

Biochemical Space: A Framework for Systemic Annotation of Biological Models¹

M. Klement, T. Děd, D. Šafránek

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

J. Červený, S. Müller, R. Steuer

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

Abstract

In this tool paper, we target the problem of unique annotation of organism-specific computational models presented in a public model database. In particular, we present Biochemical Space, a novel annotation methodology accompanied with a set of software tools that allow to create, manage and maintain the Biochemical Space content. The main idea behind is to create a transparent well-annotated reaction network of chemical entities and elemental reactions onto which the mathematical models are projected. For a given organism, the Biochemical Space represents a unifying platform for understanding of the related biological processes. The contribution of the methodology is three-fold: (i) systemic projection of models to a well-structured biological knowledge, (ii) simplification of annotation procedure, (iii) targetting several problems such as the presence of lumped model variables, combinatorial explosion in chemical modifications of entities, and hierarchical organisation of locations of individual entities. In these aspects the Biochemical Space goes beyond the features of current standards such as SBML. Application of the framework is demonstrated on a set of annotation data compiled for complex cyanobacteria processes.

Keywords: biological models, model annotation, systems biology, cyanobacteria

1 Introduction

In the last decade, many different platforms aiming to speed up and facilitate propagation of systems biology findings were revealed. They provide sharing of dynamical models. These “online” models allow us to simulate behaviour of a living organism or give us information about a single part of a living organism or its integration in large complex units. It must be said that no such functionalities would be widely available without the Internet and web.

¹ The work has been supported by the EC OP project No. CZ.1.07/2.3.00/20.0256.

To target the issues of rigorous presentation of complex biological models without congesting the users with complicated math behind the models, we have started the activity e-cyanobacterium¹. It aims at handling the domain-specific problems profoundly. It also intends to trade on advantages arising from repository and database synthesis, which all in all allow to create complex relationships between models and ontological knowledge bases, different types of simulations and analysis, and integration of models in the domain-specific context.

Biochemical Space (BCS) represents reaction networks linked to existing ontologies. Since models primarily target mechanisms behind biological processes, process-based hierarchy is used to navigate through the biochemical space. To this end, for each level of the process hierarchy, a visual representation of relevant biochemical mechanisms allowing to understand the non-trivial biological context of models is accompanying the biochemical space data.

The concept of BCS makes a crucial part of *Comprehensive Modeling Space* (CMS), a general platform for computational modelling and analysis of biological processes, first introduced in [13] as a concept for formal representation of internally consistent reduced models of oxygenic photosynthesis [15] and further refined to a general platform in [7]. In general, the main goal of BCS as a part of CMS is to simplify model-building tasks by providing simple and clear way of notation easily understandable by modellers and biologists.

Current notation syntax is compared to other available databases and their file formats focused mainly on the structure of transmitted data in a way that notation is clear and easy to remember. In contrast, *Biomodels.org* [10] is aiming at the dynamic models shared in SBML [5] format which is complex and non-editable by ordinary users without a proper tool. Similarly to *Biomodels.org*, *CellML.org* [11] is a database of dynamic models, which uses a proprietary format similar to SBML called CellML. Unlike previous, *CyanoBase* [12] is an annotation database that describes the physical structure, chromosomes, genes, etc. of cyanobacteria organisms where the data are shared only in a raw format. KEGG [6] is the closest format to BCS and was the former candidate for the intended task, unfortunately it does not support combinatorial states, locations of entities inside of reactions and hierarchy.

In contrast to the well-acclaimed standard provided by SBML [5,9] that might be also used for representation of a biochemical space, our notion of BCS totally avoids issues related with dynamical models (for that purpose we use just SBML level 2). As an annotation platform purely focused on process-level description, BCS goes beyond SBML level 2 in generalization of compartments in a hierarchy of locations, in introducing entity states, and in dealing with related combinatorial explosion. These issues are solved in detail by rule-based approaches [2,4] and there is a draft of a package for SBML level 3 in preparation (multi). Since our notation relies on a simple textual base and focuses on a simple but still reasonably precise and compact description maintainable by biologists, our notation is rather closer to KEGG.

¹ <http://www.e-cyanobacterium.org>, <http://www.cyanoteam.org>

In [7] we have presented the general database platform for domain-specific models. In this paper we focus on its inherent part formed by BCS, we describe the annotation workflow, formal structure of annotation data, and the toolset developed for its maintenance. Moreover, concrete applications of BCS are demonstrated on several processes of cyanobacteria.

