Advances in ligand-specific biosensing for structurally similar molecules

Chenggang Xi[1], Jinjin Diao[1], and Tae Seok Moon[1,2]*

[1] Department of Energy, Environmental and Chemical Engineering, Washington University in St. Louis, St. Louis, MO, United States

[2] Division of Biology and Biomedical Sciences, Washington University in St. Louis, St. Louis, MO, United States

* To whom correspondence should be addressed.

Tae Seok Moon

Tel: +1 (314) 935-5026

Email: tsmoon7@gmail.com

One Brookings Dr., Box 1180

St. Louis, MO 63130, USA

Abstract

The specificity of biological systems makes it possible to develop biosensors targeting specific metabolites, toxins, and pollutants in complex medical or environmental samples without interference from structurally similar compounds. For the last two decades, great efforts have been devoted to creating proteins or nucleic acids with novel properties through synthetic biology strategies. Beyond augmenting biocatalytic activity, expanding target substrate scopes, and enhancing enzymes' enantioselectivity and stability, an increasing research area is the enhancement of molecular specificity for genetically encoded biosensors. Here, we summarize recent advances in the development of highly specific biosensor systems and their essential applications. First, we describe the rational design principles required to create libraries containing potential mutants with less promiscuity or better specificity. Next, we review the emerging high-throughput screening techniques to engineer biosensing specificity for the desired target. Finally, we examine the computer-aided evaluation and prediction methods to facilitate the construction of ligand-specific biosensors.

Keywords: genetically encoded biosensor; directed evolution; protein engineering; molecular specificity; computational biomolecule design; machine learning; library screening; ligand-receptor interaction

Introduction

Living cells are endowed with a remarkable ability to sense and respond to a diverse array of stimuli, ranging from small molecules and ions to biomacromolecules and physical changes in their intracellular and extracellular environment. Through the integration of distinct genetic elements that enable cells to sense and respond to these stimuli, researchers have devised genetically encoded biosensors (GEBs) that fall into five major categories: allosteric transcription factor (aTF)-based biosensors [1], fluorescent protein (FP)-based biosensors [2], two-component system (TCS)-based biosensors [3], enzymatic biosensors [4], and riboswitch-based biosensors [5] **(Figure 1 & Table 1)**. These biosensors are highly versatile and have been employed for various applications, including metabolic engineering [6], environmental monitoring and remediation [7], diagnostics [8,9], and living therapeutics [10]. Each type of biosensor exhibits a unique target scope attributable to its inherent cellular function, and biosensors assembled by homologs originated from disparate organisms display distinct sensing preferences due to their respective host environments [11–13].

Although GEBs intrinsically offer greater specificity in detecting their targets compared with physicochemical methods, they are often incapable of distinguishing between multiple structurally similar chemicals that may coexist in the application environment, leading to crosstalk and undesired signal transduction, also known as promiscuity [12,14,15]. A partial solution is to focus on the most distinct ligands in the application environments [13,14]; however, the issue of sensor-target promiscuity persists. This challenge becomes especially pronounced when attempting specific biosensing in environmental and

biological samples with structurally similar chemicals, as is the case with soil, body fluids, living tissues, and similar matrices [7,9,16]. To address this, the integration of rational design and directed evolution methods emerges as a potent strategy to enhance the molecular specificity of GEBs [17].

Directed evolution represents a widely used approach to enhance the ligand-receptor specificity. Through multiple rounds of mutagenesis and screening the resulting variants [18], variants exhibiting heightened substrate specificity are isolated and further optimized [19] **(Figure 2)**. Structural and computational analysis of the sensor domain can guide the evolution process by identifying critical residues involved in substrate binding and predicting the impact of specific mutations on substrate specificity [1,20–22]. Recent advances in high-throughput screening and next-generation sequencing (NGS) technologies, combined with improvements in protein structure prediction and protein-ligand interaction modeling, have enabled the development of more refined and efficient strategies for biosensor evolution [23] **(Figure 3 & 4)**. In this regard, we summarize state-of-the-art strategies for developing highly specific GEBs that are resistant to interference from structurally similar chemicals, even when the latter constitute the substrate of the wild-type version.

1. Knowledge-based design principles for specificity control

In recent years, scientists have established a diverse range of highly specific GEBs by sophisticated design strategies **(Table 2)**. These approaches leverage either the innate specificity of genetic components or crafted macromolecular variants and chimeras for

the precise detection of specific chemical signals **(Figure 2)**. Here, we provide an overview of the rational and semi-rational design principles essential for the development, selection, and validation of specialized GEBs.

The innate specificity of different types of biosensing components

Biosensing components specialized for transmembrane signaling

An effective strategy to develop ligand-specific GEBs entails the assembly of pre-existing, specific genetic components from cell signaling systems. Natural sensing components offer an extensive array of options with varying degrees of molecular specificity for engineering GEBs **(Table 1)**. Bacteria predominantly employ TCSs to detect and adapt to fluctuations in their surroundings [3]. TCSs consist of a sensor histidine kinase (sHK) that probes a particular environmental cue and a response regulator (RR) that triggers the pertinent cellular response [24,25] **(Figure 1A)**. They can detect a broad range of inputs integral to bacterial growth, including metal cations, protons, small metabolites, and communication signals indicative of growth pressure, nutrient availability, and hormones [3]. While the majority of investigations have concentrated on the interactions between sHKs and RRs, or a RR and its promoter [26–28], only a handful of reports have assessed the full spectrum of sensing targets that sHKs are capable of distinguishing.

Most sHKs exhibit inherent specificity for their wild-type targets; while some can sense a restricted collection of chemically similar inputs, others are specific to a singular input [3,25]. For instance, *Escherichia coli* NarX demonstrates specificity for a terminal electron acceptor nitrate (NO_3^-) without interference from a closely-related compound nitrite (NO_2^-), whereas NarQ exhibits promiscuity [29]. Further examples like *Shewanella halifaxensis*

ThsS and *Shewanella baltica* TtrS are specifically activated by thiosulfate ($S_2O_3^{2-}$) and tetrathionate ($S_4O_6^{2-}$), respectively [30]. *E. coli* TorS-TorR could exclusively detect trimethylamine N-oxide ($(CH_3)_3NO$) among 117 common metabolites and solvents evaluated in a recent study [31]. The capacity of TCSs to selectively identify and react to specific stimuli renders them a compelling platform to develop biosensors for environmental signals [3].

Another class of intrinsically specific ligand-binding proteins (LBPs), G protein-coupled receptors (GPCRs), represents the most extensive category of cell surface receptors in eukaryotes **(Figure 1D)**. GPCRs facilitate cellular perception of environmental cues crucial for decision-making processes, such as proliferation, regulation of metabolism, immune functions, and neuronal circuit activities [32]. GPCRs transduce extracellular signals across the plasma membrane by activating intracellular G proteins, subsequently amplifying receptor responses through diverse downstream secondary messengers (cyclic adenosine monophosphate, inositol trisphosphate, or Ca^{2+}) that interact with their corresponding transcription factors [33]. Each GPCR type predominantly exhibits specificity for a distinct class of neurotransmitters (NT) or neuromodulators, though it is capable of binding closely related compounds with reduced activity [34]. Concurrently, various GPCR subtypes demonstrate differing degrees of ligand specificity and affinity within the same NT category, necessitating careful GPCR candidate selection for sensor development [35,36]. For instance, adrenergic receptors, which are GPCR subtypes responsive to both epinephrine and norepinephrine, can be further categorized into α- and β-adrenergic receptors based on their differential ligand sensitivity [37,38]. Likewise, dopamine receptors

can be subdivided into D1-like receptors (D1 and D5) and D2-like receptors (D2, D3, and D4), each possessing a unique ligand binding profile that ultimately constitutes the diverse specificity and sensitivity of dopamine biosensors [39–41].

TCSs and GPCRs are pre-existing biosensing components specific for their stimuli, a feature rooted in their fundamental roles in sensing the extracellular environment, which is crucial for cellular growth and cell-cell communication [42,43]. Nevertheless, these systems present certain limitations in universal applicability for biosensing purposes. Firstly, most TCSs function primarily in bacteria, while GPCRs predominantly operate in eukaryotes. Secondly, both sensors' target ranges are relatively constrained. TCSs primarily detect growth factors and stressors for cell survival, though they can sense various physical conditions [44]. The remodeling of sHKs is challenging due to their limited plasticity, thereby constraining the expansion of target compounds for TCS-based GEBs beyond the cognate signals of the specific sensor domain [45]. Endogenous ligands of GPCRs are restricted to hormones or NTs, and GPCR-based GEBs are often inhibited by synthetic antagonist drugs, complicating accurate NT detection during drug treatment [39,37,46,47]. Lastly, both GPCRs and TCSs require multistep signal transduction processes to modulate downstream gene expression, demanding increased effort to tune the biosensing performance [33,43].

Biosensing components specialized for intracellular signaling

Widely used biosensors are also based on one-component systems, consisting of an aTF and an output promoter featuring the corresponding transcription factor binding site [1]

(Figure 1B). In contrast to sHKs, aTFs are more apt to detect chemical signals rather than physical conditions due to their natural roles in regulating metabolic pathways [6]. As a result, the substrate scope of an aTF is typically broader than that of TCS, and aTFs can detect a wider range of organic compounds and their various analogs [48]. Additionally, aTFs excel at sensing intracellular metabolite levels, making them suitable tools for metabolic flux engineering and pathway evolution [49,50]. Non-cognate binding between ligands and aTFs, or aTFs and promoters, could cause crosstalk in gene regulation [13,51]. In terms of ligand specificity, aTFs that regulate the same promoter also display distinct substrate preferences when they originate from different organisms [11]. Therefore, a crucial step in developing a specific biosensor based on an aTF involves selecting the appropriate homolog, either with a preference for the target compound or with minimal activity toward the undesired target analogs.

Genome mining through sequence alignment or transcriptomic analysis has proven to be a highly effective approach to discover desired sensing elements [52,53]. The identified homolog possessing the requisite characteristics can then be introduced into the working organism for further optimization. d'Oelsnitz et al. examined six multidrug-resistant regulator candidates in *E. coli*, finding one to be highly active to tetrahydropapaverine (THP), while displaying only slight activity toward four other alkaloids [54]. They developed six specific alkaloid biosensors by using the responsive RamR regulator as a starting point for directed evolution [54,55]. In some rare instances, certain aTFs exhibit natural specificity for structurally similar chemicals, enabling their direct use as specific biosensors. For example, Diao et al. found several specific aTFs in *Rhodococcus opacus*

PD630 for aromatic compounds derived from lignin [56]. Furthermore, LhgR from *Pseudomonas putida* has been shown to recognize L-2-hydroxyglutarate (L-2-HG) as its specific effector molecule, even distinguishing it from its chiral isomer [57].

Enzymatic biosensors constitute the most established category of biosensors, despite enzymes typically being regarded as the most promiscuous building blocks for constructing biosensors [58]. Conventional output signals for these sensors include products and byproducts of enzymatic reactions, such as hydrogen peroxide (H_2O_2) [59], compounds with chromophores [60–62], redox cofactors [63], or electrical signals [64,65], making enzymatic biosensing an indirect measurement method **(Figure 1C)**. Luciferase represents one of the most commonly utilized enzymatic biosensors, converting cellular processes such as viability, protein-protein interactions, and gene expression activity into detectable light signals with the assistance of specific luciferin substrates [66]. Although enzymatic biosensors can exhibit exceptional sensitivity and rapid response, enzymatic reactions alter the target's concentration, causing the output signal to be influenced by reaction thermodynamics and kinetics. Furthermore, certain output signals are ubiquitously present in biomedical and environmental samples, challenging efforts to improve biosensor precision via protein engineering [4]. On the other hand, enzymes display the broadest substrate range and demonstrate the highest resilience to protein engineering [58]. Numerous studies have successfully modified substrate-binding specificity while preserving catalytic activities by altering ligand-binding pockets [20,67–69]. In addition, enzymatic reactions can transform noncanonical detection targets into

198 detectable compounds when combined with other biosensor types, extending the
199 chemical detection range of a biosensor through a plug-and-play modular approach [70].
200

201 ## Creation of ligand-specific, genetically encoded biosensor chimeras

202 ### Domain swapping for orthogonal signal transduction

203 For TCS-based GEBs, sensing a new signal is often achieved by replacement of entire
204 functional domains for most orthologous sHKs, owing to the lack of conserved sequence
205 within sensory domains [45]. The first chimeric sHK was constructed by fusing the sensory
206 domain of *E. coli* Tar chemoreceptor with the dimerization and histidine
207 phosphotransferase domain and catalytic and ATP-binding domain of a canonical sHK
208 EnvZ [71]. Given the innate ligand specificity, most research has concentrated on
209 enhancing signal transduction fidelity to improve biosensor performance, or on
210 transferring the developed biosensor to another organism [27,72]. Recently, Schmidl et al.
211 devised a general method for rewiring various TCSs to well-characterized output
212 promoters by modularly swapping RR DNA-binding domains (DBDs) [31]. This technique is
213 also beneficial for investigating the ligand specificity of certain uncharacterized sHKs,
214 facilitating assessments of sHKs' portability between bacterial species. As for GPCR-
215 based biosensors, several studies have crafted chimeric GPCR-G protein α subunit pairs
216 to transmit chemical signals into the downstream mitogen-activated protein kinase
217 cascade, which drives the expression of pheromone-responsive genes in yeast [73–76].
218 Researchers discovered that exogenous GPCRs exhibit orthogonality across their non-
219 cognate ligands when expressed in yeast [77]. This high degree of orthogonality enabled
220 ligand-specific mating of probiotic yeast equipped with corresponding heterogeneous

GPCRs and several ligand-specific GEBs [73,75]. However, the relatively limited dynamic range and slow kinetics have constrained the application of GPCR GEBs.

Modular assembly of one-component systems

aTFs, consisting of a ligand-binding domain (LBD) and a DBD, display a high degree of modularity for the development of chimeric biosensors [78–80]. Several ligand-specific aTF-based GEBs can be constructed by functional domain swapping [1,49]. For instance, BenR and XylS are AraC-type transcription regulators from *P. putida*, and BenR is specific to benzoic acid (Bz) among benzoate derivatives while XylS is more responsive to 3-methyl benzoic acid (3MBz) with a slightly lower activity to Bz [81]. Monteiro et al. developed a 3MBz-specific biosensor by replacing the LBD of BenR with that of XylS [81]. Similarly, Chang et al. designed aTF-based bile salts biosensors in *E. coli* by fusing LBDs of PBPs (periplasmic substrate-binding proteins) from enteropathogenic bacteria with *E. coli* DBDs, demonstrating that ligand specificity profiles of LBPs were swappable between species [82]. Furthermore, De Paepe et al. transferred the ligand specificity from an *E. coli*-incompatible *Sinorhizobium meliloti* NodD1 system to an *E. coli*-compatible *Herbaspirillum seropedicae* FdeR system [83]. They eliminated the substrate promiscuity of FdeR in *E. coli* by replacing the DBD or transcription factor binding sites of FdeR with those from the luteolin-specific regulator, NodD1. In addition, Rondon and Wilson found that the monomer−monomer interface and hinge region of the DBD are also crucial positions determining the specificity of a biosensor chimera [84].

In addition to protein-based GEBs, riboswitch-based systems also exhibit significant specificity and modularity **(Figure 1E)**. These biosensors are capable of detecting a diverse range of biological inputs, including ions, small molecules, proteins, and nucleic acids, by modularly integrating ligand-binding aptamers [85], protein-binding RNA scaffolds [86], or nucleic acid-mediated toehold switch [87]. Aptamers are short, single-stranded oligonucleotides that perceive chemical entities through the process of binding-induced alternative folding. Researchers have exploited a lot of ligand-binding aptamers, sourced either from genomic RNA pools or from random sequence libraries [88,89]. The discovery and validation process has been expedited by affinity-based enrichment techniques, such as systematic evolution of ligands by exponential enrichment (SELEX), which ensures molecular specificity for biosensor development [90] **(Figure 3)**. By integrating aptamers with various expression control elements, such as ribozymes [91], small RNA regulators [92], cis-regulatory elements [85,93], or CRISPR guide RNAs [94], along with a downstream reporter gene, researchers can construct highly orthogonal biosensing circuits. Moreover, aptamers can be coupled with electrical systems that transduce aptamer conformational changes into electrical signals. Researchers have developed a series of specific aptamer-based electrochemical biosensors using this approach [95–97].

