Metabolic modelling and synthetic pathway design

Paula Jouhten

Learning goals

After this lecture, you will be able to…

Describe what are genome-scale metabolic models

Describe how are metabolic phenotype prediction and strain design performed

Describe the steps of synthetic metabolic pathway design

Reading material

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

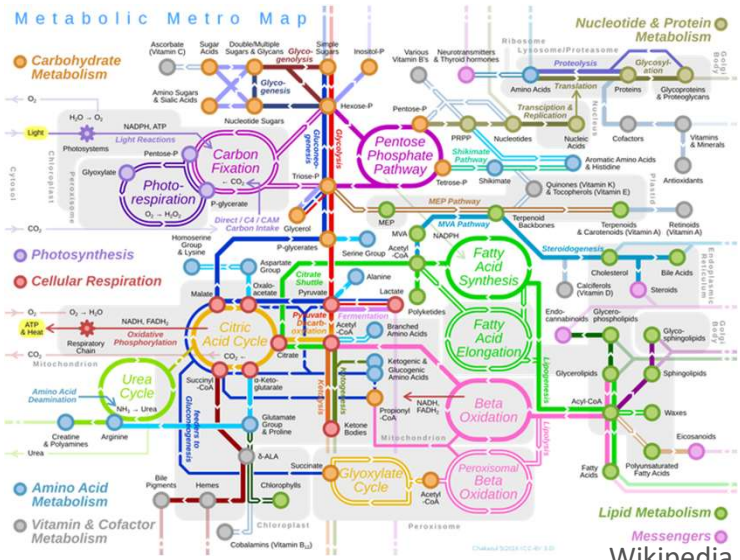
Box 2 outdated, check instead for COBRA toolbox, **COBRApy**, COBRA.jl: http://opencobra.github.io/

Metabolic modelling

Why is metabolism relevant for synthetic biology?

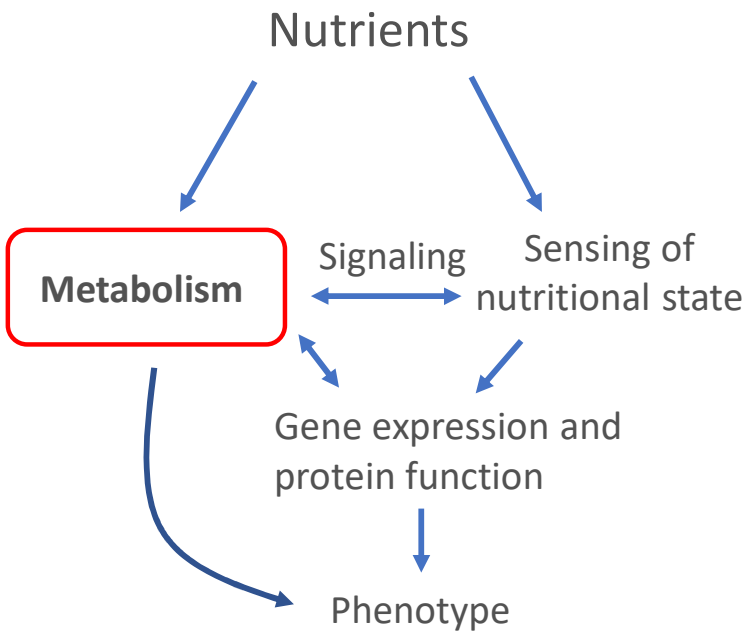
Metabolism = (bio)chemical reactions involved in sustaining a living state of cells and an organism

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



Wikipedia

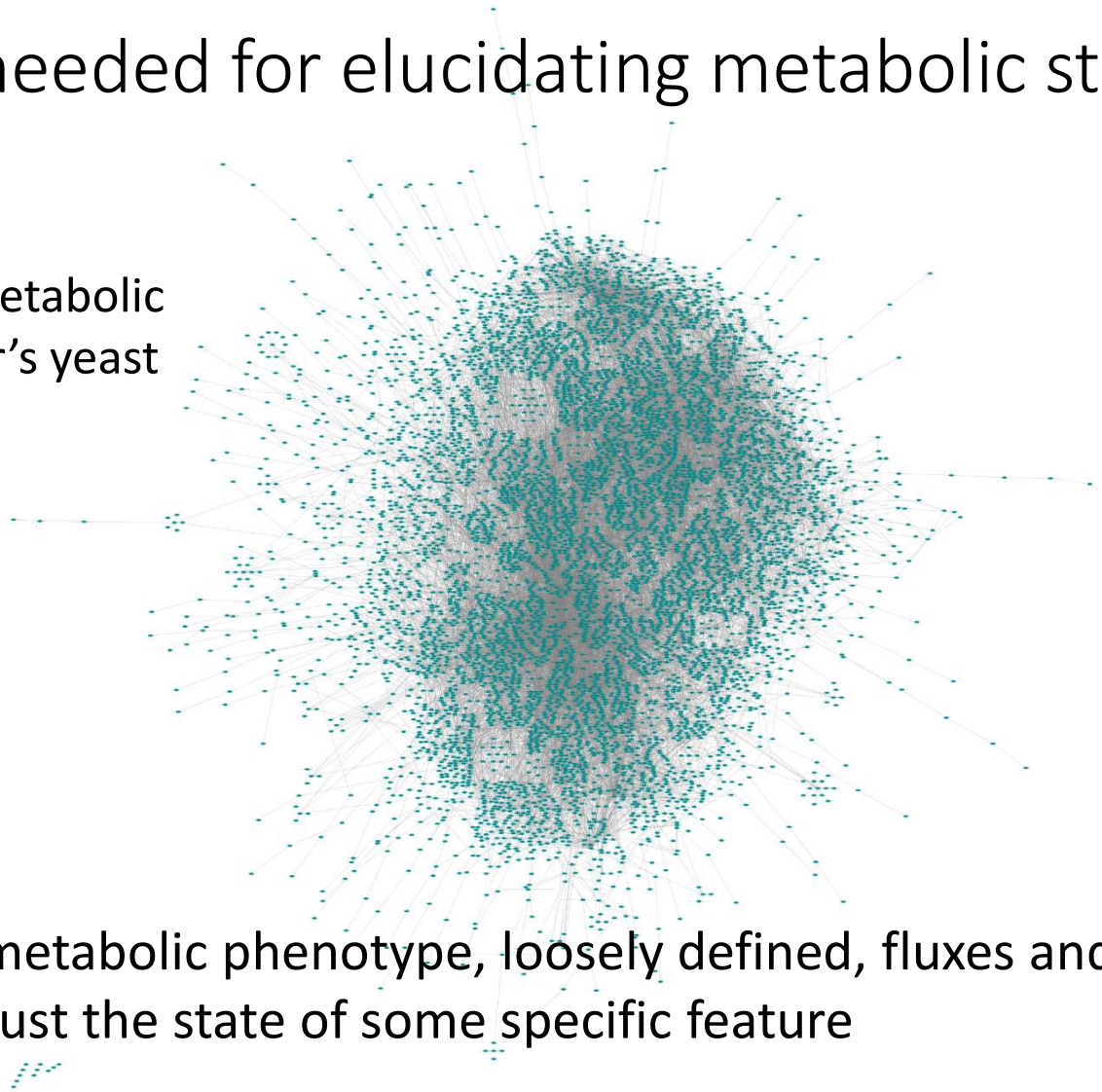
Metabolism is involved in cellular regulation



Adopted from Jaakko Mattila

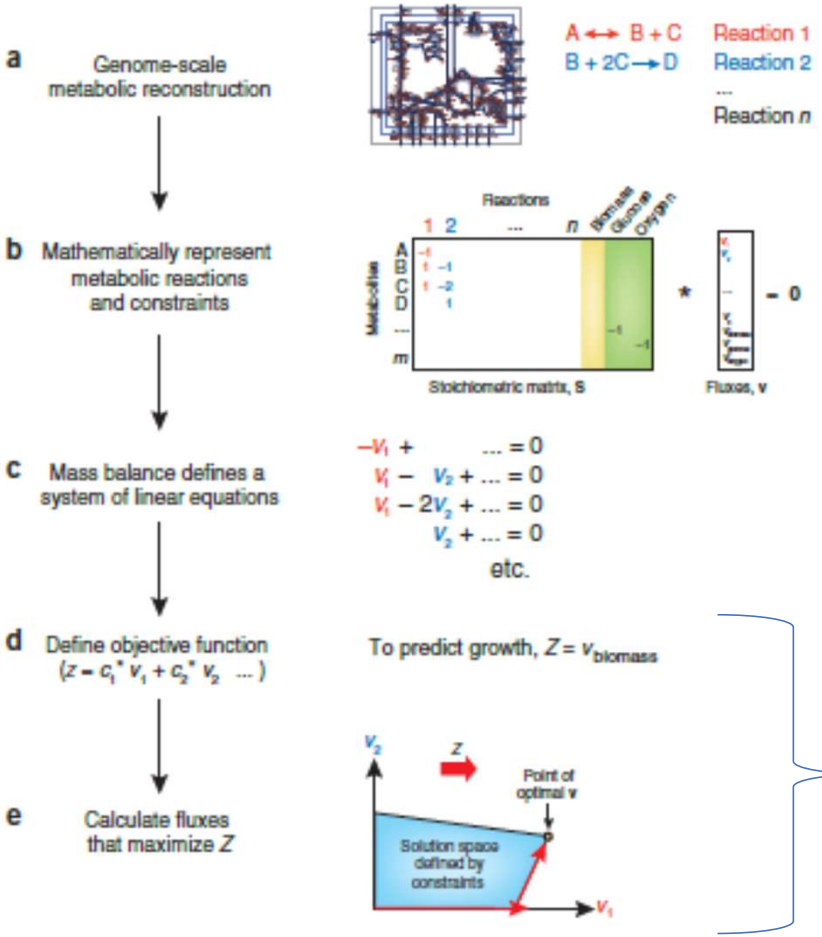
Modelling is needed for elucidating metabolic states