2 Biochemical Space

BCS provides well described biological background for mathematical models of processes taking place in specific type of organism. Complete BCS can act as a connection between existing ontologies and these models later on, which is enabled by well-defined annotations. For creation and for modifications of BCS, a textual file with specific syntax is used. This file offers a human readable format of BCS which can be easily edited in a special purpose editor and visualized on the web platform. First part of BCS is represented by set of entities while the second part contains elemental reactions working on this set. In our case, the consortium of scientists from CyanoTeam¹ is involved in modelling of several cyanobacterial processes and in creation of BCS.

Emphasis on well-defined and complete annotations is given during creation of BCS. Therefore, links to other ontologies must be filled in for each entity and reaction. Unique IDs provided by these ontologies can help to automatically detect duplicities later on in the suggested Annotation Editor. IDs are also used to create hypertext links to related ontology on the web which serves as one side of the already mentioned connection between ontologies and models. At this moment, links to KEGG, ChEBI, CyanoBase [12] and other databases are supported. One entity or reaction can have multiple links to different databases as well as to the same database. Example for the first case can be the presence of an entity in ChEBI as well as in KEGG databases. The second case can occur while linking enzymatic reactions. Then, EC number (functioning here as a descriptor of the reaction mechanism) can be affiliated with KEGG ID of the reaction. When an entity of the type protein is added to BCS, one might want to add a sequence of genes as a descriptor of this protein. One link (in our case to CyanoBase) is created for each gene separately. If more than one gene sequence is present, additional information about each sequence is put into notes. Notes generally carry internal information about an entity or a reaction. Finally, a comma is used as a separator between records within links and notes fields. In most cases, ontologies contain general information about entities and about reaction mechanisms. However, we are dealing with a single specific type of organism, therefore verbal description of the role of entity and reaction within this organism can be put down to another line.

Example of description, links and notes information for entity:

DESCRIPTION: Protein involved in hydrolysis of N-acylated or N-acetylated amino acids
LINKS: KEGG::ec3.5.1.14, CBS::slr1653, CBS::sl10100
NOTES: ChEBI link is missing

Example of link information for reaction:

LINKS: KEGG::ec3.5.1.14, KEGG::R00669

The fact that most fields in entity and reaction definitions in a BCS file are tightly coupled with information from linked ontologies is the reason why annotation information was mentioned first. In the first place, such an attribute is a name, which is taken from ontologies or follows conventional naming of biochemical compound or process. IDs of entities are likely to be represented by aliases of names which CyanoTeam agreed on. KEGG ID, ChEBI ID or internal ID is used if no reasonable ID is available. IDs of reactions are always internal (i.e., there is no need to fill in this field).

Example of ID and name of entity:

ENTITY ID: pq
ENTITY NAME: plastoquinone

Example of ID and name information for reaction:

REACTION ID:
REACTION NAME: plastoquinone reduction

Entity in our interpretation is a bounded space or part of a specific type of organism. BCS can contain entities ranging from small ones (a photon or an atom) to large ones (intracellular space). Our goal is to make BCS as simple as possible. Since states of entities (oxidised, reduced, etc.) in existing ontologies are usually treated as another entity, number of entities is quite large. To reduce such a complexity, entity states can be defined in our BCS. All states are enclosed in curly brackets and they are separated by comma. Relationship entity-state is a parent-child, respectively. All information about entity is then inherited to its state unless defined otherwise. The ID of an entity and its state in curly brackets form together unique identifier. If no state is specified, default value is the “neutral” state.

As well as in a real biological system, compartmentalisation can be done in BCS as well. One entity can be a part of another, larger entity. Then we can say that the larger entity plays a role of location for the smaller one. An entity can have one or more locations according to its occurrence. Notation in BCS file is to use IDs of locations (entities) and comma as a separator. A set of entities can be quite large and so it would be useful to cluster entities into related groups. Locations are not best suited for this task, since entity can be found in multiple places. For this purpose, class information is used. Classification describes type of an entity in a sense of functional or structural nature and entities can be easily grouped together by this attribute.

