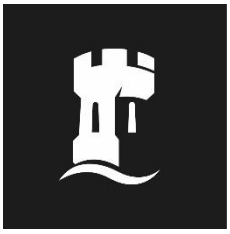


University of Nottingham
UK | CHINA | MALAYSIA

Identification, characterisation and application of inducible gene expression systems in *Cupriavidus necator* H16 and other bacteria

Ph.D. Thesis

Erik Kurt Reinhard Hanko, MSc (Hons)

2019

Thesis submitted to the University of Nottingham for the degree of Doctor of Philosophy, October 2019

Acknowledgements

Firstly, I would like to thank my supervisors Naglis Malys and Nigel Minton for taking me on as a student. Their expertise helped shape me into the scientist who I am today. I am particularly grateful for the excellent guidance and the opportunity to develop my own ideas within the scope of the project. Naglis, your door was always open for me, and, even though we had some challenging moments, especially during times when we were about to finalise a manuscript for submission, I will always look back at these moments as the ones I will miss the most. You are a true inspiration and an ideal for scientific accuracy and excellence.

I would like to acknowledge my wonderful colleagues who I have worked with during my time at the SBRC. Especially, I thank Ana Paiva, Magdalena Jonczyk, Matthew Abbott and David Onion for experimental assistance. Moreover, I would like extend a warm thank to all of my undergraduate and graduate students that have taught me at least as much as I was able to teach them. Aside from my colleagues I would like to express my gratitude to the BBSRC for the funding that enabled my research.

The most important people to thank are my family and friends. During my time in Nottingham, I was lucky to make friends with a bunch of wonderful people. I would like to thank Bashir Rumah, Christopher Stead, Jennifer Cooke, John Francis, Joshua Gascoyne, Lindsey Furness, Lorna Finch, Paul Cotton, Peter Stocks, Rupert Norman, Zuzana Grmanova, and everyone in C51 for making me feel home, for the good times, and for keeping me sane. A special thank you goes to my friends Cathleen Kmezik and Shayne Petkiewicz. No matter where you are or what time of the day it is, I can always count on you to cheer me up.

Last but not least, I would like to thank my parents Ute and Reinhard Hanko for always supporting me no matter how far from home it takes me.

Declaration

I declare that this thesis has been composed solely by myself and that it has not been submitted, in whole or in part, in any previous application for a degree. Except where stated otherwise by reference or acknowledgement, the work presented is entirely my own.

Erik Kurt Reinhard Hanko

Abstract

The production of key building block chemicals from renewable resources or waste forms a rapidly growing segment of the bioeconomy. The conversion of waste gases, such as carbon dioxide or carbon monoxide, into value-added compounds using metabolically engineered microorganisms has significant potential to maintain economic independence while reducing greenhouse gas emissions. Modification of cellular metabolism for the biosynthesis of a target molecule often requires an adjustment of gene expression, either of an endogenous or a heterologous metabolic pathway. Transcription factors are DNA-binding proteins that control gene expression at the transcriptional level in response to physical parameters, ions, or small effector molecules. They have become indispensable tools for the advancement of synthetic biology and metabolic engineering.

In this work, significant progress was made in the discovery and characterisation of transcription factor-based inducible gene expression systems for metabolic engineering of the chemolithoautotroph *Cupriavidus necator* H16 and other bacteria. Firstly, a quantitative evaluation of a range of well characterised heterolougous inducible systems in *C. necator* was undertaken. Four of them, the positively regulated L-arabinose- and L-rhamnose-inducible systems and the negatively regulated acrylate- and cumate-inducible systems, were subsequently employed for the biosynthesis of the industrially relevant building block chemical isoprene.

In addition to being used for controlling expression of structural genes, transcription factor-based inducible systems have gained increasing interest for their application as genetically encoded biosensors. Their ability to transduce the intracellular concentration of a target molecule into an output signal detectable in a

high-throughput format has the potential to revolutionise the field of microbial cell factory development. Currently, the number of compounds of biological interest by far exceeds the number of available biosensors. Here, this limitation was addressed by developing a universal genome-wide approach to identify novel transcription factor-based inducible gene expression systems. Once developed, the methodical pipeline was evaluated in the metabolically versatile *C. necator*. In total, 15 novel or little characterised inducible systems were identified and their broad host-range applicability was exemplified in three industrially relevant prokaryotes. Novel interactions between existing sensors and compounds of biological relevance were discovered by employing the largest reported library of transcription factor-based inducible systems in an automated high-throughput screen.

The same strategy, which was pursued in order to mine native inducible systems from the genome of *C. necator*, was used to source inducible systems responding to the industrially relevant platform chemicals 3-hydroxypropionic acid (3-HP) and itaconic acid. The HpdR/P_{hpdH}-3-HP-inducible system from *Pseudomonas putida* KT2440 and the ItcR/P_{ccl}-itaconic acid-inducible system from *Yersinia pseudotuberculosis* were thoroughly characterised for their regulator- and ligand-dependent orthogonality, induction kinetics and dynamics. This thesis highlights their potential to be applied as biosensors for high-throughput microbial strain development to facilitate improved 3-HP and itaconate biosynthesis.

List of publications

This thesis is based on the work presented in the following journal articles:

(I) Hanko, E. K. R, Minton, N. P., Malys N. (2019) Design, cloning and characterization of transcription factor-based inducible gene expression systems. In *Methods in Enzymology*, A. K. Shukla, ed. Elsevier. *621*, 153-169. (Corresponding to Chapter 3)

(II) Alagesan, S.*, Hanko, E. K. R.*, Malys, N.*, Ehsaan, M., Winzer, K., Minton, N. P. (2018) Functional genetic elements for controlling gene expression in *Cupriavidus necator* H16. *Appl. Environ. Microbiol. 84* (19), e00878-18. (Corresponding to Chapter 3)

(III) Hanko, E. K. R., Paiva, A. C., Jonczyk, M., Abbott, M., Minton, N. P., Malys, N. (2020) A genome-wide approach for identification and characterisation of metabolite-inducible systems. *Nat. Commun. 11* (1), 1-14. (Corresponding to Chapter 4)

(IV) Hanko, E. K. R., Minton, N. P., Malys N. (2017) Characterisation of a 3-hydroxypropionic acid-inducible system from *Pseudomonas putida* for orthogonal gene expression control in *Escherichia coli* and *Cupriavidus necator*. *Sci. Rep. 7* (1), 1724. (Corresponding to Chapter 5)

(V) Hanko, E. K. R., Minton, N. P., Malys N. (2018) A transcription factor-based biosensor for detection of itaconic acid. *ACS Synth. Biol. 7* (5), 1436-1446. (Corresponding to Chapter 6)

*Authors contributed equally to this work.

The permissions to reuse any material from the publications listed above in this thesis have been obtained.

Table of Contents

Table of figures

List of tables

Abbreviations

2-HB	2-Hydroxybutyric acid
2,4-DAPG	2,4-Diacetylphloroglucinol
3-HP	3-Hydroxypropionic acid
3-HB	3-Hydroxybutyric acid
A.U.	Arbitrary unit
BSA	Bovine serum albumin
C1	Single carbon
cAMP	Cyclic adenosine monophosphate
CBB	Calvin-Benson-Bassham
CDW	Cell dry weight
CRP	cAMP Receptor protein
DBD	DNA-binding domain
DMAPP	Dimethylallylpyrophosphate
DMSP	Dimethylsulfoniopropionate
DNA	Deoxyribonucleic acid
EMSA	Electrophoretic mobility shift assay
FACS	Fluorescence-activated cell sorting
FAEE	Fatty acid ethyl ester
FPP	Farnesyl pyrophosphate
GABA	γ-Aminobutyric acid
GC	Gas chromatography
GFP	Green fluorescent protein
GST	Glutathione S-transferase
h	Hours
HPLC	High performance liquid chromatography
HPLC-UV	HPLC coupled with ultraviolet absorption detection
IPTG	Isopropyl β-D-1-thiogalactopyranoside
LB	Lysogeny broth
LBD	Ligand-binding domain
LED	Light-emitting diode
LTTR	LysR-type transcriptional regulator

MEP/DOXP	2-C-methyl-D-erythritol 4-phosphate/1-deoxy-D-xylulose 5-phosphate
min	Minutes
mRFP	Monomeric RFP
mRNA	Messenger RNA
NADH	Nicotinamide adenine dinucleotide + hydrogen
NCBI	National Center for Biotechnology Information
NEB	New England Biolabs
OD	Optical density
PBS	Phosphate-buffered saline
PCR	Polymerase chain reaction
pH	*Potentia hydrogenii*
PHB	Poly[R-(-)-3-hydroxybutyric acid]
RBS	Ribosome binding site
RFP	Red fluorescent protein
RNA	Ribonucleic acid
RNAP	RNA polymerase
SBOL	Synthetic biology open language
TAL	Triacetic acid lactone
TFBS	Transcription factor binding site
T_m	Melting temperature
TR	Transcriptional regulator
ts	Temperature sensitive
TSS	Transcription start site
UV-Vis	Ultraviolet visible
v/v	Volume per volume
w/v	Weight per volume

1 Introduction

1.1 Context

The world is facing significant societal, economic, energy, and environmental challenges. With climate change, environmental pollution, and resource scarcity, the need for a transition from a petro- to bioeconomy has never been more important. The bioeconomy, which covers all economic sectors that are involved in the sustainable production and processing of food, materials, chemicals, and energy from waste or renewable biological resources[1], will allow economies to be less dependent on fossil resources, while reducing their negative impact on the environment[2]. In 2015, the European Commission estimated the annual revenues of the bioeconomy in the European Union to be more than two trillion Euros. Although the manufacture of bio-based chemicals, pharmaceuticals, plastics, and rubber accounted for only 177 billion Euros of the total, their production showed the highest value-added annual growth in the whole of the bioeconomy sector[1].

A large and rapidly growing segment of the bioeconomy comprises the engineering of microorganisms for the biosynthesis of chemicals, fuels, and polymers. Nature has equipped both eukaryotic and prokaryotic cells with an extensive repertoire of metabolic networks capable of synthesising more than 150.000 small molecules identified to date[3]. To facilitate the production of natural compounds, single genes or entire metabolic pathways are often transferred into microorganisms that are amenable to genetic manipulation[4], including the well characterised *Escherichia coli* or *Saccharomyces cerevisiae*. In addition, advances in *de novo* synthetic pathway design, where algorithms predict possible metabolic routes by combining enzymes from different species[5], enabled the microbial production of unnatural compounds, such as the plastic precursor 1,4-butanediol[6].

Although metabolic engineering has advanced as a field, with many success stories, including the commercial manufacture of fuels (e.g. isobutanol), chemicals (e.g. 1,3-propanediol), and pharmaceuticals (e.g. artemisinic acid), the development of microbial cell factories is still challenging due to the complexity inherent in metabolism and gene regulation[7]. Moreover, the feedstocks that are commonly utilised for microbial fermentation are mainly based on complex carbon sources which may conflict with the availability of food and scarcity of land. The use of single carbon (C1) gases as abundant and low-cost feedstocks is a promising alternative. C1 gases, including carbon monoxide (CO), carbon dioxide (CO_2), and methane (CH_4), are waste products generated by a wide range of industrial processes. In 2018, the total CO_2 emissions from fossil fuels and industry were estimated to be 37.1 gigatons, which represents an increase of 2.7% compared to the previous year[8]. In addition to industrial off-gases, organic matter, such as biomass, municipal solid waste, or organic industrial waste, can be gasified and utilised as feedstocks for microbial gas fermentation[9]. The gas mixture obtained by gasification of organic matter is called synthesis gas (syngas) and contains mainly CO, CO_2, H_2, and N_2[9]. Syngas has been historically utilised as a replacement for gasoline to power internal combustion engines[10]. It received prominence during World War II as an estimated seven million vehicles were run on syngas due to a shortage of gasoline[11]. In the last decade, there has been an increase in interest in microorganisms that are capable of utilising C1 gases, and especially syngas, as carbon source. These microorganisms include anaerobic acetogens, aerobic chemolithoautotrophic and methanotrophic bacteria, as well as photoautotrophic cyanobacteria[12] (Figure 1.1). The advantage of using C1 gas-utilising microorganisms lies in their ability to fix greenhouse gases before they are released into the atmosphere and synthesise molecules that resemble compounds

currently derived from petroleum. The process of microbial gas fermentation may therefore represent a key technology to meet national and international emission targets but also to reduce the dependency on petroleum[13]. The use of C1 gas-fixing microorganisms for such applications is hampered by the limited number of available genetic and analytical tools that enable high-throughput strain engineering and screening. This thesis presents significant steps toward the discovery, design, characterisation and implementation of transcription factor-based inducible system regulatory elements for metabolic engineering of the chemolithoautotroph *Cupriavidus necator* H16 and other bacteria.

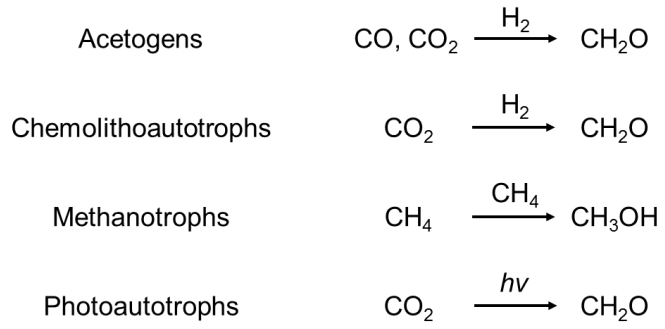


Figure 1.1 Summary of C1 gas-utilising microorganisms. The carbon- and energy sources are illustrated that are used by acetogens, chemolitoautotrophs, methanotrophs, and photoautotrophs to produce biomass. *hv*, light.

1.2 Response to changes in the environment: regulation of gene expression in bacteria

The control of gene expression through the adjustment of RNA and protein levels is essential to the regulation of cellular functions. Living organisms have evolved a variety of sensing and regulatory mechanisms to monitor and respond to changes in their local environment by adjusting the expression of their genes[14]. Coupling gene expression to physical conditions or the availability of specific metabolites in their environment gives cells the advantage of limiting the expenditure of energy and

materials to the synthesis of enzymes that are required under these growth conditions[15]. The fundamental regulatory mechanism that orchestrates gene expression is RNA polymerase (RNAP)-mediated promoter recognition and transcription initiation[16].

Although the RNAP core complex, composed of several protein subunits, is capable of mediating DNA-dependent RNA synthesis, it is unable to recognise promoters without prior binding of a sigma factor. The sigma factor, too, comprises several protein subunits with each subunit being able to interact with a discrete sequence motif upstream of the gene transcriptional start site[16]. Subsequently, the complex of RNAP and sigma factor, termed RNAP holoenzyme, is able to initiate transcription from specific promoters upon motif recognition. Transcription initiation is mainly controlled by altering the function of RNAP or by modulating the affinity or accessibility of promoters for RNAP[16].

Sigma factors interact directly with RNAP to alter its promoter specificity[17]. All bacteria possess one predominant sigma factor, also called the housekeeping sigma factor (such as σ^{70} in *E. coli*), orchestrating transcription of the majority of genes[16]. In addition to the housekeeping sigma factor, alternative sigma factors may guide the RNAP to different sets of promoters primarily in response to environmentally-induced stress, such as heat or starvation. Depending on the sigma factor, transcription initiation is controlled by the abundance of the sigma factor or the presence of an additional activator protein, like in the case of σ^{54}, which is involved in gene expression under nitrogen-limiting conditions[18].

Transcription initiation can also be controlled by modulating the promoter DNA-affinity or accessibility. The mechanisms to achieve this level of regulation range from supercoiling to proteins that attract the RNAP holoenzyme or simply block

access to the promoter[16]. Transcriptional regulators (TRs) are DNA-binding proteins that mediate gene expression in response to physical parameters (e.g., temperature, light or pH), ions, or small effector (ligand) molecules, also referred to as inducers. TRs that are controlled by small effector molecules are composed of a DNA-binding domain (DBD) and a ligand-binding domain (LBD). The DBD specifically interacts with a transcription factor binding site (TFBS), also termed operator sequence, within the promoter through non-covalent binding of the protein side chains to the exposed base pairs of the DNA groove[19]. The TFBS for the majority of DNA-binding structural motifs, such as the common helix-turn-helix motif, comprises 4-5 base pairs. The most frequent mechanism to achieve further specificity is homodimerisation or multimerisation of the TR. Therefore, promoters may contain multiple TFBSs of 4-5 base pair sequence direct or inverted repeats[16]. The LBD is required for effector binding, upon which a conformational change takes place, altering the ability of the TR to bind DNA or recruit the RNAP holoenzyme[20].

Depending on their mode of action, TRs are classed as repressors, activators or dual-function transcription factors[21]. They use one of several mechanisms to repress or activate transcription initiation[16]. Repressors bind to their target DNA sequence in the absence of the ligand. This sequence may overlap with core elements of the promoter, thus preventing the binding of RNAP to the promoter (e.g. the tetracycline repressor TetR)[22]. Repression can also be achieved by repressors simultaneously binding to distally located operators. The formation of a DNA loop prevents the promoter to be recognised by RNAP (e.g. the galactose repressor GalR)[23]. More complex mechanisms of repression are mediated by repressors that counteract activators (e.g. the anti-activator CytR)[24] or directly bind RNAP (e.g. the p4 protein from phage φ29)[25]. In contrast, activators bind to their target DNA sequence in the

presence of the ligand. Activation of transcription initiation is achieved by forming direct interactions with the RNAP holoenzyme (e.g. cAMP receptor protein-dependent activation of the *lac* promoter)[26] or by realigning suboptimally spaced core promoter elements to facilitate RNAP binding (e.g. the multidrug-responsive activator BmrR)[27]. Similarly to anti-activators, positive regulators exist that antagonise repressors (e.g. the anti-repressor Ler)[28]. Dual-function transcription factors serve as both activators and repressors. For example, expression of the *rhlAB* operon in *Pseudomonas aeruginosa* is regulated by RhlR which represses transcription in the absence of its autoinducer butanoyl-homoserine lactone but acts as activator in the presence of the ligand[29].

TRs are classed into families on the basis of their highly conserved DBDs. The LBDs are very diverse in their amino acid composition due to the variety of ligand-specific moieties. The largest family of prokaryotic DNA-binding proteins is the LysR-type transcriptional regulator (LTTR) family comprising more than 800 members identified to date[30]. LTTRs can act as activators, repressors or dual-function TRs of single genes or operons. They are often located in the opposite orientation of the cluster of genes they regulate but may also be located elsewhere on the bacterial genome. LTTRs exert diverse regulatory functions. They are involved in metabolism, quorum sensing, virulence, and nitrogen fixation, just to name a few (Table 1.1)[30].

Table 1.1 Examples of LTTRs and the functions that they are involved in.

Regulator	Origin	Function	Reference
CatM	*Acinetobacter baylyi*	Benzoate metabolism	[31]
YofA	*Bacillus subtilis*	Cell division	[32]
QseA	Enterohemorrhagic *Escherichia coli*	Quorum sensing	[33]
NodD	*Rhizobium* spp.	Nitrogen fixation	[34]
SpvR	*Salmonella typhimurium*	Virulence	[35]

1.3 Transcription regulation networks

TRs modulate the rate of gene expression in order to adjust the level of proteins that are needed under certain environmental conditions. In addition to structural genes, TRs may control expression of regulatory genes, including their own, building a complex network of interactions, called transcription regulation networks[36]. The genetic circuits that these networks are made up of are based on a small set of network motifs[37]. These motifs are universal and can be found in both prokaryotes[38] and eukaryotes[39]. The simplest network motif is the one of a TR A controlling expression of gene B in response to a stimulus (Figure 1.2a). Upon initiation of transcription, the concentration of gene product B rises until a steady-state level is reached[36]. Often, TRs repress the transcription of their own genes[40] (Figure 1.2b). This network motif is termed negative autoregulation. It can speed up the response time in cases where expression of TR gene is controlled by a strong promoter. At early stages, when levels of TR protein are low, its concentration increases rapidly until its own promoter is fully repressed, resulting in a steady-state level of TR protein that is close to its repression threshold[36]. The opposite effect is achieved when TRs activate expression of their own genes (Figure 1.2c). Positive autoregulation slows down the response time as initial low concentrations of TR protein result in low levels of TR gene expression. Once its concentration reaches an activation threshold, the TR protein is produced more rapidly[36].

Feedforward loops (FFLs) are more complex naturally occurring genetic circuits[41]. They are composed of three genes, two of which encode TRs (Figure 1.2d). Regulator A controls expression of regulator B. Expression of gene C is controlled by both TRs. As regulators can be transcriptional activators or repressors, a total of eight different types of FFL exist. Furthermore, the biological function of the FFL depends

on whether both regulators are required to activate expression of gene C or if one of them is enough. Two of the eight types of FFL, the coherent type-1 FFL and the incoherent type-1 FFL, were found to be more common than the other six types in *E. coli*[42] and yeast[41].

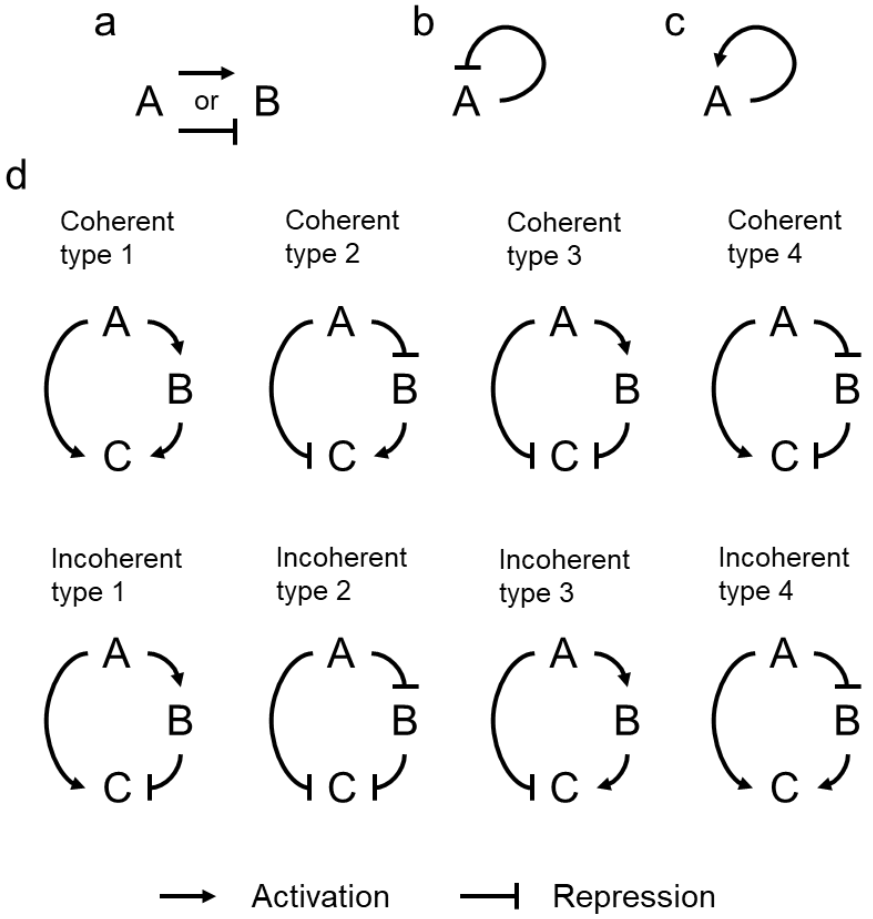


Figure 1.2 Transcription regulation network motifs. (a) Simple regulation network of a TR A controlling expression of gene B. (b) In negative autoregulation, TR A represses transcription from its own promoter. (c) In positive autoregulation, TR A activates transcription from its own promoter. (d) The eight types of feedforward loops. These networks are composed of two TRs A and B both controlling the C promter.

In case of the coherent type-1 FFL, both regulators are transcriptional activators. When both regulators are required to activate expression of gene C, production of C is delayed until the concentration of gene product B reaches the activation threshold for the C promoter. Upon removal of the stimulus, A is no longer able to activate C

expression, thus production of C is stopped immediately. This network motif makes expression of C less prone to spurious pulses of input signal[36]. For example, CRP (cAMP receptor protein) controls expression of both the gene encoding the AraC activator and the structural genes *araBAD* and *araFGH* required for arabinose utilisation in *E. coli*[43]. The delay of *araBAD* and *araFGH* expression in response to the input signal cAMP has been identified to be on the same timescale as naturally occurring oscillations of intracellular cAMP levels[44].

The opposite effect can be observed when only one of the two regulators is required to activate C expression. In the presence of a stimulus, regulator A activates expression of both genes B and C immediately. Removal of the signal, however, will result in continued expression of gene C until the concentration of regulator B is below the activation threshold for the C promoter. A transient loss of stimulus will therefore not lead to a stop of C production. In *E. coli*, expression of the operons that synthesise the flagella motor is controlled by the activators FlhDC and FliA in an additive fashion[45]. The FFL prolongs expression of the structural genes after removal of input signal, thus protecting biogenesis of the flagella[46].

Incoherent type-1 FFLs can function as pulse generators and response accelerators[36]. In this network motif, regulator A activates expression of both genes B and C (Figure 1.2d). Gene product B, however, represses transcription of gene C. In the presence of a stimulus, both proteins B and C are rapidly produced. Once the concentration of gene product B reaches the repression threshold for the C promoter, production of protein C will decrease. In the case that the C promoter is fully repressed by regulator B, the concentration of protein C will drop to zero, resulting in a pulse. In the case that regulator B only partially represses promoter C, the concentration of gene product C will reach a steady-state level. Due to a strong initial increase in levels

of protein C, before the concentration of regulator B reaches the repression threshold for the C promoter, the response time needed to activate expression of C is shorter than if it was a simply regulated system[36].

1.4 Application of transcription factor-based inducible gene expression systems

TR-based inducible systems have been historically utilised for both gene overexpression and protein production and gene silencing and the generation of conditional gene knockouts. Their modular architecture and ability to act as biological switches has allowed TR-based inducible systems to recently be harnessed for the construction of synthetic regulatory circuits, such as logic gates, toggle switches, or oscillators[47]. Furthermore, their ability to act as genetically encoded biosensors gives them the potential to contribute to a wider range of applications including diagnostics, monitor environmental pollution, and revolutionise the field of microbial cell factory development.

1.4.1 Protein production

The production of recombinant proteins was one of the earliest applications of TR-based inducible systems. In contrast to constitutive promoters, transcription from the inducible promoter can be switched on at any desired point during the cultivation, enabling the production of recombinant proteins that might negatively affect cellular functions. Of all the expression hosts available, including bacteria, yeast, insect cells or human cells, *E. coli* has been utilised by far the most[48]. In *E. coli*, the most commonly used expression systems are based on pET vectors[49]. The vector harbours the gene encoding the recombinant protein under control of the bacteriophage T7 DNA polymerase promoter containing the *lac* operator. The expression host itself has to harbour a genomic or an episomal copy of the gene encoding T7 RNAP under control

of the *lacUV5* promoter. *E. coli* naturally encodes the LacI TR, which represses transcription from the T7 DNA polymerase hybrid promoter and the *lacUV5* promoter. Upon supplementation with isopropyl β-D-1-thiogalactopyranoside (IPTG), a non-metabolisable structural analogue of allolactose, LacI dissociates from its target promoters, T7 RNAP is produced, and transcription of the target gene is initiated. The advantage of using the T7 DNA polymerase promoter is that it is not recognised by the endogenous *E. coli* RNAP which adds an additional level of regulation. Furthermore, the T7 RNAP transcribes approximately five times faster than the *E. coli* RNAP[50], thus, once it is produced, most of the protein synthesis machinery is devoted to producing the recombinant protein[48]. Basal T7 DNA polymerase promoter activity can be reduced by simultaneously expressing T7 lysozyme, which is a natural inhibitor of T7 RNAP, or by adding glucose to the cultivation medium to enhance repression of the *lacUV5* promoter by the cAMP-CRP complex[50].

1.4.2 Synthetic regulatory networks

Metabolic engineering has the potential to produce chemicals from renewable resources using biocatalysts. However, engineering of an organism for the biosynthesis of a target chemical is challenging due to the complexity of cellular metabolism and its underlying regulatory mechanisms[7]. Synthetic biology aims to design and build microbial cell factories using well-defined genetic parts, thus reducing the complexity inherent in biological systems. Ideally, the genetic elements should be programmable and their behaviour predictable, similar to the well characterised parts in traditional engineering disciplines, which follow defined physical rules. The implementation of these controllable genetic elements in a cellular system is an essential part of synthetic regulatory circuit design where cells have been

programmed to perform computational operations as a first step toward the development of a designer cell[51, 52].

The first synthetic regulatory networks, where electrical circuit analogies have been adapted to program cellular behaviour, were the genetic toggle switch[53] and the repressilator[54]. In both cases, regulatory elements of transcription factor-based inducible systems were leveraged to implement the equivalents of electronic memory storage and timekeeping in a biological system.

The toggle switch is a bistable gene-regulatory network that is controlled by two orthogonal input signals. Depending on the input signal, it will switch to one of the two stable states. Upon removal of the signal, the switch will remain at its current state indefinitely, thus mimicking a biological form of memory. The bacteriophage λ lysis-lysogeny switch[55], where bistability is achieved by two promoters that are repressed by each other's gene product, served as basis for the construction of the synthetic toggle switch[53]. In one version of the genetic toggle that was constructed by Gardner and co-workers, the bacteriophage λ P_L promoter mediates expression of *lacI* (Figure 1.3a). Its translational product represses transcription from the second promoter P_{trc2} which controls the transcription of *cl-ts*. The gene product is a temperature-sensitive (ts) Cl repressor protein which inhibits transcription from the P_L promoter. System stability is controlled by exogenous supplementation with IPTG or a transient increase in temperature[53].

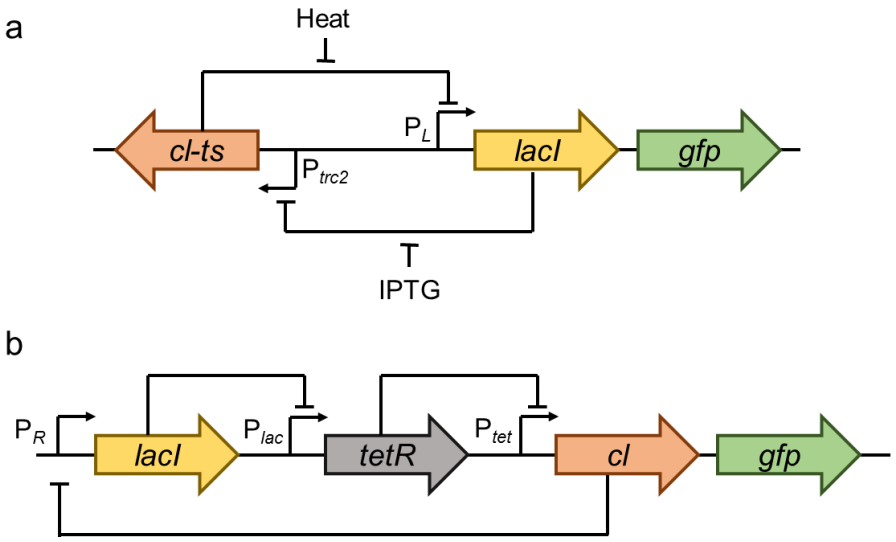


Figure 1.3 Application of TR-based inducible systems for the design of synthetic regulatory networks. (a) The toggle switch is composed of two promoters that are repressed by each other's gene product. Stabilty is achieved by addition of IPTG or an increase in temperature. (b) The repressilator is a cyclic negative feedback loop composed of three repressor proteins. The dynamics of each network is monitored by GFP fluorescence output.

In electronics, oscillating networks are based on positive and negative feedback loops that produce periodic output signals. A similar mechanism exists in biological systems. For example, circadian clocks are gene-regulatory networks that periodically express a specific set of genes[56], thus creating a cellular form of timekeeping. Elowitz and Leibler constructed a synthetic oscillatory network based on a set of three repressor proteins and their corresponding negatively regulated promoters, hence termed the repressilator[54]. In their design, the first repressor protein, LacI, inhibits expression of the second repressor gene, *tetR*, whose translational product represses the transcription of the third gene, *cI*. Cl in turn inhibits expression of *lacI*, completing the cyclic negative feedback loop (Figure 1.3b). In both cases, the toggle switch and the repressilator, a fluorescent protein reporter gene was incorporated into the synthetic circuit to monitor the dynamics of the systems. These are only two examples where regulatory elements of transcription factor-based inducible systems were utilised to build synthetic regulatory networks. Other

examples of electronics-inspired genetic devices include digital logic gates[52, 57, 58], pulse generators[59], and band-pass filters[60].

1.4.3 Genetically encoded biosensors

The throughput of the design- and build-steps of the typical design-build-test cycle of synthetic biology has increased by several orders of magnitude due to advances in *in silico* pathway design and automated genome engineering. Combinatorial DNA assembly methods enable the construction of thousands of metabolic pathway variants in a rapid manner[61]. However, the quantification of metabolites by liquid chromatography or mass spectrometry is commonly limited to fewer than 10^2 samples per instrument per day, posing a bottleneck for the evaluation of the generated constructs[62].

Linking a metabolite-responsive transcription factor-based inducible gene expression system to a genetic actuator (e.g., fluorescent reporter, selection marker or regulatory switch) overcomes such bottleneck. It enables each cell to measure the intracellular concentration of a target molecule and convert it into a detectable and quantifiable output, or alternatively, dynamically control a metabolic pathway[62].

Employing a fluorescent protein as output signal allows the multiplexed evaluation of metabolite biosynthesis within single cells in a rapid and iterative manner using high-throughput screening methods, such as fluorescence-activated cell sorting (FACS)[63]. Numerous examples exist, where genetically encoded biosensors have been used to screen libraries of metabolically engineered strains and isolate individual cells with improved product formation capabilities[63-66]. For example, Mustafi and co-workers utilised the transcriptional regulator Lrp from *Corynebacterium glutamicum* to isolate amino acid-producing mutant strains after random mutagenesis of a non-producing wild type strain using FACS[63]. In another

study, AraC, the activator protein which regulates L-arabinose utilisation in *E. coli*, was engineered to respond to triacetic acid lactone (TAL) and exploited to screen for 2-pyrone synthase enzyme mutants that led to increased TAL biosynthesis in *E. coli*[67].

When linked to a selection marker, such as an antibiotic resistance gene, the biosensor output represents a fitness advantage and can be implemented to enable a product-dependent selection. In a study of van Sint Fiet and co-workers, the NahR/P_{sal}-inducible system, responding to benzoate and 2-hydroxybenzoate, was fused to the gene encoding the tetracycline efflux pump[68]. Biocatalytically active *E. coli*, which are able to oxidise benzaldehyde and 2-hydroxybenzaldehyde, were identified through supplementation with the respective antibiotic.

In biological systems, the biosynthesis of small molecules is controlled by a complex regulatory machinery that enables optimal metabolic flux and prevents the accumulation of toxic intermediate compounds[7]. The integration of a heterologous pathway, however, may result in detrimental effects on the host metabolism. Biosensors linked to structural genes can be used as regulatory switches to dynamically control metabolic pathways and alleviate the accumulation of molecules that may interfere with other biological processes[69]. For example, Zhang and co-workers employed the fatty acyl-CoA-responsive repressor protein FadR for the biosynthesis of fatty acid ethyl esters (FAEEs) in a fatty acid-overproducing strain of *E. coli*[69]. Upon accumulation of fatty acyl-CoAs, FadR dissociates from three target promoters, resulting in the biosynthesis of ethanol, the activation of free fatty acids to fatty acyl-CoAs, and the condensation of ethanol and fatty acyl-CoAs to FAEEs. This dynamic pathway control prevented the accumulation of toxic ethanol and efficiently pulled the metabolic flux toward FAEE biosynthesis.

In addition to biosensors that are based on TR-based inducible systems, RNA switches have been developed that respond to small molecules and control gene expression at the translational level. In this case, the genetic actuator harbouring the riboswitch is expressed constitutively. However, its mRNA secondary structure prevents translation to be initiated. Upon binding of the effector molecule, the riboswitch undergoes structural changes allowing the mRNA to be processed[70].

1.5 Approaches to find a suitable transcription factor-based inducible gene expression system

After a target compound has been identified or an objective has been defined, the challenge remains to find a suitable TR-based inducible system. Online databases, such as DBD[71] (www.transcriptionfactor.org), PRODORIC[72] (www.prodoric.de), or RegPrecise[73] (regprecise.lbl.gov/RegPrecise), provide useful information on prokaryotic TRs and their corresponding target DNA sequences. Unfortunately, only a limited number of orthogonal metabolite-responsive promoters has been well characterised[74]. Inducible systems are said to be orthogonal if they are controlled exclusively by their corresponding TRs and when the TR itself is regulated solely by its unique effector molecule. Cross-reactivity caused by unspecific regulator- or ligand binding may interfere with the circuit design and generates noise[51]. In this case, a TR-promoter pair from a different organism can be used to decrease the background activity or promoter- and protein engineering may be performed to increase sensitivity. If the compound of interest is not listed, or if regulatory elements for a specific microorganism have not been researched, different strategies may be pursued to identify a suitable inducible system.

1.5.1 Strategies to identify promoters responding to a specific molecule

Two primary strategies have proven to be efficient in discovering promoters that respond to a given effector molecule[75, 76]. By using genome-wide transcriptional analyses, single genes or operons can be identified that are differentially expressed in the presence of the compound of interest. This approach was followed by Dahl and co-workers for the identification of a farnesyl pyrophosphate (FPP)-responsive promoter in *E. coli*[75]. A second strategy may involve the screening of a promoter-reporter fusion-library for reporter gene expression after supplementation with the inducer. For example, the Alon library, containing 2,000 different *E. coli* promoter-*gfpmut2* fusions, yielded the P_{mtr} promoter that responded to L-phenylalanine[76]. Both strategies have their limitations. They do not exclude promoters that are indirectly activated, nor do they guarantee the identification of the corresponding TR which is required for the application of the inducible system in a host organism different to the one the promoter was mined from.

1.5.2 Strategies to identify TR-promoter pairs responding to a specific molecule

To identify both controllable promoters and their corresponding TRs, entire DNA fragments may be analysed for their ability to mediate gene expression in response to a target molecule. Uchiyama and Miyazaki cloned DNA fragments from a metagenomic library into an open trap vector harbouring a promoter-less fluorescent protein reporter gene[77]. Following this strategy, several TR-promoter pairs were identified that responded to a range of aromatic compounds. However, it relies on the assumption that the inducible system is functional in a host organism different to the one from which it was sourced.

1.5.3 Predicting the effector molecule for a given inducible system

The reverse strategy relies on predicting the effector molecule for a putative inducible system based on genetic context or comparative genomics. In some instances, such as the majority of TRs belonging to the TetR-family, the regulated genes are divergently oriented with an intergenic region of less than 200 bp[78]. Cuthbertson and co-workers developed a methodical workflow based on this genetic arrangement[79]. Using phylogenomics, TetR-family regulators known to bind similar effectors were clustered in order to predict ligands for TRs of unknown function in the same subfamily. Following this approach, the kijanimicin-responsive TR KijR was identified. Moreover, comparative genomics have been leveraged for the prediction of TFBSs for members of the LacI- and ROK-family of TRs[80, 81]. Using the known TFBSs, target genes were identified and metabolic pathways were reconstructed, allowing putative effector molecules to be proposed. Both strategies have resulted in the identification of TR-promoter pairs and their corresponding effectors, but are limited to a specific family of TRs or class of compounds.

1.6 *Cupriavidus necator* – a promising microbial host organism for the conversion of CO_2 into value-added products

Cupriavidus necator H16, formerly known as *Ralstonia eutropha*, *Alcaligenes eutrophus*, *Wautersia eutropha*, and *Hydrogenomonas eutropha*, is a facultative chemolithoautotrophic, soil-dwelling Gram-negative β-proteobacterium. It has been extensively studied for its capacity to store large amounts of organic carbon, up to 80% of its cell dry weight (CDW)[82], in the form of poly[R-(-)-3-hydroxybutanoate] (PHB). Because PHB is a biodegradable polymer, the main focus of *C. necator* research had been the production of sustainable bioplastics[83-85]. It has gained increasing attention recently for the biosynthesis of a range of value-added

compounds, achieved by redirecting the metabolic flux from carbon storage toward the product of interest. Compounds that have been synthesised by recombinant *C. necator* include alkanes, fatty acids, alcohols, methyl ketones, and hydroxy acids[86-95].

The genome of *C. necator* H16 comprises three circular replicons, two of which are chromosomes and one a megaplasmid, with a total size of 7,416,678 bp[96]. Due to an impressive array of genes related to carbon and energy metabolism, *C. necator* is able to grow under heterotrophic, lithoautotrophic or organoautotrophic conditions[97]. Substrates that permit heterotrophic growth include tricarboxylic acid cycle intermediates, sugar acids, fatty acids, amino acids or other organic acids, alcohols, polyols, and aromatics[96, 98]. *C. necator* possesses the ability to use both organic compounds and molecular hydrogen (H_2) as sources of energy[96], utilising them to power metabolic processes and chemolithoautotrophically fix CO_2 *via* the Calvin-Benson-Bassham (CBB) cycle[99]. Alternatively, it is capable of growing organoautotrophically by oxidising formate to CO_2 by the action of a formate dehydrogenase. During this reaction, energy is generated in form of NADH. The released CO_2 is subsequently assimilated *via* the CBB cycle[100, 101]. In the absence of oxygen (O_2), it can switch to anaerobic respiration-denitrification, exploiting alternative electron acceptors, such as nitrite (NO_2^-) or nitrate (NO_3^-)[97]. Its metabolic versatility is furthermore highlighted by the existence of 688 genes for potential regulatory proteins, 521 of which are one component regulators[96]. Thus, the genome of *C. necator* itself may serve as a rich basis for mining inducible systems that can be used as regulatory elements in synthetic regulatory network design and biotechnology applications.

1.7 Inducible systems for gene expression control in *C. necator* H16

Despite its early recognition as promising host organism for the autotrophic biosynthesis of value-added compounds, significant progress in the development of genetic tools for gene expression control in *C. necator* has mainly been made in the last decade. A number of constitutive and inducible promoters for the production of homologous and heterologous proteins have been developed since. Initially limited to a handful of constitutive promoters, including the well characterised heterologous P_{lac}, P_{tac}, P_{j5}, and the native P_{phaC} and P_{pdhE}[102-104], the operational range has been vastly increased recently to span more than two orders of magnitude by promoter engineering or by investigating a broader range of promoters sourced from different organisms[105, 106]. Constitutive promoters, however, do not allow gene expression to be auto regulated or externally controlled, which plays a vital role in synthetic circuit design and the implementation of heterologous metabolic pathways. These limitations can be overcome by using transcription factor-based inducible systems. Although native promoters that are auto induced under autotrophic or phosphate-limiting growth conditions have been successfully utilised in *C. necator*[107, 108], inducible systems that can be controlled independently of growth conditions are generally more advantageous. To date, a few of these effector-inducible gene expression systems have been established in *C. necator* (Table 1.2).

Of these systems, the L-arabinose-inducible gene expression system has been utilised the most for production of value-added compounds in *C. necator*[87, 90, 92]. It exhibits several beneficial features over most of the other characterised systems. For example, in contrast to the IPTG-inducible system, it does not require additional expression of a transport system in order for the inducer to be taken up by the cell[86]. Secondly, it is not metabolised by *C. necator* unlike acetoin, thus mediating a

sustained gene expression throughout the cultivation[109]. Thirdly, P_{araBAD} demonstrates low background expression levels in the absence of L-arabinose[86]. Lastly, the inducer itself is relatively inexpensive. In addition to the L-arabinose-inducible system, the RhaSR/P_{rhaBAD}-L-rhamnose-inducible system has great potential due to low basal promoter activities. However, the range of inducer concentration over which the system can be controlled was reported to be less than one order of magnitude in minimal medium[109].

Table 1.2 Summary of effector-inducible gene expression systems employed in *C. necator*. Note that the dynamic range, the ratio of expression level in the presence of inducer to basal promoter activity, may be difficult to compare from different studies due to unstandardised evaluation methods.

Effector	Inducible system	Origin	Dynamic range	Application	Reference
Acetoin	AcoR/P_{acoE}	*C. necator* H16	ND[a]	Production of poly-hydroxyalkanoates	104
L-arabinose	AraC/P_{araBAD}	*E. coli* K12	> 12-fold	Production of methyl ketones	92, 102
m-Toluic acid	XylS/P_m	*P. putida* KT2440	N/A	Production of hydrocarbons	86
IPTG	LacI/P_{lacUV5}	*E. coli* K12	N/A	Production of hydrocarbons	86
IPTG	LacI/P_{j5}	*E. coli* K12/ bacteriophage T5	< 7-fold	N/A[b]	110
p-Cumic acid	CymR/P_{j5}	*P. putida* F1/ bacteriophage T5	> 33-fold	Production of esterase protein	110
Anhydro-tetracyline	TetR/P_{rrsC}	Tn*10*/ *C. necator* H16	> 10-fold	Controllable expression of *alsS*	111
Anhydro-tetracyline	TetR/P_{tolC}	Tn*10*/ *C. necator* H16	180-fold	Production of mevalonate	112
L-rhamnose	RhaSR/P_{rhaBAD}	*E. coli*	> 140-fold	N/A[b]	109

[a]Not determined. [b]Not applicable.

1.8 Research aims

The aims of this research can be divided into the following sections:

Identification of inducible gene expression systems - A few TR-based inducible gene expression systems exist for metabolic engineering of *C. necator* H16 and other bacteria. To make promising microbial cell factories accessible to more advanced engineering strategies, however, additional orthogonal highly inducible and ligand-specific inducible systems need to be identified. To achieve this goal, heterologous inducible systems will be investigated. Furthermore, the genomes of a range of bacteria will be explored to source inducible systems that can be specifically employed for the monitoring of industrially relevant platform chemicals, such as 3-hydroxypropionic acid and itaconic acid.

Development of a genome scale approach for inducible system mining - A generic approach will be developed to mine inducible systems from annotated bacterial genomes. The potential of this strategy will be demonstrated by applying it to identifying novel or little characterised metabolite-inducible systems in *C. necator* H16.

Parametric characterisation of inducible systems - Heterologous and native inducible systems will be quantitatively evaluated for their ability to control gene expression in a comparative manner. They will be parameterised to facilitate forward engineering efforts and their orthogonality will be determined.

Application of inducible systems - Selected inducible systems will be applied for isoprene biosynthesis in *C. necator*. Other novel inducible systems will be developed into biosensors. The itaconate biosensor will be employed for improving itaconate biosynthesis.

2 Materials and methods

2.1 Chemicals

Chemicals were purchased from Acros Organics (Thermo Fisher Scientific), Alfa Aesar (Thermo Fisher Scientific), Honeywell, Fluorochem, and Sigma-Aldrich unless stated otherwise. All chemicals are listed in Supplementary Table 1 in the *Appendix*.

2.2 Bacterial strains, cultivation media and growth conditions

2.2.1 Bacterial strains

Bacterial strains used in this work are listed in Table 2.1.

Table 2.1 Bacterial strains used in this work.

Strain	Characteristic	Reference or source
Cupriavidus necator H16	Wild type strain	DSM 428
Escherichia coli MG1655	Wild type strain	DSM 18039
Escherichia coli Rosetta DE3	F^- *ompT hsdS*$_B$(r_B^- m_B^-) *gal dcm* (DE3) pRARE (Cam^R)	Novagen
Escherichia coli TOP10	F^- *mcrA* Δ(*mrr-hsdRMS-mcrBC*) Φ80*lacZ*ΔM15 Δ*lacX*74 *recA*1 *araD*139 Δ(*araleu*)7697 *galU galK rpsL* (StrR) *endA*1 *nupG*	Thermo Fisher Scientific
Pseudomonas putida KT2440	Wild type strain	Kindly provided by Dr. Stephan Heeb

2.2.2 Cultivation media

All bacterial strains were propagated in lysogeny broth (LB). Fluorescence reporter assays were performed in minimal medium unless indicated otherwise. The M9 minimal medium[113] for cultivation of *E. coli* MG1655 was supplemented with 1

µg/mL thiamine, 20 µg/mL uracil[114] and 4 g/L glucose. *C. necator* was grown in minimal medium[115] supplemented with 4 g/L sodium gluconate unless stated otherwise. Antibiotics for plasmid maintenance were added at the following concentrations: 25 µg/mL chloramphenicol, 12.5 µg/mL tetracycline, 50 µg/mL kanamycin, or 100 µg/mL ampicillin for *E. coli*; 50 µg/mL chloramphenicol or 15 µg/mL tetracycline for *C. necator*; 25 µg/mL tetracycline for *P. putida*. The composition of the media is as follows:

LB	Tryptone	10 g/L
	Yeast extract	5 g/L
	NaCl	10 g/L
M9 minimal medium[113]	$Na_2HPO_4 \cdot 7H_2O$	12.8 g/L
	KH_2PO_4	3 g/L
	NaCl	0.5 g/L
	NH_4Cl	1 g/L
	$MgSO_4$	0.24 g/L (2 mM)
	$CaCl_2$	11.1 mg/L (0.1 mM)
C. necator minimal medium[115]	$Na_2HPO_4 \cdot 12H_2O$	9 g/L
	KH_2PO_4	1.5 g/L
	NH_4Cl	1 g/L
	$MgSO_4 \cdot 7H_2O$	0.2 g/L
	$CaCl_2 \cdot 2H_2O$	20 mg/L
	Fe(III)NH_4-citrate	1.2 mg/L
	SL7 solution	1 mL/L

SL7 solution[116] is composed of 1.3 mL/L 25% (w/v) HCl, 62 mg/L H_3BO_3, 190 mg/L $CoCl_2 \cdot 6H_2O$, 17 mg/L $CuCl_2 \cdot 2H_2O$, 100 mg/L $MnCl_2 \cdot 4H_2O$, 36 mg/L $Na_2MoO_4 \cdot 2H_2O$, 24 mg/L $NiCl_2 \cdot 6H_2O$, and 70 mg/L $ZnCl_2$.

For solid media preparation 15 g/L agar were added.

2.2.3 Growth conditions

E. coli was routinely grown at 37 °C unless indicated otherwise. For comparison, *E. coli*, *C. necator*, and *Pseudomonas putida* fluorescent protein reporter gene assays were performed at 30 °C.

2.2.4 Maintenance of strains

To prepare a permanent stock of a bacterial strain, 200 µL of an overnight culture were resuspended in the preservation medium of a CRYOBANK™ tube (COPAN Diagnostics Inc.). The supernatant was removed and the beads were stored at -80 °C. To revive the strain, a bead was streaked onto an LB agar plate or directly used to inoculate liquid medium.

2.3 Molecular cloning

2.3.1 Extraction of plasmid and genomic DNA

Plasmid DNA was purified by using the New England Biolabs (NEB) Monarch Plasmid Miniprep Kit or the QIAprep Spin Miniprep Kit (Qiagen). Microbial genomic DNA was extracted by employing the GenElute Bacterial Genomic DNA Kit (Sigma-Aldrich). All DNA purifications were performed according to the manufacturer's instructions.

2.3.2 Quantification and storage of DNA

Purified DNA was eluted in nuclease-free water and quantified using a SimpliNano™ spectrophotometer (Biochrom). DNA was stored at -20 °C.

2.3.3 Oligonucleotide primers

Oligonucleotide primers were synthesised by Eurofins Genomics or Sigma-Aldrich. Primers used for restriction enzyme-based cloning were designed to harbour at least three nucleotides upstream of the restriction site at the 5' end to allow for restriction

enzyme binding. Primers used for Hifi DNA assembly were designed to have 5' overhangs of 12 to 28 nucleotides to generate overlapping homologous DNA sequences. The annealing temperature was calculated using the NEB T_m Calculator (https://tmcalculator.neb.com). All oligonucleotide primers used in this work are listed in Supplementary Table 2.

2.3.4 Polymerase chain reaction

For cloning, DNA was amplified by polymerase chain reaction (PCR) using Phusion- or Q5 High Fidelity DNA polymerase from NEB in 50 µL reactions following the manufacturer's instructions.

2.3.5 Agarose gel electrophoresis

Agarose gel electrophoresis was performed on PCR products and DNA digests using a 1% agarose gel. The gel was stained with SYBR™ Safe DNA Gel Stain (Thermo Fisher Scientific). DNA samples were prepared by adding Purple Gel Loading Dye (6x, NEB). The 2-Log DNA Ladder (NEB) was used to determine DNA fragment sizes. Electrophoresis was performed using a Bio-Rad electrophoresis system with a PowerPac™ Universal Power Supply (Bio-Rad) at 100 V until DNA fragments were clearly separated.

2.3.6 Agarose gel imaging

Agarose gels were imaged using a Bio-Rad Molecular Imager GelDoc XR+ Imaging System running software Image Lab™ (Bio-Rad). DNA that was used for cloning was visualised using a blue light LED transilluminator (Syngene).

2.3.7 Extraction of gel-purified DNA

Gel-purified linearised DNA was extracted using the NEB Monarch DNA Extraction Kit or the Zymoclean Gel DNA Recovery Kit (Zymo Research) according to the manufacturer's instructions.

2.3.8 DNA restriction digestion

Restriction enzymes were purchased from NEB. Reactions were set up according to the manufacturer's protocol and incubated over night.

2.3.9 Hifi DNA assembly

The NEBuilder Hifi DNA Assembly Master Mix (2x) was used for vector assembly of two or more DNA fragments with overlapping homology regions. For assemblies of 2 to 3 fragments, a 1:2 vector:insert molar ratio with 50 ng of vector was used. Assemblies with 4 to 6 fragments were performed using a 1:1 vector:insert molar ratio with 50 ng of vector. Reactions were set up according to the manufacturer's instructions with a total volume of 10 µL.

2.3.10 DNA ligation

DNA ligations were performed using the NEB T4 ligase. Reactions were set up typically with a 1:3 vector:insert molar ratio following the manufacturer's instructions.

2.3.11 Plasmid construction

All plasmids generated in this thesis were constructed by employing either conventional restriction enzyme-based cloning techniques or the Hifi DNA assembly method and are based on the pBBR1 replicon[117]. Key features of all plasmids used and generated in this thesis are summarised in the individual chapters. A detailed description of how each plasmid was assembled is provided in the *Appendix* (Chapter 9.1).

2.3.12 DNA sequencing

Constructs were verified by DNA Sanger sequencing performed by Source Bioscience (Nottingham).

2.4 Preparation and transformation of microbial strains

2.4.1 Preparation and transformation of chemically competent *E. coli*

Chemical competent *E. coli* were prepared and transformed by heat shock as described by Sambrook and Russel[113].

2.4.2 Preparation and transformation of electrocompetent *C. necator*

Electrocompetent *C. necator* were prepared and transformed by following a method reported by Ausubel et al.[118]. Briefly, 25 mL of SOB medium (Hanahan's broth) were inoculated with 1 mL of a *C. necator* overnight culture in a 250-mL baffled shake flask. The culture was incubated with orbital shaking at 30 °C and 200 rpm until an OD_{600} of 0.2-0.3 was reached. Cells were harvested by centrifugation at 12.000*g* and 4 °C for 10 min. The pellet was resuspended in chilled 10 mL of buffer A [1 mM HEPES (pH 7.0)] and centrifuged as before. It was washed a second time with chilled 5 mL of buffer A and centrifuged as before. Subsequently, the pellet was resuspended in buffer B [(1 mM HEPES (pH 7.0), 10% glycerol (v/v)] to an OD_{600} of 5. Aliquots of 100 µL were stored at -80 °C.

100 µL of electrocompetent cells were added to 100 ng plasmid DNA in an electroporation cuvette (0.2 cm electrode gap, Bio-Rad) and chilled on ice for 5 min. Electroporation was performed using a Bio-Rad Micropulser at 2.5 kV/cm (200Ω, 25 µF). Immediately after electroporation, 900 µL of SOC medium was added and the suspension was transferred to a 50-mL conical centrifuge tube. After an incubation with orbital shaking at 30 °C and 200 rpm for 2 h, 100 µL of transformed cells were spread on an LB agar plate containing the respective antibiotic.

2.4.3 Preparation and transformation of electrocompetent *P. putida*

Electrocompetent *P. putida* were freshly prepared from overnight cultures grown in LB medium with orbital shaking at 30 °C and 200 rpm. Per reaction, 1 mL of cells were harvested by centrifugation at 16,000*g* for 5 min and washed three times with 1 mL of ice-cold 10% (v/v) glycerol. Electroporation was conducted in an electroporation cuvette (0.2 cm electrode gap, Bio-Rad) using a Bio-Rad Micropulser at 12.5 kV/cm by following a method reported by Sambrook and Russel[113]. Preparation and transformation of *P. putida* was performed with assistance of Ana Paiva.

2.5 Analytical techniques

2.5.1 OD_{600} measurement

Cell absorbance was measured in a 1-cm-path-length cuvette using a BioMate 3S UV-visible (UV-Vis) spectrophotometer at 600 nm (Thermo Fisher Scientific).

2.5.2 Fluorescence measurement

For quantification of red fluorescent protein (RFP) fluorescence at a single time-point, individual colonies of freshly transformed bacterial cells were used to inoculate 5 mL of LB medium in 50-mL conical centrifuge tubes. After incubation over night with orbital shaking at 30 °C and 200 rpm, *E. coli* and *P. putida* were diluted 1:50, and *C. necator* was diluted 1:20 into 5 mL of fresh LB medium. The exponentially growing cells were supplemented with inducer at an OD_{600} of 0.5. After a further incubation with orbital shaking at 30 °C and 200 rpm for 6 h, cells were pelleted by centrifugation at 16,000*g* for 4 min and resuspended in an equal volume of phosphate buffered saline (PBS). Subsequently, 100 µL of cells were transferred to a 96-well microtiter plate (flat and clear bottom, black; Greiner One International) and RFP fluorescence was quantified using an Infinite M1000 PRO microplate reader (Tecan).

Fluorescence excitation and emission wavelengths were set to 585 nm and 620 nm, respectively. The gain factor was set manually to 100%. Absorbance was measured at 600 nm to normalise fluorescence by optical density. Prior normalisation, fluorescence and absorbance values were corrected by subtracting the auto fluorescence and absorbance of the culture medium. Following changes were made if quantification of RFP fluorescence was performed in minimal medium: (i) the precultures were set up in 2 mL of minimal medium, (ii) all strains were diluted 1:20 into 5 mL of fresh minimal medium, (iii) 100 µL of the bacterial culture was directly used to measure fluorescence and absorbance.

For the growth and fluorescence time course experiments, bacterial precultures were set up as for the single time-point measurements. Cells were diluted 1:40 into 5 mL of fresh minimal medium and incubated with orbital shaking at 30 °C and 200 rpm until an OD_{600} of 0.2 was reached. Subsequently, 142.5 µL of the exponentially growing cells were transferred to a 96-well microtiter plate. To the cultures, 7.5 µL stock inducer was added at the desired concentration. Fluorescence and absorbance were quantified using the same settings as for the single time-point measurements every 5 min for 16 h.

2.5.3 HPLC-UV analysis

High-performance liquid chromatography with ultraviolet absorption detection (HPLC-UV) analysis was performed as described previously[119] with slight modifications. Briefly, to the cell-free supernatant, an equal volume of mobile phase was added which was spiked with 50 mM valeric acid as internal standard. The mobile phase was composed of 0.005 M H_2SO_4. The mixture was vortexed and subsequently passed through a 0.22 µm pore size membrane filter. Samples were analysed using a Thermo Scientific Ultimate 3000 HPLC system equipped with a Phenomenex Rezex

ROA-organic acid H+ (8%) 150 mm x 7.8 mm x 8 µm column and a diode array detector with the wavelength set at 210 nm. The column was operated at 35 °C with an isocratic flow rate of 0.5 mL/min. Samples were run for 30 min and the injection volume was 20 µL. HPLC-UV analysis was performed with assistance of Matthew Abbott.

Chapter 3

The following chapter is mainly based on the work presented in the publications:

Functional genetic elements for controlling gene expression in *Cupriavidus necator* H16

S. Alagesan*, E. K. R. Hanko*, N. Malys*, M. Ehsaan, K. Winzer & N. P. Minton (2018)

Applied and Environmental Microbiology ***84*** (19), e00878-18.[106]

*Authors contributed equally to this work

&

Design, cloning and characterization of transcription factor-based inducible gene expression systems

E. K. R. Hanko, N. P. Minton & N. Malys (2019)

in *Methods in Enzymology,* A. K. Shukla, ed. Elsevier, ***621***, 153-169.[120]

3 Evaluation of heterologous inducible systems and application in isoprene biosynthesis

3.1 Introduction

To develop and optimise biosynthetic pathways in metabolically engineered microorganisms, genome alterations often require either an adjustment of gene expression or an introduction of heterologous genes, in both cases utilising functional genetic elements, such as transcription factor-based inducible systems, to control gene expression[121]. A few studies have reported on the characterisation of inducible systems in *C. necator*. The heterologous inducible promoters P_{araBAD} (L-arabinose)[102], P_m (*m*-toluic acid)[86], P_{lac} (lactose)[86], and P_{rhaBAD} (L-rhamnose)[109], the synthetic anhydrotetracycline- and cumate-inducible promoters[110, 111], and a few native inducible promoters[104, 107, 108], have been shown to be suitable to control and drive gene expression in this microorganism. In addition, Bi and co-workers demonstrated that incorporation of a 5' mRNA stem-loop structure upstream of the ribosome binding site (RBS) sequence increased gene expression from P_{araBAD} 2.3-fold in *C. necator*[86].

A strong need to expand the synthetic biology toolbox remains, aiming to broaden the range of inducible systems for controlling gene expression in *C. necator* and to extend the repertoire of positive and negative regulators that are required for highly controllable circuits in synthetic biology and biotechnology applications. After an initial screen of nine heterologous inducible systems, responding to the ligands L-arabinose, acetate, propionate, L-rhamnose, D-mannitol, acrylate, 4-isopropylbenzoate (cumate), 2,4-diacetylphloroglucinol (2,4-DAPG), and D-xylose, two positively and two negatively regulated inducible systems were selected to be characterised in more detail. The AraC/P_{araBAD}-L-arabinose-, RhaRS/P_{rhaBAD}-L-rhamnose-, and CymR/P_{j5}-cumate-inducible systems have been employed previously

in *C. necator*[86, 109, 110]. The fourth system is regulated by AcuR, a repressor protein from *Rhodobacter sphaeroides* mediating gene expression from P_{acuRI} in the presence of acrylate[122]. Whereas expression from the L-arabinose-, cumate-, and acrylate-inducible promoters is controlled by a single TR, the L-rhamnose-inducible system differs from the other investigated systems as it comprises two regulator proteins (Figure 3.1). When present, L-rhamnose binds to RhaR, which in turn initiates expression of the operon encoding both regulatory proteins RhaR and RhaS. Subsequently, expression of the L-rhamnose catabolic operon, *rhaBAD*, is activated by the L-rhamnose-bound RhaS[123].

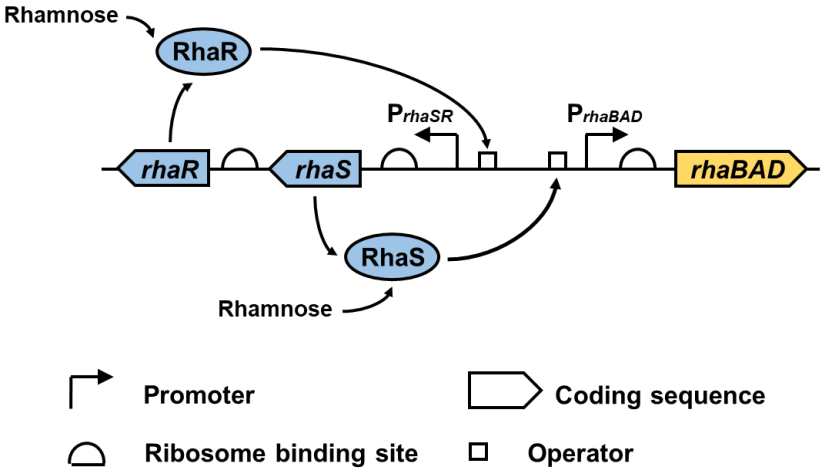


Figure 3.1 Schematic illustration of the L-rhamnose-inducible system. Expression of the genes encoding the regulators RhaS and RhaR is activated by RhaR in the presence of L-rhamnose. The promoter P_{rhaBAD} in turn is activated by L-rhamnose-bound RhaS.

Gene expression can result in different output patterns for metabolites and reporter molecules, such as fluorescent proteins. To evaluate how the variation in gene expression can affect product biosynthesis, a simple, one enzymatic reaction pathway extension based on a single gene addition was chosen as a suitable model for investigation. Among several potential targets particularly relevant to the industrial application, production of isoprene emerged as the most significant. Isoprene is a naturally occurring volatile terpenoid emitted by plants to protect against

environmental stress factors including heat, drought, or singlet reactive oxygen species[124, 125]. It is synthesised in chloroplasts by an isoprene synthase (IspS) enzyme[126]. Currently, it is mainly produced from petroleum. However, in light of environmental concerns and due to its versatile application in synthetic chemistry for the manufacture of synthetic rubber, medicines, pesticides, or as aviation fuel[127], the sustainable biological production of isoprene has gained increasing interest. IspS has been demonstrated previously to catalyse the production of isoprene from dimethylallyl pyrophosphate (DMAPP)[128, 129], which in *C. necator* can be produced via the 2-C-methyl-D-erythritol 4-phosphate/1-deoxy-D-xylulose 5-phosphate (MEP/DOXP) pathway. Notably, due to poor catalytic properties of the enzyme (high K_m and low k_{cat}), isoprene synthase is considered as one of key bottlenecks in isoprene biosynthesis[130].

The aim of this chapter was to assemble and quantitatively evaluate several heterologous inducible systems in a comparative manner. To ensure an evaluation of gene expression independent of genetic context, a modular reporter vector was designed comprising an insulated application-specific module. After nine inducible systems had been initially evaluated for their ability to mediate gene expression, four of them were selected to be quantitatively characterised for their kinetics and dynamics. From the dynamic data, which was generated by monitoring the expression output of the gene encoding the fluorescent protein over time at various inducer concentrations, the maximum protein synthesis rate of each inducible system was determined using a mathematical model. This type of parameterisation is key in order to aid in part selection and to facilitate forward engineering efforts. The four inducible systems were subsequently used to investigate how gene expression correlates with formation of the industrially relevant platform chemical isoprene. Furthermore, the

influence of the mRNA stem-loop structure was evaluated for the L-arabinose-inducible system and compared to previous findings.

3.2 Materials and methods specific to this chapter

3.2.1 Chemicals

L-arabinose, sodium acetate, propionic acid, L-rhamnose monohydrate, D-mannitol, magnesium acrylate, 4-isopropylbenzoic acid (cumic acid), 2,4-diacetylphloroglucinol (2,4-DAPG), and D-xylose were used as inducers for assaying inducible systems. Isoprene was used as a standard for isoprene yield quantification. All chemicals are listed in Supplementary Table 1 in the *Appendix*.

3.2.2 Plasmids

Key features of all plasmids used and generated in this chapter are summarised in Table 3.1. A detailed description of how each plasmid was assembled is provided in the *Appendix*. Plasmids for isoprene production were constructed to harbour *Populus alba ispS* gene optimised for *C. necator* codon usage under control of different inducible systems. The DNA sequence of *ispS* was truncated by 49 amino acids at the N-terminus and can be found in the *Appendix*. The nucleotide sequence of plasmid pEH006 has been deposited in the public version of the JBEI registry (https://public-registry.jbei.org) under the accession number JPUB_008750.

Table 3.1 Plasmids used in chapter 3.

Plasmid	Characteristic	Reference or source
pBBR1MCS-2-PphaC-eyfp-c1	Kan^r; P_{phaC}-*eyfp*	[131]
pBBR1MCS-2-RBS_1-ispS	Kan^r; P_{phaC}-RBS_1-*ispS*	[106]

pJOE7784.1	Kan^r; P_{rhaSR}-*rhaSR*; P_{rhaBAD}-*egfp*-T_{rrnB}	[132]
pJOE7801.1	Kan^r; P_{tetR}-*tetR*; P_{tetA}-*egfp*-T_{rrnB}	[132]
pKTrfp	Cm^r; P_{araC}-*araC*; P_{araBAD}-*rfp*-T_{dbl}	[86]
pNEW	Kan^r; P_{km}-*cymR*; P_{cmt}-*gfp*-T_{T7}	[133]
pEH002	Cm^r; P_{rhaSR}-*rhaSR*-T_{rrnB1}; P_{rhaBAD}-T7sl-*rfp*-T_{dbl}	This work
pEH002-ispS	Cm^r; P_{rhaSR}-*rhaSR*-T_{rrnB1}; P_{rhaBAD}-T7sl-*ispS*-T_{dbl}	This work
pEH003	Cm^r; P_{mtlR}-*mtlR*-T_{rrnB1}; P_{mtlE}-T7sl-*rfp*-T_{dbl}	This work
pEH005	Cm^r; P_{lac}-*tetR*-T_{rrnB2}; P_{tetA}-T7sl-*rfp*-T_{dbl}	This work
pEH006	Cm^r; P_{araC}-*araC*-T_{rrnB1}; P_{araBAD}-T7sl-*rfp*-T_{dbl}	This work
pEH006-ispS	Cm^r; P_{araC}-*araC*-T_{rrnB1}; P_{araBAD}-T7sl-*ispS*-T_{dbl}	This work
pEH006E	Cm^r; T_{rrnB1}-*rfp*-T_{dbl}	This work
pEH015	Cm^r; P_{prpR}-*prpR*-T_{rrnB1}; $P_{prpBCDE}$-T7sl-*rfp*-T_{dbl}	This work
pEH016	Cm^r; P_{alsR}-*alsR*-T_{rrnB1}; P_{alsSD}-T7sl-*rfp*-T_{dbl}	This work
pEH020	Cm^r; P_{lac}-*acuR*-T_{rrnB2}; P_{acuR}-T7sl-*rfp*-T_{dbl}	This work
pEH020-ispS	Cm^r; P_{lac}-*acuR*-T_{rrnB2}; P_{acuR}-T7sl-*ispS*-T_{dbl}	This work
pEH038	Cm^r; P_{lac}-*xylR*-T_{rrnB2}; P_{xylAB}-*rfp*-T_{dbl}	This work
pEH040	Cm^r; $P_{lac\text{-}}$*cymR*-T_{rrnB2}; P_{T5}-*rfp*-T_{dbl}	This work
pEH040-ispS	Cm^r; $P_{lac\text{-}}$*cymR*-T_{rrnB2}; P_{T5}-*ispS*-T_{dbl}	This work
pEH043	Cm^r; P_{lac}-*phlF*-T_{rrnB2}; P_{phlA}-*rfp*-T_{dbl}	This work
pEH176	Cm^r; P_{araC}-*araC*-T_{rrnB1}; P_{araBAD}-*rfp*-T_{dbl}	This work

3.2.3 Growth conditions

To be consistent with other measurements performed in the study of Alagesan *et al.*[106], *C. necator* strains were grown in minimal medium containing 4 g/L of fructose as carbon source. The initial experiments to determine which inducible systems are functional in *C. necator* were performed in minimal medium supplemented with 4 g/L of sodium gluconate as carbon source.

3.2.4 Fluorescence measurements

To capture the full range of fluorescence output, the gain factor was set manually to 80% unless indicated otherwise.

3.2.5 Estimation of protein synthesis rate and Hill coefficient

The rate at which the fluorescent reporter protein is produced can be calculated for each ligand concentration at any time *t* using the normalised absolute fluorescence values from the time course fluorescence measurements with formula (3.1)[74]:

$$\Delta\mathrm{RFP}_{n,t} = \frac{\mathrm{RFP}_t}{\mathrm{OD}_t} - \frac{\mathrm{RFP}_{t-1}}{\mathrm{OD}_{t-1}} \tag{3.1}$$

The time that is required from the activation of reporter gene expression to the maturation of the fluorescent protein was defined as Δt ($\Delta t \approx 30$ min for mRFP1[134], monomeric RFP). The rate of mRFP1 degradation was neglected in equation (3.1) as it was reported to demonstrate slow protein unfolding kinetics[135]. As protein synthesis rates depend not only on the genomic context of the inducible system, but also on growth rate-dependent parameters, such as the plasmid copy number and the abundance of RNAP and ribosomes[136], it is not surprising that these change over the time course of the experiment. The RFP synthesis rate as a function of time for the L-rhamnose-inducible system is provided as example in Supplementary Figure 1. From this plot, the maximum rate of RFP synthesis v_{max} can be determined. The value of v_{max} is subsequently fit to the corresponding ligand concentration using a Hill function, taking into account the basal rate of protein synthesis of the uninduced cells v_{min}[74]:

$$\left(\frac{\Delta\mathrm{RFP}}{\Delta t}\right)_{\max} = v_{\max} \cdot \frac{I^h}{I^h + {K_m}^h} + v_{\min} \tag{3.2}$$

Parameters correspond to the ligand concentration *I*, the Hill coefficient *h*, and the ligand concentration that results in half-maximal RFP synthesis K_m.

