

Systems biology

CobraMod: a pathway-centric curation tool for constraint-based metabolic models

Stefano Camborda , Jan-Niklas Weder and Nadine Töpfer*

Independent Research Group, Molecular Genetics Department, Leibniz Institute of Plant Genetics and Crop Plant Research, Gatersleben, Germany

*To whom correspondence should be addressed.

Associate Editor: Pier Luigi Martelli

Received on August 2, 2021; revised on December 20, 2021; editorial decision on February 19, 2022; accepted on February 22, 2022

Abstract

Summary: COntstraint-Based Reconstruction and Analysis of genome-scale metabolic models has become a widely used tool to understand metabolic network behavior at a large scale. However, existing reconstruction tools lack functionalities to address modellers' common objective to study metabolic networks on the pathway level. Thus, we developed CobraMod—a Python package for pathway-centric modification and extension of genome-scale metabolic networks. CobraMod can integrate data from various metabolic pathway databases as well as user-curated information. Our tool tests newly added metabolites, reactions and pathways against multiple curation criteria, suggests manual curation steps and provides the user with records of changes to ensure high quality metabolic reconstructions. CobraMod uses the visualization tool Escher for pathway representation and offers simple customization options for comparison of pathways and flux distributions. Our package enables coherent and reproducible workflows as it can be seamlessly integrated with COBRApy and Escher.

Availability and implementation: The source code can be found at <https://github.com/Toepfer-Lab/cobramod/> and can be installed with pip. The documentation including tutorials is available at <https://cobramod.readthedocs.io/>.

Contact: toepfer@ipk-gatersleben.de

1 Introduction

Genome-scale metabolic models (GEMs) and their analysis by constraint-based modeling techniques are widely used tools to study metabolic systems at a large scale. Several software tools for Constraint-Based Reconstruction and Analysis (COBRA) are available, such as the COBRA toolbox, ModelSEED, Pathway Tools, RAVEN, CarveMe and Merlin (Dias *et al.*, 2015; Heirendt *et al.*, 2019; Karp, 2002; Machado *et al.*, 2018; Seaver *et al.*, 2021; Wang *et al.*, 2018) and have been evaluated here (Faria *et al.*, 2018; Mendoza *et al.*, 2019). In recent years, the freely available and community-supported software package COBRApy has gained particular popularity (Ebrahim *et al.*, 2013). COBRApy performs commonly used COBRA methods such as flux balance analysis, flux variability analysis, gene deletion analysis and includes simple, object-oriented interfaces for model reconstruction.

Several software packages complement COBRApy by implementing extended functionalities. For instance, Cameo and MewPy offer functionalities for computational strain optimization, MEMOTE includes a suite of tests to assess GEM quality, Medusa facilitates generating and analyzing ensembles of GEMs, and the Escher visualization tool offers an user-friendly interface for designing and manipulating pathway maps (Lieven, 2020; Cardoso *et al.*, 2018; King *et al.*, 2015; Medlock *et al.*, 2020; Pereira *et al.*, 2021).

However, currently available reconstruction tools rely either on error-prone, automated reconstruction procedures or laborious, manual addition of individual reactions or reaction sets and thus preclude the extension and curation of GEMs based on their biologically meaningful subsets, i.e. the metabolic pathways.

Here, we present CobraMod, a pathway-centric curation tool for the modification and extension of GEMs. CobraMod offers a comprehensible set of functions for semi-automated network extension, testing and visualization and enables easy, user-friendly manual curation and information logging to ensure high quality network reconstructions. CobraMod is written in Python 3; it builds upon and extends COBRApy and can directly interact with Escher for pathway and flux visualization.

2 Implementation

CobraMod is an open-source package which enables modifying and extending GEMs with metabolic pathway information from various databases or user-curated datasets. Our package converts these data into native COBRApy objects and quality-checks them for multiple curation criteria before incorporating them into the model (Fig. 1A). CobraMod's main functions include downloading metabolic pathway information (`get_data`), creating COBRApy objects

