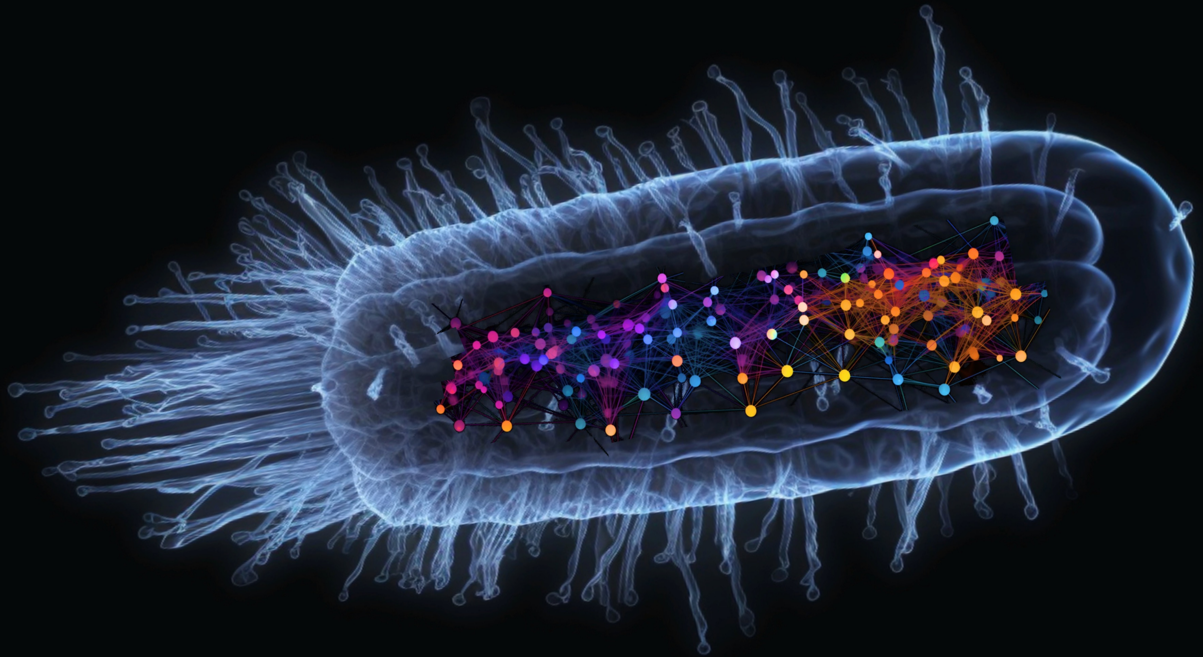


Machine learning models applied to Synthetic Biology

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

MICALIS Institute

ACS Synthetic Biology

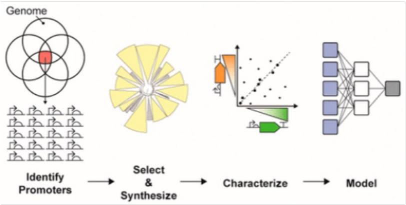
Key words: AI or Machine Learning

~400 hits

Precise Prediction of Promoter Strength Based on a De Novo Synthetic Promoter Library Coupled with Machine Learning

Mei Zhao, Zhenqi Yuan, Longtao Wu, Shenghu Zhou*, and Yu Deng*

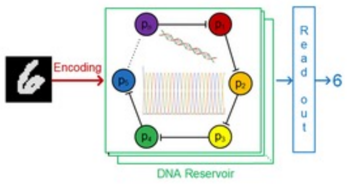
ACS Synthetic Biology 2022, 11, 1, 92-102 **(Research Article)**
Subscribed
Publication Date (Web): December 19, 2021
DOI: 10.1021/acssynbio.1c00117



Machine Learning Applied to Predicting Microorganism Growth Temperatures and Enzyme Catalytic Optima

Gang Li, Kersten S. Rabe, Jens Nielsen, and Martin K. M. Engqvist*

ACS Synthetic Biology 2019, 8, 6, 1411-1420 **(Research Article)**



Reservoir Computing Using DNA Oscillators

Xingyi Liu and Keshab K. Parhi*

ACS Synthetic Biology 2022, 11, 2, 780-787 **(Research Article)** Subscribed
Publication Date (Web): January 26, 2022
DOI: 10.1021/acssynbio.1c00483

SeqImprove: Machine-Learning-Assisted Curation of Genetic Circuit Sequence Information

Jeanet Mante, Zach Sents, Duncan Britt, William Mo, Chunxiao Liao, Ryan Greer , and Chris J. Myers*

ACS Synthetic Biology, **Articles ASAP (Technical Note)** Subscribed
Publication Date (Web): September 4, 2024
DOI: 10.1021/acssynbio.4c00292

Reinforcement Learning for Bioretrosynthesis

Mathilde Koch, Thomas Duigou, and Jean-Loup Faulon*

ACS Synthetic Biology 2020, 9, 1, 157-168 **(Research Article)**
Subscribed
Publication Date (Web): December 10, 2019
DOI: 10.1021/acssynbio.9b00447

DNA Input Classification by a Riboregulator-Based Cell-Free Perceptron

Ardjan J. van der Linden, Pascal A. Pieters , Mart W. Bartelds, Bryan L. Nathalia, Peng Yin, Wilhelm T. S. Huck*, Jongmin Kim* and Tom F. A. de Greef*

ACS Synthetic Biology 2022, 11, 4, 1510-1520 **(Research Article)** Open Access
Publication Date (Web): April 5, 2022
DOI: 10.1021/acssynbio.1c00596

Generative Artificial Intelligence GPT-4 Accelerates Knowledge Mining and Machine Learning for Synthetic Biology

Zhengyang Xiao, Wenyu Li, Hannah Moon, Garrett W. Roell*, Yixin Chen*, and Yinjie J. Tang*

ACS Synthetic Biology 2023, 12, 10, 2973-2982 **(Research Article)**
Subscribed
Publication Date (Web): September 8, 2023
DOI: 10.1021/acssynbio.3c00310

Design and Analysis of Compact DNA Strand Displacement Circuits for Analog Computation Using Autocatalytic Amplifiers

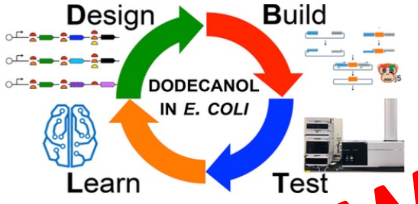
Tianqi Song, Sudhanshu Garg, Reem Mokhtar, Hieu Bui, and John Reif*

ACS Synthetic Biology 2018, 7, 1, 46-53 **(Research Article)**
Publication Date (Web): December 4, 2017
DOI: 10.1021/acssynbio.6b00390

DNA Memristors and Their Application to Reservoir Computing

Xingyi Liu and Keshab K. Parhi*

ACS Synthetic Biology 2022, 11, 6, 2202-2213 **(Research Article)** Subscribed
Publication Date (Web): May 13, 2022
DOI: 10.1021/acssynbio.2c00184



Lessons from Two Design–Build–Test–Learn Cycles of Dodecanol Production in *Escherichia coli* Aided by Machine Learning

Paul Opgenorth, Zak Costello, Takuya Okada, Garima Goyal, Yan Chen, Jennifer Gin, Veronica Benites, Markus de Raad, Trent R. Northen, Kai Deng, Samuel Deutsch, Edward E. K. Baidoo, Christopher J. Petzold, Nathan J. Hillson, Hector Garcia Martin, and Harry R. Beller*

ACS Synthetic Biology 2019, 8, 6, 1337-1351 **(Research Article)**

Analog Computation by DNA Strand Displacement Circuits

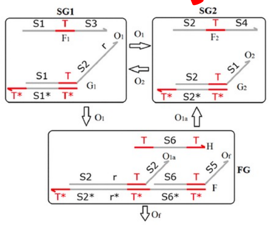
Tianqi Song, Sudhanshu Garg, Reem Mokhtar, Hieu Bui, and John Reif*

ACS Synthetic Biology 2016, 5, 8, 898-912 **(Research Article)**
Publication Date (Web): July 1, 2016
DOI: 10.1021/acssynbio.6b00144

Machine Learning of Designed Translational Control Allows Predictive Pathway Optimization in *Escherichia coli*

Adrian J. Jervis, Pablo Carbonell, Maria Vinaixa, Mark S. Dunstan, Katherine A. Hollywood, Christopher J. Robinson, Nicholas J. W. Rattray, Cunyu Yan, Neil Swainston, Andrew Currin, Rehana Sung, Helen Toogood, Sandra Taylor, Jean-Loup Faulon, Rainer Breitling, Eriko Takano, and Nigel S. Scrutton*

ACS Synthetic Biology 2019, 8, 1, 127-136 **(Research Article)**



Tuning the Performance of Synthetic Riboswitches using Machine Learning

Ann-Christin Groher, Sven Jager, Christopher Schneider, Florian Groher, Kay Hamacher*, and Beatrix Suess*

ACS Synthetic Biology 2019, 8, 1, 34-44 **(Research Article)**

Semisupervised Gaussian Process for Automated Enzyme Search

Joseph Mellor, Ioana Grigoras, Pablo Carbonell, and Jean-Loup Faulon*

ACS Synthetic Biology 2016, 5, 6, 518-528 **(Research Article)**
Publication Date (Web): March 23, 2016
DOI: 10.1021/acssynbio.5b00294

Machine-Learning-Guided Mutagenesis for Directed Evolution of Fluorescent Proteins

Yutaka Saito, Misaki Oikawa, Hikaru Nakazawa, Teppei Niide, Tomoshi Kameda, Koji Tsuda*, and Mitsuo Umetsu*

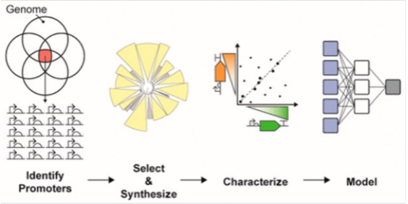
ACS Synthetic Biology 2018, 7, 9, 2014-2022 **(Letter)**

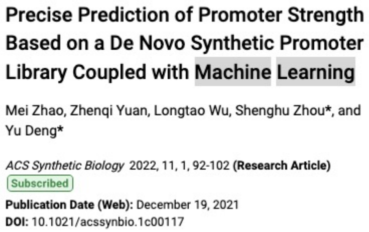
Supervised Learning in Adaptive DNA Strand Displacement Networks

Matthew R. Lakin* and Darko Stefanovic

ACS Synthetic Biology 2016, 5, 8, 885-897 **(Research Article)**
Publication Date (Web): April 25, 2016
DOI: 10.1021/acssynbio.6b00009

Supervised / Unsupervised Learning

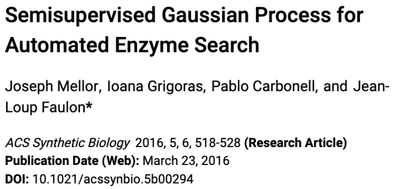



Precise Prediction of Promoter Strength Based on a De Novo Synthetic Promoter Library Coupled with Machine Learning

Mei Zhao, Zhenqi Yuan, Longtao Wu, Shenghu Zhou*, and Yu Deng*

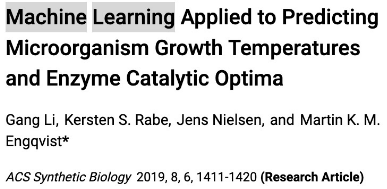
ACS Synthetic Biology 2022, 11, 1, 92-102 **(Research Article)** Subscribed

Publication Date (Web): December 19, 2021
DOI: 10.1021/acssynbio.1c00117


Semisupervised Gaussian Process for Automated Enzyme Search

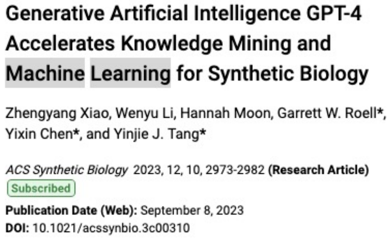
Joseph Mellor, Ioana Grigoras, Pablo Carbonell, and Jean-Loup Faulon*

ACS Synthetic Biology 2016, 5, 6, 518-528 **(Research Article)**
Publication Date (Web): March 23, 2016
DOI: 10.1021/acssynbio.5b00294


Machine Learning Applied to Predicting Microorganism Growth Temperatures and Enzyme Catalytic Optima

Gang Li, Kersten S. Rabe, Jens Nielsen, and Martin K. M. Engqvist*

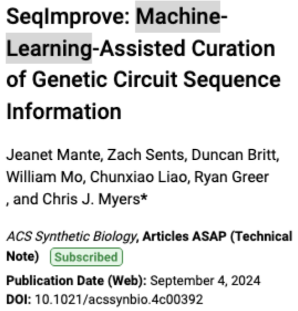
ACS Synthetic Biology 2019, 8, 6, 1411-1420 **(Research Article)**


Generative Artificial Intelligence GPT-4 Accelerates Knowledge Mining and Machine Learning for Synthetic Biology

Zhengyang Xiao, Wenyu Li, Hannah Moon, Garrett W. Roell*, Yixin Chen*, and Yinjie J. Tang*

ACS Synthetic Biology 2023, 12, 10, 2973-2982 **(Research Article)** Subscribed

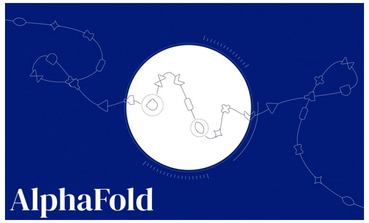
Publication Date (Web): September 8, 2023
DOI: 10.1021/acssynbio.3c00310


SeqImprove: Machine-Learning-Assisted Curation of Genetic Circuit Sequence Information

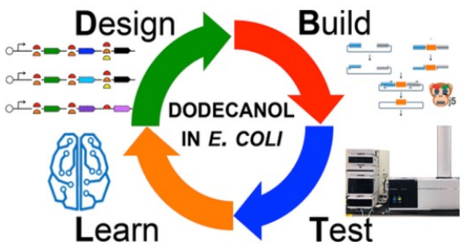
Jeanet Mante, Zach Sents, Duncan Britt, William Mo, Chunxiao Liao, Ryan Greer , and Chris J. Myers*

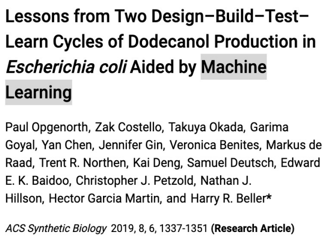
ACS Synthetic Biology, **Articles ASAP (Technical Note)** Subscribed

Publication Date (Web): September 4, 2024
DOI: 10.1021/acssynbio.4c00392



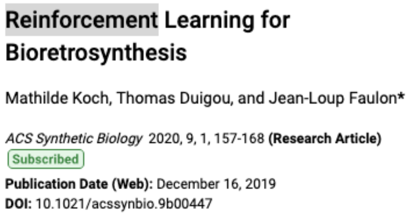
Active / Reinforcement Learning




Lessons from Two Design–Build–Test–Learn Cycles of Dodecanol Production in *Escherichia coli* Aided by Machine Learning

Paul Opgenorth, Zak Costello, Takuya Okada, Garima Goyal, Yan Chen, Jennifer Gin, Veronica Benites, Markus de Raad, Trent R. Northen, Kai Deng, Samuel Deutsch, Edward E. K. Baidoo, Christopher J. Petzold, Nathan J. Hillson, Hector Garcia Martin, and Harry R. Beller*

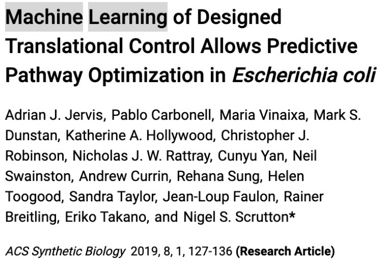
ACS Synthetic Biology 2019, 8, 6, 1337-1351 **(Research Article)**


Reinforcement Learning for Bioretrosynthesis

Mathilde Koch, Thomas Duigou, and Jean-Loup Faulon*

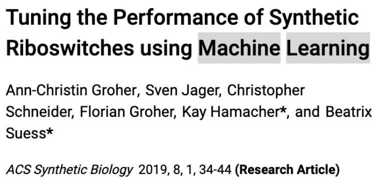
ACS Synthetic Biology 2020, 9, 1, 157-168 **(Research Article)** Subscribed

Publication Date (Web): December 16, 2019
DOI: 10.1021/acssynbio.9b00447


Machine Learning of Designed Translational Control Allows Predictive Pathway Optimization in *Escherichia coli*

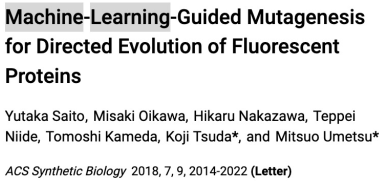
Adrian J. Jervis, Pablo Carbonell, Maria Vinaixa, Mark S. Dunstan, Katherine A. Hollywood, Christopher J. Robinson, Nicholas J. W. Rattray, Cunyu Yan, Neil Swainston, Andrew Currin, Rehana Sung, Helen Toogood, Sandra Taylor, Jean-Loup Faulon, Rainer Breitling, Eriko Takano, and Nigel S. Scrutton*

