



Editorial

Falk Schreiber*, Björn Sommer, Tobias Czauderna, Martin Golebiewski, Thomas E. Gorochoowski, Michael Hucka, Sarah M. Keating, Matthias König, Chris Myers, David Nickerson and Dagmar Waltemath

Specifications of standards in systems and synthetic biology: status and developments in 2020

<https://doi.org/10.1515/jib-2020-0022>

Received April 26, 2020; accepted April 27, 2020

Abstract: This special issue of the *Journal of Integrative Bioinformatics* presents papers related to the 10th COMBINE meeting together with the annual update of COMBINE standards in systems and synthetic biology.

Keywords: ontologies; standards; systems biology; synthetic biology.

1 Introduction

COMBINE (‘COMputational Modelling in BIOlogy’ NETWORK) [1, 2], the formal entity which coordinates the development of standards in systems and synthetic biology, celebrated 10 years of activity in 2019. COMBINE not only coordinates standard developments, but also fosters and moderates discussions; designs and implements dissemination strategies; and organises two annual community meetings – COMBINE and HARMONY.

This special issue contains two papers, one new standard, and six updates of standards. Waltemath et al. [3] discuss the first 10 years of the international coordination network for standards in systems and synthetic biology, and summarises the COMBINE meeting in Heidelberg in July 2019. Brunak et al. [4] present ongoing works and open questions towards standardisation guidelines for *in silico* approaches in personalised medicine.

COMBINE standards and associated initiatives cover a wide range of disciplines, see Figure 1. This special issue only highlights updates over the last year, namely the CellML 2.0 specification, the Systems Biology Graphical Notation - Markup Language (SBGN-ML) Milestone 3, the SBML (Systems Biology Markup Language) Level 3 Packages “Distrib” and “Multi”, the SBML (Systems Biology Markup Language) Version 3.0.0, and the SBOL (Synthetic Biology Open Language) Visual Version 2.2. Additionally, one new standard has officially

*Corresponding author: **Falk Schreiber**, Dept. of Computer and Information Science, University of Konstanz, Konstanz, Germany; and Faculty of Information Technology, Monash University, Melbourne, Australia, E-mail: falk.schreiber@uni-konstanz.de

Björn Sommer: Royal College of Arts, London, UK

Tobias Czauderna: Faculty of Information Technology, Monash University, Melbourne, Australia. <https://orcid.org/0000-0002-1788-9593>

Martin Golebiewski: Heidelberg Institute for Theoretical Studies (HITS), Heidelberg, Germany

Thomas E. Gorochoowski: School of Biological Sciences, University of Bristol, Bristol, UK

Michael Hucka: California Institute of Technology, Pasadena, USA

Sarah M. Keating: EBI Hinxton, Cambridge, UK

Matthias König: Institute for Theoretical Biology, Humboldt-University Berlin, Berlin, Germany

Chris Myers: Dept. of Electrical and Computer Engineering, University of Utah, Salt Lake City, USA

David Nickerson: Auckland Bioengineering Institute, University of Auckland, Auckland, New Zealand. <https://orcid.org/0000-0003-4667-9779>

Dagmar Waltemath: University Medicine Greifswald, Greifswald, Germany. <https://orcid.org/0000-0002-5886-5563>