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

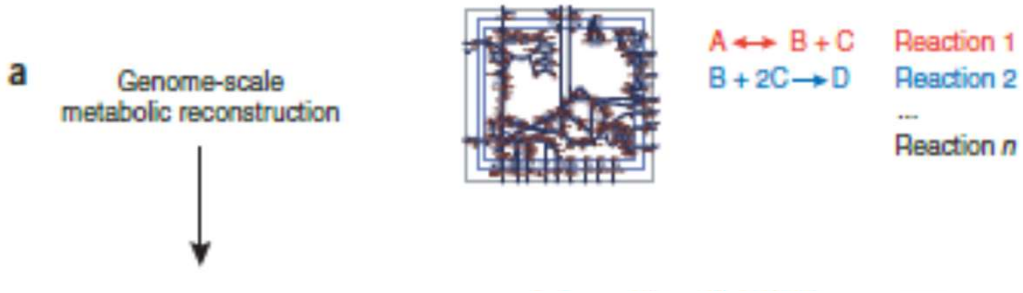
Phenotype prediction using genome-scale metabolic models



Model simulation algorithm:
Flux balance analysis (FBA)

Orth et al. (2010) Nat Biotechnol. 28:245-8. doi: 10.1038/nbt.1614.

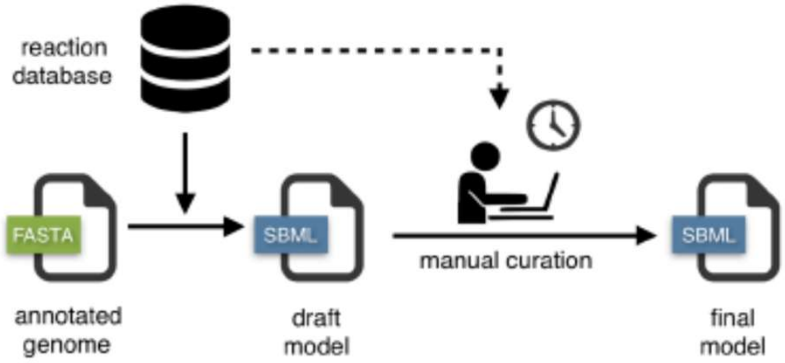
Genome-scale metabolic model reconstruction

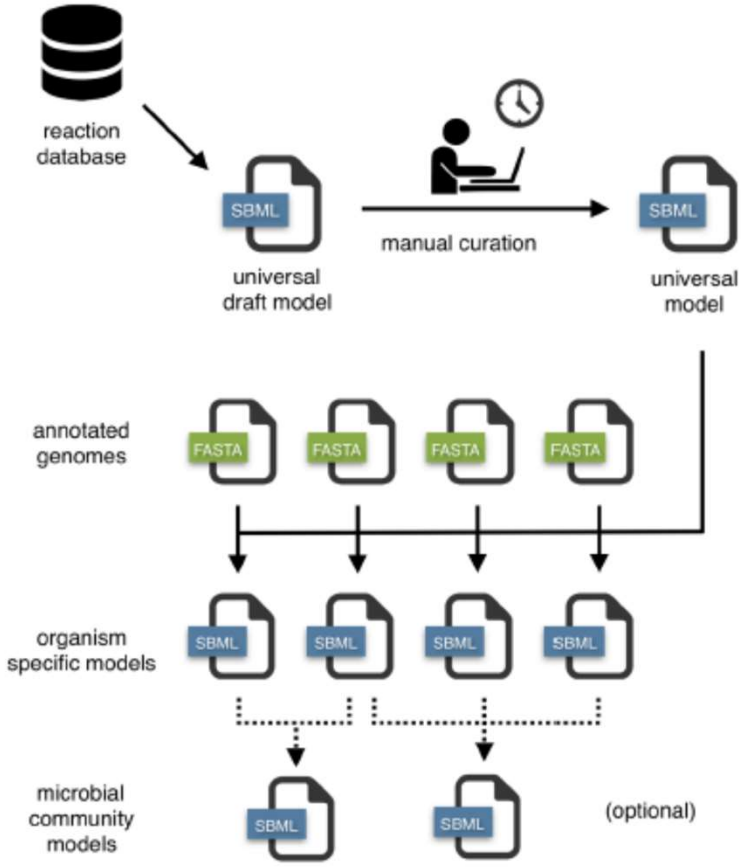


Which reactions can take place in the cells of a species?

Orth et al. (2010) Nat Biotechnol. 28:245-8. doi: 10.1038/nbt.1614.

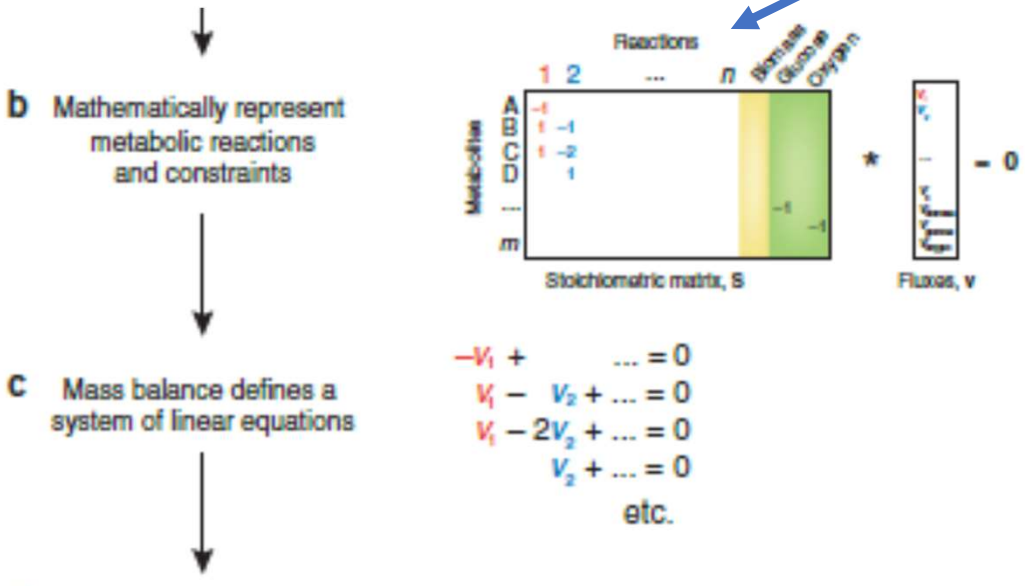
Genome-scale metabolic model reconstruction





Machado et al. (2018) Nucleic Acids Res. 46:7542-7553. doi: 10.1093/nar/gky537.

Conversion into mathematical representation

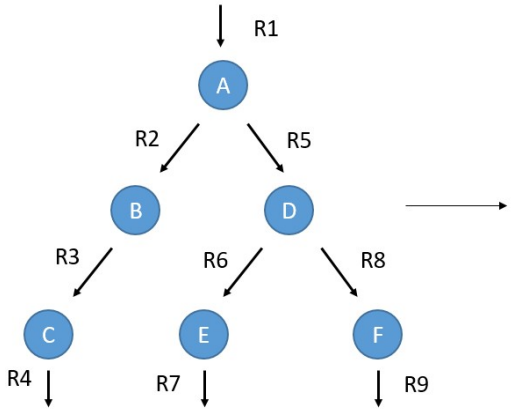


Stoichiometric matrix

Metabolite mass balances are zeros only as a metabolic steady state is assumed

Orth et al. (2010) Nat Biotechnol. 28:245-8. doi: 10.1038/nbt.1614.