3.2.6 Isoprene production

Freshly grown overnight cultures of plasmid-transformed *C. necator* H16 strains were inoculated to an OD_{600} of 0.1 into 10 mL of minimal medium containing 4 g/L fructose and 50 µg/mL chloramphenicol and incubated at 30 °C with vigorous shaking in sealed

60-mL serum bottles. The medium was supplemented with either 1.25 mM L-rhamnose, 1.25 mM L-arabinose, 0.5 mM acrylate, or 3.13 µM cumate. Gas samples from the headspace were taken for isoprene analysis 18 h after induction.

3.2.7 Analytical methods

For isoprene quantification, gas samples were collected from the headspace of sealed 60-mL serum bottles containing 10 mL of culture. Isoprene was detected by gas chromatography (GC) using instrument Focus GC (Thermo Fisher Scientific) equipped with a flame ionisation detector and an HP-AL/S column (30 m length, 0.25 mm diameter, Agilent Technologies). Nitrogen gas was used as the carrier gas at a flow rate of 2 mL/min. The injector, oven, and detector temperatures were maintained at 220 °C, 120 °C, and 250 °C, respectively. The injection volume was 1 mL. The yields of isoprene per gram of CDW were estimated from standard curves generated by analysing known quantities of isoprene. CDW was determined by washing cells from a 20-mL culture in distilled water and separation by centrifugation, followed by vacuum-freeze-drying and weighting the cell pellet using a microbalance. GC analysis was performed with assistance of Naglis Malys.

3.3 Results

3.3.1 A modular reporter vector for evaluation of inducible systems

To compare different inducible systems for their ability to control gene expression in response to their corresponding effectors, an insulated and preferably standardised plasmid backbone is desirable. Consequently, a modular reporter vector was designed that served as the basis for the construction of all inducible systems (Figure 3.2). It comprises the following features: (i) a pBBR1MCS derived broad-host range vector replicon[117], which allows for replication in Gram-negative bacteria; (ii) compatibility with the pMTL vector series enabling rapid exchange of replication origin and

selection marker[137]; (iii) various restriction sites within the application-specific module to facilitate replacement of inducible system and reporter gene, and; (iv) transcriptional terminators flanking the application-specific module to prevent background reporter gene expression. The modular reporter plasmid which has been employed as origin for all other constructs, pEH006, contains the L-arabinose-inducible system. It was used to validate the modular reporter vector design and to make the results comparable to previously reported data. The vector was constructed in such a way that the TR gene is transcribed in opposite direction of the reporter gene encoding mRFP1[134] (Figure 3.2). The gene encoding the activator protein AraC is expressed from its native promoter to maintain TR-mediated autoregulation[138]. Moreover, the L-arabinose-inducible promoter P_{araBAD} was designed to harbour the bacteriophage T7 gene *10* (T7*g10*) mRNA stem-loop structure with the aim to enhance gene expression through improved RNA stability[139]. As a control vector, an identical plasmid lacking both the AraC TR and the L-arabinose-inducible promoter was assembled (denoted pEH006E). Plasmids pEH006 and pEH006E were transformed into *C. necator* H16. The resulting strains were cultured in minimal medium supplemented with sodium gluconate as carbon source and the single time-point fluorescence of the logarithmically growing cells was determined in the absence or presence of 0.1% (w/v) L-arabinose. At 6 h after induction, the L-arabinose-inducible system demonstrated more than a 1,200-fold increase in RFP synthesis (Figure 3.3). Fluorescence of cells containing pEH006E was similar to the background fluorescence of the medium, indicating no transcriptional read-through from the vector backbone. Based on these results, its modular design and insulating backbone makes pEH006 a suitable original vector for the construction and evaluation of inducible systems.

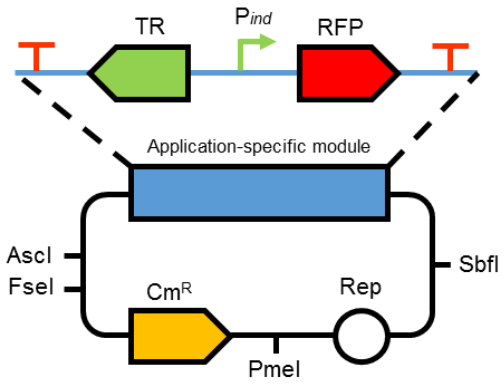


Figure 3.2 Schematic illustration of the modular reporter vector. It contains the four unique restriction sites AscI, FseI, PmeI, and SbfI which are used for modular assembly. The application-specific module harbours the *rfp* reporter gene and the inducible system composed of transcriptional regulator (TR) and inducible promoter (P_{ind}).

3.3.2 Design and construction of heterologous inducible system-reporter vectors

A significant number of transcription factor-based inducible systems has been reported to be utilised for controlling gene expression in other microorganisms[62]. In this chapter, some of these systems were selected from a variety of different prokaryotes to evaluate their potential to mediate gene expression in *C. necator* (Table 3.2). The activator-type TRs that were investigated include: (i) AlsR, which binds acetate in order to activate the acetoin forming pathway in *Bacillus subtilis*[140]; (ii) the *E. coli* propionate- and L-rhamnose-responsive PrpR and RhaRS[141, 142], and; (iii) MtlR from *Pseudomonas fluorescens* DSM50106, which activates transcription of the D-mannitol utilisation operon in response to D-mannitol, arabitol, and glucitol[132]. The repressor-type TRs that were investigated included: (i) AcuR, which regulates dimethylsulfoniopropionate (DMSP) degradation in *Rhodobacter sphaeroides* 2.4.1 and is induced by acrylate[122]; (ii) CymR, which activates transcription of the 4-isopropylbenzoic acid (cumate) degradation cluster of genes in *Pseudomonas putida* F1[143]; (iii) PhlF, which autoregulates 2,4-DAPG biosynthesis in *Pseudomonas fluorescens* CHA0, acting as an antimicrobial inhibitor of soil-borne plant

pathogens[144], and; (iv) XylR, which mediates utilisation of D-xylose in *Bacillus megaterium*[145].

Table 3.2 Summary of heterologous metabolite-inducible systems evaluated for gene expression control in *C. necator*.

Regulator name	Regulator type	Inducer	Origin	Reference
AlsR	Activator	Acetate	*B. subtilis*	[140]
PrpR	Activator	Propionate	*E. coli*	[141]
RhaRS	Activator	L-rhamnose	*E. coli*	[142]
MtlR	Activator	D-mannitol	*P. fluorescens*	[132]
AcuR	Repressor	Acrylate	*R. sphaeroides*	[122]
CymR	Repressor	Cumate	*P. putida*	[143]
PhlF	Repressor	2,4-DAPG	*P. fluorescens*	[144]
XylR	Repressor	D-xylose	*B. megaterium*	[145]

Inducible system-reporter vectors were designed identical to the L-arabinose-inducible system in the modular reporter vector pEH006, with the TR gene being transcribed in opposite direction of the *rfp* reporter gene (Figure 3.2). Similarly to AraC, the genes encoding the activator proteins are expressed from their native promoters to maintain TR-mediated autoregulation[132, 140-142]. For the negatively regulated systems, however, a constitutive promoter was required. Expression of the repressor protein must be well balanced to ensure both complete binding to the operator sequence under uninduced conditions and efficient dissociation from the operator in the presence of the ligand to allow gene transcription to be initiated. Here, the *E. coli lac* promoter including the *lacI* operator was used to control transcription of the genes encoding the transcriptional repressors. In *C. necator*, P_{lac} was shown to mediate moderate strength gene expression without the necessity of IPTG addition due to absence of LacI[102]. Expression of *rfp* is driven by the corresponding inducible promoter including TR binding sites. In cases where the transcription start site (TSS)

of the inducible promoter, including the acetate-, propionate-, L-rhamnose-, D-mannitol-, and acrylate-inducible promoter, was known, and the binding site of the TR is located upstream of the TSS, the sequence downstream of the TSS was replaced by the T7*g10* mRNA stem-loop structure sequence similarly to the L-arabinose-inducible system. For the cumate-inducible system, a synthetic promoter composed of the phage T5 promoter and the operator sequence of the *cmt* operon was employed[133]. The same strong RBS, a 20-nucleoide upstream sequence of the T7*g10*[146], was used in all of the constructs (except in case of the 2,4-DAPG-inducible system where the native RBS was employed).

3.3.3 Quantitative evaluation of heterologous inducible systems

C. necator strains carrying the different inducible systems were cultured in minimal medium supplemented with sodium gluconate as carbon source and the single time-point fluorescence of the logarithmically growing cells was determined in the absence or presence of their corresponding effectors. Inducers were used at concentrations that have been employed in previous studies and were as follows: 40 mM acetate[140], 10 mM propionate[147], 0.2% (w/v) L-rhamnose[148], 0.2% (w/v) D-mannitol[132], 2.5 mM acrylate[74], 10 µM cumate[133], 0.1 mM 2,4-DAPG[144], and 1% (w/v) D-xylose[149]. Of the analysed activator-type inducible systems, the L-rhamnose-inducible system mediated the highest induction (factor of nearly 2000-fold), mainly due to exceptionally low basal promoter activities (Figure 3.3). High background reporter gene expression in case of the positively regulated systems may result from metabolites naturally present in the cell that are structurally similar to the effector molecule. Of the analysed repressor-type inducible systems, the promoters controlled by AcuR and CymR showed the highest absolute normalised fluorescence levels. In their cases, high basal promoter activities may be caused by an insufficient amount of

repressor protein present in the cell to sufficiently repress transcription of the structural genes in the absence of the effector.

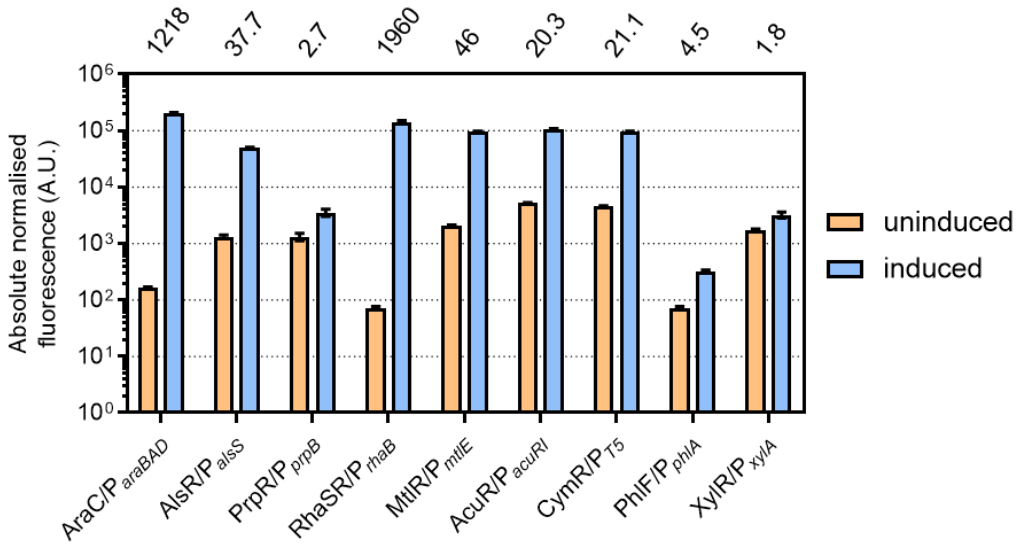


Figure 3.3 Quantitative evaluation of heterologous inducible systems. Single time-point fluorescence measurements of *C. necator* H16 carrying the L-arabinose-, acetate-, propionate-, L-rhamnose-, D-mannitol-, acrylate-, cumate-, 2,4-DAPG-, and D-xylose-inducible systems. The fluorescence output was determined in the absence of inducer and 6 h after extracellular supplementation with the corresponding effector. Ligand concentrations were as follows: 0.1% (w/v) L-arabinose, 40 mM acetate, 10 mM propionate, 0.2% (w/v) L-rhamnose, 0.2% (w/v) D-mannitol, 2.5 mM acrylate, 10 µM cumate, 0.1 mM 2,4-DAPG, and 1% (w/v) D-xylose. Induction factors are indicated. The gain factor was set manually to 100%. Error bars represent standard deviations of three biological replicates. The differences between the fluorescence output of the uninduced and induced samples are statistically significant for all evaluated inducible systems ($p < 0.01$; unpaired *t* test).

3.3.4 Induction kinetics of selected heterologous inducible systems in *C. necator* H16

Based on the results of the preliminary screen, four inducible systems, two positively and two negatively regulated ones, were selected to be further characterised. The activator-type systems controlled by L-arabinose and L-rhamnose were chosen mainly due to low basal promoter activities and high induction factors, whereas the repressor-type systems controlled by acrylate and cumate were selected based on their overall

performance. To provide a better overview of the system's architectures, their modes of action are illustrated in Figure 3.4.

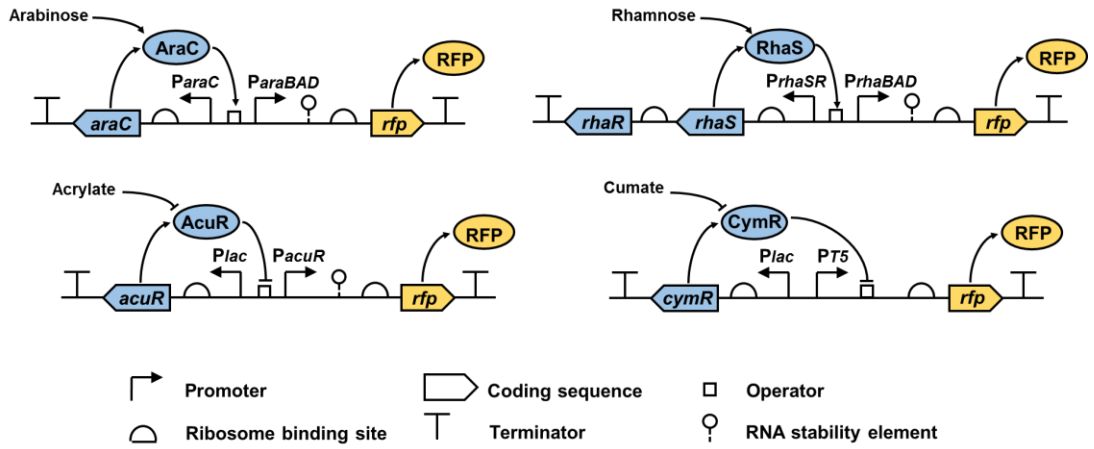


Figure 3.4 Architectures of the four selected heterologous inducible systems. Synthetic biology open language (SBOL)[150] visual representations of the reporter vector's application-specific modules that contain the L-arabinose-, L-rhamnose-, acrylate-, and cumate-inducible systems. SBOL visual icons are specified.

To gain insight into their induction kinetics, plasmids harbouring the four selected inducible systems were transformed into *C. necator* H16 and analysed for fluorescent protein reporter gene expression over time at different inducer concentrations (Figure 3.5a). The resulting dose-response curves provide information about each system's dynamic range (Figure 3.5b). For example, gene expression controlled by the L-arabinose-inducible system can be fine-tuned in the range between 0.313 and 2.5 mM L-arabinose for a linear output. The exponential increase in fluorescence output stretches more widely, between 0.016 and 1.25 mM. Furthermore, absolute normalised fluorescence values were used to calculate the RFP synthesis rate at any time point during the time course of experiment (see Chapter 3.2.5). The resulting maximum synthesis rate was fit to the corresponding inducer concentration using a Hill function (Figure 3.6), yielding key parameters, such as the maximum possible rate of RFP synthesis, the Hill coefficient, or the inducer concentration mediating half-maximal RFP synthesis K_m (Table 3.3). According to the resulting parameters, the L-rhamnose-

inducible system demonstrates the highest induction cooperativity (h = 3.88). It requires a minimum concentration of 0.156 mM L-rhamnose to be activated and achieves about 85% of maximum expression at 2.5 mM. The acrylate- and cumate-inducible systems generally require lower inducer levels to initiate gene expression. Moreover, the range of inducer concentration mediating a linear fluorescence output spans more than one order of magnitude (5 to 125 µM and 0.08 to 1.56 µM, for acrylate and cumate, respectively), and therefore can be fine-tuned more easily.

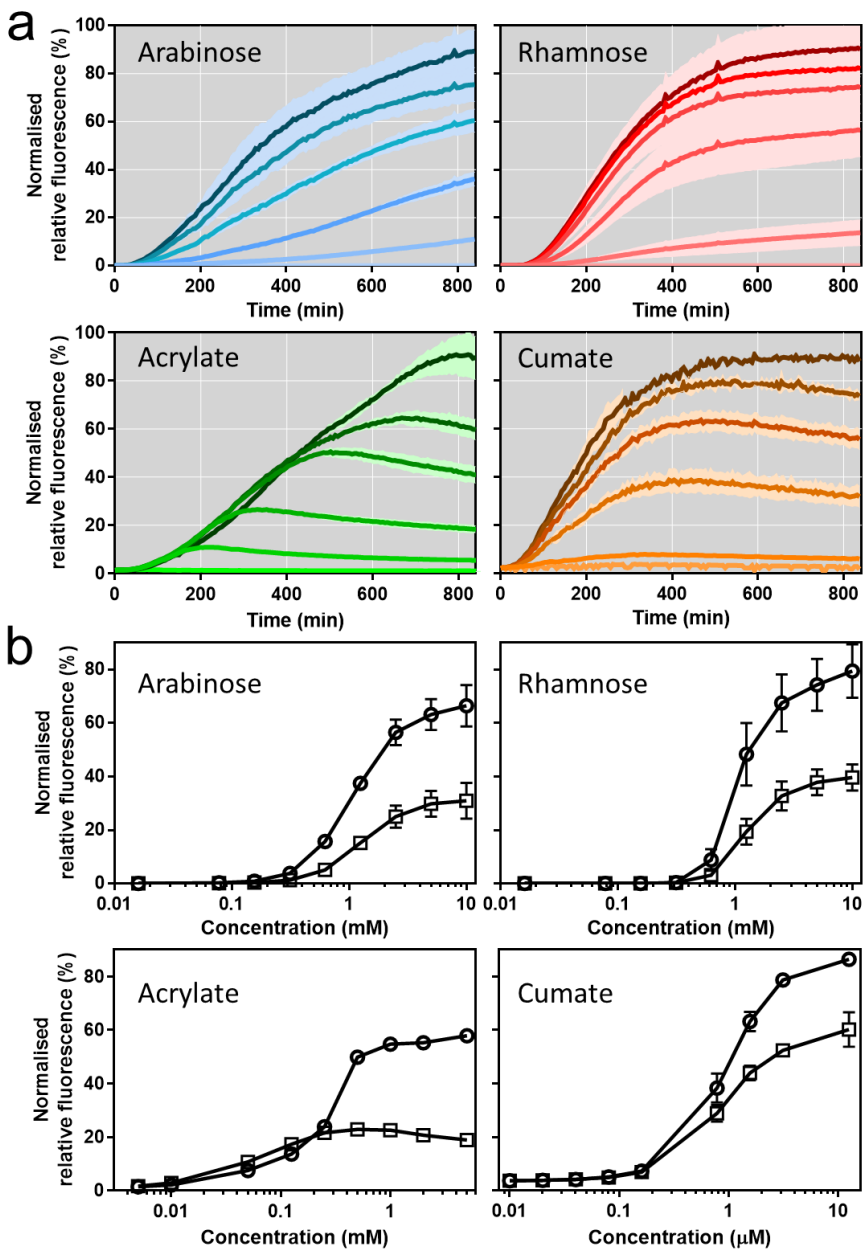


Figure 3.5 Induction kinetics and dose-response of the L-arabinose-, L-rhamnose-, acrylate-, and cumate-inducible systems. (a) Normalised relative fluorescence of *C. necator* H16 carrying the L-arabinose-, L-rhamnose-, acrylate-, and cumate-inducible systems (pEH006, pEH002, pEH020, and pEH040, respectively). Inducers were added at time zero and fluorescence was monitored for 14 h. The darker the colour shade, the higher the inducer concentration. L-arabinose was supplemented to a final concentration of 10, 2.5, 1.25, 0.625, 0.313 mM and no inducer. L-rhamnose was supplemented to a final concentration of 10, 5, 2.5, 1.25, 0.625 mM and no inducer. Acrylate was supplemented to a final concentration of 5, 1, 0.5, 0.25, 0.05 mM and no inducer. Cumate was supplemented to a final concentration of 12.5, 3.13, 1.56, 0.78, 0.16 µM and no inducer. The standard deviation of three biological replicates is illustrated as lighter colour ribbon displayed lengthwise of the induction kinetics curve. (b) Dose-response of *C. necator* H16 carrying the L-arabinose-, L-rhamnose-, acrylate-, and cumate-inducible systems 4 (square) and 8 (circle) h after inducer addition. Error bars represent standard deviations of three biological replicates.

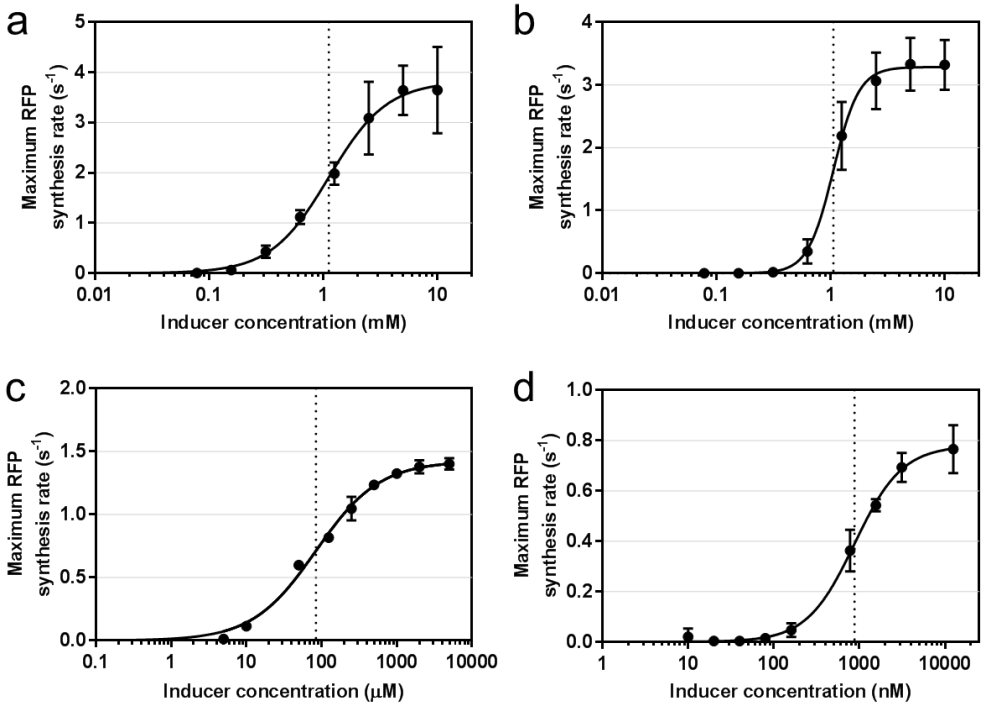


Figure 3.6 Promotor activity of the L-arabinose, L-rhamnose, acrylate-, and cumate-inducible systems. The normalised maximum rate of RFP synthesis was fit to the corresponding inducer concentration using a Hill function for the (a) L-arabinose-, (b) L-rhamnose-, (c) acrylate-, and (d) cumate-inducible systems. Error bars represent standard deviations of three biological replicates. The inducer concentration that mediates half-maximal RFP synthesis K_m is indictaed by a dotted line.

Table 3.3 Minimum and maximum reporter protein synthesis rate of the L-arabinose-, L-rhamnose-, acrylate-, and cumate-inducible systems. RFP synthesis rates were fit using a Hill function yielding key parameters, including the Hill coefficient *h* and the inducer concentration that mediates half-maximal RFP synthesis K_m.

Plasmid identifier	Inducer	Minimum RFP synthesis rate, v_{min} (s^{-1})	Maximum RFP synthesis rate, v_{max} (s^{-1})	Hill coefficient, h	Half maximal RFP synthesis, K_m (μM)	Induction factor
pEH002	L-rhamnose	0.0058 ± 0.0022	3.292 ± 0.032	3.88 ± 0.23	1057 ± 19	1960
pEH006	L-arabinose	0.0122 ± 0.0023	3.836 ± 0.085	1.70 ± 0.11	1116 ± 56	1232
pEH020	Acrylate	0.0474 ± 0.0056	1.471 ± 0.035	1.03 ± 0.09	84.32 ± 8.08	33
pEH040	Cumate	0.0472 ± 0.0076	0.828 ± 0.011	1.56 ± 0.08	0.877 ± 0.029	22

3.3.5 Acrylate is consumed by *C. necator* H16

Regardless of the inducer concentration, the absolute fluorescence, corrected by fluorescence that derives from basal promoter activity, generally increased during the time course of experiment. This was not the case for the acrylate-inducible system. At 6 h after inducer addition, the increase in absolute fluorescence output facilitated by acrylate concentrations of 1.25 mM and less was at the same level as mediated by basal P_{acuRI} activity. This behaviour is also reflected in its dose response curve (Figure 3.5b). Normalised fluorescence values for acrylate concentrations of 1.25 mM and less were lower after 8 h of induction than that after 4 h, whereas acrylate concentrations of more than 1.25 mM enabled an extended expression of the reporter gene. This type of transient gene expression was hypothesised to be caused by inducer metabolism[151]. To test whether acrylate is catabolised by *C. necator* H16, a metabolite consumption assay was performed. As predicted, acrylate was co-consumed simultaneously with the primary carbon source fructose (Figure 3.7). At 4 h after supplementation with 5

mM acrylate, 49% of the initial amount was consumed. Upon depletion of inducer, gene expression was maintained at the level of basal promoter activity. *C. necator* H16 was not able to grow in minimal medium supplemented with either of the three inducers L-arabinose, L-rhamnose, or cumate as sole carbon source.

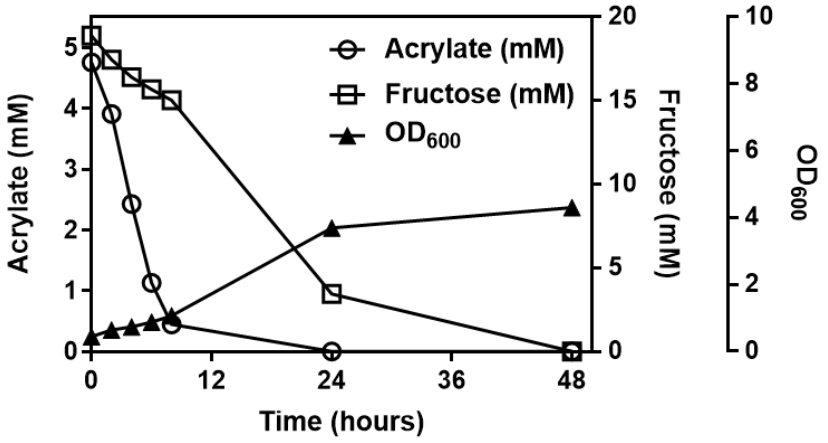


Figure 3.7 Consumption of acrylate in *C. necator* H16. Acrylate was added at time zero. Consumption of acrylate and the primary carbon source fructose was monitored by HPLC-UV analysis of the culture supernatant over the time course of 48 h. The means of three biological replicates are presented. Error bars are too small to be visible.

3.3.6 Induction factors of the selected heterologous inducible systems

In addition to their dynamic range, another important characteristic of inducible systems is their induction factor. It was calculated for cells in exponential growth phase, 6 h after the inducer was added. Dividing the maximum normalised fluorescence resulting from the highest inducer concentration by the normalised fluorescence of the uninduced sample yielded induction factors of 1232, 1960, 33, and 22 for the L-arabinose-, L-rhamnose-, acrylate-, and cumate-inducible systems, respectively (Table 3.3). The induction factors that were achieved by AcuR/P_{acuRI} and CymR/P_{cmt} were much lower than for the two positively regulated systems due to high background levels of reporter gene expression in the absence of inducer (Figure 3.8). The order from the highest to the lowest normalised absolute fluorescence achieved

by the four tested systems throughout the time course of experiment for the highest inducer concentration is as follows: L-arabinose > L-rhamnose > acrylate > cumate. The same order applies to the maximum possible RFP synthesis rate with L-arabinose demonstrating the highest value (Table 3.3). All of the systems can be considered to activate reporter gene transcription immediately after inducer addition. Fluorescence above background levels could be detected within 30 min, representing the time which is required for RFP synthesis and maturation[134].

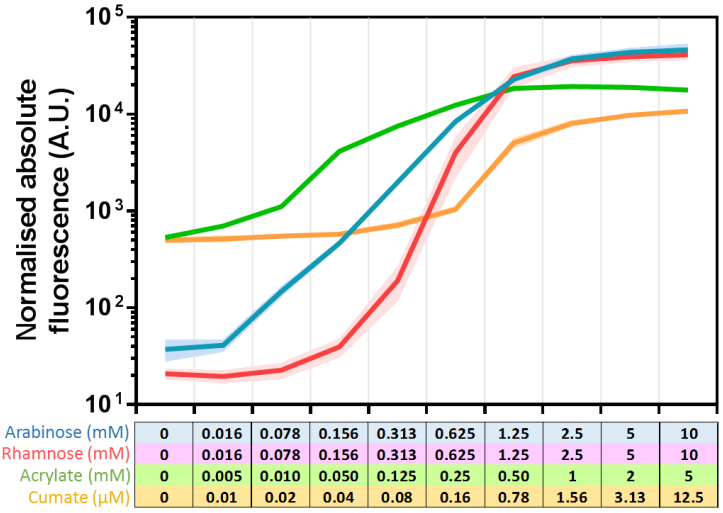


Figure 3.8 Induction dynamics of the L-arabinose-, L-rhamnose-, acrylate-, and cumate-inducible systems. Normalised absolute fluorescence of *C. necator* H16 carrying the L-arabinose-, L-rhamnose-, acrylate-, and cumate-inducible systems. Fluorescence was determined for cells in exponential growth phase 6 h after inducer addition. Inducer concentrations are indicated for each system. The standard deviation of three biological replicates is illustrated as lighter colour ribbon displayed lengthwise of the dynamics curve.

3.3.7 Influence of the T7*g10* mRNA stem-loop structure sequence on inducible gene expression

To evaluate the influence of the T7*g10* mRNA stem-loop structure sequence on inducible gene expression, it was removed from the plasmid containing the L-arabinose-inducible system, resulting in vector pEH176. The relationship between fluorescence response and inducer concentration for the newly designed L-arabinose-

inducible system was similar to the construct harbouring the stem-loop (Figure 3.9). A linear fluorescence output was achieved by addition of L-arabinose from 0.313 to 2.5 mM. The induction is considerably higher (factor of 2,900-fold), mainly due to an 8-fold lower background level of RFP synthesis. However, removing the stem-loop structure sequence also decreased absolute normalised fluorescence levels by 3.6-fold at an inducer concentration of 10 mM (Figure 3.8 and Figure 3.9).

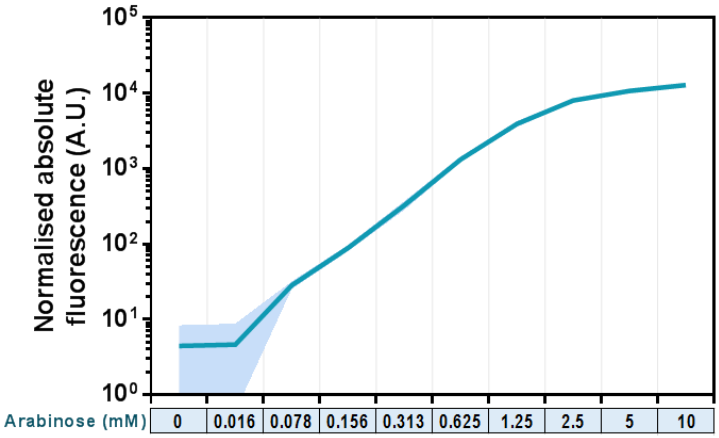


Figure 3.9 Induction dynamics of the L-arabinose-inducible system lacking the stem-loop structure sequence (pEH176). Normalised absolute fluorescence of *C. necator* H16 cells carrying pEH176 in exponential growth phase 6 h after L-arabinose addition. The standard deviation of three biological replicates is illustrated as lighter colour ribbon displayed lengthwise of the dynamics curve.

3.3.8 Control of isoprene biosynthesis using inducible promoters

To establish whether production of isoprene can be achieved in *C. necator* H16, and to investigate how gene expression of the enzyme (isoprene synthase) translates into the product of the enzymatic reaction (isoprene), the *Populus alba ispS* gene under the control of the two positively (AraC/P_{araBAD} and RhaRS/P_{rhaBAD}) and two negatively (AcuR/P_{acuRI} and CymR/P_{cmt}) regulated inducible systems was introduced into *C. necator* on a multicopy episomal vector based on the pBBR1 replicon. Induction of *ispS* expression with L-arabinose, L-rhamnose, acrylate, or cumate confirmed that isoprene was biosynthesised to different levels, resulting in up to 7 µg/g of cells (dry

weight) (Figure 3.10). Moreover, the isoprene yield showed a moderate positive linear correlation with gene expression (measured as fluorescence output) of their corresponding inducible systems (r = 0.625, only data of induced samples were used in the analysis).

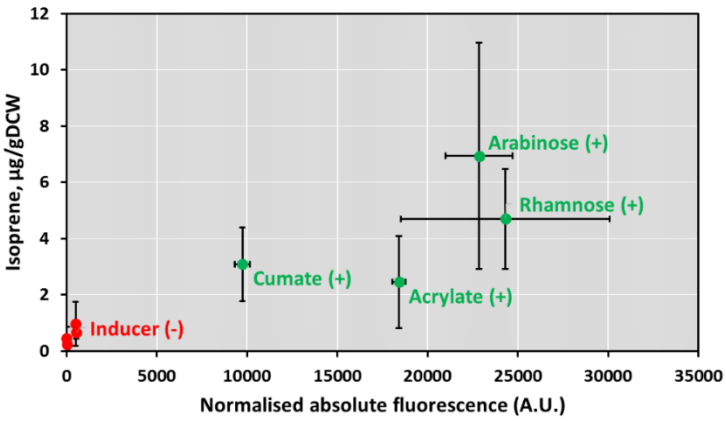


Figure 3.10 Correlation between fluorescent protein reporter gene expression levels and isoprene yields using the L-arabinose, L-rhamnose, acrylate-, and cumate-inducible systems. Normalised absolute fluorescence of *C. necator* H16 strains carrying the AraC/P_{araBAD}-L-arabinose-, RhaRS/P_{rhaBAD}-L-rhamnose-, AcuR/P_{acuRI}-acrylate-, and CymR/P_{cmt}-cumate-inducible systems is plotted against isoprene yield, resulting from strains carrying the *ispS* gene under control of the corresponding inducible system in the presence [(+), green dots] or absence [(-), red dots] of inducers. The plotted values represent the normalised fluorescence 6 h and isoprene yield 18 h after induction with 1.25 mM L-arabinsoe, 1.25 mM L-rhamnose, 0.5 mM acrylate, or 3.13 µM cumate. Error bars represent standard deviations from three biological replicates.

3.4 Discussion

Several heterologous inducible promoters, mediating gene expression in response to the effectors L-arabinose, *m*-toluic acid, lactose, L-rhamnose, and cumate, have been individually characterised previously[86, 102, 109, 110]. In this work, a more systematic approach has been taken by evaluating nine heterologous inducible systems using standardised experimental conditions. Four of them, two positively (L-arabinose and L-rhamnose) and two negatively (acrylate and cumate) regulated inducible systems were selected to be characterised in more detail. Using minimal medium and the formulated experimental design, the positively regulated inducible systems exhibited

a tighter control of gene expression than the negatively regulated ones. The highest level of reporter output was achieved using the L-arabinose-inducible system. The L-rhamnose-inducible system exhibited the highest induction factor mainly due to basal promoter activities just above the level of background autofluorescence. The acrylate- and cumate-inducible systems showed significantly lower induction and higher background levels. The application of the acrylate-inducible system in *C. necator* is limited due to effector consumption by this bacterium. The L-arabinose-, L-rhamnose-, and cumate-inducible systems are suitable for continuous activation of gene expression as well as biosensors. Notably, the cumate-inducible system appeared to be most sensitive, responding to a nanomolar to micromolar range of inducer concentrations. Overall, the systems' dose-responses, as well as the influence of the mRNA stem-loop structure sequence on gene expression, are consistent with previous induction experiments with *C. necator* and *E. coli*[74, 109, 110]. For example, the L-rhamnose-inducible system has been reported to possess a narrow range of inducer concentration over which gene expression can be fine-tuned[109]. Furthermore, the sensitivity of the cumate-inducible system could be confirmed[86, 110]. However, even though the induction factor of the L-arabinose-inducible system, obtained using minimal medium supplemented with the carbon sources sodium gluconate or fructose, is roughly the same, the induction factors of both positively regulated systems were higher than in other published studies[86, 102, 109]. This variation between independent studies may be the result of: (i) transcriptional read-through from the reporter vector backbone; (ii) the use of complex medium, such as LB, instead of minimal medium which might contain compounds that are structurally similar to the effector and cross-react with the inducible system; (iii) the use of a fluorescent reporter protein, such as GFP, whose excitation and emission wavelengths overlap with chemical species

naturally present in living cells, resulting in elevated background fluorescence levels, and; (iv) the time-point at which the sample was taken. Ideally, the fluorescence output of logarithmically growing cells should be measured. Often, however, the induction factor is calculated using samples taken at the end of the cultivation, which might result in incorrect values. Firstly, background fluorescence levels often increase at later stages of the cultivation as more chemical species are released into the culture medium due to cell lysis. Secondly, fluorescent proteins, unless tagged with a degradation signal, are very stable and will accumulate in cells even though they have entered the stationary growth phase. All the above-mentioned factors contribute to variation and exemplify the importance of comparative studies and standardised experimental conditions. Lastly, it was demonstrated that *C. necator* can be used as microbial cell factory for the biosynthesis of the industrially relevant building block chemical isoprene. The variation in gene expression by using different inducible systems was shown to correlate with product biosynthesis. Future research may be directed towards improving the metabolic flux for the autotrophic biosynthesis of isoprene.

3.5 Conclusion

A modular reporter vector was constructed that served as the basis for the evaluation of a range of heterologous transcription factor-based inducible gene expression systems in *C. neactor*. Four of them, the L-arabinose-, L-rhamnose, acrylate-, and cumate-inducible systems were characterised in detail for their range of induction. The dynamic fluorescence data was used to calculate the maximum level of protein synthesis for each of the selected inducible systems with the L-arabinose-inducible system showing the highest rate of RFP synthesis. In correlation with the fluorescence

data, the L-arabinose-inducible system mediated the highest production of isoprene in *C. necator*.

Chapter 4

The following chapter is mainly based on the work presented in the publication:

A genome-wide approach for identification and characterisation of metabolite-inducible systems

E. K. R. Hanko, A. C. Paiva, M. Jonczyk, M. Abbott, N. P. Minton & N. Malys (2020)

Nature Communications 11 (1), 1-14.[152]

The accepeted manuscript can be found in the *Appendix*.

4 A genome-wide approach for identification and characterisation of metabolite-inducible systems

4.1 Introduction

Inducible gene expression systems execute a pivotal role in establishing a sustainable balance of gene expression and protein synthesis at the genome or single pathway/circuit level in response to changes in the intra- and extracellular environment. Such systems have been historically utilised for gene overexpression and protein production. Nowadays, inducible systems and their underlying genetic elements have become essential tools in synthetic biology[153]. Initially harnessed for the design of synthetic regulatory circuits[47], metabolite-responsive TRs and their cognate inducible promoters have received increasing attention due to their application as genetically encoded biosensors[62, 154-156]. Although whole-cell biosensors or cell-free transcription/translation systems have advanced the fields of clinical diagnostics, environmental remediation, spatiotemporal regulation of signalling networks and metabolic engineering[157-161], the number of compounds that can be detected is still limited. Thus, to increase diversity and to offer unique specificities, novel inducible systems must be sought and researched.

In instances where an inducible system is to be found for a specific effector molecule, transcriptome analyses[75, 162] and the evaluation of promoter libraries[76] have proven to be efficient strategies to discover effector-responsive promoters. They do not, however, exclude promoters that are indirectly activated nor guarantee identification of their associated TRs. Some of these issues may be solved by cloning sequence clusters containing TR-promoter pairs as demonstrated in the screening of metagenome libraries[77], but this methodology relies on the inducible system being functional in an organism different to the one from which it was sourced. The reverse

strategy relies on predicting the effector molecule based on genetic context[78] or comparative genomics[80, 81]. This approach has successfully resulted in the identification of effectors and their corresponding TR-promoter pairs, but it is limited to specific families of TRs and specific classes of compounds.

This chapter addresses the deficiencies associated with the identification of metabolite-responsive inducible systems by pairing TR genes with gene clusters responsible for the catabolism of the corresponding ligand and establishes a generalised workflow to discover new native systems independent of their belonging to a specific family of regulators or the class of compounds they respond to. In total, 15 novel inducible systems are mined from the genome of *C. necator* H16 and their broad-host range of applicability demonstrated in three industrially relevant microorganisms. Also addressed is a typical issue that may arise when employing a system in a host organism different to the one from which it was sourced. To facilitate forward engineering efforts, the identified inducible systems are parameterised and their utility for controlling orthogonal gene expression demonstrated. Finally, their potential to be applied for investigation of metabolism and to expand the number of biologically detectable chemical species is highlighted by evaluating the cross-reactivity between the library of biosensors and a comprehensive list of selected compounds.

4.2 Materials and methods specific to this chapter

4.2.1 Chemicals

β-Alanine, sodium formate, xanthine, phenylglyoxylic acid, sodium salicylate, sodium benzoate, potassium sodium tartrate, sulfoacetic acid, L-kynurenine, 3-HP, L-phenylalanine, γ-aminobutyric acid (GABA), L-tyrosine disodium salt, cyclohexanecarboxylic acid, 3,4-dihydroxybenzoic acid, 3-hydroxy-2-butanone

(acetoin), L-arabinose, L-rhamnose monohydrate, magnesium acrylate, 4-isopropylbenzoic acid (cumate), itaconic acid, glycine, L-alanine, 3-aminobutanoic acid, propionic acid, butyric acid, phenyl acetate, D,L-2,3-diaminopropionic acid monohydrochloride, D,L-isoserine (D,L-3-amino-2-hydroxypropanoate), D,L-2-phenylglycine, glyoxylic acid monohydrate, D,L-mandelic acid, hippuric acid (N-benzoylglycine), 1-cyclohexene-1-carboxylic acid, shikimic acid, cyclopentanecarboxylic acid, hypoxanthine, caffeine, theobromine, uracil, tartronic acid (hydroxypropanedioate), L-α-hydroxyglutaric acid disodium salt, L-malic acid, sodium succinate dibasic hexahydrate, L-aspartic acid potassium salt, oxalacetic acid, sodium L-lactate, sodium D-lactate, levulinic acid (4-oxopentanoate), sodium fumarate dibasic, taurine, isethionic acid sodium salt, catechol, resorcinol, hydroquinone, phenol, L-glutamine, dopamine hydrochloride, cis,cis-muconic acid, 4-hydroxybenzoic acid, 2,6-dihydroxybenzoic acid, L-tryptophan, nicotinic acid, 4-hydroxyquinoline-2-carboxylic acid hydrate (kynurenic acid), D-saccharic acid potassium salt (glucarate), D-mannitol, and tricarballylic acid were used as inducers for assaying the native inducible systems. All chemicals are listed in Supplementary Table 1 in the *Appendix*.

4.2.2 Plasmids

Key features of all plasmids used and generated in this chapter are summarised in Table 4.1. A detailed description of how each plasmid was assembled is provided in the *Appendix*. The gene encoding the transcriptional regulator of the β-alanine-inducible system, *oapR*, was optimised for *E. coli* codon usage (*oapR*co*Ec*) and synthesised by Life Technologies. The corresponding DNA sequence can be found in the *Appendix*. Two versions of each plasmid containing an inducible system or an inducible promoter were constructed. One of the two versions contains a

chloramphenicol resistance gene, the other one confers resistance to tetracycline. The former version was employed for evaluation of reporter gene expression in *C. necator* and *E. coli*, whereas the latter version was used in *P. putida*.

Table 4.1 Plasmids used in chapter 4.

Plasmid	Characteristic	Reference or source
p17ACKHEP	Kan^r; $P_{H16\_RS01325}$-*oapR*co*Ec*, $P_{H16\_RS01330}$	Life Technologies
pJOE7801.1	Kan^r; P_{tetR}-*tetR*; P_{tetA}-*egfp*-T_{rrnB}	[132]
pME6000	Tet^r; P_{lac}-*lacZα*	[163]
pEH006	Cm^r; P_{araC}-*araC*-T_{rrnB1}; P_{araBAD}-T7sl-*rfp*-T_{dbl}	This work
pEH006E	Cm^r; T_{rrnb1}-*rfp*-T_{dbl}	This work
pEH010	Cm^r; $P_{H16\_RS18300}$-*hpdR*-T_{rrnb1}, $P_{H16\_RS18295}$-*rfp*-T_{dbl}	This work
pEH035	Cm^r; $P_{H16\_RS18295}$-*rfp*-T_{dbl}	This study
pEH042	Cm^r; $P_{H16\_RS08130}$-*nahR*-T_{rrnb1}, $P_{H16\_RS08125}$-*rfp*-T_{dbl}	This study
pEH052	Cm^r; $P_{H16\_RS19440}$-*acoR*-T_{rrnb1}, $P_{H16\_RS19445}$-*rfp*-T_{dbl}	This study
pEH083	Cm^r; $P_{H16\_RS10670}$-*ttdR*-T_{rrnb1}, $P_{H16\_RS10665}$-*rfp*-T_{dbl}	This study
pEH095	Cm^r; $P_{H16\_RS08125}$-*rfp*-T_{dbl}	This study
pEH096	Cm^r; $P_{H16\_RS19445}$-*rfp*-T_{dbl}	This study
pEH097	Cm^r; $P_{H16\_RS10665}$-*rfp*-T_{dbl}	This study
pEH101	Cm^r; $P_{H16\_RS01330}$-*rfp*-T_{dbl}	This study
pEH134	Cm^r; $P_{H16\_RS03160}$-*fdsR*-T_{rrnb1}, $P_{H16\_RS03165}$-*rfp*-T_{dbl}	This study
pEH136	Cm^r; $P_{H16\_RS14025}$-*kynR*-T_{rrnb1}, $P_{H16\_RS14030}$-*rfp*-T_{dbl}	This study
pEH137	Cm^r; $P_{H16\_RS18360}$-*phhR*-T_{rrnb1}, $P_{H16\_RS18365}$-*rfp*-T_{dbl}	This study
pEH147	Cm^r; $P_{H16\_RS01325}$-*oapR*-T_{rrnb1}, $P_{H16\_RS01330}$-*rfp*-T_{dbl}	This study
pEH148	Cm^r; $P_{H16\_RS09795}$-*benM*-T_{rrnb1}, $P_{H16\_RS09790}$-*rfp*-T_{dbl}	This study
pEH149	Cm^r; $P_{H16\_RS09790}$-*rfp*-T_{dbl}	This study
pEH151	Cm^r; $P_{H16\_RS03165}$-*rfp*-T_{dbl}	This study
pEH152	Cm^r; $P_{H16\_RS14030}$-*rfp*-T_{dbl}	This study
pEH153	Cm^r; $P_{H16\_RS18365}$-*rfp*-T_{dbl}	This study
pEH154	Cm^r; $P_{H16\_RS05060}$-*xdhR*-T_{rrnb1}, $P_{H16\_RS05055}$-*rfp*-T_{dbl}	This study
pEH155	Cm^r; $P_{H16\_RS05525}$-*phgR*-T_{rrnb1}, $P_{H16\_RS05530}$-*rfp*-T_{dbl}	This study
pEH156	Cm^r; $P_{H16\_RS23650}$-*gabR*-T_{rrnb1}, $P_{H16\_RS23655}$-*rfp*-T_{dbl}	This study
pEH157	Cm^r; $P_{H16\_RS13690}$-*sauR*-T_{rrnb1}, $P_{H16\_RS13695}$-*rfp*-T_{dbl}	This study

pEH158	Cm^r; $P_{H16\_RS24180}$-*hpdA*-T_{rrnb1}, $P_{H16\_RS24175}$-*rfp*-T_{dbl}	This study
pEH159	Cm^r; $P_{H16\_RS27205}$-*badR*-T_{rrnb1}, $P_{H16\_RS27200}$-*rfp*-T_{dbl}	This study
pEH160	Cm^r; $P_{H16\_RS29645}$-*H16_RS29645*-T_{rrnb1}, $P_{H16\_RS29650}$-*rfp*-T_{dbl}	This study
pEH161	Cm^r; $P_{H16\_RS30150}$-*pcaQ*-T_{rrnb1}, $P_{H16\_RS30145}$-*rfp*-T_{dbl}	This study
pEH167	Cm^r; $P_{H16\_RS05055}$-*rfp*-T_{dbl}	This study
pEH168	Cm^r; $P_{H16\_RS13695}$-*rfp*-T_{dbl}	This study
pEH169	Cm^r; $P_{H16\_RS24175}$-*rfp*-T_{dbl}	This study
pEH170	Cm^r; $P_{H16\_RS27200}$-*rfp*-T_{dbl}	This study
pEH171	Cm^r; $P_{H16\_RS30145}$-*rfp*-T_{dbl}	This study
pEH173	Cm^r; $P_{H16\_RS01325}$-*oapR*co*Ec*-T_{rrnb1}, $P_{H16\_RS01330}$-*rfp*-T_{dbl}	This study
pEH176	Cm^r; P_{araC}-*araC*-T_{rrnb1}, P_{araBAD}-*Ec*RBS-*rfp*-T_{dbl}	This study
pEH194	Tet^r; $P_{H16\_RS08130}$-*nahR*-T_{rrnb1}, $P_{H16\_RS08125}$-*rfp*-T_{dbl}	This study
pEH195	Tet^r; $P_{H16\_RS08125}$-*rfp*-T_{dbl}	This study
pEH196	Tet^r; $P_{H16\_RS01330}$-*rfp*-T_{dbl}	This study
pEH197	Tet^r; $P_{H16\_RS18300}$-*hpdR*-T_{rrnb1}, $P_{H16\_RS18295}$-*rfp*-T_{dbl}	This study
pEH198	Tet^r; $P_{H16\_RS18295}$-*rfp*-T_{dbl}	This study
pEH199	Tet^r; $P_{H16\_RS10670}$-*ttdR*-T_{rrnb1}, $P_{H16\_RS10665}$-*rfp*-T_{dbl}	This study
pEH200	Tet^r; $P_{H16\_RS10665}$-*rfp*-T_{dbl}	This study
pEH201	Tet^r; $P_{H16\_RS19445}$-*rfp*-T_{dbl}	This study
pEH202	Tet^r; $P_{H16\_RS03160}$-*fdsR*-T_{rrnb1}, $P_{H16\_RS03165}$-*rfp*-T_{dbl}	This study
pEH203	Tet^r; $P_{H16\_RS03165}$-*rfp*-T_{dbl}	This study
pEH204	Tet^r; $P_{H16\_RS05055}$-*rfp*-T_{dbl}	This study
pEH205	Tet^r; $P_{H16\_RS09795}$-*benM*-T_{rrnb1}, $P_{H16\_RS09790}$-*rfp*-T_{dbl}	This study
pEH206	Tet^r; $P_{H16\_RS09790}$-*rfp*-T_{dbl}	This study
pEH207	Tet^r; $P_{H16\_RS13695}$-*rfp*-T_{dbl}	This study
pEH208	Tet^r; $P_{H16\_RS14025}$-*kynR*-T_{rrnb1}, $P_{H16\_RS14030}$-*rfp*-T_{dbl}	This study
pEH209	Tet^r; $P_{H16\_RS14030}$-*rfp*-T_{dbl}	This study
pEH210	Tet^r; $P_{H16\_RS18360}$-*phhR*-T_{rrnb1}, $P_{H16\_RS18365}$-*rfp*-T_{dbl}	This study
pEH211	Tet^r; $P_{H16\_RS18365}$-*rfp*-T_{dbl}	This study
pEH212	Tet^r; $P_{H16\_RS24180}$-*hpdA*-T_{rrnb1}, $P_{H16\_RS24175}$-*rfp*-T_{dbl}	This study
pEH213	Tet^r; $P_{H16\_RS24175}$-*rfp*-T_{dbl}	This study
pEH214	Tet^r; $P_{H16\_RS27205}$-*badR*-T_{rrnb1}, $P_{H16\_RS27200}$-*rfp*-T_{dbl}	This study
pEH215	Tet^r; $P_{H16\_RS27200}$-*rfp*-T_{dbl}	This study
pEH216	Tet^r; $P_{H16\_RS30150}$-*pcaQ*-T_{rrnb1}, $P_{H16\_RS30145}$-*rfp*-T_{dbl}	This study
pEH217	Tet^r; $P_{H16\_RS30145}$-*rfp*-T_{dbl}	This study

pEH218	Tet^r; $P_{H16\_RS05525}$-*phgR*-T_{rrnb1}, $P_{H16\_RS05530}$-*rfp*-T_{dbl}	This study
pEH220	Tet^r; P_{araC}-*araC*-T_{rrnb1}, P_{araBAD}-*Ec*RBS-*rfp*-T_{dbl}	This study
pEH221	Tet^r; $P_{H16\_RS01325}$-*oapR*-T_{rrnb1}, $P_{H16\_RS01330}$-*rfp*-T_{dbl}	This study
pEH222	Tet^r; $P_{H16\_RS19440}$-*acoR*-T_{rrnb1}, $P_{H16\_RS19445}$-*rfp*-T_{dbl}	This study
pEH223	Tet^r; $P_{H16\_RS05060}$-*xdhR*-T_{rrnb1}, $P_{H16\_RS05055}$-*rfp*-T_{dbl}	This study
pEH224	Tet^r; $P_{H16\_RS13690}$-*sauR*-T_{rrnb1}, $P_{H16\_RS13695}$-*rfp*-T_{dbl}	This study
pEH225	Cm^r; P_{13}-*oapR*-T_{rrnb1}, $P_{H16\_RS01330}$-*rfp*-T_{dbl}	This study
pEH226	Cm^r; P_{13}-*oapR*co*Ec*-T_{rrnb1}, $P_{H16\_RS01330}$-*rfp*-T_{dbl}	This study
pEH229	Cm^r; $P_{H16\_RS05530}$-*rfp*-T_{dbl}	This study
pEH234	Tet^r; P_{13}-*oapR*-T_{rrnb1}, $P_{H16\_RS01330}$-*rfp*-T_{dbl}	This study
pEH235	Tet^r; P_{13}-*oapR*co*Ec*-T_{rrnb1}, $P_{H16\_RS01330}$-*rfp*-T_{dbl}	This study
pEH240	Cm^r; $P_{H16\_RS29645}$-*H16_RS29645*-T_{rrnb1}, $P_{H16\_RS29650}$-*H16_RS29650-H16_RS29655*-15:*rfp*-T_{dbl}	This study
pEH256	Cm^r; $P_{H16\_RS23650}$-*gabR*-T_{rrnb1}, $P_{H16\_RS23655}$-*H16_RS23655-H16_RS23660*-9::*rfp*-T_{dbl}	This study
pEH257	Tet^r; $P_{H16\_RS05530}$-*rfp*-T_{dbl}	This study
pEH263	Tet^r; $P_{H16\_RS08130}$-*nahR*-$T_{luxICDABEG}$, $P_{H16\_RS08125}$-*egfp*-T_{rrnb1}, $P_{H16\_RS30150}$-*pcaQ*-$T_{luxICDABEG}$, $P_{H16\_RS30145}$-*rfp*-T_{rrnb2}	This study
pEH266	Cm^r; $P_{H16\_RS23655}$-*H16_RS23655-H16_RS23660*-9::*rfp*-T_{dbl}	This study
pEH268	Tet^r; $P_{H16\_RS23650}$-*gabR*-T_{rrnb1}, $P_{H16\_RS23655}$-*H16_RS23655-H16_RS23660*-9::*rfp*-T_{dbl}	This study
pEH269	Tet^r; $P_{H16\_RS23655}$-*H16_RS23655-H16_RS23660*-9::*rfp*-T_{dbl}	This study

4.2.3 Growth conditions

For single time-point fluorescence measurements and flow cytometric analyses, bacterial strains were propagated in LB medium. To determine the dose-responses and to evaluate inducer-TR cross-reactivity, *C. necator* was cultivated in minimal medium containing 4 g/L sodium gluconate.

4.2.4 Fluorescence measurements

Single time-point and time course measurements were performed as described in Chapter 2.5.2. For the time course measurements, the gain factor was set to 80%. The fluorescence and absorbance values, recorded 80 min after the inducer had been added, were obtained from the time course data and used to calculate the absolute normalised fluorescence values corresponding to each inducer concentration in order to generate the dose-response curves.

4.2.5 Mathematical modelling

To obtain system parameters that can be used for synthetic circuit design, absolute normalised fluorescence values (*RFP*) were plotted as a function of inducer concentration using software GraphPad Prism 7. Subsequently, a non-linear least-squares fit was performed using the Hill function (4.1):

$$RFP(I) = b_{max} \cdot \frac{I^h}{{K_m}^h + I^h} + b_{min} \tag{4.1}$$

The parameters correspond to the maximum level of reporter output (b_{max}), the concentration of inducer (I), the Hill coefficient (h), the inducer concentration that mediates half-maximal reporter output (K_m), and the basal level of fluorescence output (b_{min}). Relative normalised fluorescence values as shown in Figure 4.7a were obtained by dividing absolute normalised fluorescence values at a specific inducer concentration, after subtraction of the absolute normalised fluorescence of the uninduced cells, by the corresponding maximum level of fluorescence output b_{max}.

The dynamic range μ was calculated with formula (4.2):

$$\mu = \frac{b_{max}}{b_{min}} \tag{4.2}$$

The corresponding standard error σ_μ was calculated using equation (4.3):

$$\sigma_\mu = \mu \cdot \sqrt{\left(\frac{\sigma_{b_{max}}}{b_{max}}\right)^2 + \left(\frac{\sigma_{b_{min}}}{b_{min}}\right)^2} \quad (4.3)$$

The standard error of the maximum level of reporter output was obtained from Prism, whereas the standard error of the basal level of fluorescence output was calculated from the absolute normalised fluorescence values of the uninduced cells.

4.2.6 Cross-reactivity screen

The activity of effectors against non-cognate promoters was evaluated using an integrated robotic platform (Beckman Coulter). The *C. necator* preculture was set up by inoculating 2 mL of chloramphenicol-containing minimal medium with a single colony of freshly transformed bacterial cells. After incubation for 18 h with orbital shaking at 30 °C and 200 rpm, the bacterial culture was diluted 1:50 in 50 mL of fresh minimal medium containing the antibiotic in 250-mL baffled shake flasks. The cells were grown for another 4 h under the same conditions until an OD_{600} of 0.15-0.2 was reached. After pouring the bacterial culture into a 250-mL reservoir (Thermo Fisher Scientific), 142.5 µL were dispensed into a black 96-well microtiter plate (the same that was used for the fluorescence measurements) using a liquid handling robotic platform (Biomek FXp, Beckman Coulter). The workflow generated using software SAMI EX (Beckman Coulter) is illustrated in Supplementary Figure 2. Inducers were dissolved to a final concentration of 100 mM (except in case of L-2-hydroxyglutarate which was dissolved to a final concentration of 10 mM) and transferred to a 96-deep-well plate (2.0 mL square wells with round bottoms, STARLAB International GmbH). Using a Biomek FXp, 7.5 µL of stock inducer were added to the *C. necator* culture in the microtiter plate. Fluorescence and absorbance measurements were taken immediately, 6, 12, and 18 h after supplementation with inducer by an integrated SpectraMax 3i plate reader (Molecular Devices). The same plate reader settings were

used as for the measurements taken using the Infinite M1000 PRO microplate reader. In between the measurements, the plates were kept in an integrated Cytomat2 shaking incubator (Thermo Fisher Scientific) at 30 °C and 600 rpm. The workflow generated using software SAMI EX (Beckman Coulter) is illustrated in Supplementary Figure 3. The cross-reactivity screen was performed with assistance of Magdalena Jonczyk.

The relative induction (in %) is calculated using equation (4.4):

$$Relative\ induction\ (\%) = 100 \cdot \left(\frac{FL_{\text{compound}} - FL_{\text{uninduced}}}{FL_{\text{primary inducer}} - FL_{\text{uninduced}}} \right) \quad (4.4)$$

FL corresponds to the OD-normalised absolute fluorescence values.

4.3 Results and discussion

4.3.1 Designing a method for identifying native inducible systems

Transcription factor-based inducible systems are composed of a TR protein and an inducible promoter, including TR and RNAP holoenzyme binding sequences. In the systems that control genes clusters associated with metabolism and catabolism in particular, the level of gene expression from the inducible promoter is often controlled by the TR which responds to small effector molecules, also referred to as ligands. To make such systems universally applicable, all three components need to be identified: the regulator, the inducible promoter, and its corresponding effector.

For the identification of inducible systems, the highly conserved genetic arrangement, typical of LysR-type transcriptional regulators (LTTRs), but not exclusive to other types of TRs, was chosen to serve as a platform for the genome scale approach. In this commonly occurring arrangement, TRs are transcribed in divergent orientation of target genes or operons[30, 164]. Once a complete list of annotated genes belonging to one species is retrieved from GenBank[165] (www.ncbi.nlm.nih.gov),

including information on coding strand orientation and protein function, it is manually screened for TRs that are oriented in the opposite direction of operons involved in metabolism. To constrain the search, the operon itself is to be composed of at least two genes encoding annotated catalytic functions associated with a distinct metabolic pathway. For each enzyme encoded by the operon, a list of metabolic substrates and products is extracted from The Comprehensive Enzyme Information System[166] (BRENDA, www.brenda-enzymes.org). By comparing potential metabolite substrates and products of each of the involved enzymes, the primary substrate is concluded which is likely to be metabolised by the operon-encoded enzymes. This compound was proposed to be the ligand that binds the TR initiating expression of genes which encode the ligand-metabolising pathway enzymes (Figure 4.1). By following this approach the TR is assigned a role in metabolism solely based on its proximity to a metabolic cluster of genes. The methodical workflow was applied in the metabolically versatile chemolithoautotrophic bacterium *C. necator* H16, resulting in the identification of 16 putative transcription factor-based inducible gene expression systems (Figure 4.2a,b; Supplementary Table 3).

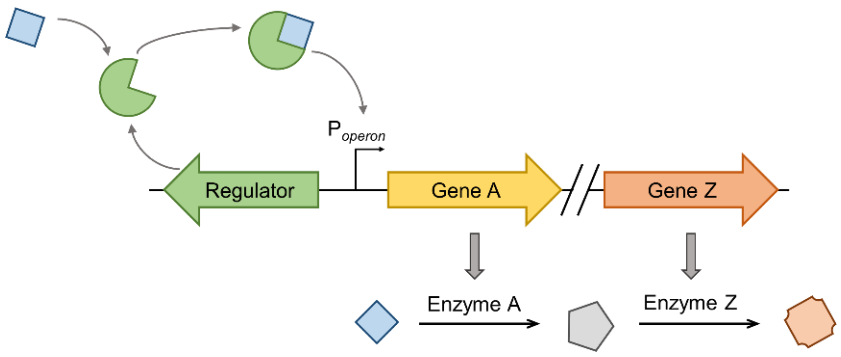
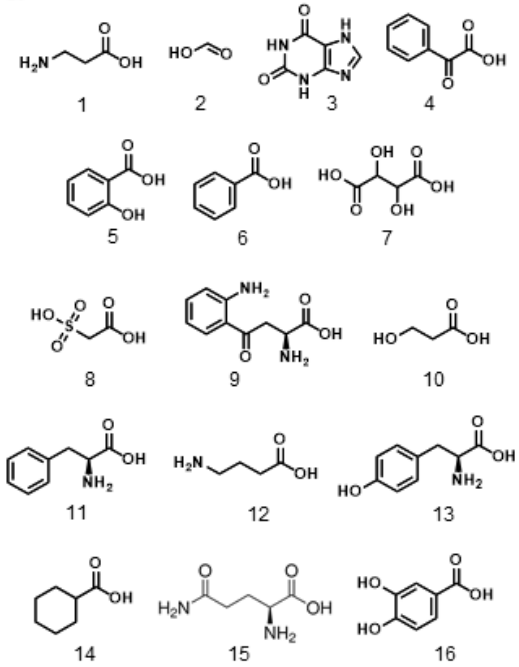


Figure 4.1 Schematic illustration of a transcriptional regulator controlling expression of a metabolic cluster of genes. The primary substrate was proposed to be the ligand. Note that the enzyme which converts the primary substrate into an intermediate product may be encoded by any of the genes in the operon.

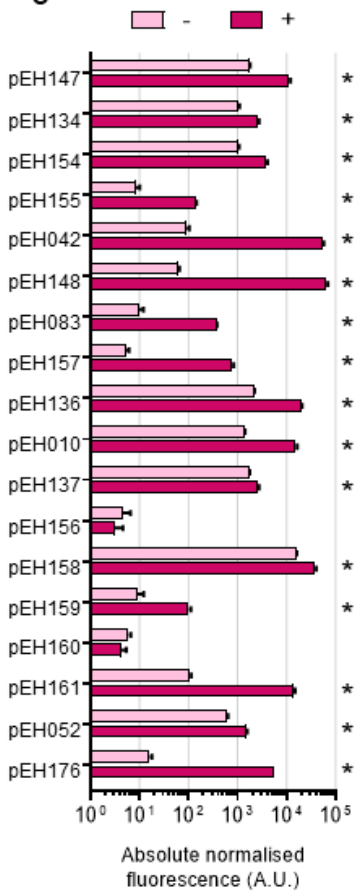
a



b

Effector	Regulator	Promoter
3-Aminopropanoate	OapR	$P_{H16\_RS01330}$
Formate	FdsR	$P_{H16\_RS03165}$
Xanthine	XdhR	$P_{H16\_RS05055}$
Phenylglyoxylate	PhgR	$P_{H16\_RS05530}$
Salicylate	NahR	$P_{H16\_RS08125}$
Benzoate	BenM	$P_{H16\_RS09790}$
Tartrate	TtdR	$P_{H16\_RS10665}$
Sulfonatoacetate	SauR	$P_{H16\_RS13605}$
L-Kynurenine	KynR	$P_{H16\_RS14030}$
3-Hydroxypropanoate	HpdR	$P_{H16\_RS18295}$
L-Phenylalanine	PhhR	$P_{H16\_RS18365}$
4-Aminobutanoate	GabR	$P_{H16\_RS23655}$
L-Tyrosine	HpdA	$P_{H16\_RS24175}$
Cyclohexanecarboxylate	BadR	$P_{H16\_RS27200}$
L-Glutamine	H16_RS29645	$P_{H16\_RS29650}$
3,4-Dihydroxybenzoate	PcaQ	$P_{H16\_RS30145}$
Acetoin	AcoR	$P_{H16\_RS19445}$
L-Arabinose	AraC	P_{araBAD}

c



d

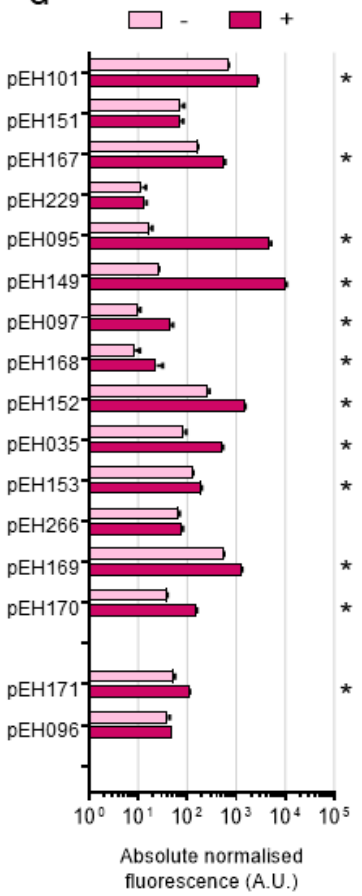


Figure 4.2 Quantitative evaluation of native inducible systems. (a) Chemical structures of the proposed primary effector molecules: β-alanine (1), formic acid (2), xanthine (3), phenylglyoxylic acid (4), salicylic acid (5), benzoic acid (6), tartaric acid (7), sulfoacetic acid (8), L-kynurenine (9), 3-hydroxypropionic acid (10), L-phenylalanine (11), γ-aminobutyric acid (12), L-tyrosine (13), cyclohexanecarboxylic acid (14), L-glutamine (15), and 3,4-dihydroxybenzoic acid (16). (b) Summary of the identified inducible systems including the inducible promoter, TR name and corresponding ligand. (c) Single time-point fluorescence measurements (arbitrary units) of *C. necator* H16 carrying the transcription factor-based inducible gene expression systems composed of TR and inducible promoter in the same order as listed in panel (b). The plasmids harbouring the individual system-reporter constructs are indicated. (d) Single time-point fluorescence measurements of *C. necator* H16 carrying the inducible 'promoter only' implementations in the same order as listed in panel (b). The plasmids harbouring the individual promoter-reporter gene constructs are indictaed. Fluorescence output was determined in the absence of inducer (-) and 6 h after extracellular supplementation with the corresponding effector to a final concentration of 5 mM (+). Error bars represent standard deviations of three biological replicates. Asterisks indicate statistically significant differences between the fluorescence output of the uninduced and induced sample ($p < 0.05$; unpaired *t* test).

4.3.2 Quantitative evaluation of TR ligand-mediated gene expression

By the genome-wide analysis identified native systems were cloned into the modular reporter vector to examine their response to the presence of the proposed compounds. The original genetic organisation was conserved by positioning the TR gene in the opposite orientation of the reporter gene encoding mRFP (Figure 4.3). The *E. coli* L-arabinose-, and the *C. necator* acetoin-inducible systems have been tested previously[102, 104] and were included for comparative purpose (Figure 4.2b).

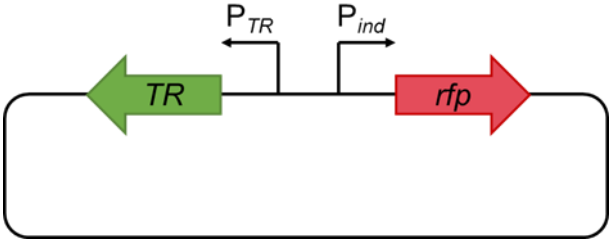


Figure 4.3 Schematic illustration of an inducible system cloned into the reporter vector. The system is composed of the transcriptional regulator (TR) gene, the TR promoter (P_{TR}), and the inducible promoter (P_{ind}).

C. necator strains carrying the inducible systems were grown in rich medium and fluorescence output of cells in late exponential growth-phase was quantified 6 h

after addition of the inducer at an OD_{600} of 1.4-2.0. It should be noted that the metabolic gene cluster, which is putatively controlled by KynR despite converting L-tryptophan into anthranilic acid and L-alanine, the intermediate compound L-kynurenine was proposed to be the effector molecule and not the primary substrate L-tryptophan based on thorough characterisation of KynR in other bacterial species[167]. In addition to the acetoin- and L-arabinose-inducible systems, which were included as positive controls, 14 of the 16 newly identified putative inducible systems showed an increase in mRFP protein synthesis after supplementation with their proposed effector molecules (Figure 4.2 and Supplementary Table 4). Systems responding to 3-aminopropanoate (β-alanine) and phenylglyoxylate have not been reported and highlight the potential of the developed methodical pipeline for mining novel biosensors. Moreover, seven of the native systems exhibited higher levels of gene expression than the commonly used heterologous L-arabinose-inducible system in the presence of their corresponding effectors. Benzoate mediated the highest induction (factor of 1063-fold) and the highest absolute normalised fluorescence with an expression level of more than 11-fold higher than AraC/P_{araBAD}. Low basal promoter activities were observed for metabolites that are neither involved in primary metabolism in *C. necator* nor likely to be present in the employed complex medium, including sulfonatoacetate, tartrate, cyclohexanecarboxylate, and phenylglyoxylate.

