Gene/Reaction Modulation

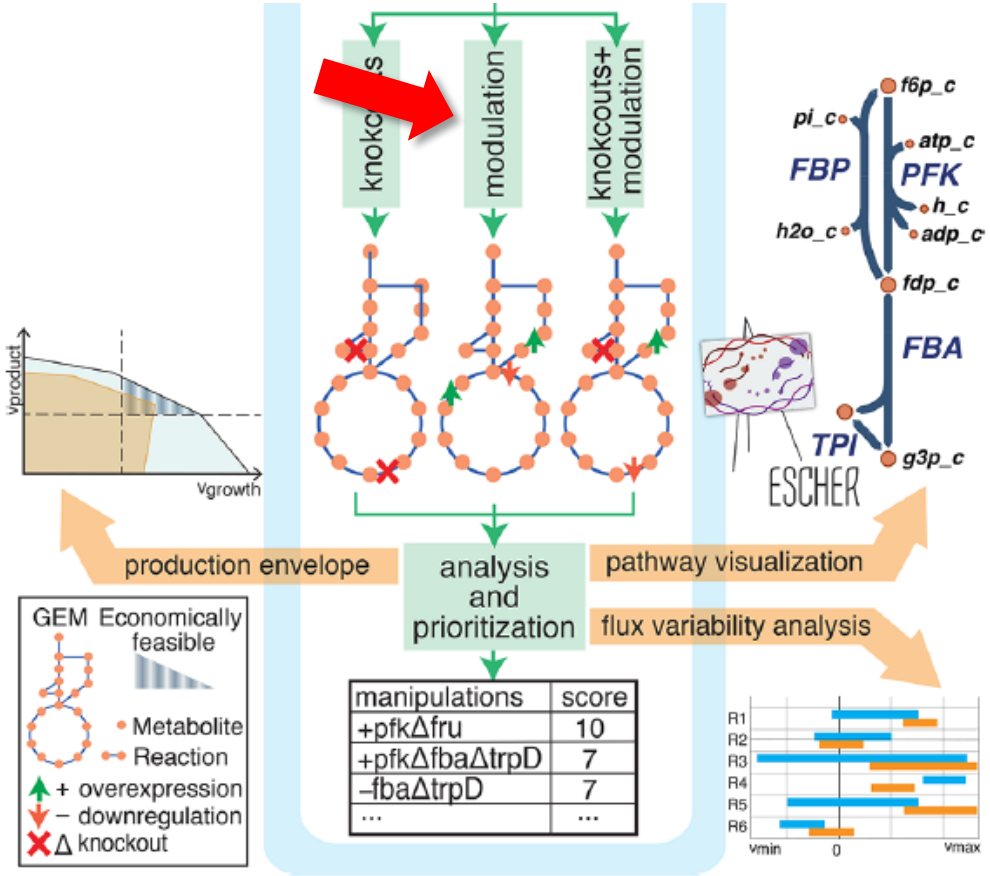
Learning Objectives

Each student should be able to:

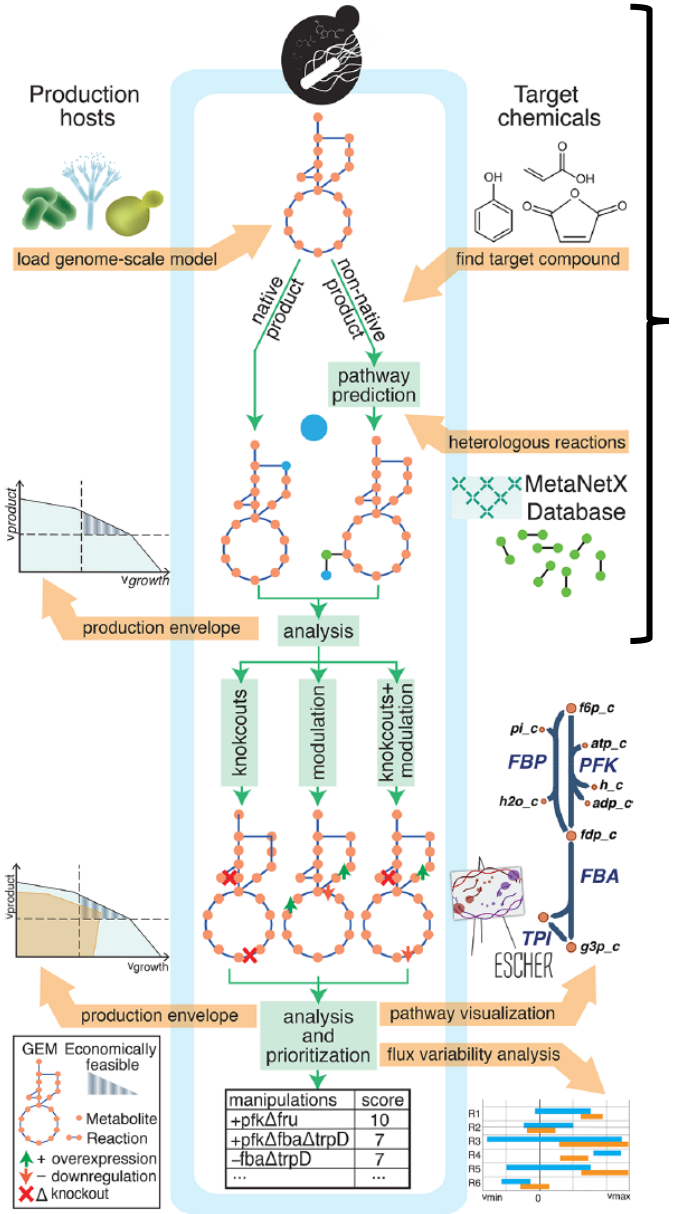
- Describe gene modulation.
- Explain the purpose of gene modulation.
- Explain "Flux Scanning based on Enforced Objective Flux" (FSEOF).
- Explain "Differential Flux Variability Analysis" (dFVA).