Toy model example



Stoichiometric matrix

Metabolites	R1	R2	R3	R4	R5	R6	R7	R8	R9
A	1	-1			-1				
B		1	-1						
C			1	-1					
D					1	-1		-1	
E						1	-1		
F								1	-1

Reactions

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

Obeying the law of conservation of mass,
metabolite mass balances constrain metabolic phenotypes

$$\frac{d\boldsymbol{X}}{dt} = \boldsymbol{S} \cdot \boldsymbol{v} = \boldsymbol{S} \cdot \boldsymbol{f}(\boldsymbol{e}(\boldsymbol{t}), \boldsymbol{s}(\boldsymbol{t}), \boldsymbol{p})$$

Steady state assumption linearizes the mass balances

$$\frac{d\boldsymbol{X}}{dt} = \boldsymbol{S} \cdot \boldsymbol{v} = \boldsymbol{S} \cdot \boldsymbol{f}(\boldsymbol{e}(\boldsymbol{t}), \boldsymbol{s}(\boldsymbol{t}), \boldsymbol{p}) \quad = 0$$

Constraints:
1) $\boldsymbol{Sv} = 0$
2) $\boldsymbol{v,lb} < \boldsymbol{v} < \boldsymbol{v,ub}$

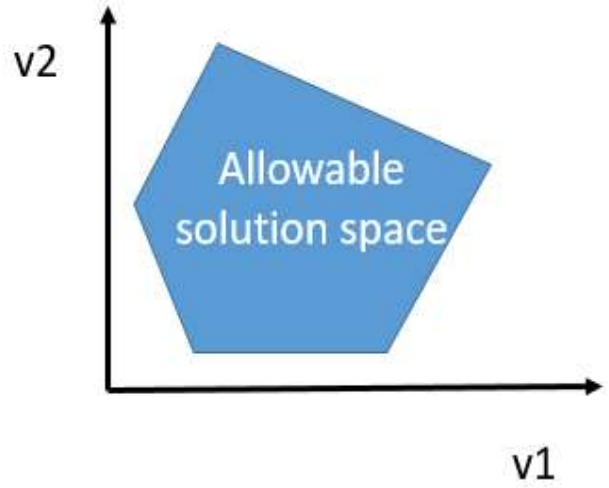
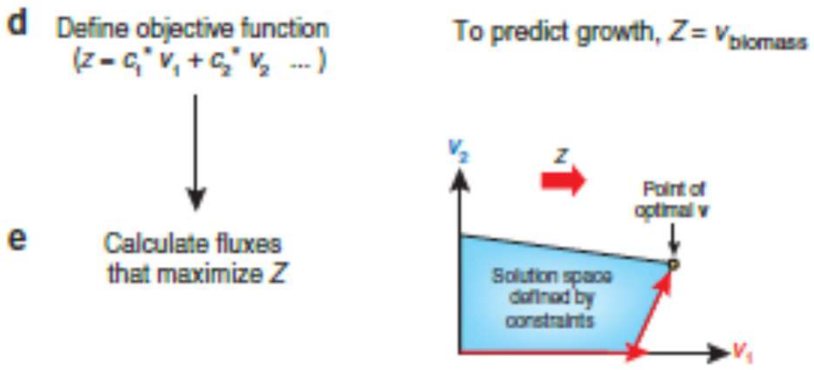


Figure modified by Tuula Tenkanen from O'Brien et al. 2015

The linear system is lighter to solve and free of kinetic equations and parameters
Additional constraints introduced to obey the second law of thermodynamics

Defining an objective function for forming an optimization problem



Choice of the objective function depends on the question

After the optimization problem is defined it can be solved using any linear optimization solver

Orth et al. (2010) Nat Biotechnol. 28:245-8. doi: 10.1038/nbt.1614.

Linear optimization can be used to identify different optimal metabolic states

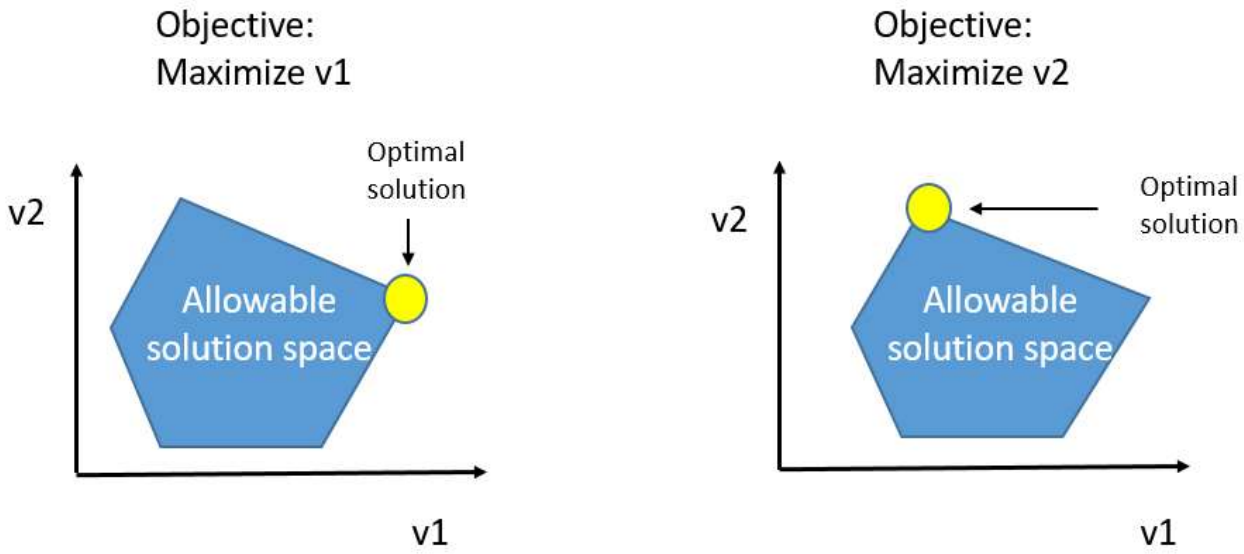


Figure modified by Tuula Tenkanen from O'Brien et al. 2015

Flux Balance Analysis (FBA)

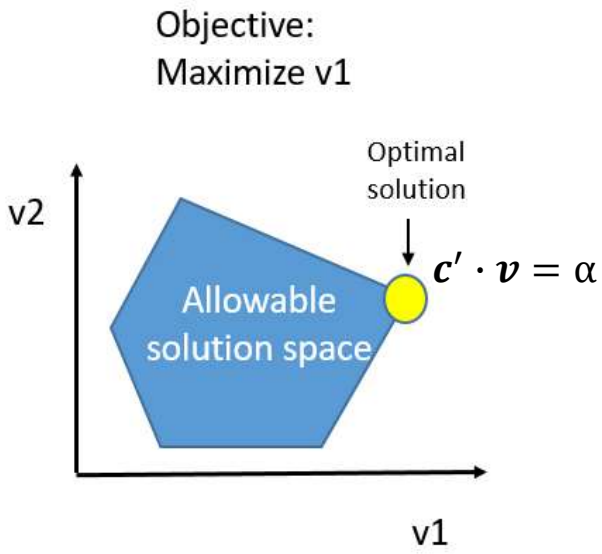
Varma and Palsson, 1993; Varma and Palsson, 1994

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

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

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

Linear optimization can be used to identify optimal metabolic states



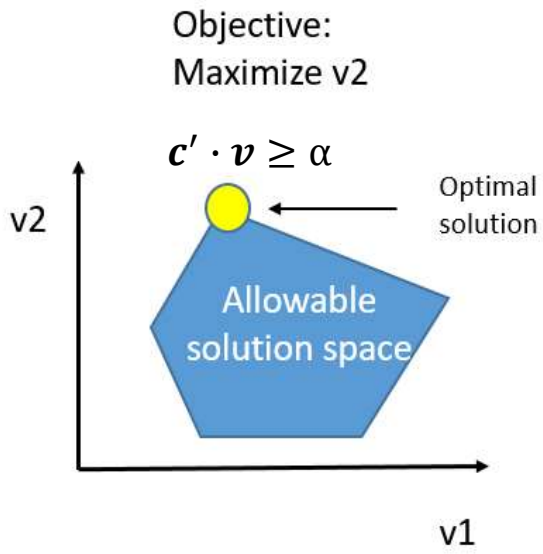


Figure modified by Tuula Tenkanen from O'Brien et al. 2015

Flux Variability Analysis (FVA)

