Modelling cellular systems

Metabolism and genetic circuits

Paula Jouhten

25.4.2023

Aalto University
School of Chemical
Engineering

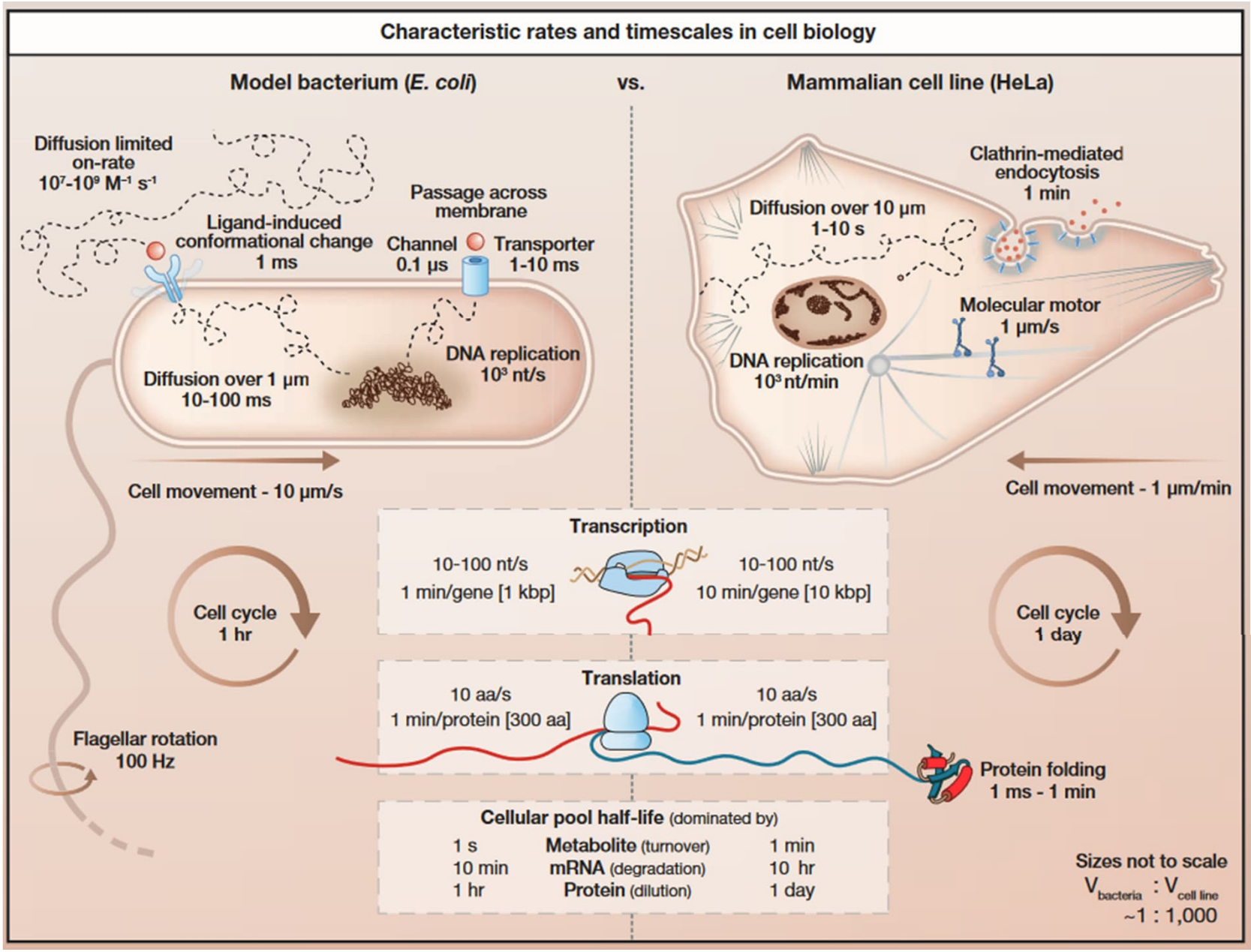
Intended learning objectives

To be able to

Describe the design processes used for genetic circuits and metabolic pathways.

Discuss the use computational design methods in applications

How fast are cellular processes ?



DOI:
10.1016/j.cell.2016.02.058

Aalto University
School of Chemical Engineering

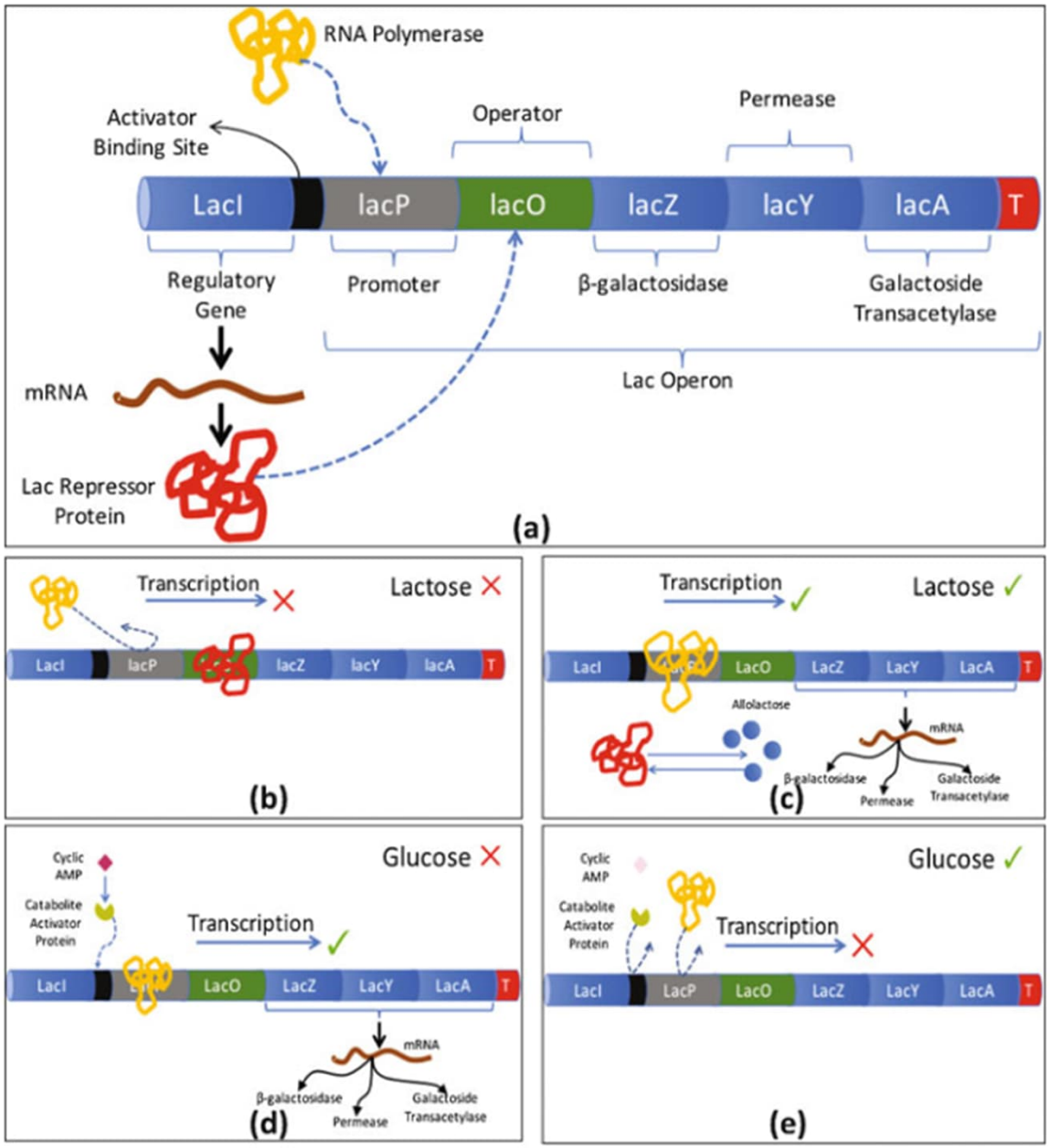
Genetic circuit

An assembly of biological parts encoded in the genome that enable cells to respond and perform functions

The functions are realized through the **Central Dogma** of molecular biology:

gene -> mRNA -> protein

Lac operon



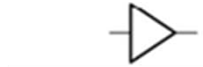
https://link-springer-com.libproxy.aalto.fi/book/10.1007/978-3-030-52355-8

Truth table and circuit diagram

Inputs		Output
Lactose	Glucose	
1	1	0
1	0	1
0	1	0
0	0	0

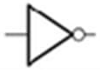
?

YES



INPUT	OUTPUT
A	
0	0
1	1

NOT



INPUT	OUTPUT
A	
0	1
1	0

AND

INPUT		OUTPUT
A	B	
0	0	0
1	0	0
0	1	0
1	1	1

OR

INPUT		OUTPUT
A	B	
0	0	0
1	0	1
0	1	1
1	1	1

XOR

INPUT		OUTPUT
A	B	
0	0	0
1	0	1
0	1	1
1	1	0

NAND

INPUT		OUTPUT
A	B	
0	0	1
1	0	1
0	1	1
1	1	0

NOR

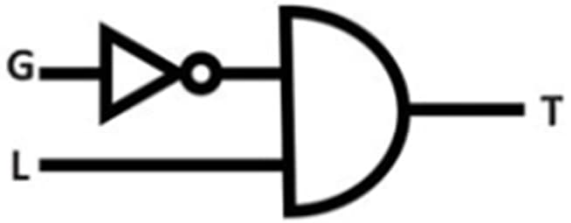
INPUT		OUTPUT
A	B	
0	0	1
1	0	0
0	1	0
1	1	0

XNOR

INPUT		OUTPUT
A	B	
0	0	1
1	0	0
0	1	0
1	1	1

https://doi.org/10.1515/rnan-2015-0003

Inputs		Output
G	L	T
0	0	0
0	1	1
1	0	0
1	1	0



Synthetic genetic circuits

- **From sensory information to biological functions**
- **If a truth table is to be converted to DNA sequence,**

What information/data/parts are needed?

Promoters, gates available, sensors, sequences,

What may be the challenges involved?

