorescent biosensors tailored to specific small molecules and peptide ligands.

Our work does have limitations. For example, in CBD sensor engineering, the nanobody scaffold we used encountered expression issues when fused with optical domains. We addressed this by introducing amino acid mutations in the scaffold sequence to stabilize protein folding, and we aim to integrate these stabilizing mutations into our nanobody libraries in the future. Additionally, there is room for improvement in kinetics. While GEFSs like GCaMP exhibit millisecond kinetics *in vivo*, our CBD sensor, N14C2, requires over 100 seconds to reach maximum fluorescence in cultured cells—likely due to mobility constraints from the display system. Slow kinetics like this limit sensor applications, and we plan to investigate the sensing mechanism further to address this issue.

Despite these limitations, our method holds significant promise. A key advantage of our pipeline is its use of entirely synthetic sensory domains, removing the need for naturally occurring protein sources. This is particularly beneficial when the ligand of interest lacks a well-studied natural binder with both allostery and resolved structures. Furthermore, since CIDs are naturally dimerized, their fusion with optical domains does not require extensive insertion site optimization. This feature enables rapid, systematic optimization of sensor performance without the need for detailed structural knowledge. Our method is highly adaptable across different scaffolds and optical domains; more importantly, insights gained from

systematic linker library screening allow us to avoid screening a large library. We can rapidly optimize linkers by testing a small, targeted subset of variants.

In conclusion, this work represents a significant advance in the engineering of GEFSs with synthetic sensory domains, offering a more generalizable and efficient approach to biosensor development.

References

1. Hochreiter, B.; Garcia, A. P.; Schmid, J. A. Fluorescent Proteins as Genetically Encoded FRET Biosensors in Life Sciences. Sensors (Basel) 2015, 15 (10), 26281–26314. https://doi.org/10.3390/s151026281.
2. Xu, X.; Gerard, A. L.; Huang, B. C.; Anderson, D. C.; Payan, D. G.; Luo, Y. Detection of Programmed Cell Death Using Fluorescence Energy Transfer. Nucleic Acids Res. 1998, 26 (8), 2034–2035. https://doi.org/10.1093/nar/26.8.2034.
3. Hwang, Y.-C.; Chen, W.; Yates, M. V. Use of Fluorescence Resonance Energy Transfer for Rapid Detection of Enteroviral Infection *in vivo*. Appl. Environ. Microbiol. 2006, 72 (5), 3710–3715. https://doi.org/10.1128/aem.72.5.3710-3715.2006.
4. Miyawaki, A.; Llopis, J.; Heim, R.; McCaffery, J. M.; Adams, J. A.; Ikura, M.; Tsien, R. Y. Fluorescent Indicators for Ca2+based on Green Fluorescent Proteins and Calmodulin. Nature 1997, 388 (6645), 882–887. https://doi.org/10.1038/42264.
5. Vinkenborg, J. L.; Koay, M. S.; Merkx, M. Fluorescent Imaging of Transition Metal Homeostasis Using Genetically Encoded Sensors. Curr. Opin. Chem. Biol. 2010, 14 (2), 231–237. https://doi.org/10.1016/j.cbpa.2009.11.022.
6. DiPilato, L. M.; Cheng, X.; Zhang, J. Fluorescent Indicators of CAMP and Epac Activation Reveal Differential Dynamics of CAMP Signaling within Discrete Subcellular Compartments. Proc. Natl. Acad. Sci. U. S. A. 2004, 101 (47), 16513–16518. https://doi.org/10.1073/pnas.0405973101.
7. Fehr, M.; Frommer, W. B.; Lalonde, S. Visualization of Maltose Uptake in Living Yeast Cells by Fluorescent Nanosensors. Proc. Natl. Acad. Sci. U. S. A. 2002, 99 (15), 9846–9851. https://doi.org/10.1073/pnas.142089199.
8. Mohsin, M.; Ahmad, A.; Iqbal, M. FRET-Based Genetically-Encoded Sensors for Quantitative Monitoring of Metabolites. Biotechnol. Lett. 2015, 37 (10), 1919–1928. https://doi.org/10.1007/s10529-015-1873-6.
9. Zhang, J.; Ma, Y.; Taylor, S. S.; Tsien, R. Y. Genetically Encoded Reporters of Protein Kinase A Activity Reveal Impact of Substrate Tethering. Proc. Natl. Acad. Sci. U. S. A. 2001, 98 (26), 14997–15002. https://doi.org/10.1073/pnas.211566798.

10. Zhou, X.; Herbst-Robinson, K. J.; Zhang, J. Visualizing Dynamic Activities of Signaling Enzymes Using Genetically Encodable Fret-Based Biosensors. *Imaging and Spectroscopic Analysis of Living Cells - Optical and Spectroscopic Techniques* **2012**, 317–340. https://doi.org/10.1016/b978-0-12-391857-4.00016-1.
11. Kneen, M.; Farinas, J.; Li, Y.; Verkman, A. S. Green Fluorescent Protein as a Noninvasive Intracellular PH Indicator. *Biophysical Journal* **1998**, *74* (3), 1591–1599. https://doi.org/10.1016/s0006-3495(98)77870-1.
12. Sankaranarayanan, S.; De Angelis, D.; Rothman, J. E.; Ryan, T. A. The Use of PHluorins for Optical Measurements of Presynaptic Activity. *Biophysical Journal* **2000**, *79* (4), 2199–2208. https://doi.org/10.1016/s0006-3495(00)76468-x.
13. Shen, Y.; Rosendale, M.; Campbell, R. E.; Perrais, D. PHuji, a PH-Sensitive Red Fluorescent Protein for Imaging of Exo- and Endocytosis. *Journal of Cell Biology* **2014**, *207* (3), 419–432. https://doi.org/10.1083/jcb.201404107.
14. Hanson, G. T.; Aggeler, R.; Oglesbee, D.; Cannon, M.; Capaldi, R. A.; Tsien, R. Y.; Remington, S. J. Investigating Mitochondrial Redox Potential with Redox-Sensitive Green Fluorescent Protein Indicators. *Journal of Biological Chemistry* **2004**, *279* (13), 13044–13053. https://doi.org/10.1074/jbc.m312846200.
15. Baird, G. S.; Zacharias, D. A.; Tsien, R. Y. Circular Permutation and Receptor Insertion within Green Fluorescent Proteins. *Proceedings of the National Academy of Sciences* **1999**, *96* (20), 11241–11246. https://doi.org/10.1073/pnas.96.20.11241.
16. Griesbeck, O.; Baird, G. S.; Campbell, R. E.; Zacharias, D. A.; Tsien, R. Y. Reducing the Environmental Sensitivity of Yellow Fluorescent Protein. *Journal of Biological Chemistry* **2001**, *276* (31), 29188–29194. https://doi.org/10.1074/jbc.m102815200.
17. Nakai, J.; Ohkura, M.; Imoto, K. A High Signal-To-Noise Ca2+ Probe Composed of a Single Green Fluorescent Protein. *Nature Biotechnology* **2001**, *19* (2), 137–141. https://doi.org/10.1038/84397.
18. Tallini, Y. N.; Ohkura, M.; Choi, B.-R.; Ji, G.; Imoto, K.; Doran, R.; Lee, J.; Plan, P.; Wilson, J.; Xin, H.-B.; Sanbe, A.; Gulick, J.; Mathai, J.; Robbins, J.; Salama, G.; Nakai, J.; Kotlikoff, M. I. Imaging Cellular Signals in the Heart *in vivo* : Cardiac Expression of the High-Signal ca $^{2+}$ Indicator

GCaMP2. *Proceedings of the National Academy of Sciences* **2006**, *103* (12), 4753–4758. https://doi.org/10.1073/pnas.0509378103.

19. Tian, L.; Hires, S. A.; Mao, T.; Huber, D.; Chiappe, M. E.; Chalasani, S. H.; Petreanu, L.; Akerboom, J.; McKinney, S. A.; Schreiter, E. R.; Bargmann, C. I.; Jayaraman, V.; Svoboda, K.; Looger, L. L. Imaging Neural Activity in Worms, Flies and Mice with Improved GCaMP Calcium Indicators. *Nature methods* **2009**, *6* (12), 875–881. https://doi.org/10.1038/nmeth.1398.
20. Akerboom, J.; Chen, T.-W.; Wardill, T. J.; Tian, L.; Marvin, J. S.; Mutlu, S.; Calderón, N. C.; Esposti, F.; Borghuis, B. G.; Sun, X. R.; Gordus, A.; Orger, M. B.; Portugues, R.; Engert, F.; Macklin, J. J.; Filosa, A.; Aggarwal, A.; Kerr, R. A.; Takagi, R.; Kracun, S. Optimization of a GCaMP Calcium Indicator for Neural Activity Imaging. *The Journal of Neuroscience* **2012**, *32* (40), 13819–13840. https://doi.org/10.1523/JNEUROSCI.2601-12.2012.
21. Chen, T.-W.; Wardill, T. J.; Sun, Y.; Pulver, S. R.; Renninger, S. L.; Baohan, A.; Schreiter, E. R.; Kerr, R. A.; Orger, M. B.; Jayaraman, V.; Looger, L. L.; Svoboda, K.; Kim, D. S. Ultrasensitive Fluorescent Proteins for Imaging Neuronal Activity. *Nature* **2013**, *499* (7458), 295–300. https://doi.org/10.1038/nature12354.
22. Dana, H.; Sun, Y.; Mohar, B.; Hulse, B. K.; Kerlin, A. M.; Hasseman, J. P.; Tsegaye, G.; Tsang, A.; Wong, A.; Patel, R.; Macklin, J. J.; Chen, Y.; Konnerth, A.; Jayaraman, V.; Looger, L. L.; Schreiter, E. R.; Svoboda, K.; Kim, D. S. High-Performance Calcium Sensors for Imaging Activity in Neuronal Populations and Microcompartments. *Nature Methods* **2019**, *16* (7), 649–657. https://doi.org/10.1038/s41592-019-0435-6.
23. Kostyuk, A. I.; Demidovich, A. D.; Kotova, D. A.; Belousov, V. V.; Bilan, D. S. Circularly Permuted Fluorescent Protein-Based Indicators: History, Principles, and Classification. *International Journal of Molecular Sciences* **2019**, *20* (17), 4200. https://doi.org/10.3390/ijms20174200.
24. Marvin, J. S.; Borghuis, B. G.; Tian, L.; Cichon, J.; Harnett, M. T.; Akerboom, J.; Gordus, A.; Renninger, S. L.; Chen, T.-W.; Bargmann, C. I.; Orger, M. B.; Schreiter, E. R.; Demb, J. B.; Gan, W.-B.; Hires, S. A.; Looger, L. L. An Optimized Fluorescent Probe for Visualizing Glutamate Neurotransmission. *Nature Methods* **2013**, *10* (2), 162–170. https://doi.org/10.1038/nmeth.2333.

25. Marvin, J. S.; Scholl, B.; Wilson, D. E.; Podgorski, K.; Kazemipour, A.; Müller, J. A.; Schoch, S.; Quiroz, F. J. U.; Rebola, N.; Bao, H.; Little, J. P.; Tkachuk, A. N.; Cai, E.; Hantman, A. W.; Wang, S. S.-H. .; DePiero, V. J.; Borghuis, B. G.; Chapman, E. R.; Dietrich, D.; DiGregorio, D. A. Stability, Affinity, and Chromatic Variants of the Glutamate Sensor IGluSnFR. *Nature Methods* **2018**, *15* (11), 936–939. https://doi.org/10.1038/s41592-018-0171-3.

26. Clements, J. D.; Westbrook, G. L. Activation Kinetics Reveal the Number of Glutamate and Glycine Binding Sites on the N-Methyl-d-Aspartate Receptor. *Neuron* **1991**, *7* (4), 605–613. https://doi.org/10.1016/0896-6273(91)90373-8.

27. Patriarchi, T.; Cho, J. R.; Merten, K.; Howe, M. W.; Marley, A.; Xiong, W.-H.; Folk, R. W.; Broussard, G. J.; Liang, R.; Jang, M. J.; Zhong, H.; Dombeck, D.; von Zastrow, M.; Nimmerjahn, A.; Gradinaru, V.; Williams, J. T.; Tian, L. Ultrafast Neuronal Imaging of Dopamine Dynamics with Designed Genetically Encoded Sensors. *Science (New York, N.Y.)* **2018**, *360* (6396). https://doi.org/10.1126/science.aat4422.

