Cell-Free Biosensors and AI integration

Paul Soudier, Léon Faure, Manish Kushwaha, Jean-Loup Faulon

Cell-Free Biosensors and AI integration

Soudier, P.[1]; Faure, L.[1]; Kushwaha, M.[1]; Faulon, JL.[1,2,3]

[1]Université Paris-Saclay, INRAE, AgroParisTech, Micalis Institute, 78350, Jouy-en-Josas, France.

[2] Université Paris-Saclay, Systems & Synthetic Biology Lab (iSSB), UMR 8030, Evry, France

[3]Manchester Institute of Biotechnology, SYNBIOCHEM center, School of Chemistry, The University of Manchester, Manchester M1 7DN, UK.

Corresponding authors: jean-loup.faulon@inrae.fr.

ABSTRACT

Cell-free biosensors hold a great potential as alternatives for traditional analytical chemistry methods providing low-cost low resource measurement of specific chemicals. However, their large scale use is limited by the complexity of their development.

In this chapter we present a standard methodology based on Computer aided design (CAD) tools that enables fast development of new cell-free biosensors based on target molecule information transduction and reporting through metabolic and genetic layers, respectively. Such systems can then be repurposed to represent complex computational problems, allowing defined multiplex sensing of various inputs and integration of artificial intelligence in synthetic biological systems.

Introduction :

Metabolite biosensors emerged as a major application of synthetic biology. Repurposing biological systems naturally present in various organisms enabled the development of synthetic sensing devices [1] giving rapid and inexpensive point of care measurement of various molecules of interest levels in a broad range of samples with applications in health [2], environment [3], industries and fundamental research [4]. The potential of these types of devices lies in the fact that they are able to deliver information on the presence and levels of certain chemicals without the need for expensive reagents, trained operators and large size facilities. Transcription factor based biosensors are a subset of those sensing devices repurposing transcription factor and inducible promoter for the quick detection of molecules of interest.

Cell-free biosensors emerged as an alternative for traditional whole-cell biosensors, solving several issues carried by these systems [5]. Cell-free biosensors are efficient, generating big signal over noise ratios[15]. They are suitable for the detection of molecules toxic for bacterial growth or molecules that are not able to cross the cell membrane. Their ability to be freeze dried on paper enables long term storage at room temperature without loss of activity pushing their use as point-of-care devices, especially in low resource communities [6]. The absence of living organisms in it is also facilitating their industrial developments with reduced regulatory issues and low biosafety concerns. Finally the ability to express each gene from distinct DNA fragments (plasmids or linear fragments) [7] and to fine tune their expression levels by varying their respective concentrations enable a fast development and optimisation of new biosensor candidates or more complex devices relying on them. This central property of straightforward tuning of gene expression is necessary for the

development of the complex information processing systems that will be described in this chapter.

The main issue encountered in the development of transcription factor (TF) based biosensors is the limited number of molecules for which an interacting transcription factor has been described. The list of these metabolites known to trigger transcriptional response, either in natural systems or in the context of synthetic biosensors, have been described in various databases [8] [9] and compiled in a dataset of small molecules triggering transcriptional and translational cellular responses [10].

In order to detect molecules for which associated transcription factors are not available, new methodologies have been developed relying on the use of metabolic pathways. It has indeed been shown that enzymes could be used to convert non-detectable molecules into molecules known to regulate a characterized transcription factor [11]. This framework has been formalized and showcased as the Sensing-Enabling Metabolic Pathway (SEMP) concept [12]. This shifted the problem from finding a TF binding the molecule of interest to finding any potential enzymatic route between the molecule of interest and the known set of detectable molecules. This was addressed by repurposing bioinformatic tools traditionally used for retrosynthesis [13] into a platform able to find potential SEMP for any molecule of interest. This web-service called SENSIPATH is a CAD program predicting pathways of up to 2 steps allowing detection for any query compounds [14]. This approach was then used for the development of a plug and play cell-free workflow where enzymes formalized as metabolic transducers were used as interchangeable modules converting various chemicals into detectable molecules called effectors [15]. Such biosensors were found to be highly efficient, the cell-free system potentialities allowing to reach optimal response for each biosensor by a

systematic tuning of each component involved (reporter, transcription factor and enzyme encoding DNA). They were also applied for the commercially relevant sensing of molecules of interest in complex real-world samples (preservative in beverage, disease biomarker and drug in clinical samples).

In addition to its biosensing potential, this framework has also then been applied to the field of information processing. Indeed, the constructed devices sensing various molecules can also be conceptualised as information processing systems, converting an input signal (the concentration of detected molecules) into an output one (the level of protein synthesised by the reporter plasmid). This processing device uses the continuous concentration of input metabolites as analog signals able to be processed by the metabolic or the genetic layer featured in the system. Multiple molecules can be continuously converted into the same effector at the same time allowing the construction of multi-input devices derived from analog adders.

Starting from this basis, we have implemented a machine learning architecture that allows integration of AI computations in cell-free systems [16]. One of the simplest architectures for signal processing is the perceptron [17]. A single-layer neural network transforms an input vector into an output value through the application of a weighted sum of the input followed by a thresholding function on the obtained result. Input data can be integrated as categorical or numerical (discrete or continuous). This signal-processing unit has been implemented in a cell-free system to enhance its multiplex sensing abilities [16]. This system is indeed capable of classifying complex samples in two categories (above or below a threshold of the reporter gene expression) based on the recognition of certain patterns of molecules present in those samples. A theoretical perceptron was first constructed and trained *"in silico"* to compute the

optimal set of weights required to solve a defined problem of classification. The perceptron was then implemented in cell-free using a molecular model, giving concentrations of enzymes coding DNA that correspond to the weight of each input (see **figure 1**). The thresholding applied through the effector biosensor dose response curve.

[[Figure 1 should be placed near here.]]

Different elements of the computing unit have a defined terminology: e.g. transducer, adder, actuator. A transducer, in cell-free terms, means one or more enzymes transforming one molecule into another. An adder consists of several reactions having different substrates but yielding the same compound. An actuator is simply a way for the cell-free system to report its activity, through the expression of a reporter gene, for example. Theoretically, many more parameters of a cell-free reaction could be implemented in such signal-processing manner. The diversity of elements can be increased, as well as the number of layers in the perceptron, and the number of detectable molecules.

Here we describe a straightforward methodology that details how to use CAD tools to identify potential new biosensors for a chosen molecule, how to build and test these biosensor candidates and finally how to repurpose these cell-free biosensors into signal processing devices to implement neural computing in biological systems.

This method to this date has been applied for the construction of a single layer 4-input perceptron but is easily scalable for more inputs and adaptable for more complex designs including multi-layer perceptron and other types of computational architectures. An example of potential implementation of a multi-layer perceptron using this system can be found in the supplementary figure 14 of the original metabolic perceptron paper [16].