Example of complete information for an entity:

ENTITY ID: HCO3
ENTITY NAME: hydrogencarbonate
CLASSIFICATION: small molecule
LOCATIONS: cyt, cell, liq
STATES: {-}
DESCRIPTION: Plays major role in carbon concentrating mechanism (CCM).
LINKS: CHEBI::17544
NOTES:
VISUAL:

An elemental reaction can also have an assigned classification. However, we treat them differently. The classification of reactions does not describe a type of reaction, but assigns a name of a high level process in which the reaction is involved. Moreover, unlike entity classification, there can be more than one classification for

one elemental reaction. Elemental in context of reactions means very detailed. Explanation of a consensus, on how detailed reactions should be considered, follows. When defining raw equation, identifiers of substrates and products are used to make notation of a reaction shorter, i.e., more readable. Then, after double colon, a location for each entity has to be specified. For example, that can play an important role for defining a reaction which acts on both sides of a membrane. Reaction itself has no location assigned, since it acts very often on a boundary of two locations (e.g., a cell membrane). All location IDs affiliated to each entity identifier then fully define where reaction acts. Information about substrates and products are separately enclosed in quotation marks. A stoichiometric number from \mathbb{N}^* can be put before first quotation mark for each entity. Implication and equivalence signs between substrates and products are used in order to depict irreversible and reversible reaction respectively. Plus sign is used as a separator between substrates and between products.

Example of complete information for a reaction:

```
REACTION ID:
REACTION NAME: plastoquinone reduction in the cytoplasmic membrane
EQUATION: "NADPH::cyt" + 5 "h{+}::cyt" + "pq::cym" =>
          "NADP{+}::cyt" + 4 "h{+}::pps" + "pqh2::cym"
CLASSIFICATION: reduction-oxidation reaction
MODIFIER: NDH1
DESCRIPTION: oxidation of NADPH and reduction of plastoquinone in the cytoplasmic membrane
LINKS:
NOTES:
VISUAL:
```

In some cases, emphasis on detail description leads to very complex BCS. Abstraction of some processes is needed to keep BCS as simple as possible. First, enzymatic reactions can be abstracted. There should be at least two different reactions in that case (one for a substrate binding and another for a catalytic step). Instead, since an enzyme does not change during this process, it is affiliated to the reaction as a modifier. Another case, where even bigger abstraction comes into account, is when several electrons play 'musical chairs' inside protein complexes. The issue is that parts of processing protein complex can have different non-stable states during a short period of time. When one tries to define all reactions among these proteins, combinatorial explosion of number of states of the complex arises. Not all of these combinations are biologically correct, but even without non-biological cases, the number of states stays still significant. For the purpose of our BCS, we come up with a solution inspired by the enzymatic reaction mentioned above. We treat a protein complex as a location on which entities change its state (not necessarily proteins) and we abstract from background processes. We can see the reaction as a change of a state of our complex (location), which is correct. In the notation of the reaction, plus sign between substrates and between products is replaced by bar sign ($\bar{}$). A plus sign can figure in a reaction together with the bar sign in case that the complex reacts with an entity, which is not located in the complex, and the complex changes its state at the same time. By the mentioned abstraction we violate the formalness, which is, e.g., prescribed by SBML, but the achieved simplification is tremendous.

Example of special reaction:

REACTION NAME: oxidation of $Qa-$ and reduction of $Qb-$ in PSII RC
 EQUATION: " $qa\{-\}::ps2$ " | " $qb\{-\}::ps2$ " \rightleftharpoons " $qa\{n\}::ps2$ " | " $qb\{2-\}::ps2$ "
 CLASSIFICATION: light reactions of photosynthesis
 MODIFIER:
 DESCRIPTION: the electron from the reduced primary quinone acceptor of PSII $Qa-$
 LINKS:
 NOTES:
 VISUAL:

The textual format of BCS is a readable and easy-to-use representation of reaction network. But, with support of a web based platform (CMS), we can enhance these features. A high level process hierarchy can be created on-line. Entities, elemental reactions and models are then mapped into this hierarchy which then becomes main connection among BCS, other ontologies and mathematical models. While creating hierarchy as well as while mapping entities and reactions inside, location and classification information are taken into account hand-to-hand with existing hierarchies in other ontologies. A whole process cannot be easily automated due to high level of abstraction so the annotators have to agree on form of hierarchy. Each level of hierarchy can have an interactive visual representation of corresponding process added.

3 Model Annotation Workflow

Dynamical models provide a very effective way for representing dynamics of biological systems. They often employ advanced mathematical techniques such as entity and reaction lumping, which may complicate the biological interpretation. The way how to describe these abstractions relies on a modeller. Therefore, precise annotations have to be created in order to make models accessible and understandable. SBML offers formal syntax which unify model description. Our aim is slightly different. CMS offers a way where models are directly mapped on biochemical space. Figure 1 describes the whole annotation process in detail, i.e. how BCS is created and how models are mapped into BCS.

