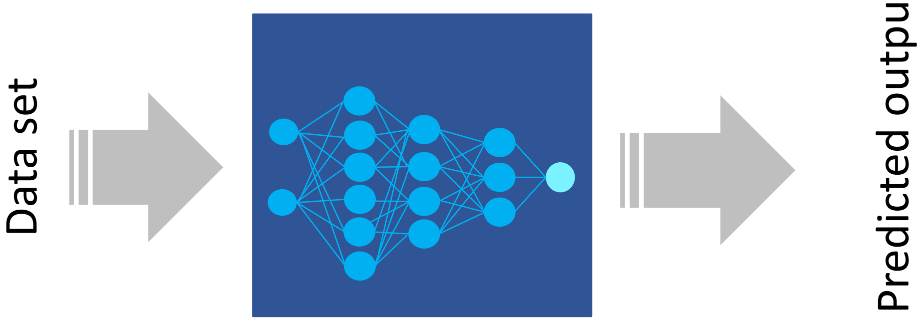
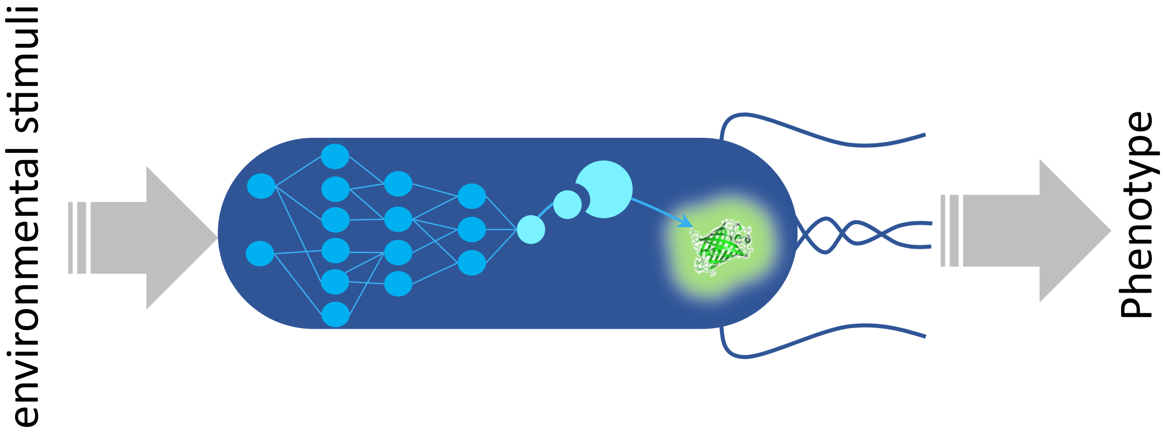
Artificial Metabolic Networks for neural computations with metabolism

Léon Faure, Bastien Mollet, Wolfram Leibermeister, and Jean-Loup Faulon

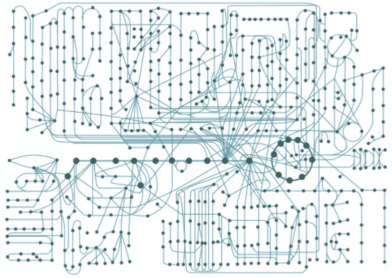




bioRxiv 2022.01.09.475487

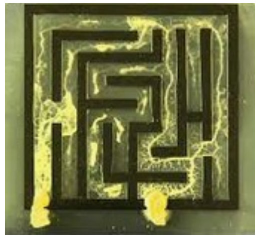
Jean-Loup.Faulon@inra.fr
www.jfaulon.com

Context

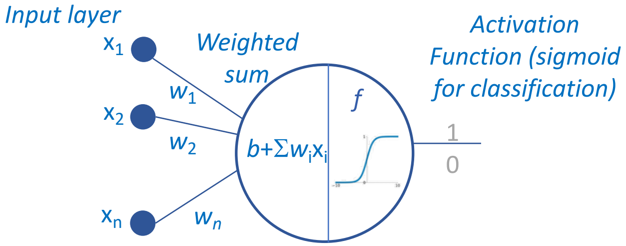


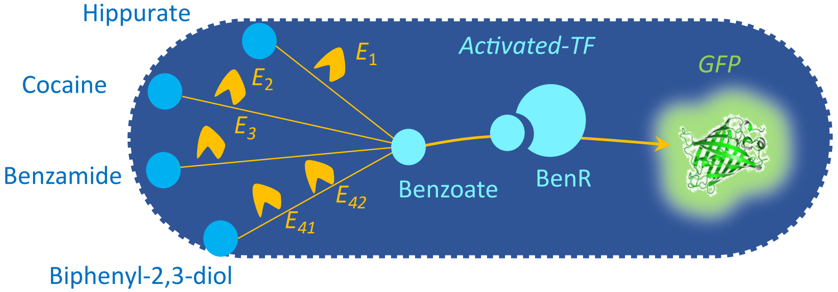
- Metabolism is to process food and it is not *a priori* an information processing apparatus

- Yet, metabolism is also involved in cell signaling – metabolic fluctuations allow signals to spread

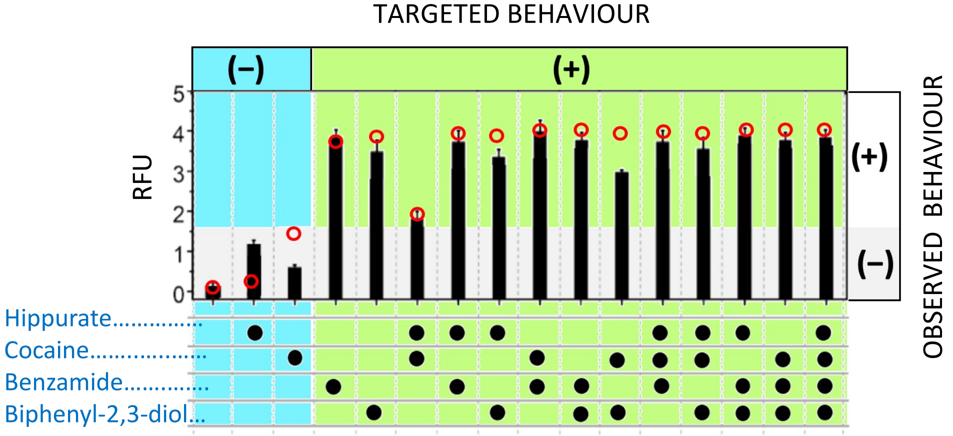


- We have built in past work a metabolic perceptron in cell-free systems enabling to classify samples based on their metabolic compositions





