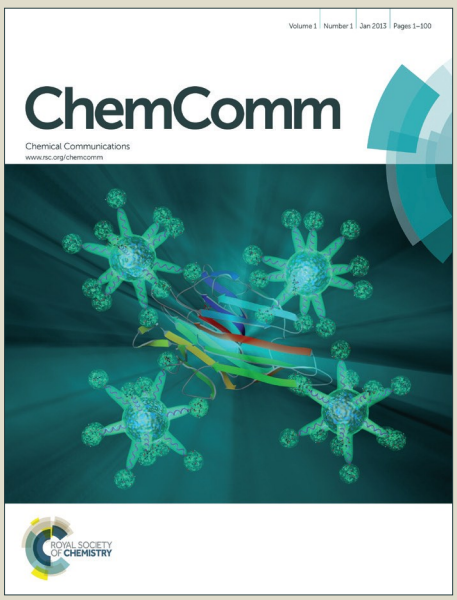
ChemComm

Accepted Manuscript



ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/chemcomm

Journal Name

COMMUNICATION

Development and substrate specificity screening of an *in vivo* biosensor for the detection of biomass derived aromatic chemical building blocks

Received 00th January 20xx,
Accepted 00th January 20xx

DOI: 10.1039/x0xx00000x

www.rsc.org/

Leopoldo F. M. Machado[a] and Neil Dixon*[,a]

Measuring substrate and/or product concentration can create a major bottleneck for synthetic and biosynthetic processes. Here we report the development and substrate screening of a whole cell biosensor to detect biomass-derived aromatic chemical building blocks, supporting the use of sustainable feedstocks in the bulk and fine chemical industries.

Valorization of biomass to create sustainable bio-synthetic routes to chemicals, plastics, monomers, waxes, fuel and energy is a central tenet of the move towards a circular bio-economy.[1] An area of particular importance is the ability to valorize low value waste/by-products such as lignin. Degradation of lignin can release chemical building blocks that can be used as substrates for the production of high value chemicals, flavors, and fragrances. [2,3] Production of high value chemicals from bio-based feedstocks can support the commercial feasibility of bio-fuels by utilizing a bio-refinery approach.[4–6] Determination of substrate and/or product concentration can create a major bottleneck for chemo-enzymatic and whole-cell biosynthetic processes, as both off-line biochemical activity screening and analytical methods can be laborious. *In vivo* biosensors provide a potential solution by enabling a real-time, intracellular read-out of activity/phenotype. [7,8] To facilitate the screening of chemical, enzymatic, and cellular processes to degrade and valorize plant biomass, we sought to develop a whole cell biosensor to detect lignin-derived substrates.

Lignin is a heterogeneous, polymeric, cross-linked material, mainly composed of monomers of *p*-coumaryl, coniferyl and sinapyl alcohols.[3] Several thermo-chemical (Kraft, Sulfite),[9] chemical (Organosolv, Alkaline hydrolysis),[10] and thermo-pressure (Steam explosion),[11] based extraction methods have been used to degrade lignin.[1,4] Enzymatic methods are currently expensive and require further optimization to be applied in large-scale.[12,13] Most well studied enzymatic processes are based on the use of isolated naturally occurring or recombinant fungal enzyme blends,[14,15] however, bacterial lignin degrading enzymes have also been identified as promising alternatives.[16,17] Degradation of lignin produces a mixture of phenylpropenoic acid monomers (e.g. *p*-coumaric, ferulic and caffeic acid). The ability to detect these lignin monomers would enable optimization of the enzymatic lignin degradation and valorization.

