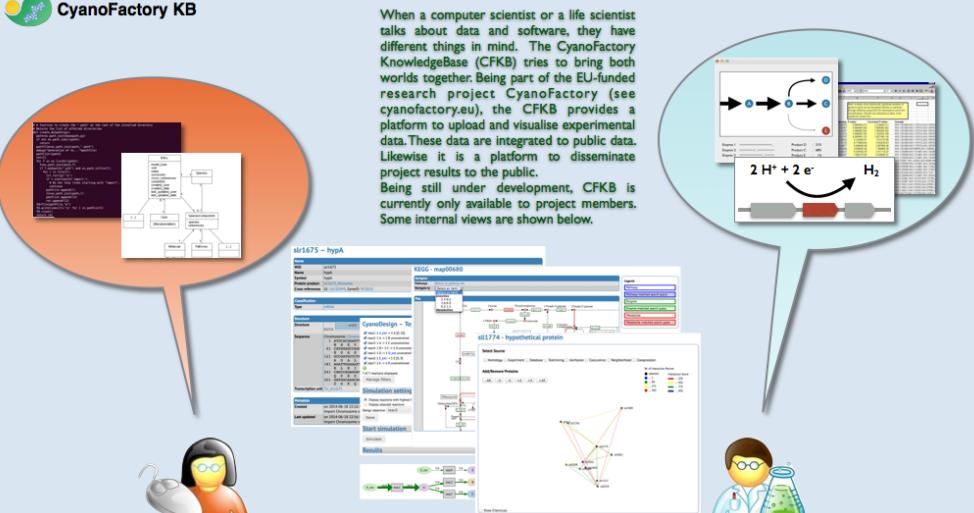


CYANOFACtORY WORKSHOP

THE CYANOFACtORY KNOWLEDGEBASE

 CyanoFactory KB

When a computer scientist or a life scientist talks about data and software, they have different things in mind. The CyanoFactory KnowledgeBase (CFKB) tries to bring both worlds together. Being part of the EU-funded research project CyanoFactory (see cyanofactory.eu), the CFKB provides a platform to upload and visualise experimental data. These data are integrated to public data. Likewise it is a platform to disseminate project results to the public. Being still under development, CFKB is currently only available to project members. Some internal views are shown below.



The interface includes a large orange speech bubble containing a screenshot of a metabolic pathway diagram with various nodes and arrows. A smaller blue speech bubble contains a screenshot of a simulation tool showing a chemical reaction: $2 \text{ H}^+ + 2 \text{ e}^- \rightarrow \text{H}_2$. At the bottom left is a cartoon character of a person using a mouse, and at the bottom right is a cartoon character of a scientist holding a flask.

UNIVERSITIES OF MITTWEIDA & VALENCIA

Copyright © 2014 CyanoFactory Workshop

PUBLISHED BY UNIVERSITIES OF MITTWEIDA & VALENCIA

CYANOFACtORY.EU

Version: December 6, 2014

Contents

The Knowledgebase 7

CyanoMaps 15

CyanoDesign 19

Bibliography 21

Introduction

THIS SMALL GUIDE shall introduce you to basic features of the CyanoFactory Knowledgebase (CyanoFactory KB). The CyanoFactory KB is currently being developed within the EU-funded CyanoFactory research consortium, mainly by the group from Mittweida/Germany¹. It is based on the WholeCell Knowledgebase, which was developed by Jonathan Karr from Stanford University².

CyanoFactory KB is a knowledge base that is embracing all data produced by and relevant to the CyanoFactory project partners. It is designed to enable comprehensive simulations of entire cells and organisms with the incorporated CyanoDesign tool³. CyanoFactory KB is currently centered around *Synechocystis* sp. PCC 6803, a gram-negative cyanobacterium that is capable of photosynthesis and will provide comprehensive, quantitative descriptions of individual mutants including:

- Cellular chemical composition
- Growth medium composition
- Gene locations, lengths, and directions
- Transcription unit organization and transcriptional regulation
- Macromolecule composition
- Reaction stoichiometry, kinetics, and catalysis
- Extensive links and cross-links to all references used to construct each database
- Experimental conditions

Currently, however, there are only few data hosted. The most important data is currently the result of re-sequencing of *Synechocystis* sp. PCC 6803 substrain Uppsala (Figure 1).

¹ Developed mainly by Gabriel Kind from Röbbe Wünschiers' group in Mittweida/Germany.

² Karr et al., 2013

³ Developed mainly by the group from Javier F. Urchueguía in Valencia/Spain.



Figure 1: SNPs from *Synechocystis* sp. PCC 6803 substrain Uppsala are shown as triangles.

The Knowledgebase

BEFORE YOU START you should know: The CyanoFactory KB requires an account for login. All project partners have their personal account. If you do not have an account yet, you will get guest access during this workshop. These accounts will be deleted later. Please use the CyanoFactory KB with the Firefox internet browser (Figure 2).

