Expanding Biosensing Abilities through Computer-Aided Design of Metabolic Pathways

DOI:
10.1021/acssynbio.5b00225

Document Version
Accepted author manuscript

ACS SyntheticBiology

Research Article

pubs.acs.org/synthbio

Expanding Biosensing Abilities through Computer-Aided Design of Metabolic Pathways

Vincent Libis,[†,‡] Baudoin Delépine,[†,‡] and Jean-Loup Faulon*[,†,‡,§]

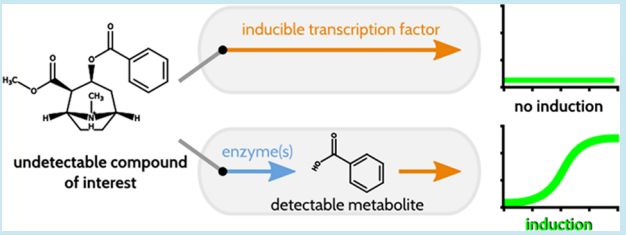
[†]Micalis Institute, INRA, AgroParisTech, Université Paris-Saclay, 78350 Jouy-en-Josas, France

[‡]Institute of Systems and Synthetic Biology, Genopole, CNRS, UEVE, Université Paris-Saclay, F-91030 Évry, France

[§]SYNBIOCHEM Center, Manchester Institute of Biotechnology, School of Chemistry, University of Manchester, Manchester M1 7DN, U.K.

Supporting Information

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 nondetectable 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.



Engineering of circuits in cells has made fast progress since the dawn of synthetic biology. New modular tools and strategies regularly expand the toolbox.[1] Just considering the progress made in the last two years, signal processing in biological systems can now rely on elements such as load drivers,[2] memory systems,[3] amplifiers,[4] coupling systems,[5] or bow-tie architectures,[6] 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.[7] 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.[8] Rational engineering of sensors through protein engineering or riboswitch engineering has been accomplished,[9,10] 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 β-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 has also been demonstrated in synthetic biology to detect an aromatic and to obtain cell to cell communication.[11−13]

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 (Figure 1). 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,[14,15] metabolic engineers created powerful computational methods in the context of small molecule production in microorganisms.[16,17] 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 a 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

Special Issue: Synthetic Biology in Europe

Received: November 6, 2015

DOI: 10.1021/acssynbio.5b00225
ACS Synth. Biol. XXXX, XXX, XXX−XXX

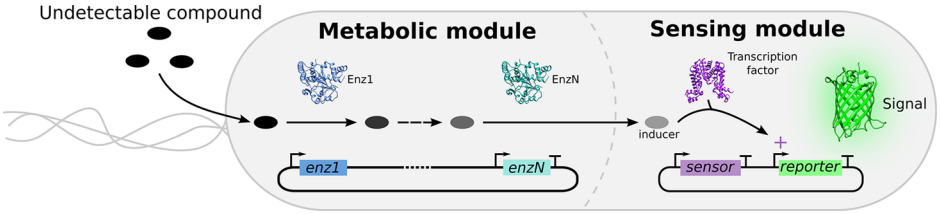


Figure 1. General concept of a cell equipped with a sensing-enabling metabolic pathway (SEMP) allowing for the detection of a new chemical. A naturally undetectable molecule is transformed by a metabolic module into an inducer molecule triggering a genetic response upon binding to a transcription factor. 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.

