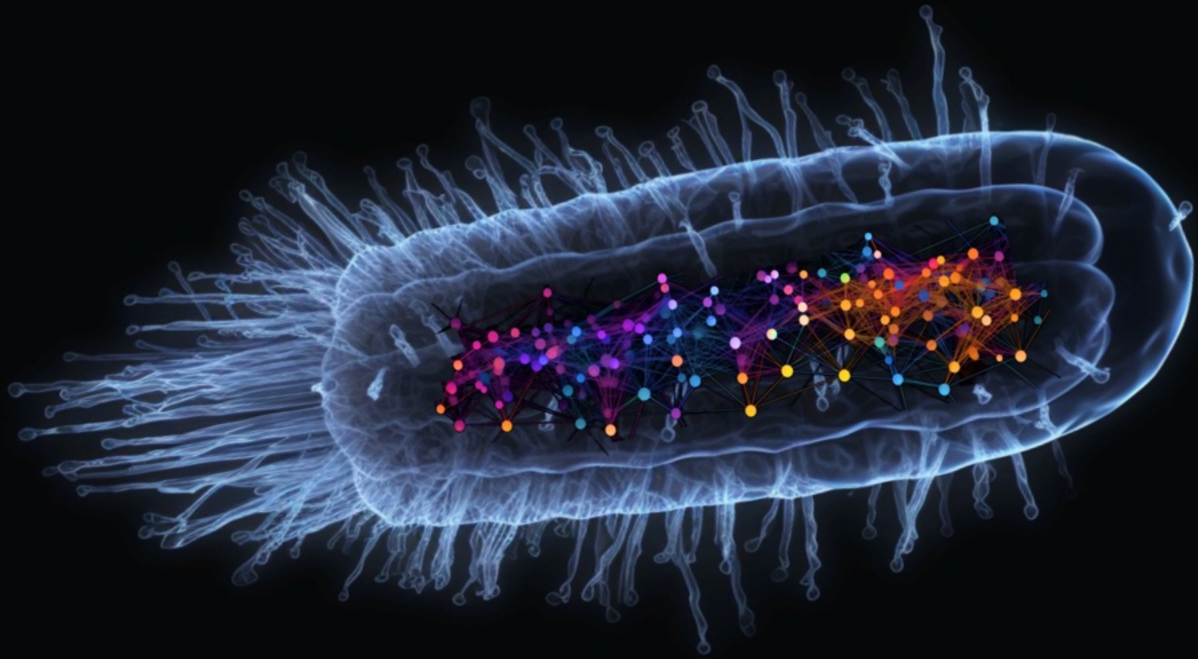


Produce, sense and compute with metabolism

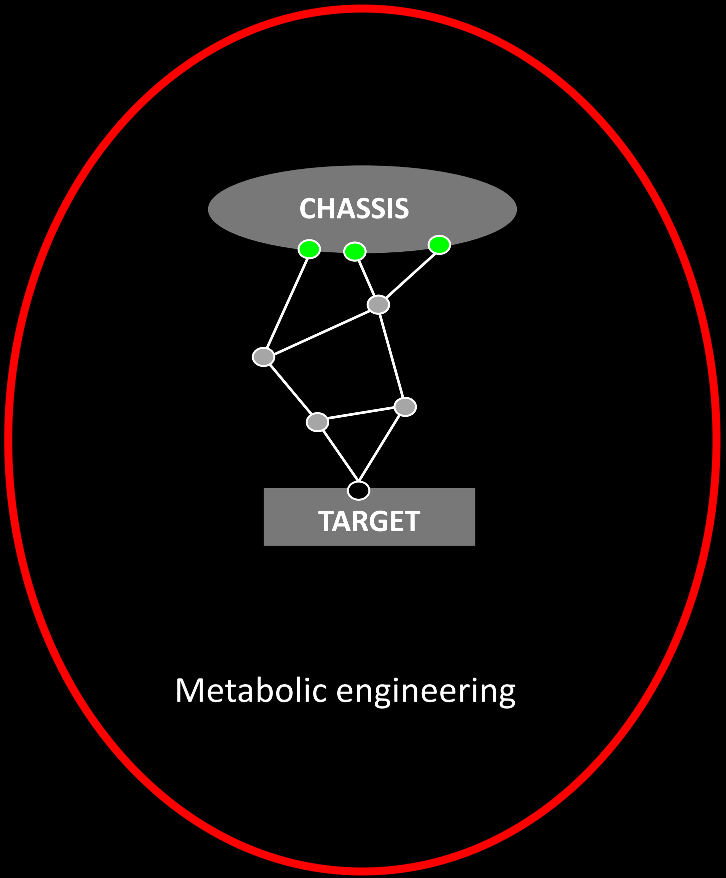
Jean-Loup Faulon – Jean-Loup.Faulon@inrae.fr – https://jfaulon.com

universite PARIS-SACLAY

MICALIS Institute

INRAE

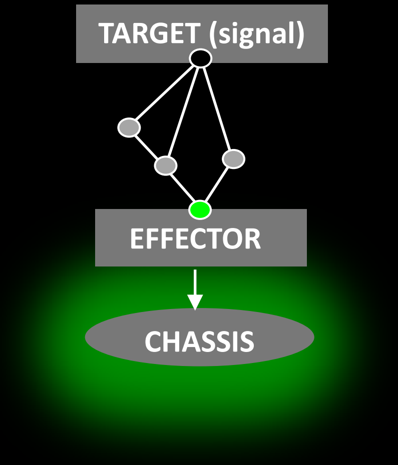
Engineering metabolic and genetic networks in vitro (cell-free) and in vivo (E. coli)



CHASSIS

TARGET

Metabolic engineering

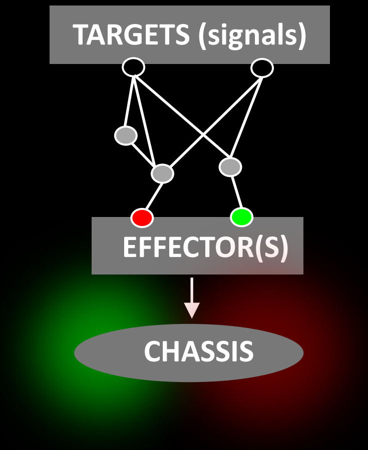


TARGET (signal)

EFFECTOR

CHASSIS

Biosensor engineering



TARGETS (signals)

EFFECTOR(S)

CHASSIS

Information processing
biocomputing

BNICE

web service

EPFL

RetroRules

INRAE

RP2Path

RetroPath2.0

INRAE

eQuilibrator

web service

WEIZMANN INSTITUTE OF SCIENCE

FBA

cobrapy

UNIVERSITY OF CALIFORNIA SAN DIEGO

DTU

cameo

Pathway ranker

web service

INRAE

PartsGenie

web service

UNIVERSITY OF LIVERPOOL

UNIVERSITAT POLITÈCNICA DE VALÈNCIA

DoE

BASIC

Imperial College London

Selenzyme

web service

MANCHESTER 1824

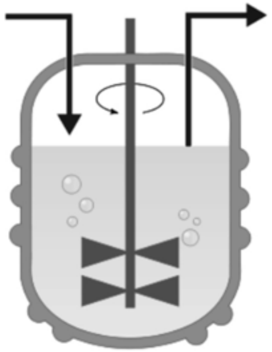
DIVA

BERKELEY LAB

DNA WEAVER

web service

THE UNIVERSITY of EDINBURGH



Benchling

OpenMS

EBERHARD KARLS UNIVERSITÄT TÜBINGEN

Workflow4metabolomics

ifb INSTITUT FRANÇAIS DE BIOINFORMATIQUE

Riffyn

Many tools that do not necessarily follow FAIR principles

CELLO

web service

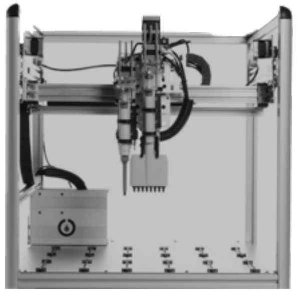
MIT

POLITÉCNICA

BioBlocks

Aq

UNIVERSITY of WASHINGTON



antha

Synthace

DNA-BOT

web service

Imperial College London

PennState JOHNS HOPKINS UNIVERSITY OREGON HEALTH & SCIENCE UNIVERSITY TACC

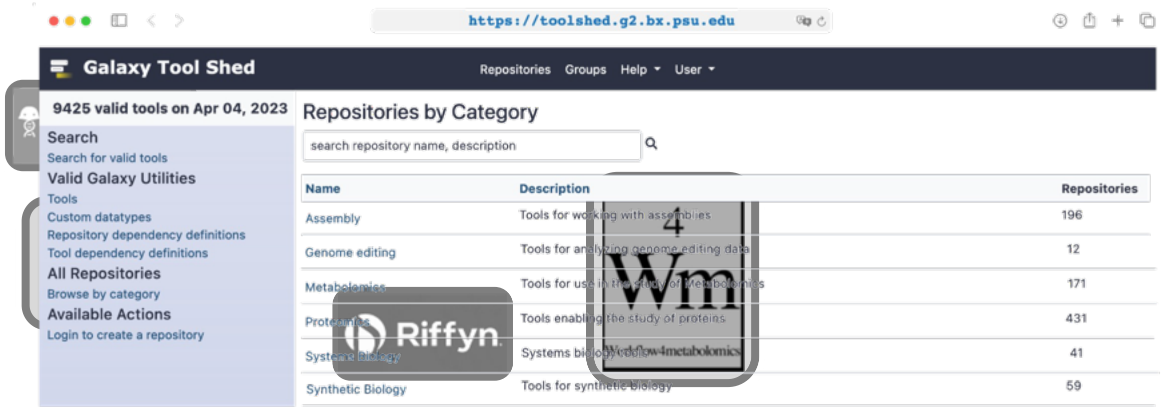
Galaxy COMMUNITY HUB

Galaxy Europe

Galaxy France

User-friendly
Widely used
Flexible
FAIR

— 9k+ tools
— 50 categories



https://toolshed.g2.bx.psu.edu

Galaxy Tool Shed Repositories Groups Help User

9425 valid tools on Apr 04, 2023

Search
- Search for valid tools

Valid Galaxy Utilities
- Tools
- Custom datatypes
- Repository dependency definitions
- Tool dependency definitions

All Repositories
- Browse by category

Available Actions
- Login to create a repository

Repositories by Category

search repository name, description

Name	Description	Repositories
Assembly	Tools for working with assemblies	196
Genome editing	Tools for analyzing genome editing data	12
Metabolomics	Tools for use in the study of Metabolomics	171
Proteomics	Tools enabling the study of proteins	431
Systems Biology	Systems biology tools	41
Synthetic Biology	Tools for synthetic biology	59

Riffyn

Workflow4metabolomics

Pathway Analysis Workflow »

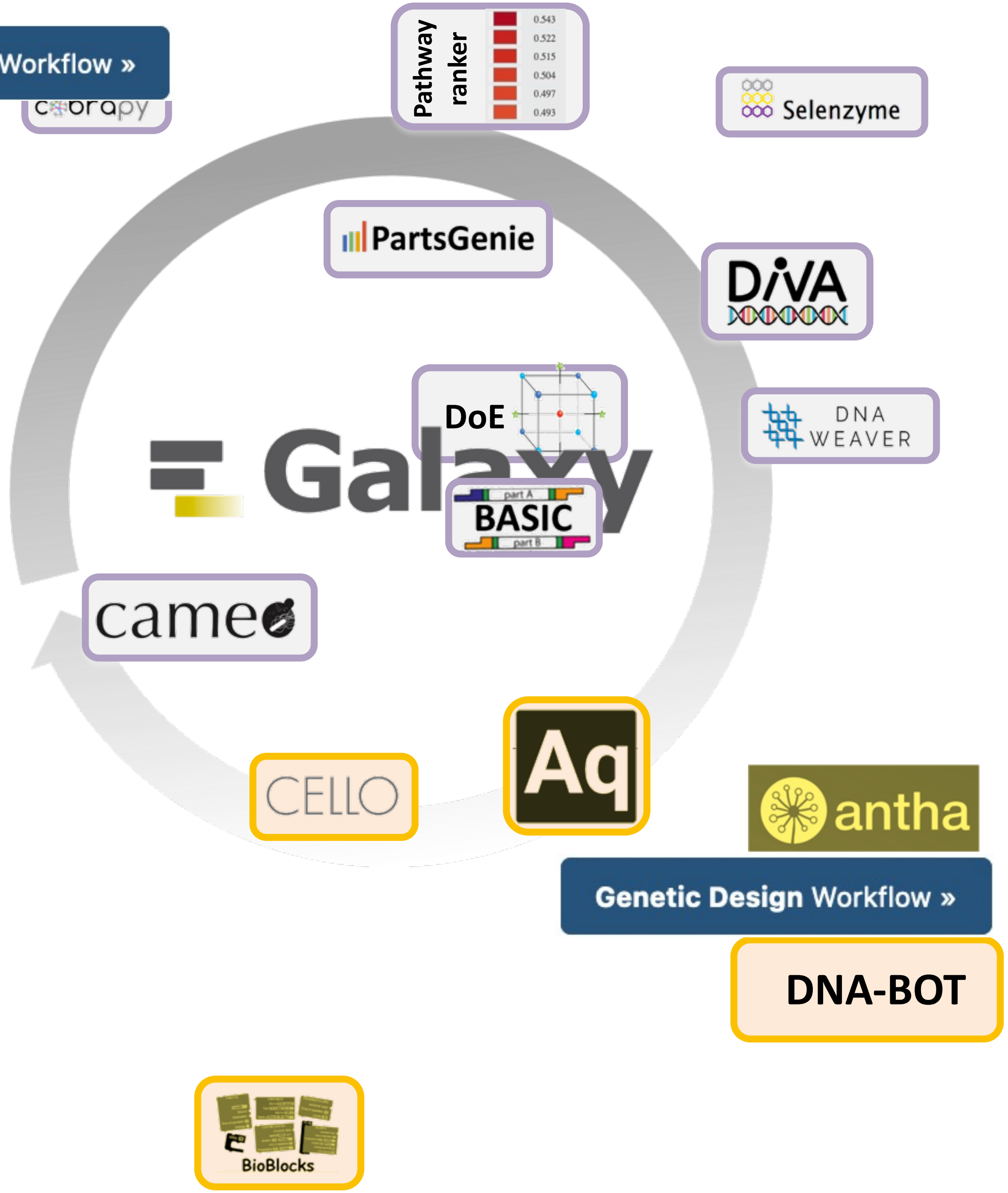
cobrapy

eQuilibrator

BNICE

RP2Path

Retrosynthesis Workflow »



Galaxy-SynBioCAD

Retrosynthesis Workflow »

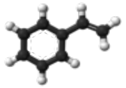
Google AI IBM Microsoft


galaxy-synbiocad.org

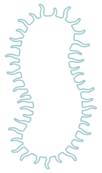
Reaction database

USPTO
MetaNetX

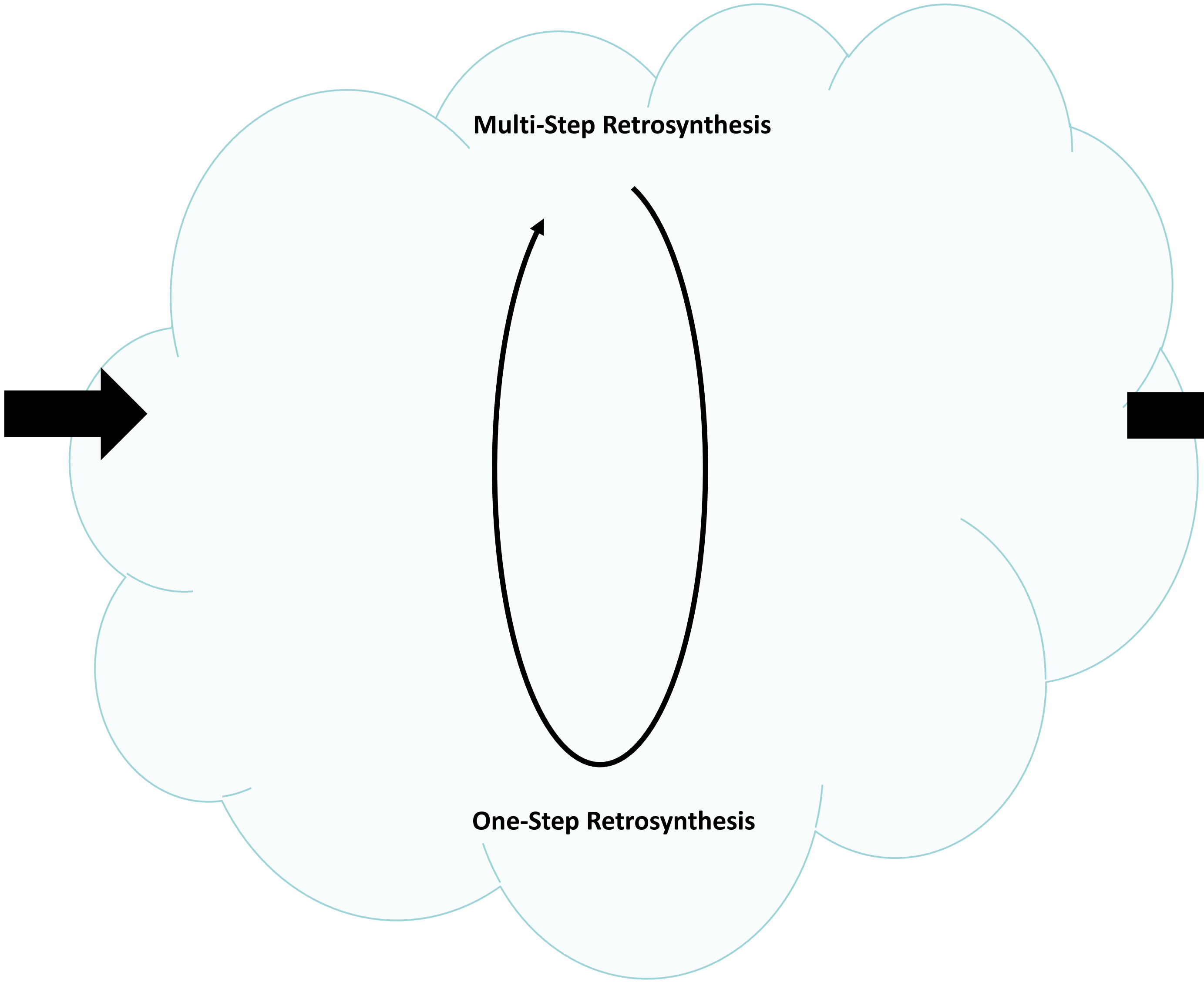
a target molecule


styrene

Available molecules

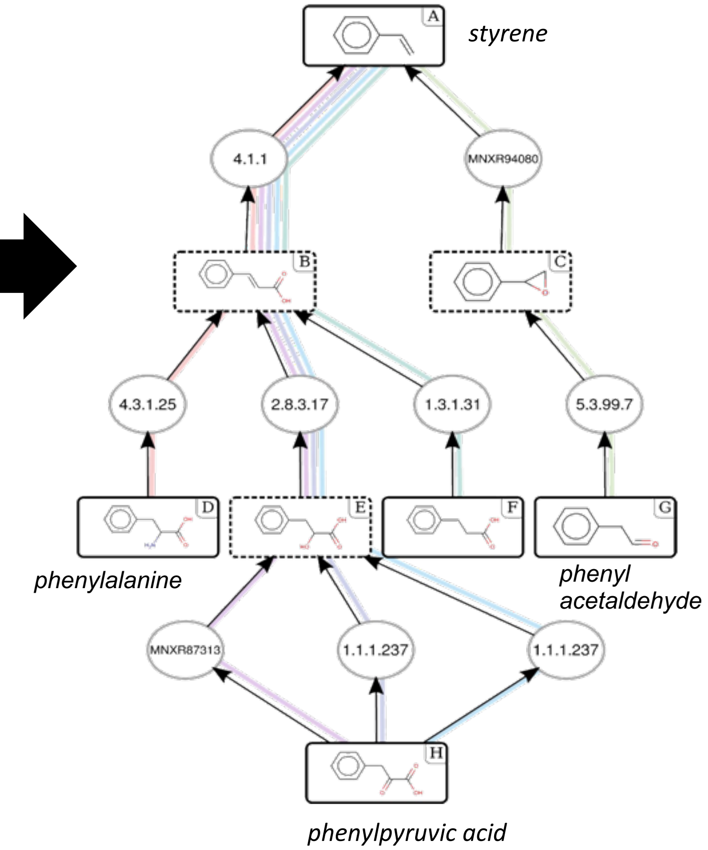


Available molecules in a host strain


Multi-Step Retrosynthesis

One-Step Retrosynthesis

A retrosynthesis tree/graph



Galaxy-SynBioCAD **Retrosynthesis Workflow »**

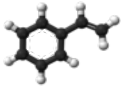
Google AI IBM Microsoft

galaxy-synbiocad.org

Reaction database

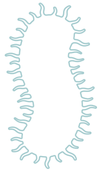
USPTO
MetaNetX

a target molecule

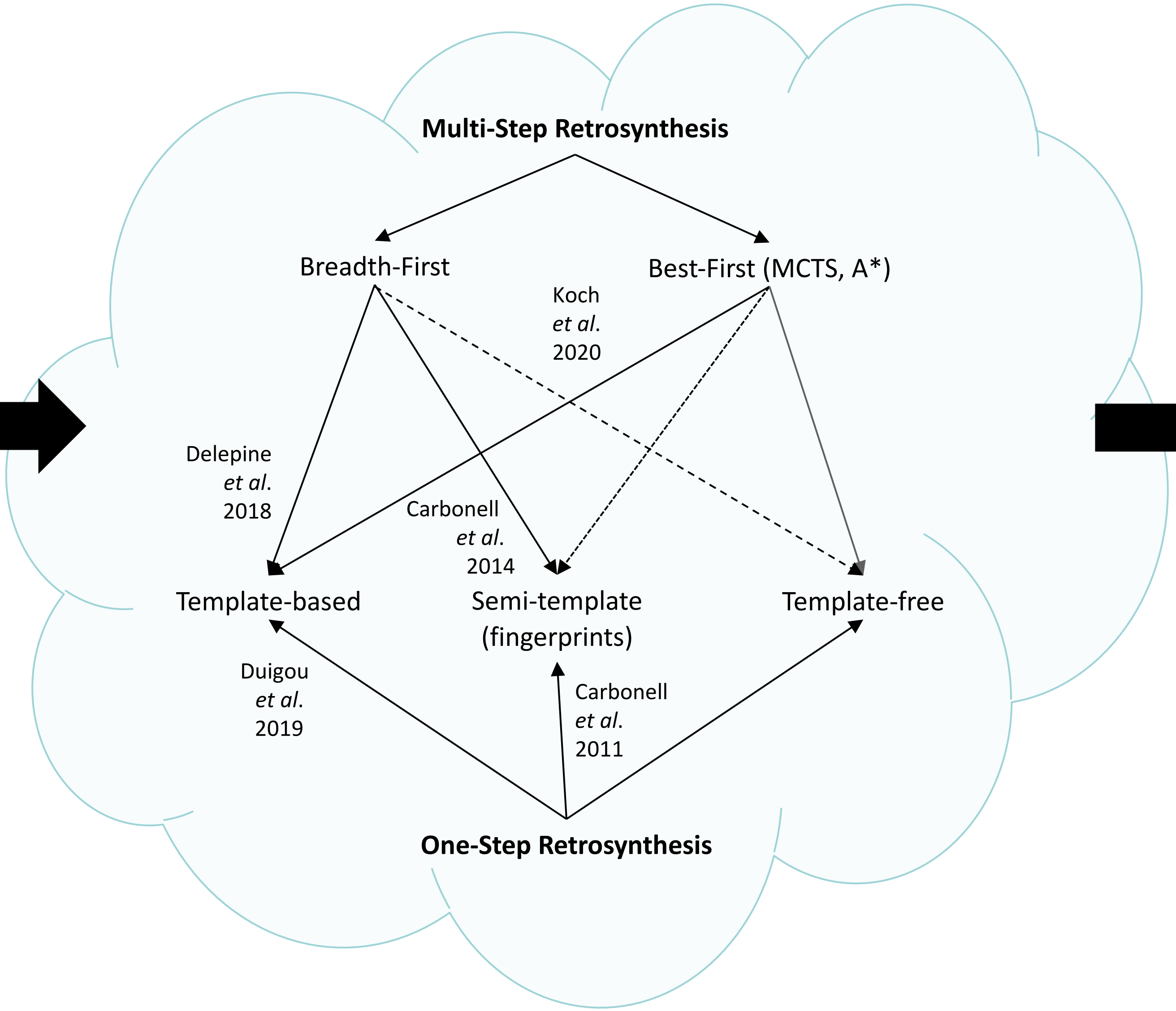


styrene

Available molecules



Available molecules in a host strain



Multi-Step Retrosynthesis

Breadth-First

Best-First (MCTS, A*)

Koch *et al.* 2020

Delepine *et al.* 2018

Carbonell *et al.* 2014

Template-based

Semi-template (fingerprints)

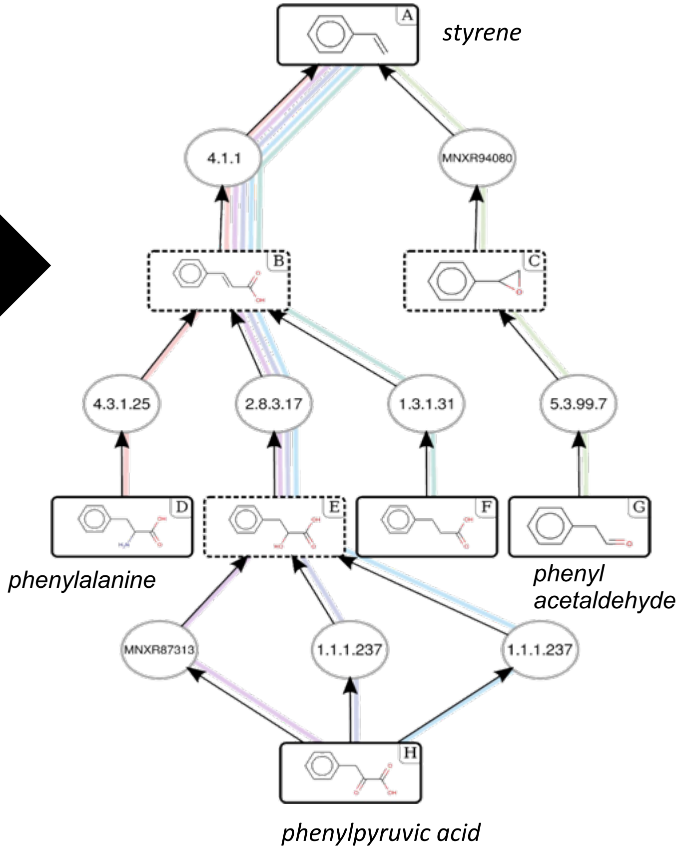
Template-free

Duigou *et al.* 2019

Carbonell *et al.* 2011

One-Step Retrosynthesis

A retrosynthesis tree/graph



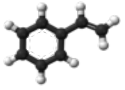
- Carbonell *et al.* ***BMC Sys. Biol***, 2011
- Carbonell *et al* ***Nucleic Acids Res***, 2014
- Delepine *et al.* ***Metabolic Eng.***, 2018
- Duigou *et al.*, ***Nucleic Acids Res.***, 2019
- Koch *et al.* ***ACS Synth Biol***, 2020