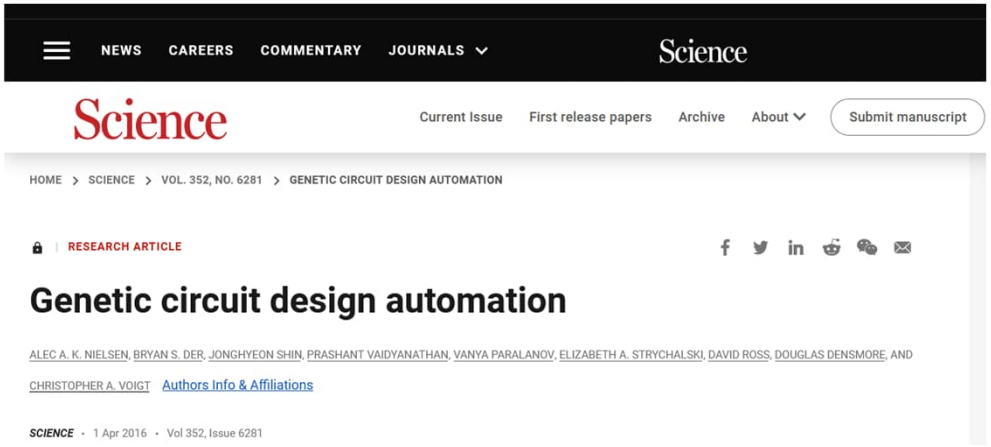
Truth table to circuit with the available parts

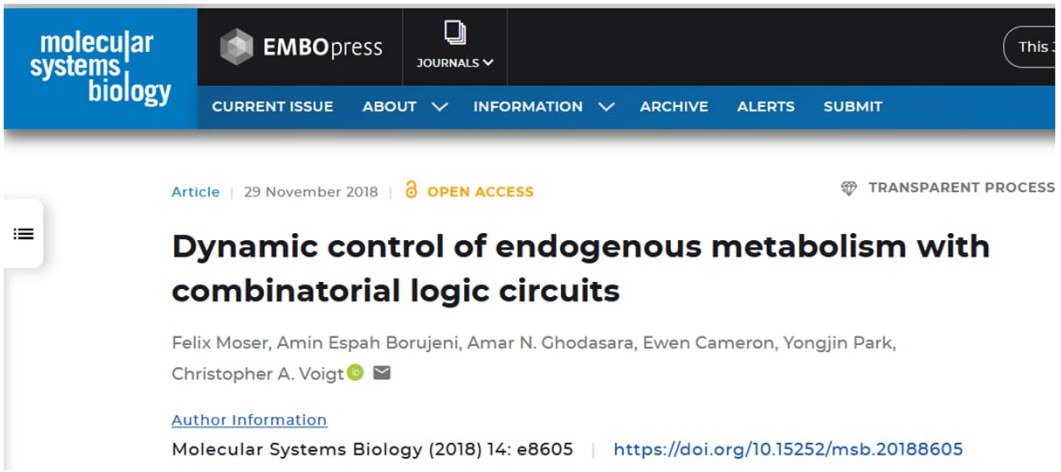
Aalto University
School of Chemical
Engineering

Reading material

Steady state modelling: Nielsen et al. (2016) Genetic circuit design automation. Science. 352:aac7341. doi: 10.1126/science.aac7341. PDF to be provided in MyCourses

Dynamic modelling: Moser et al. (2018) Dynamic control of endogenous metabolism with combinatorial logic circuits. Mol Syst Biol. 14:e8605. https://www.embopress.org/doi/full/10.15252/msb.20188605





Challenges in circuit design

- Regulator expression need to be precisely balanced for correct function
- Function of the parts can vary depending on genetic context, strain, and growth conditions
- States of circuits (their response to different inputs) can be cumbersome to characterize
- Many regulators are toxic when overexpressed, and even mild effects can combine to drive negative selection against the circuit

doi: 10.1126/science.aac7341; pdf in MyCourses

Automated circuit design input

From operation description to DNA sequence

1. Description of the desired operation
2. DNA sequences of the parts (e.g., sensors, gates)
3. Data for the sensors (e.g., ON/OFF signal strengths)
4. Data for the gate library (e.g., response functions)
5. Conditions of validity: genetic system layout, strain, operating conditions

Sensors + Simple gates -> Output promoter to control the target function

Aalto University
School of Chemical Engineering

doi: 10.1126/science.aac7341; pdf in MyCourses

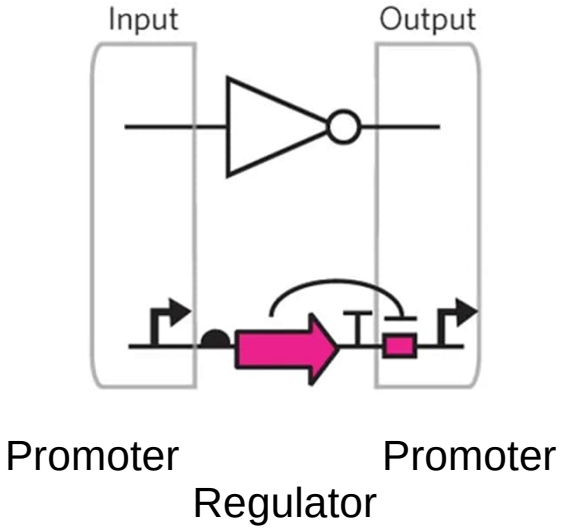
Common signal carrier for modularization

RNA polymerase flux on DNA

Output of a gate as input for next

Regulators could be transcription factors but also others like e.g., CRISPR/Cas-based regulation

NOT-gate

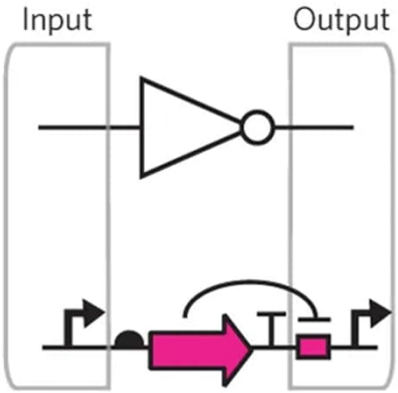


Aalto University
School of Chemical
Engineering

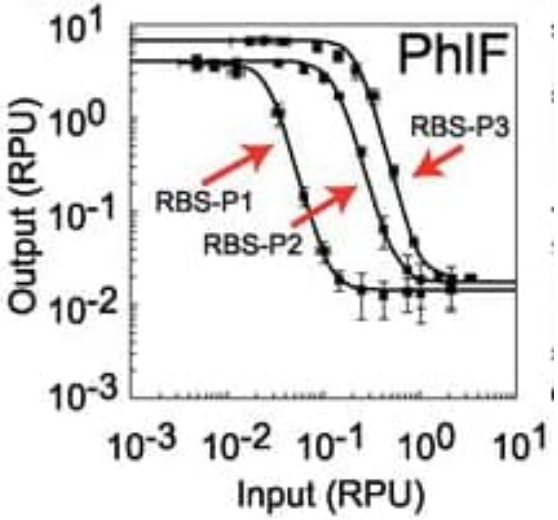
doi: 10.1126/science.aac7341; pdf in MyCourses

Response function

NOT-gate



Response function



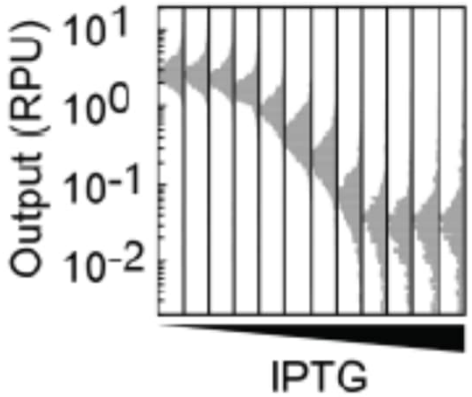
RPU (relative promoter unit)
RBS (ribosome binding site)

doi: 10.1126/science.aac7341; pdf in MyCourses

Determining a gate response function

- Standard promoter: *E. coli* BBa_J23101 constitutive promoter, output of 1 RPU
- Fluorescence measured under a range of inducer concentrations from strains in which
  1. Fluorescence protein is expressed from the standard promoter $\langle YFP \rangle_{RPU}$
  2. Autofluorescence control without fluorescence protein $\langle YFP \rangle_{0}$
  3. Fluorescence protein is expressed from the input promoter
  4. Gate controls fluorescence protein

Example for strain 4, IPTG as inducer



doi: 10.1126/science.aac7341; pdf in MyCourses

Determining a gate response function

- Convert the fluorescence readouts to RPUs for both
  1. Fluorescence protein is expressed from the input promoter
  2. Gate controls fluorescence protein

$$RPU_{input} = \frac{\langle YFP \rangle_{input} - \langle YFP \rangle_0}{\langle YFP \rangle_{RPU} - \langle YFP \rangle_0}$$

$$RPU_{gate} = \frac{\langle YFP \rangle_{gate} - \langle YFP \rangle_0}{\langle YFP \rangle_{RPU} - \langle YFP \rangle_0}$$

- Plot output as a function of input at each concentration of inducer
- Fit Hill function to the response curve

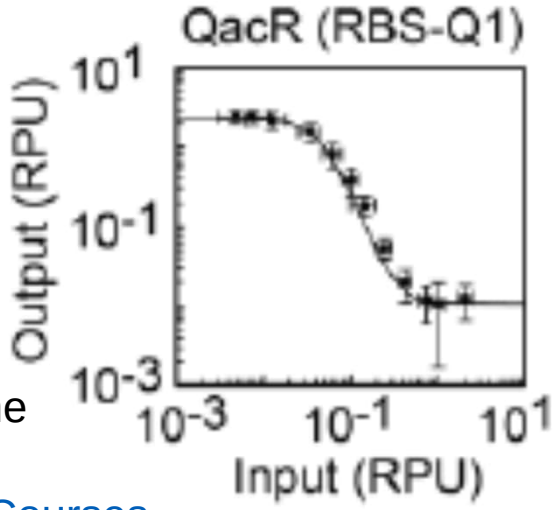
$$y = y_{min} + (y_{max} - y_{min}) \frac{K^n}{K^n + x^n}$$

where

n is the Hill coefficient

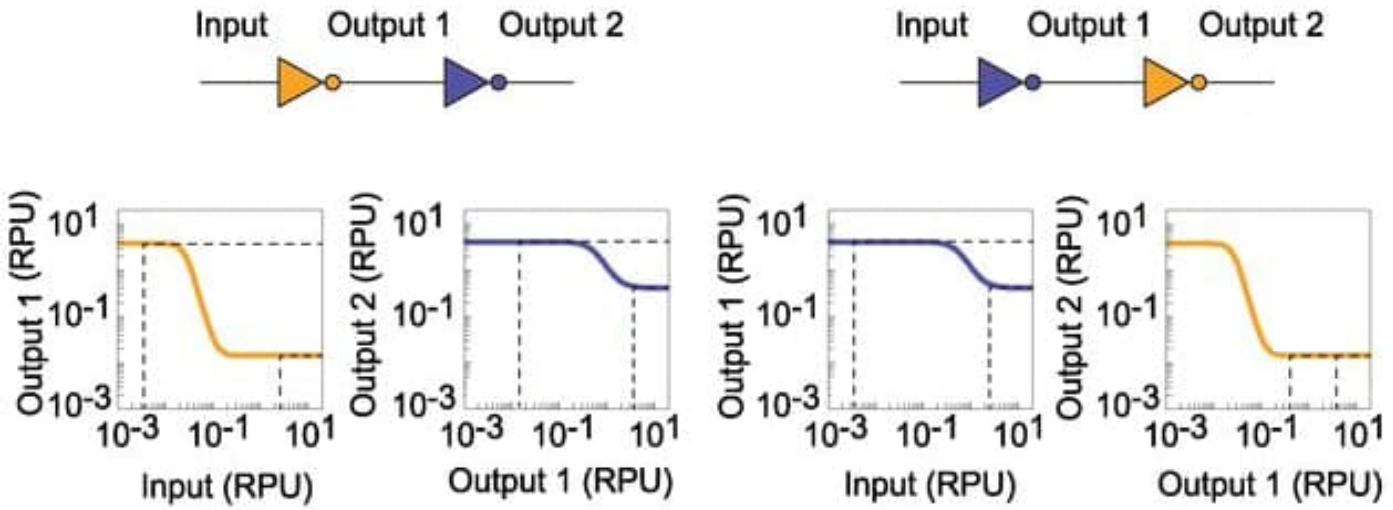
K is the threshold input level where the output is half maximum