2. Materials

2.1. Preparation of cell-free extract and buffer

1. Extract preparation equipment

In addition to common lab equipment, material for the preparation of Cell extract include large volume centrifugation equipment (for 1 L bottles) and a French Press.

2. Extract preparation media and buffers

Cells are grown in 2YTP media (31 g/L 2xYT, 40 mM potassium phosphate dibasic, 22 mM potassium phosphate monobasic)

S30A and S30B buffers are also required for the preparation of the Cell Extract

They are prepared with the following components:

- S30A = (14 mM Mg glutamate, 60 mM K-glutamate, 50 mM Tris, pH 7.7)

S30A is titrated using acetic acid, 2mM DTT are added just before use

- S30B = (14 mM Mg glutamate, 60 mM K-glutamate, ~5 mM Tris, pH 8.2)

S30B is titrated using 2 M Tris, 1 mM DTT are added just before use

3. Chemicals for the reaction buffer

    1. Solutions of the 20 canonical amino acids, 19 of them concentrated at 168 mM, except leucine that is concentrated at 140 mM
    2. Chemicals for the energy solution, including individual solutions of: HEPES pH 8, ATP, GTP CTP, UTP, tRNA, CoA, NAD, cAMP, folinic Acid, Spermidine and 3-PGA, all of them concentrated close to their limit of solubility.

3. Additional components for the reaction buffer including K-glutamate solution at 3000 mM, Mg-glutamate solution at 100 mM, PEG8000 solution at 40% w/w and DTT solution at 1M.
4. Reference plasmid pBEST-OR2-OR1-Pr-UTR1-deGFP-T500 (Addgene #40019) for reaction buffer calibration experiments

2.2. Cloning of DNA parts and production of plasmids

Additional materials for the construction and production of plasmid encoding the elements for Biosensors include :

1. High fidelity PCR Material (Q5 High-Fidelity 2X Master Mix or equivalent)
2. Template plasmid for backbone amplifications (plasmids #114597 and #114598 from Addgene) and respective primers :

   FWD primer **B**ackbone for **C**DS :

   FBC: cccGGTCTCtGCTTactttatctgagaatagtc

   REV primer **B**ackbone for **C**DS:

   RBC: cccGGTCTCtCATCatatctcttcttaaagttaaac

   FWD primer **B**ackbone for **P**romoter :

   FBP:cccGGTCTCtATGCgtaaaggcgaagagctgttc

   REV primer **B**ackbone for **P**romoter

   RBP:cccGGTCTCtTAAGaatagtaatacaggatccgaatcgtttcag
3. Thermocycler
4. TAE buffer, 50X
5. 1% agarose gel with SYBR™ Safe DNA Gel Stain or equivalent
6. Monarch® DNA Gel Extraction Kit or equivalent.
7. Thermo Scientific NanoDrop 1000 for DNA concentration determination or equivalent.

8. Thermo Scientific™ Savant™ DNA SpeedVac™ Concentrator Kits to concentrate DNA if required
9. BsaI enzyme
10. T4 DNA ligase
11. T4 DNA ligase Buffer
12. DH5α Competent Cells
13. Lysogenic Broth liquid medium for cell culture
14. Lysogenic Broth + Agar solid medium and Ampicillin for cell plating
15. Monarch® Plasmid Miniprep Kit or equivalent
16. Macherey-Nagel™ Kits NucleoBond™ Xtra Maxi for plasmid purification
17. Sequencing plasmids seqF : gataggttaaggaacgg and seqR : ttgatgcctggcaccaac for sequencing verifications of the inserts

2.3. Running Cell-Free reactions

1. Plate reader equipment : typically Synergy HTX Multi-Mode Reader from Biotek®
2. 384 Well Black Plate, Optically Clear Polymer Bottom
3. Adhesive PCR Plate Seals
4. Analytical grade chemical powder or solution for any sensed effector and target molecule.

2.4 *in silico* materials

The R programming language (version 3.2.3) was used for fitting experimental data from actuator and transducers, as well as for defining and solving classification tasks. One can perform calculations described in section 3.6. with regular computational power,

computational clusters are not required. Other programming languages can be used, but we recommend to use R in order to take advantage of our git repository containing all files from Pandi et al. work [16]. The repository is freely available here: https://github.com/brsynth/metabolic_perceptrons/tree/master/cell_free

3. Methods

3.1. Identifying sensing routes for a target molecule

[[Figure 2 should be placed near here.]]

1. Choose a target molecule to sense. Check that the molecule you want to sense isn't one of the components of the Cell-Free Buffer you are using. **(Note 1)**
2. Identify the InChi identifier or your target. To do so you can either search it by name on the database Pubchem (https://pubchem.ncbi.nlm.nih.gov/) and search for the InChi in the section Identifier. The other possibility is to go on Pubchem sketcher (https://pubchem.ncbi.nlm.nih.gov//edit3/index.html) and to draw the molecule to retrieve it's InChI by replacing SMILEs by StdInChI.
3. Go to http://sensipath.micalis.fr/ to access the online platform. Use the section "Query with a Standard InChI" and paste the InChI of your molecule.
4. Run the tool first for one step pathways. If the results are not satisfying, you can run a second attempt for two step pathways. This can be the case if no pathway to a detectable molecule (in green on the graph view) is identified or if the identified

detectable molecule is non suitable for cell-free biosensors (components of the Cell-Free Buffer) **(Note 2)**

5. Identify the promising effector molecules from the "Pathways view" section.
   If the molecule you want to sense appears in green, it means it is directly detectable without need for enzymatic conversion. For developing a biosensor for it you can skip the enzyme related steps.
6. Download the database of detectable molecules from: https://github.com/brsynth/detectable_metabolites and isolate the lines corresponding to the identified effector. First identify Transcription factor that interacts with your component. You can use column E to retrieve the names of it or, if this information is not available, use the column B that contains literature reference of papers describing potential sensing mechanisms for the molecule.
7. Once you find a TF of interest you have to find the promoters that are potentially regulated by it. To do so you can search for it in databases such as RegulonDB (http://regulondb.ccg.unam.mx/) for *E. coli,* Subtiwiki (http://subtiwiki.uni-goettingen.de/) for *B. subtilis.* You can also try a naive bibliographic search for this regulator to identify features linked to it like regulated promoters but also mechanism of action and possible existing design of biosensors using it.
8. Identify the enzymes converting the molecule you want to sense into the effector for which you found a TF. By clicking on the edges from the SENSIPATH graph view you can retrieve the references associated with the enzymatic reaction of interest. From this point using cross-references between databases or identifiers such as EC number, you can find enzyme candidates in UNIPROT with described catalysed reactions matching your expectations. You can also use the computational tool