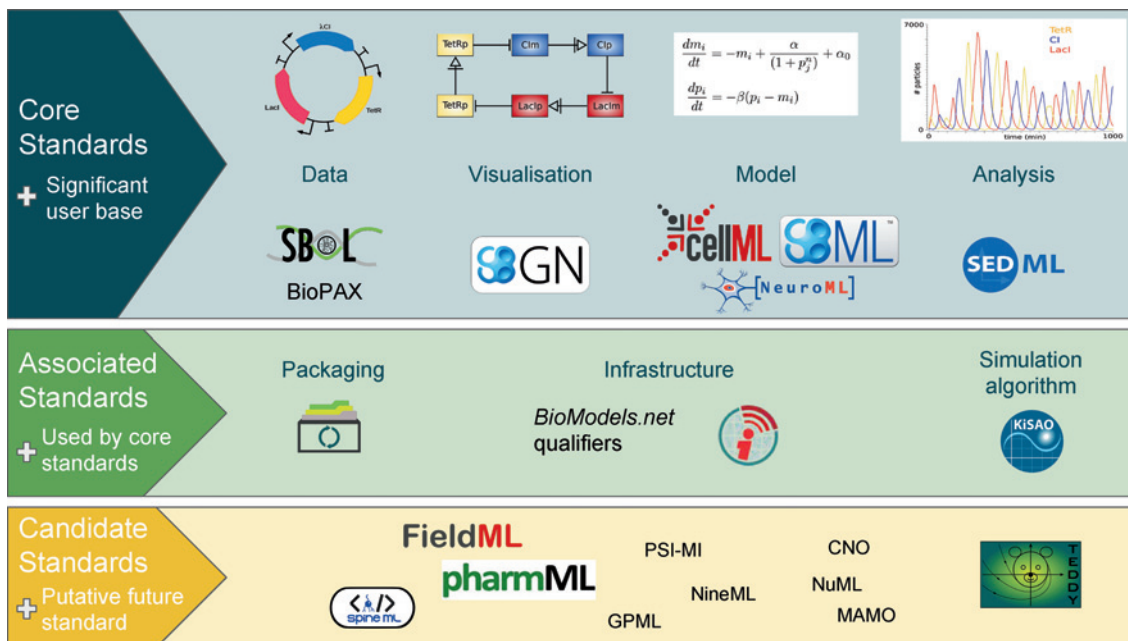


Figure 1: Standards and associated efforts in systems and synthetic biology (from Ref. [7]).

been added, Open Modelling EXchange format (OMEX) metadata specification 1.0, to harmonise the descriptions of metadata.

Further information on all standards and activities as well as links to the community websites are available from the COMBINE web site at <https://co.mbine.org/>. Detailed overviews of COMBINE, its history and its organisation have been provided, for example, by Hucka et al. [1], Myers et al. [2] or Waltemath et al. [5]. The annual special issue on COMBINE standards has become a tradition since its launch in 2016. Earlier editions provide summaries of updates for the years 2015–18 [6–9].

We hope that this editorial is helpful in identifying the relevant specification documents for standards in systems in synthetic biology in the year 2020.

2 Current versions of COMBINE standards

Please refer to the following specifications when using COMBINE standards. New specifications or updates of existing specifications are highlighted with *NEW*.

2.1 Core standards

2.1.1 BioPAX (Biological PATHway eXchange)

Biological PATHway eXchange (BioPAX) is a standard language for integration, exchange and analysis of biological pathway data. It is expressed in OWL. The current specification is listed in Table 1.

Table 1: Latest specifications for BioPAX.

Standard	Specification	Reference
BioPAX [10]	BioPAX	[11]

Table 2: Latest specifications for CellML.

Standard	Specification	Reference
CellML [13]	CellML 2.0	[12]
	CellML Metadata Framework 2.0	[14]
	CellML 1.1	[15]

Table 3: Latest specifications for BioPAX.

Standard	Specification	Reference
NeuroML [16, 17]	NeuroML version 2.0	[16]

2.1.2 CellML

The CellML language is an XML markup language to store and exchange computer-based mathematical models. The current specifications are listed in Table 2.

NEW CellML 2.0 [12]: The development of CellML 2.0 was guided by observing the use of CellML 1.1 in the community for over 10 years. The syntax of CellML 1.1 has been clarified in areas where discrepancies in model interpretation are often seen and simplified to remove features that are never used. These enhancements are primarily aimed at improving the model reuse capabilities of CellML. The single substantial addition to CellML that is introduced in CellML 2.0 is the concept of *resets*, rules that define a change in model state dependent on specified conditions being met during a simulation experiment.

2.1.3 NeuroML (Neural Open Markup Language)

The NeuroML is an XML-based description language that provides a common data format for defining and exchanging descriptions of neuronal cell and network models. The current specification is listed in Table 3.

2.1.4 SBGN (Systems Biology Graphical Notation)

The SBGN, is a set standard graphical languages to describe visually biological knowledge. It is currently made up of three languages describing Process Descriptions, Entity Relationships and Activity Flows. In addition, SBGN-ML is an XML-based file format describing the geometry of SBGN maps, while preserving their underlying biological meaning. The current specifications are listed in Table 4.

NEW SBGN-ML Milestone 3 [18] includes new developments, such as support for multiple SBGN maps within a single file, complete support for the submap glyph, and the possibility to store colours and annotations through extensions. In addition, the *language* attribute has been deprecated to add a more detailed *version* attribute and the SBGN AF perturbation glyph has been deprecated to align with the SBGN AF specification.