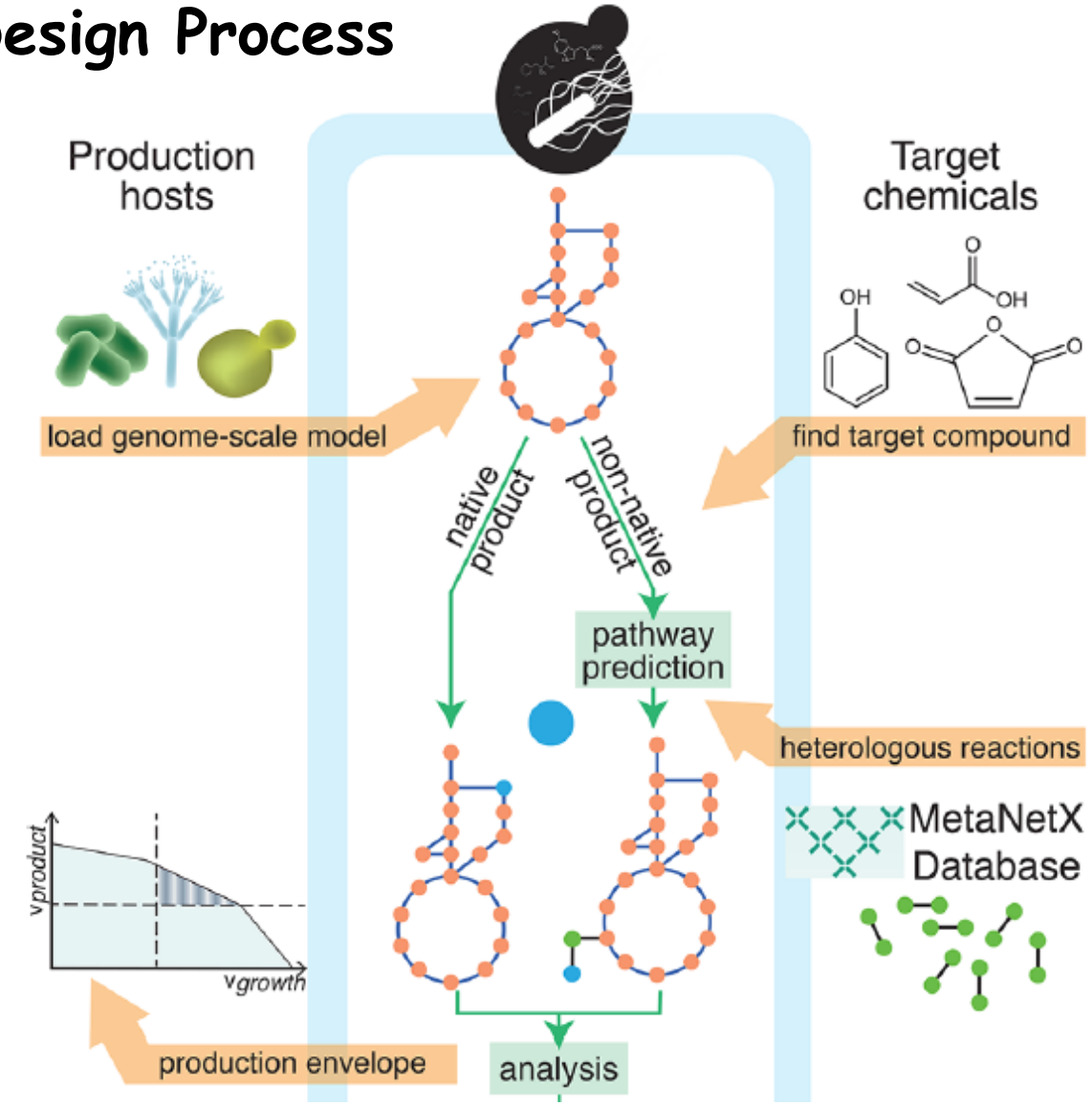
Gene/Reaction Expression Modulation

- Introduction
- Flux Scanning based on Enforced Objective Flux (FSEOF)
- Differential flux variability analysis (dFVA)