The entry page at cyanofactory.hs-mittweida.de/warehouse for the CyanoFactory KB looks as shown in Figure 3. The important part to look at is the menu (Figure 4). The following information and tools can be found:



A screenshot of the CyanoFactory KB entry page, focusing on the 'About CyanoFactory' and 'Getting started' sections. The 'About CyanoFactory' section includes a detailed description of the project's goals and the specific bacterial strain used. The 'Getting started' section provides general guidance for users.



- Home → all organisms available
- CyanoMaps → tools to map data onto metabolic pathway maps
- CyanoDesign → tool to model metabolism
- Import → opens sub-menu for data import
- Members → lists all registered users



Figure 2: The CyanoFactory KB is optimized for Firefox.

Figure 3: The entry page.



Figure 4: The menu.

- Help → all organisms available
- *wuenschi* → user specific actions and basket

In the following sections we will provide deeper insight into these functions. For CyanoMaps and CyanoDesign we devote single chapters. Besides the static menu at the top of each page, the action menus at the bottom right are important (Figure 5). Depending on the content of the current page, different **Actions** and **Export** functions are available. All pages can be exported to PDF-format or printed. Likewise, all pages can be shared via Email, Facebook, Twitter or Google+. Ultimately, a link is shared that builds the same page view for the receiver.

wuenschi – User Account Settings

Before you start working, check if your affiliations are correctly set in the user menu, which in my case is *wuenschi* (Figure 6).

Your affiliation are shown when clicking **My Profile**. The menu entry **My Baskets** lets you setup one or more baskets. These are used to collect information from the knowledgebase. Click **My Baskets** now and create three baskets named **genes**, **proteins** and **stuff** (Figure 7). You are now prepared to work with the knowledgebase and collect data with the help of the baskets.

My Baskets

| WID | Name |
|----------|---------|
| stuff | 0 items |
| proteins | 0 items |
| genes | 0 items |

Create new basket

Home – The Knowledgebase

The menu entry **Home** is the entry point to the knowledgebase. All other menu points link to either analysis or maintenance tools. Upon hovering over **Home** you get a list of available organisms, currently two. After clicking *Synechocystis* sp. PCC 6803, you get all entries stored for this organism (Figure 8). The main information available is the sequence of its chromosome and four plasmids with annotations (Figure 9).

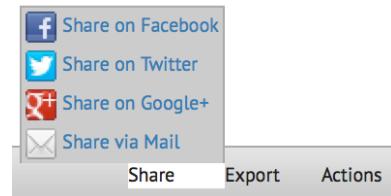


Figure 5: The action menu at the bottom right of each page.



Figure 6: User account settings.

Figure 7: All baskets for user *wuenschi*.

Welcome to the *Synechocystis* sp. PCC 6803 database!

Synechocystis sp. PCC 6803 plasmid pSYSG, complete sequence. PROVISIONAL REFSEQ: This record has not yet been subject to final NCBI review. The reference sequence was derived from AP006585. COMPLETENESS: full length.

Genetic code

Bacteria, archaea and plant plastids (11)

Content

| Content | Value | Units |
|---------------------|------------|-------|
| Compartments | 2 | |
| Genome | 5 | |
| Chromosomes | 1 | |
| Length | 3573470 nt | |
| GC-content | 47.7 % | |
| Plasmids | 4 | |
| Length | 373549 nt | |
| GC-content | 44.0 % | |
| Transcription units | 2586 | |
| Monocistrons | 1928 | |
| Polycistrons | 658 | |
| Genes | 3624 | |
| mRNA | 3575 | |
| rRNA | 6 | |
| sRNA | 0 | |
| tRNA | 43 | |
| Metabolites | 989 | |
| Amino acids | 0 | |
| Antibiotic | 0 | |
| Gases | 0 | |
| Ions | 0 | |
| Lipids | 0 | |
| Vitamins | 0 | |

| Content | Value | Units |
|----------------------------|-------|-------|
| Proteins | 3575 | |
| Monomers | 3575 | |
| DNA-binding | 0 | |
| Integral membrane | 0 | |
| Lipoprotein | 0 | |
| Secreted | 0 | |
| Terminal organelle | 0 | |
| Complexes | 0 | |
| DNA-binding | 0 | |
| Reactions | 1041 | |
| DNA damage | 0 | |
| DNA repair | 0 | |
| Metabolic | 0 | |
| Protein decay | 0 | |
| Protein modification | 0 | |
| Replication Initiation | 0 | |
| RNA decay | 0 | |
| RNA modification | 0 | |
| RNA processing | 0 | |
| Transcription | 0 | |
| Translation | 0 | |
| tRNA aminoacylation | 0 | |
| Other | 1041 | |
| Transcriptional regulation | | |
| Interactions | 0 | |
| Transcriptional regulators | 0 | |
| Regulated promoters | 0 | |
| Pathways | 84 | |
| Stimuli | 0 | |
| Quantitative parameters | 0 | |
| Cell composition | 0 | |
| Media composition | 0 | |
| Reaction K _{eq} | 0 | |
| Reaction K _m | 0 | |
| Reaction V _{max} | 0 | |
| RNA expression | 0 | |
| RNA half-lives | 0 | |
| Stimulus values | 0 | |
| Transcr. reg. activity | 0 | |
| Transcr. reg. affinity | 0 | |
| Other | 0 | |
| Processes | 0 | |
| States | 0 | |
| Mass Spectrometry Data | 0 | |