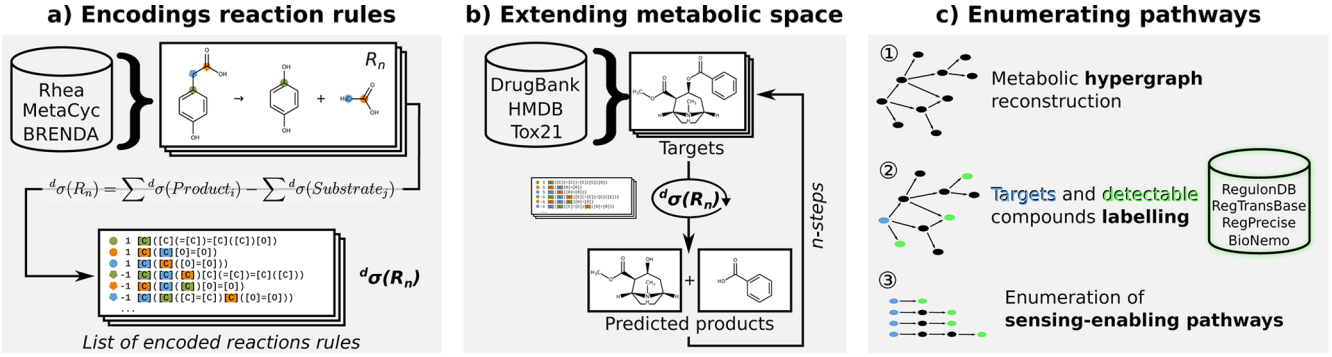


Figure 2. Sensing-enabling metabolic pathway (SEMP) design workflow. (a) Reactions are collected from biochemical databases and encoded as reaction signatures. A reaction signature describes a biochemical transformation rule as a list of fragments of the molecules that are modified during the reaction (See Methods). (b) The effect of each reaction signature is simulated on each target. If a product is predicted to form (i.e., if a 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.

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.

RESULTS

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 data sets representative of these types of molecules from three public databases: DrugBank,[18] HMDB,[19] Tox21.[20] 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,[21] Metacyc,[22] and Rhea[23]) that were merged and encoded in reaction signatures,[24] 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.[25] 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,[26] RegTransBase,[27] RegulonDB,[28] and RegPrecise.[29] An overview of the workflow is represented in Figure 2, and the labeled graph output from the DrugBank data set can be seen in Supporting Information Figure S1. 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.

In each of the considered target data sets 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

DOI: 10.1021/acssynbio.5b00225
ACS Synth. Biol. XXXX, XXX, XXX−XXX

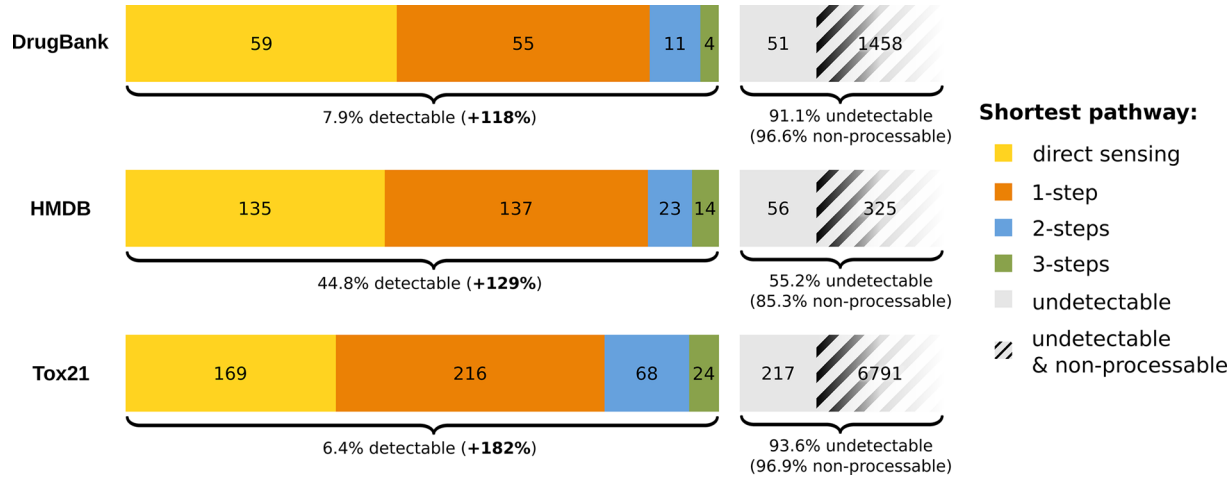


Figure 3. *In silico* prediction of detectable compounds among three data sets. Compounds labeled in the "direct sensing" category (yellow) are already present in the data set of natural transcription factor's effectors. Compounds in other colors (orange, blue, green) are reached by sensing-enabling metabolic pathways (SEMP). Nonprocessable compounds (hashed) are compounds that do not participate in known enzymatic reactions (as a product or a substrate). Data sets correspond to the DrugBank sections "approved drugs" and "illicit drugs"; HMDB's biomarker compounds are associated with a human disease and the entire Tox21 data set of putative toxic compounds.

