

SBML Level 3 Package: Flux Balance Constraints (‘fbc’)

Brett G. Olivier

b.g.olivier@vu.nl

Systems Bioinformatics
VU University Amsterdam
Amsterdam, NH, The Netherlands

Frank T. Bergmann

fbergmann@caltech.edu

Computing and Mathematical Sciences
California Institute of Technology
Pasadena, CA, US

August 10, 2012

Version 1 (Draft)

Under construction, beware fast mutating text!

The latest release, past releases, and other materials related to this specification are available at

<http://sbml.org/Documents/Specifications/Fbc>

This release of the specification is available at

http://sbml.org/Documents/Specifications/Fbc_Level_1_Version_1



Contents

1	Introduction and motivation	3
1.1	Proposal corresponding to this package specification	3
1.2	Package dependencies	4
1.3	Document conventions	4
2	Background	5
2.1	Problems with current SBML approaches	5
2.2	Past work on this problem or similar topics	5
2.3	Design goals of the current Flux Balance Constraints Package	5
3	Proposed syntax and semantics	6
3.1	Namespace URI and other declarations necessary for using this package	6
3.2	Primitive data types	6
3.2.1	Type FbcSId	6
3.2.2	Type FbcSIdRef	6
3.3	The extended Model class	6
3.3.1	The lists of objectives and flux bounds	7
3.4	The extended Species class	7
3.5	The FluxBound class	8
3.6	The Objective class	8
3.7	The FluxObjective class	8
4	Examples	9
5	Best practices	11
	Acknowledgments	12
	References	13

1 Introduction and motivation

Constraint based modelling is a widely used methodology used to analyse and study biological networks on both a small and whole organism (genome) scale. Typically these models are underdetermined and constraint based methods (e.g. linear, quadratic optimization) are used to optimise specific model properties. This is assumed to occur under a defined set of constraints (e.g. stoichiometric, metabolic) and bounds (e.g. thermodynamic, experimental and environmental) on the values that the solution fluxes can obtain.

Perhaps the most well known (and widely used) analysis method is Flux Balance Analysis (FBA; Orth et al., 2010) which is performed on Genome Scale Reconstructions (GSR's; Oberhardt et al., 2009). Using FBA a target flux is optimized (e.g. maximising a flux to biomass or minimising ATP production) while other fluxes can be bounded to simulate a selected growth environment or specific metabolic state.

As constraint based models are generally underdetermined, i.e. few or none of the kinetic rate equations and related parameters are known, it is crucial that a model definition includes the ability to define optimisation parameters such as objective functions, flux bounds and constraints ... currently this is not possible in the Systems Biology Markup Language (SBML) Level 2 or Level 3 core specification (Hucka et al., 2011, 2003).

The question of how to encode constraint based (a.k.a. 'FBA') models in SBML is not new. However, advances in the methods used to construct GSR scale models and the wider adoption of constraint based modelling in biotechnological/medical applications have led to a rapid increase in both the number of models being constructed and the tools used to analyse them.

Faced with such growth, both in number and diversity, the need for a standardised data format for the definition, exchange and annotation of constraint based models has become critical. As the core model components (e.g. species, reactions, stoichiometry) can already be efficiently described in SBML (with its significant community, software and tool support) the Flux Balance Constraints Package aims to extend SBML core by adding the elements necessary to describe current and future constraint based models.

1.1 Proposal corresponding to this package specification

This specification for Flux Balance Constraints in SBML Level 3 Version 1 is based on the proposal by the same authors, located at the following URL:

[http://sbml.org/Community/Wiki/SBML_Level_3_Proposals/Flux_Balance_Constraints_Proposal_\(2012\)](http://sbml.org/Community/Wiki/SBML_Level_3_Proposals/Flux_Balance_Constraints_Proposal_(2012))

The tracking number in the SBML issue tracking system (SBML Team, 2010) for Flux Balance Constraints package activities is 3154219. The version of the proposal used as the starting point for this specification is the version of March 2012. Previous versions of the current proposal are:

Version 3 (March 2012)

[http://sbml.org/Community/Wiki/SBML_Level_3_Proposals/Flux_Balance_Constraints_Proposal_\(2012\)](http://sbml.org/Community/Wiki/SBML_Level_3_Proposals/Flux_Balance_Constraints_Proposal_(2012))

Version 2 (March 2011)

http://sbml.org/Community/Wiki/SBML_Level_3_Proposals/Flux_Constraints_Proposal

Version 1 (February 2010)

<http://precedings.nature.com/documents/4236/version/1>

Details of earlier independent proposals are described in Section 2.

