

---

Original Article

# Development of novel metabolite-responsive transcription factors via transposon-mediated protein fusion

Andrew K.D. Younger<sup>1,†</sup>, Peter Y. Su<sup>2,†</sup>, Andrea J. Shepard<sup>2</sup>,  
Shreya V. Udani<sup>3</sup>, Thaddeus R. Cybulski<sup>4</sup>, Keith E.J. Tyo<sup>2,5,6</sup>,  
and Joshua N. Leonard<sup>2,5,6,7,\*</sup>

<sup>1</sup>Interdisciplinary Biological Sciences (IBiS) Graduate Program, Northwestern University, 2-100 Hogan Hall, 2205 Tech Drive, Evanston, IL 60208, USA, <sup>2</sup>Department of Chemical and Biological Engineering, McCormick School of Engineering and Applied Science, Northwestern University, 2145 Sheridan Road, Room E-136, Evanston, IL 60208, USA, <sup>3</sup>Department of Biomedical Engineering, McCormick School of Engineering and Applied Science, 2145 Sheridan Road, Evanston, IL 60208, USA, <sup>4</sup>Department of Physical Medicine and Rehabilitation, Northwestern University and Rehabilitation Institute of Chicago, 710 North Lake Shore Drive, #1022, Chicago, IL 60611, USA, <sup>5</sup>Chemistry of Life Processes Institute, 2170 Campus Dr, Evanston, IL 60208, USA, <sup>6</sup>Center for Synthetic Biology, Technological Institute, 2145 Sheridan Road, Evanston, IL 60208, USA, and <sup>7</sup>Robert H. Lurie Comprehensive Cancer Center, Northwestern University, 303 E. Superior L3—125, Chicago, IL 60611, USA

\*To whom correspondence should be addressed. E-mail: j-leonard@northwestern.edu

<sup>†</sup>These authors contributed equally to this work.

Edited By: Andreas Matouschek

Received 4 April 2017; Revised 16 November 2017; Editorial Decision 2 January 2018; Accepted 7 January 2018

## Abstract

Naturally evolved metabolite-responsive biosensors enable applications in metabolic engineering, ranging from screening large genetic libraries to dynamically regulating biosynthetic pathways. However, there are many metabolites for which a natural biosensor does not exist. To address this need, we developed a general method for converting metabolite-binding proteins into metabolite-responsive transcription factors—Biosensor Engineering by Random Domain Insertion (BERDI). This approach takes advantage of an *in vitro* transposon insertion reaction to generate all possible insertions of a DNA-binding domain into a metabolite-binding protein, followed by fluorescence activated cell sorting to isolate functional biosensors. To develop and evaluate the BERDI method, we generated a library of candidate biosensors in which a zinc finger DNA-binding domain was inserted into maltose binding protein, which served as a model well-studied metabolite-binding protein. Library diversity was characterized by several methods, a selection scheme was deployed, and ultimately several distinct and functional maltose-responsive transcriptional biosensors were identified. We hypothesize that the BERDI method comprises a generalizable strategy that may ultimately be applied to convert a wide range of metabolite-binding proteins into novel biosensors for applications in metabolic engineering and synthetic biology.

**Key words:** biosensor, domain insertion, metabolic engineering, synthetic biology, transcription factor, transposon

---

## Introduction

Metabolite-responsive biosensors have a wide variety of uses, from basic research and discovery, to diagnostics, to engineered biosynthesis (Khalil and Collins, 2010). Such biosensors include diverse sensing and output modalities, including fluorescent and FRET-based biosensors (Golynskiy *et al.*, 2011; Strianese *et al.*, 2012), RNA-based biosensors (Michener *et al.*, 2012; Kang *et al.*, 2014) and transcription factor biosensors (Dietrich *et al.*, 2010; Brockman and Prather, 2015; Venayak *et al.*, 2015). Transcription factor biosensors have proven to be especially powerful for bioengineering, facilitating dynamic profiling of intracellular glucaric acid (Rogers *et al.*, 2015) and malonyl-CoA (Li *et al.*, 2015), and enabling high-throughput screening of large genetic libraries constructed to achieve biosynthesis of 1-butanol, succinate and adipate (Dietrich *et al.*, 2013), benzoic acids (van Sint Fiet *et al.*, 2006), and L-Lysine (Binder *et al.*, 2012). Moreover, transcription factor biosensors have been harnessed to implement dynamic intracellular feedback control, balancing metabolic fluxes to increase production titers and yields of lycopene (Farmer and Liao, 2000), fatty acid ethyl ester (Zhang *et al.*, 2012), amorphadiene (Dahl *et al.*, 2013), 1-butanol (Dietrich, Shis, Alikhani and Keasling, 2013) and malonyl-CoA (Liu *et al.*, 2015). Notably, these examples relied upon naturally evolved transcription factor biosensors, and broader utilization of such approaches is currently restricted by the limited pool of naturally evolved (and known) metabolite-responsive biosensors. Thus, approaches for generating novel metabolite biosensors are required.

Several strategies for generating new biosensors have been explored. One approach for generating new transcription factor biosensors is fusion of the ligand-binding domain from one transcription factor to the DNA-binding domain from a different transcription factor. However, these chimeric biosensors are generally limited to fusions within families of structurally related transcription factors, such as the LacI/GalR family, in order to preserve mechanisms of ligand-responsiveness that arise from allosteric regulation (Meinhardt *et al.*, 2012; Shis *et al.*, 2014). An alternative approach is to fuse a metabolite-binding protein (which multimerizes upon ligand-binding) to a natural transcription factor, such as AraC, to generate chimeras in which metabolite binding modulates transcription factor activity (Chou and Keasling, 2013). Additionally, the binding pockets of transcription factors such as LuxR (Collins *et al.*, 2005, 2006), AraC (Tang *et al.*, 2008, 2013; Tang and Cirino, 2011) and XylR (Mohn *et al.*, 2006) have been mutagenized and evolved to bind new, albeit structurally similar, ligands. To engineer novel biosensors in eukaryotes, fusion proteins were engineered to be unstable in the absence of ligand, such that the addition of the ligand stabilized the protein and enabled it to carry out its functional role (Feng *et al.*, 2015). Recently, transposon-based random insertion was used to identify sites in the Cas9 nuclease that are permissive to peptide insertion, and then well-characterized ligand binding-induced changes in the conformation of the human estrogen receptor- $\alpha$  domain were leveraged to generate an estrogen-regulated Cas9 nuclease (Oakes *et al.*, 2016). While these findings have generated both useful biosensors and novel insights into biosensor design, most methods include some inherent limitation on the extent to which they may be generalized to build biosensors for any metabolite of interest. Therefore, there remains an outstanding need for new methods for generating novel transcription factor biosensors.

We recently reported a new strategy for converting a ligand-binding protein into a transcription factor biosensor (Younger *et al.*,

2016). In this proof-of-principle investigation, the *Escherichia coli* maltose binding protein (MBP) was genetically fused with a modular zinc finger DNA-binding domain (ZFP) to generate a novel maltose-responsive transcription factor, in which the addition of maltose alleviated transcriptional repression of an engineered promoter. This demonstration leveraged a wealth of prior knowledge pertaining to MBP; specifically the ZFP was inserted into MBP at a position that was previously identified via random fusion between MBP and TEM1  $\beta$ -lactamase (bla) to generate a maltose-regulated bla (Guntas *et al.*, 2004). Whether other fusions between MBP and a ZFP could generate a functional biosensor, and whether a functional biosensor could be generated in the absence of such prior knowledge remained open questions.