Fusing reporting proteins with ligand binding proteins

An alternative approach for developing GEBs involves fusing FPs with LBPs such as GPCRs, to probe chemical signals **(Figure 1D)**. Upon binding their corresponding ligands, GPCRs undergo rapid conformational changes, inducing alterations in the fluorescence of circularly permuted fluorescent proteins (cpFPs), fluorescence resonance

energy transfer (FRET) pairs, split FPs, or bacteriophytochrome-derived near-infrared FPs [98]. FP-based GEBs are more sensitive with superior signal-to-noise ratios compared to transcriptional activation of FP expression mediated by GPCRs [72]. Additionally, FP-based GEBs are fast-acting and straightforward to test, facilitating directed evolution for different functional domains of a LBP [32]. The first step of the evolution process requires selecting suitable GPCR scaffolds from different subtypes or species depending on the target ligand. Subsequently, a cpFP is inserted into a candidate scaffold, and the performance of these chimeras is assessed. An ideal scaffold should demonstrate efficient membrane trafficking dynamics, a high initial dynamic range after cpFP insertion, appropriate affinity, and high selectivity for the target ligand. cpFP insertion site, linker, and cpFP optimization can be performed sequentially, and further tuning can be achieved by mutating GPCRs to refine affinity and specificity [32]. To date, this method has been systematically applied to develop selective and sensitive GPCR-based FP sensors for acetylcholine [99], dopamine [39,41], norepinephrine [37], adenosine [100], serotonin [47], ATP/ADP [101], and endocannabinoid [46].

In bacteria, periplasmic substrate-binding protein (PBP) scaffolds which function similarly to GPCRs scaffolds are appealing sensor engineering candidates [32]. These proteins typically consist of two LBDs connected by a hinge region where ligand binding can induce a conserved and substantial conformational change. Swapping the LBDs of PBPs is one main method for engineering the ligand specificity of a protein scaffold. For example, Scheib et al. changed the ligand specificity profile of a promiscuous PBP PotF by grafting the binding site of a putrescine-selective homologous PotD onto PotF, which

could be an ideal scaffold for a ligand-specific biosensor [102]. Moreover, PBPs can capture a more diverse range of molecular targets than GPCR, albeit with less specificity, broadening the potential scope of FP-based GEBs. Like GPCR-based GEBs, PBP-based GEBs can detect neuromodulators along with other small molecules such as methadone [103], serotonin [104], acetylcholine [105], glycine [106], trehalose [107], glutamate [108], and maltose [109]. However, it is important to note that GPCR/PBP-based FP GEBs are not truly specific to a single input but rather exhibit high selectivity. They can detect multiple ligands in addition to the most sensitive signal, with significantly diminished sensitivity outside the realm of physiological concentrations for *in vivo* applications.

Apart from GPCRs and PBPs, FP-based biosensors could also be constructed utilizing other LBPs such as the LBDs of aTFs or enzymes, plant hormone receptors, and lipocalins [59,110,111]. Beltrán et al. rapidly developed 21 biosensors specific to their cognate signals by modifying a plant hormone receptor with a malleable ligand-binding pocket [112]. Herud-Sikimić et al. generated a FRET-based biosensor in plant tissues and reshaped the TrpR binding pocket for real-time auxin-selective biosensing [113]. Kang et al. developed a specific FRET-based biosensor based on *P. putida* aTF LhgR to detect L-2-HG, a biomarker for a variety of cancers [57]. Fan et al. developed a cpFP-based biosensor by engineering a redox relay between the active-site cysteines of human thioredoxin peroxidase to specifically sense thioredoxin redox [59]. Moreover, inverse sensing where ligand binding inhibits the fluorescence activity has also been investigated, providing another option for FP-based GEB development from different kinds of LBPs [114].

Identification of mutation hotspots in the ligand-binding domain

Crosstalk is a pervasive occurrence in biological systems, complicating the optimization of orthogonality for multi-input biosensing applications [15,115]. Enhancing the orthogonality of a ligand-receptor pair necessitates the generation of binding cavities that stabilize the target molecule's binding and eliminate the potential binding to structurally similar molecules [21]. A recent study also revealed that alterations in ligand specificity of the aTF BenM are predominantly governed by mutations in the LBD [116]. Therefore, the LBD of a biosensor is typically the region most targeted for manipulation, based on structural analysis and sequence-function relationships, when attempting to discriminate against the binding of undesired ligands [21]. Potential mutation sites can be identified through conserved residue analysis, protein structural analysis, reported functional studies, and *in silico* predictions **(Table 2)**. Techniques such as multiple sequence alignment, homology modeling, and molecular docking are commonly used to identify key residues in the LBD for specificity control [117]. Single-alanine-substitution scanning can also roughly pinpoint sensitive ligand-binding sites [118]. The subsequent evolutionary process can modify specificity from natural substrates to new ligands or eliminate undesired activity for native substrates [119,120].

For TCS-based GEBs, the most common LBD of the sHK is the Per-Arnt-Sim (PAS) domain, which is ubiquitous across all kingdoms of life [43]. The target signal is perceived by ligand binding to the PAS domain cavity, cofactor-containing PAS domains, and PAS domain-membrane interface, or by modulation of inter-PAS domain disulfide bonds [25]. However, there exists only minimal sequence conservation within these PAS domains, impeding the annotation of mutation hotspots and the protein engineering for ligand-

binding specificity [25]. Furthermore, agonists and antagonists can bind to the same domain of a sensor kinase with similar affinities [121], complicating the generation of a truly specific sHK ligand-binding pocket without interference from other molecules. To date, no successful attempt has been made to enhance the ligand specificity of an sHK by remodeling the PAS domain, given the complex multistep signal transduction process. Nevertheless, Landry et al. demonstrated a phosphatase tuning method to adjust the detection thresholds of several TCS pairs by mutating the first variable residue in the conserved transmitter domain GXGXG motif to different hydrophobic residues [122].

Bacterial aTFs are often intrinsically promiscuous biosensing generalists [13,123]. Therefore, the evolution of an aTF from a generalist to a specialist is essential to develop ligand-specific GEBs. Initial efforts to engineer inducible promoters for synthetic biology applications involved structure-guided site-saturation mutagenesis (SSM) or random mutagenesis of aTFs' LBDs [1]. Notable examples include engineering TetR [124–126], AraC [127,128], LacI [129], and LuxR [14]. In eukaryotes, ligand-activated TFs can also be evolved by structure-guided SSM to recognize synthetic compounds for conditional gene expression [17,130,131]. For example, Chockalingam et al. engineered the ligand specificity of a human estrogen receptor through a combination of random mutagenesis and SSM within the LBD of the receptor [132]. They successfully altered the receptor's specificity from its natural ligand, 17-β-estradiol, to 4,4'-dihydroxybenzil after five rounds of evolution. Recently, we demonstrated that a single mutation in the vicinity of ligand-binding sites could confer specificity to promiscuous biosensors [133]. In our work, we substantiated that TyrR, which can bind both phenylalanine and tyrosine via two distinct ligand-binding pockets, can be

engineered for specific biosensing by disrupting one of these pockets. We also engineered a TynA-FeaR biosensing system consisting of a monoamine oxidase TynA that converts various neuroactive monoamines into targets detectable by the aTF FeaR [133]. Both components were modified via SSM of key residues in the LBD predicted by homology modeling. The introduction of steric hindrance enabled the specific biosensing of phenethylamine, while the incorporation of smaller and hydrophilic residues created tyramine-specific GEBs. Similarly, Herud-Sikimić et al. engineered the binding pocket of TrpR to selectively sense auxin from 23 different indole derivatives, based on structural analysis and iterative SSM of ligand-interacting residues in the vicinity of the amino group of its substrates [113]. Specifically, they discovered that the binding poses of tryptophan and indole-3-acetic acid are distinct, and a single mutation at position 88 from serine to tyrosine (S88Y) could selectively abolish the binding of tryptophan while stabilizing the binding of indole-3-acetic acid.

In the case of FP-based GEBs, the linker region, FP domain, and receptor sequence are subjected to systematic SSM to optimize detection sensitivity and response dynamic range under varying physiological conditions [32]. Most mutations do not alter molecular specificity, except for those within LBDs of receptors [134]. Recently, Zhang et al. developed a glycine-selective FRET biosensor based on *Agrobacterium tumefaciens* PBP Atu2422 which binds to glycine, serine, and γ-aminobutyric acid [106]. They computationally designed 1,000 variants using the FoldX program and assessed them individually through ligand docking with Autodock, followed by isothermal titration calorimetry experiments. By introducing steric obstructions within the Atu2422 binding site, they successively

eliminated promiscuous binding to serine and γ-aminobutyric acid, yielding a biosensor specific to glycine. However, leucine, valine, and threonine remained as co-agonists, albeit with 10-fold lower sensitivity than glycine. Similarly, Feng et al. engineered and characterized a norepinephrine sensor based on the adrenergic receptor [37]. By introducing a T6.34K mutation, the norepinephrine detection sensitivity of the biosensor increased by 10-fold, which is 300-fold more sensitive compared to dopamine, though the sensor is still responsive to epinephrine. Borden et al. developed an acetylcholine biosensor from a PBP OpuBC of *Thermoanaerobacter sp.* X513, which bound both choline and acetylcholine with a higher affinity for choline [105]. Guided by structure modeling, they modified the binding pocket to increase hydrophobicity and aromaticity through the introduction of F219W, E174F, R178G, and K39I mutations. Consequently, the biosensor's specificity and affinity shifted towards acetylcholine, becoming insensitive to other neurochemicals except serotonin. Using this acetylcholine sensor as the starting point, Unger et al. redesigned the ligand-binding pocket of OpuBC to bind serotonin while eliminating binding to acetylcholine and choline [104]. They experimentally screened a total of 2,576 variants guided by computational design and a machine-learning model, and then combined frequent mutations with higher specificity. Finally, they obtained a variant with 5,000-fold increase in serotonin binding specificity compared to the original version by introducing 19 mutations.

A recent study by Muthusamy et al. generated a selective real-time cpFP-based biosensor for S-methadone by engineering the nicotine-binding OpuBC variant [103]. The ligand-binding pocket of OpuBC can accommodate methadone by aromatic residues F12,

Y65, Y357, and Y460, so they applied iterative SSM close to those residues and developed a S-methadone-selective variant via three mutations W436F, N11V, and L490A [103]. In another case, Z. Li et al. intended to develop a uridine diphosphate N-acetylglucosamine (UDP-GlcNAc) sensor but were unable to eliminate the UDP/UTP binding to the UDP-GlcNAc transferase backbone [110]. Nevertheless, they have succeeded in abolishing UDP-GlcNAc activity while keeping UDP/uridine triphosphate (UTP) activities by random mutagenesis for UXP biosensing.

By introducing the steric effects, π interactions, or hydrophobicity changes into the ligand-binding pocket, researchers could effectively discriminate the small ligand from the larger chemicals, while it is still difficult to distinguish large compounds from small molecules [21]. Besides, long-range effects triggered by mutations in FP linker regions [32,113], crucial allosteric communication sites [135,136], or protein-binding domains (PBDs) such as dimeric interfaces [84,129,137] may also induce changes in ligand-binding affinities. Yet, the effects of these distant mutations on molecular specificity are difficult to predict by simple inspection. Additionally, the accurate prediction of aptamer 3D structures is challenging due to their high degree of flexibility, leading most modifications of riboswitch-based biosensors to be performed through high-throughput screening instead of fine-tuning via point mutations [138]. To overcome these issues, the use of high-throughput screening and computer-aided design approaches, in conjunction with knowledge-based structural analysis, has been considered and will be discussed in subsequent sections of this review.

2. High-throughput screening and selection methods

In many instances, structure-guided single-codon mutagenesis may not consistently yield ligand-specific genetic sensors. Consequently, combinatorial libraries created by iterative random mutagenesis or pairwise SSM become the most typical way to thoroughly explore desired residue substitutions in multiple mutation hotspots. To combat issues of redundant genotypes and bias in residue substitutions due to codon redundancy, researchers have employed codon degeneracy to construct mutagenesis libraries of minimal size [139]. Nevertheless, as the number of potential mutation sites increases, the number of possible mutation combinations expands exponentially, rendering it impractical to examine all variants individually. To this end, efficient screening methods must be implemented to isolate the desired variant with the required properties from extensive libraries **(Figure 3)**.

Fluorescence-activated cell sorting (FACS)

For most GEBs, FPs, luciferases, or pigment-producing enzymes serve as quantitative reporters of biosensing readout. The conventional method for colorimetric assays involves blue-white colony screening on agar plates, which leverages β-galactosidase activity. With the application of FPs and flow cytometry, automated fluorometric sorting techniques have enabled directed evolution of various biosensors with a higher throughput. The basic workflow of FACS involves categorizing input cells based on fluorescence levels. In the absence of target ligands or the presence of undesired ligands, where biosensor readout is expected to be null, FACS picks the bottom 1-5% of cells with the lowest fluorescence levels. Meanwhile, in the presence of desired ligands where biosensor readout is expected to be robust, FACS selects the top 1-5% of cells with the

highest fluorescence levels when induced by desired ligands. Previously, FACS has been utilized as the screening method when evolving AraC to sense multiple unnatural compounds, allowing the engineered variants to function as inducible promoters for metabolic engineering [127,128,140]. Additionally, others employed FACS as the negative selection step and then test the cells individually for positive hits when targeting more than two ligands [141,142]. Recently, FACS-based screening has been used to engineer additional aTFs to sense non-native ligands or eliminate native targets, including HucR mutants for shikimic acid [143], PobR for p-nitrophenol and 3,4-dihydroxy benzoate [144], PcaV for vanillin [145], VanR for vanillic acid [146], and TtgR for resveratrol [147]. Beyond aTF-based GEBs, FACS can also facilitate the screening process for riboswitch-based and GPCR-based GEBs, yielding specific variants with high fold-change activities [74,148].

FACS not only distinguishes fluorescence in the presence or absence of desired inducers but also discerns variations in color and particle size for affinity-based selection methods, such as aptamer SELEX [149] and surface display [150,151]. Typically, target ligands or sequences are linked to their corresponding fluorescent conjugates in SELEX or surface display experiments. During the affinity-based enrichment step, the ligand-bound population exhibits greater size and higher fluorescence intensity than the unbound population, enabling FACS to sort cells by color, fluorescence intensity, or size, and is iteratively enriched. In addition, advancements in microfluidics have allowed researchers to grow large-size variant libraries individually within a single well or on a chip and test each variant in a droplet, significantly escalating selection capabilities compared to traditional 96-well plate assays. Ma et al. developed a dual-channel microfluidic droplet

screening platform and employed dual-color FACS to assess the product enantioselectivity of esterase mutants from *Archaeoglobus fulgidus* [61]. This microfluidic high-throughput screening system could evaluate more than 10^8 droplets (~10^7 enzyme variants) per day, rapidly identifying enantioselective variants. When further combined with continuous evolution, FACS-based selection could become more powerful, eliminating the need for repetitive DNA extraction and diversification from post-selection populations. Javanpour and Liu integrated a continuous hypermutation system called OrthoRep into FACS-based counter-selection cycles, reprogramming the specificity of BenM from muconic acid to adipic acid [152].

Growth-based selection

In instances where the expression platform involves selection markers such as antibiotic resistance genes [54,133], toxic protein [129], or enzymes for essential metabolism [69,137,153], growth-based selection may isolate top-performing variants. Similar to FACS-based selection, multiple rounds of counter-selection are required to eliminate variants with leaky expression or undesired activities [14]. Taylor et al. employed a hybrid selection process, combining growth-based negative selection and FACS-based positive selection for LacI, to avoid the limited resolution of flow cytometry at low fluorescence levels [129]. After identifying initial hits with single-codon substitutions that bind to the desired ligand, they shuffled beneficial mutations from promiscuous variants, enhancing specificity for gentiobiose and sucralose over the wild-type ligand isopropyl-β-D-1-thiogalactopyranoside. Similarly, Ogawa et al. utilized growth-based dual-selection by expressing two antibiotic resistance genes connected by a genetic inverter, successfully obtaining a 4-methylbenzoic acid (4MBz)-specific XylS variant through two mutations [154].

Fluorescent proteins or colorimetric markers can also aid in the selection process by identifying the brightest colonies on agar plates, providing an optimal starting point for the subsequent round of directed evolution iterations [54,55]. However, due to the toxic environment exerting selection pressure, unexpected mutations may arise from low genetic stability or PCR errors, necessitating careful selection of appropriate expression levels of selection markers or antibiotic concentrations [129,133]

Compartmentalized partnered replication (CPR) represents another high-throughput screening method that can enrich variants with enhanced expression levels during cell growth [155]. Rather than using selection markers, researchers express a thermostable DNA polymerase as the reporter and amplify positive hits via compartmentalized PCR. This process allows variants with higher polymerase levels to produce more DNA templates, preserving beneficial mutations for successive rounds of evolution. Ellefson et al. evolved an aTF TrpR to control the expression of Taq polymerase, utilizing CPR to enrich variants with specific sensing abilities for synthetic tryptophan analogs, 5- or 6-bromo-L-tryptophan [156]. Moreover, Meyer et al. systematically evolved 12 aTF-based GEBs using growth-based negative selection and CPR as a positive selection method to minimize reporter expression leakiness and crosstalk between noncognate promoters [115]. Overall, growth-based counter-selection serves as a convenient and cost-effective *in vivo* screening method, despite potentially longer time and lower throughput compared to the FACS method.