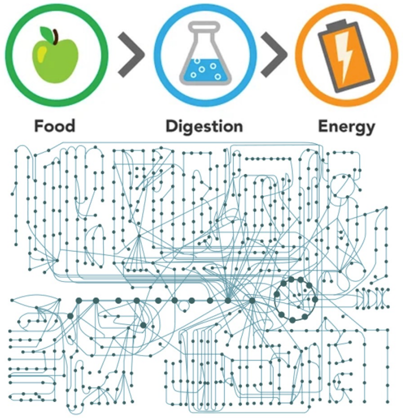
weight instantiated by DNA enzyme concentration

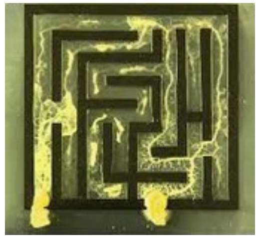


Kinetics model predictions also used to estimate DNA enzyme concentrations

- Pandi A. *et al*.., ***Nature commun.*** 10: 3880, 2019. | DOI: [10.1038/s41467-019-11889-0](https://doi.org/10.1038/s41467-019-11889-0)

Context

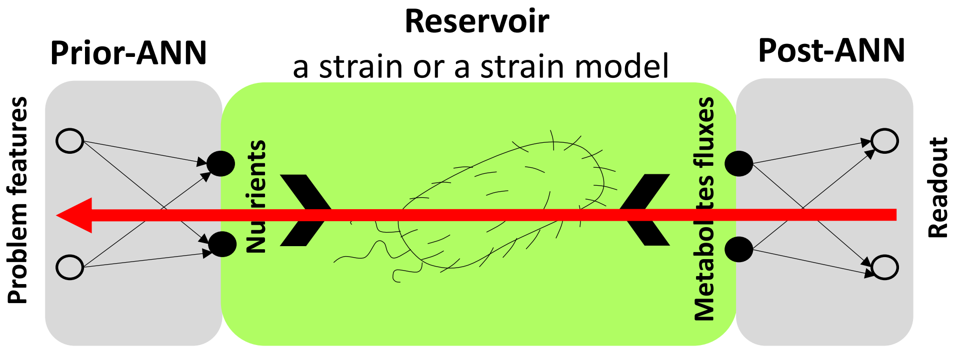




- Metabolism is to process food and it is not *a priori* an information processing apparatus
- Yet, metabolism is also involved in cell signaling – metabolic fluctuations allow signals to spread
- We have built in past work a metabolic perceptron in cell-free systems enabling to classify samples based on their metabolic compositions

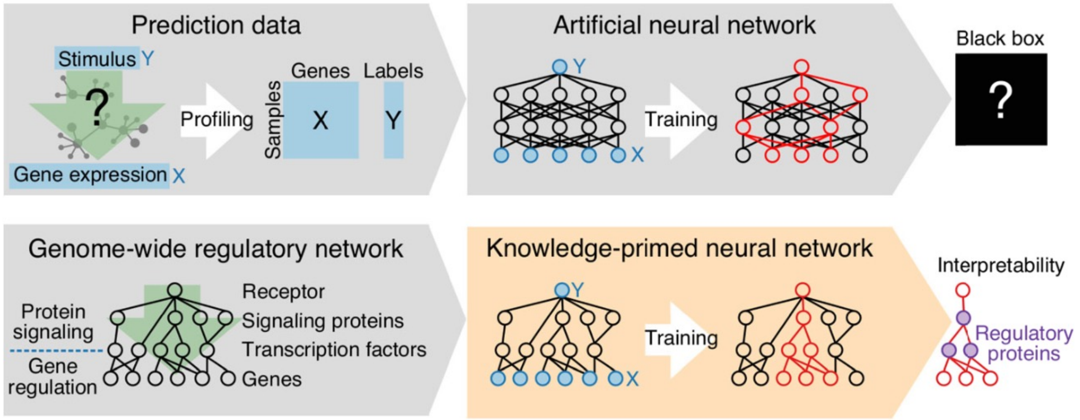
- ***Can we do this in vivo to probe if species metabolism can be diverted to tackle problems that are typically solved in silico ?***

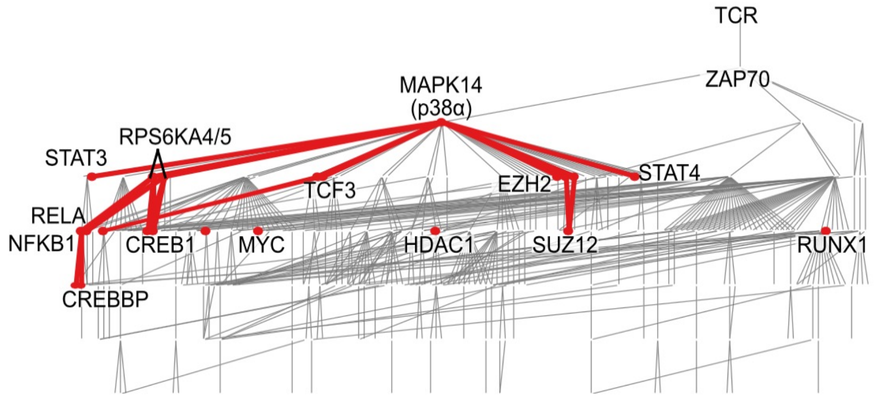
gradient backpropagation



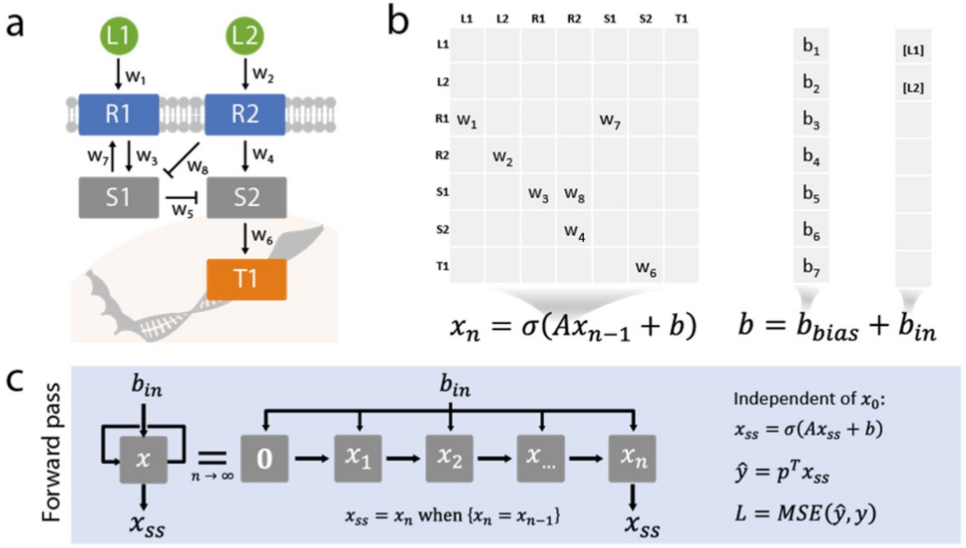
a model that allow gradient backpropagation and accurately reproduce phenotype for different media composition

