Biotechnology Journal
www.biotechnology-journal.com

Biotechnol. J. 2014, 9, 1446–1457

DOI 10.1002/biot.201400055

Research Article

Validation of RetroPath, a computer-aided design tool for metabolic pathway engineering

Tamás Fehér[1], Anne-Gaëlle Planson[1]*, Pablo Carbonell[1], Alfred Fernández-Castané[1], Ioana Grigoras[1], Ekaterina Dariy[2], Alain Perret[2] and Jean-Loup Faulon[1]*

[1] Institute of Systems and Synthetic Biology, University of Evry-Val-d'Essonne, CNRS FRE3561, Evry Cedex, France
[2] Genoscope, CEA, INSERM, CNRS UMR8030, Evry Cedex, France

Metabolic engineering has succeeded in biosynthesis of numerous commodity or high value compounds. However, the choice of pathways and enzymes used for production was many times made ad hoc, or required expert knowledge of the specific biochemical reactions. In order to rationalize the process of engineering producer strains, we developed the computer-aided design (CAD) tool RetroPath that explores and enumerates metabolic pathways connecting the endogenous metabolites of a chassis cell to the target compound. To experimentally validate our tool, we constructed 12 top-ranked enzyme combinations producing the flavonoid pinocembrin, four of which displayed significant yields. Namely, our tool queried the enzymes found in metabolic databases based on their annotated and predicted activities. Next, it ranked pathways based on the predicted efficiency of the available enzymes, the toxicity of the intermediate metabolites and the calculated maximum product flux. To implement the top-ranking pathway, our procedure narrowed down a list of nine million possible enzyme combinations to 12, a number easily assembled and tested. One round of metabolic network optimization based on RetroPath output further increased pinocembrin titers 17-fold. In total, 12 out of the 13 enzymes tested in this work displayed a relative performance that was in accordance with its predicted score. These results validate the ranking function of our CAD tool, and open the way to its utilization in the biosynthesis of novel compounds.

Received	29 JAN 2014
Revised	28 JUL 2014
Accepted	15 SEP 2014
Accepted article online	16 SEP 2014

Supporting information available online

Keywords: Computer-aided design · Flavonoid production · Pathway prediction · Rational pathway engineering · Synthetic biology

1 Introduction

The field of metabolic engineering has undergone significant advances through the development of novel strategies combining genetic engineering and systems biology to allow the production of many organic compounds using a cellular host. The diversity of molecules biosynthesized in engineered hosts has recently been extended, the feat of further extension however, remains challenging [1]. The development of novel biosynthetic routes is a multistep process not only implying the identification of biosynthetic enzymes and their assembly to create a metabolic pathway, but also characterizing the interconnection of the pathway with the host metabolism. Consequently, a number of computational tools as well as experimental techniques have been developed [2]. A large number of constraint-based frameworks have been proposed to optimize the production host for the synthesis of

Correspondence: Prof. Jean-Loup Faulon, Institute of Systems and Synthetic Biology, University of Evry-Val-d'Essonne, CNRS FRE3561, Genopole· Campus 1, Genavenir 6, 5 rue Henri Desbruères, F-91030 Evry Cedex, France
E-mail: jean-loup.faulon@issb.genopole.fr

Current address: Dr. Anne-Gaëlle Planson, INRA, AgroParisTech, UMR1319 Micalis, Jouy-en-Josas F-78350, France

Abbreviations: CHS, chalcone synthase; **CHI,** chalcone isomerase; **4CL,** coumaroyl-CoA ligase; **DW,** dry weight; **FBA,** flux balance analysis; **HPLC,** high pressure liquid chromatography; **LC-MS,** or liquid chromatography-mass spectrometry; **OD,** optical density; **PAL,** phenylalanine ammonia lyase

* Both authors contributed equally to this work.

a molecule of interest, or to determine individual performances for the predicted pathways [3]. However, in the majority of the papers published today, enzymes are chosen ad hoc, relying on information available in the literature or on expertise accumulated by the workgroup [4–6]. Several factors, such as enzyme efficiency, byproduct toxicity, and competitive pathways may influence the performance of an engineered strain. To take into account these constraints, we are using here a unified framework to design heterologous biosynthetic pathways. The method we developed, known as RetroPath, uses a retrosynthetic approach in the reaction signature space [7]. Reaction signatures are biochemical reactions coded into the molecular signature representation and are used to search and enumerate similar reactions [8]. Our tool allows one to prioritize the engineering of the most promising routes due to its ranking function [7], which quantifies the enzyme performance and compatibility between the host and the candidate exogenous genes screened from a metabolic database like KEGG [9] or MetaCyc [10], the estimation of steady-state fluxes from the in silico reconstructed model of the engineered strain [11], and the estimation of metabolite toxicity [12]. This approach, hereafter named retrosynthetic biology [12], is used here as a streamlined manufacturing pipeline for the production of compounds with therapeutic interest. RetroPath provides for this purpose a methodology that is general enough to be applied to the production of virtually any compound through metabolic engineering. Our tool does not primarily seek to bypass any of the necessary optimization steps that would at later stages be required for scaling up the engineered strains to industrial production, but rather to provide the design techniques that would univocally identify, assemble and optimize constructs that eventually would lead to the streamlined cell factories.

In the present paper, we illustrate the use of RetroPath for the production of flavonoids, which are compounds naturally produced in plants. Plant-derived natural products hold much promise for the development of new drugs and nutraceuticals [13, 14]. Modification of such compounds for drug development however, is still underexplored due to their low bioavailability and the difficulty of their chemical synthesis. Among these compounds with therapeutic interest are the flavonoids, which are plant secondary metabolites functioning as flower pigments, UV-protectants, insect repellants, or initiators of symbiosis [15]. Transferring the flavonoid synthesis pathway into recombinant microorganisms has been successfully carried out to produce various members of these drug candidates [6, 16–20]. It is worth noting that none of the enzymes used in the aforementioned works were found fully annotated in the metabolic databases, the authors, experts in the field of flavonoid production, most probably selected them based on personal experience [16, 17]. Conversely, the present work aims to assemble biosynthetic pathways for production of a compound of interest in *Escherichia coli* by systematically applying the RetroPath algorithm on an enzyme set obtained from a metabolic knowledge database.

Among flavonoids, pinocembrin, extracted from propolis, was reported to possess numerous biological activities beneficial to health such as neuroprotective, antioxidant, or antibacterial effects [21, 22]. Moreover, the flavanone pinocembrin is a starting point for the synthesis of over 8000 different biologically active molecules through the action of a myriad of enzymes, making it a key compound [23]. In this study, we use RetroPath to provide the most feasible route for microbial synthesis of pinocembrin and validate its productivity by experimentally implementing the top-ranked constructs.

2 Materials and methods

2.1 Materials

E. coli strain DH5α (Life Technologies, Darmstadt, Germany) was used for cloning, enzyme expression, and metabolite production. Plasmids pQlinkN [24] and pRSFDuet were obtained from Addgene and Merck-Novagen (Darmstadt, Germany), respectively. All chemicals were obtained from Sigma (St. Louis, MO, USA), unless specified otherwise. Antibiotics were used at the following concentrations: ampicillin (Ap): 50 μg/mL, chloramphenicol (Cm): 25 μg/mL, kanamycin (Km): 30 μg/mL.

