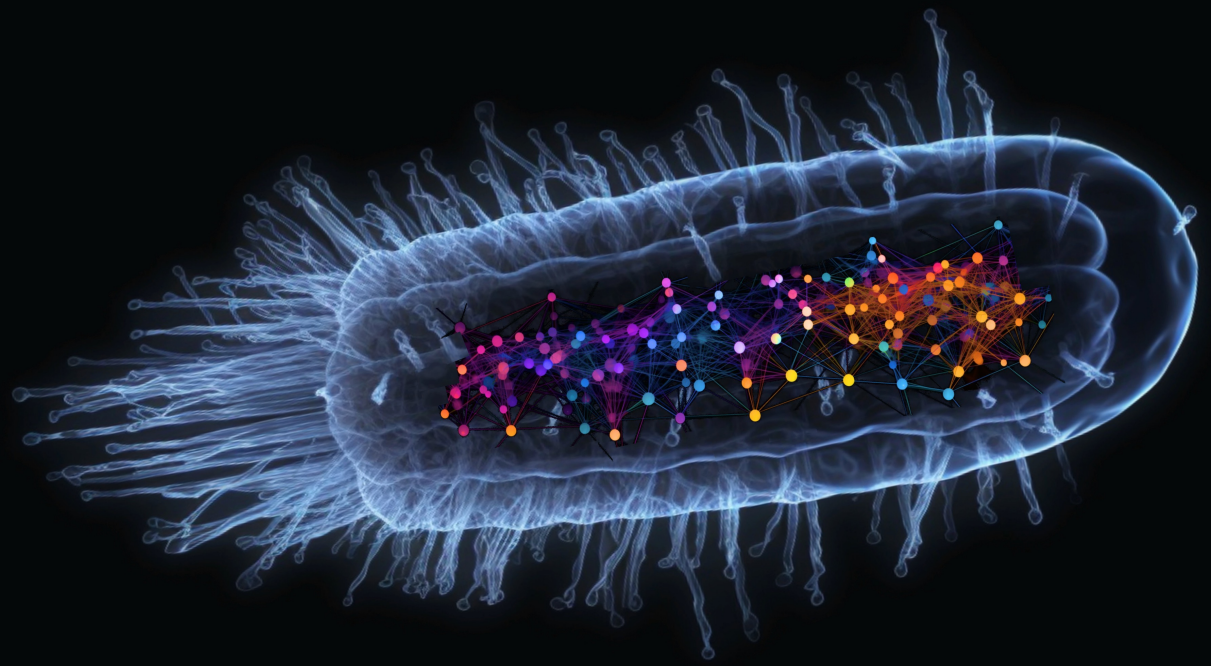


Reservoir Computing with Bacteria

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

université PARIS-SACLAY

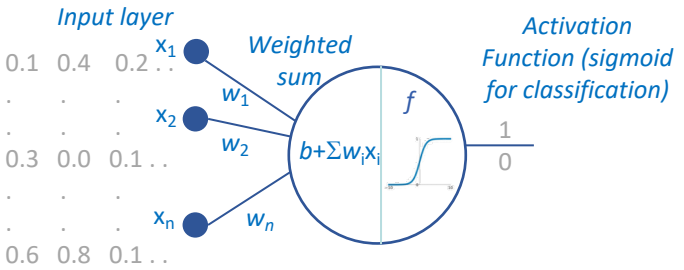
MICALIS Institute

INRAE

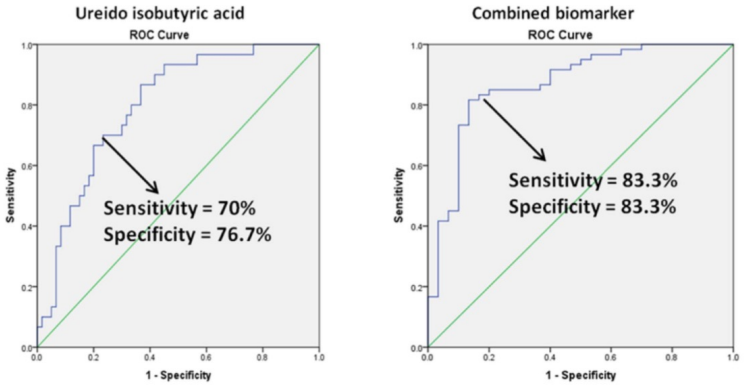
Engineering biological information processing devices

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

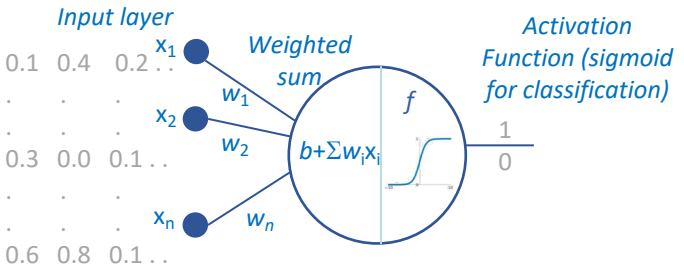
$$f\,(b+\Sigma w_i x_i\,)$$

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

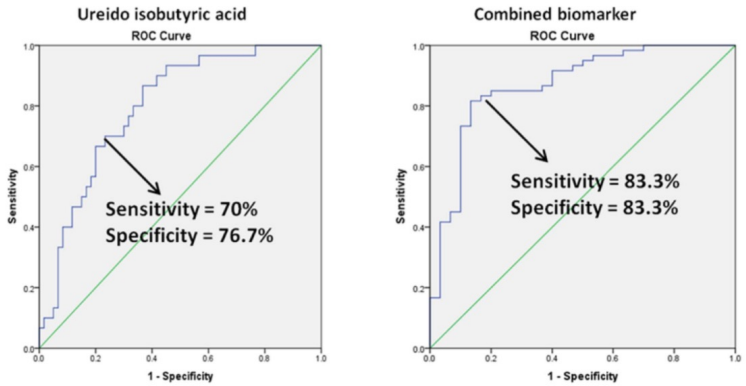
Engineering a metabolic information processing device: the concept

TRAINING THE NETWORK

Perceptron weights (w_i) are learned to increase classifier accuracy



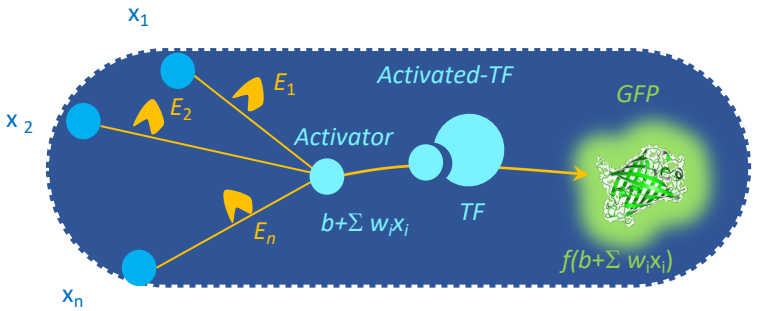
Prostate cancer $\text{PCa metabolic score} = b + \sum_{j=1}^{j} w_j x_{ij}$



- 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

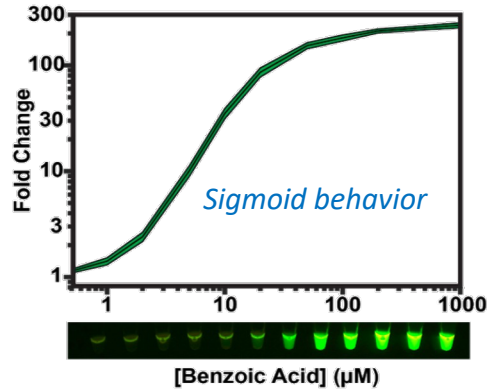


Enzymatic transformation *Reporter gene*

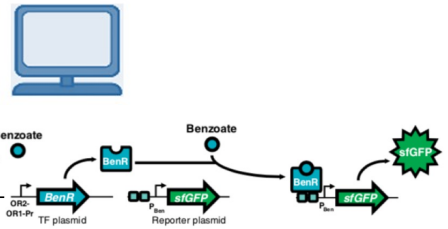
In theory (Michaelis-Menten) when $x_i << [E_i]$:

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

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



Engineering a metabolic information processing device: the concept



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

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

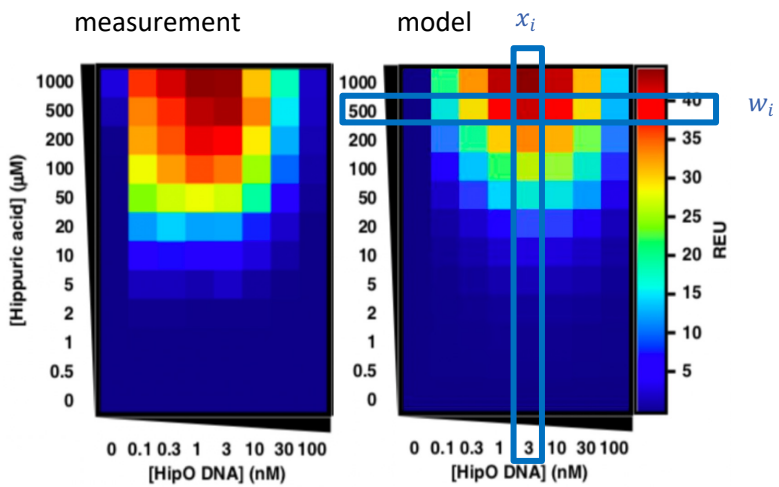
$$TF_{activated} = TF * \frac{benzoate}{benzoate + K_d^{inducer}} + 0.0005$$

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

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

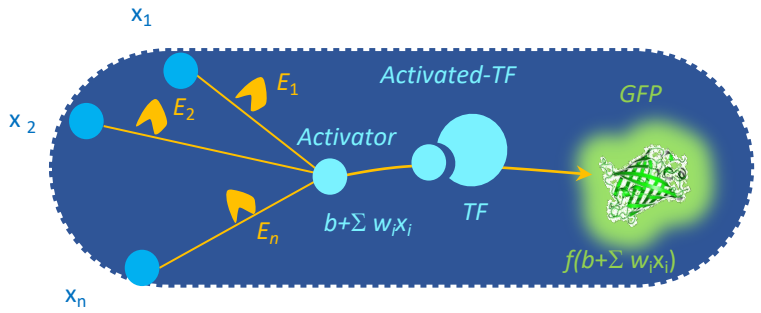
$$\frac{dmRNA}{dt} = \gamma * n * \epsilon \frac{x}{x + \chi} * \frac{K_{tox}}{K_{tox} + tox} * \frac{R_{mRNA}}{R_{mRNA} + K_{mRNA}} - \delta * mRNA$$

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



ENGINEERING THE TRAINED NETWORK

Need to actuate weighted sum and activation function

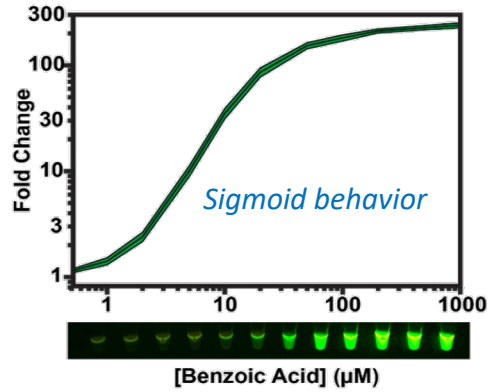


Enzymatic transformation *Reporter gene*

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

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

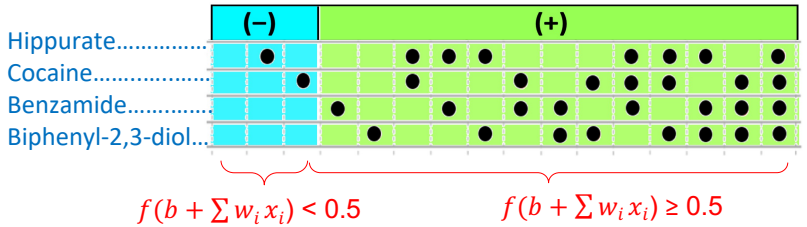
$w_i = k_i\,[E_i]$
where k_i is a kinetics constant



- Voyvodic, P.L., Pandi, A., Koch, M. *et al. Nat Commun* **10**, 1697 (2019).