ACS Synthetic Biology 2019, 8, 1, 127-136 **(Research Article)**


Tuning the Performance of Synthetic Riboswitches using Machine Learning

Ann-Christin Groher, Sven Jager, Christopher Schneider, Florian Groher, Kay Hamacher*, and Beatrix Suess*

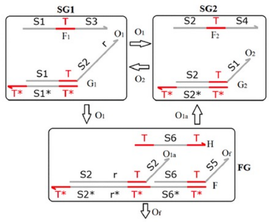
ACS Synthetic Biology 2019, 8, 1, 34-44 **(Research Article)**

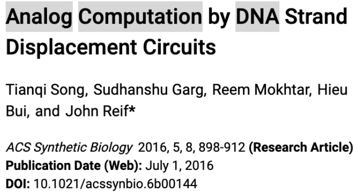

Machine-Learning-Guided Mutagenesis for Directed Evolution of Fluorescent Proteins

Yutaka Saito, Misaki Oikawa, Hikaru Nakazawa, Teppei Niide, Tomoshi Kameda, Koji Tsuda*, and Mitsuo Umetsu*

ACS Synthetic Biology 2018, 7, 9, 2014-2022 **(Letter)**

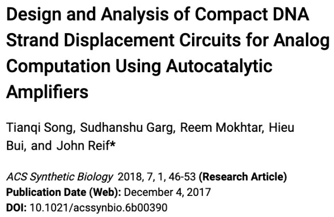
in vitro / in vivo Learning




Analog Computation by DNA Strand Displacement Circuits

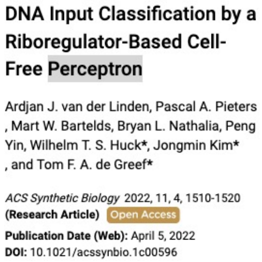
Tianqi Song, Sudhanshu Garg, Reem Mokhtar, Hieu Bui, and John Reif*

ACS Synthetic Biology 2016, 5, 8, 898-912 **(Research Article)**
Publication Date (Web): July 1, 2016
DOI: 10.1021/acssynbio.6b00144


Design and Analysis of Compact DNA Strand Displacement Circuits for Analog Computation Using Autocatalytic Amplifiers

Tianqi Song, Sudhanshu Garg, Reem Mokhtar, Hieu Bui, and John Reif*

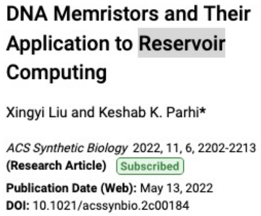
ACS Synthetic Biology 2018, 7, 1, 46-53 **(Research Article)**
Publication Date (Web): December 4, 2017
DOI: 10.1021/acssynbio.6b00390


DNA Input Classification by a Riboregulator-Based Cell-Free Perceptron

Ardjan J. van der Linden, Pascal A. Pieters , Mart W. Bartelds, Bryan L. Nathalia, Peng Yin, Wilhelm T. S. Huck*, Jongmin Kim* , and Tom F. A. de Greef*

ACS Synthetic Biology 2022, 11, 4, 1510-1520 **(Research Article)** Open Access

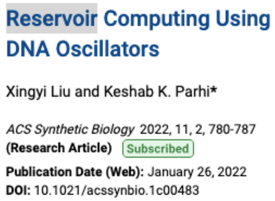
Publication Date (Web): April 5, 2022
DOI: 10.1021/acssynbio.1c00596


DNA Memristors and Their Application to Reservoir Computing

Xingyi Liu and Keshab K. Parhi*

ACS Synthetic Biology 2022, 11, 6, 2202-2213 **(Research Article)** Subscribed

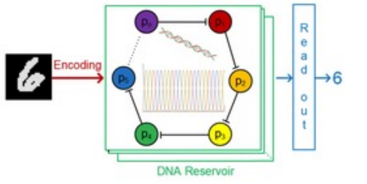
Publication Date (Web): May 13, 2022
DOI: 10.1021/acssynbio.2c00184


Reservoir Computing Using DNA Oscillators

Xingyi Liu and Keshab K. Parhi*

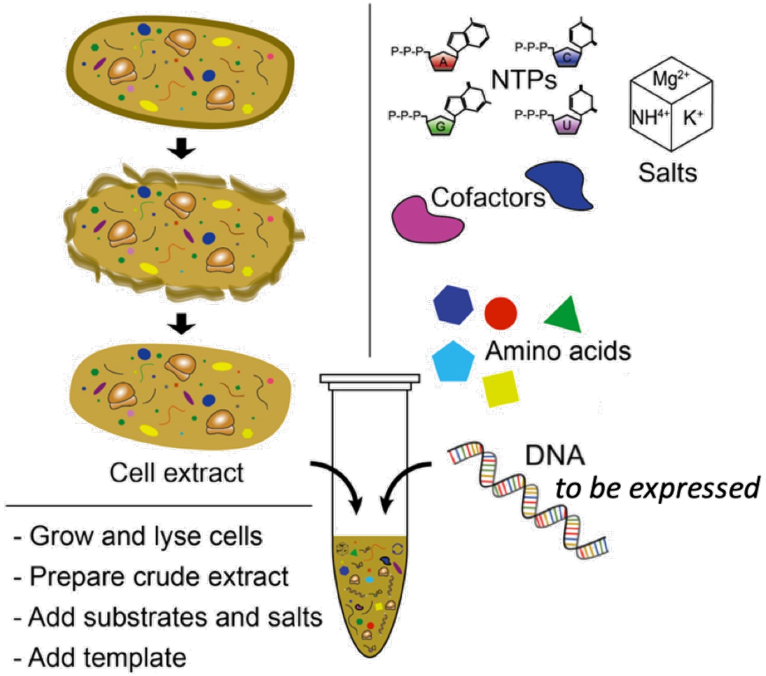
ACS Synthetic Biology 2022, 11, 2, 780-787 **(Research Article)** Subscribed

Publication Date (Web): January 26, 2022
DOI: 10.1021/acssynbio.1c00483



Active learning to optimize cell-free productivity

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

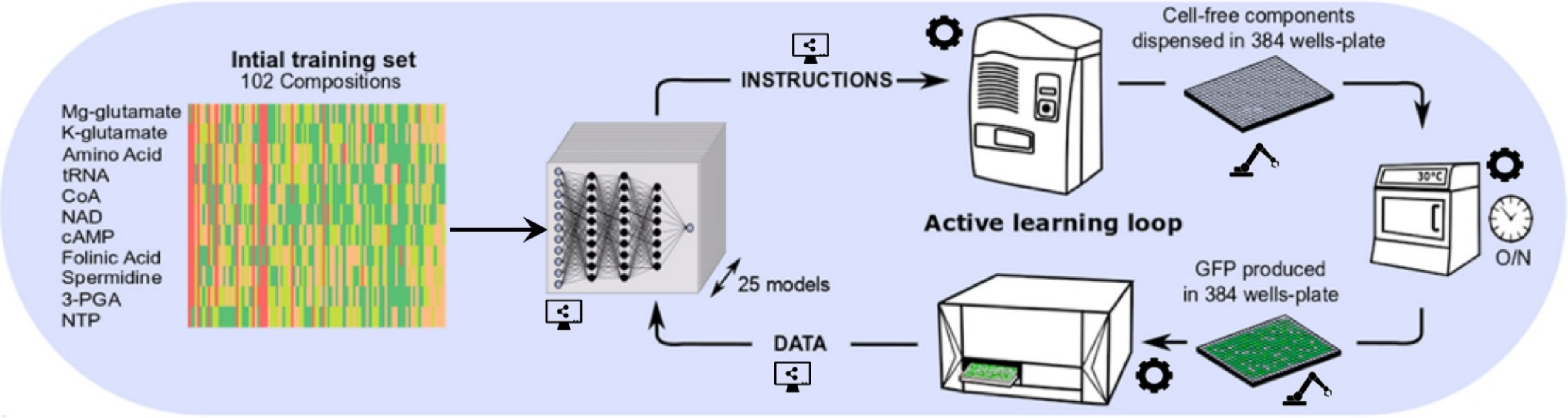
Combinatorial space = 4^{11}
= **4 194 304** compositions

⇓

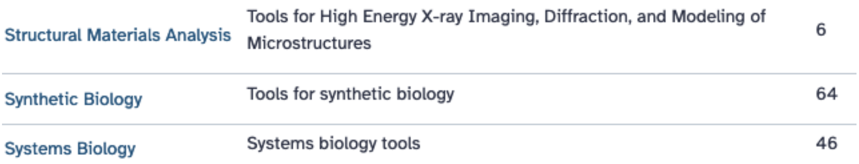
- Can we improve protein production without increasing the price of cell-free reaction?
- Can we provide efficient predictions of protein production *in vitro*?
- Can we highlight the critical parameters involve in protein production *in vitro*?

Active learning to optimize cell-free productivity

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

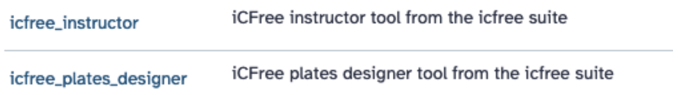


Galaxy Tool Shed →



Structural Materials Analysis	Tools for High Energy X-ray Imaging, Diffraction, and Modeling of Microstructures	6
Synthetic Biology	Tools for synthetic biology	64
Systems Biology	Systems biology tools	46

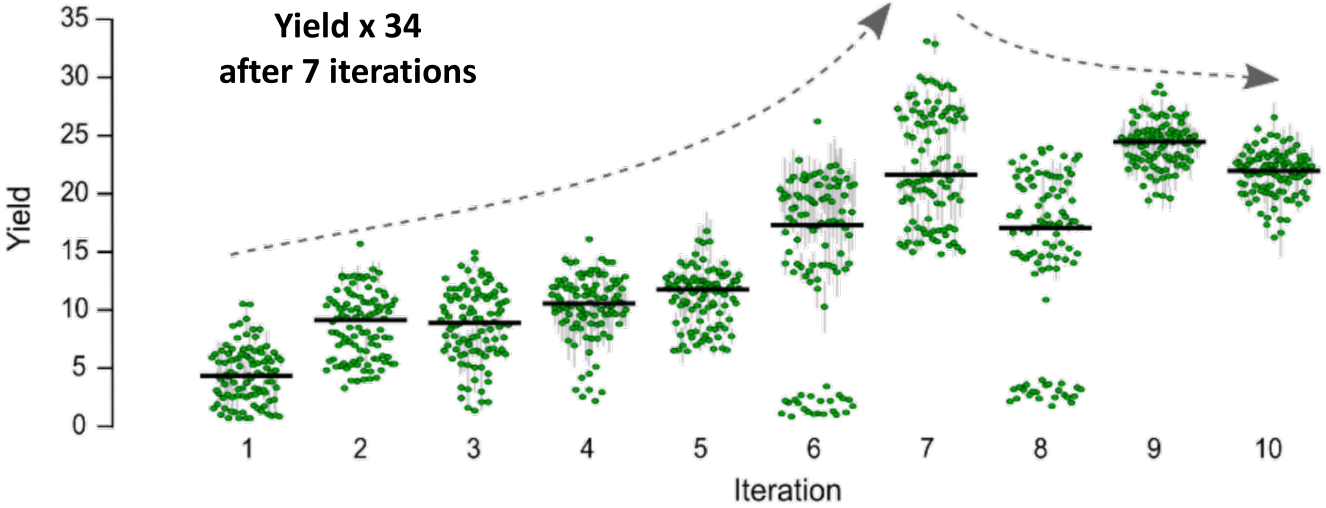
→

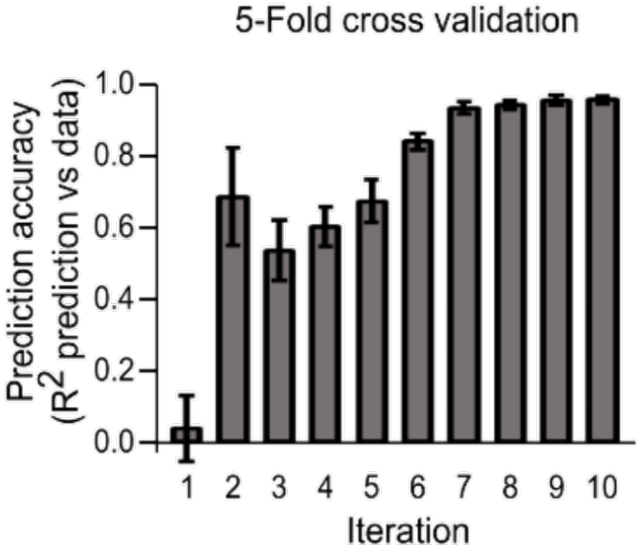


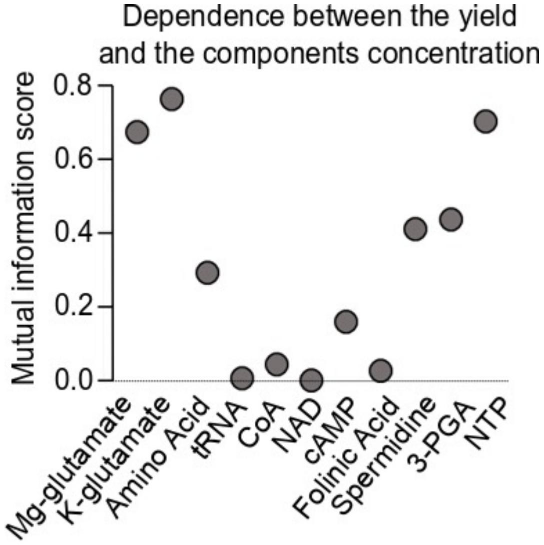
icfree_instructor	iCFree instructor tool from the icfree suite
icfree_plates_designer	iCFree plates designer tool from the icfree suite

- Herisson J, *et al. Nat Commun* **13**, 5082 (2022)

Active learning to optimize cell-free productivity

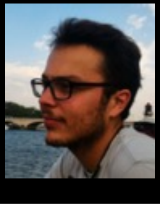


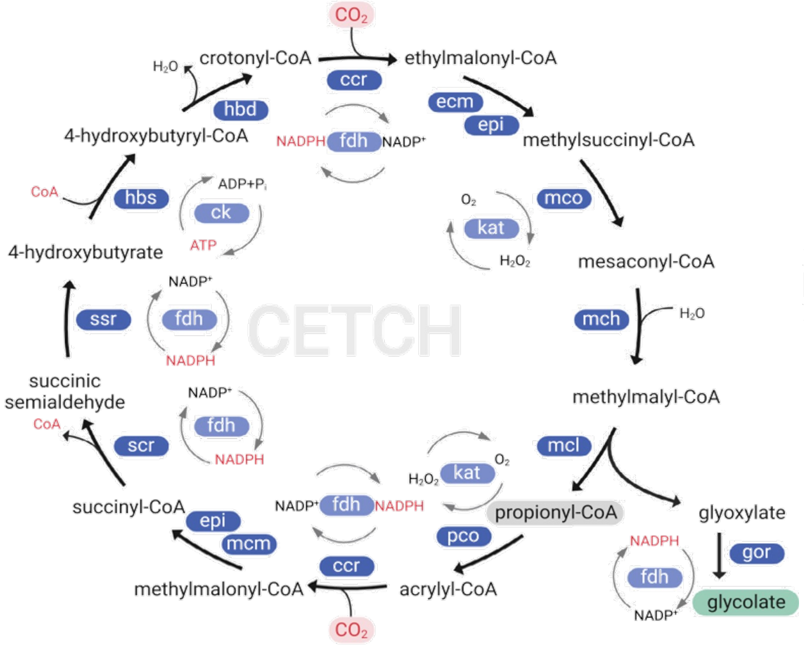


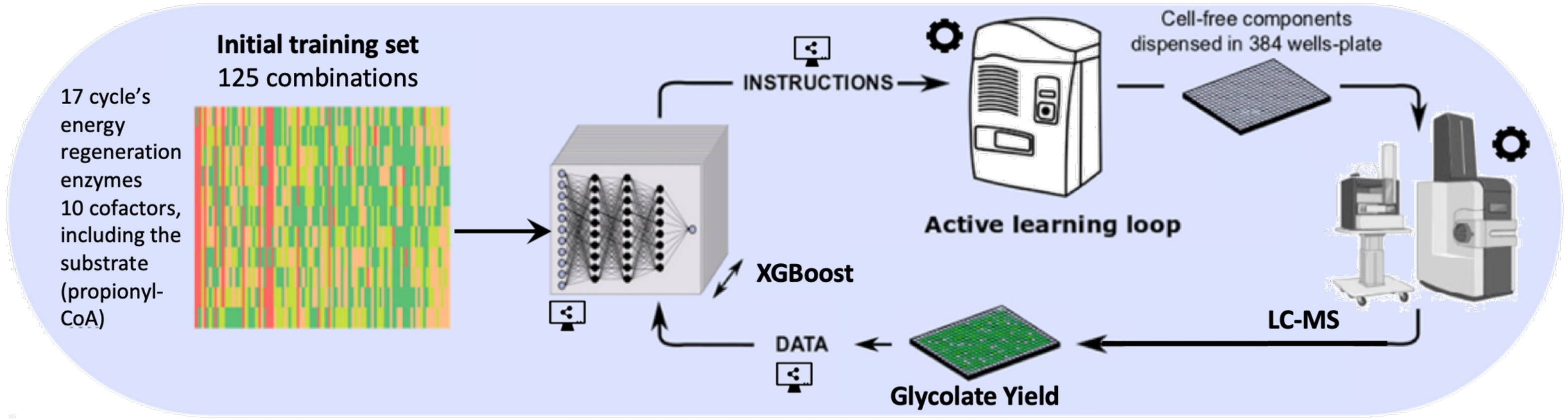


- Borkowski O, *et al. Nat Commun* **11,** 1872 (2020)