State-of-the-art (for signaling and regulatory networks): Knowledge-primed neural networks (KPNN)

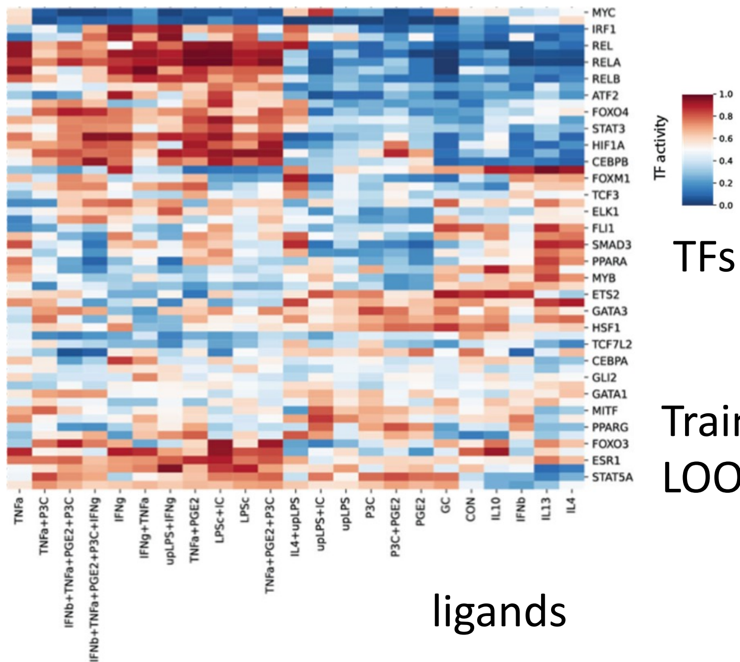




- Fortelny N., Bock C***. Genome Biol.*** 2020 Aug 3;21(1):190. doi: 10.1186/s13059-020-02100-5



- Nilsson A. *et al*. ***bioRxiv*** 2021.09.24.461703



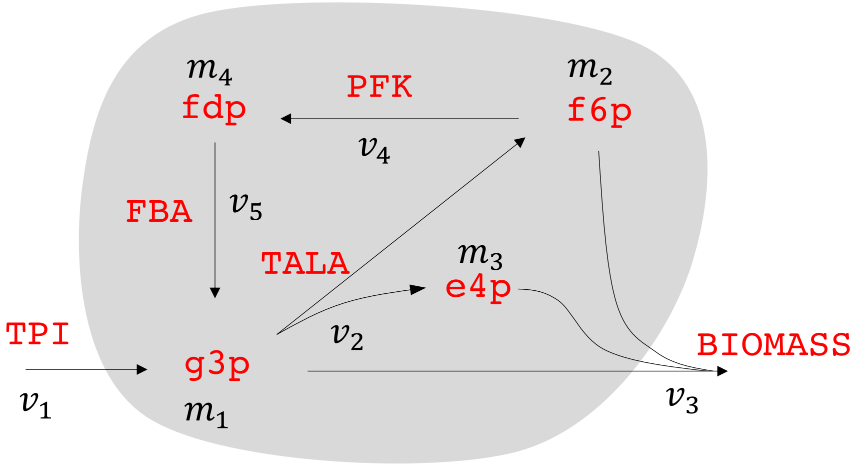
TFs

Training set: R^2 = 0.95

LOOCV: Q^2 = 0.35

ligands

Encoding metabolic network into neural network: KPNN solution



Let $V = (v_1,.., v_5)^T$ and $c = (0,0,1,0,0)^T$:
Solve Linear Program (simplex algorithm):
Max $(c^T V)$
Subjected to:

$S\,V = 0$

$0 \le V \le V^0$

where $V^0 = (0.1, \infty, \infty, \infty, \infty)^T$
and

$$S = \begin{bmatrix} 1 & -1 & -0.13 & 0 & 1 \\ 0 & 1 & -0.07 & -1 & 0 \\ 0 & 1 & -0.36 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 \end{bmatrix}$$

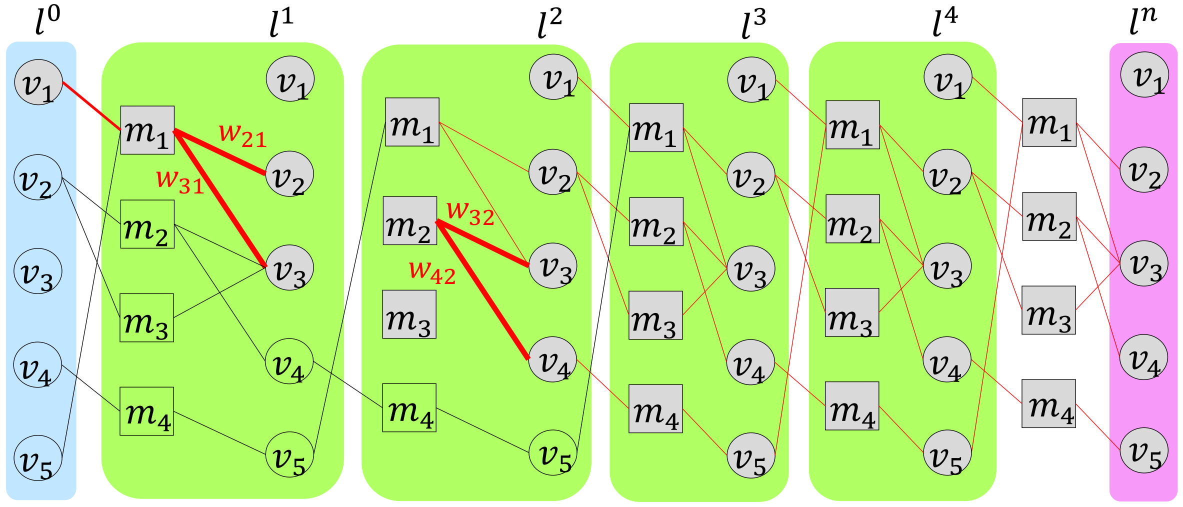
Steady state solution:
$V^{sst} = (0.10, 0.18, 0.50, 0.15, 0.15)^T$

$$\mathrm{V2M} = \mathrm{ReLU}\,(\mathrm{S}) = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

