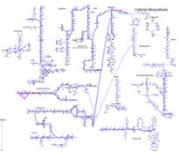


Strain Design

"What I cannot create, I do not understand"?

- Richard Feynman



Learning Objectives

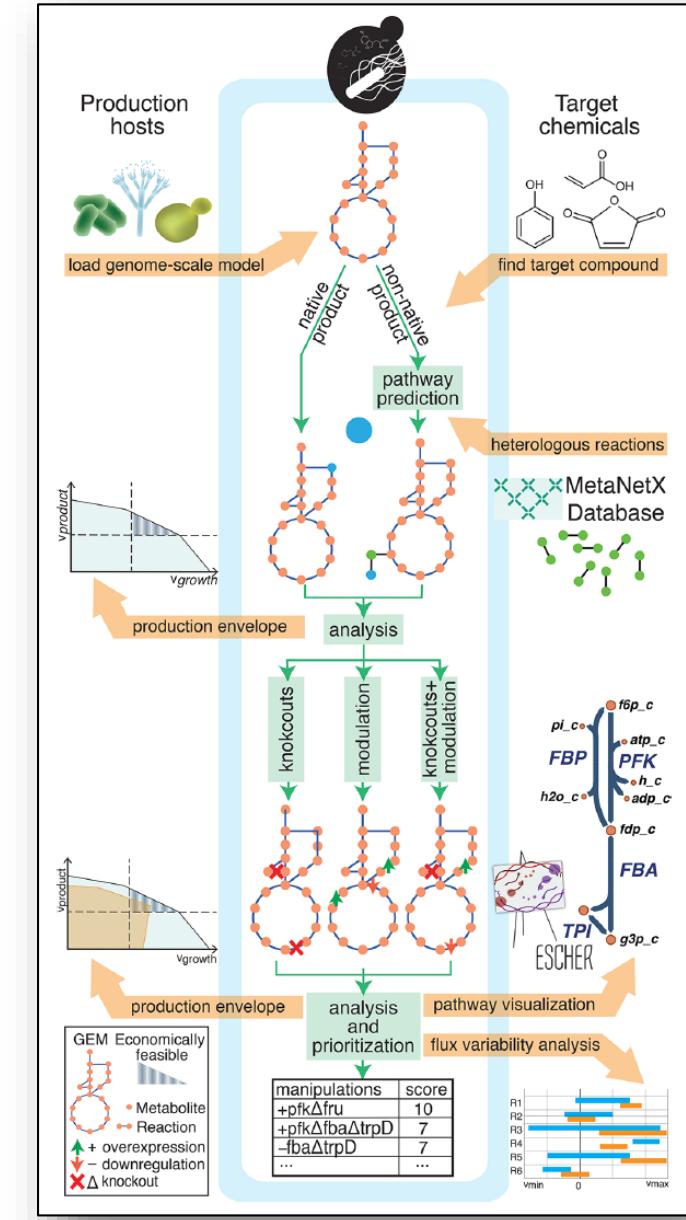
Each student should be able to:

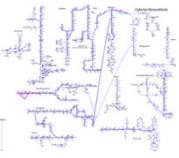
- Explain the strain design process.
- Explain how to model heterologous pathways.
- Explain how to model recombinant proteins.
- Explain the purpose of MOMA and ROOM
- Explain the purpose of the line of optimality
- Explain adaptive laboratory evolution.



Strain Design

- Introduction
- Production Hosts
- Target Chemicals
- Heterologous Pathways
- Recombinant Proteins
- Gene/Reaction Knockouts
- Gene Expression Modulation
- Adaptive Laboratory Evolution





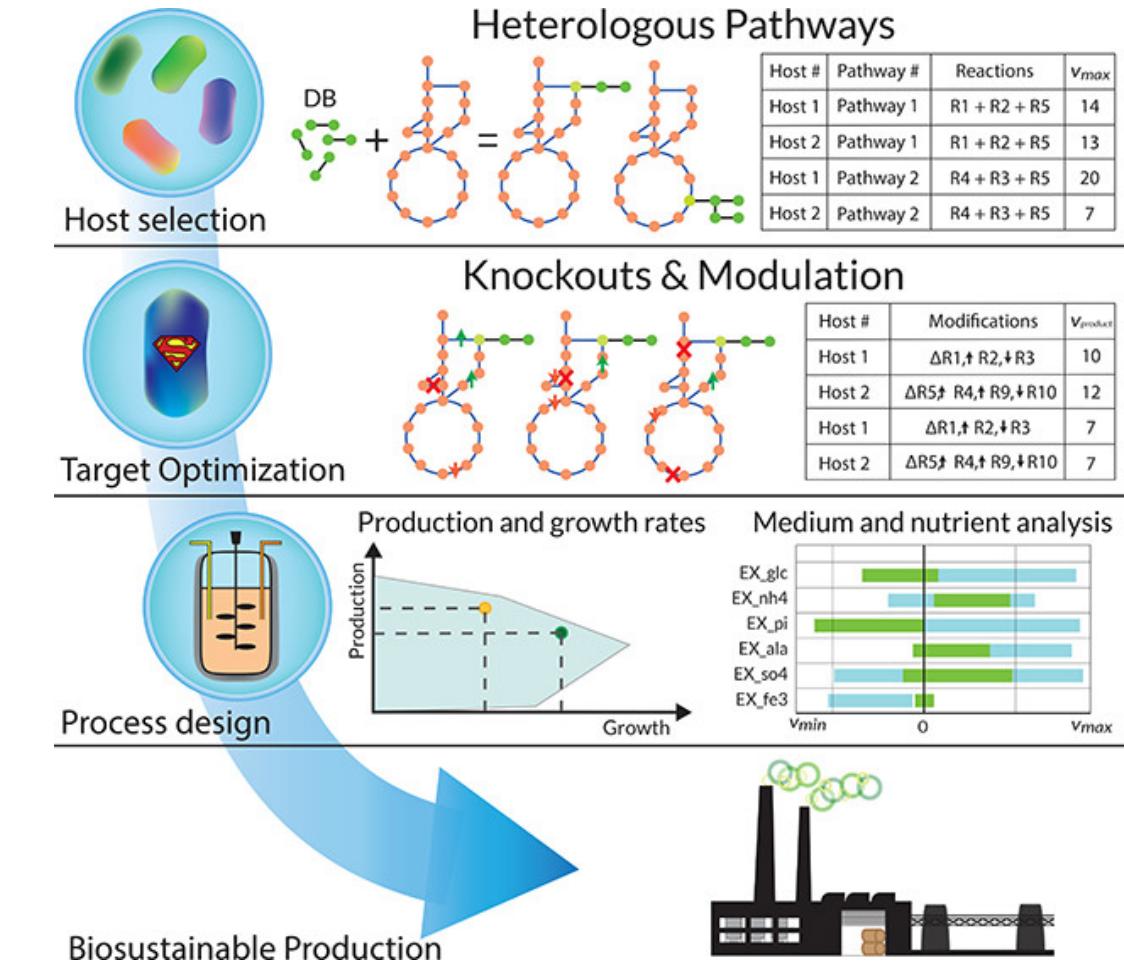
Strain Design

Heterologous Pathways

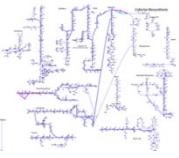
Heterologous pathways are linked series of biochemical reactions occurring in a host organism after the introduction of foreign genes. Incorporation of metabolic pathways into host organisms is a major strategy used to increase the production of valuable secondary metabolites.

Recombinant Proteins

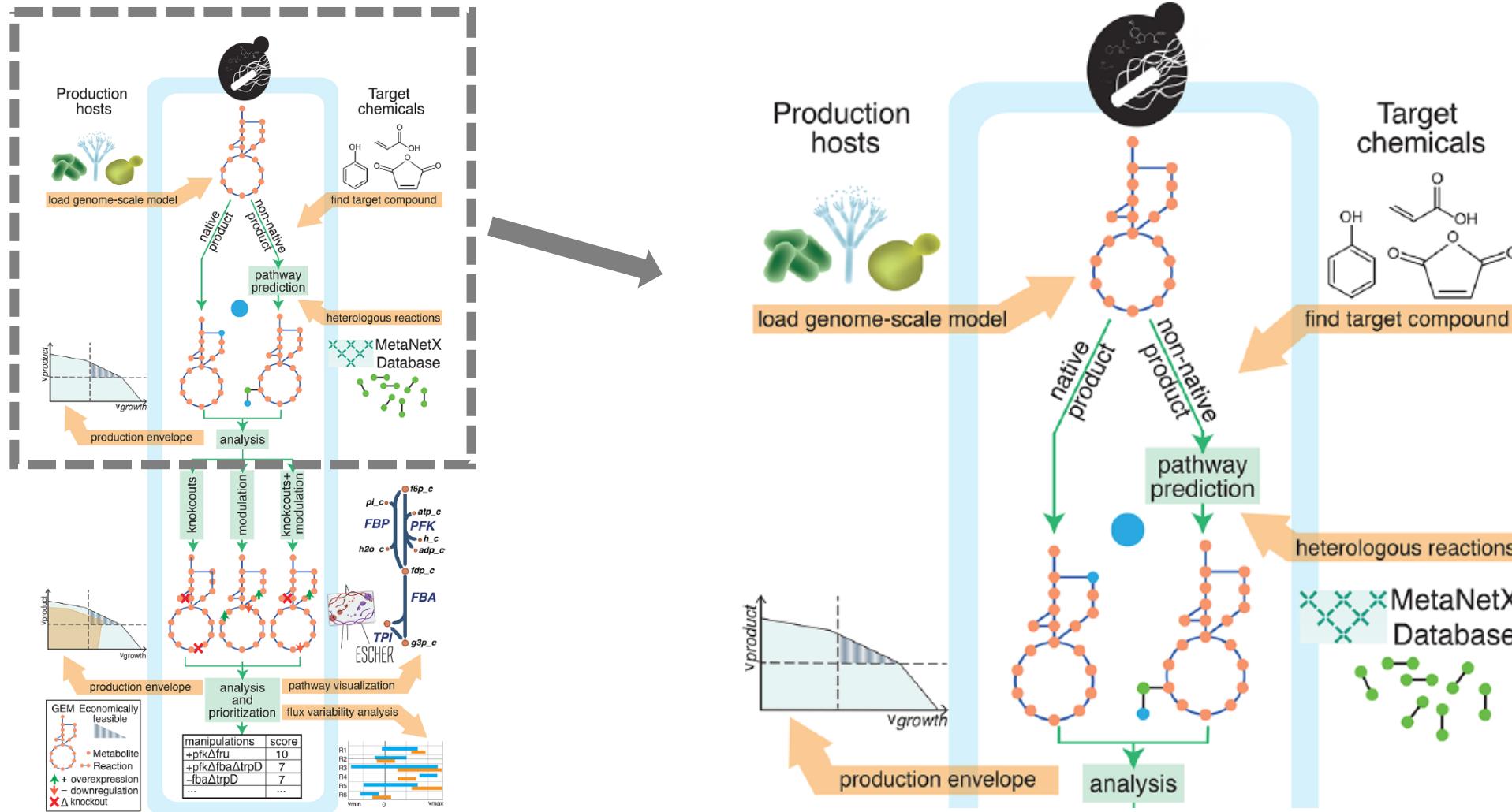
Recombinant proteins are a manipulated form of protein, generated in various ways to produce large quantities of proteins.



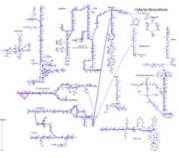
Cardoso, João GR, et al. "Cameo: a Python library for computer aided metabolic engineering and optimization of cell factories." *ACS synthetic biology* 7.4 (2018): 1163-1166.



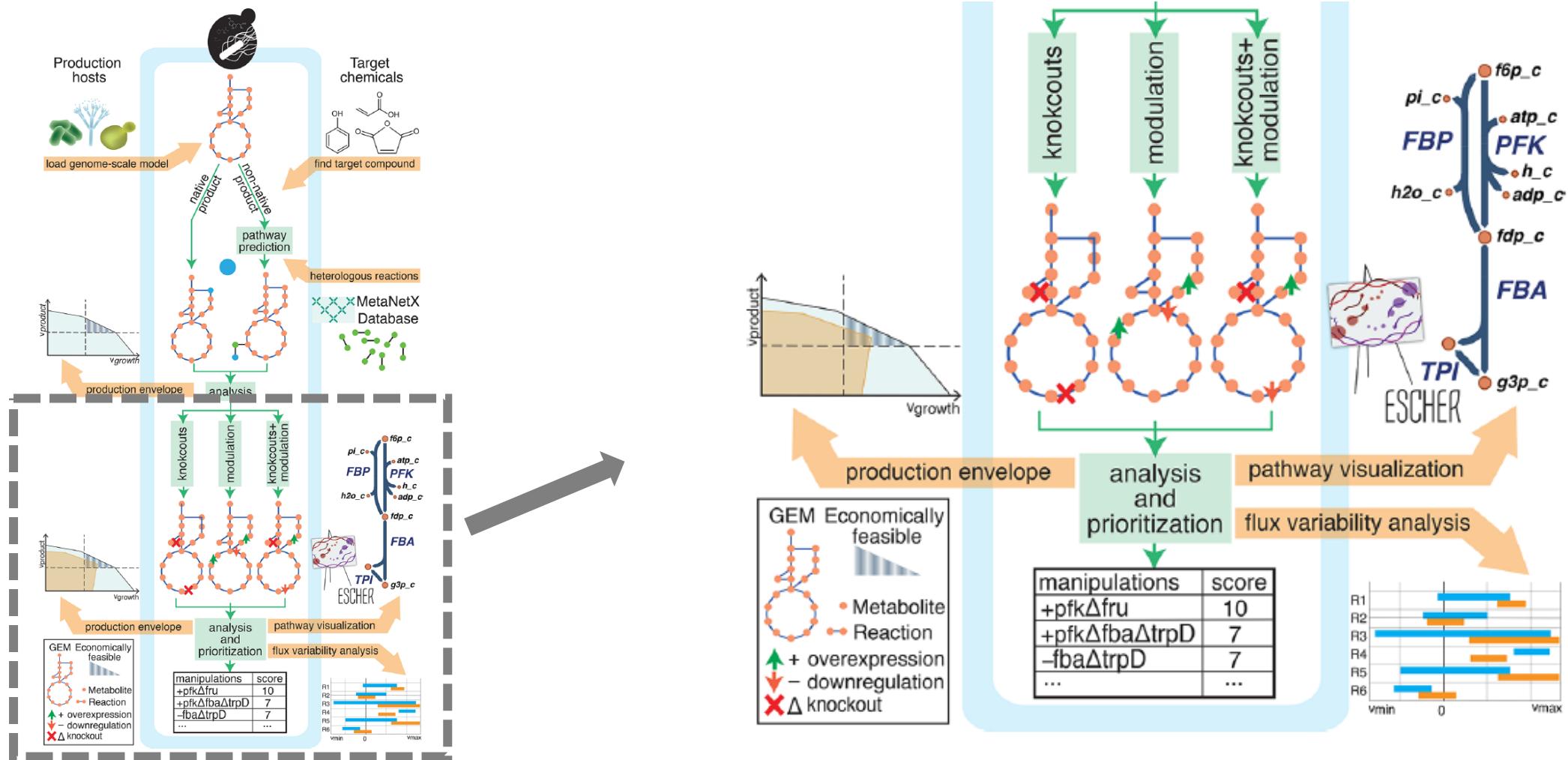
Strain Design Process



Cardoso, João GR, et al. "Cameo: a Python library for computer aided metabolic engineering and optimization of cell factories." *ACS synthetic biology* 7.4 (2018): 1163-1166.



Strain Design Process

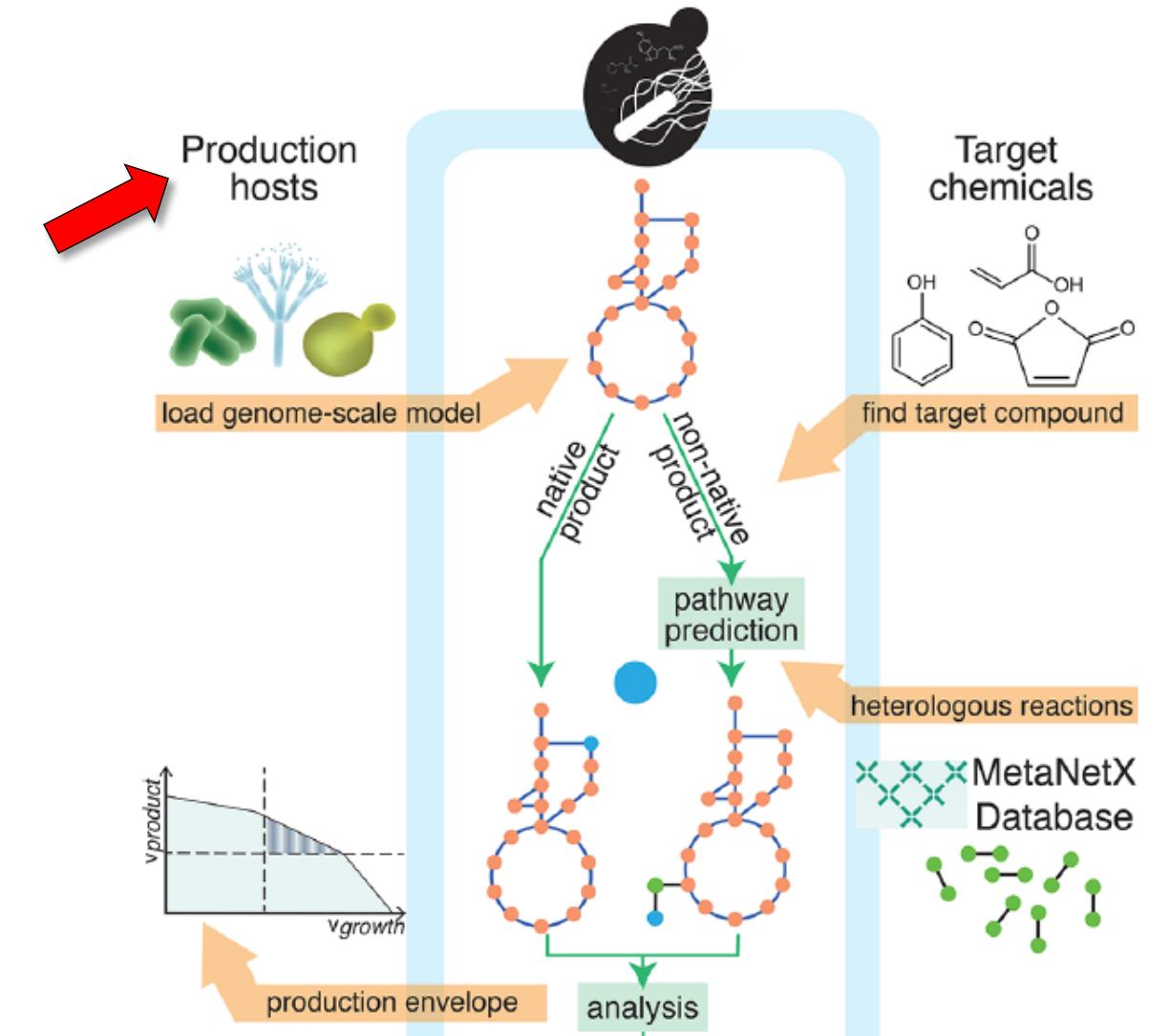


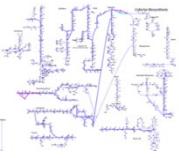
Cardoso, João GR, et al. "Cameo: a Python library for computer aided metabolic engineering and optimization of cell factories." *ACS synthetic biology* 7.4 (2018): 1163-1166.



Strain Design

- Introduction
- Production Hosts
- Target Chemicals
- Heterologous Pathways
- Recombinant Proteins
- Gene/Reaction Knockouts
- Gene Expression Modulation
- Adaptive Laboratory Evolution





E.coli as a Host Organism

1. Unparalleled fast growth kinetics. In glucose-salts media and given the optimal environmental conditions, its doubling time is about 20 min. This means that a culture inoculated with a 1/100 dilution of a saturated starter culture may reach stationary phase in a few hours. However, it should be noted that the expression of a recombinant protein may impart a metabolic burden on the microorganism, causing a considerable decrease in generation time.
2. High cell density cultures are easily achieved. The theoretical density limit of an *E. coli* liquid culture is estimated to be about 200 gram dry cell weight/liter or roughly 10^{13} viable bacteria/ml. However, exponential growth in complex media leads to densities nowhere near that number.
3. Rich complex media can be made from readily available and inexpensive components.
4. Transformation with exogenous DNA is fast and easy. Plasmid transformation of *E. coli* can be performed in as little as 5 min

Rosano, G. L. and E. A. Ceccarelli (2014). "Recombinant protein expression in Escherichia coli: advances and challenges." Frontiers in microbiology 5: 172.



BL21(DE3) *E.coli* Strain

- BL21 cells are deficient in the Lon protease which degrades many foreign proteins.
- Another missing gene is the one coding for the outer membrane protease OmpT, whose function is to degrade extracellular proteins. The liberated amino acids are then taken up by the cell. This is problematic in the expression of a recombinant protein as, after cell lysis, OmpT may digest it.
- Plasmid loss is prevented thanks to the *hsdSB* mutation which disrupts DNA methylation and degradation.
- When the genes for the desired recombinant protein placed under a T7 promoter, then T7 RNA Polymerase (RNAP) needs to be provided by the host cell. In the popular BL21(DE3) strain, the λ DE3 prophage was inserted in the chromosome of BL21 and contains the T7 RNAP gene under the lacUV5 promoter.
- The BL21(DE3) and its derivatives are by far the most used strains for protein expression.

Rosado, G. L. and E. A. Ceccarelli (2014). "Recombinant protein expression in *Escherichia coli*: advances and challenges." *Frontiers in microbiology* 5: 172.