Deep mutational scanning (DMS)

Most research has employed random mutagenesis or focused library design to evolve promiscuous sensor candidates. However, random mutagenesis often generates synonymous, redundant, or suboptimal mutations, while comprehensive screening of focused-mutagenesis designs proves expensive and labor-intensive. To overcome these limitations, DMS has emerged as a massively parallel method that can walk through a large number of mutants benefiting from deep sequencing technologies and enrich the variants bearing improved function by accessing the fitness of each variant under selection pressure [157]. DMS is a simultaneous selection and measurement process for sensor-target pairs, eliminating the need to isolate individual variants [158]. The workflow of DMS is summarized as follows: 1) Generate a biomacromolecule library through random mutagenesis or unbiased structure-guided SSM and assign a unique barcode for each variant. 2) Subject the library to a selection process, altering the density of each variant in response to selection pressure, known as enrichment. 3) Segregate the library into several subpopulations after the enrichment process according to the selected phenotype. 4) Extract DNA from each population with or without selection and utilize high-throughput sequencing to determine the frequency of each variant across different bins. 5) Derive enrichment scores from sequencing results to form a sequence-function fitness landscape, which could indicate beneficial or deleterious mutations.

When combined with artificial enrichment processes such as SELEX [159] and protein display technologies [160] as well as FACS [161,162] or growth-based selection [137,163] processes previously discussed, DMS can examine evolutionary protein-RNA, protein-DNA, RNA-ligand, or protein-ligand binding capabilities effectively [157]. CPR also has the

potential to be integrated with DMS for quantitative assessment of variants' fitness under selection pressures, preventing low sequencing fidelity caused by insufficient samples after selection. DMS has been extensively applied to quantify the epistatic mutation effects on human diseases [164], protein stability, activity or enantioselectivity [165–167], protein-protein interactions [150,168,169], and enzyme-substrate specificity [137,170].

Intriguingly, Wrenbeck et al. discovered that beneficial mutations enhancing enzyme-substrate specificity were distributed globally throughout the sequence and structure, with most residing 9-20 Å away from active sites [137]. They also observed that beneficial mutations were imbalanced among three different substrates, with a certain correlation of fitness between them. Although the "specificity-determining" mutation described in this study does not necessarily indicate exclusive binding to a single ligand, it emphasizes that advantageous mutations may extend beyond ligand-interacting residues, and DMS has proven to be an effective tool in uncovering those distal mutation hotspots. In a similar vein, Ogawa et al. conducted DMS on XylS and employed an antibiotic growth-based dual screening system to enrich XylS variants with enhanced specificity for structurally-similar 3MBz and 4MBz, respectively [171]. They randomized 213 residues within the XylS LBD using single-codon substitutions and calculated each variant's frequency to delineate the mutational fitness landscape under antibiotic selection pressure. They then focused on residue G71, which exhibited the highest fitness scores, and identified 4MBz-specific variants that lacked activity towards its native substrate. More recently, Meier et al. revealed the ligand-binding residues of ATP-binding cassette transporter EfrCD after the DMS process targeting three drugs [172]. They also discovered that a single mutation could

transform wild-type EfrCD into a Hoechst-specific importer, which could be repurposed as a biosensor by coupling it to a TCS-based regulatory system. However, a single mutation may not be sufficient to adjust the ligand-bound conformation for most LBPs. To make the method more generalizable, computational models are required to predict combinatorial mutations that enable specificity shifts based on DMS databases.

3. Computer-aided evaluation and prediction of biosensor specificity

Owing to the limited screening capacity restricted by the transformation efficiency or measurement resolution, it is infeasible to fully explore the sequence space of combinatorial mutagenesis libraries experimentally. As an alternative, computer-aided protein design can assess protein-ligand interactions for mutants of interest more efficiently and predict potentially specific variants exhibiting optimal ligand-binding compatibility [173]. At the atomic level, the objective of the forward protein engineering process is to precisely position amino acid side chains around the ligand to achieve optimal orientations. At the protein level, geometrically compatible ligand-receptor interfaces should be tailored specifically for the target ligand based on first principles [174,175]. Researchers have predominantly depended on ligand-receptor scoring functions to steer the selection of best-performing design candidates while discarding unfavorable ones. To improve their ranking performance, these functions integrated a myriad of factors such as cheminformatics of ligand-binding residues, geometric information in the form of distance matrices, functional group interaction fingerprints, and binding free energy calculations drawn from knowledge-based functions, molecular dynamics, or

Monte Carlo simulations [176]. Computational tools like Rosetta can provide precise calculation of interaction energy as well as fast prediction for protein folding and molecular docking simulations, thereby offering a higher throughput for selecting promising variants *in silico* [177]. With the development of machine learning methods, combinatorial position-specific mutations based on ligand-receptor affinity prediction may be increasingly pivotal, especially for sensors that lack accurate structural information [178].

Structure-guided computational protein design and *in silico* selection

In practice, the variant library size could grow exponentially with the increase of mutation positions, quickly exceeding the transformation efficiencies typically observed for yeast (up to 10^6 variants), *E. coli* (up to 10^7 variants), and phage (up to 10^{12} variants). Consequently, the number of sequences to be scrutinized is restricted, necessitating the optimization of the mutational space through computer-guided library design before experimental screening [179]. The typical *in silico* selection pipeline to computationally design a specific LBP is summarized as follows: 1) Obtain the structure of the sensing protein via crystallography, homology modeling, or *de novo* design of the protein backbone with desired geometries. 2) Identify active sites through molecular docking or consensus analysis and reshape the LBD using focused or saturation mutagenesis *in silico*. 3) Predict beneficial mutations by comparing ligand-binding affinities calculated via interaction energy function or molecular dynamics simulations. 4) Assess ligand specificity by computationally comparing the complementarity of each ligand and experimentally verify the top-ranking variants. Additionally, steps 3) and 4) could be performed experimentally using DMS, which yields the most stable scaffold variants for

614 further modifications [160,161,180]. This hybrid approach enables researchers to explore a
615 more extensive design space than traditional directed evolution methods.
616
617 Computational pre-screening can effectively eliminate a large number of mutants with
618 unfavorable protein scaffolds and identify those with the most complementary ligand-
619 protein interfaces. Looger et al. re-engineered the ligand-binding sites of five PBPs
620 (glucose-binding protein, ribose-binding protein, arabinose-binding protein, glutamine-
621 binding protein, and histidine-binding protein) to alter the ligand specificity for nonnative
622 substrates [181]. They generated 10^{45} to 10^{68} mutant structures *in silico* by mutating 12-18
623 amino-acid residues in direct contact with their wild-type ligands and ranked them based
624 on minimized binding energy calculated using dead-end elimination theorems. The
625 energy function considered molecular shape, chirality, hydrogen bonding, molecular
626 surface (polar, aliphatic, aromatic, charged, and cationic), and water solubility. Seventeen
627 top-ranking variants were experimentally examined and exhibited selective molecular
628 recognition for their new targets. Finally, they created a chimeric TCS biosensor based
629 on ribose-binding and glucose-binding proteins responsive to trinitrotoluene and L-
630 lactate. Furthermore, Lippow et al. engineered the substrate specificity of galactose 6-
631 oxidase to respond to glucose by a semi-rational approach, which computationally
632 predicted favorable mutants with better ligand compatibility and synthesized them for
633 experimental selection of positive hits [182]. Similarly, Jha et al. expanded the specificity of
634 an aTF PobR for 3,4-dihydroxy benzoate without interference from structurally-similar 2-
635 hydroxy benzoate using Rosetta-assisted library design [183]. However, the evolved aTF-
636 based sensor remained active for the native inducer 4-hydroxy benzoate, highlighting the

challenge of developing a truly specific biosensor by computational prediction alone. Subsequently, they screened the same library and find a specialist with switched specificity from 4-hydroxy benzoate to p-nitrophenol by four rounds of FACS [144]. Likewise, Unger et al. optimized the serotonin-binding pocket of a PBP-based biosensor using ligand docking and binding efficiency prediction for 250,000 variants through RosettaLigand [104]. They selected 18 top-ranking variants for further diversification by SSM of four crucial residues predicted by random forest modeling and combined the beneficial mutations using a generalized linear model to identify the best-performing sensor variant. However, this technique relies on high-resolution, three-dimensional protein structures for accurate binding energy calculations, which limits its broader applicability in LBP design when high-quality crystal structures are unavailable.

Alternatively, *de novo* design approaches allow researchers to define geometric positions and orientations of ligand-binding residues and search for a large number of available protein scaffolds to accommodate orchestrated ligand interactions [184] **(Figure 4A)**. Tinberg et al. pre-organized ligand-interacting residues geometrically around the ligand and placed the motif into 401 protein scaffolds [185]. The conformational fitness of each protein-ligand complex was evaluated by RosettaMatch, and top-ranking scaffolds were selected for further modification. Next, they computationally designed surrounding ligand-binding residues by RosettaDesign programs for optimal shape complementarity, ligand interaction energy, and protein stability. Finally, SSM libraries with one to three substitutions on 39 residues were experimentally screened to determine the ideal ligand-binding affinities via yeast display and FACS. Likewise, Glasgow et al. developed a FP-

based biosensor for farnesyl pyrophosphate by modeling farnesyl pyrophosphate-binding motifs within 3,463 compatible protein complex scaffolds [186]. With increasingly accurate protein folding simulations from algorithms like RoseTTAFold and AlphaFold2, scientists can efficiently generate artificial protein scaffolds featuring the desired pocket shape and tunable geometries for specific molecular interactions [187–189]. Most recently, Yeh et al. designed artificial luciferases from scratch that specifically bind synthetic luciferin diphenylterazine (DTZ) and 2-deoxycoelenterazine (h-CTZ) [190]. They selected nuclear transport factor 2-like superfamily as the topology from 4,000 small-molecule binding proteins due to their appropriate shape complementarity after ligand docking. Next, they employed a deep-learning-based 'family-wide hallucination' approach to create ligand-binding pockets that stabilize the anionic state of DTZ or h-CTZ and predicted the structure by trRosetta. Ultimately, they screened 7,648 designs for DTZ and 46 designs for h-CTZ based on ligand-binding energy, protein-ligand hydrogen bonds, shape complementarity, and contact molecular surface, and they introduced SSM to further improve the luciferases activity. These studies achieved selective biosensing by computational design of the ligand-binding pocket for the desired substrate only, demonstrating the importance of shape complementary for specific biosensing.

These computer-aided protein design methods allowed researchers to bypass bumpy mutational trajectories and tremendously circumvent laborious experimental iterations. However, few studies considered negative design to exclude structurally similar chemicals during the design process. To better improve the ligand specificity of computationally designed biosensors, researchers could consider negative designs using

683 ultra-large library docking platforms and calculate the target-to-decoy ratio to eliminate
684 promiscuous designs [191–193].
685
686 **Data-driven ligand specificity prediction using machine learning and**
687 **neural network**
688 Molecular docking simulations can provide reliable predictions for potential ligand-
689 binding; however, their accuracy remains limited [194]. To refine the ranking ability in ligand
690 docking simulations, machine learning methods have been extensively employed [195].
691 Nonetheless, the structure-oriented docking process is computationally resource-
692 intensive, especially when dealing with hundreds of thousands of variants with subtle
693 sequence changes. Instead of relying solely on the structure-guided computational
694 design, advancements in NGS techniques and machine learning models enable
695 researchers to explore larger sequence spaces and predict ligand-protein interactions
696 with minimal experimental effort [196,197] **(Figure 4B)**.
697
698 As the size of available databases continues to grow, accurate prediction of family-wide
699 enzyme-substrate compatibility via high-throughput virtual screening becomes
700 increasingly crucial for industrial biomanufacturing and drug discovery [192,198]. Currently,
701 many studies have laid the groundwork for accurately predicting enzyme-substrate
702 promiscuity by integrating sequence and structural information [198,199], enhancing enzyme
703 feature descriptors [200], expanding training databases [178], and examining various models
704 [201,202]. For example, Robinson et al. developed enzyme-substrate regression models for
705 the OleA family of thiolases, considering 153 chemical characteristics of residues within
706 12 Å of the active site [203]. Notably, they identified the cavity size of the ligand-binding

pocket as a major determinant for the binding of bulky substrates, corroborating earlier findings by Martínez-Martínez et al. [204]. Additionally, Ollikainen et al. applied Rosetta-based computational methods to enhance prediction accuracy for mutations altering enzyme specificity, accounting for the coupled flexibility of protein backbone, ligand, and ligand-binding residues [205].

Beyond the conventional approaches of empirical energy functions or force field-guided simulations, researchers have devised several machine learning or deep neural network (DNN) models to predict protein-ligand binding affinity [206–210]. These advanced prediction models demonstrate the capability to discern ligand specificity in novel proteins and predict mutations that can alter ligand specificity for new ligands [211,212]. Notably, Chatterjee et al. established an artificial intelligence-based pipeline AI-bind that predicts the binding sites and probability of diverse protein-ligand pairs, including 26 SARS-CoV-2 viral proteins and 332 human proteins [213]. In addition, Rube et al. developed a machine learning model called ProBound, which predicts TF-DNA binding affinity and quantifies sequence recognition specificity from massively parallel sequencing data [208]. Nevertheless, the specificity of other regulatory LBPs was rarely investigated by machine learning models over the past decades, possibly due to the lack of comprehensive and high-quality databases to support analyses in these areas.

As previously discussed, DMS methods have been extensively applied to map genotype-phenotype relationships and delineate mutational fitness landscapes for a broad array of proteins, thereby providing enormous training datasets [214]. Fitness scores that quantify

the mutation effects can either be determined through DNA enrichment methods following selection pressure or be computed by free energy differences derived from empirical energy functions [164]. Researchers have integrated Monte Carlo-based computational screening and growth-based selection techniques to predict and validate mutation hotspots that enhance enzyme activities by calculating the sequence density of enriched variants [215]. For the past few years, the increasing availability of massively parallel mutagenesis datasets has enabled the broader application of statistical learning for evolutionary sequence variation, resulting in more efficient predictions of sequence-function relationship [214,216,217].

Still, most machine-learning models trained on DMS datasets focused on mutation effects on protein stability, enzyme activity, antibiotic resistance, protein-protein interactions, or human diseases [164,165,217–221], rather than biosensing specificity. For example, Wu et al. trained a machine-learning model using hundreds of selected variants and predicted the fitness landscape of a nitric oxide dioxygenase from *Rhodothermus marinus*, encompassing a library with seven substitution positions [173]. Despite the model considering a mere seven mutation sites, it demonstrated substantial predictability, guiding the evolution of enzyme enantioselectivity and identifying several (S)- and (R)-selective catalysts through prediction. Furthermore, several comprehensive DMS studies have been conducted on full-length aTF sequences, employing machine-learning models to predict their allosteric communications modulated by corresponding ligands [222–226].

Attributable to advancements in mapping sequence-function relationships, mutational fitness landscapes present a more informative approach to effectively evaluate ligand specificity and predict functional mutations in comparison to conventional affinity-based prediction models. Notably, Tack et al. measured the fitness landscape of 62,472 LacI variants after antibiotic-based growth enrichment and trained a DNN model to predict transfer curves for all possible variants [225]. Their accurate predictions of the EC_{50} for selected LacI mutants shed light on the potential investigation of aTF ligand specificity by measuring mutational fitness landscapes when induced by multiple ligands. These studies have contributed to an expanding toolbox that can streamline the design-build-test-learn cycle for ligand-specific biosensor development in the future.

4. Conclusion

The accurate measurement and precise control of interactions between ligands and macromolecules have been long-standing objectives in the realm of allosteric macromolecule engineering and biosensing. In this review, we have summarized 1) rational design principles for generating suitable starting points, 2) high-throughput screening techniques to enhance biosensing performance, and 3) computational design models that facilitate the design-build-test-learn cycle **(Figure 2)**. Additionally, the design and screening methods for generating specific genetic elements could contribute to the creation of superior biosensors and also support the engineering of upstream or downstream enzymes [143], transporters [227], regulators [33], and related metabolic pathways [54,55].

Generally, any protein or nucleic-acid switch capable of undergoing allosteric conformational changes can be engineered as a GEB. Numerous studies have already improved protein-protein binding specificity, encompassing antibodies [228], protein interfaces [229], proteases [151,230], and protein inhibitors [70]. Additionally, enzyme specificity for native cofactors can be altered to accommodate synthetic compounds for conditional genetic control [63,231]. Thus, the molecular specificity of biosensors can extend to a wider array of chemical or biological targets such as proteins [232], nucleic acids [233], ions [234], pathogens [33,235], and ones beyond the small molecules discussed herein.

Still, universal engineering pipelines for specificity control of distinct GEBs remain elusive, but existing technologies have been combined to achieve novel functions that surpass the capabilities of natural evolution spanning millions of years. In the future, such synthetic evolution approaches will expedite the discovery of optimized molecular interactions, paving the way for accurate diagnostics, rapid drug discovery, and large-scale biomanufacturing. The specific genetic components outlined here possess the potential to empower researchers with the ability to control complex biosystems in a more precise and quantitative manner. By detecting a wider range of chemical or biological signals with enhanced specificity, we can tackle the problems posed by complex environments more effectively and accomplish increasingly challenging global tasks in a safer and more sustainable fashion.