Here we report the development and substrate activity screening of an *in vivo E. coli* biosensor that permits the intracellular detection of the substituted cinnamic acid scaffold (e.g. ferulic acid). The system is based on the FerC repressor, a MarR-type repressor protein that binds to the DNA sequence upstream from the *ferB* gene (feruloyl-CoA hydratase) in *Sphingobium* sp SYK-6.[18] Previous *in vitro* studies identified that interaction between FerC and two operator (IR1 and IR2) sequences upstream from the *ferB* gene is inhibited in the presence of the CoA-esters of coumaric, ferulic, and sinapic acid. In order to test a number of biosensor designs we performed promoter engineering to create variant promoter-operator sequences (**Fig. S1, ESI**†). Using the higher affinity IR2 operator site we generated three promoter-operator sequences, two chimeric phage promoters, i) T7A1 promoter based (P_{PC}), ii) lambda phage promoter based (P_{LC}), and iii) the wild-type *ferB* promoter/operator (P_{ferB}). The relative constitutive expression levels from the three promoter variants were assessed by placing them upstream from *eGFP* reporter gene (**Fig. 1**). The gene expression output was normalized to cell density (RFU/OD600), and plotted relative to the biosensor with highest expression level (P_{LC}). The P_{ferB} reporter produced the lowest relative expression level (~20%), whereas the P_{PC} and P_{LC} reporters expressed respectively intermediary (~46%) and high expression levels in *E. coli* BL21 (**Fig. 2**). The DNA encoding the ferulic acid responsive repressor (FerC) and feruloyl coA synthetase (FerA) were

[a] *Manchester Institute of Biotechnology, University of Manchester, Manchester, United Kingdom. Email: neil.dixon@manchester.ac.uk*

† Footnotes relating to the title and/or authors should appear here.
Electronic Supplementary Information (ESI) available: [details of any supplementary information available should be included here]. See DOI: 10.1039/x0xx00000x

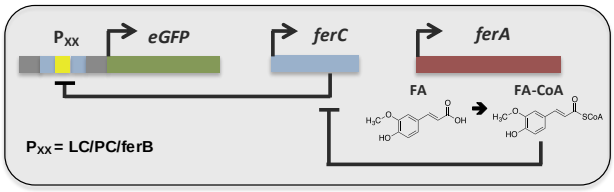


Figure 1. Biosensor design**.** Natural and engineered chimeric promoter-operators (P_{ferB}, P_{PC}, and P_{LC}) were inserted upstream from the eGFP reporter gene. FerC binds to the promoter- operator sequence(s) repressing expression of the reporter gene. In the presence of FerA, ferulic acid (FA) is converted into feruloyl-CoA (FA-CoA), which in turn de-represses FerC and activates gene expression.

cloned and expressed from constitutive promoters (**Methods, ESI†**). Expression of *ferC* resulted in repression of *eGFP* expression for all biosensor designs. The addition of ferulic acid (FA) to culture media led to de-repression of the biosensors and effective sensing of the intracellular presence of the substrate. The P_{LC} biosensor detected substrate concentrations over 13-fold sensing range, with ~23-fold signal range. Fitting with a dose response curve indicates an EC_{50} of 20.9 +/- 4.2 µM, saturation ≥100µM of ferulic acid, and incomplete de-repression (~85% of P_{LC} reporter) (**Fig. 2, Table 1**). The dose response curve for the P_{PC} biosensor indicates complete de-repression ≥40µM and an EC_{50} of 11.2 +/- 1.1 µM (**Fig. 2, Table 1**). The P_{ferB} biosensor was effectively de-repressed but had limited utility due to the small signal range. Considering the effects from the substrate perspective, FA has greater potency against the P_{PC} biosensor, but has a greater efficacy against the P_{LC} biosensor.

In order to validate the requirement of *ferA* and to provide *in vivo* validation of the previous *in vitro* observation,[18] we created *a ferA* knock-out (AKO). This allowed us to confirm *in vivo* that the active substrate is indeed the Coenzyme A (CoA) ester (**Methods, ESI†**). These strains lacked the de-repression phenotype upon addition of FA to growth media (**Fig. S2, ESI†**), confirming the essentiality of *ferA,* the FA-CoA ester as the ferC substrate, and the associated functional de-repression mechanism.[19] We next explored the dependency of the biosensor performance upon the host *E. coli* strain used by testing the system in an *E. coli* K strain (DH10B). The absolute expression levels were lower in the K strain (**Fig. S3, Table S1, ESI†**), however, the P_{LC} biosensor again produced the greatest relative signal range (~19-fold), displayed an EC_{50} of 22.8 +/- 5.2 µM, and sensing range (EC_{90}/EC_{10}) of 13-fold (**Fig. S3, and Table S1, ESI†**). As for the B strain (**Fig. 2**) incomplete de-repression was observed for the P_{LC} biosensor, whilst complete de-repression was again observed for both the P_{PC} and P_{ferB} biosensors. However, due to enhanced signal outputs of the B strain and the P_{LC} biosensor it was decided that this was the most effective combination and was used for subsequent activity screening.