Tested Cobra Toolbox Models

<http://bigg.ucsd.edu/models>

Model Number

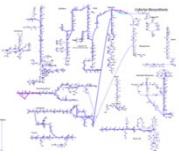
Model Number	Name
e_coli_core	Escherichia coli str. K-12 substr. MG1655
iAB_RBC_283	Homo sapiens
iAF1260	Escherichia coli str. K-12 substr. MG1655
iAF692	Methanoscarcina barkeri str. Fusaro
iAF987	Geobacter metallireducens GS-15
iAPEC01_1312	Escherichia coli APEC O1
iAT_PLT_636	Homo sapiens
iB21_1397	Escherichia coli BL21(DE3)
iBWG_1329	Escherichia coli BW2952
ic_1306	Escherichia coli CFT073
iE2348C_1286	Escherichia coli O127:H6 str. E2348/69
iEC042_1314	Escherichia coli 042
iEC55989_1330	Escherichia coli 55989
iECABU_c1320	Escherichia coli ABU 83972
iECB_1328	Escherichia coli B str. REL606
iECBD_1354	Escherichia coli 'BL21-Gold(DE3)pLysS AG'
iECD_1391	Escherichia coli BL21(DE3)
iECDH10B_1368	Escherichia coli str. K-12 substr. DH10B
iEcDH1_1363	Escherichia coli DH1
iECDH1ME8569_1439	Escherichia coli DH1
iEcE24377_1341	Escherichia coli E24377A
iECED1_1282	Escherichia coli ED1a
iECH74115_1262	Escherichia coli O157:H7 str. EC4115
iEcHS_1320	Escherichia coli HS
iECIAI1_1343	Escherichia coli IAI1
iECIAI39_1322	Escherichia coli IAI39

Model Number

Model Number	Name
iECNA114_1301	Escherichia coli NA114
iECO103_1326	Escherichia coli O103:H2 str. 12009
iECO111_1330	Escherichia coli O111:H- str. 11128
iECO26_1355	Escherichia coli O26:H11 str. 11368
iECOK1_1307	Escherichia coli IHE3034
iEcoliC_1368	Escherichia coli ATCC 8739
iECP_1309	Escherichia coli 536
iECs_1301	Escherichia coli O157:H7 str. Sakai
iECS88_1305	Escherichia coli S88
iECSE_1348	Escherichia coli SE11
iECSF_1327	Escherichia coli SE15
iEcSMS35_1347	Escherichia coli SMS-3-5
iECSP_1301	Escherichia coli O157:H7 str. TW14359
iECUMN_1333	Escherichia coli UMN026
iECW_1372	Escherichia coli W
iEKO11_1354	Escherichia coli KO11FL
iETEC_1333	Escherichia coli ETEC H10407
iG2583_1286	Escherichia coli O55:H7 str. CB9615
iHN637	Clostridium ljungdahlii DSM 13528
iIT341	Helicobacter pylori 26695
iJN678	Synechocystis sp. PCC 6803
iJN746	Pseudomonas putida KT2440
iJO1366	Escherichia coli str. K-12 substr. MG1655
iJR904	Escherichia coli str. K-12 substr. MG1655
iLF82_1304	Escherichia coli LF82
iLJ478	Thermotoga maritima MSB8

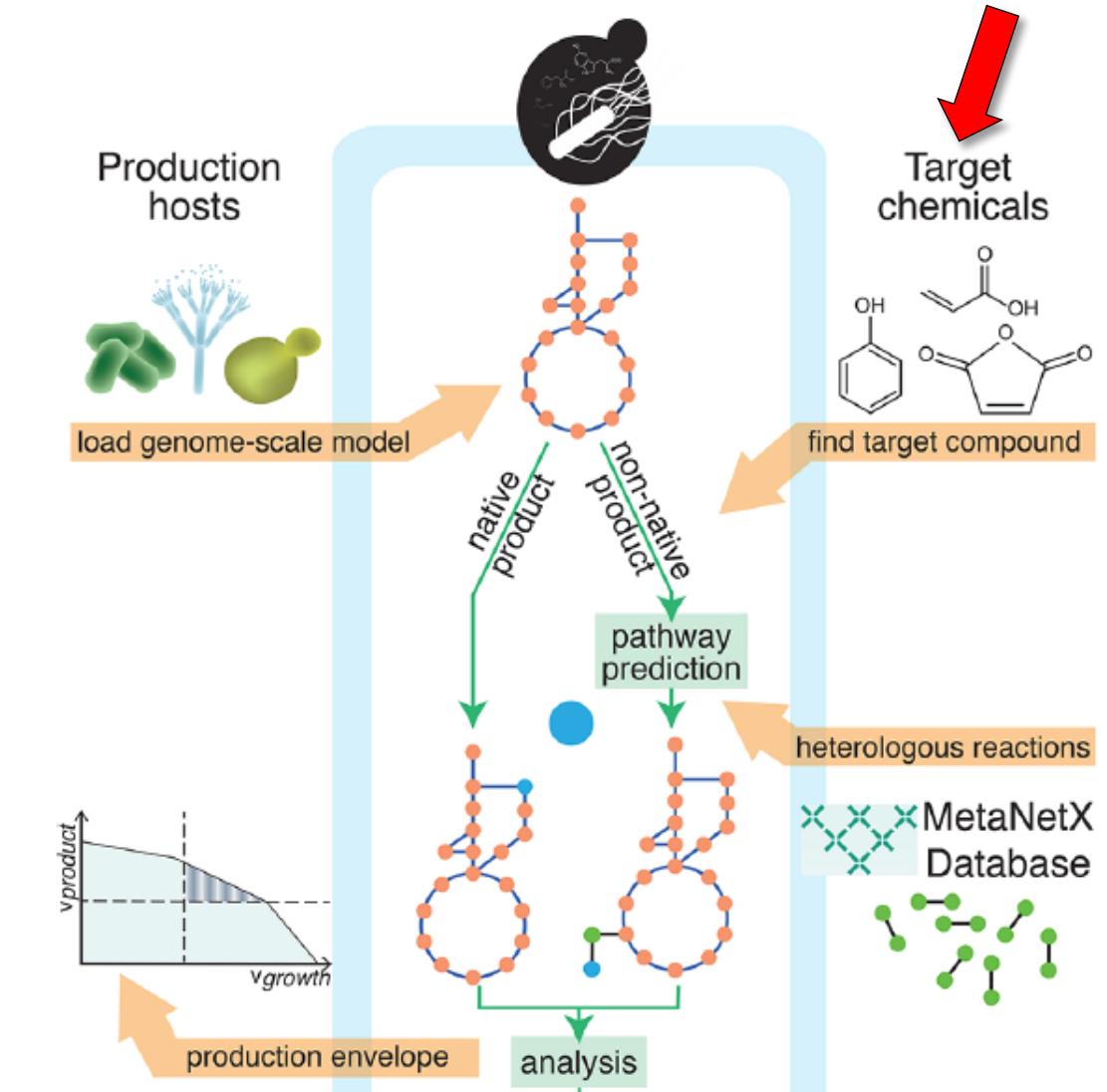
Model Number

Model Number	Name
iMM1415	Mus musculus
iMM904	Saccharomyces cerevisiae S288c
iND750	Saccharomyces cerevisiae S288c
iNJ661	Mycobacterium tuberculosis H37Rv
iNRG857_1313	Escherichia coli O83:H1 str. NRG 857C
iPC815	Yersinia pestis CO92
iRC1080	Chlamydomonas reinhardtii
iS_1188	Shigella flexneri 2a str. 2457T
iSB619	Staphylococcus aureus subsp. aureus N315
iSbBS512_1146	Shigella boydii CDC 3083-94
iSBO_1134	Shigella boydii Sb227
iSDY_1059	Shigella dysenteriae Sd197
iSF_1195	Shigella flexneri 2a str. 301
iSFV_1184	Shigella flexneri 5 str. 8401
iSFxv_1172	Shigella flexneri 2002017
iSSON_1240	Shigella sonnei Ss046
iUMN146_1321	Escherichia coli UM146
iUMNK88_1353	Escherichia coli UMNK88
iUTI89_1310	Escherichia coli UTI89
iWFL_1372	Escherichia coli W
iY75_1357	Escherichia coli str. K-12 substr. W3110
iYL1228	Klebsiella pneumoniae subsp. pneumoniae MGH 78578
iYO844	Bacillus subtilis subsp. subtilis str. 168
iZ_1308	Escherichia coli O157:H7 str. EDL933
RECON3D	Homo sapiens
STM_v1_0	Salmonella enterica subsp. enterica serovar Typhimurium str. LT2



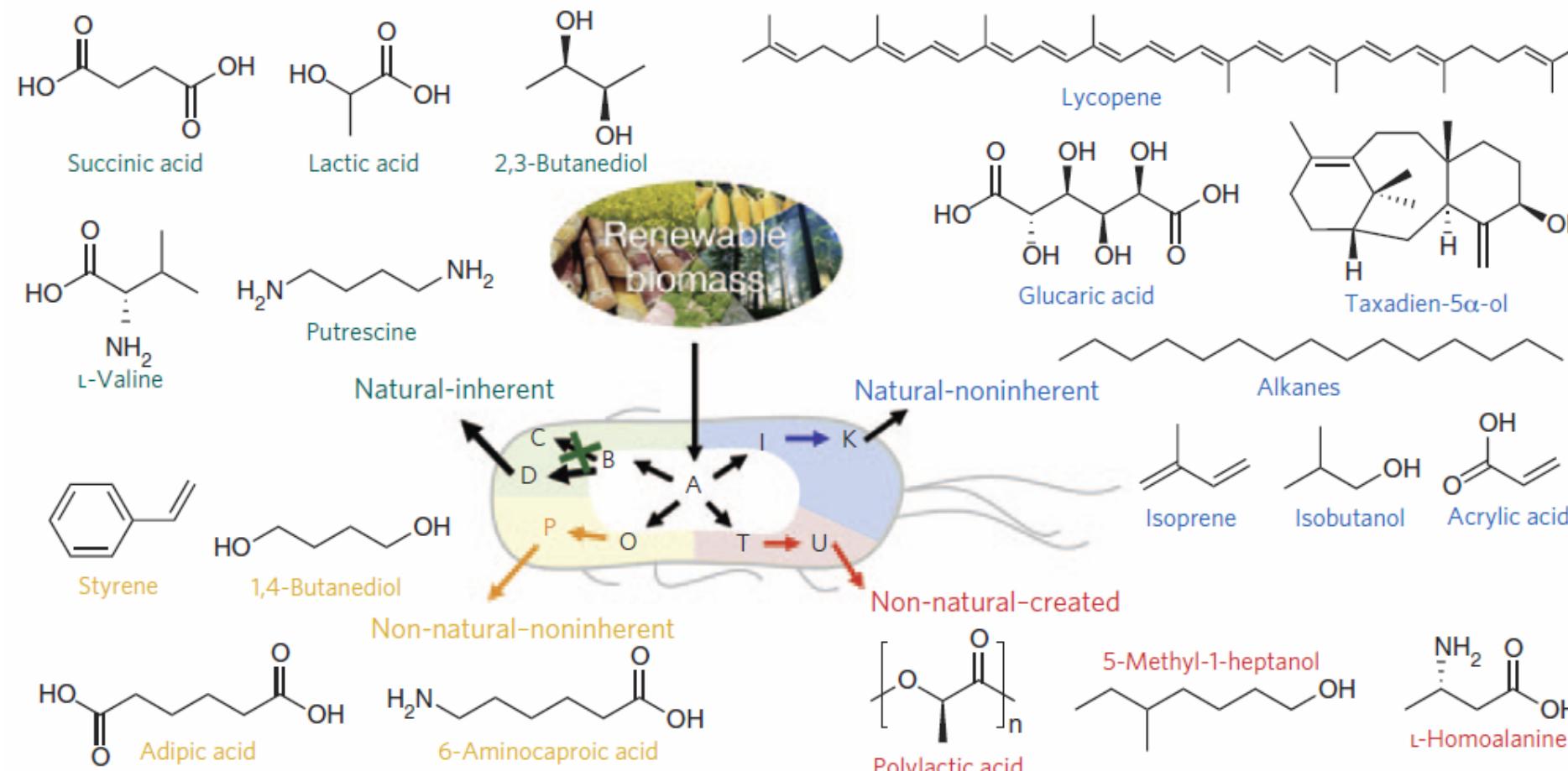
Strain Design

- Introduction
- Production Hosts
- • Target Chemicals
- Heterologous Pathways
- Recombinant Proteins
- Gene/Reaction Knockouts
- Gene Expression Modulation
- Adaptive Laboratory Evolution





Categories of Chemicals Produced by Microbial Cell Factories



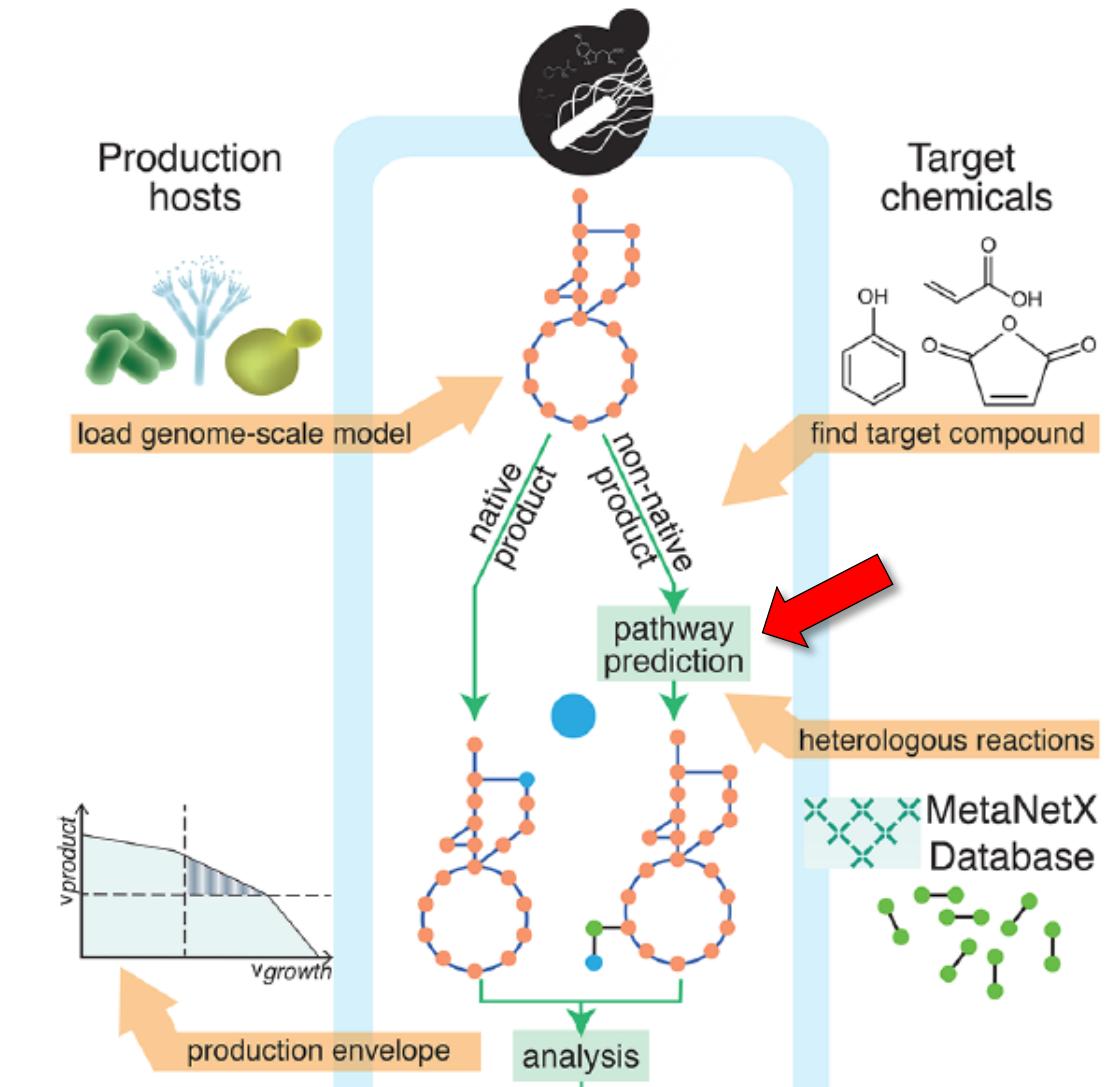
- **Natural-inherent chemicals** are endogenous metabolites in naturally isolated microorganisms and thus can be produced inherently through a native pathway.
- **Natural-noninherent chemicals** are those that are found in nature but are best produced in a heterologous host strain using noninherent pathways introduced from other hosts or metagenomes.
- **Non-natural–noninherent** chemicals are those that have not yet been found in nature but can be produced in a noninherent host strain by the establishment of heterologous pathways and enzymes, often using genes found from various sources in combination.
- **Nonnatural–created chemicals** are those that have not yet been found in nature and, owing to the lack of any known metabolic enzymes and pathways leading to their formation, can only be produced by creating synthetic enzymes and pathways with new functions.

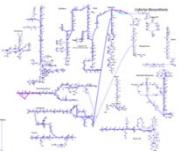
Lee, J. W., D. Na, et al. (2012). "Systems metabolic engineering of microorganisms for natural and non-natural chemicals." *Nature chemical biology* 8(6): 536-546.



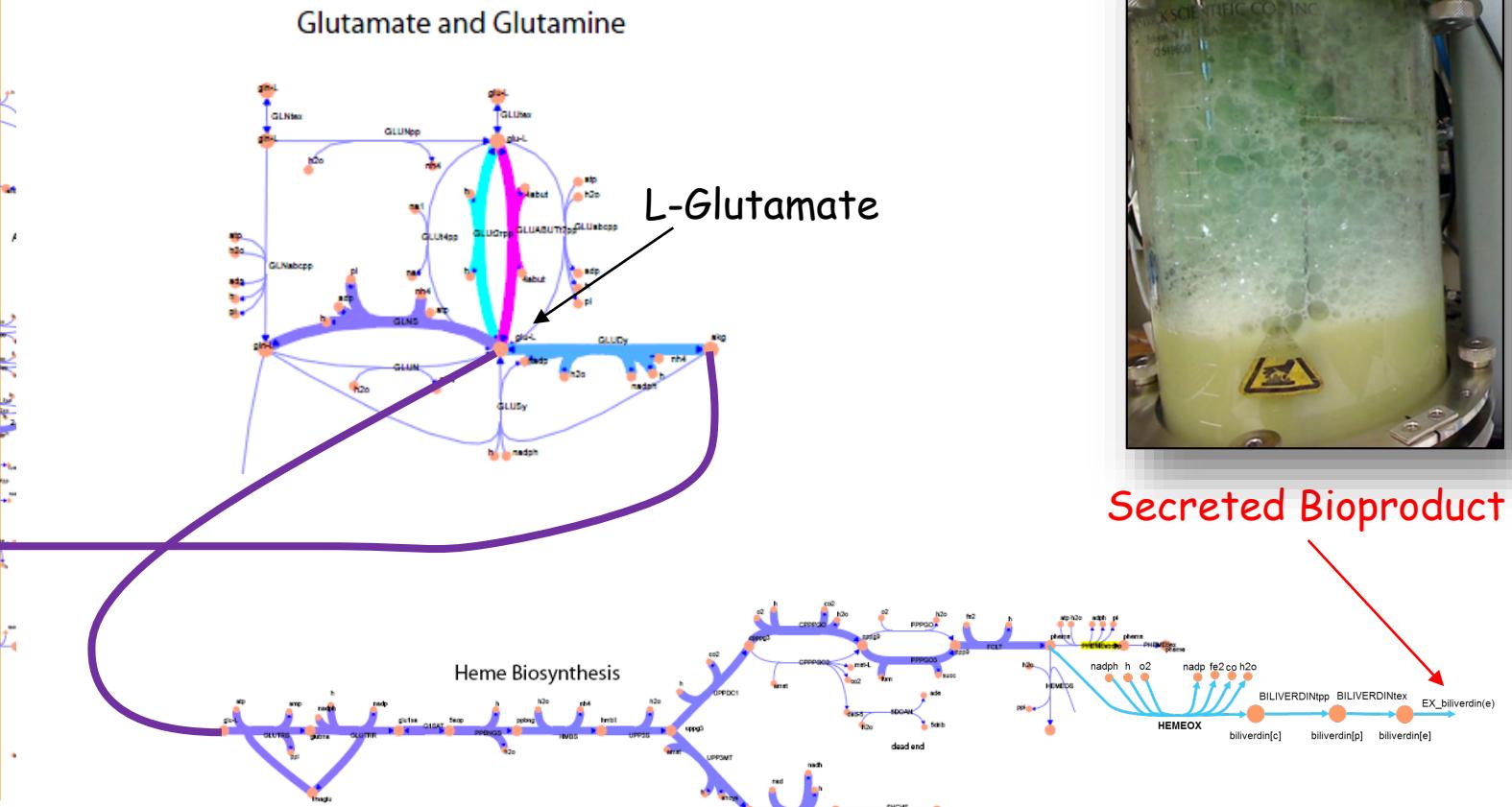
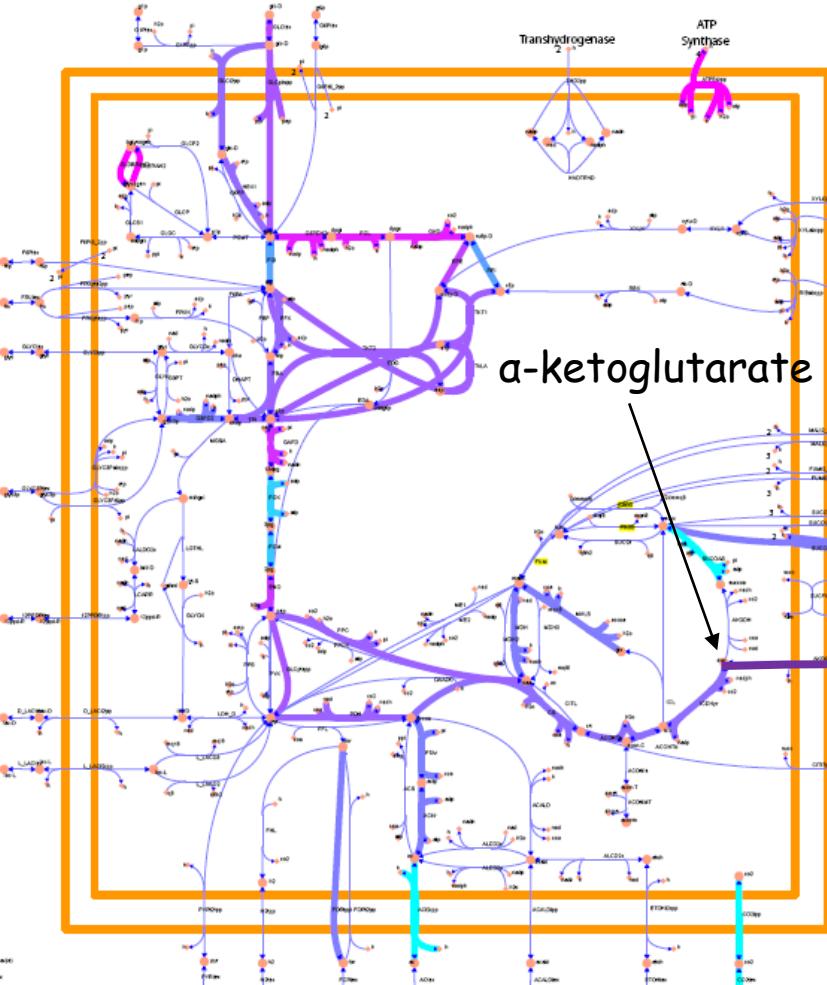
Strain Design

- Introduction
- Production Hosts
- Target Chemicals
- **Heterologous Pathways**
- Recombinant Proteins
- Gene/Reaction Knockouts
- Gene Expression Modulation
- Adaptive Laboratory Evolution