1.2 Package dependencies

The Flux Balance Constraints package has no dependencies on other SBML Level 3 packages. It is also designed with the goal of being able to work seamlessly with other SBML Level 3 packages. For example the new elements are entirely encapsulated in their own ListOf classes and any extensions to existing SBML classes are defined as optional.

1.3 Document conventions

Following the precedent set by the SBML Level 3 Core specification document, we use UML 1.0 (Unified Modeling Language; [Eriksson and Penker 1998](#); [Oestereich 1999](#)) class diagram notation to define the constructs provided by this package. We also use color in the diagrams to carry additional information for the benefit of those viewing the document on media that can display color. The following are the colors we use and what they represent:

- **Black:** Items colored black in the UML diagrams are components taken unchanged from their definition in the SBML Level 3 Core specification document.
- **Green:** Items colored green are components that exist in SBML Level 3 Core, but are extended by this package. Class boxes are also drawn with dashed lines to further distinguish them.
- **Blue:** Items colored blue are new components introduced in this package specification. They have no equivalent in the SBML Level 3 Core specification.

We also use the following typographical conventions to distinguish the names of objects and data types from other entities; these conventions are identical to the conventions used in the SBML Level 3 Core specification document:

AbstractClass: Abstract classes are classes that are never instantiated directly, but rather serve as parents of other object classes. Their names begin with a capital letter and they are printed in a slanted, bold, sans-serif typeface. In electronic document formats, the class names defined within this document are also hyperlinked to their definitions; clicking on these items will, given appropriate software, switch the view to the section in this document containing the definition of that class. (However, for classes that are unchanged from their definitions in SBML Level 3 Core, the class names are not hyperlinked because they are not defined within this document.)

Class: Names of ordinary (concrete) classes begin with a capital letter and are printed in an upright, bold, sans-serif typeface. In electronic document formats, the class names are also hyperlinked to their definitions in this specification document. (However, as in the previous case, class names are not hyperlinked if they are for classes that are unchanged from their definitions in the SBML Level 3 Core specification.)

Something, otherThing: Attributes of classes, data type names, literal XML, and generally all tokens *other* than SBML UML class names, are printed in an upright typewriter typeface. Primitive types defined by SBML begin with a capital letter; SBML also makes use of primitive types defined by XML Schema 1.0 ([Biron and Malhotra, 2000](#); [Fallside, 2000](#); [Thompson et al., 2000](#)), but unfortunately, XML Schema does not follow any capitalization convention and primitive types drawn from the XML Schema language may or may not start with a capital letter.

For other matters involving the use of UML and XML, we follow the conventions used in the SBML Level 3 Core specification document.

2 Background

2.1 Problems with current SBML approaches

While there is currently no official way of encoding constraint based models in SBML L2 there have been pragmatic approaches used by a variety of groups and applications. Arguably the best and most widely used format is that used by the COBRA toolbox (Becker et al., 2007) where the metabolic network is well defined using SBML **Reaction** and **Species** classes. However, flux bounds and reactions that take part in the objective function are defined as **LocalParameter** classes and (implicitly) rely on all tools using the same naming convention. Similarly, reaction annotations are generally stored as key-value pairs in HTML **Notes** elements which has led to different groups routinely using different keys describing the same thing. While a step in the right direction this format is not suitable for implementation in SBML Level 3.

2.2 Past work on this problem or similar topics

The problem of describing and annotating 'FBA' models in SBML has been raised at various times in the past few years. In this regard there are two known putative proposals one by Karthik Raman and the other by the Church Laboratory. As far as we are aware these proposals never developed beyond their initial presentation at SBML forum/hackathons. In 2009 the discussion was reopened by Brett Olivier at the SBML Forum held in Stanford and subsequently led to the current series of proposals (see Section 1).