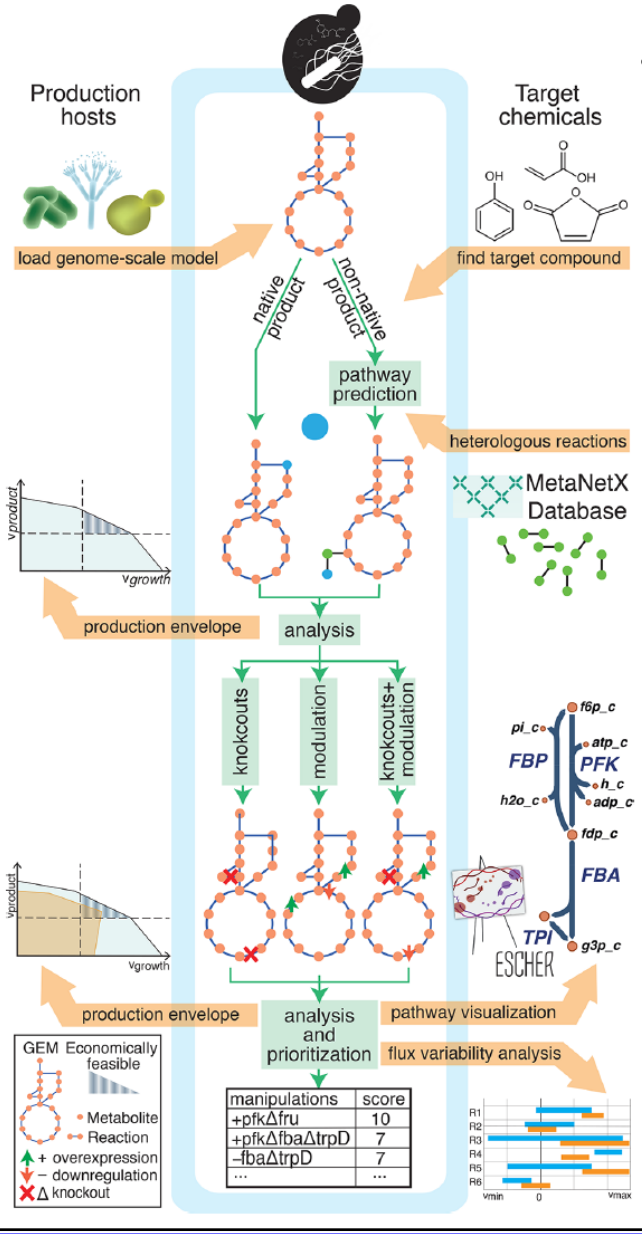
Strain Design Process



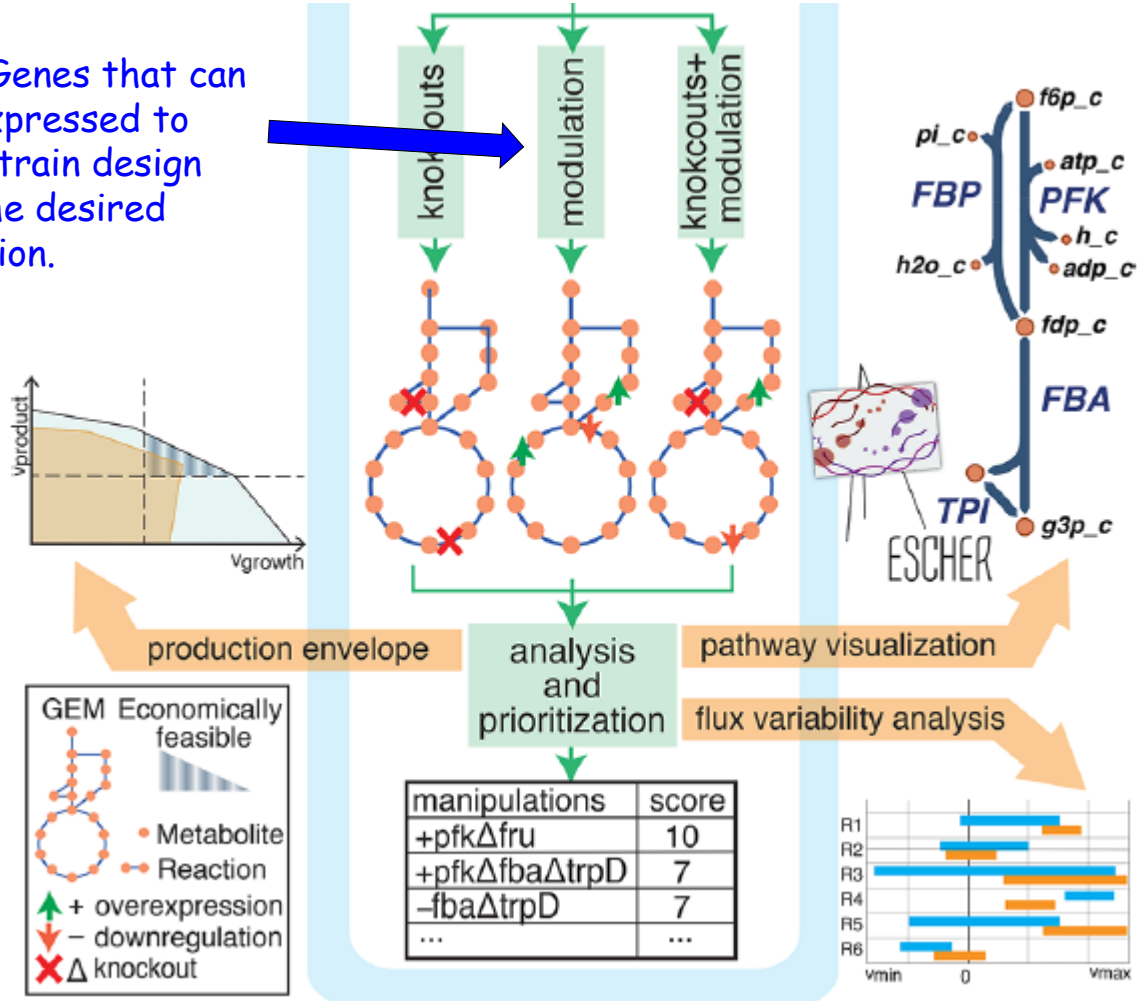


Cardoso, João GR, et al. "Cameo: a Python library for computer aided metabolic engineering and optimization of cell factories." *ACS synthetic biology* 7.4 (2018): 1163-1166.

Strain Design Process



Gene modulation: Genes that can be over or under expressed to help create a new strain design that can improve the desired bioproduct production.



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.

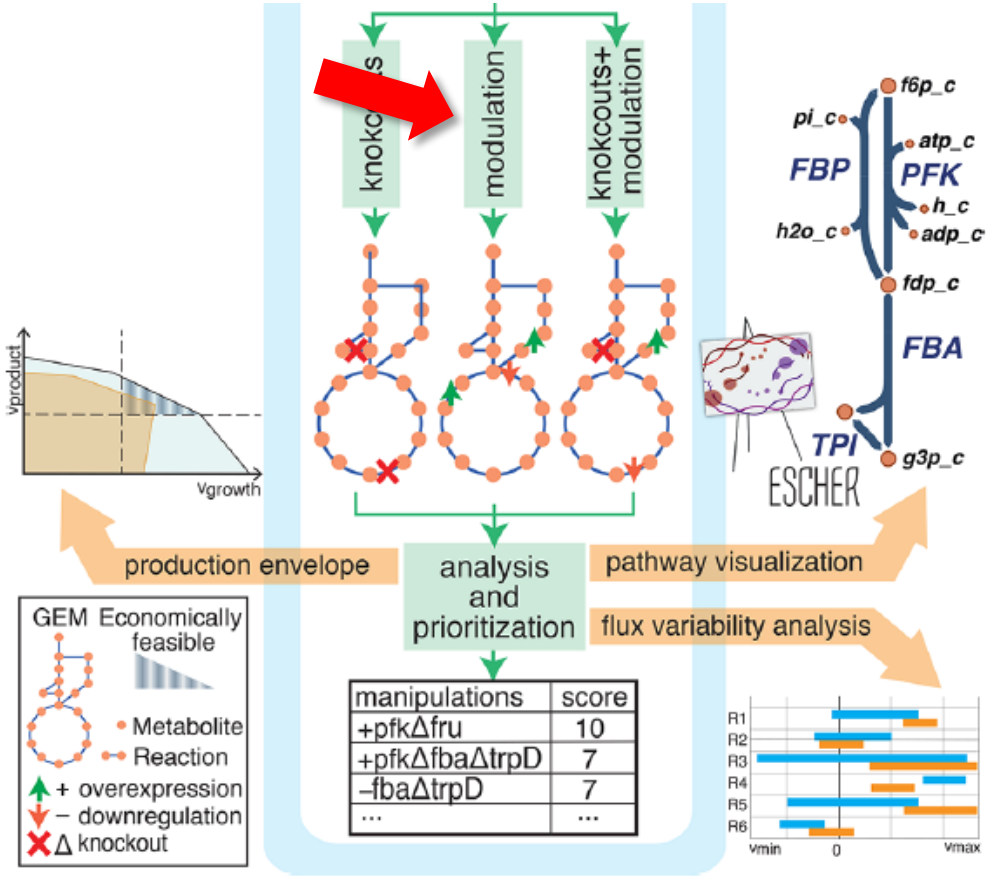
Strain Optimization

- Codon optimization of poorly expressed genes,
- Promoter supplementation,
- Altered operon order,
- Changes in plasmid copy number.

Brunk, Elizabeth, et al. "Characterizing strain variation in engineered *E. coli* using a multi-omics-based workflow." Cell systems 2.5 (2016): 335-346.

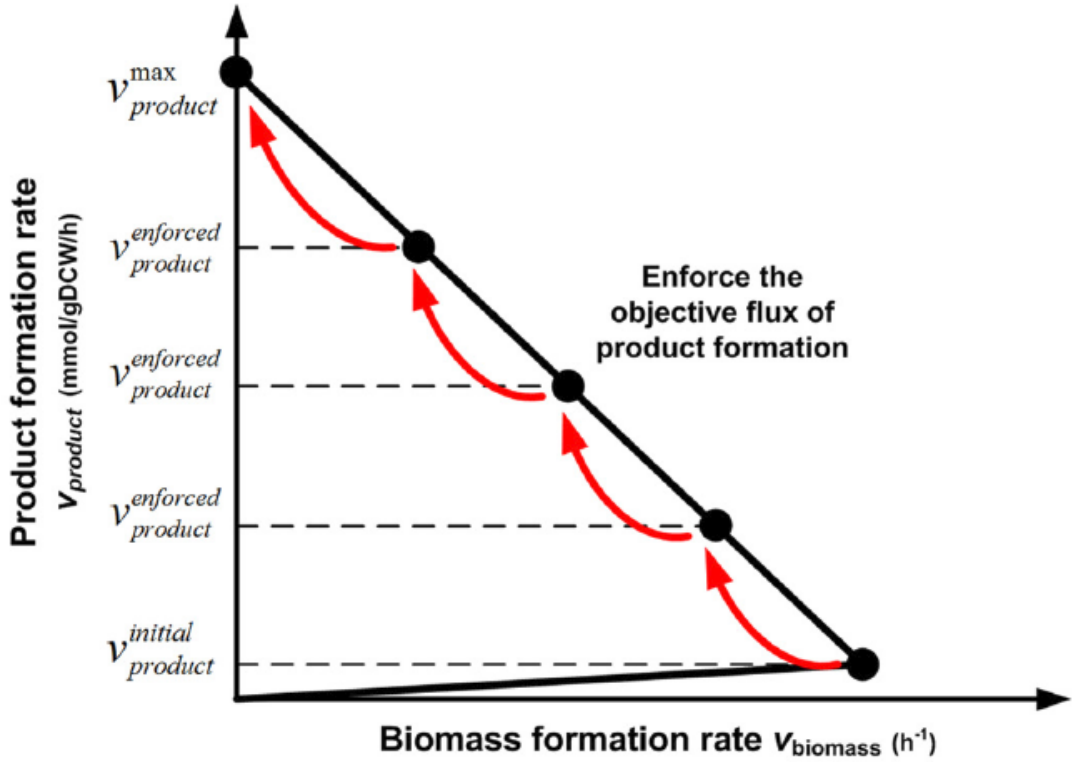
Gene/Reaction Expression Modulation

- Introduction
- Flux Scanning based on Enforced Objective Flux (FSEOF)
- Differential flux variability analysis (dFVA)