However, the putative 4-aminobutanoate (GABA)- and L-glutamine-inducible systems showed no induction, even though their proposed ligands are involved in primary metabolism in *C. necator* and likely to be present in the rich medium. It was hypothesised that translational start sites of the respective first gene in both operons are incorrectly annotated, resulting in reporter constructs with ineffectual 5' untranslated regions. To test this hypothesis, the GABA- and L-glutamine-inducible

systems were redesigned comprising the TR gene, the intergenic region, the first gene in the operon, and the intergenic region preceding the second gene cloned upstream of the reporter (Figure 4.4a). Both inducible systems resulted in increased basal promoter activities (Figure 4.4b and Supplementary Table 5). Remarkably, GabR/$P_{H16\_RS23655}$ mediated a 1.9-fold induction of gene expression in the presence of GABA.

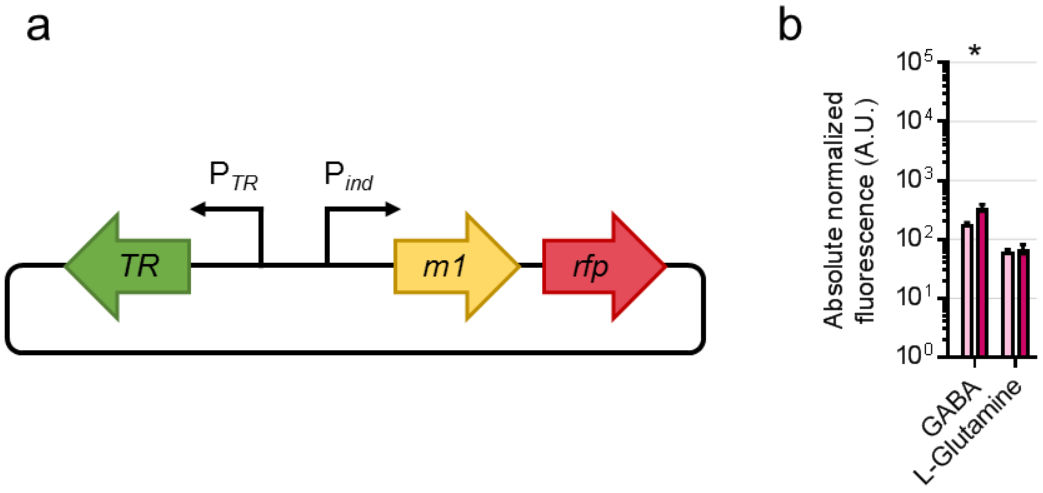


Figure 4.4 Redesigning the putative GABA- and L-glutamine-inducible systems. (a) Schematic illustration of the modular reporter vector containing a putative inducible system, the first gene in the operon (*m1*), and the intergenic region preceding the second gene of the operon upstream of the *rfp* reporter gene. (b) Single time-point fluorescence measurements of *C. necator* H16 carrying the redesigned putative GABA- and L-glutamine inducible systems (pEH256 and pEH240, respectively). The fluorescence output was determined in the absence of inducer (bright pink) and 6 h after extracellular supplementation with the corresponding effector to a final concentration of 5 mM (dark pink). Error bars represent standard deviations of three biological replicates. The asterisk indicates a statistically significant difference between the fluorescence output of the uninduced and induced sample ($p < 0.05$; unpaired *t* test).

Increase in construct size can ultimately become the limiting factor in synthetic biology as transformation efficiencies linearly decrease with increasing plasmid size[168]. Utilising an inducible system which is endogenous to the organism provides the advantage to possess a copy of the TR encoded in the genome enabling truncation of the controllable element to the sole inducible promoter, thus reducing construct size considerably. Although promoter activities under both uninduced and induced

conditions generally decreased when the TR gene was removed from the multicopy episomal vector, a majority of inducible promoters significantly facilitated gene expression in the presence of their corresponding effectors (Figure 4.2d and Supplementary Table 6). Whereas the 'promoter only' construct containing $P_{H16\_RS27200}$, which lacks the copy of the TR gene *badR*, exhibited a greater fluorescence level in the absence of effector than that of the BadR/$P_{H16\_RS27200}$-cyclohexanecarboxylate-inducible system, suggests that BadR acts as a repressor, similar to its homologue in *Rhodopseudomonas palustris*[169]. All other native TRs might be classed as activators or dual-function TRs. Furthermore, the induction factor was smaller for all of the systems as a result of an altered ratio between transcription factor binding sites and available TR proteins. In fact, the promoters controlled by formate, phenylglyoxylate, GABA, and acetoin showed no significant induction. Two of the inducible promoters, however, demonstrated an exceptionally strong activation of *rfp* expression: $P_{H16\_RS09790}$, responding to benzoate, and $P_{H16\_RS08125}$, which is activated by salicylate, mediated inductions by 403- and 292-fold, respectively. The 146 bp long intergenic region containing the benzoate-inducible promoter itself showed a stronger activation of gene expression than the majority of the 'complete' inducible systems, including the commonly used L-arabinose-inducible system (Figure 4.2c). This characteristic highlights the potential of $P_{H16\_RS09790}$ (responding to benzoate) to be employed as individual genetic element to control high levels of gene expression by reducing construct size by 7-fold. Even in cases where a TR gene cannot be mapped to a cluster of genes involved in metabolism, the methodical approach described in this study can be employed for mining endogenous metabolite-inducible promoters.

4.3.3 Orthogonal gene expression control in *Escherichia coli* and *Pseudomonas putida*

To assess the potential of the constructed biosensors to be applied in other microorganisms, the transcription factor-based inducible systems that were mined from the genome of the β-proteobacterium *C. necator* were evaluated in the industrially relevant γ-proteobacteria *Escherichia coli* TOP10 and *Pseudomonas putida* KT2440.

Regardless of their origin, the majority of systems responding to the 16 primary effectors, including acetoin, have never been tested in *E. coli* and *P. putida*. Thus, the broad host-range applicability of the identified systems is highlighted by the outcome that a total of 8 and 12 of the 16 systems mediated a significant increase in reporter gene expression after inducer addition in *E. coli* (Figure 4.5a, Supplementary Table 7) and *P. putida* (Figure 4.5c, Supplementary Table 8), respectively. Three of them, activated by salicylate, benzoate, and 3,4-dihydroxybenzoate induced by more than 75-fold in both tested microorganisms. Compared to systems that were sourced from other prokaryotes and tested in *E. coli* for controllable gene expression, PcaQ/$P_{H16\_RS30145}$ and BenM/$P_{H16\_RS09790}$ from *C. necator* outperform previously evaluated 3,4-dihydroxybenzoate- and benzoate-inducible systems by approximately 5-, and 50-fold[170, 171] (Supplementary Table 9). Specifically, benzoate mediated the highest induction (factor of 4428-fold) in *E. coli* and the highest absolute normalised fluorescence in *P. putida* with an expression level of more than 15-fold higher than AraC/P_{araBAD} demonstrating its potential for high level gene expression across different species.

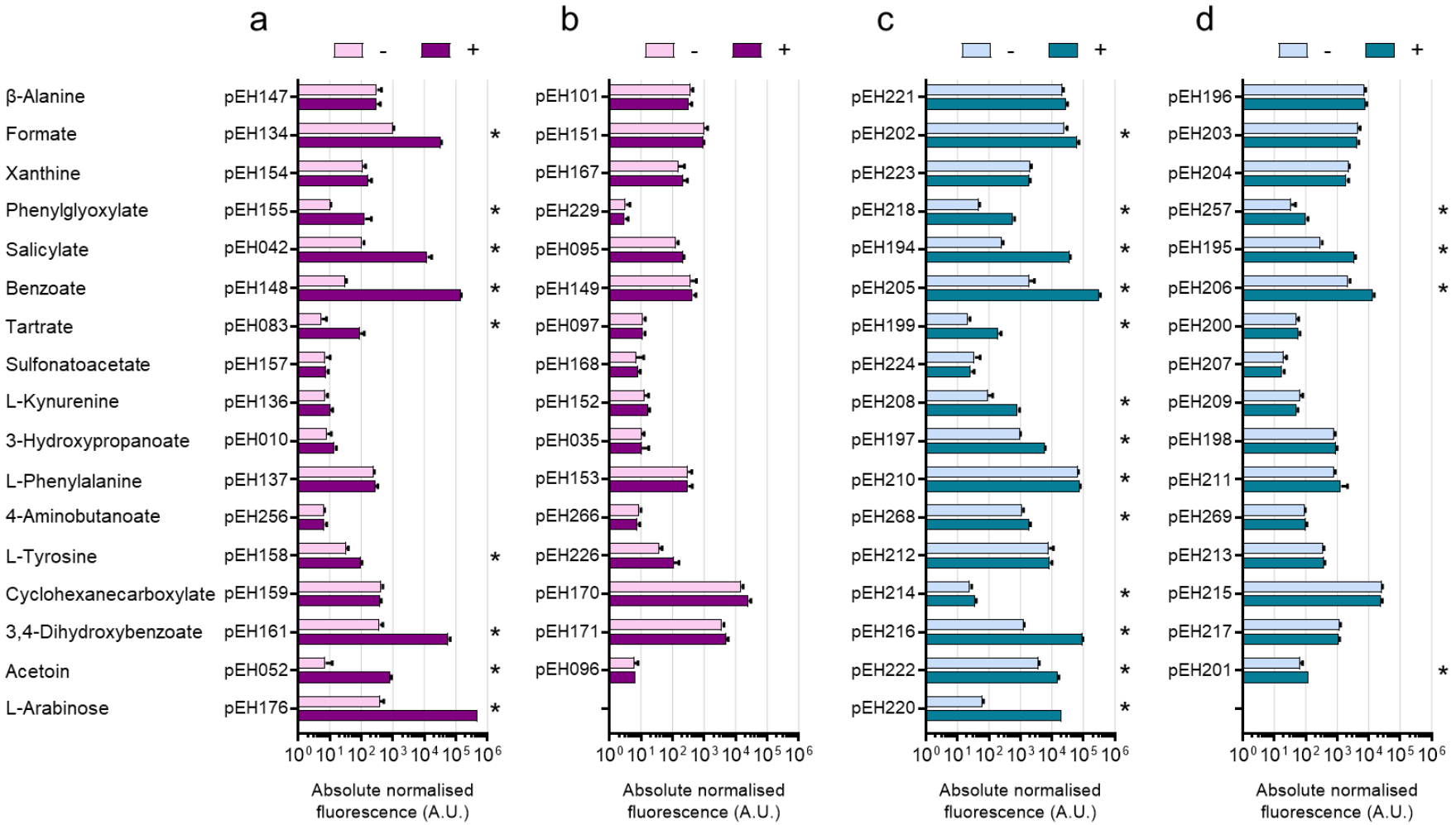

a
b
c
d
-
+
β-Alanine
Formate
Xanthine
Phenylglyoxylate
Salicylate
Benzoate
Tartrate
Sulfonatoacetate
L-Kynurenine
3-Hydroxypropanoate
L-Phenylalanine
4-Aminobutanoate
L-Tyrosine
Cyclohexanecarboxylate
3,4-Dihydroxybenzoate
Acetoin
L-Arabinose
pEH147
pEH134
pEH154
pEH155
pEH042
pEH148
pEH083
pEH157
pEH136
pEH010
pEH137
pEH256
pEH158
pEH159
pEH161
pEH052
pEH176
pEH101
pEH151
pEH167
pEH229
pEH095
pEH149
pEH097
pEH168
pEH152
pEH035
pEH153
pEH266
pEH226
pEH170
pEH171
pEH096
pEH221
pEH202
pEH223
pEH218
pEH194
pEH205
pEH199
pEH224
pEH208
pEH197
pEH210
pEH268
pEH212
pEH214
pEH216
pEH222
pEH220
pEH196
pEH203
pEH204
pEH257
pEH195
pEH206
pEH200
pEH207
pEH209
pEH198
pEH211
pEH269
pEH213
pEH215
pEH217
pEH201
*
10^0 10^1 10^2 10^3 10^4 10^5 10^6
Absolute normalised fluorescence (A.U.)

Figure 4.5 *C. necator* native systems mediate controllable gene expression in *E. coli* and *P. putida*. Single time-point fluorescence measurements of (a, b) *E. coli* and (c, d) *P. putida* carrying the *C. necator* inducible gene expression systems composed of TR and promoter (a, c) or 'promoter only' version (b, d). The plasmids harbouring the individual system- or promoter-reporter gene constructs are indicated. Fluorescence output was determined in the absence of inducer (-) and 6 h after extracellular supplementation with the corresponding primary effector to a final concentration of 5 mM (+). Error bars represent standard deviations of three biological replicates. Asterisks indicate statistically significant differences between the fluorescence output of the uninduced and induced sample ($p < 0.05$; unpaired *t* test).

To test for regulator-dependant orthogonality the 'promoter only' versions of the inducible systems were evaluated for controllable gene expression in *E. coli* and *P. putida*. Without the episomal copy of the TR gene, none of the promoters showed an increase in activity even in the presence of the effector in *E. coli* (Figure 4.5c, Supplementary Table 10), whereas in *P. putida* RFP synthesis was significantly induced from the phenylglyoxylate-, salicylate-, benzoate-, and acetoin controllable promoters (Figure 4.5d, Supplementary Table 11). Activation of reporter gene expression may be explained by cross-reactivity of chromosomally encoded TRs. A protein blast revealed homologues of BenM and AcoR to be encoded in the genome of *P. putida* (Table 4.2) which might be able to activate gene expression from the *C. necator* benzoate- and acetoin-inducible promoters, respectively. NahR and PhgR homologues could not be identified indicating that transcriptional activation from the salicylate- and phenylglyoxylate-inducible promoters may result from unspecific TR binding.

Table 4.2 *P. putida* KT2440 BenM and AcoR homologues. The genome of *P. putida* KT2440 was searched for homologues of *C. necator* H16 BenM and AcoR. Amino acid sequence identity (coverage) in %.

Regulator	*C. necator* H16		*P. putida* KT2440	
	Locus tag	Identity (coverage) in %	Locus tag	Identity (coverage) in %
BenM	H16_RS09795	100 (100)	PP_3716	50 (95)
AcoR	H16_RS19440	100 (100)	PP_0557	45 (93)

4.3.4 Engineering the β-alanine-inducible system for gene expression control in *E. coli* and *P. putida*

Each inducible system harbours several functional genetic elements which independently contribute to its overall performance. These regulatory elements, including promoters and RBSs, which control the expression of the TR gene and its associated regulon, may vary in their usage and efficiency across different species. As a consequence, inducible systems may perform differently when transferred from one into another organism. For example, the β-alanine-inducible system mediated a moderate activation of reporter gene expression in *C. necator* (Figure 4.2c), whereas in *E. coli* and *P. putida* the fluorescence output remained at basal levels even in the presence of the inducer (Figure 4.5a,c). β-Alanine is an intermediate compound for the synthesis of industrially relevant nitrogen-containing platform chemicals, including acrylamide, acrylonitrile and poly- β-alanine (also known as nylon-3)[172, 173]. Furthermore, it is a precursor of the dipeptides carnosine and anserine that have been demonstrated to improve cognitive functions and physical capacities in humans[174, 175]. To expand the host range and due to its usefulness as biosensor in synthetic biology and biotechnology applications, the β-alanine-inducible system from *C. necator* was pursued to be modified to enable its utilisation for gene expression control in *E. coli* and *P. putida*.

Although the system did not demonstrate an increase in RFP synthesis after addition of β-alanine, the promoter $P_{H16\_RS01330}$ showed the fifth highest activity of all evaluated systems under uninduced conditions in *E. coli* (Figure 4.5a). This indicated that the regulatory elements of $P_{H16\_RS01330}$, including promoter and RBS, are functional in *E. coli* and that the lack of induction is more likely to be attributed to inducer uptake or regulator gene expression. A poor ligand transport as a cause of the absence of induction could be excluded as β-alanine has been reported to be actively taken up by the cells[176]. Strikingly, the GC-content of the TR gene (71%) is significantly higher than the GC-content of the *E. coli* K12 genome (51%). To rule out that a high GC-content impairs TR synthesis, *H16_RS01325* (*oapR*) was codon-optimised for *E. coli* codon usage (pEH173, Figure 4.6a). However, this modification did not improve the response of the system to β-alanine (Figure 4.6b).

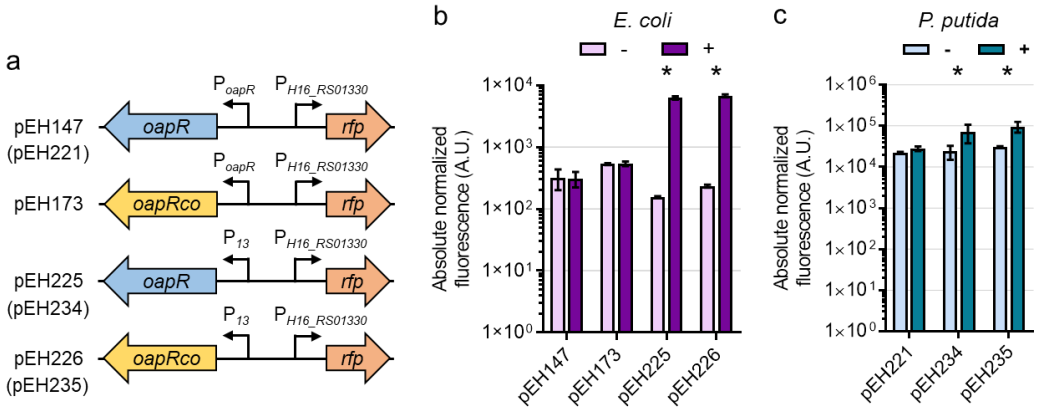


Figure 4.6 Engineering the β-alanine-inducible system. (a) Schematic illustration of the different versions of the β-alanine-inducible system and their corresponding plasmid identifiers. Absolute normalised fluorescence of (b) *E. coli* and (c) *P. putida* carrying different versions of the β-alanine-inducible system in the absence (-) and presence (+) of 5 mM β-alanine. Single time-point fluorescence measurements were taken 6 h after effector addition. Error bars represent standard deviations of three biological replicates. Asterisks indicate statistically significant differences between the fluorescence output of the uninduced and induced sample ($p < 0.05$; unpaired *t* test).

Subsequently, to ensure that the regulator is expressed in *E. coli* and *P. putida*, the *C. necator* native promoter of the TR was replaced by a host-specific promoter.

The 24 bp DNA sequence upstream of the *oapR* translational start site was replaced by the core sequence of a medium-strength insulated constitutive promoter, P_{13}[106], including the phage T7 gene *10* RBS. The substitution was implemented in both plasmids containing the native (pEH147 and pEH221) and the codon-optimised *oapR* (pEH173, Figure 4.6a). The addition of β-alanine resulted in a 40- and 29-fold increase in fluorescence output for *E. coli* cultures carrying pEH225 and pEH226, respectively (Figure 4.6b). This suggests that the original promoter CnP_{oapR} is not functional in *E. coli* and that codon optimisation might even lead to a lower TR synthesis rate which results in a decreased induction level. Moreover, a lower background reporter gene expression under uninduced conditions in case of pEH225 and pEH226 indicates that OapR, a member of the MocR family, may act as dual-function TR, repressing transcription of *oapTD* in the absence of β-alanine but acting as activator in its presence. The dual mode of action has been observed for other member of this family of TRs[177]. In *P. putida*, the promoter and RBS exchange also resulted in a 3-fold induction of gene expression for both versions the native (pEH234) and the codon-optimised (pEH235) *oapR* (Figure 4.6a,c). In contrast to *E. coli*, however, induction levels are significantly lower which may be attributed to a higher basal promoter activity in *P. putida*.

4.3.5 Characterisation of induction dynamics and homogeneity

The 16 functional native inducible systems, including the acetoin-inducible system, were subsequently evaluated for their dose-response, dynamic range, induction homogeneity, and orthogonality.

The dose-response curve of a metabolite-responsive inducible system describes the level of gene expression as a function of ligand concentration, thus indicating the range of effector concentration in which the inducible system is able to

operate. It provides key parameters which aid in part selection and the computational design of synthetic circuits. To simplify mathematical modelling approaches, effector concentrations are usually considered constant, mediating a sustained gene expression throughout the course of cell growth. In this case, however, the ligands are metabolised by *C. necator* resulting in a decrease in gene expression to basal levels once the inducer has been depleted. Therefore, the time point, at which the reporter output is correlated with the ligand concentration, must be chosen carefully. By postulating that the inducer-metabolising enzymes are not synthesised faster than the primary induction of reporter gene expression, and to account for the minimum amount of time required for RFP synthesis and maturation[134], the minimal induction interval of 80 min was determined. *C. necator* strains harbouring the inducible systems were grown in minimal medium and reporter gene expression was monitored after supplementation with the corresponding inducer at a wide range of concentrations over the time course of 16 h. The dose-responses were obtained by plotting the relative normalised fluorescence values of the 80-min minimal induction interval as a function of inducer concentration (Figure 4.7a). Data points were fit using a Hill function (see Chapter 4.2.5), taking into account the basal level of fluorescence output of the uninduced cells. Consequently, on the basis of the mathematically modelled dose-response curve, key parameters that distinguish one inducible system from another were obtained (Table 4.3).

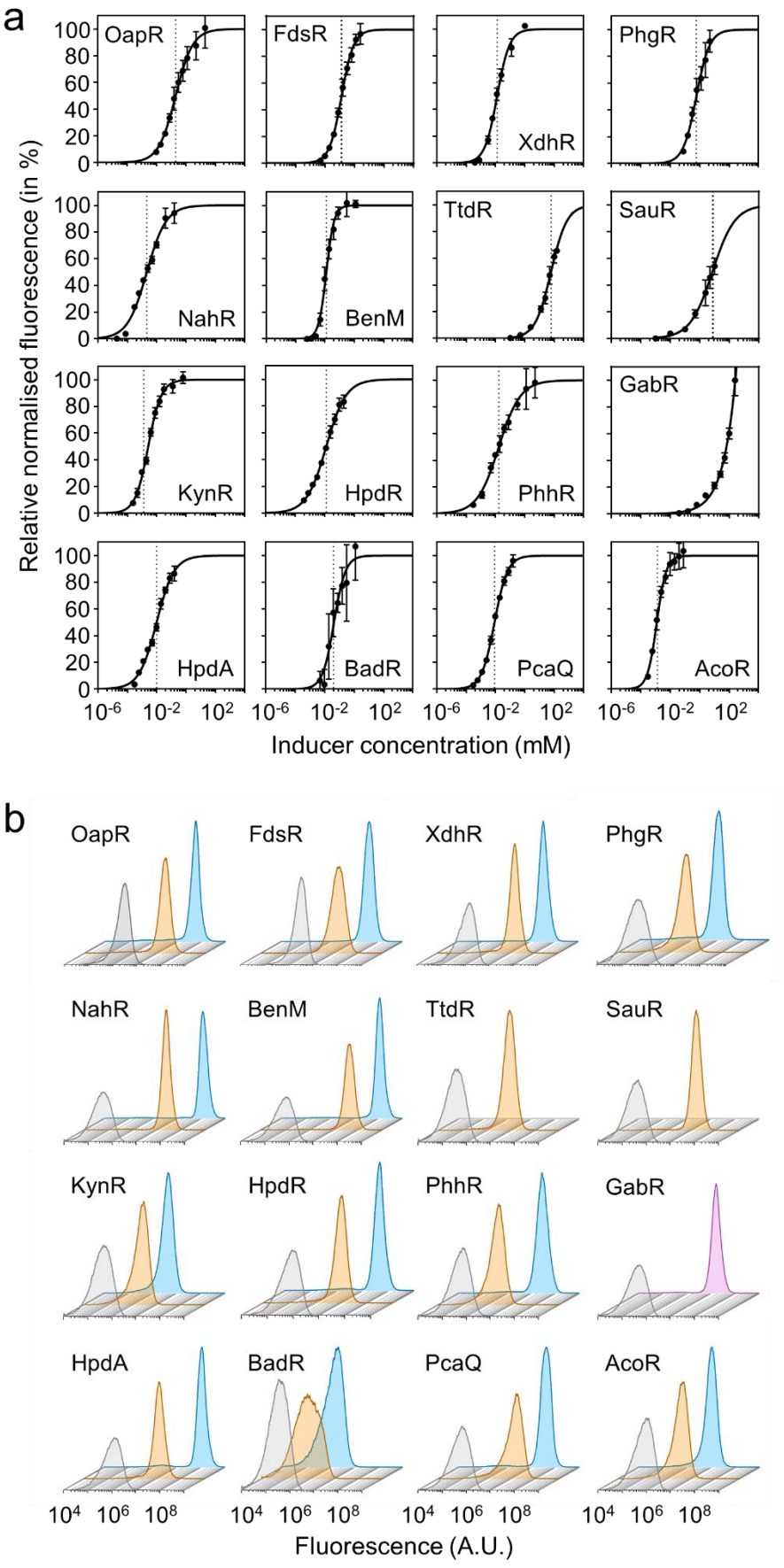

a
OapR
FdsR
XdhR
PhgR
NahR
BenM
TtdR
SauR
KynR
HpdR
PhhR
GabR
HpdA
BadR
PcaQ
AcoR
Relative normalised fluorescence (in %)
Inducer concentration (mM)
b
Fluorescence (A.U.)

Figure 4.7 Response function and induction homogeneity of the native inducible systems.(a) Relative normalised fluorescence of *C. necator* carrying the various inducible system-reporter constructs in response to different concentrations of their corresponding primary inducers. Measurements were taken 80 min after the inducer had been extracellularly added. The dose-responses were fit using a Hill function (see Chapter 4.2.5). The maximum level of reporter output b_{max} was set to 100% (except in case of the GABA-inducible system where the absolute normalised fluorescence corresponding to the highest GABA concentration tested was set to 100%). The inducer concentration that mediates half-maximal reporter output K_m is indicated by a dotted line. Error bars represent standard deviations of three biological replicates. (b) Evaluation of induction homogeneity by flow cytometry. The fluorescence intensity of 100,000 individual cells was determined for each inducible system 2 h after extracellular addition of inducer. Uninduced cells (grey) are compared to cultures supplemented with their cognate effector at final concentrations corresponding to 50% (orange) and 95% (blue; 90% in case of the phenylglyoxylate-inducible system) of the maximum level of reporter output b_{max}. The tartrate- and sulfonatoacetate-inducible systems were only evaluated for induction homogeneity at 50% of b_{max} due to solubility limits and inducer toxicity, respectively. Since b_{max} could not be calculated for the GABA-inducible system, induction homogeneity was determined using a final concentration of 250 mM (purple).

Table 4.3 Parameters of the native inducible systems.

Inducible system	Inducer	Dynamic range, in -fold[a]	K_m[b]	h[c]
OapR/P$_{H16_RS01330}$	β-Alanine	8.0 ± 0.4	201 ± 24 µM	0.75 ± 0.05
FdsR/P$_{H16_RS03165}$	Formate	4.5 ± 0.2	130 ± 7 µM	1.05 ± 0.04
XdhR/P$_{H16_RS05055}$	Xanthine	16.0 ± 1.0	13.0 ± 1.4 µM	1.03 ± 0.10
PhgR/P$_{H16_RS05530}$	Phenylglyoxylate	144.8 ± 106.5	595 ± 133 µM	0.96 ± 0.13
NahR/P$_{H16_RS08125}$	Salicylate	650.8 ± 317.4	2.12 ± 0.49 µM	0.66 ± 0.08
BenM/P$_{H16_RS09790}$	Benzoate	74.1 ± 10.7	12.6 ± 0.8 µM	1.64 ± 0.14
TtdR/P$_{H16_RS10665}$	Tartrate	370.6 ± 232.2	61.5 ± 22.9 mM	0.80 ± 0.09
SauR/P$_{H16_RS13695}$	Sulfonatoacetate	395.4 ± 193.7	7.3 ± 4.9 mM	0.57 ± 0.06
KynR/P$_{H16_RS14030}$	L-kynurenine	26.9 ± 2.1	2.64 ± 0.18 µM	0.98 ± 0.05
HpdR/P$_{H16_RS18295}$	3-Hydroxypropanoate	25.2 ± 1.7	13.2 ± 1.9 µM	0.64 ± 0.03
PhhR/P$_{H16_RS18365}$	L-phenylalanine	9.0 ± 0.8	16.9 ± 2.2 µM	0.59 ± 0.04
GabR/P$_{H16_RS23655}$	GABA	ND	ND	ND
HpdA/P$_{H16_RS24175}$	L-tyrosine	38.3 ± 5.0	9.9 ± 2.2 µM	0.72 ± 0.07
BadR/P$_{H16_RS27200}$	Cyclohexanecarboxylate	11.4 ± 10.9	43.8 ± 13.0 µM	1.11 ± 0.28
PcaQ/P$_{H16_RS30145}$	3,4-Dihydroxybenzoate	77.2 ± 29.8	8.63 ± 0.40 µM	0.98 ± 0.03
AcoR/P$_{H16_RS19445}$	Acetoin	11.1 ± 1.6	1.23 ± 0.07 µM	1.36 ± 0.10

Data are mean ± standard deviation of three biological replicates. ND – not determined. [a]Dynamic range is defined as the –fold increase in fluorescence calculated by dividing the maximum level of fluorescence output by the basal level of fluorescence output. [b]K_m represents the inducer concentration at which the half-maximal activation of the inducible system is achieved. [c]h - Hill coefficient.

One of the most important parameters when choosing an inducible system to tightly control different levels of gene expression is the dynamic range. It is defined as the maximum level of reporter output relative to basal expression levels (see Chapter 4.2.5, formula 4.2). It was found that the NahR/P$_{H16_RS08125}$-salicylate-inducible system has the highest dynamic range of the evaluated native systems

followed by SauR/$P_{H16\_RS13695}$ and TtdR/$P_{H16\_RS10665}$. In general, the dynamic range is higher than the induction level of cells grown in rich medium at an effector concentration of 5 mM (Figure 4.2c). This effect does not apply to the BenM/$P_{H16\_RS09790}$-benzoate- and PcaQ/$P_{H16\_RS30145}$-3,4-dihydroxybenzoate-inducible systems which show a lower dynamic range in minimal medium. Whereas a lower induction level in complex medium might be the consequence of structurally similar molecules that are able to interact with the respective TR, the lower dynamic range in minimal medium might be attributed to catabolic repression as it has been demonstrated to be the case for both benzoate and 3,4-dihydroxybenzoate in *P. putida*[178]. In addition to a high dynamic range in minimal medium, NahR/$P_{H16\_RS08125}$ has the lowest K_m of all evaluated native inducible systems. This parameter is defined as the inducer concentration that mediates half-maximal reporter output, suggesting that only small quantities of salicylate are needed to induce the system. Similar effector concentrations have been shown to mediate gene expression from the *P. putida* salicylate-inducible system NahR/P_{sal}[179]. In contrast to most of the inducible systems that operate in the µM-range, GabR, TtdR, and SauR seem to respond to effector concentrations three to five orders of magnitude higher than NahR. Moreover, in case of GabR, the K_m appeared to be higher than the concentration of GABA the growth medium can be supplemented with. It should be noted that the extracellular effector concentration may not necessarily correlate with the ligand concentration inside the cell ultimately dictating the level of gene expression. Ligand uptake limitations may therefore result in inaccurate parameters, as it might be the case for the GABA-, tartrate-, and sulfonatoacetate-inducible systems. For the other 13 systems ligand uptake is assumed not to be limiting. The Hill coefficient *h* indicates the range of inducer concentration over which the system results in a change in

reporter output. Inducible systems with a low Hill coefficient, such as SauR/$P_{H16\_RS13695}$, PhhR/$P_{H16\_RS18365}$, or HpdR/$P_{H16\_RS18295}$ exhibit a flatter dose-response function indicating that gene expression is tuneable over a wider range of inducer concentration. On the contrary, systems with a higher Hill coefficient, including BenM/$P_{H16\_RS09790}$ and AcoR/$P_{H16\_RS19445}$, show a steeper dose-response function suggesting that they behave more like an on/off switch (Figure 4.7a).

The homogeneity of induction was evaluated by flow cytometry. This method allows to determine whether intermediate inducer concentrations give rise to subpopulations of uninduced and fully induced cells. Their existence may indicate a more complex type of transcriptional regulation or inducer transport limitations[180]. Cell cultures of *C. necator* carrying the 16 functional native systems were subjected to inducer concentrations corresponding to 50% and 95% (90% in case of the phenylglyoxylate-inducible system) of the maximum level of fluorescence output b_{max}. The tartrate- and sulfonatoacetate-inducible systems were only evaluated for induction homogeneity at 50% of b_{max} due to solubility limits and inducer toxicity, respectively. Since b_{max} could not be calculated for the GABA-inducible system, induction homogeneity was determined using a final concentration of 250 mM. The reporter output was measured 2 h after the inducer had been extracellularly added. Each of the 16 evaluated systems demonstrated a unimodal induction behaviour after addition of the corresponding ligand at both medium and nearly saturating concentrations (Figure 4.7b). Generally, the fluorescence distribution of the uninduced cells is wider than the distribution in the presence of effector. This generalisation does not apply to the BadR/$P_{H16\_RS27200}$-cyclohexanecarboxylate-inducible system which may be a further indicator of it being a repressor-based type of inducible system.

4.3.6 Orthogonality of inducible systems

To establish whether any of these systems can be used in combination to independently control expression of more than a single gene the activity of the ligands against non-cognate promoters was evaluated. A total of 21 inducible systems was selected to test for orthogonality in *C. necator*. It includes the 16 native, the heterologous L-arabinose-, L-rhamnose-, acrylate-, and cumate-inducible systems that are described in Chapter 3, and the itaconate-inducible system that is described in Chapter 6.

To screen the 441 combinations of inducer and biosensor for cross-reactivity, an automated platform was employed (see Chapter 4.2.6). Strains of *C. necator* carrying the inducible systems were grown in minimal medium and transferred to a 96-well microtiter plate format. After the inducers had been extracellularly added, cells were cultured for 6 h before RFP fluorescence and cell density were measured. 15 of the 21 inducible systems exhibited a strong affinity to their primary ligands, showing less than 5% cross-reactivity of non-target metabolites relative to the fluorescence output mediated by the primary effector (Figure 4.8a, Supplementary Table 12). The remaining six inducible systems were activated by one or more metabolites other than their cognate inducers. This cross-reactivity may be the result of either structural resemblance, a metabolic relationship, or a combination thereof. For example, L-phenylalanine is converted into L-tyrosine by the phenylalanine 4-monooxygenase PhhA (H16_RS18365). Therefore, activation of the HpdA/$P_{H16\_RS24175}$-L-tyrosine-inducible system by L-phenylalanine is more likely to be due to biological conversion of the added compound into the primary effector rather than direct interaction of L-phenylalanine with HpdA. However, since the difference between the two molecules lies in a singly hydroxyl-group, an induction by structural resemblance cannot be entirely ruled out. The same applies to phenylglyoxylate which

activated the BenM/$P_{H16\_RS09790}$-benzoate-inducible system. Phenylglyoxylate is converted by *C. necator* into benzoate *via* a two-step reaction with benzaldehyde as intermediate compound. During the first reaction, CO_2 is generated which may also explain induction of the formate-inducible system by the structurally dissimilar phenylglyoxylate. Since CO_2 can subsequently be converted into formate[181], it is rational to postulate that the FdsR/$P_{H16\_RS03165}$-formate-inducible system is activated in this case by its primary inducer. In addition to phenylglyoxylate, the BenM/$P_{H16\_RS09790}$-benzoate-inducible system was activated by extracellular supplementation with cyclohexanecarboxylate and cumate. Cyclohexanecarboxylate shares both a structural resemblance to benzoate and downstream degradation pathways in *C. necator* which makes it difficult to conclude its cause of cross-reactivity. In cases like these, advantage can be taken of the system's transferability. Moving the inducible system from one organism into another host with a dissimilar metabolism may allow to distinguish more easily between induction by structural resemblance and metabolic relationship. To investigate the cause of cross-reactivity of the compounds that resulted in a fluorescence output of more than 10% relative to the primary inducer in *C. necator*, their induction behaviour was evaluated in *E. coli*. Single time-point fluorescence measurements of *E. coli* revealed that the BenM/$P_{H16\_RS09790}$-benzoate-inducible system is activated by addition of cyclohexanecarboxylate and cumate (Supplementary Table 12). *E. coli* has not been reported to metabolise any of these compounds suggesting that cross-reactivity is caused by structural resemblance.

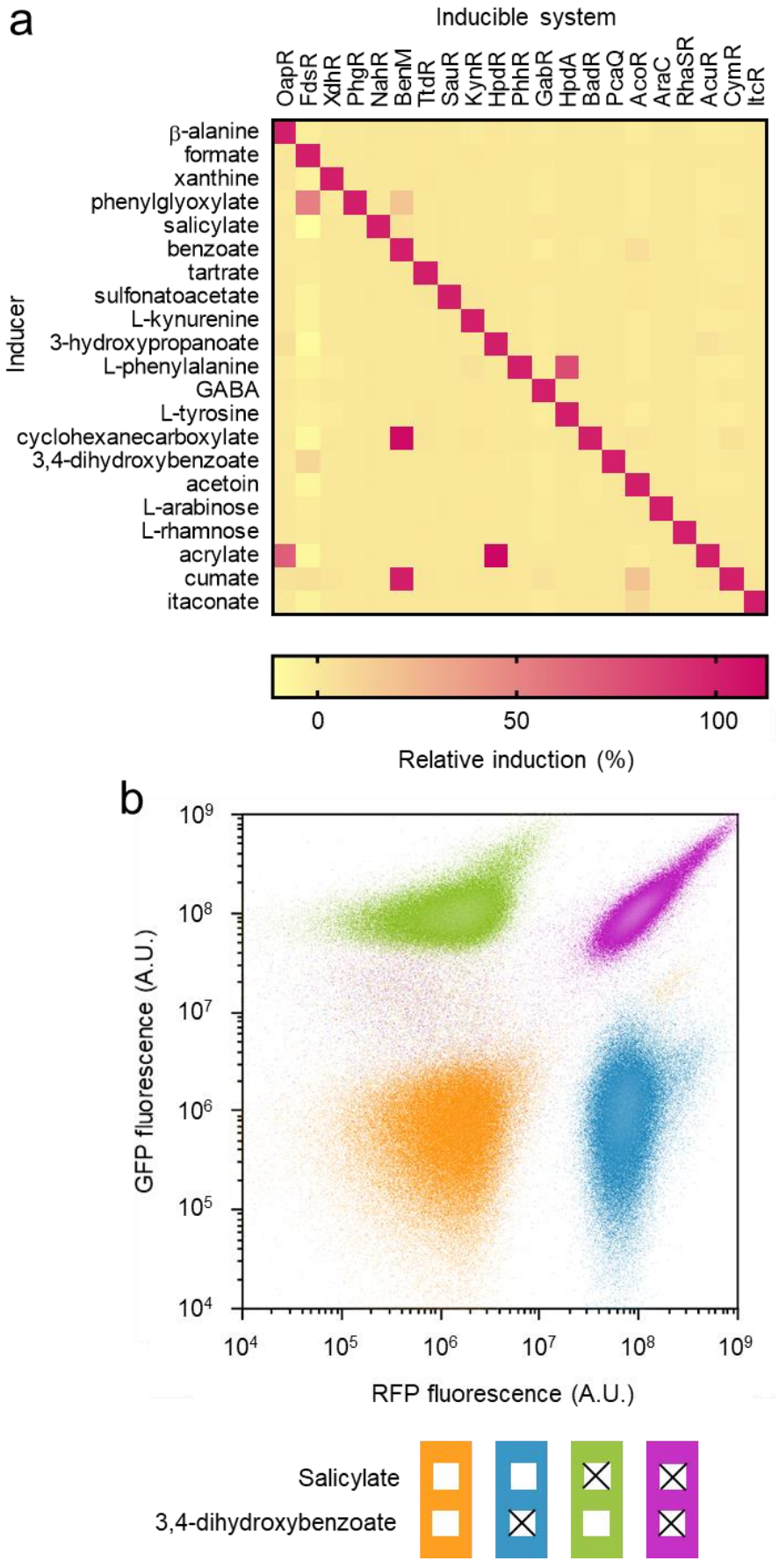

a
Inducible system
OapR
FdsR
XdhR
PhgR
NahR
BenM
TtdR
SauR
KynR
HpdR
PhhR
GabR
HpdA
BadR
PcaQ
AcoR
AraC
RhaSR
AcuR
CymR
ItcR
Inducer
β-alanine
formate
xanthine
phenylglyoxylate
salicylate
benzoate
tartrate
sulfonatoacetate
L-kynurenine
3-hydroxypropanoate
L-phenylalanine
GABA
L-tyrosine
cyclohexanecarboxylate
3,4-dihydroxybenzoate
acetoin
L-arabinose
L-rhamnose
acrylate
cumate
itaconate
0
50
100
Relative induction (%)
b
GFP fluorescence (A.U.)
RFP fluorescence (A.U.)
Salicylate
3,4-dihydroxybenzoate

Figure 4.8 Orthogonality of native inducible systems. (a) Cross-reactivity of a set of 21 inducible systems and their corresponding primary inducers. The heat map illustrates induction of reporter gene expression in the presence of the metabolite (in %) relative to the induction mediated by the corresponding primary effector. Measurements were taken 6 h after supplementation with inducer at a final concentration of 5 mM. Values represent the average of four replicates. (b) Fluorescence output of individual cells of *C. necator* containing pEH263 measured by flow cytometry. The vector pEH263 contains the 3,4-dihydroxybenzoate- and salicylate-inducible systems controlling expression of *rfp* and *egfp*, respectively. Fluorescence was determined in the absence of inducer (orange), in the presence of 3,4-dihydroxybenzoate (blue), salicylate (green), and both 3,4-dihydroxybenzoate and salicylate (purple).

More difficult to explain is the cause of activation of the β-alanine- and 3-HP-inducible systems by acrylate. It has been shown that *C. necator* is able to degrade acrylate (see Chapter 3.3.5), however, relatively little is known about its metabolism. Activation by structural resemblance to the primary effector is less likely as addition of acrylate to *E. coli* carrying the engineered β-alanine-inducible system (pEH225) only resulted in a relative induction of 2.7% in contrast to 68% in *C. necator* (Supplementary Table 12). Therefore, activation by structural resemblance of a degradation product or direct conversion into the primary effectors 3-HP and β-alanine in *C. necator* may be more likely. Acrylate can be activated into acryloyl-CoA by acyl CoA:acetate/3-ketoacid CoA transferase (H16_RS22005/H16_RS22010) or by the propionate CoA-transferase Pct (H16_RS13535)[182], which, as it has been proposed by Peplinski et al[183], can be converted into 3-HP *via* its CoA intermediate. However, pathways from acrylate or 3-HP to β-alanine or a metabolic intermediate, which is able to activate the β-alanine-inducible system, will have to be elucidated.

Based on the results of the cross-reactivity screen, two orthogonal inducible systems were employed to independently control expression of two fluorescent protein reporter genes. Plasmid pEH263 was constructed containing *rfp* under control of the 3,4-dihydroxybenzoate-, and *egfp* under control of the salicylate-inducible system (Supplementary Figure 4). The salicylate- and 3,4-dihydroxybenzoate-inducible

systems were selected due to their potential to mediate high levels of gene expression and their specificities to their corresponding inducers. Cultures of *C. necator* harbouring pEH263 were left uninduced, subjected to the individual inducers, or the combination of both at final concentrations corresponding to K_m (Table 4.3). The output of the two non-overlapping fluorescent proteins was determined by flow cytometry 2 h after the inducer/inducers had been extracellularly added. Employing every possible inducer combination, four distinct cell states could be observed (Figure 4.8b). RFP and eGFP fluorescence in the absence of both inducers remained at background levels comparable to the single-system implementations (orange population). Addition of 3,4-dihydroxybenzoate resulted in synthesis of RFP only as represented by the blue population. Similarly, the presence of salicylate activated expression of *egfp* but not *rfp* (green population). The final cell state is represented by the purple population where both inducers were added to mediate the simultaneous expression of both fluorescent protein reporter genes.

In conclusion, the identified native systems can be used in combination to independently control expression of multiple genes expanding the list of available switches in synthetic circuit design. Importantly, inducible systems for structurally similar (phenylglyoxylate, salicylate, cyclohexanecarboxylate, and 3,4-dihydroxybenzoate) and distinctive (xanthine, tartrate, sulfonatoacetate, and GABA) compounds (Figure 4.2a) were demonstrated to be fully orthogonal.

4.3.7 Screening of structurally similar and metabolically related compounds

TRs often recognise molecules that are structurally similar to their primary effectors. For example, IPTG is a commonly used structural analogue of allolactose, employed to control the expression of genes regulated by LacI. Here, the TR's specificity was investigated to identify novel ligand-TR interactions and consequently extend the

biosensor application to the detection of structurally similar or metabolically related compounds. To do so, the library of 21 native and heterologous inducible systems was screened against 46 compounds using an integrated robotic platform as described in the previous section. Of the 46 metabolites, 12 demonstrated an induction of at least 5% relative to the induction mediated by the corresponding primary effector and an absolute induction factor of at least 5 (Figure 4.9, Supplementary Table 13). Similarly to the former orthogonality screen (Figure 4.8a), it should be noted that activation of reporter gene expression by the extracellularly added compounds likely indicates a structural resemblance to the primary ligand, a metabolic relationship, or a combination thereof. To shed light on their cause of system activation, cultures of *E. coli* carrying the functional FdsR/$P_{H16\_RS03165}$-formate-, NahR/$P_{H16\_RS08125}$-salicylate-, BenM/$P_{H16\_RS09790}$-benzoate-, PcaQ/$P_{H16\_RS30145}$-3,4-dihydroxybenzoate-, AcoR/$P_{H16\_RS19445}$-acetoin-, and OapR/$P_{H16\_RS01330}$-β-alanine-inducible systems, active in this host organism, were evaluated for the same cross-reactivity (Supplementary Table 13).

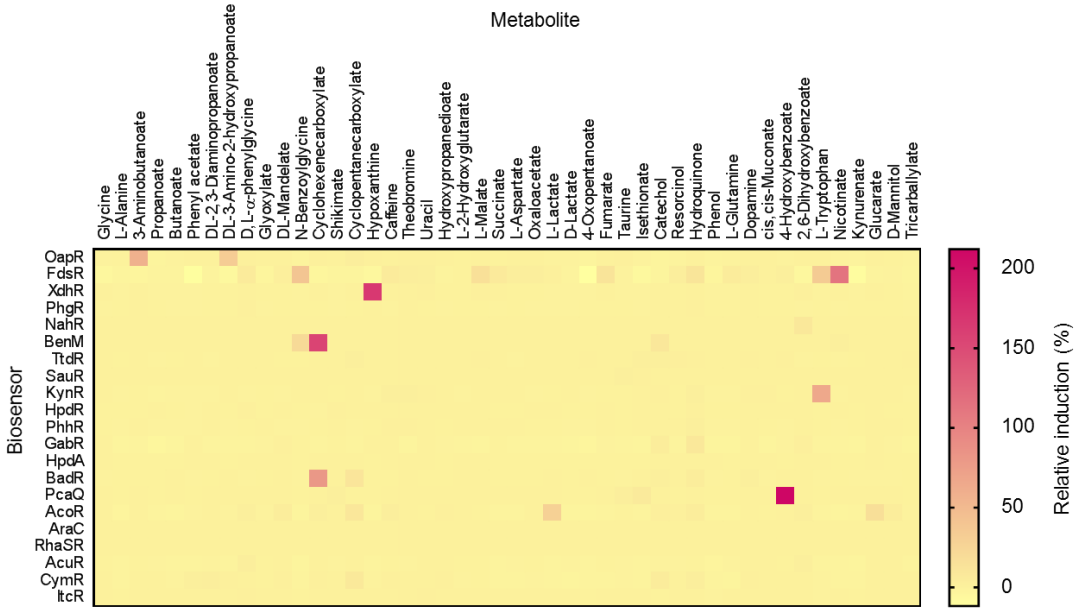


Figure 4.9 Screening of structurally similar and metabolically related compounds. Cross-reactivity of a set of 21 inducible systems and a library of 46 compounds. The heat map illustrates induction of reporter gene expression in the presence of metabolite (in %) relative to the induction mediated by the corresponding primary effector. Measurements were taken 6 h after supplementation with inducer at a final concentration of 5 mM (except in case of L-2-hydroxyglutarate which was added at a final concentration of 0.5 mM). Values represent the average of four replicates.

Based on the results of this screen, 11 of 12 metabolites may be classed into four groups. The first group comprises the compound/regulator pairs 3-aminobutanoate/OapR, DL-3-amino-2-hydroxypropanoate/OapR and cyclohexenecarboxylate/BenM. Neither of the three compounds has been reported to be metabolised by *E. coli* suggesting that they directly act as ligands. The response of the OapR/$P_{H16\_RS01330}$-β-alanine-inducible system to non-natural compounds, such as 3-aminobutanote and 3-amino-2-hydroxypropanoate, supports the initial claim of the structurally nearly identical β-alanine (3-aminopropanoate) being the primary ligand of OapR. The other three groups contain the compound/regulator pairs that did not result in an induction in *E. coli*. For example, nicotinate, hippurate, L-tryptophan, and 4-hydroxybenzoate fall into the category of metabolites that are likely to be converted into the primary effectors in *C. necator* mediating gene expression from the

FdsR/P$_{H16_RS03165}$-formate-, BenM/P$_{H16_RS09790}$-benzoate-, KynR/P$_{H16_RS14030}$-L-kynurenine-, and PcaQ/P$_{H16_RS30145}$-3,4-dihydroxybenzoate-inducible system, respectively. None of these catabolic pathways exist in *E. coli* which may explain their lack of induction. In the case of the hypoxanthine inducing the XdhR/P$_{H16_RS05055}$-xanthine-controllable system it is less clear whether hypoxanthine itself interacts with XdhR, due to structural resemblance to xanthine, or if activation of gene expression is mediated by its catabolic product, the primary effector. Interaction of hypoxanthine with XdhR could be tested by electrophoretic mobility shift assay. In *Streptomyces coelicolor*, hypoxanthine was not able to bind the regulator of the gene cluster encoding xanthine dehydrogenase enzyme[184]. However, in contrast to the *S. coelicolor* XdhR, which belongs to the TetR family of TRs[185], the *C. necator* XdhR is a LTTR which may operate in a different manner. The last group comprises the remaining compounds that could not be confirmed in *E. coli* but are likely to induce because of structural resemblance to the primary ligand. Direct interaction of cyclohexenecarboxylate and cyclopentanecarboxylate with BadR could not be confirmed due to dysfunctionality of the BadR/P$_{H16_RS27200}$-cyclohexanecarboxylate-inducible system in *E. coli*. The other two compounds, including 2,6-dihydroxybenzoate and catechol, might not be taken up by, or diffused into, *E. coli* cells and the activation of reporter gene expression by metabolically related compounds may be excluded as their consecutive degradation products, resorcinol and *cis,cis*-muconate, respectively, were not able to induce the NahR/P$_{H16_RS08125}$-salicylate- and BenM/P$_{H16_RS09790}$-benzoate-inducible systems in *C. necator* (Figure 4.9). Whole-cell extracts of *E. coli* carrying the salicylate- and benzoate-inducible systems could be used to test whether 2,6-dihydroxybenzoate and catechol are able to activate the respective systems *in vitro*.

Lastly, significant activation of the AcoR/$P_{H16\_RS19445}$-acetoin-inducible system by L-lactate was observed in *C. necator*, but not in *E. coli*. Since there is no characterised metabolic pathway for L-lactate to be converted into acetoin, and the AcoR/$P_{H16\_RS19445}$-acetoin-inducible system is highly activated by acetoin in both bacterial species, the response to L-lactate cannot be explained by either structural similarity or metabolic association.

4.4 Summary

Thus far, inducible gene expression systems have been discovered primarily by research focussed on the experimental characterisation of individual metabolic pathways and their regulation. This traditional approach has allowed the identification of a substantial number of such systems, some of which have been developed into the widely used gene expression control devices, e.g. AraC/P_{araBAD} and LacI/P_{T7}. However, this empirical approach, principally driven by interest in the pathway characterisation, usually delivers only a limited amount of new information on a specific inducible system. Recently, high throughput applications such as transcriptomics analysis, comparative genomics, and promoter or metagenome library screens, have proved to be efficient methodologies and substantially enhanced the speed of discovery of effector responsive promoters or even corresponding TR-promoter pairs[75, 157-162]. However, even these strategies suffer from several limitations, since they either do not ensure identification of all essential components of an inducible system or are TR- or ligand-type specific.

In this chapter, a methodical workflow was developed that allows inducible systems to be mined at the genome scale level enabling the extraction of information on all three components - the regulator, the inducible promoter, and its corresponding effector. To demonstrate the utility of the approach, it was applied to the genome of

the catabolically versatile *C. necator* H16. Sixteen putative inducible systems resulting from a pool that comprised over 400 TRs were identified in this single bacterium. Of these, inducible systems (OapR/$P_{H16\_RS01330}$ and PhgR/$P_{H16\_RS05530}$) responding to β-alanine and phenylglyoxylate have never previously been reported. Furthermore, tartrate- and sulfonatoacetate-inducible systems (TtdR/$P_{H16\_RS10665}$, SauR/$P_{H16\_RS13695}$) were experimentally validated. Amongst the characterised inducible systems, different types of TR were identified including LysR, AsnC, MocR, IclR and MarR. Fifteen of these were activator or dual function-type regulators, whereas the latter exhibited characteristics of a repressor (BadR/$P_{H16\_RS27200}$).

To further evaluate inducible systems, they were subjected to thorough quantitative characterisation assessing induction level, dynamics, and homogeneity. Several inducible systems exceeded induction levels of the frequently utilised L-arabinose-inducible system[171], with the BenM/$P_{H16\_RS09790}$-inducible system achieving expression level of more than 11-fold higher than AraC/P_{araBAD} in response to benzoate, highlighting its potential to be used for high level protein production in *C. necator*. Four inducible systems responding to salicylate, sulfonatoacetate, tartrate, and phenylglyoxylate exhibited a dynamic range of over 100-fold revealing that these systems are highly suitable to tightly control different levels of gene expression[186]. Along with a very high dynamic range of 650.7-fold, the NahR/$P_{H16\_RS08125}$-salicylate-inducible system responds to nM concentrations of salicylate. This degree of sensitivity is equivalent to the most sensitive of characterised inducible systems, those based on anhydrotetracycline and cumate[187, 188].

Orthogonal compatibility of inducible systems is a very important characteristic for designing multi-component and scalable circuits as well as sensory devices. These ideally require that functional genetic elements cross-react neither with

the host genetic background nor between different heterologous systems. Twelve of the newly discovered inducible systems showed distinctive response only to their primary ligands. The capacity to independently drive the expression of multiple genes was exemplified by combining the salicylate- and 3,4-dihydroxybenzoate-inducible systems exquisitely demonstrating the potential of newly characterised orthogonal switches for circuit design and other synthetic biology applications.

The utility of inducible systems was further demonstrated by applying individual promoter elements to control high levels of gene expression in *C. necator* (Figure 4.2d) and by employing TR-promoter pairs in the model bacteria *E. coli* and *P. putida* (Figure 4.5). Significantly, the adaptability of heterologous inducible system for gene expression control in other hosts was exemplified by engineering the β-alanine-inducible system for application in *E. coli* and *P. putida* (Figure 4.6).

To conclude, the genome scale approach and inducible system evaluation pipeline presented in this chapter, aids the discovery of new metabolite-controlled systems. Further, it delivers quantitative data on inducible systems dynamics and orthogonality expanding the potential of developing tuneable regulatory circuits and biosensors. Moreover, this generic approach can be utilised for mining inducible systems in any bacterial species facilitating the expansion of the toolbox for synthetic biology and biotechnology applications.

Chapter 5

The following chapter is mainly based on the work presented in the publication:

Characterisation of a 3-hydroxypropionic acid-inducible system from *Pseudomonas putida* for orthogonal gene expression control in *Escherichia coli* and *Cupriavidus necator*

E. K. R. Hanko, N. P. Minton & N. Malys (2017)

Scientific Reports ***7*** (1), 1724.[189]

5 Characterisation of a 3-hydroxypropionic acid-inducible system from *Pseudomonas putida* for orthogonal gene expression control in *Escherichia coli* and *Cupriavidus necator*

5.1 Introduction

Chemicals and fuels can be produced from renewable or waste feedstocks using metabolically engineered microorganisms, which may aid to reducing greenhouse gas emissions[90, 190, 191]. Recently, significant research efforts have been directed towards developing microorganisms for the biosynthesis of value-added chemicals, including 3-hydroxypropionic acid (3-HP)[192-195]. 3-HP is a biotechnologically attractive platform chemical that can be used as precursor for the biosynthesis of acetaldehyde, acrylate, acrylamide, methylacrylate, and 1,3-propanediol[194], as well as biodegradable polymer poly-3-HP[196]. A number of recombinant strains using *Corynebacterium glutamicum, E. coli, Klebsiella pneumoniae, Lactobacillus reuteri, Pseudomonas denitrificans, Synechocystis* sp., *Synechococcus elongates* and *Saccharomyces cerevisiae* as chassis[197-204], and a few alternative metabolic pathways have been developed for 3-HP biosynthesis using intermediates such as β-alanine, malonyl-CoA, propionyl-CoA, glycerol and lactate[205-209]. Although relatively high titres of 3-HP have been reported in *E. coli* and *K. pneumoniae*[206, 210], the challenge remains to develop a sustainable biotechnological production of this carboxylic acid[211].

3-HP can be efficiently assimilated and utilised as a carbon and energy source by bacteria. *P. denitrificans* possesses 3-HP dehydrogenase- and 3-hydroxyisobutyrate dehydrogenase activities which contribute to 3-HP degradation[198, 212]. The expression of some genes related to 3-HP metabolism in

P. denitrificans have been shown to be strongly induced by 3-HP and putatively controlled by TRs[212], which belong to the family of LTTRs[30].

LTTRs contain a conserved protein structure with a DNA-binding helix-turn-helix motif at the N-terminus and an effector-binding domain at the C-terminus. Typically activated by small effector molecules, they regulate the transcription of metabolism-related genes. Most often, LTTRs are encoded in opposite direction of the gene or operon that they regulate. They usually interact with at least two TR-binding sites in the intergenic region[30, 164]. The regulatory binding site is located 60 to 80 nucleotides upstream with respect to the transcriptional start site of the ligand-metabolising gene. The motif has been characterised as T-N_{11}-A, which can vary in length and/or both nucleotide composition. Frequently, this site overlaps with the promoter of the LTTR gene mediating negative autoregulation. The activator-binding site is usually adjacent or overlaps with the -35 box of the promoter[30]. Besides, examples with a single and multiple LTTR-binding sites have also been reported[213, 214]. Based on full length and domain analysis, in *Pseudomonas* LTTRs can be dissected into nine evolutionary different phylogenetic groups[215].

LTTR- and other TR-based-inducible (positively or negatively regulated) systems control microbial gene expression in response to the change of intracellular levels of metabolites and play an important role in governing metabolic pathways and networks. Recently, a number of such inducible systems have been adapted as genetically encoded biosensors that respond to a variety of native and non-native compounds[69, 74, 216-218]. Both native and synthetic systems, in combination with a reporter gene, have been used to screen for metabolically engineered microbial strains enabling the selection for microorganisms with improved production of target compounds[64, 219-221]. They have also been applied as metabolite-responsive gene

switches and dynamic regulators of metabolic pathways, in which levels of upstream or downstream gene expression are continuously adjusted to balance metabolic intermediate levels increasing the flux towards the product of interest[221, 222].

Pseudomonas are an attractive source for exploration of novel gene targets such as TR-based inducible systems. Particularly, *P. putida* KT2440 exhibits combinations of features characteristic to aquatic oligotrophs and terrestrial copiotrophs indicating that this bacterium has adapted functional capabilities permitting to thrive in various environments[223]. This genetically diverse microorganism contains a high affinity nutrient acquisition and metabolic efflux, as well as catabolic enzymes such as mono- and dioxygenases, oxidoreductases, and dehydrogenases. Additionally, *P. putida* possesses a wide range of gene expression control systems involving various sigma factors and regulators which form a rich basis for the exceptional metabolism versatility.

Here, a 3-HP-inducible system from *P. putida* is identified and characterised. Composed of a LTTR and a corresponding 3-HP-responsive promoter, the data in this chapter demonstrate that it can be used to control gene expression orthogonally in *E. coli* and *C. necator*. A comprehensive analysis of the promoter region is performed to establish a consensus sequence required for potential TR binding. The characterised inducible system can be exploited as 3-HP biosensor, as it shows a high specificity and a wide induction range for this compound, as well as genetic element for the construction of autoregulated metabolic pathways aiming to improve the production of bio-based 3-HP.

5.2 Materials and methods specific to this chapter

5.2.1 Chemicals

3-Hydroxypropionic acid (3-HP), glycerol, pyruvic acid, β-alanine, ethanol, 1-propanol, 2-propanol, 1-butanol, 2-butanol, 1,3-butanediol, sodium acetate, propionic acid, butyric acid, malonic acid, sodium succinate dibasic hexahydrate, D-malic acid, L-malic acid, sodium glycolate, lactic acid, D,L-2-hydroxybutyric acid sodium salt (2-HB), D-3-hydroxybutyric acid (D-3-HB), L-3-hydroxybutyric acid (L-3-HB), and sodium salicylate were used as inducers for assaying the 3-HP-inducible system. All chemicals are listed in Supplementary Table 1 in the *Appendix*.

5.2.2 Plasmids

Key features of all plasmids used and generated in this chapter are summarised in Table 5.1. A detailed description of how each plasmid was assembled is provided in the *Appendix*. The nucleotide sequence of plasmids pEH007, pEH008, pEH009, and pEH010 have been deposited in the public version of the JBEI registry (https://public-registry.jbei.org) under the accession numbers JPUB_008751-JPUB_008754.

Table 5.1 Plasmids used in chapter 5.

Plasmid	Characteristic	Reference or source
pGEX-6P-1	Amp^r; P_{tac}-*gst*	GE Healthcare
pEH006	Cm^r; P_{araC}-*araC*-T_{rrnB1}; P_{araBAD}-T7sl-*rfp*-T_{dbl}	This work
pEH007	Cm^r; *Pp*P_{mmsR}-*mmsR*-T_{rrnB1}; *Pp*P_{mmsA}-*rfp*-T_{dbl}	This work
pEH008	Cm^r; *Pp*P_{hpdR}-*hpdR*-T_{rrnB1}; *Pp*P_{hpdH}-*rfp*-T_{dbl}	This work
pEH009	Cm^r; *Cn*P_{araC}-*araC*-T_{rrnB1}; *Cn*P_{mmsA2}-*rfp*-T_{dbl}	This work
pEH010	Cm^r; *Cn*P_{hpdR}-*hpdR*-T_{rrnB1}; *Cn*P_{mmsA1}-*rfp*-T_{dbl}	This work
pEH022	Cm^r; *Pp*P_{hpdH}-*rfp*-T_{dbl}	This work
pEH034	Cm^r; *Pp*P_{mmsA}-*rfp*-T_{dbl}	This work
pEH036	Cm^r; *Pp*P_{hpdR}-*hpdR*-T_{rrnB1}; *Pp*P_{hpdH}(-118)-*rfp*-T_{dbl}	This work

pEH053	Cm^r; *Pp*P$_{hpdR}$-*hpdR*-T$_{rrnB1}$; *Pp*P$_{hpdH}$(-106)-*rfp*-T$_{dbl}$	This work
pEH068	Cm^r; *Pp*P$_{hpdR}$-*hpdR*-T$_{rrnB1}$; *Pp*P$_{hpdH}$(-118_mut2)-*rfp*-T$_{dbl}$	This work
pEH069	Cm^r; *Pp*P$_{hpdR}$-*hpdR*-T$_{rrnB1}$; *Pp*P$_{hpdH}$(-118_mut3)-*rfp*-T$_{dbl}$	This work
pEH070	Cm^r; *Pp*P$_{hpdR}$-*hpdR*-T$_{rrnB1}$; *Pp*P$_{hpdH}$(-118_mut4)-*rfp*-T$_{dbl}$	This work
pEH071	Cm^r; *Pp*P$_{hpdR}$-*hpdR*-T$_{rrnB1}$; *Pp*P$_{hpdH}$(-118_mut5)-*rfp*-T$_{dbl}$	This work
pEH072	Cm^r; *Pp*P$_{hpdR}$-*hpdR*-T$_{rrnB1}$; *Pp*P$_{hpdH}$(-118_mut6)-*rfp*-T$_{dbl}$	This work
pEH073	Cm^r; *Pp*P$_{hpdR}$-*hpdR*-T$_{rrnB1}$; *Pp*P$_{hpdH}$(-118_mut7)-*rfp*-T$_{dbl}$	This work
pEH074	Cm^r; *Pp*P$_{hpdR}$-*hpdR*-T$_{rrnB1}$; *Pp*P$_{hpdH}$(-118_mut8)-*rfp*-T$_{dbl}$	This work
pEH075	Cm^r; *Pp*P$_{hpdR}$-*hpdR*-T$_{rrnB1}$; *Pp*P$_{hpdH}$(-118_mut9)-*rfp*-T$_{dbl}$	This work
pEH076	Cm^r; *Pp*P$_{hpdR}$-*hpdR*-T$_{rrnB1}$; *Pp*P$_{hpdH}$(-118_mut10)-*rfp*-T$_{dbl}$	This work
pEH077	Cm^r; *Pp*P$_{hpdR}$-*hpdR*-T$_{rrnB1}$; *Pp*P$_{hpdH}$(-118_mut11)-*rfp*-T$_{dbl}$	This work
pEH079	Cm^r; *Pp*P$_{hpdR}$-*hpdR*-T$_{rrnB1}$; *Pp*P$_{hpdH}$(-118_mut1)-*rfp*-T$_{dbl}$	This work
pEH080	Cm^r; *Pp*P$_{hpdR}$-*hpdR*-T$_{rrnB1}$; *Pp*P$_{hpdH}$(-118_mut12)-*rfp*-T$_{dbl}$	This work
pEH089	Amp^r; P$_{tac}$-*gst-hpdR*	This work

5.2.3 Determination of consensus sequence

The nucleotide sequences of all *Pseudomonas hpdR*/*hpdH* intergenic regions were retrieved from GenBank[165] at NCBI (www.ncbi.nlm.nih.gov) and aligned using Clustal Omega[224, 225]. Subsequently, a sequence similarity motif was generated using WebLogo[226]. Putative RNAP binding sites and the *hpdH* TSS were predicted using programs BPROM[227] and NNPP[228], respectively.

5.2.4 Expression and purification of HpdR

Multiple attempts to purify the transcriptional regulator HpdR by His- or Strep-tag affinity chromatography resulted in insoluble protein, confirming previous observations that a majority of members from the LysR family of TRs tend to readily form inclusion bodies[229]. However, a small amount of soluble protein was obtained as a glutathione S-transferase (GST)-tagged version of HpdR. To decidedly ensure that potential DNA-shifts are caused by HpdR binding, *gst* was simultaneously expressed from the original vector that was employed to construct the GST-HpdR fusion protein (pGEX-6P-1), purified and subjected to electrophoretic mobility shift assay.

A 100 mL volume of fresh LB medium containing ampicillin (100 µg/mL) were inoculated 1:100 with an overnight culture of *E. coli* Rosetta DE3 harbouring either pGEX-6P-1 or pEH089. The cultures were grown in 0.5-L baffled shake flasks with orbital shaking at 37 °C and 200 rpm. At an OD_{600} of 0.5, protein expression was induced using 0.5 mM IPTG. The cultures were grown for another 3 h with orbital shaking at 20 °C and 220 rpm. Subsequently, cells were harvested by centrifugation at 6,000*g* and 4 °C for 8 min, resuspended in 2 mL of PBS and centrifuged as before. The supernatant was removed and bacterial pellets were frozen at -20 °C for 24 h.

Thawed cells were resuspended in 1 mL of lysis buffer (50 mM HEPES pH 7.5, 300 mM NaCl, 10 mM $MgSO_4$, 0.2% NP-40, 5 µL of protease inhibitor cocktail III (Calbiochem), 10 ng DNase I (Thermo Fisher Scientific)) and lysed by sonication (10 cycles à 10 micron, 30 seconds on ice in between cycles). The cell lysate was cleared by centrifugation at 21,130*g* and 4 °C for 30 min. The supernatant was added to 100 µL of glutathione sepharose 4B beads (GE Healthcare) which had been equilibrated three times with 0.5 mL of wash buffer (50 mM HEPES pH 7.5, 150 mM NaCl, 0.2% NP-40). The suspension was incubated at 150 rpm and 4 °C for 2 h on a horizontal shaker. Subsequently, the sepharose beads were pelleted by centrifugation at 500*g* at 4 °C for 5 min and washed with wash buffer. This step was repeated four times. The supernatant was removed, beads were resuspended in 100 µL of elution buffer A (50 mM HEPES pH 7.9, 10 mM reduced L-glutathione (Sigma-Aldrich), 0.2% NP-40) and incubated at 100 rpm and 4 °C for 1 h. To the suspension, 100 µL of protease cleavage buffer (50 mM HEPES pH 7.9, 300 mM NaCl, 2 mM EDTA, 2 mM DTT, 0.2% NP-40) were added together with 12 units of PreScission Protease (Protease-GST fusion, GE healthcare) and shaken at 100 rpm and 4 °C for another 16 h. The PreScission Protease-cleaved and in the supernatant released proteins (elution

fraction A) were separated from the beads by centrifugation at 500*g* and 4 °C for 5 min. Remaining unbound proteins were collected by washing twice with 150 µL of elution buffer B (50 mM HEPES pH 7.9, 150 mM NaCl, 1 mM EDTA, 1 mM DTT, 0.2% NP-40). Protein samples derived from the pGEX-6P-1 and pEH089 cultures were analysed by NuPAGE (4-12% Bis-Tris gel in MES running buffer, Invitrogen) according to the manufacturer's protocol.

5.2.5 Electrophoretic mobility shift assay (EMSA)

The purified proteins from elution fraction A were analysed for specific binding to the native *P. putida* KT2440 *hpdR*/*hpdH* intergenic region. Prior EMSA, the concentration of HpdR was determined by comparison to NuPAGE of a bovine serum albumin (BSA) dilution series.

To visualise TR binding, the *hpdR*/*hpdH* intergenic region was labelled with the fluorescent dye ATTO 700. The labelled DNA was generated by PCR of pEH008 with oligonucleotide primers EHseq018 and EHseq003. EHseq018 is ATTO 700-labelled at its 5' end. The resulting PCR product of 310 bp was gel-purified. The total volume of the DNA-protein-binding reaction was 10 µL and contained 50 pM labelled PCR product, 10 ng/µL salmon sperm DNA, 25 mM HEPES pH 7.9, 75 mM NaCl, 0.5 mM EDTA, 0.5 mM DTT, 0.1% NP-40, 150 mM KCl, and 2.5 mM $MgCl_2$. A high KCl concentration has been demonstrated to facilitate DNA-protein-complex formation[230]. Furthermore, the binding reaction contained 300 nM HpdR (elution fraction A of the culture harbouring pEH089), an equal volume of the elution fraction A derived from the culture harbouring pGEX-6P-1 or no protein. If required, 3-HP was added at the final concentration of 1 mM. Samples were incubated at 20 °C for 30 min. 10 µL of 50% (v/v) glycerol were added to each sample before being loaded onto a non-denaturing 8% Tris-Glycine gel (Invitrogen). Electrophoresis was performed in

Tris-Glycine native running buffer (Invitrogen) at 4 °C for 2 h and 120V. The gel was scanned using the Odyssey Clx infrared imaging system (Model 9120, LI-COR Biosciences) and analysed using software Image Studio Lite version 5.2 (LI-COR Biosciences).

5.2.6 Growth conditions

Single time-point and time course fluorescence measurements in *E. coli* MG1655 and *C. necator* H16 were performed in minimal medium.

5.3 Results

5.3.1 Identification of a 3-HP-inducible system in *P. putida* KT2440 and *C. necator* H16

Three enzymes have been identified in *P. denitrificans* to be involved in 3-HP degradation[212]. They are arranged in two operons. The 3-hydroxypropionate dehydrogenase (*hpdH*) in one operon and the methylmalonate-semialdehyde dehydrogenase (*mmsA*) and 3-hydroxyisobutyrate dehydrogenase (*mmsB*, also referred to as *hbdH*-4) in the second operon. In opposite direction of each operon, a gene encoding a LTTR is located which was proposed to be required for 3-HP-inducible activation of gene expression of its respective operon[212]. Homologues of the 3-HP catabolic genes have been found in various microbial genera including *Cupriavidus*[212]. Here, putative HpdH, MmsA, and MmsB homologues were identified in the genomes of *P. putida* KT2440 and *C. necator* H16 by protein-protein blast.