Acknowledgements

We thank Matthew Lima, Sandra Matteucci, and Allison Creighton for helping revise this manuscript. This work was supported by the National Science Foundation [MCB-2001743 to T.S.M.], National Institutes of Health [RO1 AT009741 to T.S.M.], the United States Department of Agriculture [2020-33522032319 to T.S.M.], and the U.S. Environmental Protection Agency [84020501 to T.S.M.]. The content is solely the responsibility of the authors and does not necessarily represent the official views of the funding agencies.

Conflict of Interest

The authors declare no conflict of interest.

Figure legend

Figure 1. Overview of Five Distinct Types of Biosensors and Their Specificity Control Methods. (A) TCS-based biosensors: Domain swapping acts as the predominant method to tune specificity. Components like SD, Dhp-RD pair, and DBD can be interchangeably utilized to facilitate orthogonal signal transduction. **(B) aTF-based Biosensors:** Both directed evolution and domain-swapping techniques serve as tools to refine ligand specificity, either by eliminating undesirable molecular interactions or adapting to non-native ligands. **(C) Enzymatic biosensors:** The sensor output could be either the product or a byproduct, and directed evolution emerges as the most effective approach to enhance substrate specificity. **(D) FP-based biosensors:** Mutagenesis within LBDs proves to be the most effective strategy for modifying ligand specificity. Performance can be enhanced by modulating both the linker and cpFP insertion site. Here, GPCR is presented as an example of ligand-binding, and orthogonal signal transduction can be achieved via chimeric Gα through domain swapping. **(E) Riboswitch-based Biosensors:** These biosensors integrate a ligand-binding aptamer with a genetic transducer and a reporter. SELEX remains the most common method to enhance aptamer specificity.
Key: TCS, two-component system; SD, sensor domain; DHp, dimerization and histidine phosphotransferase; CA, catalytic and ATP-binding; RD, receiver domain; DBD, DNA-binding domain; LBD, ligand-binding domain; PBD, protein-binding domain; aTF, allosteric transcription factor; TFBS, transcription factor binding site; FP, fluorescent protein; GPCR, G protein-coupled receptor; Gα, G protein α subunit; TM, transmembrane helices; ICL3, intracellular loop 3; cpFP, circularly permuted fluorescent protein; NIR-FP, near-infrared fluorescent protein; FRET, fluorescence resonance energy transfer; SELEX, systematic evolution of ligands by exponential enrichment.

Figure 2. The Design-Build-Test-Learn Workflow of Directed Evolution. Five genetically encoded biosensors undergo similar design-build-test-learn cycles, involving diversification, expression, and selection processes. Specific regions of the genetic elements can be intelligently chosen for subsequent mutagenesis. Techniques such as random mutagenesis, domain swapping, structure-guided site-directed saturation mutagenesis, or computation-driven focused mutation can be introduced to the selected DNA region. The expression and screening process significantly determine the throughput and robustness of the directed evolution process. Variants demonstrating specific ligand-binding will be selected and amplified for further verification or iterative rounds of selection.
Key: aTF, allosteric transcription factor; DBD, DNA-binding domain; LBD, ligand-binding domain; TCS, two-component system; SD, sensor domain; DHp, dimerization and histidine phosphotransferase; CA, catalytic and ATP-binding; FP(LBP), fluorescent protein fused with ligand-binding protein; PBD, protein-binding domain; cpFP, circularly permuted fluorescent proteins.

Figure 3. High-Throughput Screening Techniques. Following DNA diversification methods such as random mutagenesis, site-directed saturation mutagenesis, or computer-assisted focused mutation, DNA variants are tested in vitro or introduced into microbial hosts, including *E. coli* (supporting up to 10^7 variants), yeast (up to 10^6), or phages (up to 10^{12}). The screening process is dependent on the reporter gene regulated by the biosensor. When the reporter is a fluorescence protein, FACS-based selection can yield specific variants by iteratively sorting cells with desired fluorescence levels in the presence or absence of target ligands. Growth-based dual-selection can be utilized when the reporter is growth-related, such as an antibiotic resistance protein, toxic protein, or enzyme critical to survival. When the reporter is a DNA polymerase, DNA enrichment methodologies like CPR come into play. Active partner variants will trigger the expression of DNA polymerase, allowing the active partner to be amplified through PCR reactions. DMS can be coupled with positive selection via FACS or growth-based selection, followed by NGS to determine the mutational fitness landscape under varying selection pressures. Surface display and SELEX processes can also be paired with FACS or DMS to specifically enrich positive hits with superior ligand affinity. Positive hits are isolated from the screening or selection process for verification and further evolution.
Key: SSM, site-directed saturation mutagenesis; FACS, fluorescence-activated cell sorting; CPR, compartmentalized partnered replication; DMS, deep mutational scanning; NGS, next-generation sequencing; SELEX, systematic evolution of ligands by exponential enrichment.

Figure 4. Holistic View of Computer-Aided Biosensor Design. (A) Diagram of structure-based computational protein design. The potential conformational space for a protein-ligand complex could exceed 10^{50} after a series of design processes. Computational algorithms can constrain the mutational space through ligand docking ranking, free energy calculation, or molecular dynamics simulations, and the top-ranked variants are subject to experimental validation for enhanced specificity. (B) Representation of datasets of varying sizes derived from different mutagenesis libraries and screening methods for use in ligand-specificity prediction models. These models can utilize datasets ranging from hundreds to hundreds of thousands of data points to predict ligand specificity. Predictions are based on scoring protocols for ligand-receptor pairs using criteria such as ligand-binding affinity, fitness scores, or evolutionary landscape extracted from DMS data.
Key: SSM, site-directed saturation mutagenesis; LBD, ligand-binding domain, DMS, deep mutational scanning.

Table 1. Overview of Natural Biosensing Elements Discussed in This Review.
Key: GPCR, G protein-coupled receptor; TCS, two-component system; PBP, periplasmic substrate-binding protein; aTF, allosteric transcription factor; LBD, ligand-binding domain; PBD, protein-binding domain; FP, fluorescent protein; SD, sensor domain; DHp, dimerization and histidine phosphotransferase; CA, catalytic and ATP-binding; RD, receiver domain; DBD, DNA-binding domain; ABC transporter, ATP-binding cassette transporter; TFBS, transcription factor binding site; +, low; ++, medium; +++, high; -, not shown or almost none.

Signals	Allosteric components	Small molecule substrates	Sensor	Transducer	Output	Inherent specificity	Evolvability
Extracellular	**GPCR**	Chemical messengers	LBD	PBD → G protein → Second messenger	Transcription / Conformational activation of FP	++	+
	TCS	Growth factors and stressors	SD (PAS)	CA → Dhp → RR	Transcription	++	-
Periplasmic	**PBP**	Metabolites	LBD	Hinge → ABC transporter	Conformational activation of FP	-	++
	Enzyme	Metabolites	LBD	Substrate + Co-factor → Product + Byproduct	Redox, chromophore, luminescence	-	+++
Cytoplasmic	**aTF**	Metabolites	LBD	DBD → TFBS	Transcription / Conformational activation of FP	+	++
	Riboswitch	Metabolites	Aptamer	Ribozyme, cis-regulatory elements, small RNAs	Transcription, translation, fluorescence	+	+++

Table 2. Summary of Rational Design and Directed Evolution Workflow for Ligand-specific Biosensor Development. This table outlines the design-build-test-learn cycle used in the cited references. **Mutation hotspots & *in silico* analysis:** LBD, ligand-binding domain; LBS, ligand-binding sites; PBD, protein-binding domain; CDS, coding sequence; MSA, multiple sequence alignment; ML, machine learning. **Mutagenesis methods:** CAD, computer-aided design; Single SSM, single site-directed saturation mutagenesis; Pairwise SSM, pairwise multiple site-directed saturation mutagenesis; Random, Random mutagenesis. **Screening methods:** MS, mutational scanning; NS, negative selection; PS, positive selection; DS, dual selection; Deep(F), fluorescence-based deep mutational scanning; Deep(G), Growth-based deep mutational scanning; A, alanine scanning; FYW, phenylalanine, tyrosine, tryptophan scanning; ITC: isothermal titration calorimetry. **Mutations & effects:** #, number of mutations; Selective, promiscuous but diminished binding to undesired ligands; SE, steric effect; EI, electrostatic interaction; PI, polar interactions; HP, hydrophobic interaction; LE, long-range effect.

Citation	DESIGN				BUILD			TEST				LEARN					USE
	Input				Diversification			Screening				Output					
	Type	Candidate	Ligand	*In silico*	Hotspot	Mutagenesis	Library size	FACS	Growth	MS	Individual	#	Specificity	Structure-based	Data-driven	Mutation effect	Application
Z. Li et al., 2021	FP (Enzyme)	MurG	11 UDP derivatives	Crystal; MSA	LBS	Random → pairwise SSM → random	10^3				Fluorescence	1~3	UDP-GlcNAc insensitive	-	MSA	SE	Metabolism monitoring
Zhang et al., 2018	FP (PBP)	Atu2422	19 amino acids	Crystal; FoldX; Docking	LBS, linker	CAD → pairwise SSM	10^3				ITC	3	Glycine selective	Crystallography, molecular docking	-	SE	*In situ* glycine sensing
Borden et al. 2020	FP (PBP)	OpuBC	11 choline analogs and NTs	Crystal	LBD	Pairwise SSM	$>10^{10}$				Fluorescence	21	Acetylcholine selective	Crystallography	-	SE, HP	*In situ* acetylcholine sensing
Muthusamy et al., 2022	FP (PBP)	OpuBC variant	6 opioid analogs	Crystal	LBS	MS → single SSM	10^2			FYW	Fluorescence	3	S-methadone selective	Molecular docking	-	SE, HP	*In situ* drug monitoring

Unger et al., 2020	FP (PBP)	OpuBC variant	42 NTs and small molecules	Rosetta ; ML	LBS	CAD → single, pairwise SSM → ML	10^5			Fluorescence	19	Serotonin	Crystallography	ML	-	*In situ* serotonin sensing
Herud-Sikimić et al., 2021	FP (TF)	TrpR	23 indole derivatives	Docking	LBS	Single SSM → random	10^3			Fluorescence	5	Indole-3-acetic acid selective	Crystallography, molecular docking		SE, PI, HP	*In situ* auxin sensing
Looger et al., 2003	TCS (Chimeric)	5 PBPs	trinitrotoluene, lactate or serotonin analogs	Docking; Energy	LBS	CAD	10^{20}			Fluorescence	5~17	Trinitrotoluene, lactate, serotonin	Molecular docking	-	SE, PI, HP	Pollutant/Metabolite detection
d'Oelsnitz et al., 2022	TF	6 regulators	5 benzylisoquinoline alkaloids (BIAs)		LBS → CDS	Pairwise SSM → random	10^5	PS	NS		9~13	Five BIAs	Crystallography		SE, EI, HP	Enzyme evolution
Tang and Cirino, 2011	TF	AraC	4 similar compounds		LBS	Pairwise SSM	10^6	DS			4	Mevalonate	-	-	-	Heterologous pathway engineering
Tang et al., 2008	TF	AraC	5 sugars		LBS → LBD	Pairwise SSM → random	10^7	DS			4	D-arabinose	-	-	-	Inducible promoter
Wu et al., 2022	TF	BmoR	5 alcohols	Docking	LBD	Random → single SSM	10^3			Fluorescence	1~2	Ethanol insensitive	Molecular modeling and docking	-	PI, HP	Biosynthetic pathway engineering
Chockalingam et al., 2005	TF	Estrogen receptor	2 estradiol analogs	Docking	LBS	Single SSM → random	10^6		PS		7	Altered specificity	Molecular modeling and docking		SE, HP	Inducible promoter
Gallinari et al., 2005	TF	Estrogen receptor	11 estradiol analogs	Docking	LBS	Pairwise SSM	10^5			β-Galactosidase	1~5	Estradiol insensitive	Molecular docking	-	SE, HP	Inducible promoter
Li et al., 2017.	TF	HucR	6 aromatic acids		LBS	Pairwise SSM	10^5	DS			4	Shikimic acid	-	-	-	Biosynthetic pathway engineering

Taylor et al., 2016	TF	LacI	5 sugars	Rosetta	CDS	CAD, single SSM, random → shuffle	10^4	PS	NS	Deep(F)		3~5	Sucralose, gentiobiose	Crystallography, molecular docking	MSA	PI, EI, HP	Inducible promoter
Collins et al., 2006	TF	LuxR	6 acyl-homoserine lactones (HSLs)		LBD	Random	10^4		DS			2	3OCnHSL insensitive	-	MSA	EI	Engineered cell-cell communication
Ray et al., 2017	TF	MopR	6 phenol derivatives	Crystal; Docking; MS	LBS	CAD	10^2			A	ITC	2	Catechol, phenol	Crystallography, molecular docking	-	PI, HP	Metabolite detection
Y. Li et al., 2021	TF	MphR	2 macrolides	Crystal	LBS → CDS	Single SSM → random	10^3	NS	PS			4	Clarithromycin	Crystallography, molecular docking	Epistasis	SE, PI	Biosynthetic pathway engineering
Kasey et al., 2018	TF	MphR	6 macrolides	Crystal	LBS → CDS	Single SSM → pairwise SSM → random	10^3	NS	PS			4	Erythromycin selective	Crystallography, molecular docking	Epistasis	LE	Biosynthetic pathway engineering
F. M. Machado et al., 2019	TF	PcaV	9 aromatic compounds	Docking	LBS	Pairwise SSM	10^4	DS				3	Altered specificity	Molecular modeling and docking	-	PI	Metabolite detection
Jha et al., 2016	TF	PobR	2 aromatic acids	Crystal; Rosetta	LBS	Pairwise SSM	10^7	DS				8	Altered specificity	Crystallography	MSA	EI, PI, HP	Metabolite detection
Schwimmer et al., 2004	TF	RXR	2 retinoid-like compounds	Crystal	LBD	Pairwise SSM	10^4		PS			3~5	Altered specificity	-	-	SE, HP	Inducible promoter
Scholz et al., 2003	TF	TetR	5 tetracycline (tc) analogs		LBD	Random → pairwise SSM	10^4	Blue-white				1~5	4-De(dimethylamino)-6-deoxy-6-demethyl-tc	Crystallography	-	SE, PI	Inducible promoter
Henssler et al., 2004	TF	TetR	4 tc analogs		LBD	Single SSM	10^2	Blue-white				1	4-De(dimethylamino)anhydro-tc	Crystallography	-	SE, PI	Inducible promoter

Nishikawa et al., 2021.	TF	TtgR	naringenin, resveratrol	Rosetta	LBS	CAD	10^4	DS			4	Altered specificity	Crystallography	Epistasis	SE, PI, HP	Drug monitoring
D'Ambrosio et al., 2020	TF	VanR	vanillin and vanillic acid		LBD → LBS	Random → single SSM	10^4	DS			1~5	Vanillic acid insensitive	Molecular modeling and docking	-	PI, HP, LE	Metabolite detection
Ogawa et al., 2019	TF	XylS	2 aromatic acids		CDS	Random → single SSM	10^3		DS		2	Altered specificity	-	Epistasis	SE, PI	Biosynthetic pathway engineering
Ogawa et al., 2022	TF	XylS	2 aromatic acids		LBD	Single SSM	10^3			Deep(G)	1	Altered specificity	-	DMS	SE, PI	Biosynthetic pathway engineering
Monteiro et al., 2019	TF (Chimeric)	BenR + XylS	7 aromatic acids	MSA; Docking	LBS	Domain swapping	10^2			Fluorescence	-	Benzoic acid, 3-methyl benzoic acid	Molecular modeling and docking	MSA	-	Metabolite detection
De Paepe et al., 2019	TF (Chimeric)	FdeR + NodD1	3 flavonoids	MSA	TFBS	Domain swapping	10^2			Fluorescence	-	Luteolin	-	MSA	-	Metabolite detection
Rondon and Wilson, 2021	TF (Chimeric)	PurR	9 adenine derivatives	MSA	PBD, linker	Domain swapping, single SSM → random	10^6	DS			4~7	Caffeine selective	Homology modeling	MSA	SE, LE	Drug monitoring
Ellefson et al., 2018	TF (Chimeric)	TrpR	3 tryptophan derivatives	Crystal	LBS, PBD, TFBS	Pairwise SSM, domain swapping	10^7		CPR		5	5-Bromotryptophan, 6-bromotryptophan	Crystallography, molecular docking	-	SE, LE	Genetic logic gate
Rottinghaus et al., 2021	TF, Enzyme	TyrR, TynA, FeaR	10 aromatic compounds	Crystal; Docking	LBS	Single SSM	10^2		DS	Fluorescence	1	Four aromatic compounds	Molecular modeling and docking	-	SE, EI, PI, HP	Metabolite/ NT detection

933

Reference

1. De Paepe, B., Peters, G., Coussement, P., Maertens, J., and De Mey, M. (2017). Tailor-made transcriptional biosensors for optimizing microbial cell factories. J. Ind. Microbiol. Biotechnol. *44*, 623–645. 10.1007/s10295-016-1862-3.