Active learning to optimize cell-free metabolic pathways



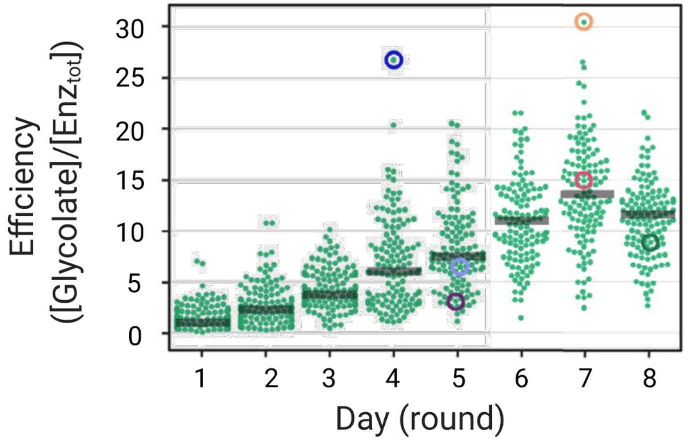


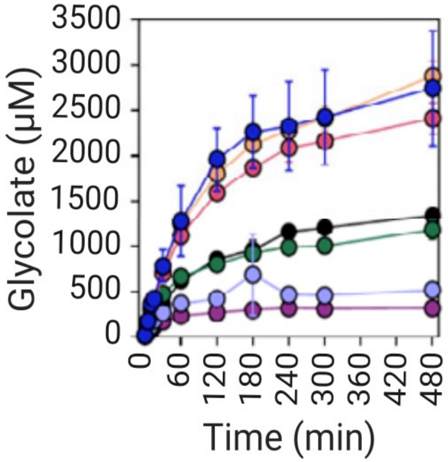


Initial training set
125 combinations

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

Cell-free components dispensed in 384 wells-plate





Top 3 best producers

Control (CETCH5.4)

Randomly selected producers

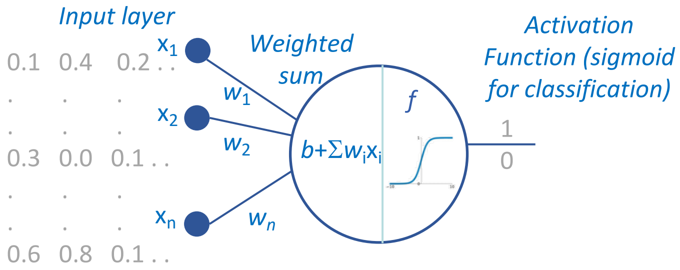
6x more efficient than the best in vitro CO_2-fixing system described to date (CETCH 5.4 , Schwander *et al. Science* 2016)

- Pandi A., *et al. Nat Commun* **13,** 3876 (2022)

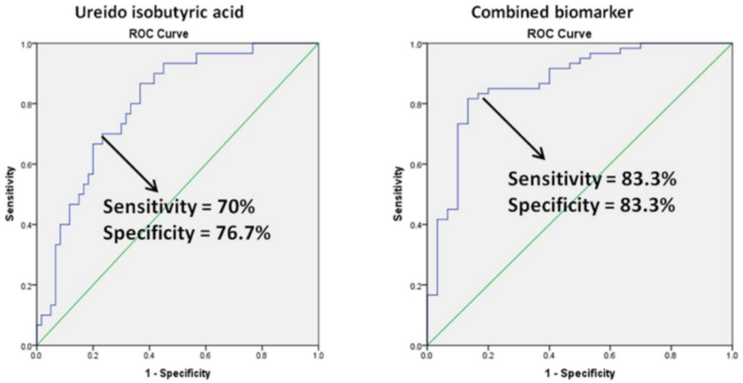
In vitro / in vivo learning: why?

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

$$\boxed{f\,(b+\Sigma w_i x_i\,)}$$

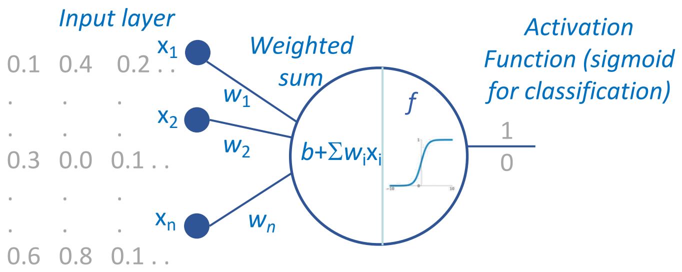
- **Is it possible to avoid biomarker concentration measurements?**

  ➢ **Engineer the trained network *in vitro* or *in vivo* and directly use it on clinical samples**

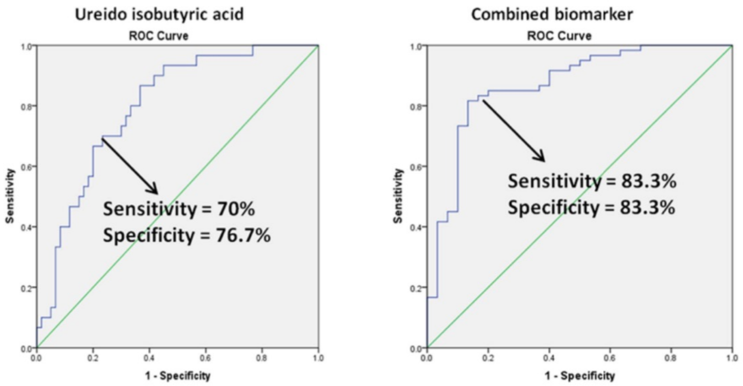
Engineering a neural metabolic network: the concept

TRAINING THE NETWORK

Perceptron weights (w_i) are learned to increase classifier accuracy



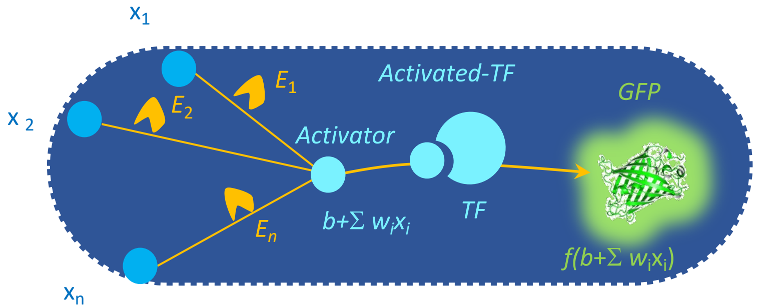
Prostate cancer $\quad \text{PCa metabolic score} = b + \sum_{j=1}^{j} w_j x_{ij}$



- 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

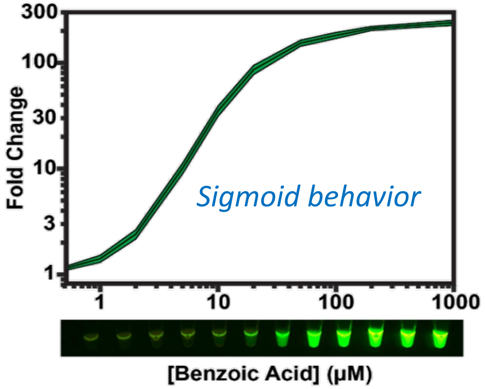


Enzymatic transformation *Reporter gene*

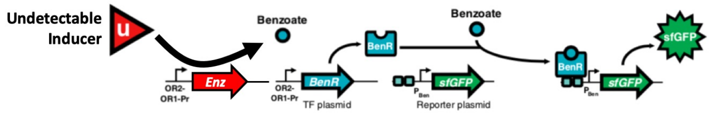
In theory (Michaelis-Menten) when $x_i << [E_i]$:

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

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



Engineering a neural metabolic network: the concept



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

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

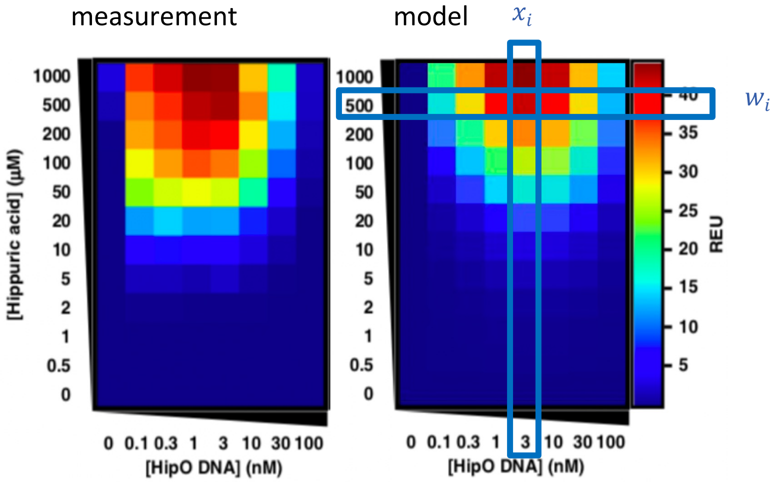
$$TF_{activated} = TF * \frac{benzoate}{benzoate + K_d^{inducer}} + 0.0005$$

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

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

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

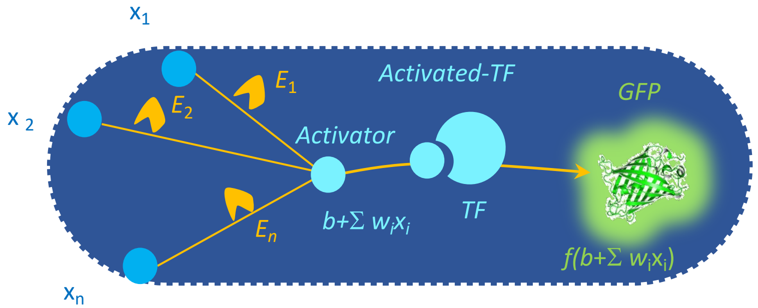
$$\frac{dprot}{dt} = \pi * mRNA * \frac{y}{y+k} * \frac{K_{tox}}{K_{tox} + tox} - \lambda * prot$$


measurement

model

ENGINEERING THE TRAINED NETWORK

Need to actuate weighted sum and activation function



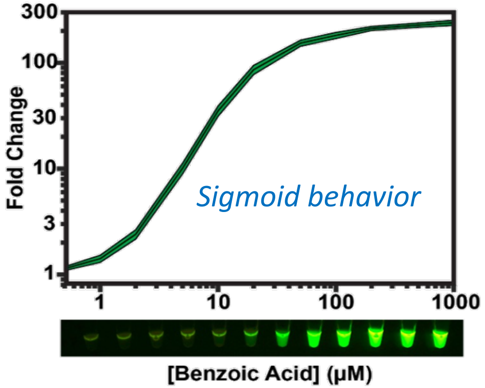
Enzymatic transformation

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



Sigmoid behavior

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

Engineering a neural metabolic network in vitro

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

$f(b+\sum w_i x_i)$

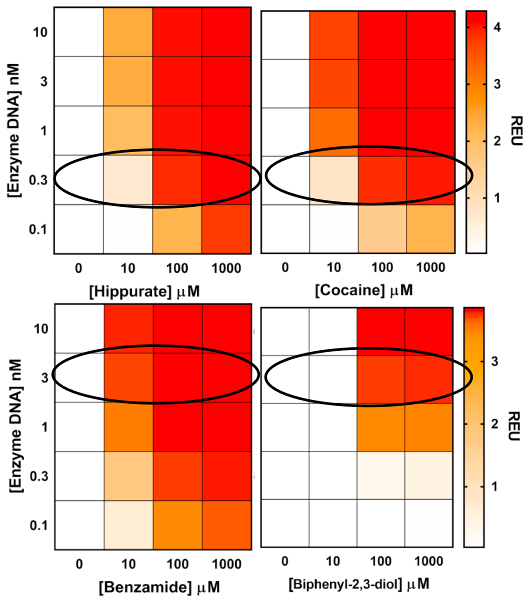
$b =$ - 0.50

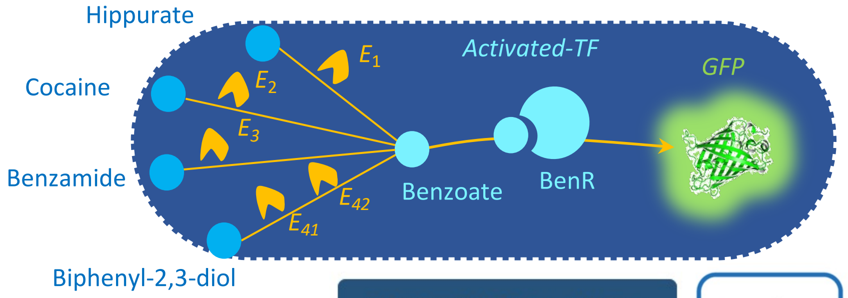
$w_1 =$ 0.25

$w_2 =$ 0.25

$w_3 =$ 0.50

$w_4 =$ 0.50

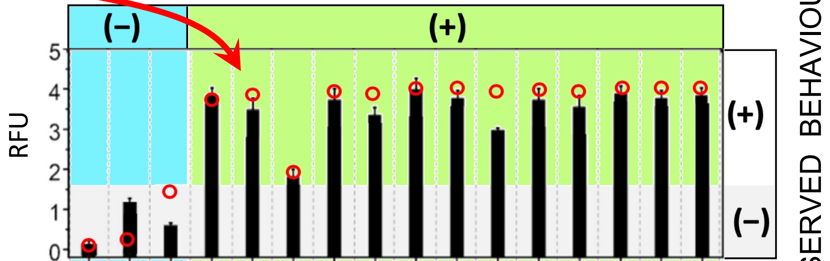




Retrosynthesis Workflow »

Galaxy

TARGETED BEHAVIOUR



Hippurate……………		●				●	●	●				●	●	●		●
Cocaine……………….			●			●			●		●	●	●		●	●
Benzamide………….				●			●		●	●		●		●	●	●
Biphenyl-2,3-diol…					●			●		●	●		●	●	●	●

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

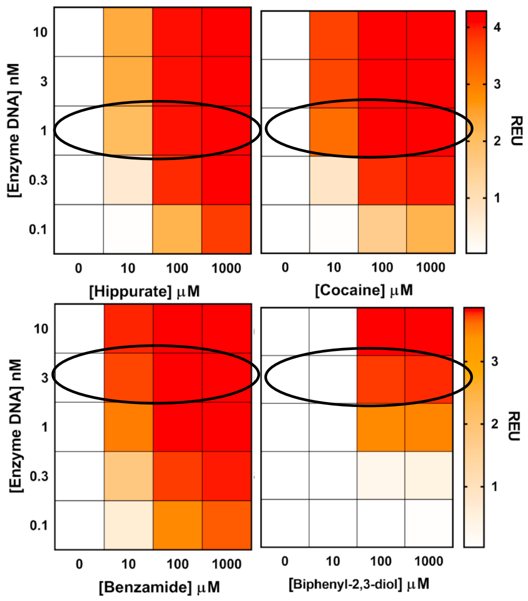
Engineering a neural metabolic network in vitro

