ACS Synthetic Biology

Article

Chimeric LysR-type transcriptional biosensors for customising ligand specificity profiles towards flavonoids

Brecht De Paepe, Jo Maertens, Bartel Vanholme, and Marjan De Mey

ACS Synth. Biol., **Just Accepted Manuscript** • DOI: 10.1021/acssynbio.8b00326 • Publication Date (Web): 18 Dec 2018

ACS Publications

is published by the American Chemical Society. 1155 Sixteenth Street N.W., Washington, DC 20036

Chimeric LysR-type transcriptional biosensors for customising ligand specificity profiles towards flavonoids

Brecht De Paepe,[†] Jo Maertens,[†] Bartel Vanholme,[‡] and Marjan De Mey*,[†]

†*Centre for Synthetic Biology, Ghent University, Coupure Links 653, B-9000 Ghent, Belgium*

‡*Department of Plant Biotechnology and Bioinformatics, Ghent University - VIB Center for Plant Systems Biology, Technologiepark 927, 9052 Ghent, Belgium*

E-mail: marjan.demey@ugent.be

Abstract

Transcriptional biosensors enable key applications in both metabolic engineering and synthetic biology. Due to nature's immense variety of metabolites, these applications require biosensors with a ligand specificity profile customised to the researcher's needs. In this work, chimeric biosensors were created by introducing parts of a donor regulatory circuit from *Sinorhizobium meliloti*, delivering the desired luteolin-specific response, into a non-specific biosensor chassis from *Herbaspirillum seropedicae*. Two strategies were evaluated for the development of chimeric LysR-type biosensors with customised ligand specificity profiles towards three closely-related flavonoids, naringenin, apigenin and luteolin. In the first strategy, chimeric promoter regions were constructed at the biosensor effector module, while in the second strategy, chimeric transcription factors were created at the biosensor detector module. Via both strategies, the biosensor repertoire was expanded with luteolin-specific chimeric biosensors demonstrating a variety of

response curves and ligand specificity profiles. Starting from the non-specific biosensor chassis, a shift from 27.5% to 95.3% luteolin specificity was achieved with the created chimeric biosensors. Both strategies provide a compelling, faster and more accessible route for the customisation of biosensor ligand specificity, compared to *de novo* design and construction of each biosensor circuit for every desired ligand specificity.

Keywords

Transcriptional biosensors, ligand specificity engineering, flavonoids, chimeric genetic circuits, *Escherichia coli*

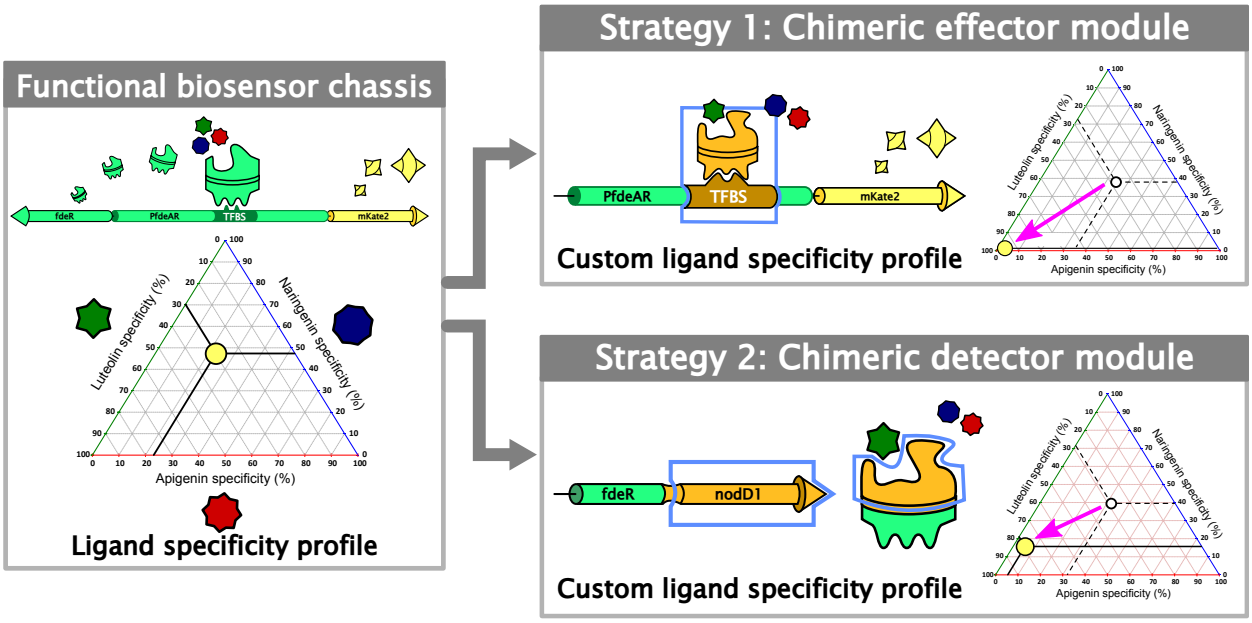


Table of contents graphic

To enable a variety of *in vivo* track and control strategies in biotechnological engineering, natural transcriptional regulatory circuits are increasingly converted into transcriptional biosensors with the desired key characteristics in terms of both the response curve and the ligand specificity profile. These characteristics are predominantly determined by the biosensor's detector and effector module (see Figure 1). The detector module consists of the transcription factor (TF) coding sequence and the corresponding (constitutive) promoter and ribosome-binding site (RBS) sequence. The effector module consists of the coding sequence for the output signal, e.g. a fluorescent protein (FP), the transcription factor binding sites (TFBSs) and corresponding promoter and RBS sequence (see Figure 1). The TF from the detector module binds the TFBS in the effector module and regulates the transcription of the output signal coding sequence of choice. Depending on this output signal, biosensors are used in applications vital for metabolic engineering and synthetic biology, such as adaptive laboratory evolution, high-throughput screening and dynamic pathway control.[1] In addition to customised response curves for these different applications, nature's immense diversity in unique small molecules implies the need for biosensors with a ligand specificity customised to the researcher's need. However, the ligand specificity profile of such a biosensor is inherent to the used natural regulatory circuit, i.e. the TF. This implies that, for every desired ligand specificity, a novel biosensor circuit should be developed, tested, modularised and optimised anew.[2] Moreover, not every natural biosensor circuit has promoter or RBS sequences compatible with the host strain of choice. These aspects greatly stall the further expansion of the currently available biosensor repertoire.

Several strategies are available to create the desired ligand specificity starting from already characterised TFs. First, protein engineering techniques, such as random and structure-based site-directed mutagenesis, enable the creation of novel TF variants with an altered ligand specificity profile. These techniques, however, tend to expand the ligand specificity profile of the TF towards the incorporation of the desired ligand, consequently, resulting in

a more promiscuous TF rather than a shift in ligand specificity.[1,3–7] Second, computational tools enable the *de novo* design of ligand-binding domains (LBDs) to create the desired ligand specificity profile. Such methods sample from an immense *in silico* mutagenic space, significantly reducing the necessary test space in the lab but require a significant amount of foreknowledge.[1,8–11] As a golden mean between random mutagenesis methods and data-driven computational methods, the creation of chimeric biosensor circuits offers a compelling alternative route for customising ligand specificity profiles.

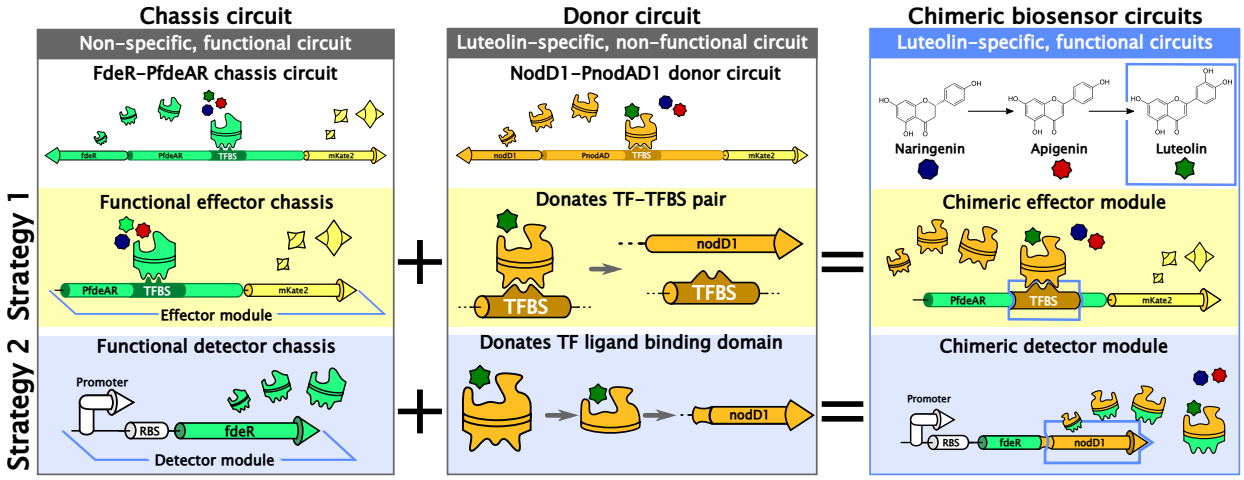


Figure 1: Schematic overview of the two strategies for the development of chimeric biosensors for custom ligand specificity profiles at either the effector module (Strategy 1) or the detector module (Strategy 2). To achieve the desired ligand specificity, parts of the luteolin-specific, *Escherichia coli*-incompatible NodD1-PnodAD1 donor circuit from *Sinorhizobium meliloti* were combined with the *E. coli*-compatible, modularised and customisable FdeR-PfdeAR chassis circuit. The boxed regions at each strategy indicate specific regulatory parts originating from a different species as the biosensor chassis and, thus, illustrate the specific chimeric nature of both biosensors. TFBS: transcription factor binding sites, RBS: ribosome-binding site.

In this context, we previously developed a functional and fully characterised naringenin-responsive biosensor circuit originating from *Herbaspirillum seropedicae* (FdeR-PfdeAR).[2,12,13] The effector and detector module of this LysR-type biosensor circuit were successfully decoupled and, subsequently, engineered to generate a collection of synthetic biosensor variants with a wide variety of response curve characteristics (see Figure 1).[2] In this work, this mod-

ularised synthetic biosensor circuit acted as a chassis circuit in which genetic parts from a different regulatory circuit were introduced to generate multiple chimeric biosensors with customised ligand specificity profiles towards distinct closely-related flavonoids.

The natural regulatory mechanism between flavonoids and rhizobial NodD TFs is one of the most well-studied ligand-TF interaction. Flavonoids comprise a large group of plant speciality metabolites with over 9000 identified compounds, demonstrating a wide variety of structures and, concurrently, biological functions.[14] For example, the central metabolite in the biosynthetic pathway of flavonoids, naringenin, which is the target of numerous efforts for microbial biosynthesis[15–25] can be converted by a two-step regiospecific hydroxylation into luteolin, with apigenin as sole intermediate (see Figure 1).[26–33]

Besides the coloration of flowers, protection against biotic and abiotic stress and auxin transport, flavonoids are essential for the symbiotic communication between leguminous plants and nitrogen-fixating rhizobia.[25,34–39] Moreover, legumes exploit the structural diversity of flavonoids to enable host specificity towards their rhizobial symbionts.[39,40] Namely, the root exudates of different legumes contain different sets of flavonoids which exclusively activate the expression of the *nod* genes of specific, compatible rhizobial partners.[41–43] The host specificity arises from the rhizobial NodD TFs which regulate the *nod* genes and are able to discriminate between specific flavonoids in these root exudates.[40,43–48] These NodD TFs belong to the large LysR-family of transcriptional regulators and have highly similar N-terminal DNA-binding domains (DBDs) but significantly diverging C-terminal ligand-binding domains (LBDs) resulting in their differences in flavonoid specificity.[45,49]

In this work, two chimera-based strategies for customising ligand specificity profiles were evaluated, focusing either on the effector module or the detector module (see Figure 1). To generate customised ligand specificity profiles, the natural diversity in specificity profiles of highly similar NodD-like TFs towards the three closely-related flavonoids, naringenin, apigenin and luteolin was exploited (see Figure 1). The NodD1-PnodAD1 circuit from *Sinorhizobium meliloti* is a LysR-type regulatory circuit (Accession NodD1: WP_010967456),

demonstrating the desired luteolin specificity as proof-of-concept, and acted as the donor of specific circuit parts for the creation of chimeric biosensor variants.[47,50] The previously developed FdeR-PfdeAR chassis circuit from *H. seropedicae* (Accession FdeR: WP_013233032) was employed as the chassis circuit in which these donor parts were introduced.[2,12,13] This proficient biosensor chassis has no distinct preference towards any of the three flavonoids.[51] First, in the chimeric effector module strategy, the concept of chimeric promoter regions was evaluated to obtain customised ligand specificity profiles within a characterised and functional biosensor framework (see Figure 1). Second, instead of redesigning an existing TF, the chimeric detector module strategy generates completely novel TFs by combining the LBDs from a donor TF, which demonstrates the desired ligand specificity, with the DBD from a TF of a fully characterised and functional biosensor chassis circuit[1,52] (see Figure 1).