2. Greenwald, E.C., Mehta, S., and Zhang, J. (2018). Genetically Encoded Fluorescent Biosensors Illuminate the Spatiotemporal Regulation of Signaling Networks. Chem. Rev. *118*, 11707–11794. 10.1021/acs.chemrev.8b00333.

3. Lazar, J.T., and Tabor, J.J. (2021). Bacterial two-component systems as sensors for synthetic biology applications. Curr. Opin. Syst. Biol. *28*, 100398. 10.1016/j.coisb.2021.100398.

4. Zhu, Z., Song, H., Wang, Y., and Zhang, Y.-H.P. (2022). Protein engineering for electrochemical biosensors. Curr. Opin. Biotechnol. *76*, 102751. 10.1016/j.copbio.2022.102751.

5. Jang, S., Jang, S., Yang, J., Seo, S.W., and Jung, G.Y. (2018). RNA-based dynamic genetic controllers: development strategies and applications. Curr. Opin. Biotechnol. *53*, 1–11. 10.1016/j.copbio.2017.10.005.

6. Zhang, J., Jensen, M.K., and Keasling, J.D. (2015). Development of biosensors and their application in metabolic engineering. Curr. Opin. Chem. Biol. *28*, 1–8. 10.1016/j.cbpa.2015.05.013.

7. Wan, X., Saltepe, B., Yu, L., and Wang, B. (2021). Programming living sensors for environment, health and biomanufacturing. Microb. Biotechnol. *14*, 2334–2342. 10.1111/1751-7915.13820.

8. Nguyen, P.Q., Soenksen, L.R., Donghia, N.M., Angenent-Mari, N.M., de Puig, H., Huang, A., Lee, R., Slomovic, S., Galbersanini, T., Lansberry, G., et al. (2021). Wearable materials with embedded synthetic biology sensors for biomolecule detection. Nat. Biotechnol. 10.1038/s41587-021-00950-3.

9. Rottinghaus, A.G., Amrofell, M.B., and Moon, T.S. (2020). Biosensing in Smart Engineered Probiotics. Biotechnol. J. *15*, 1900319. 10.1002/biot.201900319.

10. Amrofell, M.B., Rottinghaus, A.G., and Moon, T.S. (2020). Engineering microbial diagnostics and therapeutics with smart control. Curr. Opin. Biotechnol. *66*, 11–17. 10.1016/j.copbio.2020.05.006.

11. Eick, G.N., Colucci, J.K., Harms, M.J., Ortlund, E.A., and Thornton, J.W. (2012). Evolution of Minimal Specificity and Promiscuity in Steroid Hormone Receptors. PLOS Genet. *8*, e1003072. 10.1371/journal.pgen.1003072.

12. Kiel, C., Yus, E., and Serrano, L. (2010). Engineering Signal Transduction Pathways. Cell *140*, 33–47. 10.1016/j.cell.2009.12.028.

13. Kylilis, N., Tuza, Z.A., Stan, G.-B., and Polizzi, K.M. (2018). Tools for engineering coordinated system behaviour in synthetic microbial consortia. Nat. Commun. *9*, 2677. 10.1038/s41467-018-05046-2.

14. Collins, C.H., Leadbetter, J.R., and Arnold, F.H. (2006). Dual selection enhances the signaling specificity of a variant of the quorum-sensing transcriptional activator LuxR. Nat. Biotechnol. *24*, 708–712. 10.1038/nbt1209.

15. Moon, T.S., Lou, C., Tamsir, A., Stanton, B.C., and Voigt, C.A. (2012). Genetic programs constructed from layered logic gates in single cells. Nature *491*, 249–253. 10.1038/nature11516.

16. Moon, T.S. (2022). SynMADE: synthetic microbiota across diverse ecosystems. Trends Biotechnol. *40*, 1405–1414. 10.1016/j.tibtech.2022.08.010.

17. Chockalingam, K., and Zhao, H. (2005). Creating new specific ligand–receptor pairs for transgene regulation. Trends Biotechnol. *23*, 333–335. 10.1016/j.tibtech.2005.05.002.

18. Simon, A.J., d'Oelsnitz, S., and Ellington, A.D. (2019). Synthetic evolution. Nat. Biotechnol. *37*, 730–743. 10.1038/s41587-019-0157-4.

19. Packer, M.S., and Liu, D.R. (2015). Methods for the directed evolution of proteins. Nat. Rev. Genet. *16*, 379–394. 10.1038/nrg3927.

20. Harris, J.L., and Craik, C.S. (1998). Engineering enzyme specificity. Curr. Opin. Chem. Biol. *2*, 127–132. 10.1016/S1367-5931(98)80044-6.

21. Koh, J.T. (2002). Engineering Selectivity and Discrimination into Ligand-Receptor Interfaces. Chem. Biol. *9*, 17–23. 10.1016/S1074-5521(02)00087-X.

22. Wells, J.A., Powers, D.B., Bott, R.R., Graycar, T.P., and Estell, D.A. (1987). Designing substrate specificity by protein engineering of electrostatic interactions. Proc. Natl. Acad. Sci. *84*, 1219–1223. 10.1073/pnas.84.5.1219.

23. Bepler, T., and Berger, B. (2021). Learning the protein language: Evolution, structure, and function. Cell Syst. *12*, 654-669.e3. 10.1016/j.cels.2021.05.017.

24. Gao, R., Bouillet, S., and Stock, A.M. (2019). Structural Basis of Response Regulator Function. Annu. Rev. Microbiol. *73*, 175–197. 10.1146/annurev-micro-020518-115931.

25. Krell, T., Lacal, J., Busch, A., Silva-Jiménez, H., Guazzaroni, M.-E., and Ramos, J.L. (2010). Bacterial Sensor Kinases: Diversity in the Recognition of Environmental Signals. Annu. Rev. Microbiol. *64*, 539–559. 10.1146/annurev.micro.112408.134054.

26. Laub, M.T., and Goulian, M. (2007). Specificity in Two-Component Signal Transduction Pathways. Annu. Rev. Genet. *41*, 121–145. 10.1146/annurev.genet.41.042007.170548.

27. Podgornaia, A.I., and Laub, M.T. (2013). Determinants of specificity in two-component signal transduction. Curr. Opin. Microbiol. *16*, 156–162. 10.1016/j.mib.2013.01.004.

28. Skerker, J.M., Perchuk, B.S., Siryaporn, A., Lubin, E.A., Ashenberg, O., Goulian, M., and Laub, M.T. (2008). Rewiring the Specificity of Two-Component Signal Transduction Systems. Cell *133*, 1043–1054. 10.1016/j.cell.2008.04.040.

29. Rabin, R.S., and Stewart, V. (1993). Dual response regulators (NarL and NarP) interact with dual sensors (NarX and NarQ) to control nitrate- and nitrite-regulated gene expression in Escherichia coli K-12. J. Bacteriol. *175*, 3259–3268.

30. Daeffler, K.N., Galley, J.D., Sheth, R.U., Ortiz-Velez, L.C., Bibb, C.O., Shroyer, N.F., Britton, R.A., and Tabor, J.J. (2017). Engineering bacterial thiosulfate and tetrathionate sensors for detecting gut inflammation. Mol. Syst. Biol. *13*, 923. 10.15252/msb.20167416.

31. Schmidl, S.R., Ekness, F., Sofjan, K., Daeffler, K.N.-M., Brink, K.R., Landry, B.P., Gerhardt, K.P., Dyulgyarov, N., Sheth, R.U., and Tabor, J.J. (2019). Rewiring bacterial two-component systems by modular DNA-binding domain swapping. Nat. Chem. Biol. *15*, 690–698. 10.1038/s41589-019-0286-6.

32. Dong, C., Zheng, Y., Long-Iyer, K., Wright, E.C., Li, Y., and Tian, L. (2022). Fluorescence Imaging of Neural Activity, Neurochemical Dynamics, and Drug-Specific Receptor Conformation with Genetically Encoded Sensors. Annu. Rev. Neurosci. *45*, 273–294. 10.1146/annurev-neuro-110520-031137.

33. Shaw, W.M., Yamauchi, H., Mead, J., Gowers, G.-O.F., Bell, D.J., Öling, D., Larsson, N., Wigglesworth, M., Ladds, G., and Ellis, T. (2019). Engineering a Model Cell for Rational Tuning of GPCR Signaling. Cell *177*, 782-796.e27. 10.1016/j.cell.2019.02.023.

34. Jing, M., Zhang, Y., Wang, H., and Li, Y. (2019). G-protein-coupled receptor-based sensors for imaging neurochemicals with high sensitivity and specificity. J. Neurochem. *151*, 279–288. 10.1111/jnc.14855.

35. Powers, A.S., Pham, V., Burger, W.A.C., Thompson, G., Laloudakis, Y., Sexton, P.M., Paul, S.M., Christopoulos, A., Thal, D.M., Felder, C.C., et al. (2023). Structural basis of efficacy-driven ligand selectivity at GPCRs. Nat. Chem. Biol., 1–10. 10.1038/s41589-022-01247-5.

36. Yang, D., Zhou, Q., Labroska, V., Qin, S., Darbalaei, S., Wu, Y., Yuliantie, E., Xie, L., Tao, H., Cheng, J., et al. (2021). G protein-coupled receptors: structure- and function-based drug discovery. Signal Transduct. Target. Ther. *6*, 1–27. 10.1038/s41392-020-00435-w.

37. Feng, J., Zhang, C., Lischinsky, J.E., Jing, M., Zhou, J., Wang, H., Zhang, Y., Dong, A., Wu, Z., Wu, H., et al. (2019). A Genetically Encoded Fluorescent Sensor for Rapid and Specific In Vivo Detection of Norepinephrine. Neuron *102*, 745-761.e8. 10.1016/j.neuron.2019.02.037.

38. Keiser, M.J., Irwin, J.J., and Shoichet, B.K. (2010). The Chemical Basis of Pharmacology. Biochemistry *49*, 10267–10276. 10.1021/bi101540g.

39. 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., et al. (2018). Ultrafast neuronal imaging of dopamine dynamics with designed genetically encoded sensors. Science *360*, eaat4422. 10.1126/science.aat4422.

40. Sun, F., Zhou, J., Dai, B., Qian, T., Zeng, J., Li, X., Zhuo, Y., Zhang, Y., Wang, Y., Qian, C., et al. (2020). Next-generation GRAB sensors for monitoring dopaminergic activity in vivo. Nat. Methods *17*, 1156–1166. 10.1038/s41592-020-00981-9.

41. Sun, F., Zeng, J., Jing, M., Zhou, J., Feng, J., Owen, S.F., Luo, Y., Li, F., Wang, H., Yamaguchi, T., et al. (2018). A Genetically Encoded Fluorescent Sensor Enables Rapid and Specific Detection of Dopamine in Flies, Fish, and Mice. Cell *174*, 481-496.e19. 10.1016/j.cell.2018.06.042.

42. Weis, W.I., and Kobilka, B.K. (2018). The Molecular Basis of G Protein–Coupled Receptor Activation. Annu. Rev. Biochem. *87*, 897–919. 10.1146/annurev-biochem-060614-033910.

43. Zschiedrich, C.P., Keidel, V., and Szurmant, H. (2016). Molecular Mechanisms of Two-Component Signal Transduction. J. Mol. Biol. *428*, 3752–3775. 10.1016/j.jmb.2016.08.003.

44. Laub, M.T. (2010). The Role of Two-Component Signal Transduction Systems in Bacterial Stress Responses. In Bacterial Stress Responses (John Wiley & Sons, Ltd), pp. 45–58. 10.1128/9781555816841.ch4.

45. Capra, E.J., and Laub, M.T. (2012). Evolution of Two-Component Signal Transduction Systems. Annu. Rev. Microbiol. *66*, 325–347. 10.1146/annurev-micro-092611-150039.

46. Dong, A., He, K., Dudok, B., Farrell, J.S., Guan, W., Liput, D.J., Puhl, H.L., Cai, R., Wang, H., Duan, J., et al. (2022). A fluorescent sensor for spatiotemporally resolved imaging of endocannabinoid dynamics in vivo. Nat. Biotechnol. *40*, 787–798. 10.1038/s41587-021-01074-4.

47. Wan, J., Peng, W., Li, X., Qian, T., Song, K., Zeng, J., Deng, F., Hao, S., Feng, J., Zhang, P., et al. (2021). A genetically encoded sensor for measuring serotonin dynamics. Nat. Neurosci. *24*, 746–752. 10.1038/s41593-021-00823-7.

48. Wu, T., Chen, Z., Guo, S., Zhang, C., and Huo, Y.-X. (2022). Engineering Transcription Factor BmoR Mutants for Constructing Multifunctional Alcohol Biosensors. ACS Synth. Biol. 10.1021/acssynbio.1c00549.

49. Koch, M., Pandi, A., Borkowski, O., Batista, A.C., and Faulon, J.-L. (2019). Custom-made transcriptional biosensors for metabolic engineering. Curr. Opin. Biotechnol. *59*, 78–84. 10.1016/j.copbio.2019.02.016.

50. Rogers, J.K., Guzman, C.D., Taylor, N.D., Raman, S., Anderson, K., and Church, G.M. (2015). Synthetic biosensors for precise gene control and real-time monitoring of metabolites. Nucleic Acids Res. *43*, 7648–7660. 10.1093/nar/gkv616.

51. Baumgart, L.A., Lee, J.E., Salamov, A., Dilworth, D.J., Na, H., Mingay, M., Blow, M.J., Zhang, Y., Yoshinaga, Y., Daum, C.G., et al. (2021). Persistence and plasticity in bacterial gene regulation. Nat. Methods *18*, 1499–1505. 10.1038/s41592-021-01312-2.

52. Hanko, E.K.R., Paiva, A.C., Jonczyk, M., Abbott, M., Minton, N.P., and Malys, N. (2020). A genome-wide approach for identification and characterisation of metabolite-inducible systems. Nat. Commun. *11*, 1213. 10.1038/s41467-020-14941-6.

53. Henson, W.R., Campbell, T., DeLorenzo, D.M., Gao, Y., Berla, B., Kim, S.J., Foston, M., Moon, T.S., and Dantas, G. (2018). Multi-omic elucidation of aromatic catabolism in adaptively evolved Rhodococcus opacus. Metab. Eng. *49*, 69–83. 10.1016/j.ymben.2018.06.009.

54. d'Oelsnitz, S., Kim, W., Burkholder, N.T., Javanmardi, K., Thyer, R., Zhang, Y., Alper, H.S., and Ellington, A.D. (2022). Using fungible biosensors to evolve improved alkaloid biosyntheses. Nat. Chem. Biol., 1–9. 10.1038/s41589-022-01072-w.

55. d'Oelsnitz, S., Diaz, D.J., Acosta, D.J., Schechter, M.W., Minus, M.B., Howard, J.R., Do, H., Loy, J., Alper, H., and Ellington, A.D. (2023). Synthetic microbial sensing and biosynthesis of amaryllidaceae alkaloids. 2023.04.05.535710. 10.1101/2023.04.05.535710.

56. Diao, J., Carr, R., and Moon, T.S. (2022). Deciphering the transcriptional regulation of the catabolism of lignin-derived aromatics in Rhodococcus opacus PD630. Commun. Biol. *5*, 1–17. 10.1038/s42003-022-04069-2.

57. Kang, Z., Zhang, M., Gao, K., Zhang, W., Meng, W., Liu, Y., Xiao, D., Guo, S., Ma, C., Gao, C., et al. (2021). An L-2-hydroxyglutarate biosensor based on specific transcriptional regulator LhgR. Nat. Commun. *12*, 3619. 10.1038/s41467-021-23723-7.

58. Trudeau, D.L., and Tawfik, D.S. (2019). Protein engineers turned evolutionists—the quest for the optimal starting point. Curr. Opin. Biotechnol. *60*, 46–52. 10.1016/j.copbio.2018.12.002.

59. Fan, Y., Makar, M., Wang, M.X., and Ai, H. (2017). Monitoring thioredoxin redox with a genetically encoded red fluorescent biosensor. Nat. Chem. Biol. *13*, 1045–1052. 10.1038/nchembio.2417.

60. DeLoache, W.C., Russ, Z.N., Narcross, L., Gonzales, A.M., Martin, V.J.J., and Dueber, J.E. (2015). An enzyme-coupled biosensor enables (S)-reticuline production in yeast from glucose. Nat. Chem. Biol. *11*, 465–471. 10.1038/nchembio.1816.

61. Ma, F., Chung, M.T., Yao, Y., Nidetz, R., Lee, L.M., Liu, A.P., Feng, Y., Kurabayashi, K., and Yang, G.-Y. (2018). Efficient molecular evolution to generate enantioselective enzymes using a dual-channel microfluidic droplet screening platform. Nat. Commun. *9*, 1030. 10.1038/s41467-018-03492-6.

62. Romero, P.A., Tran, T.M., and Abate, A.R. (2015). Dissecting enzyme function with microfluidic-based deep mutational scanning. Proc. Natl. Acad. Sci. *112*, 7159–7164. 10.1073/pnas.1422285112.

