

Thèse de doctorat
de l'Université Sorbonne Paris Cité
Préparée à l'Université Paris Diderot

École doctorale Frontières du Vivant (ED n° 474)

Laboratoires d'accueil :

Institute of Systems and Synthetic Biology
(Genopole, CNRS, UEVE, Université
Paris-Saclay, Evry, France)

Micalis Institute
(INRA, AgroParisTech, Université Paris-
Saclay, Jouy-en-Josas, France)

New inputs for synthetic biological systems

Nouvelles stratégies d'induction pour systèmes biologiques synthétiques

Par Vincent LIBIS

Thèse de doctorat de Biologie de synthèse
Sous la direction de Jean-Loup FAULON

Présentée et soutenue publiquement à Paris le 24 Novembre 2016

Rapporteur	Marcel SALANOUBAT	CNRS
Rapporteuse	Julia FRUNZKE	Forschungszentrum Jülich
Examineur	David BIKARD	Institut Pasteur
Directeur de thèse	Jean-Loup FAULON	INRA

Abstract

Synthetic biologists program DNA with the aim of building biological systems that react under certain conditions in a predefined way. This ability could have impact in several fields, from medicine to industrial fermentation. While the scalability of synthetic biological circuits in terms of signal processing is now almost demonstrated, the variety of input signals for these circuits is limited. Because each application typically requires a circuit to react to case-specific molecules, the lack of input diversity is a major obstacle to the development of new applications. Two axes are developed over the course of this thesis to try to address input-related problems.

The main axis consists in a new strategy aiming at systematically and immediately increasing the chemical diversity of inputs for synthetic circuits. Current approaches to expand the number of potential inputs focus on re-engineering sensing systems such as riboswitches or allosteric transcription factors to make them react to previously non-detectable molecules. On the contrary, here we developed a method to transform the non-detectable molecules themselves into molecules for which sensing systems already exist. These chemical transformations are realized *in situ* by expressing synthetic metabolic pathways in the cell. In order to systematize this strategy, we leveraged computer-aided design to predict ways of detecting new molecules by digging into all known biochemical reactions. We then implemented several predictions *in vivo* that successfully enabled *E. coli* to detect new chemicals. Aside from the interest of the method for biotechnological applications, this shows that in addition to transferring matter and energy, metabolism can also play a role in transferring information, raising the question of potential occurrences of this sensing strategy in nature.

A second axis introduces a way to exempt simple programs from the need for a chemical input, and explore the use of a biological input instead. In situations where a single timely induction or repression of multiple genes is required, such as in industrial fermentation processes, we propose to replace expensive chemical induction by simultaneous infection of all the members of a growing population of cells with viral particles inputting in real-time all the necessary information for the task at hand. In the context of fermentation, we developed engineered viral particles that can dynamically reprogram the metabolism of a large population of bacteria at the optimal stage of growth and force them to produce value-added chemicals.

Résumé

Les chercheurs en biologie de synthèse programment l'ADN pour construire des systèmes biologiques capables de répondre à certaines conditions de manière prédéfinie. Cette capacité pourrait avoir un impact sur plusieurs domaines, de la médecine à la fermentation industrielle. Le traitement de signal par des circuits biologiques synthétiques est en train d'être démontré à large échelle, mais hélas la variété des signaux d'entrée capables de contrôler ces circuits est pour l'instant limitée. Ce manque de diversité est un obstacle majeur au développement de nouvelles applications car en général chaque application requiert une réponse à des signaux de nature particulière qui lui sont spécifiques. Cette thèse cherche à apporter des solutions au manque de signaux d'entrée appropriés contrôlant les circuits biologiques en développant deux nouvelles stratégies d'induction.

La première stratégie vise à étendre la diversité chimique des signaux d'entrée. A l'inverse des approches existantes, qui reposent sur la modification des systèmes de détecteurs naturels tels que les riboswitchs ou les facteurs de transcription allostériques, j'ai cherché ici à modifier directement des molécules préalablement non-déTECTABLES afin de les rendre détectables par les systèmes de détection actuels. Pour ce faire, la transformation chimique des molécules cibles est réalisée *in situ* grâce à l'expression de voies métaboliques synthétiques dans la cellule. Afin de pouvoir utiliser cette stratégie de manière systématique, j'ai employé la conception assistée par ordinateur et puisé dans l'ensemble des réactions biochimiques connues afin de prédire des voies de détecteurs pour de nouvelles molécules. J'ai ensuite implémenté *in vivo* plusieurs prédictions qui ont permis à *E. coli* de détecter de nouveaux composés. Au-delà de l'intérêt de cette méthode en biotechnologie, cela montre que le métabolisme peut jouer un rôle dans le transfert d'information, en plus de son rôle dans le transfert de matière et d'énergie, ce qui soulève la question de l'utilisation potentielle de cette stratégie de détection par la nature.

Un second axe présente une façon d'épargner l'utilisation d'inducteurs chimiques pour les programmes biologiques simples, et propose d'utiliser des inducteurs biologiques à la place. Lorsqu'une seule étape d'induction ou de répression de gènes est nécessaire, comme c'est le cas en fermentation industrielle, je propose de remplacer la coûteuse étape d'induction chimique par l'infection simultanée de toutes les cellules d'une population par des particules virales capables d'injecter en temps réel l'ensemble des informations nécessaires pour déclencher l'activité biologique recherchée. A des fins de fermentation, j'ai développé des particules virales modifiées qui reprogramment dynamiquement le métabolisme d'une large population de bactérie au moment opportun et les forcent à produire des molécules à haute valeur ajoutée.

Remerciements/Acknowledgments

I am very grateful to Julia Frunzke, Marcel Salanoubat and David Bikard for taking the time to read my thesis and for accepting to take part in my defense committee.

Merci à Jean-Loup Faulon pour m'avoir accueilli dans son laboratoire. J'ai bénéficié du cadre interdisciplinaire que je cherchais et d'une grande liberté qui m'ont fait beaucoup apprécier ces trois années de thèse. Non seulement Jean-Loup m'a permis de consacrer du temps à ma guise à des projets scientifiques variés et risqués (trop, pour certains d'entre eux qui ne peuvent pas figurer dans ce manuscrit pour cause d'échecs cuisants), mais j'ai aussi pu découvrir la création de startup et j'ai même eu l'honneur de devenir son associé. Merci de m'avoir fait confiance pour tout cela, j'ai beaucoup appris.

I am deeply grateful to the CRI and its community. I cannot overstate the impact the AIV/FDV program and the people I met there had on me. Thank you Ariel for welcoming me to this family. The list of important names is very long so I will limit it to my friends who directly contributed to this thesis through their sound advices and support: Aude, Jean, Matt, Aleksandra, Antoine. Thank you for these countless, yet precious hours at "Aux Marsouins" and elsewhere.

Thank you Jake for inspiring me, being my friend, and mentoring me.

Je remercie les membres du laboratoire de Jean-Loup Faulon avec qui j'ai travaillé, en particulier les 4 stagiaires que j'ai eu le plaisir d'encadrer : Rémi, Cécile, Amir, Thibault.

Je remercie évidemment Baudoin pour notre collaboration efficace et pour ces maintes discussions philosophiques jusqu'à des heures tardives.

Je remercie Emmanuelle Guillot-Combe et la DGA pour avoir cru en moi et financé ce projet, ainsi que la fondation Bettencourt-Schueller pour le soutien financier, depuis iGEM jusqu'à FDV.

Table of contents

Chapter 1 : Introduction	8
1.1 From genetic programs to synthetic biology.....	9
1.2 The promise of scalability	11
1.3 Current inputs	14
1.4 Overview of the thesis.....	16

Part I : Expanding the scope of chemical inputs

Contributions of the author.....	20
----------------------------------	----

Chapter 2 : Expanding biosensing abilities through computer-aided design of metabolic pathways.....	22
2.1 Abstract.....	22
2.2 Introduction.....	23
2.3 Results	25
2.3.1 The scope of detectable molecules is significantly expanded in silico via enzymatic pathways.	25
2.3.2 Implementation of SEMP in vivo expands E. coli sensing abilities.....	30
2.4 Discussion	38
2.5 Methods.....	40

Chapter 3 : Opinion on sensing new chemicals with bacterial transcription factors	51
3.1 Abstract.....	51
3.2 Introduction.....	52
3.3 Mining allosteric transcription factors in nature.....	52
3.4 Engineering the sensing scope.....	54
3.5 Regulation via indirect sensing	57
3.6 Conclusion	61

Conclusion and perspectives – Part I:	64
---	----

Part II : Phages as biological inputs

Contributions of the author.....	67
----------------------------------	----

Chapter 4 : Phage-mediated delivery of targeted sRNA constructs to knock down gene expression in <i>E. coli</i>	68
4.1 Abstract.....	68
4.2 Introduction.....	69
4.3 Protocol	72

4.3.1 Design and construction of phagemid vectors bearing sRNA silencing cassettes	72
4.3.2 Production and harvest of M13-packaged phagemid stocks	78
4.3.3 Preparation of F ⁺ target cells for silencing	78
4.3.4 Infection with packaged phagemids for silencing	79
4.4 Representative results	80
4.4.1 Silencing of mKate2 fluorescence in liquid media.....	80
4.4.2 Silencing of chloramphenicol resistance on agar plates	82
4.5 Discussion	84
Chapter 5 : On using phagemids as real-time metabolic engineering tools	86
5.1 Abstract.....	86
5.2 Introduction.....	86
5.3 Results.....	89
5.3.1 Identification of a viable design framework for phage-based metabolic engineering.....	89
5.3.2 Fermentation of lycopene and deoxyviolacein can be achieved upon phage infection. .	90
5.3.3 Engineered phages allow silencing of gene targets across the chromosome, including targets in the central metabolism and essential genes.	91
5.3.4 Characterization of fermentation features.....	93
5.4 Discussion	95
5.5 methods	96
Conclusion and perspectives – Part II:.....	99
Bibliography.....	101
Chapter 6 : Annexes.....	117
6.1 Annex 1.....	118
Sensipath: computer-aided design of sensing-enabling metabolic pathways.....	118
6.2 Annex 2:.....	125
A sense of balance: experimental investigation and modeling of a malonyl-CoA sensor in Escherichia coli	125
6.3 Annex 3.....	140
Silencing of Antibiotic Resistance in E. coli with Engineered Phage Bearing Small Regulatory RNAs	140

Chapter 1 : Introduction
Inputs of synthetic biological systems

1.1 From genetic programs to synthetic biology

When Francois Jacob depicted cells as computers running genetic programs half a century ago, the metaphor aimed at providing the public, from young biologists to philosophers, with a fresh perspective on molecular level decision making in living systems. Behind this metaphor was his and Monod's earlier concept of "genetic circuit" that provided a framework for decades of systematic molecular dissections and understanding of DNA-based decision making. Did he also anticipated that this computer metaphor would later motivate not only the understanding of biological functions, but the programming of synthetic ones?

« Une chose est certaine, si un programme informatique est une séquence d'instructions qui spécifie étape par étape les opérations à effectuer pour obtenir un résultat, et s'exprime sous une forme qui permet de l'utiliser avec une machine, l'ADN en tant que programme et les mécanismes cellulaires en tant qu'exécutants n'est pas une notion dénuée de sens. À ce moment, une cellule peut être comparée à quelque chose de semblable à un ordinateur. » Jacob F., 1970 La logique du vivant. Une histoire de l'hérédité, Gallimard, Paris.

"If a computer program is a sequence of instructions that specifies step by step the actions required to obtain a result, and comes in a form that can be used by a machine, then DNA as a program and cellular mechanisms as executing agents is a valid image. Therefore, a cell can be compared to something similar to a computer."

(Translated from the French by the author of this thesis)

What is certain is that this image played a role in motivating a few computer scientists and engineers at the dusk of the 20th century to switch their careers toward biology with the aim of converting it into an engineering field. The founding promise of the young field of synthetic biology is that the fine level of characterization of biological parts made possible by the omics era technologies, coupled with a rigorous engineering mindset (abstract hierarchical thinking, modularization and decoupling, part standardization) and toolset (automation and computer-aided design) would enable forward engineering of biology with the formidable scalability observed in other fields of engineering, especially in the mechanical and electronical counterparts. This ambition, and the conceptual shifts it would require in order to be fulfilled, rather than a technology or a scientific milestone, probably define best the birth of synthetic biology as a field. Relatively straightforward *ad hoc* modifications of DNA predate synthetic biology and led to several major medical revolutions without the need for a jargon coming from

electronics. However, in the face of exponentially increasing capabilities for reading and writing DNA, scalability matters, and “pre-synthetic” biology was arguably not scalable (see Figure 1).

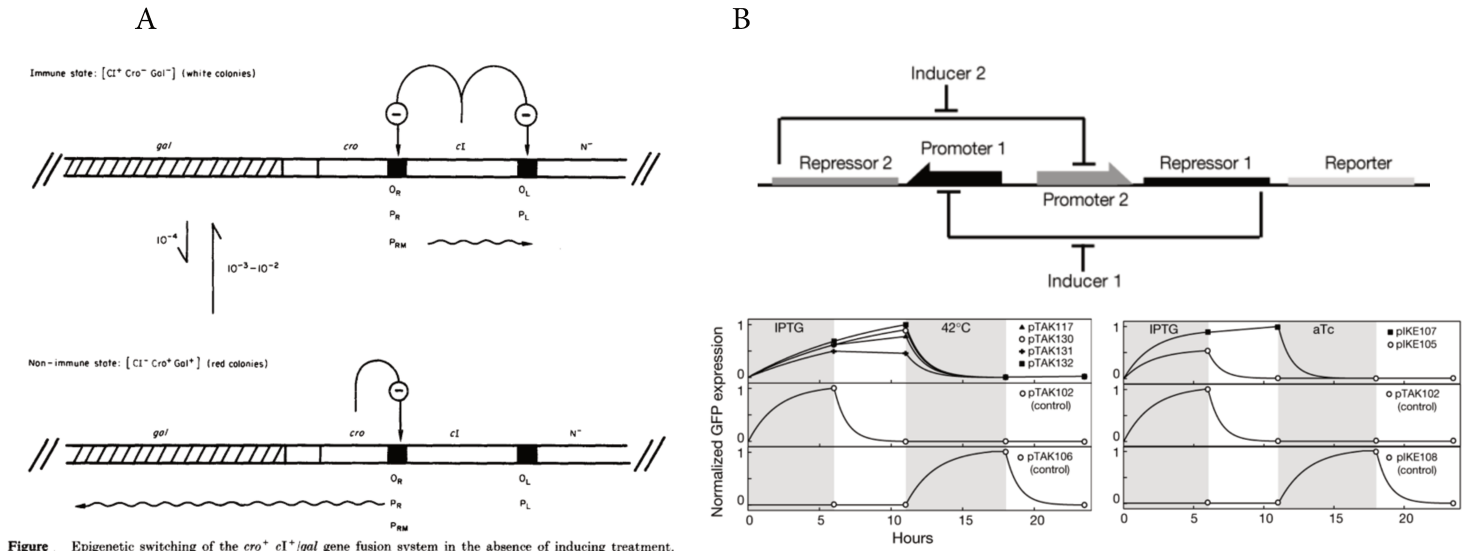


Figure 1: Epigenetic switching of the *cro*⁺ *cI*⁺/*gal* gene fusion system in the absence of inducing treatment.

Figure 1: Bistable synthetic genetic switches. A) Potentially the first “synthetic” genetic toggle switch in 1980 by Toman, Dambly and Radman[1]. B) Contrary to the 1980 version derived from the lambda phage lysis/lysogeny switch, the version by Gardner, Cantor and Collins[2] in 2000 is forward engineered with modules made of non-specialized regulatory components and the process can be guided by a simple mathematical model. In this seminal synthetic biology article, different proteins and promoters were tested in parallel designs, leading to switches with different induction mechanisms and dynamics.

The question whether the term synthetic biology will survive the hype cycle or not is probably not important (except maybe for the founders). However, the hype itself have already had an overall positive impact by triggering the interests for research of thousands of young scientists through the iGEM competition and strongly fostering interdisciplinary in biology. It helped the spread of synthetic biology’s top priorities, the use of new metrology standards, an exigence of detailed characterization and data-sharing, in a biology partially ravaged by the reproducibility crisis. It even managed to redirect some of the silicon-valley investments usually dedicated to developing apps into white and red biotechnology companies[3]. At this point, would the field be split or change name for new marketing reasons, the 15 years of efforts, of interdisciplinary cross-pollination, of contributions for science and for bioethics wouldn’t go. The new genetic tools, cloning methods, automated workflows, cloud laboratories, models, algorithms, and design frameworks developed so far, with or for the principles of

synthetic biology, will have a profound impact on biology and biotechnology in the 21st century.

1.2 The promise of scalability

Aside from the debate about the birth date or the exact nature of synthetic biology, a more interesting question is whether its founding promise is achievable. Systems biology successfully used modelling, sometimes using the conceptual framework of interacting modules, to reverse engineer several biological systems, but partial understanding of biological complexity is not necessarily sufficient to build novel complex biological systems. Evolution saves Life the need for predictions, but engineers cannot afford for their designs to fail at the same scale. Unlike in electronics, where knowledge of only a handful of physical parameters describing basic components is sufficient to build functional circuits made of billions of parts, the enthusiasm of the first biological switches and oscillators disenchanted after unsuccessful attempts at scaling them up in larger circuits. The complexity of circuits stagnated in the face of unpredictability and failure due to issues in part-interoperability and context-dependency[4]. While these issues are far from being resolved, recent advances finally pushed the limits of the size and robustness of designer circuits. In addition to improved methods in circuit design, biopart mining and characterization, these advances particularly benefited from a change of paradigm in the bioengineer's design-build-test process. New cloning methods, automation and a drastic drop in the cost of gene synthesis allowed a few well-funded centers and companies around the world to enter an era where "build" is not the limiting step anymore. In 2016, synthetic biology company Ginkgo Bioworks used 100 million base pairs of chemically synthesized genes, equaling 10% of the 2015 global volume of gene synthesis. The company, which raised 154M\$ from venture capitalists in less than two years, already placed an order of 600 million base pairs for 2017. With such capabilities, design can become a potential bottleneck and must be partially automated through computer-aided design (CAD) tools. The laboratory of C. Voigt reached the limits of human manual design abilities when they achieved layering of multiple logic gates able to perform computations on chemical input signals[5]. To overcome this limit, they developed a software that directly outputs DNA sequences corresponding to user-specified circuit function[6]. After assembling 880,000 base pairs in various signal computing circuits with up to 12 regulated promoters, most of the circuits behaved as expected without additional tuning (see **Figure 2**). Alternative approaches to scale the signal processing abilities of synthetic

circuits have relied on real-time editing of DNA to store, amplify and compute information[7–9]. CAD tools were also developed in this case to automate the design of finite state machines, a classic kind of logic circuits, with up to 16 output DNA states depending both on the nature of the chemical inputs but also on the order in which they appear[10].

Are these recent achievements the demonstration of scalability in genetic circuit design? Concerning the ability to use DNA as a computing medium to process signals in a predetermined way, we are indeed probably close to it. However, a striking aspect in all these works is the systematic use of the same chemical inputs. From the early 2000's to this day, the vast majority of contributions in the area of synthetic signal processing have been based on only half a dozen chemical inputs and their associated biosensors, in particular IPTG detected by the transcription factor LacI, arabinose by AraC, anhydrotetracycline by TetR, and N-Acyl homoserine lactones by LuxR. While proof of principle studies claim that these molecules can be easily replaced by more relevant inputs in order to tackle real-world problems, the hard truth is that very few biosensors are available to plug on these circuits. Facilitating division of labor is one of the main advantage of modularity and abstraction hierarchies, and these concepts made possible engineering adventures like the Apollo program which gathered and coordinated up to 400,000 individuals. Synthetic biology principles may allow such levels of cooperation at some point but the part of the field specialized in signal processing modules clearly outpaced the rest of the field for the moment. Achieving true scalability in biological engineering necessitates all the interdependent subfields in synthetic biology to be scalable themselves. Currently, the engineering of new inputs is lagging behind and our complex genetic constructions are like powerful computers with no mouse, no keyboard, no USB or Ethernet ports. For excelling at processing signal is only useful with relevant signals to process, new methods to develop biosensors tailor-made for relevant molecules are highly needed.

Cello design specification

Sensors	name	low	high	promoter sequence
PTac	0.003	2.8		AACGATCGTTGGCTGTGTTCGAATTAATCATC
PTet	0.001	4.4		TACTCCACGGTGGCTTTTTCCTATCAGTGA
PBAD	0.008	2.5		ACTTTTCATCTCCGCGCATTCAGAGAAGAAC

```

Verilog
module 0x21(output out,input A,B,C);
always@(C,B,A)
begin
case((C,B,A))
3'b000: {out} = 1'b0;
3'b001: {out} = 1'b0;
3'b010: {out} = 1'b1;
3'b011: {out} = 1'b0;
3'b100: {out} = 1'b0;
3'b101: {out} = 1'b0;
3'b110: {out} = 1'b0;
3'b111: {out} = 1'b1;
endcase
end
endmodule
        
```

Run

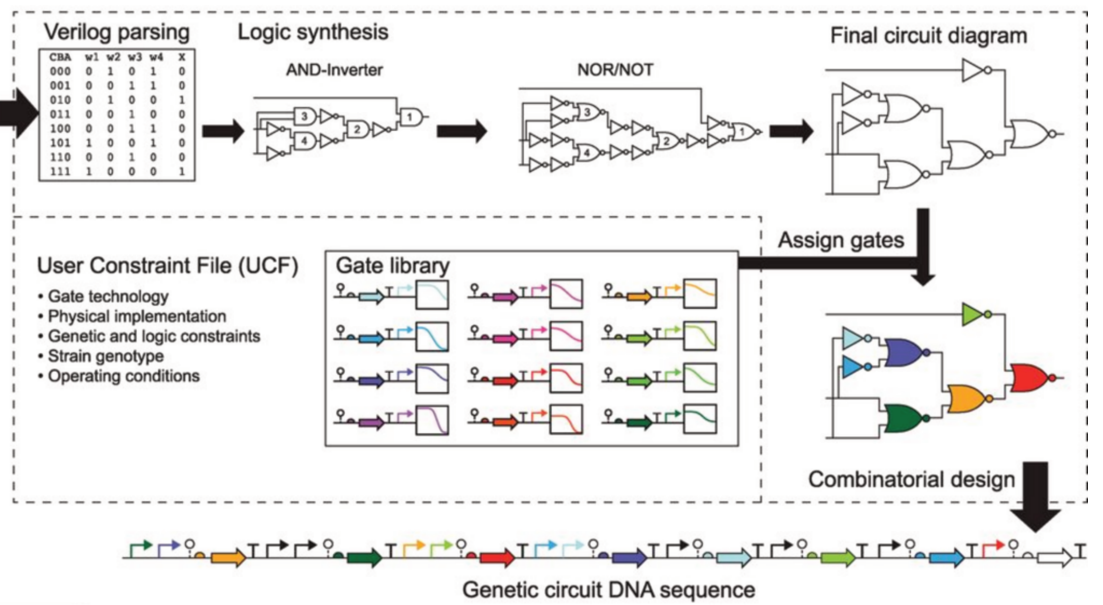
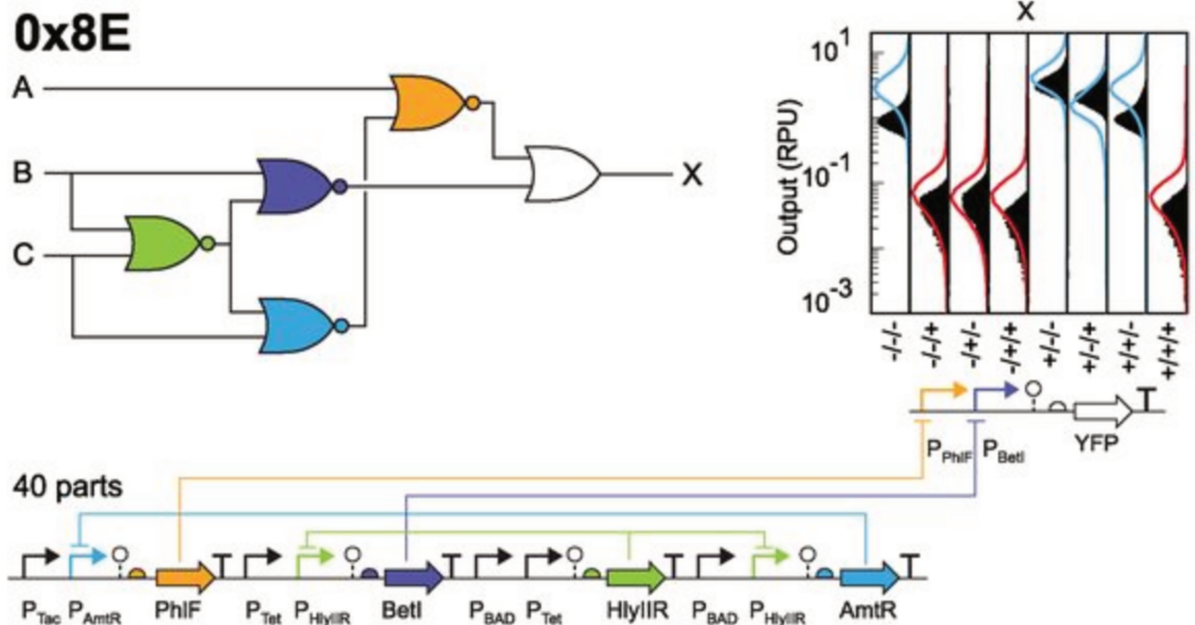


Figure 2. Genetic circuit design automation. From Nielsen et al. 2016. A) The Cello design environment accepts user requests in the form of Verilog code (hardware description language) and outputs DNA sequences. B) 0x8E, one of the 45 DNA sequences produced by Cello whose experimental FACS characterizations (black distributions) are congruent with the predicted ones (red and blue curves) in all possible combinations of input signals (IPTG/aTc/Arabinose).

1.3 Current inputs

Inputs of synthetic biological systems typically take the form of a change in the environment such as changes in chemical concentration, temperature or light, that can lead to a controlled change in the expression of one or several genes. Such change in expression can take place at the level of transcription or translation and it usually triggers a cascade of actuations in the engineered circuit. This view of inputs involves biosensors like riboswitches and transcription factors but exclude other biosensors such as aptamers or FRET sensors that do not influence gene expression even though they can be genetically encoded too.

Synthetic biologists have tried to isolate and adapt several natural biosensors to make them work into their own circuits (see **Figure 3**). Allosteric transcription factors (one-component systems) are by far the preferred system to allow input, but two-component systems have also been used[11,12], as well as GPCRs[13], and riboswitches. All these systems perform extremely well in their natural task but they are not equal in regards to portability to different hosts, ease of engineering and diversity of potential inputs. Two-component systems, from their very nature benefit from facilitated component interchangeability among a large choice of histidine-kinase (HK) domains and response regulator (RR) domains but at the same time suffer from non-specific phosphotransfer and crosstalk. More than 20 natural riboswitches have already been discovered and some of them have been successfully integrated in synthetic circuits[14]. Riboswitches could also be engineered to recognize new ligands. RNA aptamers binding to molecules of interest can easily be isolated via SELEX[15] or related methods[16] and it has been suggested that *de novo* riboswitch design could be achieved on the basis of such aptamers. However, the difficulty of the task has been initially underestimated and despite two decades of SELEX screenings very few functional riboswitches have been engineered. Most of the successful attempts consisted in reengineering versions of riboswitches binding to theophylline (the ligand for 60% of engineered riboswitches so far[17]) or neomycin and it is hard to anticipate when the know-how will reach the level necessary to develop riboswitches tailor-made for new ligands.

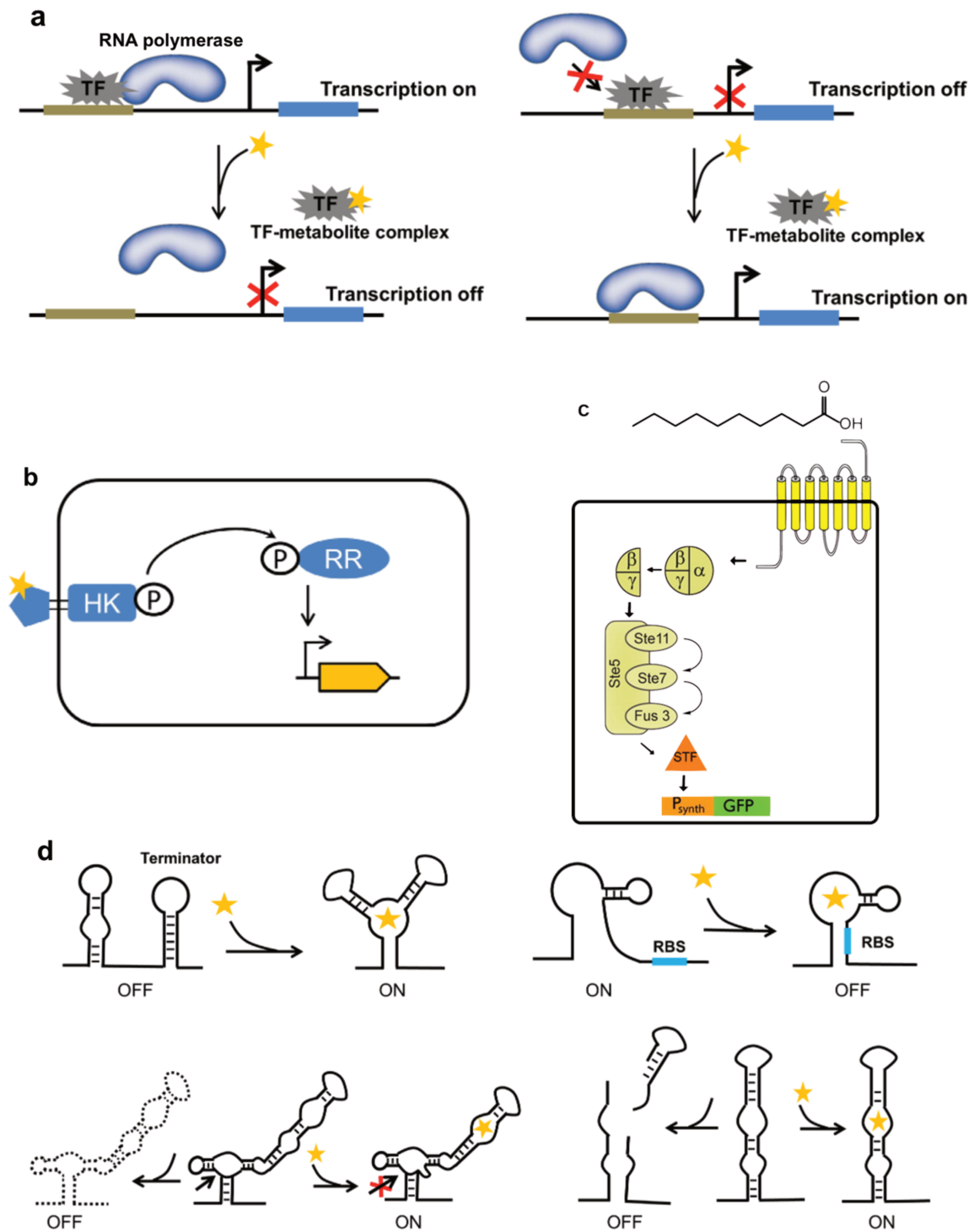


Figure 3 : Main classes of biosensors used to input signals in synthetic circuits. Figures a), b), d) adapted from Liu et al[18]. a) Basic mechanisms of allosteric transcription factors acting either as activator or repressor. b) Two-component system c) Mammalian GPCR adapted to yeast in order to detect medium chain fatty acid (figure adapted from Mukherjee et al[13]). d) Various mechanisms of regulation of translation by riboswitches.

Aside from these approaches inspired by biology, completely novel ways of influencing expression might emerge from synthetic biologists' minds. For instance, the need for post-translational control over the activity of the CRISPR-Cas9 system has led to the development of engineered versions of Cas9 inducible by light, or by chemicals such as 4-hydroxytamoxifen[19–21] or rapamycin[22]. While these works aimed at controlling gene editing, they could probably be reproduced with the nuclease-dead version dCAS9 that can behave as a repressor (CRISPRi[23]) or activator(CRISPRa[24]) of gene expression. This would open the way to a new kind of fully synthetic inducible transcription factors. Another strategy identified ligand-binding domains (LBD) that are unstable and are quickly destroyed in the cell unless a small molecule is present to stabilize them. By placing these unstable LBD in between a DNA-binding domain and a transcriptional activation domain, digoxin and progesterone-dependent control of gene expression has been achieved in a completely synthetic way[25]. This thesis later describes an approach of biosensing that was also based on rational design instead of being inspired by nature (see **Chapter 2**). However, calling it synthetic would imply it does not take place in nature, which turned out to be wrong as further investigation led to the identification of several similar natural systems (see **Chapter 3.5**).

1.4 Overview of the thesis

The research efforts deployed over the course of this thesis project aimed at trying to relieve some of the inputs-related problems in synthetic biology, whether by increasing the number of potential chemical inducers or by introducing a viral-based strategy for controlling gene expression.

In **Part I - “Expanding the scope of chemical inputs”**, I first describe a strategy that contrasts with prior approaches to develop biosensors for new chemicals: instead of trying to modify the structure of natural biosensors to make them detect new compounds, I tried to modify the chemical structure of these compounds so that they can be detected by existing biosensors. I show that information about a given compound's concentration in the medium can be faithfully transmitted to the genetic layer of an organism through covalent bonds formation and destruction catalyzed by a synthetic metabolic layer. The systematization of this strategy and evaluation of its potential is described in **Chapter 2**, which is a reproduction of the article “Expanding

biosensing abilities through computer-aided design of metabolic pathways” published in ACS Synthetic Biology in 2016 by Libis, Delépine and Faulon. In order to allow the community to access to our design abilities and foster the use of this approach for biosensor development, we then developed a webserver, described in **Annex 1 (Chapter 6)** and published as “Sensipath: computer-aided design of sensing-enabling metabolic pathways” in Nucleic Acids Research in 2016 by Delépine, Libis, Carbonell and Faulon. In **Chapter 3**, I provide context about other promising approaches to expand the chemical space detected by biosensors with a focus on what currently stands out as the most promising system: bacterial allosteric transcription factors. This part covers approaches to mine existing transcription factors in nature as well as engineering efforts based on evolution and rational design. I then review the use of indirect sensing through metabolism as an input strategy in biotechnology and also discuss the potential occurrence of this underestimated strategy in nature after listing examples in several bacteria. This part has been published as “Sensing new chemicals with bacterial transcription factors” in Current Opinion in Microbiology in 2016 by Libis*, Delépine* and Faulon. **Annex 2 (Chapter 6)** features an example of application of transcription factors in metabolic engineering, where a biosensor is used to identify variants of metabolic pathways that produce high level of malonyl-CoA in *E. coli*. This is a reproduction of the article “A sense of balance: experimental investigation and modeling of a malonyl-CoA sensor in *Escherichia coli*” by Fehér, Libis, Carbonell and Faulon published in Frontiers in Bioengineering and Biotechnology in 2015.

Part II “Phages as biological inputs” explores the potential of using heavily modified phages, as a way to control gene expression over a population of bacteria. I propose the use of such phage as a form of “biological inducers” whose release in the medium and subsequent cell infection provides control over both silencing of chromosomal targets and expression of heterologous genes. This axis originated from the article “Silencing of Antibiotic Resistance in *E. coli* with Engineered Phage Bearing Small Regulatory RNAs” which was produced in the context of the international Genetically Engineered Machine competition (iGEM), as one of the four subprojects that led the team Paris Bettencourt to win the 2013 Grand Prize of the competition at MIT in Boston. This article is reproduced in **Annex 3 (Chapter 6)** and was published in ACS Synthetic Biology in 2014 by Libis*, Bernheim*, Basier, Jaramillo-Riveri, Deyell, Aghoghogbe, Atanasković, Bencherif, Benony, Koutsoubelis, Löchner, Marinkovic, Zahra, Zegman, Lindner and Wintermute. On the basis of this demonstration of the feasibility of silencing antibiotic resistance genes with engineered phages, **Chapter 4** describes the design of such phage and their engineering to target and silence any gene

of interest on the chromosome. This work was published in the article “Phage-mediated delivery of targeted sRNA constructs to knock down gene expression in *E. coli*” in Journal of Visualized Experiments in 2016 by Bernheim*, Libis*, Lindner and Wintermute. **Chapter 5** is a preliminary work building on this concept of biological induction in the context of metabolic engineering for bioproduction of diverse chemicals. The field of metabolic engineering currently suffers from stringent design constraints imposed by the need for inducible gene circuits. I try to show that many modifications of the host metabolism typically pursued in classical metabolic engineering, i.e. timed expression of heterologous genes, down-regulation of competing pathways, overexpression of feedback resistant copies of enzymes, can be reproduced on the fly upon phage infection, thus opening new design landscapes for metabolic engineers and potentially a viable alternative to chemical induction in industrial fermentation.

Part I : Expanding the scope of chemical inputs

Contributions of the author

- Expanding biosensing abilities through computer-aided design of metabolic pathways. Libis V, Delépine B, Faulon JL. ACS Synthetic Biology, 2016
doi: 10.1021/acssynbio.5b00225 (Chapter 2) :

My contributions to this work consisted in the following: designing the project, coding an early version of the algorithm that predicted sensing-enabling metabolic pathways for cocaine and parathion (which were later successfully implemented *in vivo*). However, this first piece of code was not efficient and missed many potential pathways, thus precluding any statistics or large scale *in silico* screening of SEMP. Baudoin Delépine kindly lent his coding skills to rewrite it from scratch with the highest standards which enabled all biochemical reactions databases and all transcription factor databases to be used as source, large datasets of target molecule to be screened and clean statistics about the potential of the method to be obtained. From these new predictions and through systematic literature checking I curated several other candidate SEMP to implement, designed the DNA sequences, performed the recombinant DNA work, designed and executed the experimental characterizations. I then processed the experimental data and wrote the article.

- Sensipath: computer-aided design of sensing-enabling metabolic pathways. Delépine B, Libis V, Carbonell P, Faulon JL. Nucleic Acids Research, 2016
doi: 10.1093/nar/gkw305 (Chapter 6 – Annex 1):

I seconded Baudoin Delépine in this work by providing a biologist's point of view to help design a user friendly webserver that would foster the use of SEMP strategies among the synthetic biology community. I helped writing the documentation and the article. I also took care of parsing an additional database (EAWAG) and standardizing its entry to add them to the reaction list.

- Sensing new chemicals with bacterial transcription factors. Libis V*, Delépine B*, Faulon JL. Current Opinion in Microbiology, 2016 *contributed equally
doi: 10.1016/j.mib.2016.07.006 (Chapter 3) :

I wrote the introduction, conclusion, the first part “mining allosteric transcription factors in nature” and the third part “regulation via indirect sensing”. Part “engineering the sensing scope” was written by Baudoin Delépine who provided his expertise *with in silico* design of transcription factors.

- A sense of balance : experimental investigation and modeling of a malonyl-CoA sensor *in Escherichia coli*. Fehér T, Libis V, Carbonell P, Faulon JL.

Frontiers in Bioengineering and Biotechnology, 2015

doi: 10.3389/fbioe.2015.00046 (Chapter 6 - Annex 2) :

My contributions to this work consisted in the following: I wrote the introduction which took the form of a mini review about applications of biosensors in metabolic engineering. I had a very minor role in generating experimental data under the supervision of Tamás Fehér.

Chapter 2 : Expanding biosensing abilities through computer-aided design of metabolic pathways

This work has been published in ACS Synthetic Biology in 2016 by Vincent Libis, Baudoin Delépine and Jean-Loup Faulon.

Only very minor modifications have been made between the article and the chapter presented below.

2.1 Abstract

Detection of chemical signals is critical for cells in nature as well as in synthetic biology where they serve as inputs for designer circuits. Important progress has been made in the design of signal processing circuits triggering complex biological behaviors but the range of small molecules recognized by sensors as inputs is limited. The ability to detect new molecules will increase the number of synthetic biology applications but direct engineering of tailor-made sensors takes time. Here we describe a way to immediately expand the range of biologically detectable molecules by systematically designing metabolic pathways that transform non-detectable molecules into molecules for which sensors already exist. We leveraged computer-aided design to predict such sensing-enabling metabolic pathways and we built several new whole-cell biosensors for molecules such as cocaine, parathion, hippuric acid and nitroglycerin.

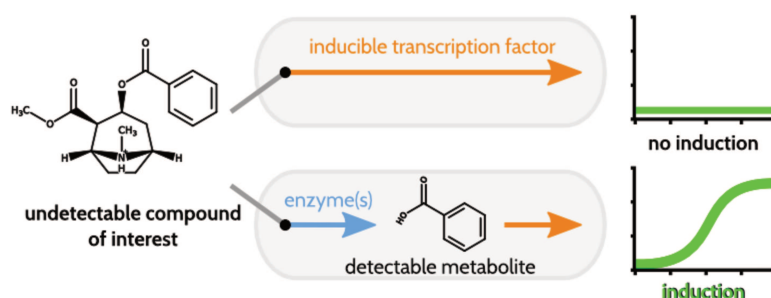


Figure 4. graphical abstract

2.2 Introduction

Engineering of circuits in cells has made fast progress since the dawn of synthetic biology. New modular tools and strategies regularly expand the toolbox[26]. Just considering progress made in the last two years, signal processing in biological systems can now rely on elements such as load drivers[27], memory systems[28], amplifiers[29], coupling systems[30] or bow-tie architectures[31] to name a few. Such tools enable the use of synthetic circuits in real life applications where the complexity of the signals encountered in the environment was until now problematic[32]. Despite these intense efforts allowing precise control of circuit behavior, the development of applications is slowed by the limited number of inputs available. Usually, inputs to the genetic layer of circuits are mediated by sensors such as transcription factors, riboswitches or two-component signaling pathways. Unfortunately, the number of organic molecules detectable by well-characterized natural sensors is relatively small[33]. Rational engineering of sensors through protein engineering or riboswitch engineering have been accomplished[34,35] but the time and effort necessary to deploy such approaches still limits the number of available tailor-made sensors.

Alternative strategies of biosensing could play a role in tackling this lack of inputs. In nature, information about a chemical signal can be indirectly conveyed through enzymatic transformations. A classic example can be observed in the Lac operon where information about the quantity of lactose in the medium is not acquired by direct interaction with a transcription factor. Instead a fraction of the available lactose is transformed by beta-galactosidase into allolactose which is the molecule detected by the transcription factor LacI. The use of metabolic transformation to convey information to the genetic layer have also been demonstrated in synthetic biology to detect an aromatic and to obtain cell to cell communication[36–38].

Here we explore the full potential of metabolism to enable detection of new molecules and thus expand the scope of chemicals that can serve as input in synthetic biology applications. We systematically search for enzymatic ways to transform undetectable molecules of interest into molecules detectable by existing biosensors (see **Figure 5**). This requires the design of tailor-made pathways out of thousands of individual enzymatic modules available in the pool of known biochemical reactions, and necessitates the development of computer-aided design (CAD) tools. As synthetic biologists developed CAD tools to guide the engineering of genetic circuits[39,40],

metabolic engineers created powerful computational methods in the context of small molecule production in microorganisms[41,42]. Among them, mathematical abstractions have been developed to represent and simulate biochemical reactions *in silico*. When fed with entire databases of known biochemical reactions, such tool can extract a set of biochemical reaction rules that can then be applied to any given substrate to generate potential products (i.e. predicted metabolites). Retrosynthesis tools can iteratively apply these rules to reconstruct natural or synthetic metabolic pathways. We leveraged this expertise to build a CAD tool exploiting 9,319 biochemical reaction rules to search for sensing-enabling metabolic pathways connecting molecules of interest to the genetic layer. We first evaluated the potential of this approach *in silico* by predicting sensing-enabling metabolic pathways (SEMP) of molecules such as drugs, biomarkers and toxics. We then tested several predictions in *E. coli* by assembling circuits made of heterologous enzymes and transcription factors. We observed the successful fluorescent response of *E. coli* to several molecules including cocaine, parathion, 2-chloro-4-nitrophenol, hippuric acid and nitroglycerin. In synergy with ongoing efforts of biosensor development, this strategy offers an immediate expansion of the scope of inputs for synthetic circuits and has the potential to open the way to new synthetic biology applications in fields such as medicine and environment.

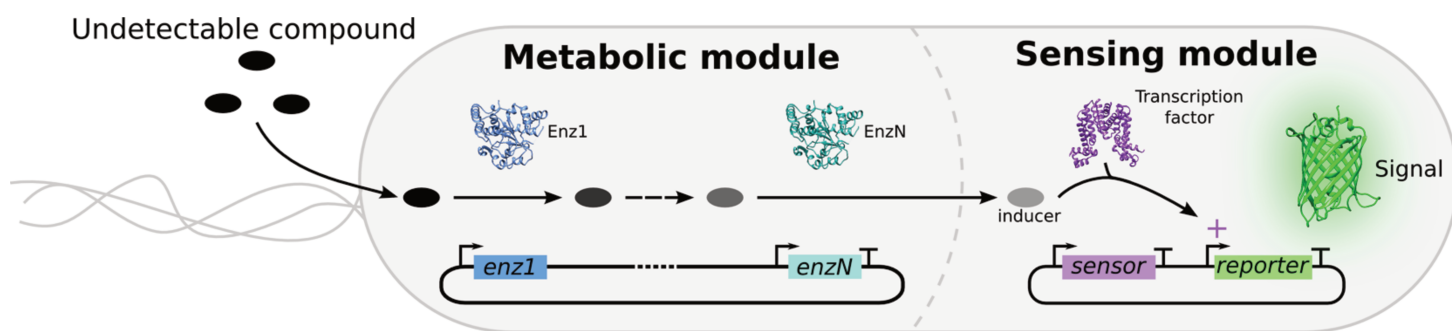


Figure 5: General concept of a cell equipped with a sensing-enabling metabolic pathway (SEMP) allowing for the detection of a new chemical. A naturally undetectable molecule is transformed by a metabolic module into an inducer molecule triggering a genetic response upon binding to a transcription factor. While the figure illustrates the case of transcription factor working as a positive activator, the same concept can be applied with a repressor or a riboswitch.

2.3 Results

2.3.1 The scope of detectable molecules is significantly expanded in silico via enzymatic pathways.

We evaluated the potential of this approach by predicting sensing pathways of target molecules such as drugs, biomarkers of human diseases and molecule with risk of toxicity for health and the environment. We gathered datasets representative of these types of molecules from three public databases: DrugBank[43], HMDB[44], Tox21[45]. Systematic design of tailor-made SEMP for these targets requires biochemical retrosynthesis and subsequent identification of inducers in the generated products. The workflow we used consists of four steps: (i) gather the whole trans-species Reactome (ts-Reactome) as a list of encoded biochemical reaction rules; (ii) submit each target to the ts-Reactome rules to generate products; (iii) iterate on the products and generate a hypergraph around targets; and (iv) screen the hypergraph for inducers and output putative SEMPs. The ts-Reactome was based on 3 major biochemical databases (BRENDA[46], Metacyc[47] and Rhea[48]) that were merged and encoded in reaction signatures[49], a mathematical representation of reactions that we developed previously. Reaction signatures have been proven useful for metabolic pathway design in the context of microbial production of value-added compounds[50]. The ts-Reactome reached a total of 9,319 unique reaction rules. When iteratively applying these reaction rules to the targets and their generated products, combinatorial explosions can be computationally demanding and we therefore limited to 3 steps the maximum length of the pathways. Once the hypergraph was generated it was colored with known inducers sourced from 4 databases of transcription factor effectors: BioNemo[51], RegTransBase[52], RegulonDB[53] and RegPrecise[54]. An overview of the workflow is represented in **Figure 6** and the labelled graph output from the DrugBank dataset can be seen in **Figure 7**. The pathways in the graph linking a target molecule to a natural transcription factor effector were automatically enumerated to allow statistical analysis and selection of proof of concept examples for *in vivo* implementation.

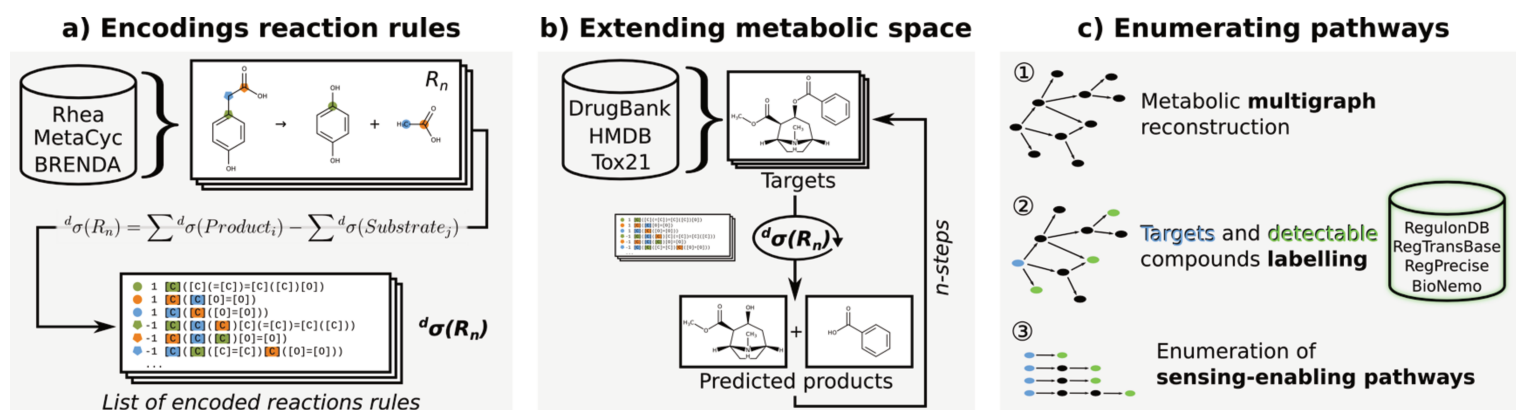


Figure 6: Sensing-enabling metabolic pathway (SEMP) design workflow. (a) Reactions are collected from biochemical databases and encoded as reaction signatures. A reaction signature describes a biochemical transformation rule as a list of fragments of the molecules that are modified during the reaction (See Methods). (b) The effect of each reaction signature is simulated on each target. If a product is predicted to form (i.e. if target's fragments match the ones of a natural substrate of the reaction) the generated structure is injected back into the target list, in case it can be further transformed. The metabolic space is extended by this iterative process until no more products are formed or if a limit is set on the number of iteration ($n=3$ in this work). (c) Finally, the extended metabolic space is represented as an oriented graph where the nodes are compounds and the edges represent enzymatic reactions. This graph is colored by labeling the compounds that are found in the transcription factor effectors databases. Pathways linking a target to a gene inducer can then be automatically enumerated.