Results

Ligand specificity mapping of the natural and synthetic biosensor circuits

To evaluate the functionality and ligand specificity profile of the natural biosensor circuit, the pNatNodD1 plasmid was designed and constructed (see Supplementary Table 3), based on the reported sequences of the luteolin-specific LysR-type NodD1-PnodAD1 regulatory circuit from *S. meliloti*. This biosensor construct consists of the complete natural bidirectional intergenic promoter region (PnodAD1), the adjacent TF coding sequence (*nodD1*) and the FP coding sequence (*mKate2*, see Figure 2a).[47,50,53] The fluorescent mKate2 protein, the output signal of choice, has a fast maturation time and a bright, far-red fluorescent signal which limits background interference due to autofluorescence of *E. coli*.[54] To define ligand specificity profiles, the three closely-related flavonoids naringenin, apigenin and luteolin were supplied separately in concentrations ranging from 0 to 100 mg/L to the *E. coli* strain containing the pNatNodD1 biosensor. No significant response was observed for any of the

three flavonoid molecules across the full concentration range (one-way ANOVA: $F = 0.170$, 0.149 and 0.253 with p-values $= 0.998$, 0.999 and $0.992 > 0.05$ for naringenin, apigenin and luteolin, respectively, see Supplementary Table 2). Consequently, the regulatory circuit NodD1-PnodAD1 in this natural configuration did not lead to any biosensor functionality in *E. coli*. This non-functional circuit was combined with the previously developed, functional naringenin-responsive NodD-like biosensor circuit, FdeR-PfdeAR, to generate multiple functional chimeric biosensor variants with the desired luteolin specificity.[2]

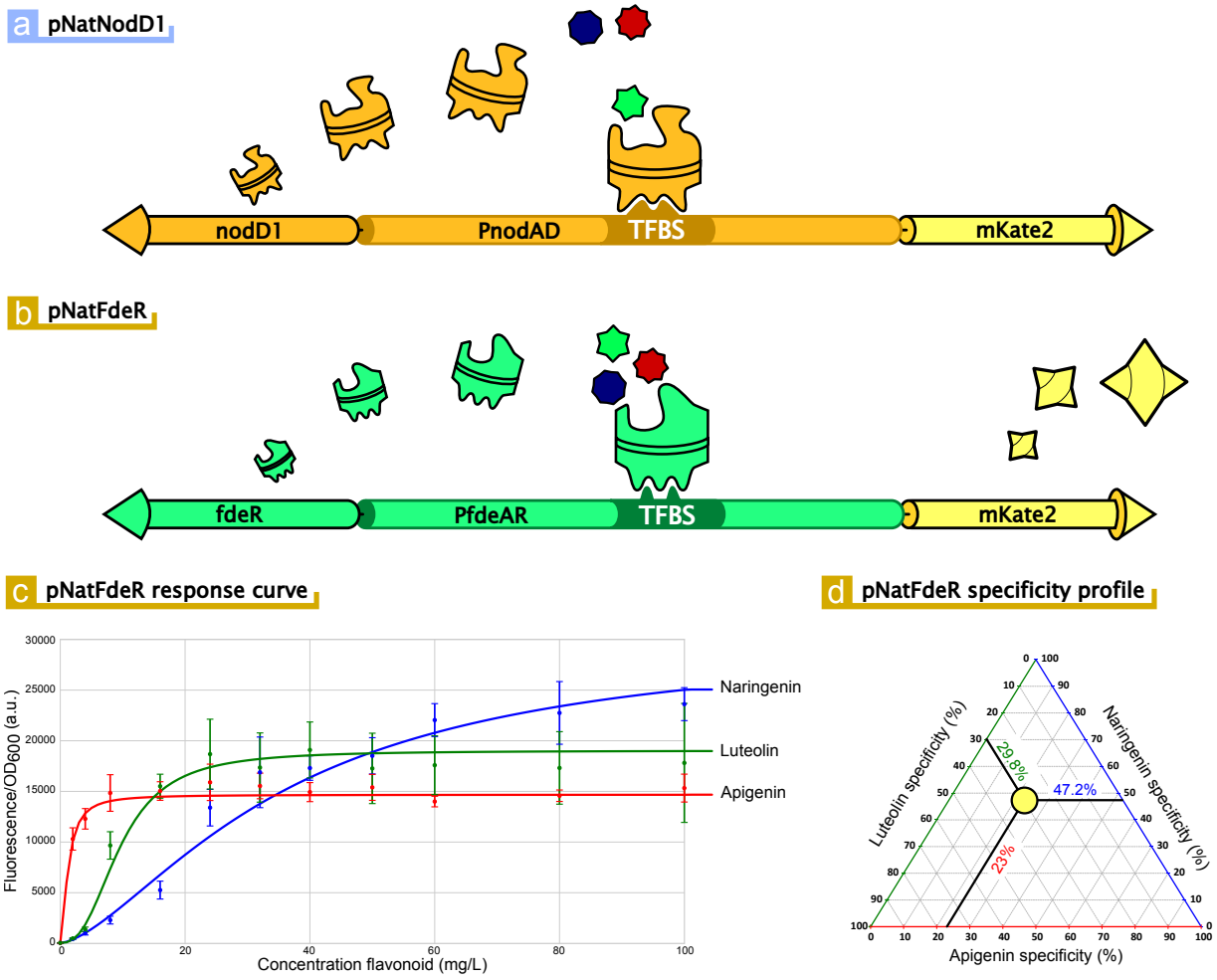


Figure 2: Schematic representation of the transcriptional biosensor circuits of **(a)** pNatNodD1 and **(b)** pNatFdeR in their naturally occurring architecture. **(c)** Naringenin-, apigenin- and luteolin-response curves (error bars represent standard errors) and fitted Hill functions for the functional pNatFdeR biosensor plasmid. **(d)** Triangular graph representing the ligand specificity profile of pNatFdeR for naringenin, apigenin and luteolin. TFBS: transcription factor binding sites.

The pNatFdeR biosensor plasmid consists of the NodD-like FdeR-PfdeA regulatory circuit from *H. seropedicae* in its natural configuration and has an identical bidirectional biosensor architecture as the pNatNodD1 plasmid (see Figure 2b).[2] Previously, this biosensor was proven to be a proficient naringenin-responsive biosensor.[2] Concentrations ranging from 0 to 100 mg/L of naringenin, apigenin and luteolin were supplied to map the responsiveness and flavonoid specificity profile. As observed in Figure 2c and d, the pNatFdeR circuit demonstrates biosensor functionality towards all three flavonoids. The maximum fluorescent response levels for naringenin, apigenin and luteolin amount up to levels of 30167 ± 2989 a.u., 14675 ± 340 a.u. and 19025 ± 1357 a.u., respectively. This translates into a ligand specificity profile of 47.2%, 23% and 29.8% for naringenin, apigenin and luteolin, respectively (see Figure 2c). In addition, distinct differences in response curve shape between each of the three flavonoids are observed. More specifically, the K-values were predicted at 35.9 ± 5.9, 1.25 ± 0.5 and 9.6 ± 1 mg/L for naringenin, apigenin and luteolin, respectively.

In the following sections, two strategies are evaluated for the development of chimeric biosensors with custom flavonoid specificity by combining the functionality of the FdeR-PfdeAR circuit (the chassis circuit) with the luteolin specificity of the NodD1-PnodAD1 circuit (the donor circuit). In the first strategy, chimeric effector modules (promoter regions) were generated. In the second strategy, chimeric detector modules (TFs) were created. To facilitate independent and unambiguous control over both modules in each strategy, synthetic biosensor variants of both pNatNodD1 and pNatFdeR were created by decoupling these modules. The modularisation of pNatFdeR was already performed and evaluated.[2] The resulting pSynFdeR plasmid has a clearly demarcated and independent detector and effector module which enables the extensive customisation of its response curve.[2] The bidirectional intergenic promoter region, PfdeAR, normally controls the expression of both the *fdeR* and *mKate2* coding sequence. However, the corresponding promoter and RBS sequence of both modules overlap and would hinder any engineering efforts specifically targeted towards either one of the modules. Therefore, in this synthetic biosensor architecture, the expression of the

fdeR coding sequence is now controlled by a constitutive, synthetic P22 promoter and RBS sequence,[55] instead of the naturally present promoter and RBS sequence in the PfdeAR intergenic region, and is decoupled from the expression level of *mKate2* by the introduction of a terminator and spacer sequence (see Figure 3b). Similarly, the pSynNodD1 biosensor circuit was constructed with an identical architecture as pSynFdeR (see Figure 3a). In this manner, NodD1 expression is controlled independent from the PnodAD1 intergenic region, by the P22 promoter and RBS sequence which are fully functional in *E. coli*.

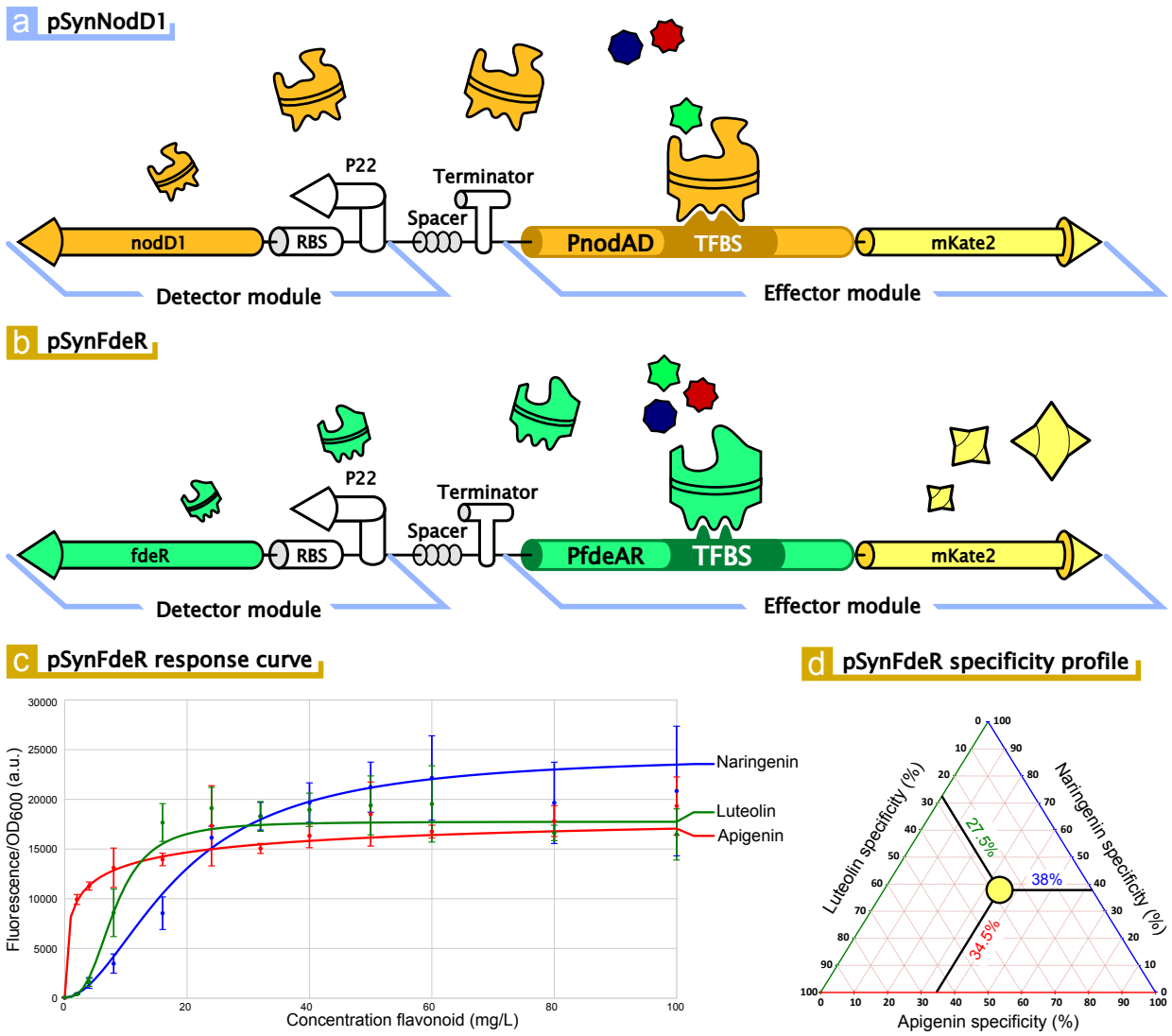


Figure 3: Schematic representation of the transcriptional biosensor circuits of **(a)** pSynNodD1 and **(b)** pSynFdeR with their modularised architecture resulting in an independent detector and effector module. **(c)** Naringenin-, apigenin- and luteolin-response curves (error bars represent standard errors) and fitted Hill functions for the functional pSynFdeR biosensor plasmid. **(d)** Triangular graph representing the ligand specificity profile of pSynFdeR for naringenin, apigenin and luteolin. TFBS: transcription factor binding sites, RBS: ribosome-binding site.