y_{min} and y_{max} are the minimum and maximum output values from the gate



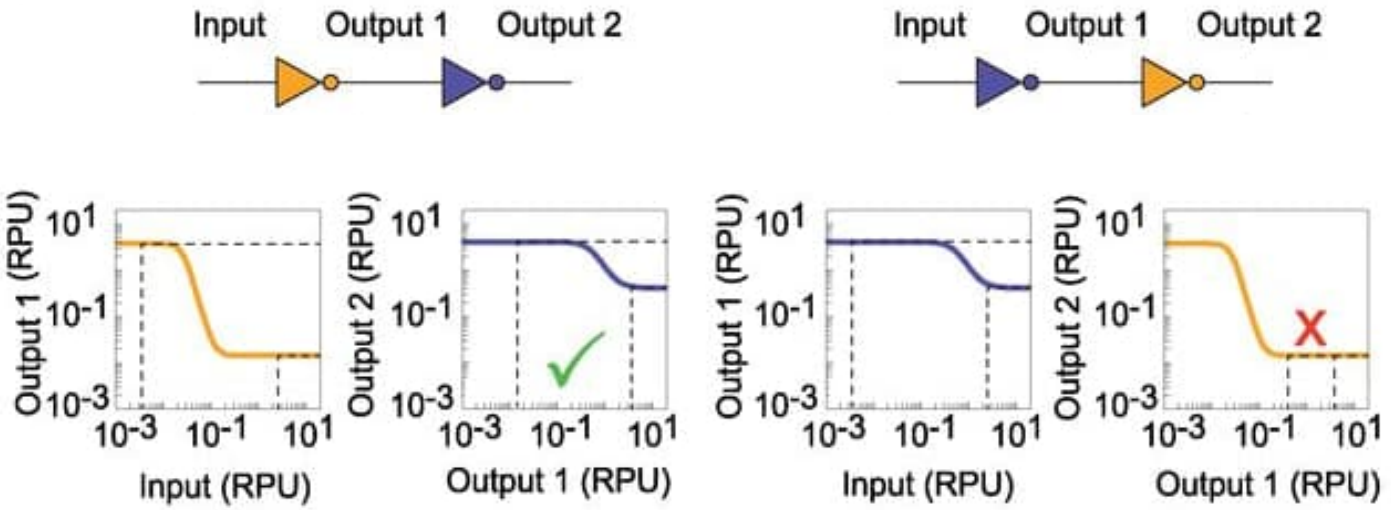
doi: 10.1126/science.aac7341; pdf in MyCourses

Which combination makes a functional circuit?



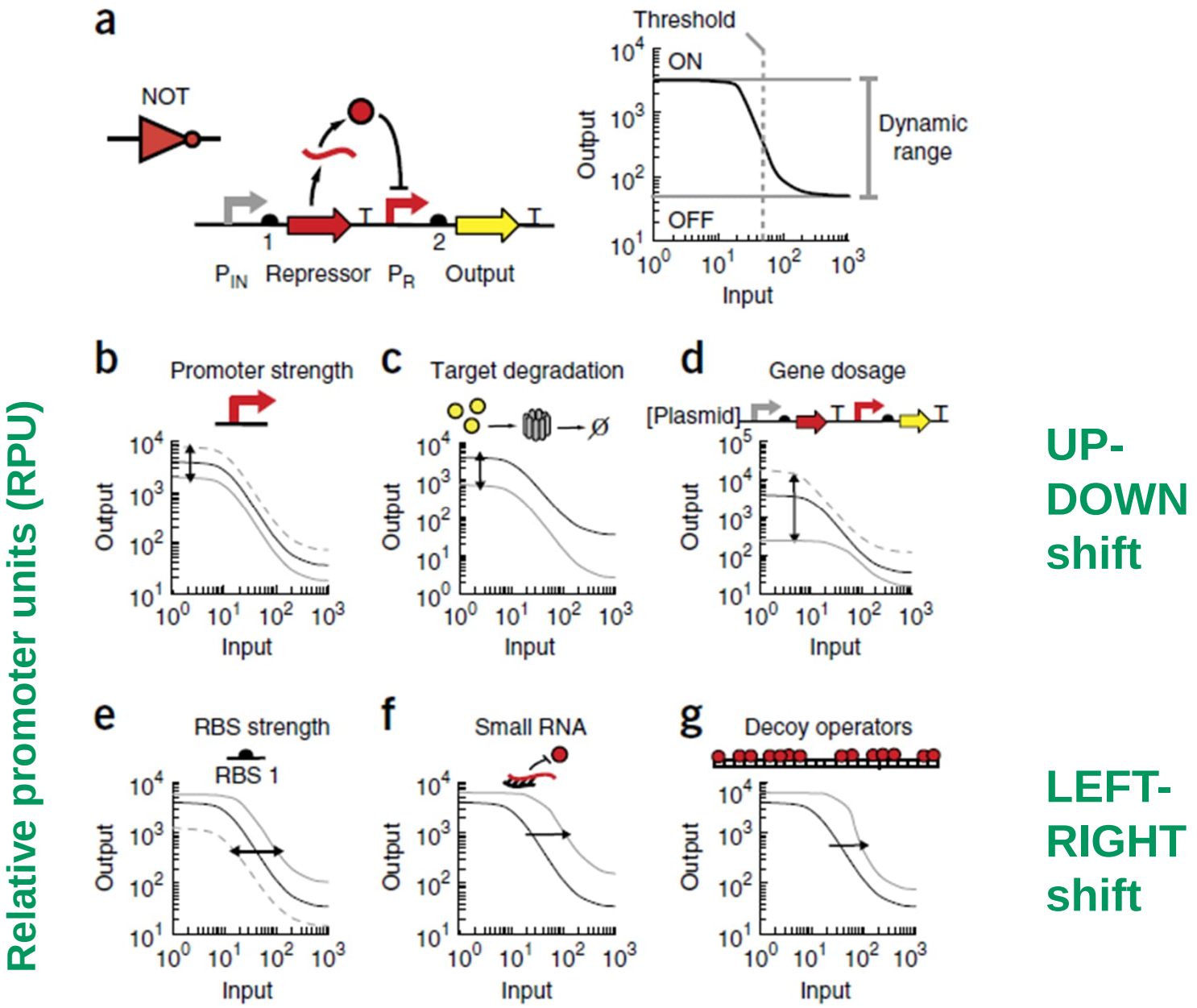
doi: 10.1126/science.aac7341; pdf in MyCourses

Response functions are essential for combining gates



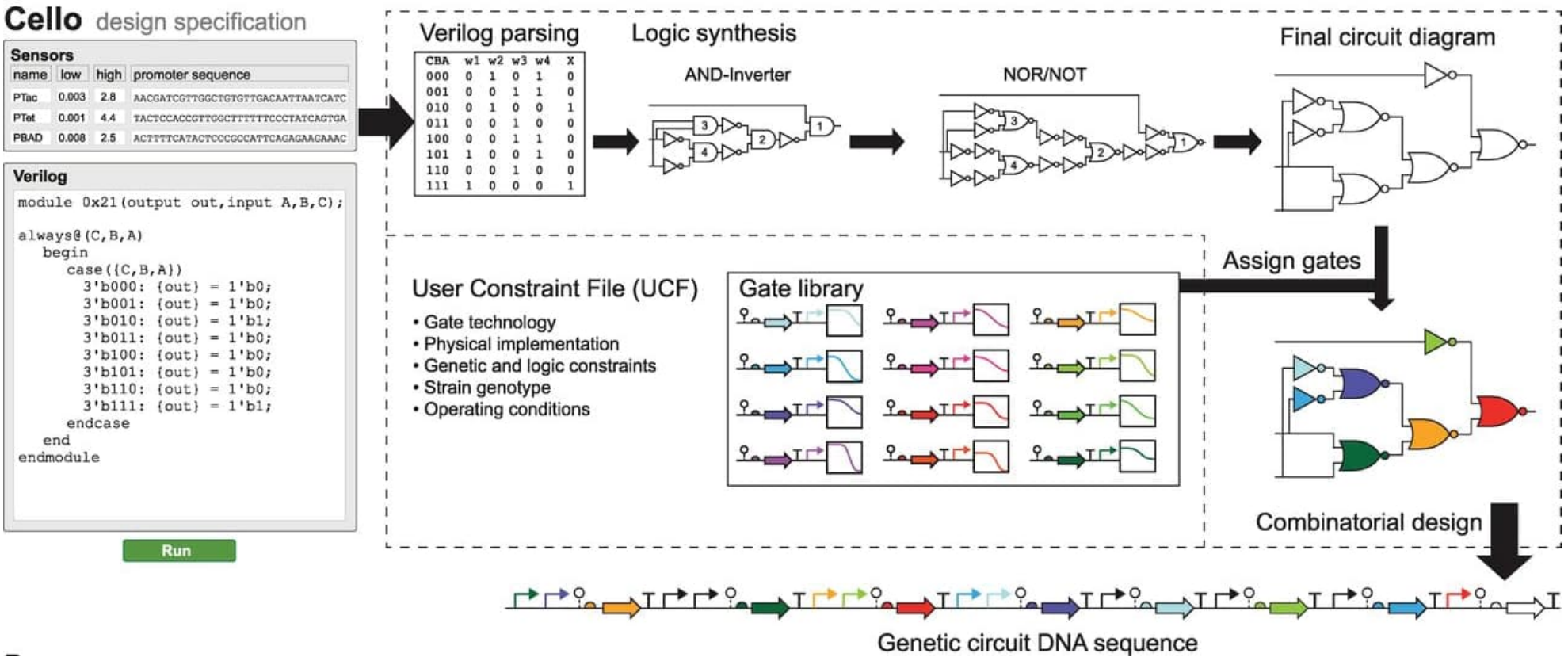
doi: 10.1126/science.aac7341; pdf in MyCourses

How to modulate the response function?



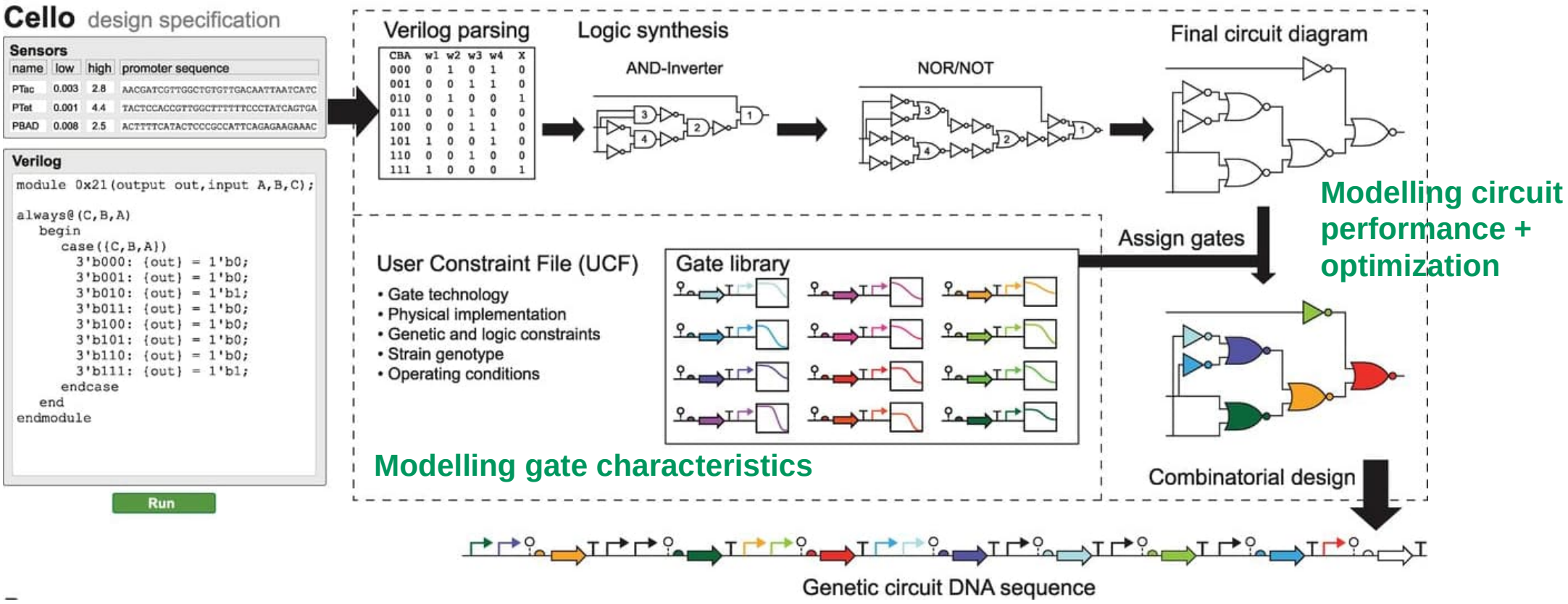
Brophy and Voigt, 2014; https://www.nature.com/articles/nmeth.2926