First, a model of biological process have to be created. All differential equations, rates, initial concentrations and other specific features are then manually uploaded into the CMS database. The annotation of the model begins. People responsible for the annotation are either modellers themselves or are members of a consortium. Modelled process should be described from the biological point of view in textual form of BCS with syntax mentioned in the previous chapter. Single file is created for every new model. Next, existing BCS is exported from database and compared with the new file in suggested annotation editor. All information missing in existing BCS is then uploaded to the database. Process is repeated for every new file. The loop ideally ends when there is nothing to add into BCS. We can approach this state when we deal with simple, well-studied unicellular organisms. However the complexity of biological systems is enormous and so there will be still things to add. Last remaining thing is to create a simplified high-level process hierarchy of reactions and entities on web and support it with an interactive visualisation for each level of the hierarchy.

As for the model mapping, modeller has to define which subset of BCS corre-

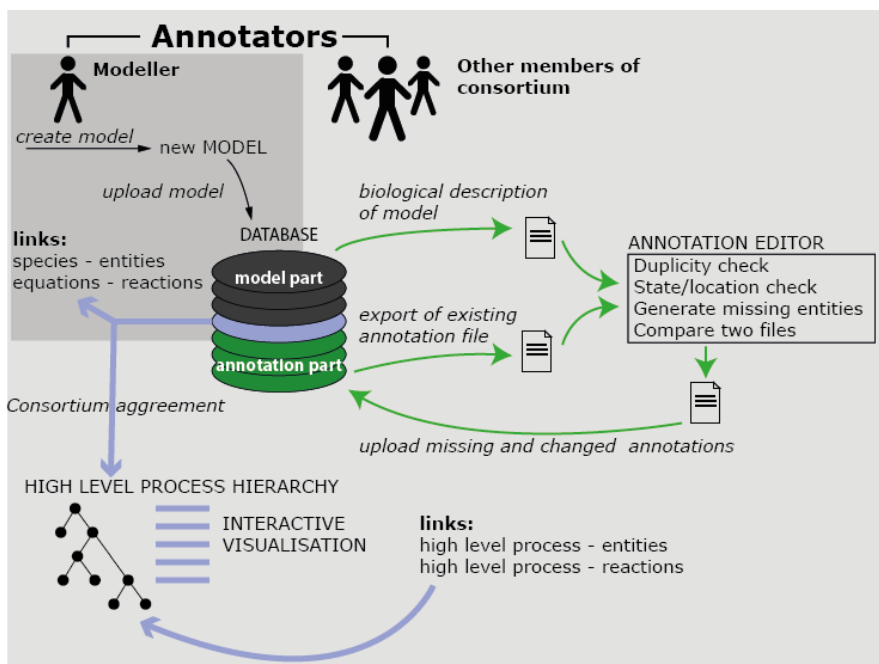


Fig. 1. Description of annotation process.

sponds to individual model parts. Entities and reactions from BCS are mapped to model species and ODEs respectively. In many cases, such direct mapping is not possible due to high level of abstraction. In case of model reaction (ODE), several reactions from BCS can be assigned to describe corresponding process. Such mapping is the final step in establishing inter-specialisation connection between biology and mathematical modelling.

4 Biochemical Space Editor and Online Tool Support

Up to this point the biochemical space was described as set of formal rules where all the required data have strictly defined format and structure. If these conditions were not met defined data could not be easily shared and understood among the scientific communities and also automatically processed by computers. For the purposes of creating, processing, sharing and automatic analysis several tools were created each focusing on a different part of the problem and the target audience of biochemical space. These tools also focus on interoperability between similar formats and ease of use. The current version of the notation of biological space was designed as textual file with specific syntax with regard to the availability of tools capable of editing of this format. In case of "emergency" or unavailability of current tools textual file can be manually edited in any text editor for those necessary adjustments with knowledge of basic rules. But ordinary text editors do not provide functionality needed to maintain the data in a consistent state and also does not assure adequate filling of required fields and coherence of data while these properties are very important for file sharing and unambiguity of other manipulation.