Selenzyme (http://selenzyme.synbiochem.co.uk/) with SENSIPATH provided references to find potential hits for other enzymes candidates **(Note 3)**

9. Retrieve DNA sequences for the identified parts. TF and Enzymes sequences can be codon optimised using any of the available tools. Promoter sequence is often defined as the 200 nucleotides before the start codon of a regulated gene. For Troubleshooting purposes, various size promoters can be synthesised to test for their response to the TF and their transcription initiation ability. **(Note 4)**

3.2. Constructing candidate biosensors plasmids

1. Synthesise the previously isolated sequence of your Transcription Factor, its regulated promoter and any enzyme required to convert the molecule you want to detect into the TF binding effector.
2. Design and orders primers for golden gate cloning of the synthesised parts. Genes (TF and enzymes) requires overhangs in the format,

   FWD : ccGGTCTCtGATG…. REV : ccGGTCTCtAAGC….

   promoters requires overhangs in the format,

   FWD : ccGGTCTCtCTTA…. REV : ccGGTCTCtGCAT….
3. Run High fidelity PCRs to amplify with the correct Golden Gate overhangs the Vectors and the synthesised inserts. The reaction typically consists of pipetting 25 µL of 2X Q5® polymerase Master Mix, 2.5 µL of FW and RV primer at 10 µM, 1µL of template DNA (at concentration around 100ng/µL) and 19 µL of water.

   This mix is then typically incubated in a thermocycler applying the following program: 30 s at 98 °C then 35 cycles with (10 s at 98 °C, 30 s at Tm +3 °C, 30 s/kb at 72 °C) then 2 minutes at 72 °C. Tm beings the lowest melting temperatures of the 2

primers and kb being the size of the amplicon in kilo bases. Use primer pairs FBP/RBP on the template plasmid 114598 to reamplify the linearized reporter backbone and the primers FBC/RBC on the template plasmid 114597 to reamplify the linearized backbone for enzyme or TF. Use the newly designed primers from step 2 on the synthesised parts from step 1 to reamplify the inserts.

4. Run an electrophoresis on the PCR product on a 1% Agarose gel stained with SYBR Safe using an appropriate DNA ladder to be able to discriminate your amplicon by its size. After an approximate time of 30 minutes at 100 V identify and cut the band of the expected size by imaging on a blue light transilluminator.
5. Recover the DNA from the Gel using a DNA Gel extraction kit, following the kit's instructions to purify your DNA fragment. Determine the titer of purified DNA using a nanodrop spectrophotometer.
6. Prepare a golden gate reaction to insert each fragment in its respective backbone **(Note 5)**. To do so you need to calculate the molarity of your DNA. You can use the tool available at https://nebiocalculator.neb.com/#!/dsdnaamt using the nanodrop determined concentrations of each purified DNA fragment. For the golden gate reaction incubate 100 fmol of insert with 50 fmol of linearized backbone in a tube with 1 µL T4 DNA ligase, 1 µL BsaI enzyme, 2 µL T4 DNA ligase Buffer and water to adjust the volume to 20 µL. Incubate the mix at the 37 °C for 1 h and then 16 °C for 5 minutes. **(Note 6)**
7. Transform DH5α competent cells with the Golden Gate Reaction product. Incubate 5 µL of golden gate product with 50 µL chemically competent cells at 4° C for 30 min.
8. Heatshock at 42 °C for 45 s.
9. Incubate at 4 °C for 3 min.
10. Add 300 µL LB media and incubate at 37 °C for 1h.

11. Finally spread 100µL of the final mix on LB Agar + Ampicillin (100 µL/mL) plates and incubate overnight at 37 °C.
12. Small scale culture for screening
    1. Select 4 colonies per assembly and inoculate them in 3mL LB medium + Ampicillin (100 µL/mL) overnight.
    2. Purify the plasmids from 2mL the bacterial cultures using a plasmid miniprep kit and use the primers seqF and seqR to check the integrity of each cassette by sanger sequencing.
    3. Save one correct clone per construction by freezing the rest of the liquid culture at -80 ºC after addition of Glycerol to final concentration of 25%.
13. Large scale culture for plasmid production. You will require big quantities (>100 µg) of each plasmid to run the cell-free reactions for the characterisation and optimisation of each biosensor candidate **(Note 7)**. To do so you have to realize large scale culture and plasmid extraction for each construction.
    1. Inoculate 300 mL of LB + Ampicillin (100 µL/mL) from the -80 °C glycerol stock and grow the cells overnight.
    2. Pellet the cells by spinning them at 6,000 g for 15 min at 4 °C.
    3. Use the Maxiprep kit to recover plasmid DNA from the pellet.
    4. After the last step of your purification, resuspend the precipitated DNA in 200 µL pure water in order to have a final solution at high concentration.
    5. Measure the final concentration using a nanodrop and adjust it at 1 µM by either diluting it with water either concentrating it using a speedvac machine.

3.3. Preparing in house Cell-Free Extract and Buffer

The method briefly described here is adapted from a widely used protocol [18]of 3-PGA powered cell-free mix with minor modifications mostly concerning the lysis method and the starting strain.

1. Extract preparation :

1. Inoculate BL21* cells from an overnight culture in 4 L of 2YTP medium.
2. Stop the culture at OD 2 and pellet the cells by centrifugation for 12 min at 5,000 x g at 4 °C in 4-L bottles.
3. Rinse the cells twice by successive resuspension/centrifugation steps with 250 mL of S30A buffer.
4. Resuspend the pellets in 40 mL of S30A buffer and transfer the suspension to pre-weighed 50-mL Falcon tubes.
5. Centrifuge the tubes at 2,000 x g at 4 °C during 8 min, discard the supernatant and after weighing the pellets freeze them overnight at -80 °C.
6. Thaw the cell-pellet on ice, weigh the pellet and resuspend them in 1 mL of S30A Buffer per gram of cell pellet.
7. Lyse the cells by passing the whole flow once through a french press at 15,000 psi. (**Note 8**)
8. Centrifuge the lysate at 12,000 g at 4 °C during 30 min.
9. Incubate the supernatant for 1 h at 37 °C with a 220 rpm shaking before a second centrifugation at 12,000 g at 4 °C during 30 min.
10. Transfer the supernatant to a 12-14 kDa MWCO dialysis cassette and incubate the cassette overnight in 2L S30B buffer at 4°C. (**Note 9)**
11. After a last centrifugation at 12,000 g 30min 4 °C, aliquot the supernatant (500 µL in 1.5-mL tubes) and flash freeze them in liquid nitrogen before storing at -80 °C.

