

E-cyanobacterium.org: A Web-based Platform for Systems Biology of Cyanobacteria

Matej Troják¹, David Šafránek¹, Jakub Hrabec¹,
Jakub Šalagovič¹, Františka Romanovská¹ and Jan Červený^{2*}

¹ Faculty of Informatics, Masaryk University, Brno, Czech Republic

² Global Change Research Centre AS CR, v. v. i., Brno, Czech Republic

Abstract. E-cyanobacterium.org is an online platform providing tools for public sharing, annotation, analysis, and visualization of dynamical models and wet-lab experiments related to cyanobacteria. The platform is unique in integrating abstract mathematical models with a precise consortium-agreed biochemical description provided in a rule-based formalism. The general aim is to stimulate collaboration between experimental and computational systems biologists to achieve better understanding of cyanobacteria.

1 Introduction

The complexity of dynamical processes occurring in biology is inherent due to the fact that living cells must be responsive to a highly dynamic environment. This applies especially to the family of photosynthetic organisms such as cyanobacteria in which the biophysical processes scale from electron transfers interacting with metabolic biochemistry to genetic regulation and back. Remarkable effort towards a consistent and coherent knowledge-base of cyanobacteria modelling is one of the key activities of CyanoNetwork³, the international network of top-leading experts on studying these unique bacteria. In cyanobacteria, biophysical processes span in vastly different time scales from femtoseconds to seasons and from individual molecules to aquatic and terrestrial ecosystems.

Such a community-wide modelling effort is notably accelerated by an interactive online platform. In particular, models need to be translated into a unified format, formalised, and uniformly annotated. This allows the models to be fully understood and re-used in the original form, compared with each other, and with wet-lab experiments. To this end, one needs a unified and flexible framework to fully represent partial models and the respective biological knowledge — the involved components as well as their interactions.

An existing resource to inspire further expansion of cyanobacteria modelling are the established web repositories of curated and well-annotated models traversing already through many branches of biology [3, 17, 19, 12]. Unfor-

* This work has been supported by the Czech National Infrastructure grant LM2015055 and by the Czech Science Foundation grant GA15-11089S.

³ <http://www.cyanoteam.org>

tunately, cyanobacteria models are strongly under-represented in these repositories, probably because of the natural differences that exist between common bacteria and phototrophic bacteria. There exist several online tools presenting biological networks or genome knowledge [4, 12, 1]. However, due to enormous complexity of biological processes, it is a challenge to develop tools presenting biological networks alongside with executable or mathematical models. Focusing on domain-specific organisms allows us to integrate the knowledge and present it in a concise and understandable form. This has been already proven on examples of well-known model organisms such as *E. coli* [14] or *C. elegans* [7]. Nevertheless, these resources do not couple the presented knowledge with modelling.

In this tool paper, we present an online platform for cyanobacteria processes — *e-cyanobacterium.org*⁴. The platform integrates several dedicated tools and is distinct in the following aspects: formal rule-based representation of biochemical interactions facilitated by cyanobacteria biochemical entities; repository of kinetic models providing basic analysis tools online (simulation, custom data sets, basic static analysis); integration of models within the rule-based description and export to SBML [10]; storage, maintenance and presentation of experimental data; content visualisation (graphical presentation of models, biochemical space and modelling/experimental data).

The presented release of *e-cyanobacterium.org* is implemented as an extension of our general database tool-kit — the so-called *comprehensive modelling platform* [15]. Key updates are the formal rule-based language BCSL, wet-lab experiments module, and improved analysis of models. Most importantly, several well-annotated and curated models developed by the consortium are provided within *e-cyanobacterium.org* including the formalisation of the respective part of biochemical space in BCSL.

2 Web Platform Overview

The platform consists of several dedicated modules (Fig. 1) all connected to a central module — *Biochemical Space* (BCS) [16] — that is the backbone of the platform. BCS provides formal description of the biological problem and it is based on the hierarchy of selected biological processes. It is accompanied with schemas representing relevant biological processes in the context of cyanobacteria. For each process, there are presented relevant models, chemical entities, and rules. Presentation of every process includes detailed information and links to relevant internal and external sources.

2.1 Biochemical Space

Biochemical Space is constructed from hierarchy of entities interacting in rules formally specified in *Biochemical Space Language* (BCSL). The main advantage of BCSL is the adoption of the most important aspects of rule-based features while still keeping the syntax human-readable and accessible to communities outside computer science. On the one hand, BCSL has executable semantics that

⁴ <http://www.e-cyanobacterium.org>

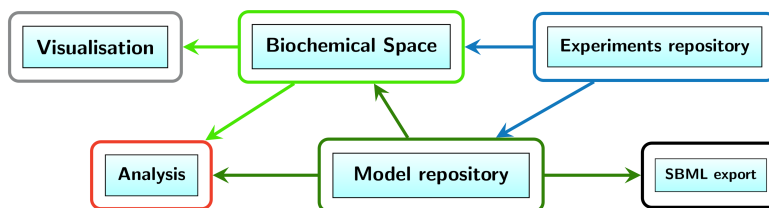


Fig. 1. Scheme of interconnections between modules of E-cyanobacterium.org platform.

allows basic analysis and consistency checking. On the other hand, the language includes constructs for detailed biological annotation reflecting the known bioinformatics databases and ontologies. BCSL is developed with the consideration of new extensions in SBML level 3. Once the SBML rule-based support becomes to be actively used in tools, our platform will provide relevant export filters.