Table 4: Latest specifications for SBGN.

Standard	Specification	Reference
SBGN [19]	SBGN Process Description Level 1 Version 2	[20]
	SBGN Entity Relationship Level 1 Version 2.0	[21]
	SBGN Activity Flow Level 1 Version 1.2	[22]
	SBGN Markup Language Version 0.3	[18]

Table 5: Latest specifications for SBGN.

Standard	Specification	Reference
SBML [23]	SBML Level 3 Core, Version 2 Release 2	[24]
	SBML Level 3 Package: Distributions, Version 1, Release 1	[25]
	SBML Level 3 Package: Flux Balance Constraints, Version 2	[26]
	SBML Level 3 Package: Groups, Version 1	[27]
	SBML Level 3 Package: Hierarchical Model Composition, Version 1	[28]
	SBML Level 3 Package: Layout, Version 1	[29]
	SBML Level 3 Package: Multistate, Multicomponent and Multicompartment Species, Version 1, Release 2	[30]
	SBML Level 3 Package: Qualitative Models, Version 1	[31]
	SBML Level 3 Package: Render, Version 1, Release 1	[32]

2.1.5 SBML (Systems Biology Markup Language)

The SBML is a computer-readable XML format for representing models of biological processes. SBML is suitable for, but not limited to, models using a process description approach. SBML development is coordinated by an elected editorial board and central developer team. The current specifications are listed in Table 5.

NEW SBML Level 3 Package: Distributions, Version 1, Release 1 [25] introduces distributions and uncertainties to SBML. Biological models often contain elements that have inexact numerical values, since they are based on values that are stochastic in nature or data that contains uncertainty. The SBML Level 3 Core specification does not include an explicit mechanism to include inexact or stochastic values in a model, but it does provide a mechanism for SBML packages to extend the Core specification and add additional syntactic constructs. The SBML Distributions package for SBML Level 3 adds the necessary features to allow models to encode information about the distribution and uncertainty of values underlying a quantity.

NEW SBML Level 3 Package: Multistate, Multicomponent and Multicompartment Species, Version 1 Release 2 [30] addresses some issues raised by users about unclear aspects of Release 1 of the specification; it also clarifies the use of XML namespaces, and updates example models.

2.1.6 SBOL (Synthetic Biology Open Language)

The SBOL is a language for the description and the exchange of synthetic biological parts, devices and systems. The current specifications are listed in Table 6.

NEW SBOL Version 3.0.0 [33] condenses and simplifies previous versions of SBOL based on experiences in deployment across a variety of scientific and industrial settings. In particular, SBOL 3.0.0, (1) separates sequence features from part/sub-part relationships, (2) renames ComponentDefinition/Component to Component/Sub-Component, (3) merges Component and Module classes, (4) ensuring consistency between data model and ontology terms, (5) extends the means to define and reference SubComponents, (6) refines requirements on object URIs, (7) enables graph-based serialisation, (8) moves Systems Biology Ontology (SBO) for Component types, (9)

Table 6: Latest specifications for SBOL.

Standard	Specification	Reference
SBOL [35]	SBOL Version 3.0.0	[33]
	SBOL Visual Version 2.2	[34]

Table 7: Latest specifications for SED-ML.

Standard	Specification	Reference
SED-ML [36]	SED-ML Level 1 Version 3	[37]

Table 8: Latest specifications of associated standards.

Associated standard	Specification	Reference
COMBINE Archive [39]	COMBINE Archive 1.0	[40]
OMEX Metadata	OMEX Metadata Version 1.0	[38]
BioModels.net qualifiers [41]		[42]
Identifiers.org URIs [43]		[44]
Systems Biology Ontology [45]	[external] Biportal	[46]
Kinetic Simulation Algorithm Ontology [45]	[external] Biportal	[47]

makes all sequence associations explicit, (10) makes interfaces explicit, (11) generalises SequenceConstraints into a general structural Constraint class, and (12) expands the set of allowed constraints.