Mahadevan et al. 2003

maximize and minimize v_i

subject to

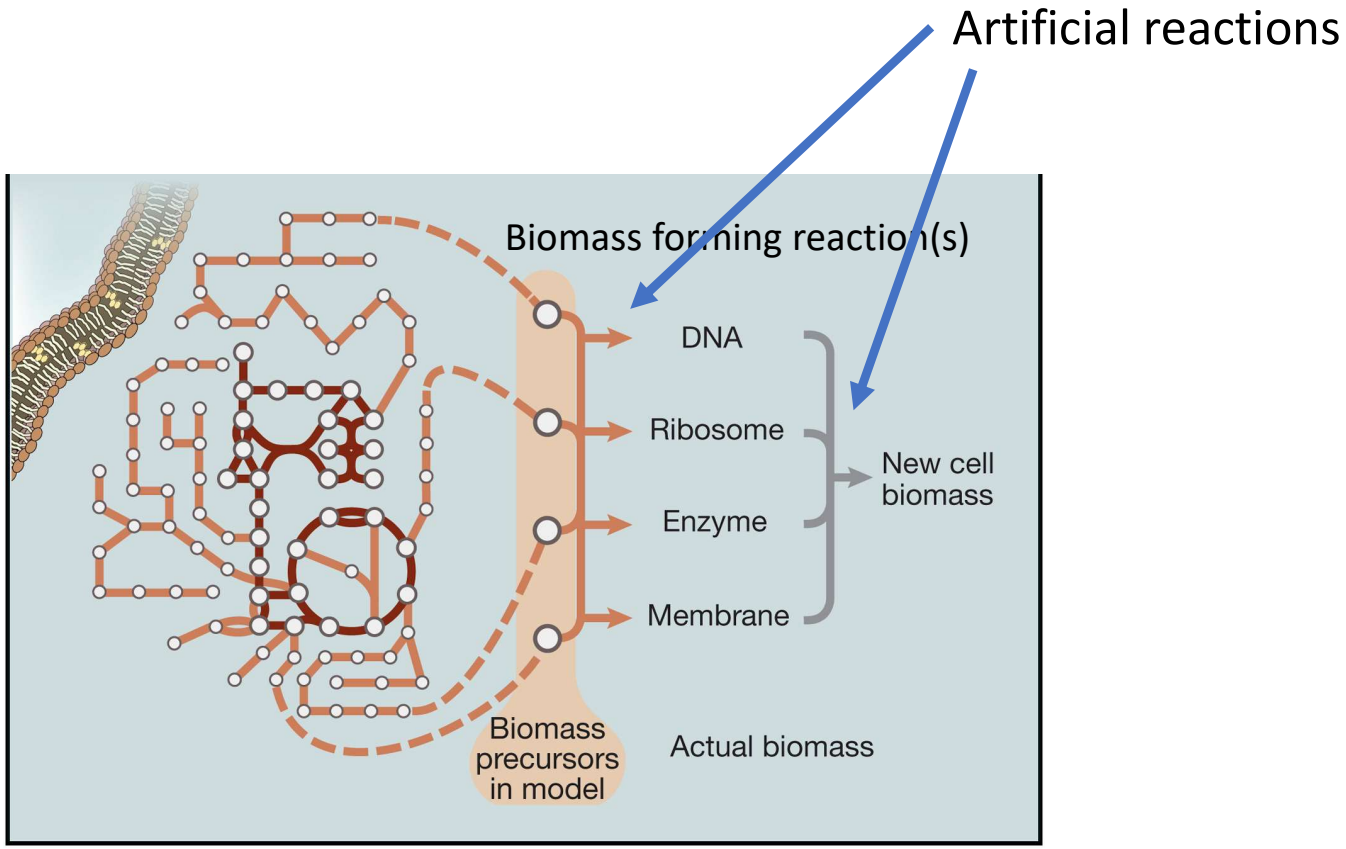
$$S \cdot v = 0$$
$$c' \cdot v \geq \alpha$$
$$v,lb < v < v,ub$$

α is the optimal value of the inital objective

⇒ **While the objective has the optimal value other fluxes may vary**

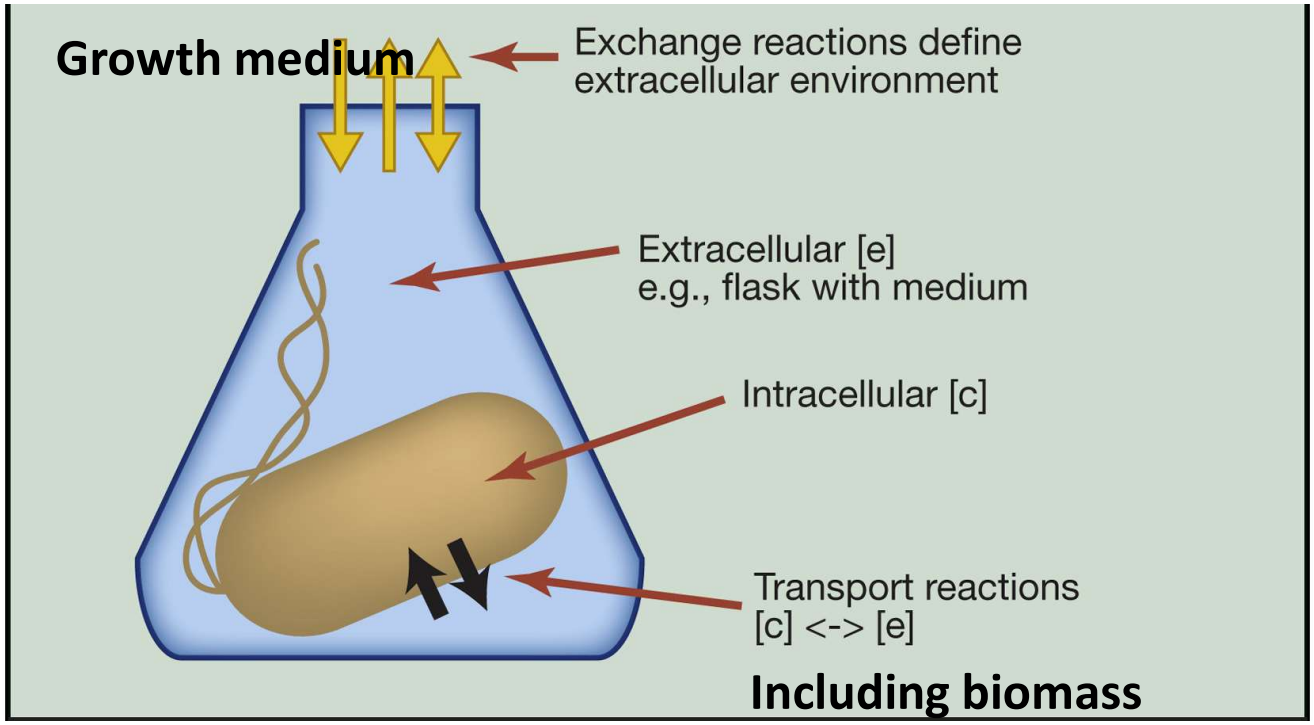
⇒ **The ones that are non-zero are essential for the optimal value of the objective**

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.