To address these questions, here we developed an efficient method for generating combinatorial fusions between a ligand-binding protein and a ZFP, followed by isolation of functional biosensors from this diverse library, which we term Biosensor Engineering by Random Domain Insertion (BERDI). We validate this method by generating multiple novel, functional biosensors and thus hypothesize that BERDI may comprise a generalizable strategy for converting metabolite-binding proteins into biosensors.

## Materials and Methods

### Bacterial strains and culturing

All experiments were conducted in DS941 Z1 *E. coli* cells (AB1157, recF143, lacI<sup>q</sup> lacZ  $\Delta$ M15, P<sub>lacIq</sub>-LacI, P<sub>N25</sub>-TetR). Cells were maintained in Lysogeny Broth (LB) Lennox formulation (10 g/L tryptone, 5 g/L yeast extract, 5 g/L NaCl) supplemented with appropriate antibiotics (Ampicillin 100  $\mu$ g/mL, Kanamycin 50  $\mu$ g/mL, and/or Chloramphenicol 34  $\mu$ g/mL). All experimental analyses were conducted in M9 minimal media (1 $\times$  M9 salts, 0.2% Cas amino acids, 2 mM MgSO<sub>4</sub>, 0.1 mM CaCl<sub>2</sub>, 1 mM Thiamine HCl) containing glycerol (0.4%) as the primary carbon source. Variable amounts of isopropyl  $\beta$ -D-1-thiogalactopyranoside (IPTG) were added, as indicated, to induce biosensor expression. Maltose monohydrate was added to the media at a final concentration of 100 mM, where indicated.

The biosensor expression vector was built using standard molecular biology techniques using parts (GFPmut3b and pTrc2) gifted by Jim Collins (MIT) (Litcofsky *et al.*, 2012). The green fluorescent protein (GFP) reporter plasmid driven by the pGo92 zinc finger-responsive promoter was previously described (Younger, Dalvie, Rottinghaus and Leonard, 2016). Custom RBS sequences for the biosensor and reporter plasmids were designed using the RBS calculator (Salis *et al.*, 2009). The camR and sacB ORFs were transferred from pKM154, which was gifted by Kenan Murphy (University of Massachusetts) (Murphy *et al.*, 2000) (Addgene plasmid #13 036), into a storage vector containing MuA transposon recognition sequences, flanked by BglII restriction sites (pAY438). The BCR-ABL1 ZFP was subcloned into a storage vector flanked by NotI restriction sites (pAY437). Description of all plasmids used in this study can be found in Supplemental Table S1, and electronic plasmid sequence files are included in Supplementary Data.

### Candidate biosensor library construction

The MuA transposase inserts its transposon randomly, and in either a forward or reverse direction, into any DNA sequence (Haapa *et al.*, 1999). Furthermore, the transposon can be inserted in any of

the three possible codon frames in MBP. A detailed description of the transposon sequence and potential scar options can be found in Supplemental Fig. S1.

First, a library representing all possible random insertions within MBP was generated. Double-stranded DNA comprising a transposon conferring chloramphenicol resistance as well as containing the *sacB* gene for negative selection with sucrose was digested out of a storage plasmid (pAY438) using BglII, gel extracted, and cleaned by ethanol precipitation/resuspension in 40  $\mu$ L of TE buffer. *In vitro* transposition reactions were carried out using the Mutation Generation System kit (Thermo Scientific # F701), as per the manufacturer's protocol. Briefly, 100 ng of purified transposon was mixed with 200 ng of target plasmid encoding MBP (pAY447), and the mixture was incubated with 1  $\mu$ L of 0.22 ng/ $\mu$ L MuA transposase for 4 h at 30°C. MuA was heat-inactivated (10 min at 75°C), and a PCR cleanup (IBI Scientific) was conducted to recover the library. The entire library was electroporated into two tubes of electrically competent *E. coli* cells (~250  $\mu$ L final volume each). Transformed cells were selected on plates containing chloramphenicol (transposon) as well as ampicillin (plasmid backbone). Serial dilutions were made at each cloning step and extrapolated to estimate library size. The MBP gene was digested out with restriction enzymes KpnI and SphI and purified by agarose gel electrophoresis to separate the band representing MBP with transposon insertion (3923 bp) from the band representing WT MBP (1122 bp). The MBP with transposon band was purified and cloned into an expression plasmid under the control of a lac-inducible promoter pTrc2 (pAY431). Finally, restriction digestion (using the NotI site present in the transposon scar) was used to replace the transposon with the sequence encoding the ZFP (BCR-ABL1), and this ligation was transformed into competent *E. coli* cells that already contained the ZFP-responsive GFP reporter plasmid (pAY430). Cells were selected with ampicillin and kanamycin for both plasmids as well as 10% sucrose to maximize loss of the transposon, yielding the naïve (unselected) candidate biosensor library.

### Microplate-based fluorescent assays and analysis

Cultures were inoculated from single colonies into 2 mL of M9 media and grown overnight to stationary phase. Overnight cultures were diluted 1:10 and grown for 1–2 h (OD<sub>600</sub> ~ 0.5). Cultures were again diluted 1:10 (OD<sub>600</sub> ~0.05), plated in black-walled clear bottom 96-well plates in biological triplicate, and induced with 30  $\mu$ M IPTG and/or 100 mM maltose. Plates with lids were incubated and shaken in a continuous double orbital pattern at 548 cpm (2 mm) inside a BioTek Synergy H1 plate reader for 10 h with GFP fluorescence and OD<sub>600</sub> absorption measurements taken every 15 min. Monochromator settings were 485/515 nm for GFP.

### Flow cytometry and fluorescence activated cell sorting

Overnight cultures (2 mL) were diluted 1:10 into a fresh 2 mL aliquot of M9 media and grown for 1–2 h (OD<sub>600</sub> ~ 0.5). Cultures were again diluted 1:10 (OD<sub>600</sub> ~0.05) in a fresh 2 mL of either M9 media, or M9 media containing 100  $\mu$ M IPTG. Cultures were grown for 4 h post-induction prior to fluorescence activated cell sorting (FACS) sorting. Cells were then diluted down to a concentration of 10<sup>7</sup> cells/mL in 4°C PBS. Sorting was performed on a BD FACS Aria II instrument (BD Biosciences, San Jose, USA) using an 85  $\mu$ m tip with a 488 nm excitation laser and a FITC emission filter (530/30 nm). This FITC channel was used for analysis of GFP

expression. Cells were first gated based upon forward and side scatter, then the population of single cells were plotted on a GFP histogram. To set a gate for recovering cells exhibiting biosensor-mediated repression of reporter (GFP) output, a distribution of GFP fluorescence in cells was obtained from the population (100 000 events), and gating was set such that no more than 1% of this 'ON' (uninduced) population would be recovered. This gate was used to recover biosensor candidates capable of reporter repression. To recover 'reversible' repressors (minimize false positive repressors), the same gating definition described above was used, but this time (in the absence of IPTG), 'ON' cells were recovered. For each round of sorting, 100 000 cells were recovered into 3 mL of M9 minimal media containing ampicillin and kanamycin, and this culture was subsequently inoculated into 50 mL of M9 containing ampicillin and kanamycin and grown overnight at 37°C. Subsequent sorts were performed (as indicated) the next day using the sorted and expanded population, as above. Traditional flow cytometry was performed on a LSRII flow cytometer (BD Biosciences) or an Attune flow cytometer (Thermo Fisher Scientific). For all flow cytometry analyses, mean fluorescent intensity was calculated based on the GFP histograms of single cells (gated by forward and side scatter) using FlowJo Software (Tree Star).