Figure 8: The home page for *Synechocystis* sp. PCC 6803.

| Content | Value | Units |
|---------------------|------------|-------|
| Compartments | 2 | |
| Genome | 5 | |
| Chromosomes | 1 | |
| Length | 3573470 nt | |
| GC-content | 47.7 % | |
| Plasmids | 4 | |
| Length | 373549 nt | |
| GC-content | 44.0 % | |
| Transcription units | 2586 | |
| Monocistrons | 1928 | |
| Polycistrons | 658 | |
| Genes | 3624 | |
| mRNA | 3575 | |
| rRNA | 6 | |
| sRNA | 0 | |
| tRNA | 43 | |

Figure 9: All content available for *Synechocystis* sp. PCC 6803.

We will now concentrate on the sequence information stored in the knowledgebase. Click on **Plasmids** in the left content table and then select plasmid **pSYSG**. You will now see all information available for this plasmid, divided into several sections (Figure 10).

Clicking at one particular gene name will show sequence details. Let's click at gene *ssr8047*, the first one listed in the feature list. The new screen shows neighbouring genes and a translation of the open reading frame (Figure 11). If present, SNPs in the *Synechocystis* sp. PCC 6803 sub-strain Uppsala will be presented as shown in Figure 1.

To get more information about the gene product, click at the **Protein product** in the **Name** section. This will open a new screen with all available metadata about the protein, as shown in Figure 12.

Among these data are physico-chemical parameters, such as:

- **Instability Index** → estimate of the stability of your protein in a test tube based on dipeptide distributions⁴; value above 40 predicts that the protein may be unstable
- **Aliphatic Index** → measure of the relative volume occupied by aliphatic side chains (alanine, valine, isoleucine, and leucine); a high value indicates thermostability of globular proteins⁵
- **GRAVY at 25C, pH 7.0** → grand average hydrophobicity⁶; extremes would be polylysine at -3.9 and polyisolsine at 4.5; positive indexed proteins are likely non-cytoplasmic
- **Half Life** → prediction of the time it takes for half of the amount

⁴ Guruprasad et al., 1990

⁵ Ikai, 1980

⁶ Kyte and Doolittle, 1982

pSYSG – Plasmid pSYSG

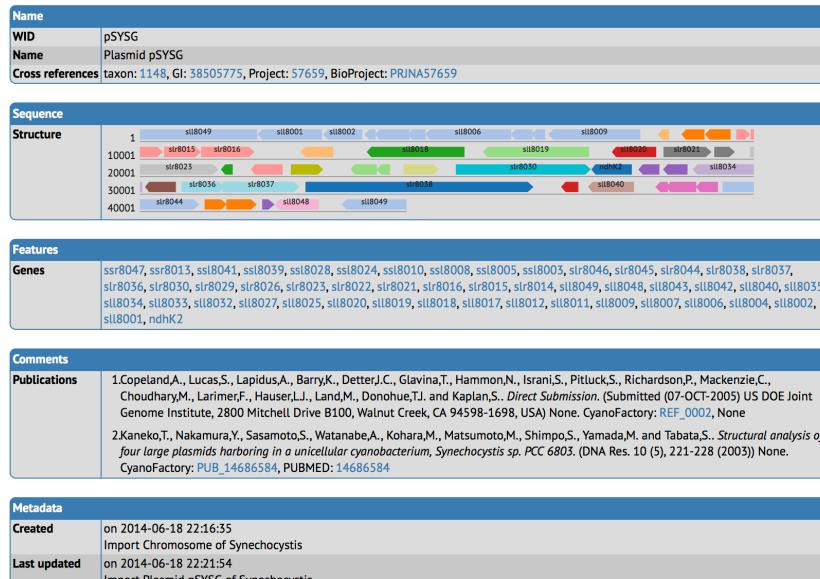


Figure 10: Overview of plasmid pSYSG.

ssr8047

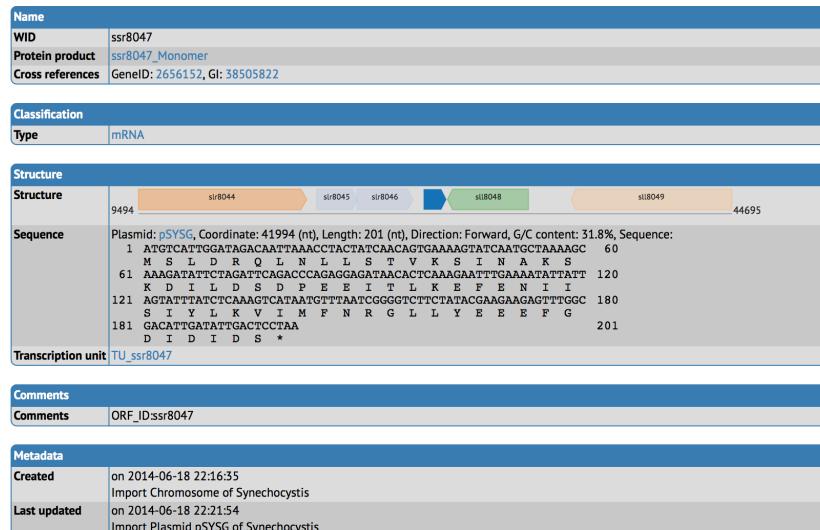


Figure 11: Sequence details for gene ssr8047.