28. Patriarchi, T.; Mohebi, A.; Sun, J.; Marley, A.; Liang, R.; Dong, C.; Puhger, K.; Mizuno, G. O.; Davis, C. M.; Wiltgen, B.; von Zastrow, M.; Berke, J. D.; Tian, L. An Expanded Palette of Dopamine Sensors for Multiplex Imaging *in vivo*. *Nature Methods* **2020**, *17* (11), 1147–1155. https://doi.org/10.1038/s41592-020-0936-3.

29. Sun, F.; Zeng, J.; Jing, M.; Zhou, J.; Feng, J.; Owen, S. F.; Luo, Y.; Li, F.; Wang, H.; Yamaguchi, T.; Yong, Z.; Gao, Y.; Peng, W.; Wang, L.; Zhang, S.; Du, J.; Lin, D.; Xu, M.; Kreitzer, A. C.; Cui, G. A Genetically Encoded Fluorescent Sensor Enables Rapid and Specific Detection of Dopamine in Flies, Fish, and Mice. *Cell* **2018**, *174* (2), 481-496.e19. https://doi.org/10.1016/j.cell.2018.06.042.

30. Lobas, M. A.; Tao, R.; Nagai, J.; Kronschläger, M. T.; Borden, P. M.; Marvin, J. S.; Looger, L. L.; Khakh, B. S. A Genetically Encoded Single-Wavelength Sensor for Imaging Cytosolic and Cell Surface ATP. *Nature Communications* **2019**, *10* (1). https://doi.org/10.1038/s41467-019-08441-5.

31. Marvin, J. S.; Kokotos, A. C.; Kumar, M.; Pulido, C.; Tkachuk, A. N.; Jocelyn Shuxin Yao; Brown, T. A.; Ryan, T. A. IATPSnFR2: A High-Dynamic-Range Fluorescent Sensor for Monitoring Intracellular ATP. *Proceedings of the National Academy of Sciences of the United States of America* **2024**, *121* (21). https://doi.org/10.1073/pnas.2314604121.

32. Belousov, V. V.; Fradkov, A. F.; Lukyanov, K. A.; Staroverov, D. B.; Shakhbazov, K. S.; Terskikh, A. V.; Lukyanov, S. Genetically Encoded Fluorescent Indicator for Intracellular Hydrogen Peroxide. *Nature Methods* **2006**, *3* (4), 281–286. https://doi.org/10.1038/nmeth866.
33. Markvicheva, K. N.; Bilan, D. S.; Mishina, N. M.; Andrey Yu. Gorokhovatsky; Vinokurov, L. M.; Sergey Lukyanov; Belousov, V. V. A Genetically Encoded Sensor for H2O2 with Expanded Dynamic Range. *Bioorganic & Medicinal Chemistry* **2011**, *19* (3), 1079–1084. https://doi.org/10.1016/j.bmc.2010.07.014.
34. Bilan, D. S.; Pase, L.; Joosen, L. P.; Andrey Yu. Gorokhovatsky; Ermakova, Y. G.; Theodorus W. J. Gadella; Grabher, C.; Schultz, C.; Sergey Lukyanov; Belousov, V. V. HyPer-3: A Genetically Encoded H_2O_2 Probe with Improved Performance for Ratiometric and Fluorescence Lifetime Imaging. **2013**, *8* (3), 535–542. https://doi.org/10.1021/cb300625g.
35. François St-Pierre; Marshall, J. D.; Yang, Y.; Gong, Y.; Schnitzer, M. J.; Lin, M. Z. High-Fidelity Optical Reporting of Neuronal Electrical Activity with an Ultrafast Fluorescent Voltage Sensor. *Nature Neuroscience* **2014**, *17* (6), 884–889. https://doi.org/10.1038/nn.3709.
36. Yang, Helen H.; St-Pierre, F.; Sun, X.; Ding, X.; Lin, Michael Z.; Clandinin, Thomas R. Subcellular Imaging of Voltage and Calcium Signals Reveals Neural Processing *in vivo*. *Cell* **2016**, *166* (1), 245–257. https://doi.org/10.1016/j.cell.2016.05.031.
37. Chamberland, S.; Yang, H. H.; Pan, M. M.; Evans, S. W.; Guan, S.; Mariya Chavarha; Yang, Y.; Salesse, C.; Wu, H.; Wu, J. C.; Clandinin, T. R.; Toth, K.; Lin, M. Z.; François St-Pierre. Fast Two-Photon Imaging of Subcellular Voltage Dynamics in Neuronal Tissue with Genetically Encoded Indicators. *eLife* **2017**, *6*. https://doi.org/10.7554/elife.25690.
38. Villette, V.; Chavarha, M.; Dimov, I. K.; Bradley, J.; Pradhan, L.; Mathieu, B.; Evans, S. W.; Chamberland, S.; Shi, D.; Yang, R.; Kim, B. B.; Ayon, A.; Jalil, A.; St-Pierre, F.; Schnitzer, M. J.; Bi, G.; Toth, K.; Ding, J.; Dieudonné, S.; Lin, M. Z. Ultrafast Two-Photon Imaging of a High-Gain Voltage Indicator in Awake Behaving Mice. *Cell* **2019**, *179* (7), 1590-1608.e23. https://doi.org/10.1016/j.cell.2019.11.004.

39. Elizabeth E.L. Lee; Bezanilla, F. Biophysical Characterization of Genetically Encoded Voltage Sensor ASAP1: Dynamic Range Improvement. *Biophysical Journal* **2017**, *113* (10), 2178–2181. https://doi.org/10.1016/j.bpj.2017.10.018.

40. Cornejo, V. H.; Ofer, N.; Yuste, R. Voltage Compartmentalization in Dendritic Spines *in vivo*. *Science* **2022**, *375* (6576), 82–86. https://doi.org/10.1126/science.abg0501.

41. S. Wenceslao Evans; Shi, D.-Q.; Mariya Chavarha; Plitt, M. H.; Jiannis Taxidis; Madruga, B.; Jiang Lan Fan; Hwang, F.-J.; Keulen, van; Carl-Mikael Suomivuori; Pang, M. M.; Su, S.; Lee, S.; Hao, Y. A.; Zhang, G.; Jiang, D.; Pradhan, L.; Roth, R. H.; Liu, Y.; Dorian, C. C. A Positively Tuned Voltage Indicator for Extended Electrical Recordings in the Brain. *Nature methods* **2023**, *20* (7), 1104–1113. https://doi.org/10.1038/s41592-023-01913-z.

42. Philipp Rühl; Nair, A. G.; Namrata Gawande; Sassrika N. C. W. Dehiwalage; Lukas Münster; Schönherr, R.; Heinemann, S. H. An Ultrasensitive Genetically Encoded Voltage Indicator Uncovers the Electrical Activity of Non-Excitable Cells. *Advanced Science* **2024**. https://doi.org/10.1002/advs.202307938.

43. Edwards, K. A. Periplasmic-Binding Protein-Based Biosensors and Bioanalytical Assay Platforms: Advances, Considerations, and Strategies for Optimal Utility. *Talanta Open* **2021**, *3*, 100038. https://doi.org/10.1016/j.talo.2021.100038.

44. Marvin, J. S.; Schreiter, E. R.; Echevarría, I. M.; Looger, L. L. A Genetically Encoded, High-Signal-To-Noise Maltose Sensor. *Proteins: Structure, Function, and Bioinformatics* **2011**, *79* (11), 3025–3036. https://doi.org/10.1002/prot.23118.

45. Keller, J. P.; Marvin, J. S.; Haluk Lacin; Lemon, W. C.; Shea, J.; Kim, S.; Lee, R. T.; Koyama, M.; Keller, P. J.; Looger, L. L. *In vivo* Glucose Imaging in Multiple Model Organisms with an Engineered Single-Wavelength Sensor. *Cell Reports* **2021**, *35* (12), 109284–109284. https://doi.org/10.1016/j.celrep.2021.109284.

46. Hu, H.; Wei, Y.; Wang, D.; Su, N.; Chen, X.; Zhao, Y.; Liu, G.; Yang, Y. Glucose Monitoring in Living Cells with Single Fluorescent Protein-Based Sensors. *RSC Advances* **2018**, *8* (5), 2485–2489. https://doi.org/10.1039/C7RA11347A.

47. Alicea, I.; Marvin, J. S.; Miklos, A. E.; Ellington, A. D.; Looger, L. L.; Schreiter, E. R. Structure of the Escherichia Coli Phosphonate Binding Protein PhnD and Rationally Optimized Phosphonate Biosensors. *Journal of Molecular Biology* **2011**, *414* (3), 356–369. https://doi.org/10.1016/j.jmb.2011.09.047.

48. Hu, H.; Gu, Y.; Xu, L.; Zou, Y.; Wang, A.; Tao, R.; Chen, X.; Zhao, Y.; Yang, Y. A Genetically Encoded Toolkit for Tracking Live-Cell Histidine Dynamics in Space and Time. *Scientific Reports* **2017**, *7* (1). https://doi.org/10.1038/srep43479.

49. Marvin, J. S.; Shimoda, Y.; Magloire, V.; Leite, M.; Kawashima, T.; Jensen, T. P.; Kolb, I.; Knott, E. L.; Novak, O.; Podgorski, K.; Leidenheimer, N. J.; Rusakov, D. A.; Ahrens, M. B.; Kullmann, D. M.; Looger, L. L. A Genetically Encoded Fluorescent Sensor for *in vivo* Imaging of GABA. *Nature Methods* **2019**, *16* (8), 763–770. https://doi.org/10.1038/s41592-019-0471-2.

50. Borden, P.; Zhang, P.; Shivange, A. V.; Marvin, J. S.; Cichon, J.; Dan, C.; Podgorski, K.; Figueiredo, A.; Ondřej Novák; Tanimoto, M.; Eiji Shigetomi; Lobas, M. A.; Kim, H.-T.; Zhu, P.; Zhang, Y.; Zheng, W.; Fan, C.; Wang, G.; Xiang, B.; Gan, L. A Fast Genetically Encoded Fluorescent Sensor for Faithful *in vivo* Acetylcholine Detection in Mice, Fish, Worms and Flies. *Social Science Research Network* **2020**. https://doi.org/10.2139/ssrn.3554080.

51. Latorraca, N. R.; Venkatakrishnan, A. J.; Dror, R. O. GPCR Dynamics: Structures in Motion. *Chemical Reviews* **2016**, *117* (1), 139–155. https://doi.org/10.1021/acs.chemrev.6b00177.

52. Jing, M.; Zhang, P.; Wang, G.; Feng, J.; Mesik, L.; Zeng, J.; Jiang, H.; Wang, S.; Looby, J. C.; Guagliardo, N. A.; Langma, L. W.; Lu, J.; Zuo, Y.; Talmage, D. A.; Role, L. W.; Barrett, P. Q.; Zhang, L. I.; Luo, M.; Song, Y.; Zhu, J. J. A Genetically Encoded Fluorescent Acetylcholine Indicator for *in vitro* and *in vivo* Studies. *Nature Biotechnology* **2018**, *36* (8), 726–737. https://doi.org/10.1038/nbt.4184.

53. Sun, F.; Zeng, J.; Jing, M.; Zhou, J.; Feng, J.; Owen, S. F.; Luo, Y.; Li, F.; Wang, H.; Yamaguchi, T.; Yong, Z.; Gao, Y.; Peng, W.; Wang, L.; Zhang, S.; Du, J.; Lin, D.; Xu, M.; Kreitzer, A. C.; Cui, G. A Genetically Encoded Fluorescent Sensor Enables Rapid and Specific Detection of Dopamine in Flies, Fish, and Mice. *Cell* **2018**, *174* (2), 481-496.e19. https://doi.org/10.1016/j.cell.2018.06.042.

54. Feng, J.; Zhang, C.; Lischinsky, J. E.; Jing, M.; Zhou, J.; Wang, H.; Zhang, Y.; Dong, A.; Wu, Z.; Wu, H.; Chen, W.; Zhang, P.; Zou, J.; Hires, S. A.; Zhu, J. J.; Cui, G.; Lin, D.; Du, J.; Li, Y. A Genetically

Encoded Fluorescent Sensor for Rapid and Specific *in vivo* Detection of Norepinephrine. *Neuron* **2019**, *102* (4), 745-761.e8. https://doi.org/10.1016/j.neuron.2019.02.037.