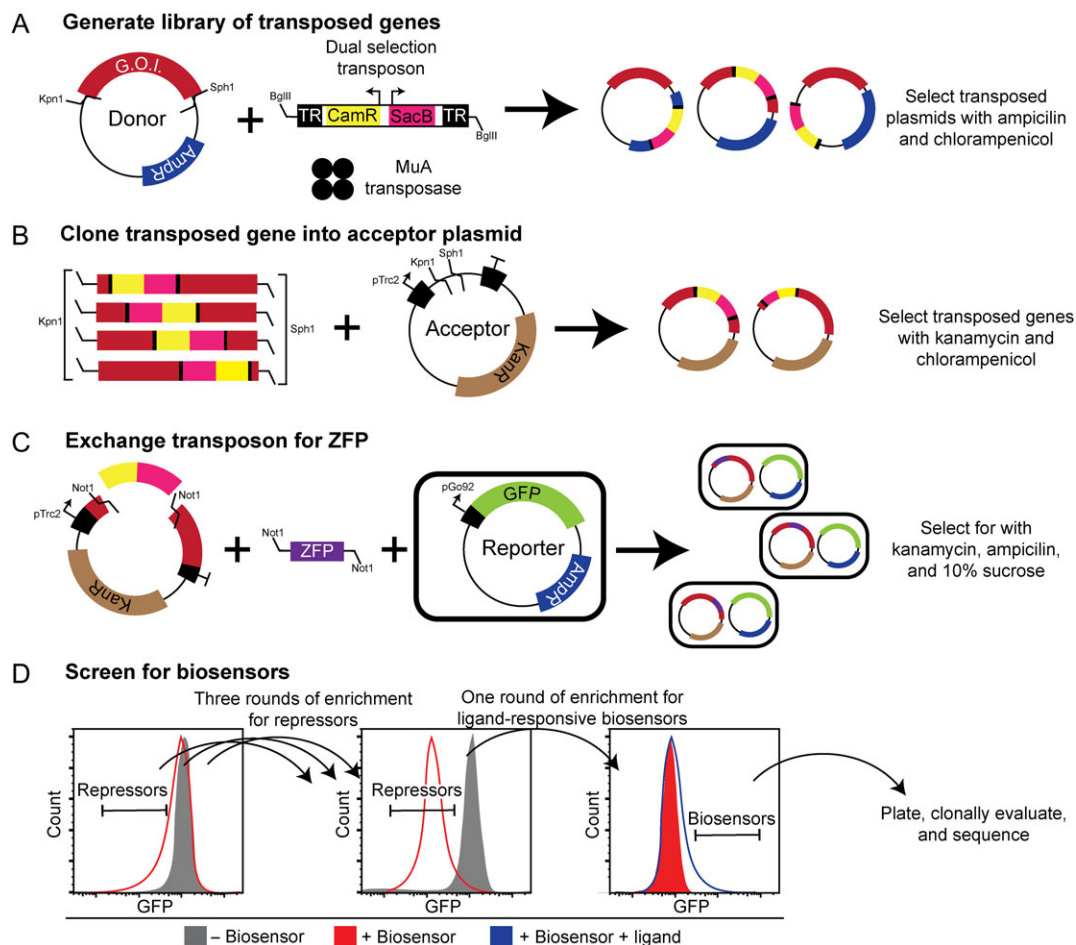
### Digest-based evaluation of library diversity

All gel electrophoresis experiments were conducted with a 1% agarose gel and run in 1× TAE (tris acetate EDTA) at 120 V. DNA was stained using SYBR Safe (Thermo Scientific) and imaged under blue light. Band sizes were estimated using a 1 kb ladder (New England BioLabs). Exposure time was adjusted to maximize the differences between the lows and highs in the gel. Plot profiles of resulting gel images were analyzed using ImageJ's 'plot profile' function. Intensity profiles for each lane were generated by subtracting the 'gray value' from an empty gel lane from the grey value evaluated along the length of the lane of interest.

## Results

### Generation of random domain insertion libraries via transposon mutagenesis

Here we sought to develop an efficient method for generating novel transcription factor biosensors, which we term BERDI. The overall BERDI strategy is summarized in Fig. 1. This method utilizes the MuA transposase to insert a transposon nonspecifically into a plasmid encoding the gene sequence for the metabolite-binding protein. Transposon mutagenesis provides a simple and efficient method for generating a library of insertions due to its non-specific and single insertion into each target DNA molecule, minimal scar sequence, and ability to generate > 10<sup>5</sup> variants in a single pot *in vitro* reaction (Haapa *et al.*, 1999; Mehta *et al.*, 2012; Nadler *et al.* 2016). The transposon is later exchanged for a ZFP coding sequence to generate a library of candidate biosensors. Transposase insertion has been used to circularly permute proteins as well as to profile proteins to identify sites permissive of peptide insertion or protein fusion (Edwards *et al.*, 2008; Segall-Shapiro *et al.*, 2011; Mehta *et al.*, 2012; Nadler *et al.*, 2016; Oakes, Nadler, Flamholz, Fellmann, Staahl, Doudna and Savage, 2016). Here, we investigated whether this technique can be used to generate novel biosensors by randomly inserting a ZFP into a ligand-binding protein. We reasoned that choosing MBP as a model ligand-binding domain would



**Fig. 1** Overview of the BERDI method for generating novel metabolite-responsive transcription factor biosensors. **(A)** Library generation—the donor plasmid containing the gene of interest, the transposon, and the transposase enzymes create a library of random insertions. **(B)** Cloning of the transposed gene. Gene containing the inserted transposon is isolated, and cloned into a similarly digested expression plasmid. **(C)** Exchanging the transposon for the ZFP. The transposon is replaced with ZFP and are transformed into cells containing the reporter GFP plasmid. **(D)** Cartoon of potential enrichment strategy for metabolite responsive biosensors using FACS. See Materials and Methods for experimental details.

be the most effective strategy for evaluating the BERDI method, since (i) we know that at least one such feasible biosensor should exist, based upon our prior work (Younger, Dalvie, Rottinghaus and Leonard, 2016); and (ii) such an approach enabled us to utilize established reporter constructs and biosensor evaluation methods.

First, a library of candidate biosensors was generated. MuA inserts randomly into target DNA molecules (in either forward or reverse direction), such that for a plasmid of length  $n$  bases, the total number of possible insertions is  $2n$ . Additionally, multiplying the three frames in which the transposon can insert by the two directions in which the transposon can insert (i.e. forward or reverse) yields six possible insertions for a given codon of target DNA. However, only one of these six insertions (forward and in-frame) will generate a productive insertion. The transposase also leaves a partially controllable scar (i.e. one has some choice in the design of this scar sequence). Therefore, we designed the transposon such that when the ZFP is inserted in frame with the rest of MBP, the resulting scars encode for linkers comprising three alanine residues on either side of the ZFP domain (see Supplemental Fig. S1B for details).