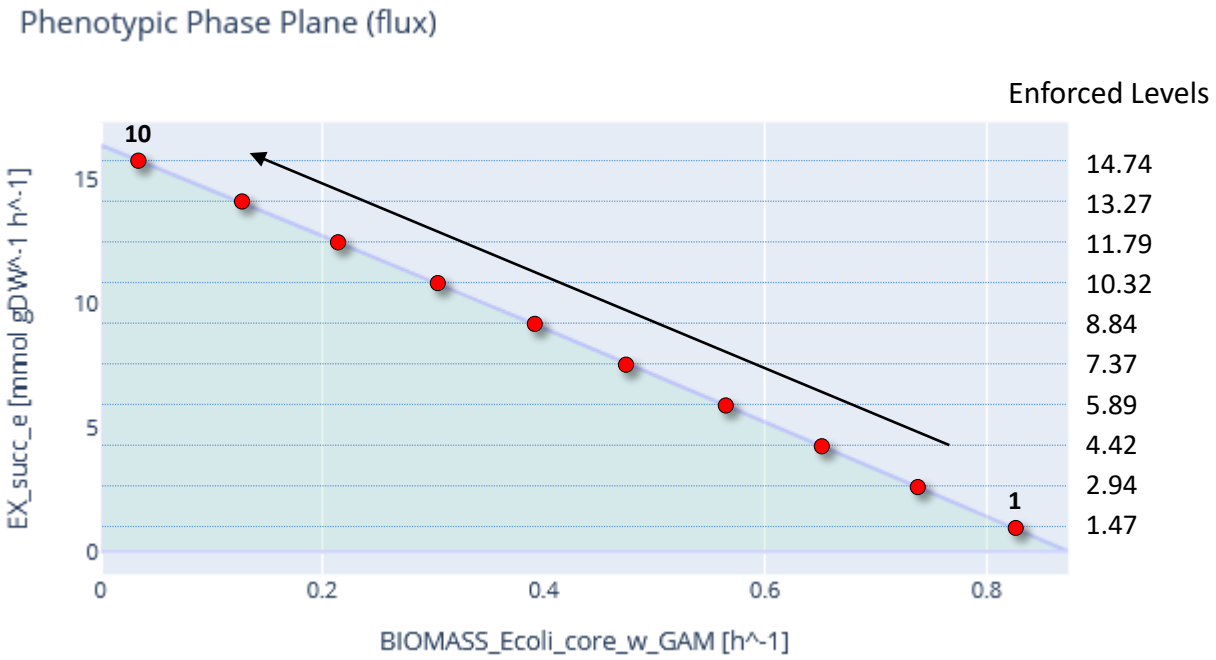
Flux Scanning based on Enforced Objective Flux (FSEOF)

- FSEOF searches for the candidate fluxes to be amplified through scanning for fluxes that increase with an enforced objective (product formation) flux under the objective function of maximizing biomass formation flux.
- During the FSEOF implementation, the three types of intracellular flux profiles are identified as increased, decreased, and unchanged. Among them, FSEOF identifies the fluxes that show an ***increased profile as the primary amplification targets***. In principle, it identifies the reactions that are coupled to the objective reaction.
- FSEOF is performed by maximizing cell growth while the bioproduct production rate (our actual objective) is gradually increased (enforced) from the initial flux value to a value adjacent to the theoretical maximum value for product formation (typically 90% of the maximum theoretical or higher value). 90% was chosen because the flux distributions obtained by constraint-based flux analysis, which represent the theoretical maximum production rate, are usually unrealistic since the biomass formation rate often becomes zero when the product formation objective function is set to its maximization during the simulation.



H. S. Choi, S. Y. Lee, T. Y. Kim, and H. M. Woo, '*In silico* identification of gene amplification targets for improvement of lycopene production.,' Appl Environ Microbiol, vol. 76, no. 10, pp. 3097–3105, May 2010.

Identify The Reactions That Increase With The Enforced Objective Levels



	1	2	3	4	5	6	7	8	9	10
ADK1	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.410338	1.527283
ENO	15.238892	15.761645	16.212407	16.577535	16.942662	17.307790	17.704415	18.602292	19.293292	19.628048
EX_succ_e	1.474575	2.949150	4.423725	5.898300	7.372875	8.847450	10.322025	11.796600	13.271175	14.745750
FBA	7.756941	8.036499	8.242515	8.361048	8.479582	8.598115	8.747793	9.393102	9.830787	9.910940
FRD7	0.000000	0.000000	0.000000	1.453235	3.175302	4.897369	6.586703	7.755141	9.127842	10.852285
GAPD	16.431132	16.838737	17.173618	17.421989	17.670359	17.918730	18.198431	18.976723	19.547784	19.761991
ICL	0.000000	0.000000	0.386057	1.231342	2.076627	2.921912	3.735322	4.041459	4.143333	3.893465
MALS	0.000000	0.000000	0.386057	1.231342	2.076627	2.921912	3.735322	4.041459	4.143333	3.893465
PDH	8.809270	8.336008	8.179868	8.400957	8.622045	8.843134	9.064539	9.290984	9.107760	8.219084
PFK	7.756941	8.036499	8.242515	8.361048	8.479582	8.598115	8.747793	9.393102	9.830787	9.910940
PGI	5.485630	6.110399	6.513173	6.651876	6.790579	6.929282	7.161108	8.874882	9.965126	9.981645
PGK	-16.431132	-16.838737	-17.173618	-17.421989	-17.670359	-17.918730	-18.198431	-18.976723	-19.547784	-19.761991
PGM	-15.238892	-15.761645	-16.212407	-16.577535	-16.942662	-17.307790	-17.704415	-18.602292	-19.293292	-19.628048
PPC	3.758320	5.012330	5.878875	6.284516	6.690157	7.095799	7.532996	8.472367	9.615323	11.108854
PPS	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.410338	1.527283
SUCCt3	1.474575	2.949150	4.423725	5.898300	7.372875	8.847450	10.322025	11.796600	13.271175	14.745750
THD2	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.119674	2.143848	2.916807	1.535161
TPI	7.756941	8.036499	8.242515	8.361048	8.479582	8.598115	8.747793	9.393102	9.830787	9.910940

FSEOF.ipynb

Reaction Fluxes that Change

Potential Overexpressed and Underexpressed Reactions

Potential Overexpressed Reactions

	1	2	3	4	5	6	7	8	9	10	Difference
EX_succ_e	1.47	2.95	4.42	5.90	7.37	8.85	10.32	11.80	13.27	14.75	13.27
SUCCt3	1.47	2.95	4.42	5.90	7.37	8.85	10.32	11.80	13.27	14.75	13.27
FRD7	0.00	0.00	0.00	1.45	3.18	4.90	6.59	7.76	9.13	10.85	10.85
PPC	3.76	5.01	5.88	6.28	6.69	7.10	7.53	8.47	9.62	11.11	7.35
PGI	5.49	6.11	6.51	6.65	6.79	6.93	7.16	8.87	9.97	9.98	4.50
ENO	15.24	15.76	16.21	16.58	16.94	17.31	17.70	18.60	19.29	19.63	4.39
ICL	0.00	0.00	0.39	1.23	2.08	2.92	3.74	4.04	4.14	3.89	3.89
MALS	0.00	0.00	0.39	1.23	2.08	2.92	3.74	4.04	4.14	3.89	3.89
GAPD	16.43	16.84	17.17	17.42	17.67	17.92	18.20	18.98	19.55	19.76	3.33
PFK	7.76	8.04	8.24	8.36	8.48	8.60	8.75	9.39	9.83	9.91	2.15
TPI	7.76	8.04	8.24	8.36	8.48	8.60	8.75	9.39	9.83	9.91	2.15
FBA	7.76	8.04	8.24	8.36	8.48	8.60	8.75	9.39	9.83	9.91	2.15
THD2	0.00	0.00	0.00	0.00	0.00	0.00	0.12	2.14	2.92	1.54	1.54
PPS	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.41	1.53	1.53
ADK1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.41	1.53	1.53

Potential Underexpressed Reactions

	1	2	3	4	5	6	7	8	9	10	Difference
PGM	-15.24	-15.76	-16.21	-16.58	-16.94	-17.31	-17.70	-18.60	-19.29	-19.63	-4.39
PGK	-16.43	-16.84	-17.17	-17.42	-17.67	-17.92	-18.20	-18.98	-19.55	-19.76	-3.33
PDH	8.81	8.34	8.18	8.40	8.62	8.84	9.06	9.29	9.11	8.22	-0.59

Reversible reactions operating in reverse, thus they need to be overexpressed

FSEOF.ipynb

Flux Scanning based on Enforced Objective Flux: Predicting Expression Modulation Targets

Cameo's Flux Scanning based on Enforced Objective Flux (FSEOF) provides algorithms to search for genes or reactions that can be over- or down-regulated in order to achieve a given biological objective.