2. Buffer preparation

1. Starting from individual solutions of chemicals dissolved in pure water (described in detail in the original paper [18]) prepare an Amino Acid (at 4X concentration) and an Energy solution mix (at 14X concentration) that will be used for buffer preparation. The Amino Acid mix has to be prepared by mixing all the 20 canonical Amino Acids at a final concentration of 6 mM except for leucine at 5 mM. Prepare Energy solution with HEPES pH 8 700 mM, ATP 21 mM, GTP 21 mM, CTP 12.6 mM, UTP 12.6 mM, tRNA 2.8 mg/ml, CoA 3.64 mM, NAD 4.62 mM, cAMP 10.5 mM, folinic Acid 0.95 mM, Spermidine 14 mM, 3-PGA 420 mM. Store each Mix at -80 ºC.
2. Using previously prepared extract, Amino acid Mix, and the Energy Solution, you will have to evaluate in this order the best concentration of: (i) Mg-glutamate, (ii) K-Glutamate, (iii) DTT, and (iv) PEG8000 to add to your buffer to optimize the protein production of your extract. You will run 4 consecutives calibration experiments of 8h cell-free reactions. For each calibration, prepare a master mix for 12 reactions by mixing 88µl extract, 66µl Amino Acid mix, 18.86µl energy solution and 13.2 µL of the Addgene plasmid #40019 concentrated at 200 nM. Add the remaining three components that you are not calibrating for from the following four (Mg-glutamate, K-glutamate, PEG8000 and DTT) at either a starting concentration (if the optimum has not been determined yet), or the optimal concentration determined in a previous step. Finally, add water to this mix to reach a final volume of 237.6 µl. Prepare each calibration reaction by mixing in PCR tubes 19.8 µl of the master mix with 2.2 µl of the tested component concentrated at 20X. Then, take 20 µl from each

PCR tube and pipette them inside individual wells of a 384 wells microplate before incubating the plate for 8h while measuring GFP signal produced (ex: 458 nm, em: 528 nm).

The concentration that leads to the highest GFP signal at 8h is identified as the optimal one to be used for future calibrations and run.

The starting concentration for each component are the following ones :(6mM for Mg-glu, 80mM for K-glu, 0 for DTT and 2% for PEG8000).

The tested concentration for each component are the following:

Mg-glu : {0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 mM}

K-glu : {0, 20, 40, 60, 80, 100, 120, 140, 160, 180 mM}

DTT : {0, 0.5, 1, 1.5, 2, 2.5, 3, 3.5, 4 mM}

PEG8000 {0, 0.5, 1, 1.5, 2, 2.5, 3, 3.5, 4%}

3. Prepare the final Buffer mix that will be used for running the cell-free reactions. The following quantities are for 1 tube of 650 µL (you need to scale up to have 2 tubes of buffer for 1 mL of final extract produced): per tube of buffer, add 128.97 µL of Amino Acid mix, 110.54 µL of Energy Solution, PEG8000, K-glu, Mg-glu and DTT according to the best determined concentration and pure water to adjust the volume to 650 µL.

3.4. Running Cell-Free reactions

All the cell-free experiments in the following parts should be run according to the same methodology. This method is for the preparation of N number of 20-µL cell free-reactions. All the following reactions have to be prepared on ice and have to be run in technical triplicate.

1. Thaw your cell free reagents (extract and buffer) on ice.

2. Pre-dilute your DNA and inducer stock: you will prepare 22-µL reactions each with 7.33 µL of extract, 9.17 µL of buffer and 5.5 µL of other components (DNA plasmids, inducer(s) and water to adjust). The easiest way to proceed if you have less than 5 different components to add per reaction is to make 20X solutions of the plasmids and inducers you want to use to add 1.1 µL of those to the reaction.
3. Prepare your cell-free Master Mix for (N+15%) reactions to compensate for the pipetting loss. Add (7.33*(N+15%)) µL of cell extract and (9.17*(N+15%)) µL of Buffer to a single 1.5-mL tube.
4. Mix by briefly vortexing.
5. Pipette 16.5 µL of that mix in N PCR tubes (in strips).
6. Add the respective other inputs (plasmids, inducers, water…) from the 20X stocks, to each PCR tube up to 22 µL.
7. Close these tubes and mix them by briefly vortexing and bench centrifugation.
8. Pipette 20 µL from each tube to a well of a black 384-well plate pre-chilled.
9. Then cover the plate with a transparent sealing film.
10. Using a plate reader, monitor the green fluorescence (ex: 458 nm, em: 528 nm) at various gains during a 12-h kinetic run.

3.5. Cell-free biosensors characterizations and optimizations

In order to obtain the best possible biosensor response for our target we need to first develop an efficient biosensor for the TF binding effector, before optimizing the enzymatic conversion of our target molecule into our effector.

The first step is aimed to screen for any potential response of designed candidate biosensors and to answer 2 questions: first is there any interaction between

the TF expression and the level of expression of the reporter and then is this interaction modulated in any way by the effector that we are trying to detect. To answer those, you need to characterize the behavior of the reporter plasmid in a cell-free reaction in presence or absence of Transcription Factor and in presence or absence of Effector. If you have identified multiple candidates for the sensing of the same molecule this first experiment can also be used to decide what are the TFs and the promoters most promising for the nexts steps.

1. Run cell-free reactions varying the quantity of added plasmids and chemicals in the following possibilities: Reporter DNA concentration at 0 or 30 nM, TF DNA concentration at 0 or 30 nM and inducer concentration at 0, 100 µM or 1 mM.
2. Use end point kinetics results at 8 hours to evaluate the potential mechanism of your TF (activation or repression) and a possible response of your system to the inducer.
3. Find the optimal plasmid concentration for the Reporter and TF that results in a maximal response for your biosensor. With 3 concentrations of inducer (0 µM, 100 µM and 1 mM), test a combination of concentrations gradients for the added plasmid DNA varying reporter and TF DNA concentration on a logarithmic scale (0, 0.1, 0.3, 1, 3, 10, 30 and 100 nM)
4. Using the previously determined best pair of plasmid concentration, evaluate the inducer dose response of the constructed sensor. Run a cell-free experiment with inducers concentration at : 0 nM, 0.5 nM, 1 nM, 2 nM, 5 nM, 10 nM, 20 nM, 50 nM, 100 nM, 200 nM, 500 nM, and 1,000 nM. Dose response curve is obtained by plotting (fluorescence at concentration x / fluorescence at concentration 0) for each datapoint. Use the final results obtained to assess the

performance of the developed sensor and its conformity with your objective of sensitivity. **(Note 10)**

5. Once you have a satisfying sensor for your effector, you can start screening for enzymes candidates to convert your target into this effector. The first experiment to do consists of screening for any potential activity on each of the selected enzymes candidates. Prepare a cell-free experiment with the previously identified best concentration for reporter and TF DNA, varying Enzyme DNA concentration at 0 nM or 10 nM and the molecule of interest at 0 mM or 1 mM.
6. Select the best enzyme candidate and optimize the expression of the enzyme on a dose response curve (same concentrations as in step 3) by varying the DNA concentration coding for the enzyme on the scale of 0, 0.1, 0.3, 1, 3, 10, 30, 100 nM.