NEW Synthetic Biology Open Language Visual (SBOL Visual) Version 2.2 [34] is a refinement to the standard harmonising the ontology used and extending the glyph library to capture new biological parts. Specifically, the changes in SBOL Visual 2.2. include, (1) the grounding of the molecular species glyphs is changed from BioPAX to SBO to better align with the use of SBO terms for interaction glyphs, (2) new glyphs are added for proteins, introns, and polypeptide regions (e. g. protein domains), (3) the prior recommended macromolecule glyph is deprecated in favour of its alternative, and (4) small polygons are proposed as alternative glyphs for simple chemicals.

2.1.7 SED-ML (Simulation Experiment Description Markup Language)

The Simulation Experiment Description Markup Language (SED-ML) is an XML-based format for encoding simulation experiments. SED-ML allows to define the models to use, the experimental tasks to run and which results to produce. SED-ML can be used with models encoded in several languages, as long as they are in XML. The current specification is listed in Table 7.

2.2 Associated standards

Associated standards provide an additional layer of semantics to COMBINE representation formats. The current specifications are listed in Table 8.

A COMBINE archive is a single file bundling the various documents necessary for a modelling and simulation project, and all relevant information. The archive is encoded using the OMEX. COMBINE archive metadata provides a harmonised, community-driven approach for annotating a variety of standardised model and data representation formats within a COMBINE archive. BioModels.net qualifiers are standardised relationships (predicates) that specify the relation between an object represented in a description language and the external resource used to annotate it. The relationship is rarely one-to-one, and the information content of an annotation is greatly increased if one knows what it represents, rather than only know it is “related to” the model component. MIRIAM Unique Resource Identifiers allow one to uniquely and unambiguously identify an entity in a stable and perennial manner. MIRIAM Registry is a set of services and resources that provide support for generating, interpreting and resolving MIRIAM URIs. Through the Identifiers.org technology, MIRIAM URIs can be dereferenced in a flexible and robust way.

MIRIAM URIs are used by SBML, SED-ML, CellML and BioPAX controlled annotation schemes. The SBO is a set of controlled, relational vocabularies of terms commonly used in Systems Biology, and in particular in computational modelling.

Each element of an SBML file carries an optional attribute `sboTerm` which value must be a term from SBO. Each symbol of SBGN is associated with an SBO term.

The Kinetic Simulation Algorithm Ontology (KISAO) describes existing algorithms and their inter-relationships through their characteristics and parameters.

Kinetic Simulation Algorithm Ontology is used in SED-ML, which allows simulation software to automatically choose the best algorithm available to perform a simulation and unambiguously refer to it.

NEW The OMEX Metadata Specification Version 1.0 [38] has been developed to harmonise the way in which computational models and other types of modelling files are annotated. Guided by consensus across the COMBINE community, the specification provides technical guidelines for encoding metadata that describes the contents of modelling projects within OMEX-formatted COMBINE archives. By helping to ensure that computable knowledge about modelling projects is encoded and shared in a consistent manner across the broader modelling community, the specification can help promote model reuse, reproducibility, discovery and semantics-based analyses.

Acknowledgments: We thank Gary D. Bader, Pdraig Gleeson, and all members of the COMBINE initiative for their help.

Author contribution: All the authors have accepted responsibility for the entire content of this submitted manuscript and approved submission.

Research funding: None declared.

Conflict of interest statement: The authors declare no conflicts of interest regarding this article.