We initially aimed to achieve a library size at least 10× greater than the maximum possible number of insertions—6288 (the

number of directions in which the transposon can insert, 2, multiplied by the size of the target plasmid, 3144 bp); this yields a target library size of 62 880 members. Our initial transposition library generated over  $8 \times 10^5$  colonies, or  $\sim 125\times$  the maximum library diversity. Next, the library was subcloned to eliminate MBP variants lacking a transposon (i.e. when the transposon inserted elsewhere in the plasmid backbone), and then the transposon was replaced with a sequence encoding the ZFP to yield a naïve candidate biosensor library. At each step, library size exceeded the target of 10× oversampling (Supplemental Table S2).

### Analyzing diversity of the naïve library

We analyzed naïve library diversity using two distinct methods. First, sequences from the naïve library were digested out of the expression vector, and these biosensor-encoding fragments were subsequently digested with a restriction enzyme recognizing a unique site in the sequence encoding the ZFP. These digests were evaluated by gel electrophoresis, yielding a smear consistent with multiple insertions (Supplemental Fig. S2). We next sequenced the plasmid content of 46 individual colonies (Supplemental Table S3).

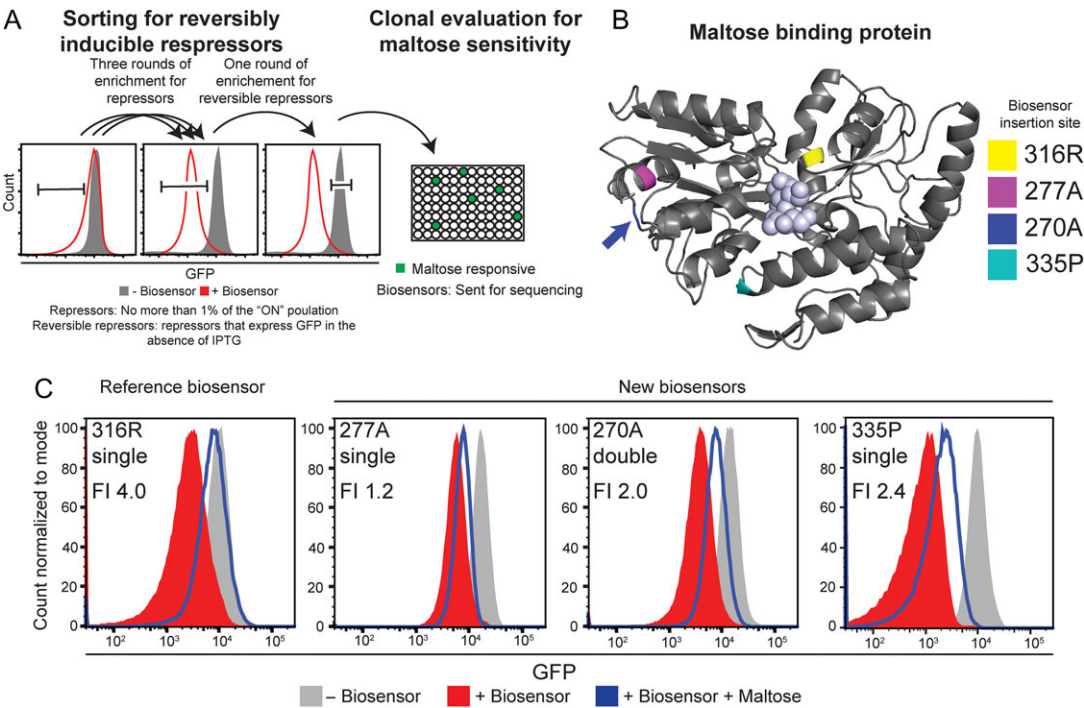
Surprisingly, this analysis revealed that 15 of the 46 colonies sequenced yielded an insertion at position 901, and the other 31 colonies yielded 11 additional insertion sites. In-frame insertions were observed at somewhat lower frequency than expected, and no bias for forward versus reverse orientation insertions was noted (Table I). While Mu transposon-based insertion is known to exhibit some sequence-associated preference (Green *et al.*, 2012), it is also possible that the overrepresented insertions conferred growth advantages (i.e. when biosensors were expressed at low levels due to leakiness from the IPTG-inducible promoter), or that some sequences were more efficiently subjected to subcloning during naïve library generation. To evaluate whether our library was sufficient to generate candidate biosensors (even in the context of the observed bias), it was carried forward for functional evaluation.

**Table I.** Frequency of insertion types observed in the naïve candidate biosensor library

| Type of insertion | Expected if unbiased (%) | Observed by colony sequencing ( <i>n</i> = 17 insertions) (%) |
|-------------------|--------------------------|---------------------------------------------------------------|
| In frame          | 33                       | 12                                                            |
| Out of frame      | 66                       | 88                                                            |
| Forward ZFP       | 50                       | 48                                                            |
| Reverse ZFP       | 50                       | 52                                                            |

Identification of functional biosensors

Our overall strategy was to first enrich for candidate biosensors capable of repressing the reporter in the absence of maltose, then to reselect to eliminate false positives (i.e. to enrich for reversible repressors), and then to identify maltose-responsive biosensors from within that pool (Fig. 2A). The naïve library was first screened by FACS to enrich for candidate biosensor clones which, when expressed at a high level (100 μM IPTG induction), repressed the GFP reporter output. The selected pool was recovered and regrown, and this process was iterated on three consecutive days to enrich for candidate biosensors that were at least capable of repressing reporter output. Next, this population of candidate biosensors was re-screened by FACS in the absence of IPTG, to recover ‘reversible’ repressors (clones which express GFP in the absence of IPTG) in order to minimize false positives. Following these four rounds of screening, the recovered cells were plated and clonally assayed for maltose-responsiveness. Candidate biosensors were initially evaluated under conditions of 30 μM IPTG to induce the expression of the biosensor and 100 mM maltose, which are conditions under which we previously observed maximal maltose-responsiveness of the reference biosensor (Younger, Dalvie, Rottinghaus and Leonard, 2016). Note that this extracellular concentration of maltose exceeds the intracellular concentration due to transport limitations. At this point, library members exhibiting both (i) at least a 2-fold repression of reporter output in the absence of maltose; and (ii) any significant maltose-responsiveness were defined as ‘functional biosensors’ and were sequenced. Of the 672 colonies analyzed, 340 demonstrated



**Fig. 2** Isolation of functional biosensors by screening. (A) Cartoon depicting the overall biosensor enrichment strategy applied. Briefly, three rounds of FACS were performed, each time sorting the induced population using a gate encompassing no more than 1% of the ‘ON’ (uninduced) population. One subsequent round of sorting was done to isolate only repressors that were reversible. Cells were then plated, and clonally evaluated for maltose responsiveness. Successful biosensors were sequenced to determine the insertions position. (B) Crystal structure of MBP is shown in gray (PDB# 1ANF)(Quiocho *et al.*, 1997), with the insertional positions (in amino acid number) of each biosensor labeled. A cluster of lavender spheres represents the ligand, maltose (space-filling model). (C) Flow cytometry of the reference biosensor compared to the three new biosensors. Biosensor production was induced with 30 μM IPTG and maltose was added at 100 mM (extracellular concentration). The insertional position (in amino acid number), and whether the ZFP is a single, or double insertion, is listed in the top left corner of each plot. Plots represent a minimum of 10 000 cells in each condition and are representative of multiple independent experiments. FI, fold-induction.

the ability to repress the reporter, but only 159 of those met both criteria for potential biosensors. These 159 colonies represented three unique biosensors (270 A, 277 A and 335 P). The other 181 repressors represented out-of-frame insertions at six positions (5E, 31 G, 188 G, 194 T, 213I and 262 V). Moreover, none of the three novel biosensors identified were detected in the original naïve library, via colony sequencing. Together, these observations indicate that our strategy substantially enriched for functional biosensor constructs, even if they were rather rare in the original naïve library.