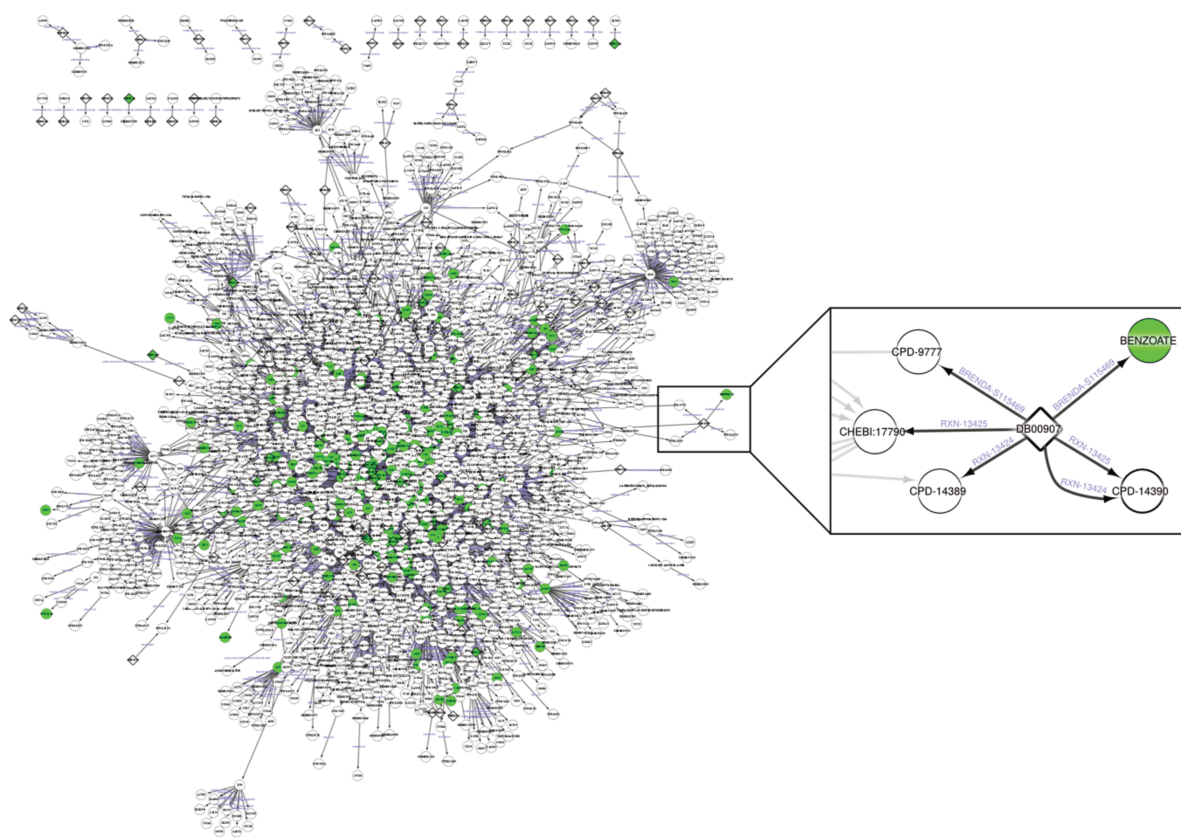


Figure 7: Extended metabolic space around the DrugBank dataset. This is an oriented hypergraph generated by the bioinformatic pipeline described in Fig. 2, here visualized with Cytoscape. Similar graphs were generated and (computationally) analyzed for the HMDB and Tox21 datasets. Nodes represent chemical compounds and edges represent biochemical reactions. Each compound is associated strictly to one node. Enzymes can be associated to different pairs of compounds, and thus can be associated to several edges. Diamond shaped nodes represent target compounds from DrugBank (illicit or accepted drugs) and circular shaped nodes represent compounds that are generated through up to three metabolic steps from the initial target compounds. Nodes colored in green are naturally detectable compounds found in the transcription factor effector databases. A zoom on the area around target compound DB00907 (cocaine) shows the predicted products of different enzymatic reactions including the natural effector benzoate. Compounds not involved with any reaction (“non-processed compounds”) are not represented.

In each of the considered target datasets a number of compounds were found to be naturally detectable by existing biosensors, respectively 59, 135 and 169 for DrugBank, HMDB and Tox21. Through metabolism, the number of detectable compounds grows to 123 (Drugbank), 280 (HMDB) and 477 (Tox21). Therefore, this approach at least doubled the number of detectable molecules that can be considered for sensor development in each of the studied datasets (Figure 8).

Most compounds in the datasets that remain undetectable are products of organic chemistry that are not processed by any known enzyme and therefore no SEMP could be predicted by this method. Remarkably, if a compound is processed by an enzyme the chances are high (>66%) that at least one SEMP will be found in 3-steps or less. Moreover, the connection with an existing biosensor is usually possible in a small number of enzymatic steps as the number of compounds that necessitate long SEMP (3 steps) is significantly lower than the ones connected through 1 and 2 step pathways. A short selection of interesting candidates for biosensor development and associated SEMPs are displayed in Table 1.

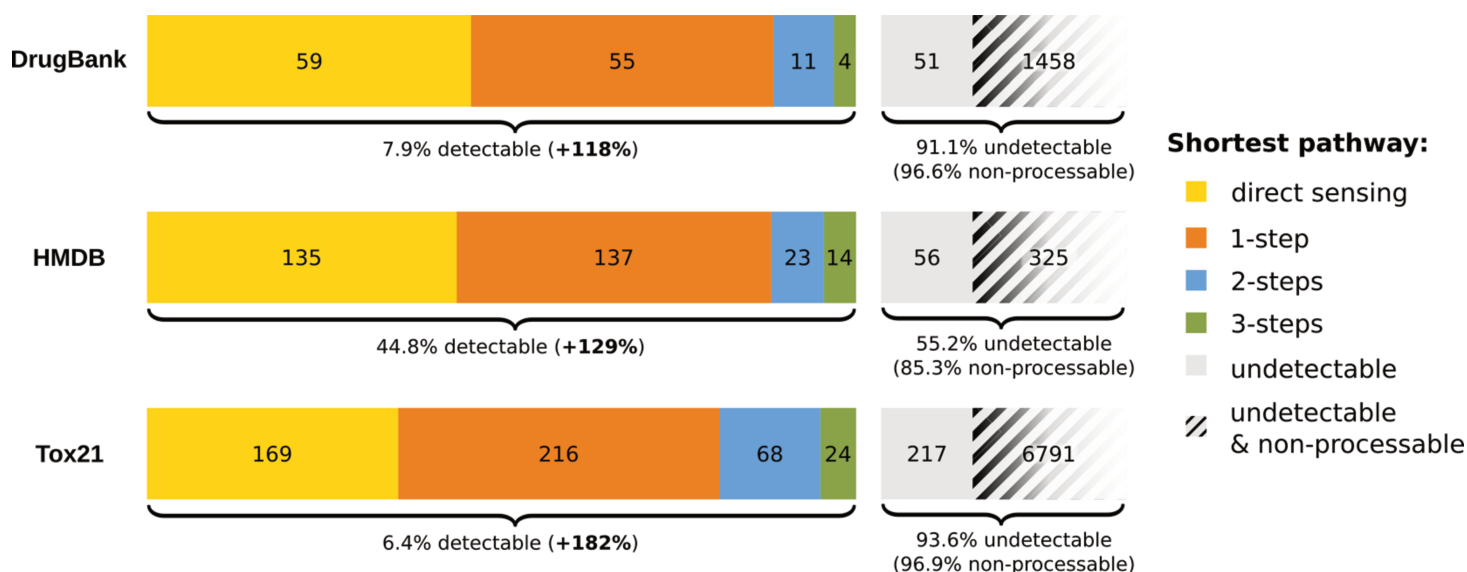


Figure 8: In silico prediction of detectable compounds among three datasets. Compounds labelled in the “direct sensing” category (yellow) are already present in the dataset of natural transcription factor’s effectors. Compounds in other colors (orange, blue, green) are reached by sensing-enabling metabolic pathways (SEMP). Non-processable compound (hashed) are compounds that do not participate to known enzymatic reactions (as a product or a substrate). Datasets correspond to DrugBank section “approved drugs” and “illicit drugs”, HMDB’s biomarker compounds associated with a human disease and the entire Tox21 dataset of putative toxic compounds.

Table 1: Sensing-enabling metabolic pathways (SEMP) examples^a

Target	Enzymes and metabolites	Sensor and effectors
Cocaine* illicit drug	Cocaine esterase (<i>Rhodococcus sp.</i>) 3.1.1.84 : Benzoate	BenR (<i>Pseudomonas putida</i>) Benzoate
Heroin illicit drug	Heroin esterase (<i>Rhodococcus sp.</i>) 3.1.-.- : Acetate (2x)	NR I (<i>Escherichia coli</i>) Acetate
Aspirin drug	Acetylsalicylate deacetylase (<i>Rattus norvegicus</i>) 3.1.1.55 : Salicylate	NahR (<i>Pseudomonas putida</i>) Salicylate
Caffeine drug	Methylxanthine N1-demethylase (<i>Pseudomonas putida</i>) 1.14.13.178 : Formaldehyde	FrmR (<i>Escherichia coli</i>) Formaldehyde
Paracetamol drug	Aryl acylamidase (<i>Rhodococcus erythropolis</i>) 3.5.1.13 : Acetate	NR I (<i>Escherichia coli</i>) Acetate
Barbituric acid drug	Bar (<i>Rhodococcus erythropolis</i>) 3.5.2.1 : Ureidomalonate N-malonylurea hydrolase (<i>Rhodococcus erythropolis</i>) 3.5.1.95 : Malonate	FapR (<i>Bacillus subtilis</i>) Malonate
Nitroglycerin* drug / explosive	NemA (<i>Escherichia coli</i>) 4.99.1.- : Nitrite (2x)	NarL (<i>Escherichia coli</i>) Nitrite
Chlorpropham* pollutant	AmpA (<i>Paracoccus sp.</i>) 3.5.1.- : 3-Chloroaniline	TadR (<i>Delftia tsuruhatensis</i>) 3-Chloroaniline
2-chloro-4-nitrophenol* pollutant	PnpA (<i>Burkholderia sp.</i>) 1.14.13.- : Chloro-1,4-benzoquinone PnpB (<i>Burkholderia sp.</i>) 1.6.5.- : Chlorohydroquinone	LinR (<i>Sphingomonas paucimobilis</i>) Chlorohydroquinone
Propanil pollutant	AmpA (<i>Paracoccus sp.</i>) 3.5.1.- : Propionate	PrpR (<i>Corynebacterium glutamicum</i>) Propionate
Parathion* pollutant	PTE (<i>Pseudomonas diminuta</i>) 3.1.8.1 : 4-nitrophenol	DmpR (<i>Pseudomonas sp.</i>) 4-nitrophenol
Hydrogen cyanide chemical warfare agent	Cyanide hydratase (<i>Gloeocercospora sorghi</i>) 4.2.1.66 : Formamide Formamidase (<i>Paracoccidioides brasiliensis</i>) 3.5.1.49 : Formate	FdsR (<i>Ralstonia eutropha</i>) Formate
Cyclosarin chemical warfare agent	PTE (<i>Pseudomonas diminuta</i>) 3.1.8.- : Cyclohexanol ChnA (<i>Acinetobacter sp.</i>) 1.1.1.245 : Cyclohexanone	ChnR (<i>Acinetobacter sp.</i>) Cyclohexanone
Hippurate* biomarker	HipO (<i>Campylobacter jejuni</i>) 3.5.1.32 : Benzoate	BenR (<i>Pseudomonas putida</i>) Benzoate

^a (*) indicates SEMP implemented and tested in vivo over the course of this work (see Figure 13).

2.3.2 Implementation of SEMP_s *in vivo* expands *E. coli* sensing abilities.

In order to validate the concept *in vivo*, we implemented in *E. coli* representative SEMP_s from the predictions related to each of the targets datasets (1 drug, 2 pollutants, 1 biomarker). For each SEMP, the genes coding for the required enzymes were cloned into a metabolic module plasmid allowing control of the enzymes' expression level by IPTG. In parallel the genes coding for the required heterologous transcription factors were cloned into a sensing module plasmid under the control of an arabinose-inducible promoter and a red fluorescent protein (RFP) was placed under the control of their associated promoters. The two modules are compatible for co-transformation in the same strain. Prior to full characterization of SEMP_s we identified optimal expression levels of the heterologous transcription factors by varying the arabinose concentration in presence of the natural effectors (see **Methods**).

Whole-cell biosensors for illicit compound detection can provide a cheap way to determine the presence of a molecule in an unknown mixture. Among the predictions we obtained by processing the DrugBank database's section "Illicit Drugs" we selected a SEMP providing *E. coli* with the ability to detect cocaine. As shown in **Table 1**, the transformation of cocaine by an esterase (CocE from *Rhodococcus sp.*) produces benzoate which can be detected by the transcription factor BenR from *Pseudomonas putida*. *E. coli* BL21(DE3) was first transformed only with the sensing module bearing the transcription factor BenR and a RFP under the control of pBEN promoter. This strain emits a strong fluorescent signal in presence of benzoate but not in presence of cocaine (**Figure 10a**). Upon co-transformation with a metabolic module containing CocE, a fluorescent response also occurs in presence of cocaine, indicating the expected extension of the sensing scope of the bacteria (**Figure 9a**). Comparison of the dose-response curves of the strain toward cocaine and the natural effector benzoate suggest that the enzymatic step have no or little impact on the biosensor performance features. The linear range of detection and dynamic range for cocaine stays in the same order of magnitude as benzoate.

Another interesting application is the development of whole-cell biosensors for monitoring chemicals in the environment. Parathion is a major environmental threat and counts among the "dirty dozen", the 12 worst offenders persistent organic pollutants according to the United Nations Environment Programme. The transformation of parathion by a phosphotriesterase (PTE from *Pseudomonas diminuta*) produces 4-nitrophenol which can be detected by the transcription factor DmpR from

Pseudomonas sp. (see **Table 1**). An *E. coli* strain harboring a sensing module based on DmpR and the associated DmpK promoter is able to detect 4-nitrophenol in a dose-dependent fashion but not parathion (**Figure 10b**). Upon addition of the metabolic module containing PTE to this strain, the extension of the sensing scope takes place and a strong fluorescent response occurs in presence of parathion (**Figure 9b**). To our surprise the fluorescent response of this strain to parathion is higher than for the native inducer above 100 μ M. The drop in signal in response to 4-nitrophenol at 300 μ M is probably due to the associated toxicity that diminishes growth (see Supplementary Fig 8). In addition to this effect on growth, flow cytometry measurements indicate a decreased fluorescence in individual cells at higher 4-nitrophenol concentration (See single cell data at 1mM in **Figure 12**). In contrast, parathion is less toxic for these cells and in fact is not toxic at these concentrations if the PTE enzyme is absent (see **Figure 16**). Moreover, despite the decrease in growth rate associated with the intracellular formation of 4-nitrophenol, the individual fluorescence of these cells in response to parathion keeps increasing gradually up to 1mM (see **Figure 12**). Partial or delayed transformation of parathion at high concentration may allow the cells to stay healthy longer and produce more signal.

Medical applications of synthetic biology often rely on biosensors for biomarkers of human diseases[55]. We chose to implement a biosensor for an interesting biomarker, hippuric acid, that is found at high concentration in the urine of a person intoxicated with toluene. As predicted by the algorithm, transformation of *E. coli* with a metabolic module harboring hippurase HipO from *Campylobacter jejuni* allows degradation of hippuric acid into benzoic acid, which is detected in a dose-dependent fashion by the BenR-based sensing module (see **Figure 9c**). We monitored the activation of this hippurate sensor and the cocaine sensor in order to investigate if SEMP's have kinetics of detection different from a regular benzoate sensor. Comparison of the response time to target compounds or to the natural inducer benzoate shows that neither HipO or CocE-mediated transformation step delays the fluorescent response (see **Figure 11**).

The engineering of transfer of matter through metabolism has made important progress during the last 20 years. Synthetic pathways as large as 23 enzymes long have been successfully implemented to divert natural carbon flux toward valuable compound production[56]. We wondered if transfer of information could as well rely on multi-enzymatic pathways and we selected a target compound that necessitates two consecutive transformations to allow detection. 2-Chloro-4-nitrophenol (2C4NP) belongs to a family of molecule with high toxicity for humans used in the chemical

industry. Upon monooxygenation and subsequent reduction performed by two enzymes from a strain of *Burkholderia sp.*, 2C4NP is transformed into chlorohydroquinone, which can be detected by LinR, a transcription factor found in *Sphingomonas paucimobilis* (see **Table 1**). All three genes were cloned into the modules and conferred *E. coli* the ability to detect 2C4NP (see **Figure 9d**), while it was not possible with the LinR-module alone (see **Figure 10d**). With our setup, the sensing module based on LinR transcription factor and a fragment of the LinE promoter from *Sphingomonas paucimobilis* had a small dynamic range in *E. coli*. This could probably be overcome by directed evolution as promoter activity can be linked to a selectable output. 2C4NP is highly toxic for *E. coli*, thus limiting our measurements to a 100 μ M upper limit. Remarkably, only the strain with the metabolic module could survive at a concentration of 2C4NP higher than 75 μ M (see **Figure 12**), this is probably due to the detoxifying effect of the two enzymes from *Burkholderia sp.*. This situation is the opposite to that observed in the case of the biosensing of parathion where the intermediate metabolite was more toxic to the cells than the initial target molecule.

Finally, an interesting case emerged from the predictions as a sensing-enabling pathway to the vasodilator drug and explosive nitroglycerin was identified with both metabolic module and sensing module component already present in wild type *E. coli*. Indeed, the promiscuous NemaA enzyme from *E. coli* is known to allow degradation of nitroglycerin into nitrites that are naturally monitored in *E. coli* by the NarL regulator. This suggests that wild type *E. coli* exhibits a fortuitous transcriptional response to nitroglycerin through a SEMP-like circuit. In order to test this hypothesis we transformed *E. coli* with a plasmid harboring a RFP fused to the native promoter pYeaR, which is controlled by NarL. This strain successfully produces a fluorescent response when cultivated in presence of nitroglycerin. To confirm the SEMP dependency of this response we transformed the same reporter plasmid into JW1642, an *E. coli* strain with a knocked-out NemaA enzyme. This strain is not anymore able to respond to nitroglycerin. Finally, we introduced a functional copy of NemaA on a second plasmid into JW1642 and observed a restoration of the fluorescent response to nitroglycerin (**Figure 9f**). This confirms that NemaA is enabling the transcriptional response to nitroglycerin of *E. coli* and represents an interesting example of a computationally elucidated naturally occurring SEMP. A summary of all the SEMP validated experimentally in this work is represented in **Figure 13**.

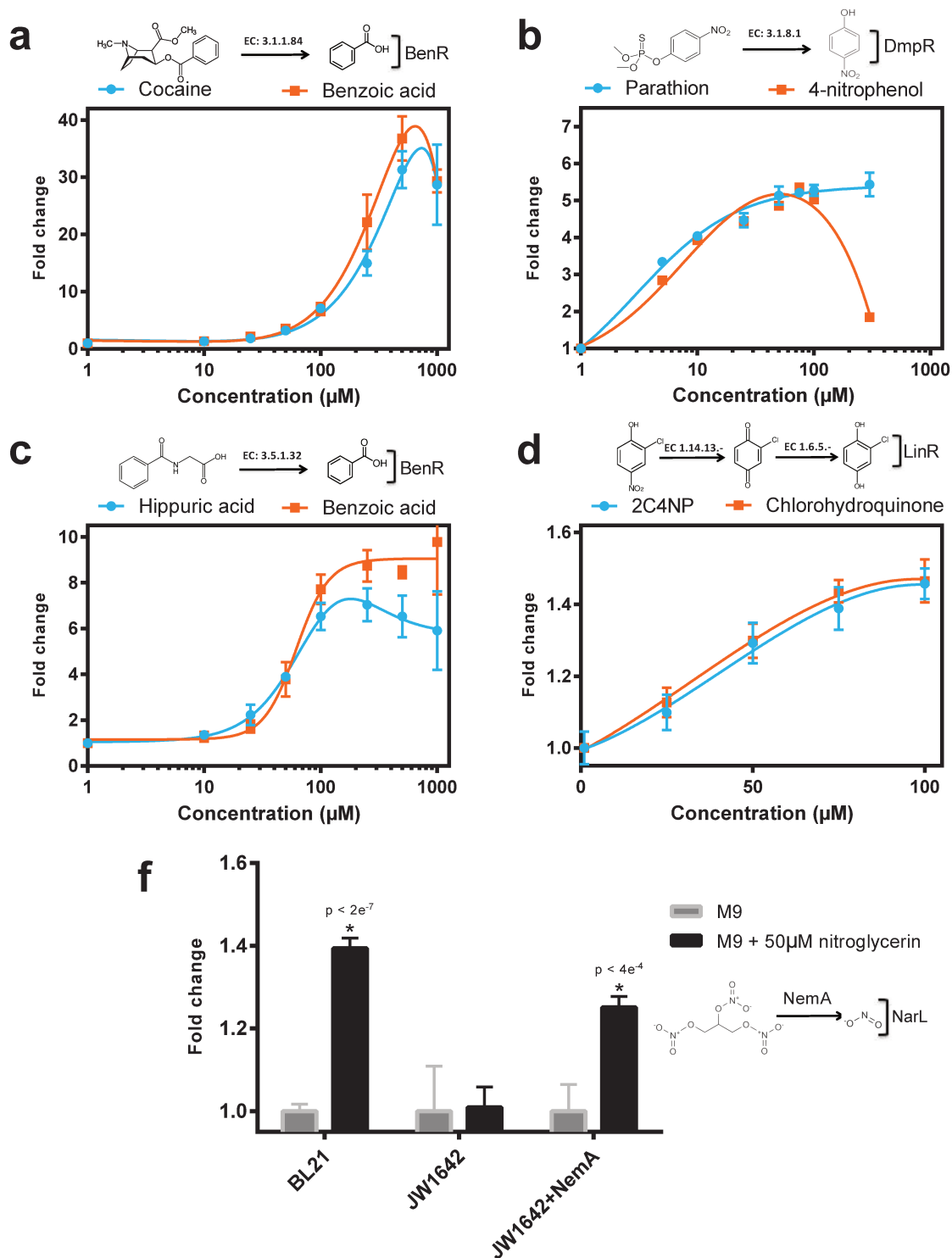


Figure 9: In vivo characterization of Sensing-Enabling Metabolic Pathways (SEMP). Cotransformation of *E. coli* with both metabolic module and sensing module confers sensing abilities toward new molecules. Dose-response relationship of engineered *E. coli* strains to the

natural effectors and to target molecules cocaine (a), parathion (b), hippurate (c), 2C4NP (d). (f) shows fluorescent response of three strains of *E. coli* harboring a RFP fused to pYeaR promoter (controlled by NarL) in M9 medium or in M9 supplemented with 50 μ M nitroglycerin. Laboratory strain BL21 responds to nitroglycerin but strain JW1642 which is knocked out for the NemA enzyme does not. However, response can be restored in JW1642 by providing a functional copy of NemA on a plasmid. Fold change represent the ratio between the fluorescence observed at a given concentration and the fluorescence of uninduced controls of the same strain, a value of 1 represents no induction. Each data point is the mean of at least 3 replicates and error bars represent standard deviations. Smooth curves represent dose-response fitting of the scattered data points. Statistical significance was determined using Student's t test with a P value cutoff of 0.005.

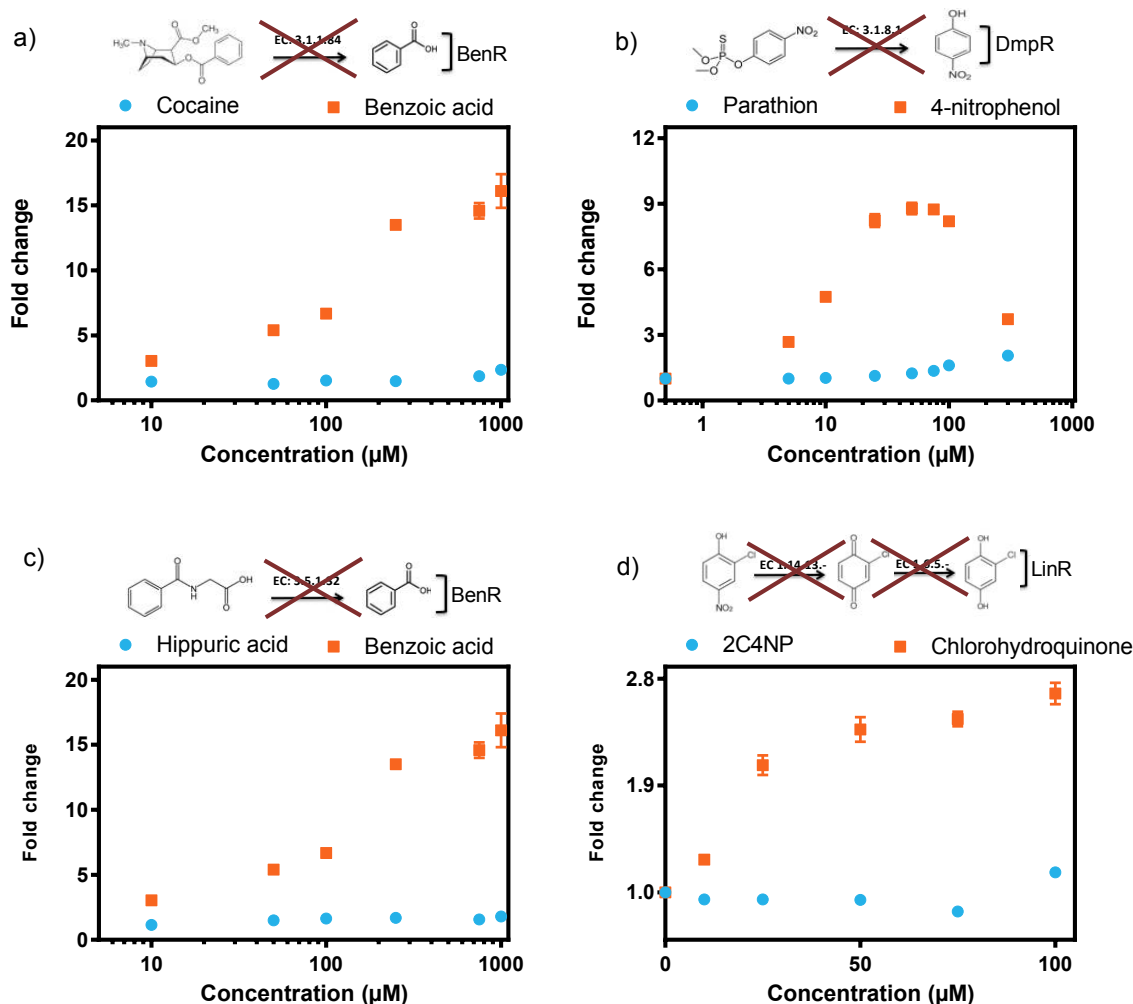


Figure 10: Dose response of control strains lacking metabolic module to the different compounds. (a)(c) Benzoate sensing module alone; (b) 4-nitrophenol sensing module alone; (d) chlorohydroquinone sensing module alone. Dose-dependent response is limited to the natural effector of transcription factors BenR, DmpR, LinR (benzoic acid, 4-nitrophenol, chlorohydroquinone) when metabolic module is not present to enable sensing scope extension to

target molecules. At high concentration of target molecules, traces of signal appeared even without metabolic module, this may be due to partial spontaneous hydrolysis, impurities from synthesis or promiscuity of the transcription factor.

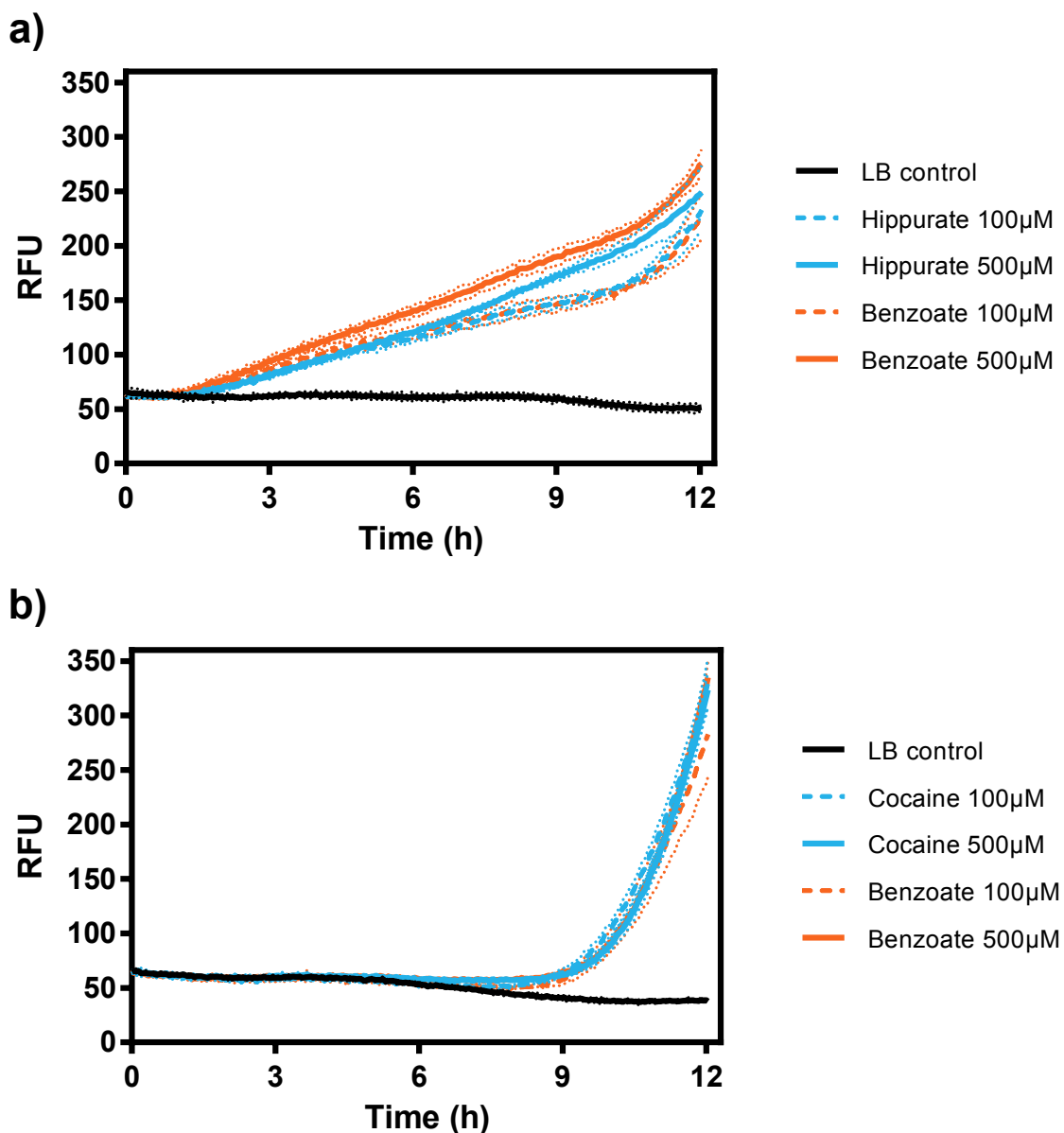


Figure 11: Monitoring of the response time to hippurate (a), cocaine (b) and benzoate of strains equipped with SEMP. No delay is observed between response time to the natural inducer benzoate and to the SEMP targets hippurate and cocaine whether in presence of 100μM or 500μM of inducers. Cells were prepared in the same way as for dose-response characterization experiments (see Methods). The longer response time for the cocaine sensor might be explained by the difference of temperature at which the experiment was performed (30°C instead of 37°C for hippurate sensor). RFU stands for relative fluorescence units and consists in the raw fluorescence value measured by the platereader. Dashed lines represent standard deviation of the three replicates.

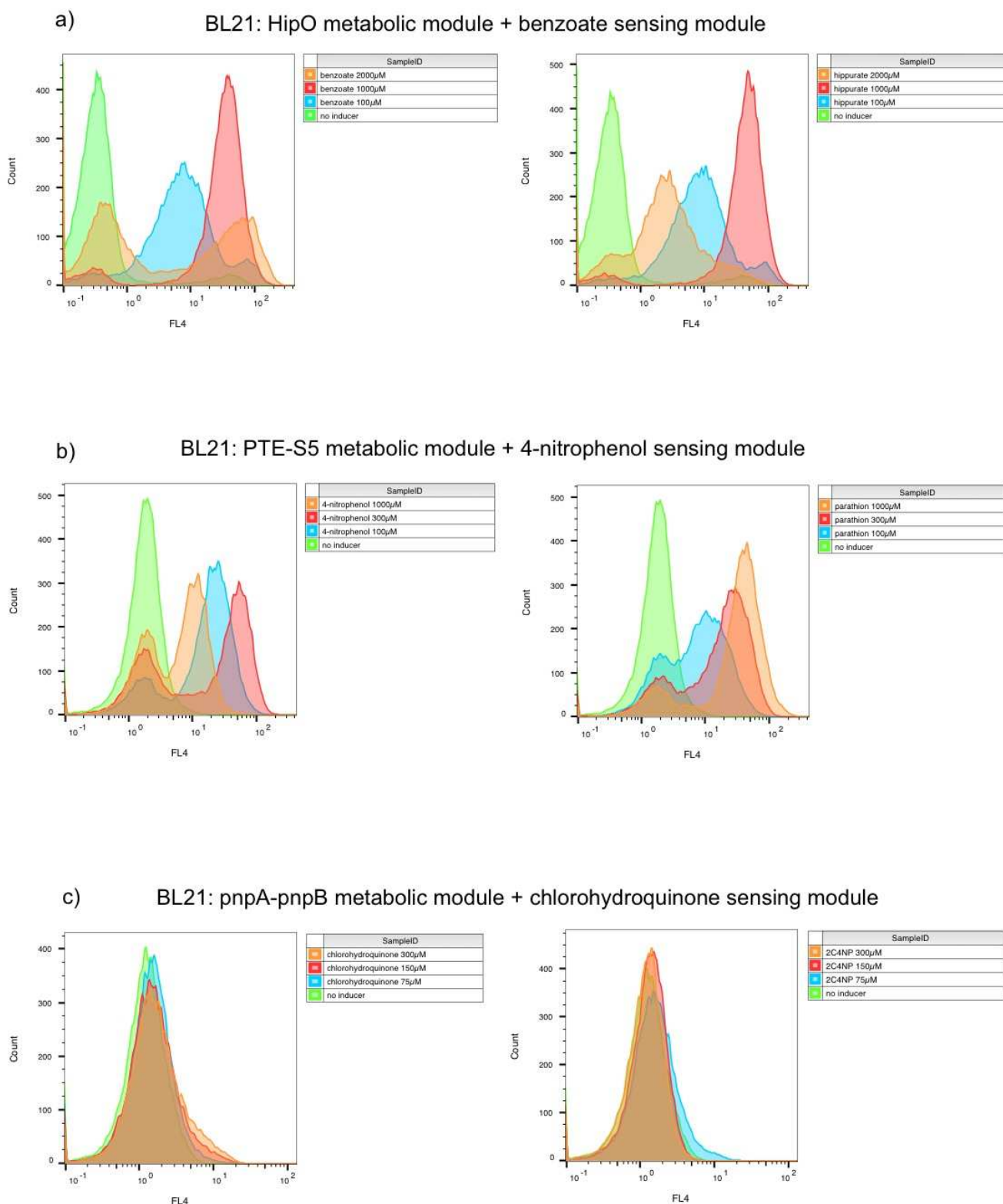


Figure 12: Flow cytometry analysis of strains equipped with SEMP in presence of toxic concentrations of inducer. Cells were cultivated in the same conditions as for dose-response characterization (see Methods) and red fluorescence was analysed on a CyFlow Space cytometer (PARTEC) counting 20,000 events for each sample. (a) Toxicity of benzoic acid leads to bimodal

distribution of cell fluorescence at 2mM and half of the population emits more fluorescence than with lower concentrations. Interestingly, hippuric acid toxicity has a different effect on cell response as fluorescence at 2mM decreases below the level observed with 100µM. (b) Toxicity of 4-nitrophenol leads to lower level of fluorescence at 2mM than at 100µM, similarly to what is observed with hippuric acid. On the contrary, proportionality of the response to parathion is maintained even at toxic levels. As shown in Figure 16, parathion by itself is not toxic unless PTE-S5 enzyme is present to transform it into 4-nitrophenol. This is probably the cause of the robust graded fluorescent response at high concentration although the exact mechanism is unknown. (c) Despite the low dynamic range of LinR, a small but graded response to chlorohydroquinone is observed while response to 2C4NP only occurs at low concentration (75µM) and is completely abolished upon reaching toxic levels.

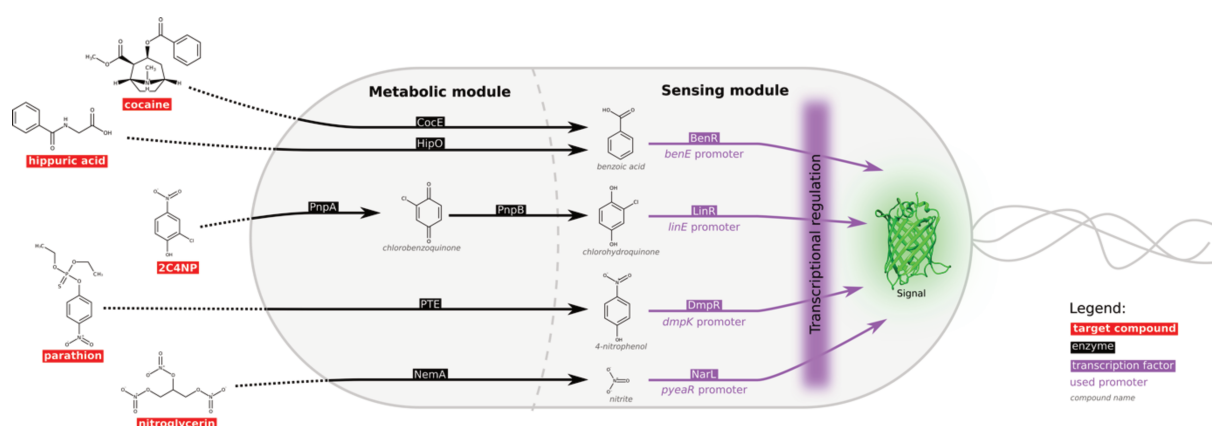


Figure 13: Summary of the SEMP successfully implemented in vivo over the course of this work. Only one pair of the represented metabolic and sensing modules was transformed in each strain. Implementation of a sixth SEMP allowing detection of the pollutant chlorpropham (see Table 1) was not successful because we were unable to obtain a response to 3-chloroaniline in *E. coli* based on the TadR transcription factor from *Delftia tsuruhatensis*. This failure may be due to a gene induction mechanism that necessitates compatibility with the host transcription machinery or that requires a co-inducing protein. Intuitively, the choice of SEMP elements coming from organisms closely related to the host in terms of evolution is expected to maximize chances of successful implementation.

2.4 Discussion

Biological sensors are central for synthetic biology to solve real-world problems. Numbers of promising systems involving biosensors have been developed for medical, environmental and industrial applications. The approach described here at least doubles the number of molecules that can be considered for sensor development in each of the studied compound classes (toxics, biomarkers and drugs). SEMP can be created without the need for time-consuming protein and riboswitch engineering or discovery of natural sensors. Moreover, the number of chemical candidates for such a sensing strategy should automatically grow in time as the number of characterized sensors and biochemical reactions will continuously increase in databases. For these reasons, we envision SEMP as an important new source of biosensors that will fruitfully be integrated within the modular synthetic biology toolbox.

Very recently a SEMP approach was used for the monitoring of a product of interest for industrial bioproduction, 3-hydroxypropionate[57]. By increasing the number of value added chemicals that can be monitored through fluorescent readout, SEMP could alleviate the bottleneck that metabolic engineers currently face with the low throughputs of conventional measurement methods. As another example of potential application, the hippuric acid sensor described here could be a starting point for a cheap way of controlling urine from workers in the paint industry, especially in the developing world where a lot of toluene intoxications occur due to lack of regular testing. However, this application requires from the cells to emit a signal upon reaching a clinically relevant threshold in complex medium. Encouragingly, such ability was achieved recently with a modular signal processing strategy (digitalization and amplification) allowing glucose detection at selected thresholds in clinical urine samples from diabetic patients[32]. In addition to the proof-of-principle examples described in this work, it is interesting to note that the detection of parathion was achieved with the phosphotriesterase variant PTE-S5[58] that is known to have an important substrate promiscuity against several organophosphorous compounds[59]. Among them are chemical warfare agents such as cyclosarin that could also be detected through the same strategy. Transformation of cyclosarin by PTE-S5 produces cyclohexanol, a compound being subsequently transformed by *Acinetobacter* sp.'s ChnA enzyme into cyclohexanone, for which a ChnR-based biosensor have already been characterized for metabolic engineering applications[60]. While cyclosarin counts among the most toxic substances ever created, its toxicity comes from the inhibition of the enzyme

Acetylcholinesterase in the brain, therefore organisms such as *E. coli* can detect it without suffering the associated toxicity.

Intuitive limitations of SEMP are common with other intracellular biosensors such as the need for sufficient membrane permeability to extracellular targets and a limited toxicity of the targets for the chassis. As a potential option to minimize these issues, we expect SEMP to be easily transplanted to artificial cells that have been shown to serve as viable chassis for both biosensors and metabolic pathways[61,62]. Specific limitations of SEMP reside in the need for non-toxic metabolic intermediates and the risk of specificity issues. These biosensors are unable to discriminate between the presence of the target molecule in the medium and any intermediates of the SEMP. Careful consideration should thus be given to the risk of crosstalk if one of the intermediate is a possible contaminant in the envisioned biosensing application. Additionally, the potential ligand promiscuity of enzymes and transcription factors must be kept in mind if high specificity is required for a particular application. Potential cases of specificity issues linked to promiscuity are depicted in **Figure 14**. While detection of multiple targets could be advantageous in environmental applications, this would usually be problematic in medical applications. Directed evolution of a SEMP guided by its fluorescent output is an option to overcome false-positive activation. Lastly, changes of physicochemical properties between the target molecule and the intermediates of the pathway might result in unexpected properties. For instance, a decrease in the permeability coefficient of metabolites resulting from the transformations could lead to their accumulation inside cells. This could lower the detection limit of the sensor.

Aside from the synthetic biology tool aspect, one may wonder if SEMP represent a motif frequently used in nature and why. To our knowledge, this has not been investigated yet. Nevertheless, recent reviews suggest that the role of the metabolic layer in integrating information about the medium or the internal state of the cell is underestimated compared to the role of the genetic layer[63,64]. While the natural occurrence of a response to nitroglycerin seems fortuitous, recent evidences concerning the Lac operon show that evolution strongly co-selected the LacI transcription factor and the side-reaction site of beta-galactosidase that leads to production of allolactose[65]. The fact that LacI never evolved to directly detect lactose suggests that this conserved SEMP motif is advantageous to the homeostasis although the mechanism is not elucidated yet.

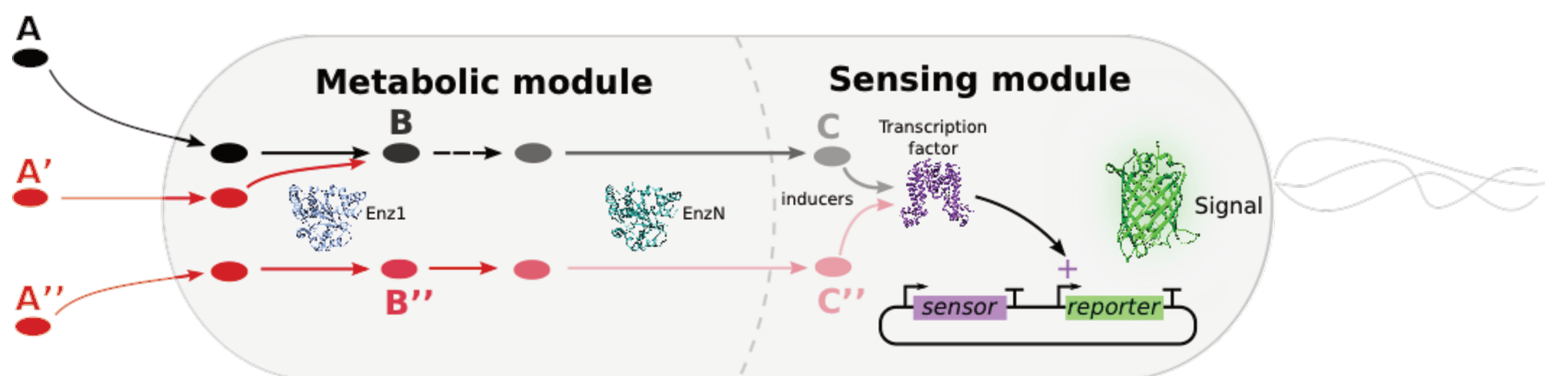


Figure 14: Summary of SEMP-associated specificity modes of failure. A specific limitation of SEMP resides in their inability to discriminate between the presence in the medium of the target molecule (A) or of any of the intermediates of the SEMP (B, C). In the context of whole cell biosensors, this can be circumvented by the use of control strains lacking enzymatic steps to allow discrimination. Another specificity issue might emerge from the fact that both transcription factors and enzymes can display high levels of ligands promiscuity. It is interesting to note that an important degree of promiscuity of several elements composing a SEMP (enzymes, transcription factors) does not necessarily translate into poor sensor discrimination abilities. Discrimination issues only emerge either if the first step of the pathway can transform several input compounds into the same output metabolite transmitted to the downstream steps of the SEMP (A') or if structurally different output metabolites can also serve as effectors to the transcription factor (A'') (in this case both the first enzyme and the transcription factor must be promiscuous).

2.5 Methods

Data source.

We retrieved 9,319 distinct encoded reactions (reaction signature diameter 12, see below) and their associated compounds from BRENDA 2015.1[46], MetaCyc 18.5[47] and Rhea v61[48]. Targets were gathered from DrugBank 4.1[43] (all approved drugs and illicit drugs datasets), the Human Metabolomic DataBase[44] (all compounds linked to a human disease) and the Distributed Structure-Searchable Toxicity (DSSTox) Database Network[45] (all compounds). Finally, we retrieved 505 distinct transcription factor effectors from BioNemo[51], RegTransBase[52], RegulonDB[53] and RegPrecise[54].

Molecular and reaction signatures.

Molecular signatures[66] (MS) are graph-based descriptors that encode the “neighborhood” of each atom of a molecule, similarly to Morgan's or ECFP fingerprints. Each kind of “neighborhood”, or atom environment, is a feature of MS. A reaction's signature (RS) is computed by subtracting the MS of the substrates to the MS of the products of the reaction[49], and takes the general form :

$$^d\sigma(R_n) = \sum_i ^d\sigma(P_i) - \sum_j ^d\sigma(S_j)$$

where $^d\sigma(P_i)$ and $^d\sigma(S_j)$ are the molecular signatures of substrate S_j and product P_i at diameter d . The variable diameter (d) of a molecular signature determines the size of the scope of the neighborhood that will be described around each atom. When this diameter is set to a small value, structurally similar molecules will share a lot of atoms with identical neighborhoods. This will allow a reaction signature to be applied to substrates that are close but different to the ones originally described in biochemical reaction databases. While small diameters allow to simulate promiscuous activities of enzymes and predict more products, it arbitrarily assumes a certain level of promiscuity from enzymes and thus lead to the generation of a higher number of incorrect predictions.. In this work we used molecular and reaction signature at a large diameter ($d=12$) to encode compounds and biochemical reactions, we thus consider enzymes largely non-promiscuous (unless several distinct reactions are reported in the databases for the same enzyme). See Carbonell et al.[42] for detailed informations about metabolic pathway design at lower diameter ($d<12$).

Compounds and reactions preprocessing.

The reactions were filtered in order to gather only biochemical reactions with a structure available for all involved compounds. The compounds were filtered using ChemAxon's Checker and Standardizer tools (JChem v.15.4.27, 2015). We let aside compounds with R-groups and the associated reactions. We performed the necessary treatments to standardize the compounds such that their molecular signature would be comparable. This involved an aromatization step and the removal of explicit hydrogens. The resulting compounds and reactions were processed to generate molecular (MS) and reaction signatures (RS). In the end, we gathered 9,319 unique RS involving more than 18000 unique MS. Target compounds were pre-treated the same way and were encoded as MS.

SEMP prediction.

Our previous work focused on the development of a synthetic pathway retrosynthesis algorithm named Retropath[42]. We built further on this basis by developing a Python pipeline adapted to predict SEMP. In our implementation, each target compound is successively used in-place of each substrate of each reaction R. If the resulting putative reaction R' has the same reaction signature as R, then we accept R' as a pathway step. In order to extend the pathway, the products of R' are then considered themselves as targets. We generated pathways with up to three steps. The result is a graph where the nodes represent compounds, and edges represent reactions. Compounds are then matched to the list transcription factor effectors compounds (Jaccard-Tanimoto coefficient[67] over 0.99). Finally, the sensing pathways are extracted from the graph with NetworkX graph library v.1.11[68]. Frequently a target can be sensed through several SEMPs leading to different sensible compounds, with different pathway lengths. For the analysis of the predictions, detectable targets were counted only once even if several SEMP were predicted.

Chemicals and reagents.

Benzoic acid, cocaine hydrochloride, hippuric acid, 2-chloro-4-nitrophenol, chlorohydroquinone, parathion-ethyl and 4-nitrophenol were purchased from Sigma (St. Louis, MO, USA). Permission to purchase cocaine hydrochloride was given by the French drug regulatory agency (Agence Nationale de Sécurité du Médicament et des Produits de Santé) to allow development of a new biosensor. Nitroglycerin was purchased from AccuStandard Europe (Niederbipp, Switzerland). Enzymes for cloning procedures (BsaI and DNA polymerase Q5) were purchased from New England BioLabs (Evry, France) and primers were purchased from Eurofins Genomics (Ebersberg, Germany).

Genetic constructs.