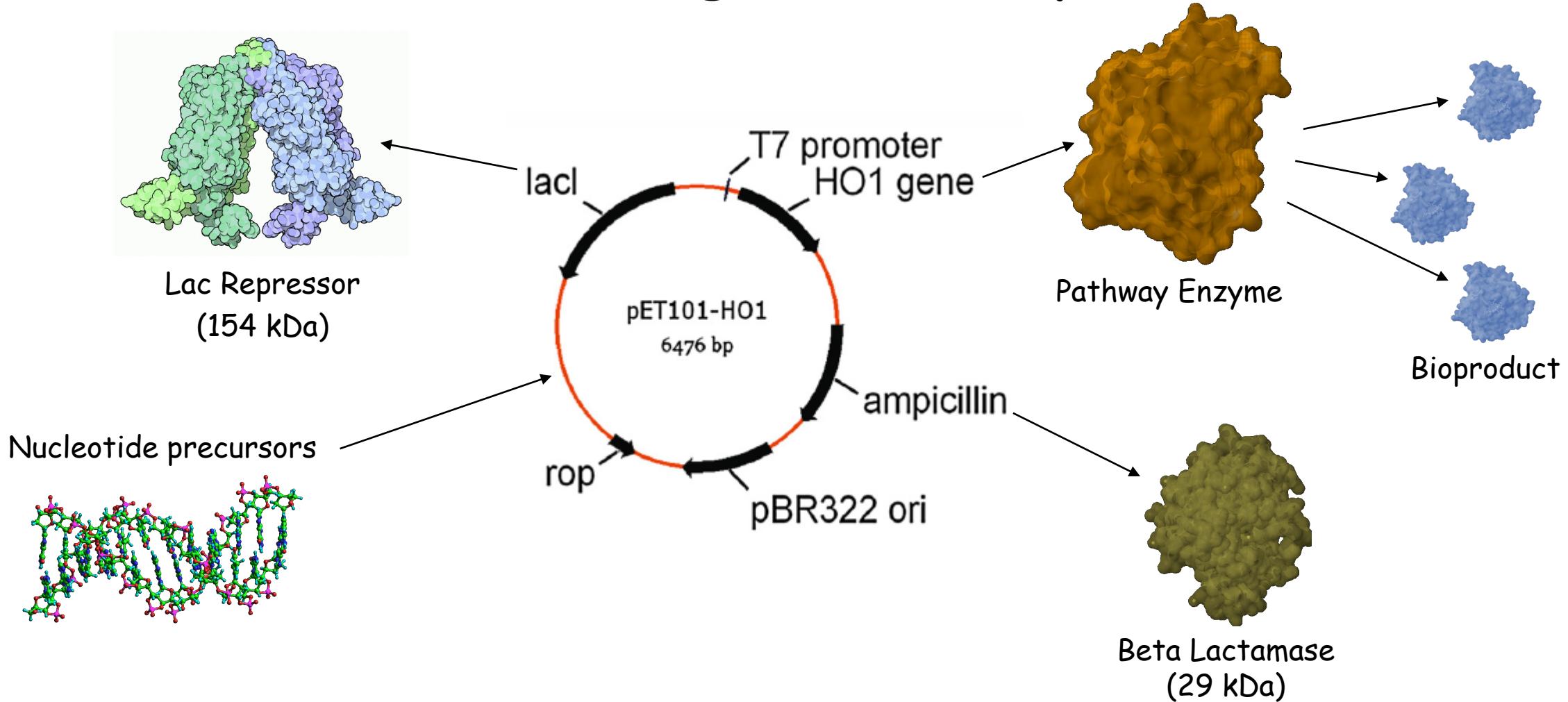
Defining the Pathway of the Desired Bioproduct



Secreted Bioproduct



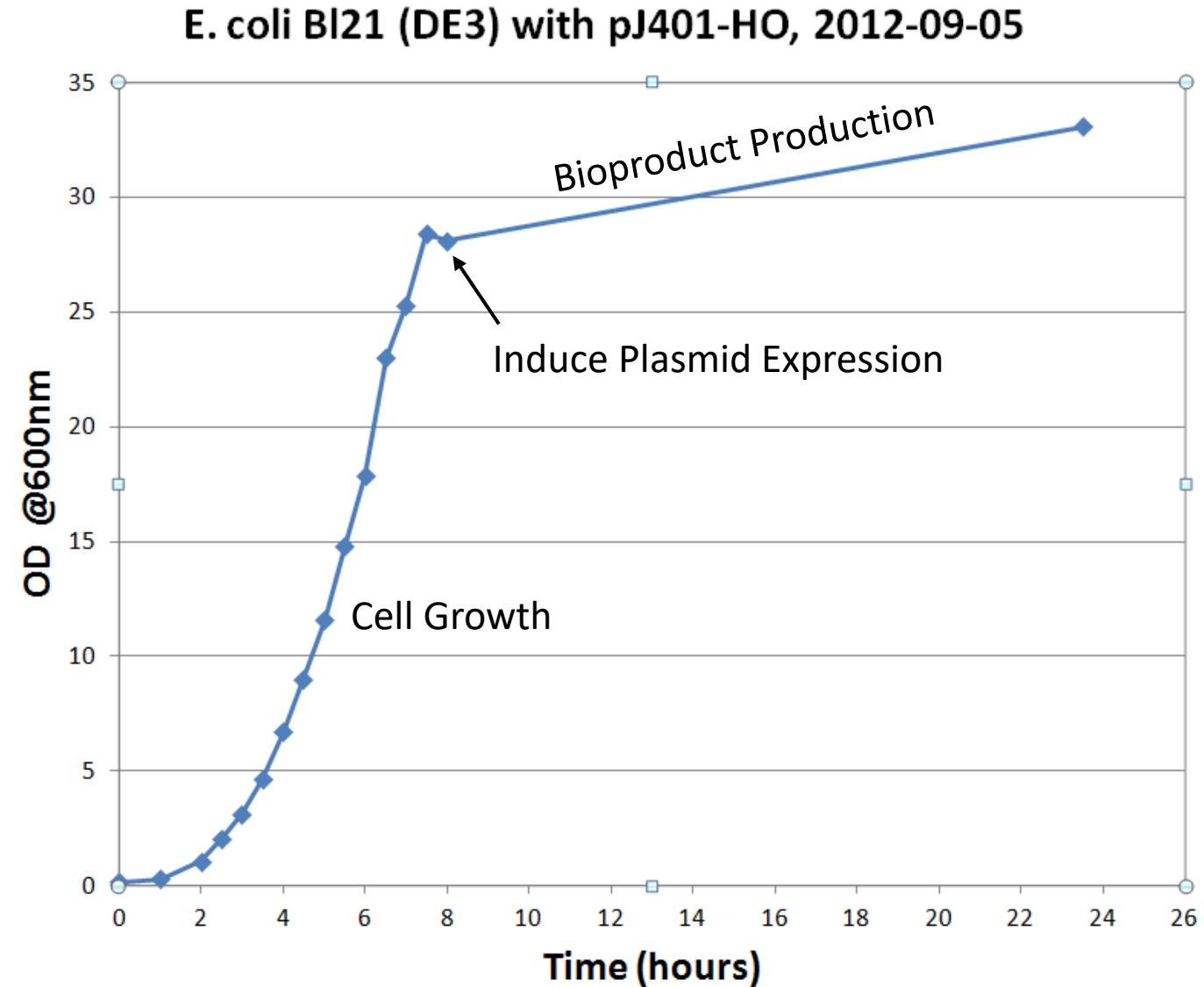
Components Produced by Plasmid for Heterologous Pathways





Bioproduct Production (Batch Process)

- The process starts by exclusively growing the biomass. The biomass grows faster without the resource load produced by the plasmids.
- The biomass increases until it reaches a chosen optical density (ex. OD₆₀₀ = 28).
- The plasmid is then induced (ex. IPTG) and then cell starts to produce the desired bioproduct.
- Since the cells resources are now used to create bioproduct instead of biomass the growth rate drops significantly.





Pathway Prediction

- Predicting heterologous pathways is an important strategy to generate new viable strains.
- This process includes the need to identify reactions that are not inherent to the host strain.
- Because total number of available reactions is very large, computer assisted pathway design becomes essential.
- Cameo implements a shortest pathways search algorithm using an universal biochemical reaction database.
 - ✓ <https://www.metanetx.org/>
- Pathway prediction
 - ✓ [Predict heterologous pathways — cameo 0.13.6 documentation](#)

Pathway 1			
	equation	lower_bound	upper_bound
MNXR81653	butanol + NAD(+) \rightleftharpoons H(+) + NADH + butanal	-1000.0	1000.0
MNXR5687	butanal + CoA + NAD(+) \rightleftharpoons H(+) + NADH + butan...	-1000.0	1000.0
Pathway 2			
	equation	lower_bound	upper_bound
MNXR81654	butanol + NADP(+) \rightleftharpoons H(+) + butanal + NADPH	-1000.0	1000.0
MNXR5687	butanal + CoA + NAD(+) \rightleftharpoons H(+) + NADH + butan...	-1000.0	1000.0
Pathway 3			
	equation	lower_bound	upper_bound
MNXR660	butanal + H2O + NAD(+) \rightleftharpoons 2.0 H(+) + NADH + b...	-1000.0	1000.0
MNXR2039	diphosphate + AMP + butanoyl-CoA \rightleftharpoons CoA + ATP...	-1000.0	1000.0
MNXR81654	butanol + NADP(+) \rightleftharpoons H(+) + butanal + NADPH	-1000.0	1000.0
Pathway 4			
	equation	lower_bound	upper_bound
MNXR660	butanal + H2O + NAD(+) \rightleftharpoons 2.0 H(+) + NADH + b...	-1000.0	1000.0
MNXR81654	butanol + NADP(+) \rightleftharpoons H(+) + butanal + NADPH	-1000.0	1000.0
MNXR980	acetyl-CoA + butanoate \rightleftharpoons butanoyl-CoA + acetate	-1000.0	1000.0





MetaNetX

Automated Model Construction and Genome Annotation for Large-Scale Metabolic Networks

<https://www.metanetx.org/>

Search MNXref

SystemsX.ch
The Swiss Initiative in Systems Biology

My Selection

- Summary
- Pick from repository
- Import model
- Upload reactions
- Delete models
- Upload genome
- Revived!

Analyze

- Flux balance (FBA)
- Groups of coupled reactions (GCR)
- Blocked reactions (BLO)
- Reaction knockout (RKO)
- Gene/peptide knockout (PKO)

Create / Modify

- Combine logically
- Split and merge
- Growth recovery (GRE)
- Build from a genome (BUILD) Revived!

Utilities

- Search/Download MNXref namespace
- MNXref ID mapper

Getting started

MetaNetX.org is an online platform for accessing, analyzing and manipulating genome-scale metabolic networks (GSM) as well as biochemical pathways. To this end, it integrates a great variety of data sources and tools.

The metabolic models

MetaNetX.org treats each model (i.e., metabolic network or pathway) as an *object* constituted by different entities:

- chemical compounds (*chem*)
- subcellular compartments (*comp*)
- species (*spec*): chemical compounds that are assigned to a subcellular compartment
- metabolic reactions (*reac*): reactions that transform species into another
- genes or peptides (*pept*)
- enzymes (*enzy*): sets of peptides (or genes) linked to a reaction with information on bounds (maximal and minimal bounds) defining the directionality such that the maximum flux can be carried by this reaction with these enzymes

Depending on the studied model (GSM or pathway), it may contain biomass production reaction(s) (identified by the identifier "*BIOMASS*" in the corresponding reaction equation), or uptake or secretion reactions (external/boundary reactions; identified by chemical species associated with the artificial "*BOUNDARY*" compartment).

Import/Export

Select a model from the repository

MetaNetX.org includes a repository of models from, for example, [BiGG](#) or the [Model SEED](#), networks from [BioCyc](#), and metabolic pathways from [Reactome](#); all reconciled in a common namespace defined by [MNXref](#).

The "Pick from repository" menu allows you to select one or more models from this repository. It is possible to search for specific models using the Search field. Search criteria can comprise taxonomy, lineage (that can be displayed using the show/hide button), or identifiers.



Constraint-based Metabolic Reconstructions & Analysis

H. Scott Hinton, 2022

19

MetaNetX
Automated Model Construction and Genome Annotation for Large-Scale Metabolic Networks

My Selection

- Summary
- Pick from repository
- Import model
- Upload reactions
- Delete models
- Upload genome
- Revived!

Analyze

- Flux balance (FBA)
- Groups of coupled reactions (GCR)
- Blocked reactions (BLO)
- Reaction knockout (RKO)
- Gene/peptide knockout (PKO)

Create / Modify

- Combine logically
- Split and merge
- Growth recovery (GRE)
- Build from a genome (BUILD) Revived!

Utilities

- Search/Download MNXref namespace
- MNXref ID mapper new
- SPARQL query
- Reset session

Documents

Search MNXref
SystemsX.ch The Swiss Initiative in Systems Biology

Compounds
No result found

Reactions

Filter: First Previous 1 Next Last

MNX reac id	MNX description
MNXR146007	$1 \text{ MNXM1104209@MNXD1} + 1 \text{ MNXM11@MNXD1} = 2 \text{ MNXM728266@MNXD1}$
MNXR146008	$1 \text{ MNXM1107976@MNXD1} + 1 \text{ MNXM11@MNXD1} = 3 \text{ MNXM1@MNXD1} + 2 \text{ MNXM728266@MNXD1}$
MNXR146009	$1 \text{ MNXM11@MNXD1} + 1 \text{ MNXM731557@MNXD1} = 2 \text{ MNXM728266@MNXD1}$
MNXR146006	$1 \text{ MNXM11@MNXD1} + 1 \text{ MNXM731559@MNXD1} = 2 \text{ MNXM728266@MNXD1}$

Showing 1 to 4 of 4 entries

Compartments
No result found

MetaNetX
Automated Model Construction and Genome Annotation for Large-Scale Metabolic Networks

Search MNXref
SystemsX.ch The Swiss Initiative in Systems Biology

My Selection

- Summary
- Pick from repository
- Import model
- Upload reactions
- Delete models
- Upload genome
- Revived!

Analyze

- Flux balance (FBA)
- Groups of coupled reactions (GCR)
- Blocked reactions (BLO)
- Reaction knockout (RKO)
- Gene/peptide knockout (PKO)

Create / Modify

- Combine logically
- Split and merge
- Growth recovery (GRE)
- Build from a genome (BUILD) Revived!

Feedback

Deprecated MNXref IDs graph

MNXR146006

Properties

MNX_ID	MNXR146006
equation	$1 \text{ diphosphate} + 1 \text{ Prephytoene diphosphate generic compartment 1} \iff 2 \text{ (2E,6E,10E)-geranylgeranyl diphosphate generic compartment 1}$ $1 \text{ MNXM11@MNXD1} + 1 \text{ MNXM731559@MNXD1} = 2 \text{ MNXM728266@MNXD1}$
is balanced?	
is transport?	
EC number	2.5.1.32 2.5.1.99
reference	biggR:PHYTEs Geranylgeranyl-diphosphate geranylgeranyl-diphosphate geranylgeranyltransferase

MNXR146006

Similar chemical equations in external resources

bigg.reaction:PHYTEs	biggR:PHYTEs	Geranylgeranyl-diphosphate geranylgeranyl-diphosphate geranylgeranyltransferase
		$1 \text{ biggM:ppi@biggC:c} + 1 \text{ biggM:prephytendp@biggC:c} = 2 \text{ biggM:gdp@biggC:c}$
bigg.reaction.R_PHYTEs	biggR:R_PHYTEs	secondary/obsolete/fantasy identifier



Lycopene Heterologous Pathways in *E. coli*

Lycopene is a natural compound that is found in bright red fruits and vegetables like tomatoes, watermelon, and grapefruit. Lycopene is a carotenoid, which are yellow, orange, or red pigments that give this color to its plants. Lycopene has been linked to many health benefits, including disease prevention and protection. (<https://www.verywellhealth.com/lycopene-health-benefits-4684446>)

Create a lycopene pathway in *E. coli* (iJO1366).

```
In [1]: from cameo import models
from cameo.strain_design import pathway_prediction
from IPython.display import display
import re
```

Load the iJO1366 *E. coli* model

```
In [2]: model = models.bigg.iJO1366
model.solver = 'glpk'

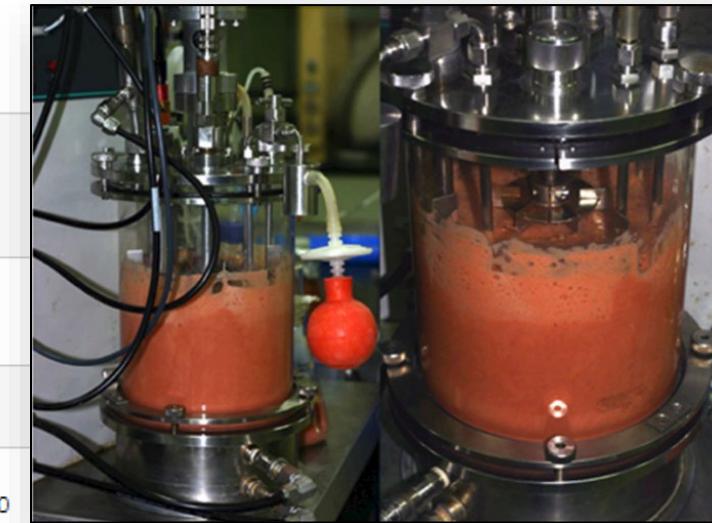
Set parameter Username
Academic license - for non-commercial use only - expires 2022-10-10
```

Identify potential reactions that can be used to create a lycopene pathway in the iJO1366 model.

```
In [3]: predictor = pathway_prediction.PathwayPredictor(model)
```

Listing the potential pathways.

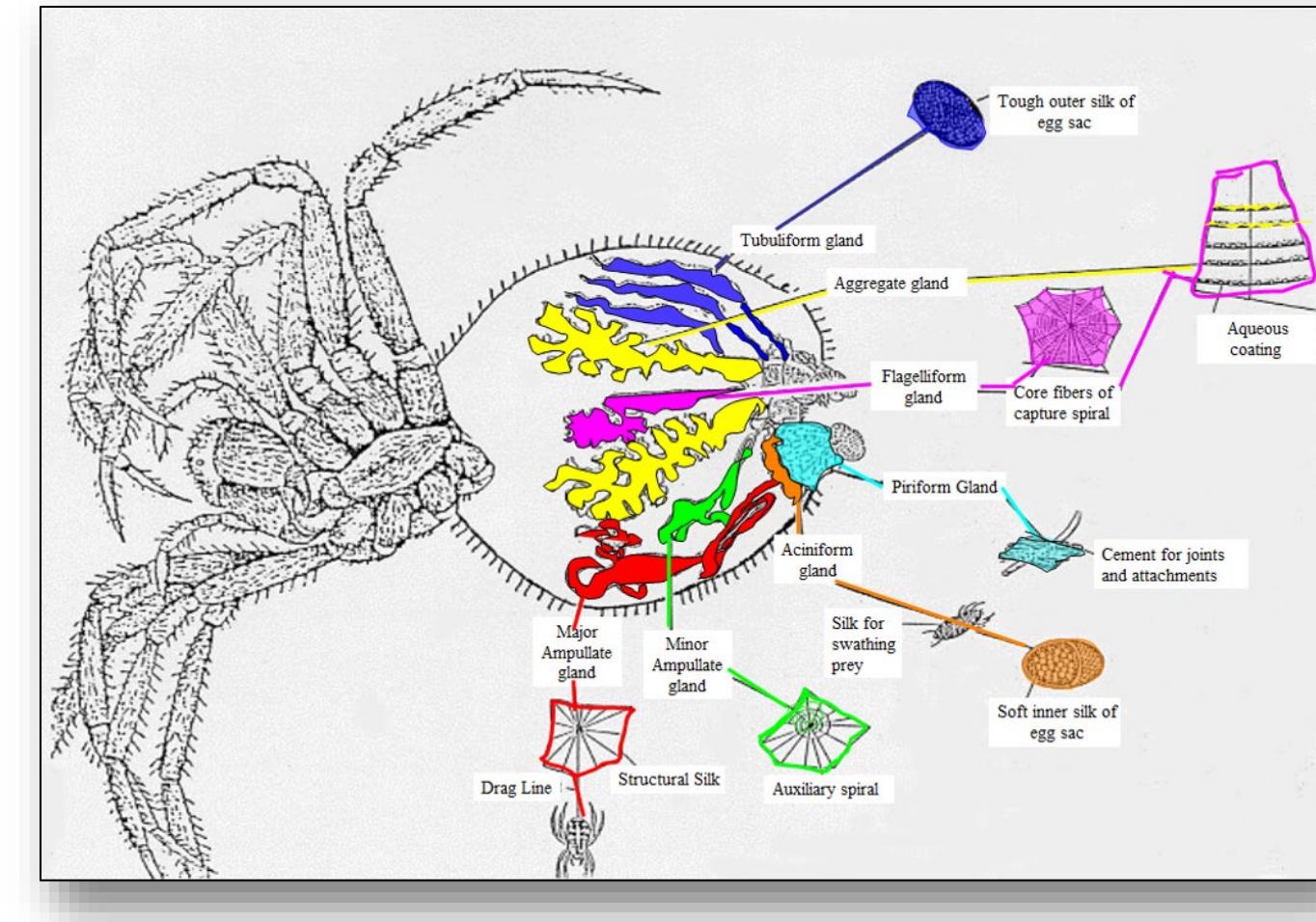
```
In [4]: pathways = predictor.run(product="lycopene", max_predictions=4)
```