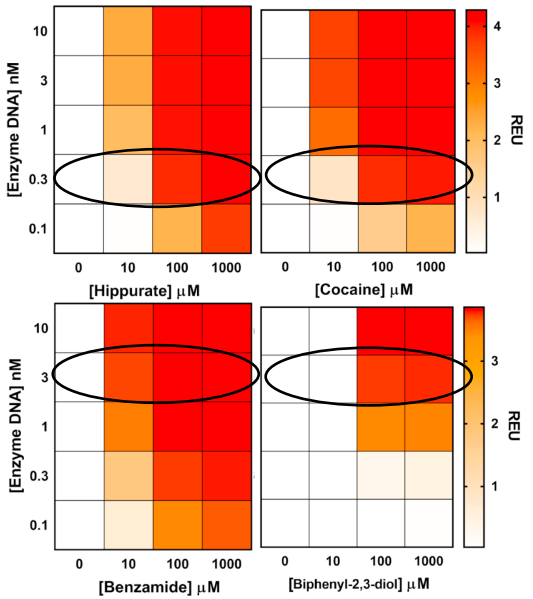
Engineering a metabolic information processing device in vitro


TARGETED BEHAVIOUR



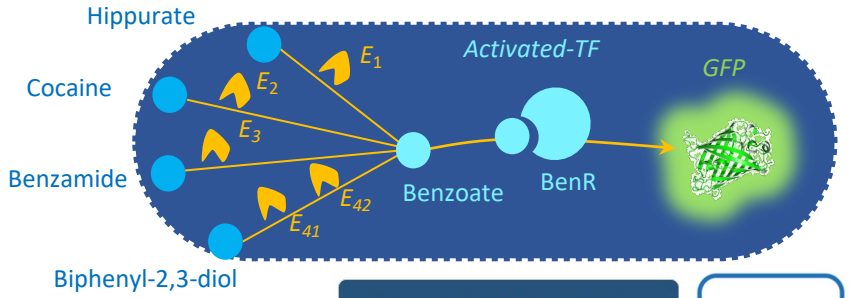
$f(b + \sum w_i x_i) < 0.5$ $f(b + \sum w_i x_i) \geq 0.5$

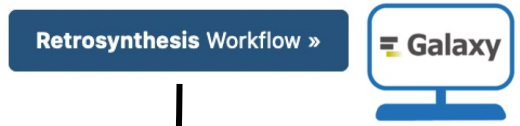
- Kinetics model is used to compute the enzyme concentration for each weight



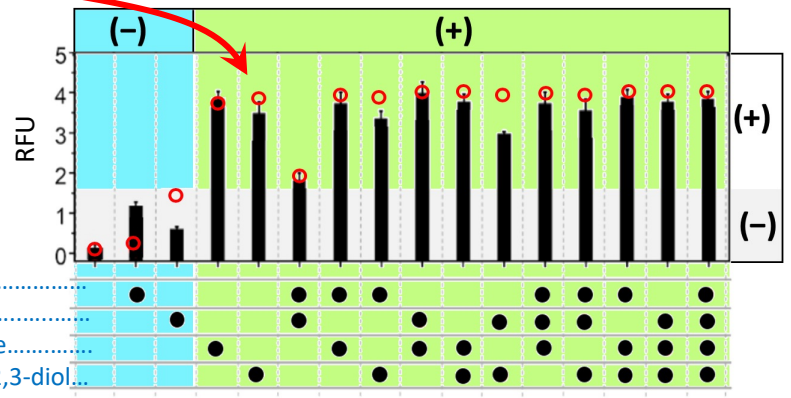
Logistic regression

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





TARGETED BEHAVIOUR

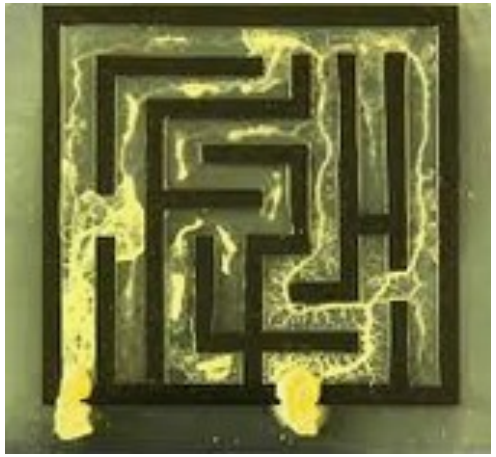


OBSERVED BEHAVIOUR

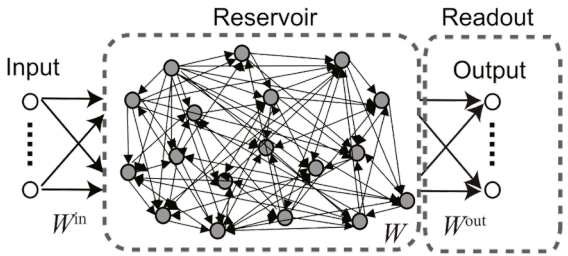
- Pandi A., Koch M. *et al*. *Nat Commun* **10**, 3880 (2019)

Engineering biological information processing devices in vivo?

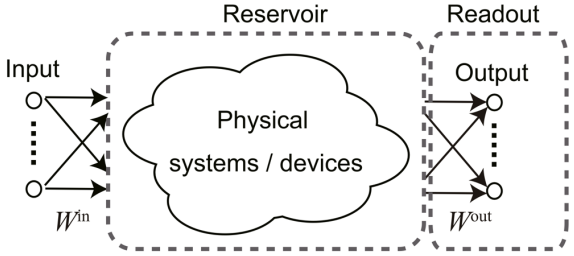
- Can we divert native metabolism to handle problems that are usually solved *in silico*?



- Ability of physical, chemical or biological devices to solve problems is studied in AI with Reservoir Computing (RC)

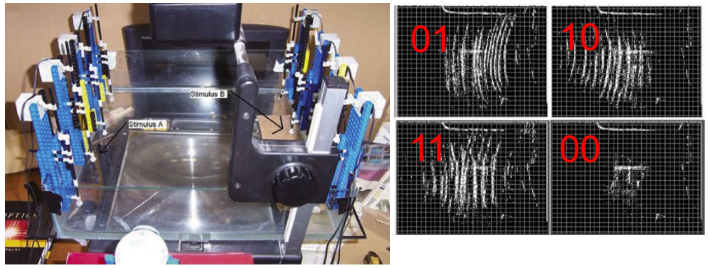


(a) Conventional RC



(b) Physical RC

Tanaka G. et al. *Neural Networks* **115**, 100 (2019)



Liquid state machine (2002)

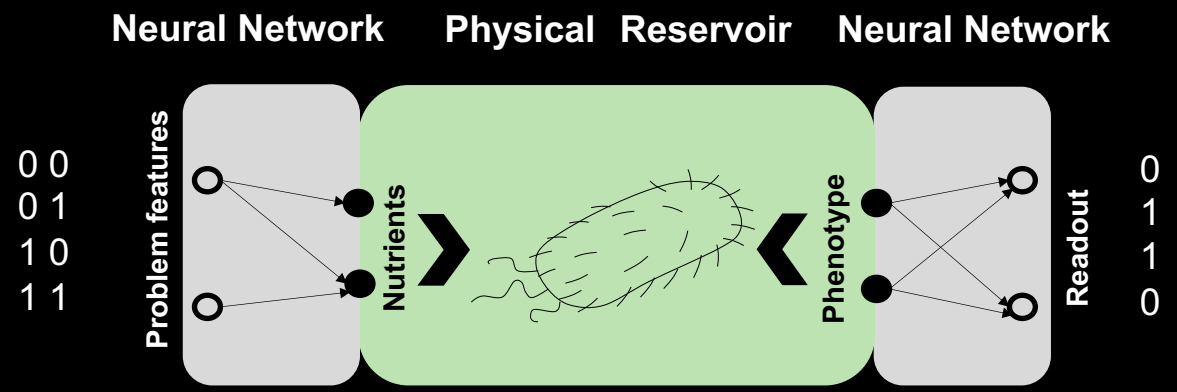
XOR

E. coli Reservoir Computer (E. coli RC)

Can we exploit *E. coli* native metabolism to build an *E. coli* RC to solve computational problems?

How complex a problem can *E. coli* RC solve?

Can we find practical uses of *E. coli* RC?



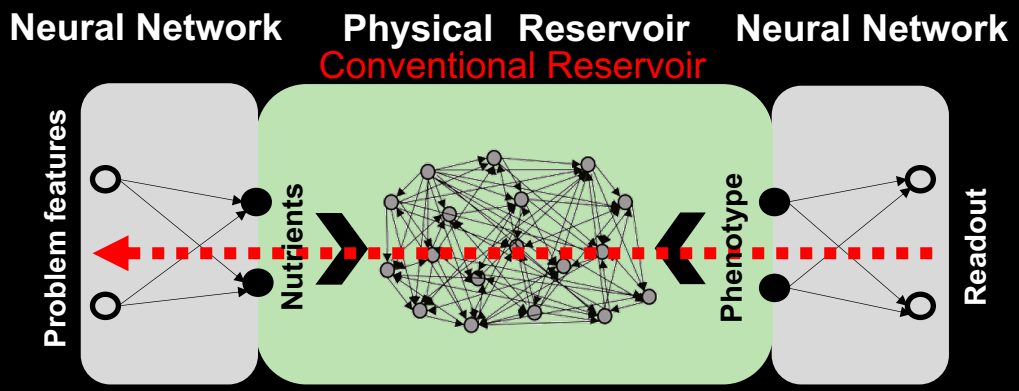
E. coli Reservoir Computer (E. coli RC)

Can we exploit *E. coli* native metabolism to build an *E. coli* RC to solve computational problems?

How complex a problem can *E. coli* RC solve?

Can we find practical uses of *E. coli* RC?

gradient backpropagation



Conventional Reservoir should:

- *accurately reproduce phenotype for different media composition*
- *enable gradient backpropagation*

The reservoir

GEnome-scale Metabolic Model (GEM/FBA)

Max ($v_{biomass}$)

Subjected to contraints:

$S\,V = 0$

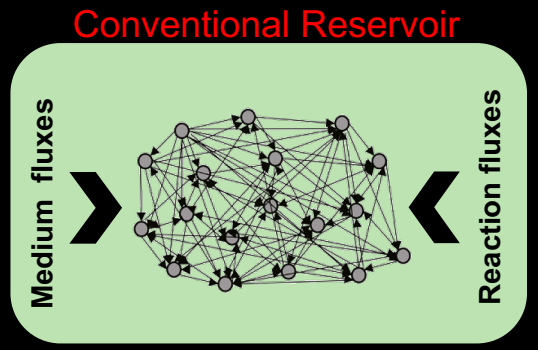
$0 \leq V \leq V_{in}$

where

$-\ V =$ set of all reaction fluxes

$-\ S\ =$ stochiometric matrix

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



The reservoir

GEnome-scale Metabolic Model (GEM/FBA)

Max $(v_{biomass})$

Subjected to contraints:

$S\,V = 0$

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

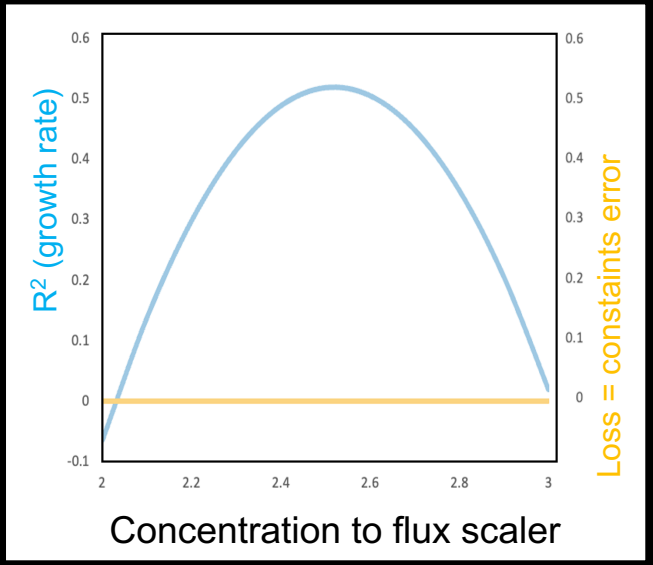
where

$-\ V$ = set of all reaction fluxes

$-\ S$ = stochiometric matrix

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

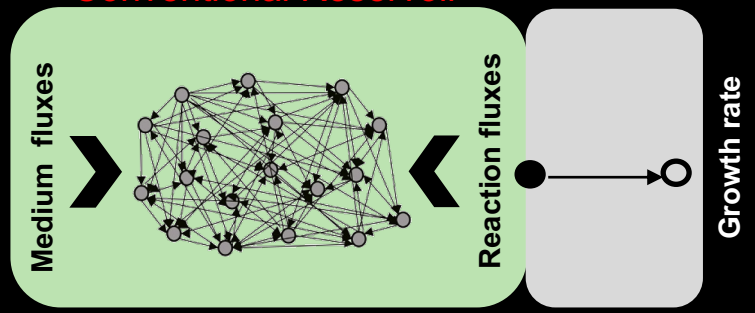
GEM/FBA growth rates vs. measured growth rate in E. coli MG1655 for 1 to 4 nutrients added to M9



Medium concentrations

?