Galaxy-SynBioCAD **Retrosynthesis Workflow »**

Breadth First – Template Based

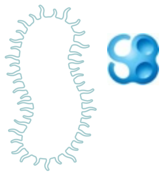
MetaNetX

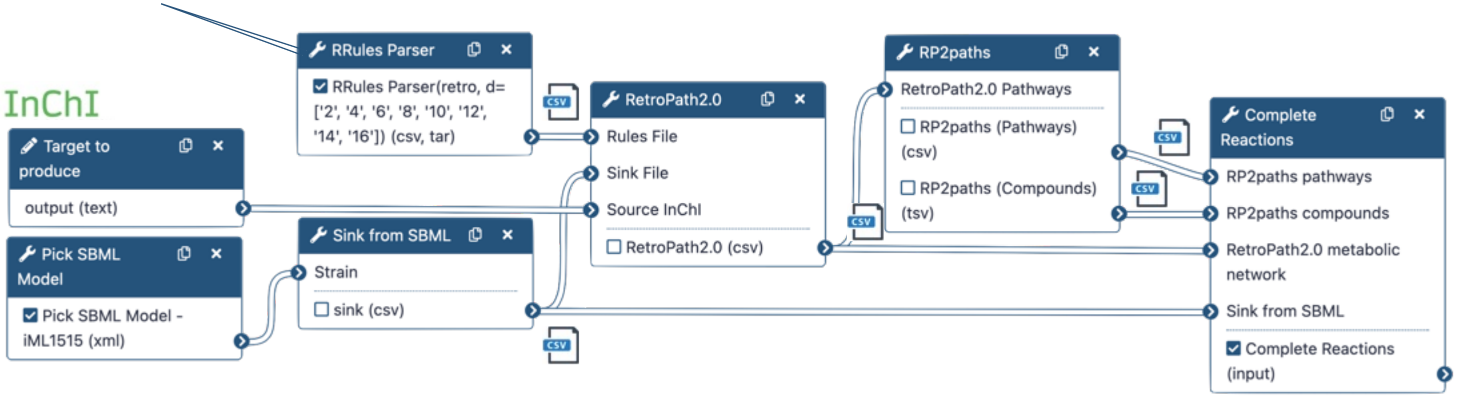
SMARTS



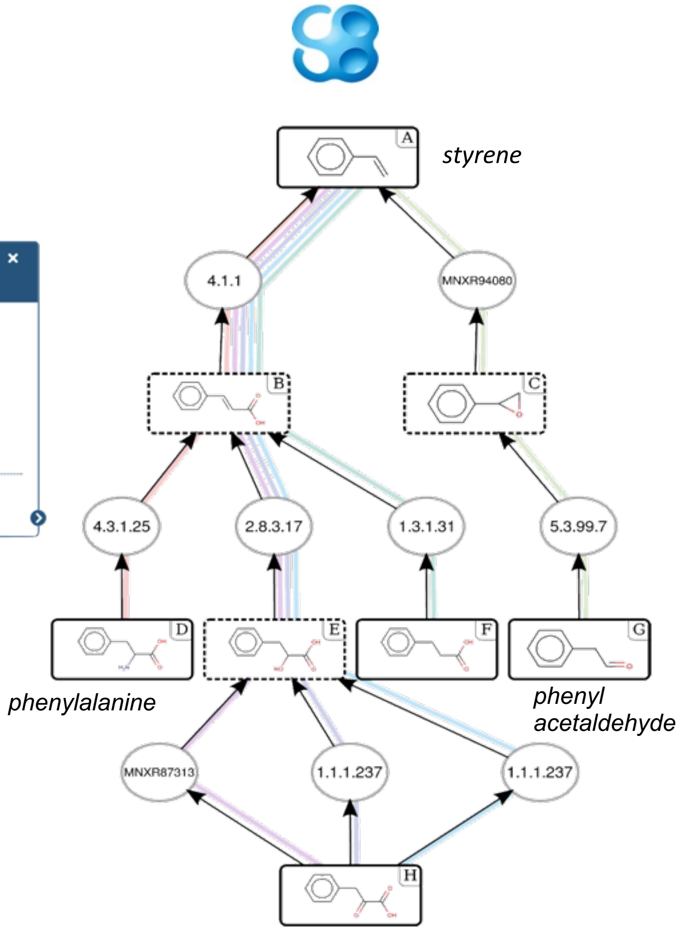
styrene

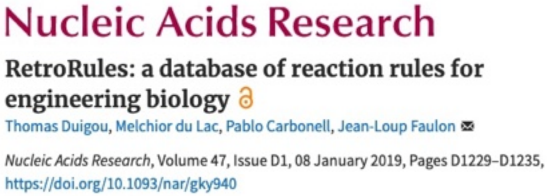
InChI





Available molecules in a host strain



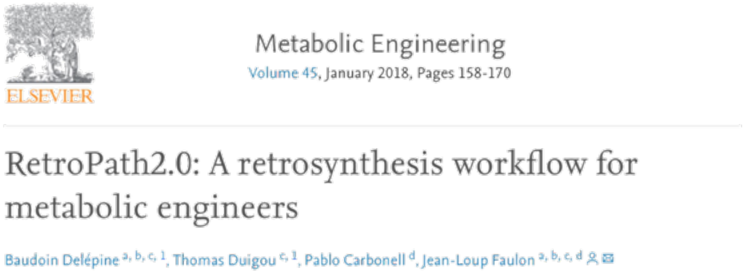


Nucleic Acids Research

RetroRules: a database of reaction rules for engineering biology

Thomas Duigou, Melchior du Lac, Pablo Carbonell, Jean-Loup Faulon

Nucleic Acids Research, Volume 47, Issue D1, 08 January 2019, Pages D1229–D1235, https://doi.org/10.1093/nar/gky940



ELSEVIER

Metabolic Engineering
Volume 45, January 2018, Pages 158-170

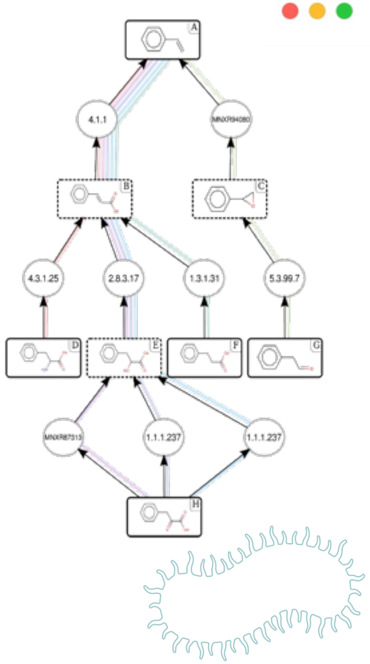
RetroPath2.0: A retrosynthesis workflow for metabolic engineers

Baudoin Delépine [a, b, c, 1], Thomas Duigou [c, 1], Pablo Carbonell [d], Jean-Loup Faulon [a, b, c, d]

Galaxy-SynBioCAD

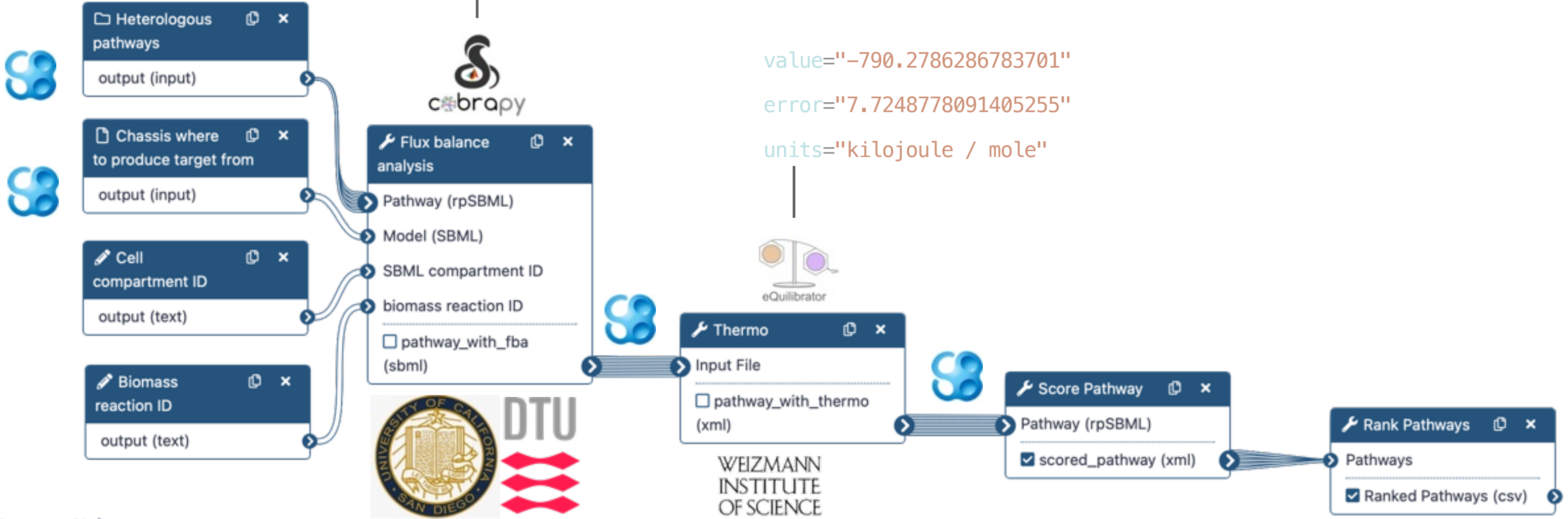
Pathway Analysis Workflow »

galaxy-synbiocad.org



value="1.6644333828047877"
units="mmol/gDW/hour"

value="-790.2786286783701"
error="7.7248778091405255"
units="kilojoule / mole"



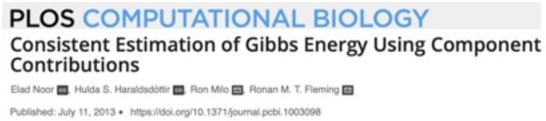

BMC Systems Biology

Software | Open Access | Published: 08 August 2013

COBRApy: COnstraints-Based Reconstruction and Analysis for Python

Ali Ebrahim, Joshua A Lerman, Bernhard O Palsson & Daniel R Hyduke

BMC Systems Biology **7**, Article number: 74 (2013) | Cite this article


PLOS COMPUTATIONAL BIOLOGY

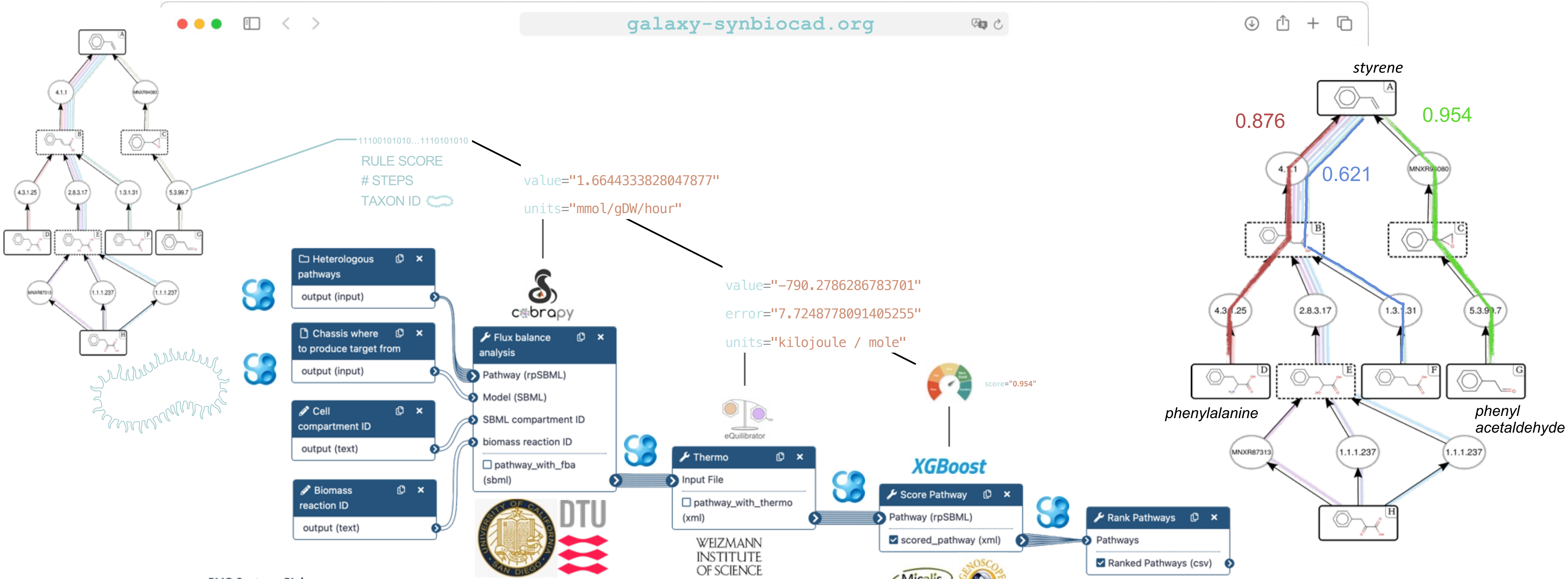
Consistent Estimation of Gibbs Energy Using Component Contributions

Elad Noor, Hulda S. Haraldsdóttir, Ron Milo, Ronan M. T. Fleming

Published: July 11, 2013 • https://doi.org/10.1371/journal.pcbi.1003098

Galaxy-SynBioCAD

Pathway Analysis Workflow »



value="1.6644333828047877"

units="mmol/gDW/hour"

value="-790.2786286783701"

error="7.7248778091405255"

units="kilojoule / mole"

score="0.954"

Heterologous pathways
output (input)

Chassis where to produce target from
output (input)

Cell compartment ID
output (text)

Biomass reaction ID
output (text)

Flux balance analysis
Pathway (rpSBML)
Model (SBML)
SBML compartment ID
biomass reaction ID
☐ pathway_with_fba (sbml)

Thermo
Input File
☐ pathway_with_thermo (xml)

Score Pathway
Pathway (rpSBML)
☑ scored_pathway (xml)

Rank Pathways
Pathways
☑ Ranked Pathways (csv)

BMC Systems Biology

Software | Open Access | Published: 08 August 2013

COBRApy: COnstraints-Based Reconstruction and Analysis for Python

Ali Ebrahim, Joshua A Lerman, Bernhard O Palsson & Daniel R Hyduke

BMC Systems Biology 7, Article number: 74 (2013) | Cite this article

PLOS COMPUTATIONAL BIOLOGY

Consistent Estimation of Gibbs Energy Using Component Contributions

Elad Noor, Hulda S. Haraldsdóttir, Ron Milo, Ronan M. T. Fleming

Published: July 11, 2013 • https://doi.org/10.1371/journal.pcbi.1003098

Nucleic Acids Research

XTMS: pathway design in an eXTended metabolic space

Pablo Carbonell, Pierre Parutto, Joan Herisson, Shashi Bhushan Pandit, Jean-Loup Faulon

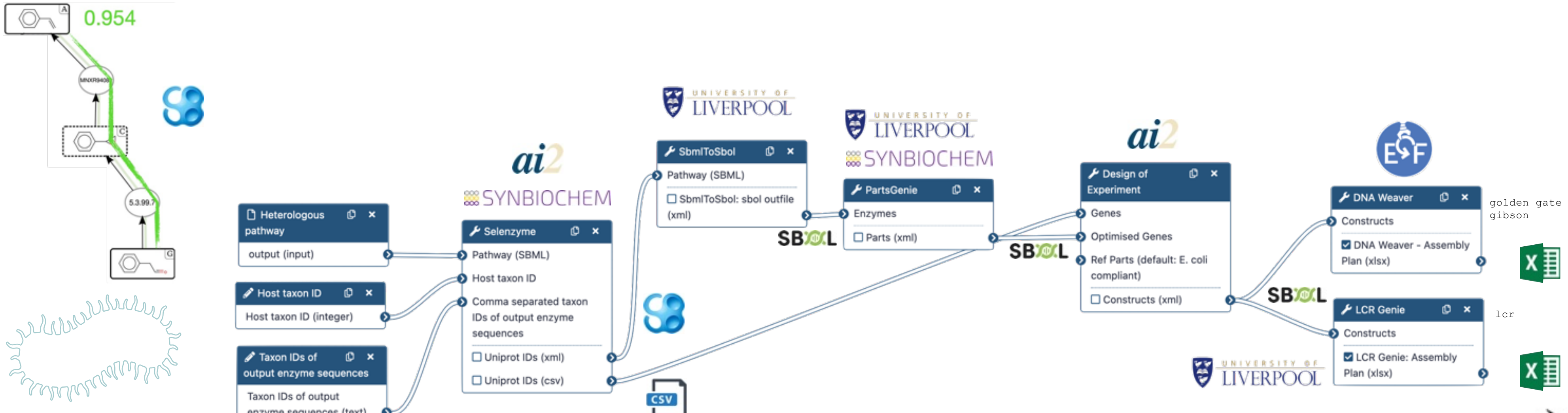
Nucleic Acids Research, Volume 42, Issue W1, 1 July 2014, Pages W389–W394, https://doi.org/10.1093/nar/gku362

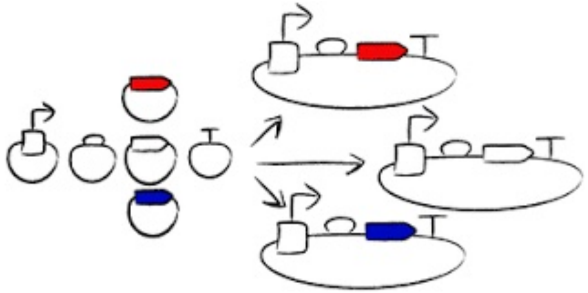
phenylpyruvic acid

Galaxy-SynBioCAD

Genetic Design Workflow »

galaxy-synbiocad.org





OptBioDes: optimal design for the SynBio toolchain

P. Carbonell, R. Breitling, J.-L. Faulon, and the SYNBIOCHEM team

IWBDA, Cambridge (UK), 2019

Selenzyme: enzyme selection tool for pathway design

Pablo Carbonell, Jerry Wong, Neil Swainston, Eriko Takano, Nicholas J Turner, Nigel S Scrutton, Douglas B Kell, Rainer Breitling, Jean-Loup Faulon

Bioinformatics, Volume 34, Issue 12, 15 June 2018, Pages 2153–2154,

PartsGenie: an integrated tool for optimizing and sharing synthetic biology parts FREE

Neil Swainston, Mark Dunstan, Adrian J Jervis, Christopher J Robinson, Pablo Carbonell, Alan R Williams, Jean-Loup Faulon, Nigel S Scrutton, Douglas B Kell

Bioinformatics, Volume 34, Issue 13, 01 July 2018, Pages 2327–2329,

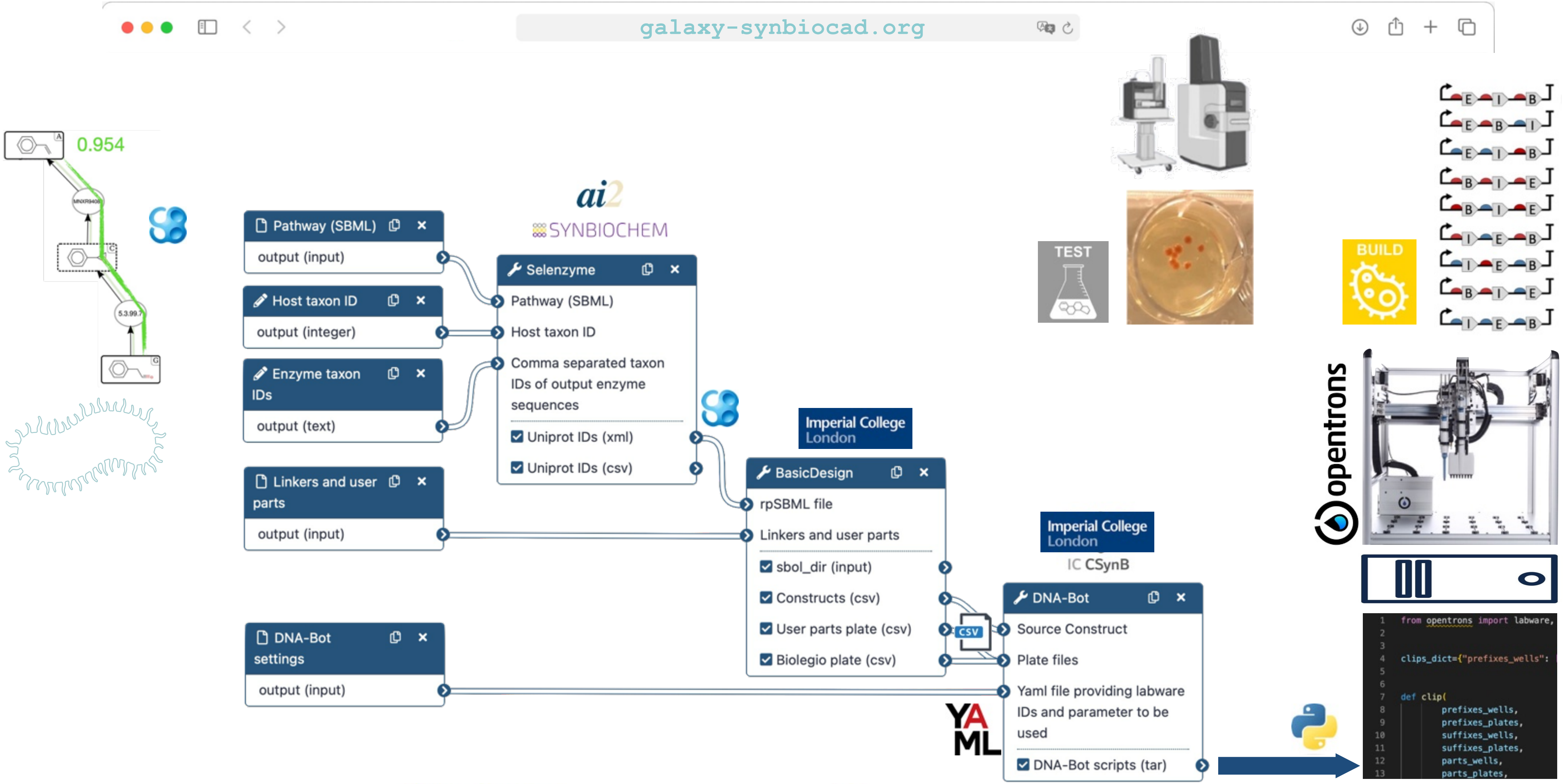
DNA-Weaver: optimal DNA assembly strategies via supply networks and shortest-path algorithms

V. Zulkower and S. Rosser

IWBDA, Cambridge (UK), 2019

Galaxy-SynBioCAD

Genetic Design (BASIC) Workflow »



```
from opentrons import labware,


clips_dict={"prefixes_wells":


def clip(
        prefixes_wells,
        prefixes_plates,
        suffixes_wells,
        suffixes_plates,
        parts_wells,
        parts_plates,
```

ACS SyntheticBiology Letter

pubs.acs.org/synthbio

BASIC: A New Biopart Assembly Standard for Idempotent Cloning Provides Accurate, Single-Tier DNA Assembly for Synthetic Biology

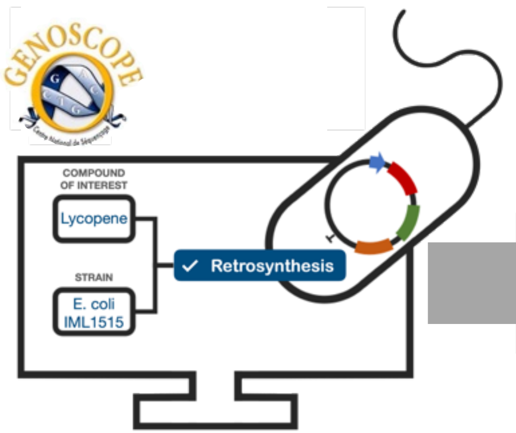
Marko Storch,[†,‖] Arturo Casini,[†,‖] Ben Mackrow,[†] Toni Fleming,[‡] Harry Trewhitt,[†] Tom Ellis,[§] and Geoff S. Baldwin[*,†]

DNA-BOT: a low-cost, automated DNA assembly platform for synthetic biology

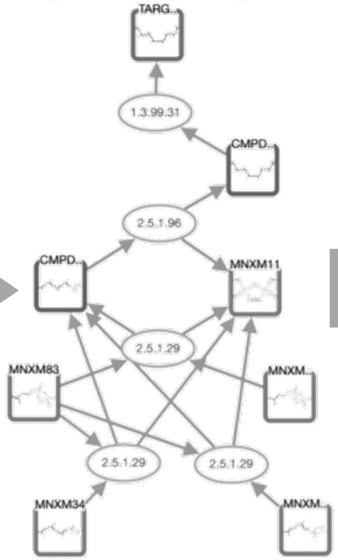
Marko Storch, Matthew C Haines, Geoff S Baldwin Author Notes

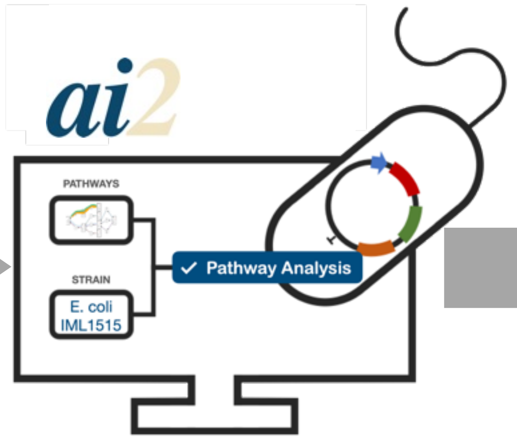
Synthetic Biology, Volume 5, Issue 1, 2020, ysaa010,