Rule description provides details of rules, explicit enumeration of substrates and products, and available annotations. Moreover, the rule is schematically visualised (Sec. 3.2) and presented with appropriate links to the process hierarchy.

Entity description provides information about associated models and rules. Links to the process hierarchy and to external sources (UniProt [2], Kegg [12], GeneOntology [1], etc.) are available.

2.2 Model Repository

Model repository is a collection of implemented mathematical models describing particular parts of biological processes. Every model is represented as a set of ordinary differential equations generated from the model reaction network. Models are integrated within BCS. In particular, each model component should be related to some BCS entity and each model reaction should be related to some BCS rule. Moreover, a model is associated with some parameter value sets (data sets) that enable simulation (Sec. 3.3) in a particular biologically-relevant scenarios. Additionally, several basic non-parameter-specific static analysis techniques (Sec. 3.4) based on model stoichiometry are also provided.

An implemented model then includes complete biological annotation (Sec. 3.1) of all components and reactions that is provided by mapping to BCS. This might help to find connections and overlaps among models. Further, implemented model can be exported to SBML (level 2).

Currently, the repository contains two models describing circadian clock (Miyoshi et al. 2007 [18], Hertel et al. 2013 [8]) and a kinetic model of metabolism (Jablonsky et al. 2014 [11]). Two other models present unpublished results – dynamics of carbon fluxes (Müller et al.) and photosynthesis (Plyusnina et al.).

2.3 Experiments Repository

Experiments repository is a tool for storage and presentation of time-series data from wet-lab experiments. Every experiment is well-grounded by precise description (device, medium, organism, etc.) and appropriate annotations. Experiments

are structured – several time series data can be attached to a single experiment. Every time series targets a specific list of measured substances together with time stamps of the individual measurements. Data can be imported/exported in simple text formats. Time series are visualised in a chart (Sec. 3.3).

Registered users can add their own experiments while keeping the selected experiments private or public. For repeated experiments, annotations and details can be cloned from previously inserted experiments.

3 Website Features

E-cyanobacterium.org provides support for modelling and analysis of biological systems. The most important fact is that all these parts are integrated within the Biochemical Space. Therefore it is possible to reveal non-trivial relations between biochemical substances and models.

3.1 Annotation

Annotation is an important task in modelling of biological systems. This is how biological knowledge is mapped to the mathematical description. Our platform considers annotation as an inherent and compulsory part of the modelling procedure. The following aspects of annotation are supported:

- BCS creation and maintenance,
- model annotation by mapping to BCS,
- experiment annotation.

BCS is being continuously extended and revised by the consortium. Researchers supplying their models are required to integrate the models within the current BCS. This gives good feedback to BCS maintainers. In the experiments repository module, emphasis is given to description of conditions under which an experiment has been performed, which is important for interpretation of the data as well as the possibility of reconstructing the measurements.

3.2 Visualisation

Static visualisation is provided by means of schemas showing the process hierarchy with most important objects from BCS. Graphical schemes are supplied with detailed information on the visualised elements of BCS. This is achieved in the information panel below the scheme. An accompanying feature is filtering displayed data to *All* or *Visible* elements. The former option extends the content to all relevant entities and rules that take the part in the displayed process but are not necessarily visualised (e.g., this applies to small “universal” molecules such as ATP, ADP, etc.). Moreover, there is a special visualisation framework for complex networks such as metabolism. This widget allows one to zoom in a specific part of the complex network. This feature is currently employed for

metabolic map of cyanobacteria. Owing to requests coming out of the consortium, we have decided to employ manually handled visualisation. In next release, we plan to provide automatically generated SBGN representation of the schemes.

Rule details (Sec. 2.1) are enriched by means of automatically generated visualisation. For every rule, a graphical scheme that displays all its substrates and products is generated.

3.3 Simulation

With every implemented model (Sec. 2.2) it is possible to execute simulations. Registered users can change initial conditions and parameters of the model and set the simulation options (the numerical solver and its parameters). To apply such settings in the simulation, they have to be saved in terms of a *custom dataset* (public or private). The simulation chart is generated for all available datasets and the platform GUI allows one to switch between them. The chart is interactive and allows one to change the axis type, zoom in the selected curves, and show a focused value on a curve.