Conventional Reservoir



Conventional Reservoir should:

- *accurately reproduce phenotype for different media composition*

Building an E. coli RC to increase mechanistic model predictability

Linear Program solved using Simplex algorithm not compatible with gradient propagation

GEnome-scale Metabolic Model (GEM/FBA)

Max $(v_{biomass})$

Subjected to contraints:

$S\,V = 0$

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

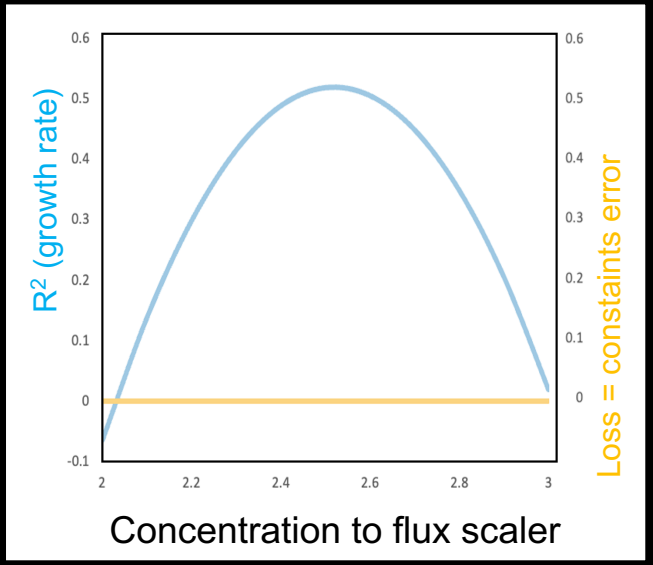
where

$-\ V$ = set of all reaction fluxes

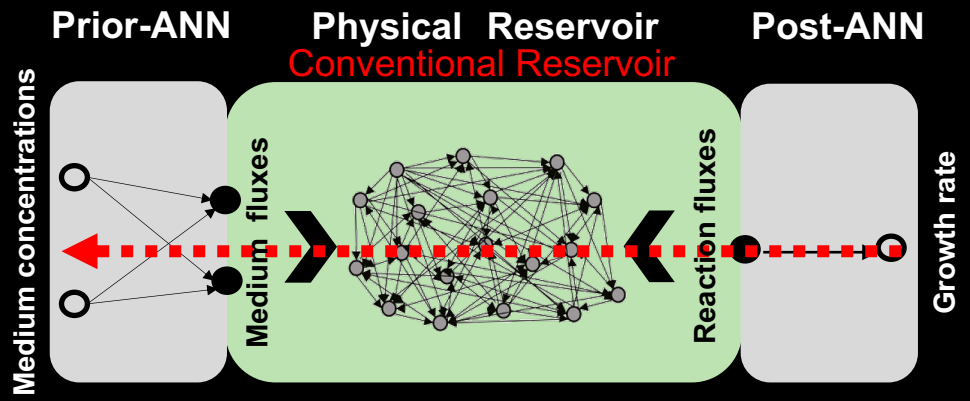
$-\ S$ = stochiometric matrix

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

GEM/FBA growth rates vs. measured growth rate in E. coli MG1655 for 1 to 4 nutrients added to M9



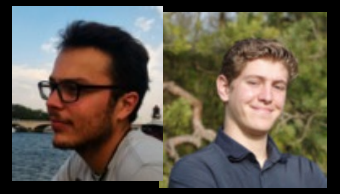
gradient backpropagation to find mapping between medium concentrations and uptake fluxes



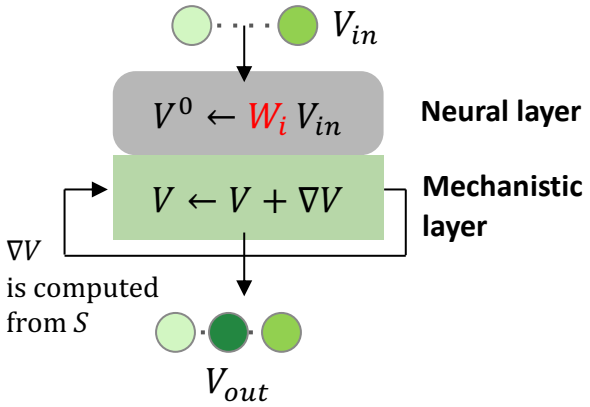
Conventional Reservoir should:

- *accurately reproduce phenotype for different media composition*
- *enable gradient backpropagation*

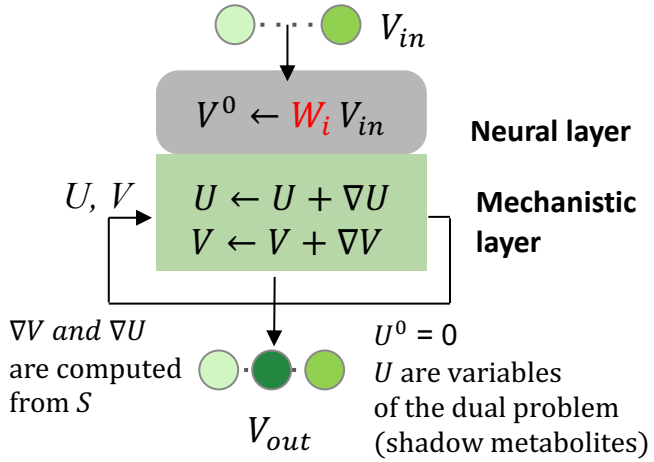
AMNs (Artificial Metabolic Networks): gradient backpropagation compatible methods surrogating FBA



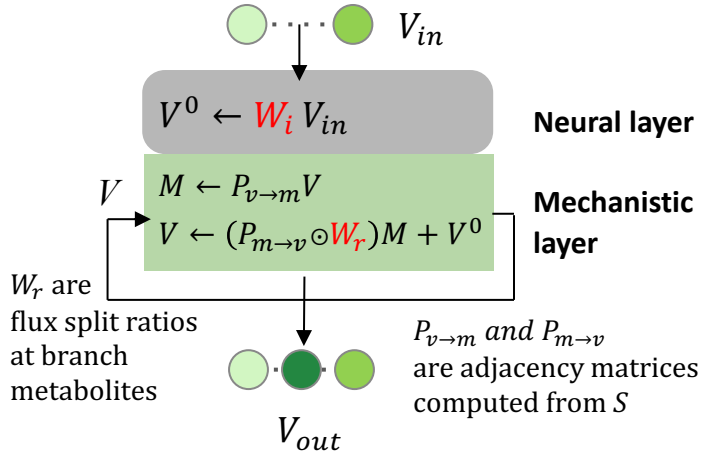
Physics informed neural network (PINN)



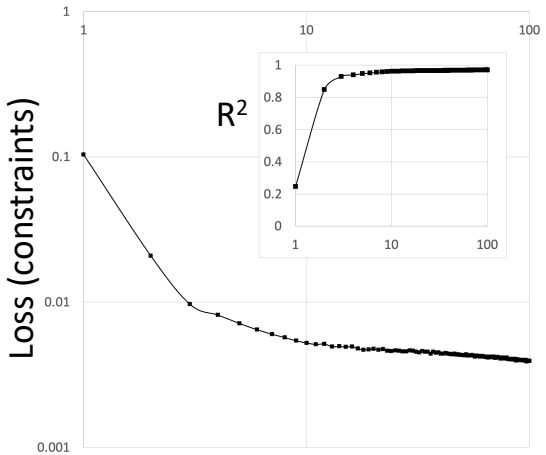
Hopfield's network

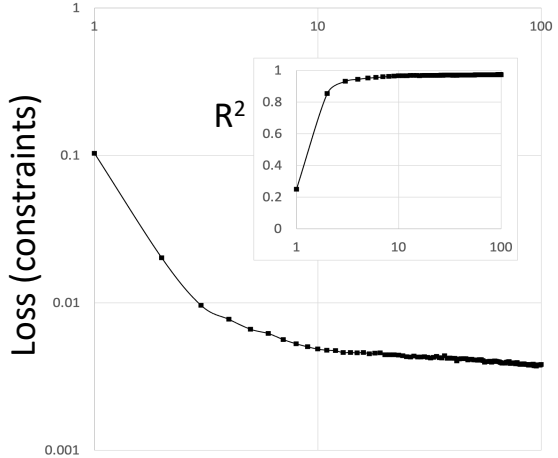


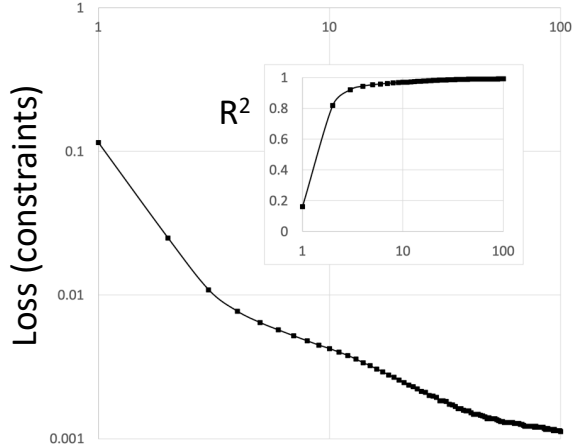
Graph neural network (GNN)



Trained on GEM/FBA (cobrapy) calculated growth rates with *E. coli* iML1515 model for 1000 different media (M9 + random combinations of nutrients among sugars, nucleotides, amino acids)