TARGETED BEHAVIOUR

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

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

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



Logistic regression

$f(b+\sum w_i x_i)$

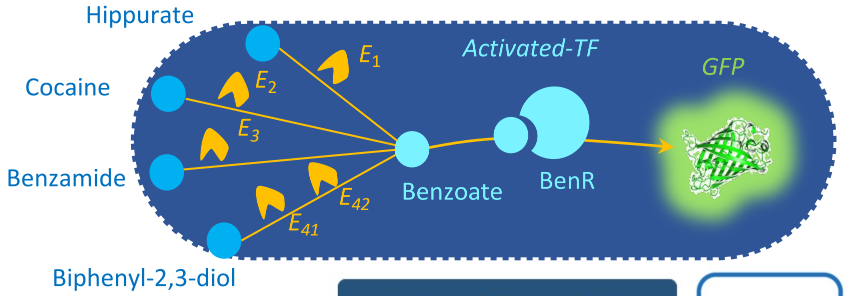
b = - 0.50

w_1 = ~~0.25~~ 0.50

w_2 = ~~0.25~~ 0.50

w_3 = ~~0.50~~ 0.50

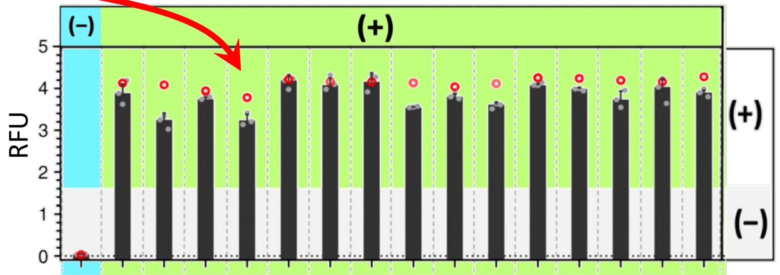
w_4 = ~~0.50~~ 0.50



Retrosynthesis Workflow »

Galaxy

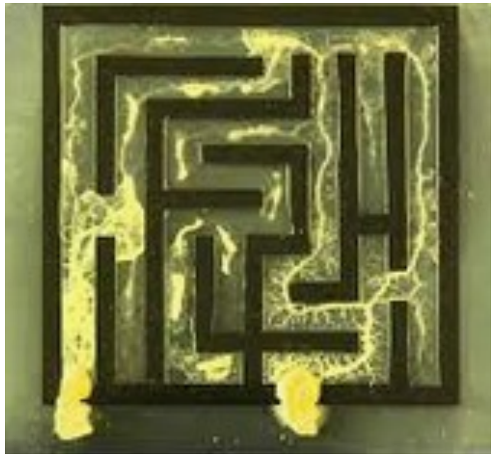
TARGETED BEHAVIOUR



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

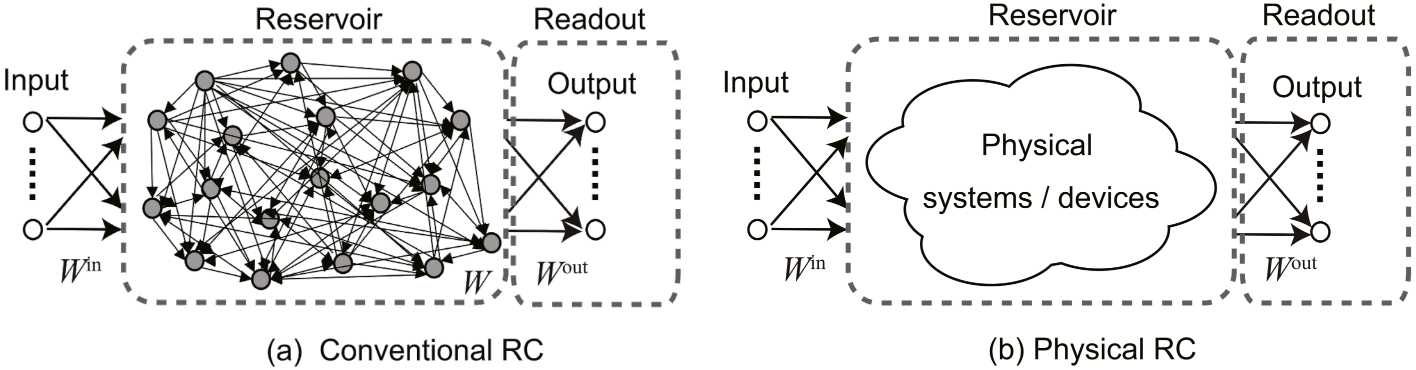
- Pandi A., Koch M. *et al. Nat Commun* **10**, 3880 (2019)

Engineering a neural metabolic network 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)



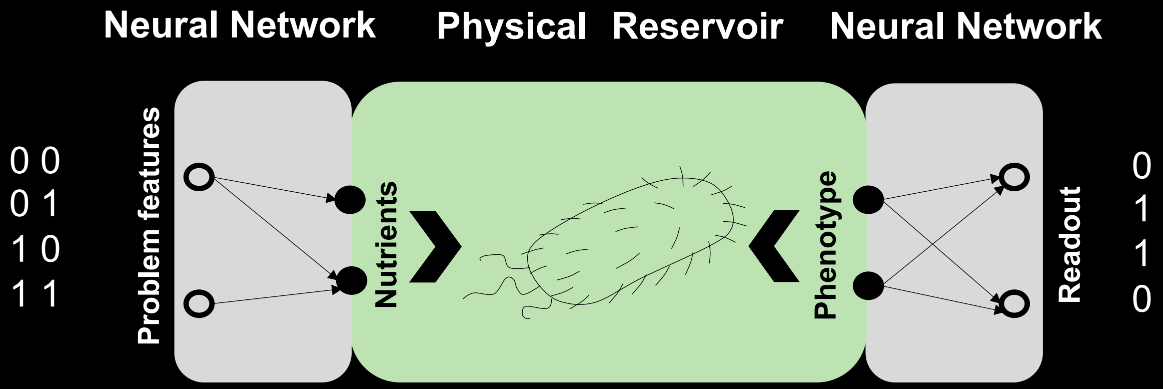
Tanaka G. et al. *Neural Networks* **115**, 100 (2019)

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?

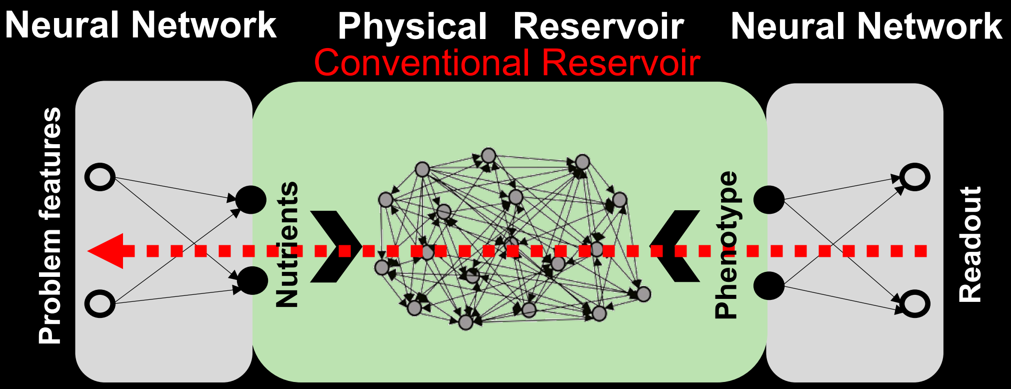


E. coli Reservoir Computer (E. coli RC)

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:

$S\,V = 0$

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

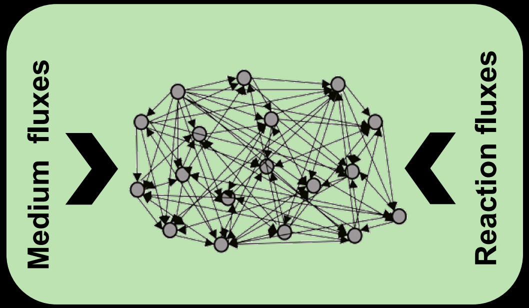
where

$-\ V\ =$ set of all reaction fluxes

$-\ \mathrm{S}\ =$ stochiometric matrix

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

Conventional Reservoir



The reservoir

GEnome-scale Metabolic Model (GEM/FBA)

Max $(v_{biomass})$

Subjected to contraints:

$S\,V = 0$

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

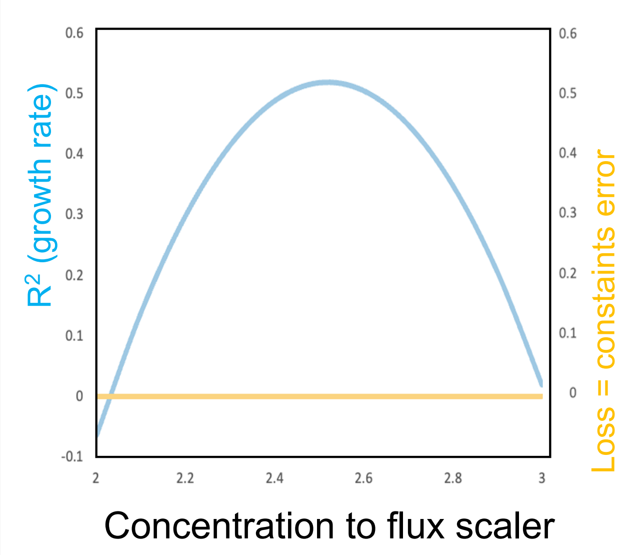
where

$-\,V$ = set of all reaction fluxes

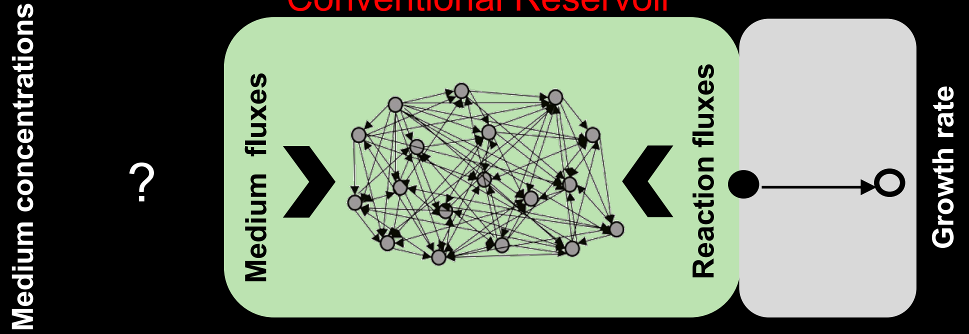
$-\,S$ = stochiometric matrix

$-V_{in}$ = 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



Conventional Reservoir



Conventional Reservoir should:

- *accurately reproduce phenotype for different media composition*

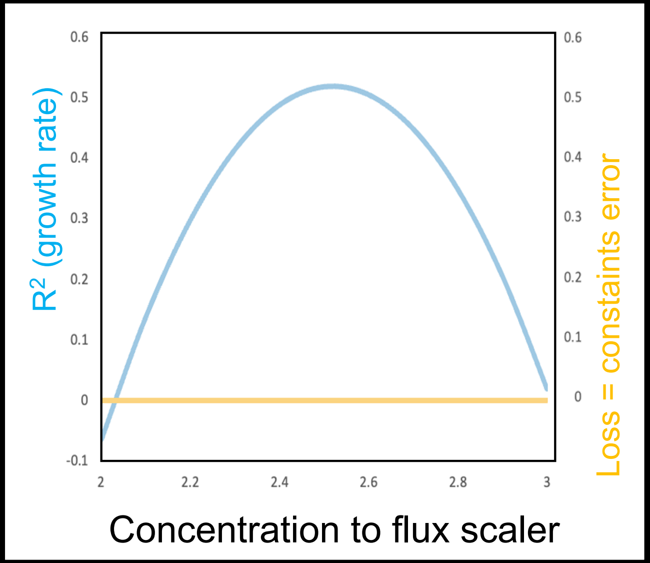
Building an E. coli RC to increase mechanistic model predictability

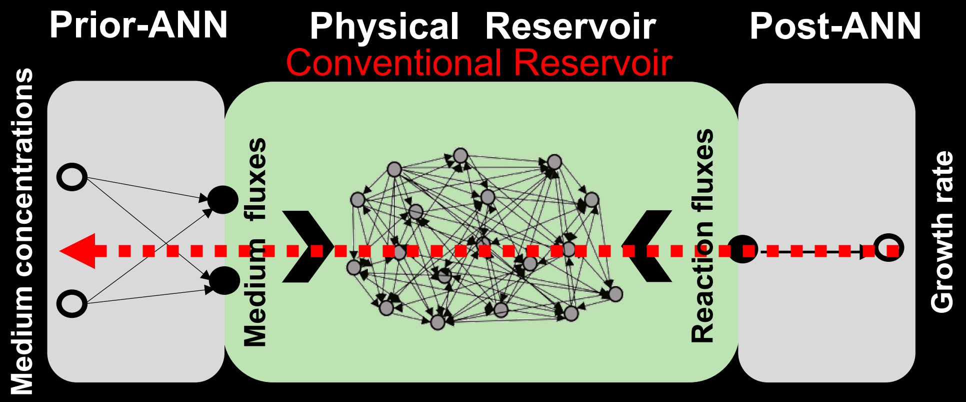
GEM/FBA is a Linear Program solved using Simplex algorithm not compatible with gradient propagation

GEnome-scale Metabolic Model (GEM/FBA)

$$\left\{\begin{array}{l} \text{Max } (v_{biomass}) \\ \\ \text{Subjected to contraints:} \\ \quad S\,V = 0 \\ \quad 0 \leq V \leq V_{in} \\ \\ \text{where} \\ - V = \text{set of all reaction fluxes} \\ - \text{S} \; = \text{stochiometric matrix} \\ - V_{in} = \text{uptake medium fluxes upper bounds} \end{array}\right.$$

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



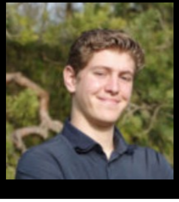


gradient backpropagation to find mapping between medium concentrations and uptake fluxes

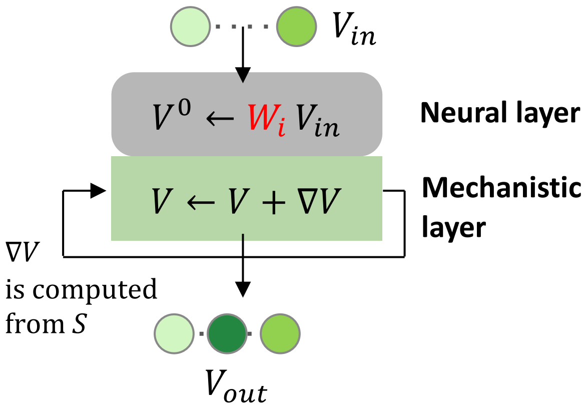
Conventional Reservoir should:

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

AMNs (Artificial Metabolic Networks): a gradient backpropagation compatible method surrogating classical mechanistic models



Physics informed neural network (PINN)



V_{in}

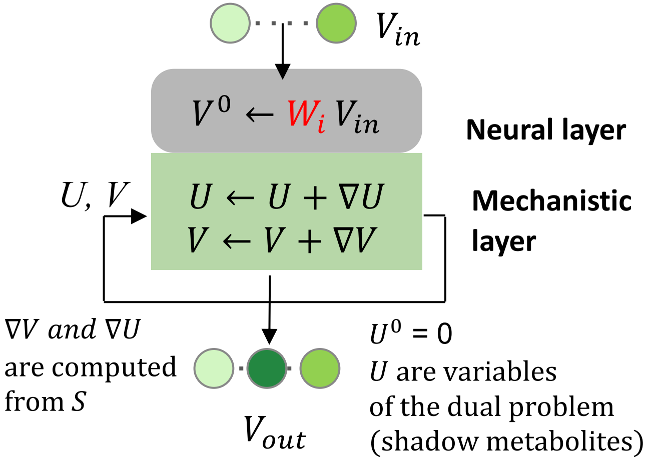
$V^0 \leftarrow W_i \, V_{in}$ Neural layer

$V \leftarrow V + \nabla V$ Mechanistic layer

∇V is computed from S

V_{out}

Hopfield's network



V_{in}

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

$U, \, V$

$U \leftarrow U + \nabla U$

$V \leftarrow V + \nabla V$ Mechanistic layer

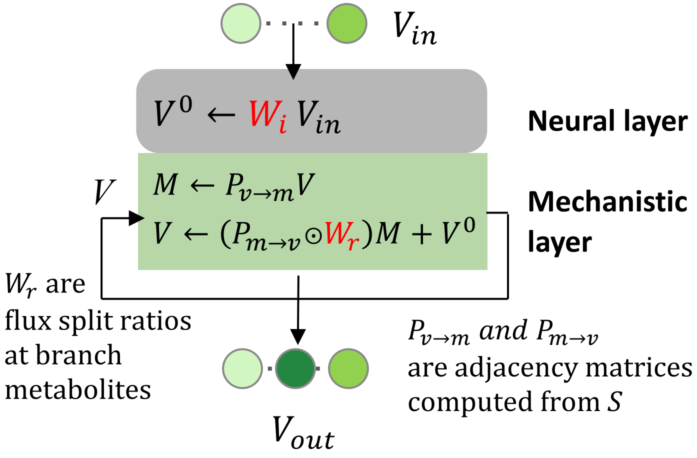
∇V *and* ∇U are computed from S

V_{out}

$U^0 = 0$

U are variables of the dual problem (shadow metabolites)

Graph neural network (GNN)