Brett Olivier (2009) SBML Level 3 FBA package discussion

http://sbml.org/images/4/4a/Olivier_sbml_forum_2009_09_04.pdf

Karthik Raman (2005) Flux annotations in SBML

<http://sbml.org/images/d/d9/Raman-flux-annotations.pdf>

Church laboratory (pre 2005) Metabolic flux model annotations

http://sbml.org/Community/Wiki/Old_known_SBML_annotations_list

2.3 Design goals of the current Flux Balance Constraints Package

3 Proposed syntax and semantics

In this section, we define the syntax and semantics of the Flux Balance Constraints package for SBML Level 3 Version 1. We expound on the various data types and constructs defined in this package, then in [Section 4 on page 9](#), we provide complete examples of using the constructs in example SBML models.

3.1 Namespace URI and other declarations necessary for using this package

Every SBML Level 3 package is identified uniquely by an XML namespace URI. For an SBML document to be able to use a given SBML Level 3 package, it must declare the use of that package by referencing its URI. The following is the namespace URI for this version of the Flux Balance Constraints package for SBML Level 3 Version 1:

`"http://www.sbml.org/sbml/level3/version1/fbc/version1"`

In addition, SBML documents using a given package must indicate whether understanding the package is required for complete mathematical interpretation of a model, or whether the package is optional. This is done using the attribute **required** on the `<sbml>` element in the SBML document. For the Flux Balance Constraints package, the value of this attribute must be set to `"true"`.

The following fragment illustrates the beginning of a typical SBML model using SBML Level 3 Version 1 and this version of the Flux Balance Constraints package:

```
<?xml version="1.0" encoding="UTF-8"?>
<sbml xmlns="http://www.sbml.org/sbml/level3/version1/core" level="3" version="1"
      xmlns:fbc="http://www.sbml.org/sbml/level3/version1/fbc/version1" fbc:required="true">
```

3.2 Primitive data types

Section 3.1 of the SBML Level 3 Version 1 Core specification defines a number of primitive data types and also uses a number of XML Schema 1.0 data types ([Biron and Malhotra, 2000](#)). More specifically we make use of **integer**, **double**, **string**, **SidRef** and **enum**. In addition we make use of two new primitives **FbcSid** and **FbcSidRef**, see [Figure 1](#) for the interrelation between these entities.

3.2.1 Type FbcSid

The type **FbcSid** is derived from **Sid** (SBML Level 3 Version 1 Core specification Section 3.1.7) and has identical syntax. The **FbcSid** type is used as the data type for the identifiers of **FluxBound** ([Section 3.5](#)) and **Objective** ([Section 3.6](#)) classes. By using a separate identifier type we differentiate them from others defined in the **SBML** model and thus ensuring data encapsulation. In addition the **Objective** class **FbcSid** provides an identifier to the **Objective** which is set as active. The equality of **FbcSid** values is determined by an exact character sequence match and therefore comparisons of these identifiers must be performed in a case-sensitive manner.

3.2.2 Type FbcSidRef

Type **FbcSidRef** is used for all attributes that refer to identifiers of type **FbcSid**. This type is derived from **FbcSid** with the restriction that the value of an attribute having type **FbcSidRef** must match the value of a **FbcSid** attribute in the current model. In the FBC package the **ListOfObjectives** has an attribute of this type that is used to refer to an instance of the **Objective** class.

3.3 The extended Model class

The **SBML Model** class is extended with the addition of two children, i.e. a **listOfFluxBounds** and a **listOfObjectives**.