Benchmarking with lycopene production in E. coli

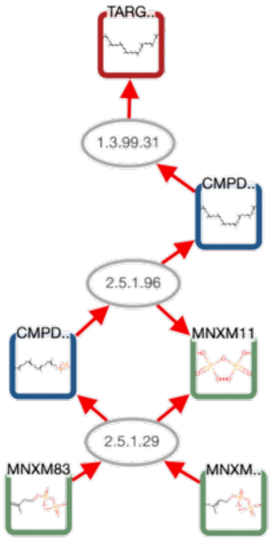


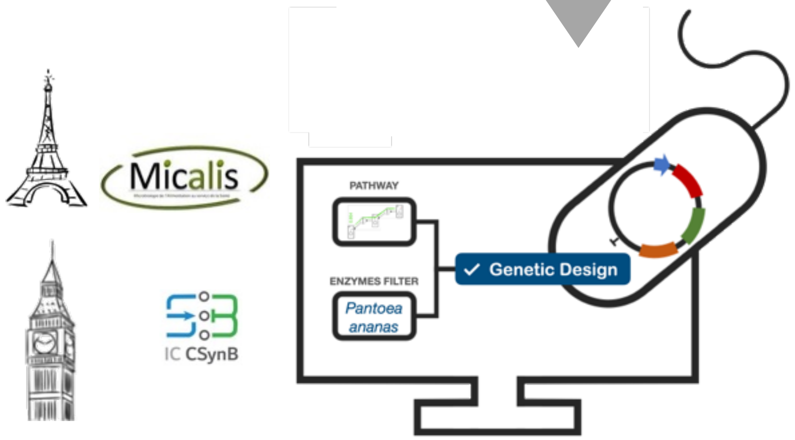
9 pathways



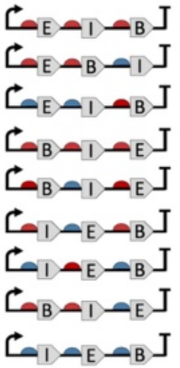


top ranked pathway



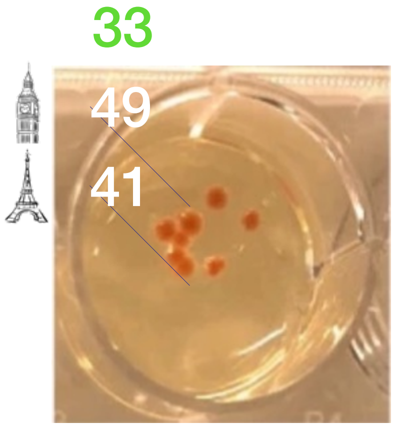


88 constructs



Varying promoter, RBS, gene order

G6: 4.389 mg/gDCW

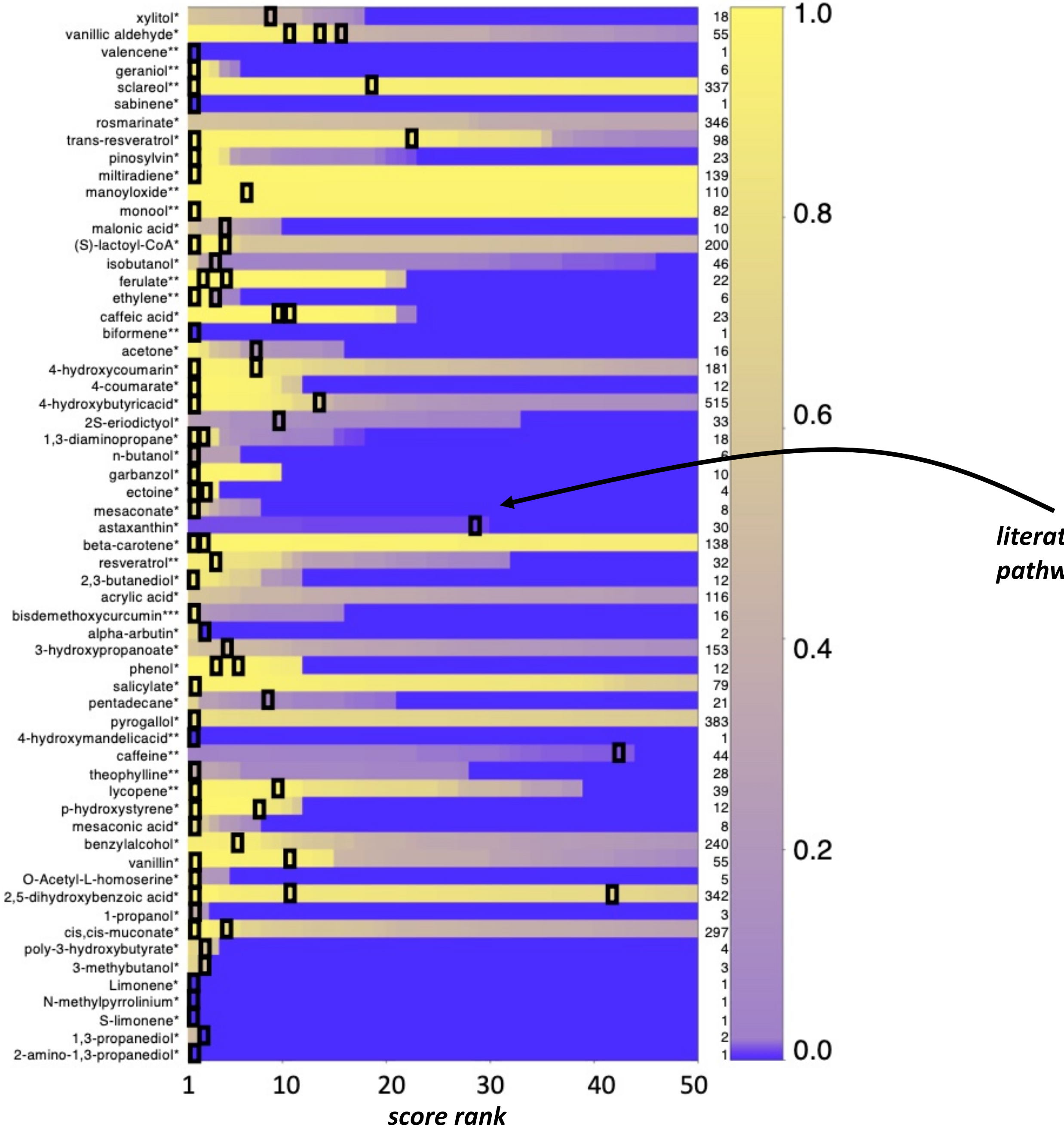


```
from opentrons import labware,


clips_dict={"prefixes_wells":


def clip(
            prefixes_wells,
            prefixes_plates,
            suffixes_wells,
            suffixes_plates,
            parts_wells,
            parts_plates,
```

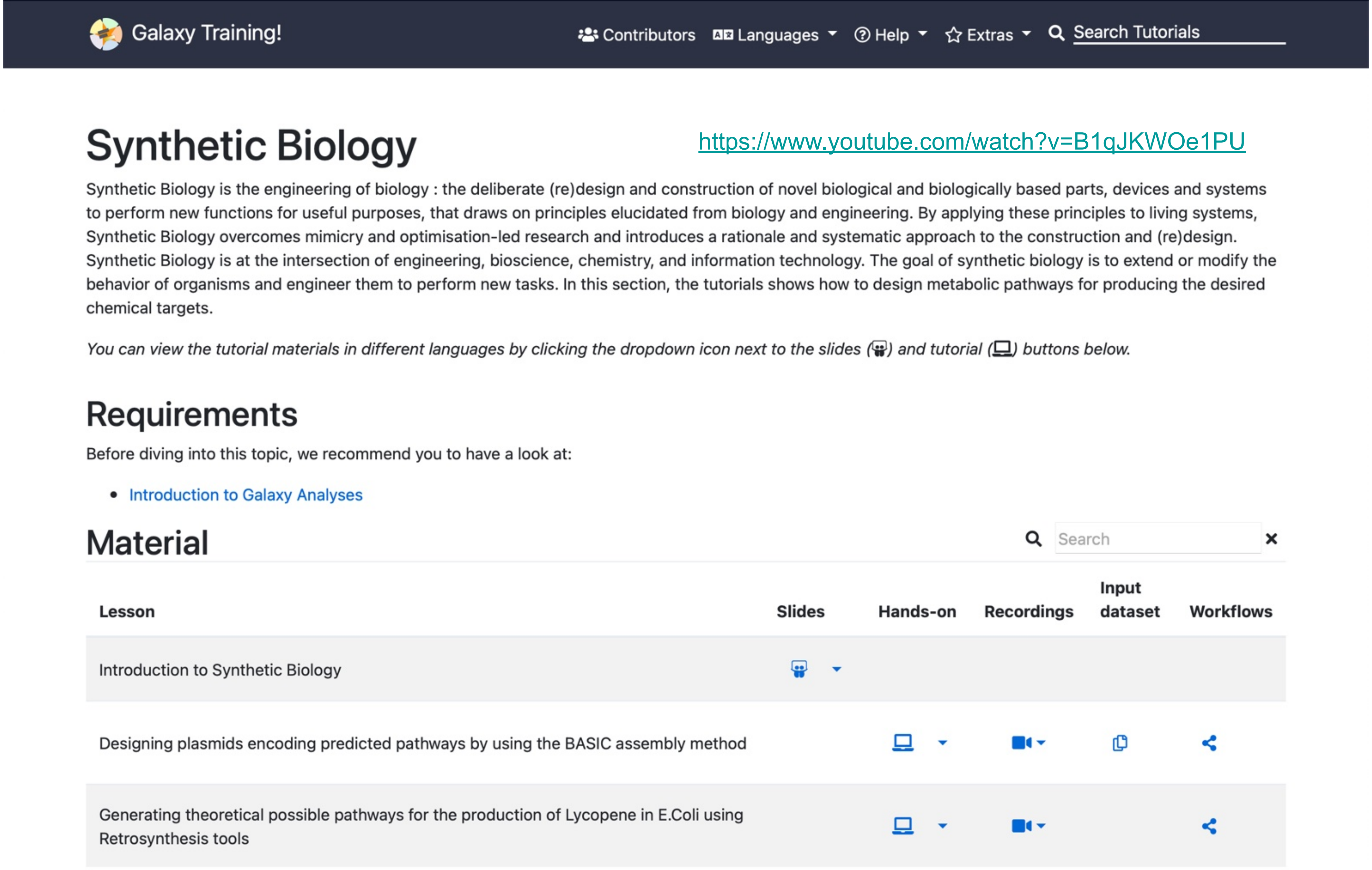
Benchmarking with literature and expert users

- Compiled a list of 80 literature pathways for various compounds and strains
- Run workflow of the same compounds and strains, generated ~8000 pathways
- Rank each pathway using the ML scored trained on expert validation trial

➢ 83% (94%) success rate in retrieving the literature pathways among the top 10 (50) workflow generated pathways



literature pathway



https://www.youtube.com/watch?v=B1qJKWOe1PU

Synthetic Biology

Synthetic Biology is the engineering of biology : the deliberate (re)design and construction of novel biological and biologically based parts, devices and systems to perform new functions for useful purposes, that draws on principles elucidated from biology and engineering. By applying these principles to living systems, Synthetic Biology overcomes mimicry and optimisation-led research and introduces a rationale and systematic approach to the construction and (re)design. Synthetic Biology is at the intersection of engineering, bioscience, chemistry, and information technology. The goal of synthetic biology is to extend or modify the behavior of organisms and engineer them to perform new tasks. In this section, the tutorials shows how to design metabolic pathways for producing the desired chemical targets.

You can view the tutorial materials in different languages by clicking the dropdown icon next to the slides (▣) and tutorial (▭) buttons below.

Requirements

Before diving into this topic, we recommend you to have a look at:

- Introduction to Galaxy Analyses

Material

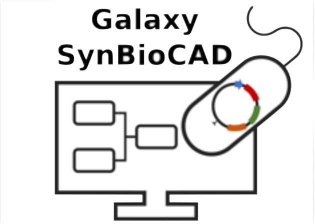
Lesson	Slides	Hands-on	Recordings	Input dataset	Workflows
Introduction to Synthetic Biology	▣ ▾				
Designing plasmids encoding predicted pathways by using the BASIC assembly method		▭ ▾	■◂ ▾	⧉	⋖
Generating theoretical possible pathways for the production of Lycopene in E.Coli using Retrosynthesis tools		▭ ▾	■◂ ▾		⋖

SynBioCAD tools can be connected to many others in the ToolShed

GitHub

SynBioPython

Galaxy



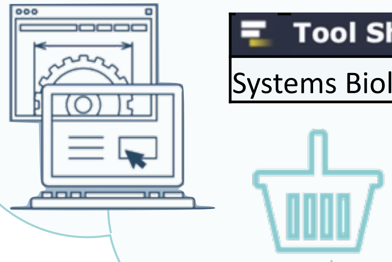
Tool Shed	Nbr Tools
Synthetic Biology	59

DATA AND
PROCESS MANAGMENT

ACTIVE LEARNING

Tool Shed	Nbr Tools
Machine Learning	62

DESIGN TOOLS STRAIN DESIGN



Tool Shed	Nbr Tools
Systems Biology	41

TEST DATA ACQUISITION

Tool Shed	Nbr Tools
Flow cytometry	45
Proteomics	431
Metabolomics	171

ROBOT DRIVERS

nature communications

Explore content ⌄ About the journal ⌄ Publish with us ⌄

nature > nature communications > articles > article

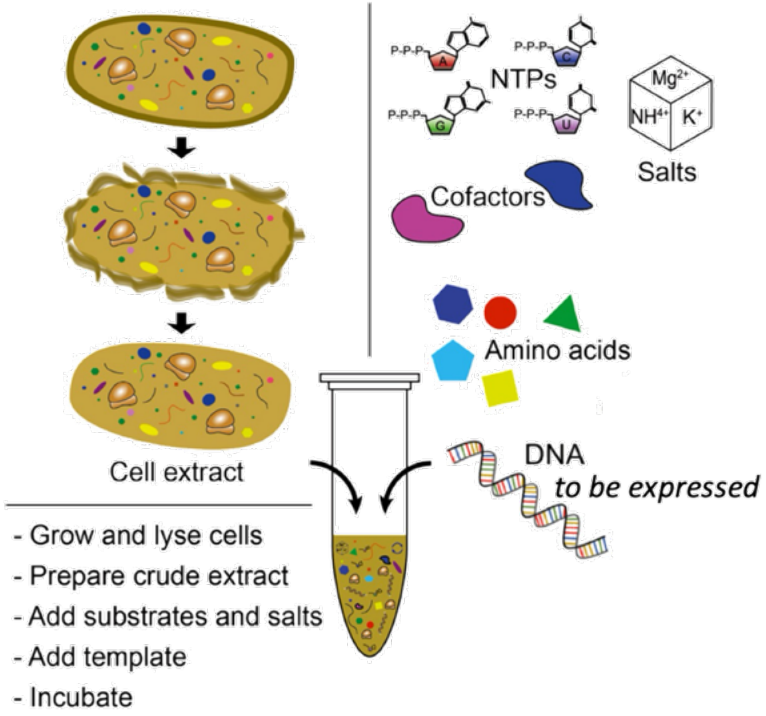
Article | Open Access | Published: 29 August 2022

The automated Galaxy-SynBioCAD pipeline for synthetic biology design and engineering

Joan Hérisson, Thomas Duigou, Melchior du Lac, Kenza Bazi-Kabbaj, Mahnaz Sabeti Azad, Gizem Buldum, Olivier Telle, Yorgo El Moubayed, Pablo Carbonell, Neil Swainston, Valentin Zulkower, Manish Kushwaha, Geoff S. Baldwin & Jean-Loup Faulon ✉

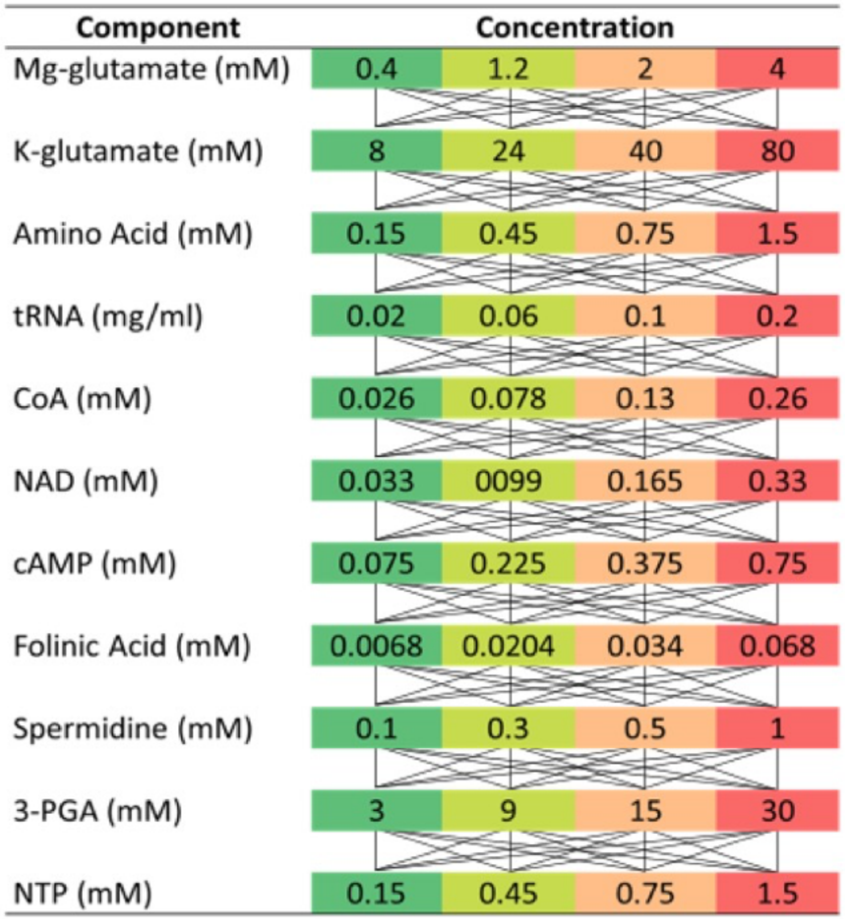
Active learning to optimize productivity in cell-free systems

Lysate-based cell-free systems (TXTL)



- Grow and lyse cells
- Prepare crude extract
- Add substrates and salts
- Add template
- Incubate

Reference composition
Sun Z.Z. *et al.* *J. Vis. Exp.* 2013



Component	Concentration			
Mg-glutamate (mM)	0.4	1.2	2	4
K-glutamate (mM)	8	24	40	80
Amino Acid (mM)	0.15	0.45	0.75	1.5
tRNA (mg/ml)	0.02	0.06	0.1	0.2
CoA (mM)	0.026	0.078	0.13	0.26
NAD (mM)	0.033	0099	0.165	0.33
cAMP (mM)	0.075	0.225	0.375	0.75
Folinic Acid (mM)	0.0068	0.0204	0.034	0.068
Spermidine (mM)	0.1	0.3	0.5	1
3-PGA (mM)	3	9	15	30
NTP (mM)	0.15	0.45	0.75	1.5

Combinatorial space = 4^{11}
= **4 194 304** compositions

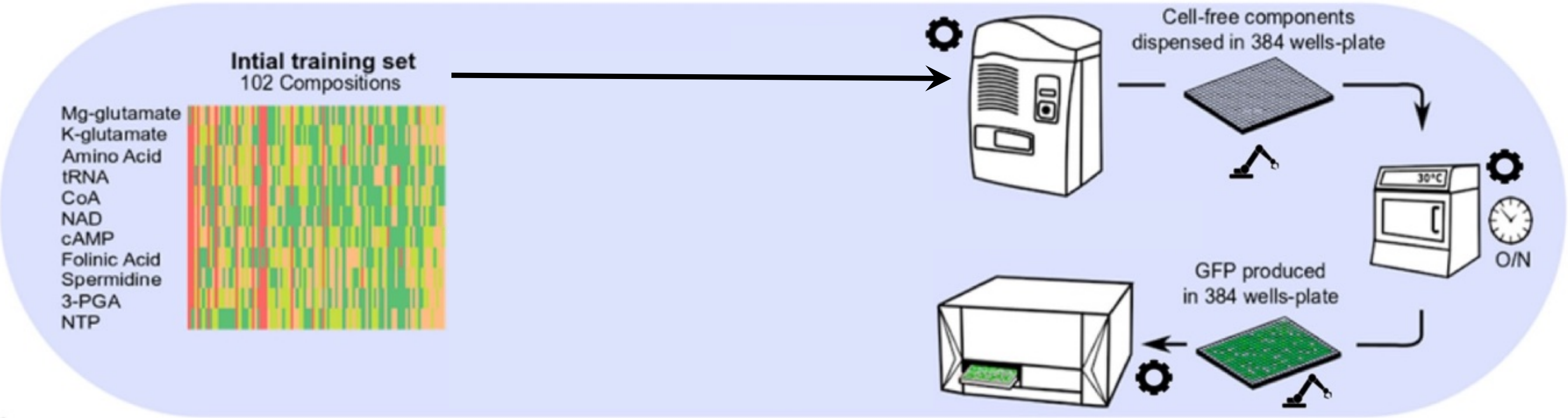
- Can we improve protein production without increasing the price of cell-free reaction?
- Can we provide efficient predictions of protein production *in vitro*?
- Can we highlight the critical parameters involve in protein production *in vitro*?

Active learning to optimize productivity in cell-free systems

- **Set up an initial batch sampling the space of possible compositions**

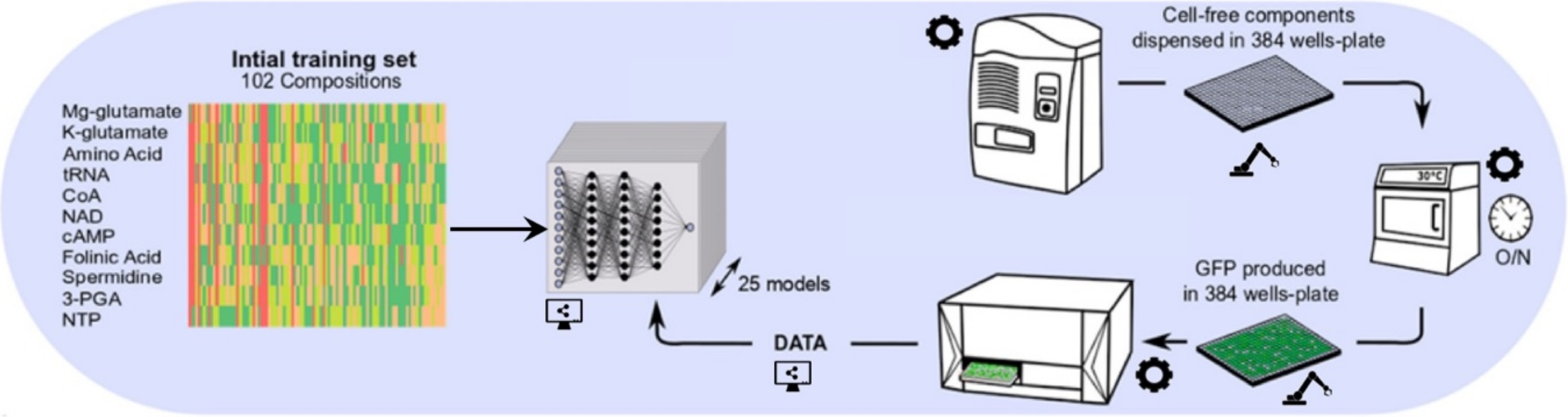
Active learning to optimize productivity in cell-free systems

- Set up an initial batch sampling the space of possible compositions
- **Measure yield level though fluorescence**



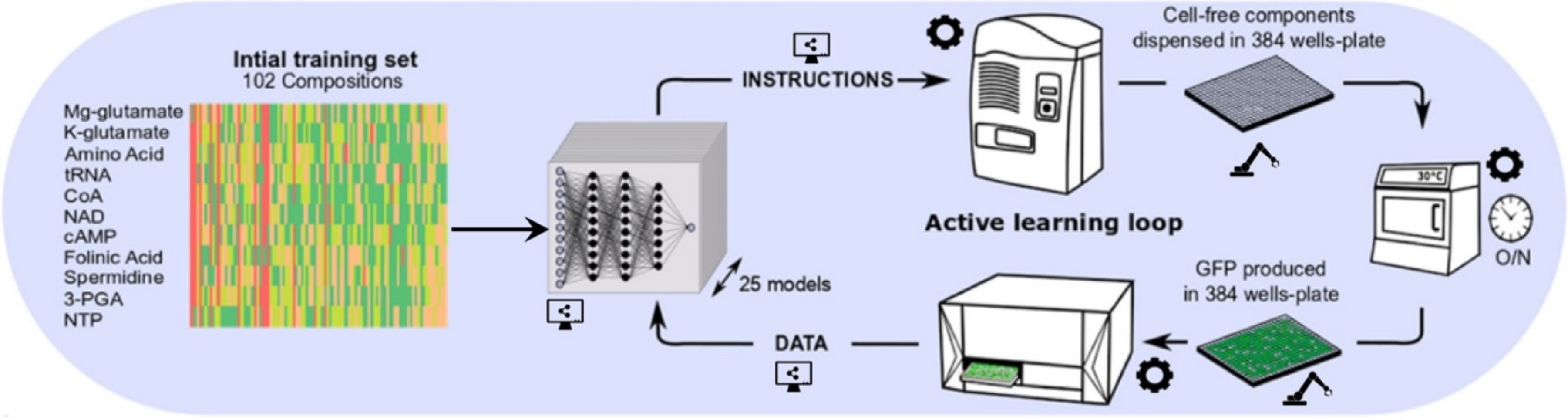
Active learning to optimize productivity in cell-free systems

- Set up an initial batch sampling the space of possible compositions
- Measure yield level though fluorescence
- **Develop a Neural Network model predicting yield from composition**



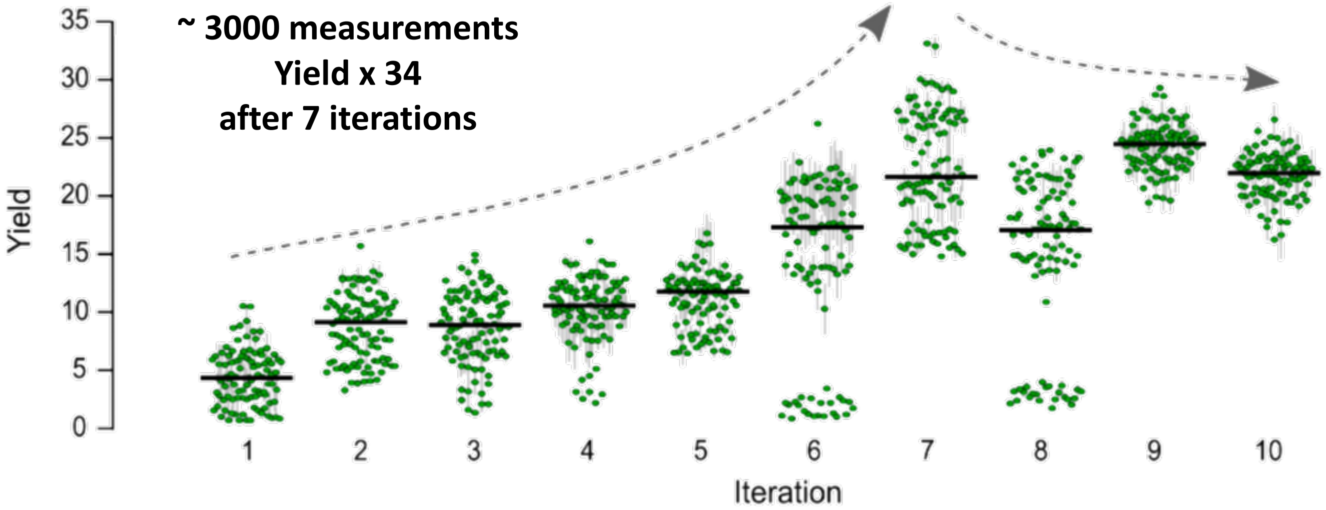
Active learning to optimize productivity in cell-free systems

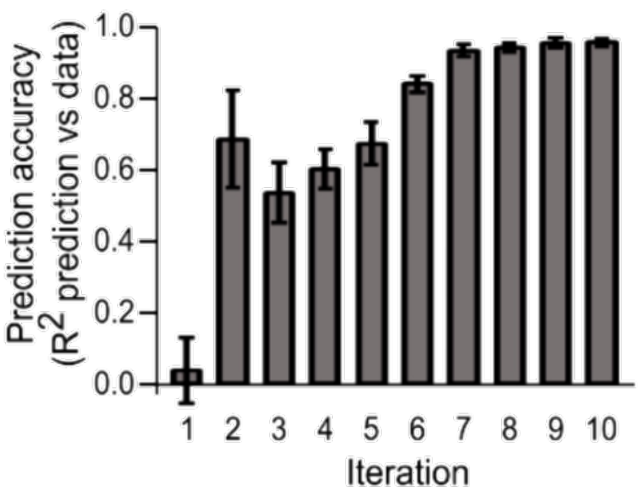
- Set up an initial batch sampling the space of possible compositions
- Measure yield level though fluorescence
- Develop a Neural Network model predicting yield from composition
- **Use the model to predict the yield for each composition not yet tested**
- **Select next batch of compositions to be measured based on exploitation vs. exploration (UCB formula)**
- **Repeat**

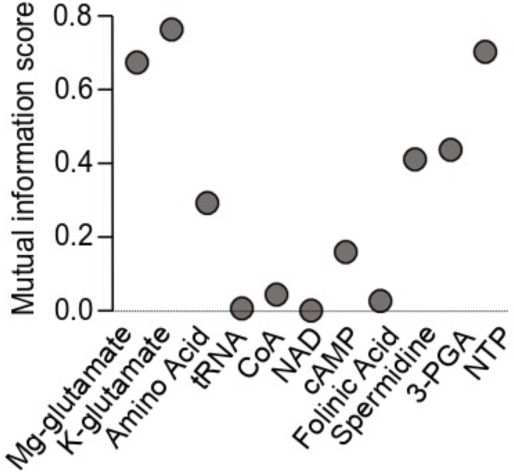


* UCB: Take the compositions having the top values $\mu + \sqrt{2}\,\sigma$, where μ is the mean predicted yield and σ the standard dev
Picking high μ values favor exploitation picking high σ values favor exploration