ssr8047_Monomer – hypothetical protein

| | |
|---|--|
| Name | |
| WID | ssr8047_Monomer |
| Name | hypothetical protein |
| Cross references | RefSeq: NP_942441.1 |
| Genetics | |
| Gene | ssr8047 |
| Structure | |
| Sequence | 1 MSLDRQLNLLSTVKSINAKSKDILDSDPEEITLKEFENIIISIYLKVIMFNRLGLYEEFG 60 61 DIDIIDS* |
| Instability index | 46.63 |
| Is stable | False |
| Aliphatic index | 114.93 |
| GRAVY (25C, pH 7.0) | -0.148 |
| Half life (OD (600 nm) = 0.3, M9 media, 36C; min) | 1200 |
| Interactions | |
| Protein/Metabolite interactions Show interactions | |
| Metadata | |
| Created | on 2014-06-18 22:16:35 Import Chromosome of Synechocystis |
| Last updated | on 2014-06-18 22:21:54 Import Plasmid pSYSG of Synechocystis |

of protein in a cell to disappear after its synthesis⁷

Now let us put this entry to our protein basket created previously. Use the **Actions** menu at the bottom right (Figure 13). Ticking the desired basket is enough. There is no button to confirm your choice. Just click into the browser again to close the window.

You can use the **Export** menu at the bottom right to export the data in different formats. Since we are not finished with the importers, yet, we did not spend a lot of effort into exporters. Fasta is working well, while the others are still buggy.

You might already have observed the section **Interactions**. Click **Show interactions** to get an overview about the metabolic environment of the current protein.

Interaction Viewer

Clicking **Show interactions** opens the protein and metabolite interaction viewer for the current protein (Figure 14). The data has been imported from the STRING database⁸ (Search Tool for the Retrieval of Interacting Genes/Proteins) for protein-protein interactions and from the STITCH database⁹ (Search Tool for Interactions of Chemicals) for metabolite-protein interaction.

The interaction graph can be moved and zoomed with the mouse. The graph is accompanied by a table at the bottom of the page. The interactions are displayed as undirected graphs. Only the most significant interaction partners are displayed. The significance is calculated based on the STRING and STITCH interaction scores, which is based on,

Figure 12: Protein details for the product of genessr8047.

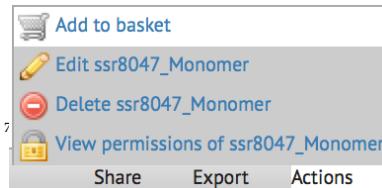


Figure 13: Adding items to your personal basket.

⁸ Franceschini et al., 2013

⁹ Kuhn et al., 2014

ssr8047 - hypothetical protein

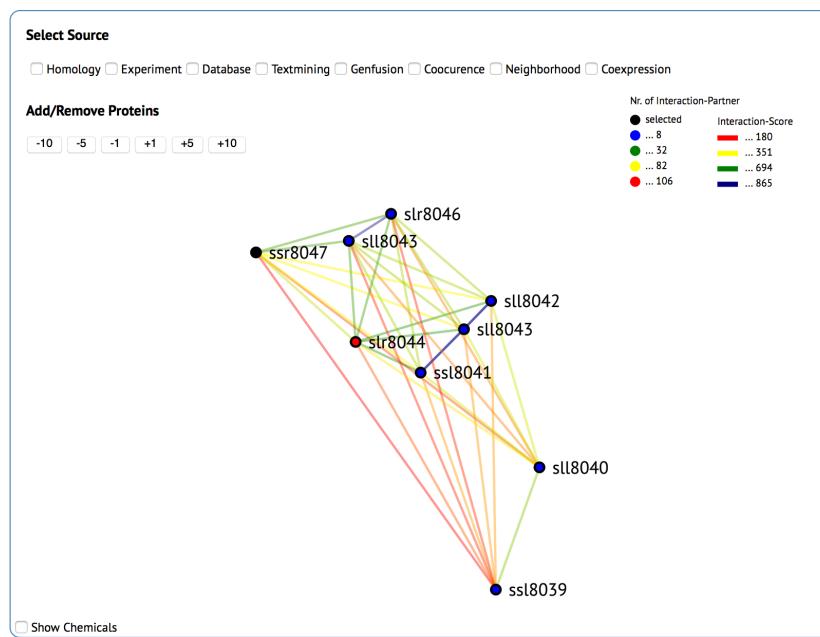


Figure 14: Interaction details for the product of gene *ssr8047*. Short edges between nodes represent higher scores.

among others, homology, coexpression and text mining.