55. Wu, Z.; He, K.; Chen, Y.; Li, H.; Pan, S.; Li, B.; Liu, T.; Xi, F.; Deng, F.; Wang, H.; Du, J.; Jing, M.; Li, Y. A Sensitive GRAB Sensor for Detecting Extracellular ATP *in vitro* and *in vivo*. *Neuron* **2022**, *110* (5), 770-782.e5. https://doi.org/10.1016/j.neuron.2021.11.027.
56. Wan, J.; Peng, W.; Li, X.; Qian, T.; Song, K.; Zeng, J.; Deng, F.; Hao, S.; Feng, J.; Zhang, P.; Zhang, Y.; Zou, J.; Pan, S.; Shin, M.; Venton, B. J.; Zhu, J. J.; Jing, M.; Xu, M.; Li, Y. A Genetically Encoded Sensor for Measuring Serotonin Dynamics. *Nature Neuroscience* **2021**, *24* (5), 746–752. https://doi.org/10.1038/s41593-021-00823-7.
57. Deng, F.; Wan, J.; Li, G.; Dong, H.; Xia, X.; Wang, Y.; Li, X.; Zhuang, C.; Zheng, Y.; Liu, L.; Yan, Y.; Feng, J.; Zhao, Y.; Xie, H.; Li, Y. Improved Green and Red GRAB Sensors for Monitoring Spatiotemporal Serotonin Release *in vivo*. *Nature methods* **2024**, *21* (4), 692–702. https://doi.org/10.1038/s41592-024-02188-8.
58. Jing, M.; Li, Y.; Zeng, J.; Huang, P.; Skirzewski, M.; Kljakic, O.; Peng, W.; Qian, T.; Tan, K.; Zou, J.; Trinh, S.; Wu, R.; Zhang, S.; Pan, S.; Hires, S. A.; Xu, M.; Li, H.; Saksida, L. M.; Prado, V. F.; Bussey, T. J. An Optimized Acetylcholine Sensor for Monitoring *in vivo* Cholinergic Activity. *Nature Methods* **2020**, *17* (11), 1139–1146. https://doi.org/10.1038/s41592-020-0953-2.
59. Qian, T.; Wang, H.; Wang, P.; Geng, L.; Mei, L.; Takuya Osakada; Wang, L.; Tang, Y.; Kania, A.; Grinevich, V.; Stoop, R.; Lin, D.; Luo, M.; Li, Y. A Genetically Encoded Sensor Measures Temporal Oxytocin Release from Different Neuronal Compartments. *Nature Biotechnology* **2023**, *41* (7), 944–957. https://doi.org/10.1038/s41587-022-01561-2.
60. Ino, D.; Hibino, H. A Fluorescent Sensor for the Real-Time Measurement of Extracellular Oxytocin Dynamics in the Brain. *Proceedings for Annual Meeting of The Japanese Pharmacological Society* **2022**, *95* (0), 1-O-005. https://doi.org/10.1254/jpssuppl.95.0_1-o-005.
61. Loïc Duffet; Kosar, S.; Mariangela Panniello; Viberti, B.; Bracey, E.; Zych, A. D.; Radoux-Mergault, A.; Zhou, X.; Dernic, J.; Luca Ravotto; Tsai, Y.-C.; Figueiredo, M.; Tyagarajan, S. K.; Weber, B.; Stoeber, M.; Gogolla, N.; Schmidt, M. H.; Adamantidis, A. R.; Fellin, T.; Burdakov, D. A Genetically Encoded

Sensor for *in vivo* Imaging of Orexin Neuropeptides. *Nature Methods* **2022**, *19* (2), 231–241. https://doi.org/10.1038/s41592-021-01390-2.

62. Wang, H.; Qian, T.; Zhao, Y.; Zhuo, Y.; Wu, C.; Takuya Osakada; Chen, P.; Chen, Z.; Ren, H.; Yan, Y.; Geng, L.; Fu, S.; Mei, L.; Li, G.; Wu, L.; Jiang, Y.; Qian, W.; Zhang, L.; Peng, W.; Xu, M. A Tool Kit of Highly Selective and Sensitive Genetically Encoded Neuropeptide Sensors. *Science* **2023**, *382* (6672). https://doi.org/10.1126/science.abq8173.
63. Sunahara, R. K.; Guan, H.-C.; O'Dowd, B. F.; Seeman, P.; Laurier, L. G.; Ng, G.; George, S. R.; Torchia, J.; Van Tol, H. H. M.; Niznik, H. B. Cloning of the Gene for a Human Dopamine D5 Receptor with Higher Affinity for Dopamine than D1. *Nature* **1991**, *350* (6319), 614–619. https://doi.org/10.1038/350614a0.
64. Park, S.-Y. .; Fung, P.; Nishimura, N.; Jensen, D. R.; Fujii, H.; Zhao, Y.; Lumba, S.; Santiago, J.; Rodrigues, A.; Chow, T. . F.; Alfred, S. E.; Bonetta, D.; Finkelstein, R.; Provart, N. J.; Desveaux, D.; Rodriguez, P. L.; McCourt, P.; Zhu, J.-K. .; Schroeder, J. I.; Volkman, B. F. Abscisic Acid Inhibits Type 2C Protein Phosphatases via the PYR/PYL Family of START Proteins. *Science* **2009**. https://doi.org/10.1126/science.1173041.
65. Wei, S.; Li, M.; Lang, X.; Robertson, N. R.; Park, S.-Y.; Cutler, S. R.; Wheeldon, I. Repurposing Plant Hormone Receptors as Chemically-Inducible Genetic Switches for Dynamic Regulation in Yeast. *Metabolic Engineering* **2024**, *83*, 102–109. https://doi.org/10.1016/j.ymben.2024.03.006.
66. Qin, Y.; Sammond, D. W.; Braselmann, E.; Carpenter, M. C.; Palmer, A. E. Development of an Optical Zn^{2+} Probe Based on a Single Fluorescent Protein. *ACS chemical biology* **2016**, *11* (10), 2744–2751. https://doi.org/10.1021/acschembio.6b00442.
67. Cunningham, B. A.; Hemperly, J. J.; Hopp, T. P.; Edelman, G. M. Favin versus Concanavalin A: Circularly Permuted Amino Acid Sequences. *Proceedings of the National Academy of Sciences* **1979**, *76* (7), 3218–3222. https://doi.org/10.1073/pnas.76.7.3218.
68. Hennecke, J.; Sebbel, P.; Rudi Glockshuber. Random Circular Permutation of DsbA Reveals Segments That Are Essential for Protein Folding and Stability 1 1Edited by R. Huber. **1999**, *286* (4), 1197–1215. https://doi.org/10.1006/jmbi.1998.2531.

69. Masahiro Iwakura; Nakamura, T.; Yamane, C.; Maki, K. Systematic Circular Permutation of an Entire Protein Reveals Essential Folding Elements. *Nature Structural & Molecular Biology* **2000**, *7* (7), 580–585. https://doi.org/10.1038/76811.

70. Yu, Y.; Lutz, S. Circular Permutation: A Different Way to Engineer Enzyme Structure and Function. *Trends in Biotechnology* **2011**, *29* (1), 18–25. https://doi.org/10.1016/j.tibtech.2010.10.004.

71. Dana, H.; Mohar, B.; Sun, Y.; Narayan, S.; Gordus, A.; Hasseman, J. P.; Tsegaye, G.; Holt, G. T.; Hu, A.; Walpita, D.; Patel, R.; Macklin, J. J.; Bargmann, C. I.; Ahrens, M. B.; Schreiter, E. R.; Jayaraman, V.; Looger, L. L.; Svoboda, K.; Kim, D. S. Sensitive Red Protein Calcium Indicators for Imaging Neural Activity. *eLife* **2016**, *5*. https://doi.org/10.7554/elife.12727.

72. Qian, Y.; Piatkevich, K. D.; Benedict Mc Larney; Abdelfattah, A.; Mehta, S.; Murdock, M. H.; Gottschalk, S.; Molina, R. S.; Zhang, W.; Chen, Y.; Wu, J.; Mikhail Drobizhev; Thomas; Zhang, J. Z.; Schreiter, E. R.; Shoham, S.; Razansky, D.; Boyden, E. S.; Campbell, R. E. A Genetically Encoded Near-Infrared Fluorescent Calcium Ion Indicator. **2019**, *16* (2), 171–174. https://doi.org/10.1038/s41592-018-0294-6.

73. Deo, C.; Abdelfattah, A. S.; Bhargava, H. K.; Berro, A. J.; Falco, N.; Farrants, H.; Moeyaert, B.; Chupanova, M.; Lavis, L. D.; Schreiter, E. R. The HaloTag as a General Scaffold for Far-Red Tunable Chemigenetic Indicators. *Nature Chemical Biology* **2021**, *17* (6), 718–723. https://doi.org/10.1038/s41589-021-00775-w.

74. Los, G. V.; Encell, L. P.; McDougall, M. G.; Hartzell, D. D.; Karassina, N.; Zimprich, C.; Wood, M. G.; Learish, R.; Ohana, R. F.; Urh, M.; Simpson, D.; Mendez, J.; Zimmerman, K.; Otto, P.; Vidugiris, G.; Zhu, J.; Darzins, A.; Klaubert, D. H.; Bulleit, R. F.; Wood, K. V. HaloTag: A Novel Protein Labeling Technology for Cell Imaging and Protein Analysis. *ACS Chemical Biology* **2008**, *3* (6), 373–382. https://doi.org/10.1021/cb800025k.

75. Grimm, J. B.; Muthusamy, A. K.; Liang, Y.; Brown, T. A.; Lemon, W. C.; Patel, R.; Lu, R.; Macklin, J. J.; Keller, P. J.; Ji, N.; Lavis, L. D. A General Method to Fine-Tune Fluorophores for Live-Cell and *in vivo* Imaging. *Nature Methods* **2017**, *14* (10), 987–994. https://doi.org/10.1038/nmeth.4403.

76. Frei, M. S.; Sanchez, S. A.; Liu, L.; Schneider, F.; Wang, Z.; Hiroyuki Hakozaki; Li, Y.; Lyons, A. C.; Rohm, T. V.; Olefsky, J. M.; Shi, L.; Johannes Schöneberg; Fraser, S. E.; Mehta, S.; Wang, Y.; Zhang,

J. Far-Red Chemigenetic Biosensors for Multi-Dimensional and Super-Resolved Kinase Activity Imaging. *bioRxiv (Cold Spring Harbor Laboratory)* **2024**. https://doi.org/10.1101/2024.02.10.579766.

77. Justin Daho Lee; Nguyen, A.; Zheyu Ruby Jin; Moghadasi, A.; Gibbs, C. E.; Wait, S. J.; Evitts, K. M.; Asencio, A.; Bremner, S. B.; Zuniga, S.; Chavan, V.; Williams, A.; Smith, N.; Regnier, M.; Young, J. E.; Mack, D.; Nance, E.; Boyle, P. M.; Berndt, A. Far-Red and Sensitive Sensor for Monitoring Real Time H_2O_2Dynamics with Subcellular Resolution and in Multi-Parametric Imaging Applications. *bioRxiv (Cold Spring Harbor Laboratory)* **2024**. https://doi.org/10.1101/2024.02.06.579232.
78. Chen, X.; Zaro, J. L.; Shen, W.-C. Fusion Protein Linkers: Property, Design and Functionality. *Advanced Drug Delivery Reviews* **2013**, *65* (10), 1357–1369. https://doi.org/10.1016/j.addr.2012.09.039.
79. Gräwe, A.; Ranglack, J.; Weyrich, A.; Stein, V. IFLinkC: An Iterative Functional Linker Cloning Strategy for the Combinatorial Assembly and Recombination of Linker Peptides with Functional Domains. *Nucleic Acids Research* **2020**, *48* (4), e24–e24. https://doi.org/10.1093/nar/gkz1210.
80. Gräwe, A.; Maarten Merkx; Stein, V. IFLinkC-X: A Scalable Framework to Assemble Bespoke Genetically Encoded Co-Polymeric Linkers of Variable Lengths and Amino Acid Composition. *Bioconjugate chemistry* **2022**, *33* (7), 1415–1421. https://doi.org/10.1021/acs.bioconjchem.2c00250.
81. Nasu, Y.; Shen, Y.; Kramer, L.; Campbell, R. E. Structure- and Mechanism-Guided Design of Single Fluorescent Protein-Based Biosensors. *Nature Chemical Biology* **2021**, *17* (5), 509–518. https://doi.org/10.1038/s41589-020-00718-x.
82. Nasu, Y.; Murphy-Royal, C.; Wen, Y.; Haidey, J. N.; Molina, R. S.; Aggarwal, A.; Zhang, S.; Kamijo, Y.; Paquet, M.-E.; Podgorski, K.; Drobizhev, M.; Bains, J. S.; Lemieux, M. J.; Gordon, G. R.; Campbell, R. E. A Genetically Encoded Fluorescent Biosensor for Extracellular L-Lactate. *Nature Communications* **2021**, *12* (1), 7058. https://doi.org/10.1038/s41467-021-27332-2.
83. Nadler, D. C.; Morgan, S.-A.; Flamholz, A.; Kortright, K. E.; Savage, D. F. Rapid Construction of Metabolite Biosensors Using Domain-Insertion Profiling. *Nature Communications* **2016**, *7* (1), 12266. https://doi.org/10.1038/ncomms12266.