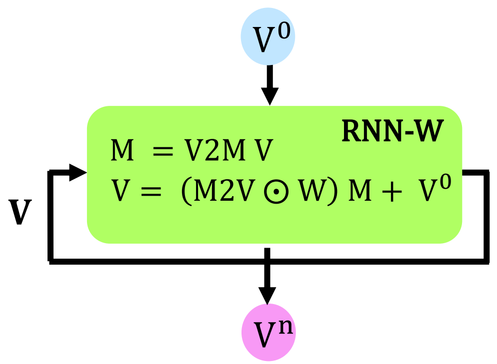
$$\mathrm{M2V} = \mathrm{ReLU}\left(\left[-1/s_{j,i}\right]\right) = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 7.75 & 14.10 & 2.77 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$\mathrm{W}^{sst} = \mathrm{ReLU}\left(\left[s_{j,i} v_i^{sst} / m_j^{sst}\right]\right) = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0.74 & 0 & 0 & 0 \\ 0.26 & 0.20 & 1 & 0 \\ 0 & 0.80 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

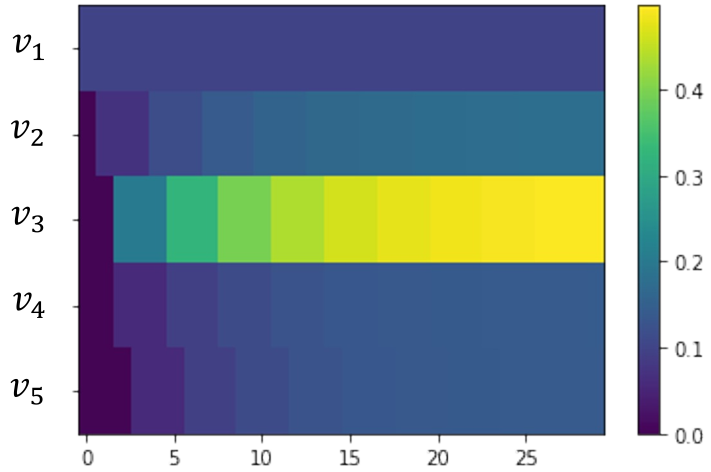
At steady state:
$M^{sst} = \mathrm{V2M}\ V^{sst} = (0.25, 0.18, 0.18, 0.15)^T$
$V^{sst} = (\mathrm{M2V} \odot \mathrm{W}^{sst})\, M^{sst} + V^0$



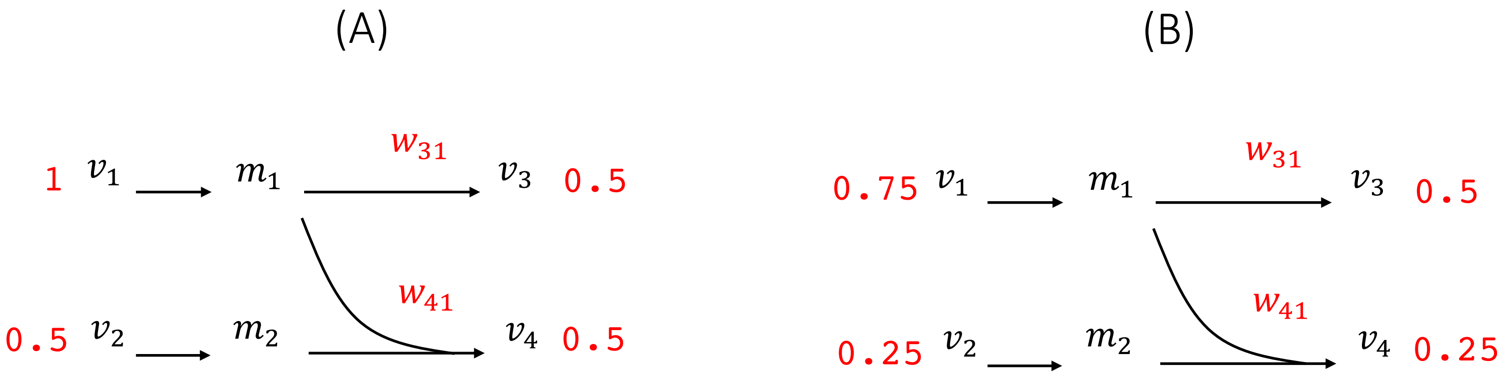
$V^0 = (0.10, 0, 0, 0, 0)^T$



$V^{30} = (0.10, 0.18, 0.50, 0.15, 0.15)^T$



Encoding metabolic network into neural network: some issues



In the two cases all flux values (v_i) satisfy the steady state constraints (SV = 0).

The weights are different in panel (A) and (B) as they depend on the intake flux values.

$w_{31} = 1/2 \;\; w_{41} = 1/2$

$w_{31} = 2/3 \;\; w_{41} = 1/3$

Encoding metabolic network into neural network: RNN cells solving LP

H. Ghasabi-Oskoei et al., *Applied Mathematics and Computation*, 175, 452–464, 2006

V^0

RNN-EB

$$R = \text{ReLU}\,(V + S^T\,M\; - c)$$
$$V = V + \delta t\,(R\; - V)$$
$$M = M +\; \delta t\,(b - S\,R)$$

V, M

V^n

c: objective function
S: stoichiometric matrix
V: flux vector
M: shadow price vector
δt: time step
$b = V2M\,V^0$

Y. Yang, J. et al., *Mathematics and Computers in Simulation*, 101, 103–112, 2014

V^0

RNN-UB

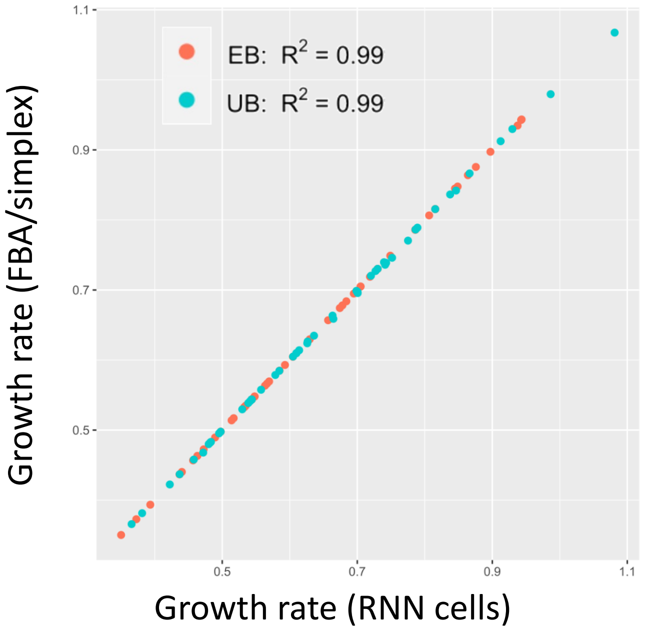
$$R = \text{ReLU}\,(M + S_{ext}V\; - b)$$
$$V = V + \delta t\,(\,P(c + S_{ext}^T R)\; - QV\,)$$
$$M = M + \frac{1}{2}\,\delta t\,(R - M)$$