In order to assess the substrate specificity of the ferulic acid biosensor, we selected the P_{LC} biosensor for further screening and tested its activity against 58 structural analogues (**Table S3, ESI†**). Five substrate analogues were identified from the screening, which had high output signal, hence good efficacy (>70%) against the biosensor (**Fig. 3A)** and displayed potencies ranging from EC_{50} 15 to 315 µM (**Table 2**). Based on the dose response curves, structure-activity relationships can be observed. Maximal potency requires a *para*-substituted phenyl ring with a hydrogen-bond donor (*p*-coumaric acid (**2**)), and a *meta*-methoxy substituent is also tolerated (ferulic acid (**1**)). Replacement of the hydroxyl substituent with an amino group (3-(4-aminophenyl)-2-propenoic acid (**3**)) results in loss of potency. Whereas, both the regioisomer of ferulic acid (3-hydroxy-4-methoxycinnamic acid (**5**)), and an additional *meta*-methoxy substituent on the phenyl ring (sinapic acid (**4**)) both result in a more significant loss of potency (EC_{50} > 200 µM). The fitting of the dose response curves also indicates an extension of the sensing range for different substrates. Ferulic acid is sensed over a 13-fold range, whereas sinapic acid presents the broadest predicted range (225-fold) (**Table S2, ESI†**). The next selection of substrates (**6-10**) (**Fig. 3B**) display a moderate signal (60-40%). Of these five substrates, one analogue (3,4-dihyroxy-5-methoxycinnamic acid (**6**)) displayed an EC_{50} of 746.5 µM with an extensive predicted sensing range (485-fold). The remaining moderately inducing substrates (2,4-dihydroxycinnamic acid (**7**), 3,4-dimethoxycinnamic acid (**8**), caffeic acid (**9**) and 4-nitrocinnamic acid (**10**)) displayed EC_{50} values ranging from 800 to 1600 µM (**Table 2**). The three remaining active substrates (3,4,5-trimethoxycinnamic acid (**11**), 3-(4-hydroxy-3-methoxyphenyl)propionic acid (**12**), and 3-methoxycinnamic acid (**13**)) displayed low signal (<40%) and low potency (>2000

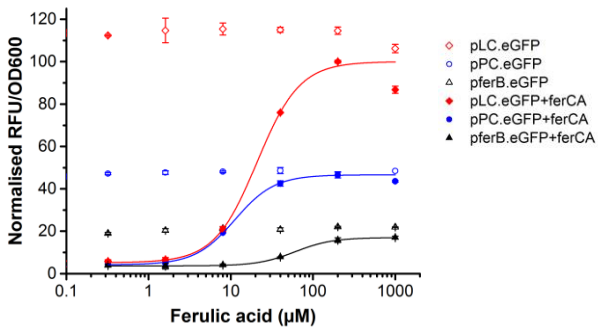


Figure 2. Biosensor performance. eGFP gene expression data in the absence (empty shapes) and presence of the *ferC* repressor (filled shapes), for the P_{ferB} (triangles), P_{PC} (circles), and P_{LC} (diamonds) biosensors in *E. coli* BL21. The fluorescent gene expression normalized to cell density (RFU/OD600) was expressed relative to the P_{LC} biosensor, and dose-response curves were fitted to increasing concentrations of ferulic acid.

Promoter	Signal range	EC50 (µM)	
P_{LC}	22.8	20.9	+/- 4.2
P_{PC}	10.4	11.2	+/- 1.1
P_{ferB}	5.0	60.5	+/- 2.3

Table 1. Signal range (max/min) and EC_{50} values from the fitted dose response curve for the three biosensor systems.

