ACS SyntheticBiology

Research Article

pubs.acs.org/synthbio

Retropath: Automated Pipeline for Embedded Metabolic Circuits

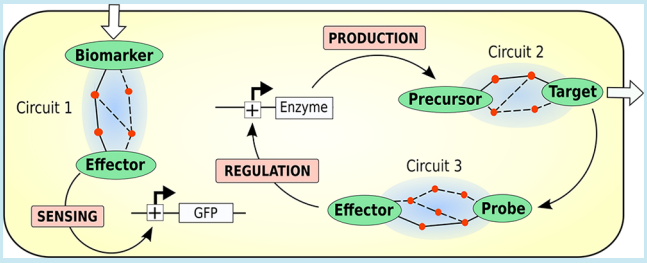
Pablo Carbonell,*[,†] Pierre Parutto,[†] Claire Baudier,[†] Christophe Junot,[‡] and Jean-Loup Faulon[†]

[†]Univ. Evry, CNRS, iSSB, F-91000 Evry, France

[‡]CEA-Centre d'Etude de Saclay, 91190 Gif-sur-Yvette, France

Supporting Information

ABSTRACT: Metabolic circuits are a promising alternative to other conventional genetic circuits as modular parts implementing functionalities required for synthetic biology applications. To date, metabolic design has been mainly focused on production circuits. Emergent applications such as smart therapeutics, however, require circuits that enable sensing and regulation. Here, we present RetroPath, an automated pipeline for embedded metabolic circuits that explores the circuit design space from a given set of specifications and selects the best circuits to implement based on desired constraints. Synthetic biology circuits embedded in a chassis organism that are capable of controlling the production, processing, sensing, and the release of specific molecules were enumerated in the metabolic space through a standard procedure. In that way, design and implementation of applications such as therapeutic circuits that autonomously diagnose and treat disease, are enabled, and their optimization is streamlined.



KEYWORDS: *synthetic biology, circuits, metabolic network, biosynthesis, biosensor, regulation*

Synthetic biology aims at translating methods and techniques from engineering into biology in order to streamline the design and implementation of devices through standardized parts.[1] To that end, metabolic circuits stand out as a promising alternative to other proposed genetic circuits[2] for developing modular parts implementing most of the basic functionalities required in advanced synthetic biology applications. To date, metabolic design has been mainly focused on production circuits that perform synthesis of exogenous compound in chassis organisms. There are hundreds of such examples in the literature.[3,4] One of the most complete engineered pathways is the one for the production of semisynthetic artemisinin in yeast.[5] Other examples of bioproduction of pharmaceutical products through heterologous pathways include taxadiene,[6] farnesol,[7] or naringenin.[8] Biosensors, in turn, are circuits capable of detecting biomarkers related to environmental or disease conditions. Even though sensing signaling pathways are mostly based on phosphorylation cascades, there are some based on metabolism regulation such as those involved in the degradation of toxic compounds such as toluene or phenol or the well-known mechanism of digestion of lactose through the *lac* operon. Besides its function in hydrolyzing lactose, *β*-galactosidase has the ability of synthesizing promiscuously allolactose, the *lac* operon inducer. This mechanism of the *lac* operon shows how enzyme promiscuity is a natural strategy for regulation that can be exploited for developing novel biosensors through metabolic circuits.

Emergent applications such as smart therapeutics are requiring such advanced circuits that enable sensing and regulation of compound delivery. Examples include a synthetic device consisting of a modified *Deinococcus radiodurans*-derived protein that sensed uric acid levels and triggered the expression of an *Aspergillus flavus* urate oxidase to eliminate uric acid in mice;[9] a synthetic circuit in *Escherichia coli* with the ability of sensing *Pseudomonas aeruginosa* quorum sensing in order to induce the production of the pyocin antibiotic and simultaneously triggering the lysis of the chassis *E. coli* for efficient delivery;[10] as well as other studies that focused on different regulation strategies such as enzyme inhibition and activation by small molecules,[11] gene expression regulation using small RNAs,[12] or dynamic feedback.[13,14] In particular, a class of such advanced circuits is processing modules based on boolean operations between signal molecules. Those metabolic circuits as well as their interplay with genetic circuits can be formalized using boolean logics,[15−17] allowing in that way the implementation of circuits depending on multiple inputs and outputs through complex logical relationships with the advantage of processing information at high speed.

Addressing the challenge of providing automated design methodologies for such advanced metabolic circuits requires an appropriate modeling of metabolic processes that should enable the efficient computation of the design space associated with any desired circuit. For that purpose, we have recently proposed a fingerprint coding representation based on changes in atom bonding environments where enzymatic reactions are processed.[18] The main goal of following this coding

Special Issue: IWBDA 2013

Received: August 31, 2013
Published: October 4, 2013

   dx.doi.org/10.1021/sb4001273 | *ACS Synth. Biol.* 2014, 3, 565−577

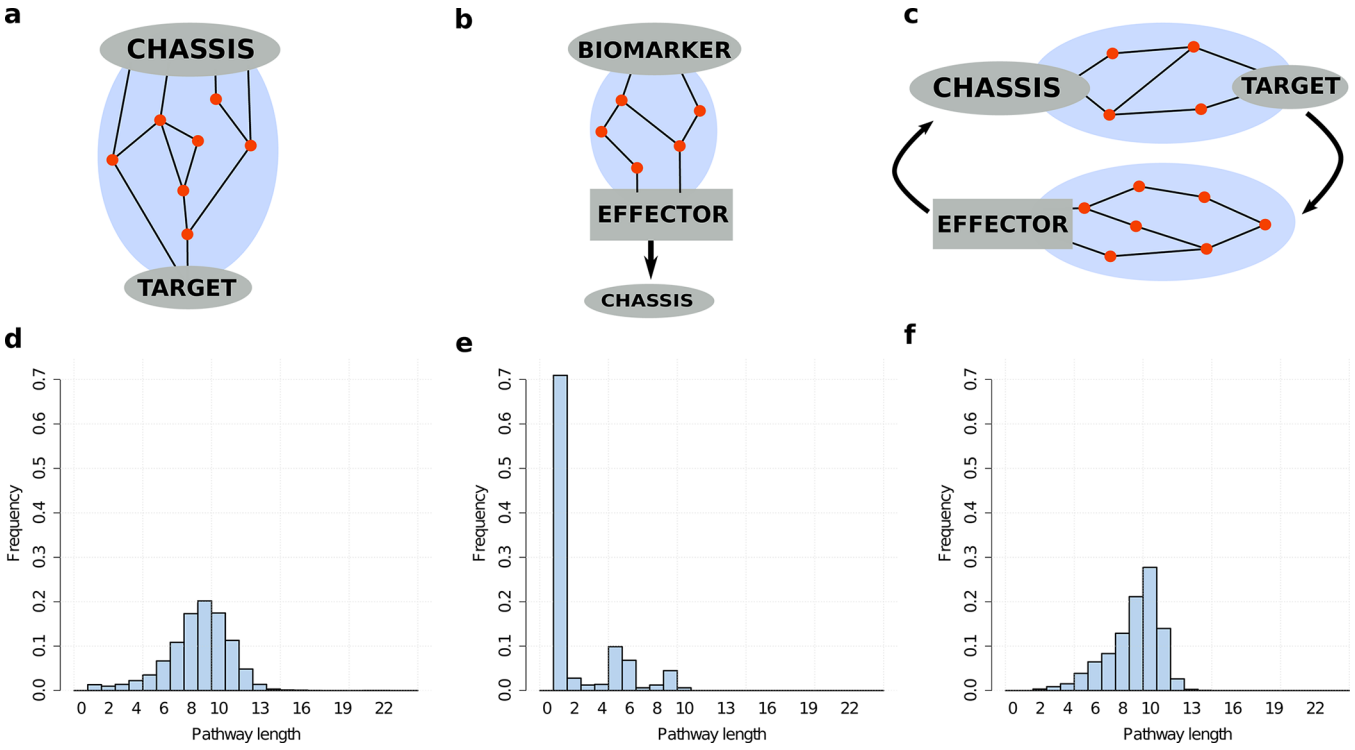


Figure 1. Metabolic circuits I/O specifications and pathways length. (a) Production: a metabolic circuit is established between the chassis and the target product. (b) Sensing: a circuit is defined between the sensed molecules (biomarkers) and the effectors of the chassis. (c) Regulation: a module connects the chassis to the target product, while a second module connects the target product to an effector of the chassis. Distribution of pathway lengths in *E. coli* for (d) bioproduction circuits, (e) biosensor circuits, (f) bioregulation circuits.

representation is to provide a technique that allows considering not only metabolic transformation reported in metabolic databases such as MetaCyc[19] but also other potential transformations that might be present due to the ability of enzymes to accept more than one substrate or to catalyze promiscuously some reaction other than its native one.[20] In this direction, enzyme promiscuity stands as a widely found basic property that should allow one to expand the design space of metabolic circuits. Such promiscuous activities, however, are often present at weaker levels in the enzyme than desired for efficient circuit operation. To optimize their performance, directed evolution of individual enzymes or full metabolic circuits would be usually needed.[21,22] Therefore, arriving at the goal of designing and constructing metabolic circuits with the desired functionalities should be attained by proceeding in a programmatic way that starts by exploring the design space. To that end, we present RetroPath, an automated pipeline allowing the design and optimization of metabolic circuits through multiple steps upon a required set of specifications.

