Introduction to Metabolic Engineering and Flux Balance Analysis

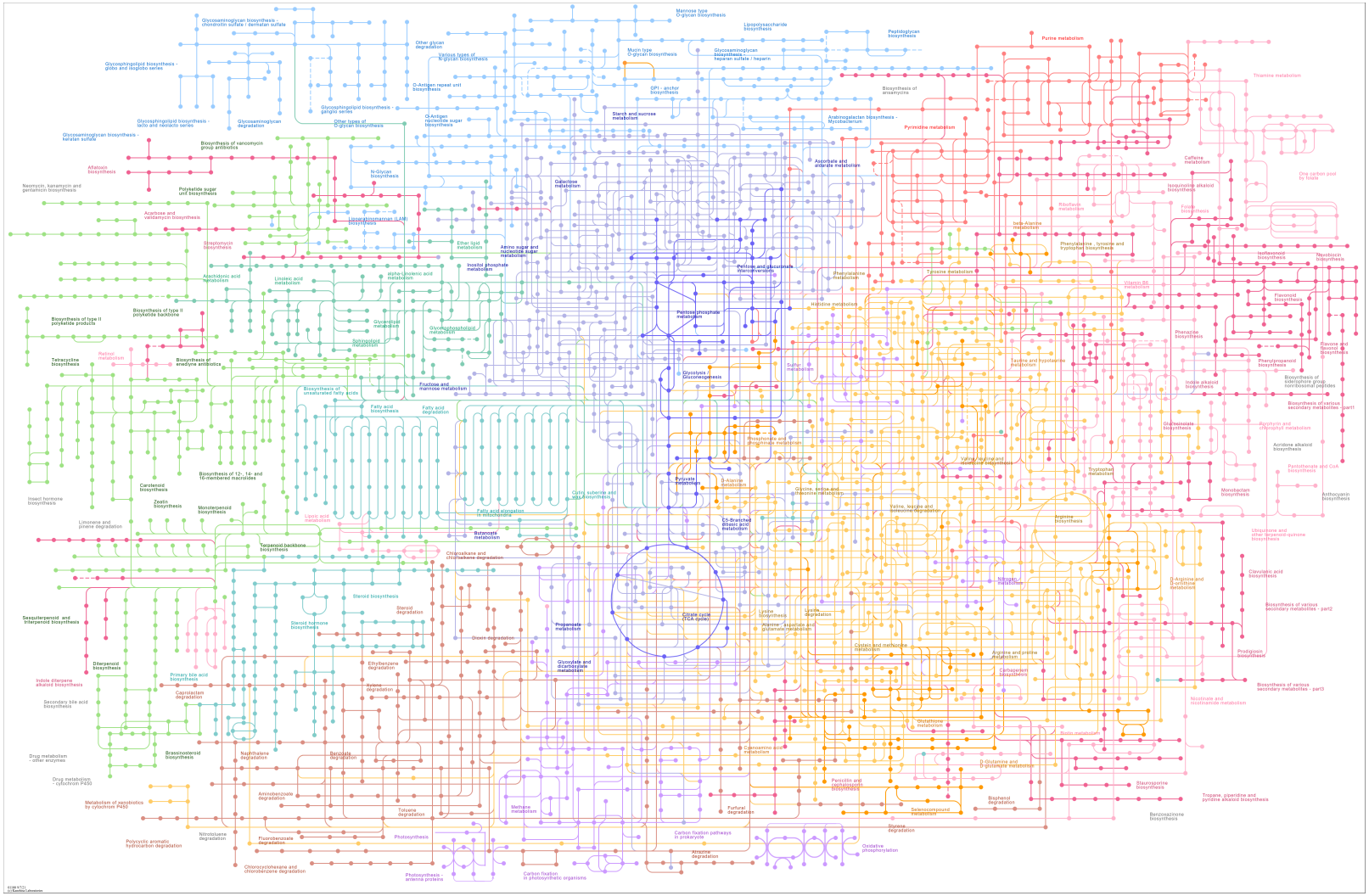
Metabolic engineering involves optimizing genetic and regulatory processes within cells or cell-free networks to enhance the production of a desired molecule or protein. This is achieved by manipulating the biochemical networks that convert raw materials into product molecules. The primary aim of metabolic engineering is to:

1. Mathematically model biochemical networks, calculate the yield (product divided substrate) of useful products and identify parts of the network that constrain the production of the products of interest.
2. Use genetic engineering techniques to modify the biochemical network to relieve constraints limiting production. The modified network can then be modeled to calculate the new product yield and to identify new constraints (back to 1).