2.2 Construction of pathways for pinocembrin and malonyl-CoA production

Genes encoding the enzymes expressed in this study are listed in Table 1. Total RNA, extracted from fresh *Arabidopsis thaliana* ecotype Columbia leaves [25] was a kind gift of Dr. Bruno Gronenborn (ISV, Gif-sur-Yvette, France). Total RNA was reverse transcribed using RevertAid Premium First Strand cDNA Synthesis kit (Thermo Fisher Scientific Biosciences GmbH-Fermentas; Vilnius, Lithuania). Genomic DNA of *Streptomyces coelicolor* and *Bacillus subtilis* 168, kind gifts of Dr. Sylvie Lautru (IGM, Orsay, France) and Dr. Hamid Nouri (iSSB, Evry, France), respectively, were prepared using the GeneJet Genomic DNA Purification Kit (Thermo Fisher Scientific Biosciences GmbH – Fermentas; Vilnius, Lithuania). For the pinocembrin-producing pathway, PCR was used to amplify *hPAL*, *Pal2*, and *hCHS* genes from the cDNA library of *A. thaliana*, as well as *h4CL* and *lCHS* from *S. coelicolor* and *B. subtilis* genomic DNA, respectively. The restriction enzymes used for gene cloning into pQLinkN can be inferred from the primer list (Supporting information, Table S1). *l4CLopt*, *m4CLopt*, *hCHIopt*, and *lCHIopt* were codon-optimized for expression in *E. coli*, synthesized by GenScript (Piscataway, NJ, USA) and cloned into

Table 1. Top-scoring genes used for pathway No. 1.1 construction

Gene tag	Encoded enzymatic activity	Source	Accession number	RetroPath score
hPAL	Phe ammonia-lyase	*Arabidopsis thaliana*	TAIR: AT2G37040	2.39
h4CL	4-Coumarate:CoA ligase	*Streptomyces coelicolor*	EMBL: CAB95894.1	1.39
l4CLopt	4-Coumarate:CoA ligase	*Streptomyces maritimus*	UniProt: Q9KHL1	1.07
m4CLopt	4-Coumarate:CoA ligase	*Arabidopsis thaliana*	TAIR: AT1G65060.1	1.08
hCHS	Chalcone synthase	*Arabidopsis thaliana*	TAIR: AT5G13930	0.96
lCHS	Chalcone synthase	*Bacillus subtilis*	SubtiList: BSU22050	0.75
hCHIopt	chalcone isomerase	*Arabidopsis thaliana*	TAIR: AT3G55120	0.69
lCHIopt	Chalcone isomerase	*Arabidopsis thaliana*	TAIR: AT5G66220	–0.07

pQLinkN using *Bam*HI and *Hin*dIII restriction/ligation. The genes were then combined into pQLinkN to make complete pinocembrin-producing pathways using ligation-independent cloning as described earlier [24]. For the pathways producing malonyl-CoA, genes *mmsA*, and *cagg1256*, as well as the *atoDA* operon were codon-optimized and synthesized by GenScript, each with *Nde*I and *Xho*I sites flanking 5′ and 3′, respectively. Plasmid pACYC*matCmatB* [26] was a kind gift of Dr. Jingwen Zhou (LBBE, Jiangnan, China), and pETM6-M.*accABCD*, carrying the genes of the *E. coli* acetyl-CoA carboxylase complex in a monocystronic form [27] was a kind present from Prof. Mattheos Koffas (Rensselaer Polytechnic Institute, Troy, NY, USA). pRSF*matCmatB* was constructed by cloning the 3032 bp *Eco*RI–*Xho*I fragment of pACYC*matCmatB* into pRSFDuet. pRSFM.*accABCD* was constructed by ligating the 4824 bp *Sal*I–*Avr*II fragment of pETM6-M.*accABCD* into *Sal*I–*Avr*II-digested pRSFDuet. pRSF*mmsA* and pRSF*cagg1256* were made by cloning the respective *Nde*I–*Xho*I-restricted synthetic *mmsA* or *cagg1256* gene into the similarly digested pRSFDuet. pRSF*matCatoDA* was constructed by ligating the *Nde*I–*Xho*I-digested synthetic *atoDA* operon into the *Nde*I–*Xho*I cleaved pRSF*matCmatB*. Nucleotide sequences of the plasmids harboring the complete pathways were verified by sequencing. All DNA manipulation and culturing was carried out according to standard protocols [28].

2.3 Conditions of culture and protein overexpression

Bacterial cultures fully grown overnight in LB medium were diluted to an optical density (OD_{600}) of 0.1 in terrific broth (TB) [28] supplemented with 0.5 M sorbitol, 5 mM betaine [29], and the adequate antibiotics, and were shaken at 37°C, 250 rpm. Protein expression was induced after 3 h with 1 mM IPTG together with the addition of 3 mM phenylalanine, 2 mg/mL sodium malonate and when indicated, 20 μg/mL cerulenin. For strains harboring pRSF*mmsA* and pRSF*cagg1256*, malonate was replaced with 3 mM of β-alanine. Cultures were sampled 24 h after induction.

2.4 Extraction of metabolites

Metabolites were extracted using a protocol adapted from Wu et al. [26]. Briefly, biomass was separated from the culture medium by centrifugation at 13 000*g* and flavonoids were extracted from 1 mL of supernatant with an equivalent volume of pre-chilled ethyl acetate. Samples were vortexed and subsequently centrifuged for 15 min at 4°C, 13 000*g*. The upper organic layer was then vacuum centrifuged for 2 h at mid temperature. Dried pellets were stored at –80°C until analysis by high pressure liquid chromatography (HPLC) or liquid chromatography-mass spectrometry (LC-MS).

2.5 HPLC analysis

HPLC analysis was carried out using a Shimadzu Prominence LC20/SIL-20AC equipped with a Kinetex XB-C18 reversed phase column (250 mm × 4.5 mm, 5 μm) and a UV–Vis detector. Mobile phase was composed of 0.1% formic acid in water (A) and 0.1% formic acid in acetonitrile (B). A linear gradient elution using a binary pump was performed as follows: 0–10 min, 10% B to 90% B; 10–20 min, 90% B to 10% B; 20–30 min, 10% B to 10% B; 30 min, stop. Samples were thawed and resuspended in 200 μL of 80% acetonitrile. The flow rate was 500 μL/min, the sample injection volume was 10 μL, and the column was thermostated at 40°C. Pinocembrin and *trans*-cinnamate were monitored at 290 nm. Quantification of metabolites was done by interpolation of the integrated peak areas using a calibration curve prepared with standard samples.

2.6 LC–MS analysis

Formal identification of flavonoids was done using LC–MS. Samples were thawed and resuspended in 80% acetonitrile and 20% of 10 mM ammonium carbonate. Chromatographic separation was performed using a normal phase method on a SeQuant ZIC pHILIC HPLC column (150 mm × 4.6 mm, 5 μm polymeric beads PEEK) from Merck (Darmstadt, Germany). The flow rate was 500 μL/min, the sample injection volume was 10 μL, and the column was thermostated at 40°C. The mobile phase

A was 10 mM of ammonium carbonate in water, and the mobile phase B was acetonitrile. The starting conditions were 20% of A and 80% of B. The following gradient profile was used: 0–2 min 20% A, 2–16 min 20–60% A, 16–24 min 60% A. The mobile phase was allowed to return to the starting conditions within 6 min, and the column was re-equilibrated for 15 min. The LTQ Orbitrap mass spectrometer was equipped with an ESI source operated in the negative ion mode. The ionization conditions were optimized for the detection of the compound of interest. The spray voltage was set to – 3.5 kV, the capillary voltage to – 47 V, tube lens offset to – 120 V, the sheath gas and auxiliary gas flow rates were 10 and 18 a.u., respectively.