Active learning to optimize productivity in cell-free systems

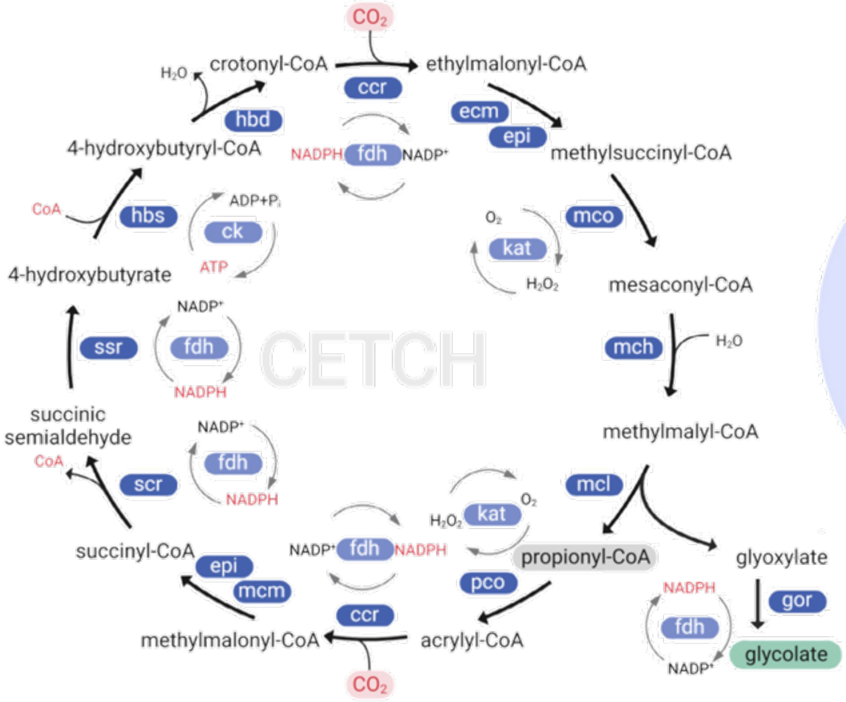

~ 3000 measurements
Yield x 34
after 7 iterations

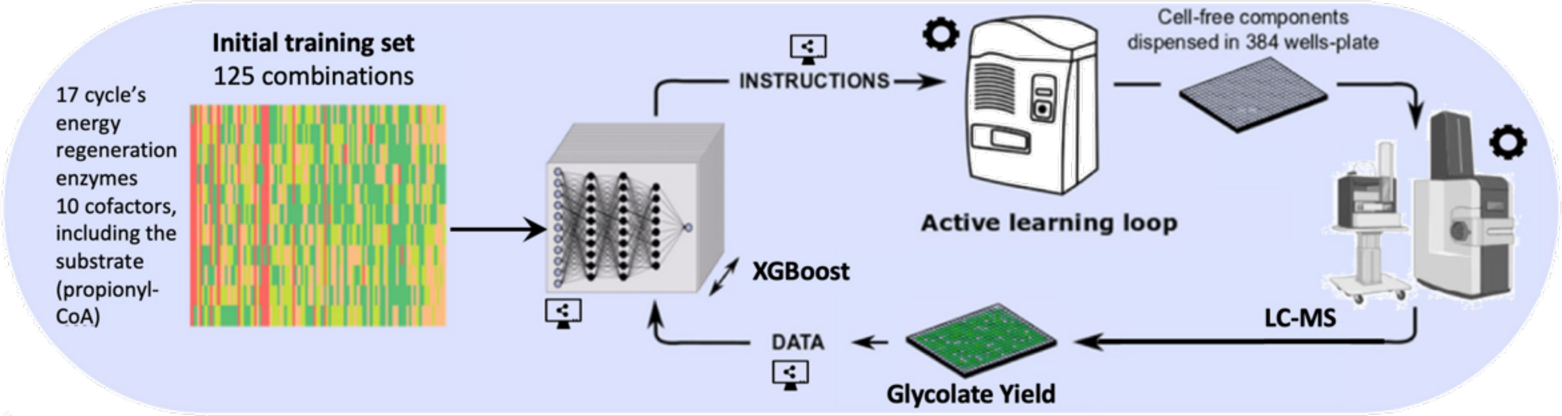

5-Fold cross validation


Dependence between the yield
and the components concentration

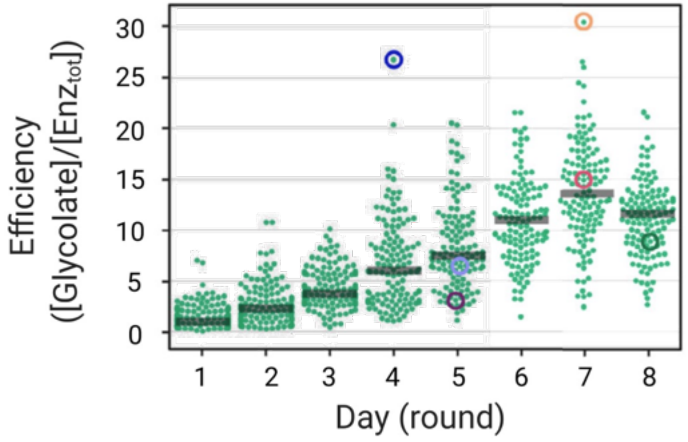
- Borkowski O, *et al.* ***Nature Communications***, 11: 1872, 2020.

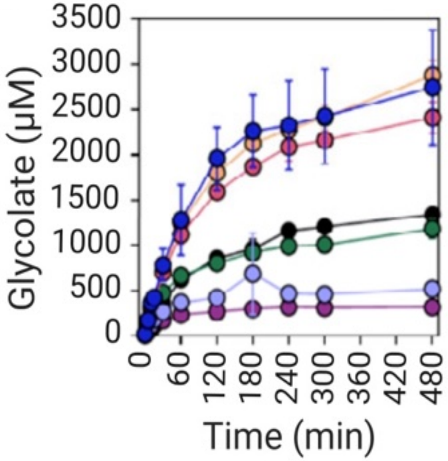
Active learning to optimize metabolic pathways in cell-free systems





17 cycle's energy regeneration enzymes 10 cofactors, including the substrate (propionyl-CoA)





Top 3 best producers

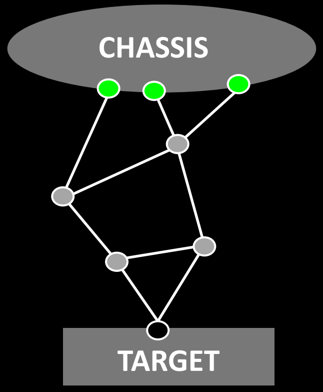
Control (CETCH5.4)

Randomly selected producers

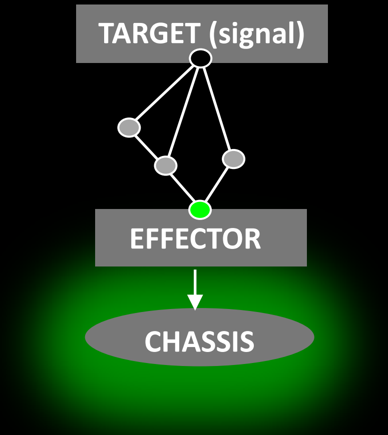
6x more efficient than the best in vitro CO_2-fixing system described to date (CETCH 5.4 , Schwander *et al. Science* 2016)

- Pandi A., *et al.* ***Nature Communications***, 13: 3876, 2022.

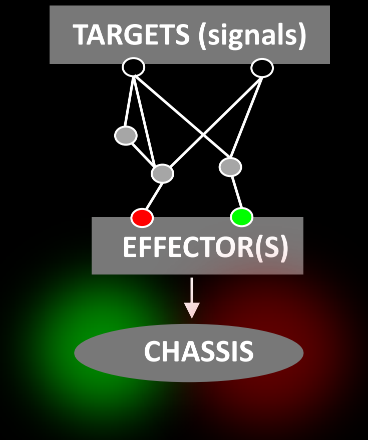
Engineering metabolic and genetic networks in vitro (cell-free) and in vivo (E. coli)



Metabolic engineering



Biosensor engineering



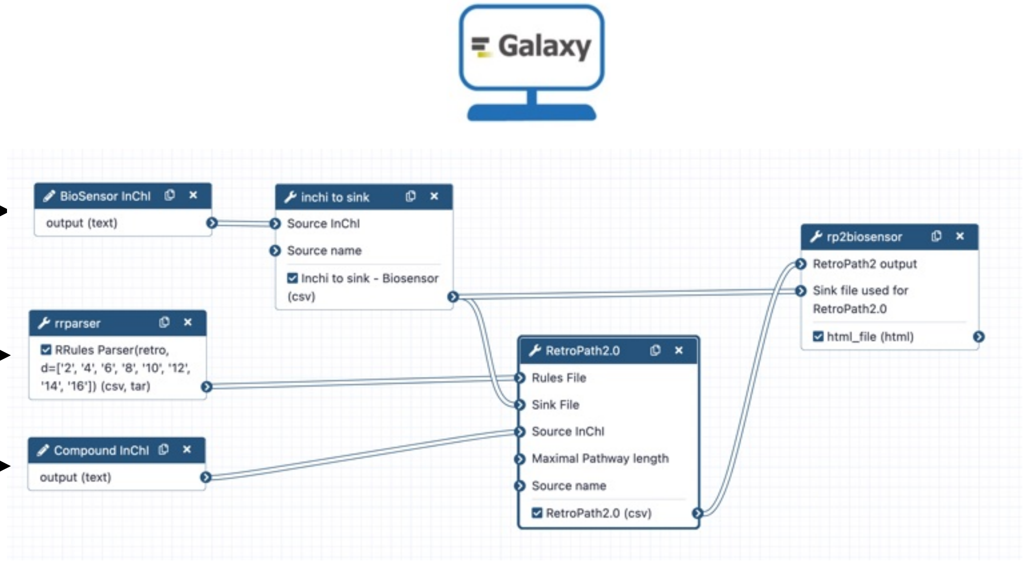
Information processing
biocomputing

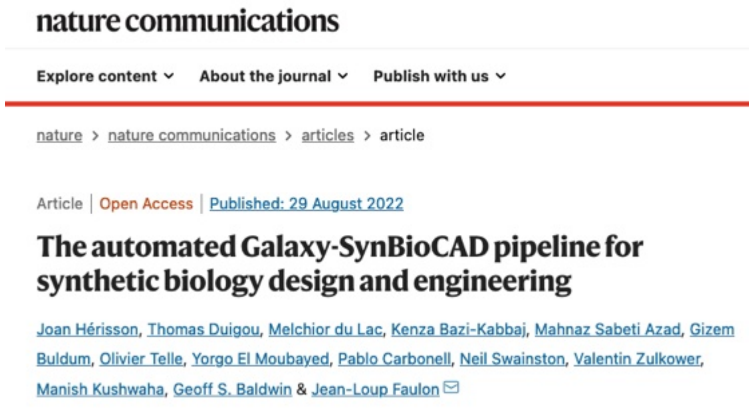
Designing biosensing circuits

Detectable Metabolites DB →
(Koch et al. DIB 2018)

Transformations rules →
RetroRules.org
(Duigou et al. NAR, 2018)

Molecule to detect →





nature communications

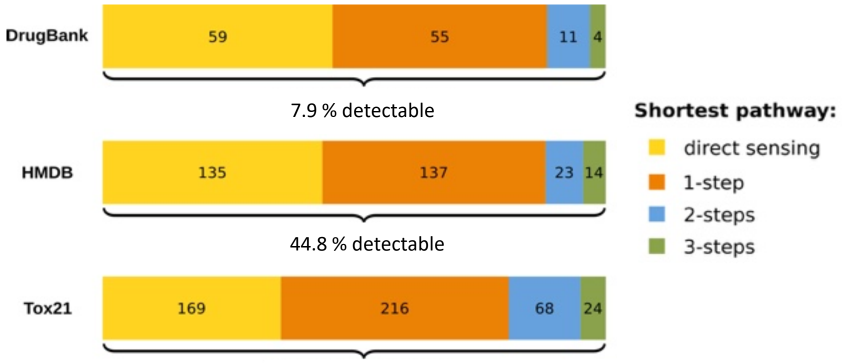
Explore content ⌄ About the journal ⌄ Publish with us ⌄

nature > nature communications > articles > article

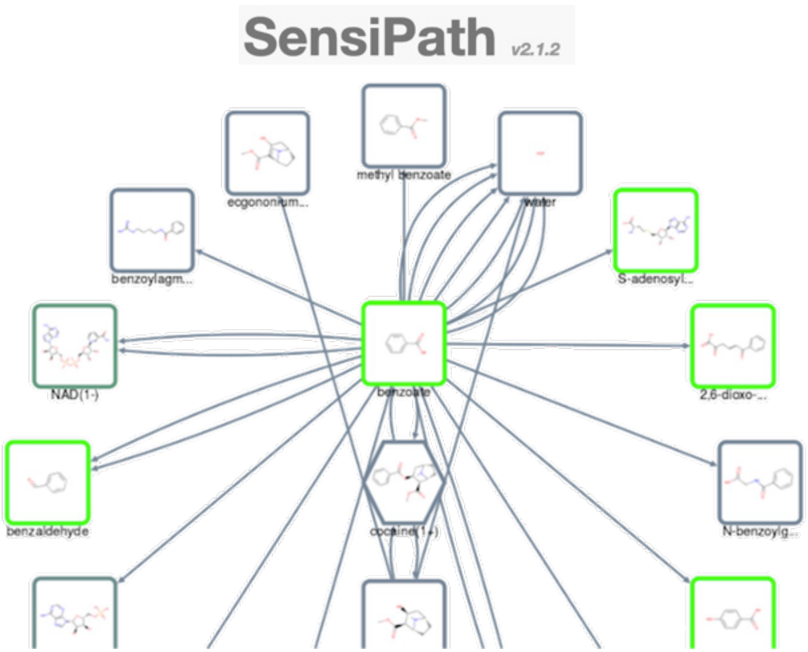
Article | Open Access | Published: 29 August 2022

The automated Galaxy-SynBioCAD pipeline for synthetic biology design and engineering

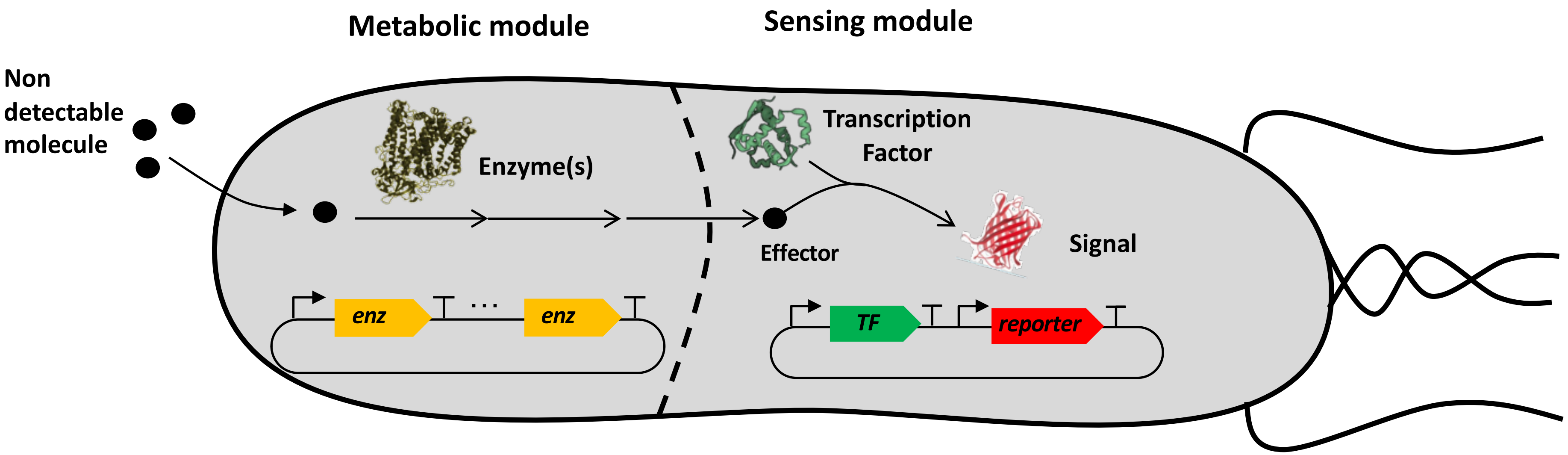
Joan Hérisson, Thomas Duigou, Melchior du Lac, Kenza Bazi-Kabbaj, Mahnaz Sabeti Azad, Gizem Buldum, Olivier Telle, Yorgo El Moubayed, Pablo Carbonell, Neil Swainston, Valentin Zulkower, Manish Kushwaha, Geoff S. Baldwin & Jean-Loup Faulon ✉

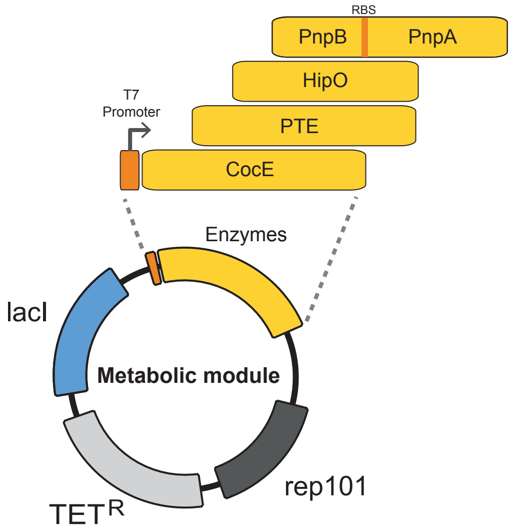


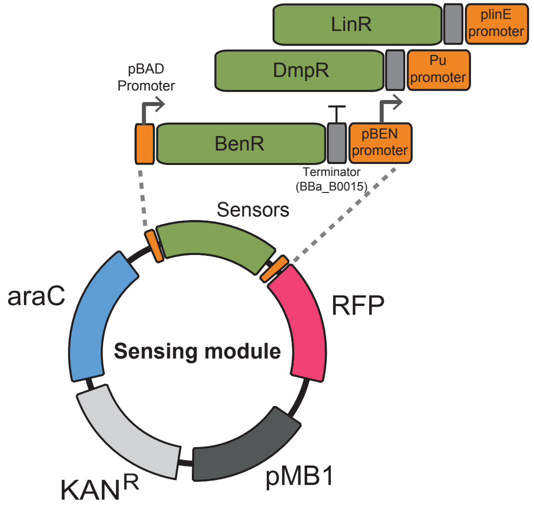
- Delepine B, et al. ***NAR*** 2016



Engineering biosensing circuits

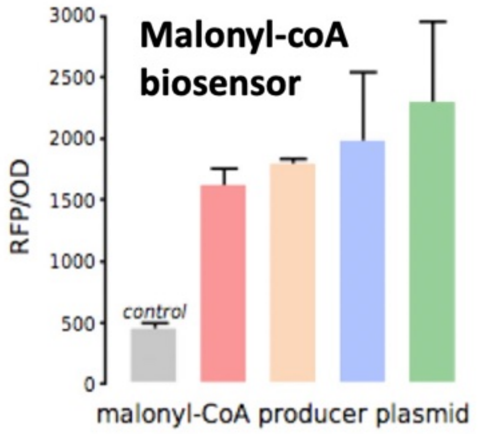


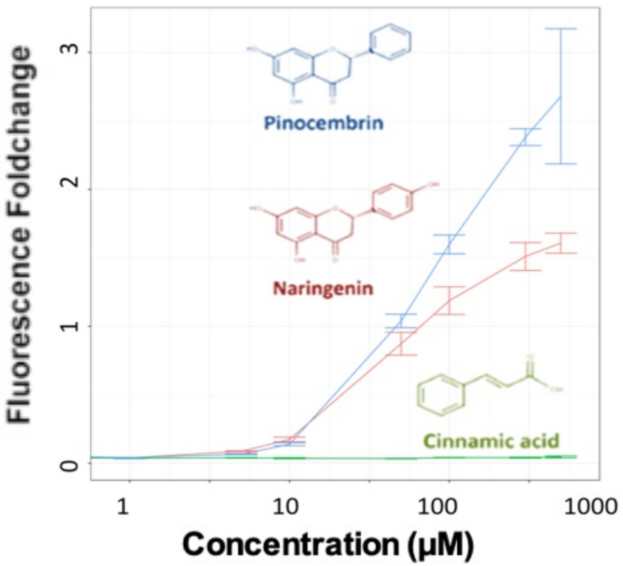


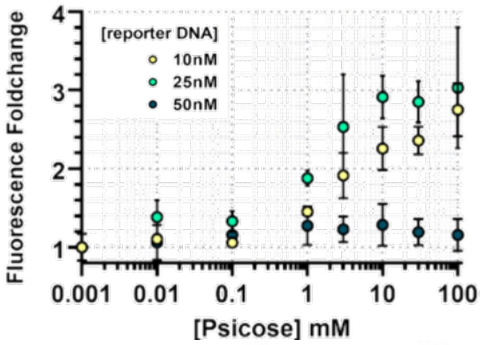


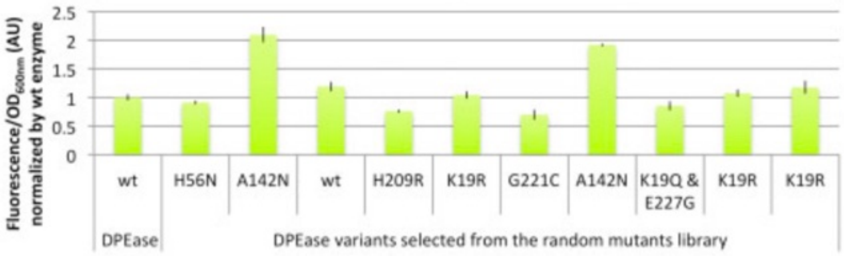
- Libis V, et al. ***ACS Synth Bio*** 2016

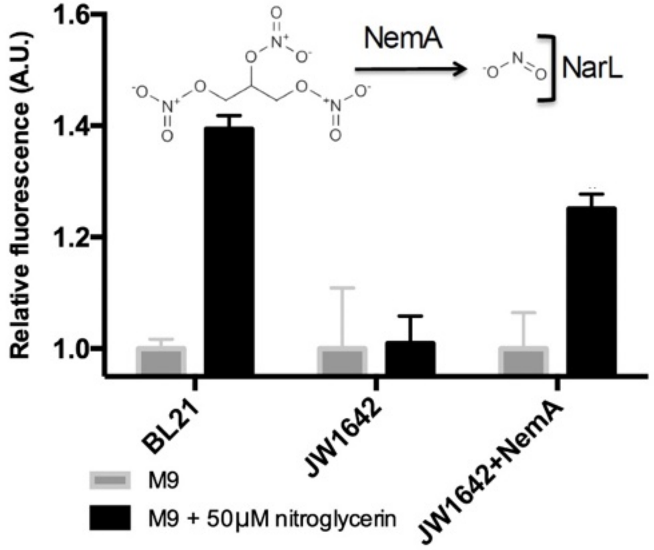
Characterizing biosensing circuits

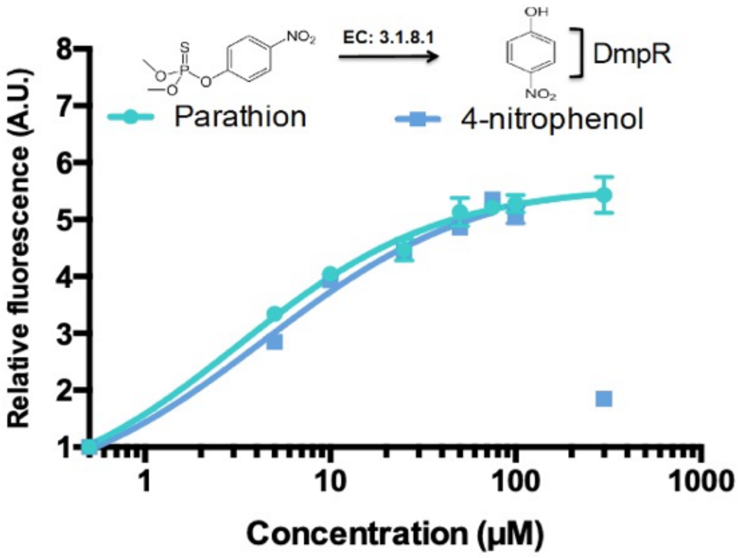


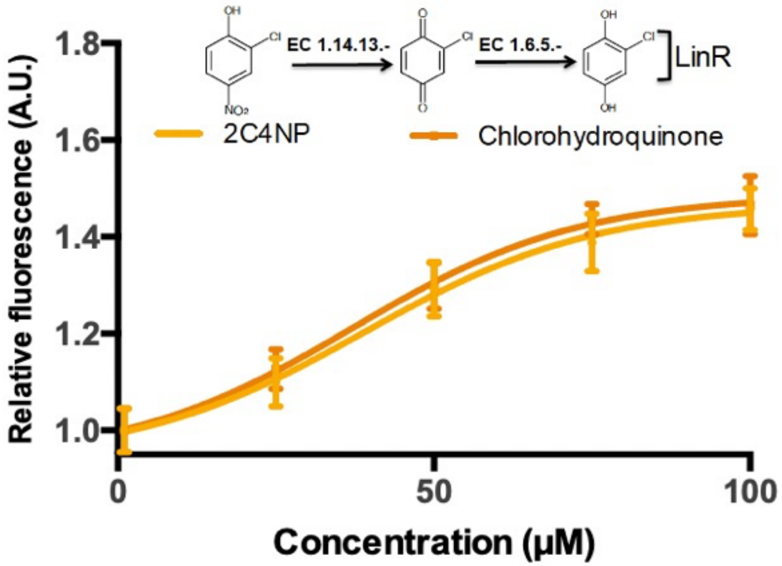






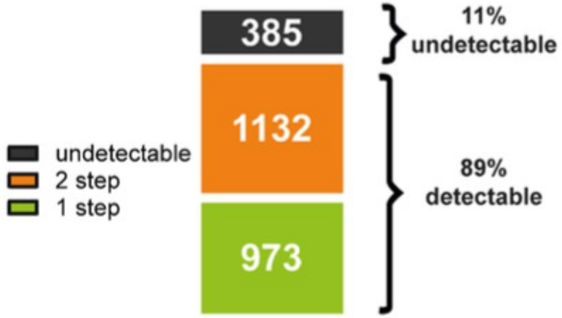




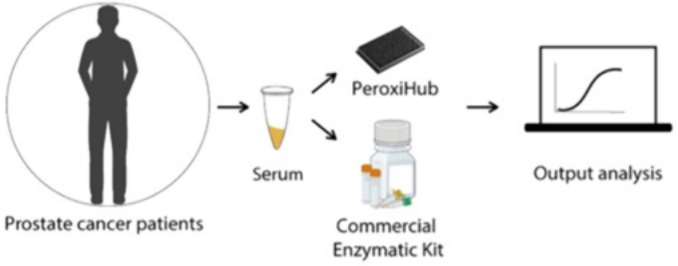


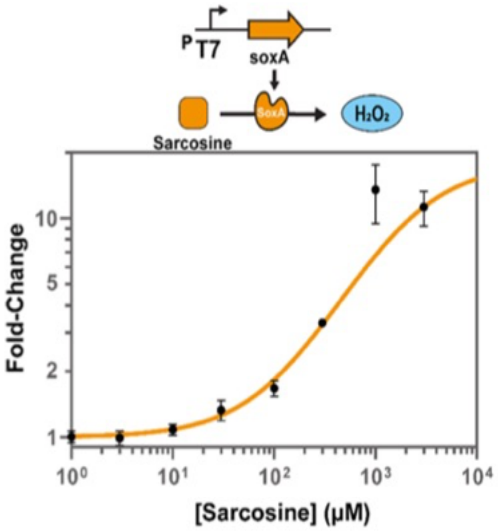
- Feher T *et al.* ***Frontier in Bioeng. & Biotech***, 2015
- Libis V. *et al*. ***Curr Opin Microbiol***, 2016
- Trabelsi H. *et al*. ***Biotech. & Bioeng.***, 2018
- Koch *M. et al.*, ***Frontiers in Bioeng & Biotech.***, 2018
- Koch M. *et al*. ***Curr Opin Biotech***, 2019
- Armetta J. *et al*. ***Synthetic Biology*** 2019 & Pandi A. *et al*. ***ACS Synth Biol***, 2019