RESULTS AND DISCUSSION

Metabolic circuits are by definition a special case of synthetic biology devices whose main function is to process chemical compounds through enzymatic transformations. Therefore, the distinctive feature of metabolic circuits is the fact that they use enzymes as their main basic constitutive parts. The most common application of metabolic circuits is to produce a desired chemical in a chassis organism by importing heterologous genes encoding for the enzymes that participate in the biosynthetic pathway. More generally, metabolic circuits can be composed of either one or a combination out of production, sensor and processing modules (Figure 1). In the case of sensor circuits, enzymes are used in order to connect through metabolic transformations a chemical of interest to another compound that can induce gene expression in the chassis. In addition, another class of metabolic circuits are the ones that process chemical compounds in order to produce other chemicals through some input/output transfer function or operation. The distinction, thus, between a specific type of metabolic circuit or another can be reduced to the type of input/output sets of metabolites that constitute the circuit's interface, whether they are endogenous or heterologous, as well as if they belong to some specific type of compounds such as biomarkers, inducers, or added-value compounds.

Such level of generalization is sought in our RetroPath framework in order to approach the metabolic circuit design problem in an automated fashion. Altogether, this problem does not differ greatly from the problem of design space exploration often found in embedded systems in electronics.[23] The basic steps are the following: (a) Define input, output, and metabolic space; (b) define specifications of the desired circuit; (c) compute scope connecting input with output sets; (d) enumerate the circuit design space; and (e) explore the design space in order to find the optimal solution(s). We describe in the following each of the steps.

Metabolic Space. In order to address the problem of determining the design space, first the fraction of the chemical space that can be processed *in vivo* using natural or synthetic biology devices, the metabolic space **M**, needs to be delimitated. It is acknowledged that current metabolic databases, although they provide a wealth of information, they are still far from complete in terms of all substrates and products that potentially could be processed by known enzymes. A glimpse into such "metabolic dark matter" is offered by

dx.doi.org/10.1021/sb4001273 | *ACS Synth. Biol.* 2014, 3, 565−577

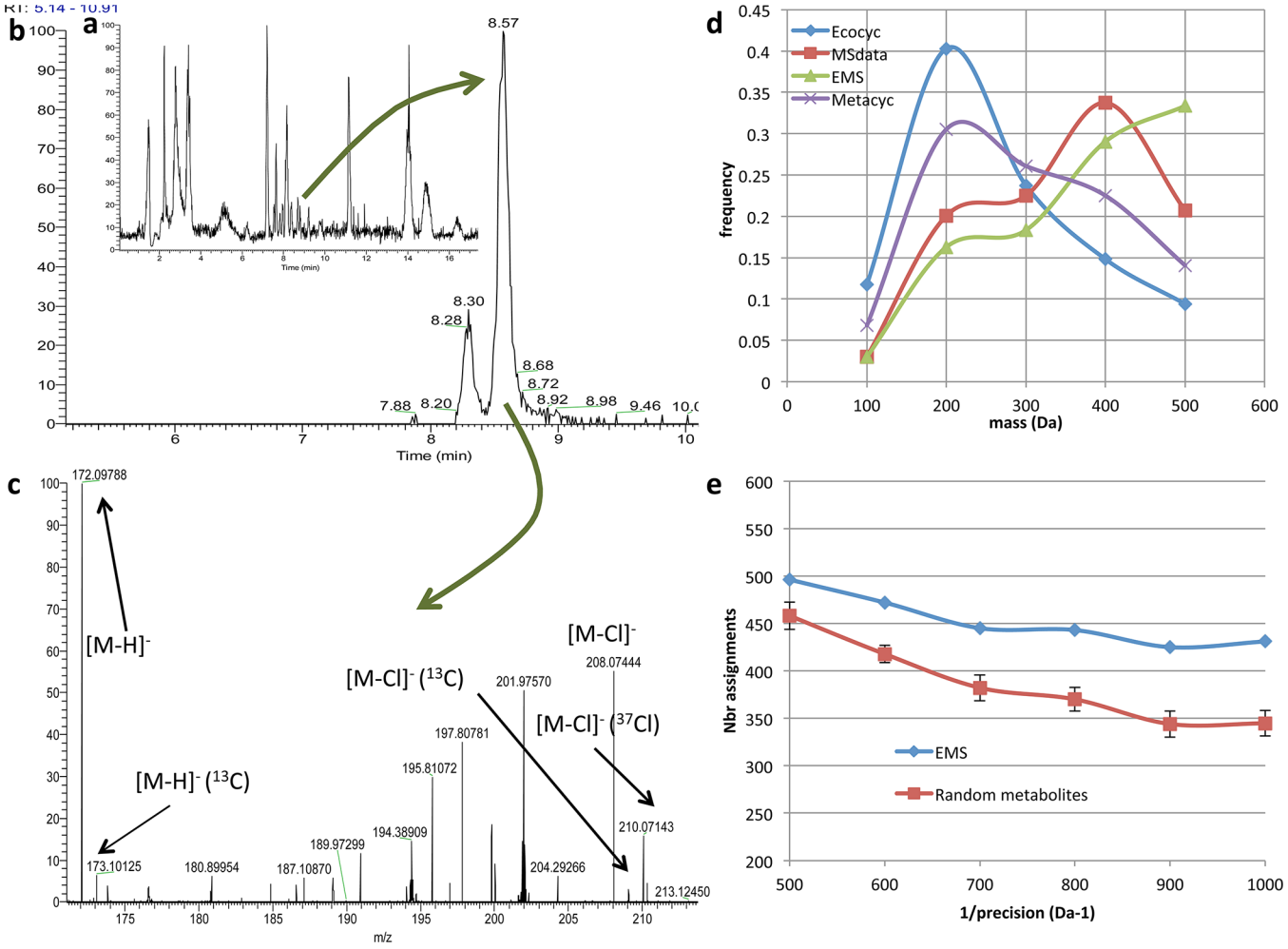


Figure 2. LC/MS metabolite identification in the expanded metabolic space. (a) Total ion LC/MS chromatogram of an *E. coli* cell extract. (b) Extracted ion chromatogram of *m/z* 172.098. (c) High resolution mass spectrum related to the peak at 8.6 min. Several ions related to the *m/z* 172.098 are annotated. Please refer to the experimental section for more details related to the LC/MS settings. (d) Comparison of masses distribution between peaks observed in the MS spectrum of an *E. coli* extract sample (MSdata), endogenous metabolites according to EcoCyc database, metabolites in MetaCyc, and masses corresponding to the expanded metabolic space (EMS). (e) Mass assignments at different levels of precission in the expanded metabolic space (EMS) compared with 5 random metabolite samples drawn from MetaCyc. EMS and random samples are of the same size.

metabolomics experiments, where a significant amount of peaks are still awaiting a proper identification. A big fraction of such unidentified compounds are admittedly originated as by-products of enzyme promiscuity. The ability of enzymes of accepting multiple substrates and of catalyzing multiple reactions provides alternative routes of biosynthesis and degradation that necessarily end up into novel compounds. The goal of quantifying such promiscuous effects, thus, is central to an appropriate determination of the metabolic space. For that purpose, we have proposed a solution based on extended connectivity fingerprints (ECFPs) that codes for changes in the atom bonding environments where the reactions are taking place.[24] Such an approach has been shown in multiple instances to be able to appropriately model enzyme promiscuity[20] and, more interestingly, to allow the computation in an efficient way of all compounds that potentially could be accessed *in vivo* through metabolic transformations. Moreover, by combining such modeling framework with the information from biological databases, the metabolic space restricted to one given chassis organism (metabolome) **S** can be also determined.

For instance, the metabolic database MetaCyc[19] in its 16.0 release contains 9623 metabolites and 10263 reactions. We have previously shown that this initial metabolic space can be expanded in the RetroPath framework through the application of the fingerprint coding.[18] For an atom bonding environment of 5 (diameter of 10),[24] we found that 25227 new compounds could be generated, within a total expanded space of 65827 reactions. The purpose of such expansion is to get a first estimate of the full extent of all compounds that can be accessed *in vivo*. However, we must assume that most of these compounds, even though they could potentially be produced through the expanded reactions, they might appear in weak concentrations, as they correspond to products of non-native reactions often with low efficiency. We analyzed more closely this issue in the case of an extract from *E. coli* BL21, a chassis organism that is often used for synthetic and metabolic engineering applications. The distribution of masses from the observed peaks in the spectrum obtained by LC/MS (shown in Figure 2a) followed a bimodal distribution (Figure 2d), with a first peak around 200 Da and a second peak of higher amplitude approximately at 400 Da. When this mass