2.7 RetroPath pathway design

Metabolic pathways were designed using the RetroPath tool, a design pipeline based on retrosynthesis [7, 8]. This software searches and enumerates all possible pathways producing the desired target compound in a host organism [30]. In the present study, the target compound is pinocembrin and the host cell is *E. coli*. Each pathway provided by RetroPath corresponds to a unique list of enzymatic reactions that potentially can lead to the production of the target compound starting from precursors available in the host organism or from added supplements in the medium. Pathways were ranked by RetroPath by defining a score function based on three criteria: enzyme performance, toxicity of intermediates, and pathway maximum yield [7]. The contribution of each term to the total score was computed as follows: (i) For each enumerated pathway P, the system provided a score $w(r)$ for the enzymes that can putatively catalyze each of the reaction steps r based on the prediction of enzyme performance through the tensor product technique [31, 32]. Such score was previously shown to parallel the ability of enzymes to catalyze multiple reactions, i.e. the prediction of enzyme's promiscuity [32], and it is therefore used here in order to evaluate the likeliness of an enzyme to catalyze a given reaction. (ii) Toxicity score for each intermediate metabolite c produced by the reactions r in the pathway P was evaluated by using the EcoliTox server [33], with the associated score given by the sum of predicted $\log(IC_{50})$. (iii) Flux balance analysis (FBA) was performed in order to compute the growth-target coupled flux, i.e. the product of maximum growth μ_{max} and maximum target compound flux v_{max} for each pathway P [8]. The associated score $S(P)$ of a given pathway P was defined by the weighted sum of the three previous terms:

$$S(P) = \lambda_{path} \sum_{r \in P} w(r) + \lambda_{tox} \sum_{r \in P} \sum_{c \in r} \log(IC_{50}(c)) + \lambda_{flux} \mu_{max} v_{max} \quad (1)$$

where the weighting coefficients were set to (λ_{path}, λ_{tox}, λ_{flux}) = (1.0, 0.4, 2.5) as in [8]. The way the weights were chosen was through an optimization algorithm that was previously described in [7].

2.8 Metabolic flux analysis

The detailed setup of metabolic flux distribution simulation is described in Section 3 of the Supporting information. Briefly, we modeled the effect of pathway insertion into metabolic fluxes by importing predicted RetroPath pathways into an in silico reconstructed model for *E. coli* [34] that was obtained from the BioModels database [35], which contains 2381 reactions and 1668 metabolites. More precisely, the pinocembrin and malonyl-CoA producing heterologous pathways were added to the model, as well as transport and exchange reactions for *trans*-cinnamate, pinocembrin, malonate, β-alanine, and phenylalanine. FBA simulations were performed using the COBRApy package [36], where measured fluxes for key intermediates and final products as well as growth were fixed to their experimental values, as additional system constraints. The obtained fitted solutions were used in order to determine phenylalanine, acetyl-CoA, and malonate or β-alanine consumption, and the drains of *trans*-cinnamate and malonyl-CoA precursors going into biomass formation.

Growth was estimated from OD_{600} measurements based on the approximate conversion between biomass and OD_{600}: biomass [gDW/L] ≅ 0.44 × OD (where DW stands for dry weight) [37]. We estimated growth rate (μ) as the difference between the measured OD and the OD_0 at the time of induction:

$$\mu = \frac{\ln(2)}{t_d} = \frac{\ln(OD / OD_0)}{\Delta t} \left[hr^{-1} \right] \quad (2)$$

where Δt is the time elapsed after induction and t_d is the doubling time.

Flux units were expressed in mmol/gDW h. Fluxes were estimated from measured concentrations [c] as follows:

$$v_c = \mu \left[hr^{-1} \right] \times \frac{[c] - [c]_0 \left[mg / l \right]}{M(c) \left[g / mol \right] \times \left[OD - OD_0 \right] \times 0.44 \left[gDW / l \right]} \left[mmol \cdot gDW^{-1} \cdot hr^{-1} \right] \quad (3)$$

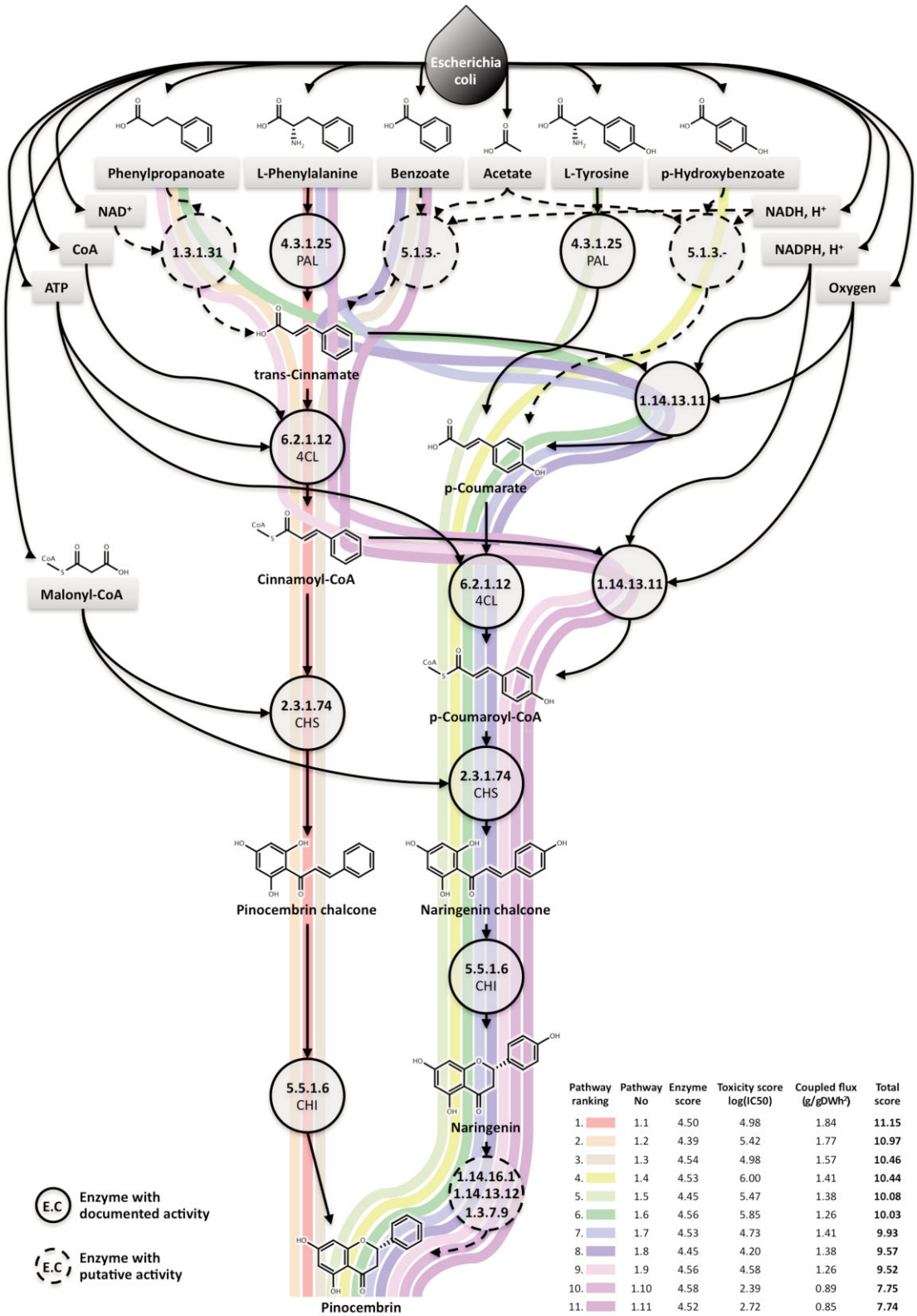
where $M(c)$ is the molar mass of the chemical compound, and $[c]_0$ is the initial concentration.

The inhibition effect on growth from *trans*-cinnamate was estimated using the EcoliTox server [33].