3.6. Design, Build and Test a perceptron

1. Define the architecture of the perceptron based on input molecules. There are 2 possibilities depending on the level of constraints you have with the input molecules. If you want to implement a computing device with a defined behavior but without any constraint on the nature of the inputs, you can reprogram the already developed benzoate based perceptron [16] to implement your desired computational function.
2. If you are interested in multiplexed sensing for defined molecules, you need to identify a common effector they can be converted into. You can use SENSIPATH with a reaction length of 2 to increase the chance to find a product accessible from the 2 or more substrates. Some central molecules

(lactate, acetate, hydrogen peroxide, ammonium…) may be useful for that purpose but a specific attention has to be put on the development of biosensors with high dynamic range for these molecules with potential high noise issues related to their central positions on the metabolic networks.

3. Create and fit the cell-free model. It can be decomposed in 2 parts: actuators and transducers.
4. An actuator is modelled with a modified Hill function; commonly used in biochemistry for simulating the binding of ligands to proteins according to the ligand concentration.

$$\square\square\square\square\square\square\square\square(\square\square\square\square\square)$$

$$= \left(\frac{(\square\square\square\square\square)^{\,h\square\square\square_{\square}}}{(\square_{\square})^{\,h\square\square\square_{\square}} + (\square\square\square\square\square)^{\,h\square\square\square_{\square}}} \times \square\square + 1\right) \times \square\square\square\square\square$$

$$+ \square\square\square \times 0.0001 \times \square\square\square\square\square$$

Function description:

- *total*: concentration of input metabolite in µM
- $\square_{\square}$: concentration of input metabolite yielding half of the maximum induction of the system (also called IC50)
- $h\square\square\square_{\square}$: Hill coefficient characterizing the cooperativity of the induction system
- *fc* : stands for fold-change, corresponds to the dynamic range (in Arbitrary Units) of the system
- *basal* : basal GFP fluorescence signal without input, i.e. the background noise of the system
- *lin* : accounts for the linearity of the system when dealing with concentrations saturating the Hill transfer function

5. A transducer is also modelled with a particular Hill function:

$$\square\square\square\square\square\square\square\square\square\square(\square\square\square\square\square) = \square\square\square\square\square_{\square\square\square\square\square\square\square} \times \left(\frac{(\square)^{\square_{\square}}}{(\square_{\square})^{\square_{\square}} + (\square)^{\square_{\square}}}\right) \times \left(\frac{(\square\square\square\square\square)^{\square_{\square\square\square\square\square}}}{(\square_{\square})^{\square_{\square\square\square\square\square}} + (\square\square\square\square\square)^{\square_{\square\square\square\square\square}}}\right)$$

Function description:

- *input:* input metabolite concentration in µM
- *range_enzyme* : coefficient characterizing the capacity of the enzyme to transduce the signal (dimensionless)
- *E*: enzyme concentration in nM
- $\square_{\square}$: Hill constant for enzyme concentration *E*
- $\square_{\square}$: Hill constant for enzyme concentration *E*
- $\square_{\square}$: Hill constant for input metabolite *input*
- $\square_{\square\square\square\square\square}$: Hill constant for input metabolite *input*

We recommend at least some inspiration from our fitting process described in the notes section **(Note 11)** as it puts an emphasis on tackling the loss of signal of the actuator when the whole system is modelled then implemented. Otherwise, one can adapt another fitting process for a particular project's needs.

6. Measure the model's performance with different metrics: Root Mean Square Deviation (RMSD); R^2; Weighted R^2 ; Error percentage...

Ensuring high metrics (e.g. above 0.9 for scores between 0-1) on many experimental data points guarantees a robust model for predicting the weights schemes for each classification task.

7. Define your classification problem. Start by defining a set of tasks (here, classifications), which outputs either 0 (OFF) or 1 (ON); as well as a set of weights to be tested (equivalent to a range of possible enzyme concentrations in the cell-free experiment, for our work it was between 0.1-10nM).
8. Continue by sampling uniformly input values for your problem, that is to be resolved by the perceptron. For example, in the case of a binary classification of solutions composed of hippurate and cocaine, sample points in a given range, either for a “low” concentration or a “high” concentration for each of the compounds. Here, let us assume we can sample between 0 and 2 μM for low concentrations and between 80 and 100 μM for high concentrations.
9. For each of the two clusters, choose to sample either in the low or high range for each compound.
10. Then, two clusters have been produced and a binary classifier can be easily defined on these points: the perceptron set of weights and its corresponding fluorescence threshold. Please find a visual example of the definition of a classification task on **figure 3**. Further is detailed how to find the set of weights and the fluorescence “decision threshold”.

[[Figure 3 should be placed near here.]]

11. Predict the best set of weights for solving this problem (aka “train” the perceptron on a classification task). Using the previously fitted (part 2) and

benchmarked (part 3) model, simulate all possible input combinations with all possible sets of weights.

12. Screen for performant set of weights, i.e. those enabling a sharp threshold between the output states of the system ("ON" or "OFF"). Several thresholds can be tested for considering an output value as ON or OFF, for all possible simulations.
13. Manually select the best set of weights and corresponding threshold, i.e. those that show the highest and clearest difference between ON and OFF behaviors, and in most scenarios. Also, prefer those showing low enzyme concentrations (to avoid resource competition). One can also test several possibilities (several sets of weights and/or thresholds) in the following cell-free experiments.
14. Implement the designs predicted as best ones, to test it in a cell-free reaction. Start by drawing a test set of chemical (input) combinations from the compositions used for the training of the *in silico* perceptron. Be cautious to evaluate enough points from the space of possibilities to capture the behavior of your system. If your perceptron is designed to solve the particular problem of binary inputs sample classifications, you can eventually build a complete test set with all the possible combinations of inputs (see **figure 1D**).
15. The designed test set can be prepared at 20X concentration in PCR tube strips to be evaluated on various perceptron implementations. We advise you to prepare a single master mix with the cell-extract, the buffer and the DNA coding for the reporter system and the various transducers at the desired concentration for each implementation of the perceptron that you designed.

16. Use this master mix to evaluate the response of the system on the chemical test set previously designed. If the perceptron does not have the expected behaviour from *in silico* analysis, try another set of weights.