dx.doi.org/10.1021/sb4001273 | *ACS Synth. Biol.* 2014, 3, 565−577

distribution is compared with the one corresponding to endogenous metabolites in *E. coli* according to the EcoCyc database, we observed that the distribution in this last database differs from the one observed in the spectrum, presenting a single peak around 200 Da. This result could indicate that such database does not contain annotations for all possible compounds that are actually produced, because of unannotated enzymes or enzyme promiscuity. Some of these unannotated metabolites might be already observed in other organisms, as in the case of the distribution of masses for all metabolites that are known to be produced in the MetaCyc database, which followed a distribution closer to the experimental one (Figure 2d). The assumption that the observed distribution might come from unannotated enzymes and promiscuous products was corroborated by the distribution of masses for all compounds present in the expanded metabolic space for a fingerprint diameter of 10, which, as shown in Figure 2d, provided a better coverage of the masses found in the spectrum. Therefore, the distribution of masses in the expanded space, which takes into account potential promiscuous activities of the enzymes, seems to describe better the actual metabolome of *E. coli*. In order to find matches between the masses found in that spectrum and masses of known metabolites, we systematically computed isotopic and adduct masses (including multimer ions) for metabolites in the EcoCyc database (Figure 2b and c), which comprised 2189 compounds. At precision of 10^{-3} Da, 359 *E. coli* compounds were assigned from the spectrum. We used then the expanded metabolic space in order to search for matches with compounds generated through the expansion. We limited the generation of new compounds to those that were catalyzed by promiscuous enzymes and for which all cosubstrates and coproducts were already known metabolites of *E. coli* BL21. Promiscuous enzymes were determined using a previously developed tool[20] and a recent list of promiscuous *E. coli* enzymes.[25] Our final metabolic space comprised 2189 metabolites already known to be in BL21 and 712 novel compounds. Next, adduct masses and isotope masses were computed for each of the compounds of the BL21 expanded metabolic space. The resulting masses were systematically searched in the MS spectra shown in Figure 2e. Finally, compound assignments were ranked based on the number of matches with the MS spectra found in a 10 s retention time window. Using that procedure, we were able of assigning 431 metabolites either already known to be in BL21 or generated through promiscuous enzyme activity. The metabolite assignment at that level of precision (i.e., 10^{-3} Da) was significantly higher than the one that was obtained in a test performed through 5 random samples of metabolites (Figure 2e), showing the ability of the expanded metabolic space of increasing the number of compounds assignments in the chassis metabolome.

Design Space. Metabolic circuit specifications define the problem that needs to be solved in our previously introduced metabolic space. Although several types of metabolic circuits with different functionalities can be conceived, all of them conceptually belong to the design space of solutions to the general problem of finding circuits within the metabolic scope between a given source **S** and the target set of metabolites **T**. Table 1 outlines typical input/output specifications (see formal details for each type of specifications in the Methods section). Moreover, other specifications can be envisaged such as maximum number of involved enzymes or desired intermediates, which altogether constitute a set of structural constraints imposed to the solutions.

Table 1. Metabolic Circuit I/O Specifications

circuit type	source	target
production	chassis	added-value chemicals
sensor	biomarkers	effectors
regulation	products	effectors
processing	signal metabolites	signal metabolites

Starting from the given set of input/output specifications, the circuit scope is computed in order to find the solution to a problem that can be stated in the following way: given an initial set of source metabolites **S** and a final set of target metabolites **T**, find the set of enzymes that are at least involved in one minimal pathway (as defined in Methods) connecting elements of **T** to the source **S**; that is, the scope should contain only enzymes that are at least essential for establishing one of the metabolic pathways. We have recently shown that the scope problem can be appropriately approached through the hypergraph formalism and the retrosynthesis technique.[26] Namely, the scope algorithm is a two-step process that allows finding the set consisting of all reactions involved in the production of a specific compound given a set of initial compounds. The first step of the algorithm (the forward step) ensures that among the list of reactions in the metabolic space, only the reachable ones are kept. This step creates a reduced graph represented as a predecessor list, which is then traversed by the backward step. Detailed descriptions of both algorithms are given in Methods. In addition, we have extended the basic scope algorithms in two different ways, (1) excluding compounds for which the reactions producing them are not stored (this allows for example to represent the cofactors or currency metabolites of the reactions, i.e., the metabolites taking part into large number of reactions, e.g. NADH, ATP, CO_2, etc.), and (2) by establishing a minimal scope mode that stops the forward step at the iteration where the target compound becomes reachable (in this manner, the network obtained is the minimal reaction network allowing to produce the target compound given a set of initial compounds). The scope algorithm, therefore, allows finding a subspace of the reaction space (minimal or maximal, depending on the scope version used) that can contain all the circuits corresponding to the given specifications.

Once the circuit scope is determined, the next step is to enumerate all metabolic circuits connecting the source to the circuit's target, which is a computationally complex problem that we have recently addressed[26] through different approaches such as elementary modes or topological analyses. The approach based on elementary flux modes (EFMs) searches for all minimal pathways in a network from which any pathway can be obtained. Such approach, originally proposed for biosynthetic pathways (production circuits), can be seamlessly extended to the general problem of enumerating the metabolic circuits *C* in Table 1 in their corresponding metabolic scopes by defining appropriate constraints into the solution types (see a detailed description in the Methods section).

These previous results showed that the number of alternative circuits performing a desired task can be quite high, and therefore, the implementation of all possible circuits in order to select the best performance might be rendered nonfeasible for practical reasons. Determining the optimal solution, however, is nontrivial. Metabolic circuits are often multistep processes involving many enzymes whose activities have to be finely controlled in concert for optimal performance. To that end, a

dx.doi.org/10.1021/sb4001273 | *ACS Synth. Biol.* 2014, 3, 565−577

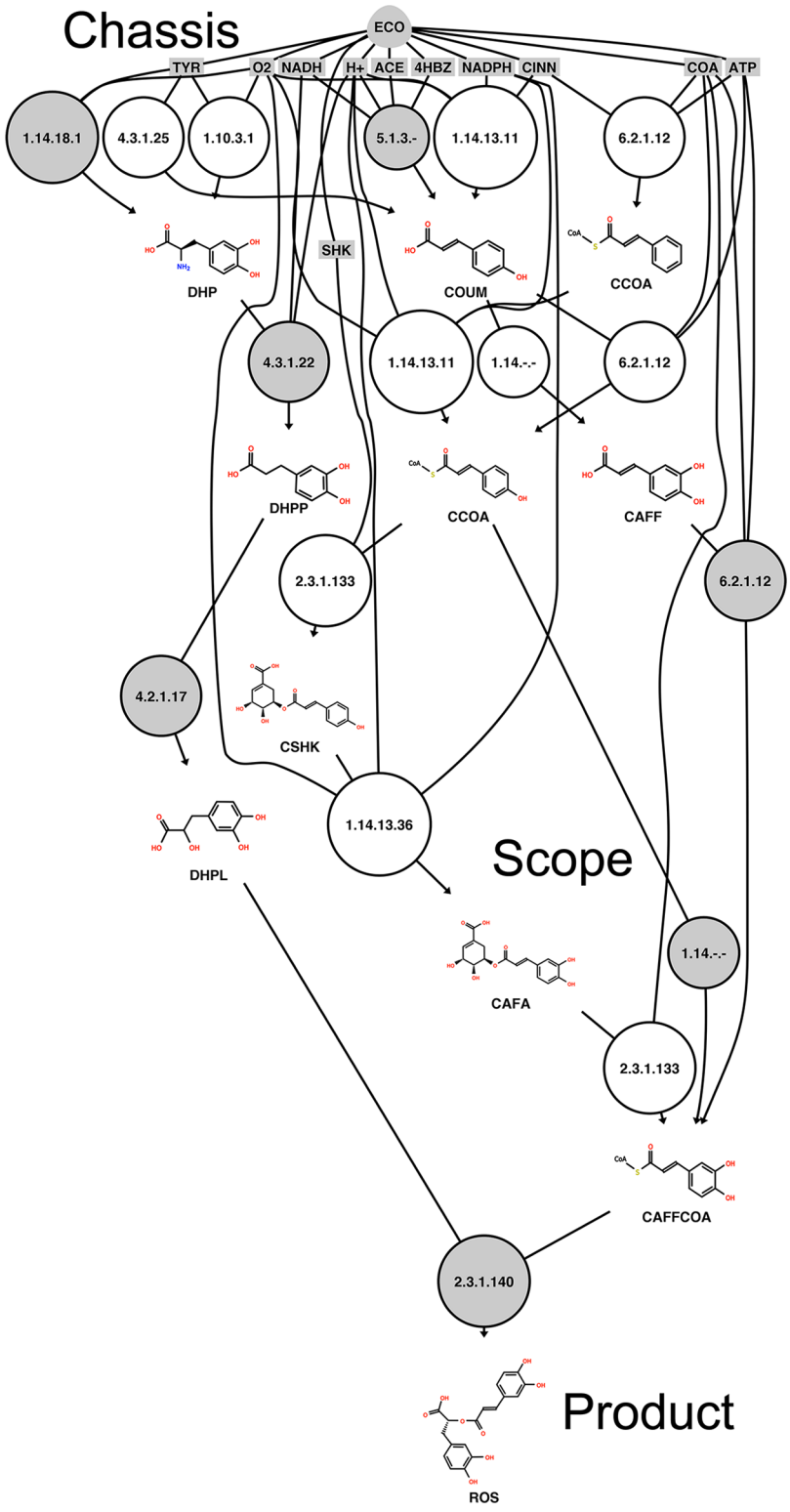


Figure 3. Metabolic scope of heterologous pathways in *E. coli* leading to the production of rosmarinate (ROS). Endogenous precursors tyrosine (TYR), acetate (ACE), 4-hydroxybenzoate (4HBZ), trans-cinnamate (CINN) and cofactors O_2, NADH, NADPH, CoA, ATP, are transformed through successive steps by catalytic steps. Alternative routes can be found because of enzyme promiscuity, such as for enzymes EC 1.14.13.11 or EC 6.2.1.12. In total, 12 pathways can be enumerated in the scope. The pathway with the best ranking is highlighted in gray. Intermediates in the pathways are 3,4-dihydroxy-L-phenylalanine (DHP), 4-coumarate (COUM), cinnamoyl-CoA (CCOA), 3-(3,4-dihydroxyphenyl)lactate (DHPL), caffeate (CAF), 3,4-dihydroxyphenylpropanoate (DHPP), p-coumaroyl-CoA (CCOA), 4-coumaroylshikimate (CSHK), 5-O-caffeoylshikimic acid (CAFA), caffeoyl-CoA (CAFFCO). Only compounds involved in pathways are represented.

circuit cost function is defined $J(C)$ that typically is computed as an aggregate function of individual costs of elements in the circuit, that is, enzymes and metabolites, as well as topological properties of the network, such as maximum expected flux,

dx.doi.org/10.1021/sb4001273 | ACS Synth. Biol. 2014, 3, 565−577

cross-talk, redox balance, etc. More precisely, in our RetroPath implementation, we are considering the following aspects:[27] (1) reaction efficiency—since reactions involved in the pathway can be either catalyzed natively or promiscuously by the enzyme, the degree of promiscuity of the enzyme, thus, has to be evaluated;[20] (2) inhibition effects, the accumulation of intermediates in the pathway can lead to inhibition of growth because of toxicity effects of the compound, and the toxicity in the chassis, thus, needs to be estimated;[28] (3) perturbation effects, since the insertion of the desired metabolic circuit necessarily leads to the perturbation in the equilibrium of metabolism fluxes. The organism readjusts its flux states in order to find an optimal state between competing objectives. Our goal, therefore, is to find the optimal circuit configuration leading to an efficient operation of the circuit through an adequate ranking of the multiple aspects influencing its performance.