4.1 Annotation Editor

A basic tool for editing of textual files in described syntax for biochemical space is AnnotationEditor. This tool is written in Java and supports modification of multiple files simultaneously and is compatible with all common operating systems. This editor was created as a necessary basis for the modification, maintenance and creation of these textual files. It also supports import of data and merging of multiple files in mentioned format. AnnotationEditor provides maintenance of two main groups of data which are entities and reactions. Besides these basic features for maintenance tool also supports important validation features required for consistency of entities definitions which are:

- Duplicity check, which verifies the occurrence of multiple entities within currently edited files
- Validation of entities states, which checks for the presence of definition for all entities' states used in the reactions and relationships
- Location check of entities which verifies definition of entities used as a location within the definition of entities and reactions
- Generation of missing entities, location and states in the case that any inconsistencies were accidentally made so that these missing items does not have to be defined manually

The editor also supports the management of reactions which contains defined entities, location and states. Modification of equation is processed in real time and changes are directly reflected in the other fields defined for the reaction. Also this tool must contain set of functions that provide extended validation of the reactions data which are:

- Duplicity check which search reactions in two different ways, either on exact match or on the basis of similarity search of equations definitions
- Validation of entities and locations, which verifies whether entities and states from equation are defined in entities definition and also whether entity's definition contains location defined in reaction
- Generation of missing entities for reaction in the case that any inconsistencies were accidentally made and these missing items does not have to be defined manually

The GUI was designed to ease use and understanding in case user has basic knowledge of functions that should be supported. The main application screen consists of the menu with items at the top which provides mentioned functionality above. The remaining part of the screen is filled by two tabs containing entities and reactions. Both tabs are divided to two parts where on the left is navigation tree and right part displays detail of the selected item from navigation tree. An interesting feature is the possibility to display graph showing the selected entity or reaction in network way connecting defined items.

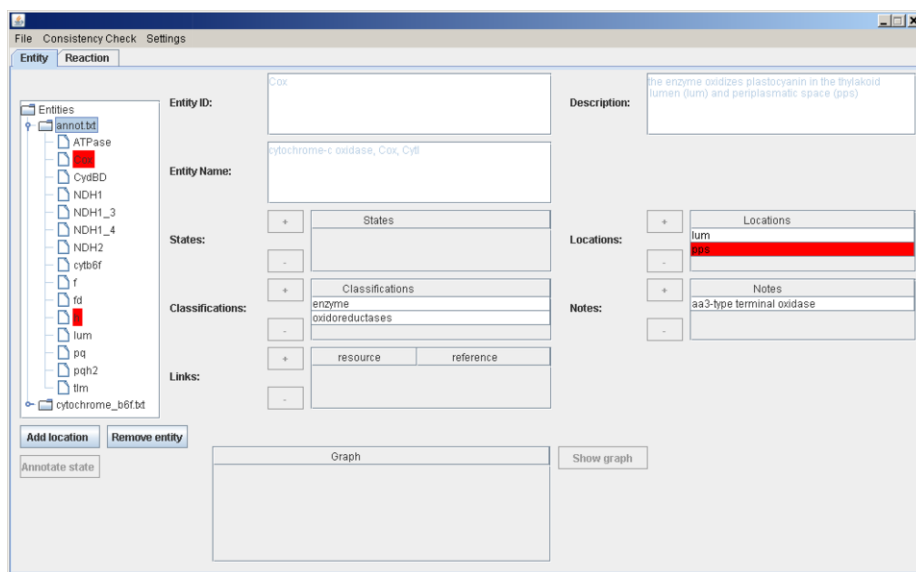


Fig. 2. Example screen of Annotation Editor.

4.2 Visualisation Editor

The basic format describing biochemical space does not support visualization and interconnection of entities and reactions to tree structure necessary for understanding of real-world distribution. Therefore, for e-cyanobacterium.org platform was created an online tool that provides additional functionality working as an extension to Annotation editor. Mentioned tool facilitates importing of already prepared and consistent textual files to an online database which can be used to transform and incorporate imported objects into the visualization structure. This process converts relatively flat structure representing entities, states and reactions from biochemical space to visualization tree showing the interconnection of individual layers or larger units and environments.

5 Case Study: Biochemical Space of Cyanobacteria

As a case study we have developed BCS for several biological processes of cyanobacteria. It makes a part of e-cyanobacterium.org², a web site focusing on a unified presentation of partial models explaining cellular processes of *Synechocystis* sp. PCC6803.