Two custom plamids were assembled to serve as vector for each module. Metabolic module vector is based on BioBrick standard vector pSB4T5 with pSC101 origin of replication and Tetracycline resistance marker, modified to harbor (i) type IIs restriction sites BsaI flanking cloning site and (ii) LacI transcription factor. Sensing module vector is based on BioBrick standard vector pSB1K3 with pMB1 origin of replication and Kanamycin resistance marker, modified to harbor (i) AraC transcription factor, (ii) pBAD promoter, (iii) type IIs restriction sites BsaI and (vi) a promoterless RFP. Coding sequences of genes BenR, CocE, HipO, were codon-optimised for *E. coli* with Jcat[69]

and natural BsaI sites were removed. These genes were synthesised by Genscript (Piscataway, NJ, USA). Coding sequence of PTE-S5, LinR, pnpA, pnpB and DmpR were extracted by PCR from plasmids pMaIc2x-PTE-S5[58], pMEU2R[70], pET-pnpA, pET-pnpB[71], and bba_k1413001 from the registry of standard biological parts (<http://parts.igem.org>), kind gifts from Prof Dan TAWFIK, Prof. Yuji NAGATA, Prof Ning-Yi ZHOU, and iGEM team EVRY 2014. Concerning sensing modules, LinR, BenR and DmpR sequences were followed by transcription terminator BBa_0015 from the registry of standard biological parts, followed respectively with sequence of the LinE gene promoter (i.e. the 153 base pairs in front of LinE CDS on the *Sphingomonas paucimobilis* UT26 chromosome), pBEN promoter (i.e. the 150 base pairs in front of BenA CDS on *Pseudomonas putida* KT2440 chromosome) or sequence of the Pu promoter (i.e. the 189 base pairs in front of DmpK CDS on *Pseudomonas sp.* CF600 plasmid pVI150). Concerning metabolic modules, T7 promoter with LacO operator and a ribosome binding site were placed in front of CocE, HipO, PTE-S5 and pnpA-pnpB CDS. All genes were amplified by PCR to add appropriate BsaI restriction sites and they were inserted in either metabolic or sensing module vectors using one-step GoldenGate assembly method[72]. In this setup the one-step assembly places the heterologous transcription factor under control of pBAD promoter and the promoterless RFP is placed under control of the heterologous promoter (see **Figure 15** for a detailed map of metabolic and sensing modules). Annotated sequences for all constructs were deposited on GenBank (accession numbers KU746628, KU746629, KU746630, KU746631, KU746632, KU746633, KU746634, KU746635 and KU746636) and are available online in Supplementary Data S2. Cloning of individual modules was done in DH5alpha and dose-response characterizations were carried out in BL21(DE3) after transformation with either sensing module only or both modules.

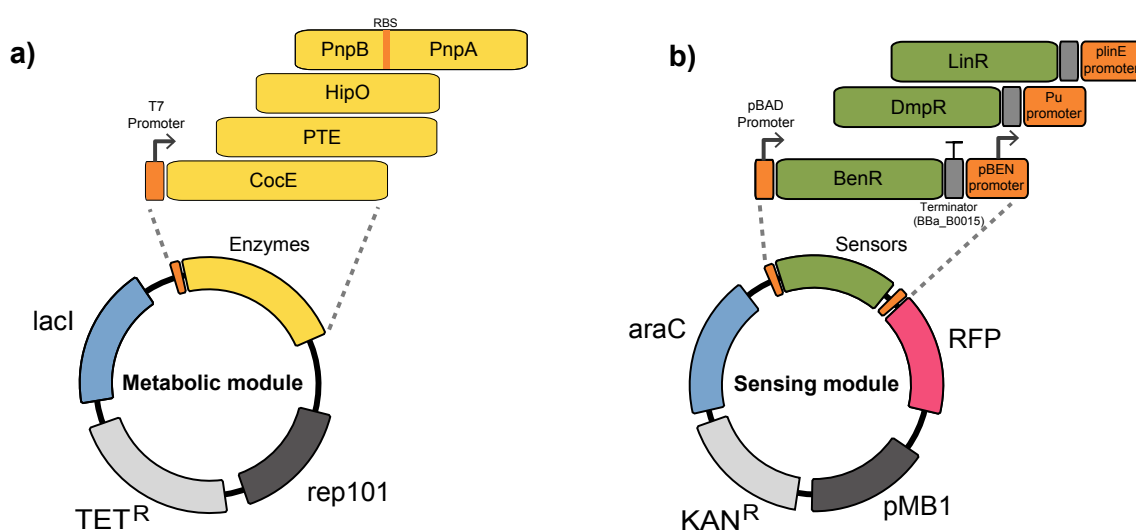


Figure 15: Architecture of metabolic and sensing modules. (a) Map of the metabolic module harboring either CocE, PTE-S5, pnpB-pnpA or HipO coding sequences. (b) Map of the sensing module harboring either BenR-pBen, DmpR-Pu or LinR-plinE sequences to control RFP. Transcriptional isolation is ensured by BBa_B0015, a double terminator from the registry of standard biological parts.

Biosensor dose-response characterization.

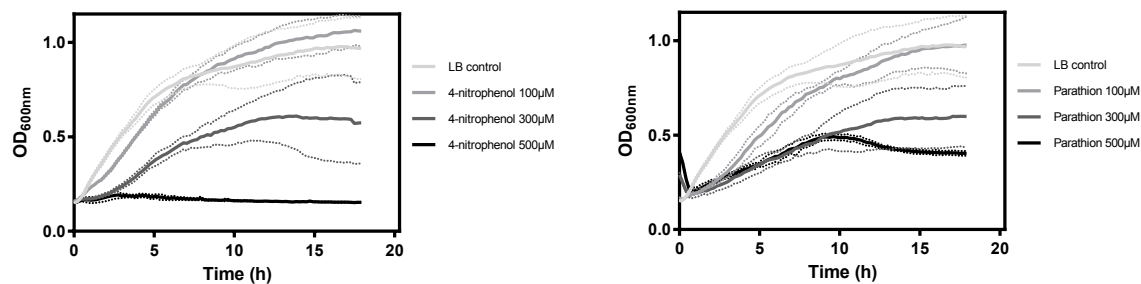
For each biosensor strain, an isolated colony of BL21(DE3) harboring the appropriate plasmid(s) was inoculated in 2mL of selective LB and grown overnight at 37°C. The overnight culture was diluted 1:100 in fresh selective LB and grown for 90 minutes under agitation at 37°C. Cells were then induced with IPTG 1mM and arabinose and further grown until OD600 reached 0.1 for parathion, 4-nitrophenol, benzoic acid, hippuric acid and cocaine sensors. For 2C4NP and chlorohydroquinone, sensor cells were grown until OD600 reached 0.4 to minimize artefacts on the signal due to the high toxicity of 2C4NP. In our setup, optimal induction levels of transcription factors were found to be obtained with arabinose levels of 0.001% (BenR) 0.1% (LinR) and none (DmpR). Candidate chemicals for biosensing were dissolved in ethanol and 2μL of different concentrations were mixed with samples of 200μL of cell culture. 2μL of pure ethanol was mixed with negative control cell samples. Cocaine sensor cells and associated control cells lacking metabolic module were grown in Eppendorf 1.5mL microtubes with vigorous agitation for 18 hours at 30°C, thus following optimal conditions described in the literature for CocE activity. All other sensor cells were grown for 18 hours with agitation at 37°C in microplate reader TECAN Infinite 500. Absorbance at OD600 and fluorescence (Exc: 580nm / Em: 610nm) was measured with microplate

reader TECAN Infinite 500. All experiments were repeated at least 3 times on different days with similar results.

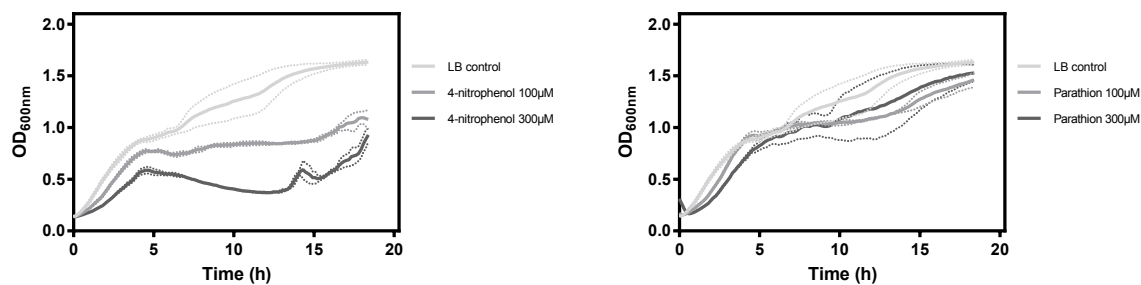
Characterization of the transcriptional response to nitroglycerin of different *E. coli* strains.

pYeaR promoter (i.e. 146 base pairs in front of YeaR CDS on *E. coli* MG1655 chromosome) was amplified by PCR from purified chromosome and fused through GoldenGate assembly with a promoterless RFP into a pACYC plasmid backbone. The resulting plasmid pYeaR-pACYC was transformed into BL21(DE3) or JW1642 from the Keio collection[73]. The coding sequence of NemaA was extracted by PCR from a BL21(DE3) strain and cloned under the control of constitutive promoter J23100 into a pCDF plasmid backbone. A strain of JW1642 was cotransformed with plasmids pYeaR-pACYC and NemaA-pCDF to investigate if it would restore the response to nitroglycerin. Cells were grown at 37°C in selective M9 minimal medium with 0.5% glucose. Overnight cultures of each strain were diluted 1:100 and grown until OD600 reached 0.1. 2µL of nitroglycerin diluted in ethanol was added to 200µL of cell samples to reach a final concentration of 50µM. 2µL of pure ethanol was added to control samples. Cells were grown for 18 hours with agitation at 37°C and absorbance at OD600 and fluorescence (Exc: 580nm / Em: 610nm) was measured in microplate reader TECAN Infinite 500.

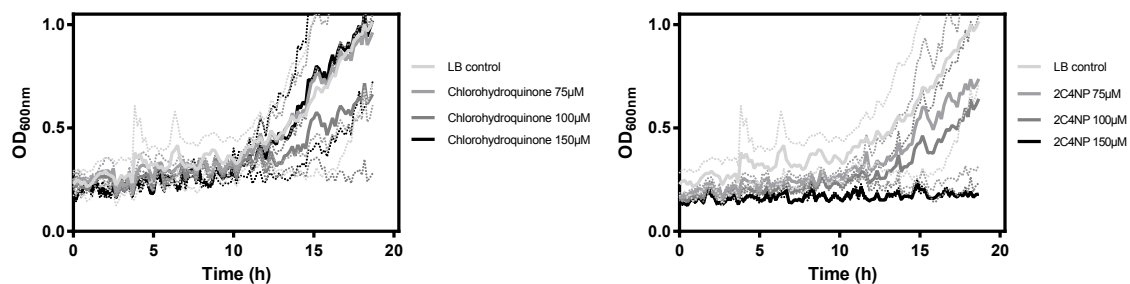
a) BL21: PTE-S5 metabolic module + 4-nitrophenol sensing module



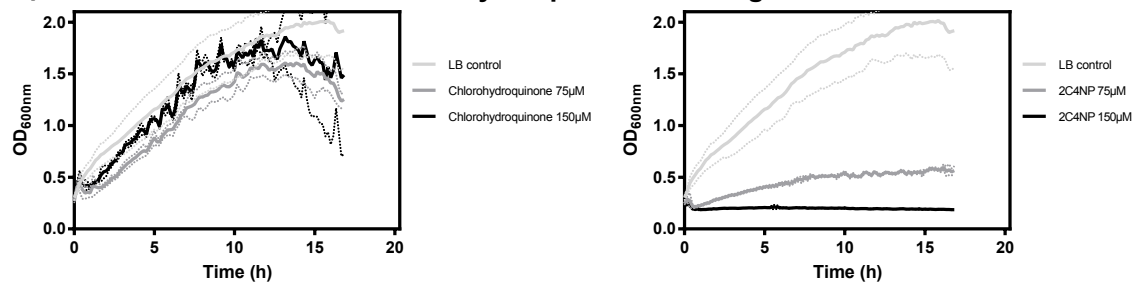
b) BL21: 4-nitrophenol sensing module



c) BL21: pnpB-pnpA metabolic module + chlorohydroquinone sensing module



d) BL21: chlorohydroquinone sensing module



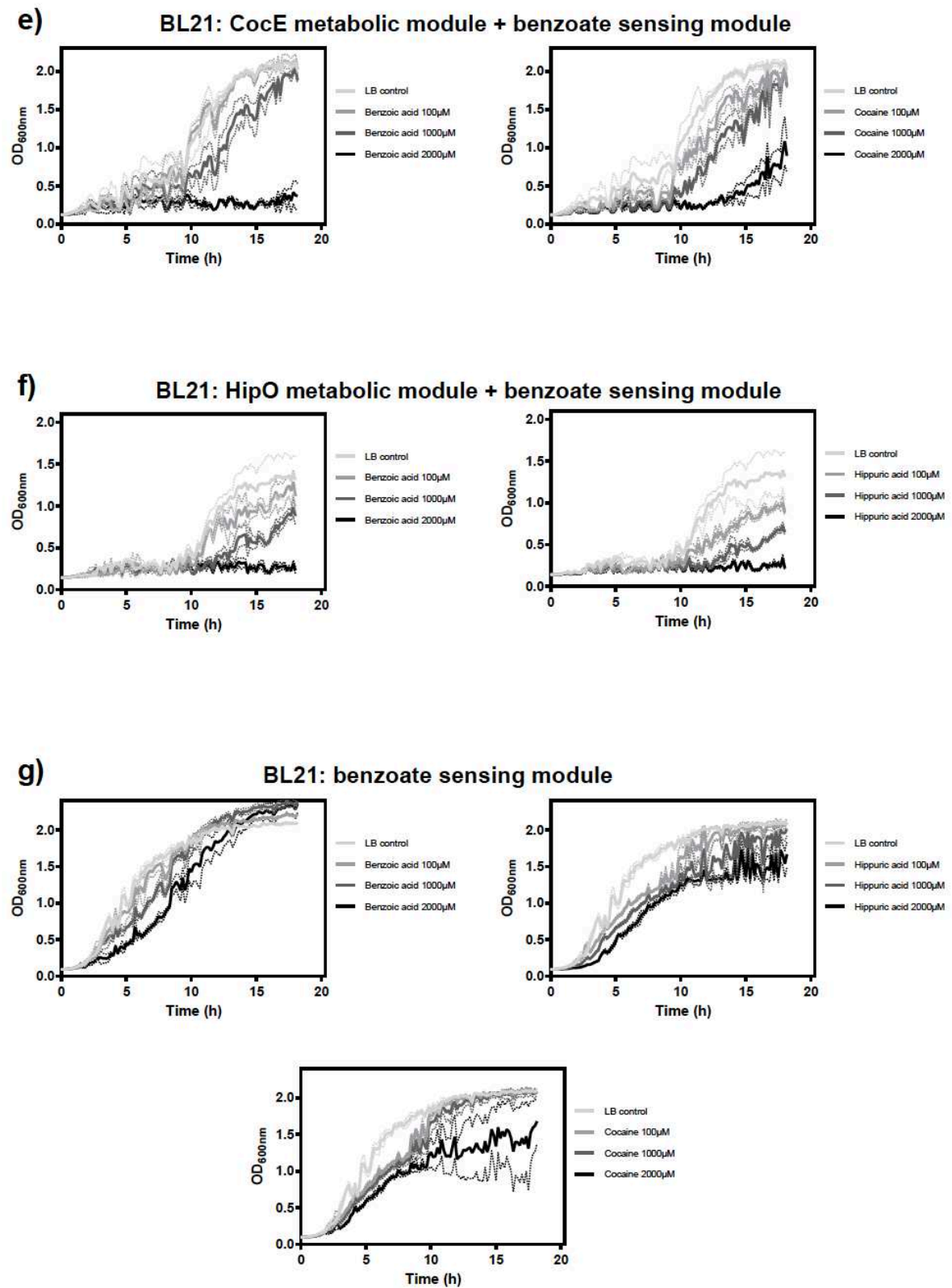
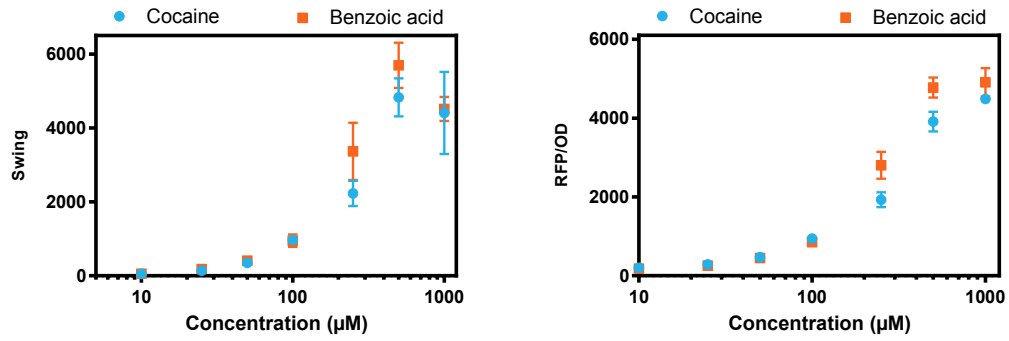
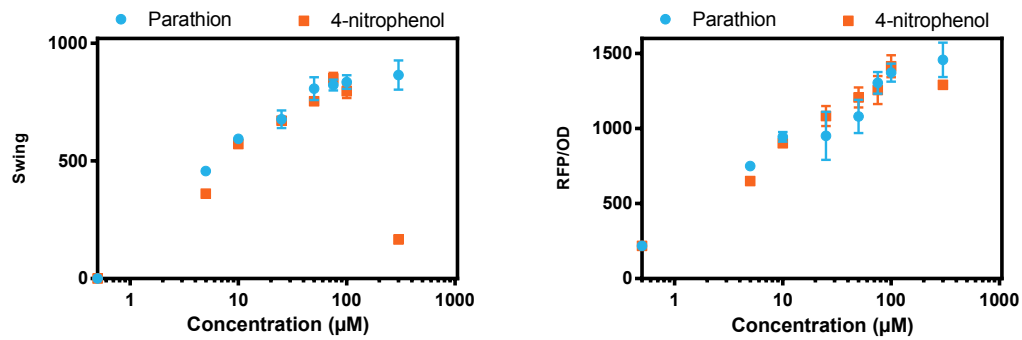


Figure 16. Growth curves of the different strains in presence of the associated compounds

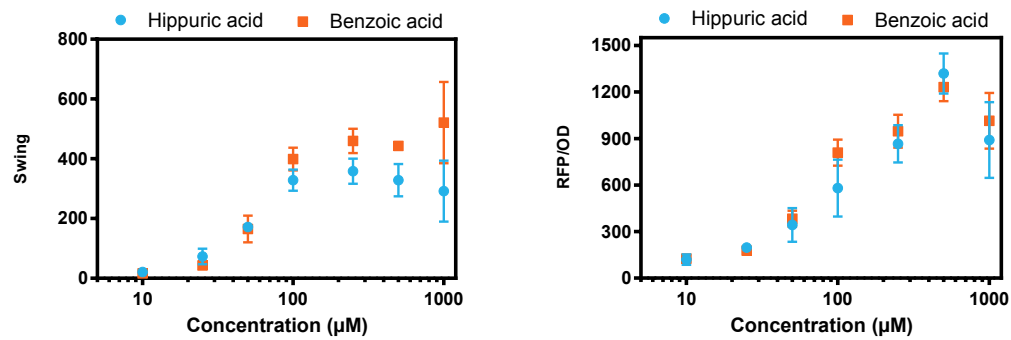
BL21: CocE metabolic module + benzoate sensing module



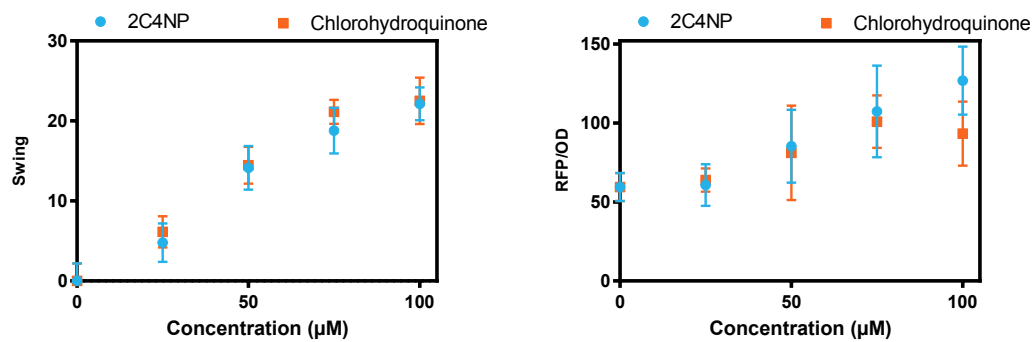
BL21: PTE-S5 metabolic module + 4-nitrophenol sensing module



BL21: HipO metabolic module + benzoate sensing module



BL21: pnpB-pnpA metabolic module + chlorohydroquinone sensing module



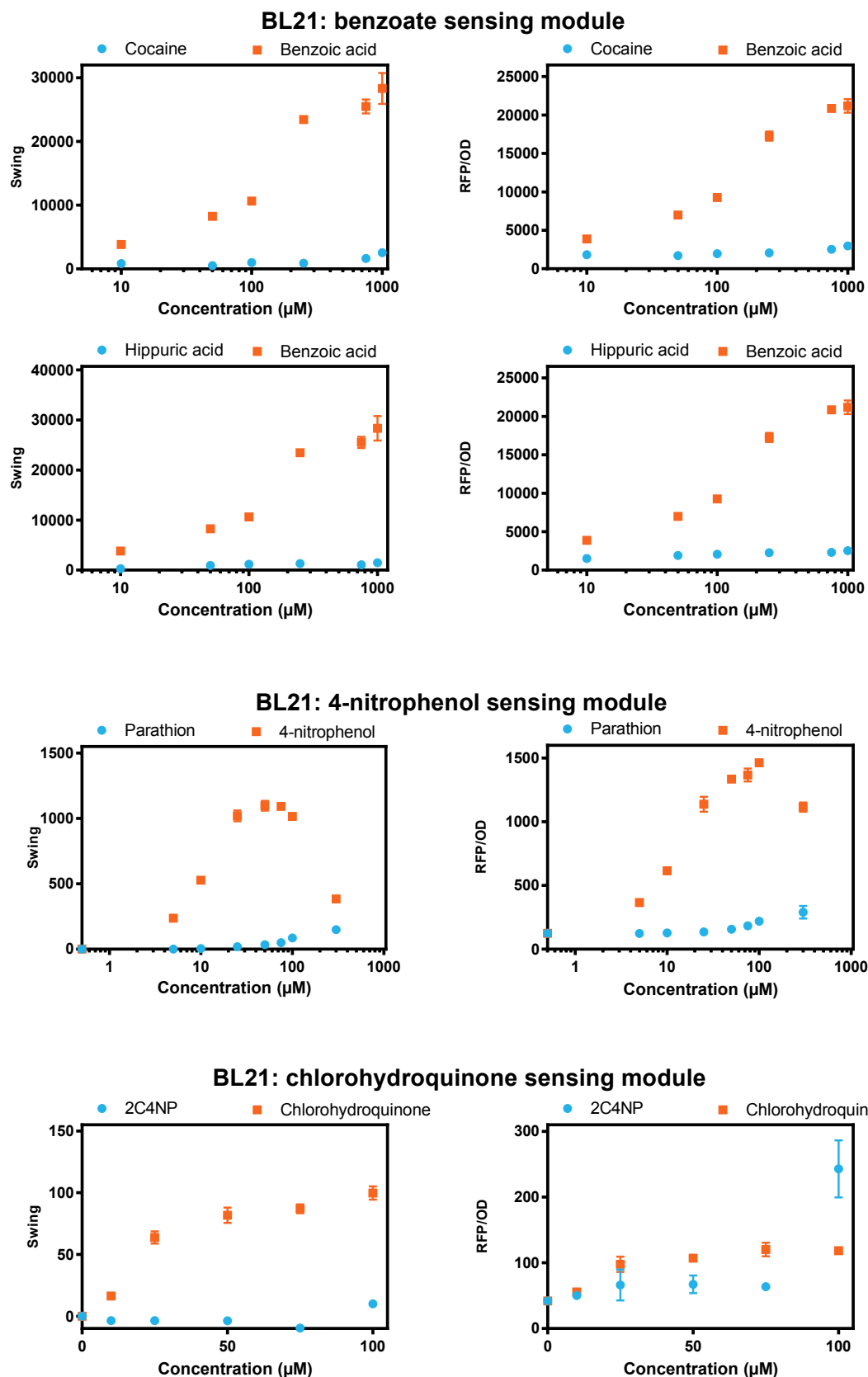


Figure 17. Swing and RFP/OD of all strains with every compounds. Swing is defined as the difference in raw fluorescence observed between induced cells and non-induced control population.

Acknowledgments

We thank Dan Tawfik, Moshe Goldsmith, Yuji Nagata and Ning-Yi Zhou, for kindly sharing their genetic constructs; Pavel Novichkov and Dmitry Rodionov for providing effector data from RegPrecise ; the INSERM U1001 lab for providing strain JW1642 ; Edwin Wintermute, Marcel Salanoubat, Matthew Deyell and Aude Bernheim for fruitful discussions.

Chapter 3 : Opinion on sensing new chemicals with bacterial transcription factors

This work has been published in Current Opinion in Microbiology in 2016 by Vincent Libis, Baudoin Delépine* and Jean-Loup Faulon. (*contributed equally)*

Only very minor modifications have been made between the article and the chapter presented below.

3.1 Abstract

Bacteria rely on allosteric transcription factors (aTFs) to sense a wide range of chemicals. The variety of effectors has contributed in making aTFs the most used input system in synthetic biological circuits. Considering their enabling role in biotechnology, an important question concerns the size of the chemical space that can potentially be detected by these biosensors. From digging into the ever changing repertoire of natural regulatory circuits, to advances in aTF engineering, we review here different strategies that are pushing the boundaries of this chemical space. We also review natural and synthetic cases of indirect sensing, where aTFs work in combination with metabolism to enable detection of new molecules.

Highlights:

- transcription factors represent the starting material of choice to match the massive need of biosensors for new chemicals in biotechnological applications.
- combination of evolution and rational design are making advances toward tailor-made transcription factors
- metabolism indirectly expands the scope of chemicals that can induce natural and synthetic circuits

3.2 Introduction

Allosteric transcription factors (aTFs) responding to chemicals have an intimate link with the history of molecular biology since the unravelling of the lac operon. The chemical diversity that they can sense is now considered critical for the development of a wide range of biotech applications. In addition to being central in the standing efforts to develop biosensors monitoring environmental pollutants [74], aTFs are now enabling synthetic biology applications in the medical field as well [75] [76]. They are also becoming critical in metabolic engineering where they enable high-throughput screening of strains producing added-value chemicals [77] [78] and optimization of metabolic pathways through dynamic regulation of the expression of key enzymes [79]. Advances in these fields should accelerate as synthetic circuits are becoming more and more predictable, modular and scalable [6] but the limited diversity of chemical inputs is a bottleneck in the development of applications. The list of well-characterized aTFs is regularly increasing to tackle this issue [33] but the needs are not met yet. While synthetic biologists dream about tailor-made aTFs, such ability will require a deep understanding of how they work and evolve in nature.

Here we showcase recent approaches aiming at broadening the scope of chemicals that can induce aTFs. Several other reviews are available for more details concerning the applications of aTFs in biotechnology [80] or comparison of aTFs to other types of sensors [18].

3.3 Mining allosteric transcription factors in nature

In nature, aTFs evolved to recognize a multitude of chemicals such as antibiotics, primary or secondary metabolites, quorum sensing molecules, or man-made pollutants. Decades of effortful individual characterization regularly expanded the known frontiers of the chemical space within reach of these proteins. This knowledge is now organized and regularly updated in several databases where couples of aTF-effectors can be found [51,52,54,81]. BioNemo for instance compiles almost a hundred xenobiotic-responding aTFs that are the fruits of a very recent evolution as most of these molecules have appeared in the environment since the industrial chemistry era [51]. Many lessons about

aTF families, mechanisms, and evolutionary potential have been learned on the way, thoroughly reviewed by De Lorenzo et al. [82] in the case of xenobiotic sensing.

Most of the natural sensing abilities remain probably uncharted as effectors are unknown for the majority of sequenced aTFs. A promising strategy consists in using comparative genomics to identify the binding sites of an aTF on the genome and thus predict its target genes which allow reconstruction of the regulon [83]. By assessing the function of the regulated metabolic pathway, known metabolites can be proposed as potential effectors for the aTF. This kind of approach has been demonstrated both with members of the LacI family [83] and of the ROK family [84].

With some aTFs, for instance those belonging to the large TetR-family, genetic context alone can be sufficient to make educated guesses about the effectors. Most aTFs from this family have their genes divergently oriented and less than 200 base pair away from the genes that they regulate [85]. Recently, predictions drawn from this feature have been coupled with phylogenomics to propose candidate effectors that were successfully validated experimentally. After building a relational tree organizing thousands of sequences from TetR-family members in subfamilies, Cuthbertson et al. [86] incorporated all known effectors on the tree and observed that similar effectors would cluster with the same subfamilies. Given the important diversity of effectors recognized by TetR-family members, the predictive aspect of such relational tree could be a powerful tool to identify new couples of aTF-effectors from the hundreds of thousands of TetR sequences available in databases.

A more direct way to dig into natural diversity is to use a screening method for substrate-induced gene expression (SIGEX). SIGEX is based on systematic cloning of fragments from metagenomic libraries into an operon-trap vector containing a promoter-less fluorescent protein. Fragments containing an aTF and associated promoter can be isolated by fluorescence-based screening in presence of potential effectors. This approach regularly yields new characterized aTFs by screening environmental DNA extracted from contaminated areas [87]. Because the chemical space recognized by nature is rapidly evolving at the contact of the human civilization, these approaches tapping directly into this rich diversity will probably continue to be a first line solution to the lack of biosensors.

3.4 Engineering the sensing scope

The regulatory function of aTFs makes them suitable for high-throughput screening protocols aiming to isolate variants induced by new chemicals. Importantly, aTF redesign experiments necessitate a dual screening strategy to avoid selecting inactivated or non-inducible mutants (see **Figure 18A**). When structures of the aTF are available (ideally co-crystalized with effectors) to guide the process, such screenings coupled with mutagenesis methods can be sufficient to isolate aTF able to recognize new chemicals [88–90].

While aTFs are composed of modules such as ligand-binding domain (LBD), DNA-binding domain (DBD) and linker domain, an important difficulty to engineer them resides in their interdependency to ensure allostery of the protein [91]. Changes in conformation induced by an effector must propagate successfully in the protein to change its affinity to DNA and this propagation can easily be disrupted by mutations. Counterintuitive lessons have been learned, for instance through the observation that mutations outside the LBD could lead to important changes in effector specificity. Nonetheless, allostery can be preserved through careful engineering guided by expert knowledge. For example, shuffling LBD and DBD from members of the same family have been demonstrated with the LacI/GalR family [92,93]. This approach could yield more functional chimeras since a new database containing structural and mutagenesis data is now available to help predicting allosteric positions in this family [94]. Hopefully, the knowledge gained about maintenance of allostery in such chimeras will help engineering LBD recognizing new chemicals in a modular way.

However, in most cases, the wiring diagram of allostery is not known and an aTF must be considered as a fragile whole. Consequently, mutagenesis efforts should not be targeted at subdomains but ideally cover the whole sequence. This put a major limit on the chances of fruitful explorations since mutagenesis campaigns are bound by the transformation efficiency of bacteria (10^8). In this context, computational design tools can be used to reduce the design space. By helping the identification of amino acids playing roles in binding, specificity or allostery, these tools can define priority targets on which mutagenesis efforts should be focused. Several recent successes suggest that these approaches outperform random mutagenesis at altering effector specificity.

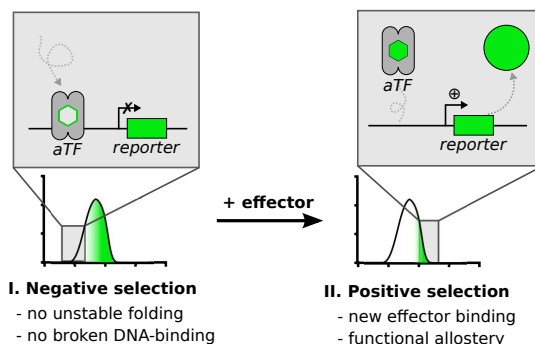
For instance, despite lacking an experimentally determined structure, Jha et al. [95] used structural bioinformatics tools to model the aTF PobR, and performed docking experiments that identified 16 important residues in the predicted LBD. The authors then used a “natural drift” heuristic in which they restricted themselves to one mutation by identified codon, and used the BLOSUM62 substitution matrix to keep only the most naturally observed substitutions. After flow cytometry screening, they found several hits with gain of function for induction by their target chemical. These variants had around ten mutations compared to the native sequence, covering most of the predicted positions.

Taylor *et al.* [35] proposed an end-to-end workflow for redesigning aTF specificity and compared several strategy to explore variants space, including computational design, random and targeted mutagenesis. They notably took advantage of a full-length single amino-acid substitution scanning of the LacI aTF to discover positions away from the LBD involved in allostery and specificity. Protein design approaches leveraging such mutational scanning data are considered to have an important potential [91,96,97]. The authors challenged their computational design abilities by developing biosensors for four chemicals with increasingly dissimilar structures from the natural LacI effectors. Interestingly, they stressed the importance of a last step of “activity maturation” consisting of making chimeras of good variants in terms of induction with good variants in terms of narrow specificity. This strategy allowed them to lose the natural induction by IPTG while maintaining the newly developed induction by their target chemicals.

An ambitious approach consists in performing part of the screening *in silico*. This is a daunting task and expert knowledge of tools initially developed in other areas such as drug design is necessary. One difficulty is that straightforward docking of effectors on *in silico* variants to predict sequences with increased affinity is usually futile. Indeed, affinity doesn’t necessarily correlate with induction abilities, and lower affinity can translate in higher induction abilities. Despite these difficulties, De Los Santos et al. [98] recently used several co-crystallized structures of the well characterized aTF QacR as starting point and performed an *in silico* screening to modify QacR specificity to recognize vanillin. They tested the best predictions with a cell-free screening protocol and obtained a few positive hits. Remarkably, the structure of vanillin is completely different from the known effectors of QacR (see **Figure 18B**). Although the cell-free *in vitro* screening selected aTFs variants that turned out to be toxic *in vivo*, this framework

could become an important source of new biosensors in the coming years if it proves adaptable to other aTFs and chemicals.

A



B

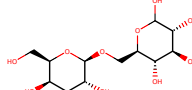
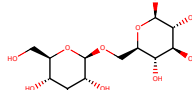
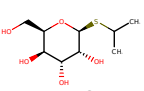
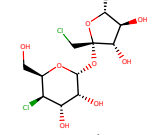
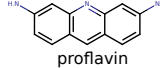
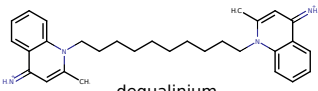
aTF	Previously known effectors	Adopted strategy	Target effector	Reference
AraC	 L-arabinose	• Site-directed saturation mutagenesis	 mevalonate	[21]
			 triacetic acid lactone	[22]
			 ectoine	[23]
PobR (IcIR family)	 4-dihydroxybenzoate	• Computational design (Rosetta), • then "natural drift".	 3-4-dihydroxybenzoate	[28]
LacI	 allolactose	Mixed: • computational design (Rosetta), • protein-wide single-amino-acid saturation mutagenesis, • error-prone PCR, • and activity maturation.	 gentiobiose	[29]
	 IPTG		 sucralose	
QacR (TetR-family)	 proflavin	• <i>In Silico</i> sequence generation, • then cell-free prototyping.	 fucose	[32]
	 dequalinium		 vanillin	

Figure 18. A) Schema of the dual selection employed while screening allosteric transcription factor (aTF) variants. Top panels represent targeted molecular state in presence or absence of the

investigated effector. The graphs represent the distribution of fluorescence intensity among a population of variants with the gates that must be used to select the cells exhibiting the appropriate response. The first step of selection (left) aims at eliminating aTF that lost the ability to repress their cognate promoter. Instead of fluorescence-based cell sorting, this step can also be replaced by a selection based on the ability to repress expression of a toxic protein [35]. The second step (right) aims at selecting aTF induced in presence of the effector. **B)** Recent examples and chemical structures of new effectors yielded by various aTF redesign strategies.

3.5 Regulation via indirect sensing

Both synthetic biologists and natural organisms can take advantage of indirect sensing through metabolism to expand the chemical space detectable by aTF. Chemicals that can be transformed *in situ* by enzymes into aTF effectors are able to induce a response at the genetic level (see **Figure 19**). Therefore, filling a need for a biosensor toward a particular chemical can sometimes be done more simply by chemically modifying the target molecule than by engineering the specificity of aTF. Thanks to the immense reservoir of known biochemical reactions, many sensing-enabling metabolic pathways (SEMP) can be designed to transform nondetectable molecules into detectable ones. Engineered bacteria harboring such proper combination of enzymes and transcription factors have seen their sensing abilities extended to previously nondetectable molecules such as lindane [99], salicylaldehyde [37] or cocaine [100] to name a few. In the context of metabolic engineering, this strategy has recently enabled the monitoring of 3-hydroxypropionate (3HP) to speed up the development of a biobased production process. Because of the lack of aTF able to sense 3HP, Rogers and Church [57] developed two alternative SEMP connected to PrpR or AcuR that allowed combinatorial testing of fermentation conditions (see **Figure 19B**). Systematic design of SEMP has been achieved by leveraging metabolic pathway databases, databases of known aTF effectors and retrosynthetic pathway design tools, leading to the development of 5 new whole cell biosensors [100]. A webserver is now available to the community to predict SEMP for chemicals on demand [101].

In nature, a growing number of SEMP-like motifs have been recently reported (see **Table 2**). The reasons why indirect sensing is selected by evolution in some cases are not understood yet. The lac operon for instance is induced indirectly in presence of lactose only after its transformation into allolactose by a secondary activity of β -galactosidase (see **Figure 19C**). This secondary activity and the LacI aTF have been

shown to be strictly co-selected [65] thus raising the question of the advantage of this SEMP motif over an evolution of LacI toward direct sensing of lactose. Another example of regulation motif with occurrence among several bacteria able to detect aromatics consists in using aromatic CoA-thioesters as effectors. In the case of 3-methylbenzoate sensing by *Azoarcus* sp. this SEMP has been well characterized and shine some lights on the type of effects these motifs can have on the genetic response profile. Juárez et al. [102] observed that the CoA-ligase producing the *bona fide* effector molecule is under the control of two promoters, one controlling the whole multi-enzyme degradation operon and a second one which is nested in the operon to add a second layer of control specifically on this enzyme. Both promoters are equally essential for growth in 3-methylbenzoate and they are both regulated by the same aTF, MbdR, but their induction kinetics and dynamic range appear different. When put in presence of 3-methylbenzoate, the expression of the CoA-ligase is first induced moderately without the rest of the operon. This leads to an increase of the 3-methylbenzoyl-CoA pool in a positive feedback fashion, ultimately triggering the expression of the whole degradation operon through induction of its main promoter (see **Figure 19D**). Positive feedbacks are usually associated with interesting effects in response to signal and such exchanges between the metabolic and the genetic layers support the growing view of metabolism taking part in the computations performed by cells [63]. How much SEMP motifs actually play a role in computing signals is unclear. The reflexion could be fueled by studies demonstrating that rich biological behaviors such as linearization of dose-response can emerge at the metabolic level without requiring genetic circuits [103]. Understanding the effects of SEMP motifs could even shed light on regulatory circuits essential in pathogenesis of bacteria such as *Mycobacterium tuberculosis* where the sensing of cholesterol is also mediated by several enzymatic steps [104,105].

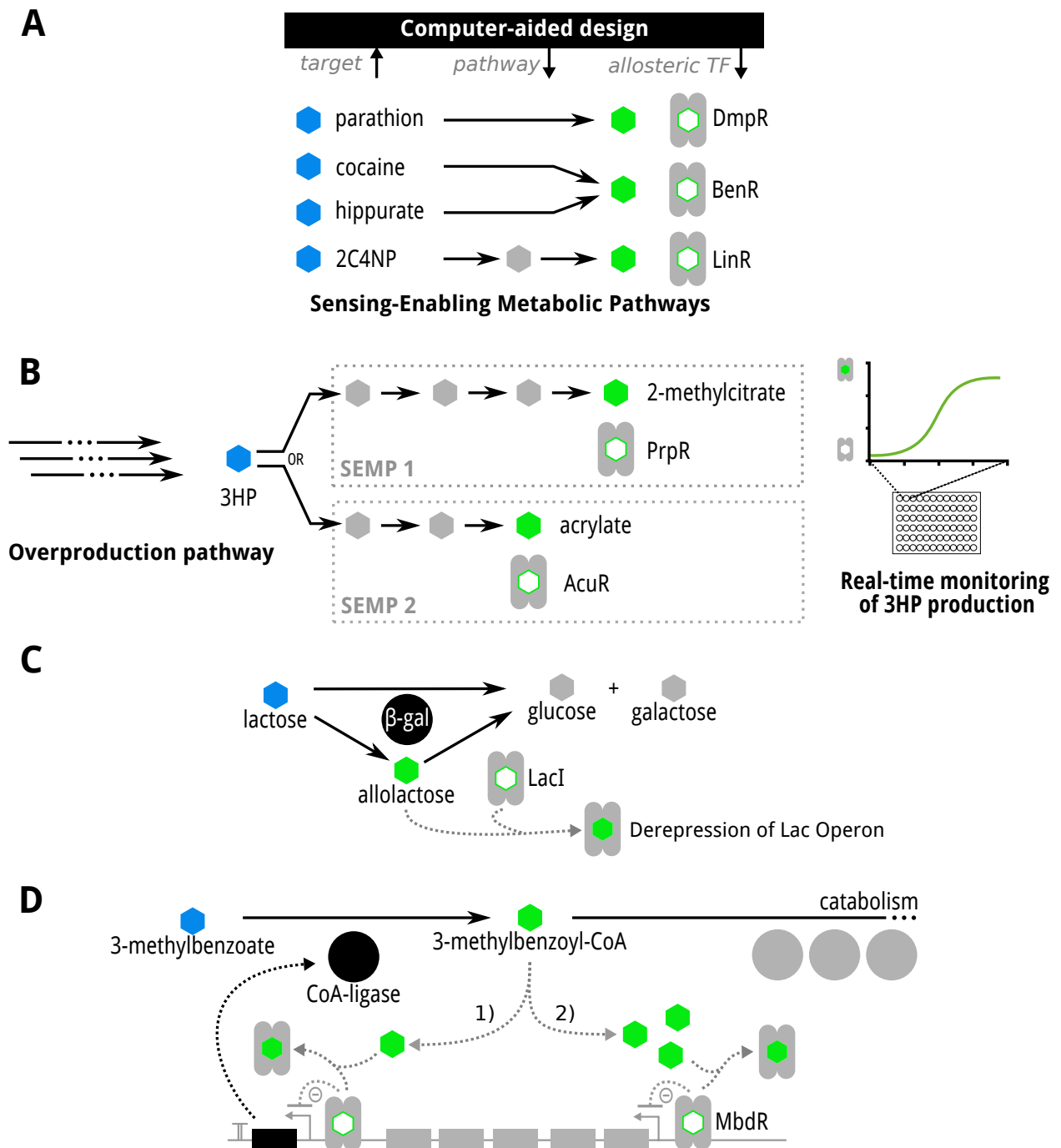


Figure 19. Examples of synthetic and natural sensing-enabling metabolic pathways (SEMP). A) Computer-aided design can leverage biochemical reaction databases to construct synthetic routes that transform a chemical of choice into a known allosteric transcription factor (aTF) effector. B) In addition to the enzymes required for production of 3-hydroxypropionate (3HP), enzymes able to transform 3HP into detectable molecules were implemented in *E. coli*. These two alternative SEMP allowed monitoring of 3HP via two different aTF. SEMP 2 exhibited the best response to 3HP and was therefore used to help the high-throughput optimization of 3HP bioproduction process. C) The lac operon necessitate the secondary activity of β -galactosidase to transform lactose into the natural effector of LacI, allolactose. D) The specific regulation of the anaerobic degradation pathway of 3-methylbenzoate in *Azoarcus* sp.. (1) The induction of the operon starts with derepression of the

nested promoter controlling specifically the CoA-ligase which in turn increase the pool of the effector 3-methylbenzoyl-CoA. (2) When the effector reach higher concentrations, the rest of the operon is transcribed and the catabolism is initiated.

Table 2 : examples of indirect sensing via metabolism

Host	Sensed compound	Intermediates	Regulator (organism)	Ref.
E. coli (natural)	lactose	allolactose	LacI	[106]
Thauera aromatica (natural)	toluene	benzylsuccinate	TutBC	[107]
Mycobacterium tuberculosis (natural)	cholesterol	cholest-4-en-3-one; 3-oxocholest-4-en-26-oyl-CoA	KstR	[104]
Mycobacterium smegmatis (natural)	cholesterol	cholest-4-en-3-one ; 3-oxo-4-cholestenoic acid	KstR	[105]
Paracoccus sp. (natural)	L-gluconate	L-5-ketogluconate ; D-idonate	LgnR	[108]
Azoarcus sp. (natural) E. coli (synthetic)	3-methylbenzoate	3-methylbenzoyl-CoA	MbdR	[102]
Sphingobium sp. (natural)	ferulate	feruloyl-CoA	FerR	[109]
Rhodopseudomonas palustris	p-Coumarate	p-Coumaroyl-CoA	CouR	[110]
Comamonas testosteroni (natural)	benzoic acid	benzoyl-CoA	GenR	[111]
Thermus thermophilus (natural)	Phenylacetic acid	phenylacetyl-CoA	PaaR	[112]
E. coli (synthetic)	salicylaldehyde	salicylate	NahR (Pseudomonas)	[37]
E. coli (synthetic)	lindane	1,2,4-trichlorobenzene	XylR (Pseudomonas putida)	[99]
E. coli (synthetic)	2-chloro-4-nitrophenol	chloro-1,4-benzoquinone ; chlorohydroquinone	LinR (Sphingomonas paucimobilis)	[100]
E. coli (synthetic)	cocaine	benzoic acid	BenR (Pseudomonas putida)	[100]
E. coli (synthetic)	parathion	4-nitrophenol	DmpR (Pseudomonas sp.)	[100]
E. coli (synthetic)	hippuric acid	benzoic acid	BenR (Pseudomonas putida)	[100]
E. coli (synthetic)	3-hydroxypropionate (3HP)	3HP-CoA ; acrylyl-CoA ; propionyl-CoA ; 2-methylcitrate	prpR (E. coli)	[57]
E. coli (synthetic)	3-hydroxypropionate (3HP)	3HP-CoA ; acrylyl-CoA ; acrylate	AcuR (Rhodobacter sphaeroides)	[57]

3.6 Conclusion

The methods described in this review represent promising progress toward “on demand” aTFs, however, they still require important expertise and efforts. Because the range of applications extends over many disciplines, one issue is that laboratories with such expertise in aTFs are often not the same as laboratories facing the needs to detect new chemicals and they do not necessarily know on which relevant target molecules to focus their efforts. An era of synthetic biology was therefore dominated by applications inspired by the availability of bioparts. A new paradigm is emerging where applications are instead inspired directly by problems but this will necessitate minimizing the needs for expert knowledge through standardized methods of biosensor development.

While different methods may lead to functional biosensors for the same chemical, the overall cost and efforts involved can vary considerably. It is therefore recommended to wisely choose the most appropriate method considering each new target chemical. If an aTF detecting a structurally similar molecule can be found in databases, evolution protocols are probably the fastest approach. In other cases, the most appropriate approach can be determined by considering the nature of the chemical, itself being usually related to the field of application. For example, environmental monitoring applications typically need biosensors for xenobiotics that are already in contact with microbial communities in the environment. Mining exposed metagenomes seems therefore more straightforward than engineering approaches. Because aTFs may evolve subsequently to enzymes, indirect sensing through metabolism is an alternative when environmental exposure is too recent for aTFs to be found. For instance, Product-Induced Gene EXpression (PIGEX) approaches using aTFs detecting likely metabolites from the target chemical can yield new enzymes from metagenomes and create SEMP biosensors at the same time [113–115].

A substantial part of metabolic engineering efforts is focused on bioproduction of natural products of secondary metabolisms. Biosensors enabling high-throughput screening of strains have chances to be immediately identified by mining microbiomes in contact with these products [116]. On the contrary metabolic engineering for xenobiotic production might require development of biosensor *de novo* through more sophisticated engineering approaches. Luckily the field of metabolic engineering is motivated by the promise of short term industrial applications, and is therefore becoming the biggest contributor of new biosensors. It is likely that approaches to

develop tailor-made biosensors will be further improved and standardized through this driving force and will in turn benefit to other fields.

Some fundamental knowledge may also be gained on the way, as in the case of indirect sensing via SEMP. Whether the role of SEMP is simply limited to sensing in nature, or whether metabolism enables some level of computation will require further investigation but synthetic biologists will certainly follow such discoveries if additional design principles can be unraveled.

Summary of important references:

Cuthbertson *et al.*[86]: Demonstration of the benefit of using predictions from genetic context coupled with phylogenomics to identify new couples of aTF-effectors. After unravelling one such couple, the author ended up characterizing a new antibiotic resistance mechanism.

Ravcheev *et al.*[83]: An example of the use of comparative genomic methods and in silico prediction of transcription factor binding sites to reconstruct regulons. Functional information of the genes in the regulon were used to propose effectors for allosteric transcription factors of the LacI family.

Juárez *et al.*[102]: An example of natural SEMP thoroughly characterized in its host and also reproduced in *E. coli*

Rogers *et al.*[57]: This metabolic engineering project took advantage of a SEMP approach to enable real-time monitoring of high value added product formation in *E. coli*.

Taylor *et al.*[35]: This project made use and compared libraries obtained through three different strategies to drive LacI specificity toward new chemicals. In addition to computational design with Rosetta, protein-wide single-amino-acid saturation mutagenesis and error-prone PCR, they used a final shuffling step to gather together the best mutations from all strategies.

Los Santos *et al.*[98]: Demonstration of the power of combining structural bioinformatic tools, expert knowledge and cell free transcription-translation assay to develop a biosensor. After screening only 27 variants, the authors identified two aTF exhibiting a specific response to vanillin.

Conclusion and perspectives – Part I:

Since the inception of this project in 2013, the needs for new chemical inputs have kept growing in the synthetic biology community. This partly results from the fact that signal processing is no longer regarded as the bottleneck to build real-world applications, and also because new uses of biosensors have been embraced by a larger community, especially for dynamic metabolic pathway regulation or screening of improved strains and pathways (see **Chapter 6 - Annex 2** for an example). Since the field of metabolic engineering currently benefits from important research efforts and fundings in both public and private sectors, the traction it exerce actively on biosensor development for new molecules is likely to last. Other potential fields of applications of biosensors such as medicine have been more passive to this point, rarely developing *ad hoc* biosensors for new molecules and mostly repurposing existing ones in an opportunistic manner. In both cases, facilitated development of tailormade biosensors through standardized approaches is likely to have an immediate impact accelerating, or enabling the development of numerous applications.