Genetic circuit design automation



doi: 10.1126/science.aac7341; pdf in MyCourses

Genetic circuit design automation



doi: 10.1126/science.aac7341; pdf in MyCourses

Gate assignment is an optimization problem

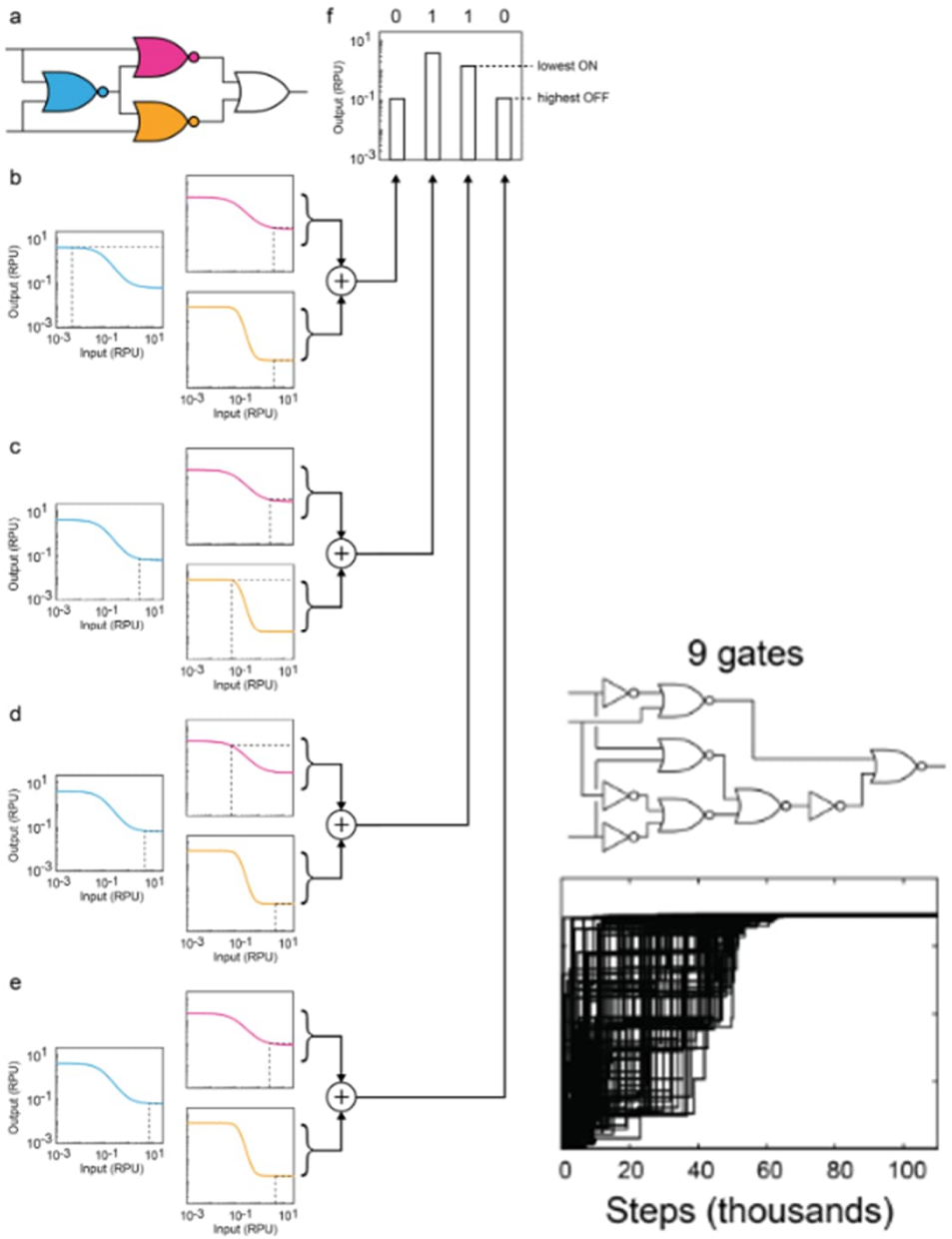
Gate assignments are scored:

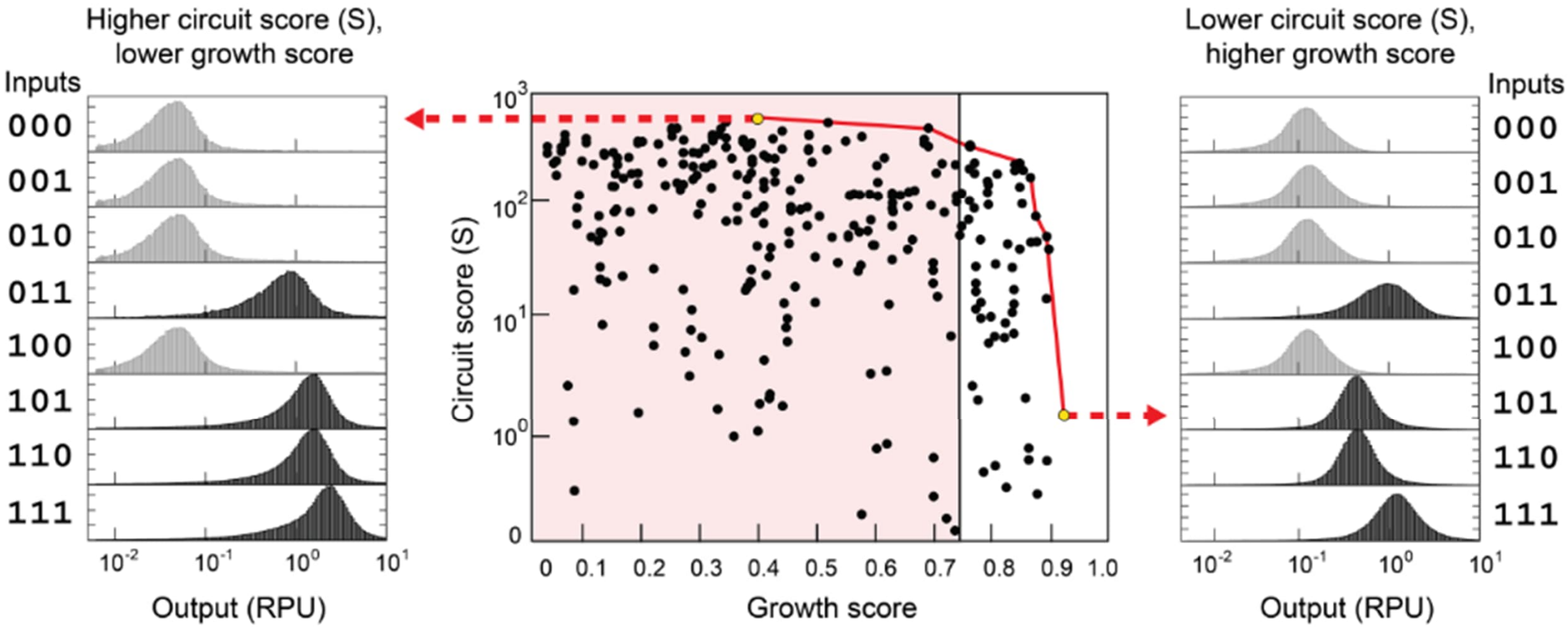
$$S = \frac{\min(ON)}{\max(OFF)}$$

Monte Carlo simulated annealing algorithm is used for optimizing the gate assignment
A swap of two gates, then calculation of S'

$$P = e^{-((S-S')/T)}$$

doi: 10.1126/science.aac7341; pdf in MyCourses



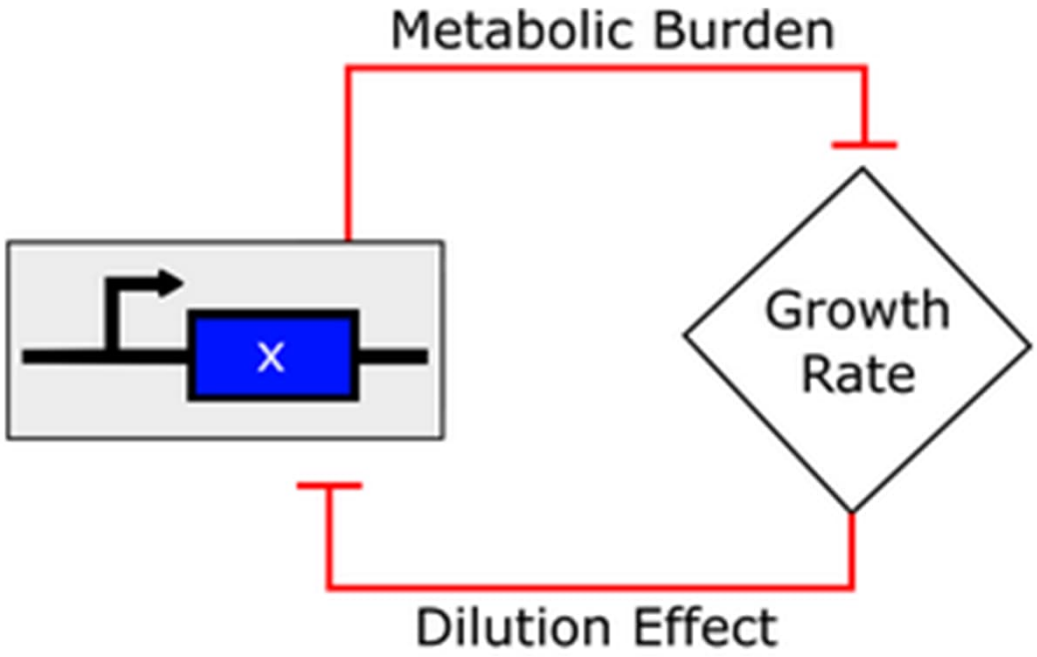


The toxicity of the whole circuit for a particular input combination is calculated as the product of normalized cell growth for each of the individual gates.