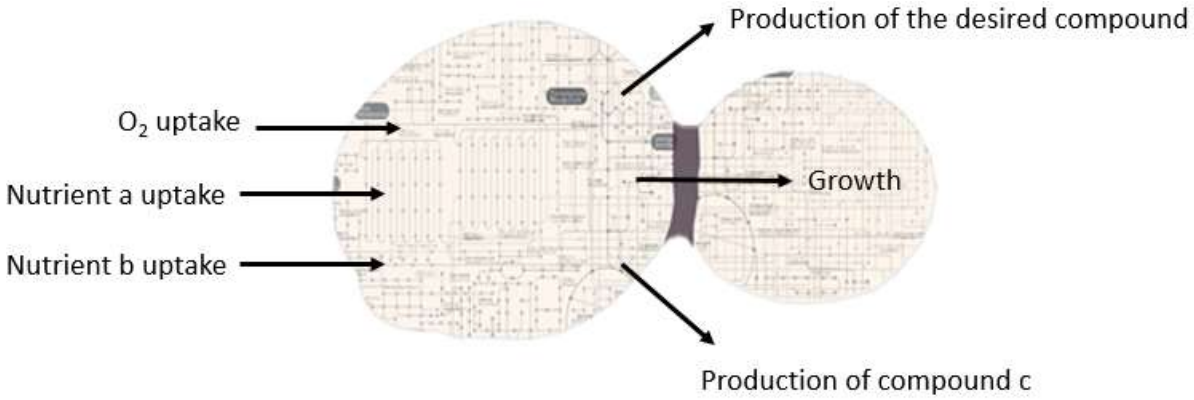
Metabolic states depend on environment



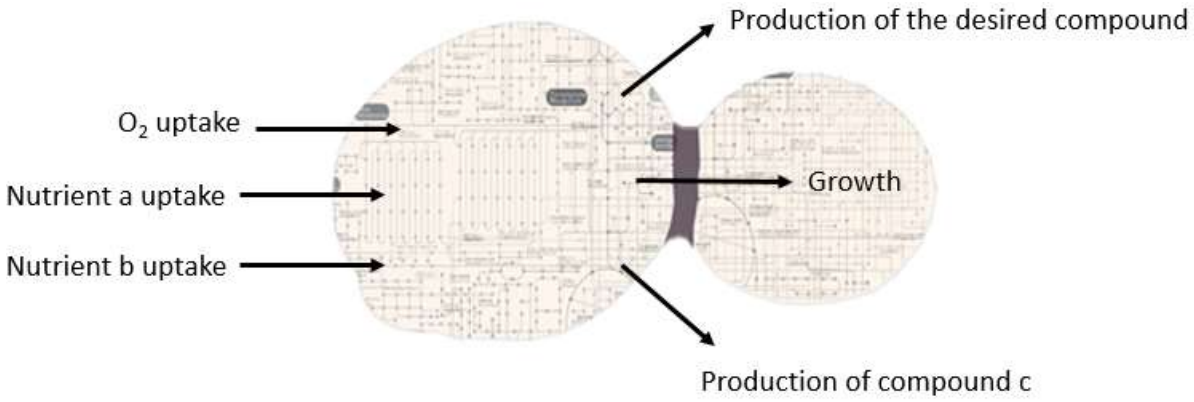
O'Brien EJ, Monk JM, Palsson BO. (2015) Cell. 161:971-987. doi: 10.1016/j.cell.2015.05.019.

Specific fluxes

- Flux units depend on how the artificial biomass producing reaction is defined
- If it is defined as mmol of precursors for generating 1 g cell dry weight (CDW), then flux units are mmol/(g CDW * h)

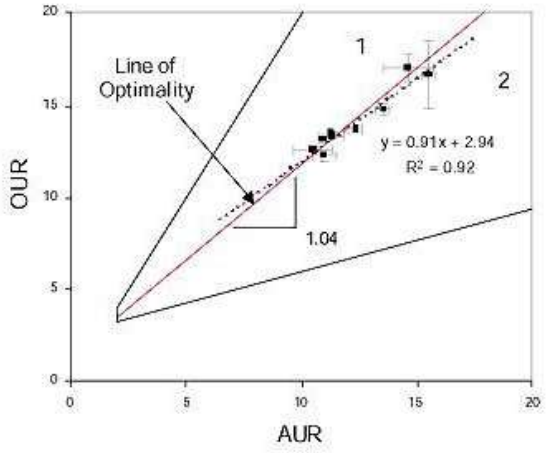


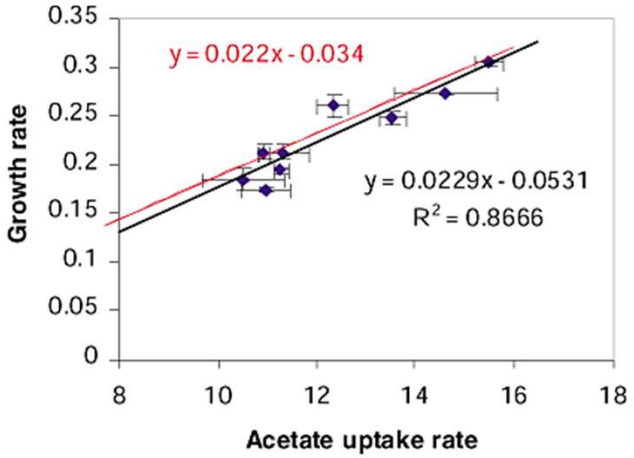
Prediction vs estimation of metabolic state?



When arbitrary constraints are used, yields can be predicted
When empirical rates are used as constraints, other rates can be estimated or predicted

FBA simulations optimizing growth predict well experimental phenotypes

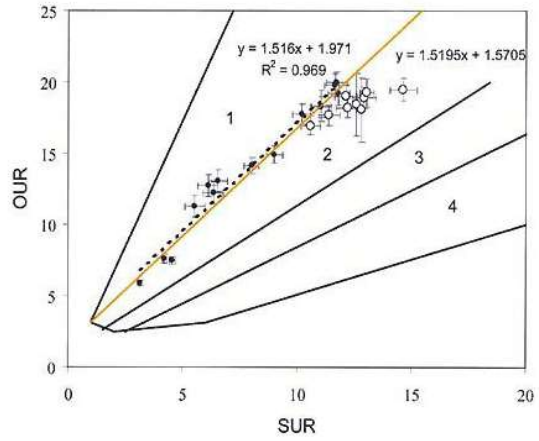


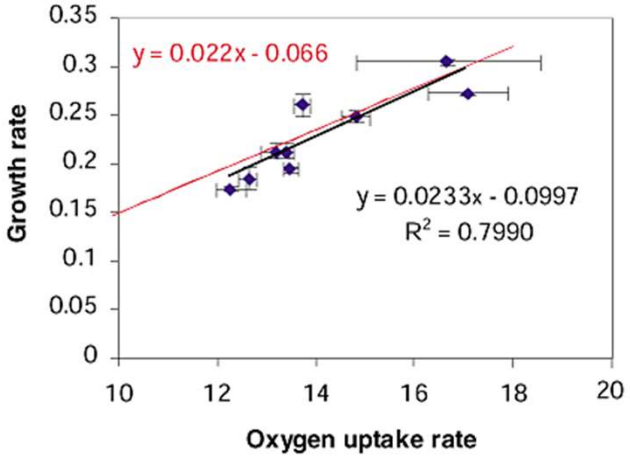


In silico predictions of *Escherichia coli* metabolic capabilities are consistent with experimental data

Jeremy S. Edwards[1,2], Rafael U. Ibarra[1], and Bernhard O. Palsson[1*]

[1]Department of Bioengineering, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0412. [2]Current address: Department of Chemical Engineering, University of Delaware, Newark, DE 19716. *Corresponding author (palsson@ucsd.edu).





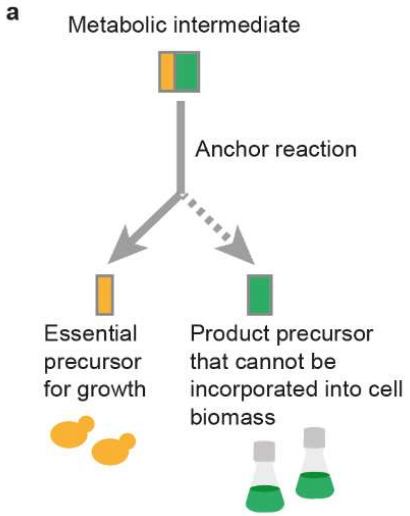
Design of strain engineering strategies

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 production

Growth-product coupling

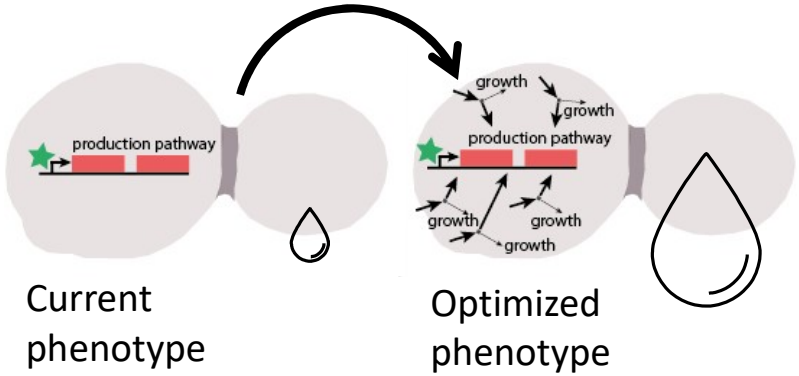
Algorithms use genome-scale metabolic models for identifying knock-out targets



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

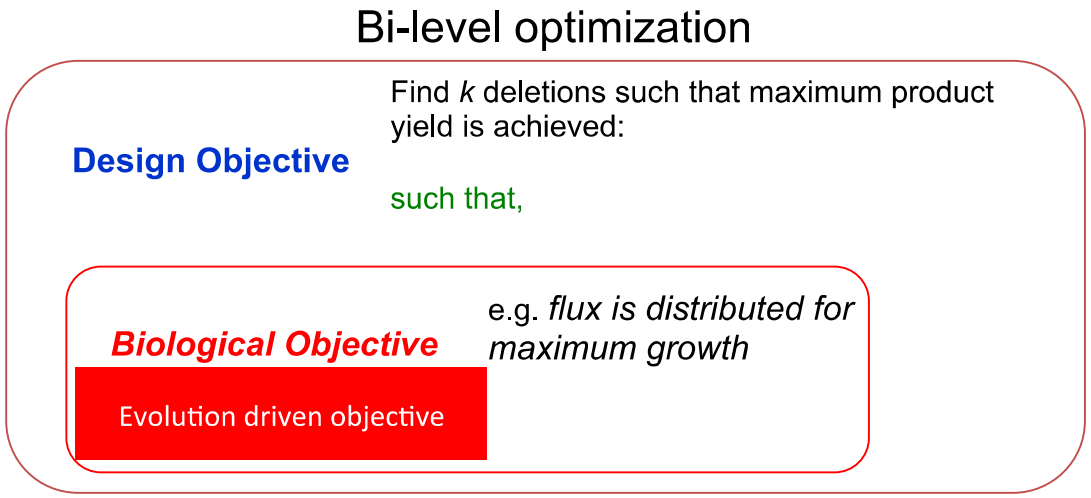
Push-pull strategies

Algorithms use genome-scale metabolic models for identifying deletion and re-regulation targets