In this context, the demonstration of the potential of sensing-enabling metabolic pathways (SEMP) to expand the scope of detectable chemicals can inspire laboratories to investigate this strategy for their own needs. The method has proven quick to implement with 5 new biosensors, each successfully characterized in a matter of weeks, in comparison to year-long efforts necessary for protein engineering approaches for instance. The only specific expertise required for the method was initially knowledge about the programming of CAD tools for metabolic pathway design, but with the publication of our user-friendly Webserver this is no longer a requirement. In view of our simulations indicating an increase in potential chemical inputs by at least a factor two, we expect SEMP to enable multiple applications in the coming years. For instance, an organophosphate pesticide biosensor similar to the one described in **Chapter 2** have been further developed with a focus on improving the colorimetric output to provide an alternative to liquid or gas chromatography methods for on-site detection[117]. Our own laboratory is now investigating the possibility to detect several other biomarkers of diseases in body fluids. A key determinant in the long term impact of the SEMP strategy is its full synergy with other biosensor development methods. Whether riboswitch engineering finally deliver new inputs or *in silico* design approaches of transcription factors end up mastering some of the mechanisms of allostery, the reach of SEMP will keep increasing proportionnaly.

It would be interesting to continue to investigate natural occurrences and mechanisms of SEMP's with a systems biology angle. An attractive approach would make use of *in silico* models of metabolism to try to identify SEMP motifs and ultimately understand some of their roles. However, the gaps in biochemical databases and inaccuracies in transcription factor effector databases might make this task challenging for the time being. For instance, several of the studies cited in **Chapter 3** for describing natural SEMP's, stated that the target chemical was wrongly believed to be the *bona fide* effector of a transcription factor at the beginning of their study. This might also be the case for some of the molecules manually curated as effectors in databases, therefore precluding systematic analysis unless rigorous molecular biology experiments confirm the nature of the inducer. Even the lac operon, that has been studied for almost 60 years, is far from being completely understood. While the purpose of this regulated system was long believed to save the cost of unnecessary protein production, an elegant synthetic biology approach recently revealed that its main purpose is to avoid the major burden caused by the permease activity of LacY. One consequence of this discovery is that it provides a new hypothesis concerning the role of the indirect sensing motif of lactose. Because of the toxicity of the LacY activity, the activation of the operon must occur only when the cell can immediately benefit from lactose consumption. As the product of lactose by LacZ, the presence of allolactose is not only a sign that lactose is present but also that LacZ is functional. This step has therefore been proposed to serve to verify the functionality of LacZ before triggering the costly expression of LacY[118]. Coupled with the direct inhibition of LacY activity by glucose[119], this system could make sure period of LacY activity are strictly minimal and always immediately compensated by the benefits of lactose catabolism.

Part II: Phages as biological inputs

Contributions of the author

- Silencing of Antibiotic Resistance in E. coli with Engineered Phage Bearing Small Regulatory RNAs. Libis V*, Bernheim A*, Basier C, Jaramillo-Riveri S, Deyell M, Aghoghogbe I, Atanasković I, Bencherif C, Benony M, Koutsoubelis N, Löchner A, Marinkovic Z, Zahra S, Zegman Y, Lindner A, Wintermute E.

ACS Synthetic Biology, 2014 *contributed equally

doi: 10.1021/sb500033d. (Chapter 6 - Annex 3):

My contributions to this work consisted in the following: I designed the project, and took care of designing the DNA sequences. I performed the recombinant DNA work with Aude Bernheim and Clovis Basier. I designed the experimental characterization with Aude Bernheim and we performed it with the help of Clovis Basier. I wrote the article with Aude Bernheim and Edwin Wintermute.

- Phage-mediated delivery of targeted sRNA constructs to knock down gene expression in E. coli. Bernheim A*, Libis V*, Lindner A, Wintermute E.

Journal of Visualized Experiments, 2016 *contributed equally

doi: 10.3791/53618 (Chapter 4):

My contributions to this work consisted in the following: I designed the DNA sequences. Aude Bernheim and I designed the protocols and performed them. I wrote the article with Aude Bernheim and Edwin Wintermute.

-Chapter 5: On using phagemids as real-time metabolic engineering tools

My contribution amount to 100% of this preliminary work at this point of the project.

Chapter 4 : Phage-mediated delivery of targeted sRNA constructs to knock down gene expression in *E. coli*.

This work has been published in Journal of Visualized Experiments in 2016 by Aude Bernheim, Vincent Libis*, Ariel Lindner and Edwin Wintermute. (*contributed equally)*

Only very minor modifications have been made between the article and the chapter presented below.

4.1 Abstract

RNA-mediated knockdowns are widely used to control gene expression. This versatile family of techniques makes use of short RNA (sRNA) that can be synthesized with any sequence and designed to complement any gene targeted for silencing. Because sRNA constructs can be introduced to many cell types directly or using a variety of vectors, gene expression can be repressed in living cells without laborious genetic modification. The most common RNA knockdown technology, RNA interference (RNAi), makes use of the endogenous RNA-induced silencing complex (RISC) to mediate sequence recognition and cleavage of the target mRNA. Applications of this technique are therefore limited to RISC-expressing organisms, primarily eukaryotes. Recently, a new generation of RNA biotechnologists have developed alternative mechanisms for controlling gene expression through RNA, and so made possible RNA-mediated gene knockdowns in bacteria. Here we describe a method for silencing gene expression in *E. coli* that functionally resembles RNAi. In this system a synthetic phagemid is designed to express sRNA, which may be designed to target any sequence. The expression construct is delivered to a population of *E. coli* cells with non-lytic M13 phage, after which it is able to stably replicate as a plasmid. Antisense recognition and silencing of the target mRNA is mediated by the Hfq protein, endogenous to *E. coli*. This protocol includes methods for designing the antisense sRNA, constructing the phagemid vector, packaging the phagemid into M13 bacteriophage, preparing a live cell population for infection, and performing the infection itself. The fluorescent protein mKate2 and the antibiotic resistance gene chloramphenicol acetyltransferase (CAT) are targeted to generate representative data and to quantify knockdown effectiveness.

4.2 Introduction

RNA-mediated gene knockdowns proceed in two stages. First, an RNA molecule is introduced to a cell line or organism of study. Second, endogenous RNA-binding proteins facilitate RNA-target recognition and produce the silencing effect. All RNA knockdown technologies benefit from the customizable nature of synthetic sRNAs, which can be easily produced to match a specific target of interest. However, the molecular details of RNA uptake and silencing vary widely across model system, constraining where and how RNA knockdowns can be applied.

In nematodes, double-stranded RNA (dsRNA) molecules can be introduced directly in the media or by feeding the worms with a population of dsRNA-expressing *E. coli* cells[120,121]. In *Drosophila*, RNAi can be achieved by microinjecting embryos with dsRNA[122], or implemented in cell lines by simply adding dsRNA to the culture medium[123]. In mammalian cell lines, synthetic small interfering RNAs (siRNAs) may be delivered to living cells by electroporation[120,121,124], packaged in liposomes[122,125], or expressed from DNA plasmid vectors[123,126]. Once the RNA species reaches the cytosol, the RNAi pathway relies on the RISC complex to process dsRNA, facilitate antisense recognition of the target, and catalyze translational repression, mRNA degradation, or heterochromatin formation, depending on the host.

Because of these requirements, classical RNAi can be performed only in organisms that take up exogenous RNA efficiently and express RISC or a RISC-like activity. Notably, this excludes the model bacterium *E. coli*, which lacks the RNAi pathway. However, recent advances in synthetic biology provide the tools to solve both the delivery problem and the silencing problem.

In this protocol, sRNA constructs are expressed in *E. coli* from a DNA vector delivered to living cells using the M13 phagemid/helper system. A phagemid is any plasmid with a phage-derived f1 origin of replication. A helper plasmid, in this case M13KO, carries all the machinery required to produce viral particles, but is not itself competent for replication and packaging. When a phagemid and helper plasmid are co-transformed, the phagemid alone is replicated at the f1 origin, packaged and secreted. The vectorized phagemid is then competent to infect live *E. coli* via the F pilus.

In this system, the silencing effect is produced by custom sRNA cassettes combining a scaffold sequence with a target-binding sequence. The target-binding sequence is 24

base pairs antisense to an mRNA target, typically at the ribosome-binding site (RBS). The scaffold sequence, developed by Na and colleagues[127], contains an Hfq-binding motif extracted from MicC, a small regulatory RNA endogenous to *E. coli*. The Hfq protein stimulates RNA-RNA binding and mRNA degradation, serving a role in this system similar to RISC in RNAi. **Figure 20** depicts a complete scheme for phage-mediated sRNA knockdowns, including the sRNA cassette structure, phagemid vectorization, and silencing mechanism.

As a method to modulate gene expression in *E. coli*, sRNA silencing is simple, rapid and versatile. The targeted *E. coli* is not burdened beyond propagating the phagemid and expressing the sRNA. This may be relevant in the context of synthetic biology or basic research where the expression of larger heterologous constructs can strain cellular resources[128]. Phagemids with new targets can be produced with a single PCR and harvested one day after phagemid transformation. Finally, nearly any mRNA can be targeted. The sRNA regulation cassette (on a standard plasmid) has been shown to work on a variety of targets in metabolism with typical repression levels >90%⁸.

This protocol reproduces and expands on earlier work using phagemid-born sRNA cassettes[129]. First, a packaged phagemid is introduced to a batch culture of *E. coli* cells and used to silence expression of the fluorescent protein mKate2. Subsequent fluorescence changes are monitored in real time. Second, knocking down the CAT gene is shown to reduce phenotypic chloramphenicol resistance on agar plates. In both cases, the phagemid itself carries a GFP marker, allowing infection rate to be measured independently of the knockdown efficiency.

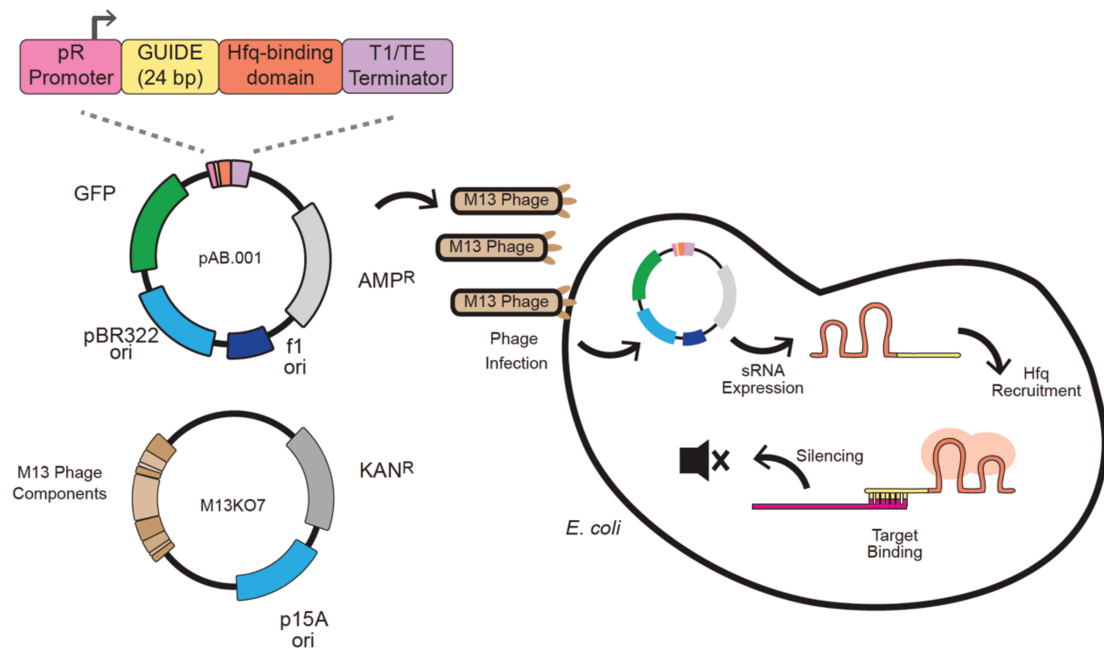


Figure 20: Gene silencing in *E. coli* with sRNA expression cassettes delivered by M13 phage. The sRNA cassette is composed of 4 modules: the pR promoter (a constitutive promoter derived from bacteriophage lambda), a 24 bp targeting domain, an Hfq binding domain extracted from MicC and a transcriptional terminator⁸. The cassette is incorporated into a phagemid bearing the M13 f1 replication origin and packaged into the M13 capsid using the helper phage M13KO7. Packaged phagemids are able to infect *E. coli* expressing the F pilus, where sRNA expression begins. The sRNA then recruits the Hfq protein (depicted in red) and binds an antisense mRNA target near the ribosome binding site, resulting in translational repression and mRNA degradation.

4.3 Protocol

4.3.1 Design and construction of phagemid vectors bearing sRNA silencing cassettes

De novo design of sRNA silencing cassettes[127]

Step 1) Identify the complete sequence of the mRNA to be silenced using a DNA sequence database. To generate the TARGET sequence, note the first 24 base pairs of coding sequence, from position +1 to +24 beginning with the start codon (e.g. ATG). Note: Silencing is less efficient when other sites or segments of the mRNA are targeted[127].

Step 2) Take the reverse complement of the TARGET sequence to produce the GUIDE sequence for the sRNA cassette. See **Table 3** for examples of TARGET and GUIDE sequences for chloramphenicol acetyltransferase (CAT).

CAT TARGET sequence	5' - ATGGAGAAAAAATCACTGGATAT - 3'
CAT GUIDE sequence	5' - ATATCCAGTGATTTTTTCTCCAT - 3'

Table 3: An example TARGET and GUIDE sequence for the CAT gene. Note the reverse complement relationship.

Step 3) To design the 292 bp complete sRNA expression cassette, arrange the pR promoter, GUIDE sequence, Hfq protein binding domain and T1/TE transcriptional terminator sequences in series (**Table 4**).

pR promoter	TAACACCGTGCGTGTTGACTATTTTACCTCTGGCGGTGATAATGGTTGC
GUIDE sequence	ATATCCAGTGATTTTTTTCTCCAT
Hfq binding domain	TTTCTGTTGGGCCATTGCATTGCCACTGATTTTCCAACATATAAAAAGA CAAGCCCGAACAGTCGTCCGGGCTTTTTTCTCGAG
T1/TE terminator	CTCGAGCCAGGCATCAAATAAAACGAAAGGCTCAGTCGAAAGACTGGGC CTTTCGTTTTATCTGTTTTGTCTGGTGAACGCTCTCTACTAGAGTCACA CTGGCTCACCTTCGGGTGGGCCTTCTGCGTTTATA

Table 4: Sequence components of the sRNA cassette. Each sequence is written 5'-3'. The complete cassette is the concatenation of these 4 elements in order and comprises 292 bp.

Step 4) Add additional cloning sites of choice to facilitate cloning of the sRNA cassette into the target vector.

Step 5) Obtain the complete sRNA cassette through commercial gene synthesis or a similar method and clone it into any phagemid vector with a functional f1 replication origin[130]. See Supporting Information for the complete sequence of the final phagemid vector.

Altering the sequence targeted by an existing sRNA expression cassette using PCR-based site directed mutagenesis[131].

Step 1) Identify the 24-bp GUIDE sequence in an existing sRNA expression cassette. Note: The annotated pAB.001 plasmid, used in this work, is available as a supplemental sequence file.

Step 2) Design forward and reverse primers with short regions of homology to the existing sRNA cassette flanking the new 24-bp GUIDE sequence. Obtain the primers through commercial oligonucleotide synthesis. Note: Primer design for site-directed mutagenesis is depicted in **Figure 21**. Exact primer template sequences are provided in **Table 5**.

Forward Primer	5' - CTGGCGGTGATAATGGTTGC[GUIDE]TTTCTGTTGGGCCATTGC - 3'
Reverse Primer	5' - GCAATGGCCCAACAGAAA[TARGET]GCAACCATTATCACCGCCAG - 3'

Table 5: Primer design to alter existing GUIDE elements. The forward primer includes the last 20 bp of the pR promoter, the new GUIDE sequence, and the first 18 bp of the Hfq-binding domain. The TARGET sequence is the exact reverse complement of the GUIDE sequence. The reverse primer is the exact reverse complement of the forward primer.

Step 3) Prepare a 5 mL culture of *E. coli* carrying the template sRNA expression phagemid. Grow the cells overnight at 37 °C with shaking, in LB media with appropriate antibiotics.

Step 4) Extract and purify the template sRNA expression phagemid from the 5 mL bacterial culture using a DNA miniprep kit or similar method[131].

Step 5) Prepare 2 PCR reactions using the template sRNA expression phagemid and high fidelity polymerase, one with the forward and one with the reverse primer (**Table 6**). Use PCR conditions as recommended by the polymerase supplier (**Table 7**). Increase the template concentration to 10-50x higher than a standard reaction to account for the fact that the single primer reaction does not produce exponential amplification.

Component	Volume
Template plasmid	500 ng
10 µM Primer	2.5 µL
10 mM dNTPs	0.5 µL
High-fidelity Polymerase (2 U/µL)	0.5 µL
5x PCR Buffer	10 µL
Total volume	50 µL

Table 6: Suggested conditions for the single-primer mutagenic PCR.

Step	Temp	Time
Initial denaturation	98 °C	30 s
30 cycles	98 °C	10 s
	55 °C	30 s
	72 °C	120 s
Final Extension	72 °C	300 s
Storage	10 °C	

Table 7: Suggested thermocycler protocol for the single-primer mutagenic PCR.

Step 6) Combine the two above PCR reactions in a microcentrifuge tube. Anneal the products by heating to 98 °C in a boiling water bath. Immediately after placing the microcentrifuge tube in the water bath, remove the heat source and allow the bath to slowly return to room temperature over 1-2 hr.

Step 7) To eliminate the unmutated template sRNA expression phagemid, add 1 µL DpnI restriction enzyme to the mixture and incubate at 37 °C for 1 hour, or the time recommended by the manufacturer for complete digestion. Note: DpnI digests only methylated target sites, which are present on host-replicated phagemids but not PCR products.

Step 8) Transform purchased or prepared[132] chemically competent *E. coli* with 1-5 µL of the annealed PCR product. Isolate single colonies of the transformed strain by selective plating on LB agar plates containing appropriate antibiotics.

Step 9) To verify the incorporation of the correct GUIDE sequence, screen the resulting colonies by colony PCR. Using a 200 µL pipette tip, collect a small quantity of cells from a single transformed colony. Mark and preserve the original colony for downstream use after verification.

Step 10) Add the collected cells to 50 µL of nuclease-free water in a microcentrifuge tube. Mix by pipetting up and down.

Step 11) Using a bench-top thermocycler or a boiling water bath, lyse the cells by heating to 95 °C for 2 minutes.

Step 12) PCR-amplify the phagemid region using 1 µL of heat-lysed cells as a DNA template. PCR conditions and thermocycler protocols are provided in **Tables 8 and 9**. See the supplemental pAB.001 sequence file for verification primer sequences.

Component	Volume
Template DNA	1 μ L
10 μ M Forward Primer	0.5 μ L
10 μ M Reverse Primer	0.5 μ L
<i>Taq</i> 2X Master Mix	25 μ L
Nuclease-free water	23 μ L
Total volume	50 μ L

Table 8: Suggested conditions for the sequence verification PCR.

Step	Temp	Time
Initial denaturation	95 °C	30 s
30 cycles	95 °C	30 s
	55 °C	30 s
	68 °C	30 s
Final Extension	68 °C	300 s
Storage	10 °C	

Table 9: Suggested thermocycler protocol for the sequence verification PCR.

Step 13) Sequence the PCR product to verify incorporation of the correct GUIDE sequence.

Step 14) Inoculate a 5 mL culture of the *E. coli* clone carrying the sequence-verified sRNA expression phagemid. Grow the cells overnight at 37 °C with shaking, in selective LB media.

Step 15) Prepare a glycerol stock of the sequence-verified clone. Add 750 μ L of the overnight culture to 250 μ L of 60% glycerol in a screw-cap cryo tube.

Step 16) Store the glycerol stock at -80 °C indefinitely. The remainder of the overnight culture may be used as a source for the sRNA expression phagemid in step 2.

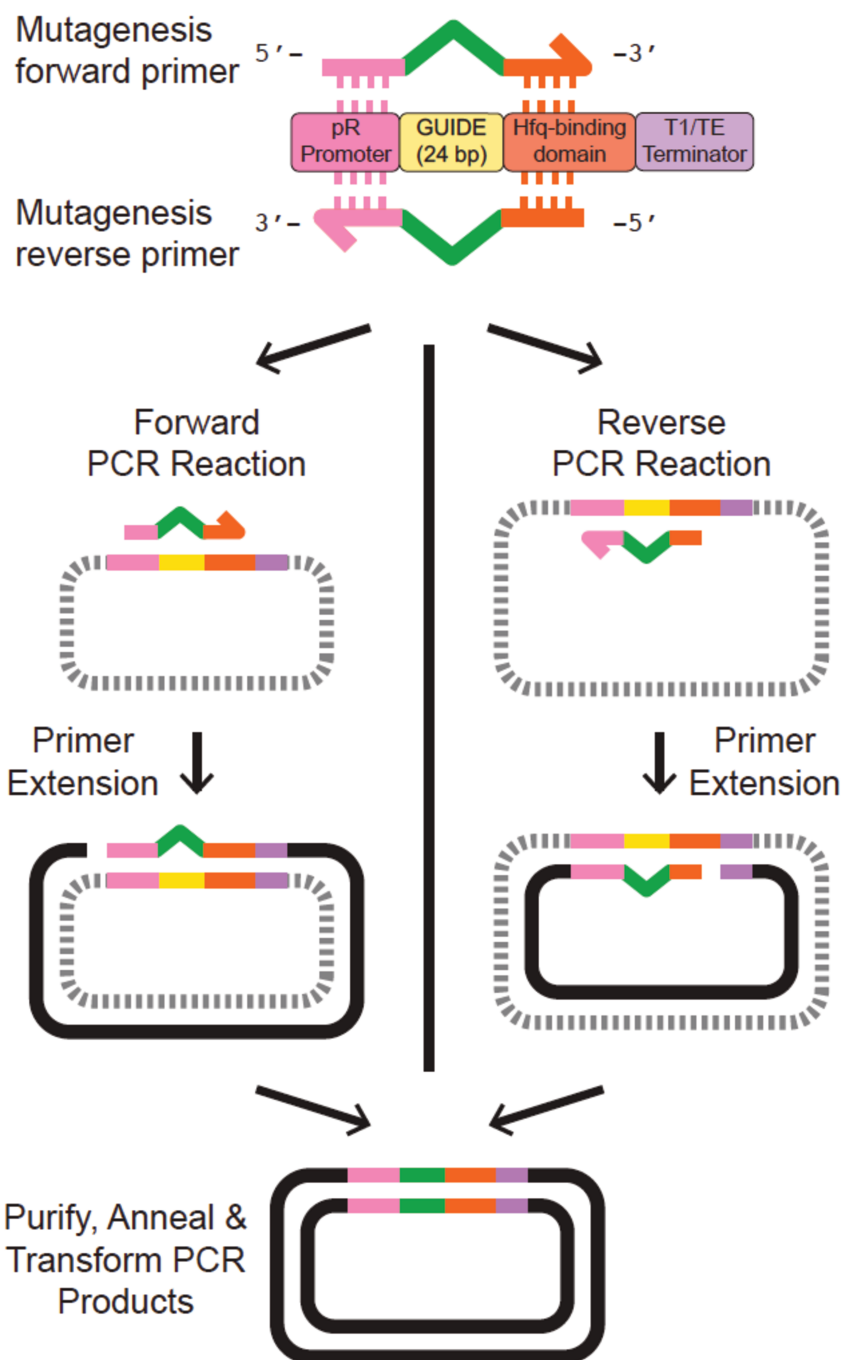


Figure 21: Primer design and site-directed mutagenesis of the sRNA target site. Two primers are designed with partial homology to an existing sRNA cassette. The forward primer contains 20 bp homologous to the pR promoter at the 5' end, followed by 24 bp representing the new GUIDE sequence, then 18 bp homologous to the Hfq-binding domain at the 3' end. The reverse primer is the exact reverse complement of the forward primer, with regions of homology to the existing sRNA cassette flanking the reverse complement of the new GUIDE sequence. Exact primer sequences are given in table 3. Separate single-primer PCR reactions with the forward and the reverse primers produce linear, single-stranded DNA with the desired modified sequence. Annealing the forward

and reverse reaction products, following clean-up as described in the protocol, results in double-stranded plasmid DNA bearing the desired modified sRNA cassette.

4.3.2 Production and harvest of M13-packaged phagemid stocks

Step 1) Prepare a 5 mL culture of *E. coli* carrying the sRNA expression phagemid. The sRNA expression phagemid may be obtained through *de novo* cloning as described in step 1.1.5, or modified from an existing phagemid and harvested in step 1.2.16. Grow the cells overnight at 37 °C with shaking, in LB media with appropriate antibiotics.

Step 2) Similarly prepare a 5 mL culture of *E. coli* carrying the M13KO7 helper plasmid. Grow the cells overnight at 37 °C with shaking, in selective LB media.

Step 3) Extract and purify the sRNA expression phagemid and helper plasmid using a DNA extraction kit or similar method[131].

Step 4) Cotransform purchased or prepared[132] chemically competent *E. coli* with 1 µL each of sRNA expression phagemid and helper plasmid. Select for cotransformants by plating on LB agar with selective antibiotics for both constructs.

Step 5) Prepare a 10 mL culture from a single colony of the cotransformed strain in LB with selective antibiotics. Incubate at 37 °C with shaking for 8-12 hours or overnight.

Step 6) Centrifuge the culture at 3300x g for 10 minutes. Collect the supernatant and filter through a 0.2 µm filter. Caution: In case of media spill, clean the area with dilute bleach (0.5%) to destroy infective phage particles.

Step 7) Store the packaged phagemid filtrate at 4 °C. Note: Samples can be maintained for days to weeks without loss of activity.

4.3.3 Preparation of F⁺ target cells for silencing

Step 1) Determine if the cells to be targeted for silencing express the F pilus[133]. Common laboratory strains of *E. coli* are annotated as F⁺ or F' to indicate the presence of the F pilus in their genome or on a plasmid. If the F pilus is already present, proceed to step 4.

Step 2) Obtain an F⁺ strain of *E. coli* such as TOP10F'. Note: Ensure that the target strain carries a unique resistance marker in order to be separated from the F-plasmid donor after conjugation.

Step 3) To introduce F pilus by conjugation with a F⁺ strain, prepare 5 mL cultures of both the target strain and F-plasmid donor[133]. Grow the cells overnight at 37 °C with shaking, in LB media with appropriate antibiotics.

Step 4) The following day, dilute both strains 1:100 in 5 mL of selective LB and continue culturing at 37 °C with shaking.

Step 5) Determine the growth phase of the cells by measuring the optical density of the culture at 600 nm (OD₆₀₀) using a benchtop spectrophotometer. Culture the cells for about 2 hours until an OD₆₀₀ of 0.3 is attained, indicating log-phase growth[134].

Step 6) Prepare 3 conjugation reactions in microcentrifuge tubes: 0.5 mL F-plasmid donor + 0.5 mL target strain, 0.5 mL F-plasmid donor + 0.5 mL LB media (negative control) and 0.5 mL target strain + 0.5 mL LB media (negative control). Allow the conjugation to proceed for 2 hours at 37 °C with shaking.

Step 7) Plate 100 µL of each conjugation reaction on selective LB agar with antibiotics specific to the F-plasmid (typically tetracycline) and target strain. Plate the negative control reactions to confirm that neither donor or recipient strain express both antibiotic resistances.

4.3.4 Infection with packaged phagemids for silencing

Step 1) Inoculate a single colony of F⁺ target cells into a 5 mL culture of LB media with appropriate antibiotics. Incubate overnight at 37 °C with shaking.

Step 2) The following day, dilute the F⁺ target cells 1:100 in 5 mL of selective LB media and continue to culture at 37 °C with shaking.

Step 3) Determine the growth phase of the cells by measuring the optical density of the culture at 600 nm (OD₆₀₀) using a benchtop spectrophotometer. Culture the cells for

about 2 hours until an OD₆₀₀ of 0.3 is attained, indicating log-phase growth[134]. Note: Expression of the F pilus and infection efficiency is highest in log phase.

Step 4) Add the M13-packaged phagemids (from step 2.6) to the target cells in a volume ratio of 1:100 to achieve nearly 99% infection of the target population. Allow the infection to proceed at 37 °C with shaking for 30 - 60 minutes.

Step 5) Assay the sRNA-silencing phenotype according to the method of choice. Note: For a fluorescent protein target, the silencing effect can be quantified directly by fluorometry[127]. Alternately, phenotypic assays may be used to observe the phenotypic consequences of gene knockdown[129].

Step 6) Prepare a glycerol stock for the sRNA-expressing phagemid host following steps 1.2.14 - 1.2.16. Note: the phagemid will propagate in the host strain indefinitely and can be maintained with antibiotic similar to a conventional plasmid.

4.4 Representative results

4.4.1 Silencing of mKate2 fluorescence in liquid media

Figure 20 depicts the scheme for sRNA-mediated knockdowns described in this work, including the sRNA cassette design, phagemid vectorization, and silencing mechanism. Following protocol 1.2, the sRNA silencing cassette of plasmid pAB.001 was altered to target mKate. The sRNA cassette was synthesized and cloned into phagemid Litmus28i_J23115-B0032-GFP, a gift from Monica Ortiz and Drew Endy[130]. This phagemid carries markers for GFP expression and kanamycin resistance, allowing successful infections to be tracked. Packaged phagemid stocks were prepared following protocol 2.

A derivative of *E. coli* K12 MG1655 carrying a constitutively expressed, chromosomally integrated mKate2 marker was prepared for phage infection by conjugation with an F plasmid donor strain following protocol 3. The cells were grown to mid log phase and the phagemid introduced following protocol 4. After phagemid infection, 200 µL cultures were transferred to a fluorescence plate reader and fluorescence was monitored continuously for 24 hours. **Figure 22** shows the effect of sRNA-mediated silencing on mKate2 expression. The strain infected with an anti-

mKate2 phagemid showed no detectable mKate2 fluorescence over background. In contrast, this strain did express the GFP marker, indicating successful uptake of the phagemid. Uninfected control cells produced mKate2 fluorescence but not GFP. An additional control, in which the anti-mKate2 targeting domain was replaced with a sequence targeting CAT, had no effect on mKate2 fluorescence.

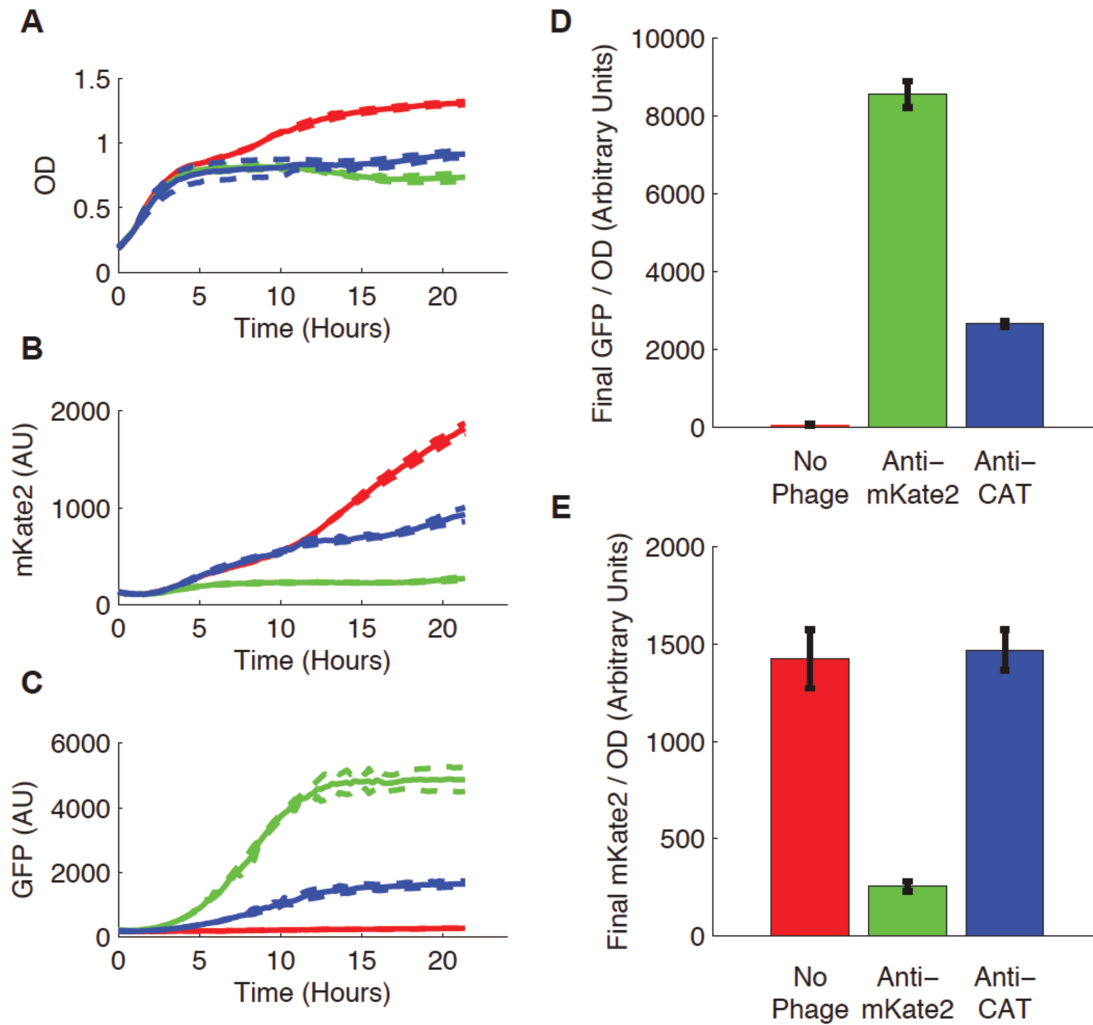


Figure 22: Knockdown of a chromosomally integrated mKate2 fluorescent reporter.

E. coli MG1655 K12 expressing mKate2 were either left untreated (red lines and bars), infected with an anti-mKate2 phagemid (green lines and bars), or infected with a control phagemid targeting CAT (blue lines and bars). A. Untreated *E. coli* grew to higher saturation densities, indicating a metabolic cost to phage infection. Dotted lines indicate standard deviations of 3 replicates. B. The mKate2 signal was reduced to near background levels in the anti-mKate2 treated strain, but not in control strains. C. GFP fluorescence, also carried by the phagemid, was detectable only in phagemid-treated controls. D, E. Final fluorescence readings after 24 hours of growth were normalized to OD. GFP signal, indicating phagemid infection, was absent in untreated controls, but also substantially reduced following anti-CAT phagemid treatment. This may indicate off-target effects of the phagemid. The mKate2 signal was reduced by anti-mKate2 phagemid treatment compared to

untreated controls. The CAT-targeted control phagemid showed no effect on mKate2 fluorescence. Error bars represent standard deviation of 3 replicates.

4.4.2 Silencing of chloramphenicol resistance on agar plates

Following protocol 1.1, an sRNA silencing cassette was produced to target CAT. A derivative of *E. coli* K12 MG1655 carrying a constitutively expressed, chromosomally integrated CAT marker was prepared for phage infection by conjugation with an F-plasmid donor strain following protocol 3. The cells were grown to mid log phase and the phagemid introduced following protocol 4. After 1 hour of incubation at 37 °C, infected cells were serially diluted and plated at a range of chloramphenicol concentrations. Plates were incubated overnight, and the proportion of resistant cells at each chloramphenicol concentration was determined by counting colony forming units (CFUs) the next day.

Figure 23 shows the effect of sRNA-mediated silencing on the chloramphenicol resistance phenotype. Uninfected cells, or cells infected with phagemid targeting mKate2, were resistant to chloramphenicol at all concentrations tested. In contrast, cells infected with phagemid targeting CAT showed reduced survival at low chloramphenicol concentrations, and nearly 99% killing at higher concentrations. The addition of ampicillin, to select for only bacteria carrying the phagemid, reduced chloramphenicol survival to undetectable levels. This is consistent with earlier work showing that escape from phage infection is a common route to escape silencing[129].

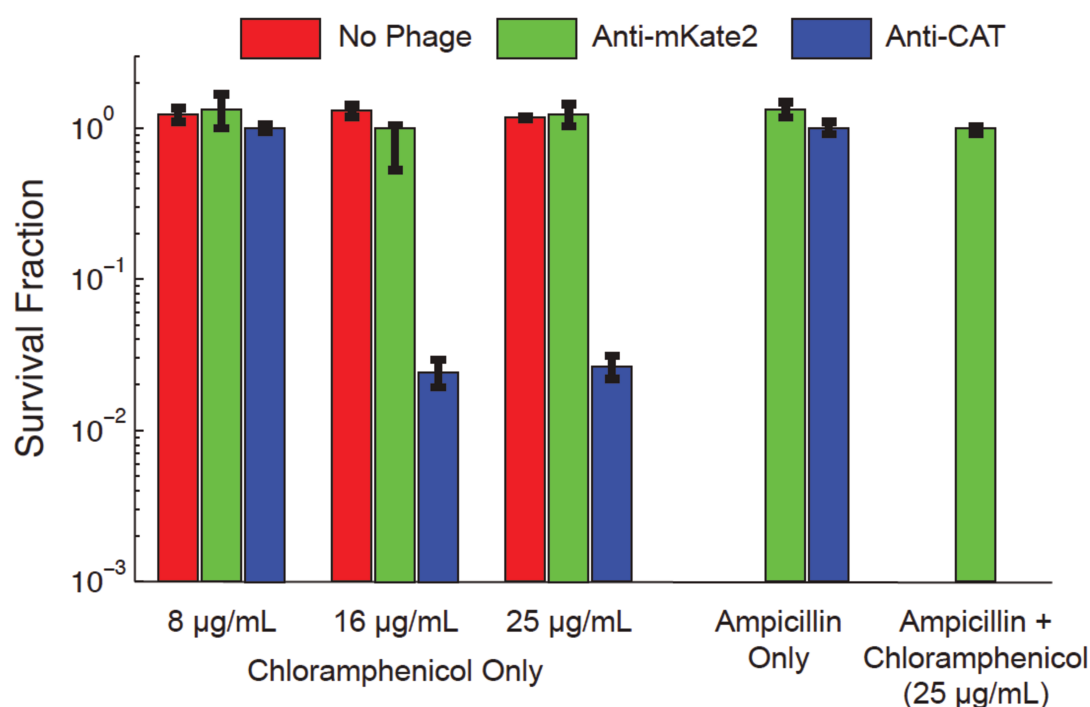


Figure 23: Knockdown of CAT restores chloramphenicol sensitivity to a genetically resistant population.

E. coli MG1655 K12 expressing a chromosomally integrated CAT gene were left untreated (red bars), treated with a control phagemid targeting mKate2 (green bars), or treated with a phagemid expressing an anti-CAT sRNA (blue bars). After 1 hour of infection, viability on the indicated antibiotics was assessed by serial dilutions and plating. Strains treated with anti-CAT phagemid were significantly killed (>90%) by chloramphenicol at higher concentrations, while control treatments were unaffected. Adding ampicillin to the culture plates positively selects for phagemid infection and eliminates uninfected cells. Under these conditions, no chloramphenicol resistant colonies were observed after anti-CAT treatment. This indicates that most survivors represent a failure of infection, rather than a failure of silencing. Error bars represent the standard deviation of 3 replicates.

4.5 Discussion

The present method achieved 80% reduction in mKate fluorescence levels compared to untargeted controls. This is in line with other RNA knockdown methods, where complete silencing is not observed and 50-90% efficiency is typical[23,135]. At the phenotypic level, CAT-targeted knockdowns were able to significantly attenuate chloramphenicol resistance, and eliminate it under some conditions.

The knockdown phenotype was detectable at the population level after only a few hours post-infection (see **Figure 22B**). This illustrates an important feature of phage-based delivery: a high knockdown frequency can be obtained directly in batch culture without prior genetic modification. Unlike conventional genetic modifications using plasmid transformation or genomic integration, phage infection does not require that a population be re-grown from a single isolated colony. This allows the effects of phage infection to be explored in populations with complex spatial dynamics[130], with pre-existing spatial structures like biofilms[136], or in genetically mixed natural populations[137].

A critical step in this method is the production of packaged phagemid at high titer. The metabolic burden associated with phage particle production may lead to high rates of mutation or plasmid loss in the phagemid production strain. It is recommended that the phagemid production strain be cultured directly from a single cotransformed colony and not refrigerated, frozen or sub-cultured prior to phage harvest. A low efficiency of co-transformation may also be observed when introducing the phagemid and helper plasmid to *E. coli* simultaneously. In this case, higher efficiencies may be obtained by transforming the helper plasmid first, then preparing competent cells bearing the helper plasmid for subsequent transformation with the phagemid.

Phagemid infection or sRNA expression also imposes a detectable metabolic burden on the target cells, and may result in some phenotypic perturbation. For example, a reduction in mKate2 fluorescence was observed even when cells were infected with a phagemid targeting CAT (see **Figure 22**). Infection with M13 is not thought to trigger systemic stress responses in *E. coli*[138], but may alter transcriptional patterns indirectly. Alternately, the GFP or ampicillin resistance markers included on the phagemid may compete for cellular resources, reducing mKate2 expression and growth[128]. Finally, the sRNA cassette itself may globally alter gene expression profiles by titrating the Hfq protein, or through off-target mRNA silencing. Off target effects are

common in *in vivo* RNAi targeting[139–141], but they have yet to be systematically investigated for this system.

One limitation of this method is that the infection efficiency is less than 100%, allowing some non-infected bacteria to persist in the population. The results of this work and earlier work[129] suggest that noninfected cells represent 1-10% of the final population, and are responsible for most of the nonsilenced phenotypes observed. A variety of routes to M13-resistance are known, with the most common being mutational loss of pilus expression[142]. In light of these limitations, controls should be used to confirm high infection rates and knockdown efficiency.

Another potential limitation for some applications is the occasional transfer of contaminating helper phage. Although M13K07 contains a mutated packaging signal, it can be packaged into phage capsid at low frequency and transferred to infected populations, resulting in cells competent for phage production and the continued spread of phage beyond the initial infection event[143]. Modifications to the helper phage have proven effective at reducing nonspecific packaging, although sometimes at the cost of reduced phage production[144].

Engineered bacteriophage have become an indispensable tool for *E. coli* synthetic biology, allowing quick delivery of new genes to a growing population. Recent work has produced intercellular communication circuits[130] or expressed transcription factors to suppress antibiotic resistance pathways[145]. The protocol presented here adds to a growing collection of tools that allow control of bacterial physiology through programmed RNAs. CRISPR-Cas nucleases, when mutated to eliminate nuclease activity, have been shown to repress transcription at RNA-guided gene targets[23,146]. In contrast, sRNA silencing works at the translational level and does not require exogenous protein expression. Next-generation biotechnologies may unite transcriptional and translational control with phage-mediated delivery to program complex phenotypes in real time.

Chapter 5 : On using phagemids as real-time metabolic engineering tools

5.1 Abstract

Bioproduction of value-added compounds by genetically modified microorganisms is usually done through a first phase of biomass propagation, followed by a phase of production of the compound of interest triggered by the addition of a chemical inducer in the fermenter. This chemical-based induction can be costly and adds strong design constraints at the DNA level by reducing the number of promoters available to control the engineered metabolic pathway. Here we present an alternative to classical induction which relies on simultaneous infection of the growing cells by engineered phages able to reroute metabolism in real time toward the production of value-added compounds. Sequestration of all the necessary heterologous genes and synthetic biology tools in the phage DNA allow replacement of the induction step by an infection step and relieves several design constraints. While phages are traditionally perceived as a threat for industrial bioproduction, we show that these parasites can be effectively turned into metabolic engineering tools.

5.2 Introduction

Bioproduction of compounds by metabolically engineered microorganisms is regarded as a way to decrease costs compared to classic ways of production for a growing number of molecules. It allows production at high yields of natural products that are normally synthesised in small quantities by plants or microorganisms, and also provides a green and sustainable alternative to petroleum-based chemistry. Considering the vast impact it could have on the broad chemical industry, the field of metabolic engineering have justified important research efforts that have led to creative solutions to increase performance of strains and bioprocesses. For instance, computer-aided design tools now help with pathway design and choice of enzymes among thousands of variants found in nature[42]. Flux balance analysis and high dimensional experimental design strategies speed up bottleneck identification in pathways. Protein engineering efforts enhance

enzymatic properties, disrupt allosteric negative feedback, or create completely novel biochemical reactions[147,148]. While recombinant DNA work was slowing down metabolic engineering projects in the past, new techniques allow combinatorial cloning in plasmids and multiplex genome engineering[149]. In parallel, biosensors are being developed to help in screening the billions of variant strains thus generated[78]. Ambitious attempts at imitating natural pathways characteristics such as compartmentalization[150,151], scaffolding[152–154] and dynamic pathway regulation[79,155] have led to increased yields, titers, and productivities.

Some of these proof of principles are then being transferred to the industry which plays an integrating role for technologies coming from very distant fields such as carbon source preparation, bioprocess engineering, chemical engineering and purification. Each of these axes necessitate *ad hoc* optimization to bring the overall cost closer to commercial viability. Because these axes remain interconnected, extensive additional metabolic engineering efforts are necessary all along the way to optimize strains studied as proof-of-concept in laboratory settings. For example, while optimizing fermentation of microorganisms that naturally produce value-added compounds amounts to recreate conditions in which the « natural program » of metabolite production of the microorganism is fully activated, genetically modified microorganisms require their production program to be written and tuned at the DNA level for optimal fermentation. A common feature that is desired for such a production program is the ability to switch between a biomass propagation phase and a value-added metabolite production phase. Separating these two distinct phases decreases the loss in yield caused by emergence of non-producer mutants over the 45 to 70 generations necessary to populate cubic-meter scale industrial fermenters[156]. Because such a number of generations theoretically warrants emergence of variants for each position in the genome, high-burden, evolutionary fragile engineered pathways are often suffering from counterselection pressure. Maintaining the engineered pathway in an OFF state during the biomass propagation phase can be executed by genetic switches that avoid enzyme expression and rerouting of fluxes leading to a decreased fitness, until a chemical inducer is added to the medium.

However, the need for genetic regulation imposes strong design constraints that are difficult to meet with the current bioparts toolbox. Two main strategies are used to trigger gene expression in response to a chemical inducer: translationnal control through riboswitches and transcriptionnal control through allosteric transcription factors. They require insertion of specific 5'UTR sequences or promoters in front of

target genes. While these insertions are easily done during the cloning of heterologous genes, insertions in front of endogenous genes necessitate important genome engineering efforts. Another issue arises from the nature of the molecules that can serve as inducers. Derived from the first inducible genetic system ever studied, promoters based on the LacO operator are usually among the first tools developed to alleviate genetic instability of synthetic pathways in new hosts[157]. Variants of IPTG-inducible promoters exhibiting different transcriptional strengths have been developed[158] and they are frequently used in laboratory settings to achieve new records in terms of yield, titer or productivity for various metabolites[159]. However the cost of IPTG prohibits its use at commercial scale[160]. This is also true for most inducers used in academic research and while it does not affect the validity of the demonstrated concepts, reproducing these advances in industrial settings will require the development of promoter with similar properties upon induction by cheaper chemicals such as copper[161]. Expanding the number of inducers for riboswitches or transcription factors represents an important effort in synthetic biology at the moment[100,162] but well characterized industrially compatible candidates are scarce. Lastly, potential promoters must be tightly repressed in absence of inducer and exhibit a large dynamic range of expression.

Here we present a potential solution to the problem of costs of chemical inducers, the scarcity of inducible systems exhibiting satisfactory dynamics and the need for genome engineering. By sequestering the synthetic DNA elements extracellularly into an engineered phage we avoid their expression until infection and thus allow any constitutive promoter to be switched from OFF to ON on demand. This bioproduction framework requires the creation of nonlytic phages able to simultaneously infect every bacteria in a fermenter, dynamically reroute their metabolism by bringing genes coding for new metabolic pathway, as well as deploy tools to downregulate competing native pathways. While ambitious, this concept provides in return the necessary timed control over metabolic modifications, offers the best dynamic ranges of gene expression, avoids leakiness issues, and avoids the design and economical constraints associated with inducible promoters (see **Figure 24**). We describe the first attempt at implementing this concept and the development of phages allowing successful bioproduction of value-added compounds such as lycopene and deoxyviolacein.

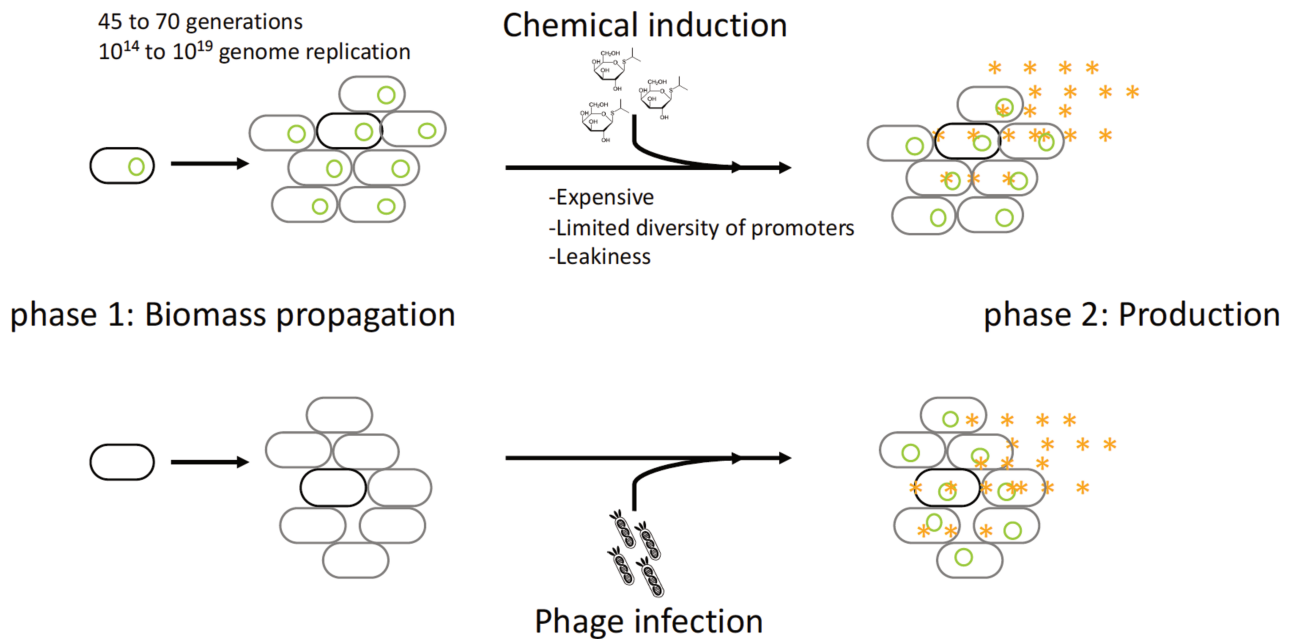


Figure 24 : standard chemical induction protocol and alternative phage-based metabolic engineering. Green circles represent heterologous genes construct in the form of plasmids, chromosomal integrations or phagemids.