After the toxicities of all the input states are calculated, the toxicity of the circuit as a whole ("growth score") is taken as the worst input state.

doi: 10.1126/science.aac7341; pdf in MyCourses

Double-negative feedback loops in host-circuit systems



Aalto University
School of Chemical Engineering

https://doi.org/10.1371/journal.pcbi.1010518

From steady state models to dynamics

If circuit input is not switch like but dynamic, dynamic modelling is useful for *in silico* screening of circuit designs

Gate response functions vs ODEs (in bold)

NOT

$$y = y_{min} + (y_{max} - y_{min})\frac{K^n}{K^n + x^n}$$

$$\boldsymbol{\frac{dy}{dt} = \alpha(y_{max} - y_{min})\frac{K^n}{K^n + x(t)^n} - \gamma(y(t) - y_{min})}$$

AND

$$y = y_{min} + (y_{max} - y_{min})\frac{x_1 {x_2}^2}{K + x_1 {x_2}^2}$$

$$\boldsymbol{\frac{dy}{dt} = \alpha(y_{max} - y_{min})\frac{x_1(t)x_2(t)^2}{K + x_1(t)x_2(t)^2} - \gamma(y(t) - y_{min})}$$

ANDN

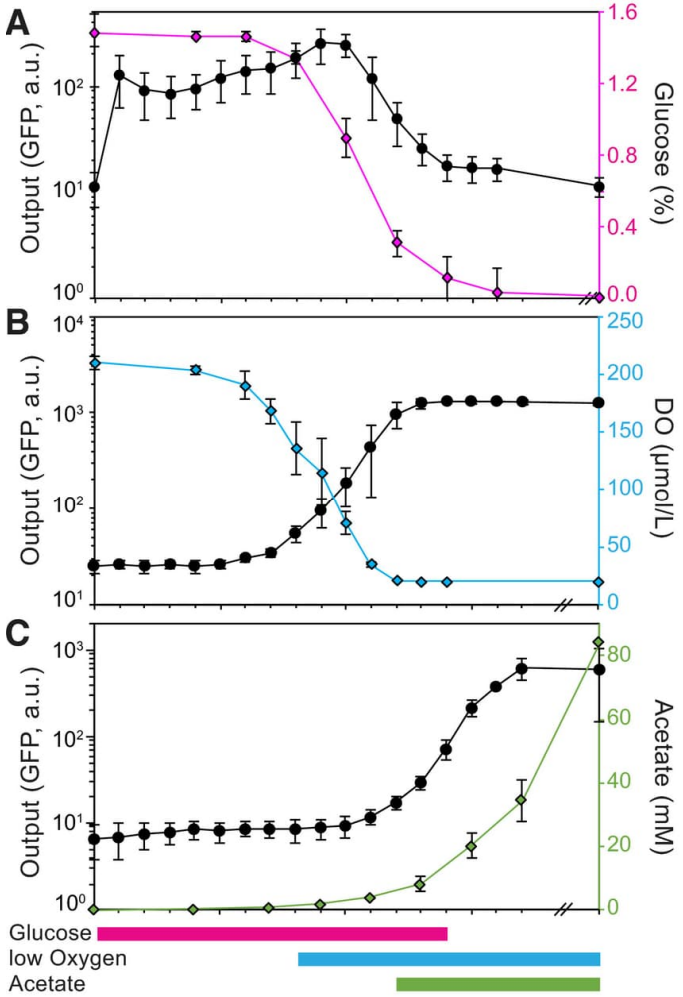
$$y = y_{min} + (x_1 - y_{min})\frac{K}{K + x_2}$$

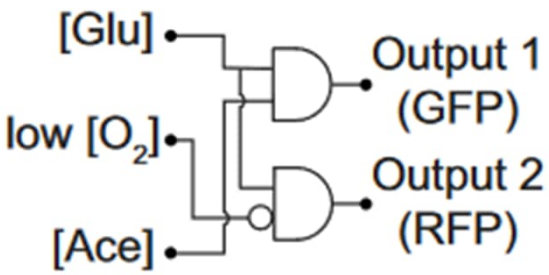
$$\boldsymbol{\frac{dy}{dt} = \alpha(x_1(t) - y_{min})\frac{K}{K + x_2(t)} - \gamma(y(t) - y_{min})}$$

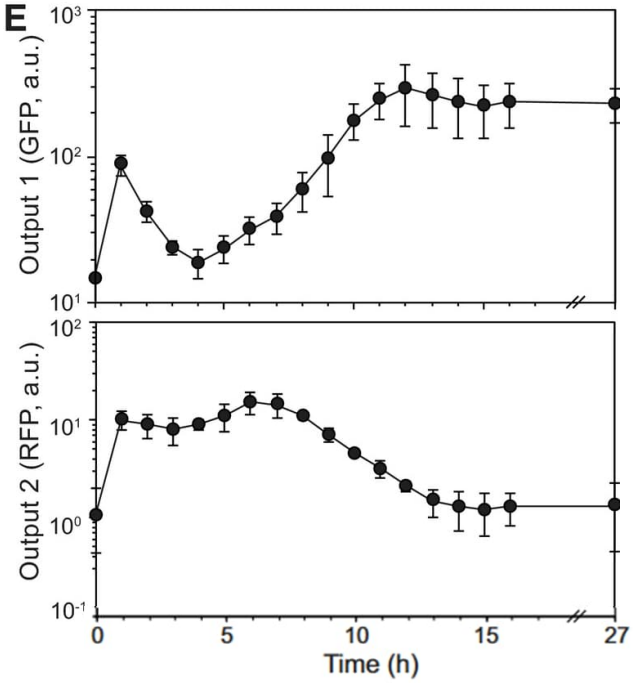
Parameters from the response function
Rate constants α and γ of turning the gate ON and OFF, respectively, 1/h (Tabor et al. 2009; Moon et al. 2012)

Adopted from: Moser et al. (2018) Mol Syst Biol. 14:e8605. doi: 10.15252/msb.20188605.

Glucose, oxygen and acetate sensors' controlled circuit dynamics predicted for *E. coli* batch culture







- ODE system solved discretely
- In each time step, the corresponding empirical values for the output activity of glucose, oxygen, and acetate sensors were assigned to the inputs

Adopted from: Moser et al. (2018) Mol Syst Biol. 14:e8605. doi: 10.15252/msb.20188605.

Metabolic modelling

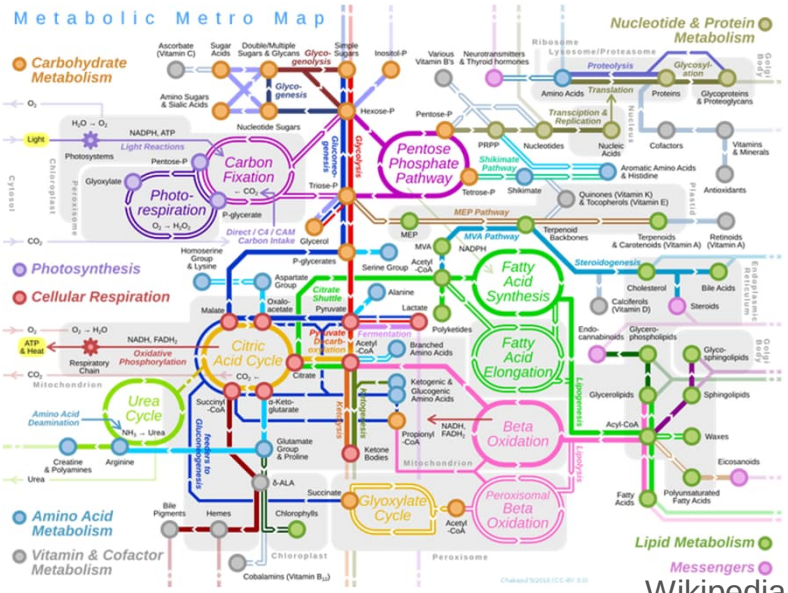
Orth JD, Thiele I, Palsson BO (2010) What is flux balance analysis? Nat Biotechnol. 28:245-8. doi: 10.1038/nbt.1614.

Box 2 outdated, check instead for COBRA toolbox, **COBRApy**, COBRA.jl: http://opencobra.github.io/

Why is metabolism relevant for synthetic biology?

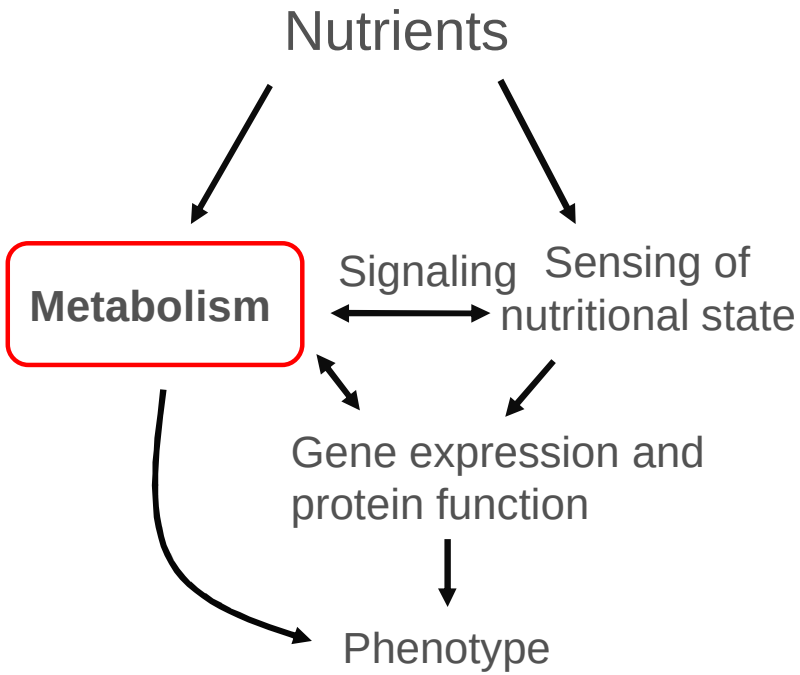
Metabolism = (bio)chemical reactions involved in sustaining a living state of cells and an organism

- Metabolism generates precursors for product compounds but also for circuit components
- Metabolism generates energy and redox power
- Metabolism is involved in cellular regulation