In the current version (March 2014), the biochemical space deployed in e-cyanobacterium.org covers the following processes of cyanobacteria: environmental processes, respiration and photosynthesis, and metabolism. Environmental processes focus on precise positioning of cyanobacteria into the context of its environment. Since the website primarily targets *in vitro* cultivation conditions in a bioreactor, we have compiled relevant elemental reactions. Processes of respiration

² <http://www.e-cyanobacterium.org>

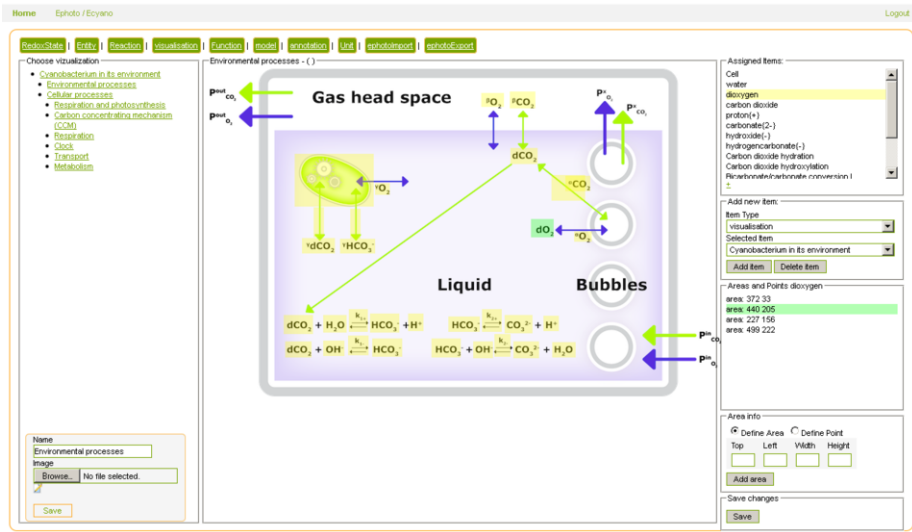


Fig. 3. Example screen of Visualisation Editor describing the environmental processes.

Carbon dioxide hydration	"CD2::liq" + "H2O::liq" <=> "HC03{-}::liq" + "h{+}::liq"
Carbon dioxide hydroxylation/dehydroxylation	"CD2::liq" + "OH{-}::liq" <=> "HC03{-}::liq"
Bicarbonate/carbonate conversion at low pH	"HC03{-}::liq" <=> "C03{2-}::liq" + "h{+}::liq"
Bicarbonate/carbonate conversion at high pH	"HC03{-}::liq" + "OH{-}::liq" <=> "C03{2-}::liq" + "H2O::liq"
Bubble-liquid carbon dioxide transfer	"CD2::bub" <=> "CD2::liq"
Carbon dioxide excretion	"CD2::cell" <=> "CD2::liq"
Bicarbonate excretion	"HC03{-}::cell" <=> "HC03{-}::liq"

Table 1

Elemental reactions representing environmental processes at the level of a bioreactor. First half of the table represents processes occurring inside the bioreactor medium whereas the second half of the table lists processes transferring the chemical entities among the compartments.

and photosynthesis cover the energetic components of cyanobacteria. Above these cellular processes, the metabolic part of the biochemical space forms a backbone that connects the bioenergetic components with metabolome and connects all key cellular processes with the general processes occurring in the environment.

5.1 Environmental processes

Our BCS targets processes occurring in the bioreactor. In fact, these processes make an interface between the environment and the cells. In particular, the interface is defined by a set of bidirectional gas exchange flows transporting carbon dioxide and dioxygen between the gas head space and the liquid (bioreactor media) containing the cells. Dioxygen and dissolved carbon dioxide are transported to/from the cells through the cell membrane. Carbon dioxide and dioxygen molecules are also considered to be transferred between the liquid and bubbles inside the liquid. Through the bubbles they are exported into the gas head space.

The processes occurring in the bioreactor media are represented by means of a set of elemental chemical reactions listed in Table 1 and visualised in Fig. 2. The chemical entities are qualified by locations that represent the considered compartments

Plastoquinone oxidation	"pqh2::t1m" + "O2::lum" + 2 "h+::lum" => "pq::t1m" + 2 "H2O::lum"
Plastoquinone reduction	"succinate::cyt" + "pq::t1m" => "fumarate::cyt" + "pqh2::t1m"
Exciton transfer	"pcb*::pbs" + "chl*::ps2" => "pcbn::pbs" + "chl*::ps2"
Photoenergy transfer	"chl*::ps1" "achln::ps1" <=> "chl*::ps1" "achl*::ps1"

Table 2

Sample of elemental reactions representing photosynthesis and respiration processes. First half of the table represents traditionally compartmentalized processes whereas the second half shows elemental reactions of combinatorial processes occurring on complex protein structures.

of the bioreactor (liq – the media, bub – bubbles, cell – the cells). Each entity is a well-known simple molecule that is adequately interlinked with its annotation term in ChEBI. Note that in this case the representation is fully compatible with SBML since the meaning of locations matches the SBML notion of compartments. The only specificity is in formal representation of reduced entities for which the state identifier in curly brackets is used. To simplify the representation we do not distinguish the dissolved and coagulate states of the gas as this would go far beyond the needs to directly represent just the bioreactor gas flow at the cell-medium interface. BCS of the bioreactor is displayed in the Entities and Reactions tabs below the process scheme after clicking on “Environmental processes” in the navigation panel of e-cyanobacteria.org homepage.