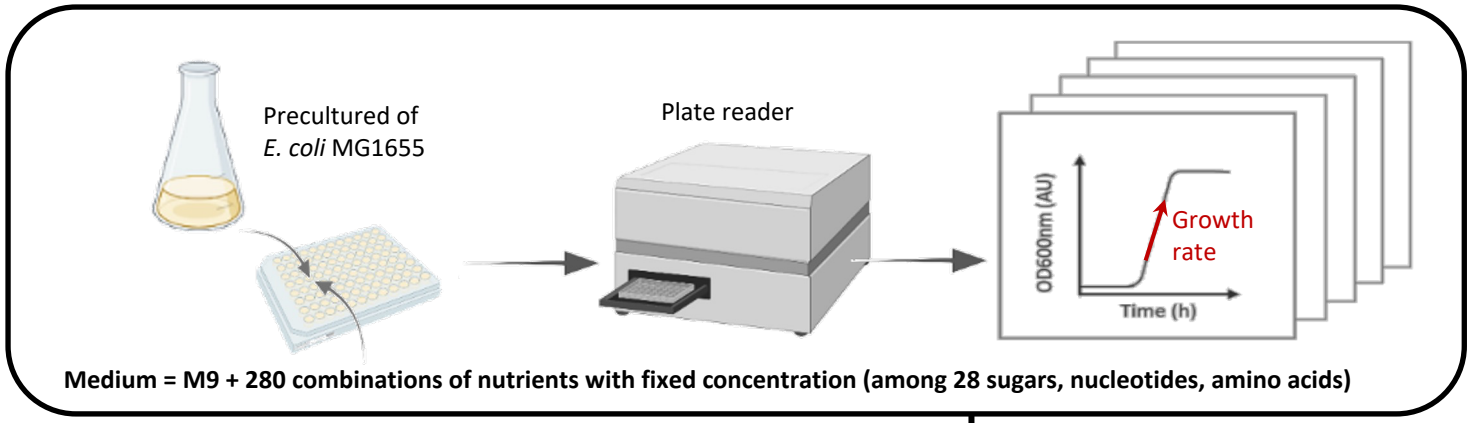


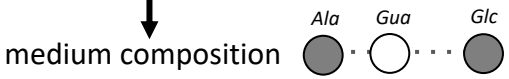




- Faure L. *et al*. *Nat Commun* **14**, 4669 (2023)

AMNs can be used as reservoir in RC to improve mechanistic model predictability

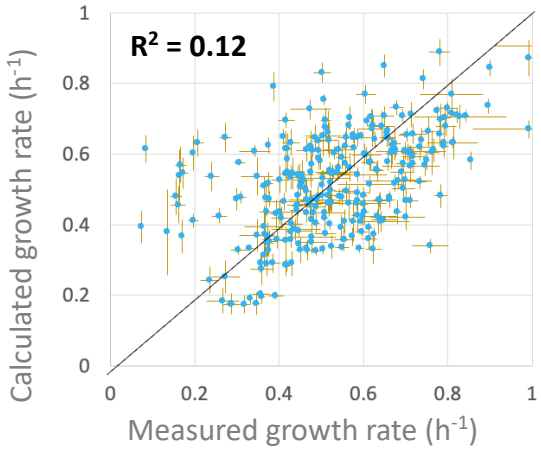




Scaler to medium fluxes

cobrapy
***E. coli* GEM/FBA**

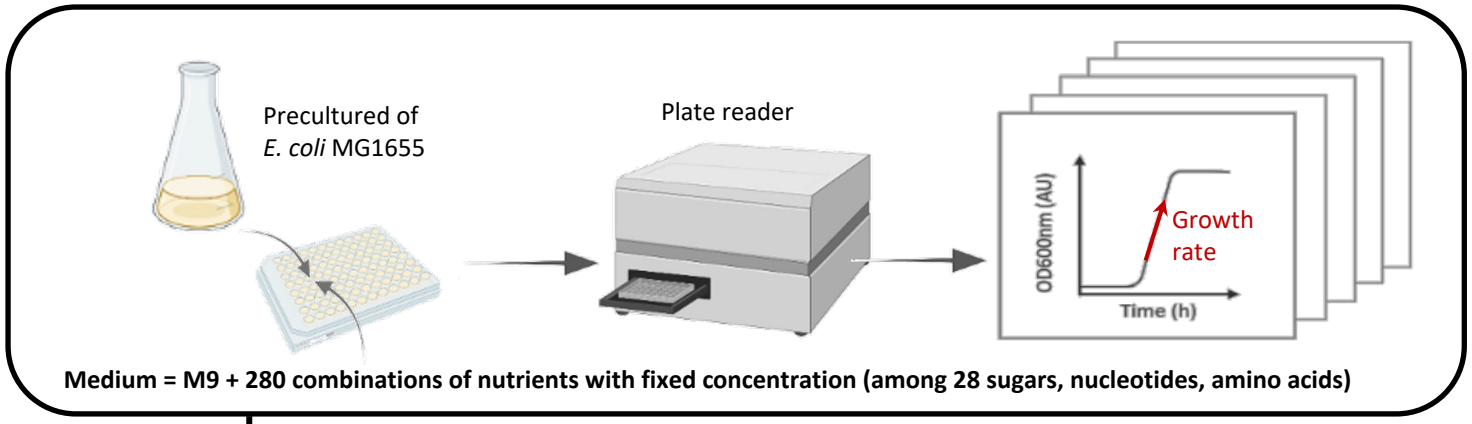
Calculated growth rate



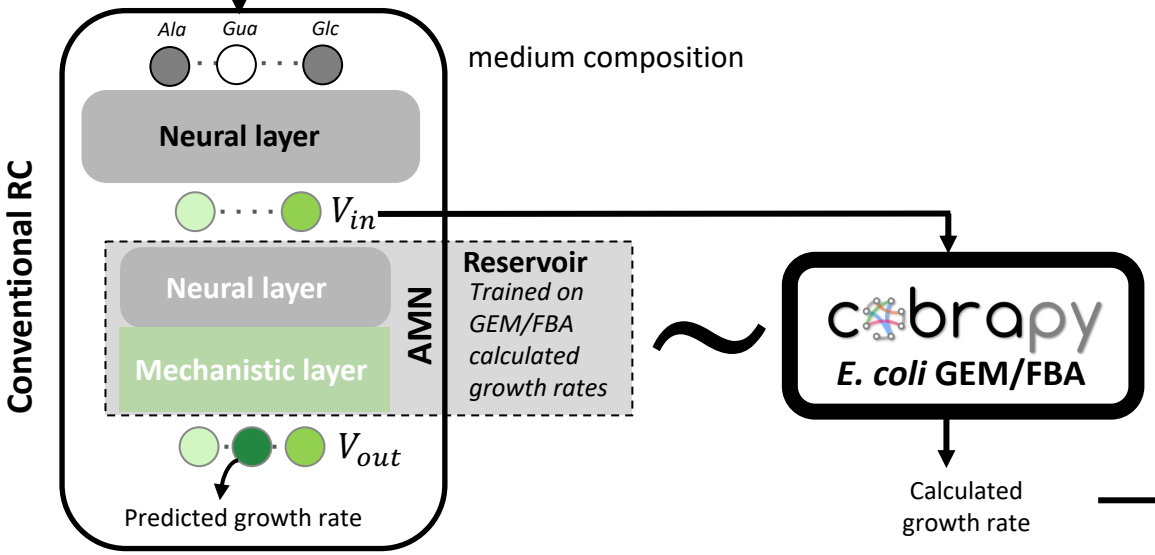
GEM/FBA results with best scaled input

- Faure L. *et al*. *Nat Commun* **14**, 4669 (2023)

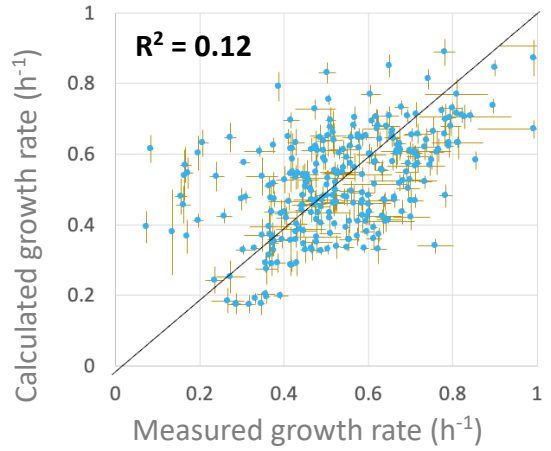
AMNs can be used as reservoir in RC to improve mechanistic model predictability

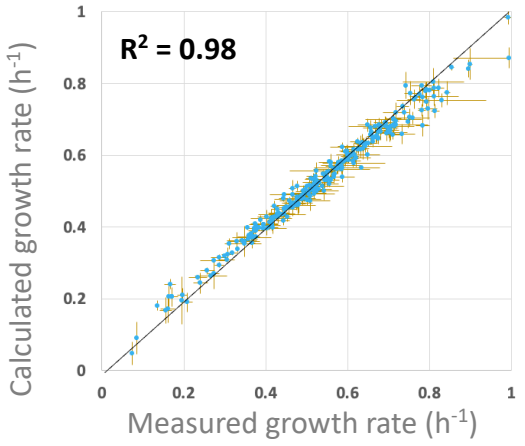
Training on 280 medium compositions



Calculated growth rate



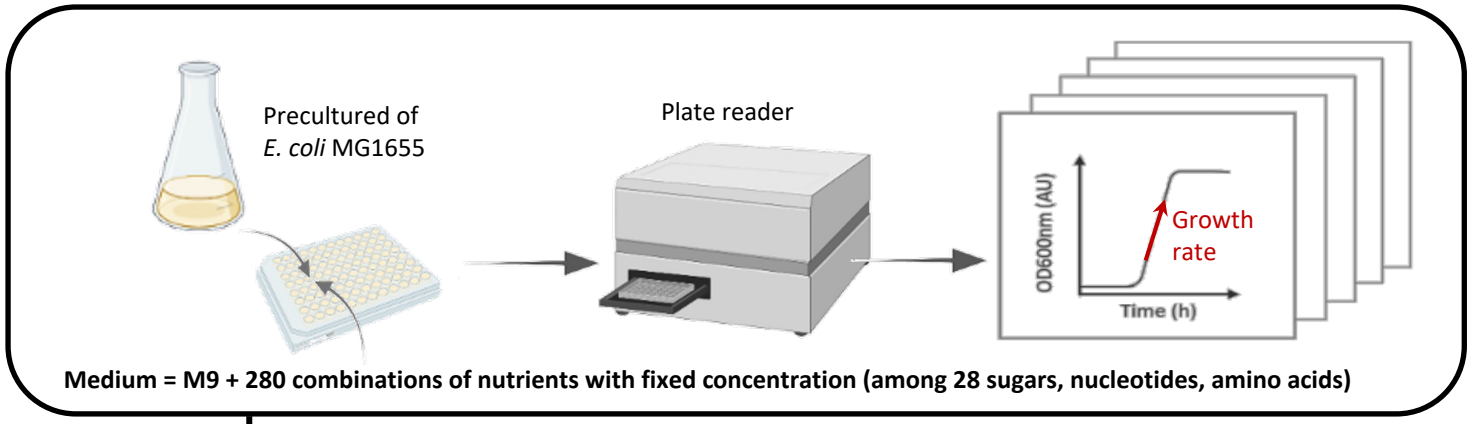
GEM/FBA results with best scaled input



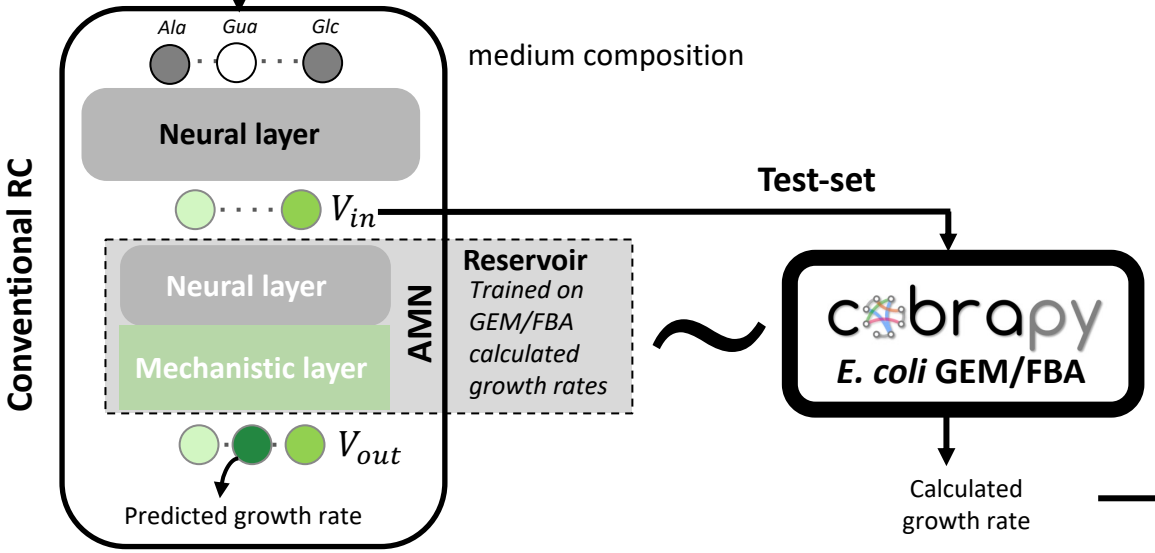
GEM/FBA results with reservoir inputs

- Faure L. *et al*. *Nat Commun* **14**, 4669 (2023)