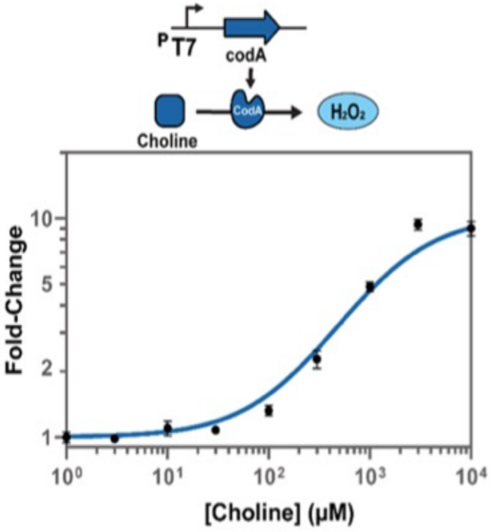
Cell-free biosensors with clinical samples

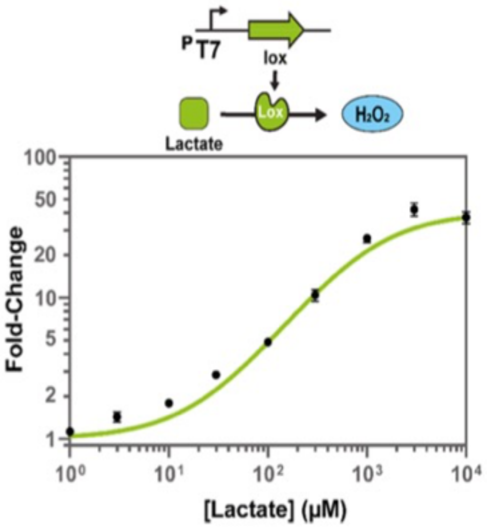


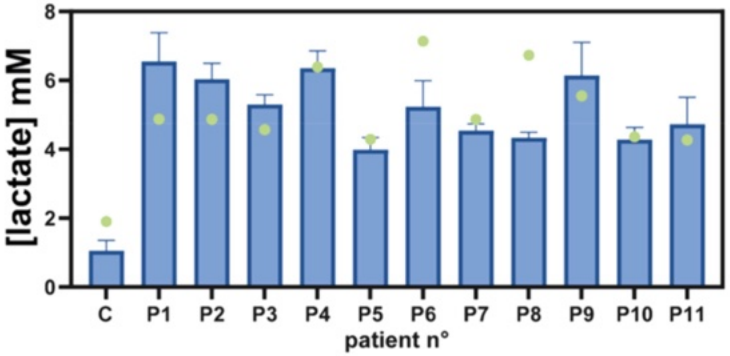
2490 biomarkers from HMDB





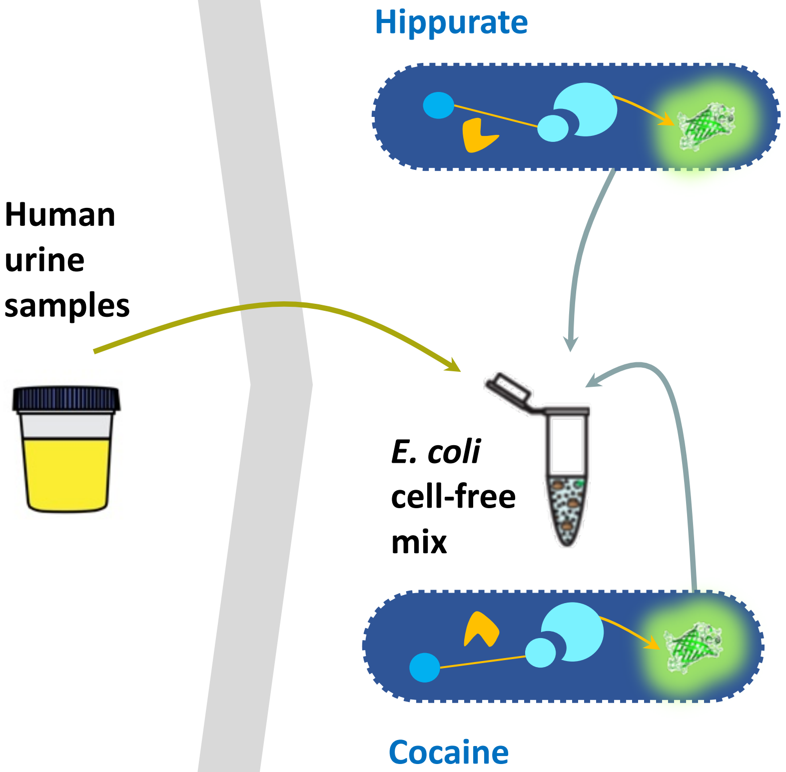






- Soudier P., et al. ***ACS Synth Bio*** 2022

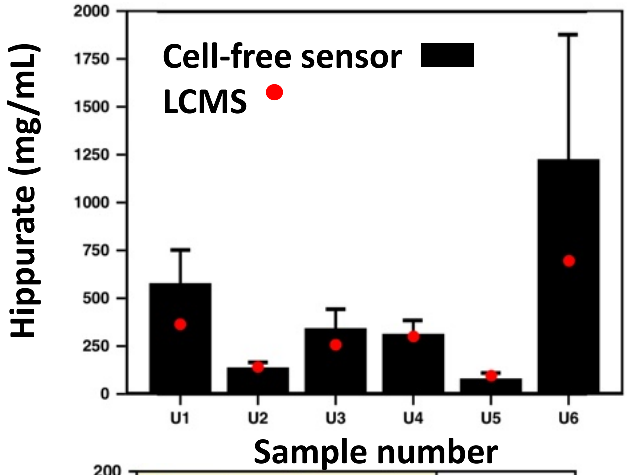
Cell-free biosensors with clinical samples

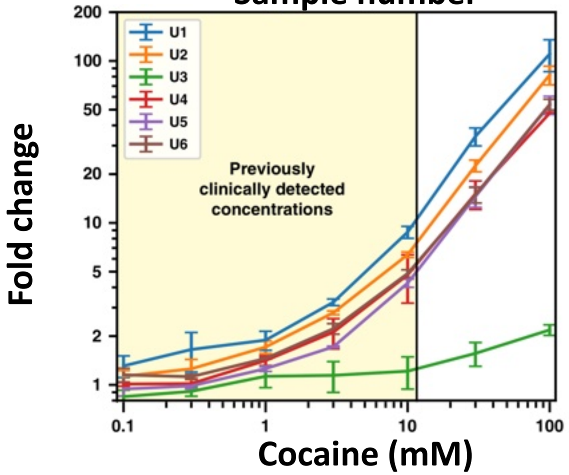

Human urine samples

Hippurate

E. coli **cell-free mix**

Cocaine



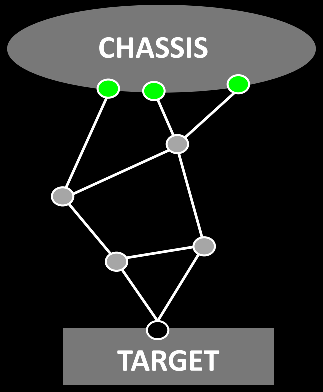


Hippurate is a degradation product of a probiotic treatment of *Phenylketonuria*

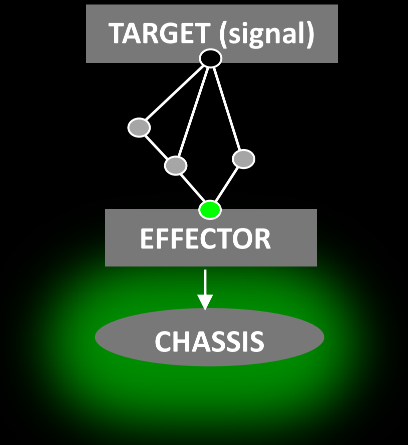

can be freeze dry

- Voyvodic, P. *et al*. ***Nature Comm.*** 2019

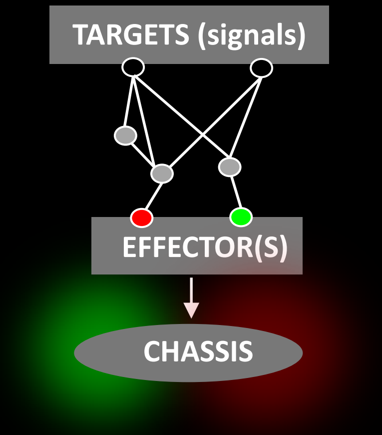
Engineering metabolic and genetic networks in vitro (cell-free) and in vivo (E. coli)



Metabolic engineering



Biosensor engineering

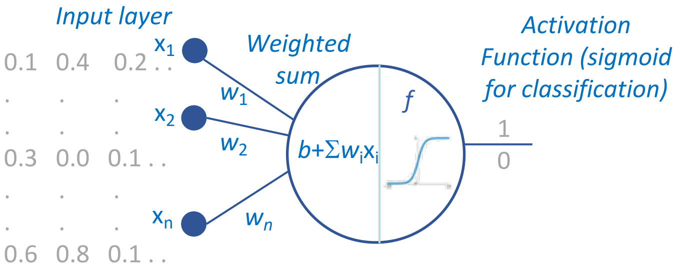


Information processing
biocomputing

Engineering a metabolic perceptron: why?

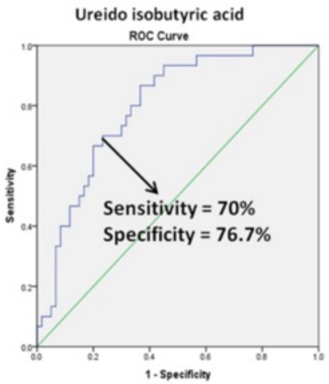
TRAINING THE NETWORK

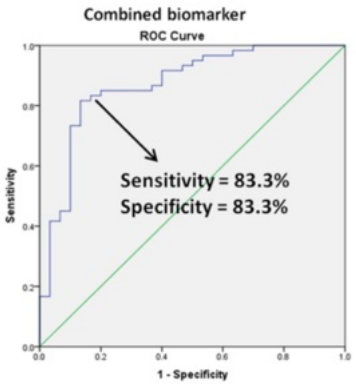
Perceptron weights (w_i) are learned to increase classifier accuracy



Prostate cancer

$$\text{PCa metabolic score} = b + \sum_{j=1}^{j} w_j x_{ij}$$





- Zang, *et al.* ***PLoS One*** 2013 and ***J Proteome Res.*** 2014
- Shen B, et al. ***Cell***. 2020
- …..

USING THE TRAINED NETWORK

To perform a diagnostic:

- Quantify a panel of biomarkers (metabolites) on clinical samples (using metabolomics)
- Feed measured biomarkers concentrations (x_i) to

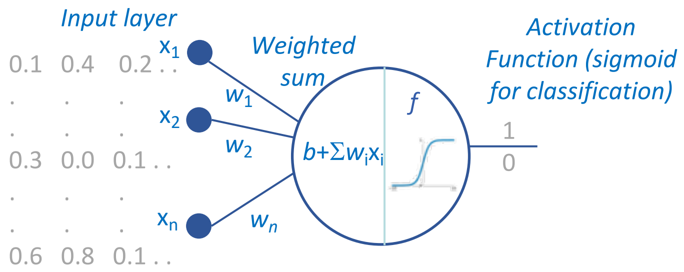
$$f\ (b\square\Sigma w_i x_i\ )$$

- **Is it possible to avoid biomarker concentration measurements?**
  - ➢ **Engineer the trained network *in vitro* or *in vivo* and directly use it on clinical samples**

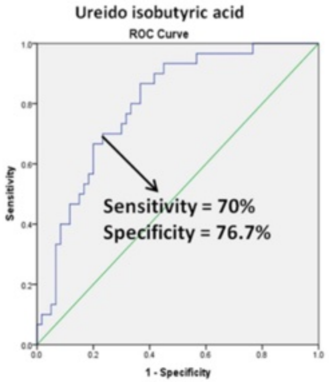
Engineering a metabolic perceptron: the concept

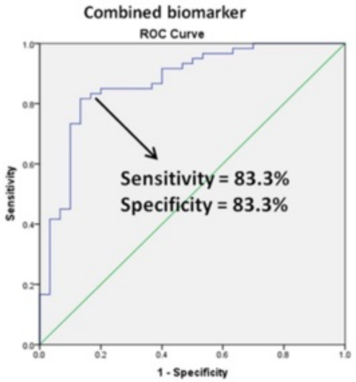
TRAINING THE NETWORK

Perceptron weights (w_i) are learned to increase classifier accuracy



Prostate cancer $\text{PCa metabolic score} = b + \sum_{j=1}^{j} w_j x_{ij}$

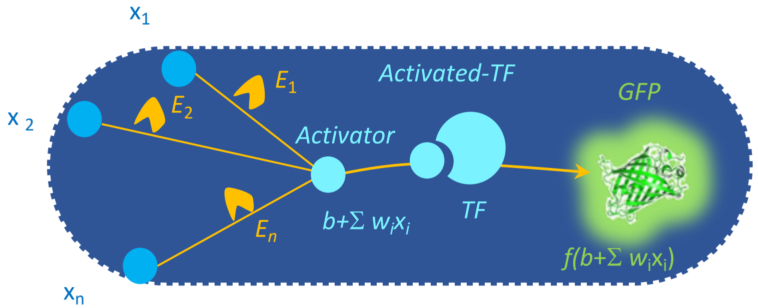




- Zang, *et al.* ***PLoS One*** 2013 and ***J Proteome Res.*** 2014
- Shen B, et al. ***Cell***. 2020
- .....

ENGINEERING THE TRAINED NETWORK

Need to actuate weighted sum and activation function

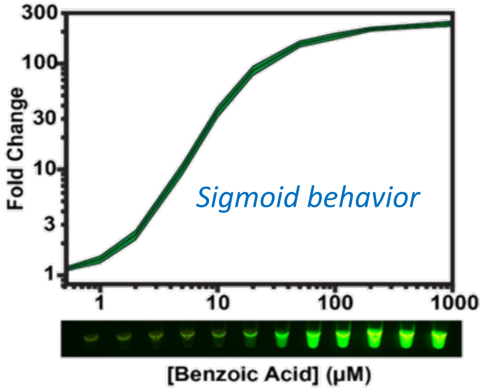


Enzymatic transformation *Reporter gene*

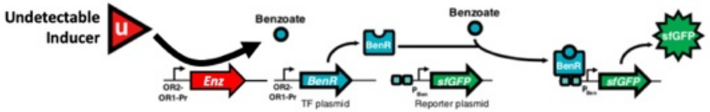
In theory (Michaelis-Menten) when $x_i \ll [E_i]$:

$$d[p] = \Sigma\, k_i [E_i] x_i\, dt$$

$w_i = k_i\ [E_i]$
where k_i is a kinetics constant



Engineering a metabolic perceptron: the concept



$$\frac{dbenzoate}{dt} = enz * \frac{k_{cat} * inducer}{inducer + K_M}$$

$$\frac{dinducer}{dt} = -enz * \frac{k_{cat} * inducer}{inducer + K_M}$$

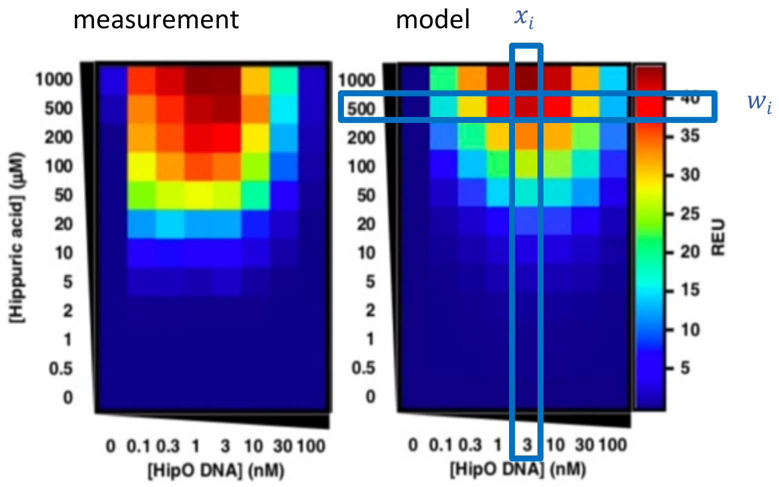
$$TF_{activated} = TF * \frac{benzoate}{benzoate + K_d^{inducer}} + 0.0005$$

$$\epsilon = \frac{TF_{activated}}{TF_{activated} + K_d^{activated}} \text{ for BenR}$$

$$\epsilon = 1 \text{ for constitutive expression}$$

$$\frac{dmRNA}{dt} = \gamma * n * \epsilon \frac{x}{x + \chi} * \frac{K_{tox}}{K_{tox} + tox} * \frac{R_{mRNA}}{R_{mRNA} + K_{mRNA}} - \delta * mRNA$$

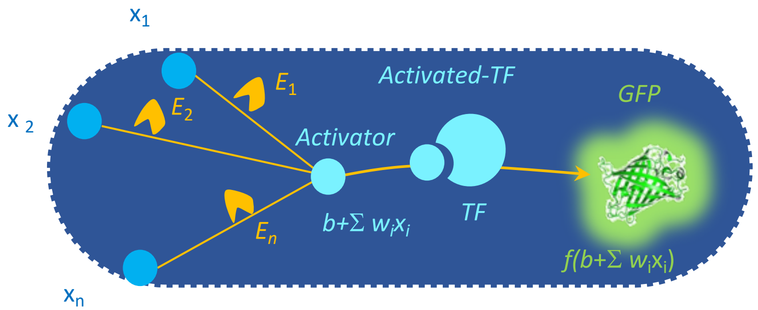
$$\frac{dprot}{dt} = \pi * mRNA * \frac{y}{y + k} * \frac{K_{tox}}{K_{tox} + tox} - \lambda * prot$$



- Pandi A.. *et al*. ***Nature Comm.*** 2019

ENGINEERING THE TRAINED NETWORK

Need to actuate weighted sum and activation function

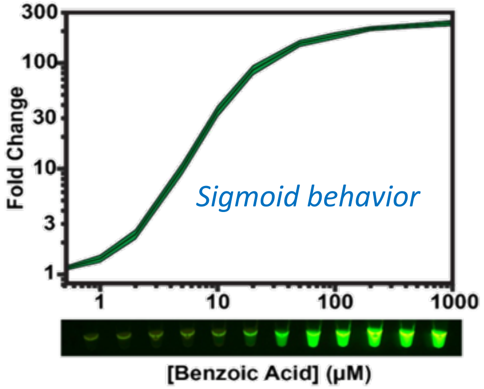


Enzymatic transformation *Reporter gene*

In theory (Michaelis-Menten) when $x_i \ll [E_i]$:

$$d[p] = \Sigma\, k_i [E_i] x_i\, dt$$

$w_i = k_i\ [E_i]$
where k_i is a kinetics constant



Engineering a metabolic perceptron

TARGETED BEHAVIOUR

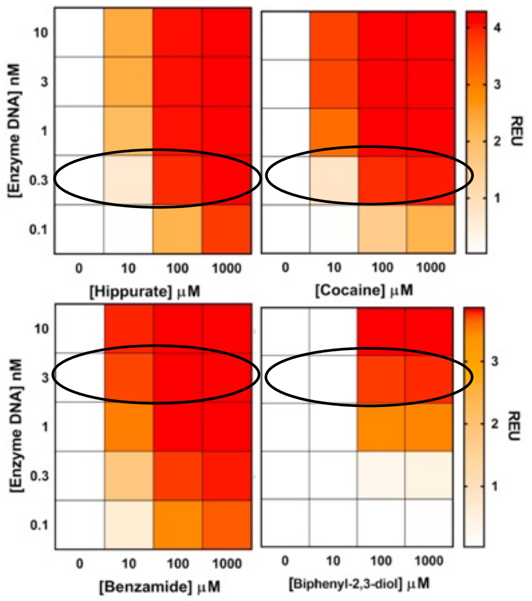
	(−)			(+)												
Hippurate……………		●				●	●	●				●	●	●		●
Cocaine……………….			●			●			●		●	●	●		●	●
Benzamide………….				●			●		●	●		●		●	●	●
Biphenyl-2,3-diol…					●			●		●	●		●	●	●	●

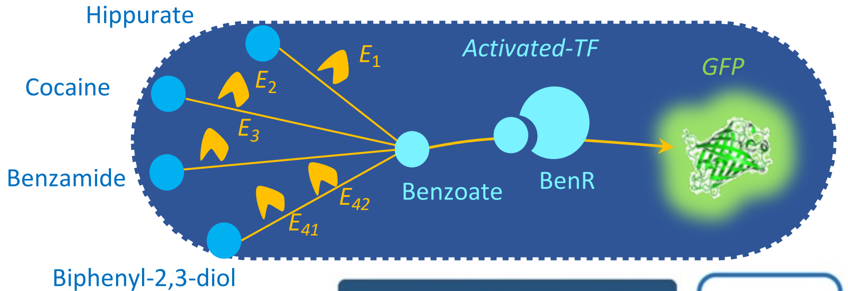
$f(b + \sum w_i x_i) < 0.5$ $f(b + \sum w_i x_i) \geq 0.5$

- Kinetics model is used to compute the enzyme concentration for each weight

Logistic regression

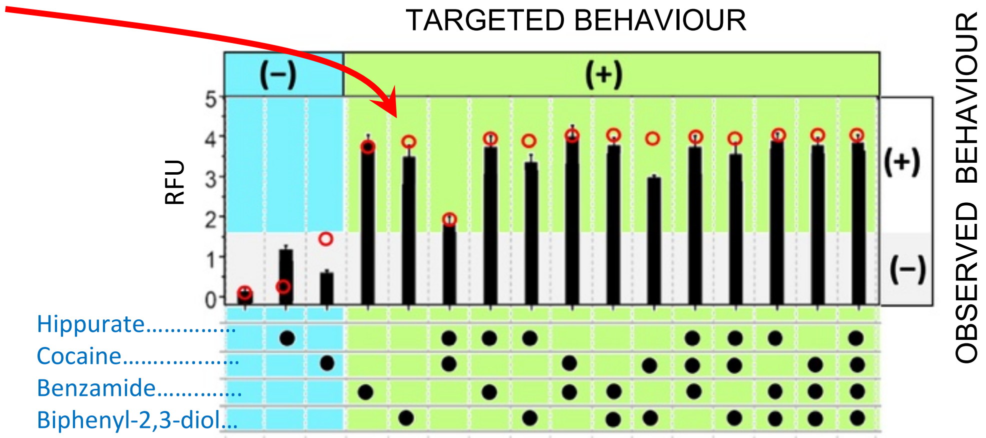
$f(b + \sum w_i x_i)$
$b = -0.50$
$w_1 = 0.25$
$w_2 = 0.25$
$w_3 = 0.50$
$w_4 = 0.50$





Retrosynthesis Workflow »

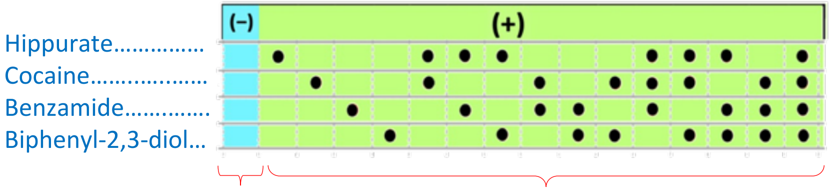
Galaxy



- Pandi A.. *et al*. ***Nature Comm.*** 2019

Engineering a metabolic perceptron

TARGETED BEHAVIOUR



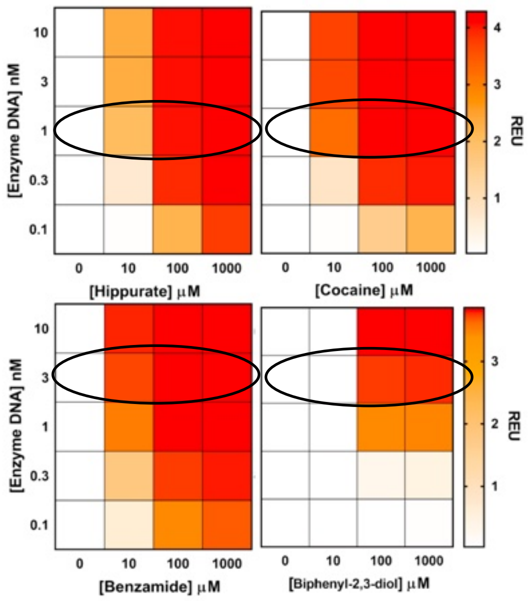
	(–)	(+)														
Hippurate..............		●				●	●	●				●	●	●		●
Cocaine..................			●			●			●		●	●	●		●	●
Benzamide.............				●			●		●	●		●		●	●	●
Biphenyl-2,3-diol...					●			●		●	●		●	●	●	●

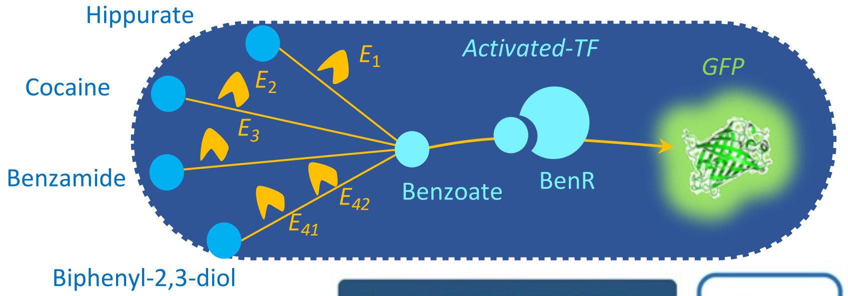
$f(b + \sum w_i x_i) < 0.5$ $f(b + \sum w_i x_i) \geq 0.5$

- Kinetics model is used to compute the enzyme concentration for each weight

Logistic regression

$f(b + \sum w_i x_i)$
$b = -0.50$
$w_1 = \cancel{0.25}\ 0.50$
$w_2 = \cancel{0.25}\ 0.50$
$w_3 = \cancel{0.50}\ 0.50$
$w_4 = \cancel{0.50}\ 0.50$

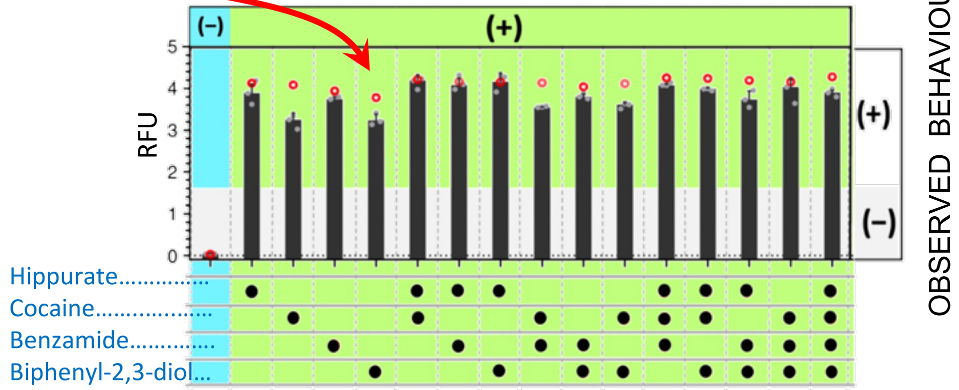




Retrosynthesis Workflow »