5.3 Results

5.3.1 Identification of a viable design framework for phage-based metabolic engineering.

This concept requires a set of particular characteristics from the phage vector. The phage must not destroy the membrane of the bacteria, and would ideally cause minimal burden on the cells to spare resources and avoid yield and productivity diminution. The phage must be able to carry large genetic cargos from a few kilobases to tens of kilobases. The phage must be able to influence the metabolism in a programmable way to reproduce modifications that metabolic engineers typically undertake such as overexpression and down regulation of certain genes. Finally, the phage must be able to infect simultaneously the vast majority of a dense population of bacteria. We identified a candidate that fulfills all these criteria in the form of a modified M13 filamentous phage. M13 phage is lysogenic and therefore naturally allow *E. coli* to

survive infection. Furthermore, we used a modified version of M13 called phagemid where all the genes of the phage have been removed from its genome and moved to a helper phage unable to encapsulate itself in viral particles. In this phagemid configuration, the empty M13 genome is left with just its origins of replications and packaging signal sequence that allow the DNA to be encapsulated in the form of infective particles if the helper phage is present in the same cell to provide the structural proteins for the capsid. Phagemids are able to infect cells and deliver their DNA but cannot replicate and thus avoid wasting cellular resources on building structural proteins. Since the M13 capsid proteins polymerise around the single stranded genome in the form of a rod of proportional length, structural constraints on the size of the cargo are minimal and genome size up to 40 kilobases have been reported[163]. Finally, we have shown previously that M13 phagemids can vectorize synthetic biology tools such as sRNA to downregulate chromosomal genes at will and that infection rate is sufficient to impose the targeted downregulation to more than 99% of a population of bacteria[129,164]. **Figure 25a** presents a design framework we propose for the phage DNA to exert metabolic engineering on infected bacteria. Three potential value-added compounds, lycopene, violaceine and pinocembrine and their associated metabolic pathways from glucose are described in **Figure 25b** with a set of suggested phage-based modifications.

5.3.2 Fermentation of lycopene and deoxyviolacein can be achieved upon phage infection.

A first generation of phagemid was built to carry heterologous metabolic genes (see **methods**). After phagemid production in phage-factory cell lines harboring the helper phage M13K07, the infective particles were isolated from the cells by filtering the supernatants. Cultures of *E. coli* MG1655 harboring the F pilus was grown to log phase in minimal medium, infected with the different phages, and further incubated for 24 hours. **Figure 25c** shows a characteristic change in color for cells infected with phages coding for lycopene and deoxyviolacein indicating that phages can graft functional heterologous pathway on the native metabolism.

5.3.3 Engineered phages allow silencing of gene targets across the chromosome, including targets in the central metabolism and essential genes.

We built phages harboring silencing modules based on sRNAs for the relevant targets previously reported to increase production of the proposed compounds of interests (see **Figure 25b**). FabB and FabF are involved in the synthesis of fatty acids by consuming malonyl-CoA, a limiting precursor in the synthesis of pinocembrine. For this reason, pinocembrine production protocols often seek to disrupt fatty acids synthesis by using CRISPRi tools[165] or by adding the antibiotic cerulenin in the medium, but its cost is prohibitive for industrial production. TrpR is a transcription factor that limits overproduction of aromatic amino acids including the precursor of deoxyviolacein, tryptophan, and the knockdown of TrpR has been shown to lead to an increased production of both tryptophan[166] and deoxyviolacein[167]. Knockdown of *tnaA* have also been shown to lead to increased violacein production by avoiding conversion of excess of tryptophan back into indole[168]. Essential genes *gdhA* and *aceE* have been suggested as downregulation targets by FBA models trying to optimize flux toward lycopene production and their impact was confirmed experimentally[169]. Lastly, we targeted the genes *rpoB* and *rpoC* coding for the RNA polymerase of *E. coli* because switching it off have been proposed as a way to block cell growth while cells continue to be metabolically active. This strategy have been applied previously to increase the yield of glycerol production in *E. coli* [170]. In order to allow cloning of such highly toxic sRNA with promoters that would be constitutively active in the target cells, we chose to use the P_R promoter from phage lambda and expressed its repressor *cI* in the cell lines used for cloning and production of the phages (see **methods**). Because the *cI* repressor is not present in the target cells, the sRNAs are strongly and immediately produced upon infection. The same concept can be extended to any constitutive promoter by designing complementary gRNAs controlling a CRISPRi system in the cloning strain and phage-factory cells. **Figure 25d** shows *E. coli* strains with and without *cI* repressor, after transformation with the phagemids bearing the new sRNAs. Three target genes, *aceE*, *gdhA*, and *rpoB*, led to strongly impaired growths on LB-agar in absence of the *cI* repressor. This shows both the efficiency of the sRNA silencing modules carried by phagemids and the possibility to manipulate phagemids that target essential genes.

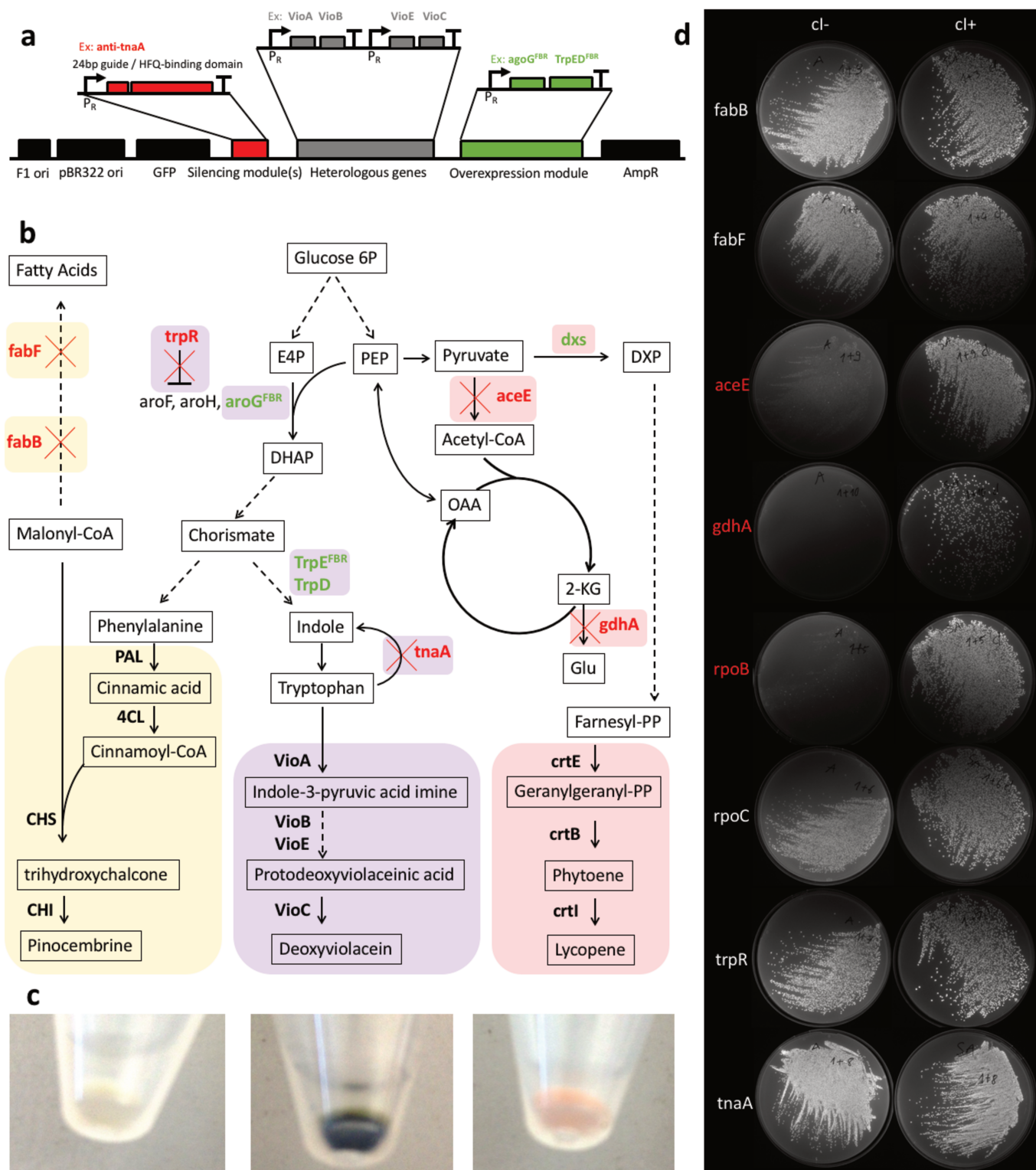


Figure 25 : Design and characterisation of phages for metabolic engineering. a) Proposed design of phagemid DNA including one or several silencing modules to downregulate competing pathways, a heterologous genes module, an overexpression module to increase amount of important homologous enzymes or introduce feedback resistant homologues. b) Examples of metabolic pathways leading to value-added compounds from glucose and potential gene targets for downregulation (red font) and overexpression of enzyme homologues (green font) including feedback resistant variants (^{FBR}). Light yellow, light violet, and light red background color indicate

new genes and regulation targets related to pinocembrine production, deoxyviolacein production, and lycopene production, respectively. **c)** Colored aspect of the cell pellets from 2mL of culture, 24 hours after infection by phages carrying lycopene or deoxyviolacein pathway. **d)** Transformants of phagemids bearing silencing module for different targets in *E. coli* Mach1 with and without basal expression of the cI repressor. Because transcription of the sRNA is controlled by the P_R promoter, competent cells harboring cI repressor into a pCDF-Duet plasmid survived transformation with all the silencing modules including those targeting essential genes (in red).

5.3.4 Characterization of fermentation features.

We varied the infection time to identify the optimal cell density that must be reached prior to infection for maximum production. **Figure 26** shows the impact of this parameter on production of either lycopene or deoxyviolacein in minimal medium with 2% glucose. In both case high cell density at infection led to suboptimal production of the compounds. This is probably due to the depletion of glucose that went into the biomass instead of yielding the compounds of interest. This could also be partially caused by the loss of the pilus taking place at the end of the log phase. Surprisingly, the optimal cell density at infection is different for each pathway. One hypothesis is that the toxicity of deoxyviolacein might contribute to a strong counterselection pressure with an observable impact when a lot of generations are still possible. We also compared production at all timepoints with and without addition of ampicillin in the medium 30 minutes after infection. Because the phages carry an ampicillin resistance gene, this step guarantees that no cell that would have escaped phage infection will grow and compete with infected cells. No significative difference in production was observed with and without ampicillin (see **Figure 26**). This indicates that engineered phages, contrary to plasmids, do not need exterior help to impose the new metabolic program to the cells. Comparison to a standard curve indicates a pic of deoxyviolacein production of 42,78mg/L which is similar to the results obtained in early recombinant works in *E. coli* with IPTG induction[171].

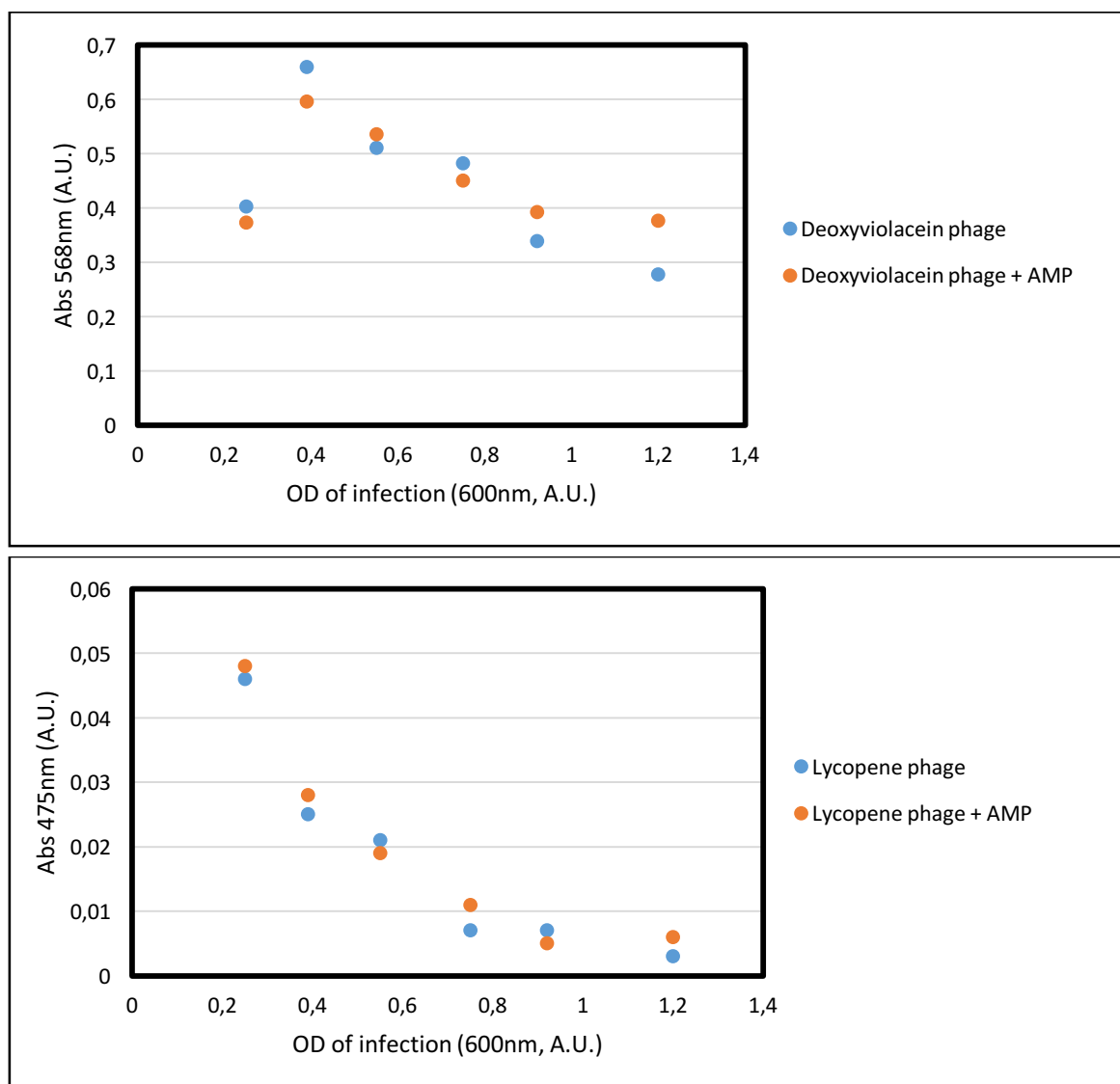


Figure 26. Relative production of deoxyviolacein and lycopene after infection at different stage of growth. Production of deoxyviolacein is assessed after extraction by ethanol and measuring the absorbance at 568nm. Production of lycopene is assessed after extraction by acetone and measuring the absorbance at 475nm. Blue dots represent data obtained in minimal medium with 2% glucose. Orange dots represent data obtained in minimal medium with 2% glucose and with addition of 100µg/mL of ampicillin 30 minutes after infection to eliminate non-infected cells. No significant differences are observed with and without ampicillin.

5.4 Discussion

In this work we explored the use of phages to reprogram dynamically the metabolism of a dense population of cells for bioproduction of value-added compounds. Viruses have been previously called natural born metabolic engineers[172], in the sense that they can take full control over the metabolism of cells and reroute it to produce their own components. Therefore, the question is whether we can reproduce this level of control in a synthetic fashion, and for our own interests. The choice of M13 as chassis for this first attempt is reminiscent of the extreme repurposings that researchers have already undertaken with this phage, from its role in the early days of sequencing and phage display technologies, to new roles as vaccine carrier, therapeutic agent, directed evolution tool, microbattery electrode component and nanomaterial building block[173,174].

The preliminary results we obtained are encouraging since observable production of two compounds have been achieved with an unmodified strain of *E. coli* in minimal medium, without inducers and without antibiotics. In particular, these results did not benefit from any fine-tuning of the strenghts of RBS and promoters, whose impact on performance typically spans several order of magnitudes. Evaluating the full potential of the method will also require advanced fermentation protocols that are developed in the field to reach high performance. For instance, switching from test tubes to bioreactor usually leads to increased performance and gram scale deoxyviolacein production records in *E. coli* were achieved with pulse wise feeding strategies over 200 hours of fedbatch fermentation protocols [167]. Aside from evaluating the impact of combination of modules (silencing, overproduction), further improvement would include removing GFP, useful for during exploratory phase of this project but detrimental for the yield, as well as avoiding unnecessary production of the ampicillin resistance protein when infecting target cells. This could be achieved either by silencing it with an sRNA targetting the phagemid transcript themselves, or by using a promoter active only in the phage-factory cells.

A potential issue that will need to be considered in cost-constrained industrial settings is the cost of producing the phages, even though they will cost much less than most chemical inducers. In our setup we used a volume of infective supernatant equal to 1% of the volume of target cells. While this represents probably a large excess of phage

and can be optimized, the cost of cultivating phage-factory cells might be non-negligible when aiming to infect large fermenters of hundreds of thousands of liters.

A major limit of this approach in the context of general metabolic engineering is the need for suitable phage vectors for the different hosts currently used for bioproduction, apart from *E. coli*. Contrary to bacteria and mammalian cells, where plenty of virus have been studied and engineered, much less is known about Yeast viruses. The rigid yeast cell wall barrier makes extracellular spread of virus difficult, and evolution therefore preferred transmission of viruses by mating which allow an exclusively intracellular route of spread[175]. While adapting our approach to yeast might thus prove intractable, extending it to other bacterial host is certainly feasible. Phagemid systems are not limited to M13, and phagemids able to infect a broad range of bacterial hosts have been developed[176]. Synthetic biologists are also making rapid progress toward engineering modularity in phage scaffolds and already allowed changes in host-specificity[177].

5.5 methods

Chemicals and reagents.

Enzymes for cloning procedures (BsaI, T4 ligase, and DNA polymerase Q5) were purchased from New England BioLabs (Evry, France) and primers were purchased from Eurofins Genomics (Ebersberg, Germany). Pinocembrine, violacein/deoxyviolacein, and lycopene were purchased from Sigma (St. Louis, MO, USA).

Genetic constructs.

BsaI sites were removed from phagemid pVKL001[129] by PCR-based site directed mutagenesis[131]. By default, the sRNA silencing module was targeting chloramphenicol acetyltransferase, whose sequence was absent in all experiments of the present work. Heterologous pathway modules were extracted and flanked with appropriate BsaI restriction sites by PCR from biobrick BBa_K274100 (CrtE, CrtB, CrtI) for the lycopene pathway, biobrick BBa_K274003 (VioA, VioB, VioC, VioE) for the deoxyviolacein pathway, plasmid pQlinkN_hPAL_h4CL_hCHS_hCHIopt from Fehér et al.[50] (PAL, 4CL, CHS, CHI) for the pinocembrine pathway. Original RBS were kept unmodified but promoters were replaced by P_R promoter from phage lambda. All heterologous pathway modules were assembled in phagemids by GoldenGate[72].

Transformation was done in competent *E. coli* Mach1 harboring a pCDF-Duet plasmid containing biobrick BBa_C0051 (cI repressor from phage lambda) to repress pR promoters and avoid counterselection during the cloning phase. Properly assembled clones were isolated by colony PCR and sequenced.

Changing sRNA target and assessment of phenotypic impact of silencing.

In order to reach 100% of successfully changed sRNA target sequences in transformants, BsaI-free version of pVKL001 was amplified by PCR in two fragments with appropriate BsaI restriction sites and the new 24 base pairs target sequence was added directly on the primers. Purification of the fragments after electrophoresis separation on gel was followed by GoldenGate assembly. The ligation product was transformed in two strains of competent *E. coli* Mach1, with and without pCDF-Duet plasmid containing the cI repressor. All the clones tested by colony PCR had properly integrated the new target sequence. Picture where taken with a GBOX-CHEMI-XX6 with appropriate filters for GFP which is encoded by the phagemid.

Phage production and isolation.

We followed the protocol described in Bernheim *et al.* [164] to produce and isolate the recombinant phages. Briefly, the new phage constructs were minipreped from cultures of Mach1 transformants, and cotransformed with plasmid M13K07 into MG1655, F+ harboring cI repressor on pCDF-Duet and plated on LB-agar with ampicillin, kanamycin, and spectinomycin. Colonies were grown in LB medium with antibiotics for 12 hours. Cultures were centrifuged and the supernatants were filtered through 0.2 μ m filters. Supernatants were stored at 4°C before infection.

Infection and fermentation.

E. coli MG1655, F+ were grown overnight from glycerol stocks in M9 minimal medium with 2% glucose. Saturated culture was diluted 1:100 in fresh minimal medium with 2% glucose in flask at 37°C and OD at 600nm was monitored. 2mL aliquots of culture were regularly sampled at different OD and individually mixed in 15mL test tubes with 20uL of filtered supernatant containing the phages of interest. Infected test tubes were incubated at 30°C with vigorous agitation for 24 hours.

Extraction and quantification of metabolites.

For extraction of deoxyviolacein, test tubes containing 2mL cultures were centrifuged and supernatants were removed. Pellets were vortexed until they broke completely, and 1 ml of pure ethanol was added. After 1 hour of shaking 200rpm 37° with intermittent

vortexing, the supernatant was collected after centrifugation and stored. The pellet was subjected to a second extraction step with 1 ml of fresh ethanol. The two supernatant fractions were pooled and deoxyviolacein was quantified through absorbance at 568nm.

For extraction of lycopene, test tubes containing 2mL cultures were centrifuged and supernatants were removed. Lycopene was extracted from pellets for 15min at 55 °C with 1 ml of acetone at 55 °C with intermittent vortexing. After centrifugation, supernatants were quantified through absorbance at 475 nm[169]. The entire extraction process was performed in reduced light conditions to prevent photo-bleaching and degradation.

Conclusion and perspectives – Part II:

The concept of loading phages with synthetic genetic cargo have attracted interests in the field of medicine over the last years, and since the beginning of the projects presented here other groups have used vectorized CRISPR systems to selectively kill bacteria carrying specific genes[178,179]. This technology has already been transferred in the private sector, in biotech companies such as Eligo Biosciences. Aside from the medical applications of gene regulations mediated by phages, no other applied field have made use of this ability yet. At this moment, the very preliminary results described in **Chapter 5** are not sufficient to evaluate precisely the potential of phages to replace inducers in bioproduction protocols. I only show that usual metabolic modifications (down regulation of native metabolic pathways, expression of heterologous ones) can be achieved within the framework of an infective phage. In the coming months, our laboratory will compare the yield, titer and productivity achievable for different molecules by this method, and with other fermentation protocols based on constitutive expression or chemically-induced expression of the same genes, in the same experimental conditions. Combinations of silencing cassettes and overexpression cassettes will also be added on top of heterologous genes. Later on, more subtle modifications could be undertaken to increase performances, such as varying promoters, RBSs, fine-tuning the strenght of repression by sRNAs or even delayed expression of some of the genes, similar to the complex dynamics of gene expression that occurs in real phages. If the results of phage-based metabolic engineering are similar to those obtained through chemical induction, this approach could have industrial relevance. Considering that the lack of inducible promoters in metabolic engineering is crucial to the point that companies such as Synpromics are exclusively dedicated to commercial promoter engineering, the possibility to bypass this bottleneck would certainly be of interest for a large community in the public and private sector.

Bibliography

1. Toman Z, Dambly C, Radman M: **Induction of a stable, heritable epigenetic change by mutagenic carcinogens: a new test system.** *IARC Sci. Publ.* 1980, 1:243–255.
2. Collins JJ, Gardner TS, Cantor CR: **Construction of a genetic toggle switch in *Escherichia coli*.** *Nature* 2000, **403**:339–342.
3. Zhang S: **Y Combinator's Sam Altman Is Bullish on Biotech Startups | WIRED.** 2015, <https://www.wired.com/2015/11/y-combinator-sam-altman-interview-on-investing-in-biotech-startups/>.
4. Cameron DE, Bashor CJ, Collins JJ: **A brief history of synthetic biology.** *Nat. Rev. Microbiol.* 2014, **12**:381–90.
5. Moon TS, Lou C, Tamsir A, Stanton BC, Voigt CA: **Genetic programs constructed from layered logic gates in single cells.** *Nature* 2012, **491**:249–53.
6. Nielsen AAK, Der BS, Shin J, Vaidyanathan P, Paralanov V, Strychalski EA, Ross D, Densmore D, Voigt CA: **Genetic circuit design automation.** *Science (80-.).* 2016, **352**:aac7341-aac7341.
7. Bonnet J, Yin P, Ortiz ME, Subsoontorn P, Endy D: **Amplifying genetic logic gates.** *Science* 2013, **340**:599–603.
8. Siuti P, Yazbek J, Lu TK: **Synthetic circuits integrating logic and memory in living cells.** *Nat. Biotechnol.* 2013, **31**:448–52.
9. Yang L, Nielsen AAK, Fernandez-Rodriguez J, McClune CJ, Laub MT, Lu TK, Voigt CA: **Permanent genetic memory with >1-byte capacity.** *Nat. Methods* 2014, **11**:1261–6.
10. Roquet N, Soleimany AP, Ferris AC, Aaronson S, Lu TK: **Synthetic recombinase-based state machines in living cells.** *Science* 2016, **353**:aad8559.
11. Ganesh I, Ravikumar S, Lee SH, Park SJ, Hong SH: **Engineered fumarate sensing *Escherichia coli* based on novel chimeric two-component system.** *J. Biotechnol.* 2013, **168**:560–6.
12. Ganesh I, Ravikumar S, Yoo I-K, Hong SH: **Construction of malate-sensing *Escherichia coli* by introduction of a novel chimeric two-component system.** *Bioprocess Biosyst. Eng.* 2015, **38**:797–804.

13. Mukherjee K, Bhattacharyya S, Peralta-Yahya P: **GPCR-Based Chemical Biosensors for Medium-Chain Fatty Acids.** *ACS Synth. Biol.* 2015, **4**:1261–9.
14. Zhou L-B, Zeng A-P: **Exploring Lysine Riboswitch for Metabolic Flux Control and Improvement of l -Lysine Synthesis in *Corynebacterium glutamicum*.** *ACS Synth. Biol.* 2015, **4**:729–734.
15. Tuerk C, Gold L: **Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase.** *Science* 1990, **249**:505–10.
16. Ellington AD, Szostak JW: **In vitro selection of RNA molecules that bind specific ligands.** *Nature* 1990, **346**:818–22.
17. McKeague M, Wong RS, Smolke CD: **Opportunities in the design and application of RNA for gene expression control.** *Nucleic Acids Res.* 2016, **44**:2987–99.
18. Liu D, Evans T, Zhang F: **Applications and advances of metabolite biosensors for metabolic engineering.** *Metab. Eng.* 2015, **31**:15–22.
19. Oakes BL, Nadler DC, Flamholz A, Fellmann C, Staahl BT, Doudna JA, Savage DF: **Profiling of engineering hotspots identifies an allosteric CRISPR-Cas9 switch.** *Nat. Biotechnol.* 2016, **34**:646–51.
20. Liu KI, Ramli MN Bin, Woo CWA, Wang Y, Zhao T, Zhang X, Yim GRD, Chong BY, Gowher A, Chua MZH, et al.: **A chemical-inducible CRISPR-Cas9 system for rapid control of genome editing.** *Nat. Chem. Biol.* 2016, doi:10.1038/nchembio.2179.
21. Davis KM, Pattanayak V, Thompson DB, Zuris JA, Liu DR: **Small molecule-triggered Cas9 protein with improved genome-editing specificity.** *Nat. Chem. Biol.* 2015, **11**:316–8.
22. Zetsche B, Volz SE, Zhang F: **A split-Cas9 architecture for inducible genome editing and transcription modulation.** *Nat. Biotechnol.* 2015, **33**:139–42.
23. Qi LS, Larson MH, Gilbert LA, Doudna JA, Weissman JS, Arkin AP, Lim WA: **Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression.** *Cell* 2013, **152**:1173–83.
24. Gilbert LA, Larson MH, Morsut L, Liu Z, Brar GA, Torres SE, Stern-Ginossar N, Brandman O, Whitehead EH, Doudna JA, et al.: **CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes.** *Cell* 2013, **154**:442–51.
25. Feng J, Jester BW, Tinberg CE, Mandell DJ, Antunes MS, Chari R, Morey KJ, Rios X, Medford JI, Church GM, et al.: **A general strategy to construct small molecule**

biosensors in eukaryotes. *Elife* 2015, **4**:4004–4009.

26. Brophy JAN, Voigt CA: **Principles of genetic circuit design.** *Nat. Methods* 2014, **11**:508–20.
27. Mishra D, Rivera PM, Lin A, Del Vecchio D, Weiss R: **A load driver device for engineering modularity in biological networks.** *Nat. Biotechnol.* 2014, **32**:1268–75.
28. Farzadfard F, Lu TK: **Genomically encoded analog memory with precise in vivo DNA writing in living cell populations.** *Science (80-.).* 2014, **346**:1256272–1256272.
29. Bonnet J, Yin P, Ortiz ME, Subsoontorn P, Endy D: **Amplifying genetic logic gates.** *Science* 2013, **340**:599–603.
30. Prindle A, Selimkhanov J, Li H, Razinkov I, Tsimring LS, Hasty J: **Rapid and tunable post-translational coupling of genetic circuits.** *Nature* 2014, **508**:387–91.
31. Prochazka L, Angelici B, Haefliger B, Benenson Y: **Highly modular bow-tie gene circuits with programmable dynamic behaviour.** *Nat. Commun.* 2014, **5**:4729.
32. Courbet A, Endy D, Renard E, Molina F, Bonnet J: **Detection of pathological biomarkers in human clinical samples via amplifying genetic switches and logic gates.** 2015, **7**:1–11.
33. Rogers JK, Guzman CD, Taylor ND, Raman S, Anderson K, Church GM: **Synthetic biosensors for precise gene control and real-time monitoring of metabolites.** *Nucleic Acids Res.* 2015, doi:10.1093/nar/gkv616.
34. Bayer TS, Smolke CD: **Programmable ligand-controlled riboregulators of eukaryotic gene expression.** *Nat. Biotechnol.* 2005, **23**:337–43.
35. Taylor ND, Garruss AS, Moretti R, Chan S, Arbing MA, Cascio D, Rogers JK, Isaacs FJ, Kosuri S, Baker D, et al.: **Engineering an allosteric transcription factor to respond to new ligands.** *Nat. Methods* 2015, **13**:177–83.
36. Silva-Rocha R, de Lorenzo V: **Engineering multicellular logic in bacteria with metabolic wires.** *ACS Synth. Biol.* 2014, **3**:204–9.
37. Xue H, Shi H, Yu Z, He S, Liu S, Hou Y, Pan X, Wang H, Zheng P, Cui C, et al.: **Design, construction, and characterization of a set of biosensors for aromatic compounds.** *ACS Synth. Biol.* 2014, **3**:1011–4.
38. Bulter T, Lee S-G, Wong WW, Fung E, Connor MR, Liao JC: **Design of artificial cell-cell communication using gene and metabolic networks.** *Proc. Natl. Acad. Sci. U. S. A.* 2004, **101**:2299–2304.

39. Beal J, Lu T, Weiss R: **Automatic compilation from high-level biologically-oriented programming language to genetic regulatory networks.** *PLoS One* 2011, **6**:e22490.
40. Clancy K, Voigt CA: **Programming cells: towards an automated “Genetic Compiler”.** *Curr. Opin. Biotechnol.* 2010, **21**:572–81.
41. Medema MH, van Raaphorst R, Takano E, Breitling R: **Computational tools for the synthetic design of biochemical pathways.** *Nat. Rev. Microbiol.* 2012, **10**:191–202.
42. Carbonell P, Parutto P, Herisson J, Pandit SB, Faulon J-L: **XTMS: pathway design in an eXTended metabolic space.** *Nucleic Acids Res.* 2014, **42**:W389-94.
43. Law V, Knox C, Djoumbou Y, Jewison T, Guo AC, Liu Y, Maciejewski A, Arndt D, Wilson M, Neveu V, et al.: **DrugBank 4.0: shedding new light on drug metabolism.** *Nucleic Acids Res.* 2014, **42**:D1091-7.
44. Wishart DS, Jewison T, Guo AC, Wilson M, Knox C, Liu Y, Djoumbou Y, Mandal R, Aziat F, Dong E, et al.: **HMDB 3.0--The Human Metabolome Database in 2013.** *Nucleic Acids Res.* 2013, **41**:D801-7.
45. Kavlock RJ, Austin CP, Tice RR: **Toxicity testing in the 21st century: implications for human health risk assessment.** *Risk Anal.* 2009, **29**:485-7-7.
46. Chang A, Schomburg I, Placzek S, Jeske L, Ulbrich M, Xiao M, Sensen CW, Schomburg D: **BRENDA in 2015: exciting developments in its 25th year of existence.** *Nucleic Acids Res.* 2015, **43**:D439-46.
47. Caspi R, Altman T, Billington R, Dreher K, Foerster H, Fulcher CA, Holland TA, Keseler IM, Kothari A, Kubo A, et al.: **The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases.** *Nucleic Acids Res.* 2014, **42**:D459-71.
48. Morgat A, Axelsen KB, Lombardot T, Alcántara R, Aimo L, Zerara M, Niknejad A, Belda E, Hyka-Nouspikel N, Coudert E, et al.: **Updates in Rhea--a manually curated resource of biochemical reactions.** *Nucleic Acids Res.* 2015, **43**:D459-64.
49. Faulon J-L, Misra M, Martin S, Sale K, Sapra R: **Genome scale enzyme-metabolite and drug-target interaction predictions using the signature molecular descriptor.** *Bioinformatics* 2008, **24**:225–33.
50. Fehér T, Planson A-G, Carbonell P, Fernández-Castané A, Grigoras I, Dariy E, Perret A, Faulon J-L: **Validation of RetroPath, a computer-aided design tool for metabolic pathway engineering.** *Biotechnol. J.* 2014, **9**:1446–57.
51. Carbajosa G, Trigo A, Valencia A, Cases I: **Bionemo: molecular information on**

biodegradation metabolism. *Nucleic Acids Res.* 2009, **37**:D598-602.

52. Cipriano MJ, Novichkov PN, Kazakov AE, Rodionov DA, Arkin AP, Gelfand MS, Dubchak I: **RegTransBase--a database of regulatory sequences and interactions based on literature: a resource for investigating transcriptional regulation in prokaryotes.** *BMC Genomics* 2013, **14**:213.
53. Salgado H, Peralta-Gil M, Gama-Castro S, Santos-Zavaleta A, Muñiz-Rascado L, García-Sotelo JS, Weiss V, Solano-Lira H, Martínez-Flores I, Medina-Rivera A, et al.: **RegulonDB v8.0: omics data sets, evolutionary conservation, regulatory phrases, cross-validated gold standards and more.** *Nucleic Acids Res.* 2013, **41**:D203-13.
54. Novichkov PS, Kazakov AE, Ravcheev DA, Leyn SA, Kovaleva GY, Sutormin RA, Kazanov MD, Riehl W, Arkin AP, Dubchak I, et al.: **RegPrecise 3.0--a resource for genome-scale exploration of transcriptional regulation in bacteria.** *BMC Genomics* 2013, **14**:745.
55. Kojima R, Aubel D, Fussenegger M: **Novel theranostic agents for next-generation personalized medicine: small molecules, nanoparticles, and engineered mammalian cells.** *Curr. Opin. Chem. Biol.* 2015, **28**:29–38.
56. Galanie S, Thodey K, Trenchard IJ, Filsinger Interrante M, Smolke CD: **Complete biosynthesis of opioids in yeast.** *Science (80-.).* 2015, **349**:1095–100.
57. Rogers JK, Church GM: **Genetically encoded sensors enable real-time observation of metabolite production.** *Proc. Natl. Acad. Sci. U. S. A.* 2016, **113**:2388–2393.
58. Roodveldt C, Tawfik DS: **Directed evolution of phosphotriesterase from *Pseudomonas diminuta* for heterologous expression in *Escherichia coli* results in stabilization of the metal-free state.** *Protein Eng. Des. Sel.* 2005, **18**:51–8.
59. Briseño-Roa L, Timperley CM, Griffiths AD, Fersht AR: **Phosphotriesterase variants with high methylphosphonate activity and strong negative trade-off against phosphotriesters.** *Protein Eng. Des. Sel.* 2011, **24**:151–9.
60. Steigedal M, Valla S: **The *Acinetobacter sp.* *chnB* promoter together with its cognate positive regulator ChnR is an attractive new candidate for metabolic engineering applications in bacteria.** *Metab. Eng.* 2008, **10**:121–9.
61. Lentini R, Santero SP, Chizzolini F, Cecchi D, Fontana J, Marchiorretto M, Del Bianco C, Terrell JL, Spencer AC, Martini L, et al.: **Integrating artificial with natural cells to translate chemical messages that direct *E. coli* behaviour.** *Nat. Commun.* 2014, **5**:4012.
62. Elani Y, Law R V, Ces O: **Vesicle-based artificial cells as chemical microreactors**

with spatially segregated reaction pathways. *Nat. Commun.* 2014, 5:5305.

63. Pflüger-Grau K, de Lorenzo V: **From the phosphoenolpyruvate phosphotransferase system to selfish metabolism: a story retraced in *Pseudomonas putida*.** *FEMS Microbiol. Lett.* 2014, **356**:144–53.
64. Chubukov V, Gerosa L, Kochanowski K, Sauer U: **Coordination of microbial metabolism.** *Nat. Rev. Microbiol.* 2014, **12**:327–40.
65. Wheatley RW, Lo S, Jancewicz LJ, Dugdale ML, Huber RE: **Structural explanation for allolactose (*lac* operon inducer) synthesis by *lacZ* β -galactosidase and the evolutionary relationship between allolactose synthesis and the *lac* repressor.** *J. Biol. Chem.* 2013, **288**:12993–3005.
66. Faulon J-L, Visco DP, Pophale RS: **The signature molecular descriptor. 1. Using extended valence sequences in QSAR and QSPR studies.** *J. Chem. Inf. Comput. Sci.* **43**:707–20.
67. Jaccard P: **Distribution de la flore alpine dans le Bassin de Dranses et dans quelques regions voisines.** *Bull. la Soc. Vaudoise des Sci. Nat.* 1901, **37**:241–272.
68. Hagberg AA, Schult DA, Swart PJ: **Exploring network structure, dynamics, and function using {NetworkX}.** In *Proceedings of the 7th Python in Science Conference (SciPy2008)*. . 2008:11–15.
69. Grote A, Hiller K, Scheer M, Munch R, Nortemann B, Hempel DC, Jahn D: **JCat: a novel tool to adapt codon usage of a target gene to its potential expression host.** *Nucleic Acids Res.* 2005, **33**:W526–W531.
70. Miyauchi K, Lee H, Fukuda M, Takagi M, Nagata Y: **Cloning and Characterization of *linR*, Involved in Regulation of the Downstream Pathway for γ - Hexachlorocyclohexane Degradation in *Sphingomonas paucimobilis* UT26.** *Appl. Environ. Microbiol.* 2002, doi:10.1128/AEM.68.4.1803.
71. Min J, Zhang J-J, Zhou N-Y: **The para-Nitrophenol Catabolic Gene Cluster is Responsible for 2-Chloro-4-nitrophenol Degradation in *Burkholderia* sp. Strain SJ98.** *Appl. Environ. Microbiol.* 2014, **80**:6212–6222.
72. Engler C, Kandzia R, Marillonnet S: **A one pot, one step, precision cloning method with high throughput capability.** *PLoS One* 2008, **3**:e3647.
73. Baba T, Ara T, Hasegawa M, Takai Y, Okumura Y, Baba M, Datsenko KA, Tomita M, Wanner BL, Mori H: **Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection.** *Mol. Syst. Biol.* 2006, **2**:2006.0008.

74. Fernandez-López R, Ruiz R, de la Cruz F, Moncalián G: **Transcription factor-based biosensors enlightened by the analyte.** *Front. Microbiol.* 2015, **6**:648.
75. Rössger K, Charpin-El-Hamri G, Fussenegger M: **A closed-loop synthetic gene circuit for the treatment of diet-induced obesity in mice.** *Nat. Commun.* 2013, **4**.
76. Kemmer C, Gitzinger M, Daoud-El Baba M, Djonov V, Stelling J, Fussenegger M: **Self-sufficient control of urate homeostasis in mice by a synthetic circuit.** *Nat. Biotechnol.* 2010, **28**:355–60.
77. Dietrich J a, Shis DL, Alikhani A, Keasling JD: **Transcription factor-based screens and synthetic selections for microbial small-molecule biosynthesis.** *ACS Synth. Biol.* 2013, **2**:47–58.
78. Raman S, Rogers JK, Taylor ND, Church GM: **Evolution-guided optimization of biosynthetic pathways.** *Proc. Natl. Acad. Sci.* 2014, doi:10.1073/pnas.1409523111.
79. Xu P, Li L, Zhang F, Stephanopoulos G, Koffas M: **Improving fatty acids production by engineering dynamic pathway regulation and metabolic control.** *Proc. Natl. Acad. Sci.* 2014, doi:10.1073/pnas.1406401111.
80. Mahr R, Frunzke J: **Transcription factor-based biosensors in biotechnology: current state and future prospects.** *Appl. Microbiol. Biotechnol.* 2015, doi:10.1007/s00253-015-7090-3.
81. Gama-Castro S, Salgado H, Santos-Zavaleta A, Ledezma-Tejeida D, Muñoz-Rascado L, García-Sotelo JS, Alquicira-Hernández K, Martínez-Flores I, Pannier L, Castro-Mondragón JA, et al.: **RegulonDB version 9.0: high-level integration of gene regulation, coexpression, motif clustering and beyond.** *Nucleic Acids Res.* 2015, **44**:gkv1156.
82. de Lorenzo V, Silva-Rocha R, Carbajosa G, Galvão TC CI: **Sensing xenobiotic compounds: lessons from bacteria that face pollutants in the environment.** In *Sensory Mechanisms in Bacteria*. . Horizon Scientific Press; 2010:81–105.
83. Ravcheev DA, Khoroshkin MS, Laikova ON, Tsoy O V, Sernova N V, Petrova SA, Rakhmaninova AB, Novichkov PS, Gelfand MS, Rodionov DA: **Comparative genomics and evolution of regulons of the LacI-family transcription factors.** *Front. Microbiol.* 2014, **5**:294.
84. Kazanov MD, Li X, Gelfand MS, Osterman AL, Rodionov DA: **Functional diversification of ROK-family transcriptional regulators of sugar catabolism in the Thermotogae phylum.** *Nucleic Acids Res.* 2013, **41**:790–803.
85. Ahn SK, Cuthbertson L, Nodwell JR: **Genome context as a predictive tool for**

identifying regulatory targets of the TetR family transcriptional regulators. *PLoS One* 2012, 7:e50562.

86. Cuthbertson L, Ahn SK, Nodwell JR: **Deglycosylation as a mechanism of inducible antibiotic resistance revealed using a global relational tree for one-component regulators.** *Chem. Biol.* 2013, 20:232–40.
87. Uchiyama T, Miyazaki K: **Metagenomic screening for aromatic compound-responsive transcriptional regulators.** *PLoS One* 2013, 8:e75795.
88. Tang S-Y, Cirino PC: **Design and application of a mevalonate-responsive regulatory protein.** *Angew. Chem. Int. Ed. Engl.* 2011, 50:1084–6.
89. Tang S-Y, Qian S, Akinterinwa O, Frei CS, Gredell JA, Cirino PC: **Screening for enhanced triacetic acid lactone production by recombinant *Escherichia coli* expressing a designed triacetic acid lactone reporter.** *J. Am. Chem. Soc.* 2013, 135:10099–103.
90. Chen W, Zhang S, Jiang P, Yao J, He Y, Chen L, Gui X, Dong Z, Tang S-Y: **Design of an ectoine-responsive AraC mutant and its application in metabolic engineering of ectoine biosynthesis.** *Metab. Eng.* 2015, 30:149–55.
91. Raman S, Taylor N, Genuth N, Fields S, Church GM: **Engineering allostery.** *Trends Genet.* 2014, 30:521–8.
92. Shis DL, Hussain F, Meinhardt S, Swint-Kruse L, Bennett MR: **Modular, multi-input transcriptional logic gating with orthogonal LacI/GalR family chimeras.** *ACS Synth. Biol.* 2014, 3:645–51.
93. Meinhardt S, Manley MW, Becker NA, Hessman JA, Maher LJ, Swint-Kruse L: **Novel insights from hybrid LacI/GalR proteins: family-wide functional attributes and biologically significant variation in transcription repression.** *Nucleic Acids Res.* 2012, 40:11139–54.
94. Sousa FL, Parente DJ, Shis DL, Hessman JA, Chazelle A, Bennett MR, Teichmann SA, Swint-Kruse L: **AlloRep: A Repository of Sequence, Structural and Mutagenesis Data for the LacI/GalR Transcription Regulators.** *J. Mol. Biol.* 2016, 428:671–8.
95. Jha RK, Chakraborti S, Kern TL, Fox DT, Strauss CEM: **Rosetta comparative modeling for library design: Engineering alternative inducer specificity in a transcription factor.** *Proteins* 2015, 83:1327–40.
96. Shin H, Cho B-K: **Rational Protein Engineering Guided by Deep Mutational Scanning.** *Int. J. Mol. Sci.* 2015, 16:23094–110.

97. Fowler DM, Fields S: **Deep mutational scanning: a new style of protein science.** *Nat. Methods* 2014, **11**:801–7.
98. Santos ELC de los, Meyerowitz JT, Mayo SL, Murray RM: **Engineering Transcriptional Regulator Effector Specificity Using Computational Design and In Vitro Rapid Prototyping: Developing a Vanillin Sensor.** [date unknown], [no volume].
99. Mohn WW, Garmendia J, Galvao TC, de Lorenzo V: **Surveying biotransformations with à la carte genetic traps: translating dehydrochlorination of lindane (gamma-hexachlorocyclohexane) into lacZ-based phenotypes.** *Environ. Microbiol.* 2006, **8**:546–55.
100. Libis V, Delépine B, Faulon J-L: **Expanding Biosensing Abilities through Computer-Aided Design of Metabolic Pathways.** *ACS Synth. Biol.* 2016, doi:10.1021/acssynbio.5b00225.
101. Delépine B, Libis V, Carbonell P, Faulon J-L: **SensiPath: computer-aided design of sensing-enabling metabolic pathways.** *Nucleic Acids Res.* 2016, doi:10.1093/nar/gkw305.
102. Juárez JF, Liu H, Zamarro MT, McMahon S, Liu H, Naismith JH, Eberlein C, Boll M, Carmona M, Díaz E: **Unraveling the specific regulation of the central pathway for anaerobic degradation of 3-methylbenzoate.** *J. Biol. Chem.* 2015, **290**:12165–83.
103. Savir Y, Tu BP, Springer M: **Competitive inhibition can linearize dose-response and generate a linear rectifier.** *Cell Syst.* 2015, **1**:238–245.
104. Ho NAT, Dawes SS, Crowe AM, Casabon I, Gao C, Kendall SL, Baker EN, Eltis LD, Lott JS: **The Structure of the Transcriptional Repressor KstR in Complex with CoA Thioester Cholesterol Metabolites Sheds Light on the Regulation of Cholesterol Catabolism in *Mycobacterium tuberculosis*.** *J. Biol. Chem.* 2016, **291**:7256–66.
105. García-Fernández E, Medrano FJ, Galán B, García JL: **Deciphering the transcriptional regulation of cholesterol catabolic pathway in mycobacteria: identification of the inducer of KstR repressor.** *J. Biol. Chem.* 2014, **289**:17576–88.
106. Jobe A, Bourgeois S: **lac Repressor-operator interaction. VI. The natural inducer of the lac operon.** *J. Mol. Biol.* 1972, **69**:397–408.
107. Coschigano PW, Bishop BJ: **Role of benzylsuccinate in the induction of the *tutE tutFDGH* gene complex of *T. aromatica* strain T1.** *FEMS Microbiol. Lett.* 2004, **231**:261–266.
108. Shimizu T, Nakamura A: **Characterization of LgnR, an IclR family transcriptional**

regulator involved in the regulation of L-gluconate catabolic genes in *Paracoccus* sp. 43P. *Microbiology* 2014, **160**:623–34.

109. Kasai D, Kamimura N, Tani K, Umeda S, Abe T, Fukuda M, Masai E: **Characterization of FerC, a MarR-type transcriptional regulator, involved in transcriptional regulation of the ferulate catabolic operon in *Sphingobium* sp. strain SYK-6.** *FEMS Microbiol. Lett.* 2012, **332**:68–75.
110. Hirakawa H, Schaefer AL, Greenberg EP, Harwood CS: **Anaerobic p-coumarate degradation by *Rhodopseudomonas palustris* and identification of CouR, a MarR repressor protein that binds p-coumaroyl coenzyme A.** *J. Bacteriol.* 2012, **194**:1960–7.
111. Chen D-W, Zhang Y, Jiang C-Y, Liu S-J: **Benzoate metabolism intermediate benzoyl coenzyme A affects gentisate pathway regulation in *Comamonas testosteroni*.** *Appl. Environ. Microbiol.* 2014, **80**:4051–62.
112. Sakamoto K, Agari Y, Kuramitsu S, Shinkai A: **Phenylacetyl coenzyme A is an effector molecule of the TetR family transcriptional repressor PaaR from *Thermus thermophilus* HB8.** *J. Bacteriol.* 2011, **193**:4388–95.
113. Choi S, Rha E, Lee SJ, Kim H, Kwon K, Jeong Y, Rhee YH, Song JJ, Kim H, Lee S: **Toward a Generalized and High-throughput Enzyme Screening System Based on Artificial Genetic Circuits.** *ACS Synth. Biol.* 2014, doi: 10.1021/sb400112u.
114. Jeong YS, Choi SL, Kyeong HH, Kim JH, Kim EJ, Pan JG, Rha E, Song JJ, Lee SG, Kim HS: **High-throughput screening system based on phenolics-responsive transcription activator for directed evolution of organophosphate-degrading enzymes.** *Protein Eng. Des. Sel.* 2012, **25**:725–731.
115. 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.* 2010, **76**:7029–35.
116. Siedler S, Stahlhut SG, Malla S, Maury J, Neves AR: **Novel biosensors based on flavonoid-responsive transcriptional regulators introduced into *Escherichia coli*.** *Metab. Eng.* 2014, **21**:2–8.
117. Chong H, Ching CB: **Development of Colorimetric-Based Whole-Cell Biosensor for Organophosphorus Compounds by Engineering Transcription Regulator DmpR.** *ACS Synth. Biol.* 2016, doi:10.1021/acssynbio.6b00061.
118. Eames M, Kortemme T: **Cost-Benefit Tradeoffs in Engineered *lac* Operons.** *Science (80-.).* 2012, **336**.