84. Akerboom, J.; Rivera, J. D. V.; Guilbe, M. M. R.; Malavé, E. C. A.; Hernandez, H. H.; Tian, L.; Hires, S. A.; Marvin, J. S.; Looger, L. L.; Schreiter, E. R. Crystal Structures of the GCaMP Calcium Sensor Reveal the Mechanism of Fluorescence Signal Change and Aid Rational Design. *Journal of Biological Chemistry* **2008**, *284* (10), 6455–6464. https://doi.org/10.1074/jbc.m807657200.

85. Akerboom, J.; Chen, T.-W.; Wardill, T. J.; Tian, L.; Marvin, J. S.; Mutlu, S.; Calderón, N. C.; Esposti, F.; Borghuis, B. G.; Sun, X. R.; Gordus, A.; Orger, M. B.; Portugues, R.; Engert, F.; Macklin, J. J.; Filosa, A.; Aggarwal, A.; Kerr, R. A.; Takagi, R.; Kracun, S. Optimization of a GCaMP Calcium Indicator for Neural Activity Imaging. *The Journal of Neuroscience* **2012**, *32* (40), 13819–13840. https://doi.org/10.1523/JNEUROSCI.2601-12.2012.

86. Rappleye, M.; Wait, S. J.; Justin Daho Lee; Siebart, J. C.; Torp, L.; Smith, N.; Muster, J.; Matreyek, K. A.; Fowler, D. M.; Berndt, A. Optogenetic Microwell Array Screening System: A High-Throughput Engineering Platform for Genetically Encoded Fluorescent Indicators. *ACS Sensors* **2023**, *8* (11), 4233–4244. https://doi.org/10.1021/acssensors.3c01573.

87. Kinney, J. B.; Anand Murugan; Callan, C. G.; Cox, E. C. Using Deep Sequencing to Characterize the Biophysical Mechanism of a Transcriptional Regulatory Sequence. **2010**, *107* (20), 9158–9163. https://doi.org/10.1073/pnas.1004290107.

88. Kosuri, S.; Goodman, D. B.; Cambray, G.; Mutalik, V. K.; Gao, Y.; Arkin, A. P.; Endy, D.; Church, G. M. Composability of Regulatory Sequences Controlling Transcription and Translation InEscherichia Coli. *Proceedings of the National Academy of Sciences* **2013**, *110* (34), 14024–14029. https://doi.org/10.1073/pnas.1301301110.

89. Sharon, E.; Kalma, Y.; Sharp, A.; Raveh-Sadka, T.; Levo, M.; Zeevi, D.; Keren, L.; Yakhini, Z.; Weinberger, A.; Segal, E. Inferring Gene Regulatory Logic from High-Throughput Measurements of Thousands of Systematically Designed Promoters. *Nature Biotechnology* **2012**, *30* (6), 521–530. https://doi.org/10.1038/nbt.2205.

90. Sharon, E.; David van Dijk; Kalma, Y.; Keren, L.; Manor, O.; Zohar Yakhini; Segal, E. Probing the Effect of Promoters on Noise in Gene Expression Using Thousands of Designed Sequences. *Genome Research* **2014**, *24* (10), 1698–1706. https://doi.org/10.1101/gr.168773.113.

91. Noderer, W. L.; Flockhart, R. J.; Bhaduri, A.; Diaz de Arce, A. J.; Zhang, J.; Khavari, P. A.; Wang, C. L. Quantitative Analysis of Mammalian Translation Initiation Sites by FACS-Seq. *Molecular Systems Biology* **2014**, *10*, 748. https://doi.org/10.15252/msb.20145136.

92. Peterman, N.; Levine, E. Sort-Seq under the Hood: Implications of Design Choices on Large-Scale Characterization of Sequence-Function Relations. *BMC Genomics* **2016**, *17* (1). https://doi.org/10.1186/s12864-016-2533-5.

93. Koberstein, J. N.; Stewart, M. L.; Mighell, T. L.; Smith, C. B.; Cohen, M. S. A Sort-Seq Approach to the Development of Single Fluorescent Protein Biosensors. *ACS Chemical Biology* **2021**, *16* (9), 1709–1720. https://doi.org/10.1021/acschembio.1c00423.

94. Saaya Hario; Le; Sugimoto, H.; Kei Takahashi-Yamashiro; Suguru Nishinami; Toda, H.; Li, S.; Marvin, J. S.; Kuroda, S.; Mikhail Drobizhev; Terai, T.; Yusuke Nasu; Campbell, R. E. High-Performance Genetically Encoded Green Fluorescent Biosensors for Intracellular L-Lactate. *ACS Central Science* **2024**. https://doi.org/10.1021/acscentsci.3c01250.

95. Stanton, B. Z.; Chory, E. J.; Crabtree, G. R. Chemically Induced Proximity in Biology and Medicine. *Science (New York, N.Y.)* **2018**, *359* (6380), eaao5902. https://doi.org/10.1126/science.aao5902.

96. Marianayagam, N. J.; Sunde, M.; Matthews, J. M. The Power of Two: Protein Dimerization in Biology. *Trends in Biochemical Sciences* **2004**, *29* (11), 618–625. https://doi.org/10.1016/j.tibs.2004.09.006.

97. Spencer, D. M.; Wandless, T. J.; Schreiber, S. L.; Crabtree, G. R. Controlling Signal Transduction with Synthetic Ligands. *Science* **1993**, *262* (5136), 1019–1024. https://doi.org/10.1126/science.7694365.

98. Kopytek, S.; Standaert, R. F.; John Cd Dyer; Hu, J. C. Chemically Induced Dimerization of Dihydrofolate Reductase by a Homobifunctional Dimer of Methotrexate. *Chemistry & Biology* **2000**, *7* (5), 313–321. https://doi.org/10.1016/s1074-5521(00)00109-5.

99. Rivera, V. M.; Clackson, T.; Natesan, S.; Pollock, R.; Amara, J. F.; Keenan, T.; Magari, S. R.; Phillips, T.; Courage, N. L.; Cerasoli, F.; Holt, D. A.; Gilman, M. A Humanized System for Pharmacologic Control of Gene Expression. *Nature Medicine* **1996**, *2* (9), 1028–1032. https://doi.org/10.1038/nm0996-1028.

100. Liang, F.-S.; Ho, W. Q.; Crabtree, G. R. Engineering the ABA Plant Stress Pathway for Regulation of Induced Proximity. Science Signaling 2011, 4 (164). https://doi.org/10.1126/scisignal.2001449.

101. Miyamoto, T.; DeRose, R.; Suarez, A.; Ueno, T.; Chen, M.; Sun, T.-P.; Wolfgang, M. J.; Mukherjee, C.; Meyers, D. J.; Inoue, T. Rapid and Orthogonal Logic Gating with a Gibberellin-Induced Dimerization System. *Nature Chemical Biology* **2012**, *8* (5), 465–470. https://doi.org/10.1038/nchembio.922.

102. Hill, Z. B.; Martinko, A. J.; Nguyen, D.; Wells, J. A. Human Antibody-Based Chemically Induced Dimerizers for Cell Therapeutic Applications. *Nature Chemical Biology* **2017**, *14* (2), 112–117. https://doi.org/10.1038/nchembio.2529.

103. Leonard, A. C.; Friedman, A. J.; Chayer, R.; Petersen, B. M.; Janty Woojuh; Xing, Z.; Cutler, S. R.; Kaar, J. L.; Shirts, M. R.; Whitehead, T. A. Rationalizing Diverse Binding Mechanisms to the Same Protein Fold: Insights for Ligand Recognition and Biosensor Design. *ACS Chemical Biology* **2024**. https://doi.org/10.1021/acschembio.4c00243.

104. Ståhl, S.; Gräslund, T.; Eriksson Karlström, A.; Frejd, F. Y.; Nygren, P.-Å.; Löfblom, J. Affibody Molecules in Biotechnological and Medical Applications. *Trends in Biotechnology* **2017**, *35* (8), 691–712. https://doi.org/10.1016/j.tibtech.2017.04.007.

105. Löfblom, J.; Feldwisch, J.; Tolmachev, V.; Carlsson, J.; Ståhl, S.; Frejd, F. Y. Affibody Molecules: Engineered Proteins for Therapeutic, Diagnostic and Biotechnological Applications. *FEBS Letters* **2010**, *584* (12), 2670–2680.

106. Orlova, A.; Tolmachev, V.; Pehrson, R.; Lindborg, M.; Tran, T.; Sandström, M.; Nilsson, F. Y.; Wennborg, A.; Abrahmsén, L.; Feldwisch, J. Synthetic Affibody Molecules: A Novel Class of Affinity Ligands for Molecular Imaging of HER2-Expressing Malignant Tumors. *Cancer Research* **2007**, *67* (5), 2178–2186. https://doi.org/10.1158/0008-5472.can-06-2887.

107. Engfeldt, T.; Renberg, B.; Brumer, H.; Nygren, P. Å.; Eriksson Karlström, A. Chemical Synthesis of Triple-Labelled Three-Helix Bundle Binding Proteins for Specific Fluorescent Detection of Unlabelled Protein. *ChemBioChem* **2005**, *6* (6), 1043–1050. https://doi.org/10.1002/cbic.200400388.

108. Stumpp, M. T.; Binz, H. K.; Amstutz, P. DARPins: A New Generation of Protein Therapeutics. *Drug Discovery Today* **2008**, *13* (15-16), 695–701. https://doi.org/10.1016/j.drudis.2008.04.013.

109. Tommi Kajander; Cortajarena, A. L.; Regan, L. Consensus Design as a Tool for Engineering Repeat Proteins. *Humana Press eBooks* **2006**, 151–170. https://doi.org/10.1385/1-59745-116-9:151.

110. Binz, H. Kaspar.; Stumpp, M. T.; Forrer, P.; Amstutz, P.; Plückthun, A. Designing Repeat Proteins: Well-Expressed, Soluble and Stable Proteins from Combinatorial Libraries of Consensus Ankyrin Repeat Proteins. *Journal of Molecular Biology* **2003**, *332* (2), 489–503. https://doi.org/10.1016/s0022-2836(03)00896-9.

111. Morselli, M.; Holton, T. R.; Pellegrini, M.; Yeates, T. O.; Arbing, M. A. Design and Construction of a Designed Ankyrin Repeat Protein (DARPin) Display Library. *Current Protocols* **2024**, *4* (1). https://doi.org/10.1002/cpz1.960.

112. Plückthun, A. Designed Ankyrin Repeat Proteins (DARPins): Binding Proteins for Research, Diagnostics, and Therapy. Annual Review of Pharmacology and Toxicology 2015, 55 (1), 489–511. https://doi.org/10.1146/annurev-pharmtox-010611-134654.

