CyanoFactory Knowledge Base & Synthetic Biology
A Plea for Human Curated Bio-databases

Gabriel Kind, Eric Zuchantke and Röbbe Wünschiers
University of Applied Sciences Mittweida, Technikumplatz 17, 09648 Mittweida, Germany

Keywords: Synthetic Biology, Big Data, Omics, Warehouses, Databases, Knowledge-base, Wisdom of the Crowd.

Abstract: Nowadays, life science research is dominated by two conditions: interdisciplinarity and high-throughput. The former leads to highly diverse datasets from a data type point of view while high-throughput yields massive amounts of data. Both aspects are reflected by the byte-growth of public bio-databases and the sheer number of specialised databases or databases of databases (i.e. data warehouses). We provide an insight to the development of a biodata knowledge base (dubbed CyanoFactory KB) targeted to bio-engineers in the field of synthetic biology and exemplify the need for data type specific data curation and cross-linking. CyanoFactory KB is unique in incorporating experimental data from a broad range of scientific methods that are based on one strain of *Synechocystis* sp. PCC 6803. The knowledge base can be accessed upon request via cyanofactory.hs-mittweida.de.

1 INTRODUCTION

Nowadays, life science research is dominated by two conditions: interdisciplinarity and high-throughput. The former leads to highly diverse datasets from a content point of view while high-throughput yields massive amounts of data. Both aspects are reflected by the byte-growth of public bio-databases and the diversity of specialised databases (see, e.g. the database issues of the NAR journal). However, quite often more data leads to less understanding. Driven by the methodology of systems biology, a holistic view of genetic and metabolic regulatory processes is demanded. One important goal is the application of these systemic data for *in silico* modelling of biological processes or, ultimately, biological systems, i.e. cells, tissues, organisms. One step towards this goal was the successful prediction of the phenotype from the genotype in *Mycoplasma genitalium* (Karr et al., 2012). The basis to solve this challenge was a database named WholeCell Knowledge Base (WholeCell KB) (Karr et al., 2013). It contains experimental results from over 900 publications and includes more than 1,900 experimentally observed parameters. Importantly, all data has been validated and curated by scientists.

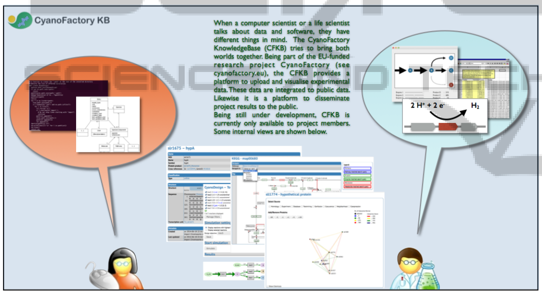
Another important manually curated database is Brenda, which contains almost 1.5 million manually curated enzyme parameters (as of July 2014, brenda-enzymes.org). In contrast, GenBank contains 174,108,750 individual and 189,080,419 whole genome shotgun sequences (as of August 2014, ncbi.nlm.nih.gov) that are partially manually uploaded but not curated. In the field of cyanobacteria research, CyanoBase is a well-known manually curated genome database, including over 5200 references (Fujisawa et al., 2014).

With the rising amount of biological data and the increasing capabilities of computer hardware, many attempts have been undertaken to automatically harvest, store, cross-link and provide biological data in databases and databases of databases (i.e. data warehouses) (Triplet and Butler, 2011). We argue that automatically generated data collections are of limited value (the common garbage-in garbage-out problem), especially in the context of large scale biological engineering as envisioned by the field of synthetic biology. In this field, computer modelling of biological processes builds the base to targeted (instead of trial-and-error) genetic engineering.

With this paper we provide an insight in the development of a biodata-warehouse (CyanoFactory KB) targeted to bio-engineers in the field of synthetic biology. We are in the extraordinary situation to work in a research consortium that consists of partners from different scientific fields (interdisciplinary) and seven different countries (multiregional) with the unifying goal to tinker the cyanobacterium *Synechocystis* sp.