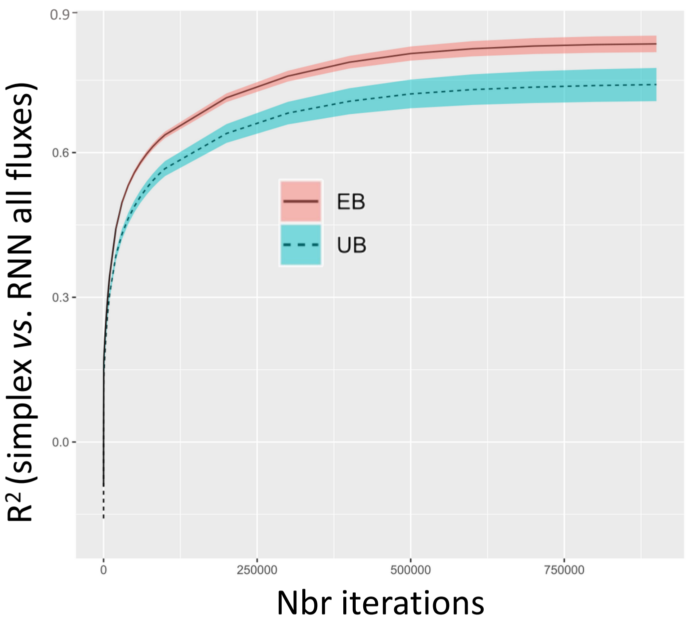
V, M

V^n

$$Q = S_{int}^T(S_{int}S_{int}^T)^{-1}$$
$$P = Q\;S_{int}\text{ - }I$$

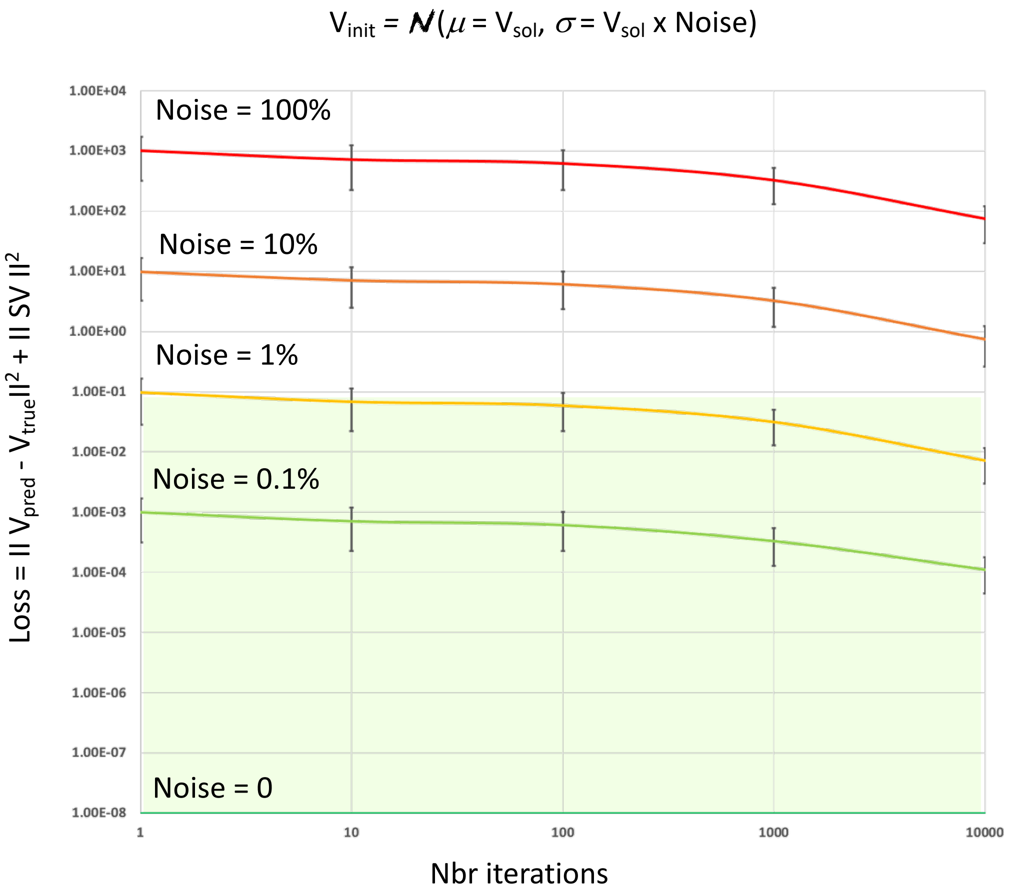
100 different growth rate generated by CobraPy for *E. coli core* model varying the concentrations of 13 different medium metabolites (mostly sugars)

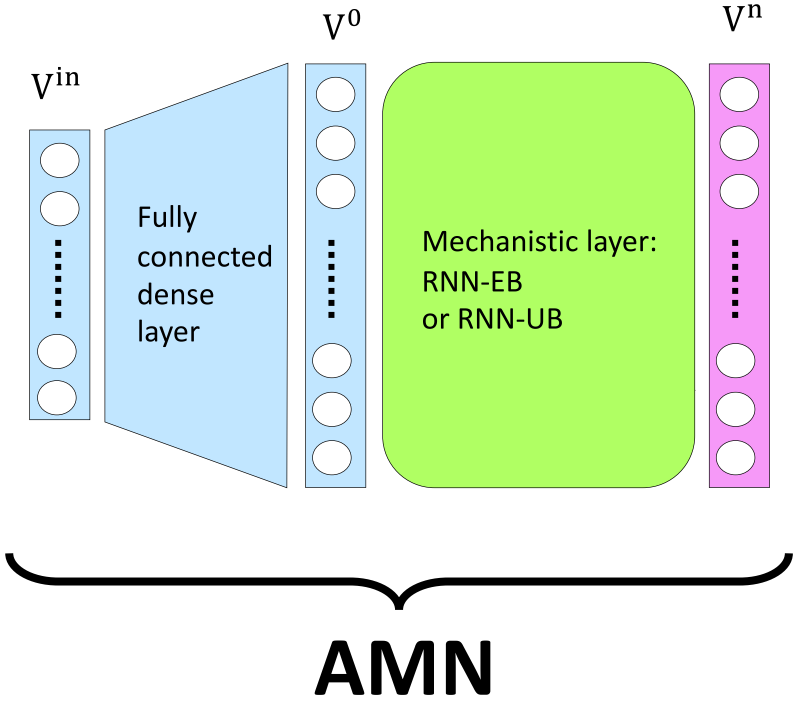




Encoding metabolic network into neural network: Artificial Metabolic Network (AMN)