(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 data sets (Figure 3).

Most compounds in the data sets that remain undetectable are products of organic chemistry that are not processed by any known enzyme and therefore no SEMPs 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 is displayed in Table 1.

Implementation of SEMPs *in vivo* expands *E. coli* sensing abilities. In order to validate the concept *in vivo*, we implemented in *E. coli* representative SEMPs from the predictions related to each of the targets data sets (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 cotransformation in the same strain. Prior to full characterization of SEMPs we identified optimal expression levels of the heterologous transcription factors by varying the arabinose concentration in the 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 the presence of benzoate but not in the presence of cocaine (Supporting Information Figure S5a). Upon cotransformation with a metabolic module containing CocE, a fluorescent response also occurs in the presence of cocaine, indicating the expected extension of the sensing scope of the bacteria (Figure 4a). 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 (Supporting Information Figure S5b). 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 the presence of parathion (Figure 4b). To our surprise the fluorescent response of this strain to parathion is higher than for the native inducer above 100 μM. The drop in signal in response to 4-nitrophenol at 300 μM is probably due to the associated toxicity that diminishes growth (see Supporting Information Figure S8). 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 1 mM in Supporting Information Figure S7).

DOI: 10.1021/acssynbio.5b00225
ACS Synth. Biol. XXXX, XXX, XXX−XXX

Table 1. Sensing-Enabling Metabolic Pathway (SEMP) Examples[a]

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

[a](*) indicates SEMPs implemented and tested *in vivo* over the course of this work (see Supporting Information Figure S3).

In contrast, parathion is less toxic for these cells and in fact is not toxic at these concentrations if the PTE enzyme is absent (Supporting Information Figure S8). 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 1 mM (Supporting Information Figure S7). 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.[30] 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 (Figure 4c). We monitored the activation of this hippurate sensor and the cocaine sensor in order to investigate if SEMPs 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 the HipO- or CocE-mediated transformation step delays the fluorescent response (see Supporting Information Figure S2).

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.[31] We wondered if transfer of information could as well rely on multienzymatic 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 (Figure 4d), while it was not possible with the LinR-module alone (Supporting Information Figure S5d). With our setup, the sensing module based on LinR transcription factor and a