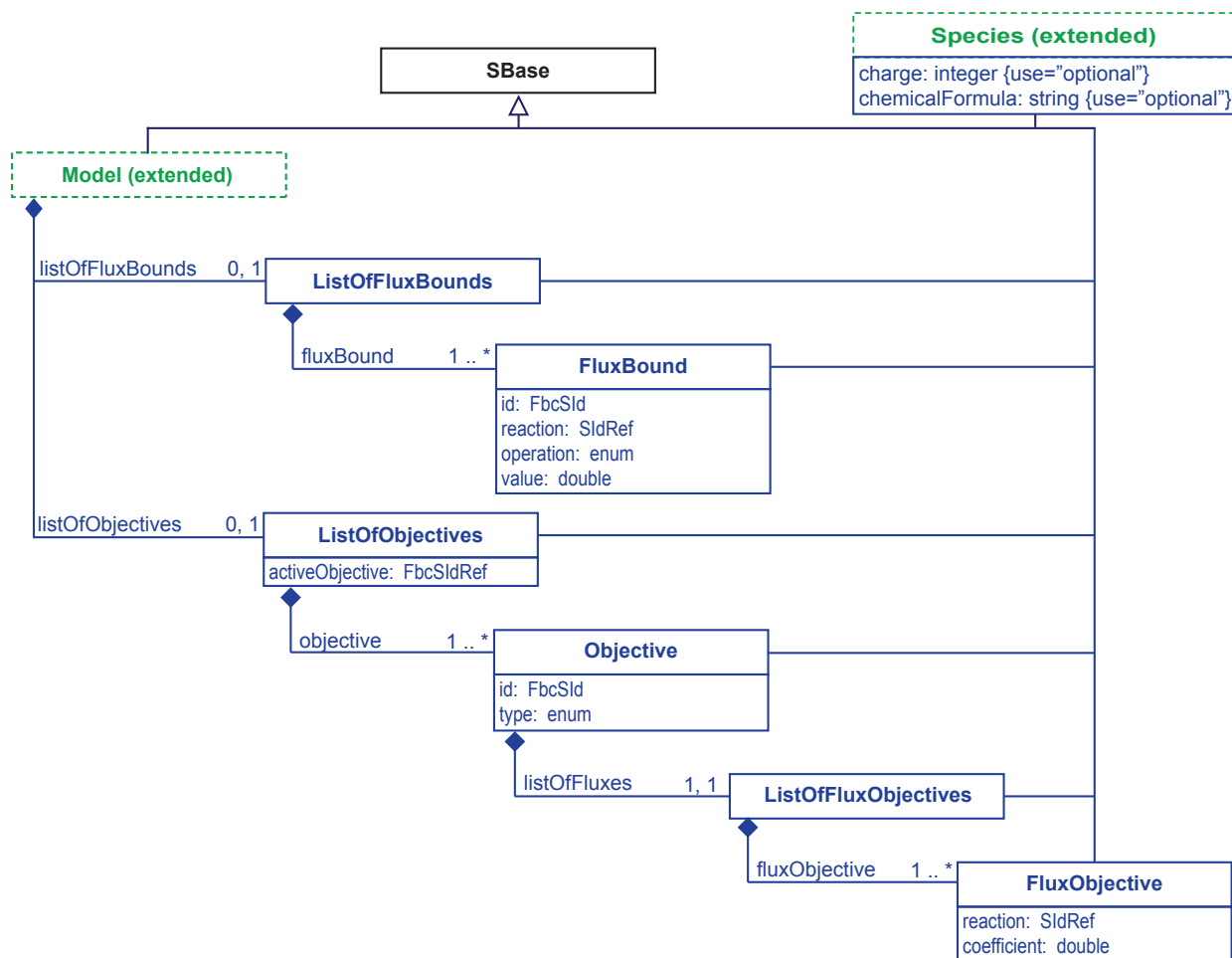


Figure 1: A UML representation of the Flux Balance Constraints package classes. See [Section 1.3](#) for conventions related to this figure.

3.3.1 The lists of objectives and flux bounds

As shown in [Figure 1](#) the **ListOfFluxBounds** and **ListOfObjectives** are derived from **SBBase** and inherit **metaid** and **sboTerm**, as well as the subcomponents for **Annotation** and **Notes**. Both of these lists are required to contain one or more elements when defined, however, the lists themselves are optional. Unlike most other **SBML ListOf__** classes, **ListOfObjectives** introduces an additional required attribute **activeObjective**

The activeObjective attribute

This attribute contains a “value” of type **FbcSIdRef** that exclusively refers to an existing **objective**. This attribute exists so that in the case where multiple **Objective**’s are defined in a single model the model will always be complete, i.e. there is a single, primary optimization target.