Similar to their biosensor counterparts with the natural architecture (i.e. pNatFdeR and pNatNodD1), the pSynFdeR biosensor plasmid demonstrates definitive biosensor functionality without a clear-cut preference in flavonoid specificity (see Figure 3c and d). Also, the pSynNodD1 biosensor plasmid does not significantly respond to any of the three flavonoids

(one-way ANOVA: $F = 1.622$, 0.342 and 0.395 with p-values = 0.128, 0.971 and $0.952 > 0.05$ for naringenin, apigenin and luteolin, respectively, see Supplementary Table 2). The ligand specificity profile of the pSynFdeR biosensor chassis towards naringenin, apigenin and luteolin corresponds to 38% ($M = 24504 \pm 2528$ a.u.), 34.5% ($M = 22308 \pm 8237$ a.u.) and 27.5% ($M = 17756 \pm 598$ a.u.), respectively (see Figure 3d). Similar to pNatFdeR, the synthetic biosensor chassis demonstrates no distinct flavonoid preference in its ligand specificity profile. In addition, comparable response curve shapes were observed for each of the three flavonoids (see Figure 3c).

Strategy 1: Customised ligand specificity profile through chimeric effector modules

To explore different strategies for customising ligand specificity profiles, the potential of chimeric detector-effector pairs was evaluated. Here, the pSynNodD1 detector module was combined with the pSynFdeR effector module presuming that NodD1 has the ability to bind the TFBS of PfdeAR and, subsequently, regulate the expression of the *mKate2* coding sequence. This is substantiated by the high similarity of the 47 basepair (bp) TFBS sequences of FdeR and NodD1 is observed (60%, see Figure 4b), in conjunction with the high amino acid similarity of the DBD of FdeR and NodD1 (61% DBD identity, 41% for the complete amino acid sequences, see Supplementary Figure 2).[56,57] In addition, both TFBSs share the common 25, 5 and 7 bp *nod* boxes and the pair of palindromic NodD consensus sequences, [AT-N_{10}-GAT]-N_7-[ATC-N_{10}-AT] (see Figure 4b).[53,56,58] The chimeric biosensor variant, pChimNodD1-PfdeAR, was created by replacing the *fdeR* coding sequence in pSynFdeR with the *nodD1* coding sequence, thus generating a chimeric TF-TFBS pair (see Figure 4a). No significant response was detected for any of three tested flavonoids (one-way ANOVA: $F = 0.656$, 0.067 and 0.163 with p-values = 0.773, 0.999 and $0.999 > 0.05$ for naringenin, apigenin and luteolin, respectively, see Supplementary Table 2).

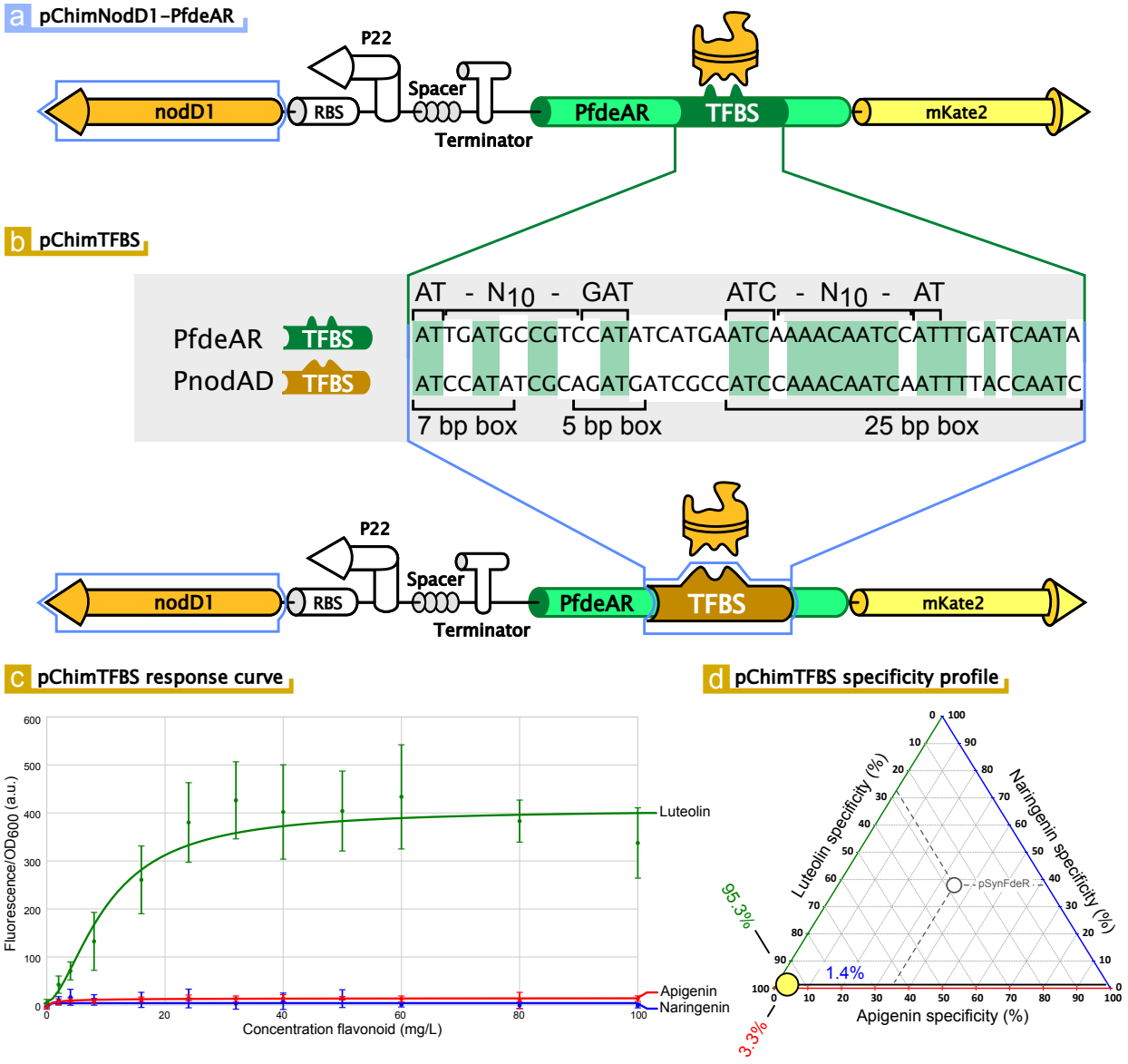


Figure 4: Schematic representation of the chimeric transcriptional biosensor circuits of **(a)** pChimNodD1-PfdeAR with its chimeric detector-effector pair consisting of the promoter region, PfdeAR, from *Herbaspirillum seropedicae* in the effector module and the *nodD1* coding sequence from *Sinorhizobium meliloti* in the detector module and **(b)** pChimTFBS with its chimeric promoter region in the effector module consisting of PfdeAR from *H. seropedicae* but with the TFBS (transcription factor binding site) of NodD1 from *S. meliloti*. **(c)** Naringenin-, apigenin- and luteolin-response curves (error bars represent standard errors) and fitted Hill functions for the functional pChimTFBS biosensor plasmid. **(d)** Triangular graph representing the ligand specificity profile of pChimTFBS (yellow circle) and pSynFdeR (white circle) for naringenin, apigenin and luteolin. RBS: ribosome-binding site.

Subsequently, the chimeric effector strategy was evaluated in which the FdeR binding

site in the PfdeAR promoter region of pChimNodD1-PfdeA was replaced with the NodD1 binding site, resulting in the pChimTFBS biosensor plasmid (see Figure 4b). In this context, the PfdeAR promoter region from the chassis circuit delivers the necessary functional *E. coli*-compatible core promoter and RBS sequence, which control the expression of the *mKate2* coding sequence. The introduction of the correct NodD1 binding sites enables NodD1 to bind and regulate the chimeric effector module. By using the PfdeAR promoter region as chassis, the critical distances between TFBS, adjacent regions and core promoter sequences remain identical to the functional FdeR-PfdeAR circuit.[1,58–64] The fluorescent response of the resulting pChimTFBS biosensor circuit was characterised for changes in concentration of the three tested flavonoids (see Figure 4c and d).

This novel chimeric biosensor exhibits distinct biosensor functionality in conjunction with a stringent luteolin specificity. The observed response curves and the corresponding fitted Hill functions are shown in Figure 4c, together with the ligand specificity profile of this chimeric biosensor in Figure 4d. The pChimTFBS biosensor has a ligand specificity profile of 1.4% ($M = 6 \pm 10$ a.u.), 3.3% ($M = 14 \pm 14$ a.u.) and 95.3% ($M = 407 \pm 45$ a.u.) for narignenin, apigenin and luteolin, respectively. In addition, the fluorescent response to naringenin and apigenin was not significantly altered by increasing the flavonoid concentration (one-way ANOVA: $F = 0.267$, 0.459 and 6.003 with p-values $= 0.989 > 0.05$, $0.919 > 0.05$ and $4.72 \times 10^{-6} < 0.05$ for naringenin, apigenin and luteolin, respectively, see Supplementary Table 2). This biosensor construct demonstrates stringent luteolin specificity and is the first reported luteolin-specific transcriptional biosensor in *E. coli* (see Figure 4d). Besides this shift in ligand specificity, a clear reduction in maximum response (M) is observed in comparison with the pSynFdeR biosensor chassis. Namely, pSynFdeR demonstrated maximum fluorescent response levels of circa 20000 a.u. for all three flavonoids, while pChimTFBS only demonstrated a maximum luteolin response level of 407 ± 45 a.u. Finally, in comparison with the pSynFdeR biosensor chassis, the leaky expression of the luteolin response of pChimTFBS is substantially lower. The predicted leaky expression of pChimTFBS corre-

sponds to 5.52 ± 7.15 a.u., in contrast to the pSynFdeR a-values of 73.6 ± 20.3 a.u. (see Supplementary Table 1).

Strategy 2: Customised ligand specificity profiles through chimeric detector modules

As the second strategy for customising ligand specificity profiles, the creation of chimeric detector modules was evaluated. Chimeric LysR-type TFs were created by combining the LBD of NodD1, delivering the desired luteolin specificity, with the DBD of FdeR, enabling these chimeric TFs to regulate the expression of *mKate2* within the established pSynFdeR chassis circuit (see Figure 5a, 6a and 7a). The LBD and DBD are linked by a hinge domain which acts as conformational signal transducer upon ligand binding. In addition, such a hinge domain has an indirect influence on TF dimerisation, TFBS affinity and TFBS specificity.[1,4,57,61,65–71] In a previous study, the effect of hinge length variations on NodD properties from *Rhizobium leguminosarum* bv. *viciae* (Accession NodD: WP_018068348) was investigated.[57] Here, it was observed that the NodD hinge domain does not interact with DNA or ligand molecules but that a wide range of regulatory phenotypes could be generated simply by altering the length of this hinge domain.