V_{in}

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

V

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

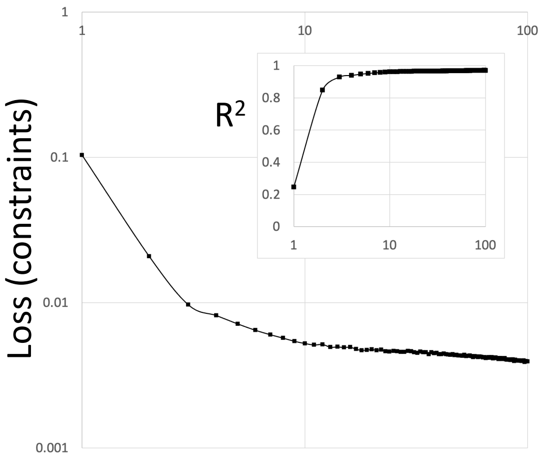
$V \leftarrow (P_{m \to v} \odot W_r) M + V^0$ Mechanistic layer

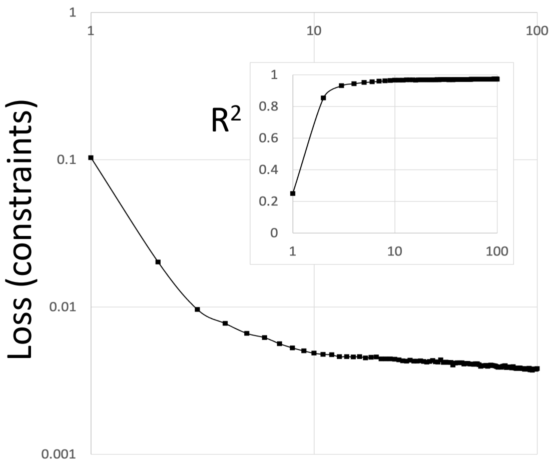
W_r are flux split ratios at branch metabolites

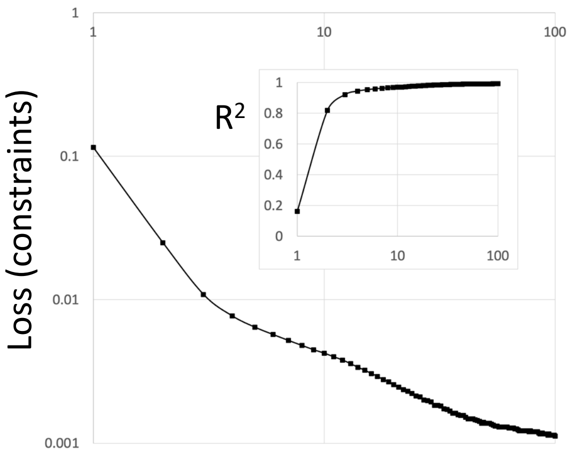
V_{out}

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

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)

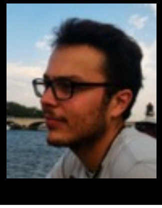


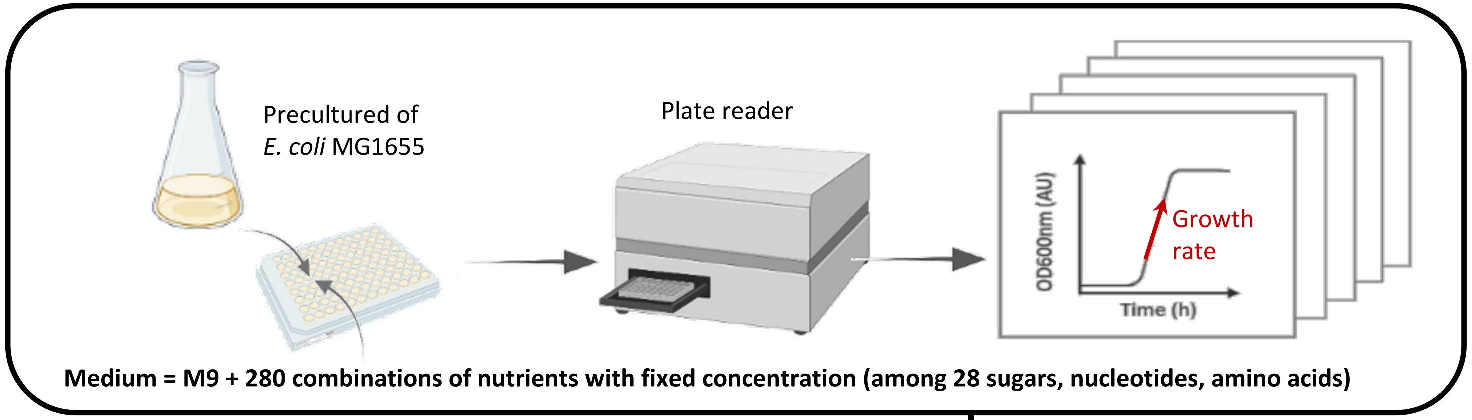




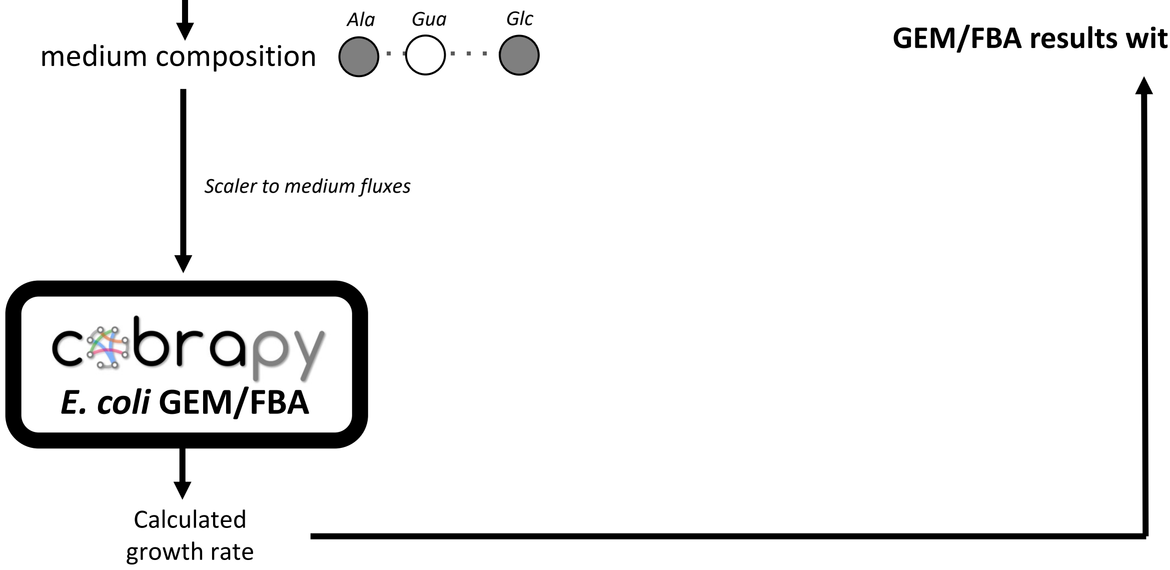
- Faure L. *et al*. *Nat Commun* **14**, 4669 (2023)

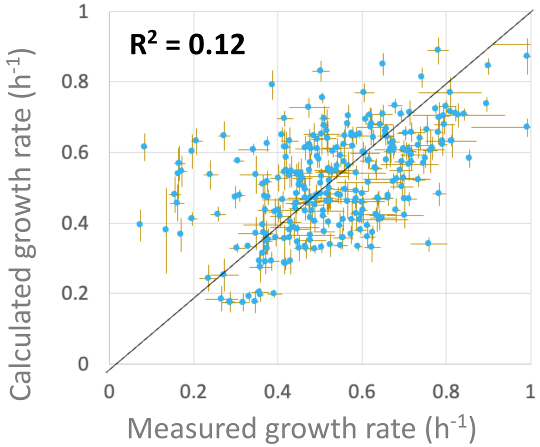
AMNs can be used as reservoir in RC to improve mechanistic model predictability





Medium = M9 + 280 combinations of nutrients with fixed concentration (among 28 sugars, nucleotides, amino acids)

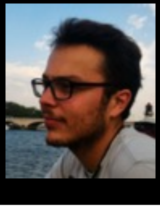


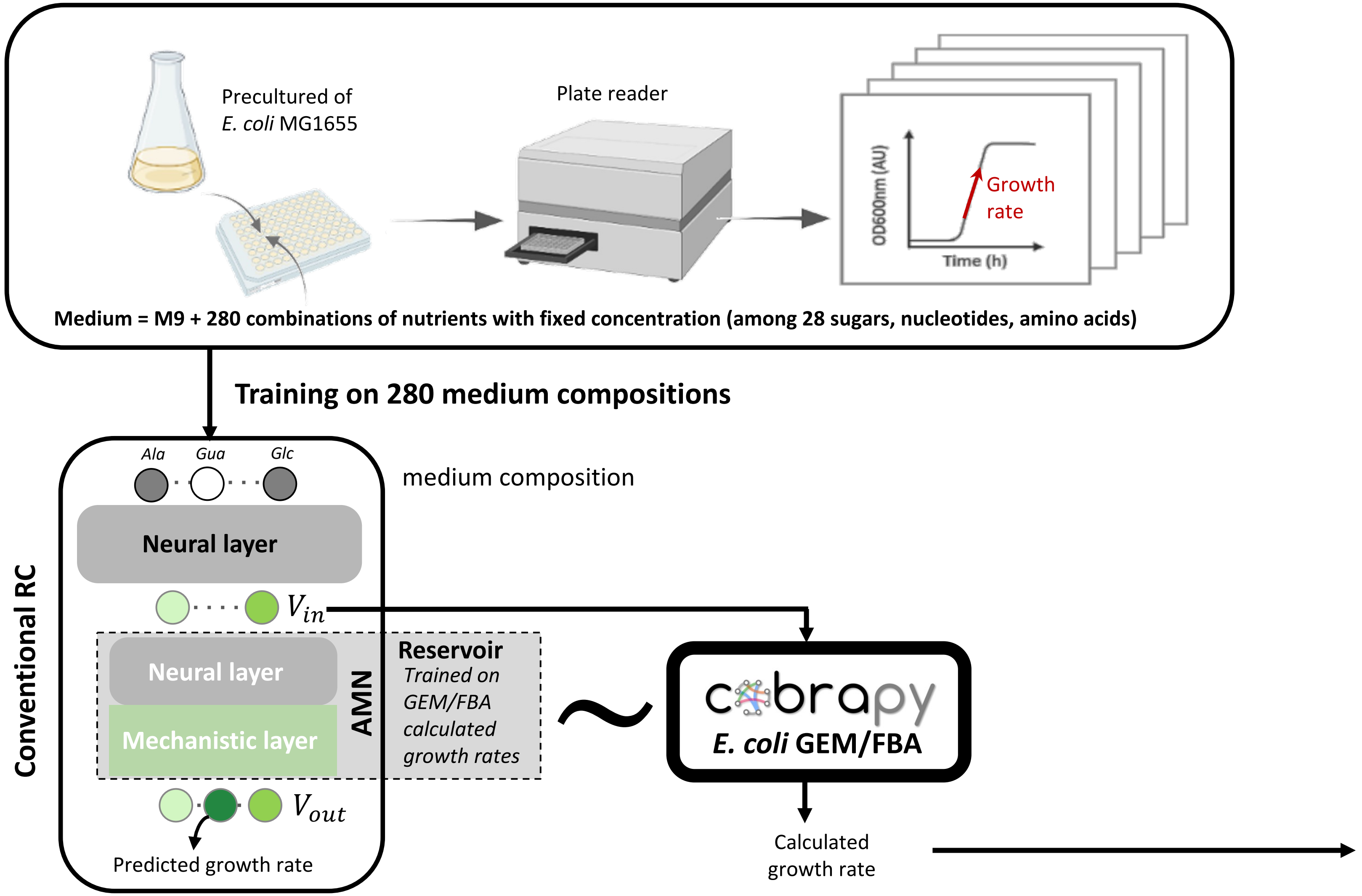


GEM/FBA results with best scaled input

- Faure L. *et al*. *Nat Commun* **14**, 4669 (2023) & Faulon et al. *bioRxiv DOI: 10.1101/2024.09.12.612674* (2024)

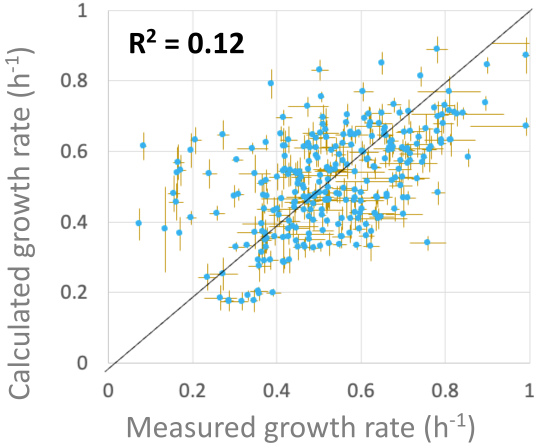
AMNs can be used as reservoir in RC to improve mechanistic model predictability



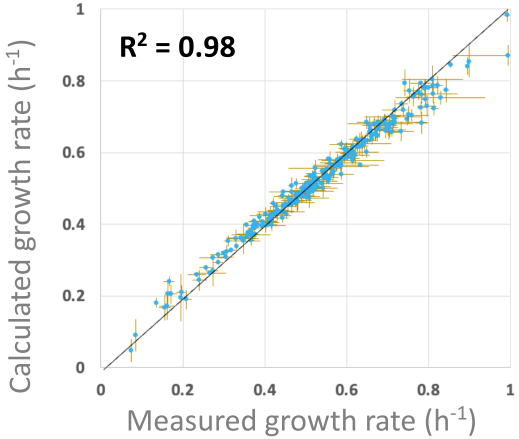


Medium = M9 + 280 combinations of nutrients with fixed concentration (among 28 sugars, nucleotides, amino acids)

Training on 280 medium compositions



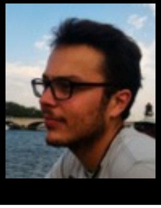
GEM/FBA results with best scaled input

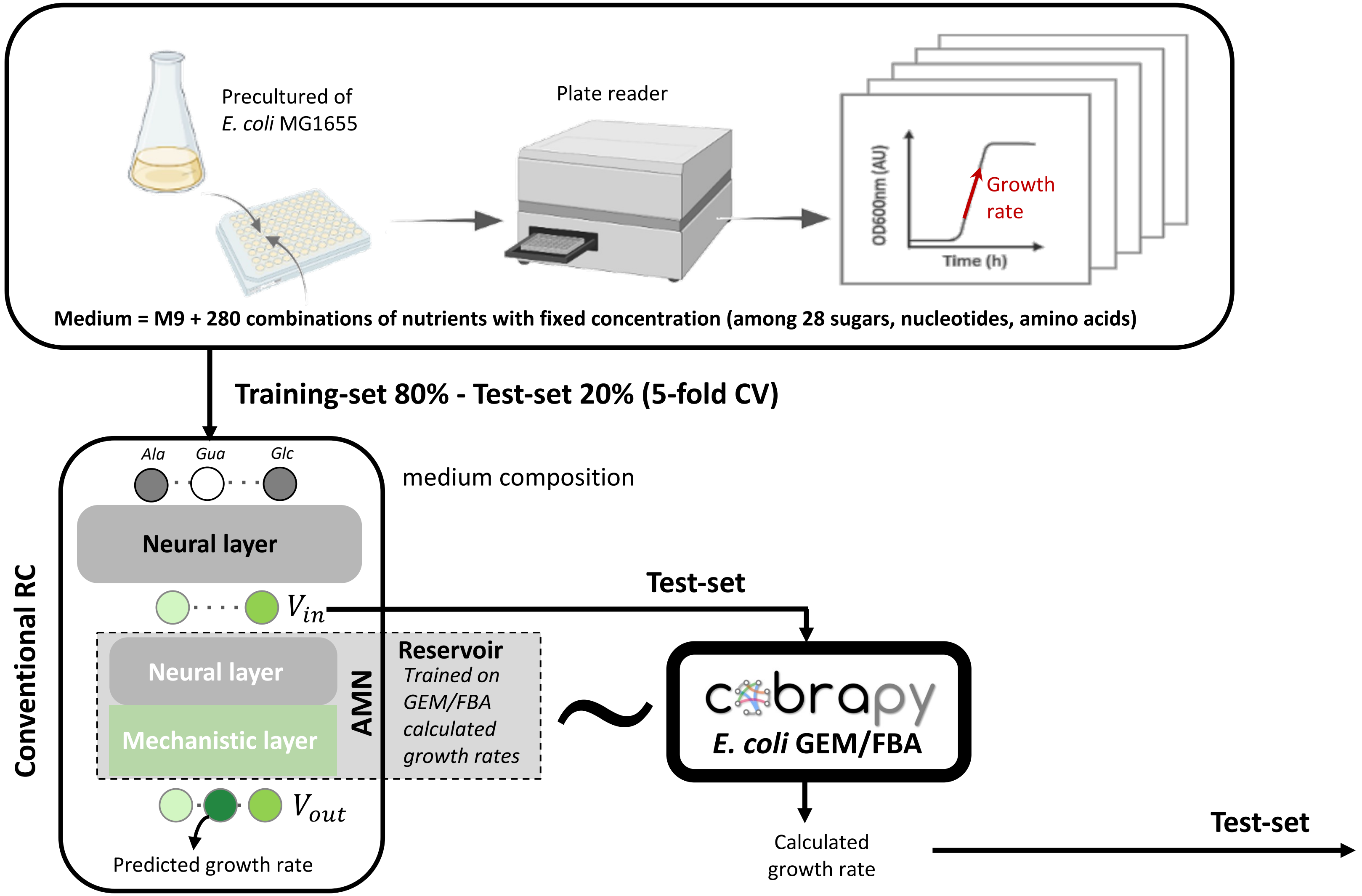


GEM/FBA results with reservoir inputs

- Faure L. *et al*. *Nat Commun* **14**, 4669 (2023) & Faulon et al. *bioRxiv DOI: 10.1101/2024.09.12.612674* (2024)

