Notes on metabolic modelling analysis

Haris Zafeiropoulos

Nov 21, 2024

CHAPTER

THREE

CONSTRAINT BASED ANALYSIS

In this notebook, we are going to discuss some basics of CBA from their mathematical, theoretical and biological perspective.

3.1 Overflow metabolism is caused by two growth-limiting constraints

Literature

- de Groot et al (2020) DOI
- Basan et. al (2015) DOI

Overflow metabolism in Escherichia coli results from efficient proteome allocation.

The authors assume that the yield of energy per carbon molecule is higher for respiration than for fermentation: $n_r > n_f$, but that fermentation is more proteome-efficient: $\epsilon_f > \epsilon_r$.

CHAPTER

FOUR

HOW TO

In this notebook, we will keep track of handy implementations for metabolic modeling related tasks.

4.1 Media and environment setup

4.1.1 Extracellular reactions

COMETS includes the capability to simulate reactions happening in the extracellular environment, without association to a specific organism. Users can implement either elementary reactions of arbitrary order based on mass-action kinetics, or enzyme-catalyzed reactions obeying Michaelis–Menten kinetics, e.g., for the simulation of extracellular enzymes.

4.1.2 Get complete medium for ModelSEED

With the term *complete medium** we describe an *in silico* object where any compound that could be used as a nutrient, it is available for the model.

To build this object for the case of ModelSEED, we need to first get all the possible compounds. And we can do this by first, getting locally the ModelSEEDDatabase repo.

Then we can explore the `Biochemistry` folder of that to retrieve all possible nutrients that could be imported in our model.

From the Biochemistry folder of the dev branch of the ModelSEEDDatabase repository, run:

```
!awk -F"\t" '$6 != 1 && $18==0  {print $5}' reaction_*.tsv   > TRANSPORT_REACTIONS.tsv
```

```
awk: fatal: cannot open file `reaction_*.tsv' for reading: No such file or␣
↪directory
```

Now, with something like the following Python chunk, you can build the complete medium and export in a `.csv` file that with the applied format, could be used for gapfilling with the `fill` command of the `gapseq` tool.

```
def write_to_gapseq_format(all_compounds, cpd2name, output_file):
    """
    Write a 3-col csv file with the compound id, its name and a boundary flux of 1000
    """
    with open(output_file, "w") as f:
        counter = 0
        for compound in all_compounds:
            if compound in cpd2name:
```

(continues on next page)

(continued from previous page)

```
                counter += 1
                f.write(f"{compound}\t{cpd2name[compound]}\t1000\n")
            else:
                print(f"Compound {compound} not found in cpd2name dictionary")

    print(f"Total compounds written: {counter}")


def process_transport_reactions(input_file, output_file=None):
    """
    Parse the TRANSPORT_REACTIONS.tsv file to export compounds that should be part of
 ↪the complete medium.
    """
    with open(input_file) as f:
        lines = f.readlines()

    ex = [line.strip() for line in lines if len(line.split(";")) == 2]

    cpd2name = {}
    all_compounds = set()

    for reaction in ex:
        compounds = reaction.split(";")
        c1 = compounds[0].split(":")[1]
        c2 = compounds[1].split(":")[1]

        if c1 == c2:
            name = compounds[0].split(":")[-1]
            all_compounds.add(c2)
            if c2 not in cpd2name:
                cpd2name[c2] = name

    if output_file is not None:
        write_to_gapseq_format(all_compounds, cpd2name, output_file)

# Main execution
if __name__ == "__main__":
    process_transport_reactions("TRANSPORT_REACTIONS.tsv", "complete_modelseed_medium.
 ↪csv")
```

```
Total compounds written: 0
```

CHAPTER

FIVE

SOFTWARE AND RESOURCES

In this page we will keep a list of the software approaches we are aware of for the various metabolic modeling tasks. Apparently, this list can never be complete, but it can be improved with your contribution! So, feel free to make a PR adding something or contact us to do that for you.