3.4 The extended Species class

The ‘FBC’ package extends the **SBML Species** class with the addition of two attributes:

- an optional attribute **charge** which contains an **integer** referring to the **Species**’ charge (as defined in **SBML Level 2**)

- an optional attribute **chemicalFormula** containing a **string** that represents the **Species'** elemental composition.

The chemicalFormula attribute

While there are many ways of referring to an elemental composition the purpose of the **chemicalFormula** attribute is to allow reaction balancing and validation which is particularly important in constraint based models. To this end it is recommended that the format of **chemicalFormula** should follow the Hill system (or notation). Here the number of carbon atoms in a molecule is indicated first, followed by the number of hydrogen atoms and then the number of all other chemical elements in alphabetical order. When the formula contains no carbon; all elements, including hydrogen, are listed alphabetically [Hill \(1900, 2012\)](#).

3.5 The FluxBound class

FluxBound is a new FBC class derived from **SBML SBBase** that inherits **metaid** and **sboTerm**, as well as the subcomponents for **Annotation** and **Notes**. The purpose of this class is to hold the a single (in)equality that provides the maximum or minimum value that a reaction flux can obtain at steady state. It is relatively straight forward and implements four attributes.

- **id** an attribute that can take exclusively contain an **FbcSId**
- **reaction** an attribute that takes an **SIdRef** which refers only to an **SBML Reaction**
- **operation** an **enum** that can take a limited set of boolean operators (see text for details on the enumerated allowed types)
- **value** an attribute that takes a **double** value representing the bound. This may include $\pm\infty$

The operation attribute

The **operation** attribute represents the (in)equality in the expression **reaction**, **operator** **value** an enumerated type that can have the following values:

- **"lessEqual"** representing \leq
- **"greaterEqual"** representing \geq
- **"less"** representing $<$
- **"greater"** representing $>$
- **"equal"** representing $=$
- **"unknown"** representing an undefined symbol

3.6 The Objective class

3.7 The FluxObjective class

4 Examples

This section contains a variety of examples of SBML Level 3 Version 1 documents employing the Flux Balance Constraints package. To begin with let's contrast some elements of an existing model, iJR904 from the BiGG database [Reed et al. \(2003\)](#); [Schellenberger et al. \(2010\)](#) that has been translated into SBML Level 3 Version 1 with FBC.

Reaction definition and flux bounds

Existing syntax:

```
1 <reaction id="R_GTHS" name="glutathione_synthetase" reversible="false">
2 <notes>
3 <html:p>Abbreviation: R_GTHS</html:p>
4 <html:p>Synonyms: _0</html:p>
5 <html:p>EC Number: 6.3.2.3</html:p>
6 <html:p>SUBSYSTEM: Cofactor and Prosthetic Group Biosynthesis</html:p>
7 <html:p>Equation: [c] : atp + glucys + gly --> adp + gthrd + h + pi</html:p>
8 <html:p>Confidence Level: 0</html:p>
9 <html:p>NCD</html:p>
10 <html:p>genes:</html:p>
11 <html:p>LOCUS:b2947#ABBREVIATION:gshB#ECNUMBERS:6.3.2.3#</html:p>
12 <html:p>proteins:</html:p>
13 <html:p>NAME:glutathione synthase#ABBREVIATION:GshB#</html:p>
14 <html:p>GENE ASSOCIATION: (b2947)</html:p>
15 </notes>
16 <listOfReactants>
17 <speciesReference species="M_atp_c" stoichiometry="1"/>
18 <speciesReference species="M_glucys_c" stoichiometry="1"/>
19 <speciesReference species="M_gly_c" stoichiometry="1"/>
20 </listOfReactants>
21 <listOfProducts>
22 <speciesReference species="M_adp_c" stoichiometry="1"/>
23 <speciesReference species="M_gthrd_c" stoichiometry="1"/>
24 <speciesReference species="M_h_c" stoichiometry="1"/>
25 <speciesReference species="M_pi_c" stoichiometry="1"/>
26 </listOfProducts>
27 <kineticLaw>
28 <math xmlns="http://www.w3.org/1998/Math/MathML">
29 <ci>FLUX_VALUE</ci>
30 </math>
31 <listOfParameters>
32 <parameter id="LOWER_BOUND" value="0" units="mmol_per_gDW_per_hr"/>
33 <parameter id="UPPER_BOUND" value="999999" units="mmol_per_gDW_per_hr"/>
34 <parameter id="OBJECTIVE_COEFFICIENT" value="0" />
35 <parameter id="FLUX_VALUE" value="0" units="mmol_per_gDW_per_hr"/>
36 </listOfParameters>
37 </kineticLaw>
38 </reaction>
```