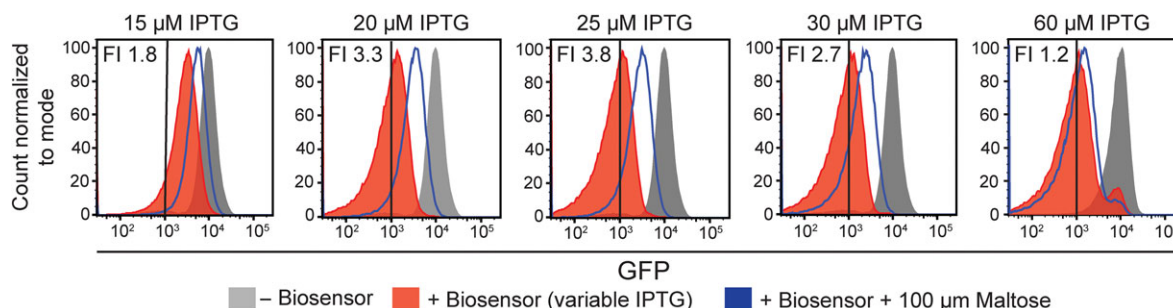
The three functional biosensors, representing a ZFP insertion at 277 A, an insertion of two ZFPs at 270 A, and a single ZFP insertion at 335 P, were next examined in greater detail. These insertions are depicted graphically in Fig. 2B, along with the position at which the ZFP was previously inserted by design in the reference biosensor, at position 316 R (Younger, Dalvie, Rottinghaus and Leonard, 2016). Interestingly, these four insertion points are distributed in three distinct regions of MBP. All four are on the outside of the protein and are either in a loop (270 A) or at the end of an  $\alpha$ -helix, near a loop (316 R, 277 A and 335 P). Given the sample size, and the lack of crystal structures of the new biosensors, it is not yet possible to predict whether they share other features that lend themselves to maltose-responsive transcriptional regulation. Performance of the new biosensors was next compared to that achieved by the reference biosensor (Fig. 2C). Both the 277 A and 270 A biosensors exhibited similar repression to that conferred by the reference biosensor (~3 to 4-fold), however, neither 277 A nor 270 A exhibited as much maltose-responsiveness as did the reference biosensor. Notably, the 335 P biosensor exhibited much better repression (~10-fold) compared to the other three constructs, while also exhibiting substantial responsiveness to maltose. To more systematically characterize biosensor responsiveness, we performed a dose response analysis varying both inducer (IPTG) and ligand (maltose) concentrations (Supplemental Fig. S3). Overall, the biosensors exhibited performance that was qualitatively consistent with that of the reference biosensor (Younger, Dalvie, Rottinghaus and Leonard, 2016); at low IPTG concentrations, little repression was observed, and at high IPTG concentrations, intracellular maltose was insufficient to alleviate repression, such that fold-induction was maximal for intermediate IPTG levels. We next investigated why our BERDI method recovered a 270 A double ZFP insertion but not a 270 A single ZFP insertion. Such a double ZFP insertion is indeed a potential product that could be generated when cloning the ZFP cassette in place of the transposon cassette (both ends of the ZFP cassette use the same NotI restriction site), but given the ligation ratios used, we expected such a double insertion to be substantially less frequent than a single

insertion. Therefore, we next generated the 270 A single ZFP insertion biosensor to evaluate its performance. First, the 270 A single ZFP construct exhibited milder repression compared to the 270 A double ZFP variant (Supplemental Fig. S4), which would explain why the single insertion variant was not (or was less) enriched during the initial three rounds of FACS screening. Furthermore, the single ZFP insertion at 270 A exhibited less responsiveness to maltose than did the double insertion biosensor variant. Taken together, these clonal observations support the conclusion that the BERDI method generated and selected for functional biosensors, based upon their performance, as intended.

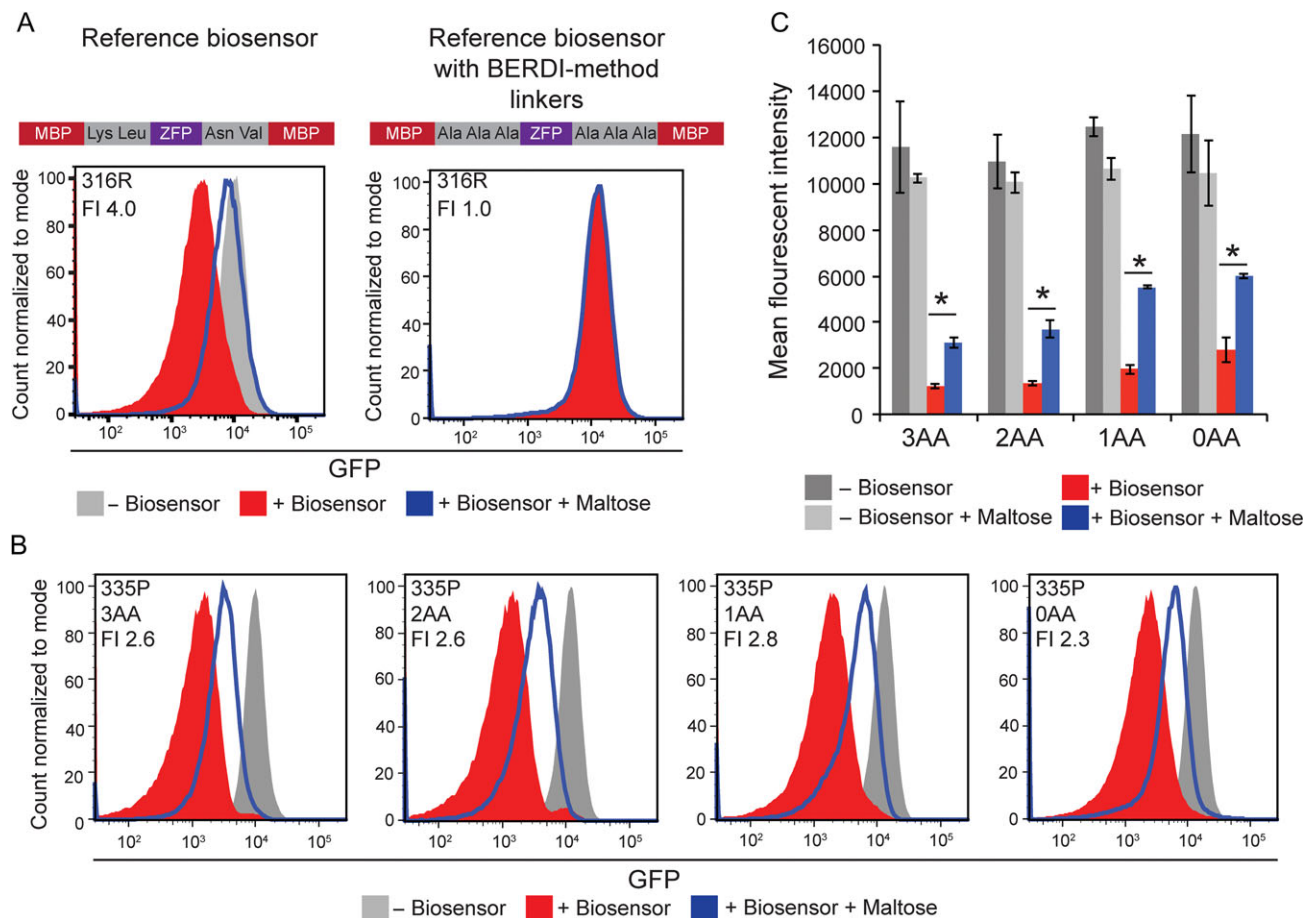
### Biosensor performance characteristics: dose and linker analysis