In *P. putida*, the 3-HP catabolic genes and their respective LTTRs are arranged as in *P. denitrificans* (Figure 5.1). In *C. necator*, however, their arrangement is different. A *P. denitrificans* MmsA homolog with 49% protein sequence identity and 96% coverage is encoded between the *C. necator hpdH* and its putative LTTR (here termed HpdR). The short intergenic region of 27 bp between the *mmsA* homolog (here

termed *mmsA*1) and *hpdH* suggests a polycistronic transcription of genes from a promoter located in the *hpdR*/*mmsA*1 intergenic region. For *mmsA* (here termed *mmsA*2) and *mmsB*, both arranged in one operon, no LTTR was found upstream of the genes. An acyl-CoA dehydrogenase, encoded by *acaD*, is annotated upstream of *mmsA*2. Both genes are oriented in the same direction. They are separated only by a short intergenic region of 44 bp suggesting an operonic arrangement of genes. In opposite orientation of *acaD*, a TR is located which is annotated to belong to the AraC family of TRs. It does not share protein sequence similarity with the *P. denitrificans* LTTR, which is putatively required to activate transcription of the *mmsAB* operon. However, it was included in the analysis for 3-HP-inducible gene expression as potential regulator of the *C. necator acaD*-*mmsA*2-*mmsB* operon.

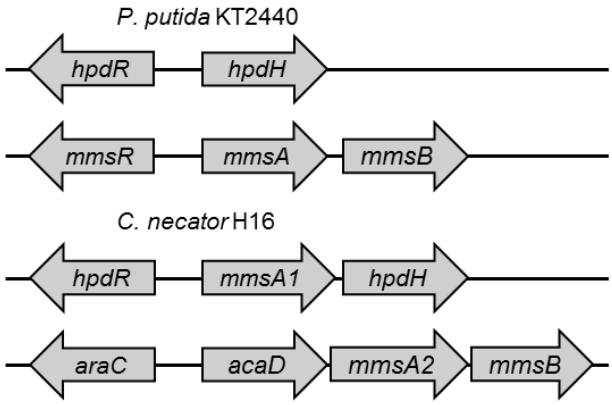


Figure 5.1 Operons putatively involved in 3-HP metabolism in *P. putida* KT2440 and *C. necator* H16. They are composed of the TR genes *hpdR*, *mmsR*, or *araC*, and divergently transcribed genes putatively required for 3-HP degradation: *hpdH*, 3-hydroxypropionate dehydrogenase; *mmsA*, methylmalonate-semialdehyde dehydrogenase; *mmsB*, 3-hydroxyisobutyrate dehydrogenase; and *acaD*, acyl-CoA dehydrogenase.

5.3.2 *P. putida* 3-HP-inducible systems outperform systems derived from *C. necator*

To evaluate whether expression of the identified catabolic gene clusters is activated in the presence of 3-HP, the putative inducible systems were cloned upstream of the *rfp* reporter gene in pEH006. They comprise the TR and the intergenic region between the

regulator gene and the translational start site of the 3-HP catabolic gene clusters. The two systems from *P. putida* KT2440 are referred to as *Pp*MmsR/P_{mmsA} (pEH007) and *Pp*HpdR/P_{hpdH} (pEH008) and those from *C. necator* H16 as *Cn*AraC/P_{acaD} (pEH009) and *Cn*HpdR/P_{mmsA1} (pEH010). The constructs were tested in *E. coli* MG1655 and *C. necator* H16 for *rfp* reporter gene expression in response to 3-HP.

Both 3-HP-inducible systems from *P. putida* exhibit statistically significant induction in *E. coli* and *C. necator* 6 h after addition of 3-HP at a final concentration of 10 mM (Table 5.2). *Pp*HpdR/P_{hpdH} demonstrates the highest level of normalised fluorescence and the strongest inductions, 23.3-fold in *E. coli* and more than 500-fold in *C. necator*. In contrast to *C. necator* in which *Cn*HpdR/P_{mmsA1} shows an 88.4-fold induction, neither of the TRs derived from *Cupriavidus* seem to be able to activate reporter gene expression from their proposed cognate promoters in *E. coli*. Besides, *Cn*AraC/P_{acaD} does not mediate a statistically significant induction after extracellular addition of 3-HP in *C. necator*.

Table 5.2 Quantitative evaluation of 3-HP-inducible systems in *E. coli* and *C. necator*. Absolute normalised fluorescence of cells carrying different versions of putative 3-HP-inducible systems from *P. putida* KT2440 (*Pp*) and *C. necator* H16 (*Cn*) in the presence or absence of 10 mM 3-HP. Each system is composed of either a putative 3-HP-inducible promoter and its corresponding TR or a 3-HP-inducible promoter only. The mean values and standard deviations represent the absolute normalised fluorescence of biological triplicates 6 h after addition of 3-HP. Asterisks indicate statistically significant induction values for $p < 0.01$ (unpaired *t* test).

Inducible system	*E. coli* MG1655			*C. necator* H16		
	Normalised absolute fluorescence		Induction ratio	Normalised absolute fluorescence		Induction ratio
	Uninduced	Induced		Uninduced	Induced	
*Pp*MmsR/P$_{mmsA}$	59 ± 3	721 ± 24	12.3*	444 ± 80	22,857 ± 1808	51.5*
*Pp*HpdR/P$_{hpdH}$	1,259 ± 68	29,351 ± 756	23.3*	304 ± 46	157,052 ± 8,409	516.6*
*Cn*AraC/P$_{acaD}$	0 ± 0	0 ± 0	0	311 ± 16	464 ± 69	1.5
*Cn*HpdR/P$_{mmsA1}$	22 ± 6	26 ± 5	1.2	847 ± 27	74,839 ± 1,763	88.4*
*Pp*P$_{mmsA}$	193 ± 16	243 ± 23	1.3	44 ± 8	42 ± 4	1.0
*Pp*P$_{hpdH}$	908 ± 60	1,077 ± 47	1.2	42 ± 1	40 ± 3	1.0

5.3.3 The *P. putida* 3-HP-inducible systems demonstrate a TR-dependent orthogonality in *E. coli* and *C. necator*

Ideally, TR-based inducible gene expression systems are specific to a certain effector and solely controlled by its corresponding TR. To examine the cross-reactivity of host-originating transcription factors on reporter gene expression, the regulator genes were removed from the plasmids that initially contained the *Pp*MmsR/P$_{mmsA}$ and *Pp*HpdR/P$_{hpdH}$ inducible systems. The new constructs are composed solely of the 3-HP-inducible promoters P$_{mmsA}$ and P$_{hpdH}$ transcriptionally fused to the *rfp* reporter gene. *Cn*HpdR/P$_{mmsA1}$ and *Cn*AraC/P$_{acaD}$ were not further investigated since they demonstrated no induction of RFP synthesis upon supplementation with 3-HP in

E. coli and only in case of *Cn*HpdR/P_{mmsA1} a statistically significant induction in *C. necator*.

In both *E. coli* and *C. necator*, neither of the promoters demonstrated a significant induction of *rfp* expression in the presence of 3-HP (Table 5.2). This suggests that their corresponding LTTRs are required for transcription activation and that neither of the two tested microorganisms have cross-reacting TR homologues. Since P_{hpdH} exhibited the highest induction level of the analysed 3-HP-inducible systems, and is controlled independently from host-originating TRs in *E. coli* and *C. necator*, it was chosen to be further characterised.

5.3.4 Identification of a conserved sequence motif within the *hpdH* promoter and *in vivo* analysis of the HpdR binding site

Control of gene expression in response to small effector molecules is mainly mediated by TRs. They interact with conserved TFBSs in the promoters of genes which are involved in effector-metabolising pathways. A simple approach to identify TFBSs is to perform a phylogenetic footprinting analysis[231]. By multiple sequence alignment of upstream regions of orthologous genes extracted from a variety of divergent species, conserved functional elements can easily be distinguished from sequences that are prone to evolve more quickly.

P. putida KT2440 HpdR and HpdH homologues were screened in other *Pseudomonas* species by searching the NCBI database for non-redundant protein sequences. Forty different *Pseudomonas* species were selected with HpdR and HpdH protein sequence identities of at least 70% (Table 5.3). To aid in the identification of the putative HpdR binding site, the consensus sequence was determined in the intergenic region as described in *5.2.3*. The *P. putida* KT2440 *hpdR*/*hpdH* intergenic region is illustrated in Figure 5.2. Upstream of the -35 region, a range of nucleotides

was identified which is highly conserved across the analysed *Pseudomonas* species. This region is illustrated as a sequence similarity motif in Figure 5.3. It corresponds to the nucleotide sequence between position -118 and -68 relative to the *P. putida* KT2440 *hpdH* translational start site. Two partially similar and conserved sequence motifs can be identified. The first one is an imperfect inverted repeat, sequence GCCCCTGTGC-N_6-GCACAGCGGC in *P. putida*, located between positions -109 and -84 and is referred to as TFBS1. The second motif is located between positions -83 and -68 and is denoted TFBS2. In order to test if the highly conserved nucleotides spanning positions -109 to -68 are essential for promoter activation mediated by HpdR in response to 3-HP, the *P. putida* KT2440 *hpdR*/*hpdH* intergenic region was truncated and subsequently mutated.

Table 5.3 *P. putida* KT2440 HpdR and HpdH homologues. List of HpdR and HpdH homologues extracted from various *Pseudomonas* species sharing at least 70% protein sequence identity with *P. putida* KT2440 HpdR and HpdH.

Species	HpdR locus tag	Size (aa)	Coverage % (identity %)	HpdH locus tag	Size (aa)	Coverage % (identity %)
P. putida KT2440	PP_0055	295	100 (100)	PP_0056	550	100 (100)
P. entomophila L48	PSEEN_RS00045	295	100 (96)	PSEEN_RS00050	549	99 (91)
P. parafulva NBRC 16636	PPA02S_RS02005	295	100 (96)	PPA02S_RS02000	548	99 (91)
P. plecoglossicida NBRC 103162	PPL01S_RS04640	295	100 (95)	PPL01S_RS04635	548	99 (95)
P. monteilii GTC 10897	APH46_RS25120	295	99 (95)	APH46_RS25115	548	99 (94)
P. mosselii SJ10	O165_RS23295	295	100 (95)	O165_RS23300	549	99 (89)
P. alkylphenolia KL28	PSAKL28_RS00210	295	100 (94)	PSAKL28_RS00215	549	99 (83)
P. vranovensis DSM 16006	H621_RS0106435	295	100 (94)	H621_RS0106440	549	99 (84)
P. japonica NBRC 103040	PJA01S_RS05195	295	100 (93)	PJA01S_RS05200	549	96 (81)
P. gingeri NCPPB 3146	PGING_RS31890	295	99 (88)	PGING_RS31885	553	96 (81)
P. fuscovaginae IRRI 6609	PF66_RS17780	295	100 (88)	PF66_RS17775	545	97 (81)
P. taetrolens DSM 21104	TU78_RS12205	295	99 (87)	TU78_RS12200	553	96 (79)
P. fragi B25	O5A_RS0119340	295	99 (86)	O5A_RS0119330	553	97 (80)

P. psychrophila HA-4	B347_RS0107760	295	99 (86)	B347_RS0107770	553	96 (80)
P. fluorescens AU12597	AA053_RS20570	295	99 (86)	AA053_RS20565	553	96 (78)
P. agarici NCPPB 2289	PAGAR_RS011339 0	295	100 (87)	PAGAR_RS011338 5	553	96 (80)
P. helleri DSM 29165	TU84_RS23875	295	100 (85)	TU84_RS23880	553	96 (80)
P. chlororaphis 30-84	PCHL3084_RS1216 5	297	100 (85)	PCHL3084_RS1216 0	548	96 (79)
P. deceptionensis DSM 26521	TR67_RS22630	295	99 (86)	TR67_RS22625	553	96 (79)
P. batumici UCM B-321	UCMB321_5547	295	100 (89)	UCMB321_5548	549	99 (80)
P. lini ZBG1	ACS73_RS06805	294	99 (85)	ACS73_RS06800	548	96 (80)
P. umsongensis UNC430CL58Col	N519_RS0117920	308	98 (85)	N519_RS0117915	549	99 (79)
P. cremoricolorata ND07	LK03_RS06240	295	100 (87)	LK03_RS06245	549	99 (81)
P. protegens Cab57	PPC_RS12670	297	99 (83)	PPC_RS12665	548	98 (77)
P. frederiksbergensis SI8	JZ00_RS11015	294	99 (84)	JZ00_RS11010	547	97 (79)
P. mediterranea TEIC1105	ADY55_RS24405	294	99 (84)	ADY55_RS24410	547	97 (80)
P. brassicacearum PA1G7	AW28_RS07090	294	99 (83)	AW28_RS07085	547	95 (81)
P. corrugata TEIC1148	ADY50_RS18415	294	99 (82)	ADY50_RS18420	547	97 (79)
P. trivialis IHBB745	AA957_RS03805	297	100 (79)	AA957_RS03810	548	96 (78)

P. orientalis DSM 17489	TU82_RS23530	296	100 (80)	TU82_RS23535	548	96 (77)
P. poae DSM 14936	TU75_RS20075	298	99 (79)	TU75_RS20070	548	96 (78)
P. tolaasii 6264	UQW_RS0110215	297	100 (79)	UQW_RS0110210	548	98 (78)
P. veronii R4	SU91_RS20320	297	99 (79)	SU91_RS20325	555	96 (78)
P. synxantha DSM 18928	TU77_RS19905	297	100 (79)	TU77_RS19910	548	99 (77)
P. marginalis ICMP 9505	AO391_RS22020	297	100 (79)	AO391_RS22025	548	96 (77)
P. rhizosphaerae IH5	LT40_RS16020	295	99 (83)	LT40_RS16015	550	98 (78)
P. aeruginosa AZPAE14918	NS34_RS16205	301	99 (73)	NS34_RS16200	557	96 (74)
P. denitrificans ATCC 13867	H681_RS18595	304	99 (73)	H681_RS18590	554	96 (73)
P. knackmussii B13	PKB_RS05640	301	99 (72)	PKB_RS05645	552	97 (72)
P. nitroreducens Aramco J	QX33_RS11735	300	99 (74)	QX33_RS11740	554	96 (73)
P. citronellolis TTU2014-008ASC	AO742_RS16215	297	99 (70)	AO742_RS16210	552	97 (73)

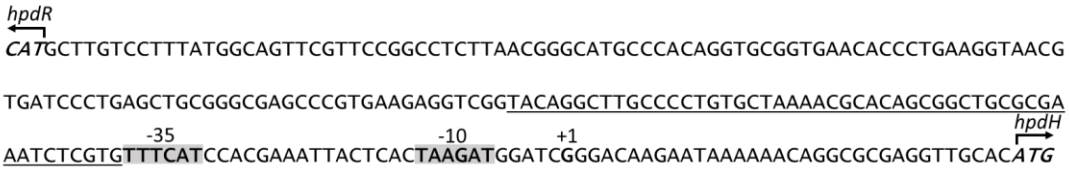


Figure 5.2 The *P. putida* KT2440 *hpdR*/*hpdH* intergenic region. Translational start sites are bold and italicised. The predicted *hpdH* -35 and -10 promoter elements are bold and highlighted in grey. The nucleotide sequence between position -118 and -68 relative to the *hpdH* translational start site is underlined.

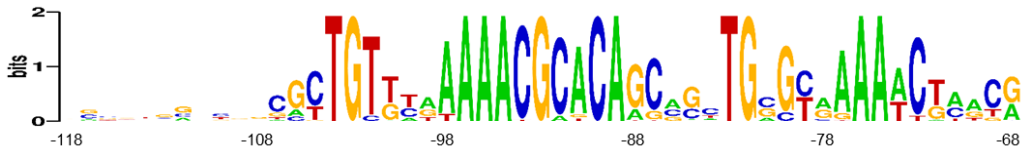


Figure 5.3 DNA sequence similarity motif of the putative HpdR binding site. The sequence similarity motif corresponds to the underlined sequence in Figure 5.2. The motif represents highly conserved nucleotides in the *hpdR*/*hpdH* intergenic regions of forty *Pseudomonas* species.

Firstly, the -233 bp long intergenic region was analysed for secondary structures using the mfold web server[232]. A palindromic sequence was identified which can form a 97 bp spanning stem-loop structure (Supplementary Figure 5). Interestingly, from the forty selected *Pseudomonas* species, this secondary structure is only present in *P. putida* KT2440 and the closely related *Pseudomonas entomophila* L48 and *Pseudomonas plecoglossicida* NBRC 103162 (Table 5.3). Likely, it has a species specific functional role, but is not evolutionary conserved. The reporter gene construct harbouring the native *P. putida* KT2440 *hpdR*/*hpdH* intergenic region is referred to as *hpdH*-233::*rfp*. The promoter truncation *hpdH*-118::*rfp* was generated by removing the palindromic sequence and incorporating an AvrII restriction site to allow further promoter modification (Figure 5.4a). It contains the 118 bp long sequence upstream of the *hpdH* translational start site. Using *hpdH*-118::*rfp*, the promoter was truncated by additional twelve nucleotides, resulting in *hpdH*-106::*rfp*. The promoter-reporter gene constructs were analysed in *E. coli* MG1655 for their

response to 3-HP (Figure 5.4b). Compared to the native intergenic region (*hpdH*-233::*rfp*), the induction level slightly increased when the palindromic sequence was removed (*hpdH*-118::*rfp*). However, further truncation of the *hpdH* promoter by 12 nucleotides (*hpdH*-106::*rfp*) resulted in a 10-fold decrease in induction compared to *hpdH*-233::*rfp*.

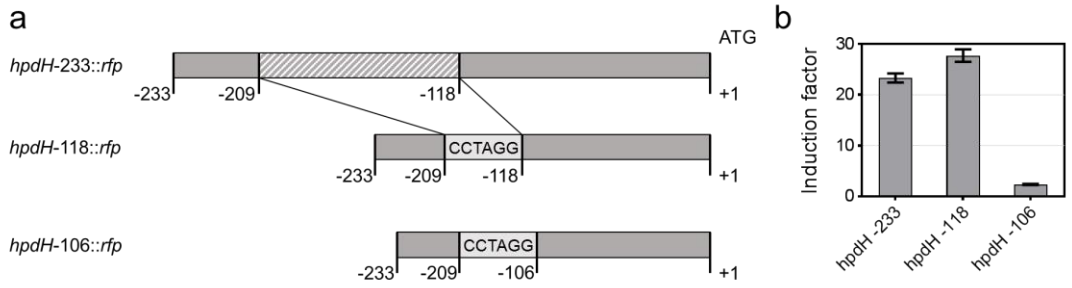


Figure 5.4 *P. putida* KT2440 *hpdH* promoter truncations and effect on 3-HP-inducible gene expression. (a) Schematic illustration of the *hpdH* promoter truncations that were evaluated for 3-HP-inducible activation of reporter gene expression. The native promoter-reporter construct is denoted *hpdH*-233::*rfp*. The bioinformatically identified palindromic sequence is illustrated as a dashed box. This sequence was replaced by an AvrII restriction site, CCTAGG, resulting in promoter truncations *hpdH*-118::*rfp* and *hpdH*-106::*rfp*. (b) Induction levels mediated by the original promoter and the promoter truncations. The promoter-reporter gene constructs were analysed in *E. coli* MG1655 for RFP synthesis in the absence and presence of 10 mM 3-HP. Error bars represent standard deviations of three biological replicates.

Since *hpdH*-118::*rfp* demonstrated similar induction levels to the native, untruncated intergenic region, it was used to generate *hpdH* promoter mutants. The use of random mutagenesis was deemed to be a combinatorial challenge, seeking to achieve a full set of completely random mutants covering the 60-bp long conserved motif. Therefore, mutations were designed in such a way that they enabled a complete coverage by employing a previously reported mutagenesis strategy[132]. In total, twelve promoter variants were constructed, harbouring single or multiple nucleotide mutations between position -118 and -68 (Figure 5.5a). They were analysed in *E. coli* MG1655 for reporter gene expression in the absence and presence of 3-HP (Figure 5.5b). By comparing the sequence similarity motif (Figure 5.3) with the induction

levels of the different promoter variants, it can be observed that the extent of nucleotide conservation correlates with their importance for inducible gene expression. Mutations 1, 2, and 3, which are upstream of the -109 to -68 region, show a minor impact on promoter activity (1.5- to 1.7-fold decrease in induction) compared to *hpdH*-118::*rfp*. Mutations 4, 8, and 11, which are located in less conserved stretches of the -109 to -68 region, decrease induction of reporter gene expression by 5.5-, 5.4-, and 1.6-fold, respectively. The remaining five mutations, 5, 6, 7, 9, and 10 alter the most conserved nucleotides in either TFBS1 or TFBS2. These mutations abolish the ability of the promoter to mediate controllable gene expression by decreasing induction levels from 11.5- (mutation 6) to 25.2-fold (mutation 7). The *in vivo* analysis of various *hpdH* promoter truncations and mutations suggests that a 50 bp long sequence upstream of the -35 region is involved in 3-HP-inducible gene expression.

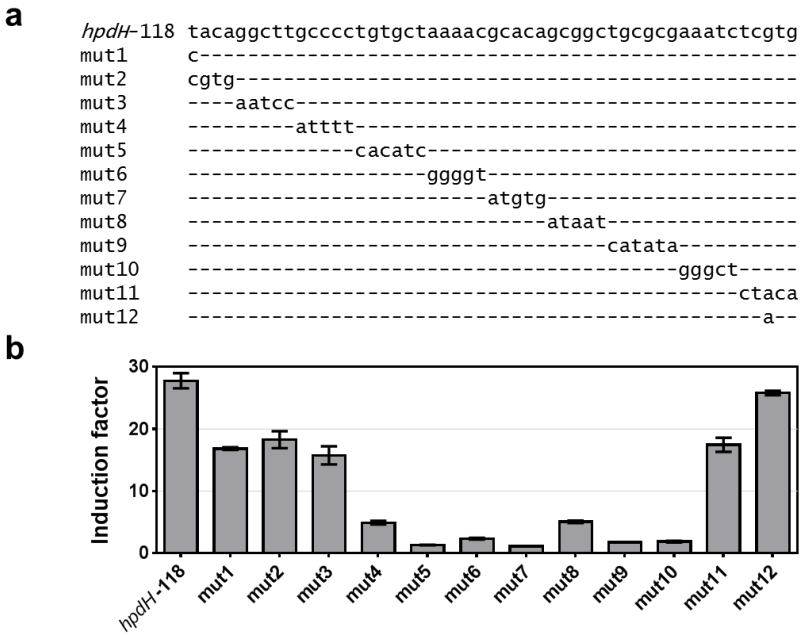


Figure 5.5 *P. putida* KT2440 *hpdH*-118::*rfp* promoter mutations. (a) Nucleotide sequences of the different *hpdH*-118::*rfp* promoter mutations that were evaluated for 3-HP-responsive activation of reporter gene expression. Unchanged nucleotides are represented as a dash. (b) Induction levels mediated by the different promoter mutations. Promoter-reporter gene constructs were evaluated in *E. coli* MG1655 for RFP synthesis in the absence and presence of 10 mM 3-HP. Error bars represent standard deviations of three biological replicates.

Finally, to confirm that HpdR binds directly to the *hpdH* promoter region, the TR was purified and an electrophoretic mobility shift assay (EMSA) was performed. A DNA shift can be observed in the absence of 3-HP (Figure 5.6), suggesting that regulator binding to the intergenic region occurs even under uninduced conditions. Addition of 3-HP to the binding reaction results in formation of a slower migrating complex 2. A similar observation was made for the *catBC* promoter, where a tighter and potentially higher-order ligand-regulator-DNA complex was proposed to facilitate activation of gene expression[233].

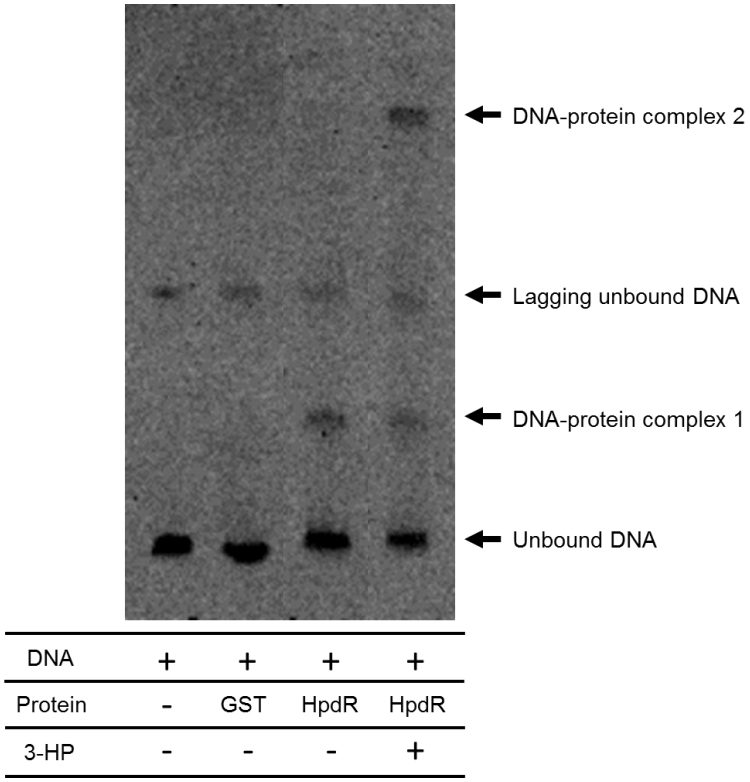


Figure 5.6 EMSA of the native *P. putida* KT2440 *hpdR*/*hpdH* intergenic region. Effect of purified TR and 3-HP on migration of labelled DNA. GST and HpdR refer to elution fractions A from pGEX-6P-1 and pEH089 cultures, respectively.

5.3.5 Determination of inducer-dependent orthogonality of the *Pp*HpdR/P$_{hpdH}$ inducible system in *E. coli* and *C. necator*

In addition to host-originating TRs, which may interfere with the heterologous inducible system, other metabolites than the primary inducer may be able to activate gene expression. These can be compounds that are involved in the cell's metabolism, components of the culture medium, or additional effectors that were added to control independent gene expression from other inducible promoters. Compounds that were investigated for cross-reactivity with the *Pp*HpdR/P$_{hpdH}$ inducible system include (Figure 5.7a): (i) the commonly used inducers L-arabinose and IPTG, (ii) fructose as an alternative carbon source, (iii) the 3-HP precursors glycerol (2), pyruvic acid (3), and β-alanine (4) (shown in blue)[234], and (iv) a broad range of compounds that are structurally similar to 3-HP such as mono- and dihydric alcohols (shown in red), mono- and dicarboxylic acids (shown in green), as well as α- and β-hydroxycarboxylic acids (shown in purple). Evaluation of these compounds may aid in the determination of structural features of the ligand that are required to interact with HpdR. Additionally, it may assist to identify a 3-HP analogue to be employed as inducer if 3-HP is metabolised as shown in *P. denitrificans*[212]. In order to achieve a sustained gene expression from the *hpdH* promoter, either genes encoding 3-HP metabolising enzymes need to be deleted or a metabolically inert analogue inducer should be employed.

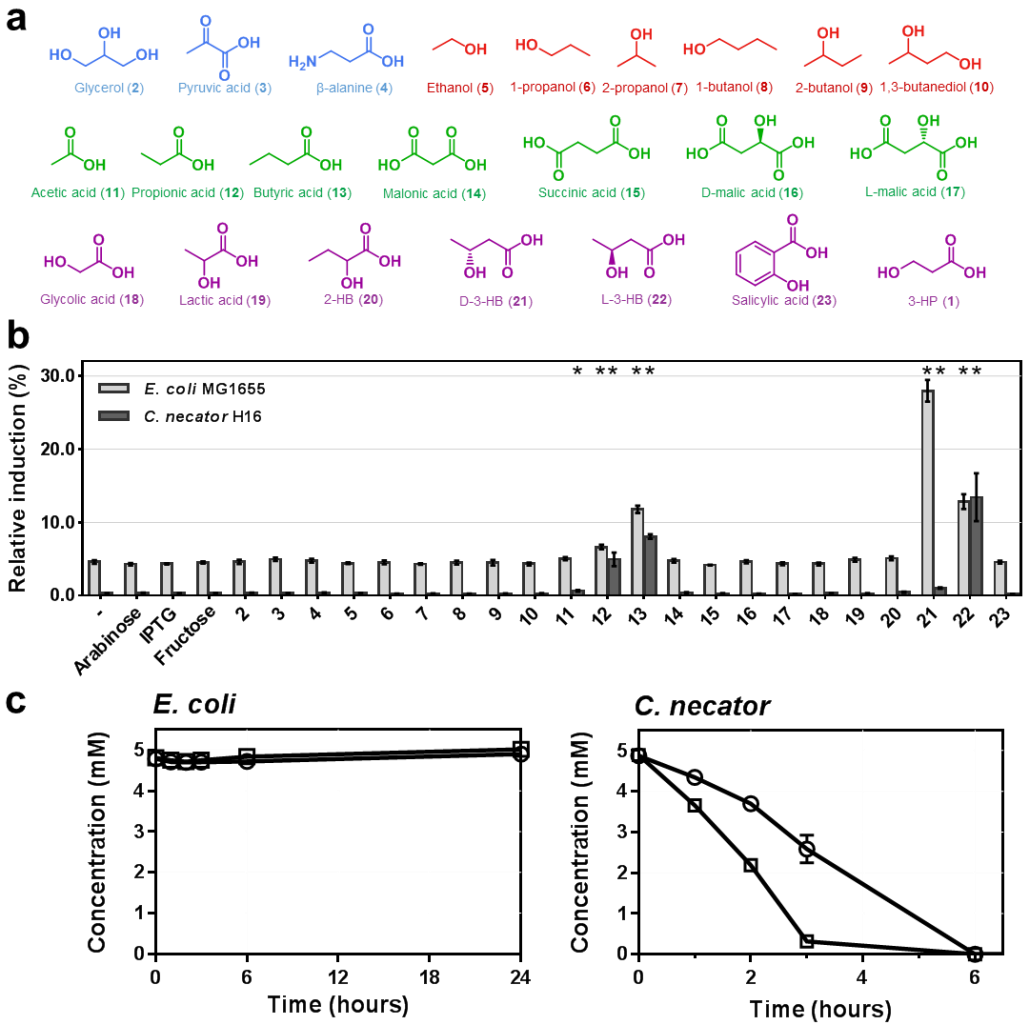


Figure 5.7 Determination of inducer-dependent orthogonality. (a) Chemical structures of the compounds that were investigated for cross-reactivity with the *Pp*HpdR/P_{hpdH} inducible system: 3-hydroxypropionic acid (1), glycerol (2), pyruvic acid (3), β-alanine (4), ethanol (5), 1-propanol (6), 2-propanol (7), 1-butanol (8), 2-butanol (9), 1,3-butanediol (10), acetic acid (11), propionic acid (12), butyric acid (13), malonic acid (14), succinic acid (15), D-malic acid (16), L-malic acid (17), glycolic acid (18), lactic acid (19), 2-hydroxybutyric acid (20), D-3-hydroxybutyric acid (21), L-3-hydroxybutyric acid (22), and salicylic acid (23). (b) Relative induction levels of the *Pp*HpdR/P_{hpdH} inducible system subjected to a variety of metabolites. The various compounds were tested in *E. coli* MG1655 (light grey) and *C. necator* H16 (dark grey) harbouring pEH008. (-) uninduced. Error bars represent standard deviations of three biological replicates. Asterisks indicate statistically significant induction values for $p < 0.01$ (unpaired *t* test). (c) Consumption of D- (compound 21, square) and L-3-HB (compound 22, circle) in *E. coli* MG1655 and *C. necator* H16. Error bars represent standard deviations of three biological replicates.

Single time-point fluorescence measurements were performed in *E. coli* MG1655 and *C. necator* H16 carrying the reporter plasmid with the native

*Pp*HpdR/P_{hpdH} inducible system (pEH008). The compounds were added at a final concentration of 5 mM, except for D,L-2-hydroxybutyrate (20) which, due to toxic effects, was used at 2.5 mM in *C. necator*. The response of the reporter system to the different molecules is illustrated as induction level relative to the one which was achieved by adding 5 mM 3-HP to the culture (Figure 5.7b). In addition to 3-HP, the monocarboxylic acids propionate (12) and butyrate (13), along with the structurally very similar D-3-hydroxybutyrate (D-3-HB, 21) and L-3-hydroxybutyrate (L-3-HB, 22) demonstrate statistically significant inductions of reporter gene expression in both *E. coli* and *C. necator*. The highest level of induction relative to 3-HP was achieved with the D-enantiomer of 3-HB in *E. coli*. It induced RFP synthesis almost to one third of the level that had been obtained with 3-HP. The observation that in *E. coli* the L-enantiomer of 3-HB only induces to half of the level of D-3-HB indicates a stereospecific preference of HpdR for the D-enantiomer.

Whereas the relative induction levels for propionate and butyrate are roughly the same in both microorganisms after RFP levels of the uninduced cultures had been subtracted, a similar correlation cannot be observed for the 3-HB enantiomers. The relative induction mediated by D-3-HB is 1% in *C. necator* as opposed to 28% in *E. coli*, whereas the L-enantiomer demonstrates a higher relative induction in *Cupriavidus*. This antagonistic behaviour in both investigated microorganisms was hypothesised to be caused by rapid metabolism of the natural D-enantiomer in *Cupriavidus*. It encodes two D-3-hydroxybutyrate dehydrogenases (E.C.1.1.1.30), which can convert D-3-HB into acetoacetate. *E. coli* MG1655 lacks these enzymes. To test this hypothesis, *C. necator* and *E. coli* were cultivated in the presence of D- or L-3-HB. Consumption of these compounds was monitored by HPLC analysis of cell-free supernatant samples.

As hypothesised, the concentration of D-3-HB in the supernatant of the *C. necator* culture decreased rapidly (Figure 5.7c). From the initial concentration of 5 mM D-3-HB, only 0.3±0.1 mM remained in the culture supernatant 3 h after it had been added. Surprisingly, the non-natural L-enantiomer was consumed as well. Its concentration does not decrease as quickly as D-3-HB, however, even L-3-HB is fully depleted 6 h after its supplementation. *E. coli* consumes neither of the 3-HB enantiomers. Varying consumption rates of D- and L-3-HB in *C. necator* may explain the discrepancy between the relative induction levels of both microorganisms. In *C. necator*, the rate of D-3-HB consumption appears to be higher than the rate of ligand-regulator-DNA complex formation, subsequently resulting in decreased gene transcription. Even though it is metabolised, the concentration of L-3-HB still remains high enough over the period of cultivation to mediate transient reporter gene expression.

5.3.6 System kinetics and dynamics

The *Pp*HpdR/P_{hpdH} inducible system was analysed for fluorescence output over time at different concentrations of 3-HP. This time course experiment was performed in both microorganisms and provides information about the time that is required to activate reporter gene expression, the influence of different inducer levels on growth, and the dynamic range[74, 158]. In both microorganisms, RFP synthesis above the level of the uninduced culture started 30 min after addition of 3-HP (Figure 5.8a). However, the profile of induction kinetics differs strongly in *E. coli* and *C. necator*. The variability can be explained by their growth kinetics and their dose response. In *E. coli*, growth is not affected by the 3-HP concentrations that were tested. In *C. necator*, however, the growth profile changes as a result of an increase in 3-HP concentration. The addition of 10 mM 3-HP has a growth-retarding effect at the beginning of

cultivation. The stationary growth phase is reached earlier in the presence of higher inducer concentrations, most likely due to an increased 3-HP metabolism as it has been reported for *P. denitrificans*[198] and other species that possess *mmsR-mmsA-hbdH* and *hpdR-hpdH* regulons[212]. The metabolism of inducer in *C. necator* in turn results in transient gene expression and only higher levels of 3-HP are able to maintain a fluorescence output over the time course of experiment. This behaviour is reflected in the dose-response curve of *C. necator*, which illustrates the correlation between inducer concentration and fluorescence output 4, 6, and 8 h after 3-HP addition (Figure 5.8b). Since 3-HP is not metabolised by *E. coli*, a sustained gene expression can be observed throughout the time course of experiment. As it can be seen in the dose-response curve, HpdR demonstrates a high induction cooperativity and gene expression can be tuned precisely in the range of 0.1-3 mM for a linear fluorescence output.

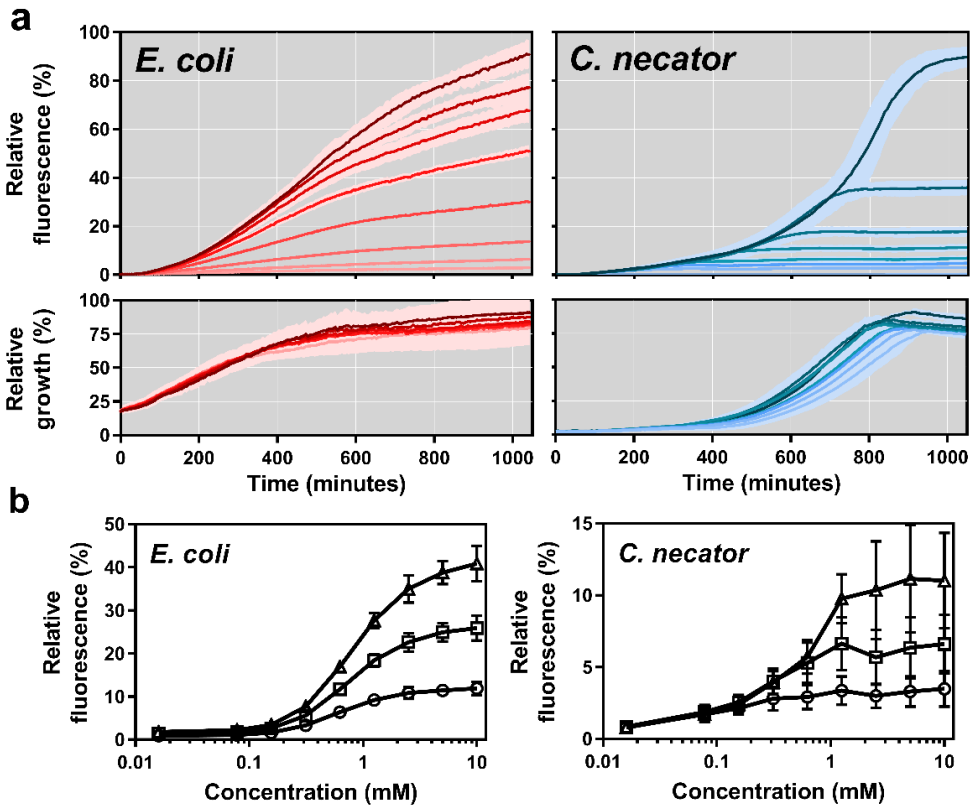


Figure 5.8 Induction dynamics and kinetics of the *Pp*HpdR/P_{hpdH} inducible system. (a) Relative fluorescence and absorbance curves of *E. coli* MG1655 and *C. necator* H16 cultures harbouring the native *Pp*HpdR/P_{hpdH} inducible system. 3-HP was added at time zero at the final concentrations of 10, 5, 2.5, 1.25, 0.625, 0.156, 0.016 mM and no inducer. The darker the shade of the colour, the higher the concentration of 3-HP. The standard errors from three biological replicates are illustrated as lighter shading above and below the induction kinetics curve. (b) Dose response curve of the *Pp*HpdR/P_{hpdH} inducible system in *E. coli* MG1655 and *C. necator* H16. It illustrates the relation between inducer concentration and fluorescence output 4 (circle), 6 (square), and 8 (triangle) h after 3-HP addition. Inducer concentrations range from 0, 0.016, 0.078, 0.156, 0.313, 0.625, 1.25, 2.5, 5, and 10 mM. Error bars represent standard deviations of three biological replicates.

5.4 Discussion

Biosustainable production of chemicals and fuels is becoming increasingly important in the global need to reduce air pollution and greenhouse gas emissions. 3-HP is an important platform chemical used as a precursor for the synthesis of acrylic acid and production of other value-added compounds. Recently, substantial metabolic

engineering efforts have been made to develop microbial cell factories and biosynthetic pathways for the biological production of 3-HP.

TR-based inducible systems are suitable for controlling metabolic pathways and screening microorganisms with improved production of target compounds. They have been shown to be indispensable in the iterative design–build–test cycles of synthetic biology and biotechnology, and can lead eventually to the development of improved and sustainable biosynthesis processes[217, 219]. In this chapter, potential 3-HP-inducible systems in *P. putida* KT2440 and *C. necator* H16 were identified, each encoding two regulons putatively involved in 3-HP metabolism. They were analysed for functionality in the well-characterised microorganism *E. coli* and the chemolithoautotrophic bacterium *C. necator* which is of industrial interest due to its ability to utilise CO_2 gas as sole carbon source.

The *mmsR-mmsA-hbdH* and *hpdR-hphH* regulons are widespread amongst proteobacteria and actinobacteria as shown previously[212]. In this chapter, it was established that neither the *PpmmsA* nor *PphpdH* promoter can be induced by the chromosomally encoded *Cn*HpdR in the absence of *P. putida* MmsR or HpdR homologues, respectively. This suggests that MmsR and HpdR homologues and/or their respective *mmsA* and *hpdH* promoter regions in *C. necator* (betaproteobacteria) and *P. putida* (gammaproteobacteria) have diverged to the extent that their corresponding species-specific LTTRs are required for transcription activation. These findings indicate that the *Pp*MmsR/P_{mmsA} and *Pp*HpdR/P_{hpdH} indcible systems can be utilised for TR-dependent orthogonal gene expression control in biotechnologically relevant microorganisms from two different bacterial classes, gammaproteobacterium *E. coli* and betaproteobacterium *C. necator*. It was observed that absolute reporter gene expression from both *P. putida* inducible systems is much higher in *C. necator* than in

E. coli, which results in overall higher induction levels in *Cupriavidus*. This behaviour may be due to a better expression of the *P. putida* transcriptional regulators in *C. necator*. The GC content of the TR genes (62% and 64% for *Pp*MmsR and *Pp*HpdR, respectively) is more similar to the median whole genome GC content of *C. necator* (66.3%) when compared to *E. coli* (50.6%). Codon-optimisation of the regulator genes for *E. coli* codon-usage may result in higher absolute reporter gene expression in this microorganism. Since *Pp*HpdR/P_{hpdH} showed the highest induction level of the analysed 3-HP-inducible systems (516.6-fold induction in *C. necator*) and is controlled independently from host-originating TRs in both analysed organisms, it was subjected to further characterisation.

Although in this chapter the aim was to uncover an orthogonal 3-HP-inducible system that can be utilised for synthetic biology and biotechnology applications, a significant attempt was made to characterise the molecular mechanisms involved in *Pp*P_{hpdH} activation and *Pp*HpdR-P_{hpdH} interaction. A thorough understanding of these mechanisms is crucial for genetic part design in order to build synthetic metabolic pathways for the production of added-value compounds such as 3-HP. The bioinformatically identified conserved 40-nucleotide sequence within P_{hpdH} and the *in vivo* analysis of various promoter mutations suggests, that two proposed TFBSs are potentially required for promoter region interaction with HpdR to mediate transcriptional activation. The length, location and the motif complexity of the TR binding site is common for almost all LysR-type inducible promoters characterised to date[164]. All *Pseudomonas* LTTRs have been clustered into 9 phylogenetic groups on the basis of full length and domain sequence analysis[215]. Interestingly, HpdR belongs to group VI, which includes only a few LysR-type proteins. None of the regulators or their potential target DNA-binding sites have been characterised to date. The novelty

of the motif identified in this study can be further highlighted by the observation that neither TFBS1 nor TFBS2 contains the typical LysR T-N_{11}-A binding motif. Furthermore, the two motifs T-N_{11}-A and TTA-$N_{7/8}$-GAA which were proposed by Zhou *et al.*[212] to be the HpdR regulatory binding site in *P. denitrificans* ATCC 13867 could neither be detected in the *P. putida hpdH* promoter upstream of the predicted -35 box, nor do these motifs seem to be conserved across the analysed *Pseudomonas* species. It should be noted that *in vitro* binding of HpdR to the native *hpdH* promoter region was confirmed by EMSA. Protein-DNA complex formation occurs even under uninduced conditions likely indicating that HpdR is involved in negative autoregulation of *hpdR* transcription. The mechanism of negative autoregulation is typical for LTTRs[30]. Interestingly, a stronger retardation effect of protein-DNA complex was observed in the presence of 3-HP. Ligand binding may facilitate formation of tighter and higher order protein-DNA complexes required for transcription activation as reported for *catBC* promoter[233].

In addition to characterisation of molecular mechanisms involved in promoter activation, the ligand-specificity of the system was determined as well as induction kinetics and dynamics. The analysis of structurally similar compounds revealed a set of ligand features that are necessary to interact with HpdR. The minimum requirement for the ligand to be functional is one carboxyl group, as in propionate (12) and butyrate (13). The presence of a hydroxyl group at the β-position enhances regulator activity. However, replacing the hydroxyl group by an amine- or a carboxyl group, changing the position of the hydroxyl group from β to α, or adding more functional groups, renders the ligand inactive. Cross-induction by these compounds appears to be caused by a lack of specificity of HpdR, rather than by their conversion into the primary inducer 3-HP. In this context, it was shown that the structurally similar 3-HB was a

metabolically inert analogue inducer in *E. coli*. However, due to their metabolism in *C. necator*, none of the compounds that were able to activate the *Pp*HpdR/P_{hpdH} inducible system can be applied as 3-HP analogue inducer for control of sustained gene expression. Metabolism and growth-retarding effects of the primary inducer were demonstrated to impact induction and growth kinetics in *C. necator*. Finally, the data suggests that the *Pp*HpdR/P_{hpdH} inducible system can be a useful genetic tool to: (i) build a 3-HP or 3-HB biosensor which reports intracellular metabolite concentrations by fluorescence output; (ii) implement directed-evolution strategies for high-throughput screening of strains with improved production titres, and; (iii) to balance enzyme levels in metabolic pathways that accumulate these compounds as final product or utilise them as intermediate compound.

5.5 Conclusion

3-HP-inducible promoters and their corresponding LTTRs were identified in *P. putida* and *C. necator*. The *Pp*HpdR/P_{hpdH}-inducible system is highly inducible by 3-HP in *E. coli* and *C. necator* (23- and 517-fold, respectively). Bioinformatics and mutagenesis analysis revealed a conserved 40-nucleotide sequence in the *hpdH* promoter, which plays a key role in HpdR-mediated transcription activation. The kinetics and dynamics of the 3-HP-inducible system were investigated. Moreover, it was shown that this system is also induced by both enantiomers of 3-HB.

Chapter 6

The following chapter is mainly based on the work presented in the publication:

A transcription factor-based biosensor for detection of itaconic acid
E. K. R. Hanko, N. P. Minton & N. Malys (2018)
ACS Synthetic Biology ***7*** (5), 1436-1446.[235]

6 A transcription factor-based biosensor for detection of itaconic acid

6.1 Introduction

The use of biological processes for the production of chemicals and fuels is a promising alternative to the traditional approach of chemical manufacture[191]. They offer the opportunity to convert renewable or waste feedstocks into higher value compounds of industrial interest[190]. Although many biological processes have the potential to replace synthetic chemistry, product titres and productivity often remain to be optimised in order to achieve economically competitive conversion rates[4, 190, 191]. To facilitate and expedite the implementation of biocatalysts with improved performance, low-cost and high-throughput microbial engineering strategies need to be developed.

Itaconic acid is an attractive platform chemical with a wide range of industrial applications, such as in rubber, detergents, or surface active agents[236]. In 2004, it was reported by the US Department of Energy to be one of the top twelve building block chemicals from biomass[237]. The C5-dicarboxylic acid can be converted into poly(acrylamide-*co*-itaconic acid) which is used as a superabsorbent for aqueous solutions, or poly(methyl methacrylate), also known as Plexiglas[193].

Itaconate is a naturally occurring metabolite formed by decarboxylation of aconitate, an intermediate of the citric acid cycle. A number of microorganisms, including *Aspergillus terreus*[238], *Ustilago maydis* (also known as *U. zeae*)[239], and *Candida sp.*[240], have been described as natural producers of itaconic acid. It is also produced as an antimicrobial compound by macrophages, mammalian immune cells[241, 242]. In *A. terreus* and macrophages, itaconate is synthesized from the tricarboxylic acid cycle intermediate *cis*-aconitate through the action of a *cis*-aconitate

decarboxylase (CadA). In contrast, in *U. maydis* it is produced via the unusual intermediate *trans*-aconitate[243]. Heterologous expression of the *A. terreus cadA* gene has demonstrated that the biosynthesis of itaconic acid can be achieved in different host organisms than the natural producer[244]. So far, the highest titre of biotechnologically produced itaconate has been obtained by fermentation of *A. terreus*[245-247]. However, due to feedback inhibition of itaconate biosynthesis at higher concentrations[248], considerable research efforts have been directed towards developing alternative microbial biocatalysts. Other microorganisms that have been investigated for the biosynthesis of itaconic acid include *Pseudozyma antarctica*, *Corynebacterium glutamicum*, *Escherichia coli*, *Saccharomyces cerevisiae*, *Yarrowia lipolytica* and species of *Candida* and *Ustilago*[240, 249-255]. Although some of these microorganisms exhibit beneficial traits, such as a high tolerance to itaconate and a low pH[250, 253], production titres need to be considerably improved.

Genetically encoded biosensors have gained increasing interest as molecular tools enabling high-throughput strain development[62]. They are composed of transcription factor-based inducible gene expression systems linked to a reporter or an antibiotic resistance gene[74, 256]. By using a fluorescent reporter gene, changes in intracellular metabolite concentrations can easily be monitored by a fluorescence output enabling the screen of millions of single-cells in a rapid manner[62]. Biosensors have been successfully applied to increase products titres of platform chemicals such as acrylate, 3-hydroxypropionate (3-HP) and glucarate[74, 217]. To date, no itaconate biosensor has been developed which could facilitate the screening process for both metabolically engineered strains and alternative feedstocks, such as biomass hydrolysates, to improve yields and decrease production costs[257].

The work described in this chapter was aimed to identify an itaconate-inducible gene expression system and to construct a fluorescence-based biosensor. Several natural compounds were screened for biosensor induction and induction kinetics measured. Moreover, the developed biosensor was exploited in the optimisation of itaconate production in *E. coli* and its output compared to analytically determined itaconate titres.

6.2 Materials and methods specific to this chapter

6.2.1 Chemicals

Itaconic acid, sodium succinate dibasic hexahydrate, D-malic acid, L-malic acid, sodium fumarate dibasic, oxalacetic acid, L-aspartic acid potassium salt, methylsuccinic acid, masaconic acid, citraconic acid, α-ketoglutaric acid, L-glutamic acid monosodium salt hydrate, sodium acetate, propionic acid, butyric acid, 3-butenoic acid, valeric acid, magnesium acrylate, methacrylic acid, tiglic acid, citric acid trisodium salt, *cis*-aconitic acid, *trans*-aconitic acid, tricarballylic acid, and D,L-Isocitric acid trisodium salt hydrate were used as inducers for assaying the itaconate-inducible system. All chemicals are listed in Supplementary Table 1 in the *Appendix*.

6.2.2 Plasmids

Key features of all plasmids used and generated in this chapter are summarised in Table 6.1. A detailed description of how each plasmid was assembled is provided in the *Appendix*. The nucleotide sequences of plasmids pEH086 and pEH177 have been deposited in the public version of the ACS registry (https://acs-registry.jbei.org) under the accession numbers ACS_000716 and ACS_000717, respectively.

Table 6.1 Plasmids used in chapter 6.

Plasmid	Characteristic	Reference or source
pBBR1MCS-2-PphaC-eyfp-c1	Kan^r; P_{phaC}-*eyfp*	[131]
pEH006	Cm^r; P_{araC}-*araC*-T_{rrnB1}; P_{araBAD}-T7sl-*rfp*-T_{dbl}	This work
pEH006E	Cm^r; T_{rrnB1}-*rfp*-T_{dbl}	This work
pEH086	Cm^r; *Yp*P_{itcR}-*itcR*-T_{rrnB1}; *Yp*P_{ccl}-*rfp*-T_{dbl}	This work
pEH164	Cm^r; P_{araC}-*araC*-T_{dbl}; P_{araBAD}-T7sl-T_{rrnB2}; *Yp*P_{itcR}-*itcR*-T_{rrnB1}; *Yp*P_{ccl}-*rfp*-T_{dbl}	This work
pEH165	Cm^r; P_{araC}-*araC*-T_{dbl}; P_{araBAD}-T7sl-*cadA*-T_{rrnB2}; *Yp*P_{itcR}-*itcR*-T_{rrnB1}; *Yp*P_{ccl}-*rfp*-T_{dbl}	This work
pEH172	Cm^r; *Yp*P_{ccl}-*rfp*-T_{dbl}	This work
pEH177	Cm^r; *Pa*P_{itcR}-*itcR*-T_{rrnB1}; *Pa*P_{ich}-*rfp*-T_{dbl}	This work
pEH178	Cm^r; *Pa*P_{ich}-*rfp*-T_{dbl}	This work

6.2.3 Production of itaconate

Real-time biosynthesis of itaconate was monitored quantitatively by HPLC-UV analysis and by fluorescence output in *E. coli* TOP10 harbouring pEH165. Single colonies of freshly transformed cells were used to inoculate 5 mL of LB medium. The preculture was incubated for 18 h with orbital shaking at 37 °C and 200 rpm. Subsequently, it was diluted 1:100 in 6 mL of fresh LB medium. The main cultures were grown in 50-mL conical centrifuge tubes with orbital shaking at 30 °C and 225 rpm. At an OD_{600} of 0.5, 50 µL of L-arabinose stock solutions were added to achieve the final concentrations of 5, 10, 25, 50, 100, 250 and 1000 µM. One sample per biological replicate remained uninduced. Samples of 0.5 mL were taken immediately, 6, 9, 12, 18, 24, and 48 h after inducer supplementation. They were directly used for evaluation by flow cytometry, OD_{600}, and fluorescence measurement. The remaining

sample was centrifuged for 5 min at 16,000*g*, and the cell-free supernatant was subjected to HPLC-UV analysis.

6.2.4 Metabolite extraction

To determine intracellular itaconate concentrations when added extracellularly or synthesised by the cells, cultures of *E. coli* TOP10 harbouring pEH164 or pEH165 were grown overnight to saturation and diluted 1:100 in 200 mL of LB medium. The main cultures were grown in 1-L non-baffled shake flasks with orbital shaking at 30 °C and 225 rpm. At an OD_{600} of 0.5, inducers were added at final concentrations of 2.5 mM itaconate or 100 µM L-arabinose to cultures of *E. coli* TOP10 harbouring pEH164 (control plasmid) or pEH165 (plasmid harbouring *cadA* gene), respectively. Samples of cells carrying pEH164 were taken 0, 6 and 12 h after addition of itaconate. Samples of cells carrying pEH165 were taken 0, 18 and 36 h after addition of L-arabinose. Each time, the culture volume corresponding to an OD_{600} of 50 was centrifuged for 10 min at 16,000*g*. The supernatant was removed and stored at -80 °C for HPLC-UV analysis. Subsequently, the cell pellet was washed once in 1 mL of PBS, transferred to a microcentrifuge tube and centrifuged as before. The supernatant was completely removed, the pellet was weighed using microbalance and frozen overnight at -80 °C.

The extraction of intracellular metabolites including itaconate was performed by modification of a previously described method[258]. Briefly, 250 µL of -40 °C cold methanol-water solution (60% v/v) was added to the wet cell pellet with the volume of 50-70 µL. Subsequently, the sample was mixed vigorously using vortex until it was completely resuspended. The cell suspension was frozen at -80 °C for 30 min, thawed on ice and vortexed vigorously for 1 min. This step was repeated three times before the sample was centrifuged at -10 °C and 26,000*g* for 20 min. The supernatant was collected and kept at -80°C. To the pellet, another 250 µL of -40 °C cold

methanol-water solution (60% v/v) was added. The cells were resuspended completely using vortex, three freeze-thaw cycles performed as above and centrifuged as before. The supernatant was pooled with the first collection and stored at -80 °C until subjected to HPLC-UV analysis.

6.2.5 Calculation of intracellular itaconate concentration

The total cell volume (V_{pellet}) in the sample was calculated by dividing the weight of wet cell pellet by the cell density of 1.105 g/mL[259]. Together with the volume of extraction solvent added to the sample, V_{pellet} was used to calculate the dilution factor required to determine the intracellular molar concentration of itaconate. Subsequently, the intracellular itaconate concentration in the cell culture ($C_{intracellular/CC}$) was calculated using equation (6.1):

$$C_{\text{intracellular/CC}} = \frac{\text{FW}_{\text{itaconic acid}} \cdot C_{\text{molar}} \cdot V_{\text{pellet}}}{V_{\text{culture}}} \qquad (6.1)$$

The remaining parameters correspond to the formula weight of itaconic acid ($FW_{itaconic\ acid}$), the intracellular molar concentration of itaconate determined by HPLC-UV analysis (C_{molar}) and the culture volume sampled ($V_{culture}$).

6.2.6 Calculation of half-maximal *rfp* expression

Because of toxicity at higher levels, the concentrations of mesaconate, *cis*-, and *trans*-aconitate, which mediate half-maximal *rfp* expression (K_i), were predicted using a phenomenological model as described previously[186]. The model describes the change in dynamic range of an inducible system as a function of inducer concentration. It assumes that: (i) the maximum dynamic range of a biosensor (μ_{max}) remains constant as long as the genetic context does not change, and (ii) K_i is dependent on metabolite-TR affinity.

The dynamic range μ for each concentration of itaconate was calculated using the absolute normalised fluorescence values from the time course experiment 6 h after itaconate addition. After subtraction of the basal output, the resulting dynamic range was fit to the corresponding inducer concentration using the Hill function (6.2):

$$\mu(I) = \mu_{\max} \cdot \frac{I^h}{K_i{}^h \cdot I^h} \tag{6.2}$$

The remaining parameters correspond to concentration of inducer (*I*), and the Hill coefficient (*h*). Subsequently, the itaconate $\mu_{\max}$ was used as fixed parameter to calculate K_i for mesaconate, *cis*-, and *trans*-aconitate employing the same Hill function. Calculations were performed using software Prism GraphPad version 7.03.

6.3 Results

6.3.1 Identification of an itaconic acid-inducible system

To build an itaconate biosensor, which can be applied across different species, both elements of a transcription factor-based inducible system, a TR and the corresponding inducible promoter, are needed. Bacterial degradation pathways, which are often activated exclusively in the presence of the compound to be degraded, represent a rich source of inducible promoters. Even though the pathway for itaconate catabolism had been known for more than 50 years[260], and a few bacteria including *Pseudomonas* spp., *Salmonella* spp., and *Micrococcus* sp. have been shown to possess enzymatic activities for itaconate degradation[261], the genes encoding these enzymes have only recently been identified in *Yersinia pestis* and *Pseudomonas aeruginosa*[262]. The pathway comprises three enzymatic reactions (Figure 6.1a). The first reaction is catalysed by itaconate CoA transferase (Ict) which converts itaconate to itaconyl-CoA. The CoA ester is subsequently hydrated to (*S*)-citramalyl-CoA by itaconyl-CoA hydratase (Ich) which is then cleaved into acetyl-CoA and pyruvate by

(*S*)-citramalyl-CoA lyase (Ccl). The production of the Ict and Ich homologues (RipA and RipB, respectively) by *Salmonella enterica* was shown to be strongly induced after macrophage infection[263]. The upregulation of *ripA* and *ripB* was suggested by Sasikaran and co-workers to result from macrophagic itaconate secretion as part of the defence mechanism against pathogenic bacteria[262, 264]. Most likely, the promoters of the gene clusters encoding the enzymes for itaconate catabolism in *Y. pestis* and *P. aeruginosa* harbour regulatory elements required for transcription of these genes in the presence of itaconate. Interestingly, a gene encoding a LysR-type transcriptional regulator (LTTR, here termed ItcR) is located in opposite direction of both the *Y. pestis ccl-ich-ict* operon (also referred to as *ripABC* operon) and the *P. aeruginosa* putative six-gene operon encoding Ich, Ict, Ccl, and three other proteins (Figure 6.1b). The genes encoding LTTRs are often transcribed in divergent orientation with respect to the cluster of genes they regulate[164], which led to the hypothesis that transcription of the *Y. pestis* and *P. aeruginosa* itaconate degradation pathway genes is mediated by their corresponding divergently oriented LTTR genes from an inducible promoter located in their intergenic regions.

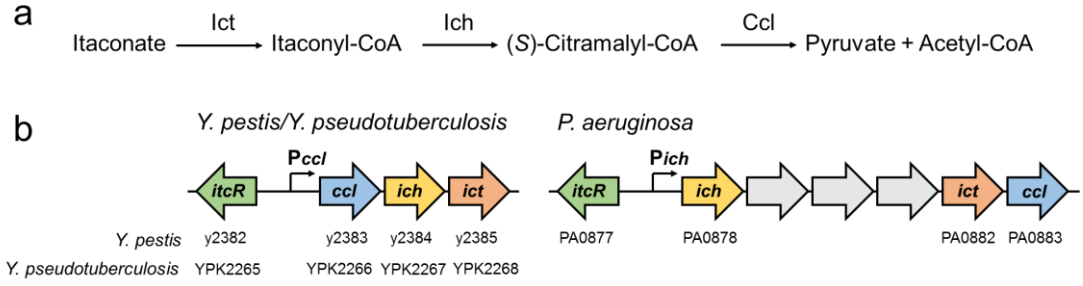


Figure 6.1 Bacterial itaconate degradation pathway. (a) The enzymes involved in bacterial itaconate degradation include itaconate CoA transferase (Ict), itaconyl-CoA hydratase (Ich), (*S*)-citramalyl-CoA lyase (Ccl). (b) The gene clusters in *Y. pestis*, *Y. pseudotuberculosis*, and *P. aeruginosa* encoding the enzymes required for itaconate catabolism. Divergently oriented LTTR genes (*itcR*) and putative itaconate-inducible promoters are depicted. Gene names and locus tags are shown under the schematic illustration of each gene cluster.

6.3.2 Itaconic acid-inducible gene expression is mediated by a LysR-type transcriptional regulator

To test the hypothesis that the itaconate degradation pathway is controlled by the transcriptional regulator and corresponding inducible promoter, both the *P. aeruginosa* PAO1 and the *Yersinia pseudotuberculosis* YPIII DNA fragments with the putative itaconate-inducible system were cloned, containing an intergenic region with promoters P_{ich} and P_{ccl}, respectively, and gene of the transcriptional regulator (*itcR*) (Figure 6.1b), into the reporter plasmid pEH006. The nucleotide sequence of the *Y. pseudotuberculosis* itaconate-inducible system is identical to the *Y. pestis* one, except for three single nucleotide polymorphisms in the *itcR* gene (YPK_2265) resulting in one amino acid difference. The nucleotide sequences of the intergenic regions containing putative itaconate-inducible promoters are provided in Supplementary Figure 6. To investigate the potential applicability of the two putative itaconate-inducible systems across different species, RFP reporter gene expression in response to itaconate was measured by fluorescence output in the model γ-proteobacterium *E. coli* MG1655 and the β-proteobacterium *C. necator* H16. Single time-point fluorescence measurements of *E. coli* and *C. necator* harbouring the putative itaconate-inducible systems, composed of transcriptional regulator and inducible promoter (ItcR/P), were performed in the absence and presence of itaconate (Figure 6.2). In both microorganisms, reporter gene expression from the *Y. pseudotuberculosis* (*Yp*) inducible system (pEH086) is induced significantly ($p < 0.01$) 6 h after supplementation with 5 mM itaconate (215-fold in *E. coli* and 105-fold in *C. necator*, Figure 6.2a,b, respectively). In contrast, the *P. aeruginosa* (*Pa*) inducible system *Pa*ItcR/P_{ich} (pEH177) does not mediate reporter gene expression in response to itaconate in *E. coli*, whereas in *C. necator* it demonstrates an 18.5-fold induction. In comparison, in *E. coli* MG1655, the level of induction mediated by the

Y. pseudotuberculosis itaconate-inducible system is considerably higher than the commonly used L-arabinose-inducible system which is subject to catabolic repression. A culture of *E. coli* MG1655 harbouring pEH006 demonstrated a 39-fold increase in RFP synthesis 6 h after addition of L-arabinose to a final concentration of 0.1% (w/v) in minimal medium.

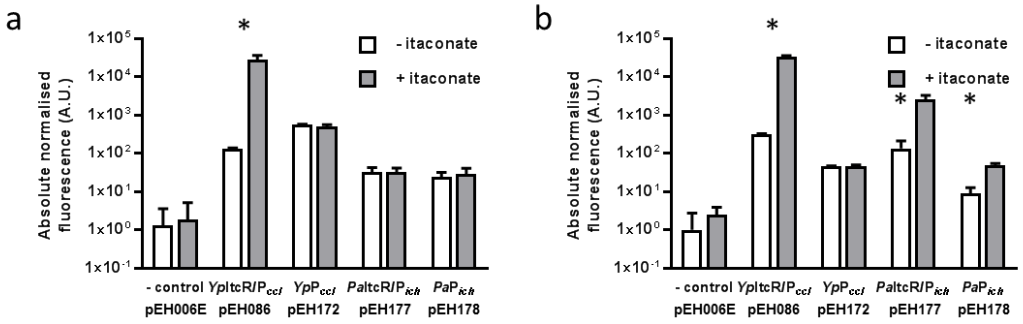


Figure 6.2 Itaconate-inducible gene expression and influence of ItcR. Absolute normalised fluorescence (arbitrary units) of (a) *E. coli* MG1655 and (b) *C. necator* H16 harbouring the *Y. pseudotuberculosis* (*Yp*) and *P. aeruginosa* (*Pa*) itaconate-inducible systems composed of promoter and transcriptional regulator (ItcR/P), and promoter only (P) implementation in the absence and presence of 5 mM itaconate. Single time-point fluorescence measurements were taken 6 h after inducer addition. The promoterless reporter construct pEH006E was employed as negative control. Error bars represent standard deviations of three biological replicates. Asterisks indicate statistically significant induction values for $p < 0.01$ (unpaired *t* test).

To confirm that itaconate-inducible reporter gene expression is indeed controlled by the episomally encoded ItcR, their genes were removed from the vectors containing *Yp*ItcR/P_{ccl} and *Pa*ItcR/P_{ich}. Single time-point fluorescence measurements were repeated for *E. coli* and *C. necator* solely harbouring the itaconate-inducible promoters in the absence and presence of itaconate (Figure 6.2). Without *Yp*ItcR, induction of reporter gene expression from the *Y. pseudotuberculosis* itaconate-inducible promoter (*Yp*P_{ccl}, pEH172) is abolished in both microorganisms. This confirms that transcription of the *Y. pseudotuberculosis* itaconate degradation pathway genes is mediated by their divergently oriented *itcR* gene and that neither of the two tested microorganisms encodes cross-reacting TR homologues. In *E. coli*, the level of

normalised fluorescence from *Pa*P$_{ich}$ (pEH178) and *Pa*ItcR/P$_{ich}$ (pEH177) is higher than the negative control, indicating that the promoter itself is active. However, the normalised fluorescence levels are of equal height, suggesting that the TR might not be produced or able to interact with its cognate operator sequence to activate gene expression in the presence of the effector. Interestingly, in *C. necator*, even though the gene of *Pa*ItcR was removed from the plasmid, reporter gene expression from *Pa*P$_{ich}$ (pEH178) is induced significantly ($p < 0.01$) after addition of itaconate indicating the presence of a cross-reacting *Pa*ItcR homologue encoded in the genome of *C. necator*.

6.3.3 Sensor characterisation

Because of its functionality in both tested microorganisms, regulator-dependent orthogonality and high level of induction, the itaconate-inducible system from *Y. pseudotuberculosis* was selected to be further characterised. The sensor was evaluated for its kinetics – the time that is required for the system to respond to a change in itaconate levels; dynamics – the range of inducer concentration that mediates a linear fluorescence output, and; inducer-dependent orthogonality – the specificity towards itaconate.

E. coli MG1655 was transformed with the plasmid harbouring the *Yp*ItcR/P$_{ccl}$-inducible system (pEH086), cultivated in M9 minimal medium, and fluorescence output was monitored over time after supplementation with different concentrations of itaconate. As can be seen from the fluorescence curve of induction kinetics, reporter gene expression is activated immediately after inducer addition, taking into account the time that is required for RFP maturation[134] (Figure 6.3a). This immediate response suggests, that the system is solely controlled by ItcR and that it is not affected by host-originating TRs. Furthermore, it suggests that itaconate is a primary inducing molecule, which starts instantly to be uptaken by or diffused into the

E. coli cells in minimal medium. It should be noted that the growth was similar for all itaconate concentrations tested.

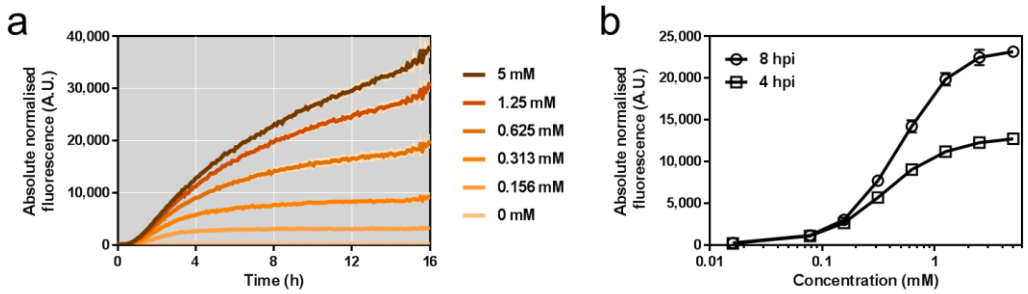


Figure 6.3 Kinetics and dynamics of the *Yp*ItcR/P_{ccl} inducible system. (a) Absolute normalised fluorescence of *E. coli* MG1655 harbouring the *Yp*ItcR/P_{ccl}-inducible system (pEH086) in response to different concentrations of itaconate added at time zero. The standard deviation of three biological replicates is shown as lighter colour ribbon displayed lengthwise of the induction kinetics curve. For the lower itaconate concentrations, the standard deviation is too small to be visible. (b) Dose-response curve of the *Yp*ItcR/P_{ccl}-inducible system in *E. coli* MG1655, illustrating the correlation between inducer concentration and fluorescence output 4 and 8 h post induction (hpi) with itaconate. Error bars represent standard deviations of three biological replicates.

The correlation between extracellular inducer concentration and fluorescence output, 4 and 8 h after itaconate supplementation, is illustrated in the dose response curve (Figure 6.3b). It indicates that gene expression can be tuned in the range of approximately 0.07 to 0.7 mM for a linear fluorescence output. The minimum concentration of exogenously added itaconate required for activation of the system is approximately 0.016 mM. A saturation of the *Yp*ItcR/P_{ccl}-inducible system on the other hand can be observed for itaconate levels above 2.5 mM. Notably, the concentration of exogenously added itaconate required to induce the system in *E. coli* MG1655 is lower in LB medium than in M9 minimal medium. At a time point 4 h after the addition of 0.016 mM itaconate, reporter gene expression is induced 7.7-fold in LB medium compared to a culture without itaconate (Figure 6.4). This is in contrast to a 1.4-fold induction in M9 minimal medium. The dose-response curve indicates that the itaconate concentration, required for a linear fluorescence output in LB medium,

ranges between approximately 0.016 and 0.16 mM (Figure 6.4). Despite a 5-fold reduced induction threshold for itaconate, the linear output range of the *Yp*ItcR/P_{ccl}-inducible system spans one order of magnitude, similar to what is observed in M9 minimal medium. This suggests that different growth conditions can contribute to the variation of both lower and upper induction thresholds, whereas the magnitude of system response is likely to remain constant.

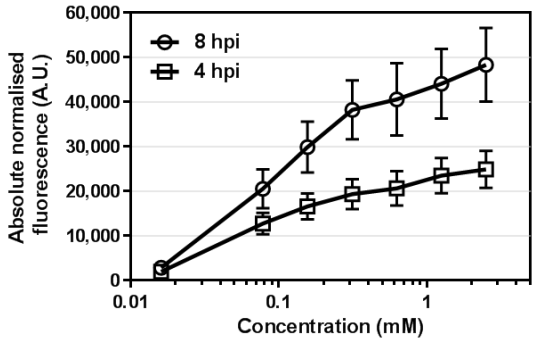


Figure 6.4 Dose-response of the *Yp*ItcR/P_{ccl}-inducible system in LB medium. Dose-response curve of the *Yp*ItcR/P_{ccl}-inducible system in *E. coli* MG1655 grown in LB medium. The graph illustrates the correlation between inducer concentration and fluorescence output 4 and 8 h post induction (hpi) with itaconate. Error bars represent standard deviations of three biological replicates.