New syntax:

```

1 <reaction metaid="meta_R_GTHS" id="R_GTHS" name="glutathione_synthetase" reversible="false">
2   <annotation>
3     <listOfKeyValueData xmlns="http://pysces.sourceforge.net/KeyValueData">
4       <data id="subsystem" type="string" value="Cofactor_and_Prosthetic_Group_Biosynthesis"/>
5       <data id="name" type="string" value="glutathione_synthase#ABBREVIATION:GshB#"/>
6       <data id="gene_association" type="string" value="(b2947)"/>
7       <data id="equation" type="string" value="[c]_atp+_glucys+_gly_-->_adp+_gthrd+_h+_pi"/>
8       <data id="genes" type="string"/>
9       <data id="proteins" type="string"/>
10      <data id="locus" type="string" value="b2947#ABBREVIATION:gshB#ECNUMBERS:6.3.2.3#"/>
11      <data id="abbreviation" type="string" value="R_GTHS"/>
12      <data id="synonyms" type="string" value="_0"/>
13      <data id="confidence_level" type="string" value="0"/>
14    </listOfKeyValueData>
15    <rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#" xmlns:dc="http://purl.org/dc/elements/1.1/" xmlns:bqbiol="http://biological-ontology.org/bqbiol#" >
16      <rdf:Description rdf:about="#meta_R_GTHS">
17        <bqbiol:is>
18          <rdf:Bag>
19            <rdf:li rdf:resource="http://identifiers.org/ec-code/6.3.2.3"/>
20          </rdf:Bag>
21        </bqbiol:is>
22      </rdf:Description>
23    </rdf:RDF>
24  </annotation>
25 </reaction>

1 <fbc:listOfFluxBounds>
2   <fbc:fluxBound fbc:id="R_GTHS_lower_bnd" fbc:reaction="R_GTHS" fbc:operation="greaterEqual" fbc:value="0"/>
3   <fbc:fluxBound fbc:id="R_GTHS_upper_bnd" fbc:reaction="R_GTHS" fbc:operation="lessEqual" fbc:value="999999"/>
4 </fbc:listOfFluxBounds>

```

Objective function definition

Existing syntax:

```

1 <reaction id="R_BiomassEcoli" name="BiomassEcoli" reversible="false">
2   <kineticLaw>
3     <math xmlns="http://www.w3.org/1998/Math/MathML">
4       <ci>FLUX_VALUE</ci>
5     </math>
6     <listOfParameters>
7       <parameter id="LOWER_BOUND" value="0" units="mmol_per_gDW_per_hr"/>
8       <parameter id="UPPER_BOUND" value="999999" units="mmol_per_gDW_per_hr"/>
9       <parameter id="OBJECTIVE_COEFFICIENT" value="1" />
10      <parameter id="FLUX_VALUE" value="0" units="mmol_per_gDW_per_hr"/>
11    </listOfParameters>
12  </kineticLaw>
13 </reaction>

```

New syntax:

```

1 <fbc:listOfObjectives fbc:activeObjective="obj1">
2   <fbc:objective fbc:id="obj1" fbc:type="maximize">
3     <fbc:listOfFluxes>
4       <fbc:fluxObjective fbc:reaction="R_BiomassEcoli" fbc:coefficient="1"/>
5     </fbc:listOfFluxes>
6   </fbc:objective>
7 </fbc:listOfObjectives>

```

5 Best practices

In this section, we recommend a number of practices for using and interpreting various constructs in the Flux Balance Constraints package. These recommendations are non-normative, but we advocate them strongly; ignoring them will not render a model invalid, but may reduce interoperability between software and models.