RNN cells loss get to acceptable values quickly when initial vector is close to optimum





Training AMN with simulated data

Training set generated by COBRApy on *E. coli core* and *iML1514* models. ANN and AMN are fitted for growth rate

SMBL strain	Size Avr. nbr. variable met. drawn	NN type Architecture Timestep	Nbr Param Nbr epoch	R^2 Train	Q^2 (5-fold validation set) IISVII^2
Ecoli_core	10000 6 [1-13]	ANN Dense n/a	3.61k 500	0.96 ± 0.02	0.95 ± 0.02 n/a
Ecoli_core	10000 6 [1-13]	AMN RNN-EB 5	4.74k 500	1.00 ± 0.00	1.00 ± 0.00 0.03 ± 0.02
iML1515	10000 5 [1-27]	AMN RNN-EB 5	289.43k 500	1.00 ± 0.00	1.00 ± 0.00 0.06 ± 0.02

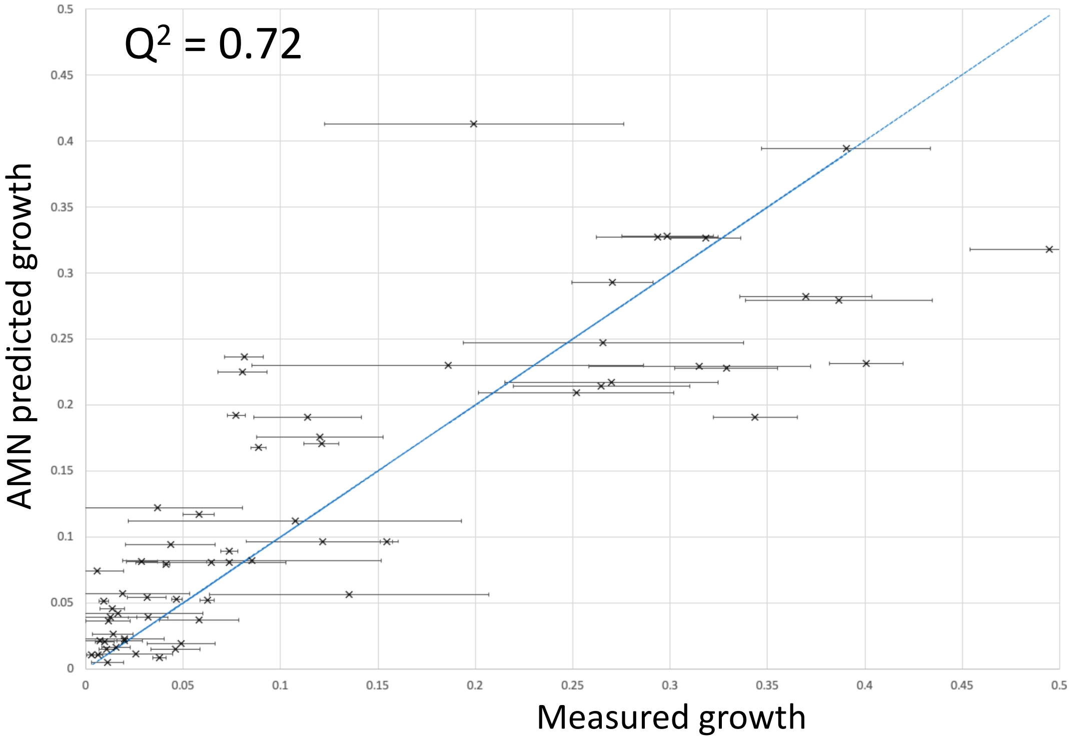
Training AMN with experimental data

73 different media composed of M9+glycerol and 27 extra variable metabolites. 1 to 5 variable metabolites are randomly drawn in each medium. All variable metabolites are added with the same concentration

EX_his__L_e_i	EX_leu__L_e_i	EX_phe__L_e_i	EX_lys__L_e_i	EX_tyr__L_e_i
EX_val__L_e_i	EX_asp__L_e_i	EX_met__L_e_i	EX_ile__L_e_i	EX_gln__L_e_i
EX_cys__L_e_i	EX_ser__L_e_i	EX_ala__L_e_i	EX_trp__L_e_i	EX_asn__L_e_i
EX_pro__L_e_i	EX_thr__L_e_i	EX_gly_e_i	EX_arg__L_e_i	EX_glu__L_e_i
EX_melib_e_i	EX_sucr_e_i	EX_tre_e_i	EX_lcts_e_i	EX_ac_e_i
EX_pyr_e_i	EX_cit_e_i			

SMBL strain	Size Method	NN type Architecture	Nbr Param Nbr epoch	R^2 Train	Q^2 (10-fold)
iML1515	73 EXP	ANN Dense	8.11k 1000	0.98	0.71
iML1515	73 EXP	AMN RNN-UB	289.43k 1000	O.89	0.72