Bioproduction Circuits. The previous methodology was applied to the MetaCyc database in order to determine all possible circuits that can be imported into *E. coli* for producing heterologous compounds. A detailed discussion about this type of circuits based on the KEGG database[29] has been presented elsewhere by the authors.[18] In a similar order of magnitude as in a previous study, we found for the MetaCyc database 2180 heterologous compounds that could be connected through production circuits to the *E. coli* chassis. We applied the previous two-step scope algorithm in order to compute these values finding an average scope of 39 reactions and maximum scope of 397 reactions. We applied next the circuit enumeration algorithm in order to enumerate the circuits contained in the previously computed scopes. Figure 1d shows the distribution of found pathway lengths. In total, we found 71307 production, which followed approximately a Gaussian distribution, with an average length around 8 enzymes. Approximately 15% of the cases contained more than 10 circuits, and 3% more than 10^2. More generally, the length of the solution circuits (in terms of reactions) can greatly vary depending on the complexity of the molecule to create and its distance to the metabolic network of the chassis organisms. Some of these compounds were value-added chemicals. For instance, we found in MetaCyc 290 saturated and unsaturated hydrocarbons that are not naturally produced by *E. coli*. Among them, RetroPath was able to connect 178 alkanes, alkenes, and alkynes ranging from C_1 to C_{40} to *E. coli* through production circuits. The solutions involved 300 alternative pathways, with a maximum of 30 alternative pathways for one single compound (pentadecane) and a maximum pathway length of 7 genes. In the expanded metabolic space, we found 33 additional hydrocarbons (excluding aromatics) that could be connected to *E. coli*. Another example, shown in Figure 3, is the metabolic scope for the production of rosmarinate, a natural phenol antioxidant carboxylic acid, in *E. coli*. The scope contains in total 12 possible pathways, where some alternative routes are found because of enzyme promiscuity, as for enzymes EC 1.14.13.11 or EC 6.2.1.12.

Biosensor Circuits. Similarly, metabolic sensor circuits that process biomarkers through heterologous transformations to *E. coli* into metabolites that can induce the expression of a desired gene were computed in the metabolic space of MetaCyc. We focused our studies on biomarkers of disease conditions that have been identified through metabolomics studies, that is, sets of metabolites (often called metabolic signatures) that can be linked to pathological states. To that end, we extracted from the accumulated knowledge of HMDB,[30] a human metabolomics database, a matrix containing 208 diseases and 408 biomarkers linked to human diseases. In order to implement sensor circuits, biomarkers need to be connected to metabolites that can induce gene expression. We selected for that purpose the list of effectors (36 activators and 42 inhibitors of transcription factors) in the RegulonDB database,[31] a comprehensive database of transcriptional regulation in *Escherichia coli*. We evaluated by these means the possibility of implementing synthetic biology devices in *E. coli* consisting of biosensor circuits for human diseases. A first step was to select those biomarkers for which RetroPath found metabolic pathways that led to the production of transcription factor effectors. By applying the described methodology for circuit enumeration under the constraints specific to biosensors (see details in Methods), we enumerated all heterologous pathways that can be implemented in the chassis linking biomarkers to inducers. We were able to find 80 biomarkers connecting to 22 inducers through a total of 331 pathways. For biosensors the average scope was of 70 reactions and maximum scope of 338 reactions. Figure 1e shows the distribution of pathways length for biosensors, which presented a bimodal distribution, with a first group composed of single-enzyme circuits, that is, biomarkers that can be transformed into an effector in a single step. This group formed the majority of the cases since in general most of the effectors could not be associated to long biosynthetic pathways and could be explained by the fact that the corresponding target compounds were already present in *E. coli*; hence, there existed reactions to directly produce them.

Furthermore, there were 57 diseases that were characterized by these 80 biomarkers and thus were found to be connected to 22 different effectors. Since it is known to which inducers the biomarkers can be connected, we were able to determine to which inducers the diseases could be linked. The procedure consisted of multiplying the matrix linking biomarkers to diseases by the matrix linking biomarkers to effectors to obtain another matrix giving the connections between diseases and effectors, as shown in the following equation:

$$\begin{matrix} & \overset{B_1 \cdots B_n}{} \\ D_1 \\ \vdots \\ D_m \end{matrix}\begin{pmatrix} & \\ & \end{pmatrix} * \begin{matrix} & \overset{E_1 \cdots E_p}{} \\ B_1 \\ \vdots \\ B_n \end{matrix}\begin{pmatrix} & \\ & \end{pmatrix} = \begin{matrix} & \overset{E_1 \cdots E_p}{} \\ D_1 \\ \vdots \\ D_m \end{matrix}\begin{pmatrix} & \\ & \end{pmatrix} \quad (1)$$

Based on this matrix, we were then able to link diseases to a specific set of effectors so that they cannot be mistaken with one another (see a description of the algorithm in Methods and a detailed example in the Supporting Information). For instance, diabetes mellitus was found to be linked to allantoin, since it can be produced from urate, a compound that is present at high concentration in subjects at risk of type 2 diabetes mellitus.[32] When reacting with water and dioxygen, urate is transformed in (S)-(+)-allantoin, which is a regulator of the *E. coli* transcriptional activator AllS and repressor AllR, which participate in the regulation of more than 10 genes that are involved in the allantoin catabolism pathway[33] (see Figure 4). Teramoto et al.[34] achieved a vector construct containing a green fluorescent protein (GFP) coding sequence next to the promoter of the *E. coli* gcl gene which is repressed by AllR. Such construct could be used to detect diabetes metillus since the presence of allantoin would repress the gcl gene promoter and thus the production of GFP, a detectable output signal.

dx.doi.org/10.1021/sb4001273 | *ACS Synth. Biol.* 2014, 3, 565−577

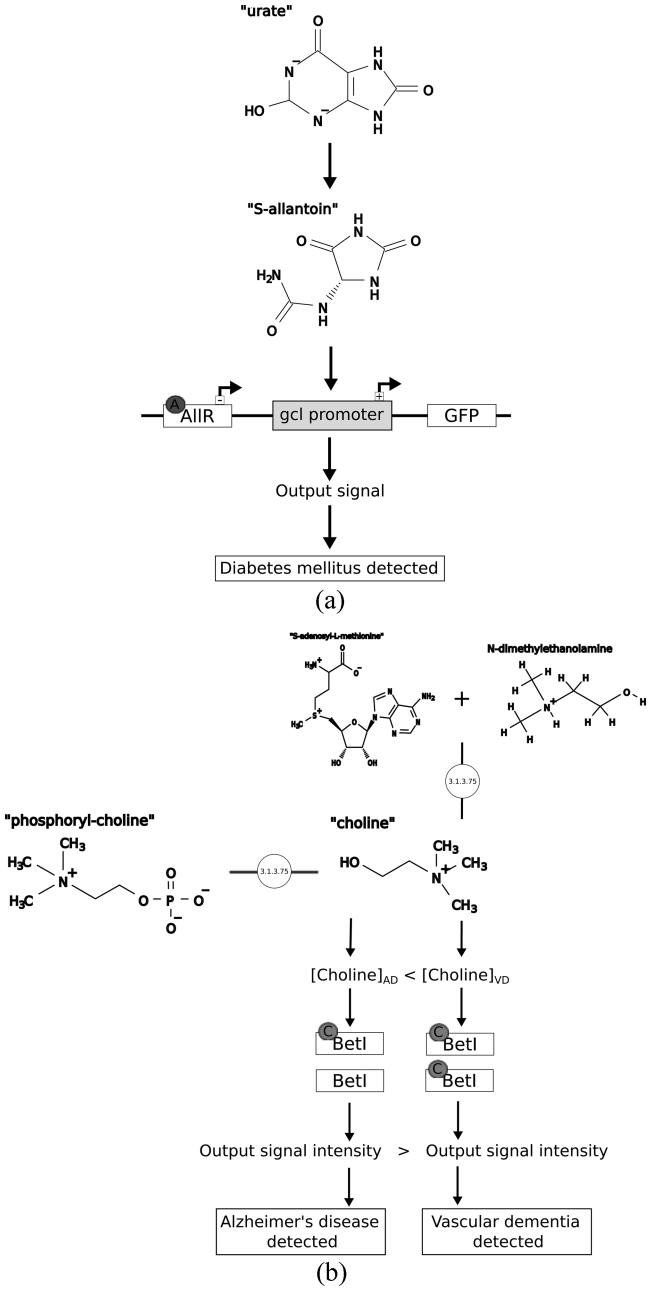


Figure 4. Examples of disease detection. (a) Scheme of the detection of diabetes mellitus: S-allantoin is produced from the reaction between urate, dioxygen, and water. Only compounds involved in the pathway are represented. Allantoin binds to AllR which will repress the gcl promoter that in the construct regulates GFP. The concentration of GFP is then decreased, which indicates that the person is at high risk of diabetes mellitus type 2. (b) Scheme of the detection of Alzheimer's disease (AD) and vascular dementia (VD). Choline is produced through enzyme EC 3.1.3.75 in the case of both diseases. It is also produced through enzyme EC 2.1.1 in the case of vascular dementia. More choline binds to betI repressor in the case of VD than for AD; thus, the gene it regulates is more repressed. The protein encoded by this gene serves as an output signal, which means the output signal intensity will be higher for AD than for VD.