4. Notes

1. Most of the molecules present in the cell-free buffer are in high concentration masking any potential response for a biosensor designed to detect them. This list includes all the 20 amino acids, HEPES, ATP, GTP, UTP, CTP, tRNA, CoA, NAD, cAMP, folinate, spermidine, 3-PGA, Magnesium, Potassium and DTT.
2. If you cannot find satisfying results with the software SENSIPATH you can try to find other SEMP using the retrosynthesis workflow Retropath [13]. It has the advantage to allow prediction for pathways with more than 2 steps or pathways using promiscuous activity of enzymes to find potential new reactions. To run retropath for biosensor design, use the list of detectable molecules [10] as Sink, the molecule you want to detect as source and the reaction rules in the forward direction. Results coming from this workflow have to be taken with more care as it relies on less reliable predictions.
3. The Selenzyme tool is predicting potential enzymes catalysing a defined reaction based on similarity of sequences, reactions or other features existing between enzymes from a well-annotated database and enzymes potentially catalysing the query reaction. The predictions should be manually checked in published literature/ datasets to evaluate if the identified enzyme can likely catalyse the given reaction.
4. The methodology described here is a standard to be used in the case where the identified promoter has limited annotation or features described. You are strongly encouraged to search for existing biosensing projects described in the literature that uses the same Transcription Factor with defined size promoters or synthetic ones

(built by inserting TF binding sequences in another promoter) as these promoters may show a better response than the natural ones.

5. Golden Gate Assembly was chosen over other methods like Gibson Assembly as it allows reusing the same primers to reamplify backbone for every new insert cloned.
6. This thermocycler protocol for Golden Gate is a variation of the fast Golden gate assembly protocol (1h 37 °C, 10 min 55 °C) that is adapted to backbone plasmid containing internal bsaI cut sites. Removing the last 55 °C step and adding one at 16 °C for 5 min avoid cutting the final assembly containing the BsaI site.
7. DNA batch can have an influence on its expression level in cell-free limitating reproducibility of results from one maxiprep to another. You are advised to purify each plasmid in a big enough quantity for running all your experiment with a single batch. If necessary, run multiple maxiprep in parallel and mix the resulting DNA to have a sufficiently large quantity of plasmid.
8. Sonication and autolysis have also been successfully tested as lysis methods on this protocol for the development of cell-free biosensors with good results. The french press method has the advantage of being easily scalable for the production of large quantity of extract but if you lack the equipment for it you can adapt the protocol to any other lysis method.
9. In our experience the dialysis did not show a major impact on behavior of the final extract. For optimisation purposes, dialysed and non dialysed extract can be screened to find the condition giving the best response for specific biosensors
10. If you plan to use this effector biosensor for indirect detection (using metabolic transducers) or for multiplex sensing (through a perceptron like architecture) you may want to have it optimise for the detection of inducers present at a lower concentration than what you need for your target molecule as the transduction of the signal through

the enzymatic layer often go with a decrease of sensitivity and an increase of potential noise.

11. Once these functions are encoded in R, the actual fitting can happen.

    First, fit the actuator experimental data (effector dose response curve obtained in part 3.5 section 3) to the Hill function model.

    To do so, fit 100 times the actuator experimental values to the actuator model. Let all parameters be able to vary. Use ordinary least-square error or the R "optim" function (that uses the Limited-memory Broyden Fletcher Goldfarb Shanno algorithm) as the objective function for the data fitting. Keep the seed of the fitting process in memory so it ensures reproducibility. Retrieve fitted parameters for each fitting (a population of 100 sets of fitted parameters will be produced). From this population, save the mean, standard deviation, standard error and confidence interval; for each parameter.

    To account for the decrease of signal in experimental actuator data when the whole system is implemented in cell-free, one needs to fit the transducer model in a specific way.

    To fit the transducers models, we need to actually fit the whole model to the whole system's cell-free experimental data.

    In order to have a coherent perceptron, we will actually constrain the actuator's parameters, then fit all transducers *and* the actuator together.

    To do so, start by constraining each of the actuator's previously fitted parameters with bounds corresponding to the 95% confidence interval; or +/- one standard deviation from the mean, in the case of following parameters: the fold-change, termed "*fc",* and the baseline, termed "*basal*".

    We call "pseudo-experimental" the data used to fit the whole system, but that originates from specific components' experimental data. These individual data were

aggregated as described in the whole-system model: combine transducers data by simply adding their output and feeding the sum to the actuator. Note that in the cell-free experimental system, the same process happens: e.g. benzoate concentrations are added as a result of all transducers yielding benzoate, and the actuator takes this aggregated benzoate concentration as an input. As a result, we obtain "pseudo-experimental" data for the whole cell-free system, with experimental data only available for each component separately.

Once your whole system is modelled and you have constrained your actuator's parameters, you can fit your transducer parameters. Initialize the whole model by drawing actuator parameters values according to a Gaussian distribution centered on the mean of the parameters estimations, with a standard deviation equal to the standard error of this parameter estimation. Let all transducer parameters vary to fit the whole system to the previously described 'pseudo-experimental' data.

Acknowledgments

PS and JLF are supported by the ANR SynBioDiag grant number ANR-18-CE33-0015. LF is supported by the French National Research Institute for Agriculture, Food, and Environment (INRAE), through the "Métaprogramme BIOLPREDICT". MK acknowledges funding support from Ile-de-France region's DIM-RFSI, INRAE's MICA department and the University of Paris-Saclay.

References

1. van der Meer JR, Belkin S (2010) Where microbiology meets microengineering: design and applications of reporter bacteria. Nat Rev Microbiol 8:511–522. https://doi.org/10.1038/nrmicro2392
2. Chang H-J, Voyvodic PL, Zúñiga A, Bonnet J (2017) Microbially derived biosensors for diagnosis, monitoring and epidemiology. Microb Biotechnol 10:1031–1035. https://doi.org/10.1111/1751-7915.12791
3. Rodriguez-Mozaz S, Alda MJL de, Marco M-P, Barceló D (2005) Biosensors for environmental monitoring: A global perspective. Talanta 65:291–297. https://doi.org/10.1016/j.talanta.2004.07.006
4. Koch M, Pandi A, Borkowski O, et al (2019) Custom-made transcriptional biosensors for metabolic engineering. Curr Opin Biotechnol 59:78–84. https://doi.org/10.1016/j.copbio.2019.02.016
5. Zhang L, Guo W, Lu Y (2020) Advances in Cell-Free Biosensors: Principle, Mechanism, and Applications. Biotechnol J 15:2000187. https://doi.org/10.1002/biot.202000187
6. Pardee K (2018) Perspective: Solidifying the impact of cell-free synthetic biology through lyophilization. Biochem Eng J 138:91–97. https://doi.org/10.1016/j.bej.2018.07.008
7. Sun ZZ, Yeung E, Hayes CA, et al (2014) Linear DNA for Rapid Prototyping of Synthetic Biological Circuits in an Escherichia coli Based TX-TL Cell-Free System. ACS Synth Biol 3:387–397. https://doi.org/10.1021/sb400131a
8. Huerta AM, Salgado H, Thieffry D, Collado-Vides J (1998) RegulonDB: A database on transcriptional regulation in Escherichia coli. Nucleic Acids Res 26:55–59. https://doi.org/10.1093/nar/26.1.55
9. Cipriano MJ, Novichkov PN, Kazakov AE, et al (2013) RegTransBase – a database of regulatory sequences and interactions based on literature: a resource for investigating

transcriptional regulation in prokaryotes. BMC Genomics 14:213. https://doi.org/10.1186/1471-2164-14-213