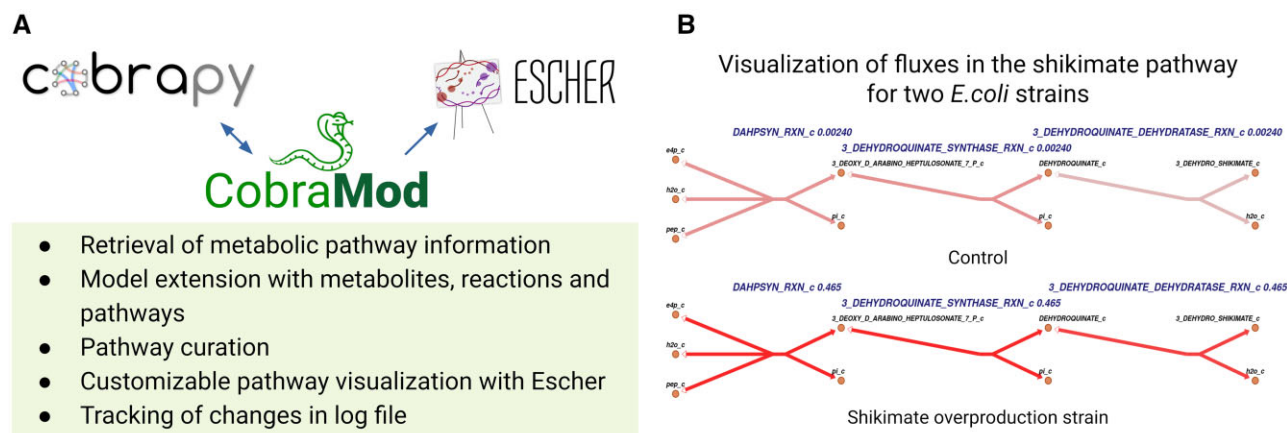


Fig. 1. CobraMod's main functionalities and pathway visualization example. (A) CobraMod's pathway-centric functionalities bridge COBRAPy methods and the visualization tool Escher. (B) Visualization of a metabolic engineering case study of the shikimate pathway in *E. coli*. Flux solutions for two strains of *E. coli* (control and engineered) are visualized. For simplicity, we represented only three reactions of the whole pathway. Reaction names and pathway fluxes are given in blue. For comparability, flux values were normalized and darker colors indicate higher flux values

(create_objects) and including new metabolites (add_metabolites), reactions (add_reactions) or pathways (add_pathway), as well as testing a reaction's capability to carry a non-zero flux (test_non_zero_flux) and pathway visualization (visualize).

2.1 Data retrieval

CobraMod supports all databases from the BioCyc collection (Karp *et al.*, 2019), the KEGG database (Kanehisa and Goto, 2000) and the BiGG Models repository (King *et al.*, 2016). The user can retrieve metabolic pathway information by specifying a database and the corresponding identifiers for metabolites, reactions or pathways. CobraMod automatically gathers gene information when obtaining information for reactions or pathways. CobraMod then downloads these datasets, stores them locally to ensure reproducibility (get_data), and transforms them into COBRAPy objects (create_object). In addition, CobraMod can integrate user-curated metabolites and reactions via text file or direct script input (add_metabolites, add_reactions).

2.2 Curation steps

CobraMod enables modifying and analyzing GEMs on the metabolic pathway level. Thus, it combines sets of reactions into pathway-objects, which the user can directly add to the model (add_pathways). Reactions and metabolites of a given pathway-object will undergo a curation process in which they are tested for duplicate elements, missing chemical formulas of the metabolites, mass balance of reactions and reaction reversibility (detailed in the documentation). To ensure that the added pathways are functional we implemented a non-zero flux test (test_non_zero_flux). During the test, CobraMod can add auxiliary source reactions and suggests manual curation steps based on these auxiliary modifications. Moreover, CobraMod offers cross-referencing and meta-data curation and is MEMOTE-compliant. Our tool offers comprehensible and user-friendly tracking of the curation process. When a pathway-object is added to the model a summary is outputted and the complete curation procedure is written to a log file. If any of the curation criteria is not met or exceptions are encountered, CobraMod passes a warning through the Python console and the log file.

2.3 Visualization

CobraMod uses Escher for pathway visualization. To this end, each pathway-object includes a visualization method (visualize) which automatically generates pathway maps of the respective set of reactions. These pathway maps can be easily customized to visualize flux distributions using default or user-defined colors and gradients (linear or quantile normalized).

3 Test case

To demonstrate CobraMod's functionalities we implemented two test cases based on *in vivo* and *in silico* overproduction studies in *Escherichia coli*. In the first example, we used a core model of *E. coli* (Orth *et al.*, 2010) to reproduce engineering strategies for improved shikimate synthesis (Chen *et al.*, 2014). Using our Escher interface, we visualized shikimate production for the control and one of the engineered strains (Fig. 1B). In a second example, we utilize a genome-scale model of *E. coli* (Monk *et al.*, 2017) to reproduce *in silico* experiments that introduce a synthetic homoserine cycle as an efficient route for methylotrophic growth (He *et al.*, 2020) and demonstrate the strength of CobraMod's pathway-centric curation procedures. The test cases with a step-by-step workflow can be found in the documentation.