Acknowledgments

We thank Herbert Sauro (University of Washington, Washington, USA), Neil Swainston (University of Manchester, Manchester, UK), Keiran Smallbone (University of Manchester), and Michael Hucka (Caltech, California, USA) for help developing this Flux Balance Constraints package for SBML Level 3. We also thank everyone who contributed ideas and discussion over the years.

References

- Becker, S. A., Feist, A. M., Mo, M. L., Hannum, G., Palsson, B. O., and Herrgard, M. J. (2007). Quantitative prediction of cellular metabolism with constraint-based models: the cobra toolbox. *Nat. Protocols*, 2(3):727–738. 1750-2799 10.1038/nprot.2007.99 10.1038/nprot.2007.99.
- Biron, P. V. and Malhotra, A. (2000). XML Schema part 2: Datatypes (W3C candidate recommendation 24 October 2000). Available via the World Wide Web at <http://www.w3.org/TR/xmlschema-2/>.
- Eriksson, H.-E. and Penker, M. (1998). *UML Toolkit*. John Wiley & Sons, New York.
- Fallside, D. C. (2000). XML Schema part 0: Primer (W3C candidate recommendation 24 October 2000). Available via the World Wide Web at <http://www.w3.org/TR/xmlschema-0/>.
- Hill, E. A. (1900). On a system of indexing chemical literature; adopted by the classification division of the u. s. patent office.1. *Journal of the American Chemical Society*, 22(8):478–494.
- Hill, E. A. (2012). Wikipedia: The hill system.
- Hucka, M., Bergmann, F. T., Hoops, S., Keating, S. M., Sahle, S., and Wilkinson, D. J. (2011).
- Hucka, M., Finney, A., Sauro, H., Bolouri, H., Doyle, J., Kitano, H., Arkin, A., Bornstein, B., Bray, D., Cornish-Bowden, A., Cuellar, A., Dronov, S., Gilles, E., Ginkel, M., Gor, V., Goryanin, I., Hedley, W., Hodgman, T., Hofmeyr, J., Hunter, P., Juty, N., Kasberger, J., Kremling, A., Kummer, U., Novère, N. L., Loew, L., Lucio, D., Mendes, P., Minch, E., Mjolsness, E., Nakayama, Y., Nelson, M., Nielsen, P., Sakurada, T., Schaff, J., Shapiro, B., shimizu, T., Spence, H., Stelling, J., Takahashi, K., Tomita, M., Wagner, J., and Wang, J. (2003). The Systems Biology Markup Language (SBML): a medium for representation and exchange of biochemical network models. *Bioinformatics*, 19(4):524–31.
- Oberhardt, M. A., Palsson, B. Ø., and Papin, J. A. (2009). Applications of genome-scale metabolic reconstructions. *Molecular Systems Biology*, 5:320.
- Oestereich, B. (1999). *Developing Software with UML: Object-Oriented Analysis and Design in Practice*. Addison-Wesley.
- Orth, J. D., Thiele, I., and Palsson, B. Ø. (2010). What is flux balance analysis? *Nature Biotechnology*, 28(3):245–248.
- Reed, J. L., Vo, T. D., Schilling, C. H., and Palsson, B. O. (2003). An expanded genome-scale model of escherichia coli k-12 (ijr904 gsm/gpr). *Genome Biol*, 4(9):R54. GM57092/GM/NIGMS NIH HHS/United States GM62791/GM/NIGMS NIH HHS/United States Journal Article Research Support, Non-U.S. Gov't Research Support, U.S. Gov't, P.H.S. England.
- SBML Team (2010). The SBML issue tracker.
- Schellenberger, J., Park, J., Conrad, T., and Palsson, B. (2010). Bigg: a biochemical genetic and genomic knowledgebase of large scale metabolic reconstructions. *BMC Bioinformatics*, 11(1):213.
- Thompson, H. S., Beech, D., Maloney, M., and Mendelsohn, N. (2000). XML Schema part 1: Structures (W3C candidate recommendation 24 October 2000). Available online via the World Wide Web at the address <http://www.w3.org/TR/xmlschema-1/>.