Having identified several novel functional biosensors, we next evaluated their performance characteristics. First, to investigate the impact of biosensor dose on reporter output repression and maltose sensitivity, the strongest repressor, 335 P, was induced at a range of IPTG concentrations (Fig. 3). At the highest IPTG concentration evaluated here (60  $\mu$ M), the repression was not significantly greater than that observed at 25 or 30  $\mu$ M IPTG, indicating that at these lower concentrations, saturating levels of the biosensor already achieve maximal repression. However, at 60  $\mu$ M IPTG, the system was not sensitive to maltose, indicating that the biosensor was in excess relative to intracellular maltose levels (which would be lower than extracellular maltose levels). As the level of IPTG (and thus biosensor expression) decreased, the sensitivity of the 335 P biosensor to maltose increased, but once the IPTG level dropped below 20  $\mu$ M IPTG, the overall repression of the reporter decreases, as expected given the decrease in biosensor protein. Thus, there exists an optimal window of biosensor expression that confers maximal maltose sensitivity while maintaining sufficient promoter repression (in the absence of maltose), corresponding to around 25  $\mu$ M IPTG (different from the 100  $\mu$ M IPTG used during selection). Each of these phenomena is consistent with observations previously characterized with the reference biosensor (Younger, Dalvie, Rottinghaus and Leonard, 2016), suggesting that at least at this functional level, the reference biosensor and this novel BERDI-generated biosensor exhibit qualitatively similar performance characteristics.

Intiguously, however, a biosensor matching the original reference biosensor (e.g. a ZFP insertion 316 R) was not recovered by the BERDI method, and we next investigated why this may be. One possible explanation is that the linkers introduced via the BERDI method differ from those included in the reference biosensor. The



**Fig. 3** Impact of biosensor expression level on performance. Response of reporter output to the addition of IPTG and IPTG along with maltose measured by flow cytometry. The maltose concentration used here was 100 mM. The vertical line at  $10^3$  GFP fluorescence units is a visual aid to facilitate comparison across the conditions. Plots represent multiple independent experiments. FI, fold-induction.



**Fig. 4** Impact of linker tuning on biosensor performance. **(A)** Comparison of the effect of amino acid linkers between reference biosensor and its transposon-created counterpart on repressibility and alleviation with maltose. **(B)** Impact of varying linker lengths for the 335 P biosensor. Biosensors were induced with 30  $\mu$ M IPTG and maltose was added at 100 mM (extracellular concentration). **(C)** Mean fluorescence intensity of the four linker variants of the 335 P biosensor measured via flow cytometry. Biosensors were induced with 30  $\mu$ M IPTG and maltose was added at 100 mM final concentration. Samples were run in biological triplicate, and error bars represent one standard deviation. (\* $P \leq 0.001$  from a two-tailed student's *t*-test). All data are representative of multiple independent experiments. FI, fold-induction.

reference biosensor has two amino linkers on either side of the ZFP—lysine and leucine on the 5' end of the ZFP insertion and an asparagine and valine on the 3' end—whereas the BERDI method introduces three alanines on either side of the ZFP (Supplemental Fig. S1). To investigate the impact of this difference in linkers, a biosensor was generated in which the ZFP was inserted at 316 R, with three alanines on each side of the ZFP, and this construct was compared to the original reference biosensor (Fig. 4A). Surprisingly, the 316 R biosensor with three Ala-linkers (mimicking the transposon scar) completely lost its ability to repress the GFP reporter. This observation could explain why the BERDI method did not recover a single ZFP insertion at 316 R, and moreover, it demonstrates the importance of linker length (and potentially composition) on impacting overall biosensor performance. To further investigate how linker length may affect biosensor performance, three variants of the 335 P biosensor (with 3AA linkers) were generated, with reduced linker lengths (0–2 AA on each side of the ZFP). As the linker length shortened, the repression of the biosensor decreased (Fig. 4B and C). Interestingly, the maltose-responsiveness also varied with linker length, such that the 1AA 335 P biosensor exhibited the best combination of promoter repression and maltose responsiveness. Altogether, these observations both validate the BERDI

method for generating novel functional biosensors and indicate that this modular biosensor design is amenable to further refinement of biosensor performance.

## Discussion

In this study, we developed and implemented the BERDI method for the generation of maltose-responsive MBP-ZFP fusion proteins in a rapid and efficient manner to find three new biosensors. The fact that multiple insertions produced a bi-functional protein is not surprising given that a previous study found twelve functional insertions for a circular permuted GFP into MBP using a similar method (Nadler, Morgan, Flamholz, Kortright and Savage, 2016). Additionally, another transposon insertion study demonstrated multiple bi-functional insertions of a cytochrome into  $\beta$ -lactamase (Edwards, Busse, Allemann and Jones, 2008), demonstrating that if multiple possible insertions exist, this method is capable of identifying them. A possible explanation for the tolerance of the proteins studied both here and in previous research is that many circularly permuted proteins are able to retain their function, demonstrated in a study that found 15 unique functional circular permutations of an adenylate kinase using transposon mutagenesis (Mehta, Liu and

Silberg, 2012). A second explanation may be that these are monomeric proteins, as homodimers would require the protein complex to tolerate two changes simultaneously. These findings emphasize the need for library based approaches, like the one described here, given the propensity for a given protein to have multiple positions where functional fusions can be created.

Although our naïve library reflected some bias with respect to insertion sites, we nonetheless found three novel functional biosensors. The three biosensors that were enriched in the screening process were not detected by colony sequencing, indicating that our screening method can isolate infrequent mutants from the initial library. Considering the observed bias, to avoid losing rare variants, it may be useful to employ a minimal library size criterion greater than the 10 $\times$  oversampling used here. In future work, it would be useful to quantify the extent to which such bias is observed across various substrates (i.e. other than MBP).