10. Koch M, Pandi A, Delépine B, Faulon J-L (2018) A dataset of small molecules triggering transcriptional and translational cellular responses. Data Brief 17:1374–1378. https://doi.org/10.1016/j.dib.2018.02.061
11. Xue H, Shi H, Yu Z, et al (2014) Design, Construction, and Characterization of a Set of Biosensors for Aromatic Compounds. ACS Synth Biol 3:1011–1014. https://doi.org/10.1021/sb500023f
12. Libis V, Delépine B, Faulon J-L (2016) Expanding Biosensing Abilities through Computer-Aided Design of Metabolic Pathways. ACS Synth Biol 5:1076–1085. https://doi.org/10.1021/acssynbio.5b00225
13. Delépine B, Duigou T, Carbonell P, Faulon J-L (2018) RetroPath2.0: A retrosynthesis workflow for metabolic engineers. Metab Eng 45:158–170. https://doi.org/10.1016/j.ymben.2017.12.002
14. Delépine B, Libis V, Carbonell P, Faulon J-L (2016) SensiPath: computer-aided design of sensing-enabling metabolic pathways. Nucleic Acids Res 44:W226–W231. https://doi.org/10.1093/nar/gkw305
15. Voyvodic PL, Pandi A, Koch M, et al (2019) Plug-and-play metabolic transducers expand the chemical detection space of cell-free biosensors. Nat Commun 10:1697. https://doi.org/10.1038/s41467-019-09722-9
16. Pandi A, Koch M, Voyvodic PL, et al (2019) Metabolic perceptrons for neural computing in biological systems. Nat Commun 10:3880. https://doi.org/10.1038/s41467-019-11889-0
17. Rosenblatt F (1957) The Perceptron, a Perceiving and Recognizing Automaton Project Para. Cornell Aeronautical Laboratory
18. Sun ZZ, Hayes CA, Shin J, et al (2013) Protocols for Implementing an Escherichia coli Based TX-TL Cell-Free Expression System for Synthetic Biology. JoVE J Vis Exp e50762. https://doi.org/10.3791/50762

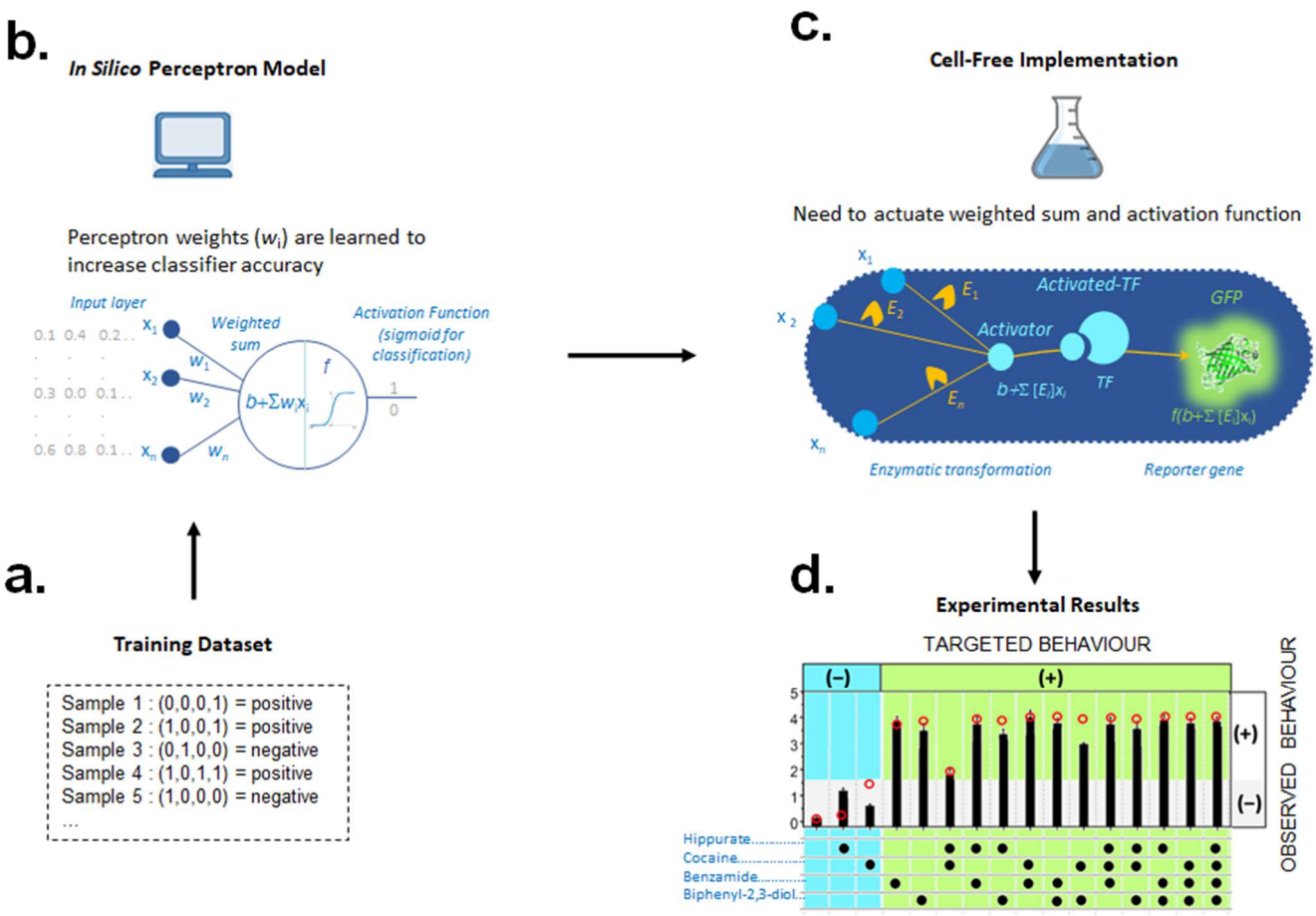

Figure 1 : Implementation of a clustering problem in a biological system through the metabolic perceptron : to obtain a genetic device operating multiplex sensing on a set of defined clustered samples an *in silico* model (b) is first trained with the input dataset (a). The results give information on the best set of DNA concentration to implement the problem in a cell free environment (C). The device is evaluated for response to a set of samples with various compositions (d).