Kind G., Zuchantke E. and Wünschiers R..
CyanoFactory Knowledge Base & Synthetic Biology - A Plea for Human Curated Bio-databases.
DOI: 10.5220/0005285802370242
In *Proceedings of the International Conference on Bioinformatics Models, Methods and Algorithms* (BIOINFORMATICS-2015), pages 237-242
ISBN: 978-989-758-070-3

PCC 6803 to produce photohydrogen. We are particularly working at the interface between experimental and computational biology, which implies different understanding of data (Figure 1). CyanoFactory KB, which is a massive expansion of the WholeCell KB, shall provide a central data hub for members of the consortium and for disseminating project data to the research community. Thus, it shall provide ways for an improved collaboration between all partners within the CyanoFactory consortium. All partners are experts in different fields from microbial biotechnology or metabolic modelling up to synthetic biology. The goal of the knowledge base is to bridge the gap between bio-engineers and bioinformaticians by providing user friendly functionalities for working with experimental data and for visualising and contextualising it in different ways. Besides experimental data, further data is obtained from other biological databases.



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.

Figure 1: Different ways of thinking about data.

CyanoFactory KB is continuously improving and adjusted to new requirements of the research partners. The warehouse is still under heavy development and gains new features every month. It is not available to the general public yet, this step is planned when the codebase is more matured and more experimental data of the partners has been integrated.

2 MATERIALS AND METHODS

CyanoFactory KB is a massive expansion of WholeCell KB (Karr et al., 2013) and was further adopted to the needs of CyanoFactory. WholeCell KB turned out to be the best suited and matured bio-warehouse system after intensive scouting and testing of various known open-source systems (Table 1).

Table 1: Intensively tested open-source biodata-warehouses.

Warehouse	Citation
BioDWH	(Töpel et al., 2008)
BioMart Central Portal	(Guberman et al., 2011)
BioWarehouse	(Lee et al., 2006)
BioXRT	(Zhang et al., 2004)
BN++	(Küntzer et al., 2007)
CoryneRegNet	(Baumbach, 2007)
DAWIS-M.D.	(Hippe et al., 2010)
InterMine	(Lyne et al., 2007)
ONDEX	(Taubert et al., 2014)
Open Genome Resource	(Klein et al., 2009)
PROFESS	(Triplet et al., 2010)
PiPa	(Arzt et al., 2011)
RAMEDIS	(Töpel et al., 2010)
WholeCell KB	(Karr et al., 2013)

Due to different responsibilities, different technologies and data types are provided and demanded by individual partners, e.g. mass spectrometry data, DNA-microarray- or RNASeq-based transcription analyses, metabolic analyses, mathematical modells, sequence based data from genetic engineering or climate and experimental data from outdoor photo-bioreactor experiments. All experiments are performed using the model organism *Synechocystis* sp. PCC 6803 as provided from the University of Uppsala/Sweden (Uppsala subtype). This subtype has been resequenced, compared to the original sequence on GenBank and analysed for genetic variations. Additional data was obtained from other biological databases: Organism related information from GenBank, pathway data from KEGG (Kanehisa et al., 2014) and Boehringer Biochemical Pathways Maps (Michal and Schomburg, 2012) and protein-protein interactions from STRING (Franceschini et al., 2013) and STITCH (Kuhn et al., 2014). Furthermore CyanoDesign provides a unique interface for metabolic modelling and analyses of the efficiency of enzymatic reactions via flux balance analysis.

3 RESULTS

CyanoFactory KB is a productive knowledge base, which handles all the information from our partners. The advantage of our solution is that, besides holding information, it provides different visualisation techniques and cross-links to other data sources.

Uploading of experimental data is supported in different formats. The warehouse provides import functionality for FASTA, GenBank and System Biology Markup Language (SBML). The import runs as a background job and is automatically merged into the current dataset upon completion. All modifications to the knowledge base are stored as revisions, therefore changes to all items can be retrieved and rolled back if necessary. Exporting is possible in the import formats and furthermore in machine readable XML or JSON formats. The access to individual resources can be restricted using permissions.

Besides the hierarchical view of the warehouse the user can group selected data in “baskets”. A user can create different baskets and group relevant items in them.

The general structure of the organism is visualized by using a chromosome viewer (Figure 2). The viewer is fully interactive and provides filtering functionalities. Additional metadata is displayed beneath the genes. In our case these are SNPs obtained from the Uppsala subtype of *Synechocystis* sp. PCC 6803. When selecting a gene or SNP additional metadata for the corresponding component is displayed.

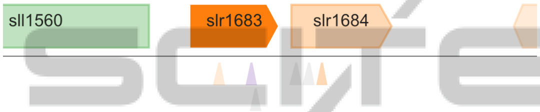


Figure 2: Chromosome viewer showing SNPs (triangles below the genes) from the Uppsala subtype of *Synechocystis* sp. PCC 6803.

