HAL
open science

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

Baudoin Delepine, Vincent Libis, Pablo Carbonell, Jean-Loup Faulon

To cite this version:

Baudoin Delepine, Vincent Libis, Pablo Carbonell, Jean-Loup Faulon. SensiPath: computer-aided design of sensing-enabling metabolic pathways. Nucleic Acids Research, 2016, 44 (W1), pp.W226-W231. 10.1093/nar/gkw305 . hal-02634192

HAL Id: hal-02634192

https://hal.inrae.fr/hal-02634192v1

Submitted on 27 May 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

W226–W231 *Nucleic Acids Research, 2016, Vol. 44, Web Server issue* *Published online 22 April 2016*
doi: 10.1093/nar/gkw305

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

Baudoin Delépine[1,2], Vincent Libis[1,2], Pablo Carbonell[3] and Jean-Loup Faulon[1,2,3,*]

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

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

ABSTRACT

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

INTRODUCTION

Synthetic biology and metabolic engineering applications often require as part of their design a way to assess the presence or to quantify the amount of a compound of interest. Genetically-encoded biosensors such as riboswitches and allosteric transcription factors offer the possibility to control the expression of a gene of choice. This feature makes them valuable for many applications (1,2) such as pollutant monitoring or high-throughput screening of optimized strains and enzymes (3–5), as expression of reporter genes like fluorescent proteins can be linked to the concentration of the compound of interest. Moreover, the ability of these biosensors to provide input at the genetic level opens the way to more complex downstream signal processing and actuation (6). Examples of applications of such circuits range from threshold activation in presence of pathological concentration levels of biomarkers (7) to the creation of a feedback control motif leading to yield improvement for a chemical producing strain (8).

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

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

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

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

© The Author(s) 2016. Published by Oxford University Press on behalf of Nucleic Acids Research.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

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

MATERIALS AND METHODS

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

Source databases

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

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

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

Reaction and compound encoding

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

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

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

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

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

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

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

Matching algorithm

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

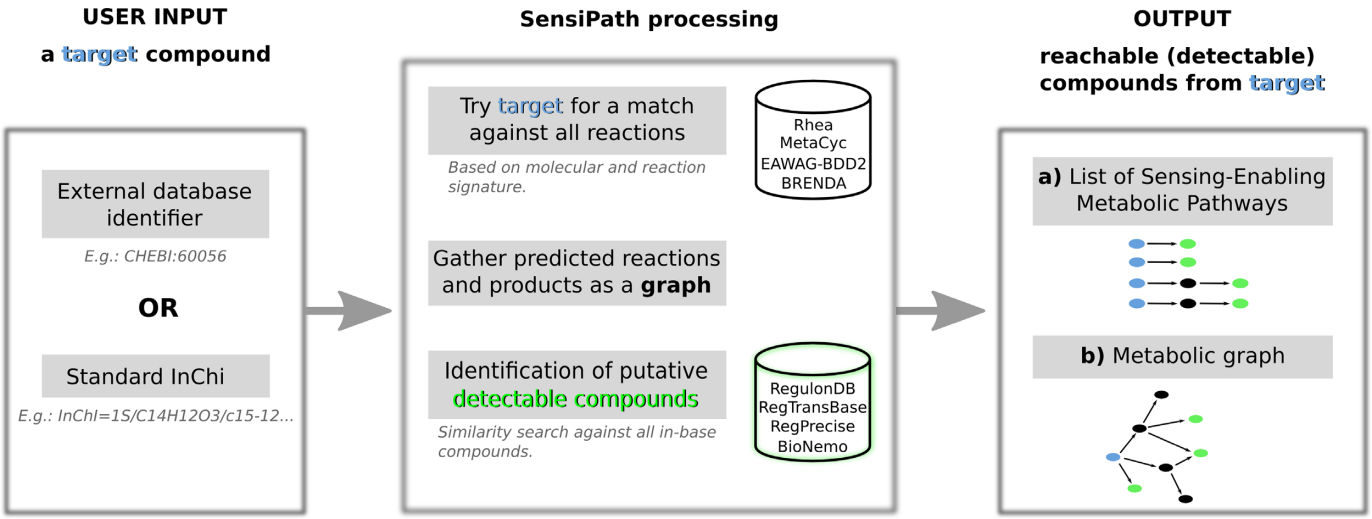

Figure 1. SensiPath flowchart. Users query a target compound (blue), either as an external database identifier or as a standard InChi. Target is processed to gather reachable compounds through enzymatic reactions; detectable compounds are annotated (green). Results are displayed both as (a) the set of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