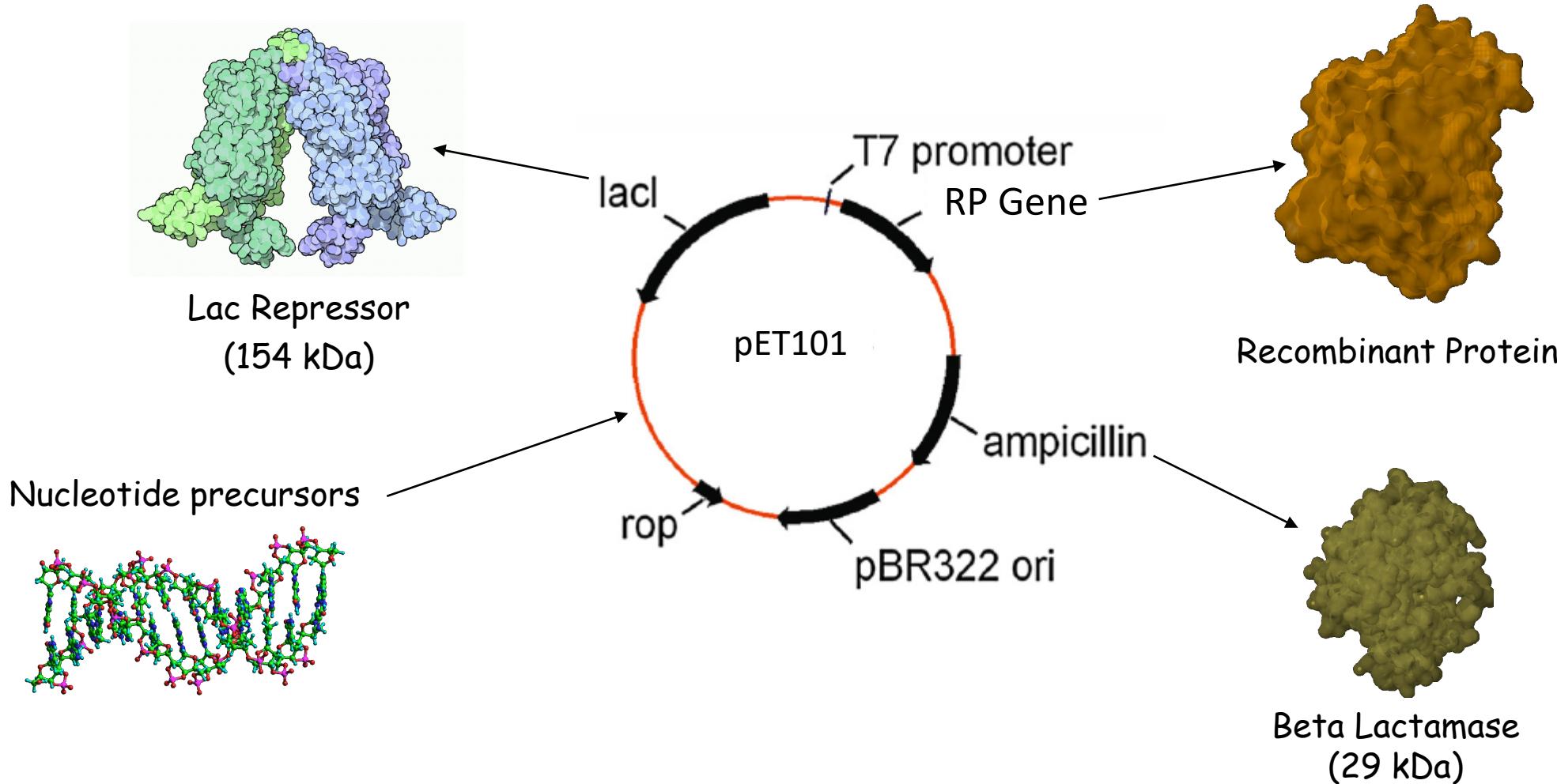
Strain Design

- Introduction
- Production Hosts
- Target Chemicals
- Heterologous Pathways
- Recombinant Proteins
- Gene/Reaction Knockouts
- Gene Expression Modulation
- Adaptive Laboratory Evolution





Components Produced by Plasmid for Recombinant Protein Production





MaSp2 Amino Acid Repeated Sequence

- MaSp2 spider silk amino acid sequence (349 Amino acids with MW of 29.5 Daltons)

KLHMT
GPGQQGPGGYGPQQGPSPGSAAAAAAAAA
GPGQQGPGGYGPQQGPGGYGPQQGPSPGSAAAAAAAAA
GPGGYGPQQGPGGYGPQQGPGGYGPQQGPSPGSAAAAAAAAA
GPGQQGPGGYGPQQGPGGYGPQQGPSPGSAAAAAAAAA
GPGQQGPGGYGPQQGPSPGSAAAAAAAAA
GPGQQGPGGYGPQQGPGGYGPQQGPSPGSAAAAAAAAA
GPGGYGPQQGPGGYGPQQGPGGYGPQQGPSPGSAAAAAAAAA
GPGQQGPGGYGPQQGPGGYGPQQGPSPGSAAAAAAAAA
GPGQQGPGGYGPQQGPSPGSAAAAAAAAA
SGGD

- Most used Amino acids
 - ✓ Glutamine - 48
 - ✓ Alanine - 72
 - ✓ Proline - 59
 - ✓ Glycine - 128

Number of each amino acid in sequence using Matlab Bioinformatics function "aaccount"

A: 72	Ala	Alanine
R: 0	Arg	Arginine
N: 0	Asn	Asparagine
D: 1	Asp	Aspartic acid (Aspartate)
C: 0	Cys	Cysteine
Q: 48	Gln	Glutamine
E: 0	Glu	Glutamic acid (Glutamate)
G: 128	Gly	Glycine
H: 1	His	Histidine
I: 0	Ile	Isoleucine
L: 1	Leu	Leucine
K: 1	Lys	Lysine
M: 1	Met	Methionine
F: 0	Phe	Phenylalanine
P: 59	Pro	Proline
S: 19	Ser	Serine
T: 1	Thr	Threonine
W: 0	Trp	Tryptophan
Y: 17	Tyr	Tyrosine
V: 0	Val	Valine



Quadrupled MaSp2 Amino Acid Sequence

- MaSp2 spider silk amino acid sequence (349 Amino acids with MW of 29.5 Daltons) multiplied 4 times (1396 Amino acids with MW of 118 kDa)

- MaSp2 spider silk Cobra reaction ($1396 * 4.306 = 6011.12$)

288.0 ala_L_c + 4.0 asp_L_c + 6011.12 atp_c + 192.0 gln_L_c + 512.0 gly_c + 6011.12 h2o_c + 4.0 his_L_c + 4.0 lys_L_c + 4.0 met_L_c + 236.0 pro_L_c + 76.0 ser_L_c + 4.0 thr_L_c + 68.0 tyr_L_c -->
 6011.12 adp_c + 6011.12 h_c + masp2_c + 6011.12 pi_c



Creating Recombinant Protein Enzyme

(MaSp2 Spider Silk Protein)

- The biosynthetic precursor balance for the sole plasmid encoded recombinant protein combined with energetic requirements of 4.306 mol ATP/mol amino acids.
- The reaction includes the number of each amino acids that makes up the protein plus the energetic requirements of ATP, ADP and Pi which is the sum of the number of amino acids multiplied by 4.306 mol ATP/mol amino acids.
- For example, the reaction formula for the MaSp2 reaction
 - ✓ $288.0 \text{ ala_L_c} + 4.0 \text{ asp_L_c} + 6011.12 \text{ atp_c} + 192.0 \text{ gln_L_c} + 512.0 \text{ gly_c} + 6011.12 \text{ h2o_c} + 4.0 \text{ his_L_c} + 4.0 \text{ lys_L_c} + 4.0 \text{ met_L_c} + 236.0 \text{ pro_L_c} + 76.0 \text{ ser_L_c} + 4.0 \text{ thr_L_c} + 68.0 \text{ tyr_L_c} \rightarrow 6011.12 \text{ adp_c} + 6011.12 \text{ h_c} + \text{masp2_c} + 6011.12 \text{ pi_c}$
 - ✓ A demand reaction for the masp2 metabolite can be set as high as 50% of the total protein mass.
 - ✓ `model.add_boundary(model.metabolites.get_by_id("masp2_c"), type="demand")`
 - ✓ `model_masp2.reactions.DM_masp2_c.bounds = lower limit, upper limit`
 - ✓ Remember that COBRA biomass includes the required components (amino acids, etc.) of the host cell reactions in the biomass function, thus the need to add the separate load of the recombinant protein enzyme to the model.



Practical Production Limits

- Maximum biomass density
 - ✓ An upper limit for cell concentration in a large scale fermenter is less than 100 gDW/L ($OD_{600} = 47$)
 - Choi, J. H., K. C. Keum, et al. (2006). "Production of recombinant proteins by high cell density culture of *Escherichia coli*." *Chemical Engineering Science* 61(3): 876-885
 - Shiloach, J. and R. Fass (2005). "Growing *E. coli* to high cell density--a historical perspective on method development." *Biotechnol Adv* 23(5): 345-357
- Protein production limit (Total protein = 550 mg per gram dry weight biomass)
 - ✓ An upper limit for the production of a recombinant protein is less than 50% of the total protein (275 mg/gDW)
 - Rosano, G. L. and E. A. Ceccarelli (2014). "Recombinant protein expression in *Escherichia coli*: advances and challenges." *Front Microbiol* 5: 172.
- *E. coli* growth is inhibited when the following nutrients are present above certain concentrations (shown in brackets)
 - ✓ Glucose (50 g/L), ammonia (3 g/L), iron (1.15 g/L), magnesium (8.7 g/L), phosphorous (10 g/L) zinc (0.038 g/L), acetate (5 g/L), and ethanol (47.3 g/L)
 - Lee, S. Y. (1996). "High cell-density culture of *Escherichia coli*." *Trends Biotechnol* 14(3): 98-105.
 - Riesenbergs, D. (1991). "High-cell-density cultivation of *Escherichia coli*." *Current opinion in biotechnology* 2(3): 380-384.
 - Goodarzi, H. (2010). "Regulatory and metabolic rewiring during laboratory evolution of ethanol tolerance in *E. coli*." *Molecular Systems Biology*, 6, 378



Spider Silk (MaSp2) Strain Design

MaSp2_design.ipynb

Design *E.coli* BL21(DE3) cells to produce the MaSp2 spider silk protein [1,2].

Set the environment

```
In [1]: from cameo import models
from cameo.strain_design import pathway_prediction
from cobrapy_bigg_client import client
import pandas as pd
pd.set_option('display.max_rows', 1000)
```

Load the *E.coli* BL21(DE3) model 'iECD_1391'

```
In [2]: model_orig = client.download_model('iECD_1391', save=False) # BL21 (DE3) E.coli model
model = model_orig.copy()
```

```
Set parameter Username
Academic license - for non-commercial use only - expires 2022-10-10
Read LP format model from file C:\Users\hinton\AppData\Local\Temp\tmpl3vm_b_y.lp
Reading time = 0.02 seconds
: 1943 rows, 5482 columns, 21200 nonzeros
```

Review the model summary of the *E.coli* BL21(DE3) model

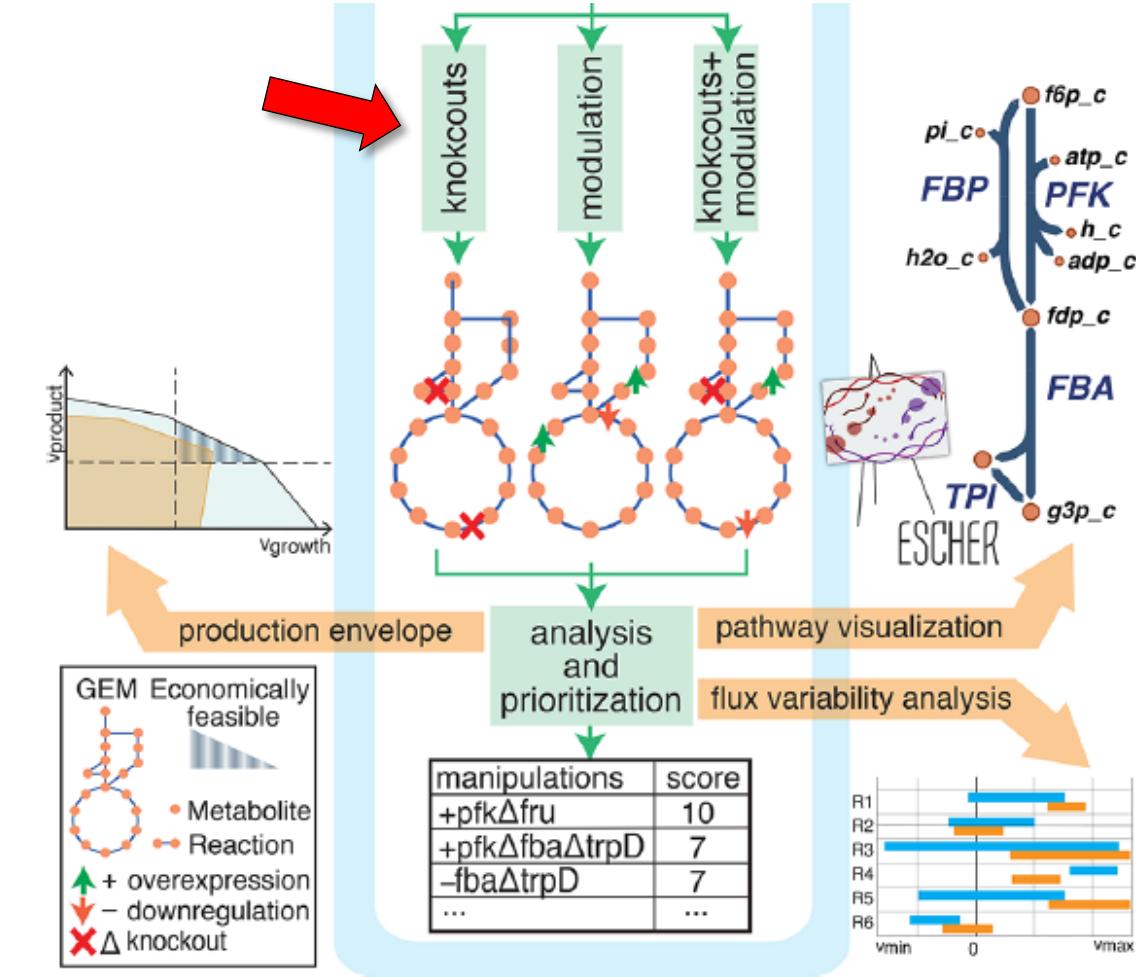
```
In [3]: model
```

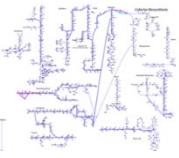
Out[3]:	Name	iECD_1391
	Memory address	0x02252e48ec40
	Number of metabolites	1943
	Number of reactions	2741
	Number of groups	0
	Objective expression	1.0*BIOMASS_Ec_iJO1366_core_53p95M - 1.0*BIOMASS_Ec_iJO1366_core_53p95M_reverse_5c8b1
	Compartments	extracellular space, cytosol, periplasm



Strain Design

- Introduction
- Production Hosts
- Target Chemicals
- Heterologous Pathways
- Recombinant Proteins
- Gene/Reaction Knockouts
- Gene Expression Modulation
- Adaptive Laboratory Evolution





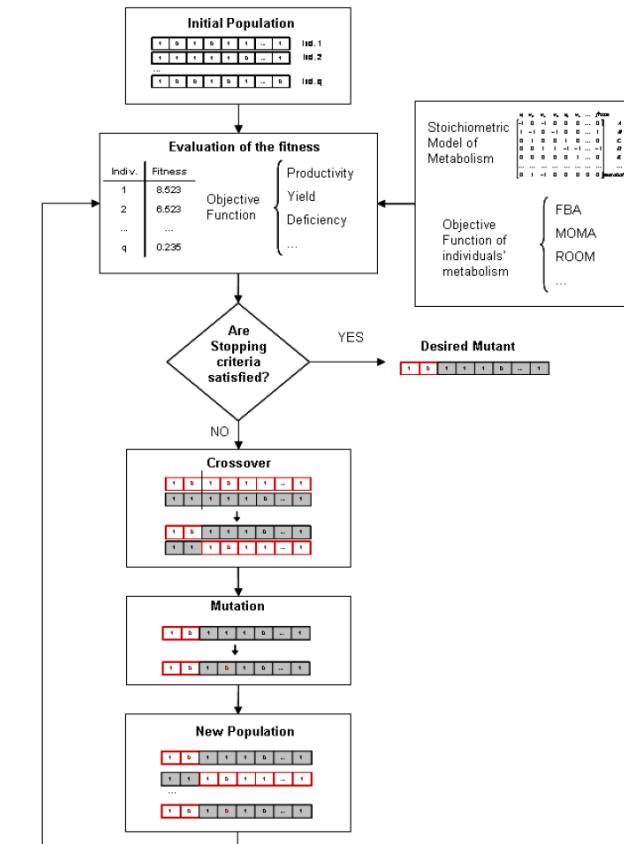
Knockout Tools

Maximize: Bioengineering Objective
(through reaction knockouts)

Subject to: Maximize: cellular objective
(over fluxes)
Subject to: Fixed substrate uptake
Network Stoichiometry
Blocked reactions identified
by the outer problem

Number of knockouts \leq limit

OptKnock

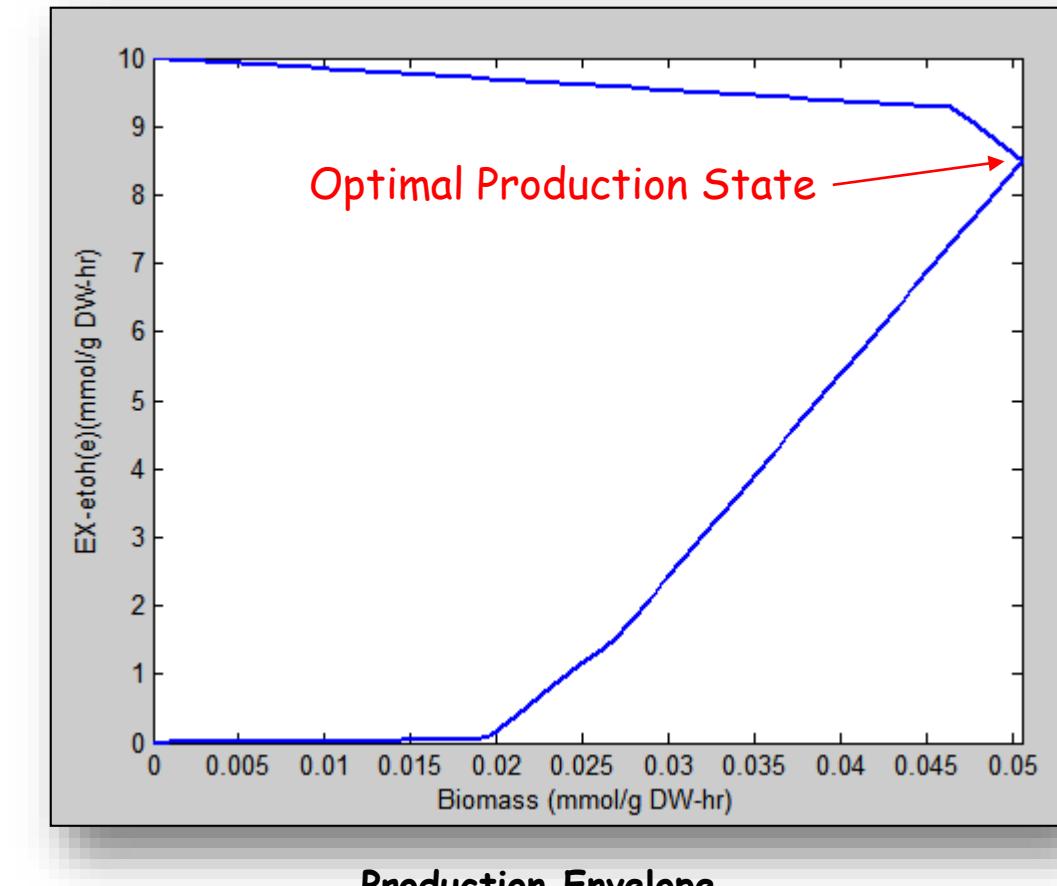


OptGene



Do Cells Really Operate at the Calculated Optimal State Of Bioproduct Production?

- It has been assumed that the mutant bacteria display an optimal metabolic state.
- Unfortunately, mutants generated artificially in the laboratory are generally not subjected to the same evolutionary pressure that shaped the wild type. Thus, a mutant is likely to initially display a **suboptimal** flux distribution that is somehow intermediate between the wild-type optimum and the mutant optimum.
- The method of minimization of metabolic adjustment (MOMA) has been developed, which is based on the same stoichiometric constraints as FBA, but relaxes the assumption of optimal growth flux for gene/reaction deletions.
- MOMA provides a mathematically tractable approximation for this intermediate suboptimal state, based on the conjecture that the mutant remains initially as close as possible to the wild type.



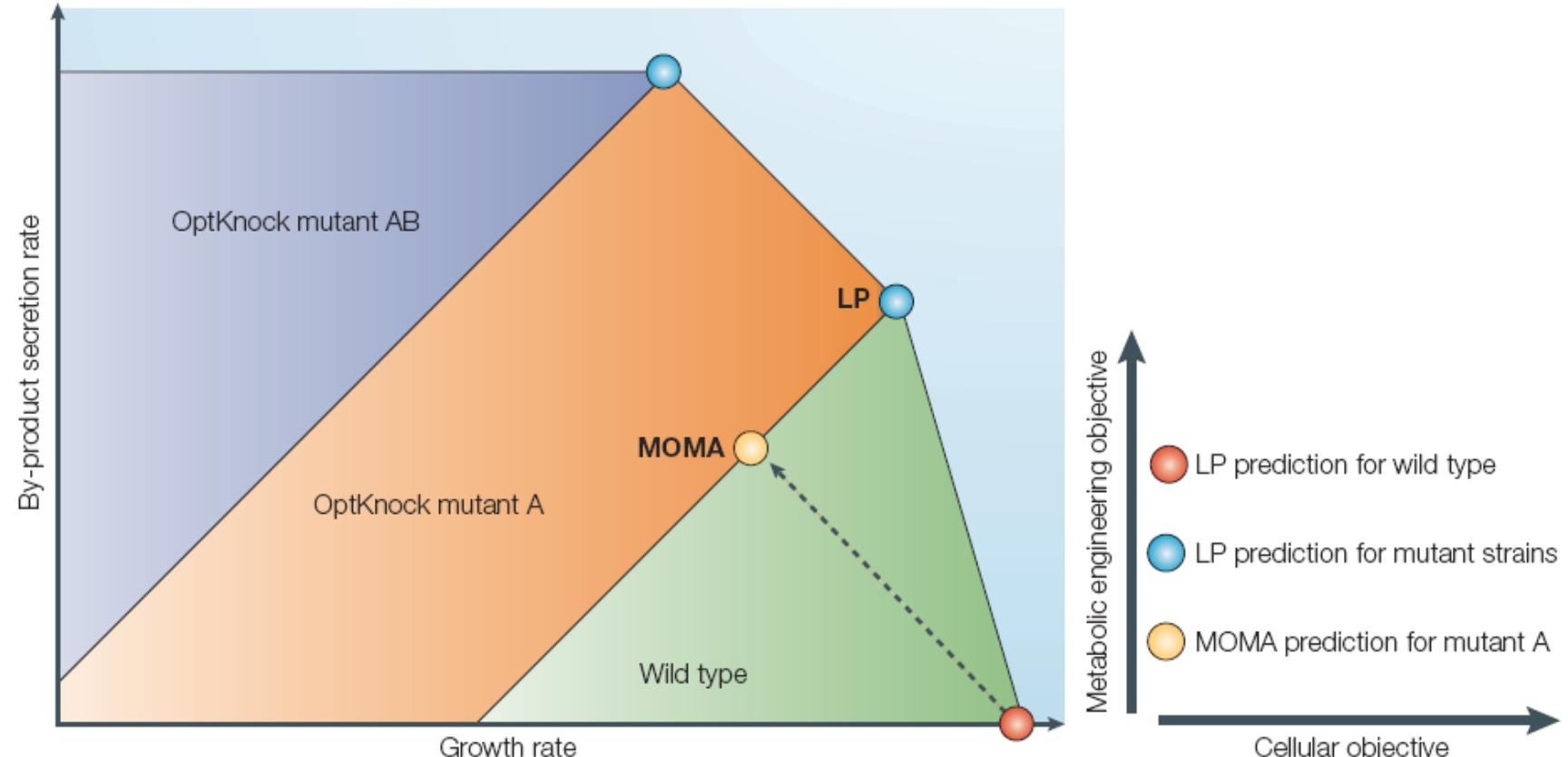
Segre, D., D. Vitkup, et al. (2002). "Analysis of optimality in natural and perturbed metabolic networks." Proceedings of the National Academy of Sciences of the United States of America 99(23): 15112-15117.