113. Kummer, L.; Chia Wen Hsu; Onur Dagliyan; MacNevin, C.; Kaufholz, M.; Zimmermann, B.; Dokholyan, N. V.; Hahn, K. M.; Plückthun, A. Knowledge-Based Design of a Biosensor to Quantify Localized ERK Activation in Living Cells. Chemistry & Biology 2013, 20 (6), 847–856. https://doi.org/10.1016/j.chembiol.2013.04.016.

114. Hamers-Casterman, C.; Atarhouch, T.; Muyldermans, S.; Robinson, G.; Hammers, C.; Songa, E. B.; Bendahman, N.; Hammers, R. Naturally Occurring Antibodies Devoid of Light Chains. *Nature* **1993**, *363* (6428), 446–448. https://doi.org/10.1038/363446a0.

115. Lauwereys, M. Potent Enzyme Inhibitors Derived from Dromedary Heavy-Chain Antibodies. *The EMBO Journal* **1998**, *17* (13), 3512–3520. https://doi.org/10.1093/emboj/17.13.3512.

116. Bencurova, E.; Pulzova, L.; Flachbartova, Z.; Bhide, M. A Rapid and Simple Pipeline for Synthesis of MRNA–Ribosome–VHH Complexes Used in Single-Domain Antibody Ribosome Display. *Molecular BioSystems* **2015**, *11* (6), 1515–1524. https://doi.org/10.1039/c5mb00026b.

117. Kruse, A. C.; Ring, A. M.; Manglik, A.; Hu, J.; Hu, K.; Eitel, K.; Hübner, H.; Pardon, E.; Valant, C.; Sexton, P. M.; Christopoulos, A.; Felder, C. C.; Gmeiner, P.; Steyaert, J.; Weis, W. I.; Garcia, K. C.; Wess, J.; Kobilka, B. K. Activation and Allosteric Modulation of a Muscarinic Acetylcholine Receptor. *Nature* **2013**, *504* (7478), 101–106. https://doi.org/10.1038/nature12735.

118. Wendel, S.; Fischer, E. C.; Martínez, V.; Seppälä, S.; Morten. A Nanobody:GFP Bacterial Platform That Enables Functional Enzyme Display and Easy Quantification of Display Capacity. *Microbial Cell Factories* **2016**, *15* (1). https://doi.org/10.1186/s12934-016-0474-y.

119. Inoue, A.; Ohmuro-Matsuyama, Y.; Tetsuya Kitaguchi; Ueda, H. Creation of a Nanobody-Based Fluorescent Immunosensor Mini Q-Body for Rapid Signal-on Detection of Small Hapten Methotrexate. *ACS Sensors* **2020**, *5* (11), 3457–3464. https://doi.org/10.1021/acssensors.0c01404.

120. Su, B.; Zhang, Z.; Sun, Z.; Tang, Z.; Xie, X.; Chen, Q.; Cao, H.; Yu, X.; Xu, Y.; Liu, X.; Hammock, B. D. Fluonanobody-Based Nanosensor via Fluorescence Resonance Energy Transfer for Ultrasensitive Detection of Ochratoxin A. *Journal of Hazardous Materials* **2022**, *422*, 126838. https://doi.org/10.1016/j.jhazmat.2021.126838.

121. Rasmussen, S. G. F.; Choi, H.-J.; Fung, J. J.; Pardon, E.; Casarosa, P.; Chae, P. S.; DeVree, B. T.; Rosenbaum, D. M.; Thian, F. S.; Kobilka, T. S.; Schnapp, A.; Konetzki, I.; Sunahara, R. K.; Gellman, S. H.; Pautsch, A.; Steyaert, J.; Weis, W. I.; Kobilka, B. K. Structure of a Nanobody-Stabilized Active State of the β2 Adrenoceptor. *Nature* **2011**, *469* (7329), 175–180. https://doi.org/10.1038/nature09648.

122. Huang, W.; Manglik, A.; Venkatakrishnan, A. J.; Laeremans, T.; Feinberg, E. N.; Sanborn, A. L.; Kato, H. E.; Livingston, K. E.; Thorsen, T. S.; Kling, R. C.; Granier, S.; Gmeiner, P.; Husbands, S. M.; Traynor, J. R.; Weis, W. I.; Steyaert, J.; Dror, R. O.; Kobilka, B. K. Structural Insights into μ-Opioid Receptor Activation. *Nature* **2015**, *524* (7565), 315–321. https://doi.org/10.1038/nature14886.

123. Irannejad, R.; Tomshine, J. C.; Tomshine, J. R.; Chevalier, M.; Mahoney, J. P.; Steyaert, J.; Rasmussen, S. G. F.; Sunahara, R. K.; El-Samad, H.; Huang, B.; von Zastrow, M. Conformational

Biosensors Reveal GPCR Signalling from Endosomes. *Nature* **2013**, *495* (7442), 534–538. https://doi.org/10.1038/nature12000.

124. Kroning, K. E.; Wang, W. Designing a Single Protein-Chain Reporter for Opioid Detection at Cellular Resolution. *Angewandte Chemie* **2021**, *133* (24), 13470–13477. https://doi.org/10.1002/ange.202101262.

125. Devina Wongso; Dong, J.; Ueda, H.; Tetsuya Kitaguchi. Flashbody: A next Generation Fluobody with Fluorescence Intensity Enhanced by Antigen Binding. *Analytical Chemistry* **2017**, *89* (12), 6719–6725. https://doi.org/10.1021/acs.analchem.7b00959.

126. Sekhon, H.; Ha, J.-H.; Presti, M. F.; Procopio, S. B.; Jarvis, A. R.; Mirsky, P. O.; John, A. M.; Loh, S. N. Adaptable, Turn-on Maturation (ATOM) Fluorescent Biosensors for Multiplexed Detection in Cells. *Nature Methods* **2023**, *20* (12), 1920–1929. https://doi.org/10.1038/s41592-023-02065-w.

127. Koide, A.; Bailey, C. W.; Huang, X.; Koide, S. The Fibronectin Type III Domain as a Scaffold for Novel Binding Proteins. *Journal of Molecular Biology* **1998**, *284* (4), 1141–1151. https://doi.org/10.1006/jmbi.1998.2238.

128. Koide, A.; Gilbreth, R. N.; Esaki, K.; Tereshko, V.; Koide, S. High-Affinity Single-Domain Binding Proteins with a Binary-Code Interface. *Proceedings of the National Academy of Sciences* **2007**, *104* (16), 6632–6637. https://doi.org/10.1073/pnas.0700149104.

129. Koide, A.; Wojcik, J.; Gilbreth, R.; Hoey, R. J.; Koide, S. Teaching an Old Scaffold New Tricks: Monobodies Constructed Using Alternative Surfaces of the FN3 Scaffold. *Journal of Molecular Biology* **2012**, *415* (2), 393–405. https://doi.org/10.1016/j.jmb.2011.12.019.

130. Limsakul, P.; Peng, Q.; Wu, Y.; Allen, M. E.; Liang, J.; Remacle, A. G.; Lopez, T.; Ge, X.; Kay, B. K.; Zhao, H.; Strongin, A. Y.; Yang, X.-L.; Lu, S.; Wang, Y. Directed Evolution to Engineer Monobody for FRET Biosensor Assembly and Imaging at Live-Cell Surface. *Cell Chemical Biology* **2018**, *25* (4), 370-379.e4. https://doi.org/10.1016/j.chembiol.2018.01.002.

131. Gulyani, A.; Vitriol, E.; Allen, R.; Wu, J.; Gremyachinskiy, D.; Lewis, S.; Dewar, B.; Graves, L. M.; Kay, B. K.; Kuhlman, B.; Elston, T.; Hahn, K. M. A Biosensor Generated via High-Throughput Screening Quantifies Cell Edge Src Dynamics. *Nature Chemical Biology* **2011**, *7* (7), 437–444. https://doi.org/10.1038/nchembio.585.

132. H. Mario Geysen; Meloen, R. H.; Barteling, S. J. Use of Peptide Synthesis to Probe Viral Antigens for Epitopes to a Resolution of a Single Amino Acid. *Proceedings of the National Academy of Sciences of the United States of America* **1984**, *81* (13), 3998–4002. https://doi.org/10.1073/pnas.81.13.3998.

133. Huse, W. D.; Sastry, L.; Iverson, S. A.; Kang, A. S.; Alting-Mees, M.; Burton, D. R.; Benkovic, S. J.; Lerner, R. A. Generation of a Large Combinatorial Library of the Immunoglobulin Repertoire in Phage Lambda. *Science* **1989**, *246* (4935), 1275–1281. https://doi.org/10.1126/science.2531466.

134. Sastry, L.; M Alting-Mees; Huse, W. D.; Short, J. M.; Sorge, J. A.; Hay, B. N.; Janda, K. D.; Benkovic, S. J.; Lerner, R. A. Cloning of the Immunological Repertoire in Escherichia Coli for Generation of Monoclonal Catalytic Antibodies: Construction of a Heavy Chain Variable Region-Specific CDNA Library. *Proceedings of the National Academy of Sciences* **1989**, *86* (15), 5728–5732. https://doi.org/10.1073/pnas.86.15.5728.

135. Orlandi, R.; Detlef Güssow; Jones, P.; Winter, G. Cloning Immunoglobulin Variable Domains for Expression by the Polymerase Chain Reaction. *Proceedings of the National Academy of Sciences of the United States of America* **1989**, *86* (10), 3833–3837. https://doi.org/10.1073/pnas.86.10.3833.

136. Begent, R. H. J.; Verhaar, M. J.; Chester, K. A.; Casey, J. L.; Green, A. J.; Napier, M. P.; Hopestone, L. D.; Cushen, N.; Keep, P. A.; Johnson, C. J.; Hawkins, R. E.; Hilson, A. J. W.; Robson, L. Clinical Evidence of Efficient Tumor Targetting Based on Single–Chain Fv Antibody Selected from a Combinatorial Library. *Nature Medicine* **1996**, *2* (9), 979–984. https://doi.org/10.1038/nm0996-979.

137. Pardon, E.; Laeremans, T.; Triest, S.; Rasmussen, S. G. F.; Wohlkönig, A.; Ruf, A.; Muyldermans, S.; Hol, W. G. J.; Kobilka, B. K.; Steyaert, J. A General Protocol for the Generation of Nanobodies for Structural Biology. *Nature protocols* **2014**, *9* (3), 674–693. https://doi.org/10.1038/nprot.2014.039.

138. Yan, J.; Li, G.; Hu, Y.; Ou, W.; Wan, Y. Construction of a Synthetic Phage-Displayed Nanobody Library with CDR3 Regions Randomized by Trinucleotide Cassettes for Diagnostic

Applications. *Journal of Translational Medicine* **2014**, *12* (1). https://doi.org/10.1186/s12967-014-0343-6.

139. Moutel, S.; Bery, N.; Bernard, V.; Keller, L.; Lemesre, E.; de Marco, A.; Ligat, L.; Rain, J.-C.; Favre, G.; Olichon, A.; Perez, F. NaLi-H1: A Universal Synthetic Library of Humanized Nanobodies Providing Highly Functional Antibodies and Intrabodies. *eLife* **2016**, *5*. https://doi.org/10.7554/elife.16228.

140. Lerner, R. A. Manufacturing Immunity to Disease in a Test Tube: The Magic Bullet Realized. *ChemInform* **2007**, *38* (13). https://doi.org/10.1002/chin.200713273.

141. Kille, S.; Acevedo-Rocha, C. G.; Parra, L. P.; Zhang, Z.-G.; Opperman, D. J.; Reetz, M. T.; Acevedo, J. P. Reducing Codon Redundancy and Screening Effort of Combinatorial Protein Libraries Created by Saturation Mutagenesis. *ACS Synthetic Biology* **2012**, *2* (2), 83–92. https://doi.org/10.1021/sb300037w.

142. Wells, J. A.; Vasser, M.; Powers, D. B. Cassette Mutagenesis: An Efficient Method for Generation of Multiple Mutations at Defined Sites. *Gene* **1985**, *34* (2-3), 315–323. https://doi.org/10.1016/0378-1119(85)90140-4.

143. Derbyshire, K. M.; Salvo, J. J.; Grindley, N. D. F. A Simple and Efficient Procedure for Saturation Mutagenesis Using Mixed Oligodeoxynucleotides. *Gene* **1986**, *46* (2-3), 145–152. https://doi.org/10.1016/0378-1119(86)90398-7.