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

Growth-product coupling aligns biological and engineering objectives

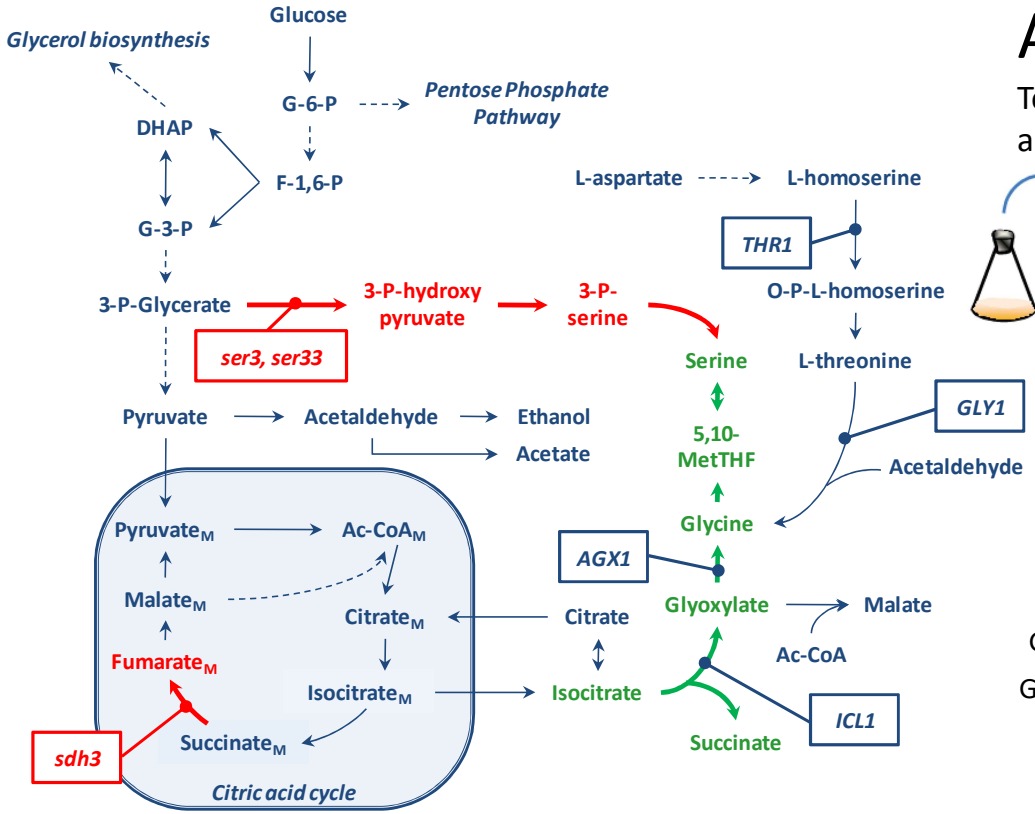


OptKnock: Burgard et al. (2003)
OptGene: Patil et al. (2005)

Slide modified from Kiran Patil

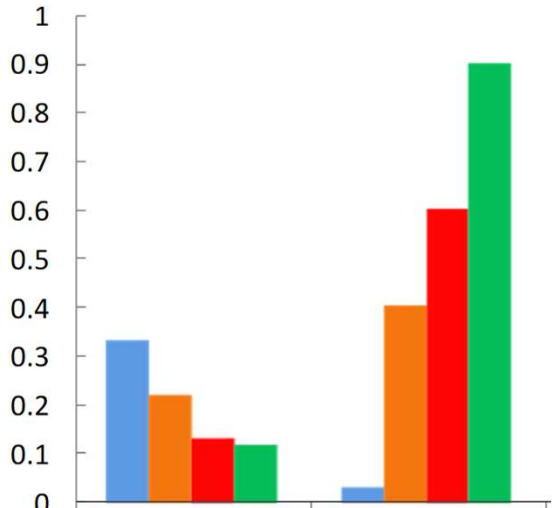
Growth-product coupling allows using adaptive laboratory evolution for improving production

Proof of concept: succinate production in *S. cerevisiae*



ALE

To recover from Gly auxotrophy



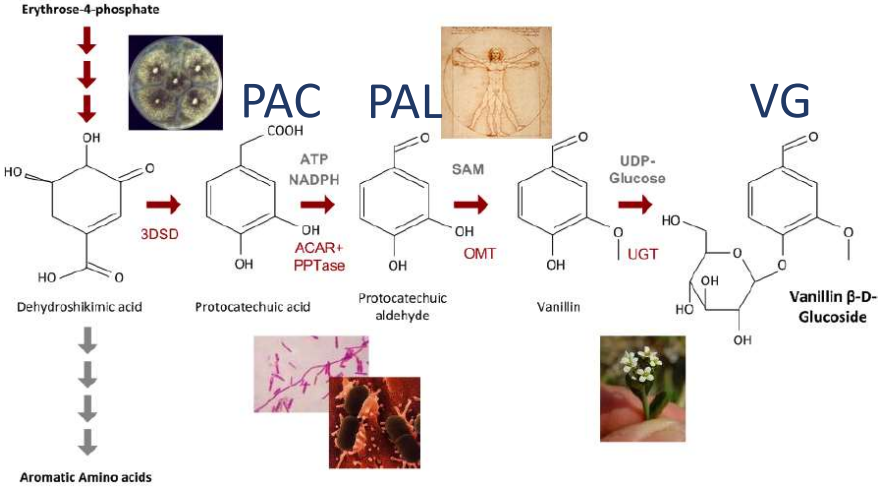
	Specific Growth Rate (1/h)	Maximum Titer (g/L)
Gly auxotrophic / Gly prototrophic		
REF	0.33	0.03
8D	0.22	0.40
8D Evolved	0.13	0.60
8D Evolved + pICL1	0.12	0.90

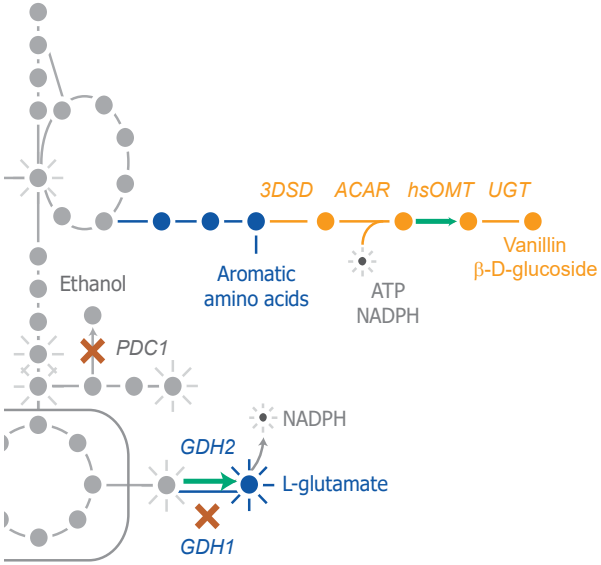
Otero *et al.* PLoS One. (2013) **8**:e54144.