We also managed to discriminate neurological disorders that can share similar symptoms by linking each disease to a different set of effectors. These neurological disorders could be linked to different or similar effectors (Table 2). The algorithm

Table 2. Diseases and Effectors to Which They Can Be Linked

effectors	acetate	allantoin	glycollate	choline
Alzheimer's disease				x
epilepsy	x		x	
meningitis		x		
multiplesclerosis	x			
vascular dementia				x

managed to select one different effector to be linked to each disease wherever was possible (Table 3). We particularly

Table 3. Transcription Factor Effectors Chosen by the Algorithm to Characterize Each Neurological Disorder and Biomakers from Which These Regulators Can Be Produced

diseases	biomarker(s)	effector
Alzheimer's disease	phosphorylcholine, S-adenosylmethionine	choline
epilepsy	methionine	glycollate
meningitis	allantoin	allantoin
multiplesclerosis	L-cysteine	acetate
vascular dementia	phosphorylcholine	choline

focused on differentiating vascular dementia and Alzheimer's diseases, which can be mistaken for each other since they share the common effector choline. They are both characterized by a more elevated level of phosphoryl-choline compared to healthy patients,[35] but patients with vascular dementia have a higher level of S-adenosylmethionine than Alzheimer patients. Indeed, cerebrospinal fluid of Alzheimer patients has a lower concentration of S-adenosylmethionine than cerebrospinal fluid of control subjects.[36] Thus, we can detect both diseases by using the reaction catalyzed by the phosphocholine phosphatase enzyme which transforms phosphoryl-choline into choline:

$$H_2O + \text{phosphorylcholine} \rightarrow \text{choline} + PO_4^{3-}$$

Then, we can distinguish vascular dementia from Alzheimer's disease using the reaction catalyzed by the ethanolamine N-methyltransferase enzyme, which transforms S-adenosylmethionine into choline and adenosyl-homocys:

$$\text{S-adenosylmethionine} + \text{N-dimethylethanolamine} \rightarrow \text{choline} + \text{adenosyl-homosys} + H^+$$

This means that in the case of vascular dementia, more choline would be produced than in the case of Alzheimer's disease. Choline binds to the betI promoters which then represses the betA and betT genes in *E. coli*.[37] Such repressor could be used to regulate a gene triggering a detectable signal such as GFP. Thus, the output signal intensity in the case of Alzheimer's disease would be higher than in the case of vascular dementia given that the repression increases along with choline concentration (Figure 4). One must note that the genes coding for the enzymes phosphocholine phosphatase and ethanolamine N-methyltransferase have to be implemented in *E. coli*.

Bioregulation Circuits. Regulation circuits are at least composed of two modules, one production module and one sensor module. These type of circuits can be tuned in order to control the production level of a desired compound. To that end, the sensor module implements a feedback loop by measuring the presence of some chemical that can be related to

dx.doi.org/10.1021/sb4001273 | ACS Synth. Biol. 2014, 3, 565−577

the concentration of the target product and by connecting such sensor into an effector circuit that should eventually regulate the expression of some element linked to the production circuit. Several mechanisms of regulation are possible. Here, we focused on mechanisms of regulation that are based on metabolic circuits that connect the compound of interest to an inducer through enzymatic transformations. This case, therefore, is a combination of the previous cases of metabolic production and sensing. Starting from the previous list of heterologous compounds that can be produced in *E. coli* and through the application of the described enumeration methodology, we found 157 heterologous products in *E. coli* that can be connected to activators and 18 products that can be connected to inhibitors through metabolic pathways (Table 4).

Table 4. Summary of the Results for Regulation Circuits

effector type	total products	total pathways
activator	157	1859
inhibitor	18	65

The scope for bioregulators contained an average of 90 reactions and a maximum of 483 reactions. The distribution of pathway lengths for the bioregulation circuits is shown in Figure 1f, with a total of 55613 enumerated circuits. Since they are a composition of the Gaussian distribution typically seen in production circuits and the distribution for biosensors, the curve follows approximately a skewed Gaussian, with an average pathway length around 8. Interestingly, the number of enzymes needed to form such synthetic bioregulators seemed to be higher than the ones that are naturally found in organisms. For instance, we looked at the regulation circuits found in *E. coli* in the EcoCyc database. The lengths of these pathways are shown in Table 5, with a maximum length of 9 and an average length of approximately 4.

Table 5. Length of Natural Regulation Circuits for Amino Acid Biosynthesis

amino acid	regulation loop length
alanine	1,2,2,2
arginine	8
asparagine	1
glycine	1,4
leucine	5
lysine	9
phenylalanine	3
tryptophan	6
tyrosine	3

One example of this type of regulation circuits is given in Figure 5, where the production in *E. coli* of 4-coumaroyl-CoA, a precursor for flavonoids,[38] is regulated through the insertion of the eukaryote pathway for the biosynthesis of 4-hydroxybenzoate,[39] which, in turn, could be used to induce the gene expression[40] regulating the pathway. This circuit could be therefore used to regulate the production of biosynthetic pathways for flavanoids derivatives, which constitute a family of compounds whose efficient production in *E. coli* is of interest because of their properties for therapeutics, cosmetics and food industry.

Bioprocessing Circuits. Metabolic circuits can be used as well for processing signals and for implementing logical

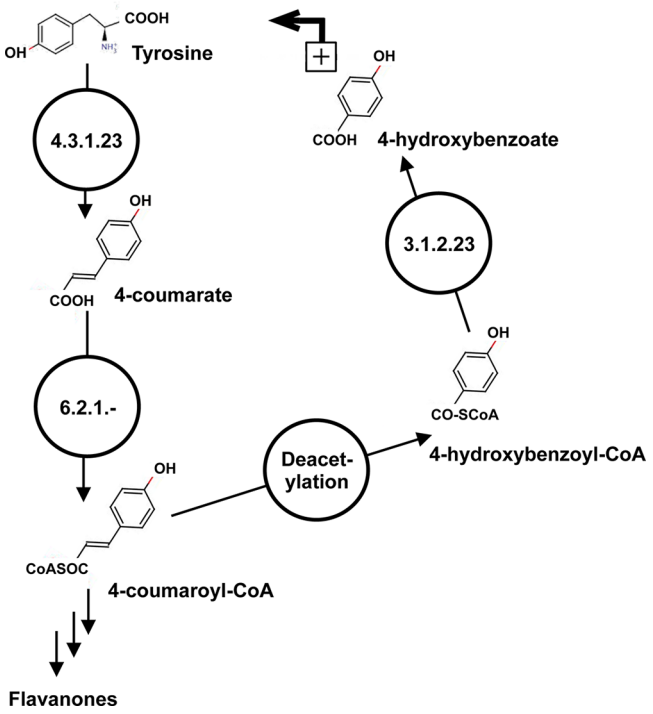


Figure 5. Example of regulation of production of compounds. 4-coumaroyl-CoA, a main precursor for the flavonoid biosynthesis pathway, can be sensed through a 2-step circuit that converts the precursor into 4-hydroxybenzoyl-CoA and then into 4-hydroxybenzoate, an effector. 4-coumaroyl-CoA can be heterologously produced in *E. coli* in 2 steps from tyrosine through the 4-coumarate intermediate. Only compounds involved in the pathway are represented. Scheme of regulation of production of 4-coumaroyl, a main precursor for flavonoids, through a sensing circuit that converts it into 4-hydroxybenzoate, an inducer.

circuits.[41] For instance, we have enumerated in the MetaCyc database the number of orthogonal metabolic AND and OR logic gates that could be implemented in *E. coli*, that is, logic gates implemented through heterologous enzymes having no cross-talk with endogenous metabolites of the chassis. In total, we found 2595 OR gates (i.e., a signal metabolite that can be produced by an enzyme from different substrates) that could be implemented through 374 heterologous enzymes and 7904 AND gates (i.e., a signal metabolite that is produced through a reaction involving multiple signal metabolites) that could be implemented through 481 heterologous enzymes. An example of such type of logic relationship between signal metabolites is given in Figure 6, where and AND gate is used in combination with biosensing. More complex logic operations and other metabolic logic gates might be possible to implement[15] by means of other metabolic circuits.