Chimeric NodD1-FdeR TFs with differing hinge domains were created to evaluate the use of chimeric TFs for customised flavonoid specificity and to assess the effects of the hinge domain on the overall biosensor performance. More specifically, two chimeric TFs, ChimTF-FdeR and ChimTF-NodD1, were created with the hinge domain originating from either FdeR (see Figure 5a) or NodD1 (see Figure 6a), respectively. First, the hinge domain of NodD1 from *S. meliloti* (85'IAWDPLNPAQSD) was determined based on the reported amino acid sequence of the hinge domain of NodD from *Rhizobium leguminosarum* bv. *viciae* (85'IAWDPINPAESD, 83% hinge domain identity, see Supplementary Figure 1).[57] Second, the newly identified NodD1 hinge domain was used to pinpoint the FdeR hinge domain (84'IAALPAFVPAEST, 41% identity for the full amino acid sequences, see Supplementary

Figure 2). These identified hinge domains are clearly located at the end of the helix-turn-helix DBD, quintessential of the LysR-family of TFs (see Supplementary Figure 1 and 2).[62] Finally, during the construction of the ChimTF-FdeR, a third chimeric TF variant, dubbed ChimTF-FdeRMut, was picked up with a mutation at the FdeR-based hinge domain (see Figure 7a). In this mutated hinge domain (84'IAALPAFVNLAESTR), the conformationally rigid proline residue at position 92 is substituted with a polar, uncharged asparagine residue (P92N). In addition, this mutated hinge domain is elongated by the insertion of a hydrophobic leucine residue adjacent to the P92N substitution (see Figure 7a).

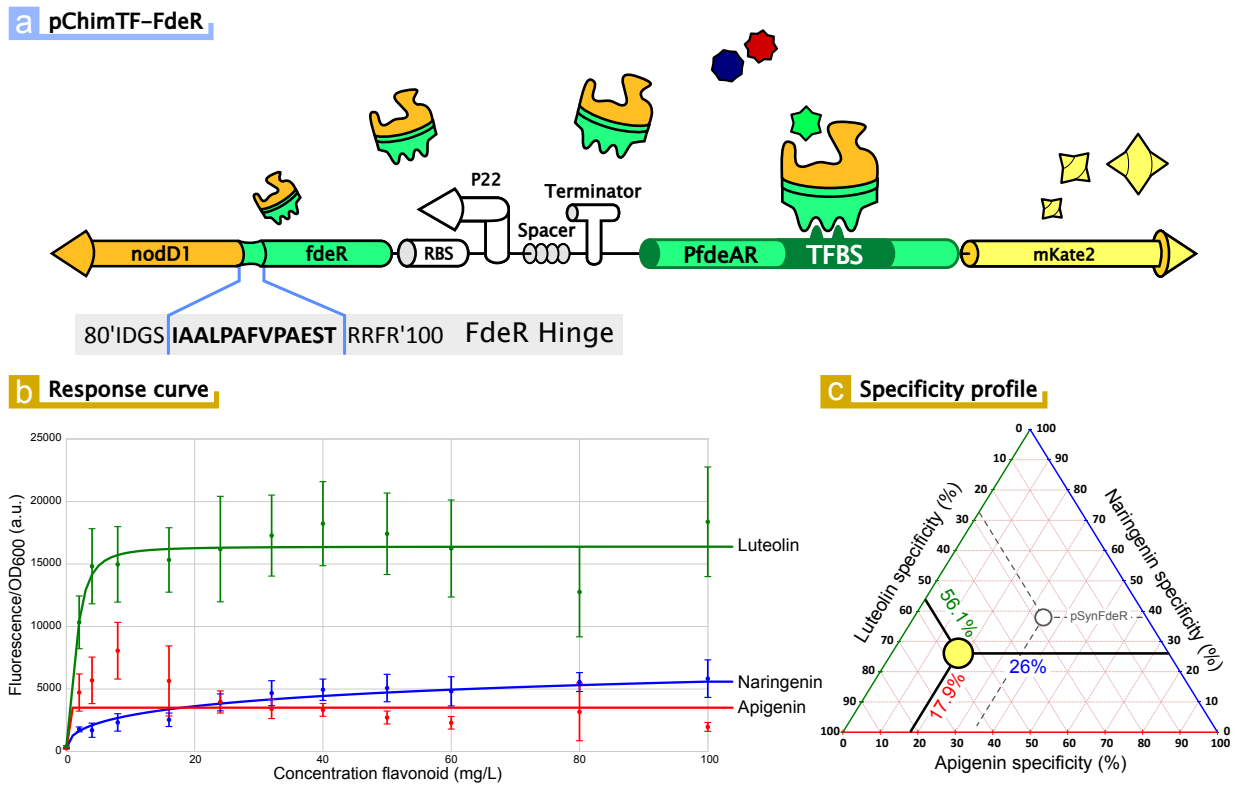


Figure 5: **(a)** Schematic representation of the chimeric transcriptional biosensor circuit of pChimTF-FdeR with its chimeric transcription factor consisting of the DNA-binding domain and hinge domain of FdeR from *Herbaspirillum seropedicae* and the ligand-binding domain of NodD1 from *Sinorhizobium meliloti*. **(b)** Naringenin-, apigenin- and luteolin-response curves (error bars represent standard errors) and fitted Hill functions for the functional pChimTF-FdeR biosensor plasmid. **(c)** Triangular graph representing the ligand specificity profile of pChimTF-FdeR (yellow circle) and pSynFdeR (white circle) for naringenin, apigenin and luteolin. TFBS: transcription factor binding sites, RBS: ribosome-binding site.

The pChimTF-FdeR biosensor circuit demonstrated biosensor functionality upon induction by all three flavonoids (naringenin, apigenin and luteolin), in addition to an increased luteolin specificity in comparison to pSynFdeR (see Figure 5b and c). With a ligand specificity profile of 26.0% ($M = 7617 \pm 5133$ a.u.), 17.9% ($M = 5232 \pm 4322$ a.u.) and 56.1% ($M = 16394 \pm 1364$ a.u.) for naringenin, apigenin and luteolin, respectively, the stringent luteolin specificity of the natural NodD1 TF is only partially transferred to the chimeric biosensor (see Figure 5c). Further, the maximum luteolin response level of pChimTF-FdeR ($M = 16394 \pm 1364$ a.u.) is in the same order of magnitude as the maximum response levels of the pSynFdeR biosensor, containing the natural *fdeR* coding sequence (circa 20000 a.u., see Figure 3a). In contrast, the maximum luteolin response level of the pChimTFBS biosensor, containing the natural *nodD1* coding sequence, corresponds to 407 ± 45 a.u. (see Figure 4a). In addition, the leaky expression observed in pChimTF-FdeR corresponds to 406 ± 66.5 a.u. for luteolin which is higher than for pSynFdeR (73.6 ± 20.3 a.u., see Supplementary Table 1). Finally, in comparison to both pSynFdeR and pChimTFBS, the pChimTF-FdeR biosensor displays a lower K-value of 1.52 ± 0.73 mg/L.

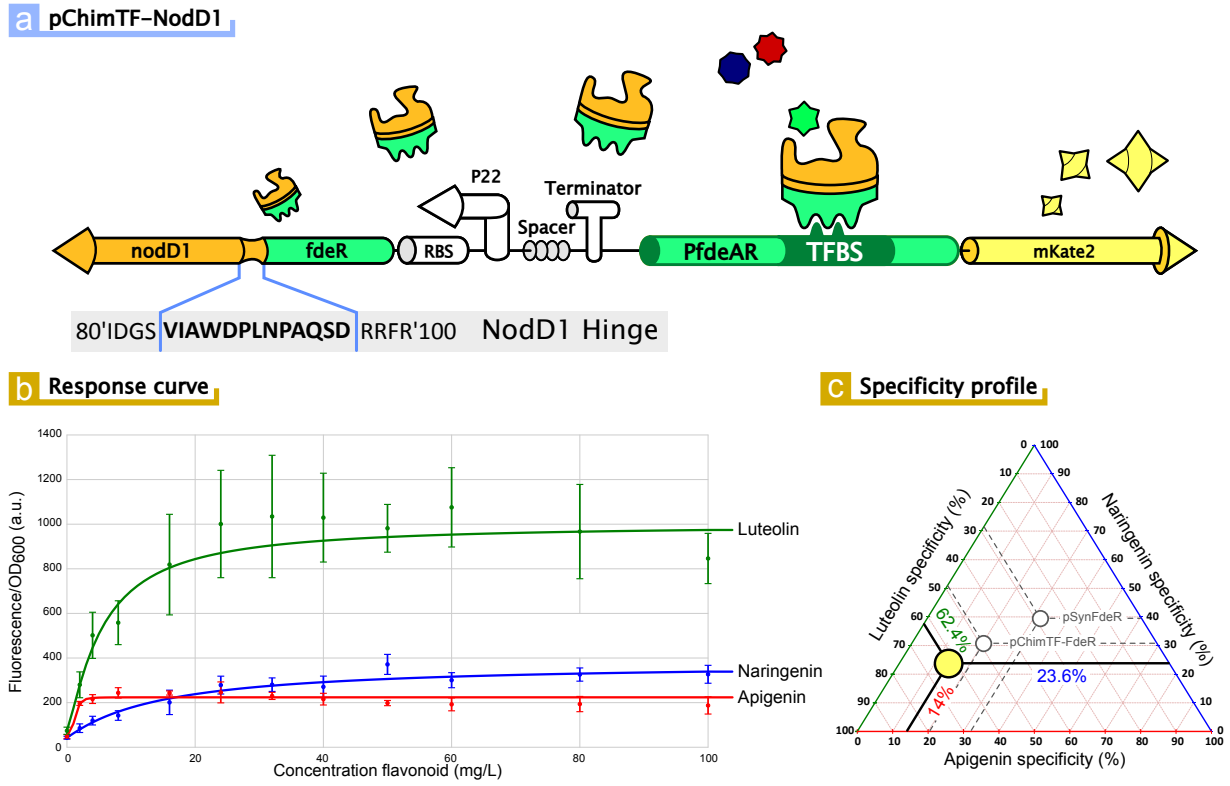


Figure 6: **(a)** Schematic representation of the chimeric transcriptional biosensor circuit of pChimTF-NodD1 with its chimeric transcription factor consisting of the DNA-binding domain of FdeR from *Herbaspirillum seropedicae* and the ligand-binding domain and hinge domain of NodD1 from *Sinorhizobium meliloti*. **(b)** Naringenin-, apigenin- and luteolin-response curves (error bars represent standard errors) and fitted Hill functions for the functional pChimTF-NodD1 biosensor plasmid. **(c)** Triangular graph representing the ligand specificity profile of pChimTF-NodD1 (yellow circle), pSynFdeR and pChimTF-FdeR (white circles) for naringenin, apigenin and luteolin. TFBS: transcription factor binding sites, RBS: ribosome-binding site.

The pChimTF-NodD1 chimeric biosensor, containing ChimTF-NodD1 with the NodD1 hinge domain, also demonstrates biosensor functionality towards the three flavonoids (see Figure 6b). This validates that the hinge domain, in addition to the LBD (see pChimTF-FdeR), can be altered to create chimeric TFs without loss of function and with differing characteristics. Similar to pChimTF-FdeR, an increase in luteolin specificity is observed in comparison to pSynFdeR by changing the LBD and hinge domain in the biosensor chassis (from 27.5% to 62.4% in pSynFdeR and pChimTF-NodD1, respectively, see Figure 6c). On the other hand, also the stringent luteolin specificity of NodD1, observed in pChimTFBS

(95.3%), was not achieved (23.6%, 14% and 62.4% naringenin, apigenin and luteolin specificity, respectively, in pChimTF-NodD1). Further, two pChimTF-NodD1 characteristics are distinctly different from pChimTF-FdeR, namely the decrease in maximum luteolin response level and the increase in K-value for luteolin response. First, instead of a maximum luteolin response level of 16394 ± 1364 a.u. as observed with pChimTF-FdeR (see Figure 5b), the pChimTF-NodD1 biosensor reaches a maximum response of 998 ± 105 a.u. (see Figure 6b). Second, the luteolin response curve of pChimTF-NodD1 displays a higher K-value (5.5 ± 1.9 mg/L) than that of pChimTF-FdeR ($K = 1.52 \pm 0.73$ mg/L). With the only difference between these two biosensor circuits being the hinge domain, the hinge domain amino acid sequence clearly affects the response characteristics. Finally, pChimTF-NodD1 can be considered as a hybrid between pChimTFBS (containing the natural NodD1 TF) and pChimTF-FdeR (containing only the NodD1 LBD). In this context, it is remarkable that the K-value for luteolin ($K = 5.5 \pm 1.9$ mg/L), the luteolin specificity (62.4%), maximum luteolin response ($M = 998 \pm 105$ a.u.) and leaky expression (74.1 ± 16.1 a.u.) of pChimTF-NodD1 all correspond to values between the values determined for pChimTF-FdeR and pChimTFBS (see Supplementary Table 1).