In addition, the analysis of extracellular and intracellular itaconate by using HPLC-UV spectroscopy shows no significant change in the itaconate concentration during the 12-hour period in the actively growing *E. coli* culture (Table 6.2). This demonstrates that itaconate is not metabolised and therefore is a primary inducing molecule. Moreover, the analysis confirms that itaconate is taken up by or diffuses into the *E. coli* cell and reaches a relatively high concentration of at least 1.3 mM after 6 h. It should be noted that the actual intracellular molar concentration could be even higher, since the approximation is based on the assumption that the intracellular cell volume is equal to the total cell volume including the space occupied by cell membranes, lipids, etc. Interestingly, the intracellular itaconate concentration

becomes reduced when *E. coli* cells reach the stationary phase (12-h time point, Table 6.2); however, the total itaconate concentration in the culture remains unchanged.

Table 6.2 Extracellularly added and intracellularly produced itaconate and its distribution between the supernatant and cells of an *E. coli* culture grown in LB medium.

Itaconate extracellularly added					
	Molar concentration (mM)[a]		Concentration in cell culture (mg/L)		
Time (h)	Extracellular	Intracellular	Resulting from supernatant	Resulting from cells	Total
0	2.5[b]	nd[c]	325.253	nd	325.253
6	2.454 ± 0.050	1.309 ± 0.132	319.242 ± 6.437	0.685 ± 0.067	319.927 ± 6.437
12	2.462 ± 0.059	0.551 ± 0.058	320.243 ± 7.715	0.411 ± 0.074	320.654 ± 7.715
Itaconate intracellularly produced					
	Molar concentration (mM)		Normalised concentration in cell culture (mg/L/OD)		
Time (h)	Extracellular	Intracellular	Resulting from supernatant (% of total)	Resulting from cells (% of total)	Total
	nd	nd	nd	nd	nd
18	0.071 ± 0.021	0.145 ± 0.008	2.149 ± 0.638 (98.67)	0.029 ± 0.003 (1.33)	2.178
36	0.242 ± 0.117	0.241 ± 0.181	7.010 ± 2.506 (99.38)	0.044 ± 0.028 (0.62)	7.054

[a]Arithmetic mean ± standard deviation is derived using data of three biological replicates. [b]Itaconate concentration added to the cell culture at 0 h time point. [c]Not detected.

6.3.4 Sensor specificity

The *Yp*ItcR/P_{ccl}-inducible system was analysed for cross-reactivity by metabolites that may activate reporter gene expression in the absence of the primary inducing molecule itaconate. These can be exogenously added compounds or intermediates naturally involved in cellular metabolism. Compounds that were investigated for cross-reactivity mainly include citric acid cycle intermediates and structurally similar variants thereof (Figure 6.5a). Evaluation of these molecules may shed light on structural features required for TR-binding and TR affinity toward itaconate. Furthermore, screening potential candidate compounds might expand the list of metabolites to be detected by TR-based controllable systems and offer the possibility to be utilised as analogue inducers to control gene expression.

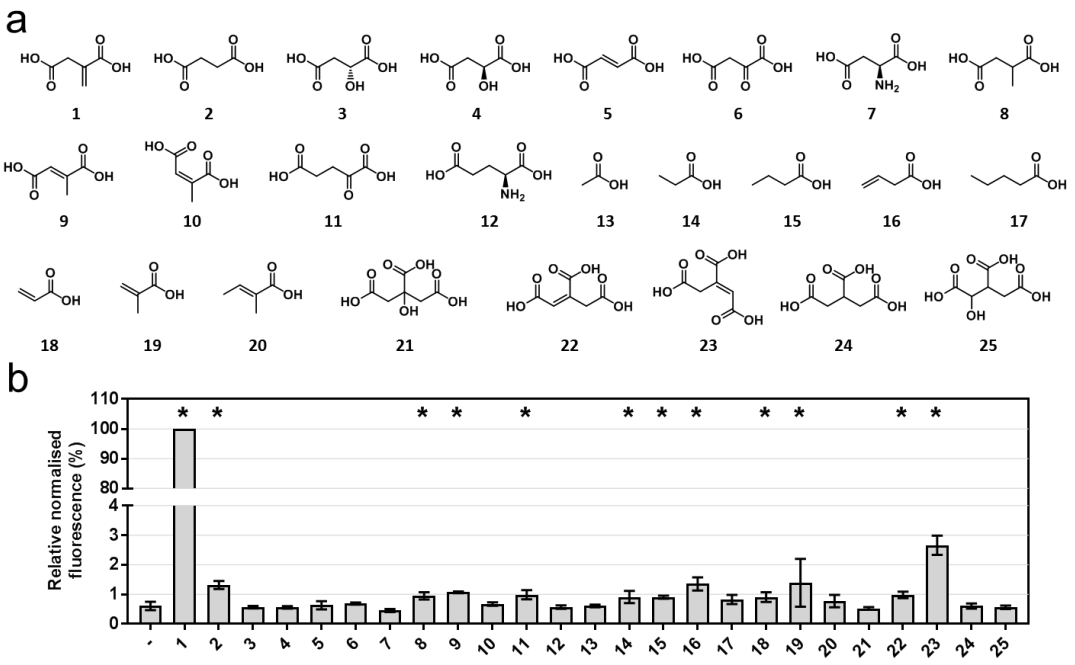


Figure 6.5 Inducer-dependent orthogonality of the *Yp*ItcR/P_{ccl}-inducible system. (a) Compounds that were investigated for cross-reactivity with the *Yp*ItcR/P_{ccl}-inducible system: itaconic acid (1), succinic acid (2), D-malic acid (3), L-malic acid (4), fumaric acid (5), oxaloacetic acid (6), L-aspartic acid (7), methylsuccinic acid (8), mesaconic acid (9), citraconic acid (10), α-ketoglutaric acid (11), L-glutamic acid (12), acetic acid (13), propionic acid (14), butyric acid (15), 3-butenoic acid (16), valeric acid (17), acrylic acid (18), methacrylic acid (19), tiglic acid (20), citric acid (21), *cis*-aconitic acid (22), *trans*-aconitic acid (23), tricarballylic acid (24), isocitric acid (25). (b) Normalised fluorescence (in %) of *E. coli* MG1655 harbouring the *Yp*ItcR/P_{ccl}-inducible system 12 h after addition of different compounds at a final concentration of 5 mM, relative to the fluorescence output obtained by adding 5 mM itaconate. (-), uninduced sample. Error bars represent standard deviations of three biological replicates. Asterisks indicate statistically significant induction values for $p < 0.01$ (unpaired *t* test).

The fluorescence output from cultures of *E. coli* MG1655 harbouring the *Yp*ItcR/P_{ccl}-inducible system, and cultivated in M9 minimal medium, was monitored over time after individual addition of each compound at a final concentration of 5 or 10 mM. Normalised fluorescence levels (in %), relative to the output obtained by adding 5 mM itaconate, were determined 12 h after compound supplementation. In addition to the primary inducer itaconate and under the assumption that all tested metabolites are able to enter the cell, the compounds succinate (2), methylsuccinate (8), mesaconate (9), α-ketoglutarate (11), propanoate (14), butanoate (15), 3-butenoate

(16), acrylate (18), methacrylate (19), *cis*-aconitate (22), and *trans*-aconitate (23) induce reporter gene expression at a final concentration of 5 mM with high statistical significance ($p < 0.01$) (Figure 6.5b). Of these eleven compounds, succinate, mesaconate, propanoate, butanoate, 3-butenoate, *cis*-aconitate, and *trans*-aconitate demonstrated a significant increase in RPF synthesis at a final concentration of 10 mM (Figure 6.6). Increased activation of reporter gene expression suggests that these inducers may exhibit a weak binding to ItcR inducing the system to some extent. The highest level of cross-reactivity is mediated by *trans*-aconitate. At a concentration of 10 mM, it reached 9.9% of the absolute normalised fluorescence that was achieved by using 5 mM itaconate. Since *E. coli* has not been reported to encode a *trans*-aconitate decarboxylase, converting *trans*-aconitate into itaconate, induction of reporter gene expression from *Yp*P_{ccl} is more likely to be caused by ItcR promiscuity rather than by decarboxylation of *trans*-aconitate forming itaconate.

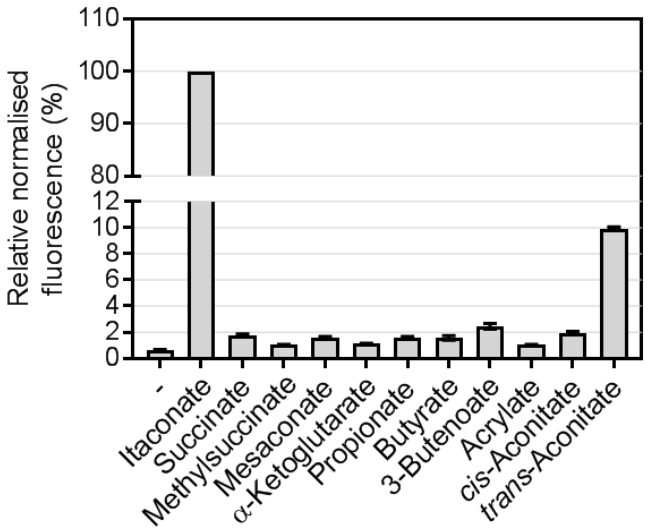


Figure 6.6 Specificity of the *Yp*ItcR/P_{ccl}-inducible system at effector concentrations of 10 mM. Normalised fluorescence (in %) of *E. coli* MG1655 harbouring the *Yp*ItcR/P_{ccl}-inducible system 12 h after addition of different compounds at final concentrations of 10 mM, relative to the fluorescence output obtained by adding 5 mM itaconate (second bar). (-), uninduced sample. Error bars represent standard deviations of three biological replicates.

cis-Aconitate and *trans*-aconitate showed more than a 2-fold change in induction level when the inducer concentration was increased by 2-fold from 5 to 10 mM suggesting that these compounds may activate the system at higher concentrations. To obtain a more accurate resolution of their dose-responses, the *Yp*ItcR/P_{ccl}-inducible system was subjected to a range of concentrations of *cis*-, and *trans*-aconitate. Since mesaconate has been previously shown to act as CoA acceptor by *Yp*Ict, with the second lowest K_m after itaconate[262], this compound was also included in the dose-response experiment.

It did not prove possible to obtain a saturation in fluorescence output when using mesaconate, *cis*-, or *trans*-aconitate as the inducer. All three inducers demonstrated some degree of toxicity inhibiting cell growth at higher concentrations. However, based on a phenomenological model for metabolite biosensors[186], it can be postulated that the maximal dynamic range of an inducible system, which is the maximal level of expression relative to basal promoter activity, is not affected by metabolite-TR affinity. Therefore, the maximal dynamic range calculated for itaconate as inducer was employed to fit the dynamic range data for mesaconate, *cis*-, and *trans*-aconitate using a Hill function (Figure 6.7). The resulting K_i, the extracellularly added inducer concentration which mediates half-maximal *rfp* expression, is different for each of these compounds. They reveal that mesaconate, *cis*- and *trans*-aconitate K_i values are higher (45.2 mM, 31.1 mM, and 13.2 mM, respectively) and therefore activate the *Yp*ItcR/P_{ccl}-inducible system at much higher extracellular concentrations than itaconate (K_i = 0.43 mM).

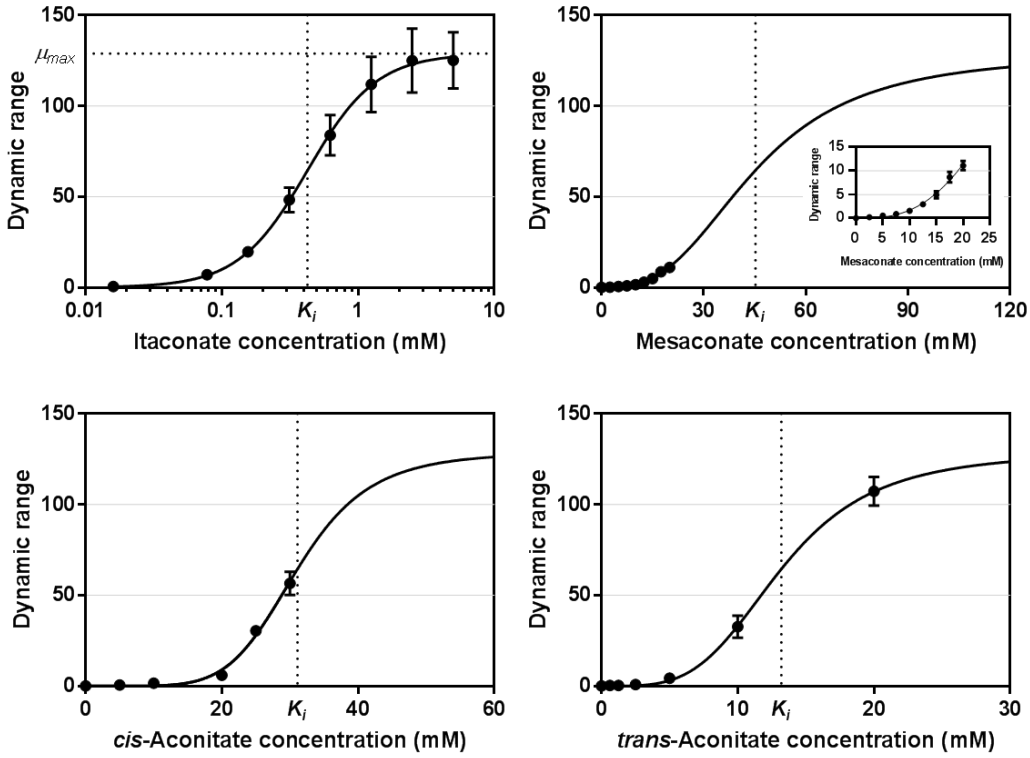


Figure 6.7 Dynamic range of the *Yp*ItcR/P_{ccl}-inducible system in response to different inducers. Dynamic range of the *Yp*ItcR/P_{ccl}-inducible system in *E. coli* MG1655 in response to various concentrations of itaconate, mesaconate, *cis*-aconitate, and *trans*-aconitate 6 h after inducer addition. The dynamic range was fit to the corresponding inducer concentration using a Hill function. The maximum dynamic range (μ_{max}) is indicated for itaconate. For the other three inducers, the values for μ were extrapolated to reach μ_{max} using the available data points. The inducer concentration mediating half-maximal *rfp* expression (K_i) is indicated for each compound. Error bars represent standard deviations of three biological replicates.

6.3.5 Biosensor-assisted optimisation of itaconic acid production

Itaconic acid can be synthesized by decarboxylation of the citric acid cycle intermediate *cis*-aconitic acid. This reaction is catalysed by *cis*-aconitate decarboxylase (CadA). The *A. terreus cadA* gene has previously been expressed in *E. coli* for the biosynthesis of itaconate by using either a constitutive promoter, or an inducible T7 polymerase-based expression system[244, 251, 265]. Overexpression of *cadA* was reported to impair cellular growth[265], suggesting that fine-tuning of CadA levels is essential to ensure optimal metabolic flux. Even though the pathway for itaconate biosynthesis in *E. coli* solely requires the introduction of one additional gene,

balancing its expression and quantitatively evaluating its impact on itaconate production can be laborious when using standard analytical techniques. It was of interest to apply the *Yp*ItcR/P_{ccl}-inducible system to monitor itaconate production by fluorescence output in response to different levels of CadA.

A single plasmid (pEH165) was constructed that contains two modules: one for itaconate production and one for itaconate sensing (Figure 6.8). The *A. terreus cadA* (ATEG_09971) gene was cloned downstream of the L-arabinose-inducible system and a T7 mRNA stem-loop structure sequence, which was incorporated to enhance *cadA* mRNA stability[86]. The itaconate sensing module contains the *Yp*ItcR/P_{ccl}-inducible system in combination with the *rfp* reporter gene. Addition of L-arabinose to cells harbouring this plasmid was expected to initiate *cadA* expression, resulting in biosynthesis of itaconate and subsequent activation of reporter gene expression. *E. coli* TOP10 was transformed with plasmid pEH165 and cells in early exponential growth phase were transferred to a 96-well microtiter plate. Subsequently, growth and fluorescence were monitored over time after supplementation with different concentrations of L-arabinose ranging from 1 to 1000 µM. As it can be seen in the fluorescence curve of induction kinetics, higher concentrations of L-arabinose mediate a faster fluorescence output (Figure 6.9a). Reporter gene expression above background levels can be observed 150 min after addition of 100 µM L-arabinose, whereas 10 µM require about one hour more. The dose-response curve indicates that maximum absolute normalised fluorescence is achieved by supplementation with 250 µM L-arabinose (Figure 6.9b). This suggests that expression of *cadA* can be fine-tuned when using inducer concentrations in the range between 1 and 100 µM. L-arabinose concentrations of 0.5 and 1 mM, however, appear to negatively impact reporter gene expression, indicating a drop in itaconate levels. The negative effect of high inducer

levels becomes even more evident from the absorbance data, showing that L-arabinose concentrations of 250 µM and more reduce cell density considerably (Figure 6.9c). Most likely, this behaviour results from an increased metabolic burden caused by overproduction of CadA, as mentioned earlier[265].

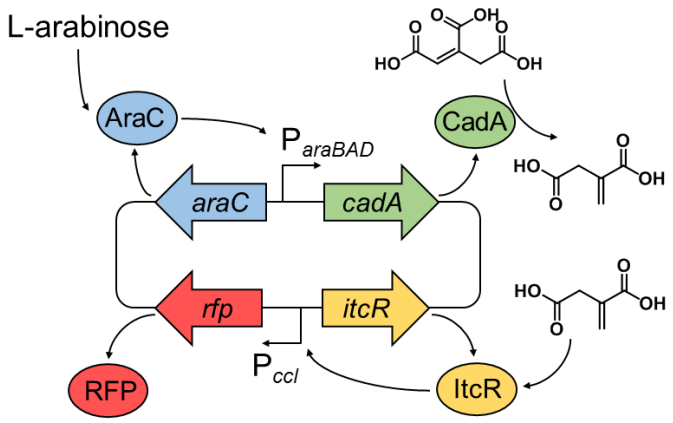


Figure 6.8 Schematic illustration of plasmid pEH165. It contains both an itaconate production- and sensing module. Exogenous addition of L-arabinose initiates synthesis of the *cis*-aconitate decarboxylase CadA which converts *cis*-aconitate into itaconate. Reporter gene expression is subsequently mediated by ItcR in the presence of itaconate.

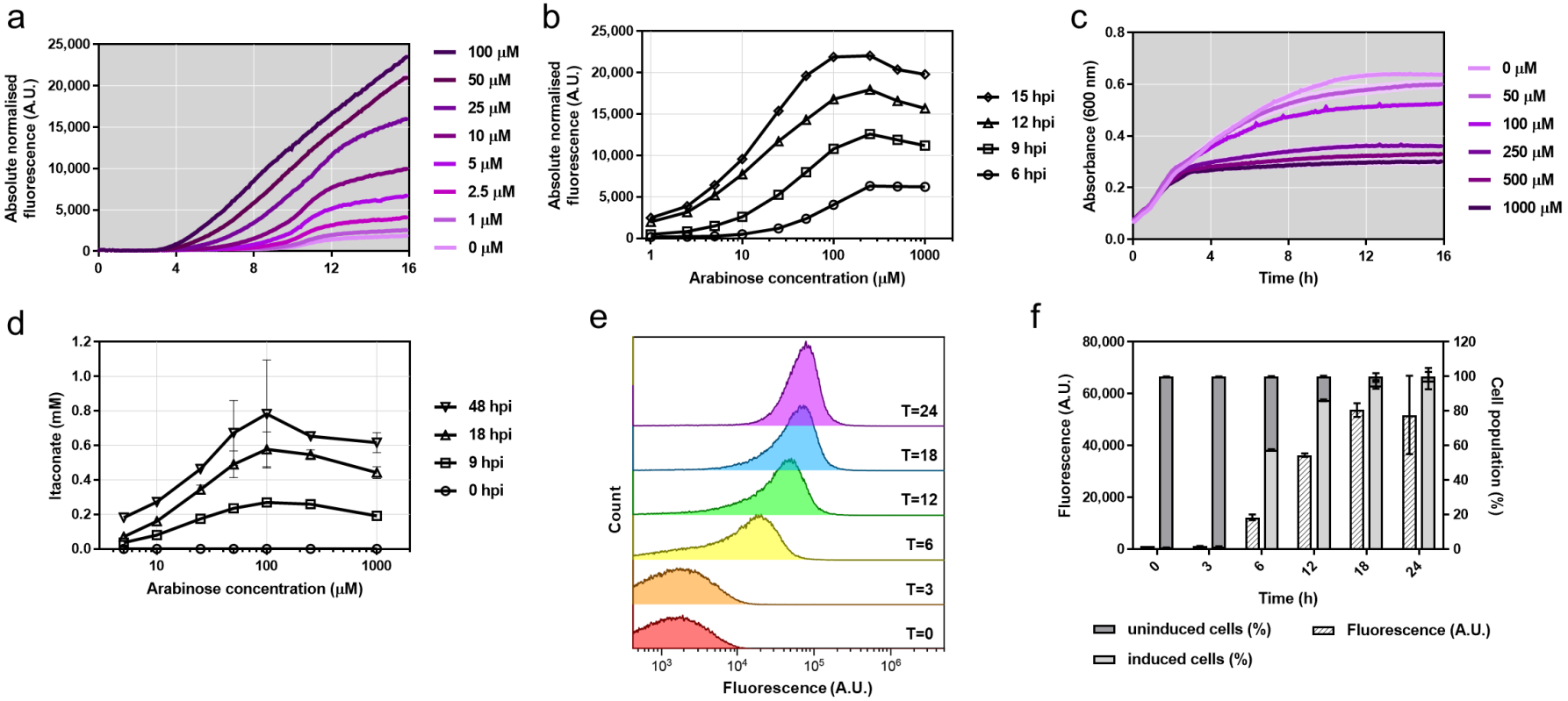

a
Absolute normalised fluorescence (A.U.)
25,000
20,000
15,000
10,000
5,000
0
0
4
8
12
16
100 μM
50 μM
25 μM
10 μM
5 μM
2.5 μM
1 μM
0 μM
b
Absolute normalised fluorescence (A.U.)
25,000
20,000
15,000
10,000
5,000
0
1
10
100
1000
Arabinose concentration (μM)
15 hpi
12 hpi
9 hpi
6 hpi
c
Absorbance (600 nm)
0.8
0.6
0.4
0.2
0.0
0
4
8
12
16
Time (h)
0 μM
50 μM
100 μM
250 μM
500 μM
1000 μM
d
Itaconate (mM)
1.2
1.0
0.8
0.6
0.4
0.2
0.0
10
100
1000
Arabinose concentration (μM)
48 hpi
18 hpi
9 hpi
0 hpi
e
Count
T=24
T=18
T=12
T=6
T=3
T=0
10^3
10^4
10^5
10^6
Fluorescence (A.U.)
f
Fluorescence (A.U.)
80,000
60,000
40,000
20,000
0
Cell population (%)
120
100
80
60
40
20
0
0
3
6
12
18
24
Time (h)
uninduced cells (%)
induced cells (%)
Fluorescence (A.U.)

Figure 6.9 Biosensor-assisted optimisation of itaconate production. (a) Absolute normalised fluorescence of *E. coli* TOP10 carrying pEH165, grown in microtiter plates, in response to 1-100 µM of L-arabinose supplemented at time zero. The means of three biological replicates are presented. Error bars are too small to be visible. (b) Dose-response curve of *E. coli* TOP10 carrying pEH165, grown in microtiter plates, 6, 9, 12, and 15 h post induction (hpi) with 1-1000 µM of L-arabinose. The means of three biological replicates are presented. Error bars are too small to be visible. (c) Absorbance at 600 nm of *E. coli* TOP10 carrying pEH165, grown in microtiter plates, in response to 50-1000 µM of L-arabinose supplemented at time zero. The means of three biological replicates are presented. The standard deviation for 50 µM of inducer is illustrated as lighter colour ribbon displayed lengthwise the growth curve. The error bars for the other inducer concentrations are too small to be visible. (d) Itaconate titres of *E. coli* TOP10 carrying pEH165, grown in small-volume cultures, 0, 9, 18, and 48 h post induction with 5, 10, 25, 50, 100, 250, and 1000 µM of L-arabinose. Error bars represent standard deviations of three biological replicates. (e) Flow cytometric analysis of *E. coli* TOP 10 carrying pEH165, grown in small-volume cultures, in response to 100 µM of L-arabinose. Samples were taken 0, 3, 6, 12, 18 and 24 h after inducer addition. For the time points T=6, T=12, T=18 and T=24, fluorescence from more than 99% of cells are displayed in the histogram, whereas for time points T=0 and T=3, less than 25% of cells are below 429 A. U. fluorescence threshold in the histogram. (f) Fluorescence intensity (median) and percentage of uninduced and induced cells corresponding to the data presented in panel e. Error bars represent standard deviations of three biological replicates.

To quantitatively validate the data which was generated from cultures grown in microtiter plates, the experiment was repeated in small culture volumes. *E. coli* TOP10 pEH165 was grown in 50-mL culture tubes and expression of *cadA* was initiated by supplementation with different concentrations of L-arabinose. To determine itaconate titres, samples were subjected to the analysis using HPLC-UV. The highest itaconate concentration was achieved in cultures containing 100 µM L-arabinose, resulting in 0.78 ± 0.31 mM itaconate 48 h after inducer addition (Figure 6.9d). This represents a 4.3-fold improvement over cultures containing only 5 µM L-arabinose. It should be noted that these and the data in Table 6.2 demonstrate that the intracellularly synthesised itaconate was actively excreted or diffused into the media.

Addition of an excessive amount of 1 mM inducer also decreased itaconate levels by 1.3-fold. Therefore, the quantitative data obtained from the small-volume

cultures match well with the fluorescence output measured in the microtiter plate (compare Figure 6.9b and Figure 6.9d). Particularly when itaconate titres are OD-normalised, 250 µM L-arabinose results in the highest OD-normalised itaconate titre (Figure 6.10). This experiment illustrates that *cadA* expression needs to be carefully fine-tuned to guarantee both optimal metabolic flux and viability of cells.

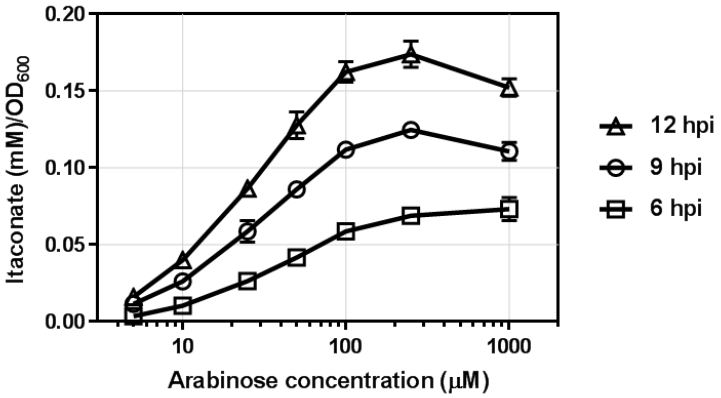


Figure 6.10 OD-normalised itaconate titres of *E. coli* TOP10 harbouring pEH165. *E. coli* was grown in small-volume cultures and samples were taken 6, 9, and 12 h post induction (hpi) with 5, 10, 25, 50, 100, 250, and 1000 µM of L-arabinose. Error bars represent standard deviations of three biological replicates.

Moreover, using 100 µM of L-arabinose yields itaconate concentrations of 0.24, 0.56 and 0.78 mM after 9, 18 and 48 h post induction, respectively (Figure 6.9d). These itaconate concentrations fall within the linear range of dose response (Figure 6.3b) and result in a fluorescence output with a unimodal distribution suggesting that almost all cells in the population were activated (Figure 6.9e,f). As demonstrated here, the itaconate biosensor can be employed to facilitate a fluorescence-based high-throughput screen to evaluate various conditions for their impact on itaconate biosynthesis.

6.3.6 Correlation between biosensor output and itaconate concentration

In addition to HPLC-UV analysis, the samples from the small-volume cultures of *E. coli* TOP10 carrying pEH165 were analysed for fluorescence output. The obtained data were used to evaluate whether quantitatively determined itaconate titres correlate with reporter gene expression from the biosensor. The five tested inducer concentrations that did not impair bacterial growth produced a 59-fold range in fluorescence after 6 h (Figure 6.11a). The addition of 25, 50 and 100 µM L-arabinose resulted in itaconate titres that were sufficiently high to be detected by the biosensor. Notably, in the linear response range of the *Yp*ItcR/P_{ccl}-inducible system, the fluorescence output shows a high level of correlation with HPLC-UV-measured extracellular itaconate titres (Figure 6.11a) and unimodal fluorescence distribution in the cell population (Figure 6.11b). L-arabinose concentrations of 5 and 10 µM result in a bimodal fluorescence response, suggesting an all-or-none induction in which intermediate inducer concentrations give rise to subpopulations. However, when different levels of itaconate are synthesised in the range between 0.1 and 0.78 mM, which corresponds to the linear response range, the fluorescence output becomes unimodal (Figure 6.9d,e and Figure 6.11a). This confirms that for itaconate levels in the linear range, the *Yp*ItcR/P_{ccl}-inducible system mediates a homogenous induction of cells, exemplifying its potential to fine-tune gene expression across cell populations and to be utilised as a quantitatively reliable biosensor.

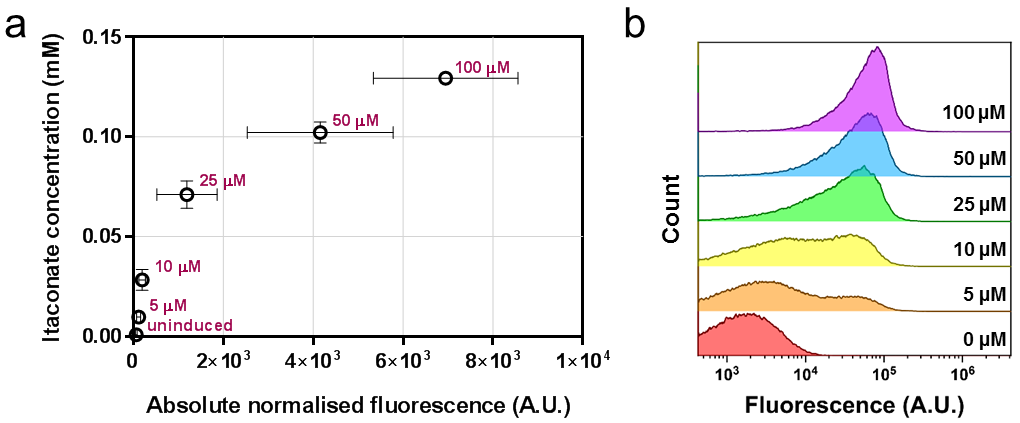


Figure 6.11 Correlation between biosensor output and itaconate concentration. (a) Absolute normalised fluorescence values of *E. coli* TOP10 carrying pEH165 are correlated with their corresponding itaconate concentration in the culture supernatant. Samples were taken 6 h after inducer addition. The different concentrations of exogenously added L-arabinose, ranging from 5 to 100 µM, are highlighted. Error bars represent standard deviations of three biological replicates. (b) Flow cytometric analysis of samples from panel a. For L-arabinose (inducer) concentrations of 25, 50, and 100 µM, fluorescence from more than 99% of cells are displayed in the histogram, whereas for concentrations of 10, 5 and 0 µM, less than 2, 10 and 25% of cells, respectively, are below 429 A. U. fluorescence threshold in the histogram.

6.4 Discussion

Itaconic acid is an important platform chemical that can easily be incorporated into polymers and has the potential to replace petrochemical-based acrylic or methacrylic acid. A number of microorganisms have been developed for the biosynthesis of itaconate including *A. terreus*, *E. coli*, and *S. cerevisiae*[247, 251, 252]. However, the number of strains and conditions that can be tested for increased itaconate titres are currently limited due to the lack of high-throughput screening methods. In the previous chapters, the highly conserved genetic arrangement typical of LTTRs was chosen to serve as the basis for mining inducible systems associated with metabolic gene clusters. This strategy was further expanded to identify itaconate-inducible promoters and their corresponding LTTRs in the genomes of *Y. pseudotuberculosis* YPIII and *P. aeruginosa* PAO1. Their functionality was evaluated in the model γ-proteobacterium *E. coli* and the β-proteobacterium *C. necator*.

Of the two systems, the *Y. pseudotuberculosis* ItcR/P_{ccl}-inducible system was demonstrated to be orthogonal to host-originating transcriptional regulation in both tested microorganisms. The *P. aeruginosa* P_{ich}, however, was induced by itaconate even in the absence of its cognate TR in *C. necator*. A *Pa*ItcR homology search in *C. necator* revealed the presence of several chromosomally encoded LTTRs exhibiting 40-50% protein sequence identity (96-98% coverage). One of the LTTR genes is located within close proximity to the cluster that includes genes potentially involved in itaconate degradation similar to *P. aeruginosa* (Supplementary Figure 7). *C. necator* ItcR homologues can potentially activate gene expression from the heterologous *P. aeruginosa* itaconate-inducible promoter even in the absence of its corresponding LTTR. However, since both the induction level and the absolute normalised fluorescence in the presence of itaconate are higher in the plasmid carrying *Pa*ItcR/P_{ich} (pEH177) than the one harbouring the inducible promoter *Pa*P_{ich} (pEH178) alone (by 3.5- and 52-fold, respectively), it can be concluded that *Pa*ItcR is involved in activation of gene expression of the itaconate degradation cluster of genes in *P. aeruginosa* and therefore enables persistence in macrophages similar to *Yp*ItcR. The finding that expression of the genes encoding enzymes involved in itaconate catabolism is mediated by their divergently oriented LTTR genes in *Y. pseudotuberculosis* and *P. aeruginosa* may aid in developing new antimicrobial agents.

Because of its regulator-dependent orthogonality and high dynamic range in both *E. coli* and *C. necator*, the itaconate-inducible system from *Y. pseudotuberculosis* was selected to be characterised in more detail. It was observed in minimal medium that itaconate concentrations above 2.5 mM resulted in a saturation of the *Yp*ItcR/P_{ccl}-inducible system However, in order for this system to be applied as

biosensor for concentrations higher than 2.5 mM, its elements require modification. This is commonly accomplished by promoter or protein engineering, both strategies aiming to alter the binding affinity of the TR for either the operator sequence or the ligand itself[186, 266, 267].

In addition to its regulator-dependent orthogonality, this system was subjected to a range of compounds that are structurally similar to itaconate to determine its ligand-specificity. Three of the investigated metabolites, namely mesaconate, *cis*-, and *trans*-aconitate were shown to activate gene expression from the *Yp*ItcR/P_{ccl}-inducible system in addition to itaconate. As *E. coli* has not been reported to metabolise mesaconate or *trans*-aconitate, induction of the *Yp*ItcR/P_{ccl}-inducible system may result from structural resemblance of all three compounds to the primary effector. Indeed, mesaconate, *cis*-, and *trans*-aconitate have structural similarities to itaconate, with the latter two harbouring the complete itaconate element. However, the observation that all three compounds have a much higher K_i than itaconate suggests, that for maximal activation of the *Yp*ItcR/P_{ccl}-inducible system, the unmodified itaconate structure is indispensable. On the other hand, it cannot be excluded that the change in inducer dynamic range is affected by the differential uptake of these compounds by the *E. coli* cell.

It should be noted that acetate, propanoate, butanoate, methylsuccinate, and mesaconate have previously been demonstrated to act as CoA acceptors by *Yp*Ict, albeit at a much higher K_m than itaconate[262], suggesting that these compounds might be secondary effectors of the *Yp*ItcR/P_{ccl}-inducible system. Interestingly, their level of induction correlates with their ability to act as CoA acceptors, with acetate, propanoate, and butanoate having a higher, and mesaconate having a lower K_m[262]. Furthermore, the catalytic efficiency (k_{cat}/K_m) of *Yp*Ict with itaconate, mesaconate,

methylsuccinate, butanoate, propanoate, and acetate[262], shows a high level of direct correlation with the level of induction by these compounds. This suggests that there might be a structural evolutionary link between enzyme (*Yp*Ict) and transcriptional regulator (ItcR), where both proteins have coevolved enabling a hierarchical ranking of metabolites as enzyme substrates and TR activators in the following order: itaconate > mesaconate > methylsuccinate > butanoate > propanoate > acetate. The direct correlation between catalytic efficiency and level of induction potentially ensures that the hierarchy is supported at the gene expression and enzyme activity levels by securing the highest level of *Yp*Ict synthesis and highest catalytic efficiency when itaconate is present in the environment. Overall, the *Yp*ItcR/P_{ccl}-inducible system demonstrates a high specificity toward itaconate and may therefore be used in combination with other inducible systems as outlined in Chapter 4 to orthogonally control gene expression in biosynthetic pathways composed of multiple genes.

To conclude, this chapter highlights the potential of the *Yp*ItcR/P_{ccl}-inducible system to be applied as biosensor for high-throughput microbial strain development in order to facilitate improved itaconate biosynthesis. It expands the genetic toolbox for engineering *C. necator*, however, to unlock its full potential, itaconate metabolic pathways must be identified and eliminated.

6.5 Conclusion

Itaconic acid-inducible promoters and their corresponding LTTRs were identified in *Y. pseudotuberculosis* and *P. aeruginosa*. The *Yp*ItcR/P_{ccl}-inducible system is highly inducible by itaconate in *E. coli* and *C. necator* (215- and 105-fold, respectively). In addition to itaconate, the genetically encoded biosensor is capable of detecting mesaconate, *cis*-, and *trans*-aconitate in a dose-dependednt manner. The fluorescence-based biosensor was applied in *E. coli* to identify the optimum expression level of

cadA, the protein product of which catalyses the conversion of *cis*-aconitate into itaconate. The fluorescence output is shown to correlate well with itaconate concentrations quantified using HPLC-UV.

7 Conclusions and outlook

The main research aims of this thesis were the identification, characterisation, and application of orthogonal and highly inducible transcription factor-based controllable gene expression systems for metabolic engineering of *C. necator* H16 and other bacteria. Specifically, inducible systems responding to the industrially relevant building block chemicals 3-HP and itaconate were to be identified and assembled into biosensors. A genome-wide approach to mining inducible systems from annotated bacterial genomes was to be developed and applied in *C. necator* to expand the range of metabolites that can be detected using biosensors. The inducible systems identified by following this strategy, along with previously reported heterologous inducible systems, were to be quantitatively evaluated for their ability to control gene expression in a comparative manner, parameterised to facilitate forward engineering efforts and analysed for their orthogonality. Selected heterologous inducible systems were to be applied for production of isoprene in *C. necator* and the itaconate biosensor was to be leveraged to improve microbial itaconate biosynthesis.

In the past, inducible gene expression systems have been successfully implemented in the design, construction, and evaluation of metabolically engineered strains. These success stories, however, are based on a few inducible systems that have been reused across different research groups for more than two decades. Although the number of available metabolite-inducible systems continuously increases, the differences between their evaluation methods limits these regulatory elements from being used to their full potential. More comparative studies are needed to aid in part selection. For example, the benzoate-inducible system sourced from the genome of *C. necator* may be a more cost-effective alternative for high levels of protein production in a wide range of bacteria. Compared to inducers that mediate a similar

expression output at 5 mM, it is approximately 20 (L-arabinose) to 60 (L-rhamnose) times less expensive. Taking the minimum concentration of inducer that is required to fully induce the system into consideration, these factors may even be multiplied by at least one order of magnitude. The developed library of metabolite-responsive inducible systems represents only a small fraction of available biosensors. Its expansion offers significant potential to be applied as universal synthetic biology tool kit. The modular design of the reporter vectors enables rapid exchange of selection markers and replicons, thus allowing the evaluation of metabolite-induced gene expression in any prokaryotic organism.

In this context, the developed genome-wide approach to identifying native inducible systems led to the discovery of 15 novel or little characterised transcription factor-based inducible systems in *C. necator*. To simplify the manual screen of the genome for catabolic gene clusters, the search was constrained to annotated genes. This limitation, though necessary to conclude the ultimate effector molecule, reveals the immense potential for the discovery of even more regulatory switches since the majority of genes had not been annotated. With the rapidly growing genomic data these new capabilities offer a platform to vastly increase the number of biologically detectable molecules.

The orthogonality screen yielded a number of inducible systems that can be used in combination to independtly control gene expression or built synthetic regulatory networks. As a consequence of the screen, the cross-reactivity of a number of metabolites with non-cognate inducible systems was investigated. The exploration of metabolism by using metabolite-responsive biosensors as devices to uncover unknown catalytic functions represents a completely new field of molecular biology. Once a more sophisticated libraray of transcription factor-based inducible systems

becomes available, the kinetics and dynamics of biocatalytical conversions may be easily visualised *in vivo* using fluorescence-based biosensors.

A significant part of this research was dedicated to the identification of inducibe systems and the construction of biosensors that respond to the platform chemicals 3-HP and itaconic acid. Their dynamic ranges may suffice to act as switches for dynamic pathway control, however, in order to determine product titres of already high-producing strains, their dynamic ranges may have to be modulated. This will be achieved through promoter- or protein engineering, both strategies aiming to change the TR ligand- or DNA-binding affinity. In addition to addressing the range of ligand concentration that results in a change in expression output, future engineering efforts should be directed towards decreasing unspecific basal promoter activities of inducible systems. High levels of background gene expression in the absence of ligand limit their application. Naturally occurring regulatory circuits, such as feedforward loops, may be used as blueprints for the design of novel synthetic controllable systems that are more robust and less prone to spurious activation.

C. necator is an attractive biocatalyst for the conversion of CO_2 into these building block chemicals. Unfortunately, it possesses multiple gene clusters that are involved in the catabolism of 3-HP and itaconate. To facilitate their biosynthesis in this host organism these gene clusters must be identified and deleted. In contrast to 3-HP and itaconate, isoprene is a volatile compound and accumulates in the headspace of the bacterial culture. Production of isoprene may be improved by introducing the mevalonate pathway, the eukaryotic counterpart of the MEP/DOXP pathway, to bypass the natural negative feedback mechanism that prevents the biosynthesis of isoprenoids.

To conclude, this thesis advances our abilities to control and evaluate biological designs. On the basis of catabolic gene clusters, a range of novel transcription factor-based inducible gene expression systems were discovered, including the phenylglyoxylate-, β-alanine-, 3-HP-, and itaconate-inducible systems. The identified and characterised regulatory switches will aid in the engieneering of microbial cell factories for the sustainable production of bio-based chemcials from renewable resources or waste.

8 References

1. European Commission, Directorate-General for Research and Innovation. A sustainable Bioeconomy for Europe: Strengthening the connection between economy, society and the environment. (Publications Office of the European Union, Brussels; 2018).
2. Langeveld, H., Sanders, J. & Meeusen, M. The biobased economy: Biofuels, materials and chemicals in the post-oil era. (Earthscan, London; 2012).
3. Beutler, J.A. Natural products as a foundation for drug discovery. *Curr. Protoc. Pharmacol.* **46**, 9.11. 11-19.11. 21 (2009).
4. Keasling, J.D. Manufacturing molecules through metabolic engineering. *Science* **330**, 1355-1358 (2010).
5. Cho, A., Yun, H., Park, J.H., Lee, S.Y. & Park, S. Prediction of novel synthetic pathways for the production of desired chemicals. *BMC Syst. Biol.* **4**, 35 (2010).
6. Yim, H. et al. Metabolic engineering of *Escherichia coli* for direct production of 1, 4-butanediol. *Nat. Chem. Biol.* **7**, 445 (2011).
7. Nielsen, J. & Keasling, J.D. Engineering cellular metabolism. *Cell* **164**, 1185-1197 (2016).
8. Le Quéré, C. et al. Global carbon budget 2018. *Earth Syst. Sci. Data* **10** (2018).
9. Liew, F. et al. Gas fermentation—a flexible platform for commercial scale production of low-carbon-fuels and chemicals from waste and renewable feedstocks. *Front. Microbiol.* **7**, 694 (2016).
10. Hagos, F.Y., Aziz, A.R.A. & Sulaiman, S.A. Trends of syngas as a fuel in internal combustion engines. *Adv. Mech. Eng.* **6**, 401587 (2014).
11. Sridhar, G., Paul, P. & Mukunda, H. Biomass derived producer gas as a reciprocating engine fuel—an experimental analysis. *Biomass Bioenergy* **21**, 61-72 (2001).
12. Humphreys, C.M. & Minton, N.P. Advances in metabolic engineering in the microbial production of fuels and chemicals from C1 gas. *Curr. Opin. Biotechnol.* **50**, 174-181 (2018).
13. Heijstra, B.D., Leang, C. & Juminaga, A. Gas fermentation: cellular engineering possibilities and scale up. *Microb. Cell Fact.* **16**, 60 (2017).
14. Seshasayee, A.S., Bertone, P., Fraser, G.M. & Luscombe, N.M. Transcriptional regulatory networks in bacteria: from input signals to output responses. *Curr. Opin. Microbiol.* **9**, 511-519 (2006).
15. Kalisky, T., Dekel, E. & Alon, U. Cost–benefit theory and optimal design of gene regulation functions. *Phys. Biol.* **4**, 229 (2007).
16. Browning, D.F. & Busby, S.J. Local and global regulation of transcription initiation in bacteria. *Nat. Rev. Microbiol.* **14**, 638 (2016).
17. Feklístov, A., Sharon, B.D., Darst, S.A. & Gross, C.A. Bacterial sigma factors: a historical, structural, and genomic perspective. *Annu. Rev. Microbiol.* **68**, 357-376 (2014).
18. Yang, Y. et al. Structures of the RNA polymerase-σ54 reveal new and conserved regulatory strategies. *Science* **349**, 882-885 (2015).
19. Brennan, R.G. & Matthews, B.W. The helix-turn-helix DNA binding motif. *J. Biol. Chem.* **264**, 1903-1906 (1989).

20. Huffman, J.L. & Brennan, R.G. Prokaryotic transcription regulators: more than just the helix-turn-helix motif. *Curr. Opin. Struct. Biol.* **12**, 98-106 (2002).

21. Raibaud, O. & Schwartz, M. Positive control of transcription initiation in bacteria. *Annu. Rev. Genet.* **18**, 173-230 (1984).

22. Orth, P., Schnappinger, D., Hillen, W., Saenger, W. & Hinrichs, W. Structural basis of gene regulation by the tetracycline inducible Tet repressor–operator system. *Nat. Struct. Biol.* **7**, 215-219 (2000).

23. Swint-Kruse, L. & Matthews, K.S. Allostery in the LacI/GalR family: variations on a theme. *Curr. Opin. Microbiol.* **12**, 129-137 (2009).

24. Valentin-Hansen, P., Søgaard-Andersen, L. & Pedersen, H. A flexible partnership: the CytR anti-activator and the cAMP–CRP activator protein, comrades in transcription control. *Mol. Microbiol.* **20**, 461-466 (1996).

25. Monsalve, M., Mencía, M., Salas, M. & Rojo, F. Protein p4 represses phage phi 29 A2c promoter by interacting with the alpha subunit of *Bacillus subtilis* RNA polymerase. *PNAS* **93**, 8913-8918 (1996).

26. Lawson, C.L. et al. Catabolite activator protein: DNA binding and transcription activation. *Curr. Opin. Struct. Biol.* **14**, 10-20 (2004).

27. Heldwein, E.E.Z. & Brennan, R.G. Crystal structure of the transcription activator BmrR bound to DNA and a drug. *Nature* **409**, 378-382 (2001).

28. Sperandio, V. et al. Activation of enteropathogenic *Escherichia coli* (EPEC) LEE2 and LEE3 operons by Ler. *Mol. Microbiol.* **38**, 781-793 (2000).

29. Medina, G., Juárez, K., Valderrama, B. & Soberón-Chávez, G. Mechanism of *Pseudomonas aeruginosa* RhlR transcriptional regulation of the *rhlAB* promoter. *J. Bacteriol.* **185**, 5976-5983 (2003).

30. Maddocks, S.E. & Oyston, P.C. Structure and function of the LysR-type transcriptional regulator (LTTR) family proteins. *Microbiology* **154**, 3609-3623 (2008).

31. Ezezika, O.C., Collier-Hyams, L.S., Dale, H.A., Burk, A.C. & Neidle, E.L. CatM regulation of the *benABCDE* operon: functional divergence of two LysR-type paralogs in *Acinetobacter baylyi* ADP1. *Appl. Environ. Microbiol.* **72**, 1749-1758 (2006).

32. Lu, Z., Takeuchi, M. & Sato, T. The LysR-type transcriptional regulator YofA controls cell division through the regulation of expression of *ftsW* in *Bacillus subtilis*. *J. Bacteriol.* **189**, 5642-5651 (2007).

33. Sperandio, V., Li, C.C. & Kaper, J.B. Quorum-sensing *Escherichia coli* regulator A: a regulator of the LysR family involved in the regulation of the locus of enterocyte effacement pathogenicity island in enterohemorrhagic *E. coli*. *Infect. Immun.* **70**, 3085-3093 (2002).

34. Schlaman, H., Okker, R.J. & Lugtenberg, B.J. Regulation of nodulation gene expression by NodD in Rhizobia. *J. Bacteriol.* **174**, 5177 (1992).

35. Sheehan, B.J. & Dorman, C.J. *In vivo* analysis of the interactions of the LysR-like regulator SpvR with the operator sequences of the *spvA* and *spvR* virulence genes of *Salmonella typhimurium*. *Mol. Microbiol.* **30**, 91-105 (1998).

36. Alon, U. Network motifs: theory and experimental approaches. *Nat. Rev. Genet.* **8**, 450-461 (2007).

37. Milo, R. et al. Network motifs: simple building blocks of complex networks. *Science* **298**, 824-827 (2002).

38. Eichenberger, P. et al. The program of gene transcription for a single differentiating cell type during sporulation in *Bacillus subtilis*. *PLoS Biol.* **2** (2004).

39. Lee, T.I. et al. Transcriptional regulatory networks in *Saccharomyces cerevisiae*. *Science* **298**, 799-804 (2002).

40. Thieffry, D., Huerta, A.M., Pérez-Rueda, E. & Collado-Vides, J. From specific gene regulation to genomic networks: a global analysis of transcriptional regulation in *Escherichia coli*. *Bioessays* **20**, 433-440 (1998).

41. Mangan, S. & Alon, U. Structure and function of the feed-forward loop network motif. *PNAS* **100**, 11980-11985 (2003).

42. Ma, H.-W. et al. An extended transcriptional regulatory network of *Escherichia coli* and analysis of its hierarchical structure and network motifs. *Nucleic Acids Res.* **32**, 6643-6649 (2004).

43. Ogden, S., Haggerty, D., Stoner, C.M., Kolodrubetz, D. & Schleif, R. The *Escherichia coli* L-arabinose operon: binding sites of the regulatory proteins and a mechanism of positive and negative regulation. *PNAS* **77**, 3346-3350 (1980).

44. Mangan, S., Zaslaver, A. & Alon, U. The coherent feedforward loop serves as a sign-sensitive delay element in transcription networks. *J. Mol. Biol.* **334**, 197-204 (2003).

45. Kalir, S. & Alon, U. Using a quantitative blueprint to reprogram the dynamics of the flagella gene network. *Cell* **117**, 713-720 (2004).

46. Kalir, S., Mangan, S. & Alon, U. A coherent feed-forward loop with a SUM input function prolongs flagella expression in *Escherichia coli*. *Mol. Syst. Biol.* **1** (2005).

47. Brophy, J.A. & Voigt, C.A. Principles of genetic circuit design. *Nat. Methods* **11**, 508 (2014).

48. Gräslund, S. et al. Protein production and purification. *Nat. Methods* **5**, 135 (2008).

49. Studier, F.W. & Moffatt, B.A. Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J. Mol. Biol.* **189**, 113-130 (1986).

50. Sørensen, H.P. & Mortensen, K.K. Advanced genetic strategies for recombinant protein expression in *Escherichia coli*. *J. Biotechnol.* **115**, 113-128 (2005).

51. Voigt, C.A. Genetic parts to program bacteria. *Curr. Opin. Biotechnol.* **17**, 548-557 (2006).

52. Moon, T.S., Lou, C., Tamsir, A., Stanton, B.C. & Voigt, C.A. Genetic programs constructed from layered logic gates in single cells. *Nature* **491**, 249 (2012).

53. Gardner, T.S., Cantor, C.R. & Collins, J.J. Construction of a genetic toggle switch in *Escherichia coli*. *Nature* **403**, 339 (2000).

54. Elowitz, M.B. & Leibler, S. A synthetic oscillatory network of transcriptional regulators. *Nature* **403**, 335 (2000).

55. Ptashne, M. A genetic switch: phage λ and higher organisms. (Cell, Cambridge, Massachusetts; 1992).

56. Dunlap, J.C. Molecular bases for circadian clocks. *Cell* **96**, 271-290 (1999).

57. Anderson, J.C., Voigt, C.A. & Arkin, A.P. Environmental signal integration by a modular AND gate. *Mol. Syst. Biol.* **3** (2007).

58. Stanton, B.C. et al. Genomic mining of prokaryotic repressors for orthogonal logic gates. *Nat. Chem. Biol.* **10**, 99 (2014).

59. Basu, S., Mehreja, R., Thiberge, S., Chen, M.-T. & Weiss, R. Spatiotemporal control of gene expression with pulse-generating networks. *PNAS* **101**, 6355-6360 (2004).

60. Basu, S., Gerchman, Y., Collins, C.H., Arnold, F.H. & Weiss, R. A synthetic multicellular system for programmed pattern formation. *Nature* **434**, 1130 (2005).

61. Wang, H.H. et al. Programming cells by multiplex genome engineering and accelerated evolution. *Nature* **460**, 894-898 (2009).

62. Rogers, J.K., Taylor, N.D. & Church, G.M. Biosensor-based engineering of biosynthetic pathways. *Curr. Opin. Biotechnol.* **42**, 84-91 (2016).

63. Mustafi, N., Grünberger, A., Kohlheyer, D., Bott, M. & Frunzke, J. The development and application of a single-cell biosensor for the detection of L-methionine and branched-chain amino acids. *Metab. Eng.* **14**, 449-457 (2012).

64. Mahr, R. et al. Biosensor-driven adaptive laboratory evolution of L-valine production in *Corynebacterium glutamicum*. *Metab. Eng.* **32**, 184-194 (2015).

65. Liu, Y.n. et al. Developing a high-throughput screening method for threonine overproduction based on an artificial promoter. *Microb. Cell Fact.* **14**, 121 (2015).

66. Schendzielorz, G. et al. Taking control over control: use of product sensing in single cells to remove flux control at key enzymes in biosynthesis pathways. *ACS Synth. Biol.* **3**, 21-29 (2013).

67. Tang, S.-Y. et al. Screening for enhanced triacetic acid lactone production by recombinant *Escherichia coli* expressing a designed triacetic acid lactone reporter. *J. Am. Chem. Soc.* **135**, 10099-10103 (2013).

68. van Sint Fiet, S., van Beilen, J.B. & Witholt, B. Selection of biocatalysts for chemical synthesis. *PNAS* **103**, 1693-1698 (2006).

69. Zhang, F., Carothers, J.M. & Keasling, J.D. Design of a dynamic sensor-regulator system for production of chemicals and fuels derived from fatty acids. *Nat. Biotechnol.* **30**, 354 (2012).

70. Serganov, A. & Nudler, E. A decade of riboswitches. *Cell* **152**, 17-24 (2013).

71. Wilson, D., Charoensawan, V., Kummerfeld, S.K. & Teichmann, S.A. DBD—taxonomically broad transcription factor predictions: new content and functionality. *Nucleic Acids Res.* **36**, D88-D92 (2007).

72. Münch, R. et al. PRODORIC: prokaryotic database of gene regulation. *Nucleic Acids Res.* **31**, 266-269 (2003).

73. Novichkov, P.S. et al. RegPrecise 3.0–a resource for genome-scale exploration of transcriptional regulation in bacteria. *BMC Genomics* **14**, 745 (2013).

74. Rogers, J.K. et al. Synthetic biosensors for precise gene control and real-time monitoring of metabolites. *Nucleic Acids Res.* **43**, 7648-7660 (2015).

75. Dahl, R.H. et al. Engineering dynamic pathway regulation using stress-response promoters. *Nat. Biotechnol.* **31**, 1039 (2013).

76. Mahr, R., von Boeselager, R.F., Wiechert, J. & Frunzke, J. Screening of an *Escherichia coli* promoter library for a phenylalanine biosensor. *Appl. Microbiol. Biotechnol.* **100**, 6739-6753 (2016).

77. Uchiyama, T. & Miyazaki, K. Metagenomic screening for aromatic compound-responsive transcriptional regulators. *PloS One* **8**, e75795 (2013).

78. Ahn, S.K., Cuthbertson, L. & Nodwell, J.R. Genome context as a predictive tool for identifying regulatory targets of the TetR family transcriptional regulators. *PloS One* **7**, e50562 (2012).

79. Cuthbertson, L., Ahn, S.K. & Nodwell, J.R. Deglycosylation as a mechanism of inducible antibiotic resistance revealed using a global relational tree for one-component regulators. *Chem. Biol.* **20**, 232-240 (2013).

80. Ravcheev, D.A. et al. Comparative genomics and evolution of regulons of the LacI-family transcription factors. *Front. Microbiol.* **5**, 294 (2014).

81. Kazanov, M.D., Li, X., Gelfand, M.S., Osterman, A.L. & Rodionov, D.A. Functional diversification of ROK-family transcriptional regulators of sugar catabolism in the Thermotogae phylum. *Nucleic Acids Res.* **41**, 790-803 (2012).

82. Grousseau, E. et al. Impact of sustaining a controlled residual growth on polyhydroxybutyrate yield and production kinetics in *Cupriavidus necator*. *Bioresour. Technol.* **148**, 30-38 (2013).

83. Chen, G.-Q. A microbial polyhydroxyalkanoates (PHA) based bio-and materials industry. *Chem. Soc. Rev.* **38**, 2434-2446 (2009).

84. Jendrossek, D. Polyhydroxyalkanoate granules are complex subcellular organelles (carbonosomes). *J. Bacteriol.* **191**, 3195-3202 (2009).

85. Reinecke, F. & Steinbüchel, A. *Ralstonia eutropha* strain H16 as model organism for PHA metabolism and for biotechnological production of technically interesting biopolymers. *J. Mol. Microbiol. Biotechnol.* **16**, 91-108 (2009).

86. Bi, C. et al. Development of a broad-host synthetic biology toolbox for *Ralstonia eutropha* and its application to engineering hydrocarbon biofuel production. *Microb. Cell Fact.* **12**, 107 (2013).

87. Crépin, L., Lombard, E. & Guillouet, S.E. Metabolic engineering of *Cupriavidus necator* for heterotrophic and autotrophic alka(e)ne production. *Metab. Eng.* **37**, 92-101 (2016).

88. Chen, J.S. et al. Production of fatty acids in *Ralstonia eutropha* H16 by engineering β-oxidation and carbon storage. *PeerJ* **3**, e1468 (2015).

89. Li, H. et al. Integrated electromicrobial conversion of CO_2 to higher alcohols. *Science* **335**, 1596-1596 (2012).

90. Grousseau, E., Lu, J., Gorret, N., Guillouet, S.E. & Sinskey, A.J. Isopropanol production with engineered *Cupriavidus necator* as bioproduction platform. *Appl. Microbiol. Biotechnol.* **98**, 4277-4290 (2014).

91. Torella, J.P. et al. Efficient solar-to-fuels production from a hybrid microbial–water-splitting catalyst system. *PNAS* **112**, 2337-2342 (2015).

92. Müller, J. et al. Engineering of *Ralstonia eutropha* H16 for autotrophic and heterotrophic production of methyl ketones. *Appl. Environ. Microbiol.* **79**, 4433-4439 (2013).

93. Ewering, C., Heuser, F., Benölken, J.K., Brämer, C.O. & Steinbüchel, A. Metabolic engineering of strains of *Ralstonia eutropha* and *Pseudomonas putida* for biotechnological production of 2-methylcitric acid. *Metab. Eng.* **8**, 587-602 (2006).

94. Hoefel, T., Wittmann, E., Reinecke, L. & Weuster-Botz, D. Reaction engineering studies for the production of 2-hydroxyisobutyric acid with recombinant *Cupriavidus necator* H16. *Appl. Microbiol. Biotechnol.* **88**, 477-484 (2010).

95. Windhorst, C. & Gescher, J. Efficient biochemical production of acetoin from carbon dioxide using *Cupriavidus necator* H16. *Biotechnol. Biofuels* **12**, 163 (2019).

96. Pohlmann, A. et al. Genome sequence of the bioplastic-producing "Knallgas" bacterium *Ralstonia eutropha* H16. *Nat. Biotechnol.* **24**, 1257 (2006).

97. Cramm, R. Genomic view of energy metabolism in *Ralstonia eutropha* H16. *J. Mol. Microbiol. Biotechnol.* **16**, 38-52 (2009).

98. Johnson, B. & Stanier, R. Dissimilation of aromatic compounds by *Alcaligenes eutrophus*. *J. Bacteriol.* **107**, 468-475 (1971).

99. Bowien, B. & Kusian, B. Genetics and control of CO_2 assimilation in the chemoautotroph *Ralstonia eutropha*. *Arch. Microbiol.* **178**, 85-93 (2002).

100. Friedrich, C.G., Bowien, B. & Friedrich, B. Formate and oxalate metabolism in *Alcaligenes eutrophus*. *Microbiology* **115**, 185-192 (1979).

101. Oh, J.-I. & Bowien, B. Structural analysis of the *fds* operon encoding the NAD+-linked formate dehydrogenase of *Ralstonia eutropha*. *J. Biol. Chem.* **273**, 26349-26360 (1998).

102. Fukui, T., Ohsawa, K., Mifune, J., Orita, I. & Nakamura, S. Evaluation of promoters for gene expression in polyhydroxyalkanoate-producing *Cupriavidus necator* H16. *Appl. Microbiol. Biotechnol.* **89**, 1527-1536 (2011).

103. Gruber, S., Hagen, J., Schwab, H. & Koefinger, P. Versatile and stable vectors for efficient gene expression in *Ralstonia eutropha* H16. *J. Biotechnol.* **186**, 74-82 (2014).

104. Delamarre, S.C. & Batt, C.A. Comparative study of promoters for the production of polyhydroxyalkanoates in recombinant strains of *Wautersia eutropha*. *Appl. Microbiol. Biotechnol.* **71**, 668-679 (2006).

105. Johnson, A.O., Gonzalez-Villanueva, M., Tee, K.L. & Wong, T.S. An Engineered Constitutive Promoter Set with Broad Activity Range for *Cupriavidus necator* H16. *ACS Synth. Biol.* **7**, 1918-1928 (2018).

106. Alagesan, S. et al. Functional genetic elements for controlling gene expression in *Cupriavidus necator* H16. *Appl. Environ. Microbiol.* **84**, e00878-00818 (2018).

107. Lütte, S. et al. Autotrophic production of stable-isotope-labeled arginine in *Ralstonia eutropha* strain H16. *Appl. Environ. Microbiol.* **78**, 7884-7890 (2012).

108. Barnard, G.C., McCool, J.D., Wood, D.W. & Gerngross, T.U. Integrated recombinant protein expression and purification platform based on *Ralstonia eutropha*. *Appl. Environ. Microbiol.* **71**, 5735-5742 (2005).

109. Sydow, A. et al. Expanding the genetic tool box for *Cupriavidus necator* by a stabilized L-rhamnose inducible plasmid system. *J. Biotechnol.* **263**, 1-10 (2017).

110. Gruber, S. et al. Design of inducible expression vectors for improved protein production in *Ralstonia eutropha* H16 derived host strains. *J. Biotechnol.* **235**, 92-99 (2016).

111. Li, H. & Liao, J.C. A synthetic anhydrotetracycline-controllable gene expression system in *Ralstonia eutropha* H16. *ACS Synth. Biol.* **4**, 101-106 (2014).

112. Aboulnaga, E.A., Zou, H., Selmer, T. & Xian, M. Development of a plasmid-based, tunable, *tolC*-derived expression system for application in *Cupriavidus necator* H16. *J. Biotechnol.* **274**, 15-27 (2018).

113. Sambrook, J. & Russell, D.W. Molecular Cloning: A Laboratory Manual, Edn. 3rd. (Cold Spring Harbor Laboratory Press, New York; 2001).

114. Jensen, K.F. The *Escherichia coli* K-12" wild types" W3110 and MG1655 have an *rph* frameshift mutation that leads to pyrimidine starvation due to low *pyrE* expression levels. *J. Bacteriol.* **175**, 3401-3407 (1993).

115. Schlegel, H., Kaltwasser, H. & Gottschalk, G. Ein Submersverfahren zur Kultur wasserstoffoxydierender Bakterien: wachstumsphysiologische Untersuchungen. *Archiv für Mikrobiologie* **38**, 209-222 (1961).

116. Trüper, H.G. & Pfennig, N. Characterization and identification of the anoxygenic phototrophic bacteria. (Springer, Berlin; 1981).

117. Kovach, M., Phillips, R., Elzer, P. & Peterson, K. pBBR1MCS: a broad-host-range cloning vector. *BioTechniques* **16**, 800-802 (1994).

118. Ausubel, F. et al. Current protocols in molecular biology. (John Wiley & Sons, New York; 2003).

119. Sheng, L., Kovács, K., Winzer, K., Zhang, Y. & Minton, N.P. Development and implementation of rapid metabolic engineering tools for chemical and fuel production in *Geobacillus thermoglucosidasius* NCIMB 11955. *Biotechnol. Biofuels* **10**, 5 (2017).

120. Hanko, E.K., Minton, N.P. & Malys, N. Design, cloning and characterization of transcription factor-based inducible gene expression systems. *Methods Enzymol.* **621**, 153-169 (2019).

121. Mutalik, V.K. et al. Quantitative estimation of activity and quality for collections of functional genetic elements. *Nat. Methods* **10**, 347 (2013).

122. Sullivan, M.J. et al. Unusual regulation of a leaderless operon involved in the catabolism of dimethylsulfoniopropionate in *Rhodobacter sphaeroides*. *PLoS One* **6**, e15972 (2011).

123. Egan, S.M. & Schleif, R.F. A regulatory cascade in the induction of *rhaBAD*. *J. Mol. Biol.* **234**, 87-98 (1993).

124. Singsaas, E.L., Lerdau, M., Winter, K. & Sharkey, T.D. Isoprene increases thermotolerance of isoprene-emitting species. *Plant Physiol.* **115**, 1413-1420 (1997).

125. Vickers, C.E. et al. Isoprene synthesis protects transgenic tobacco plants from oxidative stress. *Plant, Cell Environ.* **32**, 520-531 (2009).

126. Miller, B., Oschinski, C. & Zimmer, W. First isolation of an isoprene synthase gene from poplar and successful expression of the gene in *Escherichia coli*. *Planta* **213**, 483-487 (2001).

127. Zhao, Y. et al. Biosynthesis of isoprene in *Escherichia coli* via methylerythritol phosphate (MEP) pathway. *Appl. Microbiol. Biotechnol.* **90**, 1915 (2011).

128. Silver, G.M. & Fall, R. Enzymatic synthesis of isoprene from dimethylallyl diphosphate in aspen leaf extracts. *Plant Physiol.* **97**, 1588-1591 (1991).

129. Silver, G.M. & Fall, R. Characterization of aspen isoprene synthase, an enzyme responsible for leaf isoprene emission to the atmosphere. *J. Biol. Chem.* **270**, 13010-13016 (1995).

130. Vickers, C.E. & Sabri, S. Isoprene. (Springer, Berlin; 2015).

131. Pfeiffer, D. & Jendrossek, D. Interaction between poly (3-hydroxybutyrate) granule-associated proteins as revealed by two-hybrid analysis and identification of a new phasin in *Ralstonia eutropha* H16. *Microbiology* **157**, 2795-2807 (2011).

132. Hoffmann, J. & Altenbuchner, J. Functional characterization of the mannitol promoter of *Pseudomonas fluorescens* DSM 50106 and its application for a mannitol-inducible expression system for *Pseudomonas putida* KT2440. *PloS One* **10**, e0133248 (2015).

133. Choi, Y.J. et al. Novel, versatile, and tightly regulated expression system for *Escherichia coli* strains. *Appl. Environ. Microbiol.* **76**, 5058-5066 (2010).

134. Campbell, R.E. et al. A monomeric red fluorescent protein. *PNAS* **99**, 7877-7882 (2002).

135. Stepanenko, O.V. et al. Comparative studies on the structure and stability of fluorescent proteins EGFP, zFP506, mRFP1,"dimer2", and DsRed1. *Biochemistry* **43**, 14913-14923 (2004).

136. Klumpp, S., Zhang, Z. & Hwa, T. Growth rate-dependent global effects on gene expression in bacteria. *Cell* **139**, 1366-1375 (2009).

137. Heap, J.T., Pennington, O.J., Cartman, S.T. & Minton, N.P. A modular system for *Clostridium* shuttle plasmids. *J. Microbiol. Methods* **78**, 79-85 (2009).

138. Hahn, S. & Schleif, R. In vivo regulation of the *Escherichia coli araC* promoter. *J. Bacteriol.* **155**, 593-600 (1983).

139. Miertens, N., Remaut, E. & Fiers, W. Increased stability of phage T7*g10* mRNA is mediated by either a 5'-or a 3'-terminal stem-loop structure. *Biol. Chem.* **377**, 811-818 (1996).

140. Frädrich, C. et al. The transcription factor AlsR binds and regulates the promoter of the *alsSD* operon responsible for acetoin formation in *Bacillus subtilis*. *J. Bacteriol.* **194**, 1100-1112 (2012).

141. Lee, S.K. & Keasling, J.D. A propionate-inducible expression system for enteric bacteria. *Appl. Environ. Microbiol.* **71**, 6856-6862 (2005).

142. Tobin, J. & Schleif, R. Positive regulation of the *Escherichia coli* L-rhamnose operon is mediated by the products of tandemly repeated regulatory genes. *J. Mol. Biol.* **196**, 789-799 (1987).

143. Eaton, R.W. *p*-Cymene catabolic pathway in *Pseudomonas putida* F1: cloning and characterization of DNA encoding conversion of *p*-cymene to *p*-cumate. *J. Bacteriol.* **179**, 3171-3180 (1997).

144. Schnider-Keel, U. et al. Autoinduction of 2, 4-diacetylphloroglucinol biosynthesis in the biocontrol agent *Pseudomonas fluorescens* CHA0 and repression by the bacterial metabolites salicylate and pyoluteorin. *J. Bacteriol.* **182**, 1215-1225 (2000).

145. Rygus, T., Scheler, A., Allmansberger, R. & Hillen, W. Molecular cloning, structure, promoters and regulatory elements for transcription of the *Bacillus megaterium* encoded regulon for xylose utilization. *Arch. Microbiol.* **155**, 535-542 (1991).

146. Balzer, D., Ziegelin, G., Pansegrau, W., Kruft, V. & Lanka, E. KorB protein of promiscuous plasmid RP4 recognizes inverted sequence repetitions in regions essential for conjugative plasmid transfer. *Nucleic Acids Res.* **20**, 1851-1858 (1992).

147. Lee, S.K. & Keasling, J.D. A Salmonella-based, propionate-inducible, expression system for *Salmonella enterica*. *Gene* **377**, 6-11 (2006).

148. Cardona, S.T. & Valvano, M.A. An expression vector containing a rhamnose-inducible promoter provides tightly regulated gene expression in *Burkholderia cenocepacia*. *Plasmid* **54**, 219-228 (2005).

149. Lewis, P.J. & Marston, A.L. GFP vectors for controlled expression and dual labelling of protein fusions in Bacillus subtilis. *Gene* **227**, 101-109 (1999).

150. Peccoud, J. et al. Essential information for synthetic DNA sequences. *Nat. Biotechnol.* **29**, 22 (2011).

151. Kelly, C.n.L. et al. Synthetic chemical inducers and genetic decoupling enable orthogonal control of the *rhaBAD* promoter. *ACS Synth. Biol.* **5**, 1136-1145 (2016).

152. Hanko, E.K. et al. A genome-wide approach for identification and characterisation of metabolite-inducible systems. *Nat. Commun.* **11**, 1-14 (2020).

153. Khalil, A.S. & Collins, J.J. Synthetic biology: applications come of age. *Nat. Rev. Genet.* **11**, 367 (2010).

154. D'ambrosio, V. & Jensen, M.K. Lighting up yeast cell factories by transcription factor-based biosensors. *FEMS Yeast Res.* **17** (2017).

155. Chen, Y. et al. Tuning the dynamic range of bacterial promoters regulated by ligand-inducible transcription factors. *Nat. Commun.* **9**, 64 (2018).

156. Wan, X., Marsafari, M. & Xu, P. Engineering metabolite-responsive transcriptional factors to sense small molecules in eukaryotes: current state and perspectives. *Microb. Cell Fact.* **18**, 61 (2019).

157. Chang, H.J., Voyvodic, P.L., Zúñiga, A. & Bonnet, J. Microbially derived biosensors for diagnosis, monitoring and epidemiology. *Microb. Biotechnol.* **10**, 1031-1035 (2017).

