Systems biology

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

Stefano Camborda¹, Jan-Niklas Weder¹ and Nadine Töpfer¹*

¹Leibniz Institute of Plant Genetics and Crop Plant Research, Gatersleben, Germany.

*To whom correspondence should be addressed.

Associate Editor: XXXXXXX
Received on XXXXX; revised on XXXXX; accepted on XXXXX

Abstract

Summary: Constraint-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: 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 et al., 2020; 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 analysing ensembles of GEMs, and the Escher visualization tool offers an user-friendly interface for designing and manipulating pathway maps (Beber & Lieven, 2020; Pereira et al., 2021; Cardoso et al., 2018; King et al., 2015; Medlock et al., 2020). 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 comprehensive 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.

© The Author(s) 2022. Published by Oxford University Press.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.
2. Implementation

CobraMod is an open-source package which enables modifying and extending GEMs with metabolic pathway information from various databases or user-curated data sets. 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 (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 & 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 the pathways. CobraMod then downloads these data sets, stores them locally to ensure reproducibility (get_data), and transforms them into COBRApy objects (create_object). Additionally, 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 analysing 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 E. 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 methyloptrophic 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.

Acknowledgements

The authors thank Mithil Gaikwad for feedback on this package.
References

Beber, M., & Lieven, C. (2020). memote—The genome-scale metabolic model test suite (Version 0.11.1). Novo Nordisk Foundation Center for BioSustainability, Technical University of Denmark.

Cardoso, J. G. R., Jensen, K., Lieven, C., Larke Hansen, A. S., Galkina, S., Beber, M., ... & Sonnenschein, N. (2018). Cameo: A Python Library for Computer Aided Metabolic Engineering and Optimization of Cell Factories. ACS Synthetic Biology, 7(4), 1363–1366. doi: 10.1021/acsynt.8b00423

Chen, Z., Li, M., Zhou, L., Shen, W., Algasan, G., Fan, Y., & Wang, Z. (2014). Metabolic engineering of Escherichia coli for improving shikimate synthesis from glucose. Bioresource Technology, 166, 64–71. doi: 10.1016/j.biortech.2014.05.035

Dias, O., Rocha, M., Ferreira, E. C., & Rocha, I. (2015). Reconstructing genome-scale metabolic models with merlin. Nucleic Acids Research, 43(8), 3899–3910. doi: 10.1093/nar/gkv294

Ebrahim, A., Lerman, J. A., Palsson, B. O., & Hyduke, D. R. (2013). Cobrapy: Constraints-based reconstruction and analysis for python. BMC Systems Biology, 7(1), 74. doi: 10.1186/1752-0509-7-74

Faria, J. P., Rocha, M., Rocha, I., & Henry, C. S. (2018). Methods for automated genome-scale metabolic model reconstruction. Biochemical Society Transactions, 46(4), 931–936. doi: 10.1042/BST20170246

He, H., Hoper, R., Dodenhoff, M., Marlütre, P., & Bar-Even, A. (2020). An optimized methylotol assimilation pathway relying on promiscuous formaldehyde-condensing aldolases in E. coli. Metabolic Engineering, 60, 1–13. doi: 10.1016/j.meteng.2020.03.002

Heirendt, L., Arreckx, S., Pfarr, T., Mendoza, S. N., Richelle, A., Heinken, A., ... Fleming, R. M. T. (2019). Creation and analysis of biochemical constraint-based models using the COBRA Toolbox v.3.0. Nature Protocols, 14(3), 639–702. doi: 10.1038/s41596-018-0094-2

Kamisawa, M., & Goto, S. (2000). Kegg: Kyoto encyclopedia of genes and genomes. Nucleic Acids Research, 28(1), 27–30. doi: 10.1093/nar/28.1.27

Karp, P. D., Paley, S. M., Mushford, P. E., Krummenacker, M., Billington, R., Kohari, A., ... Caspi, R. (2020). Pathway Tools version 24.0: Integrated software for Pathway/Genome Informatics and Systems Biology. ArXiv Preprints, 1510.03964 [q-bio]. Retrieved from http://arxiv.org/abs/1510.03964

Karp, P. D., Billington, R., Caspi, R., Fulcher, C. A., Lahmmerre, M., Kohari, A., ... Subhraveti, P. (2019). The BioCyc collection of microbial genomes and metabolic pathways. Bioinformatics, 35(20), 3813–3814. doi: 10.1093/bioinformatics/btz540

Kanehisa, M., & Goto, S. (2000). Kegg: Kyoto encyclopedia of genes and genomes. Nucleic Acids Research, 28(1), 27–30. doi: 10.1093/nar/28.1.27

King, Z. A., Dräger, A., Ebrahim, A., Sonnenschein, N., Lewis, N. E., & Palsson, B. O. (2015). Escher: A web application for building, sharing, and embedding data-rich visualizations of biological pathways. PLOS Computational Biology, 11(5), e1004321. doi: 10.1371/journal.pcbi.1004321

King, Z. A., Lu, J., Dräger, A., Miller, P., Federowicz, S., Lerman, J. A., ... Lewis, N. E. (2016). BiGG Models: A platform for integrating, standardizing and sharing genome-scale models. Nucleic Acids Research, 44(D1), D515–D522. doi: 10.1093/nar/gkv1049

Machado, D., Andrejev, S., Tramontano, M., & Patil, K. R. (2018). Fast automated reconstruction of genome-scale metabolic models for microbial species and communities. Nucleic Acids Research, 46(15), 7542–7553. doi: 10.1093/nar/gky537

Meedsk, G. L., Moutinho, T. J., & Papin, J. A. (2020). Medusa: Software to build and analyze ensembles of genome-scale metabolic network reconstructions. PLOS Computational Biology, 16(4), e1007847. doi: 10.1371/journal.pcbi.1007847

Mendoza, S. N., Olivier, B. G., Molenaar, D., & Teusink, B. (2019). A systematic assessment of current genome-scale metabolic reconstruction tools. Genome Biology, 20(1), 158. doi: 10.1186/s13059-019-1769-1

Mord, J. M., Lloyd, C. J., Brunk, E., Mih, N., Sastry, A., King, Z., ... Palsson, B. O. (2017). iML1515, a knowledgebase that computes Escherichia coli traits. Nature Biotechnology, 35(10), 904–908. doi: 10.1038/nbt.3956

Orth, J. D., Fleming, R. M. T., & Palsson, B. O. (2010). Reconstruction and Use of Microbial Metabolic Networks: The Core Escherichia coli Metabolic Model as an Educational Guide. EcoSal Plus, 4(1). doi: 10.1128/ecosalplus.4.1.1

Pereira, V., Cruz, F., & Rocha, M. (2021). MWPYpy: A computational strain optimization workbench in Python. Bioinformatics, 37(16), 2494–2496. doi: 10.1093/bioinformatics/btab013

Sever, S. M., Liu, F., Zeng, Q., Jeffries, J., Faria, J. P., Edirisinghe, J. N., ... Henry, C. S. (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 Research, 49(D1), D575–D588. doi: 10.1093/nar/gkaa746

Wang, H., Marcišauskas, S., Sánchez, B. J., Domenzain, I., Hermansson, D., Ågren, R., ... Kerkhoven, E. J. (2018). RAVEN 2.0: A versatile toolbox for metabolic network reconstruction and a case study on Streptomyces coelicolor. PLOS Computational Biology, 14(10), e1006541. doi: 10.1371/journal.pcbi.1006541