CV predictions with AMN

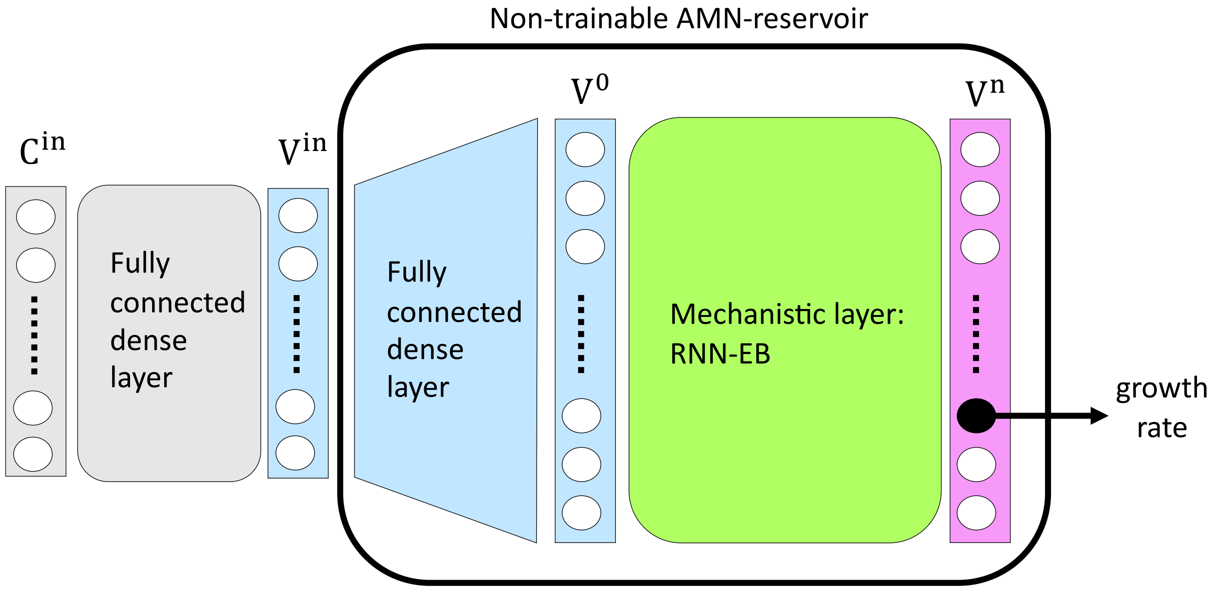


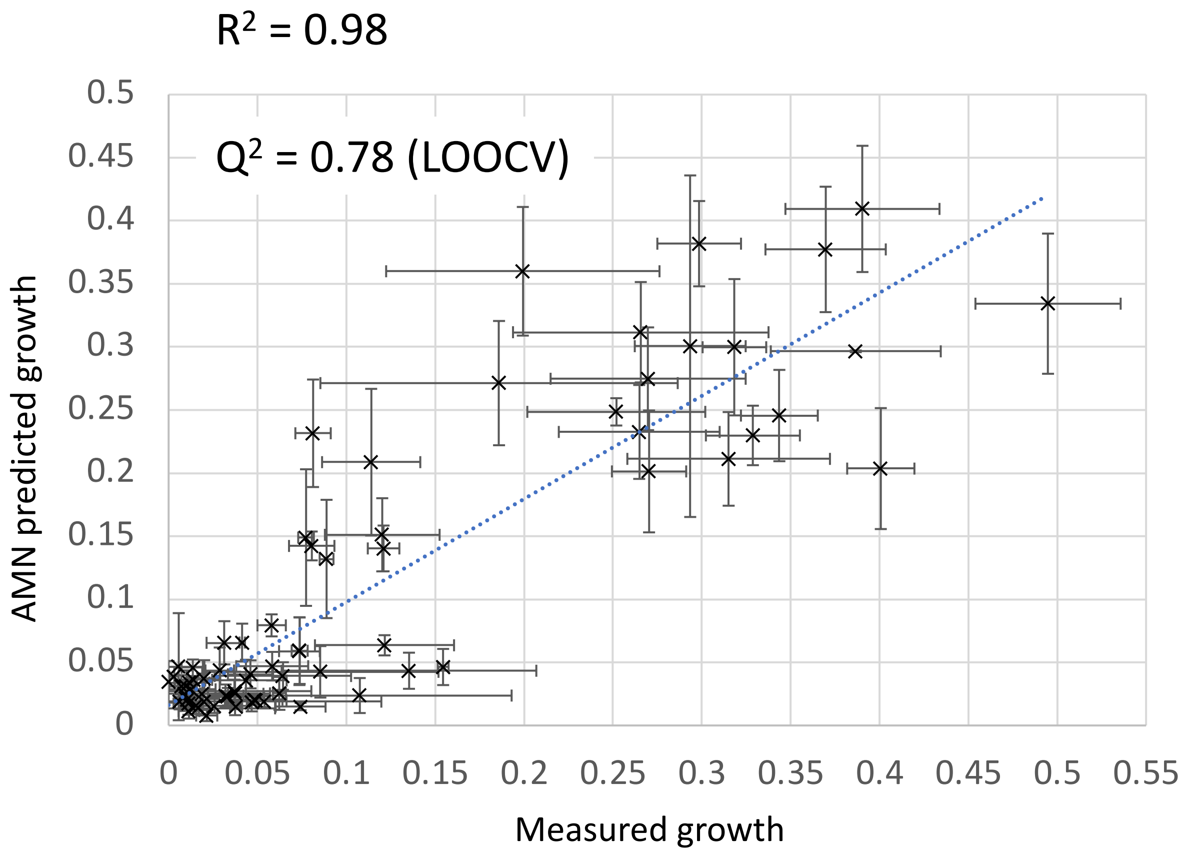
Predicting medium intake fluxes with AMN

73 different media composed of M9+glycerol and 27 extra variable metabolites. 1 to 5 variable metabolites are randomly drawn in each medium. All variable metabolites are added with the same concentration

EX_his__L_e_i	EX_leu__L_e_i	EX_phe__L_e_i	EX_lys__L_e_i	EX_tyr__L_e_i
EX_val__L_e_i	EX_asp__L_e_i	EX_met__L_e_i	EX_ile__L_e_i	EX_gln__L_e_i
EX_cys__L_e_i	EX_ser__L_e_i	EX_ala__L_e_i	EX_trp__L_e_i	EX_asn__L_e_i
EX_pro__L_e_i	EX_thr__L_e_i	EX_gly_e_i	EX_arg__L_e_i	EX_glu__L_e_i
EX_melib_e_i	EX_sucr_e_i	EX_tre_e_i	EX_lcts_e_i	EX_ac_e_i
EX_pyr_e_i	EX_cit_e_i			

SMBL strain	**R^2 (5-fold)**	**Q^2 (5-fold) IISVII²**
iML1515	1.00 ± 0.00	1.00 ± 0.00 0.06 ± 0.02