Resources for biochemical network information:

- Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 2000 Jan 1;28(1):27-30. doi: 10.1093/nar/28.1.27. PMID: 10592173; PMCID: PMC102409.
- Karp, Peter D et al. "The BioCyc collection of microbial genomes and metabolic pathways." Briefings in bioinformatics vol. 20,4 (2019): 1085-1093. doi:10.1093/bib/bbx085
- Gama-Castro, Socorro, et al. "RegulonDB version 9.0: high-level integration of gene regulation, coexpression, motif clustering and beyond." Nucleic acids research vol. 44, D1 (2016): D133-43. doi:10.1093/nar/gkv1156

Fig 1. The overall metabolic map from the KEGG database. Each dot (*node*) is a metabolite, each line (*edge*) is a metabolic reaction.



What is Flux Balance Analysis (FBA)?

Flux balance analysis (FBA) is a mathematical modeling and analysis approach which computes the flow (or *flux*) of carbon and energy throughout a metabolic network. FBA, a member of the constraint based family of mathematical modeling tools, is a widely used approach to compute metabolic flux. However, there are alternatives to FBA, such as metabolic flux analysis (MFA), but these alternatives vary more in the solution

approach than the structure of the estimation problem.

Let's look at the following reference to understand better the different components of a flux balance analysis problem:

- Orth, J., Thiele, I. & Palsson, B. What is flux balance analysis?. Nat Biotechnol 28, 245–248 (2010). https://doi.org/10.1038/nbt.1614

Computational tools for working with flux balance analysis models:

- Heirendt, Laurent et al. “Creation and analysis of biochemical constraint-based models using the COBRA Toolbox v.3.0.” Nature Protocols vol. 14,3 (2019): 639-702. doi:10.1038/s41596-018-0098-2

Flux balance analysis problem structure

The FBA problem is typically encoded as a linear programming (LP) problem of the form:

$$\max_{v} \sum_{i \in \mathcal{R}} c_i v_i$$

subject to the constraints:

$$\begin{aligned} \sum_{j \in \mathcal{R}} \sigma_{ij} v_j = 0 \qquad & \forall i \in \mathcal{M} \\ \mathcal{L}_i \leq v_i \ \leq \mathcal{U}_i \qquad & \forall i \in \mathcal{R} \\ \ldots = \ldots & \end{aligned}$$

where $\mathbf{S}$ denotes the stoichiometric matrix, c_i denote the objective coefficients, $\mathbf{v}$ denotes the metabolic flux (the unknown that we are trying to estimate), and $\mathcal{L}$ (or $\mathcal{U}$) denote the permissible lower (or upper) bounds on the *unknown* metabolic flux. The first set of constraints enforces the conservation of mass, while the second imparts thermodynamic and kinetic information into the calculation. Finally, there are potentially other types of constraints (both linear and nonlinear) that can be used in this type of problem (we will not cover these here, but

these additional constraints may be important in specific applications).

What is a stoichiometric matrix?

The basis for flux balance calculations is the stoichiometric matrix. The stochiometric matrix is a digital representation of the biochemistry that can occur insides cells (such as that shown in Fig. 1). The stochiometric matrix, denoted by $\mathbf{S}$, is a $\mathcal{M} \times \mathcal{R}$ array of stoichiometric coefficients, denoted by the symbol σ_{ij} (i is the row index, j is the col index). Each of the $\mathcal{M}$ rows of $\mathbf{S}$ describes a particular metabolite (*node* in the metabolic network), while each of the $\mathcal{R}$ columns corresponds to a metabolic reaction (*edge* in the metabolic network). Thus:

- A stoichiometric coefficient $\sigma_{ij} > 0$ implies that metabolite i is **produced** by reaction j
- A stoichiometric coefficient $\sigma_{ij} = 0$ implies that metabolite i is **not connected** to reaction j
- A stoichiometric coefficient $\sigma_{ij} < 0$ implies that metabolite i is **consumed** by reaction j