144. Oliphant, A.; Nussbaum, A. L.; Struhl, K. Cloning of Random-Sequence Oligodeoxynucleotides. *Gene* **1986**, *44* (2-3), 177–183. https://doi.org/10.1016/0378-1119(86)90180-0.

145. Virnekäs, B.; Ge, L.; Plückthun, A.; Schneider, K. C.; Wellnhofer, G.; Moroney, S. E. Trinucleotide Phosphoramidites: Ideal Reagents for the Synthesis of Mixed Oligonucleotides for Random Mutagenesis. *Nucleic Acids Research* **1994**, *22* (25), 5600–5607. https://doi.org/10.1093/nar/22.25.5600.

146. Smith, G. P.; Petrenko, V. A. Phage Display. *Chemical Reviews* **1997**, *97* (2), 391–410. https://doi.org/10.1021/cr960065d.

147. Lee, S. Y.; Choi, J. H.; Xu, Z. Microbial Cell-Surface Display. *Trends in Biotechnology* **2003**, *21* (1), 45–52. https://doi.org/10.1016/s0167-7799(02)00006-9.

148. Kondo, A.; Ueda, M. Yeast Cell-Surface Display—Applications of Molecular Display. *Applied Microbiology and Biotechnology* **2004**, *64* (1), 28–40. https://doi.org/10.1007/s00253-003-1492-3.

149. Lipovsek, D.; Plückthun, A. In-Vitro Protein Evolution by Ribosome Display and MRNA Display. *Journal of Immunological Methods* **2004**, *290* (1-2), 51–67. https://doi.org/10.1016/j.jim.2004.04.008.

150. Kehoe, J. W.; Kay, B. K. Filamentous Phage Display in the New Millennium. *Chemical Reviews* **2005**, *105* (11), 4056–4072. https://doi.org/10.1021/cr000261r.

151. Boder, E. T.; K. Dane Wittrup. Yeast Surface Display for Directed Evolution of Protein Expression, Affinity, and Stability. *Methods in enzymology* **2000**, *328*, 430–444. https://doi.org/10.1016/s0076-6879(00)28410-3.

152. Daugherty, P. S. Protein Engineering with Bacterial Display. *Current Opinion in Structural Biology* **2007**, *17* (4), 474–480. https://doi.org/10.1016/j.sbi.2007.07.004.

153. Harvey, B. R.; Georgiou, G.; Hayhurst, A.; Jeong, K. J.; Iverson, B. L.; Rogers, G. K. Anchored Periplasmic Expression, a Versatile Technology for the Isolation of High-Affinity Antibodies from Escherichia Coli-Expressed Libraries. *Proceedings of the National Academy of Sciences* **2004**, *101* (25), 9193–9198. https://doi.org/10.1073/pnas.0400187101.

154. He, M.; Taussig, M. J. Ribosome Display: Cell-Free Protein Display Technology. *Briefings in Functional Genomics & Proteomics* **2002**, *1* (2), 204–212. https://doi.org/10.1093/bfgp/1.2.204.

155. Pande, J.; Szewczyk, M. M.; Grover, A. K. Phage Display: Concept, Innovations, Applications and Future. *Biotechnology Advances* **2010**, *28* (6), 849–858. https://doi.org/10.1016/j.biotechadv.2010.07.004.

156. Gera, N.; Hussain, M.; Rao, B. M. Protein Selection Using Yeast Surface Display. *Methods* **2013**, *60* (1), 15–26. https://doi.org/10.1016/j.ymeth.2012.03.014.

157. VanAntwerp, J. J.; Wittrup, K. D. Fine Affinity Discrimination by Yeast Surface Display and Flow Cytometry. *Biotechnology Progress* **2000**, *16* (1), 31–37. https://doi.org/10.1021/bp990133s.

158. Kang, S.; Davidsen, K.; Gomez-Castillo, L.; Jiang, H.; Fu, X.; Li, Z.; Liang, Y.; Jahn, M.; Moussa, M.; DiMaio, F.; Gu, L. COMBINES-CID: An Efficient Method for de Novo Engineering of Highly Specific Chemically Induced Protein Dimerization Systems. *Journal of the American Chemical Society* **2019**, *141* (28), 10948–10952. https://doi.org/10.1021/jacs.9b03522.

159. Muyldermans, S. Nanobodies: Natural Single-Domain Antibodies. Annu. Rev. *Biochem.* **2013**, 82 (1), 775–797. https://doi.org/10.1146/annurev-biochem-063011-092449.

160. Weber, E.; Esch, F. S.; P Böhlen; Paterson, S.; Corbett, A. D.; McKnight, A. T.; Kosterlitz, H. W.; Barchas, J. D.; Evans, C. J. Metorphamide: Isolation, Structure, and Biologic Activity of an Amidated Opioid Octapeptide from Bovine Brain. *Proceedings of the National Academy of Sciences* **1983**, *80* (23), 7362–7366. https://doi.org/10.1073/pnas.80.23.7362.

161. Douglass, E. F.; Miller, C. J.; Sparer, G.; Shapiro, H.; Spiegel, D. A. A Comprehensive Mathematical Model for Three-Body Binding Equilibria. *Journal of the American Chemical Society* **2013**, *135* (16), 6092–6099. https://doi.org/10.1021/ja311795d.

162. Dingus, J. G.; Tang Jc; Ryoji Amamoto; Wallick, G. K.; Cepko, C. L. A General Approach for Stabilizing Nanobodies for Intracellular Expression. *eLife* **2022**, *11*. https://doi.org/10.7554/elife.68253.

163. Kubitschke, M.; Müller, M.; Wallhorn, L.; Pulin, M.; Mittag, M.; Pollok, S.; Ziebarth, T.; Bremshey, S.; Gerdey, J.; Claussen, K. C.; Renken, K.; Groß, J.; Gneiße, P.; Meyer, N.; Wiegert, J. S.; Reiner, A.; Fuhrmann, M.; Masseck, O. A. Next Generation Genetically Encoded Fluorescent Sensors for Serotonin. *Nature Communications* **2022**, *13* (1), 7525. https://doi.org/10.1038/s41467-022-35200-w.

164. Morita, M.; Nakamura, S.; Shimizu, K. Highly Accurate Method for Ligand-Binding Site Prediction in Unbound State (Apo) Protein Structures. *Proteins: Structure, Function, and Bioinformatics* **2008**, *73* (2), 468–479. https://doi.org/10.1002/prot.22067.

165. Bryant, P.; Pozzati, G.; Elofsson, A. Improved Prediction of Protein-Protein Interactions Using AlphaFold2. *Nature Communications* **2022**, *13* (1), 1265. https://doi.org/10.1038/s41467-022-28865-w.

166. Steiner, P. J.; Swift, S. D.; Bedewitz, M.; Wheeldon, I.; Cutler, S. R.; Nusinow, D. A.; Whitehead, T. A. A Closed Form Model for Molecular Ratchet-Type Chemically Induced Dimerization Modules. *Biochemistry* **2022**, *62* (2), 281–291. https://doi.org/10.1021/acs.biochem.2c00172.

167. Wang, Q.; Shui, B.; Kotlikoff, M. I.; Sondermann, H. Structural Basis for Calcium Sensing by GCaMP2. *Structure (London, England : 1993)* **2008**, *16* (12), 1817–1827. https://doi.org/10.1016/j.str.2008.10.008.

168. Nasu, Y.; Aggarwal, A.; Le, G. N. T.; Vo, C. T.; Kambe, Y.; Wang, X.; Beinlich, F. R. M.; Lee, A. B.; Ram, T. R.; Wang, F.; Gorzo, K. A.; Kamijo, Y.; Boisvert, M.; Nishinami, S.; Kawamura, G.; Ozawa, T.; Toda, H.; Gordon, G. R.; Ge, S.; Hirase, H. Lactate Biosensors for Spectrally and Spatially Multiplexed Fluorescence Imaging. *Nature Communications* **2023**, *14* (1), 6598. https://doi.org/10.1038/s41467-023-42230-5.

169. Grimm, J. B.; Tkachuk, A. N.; Xie, L.; Choi, H.; Mohar, B.; Falco, N.; Schaefer, K.; Patel, R.; Zheng, Q.; Liu, Z.; Lippincott-Schwartz, J.; Brown, T. A.; Lavis, L. D. A General Method to Optimize and Functionalize Red-Shifted Rhodamine Dyes. *Nature Methods* **2020**, *17* (8), 815–821. https://doi.org/10.1038/s41592-020-0909-6.

170. Roehr, J. T.; Dieterich, C.; Reinert, K. Flexbar 3.0 – SIMD and Multicore Parallelization. *Bioinformatics* **2017**, *33* (18), 2941–2942. https://doi.org/10.1093/bioinformatics/btx330.

171. Crooks, G. E. WebLogo: A Sequence Logo Generator. *Genome Research* **2004**, *14* (6), 1188–1190. https://doi.org/10.1101/gr.849004.

172. Cao, S.; Kang, S.; Mao, H.; Yao, J.; Gu, L.; Zheng, N. Defining Molecular Glues with a Dual-Nanobody Cannabidiol Sensor. *Nature Communications* **2022**, *13* (1). https://doi.org/10.1038/s41467-022-28507-1.

173. Newton, M. S.; Cabezas-Perusse, Y.; Tong, C. L.; Seelig, B. *In vitro* Selection of Peptides and Proteins—Advantages of MRNA Display. *ACS Synthetic Biology* **2019**, *9* (2), 181–190. https://doi.org/10.1021/acssynbio.9b00419.

174. Kurz, M.; Gu, K.; Al-Gawari, A.; Lohse, P. A. CDNA–Protein Fusions: Covalent Protein–Gene Conjugates for the *in vitro* Selection of Peptides and Proteins. *ChemBioChem* **2001**, *2* (9), 666–672. https://doi.org/10.1002/1439-7633(20010903)2:9%3C666::aid-cbic666%3E3.0.co;2-#.

Appendix

Construct	DNA sequence
CA14v1-pBAD	aagaaggagatatggatccCATATGCAAGTCCAATTACAAGAGTCTGGAGGGGGTCTGGTACAACCTGGCGGCAGCCTG CGCTTGTCTTGCGCCGCAAGTGGTAGTACGTCCCGTCAATACGACATGGGATGGTTCCGTCAGGCACCCGGTAAG GAGCGCGAGTTTGTTAGTGCGATCTCAAGTAATCAAGACCAGCCCCCCTATTATGCCGACAGCGTTAAAGGTCGCT TCACGATTTCGCGTGACAATGCCAAGAACACCGTATATTTACAAATGAATAGCTTAAAGCCAGAGGACACTGCGACC TATTATTGTGCATTCAAACAGCATCACGCGAATGGGGCATACTGGGGACAGGGGACACAAGTGACCGTTTCCTCGA AGCTTggtggcggtagcgagaatt
CA14v2-pBAD	aagaaggagatatggatccCATATGCAAGTCCAATTACAAGAGTCTGGAGGGGGTCTGGTACAACCTGGCGGCAGCCTG CGCTTGTCTTGCGCCGCAAGTGGTAGTACGTCCCGTCAATACGACATGGGATGGTTCCGTCAGGCACCCGGTAAG GAGCGCGAGTTTGTTGCGGCGATCTCAAGTAATCAAGACCAGCCCCCCTATTATGCCGACAGCGTTAAAGGTCGCT TCACGATTTCGCGTGACAATGCCAAGAACACCGTATATTTACAAATGAATAGCTTAAAGCCAGAGGACACTGCGACC TATTATTGTGCATTCAAACAGCATCACGCGAATGGGGCATACTGGGGACAGGGGACACAAGTGACCGTTTCCTCGA AGCTTggtggcggtagcgagaatt
DB21v1-pBAD	aagaaggagatatggatccCATATGCAGGTGCAGTTACAAGAGTCAGGTGGAGGGTTGGTACAGCCTGGTGGGTCCCTG CGCCTGAGTTGTGCCGCCAGCGGAACCACCTACGGGCAGACAAACATGGGATGGTTTCGTCAAGCTCCAGGGAA AGAGCGTGAGTTTGTCTCTGCAATTTCAGGCTTGCAAGGACGTGACTTATACTACGCCGATAGTGTTAAGGGTCGTTT CACTATCTCGCGCGATAACGCAAAGAATACTGTGTATTTGCAGATGAATAGCTTGAAGCCTGAGGACACTGCCACAT ATTATTGTGCATTTCACGATTTCCTGCGTATGTGGGAATACTGGGGCCAGGGCACGCAGGTCACCGTAAGCTCCAA GCTTggtggcggtagcgagaatt
DB21v2-pBAD	aagaaggagatatggatccCATATGCAGGTGCAGTTACAAGAGTCAGGTGGAGGGTTGGTACAGCCTGGTGGGTCCCTG CGCCTGAGTTGTGCCGCCAGCGGAACCACCTACGGGCAGACAAACATGGGATGGTTTCGTCAAGCTCCAGGGAA AGAGCGTGAGTTTGTCGCGGCAATTTCAGGCTTGCAAGGACGTGACTTATACTACGCCGATAGTGTTAAGGGTCGTT TCACTATCTCGCGCGATAACGCAAAGAATACTGTGTATTTGCAGATGAATAGCTTGAAGCCTGAGGACACTGCCACA TATTATTGTGCATTTCACGATTTCCTGCGTATGTGGGAATACTGGGGCCAGGGCACGCAGGTCACCGTAAGCTCCAA GCTTggtggcggtagcgagaatt

Table S1. DNA sequences of stabilized nanobody CID pair. Sequences in lowercase are overlaps with pBAD expression vector.