RetroPath: An Automated Design Software for Metabolic Circuits. Previous results showed the usefulness of computer-aided tools for automated biodesign of metabolic circuits. To that end, we have developed RetroPath, an automated software that implements the described design methodologies for metabolic circuits, allowing the user to explore different types of circuits based on the specifications, as detailed in Methods (see Figure 7). RetroPath consists of several modules in charge of the different presented algorithms: (1) the scope and stoichiometric matrix construction algorithms are implemented as a C++ program; (2) the elementary flux modes computations relies on the efmtool

dx.doi.org/10.1021/sb4001273 | *ACS Synth. Biol.* 2014, 3, 565−577

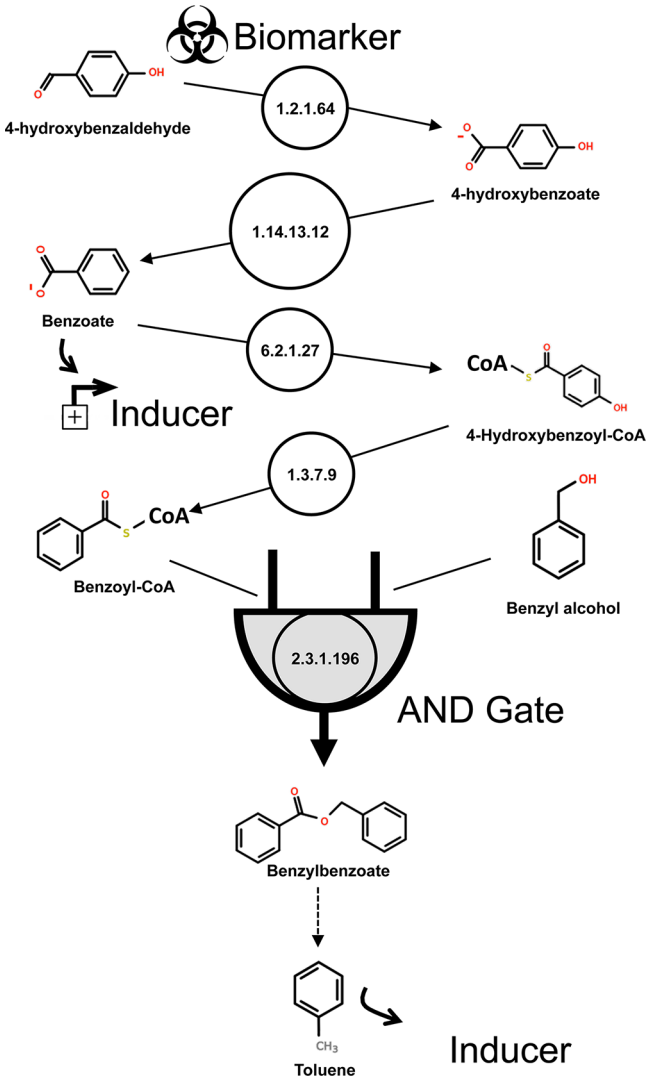


Figure 6. Example of advanced metabolic circuit. 4-hydroxybenzaldehyde, a prostate cancer biomarker is transformed into 4-hydroxybenzoate and benzoate through enzymes EC 1.2.1.64 and EC 1.14.13.12. These molecules can be used as inducers, constituting a biosensor circuit. Furthermore, benzoate can be transformed into 4-hydroxy benzoyl CoA and further into benzoyl-CoA, through EC 6.2.1.27 and EC 1.3.7.9, respectively. This last compound can be used to form an AND logic gate in combination with benzyl-alcohol to produce benzylbenzoate through enzyme EC 2.3.1.196, that could be further degraded into toluene, another inducer. Only compounds involved in the pathway are represented.

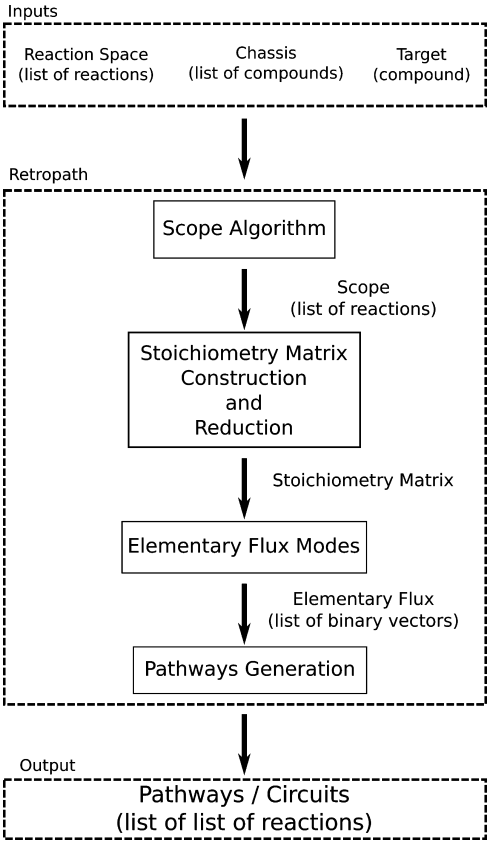


Figure 7. RetroPath pipeline. The inputs are taken from a given metabolic database and they are processed in order to generate the scope, generate the stoichiometry matrix, compute the EMFs and enumerate the pathways. The output can be further processed to create an image of the pathways or converted into the desired format.

software,[42] written in Java. Retropath requires as input the metabolic space and the corresponding constraints of the circuit, defined through two formats: (1) the compound file format as used by RetroPath simply consists in a file listing the compounds as found in the database in use; (2) the reaction file format is specified in an EBNF format presented in Supporting Infomration Figure S1. Regarding the output, metabolic circuits can be exported into SBML format[43] (see Supporting Information Figure S2), allowing them to be embedded into *in silico* reconstruction models of chassis organisms such as *E. coli*. This feature allows simulation of distribution of fluxes. In addition, DNA sequences encoding for the corresponding enzymes can be annotated in the model through SBOL (Synthetic Biology Open Language) definitions[44] (Supporting Information Figure S2). Access to RetroPath is currently availble upon request.

METHODS

Definitions. The **metabolic space**, denoted M, represents all the possible compounds C and allowed transformations (reactions) R between compounds. It is represented as an ordered pair: $M = (C,R)$. A **reaction** is defined as a transformation from a set of substrate compounds to as set of product compounds: $R{:}C_1 \rightarrow C_2$. A reaction is represented as an ordered pair: $R = (C_1,C_2)$. $C_1 \subset C$ represents the substrate compounds and $C_2 \subset C$ the product compounds. A **chassis** is a subset of the metabolic space M that corresponds to the metabolic network of a host organism for metabolic circuits. Such subset is denoted by $O = (C_O,R_O) \subset M$ and contains all metabolites and reactions that are considered to be initially present. **Currency metabolites** consist of the set $C_K \subset C$ that contains all compounds that act as cofactors in the metabolic reactions. A **metabolic circuit** $S = (C_S,R_S) \subset M = (C,R)$ is a subset of the metabolic space M that verifies a list of constraints λ and some given specifications ε. The **circuit interface** consists of two sets of compounds: (1) the **precursor set** $Prec \subset C_S$ are all compounds that need to be available in the medium in order to make viable all the reactions R_s in the circuit; (2) the **product set** $Prod \subset C_s$ are all compounds that are produced by the reactions R_s in the circuit and not further consumed. Compounds that do not belong to these two subsets are

dx.doi.org/10.1021/sb4001273 | *ACS Synth. Biol.* 2014, 3, 565−577

defined as *inner compounds* of the circuit $C_I \subset C$: $C_I \cap u \cap y = \varnothing$ (see definitions for u and y below).

Circuit Specifications. The specifications of a metabolic circuit dictate relationships between selected precursors, known as the input set $u \subset Pred$; and selected products, known as the output set $y \subset Prod$. Without loss of generality, a circuit specification is defined by some relationship between the input and the output that the circuit should implement: ε_S: $y = S(u)$. General constraints on metabolic circuits, denoted by λ, should be verified by any metabolic circuit. They can be classified into three categories: (1) **Compound constraints**—these are constraints that are usually defined in order to limit the possible type of compounds appearing in the circuits. The most common compound constraint is that of heterogeneity; that is, inner compounds should not contain chassis compounds (excluded currency compounds): λ_{CH}: $(C_I \setminus C_K) \cap C_O = \varnothing$. (2) **Reactions constraints**—these constraints affect which metabolic reactions can appear in the circuits. A basic requirement is that of heterogeneity of reactions; that is, reactions in the circuit do not appear in the chassis: λ_{RH}: $R_S \cap R_O = \varnothing$. (3) **Topology constraints**—these constraints affect the structures of the accepted solution networks. Typically, in order to avoid metabolic burden, we require a metabolic circuit to implement a "minimal pathway",[26] which consists of a set of reactions where the removal of any of them would render the circuit nonviable, that is, λ_{MP}:$S^* = (C_s^*, R_s^*)$ s.t $\nexists S' = (C_{S'}, R_{S'} | R_{S'} \subset R_S \wedge y = S'(u))$.

Besides the general constraints previously described, λ_{CH}, λ_{RH}, and λ_{MP}, which, unless stated, are common to all circuits, some others might appear depending on the specifications of the metabolic circuit: (1) **Production circuits**—the objective of production circuits is to produce a metabolite in the chassis by importing genes from other organisms. Specification: $u \subset C_O$; $y \subset C \setminus C_O$. (2) **Biosensing circuits**—here, the objective is to transform a set of compounds of interest (biomarkers) into compounds that can regulate gene expression. Specification: $u \subset C$ biomarkers; $y \subset C$ effectors of gene expression. (3) **Bioregulation circuits**—they are formed by the interconnection of two metabolic circuits, one production circuit S_1 and one biosensing circuit S_2, which should be orthogonal. Specification: $S_1 \perp S_2$. (4) **Bioprocessing circuits**—in these circuits, the relationship between the input signal compounds and the output signal compounds should follow a particular transfer function. Specification: $y = f(u)$.