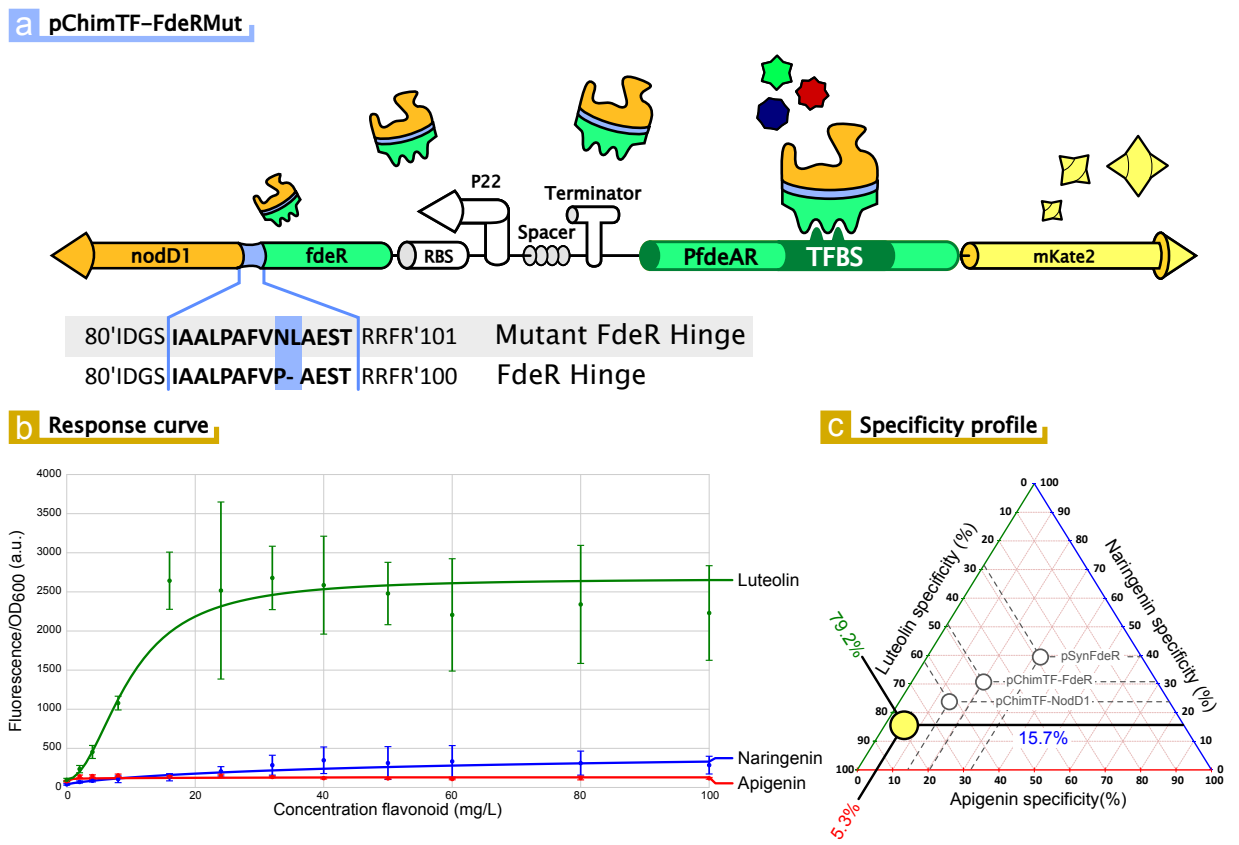


Figure 7: **(a)** Schematic representation of the chimeric transcriptional biosensor circuit of pChimTF-NodD1 with its chimeric transcription factor consisting of the DNA-binding domain and mutated hinge domain of FdeR from *Herbaspirillum seropedicae* and the ligand-binding domain of NodD1 from *Sinorhizobium meliloti*. **(b)** Naringenin-, apigenin- and luteolin-response curves (error bars represent standard errors) and fitted Hill functions for the functional pChimTF-FdeRMut biosensor plasmid. **(c)** Triangular graph representing the ligand specificity profile of pChimTF-FdeRMut (yellow circle), pSynFdeR, pChimTF-FdeR and pChimTF-NodD1 (white circles) for naringenin, apigenin and luteolin. TFBS: transcription factor binding sites, RBS: ribosome-binding site.

The pChimTF-FdeRMut biosensor circuit is composed of the chimeric NodD1-FdeR TF with the mutated FdeR hinge domain to regulate the PfdeAR promoter region (see Figure 7a). Despite significant alterations in this hinge domain and the potential effects on the conformation of the TF and LBD-DBD signal transduction, this circuit demonstrated biosensor functionality in response to the three flavonoids (see Figure 7b). This biosensor exhibited naringenin, apigenin and luteolin specificities of 15.8%, 5.1% and 79.2%, respectively. An increase in luteolin specificity was observed in comparison to the biosensor chassis, pSynFdeR,

as well as the previously discussed pChimTF-FdeR and pChimTF-NodD1 chimeric biosensors (see Figure 7b and c). On the other hand, this mutated hinge domain resulted in a biosensor circuit with a maximum luteolin response level, M, of 2673 ± 263 a.u., circa 6 times lower as observed in pChimTF-FdeR but circa 2.7 times higher as observed in pChimTF-NodD1. Further, the leaky expression of this biosensor corresponds to values similar to those observed in the pSynFdeR and pChimTF-NodD1 biosensor circuits (102 ± 18.6 a.u. for luteolin, see Supplementary Table 1). Finally, the luteolin response curve of pChimTF-FdeRMut displays an even higher K-value (9.79 ± 1.33 mg/L) than observed in pChimTF-NodD1 and pChimTF-FdeR.

Material and methods

Strains and growth conditions

E. coli TOP10 cells (Invitrogen, Carlsbad, U.S.A.) were used for plasmid construction and growth experiment purposes. Unless otherwise stated, all products were purchased from Sigma-Aldrich (Diegem, Belgium). For plasmid construction, strains were grown in lysogeny broth (LB) at 37 °C with shaking. LB was composed of 1% tryptone peptone (Difco, Erembodegem, Belgium), 0.5% yeast extract (Difco), 1% sodium chloride (VWR, Leuven, Belgium) and 25 μg/mL chloramphenicol as antibiotic. LB agar plates contained the same components as LB with the addition of 1% agar. For *in vivo* fluorescence experiments, MOPS EZ Rich Defined medium was used (Teknova, Hollister, Canada) with 25 μg/mL chloramphenicol and 2% glucose as carbon source.

Plasmid construction

All plasmids used in this work were constructed using Circular Polymerase Extension Cloning (CPEC) assembly and are listed in Table 1.[72] DNA oligonucleotides were purchased from IDT (Leuven, Belgium) and DNA sequences of every constructed plasmid were verified using se-

quencing services (Macrogen Inc., Amsterdam, The Netherlands). All plasmids are low-copy vectors with a pSC101 origin of replication[73] and a chloramphenicol resistance marker.[74] All plasmid nucleotide sequences are listed in Supplementary Tables 3 – 7. All biosensor plasmids express a TF which regulates the expression of the reporter gene, *mKate2*.[54] Plasmid pBlank is an empty pSC101Cm plasmid as background reference. Plasmid pRef is a pSC101Cm plasmid which constitutively expresses the *mKate2* coding sequence as a reference plasmid (see Supplementary Table 7).

In vivo fluorescence experiments

Strains, as 6 biological replicates (n = 6 for each concentration of flavonoid), were inoculated in 150 μL LB and grown overnight on a Compact Digital Microplate Shaker (Thermo Scientific) at 800 rpm and 37 °C. Subsequently, these cultures were 1:200 diluted in 150 μL of fresh MOPS EZ Rich Defined medium containing concentrations ranging from 0 to 100 mg/L (0, 2, 4, 8, 16, 24, 32, 40, 50, 60, 80 and 100 mg/L) of either naringenin, apigenin or luteolin and, subsequently, grown on a Compact Digital Microplate Shaker for 24 h at 800 rpm and 37 °C. Finally, fluorescence and optical density were measured using a Tecan Infinite 200 Pro. For measuring mKate2 fluorescence excitation and emission wavelengths were set at 588 nm and 633 nm, respectively. Optical density was measured at a wavelength of 600 nm.

Data processing and statistical analysis

For fluorescence measurements, MOPS EZ Rich Defined medium without cell culture was used to correct for background fluorescence and optical density of the medium (FP_{med} and OD_{med}, respectively) for each imposed ligand concentration. To account for background fluorescence and optical density of the cell culture, *E. coli* TOP10 cells containing pBlank, an empty pSC101Cm plasmid, were used (FP_{pBlank} and OD_{pBlank}, respectively). Thus, the correction of the fluorescent output signal, normalised for optical density, was calculated as

follows:

$$\left(\frac{FP}{OD}\right)_{cor} = \frac{FP - FP_{med}}{OD - OD_{med}} - \frac{FP_{pBlank} - FP_{med}}{OD_{pBlank} - OD_{med}}$$

for every biosensor strain, with 6 biological replicates and each imposed ligand concentration. To account for interplate variability, *E. coli* TOP10 cells containing the pRef plasmid were used as a reference strain which constitutively expresses the *mKate2* coding sequence (see Supplementary Table 7).

The resulting mean $\left(\frac{\overline{FP}}{\overline{OD}}\right)_{cor}$ value (n = 6), and corresponding standard error, for each concentration of the flavonoid ligand molecules is expressed in absolute units (a.u.) and are fitted with the following Hill function using the weighted non-linear least squares algorithm (SciPy, curve_fit, Levenberg-Marquardt algorithm):[2,75–78]

$$\left(\frac{\overline{FP}}{\overline{OD}}\right)_{cor} = f(C) = a + k\left(\frac{C^n}{C^n + K^n}\right)$$

with

C = the concentration of flavonoid ligand (mg/L)

a = the basal normalised fluorescent signal (leaky expression, a.u.)

k = the maximum normalised fluorescent signal relative to a (a.u.)

$M = a + k$ = the maximum normalised fluorescent signal (a.u.)

n = the Hill coefficient (cooperativity, sigmoid character)

K = the Hill constant (TF-ligand affinity, mg/L)

For each of the estimated parameters, at least five different initial guesses were chosen across different orders of magnitude which gave 200 different combinations resulting in highly similar

parameter estimates. In addition, every algorithm was run at least three times for each biosensor response curve fitting and compared. Standard errors on parameter estimation were determined by calculating the square root of the variance of each parameter from the covariance matrix.

The ligand specificity profiles, within the reference frame of the discussed three flavonoids, were defined and calculated based on the maximum normalized fluorescent output signal, M, as this parameter is one of the most important parameters in biosensor applications in synthetic biology and metabolic engineering and enables a more definite discrepancy between functional and non-functional biosensors. In addition, as leaky expression (a) is generally low for the discussed biosensor variants, M also reflects the dynamic range in the context of this research article. More specifically, the ligand specificity of a biosensor for one of the three flavonoids was defined as the ratio of M of that flavonoid to the sum of M-values of each of the three flavonoids, calculated as follows:

$$\mathrm{Spec}_{Nar} = \left(\frac{M_{Nar}}{M_{Nar} + M_{Api} + M_{Lut}} \right)$$

whereby $\mathrm{Spec}_{Nar} + \mathrm{Spec}_{Api} + \mathrm{Spec}_{Lut} = 100\%$ and with Nar: Naringenin, Api: Apigenin and Lut: Luteolin.

In contrast to the specificity definition used in the field of analytic biosensor devices for medical diagnostics (specificity relates to false positive detection), this definition of ligand specificity focuses on the extent of TF-ligand interaction and the resulting transcription initiation strength, in analogy with specificity definitions in enzyme catalysis, gene regulation and molecular biology.[79] This response-based ligand specificity definition allows for a straight-forward visualization and comparison of the maximum response of a biosensor variant for a specific flavonoid relative to the response this biosensor demonstrates for the other flavonoids.