AMNs can be used as reservoir in RC to improve mechanistic model predictability

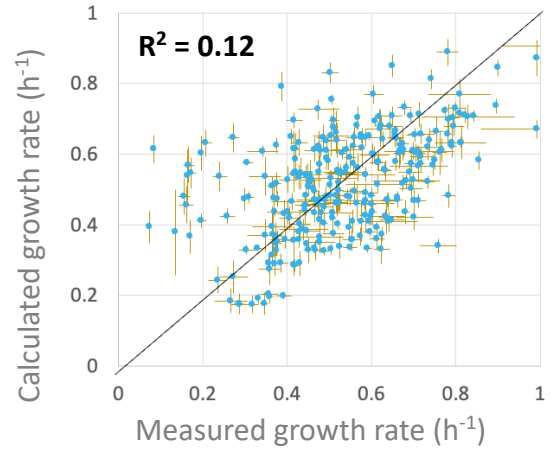
Training-set 80% - Test-set 20% (5-fold CV)



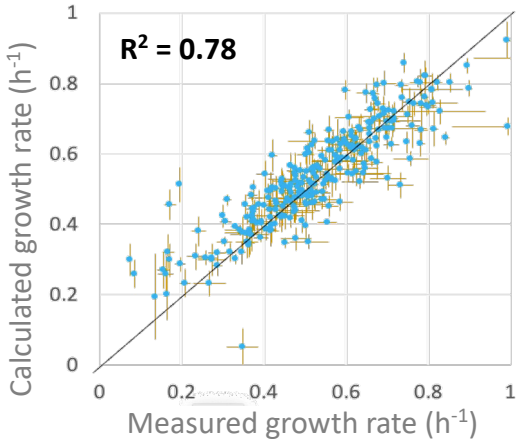
Test-set

Calculated growth rate

Test-set



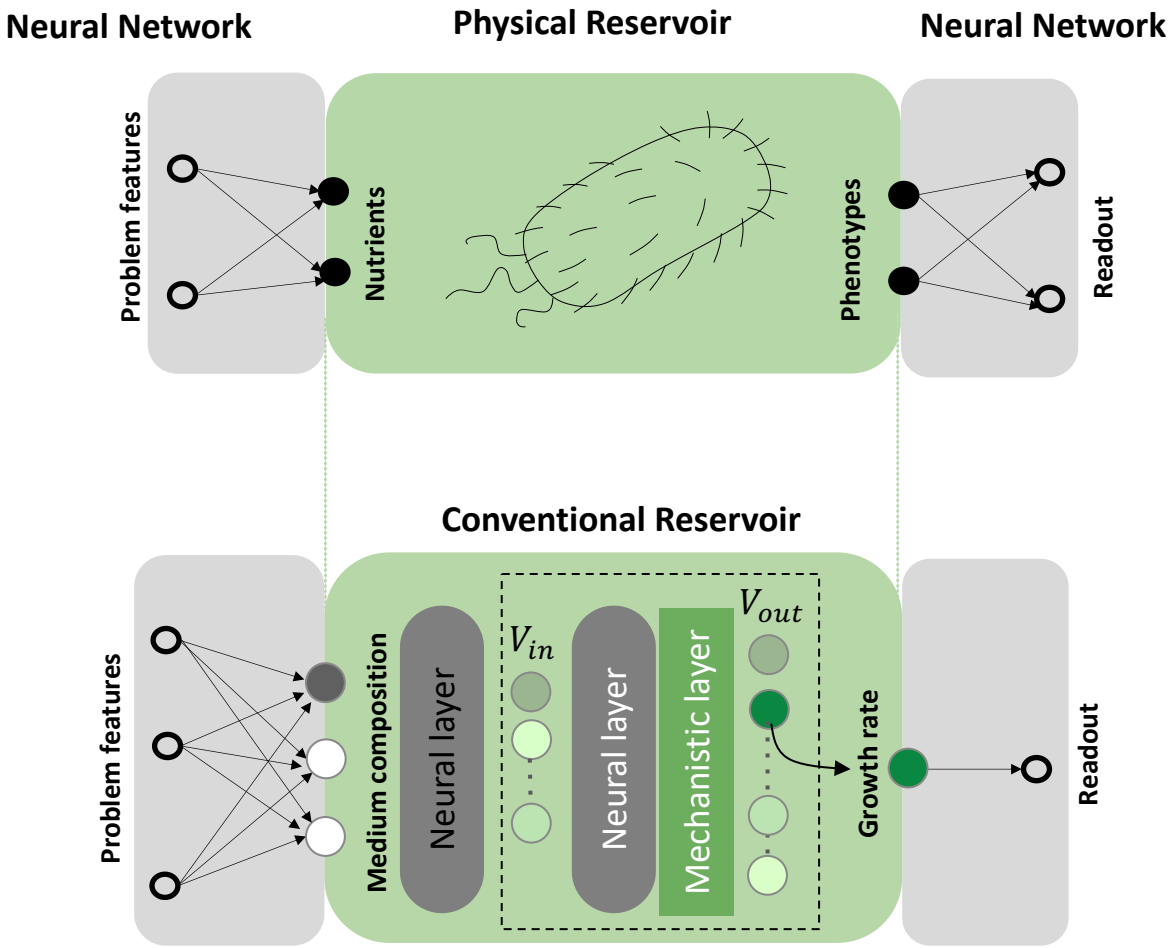
GEM/FBA results with best scaled input



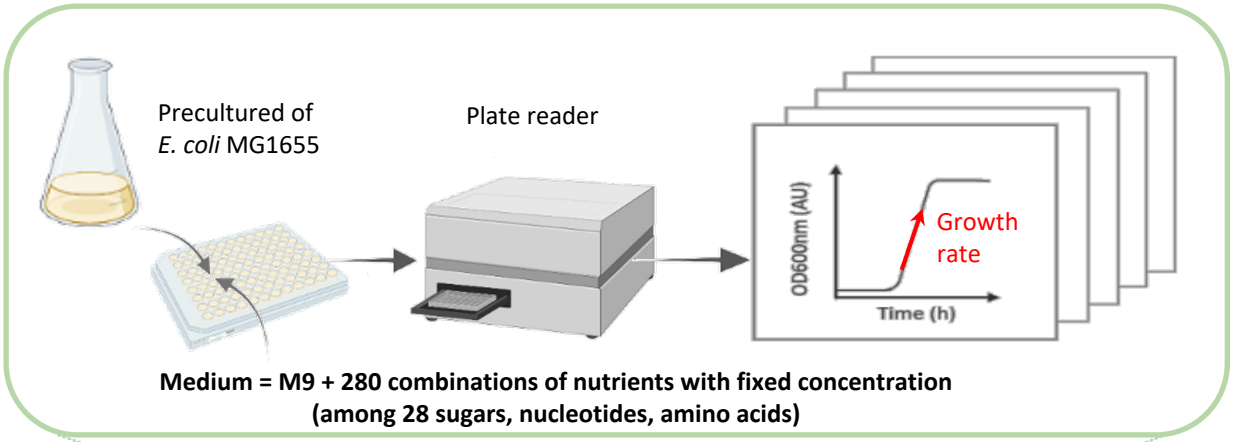
GEM/FBA results with reservoir inputs

- Faure L. *et al*. *Nat Commun* **14**, 4669 (2023)

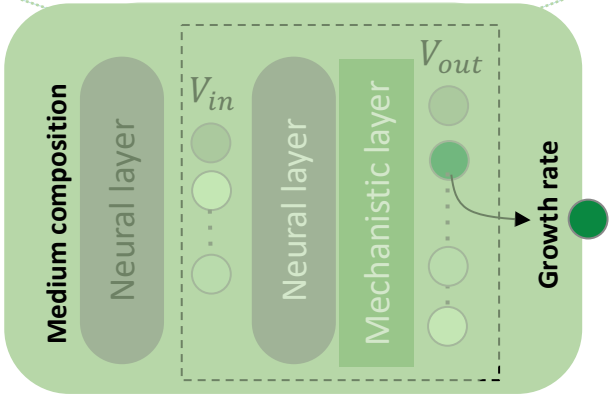
Can E. coli RC be used to solve a classical machine learning problem?



Building an E. coli RC to solve a regression problem



Conventional Reservoir

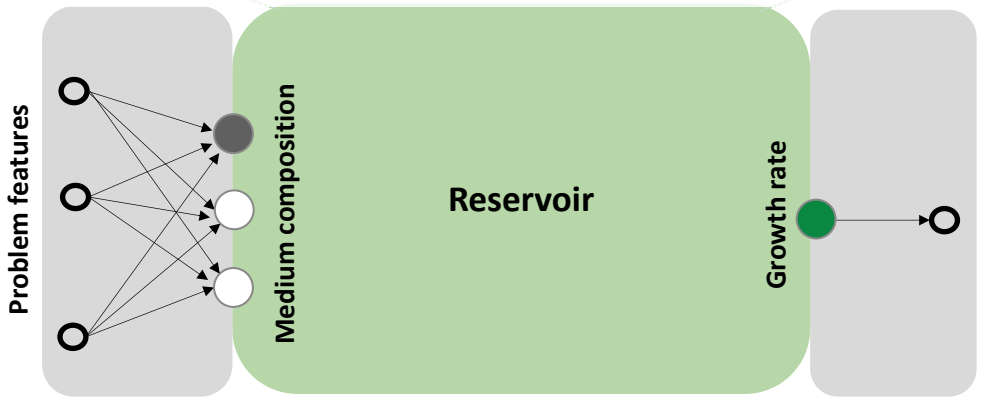


Using E. coli RC to solve a regression problem

OpenML
A worldwide machine learning lab

Example of regression problem : OpenML 'Energy Efficiency' dataset (768 instances, X = 8 features, y = % efficiency)

Compactness	Surface Area	Orientation
0.98	514.5	3



% efficiency
15.55

Glc	Xyl	Succ	Trp
1	0	1	1

μ
0.71

- Mollet, B., Ahavi P. *et al*. *Lecture Notes in Computer Science* (2025)

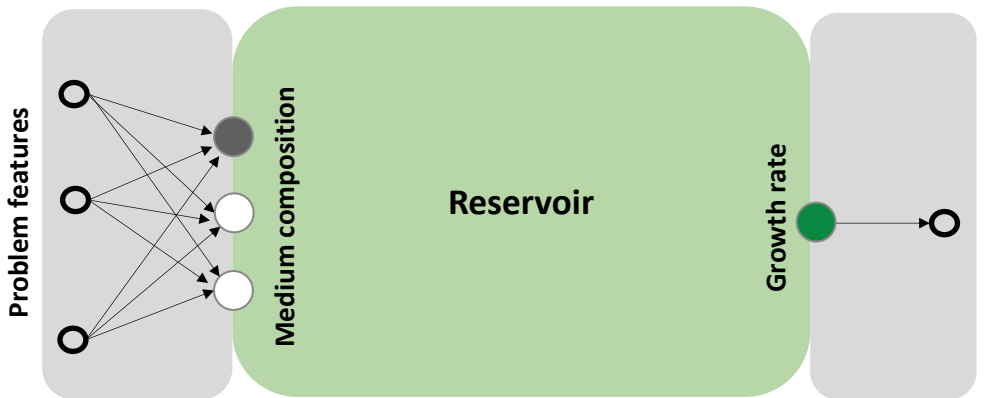
Using E. coli RC to solve a regression problem