Due to possible mutations in *Synechocystis* sp. PCC 6803 over time the strand used in experiments by the Uppsala university was resequenced, analyzed and the sequence modifications uploaded into the knowledge base. The newly sequenced genome was aligned to the reference sequence. Many genes of the Uppsala strain contain SNPs (Table 2). It is currently investigated whether these inflict noticeable changes to the metabolic system of the organism.

Table 2: Mutations in chromosome of *Synechocystis* sp. PCC 6803.

	Chromosome
Number of SNPs	732
SNPs on Genes	511
Genes	6462
Modified Genes	1028

The metabolic processes of *Synechocystis* sp. PCC 6803 and interactions of biochemical components are visualized using the *Process Description Language* of the Systems Biology Graphical Notation (SBGN) (Le Novere et al., 2009). SBGN represents the metabolic model of the organism in a way detailed enough for biochemists and is machine readable, therefore supporting mathematical simulations inside the model.

It should, however, be noted that SBGN proofed to be confusing to the human eye. Thus, traditional visualisations such as Boehringer Pathway Maps (Michal and Schomburg, 2012) or KEGG maps (Kanehisa et al., 2014) are to be preferred.

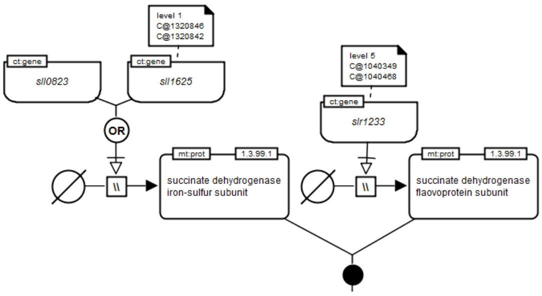


Figure 3: Part of the metabolic model of *Synechocystis* sp. PCC 6803 rendered in Systems Biology Graphical Notation (SBGN).

Pathway Visualisation

The enzymes and metabolites of *Synechocystis* sp. PCC 6803 are displayed on different pathway maps. Biochemical Pathways provides a detailed overview about chemical reactions. A small excerpt with highlighted enzymes contained in *Synechocystis* sp. PCC 6803 is visualised in Figure 4.

All found enzymes and metabolites are highlighted on individual KEGG pathway maps (Figure 5). When an item was detected the image is filled in green for enzymes and red for metabolites. Pathways are cross-linked to each other.

Custom searches, independent of the organism, are supported on all pathway maps. All search queries can be saved for later reuse.

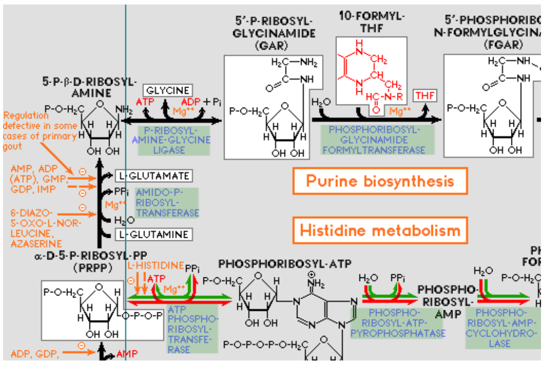

5'-P-RIBOSYL-GLYCINAMIDE (GAR)
10-FORMYL-THF
5'-PHOSPHORIBO N-FORMYLGLYCINA (FGAR)
5-P-β-D-RIBOSYL-AMINE
GLYCINE
ATP
ADP + Pi
THF
Regulation defective in some cases of primary gout
P-RIBOSYL-AMINE-GLYCINE LIGASE
PHOSPHORIBOSYL-GLYCINAMIDE FORMYLTRANSFERASE
AMP, ADP (ATP), GMP, GDP, IMP
L-GLUTAMATE
AMIDO-P-RIBOSYL-TRANSFERASE
6-DIAZO-5-OXO-L-NOR-LEUCINE, AZASERINE
L-GLUTAMINE
Purine biosynthesis
Histidine metabolism
α-D-5-P-RIBOSYL-PP (PRPP)
L-HISTIDINE
PHOSPHORIBOSYL-ATP
PHOSPHO-RIBOSYL-AMP
ATP PHOSPHO-RIBOSYL-TRANSFERASE
PHOSPHO-RIBOSYL-ATP-PYROPHOSPHATASE
PHOSPHO-RIBOSYL-AMP-CYCLOHYDROLASE
ADP, GDP
AMP

Figure 4: Boehringer Pathway Map.