In the model repository, there is a kinetic model (Müller et al. 2014) describing the dynamics of the bioreactor processes. The model is mapped directly onto the BCS its state variables represent the individual chemical entities. The platform [7] allows to simulate the model and provides several time-course profiles that are currently being experimentally validated.

5.2 Respiration and photosynthesis

A crucial part of cyanobacteria BCS is made by elemental reactions representing photosynthesis and respiration. Both processes occur in a specific folds of the cell membrane called thylakoid membrane and are interlinked since they share several biological entities, e.g., thylakoid membrane transporters (plastoquinone). Photosynthesis serves as the source of energy taken from light and transferred into production of ATP and NADPH molecules with oxygen resulting as a by-product. Respiration has the same outcome but requires NADPH instead of light. In contrast to photosynthesis, respiration also occurs in the cytoplasmic membrane.

Entities of photosynthesis and respiration BCS are represented by several complex proteins (enzymes) residing on the thylakoid membrane (t1m) in the cell. Since the thylakoid membrane encloses the inner-membrane space called lumen (lum) where H_2O molecules are processed, there are basically three locations defined for this set of entities. Reactions occurring in the lumen, cytosol and inbetween the thylakoid membrane and these locations have classical form (examples are given in Table 2). However, electron transfer reactions occurring in the structure of complex processes (e.g., photosystems, cytochromes) lead to combinatorial explosion of all possible conformations. To avoid that, these reactions are represented by employing advanced BCS constructs (i.e., the operator '|'). Several examples are given in

Table 2, e.g., there is an example of a photoenergy transfer reaction occurring on the photosystem I where the excitation of chlorophyll is passed to a light-harvesting antenna and back. The ground state $\{n\}$ is changed to excited state $\{*\}$ and vice versa. Special entities "chl" and "achl" represent parts making a conformation of a complex entity. Location for these entities is now an enzyme – photosystem I (ps1). Another reaction, the exciton transfer, describes a similar situation but now between conformations of two different complexes phycobilisome (pbs) and photosystem II (ps2). The traditional '+' is used in that case. All the respective reactions are displayed in the Entities and Reactions tabs below the process scheme after clicking on "Respiration and photosynthesis" in the navigation panel of e-cyanobacteria.org homepage.

The model repository contains the model Plyusnina et al. 2014 that represents kinetics of photosynthesis and respiration. In this case, the model employs lumping of internal states and reduces the dynamics to eight state variables running on a slow time scale. BCS is therefore necessary to describe the mechanisms incorporated in the model but lost in lumping.

5.3 Metabolism

Metabolism makes the backbone of cyanobacteria cellular processes [14]. Genome-wide reconstruction of metabolic network for *Synechocystis* sp. PCC6803 has been published in [8]. The largest part of cyanobacteria BCS covers this network. Most of the metabolic reactions occur in the cell cytoplasm. Connection with respiration and photosynthesis processes at the level of BCS is interfaced through the shared simple molecules (ATP, ADP, NADPH, etc.) and through the enzyme complexes (photosystem I, photosystem II, etc.) that make locations for electron transfer reactions covered in respiration and photosynthesis.

Entities of metabolism BCS represent enzymes driving the metabolic reactions and individual simple molecules acting as metabolites. Enzymes are assigned to reactions as modifiers (catalysers) and they are annotated by EC numbers and genes via crosslinking to CyanoBase. This makes the genome-wide projection of the metabolism BCS. Since the reactions in metabolism have primarily the form of catalytic reactions and there are no combinatorial complexes, this part of BCS is compliant with SBML. Metabolism BCS is accessible on e-cyanobacterium.org home page after clicking on the Metabolism process in the navigation panel left.

Since there is currently no kinetic model describing the dynamics of the entire metabolic network due to unavailability of kinetic constants values, we currently work on kinetic models that cover small parts of the network. Additionally, we plan to enrich the model repository with a mathematical model [14] amenable for flux-balance analysis. This model will directly map onto the metabolism BCS with the exceptions of boundary conditions and metabolome composition variables that are model-specific and not covered in the biology description provided by the BCS.