Simulation data are accessible by exporting them in several data formats. Similar options are also available in charts visualising experimental data in the experiments repository. Additionally, the simulated model including the selected dataset can be downloaded as an Octave file [6]. Additionally, the platform allows one to use different externally called numerical solvers for simulation. In particular, model administrators can select either Octave (default), COPASI [9], or remote simulation with COPASI web services [13].

3.4 Static analysis

At the current stage, there are three static analysis tasks available. *Matrix analysis* produces incidence matrix, *Conservation analysis* produces mass conservation analysis (moiety conservation), and *Modes analysis* produces elementary flux modes. For these tasks, the well-acclaimed third party tool COPASI [9] is used which allows one to download the task data by means of an SBRML file [5].

4 Conclusion

E-cyanobacterium.org provides several features that contribute to production and presentation of models targeting cyanobacteria. The most principal effort is to interlink biological knowledge with benefits of computational systems biology tools. This is enabled by means of a novel formal notation – Biochemical Space – which allows us to integrate computational models with the biological knowledge and wet-lab experiments.

For future work, we plan to improve the mapping between mathematics and biology, to enhance the website with more analysis tools (currently we are implementing an interface for existing property monitoring and robustness analysis tools such as [20]), and to automatise the comparison of models against experimental data. Moreover, biochemical space of cyanobacteria is continuously being extended and improved with interactive visualisations of reaction networks based on formal description provided in Biochemical Space Language.

References

1. Ashburner, M., et al.: Gene ontology: tool for the unification of biology. the gene ontology consortium. *Nature genetics* 25(1), 2529 (2000)
2. Bateman, A., et al.: Uniprot: a hub for protein information. *Nucleic Acids Research* 43(D1), D204–D212 (2015)
3. Beard, D.A., et al.: Cellml metadata standards, associated tools and repositories. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* 367(1895), 1845–1867 (2009)
4. Croft, D., et al.: The reactome pathway knowledgebase. *Nucleic Acids Research* 42(D1), D472–D477 (2014)
5. Dada, J.O., et al.: Sbrml: a markup language for associating systems biology data with models. *Bioinformatics* 26(7), 932–938 (2010)
6. Eaton, J.W., et al.: GNU Octave version 3.0.1 manual: a high-level interactive language for numerical computations. CreateSpace Independent Publishing Platform (2009)
7. Harris, T.W., et al.: Wormbase: a comprehensive resource for nematode research. *Nucleic Acids Research* 38(suppl 1), D463–D467 (2010)
8. Hertel, S., et al.: Revealing a two-loop transcriptional feedback mechanism in the cyanobacterial circadian clock. *PLoS Comput Biol* 9(3), 1–16 (2013)
9. Hoops, S., et al.: Copasi – a complex pathway simulator. *Bioinformatics* 22(24), 3067–3074 (2006)
10. Hucka, M., et al.: The systems biology markup language (sbml): a medium for representation and exchange of biochemical network models. *Bioinformatics* 19(4), 524–531 (2003)
11. Jablonsky, J., et al.: Multi-level kinetic model explaining diverse roles of isozymes in prokaryotes. *PLoS ONE* 9(8), 1–8 (2014)
12. Kanehisa, M., et al.: Kegg as a reference resource for gene and protein annotation. *Nucleic Acids Research* 44(D1), D457–D462 (2016)
13. Kent, E., et al.: Condor-copasi: high-throughput computing for biochemical networks. *BMC Systems Biology* 6(1), 1–13 (2012)
14. Keseler, I.M., et al.: Ecocyc: fusing model organism databases with systems biology. *Nucleic Acids Research* 41(D1), D605–D612 (2013)
15. Klement, M., et al.: A comprehensive web-based platform for domain-specific biological models. *Electronic Notes in Theoretical Computer Science* 299(0), 61 – 67 (2013)
16. Klement, M., et al.: Biochemical space: A framework for systemic annotation of biological models. *Electronic Notes in Theoretical Computer Science* 306(0), 31 – 44 (2014), proceedings of the 5th International Workshop on Interactions between Computer Science and Biology (CS2Bio14)
17. Le Novère, N., et al.: Biomodels database: a free, centralized database of curated, published, quantitative kinetic models of biochemical and cellular systems. *Nucleic Acids Res* 34, D689–91 (2006)
18. Miyoshi, F., et al.: A mathematical model for the kai-proteinbased chemical oscillator and clock gene expression rhythms in cyanobacteria. *Journal of Biological Rhythms* 22(1), 69–80 (2007)
19. Olivier, B.G., et al.: Web-based kinetic modelling using jws online. *Bioinformatics* 20(13), 2143–2144 (2004)
20. Rizk, A., et al.: A general computational method for robustness analysis with applications to synthetic gene networks. *Bioinformatics* 25(12), i169–i178 (2009)