You can add or remove more interacting protein in step of one, five or ten proteins. Addition of proteins to the graph is based on their interaction score. By clicking **Show Chemicals**, metabolites are added to the graph. When you click on a particular protein or metabolite in the graph, it will move into focus. No new network is created. Clicking into empty space will recover the full graph.

| <input checked="" type="checkbox"/> Proteins <input checked="" type="checkbox"/> Chemicals | ID | Name | Annotation | Homology | Experiment | Database | Textmining | Genfusion | Cooccurrence | Neighborhood | Coexpression | Score |
|--|-----------------|----------------------|------------|----------|------------|----------|------------|-----------|--------------|--------------|--------------|-------|
| P18951 | sll8043 | transposase | | | | | | . | | . | | 865 |
| P18950 | str8044 | hypothetical protein | | | | | | | | . | | 606 |
| P19305 | ssr8047 | hypothetical protein | | | | | | | | . | | 599 |
| P17244 | sll8042 | transposase | . | | | | | | . | . | | 479 |
| P17245 | sll8043 | transposase | | | | | . | | . | . | | 468 |
| P19144 | ssl8041 | transposase | | | | | | | | . | | 468 |
| P17243 | sll8040 | hypothetical protein | | | | | | | | . | | 264 |
| CID100001030 | 1,2-propanediol | | | | | | . | | | | | 230 |
| CID100134919 | propanediol | | | | | | . | | | | | 230 |
| CID103348130 | R up | | | | | | . | | | | | 226 |
| CID100002061 | tyrphostin 25 | | | | | | . | | | | | 209 |
| P19143 | ssl8039 | hypothetical protein | | | | | | | | . | | 204 |
| CID105496819 | 1ppm | | | | | | . | | | | | 194 |
| CID124820124 | M5 S | | | | | | . | | | | | 194 |

Load more

Colors of the edges and nodes correlate to the interaction score and number of interaction partners, respectively (Figure 16).

After selecting a protein or metabolite from the network, a small table

Figure 15: Interaction details for the product of gene *ssr8047*. Here, metabolites have been added.

Nr. of Interaction-Partner



Figure 16: Legend to the interaction graph.

with all connected proteins or metabolites is shown at the right. Here you can select an item and thereby create a new network around this item.

Search Field

Finally, a search field is available at the top left of the knowledgebase (Figure 17).

Here you can enter query terms that will be searched for in the gene description.

Import – Filling the Knowledgebase

Well, this will be a rather short section.

Importing data to the CyanoFactory KB is of course a crucial feature. But it is also the most difficult one in terms of data integrity. For the time being, we restrict this function to selected users.

Members – Who has Access?

This page lists all assigned user accounts to the database. After clicking a name, you get the member's institution and Email address, too.

Help – Hmm, Help?

The Help menu does not really help yet. It presently holds licence information. But this will change in the future. If you have questions, do not hesitate to ask Gabriel (gkind@hs-mittweida.de).

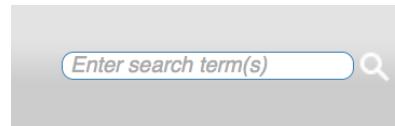


Figure 17: The search field at the top right of the knowledgebase.



Figure 18: Data import is currently restricted to selected users.

CyanoMaps

CYANOMAPS HARBOURS TOOLS to display metabolic and regulatory network data (Figure 19). It can be regarded as a collection of generic stand-alone tools under the umbrella of the CyanoFactory KB. Both, KEGG Pathways and SBGN of *Synechocystis* are linked to the knowledgebase. Biochemical Pathways is currently disconnected from the rest of the knowledgebase.

The intention of developing CyanoMaps was the generation of a platform for highlighting selected enzymes or metabolites in metabolic maps. This has currently been implemented for KEGG maps¹⁰ (Kyoto Encyclopedia of Genes and Genomes) and the Boehringer Map, originally developed by Gerhard Michal from the Boehringer Company¹¹.

The SBGN maps¹² (System Biology Graphical Notation) was developed as part of CyanoFactory as an attempt to bring all regulatory and metabolic information about hydrogen metabolism of *Synechocystis* sp. PCC 6803 together. The SBGN visualization itself is an attempt to visualize biological processes in an human and computer readable way. It has, however, been decided by the CyanoFactory consortium to discontinue its development.



Figure 19: The CyanoMaps menu.

¹⁰ Kanehisa et al., 2014

¹¹ Michal and Schomburg, 2012

¹² Le Novere et al., 2009

Biochemical Pathways – Boehringer Maps

When you open Biochemical Pathways, the page shown in Figure 20 will be displayed – though the map has already been moved. The main part of the page is occupied by the actual viewer. You can use the mouse to pan and zoom the map. At the right-hand side you'll find an input form. When you load the Biochemical Pathways page for the first time, the form is pre-filled with five EC-numbers and one metabolite. The matches for these query terms are highlighted in blue within the map²⁰. You can, however, define other colors by appending a color name, separated by the hash character.