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





Precultured of *E. coli* MG1655

Plate reader

Medium = M9 + 280 combinations of nutrients with fixed concentration (among 28 sugars, nucleotides, amino acids)

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

medium composition

Neural layer

V_{in}

Test-set

Conventional RC

Neural layer

Mechanistic layer

AMN

Reservoir
Trained on GEM/FBA calculated growth rates

~

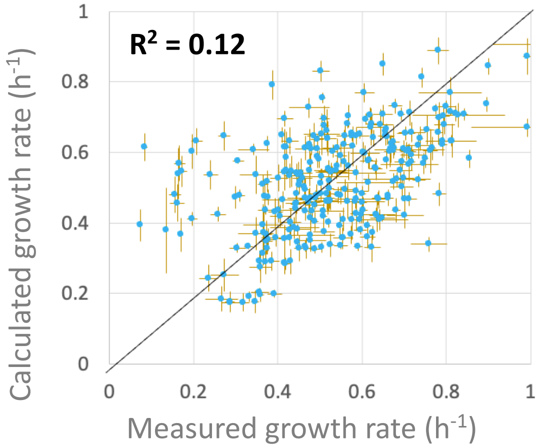
***E. coli* GEM/FBA**

V_{out}

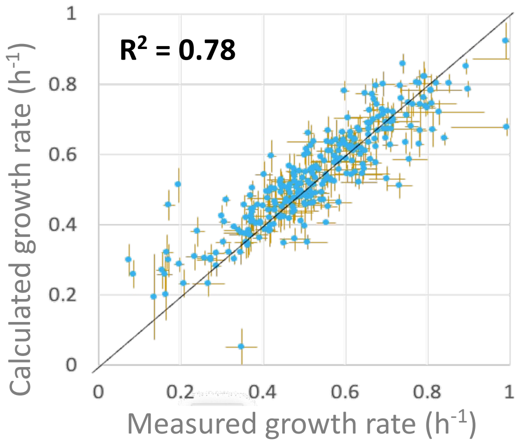
Predicted growth rate

Calculated growth rate

Test-set



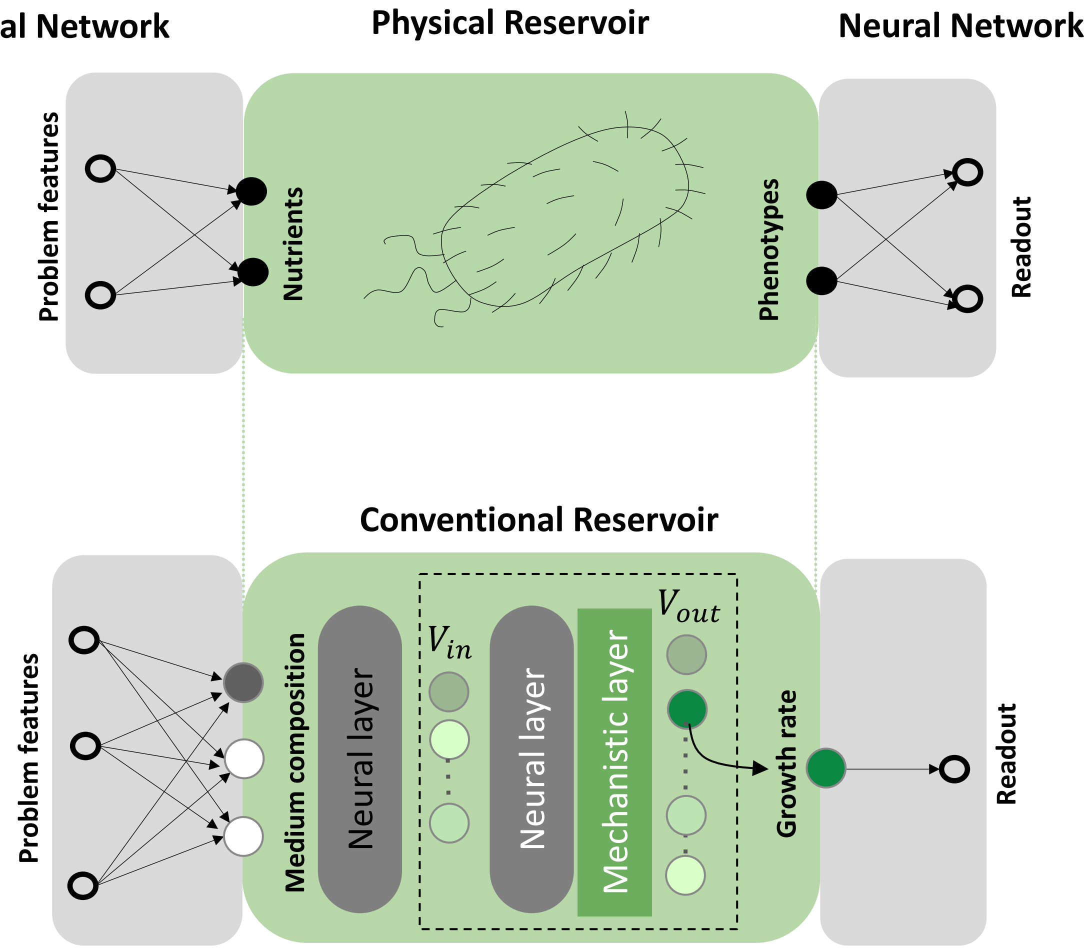
GEM/FBA results with best scaled input



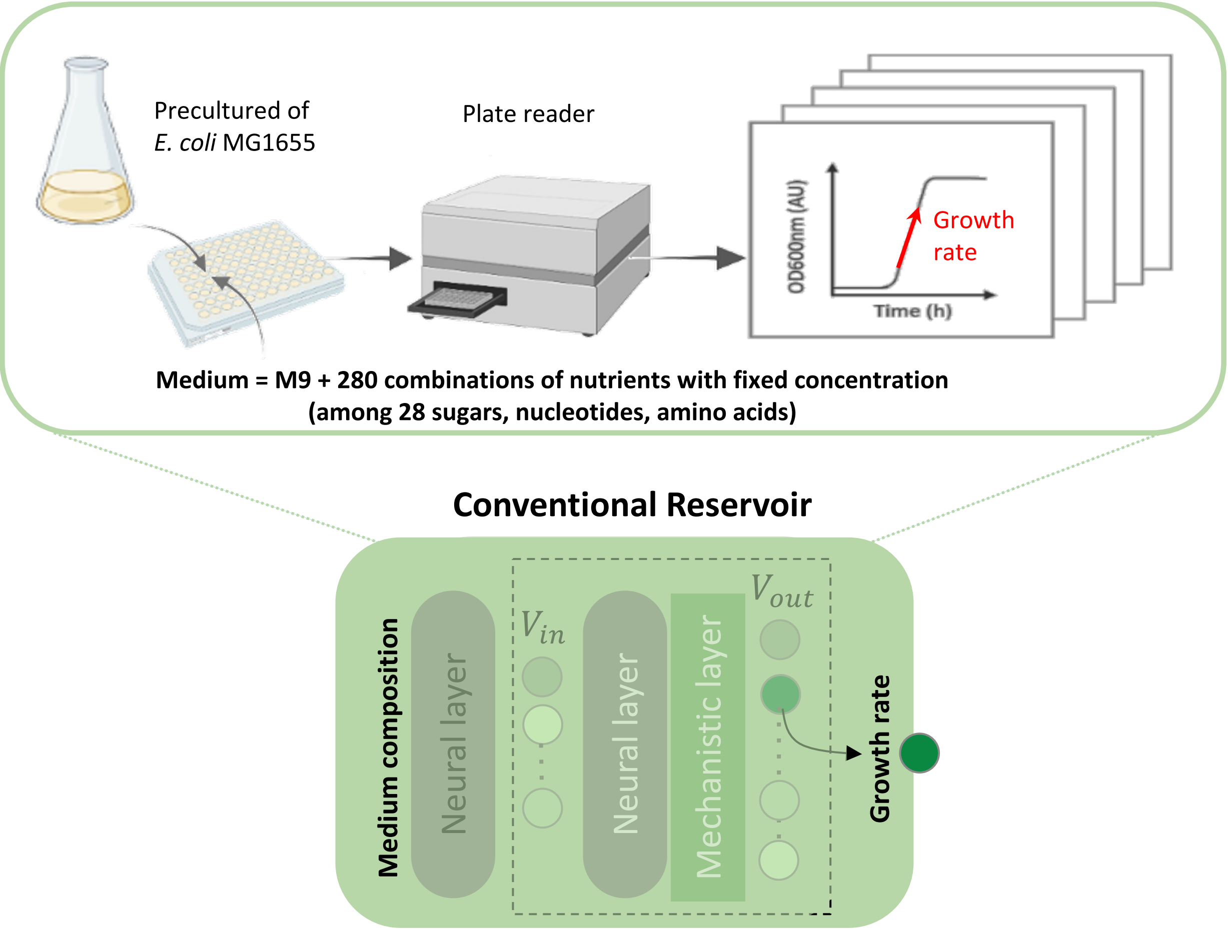
GEM/FBA results with reservoir inputs

- Faure L. *et al*. *Nat Commun* **14**, 4669 (2023) & Faulon et al. *bioRxiv DOI: 10.1101/2024.09.12.612674* (2024)

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



Building an E. coli RC to solve a regression problem



Precultured of
E. coli MG1655

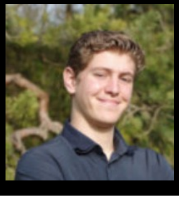
Plate reader

Medium = M9 + 280 combinations of nutrients with fixed concentration (among 28 sugars, nucleotides, amino acids)

Conventional Reservoir

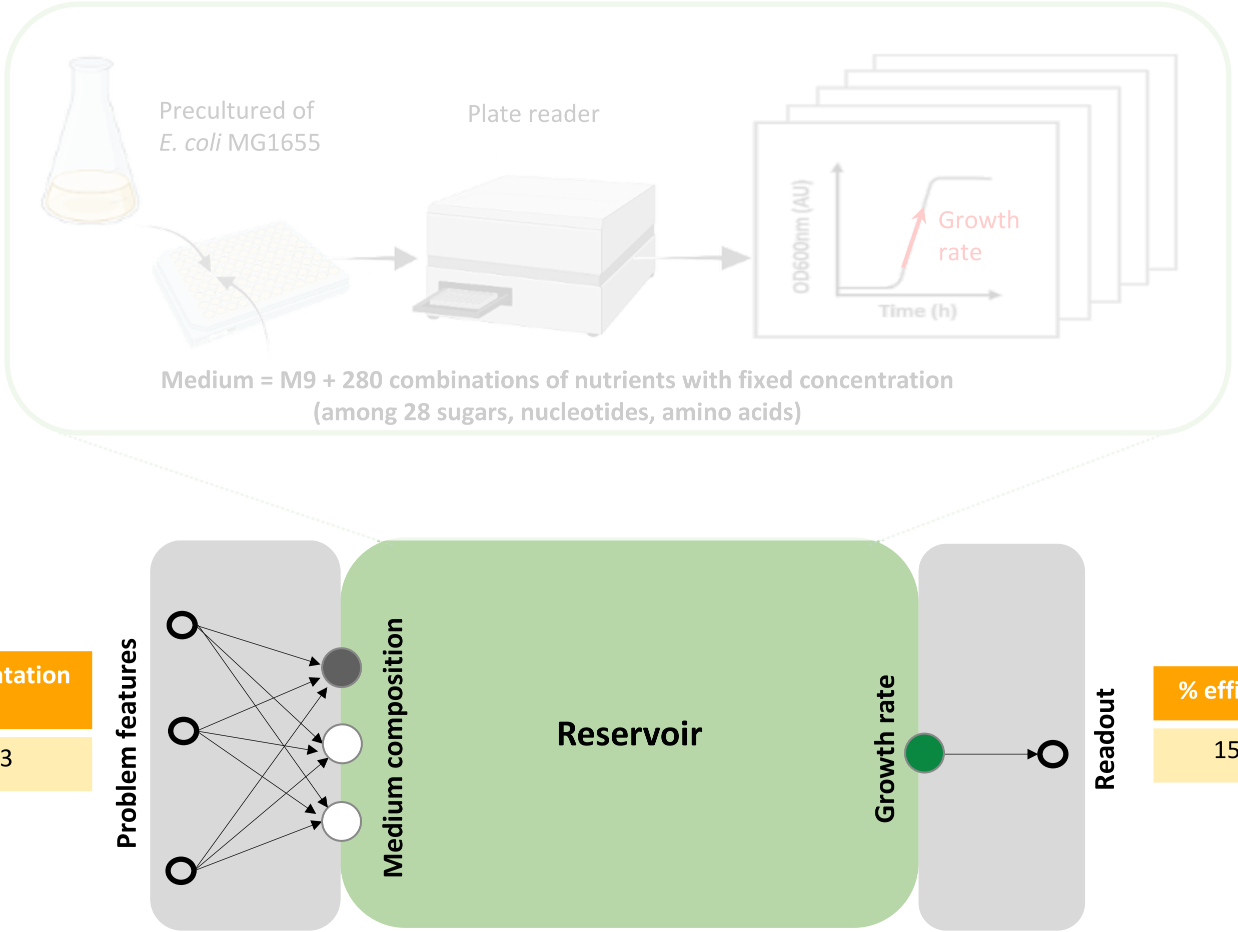
Using E. coli RC to solve a regression problem

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

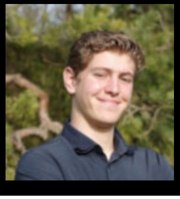
Glc	Xyl	Succ	Trp
1	0	1	1

µ
0.71

- Faulon et al. *bioRxiv DOI: 10.1101/2024.09.12.612674* (2024)

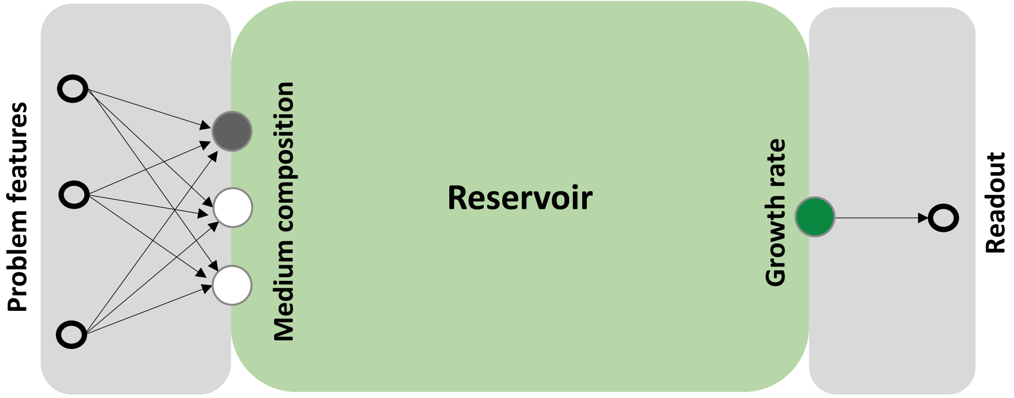
Using E. coli RC to solve a regression problem

