Evaluating and ranking a set of Pathways based on multiple metrics

Desi Arti

NPM 2206130706

PATHWAY ANALYSIS

The Pathway Analysis process allows the assessment of a set of diverse pathways generated through the RetroSynthesis workflow within a host organism (E. coli iML1515). These workflows are accessible through the Galaxy SynbioCAD platform. The objective is to guide users on the most theoretically effective pathways by evaluating them according to four criteria: flux of the target product, thermodynamic viability, pathway length, and enzyme accessibility.

PATHWAY ANALYSIS

Evaluating and ranking a set of pathways involves assessing their effectiveness and suitability using various metrics. This process is crucial for identifying the most promising pathways in a given context. Multiple metrics are employed to comprehensively analyze and compare these pathways.

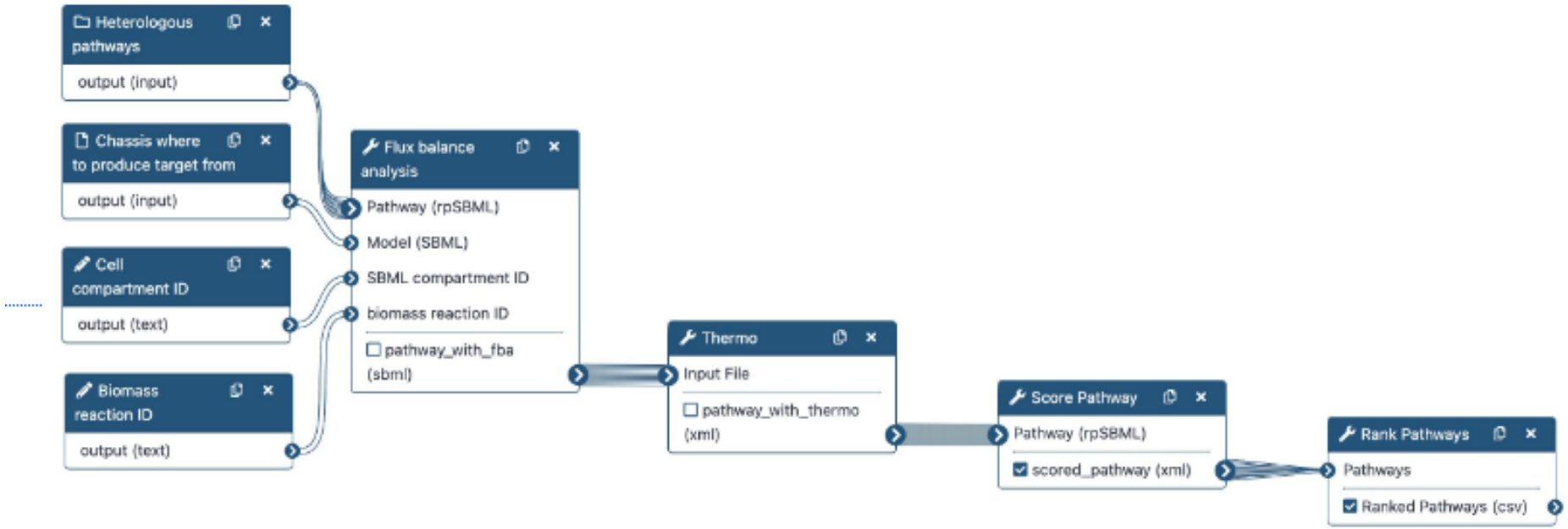
Common Metrics

1. **Production Flux:** The rate at which the desired product is produced within a pathway.
2. **Thermodynamics:** The assessment of the thermodynamic feasibility of each pathway, considering factors like energy requirements and stability.
3. **Pathway Length:** The number of steps or reactions involved in a pathway. Shorter pathways are often preferred for efficiency.
4. **Enzyme Availability:** The availability and compatibility of enzymes involved in each pathway.
5. **Reaction SMARTS:** Evaluating the specific chemical reactions within each pathway using SMARTS patterns.

To systematically rank these pathways, a global score is computed by combining these metrics. This can be achieved through the application of machine learning models, which weigh and integrate the various metrics based on their relative importance. The outcome is a ranked list of pathways, with the highest-scoring pathways considered the most promising or optimal for a particular objective.

This comprehensive evaluation and ranking process are often employed in the field of synthetic biology, metabolic engineering, and pathway optimization to guide researchers toward selecting pathways that are not only theoretically sound but also align with specific criteria or goals.

Step of Workflow



•Step 1 Get the Data

1. Create a new history for this tutorial named *Pathway Analysis*.
2. Import the files from Zenodo :

```
https://zenodo.org/record/6628296/files/rp_001_0001.xml
https://zenodo.org/record/6628296/files/rp_001_0006.xml
https://zenodo.org/record/6628296/files/rp_001_0011.xml
https://zenodo.org/record/6628296/files/rp_002_0001.xml
https://zenodo.org/record/6628296/files/rp_002_0006.xml
https://zenodo.org/record/6628296/files/rp_002_0011.xml
https://zenodo.org/record/6628296/files/rp_003_0001.xml
https://zenodo.org/record/6628296/files/rp_003_0116.xml
https://zenodo.org/record/6628296/files/rp_003_0231.xml
https://zenodo.org/record/6628296/files/SBML_Model_iML1515.xml
```

Copy

Tip: Importing via links

3. Create a list or collection named `Heterologous pathways` and composed of the 9 rpSBML pathways.

Tip: Creating a dataset collection

Step 2 Calculating the Flux of a Target using FBA

1. Run **Flux balance analysis** (Galaxy version 5.12.1) with the following parameters:
    - *"Pathway (rpSBML)"*: Select `Heterologous pathways` (Input dataset collection) from your current history.
    - *"Model (SBML)"*: Select `SBML_Model_iML1515.xml` (Input dataset) from your current history.
    - *"SBML compartment ID"*: Leave the default value `c`.