158. Van Der Meer, J.R. & Belkin, S. Where microbiology meets microengineering: design and applications of reporter bacteria. *Nat. Rev. Microbiol.* **8**, 511 (2010).

159. Wen, K.Y. et al. A cell-free biosensor for detecting quorum sensing molecules in *P. aeruginosa*-infected respiratory samples. *ACS Synth. Biol.* **6**, 2293-2301 (2017).

160. Mahr, R. & Frunzke, J. Transcription factor-based biosensors in biotechnology: current state and future prospects. *Appl. Microbiol. Biotechnol.* **100**, 79-90 (2016).

161. Greenwald, E.C., Mehta, S. & Zhang, J. Genetically encoded fluorescent biosensors illuminate the spatiotemporal regulation of signaling networks. *Chem. Rev.* **118**, 11707-11794 (2018).

162. van Hijum, S.A., Medema, M.H. & Kuipers, O.P. Mechanisms and evolution of control logic in prokaryotic transcriptional regulation. *Microbiol. Mol. Biol. Rev.* **73**, 481-509 (2009).

163. Lefebre, M.D. & Valvano, M.A. Construction and evaluation of plasmid vectors optimized for constitutive and regulated gene expression in *Burkholderia cepacia* complex isolates. *Appl. Environ. Microbiol.* **68**, 5956-5964 (2002).

164. Oliver, P., Peralta-Gil, M., Tabche, M.-L. & Merino, E. Molecular and structural considerations of TF-DNA binding for the generation of biologically meaningful and accurate phylogenetic footprinting analysis: the LysR-type transcriptional regulator family as a study model. *BMC Genomics* **17**, 686 (2016).

165. Benson, D.A. et al. GenBank. *Nucleic Acids Res.* **42**, D32-D37 (2013).

166. Jeske, L., Placzek, S., Schomburg, I., Chang, A. & Schomburg, D. BRENDA in 2019: a European ELIXIR core data resource. *Nucleic Acids Res.* **47**, D542-D549 (2018).

167. Knoten, C.A., Hudson, L.L., Coleman, J.P., Farrow, J.M. & Pesci, E.C. KynR, a Lrp/AsnC-type transcriptional regulator, directly controls the kynurenine pathway in *Pseudomonas aeruginosa*. *J. Bacteriol.* **193**, 6567-6575 (2011).

168. Hanahan, D. Studies on transformation of *Escherichia coli* with plasmids. *J. Mol. Biol.* **166**, 557-580 (1983).

169. Hirakawa, H., Hirakawa, Y., Greenberg, E.P. & Harwood, C.S. BadR and BadM proteins transcriptionally regulate two operons needed for anaerobic benzoate degradation by *Rhodopseudomonas palustris*. *Appl. Environ. Microbiol.* **81**, 4253-4262 (2015).

170. Uchiyama, T. & Miyazaki, K. Product-induced gene expression, a product-responsive reporter assay used to screen metagenomic libraries for enzyme-encoding genes. *Appl. Environ. Microbiol.* **76**, 7029-7035 (2010).

171. Jha, R.K., Chakraborti, S., Kern, T.L., Fox, D.T. & Strauss, C.E. Rosetta comparative modeling for library design: engineering alternative inducer specificity in a transcription factor. *Proteins: Struct. Funct. Bioinform.* **83**, 1327-1340 (2015).

172. Könst, P.M., Franssen, M.C., Scott, E.L. & Sanders, J.P. A study on the applicability of L-aspartate α-decarboxylase in the biobased production of nitrogen containing chemicals. *Green Chem.* **11**, 1646-1652 (2009).

173. Song, C.W., Lee, J., Ko, Y.-S. & Lee, S.Y. Metabolic engineering of *Escherichia coli* for the production of 3-aminopropionic acid. *Metab. Eng.* **30**, 121-129 (2015).

174. Rokicki, J. et al. Daily carnosine and anserine supplementation alters verbal episodic memory and resting state network connectivity in healthy elderly adults. *Front. Aging Neurosci.* **7**, 219 (2015).

175. Szcześniak, D., Budzeń, S., Kopeć, W. & Rymaszewska, J. Anserine and carnosine supplementation in the elderly: Effects on cognitive functioning and physical capacity. *Arch. Geronto.l Geriatr.* **59**, 485-490 (2014).

176. Schneider, F., Krämer, R. & Burkovski, A. Identification and characterization of the main β-alanine uptake system in *Escherichia coli*. *Appl. Microbiol. Biotechnol.* **65**, 576-582 (2004).

177. Belitsky, B.R. *Bacillus subtilis* GabR, a protein with DNA-binding and aminotransferase domains, is a PLP-dependent transcriptional regulator. *J. Mol. Biol.* **340**, 655-664 (2004).

178. Morales, G. et al. The *Pseudomonas putida* Crc global regulator controls the expression of genes from several chromosomal catabolic pathways for aromatic compounds. *J. Bacteriol.* **186**, 1337-1344 (2004).

179. Cebolla, A., Sousa, C. & de Lorenzo, V. Effector specificity mutants of the transcriptional activator NahR of naphthalene degrading *Pseudomonas* define protein sites involved in binding of aromatic inducers. *J. Biol. Chem.* **272**, 3986-3992 (1997).

180. Khlebnikov, A., Risa, Ø., Skaug, T., Carrier, T.A. & Keasling, J. Regulatable arabinose-inducible gene expression system with consistent control in all cells of a culture. *J. Bacteriol.* **182**, 7029-7034 (2000).

181. Walker, L.M., Li, B., Niks, D., Hille, R. & Elliott, S.J. Deconvolution of reduction potentials of formate dehydrogenase from *Cupriavidus necator*. *J. Biol. Inorg. Chem.* **24**, 889-898 (2019).

182. Volodina, E., Schürmann, M., Lindenkamp, N. & Steinbüchel, A. Characterization of propionate CoA-transferase from *Ralstonia eutropha* H16. *Appl. Microbiol. Biotechnol.* **98**, 3579-3589 (2014).

183. Peplinski, K., Ehrenreich, A., Döring, C., Bömeke, M. & Steinbüchel, A. Investigations on the microbial catabolism of the organic sulfur compounds TDP and DTDP in *Ralstonia eutropha* H16 employing DNA microarrays. *Appl. Microbiol. Biotechnol.* **88**, 1145-1159 (2010).

184. Sivapragasam, S. & Grove, A. *Streptomyces coelicolor* XdhR is a direct target of (p) ppGpp that controls expression of genes encoding xanthine dehydrogenase to promote purine salvage. *Mol. Microbiol.* **100**, 701-718 (2016).

185. Hillerich, B. & Westpheling, J. A new TetR family transcriptional regulator required for morphogenesis in *Streptomyces coelicolor*. *J. Bacteriol.* **190**, 61-67 (2008).

186. Mannan, A.A., Liu, D., Zhang, F. & Oyarzún, D.A. Fundamental design principles for transcription-factor-based metabolite biosensors. *ACS Synth. Biol.* **6**, 1851-1859 (2017).

187. Lutz, R. & Bujard, H. Independent and tight regulation of transcriptional units in *Escherichia coli* via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements. *Nucleic Acids Res.* **25**, 1203-1210 (1997).

188. Kaczmarczyk, A., Vorholt, J.A. & Francez-Charlot, A. Cumate-inducible gene expression system for Sphingomonads and other *Alphaproteobacteria*. *Appl. Environ. Microbiol.* **79**, 6795-6802 (2013).

189. Hanko, E.K., Minton, N.P. & Malys, N. Characterisation of a 3-hydroxypropionic acid-inducible system from *Pseudomonas putida* for orthogonal gene expression control in *Escherichia coli* and *Cupriavidus necator*. *Sci. Rep.* **7**, 1724 (2017).

190. Latif, H., Zeidan, A.A., Nielsen, A.T. & Zengler, K. Trash to treasure: production of biofuels and commodity chemicals via syngas fermenting microorganisms. *Curr. Opin. Biotechnol.* **27**, 79-87 (2014).

191. Clomburg, J.M., Crumbley, A.M. & Gonzalez, R. Industrial biomanufacturing: the future of chemical production. *Science* **355**, aag0804 (2017).

192. Erickson, B. & Winters, P. Perspective on opportunities in industrial biotechnology in renewable chemicals. *Biotechnol. J.* **7**, 176-185 (2012).

193. Choi, S., Song, C.W., Shin, J.H. & Lee, S.Y. Biorefineries for the production of top building block chemicals and their derivatives. *Metab. Eng.* **28**, 223-239 (2015).

194. Becker, J., Lange, A., Fabarius, J. & Wittmann, C. Top value platform chemicals: bio-based production of organic acids. *Curr. Opin. Biotechnol.* **36**, 168-175 (2015).

195. Valdehuesa, K.N.G. et al. Recent advances in the metabolic engineering of microorganisms for the production of 3-hydroxypropionic acid as C3 platform chemical. *Appl. Microbiol. Biotechnol.* **97**, 3309-3321 (2013).

196. Andreeßen, B., Taylor, N. & Steinbüchel, A. Poly (3-hydroxypropionate): a promising alternative to fossil fuel-based materials. *Appl. Environ. Microbiol.* **80**, 6574-6582 (2014).

197. Rathnasingh, C., Raj, S.M., Jo, J.E. & Park, S. Development and evaluation of efficient recombinant *Escherichia coli* strains for the production of 3-hydroxypropionic acid from glycerol. *Biotechnol. Bioeng.* **104**, 729-739 (2009).

198. Zhou, S., Catherine, C., Rathnasingh, C., Somasundar, A. & Park, S. Production of 3-hydroxypropionic acid from glycerol by recombinant *Pseudomonas denitrificans*. *Biotechnol. Bioeng.* **110**, 3177-3187 (2013).

199. Burgé, G. et al. Diversity of *Lactobacillus reuteri* strains in converting glycerol into 3-hydroxypropionic acid. *Appl. Biochem. Biotechnol.* **177**, 923-939 (2015).

200. Chen, Y., Bao, J., Kim, I.-K., Siewers, V. & Nielsen, J. Coupled incremental precursor and co-factor supply improves 3-hydroxypropionic acid production in *Saccharomyces cerevisiae*. *Metab. Eng.* **22**, 104-109 (2014).

201. Lan, E.I. et al. Metabolic engineering of cyanobacteria for photosynthetic 3-hydroxypropionic acid production from CO_2 using *Synechococcus elongatus* PCC 7942. *Metab. Eng.* **31**, 163-170 (2015).

202. Zhao, L., Lin, J., Wang, H., Xie, J. & Wei, D. Development of a two-step process for production of 3-hydroxypropionic acid from glycerol using *Klebsiella pneumoniae* and Gluconobacter oxydans. *Bioprocess Biosystems Eng.* **38**, 2487-2495 (2015).

203. Wang, Y. et al. Biosynthesis of platform chemical 3-hydroxypropionic acid (3-HP) directly from CO_2 in cyanobacterium *Synechocystis* sp. PCC 6803. *Metab. Eng.* **34**, 60-70 (2016).

204. Chen, Z. et al. Metabolic engineering of *Corynebacterium glutamicum* for the production of 3-hydroxypropionic acid from glucose and xylose. *Metab. Eng.* **39**, 151-158 (2017).

205. Borodina, I. et al. Establishing a synthetic pathway for high-level production of 3-hydroxypropionic acid in *Saccharomyces cerevisiae* via β-alanine. *Metab. Eng.* **27**, 57-64 (2015).

206. Kim, K., Kim, S.-K., Park, Y.-C. & Seo, J.-H. Enhanced production of 3-hydroxypropionic acid from glycerol by modulation of glycerol metabolism in recombinant *Escherichia coli*. *Bioresour. Technol.* **156**, 170-175 (2014).

207. Rathnasingh, C. et al. Production of 3-hydroxypropionic acid via malonyl-CoA pathway using recombinant *Escherichia coli* strains. *J. Biotechnol.* **157**, 633-640 (2012).

208. Gokarn, R.R. et al. 3-Hydroxypropionic acid and other organic compounds. US Patent 7,186,541 B182 (2007).

209. Luo, H. et al. Production of 3-hydroxypropionic acid via the propionyl-CoA pathway using recombinant *Escherichia coli* strains. *PloS One* **11**, e0156286 (2016).

210. Li, Y., Wang, X., Ge, X. & Tian, P. High production of 3-hydroxypropionic acid in *Klebsiella pneumoniae* by systematic optimization of glycerol metabolism. *Sci. Rep.* **6**, 26932 (2016).

211. Chen, Y. & Nielsen, J. Biobased organic acids production by metabolically engineered microorganisms. *Curr. Opin. Biotechnol.* **37**, 165-172 (2016).

212. Zhou, S., Ainala, S.K., Seol, E., Nguyen, T.T. & Park, S. Inducible gene expression system by 3-hydroxypropionic acid. *Biotechnol. Biofuels* **8**, 169 (2015).

213. MacLean, A.M., Anstey, M.I. & Finan, T.M. Binding site determinants for the LysR-type transcriptional regulator PcaQ in the legume endosymbiont *Sinorhizobium meliloti*. *J. Bacteriol.* **190**, 1237-1246 (2008).

214. Porrúa, O. et al. An A-tract at the AtzR binding site assists DNA binding, inducer-dependent repositioning and transcriptional activation of the P*atzDEF* promoter. *Mol. Microbiol.* **90**, 72-87 (2013).

215. Reen, F.J., Barret, M., Fargier, E., O'Muinneacháin, M. & O'Gara, F. Molecular evolution of LysR-type transcriptional regulation in *Pseudomonas aeruginosa*. *Mol. Phylogen. Evol.* **66**, 1041-1049 (2013).

216. Michener, J.K., Thodey, K., Liang, J.C. & Smolke, C.D. Applications of genetically-encoded biosensors for the construction and control of biosynthetic pathways. *Metab. Eng.* **14**, 212-222 (2012).

217. Rogers, J.K. & Church, G.M. Genetically encoded sensors enable real-time observation of metabolite production. *PNAS* **113**, 2388-2393 (2016).

218. Zhang, J., Jensen, M.K. & Keasling, J.D. Development of biosensors and their application in metabolic engineering. *Curr. Opin. Chem. Biol.* **28**, 1-8 (2015).

219. Dietrich, J.A., Shis, D.L., Alikhani, A. & Keasling, J.D. Transcription factor-based screens and synthetic selections for microbial small-molecule biosynthesis. *ACS Synth. Biol.* **2**, 47-58 (2012).

220. Eggeling, L., Bott, M. & Marienhagen, J. Novel screening methods—biosensors. *Curr. Opin. Biotechnol.* **35**, 30-36 (2015).

221. Qian, S. & Cirino, P.C. Using metabolite-responsive gene regulators to improve microbial biosynthesis. *Curr. Opin. Chem. Eng.* **14**, 93-102 (2016).

222. Fang, M. et al. Intermediate-sensor assisted push–pull strategy and its application in heterologous deoxyviolacein production in *Escherichia coli*. *Metab. Eng.* **33**, 41-51 (2016).

223. Dos Santos, V.M., Heim, S., Moore, E., Strätz, M. & Timmis, K. Insights into the genomic basis of niche specificity of *Pseudomonas putida* KT2440. *Environ. Microbiol.* **6**, 1264-1286 (2004).

224. Sievers, F. et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7** (2011).

225. Goujon, M. et al. A new bioinformatics analysis tools framework at EMBL–EBI. *Nucleic Acids Res.* **38**, W695-W699 (2010).

226. Crooks, G.E., Hon, G., Chandonia, J.-M. & Brenner, S.E. WebLogo: a sequence logo generator. *Genome Res.* **14**, 1188-1190 (2004).

227. Solovyev, V. & Salamov, A. Automatic annotation of microbial genomes and metagenomic sequences. (Nova Science Publishers, New York; 2011).

228. Reese, M.G. Application of a time-delay neural network to promoter annotation in the *Drosophila melanogaster* genome. *Comput. Chem.* **26**, 51-56 (2001).

229. Jørgensen, C. & Dandanell, G. Isolation and Characterization of Mutations in theEscherichia coli Regulatory Protein XapR. *J. Bacteriol.* **181**, 4397-4403 (1999).

230. Pande, S.G. et al. Lactobacillus brevis responds to flavonoids through KaeR, a LysR-type of transcriptional regulator. *Mol. Microbiol.* **81**, 1623-1639 (2011).

231. Hughes, J.R. et al. Annotation of cis-regulatory elements by identification, subclassification, and functional assessment of multispecies conserved sequences. *PNAS* **102**, 9830-9835 (2005).

232. Zuker, M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* **31**, 3406-3415 (2003).

233. Parsek, M., Shinabarger, D., Rothmel, R. & Chakrabarty, A. Roles of CatR and *cis,cis*-muconate in activation of the *catBC* operon, which is involved in benzoate degradation in *Pseudomonas putida*. *J. Bacteriol.* **174**, 7798-7806 (1992).

234. Jiang, X., Meng, X. & Xian, M. Biosynthetic pathways for 3-hydroxypropionic acid production. *Appl. Microbiol. Biotechnol.* **82**, 995-1003 (2009).

235. Hanko, E.K., Minton, N.P. & Malys, N. A transcription factor-based biosensor for detection of itaconic acid. *ACS Synth. Biol.* **7**, 1436-1446 (2018).

236. Okabe, M., Lies, D., Kanamasa, S. & Park, E.Y. Biotechnological production of itaconic acid and its biosynthesis in *Aspergillus terreus*. *Appl. Microbiol. Biotechnol.* **84**, 597-606 (2009).

237. Werpy, T. et al. Top value added chemicals from biomass. Volume 1-Results of screening for potential candidates from sugars and synthesis gas. (Department of Energy, Washington DC, 2004).

238. Bentley, R. & Thiessen, C.P. Biosynthesis of itaconic acid in *Aspergillus terreus* I. Tracer studies with C^{14}-labeled substrates. *J. Biol. Chem.* **226**, 673-687 (1957).

239. Haskins, R., Thorn, J. & Boothroyd, B. Biochemistry of the Ustilaginales: XI. Metabolic products of *Ustilago zeae* in submerged culture. *Can. J. Microbiol.* **1**, 749-756 (1955).

240. Tabuchi, T., Sugisawa, T., Ishidori, T., Nakahara, T. & Sugiyama, J. Itaconic acid fermentation by a yeast belonging to the genus *Candida*. *Agric. Biol. Chem.* **45**, 475-479 (1981).

241. Strelko, C.L. et al. Itaconic acid is a mammalian metabolite induced during macrophage activation. *J. Am. Chem. Soc.* **133**, 16386-16389 (2011).

242. Cordes, T., Michelucci, A. & Hiller, K. Itaconic acid: the surprising role of an industrial compound as a mammalian antimicrobial metabolite. *Annu. Rev. Nutr.* **35**, 451-473 (2015).

243. Geiser, E. et al. *Ustilago maydis* produces itaconic acid via the unusual intermediate *trans*-aconitate. *Microb. Biotechnol.* **9**, 116-126 (2016).

244. Li, A. et al. A clone-based transcriptomics approach for the identification of genes relevant for itaconic acid production in *Aspergillus*. *Fungal Genet. Biol.* **48**, 602-611 (2011).

245. Kuenz, A., Gallenmüller, Y., Willke, T. & Vorlop, K.-D. Microbial production of itaconic acid: developing a stable platform for high product concentrations. *Appl. Microbiol. Biotechnol.* **96**, 1209-1216 (2012).

246. Hevekerl, A., Kuenz, A. & Vorlop, K.-D. Filamentous fungi in microtiter plates—an easy way to optimize itaconic acid production with *Aspergillus terreus*. *Appl. Microbiol. Biotechnol.* **98**, 6983-6989 (2014).

247. Krull, S., Hevekerl, A., Kuenz, A. & Prüße, U. Process development of itaconic acid production by a natural wild type strain of *Aspergillus terreus* to reach industrially relevant final titers. *Appl. Microbiol. Biotechnol.* **101**, 4063-4072 (2017).

248. Kanamasa, S., Dwiarti, L., Okabe, M. & Park, E.Y. Cloning and functional characterization of the *cis*-aconitic acid decarboxylase (CAD) gene from *Aspergillus terreus*. *Appl. Microbiol. Biotechnol.* **80**, 223-229 (2008).

249. Levinson, W.E., Kurtzman, C.P. & Kuo, T.M. Production of itaconic acid by *Pseudozyma antarctica* NRRL Y-7808 under nitrogen-limited growth conditions. *Enzyme Microb. Technol.* **39**, 824-827 (2006).

250. Otten, A., Brocker, M. & Bott, M. Metabolic engineering of *Corynebacterium glutamicum* for the production of itaconate. *Metab. Eng.* **30**, 156-165 (2015).

251. Harder, B.-J., Bettenbrock, K. & Klamt, S. Model-based metabolic engineering enables high yield itaconic acid production by *Escherichia coli*. *Metab. Eng.* **38**, 29-37 (2016).

252. Blazeck, J. et al. Metabolic engineering of *Saccharomyces cerevisiae* for itaconic acid production. *Appl. Microbiol. Biotechnol.* **98**, 8155-8164 (2014).

253. Blazeck, J. et al. Metabolic engineering of *Yarrowia lipolytica* for itaconic acid production. *Metab. Eng.* **32**, 66-73 (2015).

254. Guevarra, E.D. & Tabuchi, T. Accumulation of itaconic, 2-hydroxyparaconic, itatartaric, and malic acids by strains of the genus *Ustilago*. *Agric. Biol. Chem.* **54**, 2353-2358 (1990).

255. Zambanini, T. et al. Efficient itaconic acid production from glycerol with *Ustilago vetiveriae* TZ1. *Biotechnol. Biofuels* **10**, 131 (2017).

256. Raman, S., Rogers, J.K., Taylor, N.D. & Church, G.M. Evolution-guided optimization of biosynthetic pathways. *PNAS* **111**, 17803-17808 (2014).

257. Klement, T. & Büchs, J. Itaconic acid–a biotechnological process in change. *Bioresour. Technol.* **135**, 422-431 (2013).

258. Duportet, X., Aggio, R.B.M., Carneiro, S. & Villas-Bôas, S.G. The biological interpretation of metabolomic data can be misled by the extraction method used. *Metabolomics* **8**, 410-421 (2012).

259. Martinez-Salas, E., Martin, J. & Vicente, M. Relationship of *Escherichia coli* density to growth rate and cell age. *J. Bacteriol.* **147**, 97-100 (1981).

260. Cooper, R. & Kornberg, H. The utilization of itaconate by *Pseudomonas* sp. *Biochem. J.* **91**, 82 (1964).

261. Martin, W.R., Frigan, F. & Bergman, E.H. Noninductive metabolism of itaconic acid by *Pseudomonas* and *Salmonella* species. *J. Bacteriol.* **82**, 905-908 (1961).

262. Sasikaran, J., Ziemski, M., Zadora, P.K., Fleig, A. & Berg, I.A. Bacterial itaconate degradation promotes pathogenicity. *Nat. Chem. Biol.* **10**, 371 (2014).

263. Shi, L. et al. Proteomic analysis of *Salmonella enterica* serovar Typhimurium isolated from RAW 264.7 macrophages: Identification of a novel protein that contributes to the replication of serovar Typhimurium inside macrophages. *J. Biol. Chem.* **281**, 29131-29140 (2006).

264. Michelucci, A. et al. Immune-responsive gene 1 protein links metabolism to immunity by catalyzing itaconic acid production. *PNAS* **110**, 7820-7825 (2013).

265. Vuoristo, K.S. et al. Metabolic engineering of itaconate production in *Escherichia coli*. *Appl. Microbiol. Biotechnol.* **99**, 221-228 (2015).

266. Blazeck, J. & Alper, H.S. Promoter engineering: recent advances in controlling transcription at the most fundamental level. *Biotechnol. J.* **8**, 46-58 (2013).

267. Taylor, N.D. et al. Engineering an allosteric transcription factor to respond to new ligands. *Nat. Methods* **13**, 177 (2016).

268. Mitchell, R.J. & Gu, M.B. Construction and evaluation of *nagR-nagAa::lux* fusion strains in biosensing for salicylic acid derivatives. *Appl. Biochem. Biotechnol.* **120**, 183-197 (2005).

269. Zhang, D. et al. Characterization and modeling of transcriptional cross-regulation in *Acinetobacter baylyi* ADP1. *ACS Synth. Biol.* **1**, 274-283 (2012).

270. Xue, H. et al. Design, construction, and characterization of a set of biosensors for aromatic compounds. *ACS Synth. Biol.* **3**, 1011-1014 (2014).

271. Calero, P., Jensen, S.I. & Nielsen, A.T. Broad-host-range ProUSER vectors enable fast characterization of inducible promoters and optimization of *p*-coumaric acid production in *Pseudomonas putida* KT2440. *ACS Synth. Biol.* **5**, 741-753 (2016).

272. Nguyen, N.H., Kim, J.-R. & Park, S. Development of Biosensor for 3-Hydroxypropionic Acid. *Biotechnol. Bioprocess Eng.* **24**, 109-118 (2019).

273. Liu, Y. et al. Biosensor-based evolution and elucidation of a biosynthetic pathway in *Escherichia coli*. *ACS Synth. Biol.* **6**, 837-848 (2017).

274. Chou, H.H. & Keasling, J.D. Programming adaptive control to evolve increased metabolite production. *Nat. Commun.* **4**, 1-8 (2013).

275. Jha, R.K. et al. A protocatechuate biosensor for *Pseudomonas putida* KT2440 via promoter and protein evolution. *Metab. Eng. Commun.* **6**, 33-38 (2018).

276. Jha, R.K., Kern, T.L., Fox, D.T. & M. Strauss, C.E. Engineering an *Acinetobacter* regulon for biosensing and high-throughput enzyme screening in *E. coli* via flow cytometry. *Nucleic Acids Res.* **42**, 8150-8160 (2014).

9 Appendix

9.1 Construction of plasmids

pEH002 was constructed by restriction enzyme-based cloning. Oligonucleotide primers EH007_f and EH008_r were used to amplify the L-rhamnose-inducible system RhaSR/P_{rhaBAD} from pJOE7784.1, and cloned into pEH006 by AatII and XbaI restriction sites.

pEH002-ispS was constructed by restriction enzyme-based cloning. Gene *ispS* was cloned from pBBR1MCS-2-RBS_1-ispS into pEH002 by NdeI and BamHI restriction sites.

pEH003 was constructed by restriction enzyme-based cloning. Oligonucleotide primers EH009_f and EH010_r were used to amplify the D-mannitol-inducible system MtlR/P_{mtlE} from pJOE7771.1, and cloned into pEH006 by AatII and XbaI restriction sites.

pEH005 served as the backbone for assembly of negatively regulated inducible systems. It was constructed by using the NEBuilder Hifi DNA assembly method. Oligonucleotide primers EH025_f and EH026_r, EH0027_f and EH028_r were used to amplify the vector backbone and *tetR* from pEH006 and pJOE7801.1, respectively. Primer overhangs were designed to contain P_{lac} and P_{tetA}.

pEH006 served as the backbone for assembly of positively regulated inducible systems. It was constructed by using the NEBuilder Hifi DNA assembly method. Oligonucleotide primers EH001_f and EH002_r, EH0003_f and EH004_r, EH005_f and EH006_r were used to amplify the replication origin and the chloramphenicol resistance marker, the *rfp* reporter gene, and the L-arabinose-inducible system from pBBR1MCS-2-PphaC-eyfp-c1 and pKTrfp, respectively. The primer overhangs were designed to contain PmeI, FseI, AscI, and SbfI restriction sites to allow for modular assembly as well as AatII and NdeI restriction sites to be able to replace the inducible system. Since the transcriptional start site (+1) of the L-arabinose-inducible promoter was known, the sequence between +1 and the translational start site was replaced by a T7*g10* mRNA stem-loop structure and a strong RBS, a 27-nucleotide upstream sequence of the bacteriophage T7 gene *10*.

pEH006-ispS was constructed by restriction enzyme-based cloning. Gene *ispS* was cloned from pBBR1MCS-2-RBS_1-ispS into pEH006 by NdeI and BamHI restriction sites.

pEH006E was constructed by restriction enzyme-based cloning. Oligonucleotide primers EH013_f and EH148_r were used to amplify the *rfp* reporter gene from pKTrfp, and cloned into pEH006 by AatII and SbfI restriction sites.

pEH007 was constructed by restriction enzyme-based cloning. Oligonucleotide primers EH017_f and EH018_r were used to amplify the putative 3-HP-inducible system MmsR/P_{mmsA} from *P. putida* KT2440 genomic DNA, and cloned into pEH006 by AatII and NdeI restriction sites.

pEH008 was constructed by restriction enzyme-based cloning. Oligonucleotide primers EH019_f and EH020_r were used to amplify the putative 3-HP-inducible system HpdR/P_{hpdH} from *P. putida* KT2440 genomic DNA, and cloned into pEH006 by AatII and NdeI restriction sites.

pEH009 was constructed by restriction enzyme-based cloning. Oligonucleotide primers EH021_f and EH022_r were used to amplify the putative 3-HP-inducible system AraC/P_{acaD} from *C. necator* H16 genomic DNA, and cloned into pEH006 by AatII and NdeI restriction sites.

pEH010 was constructed by restriction enzyme-based cloning. Oligonucleotide primers EH023_f and EH024_r were used to amplify the putative 3-HP-inducible system HpdR/P_{mmsA1} from *C. necator* H16 genomic DNA, and cloned into pEH006 by AatII and NdeI restriction sites.

pEH015 was constructed by restriction enzyme-based cloning. Oligonucleotide primers EH041_f and EH042_r were used to amplify the propionate-inducible system PrpR/$P_{prpBCDE}$ from *E. coli* MG1655 genomic DNA, and cloned into pEH006 by AatII and XbaI restriction sites.

pEH016 was constructed by restriction enzyme-based cloning. Oligonucleotide primers EH043_f and EH044_r were used to amplify the acetate-inducible system AlsR/P_{alsSD} from *B. subtilis* 168 genomic DNA, and cloned into pEH006 by AatII and XbaI restriction sites.

pEH020 was constructed by using the NEBuilder Hifi DNA assembly method. Oligonucleotide primers EH048_f and EH051_r, EH056_f and EH055_r were used to amplify the vector backbone for negatively regulated inducible systems and the acrylate-inducible system AcuR/P_{acuRI} from pEH005 and *R. sphaeroides* 2.4.1 genomic DNA, respectively.

pEH020-ispS was constructed by restriction enzyme-based cloning. Gene *ispS* was cloned from pBBR1MCS-2-RBS_1-ispS into pEH020 by NdeI and BamHI restriction sites.

pEH022 was constructed by restriction enzyme-based cloning. Oligonucleotide primers EH059_f and EH060_r were used to amplify the putative 3-HP-inducible promoter P_{hpdH} from *P. putida* KT2440 genomic DNA, and cloned into pEH006 by AatII and NdeI restriction sites.

pEH034 was constructed by restriction enzyme-based cloning. Oligonucleotide primers EH096_f and EH095_r were used to amplify the putative 3-HP-inducible promoter P_{mmsA} from *P. putida* KT2440 genomic DNA, and cloned into pEH006 by AatII and NdeI restriction sites.

pEH035 was constructed by restriction enzyme-based cloning. Oligonucleotide primers EH097_r and EH098_f were used to amplify the putative 3-hydroxypropanoate-inducible promoter from *C. necator* H16 genomic DNA, and cloned into pEH006 by AatII and NdeI restriction sites.

pEH036 (*hpdH*-118::*rfp*) was constructed by using the NEBuilder Hifi DNA assembly method. Oligonucleotide primers EH061_f and EH099_r, EH100_f and EH062_r were used to amplify the replication origin, the chloramphenicol resistance marker and *hpdR*, as well as the truncated *hpdR*/*hpdH* intergenic region and the *rfp* reporter gene from pEH008. The bioinformatically identified mRNA stem-loop structure sequence within the *hpdR*/*hpdH* intergenic region was replaced by an AvrII restriction site to allow consecutive promoter modifications.

pEH038 was constructed by using the NEBuilder Hifi DNA assembly method. Oligonucleotide primers EH048_f and EH012_r, EH011_f and EH109_r, EH106_f and EH105_r, EH108_f and EH107_r were used to amplify the vector backbone for negatively regulated inducible systems and the xylose-inducible sytem XylR/P_{xylAB} from pEH005 and *Bacillus megaterium* DSM319 genomic DNA, respectively.

pEH040 was constructed by using the NEBuilder Hifi DNA assembly method. Oligonucleotide primers EH048_f and EH012_r, EH011_f and EH109_r, EH112_f and EH111_r, EH113_f and EH114_r were used to amplify the vector backbone for negatively regulated inducible systems, the transcriptional regulator *cymR*, and the cumate-inducible promoter P_{cmt} from pEH005 and pNEW, respectively.

pEH040-ispS was constructed by restriction enzyme-based cloning. Oligonucleotide primers P009_ispS_f and P010_ispS_r were used to amplify *ispS* from pBBR1MCS-2-RBS_1-ispS, and cloned into pEH040 by NdeI and AflII restriction sites.

pEH042 was constructed by restriction enzyme-based cloning. Oligonucleotide primers EH118_r and EH119_f were used to amplify the putative salicylate-inducible system from *C. necator* H16 genomic DNA, and cloned into pEH006 by AscI and NdeI restriction sites.

pEH043 was constructed by using the NEBuilder Hifi DNA assembly method. Oligonucleotide primers EH048_f and EH051_r, EH121_f and EH120_r, EH123_f and EH122_r were used to amplify the vector backbone for negatively regulated inducible systems and the 2,4-DAPG-inducible system PhlF/P_{phlA} from pEH005 and *Pseudomonas protegens* CHA0 genomic DNA, respectively.

pEH052 was constructed by using the NEBuilder Hifi DNA assembly method. Oligonucleotide primers EH133_r and EH134_f, EH135_r and EH136_f, EH051_r and EH015_f were used to amplify *acoR* and the acetoin-inducible promoter from *C. necator* H16 genomic DNA, and the vector backbone from pEH006.

pEH053 (*hpdH*-106::*rfp*) was constructed by restriction enzyme-based cloning. Oligonucleotide primers EH138_f and EH139_r were used to amplify the 106 bp long DNA sequence upstream of the *hpdH* translational start site from pEH008, and cloned into pEH036 by AvrII and NdeI restriction sites.

pEH068 (*hpdH*-118_mut2::*rfp*) was constructed by restriction enzyme-based cloning. Oligonucleotide primers EH175_f and EH165_r were used to incorporate mutation 2 (Figure 5.5a) into the 118 bp long DNA sequence upstream of the *hpdH* translational

start site. The sequence was amplified from pEH008 and cloned into pEH036 by AvrII and NdeI restriction sites.

pEH069 (*hpdH*-118_mut3::*rfp*) was constructed by restriction enzyme-based cloning. Oligonucleotide primers EH175_f and EH166_r were used to incorporate mutation 3 (Figure 5.5a) into the 118 bp long DNA sequence upstream of the *hpdH* translational start site. The sequence was amplified from pEH008 and cloned into pEH036 by AvrII and NdeI restriction sites.

pEH070 (*hpdH*-118_mut4::*rfp*) was constructed by restriction enzyme-based cloning. Oligonucleotide primers EH175_f and EH167_r were used to incorporate mutation 4 (Figure 5.5a) into the 118 bp long DNA sequence upstream of the *hpdH* translational start site. The sequence was amplified from pEH008 and cloned into pEH036 by AvrII and NdeI restriction sites.

pEH071 (*hpdH*-118_mut5::*rfp*) was constructed by restriction enzyme-based cloning. Oligonucleotide primers EH175_f and EH168_r were used to incorporate mutation 5 (Figure 5.5a) into the 118 bp long DNA sequence upstream of the *hpdH* translational start site. The sequence was amplified from pEH008 and cloned into pEH036 by AvrII and NdeI restriction sites.

pEH072 (*hpdH*-118_mut6::*rfp*) was constructed by restriction enzyme-based cloning. Oligonucleotide primers EH175_f and EH169_r were used to incorporate mutation 6 (Figure 5.5a) into the 118 bp long DNA sequence upstream of the *hpdH* translational start site. The sequence was amplified from pEH008 and cloned into pEH036 by AvrII and NdeI restriction sites.

pEH073 (*hpdH*-118_mut7::*rfp*) was constructed by restriction enzyme-based cloning. Oligonucleotide primers EH175_f and EH170_r were used to incorporate mutation 7 (Figure 5.5a) into the 118 bp long DNA sequence upstream of the *hpdH* translational start site. The sequence was amplified from pEH008 and cloned into pEH036 by AvrII and NdeI restriction sites.

pEH074 (*hpdH*-118_mut8::*rfp*) was constructed by restriction enzyme-based cloning. Oligonucleotide primers EH175_f and EH171_r were used to incorporate mutation 8 (Figure 5.5a) into the 118 bp long DNA sequence upstream of the *hpdH* translational start site. The sequence was amplified from pEH008 and cloned into pEH036 by AvrII and NdeI restriction sites.

pEH075 (*hpdH*-118_mut9::*rfp*) was constructed by restriction enzyme-based cloning. Oligonucleotide primers EH175_f and EH172_r were used to incorporate mutation 9 (Figure 5.5a) into the 118 bp long DNA sequence upstream of the *hpdH* translational start site. The sequence was amplified from pEH008 and cloned into pEH036 by AvrII and NdeI restriction sites.

pEH076 (*hpdH*-118_mut10::*rfp*) was constructed by restriction enzyme-based cloning. Oligonucleotide primers EH175_f and EH173_r were used to incorporate mutation 10 (Figure 5.5a) into the 118 bp long DNA sequence upstream of the *hpdH* translational start site. The sequence was amplified from pEH008 and cloned into pEH036 by AvrII and NdeI restriction sites.

pEH077 (*hpdH*-118_mut11::*rfp*) was constructed by restriction enzyme-based cloning. Oligonucleotide primers EH175_f and EH174_r were used to incorporate mutation 11 (Figure 5.5a) into the 118 bp long DNA sequence upstream of the *hpdH* translational start site. The sequence was amplified from pEH008 and cloned into pEH036 by AvrII and NdeI restriction sites.

pEH079 (*hpdH*-118_mut1::*rfp*) was constructed by restriction enzyme-based cloning. Oligonucleotide primers EH175_f and EH178_r were used to incorporate mutation 1 (Figure 5.5a) into the 118 bp long DNA sequence upstream of the *hpdH* translational start site. The sequence was amplified from pEH008 and cloned into pEH036 by AvrII and NdeI restriction sites.

pEH080 (*hpdH*-118_mut12::*rfp*) was constructed by restriction enzyme-based cloning. Oligonucleotide primers EH175_f and EH179_r were used to incorporate mutation 12 (Figure 5.5a) into the 118 bp long DNA sequence upstream of the *hpdH* translational start site. The sequence was amplified from pEH008 and cloned into pEH036 by AvrII and NdeI restriction sites.

pEH083 was constructed by restriction enzyme-based cloning. Oligonucleotide primers EH182_r and EH183_f were used to amplify the putative tartrate-inducible system from *C. necator* H16 genomic DNA, and cloned into pEH006 by AscI and NdeI restriction sites.

pEH086 was constructed by restriction enzyme-based cloning. Oligonucleotide primers EH191_f and EH190_r were used to amplify the itaconate-inducible system ItcR/P_{ccl} from *Y. pseudotuberculosis* YPIII genomic DNA, and cloned into pEH006 by AatII and NdeI restriction sites.

pEH089 was constructed by using the NEBuilder Hifi DNA assembly method. Oligonucleotide primers EH197_f and EH196_r were used to amplify *hpdR* from *P. putida* KT2440 genomic DNA, and cloned into pGEX-6P-1 by BamHI and NotI restriction sites.

pEH095 was constructed by restriction enzyme-based cloning. Oligonucleotide primers EH210_r and EH119_f were used to amplify the putative salicylate-inducible promoter from *C. necator* H16 genomic DNA, and cloned into pEH006 by AatII and NdeI restriction sites.

pEH096 was constructed by restriction enzyme-based cloning. Oligonucleotide primers EH211_r and EH212_f were used to amplify the acetoin-inducible promoter from *C. necator* H16 genomic DNA, and cloned into pEH006 by AatII and NdeI restriction sites.

pEH097 was constructed by restriction enzyme-based cloning. Oligonucleotide primers EH213_r and EH183_f were used to amplify the putative tartrate-inducible promoter from *C. necator* H16 genomic DNA, and cloned into pEH006 by AatII and NdeI restriction sites.

pEH101 was constructed by restriction enzyme-based cloning. Oligonucleotide primers N31_asp_r and EH217_f were used to amplify the putative

β-alanine-inducible promoter from *C. necator* H16 genomic DNA, and cloned into pEH006 by AatII and NdeI restriction sites.

pEH134 was constructed by using the NEBuilder Hifi DNA assembly method. Oligonucleotide primers EH258_r and EH259_f were used to amplify the putative formate-inducible system from *C. necator* H16 genomic DNA. The PCR product was combined with AatII/NdeI digested pEH006.

pEH136 was constructed by restriction enzyme-based cloning. Oligonucleotide primers EH262_r and EH263_f were used to amplify the putative L-kynurenine-inducible system from *C. necator* H16 genomic DNA, and cloned into pEH006 by AatII and NdeI restriction sites.

pEH137 was constructed by restriction enzyme-based cloning. Oligonucleotide primers EH264_r and EH265_f were used to amplify the putative L-phenylalanine-inducible system from *C. necator* H16 genomic DNA, and cloned into pEH006 by AatII and NdeI restriction sites.

pEH147 was constructed by restriction enzyme-based cloning. Oligonucleotide primers N31_asp_r and N30_asp_f were used to amplify the putative β-alanine-inducible system from *C. necator* H16 genomic DNA, and cloned into pEH006 by AatII and NdeI restriction sites.

pEH148 was constructed by restriction enzyme-based cloning. Oligonucleotide primers N36_f and N38_r were used to amplify the putative benzoate-inducible system from *C. necator* H16 genomic DNA, and cloned into pEH006 by AatII and NdeI restriction sites.

pEH149 was constructed by restriction enzyme-based cloning. Oligonucleotide primers N37_f and N38_r were used to amplify the putative benzoate-inducible promoter from *C. necator* H16 genomic DNA, and cloned into pEH006 by AatII and NdeI restriction sites.

pEH151 was constructed by using the NEBuilder Hifi DNA assembly method. Oligonucleotide primers EH272_r and EH259_f were used to amplify the putative formate-inducible promoter from *C. necator* H16 genomic DNA. The PCR product was combined with AatII/NdeI digested pEH006.

pEH152 was constructed by restriction enzyme-based cloning. Oligonucleotide primers EH273_r and EH263_f were used to amplify the putative L-kynurenine-inducible promoter from *C. necator* H16 genomic DNA, and cloned into pEH006 by AatII and NdeI restriction sites.

pEH153 was constructed by restriction enzyme-based cloning. Oligonucleotide primers EH274_r and EH265_f were used to amplify the putative L-phenylalanine-inducible promoter from *C. necator* H16 genomic DNA, and cloned into pEH006 by AatII and NdeI restriction sites.

pEH154 was constructed by using the NEBuilder Hifi DNA assembly method. Oligonucleotide primers EH275_r and EH276_f were used to amplify the putative

xanthine-inducible system from *C. necator* H16 genomic DNA. The PCR product was combined with AatII/NdeI digested pEH006.

pEH155 was constructed by restriction enzyme-based cloning. Oligonucleotide primers EH277_r and EH278_f were used to amplify the putative phenylglyoxylate-inducible system from *C. necator* H16 genomic DNA, and cloned into pEH006 by AatII and NdeI restriction sites.

pEH156 was constructed by using the NEBuilder Hifi DNA assembly method. Oligonucleotide primers EH279_r and EH280_f were used to amplify the putative GABA-inducible system from *C. necator* H16 genomic DNA. The PCR product was combined with AatII/NdeI digested pEH006.

pEH157 was constructed by restriction enzyme-based cloning. Oligonucleotide primers EH281_r and EH282_f were used to amplify the putative sulfonatoacetate-inducible system from *C. necator* H16 genomic DNA, and cloned into pEH006 by AatII and NdeI restriction sites.

pEH158 was constructed by restriction enzyme-based cloning. Oligonucleotide primers EH283_r and EH284_f were used to amplify the putative L-tyrosine-inducible system from *C. necator* H16 genomic DNA, and cloned into pEH006 by AatII and NdeI restriction sites.

pEH159 was constructed by restriction enzyme-based cloning. Oligonucleotide primers EH285_r and EH286_f were used to amplify the putative cyclohexanecarboxylate-inducible system from *C. necator* H16 genomic DNA, and cloned into pEH006 by AatII and NdeI restriction sites.

pEH160 was constructed by using the NEBuilder Hifi DNA assembly method. Oligonucleotide primers EH287_r and EH288_f were used to amplify the putative L-glutamine-inducible system from *C. necator* H16 genomic DNA. The PCR product was combined with AatII/NdeI digested pEH006.

pEH161 was constructed by restriction enzyme-based cloning. Oligonucleotide primers EH289_r and EH290_f were used to amplify the putative 3,4-dihydroxybenzoate-inducible system from *C. necator* H16 genomic DNA, and cloned into pEH006 by AatII and NdeI restriction sites.

pEH164 contains both the itaconate biosensor composed of *Yp*ItcR-P_{ccl}-*rfp* and the L-arabinose-inducible system including restriction sites for subsequent integration of the *cis*-aconitate decarboxylase *cadA* gene (ATEG_09971) downstream of P_{araBAD}. It was constructed by using the NEBuilder Hifi DNA assembly method. Oligonucleotide primers EH011_f and EH075_r, EH015_f and EH012_r, EH078_f and EH190_r, EH083_f and EH079_r were used to amplify the replication origin and the chloramphenicol resistance marker, *Yp*ItcR-P_{ccl}-*rfp*, and the L-arabinose-inducible system from pBBR1MCS-2-PphaC-eyfp-c1, pEH086, and pEH006, respectively.

pEH165 contains both the itaconate production system AraC-P_{araBAD}-*cadA* and the itaconate biosensor *Yp*ItcR-P_{ccl}-*rfp*. It was constructed by using the NEBuilder Hifi DNA assembly method. Oligonucleotide primers EH294_f and EH293_r, EH296_f

and EH295_r were used to amplify exon 1 and 2 of ATEG_09971 from *A. terreus* NIH2642 genomic DNA, and combined with BglII/SbfI digested pEH164.

pEH167 was constructed by restriction enzyme-based cloning. Oligonucleotide primers EH297_r and EH276_f were used to amplify the putative xanthine-inducible promoter from *C. necator* H16 genomic DNA, and cloned into pEH006 by AatII and NdeI restriction sites.

pEH168 was constructed by restriction enzyme-based cloning. Oligonucleotide primers EH298_r and EH282_f were used to amplify the putative sulfonatoacetate-inducible promoter from *C. necator* H16 genomic DNA, and cloned into pEH006 by AatII and NdeI restriction sites.

pEH169 was constructed by restriction enzyme-based cloning. Oligonucleotide primers EH299_r and EH284_f were used to amplify the putative L-tyrosine-inducible promoter from *C. necator* H16 genomic DNA, and cloned into pEH006 by AatII and NdeI restriction sites.

pEH170 was constructed by restriction enzyme-based cloning. Oligonucleotide primers EH300_r and EH286_f were used to amplify the putative cyclohexanecarboxylate-inducible promoter from *C. necator* H16 genomic DNA, and cloned into pEH006 by AatII and NdeI restriction sites.

pEH171 was constructed by restriction enzyme-based cloning. Oligonucleotide primers EH301_r and EH290_f were used to amplify the putative 3,4-dihydroxybenzoate-inducible promoter from *C. necator* H16 genomic DNA, and cloned into pEH006 by AatII and NdeI restriction sites.

pEH172 was constructed by restriction enzyme-based cloning. Oligonucleotide primers EH191_f and EH302_r were used to amplify the itaconate-inducible promoter P_{ccl} from *Y. pseudotuberculosis* YPIII genomic DNA, and cloned into pEH006 by AatII and NdeI restriction sites.

pEH173 was constructed by restriction enzyme-based cloning. The β-alanine-inducible system which was optimised for *E. coli* codon usage was cut from p17ACKHEP and cloned into pEH006 by AatII and NdeI restriction sites.

pEH176 was constructed by restriction enzyme-based cloning. Oligonucleotide primers EH310_f and EH016_r were used to amplify the L-arabinose-inducible system AraC/P_{araBAD} from *E. coli* MG1655 genomic DNA, and cloned into pEH006 by AatII and NdeI restriction sites.

pEH177 was constructed by restriction enzyme-based cloning. Oligonucleotide primers EH312_f and EH311_r were used to amplify the itaconate-inducible system ItcR/P_{ich} from *P. aeruginosa* PAO1 genomic DNA, and cloned into pEH006 by AatII and NdeI restriction sites.

pEH178 was constructed by restriction enzyme-based cloning. Oligonucleotide primers EH312_f and EH313_r were used to amplify the itaconate-inducible promoter P_{ich} from *P. aeruginosa* PAO1 genomic DNA, and cloned into pEH006 by AatII and NdeI restriction sites.

pEH194 was constructed by restriction enzyme-based cloning. Oligonucleotide primers EH344_f and EH345_r were used to amplify the tetracycline resistance marker from pME6000, and cloned into pEH042 by PmeI and AscI restriction sites.

pEH195 was constructed by restriction enzyme-based cloning. Oligonucleotide primers EH344_f and EH345_r were used to amplify the tetracycline resistance marker from pME6000, and cloned into pEH095 by PmeI and AscI restriction sites.

pEH196 was constructed by restriction enzyme-based cloning. Oligonucleotide primers EH344_f and EH345_r were used to amplify the tetracycline resistance marker from pMTL74311, and cloned into pEH101 by PmeI and AscI restriction sites.

pEH197 was constructed by restriction enzyme-based cloning. Oligonucleotide primers EH344_f and EH345_r were used to amplify the tetracycline resistance marker from pME6000, and cloned into pEH010 by PmeI and AscI restriction sites.

pEH198 was constructed by restriction enzyme-based cloning. Oligonucleotide primers EH344_f and EH345_r were used to amplify the tetracycline resistance marker from pME6000, and cloned into pEH035 by PmeI and AscI restriction sites.

pEH199 was constructed by restriction enzyme-based cloning. Oligonucleotide primers EH344_f and EH345_r were used to amplify the tetracycline resistance marker from pMTL74311, and cloned into pEH083 by PmeI and AscI restriction sites.

pEH200 was constructed by restriction enzyme-based cloning. Oligonucleotide primers EH344_f and EH345_r were used to amplify the tetracycline resistance marker from pMTL74311, and cloned into pEH097 by PmeI and AscI restriction sites.

pEH201 was constructed by restriction enzyme-based cloning. Oligonucleotide primers EH344_f and EH345_r were used to amplify the tetracycline resistance marker from pMTL74311, and cloned into pEH096 by PmeI and AscI restriction sites.

pEH202 was constructed by restriction enzyme-based cloning. Oligonucleotide primers EH344_f and EH345_r were used to amplify the tetracycline resistance marker from pMTL74311, and cloned into pEH134 by PmeI and AscI restriction sites.

pEH203 was constructed by restriction enzyme-based cloning. Oligonucleotide primers EH344_f and EH345_r were used to amplify the tetracycline resistance marker from pMTL74311, and cloned into pEH151 by PmeI and AscI restriction sites.

pEH204 was constructed by restriction enzyme-based cloning. Oligonucleotide primers EH344_f and EH345_r were used to amplify the tetracycline resistance marker from pMTL74311, and cloned into pEH167 by PmeI and AscI restriction sites.

pEH205 was constructed by restriction enzyme-based cloning. Oligonucleotide primers EH344_f and EH345_r were used to amplify the tetracycline resistance marker from pMTL74311, and cloned into pEH148 by PmeI and AscI restriction sites.

pEH206 was constructed by restriction enzyme-based cloning. Oligonucleotide primers EH344_f and EH345_r were used to amplify the tetracycline resistance marker from pMTL74311, and cloned into pEH149 by PmeI and AscI restriction sites.

pEH207 was constructed by restriction enzyme-based cloning. Oligonucleotide primers EH344_f and EH345_r were used to amplify the tetracycline resistance marker from pMTL74311, and cloned into pEH168 by PmeI and AscI restriction sites.

pEH208 was constructed by restriction enzyme-based cloning. Oligonucleotide primers EH344_f and EH345_r were used to amplify the tetracycline resistance marker from pMTL74311, and cloned into pEH136 by PmeI and AscI restriction sites.

pEH209 was constructed by restriction enzyme-based cloning. Oligonucleotide primers EH344_f and EH345_r were used to amplify the tetracycline resistance marker from pMTL74311, and cloned into pEH152 by PmeI and AscI restriction sites.

pEH210 was constructed by restriction enzyme-based cloning. Oligonucleotide primers EH344_f and EH345_r were used to amplify the tetracycline resistance marker from pMTL74311, and cloned into pEH137 by PmeI and AscI restriction sites.

pEH211 was constructed by restriction enzyme-based cloning. Oligonucleotide primers EH344_f and EH345_r were used to amplify the tetracycline resistance marker from pMTL74311, and cloned into pEH153 by PmeI and AscI restriction sites.

pEH212 was constructed by restriction enzyme-based cloning. Oligonucleotide primers EH344_f and EH345_r were used to amplify the tetracycline resistance marker from pMTL74311, and cloned into pEH158 by PmeI and AscI restriction sites.

pEH213 was constructed by restriction enzyme-based cloning. Oligonucleotide primers EH344_f and EH345_r were used to amplify the tetracycline resistance marker from pMTL74311, and cloned into pEH169 by PmeI and AscI restriction sites.

pEH214 was constructed by restriction enzyme-based cloning. Oligonucleotide primers EH344_f and EH345_r were used to amplify the tetracycline resistance marker from pMTL74311, and cloned into pEH159 by PmeI and AscI restriction sites.

pEH215 was constructed by restriction enzyme-based cloning. Oligonucleotide primers EH344_f and EH345_r were used to amplify the tetracycline resistance marker from pMTL74311, and cloned into pEH170 by PmeI and AscI restriction sites.

pEH216 was constructed by restriction enzyme-based cloning. Oligonucleotide primers EH344_f and EH345_r were used to amplify the tetracycline resistance marker from pMTL74311, and cloned into pEH161 by PmeI and AscI restriction sites.

pEH217 was constructed by restriction enzyme-based cloning. Oligonucleotide primers EH344_f and EH345_r were used to amplify the tetracycline resistance marker from pMTL74311, and cloned into pEH171 by PmeI and AscI restriction sites.

pEH218 was constructed by restriction enzyme-based cloning. Oligonucleotide primers EH344_f and EH345_r were used to amplify the tetracycline resistance marker from pMTL74311, and cloned into pEH155 by PmeI and AscI restriction sites.

pEH220 was constructed by restriction enzyme-based cloning. Oligonucleotide primers EH344_f and EH345_r were used to amplify the tetracycline resistance marker from pME6000, and cloned into pEH176 by PmeI and AscI restriction sites.

pEH221 was constructed by restriction enzyme-based cloning. Oligonucleotide primers EH344_f and EH345_r were used to amplify the tetracycline resistance marker from pMTL74311, and cloned into pEH147 by PmeI and FseI restriction sites.

pEH222 was constructed by restriction enzyme-based cloning. Oligonucleotide primers EH344_f and EH345_r were used to amplify the tetracycline resistance marker from pMTL74311, and cloned into pEH052 by PmeI and FseI restriction sites.

pEH223 was constructed by restriction enzyme-based cloning. Oligonucleotide primers EH344_f and EH345_r were used to amplify the tetracycline resistance marker from pMTL74311, and cloned into pEH154 by PmeI and FseI restriction sites.

pEH224 was constructed by restriction enzyme-based cloning. Oligonucleotide primers EH344_f and EH345_r were used to amplify the tetracycline resistance marker from pMTL74311, and cloned into pEH157 by PmeI and FseI restriction sites.

pEH225 was constructed by restriction enzyme-based cloning. Oligonucleotide primers N30_asp_f and N39_r were used to amplify *oapR* gene from pEH147 plasmid DNA with the core sequence of constitutive promoter P_{13} including the phage T7 gene *10* RBS, and cloned into pEH147 by AatII and BsrGI restriction sites.

pEH226 was constructed by restriction enzyme-based cloning. Oligonucleotide primers N30_asp_f and N40_r were used to amplify *oapR* gene optimised for *E. coli* codon usage from p17ACKHEP plasmid DNA with the core sequence of constitutive promoter P_{13} including the phage T7 gene *10* RBS, and cloned into pEH147 by AatII and BsrGI restriction sites.

pEH229 was constructed by restriction enzyme-based cloning. Oligonucleotide primers EH346_r and EHseq026_f were used to amplify the putative phenylglyoxylate-inducible promoter from pEH155, and cloned into pEH006 by AatII and NdeI restriction sites.

pEH234 was constructed by restriction enzyme-based cloning. Oligonucleotide primers EH344_f and EH345_r were used to amplify the tetracycline resistance marker from pME6000, and cloned into pEH225 by PmeI and FseI restriction sites.

pEH235 was constructed by restriction enzyme-based cloning. Oligonucleotide primers EH344_f and EH345_r were used to amplify the tetracycline resistance marker from pME6000, and cloned into pEH226 by PmeI and AscI restriction sites.

pEH240 was constructed by using the NEBuilder Hifi DNA assembly method. Oligonucleotide primers EH360_r and EH361_f were used to amplify the putative L-glutamine-inducible system, *H16_RS29650*, and the intergenic region preceding *H16_RS29655* from *C. necator* H16 genomic DNA. The PCR product was combined with AatII/NdeI digested pEH006.

pEH256 was constructed by using the NEBuilder Hifi DNA assembly method. Oligonucleotide primers EH397_r and EH398_f were used to amplify the putative GABA-inducible system, *H16_RS23655*, and the intergenic region preceding *H16_RS23660* from *C. necator* H16 genomic DNA. The PCR product was combined with AatII/NdeI digested pEH006.

pEH257 was constructed by restriction enzyme-based cloning. The tetracycline resistance marker from pMTL74311 was cloned into pEH229 by AscI and PmeI restriction sites.

pEH263 was constructed by using the NEBuilder Hifi DNA assembly method. Oligonucleotide primers EH421_f and EH415_r, EH423_f and EH422_r, EH425_f and EH424_r, EH420_f and EH426_r were used to amplify *egfp*, the salicylate-inducible system, the 3,4-dihydroxybenzoate-inducible system and *rfp* from pJOE7801.1, pEH042, and pEH161. The PCR products were combined with AscI and SbfI digested pEH220.

pEH266 was constructed by restriction enzyme-based cloning. Oligonucleotide primers EH436_f and EH435_r were used to amplify the putative GABA-inducible promoter, *H16_RS23655*, and the intergenic region preceding *H16_RS23660* from *C. necator* H16 genomic DNA, and cloned into pEH006 by NdeI and AscI restriction sites.

pEH268 was constructed by using the NEBuilder Hifi DNA assembly method. Oligonucleotide primers EH441_r and EH398_f were used to amplify the putative GABA-inducible system, *H16_RS23655*, and the intergenic region preceding *H16_RS23660* from *C. necator* H16 genomic DNA. The PCR product was combined with AscI/NdeI digested pEH220.

pEH269 was constructed by restriction enzyme-based cloning. Oligonucleotide primers EH436_f and EH435_r were used to amplify the putative GABA-inducible promoter, *H16_RS23655*, and the intergenic region preceding *H16_RS23660* from *C. necator* H16 genomic DNA, and cloned into pEH220 by NdeI and AscI restriction sites.

9.2 Codon optimised DNA sequences

P. alba ispS optimised for *C. necator* codon usage:

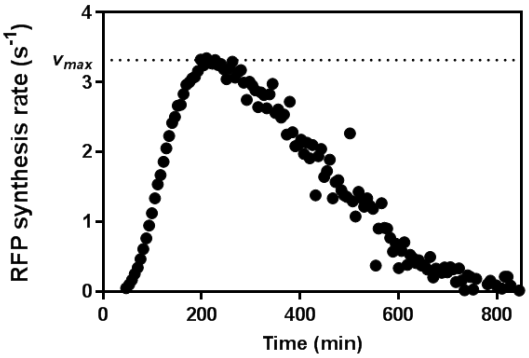
ATGGAAGCCCGCCGCTCGGCCAACTACGAGCCGAACTCGTGGGACTACGACTACCT
GCTGTCGTCGGACACCGACGAGTCGATCGAGGTGTACAAGGACAAGGCCAAGAAGC
TGGAAGCCGAGGTCCGCCGCGAGATCAACAACGAGAAGGCCGAGTTCCTGACGCTG
CTGGAACTGATCGACAACGTGCAGCGCCTGGGCCTGGGCTACCGCTTCGAAAGCGA
CATCCGCGGCGCCCTGGACCGCTTCGTGAGCAGCGGCGGCTTCGACGCCGTGACCA
AGACCTCGCTGCATGGCACCGCGCTGTCGTTCCGCCTGCTGCGCCAGCACGGCTTC
GAGGTGTCGCAGGAAGCCTTCTCGGGCTTCAAGGACCAGAACGGCAACTTCCTGGA
AAACCTGAAGGAAGATATCAAGGCCATCCTGTCGCTGTACGAGGCCAGCTTCCTGG
CGCTGGAAGGCGAGAACATCCTGGACGAGGCCAAGGTGTTCGCCATCTCGCATCTG
AAGGAACTGTCGGAGGAAAAGATCGGCAAGGAACTGGCCGAACAGGTGAACCATGC
GCTGGAACTGCCGCTGCACCGCCGCACGCAGCGCCTGGAAGCCGTGTGGTCGATCG
AAGCCTACCGCAAGAAGGAAGACGCCAACCAGGTCCTGCTGGAACTGGCCATCCTG
GACTACAACATGATCCAGTCGGTGTACCAGCGCGACCTGCGCGAAACCAGCCGCTG
GTGGCGCCGCGTCGGCCTGGCCACCAAGCTGCACTTCGCCCGCGACCGCCTGATCG
AGTCGTTCTACTGGGCCGTGGGCGTCGCCTTCGAGCCGCAGTATTCGGACTGCCGC
AACTCGGTGGCCAAGATGTTCAGCTTCGTGACCATCATCGACGACATCTACGACGT

GTACGGCACCCTGGACGAACTGGAACTGTTCACCGACGCGGTGGAACGCTGGGACG
TGAACGCCATCAACGACCTGCCGGACTATATGAAGCTGTGCTTCCTGGCGCTGTAC
AACACCATCAACGAGATCGCCTACGACAATCTGAAGGACAAGGGCGAAAATATCCT
GCCGTACCTGACCAAGGCCTGGGCCGACCTGTGCAACGCCTTCCTGCAGGAAGCGA
AGTGGCTGTATAACAAGTCGACCCCGACCTTCGACGACTACTTCGGCAACGCGTGG
AAGTCGTCGTCGGGCCCGCTGCAGCTGGTGTTCGCGTACTTCGCCGTGGTGCAGAA
CATCAAGAAGGAAGAGATCGAGAACCTGCAGAAGTACCACGACACCATCTCGCGCC
CGTCGCACATCTTCCGCCTGTGCAATGACCTGGCCTCGGCCTCGGCGGAAATCGCC
CGCGGCGAAACCGCCAACAGCGTGTCGTGCTACATGCGCACCAAGGGCATCTCGGA
AGAACTGGCGACCGAGTCGGTGATGAACCTGATCGACGAAACCTGGAAGAAGATGA
ACAAGGAAAAGCTGGGCGGCTCGCTGTTCGCCAAGCCGTTCGTGGAAACCGCGATC
AATCTGGCCCGCCAGTCGCACTGCACCTACCACAACGGCGACGCGCACACCTCGCC
GGATGAGCTGACCCGCAAGCGCGTGCTGAGCGTGATCACCGAGCCGATCCTGCCGT
TCGAGCGCTGA

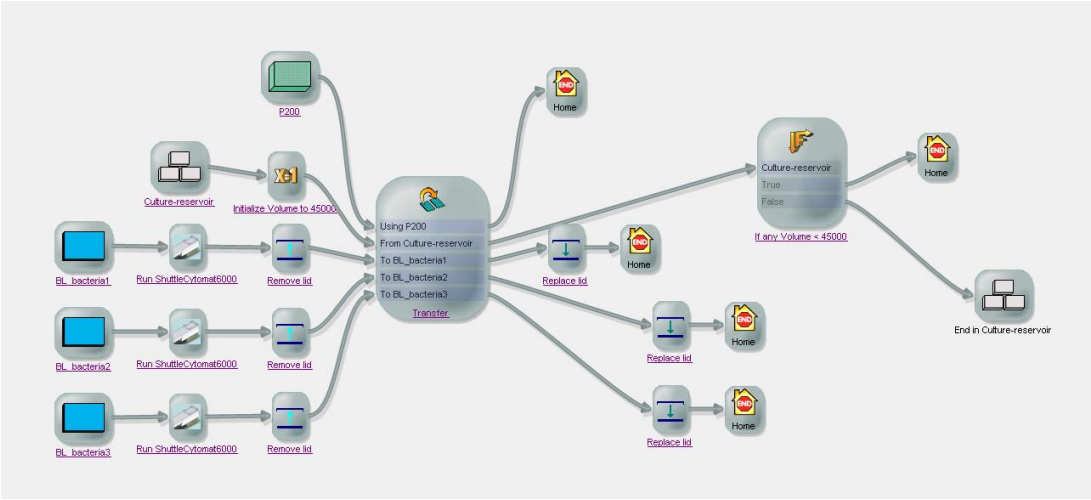
C. necator oapR optimised for *E. coli* codon usage:

ATGCTGACCCTGAATCTGACCCGTACACGTCGTGATGGTGATACCCTGACCGAGCA
GATTGTTGCAGGTATTGCAGCACTGGTTGAACAGCGTGCACTGCGTGCAGGCACCG
CACTGCCGAGCGTTCGTCGTTTTGCACAGCATCATCATGTTAGCACCTTTACCGTT
GCCGAAGCCTATGGTCGTCTGACCGCACTGGGTTATCTGGCAGCACGTCCTGGTAG
CGGTTATACCGTTGCACATCGTCATGCACCGGCAGGTCATGCCCGTGCACCGCAGT
GGGAAGCACCGGGTCTGAATGCAGCATGGCTGCTGAGTGATGTTTTTGCAGATCAT
AGCGTTCCGATTAAAGCCGGTGCAGGTTGGCTGCCTGGTGATTGGCTGAATGAAGA
AGGCCTGCATCAGGCCATGCGTGCAAGCGCACGTGTTCCGGCAGCACAGCTGAGCG
GTTATGGTCATCCGTATGGTTTTGCACCGCTGCGTGAACACATTGCAGCAGGTCTG
GGTCAGTATGGTATTCCGCTGCAGGCACAGCAGGTTGTTCTGACCCAGGGTGCAAC
CCAGGCACTGGATCTGGTTGTTCGTACCCTGCTGCGTGCCGGTGATCGTGTTCTGG
TTGAAAGCCCGTGTTATTGTAATCTGCTGCAGATCCTGCGTCTGGCAGGTCTGCGT
GTTGTTGGTGTTCCGCGTAGCGCAGCCGGTCTGGATACCGATGCACTGGATGATGC
AATTCGTGCACATGCACCGCGTGCGCTGTTTATTAACACCGTGCTGCAGAATCCGA
CCGGTGCAAGCCTGAGCAGCATGAATGCATTTCGTGTGCTGCAGCTGGCAGAACAG
CATCGTCTGCTGGTTGTTGAAGATGATATTTATCGTGAACTGGCTCCGGCAGGTAG
CCCGATGCTGGCAGCAATGGATGGTCTGAGCCAGGTTGTTTATATCAATGGTTTTA
GCAAAACGATTACCCCGAGCCTGCGCGTTGGTTATTTAGCAGCAAGTCCGGATCTG
GCAAAAGCATTTGCACGTACCAAAATGGCAGTTGGTCTGACCAGCAGCGAAGTTAC
CGAACGTCTGGTTTATAGCGTGCTGACCAGCGGTCATTATGGTCGCCATGTTGCAG
CCCTGGCGGAACGTCTGCGTGCCCAGCAGGATCGTGTTACCGAAAAAATGGAAGCA
CATGGTCTGGAAGTTCTGCTGCGTCCTGAAGGTGGTATGTTTGCATGGGCACGCCT
GACCGAAGCCGCACAGGCACGTCTGCAGGCAAGCCGTCGTGGTGGTCCGCTGCATG
GTAATCGTCTGGCAACCCTGGCGCTGGAATATGGTATTTGGCTGGCACCGGGTAGC
TATTTTGAACCGGATGAAACCGATAGTCCGTGGATTCGTTTTAATGTTGCAACCGG
TGATGCTCCGCAGCTGTGGCAGTTTTTTGATCGCCTGGCGCAGGCACCTCGTGCAG
CATAA

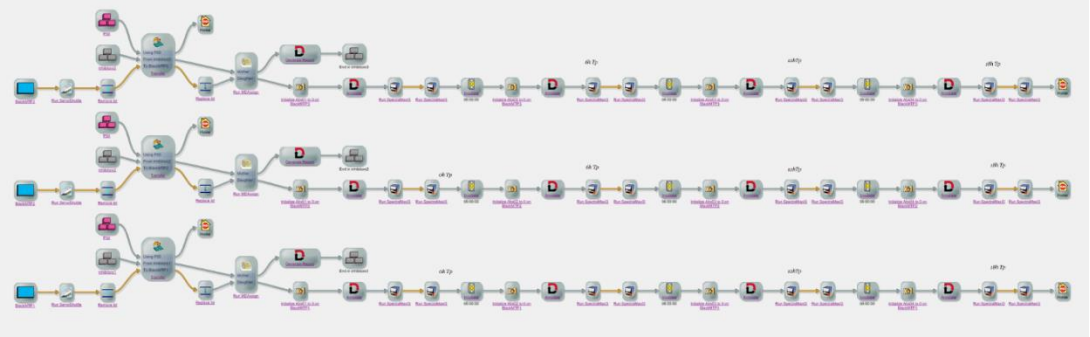
9.3 Supplementary Figures



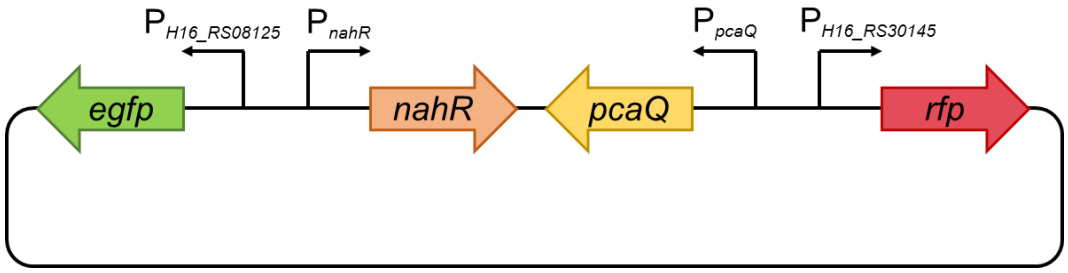
Supplementary Figure 1 RFP synthesis rate as a function of time for the L-rhamnose-inducible system. The RFP synthesis rate of a culture of *C. necator* H16 carrying pEH002 after supplementation with 10 mM L-rhamnose was calculated using formula (3.1). v_{max}, the maximum rate of RFP synthesis.



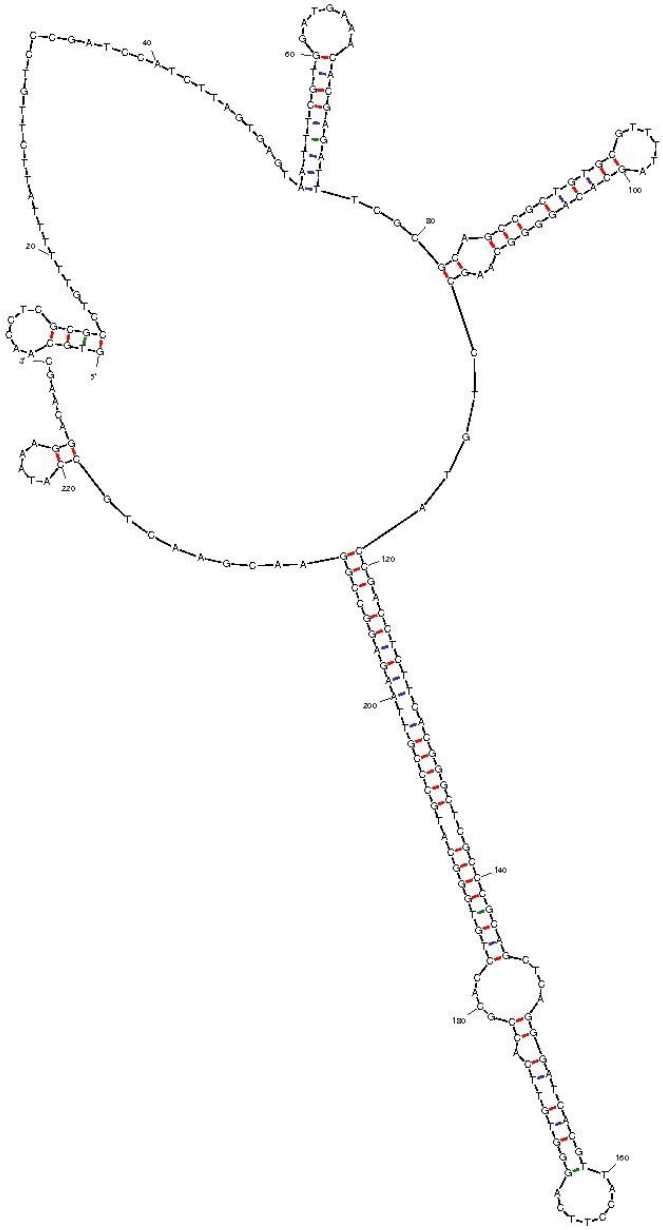
Supplementary Figure 2 Workflow illustrating the automated dispensing of bacterial culture. In total, one strain was dispensed into three black microtiter plates to test 67 compounds for cross-reactivity (four replicates per compound). The workflow was generated using software SEMI EX (Beckman Coulter).