DOI: 10.1021/acssynbio.5b00225
ACS Synth. Biol. XXXX, XXX, XXX−XXX

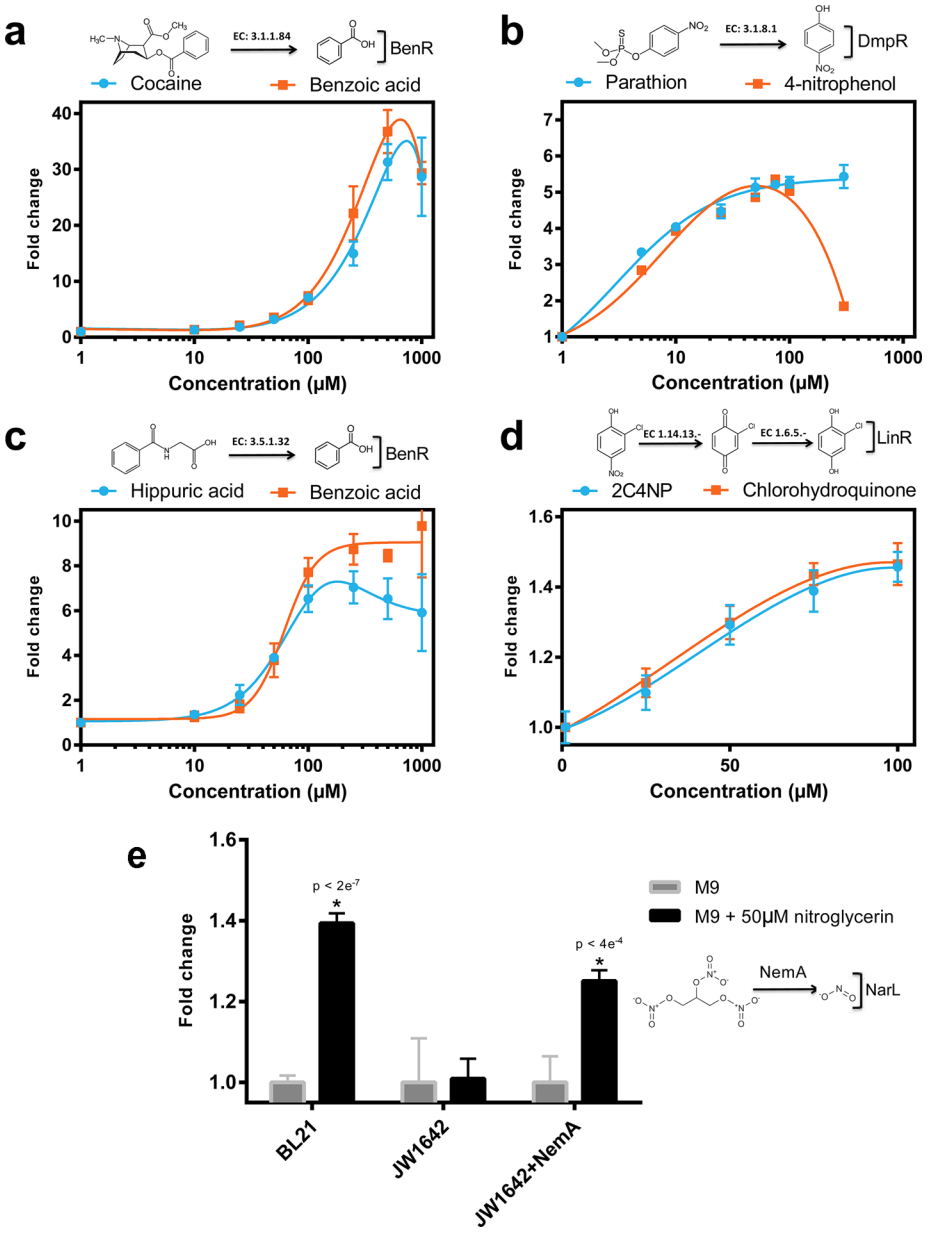


Figure 4. *In vivo* characterization of sensing-enabling metabolic pathways (SEMP). Cotransformation of *E. coli* with both the metabolic module and sensing module confers sensing abilities toward new molecules. The dose−response relationship of engineered *E. coli* strains to the natural effectors and to the target molecules cocaine (a), parathion (b), hippurate (c), and 2C4NP (d). (e) shows the 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 μM 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 represents 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.

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 μM upper limit. Remarkably, only the strain with the metabolic module could survive at a concentration of 2C4NP higher than 75 μM (Supporting Information Figure S8), 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 NemA 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

DOI: 10.1021/acssynbio.5b00225
ACS Synth. Biol. XXXX, XXX, XXX−XXX

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 the 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 NemA enzyme. This strain is not anymore able to respond to nitroglycerin. Finally, we introduced a functional copy of NemA on a second plasmid into JW1642 and observed a restoration of the fluorescent response to nitroglycerin (Figure 4e). This confirms that NemA 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 Supporting Information Figure S3.

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). SEMPs 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 SEMPs 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.[32] By increasing the number of value added chemicals that can be monitored through fluorescent readout, SEMPs 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 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.[7]

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,[33] which is known to have an important substrate promiscuity against several organophosporous compounds.[34] 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 has already been characterized for metabolic engineering applications.[35] 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 SEMPs 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 SEMPs to be easily transplanted to artificial cells that have been shown to serve as viable chassis for both biosensors and metabolic pathways.[36,37] Specific limitations of SEMPs reside in the need for nontoxic 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 Supporting Information Figure S6. 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 SEMPs 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.[38,39] While the natural occurrence of a response to nitroglycerin seems fortuitous, recent evidence concerning the Lac operon shows that evolution strongly coselected the LacI transcription factor and the side-reaction site of β-galactosidase that leads to production of allolactose.[40] 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.