Method of Minimization of Metabolic Adjustment (MOMA)

Segre, D., D. Vitkup, et al. (2002). "Analysis of optimality in natural and perturbed metabolic networks."

- Uses the same steady state flux cone as FBA.
- Relaxes the assumption of maximal optimal growth.
- MOMA searches the flux distribution in the "mutant flux space" which is closest to the optimal flux distribution in the "wild-type flux space."
- Typically returns suboptimal flux distribution between wild type optimum and mutant optimum



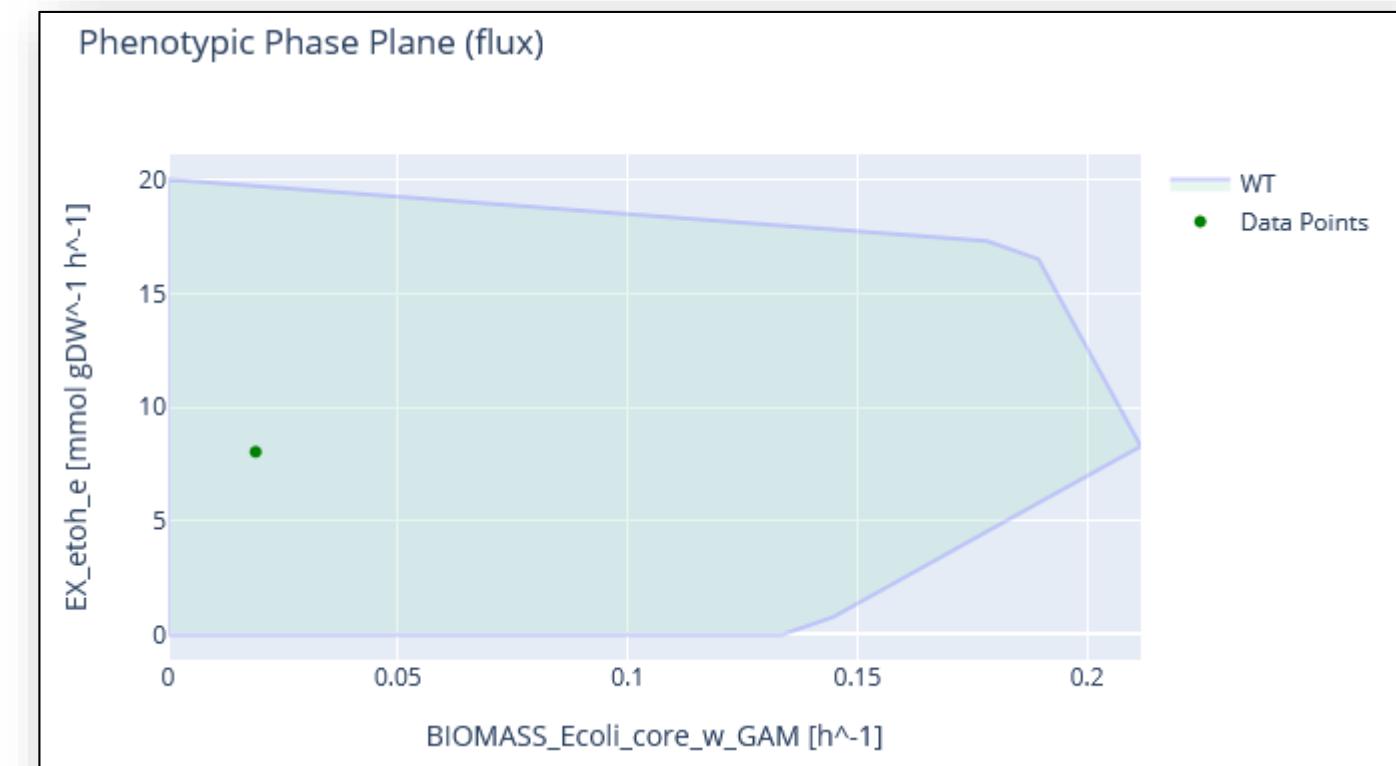
Segre, D., D. Vitkup, et al. (2002). "Analysis of optimality in natural and perturbed metabolic networks." Proceedings of the National Academy of Sciences of the United States of America 99(23): 15112-15117.

Price, N. D., J. L. Reed, et al. (2004). "Genome-scale models of microbial cells: evaluating the consequences of constraints." Nature reviews. Microbiology 2(11): 886-897.

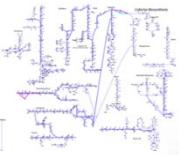


Regulatory On-Off Minimization (ROOM)

- Regulatory On-Off Minimization (ROOM) is a model for predicting the behavior of metabolic networks in response to gene knockouts.
- Since regulatory constraints are not explicitly incorporated into metabolic network models, ROOM implicitly accounts for regulatory changes by identifying significant flux changes in the respective metabolic reactions.
- ROOM minimizes the total number of significant flux changes from the wild-type flux distribution.
- Since ROOM acts to minimize the number of significant flux changes, a significant change in growth is unlikely.
- ROOM is shown to provide more accurate flux predictions than FBA and MOMA for the final metabolic steady state.
- ROOM is computationally more complex than MOMA



Tomer Shlomi, Omer Berkman and Eytan Ruppin, "Regulatory on/off minimization of metabolic flux changes after genetic perturbations", PNAS 2005 102 (21) 7695-7700;



MOMA & ROOM

MOMA_ROOM_core_model.ipynb

Set environment

```
In [1]: import escher
from cameo.visualization.plotting.with_plotly import PlotlyPlotter
plotter = PlotlyPlotter()

from cameo import load_model
model = load_model('e_coli_core')
model.reactions.EX_o2_e.lower_bound = -0 # Set the model for anaerobic operation
```

Set parameter Username

Academic license - for non-commercial use only - expires 2022-10-10

Reference Model

```
In [2]: from cameo import fba
%time WT_result = fba(model)
WT_growth = WT_result.fluxes.BIOMASS_Ecoli_core_w_GAM
WT_ethanol_production = WT_result.fluxes.EX_etoh_e
print('FBA Growth = ',WT_growth,'      ', 'FBA Ethanol Production = ',WT_ethanol_production)
```

Wall time: 5.99 ms

FBA Growth = 0.21166294973531055

FBA Ethanol Production = 8.279455380486583

Plot a phenotypic phase plane setting the objective to maximize the production of ethanol. This illustrates the boundaries for the operational solution space of the cell as a function of the growth rate and the ethanol production.

Plot the location of the FBA optimized cell operation. Note that the ethanol production is coupled to the growth of the cell. The more the cell grows the more ethanol is produced.



Spider Silk (MaSp2) Protein MOMA Analysis

Set environment

```
In [1]: import cobra.test
import escher
from cameo.visualization.plotting.with_plotly import PlotlyPlotter
from cobrapy_bigg_client import client
import pandas as pd
pd.set_option('display.max_rows', 1000)
plotter = PlotlyPlotter()
```

Reference Model

Create a wild type model that can be used as the reference model for the MOMA analysis. The wild type model in this case does not include the plasmids that create the MaSp2 proteins. It is the basic BL21(DE3) model found in the BIGG database called "iECD_1391".

```
In [2]: model_WT = client.download_model('iECD_1391', save=False) # BL21(DE3) E.coli model
solution_WT = model_WT.optimize()
growth_WT = solution_WT.fluxes.BIOMASS_Ec_iJO1366_core_53p95M
masp2_production_WT = 0
print('Wild Type Growth =',growth_WT, ' Wild Type MaSp2 Production =',masp2_production_WT)
```

```
Set parameter Username
Academic license - for non-commercial use only - expires 2022-10-10
Wild Type Growth = 0.9756145115851279      Wild Type MaSp2 Production = 0
```

Mutant Model

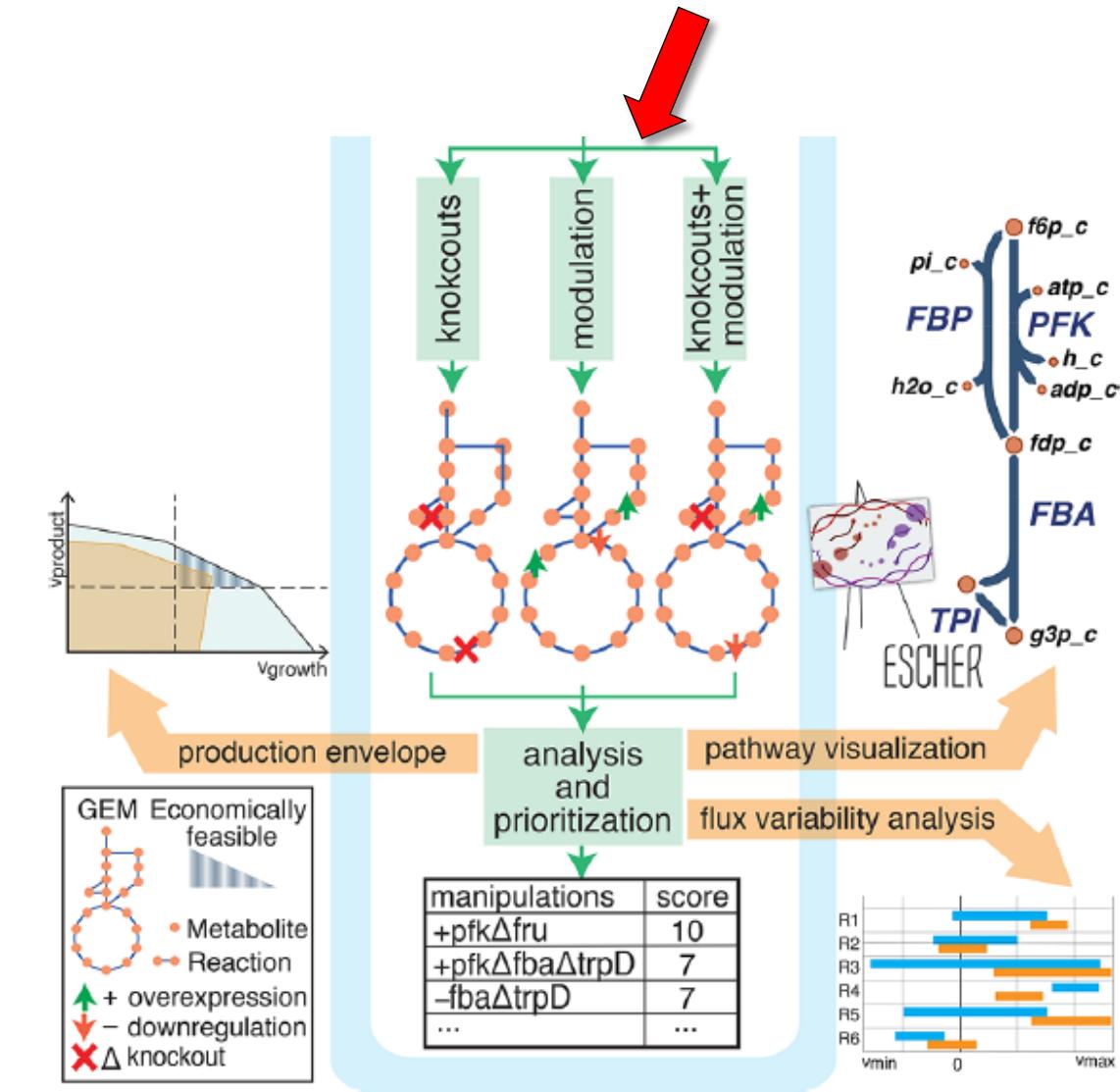
The mutant model is the BL21(DE3) model that includes the MaSp2 pathway ('iECD_1391_MaSp2.json').

MaSp2_MOMA.ipynb



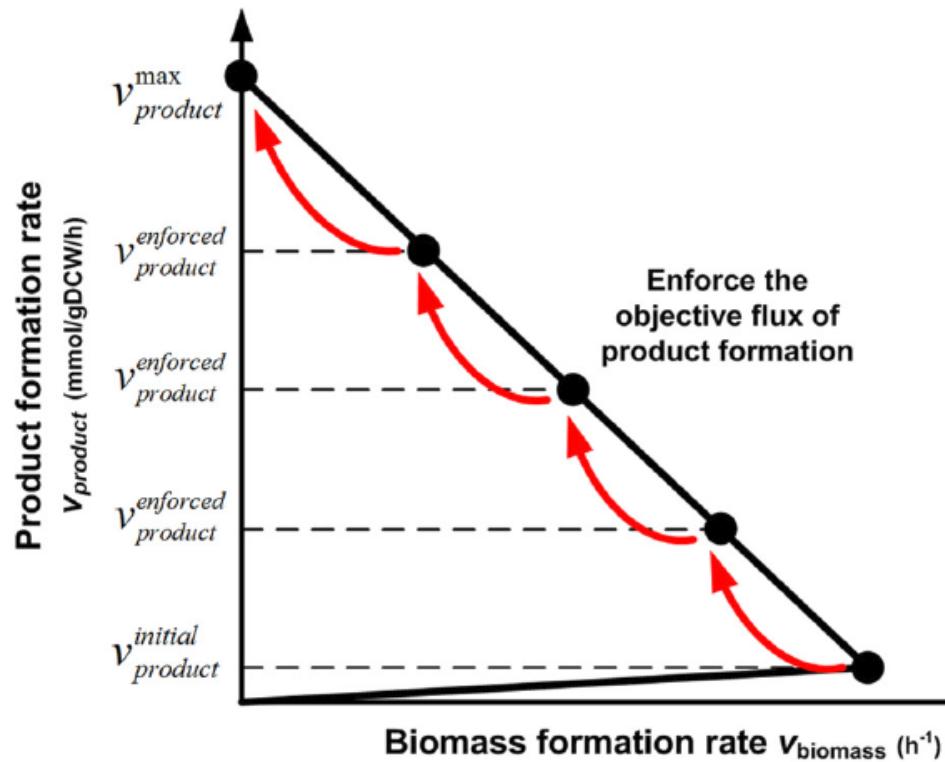
Strain Design

- Introduction
- Production Hosts
- Target Chemicals
- Heterologous Pathways
- Recombinant Proteins
- Gene/Reaction Knockouts
- Gene Expression Modulation
- Adaptive Laboratory Evolution



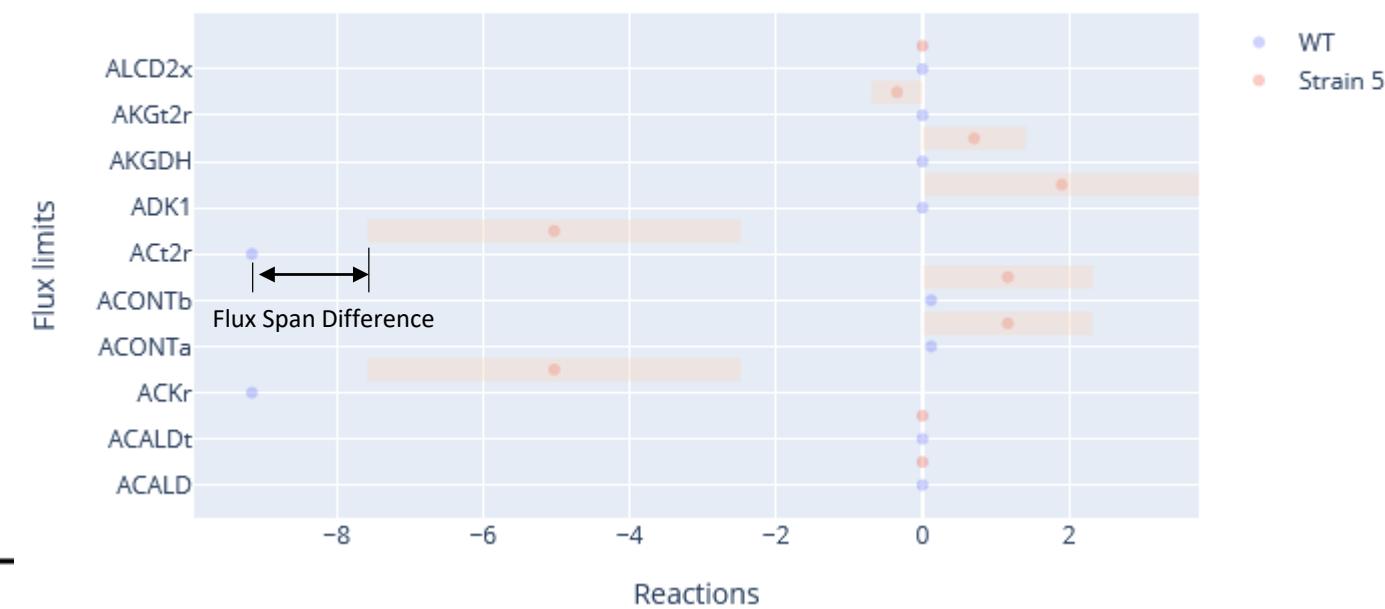


Gene Expression Modulation



Flux Scanning based on Enforced Objective Flux (FSEOF)

Compare WT solution 5

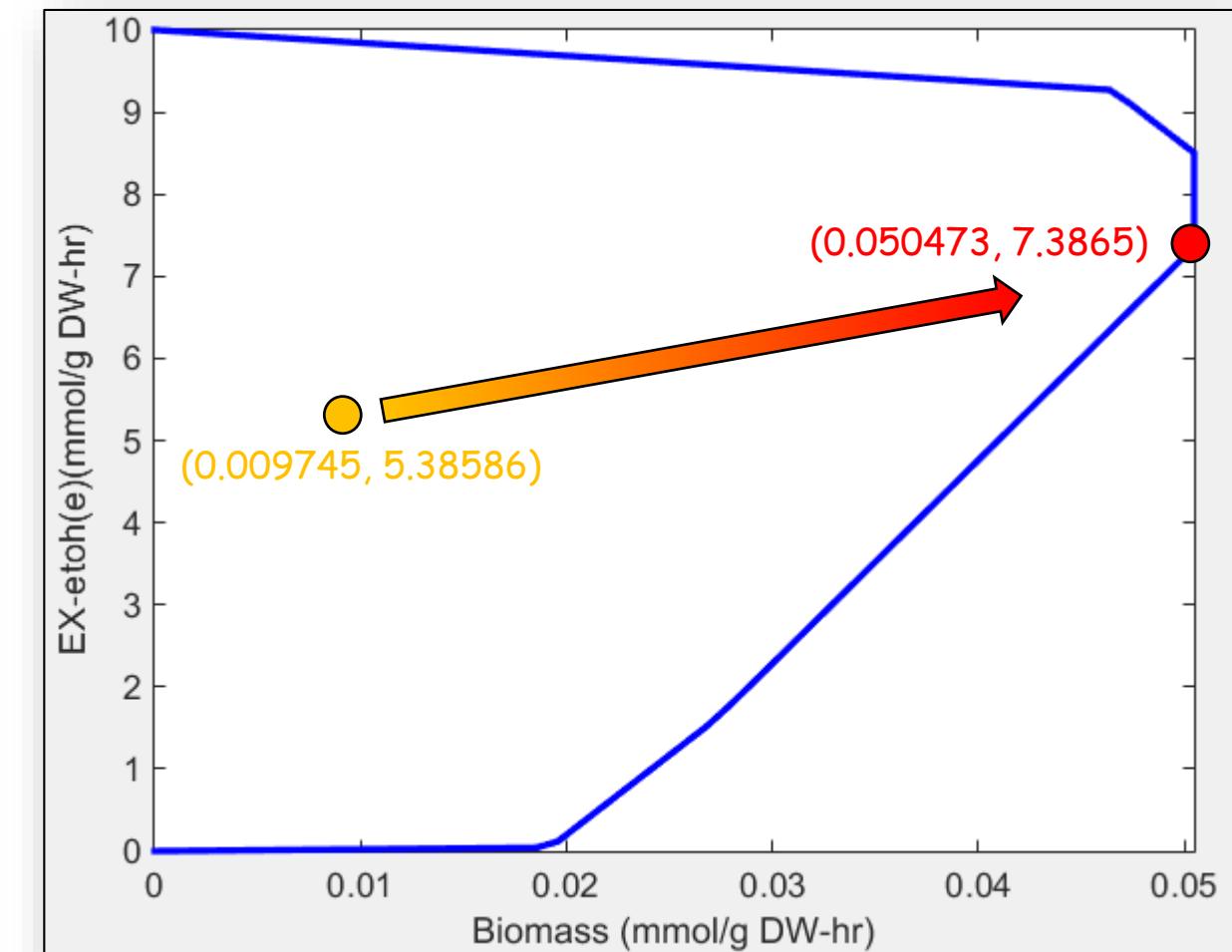


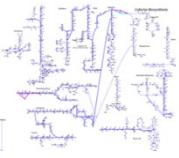
Differential Flux Variability Analysis



Strain Design

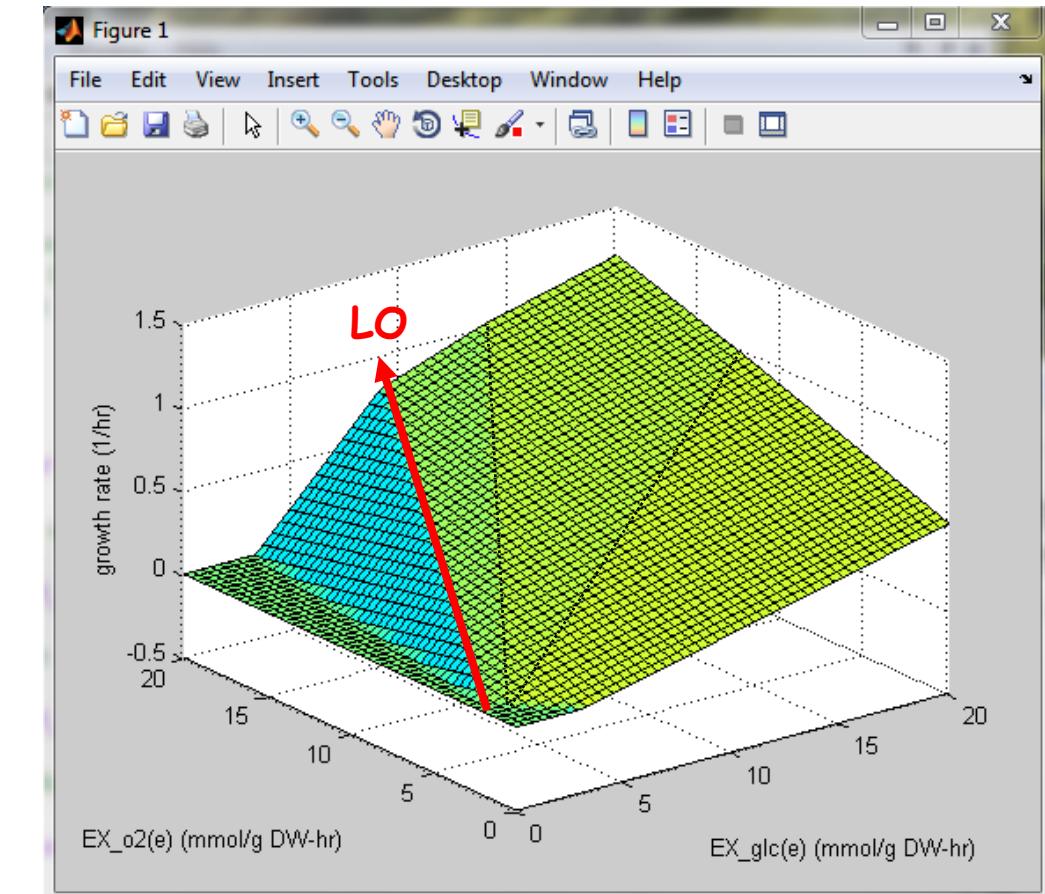
- Introduction
- Production Hosts
- Target Chemicals
- Heterologous Pathways
- Recombinant Proteins
- Gene/Reaction Knockouts
- Gene Expression Modulation
- • Adaptive Laboratory Evolution