Slide from Kiran Patil

Pathway optimization improved vanillin production only after designed optimization of network

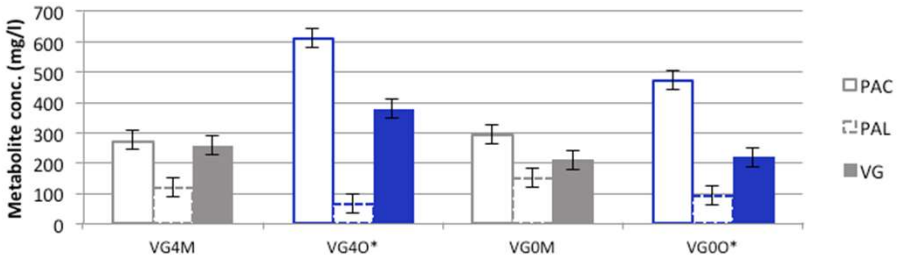
Synthetic vanillin pathway





Experimental validation

Overall 5-fold productivity improvement



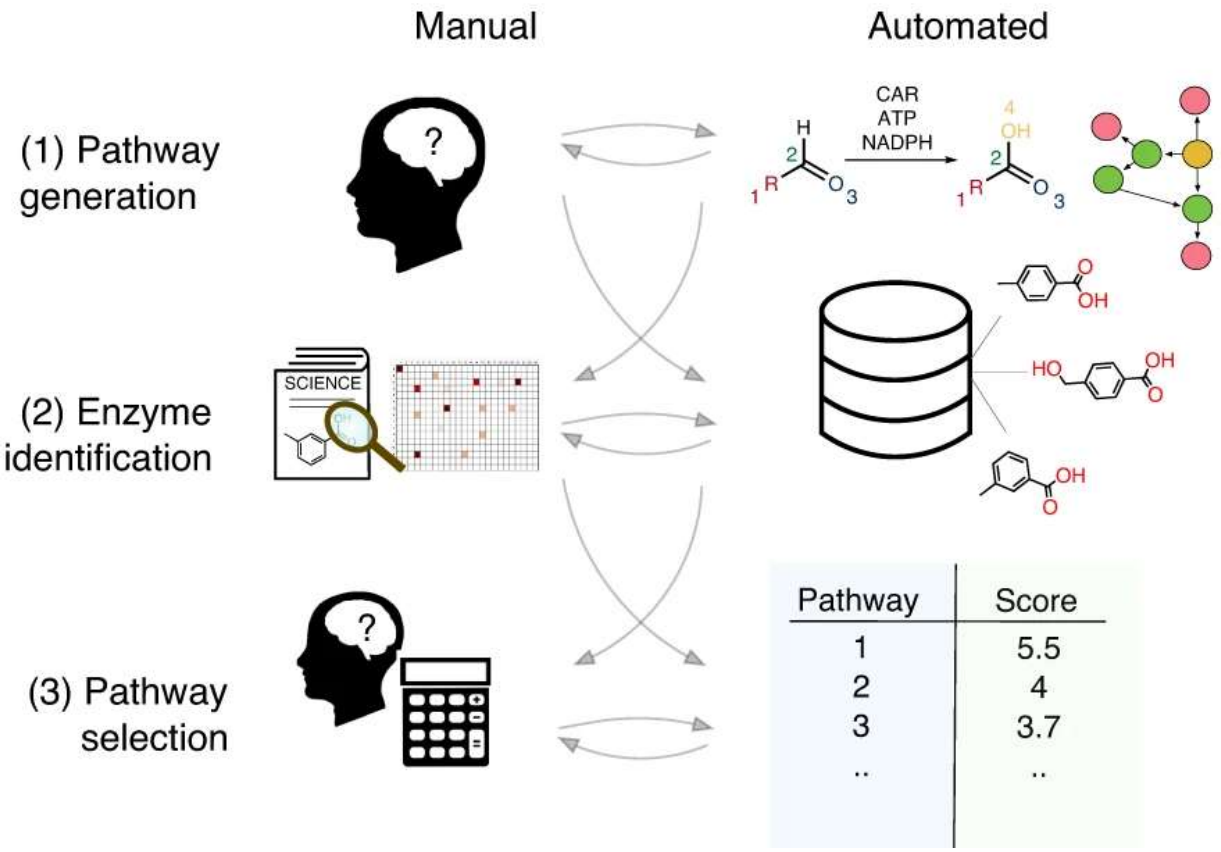
Strains with overexpressed O-methyltransferase in blue

Brochado et al. (2011, 2013). Dr. Kiran Patil in collaboration with Evolva A/S (Denmark)

Slide from Kiran Patil

Synthetic pathway design

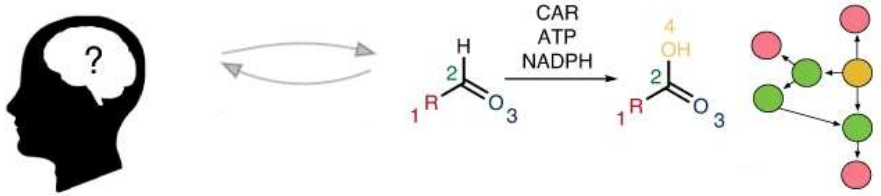
Synthetic pathway design



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

Biochemical pathway generation

(1) Pathway generation



- Can be defined as a retrosynthesis problem from desired compound back to precursors in microbial cells
- Such pathways can be searched through known biochemical reactions from data bases like Kegg, Metacyc, Rhea
- They can also be searched through potential reactions that enzymes could catalyze defined by reaction rules

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

Reaction rules model possible enzyme catalyzed reactions

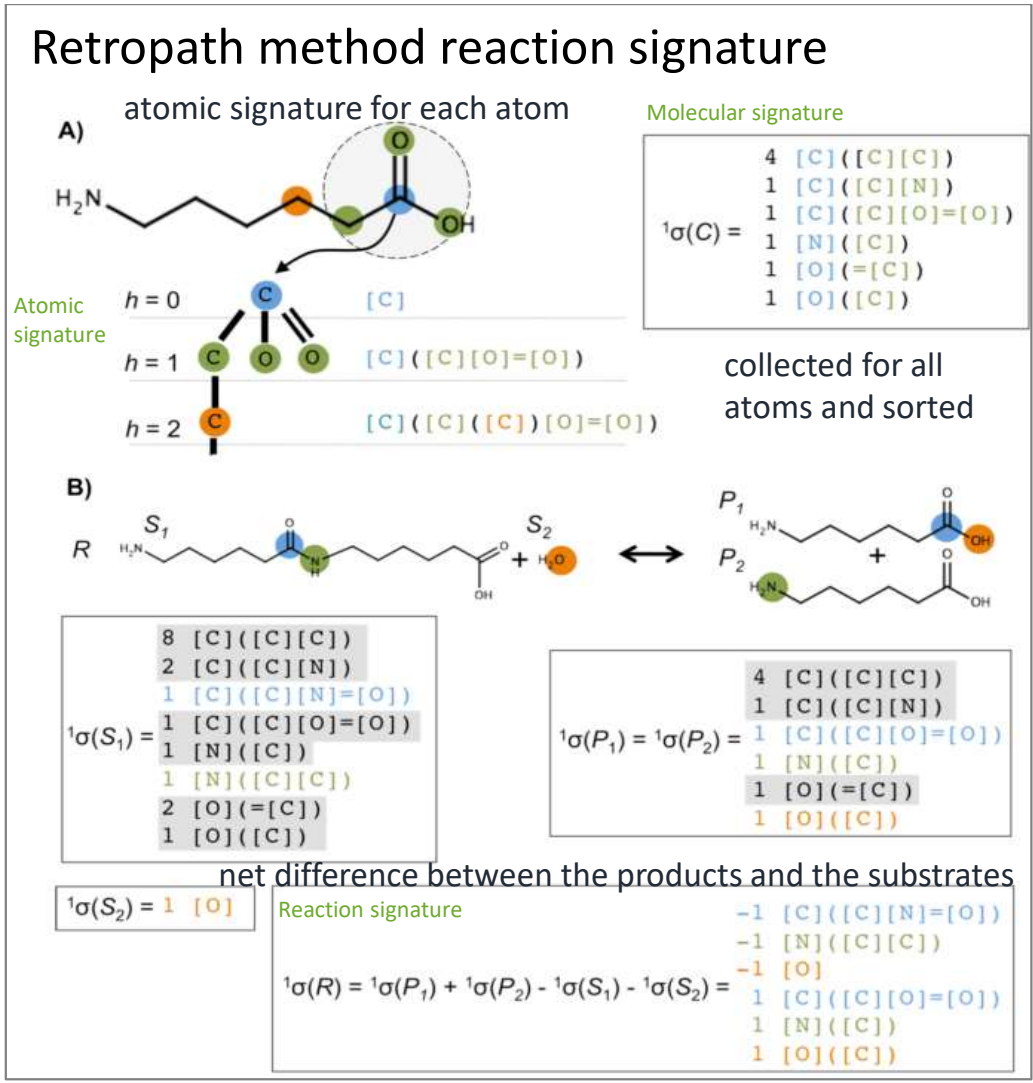
- Rules model similarities to known reactions (i.e. similarities of reactants)
- Assume that if the core of the reaction (where the bonds break) remains the same then an enzyme could be found/built for the novel reaction
- Define different dimensions of the core

- Reaction rules create extended metabolic space

Table 1 Reactions in the EMRS

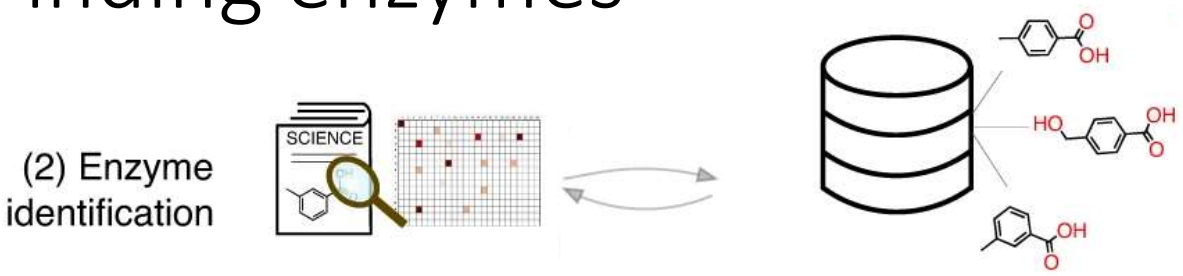
height *h*	reactions	% increase from canonical
2	9083	17.72%
3	7882	2.15%
4	7800	1.09%
5	7752	0.47%
6	7725	0.12%
canonical	7716	0%

Number of novel generated putative reactions in the EMRS for different heights *h*.