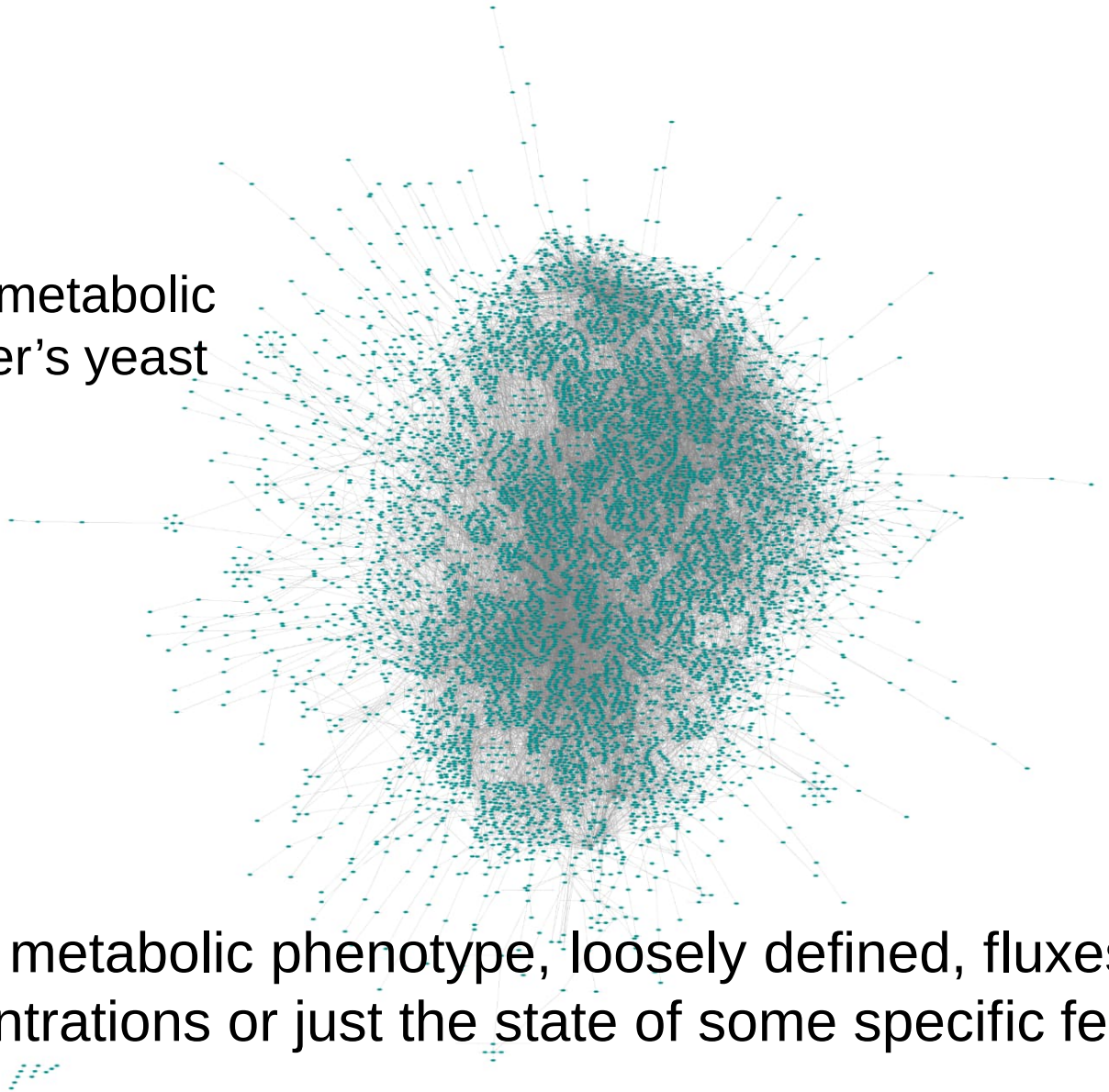
Wikipedia

Metabolism is involved in cellular regulation



Adopted from Jaakko Mattila

Genome-scale metabolic network of Baker’s yeast



Metabolic state = metabolic phenotype, loosely defined, fluxes and metabolite concentrations or just the state of some specific feature

Genome-scale metabolic model reconstruction

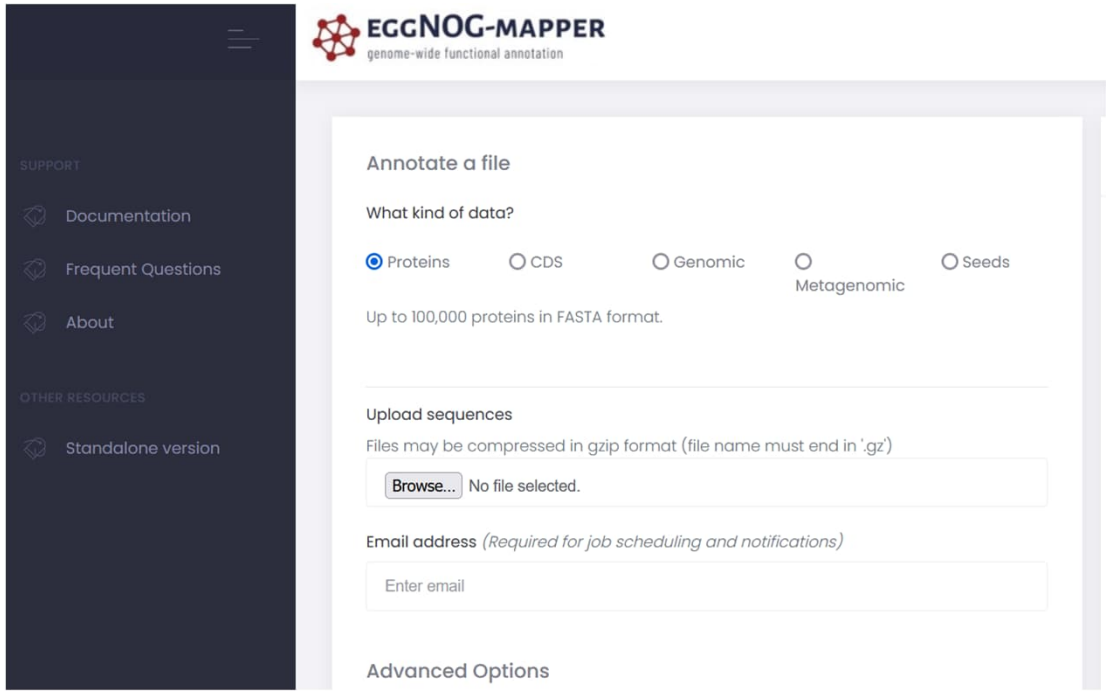
Which reactions can take place in the cells of a species?

Genome-scale metabolic model reconstruction

Which reactions can take place in the cells of a species?

1. Genome sequencing and assembly
2. Gene prediction
3. Protein functional annotation

   Using information from orthologous genes (i.e., diversified via speciation)

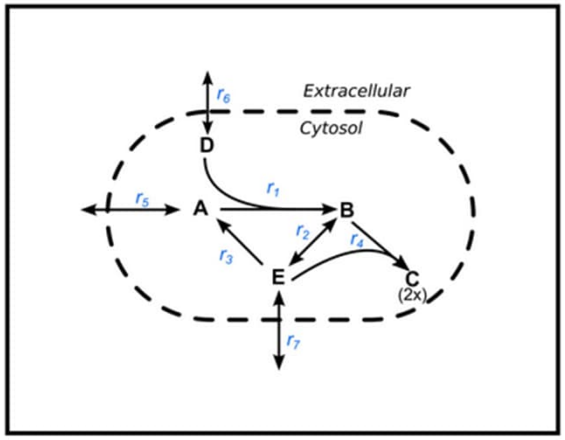


http://eggnog-mapper.embl.de/

Genome-scale metabolic model reconstruction

Conversion into mathematcal form

1. Unifying metabolite naming across reactions
2. Gathering reaction stoichiometries
3. Converting into matrix [metabolites x reactions]



r_1: A + D ⟶ B
r_2: 2B ⟷ 2E
r_3: E ⟶ A
r_4: B + E ⟶ 2C
r_5: ⟷ A
r_6: ⟷ D
r_7: ⟷ E
} transport reactions

$$\boldsymbol{S} = \begin{array}{c|ccccccc} & r_1 & r_2 & r_3 & r_4 & r_5 & r_6 & r_7 \\ \hline A & -1 & 0 & 1 & 0 & 1 & 0 & 0 \\ B & 1 & -2 & 0 & -1 & 0 & 0 & 0 \\ C & 0 & 0 & 0 & 2 & 0 & 0 & 0 \\ D & -1 & 0 & 0 & 0 & 0 & 1 & 0 \\ E & 0 & 2 & -1 & -1 & 0 & 0 & 1 \end{array}$$

$$\vec{v} = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \\ v_6 \\ v_7 \end{bmatrix}$$

doi.org/10.3389/fmicb.2016.00907

Genome-scale metabolic model reconstruction

Considering steady state operation

1. Considering steady state operation assumes constant metabolite concentrations and fluxes
2. System of linear equations with fluxes as variables is formed
3. The linear equations enforce **mass conservation** and flux bounds can be set to obey **thermodynamic feasibility**

= Constraints for metabolic states in specific cells formulated from first-principles

Dynamic situation

$$\frac{d\boldsymbol{c}}{dt} = \boldsymbol{S} \cdot \boldsymbol{v} = \boldsymbol{S} \cdot \boldsymbol{f}(\boldsymbol{e}(t), \boldsymbol{c}(t), \boldsymbol{z})$$

S stoichiometric matrix, **c** concentrations, **v** fluxes, **e** enzyme abundances, **z** parameters

Steady state

$$\boldsymbol{S} \cdot \boldsymbol{v} = \mathbf{0} = \begin{cases} \frac{dA}{dt} = -v_1 + v_3 + v_5 & 0 \le v_1 < \infty \\ \frac{dB}{dt} = v_1 - 2v_2 - v_4 & -\infty < v_2 < \infty \\ \frac{dC}{dt} = 2v_4 & 0 \le v_3 < \infty \\ \frac{dD}{dt} = -v_1 + v_6 & 0 \le v_4 < \infty \\ \frac{dE}{dt} = 2v_2 - v_3 - v_4 + v_7 & 0 \le v_5 \le \infty \\ & -\infty < v_6 < \infty \\ & 0 \le v_7 \le \infty \end{cases}$$

doi.org/10.3389/fmicb.2016.00907

Space of feasible metabolic states

Fluxes variables, metabolites not in the steady state system

Number of metabolites (equations) < number of fluxes (variables) = underdetermined system

Constraints:

1) $\boldsymbol{Sv} = 0$

2) $\boldsymbol{v,lb} < \boldsymbol{v} < \boldsymbol{v,ub}$

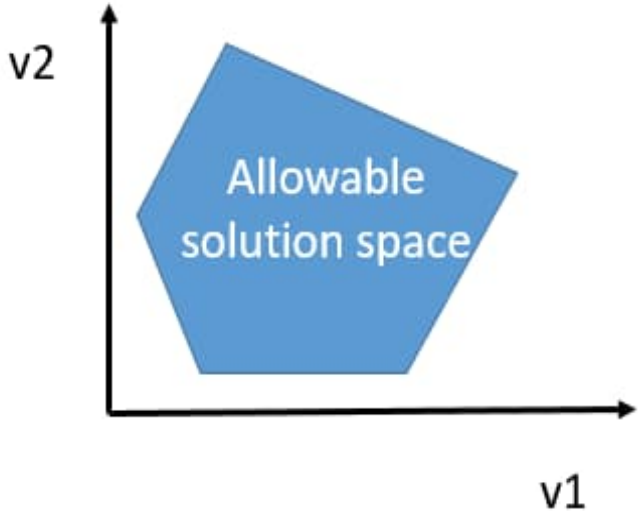
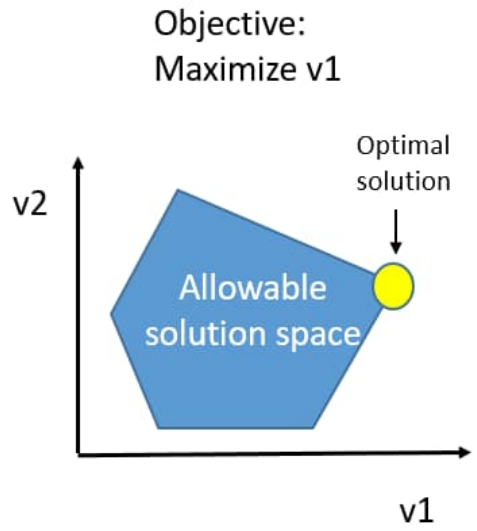


Figure modified by Tuula Tenkanen from O'Brien et al. 2015

Linear optimization to identify optimal states



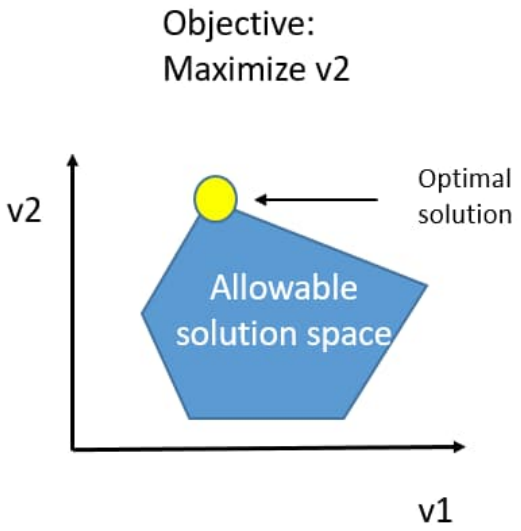


Figure modified by Tuula Tenkanen from O'Brien et al. 2015

Varma and Palsson, 1993; Varma and Palsson, 1994

Flux Balance Analysis (FBA)