Circuit Scope. The scope S_c of a circuit specification $y = S(u)$ is a metabolic space that contains all reactions involved in at least one circuit that in addition to the specification, verifies the circuit constraints. The scope algorithm ensures that the circuit verifies the compound and reaction constraints λ_{CH}, λ_{RH}. The algorithm consists of two steps: (1) The **forward step** (Algorithm 1), starts from the set of initial compounds (input set u) and iteratively finds newly reachable compounds. The iterations stop when a fixpoint is reached; that is, no newly reachable compounds were found in the last iteration. For each compound, the algorithm keeps track of which reaction produced it. A compound is reachable if there exists a reachable reaction that can produce it, that is, a reaction for which all its substrates are reachable. (2) The **backward step** (Algorithm 2), starts from the compound of interest (output set y) and goes back to the initial compounds as follows: For each reaction that can produce the target compound, add it to the scope and recursively apply the same procedure on each substrate of the reaction. The recursion stops when initial compounds are reached (there is no reaction to produce it).

```
Data:
init: Set of Initial Compounds
reactions: List of reactions
Result: pred: Map(compound → list of reactions)
/* code starts here                                          */
pred ← ∅;
unreached_reactions ← reactions;
while not fixpoint do
    for r ∈ unreached_reactions do
        if reachable(pred, r) then
            unreached_reactions ← unreached_reactions \ {r};
            for p ∈ products(r) do
                pred[p] ← pred[p] ∪ r;
            end
        end
    end
end
return pred
```

Algorithm 1.

Forward step of the scope algorithm, the condition $reachable(r)$ can be translated as $\forall s \in substrates(r), s \in pred$ and denotes the fact that all the substrates of the reaction r have been reached.

```
Function backward:
    Data:
    pred: Map(compound → list of reactions)
    comp: Compound
    Result: scope: list of reactions
    /* Code starts here                                      */
    scope ← ∅;
    if pred[comp] then
        scope ← Scope ∪ pred[comp];
        for r ∈ pred[comp] do
            for s ∈ substrates(r) do
                scope ← scope ∪ backward(pred, s);
            end
        end
    return scope
end
```

Algorithm 2.

Backward step of the scope algorithm, enumerating all reactions involved in the production of a target compound.

Circuit Enumeration. From the scope S_c of a circuit specification $y = S(u)$, the enumeration of all viable circuits can be enumerated by following several algorithms.[26] One of the possible algorithm is the one based on the decomposition of the reactions that span the scope S_c into their elementary flux modes (EFMs).[45] EFMs are the set of minimal pathways that are nontrivial solutions to the steady-state equation whose combination can describe any possible path in the network. Several toolboxes exist that allow to efficiently compute the elementary modes from the stoichiometric matrix $\mathbf{S}$ by computing the nontrivial solutions to the steady-state equation: $\mathbf{Sv} = \mathbf{0}$, where $\mathbf{S}$ is a matrix where each row corresponds to a compound in the circuit C_s and each column to a reaction $\boldsymbol{R}_s$ and the value of each cell is the stoichiometric coefficient. In order to solve the circuit enumeration problem, the stoichiometric matrix is constructed in a specific manner starting from the scope of the target compound: (1) remove all rows representing initial compounds; (2) remove all rows representing compounds that are produced by a reaction but never used in any other; (3) merge equivalent column that can have been created because of steps 1 and 2 (it consists in deleting redundant columns and renaming the remaining

dx.doi.org/10.1021/sb4001273 | ACS Synth. Biol. 2014, 3, 565−577

column with the names of all reactions); (4) add an additional column to create a flux out for the target compound. By following these steps, one can obtain a minimal matrix modeling the circuit enumeration problem that is suitable to use with the EFM algorithm. Since we are only interested in solutions that involve the input set u and the output set y, further simplifications can be done in the matrix for faster computation by removing precursors and products that do not form part of the input/output sets of interest. In that way, each EFMs provides a solution that connects the input u to the output y and solutions can be filtered in order to verify the constraint of minimal pathway λ_{MP}.

Disease-Biomarker Biosensor Detection. In the case of biosensor circuits, the problem of circuit enumeration is augmented with the problem of finding a right combination of effectors for detecting multiple diseases. To detect a certain group of diseases, we built an algorithm that assigns effectors or combination of effectors to biomarkers associated with diseases states so that the disease can be easily detected. The starting point is a matrix showing which effectors can be linked to which disease. Such matrix is computed through the application of the RetroPath workflow for determining circuits that can connect the input set of 408 biomarkers linked to human diseases to the set of 78 effectors from RegulonDB with the ability of activating or inhibiting transcription factors. In total, RetroPath found 80 biomarkers connected to 22 effectors. From this matrix, the algorithm finds an optimal way to link a disease to a unique set of effectors so that it cannot be mistaken with one another. It is composed of three steps and returns for each disease one effector or a set of effectors which will be used to detect the disease by inducing the expression of one or more desired genes (a detailed example can be found in the Supporting Information):

a. In the first step, the algorithm finds diseases that can be characterized by only one effector. After selecting such effector and disease couple, it deletes the row and column representing them from the previous matrix. It redoes it for the newly formed matrix since deleting a disease can "artificially" make an effector now linked to only one disease.
b. The second step consists in reordering the rows of the matrix such that diseases that can be characterized by fewer effectors than the other ones are at the top and the ones that are characterized by a more important number of effectors are at the bottom.
c. The last step has for goal to assign a unique combination of effectors to characterize each disease (Algorithm 3). The idea is to proceed per round for a particular disease. In the first one, the algorithm searches for one effector that is not already linked to another disease. If they are all already linked to another disease, then it starts searching for a combination of two effectors that does not already characterize a disease, etc. It continues in that way until it finds the right combination or when the only combination possible is the combination of all the effectors that can be linked to this disease.

Data Sets. The metabolic space was computed using the MetaCyc database[19] in its 16.0 release containing 9623 metabolites and 10263 reactions. Disease biomarkers were obtained from the HMDB database,[30] comprising 208 metabolite biomarkers and 408 disease conditions. Effectors (36 activators and 42 inhibitors of transcription factors) in *Escherichia coli* were taken from the RegulonDB database.[31]

```
Function assignCombinations:
    Data: M: matrix linking diseases (rows) to effectors (columns)
    Result: linkDE: Map (Disease → list of Effectors)
    linkDE ← ∅ ;
    used ← ∅ ;
    for D ∈ rows(M) do
        while D ∉ linkDE ∧ not all combinations tried do
            combination ← next_comb(D, M);
            if combination ∉ used then
                linkDE ← linkDE ∪ {D, combination} ;
                used ← used ∪ {combination} ;
            end
        end
    end
    return linkDE;
end
```

Algorithm 3.

Algorithm of the third step: assigns a unique combination of effectors to characterize each disease. *next_comb* is a function giving a new possibe combination of effectors linked to a disease D given the matrix M.

Experimental Procedure for *E. coli* Metabolite Identification. We cultured *E. coli* BL21 strains at 37 °C and lysed the cells through a freeze/defreeze cycle. The sample was then subjected to LC/MS analyses. Analyses were carried out with a Nexera LC system (Shimadzu, Marne la Vallée, France) coupled with an Exactive mass spectrometer (Thermo-Fisher scientifics, Courtaboeuf, France). The chromatographic separation was performed on a Discovery HSF5 pentafluorophenylpropyl (PFPP) column (2.1 mm × 150 mm, 5 μm) from Sigma-Aldrich (Saint Quentin Fallavier, France), equipped with an online prefilter (Interchim, Montlucon, France). The mobile phases were (A) 100% water and (B) 100% acetonitrile, containing 0.1% formic acid. Elution was performed at a flow rate of 250 μL/min, using the following gradient conditions: 0−2 min, 5% B; 2−20 min, from 5 to 100% B; 20−24 min, 100% B; 24.1−30 min, 5% B. Bacterial extract (10 μL) was injected into the LC/MS system. Mass spectrometric detection was performed using an Orbitrap-Exactive mass spectrometer (Thermo Fisher Scientifics, Courtaboeuf, France) fitted with an electrospray source (ESI) operated in positive ion mode, at a mass resolving power of 100 000 ($m/\Delta m$, full width at half-maximum of peak intensity, for an ion at 400 Da).

■ ASSOCIATED CONTENT

Supporting Information

Section S1. Example of the algorithm that finds unique combinations of effectors to characterize disease. Figure S1. Syntax used to represent a metabolic reaction in the RetroPath software. Figure S2. Pseudocode following the SBML data model structure and data fields for the metabolic circuit. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*Phone: +33 01 69 47 53 90. Fax: +33 01 69 47 44 37. E-mail: pablo.carbonell@issb.genopole.fr.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

Genopole through an ATIGE grant; PRES UniverSud Paris through a student scholarship; Bio-Design Automation Consortium, Inc. (BDAC) through a student scholarship; abSYNTH facilities; and ANR through a Chair of Excellence.