In [1]:
```
from cameo import models
from cameo.visualization.plotting.with_plotly import PlotlyPlotter
import pandas as pd
pd.set_option('display.max_colwidth',None)
pd.set_option('display.max_rows', 1000)
```

Load the *E. coli* core model.

In [2]:
```
model = models.bigg.e_coli_core
plotter = PlotlyPlotter()
```

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

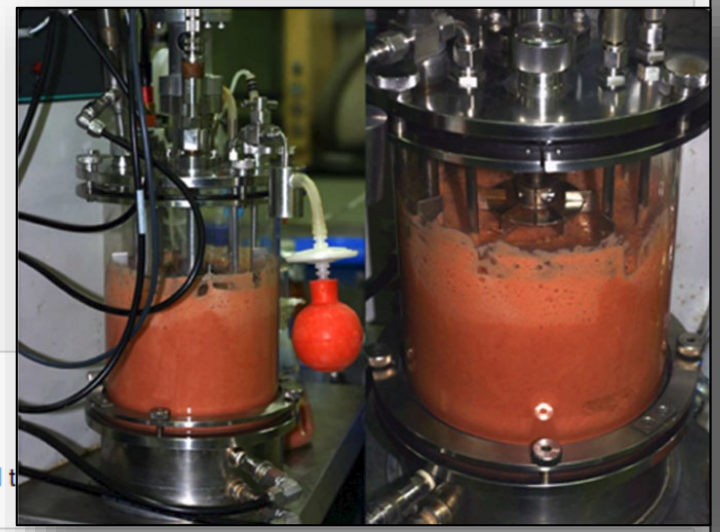
FSEOF searches for the candidate fluxes to be amplified through scanning for those fluxes that increase with enforced objective (product formation) flux under the objective function of maximizing biomass formation flux. In this case, gene amplification refers to an increase in the number of copies of a gene sequence. During the FSEOF implementation, three types of intracellular flux profiles are typically identified increased, decreased, and unchanged. Among them, FSEOF identifies the fluxes showing the increased profile as the primary amplification targets. In principle, it identifies the reactions that are coupled to the objective reaction.[1]

FSEOF.ipynb

Flux Scanning based on Enforced Objective Flux: Lycopene Production

Cameo's Flux Scanning based on Enforced Objective Flux (FSEOF) provides algorithms to search for genes or reactions that can be over- or down-regulated in order to achieve a given biological objective. This notebook will explore what reactions/genes can be over- or underexpressed to increase the production of lycopene.

In [1]:
```
import cobra.test
from cobra.flux_analysis import flux_variability_analysis
from cobrapy_bigg_client import client
from cameo import models
from cameo.visualization.plotting.with_plotly import PlotlyPlotter
from cameo import phenotypic_phase_plane
import pandas as pd
import numpy as np
import escher
from escher import Builder
plotter = PlotlyPlotter()
pd.set_option('display.max_colwidth',None)
pd.set_option('display.max_rows', 1000)
```



Load the model

Load the 'iJO1366_lycopene.json' model. This is model had a seven reaction pathway added t

In [2]:
```
model_orig = cobra.io.load_json_model("iJO1366_lycopene.json")
#model_orig.reactions.DM_lycop_c.bounds = 0.5,1000 # Plasmid impact
model = model_orig.copy()
```

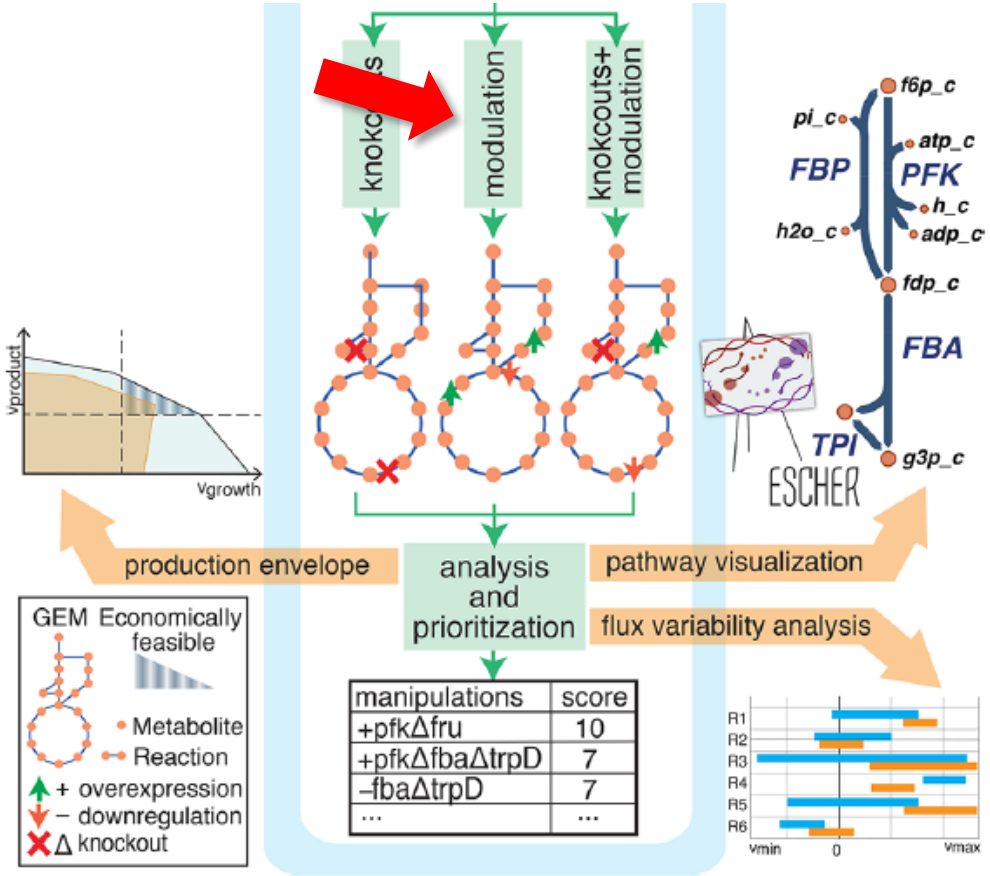
```
Set parameter Username
Academic license - for non-commercial use only - expires 2022-10-10
Read LP format model from file C:\Users\hinton\AppData\Local\Temp\tmph4rsydxu.lp
Reading time = 0.02 seconds
: 1812 rows, 5182 columns, 20436 nonzeros
```

Looking at a summary of the model's operation in it's native state.