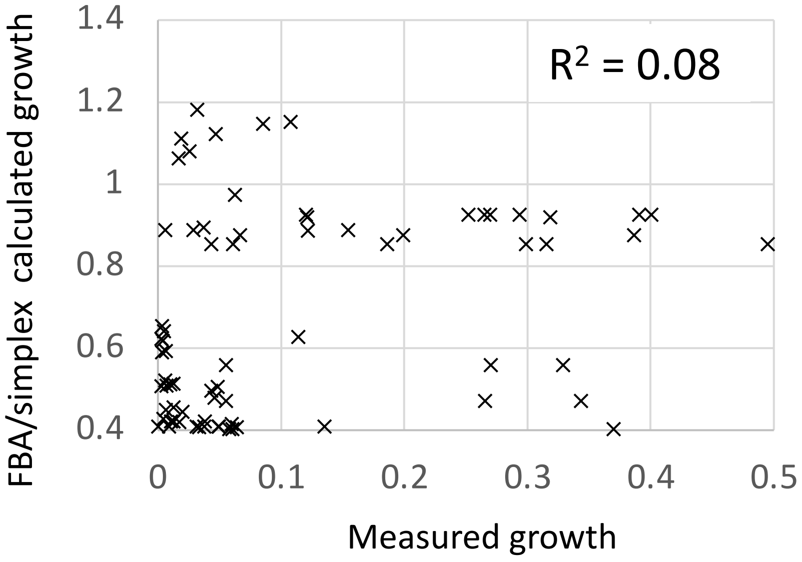


Matching FBA/simplex calculations with experimental data

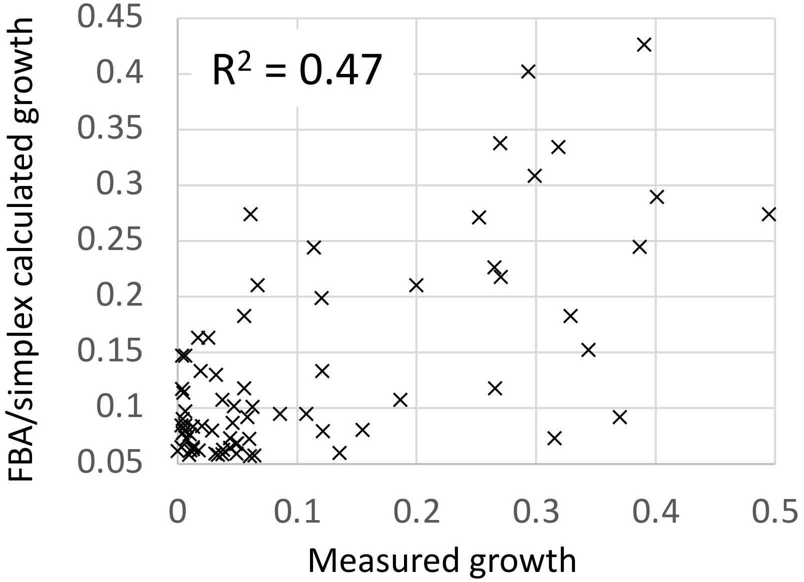
73 different media composed of M9+glycerol and 27 extra variable metabolites. 1 to 5 variable metabolites are randomly drawn in each medium. All variable metabolites are added with the same concentration

EX_his__L_e_i	EX_leu__L_e_i	EX_phe__L_e_i	EX_lys__L_e_i	EX_tyr__L_e_i
EX_val__L_e_i	EX_asp__L_e_i	EX_met__L_e_i	EX_ile__L_e_i	EX_gln__L_e_i
EX_cys__L_e_i	EX_ser__L_e_i	EX_ala__L_e_i	EX_trp__L_e_i	EX_asn__L_e_i
EX_pro__L_e_i	EX_thr__L_e_i	EX_gly_e_i	EX_arg__L_e_i	EX_glu__L_e_i
EX_melib_e_i	EX_sucr_e_i	EX_tre_e_i	EX_lcts_e_i	EX_ac_e_i
EX_pyr_e_i	EX_cit_e_i			

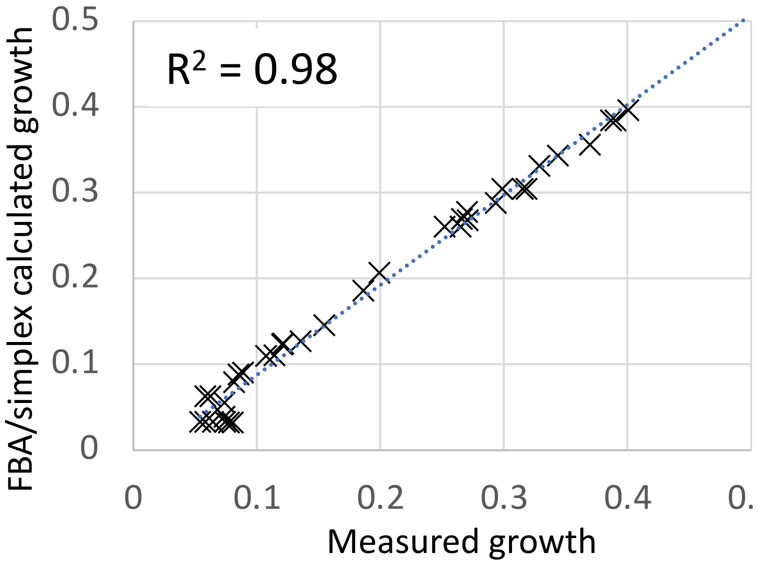
Upper bound = 10 when metabolite is present in medium, 0 otherwise



Upper bound optimized to minimize MSE between calculated and measured growth



Extract bound = calculated by ANN+AMN-reservoir

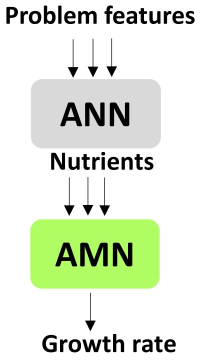


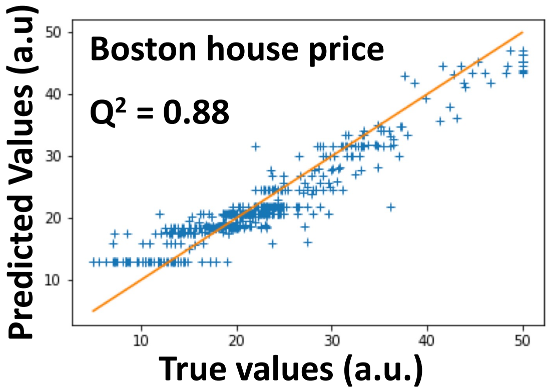
Acknowledgement

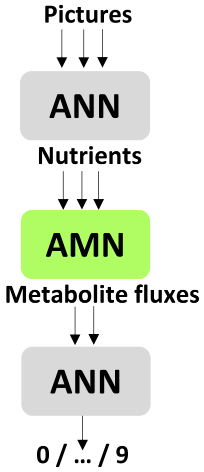
INRAE

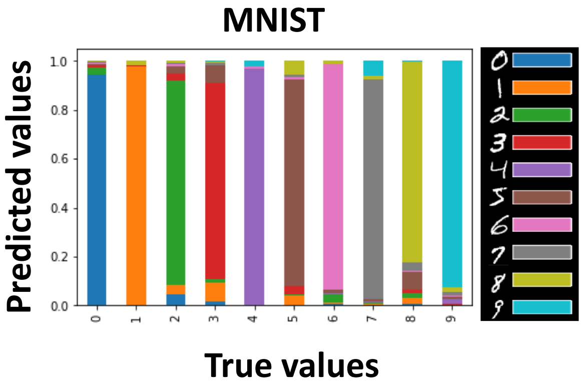
anr agence nationale de la recherche

ANR-21-CE45-0021









Léon Faure (U. Paris Saclay)

Bastien Mollet (ENS)

Aymeric Gaudin (CentraleSupelec)

Wolfram Leibermeister (INRAe)