Our random domain insertion method also generated some apparent repressors due to out-of-frame insertions. We hypothesize that these variants suppress transcription because the DNA-binding domains are still expressed due to non-specific translation initiation—the start codon of the ZFP remained in the final constructs. Therefore, it is possible that ribosomes translated the full ZFP along with the downstream portion of MBP (out of frame), leading to a functional repressor. Thus, in future implementations of the BERDI method, it may be desirable to remove the start codon from the ZFP prior to library construction to minimize this issue and further enrich for productive biosensors over these false positives. In any event, this phenomenon did not preclude our identification of functional biosensors, but rather it necessitated clonal analysis of more candidate biosensors in the final step of the selection. Should the problem of false positive repressors prove intractable, an alternative solution would be to add a FACS-based screen to enrich for ligand-responsive biosensors prior to clonal analysis.

One of our newly discovered biosensors (270 A) had a double ZFP insertion, which outperformed a similar biosensor comprising a single ZFP insertion at this position. It is possible that the presence of two ZFP domains, if correctly folded, increased repression due to the higher potential conformational shift. While this phenomenon is neither problematic or advantageous, it is inherently tied to our use of a unique restriction site within the transposon recognition sequences during library generation, such that it seems reasonable to accept this rare occurrence as a possible event that may occur during library creation. If this phenomenon were to prove problematic, library generation could be performed with a greater ratio of backbone plasmid to ZFP cassette during the applicable ligation step. Alternatively, mutagenesis of the transposon recognition sequences might reveal an alternative method that removes this possibility, although such a study is outside the scope of this investigation, and our results suggest that such a modification of the library generation method is not necessary.

Our library design strategy imposes constraints on linker sequence, which impacts which biosensors (i.e. which insertion sites) are recovered during functional screening. We hypothesized that using too long of a linker would reduce the degree to which ligand binding induces conformational changes that are translated through the protein (e.g. via an allosteric mechanism) to impair DNA binding. Conversely, using no linker may prevent the ZFP from folding in a conformation conferring DNA binding in the absence of ligand. Our use of the  $\mu$  transposon system for random insertion also restricts which types of sequences may be created as linkers upon in-

frame ZFP insertion (Supplemental Fig. S1). Given the options, we hypothesized that designing our library to include three-alanine linkers on each side of the ZFP may enable feasible protein fusions. The performance of the linker variants of the 335 P biosensor largely supported our understanding of how linker length impacts biosensor function; shortening linkers reduced biosensor-mediated repression and changed the maltose responsiveness, potentially indicating that the ZFP bound DNA to a lesser extent. The impact of linker choice on biosensor function was also highlighted by the fact that the reference biosensor (an insertion at position 316 R) was not recovered when the ZFP was flanked by three-alanine linkers. Although the reference biosensor bound DNA when a lysine–leucine linker flanked the N-terminus of the ZFP and an asparagine–valine linker flanked the C-terminus of the ZFP, when these linkers were each replaced with three-alanine linkers, promoter repression was ablated, implying that both length and composition of the linkers impact performance. The specific linkers are likely to impact every biosensor differently, and this phenomenon may also provide an additional handle for tuning biosensor performance.

Although the BERDI method generated biosensors with performance characteristics that may be suitable for use in high-throughput screening (as is evidenced by the recovery of said biosensors), there exist many opportunities for potentially improving biosensor performance by either design-driven protein engineering or directed evolution. For example, based upon our observations, systematically varying linker sequences and lengths for each candidate insertion site identified by the BERDI method may improve repression (in the ligand-free state) and/or fold-induction upon the addition of ligand. In addition, directed evolution by random mutagenesis is well-suited to improving an existing protein function (and less well-suited to creating a novel function altogether), such that this method may be applied to biosensor candidates identified by the BERDI method. Thus, BERDI may best be viewed as a method to generate candidate biosensors, constraining the protein engineering search space sufficiently to enable application-specific optimizations.

When applying the BERDI method to novel substrates, it may be useful to evaluate alternative versions of the selection algorithm. The three novel biosensors described here were all found by first sorting for inducible repressors, then evaluating individual clones for maltose responsiveness. However, if ligand-responsive biosensors prove to be exceptionally rare, it could be useful to use FACS to enrich for ligand-responsive biosensors, prior to clonal analysis, as noted above. Additionally, instead of using GFP and FACS as the screening system for ligand-responsive biosensor, GFP could be replaced with a gene conferring a survival selection (e.g. antibiotic resistance), such that growth on an antibiotic could be used as a way to enrich for rare, ligand-responsive, biosensor variants.

Looking forward, we hypothesize that the BERDI method may be generalized to convert other metabolite-binding proteins into metabolite-regulated transcription factors. Since the BERDI method enables the development of novel biosensors without relying upon prior knowledge about permissive sites within the ligand-binding protein, it may be applied to virtually any protein. However, while the procedure for generating a library of candidate biosensors is certainly generalizable, whether such an approach would yield functional biosensors remains to be evaluated, and may well vary with protein family, structure or other variables. Rigorously evaluating overall generalizability, and the limits of generalizability, will require application of the BERDI method to a wide variety of metabolite-binding proteins, to accumulate a meaningful number of both

successes and failures. Such insights would be useful for identifying applications for which BERDI may be well-suited, versus applications that would be better served by biosensors that operate via a distinct mechanism, such as Cas9 fusions leveraging domains that change conformation upon ligand binding (Oakes, Nadler, Flamholz, Fellmann, Staahl, Doudna and Savage, 2016). Ultimately, we anticipate that this approach may be extended to generate novel biosensors for a range of applications in microbiology and synthetic biology.

## Supplementary Data

Supplementary data are available at *Protein Engineering, Design & Selection* online.

## Acknowledgements

This work was supported by the National Science Foundation (MCB-1341414 to J.N.L., DGE-1 324 585 to P.Y.S.); the Environmental Protection Agency (STAR Fellowship F13A30124 to A.K.D.Y.); National Institutes of Health (TRC is supported by 1R01MH103910-01, K.T supported by 5R01MH103910-02); and Northwestern University (Undergraduate Research Grant to A.J.S.). A.K.D.Y. was supported in part by the Northwestern University Graduate School Cluster in Biotechnology, Systems, and Synthetic Biology, which is affiliated with the Biotechnology Training Program. Flow cytometry experiments were conducted at the Robert H. Lurie Flow Cytometry Core Facility. Traditional sequencing services were performed at the Northwestern University Genomics Core Facility.