6 Conclusions

We have presented the notion of biochemical space as an annotation base for complex cellular processes that differs from existing formalisms in several aspects. First, in contrast to model description languages, it does not incorporate model-level details (kinetic or rate functions, location volume). It rather fits as a well-organized reaction network annotation platform onto which kinetic models are mapped. Second, in contrast to existing annotation formalisms and ontologies it precises the notation regarding the hierarchical organisation of object locations and dealing with combinatorial object states. In conclusion, it can be seen as a stand-alone interface that clearly connects the mathematical models with the existing annotation databases.

We have presented a portion of a concrete BCS defined for cyanobacteria processes. In future work this particular BCS will be further refined with processes covering carbon-concentrating [1] and circadian clock [3].

References

- [1] Burnap, R., R. Nambudiri and S. Holland, *Regulation of the carbon-concentrating mechanism in the cyanobacterium synechocystis sp. pcc6803 in response to changing light intensity and inorganic carbon availability*, Photosynthesis Research **118** (2013), pp. 115–124.
- [2] Danos, V. and C. Laneve, *Formal molecular biology*, Theor. Comput. Sci. **325** (2004), pp. 69–110.
- [3] Hertel, S., C. Brettschneider and I. M. Axmann, *Revealing a two-loop transcriptional feedback mechanism in the cyanobacterial circadian clock*, PLoS Comput Biol **9** (2013), p. e1002966.
- [4] Hlavacek, W. S., J. R. Faeder, M. L. Blinov, R. G. Posner, M. Hucka and W. Fontana, *Rules for modeling signal-transduction systems*, Sci. STKE **2006** (2006), p. re6.
- [5] Hucka, M. et al., *The systems biology markup language (sbml): a medium for representation and exchange of biochemical network models*, Bioinformatics **19** (2003), pp. 524–531.
- [6] Kanehisa, M. and S. Goto, *Kegg: Kyoto encyclopedia of genes and genomes*, Nucleic Acids Research **28** (2000), pp. 27–30.
- [7] Klement, M., D. Šafránek, T. Děd, A. Pejznoch, L. Nedbal, R. Steuer, J. Červený and S. Müller, *A comprehensive web-based platform for domain-specific biological models*, Electronic Notes in Theoretical Computer Science **299** (2013), pp. 61 – 67.
- [8] Knoop, H., Y. Zilliges, W. Lockau and R. Steuer, *The metabolic network of synechocystis sp. pcc 6803: Systemic properties of autotrophic growth*, Plant Physiology **154** (2010), pp. 410–422.
- [9] Le Novère, N. et al., *Minimum information requested in the annotation of biochemical models (MIRIAM)*, Nat Biotech **23** (2005), pp. 1509–1515.
- [10] Li, C., M. Donizelli, N. Rodriguez, H. Dharuri, L. Endler, V. Chelliah, L. Li, E. He, A. Henry, M. I. Stefan, J. L. Snoep, M. Hucka, N. Le Novère and C. Laibe, *BioModels Database: An enhanced, curated and annotated resource for published quantitative kinetic models.*, BMC Systems Biology **4** (2010), p. 92.
- [11] Lloyd, C. M., J. R. Lawson, P. J. Hunter and P. F. Nielsen, *The cellml model repository*, Bioinformatics **24** (2008), pp. 2122–2123.
- [12] Nakao, M., S. Okamoto, M. Kohara, T. Fujishiro, T. Fujisawa, S. Sato, S. Tabata, T. Kaneko and Y. Nakamura, *Cyanobase: the cyanobacteria genome database update 2010*, Nucleic Acids Research **38** (2010), pp. D379–D381.
- [13] Nedbal, L., J. Červený and H. Schmidt, *Scaling and integration of kinetic models of photosynthesis: Towards comprehensive e-photosynthesis*, in: *Photosynthesis in silico*, Advances in Photosynthesis and Respiration **29**, Springer, 2009 pp. 17–29.

- [14] Steuer, R., H. Knoop and R. Machn, *Modelling cyanobacteria: from metabolism to integrative models of phototrophic growth*, *Journal of Experimental Botany* **63** (2012), pp. 2259–2274.
- [15] Šafránek, D., J. Červený, M. Klement, J. Pospíšilová, L. Brim, D. Lazár and L. Nedbal, *E-photosynthesis: Web-based platform for modeling of complex photosynthetic processes*, *Biosystems* **103** (2011), pp. 115–124.