119. Osumi T, Saier MH: **Regulation of lactose permease activity by the phosphoenolpyruvate:sugar phosphotransferase system: evidence for direct binding of the glucose-specific enzyme III to the lactose permease.** *Proc. Natl. Acad. Sci. U. S. A.* 1982, **79**:1457–61.
120. Ohkumo T, Masutani C, Eki T, Hanaoka F: **Use of RNAi in *C. elegans*.** *Methods Mol. Biol.* 2008, **442**:129–37.
121. Fire A, Xu S, Montgomery MK, Kostas SA, Driver SE, Mello CC: **Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*.** *Nature* 1998, **391**:806–11.
122. Iordanou E, Chandran RR, Blackstone N, Jiang L: **RNAi interference by dsRNA injection into *Drosophila* embryos.** *J. Vis. Exp.* 2011, doi:10.3791/2477.
123. Ramadan N, Flockhart I, Booker M, Perrimon N, Mathey-Prevot B: **Design and implementation of high-throughput RNAi screens in cultured *Drosophila* cells.** *Nat. Protoc.* 2007, **2**:2245–64.
124. Tsong TY: **Electroporation of cell membranes.** *Biophys. J.* 1991, **60**:297–306.
125. Kim WJ, Chang C-W, Lee M, Kim SW: **Efficient siRNA delivery using water soluble lipopolymer for anti-angiogenic gene therapy.** *J. Control. Release* 2007, **118**:357–63.
126. Shi Y: **Mammalian RNAi for the masses.** *Trends Genet.* 2003, **19**:9–12.
127. Na D, Yoo SM, Chung H, Park H, Park JH, Lee SY: **Metabolic engineering of *Escherichia coli* using synthetic small regulatory RNAs.** *Nat. Biotechnol.* 2013, **31**:170–4.
128. Ceroni F, Algar R, Stan G-B, Ellis T: **Quantifying cellular capacity identifies gene expression designs with reduced burden.** *Nat. Methods* 2015, **12**:415–8.
129. Libis VK, Bernheim AG, Basier C, Jaramillo-Riveri S, Deyell M, Aghoghogbe I, Atanaskovic I, Bencherif AC, Benony M, Koutsoubelis N, et al.: **Silencing of Antibiotic Resistance in *E. coli* with Engineered Phage Bearing Small Regulatory RNAs.** *ACS Synth. Biol.* 2014, **3**:1003–1006.
130. Ortiz ME, Endy D: **Engineered cell-cell communication via DNA messaging.** *J. Biol. Eng.* 2012, **6**:16.
131. Edelheit O, Hanukoglu A, Hanukoglu I: **Simple and efficient site-directed mutagenesis using two single-primer reactions in parallel to generate mutants for protein structure-function studies.** *BMC Biotechnol.* 2009, **9**:61.

132. Chung CT, Miller RH: **Preparation and storage of competent *Escherichia coli* cells.** *Methods Enzymol.* 1993, **218**:621–7.
133. Phornphisutthimas S, Thamchaipenet A, Panijpan B: **Conjugation in *Escherichia coli*.** *Biochem. Mol. Biol. Educ.* 2007, **35**:440–445.
134. Sezonov G, Joseleau-Petit D, D'Ari R: ***Escherichia coli* physiology in Luria-Bertani broth.** *J. Bacteriol.* 2007, **189**:8746–9.
135. Mittal V: **Improving the efficiency of RNA interference in mammals.** *Nat. Rev. Genet.* 2004, **5**:355–65.
136. Lu TK, Collins JJ: **Dispersing biofilms with engineered enzymatic bacteriophage.** *Proc. Natl. Acad. Sci. U. S. A.* 2007, **104**:11197–202.
137. Yosef I, Manor M, Kiro R, Qimron U: **Temperate and lytic bacteriophages programmed to sensitize and kill antibiotic-resistant bacteria.** *Proc. Natl. Acad. Sci. U. S. A.* 2015, **112**:7267–72.
138. Karlsson F, Malmborg-Hager A-C, Albrekt A-S, Borrebaeck CAK: **Genome-wide comparison of phage M13-infected vs. uninfected *Escherichia coli*.** *Can. J. Microbiol.* 2005, **51**:29–35.
139. Senthil-Kumar M, Mysore KS: **Caveat of RNAi in plants: the off-target effect.** *Methods Mol. Biol.* 2011, **744**:13–25.
140. Jackson AL, Linsley PS: **Recognizing and avoiding siRNA off-target effects for target identification and therapeutic application.** *Nat. Rev. Drug Discov.* 2010, **9**:57–67.
141. Cho SW, Kim S, Kim Y, Kweon J, Kim HS, Bae S, Kim J-S: **Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases.** *Genome Res.* 2014, **24**:132–41.
142. Hagens S, Bläsi U: **Genetically modified filamentous phage as bactericidal agents: a pilot study.** *Lett. Appl. Microbiol.* 2003, **37**:318–23.
143. Kasman LM, Kasman A, Westwater C, Dolan J, Schmidt MG, Norris JS: **Overcoming the phage replication threshold: a mathematical model with implications for phage therapy.** *J. Virol.* 2002, **76**:5557–64.
144. Chasteen L, Ayriss J, Pavlik P, Bradbury ARM: **Eliminating helper phage from phage display.** *Nucleic Acids Res.* 2006, **34**:e145.
145. Lu TK, Collins JJ: **Engineered bacteriophage targeting gene networks as adjuvants for antibiotic therapy.** *Proc. Natl. Acad. Sci. U. S. A.* 2009, **106**:4629–34.

146. Bikard D, Jiang W, Samai P, Hochschild A, Zhang F, Marraffini LA: **Programmable repression and activation of bacterial gene expression using an engineered CRISPR-Cas system.** *Nucleic Acids Res.* 2013, **41**:7429–37.
147. Röthlisberger D, Khersonsky O, Wollacott AM, Jiang L, DeChancie J, Betker J, Gallaher JL, Althoff EA, Zanghellini A, Dym O, et al.: **Kemp elimination catalysts by computational enzyme design.** *Nature* 2008, **453**:190–5.
148. Siegel JB, Zanghellini A, Lovick HM, Kiss G, Lambert AR, St Clair JL, Gallaher JL, Hilvert D, Gelb MH, Stoddard BL, et al.: **Computational design of an enzyme catalyst for a stereoselective bimolecular Diels-Alder reaction.** *Science* 2010, **329**:309–13.
149. Wang HH, Isaacs FJ, Carr PA, Sun ZZ, Xu G, Forest CR, Church GM: **Programming cells by multiplex genome engineering and accelerated evolution.** *Nature* 2009, **460**:894–8.
150. Lawrence AD, Frank S, Newnham S, Lee MJ, Brown IR, Xue W-F, Rowe ML, Mulvihill DP, Prentice MB, Howard MJ, et al.: **Solution Structure of a Bacterial Microcompartment Targeting Peptide and Its Application in the Construction of an Ethanol Bioreactor.** *ACS Synth. Biol.* 2014, **3**:454–465.
151. Avalos JL, Fink GR, Stephanopoulos G: **Compartmentalization of metabolic pathways in yeast mitochondria improves the production of branched-chain alcohols.** *Nat. Biotechnol.* 2013, **31**:335–341.
152. Dueber JE, Wu GC, Malmirchegini GR, Moon TS, Petzold CJ, Ullal A V, Prather KLJ, Keasling JD: **Synthetic protein scaffolds provide modular control over metabolic flux.** *Nat. Biotechnol.* 2009, **27**:753–9.
153. Myhrvold C, Polka JK, Silver PA: **Synthetic Lipid-Containing Scaffolds Enhance Production by Colocalizing Enzymes.** *ACS Synth. Biol.* 2016, doi:10.1021/acssynbio.6b00141.
154. Delebecque CJ, Lindner AB, Silver PA, Aldaye FA: **Organization of intracellular reactions with rationally designed RNA assemblies.** *Science* 2011, **333**:470–4.
155. Zhang F, Carothers JM, Keasling JD: **Design of a dynamic sensor-regulator system for production of chemicals and fuels derived from fatty acids.** *Nat. Biotechnol.* 2012, **30**:354–9.
156. Sandoval CM, Ayson M, Moss N, Lieu B, Jackson P, Gaucher SP, Horning T, Dahl RH, Denery JR, Abbott DA, et al.: **Use of pantothenate as a metabolic switch increases the genetic stability of farnesene producing *Saccharomyces cerevisiae*.** *Metab. Eng.* 2014, **25**:215–26.

157. Nozzi NE, Atsumi S: **Genome Engineering of the 2,3-Butanediol Biosynthetic Pathway for Tight Regulation in Cyanobacteria.** *ACS Synth. Biol.* 2015, 4:1197–1204.
158. Jones JA, Vernacchio VR, Lachance DM, Lebovich M, Fu L, Shirke AN, Schultz VL, Cress B, Linhardt RJ, Koffas MAG: **ePathOptimize: A Combinatorial Approach for Transcriptional Balancing of Metabolic Pathways.** *Sci. Rep.* 2015, 5:11301.
159. Wu J, Zhang X, Dong M, Zhou J: **Stepwise modular pathway engineering of *Escherichia coli* for efficient one-step production of (2S)-pinocembrin.** *J. Biotechnol.* 2016, 231:183–92.
160. Wynn JP, Hanchar R, Kleff S, Senyk D, Tiedje T: **Biobased Technology Commercialization: The Importance of Lab to Pilot Scale-up.** In *Metabolic Engineering for Bioprocess Commercialization.* . Springer International Publishing; 2016:101–119.
161. Paddon CJ, Westfall PJ, Pitera DJ, Benjamin K, Fisher K, McPhee D, Leavell MD, Tai a, Main a, Eng D, et al.: **High-level semi-synthetic production of the potent antimalarial artemisinin.** *Nature* 2013, 496:528–32.
162. Libis V, Delépine B, Faulon J-L: **Sensing new chemicals with bacterial transcription factors.** *Curr. Opin. Microbiol.* 2016, 33:105–112.
163. Messing J: **M13 cloning vehicles. Their contribution to DNA sequencing.** *Methods Mol. Biol.* 1993, 23:9–22.
164. Bernheim AG, Libis VK, Lindner AB, Wintermute EH: **Phage-mediated Delivery of Targeted sRNA Constructs to Knock Down Gene Expression in *E. coli*.** *J. Vis. Exp.* 2016, doi:10.3791/53618.
165. Wu J, Du G, Chen J, Zhou J: **Enhancing flavonoid production by systematically tuning the central metabolic pathways based on a CRISPR interference system in *Escherichia coli*.** *Sci. Rep.* 2015, 5:13477.
166. Gu P, Yang F, Kang J, Wang Q, Qi Q, Bongaerts J, Kramer M, Muller U, Raeven L, Wubbolts M, et al.: **One-step of tryptophan attenuator inactivation and promoter swapping to improve the production of L-tryptophan in *Escherichia coli*.** *Microb. Cell Fact.* 2012, 11:30.
167. Rodrigues AL, Becker J, de Souza Lima AO, Porto LM, Wittmann C: **Systems metabolic engineering of *Escherichia coli* for gram scale production of the antitumor drug deoxyviolacein from glycerol.** *Biotechnol. Bioeng.* 2014, 111:2280–2289.

168. Fang M-Y, Zhang C, Yang S, Cui J-Y, Jiang P-X, Lou K, Wachi M, Xing X-H, Ryan K, Drennan C, et al.: **High crude violacein production from glucose by *Escherichia coli* engineered with interactive control of tryptophan pathway and violacein biosynthetic pathway.** *Microb. Cell Fact.* 2015, 14:8.
169. Alper H, Jin Y-S, Moxley JF, Stephanopoulos G: **Identifying gene targets for the metabolic engineering of lycopene biosynthesis in *Escherichia coli*.** *Metab. Eng.* 2005, 7:155–164.
170. Izard J, Gomez Balderas CDC, Ropers D, Lacour S, Song X, Yang Y, Lindner AB, Geiselmann J, de Jong H, Alper H, et al.: **A synthetic growth switch based on controlled expression of RNA polymerase.** *Mol. Syst. Biol.* 2015, 11:840.
171. Jiang P, Wang H, Zhang C, Lou K, Xing X-H: **Reconstruction of the violacein biosynthetic pathway from *Duganella sp.* B2 in different heterologous hosts.** *Appl. Microbiol. Biotechnol.* 2010, 86:1077–1088.
172. Maynard ND, Gutschow M V, Birch EW, Covert MW: **The virus as metabolic engineer.** *Biotechnol. J.* 2010, 5:686–94.
173. Henry KA, Arbabi-Ghahroudi M, Scott JK: **Beyond phage display: non-traditional applications of the filamentous bacteriophage as a vaccine carrier, therapeutic biologic, and bioconjugation scaffold.** *Front. Microbiol.* 2015, 6:755.
174. Pires DP, Cleto S, Sillankorva S, Azeredo J, Lu TK: **Genetically Engineered Phages: a Review of Advances over the Last Decade.** *Microbiol. Mol. Biol. Rev.* 2016, 80:523–43.
175. Schmitt MJ, São-José C, Santos MA: **Phages of Yeast and Bacteria.** In *Biology of Microorganisms on Grapes, in Must and in Wine.* . Springer Berlin Heidelberg; 2009:89–109.
176. Piekarowicz A, Kłyż A, Majchrzak M, Szczêsna E, Piechucki M, Kwiatek A, Mangel TK, Stein DC: ***Neisseria gonorrhoeae* filamentous phage NgoΦ6 is capable of infecting a variety of Gram-negative bacteria.** *J. Virol.* 2014, 88:1002–10.
177. Ando H, Lemire S, Pires DP, Lu TK: **Engineering Modular Viral Scaffolds for Targeted Bacterial Population Editing.** *Cell Syst.* 2015, 1:187–196.
178. Citorik RJ, Mimee M, Lu TK: **Sequence-specific antimicrobials using efficiently delivered RNA-guided nucleases.** *Nat. Biotechnol.* 2014, 32:1141–5.
179. Bikard D, Euler CW, Jiang W, Nussenzweig PM, Goldberg GW, Duportet X, Fischetti VA, Marraffini LA: **Exploiting CRISPR-Cas nucleases to produce sequence-specific antimicrobials.** *Nat. Biotechnol.* 2014, 32:1146–50.

Chapter 6 : Annexes

6.1 Annex 1

**Sensipath: computer-aided design of sensing-enabling
metabolic pathways**

SensiPath: computer-aided design of sensing-enabling metabolic pathways

Baudoin Delépine^{1,2}, Vincent Libis^{1,2}, Pablo Carbonell³ and Jean-Loup Faulon^{1,2,3,*}

¹ISSB, Genopole, CNRS, UEVE, Université Paris Saclay, 91000 Évry, France, ²Micalis Institute, INRA, AgroParisTech, Université Paris Saclay, 78350 Jouy-en-Josas, France and ³SYNBIOCHEM Centre, Manchester Institute of Biotechnology, University of Manchester, M1 7DN Manchester, UK

Received February 9, 2016; Revised April 4, 2016; Accepted April 12, 2016

ABSTRACT

Genetically-encoded biosensors offer a wide range of opportunities to develop advanced synthetic biology applications. Circuits with the ability of detecting and quantifying intracellular amounts of a compound of interest are central to whole-cell biosensors design for medical and environmental applications, and they also constitute essential parts for the selection and regulation of high-producer strains in metabolic engineering. However, the number of compounds that can be detected through natural mechanisms, like allosteric transcription factors, is limited; expanding the set of detectable compounds is therefore highly desirable. Here, we present the SensiPath web server, accessible at <http://sensipath.micalis.fr>. SensiPath implements a strategy to enlarge the set of detectable compounds by screening for multi-step enzymatic transformations converting non-detectable compounds into detectable ones. The SensiPath approach is based on the encoding of reactions through signature descriptors to explore sensing-enabling metabolic pathways, which are putative biochemical transformations of the target compound leading to known effectors of transcription factors. In that way, SensiPath enlarges the design space by broadening the potential use of biosensors in synthetic biology applications.

INTRODUCTION

Synthetic biology and metabolic engineering applications often require as part of their design a way to assess the presence or to quantify the amount of a compound of interest. Genetically-encoded biosensors such as riboswitches and allosteric transcription factors offer the possibility to control the expression of a gene of choice. This feature makes them valuable for many applications (1,2) such as pollutant monitoring or high-throughput screening of optimized

strains and enzymes (3–5), as expression of reporter genes like fluorescent proteins can be linked to the concentration of the compound of interest. Moreover, the ability of these biosensors to provide input at the genetic level opens the way to more complex downstream signal processing and actuation (6). Examples of applications of such circuits range from threshold activation in presence of pathological concentration levels of biomarkers (7) to the creation of a feedback control motif leading to yield improvement for a chemical producing strain (8).

There is thus a critical need for biosensors, but it appears that current strategies for finding new biosensors may not be sufficient to answer all the needs. Although remarkable progress has been made in the field of genetically encoded biosensor design (9–11) and genome mining (12), the number of chemicals that can be detected is still limited and thus constitute a bottleneck in the development of synthetic biology applications.

New strategies of biosensing can be considered to tackle this issue. One of them relies on indirect sensing by transforming the molecule of interest into a detectable one. Such strategy has been successfully used with the help of enzymes to transform a key metabolite such as L-tyrosine (13) or L-DOPA (14) into pigments and thus allowing high-throughput screening of overproducers. The same strategy can also be employed to transform the molecule of interest into a molecule for which a genetically-encoded biosensor is available (15,16). We recently demonstrated that this approach could be attempted in a systematic fashion by combining information on the available biosensors and automatic design of enzymatic networks. This led to the development of five new whole-cell biosensors for pollutants (parathion, 2C4NP), biomarker (hippuric acid) and drugs (cocain, nitroglycerin) (17).

In order to open this untapped source of biosensors for synthetic biologists, we hereby present SensiPath (<http://sensipath.micalis.fr>), a web-based tool assisting the design of sensing-enabling metabolic pathways (SEMPs). SensiPath will serve users wishing to perform cell-mediated detection of a compound when no direct-sensing solution is feasible. The primary objective of SensiPath, thus, is to en-

*To whom correspondence should be addressed. Tel: +33 1 34 65 27 71; Email: jlfaulon@jouy.inra.fr

large the number of detectable compounds for synthetic biology applications. The algorithms we implemented to simulate biochemical reactions are derived from the well-tested RetroPath (18). It notably allows to take advantage of enzymatic promiscuity, i.e. the ability that enzymes have to process structurally similar substrates, thus yielding more results. SensiPath is built from a comprehensive list of more than 100 000 compounds and 87 000 reactions from four metabolic databases, covering most of the known metabolism. We also collected a large dataset of more than 500 detectable compounds for which intracellular biosensors exist from several gene expression regulation databases, focusing our search on allosteric transcription factors.

MATERIALS AND METHODS

Figure 1 shows an overview of how SensiPath works, the details are exposed in the following subsections. SensiPath is based on a comprehensive internal database of biochemical reactions and compounds encoded as chemical signatures. Once a compound query is submitted, it performs a search in order to find a match against all the enzymatic reactions that we have collected in our database. The search is carried out in order to predict reachable compounds from the target. This search generates a metabolic graph at up to two enzymatic steps away from the target, in which nodes are compounds and edges are reactions. Detectable compounds are identified and annotated by a score of similarity based on searching against the list of known detectable compounds in the database. For later reference, all SensiPath sources in its current online version are available on FigShare (<https://dx.doi.org/10.6084/m9.figshare.3144616.v1>) in addition of our list of detectable compounds (<https://dx.doi.org/10.6084/m9.figshare.3144715.v1>).

Source databases

SensiPath predictions are based on imported data from metabolic and gene expression regulation databases. We gathered data from multiple sources to cover most of available knowledge in current databases.

Reactions. Known biochemical reactions were extracted from main common reaction databases (Rhea (v66, <http://www.rhea-db.org>) (19), MetaCyc (v19.1, <http://metacyc.org>) (20), BRENDA (v15.2, <http://www.brenda-enzymes.info>) (21) as well as from a more specialized database, the Biocatalysis/Biodegradation Database (<http://eawag-bbd.ethz.ch>, accessed in December 2015) (22). We considered only reactions for which structures of all reactants were available, fully defined and valid. Overall, we collected more than 100 000 compounds and 87 000 reactions with references to external databases.

Detectable compounds. We gathered a list of 504 putative detectable compounds focusing our search on effectors of allosteric transcription factors from prokaryotes. Data were collected from several gene expression regulation databases: RegulonDB (v9.0, <http://regulondb.ccg.unam.mx>) (23), RegPrecise (v4, <http://regprecise.lbl.gov>) (24), RegTransBase (v7, <http://regtransbase.lbl.gov>) (25) and BioNemo (v6.0, <http://bionemo.bioinfo.cnio.es>) (26).

Reaction and compound encoding

In order to encode the reactions we first normalized the compounds, next computed molecular signatures and finally computed reaction signatures.

Compound normalization. The representation of compounds must be normalized in order to improve the performance of the encoding method. In particular, compounds were represented under their aromatic form while charges and hydrogens were removed; stereochemistry was kept.

Molecular signature. All compounds were encoded internally through their molecular signature (27). The molecular signature of a compound is a list of overlapping molecular fragments, each of them centred on a distinct atom. Thus, fragments represent atom neighbourhood (also called atomic signature or atomic environment) in terms of atom and bond type. Basically, a molecular signature is similar to the extended connectivity circular fingerprint (ECFP) (28). We used fragments (atomic signatures) with an environment diameter of 12 bonds.

Reaction signature. All biochemical reactions were represented internally by reaction signatures (29). The reaction signature $\sigma(R)$ is defined in a vector space as the sum of molecular signatures of products less the sum of molecular signatures of substrates:

$${}^d\sigma(R_n) = \sum_i {}^d\sigma(P_i) - \sum_j {}^d\sigma(S_j)$$

where ${}^d\sigma(P_i)$ and ${}^d\sigma(S_j)$ are the molecular signatures of substrate S_j and product P_i at diameter d .

This approach allows us to encode biochemical reactions by looking at the changes occurring at the reaction center. Note that the specificity of a reaction signature is determined by the diameter of the molecular signature, as lower diameters encode multiple compounds while higher diameters are specific. Therefore, reaction signatures have been shown as a handy way to model enzymatic substrate promiscuity (18,29,30), i.e. the ability that enzymes have to process structurally similar substrates. Our chosen diameter of 12 assumes a relatively low degree of enzymatic promiscuity for the encoded reactions.

Matching algorithm

After integrating reaction signatures in our database, we can predict on-the-fly if a compound can act as substrate of a reaction by using a new implementation of the RetroPath forward algorithm (18). If a compound C has a list of fragments (atomic signatures and their respective occurrence) embedding the substrate fragments contained in a reaction signature R (i.e. the negative part of reaction signature), then the compound is said to match the reaction. The sum of the signatures of compound C and those of the reaction R generates a new list of (positive) fragments P, representing the putative products generated by the reaction signature acting upon compound C. If we can retrieve a set of known compounds from those fragments, then the reaction is accepted and C is considered a valid substrate for R to produce P.

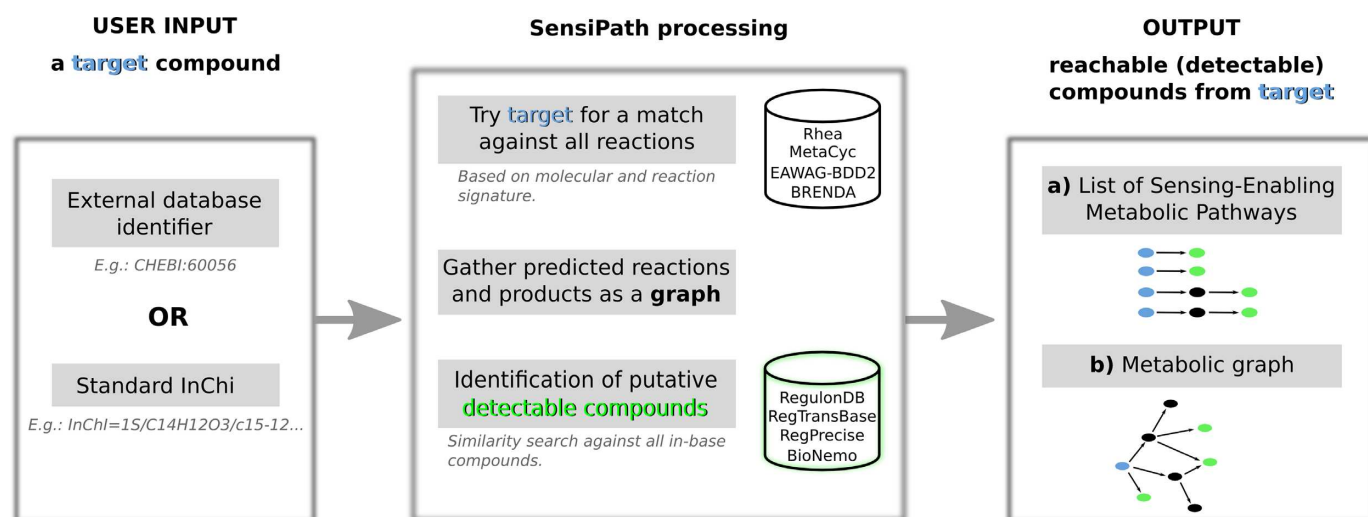


Figure 1. SensiPath flowchart. Users query a target compound (blue), either as an external database identifier or as a standard InChi. Target is processed to gather reachable compounds through enzymatic reactions; detectable compounds are annotated (green). Results are displayed both as (a) the set of pathways leading to recognized detectable compounds; and (b) the whole computed graph around the target.

Metabolic graph

Pathways are handled as a graph (where nodes are compounds and edges reactions) with NetworkX python library (31).

Similarity search

In order to annotate compounds structurally similar to detectable compounds in predicted metabolic graphs, we pre-computed the similarities between all compounds and detectable ones. Indeed, promiscuous detection of structurally similar compounds may not be reported in databases and should be checked in the literature if no suitable detectable compound is found by SensiPath.

Similarity was evaluated with RDKit python library (<http://www.rdkit.org/>), representing compounds with RDKit's ECFP4 fingerprint implementation and a Jaccard-Tanimoto index (32). A Tanimoto of one is a perfect match.

Web server implementation

SensiPath web server is a Docker application running the following standard software packages: Nginx, gUnicorn, Django and Postgres. Data and matching functions are stored in the database.

INPUT AND OUTPUT

Input

Users query SensiPath with the compound they wish to detect (Figure 1, left panel), either as an identifier from an external database (e.g. ChEBI available at <https://www.ebi.ac.uk/chebi/>) or as a standard InChi (<http://www.inchi-trust.org/>). InChi is a IUPAC string representation of compounds and can be easily obtained from compound databases. Users can choose to search for detectable compounds that are at one or two enzymatic steps away from their target.

Output

SensiPath displays its results in two views; (i) pathway view: the set of pathways leading to recognized detectable compounds (Figure 2A); and (ii) graph view: the whole computed graph around the target (Figure 2B), also available for download as a standard Graph Markup Language file.

CASE STUDIES

Five examples of SEMP were characterized experimentally in *Escherichia coli* by our group to validate SEMP concept. For a model bacteria such as *E. coli*, *in vivo* implementation of SEMP only requires basic molecular biology knowledge. As an example, we describe here the design steps required to build a strain of *E. coli* able to detect the drug cocaine, and a strain able to detect the pollutant parathion with the help of SensiPath. A detailed *in vivo* characterization for these examples is described elsewhere (17).

We refer to 'the metabolic module' as the genetic parts providing the enzymatic transformations and to 'the sensing module' as the gene circuit consisting of the transcription factor, its responsive promoter and the reporter gene.

Cocaine detection

While several studies have shown interest in detecting cocaine in biological samples, they rely on aptamers and nanoparticles sensors (33,34), which do not allow the signal to be transferred to the genetic layer of a living organism, a requirement for further *in situ* signal processing.

Here, we show how SensiPath was used in order to design a SEMP that detects cocaine *in vivo*. To that end, SensiPath web server is queried using a chemical identifier of cocaine, either through CHEBI:60056 or InChi=1S/C17H21NO4/c1-18-12-8-9-13(18)15(17(20)21-2)14(10-12)22-16(19)11-6-4-3-5-7-11/h3-7,12-15H,8-10H2,1-2H3/t12-,13+,14-,15+/m0/s1. SensiPath founds a candidate SEMP allowing detection that is one enzymatic

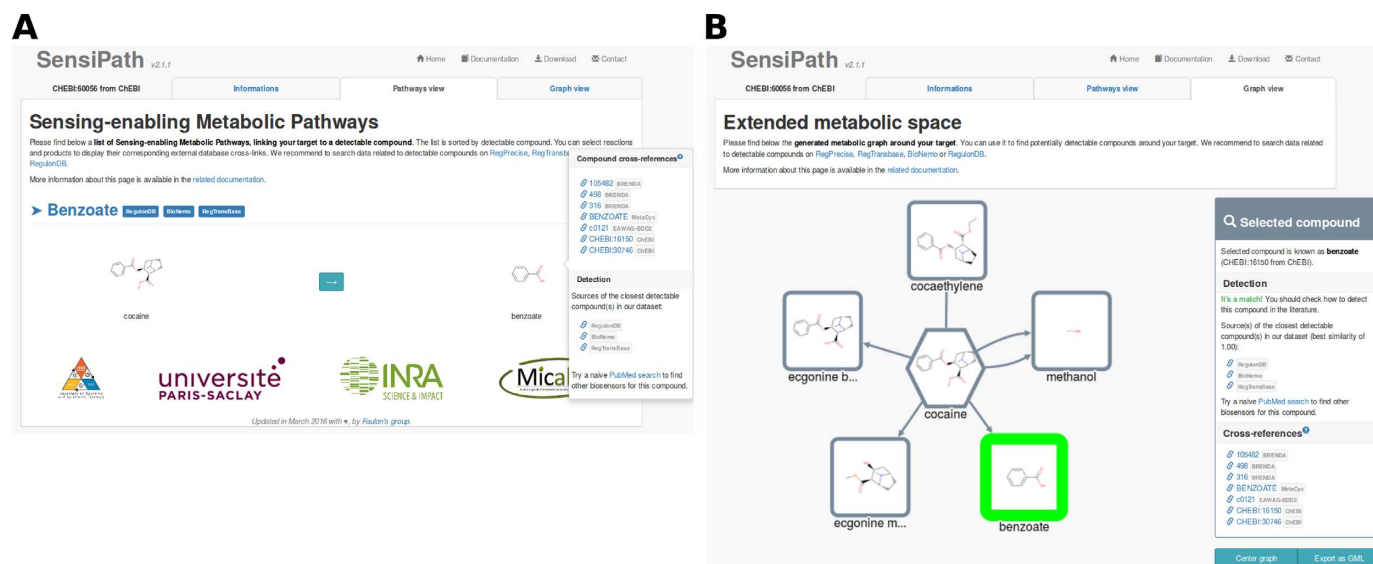


Figure 2. Results pages for querying 'CHEBI:60056' at one step. **(A)** Pathway view. List of sensing-enabling metabolic pathways, sorted by detectable compound. Clicking reactions (arrows) or products will display a list of cross-references to relevant databases. Cross-references which are displayed together are considered identical by SensiPath. **(B)** Graph view. Interactive computed graph around the target (hexagonal node). Putative detectable compounds have a green border. Selected elements (edge or node) have a bold border. Again, clicking elements will display their respective cross-references (right panel). Note that reactions leading to several products have duplicated edges; for instance, benzoate and ecgonine methyl ester are both products of the same biochemical reaction (hydrolysis by a cocaine esterase).

step away from the target. On the Graph view (Figure 2B), the five different products obtained through known enzymatic activities on cocaine are displayed. Clicking on an edge of the graph provides a link to databases providing information on each reaction. One of these compounds has a green border indicating that a biosensor is known to interact with an identical or highly similar chemical structure. This suggests that the information of the presence of cocaine in the medium can be transferred to the genetic layer and thus constitutes a putative SEMP. All found SEMPs are summarized on the Pathway view (Figure 2A). In the present case study, cocaine can be hydrolysed and forms the detectable molecule benzoate. Clicking on the arrow that represents the enzymatic transformation will display cross reference links to external databases of enzymatic transformations. It is strongly recommended to carefully check the bibliography that motivated the annotation of the reaction in the database, since important results might be omitted or misrepresented due to an incorrect curation process. In the case of cocaine hydrolysis, several publications confirm the benzoate conversion and databases such as Rhea and MetaCyc provide a direct link to Uniprot or GenBank where the sequence coding for the enzyme can be found (GenBank AF173165.1). This sequence can be synthesized and cloned into an expression vector of choice to constitute the metabolic module part of the SEMP.

In parallel, a query on BioNemo or RegTransBase for benzoate (or benzoic acid), the compound reported by SensiPath as having a biosensor, leads to several potential transcription factors that are known to interact with this compound (BenM, BenR, CbdS, PcaR, TcbR, CatR, BadR and XylS). In our experimental implementation, we chose the couple composed of BenR and its responsive promoter

pBen from *Pseudomonas putida* KT2440 after a quick assessment of the available literature.

The sequence of pBen can then be synthesized and cloned in front of a reporter gene of choice (e.g. a fluorescent protein) in addition to the transcription factor coding sequence in order to form the sensing module part of the SEMP. To maximize the chances of proper expression of the heterologous proteins, we recommend to perform a step of codon-optimization on all the coding sequences, to place them under control of inducible promoters and to use strain such as BL21(DE3) due to its efficient protein expression capabilities.

Parathion detection

Synthetic biology application of biosensors in the field of environmental protection could take the form of micro-organisms programmed with a 'seek and destroy' behaviour toward pollutants (35). However, the task of engineering tailor-made biosensors for pollutants has been difficult to date (36).

Parathion is listed as one of the twelve worst offenders persistent organic pollutants according to the United Nations Environment Program and could benefit from such synthetic biology applications provided that a biosensor is available.

A request on Sensipath for parathion, with identifier CHEBI:27928, leads to the identification of a 1-step SEMP that depends on a phosphotriesterase (PTE) allowing transformation of parathion into 4-nitrophenol. As in the previous cocaine example, the proposed transformation could be verified in the literature (37). We have experimentally validated this SEMP with a metabolic module based on the PTE coding sequence coupled with the sensing mod-

ule made up of the transcription factor DmpR and its responsive promoter Pu from *Pseudomonas* sp. CF600. However, both PTE enzyme and DmpR promoter are known to be promiscuous, and other pollutants harbouring phenolic structures could activate DmpR. As this could impair applications requiring a high specificity, alternative SEMP were also explored.

Interestingly, with a 2-steps query, SensiPath's Pathway view shows that 4-nitrophenol can be an intermediate compound to another SEMP based on nitrite detection. Indeed a second enzymatic step mediated by a monooxygenase (38) is able to further transform 4-nitrophenol into nitrite, which is known to interact with regulators such as NarL from *E. coli*. This alternative offers the possibility of developing a more specific biosensor, effectively discarding any risk of cross-activation by phenolic compounds, as long as they do not have a nitro group. Going further with this idea, high specificity target detection could be guaranteed by building up combinations of alternative SEMPs in one or several strains.

DISCUSSION

The development of novel biosensors is presently needed in order to enlarge the set of detectable and observable metabolites that are available for synthetic biology applications such as in health, environment or fine chemical production. In that direction, the SensiPath web server provides synthetic biologists with new solutions to build circuits having the ability of triggering a genetic response when a compound of interest is present. Our biosensor design solution is based on the strategy, not fully explored previously, of performing an *in silico* screening for enzymatic pathways linking the target to known detectable compounds. The originality of the approach lies in the systematic search through a full enumeration that SensiPath carries out, allowing discovery of novel sensing pathway candidates in the metabolic space. Resulting SEMPs are appealing for synthetic biologists because they can be easily built using conventional DNA assembling techniques and tested *in vivo*. SensiPath thus provides an easy way to explore right out of the box multiple biosensor constructs.

Depending on the application, the reliability of the candidate SEMPs identified by our method may vary. Limitations of the SEMP method include the need for the target compound to be able to co-localize with the enzyme (i.e. to enter the cells or to be internally produced in the cell), and the need for enzymatic products of the sensing pathway to be not too toxic to the cell. Such issues need to be addressed in a case-by-case manner, since they greatly depend on the application and on physico-chemical properties that are not always known for the compound. Other potential limitations of the method hold with regards to the choice of the biosensor. Although some information about the degree of promiscuity of transcription factors may be available from databases and literature, this aspect should be carefully considered in each application, especially if the final application requires a high level of specificity. The choice of the biosensor should also take into account dose response parameters such as the dynamic range and linear range of detection. SEMP's properties will depend on the actual proper-

ties of the biosensor, an information that therefore should be considered and retrieved from the available literature. In addition, promoter sequences responding to transcription factors may not be always found in databases, often requiring an investigation of associated references. This information nevertheless is progressively becoming more available through repositories like the Registry of Standard Biological Parts (http://parts.igem.org/Main_Page).

In conclusion, we believe that the SEMP detection method is an interesting alternative worth considering with respect to tailored solutions such as rational design (10,11) or genome mining (12). To the authors acknowledgement, this is the first time a web-based tool is proposed to design biosensors based on the SEMPs approach. Other tools (such as M-path (39) or BioSynther (40) to name a few) proposed finding pathways from one compound to another, but they did not include any detectability concept in the way it was considered here. In that sense, SensiPath and SEMPs will surely contribute to the design of new synthetic biology applications. Moreover, we should expect in the next years to see the broadness of applicability of SEMPs to increase in parallel with progress in reaction and gene expression regulation knowledge sources.

AVAILABILITY

SensiPath is available online at <http://sensipath.micalis.fr>. A stand-alone snapshot of SensiPath at the time this manuscript was written is available on FigShare at <https://dx.doi.org/10.6084/m9.figshare.3144616.v1>.

ACKNOWLEDGEMENT

The authors would like to thank P. Novichkov and D. Rodionov for providing effector data from RegPrecise; P. Vessiere for helpful insights on database implementation; T. Duigou and JM. Batto for setting up the server.

FUNDING

Structure et Dynamique des Systèmes Vivants doctoral school, Université Paris Saclay; Direction Générale de l'Armement (DGA-MRIS Scholarship); Biotechnology and Biological Sciences Research Council, Centre for synthetic biology of fine and speciality chemicals [BB/M017702/1]; GIP Genopole; Toulouse White Biotech (TWB); and French National Research Agency [ANR-15-CE1-0008]. Funding for open access charge: Biotechnology and Biological Sciences Research Council [BB/M017702/1].

Conflict of interest statement. None declared.

REFERENCES

1. Liu,D., Evans,T. and Zhang,F. (2015) Applications and advances of metabolite biosensors for metabolic engineering. *Metab. Eng.*, **31**, 35–43.
2. Mahr,R. and Frunzke,J. (2016) Transcription factor-based biosensors in biotechnology: current state and future prospects. *Appl. Microbiol. Biotechnol.*, **100**, 79–90.
3. Mahr,R., Gätgens,C., Gätgens,J., Polen,T., Kalinowski,J. and Frunzke,J. (2015) Biosensor-driven adaptive laboratory evolution of l-valine production in *Corynebacterium glutamicum*. *Metab. Eng.*, **32**, 184–194.

4. Boock, J.T., Gupta, A. and Prather, K.L. (2015) Screening and modular design for metabolic pathway optimization. *Curr. Opin. Biotechnol.*, **36**, 189–198.
5. de las Heras, A. and de Lorenzo, V. (2011) In situ detection of aromatic compounds with biosensor *Pseudomonas putida* cells preserved and delivered to soil in water-soluble gelatin capsules. *Anal. Bioanal. Chem.*, **400**, 1093–1104.
6. Brophy, J.A.N. and Voigt, C.A. (2014) Principles of genetic circuit design. *Nat. Methods*, **11**, 508–520.
7. Courbet, A., Endy, D., Renard, E., Molina, F. and Bonnet, J. (2015) Detection of pathological biomarkers in human clinical samples via amplifying genetic switches and logic gates. *Sci. Transl. Med.*, **7**, 289ra83.
8. Xu, P., Li, L., Zhang, F., Stephanopoulos, G. and Koffas, M. (2014) Improving fatty acids production by engineering dynamic pathway regulation and metabolic control. *Proc. Natl. Acad. Sci. U.S.A.*, **111**, 11299–11304.
9. Tang, S.-Y., Fazelinia, H. and Cirino, P.C. (2008) AraC regulatory protein mutants with altered effector specificity. *J. Am. Chem. Soc.*, **130**, 5267–5271.
10. Taylor, N.D., Garruss, A.S., Moretti, R., Chan, S., Arbing, M.A., Cascio, D., Rogers, J.K., Isaacs, F.J., Kosuri, S., Baker, D. *et al.* (2016) Engineering an allosteric transcription factor to respond to new ligands. *Nat. Methods*, **13**, 177–183.
11. Jha, R.K., Chakraborti, S., Kern, T.L., Fox, D.T. and Strauss, C.E.M. (2015) Rosetta comparative modeling for library design: engineering alternative inducer specificity in a transcription factor. *Proteins*, **7**, 1327–1367.
12. Uchiyama, T. and Watanabe, K. (2008) Substrate-induced gene expression (SIGEX) screening of metagenome libraries. *Nat. Protoc.*, **3**, 1202–1212.
13. Santos, C.N.S. and Stephanopoulos, G. (2008) Melanin-based high-throughput screen for L-tyrosine production in *Escherichia coli*. *Appl. Environ. Microbiol.*, **74**, 1190–1197.
14. DeLoache, W.C., Russ, Z.N., Narcross, L., Gonzales, A.M., Martin, V.J.J. and Dueber, J.E. (2015) An enzyme-coupled biosensor enables (S)-reticuline production in yeast from glucose. *Nat. Chem. Biol.*, **11**, 465–471.
15. Mohn, W.W., Garmendia, J., Galvao, T.C. and de Lorenzo, V. (2006) Surveying biotransformations with à la carte genetic traps: translating dehydrochlorination of lindane (gamma-hexachlorocyclohexane) into lacZ-based phenotypes. *Environ. Microbiol.*, **8**, 546–555.
16. Rogers, J.K. and Church, G.M. (2016) Genetically encoded sensors enable real-time observation of metabolite production. *PNAS*, **9**, 2388–2481.
17. Libis, V., Delépine, B. and Faulon, J.-L. (2016) Expanding biosensing abilities through computer-aided design of metabolic pathways. *ACS Synth. Biol.*, doi:10.1021/acssynbio.5b00225.
18. Carbonell, P., Parutto, P., Baudier, C., Junot, C. and Faulon, J.-L. (2014) Retropath: automated pipeline for embedded metabolic circuits. *ACS Synth. Biol.*, **3**, 565–577.
19. Morgat, A., Axelsen, K.B., Lombardot, T., Alcántara, R., Aimo, L., Zerara, M., Niknejad, A., Belda, E., Hyka-Nouspikel, N., Coudert, E. *et al.* (2015) Updates in Rhea—a manually curated resource of biochemical reactions. *Nucleic Acids Res.*, **43**, D459–D464.
20. Caspi, R., Billington, R., Ferrer, L., Foerster, H., Fulcher, C.A., Keseler, I.M., Kothari, A., Krummenacker, M., Latendresse, M., Mueller, L.A. *et al.* (2016) The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Res.*, **44**, D471–D480.
21. Chang, A., Schomburg, I., Placzek, S., Jeske, L., Ulbrich, M., Xiao, M., Sensen, C.W. and Schomburg, D. (2015) BRENDA in 2015: exciting developments in its 25th year of existence. *Nucleic Acids Res.*, **43**, D439–D446.
22. Gao, J., Ellis, L.B.M. and Wackett, L.P. (2010) The University of Minnesota Biocatalysis/Biodegradation Database: improving public access. *Nucleic Acids Res.*, **38**, D488–D491.
23. Salgado, H., Peralta-Gil, M., Gama-Castro, S., Santos-Zavaleta, A., Muñoz-Rascado, L., García-Sotelo, J.S., Weiss, V., Solano-Lira, H., Martínez-Flores, I., Medina-Rivera, A. *et al.* (2013) RegulonDB v8.0: omics data sets, evolutionary conservation, regulatory phrases, cross-validated gold standards and more. *Nucleic Acids Res.*, **41**, D203–D213.
24. Novichkov, P.S., Kazakov, A.E., Ravcheev, D.A., Leyn, S.A., Kovaleva, G.Y., Sutormin, R.A., Kazanov, M.D., Riehl, W., Arkin, A.P., Dubchak, I. *et al.* (2013) RegPrecise 3.0—a resource for genome-scale exploration of transcriptional regulation in bacteria. *BMC Genomics*, **14**, 745.
25. Cipriano, M.J., Novichkov, P.N., Kazakov, A.E., Rodionov, D.A., Arkin, A.P., Gelfand, M.S. and Dubchak, I. (2013) RegTransBase—a database of regulatory sequences and interactions based on literature: a resource for investigating transcriptional regulation in prokaryotes. *BMC Genomics*, **14**, 213.
26. Carbajosa, G., Trigo, A., Valencia, A. and Cases, I. (2009) Bionemo: molecular information on biodegradation metabolism. *Nucleic Acids Res.*, **37**, D598–D602.
27. Carbonell, P., Carlsson, L. and Faulon, J.-L. (2013) Stereo signature molecular descriptor. *J. Chem. Inf. Model.*, **53**, 887–897.
28. Rogers, D. and Hahn, M. (2010) Extended-connectivity fingerprints. *J. Chem. Inf. Model.*, **50**, 742–754.
29. Faulon, J.-L., Misra, M., Martin, S., Sale, K. and Sapra, R. (2008) Genome scale enzyme-metabolite and drug-target interaction predictions using the signature molecular descriptor. *Bioinformatics*, **24**, 225–233.
30. Carbonell, P., Parutto, P., Herisson, J., Pandit, S.B. and Faulon, J.-L. (2014) XTMS: pathway design in an eXTended metabolic space. *Nucleic Acids Res.*, **42**, W389–W394.
31. Hagberg, A.A., Schult, D.A. and Swart, P.J. (2008) Exploring network structure, dynamics, and function using NetworkX. In: Varoquaux, G., Vaught, T. and Millman, J. (eds). *Proceedings of the 7th Python in Science Conference*. Pasadena, pp. 11–15.
32. Jaccard, P. (1901) Distribution de la flore alpine dans le bassin des Dranses et dans quelques régions voisines. *Bull. Soc. Vaud. Sci. Nat.*, **37**, 241–272.
33. Neves, M.A.D., Blaszykowski, C., Bokhari, S. and Thompson, M. (2015) Ultra-high frequency piezoelectric aptasensor for the label-free detection of cocaine. *Biosens. Bioelectron.*, **72**, 383–392.
34. Emrani, A.S., Danesh, N.M., Ramezani, M., Taghdisi, S.M. and Abnous, K. (2016) A novel fluorescent aptasensor based on hairpin structure of complementary strand of aptamer and nanoparticles as a signal amplification approach for ultrasensitive detection of cocaine. *Biosens. Bioelectron.*, **79**, 288–293.
35. Sinha, J., Reyes, S.J. and Gallivan, J.P. (2010) Reprogramming bacteria to seek and destroy an herbicide. *Nat. Chem. Biol.*, **6**, 464–470.
36. Sinha, J., Reyes, S. and Gallivan, J.P. (2014) Retraction: reprogramming bacteria to seek and destroy an herbicide. *Nat. Chem. Biol.*, **10**, 239–239.
37. Roodveldt, C. and Tawfik, D.S. (2005) Directed evolution of phosphotriesterase from *Pseudomonas diminuta* for heterologous expression in *Escherichia coli* results in stabilization of the metal-free state. *Protein Eng. Des. Sel.*, **18**, 51–58.
38. Zhang, J.-J., Liu, H., Xiao, Y., Zhang, X.-E. and Zhou, N.-Y. (2009) Identification and characterization of catabolic para-nitrophenol 4-monooxygenase and para-benzoquinone reductase from *Pseudomonas* sp. strain WBC-3. *J. Bacteriol.*, **191**, 2703–2710.
39. Araki, M., Cox, R.S., Makiguchi, H., Ogawa, T., Taniguchi, T., Miyaoku, K., Nakatsui, M., Hara, K.Y. and Kondo, A. (2015) M-path: a compass for navigating potential metabolic pathways. *Bioinformatics*, **31**, 905–911.
40. Tu, W., Zhang, H., Liu, J. and Hu, Q.-N. (2016) BioSynther: a customized biosynthetic potential explorer. *Bioinformatics*, **32**, 472–473.