> 💬 Comment: Choose a compartment corresponding to your model
>
> You can specify the compartment from which the chemical species were extracted. The default is `c`, the BiGG code for the cytoplasm.

    - *"biomass reaction ID"*: Specify the biomass reaction ID that will be used for the "fraction" simulation, type `R_BIOMASS_Ec_iML1515_core_75p37M`.

Step 3 Compute thermodynamics values

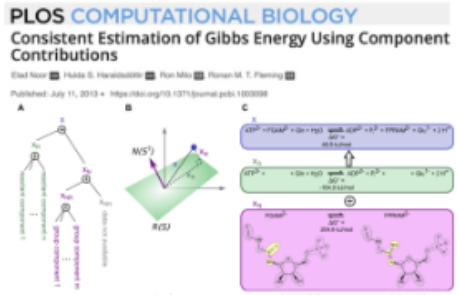
rpFBA

rpThermo

Optimisation of target production

E_{form}

$\Delta G = \Delta H - T\Delta S$



PLOS COMPUTATIONAL BIOLOGY

Consistent Estimation of Gibbs Energy Using Component Contributions

```
> coeffs = optimise(reactions, target)

Reactions
  |- rxn_1: 1.0 MNXM188 + 1.0 MNXM4 + 1.0 MNXM6 + 3.0 MNXM1 --> 1.0 CMPD_0000000004 + 1.0 CMPD_0000000003 + 1.0 MNXM13 + 1.0 MNXM15 + 1.0 MNXM5
  |- rxn_2: 1.0 MNXM4 + 2.0 CMPD_0000000003 --> 2.0 MNXM1 + 1.0 TARGET_0000000001
  |- rxn_3: 1.0 MNXM4 + 1.0 MNXM6 + 3.0 CMPD_0000000004 --> 1.0 MNXM13 + 1.0 MNXM5

Intermediate compounds: CMPD_0000000003, CMPD_0000000004
Reactions: rxn_1, rxn_2*, rxn_3 (* to be optimised)
[3., 1.5, 1.]


> pseudo_reaction(reactions, coeffs)

Reactions
  |- rxn_1: 3.0 MNXM188 + 3.0 MNXM4 + 3.0 MNXM6 + 9.0 MNXM1 --> 3.0 CMP
  |- rxn_2: 1.5 MNXM4 + 3.0 CMPD_0000000003 --> 3.0 MNXM1 + 1.5 TARGET_
  |- rxn_3: 1.0 MNXM4 + 1.0 MNXM6 + 3.0 CMPD_0000000004 --> 1.0 MNXM13 +

Pseudo-Reaction
  |- 7.5 MNXM1 + 3.0 MNXM188 + 5.5 MNXM4 + 4.0 MNXM6 --> 4
```

1. **Thermo** (Galaxy version 5.12.1) with the following parameters:
   - *"Input File"*: `pathway_with_fba` (output of **Flux balance analysis**)

Step 4 Compute the Global Score

1. **Score Pathway** (Galaxy version 5.12.1) with the following parameters:
   - *"Pathway (rpSBML)"*: `pathway_with_thermo` (output of **Thermo**)

> Comment
>
> This tool will output a new annotated SBML file representing the pathway, containing the `global_score` annotation.

Step 5 Rank Annotate Pathway

1. **Score Pathway** (Galaxy version 5.12.1) with the following parameters:
    - *"Pathway (rpSBML)"*: `pathway_with_thermo` (output of **Thermo**)

> Comment
>
> This tool will output a new annotated SBML file representing the pathway, containing the `global_score` annotation.

Step 6 Run the **Pathway Analysis Workflow**

Click on *Workflow* on the top menu bar of Galaxy. You will see **Pathway Analysis Workflow**

Click on the ▶ (*Run workflow*) button next to your workflow

Provide the workflow with the following parameters:

- *"Heterologous pathways"*: Select `Heterologous pathways` (Input dataset collection) from your current history.
- *"Chassis where to produce target from"*: Select `SBML_Model_iML1515.xml` (Input dataset) from your current history.
- *"Cell compartment ID"*: Enter value `c`.
- *"biomass reaction ID"*: Specify the biomass reaction ID that will be restricted in the "fraction" simulation type `R_BIOMASS_Ec_iML1515_core_75p37M`.

Conclusion

- In the process of choosing optimal pathways for lycopene production in E. coli, certain metrics need assessment, including the production flux of the target and the thermodynamics of the pathway. A comprehensive score is then calculated by integrating these factors with others (such as pathway length, enzyme availability score, and reaction SMARTS) through a machine learning model. These stages are accomplished through the tools provided in the described Pathway Analysis workflow.

Refences

1.Flamholz, A., E. Noor, A. Bar-Even, and R. Milo, 2011 **eQuilibrator–the biochemical thermodynamics calculator**. Nucleic Acids Research 40: D770–D775. 10.1093/nar/gkr874

2.Ebrahim, A., J. A. Lerman, B. O. Palsson, and D. R. Hyduke, 2013 **COBRApy: COnstraints-Based Reconstruction and Analysis for Python**. BMC Systems Biology 7: 10.1186/1752-0509-7-74

3.Hérisson, J., T. Duigou, M. du Lac, K. Bazi-Kabbaj, M. S. Azad *et al.*, 2022 **Galaxy-SynBioCAD: Automated Pipeline for Synthetic Biology Design and Engineering**. 10.1101/2022.02.23.481618