63. Meng, D., Liu, M., Su, H., Song, H., Chen, L., Li, Q., Liu, Y., Zhu, Z., Liu, W., Sheng, X., et al. (2023). Coenzyme Engineering of Glucose-6-phosphate Dehydrogenase on a Nicotinamide-Based Biomimic and Its Application as a Glucose Biosensor. ACS Catal., 1983–1998. 10.1021/acscatal.2c04707.

64. Bollella, P., and Katz, E. (2020). Enzyme-Based Biosensors: Tackling Electron Transfer Issues. Sensors *20*, 3517. 10.3390/s20123517.

65. Li, Z., Zheng, Y., Gao, T., Liu, Z., Zhang, J., and Zhou, G. (2018). Fabrication of biosensor based on core–shell and large void structured magnetic mesoporous microspheres immobilized with laccase for dopamine detection. J. Mater. Sci. *53*, 7996–8008. 10.1007/s10853-018-2165-z.

66. Azad, T., Janse van Rensburg, H.J., Morgan, J., Rezaei, R., Crupi, M.J.F., Chen, R., Ghahremani, M., Jamalkhah, M., Forbes, N., Ilkow, C., et al. (2021). Luciferase-Based Biosensors in the Era of the COVID-19 Pandemic. ACS Nanosci. Au *1*, 15–37. 10.1021/acsnanoscienceau.1c00009.

67. Lu, J., Wang, Z., Jiang, Y., Sun, Z., and Luo, W. (2022). Modification of the substrate specificity of leucine dehydrogenase by site-directed mutagenesis based on biocomputing strategies. Syst. Microbiol. Biomanufacturing. 10.1007/s43393-022-00116-5.

68. Murphy, P.M., Bolduc, J.M., Gallaher, J.L., Stoddard, B.L., and Baker, D. (2009). Alteration of enzyme specificity by computational loop remodeling and design. Proc. Natl. Acad. Sci. *106*, 9215–9220. 10.1073/pnas.0811070106.

69. Rix, G., Watkins-Dulaney, E.J., Almhjell, P.J., Boville, C.E., Arnold, F.H., and Liu, C.C. (2020). Scalable continuous evolution for the generation of diverse enzyme variants encompassing promiscuous activities. Nat. Commun. *11*, 5644. 10.1038/s41467-020-19539-6.

70. Voyvodic, P.L., Pandi, A., Koch, M., Conejero, I., Valjent, E., Courtet, P., Renard, E., Faulon, J.-L., and Bonnet, J. (2019). Plug-and-play metabolic transducers expand the chemical detection space of cell-free biosensors. Nat. Commun. *10*, 1697. 10.1038/s41467-019-09722-9.

71. Utsumi, R., Brissette, R.E., Rampersaud, A., Forst, S.A., Oosawa, K., and Inouye, M. (1989). Activation of Bacterial Porin Gene Expression by a Chimeric Signal Transducer in Response to Aspartate. Science *245*, 1246–1249. 10.1126/science.2476847.

72. Wu, Z., Lin, D., and Li, Y. (2022). Pushing the frontiers: tools for monitoring neurotransmitters and neuromodulators. Nat. Rev. Neurosci. *23*, 257–274. 10.1038/s41583-022-00577-6.

73. Jensen, E.D., Deichmann, M., Ma, X., Vilandt, R.U., Schiesaro, G., Rojek, M.B., Lengger, B., Eliasson, L., Vento, J.M., Durmusoglu, D., et al. (2022). Engineered cell differentiation and sexual reproduction in probiotic and mating yeasts. Nat. Commun. *13*, 6201. 10.1038/s41467-022-33961-y.

74. Kapolka, N.J., Taghon, G.J., Rowe, J.B., Morgan, W.M., Enten, J.F., Lambert, N.A., and Isom, D.G. (2020). DCyFIR: a high-throughput CRISPR platform for multiplexed G protein-coupled receptor profiling and ligand discovery. Proc. Natl. Acad. Sci. *117*, 13117–13126. 10.1073/pnas.2000430117.

75. Lengger, B., Hoch-Schneider, E.E., Jensen, C.N., Jakočiūnas, T., Petersen, A.A., Frimurer, T.M., Jensen, E.D., and Jensen, M.K. (2022). Serotonin G Protein-Coupled Receptor-Based Biosensing Modalities in Yeast. ACS Sens. *7*, 1323–1335. 10.1021/acssensors.1c02061.

76. Scott, B.M., Gutiérrez-Vázquez, C., Sanmarco, L.M., da Silva Pereira, J.A., Li, Z., Plasencia, A., Hewson, P., Cox, L.M., O'Brien, M., Chen, S.K., et al. (2021). Self-tunable engineered yeast probiotics for the treatment of inflammatory bowel disease. Nat. Med. *27*, 1212–1222. 10.1038/s41591-021-01390-x.

77. Billerbeck, S., Brisbois, J., Agmon, N., Jimenez, M., Temple, J., Shen, M., Boeke, J.D., and Cornish, V.W. (2018). A scalable peptide-GPCR language for engineering multicellular communication. Nat. Commun. *9*, 5057. 10.1038/s41467-018-07610-2.

78. Juárez, J.F., Lecube-Azpeitia, B., Brown, S.L., Johnston, C.D., and Church, G.M. (2018). Biosensor libraries harness large classes of binding domains for construction of allosteric transcriptional regulators. Nat. Commun. *9*, 3101. 10.1038/s41467-018-05525-6.

79. Rondon, R.E., Groseclose, T.M., Short, A.E., and Wilson, C.J. (2019). Transcriptional programming using engineered systems of transcription factors and genetic architectures. Nat. Commun. *10*, 4784. 10.1038/s41467-019-12706-4.

80. Shis, D.L., Hussain, F., Meinhardt, S., Swint-Kruse, L., and Bennett, M.R. (2014). Modular, Multi-Input Transcriptional Logic Gating with Orthogonal LacI/GalR Family Chimeras. ACS Synth. Biol. *3*, 645–651. 10.1021/sb500262f.

81. Monteiro, L.M.O., Arruda, L.M., Sanches-Medeiros, A., Martins-Santana, L., Alves, L. de F., Defelipe, L., Turjanski, A.G., Guazzaroni, M.-E., de Lorenzo, V., and Silva-Rocha, R. (2019). Reverse Engineering of an Aspirin-Responsive Transcriptional Regulator in *Escherichia coli*. ACS Synth. Biol. *8*, 1890–1900. 10.1021/acssynbio.9b00191.

82. Chang, H.-J., Zúñiga, A., Conejero, I., Voyvodic, P.L., Gracy, J., Fajardo-Ruiz, E., Cohen-Gonsaud, M., Cambray, G., Pageaux, G.-P., Meszaros, M., et al. (2021). Programmable receptors enable bacterial biosensors to detect pathological biomarkers in clinical samples. Nat. Commun. *12*, 5216. 10.1038/s41467-021-25538-y.

83. De Paepe, B., Maertens, J., Vanholme, B., and De Mey, M. (2019). Chimeric LysR-Type Transcriptional Biosensors for Customizing Ligand Specificity Profiles toward Flavonoids. ACS Synth. Biol. *8*, 318–331. 10.1021/acssynbio.8b00326.

84. Rondon, R., and Wilson, C.J. (2021). Engineering Alternate Ligand Recognition in the PurR Topology: A System of Novel Caffeine Biosensing Transcriptional Antirepressors. ACS Synth. Biol. *10*, 552–565. 10.1021/acssynbio.0c00582.

85. Harbaugh, S.V., Silverman, A.D., Chushak, Y.G., Zimlich, K., Wolfe, M., Thavarajah, W., Jewett, M.C., Lucks, J.B., and Chávez, J.L. (2022). Engineering a Synthetic Dopamine-Responsive Riboswitch for *In Vitro* Biosensing. ACS Synth. Biol., acssynbio.1c00560. 10.1021/acssynbio.1c00560.

86. Vezeau, G.E., Gadila, L.R., and Salis, H.M. (2023). Automated design of protein-binding riboswitches for sensing human biomarkers in a cell-free expression system. Nat. Commun. *14*, 2416. 10.1038/s41467-023-38098-0.

87. Hong, F., Ma, D., Wu, K., Mina, L.A., Luiten, R.C., Liu, Y., Yan, H., and Green, A.A. (2020). Precise and Programmable Detection of Mutations Using Ultraspecific Riboregulators. Cell *180*, 1018-1032.e16. 10.1016/j.cell.2020.02.011.

88. Dykstra, P.B., Kaplan, M., and Smolke, C.D. (2022). Engineering synthetic RNA devices for cell control. Nat. Rev. Genet. 10.1038/s41576-021-00436-7.

89. Kavita, K., and Breaker, R.R. (2023). Discovering riboswitches: the past and the future. Trends Biochem. Sci. *48*, 119–141. 10.1016/j.tibs.2022.08.009.

90. Townshend, B., Xiang, J.S., Manzanarez, G., Hayden, E.J., and Smolke, C.D. (2021). A multiplexed, automated evolution pipeline enables scalable discovery and characterization of biosensors. Nat. Commun. *12*, 1437. 10.1038/s41467-021-21716-0.

91. Yokobayashi, Y. (2020). High-Throughput Analysis and Engineering of Ribozymes and Deoxyribozymes by Sequencing. Acc. Chem. Res. *53*, 2903–2912. 10.1021/acs.accounts.0c00546.

92. Su, Y., and Hammond, M.C. (2020). RNA-based fluorescent biosensors for live cell imaging of small molecules and RNAs. Curr. Opin. Biotechnol. *63*, 157–166. 10.1016/j.copbio.2020.01.001.

93. Thavarajah, W., Silverman, A.D., Verosloff, M.S., Kelley-Loughnane, N., Jewett, M.C., and Lucks, J.B. (2020). Point-of-Use Detection of Environmental Fluoride via a Cell-Free Riboswitch-Based Biosensor. ACS Synth. Biol. *9*, 10–18. 10.1021/acssynbio.9b00347.

94. Iwasaki, R.S., Ozdilek, B.A., Garst, A.D., Choudhury, A., and Batey, R.T. (2020). Small molecule regulated sgRNAs enable control of genome editing in E. coli by Cas9. Nat. Commun. *11*, 1394. 10.1038/s41467-020-15226-8.

95. Nakatsuka, N., Yang, K.-A., Abendroth, J.M., Cheung, K.M., Xu, X., Yang, H., Zhao, C., Zhu, B., Rim, Y.S., Yang, Y., et al. (2018). Aptamer–field-effect transistors overcome Debye length limitations for small-molecule sensing. Science *362*, 319–324. 10.1126/science.aao6750.

96. Wang, B., Zhao, C., Wang, Z., Yang, K.-A., Cheng, X., Liu, W., Yu, W., Lin, S., Zhao, Y., Cheung, K.M., et al. (2022). Wearable aptamer-field-effect transistor sensing system for noninvasive cortisol monitoring. Sci. Adv. *8*, eabk0967. 10.1126/sciadv.abk0967.

97. Zhao, C., Cheung, K.M., Huang, I.-W., Yang, H., Nakatsuka, N., Liu, W., Cao, Y., Man, T., Weiss, P.S., Monbouquette, H.G., et al. (2021). Implantable aptamer–field-effect transistor neuroprobes for in vivo neurotransmitter monitoring. Sci. Adv. 10.1126/sciadv.abj7422.

98. Zhou, X., Mehta, S., and Zhang, J. (2020). Genetically Encodable Fluorescent and Bioluminescent Biosensors Light Up Signaling Networks. Trends Biochem. Sci. *45*, 889–905. 10.1016/j.tibs.2020.06.001.

99. Jing, M., Zhang, P., Wang, G., Feng, J., Mesik, L., Zeng, J., Jiang, H., Wang, S., Looby, J.C., Guagliardo, N.A., et al. (2018). A genetically encoded fluorescent acetylcholine indicator for in vitro and in vivo studies. Nat. Biotechnol. *36*, 726–737. 10.1038/nbt.4184.

100. Peng, W., Wu, Z., Song, K., Zhang, S., Li, Y., and Xu, M. (2020). Regulation of sleep homeostasis mediator adenosine by basal forebrain glutamatergic neurons. Science *369*, eabb0556. 10.1126/science.abb0556.

101. Wu, Z., He, K., Chen, Y., Li, H., Pan, S., Li, B., Liu, T., Xi, F., Deng, F., Wang, H., et al. (2022). A sensitive GRAB sensor for detecting extracellular ATP in vitro and in vivo. Neuron *110*, 770-782.e5. 10.1016/j.neuron.2021.11.027.

102. Scheib, U., Shanmugaratnam, S., Farías-Rico, J.A., and Höcker, B. (2014). Change in protein-ligand specificity through binding pocket grafting. J. Struct. Biol. *185*, 186–192. 10.1016/j.jsb.2013.06.002.

103. Muthusamy, A.K., Kim, C.H., Virgil, S.C., Knox, H.J., Marvin, J.S., Nichols, A.L., Cohen, B.N., Dougherty, D.A., Looger, L.L., and Lester, H.A. (2022). Three Mutations Convert the Selectivity of a Protein Sensor from Nicotinic Agonists to S-Methadone for Use in Cells, Organelles, and Biofluids. J. Am. Chem. Soc. *144*, 8480–8486. 10.1021/jacs.2c02323.

104. Unger, E.K., Keller, J.P., Altermatt, M., Liang, R., Matsui, A., Dong, C., Hon, O.J., Yao, Z., Sun, J., Banala, S., et al. (2020). Directed Evolution of a Selective and Sensitive Serotonin Sensor via Machine Learning. Cell *183*, 1986-2002.e26. 10.1016/j.cell.2020.11.040.

105. Borden, P.M., Zhang, P., Shivange, A.V., Marvin, J.S., Cichon, J., Dan, C., Podgorski, K., Figueiredo, A., Novak, O., Tanimoto, M., et al. (2020). A fast genetically encoded fluorescent sensor for faithful in vivo acetylcholine detection in mice, fish, worms and flies. 2020.02.07.939504. 10.1101/2020.02.07.939504.

106. Zhang, W.H., Herde, M.K., Mitchell, J.A., Whitfield, J.H., Wulff, A.B., Vongsouthi, V., Sanchez-Romero, I., Gulakova, P.E., Minge, D., Breithausen, B., et al. (2018). Monitoring hippocampal glycine with the computationally designed optical sensor GlyFS. Nat. Chem. Biol. *14*, 861–869. 10.1038/s41589-018-0108-2.

107. Nadler, D.C., Morgan, S.-A., Flamholz, A., Kortright, K.E., and Savage, D.F. (2016). Rapid construction of metabolite biosensors using domain-insertion profiling. Nat. Commun. *7*, 12266. 10.1038/ncomms12266.

108. 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., et al. (2013). An optimized fluorescent probe for visualizing glutamate neurotransmission. Nat. Methods *10*, 162–170. 10.1038/nmeth.2333.

109. Marvin, J.S., Schreiter, E.R., Echevarría, I.M., and Looger, L.L. (2011). A genetically encoded, high-signal-to-noise maltose sensor. Proteins Struct. Funct. Bioinforma. *79*, 3025–3036. 10.1002/prot.23118.

110. Li, Z., Zhang, J., and Ai, H. (2021). Genetically Encoded Green Fluorescent Biosensors for Monitoring UDP-GlcNAc in Live Cells. ACS Cent. Sci. *7*, 1763–1770. 10.1021/acscentsci.1c00745.

111. Nasu, Y., Shen, Y., Kramer, L., and Campbell, R.E. (2021). Structure- and mechanism-guided design of single fluorescent protein-based biosensors. Nat. Chem. Biol. 10.1038/s41589-020-00718-x.

112. Beltrán, J., Steiner, P.J., Bedewitz, M., Wei, S., Peterson, F.C., Li, Z., Hughes, B.E., Hartley, Z., Robertson, N.R., Medina-Cucurella, A.V., et al. (2022). Rapid biosensor development using plant hormone receptors as reprogrammable scaffolds. Nat. Biotechnol. 10.1038/s41587-022-01364-5.

113. Herud-Sikimić, O., Stiel, A.C., Kolb, M., Shanmugaratnam, S., Berendzen, K.W., Feldhaus, C., Höcker, B., and Jürgens, G. (2021). A biosensor for the direct visualization of auxin. Nature *592*, 768–772. 10.1038/s41586-021-03425-2.

114. Zhao, Y., Shen, Y., Wen, Y., and Campbell, R.E. (2020). High-Performance Intensiometric Direct- and Inverse-Response Genetically Encoded Biosensors for Citrate. ACS Cent. Sci. *6*, 1441–1450. 10.1021/acscentsci.0c00518.

115. Meyer, A.J., Segall-Shapiro, T.H., Glassey, E., Zhang, J., and Voigt, C.A. (2019). Escherichia coli “Marionette” strains with 12 highly optimized small-molecule sensors. Nat. Chem. Biol. *15*, 196–204. 10.1038/s41589-018-0168-3.