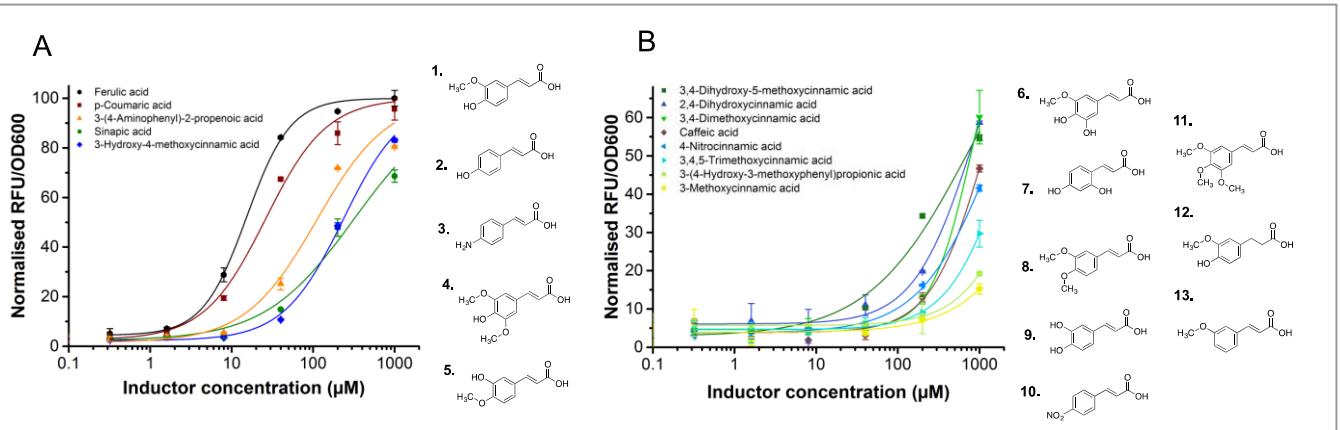


Figure 3. Biosensor responsive compounds. Dose response curves for different compounds using the P_{LC} biosensor system in *E. coli* BL21 and their respective molecular structure*. The fluorescent gene expression normalized to cell density (RFU/OD600) was expressed relative the pLC biosensor response curve with ferulic acid. Five compounds (**1-5**) generated high levels of expression with efficacy superior to 70% (**A**). Eight compounds generated moderate levels of expression with efficacy ranging from 10% to 55% (**6-13**) (**B**). *Compounds numbers with respective names are described in Table 2 and Table S3, ESI†.

μM). The remaining 45 analogues displayed no observable activity against the FerC biosensor (**Fig. S4**, **Table S3, ESI**†).

Consideration of the structure-activity relationships indicates modest phenyl substituent change can result in dramatic potency changes, for example an additional methyl group between 3,4,5-trimethoxycinnamic acid (**11**) and sinapic acid (**4**) results in 8-fold change in potency. Similarly it can be observed that the substrate (**6**) 3,4-dihyroxy-5-methoxycinnamic acid closely related to sinapic acid (**4**), displays a slightly reduced potency (2.5-fold), whereas the caffeic acid (**9**) displays a 55-fold loss in potency relative to the closely related ferulic acid (**1**). The low activity observed for 3-(4-hydroxy-3-methoxyphenyl)propionic acid (**12**) compared to ferulic acid (**1**) demonstrates the importance of the α-β unsaturated functionality. A number of observations can be made for the analogues that displayed no activity (**Fig S4, Table S3, ESI**†). Most interestingly, the un-substituted cinnamic acid (**14**) displayed no activity; in addition, the phenyl ring could not be exchanged for any other aromatic ring system (**23-28**); the presence of the carboxylic acid was confirmed by the lack of activity (**30-34**); the distance between the phenyl ring and carboxylic acid was confirmed (**39-46**) and the necessity of the α-β unsaturation for activity was confirmed (**16-22**). Finally, phenyl rings substituted with electron withdrawing halide groups (**50-54**), or electron donating methyl groups (4-Methylcinnamic acid (**29**)) were also devoid of activity.