6.2 Annex 2:

A sense of balance: experimental investigation and modeling of a malonyl-CoA sensor in *Escherichia coli*



A sense of balance: experimental investigation and modeling of a malonyl-CoA sensor in *Escherichia coli*

Tamás Fehér^{1,2}, Vincent Libis^{1,3}, Pablo Carbonell^{1,4,5} and Jean-Loup Faulon^{1,5}*

¹ Institute of Systems and Synthetic Biology, University of Evry Val d'Essonne, Evry, France

² Institute of Biochemistry, Biological Research Centre of the Hungarian Academy of Sciences, Szeged, Hungary

³ Paris Diderot University, Paris, France

⁴ Research Programme on Biomedical Informatics (GRIB), Department of Experimental and Health Sciences, Hospital del Mar Medical Research Institute (IMIM), Universitat Pompeu Fabra, Barcelona, Spain

⁵ SYNBIOCHEM Center, Manchester Institute of Biotechnology, School of Chemistry, University of Manchester, Manchester, UK

Edited by:

Jean Marie François, CNRS, France

Reviewed by:

Mattheos Koffas, Rensselaer

Polytechnic Institute, USA

Shota Isogai, Technology Research

Association of Highly Efficient Gene
Design, Japan

*Correspondence:

Jean-Loup Faulon, 5 rue Henri
Desbruères, Genavenir 6, F-91030,
Evry Cedex, France
e-mail: jean-loup.faulon@issb.
genopole.fr

Production of value-added chemicals in microorganisms is regarded as a viable alternative to chemical synthesis. In the past decade, several engineered pathways producing such chemicals, including plant secondary metabolites in microorganisms have been reported; upscaling their production yields, however, was often challenging. Here, we analyze a modular device designed for sensing malonyl-CoA, a common precursor for both fatty acid and flavonoid biosynthesis. The sensor can be used either for high-throughput pathway screening in synthetic biology applications or for introducing a feedback circuit to regulate production of the desired chemical. Here, we used the sensor to compare the performance of several predicted malonyl-CoA-producing pathways, and validated the utility of malonyl-CoA reductase and malonate-CoA transferase for malonyl-CoA biosynthesis. We generated a second-order dynamic linear model describing the relation of the fluorescence generated by the sensor to the biomass of the host cell representing a filter/amplifier with a gain that correlates with the level of induction. We found the time constants describing filter dynamics to be independent of the level of induction but distinctively clustered for each of the production pathways, indicating the robustness of the sensor. Moreover, by monitoring the effect of the copy-number of the production plasmid on the dose-response curve of the sensor, we managed to coarse-tune the level of pathway expression to maximize malonyl-CoA synthesis. In addition, we provide an example of the sensor's use in analyzing the effect of inducer or substrate concentrations on production levels. The rational development of models describing sensors, supplemented with the power of high-throughput optimization provide a promising potential for engineering feedback loops regulating enzyme levels to maximize productivity yields of synthetic metabolic pathways.

Keywords: malonyl-CoA, dynamic pathway regulation, high-throughput screening, synthetic regulatory circuit, fluorescent reporter circuit, sensor-actuator circuit

INTRODUCTION

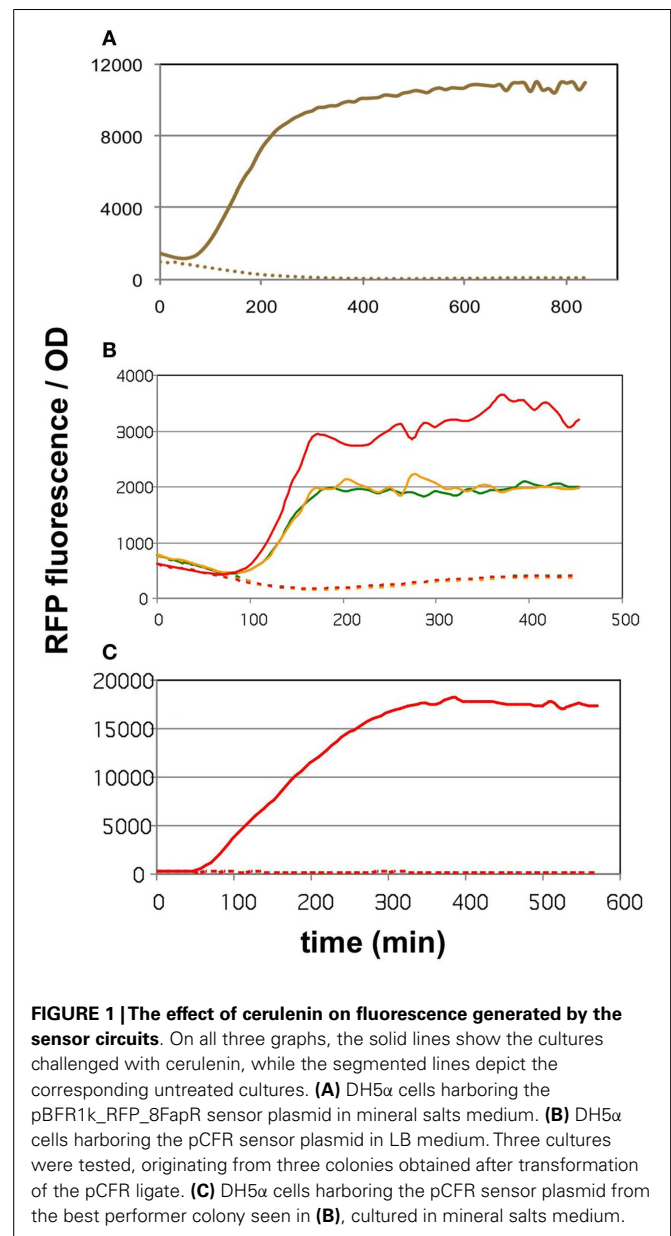
Natural metabolic pathways are highly regulated and can adapt dynamically to changes in the levels of chemicals within and around the cell (Chubukov et al., 2014). For this purpose, cells have developed a high number of sensors to monitor and control their own metabolic state. This stringent regulatory process is impaired or lost when metabolic engineers insert heterologous enzymes into artificial pathways, usually at non-optimal levels of activities, leading to accumulation of intermediate metabolites and reduced growth rate (Holtz and Keasling, 2010). To address this issue, several projects have attempted to mimic natural regulation by adding such sensors to engineered metabolic pathways. The advantages of using natural sensors such as transcription factors to implement synthetic dynamic regulation in heterologous pathways have been demonstrated in several studies: following pioneering work on lycopene synthesis regulation (Farmer and Liao, 2000), this strategy also led recently to higher yields in fatty

acid synthesis with the help of fatty acids (Zhang et al., 2012) and malonyl-CoA (Xu et al., 2014; Liu et al., 2015) biosensors.

Such biosensor-based approaches allow the bacterial factory to artificially monitor its own level of metabolites and to modify expression levels of certain enzymes in order to balance pathway function. They can also be used to devise external monitoring systems to aid the selection of cells displaying improved biological functions. A major application for this monitoring is the screening of libraries of variant strains in directed evolution of enzymes (Michener and Smolke, 2014). Screening of producers is still a major bottleneck in metabolic engineering as quantification of the compound of interest usually relies on low-throughput measurement techniques such as liquid or gas chromatography. However, the typical throughput necessary for directed evolution applications or combinatorial libraries' screening is several orders of magnitude higher (Dietrich et al., 2010). Therefore, *in vivo* biosensing offers a relevant non-destructive, cost-effective,

and high-throughput alternative for monitoring production levels (Schallmeyer et al., 2014). This strategy has proven successful in identifying high-producing strains, via the coupling of biosensor to reporter genes such as fluorescent proteins or fitness-related proteins. The screening can either be done at the colony level, in microtiter plates or by fluorescence-activated cell sorting (FACS). For instance, screening of *Escherichia coli* colonies expressing a transcription factor evolved to detect mevalonate allowed successful identification of high producers of mevalonate in a library of mevalonate-synthesis variants (Tang and Cirino, 2011). Theoretically, FACS may allow even higher throughputs and was notably used to evolve the caffeine demethylase enzyme in yeast cells harboring riboswitches able to detect theophylline, the enzyme's product (Smolke and Michener, 2012). FACS was also used in combination with transcription factor-based sensors to measure L-lysine production in *C. glutamicum* (Binder et al., 2012) and detect production of Kaempferol in engineered *E. coli* (Siedler et al., 2014). Alternatively, growth rate can be coupled to the production of the product of interest via transcription factor-dependent expression of antibiotic resistance proteins. This strategy has been used in a proof of principle experiment to enrich a population of *E. coli* for cells harboring the pathway for 1-butanol-production (Dietrich et al., 2013). Later, the power of multiplex automated genome engineering was combined with this approach to optimize the production of naringenin and glucaric acid in *E. coli* (Raman et al., 2014).

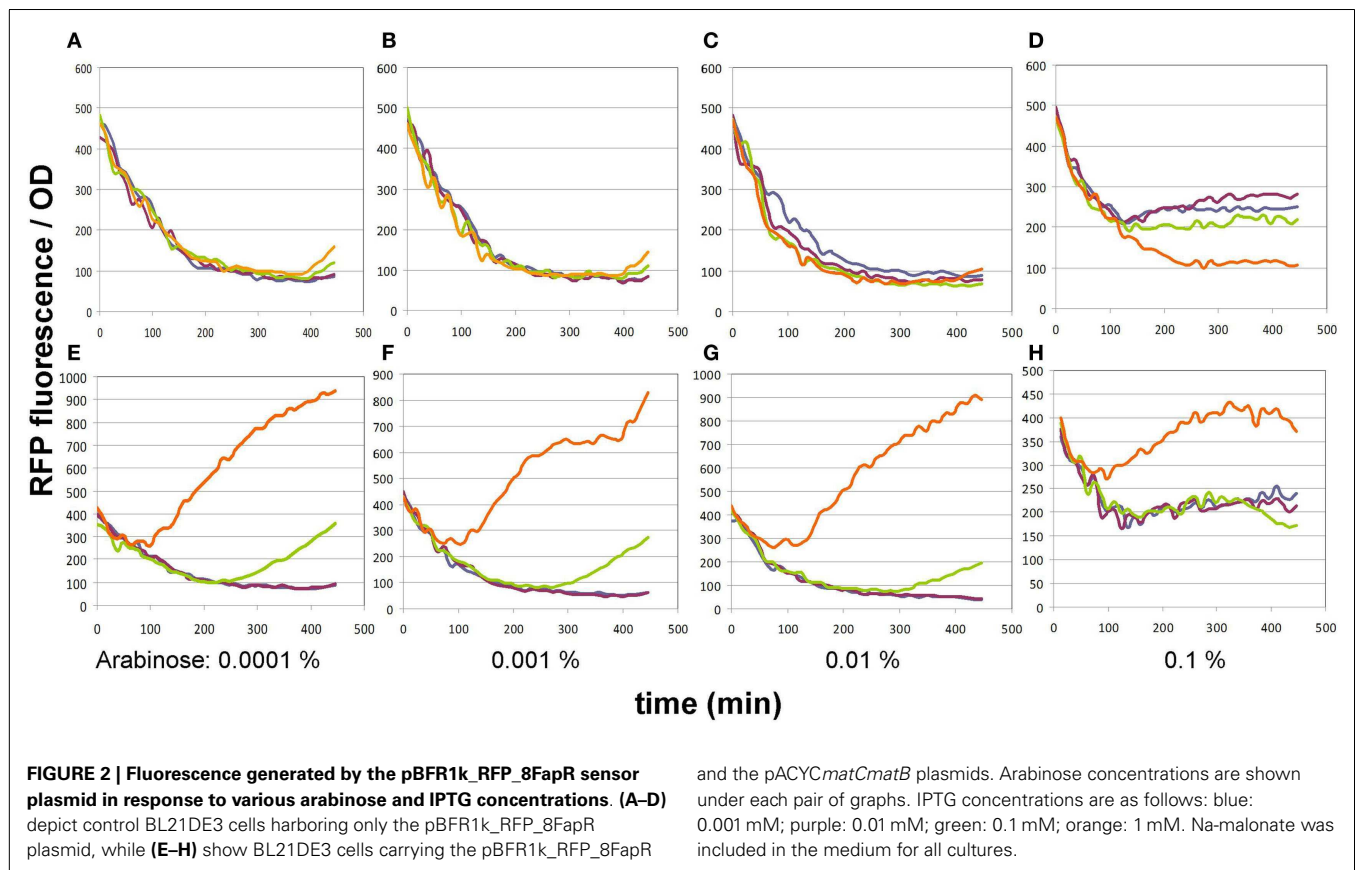
Because both pathway dynamic regulation and screening rely on the same type of biosensors, a synergy exists between these two approaches. Illustrating such synergy, here we adapted a malonyl-CoA biosensor previously developed for a dynamic regulation purpose (Liu et al., 2015) to a new screening goal. Transcription factor-based biosensing is particularly interesting for monitoring malonyl-CoA as it is difficult and time consuming to quantify this molecule *in vivo* via standard LC/MS. In addition to being an important step for fatty acid synthesis in *E. coli*, malonyl-CoA production is also a limiting step for the production of plant flavonoids in engineered *E. coli*. Our group is working on the development of computer-assisted design of metabolic pathways and is seeking high-throughput techniques to test predicted pathways and enzymes for continual feedback-optimization of the prediction algorithm. In our earlier proof of concept work, we used our software RetroPath (Carbonell et al., 2011) to identify and rank pathways capable of producing the high-value flavonoid pinocembrin in *E. coli*. During the implementation of multiple constructs, Malonyl-CoA, was identified as a bottleneck, preventing high-yield production. Here, we used the biosensor based on transcription factor FapR to analyze the malonyl-CoA-producing pathways that were proposed by our software to remove the bottleneck. We identified optimal conditions for malonyl-CoA monitoring by fine-tuning levels of expression of the transcription factor and we fitted our experimental data to a model giving insights on proper interpretation of output profiles encountered in such screening protocols. This approach helped us identify the best candidate pathways, validate two novel routes of malonyl-CoA synthesis, and optimize enzyme-expression levels through modification of plasmid copy-number.



MATERIALS AND METHODS

STRAINS AND PLASMIDS

Plasmid construction was done in *E. coli* strain DH5α, and all malonyl-CoA sensing experiments were carried out in strain BL21DE3. Plasmid pBFR1k_RFP_8FapR (Liu et al., 2015) was a kind gift of Dr. Fuzhong Zhang (Washington University, St. Louis, MO, USA). To construct plasmid pCFR, pBFR1k_RFP_8FapR was PCR-amplified using primers pBFR-FW (ACTGTCGACGAAACGATCCTCATCCTG) and pBFR-Rev (GCTCTAGATTCTTCTGAGCGGGACTCTG), and the Cm-resistance marker of pSG76-CS (GeneBank: AF402780.1) was amplified with primers pSGCSH-FW (GCTCTAGAGTGAGGCA CCAATAACTG) and pSGCSH-Rev (ACTGTCGACGATCGGCA CGTAAGAGGTTTC). Both PCR products were double-digested



with *Xba*I and *Sal*I, and were ligated using T4 ligase. Plasmid pMSD8 was a kind gift of Prof. John Cronan (University of Illinois, Urbana, IL, USA), and plasmids pETM6-MaccABCD and pETM6-PaccABCD (Xu et al., 2013) were kindly provided by Prof. Matheos Koffas (Rensselaer Polytechnic Institute, Troy, NY, USA). The construction of plasmids pRSFmatCmatB, pRSFmmsA, pRSFcagg1256, pRSFMaccABCD, and pRSFmatCatoDA was described elsewhere (Fehér et al., 2014). Plasmid pACYCmatCmatB (Wu et al., 2013) was a kind gift of Dr. Jingwen Zhou (LBBE, Jiangnan, China). pACYCMaccABCD was constructed by double-digesting pRSFMaccABCD with *Bam*HI and *Avr*II, and ligating the 4855 bp-long fragment with the similarly digested pACYCmatCmatB. pACYCmatCatoDA was engineered by double-digesting pRSFmatCatoDA with *Nde*I and *Xho*I, and ligating the 1408 bp fragment with the 5273 bp fragment of the similarly digested pACYCmatCmatB. pACYCmmsA and pACYCcagg1256 were constructed by double-digesting pRSFmmsA and pRSFcagg1256, respectively, both with *Eco*RI and *Xho*I, and ligating the respective 1685 and 3848 bp fragments with the similarly digested pACYCmatCmatB. Restriction endonucleases and T4 ligase were obtained from Thermo Fisher Scientific (Waltham, MA, USA), Q5 DNA polymerase for PCR was from New England Biolabs.

FLUORESCENCE MEASUREMENT OF BACTERIAL CULTURES

Each bacterial strain to be measured was grown overnight in mineral salts medium (Hall, 1998) supplemented with 0.2% glucose

and the appropriate antibiotics. On the day of the measurement, the cultures were diluted 20-fold in fresh medium, and dispensed into clear-bottomed black-walled 96-well plates (Costar Ref. 3603, Corning, NY, USA). Growth and measurement of the cultures took place in a TECAN Infinite 500 fluorescent reader, shaken and incubated at 37°C. Optical density was determined at 600 nm. Fluorescence intensity was recorded using an excitation wavelength of 580 ± 10 nm, and an emission wavelength of 610 ± 5 nm, with a gain set to 40. Readings were taken every 12 min. Antibiotics were used at the following concentrations: ampicillin (Ap): 50 µg/ml, chloramphenicol (Cm): 25 µg/ml, kanamycin (Km): 30 µg/ml. Na-malonate was administered at 2 mg/ml, cerulenin was used at 20 µg/ml, and β-alanine at 3 mM end concentrations. For characterization of the sensors' response to malonyl-CoA, IPTG was administered in a concentration-series of 0.01, 0.1, 0.3, 0.6, 1, and 10 mM. Arabinose was used in a log10 series from 10^{-4} to 10^{-1} % during optimization and at 10^{-2} % for comparison of alternate malonyl-CoA producer constructs. All chemicals were obtained from SIGMA (St. Louis, MO, USA).

IDENTIFICATION OF THE DYNAMIC MODEL OF THE SENSOR

In order to determine the dynamic response of the sensor, we consider three approximate dynamic models.

Zero-pole model

This model considers that the dynamics between biomass $X(t)$ and fluorescence $R(t)$ can be approximately described by an integral

time constant τ_p , a derivative time constant τ_z , and a gain K :

$$\tau_p \frac{dR(t)}{dt} = -R(t) + K \left(X(t) - \tau_z \frac{dX(t)}{dt} \right)$$

The transfer function between the input and the output expressed through the Laplace transform notation is:

$$\frac{R(s)}{X(s)} = K \frac{1 - \tau_z s}{1 + \tau_p s}$$

1 zero-2 poles model

In this model, we consider an additional integral time constant τ_{p2} between the input $X(t)$ and the output $R(t)$, represented in the following equation through an auxiliary state variable $X_1(t)$:

$$\begin{aligned} \tau_{p1} \frac{dX_1(t)}{dt} &= -X_1(t) + K \left(X(t) - \tau_z \frac{dX(t)}{dt} \right) \\ \tau_{p2} \frac{dR(t)}{dt} &= -R(t) + X_1(t) \end{aligned}$$

In Laplace transform notation:

$$\frac{R(s)}{X(s)} = K \frac{1 - \tau_z s}{(1 + \tau_{p1}s)(1 + \tau_{p2}s)}$$

1 zero-3 poles model

In this model, we consider an additional integral time constant τ_{p3} between the input $X(t)$ and the output $R(t)$, represented in the following equation through an additional state variable $X_2(t)$:

$$\begin{aligned} \tau_{p1} \frac{dX_1(t)}{dt} &= -X_1(t) + K \left(X(t) - \tau_z \frac{dX(t)}{dt} \right) \\ \tau_{p2} \frac{dX_2(t)}{dt} &= -X_2(t) + X_1(t) \\ \tau_{p3} \frac{dR(t)}{dt} &= -R(t) + X_2(t) \end{aligned}$$

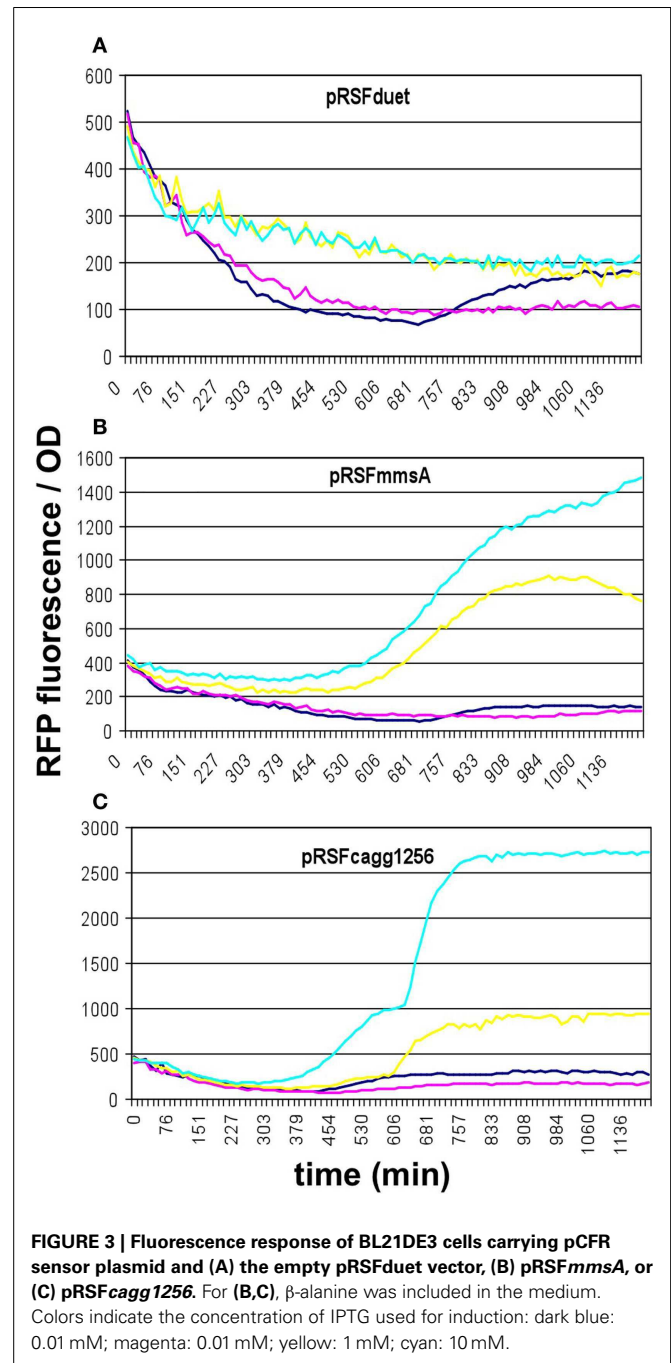
In Laplace transform notation:

$$\frac{R(s)}{X(s)} = K \frac{1 - \tau_z s}{(1 + \tau_{p1}s)(1 + \tau_{p2}s)(1 + \tau_{p3}s)}$$

In order to identify the model from the samples, we used an approximate discrete model based on the bilinear transform (Oppenheim and Schaffer, 2010):

$$s \leftarrow \frac{2}{T_s} \frac{z - 1}{z + 1}$$

where z^{-1} corresponds to a pure delay in the sample and T_s is the sampling frequency. Parameters of the discrete model were fitted by linear regression using the R package with the dynlm package (Zeileis, 2014), which was also employed to perform the simulations.



RESULTS

EXPERIMENTAL ASSESSMENT OF MALONYL-CoA SENSING

Malonyl-CoA is an important building block from the aspect of metabolic engineering, for it is required for the biosynthesis of fatty acids, polyketides, flavonoids, and other compounds (Fowler et al., 2009; Xu et al., 2011). In an earlier work, we have predicted and tested several alternative pathways that yield malonyl-CoA in order to boost pinocembrin production in *E. coli* (Fehér et al., 2014). Briefly, one candidate for malonyl-CoA synthase (*matB*, EC 6.2.1), one for malonate-CoA transferase (*atoDA*, EC 2.8.3.3),

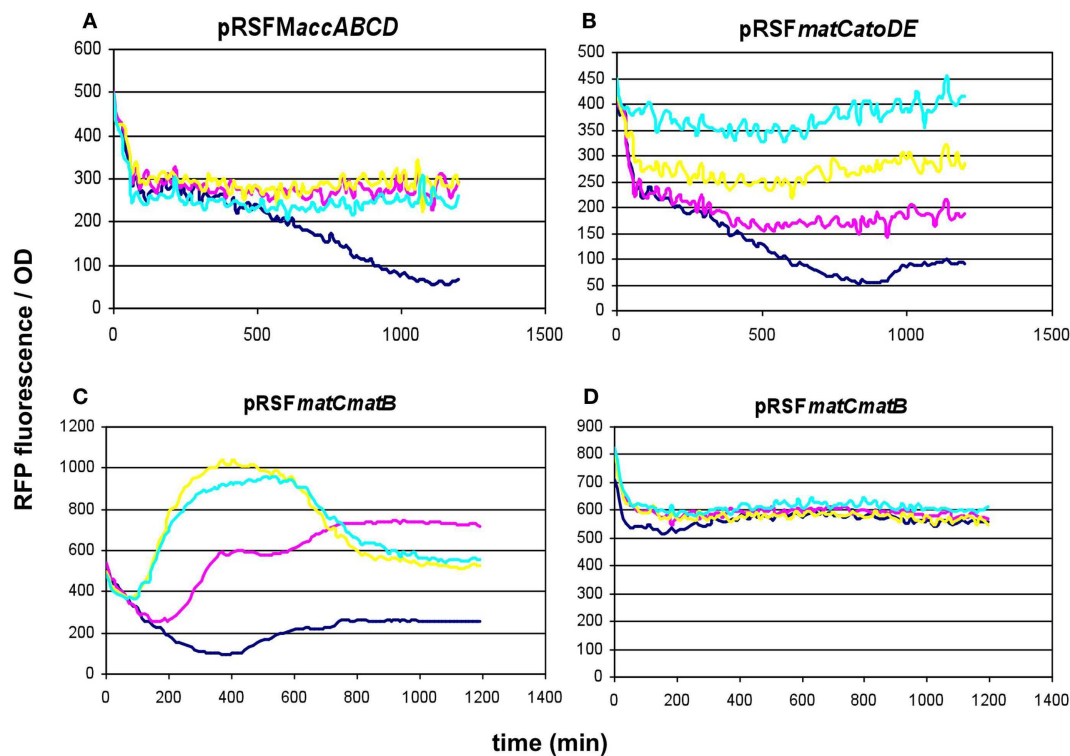


FIGURE 4 | Fluorescence response of BL21DE3 cells carrying pCFR sensor plasmid and (A) pRSFMaccABCD or (B) pRSFmatCatoDA. (C,D) show two distinct fluorescence responses to IPTG induction of cells carrying

pRSFmatCmatB. Na-malonate was included in the medium for (B–D). Colors indicate the concentration of IPTG used for induction: dark blue: 0.01 mM; magenta: 0.1 mM; yellow: 1 mM; cyan: 10 mM.

two for malonyl-CoA reductase (*mmsA* and *cagg1256*, EC 1.2.1.75) as well as one for the acetyl-CoA carboxylase complex (*accABCD* from *E. coli*, EC 6.4.1.2) were expressed together with the complete pinocembrin pathway, and their efficiencies were deduced from the resulting normalized pinocembrin titers. Here, we used the molecular sensor as a more direct method to investigate the levels of malonyl-CoA produced by the alternative enzymes.

Verifying the function of the malonyl-CoA sensor

The core of the malonyl-CoA sensor is an RFP gene driven by the synthetic pFR1 promoter (Liu et al., 2015). pFR1 is a combination of P_{A1} promoter from phage T7 (Deuschle et al., 1986) and two FapR-binding segments flanking the –10 region. The FapR protein, originating from *Bacillus subtilis*, dissociates from the DNA upon the binding of malonyl-CoA (Schujman et al., 2003), thereby allowing the *E. coli* RNA polymerase to transcribe the downstream sequences. The transcription of FapR is driven by a pAra promoter (controlled by the AraC protein), pFR1 is therefore increasingly repressible by elevating the L-arabinose levels. The AraC and FapR transcription factors, as well as RFP are all encoded on pBFR1k_RFP_8FapR, a kanamycin-resistant plasmid constructed by the workgroup of Fuzhong Zhang (Liu et al., 2015). The authors validated the quantitative nature of malonyl-CoA sensing by generating a calibration curve that correlated the fluorescence levels to LC-MS measurements of intracellular malonyl-CoA. In order to use this sensor to compare the malonyl-CoA productivity of our

construct collection available on the kanamycin-resistant pRSF vector, we generated a chloramphenicol-resistant version of the plasmid, which we call pCFR (plasmids used in this study are listed in Table S1 in Supplementary Material). The functionality of pCFR, as well as that of the original sensor plasmid was benchmarked with the help of cerulenin. This is an inhibitor of 3-ketoacyl-ACP synthase I and II (D'Agnolo et al., 1973), which participate in the committing step of free fatty acid biosynthesis. The inhibition of these enzymes results in a substantial intracellular accumulation of malonyl-CoA (Davis et al., 2000), and should therefore cause an increase in RFP fluorescence. Indeed, *E. coli* cells carrying either pBFR1k_RFP_8FapR or pCFR responded to cerulenin with a sharp increase in OD-normalized fluorescence (Figures 1A,B). As a matter of fact, a small scale screening of four pCFR-carrying clones allowed the identification of the most effectively functioning plasmid, which was chosen for all downstream experiments (Figure 1B). As expected, the ratio of RFP/OD values of cerulenin-treated to control cultures was higher in minimal medium than in LB broth (Figure 1C), due to the relatively high background-fluorescence of LB medium (Xu et al., 1999). For this reason, all further measurements were carried out in minimal medium.

Optimizing sensitivity toward malonyl-CoA detection

In its original publication by Liu et al. (2015), the malonyl-CoA sensor was characterized in a split form: the FapR and RFP

genes were carried by two separate plasmids. Therefore, we re-characterized pBFR1k_RFP_8FapR in order to find the optimal conditions of comparing various malonyl-CoA producer enzymes using the single-plasmid sensor construct. The expression level of the FapR repressor was varied by administering a dilution series of arabinose, while the expression level of the malonate transporter (*matC*) and the *matB* from plasmid pACYC*matCmatB* was controlled by altering the IPTG concentration. A 4 × 4 concentration matrix was thus applied to cells carrying either both plasmids or just the sensor plasmid. As apparent from the OD-normalized fluorescence values shown on **Figure 2**, an arabinose concentration of 0.01% proved to be optimal, taking both specificity and sensitivity into account, similarly to the originally published sensor system. At lower arabinose concentrations, specificity deteriorated, since the sensor plasmid produced slight increases in fluorescence in response to IPTG even in the absence of the malonyl-CoA producer plasmid. This could indicate the promoter's non-specific binding to LacI, which is competitively inhibited by sufficient quantities of the FapR protein. Increasing arabinose levels to 0.1% strongly reduced fluorescence levels of the sample, leading to a decrease of sensitivity and of the signal to noise ratio, possibly resulting from an over-repression by FapR. The initial decrease in fluorescence/OD values, seen on all graphs is likely due to the delay of RFP expression compared to culture growth.

Comparison of malonyl-CoA-producing construct collection

To compare the malonyl-CoA production efficiencies of alternative malonyl-CoA-producing pathways, at least one representative enzyme for each pathway was cloned, as described earlier (Fehér et al., 2014). Briefly, the *matB*, *atoDA*, *mmsA*, and *cagg1256* genes, as well as the *accABCD* gene complex were expressed from the high-copy plasmid pRSFduet. To provide sufficient amounts of substrate, a MatC was co-expressed with *matB* and *atoDA*, and the medium was supplemented with Na-malonate. To elevate the substrate levels for *mmsA* and *cagg1256*, cultures were supplemented with β-alanine, which is converted to 3-oxopropanoate (malonate semialdehyde) by the cell's endogenous 4-aminobutyrate aminotransferase. In every case, the fluorescence/OD values were monitored over time after growing the cells in various IPTG concentrations. Cells carrying the empty pRSFduet plasmid (besides the sensor plasmid) were used as a negative control. Based on their response, the strains carrying the various constructs fell into either one of the two following categories: (i) the cells exhibited growth and a dose-dependent fluorescence/OD response to IPTG (pRSF*mmsA* and pRSF*cagg*) (**Figure 3**) or (ii) the cells exhibited absolutely no growth upon IPTG induction in minimal medium (pRSF*macABCD* and pRSF*atoDA*) (**Figures 4A,B**). It is important to note that certain cells of the latter category also gave a seemingly dose-dependent response to IPTG (such as pRSF*atoDA*), but their fluorescence and OD values were practically unchanged during the course of the experiment, most probably indicating that the trend in their ratio is an artifact (Figure S1 in Supplementary Material). Interestingly, both types of responses were seen with pRSF*matCmatB*-carrying cells, in a quite irreproducible manner (see Discussion and **Figures 4C,D** and Figure S2 in Supplementary Material). The growth curves obtained by inducing strains

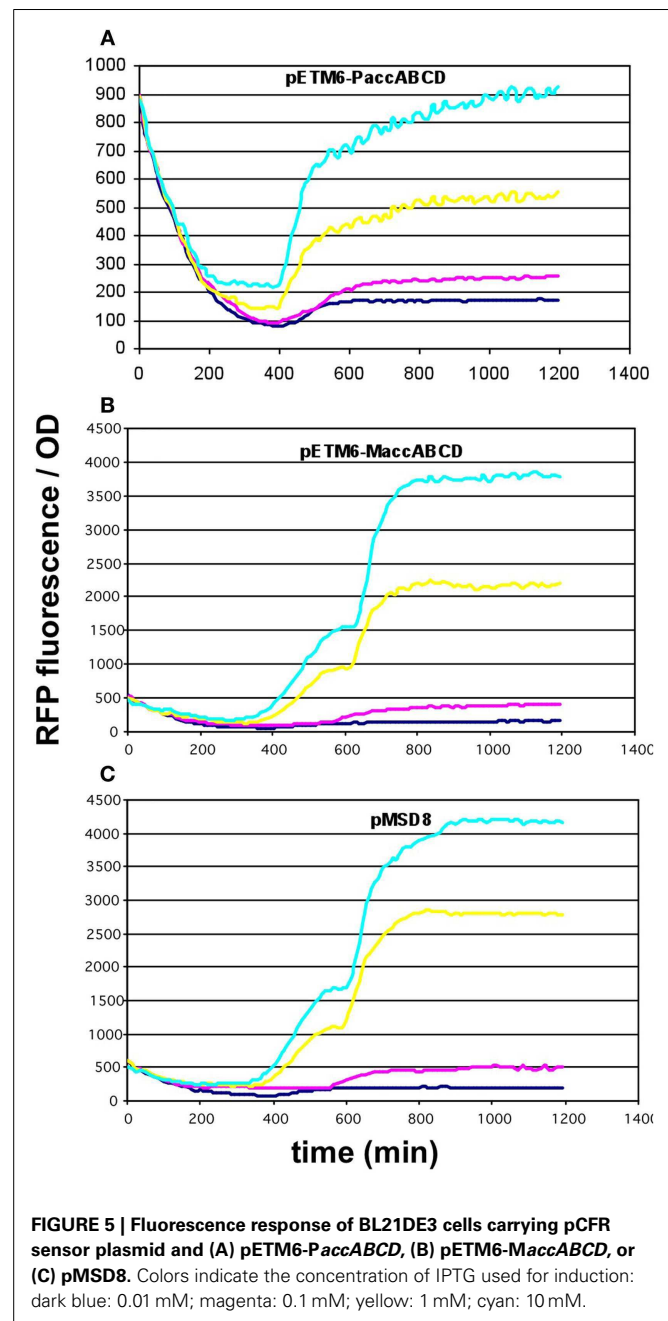


FIGURE 5 | Fluorescence response of BL21DE3 cells carrying pCFR sensor plasmid and (A) pETM6-PaccABCD, (B) pETM6-MaccABCD, or (C) pMSD8. Colors indicate the concentration of IPTG used for induction: dark blue: 0.01 mM; magenta: 0.1 mM; yellow: 1 mM; cyan: 10 mM.

carrying various producer plasmids are summarized on Figure S3 in Supplementary Material.

Investigating lower-copy-number alternative constructs

Since the overexpression of certain members of the *accABCD* complex (Davis et al., 2000), as well as protein overproduction in general (Flores et al., 2004) have been described to have a toxic effect on the cells, the overexpression of certain constructs was repeated using lower-copy vectors. For the *accABCD* complex, pMSD8, an ultra-low-copy variant was available expressing the four genes as a T7-driven operon. In addition, two medium-copy alternatives, pETM6-PaccABCD and

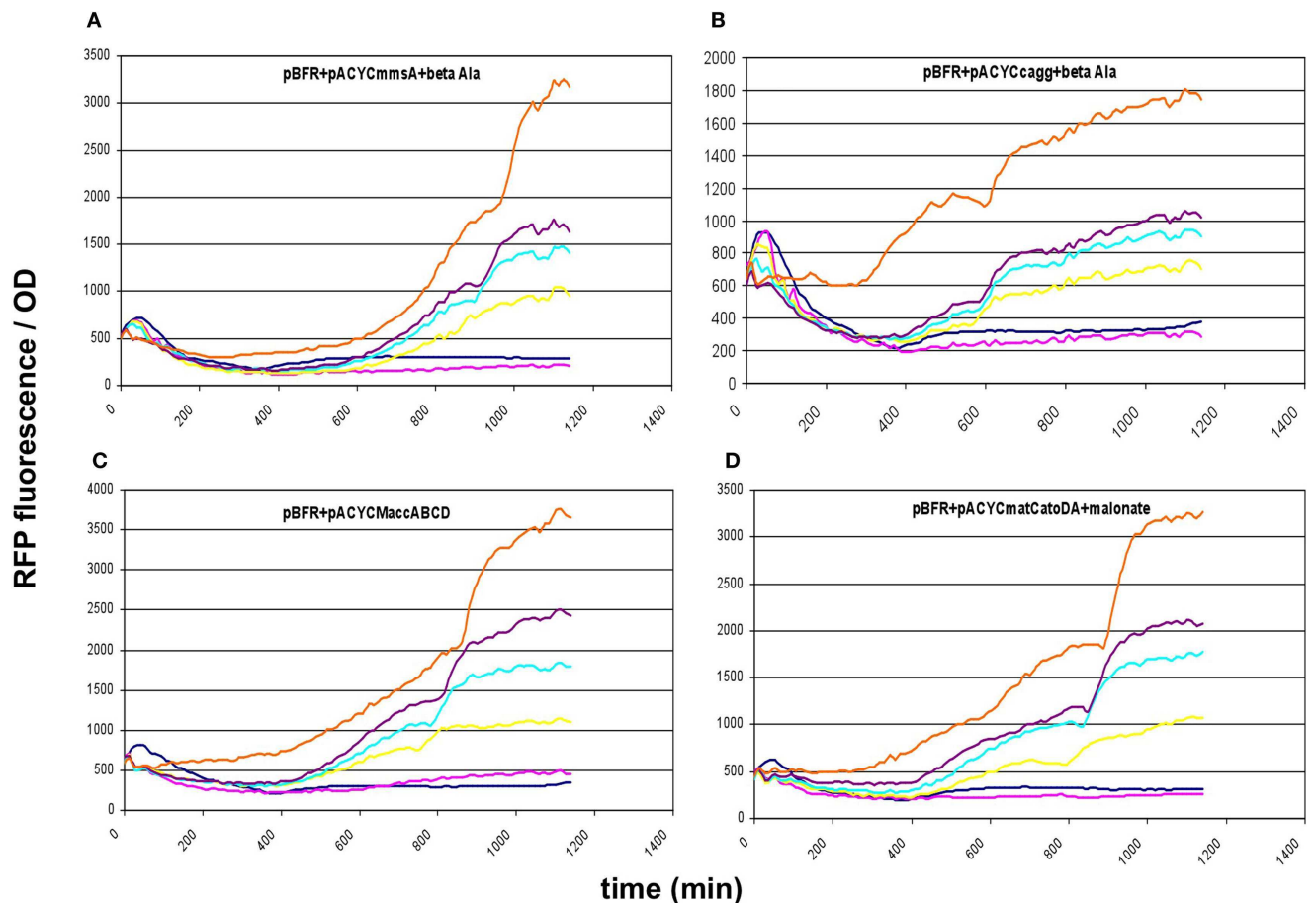


FIGURE 6 | Fluorescence response of BL21DE3 cells carrying pBFR1k_RFP_8FapR sensor plasmid and (A) pACYCmmsA, (B) pACYCcagg1256, (C) pACYCmaccABCD, or (D) pACYCmatCatoDA. The medium was supplemented with β -alanine for pACYCmmsA and

pACYCcagg1256, and with Na-malonate for pACYCmatCatoDA. Colors indicate the concentration of IPTG used for induction: dark blue: 0.01 mM; magenta: 0.1 mM; yellow: 0.3 mM; cyan: 0.6 mM; purple: 1 mM; orange: 10 mM.

pETM6-MaccABCD, carrying the genes as a pseudo-operon (separate promoters) and in monocistronic form (separate promoters and terminators), respectively, were also obtained and tested (Xu et al., 2012, 2013). All three plasmids displayed a clear dose-dependent fluorescence/OD pattern, well above the levels of the negative control (Figures 5A–C). For *matCmatB*, pACYCmatCmatB was readily available as a low-copy alternative that functioned in a quite well-reproducible fashion (Figure 2), when tested with the pBFR1k_RFP_8FapR sensor plasmid.

These results prompted us to subclone our pathway collection into a pACYC backbone, thereby reducing the copy-number of all constructs to ~ 15 . When expressed this way, all five constructs (*matCmatB*, *mmsA*, *cagg1256*, *M. accABCD*, and *matCatoDA*) allowed cell growth and displayed a scalable, IPTG-dependent fluorescence (Figures 2 and 6). As it turned out, the concise nature of the dataset obtained with the pACYC-collection allowed us to fit various models describing the response of the sensor to malonyl-CoA production (see below).

Quantification and reproducibility of malonyl-CoA levels

Since *E. coli* strain BL21DE3, the host used in our experiments is not deleted for the genes of arabinose catabolism, it breaks down arabinose when the medium is depleted for glucose. This likely causes a decrease in FapR expression, and a consequent derepression of RFP at the end of growth. As a result, a second, sudden increase in fluorescence was seen in many samples upon the transition to stationary phase, as apparent from juxtaposing the temporal evolution of fluorescence and OD (Figure S4 in Supplementary Material). To avoid this effect, we analyzed fluorescence levels corresponding to the log-phase ($OD = 0.6$), a physiological state when arabinose catabolism is still inhibited by the catabolite repression of *E. coli*. Figure S5 in Supplementary Material shows some of the fluorescence vs. OD plots generated to obtain such readings.

Prior to comparing the performance of various malonyl-CoA-producing constructs, we investigated the reproducibility of our measurements at three different levels. First, we compared the fluorescence values of parallel clones, resulting from

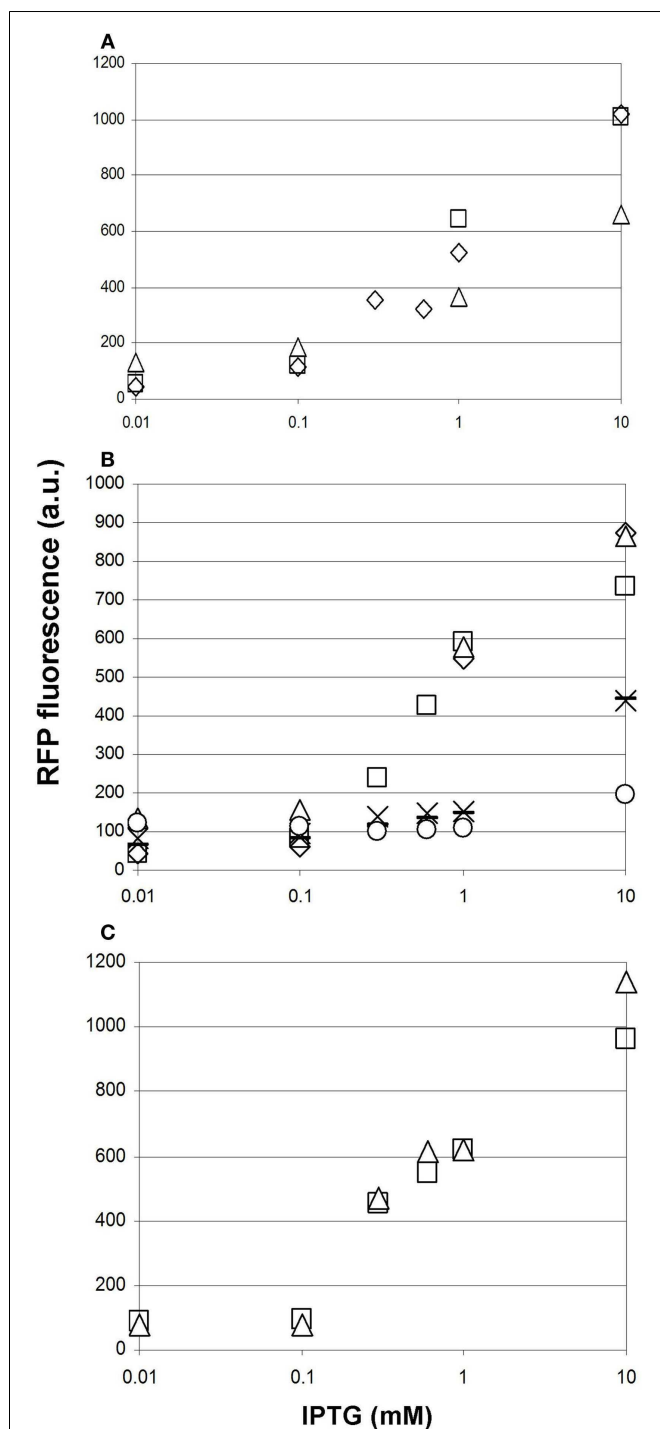


FIGURE 7 | Reproducibility of measurements of malonyl-CoA production in BL21DE3 cells using the fluorescent sensor plasmid pCFR and various production plasmids. The IPTG-dependence of fluorescence measured at OD = 0.6 is shown in every case. **(A)** Two cultures harboring pMSD8, originating from two distinct colonies after transformation are depicted by open squares and open triangles. Open diamonds represent the re-measurement of the strain corresponding to the open squares, 1 day later. **(B)** Six cultures harboring ETM6-MaccABCD measured on different days. **(C)** Two cultures of the same clone of pACYCmatCmatB originating from the same starter culture, measured in the same run, in the presence of Na-malonate.

the transformation of a given producer construct. These displayed a strong variation in their response to IPTG, which was not unexpected taking into account their heterologous colony morphologies (**Figure 7A**). Second, we re-tested the best performing clones after their storage in the form of glycerol stocks. Reproducibility was substantially better, being the best for the lowest copy-number pMSD8 plasmid (**Figure 7A**). For other constructs, the performance seemed to fall into two distinct clusters: it either matched that of earlier measurements, or deteriorated to a lower, near-zero value, as demonstrated by the dose–response curves of pETM6-MaccABCD (**Figure 7B**). This phenomenon was reminiscent of a burden-relieving mutation that swept through the population early after induction. The third level of reproducibility corresponds to comparing the performance of a given strain in the same microtiter plate in the same measurement-run, starting from the same overnight seed-culture (**Figure 7C**). The high reproducibility of such measurements was sufficient to draw conclusions on the effect of providing the substrates for malonyl-CoA production (see below).

Investigating the effect of administering the substrate

To test whether providing the substrates of malonyl-CoA synthesis in the growth medium offers a measurable advantage in the product levels, we repeated the experiment for, pACYCmatCmatB and pACYCmatCatoDA with and without Na-malonate as well as for pRSFmmsA and pRSFcagg with and without β -alanine. For pACYCmatCmatB, we observed significantly boosted production at 0.3 and 0.6 mM IPTG (**Figure 8A**). However, the presence of Na-malonate significantly hindered the performance of pACYCmatCatoDA at IPTG concentrations of 0.3 mM and above (**Figure 8B**). Similarly, supplementing β -alanine to pRSFmmsA seemed to have a negative effect at 0.6 mM IPTG induction (**Figure 8C**). No significant effect was seen for adding the substrate to pRSFcagg (not shown).

Investigating the efficiencies of malonyl-CoA production

The four alternative pathways for malonyl-CoA production, as well as the top-scoring enzymes that catalyze the required steps were sought for by RetroPath, our retrosynthetic biology tool developed for metabolic engineering (Carbonell et al., 2011). As opposed to our initial plans, the growth-inhibitory effect of certain pRSF-based constructs prohibited us to validate the ranking of the alternative pathways by direct comparison of their measured fluorescence values. Indeed, the toxicity of overproducing malonyl-CoA by overexpression of acetyl-CoA carboxylase has been described before (Davis et al., 2000). The performance of the readily available low-copy variants of accABCD, however (pMSD8, pETM6-PaccABCD, and pETM6-MaccABCD), indicated the importance of using lower-copy expression vectors. After recloning our gene-set into pACYC-derived plasmids, matCatoDA, MaccABCD, mmsA, cagg1256, and matCmatB all permitted growth when expressed. Importantly, based on their fluorescence values, these constructs all produced significantly more malonyl-CoA when induced with 1 mM of IPTG, than the control cell harboring no production (see below). Ranking the alternative malonyl-CoA-producing pathways was not feasible due to the high variation of the data. Important practical conclusions could be nevertheless made, listed in the Section “Discussion.”

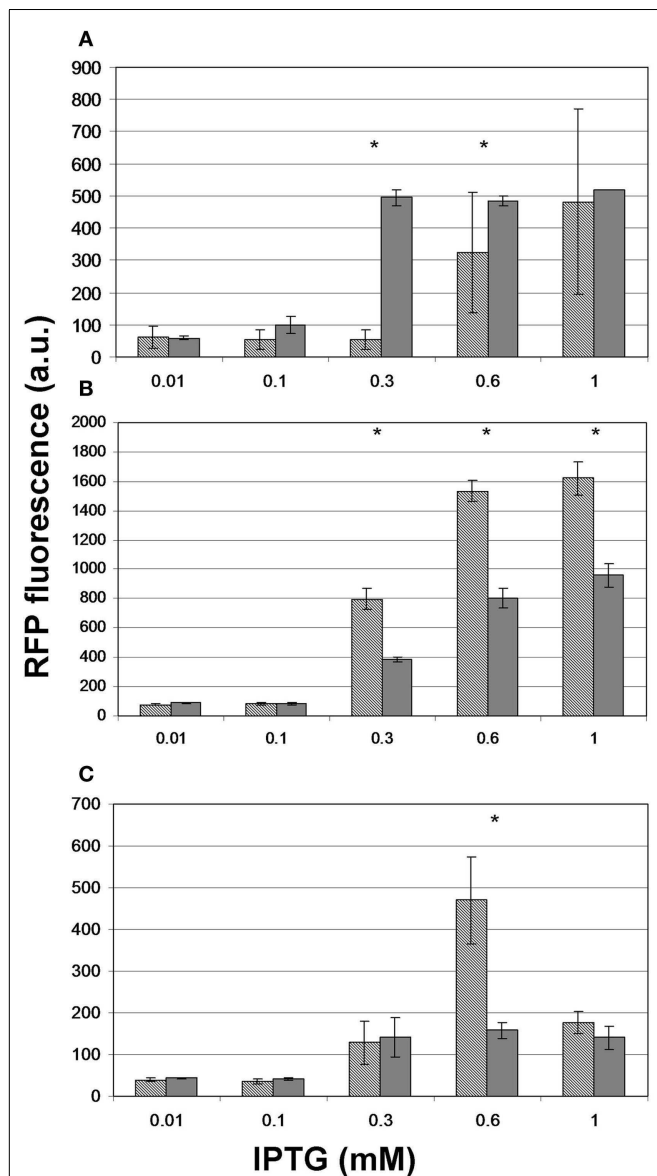


FIGURE 8 | The effect of administering the substrate on the level of fluorescence obtained upon induction of malonyl-CoA production. Hashed bars represent the effect of IPTG on the fluorescence obtained at OD = 0.6 with constructs **(A)** pACYC $matCmatB$, **(B)** pACYC $matCatoDA$, and **(C)** pRSF $mmsA$. Gray bars depict the peak fluorescence detected when inducing the same constructs in the presence of their substrate: Na-malonate for $matCmatB$ and $matCatoDA$ and β -alanine for $mmsA$. The sensor plasmid was pBFR1k_RFP_8FapR for **(A,B)** and pCFR for **(C)**. Asterisks indicate a significantly different fluorescence caused by the substrate ($p < 0.05$).