FSOEF_lycopene.ipynb

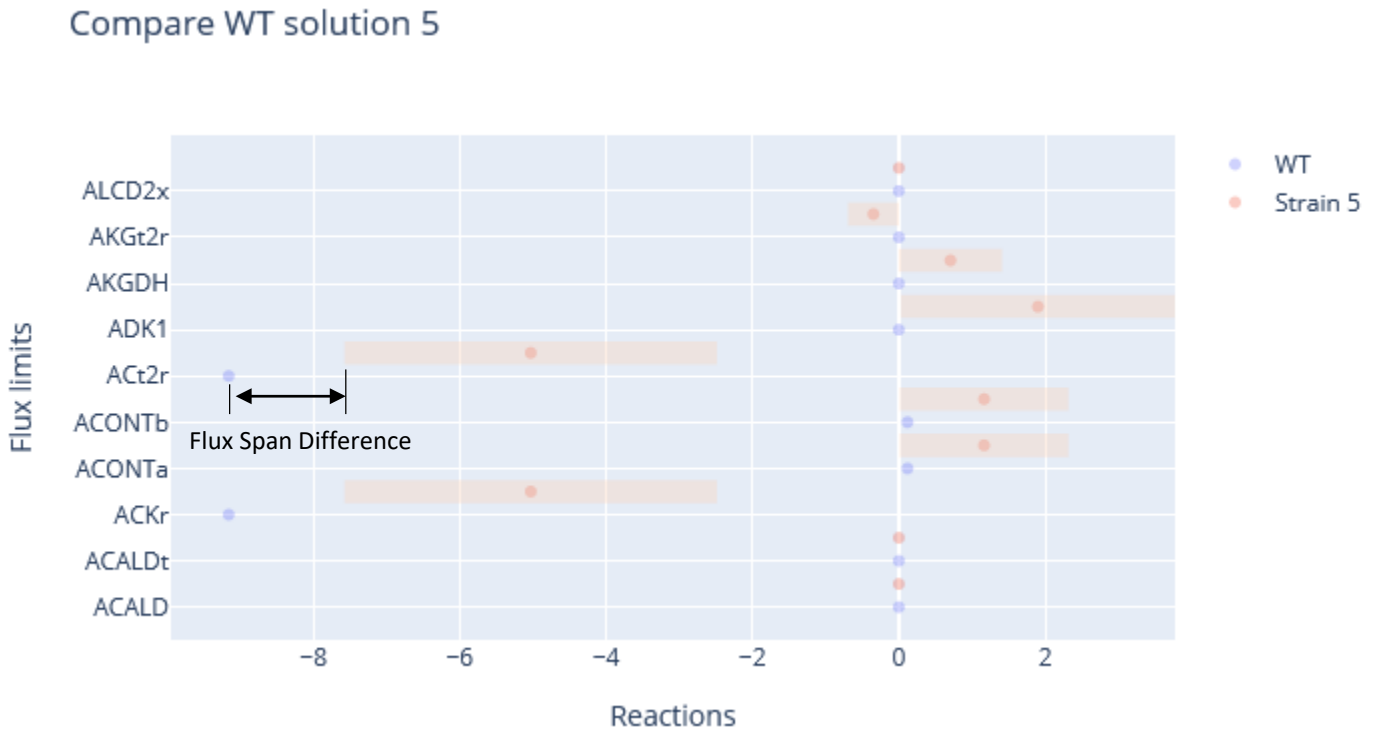
Gene/Reaction Expression Modulation

- Introduction
- Flux Scanning based on Enforced Objective Flux (FSEOF)
- Differential flux variability analysis (dFVA)

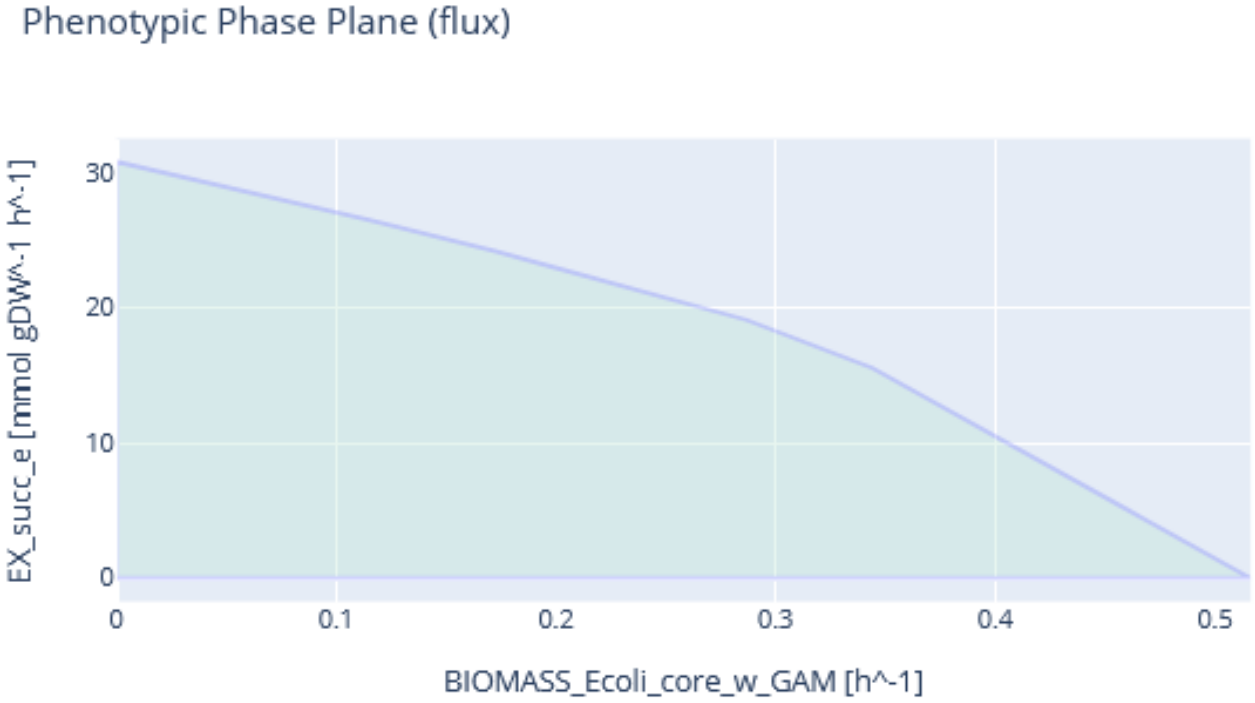


Differential Flux Variability Analysis

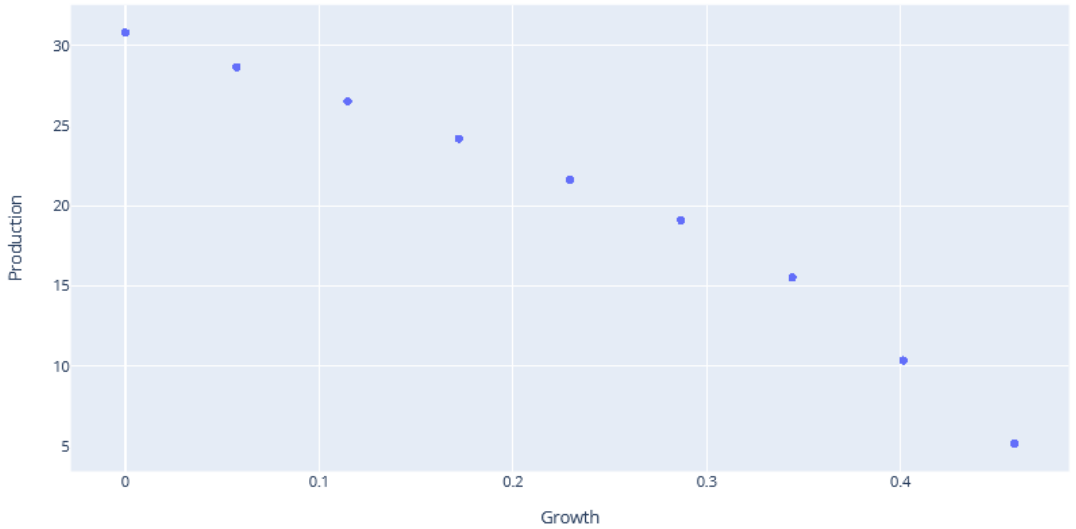
- The assumption is that the reactions in the new mutant, created by setting the target metabolite or reaction as the objective function, will have different flux spans than the original (reference) model that uses the biomass as the objective function.
  - ✓ The reference model represents the natural optimization process used by cells to maximize their growth.
  - ✓ The mutant model represents what reactions and pathways need to be changed to increase production of the target metabolite or reaction, even at the cost of decreased growth.
- By identifying these gaps between the different flux spans, the key reactions/enzymes can be altered to increase or decrease the flux to better match the desired mutants flux span.
  - ✓ The reactions identified for expression modulation are identified by large normalized gaps