Metabolic Modelling

Metabolic modelling is provided as part of CyanoDesign. Flux balance analysis (FBA) is used for the reconstruction of metabolic networks of organisms (Figure 6). A metabolic network consists of multiple enzymatic reactions with metabolites contained in

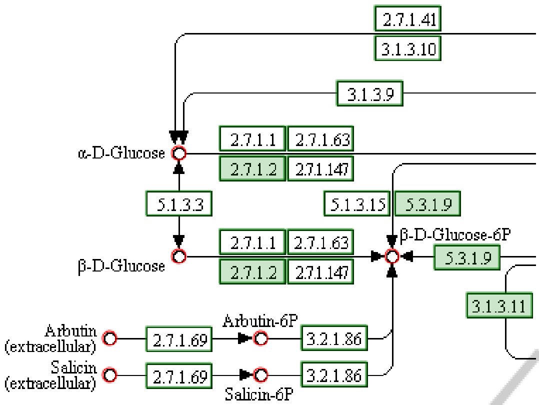


Figure 5: KEGG Pathway Map.

a stoichiometric matrix (positive for production, negative for degradation). This network is solved using a linear solving method. The motivation behind CyanoDesign is allowing the bio-engineer to change the metabolic network *in silico* and to get a prediction how the organism will behave. A modelling approach saves valuable time because it gives hints how mutants of the organism behave, resulting in a high amount of saved work in the lab. FBA is done using the library PyNetMet (Gamermann et al., 2014). For improved quality of the simulation results the addition of more advanced algorithms like “Minimization of Metabolic Adjustment” (MOMA) is planned.

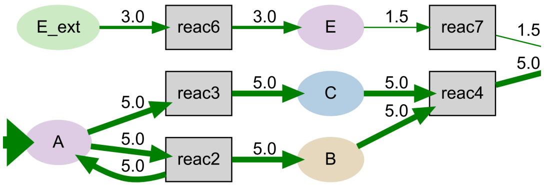


Figure 6: Metabolic model with calculated fluxes. The calculated amount of flux is displayed on top of the arrows and indicated by there thickness.

Interactions

CyanoInteraction provides visualisation of protein-protein-interactions and protein-metabolite-interactions of selected proteins and metabolites from *Synechocystis* sp. PCC 6803 (Figure 7). The dataset used is based on data from STRING (Franceschini et al., 2013) and STITCH (Kuhn et al., 2014). 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. This score is based on, among others, homology, coexpression and text mining. The visualisation is completely interactive and adjustable to the users needs.

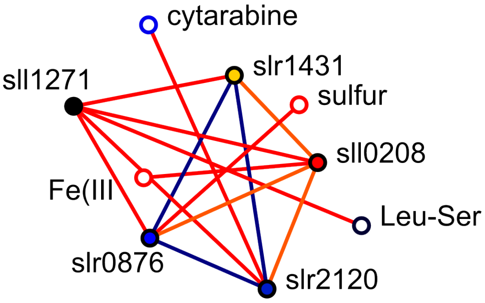


Figure 7: Graph displaying protein and chemical interactions. Shorter edges between nodes mean higher scores.

4 CONCLUSION

With the help of the web-based CyanoFactory knowledge base it is possible to concentrate information from a single organism and its derived mutants. The more information and the more cross-links between these information can be established, the better is the understanding of the selected organism. The CyanoFactory KB is still under development. Thus, later versions of the warehouse will see more functions and experimental data.

From previous interactions between, and from experiences by the partners in the research consortium it became clear that automatically collected data sets are too error prone. Thus, the CyanoFactory KB sets the stage to examine the effort and value of human curated databases and warehouses.