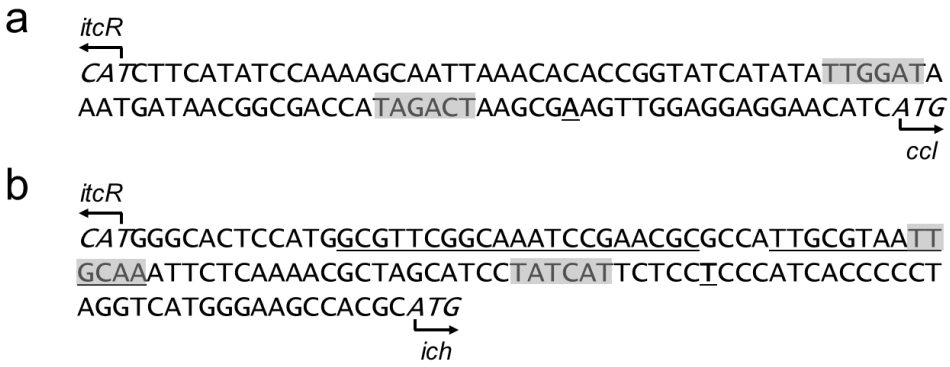
Supplementary Figure 3 Workflow illustrating the automated addition of inducer, fluorescence measurement, and incubation. In total, three black microtiter plates were required to test one strain against 67 compounds (four replicates per compound). The workflow was generated using software SEMI EX (Beckman Coulter).



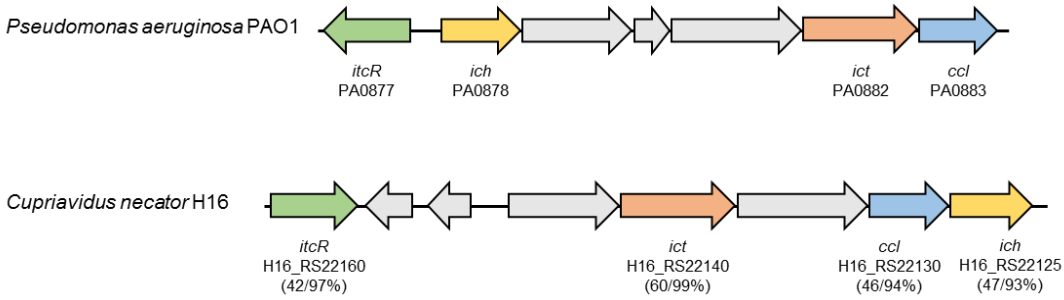
Supplementary Figure 4 Schematic illustration of pEH263. The vector contains the *egfp* reporter gene under control of the salicylate-inducible system NahR/$P_{H16\_RS08125}$ and *rfp* under control of the 3,4-dihydroxybenzoate-inducible system PcaQ/$P_{H16\_RS30145}$.



Supplementary Figure 5 Secondary structure analysis of the *P. putida* KT2440 *hpdR*/*hpdH* intergenic region. The *hpdR*/*hpdH* intergenic region was analysed using the mfold web server for potential secondary structures. Default settings were chosen to predict nucleic acid folding. The 5'- and 3'-ends represent the translational start sites of *hpdH* and *hpdR*, respectively. The calculated minimum free energy, Δ*G*, is -41.66 kcal/mol.



Supplementary Figure 6 Nucleotide sequences of intergenic regions containing putative itaconate-inducible promoters. The (a) *Y. pseudotuberculosis* YPIII *itcR*/*ccl* and (b) *P. aeruginosa* PAO1 *itcR*/*ich* intergenic regions. Translational start sites are italicised. Transcriptional start sites of genes *ccl* and *ich* were predicted using program NNPP[228] and are underlined and bold. The putative promoter -35 and -10 sequences were annotated manually on the basis of typical characteristics of bacterial promoters and shaded in grey. The putative ItcR Binding site in the *P. aeruginosa itcR*/*ich* intergenic region is underlined. Speculative annotation of the ItcR binding site is based on palindromic sequences upstream of the putative -35 and -10 promoter boxes. Palindromic sequences were identified using mfold[232] web server.



Supplementary Figure 7 Gene cluster putatively involved in itaconate degradation in *C. necator* H16. *C. necator* H16 gene cluster encoding the homologues of *P. aeruginosa* PAO1 Ict, Ich, and Ccl. An ItcR homolog is located in close proximity to the gene cluster involved in itaconate degradation in the same orientation. Gene names and locus tags are shown. Protein sequence identity and coverage is indicated in brackets.

9.4 Supplementary Tables

Supplementary Table 1 Chemicals used in this work.

Chemical	Supplier	Catalogue number
1-butanol	Sigma-Aldrich	B7906
1-Cyclohexene-1-carboxylic acid	Sigma-Aldrich	328367
1,3-butanediol	Sigma-Aldrich	309443
1-propanol	Sigma-Aldrich	279544
2-butanol	Sigma-Aldrich	294810
2-propanol	Sigma-Aldrich	I9516
2,6-Dihydroxybenzoic acid	Acros Organics, Thermo Fisher Scientific	114880250
3-Aminobutanoic acid	Sigma-Aldrich	A44207
3-butenoic acid	Sigma-Aldrich	134716
3-Hydroxy-2-butanone	Alfa Aesar, Thermo Fisher Scientific	A13752
3-Hydroxypropionic acid	Fluorochem	147100
3,4-Dihydroxybenzoic acid	Sigma-Aldrich	37580
4-Hydroxybenzoic acid	Acros Organics, Thermo Fisher Scientific	120991000
4-Hydroxyquinoline-2-carboxylic acid hydrate	Alfa Aesar, Thermo Fisher Scientific	A12602
4-isopropylbenzoic acid	Acros Organics, Thermo Fisher Scientific	412800050
α-ketoglutaric acid	Alfa Aesar, Thermo Fisher Scientific	A10256
β-Alanine	Sigma-Aldrich	146064
γ-Aminobutyric acid	Sigma-Aldrich	A2129
Butyric acid	Alfa Aesar, Thermo Fisher Scientific	L13189
Caffeine	Alfa Aesar, Thermo Fisher Scientific	39214
Catechol	Acros Organics, Thermo Fisher Scientific	158980050
cis-Aconitic acid	Sigma-Aldrich	A3412

cis,cis-Muconic acid	Acros Organics, Thermo Fisher Scientific	297760025
Citraconic acid	Sigma-Aldrich	C82604
Citric acid trisodium salt	Sigma-Aldrich	C3674
Cyclohexanecarboxylic acid	Alfa Aesar, Thermo Fisher Scientific	A14693
Cyclopentanecarboxylic acid	Alfa Aesar, Thermo Fisher Scientific	A12375
D-3-hydroxybutyric acid	Sigma-Aldrich	54920
D-Malic acid	Alfa Aesar, Thermo Fisher Scientific	A11688
D-Mannitol	Sigma-Aldrich	M9546
D-Saccharic acid potassium salt	Sigma-Aldrich	S4140
D,L-2-Hydroxybutyric acid sodium salt	Alfa Aesar, Thermo Fisher Scientific	A18636
D,L-2-Phenylglycine	Alfa Aesar, Thermo Fisher Scientific	B21129
D,L-2,3-Diaminopropionic acid monohydrochloride	Alfa Aesar, Thermo Fisher Scientific	L09485
D,L-Isocitric acid trisodium salt hydrate	Sigma-Aldrich	I1252
D,L-Isoserine	Sigma-Aldrich	286338
D,L-Mandelic acid	Acros Organics, Thermo Fisher Scientific	125311000
Dopamine hydrochloride	Alfa Aesar, Thermo Fisher Scientific	A11136
Ethanol	Honeywell	32221
Fumaric acid	Sigma-Aldrich	47910
Glycerol	Honeywell	49770
Glycine	Sigma-Aldrich	1042010100
Glyoxylic acid monohydrate	Sigma-Aldrich	G10601
Hippuric acid	Acros Organics, Thermo Fisher Scientific	150270050
Hydroquinone	Acros Organics, Thermo Fisher Scientific	219930500
Hypoxanthine	Sigma-Aldrich	H9377
Isethionic acid sodium salt	Sigma-Aldrich	220078

Isoprene	Alfa Aesar, Thermo Fisher Scientific	L14619
Itaconic acid	Sigma-Aldrich	I29204
L-3-hydroxybutyric acid	Sigma-Aldrich	54925
L-α-Hydroxyglutaric acid disodium salt	Sigma-Aldrich	90790
L-Alanine	Sigma-Aldrich	A26802
L-Arabinose	Acros Organics, Thermo Fisher Scientific	365181000
L-Aspartic acid potassium salt	Sigma-Aldrich	A6558
L-glutamic acid monosodium salt hydrate	Sigma-Aldrich	G1626
L-Glutamine	Sigma-Aldrich	G3126
L-Kynurenine	Sigma-Aldrich	K8625
L-Malic acid	Sigma-Aldrich	M1000
L-Phenylalanine	Sigma-Aldrich	P2126
L-Rhamnose monohydrate	Sigma-Aldrich	R3875
L-Tryptophan	Sigma-Aldrich	T0271
L-Tyrosine disodium salt	Acros Organics, Thermo Fisher Scientific	426410250
Lactic acid	Sigma-Aldrich	W261106
Levulinic acid	Sigma-Aldrich	L2009
Malonic acid	Sigma-Aldrich	M1296
Magnesium acrylate	Alfa Aesar, Thermo Fisher Scientific	42002
Mesaconic acid	Sigma-Aldrich	131040
Methacrylic acid	Sigma-Aldrich	155721
Methylsuccinic acid	Sigma-Aldrich	M81209
Nicotinic acid	Acros Organics, Thermo Fisher Scientific	128291000
Oxalacetic acid	Acros Organics, Thermo Fisher Scientific	416600050
Phenol	Acros Organics, Thermo Fisher Scientific	149340500
Phenyl acetate	Acros Organics, Thermo Fisher Scientific	148771000

Phenylglyoxylic acid	Alfa Aesar, Thermo Fisher Scientific	L00648
Potassium sodium tartrate	Sigma-Aldrich	1551140
Propionic acid	Alfa Aesar, Thermo Fisher Scientific	L04210
Pyruvic acid	Sigma-Aldrich	P76209
Resorcinol	Acros Organics, Thermo Fisher Scientific	132290500
Shikimic acid	Acros Organics, Thermo Fisher Scientific	132700010
Sodium acetate	Sigma-Aldrich	S2889
Sodium benzoate	Alfa Aesar, Thermo Fisher Scientific	A15946
Sodium D-lactate	Sigma-Aldrich	71716
Sodium formate	Sigma-Aldrich	71539
Sodium fumarate dibasic	Sigma-Aldrich	F1506
Sodium glycolate	Alfa Aesar, Thermo Fisher Scientific	A12341
Sodium L-lactate	Alfa Aesar, Thermo Fisher Scientific	L14500
Sodium salicylate		
Sodium succinate dibasic hexahydrate	Sigma-Aldrich	S2378
Sulfoacetic acid	Sigma-Aldrich	242802
Tartronic acid	Sigma-Aldrich	83620
Taurine	Alfa Aesar, Thermo Fisher Scientific	A12403
Theobromine	Sigma-Aldrich	T4500
Tiglic acid	Alfa Aesar, Thermo Fisher Scientific	A17056
trans-Aconitic acid	Sigma-Aldrich	122750
Tricarballylic acid	Acros Organics, Thermo Fisher Scientific	139360050
Uracil	Sigma-Aldrich	U0750
Valeric acid	Sigma-Aldrich	240370
Xanthine	Acros Organics, Thermo Fisher Scientific	149170050

Supplementary Table 2 Oligonucleotide primers used in this work. Restriction sites that were incorporated for cloning are underlined.

Primer name	Primer sequence (5' → 3')
EH001_f	tggtgagaatccaagcttccattcaggtcgaggtggccc
EH002_r	ttatacctagggcgttcggctgcggctggcgctgggcct
EH003_f	gccgcagccgaacgccctaggtataaacgcagaaaggccca
EH004_r	tgtttctccatagggagaccacaacggtttccctctagaaataattttggaattcaaaagatcttttaagaag gagatatacatatgg
EH005_f	accgttgtggtctccctatggagaaacagtagagagttgcgataaaaagcg
EH006_r	cctcgacctgaatggaagcttggattctcaccaataaaaaacgcccgg
EH007_f	atttctagagggaaaccgttgtggtctccctacgaccagtctaaaaagcgcct
EH008_r	atcgacgtcttaatctttctgcgaattgagatgacgc
EH009_f	atttctagagggaaaccgttgtggtctccctgcgttgattacagccttcaaacg
EH010_r	atcgacgtctcaggccaggttttgttccg
EH011_f	aatccaagcgtttaaacggaggcagacaaggtatagggc
EH012_r	tctgcctccgtttaaacgcttggattctcaccaataaaaaacgc
EH013_f	ggatgacctgcaggtataaacgcagaaaggcccacc
EH015_f	gcccagtctttcgactgagcctttcgttttatggcgcgccaggccggccgatatctggcgaaaatgagac gttg
EH016_r	cgaaaggctcagtcgaaagactgggcctttcgttttatgacgtcttatgacaacttgacggctacatcattc
EH017_f	cgccatatgcgttctccttggaattgttgtc
EH018_r	tatgacgtctcagtcctgggcaaagcg
EH019_f	cgccatatgcaacctcgcgcctg
EH020_r	tatgacgtcctactcggctagcaactcgc
EH021_f	cgccatatgagatgagggtgggggtttg
EH022_r	tatgacgtctcaaccttccgccacgc
EH023_f	cgccatatggctggcttctgcaaggatg
EH024_r	tatgacgtcctaagtccggaacaccgact
EH025_f	tcagaaggccatcctgacggatggccttttggcgcgccaggccggccgata
EH026_r	catacgagccggaagcataaagtgtaaagccgatcgttgttgacactctatcattgatagagttatttttacc acgggagaccacaacggtttcc
EH027_f	tacactttatgcttccggctcgtatgttgtgtggaattgtgagcggataacaatttcacacaggaaacagcta tgacaaagttgcagccgaat
EH028_r	ggccatccgtcaggatggccttctgacgtctcaatcgtcacccttttctcgg
EH041_f	atttctagagggaaaccgttgtggtctcccgataaagacaaagcaaggggtgtg

EH042_r	tatgacgtctcagcttttcagccgccg
EH043_f	atttctagagggaaaccgttgtggtctccctaagtttcactatacactctttggaaattgacc
EH044_r	tatgacgtctcatgtacctgcatcactctctttagt
EH048_f	gacgtcagaaggccatcct
EH051_r	atggcgagtagcgaagacgttatcaa
EH055_r	atccgtcaggatggccttctgacgtcttattcctgtgtccgggtcacg
EH056_f	tacactttatgcttccggctcgtatgttgtgtggaattgtgagcggataacaatttcacacaggaaacagcta tgcctctgacagacaccc
EH059_f	gctactcgccatatgcaacctcgcgcctg
EH060_r	tcgttttatgacgtcgcttgtcctttatggcagttcg
EH061_f	acttttacgcaacgcataattgttgt
EH062_r	acaacaattatgcgttgcgtaaaagt
EH075_r	aaggccatcctgacggatggccttttcctgcaggtcatcccaggtggcact
EH078_f	gttgtcataacctaggtataaacgcagaaaggcccacc
EH079_r	tgcgtttatacctaggttatgacaacttgacggctacatcat
EH083_f	aaaaggccatccgtcaggatggccttctatgtatatctccttcttaaaagatcttttgaattcca
EH095_r	tcgttttatgacgtcgatgatccggtttttttgtgcgt
EH096_f	gctactcgccatatgcgttctccttggaattgttgtc
EH097_r	tcgttttatgacgtcggcagtgtcggggcgaaa
EH098_f	gctactcgccatatggctggcttctgcaaggatg
EH099_r	gcctgtacctaggaacgaactgccataaaggacaag
EH100_f	gtcctttatggcagttcgttcctaggtacaggcttgcccctgtg
EH105_r	atccgtcaggatggccttctgacgtcctaacttataggggtaacacttaaaaaagaatcaat
EH106_f	tccggctcgtatgttgtgtggaattgtgagcggataacaatttcacacaggaaacagctatggttattattca aattgcagatcaagct
EH107_r	caattccacacaacatacgagccggaagcataaagtgtaaagccgatcgcattgaaataaacatttatttt gtatatgatgagataaagttagtttatt
EH108_f	cgccatatgtatatctccttcttaaattaagtgaacaagtttatccatcaactatcttaattg
EH109_r	tttaagaaggagatatacatatggcgagtagc
EH111_r	atccgtcaggatggccttctgacgtcctagcgcttgaatttcgcgtac
EH112_f	ccggctcgtatgttgtgtggaattgtgagcggataacaatttcacacaggaaacagctatggtgatcatga gtccaaagagaa
EH113_f	acaattccacacaacatacgagccggaagcataaagtgtaaagccgatcgaaatcataaaaaatttatttg ctttgtgagcg
EH114_r	atgtatatctccttcttaaagttaaacaaaattatttctagtaac

EH118_r ggccggcctggcgcgccataaaacgaaaggctcagtcgaaagactgggcctttcgttttatgacgtcctaatccgaaaacaagccgacg

EH119_f gctactcgccatatgtgtctccggctatgtctcttcg

EH120_r atccgtcaggatggccttctgacgtctcaacgttgcgtaccaggac

EH121_f tcgtatgttgtgtggaattgtgagcggataacaatttcacacaggaaacagctatggctcgtacccccctctc

EH122_r cgctcacaattccacacaacatacgagccggaagcataaagtgtaaagccgatcgaataaaaaatctccaggcagggc

EH123_f ttgataacgtcttcgctactcgccatatgcctcttgattccattcttttcagaaaact

EH133_r cgaaaggctcagtcgaaagactgggcctttcgttttatgacgtcttatgcgtcggtgggctc

EH134_f gcgcgtgcgccgtgctgcctgccggcagcctgaagacctcatcagc

EH135_r gccggcaggcagcacggcgcacgcgcgccgtctgtgcgatacggttgtcc

EH136_f ttgataacgtcttcgctactcgccatatgctgtctcctgttgtcgt

EH138_f aacacgttcccatttgaaaccttc

EH139_r agttcgttcctaggcctgtgctaaaacgcacagc

EH148_r tcgttttatgacgtcatggcgagtagcgaagacgtta

EH165_r cagttcgttcctaggcgtgggcttgcccctgtgctaaaa

EH166_r cagttcgttcctaggtacaaatccgcccctgtgctaaaacgcac

EH167_r cagttcgttcctaggtacaggcttatttttgtgctaaaacgcacagcgg

EH168_r cagttcgttcctaggtacaggcttgcccccacatcaaaacgcacagcggctgcg

EH169_r cagttcgttcctaggtacaggcttgcccctgtgctggggtgcacagcggctgcgcgaaat

EH170_r cagttcgttcctaggtacaggcttgcccctgtgctaaaacatgtggcggctgcgcgaaatctcgt

EH171_r cagttcgttcctaggtacaggcttgcccctgtgctaaaacgcacaataattgcgcgaaatctcgtgtttc

EH172_r cagttcgttcctaggtacaggcttgcccctgtgctaaaacgcacagcggccatataaaatctcgtgtttcatccacgaaat

EH173_r cagttcgttcctaggtacaggcttgcccctgtgctaaaacgcacagcggctgcgcggggcttcgtgtttcatccacgaaattactcac

EH174_r cagttcgttcctaggtacaggcttgcccctgtgctaaaacgcacagcggctgcgcgaaatcctacatttcatccacgaaattactcactaagatgga

EH175_f ataacgtcttcgctactcgccatatgcaacctcgcgcctgtttttat

EH178_r cagttcgttcctagg**c**acaggcttgcccctgtgctaa

EH179_r cagttcgttcctaggtacaggcttgcccctgtgctaaaacgcacagcggctgcgcgaaatctcatgtttcatccacgaaattactcactaagatg

EH182_r ggccggcctggcgcgccataaaacgaaaggctcagtcgaaagactgggcctttcgttttatgacgtcttacttcccaaactcactgaacctgg

EH183_f gctactcgccatatgtgtctcctggtccagatcga

EH190_r aggctcagtcgaaagactgggcctttcgttttatgacgtctcaaggaaacacggtcaggaca

EH191_f	gctactcgccatatggttcctcctccaacttcgct
EH196_r	ttctgttccaggggcccctgggatccatgttcgactggaatgatctgc
EH197_f	gatcgtcagtcagtcacgatgcggccgcctactcggctagcaactcgc
EH210_r	tcgttttatgacgtcggtgccgtctattattgattttatgaatgt
EH211_r	tcgttttatgacgtccggcaggcagcacggc
EH212_f	gctactcgccatatgctgtctcctgttgtcgttgtct
EH213_r	tcgttttatgacgtctccatctttaatttcaggttaacaatggt
EH217_f	tcgttttatgacgtccaggcactcgggcaggcgtga
EH258_r	cagtcgaaagactgggcctttcgttttatgacgtcctagcgcgcctcagcgtaat
EH259_f	ttgataacgtcttcgctactcgccatatggtctcctgtgaggcattgcgc
EH262_r	tcgttttatgacgtcctaggcgccctccgccgc
EH263_f	gctactcgccatatgggtgcgcggcgcctttg
EH264_r	tcgttttatgacgtctcagcgcaggtgcgacag
EH265_f	gctactcgccatatgtgcgcgtctccttgtaatgc
EH272_r	cagtcgaaagactgggcctttcgttttatgacgtcatgtagaaaatagcttattgaagggccgct
EH273_r	tcgttttatgacgtctcgaatttccttcgactaatcatcaattcttcgcattttatgc
EH274_r	tcgttttatgacgtcacacaaatcccgcattcttcagtt
EH275_r	cagtcgaaagactgggcctttcgttttatgacgtctcagccggcctctgcctc
EH276_f	ttgataacgtcttcgctactcgccatatgcttctctctcttggtgatggcgga
EH277_r	tcgttttatgacgtctcagcccacgtccgatgtc
EH278_f	gctactcgccatatgcgtggaggctccagtgcg
EH279_r	cagtcgaaagactgggcctttcgttttatgacgtctcaacgctgcggggcgcg
EH280_f	ttgataacgtcttcgctactcgccatatggcggtaaggactgtcggcaccggc
EH281_r	tcgttttatgacgtcctattcgcagaacgtcttcaccatctc
EH282_f	gctactcgccatatgttcgactccgcggtttgga
EH283_r	tcgttttatgacgtctcaggcctccagcggcac
EH284_f	gctactcgccatatgtttctcactaaaggggcgggaaataatattgt
EH285_r	tcgttttatgacgtctcagtcccctgggctgcc
EH286_f	gctactcgccatatggttctcctttcaggattccggtgct
EH287_r	tcgttttatgacgtctcactccccttgtcgcggatagatggtgatg
EH288_f	gctactcgccatatgcccggctggcgccgcaag
EH289_r	tcgttttatgacgtctcactgagcggaagcagg
EH290_f	gctactcgccatatggcagtctcctcgtcgttg
EH293_r	cggtttccctctagaaataattttggaattcaaaagatcttttaagaaggagatataaccatgaccaagcaat ctgcgga
EH294_f	tgcaacaggccccagtttctgcccatatccaatcaccctg

EH295_r	gtgattggatatgggcagaaactggggcctgttgca
EH296_f	gcgcacatttccccgaaaagtgccacctgggatgacctgcaggaaaaggccatccgtcaggatggcctt ctttataccagtggcgatttcacg
EH297_r	tcgttttatgacgtcggcgatggcggccctccg
EH298_r	tcgttttatgacgtcggatcaaatttaaaggattccgggacg
EH299_r	tcgttttatgacgtctttgatttcctcacatcagagattcacg
EH300_r	tcgttttatgacgtcagtctttttcacagtgttggcgtcttg
EH301_r	tcgttttatgacgtcgcctaaaagttatcgtatttaacagtaatcttaatttaca
EH302_r	tcgttttatgacgtccttcatatccaaaagcaattaaacacac
EH310_f	gctactcgccatatgtatatctccttcttaaaagatcttttgaattcccaaaaaaacgggtatggagaaac
EH311_r	tcgttttatgacgtcctacaggctgtctcccaccct
EH312_f	gctactcgccatatggcttcccatgacctaggg
EH344_f	actagtactgtttaaaccgctcacaattccaca
EH345_r	tcgttttatggcgcgccaggccggccaattagaaggccgccagagagg
EH346_r	tcgttttatgacgtcactggtccgcaagctttgtcac
EH360_r	cagtcgaaagactgggcctttcgttttatgacgtctcactccccttgtcgcggat
EH361_f	ttgataacgtcttcgctactcgccatatggtgattccttcttcaaaagaattccaggt
EH397_r	cagtcgaaagactgggcctttcgttttatgacgtcttgtctctcggcacggttc
EH398_f	ttgataacgtcttcgctactcgccatatgttccccttattcggccag
EH415_r	ggcgctcctgcggccggcctggcgcgccataaaacgaaaggctcagtcgaaagactgggcctttcgttt tatttacttgtacagctcgtccatgc
EH420_f	aaaagtgccacctgggatgacctgcaggaaaaggccatccgtcaggatggccttctggatccttaagca ccggtg
EH421_f	ccggagacacatatggtgagcaagggcgaggagctgt
EH422_r	cttgctcaccatatgtgtctccggctatgtct
EH423_f	aaataataaaaaagccggattaataatctggctttttatattctctctaatccgaaaacaagccgacg
EH424_r	attaatccggctttttttattatttcactgagcggaagcaggc
EH425_f	gctactcgccatcgggcagtctcctcgtcg
EH426_r	ggagactgcccgatggcgagtagcgaagacg
EH435_r	gccggcctggcgcgccataaaacgaaaggctcagtcgaaagactgggcctttcgttttatgacgtcattg gcaccatcggaattgc
EH436_f	ctactcgccatatgttcccctttattcggccag
EH441_r	ttctggcgctcctgcggccggcctggcgcgccataaaacga
EHseq003	catacgaactttgaaacgcatgaact
EHseq018	cgcagatcattccagtcgaaca
EHseq026_f	gtctgggtaccttcgtacgga

P009_ispS_f	gacggcgaaggagatatacatatgg
P010_ispS_r	tatacttaagtctagattcagcgctcgaacgg
N30_asp_f	tcgttttatgacgtcttacatgctgcccgggtgctac
N31_asp_r	gctactcgccatatgtcctcctagggggaatcgg
N36_f	tttatgacgtcttacgttgcgccttattccttcgg
N37_f	tttatgacgtcttaaccgccacgaagtaacgc
N38_r	atatatcatatgggtctccatagttgtggttggg
N39_r	atatgtacaattccctttttaatcatccggctcgtataatgtgtggagacttgaattcactagtttaactttaaga aggagatatatatatgctgaccctgaacctgac
N40_r	atatgtacaattccctttttaatcatccggctcgtataatgtgtggagacttgaattcactagtttaactttaaga aggagatatatatatgctgaccctgaatctgaccc

Supplementary Table 3 Putative transcription factor-based inducible systems mined from the genome of *C. necator* H16.

Locus tag regulator	Regulator name	Regulator family	Inducible promoter
H16_RS01325	OapR	MocR	$P_{H16\_RS01330}$
H16_RS03160	FdsR	LysR	$P_{H16\_RS03165}$
H16_RS05060	XdhR	LysR	$P_{H16\_RS05055}$
H16_RS05525	PhgR	LysR	$P_{H16\_RS05530}$
H16_RS08130	NahR	LysR	$P_{H16\_RS08125}$
H16_RS09795	BenM	LysR	$P_{H16\_RS09790}$
H16_RS10670	TtdR	LysR	$P_{H16\_RS10665}$
H16_RS13690	SauR	IclR	$P_{H16\_RS13695}$
H16_RS14025	KynR	AsnC	$P_{H16\_RS14030}$
H16_RS18300	HpdR	LysR	$P_{H16\_RS18295}$
H16_RS18360	PhhR	AsnC	$P_{H16\_RS18365}$
H16_RS23650	GabR	MocR	$P_{H16\_RS23655}$
H16_RS24180	HpdA	AsnC	$P_{H16\_RS24175}$
H16_RS27205	BadR	MarR	$P_{H16\_RS27200}$
H16_RS29645	N/A	XRE	$P_{H16\_RS29650}$
H16_RS30150	PcaQ	LysR	$P_{H16\_RS30145}$

Supplementary Table 4 Absolute normalised fluorescence values of *C. necator* carrying the various transcription factor-based inducible gene expression system-reporter constructs in the presence and absence of their proposed effector molecule. Single time-point fluorescence measurements were taken 6 h after supplementation with inducer at a final concentration of 5 mM. *C. necator* was grown in LB medium. The standard deviation was calculated from the absolute normalised fluorescence values of three biological replicates.

Plasmid identifier	Effector	Absolute normalised fluorescence (no inducer)	Absolute normalised fluorescence (+ inducer)	Induction factor
pEH147	3-Aminopropanoate	1774 ± 37	11413 ± 682	6.4
pEH134	Formate	1062 ± 41	2564 ± 190	2.4
pEH154	Xanthine	1044 ± 35	3776 ± 332	3.6
pEH155	Phenylglyoxylate	8 ± 2	139 ± 6	17.4
pEH042	Salicylate	90 ± 14	55943 ± 1955	622
pEH148	Benzoate	61 ± 4	64847 ± 4986	1063
pEH083	Tartrate	10 ± 2	374 ± 9	37.4
pEH157	Sulfonatoacetate	5 ± 1	772 ± 61	154
pEH136	L-Kynurenine	2178 ± 61	34165 ± 591	15.7
pEH010	3-Hydroxypropanoate	1389 ± 41	15615 ± 835	11.2
pEH137	L-Phenylalanine	1726 ± 53	2626 ± 127	1.5
pEH156	4-Aminobutanoate	4 ± 2	3 ± 1	0.8
pEH158	L-Tyrosine	15932 ± 343	38853 ± 2021	2.4
pEH159	Cyclohexanecarboxylate	9 ± 3	101 ± 11	11.2
pEH160	L-Glutamine	6 ± 1	4 ± 1	0.7
pEH161	3,4-Dihydroxybenzoate	109 ± 5	13778 ± 1212	126
pEH052	Acetoin	601 ± 38	1493 ± 101	2.5
pEH176	L-Arabinose	16 ± 2	5464 ± 256	342

Supplementary Table 5 Absolute normalised fluorescence values of *C. necator* carrying the redesigned 4-aminobutanoate (GABA)- and L-glutamine-inducible gene expression system-reporter constructs in the presence and absence of GABA and L-glutamine, respectively. Single time-point fluorescence measurements were taken 6 h after supplementation with inducer at a final concentration of 5 mM. *C. necator* was grown in LB medium. The standard deviation was calculated from the absolute normalised fluorescence values of three biological replicates.

Plasmid identifier	Effector	Absolute normalised fluorescence (no inducer)	Absolute normalised fluorescence (+ inducer)	Induction factor
pEH256	4-Aminobutanoate	180 ± 10	346 ± 47	1.9
pEH240	L-Glutamine	63 ± 4	69 ± 12	1.1

Supplementary Table 6 Absolute normalised fluorescence values of *C. necator* carrying the various inducible promoter-reporter gene constructs in the presence and absence of their proposed effector molecule. Single time-point fluorescence measurements were taken 6 h after supplementation with inducer at a final concentration of 5 mM. *C. necator* was grown in LB medium. The standard deviation was calculated from the absolute normalised fluorescence values of three biological replicates.

Plasmid identifier	Effector	Absolute normalised fluorescence (no inducer)	Absolute normalised fluorescence (+ inducer)	Induction factor
pEH101	3-Aminopropanoate	685 ± 16	2731 ± 55	4.0
pEH151	Formate	73 ± 13	73 ± 9	1.0
pEH167	Xanthine	162 ± 6	580 ± 11	3.6
pEH229	Phenylglyoxylate	12 ± 3	14 ± 1	1.2
pEH095	Salicylate	17 ± 2	4959 ± 211	292
pEH149	Benzoate	26 ± 1	10485 ± 157	403
pEH097	Tartrate	10 ± 1	46 ± 6	4.6
pEH168	Sulfonatoacetate	9 ± 2	24 ± 8	2.7
pEH152	L-Kynurenine	270 ± 16	2461 ± 44	9.1
pEH035	3-Hydroxypropanoate	87 ± 9	505 ± 35	5.8
pEH153	L-Phenylalanine	131 ± 2	192 ± 6	1.5
pEH266	4-Aminobutanoate	65 ± 5	76 ± 5	1.2
pEH169	L-Tyrosine	541 ± 23	1278 ± 54	2.4
pEH170	Cyclohexanecarboxylate	39 ± 1	150 ± 7	3.8
pEH171	3,4-Dihydroxybenzoate	53 ± 4	110 ± 4	2.1
pEH096	Acetoin	40 ± 4	50 ± 7	1.3

Supplementary Table 7 Absolute normalised fluorescence values of *E. coli* carrying the various transcription factor-based inducible gene expression system-reporter constructs in the presence and absence of their proposed effector molecule. Single time-point fluorescence measurements were taken 6 h after supplementation with inducer at a final concentration of 5 mM. *E. coli* was grown in LB medium. The standard deviation was calculated from the absolute normalised fluorescence values of three biological replicates.

Plasmid identifier	Effector	Absolute normalised fluorescence (no inducer)	Absolute normalised fluorescence (+ inducer)	Induction factor
pEH147	3-Aminopropanoate	319 ± 117	311 ± 88	1.0
pEH134	Formate	1067 ± 12	33722 ± 1223	31.6
pEH154	Xanthine	113 ± 24	165 ± 49	1.5
pEH155	Phenylglyoxylate	11 ± 1	130 ± 79	11.8
pEH042	Salicylate	109 ± 15	12697 ± 4341	116
pEH148	Benzoate	32 ± 1	141707 ± 4977	4428
pEH083	Tartrate	6 ± 2	93 ± 33	15.5
pEH157	Sulfonatoacetate	7 ± 3	8 ± 1	1.1
pEH136	L-Kynurenine	7 ± 1	10 ± 2	1.4
pEH010	3-Hydroxypropanoate	8 ± 3	14 ± 2	1.8
pEH137	L-Phenylalanine	247 ± 11	285 ± 51	1.2
pEH256	4-Aminobutanoate	7 ± 1	7 ± 1	1.0
pEH158	L-Tyrosine	33 ± 6	102 ±4	3.1
pEH159	Cyclohexanecarboxylate	445 ± 36	417 ± 12	0.9
pEH161	3,4-Dihydroxybenzoate	388 ± 95	57452 ± 8153	148
pEH052	Acetoin	7 ± 5	824 ± 84	118
pEH176	L-Arabinose	392 ± 121	508015 ± 60936	1296

Supplementary Table 8 Absolute normalised fluorescence values of *P. putida* carrying the various transcription factor-based inducible gene expression system-reporter constructs in the presence and absence of their proposed effector molecule. Single time-point fluorescence measurements were taken 6 h after supplementation with inducer at a final concentration of 5 mM. *P. putida* was grown in LB medium. The standard deviation was calculated from the absolute normalised fluorescence values of three biological replicates.

Plasmid identifier	Effector	Absolute normalised fluorescence (no inducer)	Absolute normalised fluorescence (+ inducer)	Induction factor
pEH221	3-Aminopropanoate	21987 ± 1369	27923 ± 3472	1.3
pEH202	Formate	25978 ± 4736	63325 ± 10034	2.4
pEH223	Xanthine	2102 ± 157	1915 ± 139	0.9
pEH218	Phenylglyoxylate	48 ± 2	601 ± 57	12.5
pEH194	Salicylate	266 ± 18	38100 ± 763	143
pEH205	Benzoate	1949 ± 855	320995 ± 29072	165
pEH199	Tartrate	22 ± 3	198 ± 45	9
pEH224	Sulfonatoacetate	36 ± 16	27 ± 7	0.8
pEH208	L-Kynurenine	94 ± 38	843 ± 99	9.0
pEH197	3-Hydroxypropanoate	978 ± 35	5901 ± 240	6.0
pEH210	L-Phenylalanine	67340 ± 2748	79256 ± 2008	1.2
pEH268	4-Aminobutanoate	1147 ± 98	1998 ± 102	1.7
pEH212	L-Tyrosine	7969 ± 3330	8642 ± 1436	1.1
pEH214	Cyclohexanecarboxylate	26 ± 3	38 ± 1	1.5
pEH216	3,4-Dihydroxybenzoate	1266 ± 51	95517 ± 803	75.4
pEH222	Acetoin	3865 ± 84	16067 ± 783	4.2
pEH220	L-Arabinose	62 ± 5	21137 ± 2616	341

Supplementary Table 9 Previously employed transcription factor-based inducible gene expression systems. The dataset may not be complete.

Effector	Inducible system	Source	Host	Induction factor	Application	Reference
Salicylate	NagR/P_{nagAa}	*Ralstonia* sp. U2	*Escherichia coli* RFM443	N/A[a]	Evaluation of reporter gene expression	[268]
Salicylate	SalR/P_{salA}	*Acinetobacter baylyi* ADP1	*Acinetobacter baylyi* ADP1	N/A[a]	Evaluation of cross-regulation	[269]
Salicylate	NahR/P_{sal}	*Pseudomonas putida* NAH7	*Escherichia coli* Top10	<120	Evaluation of reporter gene expression	[270]
Salicylate	NahR/P_{sal}	*Pseudomonas putida* NAH7	*Escherichia coli* SAL1	~1.8	Evaluation of reporter gene expression	[179]
Salicylate	NahR/P_{sal}	*Pseudomonas putida* KT2440	*Pseudomonas putida* KT2440	N/A[a]	Evaluation of reporter gene expression	[271]
Benzoate	BenM/P_{benA}	*Acinetobacter baylyi* ADP1	*Acinetobacter baylyi* ADP1	N/A[a]	Evaluation of cross-regulation	[269]
Benzoate	NahR/P_{sal}	*Pseudomonas putida* NAH7	*Escherichia coli* DH10B	N/A[a]	Selection of biocatalysts	[68]
Benzoate	BenR/P_{benA}	Metagneome	*Escherichia coli* JM109	~84	Screening for amidases	[170]
3-hydroxypropanoate	MmsR/P_{mmsA}	*Pseudomonas denitrificans* ATCC 13867	*Pseudomonas denitrificans* ATCC 13867	<100	Evaluation of reporter gene expression	[272]
3-hydroxypropanoate	MmsR/P_{mmsA}	*Pseudomonas denitrificans* ATCC 13867	*Escherichia coli*	14	Evaluation of reporter gene expression	[272]
3-hydroxypropanoate	MmsR/P_{mmsA}	*Pseudomonas denitrificans* ATCC 13867	*Pseudomonas putida*	28	Evaluation of reporter gene expression	[272]
L-phenylalanine	Tyr/P_{tyrP}	*Escherichia coli*	*Escherichia coli*	N/A[a]	Phenotype screening	[273]

L-phenylalanine	Tyr/P_{mtr}	*Escherichia coli* MG1655	*Escherichia coli* MG1655	4.5	Phenotype screening	[76]
L-tyrosine	TyrR/P_{aroF}	*Escherichia coli*	*Escherichia coli*	N/A[a]	Phenotype screening	[274]
3,4-dihydroxybenzoate	PobR/P_{pobA}	*Acinetobacter* sp. ADP1	*Escherichia coli* BL21	8-30	Evaluation of reporter gene expression	[171]
3,4-dihydroxybenzoate	PcaU/P_{pcaI}	*Acinetobacter baylyi* ADP1	*Pseudomonas putida* KT2440	>12	Evaluation of reporter gene expression	[275]
3,4-dihydroxybenzoate	PcaU/P_{pcaI}	*Acinetobacter baylyi* ADP1	*Escherichia coli* BL21	14	Evaluation of reporter gene expression	[276]

[a]Not available.

Supplementary Table 10 Absolute normalised fluorescence values of *E. coli* carrying the various inducible promoter-reporter gene constructs in the presence and absence of their proposed effector molecule. Single time-point fluorescence measurements were taken 6 h after supplementation with inducer at a final concentration of 5 mM. *E. coli* was grown in LB medium. The standard deviation was calculated from the absolute normalised fluorescence values of three biological replicates.

Plasmid identifier	Effector	Absolute normalised fluorescence (no inducer)	Absolute normalised fluorescence (+ inducer)	Induction factor
pEH101	3-Aminopropanoate	377 ± 67	328 ± 81	0.9
pEH151	Formate	1021 ± 280	958 ± 41	0.9
pEH167	Xanthine	155 ± 83	218 ± 82	1.4
pEH229	Phenylglyoxylate	3 ± 1	3 ± 1	1.0
pEH095	Salicylate	131 ± 20	220 ± 12	1.7
pEH149	Benzoate	378 ± 196	441 ± 109	1.2
pEH097	Tartrate	12 ± 2	12 ± 2	1.0
pEH168	Sulfonatoacetate	7 ± 5	8 ± 1	1.1
pEH152	L-Kynurenine	13 ± 4	17 ± 2	1.3
pEH035	3-Hydroxypropanoate	10 ± 2	11 ± 7	1.1
pEH153	L-Phenylalanine	305 ± 106	314 ± 111	1.0
pEH266	4-Aminobutanoate	9 ± 1	8 ± 1	0.9
pEH169	L-Tyrosine	37 ± 10	111 ± 50	3.0
pEH170	Cyclohexanecarboxylate	15230 ± 2135	25748 ± 4854	1.7
pEH171	3,4-Dihydroxybenzoate	3806 ± 419	5011 ± 815	1.3
pEH096	Acetoin	6 ± 2	7 ± 1	1.2

Supplementary Table 11 Absolute normalised fluorescence values of *P. putida* carrying the various inducible promoter-reporter gene constructs in the presence and absence of their proposed effector molecule. Single time-point fluorescence measurements were taken 6 h after supplementation with inducer at a final concentration of 5 mM. *P. putida* was grown in LB medium. The standard deviation was calculated from the absolute normalised fluorescence values of three biological replicates.

Plasmid identifier	Effector	Absolute normalised fluorescence (no inducer)	Absolute normalised fluorescence (+ inducer)	Induction factor
pEH196	3-Aminopropanoate	7665 ± 409	7887 ± 826	1.0
pEH203	Formate	4569 ± 824	4358 ± 499	1.0
pEH204	Xanthine	2320 ± 120	2008 ± 320	0.9
pEH257	Phenylglyoxylate	34 ± 13	103 ± 14	3.0
pEH195	Salicylate	300 ± 33	3681 ± 165	12.3
pEH206	Benzoate	2287 ± 307	14062 ± 1092	6.1
pEH200	Tartrate	52 ± 8	60 ± 6	1.2
pEH207	Sulfonatoacetate	20 ± 5	18 ± 3	0.9
pEH209	L-Kynurenine	66 ± 16	52 ± 5	0.8
pEH198	3-Hydroxypropanoate	801 ± 63	939 ± 53	1.2
pEH211	L-Phenylalanine	836 ± 32	1346 ± 789	1.6
pEH269	4-Aminobutanoate	92 ± 4	101 ± 8	1.1
pEH213	L-Tyrosine	377 ± 6	382 ± 26	1.0
pEH215	Cyclohexanecarboxylate	26722 ± 1051	25438 ± 1810	1.0
pEH217	3,4-Dihydroxybenzoate	1238 ± 32	1169 ± 29	0.9
pEH201	Acetoin	67 ± 12	122 ± 11	1.8

Supplementary Table 12 Induction of metabolites relative to the induction achieved using the primary effectors in *C. necator* and *E. coli*. Relative induction in *C. necator* was calculated using fluorescence values from the orthogonality screen performed in minimal medium. Relative induction in *E. coli* was calculated using single time-point fluorescence measurements from cultures grown in LB medium.

Compound	Inducible system	Relative induction in *C. necator* (in %)	Relative induction in *E. coli* (in %)
Acrylate	OapR/$P_{H16\_RS01330}$	68	2.7*
Phenylglyoxylate	FdsR/$P_{H16\_RS03165}$	52	0
Phenylglyoxylate	BenM/$P_{H16\_RS09790}$	17	0.19
Cyclohexanecarboxylate	BenM/$P_{H16\_RS09790}$	110	4.3
Cumate	BenM/$P_{H16\_RS09790}$	100	2.0
Acrylate	HpdR/$P_{H16\_RS18295}$	113	N/A
L-phenylalanine	HpdA/$P_{H16\_RS24175}$	77	0
Cumate	AcoR/$P_{H16\_RS19445}$	18	0

*The engineered β-alanine-inducible system was used for evaluation of orthogonality in *E. coli*.

Supplementary Table 13 Induction of metabolites relative to the induction achieved using the primary effectors in *C. necator* and *E. coli*. Relative induction in *C. necator* was calculated using fluorescence values from the orthogonality screen performed in minimal medium. Relative induction in *E. coli* was calculated using single time-point fluorescence measurements from cultures grown in LB medium.

Compound	Inducible system	Relative induction in *C. necator* (in %)	Relative induction in *E. coli* (in %)
3-Aminobutanoate	OapR/$P_{H16\_RS01330}$	58	44*
D,L-3-Amino-2-hydroxypropanoate	OapR/$P_{H16\_RS01330}$	33	0.4*
Nicotinate	FdsR/$P_{H16\_RS03165}$	113	0
Hypoxanthine	XdhR/$P_{H16\_RS05055}$	166	N/A
2,6-dihydroxybenzoate	NahR/$P_{H16\_RS08125}$	8	0.2
Hippurate	BenM/$P_{H16\_RS09790}$	21	0
Cyclohexene-1-carboxylate	BenM/$P_{H16\_RS09790}$	156	94
Catechol	BenM/$P_{H16\_RS09790}$	10	0.2
L-tryptophan	KynR/$P_{H16\_RS14030}$	66	N/A
Cylohexene-1-carboxylate	BadR/$P_{H16\_RS27200}$	80	N/A
Cyclopentane-1-carboxylate	BadR/$P_{H16\_RS27200}$	11	N/A
4-Hydroxybenzoate	PcaQ/$P_{H16\_RS30145}$	212	0.1
L-Lactate	AcoR/$P_{H16\_RS19445}$	29	0

*The engineered β-alanine-inducible system was used for evaluation of orthogonality in *E. coli*.

9.5 Manuscript: A genome-wide approach for identification and characterisation of metabolite-inducible systems

E. K. R. Hanko, A. C. Paiva, M. Jonczyk, M. Abbott, N. P. Minton & N. Malys (2020). *Nature Communications 11* (1), 1-14.

nature COMMUNICATIONS

ARTICLE

https://doi.org/10.1038/s41467-020-14941-6 OPEN

A genome-wide approach for identification and characterisation of metabolite-inducible systems

Erik K.R. Hanko [1], Ana C. Paiva [2], Magdalena Jonczyk [1], Matthew Abbott [1], Nigel P. Minton [1] & Naglis Malys [1✉]

Inducible gene expression systems are vital tools for the advancement of synthetic biology. Their application as genetically encoded biosensors has the potential to contribute to diagnostics and to revolutionise the field of microbial cell factory development. Currently, the number of compounds of biological interest by far exceeds the number of available biosensors. Here, we address this limitation by developing a generic genome-wide approach to identify transcription factor-based inducible gene expression systems. We construct and validate 15 functional biosensors, provide a characterisation workflow to facilitate forward engineering efforts, exemplify their broad-host-range applicability, and demonstrate their utility in enzyme screening. Previously uncharacterised interactions between sensors and compounds of biological relevance are identified by employing the largest reported library of metabolite-responsive biosensors in an automated high-throughput screen. With the rapidly growing genomic data these innovative capabilities offer a platform to vastly increase the number of biologically detectable molecules.

[1] BBSRC/EPSRC Synthetic Biology Research Centre (SBRC), School of Life Sciences, The University of Nottingham, Nottingham NG7 2RD, UK. [2] Centre for Biomolecular Sciences, School of Life Sciences, The University of Nottingham, Nottingham NG7 2RD, UK. ✉email: n.malys@gmail.com

Inducible gene expression systems execute a pivotal role in establishing a sustainable balance of gene expression and protein synthesis at the genome or single pathway/circuit level in response to changes in the intra- and extracellular environment. Such systems have been historically utilised for gene overexpression and protein production. Nowadays, inducible systems and their underlying genetic elements have become essential tools in synthetic biology[1]. Initially harnessed for the design of synthetic regulatory circuits[2], metabolite-responsive transcriptional regulators (TRs) and their cognate inducible promoters have received increasing attention due to their application as genetically encoded biosensors[3–6]. Although whole-cell biosensors or cell-free transcription/translation systems have advanced the fields of clinical diagnostics, environmental remediation, spatiotemporal regulation of signalling networks and metabolic engineering[7–11], the number of compounds that can be detected is still limited. Thus, to increase the diversity and to offer unique specificities, previously uncharacterised inducible systems must be sought and researched.

In instances, where an inducible system is to be found for a specific effector molecule, transcriptome analyses[12,13] and the evaluation of promoter libraries[14] have proven to be efficient strategies to discover effector-responsive promoters. They do not, however, exclude promoters that are indirectly activated nor guarantee identification of their associated TRs. Some of these issues may be solved by cloning sequence clusters containing TR-promoter pairs as demonstrated in the screening of metagenome libraries[15], but this methodology relies on the inducible system being functional in an organism different to the one it was sourced from. The reverse strategy relies on predicting the effector molecule based on genetic context[16] or comparative genomics[17,18]. This approach has successfully resulted in the identification of effectors and their corresponding TR-promoter pairs, but is limited to specific families of TRs and specific classes of compounds.

In this work, we address the deficiencies associated with the identification of metabolite-responsive-inducible systems by interconnecting information on ligand metabolism, TR genes and gene clusters responsible for the catabolism of the corresponding ligand. A generalised genome-wide approach is established to discover previously uncharacterised systems independent of their belonging to a specific family of regulators, the class of compounds they respond to or bacterial species utilised as a genetic resource. The discovered systems are validated for their response to proposed ligands and a comprehensive characterisation is performed. Specifically, we demonstrate their broad-host-range applicability in three industrially relevant microorganisms and address a typical issue that may arise, when employing a system in a host organism different to the one it was mined from. To facilitate forward engineering efforts, the identified inducible systems are parameterised and we demonstrate their utility for controlling orthogonal gene expression. We highlight their potential to be applied for investigation of metabolism and to expand the number of biologically detectable chemical species by evaluating the cross-reactivity between the library of biosensors and a comprehensive list of selected compounds. Finally, the biosensor responding to the industrially important intermediate compound β-alanine is applied to screen a library of L-aspartate 1-decarboxylase homologues and enzymes with superior activities are identified.

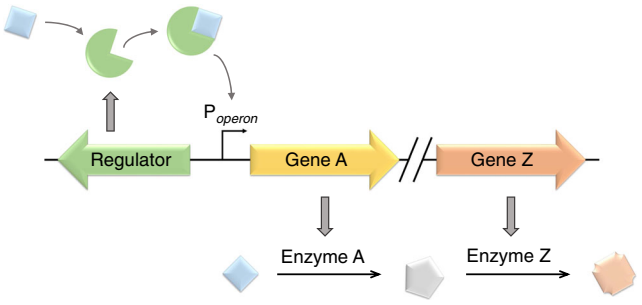


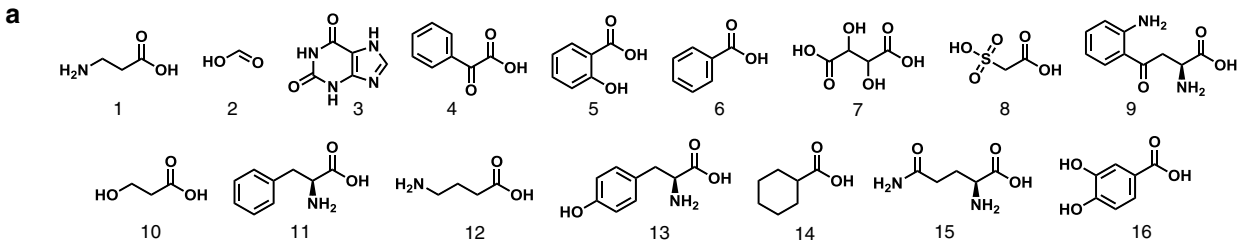
Fig. 1 Schematic illustration of a TR controlling expression of a metabolic cluster of genes. The primary substrate was proposed to be the ligand (light blue diamond). Note that the enzyme that converts the primary substrate into an intermediate product (grey regular pentagon) may be encoded by any of the genes in the operon. TR gene and protein are shown as green left arrow and pie, respectively. Metabolic cluster genes are yellow and orange right arrows. Product of metabolic conversion is shown as an orange octagon.

Results

A genome-wide approach to identify inducible systems. Transcription factor-based-inducible systems are composed of a TR protein and an inducible promoter, including TR and RNA polymerase binding sequences. In the systems that control gene clusters associated with metabolism and catabolism in particular, the level of gene expression from the inducible promoter is often controlled by the TR that responds to small effector molecules, also referred to as ligands. To make such systems universally applicable, all three components need to be identified: the regulator, the inducible promoter and its corresponding effector.

For the identification of inducible systems, we chose the highly conserved genetic arrangement, typical of LysR-type TRs (LTTRs), but not exclusive to other types of TRs, to serve as a platform for the genome scale approach. In this commonly occurring arrangement, TRs are transcribed in divergent orientation of target genes or operons[19,20]. Once a complete list of annotated genes belonging to one species is retrieved from GenBank[21] (www.ncbi.nlm.nih.gov), including information on coding strand orientation and protein function, it is screened for TRs that are oriented in the opposite direction of operons involved in metabolism of any or specific ligands. To constrain the search, the operon itself is to be composed of at least two genes encoding annotated catalytic functions associated with a distinct metabolic pathway. For each enzyme encoded by the operon, a list of metabolic substrates and products is extracted from The Comprehensive Enzyme Information System[22] (BRENDA, www.brenda-enzymes.org). By comparing potential metabolite substrates and products of each of the involved enzymes, the primary substrate is concluded that is likely to be metabolised by the operon-encoded enzymes. This compound was proposed to be the ligand that binds the TR, initiating expression of genes that encode the ligand-metabolising pathway enzymes (Fig. 1). By following this approach, the TR is assigned a role in metabolism solely based on its proximity to a metabolic cluster of genes. To exemplify the utility of the approach, it was applied in the chemolithoautotrophic bacterium *Cupriavidus necator* H16, known for its metabolic versatility and diverse gene expression regulation. Consequently, 16 putative metabolite-responsive transcription factor-based inducible gene expression systems were identified (Fig. 2a, b; Supplementary Table 1). Their genomic organisation is illustrated in Supplementary Fig. 1.

Validation of inducible systems and quantitative evaluation. By the genome-wide analysis identified native systems were cloned into a modular reporter vector to examine their response to the presence of the proposed compounds. The original genetic organisation was conserved by positioning the TR-coding sequence in the opposite orientation of a reporter gene



b

Effector	Regulator	Promoter
3-Aminopropanoate	OapR	$P_{H16\_RS01330}$
Formate	FdsR	$P_{H16\_RS03165}$
Xanthine	XdhR	$P_{H16\_RS05055}$
Phenylglyoxylate	PhgR	$P_{H16\_RS05530}$
Salicylate	NahR	$P_{H16\_RS08125}$
Benzoate	BenM	$P_{H16\_RS09790}$
Tartrate	TtdR	$P_{H16\_RS10665}$
Sulfonatoacetate	SauR	$P_{H16\_RS13695}$
L-Kynurenine	KynR	$P_{H16\_RS14030}$
3-Hydroxypropanoate	HpdR	$P_{H16\_RS18295}$
L-Phenylalanine	PhhR	$P_{H16\_RS18365}$
4-Aminobutanoate	GabR	$P_{H16\_RS23655}$
L-Tyrosine	HpdA	$P_{H16\_RS24175}$
Cyclohexanecarboxylate	BadR	$P_{H16\_RS27200}$
L-Glutamine	H16_RS29645	$P_{H16\_RS29650}$
3,4-Dihydroxybenzoate	PcaQ	$P_{H16\_RS30145}$
Acetoin	AcoR	$P_{H16\_RS19445}$
L-Arabinose	AraC	P_{araBAD}

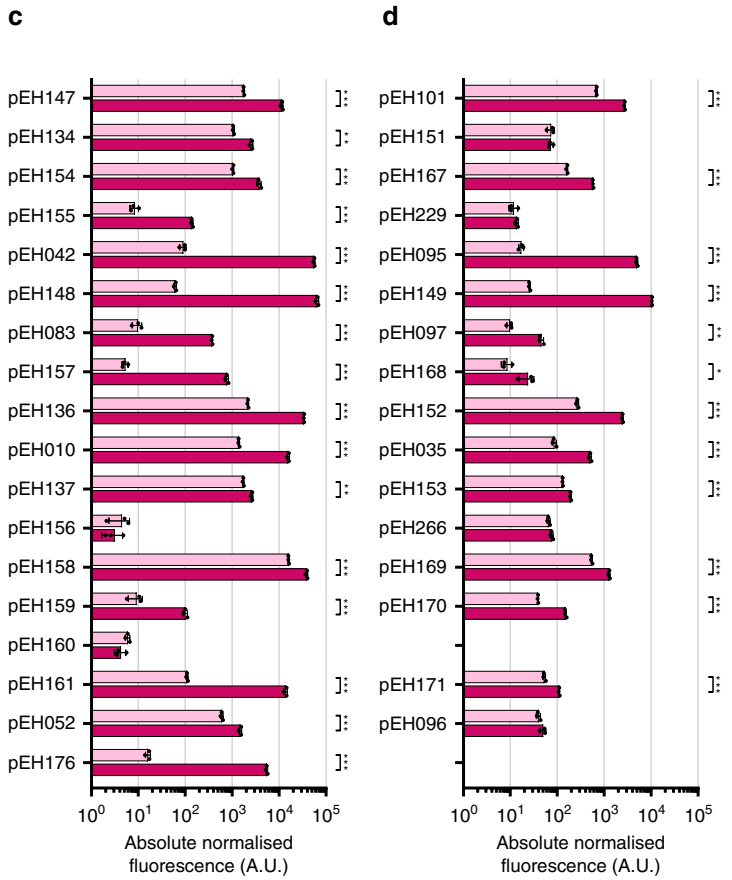


Fig. 2 Quantitative evaluation of native inducible systems. a Chemical structures of the proposed primary effector molecules: β-alanine (1), formic acid (2), xanthine (3), phenylglyoxylic acid (4), salicylic acid (5), benzoic acid (6), tartaric acid (7), sulfoacetic acid (8), L-kynurenine (9), 3-hydroxypropionic acid (10), L-phenylalanine (11), γ-aminobutyric acid (12), L-tyrosine (13), cyclohexanecarboxylic acid (14), L-glutamine (15) and 3,4-dihydroxybenzoic acid (16). **b** Summary of the identified inducible systems, including the inducible promoter, TR name and corresponding ligand. **c** Single time-point RFP fluorescence measurements (arbitrary units) of *C. necator* H16 carrying the transcription factor-based inducible gene expression systems composed of TR and inducible promoter in the same order as listed in **b**. The plasmids harbouring the individual system–reporter constructs are indicated. **d** Single time-point RFP fluorescence measurements of *C. necator* H16 carrying the inducible 'promoter only' implementations in the same order as listed in **b**. The plasmids harbouring the individual promoter–reporter gene constructs are indicated. Fluorescence output was determined in the absence of inducer (light magenta) and 6 h after extracellular supplementation with the corresponding effector to a final concentration of 5 mM (dark magenta). Data are mean ± SD, $n = 3$, $^*p < 0.05$, $^{**}p < 0.001$, $^{***}p \leq 0.0001$, unpaired two-tailed *t*-test. Source data are provided as a Source Data file.

encoding a monomeric red fluorescent protein (mRFP)[23] (Supplementary Fig. 2). This arrangement enables reporter protein synthesis in response to exogenous supplementation with the proposed ligand to be measured by fluorescence output. The *Escherichia coli* L-arabinose-, and the *C. necator* acetoin-inducible systems have been tested previously[24,25] and were included for comparative purpose (Fig. 2b).

C. necator strains carrying the inducible systems were grown in rich medium and fluorescence output of the logarithmically growing cells was quantified 6 h after addition of the inducer. It should be noted that the metabolic gene cluster, which is putatively controlled by KynR, despite converting L-tryptophan into anthranilic acid and L-alanine, the intermediate compound L-kynurenine was proposed to be the effector molecule and not the primary substrate L-tryptophan based on the thorough characterisation of KynR in other bacterial species[26]. Of the 16 analysed putative inducible systems, 14 showed a significant increase in mRFP protein synthesis after supplementation with their proposed effector molecules (Fig. 2c). Systems responding to 3-aminopropanoate (β-alanine) and phenylglyoxylate have not been reported, and highlight the potential of the developed approach for mining biosensors. Moreover, seven of the native systems exhibited higher levels of gene expression than the commonly used heterologous L-arabinose-inducible system in the presence of their corresponding effectors. Benzoate mediated the highest induction (factor of 1063-fold) and the highest absolute

normalised fluorescence with an expression level of >11-fold higher than AraC/P_{araBAD}. Low basal promoter activities were observed for metabolites that are neither involved in primary metabolism in *C. necator* nor likely to be present in the employed complex medium, including sulfonatoacetate, tartrate, cyclohexanecarboxylate and phenylglyoxylate.

However, the putative 4-aminobutanoate (GABA)- and L-glutamine-inducible systems showed no induction, even though their proposed ligands are involved in primary metabolism in *C. necator* and likely to be present in the rich medium. We hypothesised that translational start sites of the respective first gene in both operons are incorrectly annotated, resulting in reporter constructs with ineffectual 5′ untranslated regions. To test this hypothesis, the GABA- and L-glutamine-inducible systems were redesigned comprising the TR gene, the intergenic region, the first gene in the operon and the intergenic region preceding the second gene cloned upstream of the reporter (Supplementary Fig. 3a). Both inducible systems resulted in increased basal promoter activities (Supplementary Fig. 3b). Remarkably, GabR/$P_{H16\_RS23655}$ mediated a 1.9-fold induction of gene expression in the presence of GABA.

Increase in construct size can ultimately become the limiting factor in synthetic biology. As it has been reported, the transformation efficiency linearly decreases with increasing plasmid size[27]. Utilising an inducible system that is endogenous to the organism provides the advantage to possess a copy of the TR encoded in the genome enabling truncation of the controllable element to the sole inducible promoter, thus reducing construct size considerably. Despite that promoter activities under both uninduced and induced conditions generally decreased when the TR gene was removed from the multicopy episomal vector, a majority of inducible promoters significantly facilitated gene expression in the presence of their corresponding effectors (Fig. 2d). Whereas the 'promoter only' construct containing $P_{H16\_RS27200}$, which lacks the copy of the TR gene *badR*, exhibited a greater fluorescence level in the absence of effector than that of the BadR/$P_{H16\_RS27200}$-inducible system, suggests that BadR acts as a repressor, similarly to its homologue in *Rhodopseudomonas palustris*[28]. All other native TRs might be classed as activators or dual-function TRs. Furthermore, the induction factor was smaller for all of the systems, as a result of an altered ratio between transcription factor binding sites and available TR proteins. In fact, the promoters controlled by formate, phenylglyoxylate, GABA and acetoin showed no significant induction. Two of the inducible promoters, however, demonstrated an exceptionally strong activation of *rfp* expression: $P_{H16\_RS09790}$, responding to benzoate, and $P_{H16\_RS08125}$, which is activated by salicylate, mediated inductions by 403- and 292-fold, respectively. The 146 bp long intergenic region containing the benzoate-inducible promoter itself showed a stronger activation of gene expression than the majority of the 'complete' inducible systems, including the commonly used L-arabinose-inducible system (Fig. 2c). This characteristic highlights the potential of $P_{H16\_RS09790}$ to be employed as individual genetic element to control high levels of gene expression by reducing construct size by sevenfold. Even in cases, where a TR gene cannot be mapped to a cluster of genes involved in metabolism, the methodical approach described in this study can be employed for mining endogenous metabolite-inducible promoters.

Demonstration of broad-host-range applicability. To assess the potential of the constructed biosensors to be applied in other microorganisms, we evaluated the transcription factor-based-inducible systems that were mined from the genome of the β-proteobacterium *C. necator* in the industrially relevant γ-proteobacteria *E. coli* and *Pseudomonas putida*.

Regardless of their origin, the majority of systems responding to the 16 primary effectors, including acetoin, has never been tested in *E. coli* and *P. putida*. Thus, the broad-host-range applicability of the identified systems is highlighted by the outcome that a total of 8 and 12 of the 16 systems mediated a significant increase in reporter gene expression after inducer addition in *E. coli* (Fig. 3a) and *P. putida* (Fig. 3c), respectively. Three of them, activated by salicylate, benzoate and 3,4-dihydroxybenzoate induced by >75-fold in both tested microorganisms. Compared to systems that were sourced from other prokaryotes and tested in *E. coli* for controllable gene expression, PcaQ/$P_{H16\_RS30145}$ and BenM/$P_{H16\_RS09790}$ from *C. necator* outperform previously evaluated 3,4-dihydroxybenzoate- and benzoate-inducible systems by ~5- and 50-fold[15,29] (Supplementary Table 2). Specifically, benzoate mediated the highest induction (factor of 4428-fold) in *E. coli* and the highest absolute normalised fluorescence in *P. putida* with an expression level of >15-fold higher than AraC/P_{araBAD} demonstrating its potential for high-level inducible gene expression across different species.

To test for regulator-dependant orthogonality the 'promoter only' versions of the inducible systems were evaluated for controllable gene expression in *E. coli* and *P. putida*. Without the episomal copy of the TR gene, none of the promoters showed an increase in activity even in the presence of the effector in *E. coli* (Fig. 3b), whereas in *P. putida* RFP synthesis was significantly induced from the phenylglyoxylate-, salicylate-, benzoate- and acetoin-controllable promoters (Fig. 3d). Activation of reporter gene expression may be explained by cross-reactivity of chromosomally encoded TRs. A protein blast revealed homologues of BenM and AcoR to be encoded in the genome of *P. putida* (Supplementary Table 3), which might be able to activate gene expression from the *C. necator* benzoate- and acetoin-inducible promoters, respectively. NahR and PhgR homologues could not be identified indicating that transcriptional activation from the salicylate- and phenylglyoxylate-inducible promoters may result from unspecific TR binding.

Engineering the β-alanine-inducible system. Each inducible system harbours several functional genetic elements that independently contribute to its overall performance. These regulatory elements, including promoters and ribosome binding sites (RBSs), which control the expression of the TR gene and its associated regulon, may vary in their usage and efficiency across different species. As a consequence, inducible systems may perform differently when transferred from one into another organism. For example, the β-alanine-inducible system mediated a moderate activation of reporter gene expression in *C. necator*, whereas in *E. coli* and *P. putida* the fluorescence output remained at basal levels even in the presence of the inducer (Fig. 3a, c). β-Alanine is an intermediate compound for the synthesis of industrially relevant nitrogen-containing platform chemicals, including acrylamide, acrylonitrile and poly-β-alanine (also known as nylon-3)[30,31]. Furthermore, it is a precursor of the dipeptides carnosine and anserine that have been demonstrated to improve cognitive functions and physical capacities in humans[32,33]. To expand the host range and due to its usefulness as biosensor in synthetic biology and biotechnology applications, the β-alanine-inducible system from *C. necator* was pursued to be modified to enable its utilisation for gene expression control in *E. coli* and *P. putida*.

Although the system did not demonstrate an increase in RFP synthesis after addition of β-alanine, the promoter $P_{H16\_RS01330}$ showed the fifth highest activity of all evaluated systems under

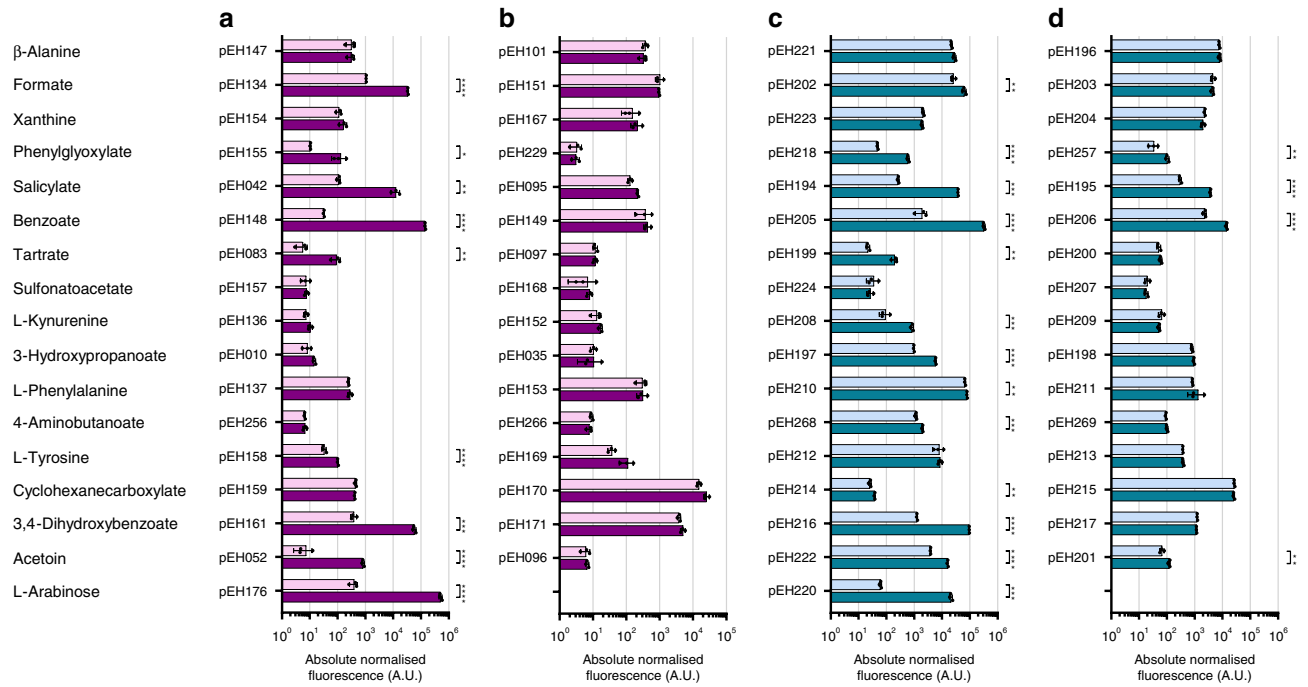


Fig. 3 *C. necator* native systems mediate controllable gene expression in *E. coli* and *P. putida*. Single time-point fluorescence measurements of **a**, **b** *E. coli* and **c**, **d** *P. putida* carrying the *C. necator*-inducible gene expression systems composed of TR and promoter **a**, **c** or 'promoter only' version **b**, **d**. The plasmids harbouring the individual system- or promoter-reporter gene constructs are indicated. RFP fluorescence output was determined in the absence of inducer (light purple and light blue) and 6 h after extracellular supplementation with the corresponding primary effector to a final concentration of 5 mM (dark purple and dark blue). Data are mean ± SD, $n = 3$, $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p \leq 0.001$, $^{****}p \leq 0.0001$, unpaired two-tailed *t*-test. Source data are provided as a Source Data file.