■ REFERENCES

(1) Khalil, A. S., and Collins, J. J. (2010) Synthetic biology: Applications come of age. *Nat. Rev. Genet. 11*, 367−379.
(2) Nandagopal, N., and Elowitz, M. B. (2011) Synthetic biology: Integrated gene circuits. *Science 333*, 1244−1248.
(3) Papini, M., Salazar, M., and Nielsen, J. (2010) Systems biology of industrial microorganisms. *Adv. Biochem. Eng. Biot. 120*, 51−99.
(4) Curran, K. A., and Alper, H. S. (2012) Expanding the chemical palate of cells by combining systems biology and metabolic engineering. *Metab. Eng. 14*, 289−297.
(5) Paddon, C. J., et al. (2013) High-level semi-synthetic production of the potent antimalarial artemisinin. *Nature 496*, 528−532.
(6) Ajikumar, P. K., Xiao, W.-H., Tyo, K. E. J., Wang, Y., Simeon, F., Leonard, E., Mucha, O., Phon, T. H., Pfeifer, B., and Stephanopoulos, G. (2010) Isoprenoid pathway optimization for taxol precursor overproduction in *Escherichia coli*. *Science 330*, 70−74.
(7) Chandran, S. S., Kealey, J. T., and Reeves, C. D. (2011) Microbial production of isoprenoids. *Process Biochem. 46*, 1703−1710.
(8) Santos, C. N., Koffas, M., and Stephanopoulos, G. (2011) Optimization of a heterologous pathway for the production of flavonoids from glucose. *Metab. Eng. 13*, 392−400.
(9) Kemmer, C., Gitzinger, M., Daoud-El Baba, M., Djonov, V., Stelling, J., and Fussenegger, M. (2010) Self-sufficient control of urate homeostasis in mice by a synthetic circuit. *Nat. Biotechnol. 28*, 355−360.
(10) Saeidi, N., Wong, C. K. K., Lo, T.-M. M., Nguyen, H. X. X., Ling, H., Leong, S. S. J. S., Poh, C. L. L., and Chang, M. W. W. (2011) Engineering microbes to sense and eradicate *Pseudomonas aeruginosa*, a human pathogen. *Mol. Syst. Biol. 7*, 521.
(11) Gutteridge, A., Kanehisa, M., and Goto, S. (2007) Regulation of metabolic networks by small molecule metabolites. *BMC Bioinformatics 8*, 88.
(12) 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.
(13) Anesiadis, N., Kobayashi, H., Cluett, W. R., and Mahadevan, R. (2013) Analysis and design of a genetic circuit for dynamic metabolic engineering. *ACS Synth. Biol. 2*, 442−452.
(14) Ang, J., and McMillen, D. R. (2013) Physical constraints on biological integral control design forhomeostasis and sensory adaptation. *Biophys. J. 104*, 505−515.
(15) HalÃąmek, J., Bocharova, V., Arugula, M. A., Strack, G., Privman, V., and Katz, E. (2011) Realization and properties of biochemical-computing biocatalytic XOR Gate based on enzyme inhibition by a substrate. *J. Phys. Chem. B 115*, 9838−9845.
(16) Silva-Rocha, R., Tamames, J., dos Santos, V. M., and de Lorenzo, V. (2011) The logicome of environmental bacteria: Merging catabolic and regulatory events with Boolean formalisms. *Environ. Microbiol. 13*, 2389−2402.
(17) Shlomi, T., Eisenberg, Y., Sharan, R., and Ruppin, E. (2007) A genome-scale computational study of the interplay between transcriptional regulation and metabolism. *Mol. Syst. Biol. 3*, 101.
(18) 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+.
(19) Caspi, R., et al. (2012) The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Res. 40*, D742−D753.
(20) Carbonell, P., and Faulon, J.-L. (2010) Molecular signatures-based prediction of enzyme promiscuity. *Bioinformatics 26*, 2012−2019.
(21) Cobb, R. E., Sun, N., and Zhao, H. (2013) Directed evolution as a powerful synthetic biology tool. *Methods 60*, 81−90.
(22) Abatemarco, J., Hill, A., and Alper, H. S. (2013) Expanding the metabolic engineering toolbox with directed evolution. *Biotechnol. J.*, DOI: 10.1002/biot.201300021.
(23) Kang, E., Jackson, E., and Schulte, W. (2011) An approach for effective design space exploration. *Proc. 16th Monterey Conf. Found. of Comp. Soft.*, 33−54.
(24) Carbonell, P., Carlsson, L., and Faulon, J.-L. (2013) Stereo signature molecular descriptor. *J. Chem. Inf. Model. 53*, 887−897.
(25) Nam, H., Lewis, N. E., Lerman, J. A., Lee, D.-H. H., Chang, R. L., Kim, D., and Palsson, B. O. (2012) Network context and selection in the evolution to enzyme specificity. *Science 337*, 1101−1104.
(26) Carbonell, P., Fichera, D., Pandit, S., and Faulon, J. L. (2012) Enumerating metabolic pathways for the production of heterologous target chemicals in chassis organisms. *BMC Syst. Biol. 6*, 10+.
(27) Carbonell, P., Planson, A.-G. G., and Faulon, J.-L. L. (2013) Retrosynthetic design of heterologous pathways. *Methods Mol. Biol. 985*, 149−173.
(28) Planson, A.-G., Carbonell, P., Paillard, E., Pollet, N., and Faulon, J.-L. (2012) Compound toxicity screening and structure−activity relationship modeling in *Escherichia coli*. *Biotechnol. Bioeng. 109*, 846−850.
(29) Kanehisa, M., Araki, M., Goto, S., Hattori, M., Hirakawa, M., Itoh, M., Katayama, T., Kawashima, S., Okuda, S., Tokimatsu, T., and Yamanishi, Y. (2008) KEGG for linking genomes to life and the environment. *Nucleic Acids Res. 36*, D480−484.
(30) Wishart, D. S., et al. (2013) HMDB 3.0−The Human Metabolome Database in 2013. *Nucleic Acids Res. 41*, D801−D807.
(31) Salgado, H., 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.
(32) Costa, A., Bedini, J., Conget, I., et al. (2002) Uric acid concentration in subjects at risk of type 2 diabetes mellitus: Relationship to components of the metabolic syndrome. *Metabolism 51*, 372−375.
(33) Rintoul, M. R., Cusa, E., Baldomà, L., Badia, J., Reitzer, L., and Aguilar, J. (2002) Regulation of the *Escherichia coli* allantoin regulon: Coordinated function of the repressor AllR and the activator AllS. *J. Mol. Biol. 324*, 599−610.
(34) Teramoto, J., Yamanishi, Y., Magdy, E.-S. H., Hasegawa, A., Kori, A., Nakajima, M., Arai, F., Fukuda, T., and Ishihama, A. (2010) Single live-bacterial cell assay of promoter activity and regulation. *Genes to Cells 15*, 1111−1122.
(35) Walter, A., Korth, U., Hilgert, M., Hartmann, J., Weichel, O., Hilgert, M., Fassbender, K., Schmitt, A., and Klein, J. (2004) Glycerophosphocholine is elevated in cerebrospinal fluid of Alzheimer patients. *Neurobiology of Aging 25*, 1299−1303.
(36) Linnebank, M., Popp, J., Smulders, Y., Smith, D., Semmler, A., Farkas, M., Kulic, L., Cvetanovska, G., Blom, H., Stoffel-Wagner, B., et al. (2010) S-adenosylmethionine is decreased in the cerebrospinal fluid of patients with Alzheimer's disease. *Neurodegenerative Diseases 7*, 373−378.
(37) Rkenes, T., Lamark, T., and Strøm, A. R. (1996) DNA-binding properties of the BetI repressor protein of *Escherichia coli*: The inducer choline stimulates BetI-DNA complex formation. *J. Bacteriol. 178*, 1663−1670.
(38) Kaneko, M., Hwang, E. I., Ohnishi, Y., and Horinouchi, S. (2003) Heterologous production of flavanones in *Escherichia coli*: Potential for combinatorial biosynthesis of flavonoids in bacteria. *J. Ind. Microbiol. Biot. 30*, 456−61.
(39) Meganathan, R. (2001) Ubiquinone biosynthesis in microorganisms. *FEMS Microbiol. LetT. 203*, 131−139.
(40) Egland, P. G., Harwood, C. S., and Harwood, C. S. (2000) HbaR, a 4-Hydroxybenzoate Sensor and FNR-CRP Superfamily Member, Regulates Anaerobic 4-Hydroxybenzoate Degradation by Rhodopseudomonas palustris HbaR, a 4-Hydroxybenzoate Sensor and FNR-CRP Superfamily Member, Regulates Anaerobic 4-Hydroxybenzoate Deg. *J. Bacteriol. 182*, 100−106.

(41) Niazov, T., Baron, R., Katz, E., Lioubashevski, O., and Willner, I. (2006) Concatenated logic gates using four coupled biocatalysts operating in series. *Proc. Nat. Acad. Sci. 103*, 17160−17163.
(42) Terzer, M., and Stelling, J. (2008) Large-scale computation of elementary flux modes with bit pattern trees. *Bioinformatics 24*, 2229−2235.
(43) Hucka, M., et al. (2003) The systems biology markup language (SBML): A medium for representation and exchange of biochemical network models. *Bioinformatics 19*, 524−531.
(44) Galdzicki, M., Rodriguez, C., Chandran, D., Sauro, H. M., and Gennari, J. H. (2011) Standard biological parts knowledgebase. *PloS ONE 6*, e17005.
(45) Zanghellini, J., Ruckerbauer, D. E., Hanscho, M., and Jungreuthmayer, C. (2013) Elementary flux modes in a nutshell: Properties, calculation, and applications. *Biotechnol J.*, 1009−1016.