Metabolic graph

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

Similarity search

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

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

Web server implementation

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

INPUT AND OUTPUT

Input

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

Output

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

CASE STUDIES

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

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

Cocaine detection

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

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

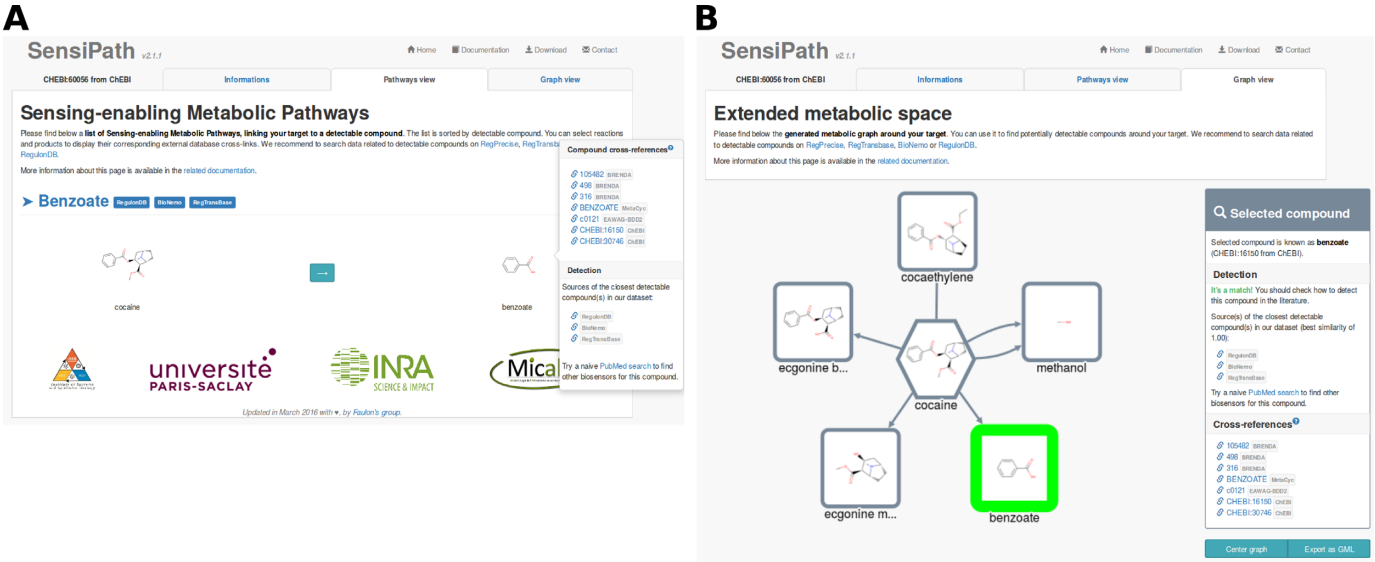

Figure 2. Results pages for querying 'CHEBI:60056' at one step. (**A**) Pathway view. List of sensing-enabling metabolic pathways, sorted by detectable compound. Clicking reactions (arrows) or products will display a list of cross-references to relevant databases. Cross-references which are displayed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

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

In parallel, a query on BioNemo or RegTransBase for benzoate (or benzoic acid), the compound reported by SensiPath as having a biosensor, leads to several potential transcription factors that are known to interact with this compound (BenM, BenR, CbdS, PcaR, TcbR, CatR, BadR and XylS). In our experimental implementation, we chose the couple composed of BenR and its responsive promoter pBen from *Pseudomonas putida* KT2440 after a quick assessment of the available literature.

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

Parathion detection

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

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

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

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

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

DISCUSSION

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

Depending on the application, the reliability of the candidate SEMPs identified by our method may vary. Limitations of the SEMP method include the need for the target compound to be able to co-localize with the enzyme (i.e. to enter the cells or to be internally produced in the cell), and the need for enzymatic products of the sensing pathway to be not too toxic to the cell. Such issues need to be addressed in a case-by-case manner, since they greatly depend on the application and on physico-chemical properties that are not always known for the compound. Other potential limitations of the method hold with regards to the choice of the biosensor. Although some information about the degree of promiscuity of transcription factors may be available from databases and literature, this aspect should be carefully considered in each application, especially if the final application requires a high level of specificity. The choice of the biosensor should also take into account dose response parameters such as the dynamic range and linear range of detection. SEMP's properties will depend on the actual properties of the biosensor, an information that therefore should be considered and retrieved from the available literature. In addition, promoter sequences responding to transcription factors may not be always found in databases, often requiring an investigation of associated references. This information nevertheless is progressively becoming more available through repositories like the Registry of Standard Biological Parts (http://parts.igem.org/Main_Page).

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

AVAILABILITY

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

ACKNOWLEDGEMENT

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

FUNDING

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

Conflict of interest statement. None declared.

REFERENCES

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

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