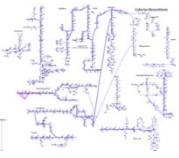




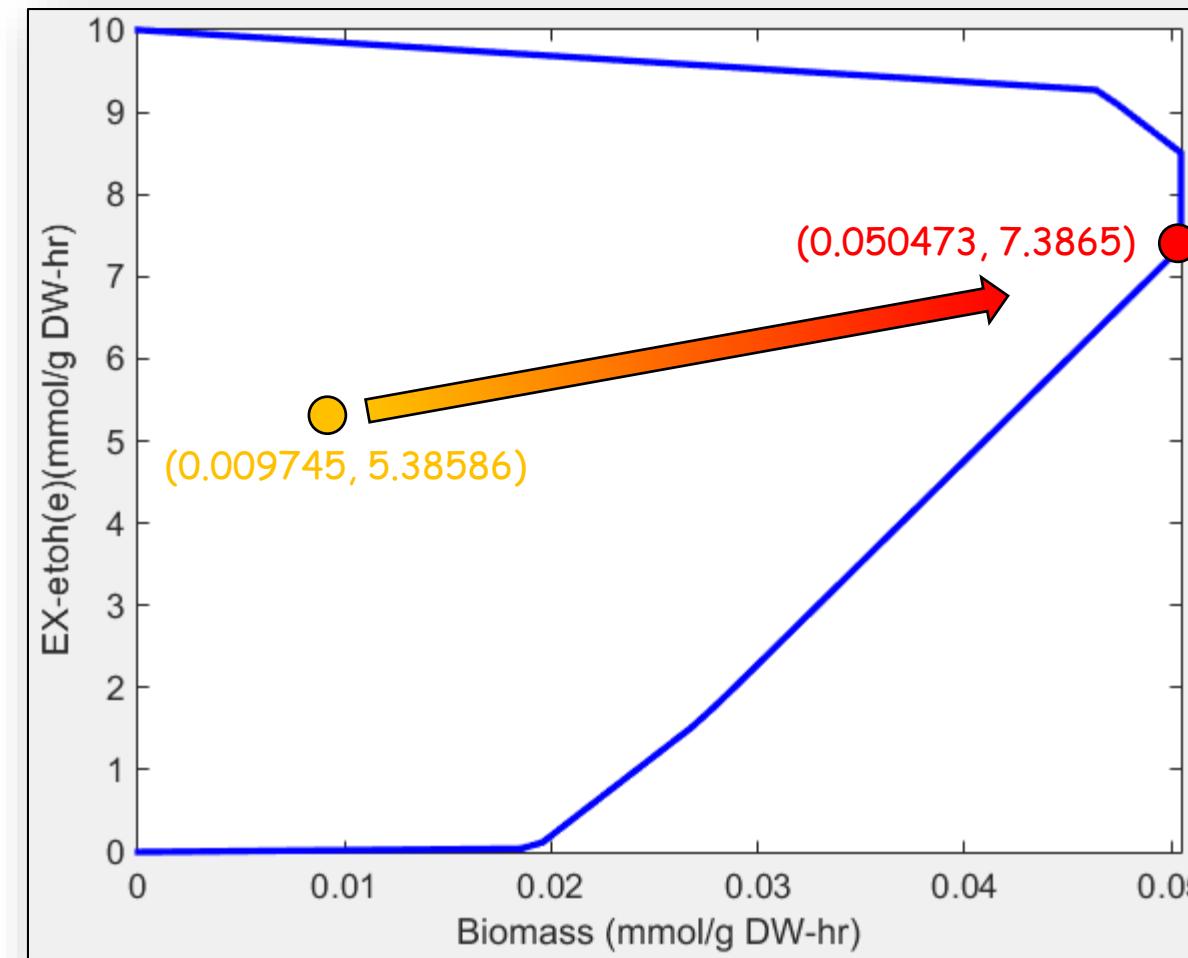
Line of Optimality

- The line of optimality (LO) is defined as a line representing the optimal relation between the two metabolic fluxes used to create a phenotype phase plane.
- The line of optimality is determined by specifying an uptake rate of the substrate along the x-axis and then allowing any value for the flux along the y-axis. Linear Programming can then be used to calculate the optimal value of the objective as a function of the y-axis flux. Once the objective is determined, the corresponding flux value for the y-axis is used to plot the line of optimality (LO).
- The LO defines the optimal utilization of the metabolic pathways without limitations on the availability of the substrates.



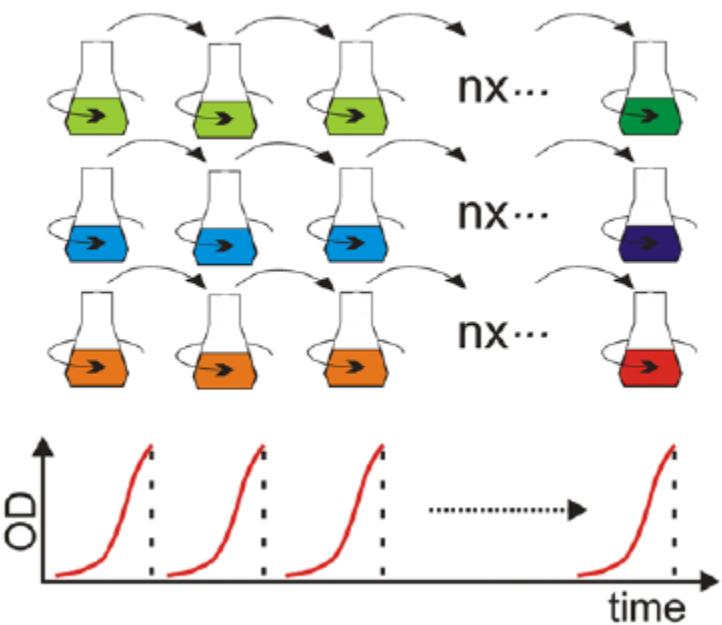
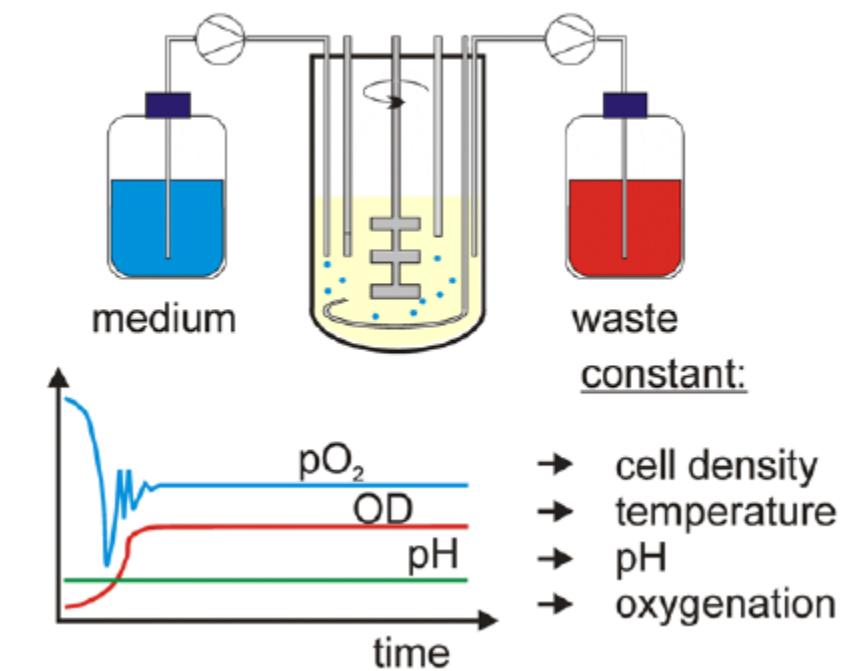


Desired Cell Evolution

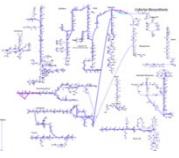




Adaptive Laboratory Evolution Methods

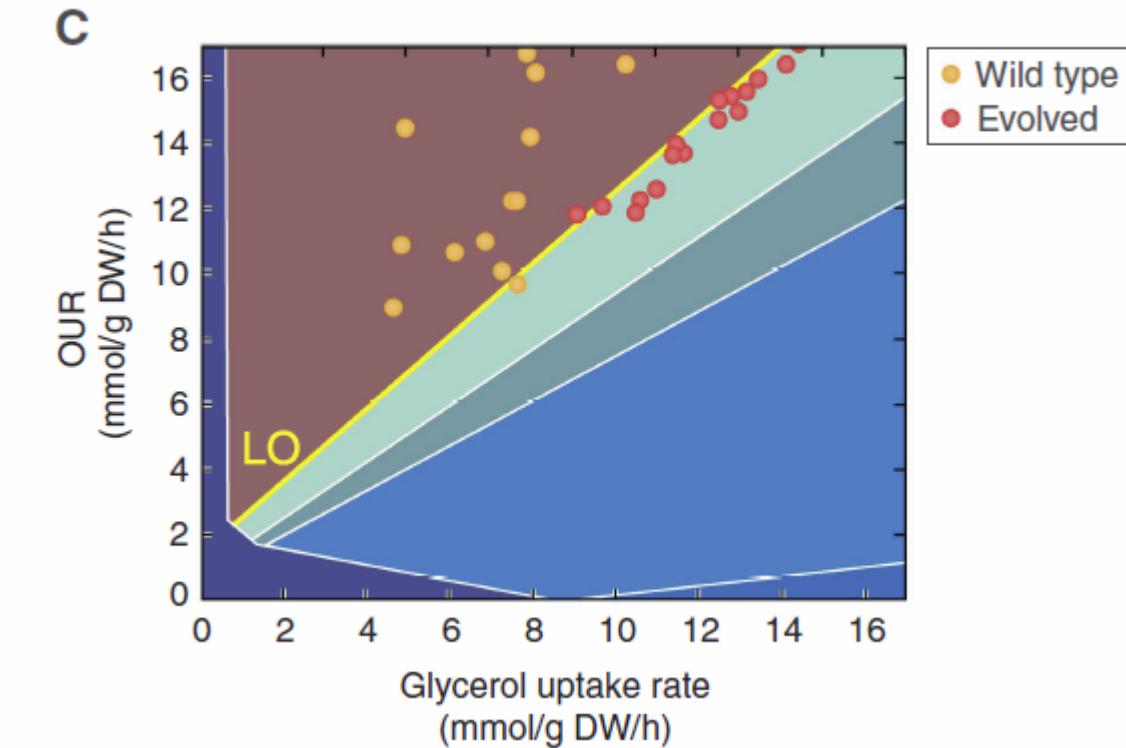
a serial dilution**b chemostat**

Dragosits, M. and D. Mattanovich (2013). "Adaptive laboratory evolution -- principles and applications for biotechnology." *Microbial cell factories* 12: 64.



Adaptive Laboratory Evolution

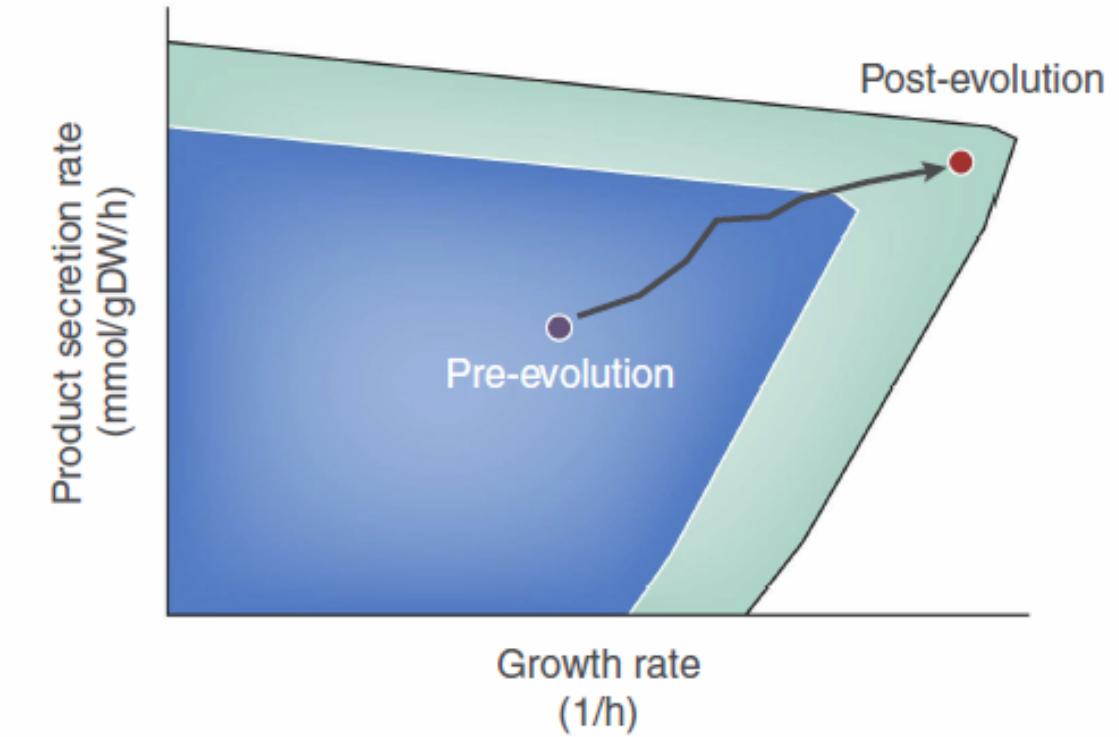
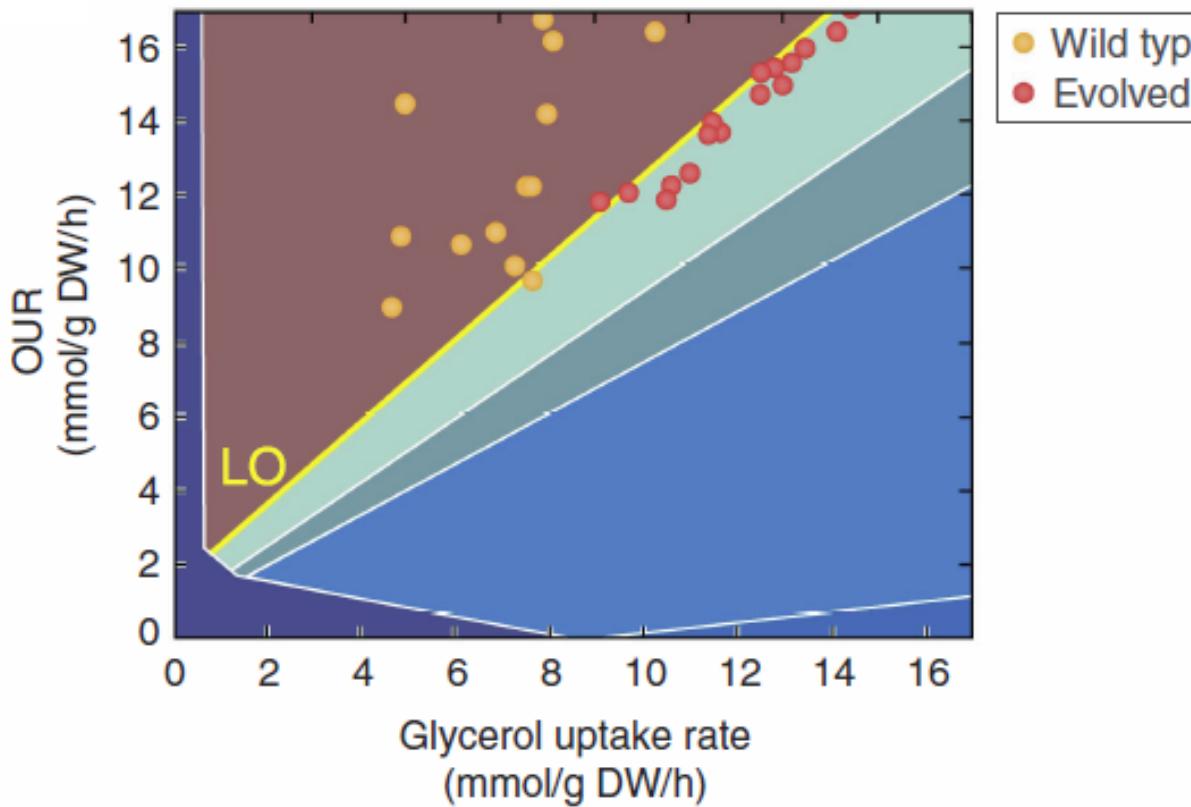
- Cells will modify their genotype to optimize their fitness in the given environment.
- FBA results can be presented on phenotypic phase-plane (PPP) plots of model-predicted optimal biomass production (growth rate) versus carbon source uptake rate (SUR) and oxygen uptake rate (OUR).
- The line of optimality (LO) describes the most efficient ratio of SUR and OUR for biomass synthesis.
- Under several conditions, the experimentally measured *E.coli* phenotype corresponds to the LO of the PPP
- When *E.coli* growth is not consistent with the LO, populations migrate toward the LO through adaptive evolution.
- Adaptive evolution outcomes have shown that evolved strains exhibit a general pattern of increased expression of genes and proteins associated with the optimal flux distribution, and decreased expression of genes and proteins associated with unused pathways
- The frequent mutation of transcriptional regulators is consistent with recent evidence showing that regulatory networks evolve faster than other networks, such as genetic networks, protein interaction networks, and metabolic networks



Conrad, T. M., N. E. Lewis, et al. (2011). "Microbial laboratory evolution in the era of genome-scale science." *Molecular Systems Biology* 7: 509.



Adaptive Laboratory Evolution

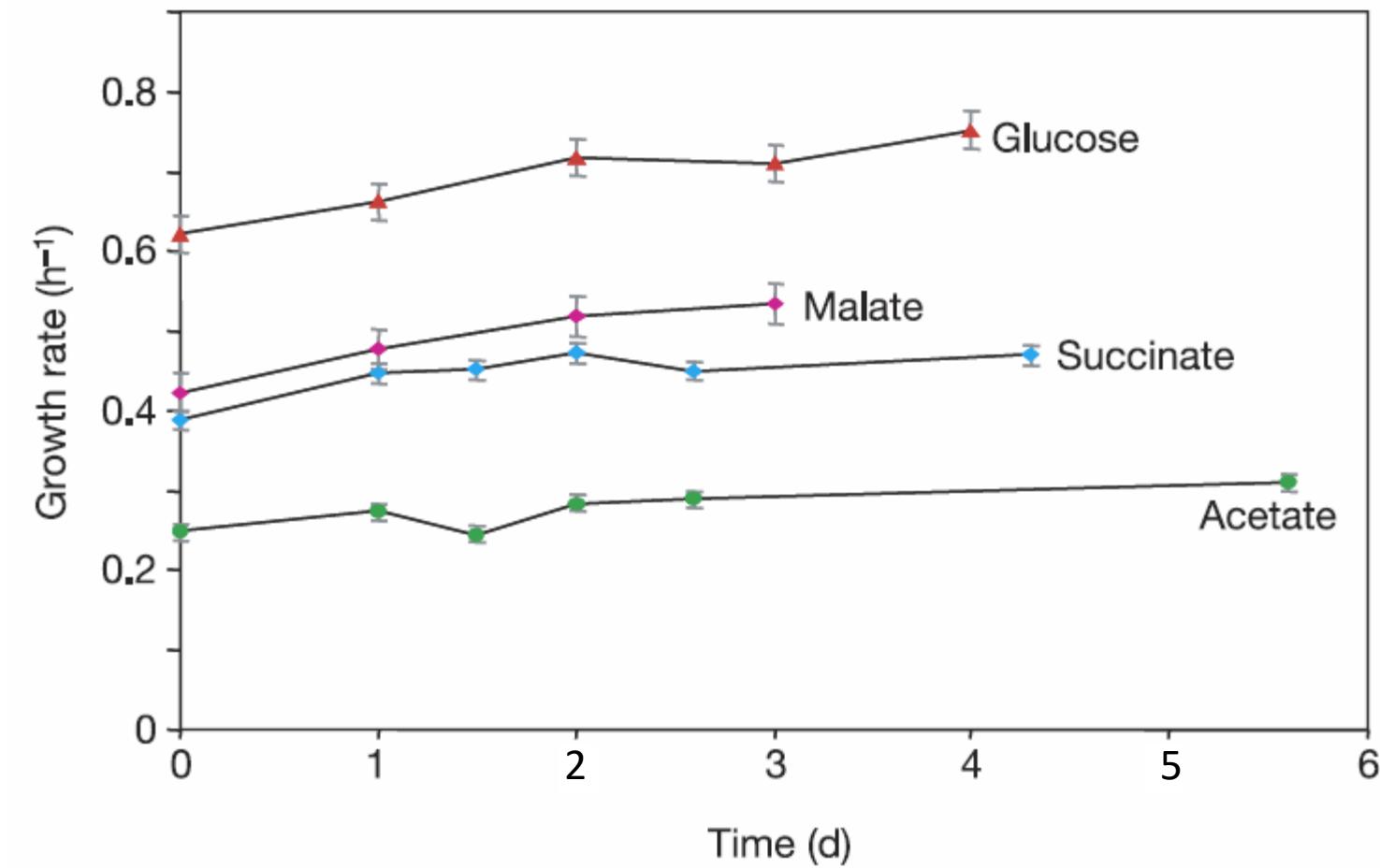


Conrad, T. M., N. E. Lewis, et al. (2011). "Microbial laboratory evolution in the era of genome-scale science." *Molecular Systems Biology* 7: 509.