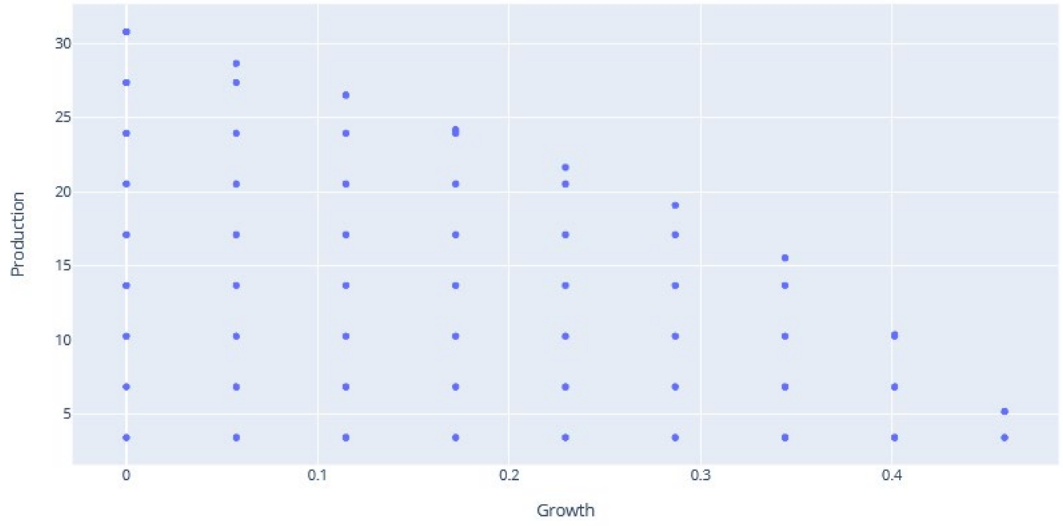
dFVA Sample Space



Production Envelope



Surface Samples



Grid Samples

Table Created by dFVA

solution 5

biomass: 0.229446

production: 21.6437

A table for each point on the production envelope

Underexpressed

reaction	lower_bound	upper_bound	gaps	normalized_gaps	biomass	production	KO	flux_reversal	suddenly_essential	free_flux	excluded
PYK	0.000000	0.000000	-16.107963	-31.201635	0.229446	21.643722	True	False	False	False	False
ACALD	0.000000	0.000000	-15.803528	-30.611935	0.229446	21.643722	True	False	False	False	False
ALCD2x	0.000000	0.000000	-15.803528	-30.611935	0.229446	21.643722	True	False	False	False	False
ETOHt2r	0.000000	0.000000	-15.803528	-30.611935	0.229446	21.643722	True	False	False	False	False
PFL	8.958866	8.958866	-25.686653	-28.063853	0.229446	21.643722	False	False	False	False	False
FORt	-8.958866	-8.958866	-25.686653	-28.063853	0.229446	21.643722	False	False	False	False	False
ATPS4r	-5.568458	-5.568458	-6.631708	0.637021	0.229446	21.643722	False	False	False	False	False
ADK1	3.373517	3.373517	3.373517	14.702870	0.229446	21.643722	False	False	False	False	False
PPS	3.373517	3.373517	3.373517	14.702870	0.229446	21.643722	False	False	False	False	False
ATPM	8.390000	8.390000	0.000000	20.314621	0.229446	21.643722	False	False	False	False	False
PDH	7.017642	7.017642	7.017642	30.585141	0.229446	21.643722	False	False	False	False	False
ACt2r	-14.869041	-14.869041	-1.481148	33.133223	0.229446	21.643722	False	False	False	False	False
ACKr	-14.869041	-14.869041	-1.481148	33.133223	0.229446	21.643722	False	False	False	False	False
PTAr	14.869041	14.869041	-1.481148	33.133223	0.229446	21.643722	False	False	False	False	False
PFK	19.771770	19.771770	0.285288	48.425794	0.229446	21.643722	False	False	False	False	False