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



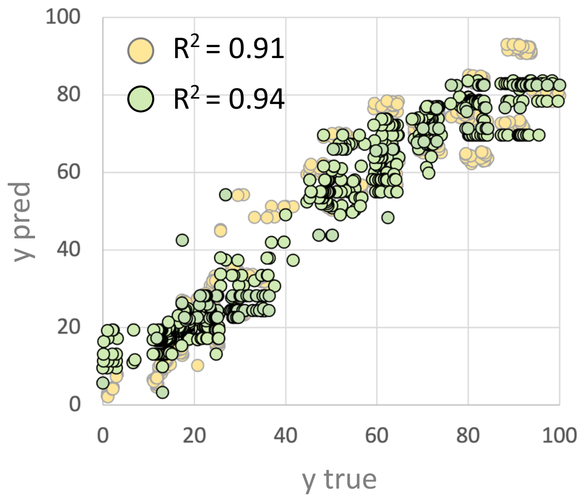
Glc	Xyl	Succ	Trp
1	0	1	1

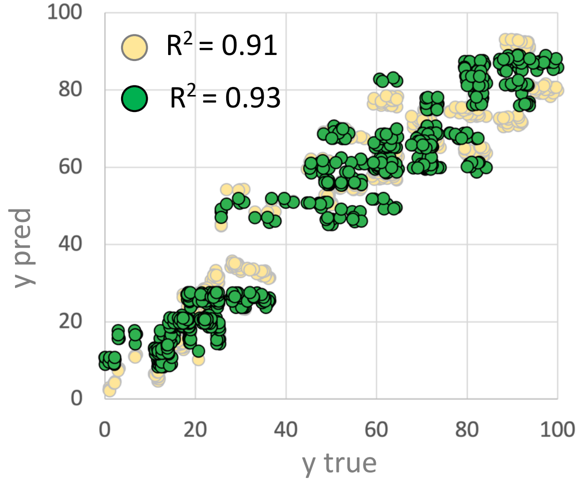
μ
0.71

% efficiency
15.55

○ MLR
○ Conventional RC
● Physical RC

5-fold CV

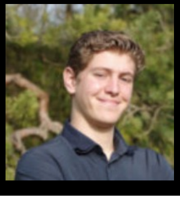




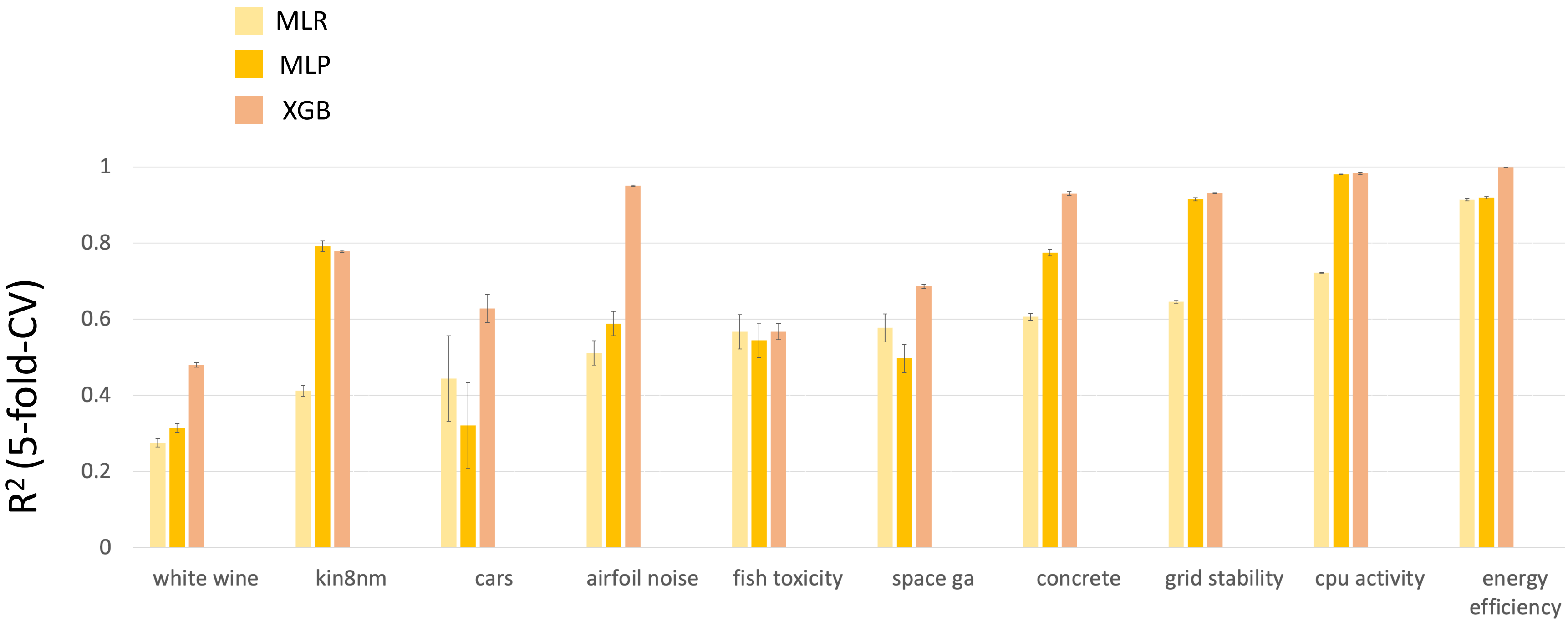
- Faulon et al. *bioRxiv DOI: 10.1101/2024.09.12.612674* (2024)

Using E. coli RC to solve many regression problems

OpenML
A worldwide machine learning lab



10 OpenML regression problems of increasing difficulty

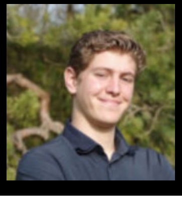


- Faulon et al. *bioRxiv DOI: 10.1101/2024.09.12.612674* (2024)

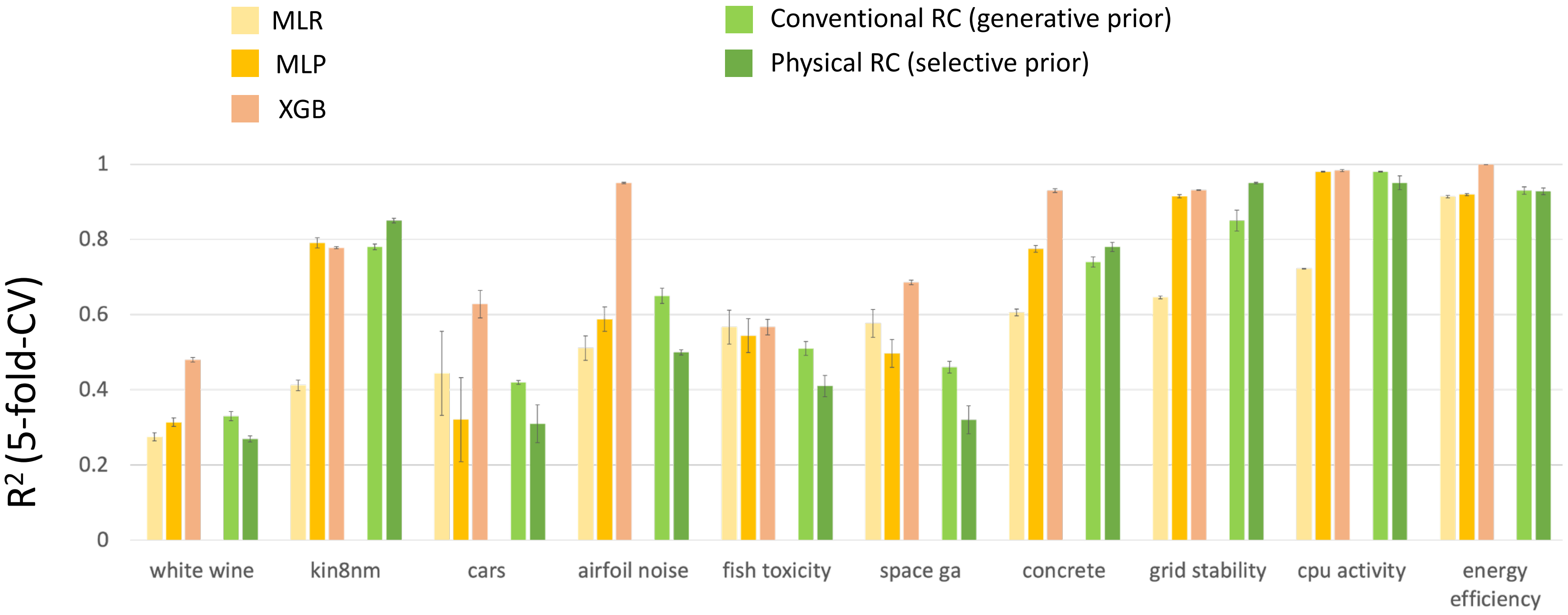
Using E. coli RC to solve many regression problems

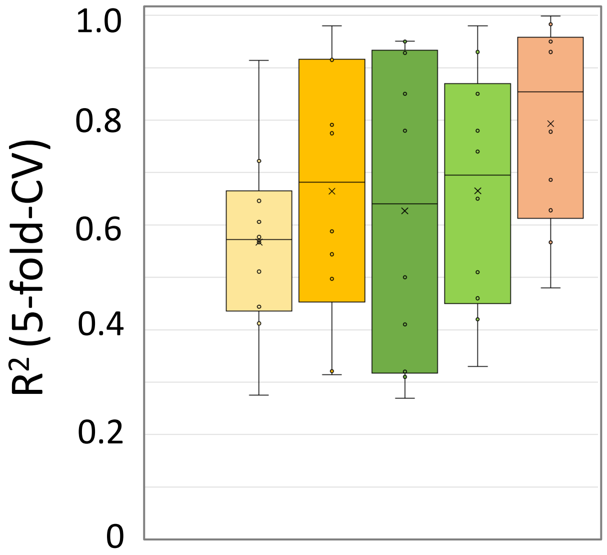
OpenML

A worldwide machine learning lab



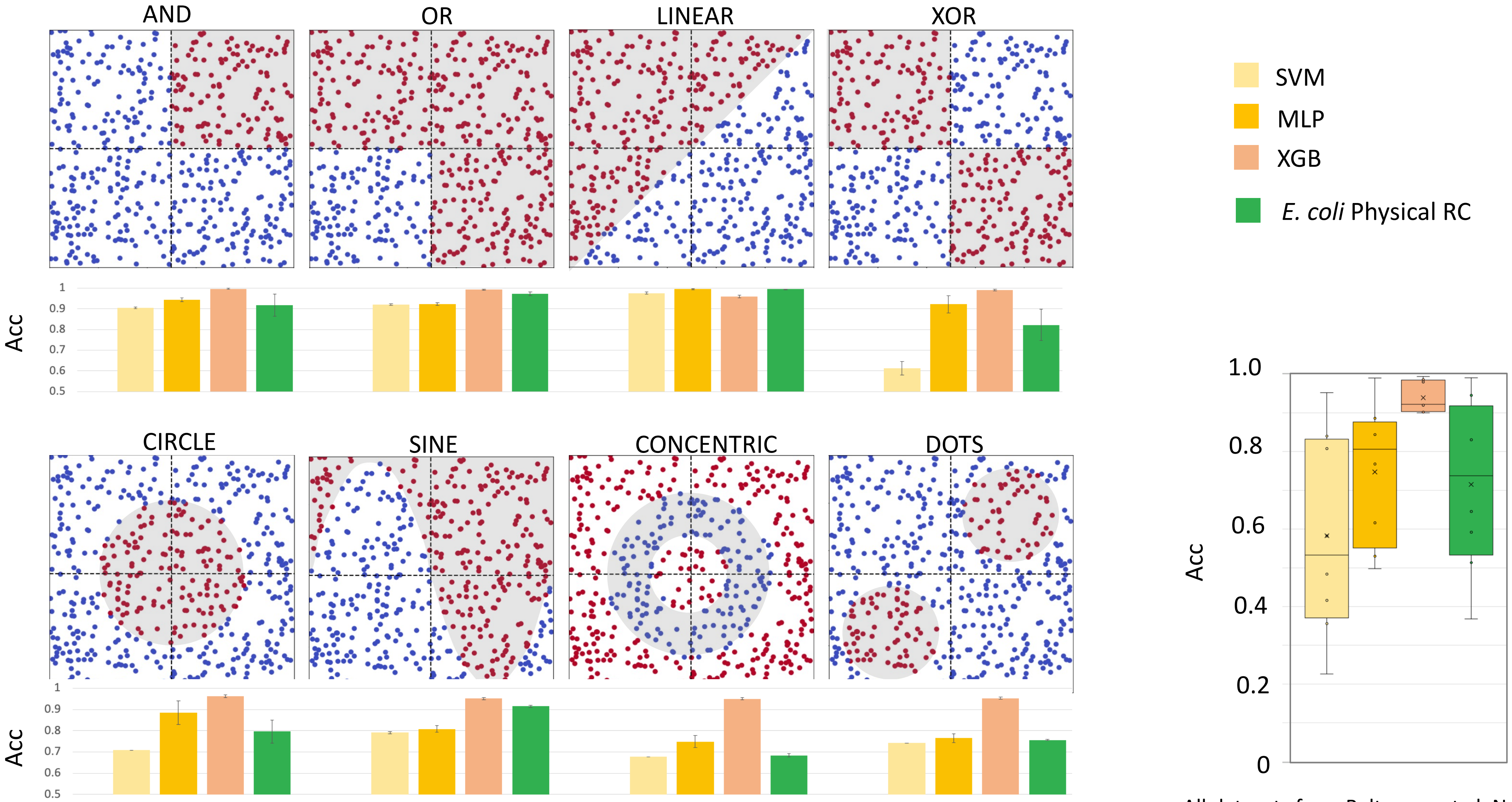
10 OpenML regression problems of increasing difficulty





- Faulon et al. *bioRxiv DOI: 10.1101/2024.09.12.612674* (2024)

Using E. coli RC to solve classification problems



- All datasets from Baltussen et al. Nature 2024

- Faulon et al. *bioRxiv DOI: 10.1101/2024.09.12.612674* (2024)

Can E. coli RC solve a concrete classification problem ?

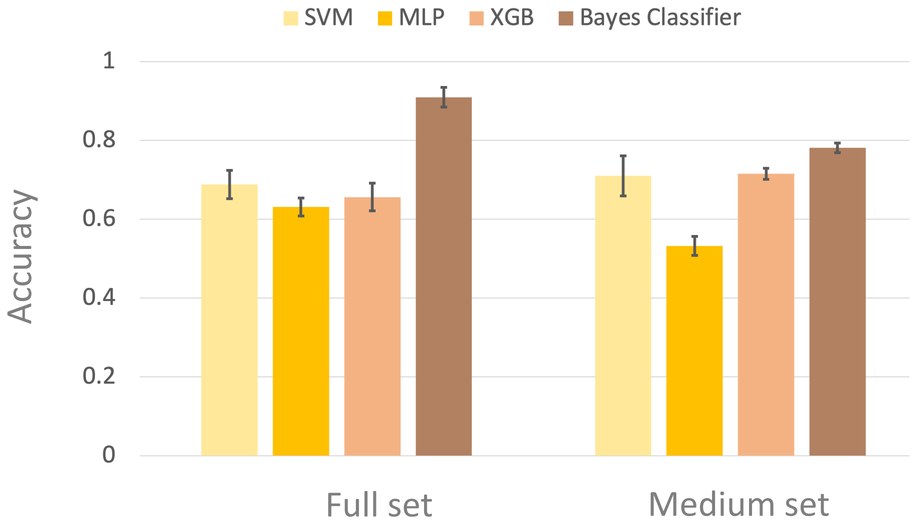
The problem:

- 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)

Classifier performances (20-fold CV results)



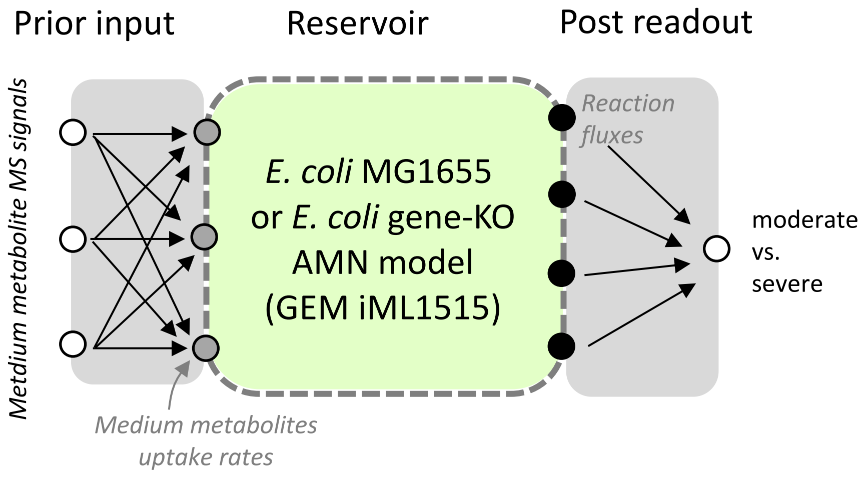
Accuracy = 0.84 in Shen et al. *Cell* 2020; 182(1): 59–72