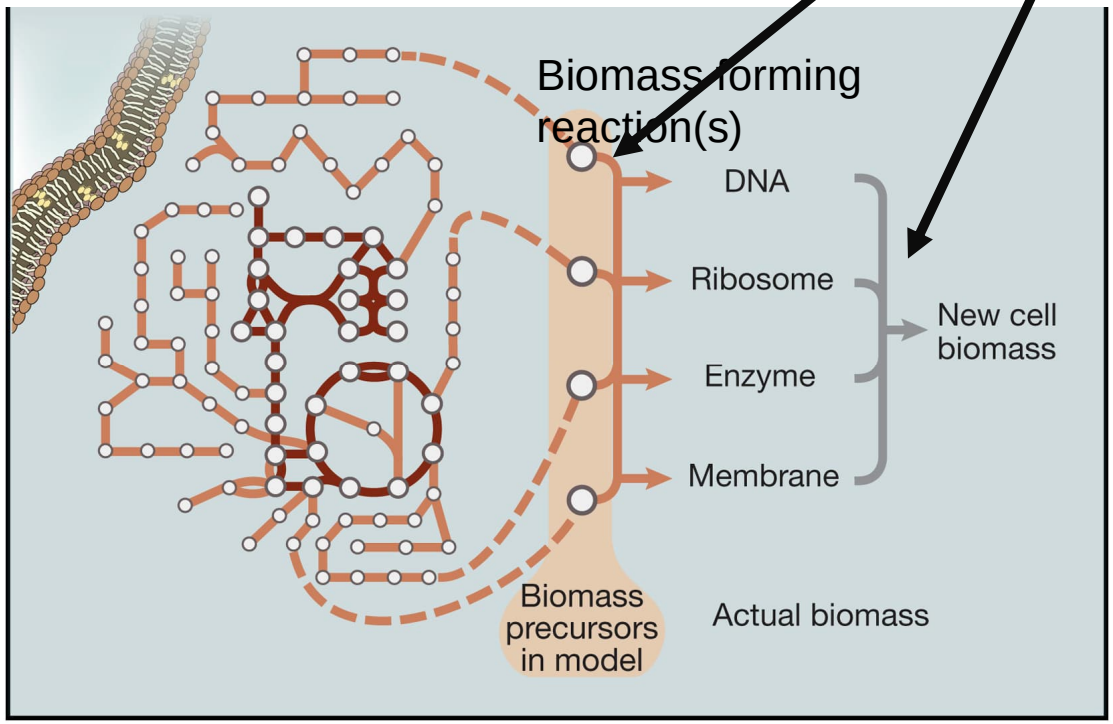
maximize (or minimize) $\boldsymbol{a}' \cdot \boldsymbol{v}$

subject to $\boldsymbol{S} \cdot \boldsymbol{v} = 0$

$\boldsymbol{v}_{lb} < \boldsymbol{v} < \boldsymbol{v}_{ub}$

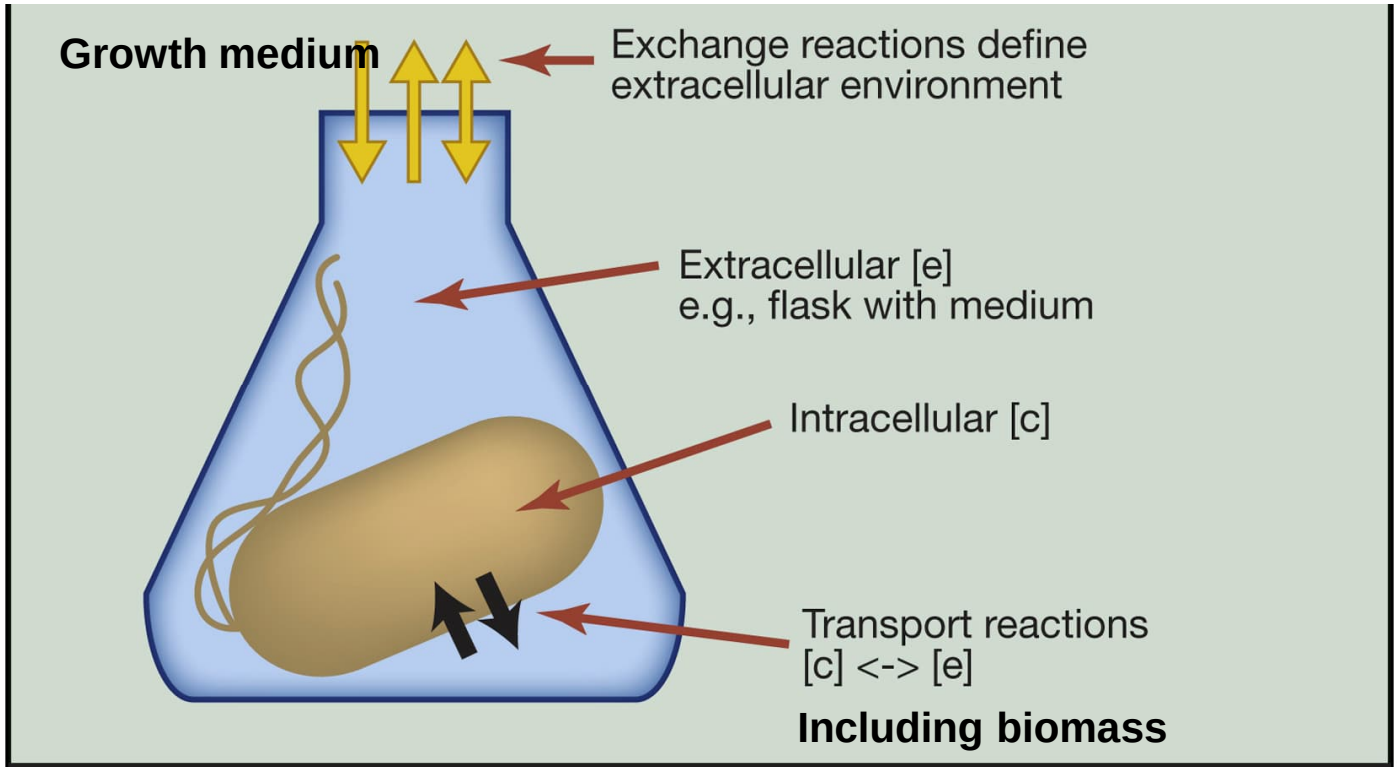
S	stoichiometric matrix
$\boldsymbol{a}$	objective coefficients
$\boldsymbol{v}$	fluxes (specific rates)
$\boldsymbol{v}_{lb}$	flux lower bounds
$\boldsymbol{v}_{ub}$	flux upper bounds

Artificial reactions forming biomass allow growth simulations



O'Brien EJ, Monk JM, Palsson BO. (2015) Cell. 161:971-987. doi: 10.1016/j.cell.2015.05.019.

Metabolic states depend on environment



O'Brien EJ, Monk JM, Palsson BO. (2015) Cell. 161:971-987. doi: 10.1016/j.cell.2015.05.019.

Genome-scale metabolic model applications?

Prediction of metabolic phenotypes from genotype

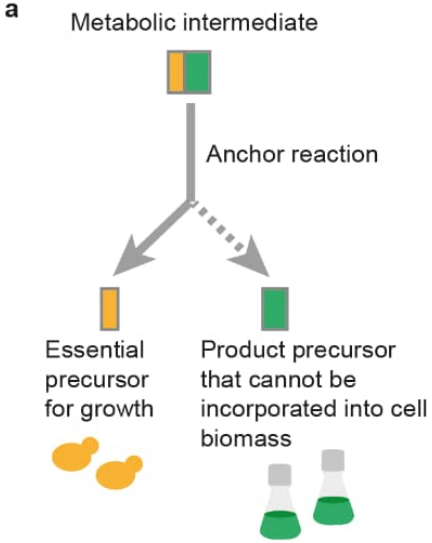
1. Prediction of true (= optimal growth) metabolic states on different nutrients
2. Prediction of optimal product yields
3. Strain design to improve production
4. ...

In silico design of engineering strategies using genome-scale metabolic models

- Growth-product coupling: the cells can only grow if they produce
- Push-pull strategies: expression levels are modified to push and pull more resources to product synthesis

Growth-product coupling

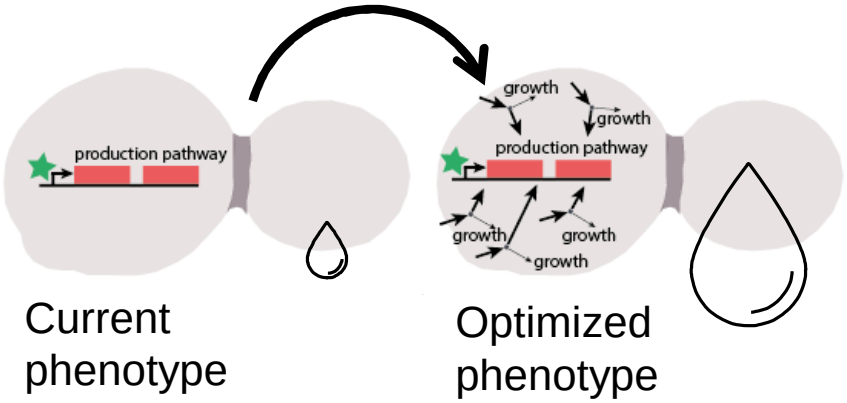
Algorithms use model simulations for identifying knock-out targets



Jouhten P. et al. Metab Eng. (2017)