3 Results and discussion

3.1 Predicting pathways for pinocembrin production in *E. coli*

Here, we use the term pathway to designate a series of enzymatic steps (defined by their EC numbers) connect-



Pathway ranking	Pathway No	Enzyme score	Toxicity score log(IC50)	Coupled flux (g/gDWh²)	Total score
1.	1.1	4.50	4.98	1.84	**11.15**
2.	1.2	4.39	5.42	1.77	**10.97**
3.	1.3	4.54	4.98	1.57	**10.46**
4.	1.4	4.53	6.00	1.41	**10.44**
5.	1.5	4.45	5.47	1.38	**10.08**
6.	1.6	4.56	5.85	1.26	**10.03**
7.	1.7	4.53	4.73	1.41	**9.93**
8.	1.8	4.45	4.20	1.38	**9.57**
9.	1.9	4.56	4.58	1.26	**9.52**
10.	1.10	4.58	2.39	0.89	**7.75**
11.	1.11	4.52	2.72	0.85	**7.74**

Figure 1. Pathways enumerated and ranked by RetroPath leading to production of pinocembrin in *E. coli*. Solid circles correspond to enzymes with documented activity, while dotted circles correspond to enzymes with putative activity predicted by RetroPath. Endogenous *E. coli* metabolites are displayed in gray boxes. The inset table shows the ranking of pinocembrin pathways based on a score function consisting of three terms: enzyme efficiency, metabolite toxicity, and calculated growth-target coupled flux.

ing designated source metabolites to target compounds. Since usually more than one enzyme is available for each step, a pathway can be assembled using multiple combinations of enzymes, referred to as constructs. Under such definitions, we applied RetroPath to design and optimize the pathway and the associated constructs leading to the heterologous production of the flavonoid pinocembrin in *E. coli*. In total, RetroPath found and ranked 11 heterologous pathways connecting the endogenous metabolites of *E. coli* to pinocembrin. Figure 1 depicts the map of these pathways, together with their connection to the metabolome of the chassis cell. The contribution of each score term (enzyme efficiency, metabolite toxicity, and product flux) to the total pathway score is provided in the inset of Fig. 1. The top ranked pathway (No. 1.1) uses phenylalanine as main precursor, which is converted into *trans*-cinnamate by phenylalanine/tyrosine ammonia lyase (PAL/TAL). The *trans*-cinnamate is subsequently transformed into cinnamoyl-CoA by 4-coumaroyl-CoA ligase (4CL), and then into pinocembrin chalcone by a chalcone synthase (CHS) and further into pinocembrin by chalcone isomerase (CHI). This is the natural pathway of flavonoid synthesis, and has already been demonstrated to function appropriately in recombinant *E. coli* cells [5, 19]. However, within the natural pathway, there is still an impressive number of possible enzymes combinations (i.e. constructs) depending on which species the enzyme originated from. The choice of enzymes proposed by RetroPath is a critical feature that reduces the number of constructs one needs to implement to only those that have the best-predicted performances.

For the top ranked pathway (No. 1.1), RetroPath compiled the list of genes available for each enzymatic step. To test all gene-combinations supporting this pathway (Supporting information, Table S2A), one would need to assemble more than 8.8 million constructs (110 × 112 × 45 × 16). We narrowed this list by using RetroPath (see Section 2.7), which ranked the genes for each reaction according to the expected performance of the encoded enzymes. The score of the enzymes taken into consideration for each step is shown in Supporting information, Table S2B. We privileged using the wild type alleles recovered from genomic DNA or cDNA libraries. If any of these failed to give a band on a Coomassie-stained polyacrylamide gel upon induction, it was codon-optimized and commercially synthesized for expression in *E. coli*, and was marked by an "opt" suffix in its tag. Our initial strategy was testing at least two alternative enzymes for each step. For PAL however, we were in the advantageous situation that its product (*trans*-cinnamate) is directly measurable by HPLC. The two highest-scoring candidates *A. thaliana*'s *Pal1* and *Pal2* produced 87.98 and 0.22 mg/L *trans*-cinnamate, respectively, when expressed individually. Therefore, we decided to include only the prior in the final constructs, since it is the better *trans*-cinnamate producer. In addition, plant-derived 4CL enzymes have been described to have low cinnamoyl-CoA ligase activity [38]. We therefore included h4CL as a positive control, since it had already proven to be effective in this pathway [18]. h4CL is not present in KEGG, our initial enzyme source, but exhibited a significantly high score when included in the list of queried enzymes. Altogether, genes for one enzyme displaying PAL activity, three displaying 4CL, two displaying CHS and two exhibiting CHI activities were cloned to make a full pathway (Table 1). To simplify their nomenclature, we differentiate them with the letters "h," "m," or "l" referring to their scores being "high," "medium," or "low." These enzymes can be assembled in 12 combinations, depicted in Supporting information, Table S3.

3.2 Assembly and analysis of pinocembrin-producing pathways

Based on the provided scores of the individually cloned genes, we set out to implement the highest-ranked pathway (No. 1.1) in the *E. coli* chassis. Genes were assembled into 12 possible constructs, i.e. into 12 possible combinations of enzymes corresponding to steps in the highest-ranked pathway (Supporting information, Table S3) using the rapid method of ligation-independent cloning [24].

The performance of each construct was assessed experimentally by measuring the concentration of pinocembrin in the culture medium 24 h after induction, as described in Section 2. For the top-ranked constructs HHHH and HHHL, we confirmed the effective production of pinocembrin at low titers using LC–MS analysis (Supporting information, Fig. S1 and Table S4). However, we observed that *trans*-cinnamate, the first intermediate in the pathway, was produced by both strains in significant amounts (38.2 and 53.6 mg/L for HHHH and HHHL, respectively) compared to the negative control (strain with induced empty plasmid, DH5α + pQlinkN in Supporting information, Table S4). Therefore, the *trans*-cinnamate produced from phenylalanine by overexpression of *hPAL*, was not efficiently being consumed to produce pinocembrin and was accumulated in the medium.

3.3 Network-level optimization of pinocembrin production

FBA provided a possible explanation for the low production of pinocembrin obtained in the initial tests, and guided us toward pathway optimization using RetroPath. The analysis was performed on an *E. coli* in silico model constrained with experimental measurements for *trans*-cinnamate, pinocembrin, and biomass (see Section 3 of Supporting information for details). Supporting information Fig. S2 shows the main fluxes involved in the pathway. The model predicted that under the observed conditions of low pinocembrin yield, most of *trans*-cinnamate was consumed for growth, while the rest was accumulat-