116. Snoek, T., Chaberski, E.K., Ambri, F., Kol, S., Bjørn, S.P., Pang, B., Barajas, J.F., Welner, D.H., Jensen, M.K., and Keasling, J.D. (2020). Evolution-guided engineering of small-molecule biosensors. Nucleic Acids Res. *48*, e3. 10.1093/nar/gkz954.

117. Pham, C., Stogios, P.J., Savchenko, A., and Mahadevan, R. (2022). Advances in engineering and optimization of transcription factor-based biosensors for plug-and-play small molecule detection. Curr. Opin. Biotechnol. *76*, 102753. 10.1016/j.copbio.2022.102753.

118. Ray, S., Panjikar, S., and Anand, R. (2017). Structure Guided Design of Protein Biosensors for Phenolic Pollutants. ACS Sens. *2*, 411–418. 10.1021/acssensors.6b00843.

119. Galvão, T.C., and de Lorenzo, V. (2006). Transcriptional regulators à la carte: engineering new effector specificities in bacterial regulatory proteins. Curr. Opin. Biotechnol. *17*, 34–42. 10.1016/j.copbio.2005.12.002.

120. Libis, V., Delépine, B., and Faulon, J.-L. (2016). Sensing new chemicals with bacterial transcription factors. Curr. Opin. Microbiol. *33*, 105–112. 10.1016/j.mib.2016.07.006.

121. Busch, A., Lacal, J., Martos, A., Ramos, J.L., and Krell, T. (2007). Bacterial sensor kinase TodS interacts with agonistic and antagonistic signals. Proc. Natl. Acad. Sci. *104*, 13774–13779. 10.1073/pnas.0701547104.

122. Landry, B.P., Palanki, R., Dyulgyarov, N., Hartsough, L.A., and Tabor, J.J. (2018). Phosphatase activity tunes two-component system sensor detection threshold. Nat. Commun. *9*, 1433. 10.1038/s41467-018-03929-y.

123. Machado, L.F.M., and Dixon, N. (2016). Development and substrate specificity screening of an in vivo biosensor for the detection of biomass derived aromatic chemical building blocks. Chem. Commun. *52*, 11402–11405. 10.1039/C6CC04559F.

124. Henssler, E.-M., Scholz, O., Lochner, S., Gmeiner, P., and Hillen, W. (2004). Structure-Based Design of Tet Repressor To Optimize a New Inducer Specificity. Biochemistry *43*, 9512–9518. 10.1021/bi049682j.

125. Henßler, E.-M., Bertram, R., Wisshak, S., and Hillen, W. (2005). Tet repressor mutants with altered effector binding and allostery. FEBS J. *272*, 4487–4496. 10.1111/j.1742-4658.2005.04868.x.

126. Scholz, O., Köstner, M., Reich, M., Gastiger, S., and Hillen, W. (2003). Teaching TetR to Recognize a New Inducer. J. Mol. Biol. *329*, 217–227. 10.1016/S0022-2836(03)00427-3.

127. Tang, S.-Y., Fazelinia, H., and Cirino, P.C. (2008). AraC Regulatory Protein Mutants with Altered Effector Specificity. J. Am. Chem. Soc. *130*, 5267–5271. 10.1021/ja7109053.

128. Tang, S.-Y., and Cirino, P.C. (2011). Design and Application of a Mevalonate-Responsive Regulatory Protein. Angew. Chem. Int. Ed. *50*, 1084–1086. 10.1002/anie.201006083.

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

130. Gallinari, P., Lahm, A., Koch, U., Paolini, C., Nardi, M.C., Roscilli, G., Kinzel, O., Fattori, D., Muraglia, E., Toniatti, C., et al. (2005). A Functionally Orthogonal Estrogen Receptor-Based Transcription Switch Specifically Induced by a Nonsteroid Synthetic Ligand. Chem. Biol. *12*, 883–893. 10.1016/j.chembiol.2005.05.018.

131. Schwimmer, L.J., Rohatgi, P., Azizi, B., Seley, K.L., and Doyle, D.F. (2004). Creation and discovery of ligand–receptor pairs for transcriptional control with small molecules. Proc. Natl. Acad. Sci. *101*, 14707–14712. 10.1073/pnas.0400884101.

132. Chockalingam, K., Chen, Z., Katzenellenbogen, J.A., and Zhao, H. (2005). Directed evolution of specific receptor–ligand pairs for use in the creation of gene switches. Proc. Natl. Acad. Sci. *102*, 5691–5696. 10.1073/pnas.0409206102.

133. Rottinghaus, A.G., Xi, C., Amrofell, M.B., Yi, H., and Moon, T.S. (2021). Engineering ligand-specific biosensors for aromatic amino acids and neurochemicals. Cell Syst. *0*. 10.1016/j.cels.2021.10.006.

134. Sarkar, C.A., Dodevski, I., Kenig, M., Dudli, S., Mohr, A., Hermans, E., and Plückthun, A. (2008). Directed evolution of a G protein-coupled receptor for expression, stability, and binding selectivity. Proc. Natl. Acad. Sci. *105*, 14808–14813. 10.1073/pnas.0803103105.

135. Chen, K.-Y.M., Keri, D., and Barth, P. (2020). Computational design of G Protein-Coupled Receptor allosteric signal transductions. Nat. Chem. Biol. *16*, 77–86. 10.1038/s41589-019-0407-2.

136. Glasgow, A., Hobbs, H.T., Perry, Z.R., Wells, M.L., Marqusee, S., and Kortemme, T. (2023). Ligand-specific changes in conformational flexibility mediate long-range allostery in the lac repressor. Nat. Commun. *14*, 1179. 10.1038/s41467-023-36798-1.

137. Wrenbeck, E.E., Azouz, L.R., and Whitehead, T.A. (2017). Single-mutation fitness landscapes for an enzyme on multiple substrates reveal specificity is globally encoded. Nat. Commun. *8*, 15695. 10.1038/ncomms15695.

138. Sun, D., Sun, M., Zhang, J., Lin, X., Zhang, Y., Lin, F., Zhang, P., Yang, C., and Song, J. (2022). Computational tools for aptamer identification and optimization. TrAC Trends Anal. Chem. *157*, 116767. 10.1016/j.trac.2022.116767.

139. Mena, M.A., and Daugherty, P.S. (2005). Automated design of degenerate codon libraries. Protein Eng. Des. Sel. *18*, 559–561. 10.1093/protein/gzi061.

140. Tang, S.-Y., Qian, S., Akinterinwa, O., Frei, C.S., Gredell, J.A., and Cirino, P.C. (2013). Screening for Enhanced Triacetic Acid Lactone Production by Recombinant Escherichia coli Expressing a Designed Triacetic Acid Lactone Reporter. J. Am. Chem. Soc. *135*, 10099–10103. 10.1021/ja402654z.

141. Kasey, C.M., Zerrad, M., Li, Y., Cropp, T.A., and Williams, G.J. (2018). Development of Transcription Factor-Based Designer Macrolide Biosensors for Metabolic Engineering and Synthetic Biology. ACS Synth. Biol. *7*, 227–239. 10.1021/acssynbio.7b00287.

142. Li, Y., Reed, M., Wright, H.T., Cropp, T.A., and Williams, G.J. (2021). Development of Genetically Encoded Biosensors for Reporting the Methyltransferase-Dependent Biosynthesis of Semisynthetic Macrolide Antibiotics. ACS Synth. Biol. *10*, 2520–2531. 10.1021/acssynbio.1c00151.

143. Li, H., Liang, C., Chen, W., Jin, J.-M., Tang, S.-Y., and Tao, Y. (2017). Monitoring in vivo metabolic flux with a designed whole-cell metabolite biosensor of shikimic acid. Biosens. Bioelectron. *98*, 457–465. 10.1016/j.bios.2017.07.022.

144. Jha, R.K., Kern, T.L., Kim, Y., Tesar, C., Jedrzejczak, R., Joachimiak, A., and Strauss, C.E.M. (2016). A microbial sensor for organophosphate hydrolysis exploiting an engineered specificity switch in a transcription factor. Nucleic Acids Res. *44*, 8490–8500. 10.1093/nar/gkw687.

145. F. M. Machado, L., Currin, A., and Dixon, N. (2019). Directed evolution of the PcaV allosteric transcription factor to generate a biosensor for aromatic aldehydes. J. Biol. Eng. *13*, 91. 10.1186/s13036-019-0214-z.

146. D'Ambrosio, V., Pramanik, S., Goroncy, K., Jakočiūnas, T., Schönauer, D., Davari, M.D., Schwaneberg, U., Keasling, J.D., and Jensen, M.K. (2020). Directed evolution of VanR biosensor specificity in yeast. Biotechnol. Notes *1*, 9–15. 10.1016/j.biotno.2020.01.002.

147. Nishikawa, K.K., Hoppe, N., Smith, R., Bingman, C., and Raman, S. (2021). Epistasis shapes the fitness landscape of an allosteric specificity switch. Nat. Commun. *12*, 5562. 10.1038/s41467-021-25826-7.

148. Townshend, B., Kennedy, A.B., Xiang, J.S., and Smolke, C.D. (2015). High-throughput cellular RNA device engineering. Nat. Methods *12*, 989–994. 10.1038/nmeth.3486.

149. Mayer, G., Ahmed, M.-S.L., Dolf, A., Endl, E., Knolle, P.A., and Famulok, M. (2010). Fluorescence-activated cell sorting for aptamer SELEX with cell mixtures. Nat. Protoc. *5*, 1993–2004. 10.1038/nprot.2010.163.

150. Naftaly, S., Cohen, I., Shahar, A., Hockla, A., Radisky, E.S., and Papo, N. (2018). Mapping protein selectivity landscapes using multi-target selective screening and next-generation sequencing of combinatorial libraries. Nat. Commun. *9*, 3935. 10.1038/s41467-018-06403-x.

151. Pethe, M.A., Rubenstein, A.B., and Khare, S.D. (2019). Data-driven supervised learning of a viral protease specificity landscape from deep sequencing and molecular simulations. Proc. Natl. Acad. Sci. *116*, 168–176. 10.1073/pnas.1805256116.

152. Javanpour, A.A., and Liu, C.C. (2021). Evolving Small-Molecule Biosensors with Improved Performance and Reprogrammed Ligand Preference Using OrthoRep. ACS Synth. Biol. *10*, 2705–2714. 10.1021/acssynbio.1c00316.

153. Wu, S., Xiang, C., Zhou, Y., Khan, M.S.H., Liu, W., Feiler, C.G., Wei, R., Weber, G., Höhne, M., and Bornscheuer, U.T. (2022). A growth selection system for the directed evolution of amine-forming or converting enzymes. Nat. Commun. *13*, 7458. 10.1038/s41467-022-35228-y.

154. Ogawa, Y., Katsuyama, Y., Ueno, K., and Ohnishi, Y. (2019). Switching the Ligand Specificity of the Biosensor XylS from meta to para-Toluic Acid through Directed Evolution Exploiting a Dual Selection System. ACS Synth. Biol. *8*, 2679–2689. 10.1021/acssynbio.9b00237.

155. Ellefson, J.W., Meyer, A.J., Hughes, R.A., Cannon, J.R., Brodbelt, J.S., and Ellington, A.D. (2014). Directed evolution of genetic parts and circuits by compartmentalized partnered replication. Nat. Biotechnol. *32*, 97–101. 10.1038/nbt.2714.

156. Ellefson, J.W., Ledbetter, M.P., and Ellington, A.D. (2018). Directed evolution of a synthetic phylogeny of programmable Trp repressors. Nat. Chem. Biol. *14*, 361–367. 10.1038/s41589-018-0006-7.

157. Araya, C.L., and Fowler, D.M. (2011). Deep mutational scanning: assessing protein function on a massive scale. Trends Biotechnol. *29*, 435–442. 10.1016/j.tibtech.2011.04.003.

158. Wrenbeck, E.E., Faber, M.S., and Whitehead, T.A. (2017). Deep sequencing methods for protein engineering and design. Curr. Opin. Struct. Biol. *45*, 36–44. 10.1016/j.sbi.2016.11.001.

159. Jolma, A., Kivioja, T., Toivonen, J., Cheng, L., Wei, G., Enge, M., Taipale, M., Vaquerizas, J.M., Yan, J., Sillanpää, M.J., et al. (2010). Multiplexed massively parallel SELEX for characterization of human transcription factor binding specificities. Genome Res. *20*, 861–873. 10.1101/gr.100552.109.

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

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

162. Schlinkmann, K.M., Honegger, A., Türeci, E., Robison, K.E., Lipovšek, D., and Plückthun, A. (2012). Critical features for biosynthesis, stability, and functionality of a G protein-coupled receptor uncovered by all-versus-all mutations. Proc. Natl. Acad. Sci. *109*, 9810–9815. 10.1073/pnas.1202107109.

163. Jacquier, H., Birgy, A., Le Nagard, H., Mechulam, Y., Schmitt, E., Glodt, J., Bercot, B., Petit, E., Poulain, J., Barnaud, G., et al. (2013). Capturing the mutational landscape of the beta-lactamase TEM-1. Proc. Natl. Acad. Sci. *110*, 13067–13072. 10.1073/pnas.1215206110.

164. Hopf, T.A., Ingraham, J.B., Poelwijk, F.J., Schärfe, C.P.I., Springer, M., Sander, C., and Marks, D.S. (2017). Mutation effects predicted from sequence co-variation. Nat. Biotechnol. *35*, 128–135. 10.1038/nbt.3769.

165. Gray, V.E., Hause, R.J., Luebeck, J., Shendure, J., and Fowler, D.M. (2018). Quantitative Missense Variant Effect Prediction Using Large-Scale Mutagenesis Data. Cell Syst. *6*, 116-124.e3. 10.1016/j.cels.2017.11.003.

166. Pokusaeva, V.O., Usmanova, D.R., Putintseva, E.V., Espinar, L., Sarkisyan, K.S., Mishin, A.S., Bogatyreva, N.S., Ivankov, D.N., Akopyan, A.V., Avvakumov, S.Y., et al. (2019). An experimental assay of the interactions of amino acids from orthologous sequences shaping a complex fitness landscape. PLOS Genet. *15*, e1008079. 10.1371/journal.pgen.1008079.

167. van der Meer, J.-Y., Poddar, H., Baas, B.-J., Miao, Y., Rahimi, M., Kunzendorf, A., van Merkerk, R., Tepper, P.G., Geertsema, E.M., Thunnissen, A.-M.W.H., et al. (2016). Using mutability landscapes of a promiscuous tautomerase to guide the engineering of enantioselective Michaelases. Nat. Commun. *7*, 10911. 10.1038/ncomms10911.

168. Diss, G., and Lehner, B. (2018). The genetic landscape of a physical interaction. eLife *7*, e32472. 10.7554/eLife.32472.

169. Fowler, D.M., Araya, C.L., Fleishman, S.J., Kellogg, E.H., Stephany, J.J., Baker, D., and Fields, S. (2010). High-resolution mapping of protein sequence-function relationships. Nat. Methods *7*, 741–746. 10.1038/nmeth.1492.

170. Melnikov, A., Rogov, P., Wang, L., Gnirke, A., and Mikkelsen, T.S. (2014). Comprehensive mutational scanning of a kinase in vivo reveals substrate-dependent fitness landscapes. Nucleic Acids Res. *42*, e112. 10.1093/nar/gku511.

171. Ogawa, Y., Katsuyama, Y., and Ohnishi, Y. (2022). Engineering of the Ligand Specificity of Transcriptional Regulator XylS by Deep Mutational Scanning. ACS Synth. Biol. *11*, 473–485. 10.1021/acssynbio.1c00564.

172. Meier, G., Thavarasah, S., Ehrenbolger, K., Hutter, C.A.J., Hürlimann, L.M., Barandun, J., and Seeger, M.A. (2022). Deep mutational scan of a drug efflux pump reveals its structure–function landscape. Nat. Chem. Biol., 1–11. 10.1038/s41589-022-01205-1.

173. Wu, Z., Kan, S.B.J., Lewis, R.D., Wittmann, B.J., and Arnold, F.H. (2019). Machine learning-assisted directed protein evolution with combinatorial libraries. Proc. Natl. Acad. Sci. *116*, 8852–8858. 10.1073/pnas.1901979116.

174. Havranek, J.J. (2010). Specificity in Computational Protein Design*. J. Biol. Chem. *285*, 31095–31099. 10.1074/jbc.R110.157685.

175. Lovelock, S.L., Crawshaw, R., Basler, S., Levy, C., Baker, D., Hilvert, D., and Green, A.P. (2022). The road to fully programmable protein catalysis. Nature *606*, 49–58. 10.1038/s41586-022-04456-z.

176. Wang, D.D., Zhu, M., and Yan, H. (2021). Computationally predicting binding affinity in protein–ligand complexes: free energy-based simulations and machine learning-based scoring functions. Brief. Bioinform. *22*, bbaa107. 10.1093/bib/bbaa107.

177. Leman, J.K., Weitzner, B.D., Lewis, S.M., Adolf-Bryfogle, J., Alam, N., Alford, R.F., Aprahamian, M., Baker, D., Barlow, K.A., Barth, P., et al. (2020). Macromolecular modeling and design in Rosetta: recent methods and frameworks. Nat. Methods *17*, 665–680. 10.1038/s41592-020-0848-2.