METHODS

Data source. We retrieved 9,319 distinct encoded reactions (reaction signature diameter 12, see below) and their associated compounds from BRENDA 2015.1,[21] MetaCyc 18.5,[22] and Rhea v61.[23] Targets were gathered from DrugBank 4.1[18] (all approved drugs and illicit drugs data sets), the Human Metabolomic DataBase[19] (all compounds linked to a human disease) and the Distributed Structure-Searchable Toxicity (DSSTox) Database Network[20] (all compounds). Finally, we retrieved 505 distinct transcription factor effectors from BioNemo,[26] RegTransBase,[27] RegulonDB,[28] and RegPrecise.[29]

Molecular and reaction signatures. Molecular signatures[41] (MS) are graph-based descriptors that encodes the "neighborhood" of each atom of a molecule, similarly to

DOI: 10.1021/acssynbio.5b00225
ACS Synth. Biol. XXXX, XXX, XXX−XXX

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,[24] 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 allows 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 nonpromiscuous (unless several distinct reactions are reported in the databases for the same enzyme). See Carbonell et al.[17] 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 pretreated 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.[17] 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[42] over 0.99). Finally, the sensing pathways are extracted from the graph with NetworkX graph library v.1.11.[43] Frequently a target can be sensed through several SEMPs leading to different sensable compounds, with different pathway lengths. For the analysis of the predictions, detectable targets were counted only once even if several SEMP were predicted.

Request for predictions on custom list of compounds can be addressed to jean-loup.faulon@jouy.inra.fr.

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-optimized for *E. coli* with Jcat[44] and natural BsaI sites were removed. These genes were synthesized by Genscript (Piscataway, NJ, USA). Coding sequence of PTE-S5, LinR, pnpA, pnpB and DmpR were extracted by PCR from plasmids pMaIc2x-PTE-S5,[33] pMEU2R,[45] pET-pnpA, pET-pnpB,[46] 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.[47] 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 Supporting Information Figure S4 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 in Supporting Information Data File 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.

Biosensor dose−response characterization. For each biosensor strain, an isolated colony of BL21(DE3) harboring the appropriate plasmid(s) was inoculated in 2 mL of selective

DOI: 10.1021/acssynbio.5b00225
ACS Synth. Biol. XXXX, XXX, XXX−XXX

LB and grown overnight at 37 °C. The overnight culture was diluted 1:100 in fresh selective LB and grown for 90 min under agitation at 37 °C. Cells were then induced with IPTG 1 mM 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 artifacts 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 2 μL of different concentrations were mixed with samples of 200 μL of cell culture. 2 μL of pure ethanol was mixed with negative control cell samples. Cocaine sensor cells and associated control cells lacking metabolic module were grown in Eppendorf 1.5 mL microtubes with vigorous agitation for 18 h hours at 30 °C, thus following optimal conditions described in the literature for CocE activity. All other sensor cells were grown for 18 h with agitation at 37 °C in microplate reader TECAN Infinite 500. Absorbance at OD600 and fluorescence (Exc: 580 nm/Em: 610 nm) 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.[48] The coding sequence of NemA 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 NemA-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. 2 μL of nitroglycerin diluted in ethanol was added to 200 μL of cell samples to a reach a final concentration of 50 μM. 2 μL of pure ethanol was added to control samples. Cells were grown for 18 h with agitation at 37 °C and absorbance at OD600 and fluorescence (Exc: 580 nm/Em: 610 nm) was measured in microplate reader TECAN Infinite 500.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acssynbio.5b00225.

Figures S1−S9 and Data files S1 and S2. (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: jean-loup.faulon@jouy.inra.fr.

Author Contributions

V.L., B.D., and J.-L.F. designed the study. V.L. designed, built, and characterized the biosensors. B.D. adapted the algorithms and conducted simulations and analysis. All authors participated in the interpretation of the results and in the preparation of the manuscript.

Notes

The authors declare no competing financial interest.