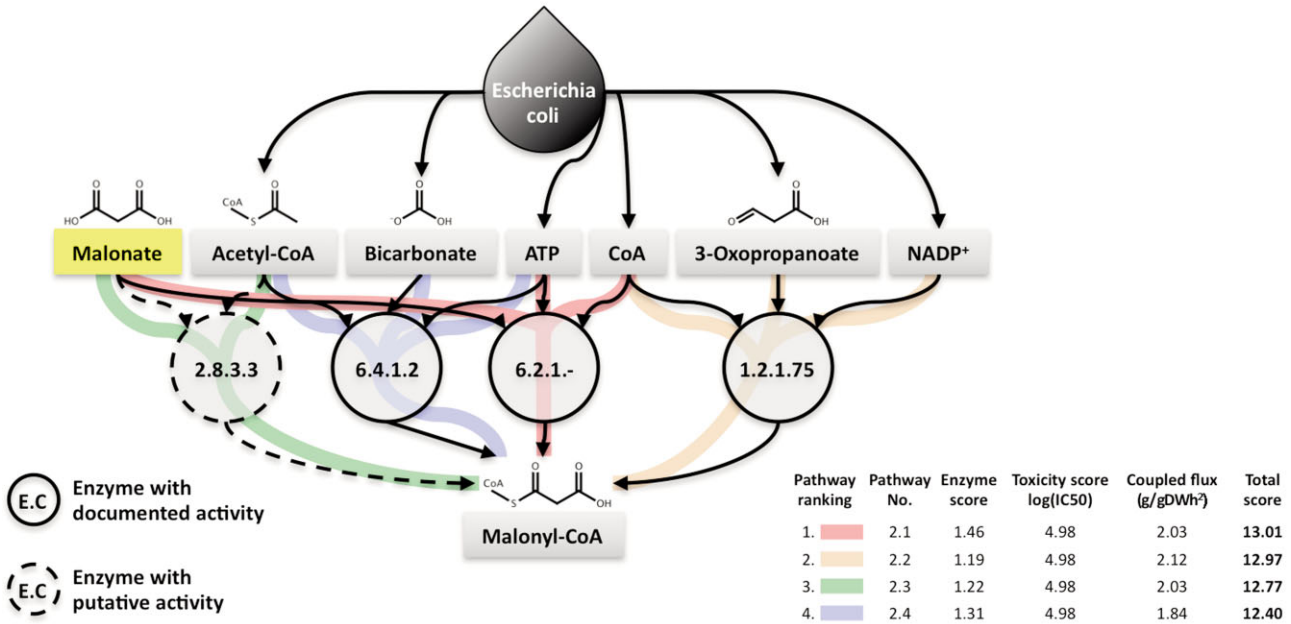


Figure 2. Pathways enumerated by RetroPath for the production of malonyl-CoA. Four 1-step pathways are possible: EC 6.2.1.- and EC 2.8.3.3, which consume malonate added as a supplement, as well as EC 6.4.1.2 and EC 1.2.1.75. Solid circles correspond to enzymes with documented activity, while dotted circles correspond to enzymes with putative activity, predicted by RetroPath. Endogenous and heterologous *E. coli* metabolites are displayed in gray and yellow boxes, respectively.

ed in the medium because the excess of *trans*-cinnamate was not compensated by the availability of malonyl-CoA, mostly consumed in the competitive pathways of fatty acids synthesis. Therefore, in order to boost the production of pinocembrin, the available pool of malonyl-CoA needed to be increased. We thus returned to the design task of our metabolic engineering pipeline and used RetroPath to enumerate all pathways enabling the production of malonyl-CoA. RetroPath ranked four heterologous pathways producing malonyl-CoA (Fig. 2), which are as follows:

2.1. Malonyl-CoA production from malonate through malonyl-CoA synthase (EC 6.2.1.) using CoA and ATP as cofactors (pathway score 13.01). This reaction is not annotated for *E. coli*, but the product of the *matB* gene [39], originating from *Rhizobium trifolii* has been applied successfully for this purpose [4, 26].

2.2. Malonyl-CoA reductase (EC 1.2.1.75, pathway score 12.97). It requires 3-oxopropanoate, CoA and $NADP^+$ as substrates. This reaction is not annotated for *E. coli*.

2.3. Malonate-CoA transferase (EC 2.8.3.3). This enzyme can also produce malonyl-CoA from malonate but requires acetyl-CoA as a co-substrate (pathway score 12.77). This reaction is not annotated for *E. coli*, either.

2.4. Acetyl-CoA carboxylase (EC 6.4.1.2, pathway score 12.40), producing malonyl-CoA from acetyl-CoA and bicarbonate. In *E. coli*, a four-enzyme complex comprising carboxyltransferase α, biotin-carboxyl-carrier-protein, biotin carboxylase, and carboxyltransferase β is responsible for this step. The endogenous complex [40], a similar complex of *Photorabdus luminescens* [16], as well as a two-enzyme complex from *Corynebacterium glutamicum* [5] have all been successfully expressed in *E. coli*.

These pathways, integrated with pathway 1.1, are summarized in Table 2. To verify the pathway-predicting function of RetroPath, we implemented all four malonyl-CoA producer pathways using the top-scoring genes available for each (Supporting information, Table S5). For reaction 1.2.1.75, besides testing *cagg1256*, we also implemented *mmsA*, which was discovered by running the query after a recent update of the KEGG-database. The transmembrane dicarboxylate carrier protein, encoded by *matC* was co-expressed with *matB* and *atoDA* to grant sufficient uptake of malonate by the cells. In order to provide the substrate for *mmsA* and *cagg1256*, cultures were supplemented with β-alanine, which is converted to 3-oxopropanoate (malonate semialdehyde) by the cell's endogenous 4-aminobutyrate aminotransferase. The absolute pinocembrin titers achieved with HHHH together with these supplementary pathways are shown on Supporting information, Fig. S3. It is apparent that pathway 2.1, which has the highest score yielded the greatest titer (4.22 mg/L).

In order to divert further the malonyl-CoA excess into the flavonoid pathway, cerulenin was added into the medium to inhibit fatty acid formation [4, 41]. The inhibitory effect of cerulenin on the production of fatty acids was estimated again by flux analysis (Supporting information, Fig. S2) leading to a lower consumption of

Table 2. Ranking of the pinocembrin-producing pathways derived from pathway 1.1 (Table 1) with an additional enzyme to increase production of the malonyl-CoA precursor

Rank	No.	Pathway	Enzyme score	Toxicity score log(IC_{50})	Coupled flux (g/gDW h^2)	Total score
1	2.1	Malonate + CoA + ATP → malonyl-CoA L-phenylalanine → *trans*-cinnamate → cinnamoyl-CoA → pinocembrin chalcone → pinocembrin	1.46	4.98	2.03	13.01
2	2.2	3-Oxopropanoate + CoA + NADP+ → malonyl-CoA L-phenylalanine → *trans*-cinnamate → cinnamoyl-CoA → pinocembrin chalcone → pinocembrin	1.19	4.98	2.12	12.97
3	2.3	Malonate + acetyl-CoA → malonyl-CoA L-phenylalanine → *trans*-cinnamate → cinnamoyl-CoA → pinocembrin chalcone → pinocembrin	1.22	4.98	2.03	12.77
4	2.4	CCCP[a)] + acetyl-CoA → malonyl-CoA L-phenylalanine → *trans*-cinnamate → cinnamoyl-CoA → pinocembrin chalcone → pinocembrin	1.31	4.98	1.84	12.40

a) CCCP, caroxybiotin-carboxyl-carrier protein

malonyl-CoA for fatty acids biosynthesis, which in turn led to an increase in the flux for production of pinocembrin and a larger accumulation of *trans*-cinnamate. By combining the best performing malonyl-CoA producer (*matCmatB* cassette) with the cerulenin inhibition, measurements were carried out for all 12 pinocembrin production constructs, obtaining in this case higher pinocembrin titers (Supporting information, Table S7), which were routinely detectable thereafter using HPLC (Supporting information, Fig. S4). Estimated flux values are provided in Supporting information, Table S6.

3.4 Validation of the predicted enzyme ranking and evaluation of results

The accuracy of enzyme-score predictions was verified using the concentrations of the intermediate metabolite *trans*-cinnamate, and the target compound pinocembrin. In the case of using *trans*-cinnamate levels for validation, the PAL score is taken as it is, since it is a contributor, while the 4CL score is subtracted, since it is a consumer of *trans*-cinnamate. In the two cases where pinocembrin was produced, we also subtracted the scores for CHS and CHI as these enzymes also decrease the levels of *trans*-cinnamate. We obtain a good agreement between the RetroPath scores combined this way and the experimentally measured *trans*-cinnamate levels (Fig. 3A). Specific *trans*-cinnamate levels for each construct are given in Supporting information, Table S7. The constructs containing the *l4CLopt* gene from *S. maritimus* (H**L**HH, H**L**HL, H**L**LH, and H**L**LL), which had a lower score correspond to a higher accumulation of *trans*-cinnamate, which might indicate that less *trans*-cinnamate was transformed. The constructs that contained the 4CL gene with the highest score, *h4CL* from *S. coelicolor* (H**H**HH, H**H**HL, H**H**LH, and H**H**LL) led to less *trans*-cinnamate accumulation indicating its more efficient transformation, although they showed more variability.