Growth rate of *E.coli* K-12 on Glucose, Malate, Succinate, & Acetate

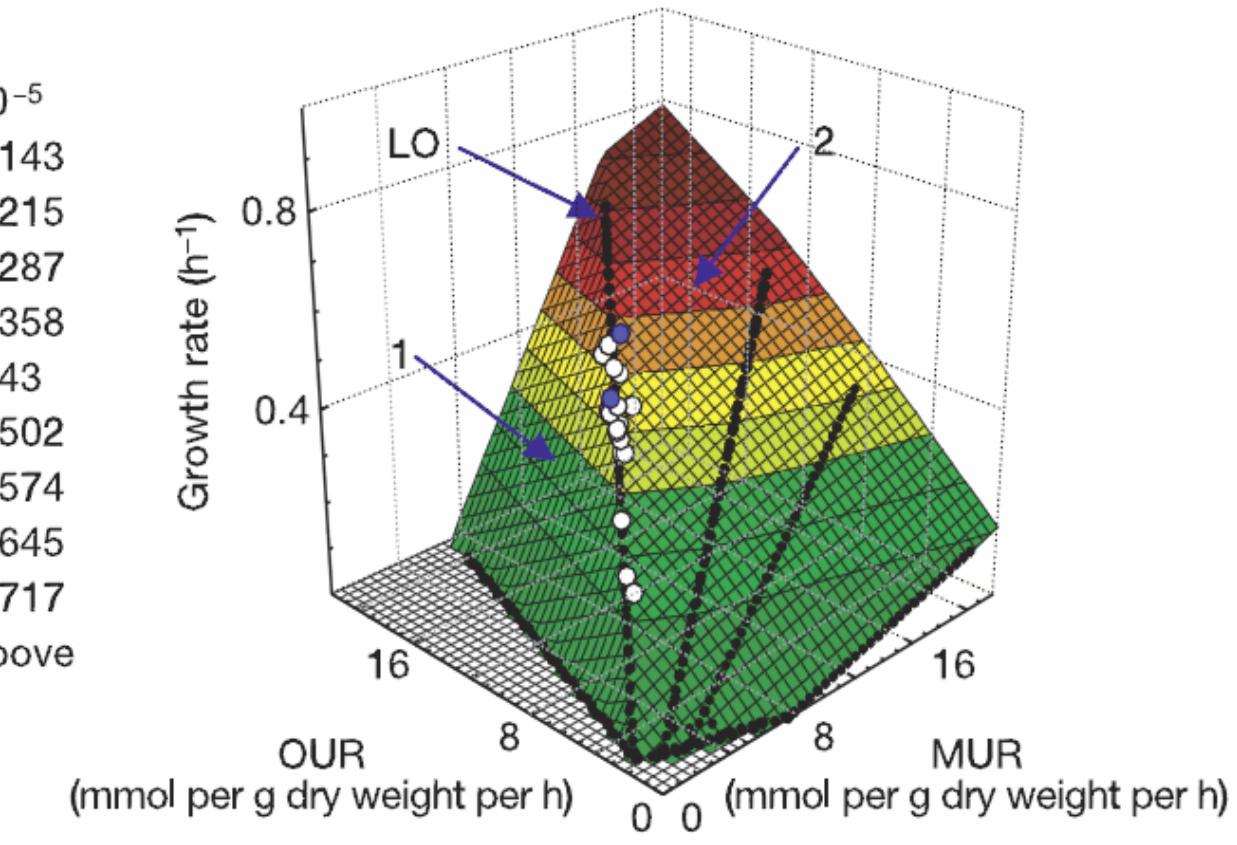
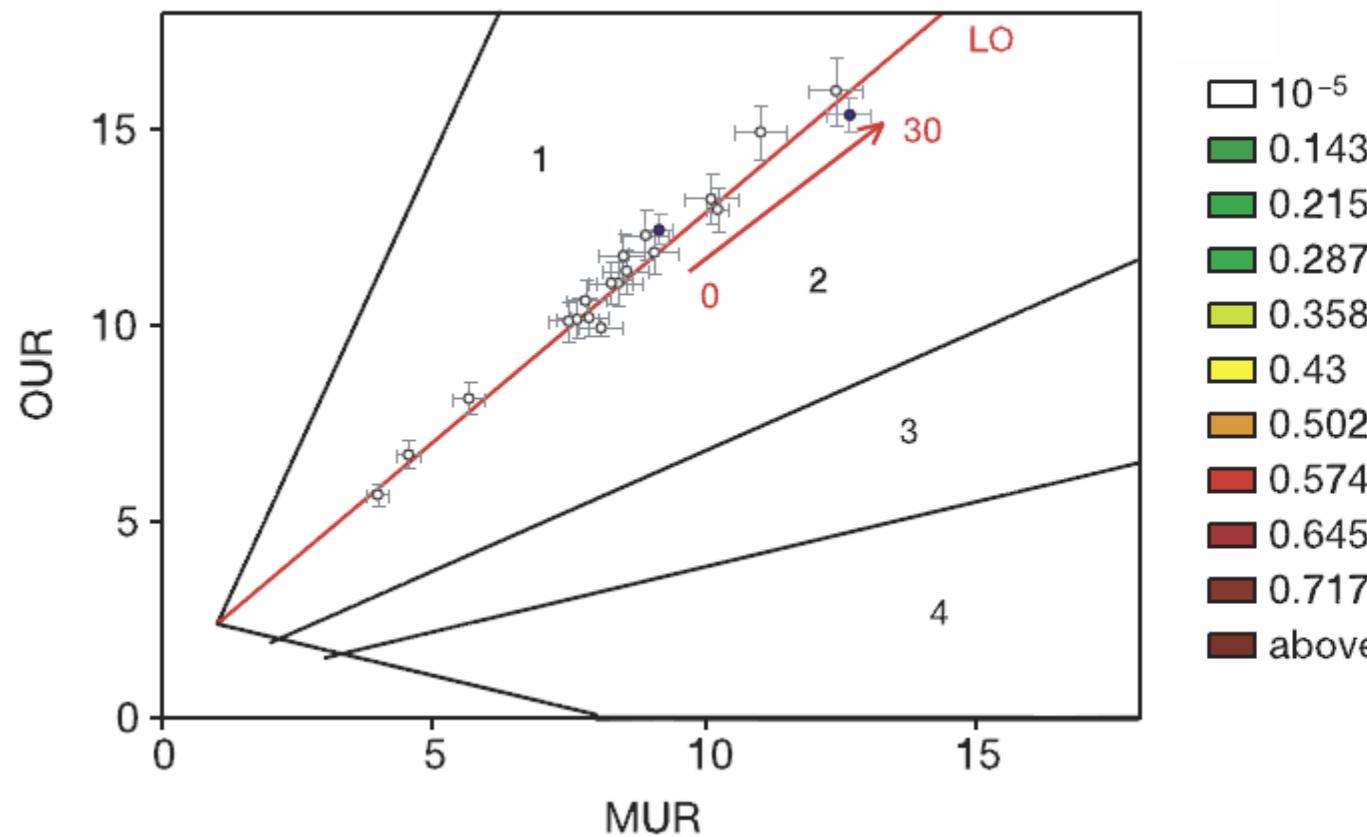
- Growth rate (exponential phase) during adaptive evolution on glucose, malate, succinate and acetate.
- The increases in growth rate over time were as follows:
 - ✓ glucose (18%),
 - ✓ malate (21%),
 - ✓ succinate (17%) and
 - ✓ acetate (20%).
- The number of generations for each adaptive evolution was: glucose (500), malate (500), succinate (1,000) and acetate (700).



Ibarra, R. U., J. S. Edwards, et al. (2002). "Escherichia coli K-12 undergoes adaptive evolution to achieve *in silico* predicted optimal growth." Nature 420(6912): 186-189.



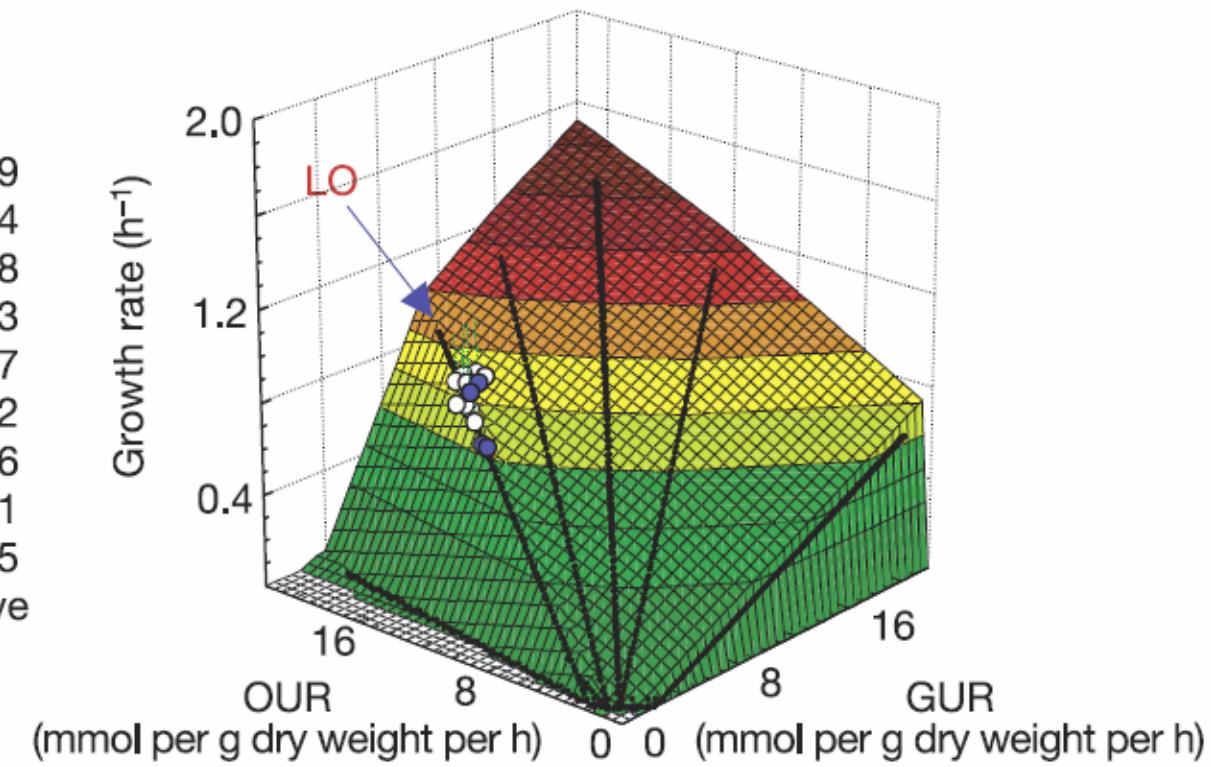
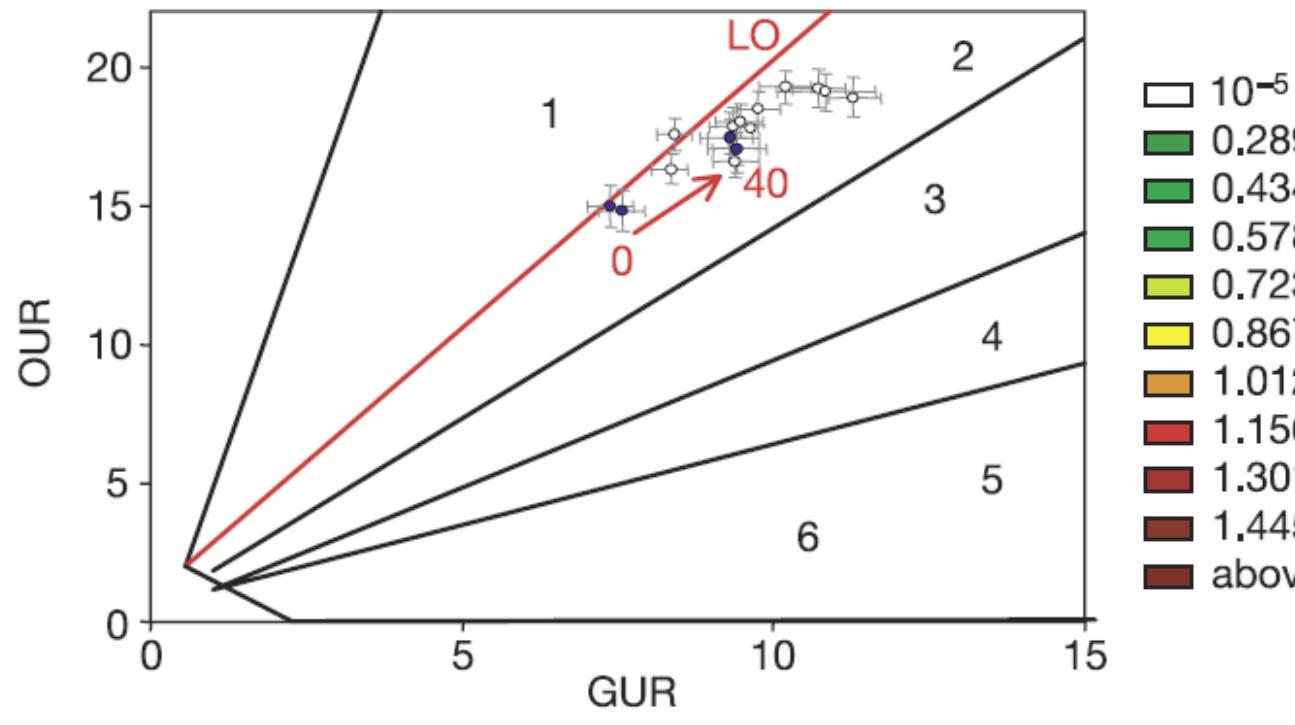
Growth of *E.coli* K-12 on Malate



Ibarra, R. U., J. S. Edwards, et al. (2002). "Escherichia coli K-12 undergoes adaptive evolution to achieve *in silico* predicted optimal growth." Nature 420(6912): 186-189.



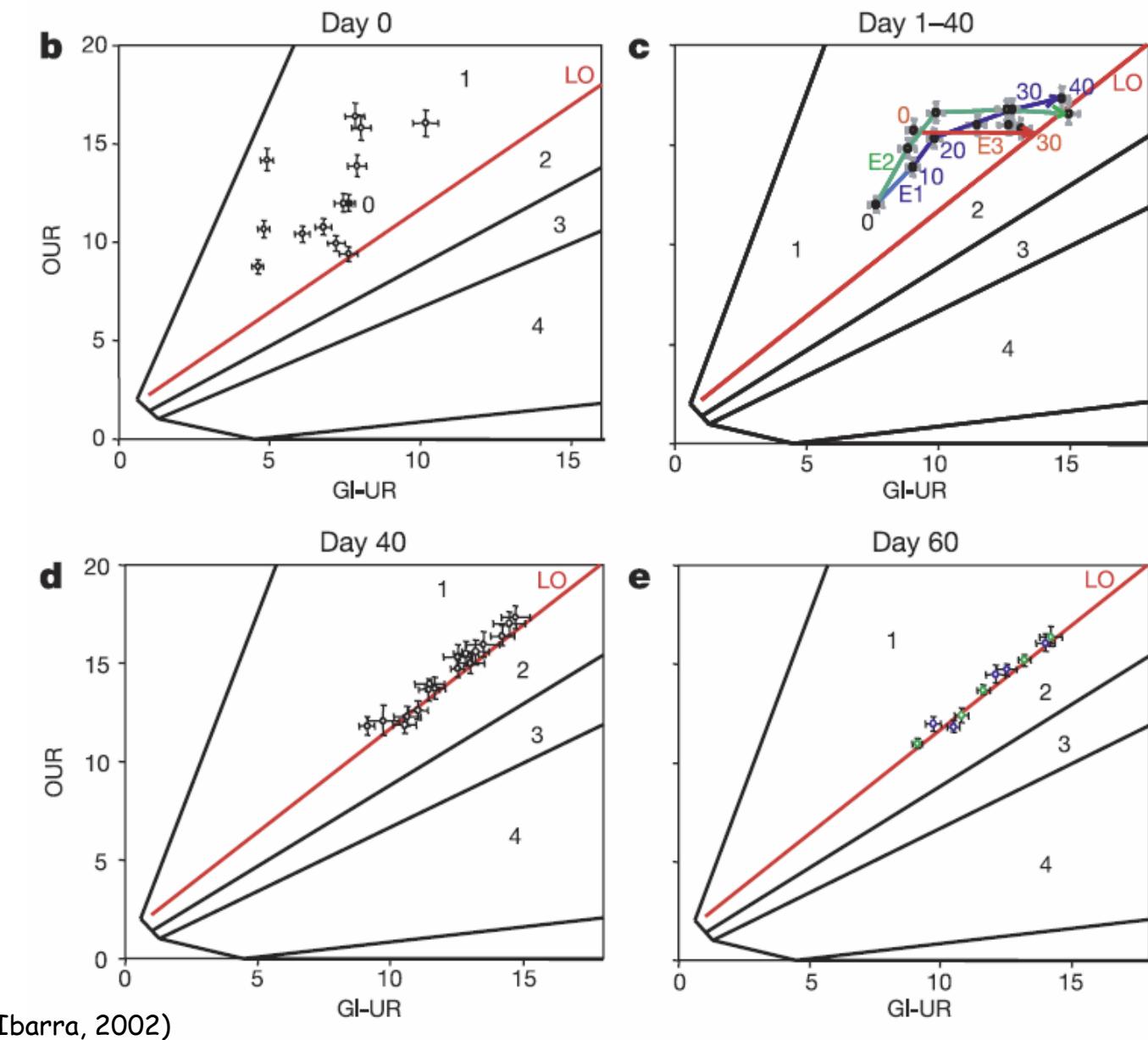
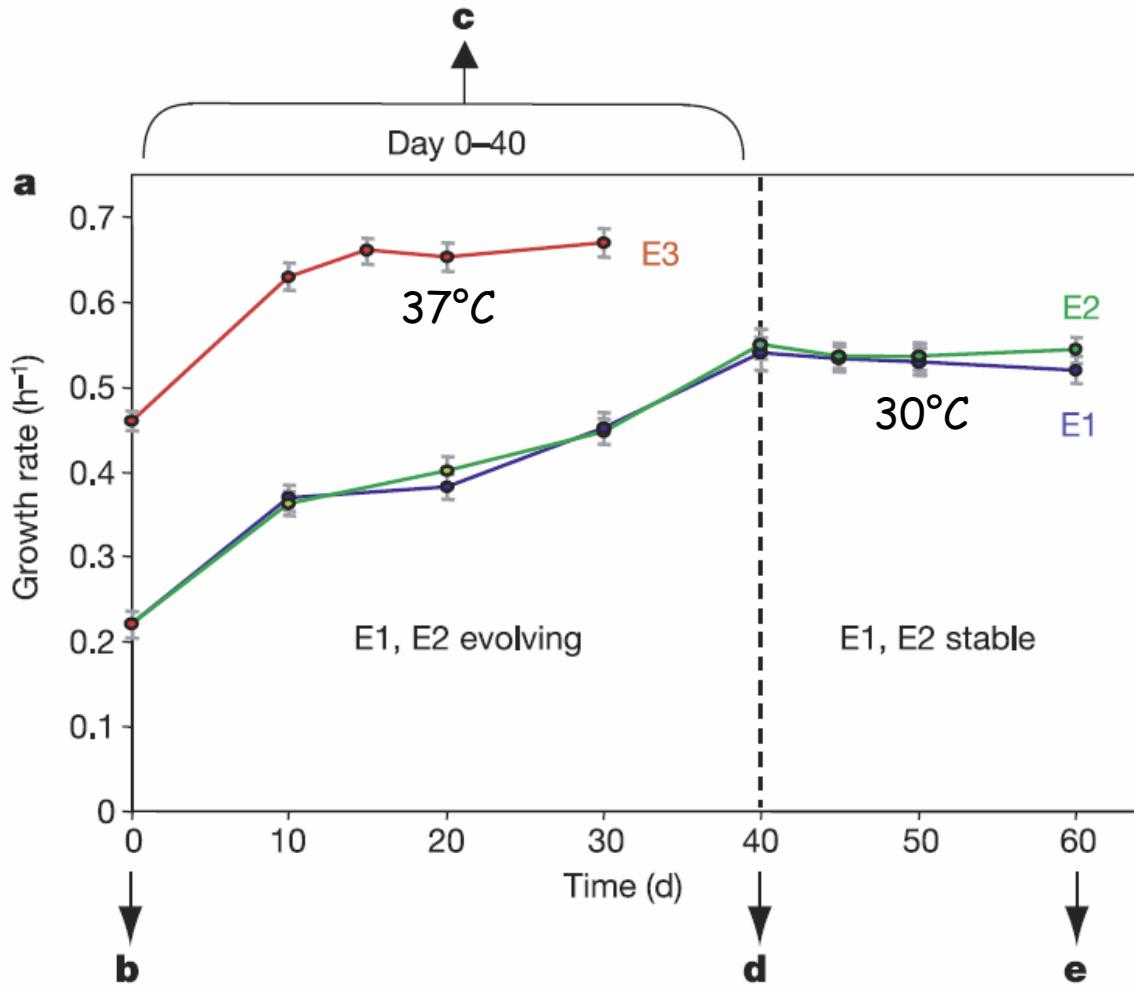
Growth of *E.coli* K-12 on Glucose

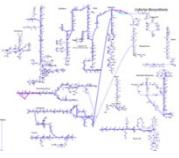


Ibarra, R. U., J. S. Edwards, et al. (2002). "Escherichia coli K-12 undergoes adaptive evolution to achieve *in silico* predicted optimal growth." Nature 420(6912): 186-189.