In the top sections of the page the overall number of matches and mismatches for your query terms is shown. The dropdown menu in the **Navigator** section lets you jump directly to an enzyme or metabolite

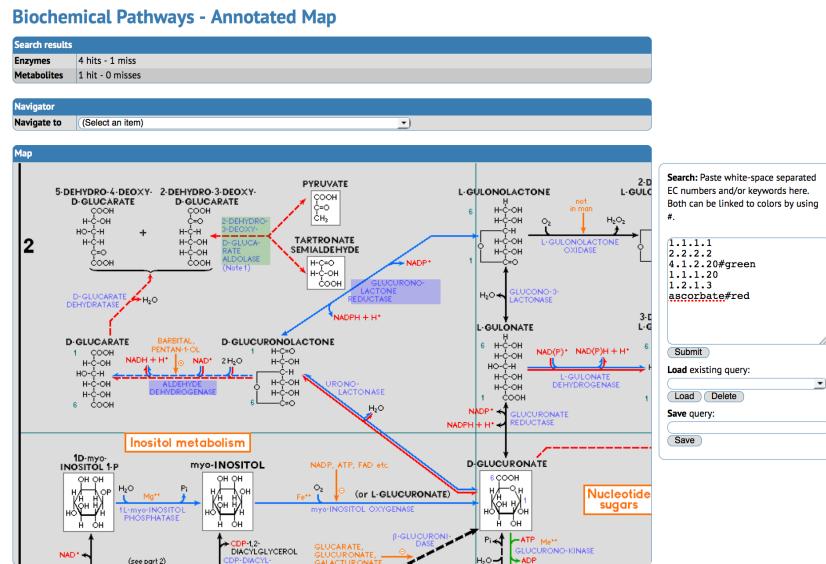


Figure 20: The Boehringer pathway map.

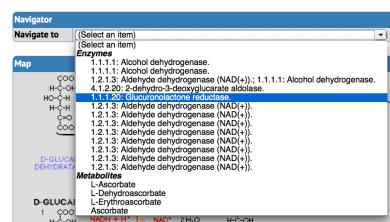


Figure 21: The navigator menu.

(Figure 21). Duplicate entries indicate that the item occurs at different positions in the Boehringer Map. Highlighted enzymes are hyperlinked to the BRENDA database. It is currently not possible to save the complete map.

The query form can also be used to save your queries. Note, however, that there is a size limit for the query to be stored. We hope to resolve this issue in later releases.

KEGG Pathways – KEGG Maps

KEGG Pathways are a collection of manually drawn pathway maps representing molecular interaction and reaction networks. Although enzyme and metabolite mapping and highlighting can be performed at the KEGG server, we integrated KEGG maps within the CyanoFactory KB and **linked** them to our data. When you open KEGG Pathways you will first see a list as shown in Figure 22. At the top section all three overview maps are listed, while the bottom section list all detailed maps. The overview maps are:

- map01100 → metabolic pathways
 - map01110 → biosynthesis of secondary metabolites
 - map01120 → microbial metabolism in diverse environments

Within these maps, all enzymes, i.e. edges, that are found in the CyanoFactory KB are colored in green (Figure 23). The color of the metabolites has no meaning in this overview maps. All enzymes and metabolites are hyperlinked to the KEGG database.

Besides this browsing through maps, you can use the query panel to search for enzymes and metabolites within KEGG maps.

KEGG Maps

| WID | Name |
|----------|--|
| map01100 | |
| map01110 | |
| map01120 | |
| | |
| WID | Name |
| map00010 | Glycolysis / Gluconeogenesis |
| map00020 | Citrate cycle (TCA cycle) |
| map00030 | Pentose phosphate pathway |
| map00040 | Pentose and glucuronate interconversions |
| map00051 | Fructose and mannose metabolism |
| map00052 | Galactose metabolism |
| map00053 | Ascorbate and aldarate metabolism |
| map00061 | Fatty acid biosynthesis |
| map00062 | Fatty acid elongation |
| map00071 | Fatty acid metabolism |
| map00072 | Synthesis and degradation of ketone bodies |

Figure 22: Truncated list of all KEGG pathways.

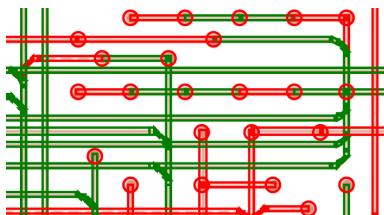


Figure 23: Green colored edges represent enzymes present in *Synechocystis* sp. PCC 6803.

Figure 24 shows the result of a query for four enzymes and two metabolites. After you execute the query, all pathway maps that contain matches to your query terms are shown. From this list, map00680 (methane metabolism) has been chosen for Figure 24. If no color attributes are joined to the query terms, the default color scheme will be used. However, you can override this setting by attaching colors as shown in the legend of Figure 25.