Table 1: Plasmids constructed, characterised and examined in this work with the corresponding transcription factor (TF), regulatory promoter region and transcription factor binding sites (TFBS). Chimeric DNA sequences are depicted in bold. FdeR-m-'NodD1 represents the TF with the DNA-binding domain of FdeR and ligand-binding domain of NodD1, connected by a mutated version of the FdeR-hinge sequence. N.A.: not applicable.

Plasmid name	**Plasmid structure**	**TF**	**Regulatory promoter region**	**TFBS**
pNatFdeR	pSC101Cm-*fdeR*-PfdeAR-*mKate2*	FdeR	*PfdeAR*	*fdeR*
pNatNodD1	pSC101Cm-*nodD1*-PnodAD1-*mKate2*	NodD1	*PnodAD1*	*nodD1*
pSynFdeR	pSC101Cm-P22-*fdeR*-T-PfdeA(R)-*mKate2*	FdeR	*PfdeA(R)*	*fdeR*
pSynNodD1	pSC101Cm-P22-*nodD1*-T-PnodA(D1)-*mKate2*	NodD1	*PnodA(D1)*	*nodD1*
pChimNodD1-PfdeA	pSC101Cm-P22-*nodD1*-T-PnodA(D)-*mKate2*	NodD1	*PfdeA(R)*	*fdeR*
pChimTFBS-NodD1	pSC101Cm-P22-*nodD1*-T-ChimPfdeA(R)-*mKate2*	NodD1	***ChimPfdeA(R)/PnodD1***	***nodD1***
pChimTF-NodD1	pSC101Cm-P22-*fdeR'-nodD1*-T-PfdeA(R)-*mKate2*	**FdeR'-NodD1**	*PfdeA(R)*	*fdeR*
pChimTF-FdeR	pSC101Cm-P22-*fdeR-'nodD1*-T-PfdeA(R)-*mKate2*	**FdeR-'NodD1**	*PfdeA(R)*	*fdeR*
pChimTF-FdeRMut	pSC101Cm-P22-*fdeR-m-nodD1*-T-PfdeA(R)-*mKate2*	**FdeR-m-'NodD1**	*PfdeA(R)*	*fdeR*
pBlank	pSC101Cm	N.A.	N.A.	N.A.
pRef	pSC101Cm-ProB-BBa_B0032-*mKate2*	N.A.	N.A.	N.A.

Sequence alignments

The structure-based amino acid sequence alignments of the TFs were constructed with the online PROMALS3D tool (`http://prodata.swmed.edu/promals3d`, see Supplementary Figure 1 and 2).[80] The nucleotide sequence alignment of the TFBS and promoter regions were constructed using the online MUSCLE tool (MUltiple Sequence Comparison by Log- Expectation, `http://www.ebi.ac.uk/Tools/msa/muscle/`, see Figure 4b).[81]

Discussion

The use of transcriptional biosensors as *in vivo* track and control tools is sparking a much needed surge in high-throughput applications for metabolic engineering and synthetic biology. However, the lack of general biosensor engineering principles prevents the currently limited biosensor repertoire from expanding towards a collection of custom-built metabolite-specific biosensors. In this work, the feasibility and effectiveness of two strategies for customising biosensor ligand specificity profiles through the development of chimeric transcriptional biosensor circuits was demonstrated. Both proposed strategies enabled the creation of multiple chimeric biosensor circuits demonstrating the desired traits. In the first strategy, functioning chimeric effector modules were created without altering the detector module. In the second strategy, functioning chimeric detector modules were created without altering the effector module. To enable the independent engineering of these biosensor modules and to exploit an otherwise *E. coli*-incompatible regulatory circuit, a synthetic modularised biosensor was successfully used as a chassis circuit for the creation of these chimeric circuits. As proof-of-concept, the ligand specificity profiles towards the three closely-related flavonoids, naringenin, apigenin and luteolin of the non-specific biosensor chassis were customised to acquire stringent luteolin-specific biosensors instead.

The desired stringent luteolin specificity was delivered by the LysR-type NodD1-PnodAD1 regulatory circuit from *S. meliloti*. On the other hand, the desired functionality, *E. coli*-

compatibility and customisability was delivered by the LysR-type FdeR-PfdeAR biosensor circuit from *H. seropedicae*. As a starting point, these two transcriptional regulatory circuits were converted into biosensor circuits both in their naturally occurring genetic configuration (pNatNodD1 and pNatFdeR) and in the modularised synthetic architecture (pSynNodD1 and pSynFdeR). These four biosensors were characterised for ligand specificity towards the three flavonoids (naringenin, apigenin and luteolin). Both the natural and synthetic NodD1-PnodAD1-based circuits did not demonstrate any response towards these flavonoids. This lack of biosensor functionality is likely the result of inefficient or absent recognition of the *S. meliloti* intergenic bidirectional promoter region by the *E. coli* transcriptional or translational machinery.[82] Because in pSynNodD1 the expression of the TF is assured by a synthetic *E. coli* promoter and RBS sequence, the lack of functionality of both circuits can not solely be attributed to the lack of TF expression. On the other hand, pNatFdeR and pSynFdeR both demonstrated biosensor functionality albeit without a preferred flavonoid specificity. For both FdeR-PfdeAR-based biosensors, the differences in K-values for each flavonoid suggest differences in affinity of the TF (FdeR) towards these flavonoids. For naringenin, a higher K-value was observed in comparison to apigenin and luteolin. Subsequently, pSynFdeR biosensor was used as the chassis circuit to receive specific parts of the NodD1-PnodAD1 regulatory circuit to generate multiple chimeric biosensors.

An important aspect specific to the LysR-type transcriptional regulatory circuits discussed in this research article, is their underlying dual activation-repression mechanism which complicates the generation of clear-cut mechanistical hypotheses and conclusions based on the different Hill parameters, e.g. cooperativity.[61,62] More specifically, the TFs are constantly bound to their TFBS (as multimer) and generate a bend in the DNA strand, independent of their ligand molecule presence. This results in repression and, concurrently, the low leaky expression levels which are observed for all developed biosensor variants. If ligand molecules are present and bound to the TFs, the DNA bend is relaxed, while the TFs remain bound, upon which RNAP can initiate transcription.[61,62] To gain more information on this under-

lying mechanism, biophysical experminents would be imperative.

As a starting point for customised biosensor specificity profiles, the pChimNodD1-PfdeAR biosensor was created by combining the complete NodD1-PnodAD1 detector module with the complete FdeR-PfdeAR effector module. Besides the high similarity of the TFBS nucleotide sequence as well as the TF amino acid sequence, this strategy is substantiated by the reported ability of certain NodD TFs to regulate the *nod* operon promoter regions (PnodAD) in other rhizobia and thus transfer their host specificity profiles.[40,47,83] However, no significant flavonoid response was observed with pChimNodD1-PfdeAR (see Supplementary Table 2). This indicates that, despite the significant similarity of the donor and chassis circuit modules, NodD1 does not have the ability to interact with the PfdeAR promoter region in a similar manner as FdeR and, therefore, can not regulate transcription of the *mKate2* coding sequence in response to changes in flavonoid concentration.

In the first strategy, pChimTFBS was created from pChimNodD1-PfdeAR by replacing the natural TFBS in PfdeAR with the NodD1 binding sites, thus creating a chimeric effector module. By giving NodD1 its own TFBS within the chassis of the functioning pSynFdeR biosensor chassis, the resulting chimeric pChimTFBS biosensor demonstrated distinct biosensor functionality. Moreover, the luteolin specificity shifted from 27.5% with pSynFdeR to 95.3% with pChimTFBS. This indicates that the stringent luteolin specificity of the natural NodD1-PnodAD1 circuit was transferred into the pSynFdeR chassis. However, the maximum luteolin response was circa 50 times lower as for pSynFdeR. This discrepancy could be the result of differences in recruitment of RNA polymerase (RNAP) by the TF, NodD1, in comparison to FdeR. Another cause could be differences in the regulatory DNA-bending mechanism, typical of LysR-type circuits, as the result of the non-natural chimeric promoter region and, therefore indirectly, RNAP recruitment.[1,47] On the other hand, leaky expression was clearly reduced in comparison to pSynFdeR which also could be the result of reduced RNAP recruitment.

In the second strategy, chimeric detector modules were created, i.e. chimeric TFs, for

customising ligand specificity profiles. The LBD of NodD1 was combined with the DBD of FdeR within the pSynFdeR biosensor chassis. The hinge domains from both FdeR and NodD1 were evaluated, resulting in pChimTF-FdeR and pChimTF-NodD1, respectively. In addition, a mutated variant of the FdeR hinge domain was identified and evaluated, resulting in pChimTF-FdeRMut. All three chimeric TFs were able to regulate the PfdeAR promoter region and, therefore, generated biosensor functionality in their respective circuits. In addition, all three chimeric TFs led to a shift in ligand specificity towards luteolin in comparison to pSynFdeR, albeit not as stringent as observed in pChimTFBS (56.1%, 62.4%, 79.2%, 95.3% and 27.5% for pChimTF-FdeR, pChimTF-NodD1, pChimTF-FdeRMut, pChimTFBS and pSynFdeR, respectively). This indicates that the luteolin specificity of NodD1, is either only partially determined by the used NodD1 LBD, is reduced by the presence of the non-natural DBD (and hinge domain) of FdeR. Interestingly, distinct differences were observed between the response curves of the three chimeric biosensor circuits.

The maximum luteolin response level of pChimTF-FdeR is almost identical to that of pSynFdeR which highlights the advantage of using pSynFdeR as a chassis circuit. However, the leaky expression of pChimTF-FdeR is higher than that of pSynFdeR which indicates that changing the LBD has an influence on the repressive capacity of this TF. This phenomenon could be the result of altered dimerisation or RNAP recruitment in the absence of flavonoids.[61] Notably, LysR-type TFs display a dual regulatory mechanism, either repressing or activating transcription in either the absence or presence of the correct ligand molecule.[62] In contrast, the leaky expression of pChimTF-NodD1 is similar to that of pSynFdeR which hints to the importance of pairing up the NodD1 LBD with its natural hinge domain. Further, the luteolin specificity, the maximum luteolin response and K-value for luteolin response of pChimTF-NodD1 all correspond to values between those observed with pChimTFBS and pChimTF-FdeR. This suggests that incorporating the accompanying hinge domain of the NodD1 LBD, results in biosensor characteristics tending towards those of the pChimTFBS biosensor, comprising the natural NodD1 TF, and away from more FdeR-like chimeric cir-

cuits, i.e. pChimTF-FdeR. Finally, pChimTF-FdeRMut demonstrated the highest luteolin specificity of the three chimeric TF circuits, has a maximum response level circa 2.7 times higher as pChimTF-NodD1 and has a higher K-value, more similar to pChimTFBS. The mutated hinge domain seems to unlock the potential for higher luteolin specificity from within the NodD1 LBD and higher maximum luteolin response from within the FdeR DBD. This could be attributed to the reduction of potentially imperfect chimeric LBD-DBD interactions at the protein interface as a result of the longer and likely more flexible hinge domain (84'IAALPAFVNLAESTR instead of 84'IAALPAFVPAESTR) as proline is known to be a rigid amino acid residue. Further, the longer hinge domain could aid in exposing activating sites which are otherwise less available to RNAP in this chimeric conformation.[57] Because the only difference between these three chimeric circuits lies in the hinge domain, it is evident that the hinge domain has an influence on both the response curve characteristics as well as on the ligand specificity profile. This substantiates the fact that the TF hinge domain is a compelling target for various biosensor engineering efforts to further fine-tune both the ligand specificity and the response curve. Computational protein modelling tools could aid in elaborating on the contribution of the hinge domain in inter-domain contacts and the general underlying mechanisms of transcriptional regulation.

In this study, several chimera-based engineering strategies were evaluated to successfully alter the ligand specificity of a LysR-type biosensor chassis circuit in *E. coli*. However, for this NodD1-based proof of concept, the discussed results do not exclude the possibility of competitive interference between ligands in the context of ligand mixtures, as was demonstrated by Peck *et al.* (2006). Despite the fact that only luteolin is capable of inducing nod gene transcription, these authors demonstrated, on the ÂţM-scale, that non-inducing flavonoids can act as competitive inhibitors for luteolin-specific induction of NodD1-based transcription regulation in *S. meliloti*.[47] Therefore, for this specific case-study, this competitive inhibition may have an influence on the applicability of the developed biosensors for specific in vivo applications. To this end, additional research is imperative to fully map the

multi-dimensional response landscape of the different biosensors for different combinations of ligands and concentrations but will result in combinatorial explosion in terms of experimental set-up.