OpenML
A worldwide machine learning lab



Example of regression problem : OpenML 'Energy Efficiency' dataset (768 instances, X = 8 features, y = % efficiency)

Compactness	Surface Area	Orientation
0.98	514.5	3



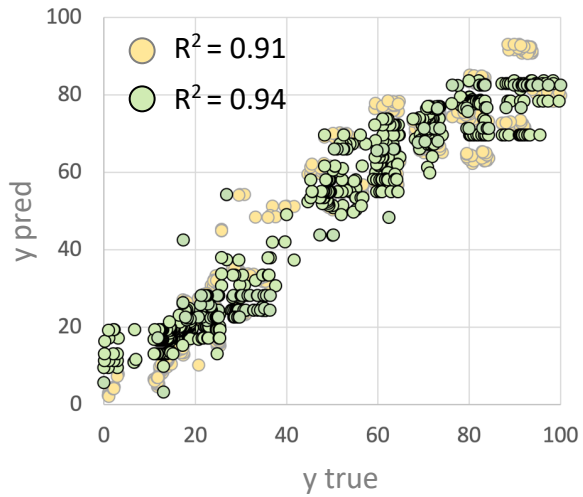
% efficiency
15.55

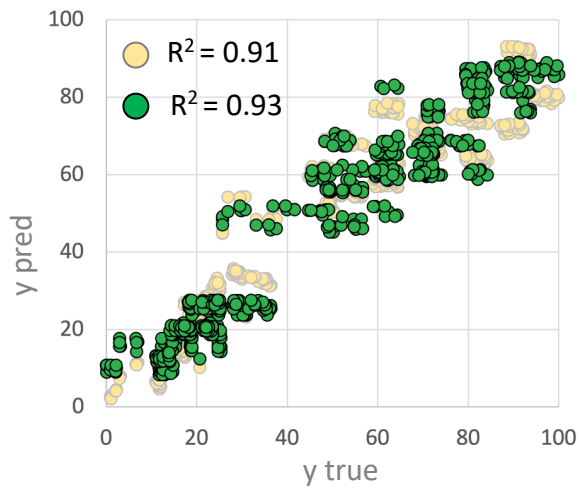
Glc	Xyl	Succ	Trp
1	0	1	1

μ
0.71

MLR
Conventional RC
Physical RC

5-fold CV





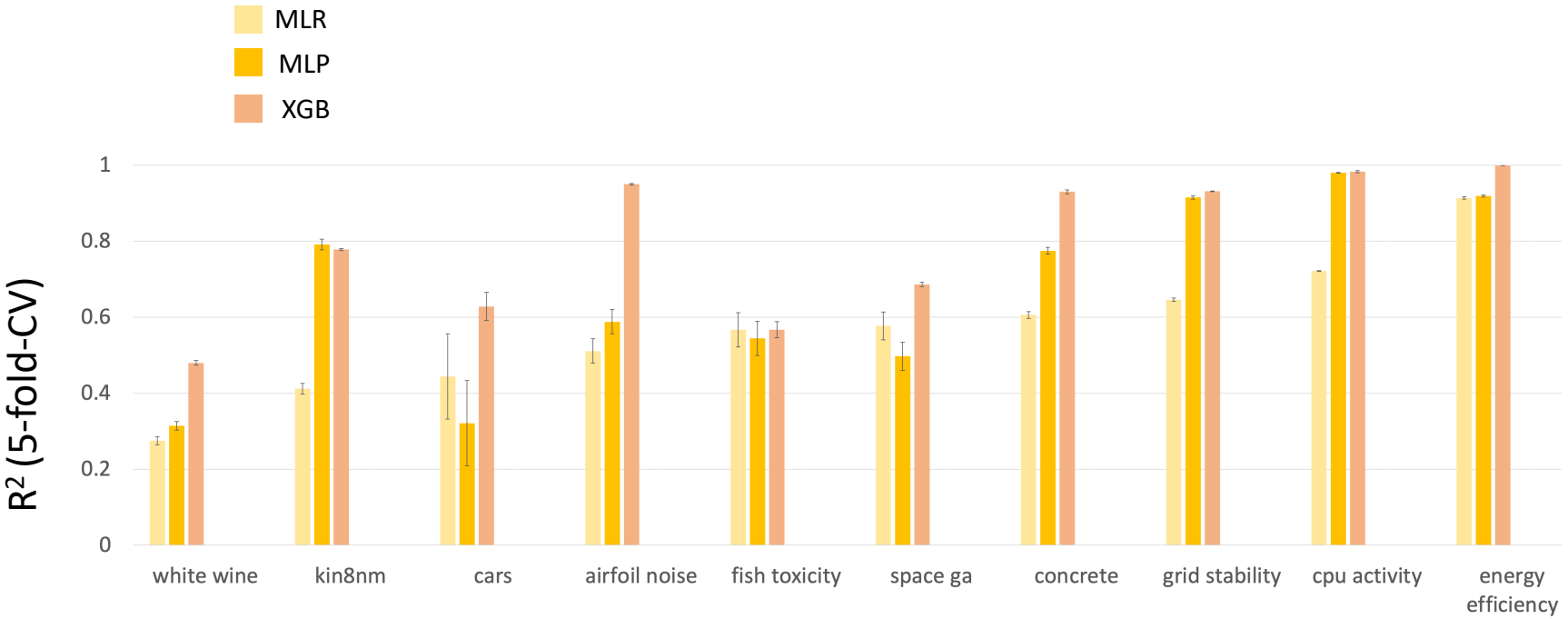
- Mollet, B., Ahavi P. *et al*. *Lecture Notes in Computer Science* (2025) and Ahavi P et al. *bioRxiv DOI: 10.1101/2024.09.12.612674* (2025)

Using E. coli RC to solve many regression problems

OpenML
A worldwide machine learning lab



10 OpenML regression problems of increasing difficulty



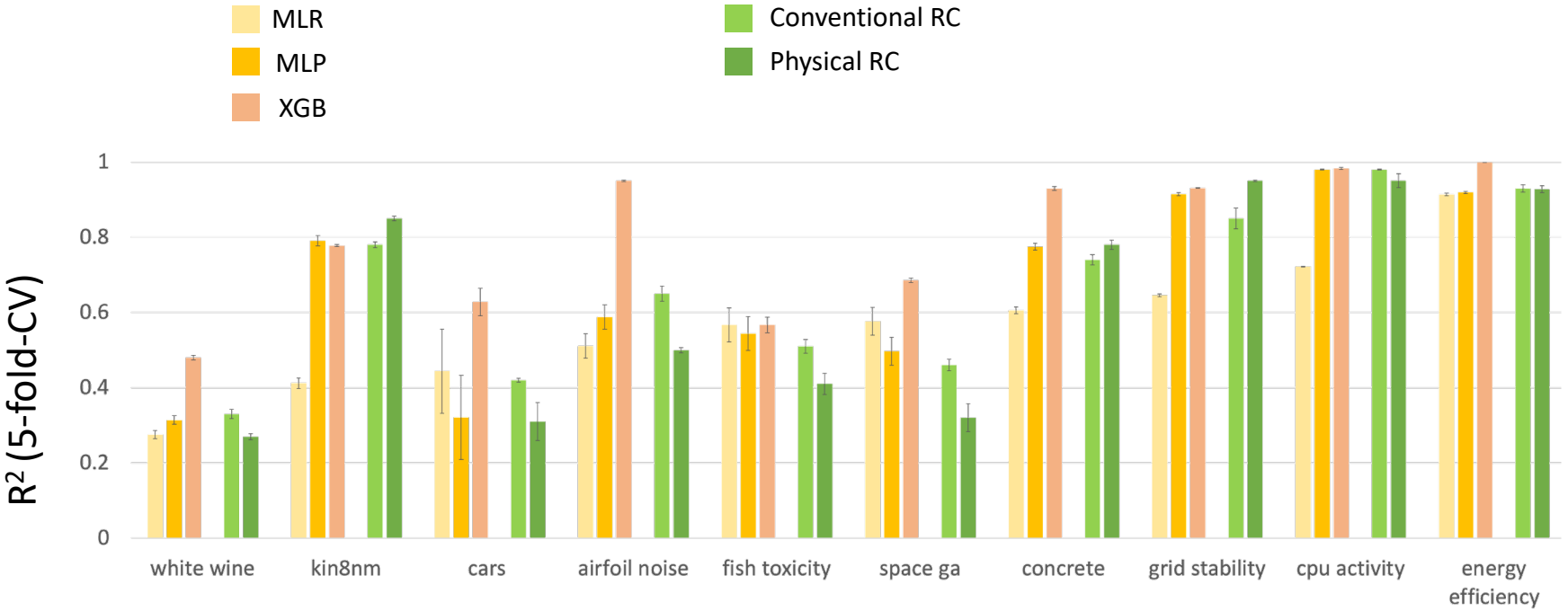
- Mollet, B., Ahavi P. *et al*. *Lecture Notes in Computer Science (*2025) and Ahavi P et al. *bioRxiv DOI: 10.1101/2024.09.12.612674* (2025)

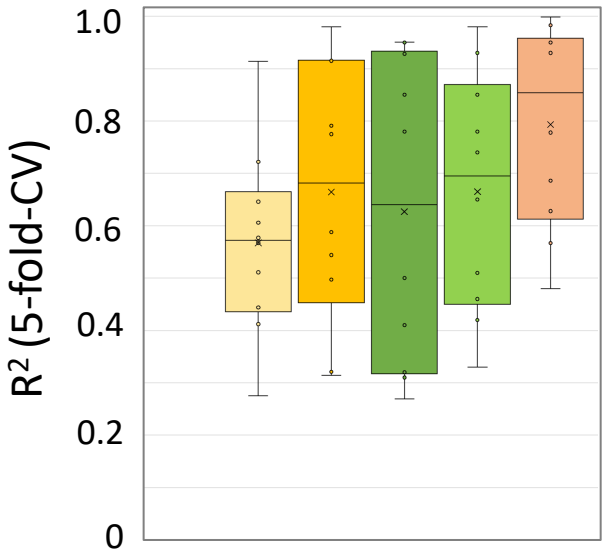
Using E. coli RC to solve many regression problems