■ 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; and Edwin Wintermute, Marcel Salanoubat, Matthew Deyell, and Aude Bernheim for fruitful discussions. V.L is supported by a DGA (French Ministry of Defence) graduate fellowship and received additional support from the "Frontières du Vivant" doctoral school, Programme Bettencourt. B.D. is supported by a graduate fellowship from SDSV doctoral school (Université Paris Saclay). J.-L.F. is supported by GIP Genopole, Toulouse White Biotech (TWB), Institut National de la Recherche Agronomique, and the University of Manchester through BBSRC/EPSRC grant BB/M017702/1.

■ REFERENCES

(1) Brophy, J. A. N., and Voigt, C. A. (2014) Principles of genetic circuit design. *Nat. Methods 11*, 508−20.

(2) Mishra, D., Rivera, P. M., Lin, A., Del Vecchio, D., and Weiss, R. (2014) A load driver device for engineering modularity in biological networks. *Nat. Biotechnol. 32*, 1268−75.

(3) Farzadfard, F., and Lu, T. K. (2014) Genomically encoded analog memory with precise in vivo DNA writing in living cell populations. *Science (Washington, DC, U. S.) 346*, 1256272−1256272.

(4) Bonnet, J., Yin, P., Ortiz, M. E., Subsoontorn, P., and Endy, D. (2013) Amplifying genetic logic gates. *Science 340*, 599−603.

(5) Prindle, A., Selimkhanov, J., Li, H., Razinkov, I., Tsimring, L. S., and Hasty, J. (2014) Rapid and tunable post-translational coupling of genetic circuits. *Nature 508*, 387−91.

(6) Prochazka, L., Angelici, B., Haefliger, B., and Benenson, Y. (2014) Highly modular bow-tie gene circuits with programmable dynamic behaviour. *Nat. Commun. 5*, 4729.

(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) Rogers, J. K., Guzman, C. D., Taylor, N. D., Raman, S., Anderson, K., and Church, G. M. (2015) Synthetic biosensors for precise gene control and real-time monitoring of metabolites. *Nucleic Acids Res. 43*, 7648.

(9) Bayer, T. S., and Smolke, C. D. (2005) Programmable ligand-controlled riboregulators of eukaryotic gene expression. *Nat. Biotechnol. 23*, 337−43.

(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., Fields, S., Church, G. M., and Raman, S. (2015) Engineering an allosteric transcription factor to respond to new ligands. *Nat. Methods 13*, 177−83.

(11) Silva-Rocha, R., and de Lorenzo, V. (2014) Engineering multicellular logic in bacteria with metabolic wires. *ACS Synth. Biol. 3*, 204−9.

(12) Xue, H., Shi, H., Yu, Z., He, S., Liu, S., Hou, Y., Pan, X., Wang, H., Zheng, P., Cui, C., Viets, H., Liang, J., Zhang, Y., Chen, S., Zhang, H. M., and Ouyang, Q. (2014) Design, construction, and characterization of a set of biosensors for aromatic compounds. *ACS Synth. Biol. 3*, 1011−4.

(13) Bulter, T., Lee, S.-G., Wong, W. W., Fung, E., Connor, M. R., and Liao, J. C. (2004) Design of artificial cell-cell communication using gene and metabolic networks. *Proc. Natl. Acad. Sci. U. S. A. 101*, 2299−2304.

(14) Beal, J., Lu, T., and Weiss, R. (2011) Automatic compilation from high-level biologically-oriented programming language to genetic regulatory networks. *PLoS One 6*, e22490.