Galaxy

TARGETED BEHAVIOUR



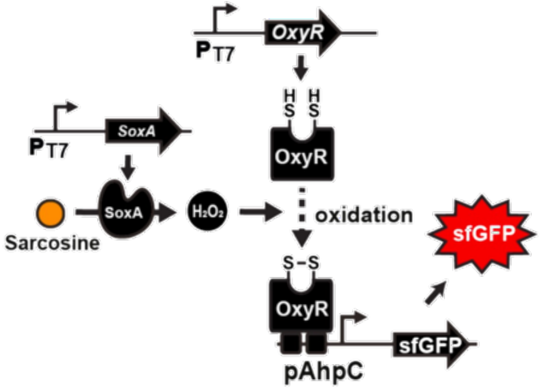
	(–)	(+)														
Hippurate..............		●				●	●	●				●	●	●		●
Cocaine..................			●			●			●		●	●	●		●	●
Benzamide.............				●			●		●	●		●		●	●	●
Biphenyl-2,3-diol...					●			●		●	●		●	●	●	●

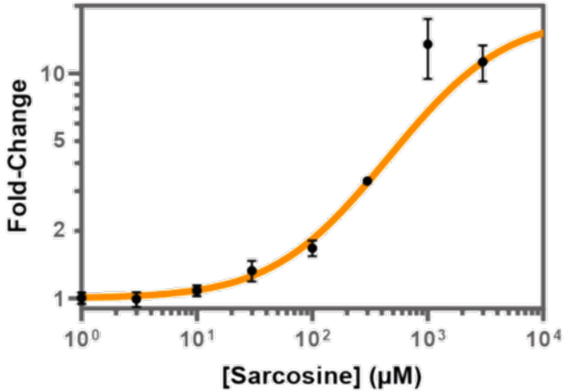
- Pandi A.. *et al*. ***Nature Comm.*** 2019

Toward engineering a multimodal perceptron

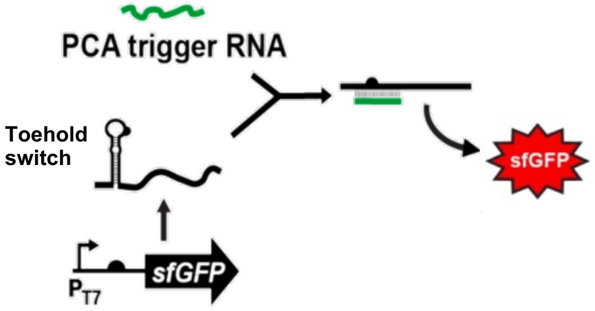
prostate cancer biomarkers

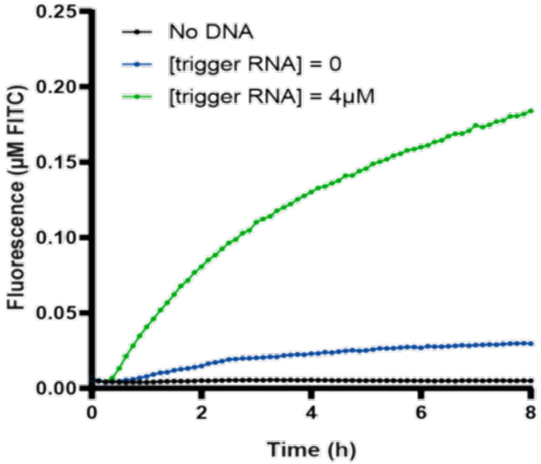
Metabolite (sarcosine)



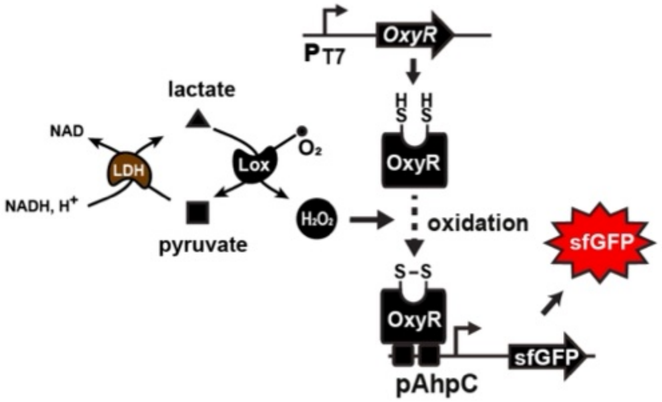


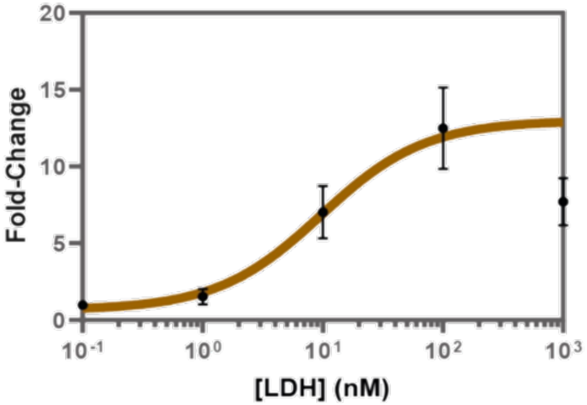
RNA (PCA3)





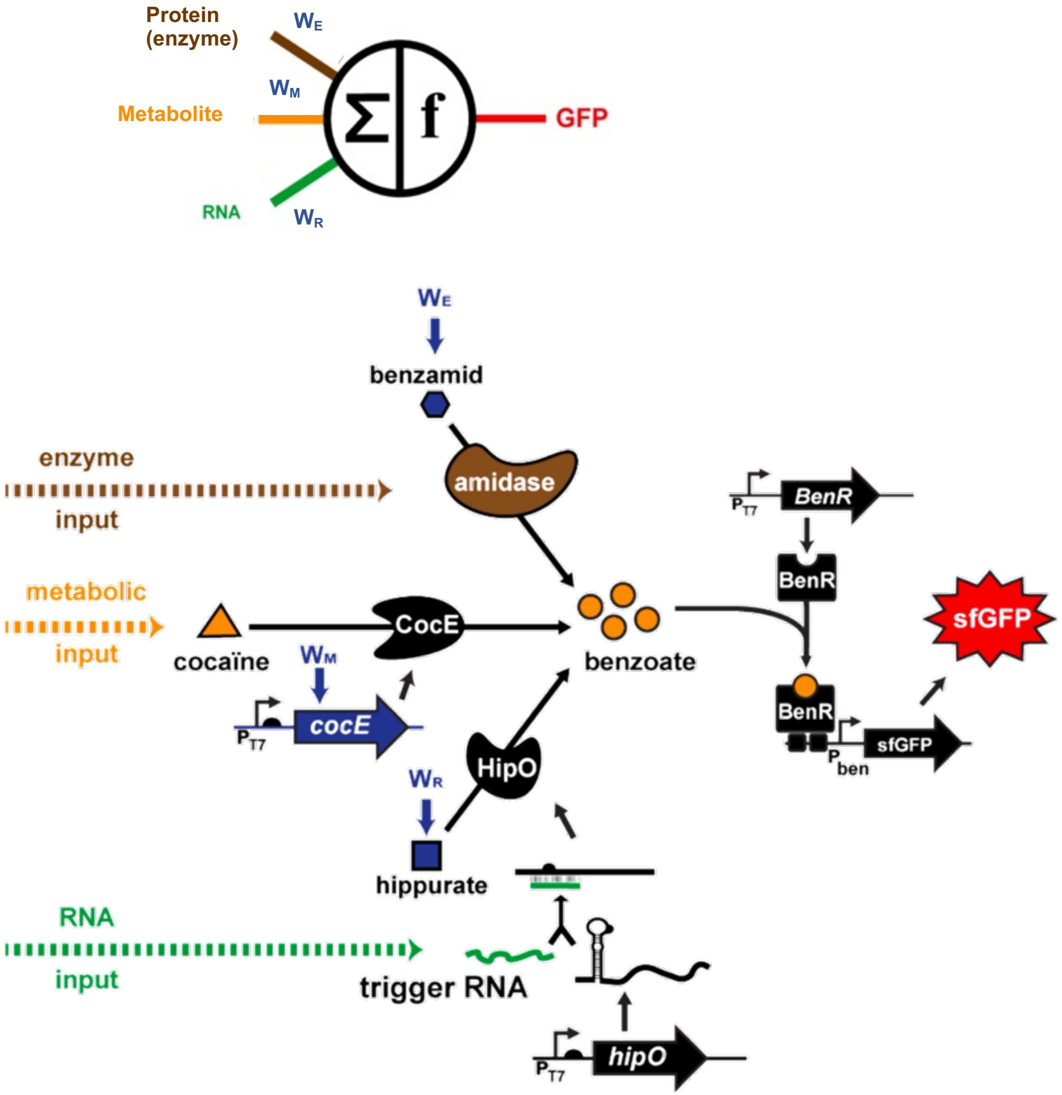
Protein (LDH)

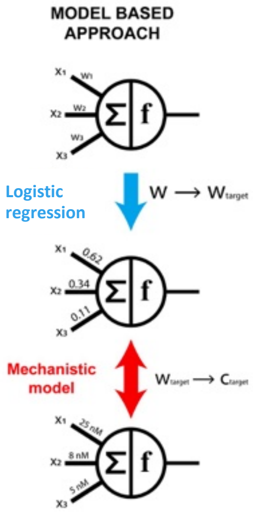




- Soudier, P.. ***PhD. thesis*** 2022

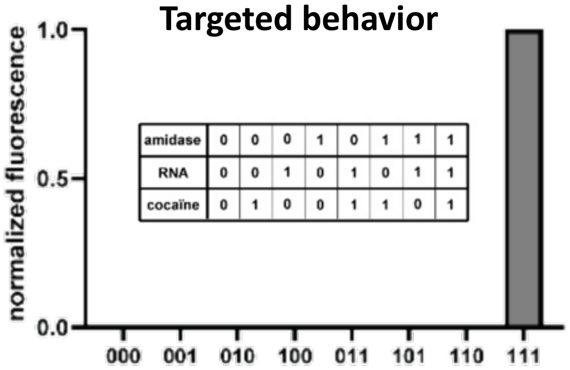
Engineering a multimodal perceptron



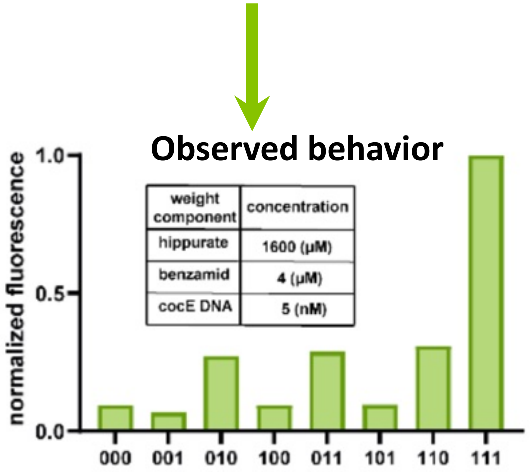


- Pandi A. *et al*. ***Nature Comm.*** 2019

- Borkowski O. *et al*. ***Nature Comm.*** 2020 & Pandi A. *et al*. ***Nature Comm.*** 2022



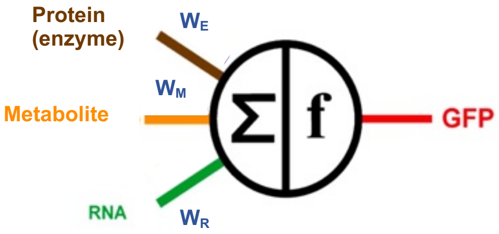
amidase	0	0	0	1	0	1	1	1
RNA	0	0	1	0	1	0	1	1
cocaïne	0	1	0	0	1	1	0	1

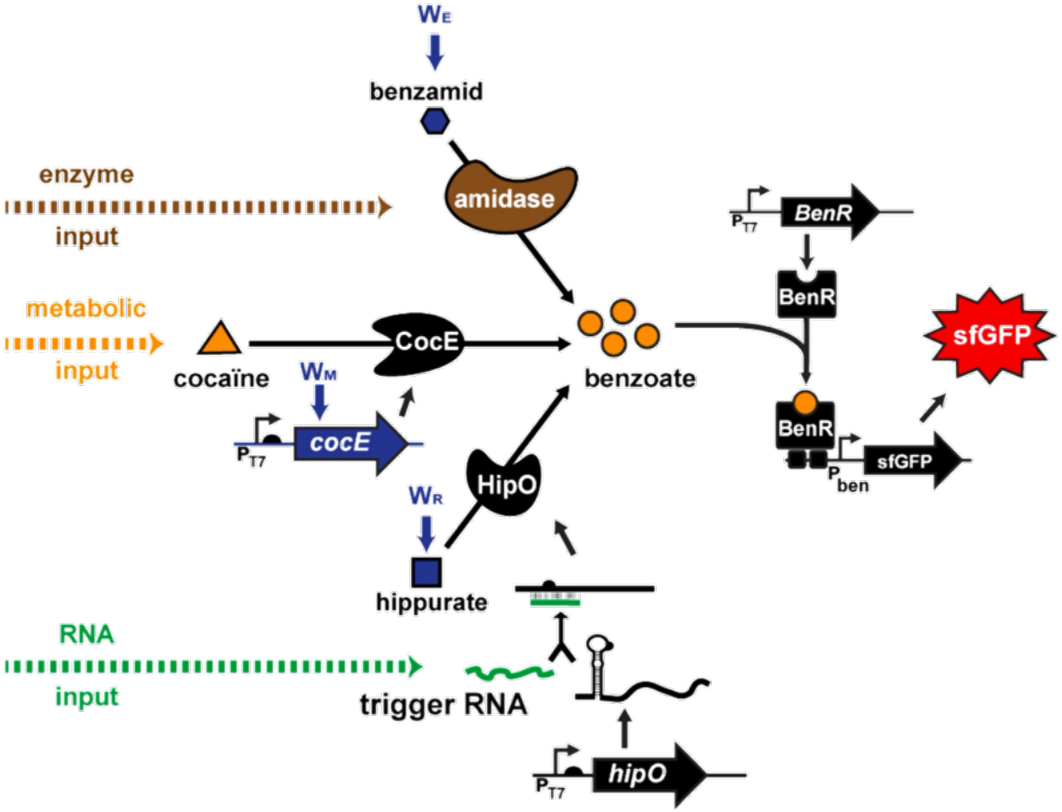


weight component	concentration
hippurate	1600 (µM)
benzamid	4 (µM)
cocE DNA	5 (nM)

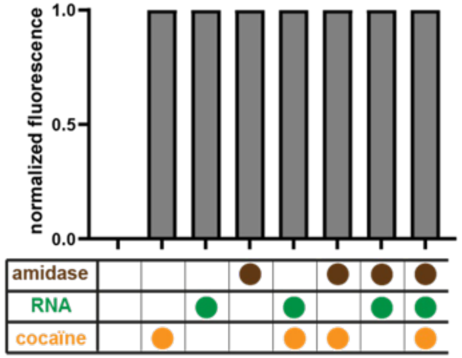
- Soudier, P.. ***PhD. thesis*** 2022

Engineering a multimodal perceptron

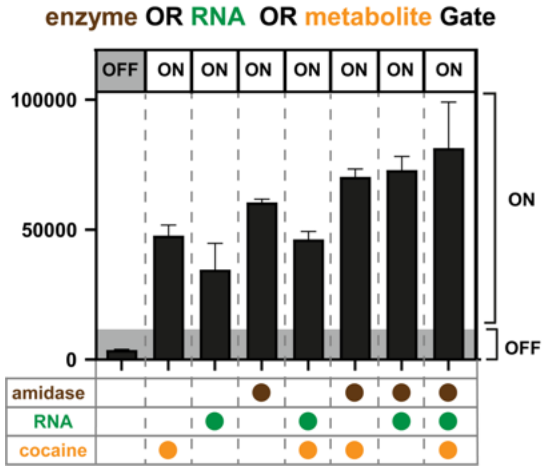




Targeted behavior

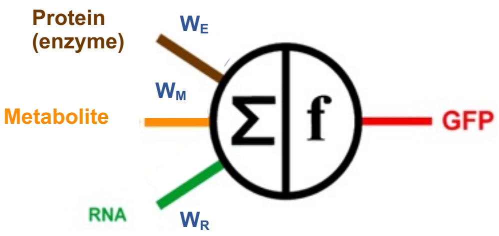


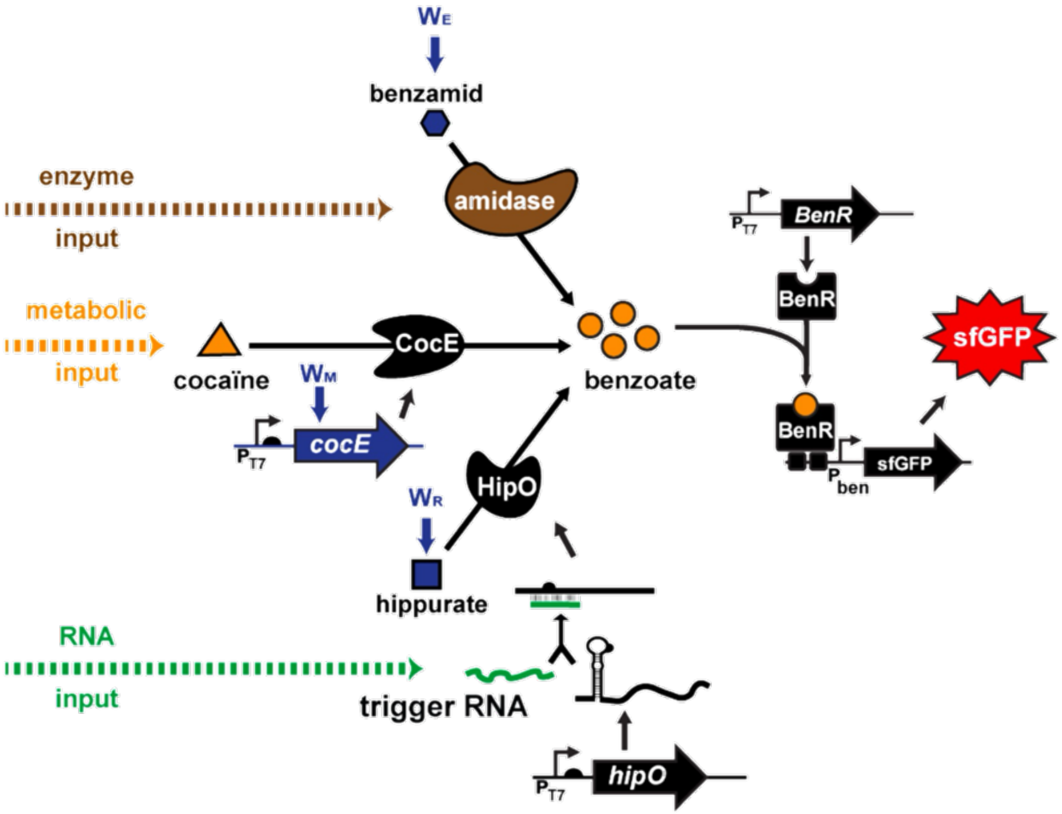
Observed behavior

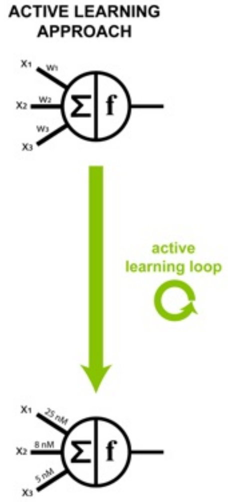


- Soudier, P.. *PhD. thesis* 2022

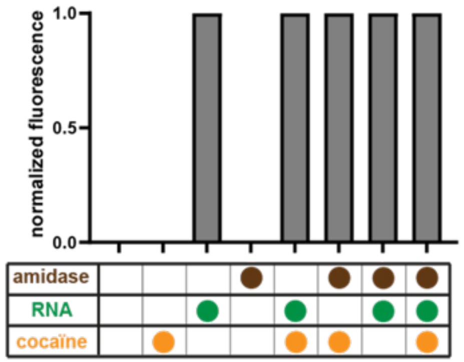
Engineering a multimodal perceptron



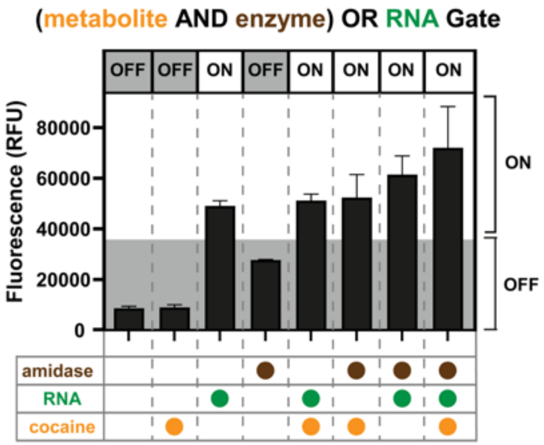




Targeted behavior

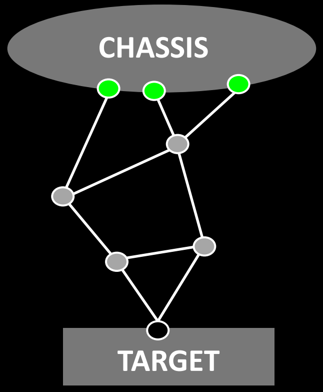


Observed behavior

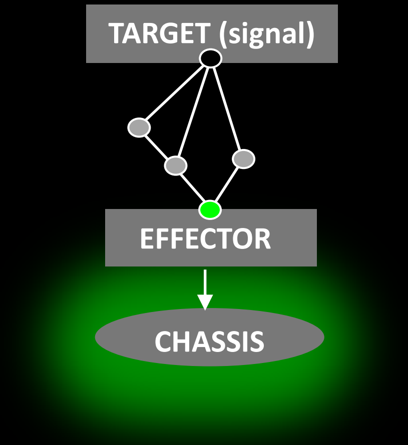


- Soudier, P.. ***PhD. thesis*** 2022

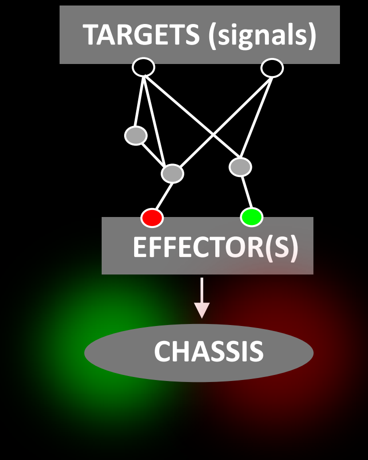
Engineering metabolic and genetic networks in vitro (cell-free) and in vivo (E. coli)



Metabolic engineering



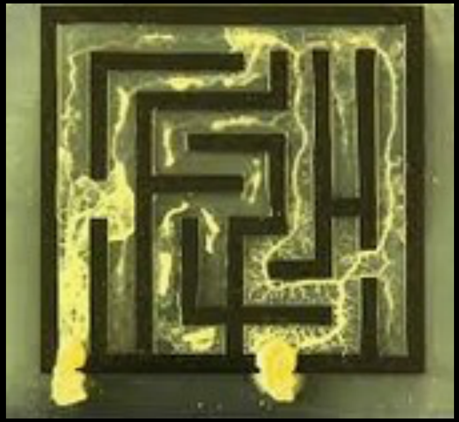
Biosensor engineering

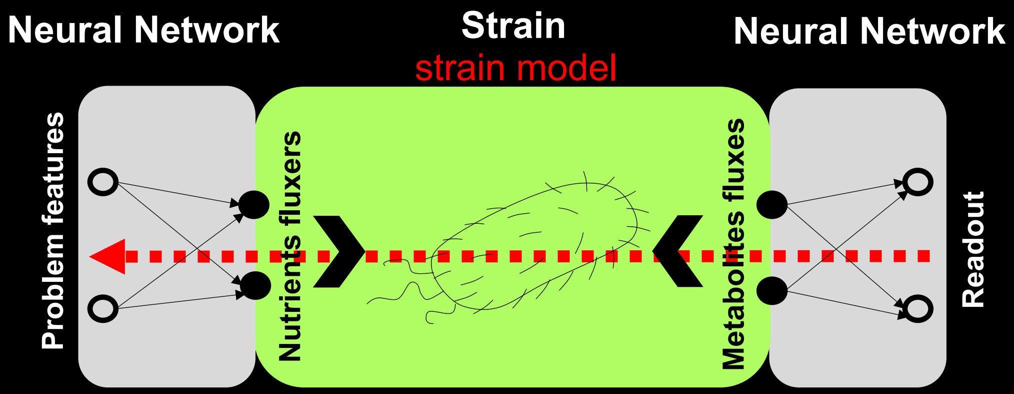


Information processing
biocomputing

Engineering complex metabolic devices in vivo?

- Can we divert native metabolism to handle problems that are usually solved *in silico*?





gradient backpropagation

A model allowing gradient backpropagation accurately reproducing phenotype for different media composition

Engineering complex metabolic devices in vivo?