Among the genes chosen for the pinocembrin producing constructs, only two, *l4CLopt* and *lCHS* were found to lack activity, narrowing the number of successful constructs to four (HHHH, HHHL, HMHH, and HMHL). For proper comparison of these constructs, we normalized the pinocembrin production titers with the biomass and the fermentation time (Fig. 3B) to obtain a parameter we refer to as efficiency. Our analysis revealed HHHH to be the most efficient producer, supporting the construct-scores predicted by RetroPath. In addition, we found that for each enzymatic step, inserting an enzyme with a higher score led to a construct with a more efficient performance (Fig. 3C). It is apparent from Fig. 3C that the impact of 4CL score on construct efficiency is higher than that of CHS or CHI. Additionally, the scores of the CHI enzymes did not provide further improvement in predicting pinocembrin production, since the choice of the gene for this last step did not significantly alter construct efficiency (Supporting information, Table S8). This is due most likely to the fact that only 16 enzyme sequences were available in databases in order to build the training set (accuracy and specificity <70%). This issue of low performance for intramolecular lyase activity (EC 5.5) prediction has been previously reported by several groups using different learning techniques [31, 42, 43].

From the industrial point of view, the maximal production titer may in certain cases be of greater importance than the efficiency normalized by the biomass. In this aspect, the HHHL, and not the HHHH construct, producing extracellular pinocembrin levels of 24.14 mg/L

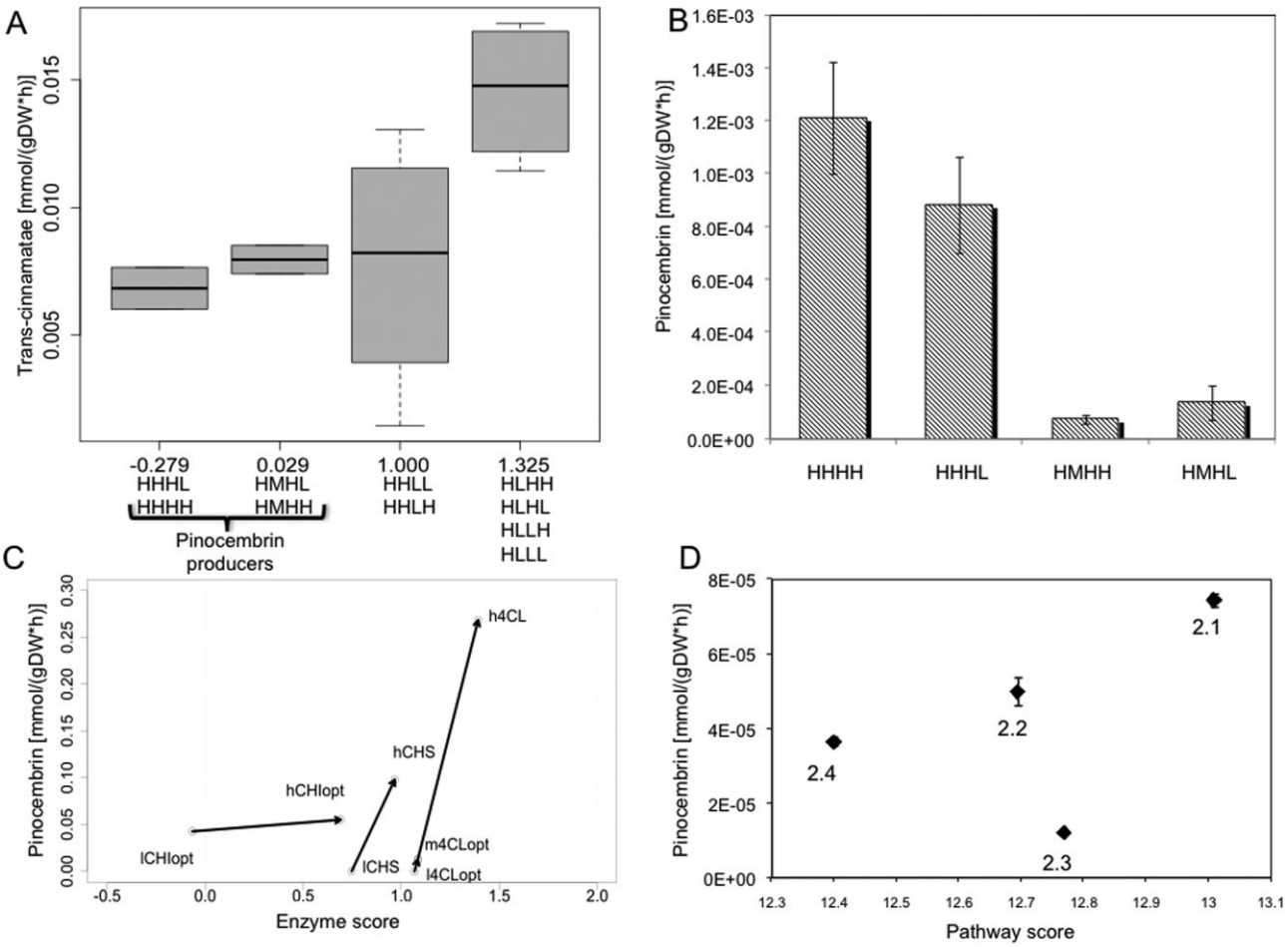


Figure 3. (**A**) The accumulation of *trans*-cinnamate in the medium versus the calculated RetroPath scores. A higher score means higher predicted *trans*-cinnamate levels, as described in the text. Cerulenin was administered in all cases. (**B**) The efficiency of pinocembrin production for the four strains carrying the successful constructs. The *matCmatB* cassette was present in each strain, and cerulenin was administered in all cases. (**C**) Effect of individual enzyme scores on construct efficiencies. Variation of average efficiencies and average enzyme scores associated with enzyme change at various positions of the pathway. The letter preceding the gene names is "l" for low, "m" for medium, and "h" for high predicted activities, as described in Section 3.1. The *matCmatB* cassette was present in each strain, and cerulenin was administered in all cases. (**D**) Validation of the pathway ranking function. Measured pinocembrin production efficiencies are shown versus the predicted scores of pathways listed in Table 2. In each case, the HHHH construct was present in the cell, with the appropriate pathway for malonyl-CoA production present on a second, pRSFDuet-based plasmid. The scores, as well as the efficiencies for *mmsA* and *cagg1256* were averaged, since both correspond to pathway 2.2. No cerulenin was added in this experiment. Pathway numbering corresponds to that of Table 2.

(±5.47 mg/L) proved to be the best (Supporting information, Table S8), and was comparable to other metabolic engineering projects targeting flavanones (Supporting information, Table S9). The reason for this discrepancy among our top-producers is the fact that a uniform protocol of culturing and induction was used for all of our strains. The strains carrying the HHHH construct, due to their slower growth, would most probably need longer fermentation times to generate a proper biomass prior to induction to surpass HHHL in absolute pinocembrin titers. Such an optimization is not included in RetroPath in its current version, nor is the optimization of expression levels of the inserted heterologous enzymes.