This is the first contribution of LysR-type chimeric biosensors to the biosensor repertoire which, until now, only contained chimeras from the LacI- and XylR-family.[1,67,68,71,84,85] LysR-type regulators comprise the largest known family of prokaryotic TFs and are able to detect a wide variety of molecules such as arginine, acetic acid, malonate, salicylates, chitobiose and flavonoids.[62,86] Therefore, the proposed strategies could also be used to generate various other chimeric LysR-type biosensors for different ligand molecules, whether or not within the FdeR-PfdeAR chassis used in this work. Moreover, in a recent study, Skjoedt *et al.* demonstrated that the FdeR-PfdeAR circuit can also be adopted as a naringenin biosensor in the model eukaryotic organism *Saccharomyces cerevisiae*.[87] This suggests that the discussed strategies are likely to be transferable to yeast and, potentially to other eukaryotes, for the development of customised LysR-based biosensors.

Besides establishing two distinct engineering strategies, the developed synthetic biosensors are useful contributions to the biosensor repertoire and are the first reported luteolin-specific biosensors. Because of their variable biosensor characteristics and the availability of response curve engineering principles,[2] these biosensors could be applied in various synthetic biology and metabolic engineering strategies, e.g. for the development of microbial cell factories (MCF) for luteolin biosynthesis.

Acknowledgements

Brecht De Paepe holds a Ph.D. Grant from the Institute for the Promotion of Innovation through Science and Technology in Flanders (IWT-Vlaanderen). This research was also supported by the project "CondEx" (FWO-G.0321.13 N) and by the BOF-IOP project (Bijzonder Onderzoeksfonds-Interdisciplinair Onderzoeksproject), "MLSB"(BOF16/IOP/040).

Author contributions

Brecht De Paepe designed and performed all experiments, analysed the data and wrote the manuscript. All authors were involved in the conception and design of this work and in drafting, discussing and correcting the manuscript. All authors revised the manuscript critically.

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: ...

Table S1: Overview of Hill parameter values for all relevant biosensor constructs discussed in this study; Table S2: Overview of the statistical analysis values (F- and corresponding p-values) for the significance testing of the effect of flavonoid supplementation on specific biosensor response for all relevant biosensor constructs discussed in this study; Tables S3-S7: Annotated nucleotide sequences of pNatNodD1, pSynNodD1, pChimNodD1-PfdeAR, pChimTFBS and pRef; Table S8: Amino acid sequences of the discussed natural TFs used in the pNatFdeR, pNatNodD1, pSynFdeR, pSynNodD1, pChimNodD1-PfdeAR and pChimTFBS biosensors; Table S9: Amino acid sequences of the newly created chimeric TFs used in the pChimTF-FdeR, pChimTF-NodD1 and pChimTF-FdeRMut chimeric biosensors; Figure S1: Amino acid sequence alignment of the NodD and NodD1 protein sequences; Figure S2: Amino acid sequence alignment of the FdeR and NodD1 protein sequences.

References

(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.

(2) De Paepe, B., Maertens, J., Vanholme, B., and De Mey, M. (2018) Modularization and Response Curve Engineering of a Naringenin-Responsive Transcriptional Biosensor. *ACS Synth. Biol. 7*, 1303–1314.

(3) Aharoni, A., Gaidukov, L., Khersonsky, O., Gould, S. M., Roodveldt, C., and Tawfik, D. S. (2004) The 'evolvability' of promiscuous protein functions. *Nat. Genet. 37*, 73.

(4) Galvão, T. C., Mencía, M., and De Lorenzo, V. (2007) Emergence of novel functions in transcriptional regulators by regression to stem protein types. *Mol. Microbiol. 65*, 907–919.

(5) Jensen, R. A. (1976) Enzyme recruitment in evolution of new function. *Annu. Rev. Microbiol. 30*, 409–425.

(6) de Lorenzo, V., Pérez-Martín, J., Lorenzo, V., and Pérez-Martín, J. (1996) Regulatory noise in prokaryotic promoters: how bacteria learn to respond to novel environmental signals. *Mol. Microbiol 19*, 1177–84.

(7) McIver, J., Djordjevic, M. A., Weinman, J. J., Bender, G. L., and Rolfe, B. G. (1989) Extension of host range of *Rhizobium leguminosarum* bv. *trifolii* caused by point mutations in *nodD* that result in alterations in regulatory function and recognition of inducer molecules. *Mol. Plant-Microbe Interact. 2*, 97–106.

(8) 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.

(9) 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) Dietrich, J. a., McKee, A. E., and Keasling, J. D. (2010) High-throughput metabolic engineering: advances in small-molecule screening and selection. *Annu. Rev. Biochem. 79*, 563–90.

(11) 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 83*, 1327–1340.

(12) Marin, A. M., Souza, E. M., Pedrosa, F. O., Souza, L. M., Sassaki, G. L., Baura, V. A., Yates, M. G., Wassem, R., and Monteiro, R. A. (2013) Naringenin degradation by the endophytic diazotroph *Herbaspirillum seropedicae* SmR1. *Microbiology 159*, 167–75.

(13) Siedler, S., Stahlhut, S. G., Malla, S., Maury, J., and Neves, A. R. (2014) Novel biosensors based on flavonoid-responsive transcriptional regulators introduced into *Escherichia coli*. *Metab. Eng. 21*, 2–8.

(14) Falcone Ferreyra, M. L., Rius, S. P., and Casati, P. (2012) Flavonoids: biosynthesis, biological functions, and biotechnological applications. *Front. Plant Sci. 3*, 222.

(15) Forkmann, G., and Martens, S. (2001) Metabolic engineering and applications of flavonoids. *Curr. Opin. Biotechnol. 12*, 155–160.

(16) Wang, Y., Chen, S., and Yu, O. (2011) Metabolic engineering of flavonoids in plants and microorganisms. *Appl. Microbiol. Biotechnol. 91*, 949–956.

(17) Santos, C. N. S., Koffas, M., and Stephanopoulos, G. (2011) Optimization of a heterologous pathway for the production of flavonoids from glucose. *Metab. Eng. 13*, 392–400.

(18) Trantas, E. A., Koffas, M. A. G., Xu, P., and Ververidis, F. (2015) When plants produce not enough or at all: metabolic engineering of flavonoids in microbial hosts. *Front. Plant Sci. 6*, 7.

(19) Leonard, E., Yan, Y., Fowler, Z. L., Li, Z., Lim, C.-G., Lim, K.-H., and Koffas, M. A. G. (2008) Strain improvement of recombinant *Escherichia coli* for efficient production of plant flavonoids. *Mol. Pharm. 5*, 257–65.

(20) Kaneko, M., Hwang, E., Ohnishi, Y., and Horinouchi, S. (2003) Heterologous production of flavanones in *Escherichia coli*: potential for combinatorial biosynthesis of flavonoids in bacteria. *J. Ind. Microbiol. Biotechnol. 30*, 456–461.

(21) Wu, J., Zhou, T., Du, G., Zhou, J., and Chen, J. (2014) Modular optimization of heterologous pathways for de novo synthesis of (2S)-naringenin in *Escherichia coli*. *PloS ONE 9*, e101492.

(22) Wu, J., Du, G., Zhou, J., and Chen, J. (2014) Systems metabolic engineering of microorganisms to achieve large-scale production of flavonoid scaffolds. *J. Biotechnol. 188C*, 72–80.

(23) Wang, Z., and Cirino, P. C. (2016) New and improved tools and methods for enhanced biosynthesis of natural products in microorganisms. *Curr Opin. Biotechnol. 42*, 159–168.

(24) Xu, P., Bhan, N., and Koffas, M. A. (2013) Engineering plant metabolism into microbes: from systems biology to synthetic biology. *Curr. Opin. Biotechnol. 24*, 291–299.

(25) Delmulle, T., De Maeseneire, S. L., and De Mey, M. (2017) Challenges in the microbial production of flavonoids. *Phytochem. Rev.* 1–19.

(26) López-Lázaro, M. (2009) Distribution and biological activities of the flavonoid luteolin. *Mini Rev. Med. Chem. 9*, 31–59.

(27) Seelinger, G., Merfort, I., and Schempp, C. M. (2008) Anti-oxidant, anti-inflammatory and anti-allergic activities of luteolin. *Planta Medica 74*, 1667–77.

(28) Theoharides, T. (2009) Luteolin as a therapeutic option for multiple sclerosis. *J. Neuroinflammation 6*, 29.

(29) Brown, J. E., and Rice-Evans, C. A. (1998) Luteolin-rich artichoke extract protects low density lipoprotein from oxidation in vitro. *Free Radic. Res. 29*, 247–255.

(30) Byun, S., Lee, K. W., Jung, S. K., Lee, E. J., Hwang, M. K., Lim, S. H., Bode, A. M., Lee, H. J., and Dong, Z. (2010) Luteolin inhibits protein kinase Cϵ and c-Src activities and UVB-induced skin cancer. *Cancer Res. 70*, 2415–2423.

(31) Jang, S., Kelley, K. W., and Johnson, R. W. (2008) Luteolin reduces IL-6 production in microglia by inhibiting JNK phosphorylation and activation of AP-1. *Proc. Natl. Acad. Sci. U.S.A. 105*, 7534–7539.

(32) Zhao, G., Qin, G. W., Wang, J., Chu, W. J., and Guo, L. H. (2010) Functional activation of monoamine transporters by luteolin and apigenin isolated from the fruit of *Perilla frutescens* (L.) Britt. *Neurochem. Int. 56*, 168–176.

(33) Yu, M.-C., Chen, J.-H., Lai, C.-Y., Han, C.-Y., and Ko, W.-C. (2010) Luteolin, a non-selective competitive inhibitor of phosphodiesterases 1âĂŞ5, displaced [3H]-rolipram from high-affinity rolipram binding sites and reversed xylazine/ketamine-induced anesthesia. *Eur. J. Pharmacol. 627*, 269–275.

(34) Demain, A. L., and Fang, A. (2000) The natural functions of secondary metabolites. *Adv. Biochem. Engin./Biotechnol. 69*, 1–39.

(35) Bradshaw, H. D., and Schemske, D. W. (2003) Allele substitution at a flower colour locus produces a pollinator shift in monkeyflowers. *Nature 426*, 176–178.

(36) Mol, J., Grotewold, E., and Koes, R. (1998) How genes paint flowers and seeds. *Trends Plant Sci. 3*, 212–217.

(37) Winkel-Shirley, B. (2002) Biosynthesis of flavonoids and effects of stress. *Curr. Opin. Plant Biol. 5*, 218–223.

(38) Stapleton, A. E., and Walbot, V. (1994) Flavonoids can protect maize DNA from the induction of ultraviolet radiation damage. *Plant Physiol. 105*, 881–9.

(39) Liu, C.-W., and Murray, J. (2016) The role of flavonoids in nodulation host-range specificity: An update. *Plants 5*, 33.

(40) Wang, D., Yang, S., Tang, F., and Zhu, H. (2012) Symbiosis specificity in the legume - rhizobial mutualism. *Cell. Microbiol. 14*, 334–342.

(41) Zaat, S. A., Wijffelman, C. A., Mulders, I. H., van Brussel, A. A., and Lugtenberg, B. J. (1988) Root exudates of various host plants of *Rhizobium leguminosarum* contain different sets of inducers of *Rhizobium* nodulation genes. *Plant Phys. 86*, 1298–303.

(42) Göttfert, M. (1993) Regulation and function of rhizobial nodulation genes. *FEMS Microbiol. Lett. 104*, 39–63.

(43) Spaink, H. P., Wijffelman, C. A., Pees, E., Okker, R. J. H., and Lugtenberg, B. J. J. (1987) *Rhizobium* nodulation gene *nodD* as a determinant of host specificity. *Nature 328*, 337–340.

(44) Rossen, L., Davis, E. O., and Johnston, A. W. (1987) Plant-induced expression of *Rhizobium* genes involved in host specificity and early stages of nodulation. *Trends Biochem. Sci. 12*, 430–433.

(45) Schell, M. A., and Sukordhaman, M. (1989) Evidence that the transcription activator encoded by the *Pseudomonas putida nahR* gene is evolutionarily related to the transcription activators encoded by the *Rhizobium nodD* genes. *J. Bacteriol. 171*, 1952–9.