MODELING THE MALONYL-CoA SENSOR

Characterization of the sensor response

In order to characterize the sensor, we modeled the sensor response to the production of malonyl-CoA, its measured signal. Ideally, the sensor once calibrated should display a good reproducibility as well as a stable sensitivity for its use in applications such as screening or feedback regulation. For that purpose, it is desirable to keep the

dynamic response of the sensor uncoupled from any parameter variation so that the relationship with the measured signal and its output corresponds basically to a linear one. The time of response of such linear sensor, on the other hand, should be kept below some reasonable limit in order to avoid large lags introducing potential instability into the system. Here, as we are basing our model on the observations of OD and RFP, we need to consider in our model the relation between these variables with respect to the malonyl-CoA concentration, the signal that is measured by the sensor. In this study, the model necessarily will be kept simple, as our purpose is to provide a set of parameters that can be easily measured in order to calibrate the sensor.

Similarly to other proposed models (Anesiadis et al., 2008), we consider that the signal undergoes several transformation stages in cascade-mode going from the biomass to the fluorescence levels:

1. We assume that the variation on concentration of malonyl-CoA $M(t)$ depends on its instant production, given by the rate v_m , which is specific to each selected malonyl-CoA-producing construct or pathway, multiplied by the biomass $X(t)$ minus the malonyl-CoA that is consumed for growth defined as a constant rate v_g that multiplies the variation in time of biomass and another term for degradation, defined by a rate constant γ_m :

$$\frac{dM(t)}{dt} = v_m X(t) - v_g \frac{dX(t)}{dt} - \gamma_m M(t) \quad (1)$$

During the exponential growth phase, this previous equation is simplified as:

$$\frac{dM(t)}{dt} = (v_m - \mu v_g) X(t) - \gamma_m M(t) \quad (2)$$

where μ is the specific growth rate.

2. Malonyl-CoA binds to transcription factor FapR to form a complex whose concentration $C(t)$ depends on a dissociation constant K_d [2.4 μ M according to Schujman et al. (2006)]. The concentration of FapR $F(t)$ is considered constant depending on the concentration levels of the inducer L-arabinose:

$$F + M \xrightleftharpoons[k_r]{k_f} C$$

$$\frac{dC(t)}{dt} = k_f FM(t) - k_r C(t) \quad (3)$$

3. The change on concentration of RFP $R(t)$, depends on the concentration of the complex $C(t)$ through a constant κ associated to the promoter strength (Oyarzún and Stan, 2013) and a decay constant γ_r :

$$\frac{dR(t)}{dt} = \kappa C(t) - \gamma_r R(t) \quad (4)$$

Under the assumptions given by the model described by Eqs 1–4, the response of the sensor can be approximated as a cascade of filters as shown in **Figure 9**. The advantage of using such approximation is that we can assume that the dynamics

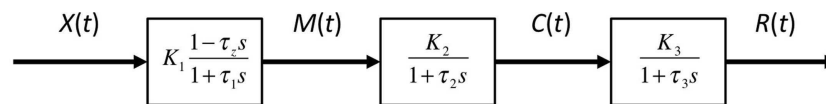


FIGURE 9 | An approximate model of the response behavior of the sensor based on a cascade of first-order filters. The biomass $X(t)$ is assumed to be approximately related to the concentration of malonyl-CoA $M(t)$ through a first-order pole-zero filter. The sensor complex malonyl-CoA-FapR concentration $C(t)$ is then related to $M(t)$ through a first-order filter and finally, the RFP concentration $R(t)$ is again related to $C(t)$ through another first-order filter.

due to the time constants of these three filters occur at different time scales and therefore, under the appropriate conditions, time-scale separations can be performed. Based on this principle, we can approximate the model by considering main time constants present in the model, i.e., by estimating parameters of a cascade of first-order models (see Materials and Methods).

Sensor model fitting

In our sensor model, we are assuming that under same inducer and regulatory conditions on a strain population, malonyl-CoA levels should change basically because of the change in the rate of production of this metabolite by the strain (v_m in Eq. 1), which would correspond to a different gain of the sensor, i.e., a different measured RFP and biomass X in steady state.

Such gain according to the sensor model in Eqs 1–4, the gain or relationship in steady state between RFP and X is given by the following relationship between model parameters:

$$\frac{R}{X} = K_d \frac{v_m}{\gamma_m \gamma_r} \quad (5)$$

Regarding model dynamics, we considered three simplified models for the response dynamics depending on the time constants that are considered (see Materials and Methods): (1) a first-order with integral and derivative terms; (2) a second-order model consisting of a first-order model with integral and derivative terms and an additional first-order integral term; and (3) a third-order model with the integral and derivative terms and two additional first-order integral terms.

As shown in **Figure 10**, measured responses were in general fitted with an increasing level of accuracy to each of the three approximate sensor models. For some constructs, however, measurements could not be successfully fitted to model 3 (see Figure S4 in Supplementary Material). Therefore, we decided to use model 2 (the one considering one derivative and two integral terms) as the reference model in order to characterize the dynamics of the sensor. **Table 1** provides a list of estimated parameters for different constructs. Time constants ranged between approximately 1 and 10 h.

Dependence of sensor parameters on induction

According to our simplified model of the sensor consisting on a second-order transfer function, gain between biomass and RFP should depend on the level of induction of the malonyl-CoA-producing enzyme. Time constants, on the contrary, should not depend on the level of induction, but should be constitutive parameters of the sensor that can be calibrated for each construct. As shown in **Figure 11**, we found effectively a good

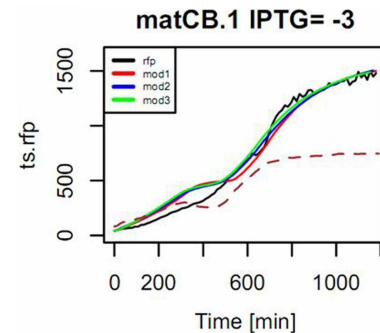


FIGURE 10 | Example of the mode of fitting models to describe the response of the pBFR1k_RFP_8FapR sensor in the presence of pACYCmatCmatB, induced with 1 mM of IPTG. Na-malonate was included in the medium. The input signal is the measured OD (shown in dotted line and scaled to 50% of maximum in the plot) and the simulated response of the model for the observed output RFP (in black) is shown for a fitting to a first-order model with input derivative (red), a second-order with input derivative (blue), and a third-order with input derivative (green).

Table 1 | Gain and integral time constants of the sensor fitted to model 2 estimated for different constructs for IPTG = 1×10^{-5} .

Construct	K	τ_1	τ_2
matCmatB	250.84	290.11	12.69
mmsA	300.53	288.31	15.67
cagg	362.82	525.90	17.39
acc	320.20	417.92	11.49
matCatoDA	321.12	156.95	11.14

correlation between IPTG levels and sensor gain ($r = 0.72$, p -value = 8.1×10^{-6}), indicating that the sensor is responsive to changes in induction of the enzyme. Time constants, as expected, were not significantly affected by changes in the IPTG concentration. These time constants, on the other hand, appear as distinctively clustered for each construct, showing the robustness of the sensor (**Figure 12**).

DISCUSSION

In this work, we have successfully implemented a malonyl-CoA sensor for the partial optimization of malonyl-CoA production in *E. coli* BL21DE3. Initially, we screened several conditions to be used for the measurements, and found the optimal one to match that described earlier (Liu et al., 2015), despite the possible changes in relative gene dosage caused by the single-plasmid vs.

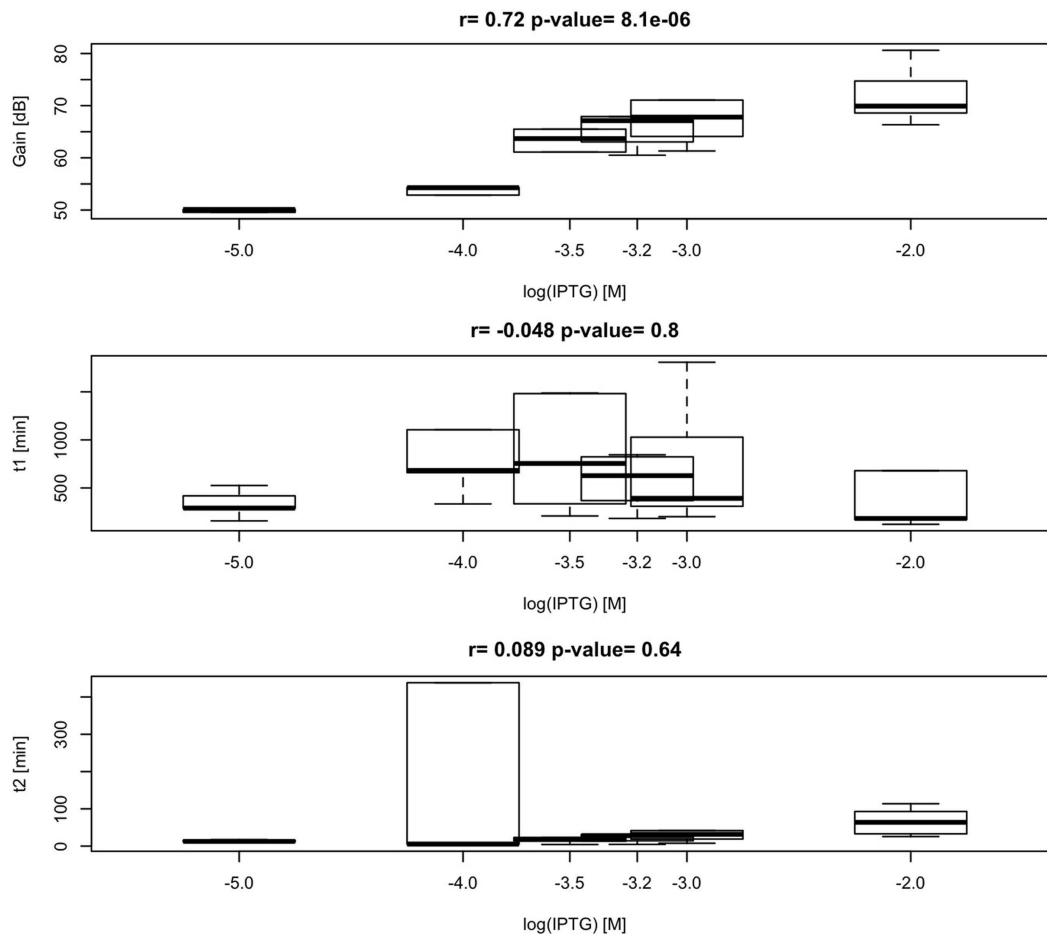


FIGURE 11 | Variability and correlation between model parameters K , τ_p , τ_z of the sensor, and IPTG levels 0.01, 0.1, 0.3, 0.6, 1, and 10 mM. Calculated correlation between parameters K , τ_p , τ_z , and IPTG levels are shown.

dual-plasmid layout. We constructed a chloramphenicol-resistant version of the plasmid as well, to compare our construct collection encoding four alternative pathways for malonyl-CoA production, predicted by RetroPath. Besides their comparison, our aim was to use malonyl-CoA sensing to seek optimal conditions for efficient malonyl-CoA production, previously seen to be hindered by growth retardation (Fehér et al., 2014). Our current results also indicated, that the viability of the cells can be, in many cases strongly impaired by the high expression levels of the original pRSF-based constructs (Figure S3 in Supplementary Material). It was also apparent, that the lower-copy alternatives available for the *accABCD* gene complex (pMSD8, pETM6-*MaccABCD*, and pETM6-*PaccABCD*) outperformed all or most members of the RSF-collection. Therefore, a partial optimization of expression levels was carried out by using smaller copy-number vectors to obtain viable cells with measurable malonyl-CoA production upon induction, leading to a successful circumvention of the possible toxicity caused by the high-copy expression vectors. Due to the relatively high variation of the resulting data, comparison of these constructs was not conclusive enough to validate

all our predicted rankings. However, the obtained measurements were quite useful for the practical purpose of selecting the most efficient implementation of our malonyl-CoA-producing constructs. From this aspect, the best performer of our construct set (in terms of fluorescence at $\text{OD} = 0.6$) turned out to be pACYC*matCatoDA* carrying the malonate-CoA transferase on a P15a-derived low-copy-number plasmid (Figure 13). Its performance could be nonetheless matched by the expression of the acetyl-CoA carboxylase complex, which, in turn was indiscernible from the *matB*. The two genes encoding malonyl-CoA reductase (*mmsA* and *cagg1256*) performed significantly weaker. Importantly, the activity of these constructs uncovered two pathways, malonyl-CoA reductase and malonate-CoA transferase, which have not been implemented before in boosting malonyl-CoA biosynthesis.

When comparing the variation of measured fluorescence values, we observed a consistent decrease in the coefficient of variation for all constructs when recloned into the low-copy pACYC vector, possibly marking an increased genetic stability (Table S2 in Supplementary Material). In theory, this change could

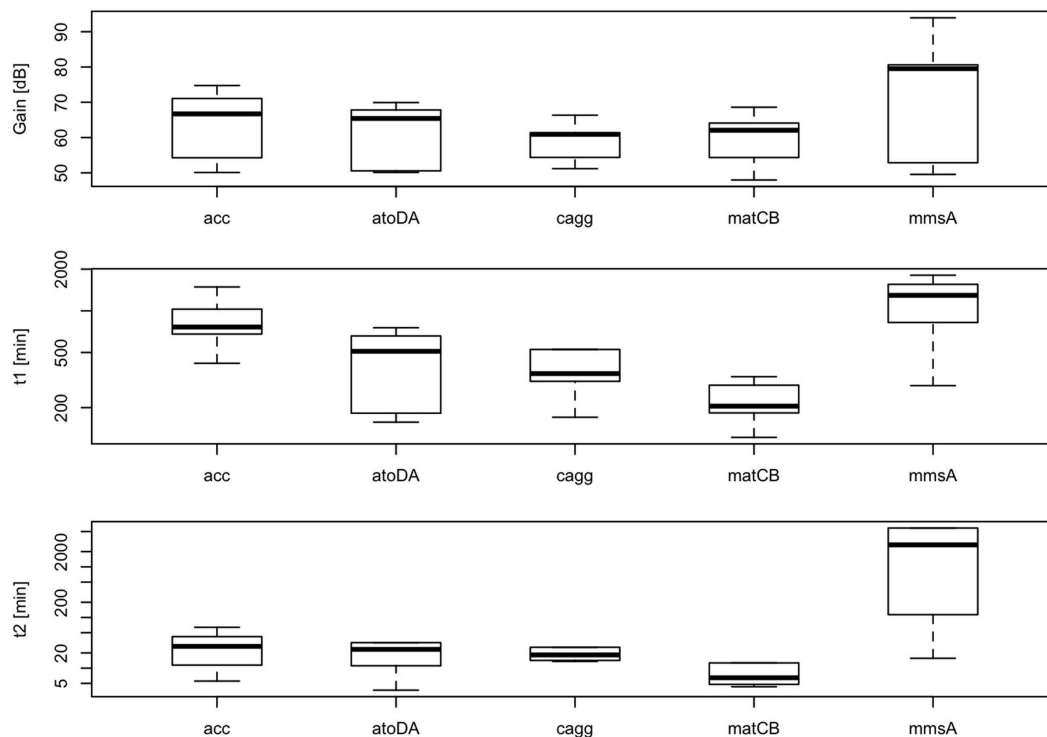


FIGURE 12 | Variability of parameters K , τ_p , τ_z of the sensor for each construct at several IPTG concentrations (0.01, 0.1, 0.3, 0.6, 1, and 10 mM). Calculated correlation between parameters K , τ_p , τ_z , and IPTG levels are shown.

also be caused by the fact of using different sensor plasmids in the comparison. However, no such difference was seen in the case of the pET-based plasmids when switching from pCFR to pBFR1k_RFP_8FapR used for sensing (Table S2 in Supplementary Material), indicating that it is the change in copy-number of the malonyl-CoA producer plasmid, and not the change in the sensor plasmid that caused the decrease of variation. This variation can be, at least in part, explained by burden-relieving mutations that sweep through the population early after induction. In fact, such an expansion of a non-fluorescent subpopulation was seen real time during one of our tests concerning pRSFmatCmatB (Figure 4C), which was the least reproducible among all of our constructs. In that specific experiment, the fluorescence of the population ceased to grow after a certain point of time, despite the accelerating increase of the OD, leading to a decrease of the fluorescence/OD ratio (Figure S2 in Supplementary Material). The fact that no such effect is seen in Figure 4D indicates the random and incidental nature of this phenomenon, supporting the assumption that it is caused by a mutation. The specific mutations inactivating the production were not sought for.

In their original publication, the designers of this sensor detected a saturation of malonyl-CoA levels at 25 μ M, when inducing the expression of the *acc* operon at different strengths (Liu et al., 2015). This phenomenon was not investigated further, but could have been caused by the toxicity of the produced compound. Although we did not observe an unambiguous plateau of fluorescence values in our experiments, the possibility of such a scenario

should be taken into account when using similar systems in the future. One possible solution for such cases could be to transform the molecule of interest into a compound *in vivo* that is better tolerated by the cell. The prerequisites of this strategy are to have a high-capacity, unsaturated pathway for transformation as well as a sensory device for the downstream product.

To characterize further the dynamic response of the sensor, we proposed and validated a second-order model linking biomass to fluorescence. We showed that dynamic parameters were specific to each construct and that they were not significantly altered by changes in the production rate of malonyl-CoA. This result shows the robustness of the sensor for its use in pathway regulation, since a dynamic response uncoupled from the measured signal is often necessary in order to assure the stability of the feedback loop. Moreover, in such regulation strategies, knowledge of the sensor's parameters is of critical importance in the choice of one particular control architecture over another (Stevens and Carothers, 2014). Therefore, we hope our effort of modeling experimental data provided by this malonyl-CoA sensor will facilitate future developments in dynamic regulation of pathways such as the ones involved in fatty acid or flavonoid production.

Our most unexpected findings arose when we investigated the effect of providing the substrates of malonyl-CoA production on the fluorescence levels produced by the sensor. On one hand, providing Na-malonate to *matB*, resulted in an increased fluorescence, possibly indicating more product formation. On the other hand however, administering β -alanine or Na-malonate to *mmsA* and

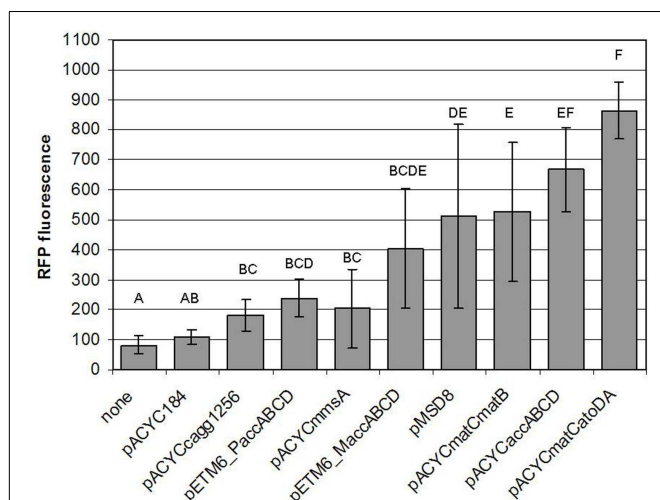


FIGURE 13 | Comparison of the malonyl-CoA levels produced by various constructs. Fluorescence observed at OD = 0.6, generated by pBFR1k_RFP_8FapR upon induction of the corresponding producer plasmid with 1 mM IPTG in a BL21DE3 host is depicted. Na-malonate was added for pACYCmatCmatB and pACYCmatCatoDA, and β -alanine was administered for pACYCmmsA and pACYCagg1256. Each letter represents a value significantly differing from the others [i.e., if two values were found to significantly differ from each other ($p < 0.05$), then the letters corresponding to their bars do not overlap].

atoDA, respectively resulted in an opposite effect. This phenomenon was especially pronounced for *atoDA*, which was notably the most effective malonyl-CoA producer among the constructs tested in this study. We can only speculate on the mechanism, which may even be a direct effect on the sensor plasmid, and may not represent an actually reduced malonyl-CoA synthesis. According to our hypothesis, it could be due to the fact that *atoDA* (as well as *matB*) are consumers of cellular acetyl-CoA, thereby causing an energy depletion. This is probably more severe if we provide the missing substrate (malonate), with very high enzyme concentrations, and thereby push the reaction toward the product. Perhaps, this energy depletion inhibits RFP production more than the newly produced malonyl-CoA would elevate it. Clearing this issue nevertheless requires further experiments.

When summarizing the observed results, several conclusions important to the metabolic engineer in general can be drawn. First, it is useful to test the performance of several parallel clones after transformation of the constructs, and choose the best for further experiments. Second, the reproducibility of cultures re-grown from glycerol stocks is acceptable, but repeated measurements are advisable to uncover the repeated emergence of deleterious mutants, possibly indicating genetic instability. Third, this instability was minimal when using the lowest copy-number vector for expressing the producer construct. This provides a further argument to start testing various alternative enzymes by expressing them from low-copy vectors with gradually controllable promoters, and varying ribosome binding sites to increase expression and find a construct that is optimal both in efficiency and stability. And finally, the variation among measurements obtained in the same run were small enough to test the effect of adding the substrates

of the tested pathways to the cell culture, and obtain the optimal IPTG levels for most effective utilization. This approach could be useful in the future to optimize the substrate concentration itself, as well as any other component of the culture medium or the induction process in a high throughput, combinatorial fashion.

ACKNOWLEDGMENTS

We thank Prof. John Cronan, Prof. Mattheos Koffas, Prof. Fuzhong Zhang, and Dr. Jingwen Zhou for providing their plasmids and Dr. Brian Jester for helpful discussions. PC is supported by Genopole through an ATIGE Grant by PRES UniverSud Paris, by Agence Nationale de la Recherche, and by UPFellows program with the support of the Marie Curie COFUND program. VL is supported by a DGA (French Ministry of Defense) graduate fellowship. The authors acknowledge funding provided by Toulouse White Biotech (TWB) and the GIP GENOPOLE.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at <http://journal.frontiersin.org/article/10.3389/fbioe.2015.00046>

REFERENCES

- Anesiadis, N., Cluett, W. R., and Mahadevan, R. (2008). Dynamic metabolic engineering for increasing bioprocess productivity. *Metab. Eng.* 10, 255–266. doi:10.1016/j.ymben.2008.06.004
- Binder, S., Schendzielorz, G., Stäbler, N., Krumbach, K., Hoffmann, K., Bott, M., et al. (2012). A high-throughput approach to identify genomic variants of bacterial metabolite producers at the single-cell level. *Genome Biol.* 13, R40. doi:10.1186/gb-2012-13-5-r40
- Carbonell, P., Planson, A.-G., Fichera, D., and Faulon, J.-L. (2011). A retrosynthetic biology approach to metabolic pathway design for therapeutic production. *BMC Syst. Biol.* 5:122. doi:10.1186/1752-0509-5-122
- Chubukov, V., Gerosa, L., Kochanowski, K., and Sauer, U. (2014). Coordination of microbial metabolism. *Nat. Rev. Microbiol.* 12, 327–340. doi:10.1038/nrmicro3238
- D'Agnolo, G., Rosenfeld, I. S., Awaya, J., Omura, S., and Vagelos, P. R. (1973). Inhibition of fatty acid synthesis by the antibiotic cerulenin. Specific inactivation of beta-ketoacyl-acyl carrier protein synthetase. *Biochim. Biophys. Acta* 326, 155–156. doi:10.1016/0005-2760(73)90241-5
- Davis, M. S., Solbiati, J., and Cronan, J. E. (2000). Overproduction of acetyl-CoA carboxylase activity increases the rate of fatty acid biosynthesis in *Escherichia coli*. *J. Biol. Chem.* 275, 28593–28598. doi:10.1074/jbc.M004756200
- Deuschle, U., Kammerer, W., Gentz, R., and Bujard, H. (1986). Promoters of *Escherichia coli*: a hierarchy of in vivo strength indicates alternate structures. *EMBO J.* 5, 2987–2994.
- Dietrich, J. A., McKee, A. E., and Keasling, J. D. (2010). High-throughput metabolic engineering: advances in small-molecule screening and selection. *Annu. Rev. Biochem.* 79, 563–590. doi:10.1146/annurev-biochem-062608-095938
- Dietrich, J. A., Shis, D. L., Alikhani, A., and Keasling, J. D. (2013). Transcription factor-based screens and synthetic selections for microbial small-molecule biosynthesis. *ACS Synth. Biol.* 2, 47–58. doi:10.1021/sb300091d
- Farmer, W. R., and Liao, J. C. (2000). Improving lycopene production in *Escherichia coli* by engineering metabolic control. *Nat. Biotechnol.* 18, 533–537. doi:10.1038/75398
- Fehér, T., Planson, A.-G., Carbonell, P., Fernández-Castané, A., Grigoras, I., Dariy, E., et al. (2014). Validation of RetroPath, a computer-aided design tool for metabolic pathway engineering. *Biotechnol. J.* 9, 1446–1457. doi:10.1002/biot.201400055
- Flores, S., de Anda-Herrera, R., Gosset, G., and Bolívar, F. G. (2004). Growth-rate recovery of *Escherichia coli* cultures carrying a multicopy plasmid, by engineering of the pentose-phosphate pathway. *Biotechnol. Bioeng.* 87, 485–494. doi:10.1002/bit.20137
- Fowler, Z. L., Gikandi, W. W., and Koffas, M. A. G. (2009). Increased malonyl coenzyme A biosynthesis by tuning the *Escherichia coli* metabolic network and its application to flavanone production. *Appl. Environ. Microbiol.* 75, 5831–5839. doi:10.1128/AEM.00270-09

- Hall, B. G. (1998). Activation of the *bgl* operon by adaptive mutation. *Mol. Biol. Evol.* 15, 1–5. doi:10.1093/oxfordjournals.molbev.a025842
- Holtz, W. J., and Keasling, J. D. (2010). Engineering static and dynamic control of synthetic pathways. *Cell* 140, 19–23. doi:10.1016/j.cell.2009.12.029
- Liu, D., Xiao, Y., Evans, B., and Zhang, F. (2015). Negative feedback regulation of fatty acid production based on a malonyl-CoA sensor-actuator. *ACS Synth. Biol.* 4, 132–140. doi:10.1021/sb400158w
- Michener, J. K., and Smolke, C. D. (2014). Synthetic RNA switches for yeast metabolic engineering: screening recombinant enzyme libraries. *Methods Mol. Biol.* 1152, 125–136. doi:10.1007/978-1-4939-0563-8_7
- Oppenheim, A. V., and Schafer, R. W. (2010). *Discrete-Time Signal Processing*, 3rd Edn. Upper Saddle River, NJ: Pearson Higher Education, Inc., 504.
- Oyarzún, D. A., and Stan, G.-B. V. (2013). Synthetic gene circuits for metabolic control: design trade-offs and constraints. *J. R. Soc. Interface* 10:20120671. doi:10.1098/rsif.2012.0671
- Raman, S., Rogers, J. K., Taylor, N. D., and Church, G. M. (2014). Evolution-guided optimization of biosynthetic pathways. *Proc. Natl. Acad. Sci. U.S.A.* 111, 17803–17808. doi:10.1073/pnas.1409523111
- Schallmeyer, M., Frunzke, J., Eggeling, L., and Marienhagen, J. (2014). Looking for the pick of the bunch: high-throughput screening of producing microorganisms with biosensors. *Curr. Opin. Biotechnol.* 26, 148–154. doi:10.1016/j.copbio.2014.01.005
- Schujman, G. E., Guerin, M., Buschiazzo, A., Schaeffer, F., Llarrull, L. I., Reh, G., et al. (2006). Structural basis of lipid biosynthesis regulation in Gram-positive bacteria. *EMBO J.* 25, 4074–4083. doi:10.1038/sj.emboj.7601284
- Schujman, G. E., Paoletti, L., Grossman, A. D., and de Mendoza, D. (2003). FapR, a bacterial transcription factor involved in global regulation of membrane lipid biosynthesis. *Dev. Cell* 4, 663–672. doi:10.1016/S1534-5807(03)00123-0
- Siedler, S., Stahlhut, S. G., Malla, S., Maury, J., and Neves, A. R. (2014). Novel biosensors based on flavonoid-responsive transcriptional regulators introduced into *Escherichia coli*. *Metab. Eng.* 21, 2–8. doi:10.1016/j.ymben.2013.10.011
- Smolke, C. D., and Michener, J. K. (2012). High-throughput enzyme evolution in *Saccharomyces cerevisiae* using a synthetic RNA switch. *Metab. Eng.* 14, 306–316. doi:10.1016/j.ymben.2012.04.004
- Stevens, J. T., and Carothers, J. M. (2014). Designing RNA-based genetic control systems for efficient production from engineered metabolic pathways. *ACS Synth. Biol.* 4, 107–115. doi:10.1021/sb400201u
- Tang, S.-Y., and Cirino, P. C. (2011). Design and application of a mevalonate-responsive regulatory protein. *Angew. Chem. Int. Ed. Engl.* 50, 1084–1086. doi:10.1002/anie.201006083
- Wu, J., Du, G., Zhou, J., and Chen, J. (2013). Metabolic engineering of *Escherichia coli* for (2S)-pinocembrin production from glucose by a modular metabolic strategy. *Metab. Eng.* 16, 48–55. doi:10.1016/j.ymben.2012.11.009
- Xu, P., Gu, Q., Wang, W., Wong, L., Bower, A. G. W., Collins, C. H., et al. (2013). Modular optimization of multi-gene pathways for fatty acids production in *E. coli*. *Nat. Commun.* 4, 1409. doi:10.1038/ncomms2425
- Xu, P., Li, L., Zhang, F., Stephanopoulos, G., and Koffas, M. (2014). Improving fatty acids production by engineering dynamic pathway regulation and metabolic control. *Proc. Natl. Acad. Sci. U.S.A.* 111, 11299–11304. doi:10.1073/pnas.1406401111
- Xu, P., Ranganathan, S., Fowler, Z. L., Maranas, C. D., and Koffas, M. A. G. (2011). Genome-scale metabolic network modeling results in minimal interventions that cooperatively force carbon flux towards malonyl-CoA. *Metab. Eng.* 13, 578–587. doi:10.1016/j.ymben.2011.06.008
- Xu, P., Vansiri, A., Bhan, N., and Koffas, M. A. G. (2012). ePathBrick: a synthetic biology platform for engineering metabolic pathways in *E. coli*. *ACS Synth. Biol.* 1, 256–266. doi:10.1021/sb300016b
- Xu, Y., Piston, D. W., and Johnson, C. H. (1999). A bioluminescence resonance energy transfer (BRET) system: application to interacting circadian clock proteins. *Proc. Natl. Acad. Sci. U.S.A.* 96, 151–156. doi:10.1073/pnas.96.1.151
- Zeileis, A. (2014). dynlm: Dynamic Linear Regression. R package version 0.3-3. Available from: <http://CRAN.R-project.org/package=dynlm>
- Zhang, F., Carothers, J. M., and Keasling, J. D. (2012). Design of a dynamic sensor-regulator system for production of chemicals and fuels derived from fatty acids. *Nat. Biotechnol.* 30, 354–359. doi:10.1038/nbt.2149

Conflict of Interest Statement: The authors certify that there is no conflict of interest with any financial organization regarding the material discussed in the manuscript. The Associate Editor, Jean Marie François, declares that, despite hosting a Frontiers Research Topic alongside the author Pablo Carbonell, the review process was handled objectively and no conflict of interest exists.

Received: 02 September 2014; accepted: 23 March 2015; published online: 08 April 2015.

Citation: Fehér T, Libis V, Carbonell P and Faulon J-L (2015) A sense of balance: experimental investigation and modeling of a malonyl-CoA sensor in *Escherichia coli*. *Front. Bioeng. Biotechnol.* 3:46. doi: 10.3389/fbioe.2015.00046

This article was submitted to *Synthetic Biology*, a section of the journal *Frontiers in Bioengineering and Biotechnology*.

Copyright © 2015 Fehér, Libis, Carbonell and Faulon. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

6.3 Annex 3

Silencing of Antibiotic Resistance in *E. coli* with Engineered Phage Bearing Small Regulatory RNAs

Silencing of Antibiotic Resistance in *E. coli* with Engineered Phage Bearing Small Regulatory RNAs

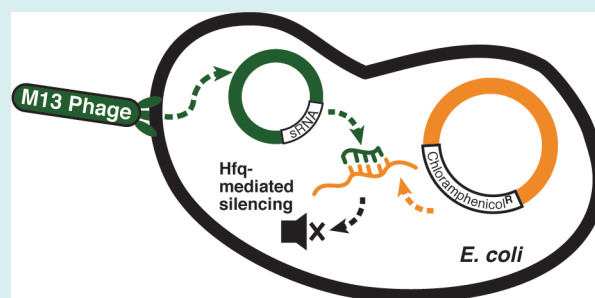
Vincent K. Libis,^{†,§} Aude G. Bernheim,^{†,§} Clovis Basier,[†] Sebastián Jaramillo-Riveri,[†] Matthew Deyell,[†] Idonnya Aghoghogbe,[†] Iva Atanaskovic,[†] Amel Camélia Bencherif,[†] Marguerite Benony,[†] Nicolas Koutsoubelis,[†] Anne C. Löchner,[†] Zoran S. Marinkovic,[†] Sarah Zahra,[†] Yonatan Zegman,[†] Ariel B. Lindner,^{*,†,‡} and Edwin H. Wintermute^{*,†,‡}

[†]2013 Paris Bettencourt iGEM team, 75014 Paris, France

[‡]U1001 Institut National de la Santé et de la Recherche Médicale, 75654 Paris, France

Supporting Information

ABSTRACT: In response to emergent antibiotic resistance, new strategies are needed to enhance the effectiveness of existing antibiotics. Here, we describe a phagemid-delivered, RNA-mediated system capable of directly knocking down antibiotic resistance phenotypes. Small regulatory RNAs (sRNAs) were designed to specifically inhibit translation of chloramphenicol acetyltransferase and kanamycin phosphotransferase. Nonlytic phagemids coding for sRNA expression were able to infect and restore chloramphenicol and kanamycin sensitivity to populations of otherwise resistant *E. coli*. This modular system could easily be extended to other bacteria with resistance profiles that depend on specific transcripts.



The emergence of resistance to conventional antibiotics has motivated interest in techniques to extend or complement their effectiveness.¹ Bacteriophage act to kill bacteria through mechanisms distinct from those of small-molecule antibiotics. Therefore, bacteriophage may serve as complementary therapeutics or bacterial control agents in some contexts.

As with classical antibiotics, the direct application of bacteriophage to cause cell death rapidly selects for resistant strains.² Combination therapies that employ multiple phage or drugs can slow the evolution of resistance. Previous work has shown that bacteriophage and antibiotics can act synergistically, with modified bacteriophage altering host physiology to increase drug sensitivity.³

In this work, we developed an M13 bacteriophage to knock-down the expression of two genes responsible for antibiotic resistance in *E. coli*: chloramphenicol acetyltransferase and kanamycin phosphotransferase (Figure 1). We used rationally designed, sequence-specific sRNAs to target and silence the transcript for each gene directly.⁴ Because such sRNAs can be designed to complement any mRNA, the approach could extend to any antibiotic resistance bestowed by specific transcripts.

The system is RNA-mediated and does not require additional protein expression. Further, the construct is delivered by a nonlytic phagemid system and targets only antibiotic-resistance transcripts. By limiting the disruption of host physiology we have sought to minimize the fitness burden imposed by the construct. In this way, the evolutionary pressures favoring

phage-resistance are minimized until a lethal dose of antibiotic can be applied to a formerly resistant population.

■ RESULTS AND DISCUSSION

Following the design of Na et al.,⁴ we created two sRNA expression constructs targeting chloramphenicol acetyltransferase and kanamycin phosphotransferase (Figure 2). These constructs were incorporated into an M13 phagemid, which was then assembled and released by *E. coli* with the help of an M13 phagemid helper plasmid. Because phagemids do not code for viral structural proteins, we reasoned that a phagemid-based delivery would minimize the physiological disruption and fitness cost of our construct to the infected host.⁵

As described in the Methods, populations of chloramphenicol or kanamycin resistant *E. coli* were infected with phagemid bearing the relevant sRNA silencing construct. Antibiotic sensitivity was assayed by plating on chloramphenicol or kanamycin at a range of concentrations.

Our construct restored substantial antibiotic sensitivity to otherwise resistant populations (Figure 3). At high chloramphenicol concentrations (500 $\mu\text{g/mL}$), we observed >99% killing for infected populations, but no significant killing in control populations (Figure 3a). Lower chloramphenicol concentrations killed less efficiently, but the effect of the

Special Issue: iGEM 2013

Received: January 17, 2014

Published: December 19, 2014

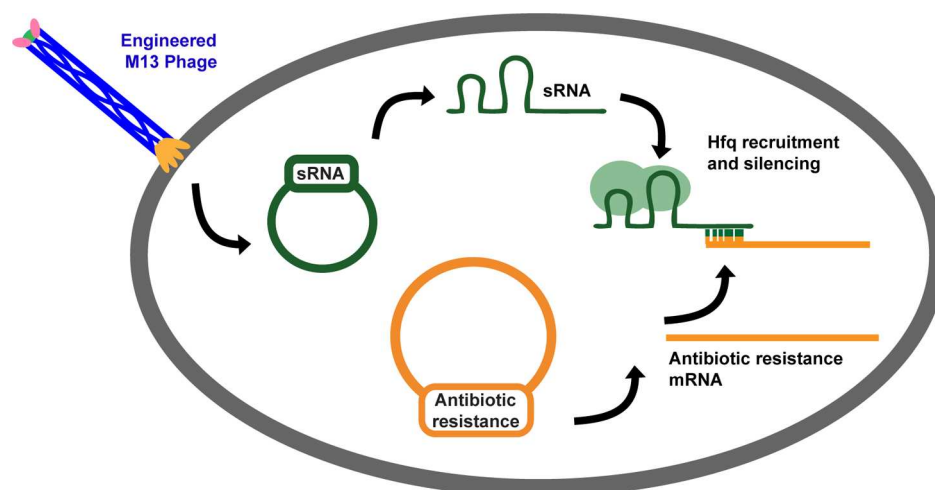


Figure 1. Scheme for antibiotic silencing with phagemid-delivered sRNA. M13 bacteriophage capsules infect the *E. coli* host and deliver a modified nonreplicating phagemid bearing the sRNA expression construct. A 24 bp complementary sequence in the sRNA targets the antibiotic resistance mRNA, while an MicC scaffold sequence recruits the Hfq protein to promote hybridization and degradation.⁴

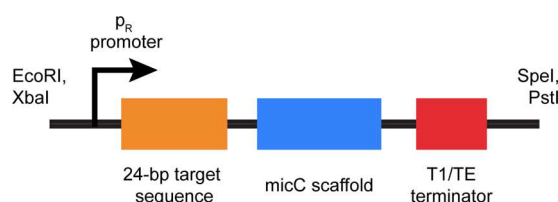


Figure 2. Structure of the sRNA module for the silencing of chloramphenicol acetyltransferase. The entire 419 bp construct was synthesized and cloned into a phagemid vector with BioBrick-compatible restriction sites. The 24 bp target sequence is complementary to the +1 to +24 sequence of the target gene. The modular design is easily changed to target other resistance-associated mRNAs.

sRNA was still evident at 250 $\mu\text{g/mL}$. Similar results were obtained for kanamycin (Figure 3b).

We observed significant variability among biological replicates in our plating assay, particularly at intermediate antibiotic concentrations (Supporting Information Table S1). Plating efficiency may have been affected by the stress of high-antibiotic concentrations, even in genetically resistant cells. Other sources of variation may include variable titers or the accumulation of mutations in phage collected from overnight cultures. In the case of the Kanamycin-silencing construct, phagemids were produced using the M13K07 phagemid-helper plasmid. This plasmid itself contains a target for the expressed sRNA. Although the plasmid was maintained at kanamycin concentrations below the levels at which we observed reduced resistance, this interaction may have contributed to the accumulation of mutations.

Because our observed killing efficiencies were less than 100%, we next sought to investigate the ways in which bacteria may escape the killing effect. A cell may escape killing in our system either because it does not become infected with the phagemid or because the phagemid does not effectively silence resistance. Because our phagemid also codes for GFP expression, we were able to estimate the infection rates of the surviving cells.

We scored GFP fluorescence, indicating phagemid infection, in 71% of surviving colonies (Figure 3c). Thus, in the majority of cases, cell survival is due to a failure of the sRNA to adequately silence gene expression. This may be due

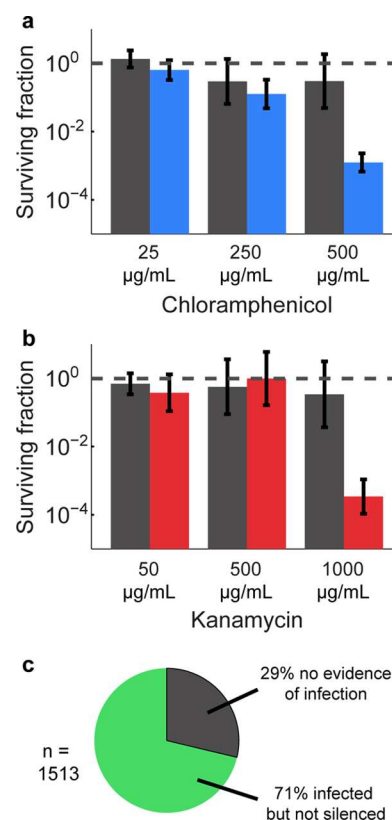


Figure 3. Engineered phagemids restore antibiotic sensitivity to a genetically resistant population. Bacterial populations carrying plasmid-borne resistance cassettes for chloramphenicol (a) and kanamycin (b) were infected with sRNA-expressing phagemids. In the infected populations, significant killing was observed especially at high antibiotic concentrations. In control populations with no sRNA-silencing construct, no significant killing was observed. Dotted lines indicate 100% survival. Error bars represent 95% confidence intervals. (c) Failure modes for sRNA-mediated silencing. Phagemid-encoded GFP was used to score surviving colonies for the presence of the phagemid. The expression of GFP in 71% of surviving colonies (68–74 at 95% confidence) indicates that phagemid infection was successful but sRNA-mediated silencing did not act to restore antibiotic sensitivity.

insufficient production of the sRNA. The pACYCDuet-1 and pCOLADuet-1 vectors that were used to convey antibiotic resistance are estimated by the manufacturer (Novagen) to be present at copy numbers of 10–12 and 20–40, respectively. The sRNA-targeted transcripts may be in excess of sRNA production. Alternately, the sRNA-mRNA complex may be too slow to form or too weakly bound to completely silence transcription. Finally, mutations in the antibiotic resistance genes could disrupt sRNA targeting.

Future work could address in greater detail the mechanism by which resistant *E. coli* escape silencing. Some routes of escape are easily addressed with simple and modular sRNA design. For example, point mutations in the antibiotic resistance gene could be easily matched with a redesigned sRNA. Other escape mechanisms, such as overexpression of the target mRNA or resistance to phage infection, would be more problematic. In particular, mutations in the pili are known to confer resistance to M13 phage.²

The counter-selection of our construct in a bacterial population will be in proportion to the fitness burden it imposes on the host. We have taken steps to reduce the burden of our construct by avoiding protein expression, by minimizing the size of the coding sequence and by targeting only antibiotic resistance transcripts. The targeted genes are nonfunctional in any antibiotic-free medium and costly to translate.⁶ A low-burden construct should be more easily maintained in bacterial population prior to the administration of lethal antibiotic doses. However, further population genetic modeling will be required to minimize counterselection and account for the specific dynamics of a particular host–phage system.

Phage therapy and phage-based biocontrol predate the development of conventional antibiotics.⁷ Currently, broad-spectrum small molecule drugs remain the antibacterial weapon of choice in most cases, despite recent challenges in the form of evolved resistance. Integral to the general success of antibiotics was the application of synthetic chemistry to improve the stability, specificity, and activity of naturally occurring antibiotic compounds. Increasingly, Synthetic Biology provides the tools to enhance wild phage isolates.⁸ Combination therapies featuring semisynthetic antibiotics and designer phage may synergistically increase the value of both approaches.

METHODS

Design of an sRNA to Target Antibiotic Resistance.

The method of Na et al.⁴ was used to design sRNA transcripts to silence chloramphenicol acetyltransferase and kanamycin phosphotransferase. A 24 bp sequence complementary to the +1 to +24 area of each target gene was inserted within the MicC scaffold sequence, which promotes RNA hybridization through recruitment of the Hfq protein. The P_R promoter and T1/TE terminators were included to control the transcription of the regulatory RNA. With the addition of BioBrick-compatible restriction cloning sites, the complete 419 bp module was synthesized by IDT. Restriction cloning was used to integrate the sRNA-expression module into Litmus28i_J23115-B0032-GFP, an M13-derived phagemid carrying GFP.^{9,10}

Construction of Phage and Antibiotic Resistant Strains. Phage-producing cells were created by cotransforming *E. coli* MG1655 with a phagemid and phagemid-helper plasmid M13K07 (NEB). After overnight culture of the phage-producers in selective LB, phage were isolated by centrifugation and filtration at 0.2 μ m.

Antibiotic resistant strains were created by transforming *E. coli* MG1655 with pACYCDuet-1 or pCOLADuet-1 plasmids (Novagen) to convey resistance to chloramphenicol or kanamycin, respectively. An F plasmid was also introduced to both strains by conjugation, as F pilus expression is required for sensitivity to M13.

Antibiotic Sensitivity Assay. Resistant *E. coli* were grown to log phase (OD₆₀₀ = 0.7) in selective LB. One milliliter of culture was mixed with 100 μ L of harvested phage or plain LB as a negative control. Phage infection was allowed to proceed for 30 min at 37 °C with agitation. Serial dilutions were then plated on selective LB agar with defined antibiotic concentrations. Sensitivity to a given antibiotic was assayed by comparing CFU counts at a specific antibiotic concentration to counts on antibiotic-free LB agar.

Phage Infection Assay. Following infection and plating, the number of phagemid-carrying *E. coli* was estimated by scoring GFP fluorescence of single colonies by eye using an Illumatool model LT-9900 (Lighttools Research) with appropriate filters for GFP.

Statistical Analysis. For the antibiotic sensitivity assay, the survival fraction was calculated as a CFU count on antibiotic plates compared to a matched antibiotic-free control. Confidence intervals were estimated using the standard error of the mean for the log-transformed survival fraction. For the phage infection assay, confidence was estimated as a binomial proportion confidence interval.

ASSOCIATED CONTENT

Supporting Information

Exact DNA sequences for the sRNA-expressing phagemids created for this study; raw survival data for the antibiotic sensitivity assay depicted in Figure 3a, b. This material is available free of charge via the Internet at <http://pubs.acs.org>. Further discussion of this project and its context within the larger work of the 2013 Paris Bettencourt iGEM team is available via our team Web site at http://2013.igem.org/Team:Paris_Bettencourt.

AUTHOR INFORMATION

Corresponding Authors

*E-mail: ariel.lindner@inserm.fr.

*E-mail: ehwintermute@gmail.com.

Author Contributions

[§]V.K.L. and A.G.B. contributed equally to this work. V.K.L., A.G.B., and E.H.W. designed the experiments. A.G.B., V.K.L., and C.B. conducted the experiments. All authors contributed to the analysis and interpretation of results. V.K.L., A.G.B., and E.H.W. prepared the manuscript.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

Financial support for the Paris Bettencourt iGEM team was provided by the Fondation Bettencourt Schueller. We acknowledge all the sponsors of the 2013 Paris Bettencourt iGEM team for their contributions this project: http://2013.igem.org/Team:Paris_Bettencourt. We thank the INSERM U1001 research unit and Chantal Lotton for technical assistance and advice. Phagemid Litmus28i_J23115-B0032-GFP was provided by Monica Ortiz and Drew Endy of Stanford.

■ REFERENCES

- (1) Donadio, S., Maffioli, S., Monciardini, P., Sosio, M., and Jabes, D. (2010) Antibiotic discovery in the twenty-first century: Current trends and future perspectives. *J. Antibiotics* 63, 423–430.
- (2) Hagens, S., and Bläsi, U. (2003) Genetically modified filamentous phage as bactericidal agents: A pilot study. *Lett. Appl. Microbiol.* 37, 318–323.
- (3) Lu, T. K., and Collins, J. J. (2009) Engineered bacteriophage targeting gene networks as adjuvants for antibiotic therapy. *Proc. Natl. Acad. Sci. U.S.A.* 106, 4629–4634.
- (4) Na, D., Yoo, S. M., Chung, H., Park, H., Park, J. H., and Lee, S. Y. (2013) Metabolic engineering of *Escherichia coli* using synthetic small regulatory RNAs. *Nat. Biotechnol.* 31, 170–174.
- (5) Qi, H., Lu, H., Qiu, H.-J., Petrenko, V., and Liu, A. (2012) Phagemid vectors for phage display: Properties, characteristics, and construction. *J. Mol. Biol.* 417, 129–143.
- (6) Andersson, D. I., and Hughes, D. (2010) Antibiotic resistance and its cost: Is it possible to reverse resistance? *Nat. Rev. Microbiol.* 8, 260–271.
- (7) Hérelle, F. and Smith, G. (1926) *Bacteriophage and Its Behavior*, The Williams & Wilkins Company, Baltimore, MD.
- (8) Lu, T. K., and Collins, J. J. (2007) Dispersing biofilms with engineered enzymatic bacteriophage. *Proc. Natl. Acad. Sci. U.S.A.* 104, 11197–11202.
- (9) Ortiz, M. E., and Endy, D. (2012) Engineered cell–cell communication via DNA messaging. *J. Biol. Eng.* 6, 16 DOI: 10.1186/1754-1611-6-16.
- (10) Evans, P. D., Cook, S. N., Riggs, P. D., and Noren, C. J. (1995) LITMUS: Multipurpose cloning vectors with a novel system for bidirectional in vitro transcription. *BioTechniques* 19, 130–135.