DOI: 10.1021/acssynbio.5b00225
ACS Synth. Biol. XXXX, XXX, XXX−XXX

(15) Clancy, K., and Voigt, C. A. (2010) Programming cells: towards an automated "Genetic Compiler. *Curr. Opin. Biotechnol. 21*, 572−81.
(16) Medema, M. H., van Raaphorst, R., Takano, E., and Breitling, R. (2012) Computational tools for the synthetic design of biochemical pathways. *Nat. Rev. Microbiol. 10*, 191−202.
(17) 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*, W389−94.
(18) Law, V., Knox, C., Djoumbou, Y., Jewison, T., Guo, A. C., Liu, Y., Maciejewski, A., Arndt, D., Wilson, M., Neveu, V., Tang, A., Gabriel, G., Ly, C., Adamjee, S., Dame, Z. T., Han, B., Zhou, Y., and Wishart, D. S. (2014) DrugBank 4.0: shedding new light on drug metabolism. *Nucleic Acids Res. 42*, D1091−7.
(19) Wishart, D. S., Jewison, T., Guo, A. C., Wilson, M., Knox, C., Liu, Y., Djoumbou, Y., Mandal, R., Aziat, F., Dong, E., Bouatra, S., Sinelnikov, I., Arndt, D., Xia, J., Liu, P., Yallou, F., Bjorndahl, T., Perez-Pineiro, R., Eisner, R., Allen, F., Neveu, V., Greiner, R., and Scalbert, A. (2013) HMDB 3.0–The Human Metabolome Database in 2013. *Nucleic Acids Res. 41*, D801−7.
(20) Kavlock, R. J., Austin, C. P., and Tice, R. R. (2009) Toxicity testing in the 21st century: implications for human health risk assessment. *Risk Anal. 29*, 485−7 (discussion 492-7)..
(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*, D439−46.
(22) Caspi, R., Altman, T., Billington, R., Dreher, K., Foerster, H., Fulcher, C. A., Holland, T. A., Keseler, I. M., Kothari, A., Kubo, A., Krummenacker, M., Latendresse, M., Mueller, L. A., Ong, Q., Paley, S., Subhraveti, P., Weaver, D. S., Weerasinghe, D., Zhang, P., and Karp, P. D. (2014) The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases. *Nucleic Acids Res. 42*, D459−71.
(23) Morgat, A., Axelsen, K. B., Lombardot, T., Alcántara, R., Aimo, L., Zerara, M., Niknejad, A., Belda, E., Hyka-Nouspikel, N., Coudert, E., Redaschi, N., Bougueleret, L., Steinbeck, C., Xenarios, I., and Bridge, A. (2015) Updates in Rhea—a manually curated resource of biochemical reactions. *Nucleic Acids Res. 43*, D459−64.
(24) 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*, 225−33.
(25) Fehér, T., Planson, A.-G., Carbonell, P., Fernández-Castané, A., Grigoras, I., Dariy, E., Perret, A., and Faulon, J.-L. (2014) Validation of RetroPath, a computer-aided design tool for metabolic pathway engineering. *Biotechnol. J. 9*, 1446−57.
(26) Carbajosa, G., Trigo, A., Valencia, A., and Cases, I. (2009) Bionemo: molecular information on biodegradation metabolism. *Nucleic Acids Res. 37*, D598−602.
(27) 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.
(28) Salgado, H., Peralta-Gil, M., Gama-Castro, S., Santos-Zavaleta, A., Muñiz-Rascado, L., García-Sotelo, J. S., Weiss, V., Solano-Lira, H., Martínez-Flores, I., Medina-Rivera, A., Salgado-Osorio, G., Alquicira-Hernández, S., Alquicira-Hernández, K., López-Fuentes, A., Porrón-Sotelo, L., Huerta, A. M., Bonavides-Martínez, C., Balderas-Martínez, Y. I., Pannier, L., Olvera, M., Labastida, A., Jiménez-Jacinto, V., Vega-Alvarado, L., Del Moral-Chávez, V., Hernández-Alvarez, A., Morett, E., and Collado-Vides, J. (2013) RegulonDB v8.0: omics data sets, evolutionary conservation, regulatory phrases, cross-validated gold standards and more. *Nucleic Acids Res. 41*, D203−13.
(29) 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., and Rodionov, D. A. (2013) RegPrecise 3.0—a resource for genome-scale exploration of transcriptional regulation in bacteria. *BMC Genomics 14*, 745.
(30) Kojima, R., Aubel, D., and Fussenegger, M. (2015) Novel theranostic agents for next-generation personalized medicine: small molecules, nanoparticles, and engineered mammalian cells. *Curr. Opin. Chem. Biol. 28*, 29−38.