It is important to note the importance to put experimental and computational results about one particular bacterial strain under one umbrella. This demands management decisions within a research consortium or even a research community. The research consortium CyanoFactory (cyanofactory.eu) works with the model organism *Synechocystis* sp. PCC 6803. The Kazusa strain of *Synechocystis* sp. PCC 6803 was the first photosynthetic prokaryote whose genome sequence has been determined in 1996 (Kaneko et al., 1996). Besides the sister strains PCC (Pasteur Culture Collection), ATCC (American Type Culture Collection) and GT (glucose tolerant), the Kazusa strain has been derived from one original California freshwater isolate from 1971, the Berkeley strain (Stanier et al., 1971). Recently, it has been shown that all sister strains can be distinguished by single nucleotide polymorphisms and indels (Ikeuchi and Tabata, 2001; Kanesaki et al., 2012; Trautmann et al., 2012). Furthermore, many sub- or laboratory strains have been derived from all four strains. This leads to experimental and computational results based on different genetic backgrounds. Ultimately, this may lead to non-comparable results. Thus, when integrating data in a knowledge base, detailed information about the experimental setup and the genetic background are necessary. Only then, valid and consistent metabolic models can be derived.

The CyanoFactory KB takes a first step in the direction of strain specific databases. This, however, requires a high investment in man-month of skilled personnel. Another important problem that we are currently addressing is the way of data integration from diverse data sources. While the systems biology markup language (SBML) provides a good foundation for data exchange, strong effort has to be invested in data-pipeline setup. On the long run, research consortia need special funding solely directed to manual data(base) curation. In return, a sustainable and coherent data source for follow-up research can be established.

ACKNOWLEDGEMENT

We want to thank Eric Frenzel for the creation of the SBGN networks. This project has received funding from the European Union's Seventh Programme for research, technological development and demonstration under grant agreement No 308518, CyanoFactory.

REFERENCES

Arzt, S., Starlinger, J., Arnold, O., Kröger, S., Jaeger, S., and Leser, U. (2011). Pipa: Custom integration of protein interactions and pathways. In *Workshop Daten In den Lebenswissenschaften, Berlin, Germany*. Citeseer.

Baumbach, J. (2007). CoryneRegNet 4.0 – A reference database for corynebacterial gene regulatory networks. *BMC Bioinformatics*, 8(1):429.

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

Fujisawa, T., Okamoto, S., Katayama, T., Nakao, M., Yoshimura, H., Kajiya-Kanegae, H., Yamamoto, S., Yano, C., Yanaka, Y., Maita, H., Kaneko, T., Tabata, S., and Nakamura, Y. (2014). CyanoBase and RhizoBase: databases of manually curated annotations for cyanobacterial and rhizobial genomes. *Nucleic Acids Research*, 42(Database issue):D666–70.

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

Guberman, J. M., Ai, J., Arnaiz, O., Baran, J., Blake, A., Baldock, R., Chelala, C., Croft, D., Cros, A., Cutts, R. J., Di Génova, A., Forbes, S., Fujisawa, T., Gadaleta, E., Goodstein, D. M., Gundem, G., Haggarty, B., Haider, S., Hall, M., Harris, T., Haw, R., Hu, S., Hubbard, S., Hsu, J., Iyer, V., Jones, P., Katayama, T., Kinsella, R., Kong, L., Lawson, D., Liang, Y., Lopez-Bigas, N., Luo, J., Lush, M., Mason, J., Moreews, F., Ndegwa, N., Oakley, D., Perez-Llamas, C., Primig, M., Rivkin, E., Rosanoff, S., Shepherd, R., Simon, R., Skarnes, B., Smedley, D., Sperling, L., Spooner, W., Stevenson, P., Stone, K., Teague, J., Wang, J., Wang, J., Whitty, B., Wong, D. T., Wong-Erasmus, M., Yao, L., Youens-Clark, K., Yung, C., Zhang, J., and Kasprzyk, A. (2011). BioMart Central Portal: an open database network for the biological community. *Database*, 2011(0):bar041–bar041.

Hippe, K., Kormeier, B., Töpel, T., and Janowski, S. (2010). DAWIS-MD-A Data Warehouse System for Metabolic Data. *GI Jahrestagung*.

Ikeuchi, M. and Tabata, S. (2001). Synechocystis sp. PCC 6803 - a useful tool in the study of the genetics of cyanobacteria. *Photosynthesis research.*, 70(1):73–83.

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

Kaneko, T., Sato, S., Kotani, H., Tanaka, A., Asamizu, E., Nakamura, Y., Miyajima, N., Hirosawa, M., Sugiura, M., Sasamoto, S., Kimura, T., Hosouchi, T., Matsuno, A., Muraki, A., Nakazaki, N., Naruo, K., Okumura, S., Shimpo, S., Takeuchi, C., Wada, T., Watanabe, A., Yamada, M., Yasuda, M., and Tabata, S. (1996). Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis sp. strain PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions. *DNA Research: An International Journal for Rapid Publication of Reports on Genes and Genomes*, 3(3):109–136.