Name	DNA sequence
VgsVLR-F2	TCAGTAATTTTCCAATATGAGAACATCAATTGTACGAAGACggtggctataacgtctttatcatggccgacaagcagaagaacggcatcaaggcgaacttcaagatccgccacaacatcgaggacggcggcgtgcagctcgcctatcactaccagcagaacacccccatcggcgacggccccgtgctgctgcccgacaaccactacctgagcgtgcagtccaaactgagcaaagaccccaacgagaagcgcgatcacatggtcctgctggagttcgtgaccgccgccgggatcactctcg
VgsVLR-F3	TCAGTAATTTTCCAATATGAGAACATCAATTGTACGAAGACggtggcAGTtataacgtctttatcatggccgacaagcagaagaacggcatcaaggcgaacttcaagatccgccacaacatcgaggacggcggcgtgcagctcgcctatcactaccagcagaacacccccatcggcgacggccccgtgctgctgcccgacaaccactacctgagcgtgcagtccaaactgagcaaagaccccaacgagaagcgcgatcacatggtcctgctggagttcgtgaccgccgccgggatcactctcg
VgsVLR-F4	TCAGTAATTTTCCAATATGAGAACATCAATTGTACGAAGACggtggcGGAAGTtataacgtctttatcatggccgacaagcagaagaacggcatcaaggcgaacttcaagatccgccacaacatcgaggacggcggcgtgcagctcgcctatcactaccagcagaacacccccatcggcgacggccccgtgctgctgccccgacaaccactacctgagcgtgcagtccaaactgagcaaagaccccaacgagaagcgcgatcacatggtcctgctggagttcgtgaccgccgccgggatcactctcg
VgsVLR-F5	TCAGTAATTTTCCAATATGAGAACATCAATTGTACGAAGACggtggcGGCGGAAGTtataacgtctttatcatggccgacaagcagaagaacggcatcaaggcgaacttcaagatccgccacaacatcgaggacggcggcgtgcagctcgcctatcactaccagcagaacacccccatcggcgacggccccgtgctgctgcccgacaaccactacctgagcgtgcagtccaaactgagcaaagaccccaacgagaagcgcgatcacatggtcctgctggagttcgtgaccgccgccgggatcactctcg
VgsVLR-F6	TCAGTAATTTTCCAATATGAGAACATCAATTGTACGAAGACggtggcTCCGGCGGAAGTtataacgtctttatcatggccgacaagcagaagaacggcatcaaggcgaacttcaagatccgccacaacatcgaggacggcggcgtgcagctcgcctatcactaccagcagaacacccccatcggcgacggccccgtgctgctgcccgacaaccactacctgagcgtgcagtccaaactgagcaaagaccccaacgagaagcgcgatcacatggtcctgctggagttcgtgaccgccgccgggatcactctcg
VgsVLR-F7	TCAGTAATTTTCCAATATGAGAACATCAATTGTACGAAGACggtggcTCCGGAGGCGGAAGTtataacgtctttatcatggccgacaagcagaagaacggcatcaaggcgaacttcaagatccgccacaacatcgaggacggcggcgtgcagctcgcctatcactaccagcagaacacccccatcggcgacggccccgtgctgctgcccgacaaccactacctgagcgtgcagtccaaactgagcaaagaccccaacgagaagcgcgatcacatggtcctgctggagttcgtgaccgccgccgggatcactctcg
VgsVLR-F8	TCAGTAATTTTCCAATATGAGAACATCAATTGTACGAAGACggtggcTCCGGTGGAGGCGGAAGTtataacgtctttatcatggccgacaagcagaagaacggcatcaaggcgaacttcaagatccgccacaacatcgaggacggcggcgtgcagctcgcctatcactaccagcagaacacccccatcggcgacggccccgtgctgctgcccgacaaccactacctgagcgtgcagtccaaactgagcaaagaccccaacgagaagcgcgatcacatggtcctgctggagttcgtgaccgccgccgggatcactctcg
VgsVLR-F9	TCAGTAATTTTCCAATATGAGAACATCAATTGTACGAAGACggtggcGGCTCCGGTGGAGGCGGAAGTtataacgtctttatcatggccgacaagcagaagaacggcatcaaggcgaacttcaagatccgccacaacatcgaggacggcggcgtgcagctcgcctatcactaccagcagaacacccccatcggcgacggccccgtgctgctgcccgacaaccactacctgagcgtgcagtccaaactgagcaaagaccccaacgagaagcgcgatcacatggtcctgctggagttcgtgaccgccgccgggatcactctcg
VgsVLR-F10	TCAGTAATTTTCCAATATGAGAACATCAATTGTACGAAGACggtggcGGTGGCTCCGGTGGAGGCGGAAGTtataacgtctttatcatggccgacaagcagaagaacggcatcaaggcgaacttcaagatccgccacaacatcgaggacggcggcgtgcagctcgcctatcactaccagcagaacacccccatcggcgacggccccgtgctgctgcccgacaaccactacctgagcgtgcagtccaaactgagcaaagacccccaacgagaagcgcgatcacatggtcctgctggagttcgtgaccgccgccgggatcactctcg

Name	DNA sequence
VgsVLR-F11	TCAGTAATTTTCCAATATGAGAACATCAATTGTACGAAGACggtggcTCTGGTGGCTCCGGTGGAGGCGGAAGTtataacgtctttatcatggccgacaagcagaagaacggcatcaaggcgaacttcaagatccgccacaacatcgaggacggcggcgtgcagctcgcctatcactaccagcagaacacccccatcggcgacggccccgtgctgctgcccgacaaccactacctgagcgtgcagtccaaactgagcaaagaccccaacgagaagcgcgatcacatggtcctgctggagttcgtgaccgccgccgggatcactctcg
VgsVLR-F12	TCAGTAATTTTCCAATATGAGAACATCAATTGTACGAAGACggtggcTCTGGAGGTGGCTCCGGTGGAGGCGGAAGTtataacgtctttatcatggccgacaagcagaagaacggcatcaaggcgaacttcaagatccgccacaacatcgaggacggcggcgtgcagctcgcctatcactaccagcagaacacccccatcggcgacggccccgtgctgctgcccgacaaccactacctgagcgtgcagtccaaactgagcaaagaccccaacgagaagcgcgatcacatggtcctgctggagttcgtgaccgccgccgggatcactctcg
VgsVLR-F13	TCAGTAATTTTCCAATATGAGAACATCAATTGTACGAAGACggtggcTCTGGAGGAGGTGGCTCCGGTGGAGGCGGAAGTtataacgtctttatcatggccgacaagcagaagaacggcatcaaggcgaacttcaagatccgccacaacatcgaggacggcggcgtgcagctcgcctatcactaccagcagaacacccccatcggcgacggccccgtgctgctgcccgacaaccactacctgagcgtgcagtccaaactgagcaaagaccccaacgagaagcgcgatcacatggtcctgctggagttcgtgaccgccgccgggatcactctcg
VgsVLR-F14	TCAGTAATTTTCCAATATGAGAACATCAATTGTACGAAGACggtggcGGGTCTGGAGGAGGTGGCTCCGGTGGAGGCGGAAGTtataacgtctttatcatggccgacaagcagaagaacggcatcaaggcgaacttcaagatccgccacaacatcgaggacggcggcgtgcagctcgcctatcactaccagcagaacacccccatcggcgacggccccgtgctgctgcccgacaaccactacctgagcgtgcagtccaaactgagcaaagaccccaacgagaagcgcgatcacatggtcctgctggagttcgtgaccgccgccgggatcactctcg
VgsVLR-F15	TCAGTAATTTTCCAATATGAGAACATCAATTGTACGAAGACggtggcGGAGGGTCTGGAGGAGGTGGCTCCGGTGGAGGCGGAAGTtataacgtctttatcatggccgacaagcagaagaacggcatcaaggcgaacttcaagatccgccacaacatcgaggacggcggcgtgcagctcgcctatcactaccagcagaacacccccatcggcgacggccccgtgctgctgcccgacaaccactacctgagcgtgcagtccaaactgagcaaagaccccaacgagaagcgcgatcacatggtcctgctggagttcgtgaccgccgccgggatcactctcg
VgsVLR-F16	TCAGTAATTTTCCAATATGAGAACATCAATTGTACGAAGACggtggcAGCGGAGGGTCTGGAGGAGGTGGCTCCGGTGGAGGCGGAAGTtataacgtctttatcatggccgacaagcagaagaacggcatcaaggcgaacttcaagatccgccacaacatcgaggacggcggcgtgcagctcgcctatcactaccagcagaacacccccatcggcgacggccccgtgctgctgcccgacaaccactacctgagcgtgcagtccaaactgagcaaagaccccaacgagaagcgcgatcacatggtcctgctggagttcgtgaccgccgccgggatcactctcg
VgsVLR-F17	TCAGTAATTTTCCAATATGAGAACATCAATTGTACGAAGACggtggcAGCGGTGGAGGGTCTGGAGGAGGTGGCTCCGGTGGAGGCGGAAGTtataacgtctttatcatggccgacaagcagaagaacggcatcaaggcgaacttcaagatccgccacaacatcgaggacggcggcgtgcagctcgcctatcactaccagcagaacacccccatcggcgacggccccgtgctgctgcccgacaaccactacctgagcgtgcagtccaaactgagcaaagaccccaacgagaagcgcgatcacatggtcctgctggagttcgtgaccgccgccgggatcactctcg
VgsVLR-F18	TCAGTAATTTTCCAATATGAGAACATCAATTGTACGAAGACggtggcAGCGGCGGTGGAGGGTCTGGAGGAGGTGGCTCCGGTGGAGGCGGAAGTtataacgtctttatcatggccgacaagcagaagaacggcatcaaggcgaacttcaagatccgccacaacatcgaggacggcggcgtgcagctcgcctatcactaccagcagaacacccccatcggcgacggccccgtgctgctgcccgacaaccactacctgagcgtgcagtccaaactgagcaaagaccccaacgagaagcgcgatcacatggtcctgctggagttcgtgaccgccgccgggatcactctcg
VgsVLR-F19	TCAGTAATTTTCCAATATGAGAACATCAATTGTACGAAGACggtggcGGTAGCGGCGGTGGAGGGTCTGGAGGAGGTGGCTCCGGTGGAGGCGGAAGTtataacgtctttatcatggccgacaagcagaagaacggcatcaaggcgaacttcaagatccgccacaacatcgaggacggcggcgtgcagctcgcctatcactaccagcagaacacccccatcggcgacggccccgtgctgctgcccgacaaccactacctgagcgtgcagtccaaactgagcaaagaccccaacgagaagcgcgatcacatggtcctgctggagttcgtgaccgccgccgggatcactctcg
VgsVLR-F20	TCAGTAATTTTCCAATATGAGAACATCAATTGTACGAAGACggtggcGGAGGTAGCGGCGGTGGAGGGTCTGGAGGAGGTGGCTCCGGTGGAGGCGGAAGTtataacgtctttatcatggccgacaagcagaagaacggcatcaaggcgaacttcaagatccgccacaacatcgaggacggcggcgtgcagctcgcctatcactaccagcagaacacccccatcggcgacggccccgtgctgctgcccgacaaccactacctgagcgtgcagtccaaactgagcaaagaccccaacgagaagcgcgatcacatggtcctgctggagttcgtgaccgccgccgggatcactctcg