## References

- Binder, S., Schendzielorz, G., Stabler, N., Krumbach, K., Hoffmann, K., Bott, M. and Eggeling, L. (2012) *Genome Biol.*, **13**, R40. doi:10.1186/gb-2012-13-5-r40.
- Brockman, I.M. and Prather, K.L. (2015) *Biotechnol. J.*, **10**, 1360–1369. doi:10.1002/biot.201400422.
- Chou, H.H. and Keasling, J.D. (2013) *Nat. Commun.*, **4**, 2595. doi:10.1038/ncomms3595.
- Collins, C.H., Arnold, F.H. and Leadbetter, J.R. (2005) *Mol. Microbiol.*, **55**, 712–723. doi:10.1111/j.1365-2958.2004.04437.x.
- Collins, C.H., Leadbetter, J.R. and Arnold, F.H. (2006) *Nat. Biotechnol.*, **24**, 708–712. doi: 10.1038/nbt1209.
- Dahl, R.H., Zhang, F., Alonso-Gutierrez, J. et al. (2013) *Nat. Biotechnol.*, **31**, 1039–1046. doi:10.1038/nbt.2689.
- Dietrich, J.A., McKee, A.E. and Keasling, J.D. (2010) *Annu. Rev. Biochem.*, **79**, 563–590. doi: 10.1146/annurev-biochem-062608-095938.
- Dietrich, J.A., Shis, D.L., Alikhani, A. and Keasling, J.D. (2013) *ACS Synth. Biol.*, **2**, 47–58. doi:10.1021/sb300091d.
- Edwards, W.R., Busse, K., Allemann, R.K. and Jones, D.D. (2008) *Nucleic Acids Res.*, **36**, e78. doi:10.1093/nar/gkn363.
- Farmer, W.R. and Liao, J.C. (2000) *Nat. Biotechnol.*, **18**, 533–537. doi:10.1038/75398.
- Feng, J., Jester, B.W., Tinberg, C.E. et al. (2015) *Elife*, **4**. doi:10.7554/eLife.10606.
- Golynskiy, M.V., Koay, M.S., Vinkenborg, J.L. and Merckx, M. (2011) *Chembiochem*, **12**, 353–361. doi:10.1002/cbic.201000642.
- Green, B., Bouchier, C., Fairhead, C., Craig, N.L. and Cormack, B.P. (2012) *Mob. DNA*, **3**, 3. doi:10.1186/1759-8753-3-3.
- Guntas, G., Mitchell, S.F. and Ostermeier, M. (2004) *Chem. Biol.*, **11**, 1483–1487. doi:10.1016/j.chembiol.2004.08.020.
- Haapa, S., Suomalainen, S., Eerikainen, S., Airaksinen, M., Paulin, L. and Savilahti, H. (1999) *Genome Res.*, **9**, 308–315.
- Kang, Z., Zhang, C., Zhang, J., Jin, P., Zhang, J., Du, G. and Chen, J. (2014) *Appl. Microbiol. Biotechnol.*, **98**, 3413–3424. doi: 10.1007/s00253-014-5569-y.
- Khalil, A.S. and Collins, J.J. (2010) *Nat. Rev. Genet.*, **11**, 367–379. doi:10.1038/nrg2775.
- Li, S., Si, T., Wang, M. and Zhao, H. (2015) *ACS Synth. Biol.*, **4**, 1308–1315. doi:10.1021/acssynbio.5b00069.
- Litcofsky, K.D., Afeyan, R.B., Krom, R.J., Khalil, A.S. and Collins, J.J. (2012) *Nat. Methods*, **9**, 1077–1080. doi:10.1038/nmeth.2205.
- Liu, D., Xiao, Y., Evans, B.S. and Zhang, F. (2015) *ACS Synth. Biol.*, **4**, 132–140. doi: 10.1021/sb400158w.
- Mehta, M.M., Liu, S. and Silberg, J.J. (2012) *Nucleic Acids Res.*, **40**, e71. doi: 10.1093/nar/gks060.
- Meinhardt, S., Manley, M.W., Jr., Becker, N.A., Hessman, J.A., Maher, L.J., 3rd and Swint-Kruse, L. (2012) *Nucleic Acids Res.*, **40**, 11139–11154. doi:10.1093/nar/gks806.
- Michener, J.K., Thodey, K., Liang, J.C. and Smolke, C.D. (2012) *Metab. Eng.*, **14**, 212–222. doi: 10.1016/j.ymben.2011.09.004.
- Mohn, W.W., Garmendia, J., Galvao, T.C. and de Lorenzo, V. (2006) *Environ. Microbiol.*, **8**, 546–555. doi:10.1111/j.1462-2920.2006.00983.x.
- Murphy, K.C., Campellone, K.G. and Poteete, A.R. (2000) *Gene*, **246**, 321–330.
- Nadler, D.C., Morgan, S.A., Flamholz, A., Kortright, K.E. and Savage, D.F. (2016) *Nat. Commun.*, **7**, 12266. doi:10.1038/ncomms12266.
- Oakes, B.L., Nadler, D.C., Flamholz, A., Fellmann, C., Staahl, B.T., Doudna, J.A. and Savage, D.F. (2016) *Nat. Biotechnol.*, **34**, 646–651. doi:10.1038/nbt.3528.
- Quiocho, F.A., Spurlino, J.C. and Rodseth, L.E. (1997) *Structure*, **5**, 997–1015.
- Rogers, J.K., Guzman, C.D., Taylor, N.D., Raman, S., Anderson, K., Church, G.M. (2015) *Nucleic Acids Res.* doi:10.1093/nar/gkv616
- Salis, H.M., Mirsky, E.A. and Voigt, C.A. (2009) *Nat. Biotechnol.*, **27**, 946–950. doi:10.1038/nbt.1568.
- Segall-Shapiro, T.H., Nguyen, P.Q., Dos Santos, E.D., Subedi, S., Judd, J., Suh, J. and Silberg, J.J. (2011) *J. Mol. Biol.*, **406**, 135–148. doi:10.1016/j.jmb.2010.11.057.
- Shis, D.L., Hussain, F., Meinhardt, S., Swint-Kruse, L. and Bennett, M.R. (2014) *ACS Synth. Biol.*, **3**, 645–651. doi:10.1021/sb500262f.
- Strianese, M., Staiano, M., Ruggiero, G., Labella, T., Pellicchia, C. and D'Auria, S. (2012) *Methods Mol. Biol.*, **875**, 193–216. doi:10.1007/978-1-61779-806-1\_9.
- Tang, S.Y. and Cirino, P.C. (2011) *Angew. Chem. Int. Ed. Engl.*, **50**, 1084–1086. doi:10.1002/anie.201006083.
- Tang, S.Y., Fazelinia, H. and Cirino, P.C. (2008) *J. Am. Chem. Soc.*, **130**, 5267–5271. doi:10.1021/ja7109053.
- Tang, S.Y., Qian, S., Akinterinwa, O., Frei, C.S., Gredell, J.A. and Cirino, P.C. (2013) *J. Am. Chem. Soc.*, **135**, 10099–10103. doi:10.1021/ja402654z.
- van Sint Fiet, S., van Beilen, J.B. and Witholt, B. (2006) *Proc. Natl. Acad. Sci. USA*, **103**, 1693–1698. doi: 10.1073/pnas.0504733102.
- Venayak, N., Anesiadis, N., Cluett, W.R. and Mahadevan, R. (2015) *Curr. Opin. Biotechnol.*, **34**, 142–152. doi:10.1016/j.copbio.2014.12.022.
- Younger, A.K., Dalvie, N.C., Rottinghaus, A.G. and Leonard, J.N. (2016) *ACS Synth. Biol.* doi:10.1021/acssynbio.6b00184.
- Zhang, F., Carothers, J.M. and Keasling, J.D. (2012) *Nat. Biotechnol.*, **30**, 354–359. doi:10.1038/nbt.2149.