Classical mechanistic model (FBA):

Max $(v_{biomass})$

Subjected to:

$S\,V = 0$

$0 \leq V \leq V_{in}$

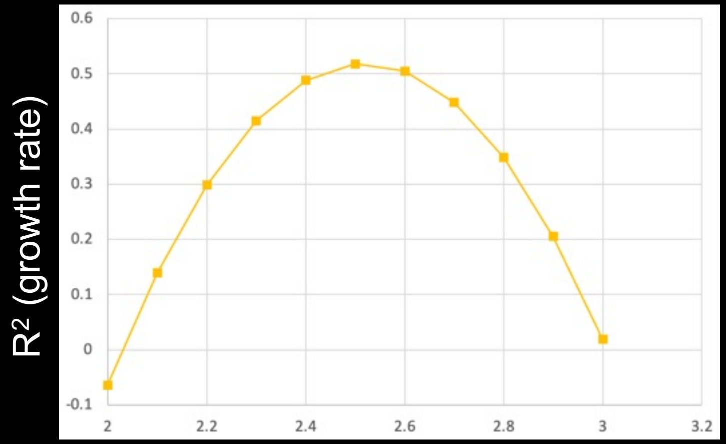
where

$-\ V$ = set of all reaction fluxes

$-$ S = stochiometric matrix

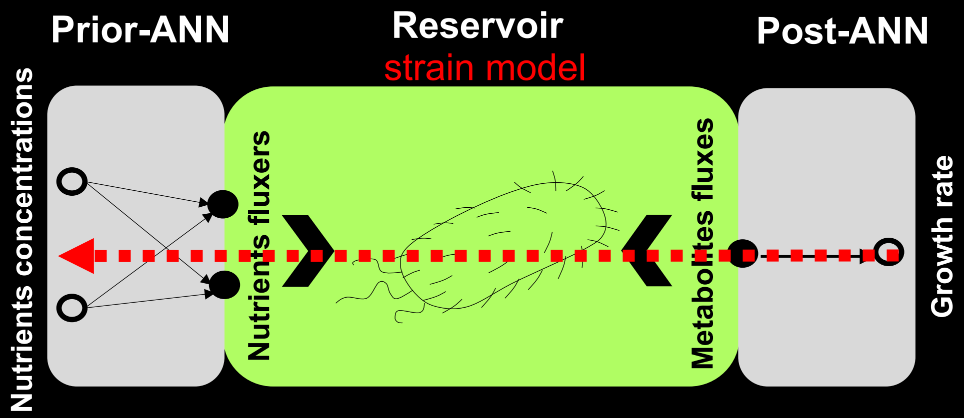
$-V_{in}$ = uptake medium fluxes upper bounds

FBA (Cobrapy) growth rates vs. measured growth rate in E. coli DH5-alpha for 1 to 4 nutrients added to M9



Concentration to flux scaler

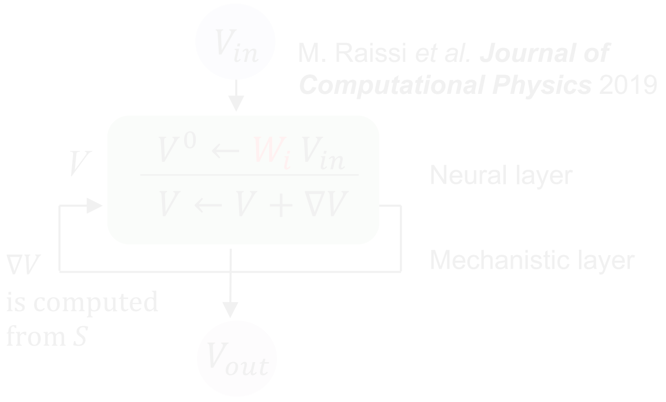
gradient backpropagation



a model allowing gradient backpropagation that accurately reproduce phenotype for different media composition

Hybrid models:
gradient backpropagation compatible solutions surrogating classical mechanistic models

Physics informed neural network



M. Raissi *et al. **Journal of Computational Physics*** 2019

V_{in}

$V^0 \leftarrow W_i\, V_{in}$

$V \leftarrow V + \nabla V$

V

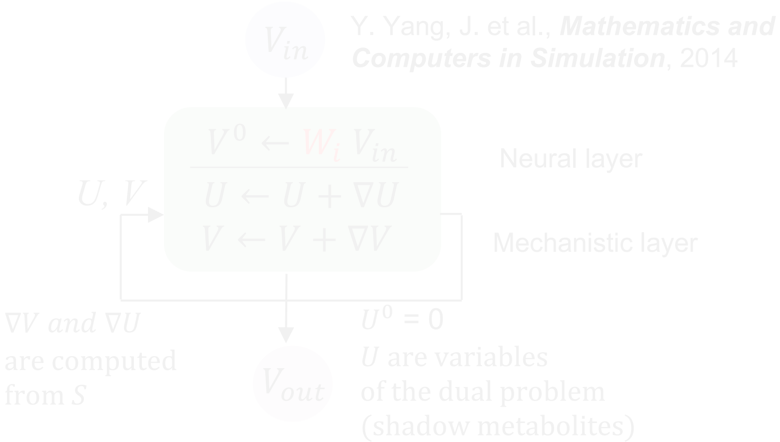
Neural layer

Mechanistic layer

∇V is computed from S

V_{out}

Hopfield's network



Y. Yang, J. et al., ***Mathematics and Computers in Simulation***, 2014

V_{in}

$V^0 \leftarrow W_i\, V_{in}$

$U \leftarrow U + \nabla U$

$V \leftarrow V + \nabla V$

U, V

Neural layer

Mechanistic layer

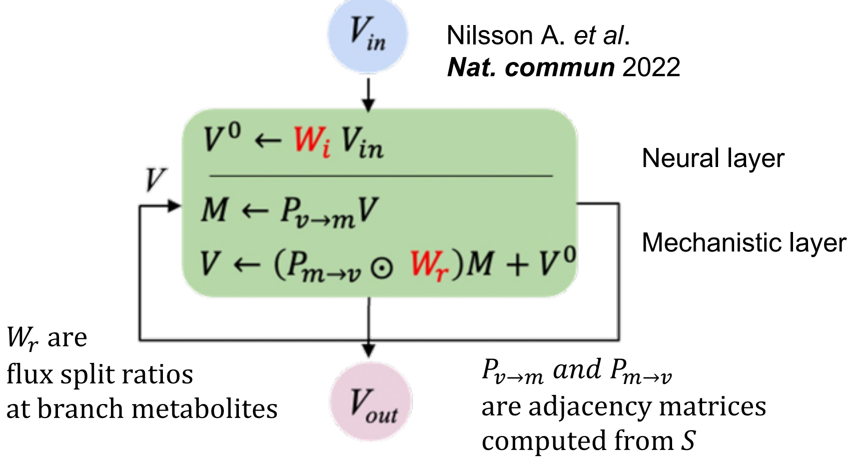
∇V *and* ∇U are computed from S

V_{out}

$U^0 = 0$

U are variables of the dual problem (shadow metabolites)

Knowledge primed neural network



Nilsson A. *et al*.
Nat. commun 2022

V_{in}

$V^0 \leftarrow W_i\, V_{in}$

$M \leftarrow P_{v\rightarrow m} V$

$V \leftarrow (P_{m\rightarrow v} \odot W_r) M + V^0$

V

Neural layer

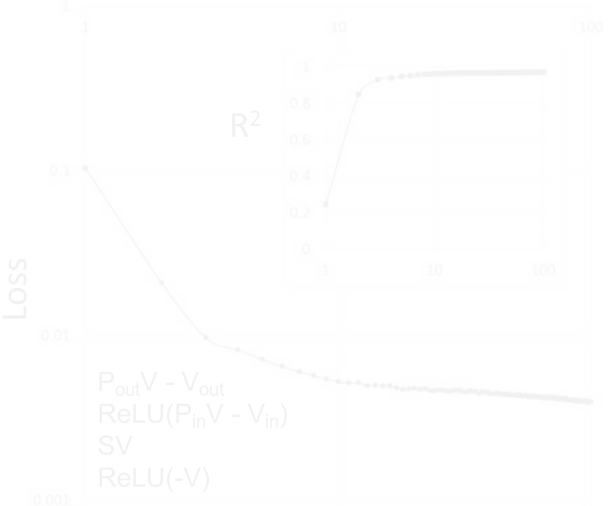
Mechanistic layer

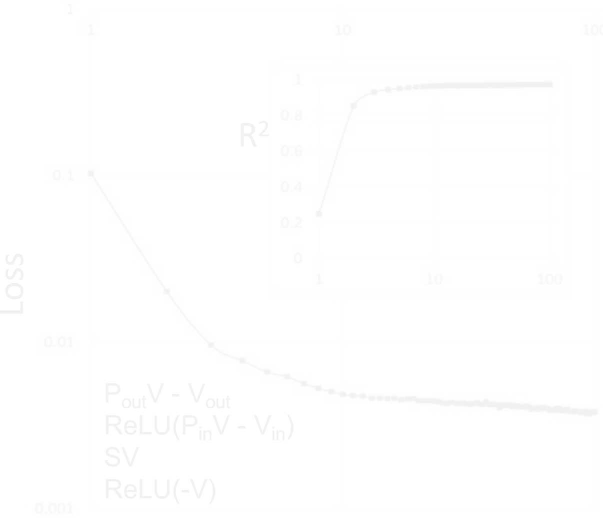
W_r are flux split ratios at branch metabolites

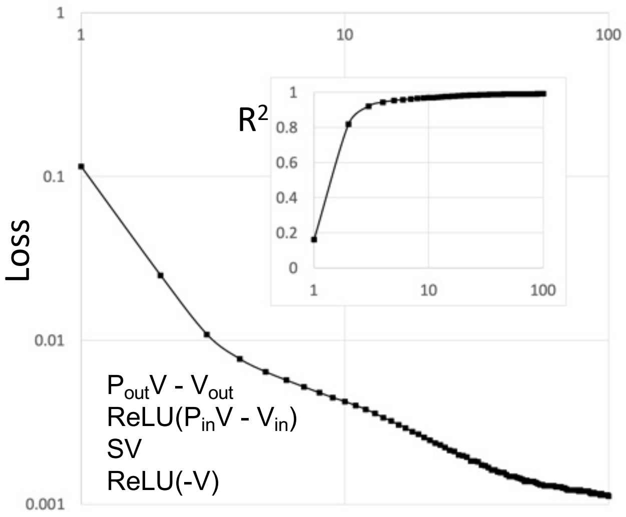
V_{out}

$P_{v\rightarrow m}$ *and* $P_{m\rightarrow v}$ are adjacency matrices computed from S

Trained on FBA simulated growth rates with *E. coli core* model for 1000 different media (media = minimal medium + 1 to 6 metabolites chosen at random among 13)



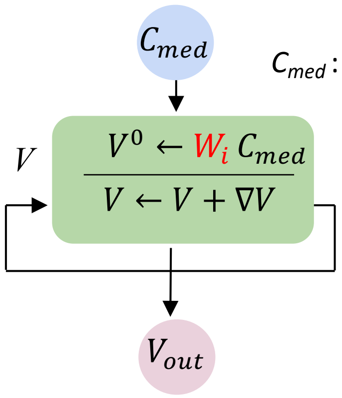




- Faure L. *et al*. ***Nat Communications***, 2023

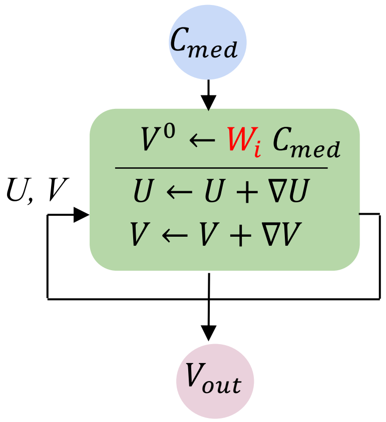
Hybrid models:
gradient backpropagation compatible solutions surrogating classical mechanistic models

Physics informed neural network

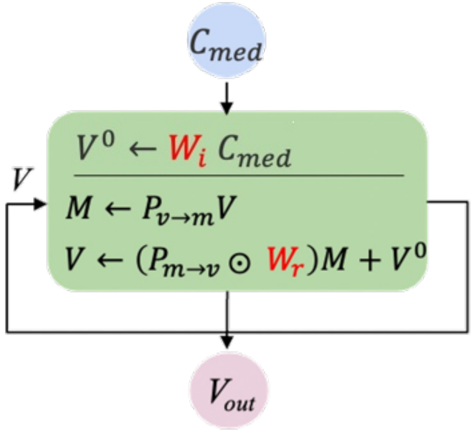


C_{med} : medium metabolite concentrations

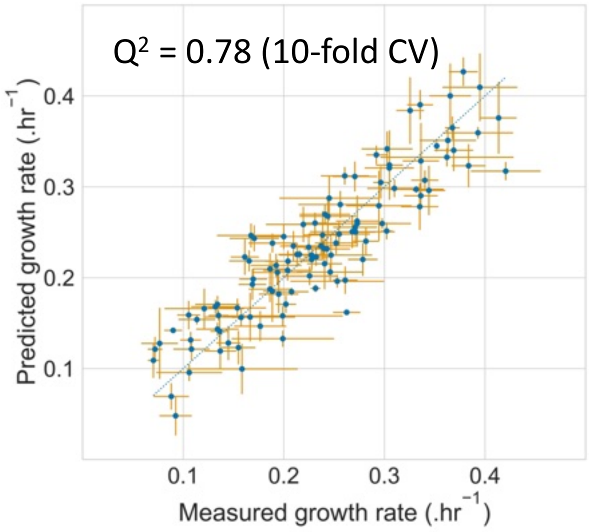
Hopfield’s network

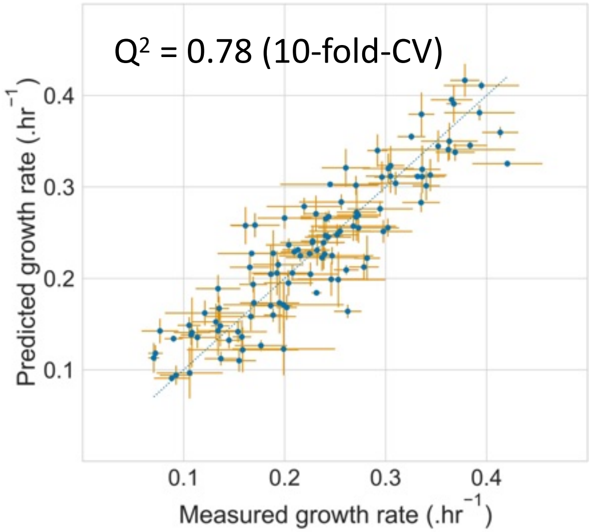


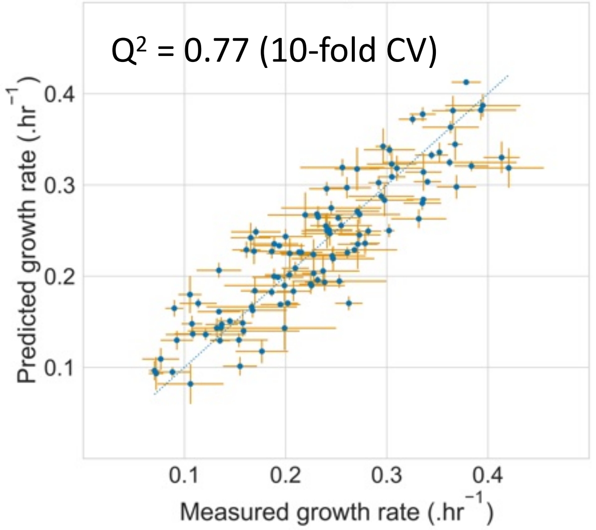
Knowledge primed neural network



Prediction after being trained on an experimental data set where growth rates were measured for 110 different media compositions for *E. coli* DH5-alpha *strain* (media = M9 + 4 nutrients chosen at random among 10)

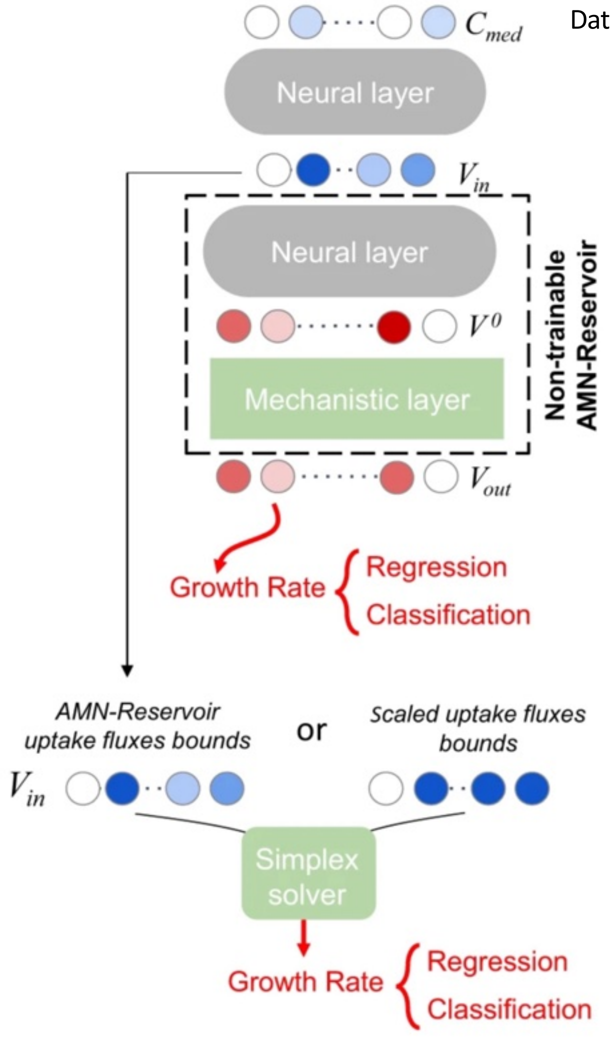






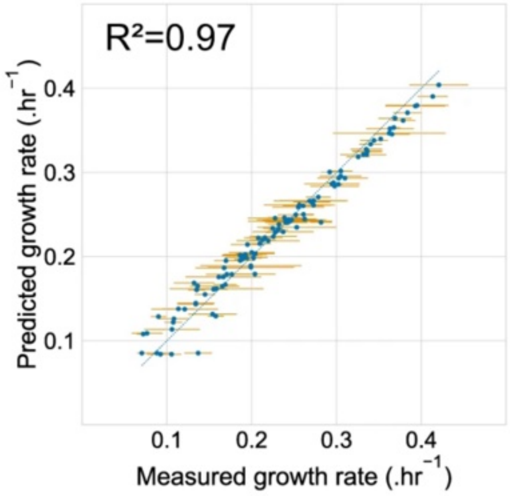
- Faure L. *et al*. ***Nat Communications***, 2023

Hybrid models:
can be used to parameterize mechanistic models

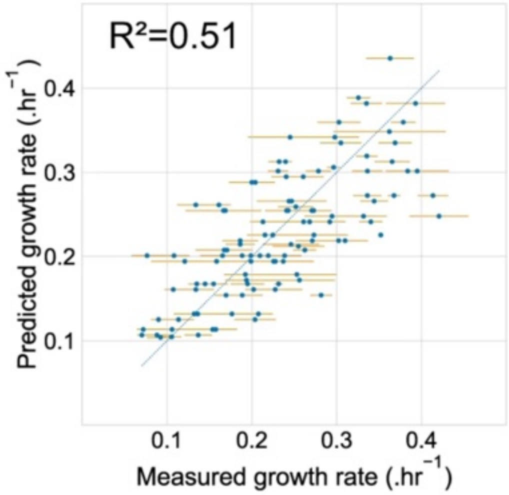


Data set 110 different media compositions for *E. coli* DH5-alpha for which growth rates have been measured

Hybrid model results

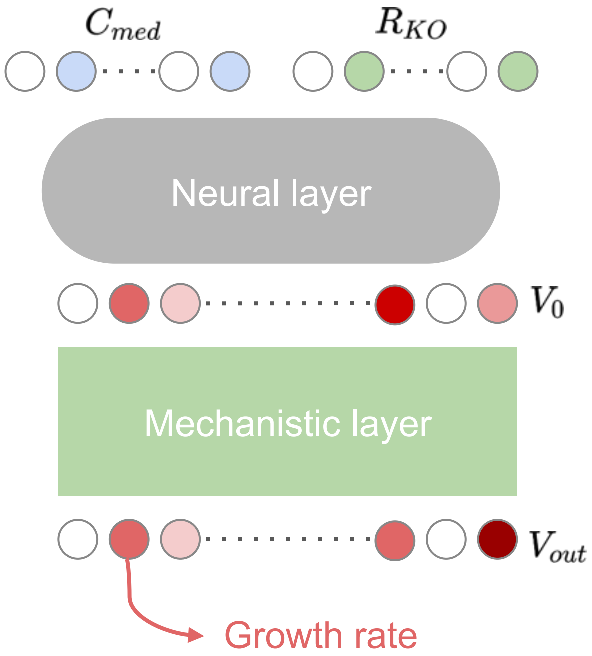


Mechanistic model results



- Faure L. *et al*. ***Nat Communications***, 2023

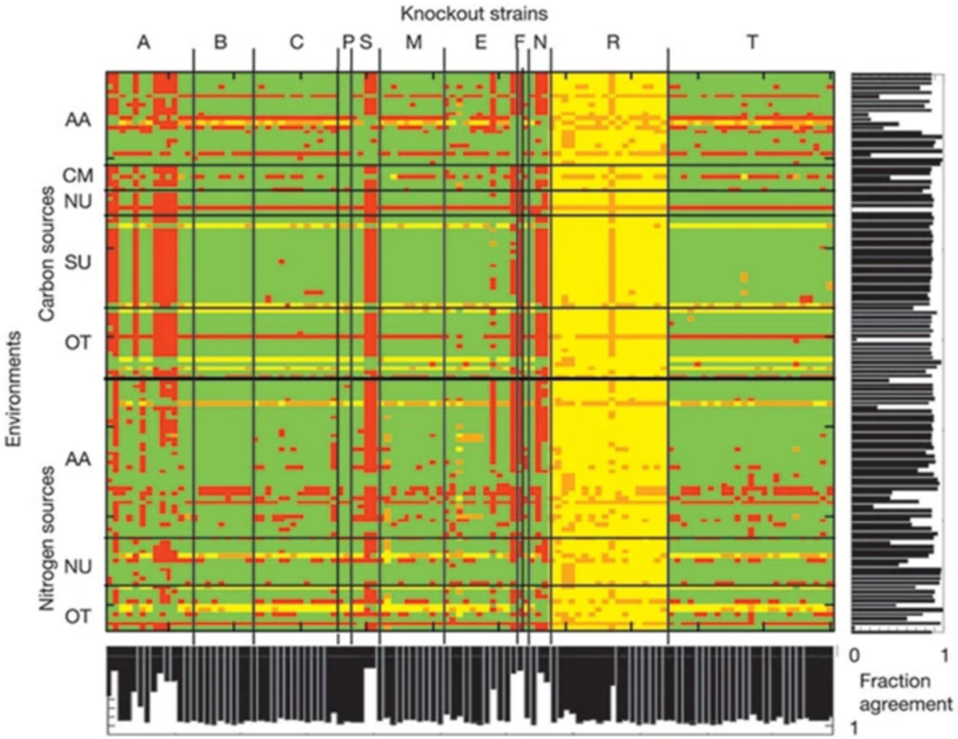
Hybrid models:
can integrate gene regulation



Loss

$P_{out}V - V_{out}$
$\mathrm{ReLU}(P_{in}V - V_{in})$
SV
$\mathrm{ReLU}(-V)$
$\mathbf{VR_{KO}}$

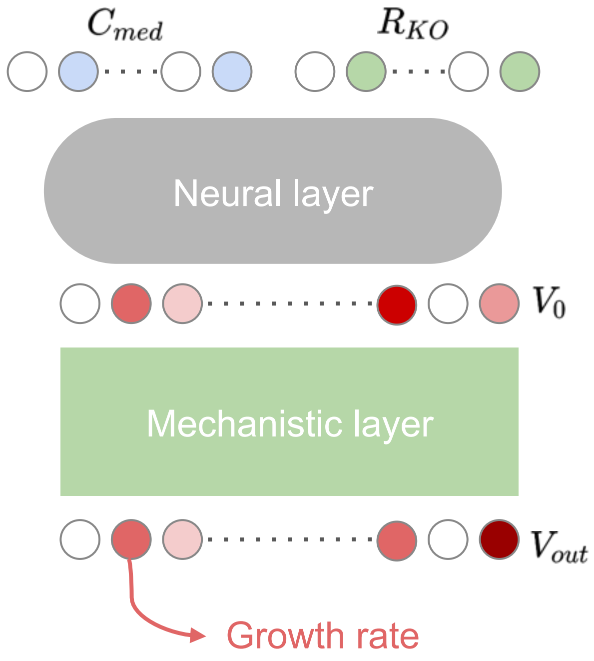
Dataset from ASAP database (Glasner D. *et al*. *NAR* 2003)



- 120 metabolic genes targeted (KOs), targeting 127 reactions
- 145 conditions (1 or 2 substrates, some with added succinate to enable growth)
- Each KO screened in all conditions (145*120=17400)

- Faure L. *et al*. ***Nat Communications***, 2023

Hybrid models:
can integrate gene regulation

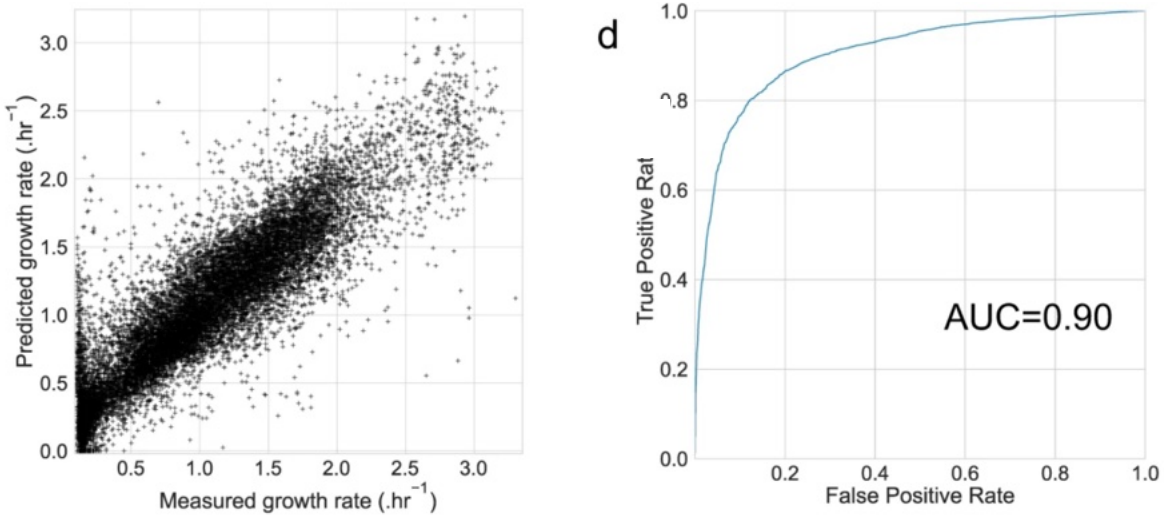


Loss

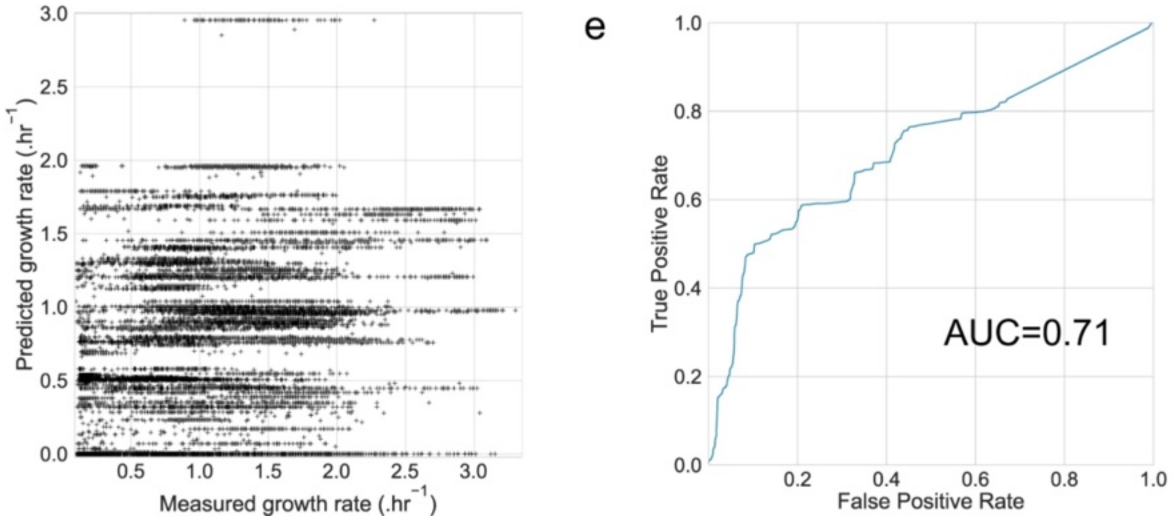
$P_{out}V - V_{out}$
$\mathrm{ReLU}(P_{in}V - V_{in})$
SV
$\mathrm{ReLU}(-V)$
$\mathbf{VR_{KO}}$

Dataset from ASAP database (Glasner D. *et al*. *NAR* 2003)