In the current work, RetroPath was nevertheless used in a round of network-level optimization to select the most efficient pathway to boost malonyl-CoA production, and eliminate the bottleneck of the CHS step. To that end, we note that in our work, the flux balance model was only applied to identify this bottleneck, and was not used for the validation of RetroPath at any level. Implementing the four malonyl-CoA producer pathways, and comparing their effect on pinocembrin production efficiency provided a good validation of the pathway-ranking function (Fig. 3D). We emphasize that the four routes of malonyl-CoA synthesis use different substrates and employ enzymes designated by different EC numbers, supporting

the fact that they are alternative metabolic pathways, and not alternative enzymes catalyzing the same reaction. Our results highlight the ability of the scoring process to rank pathways even in cases where predicted scores are relatively close in terms of magnitude (score differences of approximately 2%). Based on the validation results from Fig. 3C and D, we would recommend considering pathway scores as significantly different if they differ in at least 2% of the averaged score per gene, which is computed by dividing the total score by the number of enzymatic steps. Besides the high sensitivity of the predictions displayed above, we also found the scoring to be quite robust: the coefficients for the contributions of enzyme score (1) toxicity score (0.4) and flux score (2.5) can be changed ±20% without changing the rank order seen in Figs. 1 and 2 (data not shown). During the validation of pathway ranking, only pathway 2.8.3.3 turned out to be an outlier, with *atoDA* performing much weaker than expected. Since this was the only putative pathway among the four alternatives, it may indicate the need of further optimization of the pathway scoring function by introducing a penalty for putative steps. Malonyl-CoA, the compound targeted in the optimization process, was chosen based on its position on the metabolic map. In the future, this choice could also be automated by incorporating existing methods such as OptKnock [44] or Redirector [45] into our pipeline.

4 Concluding remarks

Our results underline several factors that future users of RetroPath must be aware of when planning to experimentally investigate their predictions. First, due to the nature of automated pipelines, false positive hits or score differences lying below the threshold of significance for proper ranking are probably inevitable in some cases. Second, a certain percentage of the predicted candidates are "lost" due to problematic cloning or expression. The continuously decreasing price of gene synthesis will aid the evasion of the latter caveat, multiple candidates for each enzymatic step should nevertheless be considered. Third, several enzymes proven to be effective in this pathway were missing or not annotated in metabolic knowledge bases. Besides emphasizing the need to update these databases, this information highlights the necessity of integrating the data from multiple resources. And fourth, classical metabolic engineering experiments, i.e. optimizing gene expression, culture conditions, and extraction protocols will still be required to fully exploit the capacities of the top-ranked pathways.

Despite these difficulties, the goal of assembling a pathway to produce a target compound using retrosynthetic biology was nevertheless successfully reached. RetroPath predicted and ranked 11 pathways linking endogenous *E. coli* metabolites to pinocembrin. To validate the enzyme-ranking function of RetroPath, we implemented the top-ranking pathway using various enzymes. We found a good correlation between enzyme scores and their contribution to construct performance (Fig. 3C), with the top-ranking construct being the best producer of pinocembrin. These results demonstrate the power of the enzyme ranking function, since it was necessary to test only a small number of constructs to find target-producing hits despite the initial pool of nine million enzyme combinations. The pathway ranking function was also validated by implementing four pathways for malonyl-CoA production, three of which performed according to our predictions. In total, 12 out of the 13 enzymes (2 PAL, 3 4CL, 2 CHS, 2CHI, and 4 for malonyl-CoA synthesis) tested in this work displayed a relative performance that was in accordance with its predicted score (Fig. 3A, 3C, and 3D). Notably, choosing the highest-ranking pathway and the top-scoring enzyme available for each step would have led to the best performing strain in the first place, which in our opinion is a strong indicator of the value of our predictions for the metabolic engineering community.

The support that our tool provided in pathway and construct ranking significantly shortened the design phase and alleviated the need for either expert knowledge concerning the engineered pathway or numerous trial-and-error experiments. We believe therefore that its adoption would substantially accelerate projects targeting the biosynthesis of other compounds beyond our tested flavonoid pathways. Furthermore, a remarkable feature of our method is that it considers and screens for promiscuous enzymatic activities when predicting novel pathways, a capability that notably allows its use to engineer pathways that did not exist before in nature; opening in that way the possibility of producing non-natural compounds as a result [46, 47]. In addition, RetroPath could be used to select the chassis cell itself for more effective production of natural or non-natural compounds. To that end, our forthcoming aim is to validate and explore further the advanced capabilities of RetroPath, building upon the promising results showcased in the present work, in order to extend its applicability into an ever-wider range of metabolic engineering and synthetic biology projects.

We thank Elodie Paillard and Fred Green (iSSB, Evry, France) for technical assistance, as well as Cyrille Pauthenier (iSSB, Evry, France) for helpful discussions. This project was financed by the ATIGE and the ANR Chair of Excellence grants.

The authors declare no financial or commercial conflict of interest.