178. Pertusi, D.A., Moura, M.E., Jeffryes, J.G., Prabhu, S., Walters Biggs, B., and Tyo, K.E.J. (2017). Predicting novel substrates for enzymes with minimal experimental effort with active learning. Metab. Eng. *44*, 171–181. 10.1016/j.ymben.2017.09.016.

179. Voigt, C.A., Mayo, S.L., Arnold, F.H., and Wang, Z.-G. (2001). Computational method to reduce the search space for directed protein evolution. Proc. Natl. Acad. Sci. *98*, 3778–3783. 10.1073/pnas.051614498.

180. Dou, J., Vorobieva, A.A., Sheffler, W., Doyle, L.A., Park, H., Bick, M.J., Mao, B., Foight, G.W., Lee, M.Y., Gagnon, L.A., et al. (2018). De novo design of a fluorescence-activating β-barrel. Nature *561*, 485–491. 10.1038/s41586-018-0509-0.

181. Looger, L.L., Dwyer, M.A., Smith, J.J., and Hellinga, H.W. (2003). Computational design of receptor and sensor proteins with novel functions. Nature *423*, 185–190. 10.1038/nature01556.

182. Lippow, S.M., Moon, T.S., Basu, S., Yoon, S.-H., Li, X., Chapman, B.A., Robison, K., Lipovšek, D., and Prather, K.L.J. (2010). Engineering Enzyme Specificity Using Computational Design of a Defined-Sequence Library. Chem. Biol. *17*, 1306–1315. 10.1016/j.chembiol.2010.10.012.

183. Jha, R.K., Chakraborti, S., Kern, T.L., Fox, D.T., and Strauss, C.E.M. (2015). Rosetta comparative modeling for library design: Engineering alternative inducer specificity in a transcription factor. Proteins Struct. Funct. Bioinforma. *83*, 1327–1340. 10.1002/prot.24828.

184. Lucas, J.E., and Kortemme, T. (2020). New computational protein design methods for de novo small molecule binding sites. PLOS Comput. Biol. *16*, e1008178. 10.1371/journal.pcbi.1008178.

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

186. Glasgow, A.A., Huang, Y.-M., Mandell, D.J., Thompson, M., Ritterson, R., Loshbaugh, A.L., Pellegrino, J., Krivacic, C., Pache, R.A., Barlow, K.A., et al. (2019). Computational design of a modular protein sense-response system. Science *366*, 1024–1028. 10.1126/science.aax8780.

187. Baek, M., DiMaio, F., Anishchenko, I., Dauparas, J., Ovchinnikov, S., Lee, G.R., Wang, J., Cong, Q., Kinch, L.N., Schaeffer, R.D., et al. (2021). Accurate prediction of protein structures and interactions using a three-track neural network. Science *373*, 871–876. 10.1126/science.abj8754.

188. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., et al. (2021). Highly accurate protein structure prediction with AlphaFold. Nature *596*, 583–589. 10.1038/s41586-021-03819-2.

189. Pan, X., Thompson, M.C., Zhang, Y., Liu, L., Fraser, J.S., Kelly, M.J.S., and Kortemme, T. (2020). Expanding the space of protein geometries by computational design of de novo fold families. Science *369*, 1132–1136. 10.1126/science.abc0881.

190. Yeh, A.H.-W., Norn, C., Kipnis, Y., Tischer, D., Pellock, S.J., Evans, D., Ma, P., Lee, G.R., Zhang, J.Z., Anishchenko, I., et al. (2023). De novo design of luciferases using deep learning. Nature *614*, 774–780. 10.1038/s41586-023-05696-3.

191. Cavasotto, C.N., and Di Filippo, J.I. (2023). The Impact of Supervised Learning Methods in Ultralarge High-Throughput Docking. J. Chem. Inf. Model. *63*, 2267–2280. 10.1021/acs.jcim.2c01471.

192. Gorgulla, C., Boeszoermenyi, A., Wang, Z.-F., Fischer, P.D., Coote, P.W., Padmanabha Das, K.M., Malets, Y.S., Radchenko, D.S., Moroz, Y.S., Scott, D.A., et al. (2020). An open-source drug discovery platform enables ultra-large virtual screens. Nature *580*, 663–668. 10.1038/s41586-020-2117-z.

193. Lyu, J., Wang, S., Balius, T.E., Singh, I., Levit, A., Moroz, Y.S., O'Meara, M.J., Che, T., Algaa, E., Tolmachova, K., et al. (2019). Ultra-large library docking for discovering new chemotypes. Nature *566*, 224–229. 10.1038/s41586-019-0917-9.

194. Borgo, B., and Havranek, J.J. (2014). Motif-directed redesign of enzyme specificity. Protein Sci. *23*, 312–320. 10.1002/pro.2417.

195. Crampon, K., Giorkallos, A., Deldossi, M., Baud, S., and Steffenel, L.A. (2022). Machine-learning methods for ligand–protein molecular docking. Drug Discov. Today *27*, 151–164. 10.1016/j.drudis.2021.09.007.

196. Feehan, R., Montezano, D., and Slusky, J.S.G. (2021). Machine learning for enzyme engineering, selection and design. Protein Eng. Des. Sel. *34*, gzab019. 10.1093/protein/gzab019.

197. Mazurenko, S., Prokop, Z., and Damborsky, J. (2020). Machine Learning in Enzyme Engineering. ACS Catal. *10*, 1210–1223. 10.1021/acscatal.9b04321.

198. Goldman, S., Das, R., Yang, K.K., and Coley, C.W. (2022). Machine learning modeling of family wide enzyme-substrate specificity screens. PLOS Comput. Biol. *18*, e1009853. 10.1371/journal.pcbi.1009853.

199. Röttig, M., Rausch, C., and Kohlbacher, O. (2010). Combining Structure and Sequence Information Allows Automated Prediction of Substrate Specificities within Enzyme Families. PLOS Comput. Biol. *6*, e1000636. 10.1371/journal.pcbi.1000636.

200. Ferrario, V., Siragusa, L., Ebert, C., Baroni, M., Foscato, M., Cruciani, G., and Gardossi, L. (2014). BioGPS Descriptors for Rational Engineering of Enzyme Promiscuity and Structure Based Bioinformatic Analysis. PLOS ONE *9*, e109354. 10.1371/journal.pone.0109354.

201. Holderbach, S., Adam, L., Jayaram, B., Wade, R.C., and Mukherjee, G. (2020). RASPD+: Fast Protein-Ligand Binding Free Energy Prediction Using Simplified Physicochemical Features. Front. Mol. Biosci. *7*.

202. Kroll, A., Ranjan, S., Engqvist, M.K.M., and Lercher, M.J. (2023). A general model to predict small molecule substrates of enzymes based on machine and deep learning. Nat. Commun. *14*, 2787. 10.1038/s41467-023-38347-2.

203. Robinson, S.L., Smith, M.D., Richman, J.E., Aukema, K.G., and Wackett, L.P. (2020). Machine learning-based prediction of activity and substrate specificity for OleA enzymes in the thiolase superfamily. Synth. Biol. *5*, ysaa004. 10.1093/synbio/ysaa004.

204. Martínez-Martínez, M., Coscolín, C., Santiago, G., Chow, J., Stogios, P.J., Bargiela, R., Gertler, C., Navarro-Fernández, J., Bollinger, A., Thies, S., et al. (2018). Determinants and Prediction of Esterase Substrate Promiscuity Patterns. ACS Chem. Biol. *13*, 225–234. 10.1021/acschembio.7b00996.

205. Ollikainen, N., Jong, R.M. de, and Kortemme, T. (2015). Coupling Protein Side-Chain and Backbone Flexibility Improves the Re-design of Protein-Ligand Specificity. PLOS Comput. Biol. *11*, e1004335. 10.1371/journal.pcbi.1004335.

206. Dong, L., Qu, X., and Wang, B. (2022). XLPFE: A Simple and Effective Machine Learning Scoring Function for Protein–Ligand Scoring and Ranking. ACS Omega *7*, 21727–21735. 10.1021/acsomega.2c01723.

207. Jiménez, J., Škalič, M., Martínez-Rosell, G., and De Fabritiis, G. (2018). KDEEP: Protein–Ligand Absolute Binding Affinity Prediction via 3D-Convolutional Neural Networks. J. Chem. Inf. Model. *58*, 287–296. 10.1021/acs.jcim.7b00650.

208. Rube, H.T., Rastogi, C., Feng, S., Kribelbauer, J.F., Li, A., Becerra, B., Melo, L.A.N., Do, B.V., Li, X., Adam, H.H., et al. (2022). Prediction of protein–ligand binding affinity from sequencing data with interpretable machine learning. Nat. Biotechnol., 1–8. 10.1038/s41587-022-01307-0.

209. Volkov, M., Turk, J.-A., Drizard, N., Martin, N., Hoffmann, B., Gaston-Mathé, Y., and Rognan, D. (2022). On the Frustration to Predict Binding Affinities from Protein–Ligand Structures with Deep Neural Networks. J. Med. Chem. *65*, 7946–7958. 10.1021/acs.jmedchem.2c00487.

210. Zheng, L., Fan, J., and Mu, Y. (2019). OnionNet: a Multiple-Layer Intermolecular-Contact-Based Convolutional Neural Network for Protein–Ligand Binding Affinity Prediction. ACS Omega *4*, 15956–15965. 10.1021/acsomega.9b01997.

211. Aldeghi, M., Gapsys, V., and de Groot, B.L. (2018). Accurate Estimation of Ligand Binding Affinity Changes upon Protein Mutation. ACS Cent. Sci. *4*, 1708–1718. 10.1021/acscentsci.8b00717.

212. Wang, D.D., Ou-Yang, L., Xie, H., Zhu, M., and Yan, H. (2020). Predicting the impacts of mutations on protein-ligand binding affinity based on molecular dynamics simulations and machine learning methods. Comput. Struct. Biotechnol. J. *18*, 439–454. 10.1016/j.csbj.2020.02.007.

213. Chatterjee, A., Walters, R., Shafi, Z., Ahmed, O.S., Sebek, M., Gysi, D., Yu, R., Eliassi-Rad, T., Barabási, A.-L., and Menichetti, G. (2023). Improving the generalizability of protein-ligand binding predictions with AI-Bind. Nat. Commun. *14*, 1989. 10.1038/s41467-023-37572-z.

214. Kinney, J.B., and McCandlish, D.M. (2019). Massively Parallel Assays and Quantitative Sequence–Function Relationships. Annu. Rev. Genomics Hum. Genet. *20*, 99–127. 10.1146/annurev-genom-083118-014845.

215. Hayes, R.J., Bentzien, J., Ary, M.L., Hwang, M.Y., Jacinto, J.M., Vielmetter, J., Kundu, A., and Dahiyat, B.I. (2002). Combining computational and experimental screening for rapid optimization of protein properties. Proc. Natl. Acad. Sci. *99*, 15926–15931. 10.1073/pnas.212627499.

216. Diaz, D.J., Kulikova, A.V., Ellington, A.D., and Wilke, C.O. (2023). Using machine learning to predict the effects and consequences of mutations in proteins. Curr. Opin. Struct. Biol. *78*, 102518. 10.1016/j.sbi.2022.102518.

217. Wittmann, B.J., Johnston, K.E., Almhjell, P.J., and Arnold, F.H. (2022). evSeq: Cost-Effective Amplicon Sequencing of Every Variant in a Protein Library. ACS Synth. Biol. 10.1021/acssynbio.1c00592.

218. Gelman, S., Fahlberg, S.A., Heinzelman, P., Romero, P.A., and Gitter, A. (2021). Neural networks to learn protein sequence–function relationships from deep mutational scanning data. Proc. Natl. Acad. Sci. *118*, e2104878118. 10.1073/pnas.2104878118.

219. Khersonsky, O., Lipsh, R., Avizemer, Z., Ashani, Y., Goldsmith, M., Leader, H., Dym, O., Rogotner, S., Trudeau, D.L., Prilusky, J., et al. (2018). Automated Design of Efficient and Functionally Diverse Enzyme Repertoires. Mol. Cell *72*, 178-186.e5. 10.1016/j.molcel.2018.08.033.

220. Luo, Y., Jiang, G., Yu, T., Liu, Y., Vo, L., Ding, H., Su, Y., Qian, W.W., Zhao, H., and Peng, J. (2021). ECNet is an evolutionary context-integrated deep learning framework for protein engineering. Nat. Commun. *12*, 5743. 10.1038/s41467-021-25976-8.

221. Riesselman, A.J., Ingraham, J.B., and Marks, D.S. (2018). Deep generative models of genetic variation capture the effects of mutations. Nat. Methods *15*, 816–822. 10.1038/s41592-018-0138-4.

222. Garruss, A.S., Collins, K.M., and Church, G.M. (2021). Deep representation learning improves prediction of LacI-mediated transcriptional repression. Proc. Natl. Acad. Sci. *118*, e2022838118. 10.1073/pnas.2022838118.

223. Leander, M., Yuan, Y., Meger, A., Cui, Q., and Raman, S. (2020). Functional plasticity and evolutionary adaptation of allosteric regulation. Proc. Natl. Acad. Sci. *117*, 25445–25454. 10.1073/pnas.2002613117.

224. Mathy, C.J.P., Mishra, P., Flynn, J.M., Perica, T., Mavor, D., Bolon, D.N.A., and Kortemme, T. (2023). A complete allosteric map of a GTPase switch in its native cellular network. Cell Syst. *14*, 237-246.e7. 10.1016/j.cels.2023.01.003.

225. Tack, D.S., Tonner, P.D., Pressman, A., Olson, N.D., Levy, S.F., Romantseva, E.F., Alperovich, N., Vasilyeva, O., and Ross, D. (2021). The genotype-phenotype landscape of an allosteric protein. Mol. Syst. Biol. *17*, e10179. 10.15252/msb.202010179.

226. Tonner, P.D., Pressman, A., and Ross, D. (2022). Interpretable modeling of genotype–phenotype landscapes with state-of-the-art predictive power. Proc. Natl. Acad. Sci. *119*, e2114021119. 10.1073/pnas.2114021119.

227. Srikant, S., Gaudet, R., and Murray, A.W. (2020). Selecting for Altered Substrate Specificity Reveals the Evolutionary Flexibility of ATP-Binding Cassette Transporters. Curr. Biol. *30*, 1689-1702.e6. 10.1016/j.cub.2020.02.077.

228. Makowski, E.K., Kinnunen, P.C., Huang, J., Wu, L., Smith, M.D., Wang, T., Desai, A.A., Streu, C.N., Zhang, Y., Zupancic, J.M., et al. (2022). Co-optimization of therapeutic antibody affinity and specificity using machine learning models that generalize to novel mutational space. Nat. Commun. *13*, 3788. 10.1038/s41467-022-31457-3.

229. Gainza, P., Wehrle, S., Van Hall-Beauvais, A., Marchand, A., Scheck, A., Harteveld, Z., Buckley, S., Ni, D., Tan, S., Sverrisson, F., et al. (2023). De novo design of protein interactions with learned surface fingerprints. Nature, 1–9. 10.1038/s41586-023-05993-x.

230. Lu, C., Lubin, J.H., Sarma, V.V., Stentz, S.Z., Wang, G., Wang, S., and Khare, S.D. (2023). Prediction and Design of Protease Enzyme Specificity Using a Structure-Aware Graph Convolutional Network. 2023.02.16.528728. 10.1101/2023.02.16.528728.

231. Sellés Vidal, L., Murray, J.W., and Heap, J.T. (2021). Versatile selective evolutionary pressure using synthetic defect in universal metabolism. Nat. Commun. *12*, 6859. 10.1038/s41467-021-27266-9.

232. Quijano-Rubio, A., Yeh, H.-W., Park, J., Lee, H., Langan, R.A., Boyken, S.E., Lajoie, M.J., Cao, L., Chow, C.M., Miranda, M.C., et al. (2021). De novo design of modular and tunable protein biosensors. Nature *591*, 482–487. 10.1038/s41586-021-03258-z.

233. Collins, S.P., Rostain, W., Liao, C., and Beisel, C.L. (2021). Sequence-independent RNA sensing and DNA targeting by a split domain CRISPR–Cas12a gRNA switch. Nucleic Acids Res. *49*, 2985–2999. 10.1093/nar/gkab100.

234. Kim, H.J., Lim, J.W., Jeong, H., Lee, S.-J., Lee, D.-W., Kim, T., and Lee, S.J. (2016). Development of a highly specific and sensitive cadmium and lead microbial biosensor using synthetic CadC-T7 genetic circuitry. Biosens. Bioelectron. *79*, 701–708. 10.1016/j.bios.2015.12.101.

235. Ostrov, N., Jimenez, M., Billerbeck, S., Brisbois, J., Matragrano, J., Ager, A., and Cornish, V.W. (2017). A modular yeast biosensor for low-cost point-of-care pathogen detection. Sci. Adv. *3*, e1603221. 10.1126/sciadv.1603221.