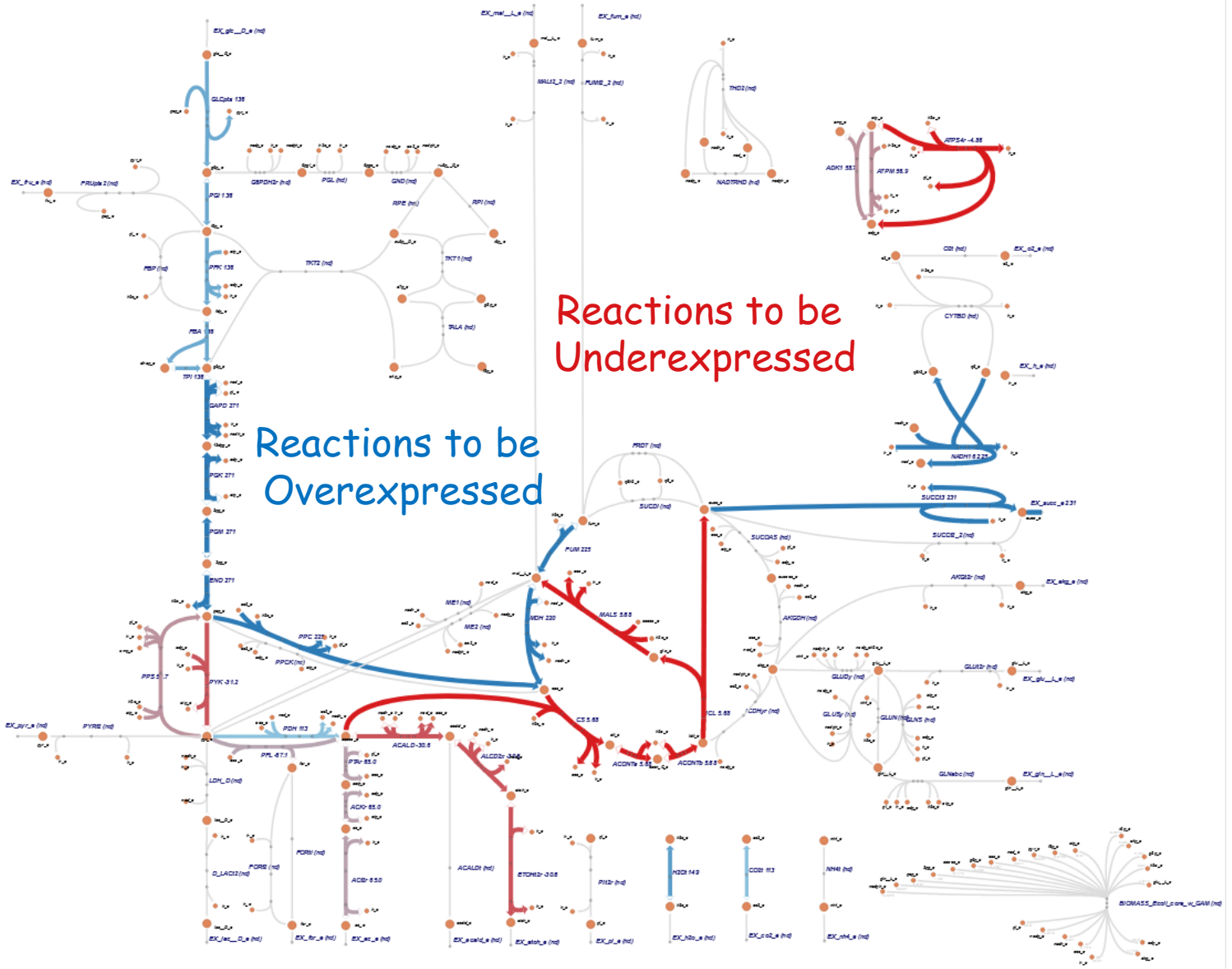
Overexpressed

dFVA.ipynb

Data Generated by Differential Flux Variability Analysis

- **reaction** - strain reactions that have a normalized_gaps > 0
- **lower_bound** - lower bound of reaction flux span
- **upper_bound** - upper bound of reaction flux span
- **gaps** - the distance between the reaction flux span of a given strain to the flux span of the reference strain
- **normalized_gaps** - calculated by dividing the strains gaps by the biomass function of the specific strain
- **biomass** - growth or biomass function
- **production** - the bioproduct production for a given strain
- **KO** - potential reaction knockout target (True, False)
- **flux_reversal** - does the flux direction change compared to the reference strain (True, False)
- **suddenly_essential** - has the strain changed the status of a reaction to essential (True, False)
- **free_flux** - reactions that has no flux limits ([-1000,1000] - reversible, [-1000,0] or [0,1000] - irreversible
- **reactions that are excluded from the analysis** (e.g. DM_succ_e, exchange reactions)

Escher Map of dFVA Results



Differential Flux Variability Analysis: Predicting Expression Modulation Targets

Cameo provides algorithms to search for genes or reactions that can be over- or down-regulated in order to achieve a given biological objective.

In [1]:
```
from cameo import models
from cameo.visualization.plotting.with_plotly import PlotlyPlotter
plotter = PlotlyPlotter()
import pandas as pd
pd.set_option('display.max_colwidth',None)
pd.set_option('display.max_rows', 1000)
```

Load the *E. coli* core model and create an anaerobic environment.

In [2]:
```
model = models.bigg.e_coli_core
model.reactions.EX_o2_e.lower_bound = 0
model.reactions.EX_glc__D_e.lower_bound = -20
```

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

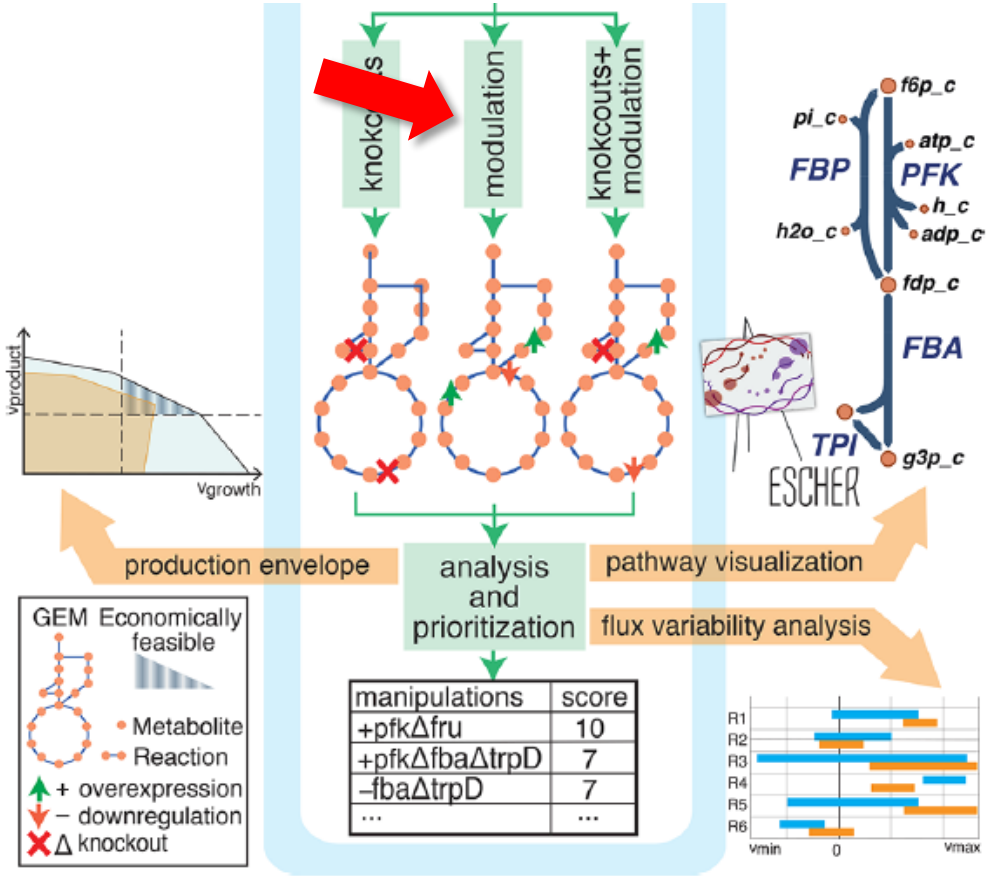
Differential flux variability analysis compares flux ranges of a reference model to a set of models that have been parametrized to lie on a grid of evenly spaced points in the *n*-dimensional production envelope (*n* being the number of reaction bounds to be varied). Below are the python packages that need to be loaded.

In [3]:
```
from cameo.flux_analysis.analysis import phenotypic_phase_plane
from cameo.flux_analysis.analysis import flux_variability_analysis
from cameo.strain_design.deterministic import DifferentialFVA
import plotly.express as px
```

dFVA.ipynb

Gene/Reaction Expression Modulation

- Introduction
- Flux Scanning based on Enforced Objective Flux (FSEOF)
- Differential flux variability analysis (dFVA)



Review Questions

What is gene modulation?

What is the purpose of gene modulation?

What is gene amplification?

What is the purpose of Flux Scanning based on Enforced Objective Flux (FSEOF)?

What is the "enforced objective" in FSEOF?

What is the purpose of Differential Flux Variability Analysis (dFVA)?

What is the flux span difference in dFVA?

What is the purpose of normalized gaps?

What is the relationship between a production envelope of a specific bioproduct and the surface sample space?

What is the relationship between a production envelope of a specific bioproduct and the grid sample space?

References

1. Raajaraam-2022-A Computational Framework to Identify Metabolic Engineering Strategies for the Co-Production of Metabolites (GitHub - RamanLab/co-FSEOF)
2. H. S. Choi, S. Y. Lee, T. Y. Kim, and H. M. Woo, '*In silico* identification of gene amplification targets for improvement of lycopene production.,' Appl Environ Microbiol, vol. 76, no. 10, pp. 3097–3105, May 2010.
3. https://nbviewer.org/github/biosustain/cameo-notebooks/blob/master/06-predict-gene-modulation-targets.ipynb