References

1. Hucka M, Nickerson DP, Bader GD, Bergmann FT, Cooper J, Demir E, et al. Promoting coordinated development of community-based information standards for modeling in biology: the COMBINE initiative. *Front Bioeng Biotechnol* 2015;3:19.
2. Myers C, Bader GD, Gleeson P, Golebiewski M, Hucka M, Le Novère N, et al. A brief history of COMBINE. In: *Proceedings of the 2017 winter simulation conference*; 2017:884–95 pp. <https://doi.org/10.1109/wsc.2017.8247840>.
3. Waltemath D, Golebiewski M, Blinov ML, Gleeson P, Hermjakob H, Inau ET, et al. The first 10 years of the international coordination network for standards in systems and synthetic biology (COMBINE). *J Integr Bioinf* 2020;17:5.
4. Brunak S, Collin CB, Cathaoir KEO, Golebiewski M, Kirschner M, Kockum I, et al. Towards standardization guidelines for in silico approaches in personalized medicine. *J Integr Bioinf* 2020;17:6.
5. Waltemath D, Bergmann FT, Chaouiya C, Czauderna T, Gleeson P, Goble C, et al. Meeting report from the fourth meeting of the computational modeling in biology network (COMBINE). *Stand Genom Sci* 2014;9:1285–301.
6. Schreiber F, Bader GD, Golebiewski M, Hucka M, Kormeier B, Le Novère L, et al. Specifications of standards in systems and synthetic biology. *J Integr Bioinf* 2015;12:258.
7. Schreiber F, Bader GD, Gleeson P, Golebiewski M, Hucka M, Le Novère N, et al. Specifications of standards in systems and synthetic biology: status and developments in 2016. *J Integr Bioinf* 2016;13:289.
8. Schreiber F, Bader GD, Gleeson P, Golebiewski M, Hucka M, Keating SM, et al. Specifications of standards in systems and synthetic biology: status and developments in 2018. *J Integr Bioinf* 2018;15:13.
9. Schreiber F, Sommer B, Bader GD, Gleeson P, Golebiewski M, Hucka M, et al. Specifications of standards in systems and synthetic biology: status and developments in 2019. *J Integr Bioinf* 2019;16:35.
10. Demir E, Cary MP, Paley S, Fukuda K, Lemer C, Vastrik I, et al. The BioPAX community standard for pathway data sharing. *Nat Biotechnol* 2010;28:935–42.
11. BioPax 2017. <http://www.biopax.org/>.
12. Clerx M, Cooling MT, Cooper J, Garny A, Moyle K, Nickerson DP, et al. CellML 2.0. *J Integr Bioinf* 2020;17:21.
13. Cuellar AA, Lloyd CM, Nielsen PF, Bullivant D, Nickerson D, Hunter P. An overview of CellML 1.1, a biological model description language. *Simulation* 2003;79:740–7.
14. Cooling MT, Hunter PJ. The CellML metadata framework 2.0 specification. *J Integr Bioinf* 2015;12:86–103.
15. Cuellar AA, Hedley W, Nelson M, Lloyd CM, Halstead MDB, Bullivant DP, et al. The CellML 1.1 specification. *J Integr Bioinf* 2015;12:4–85.
16. Cannon RC, Gleeson P, Crook S, Ganapathy G, Marin B, Piasini E, et al. LEMS: a language for expressing complex biological models in concise and hierarchical form and its use in underpinning NeuroML 2. *Front Neuroinf* 2014;8. <https://doi.org/10.3389/fninf.2014.00079>.
17. Gleeson P, Crook S, Cannon RC, Hines ML, Billings GO, Farinella M, et al. NeuroML: a language for describing data driven models of neurons and networks with a high degree of biological detail. *PLoS Comput Biol* 2010;6:e1000815.
18. Bergmann FT, Czauderna T, Dogrusoz U, Rougny A, Dräger A, Toure V, et al. Systems biology graphical notation markup language (SBGNML) Version 0.3. *J Integr Bioinf* 2020;17:16.
19. Le Novère N, Hucka M, Mi H, Moodie S, Schreiber F, Anatoly S, et al. The systems biology graphical notation. *Nat Biotechnol* 2009;27:735–41.