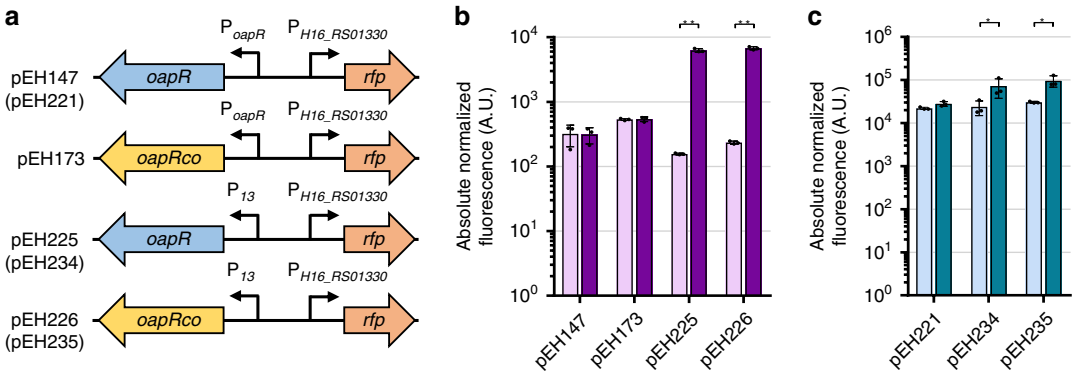


Fig. 4 Engineering the β-alanine-inducible system. a Schematic illustration of the different versions of the β-alanine-inducible system and their corresponding plasmid identifiers. Absolute normalised fluorescence of **b** *E. coli* and **c** *P. putida* carrying different versions of the β-alanine-inducible system-reporter construct in the absence (light purple and light blue) and presence (dark purple and dark blue) of 5 mM β-alanine. Single time-point RFP fluorescence measurements were taken 6 h after effector addition. Data are mean ± SD, $n = 3$, $^{*}p < 0.05$, $^{**}p < 0.0001$, unpaired two-tailed *t*-test. Source data are provided as a Source Data file.

uninduced conditions in *E. coli* (Fig. 3a). This indicated that the regulatory elements of $P_{H16\_RS01330}$ are functional in *E. coli* and that the lack of induction is more likely to be attributed to inducer uptake or regulator gene expression. A poor ligand transport as a cause of the absence of induction could be excluded, as β-alanine has been reported to be actively taken up by the cells[34]. Strikingly, the GC-content of the TR-coding sequence (71%) is significantly higher than the GC-content of the *E. coli* K12 genome (51%). To rule out that a high GC-content impairs TR synthesis, *H16_RS01325* (*oapR*) was codon optimised for *E. coli* codon usage (pEH173, Fig. 4a). However, this modification did not improve the response of the system to β-alanine (Fig. 4b).

Subsequently, to ensure that the regulator is expressed in *E. coli* and *P. putida*, the *C. necator* native promoter of the TR was replaced by a host-specific promoter. The 24 bp DNA sequence upstream of the *oapR* translational start site was replaced by the core sequence of a medium-strength insulated constitutive promoter, P_{13} (ref. [35]), including the phage T7 gene *10* RBS. The substitution was implemented in both plasmids containing the native (pEH147 and pEH221) and the codon-optimised *oapR* (pEH173, Fig. 4a). The addition of β-alanine resulted in a 40- and 29-fold increase in fluorescence output for *E. coli* cultures carrying pEH225 and pEH226, respectively (Fig. 4b). This suggests that the original promoter CnP_{oapR} is not functional in *E. coli* and that codon optimisation might even lead to a lower TR

synthesis rate, which results in a decreased induction level. Moreover, a lower background reporter gene expression under uninduced conditions in case of pEH225 and pEH226 indicates that OapR, a member of the MocR family, may act as dual-function TR, repressing transcription of *oapTD* in the absence of β-alanine, but acting as activator in its presence. The dual mode of action has been observed for other members of this family of TRs[36]. In *P. putida*, the promoter and RBS exchange also resulted in a threefold induction of gene expression for both versions the native (pEH234) and the codon-optimised (pEH235) *oapR* (Fig. 4a, c). In contrast to *E. coli*, however, induction levels are significantly lower that may be attributed to a higher basal promoter activity in *P. putida*.

Parameterisation of inducible systems. The 16 functional native inducible systems, including the acetoin-inducible system, were subsequently evaluated for their dose–response, dynamic range, induction homogeneity and orthogonality. To obtain a preliminary overview of the kinetics of induction and to assess the effect of the extracellularly added ligand on cell growth, *C. necator* strains carrying plasmids with inducible systems were grown in minimal medium (MM) supplemented with the corresponding inducer at a final concentration of 5 mM, and fluorescence and absorbance were monitored over time. For all inducible systems, increasing fluorescence above the level of the uninduced culture was observed within 30 min after the effector had been added (Supplementary Fig. 4). Maximum fluorescence was reached within 2–4 h for most of the systems with exception of the sulfonatoacetate-, L-phenylalanine-, and cyclohexanecarboxylate-inducible systems continuing to produce the reporter protein after 6 h. Most of the compounds had a beneficial effect on growth and no toxicity was observed for any effector at the tested concentration of 5 mM (Supplementary Fig. 5). L-tyrosine and 3,4-dihydroxybenzoate had the most significant impact on growth resulting in more than a twofold increase in cell density most likely due to ligand catabolism. Therefore, the effector consumption plays an important role in the kinetics of induction. In order to parameterise the identified systems, assumptions must be implemented that account for these factors, including ligand uptake and metabolism.

The dose–response curve of a metabolite-responsive-inducible system describes the level of gene expression as a function of ligand concentration, thus indicating the range of effector concentration in which the inducible system is able to operate. It provides key parameters that aid in part selection and the computational design of synthetic circuits. To simplify mathematical modelling approaches, effector concentrations are usually considered constant, mediating a sustained gene expression throughout the course of cell growth. In this study, however, the ligands are metabolised by *C. necator* resulting in a decrease in gene expression to basal levels once the inducer has been depleted. Therefore, the time point, at which the reporter output is correlated with the ligand concentration, must be chosen carefully. By postulating that the inducer-metabolising enzymes are not synthesised faster than the primary induction of reporter gene expression, and to account for the minimum amount of time required for RFP synthesis and maturation, the minimal induction interval of 80 min was determined. *C. necator* strains harbouring the inducible systems were grown in MM and reporter gene expression was monitored after supplementation with the corresponding inducer at a wide range of concentrations over time. The dose–responses were obtained by plotting the relative normalised fluorescence values of the 80-min minimal induction interval as a function of inducer concentration (Fig. 5a). Data points were fit using a Hill function (see Methods section), taking into account the basal level of fluorescence output of the uninduced cells. Consequently, on the basis of the mathematically modelled dose–response curve, key parameters that distinguish one inducible system from another were obtained (Table 1).

One of the most important parameters when choosing an inducible system to tightly control different levels of gene expression is the dynamic range. It is defined as the maximum level of reporter output relative to basal expression levels (see Methods section, formula (3)). We found that the salicylate-inducible system has the highest dynamic range of the evaluated native systems followed by SauR/$P_{H16\_RS13695}$ and TtdR/$P_{H16\_RS10665}$. In general, the dynamic range is higher than the induction level of cells grown in rich medium at an effector concentration of 5 mM (Fig. 2c). This effect does not apply to the benzoate- and 3,4-dihydroxybenzoate-inducible systems, which show a lower dynamic range in MM. Whereas a lower induction level in complex medium might be the consequence of structurally similar molecules that are able to interact with the respective TR, the lower dynamic range in MM might be attributed to catabolic repression as it has been demonstrated to be the case for both benzoate and 3,4-dihydroxybenzoate in *P. putida*[37]. In addition to a high dynamic range in MM, NahR/$P_{H16\_RS08125}$ has the lowest K_m of all evaluated native inducible systems. This parameter is defined as the inducer concentration that mediates half-maximal reporter output, suggesting that only small quantities of salicylate are needed to induce the system. Similar effector concentrations have been shown to mediate gene expression from the *P. putida* salicylate-inducible system NahR/P_{sal}[38]. In contrast to most of the inducible systems that operate in the µM-range, GabR, TtdR and SauR seem to respond to effector concentrations three to five orders of magnitude higher than NahR. Moreover, in case of GabR, the K_m appeared to be higher than the concentration of GABA the growth medium can be supplemented with. It should be noted that the extracellular effector concentration may not necessarily correlate with the ligand concentration inside the cell ultimately dictating the level of gene expression. Ligand uptake limitations may therefore result in inaccurate parameters, as it might be the case for the GABA-, tartrate- and sulfonatoacetate-inducible systems. For the other 13 systems ligand uptake is assumed not to be limiting. The Hill coefficient h indicates the range of inducer concentration over which the system results in a change in reporter output. Inducible systems with a low Hill coefficient, such as SauR/$P_{H16\_RS13695}$, PhhR/$P_{H16\_RS18365}$ or HpdR/$P_{H16\_RS18295}$ exhibit a flatter dose–response function, indicating that gene expression is tuneable over a wider range of inducer concentration. On the contrary, systems with a higher Hill coefficient, including BenM/$P_{H16\_RS09790}$ and AcoR/$P_{H16\_RS19445}$, show a steeper dose–response function suggesting that they behave more like an on/off switch.

The homogeneity of induction was evaluated by flow cytometry. This method allows to determine whether intermediate inducer concentrations give rise to subpopulations of uninduced and fully induced cells. Their existence may indicate a more complex type of transcriptional regulation or inducer transport limitations[39]. Cell cultures of *C. necator* carrying the 16 functional native system–reporter constructs were subjected to inducer concentrations corresponding to 50 and 95% of the maximum level of fluorescence output b_{max} (except in case of the phenylglyoxylate-, tartrate-, sulfonatoacetate- and GABA-inducible systems; refer to the legend of Fig. 5b; Supplementary Fig. 6). The reporter output was measured 2 h after the inducer had been extracellularly added. Each of the 16 evaluated systems demonstrated a unimodal induction behaviour after addition of the corresponding ligand at both medium and nearly saturating concentrations. Generally, the fluorescence

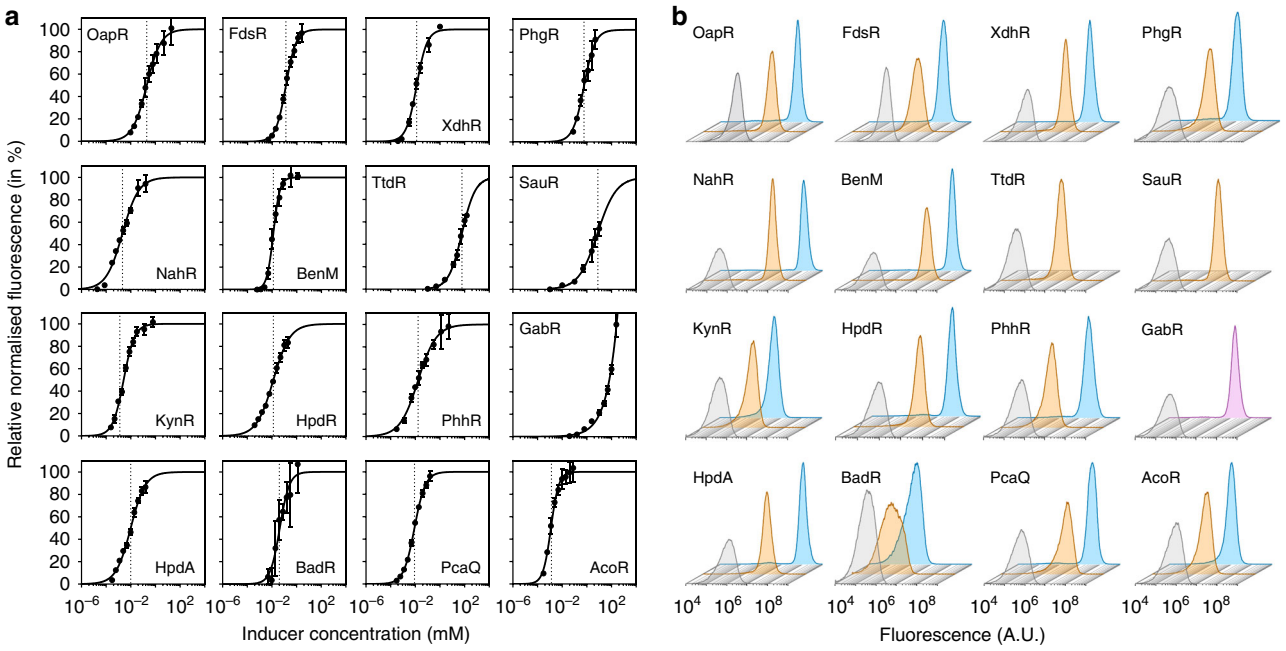


Fig. 5 Response function and induction homogeneity. a Relative normalised fluorescence of *C. necator* carrying the various inducible system–reporter constructs in response to different concentrations of their corresponding primary inducers. Measurements were taken 80 min after the inducer had been extracellularly added. The dose–responses were fit using a Hill function (see Methods section). The maximum level of reporter output b_{max} was set to 100% (except in case of the GABA-inducible system where the absolute normalised fluorescence corresponding to the highest GABA concentration tested was set to 100%). The inducer concentration that mediates half-maximal reporter output K_m is indicated by a dotted line. Data are mean ± SD, $n = 3$, non-linear regression. Source data are provided as a Source Data file. **b** Evaluation of induction homogeneity by flow cytometry. The fluorescence intensity of 100,000 individual cells was determined for each inducible system–reporter construct 2 h after extracellular addition of inducer. Uninduced cells (grey) are compared to cultures supplemented with their cognate effector at final concentrations corresponding to 50% (orange) and 95% (blue; 90% in case of the phenylglyoxylate-inducible system) of the maximum level of reporter output b_{max}. The tartrate- and sulfonatoacetate-inducible systems were only evaluated for induction homogeneity at 50% of b_{max} due to solubility limits and inducer toxicity, respectively. Since b_{max} could not be calculated for the GABA-inducible system, induction homogeneity was determined using a final concentration of 250 mM (purple).

Table 1 Parameters of the native inducible systems.

Inducible system	Inducer	Dynamic range, in -fold[a]	K_m[b]	h[c]
OapR/$P_{H16\_RS01330}$	β-Alanine	8.0 ± 0.4	201 ± 24 µM	0.75 ± 0.05
FdsR/$P_{H16\_RS03165}$	Formate	4.5 ± 0.2	130 ± 7 µM	1.05 ± 0.04
XdhR/$P_{H16\_RS05055}$	Xanthine	16.0 ± 1.0	13.0 ± 1.4 µM	1.03 ± 0.10
PhgR/$P_{H16\_RS05530}$	Phenylglyoxylate	144.8 ± 106.5	595 ± 133 µM	0.96 ± 0.13
NahR/$P_{H16\_RS08125}$	Salicylate	650.8 ± 317.4	2.12 ± 0.49 µM	0.66 ± 0.08
BenM/$P_{H16\_RS09790}$	Benzoate	74.1 ± 10.7	12.6 ± 0.8 µM	1.64 ± 0.14
TtdR/$P_{H16\_RS10665}$	Tartrate	370.6 ± 232.2	61.5 ± 22.9 mM	0.80 ± 0.09
SauR/$P_{H16\_RS13695}$	Sulfonatoacetate	395.4 ± 193.7	7.3 ± 4.9 mM	0.57 ± 0.06
KynR/$P_{H16\_RS14030}$	L-kynurenine	26.9 ± 2.1	2.64 ± 0.18 µM	0.98 ± 0.05
HpdR/$P_{H16\_RS18295}$	3-Hydroxypropanoate	25.2 ± 1.7	13.2 ± 1.9 µM	0.64 ± 0.03
PhhR/$P_{H16\_RS18365}$	L-phenylalanine	9.0 ± 0.8	16.9 ± 2.2 µM	0.59 ± 0.04
GabR/$P_{H16\_RS23655}$	GABA	ND	ND	ND
HpdA/$P_{H16\_RS24175}$	L-tyrosine	38.3 ± 5.0	9.9 ± 2.2 µM	0.72 ± 0.07
BadR/$P_{H16\_RS27200}$	Cyclohexanecarboxylate	11.4 ± 10.9	43.8 ± 13.0 µM	1.11 ± 0.28
PcaQ/$P_{H16\_RS30145}$	3,4-Dihydroxybenzoate	77.2 ± 29.8	8.63 ± 0.40 µM	0.98 ± 0.03
AcoR/$P_{H16\_RS19445}$	Acetoin	11.1 ± 1.6	1.23 ± 0.07 µM	1.36 ± 0.10

Data are mean ± SD, $n = 3$. Source data are provided as a Source Data file. ND - not determined.
[a]Dynamic range is defined as the -fold increase in fluorescence calculated by dividing the maximum level of fluorescence output by the basal level of fluorescence output.
[b]K_m represents the inducer concentration at which the half-maximal activation of the inducible system is achieved.
[c]h—Hill coefficient.

distribution of the uninduced cells is wider than the distribution in the presence of effector. This generalisation does not apply to the cyclohexanecarboxylate-inducible system that may be a further indicator of it being a repressor-based type of inducible system.

Orthogonality of inducible systems. To establish whether any of these systems can be used in combination to independently control expression of more than a single gene the activity of the ligands against non-cognate promoters was evaluated. A total of 21 inducible systems was selected to test for orthogonality in

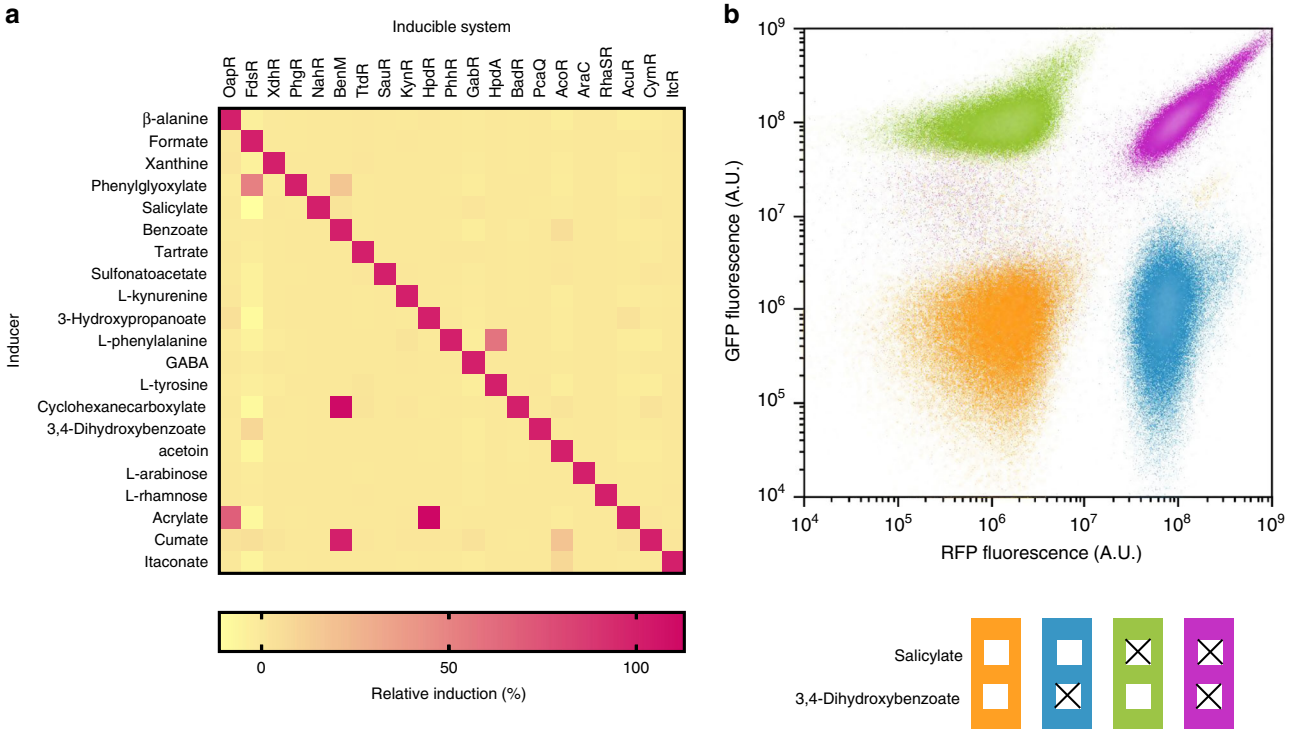


Fig. 6 Orthogonality of inducible systems. a Cross-reactivity of a set of 21 inducible systems and their corresponding primary inducers. The heat map illustrates induction of reporter gene expression in the presence of the metabolite (in %) relative to the induction mediated by the corresponding primary effector. Measurements were taken 6 h after supplementation with inducer at a final concentration of 5 mM. Data are mean, $n = 4$. Source data are provided as a Source Data file. **b** Fluorescence output of individual cells of *C. necator* carrying pEH263 measured by flow cytometry. The vector pEH263 contains the 3,4-dihydroxybenzoate- and salicylate-inducible systems controlling expression of *rfp* and *egfp*, respectively. Fluorescence was determined in the absence of inducer (orange), in the presence of 3,4-dihydroxybenzoate (blue), salicylate (green), and both 3,4-dihydroxybenzoate and salicylate (purple). Effector concentrations correspond to their respective K_m's and fluorescence was measured 2 h after inducer addition.

C. necator. It includes the 16 native and the previously characterised heterologous L-arabinose-, L-rhamnose-, acrylate-, cumate- and itaconate-inducible systems[35,40].

To screen the 441 combinations of inducer and biosensor for cross-reactivity, we employed an automated platform (see Methods section). Strains of *C. necator* harbouring the inducible systems were grown in MM and transferred to a 96-well microtiter plate format. After the inducers had been extracellularly added, cells were cultured for 6 h before RFP fluorescence and cell density were measured. A total of 15 of the 21 inducible systems exhibited a strong affinity to their primary ligands, showing <5% cross-reactivity of non-target metabolites relative to the fluorescence output mediated by the primary effector (Fig. 6a, Supplementary Table 4). The remaining six inducible systems were activated by one or more metabolites other than their cognate inducers. This cross-reactivity may be the result of either structural resemblance, a metabolic relationship or a combination thereof. For example, L-phenylalanine is converted into L-tyrosine by the phenylalanine 4-monooxygenase PhhA (H16_RS18365). Therefore, activation of the L-tyrosine-inducible system by L-phenylalanine is more likely to be due to biological conversion of the added compound into the primary effector rather than direct interaction of L-phenylalanine with HpdA. However, since the difference between the two molecules lies in a singly hydroxyl-group, an induction by structural resemblance cannot be entirely ruled out. The same applies to phenylglyoxylate that activated the benzoate-inducible system. Phenylglyoxylate is converted by *C. necator* into benzoate via a two-step reaction with benzaldehyde as intermediate compound. During the first reaction, carbon dioxide is generated that may also explain induction of the formate-inducible system by the structurally dissimilar phenylglyoxylate. Since carbon dioxide can subsequently be converted into formate[41], it is rational to postulate that the FdsR/$P_{H16\_RS03165}$-inducible system is activated in this case by its primary inducer. In addition to phenylglyoxylate, the benzoate-inducible system was activated by extracellular supplementation with cyclohexanecarboxylate and cumate. Cyclohexanecarboxylate shares both a structural resemblance to benzoate and downstream degradation pathways in *C. necator*, which makes it difficult to conclude its cause of cross-reactivity. In cases like these, we can take advantage of the system's transferability. Moving the inducible system from one organism into another host with a dissimilar metabolism may allow to distinguish more easily between induction by structural resemblance and metabolic relationship. To investigate the cause of cross-reactivity of the compounds that resulted in a fluorescence output of >10% relative to the primary inducer in *C. necator*, their induction behaviour was evaluated in *E. coli*. Single time-point fluorescence measurements of *E. coli* revealed that the benzoate-inducible system is activated by addition of cyclohexanecarboxylate and cumate (Supplementary Table 4). *E. coli* has not been reported to metabolise any of these compounds suggesting that cross-reactivity is caused by structural resemblance.

More difficult to explain is the cause of activation of the β-alanine- and 3-hydroxypropanoate (HP)-inducible systems by acrylate. It has been shown that *C. necator* is able to degrade acrylate[35]; however, relatively little is known about its metabolism. Activation by structural resemblance to the primary effector is less likely as addition of acrylate to *E. coli* carrying the engineered β-alanine-inducible system (pEH225) only resulted in

a relative induction of 2.7% in contrast to 68% in *C. necator* (Supplementary Table 4). Therefore, activation by structural resemblance of a degradation product or direct conversion into the primary effectors 3-HP and β-alanine in *C. necator* may be more likely. Acrylate can be activated into acryloyl-CoA by acyl CoA:acetate/3-ketoacid CoA transferase (H16_RS22005/H16_RS22010) or by the propionate CoA transferase Pct (H16_RS13535)[42], which, as it has been proposed by Peplinski et al.[43], can be converted into 3-HP via its CoA intermediate. However, pathways from acrylate or 3-HP to β-alanine or a metabolic intermediate, which is able to activate the β-alanine-inducible system, will have to be elucidated.

Based on the results of the cross-reactivity screen, two orthogonal inducible systems were employed to independently control expression of two fluorescent protein reporter genes. Plasmid pEH263 was constructed containing *rfp* under control of the 3,4-dihydroxybenzoate-, and *egfp* under control of the salicylate-inducible system (Supplementary Fig. 7). Cultures of *C. necator* carrying pEH263 were left uninduced, subjected to the individual inducers or the combination of both at final concentrations corresponding to K_m (Table 1). The output of the two non-overlapping fluorescent proteins was determined by flow cytometry 2 h after the inducer/inducers had been extracellularly added. Employing every possible inducer combination, four distinct cell states could be observed (Fig. 6b, Supplementary Fig. 8). RFP and enhanced green fluorescent protein (eGFP) fluorescence in the absence of both inducers remained at background levels comparable to the single-system implementations (orange population). Addition of 3,4-dihydroxybenzoate resulted in synthesis of RFP only as represented by the blue population. Similarly, the presence of salicylate activated expression of *egfp* but not *rfp* (green population). The final cell state is represented by the purple population, where both inducers were added to mediate the simultaneous expression of both fluorescent protein reporter genes.

In conclusion, the identified native systems can be used in combination to independently control expression of multiple genes expanding the list of available switches in synthetic circuit design. Importantly, inducible systems for structurally similar (phenylglyoxylate, salicylate, cyclohexanecarboxylate and 3,4-dihydroxybenzoate) and distinctive (xanthine, tartrate, sulfonatoacetate and GABA) compounds were demonstrated to be fully orthogonal.

Screening structurally and metabolically related compounds. TRs often recognise molecules that are structurally similar to their primary effectors. For example, isopropyl β-D-1-thiogalactopyranoside is a commonly used structural analogue of allolactose, employed to control the expression of genes regulated by LacI. Here, the TR's specificity was investigated to identify previously uncharacterised ligand–TR interactions and consequently extend the biosensor application to the detection of structurally similar or metabolically related compounds. To do so, the library of 21 native and heterologous inducible systems was screened against 46 compounds using an integrated robotic platform, as described in the previous section. Of the 46 metabolites, 12 demonstrated an induction of at least 5% relative to the induction mediated by the corresponding primary effector and an absolute induction factor of at least five (Fig. 7, Supplementary Table 5). Similarly to the former orthogonality screen (Fig. 6a), it should be noted that activation of reporter gene expression by the extracellularly added compounds likely indicates a structural resemblance to the primary ligand, a metabolic relationship or a combination thereof. To shed light on their cause of system activation, cultures of *E. coli* carrying the functional formate-, salicylate-, benzoate-, 3,4-dihydroxybenzoate-, acetoin- and β-alanine-inducible systems, active in this host organism, were evaluated for the same cross-reactivity (Supplementary Table 5).

Based on the results of this screen, 11 out of 12 metabolites may be classed into four groups. The first group comprises the compound/regulator pairs 3-aminobutanoate/OapR, DL-3-amino-2-hydroxypropanoate/OapR and cyclohexenecarboxylate/BenM. Neither of the three compounds has been reported to be metabolised by *E. coli* suggesting that they directly act as ligands. The response of the OapR/$P_{H16\_RS01330}$-inducible system to non-natural compounds, such as 3-aminobutanote and 3-amino-2-hydroxypropanoate, strongly supports our initial claim of the structurally nearly identical β-alanine (3-aminopropanoate) being the primary ligand of OapR. The other three groups contain the compound/regulator pairs that did not result in an induction in *E. coli*. For example, nicotinate, N-benzoylglycine, L-tryptophan and 4-hydroxybenzoate fall into the category of metabolites that are likely to be converted into the primary effectors in *C. necator* mediating gene expression from the formate-, benzoate-, L-kynurenine- and 3,4-dihydroxybenzoate-inducible system, respectively. None of these catabolic pathways exist in *E. coli*, which may explain their lack of induction. In the case of the hypoxanthine inducing the XdhR/$P_{H16\_RS05055}$-controllable system, it is less clear whether hypoxanthine itself interacts with XdhR, due to structural resemblance to xanthine, or if activation of gene expression is mediated by its catabolic product, the primary effector. In *Streptomyces coelicolor*, hypoxanthine was not able to bind the regulator of the gene cluster encoding xanthine dehydrogenase enzyme[44]. However, in contrast to the *S. coelicolor* XdhR, which belongs to the TetR family of TRs[45], the *C. necator* XdhR is a LTTR that may operate in a different manner. The last group comprises the remaining compounds that could not be confirmed in *E. coli*, but are likely to induce because of structural resemblance to the primary ligand. Direct interaction of cyclohexenecarboxylate and cyclopentanecarboxylate with BadR could not be confirmed due to dysfunctionality of BadR/$P_{H16\_RS27200}$ in *E. coli*. The other two compounds, including 2,6-dihydroxybenzoate and catechol, might not be taken up by, or diffused into, *E. coli* cells and the activation of reporter gene expression by metabolically related compounds may be excluded as their consecutive degradation products, resorcinol and *cis*,*cis*-muconate, respectively, were not able to induce the salicylate- and benzoate-inducible systems in *C. necator* (Fig. 7).

Lastly, significant activation of the acetoin-inducible system by L-lactate was observed in *C. necator*, but not in *E. coli*. Since there is no characterised metabolic pathway for L-lactate to be converted into acetoin, and the AcoR/$P_{H16\_RS19445}$-inducible system is highly activated by acetoin in both bacterial species, the response to L-lactate cannot be explained by either structural similarity or metabolic association.

Biosensor-assisted screening of enzyme variants. The β-alanine biosensor was employed to screen a library of six L-aspartate 1-decarboxylase (PanD) homologues for their ability to convert L-aspartate into β-alanine. The whole-cell enzymatic conversion of L-aspartate into β-alanine has been previously demonstrated in *E. coli*[46]. In this study, *C. necator* was selected as biocatalyst as it may allow for the autotrophic biosynthesis of this industrially relevant intermediate compound. To comparatively screen homologues of PanD, plasmids were constructed that contain the original β-alanine biosensor and each a *panD* gene under control of the L-arabinose-inducible promoter (Supplementary Fig. 9).

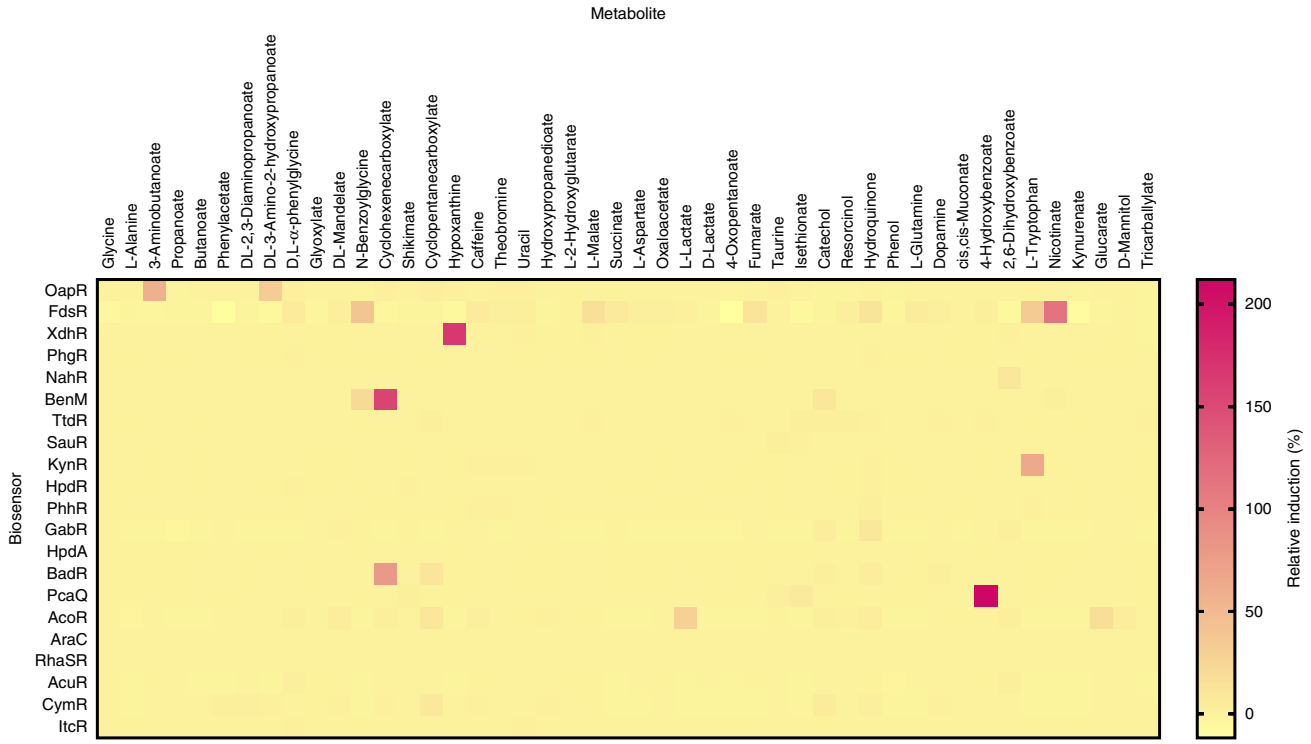


Fig. 7 Screening of structurally similar and metabolically related compounds. Cross-reactivity of a set of 21 inducible systems and a library of 46 compounds. The heat map illustrates induction of reporter gene expression in the presence of metabolite (in %) relative to the induction mediated by the corresponding primary effector. Measurements were taken 6 h after supplementation with inducer at a final concentration of 5 mM (except in case of L-2-hydroxyglutarate that was added at a final concentration of 0.5 mM). Data are mean, $n = 4$. Source data are provided as a Source Data file.

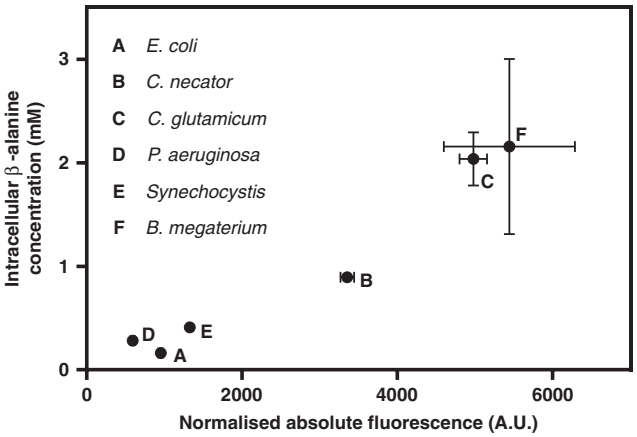


Fig. 8 Correlation between biosensor output and intracellular β-alanine concentration. Absolute normalised fluorescence values of *C. necator* cells, carrying plasmids containing the β-alanine biosensor and each a gene encoding a different homologue of L-aspartate 1-decarboxylase enzyme, correlated with their corresponding intracellular β-alanine concentration quantified by HPLC-UV. Cells were grown in MM. Samples were taken 6 h after supplementation with L-aspartate and L-arabinose. Data are mean ± SD, $n = 3$. Source data are provided as a Source Data file.

C. necator cells carrying these plasmids were grown in MM supplemented with both L-aspartate and L-arabinose to ensure an excess of substrate and to initiate the expression of the *panD* variants, respectively. The six selected homologues resulted in a ninefold range in fluorescence after 6 h (Fig. 8). To determine whether the biosensor output is indicative of product formation, metabolites were extracted from the cell pellets and the β-alanine concentrations were quantified using high-performance liquid chromatography coupled with ultraviolet spectroscopy (HPLC-UV). Intracellular β-alanine titres correlated well with fluorescence levels (Fig. 8), highlighting the utility of this biosensor to aid in the selection of enzyme variants. Of the six tested homologues, the *Bacillus megaterium* PanD demonstrated the highest fluorescence and product titre. Moreover, the obtained results are in accordance with previous findings in *E. coli*, where PanD from *Corynebacterium glutamicum* has been shown to be significantly more active than the native *E. coli* PanD[46].

Discussion

Thus far, inducible gene expression systems have been discovered primarily by research focussed on the experimental characterisation of individual metabolic pathways and their regulation. This traditional approach has allowed the identification of a substantial number of such systems, some of which have been developed into the widely used gene expression control devices, e.g., AraC/P_{araBAD} and LacI/P_{T7}. However, this empirical approach, principally driven by interest in the pathway characterisation, usually delivers only a limited amount of original information on a specific inducible system. Recently, high-throughput applications, such as transcriptomics analysis, comparative genomics and promoter or metagenome library screens, have proved to be efficient methodologies and substantially enhanced the speed of discovery of effector-responsive promoters or even corresponding TR-promoter pairs[7–13]. However, even these strategies suffer from several limitations, since they either do not ensure identification of all essential components of an inducible system or are TR- or ligand-type specific.

In this study, we developed a methodical approach that allows inducible systems to be mined at the genome scale level enabling the extraction of information on all three components—the regulator, the inducible promoter and its corresponding effector. To

demonstrate the utility of our approach, the method was applied to the genome of the catabolically versatile *C. necator* H16. Sixteen putative inducible systems resulting from a pool that comprised over 400 TRs were identified in this single bacterium. With exception of the acetoin- and 3-HP-inducible systems[25,47], we identify and characterise 14 systems from *C. necator*. Two previously proposed regulators[48,49], responding to tartrate and sulfonatoacetate, were experimentally validated to be involved in transcription activation of their corresponding metabolic genes in this study. Furthermore, here, we report inducible systems (OapR/$P_{H16\_RS01330}$ and PhgR/$P_{H16\_RS05530}$) responding to β-alanine and phenylglyoxylate. Both compounds play an important role as building blocks in chemical synthesis or food biotechnology[31,50]. Among the characterised inducible systems, different types of TR were identified, including LysR, AsnC, MocR, IclR and MarR. Fifteen of these were activator or dual-function-type regulators, whereas the latter exhibited characteristics of a repressor (BadR/$P_{H16\_RS27200}$).

To further evaluate inducible systems, they were subjected to thorough quantitative characterisation assessing induction level, dynamics and homogeneity. Several inducible systems exceeded induction levels of the frequently utilised L-arabinose-inducible system[29], with the BenM/$P_{H16\_RS09790}$-inducible system achieving expression level of more than 11-fold higher than AraC/P_{araBAD} in response to benzoate in *C. necator*. Four inducible systems responding to salicylate, sulfonatoacetate, tartrate and phenylglyoxylate exhibited a dynamic range of over 100-fold revealing that these systems are highly suitable to tightly control different levels of gene expression[51]. Along with a very high dynamic range of 650.7-fold, the NahR/$P_{H16\_RS08125}$-inducible system responds to nM concentrations of salicylate. This degree of sensitivity is equivalent to the most sensitive of characterised inducible systems, those based on anhydrotetracycline and cumate[52,53].

Orthogonal compatibility of inducible systems is a very important characteristic for designing multi-component and scalable circuits as well as sensory devices. These ideally require that functional genetic elements cross-react neither with the host genetic background nor between different heterologous systems. Twelve of the discovered inducible systems showed distinctive response only to their primary ligands. The capacity to independently drive the expression of multiple genes was exemplified by combining the salicylate- and 3,4-dihydroxybenzoate-inducible systems exquisitely demonstrating the potential of the characterised orthogonal switches for circuit design and other synthetic biology applications.

The utility of inducible systems was further demonstrated by applying individual promoter elements to control high levels of gene expression in *C. necator* (Fig. 2d) and by employing TR-promoter pairs in the model bacteria *E. coli* and *P. putida* (Fig. 3). Significantly, the adaptability of heterologous inducible system for gene expression control in other hosts was exemplified by engineering the β-alanine-inducible system for application in *E. coli* and *P. putida* (Fig. 4). In addition, the utility of the β-alanine biosensor was demonstrated by applying it to screen variants of L-aspartate 1-decarboxylase from different species and by identifying the *B. megaterium* homologue as most prominent to convert L-aspartate into β-alanine.

To conclude, the genome scale approach and inducible system evaluation pipeline presented in this paper, aids the discovery of metabolite-controlled systems. Further, it delivers quantitative data on inducible systems dynamics and orthogonality expanding the potential of developing tuneable regulatory circuits and biosensors. Furthermore, this generic approach can be utilised for mining inducible systems in any bacterial species, facilitating the expansion of the toolbox for synthetic biology and biotechnology applications.

Methods

Chemicals. All chemicals employed as ligands in this study are listed in Supplementary Table 6.

Base strains and media. All strains used in this study are listed in Supplementary Table 7. *E. coli* TOP10 (Life Technologies) was used for cloning and plasmid propagation. For single time-point fluorescence measurements and flow cytometric analyses, bacterial strains were propagated in Luria-Bertani (LB) medium[54]. To determine the dose–responses and to evaluate inducer TR cross-reactivity, *C. necator* was cultivated in MM containing 1 g/L NH_4Cl, 9 g/L $Na_2HPO_4{\cdot}12H_2O$, 1.5 g/L KH_2PO_4, 0.2 g/L $MgSO_4{\cdot}7H_2O$, 0.02 g/L $CaCl_2{\cdot}2H_2O$ and 0.0012 g/L $(NH_4)5[Fe(C_6H_4O_7)2]$ with 1 mL/L trace element solution SL7 (1.3 mL/L 25% (w/v) HCl, 0.07 g/L $ZnCl_2$, 0.1 g/L $MnCl_2{\cdot}4H_2O$, 0.062 g/L H_3BO_3, 0.190 g/L $CoCl_2{\cdot}6H_2O$, 0.017 g/L $CuCl_2{\cdot}2H_2O$, 0.024 g/L $NiCl_2{\cdot}6H_2O$ and 0.036 g/L $Na_2MoO_4{\cdot}2H_2O$) supplemented with 0.4% (w/v) sodium gluconate. If required, antibiotics were added to the growth medium at the following concentrations: 12.5 µg/mL tetracycline or 25 µg/mL chloramphenicol for *E. coli*, 50 µg/mL chloramphenicol for *C. necator* and 25 µg/mL tetracycline for *P. putida*. For solid media preparation, 15 g/L agar was added.

Cloning and transformation. Plasmid DNA was purified by using the QIAprep Spin Miniprep Kit (Qiagen). Microbial genomic DNA was extracted by employing the GenElute Bacterial Genomic DNA Kit (Sigma). DNA was amplified by PCR in 50 µL reactions using the Phusion High-Fidelity DNA polymerase from New England BioLabs (NEB). The Zymoclean Gel DNA Recovery Kit was employed to extract gel-purified linearised DNA. The NEBuilder Hifi DNA Assembly Master Mix, restriction enzymes and T4 DNA Ligase were purchased from NEB. All PCR-, digestion- and ligation reactions were set up according to the manufacturer's instructions.

For *E. coli* transformations, 50 µL of chemically competent *E. coli* TOP10 were mixed with 50 ng plasmid DNA, incubated on ice for 30 min, followed by a heat shock at 42 °C for 2 min and a subsequent incubation on ice for 2 min[54]. Cells were recovered in 450 µL of Super Optimal broth with Catabolite repression (SOC) medium (Invitrogen) at 37 °C for 1 h, plated on LB agar containing the appropriate antibiotic and incubated over night at 37 °C.

For *C. necator* transformations, 100 ng plasmid DNA were added to 100 µL of electrocompetent *C. necator* H16 in a pre-chilled electroporation cuvette (0.2 cm gap width, Bio-Rad) and incubated on ice for 5 min[55]. Electroporation was performed using a Bio-Rad Micropulser at 2.5 kV. Cells were recovered in 1 mL of SOC medium at 30 °C for 2 h, plated on LB agar containing the appropriate antibiotic and incubated at 30 °C for 2 days.

Electrocompetent *P. putida* KT2440 were freshly prepared from overnight cultures grown in LB medium at 30 °C and 200 rpm. A volume of 1 mL of cells were harvested by centrifugation at 16,000 × *g* for 5 min and washed with 1 mL of ice-cold 10% (v/v) glycerol. This step was repeated twice. Electroporation was conducted in an electroporation cuvette (0.2 cm gap width, Bio-Rad) using a Bio-Rad Micropulser at 12.5 kV[54]. Transformants were recovered in 1 mL of LB medium at 30 °C for 2 h, plated on LB agar containing the appropriate antibiotic and incubated over night at 30 °C.

Plasmid construction. Oligonucleotide primers were synthesised by Sigma-Aldrich (Supplementary Table 8). The *oapR* coding sequence was optimised for *E. coli* codon usage and synthesised by Life Technologies. The sequence can be found in the Supplementary Methods. Plasmids constructed by employing either the NEBuilder Hifi DNA assembly method or restriction enzyme-based cloning techniques[54] were validated by Sanger sequencing (Source BioScience, Nottingham, UK). A detailed assembly description for each plasmid is provided in the Supplementary Methods. Two versions of each plasmid containing an inducible system or an inducible promoter were constructed. One of the two versions contains a chloramphenicol resistance gene, the other one confers resistance to tetracycline. The former version was employed for evaluation of reporter gene expression in *C. necator* and *E. coli*, whereas the latter version was used in *P. putida*. To quantitatively evaluate the activity of various aspartate 1-decarboxylase (PanD) variants to convert L-aspartate into β-alanine, vectors were constructed that contain the original β-alanine biosensor and each one homologue of *panD* under control of the L-arabinose-inducible system. The *panD* genes were amplified from genomic DNA of the following species: *E. coli* MG1655, *C. necator* H16, *C. glutamicum* ATCC13032, *Pseudomonas aeruginosa* PAO1, *Synechocystis* sp. PCC6803 and *B. megaterium* DSM319. Key features of all plasmids used and generated in this study are summarised in Supplementary Table 9. The nucleotide sequences of the plasmids containing the 15 functional inducible systems pEH147, pEH134, pEH154, pEH155, pEH042, pEH148, pEH083, pEH157, pEH136, pEH137, pEH256, pEH158, pEH159, pEH161 and pEH052 have been deposited in the public version of the JBEI registry (https://public-registry.jbei.org) under the accession numbers JPUB_014465-JPUB_014479, respectively. The nucleotide sequence of plasmid pEH010 has been deposited under accession number JPUB_008754[47].

Fluorescence measurements. For quantification of RFP fluorescence at a single time point, individual colonies of freshly transformed bacterial cells were used to

inoculate 5 mL of LB containing the appropriate antibiotic in 50-mL conical centrifuge tubes. After incubation over night with orbital shaking at 200 rpm and 30 °C, *E. coli* and *P. putida* were diluted 1:50, and *C. necator* was diluted 1:20 into 5 mL of fresh LB medium containing the respective antibiotic. The exponentially growing cells with an OD_{600} of 0.05–0.1 were supplemented with inducer to achieve a final concentration of 5 mM 4 h after the main culture had been set up or left uninduced. After a further incubation with orbital shaking at 30 °C and 200 rpm for 6 h, uninduced and induced cells with an OD_{600} of 1.4-2.0 were pelleted by centrifugation at 16,000 × *g* for 4 min and resuspended in an equal volume of phosphate buffered saline (PBS). Subsequently, 100 µL of cells were transferred to a 96-well microtiter plate (flat and clear bottom, black; Greiner One International) and RFP fluorescence was quantified using an Infinite M1000 PRO microplate reader (Tecan). Fluorescence excitation and emission wavelengths were set to 585 nm and 620 nm, respectively. The gain factor was set manually to 100%. Absorbance was measured at 600 nm to normalise fluorescence by optical density. Prior normalisation, fluorescence and absorbance values were corrected by subtracting the auto fluorescence and absorbance of the culture medium.

To determine the dose–response of each individual inducible system, RFP fluorescence and absorbance were quantified over time. The precultures of *C. necator* cells were prepared as for the single time-point measurements in 2 mL of MM containing the appropriate antibiotic. After incubation over night, cells were diluted 1:20 into 5 mL of fresh MM containing the respective antibiotic and grown with orbital shaking at 30 °C and 200 rpm. At an OD_{600} of 0.2, 142.5 µL of the exponentially growing cells were transferred to a 96-well microtiter plate. Each well was supplemented with 7.5 µL of stock inducer at the desired concentration. Fluorescence and absorbance were measured every 5 min over the time course of 6 h using the same excitation and emission wavelengths as for the single time-point measurements with the gain factor set to 80%. The fluorescence and absorbance values, recorded 80 min after the inducer had been added, were obtained from the time course data and used to calculate the absolute normalised fluorescence values corresponding to each inducer concentration.

Production of β-alanine and metabolite extraction. Real-time biosynthesis of β-alanine was monitored quantitatively by HPLC and by fluorescence output of *C. necator* harbouring both the β-alanine biosensor and different homologues of the gene encoding aspartate 1-decarboxylase (PanD). Single colonies of freshly transformed cells were used to inoculate 2 mL of MM in 50-mL conical centrifuge tubes. The preculture was incubated for 18 h with orbital shaking at 30 °C and 200 rpm. Subsequently, it was diluted 1:20 in 6 mL of fresh MM and incubated under the same conditions. At an OD_{600} of 0.15–0.2, L-arabinose and L-aspartate were added to achieve the final concentrations of 1 mM and 20 mM, respectively. Samples of 0.6 mL were taken immediately, 2, 4 and 6 h after compound supplementation and used for absorbance and fluorescence measurements with the gain factor set to 80%. To quantify metabolites in the supernatant, cells were pelleted by centrifugation at 16,000 *g* for 5 min. A volume of 50 µL of the cell-free supernatant was kept at −80 °C until subjected to HPLC analysis.

To determine the intracellular concentration of β-alanine 6 h after supplementation with L-arabinose and L-aspartate, 1.5 mL of the remaining culture was pelleted by centrifugation at 16,000 *g* for 5 min in a preweighed 1.5-mL microcentrifuge tube. The supernatant was discarded and the same step repeated with another 1.5 mL of the culture. The cell pellet was resuspended in 1 mL of PBS and centrifuged as before. Subsequently, the supernatant was completely removed, the weight of the wet cell pellet determined using a fine balance and frozen over night at −80 °C. To extract metabolites, 50 µL of −40 °C methanol–water solution (60% v/v) was added to the pellet. The sample was mixed vigorously using vortex until completely resuspended. Cells were frozen at −80 °C for 30 min, thawed on ice and vortexed vigorously for 1 min. The freeze-thaw cycle was repeated three times. Subsequently, cells were pelleted by centrifugation at 26,000 *g* for 20 min at −10 °C. The supernatant was removed and kept at −80 °C. Cells were resuspended in another 50 µL of −40 °C methanol–water solution (60% v/v), subjected to three freeze-thaw cycles and centrifuged as before. The supernatant was pooled with the first collection and stored at −80 °C until used for HPLC analysis.

Quantification of amino acids. β-Alanine was quantified using a Dionex UltiMate 3000 HPLC system (Thermo Scientific) equipped with a Kinetex 5 µm EVO C18 100 Å LC 150 mm × 4.6 mm column (Phenomenex) and a photo diode array (UV-VIS) detector measuring the absorption at 338 nm and 210 nm. The sample was prepared by adding 950 µL of 50% methanol diluent (HPLC-grade) to 50 µL of either the cell-free supernatant or the extract containing the intracellular metabolites. After the sample was mixed by vortexing, it was filtered into a HPLC vial using a 0.2 µm syringe filter. Derivatisation of amino acids with fluoraldehyde *o*-phthalaldehyde (OPA) was performed automatically by the autosampler before injection. The autosampler programming instructions can be found in Supplementary Table 10. Preparation of OPA reagent was adapted from Roth[56] with slight modifications: the reagent is composed of OPA (0.8 g/L) diluted in HPLC-grade methanol (10 mL/L), added to a 0.1 M KH_2PO_4 buffer at pH 10.4 supplemented with β-mercaptoethanol (2 mL/L). The HPLC method was adapted from Phenomenex HPLC application ID 23092 (https://phenomenex.com/Application/Detail/23092). Briefly, two mobile phases were used: mobile phase A was 20 mM KH_2PO_4 adjusted to pH 7.2 with KOH, while mobile phase B was methanol/acetonitrile (50/50 v/v, HPLC-grade). All samples and reagents were kept at 4 °C throughout the analysis. The column was operated at 30 °C. The injection volume was 5 µL and samples were run for 23 min. The separation was achieved with a flow rate of 1 mL/min and followed the gradient programme in Supplementary Table 11. Data analysis was performed using Chromeleon 7 (Thermo Scientific). Metabolite concentrations were quantified using calibration curves generated from running standards of known concentrations which were prepared the same as the samples.

Calculation of intracellular β-alanine concentration. The volume of the cell pellet (V_{pellet}) was calculated by dividing the weight of the wet cell pellet by the cell density of 1.105 g/mL[57]. The cell density of *C. necator* was denoted to be the same as of *E. coli*, assuming that under the conditions tested no polyhydroxybutyrate (PHB) had been accumulated. Together with the volume of extraction solvent added to the pellet ($V_{solvent}$), V_{pellet} was used to calculate the intracellular molar concentration of β-alanine using formula (1):

$$C_{intracellular} = \left(\frac{V_{pellet} + V_{solvent}}{V_{pellet}}\right) \times C_{extract} \quad (1)$$

The remaining parameters correspond to the intracellular molar concentration of β-alanine ($C_{intracellular}$) and the molar concentration of β-alanine in the extract ($C_{extract}$) determined by HPLC-UV analysis.

Mathematical modelling. To obtain system parameters that can be used for synthetic circuit design, absolute normalised fluorescence values (RFP) were plotted as a function of inducer concentration using software GraphPad Prism 7. Subsequently, a non-linear least-squares fit was performed using the Hill function (2):

$$\mathrm{RFP}(I) = b_{max} \times \frac{I^h}{K_m^h + I^h} + b_{min} \quad (2)$$

The parameters correspond to the maximum level of reporter output (b_{max}), the concentration of inducer (I), the Hill coefficient (h), the inducer concentration that mediates half-maximal reporter output (K_m) and the basal level of fluorescence output (b_{min}). Relative normalised fluorescence values as shown in Fig. 5a were obtained by dividing absolute normalised fluorescence values at a specific inducer concentration after subtraction of the absolute normalised fluorescence of the uninduced cells by the corresponding maximum level of fluorescence output b_{max}.

The dynamic range μ was calculated with formula (3):

$$\mu = \frac{b_{max}}{b_{min}} \quad (3)$$

The corresponding standard deviation σ_μ was calculated using Eq. (4):

$$\sigma_\mu = \mu \times \sqrt{\left(\frac{\sigma_{b_{max}}}{b_{max}}\right)^2 + \left(\frac{\sigma_{b_{min}}}{b_{min}}\right)^2} \quad (4)$$

The standard deviation of the maximum level of reporter output was obtained from Prism, whereas the standard deviation of the basal level of fluorescence output was calculated from the absolute normalised fluorescence values of the uninduced cells.

Flow cytometry. For evaluation of induction homogeneity and cross-reactivity by flow cytometry, single colonies of freshly transformed *C. necator* cells were used to inoculate 5 mL of LB containing 50 µg/mL chloramphenicol in 50-mL conical centrifuge tubes. After incubation over night with orbital shaking at 30 °C and 200 rpm, cells were diluted 1:50 in 5 mL of fresh LB medium containing the antibiotic. Inducers were added to the logarithmically growing cells 4 h after further incubation. Samples were taken 2 h after supplementation with inducer. Cells were pelleted by centrifugation at 4,000 *g* for 5 min and resuspended in cold and sterile filtered PBS to an OD_{600} of 0.01. The cells were kept on ice until analysed using an Astrios EQ flow cytometer (Beckman Coulter). RFP fluorescence was measured with a 561 nm laser and a 614/20 nm emission band-pass filter. GFP fluorescence was quantified with a 488 nm laser and a 516/28 nm emission band-pass filter. The voltage of the photomultiplier tube was set to 400 V and the area and height gain was adjusted to 1.0. At least 100,000 events were collected for each sample. Data analysis was performed using software Kaluza 1.5 (Beckman Coulter).

Cross-reactivity screen. The activity of effectors against non-cognate promoters was evaluated using an integrated robotic platform (Beckman Coulter). The *C. necator* preculture was set up by inoculating 2 mL of chloramphenicol-containing MM with a single colony of freshly transformed bacterial cells. After incubation for 18 h with orbital shaking at 30 °C and 200 rpm, the bacterial culture was diluted 1:50 in 50 mL of fresh MM containing the antibiotic in 250-mL baffled shake flasks. The cells were grown for another 4 h under the same conditions until an OD_{600} of 0.15–0.2 was reached. After pouring the bacterial culture into a 250-mL reservoir (Thermo Fisher), 142.5 µL were dispensed into a black 96-well microtiter plate (the same that was used for the fluorescence measurements) using a liquid handling robotic platform (Biomek FXp, Beckman Coulter). The workflow generated using

software SAMI EX (Beckman Coulter) is illustrated in Supplementary Fig. 10. Inducers were dissolved to a final concentration of 100 mM (except in case of L-2-hydroxyglutarate which was dissolved to a final concentration of 10 mM) and transferred to a 96-deep-well plate (2.0 mL square wells with round bottoms, STARLAB International GmbH). Using a Biomek FXp, 7.5 µL of stock inducer were added to the *C. necator* culture in the microtiter plate. Fluorescence and absorbance measurements were taken immediately, 6, 12 and 18 h after supplementation with inducer by an integrated SpectraMax 3i plate reader (Molecular Devices). The same plate reader settings were used as for the measurements taken using the Infinite M1000 PRO microplate reader. In between the measurements, the plates were kept in an integrated Cytomat2 shaking incubator (Thermo Fisher Scientific) at 30 °C and 600 rpm. The workflow generated using software SAMI EX (Beckman Coulter) is illustrated in Supplementary Fig. 11.

The relative induction (in %) is calculated using Eq. (5):

$$\textit{Relative induction}\ (\%) = 100\times\left(\frac{FL_{\text{compound}} - FL_{\text{uninduced}}}{FL_{\text{primary inducer}} - FL_{\text{uninduced}}}\right) \quad (5)$$

FL corresponds to the OD-normalised absolute fluorescence values.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

All data generated or analysed during this study are included in this published article (and its Supplementary Information files). The nucleotide sequences of the plasmids containing the 16 functional inducible systems have been deposited in the public version of the JBEI registry (https://public-registry.jbei.org) under the accession numbers JPUB_014465-JPUB_014479 and JPUB_008754. The nucleotide sequences of the plasmids containing the 15 functional inducible systems pEH042, pEH052, pEH083, pEH134, pEH136, pEH137, pEH147, pEH148, pEH154, pEH155, pEH157, pEH158, pEH159, pEH161 and pEH256 have been deposited NCBI GenBank under the accession numbers MT024789, MT024790, MT024791, MT024792, MT024793, MT024794, MT024795, MT024796, MT024797, MT024798, MT024799, MT024800, MT024801, MT024802 and MT024803, respectively. The source data underlying Fig. 2c, d, 3a–d, 4b, c, 5a, 6a, 7, 8, Table 1, Supplementary Figs. 3b, 4, 5, and Supplementary Tables 4 and 5 are provided as a Source Data file or available from the corresponding authors upon reasonable request.

Received: 19 September 2019; Accepted: 7 February 2020;
Published online: 05 March 2020

References

1. Khalil, A. S. & Collins, J. J. Synthetic biology: applications come of age. *Nat. Rev. Genet.* **11**, 367–379 (2010).
2. Brophy, J. A. & Voigt, C. A. Principles of genetic circuit design. *Nat. Methods* **11**, 508–520 (2014).
3. Rogers, J. K., Taylor, N. D. & Church, G. M. Biosensor-based engineering of biosynthetic pathways. *Curr. Opin. Biotechnol.* **42**, 84–91 (2016).
4. D'Ambrosio, V. & Jensen, M. K. Lighting up yeast cell factories by transcription factor-based biosensors. *FEMS Yeast Res.* **17**, fox076 (2017).
5. Chen, Y. et al. Tuning the dynamic range of bacterial promoters regulated by ligand-inducible transcription factors. *Nat. Commun.* **9**, 64 (2018).
6. Wan, X., Marsafari, M. & Xu, P. Engineering metabolite-responsive transcriptional factors to sense small molecules in eukaryotes: current state and perspectives. *Microb. Cell Fact.* **18**, 61 (2019).
7. Chang, H. J., Voyvodic, P. L., Zúñiga, A. & Bonnet, J. Microbially derived biosensors for diagnosis, monitoring and epidemiology. *Microb. Biotechnol.* **10**, 1031–1035 (2017).
8. Van Der Meer, J. R. & Belkin, S. Where microbiology meets microengineering: design and applications of reporter bacteria. *Nat. Rev. Microbiol.* **8**, 511–522 (2010).
9. Wen, K. Y. et al. A cell-free biosensor for detecting quorum sensing molecules in *P. aeruginosa*-infected respiratory samples. *ACS Synth. Biol.* **6**, 2293–2301 (2017).
10. Mahr, R. & Frunzke, J. Transcription factor-based biosensors in biotechnology: current state and future prospects. *Appl. Microbiol. Biotechnol.* **100**, 79–90 (2016).
11. Greenwald, E. C., Mehta, S. & Zhang, J. Genetically encoded fluorescent biosensors illuminate the spatiotemporal regulation of signaling networks. *Chem. Rev.* **118**, 11707–11794 (2018).
12. van Hijum, S. A., Medema, M. H. & Kuipers, O. P. Mechanisms and evolution of control logic in prokaryotic transcriptional regulation. *Microbiol. Mol. Biol. Rev.* **73**, 481–509 (2009).
13. Dahl, R. H. et al. Engineering dynamic pathway regulation using stress-response promoters. *Nat. Biotechnol.* **31**, 1039–1046 (2013).
14. Mahr, R., von Boeselager, R. F., Wiechert, J. & Frunzke, J. Screening of an *Escherichia coli* promoter library for a phenylalanine biosensor. *Appl. Microbiol. Biotechnol.* **100**, 6739–6753 (2016).
15. Uchiyama, T. & Miyazaki, K. Product-induced gene expression, a product-responsive reporter assay used to screen metagenomic libraries for enzyme-encoding genes. *Appl. Environ. Microbiol.* **76**, 7029–7035 (2010).
16. Ahn, S. K., Cuthbertson, L. & Nodwell, J. R. Genome context as a predictive tool for identifying regulatory targets of the TetR family transcriptional regulators. *PloS ONE* **7**, e50562 (2012).
17. Ravcheev, D. A. et al. Comparative genomics and evolution of regulons of the LacI-family transcription factors. *Front. Microbiol.* **5**, 294 (2014).
18. Kazanov, M. D., Li, X., Gelfand, M. S., Osterman, A. L. & Rodionov, D. A. Functional diversification of ROK-family transcriptional regulators of sugar catabolism in the Thermotogae phylum. *Nucleic Acids Res.* **41**, 790–803 (2012).
19. Maddocks, S. E. & Oyston, P. C. Structure and function of the LysR-type transcriptional regulator (LTTR) family proteins. *Microbiology* **154**, 3609–3623 (2008).
20. Oliver, P., Peralta-Gil, M., Tabche, M.-L. & Merino, E. Molecular and structural considerations of TF-DNA binding for the generation of biologically meaningful and accurate phylogenetic footprinting analysis: the LysR-type transcriptional regulator family as a study model. *BMC Genomics* **17**, 686 (2016).
21. Benson, D. A. et al. GenBank. *Nucleic Acids Res.* **42**, D32–D37 (2013).
22. Jeske, L., Placzek, S., Schomburg, I., Chang, A. & Schomburg, D. BRENDA in 2019: a European ELIXIR core data resource. *Nucleic Acids Res.* **47**, D542–D549 (2018).
23. Campbell, R. E. et al. A monomeric red fluorescent protein. *Proc. Natl Acad. Sci. USA* **99**, 7877–7882 (2002).
24. Delamarre, S. C. & Batt, C. A. Comparative study of promoters for the production of polyhydroxyalkanoates in recombinant strains of *Wautersia eutropha*. *Appl. Microbiol. Biotechnol.* **71**, 668–679 (2006).
25. Fukui, T., Ohsawa, K., Mifune, J., Orita, I. & Nakamura, S. Evaluation of promoters for gene expression in polyhydroxyalkanoate-producing *Cupriavidus necator* H16. *Appl. Microbiol. Biotechnol.* **89**, 1527–1536 (2011).
26. Knoten, C. A., Hudson, L. L., Coleman, J. P., Farrow, J. M. & Pesci, E. C. KynR, a Lrp/AsnC-type transcriptional regulator, directly controls the kynurenine pathway in *Pseudomonas aeruginosa*. *J. Bacteriol.* **193**, 6567–6575 (2011).
27. Hanahan, D. Studies on transformation of *Escherichia coli* with plasmids. *J. Mol. Biol.* **166**, 557–580 (1983).
28. Hirakawa, H., Hirakawa, Y., Greenberg, E. P. & Harwood, C. S. BadR and BadM proteins transcriptionally regulate two operons needed for anaerobic benzoate degradation by *Rhodopseudomonas palustris*. *Appl. Environ. Microbiol.* **81**, 4253–4262 (2015).
29. Jha, R. K., Chakraborti, S., Kern, T. L., Fox, D. T. & Strauss, C. E. Rosetta comparative modeling for library design: engineering alternative inducer specificity in a transcription factor. *Proteins* **83**, 1327–1340 (2015).
30. Könst, P. M., Franssen, M. C., Scott, E. L. & Sanders, J. P. A study on the applicability of L-aspartate α-decarboxylase in the biobased production of nitrogen containing chemicals. *Green Chem.* **11**, 1646–1652 (2009).
31. Song, C. W., Lee, J., Ko, Y.-S. & Lee, S. Y. Metabolic engineering of *Escherichia coli* for the production of 3-aminopropionic acid. *Metab. Eng.* **30**, 121–129 (2015).
32. Rokicki, J. et al. Daily carnosine and anserine supplementation alters verbal episodic memory and resting state network connectivity in healthy elderly adults. *Front. Aging Neurosci.* **7**, 219 (2015).
33. Szcześniak, D., Budzeń, S., Kopeć, W. & Rymaszewska, J. Anserine and carnosine supplementation in the elderly: effects on cognitive functioning and physical capacity. *Arch. Gerontol. Geriatr.* **59**, 485–490 (2014).
34. Schneider, F., Krämer, R. & Burkovski, A. Identification and characterization of the main β-alanine uptake system in *Escherichia coli*. *Appl. Microbiol. Biotechnol.* **65**, 576–582 (2004).
35. Alagesan, S. et al. Functional genetic elements for controlling gene expression in *Cupriavidus necator* H16. *Appl. Environ. Microbiol.* **84**, e00878–00818 (2018).
36. Belitsky, B. R. *Bacillus subtilis* GabR, a protein with DNA-binding and aminotransferase domains, is a PLP-dependent transcriptional regulator. *J. Mol. Biol.* **340**, 655–664 (2004).
37. Morales, G. et al. The *Pseudomonas putida* Crc global regulator controls the expression of genes from several chromosomal catabolic pathways for aromatic compounds. *J. Bacteriol.* **186**, 1337–1344 (2004).
38. Cebolla, A., Sousa, C. & de Lorenzo, V. Effector specificity mutants of the transcriptional activator NahR of naphthalene degrading *Pseudomonas* define protein sites involved in binding of aromatic inducers. *J. Biol. Chem.* **272**, 3986–3992 (1997).

39. Khlebnikov, A., Risa, Ø., Skaug, T., Carrier, T. A. & Keasling, J. Regulatable arabinose-inducible gene expression system with consistent control in all cells of a culture. *J. Bacteriol.* **182**, 7029–7034 (2000).
40. Hanko, E. K. R., Minton, N. P. & Malys, N. A transcription factor-based biosensor for detection of itaconic acid. *ACS Synth. Biol.* **7**, 1436–1446 (2018).
41. Walker, L. M., Li, B., Niks, D., Hille, R. & Elliott, S. J. Deconvolution of reduction potentials of formate dehydrogenase from *Cupriavidus necator*. *J. Biol. Inorg. Chem.* **24**, 889–898 (2019).
42. Volodina, E., Schürmann, M., Lindenkamp, N. & Steinbüchel, A. Characterization of propionate CoA-transferase from *Ralstonia eutropha* H16. *Appl. Microbiol. Biotechnol.* **98**, 3579–3589 (2014).
43. Peplinski, K., Ehrenreich, A., Döring, C., Bömeke, M. & Steinbüchel, A. Investigations on the microbial catabolism of the organic sulfur compounds TDP and DTDP in *Ralstonia eutropha* H16 employing DNA microarrays. *Appl. Microbiol. Biotechnol.* **88**, 1145–1159 (2010).
44. Sivapragasam, S. & Grove, A. *Streptomyces coelicolor* XdhR is a direct target of (p) ppGpp that controls expression of genes encoding xanthine dehydrogenase to promote purine salvage. *Mol. Microbiol.* **100**, 701–718 (2016).
45. Hillerich, B. & Westpheling, J. A new TetR family transcriptional regulator required for morphogenesis in *Streptomyces coelicolor*. *J. Bacteriol.* **190**, 61–67 (2008).
46. Li, H. et al. β-Alanine production using whole-cell biocatalysts in recombinant *Escherichia coli*. *Mol. Catal.* **449**, 93–98 (2018).
47. Hanko, E. K. R., Minton, N. P. & Malys, N. Characterisation of a 3-hydroxypropionic acid-inducible system from *Pseudomonas putida* for orthogonal gene expression control in *Escherichia coli* and *Cupriavidus necator*. *Sci. Rep.* **7**, 1724 (2017).
48. Crouzet, P. & Otten, L. Sequence and mutational analysis of a tartrate utilization operon from *Agrobacterium vitis*. *J. Bacteriol.* **177**, 6518–6526 (1995).
49. Weinitschke, S. et al. Sulfoacetate is degraded via a novel pathway involving sulfoacetyl-CoA and sulfoacetaldehyde in *Cupriavidus necator* H16. *J. Biol. Chem.* **285**, 35249–35254 (2010).
50. Tang, C.-D. et al. Biosynthesis of phenylglyoxylic acid by LhDMDH, a novel d-mandelate dehydrogenase with high catalytic activity. *J. Agric. Food Chem.* **66**, 2805–2811 (2018).
51. Mannan, A. A., Liu, D., Zhang, F. & Oyarzún, D. A. Fundamental design principles for transcription-factor-based metabolite biosensors. *ACS Synth. Biol.* **6**, 1851–1859 (2017).
52. Lutz, R. & Bujard, H. Independent and tight regulation of transcriptional units in *Escherichia coli* via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements. *Nucleic Acids Res.* **25**, 1203–1210 (1997).
53. Kaczmarczyk, A., Vorholt, J. A. & Francez-Charlot, A. Cumate-inducible gene expression system for Sphingomonads and other *Alphaproteobacteria*. *Appl. Environ. Microbiol.* **79**, 6795–6802 (2013).
54. Sambrook, J. & Russell, D. W. *Molecular Cloning: A Laboratory Manual* 3rd. edn, (Cold Spring Harbor Laboratory Press, New York, 2001).
55. Ausubel, F. M. *Current Protocols in Molecular Biology* (Wiley, 2003).
56. Roth, M. Fluorescence reaction for amino acids. *Anal. Chem.* **43**, 880–882 (1971).
57. Martinez-Salas, E., Martin, J. & Vicente, M. Relationship of *Escherichia coli* density to growth rate and cell age. *J. Bacteriol.* **147**, 97–100 (1981).

Acknowledgements

This work was supported by the Biotechnology and Biological Sciences Research Council (BBSRC; grant number BB/L013940/1) and the Engineering and Physical Sciences Research Council (EPSRC) under the same grant number. We thank The University of Nottingham for providing SBRC-DTProg Ph.D. studentship to E.K.R.H. and Ph.D. studentship to A.C.P., Stephan Heeb for gifting strain *P. putida* KT2440, Toby Pendlebury and Simona Lukosiute for assistance with plasmid assembly, David Onion for assistance with flow cytometry, R. Biedendieck, S. Grigoriou and S. Craig for gifting microbial genomic DNA, and all members of the SBRC who helped to carry out this research.

Author contributions

E.K.R.H. and N.M. conceived the project, designed and performed the experiments, analysed the results and wrote the manuscript. A.C.P. conducted fluorescence assays in *P. putida*. M.J. assisted with the robotic platform for cross-reactivity screen. M.A. designed the HPLC method and assisted with HPLC analysis. N.P.M. analysed the results and wrote the manuscript. All authors read and approved the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41467-020-14941-6.

Correspondence and requests for materials should be addressed to N.M.

Peer review information *Nature Communications* thanks the anonymous reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.