(31) Galanie, S., Thodey, K., Trenchard, I. J., Filsinger Interrante, M., and Smolke, C. D. (2015) Complete biosynthesis of opioids in yeast. *Science (Washington, DC, U. S.) 349*, 1095−100.
(32) Rogers, J. K., and Church, G. M. (2016) Genetically encoded sensors enable real-time observation of metabolite production. *Proc. Natl. Acad. Sci. U. S. A. 113*, 2388−2393.
(33) 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*, 51−8.
(34) Briseño-Roa, L., Timperley, C. M., Griffiths, A. D., and Fersht, A. R. (2011) Phosphotriesterase variants with high methylphosphonatase activity and strong negative trade-off against phosphotriesters. *Protein Eng., Des. Sel. 24*, 151−9.
(35) Steigedal, M., and Valla, S. (2008) 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. 10*, 121−9.
(36) Lentini, R., Santero, S. P., Chizzolini, F., Cecchi, D., Fontana, J., Marchioretto, M., Del Bianco, C., Terrell, J. L., Spencer, A. C., Martini, L., Forlin, M., Assfalg, M., Dalla Serra, M., Bentley, W. E., and Mansy, S. S. (2014) Integrating artificial with natural cells to translate chemical messages that direct E. coli behaviour. *Nat. Commun. 5*, 4012.
(37) Elani, Y., Law, R. V., and Ces, O. (2014) Vesicle-based artificial cells as chemical microreactors with spatially segregated reaction pathways. *Nat. Commun. 5*, 5305.
(38) Pflüger-Grau, K., and de Lorenzo, V. (2014) From the phosphoenolpyruvate phosphotransferase system to selfish metabolism: a story retraced in Pseudomonas putida. *FEMS Microbiol. Lett. 356*, 144−53.
(39) Chubukov, V., Gerosa, L., Kochanowski, K., and Sauer, U. (2014) Coordination of microbial metabolism. *Nat. Rev. Microbiol. 12*, 327−40.
(40) Wheatley, R. W., Lo, S., Jancewicz, L. J., Dugdale, M. L., and Huber, R. E. (2013) 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. 288*, 12993−3005.
(41) Faulon, J.-L., Visco, D. P., and Pophale, R. S. (2003) The signature molecular descriptor. 1. Using extended valence sequences in QSAR and QSPR studies. *J. Chem. Inf. Model. 43*, 707−20.
(42) Jaccard, P. (1901) Distribution de la flore alpine dans le Bassin de Dranses et dans quelques regions voisines. *Bull. la Soc. Vaudoise des Sci. Nat. 37*, 241−272.
(43) Hagberg, A. A., Schult, D. A., and Swart, P. J. (2008) Exploring network structure, dynamics, and function using {NetworkX}, in *Proceedings of the 7th Python in Science Conference (SciPy2008)*, pp 11−15. Pasadena, CA USA.
(44) Grote, A., Hiller, K., Scheer, M., Munch, R., Nortemann, B., Hempel, D. C., and Jahn, D. (2005) JCat: a novel tool to adapt codon usage of a target gene to its potential expression host. *Nucleic Acids Res. 33*, W526−W531.
(45) Miyauchi, K., Lee, H., Fukuda, M., Takagi, M., and Nagata, Y. (2002) Cloning and Characterization of linR. *Involved in Regulation of the Downstream Pathway for γ -Hexachlorocyclohexane Degradation in Sphingomonas paucimobilis UT26 68*, 1803−1807.
(46) Min, J., Zhang, J.-J., and Zhou, N.-Y. (2014) The para-Nitrophenol Catabolic Gene Cluster is Responsible for 2-Chloro-4-nitrophenol Degradation in Burkholderia sp. Strain SJ98. *Appl. Environ. Microbiol. 80*, 6212−6222.
(47) Engler, C., Kandzia, R., and Marillonnet, S. (2008) A one pot, one step, precision cloning method with high throughput capability. *PLoS One 3*, e3647.
(48) Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., Datsenko, K. A., Tomita, M., Wanner, B. L., and Mori, H. (2006)

DOI: 10.1021/acssynbio.5b00225
ACS Synth. Biol. XXXX, XXX, XXX−XXX

Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol. Syst. Biol. 2*, 2006.0008.

DOI: 10.1021/acssynbio.5b00225
ACS Synth. Biol. XXXX, XXX, XXX−XXX