KEGG - map00680

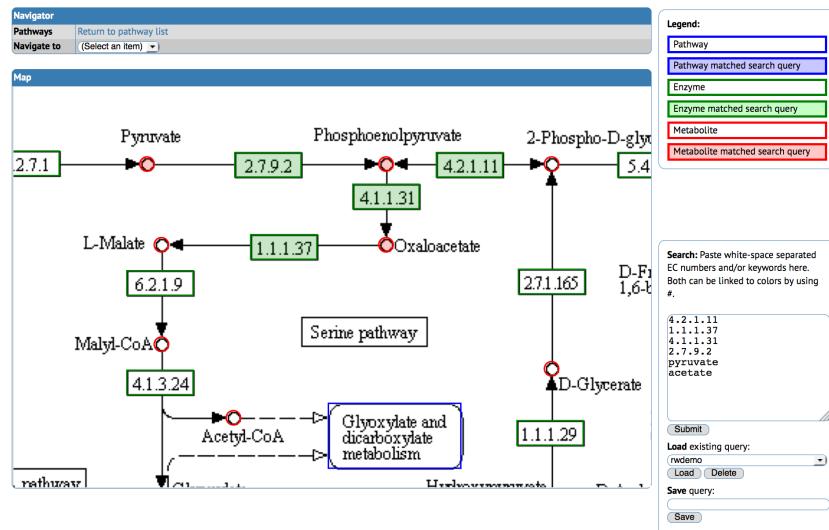


Figure 24: The KEGG pathway map query tool.

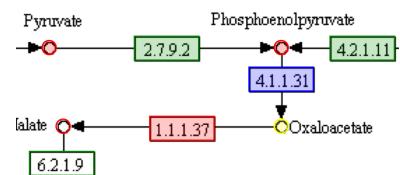


Figure 25: When colors are set in the query, the default color scheme as stated in the legend will be overridden. The query from Figure 24 had been changed to:

```
4.2.1.11
1.1.1.37#red
4.1.1.31#blue
2.7.9.2
pyruvate
acetate#yellow
```

The query form can also be used to save your queries. Note, however, that there is a size limit for the query to be stored. We hope to resolve this issue in later releases.

SBGN of Synechocystis – A new Approach

A final option to look at pathway information with the CyanoFactory KB is the application of the systems biology graphical notation (SBGN). The respective application is shown in Figure 26.

SBGN of Synechocystis

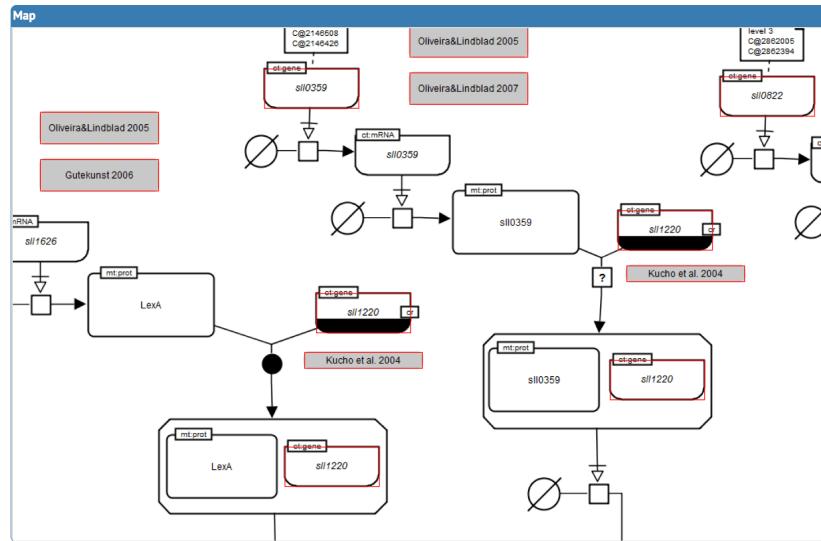


Figure 26: The KEGG pathway map query tool.

Within this map, publications are hyperlinked to PubMed, while genes are linked to the CyanoFactory KB. I will not go into details of the graphical notation here but refer to Novere *et al.*¹³.

¹³ Le Novere et al., 2009

CyanoDesign

PERFORMING GENOME-SCALE METABOLIC MODELLING is the task, CyanoDesign is being developed for. By default, two models are pre-installed in the CyanoFactory KB: iSyn811¹⁴ for *Synechocystis* sp. PCC 6803 and a toy model (Figure 27).

¹⁴ Montagud et al., 2011

CyanoDesign – Toy Model

- reac1: 1 A_ext -> 1 A [0, 10]
- reac2: 1 A -> 1 B unconstrained
- reac3: 1 A -> 1 C unconstrained
- reac4: 1 B + 1 C -> 1 D unconstrained
- reac5: 1 D -> 1 D_ext unconstrained
- reac6: 1 E_ext -> 1 E [0, 3]
- reac7: 1 E -> 1 D unconstrained

7 of 7 reactions displayed.