20. Rougny A, Toure V, Moodie S, Balaur I, Czauderna T, Borlinghaus H, et al. Systems biology graphical notation: process description language level 1 version 2. *J Integr Bioinf* 2019;16:22.
21. Sorokin AA, Le Novère N, Luna A, Czauderna T, Demir E, Haw R, et al. Systems biology graphical notation: entity relationship language level 1 version 2. *J Integr Bioinf* 2015;12:281–339.
22. Mi H, Schreiber F, Moodie SL, Czauderna T, Demir E, Haw R, et al. Systems biology graphical notation: activity flow language level 1 version 1.2. *J Integr Bioinf* 2015;12:340–81.
23. Hucka M, Finney A, Sauro HM, Bolouri H, Doyle JC, Kitano H, et al. The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. *Bioinformatics* 2003;19:524–31.
24. Hucka M, Bergmann FT, Chaouiya C, Dräger A, Hoops S, Keating SM, et al. The systems biology markup language (SBML): language specification for level 3 version 2 core release 2. *J Integr Bioinf* 2019;16:21.
25. Smith LP, Moodie SL, Bergmann FT, Gillespie C, Keating SM, König M, et al. SBML level 3 package: distributions, version 1, release 1. *J Integr Bioinf* 2020;17:18.
26. Olivier BG, Bergmann FT. SBML level 3 package: flux balance constraints, version 2. *J Integr Bioinf* 2018;15:82.
27. Hucka M, Smith LP. The systems biology markup language (SBML) level 3 package: groups, version 1 release 1. *J Integr Bioinf* 2016;13:290.
28. Smith LP, Hucka M, Hoops S, Finney A, Ginkel M, Myers CJ, et al. SBML level 3 package: hierarchical model composition, version 1 release 3. *J Integr Bioinf* 2015;12:603–59.
29. Gauges R, Rost U, Sahle S, Wengler K, Bergmann FT. The systems biology markup language (SBML) level 3 package: Layout, version 1 core. *J Integr Bioinf* 2015;12:550–602.
30. Zhang F, Smith LP, Blinov ML, Faeder J, Hlavacek WS, Tapia JJ, et al. SBML Level 3 package: multistate, multicomponent and multicompartments species, version 1, release 2. *J Integr Bioinf* 2020;17:15.
31. Chaouiya C, Keating SM, Bérenguier D, Naldi A, Thieffry D, van Iersel MP, et al. The systems biology markup language (SBML) level 3 package: qualitative models, version 1, release 1. *J Integr Bioinf* 2015;12:691–703.
32. Bergmann FT, Keating SM, Gauges R, Sahle S, Wengler K. SBML level 3 package: render, version 1, release 1. *J Integr Bioinf* 2018;15:78.
33. Baig H, Fontanarossa P, Kulkarni V, McLaughlin JA, Vaidyanathan P, Bartley B, et al. Synthetic biology open language (SBOL) version 3.0.0. *J Integr Bioinf* 2020;17:17.
34. Baig H, Fontanarossa P, Kulkarni V, McLaughlin JA, Vaidyanathan P, Bartley B, et al. Synthetic biology open language visual (SBOL visual) Version 2.2. *J Integr Bioinf* 2020;17:14.
35. Galdzicki M, Clancy KP, Oberortner E, Pocock M, Quinn JY, Rodriguez CA, et al. The synthetic biology open language (SBOL) provides a community standard for communicating designs in synthetic biology. *Nat Biotechnol* 2014;32:545–50.
36. Waltemath D, Adams R, Bergmann FT, Hucka M, Kolpakov F, Miller AK, et al. Reproducible computational biology experiments with SED-ML – the Simulation experiment description markup language. *BMC Sys Biol* 2011;5:198.
37. Bergmann F, Cooper J, König M, Moraru I, Nickerson D, Novère NL, et al. Simulation experiment description markup language (SED-ML) level 1 version 3 (L1V3). *J Integr Bioinf* 2018;15:86.
38. Neal ML, Gennari JH, Waltemath D, Nickerson DP, König M. OMEX metadata specification version 1.0. *J Integr Bioinf* 2020;17:20.
39. Bergmann FT, Adams R, Moodie S, Cooper J, Glont M, Golebiewski M, et al. COMBINE archive and OMEX format: one file to share all information to reproduce a modeling project. *BMC Bioinf* 2014;15:369.
40. Bergmann FT, Rodríguez N, Le Novère N. COMBINE archive specification version 1. *J Integr Bioinf* 2015;12:104–18.
41. Li C, Courtot M, Le Novère N, Laibe C. BioModels. Net web services, a free and integrated toolkit for computational modelling software. *Briefings Bioinf*;11 2010:270–7.
42. BioModels Net qualifiers. 2020. <http://co.mbine.org/standards/qualifiers>.
43. Juty N, Le Novère N, Laibe C. Identifiers. Org and miriam registry: community resources to provide persistent identification. *Nucleic Acids Res* 2012;40:D580–6.
44. Identifiers.org. 2020. <http://identifiers.org/>.
45. Courtot M, Juty N, Knüpfer C, Waltemath D, Zhukova A, Dräger A, et al. Controlled vocabularies and semantics in systems biology. *Mol Syst Biol* 2011;7:543.
46. BioPortal. Systems biology ontology on bioportal, 2019. <http://bioportal.bioontology.org/ontologies/SBO>.
47. BioPortal. Kinetic simulation algorithm ontology on bioportal, 2019. <http://bioportal.bioontology.org/ontologies/KISAO>.