OpenML
A worldwide machine learning lab



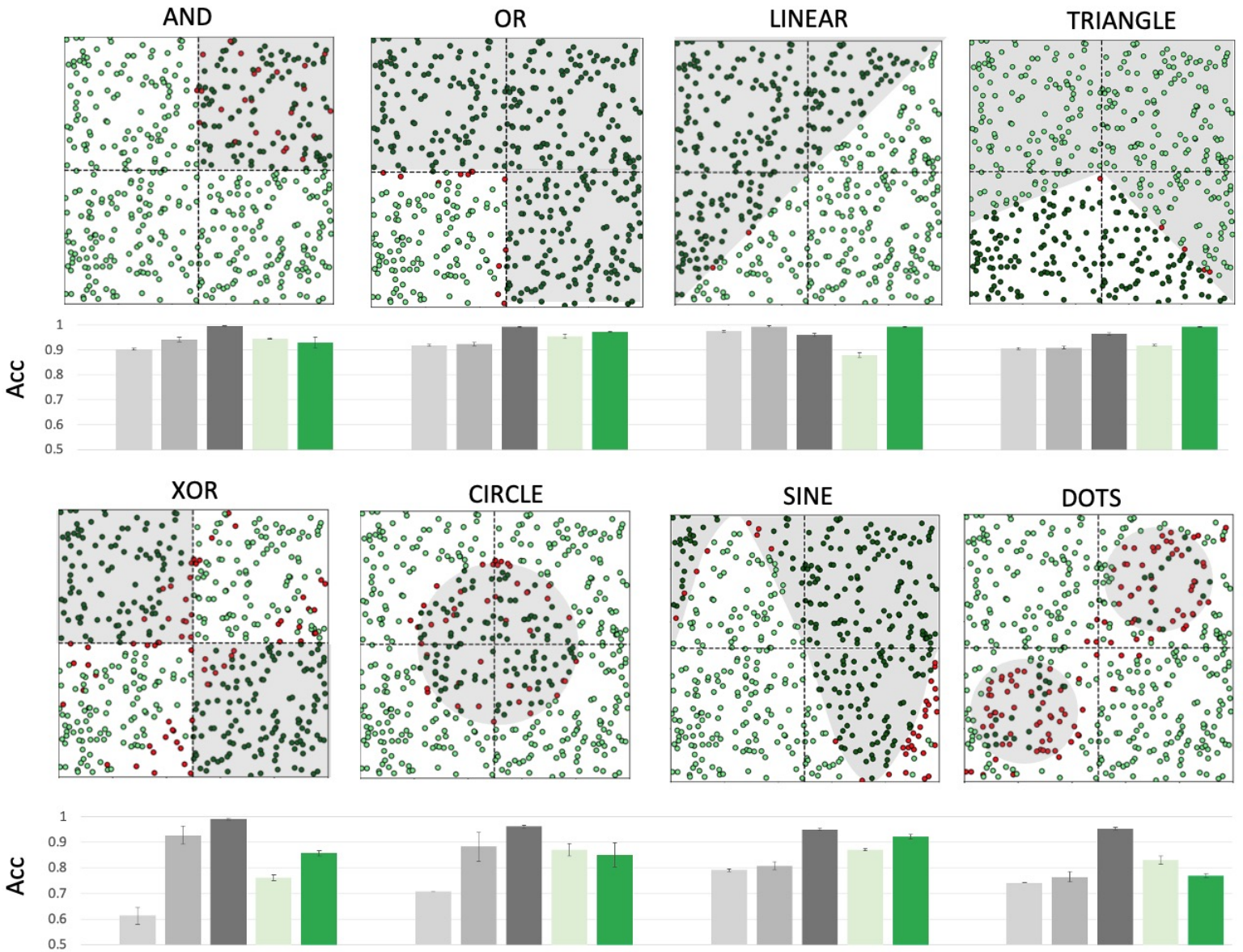
10 OpenML regression problems of increasing difficulty

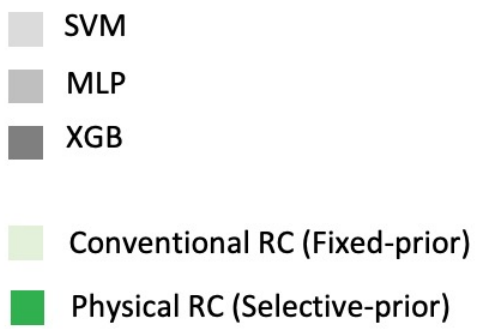


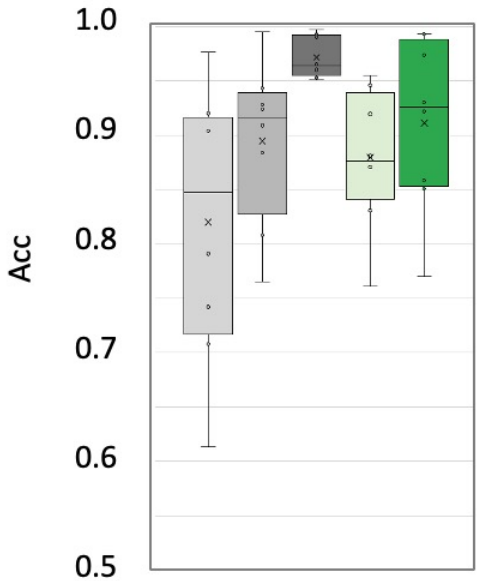


- Mollet, B., Ahavi P. *et al*. *Lecture Notes in Computer Science (*2025) and Ahavi P et al. *bioRxiv DOI: 10.1101/2024.09.12.612674* (2025)

Using E. coli RC to solve classification problems

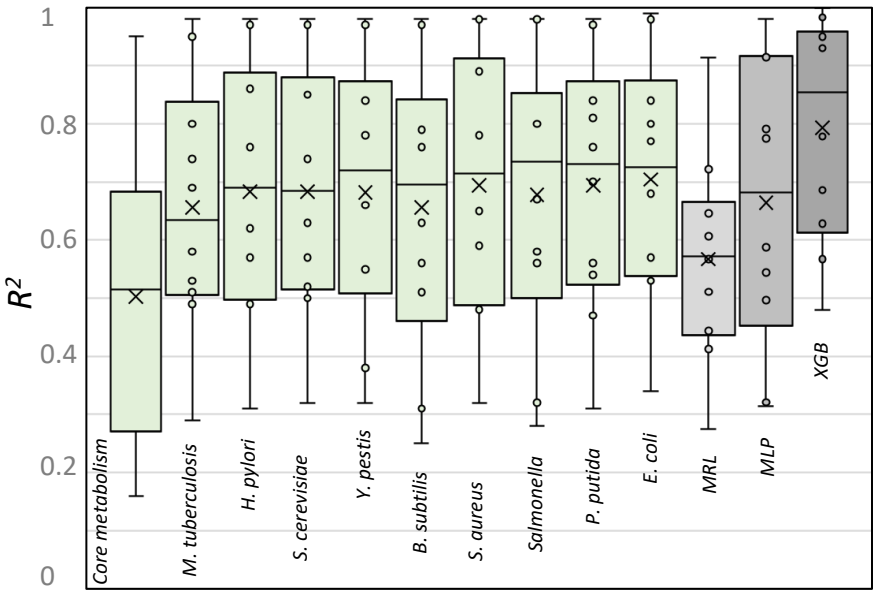


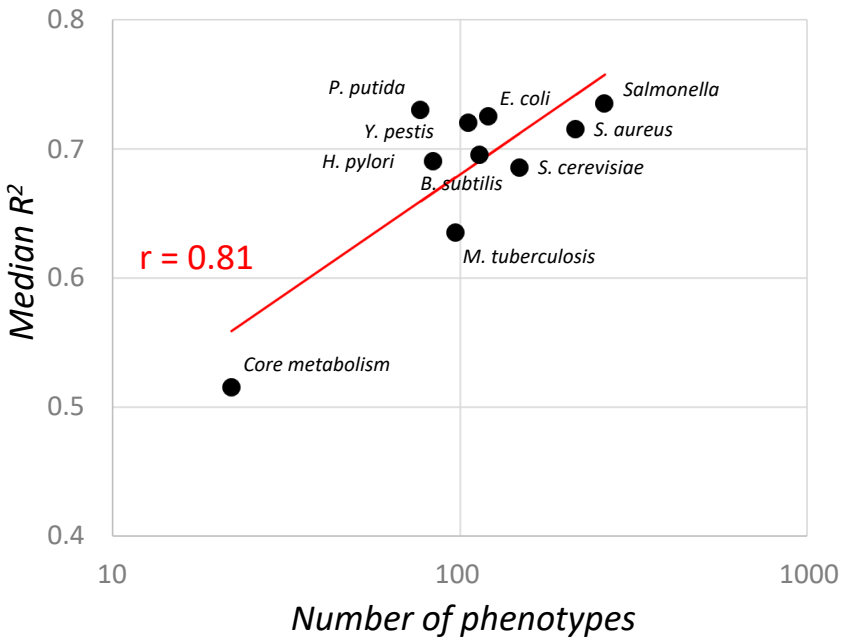


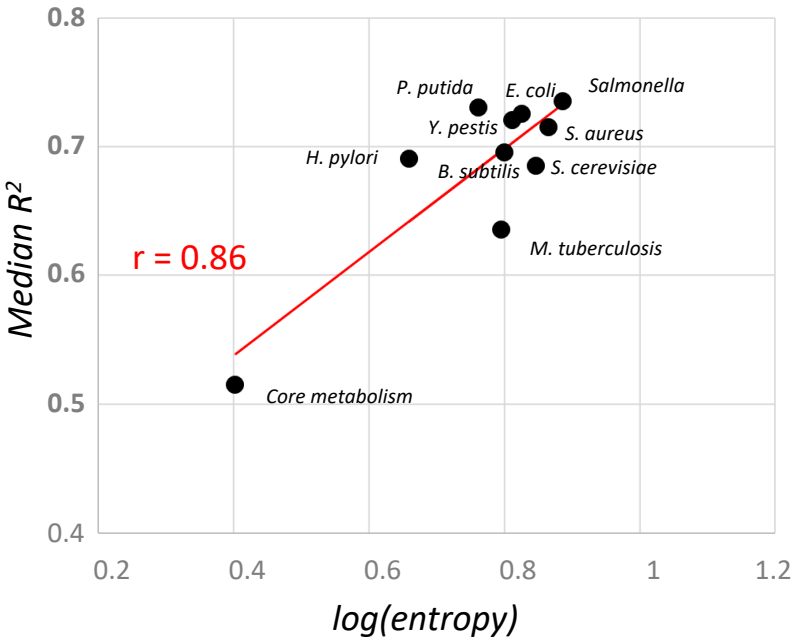


- Ahavi P et al. *bioRxiv DOI: 10.1101/2024.09.12.612674* (2025)

Relationship between GEM RC capacity to solve ML problems and phenotype diversity







- 10 AMN reservoirs trained on 10 GEMs
- Growth rates were acquired running FBA for 10k different media composition (selected at random among 28 sugars, amino and nucleic acids)
- Box plots are for 10 regression OpenML problems

- Number of phenotype is the number of different flux distributions calculated running FBA for 10k different media composition

- $H = \sum_i p_i \, log_2(p_i)$ where p_i is the probability of appearance of phenotype i

- Ahavi P et al. *bioRxiv DOI: 10.1101/2024.09.12.612674* (2025)

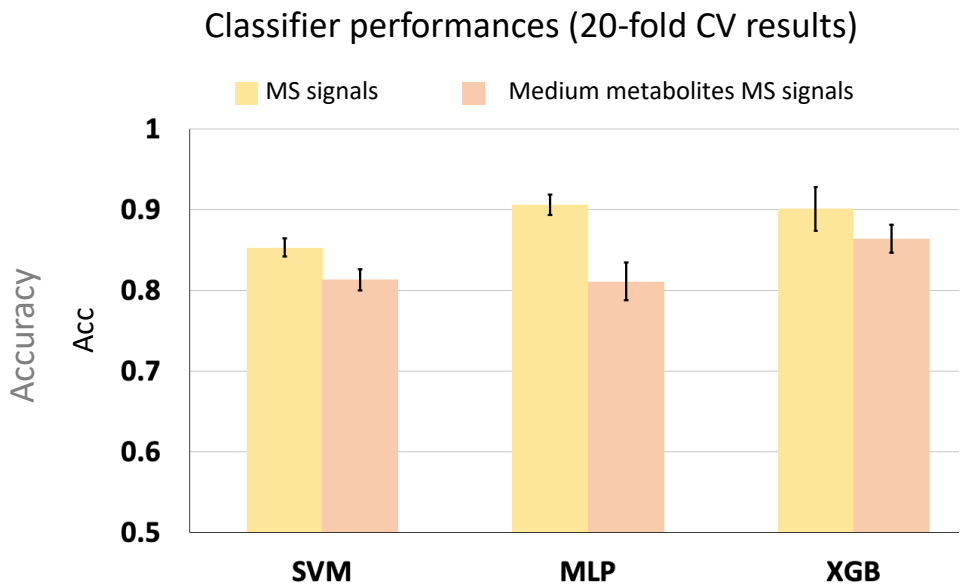
Can E. coli RC solve a concrete classification problem ?

The problem:

- Blood sample are collected for Covid-19 patients once they enter the hospital
- Metabolomics analyses are carried out on the samples
- Can we predict from the analyses if the disease outcome will be severe or moderate?