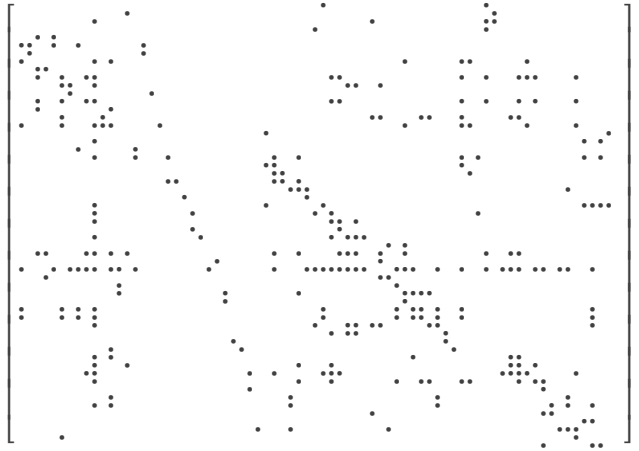
Example: Core stoichiometric matrix from *Escherichia coli*: Orth, Jeffrey D et al. “Reconstruction and Use of Microbial Metabolic Networks: the Core Escherichia coli Metabolic Model as an Educational Guide.” EcoSal Plus vol. 4,1 (2010): 10.1128/ecosalplus.10.2.1. doi:10.1128/ecosalplus.10.2.1

```
begin

    # what is the model name -
    model_file_name = "modelReg.mat"
    model_name = "modelReg"

    # where is model file?
    _PATH_TO_MODEL_FILE = joinpath(_PATH_TO_DATA, model_file_name)

    # load the mat file -> get the cobra_dictionary
    file = matopen(_PATH_TO_MODEL_FILE)
    cobra_dictionary = read(file, model_name)
    close(file)
end
```

```
KeySet(["c", "mets", "subSystems", "b", "metFormulas", "rxnGeneMat", "ub", "regulatoryGenes", "regu
```

```
keys(cobra_dictionary)
```

```
stm_sparse = 72×95 SparseArrays.SparseMatrixCSC{Float64, Int64} with 360 stored entries:
```



```
# get the stoichiometric matrix
stm_sparse = cobra_dictionary["S"]
```

```
stm_full = 72×95 Matrix{Float64}:
 0.0  0.0   0.0  0.0  0.0   0.0  0.0  0.0  …   0.0   0.0   0.0  0.0   0.0   0.0  0.0
 0.0  0.0   0.0  0.0  0.0   0.0  0.0  0.0       0.0   0.0   0.0  0.0   0.0   0.0  0.0
 0.0  0.0   0.0  0.0  0.0   0.0  0.0  0.0       0.0   0.0   0.0  0.0   0.0   0.0  0.0
 0.0  0.0   0.0  0.0  0.0   0.0  0.0  0.0       0.0   0.0   0.0  0.0   0.0   0.0  0.0
 0.0  0.0   0.0  0.0  0.0   0.0  0.0  0.0       0.0   0.0   0.0  0.0   0.0   0.0  0.0
 0.0  0.0  -1.0  0.0  0.0   1.0  0.0  0.0  …   0.0   0.0   0.0  0.0   0.0   0.0  0.0
 0.0  0.0   0.0  0.0  0.0  -1.0  0.0  0.0       0.0   0.0   0.0  0.0   0.0   0.0  0.0
 ⋮                         ⋮                ⋱              ⋮
 0.0  0.0   0.0  0.0  0.0   0.0  0.0  0.0       0.0   0.0   0.0  0.0   0.0   0.0  0.0
 0.0  0.0   0.0  0.0  0.0   0.0  0.0  0.0       0.0   0.0  -1.0  0.0   1.0   0.0  0.0
 0.0  0.0   0.0  0.0  0.0   0.0  0.0  0.0      -1.0  -1.0   0.0  0.0   0.0   0.0  0.0
 0.0  0.0   0.0  0.0  0.0   0.0  0.0  0.0       0.0   0.0   0.0  0.0   0.0   0.0  0.0
 0.0  0.0   0.0  0.0  0.0   0.0  0.0  1.0  …   0.0   1.0   0.0  0.0   0.0   0.0  0.0
 0.0  0.0   0.0  0.0  0.0   0.0  0.0  0.0       0.0   0.0   0.0  0.0  -1.0  -1.0  0.0
```

```
stm_full = Matrix(stm_sparse)
```

```
mets = 72×1 Matrix{Any}:
        "C3H4O10P2"
        "C3H4O7P"
        "C3H4O7P"
        "C6H10O10P"
        "C6H9O9P"
        "C2H3O2"
        "C2H3O2"
        ⋮
        "C5H9O8P"
        "C7H13O10P"
        "C4H4O4"
        "C4H4O4"
        "C25H35N7O19P3S"
        "C5H9O8P"
```

```
1 mets = cobra_dictionary["metFormulas"]
```

```
rxns = 95×1 Matrix{Any}:
        "ACALD"
        "ACALDt"
        "ACKr"
        "ACONTa"
        "ACONTb"
        "ACt2r"
        "ADK1"
        ⋮
        "SUCOAS"
        "TALA"
        "THD2"
        "TKT1"
        "TKT2"
        "TPI"
```

```
1 rxns = cobra_dictionary["rxns"]
```