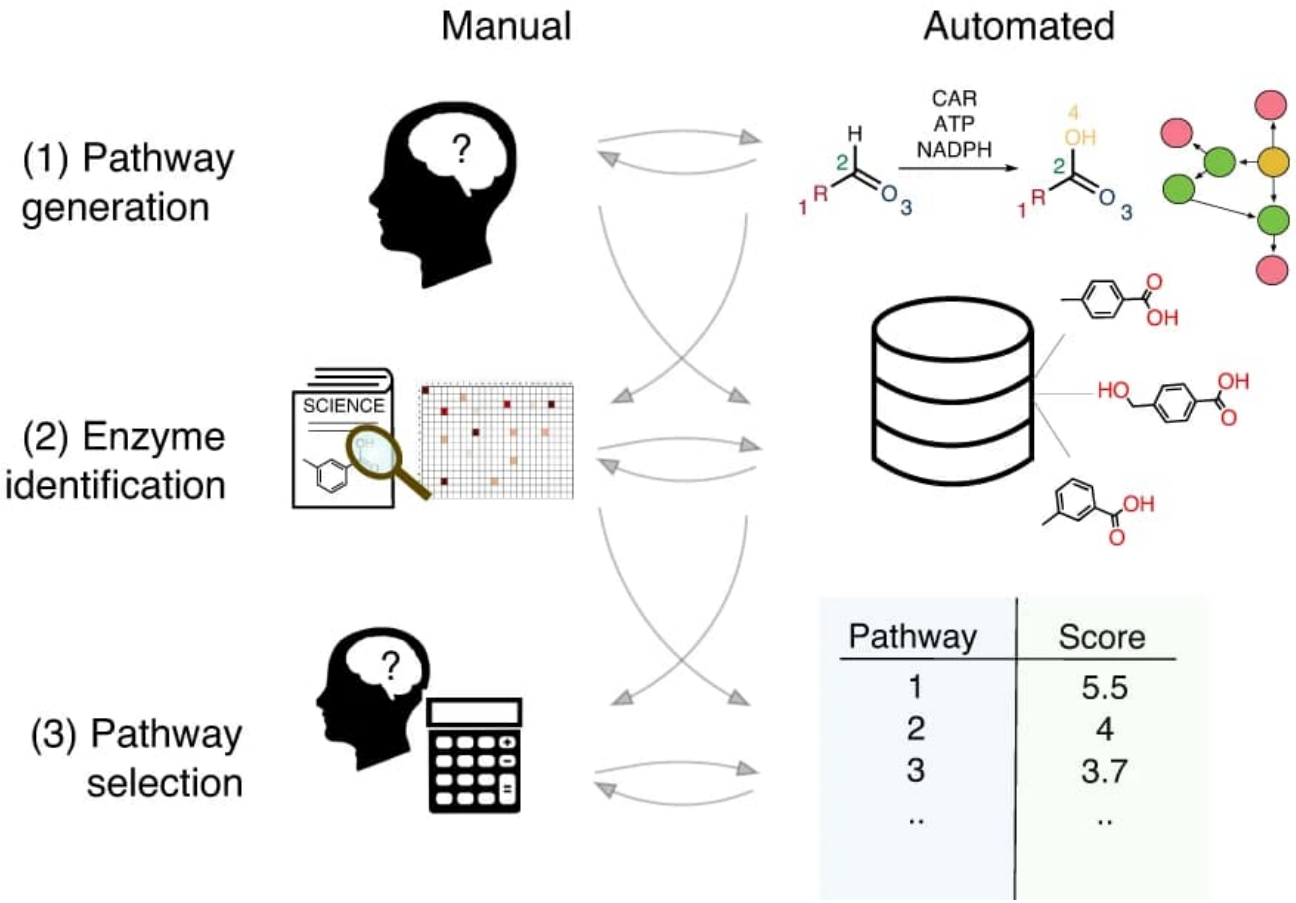
Push-pull strategies

Algorithms use model simulations for identifying deletion and re-regulation targets



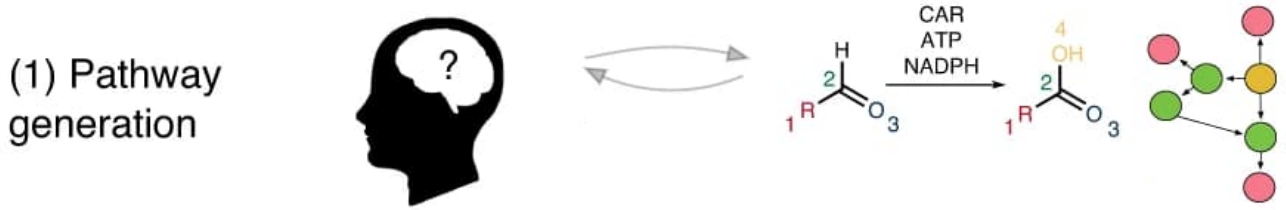
Jouhten P. et al. *unpublished work* with Kiran Patil, EMBL Heidelberg

Synthetic pathway design



Finnigan et al. (2021) Nat Catal 4:98-104. doi: 10.1038/s41929-020-00556-z.

Pathway generation



- A retrosynthesis (routes from desired compound back to precursors) problem
- Known biochemical reactions are available in databases like Kegg, Metacyc, Rhea
- Potential reactions that enzymes could also catalyze can be computationally designed

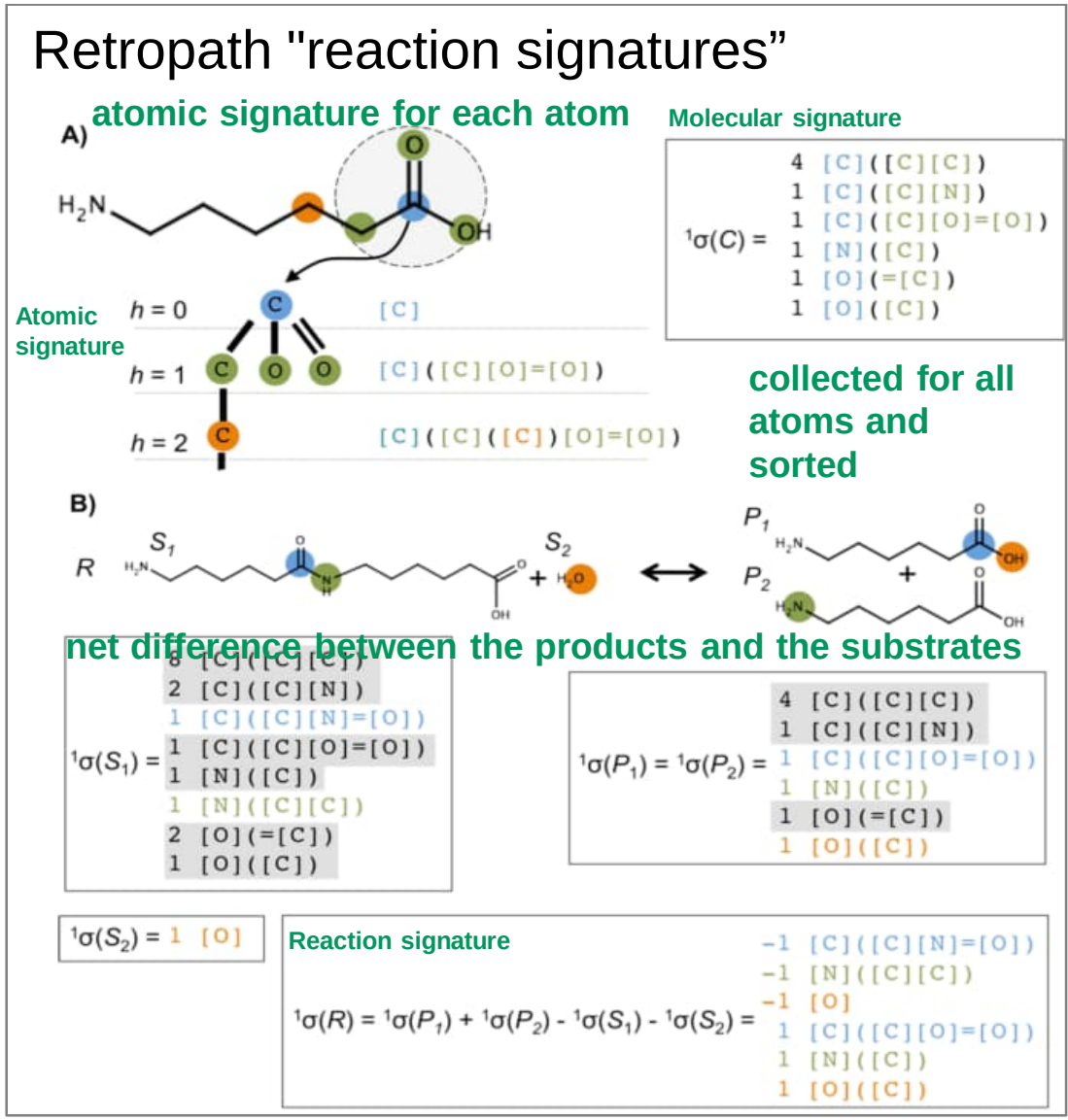
Finnigan et al. (2021) Nat Catal 4:98-104. doi: 10.1038/s41929-020-00556-z.

Potential enzyme catalyzed reactions may follow rules of known reactions

- **Reaction rules** model known reactions
- Prediction of potential reactions assume that the core of the reaction (where the bonds break) remains the same
- Define different **dimensions of the core**
- **Extended metabolic space** is formed of endogeneous and potential reactions (given a specific dimension)

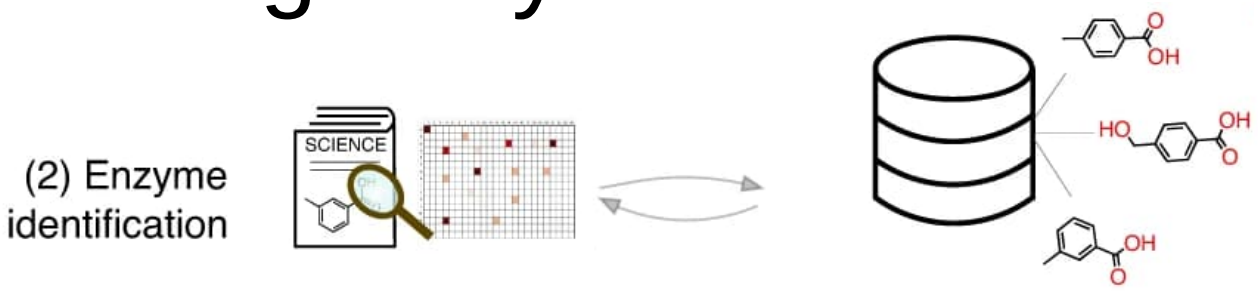
Extended metabolic space (height = dimension)

height *h*	reactions	% increase from canonical
2	9083	17.72%
3	7882	2.15%
4	7800	1.09%
5	7752	0.47%
6	7725	0.12%
canonical	7716	0%



Carbonell, P., Planson, A.-G., Fichera, D., & Faulon, J.-L. (2011). BMC Systems Biology, 5(1), 122.

Finding enzymes



- The reactions are realized through the Central Dogma:

  gene -> mRNA -> protein

- Known reactions: **sequence and structure similarity-based** search using known sequences as seeds
- Potential reactions: reaction rules can be used for identifying seed sequences that may encode also the desired activity (due to promiscuity)

Finnigan et al. (2021) Nat Catal 4:98-104. doi: 10.1038/s41929-020-00556-z.

Novel protein design is coming within reach

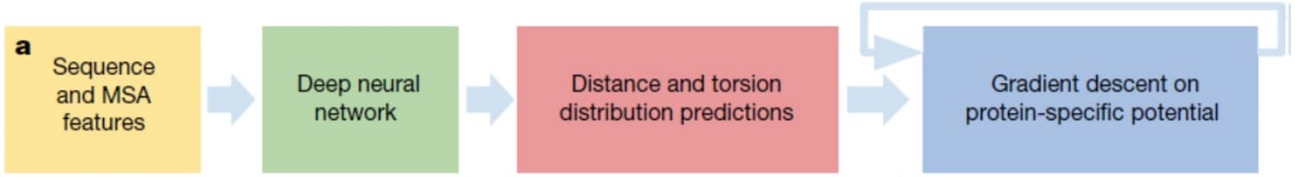
AlphaFold by DeepMind is a breakthrough in natural protein folding prediction

Article

Improved protein structure prediction using potentials from deep learning

https://doi.org/10.1038/s41586-019-1923-7
Received: 2 April 2019
Accepted: 10 December 2019
Published online: 15 January 2020

Andrew W. Senior[1,4*], Richard Evans[1,4], John Jumper[1,4], James Kirkpatrick[1,4], Laurent Sifre[1,4], Tim Green[1], Chongli Qin[1], Augustin Žídek[1], Alexander W. R. Nelson[1], Alex Bridgland[1], Hugo Penedones[1], Stig Petersen[1], Karen Simonyan[1], Steve Crossan[1], Pushmeet Kohli[1], David T. Jones[2,3], David Silver[1], Koray Kavukcuoglu[1] & Demis Hassabis[1]



https://www.deepmind.com/research/highlighted-research/alphafold

Synthetic pathway design

(3) Pathway selection

Pathway	Score
1	5.5
2	4
3	3.7
..	..

- Criteria e.g., theoretical yield, thermodynamics of reactions, pathway length, number of new-to-nature reactions, toxicity
- Prediction of performance within native metabolism using genome-scale metabolic models

Finnigan et al. (2021) Nat Catal 4:98-104. doi: 10.1038/s41929-020-00556-z.