	Tested compounds	Induction	Signal range	EC50 (μM)	
1	Trans-Ferulic acid	H	26.2	15.3	+/- 0.9
2	p-Coumaric acid	H	25.0	26.1	+/- 3.8
3	3-(4-Aminophenyl)-2-propenoic acid	H	28.1	110.2	+/- 25.7
4	Sinapic acid	H	15.4	314.4	+/- 55.8
5	3-Hydroxy-4-methoxycinnamic acid	H	33.5	234.3	+/- 15.4
6	3,4-Dihydroxy-5-methoxycinnamic acid	M	14.8	746.5	+/- 117.7
7	2,4-Dihydroxycinnamic acid	M	9.6	823.4	+/- 45.6
8	3,4-Dimethoxycinnamic acid	M	11.4	825.0	+/- 13.2
9	Caffeic acid	M	11.2	1176.4	+/- 88.5
10	4-Nitrocinnamic acid	M	9.5	1564.5	+/- 96.8
11	3,4,5-Trimethoxycinnamic acid	L	6.7	2364.9	+/- 264.9
12	3-(4-Hydroxy-3-methoxyphenyl) propionic acid	L	3.3	4687.1	+/- 1189.6
13	3-Methoxycinnamic acid	L	4.8	9251.1	+/- 11084

Table 2. Signal range (max/min) and EC_{50} values from a dose response curve fitting for all responsive compounds tested.

To demonstrate the utility of the biosensor in both a practical application and against complex substrate mixtures, we screened for biosensor activity following enzyme treatments against a number of different biomass sources. In total, we used the biosensor to assess the activity of three feruloyl esterases (**EC** 3.1.1.73, CAZy CE1) against three different biomass sources (**Fig. 4**). After enzyme treatment with ***CE1-3*** (from *C. thermocellum* DSM 1313) against wheat flour biomass, the biosensor screening confirmed efficient release of ferulic acid and/or closely related analogues, whereas treatment with enzymes ***CE1-1*** (from *A. cellulolyticus* CD2) and ***CE1-2*** (from *C. thermocellum)* resulted in reduced activity indicating only partial release. A similar relative activity profile was observed for the 3 enzymes against micronized oat husk biomass, however the total signal was reduced (>50%) suggesting lower levels of enzymatic release from this source. Thirdly, no activity was detected against Kraft lignin biomass. The observed lower level and lack of activity may be due to the more recalcitrant nature of the particular biomass source or, in the case of Kraft lignin, can be associated with the chemical pre-treatment process.[9]

In conclusion, the developed biosensor is able to detect 13 substituted cinnamic acid based compounds. The ferulic acid substrate is detected over a 13-fold sensitivity range, with >25-fold signal read-out range, and four other compounds display similar efficacy. The defined substrate specificity of this biosensor will enable its use in the identification and optimization of chemical and enzymatic processes. For example, processes which enable the de-polymerization of lignin and release of chemical building blocks, in addition to processes which use these chemical building blocks as substrates for the production of high value chemicals including vanillin and flavonoids. [2,6,20] Further applications include use of the biosensor in screening for the production of value-added compounds, and for substrate/product transport across biological membranes, which we are actively pursuing. This

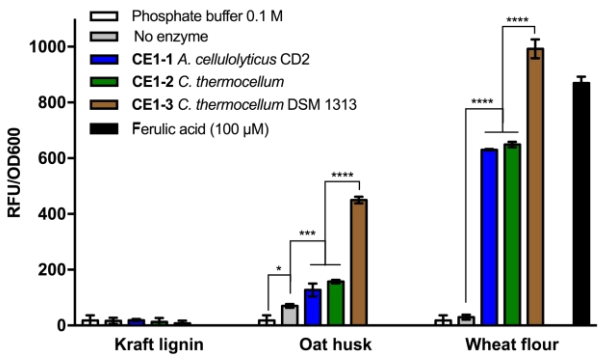


Figure 4. Biomass/lignin degradation screening with the pLC biosensor system in *E. coli* BL21. Different lignin sources were submitted to treatment with three feruloyl esterase enzymes (CE1) or absence of enzyme and the supernatants were tested with the pLC biosensor. The relative fluorescence to cell density (RFU/OD600) is shown for the feedstock treatments, the phosphate buffer alone or ferulic acid at 100 mM. (* P<0.05, *** P<0.001, **** P<0.0001 analyzed by one-way ANOVA followed by Tukey's Multiple Comparison Test).

combination of applications will support the chemical using industries to source chemical building blocks from alternative sustainable bio-based feedstocks.