Additional resources for the stoichiometric matrix

- Lecture 3: The genome reconstruction process. Systems Biology: Constraint-based Reconstruction and Analysis, Cambridge University Press, 2015
- Lecture 9: Properties of the Stoichiometric Matrix. Systems Biology: Constraint-based Reconstruction and Analysis, Cambridge University Press, 2015

What are flux bounds constraints?

Flux bounds constraints control the permissible ranges of a metabolic reaction rate (flux). Flux bounds constraints can incorporate thermodynamic information (whether a reaction is reversible) and kinetic information.

Resources for thermodynamics in the constraint-based world:

- Peres, Sabine, and Vincent Fromion. "Thermodynamic Approaches in Flux Analysis." Methods in molecular biology (Clifton, N.J.) vol. 2088 (2020): 359-367. doi:10.1007/978-1-0716-0159-4_17
- Flamholz, Avi, et al. "eQuilibrator–the biochemical thermodynamics calculator." Nucleic acids research vol. 40, Database issue (2012): D770-5. doi:10.1093/nar/gkr874

Let's do some sample calculations using eQuilibrator

Thermodynamic information tells you the directionality of the reaction (is the reaction reversible), but to get the permissible magnitudes of the reaction, we need to compute the kinetics. We (and others) formulate flux bounds as the product of successive corrections to the maximum possible rate, e.g., for an irreversible reaction:

$$0 \leq v_i \leq V^{\circ}_{max,j} \left(\frac{\epsilon}{\epsilon^{\circ}}\right) \theta_j(\ldots) f_j(\ldots)$$

where $V^{\circ}_{max,j}$ denotes the maximum reaction velocity computed for enzyme j at some characteristic enzyme concentration (and full activity), the ratio $\epsilon/\epsilon^{\circ}$ is a correction for enzyme concentration, $\theta_j(\ldots) \in [0,1]$ is an enzyme activity function (or measurement) and $f_j(\ldots)$ is a function describing the substrate dependence of the reaction rate j. Both $\theta_j(\ldots)$ and $f_j(\ldots)$ could have associated parameters, e.g., saturation or binding constants, etc.

What is the objective of an *E. coli* cell?

Short answer: maximize the growth rate

Longer answer: This question raged in the metabolic engineering community twenty years ago but is mostly forgotten today. First, there are long-standing arguments from ecology (or even from the microeconomics perspective) as to why *E.coli* would be goal-oriented. However, Palsson and coworkers largely put this question to rest (at least for *E.coli*) with the publication:

- Ibarra, Rafael U et al. "Escherichia coli K-12 undergoes adaptive evolution to achieve in silico predicted optimal growth." Nature vol. 420,6912 (2002): 186-9. doi:10.1038/nature01149

What about cell-free networks? No cells, but most of the biochemistry. Is there an objective? This is an open question. However, initial work in this area suggests the maximization of translation, at least for single gene systems:

- Vilkhovoy, Michael, et al. "Sequence-Specific Modeling of E. coli Cell-Free Protein Synthesis." ACS Synthetic Biology vol. 7,8 (2018): 1844-1857. doi:10.1021/acssynbio.7b00465

Flux balance analysis case studies and examples

- Wayman, Joseph A et al. "Improving designer glycan production in Escherichia coli through model-guided metabolic engineering." Metabolic engineering communications vol. 9 e00088. 29 Mar. 2019, doi:10.1016/j.mec.2019.e00088
- Shimpi AA, Tan ML, Vilkhovoy M, Dai D, Roberts LM, Kuo JC, Huang L, Varner JD, Paszek M, Fischbach C. Convergent Approaches to Delineate the Metabolic Regulation of Tumor Invasion by Hyaluronic Acid Biosynthesis. Adv Healthc Mater. 2023 Jun;12(14):e2202224. doi: 10.1002/adhm.202202224. Epub 2022 Dec 21. PMID: 36479976; PMCID: PMC10238572.