Hybrid model results

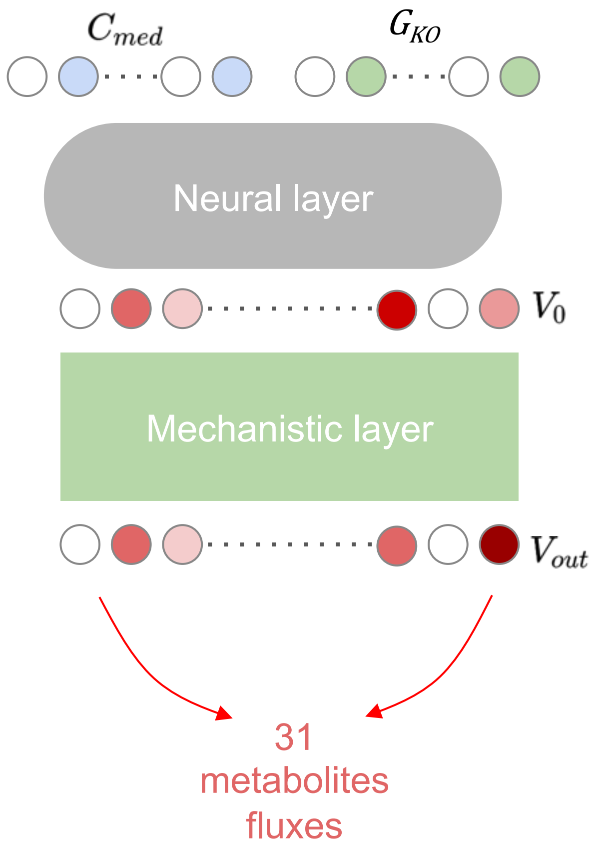


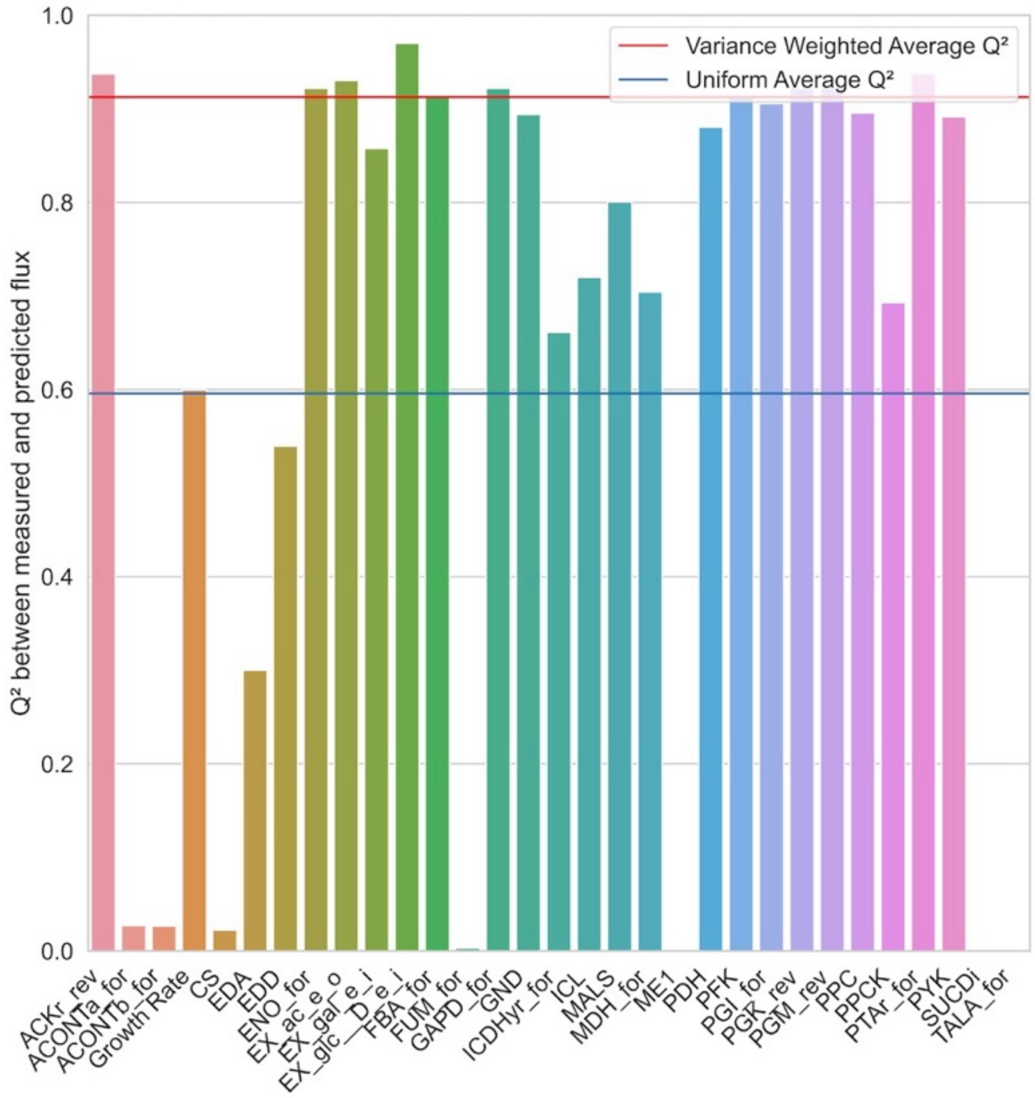
Mechanistic model results



- Faure L. *et al*. ***Nat Communications***, 2023

Hybrid models:
can learn more than the growth rate



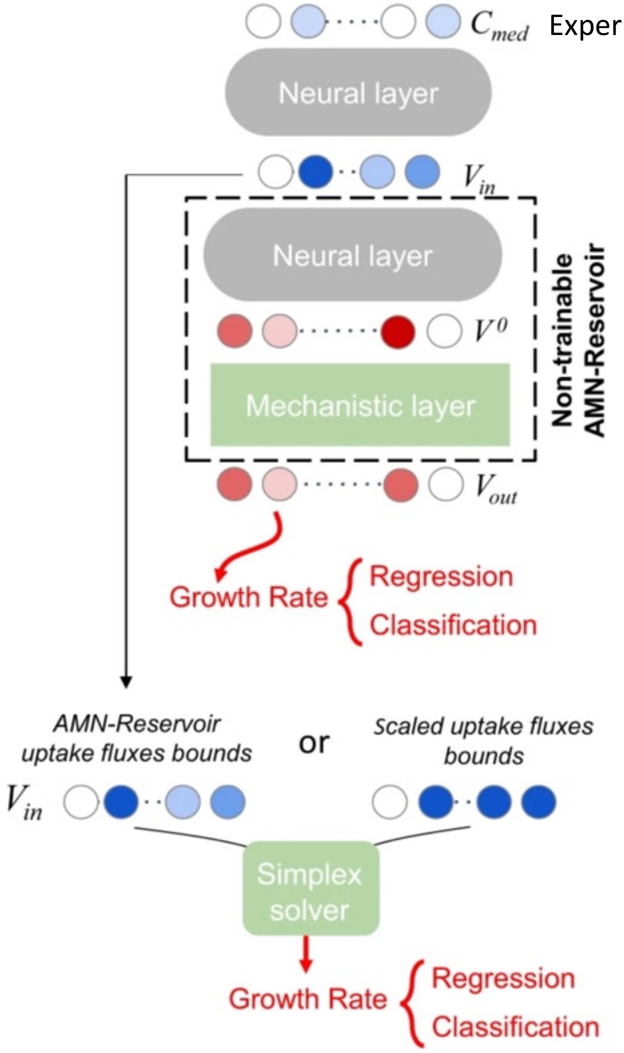


E. coli dataset from Rijsewijk *et al. Mol. Syst. Biol.* **7**, 477, 2011

- 128 experiments, each containing 31 measured fluxes.
- 2 media compositions (glucose or galactose as carbon source)
- 64 regulator gene KOs mutants (GKO)

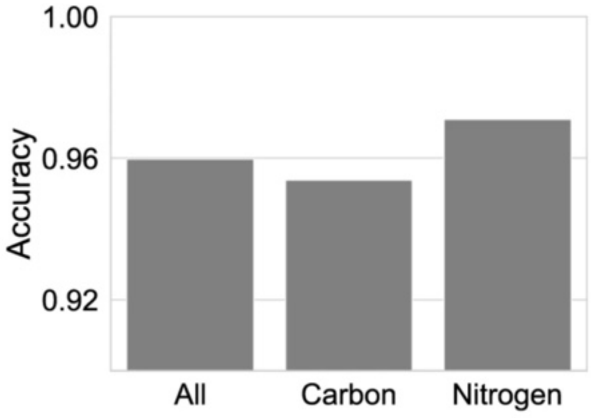
- Faure L. *et al*. ***Nat Communications***, 2023

Hybrid models:
Application to P. putida

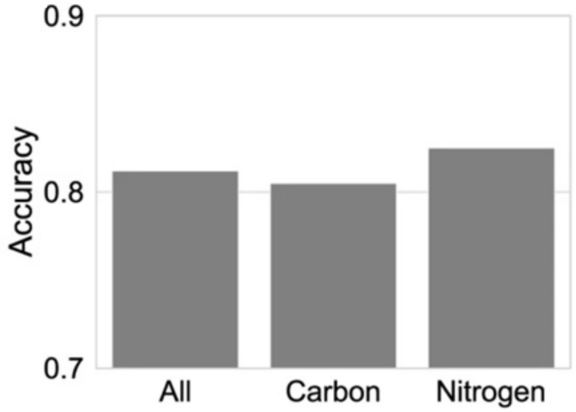


Experimental data from Nogales J. et al. *Applied Microbiology* 2019

Hybrid model results



Mechanistic model results

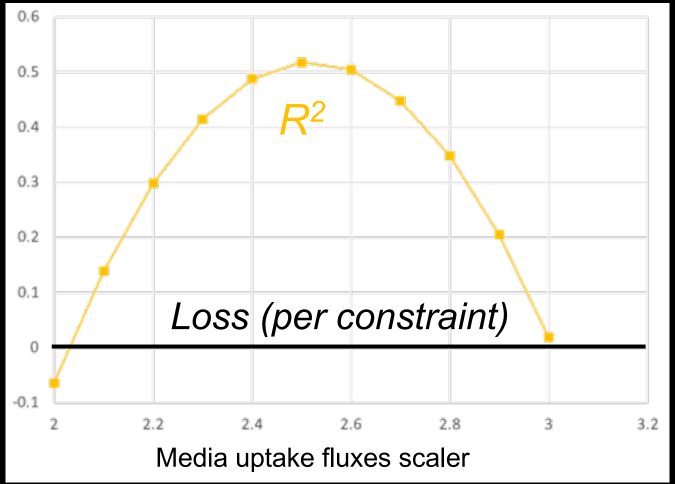


- Faure L. *et al*. ***Nat Communications***, 2023

Hybrid model: a new modelling paradigm

Training set : reaction flux for different media compositions *(strain E coli MG1655, model = iML1515, media = M9 + 4 nutrients chosen at random among 10, measured reaction flux is chosen at random among all reactions).*

Mechanistic model (FBA Cobrapy)
Training set not needed

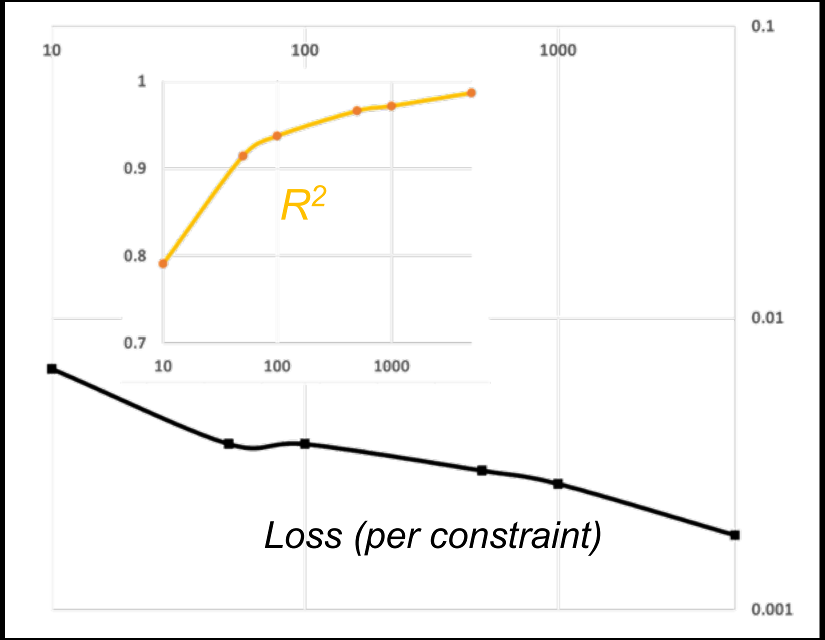


Mechanistic constraints are respected but poor match with measured (training) data

Black –box neural model (dense-ANN)
> 500k entries needed in training

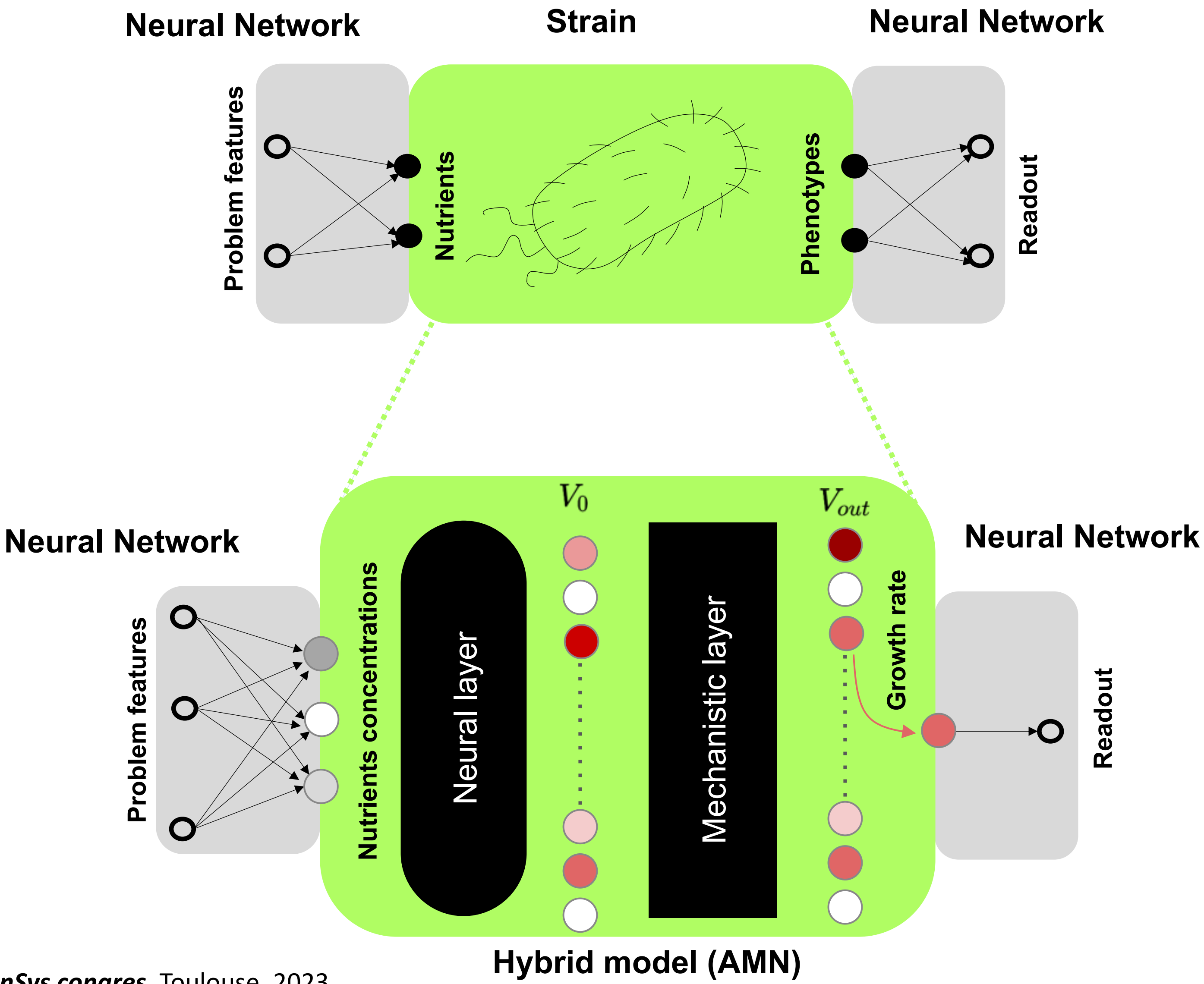
Predictions match measured training data but do not fit mechanistic constraints

Hybrid-model
Less than 1000 entries in training are enough



- Faure L. *et al*. ***Nat Communications***, 2023

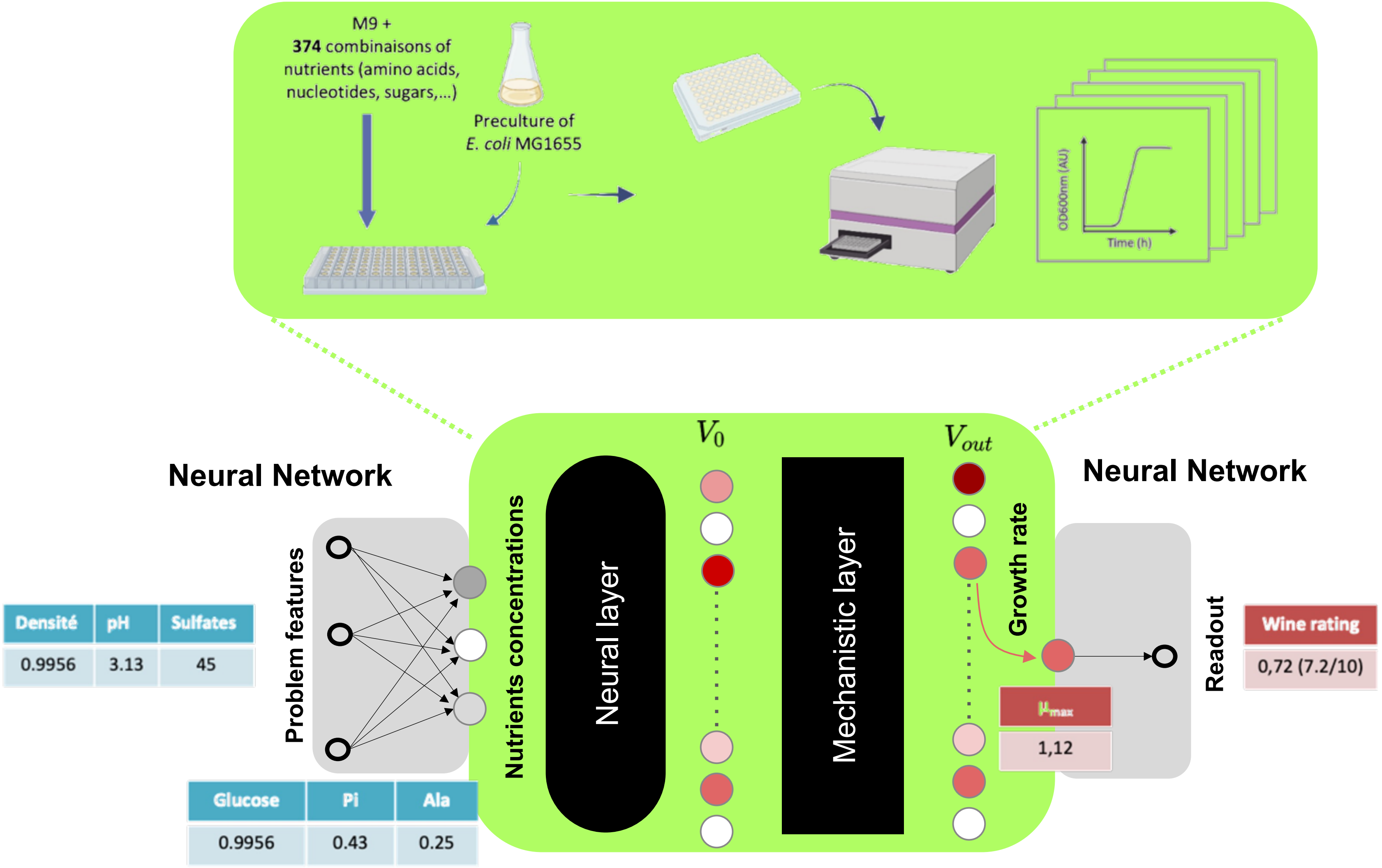
Can microorganism metabolism be diverted to solve classical machine learning problems?



- Ahavi P, Mollet B *et al*. ***BioSynSys congres***, Toulouse, 2023

Can microorganism metabolism be diverted to solve classical machine learning problems?

Example : White wine quality dataset (score from 1 to 10)



Densité	pH	Sulfates
0.9956	3.13	45

Glucose	Pi	Ala
0.9956	0.43	0.25

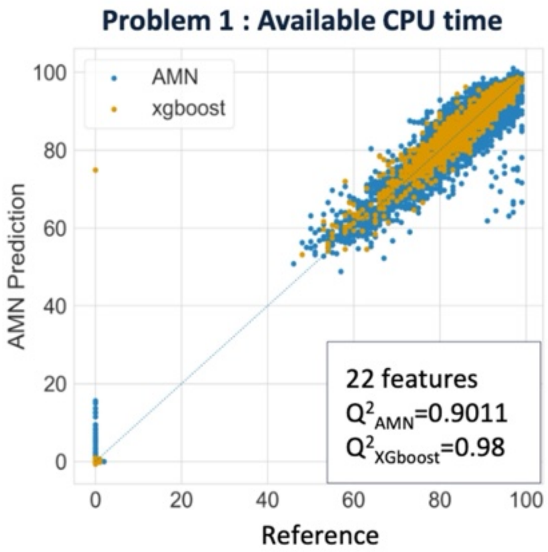
μ_{max}
1,12

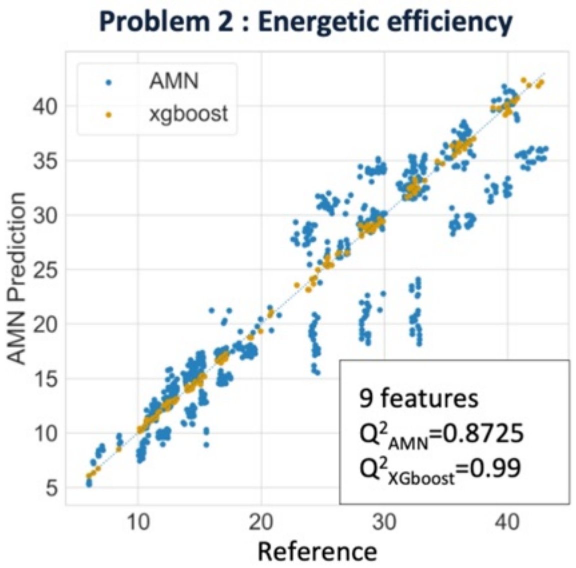
Wine rating
0,72 (7.2/10)

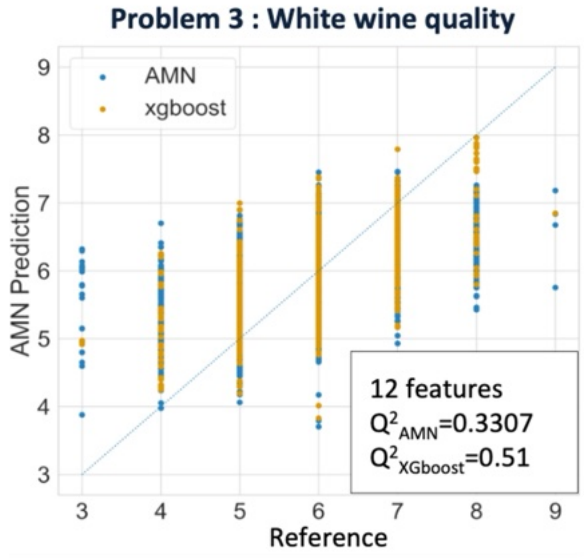
Hybrid model (AMN)

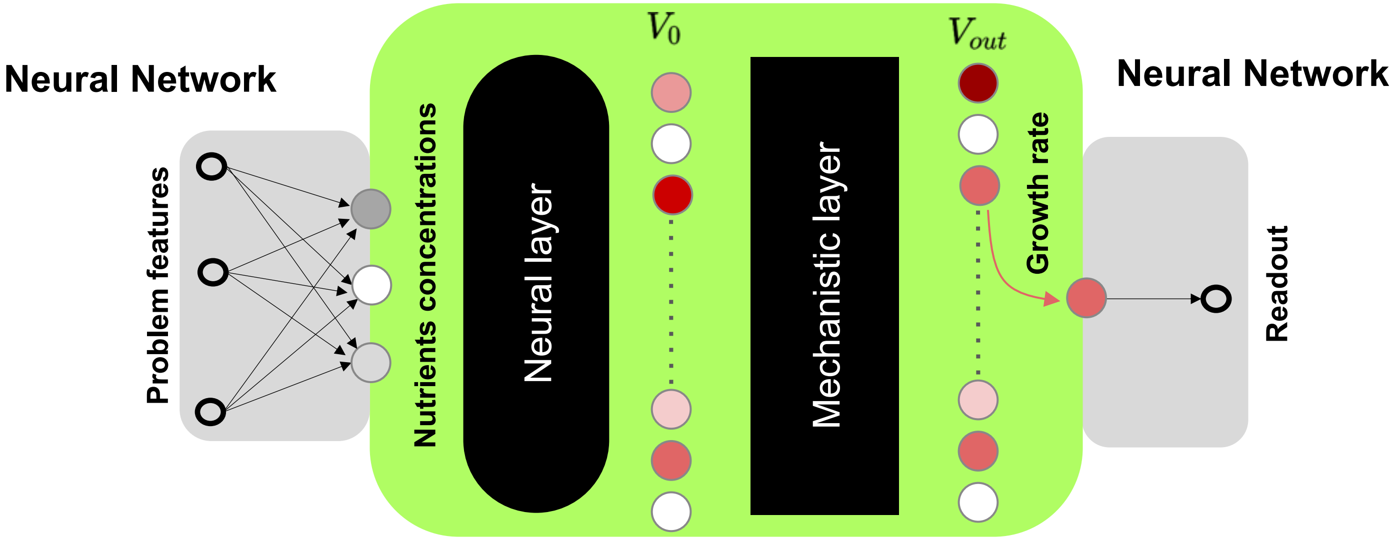
- Ahavi P, Mollet B *et al*. ***BioSynSys congres***, Toulouse, 2023

Can microorganisms solve classical machine learning problems?









Hybrid model (AMN)

- Ahavi P, Mollet B *et al.* ***BioSynSys congres***, Toulouse, 2023

More pragmatic (and funded) applications for hybrid modelling

Bioproduction

- Find the best gene deletion in *P. putida* to produce terpenes, biodegradable polyesters (polylactic acid, polyethylene furanoate), and methylacrylate (a building block for plexiglass)

HORIZON BIOS SUSTAINABLE BIO-MANUFACTURING

Biodegradation for biotherapy

- Find the best gene deletion in *E. coli Nissle* 1917 engineered strain to degrade p-cresol and 4-ethylphenol (two metabolites involved in Autism)

anr

AMN
SynBioDiag

Diagnostic

- Find the best gene deletion in *E. coli DH5 alpha* to classify (benign from severe) Covid-19 samples

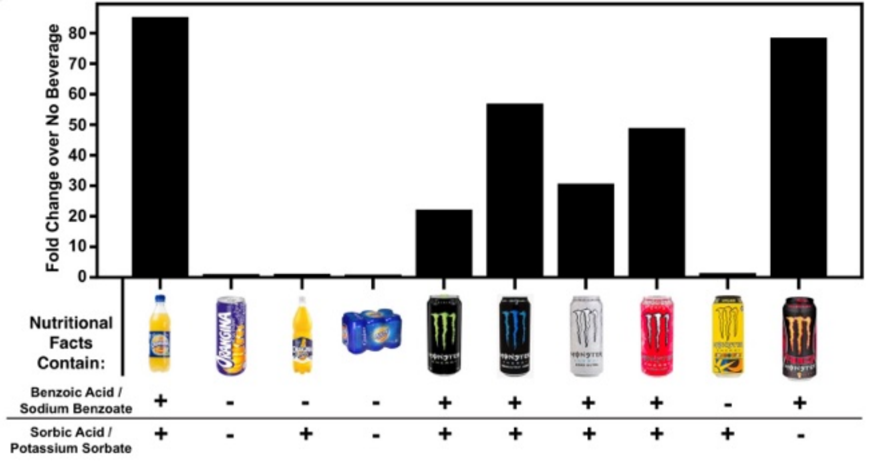
- Find the best gene deletion in *E. coli* pilot strain (an engineered strain with CTCs receptors breast cancer) which upon binding produces violacein derivatives (molecules detectable by SERS)

HORIZON BIO CELLPHE

★ Molecular Biology

★ Computational Biology

Benzoic acid (E210) is used as a food preservative
Biosensor detection in commercial beverages



We have internships and PhD scholarships to propose

Jean-Loup.Faulon@inrae.fr

https://jfaulon.com