This research was supported by the Biotechnology and Biological Sciences Research Council (BBSRC) grants [BB/L026244/1]. ND holds a BBSRC David Phillips Fellowship [BB/K014773/1]. LM is supported via Science without Borders / Ciência sem fronteiras scheme from the CNPq, Brazil (233608/2014-1). We would like to thank Prof David Leys Group for use of cinnamic acid compound library, and Prof Andrew Doig for useful discussions regarding data fitting.

Notes and references

All assays were performed by addition of substrates/ supernatants directly to culture of the biosensor containing *E. coli* strain(s), freshly grown to the appropriate cell density (OD 0.6). eGFP expression was monitored after three hours growth/induction at 37°C with shaking (1000 RPM). Cells were centrifuged, washed and re-suspended with PBS buffer. The expression output was then analyzed by monitoring the fluorescence normalised to cell density (RFU/OD600) in a multimode plate reader (**Methods, ESI†**).

1 A. J. Ragauskas, G. T. Beckham, M. J. Biddy, R. Chandra, F. Chen, M. F. Davis, B. H. Davison, R. A. Dixon, P. Gilna, M. Keller, P. Langan, A. K. Naskar, J. N. Saddler, T. J. Tschaplinski, G. A. Tuskan and C. E. Wyman, *Science*, 2014, **344**, 1246843.
2 M. Fache, B. Boutevin and S. Caillol, *ACS Sustain. Chem. Eng.*, 2016, **4**, 35–46.
3 J. Zakzeski, P. C. A. Bruijnincx, A. L. Jongerius and B. M. Weckhuysen, *Chem. Rev.*, 2010, **110**, 3552–99.
4 D. Esposito and M. Antonietti, *Chem Soc Rev*, 2015, **44**, 5821–5835.
5 G. T. Beckham, C. W. Johnson, E. M. Karp, D. Salvachúa and D. R. Vardon, *Curr. Opin. Biotechnol.*, 2016, **42**, 40–53.
6 T. D. Bugg and R. Rahmanpour, *Curr Opin Chem Biol*, 2015, **29**, 10–17.
7 D. Liu, T. Evans and F. Zhang, *Metab. Eng.*, 2015, **31**, 35–43.
8 K. M. Polizzi and C. Kontoravdi, *Curr Opin Biotechnol*, 2015, **31**, 50–56.
9 H. Chung and N. R. Washburn, *Green Mater.*, 2013, **1**, 137–160.
10 M. Zhang, Y. Xu and K. Li, *J. Appl. Polym. Sci.*, 2007, **106**, 630–636.
11 G. J. M. Rocha, A. R. Gonçalves, B. R. Oliveira, E. G. Olivares and C. E. V. Rossell, *Ind. Crops Prod.*, 2012, **35**, 274–279.
12 T. D. H. Bugg, M. Ahmad, E. M. Hardiman and R. Rahmanpour, *Nat. Prod. Rep.*, 2011, **28**, 1883.
13 R. Rahmanpour and T. D. H. Bugg, *Arch. Biochem. Biophys.*, 2015, **574**, 93–98.
14 C. Sánchez, *Biotechnol. Adv.*, 2009, **27**, 185–194.
15 D. W. S. Wong, *Appl. Biochem. Biotechnol.*, 2009, **157**, 174–209.
16 T. D. H. Bugg, M. Ahmad, E. M. Hardiman and R. Singh, *Curr. Opin. Biotechnol.*, 2011, **22**, 394–400.
17 M. E. Brown and M. C. Y. Chang, *Curr. Opin. Chem. Biol.*, 2014, **19**, 1–7.
18 D. Kasai, N. Kamimura, K. Tani, S. Umeda, T. Abe, M. Fukuda and E. Masai, *FEMS Microbiol. Lett.*, 2012, **332**, 68–75.
19 H. Otani, P. J. Stogios, X. Xu, B. Nocek, S.-N. Li, A. Savchenko and L. D. Eltis, *Nucleic Acids Res.*, 2015, gkv955.
20 D. R. Vardon, M. A. Franden, C. W. Johnson, E. M. Karp, M. T. Guarnieri, J. G. Linger, M. J. Salm, T. J. Strathmann and G. T. Beckham, *Energy Environ. Sci.*, 2015, **8**, 617–628.