Using E. coli RC for classification

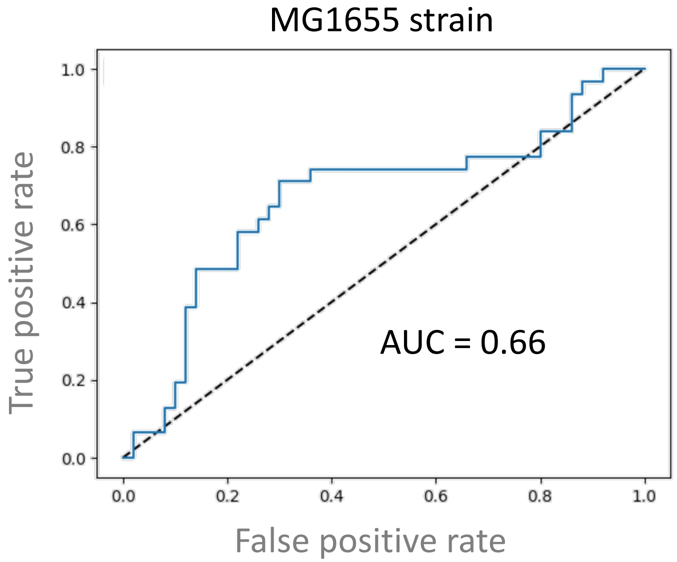
The problem:

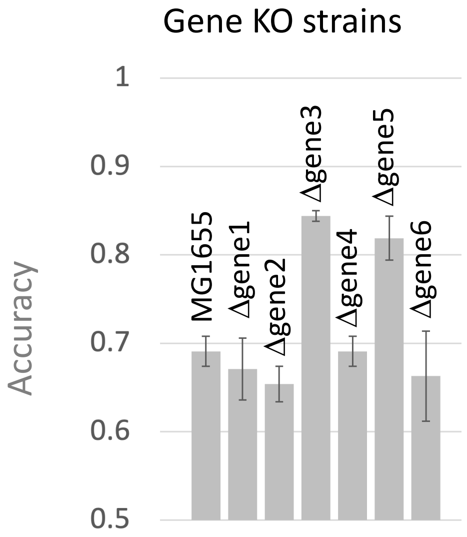
- Blood sample are collected for Covid-19 patients once they enter the hospital
- ~~Metabolomics analyses are carried out on the samples~~
- ~~Can we predict from the analyses if the disease outcome will be severe or moderate?~~
- Can we use an *E. coli* RC grown on the patient's sample to predict if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

Conventional RC to predict disease outcome from phenotype

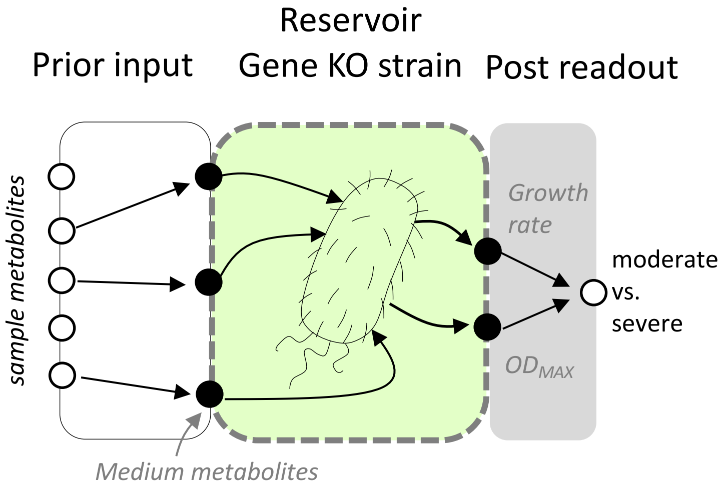


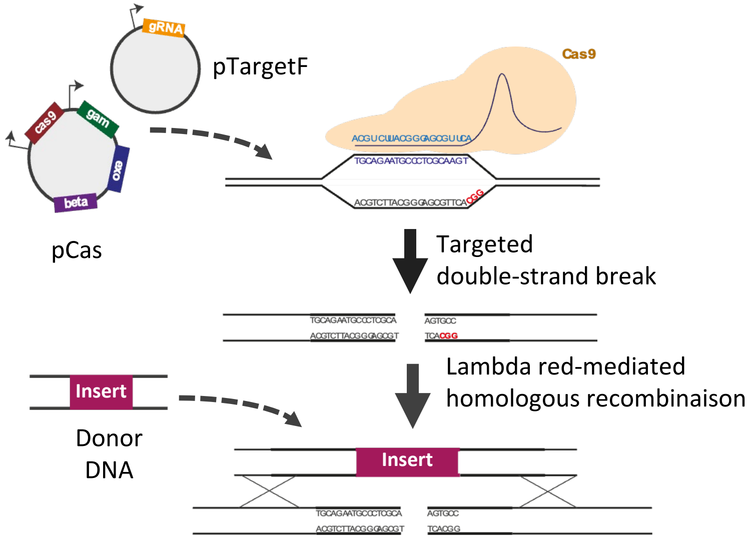
Prior input
Reservoir
Post readout
Metdium metabolite MS signals
E. coli MG1655 or *E. coli* gene-KO AMN model (GEM iML1515)
Reaction fluxes
moderate vs. severe
Medium metabolites uptake rates



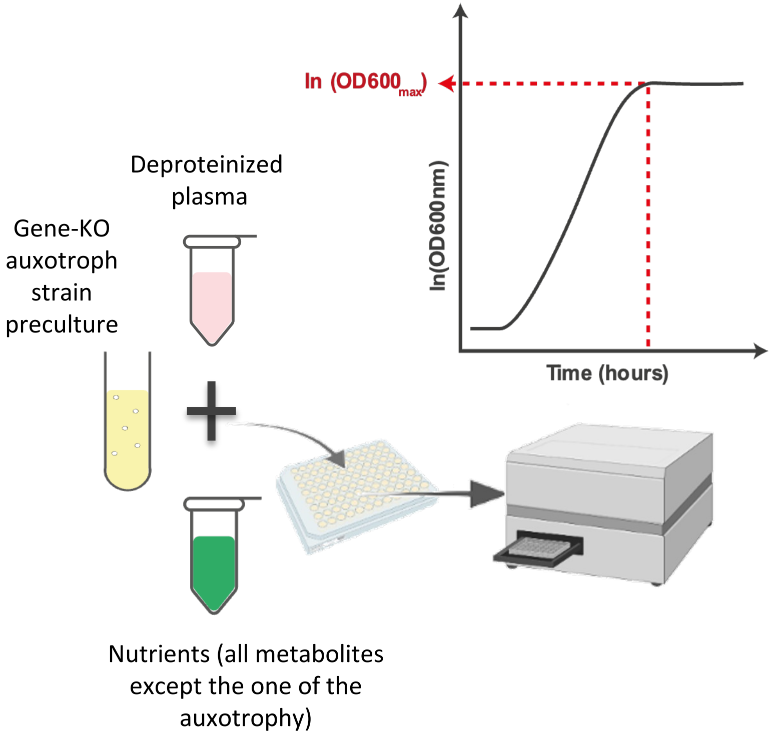


Building an *E. coli* physical RC for classification





gRNA
pTargetF
cas9
gam
exo
beta
pCas
Cas9
Targeted double-strand break
Lambda red-mediated homologous recombinaison
Insert
Donor DNA
Insert



ln (OD600max)
ln(OD600nm)
Time (hours)
Deproteinized plasma
Gene-KO auxotroph strain preculture
+
Nutrients (all metabolites except the one of the auxotrophy)

- Gene-KO *E. coli* Physical RC to predict disease outcome from growth rate and OD_{MAX}

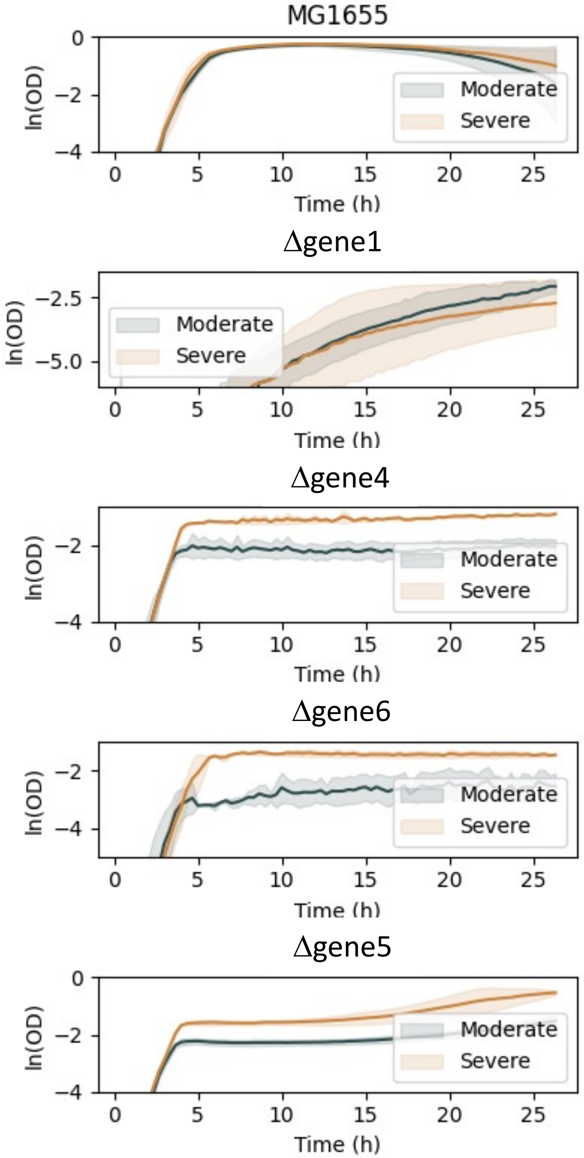
- CRISPR-Cas9/Lambda red system

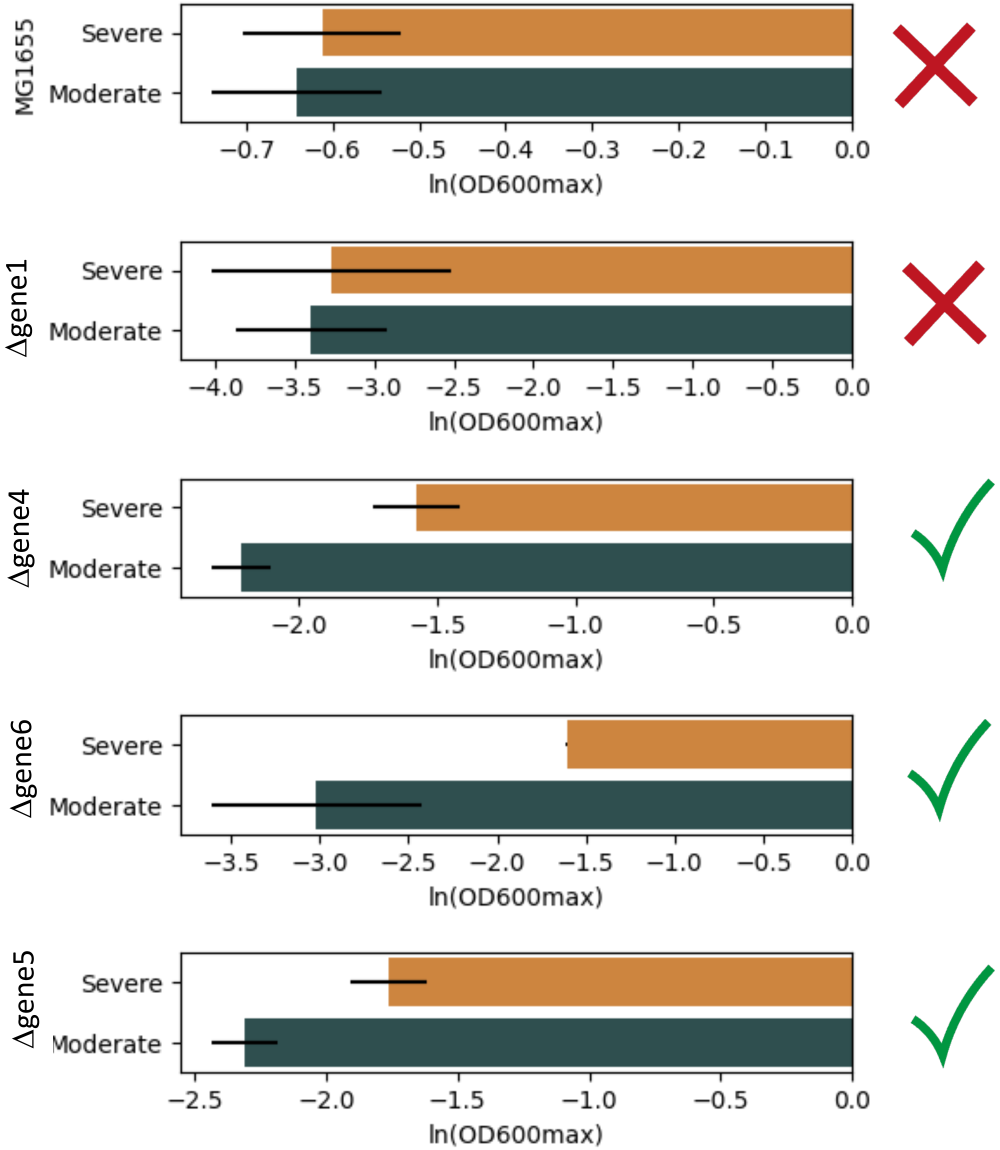
  - Jiang et al. , *Appl Environ Microbiol*, 2015
  - Scarless, Efficient, Multiplexable
  - ~80 KOs built

- Gene deletions force *E. coli* to collect specific nutrients from the plasma in order to grow

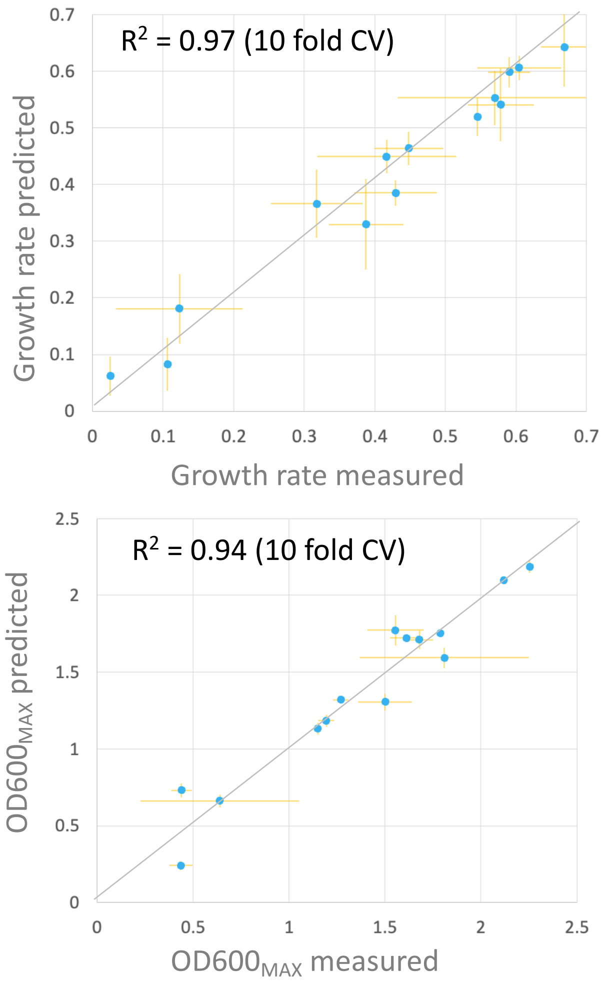
- According to conventional RC, differences of nutrients concentration in the plasma should result in different growth curve

Benchmarking E. coli physical RC for classification





Conventional reservoir predictions (8 moderate – 6 severe)



Perspectives

- Supervised Learning & Active learning
  - New generative AI models (transformer) for retro-(bio)synthesis and to generate sequences
  - LLMs (like GPT4) to drive biofoundries
  - Active Learning / Transfer Learning / Hybrid Learning to cope with small training set sizes

- *in vitro/in vivo* learning
  - 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: perhaps one should to consider building devices top-down, *i.e*. exploiting/modifying hosts rather than plugging orthogonal devices.

Acknowledgments

anr

FRANCE 2030

AMN
SynBioDiag

B-BEST

HORIZON

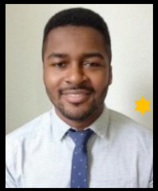
BIOS SUSTAINABLE BIO-MANUFACTURING

université PARIS-SACLAY

INRAE

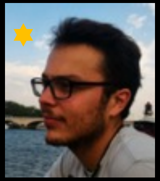
DGA

l'X ÉCOLE POLYTECHNIQUE


**Paul
Ahvi**

**An
Hoang**

**Bastien
Mollet**


**Léon
Faure**

**Amir
Pandi**

**Mathilde
Koch**

**Olivier
Borkowski**

**Anne
Giralt**

**Joan
Hérisson**

★ Molecular Biology
★ Computational Biology

Jérôme Bonnet’s group

CBS Inserm

Tobby Erb’s group

MAX PLANCK INSTITUTE
FOR TERRESTRIAL MICROBIOLOGY