5 References

[1] Curran, K. A., Alper, H. S., Expanding the chemical palate of cells by combining systems biology and metabolic engineering. *Metab. Eng.* 2012, *14*, 289–297.
[2] Shin, J. H., Kim, H. U., Kim, D. I., Lee, S. Y., Production of bulk chemicals via novel metabolic pathways in microorganisms. *Biotechnol. Adv.* 2012, *31*, 925–935.
[3] Copeland, W. B., Bartley, B. A., Chandran, D., Galdzicki, M. et al., Computational tools for metabolic engineering. *Metab. Eng.* 2012, *14*, 270–280.
[4] Leonard, E., Yan, Y., Fowler, Z. L., Li, Z. et al., Strain improvement of recombinant *Escherichia coli* for efficient production of plant flavonoids. *Mol. Pharm.* 2008, *5*, 257–265.
[5] Miyahisa, I., Kaneko, M., Funa, N., Kawasaki, H. et al., Efficient production of (2S)-flavanones by *Escherichia coli* containing an artificial biosynthetic gene cluster. *Appl. Microbiol. Biotechnol.* 2005, *68*, 498–504.
[6] Santos, C. N., Koffas, M., Stephanopoulos, G., Optimization of a heterologous pathway for the production of flavonoids from glucose. *Metab. Eng.* 2011, *13*, 392–400.
[7] Carbonell, P., Planson, A. G., Fichera, D., Faulon, J. L., A retrosynthetic biology approach to metabolic pathway design for therapeutic production. *BMC Syst. Biol.* 2011, *5*, 122.
[8] Carbonell, P., Planson, A. G., Faulon, J. L., Retrosynthetic design of heterologous pathways. *Methods Mol. Biol.* 2013, *985*, 149–173.
[9] Kanehisa, M., Goto, S., Sato, Y., Furumichi, M., Tanabe, M., KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res.* 2012, *40*, D109–D114.
[10] Caspi, R., Altman, T., Dreher, K., Fulcher, C. A. et al., The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Res.* 2012, *40*, D742–D753.
[11] Curran, K. A., Crook, N. C., Alper, H. S., Using flux balance analysis to guide microbial metabolic engineering. *Methods Mol. Biol.* 2012, *834*, 197–216.
[12] Planson, A. G., Carbonell, P., Grigoras, I., Faulon, J. L., A retrosynthetic biology approach to therapeutics: From conception to delivery. *Curr. Opin. Biotechnol.* 2012, *23*, 948–956.
[13] Chin, Y. W., Balunas, M. J., Chai, H. B., Kinghorn, A. D., Drug discovery from natural sources. *AAPS J.* 2006, *8*, E239–E253.
[14] Cragg, G. M., Newman, D. J., Natural products: A continuing source of novel drug leads. *Biochim. Biophys. Acta* 2013, *1830*, 3670–3695.
[15] Dixon, R. A., Lamb, C. J., Masoud, S., Sewalt, V. J., Paiva, N. L., Metabolic engineering: Prospects for crop improvement through the genetic manipulation of phenylpropanoid biosynthesis and defense responses – a review. *Gene* 1996, *179*, 61–71.
[16] Leonard, E., Lim, K. H., Saw, P. N., Koffas, M. A., Engineering central metabolic pathways for high-level flavonoid production in *Escherichia coli*. *Appl. Environ. Microbiol.* 2007, *73*, 3877–3886.
[17] Miyahisa, I., Funa, N., Ohnishi, Y., Martens, S. et al., Combinatorial biosynthesis of flavones and flavonols in *Escherichia coli*. *Appl. Microbiol. Biotechnol.* 2006, *71*, 53–58.
[18] Hwang, E. I., Kaneko, M., Ohnishi, Y., Horinouchi, S., Production of plant-specific flavanones by *Escherichia coli* containing an artificial gene cluster. *Appl. Environ. Microbiol.* 2003, *69*, 2699–2706.
[19] Kaneko, M., Hwang, E. I., Ohnishi, Y., Horinouchi, S., Heterologous production of flavanones in *Escherichia coli*: Potential for combinatorial biosynthesis of flavonoids in bacteria. *J. Ind. Microbiol. Biotechnol.* 2003, *30*, 456–461.
[20] Watts, K. T., Lee, P. C., Schmidt-Dannert, C., Exploring recombinant flavonoid biosynthesis in metabolically engineered *Escherichia coli*. *ChemBioChem* 2004, *5*, 500–507.
[21] Liu, R., Wu, C. X., Zhou, D., Yang, F. et al., Pinocembrin protects against beta-amyloid-induced toxicity in neurons through inhibiting receptor for advanced glycation end products (RAGE)-independent signaling pathways and regulating mitochondrion-mediated apoptosis. *BMC Med.* 2012, *10*, 105.
[22] Soromou, L. W., Zhang, Y., Cui, Y., Wei, M. et al., Subinhibitory concentrations of pinocembrin exert anti-*Staphylococcus aureus* activity by reducing alpha-toxin expression. *J. Appl. Microbiol.* 2013, *115*, 41–49.
[23] Fowler, Z. L., Koffas, M. A., Biosynthesis and biotechnological production of flavanones: Current state and perspectives. *Appl. Microbiol. Biotechnol.* 2009, *83*, 799–808.
[24] Scheich, C., Kummel, D., Soumailakakis, D., Heinemann, U., Bussow, K., Vectors for co-expression of an unrestricted number of proteins. *Nucleic Acids Res.* 2007, *35*, e43.
[25] Grigoras, I., Timchenko, T., Gronenborn, B., Transcripts encoding the nanovirus master replication initiator proteins are terminally redundant. *J. Gen. Virol.* 2008, *89*, 583–593.
[26] Wu, J., Du, G., Zhou, J., Chen, J., Metabolic engineering of *Escherichia coli* for (2S)-pinocembrin production from glucose by a modular metabolic strategy. *Metab. Eng.* 2013, *16*, 48–55.
[27] Xu, P., Vansiri, A., Bhan, N., Koffas, M. A., ePathBrick: A synthetic biology platform for engineering metabolic pathways in *E. coli*. *ACS Synth. Biol.* 2014, *1*, 256–266.
[28] Sambrook, J., Fritch, E. F., Maniatis, T., *Molecular Cloning. A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY 1987.
[29] Blackwell, J. R., Horgan, R., A novel strategy for production of a highly expressed recombinant protein in an active form. *FEBS Lett.* 1991, *295*, 10–12.
[30] Carbonell, P., Fichera, D., Pandit, S. B., Faulon, J. L., Enumerating metabolic pathways for the production of heterologous target chemicals in chassis organisms. *BMC Syst. Biol.* 2012, *6*, 10.
[31] Faulon, J. L., Misra, M., Martin, S., Sale, K., Sapra, R., Genome scale enzyme-metabolite and drug-target interaction predictions using the signature molecular descriptor. *Bioinformatics* 2008, *24*, 225–233.
[32] Carbonell, P., Faulon, J. L., Molecular signatures-based prediction of enzyme promiscuity. *Bioinformatics* 2010, *26*, 2012–2019.
[33] Planson, A. G., Carbonell, P., Paillard, E., Pollet, N., Faulon, J. L., Compound toxicity screening and structure–activity relationship modeling in *Escherichia coli*. *Biotechnol. Bioeng.* 2012, *109*, 846–850.
[34] Feist, A. M., Henry, C. S., Reed, J. L., Krummenacker, M. et al., A genome-scale metabolic reconstruction for *Escherichia coli* K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. *Mol. Syst. Biol.* 2007, *3*, 121.
[35] Li, C., Donizelli, M., Rodriguez, N., Dharuri, H. et al., BioModels database: An enhanced, curated and annotated resource for published quantitative kinetic models. *BMC Syst. Biol.* 2010, *4*, 92.
[36] Ebrahim, A., Lerman, J. A., Palsson, B. O., Hyduke, D. R., COBRApy: COnstraints-Based Reconstruction and Analysis for python. *BMC Syst. Biol.* 2013, *7*, 74.
[37] Conrad, T. M., Frazier, M., Joyce, A. R., Cho, B. K. et al., RNA polymerase mutants found through adaptive evolution reprogram *Escherichia coli* for optimal growth in minimal media. *Proc. Natl. Acad. Sci. USA* 2010, *107*, 20500–20505.
[38] Kaneko, M., Ohnishi, Y., Horinouchi, S., Cinnamate: Coenzyme A ligase from the filamentous bacterium *Streptomyces coelicolor* A3(2). *J. Bacteriol.* 2003, *185*, 20–27.
[39] An, J. H., Lee, G. Y., Jung, J. W., Lee, W., Kim, Y. S., Identification of residues essential for a two-step reaction by malonyl-CoA synthetase from *Rhizobium trifolii*. *Biochem. J.* 1999, *344*, 159–166.

[40] Davis, M. S., Solbiati, J., Cronan, J. E., Jr., Overproduction of acetyl-CoA carboxylase activity increases the rate of fatty acid biosynthesis in *Escherichia coli*. *J. Biol. Chem.* 2000, *275*, 28593–28598.

[41] Heath, R. J., Rock, C. O., Regulation of malonyl-CoA metabolism by acyl-acyl carrier protein and beta-ketoacyl-acyl carrier protein synthases in *Escherichia coli*. *J. Biol. Chem.* 1995, *270*, 15531–15538.

[42] Shen, H. B., Chou, K. C., EzyPred: A top-down approach for predicting enzyme functional classes and subclasses. *Biochem. Biophys. Res. Commun.* 2007, *364*, 53–59.

[43] Matsuta, Y., Ito, M., Tohsato, Y., ECOH: An enzyme commission number predictor using mutual information and a support vector machine. *Bioinformatics* 2013, *29*, 365–372.

[44] Burgard, A. P., Pharkya, P., Maranas, C. D., Optknock: A bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. *Biotechnol. Bioeng.* 2003, *84*, 647–657.

[45] Rockwell, G., Guido, N. J., Church, G. M., Redirector: Designing cell factories by reconstructing the metabolic objective. *PLoS Comput. Biol.* 2013, *9*, e1002882.

[46] Carbonell, P., Parutto, P., Baudier, C., Junot, C., Faulon, J. L., RetroPath: Automated pipeline for embedded metabolic circuits. *ACS Synth. Biol.* 2014, *3*, 565–577.

[47] Carbonell, P., Parutto, P., Herisson, J., Pandit, S. B., Faulon, J. L., XTMS: Pathway design in an eXTended metabolic space. *Nucleic Acids Res.* 2014, DOI:10.1093/nar/gku362.

This issue of Biotechnology Journal is a focus issue on non-viral gene transfer and is edited by Tristan Montier, Chantal Pichon, Marc Blondel and Patrick Midoux. The cover was chosen with the intention of providing an essence of nucleic acid transfer technology.
Image: DNA molecules © vitstudio Fotolia.com

Biotechnology Journal – list of articles published in the November 2014 issue.

Special Articles

Regular Articles