Name	DNA sequence
VgsVLR-R2	caccctgacctacggcgtgcagtgcttcagccgctaccccgaccacatgaagcagcacgacttcttcaa gtccgccatgcccgaaggctacattcaggagcgcaccatcttcttcaaggacgacggcaactataagac acgcgctgaggttaagttcgagggcgacactctggttaaccgcatcgagctgaagggcatcgacttcaagg aggacggcaacatcctgggccataagcttgaatataacttcaacggcggtttGTCTTCCAAACGAGG TGATTTCAACAAATTTTGTTGATGAT
VgsVLR-R3	caccctgacctacggcgtgcagtgcttcagccgctaccccgaccacatgaagcagcacgacttcttcaa gtccgccatgcccgaaggctacattcaggagcgcaccatcttcttcaaggacgacggcaactataagac acgcgctgaggttaagttcgagggcgacactctggttaaccgcatcgagctgaagggcatcgacttcaagg aggacggcaacatcctgggccataagcttgaatataacttcaacTCCggcggtttGTCTTCCAAACG AGGTGATTTCAACAAATTTTGTTGATGAT
VgsVLR-R4	caccctgacctacggcgtgcagtgcttcagccgctaccccgaccacatgaagcagcacgacttcttcaa gtccgccatgcccgaaggctacattcaggagcgcaccatcttcttcaaggacgacggcaactataagac acgcgctgaggttaagttcgagggcgacactctggttaaccgcatcgagctgaagggcatcgacttcaagg aggacggcaacatcctgggccataagcttgaatataacttcaacTCCGGAggcggtttGTCTTCCAA ACGAGGTGATTTCAACAAATTTTGTTGATGAT
VgsVLR-R5	caccctgacctacggcgtgcagtgcttcagccgctaccccgaccacatgaagcagcacgacttcttcaa gtccgccatgcccgaaggctacattcaggagcgcaccatcttcttcaaggacgacggcaactataagac acgcgctgaggttaagttcgagggcgacactctggttaaccgcatcgagctgaagggcatcgacttcaagg aggacggcaacatcctgggccataagcttgaatataacttcaacTCCGGAGGTggcggtttGTCTTC CAAACGAGGTGATTTCAACAAATTTTGTTGATGAT
VgsVLR-R6	caccctgacctacggcgtgcagtgcttcagccgctaccccgaccacatgaagcagcacgacttcttcaa gtccgccatgcccgaaggctacattcaggagcgcaccatcttcttcaaggacgacggcaactataagac acgcgctgaggttaagttcgagggcgacactctggttaaccgcatcgagctgaagggcatcgacttcaagg aggacggcaacatcctgggccataagcttgaatataacttcaacTCCGGAGGTTCTggcggtttGTC TTCCAAACGAGGTGATTTCAACAAATTTTGTTGATGAT
VgsVLR-R7	caccctgacctacggcgtgcagtgcttcagccgctaccccgaccacatgaagcagcacgacttcttcaa gtccgccatgcccgaaggctacattcaggagcgcaccatcttcttcaaggacgacggcaactataagac acgcgctgaggttaagttcgagggcgacactctggttaaccgcatcgagctgaagggcatcgacttcaagg aggacggcaacatcctgggccataagcttgaatataacttcaacTCCGGAGGTGGCTCTggcggttt GTCTTCCAAACGAGGTGATTTCAACAAATTTTGTTGATGAT
VgsVLR-R8	caccctgacctacggcgtgcagtgcttcagccgctaccccgaccacatgaagcagcacgacttcttcaa gtccgccatgcccgaaggctacattcaggagcgcaccatcttcttcaaggacgacggcaactataagac acgcgctgaggttaagttcgagggcgacactctggttaaccgcatcgagctgaagggcatcgacttcaagg aggacggcaacatcctgggccataagcttgaatataacttcaacTCCGGAGGTGGCGGATCTggc ggtttGTCTTCCAAACGAGGTGATTTCAACAAATTTTGTTGATGAT
VgsVLR-R9	caccctgacctacggcgtgcagtgcttcagccgctaccccgaccacatgaagcagcacgacttcttcaa gtccgccatgcccgaaggctacattcaggagcgcaccatcttcttcaaggacgacggcaactataagac acgcgctgaggttaagttcgagggcgacactctggttaaccgcatcgagctgaagggcatcgacttcaagg aggacggcaacatcctgggccataagcttgaatataacttcaacTCCGGAGGTGGCGGATCTGG Tggcggtttgtcttccaaacgaggtgatttcaacaaattttgttgatgat
VgsVLR-R10	caccctgacctacggcgtgcagtgcttcagccgctaccccgaccacatgaagcagcacgacttcttcaa gtccgccatgcccgaaggctacattcaggagcgcaccatcttcttcaaggacgacggcaactataagac acgcgctgaggttaagttcgagggcgacactctggttaaccgcatcgagctgaagggcatcgacttcaagg aggacggcaacatcctgggccataagcttgaatataacttcaacTCCGGAGGTGGCGGATCTGG TGGAggcggtttGTCTTCCAAACGAGGTGATTTCAACAAATTTTGTTGATGAT
VgsVLR-R11	caccctgacctacggcgtgcagtgcttcagccgctaccccgaccacatgaagcagcacgacttcttcaa gtccgccatgcccgaaggctacattcaggagcgcaccatcttcttcaaggacgacggcaactataagac acgcgctgaggttaagttcgagggcgacactctggttaaccgcatcgagctgaagggcatcgacttcaagg aggacggcaacatcctgggccataagcttgaatataacttcaacTCCGGAGGTGGCGGATCTGG TGGAAGTggcggtttGTCTTCCAAACGAGGTGATTTCAACAAATTTTGTTGATGAT

Name	DNA sequence
VgsVLR-R12	caccctgacctacggcgtgcagtgcttcagccgctaccccgaccacatgaagcagcacgacttcttcaa gtccgccatgcccgaaggctacattcaggagcgcaccatcttcttcaaggacgacggcaactataagac acgcgctgaggttaagttcgagggcgacactctggttaaccgcatcgagctgaagggcatcgacttcaagg aggacggcaacatcctgggccataagcttgaatataacttcaacTCCGGAGGTGGCGGATCTGG TGGAGGAAGTggcggtttGTCTTCCAAACGAGGTGATTTCAACAAATTTTGTTGATGAT
VgsVLR-R13	caccctgacctacggcgtgcagtgcttcagccgctaccccgaccacatgaagcagcacgacttcttcaa gtccgccatgcccgaaggctacattcaggagcgcaccatcttcttcaaggacgacggcaactataagac acgcgctgaggttaagttcgagggcgacactctggttaaccgcatcgagctgaagggcatcgacttcaagg aggacggcaacatcctgggccataagcttgaatataacttcaacTCCGGAGGTGGCGGATCTGG TGGAGGAGGCAGTggcggtttGTCTTCCAAACGAGGTGATTTCAACAAATTTTGTTGATGAT
VgsVLR-R14	caccctgacctacggcgtgcagtgcttcagccgctaccccgaccacatgaagcagcacgacttcttcaa gtccgccatgcccgaaggctacattcaggagcgcaccatcttcttcaaggacgacggcaactataagac acgcgctgaggttaagttcgagggcgacactctggttaaccgcatcgagctgaagggcatcgacttcaagg aggacggcaacatcctgggccataagcttgaatataacttcaacTCCGGAGGTGGCGGATCTGG TGGAGGAGGCAGTGGCggcggtttGTCTTCCAAACGAGGTGATTTCAACAAATTTTGTT GATGAT
VgsVLR-R15	caccctgacctacggcgtgcagtgcttcagccgctaccccgaccacatgaagcagcacgacttcttcaa gtccgccatgcccgaaggctacattcaggagcgcaccatcttcttcaaggacgacggcaactataagac acgcgctgaggttaagttcgagggcgacactctggttaaccgcatcgagctgaagggcatcgacttcaagg aggacggcaacatcctgggccataagcttgaatataacttcaacTCCGGAGGTGGCGGATCTGG TGGAGGAGGCAGTGGCGGTggcggtttGTCTTCCAAACGAGGTGATTTCAACAAATTTT GTTGATGAT
VgsVLR-R16	caccctgacctacggcgtgcagtgcttcagccgctaccccgaccacatgaagcagcacgacttcttcaa gtccgccatgcccgaaggctacattcaggagcgcaccatcttcttcaaggacgacggcaactataagac acgcgctgaggttaagttcgagggcgacactctggttaaccgcatcgagctgaagggcatcgacttcaagg aggacggcaacatcctgggccataagcttgaatataacttcaacTCCGGAGGTGGCGGATCTGG TGGAGGAGGCAGTGGCGGTAGCggcggtttGTCTTCCAAACGAGGTGATTTCAACAAA TTTTGTTGATGAT
VgsVLR-R17	caccctgacctacggcgtgcagtgcttcagccgctaccccgaccacatgaagcagcacgacttcttcaa gtccgccatgcccgaaggctacattcaggagcgcaccatcttcttcaaggacgacggcaactataagac acgcgctgaggttaagttcgagggcgacactctggttaaccgcatcgagctgaagggcatcgacttcaagg aggacggcaacatcctgggccataagcttgaatataacttcaacTCCGGAGGTGGCGGATCTGG TGGAGGAGGCAGTGGCGGTGGAAGCggcggtttGTCTTCCAAACGAGGTGATTTCAA CAAATTTTGTTGATGAT
VgsVLR-R18	caccctgacctacggcgtgcagtgcttcagccgctaccccgaccacatgaagcagcacgacttcttcaa gtccgccatgcccgaaggctacattcaggagcgcaccatcttcttcaaggacgacggcaactataagac acgcgctgaggttaagttcgagggcgacactctggttaaccgcatcgagctgaagggcatcgacttcaagg aggacggcaacatcctgggccataagcttgaatataacttcaacTCCGGAGGTGGCGGATCTGG TGGAGGAGGCAGTGGCGGTGGAGGCAGCggcggtttGTCTTCCAAACGAGGTGATTT CAACAAATTTTGTTGATGAT
VgsVLR-R19	caccctgacctacggcgtgcagtgcttcagccgctaccccgaccacatgaagcagcacgacttcttcaa gtccgccatgcccgaaggctacattcaggagcgcaccatcttcttcaaggacgacggcaactataagac acgcgctgaggttaagttcgagggcgacactctggttaaccgcatcgagctgaagggcatcgacttcaagg aggacggcaacatcctgggccataagcttgaatataacttcaacTCCGGAGGTGGCGGATCTGG TGGAGGAGGCAGTGGCGGTGGAGGCAGCGGAggcggtttGTCTTCCAAACGAGGTG ATTTCAACAAATTTTGTTGATGAT
VgsVLR-R20	caccctgacctacggcgtgcagtgcttcagccgctaccccgaccacatgaagcagcacgacttcttcaa gtccgccatgcccgaaggctacattcaggagcgcaccatcttcttcaaggacgacggcaactataagac acgcgctgaggttaagttcgagggcgacactctggttaaccgcatcgagctgaagggcatcgacttcaagg aggacggcaacatcctgggccataagcttgaatataacttcaacTCCGGAGGTGGCGGATCTGG TGGAGGAGGCAGTGGCGGTGGAGGCAGCGGAGGCggcggtttGTCTTCCAAACGAG GTGATTTCAACAAATTTTGTTGATGAT
VgsVLR-M-Ext	tgctggagttcgtgaccgccgccgggatcactctcggcatggacgagctgtacaagggcggtaccggagg gagcatggtgagcaagggcgaggagctgttcaccggggtggtgcccatcctggtcgagctggacggcgac gtaaacggccacaagttcagcgtgtccggcgagggcgagggcgatgccacctacggcaagctgaccctg aagttcatctgcaccaccggcaagctgcccgtgccctggcccaccctcgtgaccaccctgacctacggc gtgcagtgcttcagccgctacc

Table S2. DNA fragments used in overlap-extension PCR to generate the linker library.