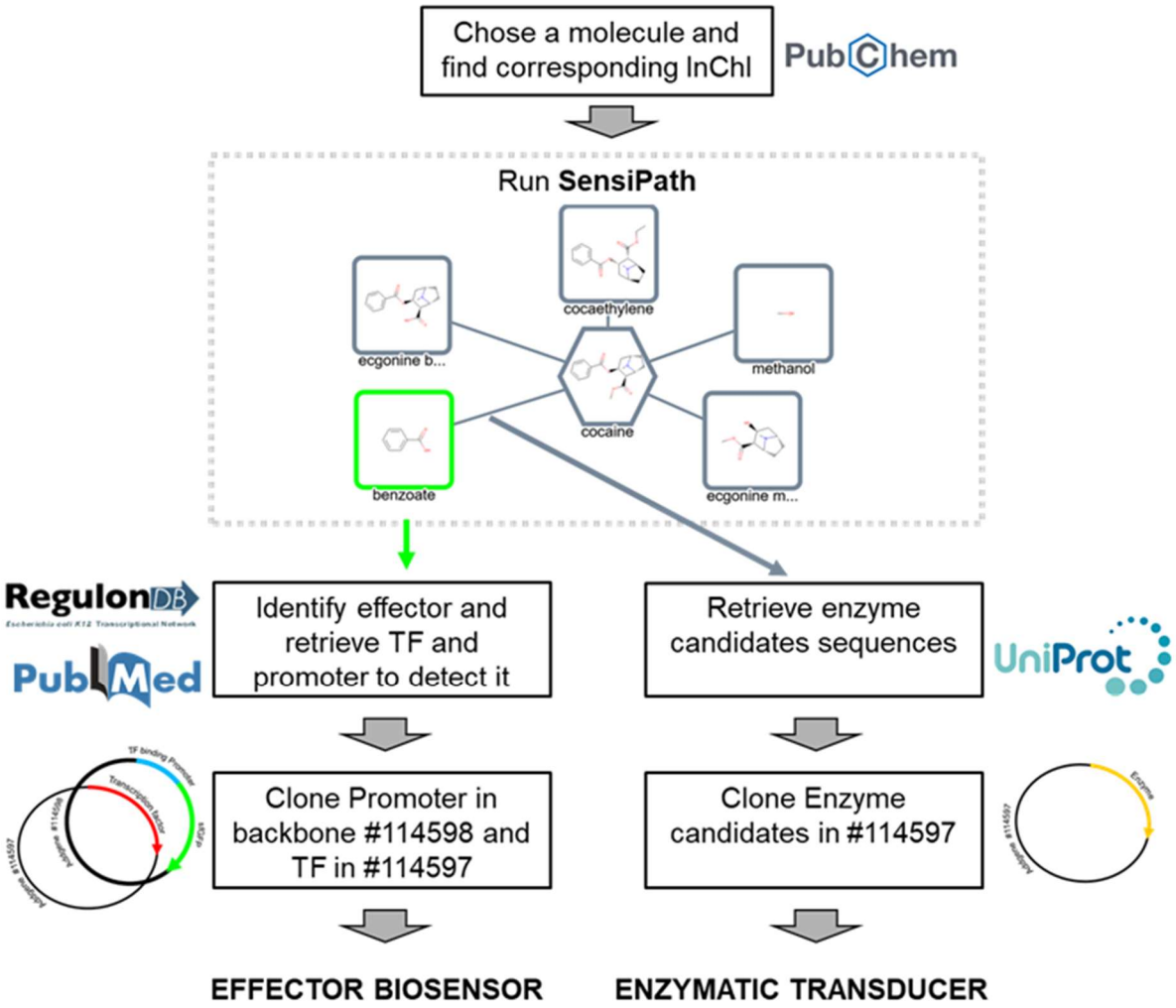

Figure 2 : CAD pipeline for Cell-Free biosensor Design. Main steps of the biosensor design and construction process are presented, from the identification of the chemicals identifiers for the target molecule, to the cloning of the DNA parts in plasmids. The sensipath results showed here are the output given for a query of 1 step detection potentialities for cocaine.

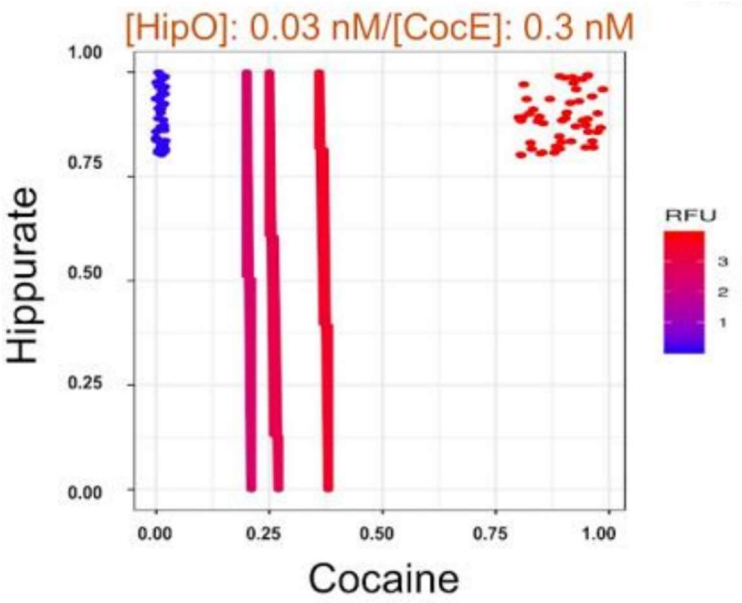

Figure 3 : Visual example of a binary clustering with a metabolic perceptron

This figure has been copied from supplementary material of Pandi et al. [16] with authorization from the authors. The X-axis shows the normalized concentration of cocaine in a sample, the Y-axis shows the normalized concentration of Hippurate in a sample.

Here the blue points drop into the first cluster (“high Hippurate, low cocaine”) and the red ones drop into the second cluster (“high cocaine, high hippurate”). Colors of the points correspond to the predicted RFU values by the fitted perceptron (after section 3.6, part 3. is done). The 3 vertical lines correspond to 3 isofluorescence lines; equivalent to 3 fluorescence thresholds, each of them being associated with one set of weights of the perceptron (i.e. [HipO]: 0.03nM and [CocE]: 0.3 nM).