Kanesaki, Y., Shiwa, Y., Tajima, N., Suzuki, M., Watanabe, S., Sato, N., Ikeuchi, M., and Yoshikawa, H. (2012). Identification of substrain-specific mutations by massively parallel whole-genome resequencing of Synechocystis sp. PCC 6803. *DNA Research: An International Journal for Rapid Publication of Reports on Genes and Genomes*, 19(1):67–79.

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

Karr, J. R., Sanghvi, J. C., Macklin, D. N., Gutschow, M. V., Jacobs, J. M., Bolival Jr., B., Assad-Garcia, N., Glass, J. I., and Covert, M. W. (2012). A Whole-Cell Computational Model Predicts Phenotype from Genotype. *Trends in Genetics*, 150(2):389–401.

Klein, J., Münch, R., Biegler, I., Haddad, I., Retter, I., and Jahn, D. (2009). Strepto-DB, a database for comparative genomics of group A (GAS) and B (GBS) streptococci, implemented with the novel database platform 'Open Genome Resource' (OGeR). *Nucleic Acids Research*, 37(Database issue):D494–8.

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

Küntzer, J., Backes, C., Blum, T., Gerasch, A., Kaufmann, M., Kohlbacher, O., and Lenhof, H.-P. (2007). BNDB - the Biochemical Network Database. *BMC Bioinformatics*, 8(1):367.

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

Lee, T. J., Pouliot, Y., Wagner, V., Gupta, P., Stringer-Calvert, D. W. J., Tenenbaum, J. D., and Karp, P. D. (2006). BioWarehouse: a bioinformatics database warehouse toolkit. *BMC Bioinformatics*, 7(1):170.

Lyne, R., Smith, R., Rutherford, K., Wakeling, M., Varley, A., Guillier, F., Janssens, H., Ji, W., Mclaren, P., North, P., Rana, D., Riley, T., Sullivan, J., Watkins, X., Woodbridge, M., Lilley, K., Russell, S., Ashburner, M., Mizuguchi, K., and Micklem, G. (2007). FlyMine: an integrated database for Drosophila and Anopheles genomics. *Genome Biology*, 8(7):R129.

Michal, G. and Schomburg, D., editors (2012). *Biochemical Pathways*. An Atlas of Biochemistry and Molecular Biology. Wiley.

Stanier, R. Y., Kunisawa, R., Mandel, M., and Cohen-Bazire, G. (1971). Purification and properties of unicellular blue-green algae (order Chroococcales). *Bacteriological reviews*, 35(2):171–205.

Taubert, J., Hassani-Pak, K., Castells-Brooke, N., and Rawlings, C. J. (2014). Ondex Web: web-based visualization and exploration of heterogeneous biological networks. *Bioinformatics (Oxford, England)*, 30(7):1034–1035.

Töpel, T., Kormeier, B., Klassen, A., and Hofestädt, R. (2008). BioDWH: a data warehouse kit for life science data integration. *Journal of Integrative Bioinformatics*, 5(2).

Töpel, T., Scheible, D., Trefz, F., and Hofestädt, R. (2010). RAMEDIS: a comprehensive information system for variations and corresponding phenotypes of rare metabolic diseases. *Human mutation*, 31(1):E1081–8.

Trautmann, D., Voss, B., Wilde, A., Al-Babili, S., and Hess, W. R. (2012). Microevolution in cyanobacteria: re-sequencing a motile substrain of Synechocystis sp. PCC 6803. *DNA Research: An International Journal for Rapid Publication of Reports on Genes and Genomes*, 19(6):435–448.

Triplet, T. and Butler, G. (2011). Systems Biology Warehousing: Challenges and Strategies toward Effective Data Integration. *DBKDA 2011 : The Third International Conference on Advances in Databases, Knowledge, and Data Applications*, pages 34–40.

Triplet, T., Shortridge, M. D., Griep, M. A., Stark, J. L., Powers, R., and Revesz, P. (2010). PROFESS: a PROtein function, evolution, structure and sequence database. *Database*, 2010(0):baq011–baq011.

Zhang, J., Duggan, G. E., Khaja, R., and Scherer, S. W. (2004). Bioxrt: a novel platform for developing online biological databases based on the cross-referenced tables model. In *3rd Canadian Working Conference on Computational Biology, Markham, Canada*.