CHU Grenoble-Alpes cohort (training set):

- 81 patients
- 624 molecules detected (56 *E. coli* medium molecules)
- severe (31) – moderate (50)



Accuracy = 0.84 in Shen et al. *Cell* 2020; 182(1): 59–72

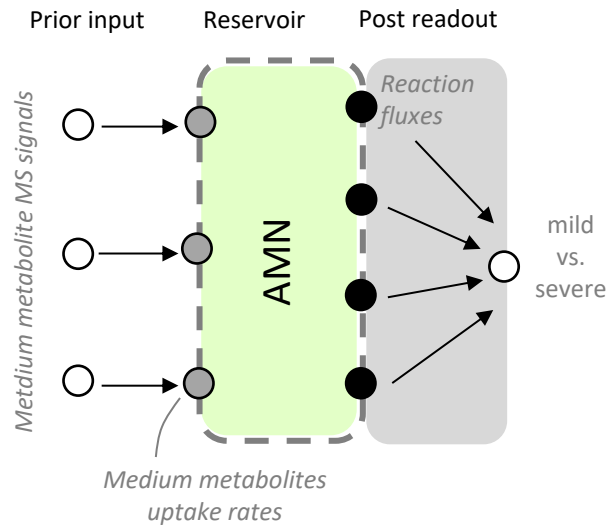
- Ahavi P et al. *bioRxiv DOI: 10.1101/2024.09.12.612674* (2025)

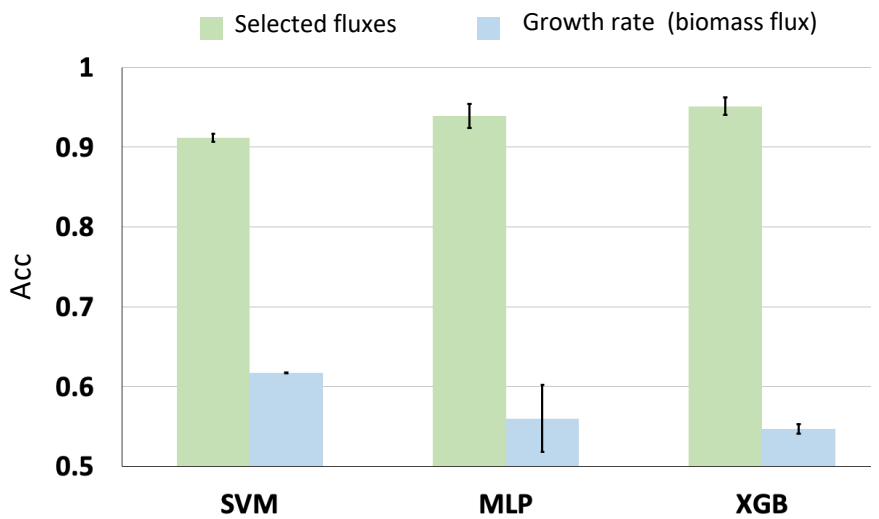
Using E. coli RC for classification

The problem:

- Blood sample are collected for Covid-19 patients once they enter the hospital
- ~~Metabolomics analyses are carried out on the samples~~
- ~~Can we predict from the analyses if the disease outcome will be severe or moderate?~~
- Can we use an *E. coli* RC grown on the patient's sample to predict if the disease outcome will be severe or moderate ?

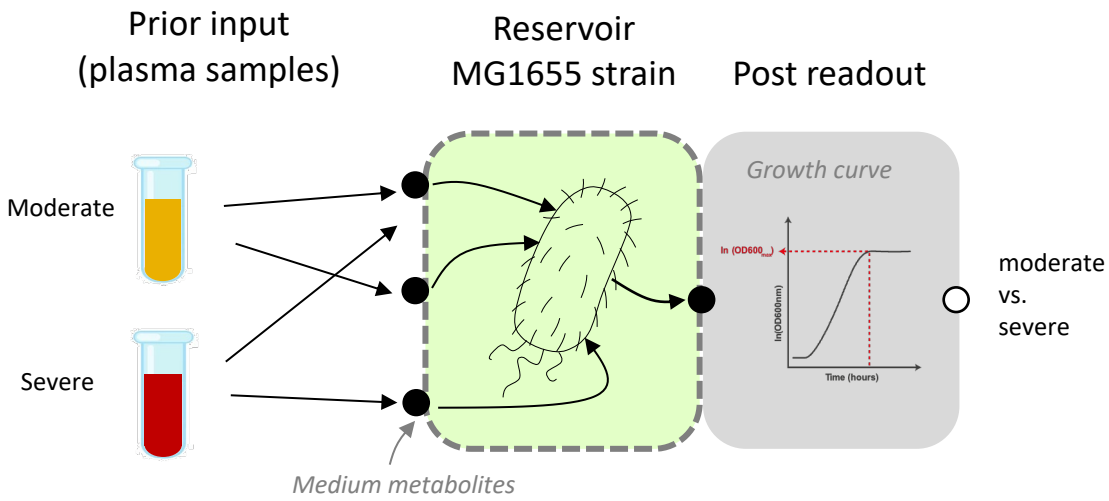
Conventional RC to predict disease outcome from phenotype

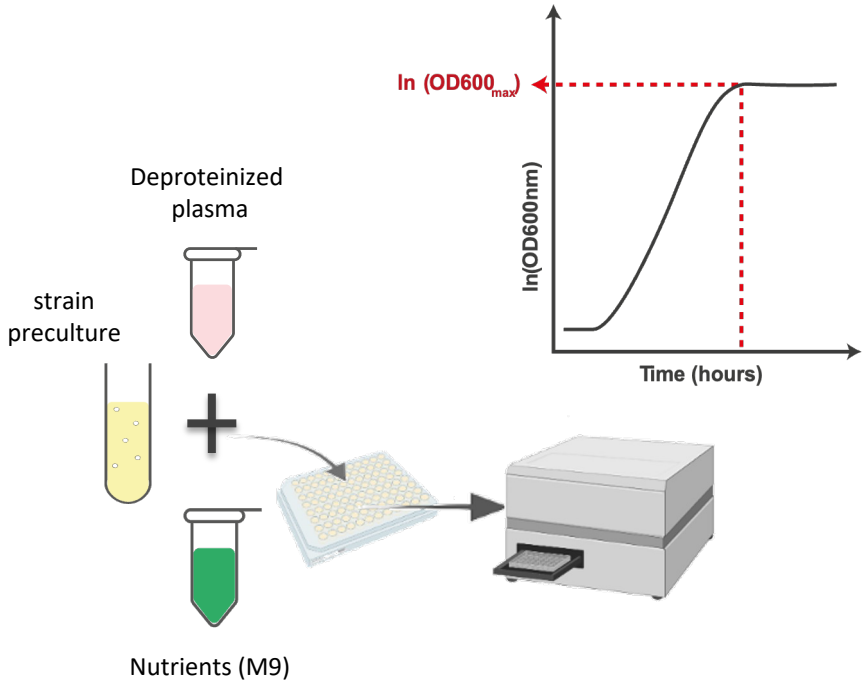




- Ahavi P et al. *bioRxiv DOI: 10.1101/2024.09.12.612674* (2025)

Building an E. coli physical RC for classification





- According to conventional RC, differences of nutrients concentration in the plasma should result in different phenotype
- *Build an E. coli* Physical RC to predict disease outcome from growth curve

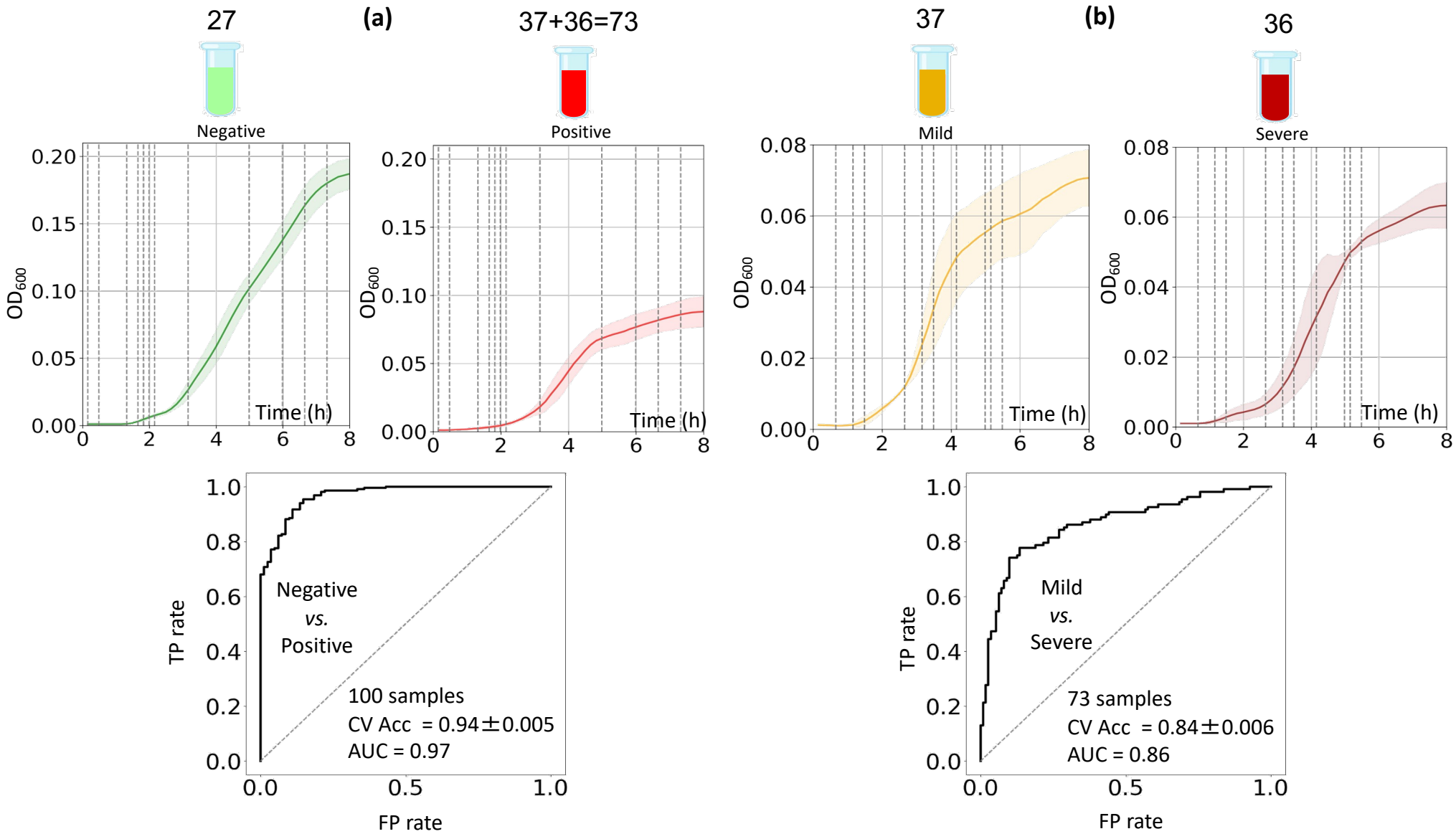
- Ahavi P et al. *bioRxiv DOI: 10.1101/2024.09.12.612674* (2025)

Benchmarking E. coli physical RC for classification

Prior input

Reservoir

Post readout



- Ahavi P et al. *bioRxiv DOI: 10.1101/2024.09.12.612674* (2025)

Perspective: a new concept in synthetic biology

- Decades of research and development in Synthetic Biology to build bottom-up computing devices (digital, analog, neural,…)… but many difficulties

- Most devices were inspired from natural biological networks: one should to consider building devices top-down, *i.e*. exploiting/modifying hosts rather than plugging orthogonal devices

Acknowledgments

AMN
SynBioDiag

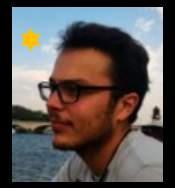
B-BEST

HORIZON BIOS SUSTAINABLE BIO-MANUFACTURING


Paul Ahvi


An Hoang


Bastien Mollet


Léon Faure


Amir Pandi


Mathilde Koch

Molecular Biology
Computational Biology

Audrey Le Gouellec

Jérôme Bonnet

Wolfram Liebermeister