4 Conclusion

CobraMod offers user-friendly, pathway-centric extension, curation and flux visualization for large-scale metabolic networks. It thus addresses a common modeller's objective to study metabolic network behavior on the pathway level. CobraMod employs as much automation as possible and suggests necessary manual curation steps to ensure high quality metabolic reconstructions. Our tool can be directly linked with COBRAPy and the Escher visualization tool and thus enables coherent and reproducible workflows.

Acknowledgement

The authors thank Mithil Gaikwad for feedback on this package.

Financial Support: none declared.

Conflict of Interest: none declared.

References

- Cardoso, J.G.R. *et al.* (2018) Cameo: a Python library for computer aided metabolic engineering and optimization of cell factories. *ACS Synthetic Biol.*, 7, 1163–1166.
- Chen, X. *et al.* (2014) Metabolic engineering of *Escherichia coli* for improving shikimate synthesis from glucose. *Bioresource Technol.*, 166, 64–71.
- Dias, O. *et al.* (2015) Reconstructing genome-scale metabolic models with merlin. *Nucleic Acids Res.*, 43, 3899–3910.
- Ebrahim, A. *et al.* (2013) Cobrapy: constraints-based reconstruction and analysis for python. *BMC Syst. Biol.*, 7, 74.
- Faria, J.P. *et al.* (2018) Methods for automated genome-scale metabolic model reconstruction. *Biochem. Soc. Trans.*, 46, 931–936.

- He,H. *et al.* (2020) An optimized methanol assimilation pathway relying on promiscuous formaldehyde-condensing aldolases in *E. coli*. *Metab. Eng.*, **60**, 1–13.
- Heirendt,L. *et al.* (2019) Creation and analysis of biochemical constraint-based models using the COBRA Toolbox v.3.0. *Nat. Protoc.*, **14**, 639–702.
- Kanehisa,M. and Goto,S. (2000) Kegg: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.*, **28**, 27–30.
- Karp,P.D. *et al.* The Pathway Tools software. *Bioinformatics*, (2002) **18**, Suppl 1S, 225–232.
- Karp,P.D. *et al.* (2019) The BioCyc collection of microbial genomes and metabolic pathways. *Brief. Bioinf.*, **20**, 1085–1093.
- King,Z.A. *et al.* (2015) Escher: a web application for building, sharing, and embedding data-rich visualizations of biological pathways. *PLoS Comput. Biol.*, **11**, e1004321.
- King,Z.A. *et al.* (2016) BiGG Models: a platform for integrating, standardizing and sharing genome-scale models. *Nucleic Acids Res.*, **44**, D515–D522.
- Lieven,C. *et al.* (2020) MEMOTE for standardized genome-scale metabolic model testing. *Nature Biotechnology*, **38**, 272–276. 10.1038/s41587-020-0446-y
- Machado,D. *et al.* (2018) Fast automated reconstruction of genome-scale metabolic models for microbial species and communities. *Nucleic Acids Res.*, **46**, 7542–7553.
- Medlock,G.L. *et al.* (2020) Medusa: software to build and analyze ensembles of genome-scale metabolic network reconstructions. *PLoS Comput. Biol.*, **16**, e1007847.
- Mendoza,S.N. *et al.* (2019) A systematic assessment of current genome-scale metabolic reconstruction tools. *Genome Biol.*, **20**, 158.
- Monk,J.M. *et al.* (2017) IML1515, a knowledgebase that computes *Escherichia coli* traits. *Nat. Biotechnol.*, **35**, 904–908.
- Orth,J.D. *et al.* (2010) Reconstruction and use of microbial metabolic networks: the core *Escherichia coli* metabolic model as an educational guide. *EcoSal Plus*, **4**, <https://doi.org/10.1128/ecosalplus.10.2.1>.
- Pereira,V. *et al.* (2021) MEWpy: a computational strain optimization workbench in Python. *Bioinformatics*, **37**, 2494–2496.
- Seaver,S.M.D. *et al.* (2021) The ModelSEED Biochemistry Database for the integration of metabolic annotations and the reconstruction, comparison and analysis of metabolic models for plants, fungi and microbes. *Nucleic Acids Res.*, **49**, D575–D588.
- Wang,H. *et al.* (2018) RAVEN 2.0: A versatile toolbox for metabolic network reconstruction and a case study on *Streptomyces coelicolor*. *PLoS Comput. Biol.*, **14**, e1006541.