[Manage filters](#)

Simulation settings

- Display reactions with highest flux
- Display selected reactions

Main objective:

[Save](#)

Select visible reactions

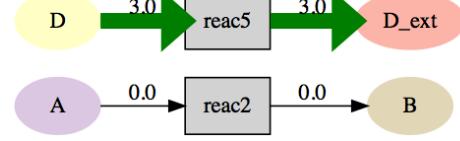
reac2

reac5

Start simulation

[Simulate](#)

Results



After hitting Simulate a flux balance analysis is triggered and the result shown for all selected reactions at the bottom of the page. This

Figure 27: CyanoDesign facilitates simulation of metabolism. Here a *virtual mutant* is generated by *knocking-out* reaction 5.

simulation is based on the PyNetMet algorithm¹⁵.

A details guide on how to use CyanoDesign is presented in an extra tutorial (Figure 28),

¹⁵ Gamermann et al., 2014

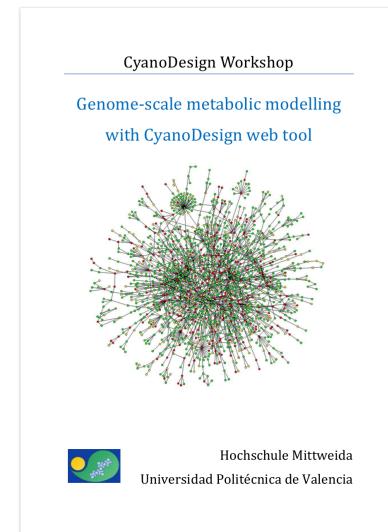


Figure 28: This tutorial explains usage of CyanoDesign.

Bibliography

A Bachmair, D Finley, and A Varshavsky. In vivo half-life of a protein is a function of its amino-terminal residue. *Science (New York, N.Y.)*, 234(4773):179–186, October 1986.

A. Franceschini, D. Szklarczyk, S. Frankild, M. Kuhn, M. Simonovic, A. Roth, J. Lin, P. Minguez, P. Bork, C. von Mering, and L. J. Jensen. STRING v9.1: protein-protein interaction networks, with increased coverage and integration. *Nucleic Acids Res.*, 41(Database issue):D808–815, Jan 2013.

D. Gamermann, A. Montagud, R. A. Jaime Infante, J. Triana, P. F. de Córdoba, and Urchueguía. PyNetMet: Python tools for efficient work with networks and metabolic models. *Computational and Mathematical Biology*, 3(5):1–11, 2014.

K Guruprasad, B V Reddy, and M W Pandit. Correlation between stability of a protein and its dipeptide composition: a novel approach for predicting in vivo stability of a protein from its primary sequence. *Protein engineering*, 4(2):155–161, December 1990.

A Ikai. Thermostability and aliphatic index of globular proteins. *Journal of biochemistry*, 88(6):1895–1898, December 1980.

M. Kanehisa, S. Goto, Y. Sato, M. Kawashima, M. Furumichi, and M. Tanabe. Data, information, knowledge and principle: back to metabolism in KEGG. *Nucleic Acids Res.*, 42(Database issue):199–205, Jan 2014.

J. R. Karr, J. C. Sanghvi, D. N. Macklin, A. Arora, and M. W. Covert. WholeCellKB: model organism databases for comprehensive whole-cell models. *Nucleic Acids Res.*, 41(Database issue):D787–792, Jan 2013.

M. Kuhn, D. Szklarczyk, S. Pletscher-Frankild, T. H. Blicher, C. von Mering, L. J. Jensen, and P. Bork. STITCH 4: integration of protein-chemical interactions with user data. *Nucleic Acids Res.*, 42(Database issue):D401–407, Jan 2014.

J Kyte and R F Doolittle. A simple method for displaying the hydrophobic character of a protein. *Journal of molecular biology*, 157(1): 105–132, May 1982.

N. Le Novere, M. Hucka, H. Mi, S. Moodie, F. Schreiber, A. Sorokin, E. Demir, K. Wegner, M. I. Aladjem, S. M. Wimalaratne, F. T. Bergman, R. Gauges, P. Ghazal, H. Kawaji, L. Li, Y. Matsuoka, A. Villeger, S. E. Boyd, L. Calzone, M. Courtot, U. Dogrusoz, T. C. Freeman, A. Funahashi, S. Ghosh, A. Jouraku, S. Kim, F. Kolpakov, A. Luna, S. Sahle, E. Schmidt, S. Watterson, G. Wu, I. Goryanin, D. B. Kell, C. Sander, H. Sauro, J. L. Snoep, K. Kohn, and H. Kitano. The Systems Biology Graphical Notation. *Nat. Biotechnol.*, 27(8):735–741, Aug 2009.

Gerhard Michal and Dietmar Schomburg, editors. *Biochemical Pathways*. An Atlas of Biochemistry and Molecular Biology. Wiley, September 2012.

Arnaud Montagud, Aleksei Zelezniak, Emilio Navarro, Pedro Fernández de Córdoba, Javier F Urchueguía, and Kiran Raosaheb Patil. Flux coupling and transcriptional regulation within the metabolic network of the photosynthetic bacterium Synechocystis sp. PCC6803. *Biotechnology journal*, 6(3):1–13, January 2011.