Carbonell, P., Planson, A.-G., Fichera, D., & Faulon, J.-L. (2011). A retrosynthetic biology approach to metabolic pathway design for therapeutic production. BMC Systems Biology, 5(1), 122.

Finding enzymes

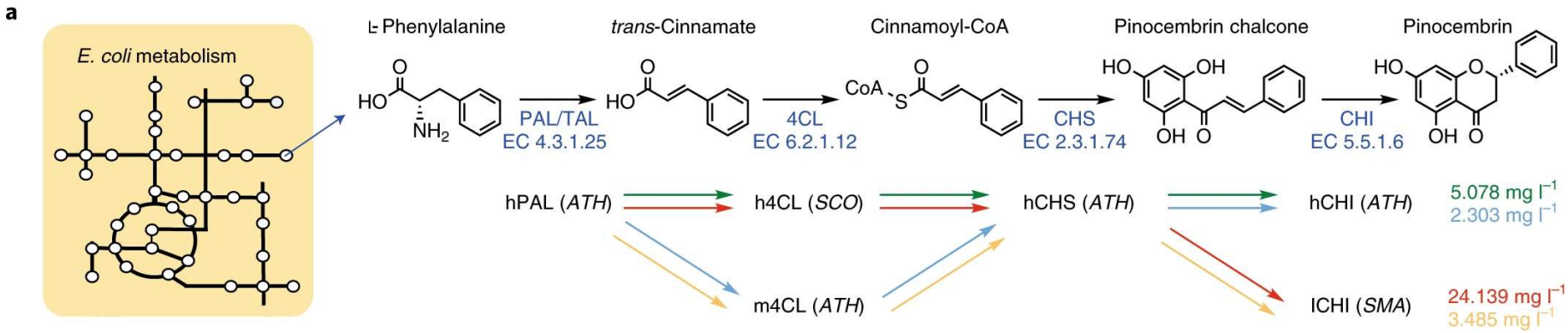


- If the reactions were already known and annotated with enzyme sequences, more candidate sequences found from sequence resources using similarity-based search
- If many sequences encoding the desired enzymatic activity are known, likely important sequence features for the activity can be identified
- If no sequence is known, reaction rules can be used for identifying sequences that may encode also the desired activity (i.e. due to promiscuity)

If substantial **sequence similarity** is observed, the sequences are likely **homologous (i.e. share ancestry in evolution)**

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

Synthetic pathway to pinocembrin to *E.coli*



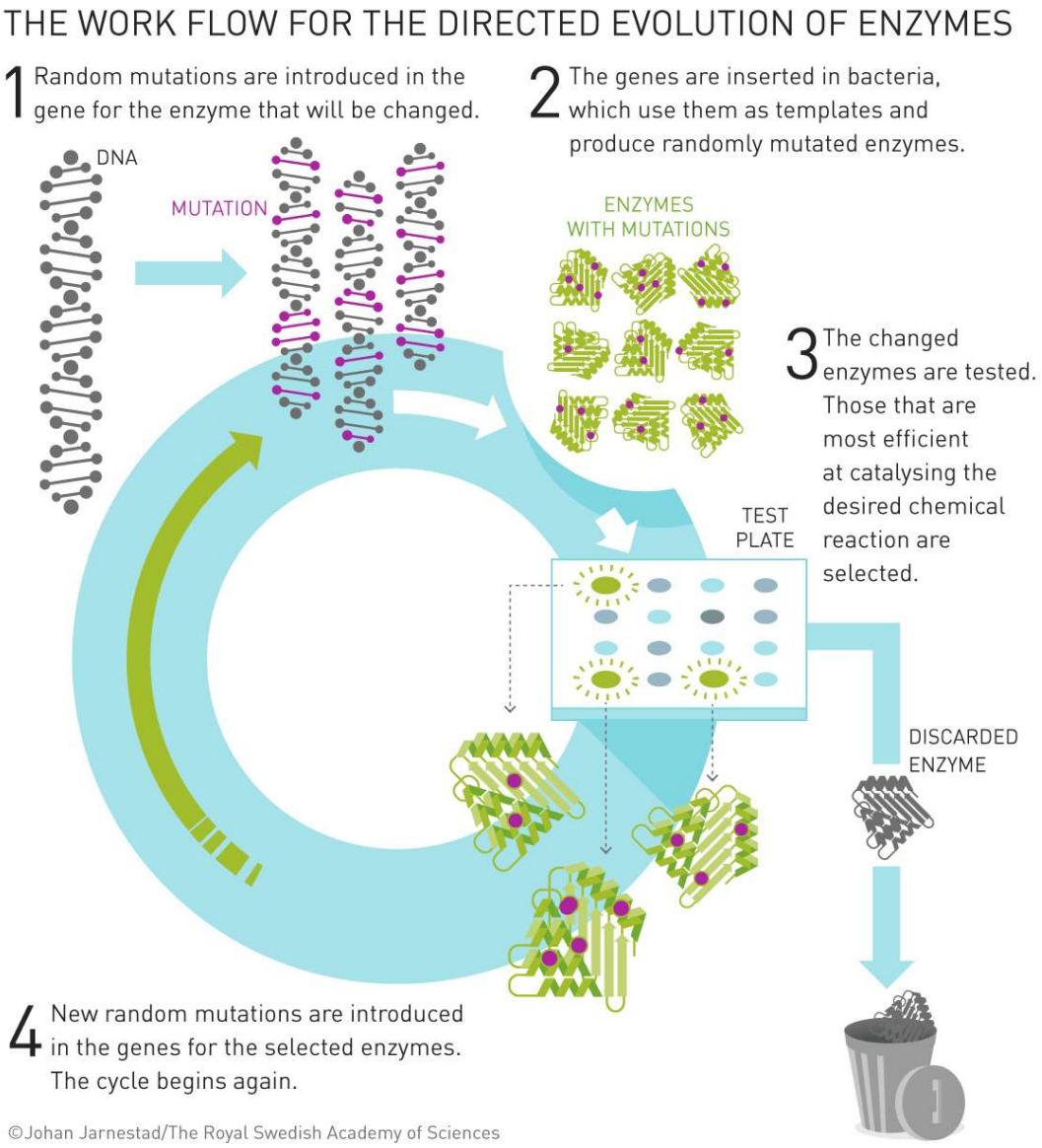
Alternative enzyme options result in different pinocembrin titers

Pathway optimization could involve optimizing the enzyme levels or the actual enzymes

Figure from Lee et al. Nature Catalysis 2,18–33(2019) but data from Feher, T. et al. Biotechnol. J. 9, 1446–1457 (2014).

Frances H. Arnold received the Noble prize for directed evolution of proteins in 2018

https://www.quantamagazine.org/frances-arnold-george-smith-and-gregory-winter-win-chemistry-nobel-for-directing-evolution-20181003/



Novel protein design is coming within reach

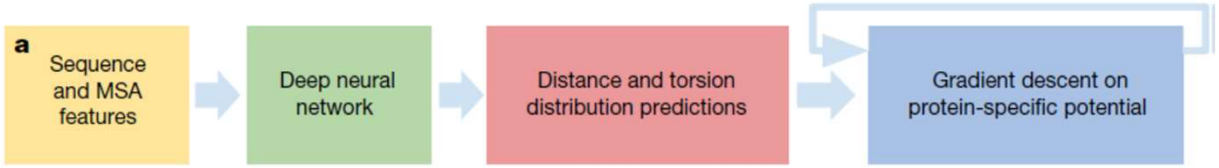
AlphaFold by DeepMind is a breakthrough in natural protein folding prediction

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]



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

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