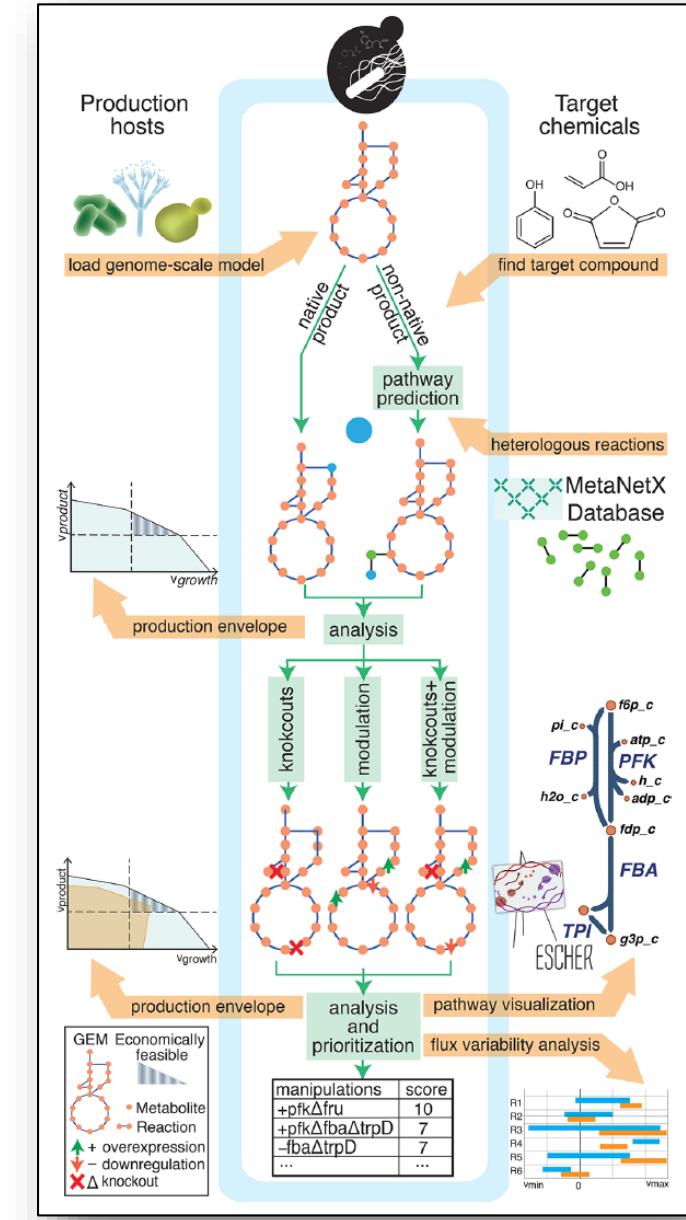
Growth of *E.coli* K-12 on Glycerol

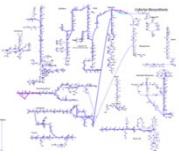




Strain Design

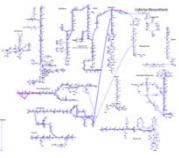
- Introduction
- Production Hosts
- Target Chemicals
- Heterologous Pathways
- Recombinant Proteins
- Gene/Reaction Knockouts
- Gene Expression Modulation
- Adaptive Laboratory Evolution





Review Questions

1. What is the difference between the CAMEO and the COBRApy Python packages?
2. What are heterologous pathways?
3. What are recombinant proteins?
4. What is the strain design process?
5. What are production hosts?
6. What are the reasons that *E.coli* is often used as a production host?
7. What are the categories of chemicals that can be produced by microbial cell factories?
8. What is the purpose of pathway prediction?
9. Why is it important to understand the practical production limits?
10. What is MOMA?
11. What is ROOM?
12. What is the line of optimality?
13. What is adaptive laboratory evolution?
14. What are the number of generations required for adaptive laboratory evolution?



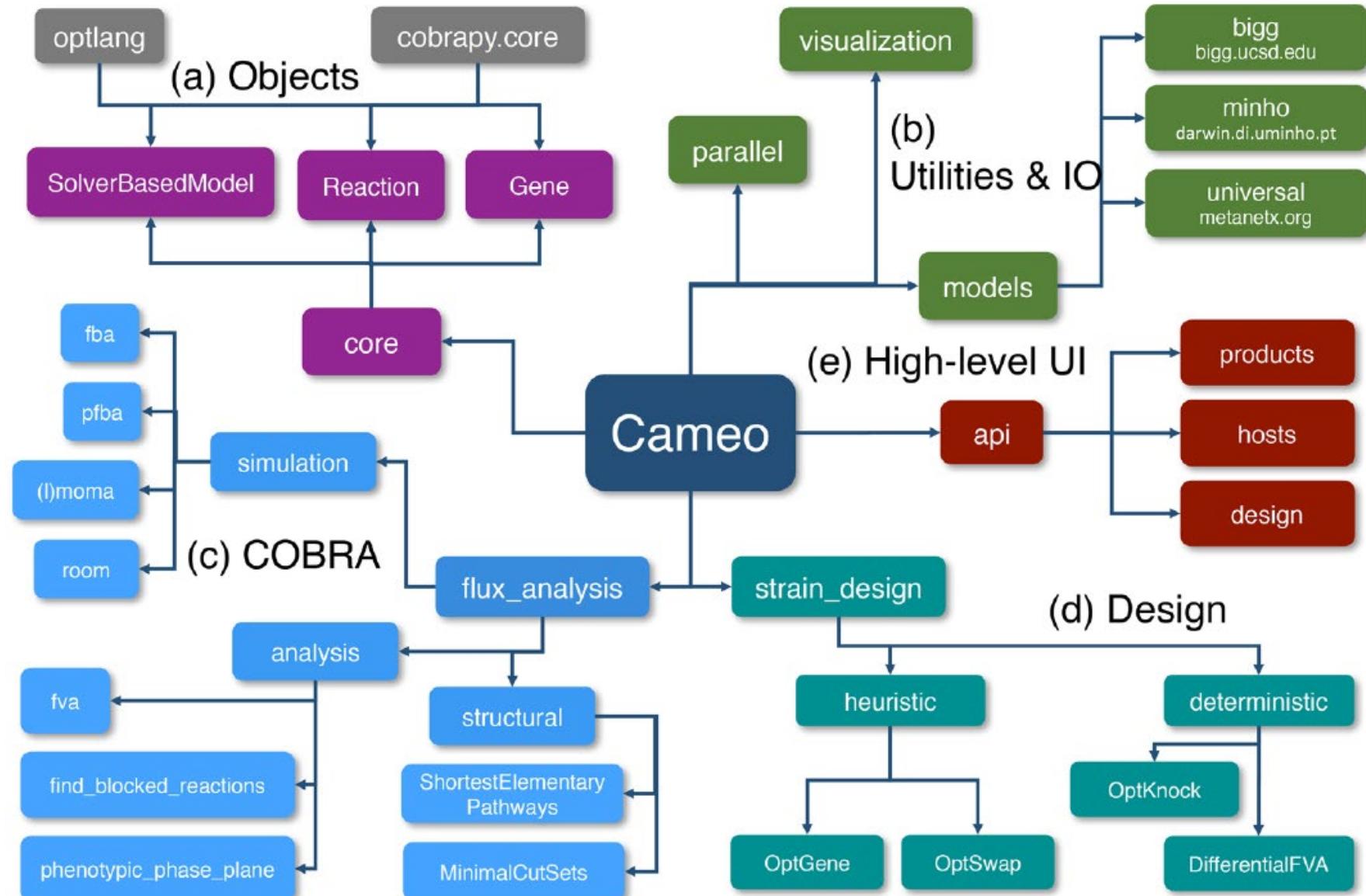
References

1. Cardoso, João GR, et al. "Cameo: a Python library for computer aided metabolic engineering and optimization of cell factories." *ACS synthetic biology* 7.4 (2018): 1163-1166.
2. Rosano, G. L. and E. A. Ceccarelli (2014). "Recombinant protein expression in *Escherichia coli*: advances and challenges." *Frontiers in microbiology* 5: 172.
3. Monk, J. M., P. Charusanti, et al. (2013). *Proceedings of the National Academy of Sciences of the United States of America* 110(50): 20338-20343.
4. Lee, J. W., D. Na, et al. (2012). "Systems metabolic engineering of microorganisms for natural and non-natural chemicals." *Nature chemical biology* 8(6): 536-546.
5. Chen, D., Brown, J.D., Kawasaki, Y. et al. Scalable production of biliverdin IXa by *Escherichia coli*. *BMC Biotechnol* 12, 89 (2012).
6. Pharkya, P., A. P. Burgard, et al. (2004). "OptStrain: a computational framework for redesign of microbial production systems." *Genome research* 14(11): 2367-2376.
7. Chou, C. H., W. C. Chang, et al. (2009). "FMM: a web server for metabolic pathway reconstruction and comparative analysis." *Nucleic Acids Res* 37(Web Server issue): W129-134.
8. Choi, Hyung Seok, et al. "In silico identification of gene amplification targets for improvement of lycopene production." *Applied and environmental microbiology* 76.10 (2010): 3097-3105.
9. Choi, J. H., K. C. Keum, et al. (2006). "Production of recombinant proteins by high cell density culture of *Escherichia coli*." *Chemical Engineering Science* 61(3): 876-885
10. Shiloach, J. and R. Fass (2005). "Growing *E. coli* to high cell density--a historical perspective on method development." *Biotechnol Adv* 23(5): 345-357
11. Lee, S. Y. (1996). "High cell-density culture of *Escherichia coli*." *Trends Biotechnol* 14(3): 98-105.
12. Riesenbeck, D. (1991). "High-cell-density cultivation of *Escherichia coli*." *Current opinion in biotechnology* 2(3): 380-384.
13. Goodarzi, H. (2010). "Regulatory and metabolic rewiring during laboratory evolution of ethanol tolerance in *E. coli*." *Molecular Systems Biology*, 6, 378
14. Segre, D., D. Vitkup, et al. (2002). "Analysis of optimality in natural and perturbed metabolic networks." *Proceedings of the National Academy of Sciences of the United States of America* 99(23): 15112-15117.
15. Price, N. D., J. L. Reed, et al. (2004). "Genome-scale models of microbial cells: evaluating the consequences of constraints." *Nature reviews. Microbiology* 2(11): 886-897.
16. Tomer Shlomi, Omer Berkman and Eytan Ruppin, "Regulatory on/off minimization of metabolic flux changes after genetic perturbations", PNAS 2005 102 (21) 7695-7700;
17. Dragosits, M. and D. Mattanovich (2013). "Adaptive laboratory evolution -- principles and applications for biotechnology." *Microbial cell factories* 12: 64.
18. Conrad, T. M., N. E. Lewis, et al. (2011). "Microbial laboratory evolution in the era of genome-scale science." *Molecular Systems Biology* 7: 509.
19. Ibarra, R. U., J. S. Edwards, et al. (2002). "Escherichia coli K-12 undergoes adaptive evolution to achieve in silico predicted optimal growth." *Nature* 420(6912): 186-189

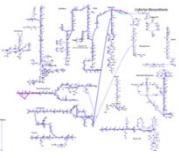


CAMEO Toolbox

- Enhances the COBRApy tools
- Includes tutorials
 - ✓ <https://cameo.bio/tutorials.html>
 - ✓ [Jupyter Notebook Viewer \(nbviewer.org\)](https://nbviewer.org/)
 - ✓ [GitHub - biosustain/cameo-notebooks: Example notebooks for Cameo](https://github.com/biosustain/cameo-notebooks)
- Repository
 - ✓ [GitHub - biosustain/cameo: cameo - computer aided metabolic engineering & optimization](https://github.com/biosustain/cameo-computer-aided-metabolic-engineering-and-optimization)



Cardoso, João GR, et al. "Cameo: a Python library for computer aided metabolic engineering and optimization of cell factories." *ACS synthetic biology* 7.4 (2018): 1163-1166.



E. coli Genome-scale Reconstructions

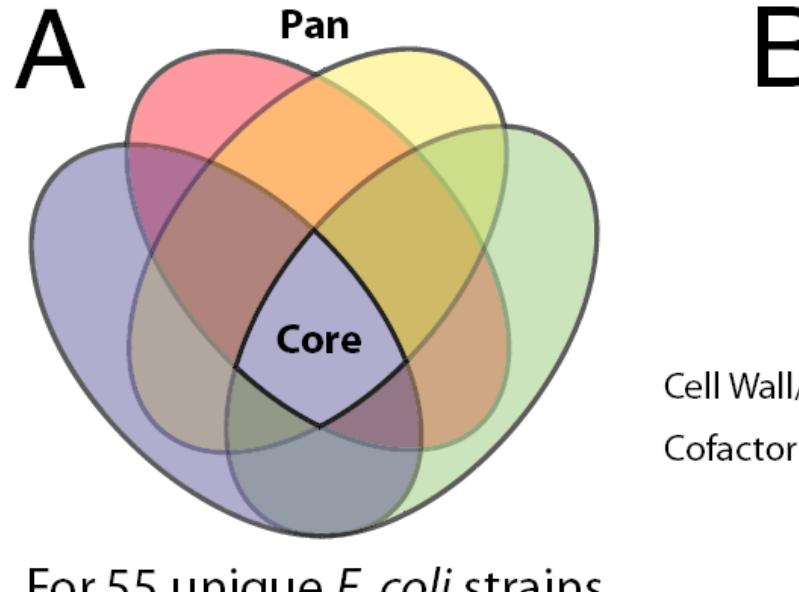
- Escherichia coli 042
- Escherichia coli 536
- Escherichia coli 55989
- Escherichia coli ABU 83972
- Escherichia coli APEC O1
- Escherichia coli ATCC 8739
- Escherichia coli B str. REL606
- Escherichia coli BL21(DE3) AM946981
- Escherichia coli BL21(DE3) BL21-Gold(DE3)pLySS AG
- **Escherichia coli BL21(DE3) CP001509**

**Strain used
at USU**
- Escherichia coli BW2952
- Escherichia coli CFT073
- Escherichia coli DH1
- Escherichia coli DH1 ME8569
- Escherichia coli E24377A
- Escherichia coli ED1a
- Escherichia coli ETEC H10407
- Escherichia coli HS
- Escherichia coli IAI1
- Escherichia coli IAI39
- Escherichia coli IHE3034
- Escherichia coli KO11FL
- Escherichia coli LF82
- Escherichia coli NA114
- Escherichia coli O103:H2 str. 12009
- Escherichia coli O111:H- str. 11128
- Escherichia coli O127:H6 str. E2348/69
- Escherichia coli O157:H7 EDL933
- Escherichia coli O157:H7 str. EC4115
- Escherichia coli O157:H7 str. Sakai
- Escherichia coli O157:H7 str. TW14359
- Escherichia coli O26:H11 str. 11368
- Escherichia coli O55:H7 str. CB9615
- Escherichia coli O83:H1 str. NRG 857C
- Escherichia coli S88
- Escherichia coli SE11
- Escherichia coli SE15
- Escherichia coli SMS-3-5
- Escherichia coli str. K-12 substr. DH10B
- Escherichia coli str. K-12 substr. MG1655
- Escherichia coli str. K-12 substr. W3110
- Escherichia coli UM146
- Escherichia coli UMN026
- Escherichia coli UMNK88
- Escherichia coli UTI89
- Escherichia coli W
- Escherichia coli W CP002185
- Escherichia coli K-12 MG1655

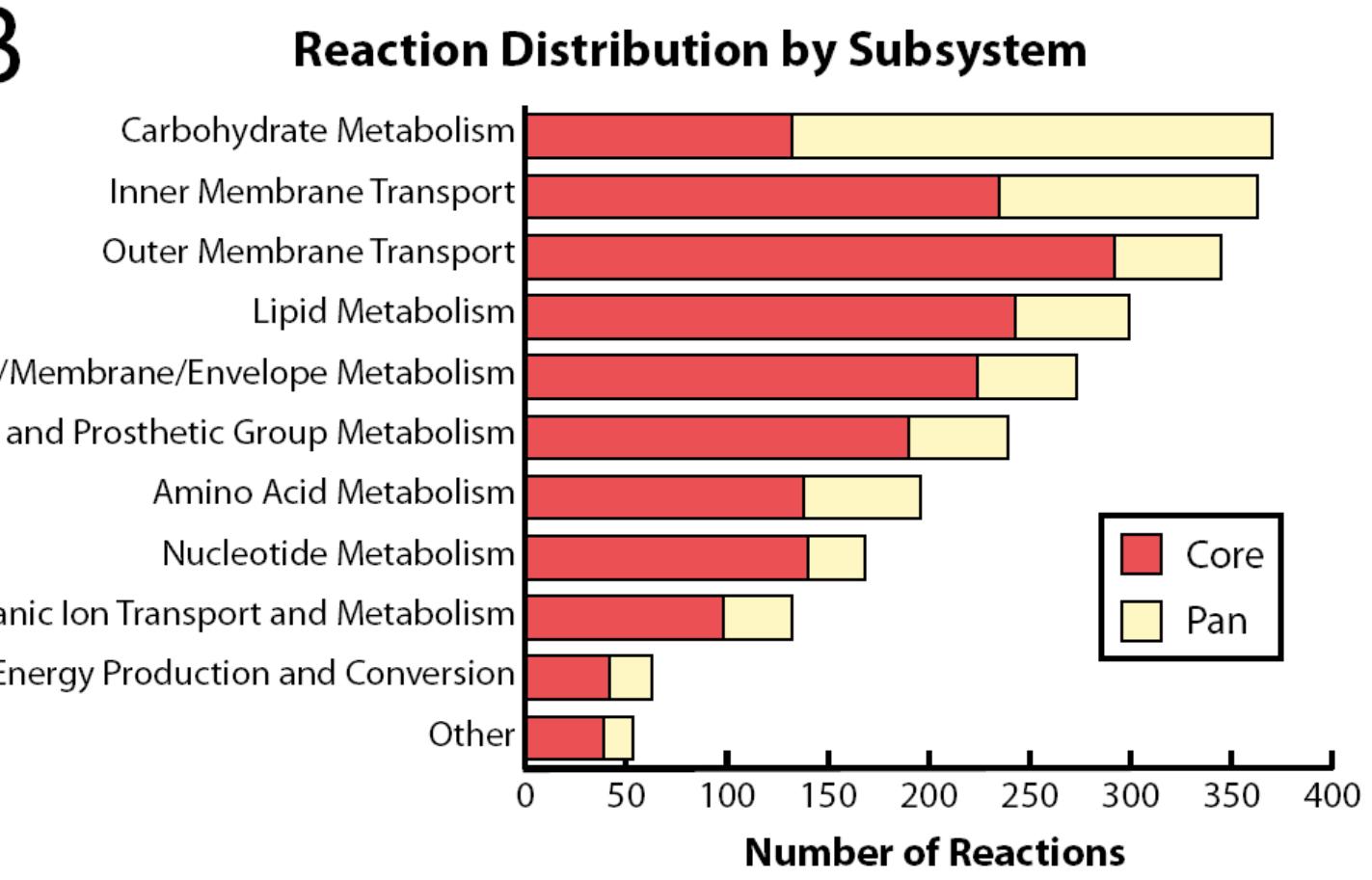
Monk, J. M., P. Charusanti, et al. (2013). Proceedings of the National Academy of Sciences of the United States of America 110(50): 20338-20343.



Core and Pan Metabolic Capabilities of the *E. coli* Species.



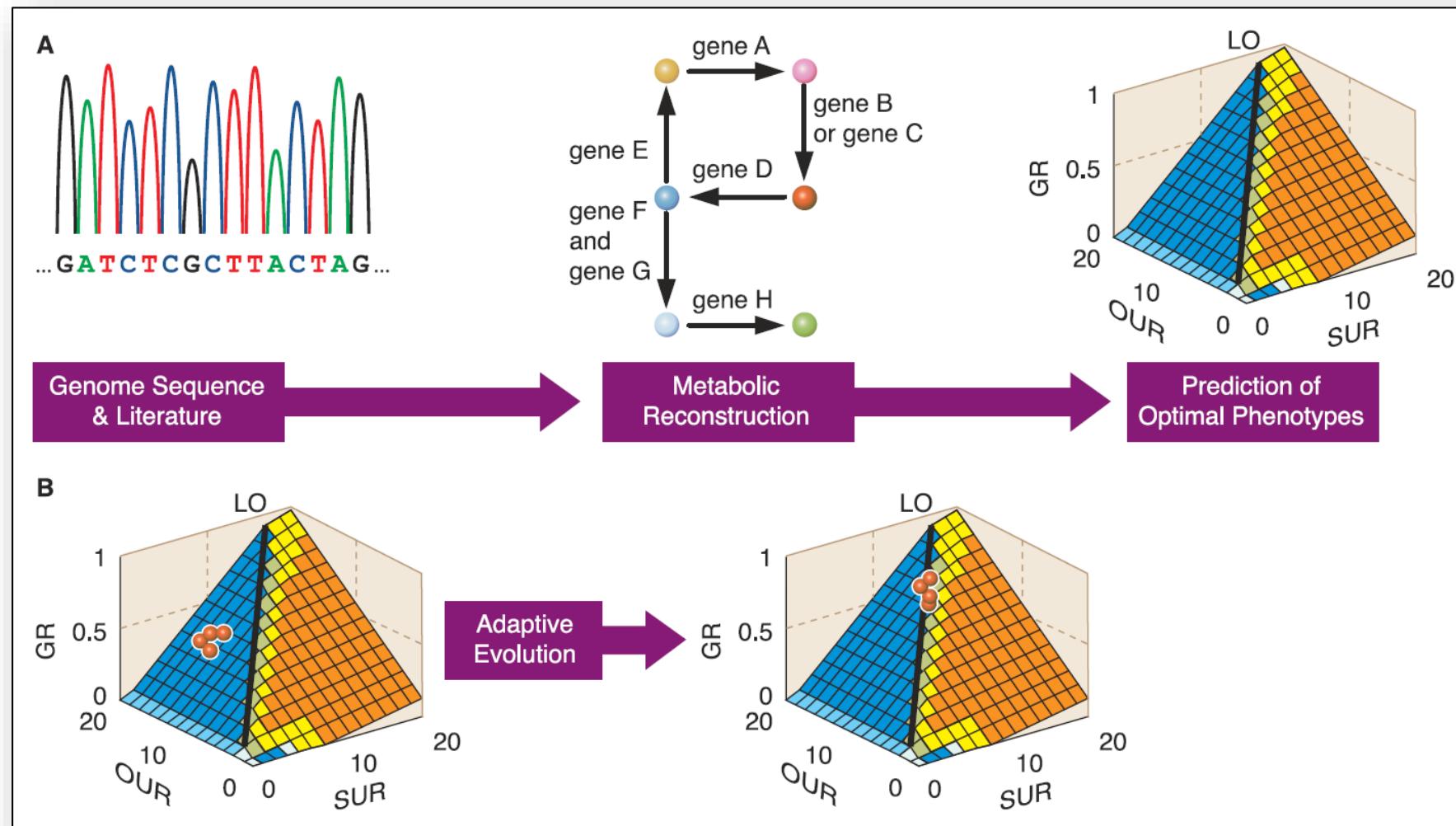
	Core:	Pan:
Genes:	965	1,460
Reactions:	1,773	2,501
Metabolites:	1,665	2,043



Monk, J. M., P. Charusanti, et al. (2013). Proceedings of the National Academy of Sciences of the United States of America 110(50): 20338-20343.



Adaptive Laboratory Evolution



B. Palsson (2010). "Adaptive Laboratory Evolution." *Microbe*, 6(2):69-74