(46) Horvath, B., Bachem, C. W., Schell, J., and Kondorosi, A. (1987) Host-specific regulation of nodulation genes in *Rhizobium* is mediated by a plant-signal, interacting with the *nodD* gene product. *EMBO J. 6*, 841–8.

(47) Peck, M. C., Fisher, R. F., and Long, S. R. (2006) Diverse flavonoids stimulate NodD1 binding to nod gene promoters in *Sinorhizobium meliloti*. *J. Bacteriol. 188*, 5417–5427.

(48) Györgypal, Z., Iyer, N., and Kondorosi, A. (1988) Three regulatory *nodD* alleles of diverged flavonoid-specificity are involved in host-dependent nodulation by *Rhizobium meliloti*. *Mol. Gen. Genet. 212*, 85–92.

(49) Schlaman, H. R. M., Okker, R. J. H., and Lugtenberg, B. J. J. (1992) Regulation of nodulation gene expression by *nodD* in rhizobia. *J. Bacteriol. 174*, 5177–5182.

(50) Peters, N. K., Frost, J. W., and Long, S. R. (1986) A plant flavone, luteolin, induces expression of *Rhizobium meliloti* nodulation genes. *Science 233*, 977–80.

(51) Wassem, R., Marin, A. M., Daddaoua, A., Monteiro, R. A., Chubatsu, L. S., Ramos, J., Deakin, W. J., Broughton, W. J., Pedrosa, F. O., and Souza, E. M. (2017) A NodD-like protein activates transcription of genes involved with naringenin degradation in a flavonoid-dependent manner in *Herbaspirillum seropedicae*. *Environ. Microbiol. 19*, 1030–1040.

(52) Raman, S., Taylor, N., Genuth, N., Fields, S., and Church, G. M. (2014) Engineering allostery. *Trends Genet. 30*, 521–528.

(53) Fisher, R. F., Egelhoff, T. T., Mulligan, J. T., and Long, S. R. (1988) Specific binding of proteins from *Rhizobium meliloti* cell-free extracts containing NodD to DNA sequences upstream of inducible nodulation genes. *Genes Dev. 2*, 282–293.

(54) Shcherbo, D., Murphy, C. S., Ermakova, G. V., Solovieva, E. A., Chepurnykh, T. V., Shcheglov, A. S., Verkhusha, V. V., Pletnev, V. Z., Hazelwood, K. L., Roche, P. M.,

Lukyanov, S., Zaraisky, A. G., Davidson, M. W., and Chudakov, D. M. (2009) Far-red fluorescent tags for protein imaging in living tissues. *Biochem. J. 418*, 567–574.

(55) De Mey, M., Maertens, J., Lequeux, G. J., Soetaert, W. K., and Vandamme, E. J. (2007) Construction and model-based analysis of a promoter library for *E. coli*: an indispensable tool for metabolic engineering. *BMC Biotechnol. 7*, 34.

(56) Suominen, L., Paulin, L., Saano, A., Saren, A. M., Tas, E., and Lindström, K. (1999) Identification of nodulation promoter (*nod*-box) regions of *Rhizobium galegae*. *FEMS Microbiol. Lett. 177*, 217–23.

(57) Hou, B., Li, F., Yang, X., and Hong, G. (2009) The properties of NodD were affected by mere variation in length within its hinge region. *Acta Biochim. Biophys. Sin. 41*, 963–71.

(58) Feng, J., Li, Q., Hu, H.-L., Chen, X.-C., and Hong, G.-F. (2003) Inactivation of the *nod* box distal half-site allows tetrameric NodD to activate *nodA* transcription in an inducer-independent manner. *Nucleic Acids Res. 31*, 3143–56.

(59) Dunn, T. M., Hahn, S., Ogden, S., and Schleif, R. F. (1984) An operator at -280 base pairs that is required for repression of *araBAD* operon promoter: addition of DNA helical turns between the operator and promoter cyclically hinders repression. *Proc. Natl. Acad. Sci. U.S.A. 81*, 5017–20.

(60) Barnard, A., Wolfe, A., and Busby, S. (2004) Regulation at complex bacterial promoters: how bacteria use different promoter organizations to produce different regulatory outcomes. *Curr. Opin. Microbiol. 7*, 102–108.

(61) Chen, X. C., Feng, J., Hou, B. H., Li, F. Q., Li, Q., and Hong, G. F. (2005) Modulating DNA bending affects NodD-mediated transcriptional control in *Rhizobium leguminosarum*. *Nucleic Acids Res. 33*, 2540–2548.

(62) Maddocks, S. E., and Oyston, P. C. F. (2008) Structure and function of the LysR-type transcriptional regulator (LTTR) family proteins. *Microbiology 154*, 3609–23.

(63) van Hijum, S. A. F. T., Medema, M. H., and Kuipers, O. P. (2009) Mechanisms and evolution of control logic in prokaryotic transcriptional regulation. *Microbiol. Mol. Biol. Rev. 73*, 481–509, Table of Contents.

(64) Bond, L. M., Peters, J. P., Becker, N. A., Kahn, J. D., and Maher, L. J. (2010) Gene repression by minimal *lac* loops in vivo. *Nucleic Acids Res. 38*, 8072–82.

(65) 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.

(66) 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.

(67) Meinhardt, S., and Swint-Kruse, L. (2008) Experimental identification of specificity determinants in the domain linker of a LacI/GalR protein: Bioinformatics-based predictions generate true positives and false negatives. *Proteins 73*, 941–957.

(68) Meinhardt, S., Manley, M. W., Becker, N. A., Hessman, J. A., Maher, L. J., and Swint-Kruse, L. (2012) Novel insights from hybrid LacI/GalR proteins: Family-wide functional attributes and biologically significant variation in transcription repression. *Nucleic Acids Res. 40*, 11139–11154.

(69) Swint-Kruse, L., Zhan, H., Fairbanks, B. M., Maheshwari, A., and Matthews, K. S. (2003) Perturbation from a distance: Mutations that alter LacI function through long-range effects. *Biochemistry 42*, 14004–14016.

(70) Zhan, H., Taraban, M., Trewhella, J., and Swint-Kruse, L. (2008) Subdividing repressor function: DNA binding affinity, selectivity, and allostery can be altered by amino acid

substitution of nonconserved residues in a LacI/GalR homologue. *Biochemistry 47*, 8058–8069.

(71) Tungtur, S., Egan, S. M., and Swint-Kruse, L. (2007) Functional consequences of exchanging domains between LacI and PurR are mediated by the intervening linker sequence. *Proteins 68*, 375–388.

(72) Quan, J., and Tian, J. (2009) Circular polymerase extension cloning of complex gene libraries and pathways. *PloS ONE 4*, e6441.

(73) Kazuo, Y., and Mitsuyo, Y. (1984) The replication origin of pSC101: the nucleotide sequence and replication functions of the ori region. *Gene 29*, 211–219.

(74) Alton, N. K., and Vapnek, D. (1979) Nucleotide sequence analysis of the chloramphenicol resistance transposon Tn9. *Nature 282*, 864–9.

(75) Hill, A. V. (1913) The combinations of haemoglobin with oxygen and with carbon monoxide. I. *Biochem. J. 7*, 471–480.

(76) Wang, B., Barahona, M., and Buck, M. (2013) A modular cell-based biosensor using engineered genetic logic circuits to detect and integrate multiple environmental signals. *Biosens. Bioelectron. 40*, 368–76.

(77) 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.

(78) Alon, U. *An introduction to systems biology : design principles of biological circuits*; Chapman & Hall/CRC, 2007; p 301.

(79) Eaton, B. E., Gold, L., and Zichi, D. A. (1995) Let's get specific: the relationship between specificity and affinity. *Chemistry and Biology 2*, 633–638.

(80) Pei, J., Kim, B.-H., and Grishin, N. V. (2008) PROMALS3D: a tool for multiple protein sequence and structure alignments. *Nucleic Acids Res. 36*, 2295–2300.

(81) Edgar, R. C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res. 32*, 1792–1797.

(82) Fisher, R. F., Brierley, H. L., Mulligan, J. T., and Long, S. R. (1987) Transcription of *Rhizobium meliloti* nodulation genes. Identification of a *nodD* transcription initiation site in vitro and in vivo. *J. Biol. Chem. 262*, 6849–55.

(83) Bender, G. L., Nayudu, M., Le Strange, K. K., and Rolfe, B. G. (1988) The *nodD1* gene from *Rhizobium* strain NGR234 is a key determinant in the extension of host rang to the nonlegume Paraspona. *Mol. Plant-Microbe Interact. 1*, 259–266.

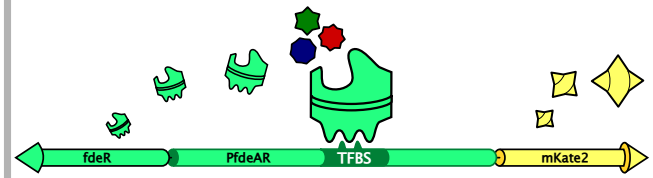
(84) Garmendia, J., Devos, D., Valencia, A., and De Lorenzo, V. (2001) À la carte transcriptional regulators: Unlocking responses of the prokaryotic enhancer-binding protein XylR to non-natural effectors. *Mol. Microbiol. 42*, 47–59.

(85) Juarez, J. F., Lecube-Azpeitia, B., Brown, S. L., and Church, G. M. (2017) Biosensor libraries harness large classes of binding domains for allosteric transcription regulators. *bioRxiv* 193029.

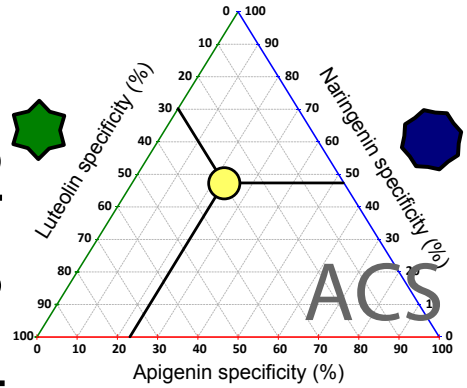
(86) Hammer, S. K., and Avalos, J. L. (2016) Metabolic engineering: Biosensors get the green light. *Nat. Chem. Biol. 12*, 894–895.

(87) Skjoedt, M. L., Snoek, T., Kildegaard, K. R., Arsovska, D., Eichenberger, M., Goedecke, T. J., Rajkumar, A. S., Zhang, J., Kristensen, M., Lehka, B. J., Siedler, S., Borodina, I., Jensen, M. K., and Keasling, J. D. (2016) Engineering prokaryotic transcriptional activators as metabolite biosensors in yeast. *Nat. Chem. Biol. 12*, 951–958.

Functional biosensor chassis



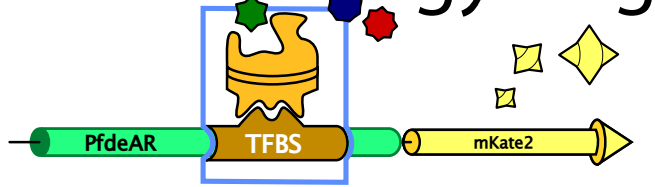
1
2
3
4



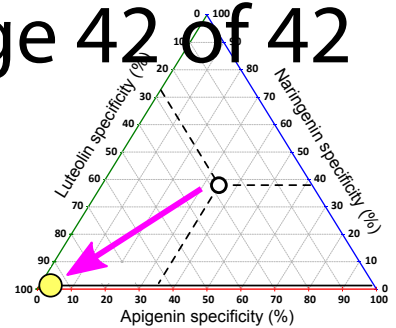
5 **Ligand specificity profile**

6

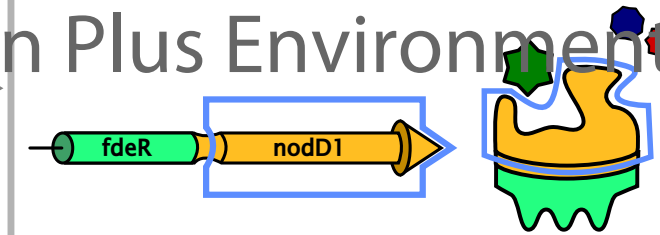
Strategy 1: Chimeric effector module



Custom ligand specificity profile



Strategy 2: Chimeric detector module



Custom ligand specificity profile