5.1 Reconstruction, gap-flling, validation

- DEMETER
-

5.2 Further constraints

5.2.1 Thermodynamics

Tool	Description	Architecture	Repo	DOI
pyTFA	implementations of the original thermodynamics-based Flux Analysis (TFA) paper. Specifically, they include explicit formulation of Gibbs energies and metabolite concentrations, which enables straightforward integration of metabolite concentration measurements.	standalone	GitHu	OA
OptMDFpathway**	integration of thermodynamic information in metabolic models to assess the feasibility of flux distributions by thermodynamic driving forces. Extends the framework of Max-min Driving Force (MDF) for thermodynamic pathway analysis. Identifies both the optimal MDF for a desired phenotypic behavior and the respective pathway that supports the optimal driving force.	standalone	GitHu	OA

- matTFA is the MATLAB version of pyTFA ** part of the CNApy library

5.2.2 Enzymes

Tool	Description	Architecture	Repo	DOI
AutoPACMEN	Retrieves kcat data and adds protein allocation constraints to stoichiometric metabolic models according to the sMOMENT method.	standalone	GitHu	OA
ETGEMs	Construction of enzymatic and thermodynamic constrained GEMs in a single Pyomo* modelling framework.	scripts	GitHu	OA

*Python-based, open-source optimization modeling language

5.2.3 Transporter annotations

Tool	Description	Architecture	Repo	DOI
SPOT	machine learning model that can successfully predict specific substrates for arbitrary transport protein	standalone	GitHu	OA
TranSyT	Identifies transport systems and the compounds carried across membranes, based on the annotations of the Transporter Classification Database (TCDB). Generates the respective transport reactions while providing the respective Gene-Protein-Reaction associations.	webservice	GitHu	OA

- They showed that the majority of TCDB entries are of low UniProt-based annotation scores.

CHAPTER

SIX

DATABASES

Tool	Description	Link	DOI
equilibrator	estimated thermodynamic constants	link	OA
SABIO-RK	curated information about biochemical reactions, their kinetic rate equations with parameters and experimental conditions.	link	OA

has also an API library `equilibrator-api` ReadTheDocs | GitLab

and a cache one `equilibrator-cache`: A database application for caching data used by eQuilibrator and related projects. Stored data includes compound names, structures, protonation state information, reaction and enzyme info, and cross-references to other databases. All compounds stored in equilibrator-cache are cross-referenced using InChIKey. GitLab

6.1 Topological approaches

- QFCA
- **FluxModeCalculator**

Efficient Elementary flux mode (EFM) calculation

paper | GitHub (part of the `MCS` repo)

6.2 Dynamic approaches

- **dfba** GitLab repo Documentation paper A recent approach for dynamic FBA that considers the solution non-uniqueness.

-COMETS

-BacArena

6.3 Community modelling

- mergem
- PyCoMo
-

6.4 Data integration

- FASTCROMICS
- TRFBA

tellurium:

https://tellurium.analogmachine.org

A Python Environment for Reproducible Dynamical Modeling of Biological Networks

6.5 Other resources and tools of interest

Tool	Description	Architecture	Repo	DOI
gRodon2	predicts maximal growth rates using genomic data	R-package	GitHub	OA

CHAPTER

SEVEN

BIOLOGICAL INSIGHT

Here I keep track of bio facts that can be of use in my *in silico* explorations.

Apparently, this notebook will make much less sense since it will include pieces of information that do not necessarily lead somewhere :sweat_smile:

Pseudomonas aeruginosa: May rely on glutamine synthetase and nitrate/nitrite reduction for glutamine production.

CHAPTER

EIGHT

BIBLIOGRAPHY

BIBLIOGRAPHY

[1] Bernhard Palsson. *Systems biology*. Cambridge university press, 2015.

[2] Gregory Stephanopoulos. Metabolic fluxes and metabolic engineering. *Metabolic engineering*, 1(1):1–11, 1999.

[3] Bernhard Ø Palsson. *Systems biology: properties of reconstructed networks*. Cambridge university press, 2006.

[4] Bernhard Ø Palsson. *Systems biology: simulation of dynamic network states*. Cambridge University Press, 2011.

[5] Neema Jamshidi and Bernhard Ø Palsson. Mass action stoichiometric simulation models: incorporating kinetics and regulation into stoichiometric models. *Biophysical journal*, 98(2):175–185, 2010.

PROOF INDEX