

## **MAMMOTH: a new database for curated MAThematical Models of bioMOlecular sysTems**

Fedor Kazantsev<sup>1,2\*</sup>, Ilya Akberdin<sup>1,5</sup>, Sergey Lashin<sup>1,2</sup>, Natalia Ree<sup>1</sup>, Vladimir Timonov<sup>2</sup>, Alexander Ratushny<sup>3,4</sup>, Tamara Khlebodarova<sup>1</sup>, Vitaly Likhoshvai<sup>\*1,2</sup>

<sup>1</sup>Institute of Cytology and Genetics SB RAS, Lavrentyev Avenue., 10, Novosibirsk, 630090, Russia,

<sup>2</sup>Novosibirsk State University, Pirogova str. 2, Novosibirsk, 630090, Russia, <sup>3</sup>Center for Infectious Disease Research (formerly Seattle Biomedical Research Institute), Seattle, WA 98109, USA,

<sup>4</sup>Institute for Systems Biology, Seattle, WA 98109-5234, USA, <sup>5</sup>San Diego State University, San Diego, CA 92182-4614, USA

\*Corresponding author: Tel: +7 (383) 363-49-63\*1331; Email: kazfdr@bionet.nsc.ru

\*Correspondence about models may also be addressed to Vitaly Likhoshvai. Tel: +7 (383) 363-49-63\*3102; Email: likho@bionet.nsc.ru

### **ABSTRACT**

Living systems have a complex hierarchical organization that can be viewed as a set of dynamically interacting subsystems. Thus, to simulate the internal nature and dynamics of the whole biological system we should use the iterative way for a model reconstruction, which is a consistent composition and combination of its elementary subsystems. In accordance with this bottom-up approach, we have developed MAMMOTH (MAThematical Models of bioMOlecular sysTems) database that allows integrating manually curated mathematical models of biomolecular systems, which are fit to the experimental data. The database entries are organized as building blocks in a way that the model parts can be used in different combinations to describe systems with higher organizational level (metabolic pathways and/or transcription regulatory networks). The database supports export of single model or their combinations in SBML or Mathematica standards. The database currently contains more than 100 mathematical models for *Escherichia coli* elementary subsystems (enzymatic reactions and gene expression regulatory processes) that can be combined in at least 5100 complex/sophisticated models concerning such biological processes as: *de novo* nucleotide biosynthesis, aerobic/anaerobic respiration, and nitrate/nitrite utilization in *E. coli*. All current models are functionally interconnected and sufficiently complement public model resources.

**Database URL:** <http://mammoth.biomodelsgroup.ru>

**Keywords:**

Mathematical models; generalized Hill functions; respiration; *de novo* nucleotide synthesis; gene network, gene expression.

## INTRODUCTION

Recent developments in experimental technologies have revolutionized biology and allowed researchers to generate huge amounts of highly resolved multiscale quantitative data in their respective biological fields. As a result, it caused a renaissance of the mathematical modeling of biological systems (1–4). Nowadays, mathematical modeling approaches are more often considered as general frameworks for the integration and analysis of experimental data and iterative investigation of dynamical biological systems (5–15). The general type of the model is determined on basis of the available information, used qualitative or quantitative data and the problem to solve. The assigned determinants for system behaviour define the required model variables. Mathematical models of dynamical biological systems can be formulated as systems of ordinary, partial, or delayed differential equations (ODE, PDE, DDE, respectively), stochastic equations, discrete operations or by hybrid models (4,13–23). The choice of the formalism usually depends on the complexity of biological system in question and availability of experimental data.

Currently, there are more than 1000 different published models of complex biomolecular systems (BMS), which are developed for several organisms (24–28). The number of such models is continuously increasing. However, reuse, expansion, or modification of the developed models is a labor-intensive process (29–32). Usually the models are designed as complete integrated systems and there are no tools available for researchers to modify their structural and/or functional contents in an automatic or semiautomatic way (31). The decomposition of the models into their elementary subsystems is an alternative approach to solve this problem. Such elementary subsystems can be viewed as “building blocks” and can be used to describe certain biomolecular functions and/or elementary processes of biological systems (14,33–36). The “functional network” represents a minimal model of the subsystem that is sufficient to perform a given function (37).

We have developed the MAMMOTH database which is designed to store mathematical models of an arbitrary biomolecular subsystem using the above-mentioned building blocks approach. The current version of the database contains more than 100 mathematical models of the *E.coli* biomolecular “elementary” subsystems, developed by authors of the present study using the original approach of generalized Hill’s functions (21). It should be mentioned that these “elementary” models then can be combined in various composite “complex” models, the total amount of which may be estimated at least to 5100 (which are consists of two and more connected "elementary" models). The

models curated and stored in MAMMOTH are not presented in Biomodels.net (<http://www.ebi.ac.uk/biomodels-main/>) or Sabio-RK (<http://sabio.villa-bosch.de>) databases. MAMMOTH enables an automatic creation of mathematical models with different combinations of elementary subsystem. This database also allows export of the created models into different model standards to share them with other researchers and/or to integrate them in more complex models.

## RESULTS

### MAMMOTH database

The MAMMOTH database is designed for storage and accumulation of *elementary subsystem models* of biomolecular systems. *Elementary subsystem model* is the “functional network” model where the entities of the network and the mathematical model parameters are specified. The dynamics of each *elementary subsystem model* is related to published experimental data, which are referenced in the corresponding “model description” field. The database currently contains more than 100 mathematical models of BMS elementary subsystems: enzymatic reactions and gene expression regulation processes of *de novo* nucleotides biosynthesis, aerobic/anaerobic respiration and nitrate/nitrite utilization processes in *E.coli* cell. Every model consists of three parts: (1) the structural model (Figure 1a), which describes entities of the process, stoichiometric parameters and links on the public sources (EcoCyc, KEGG, PubChem); (2) the mathematical model (Figure 1b), which describes rate equation; (3) the kinetic parameter values (Figure 1c). Each mathematical model in the database has the description field where one can find references on articles which were the basis for the model development or which contain a quantitative experimental data used for model parameters fitting. The mathematical models were manually curated and verified by expert authors. The majority of the models were reconstructed using generalized Hill functions (21). These functions enable flexible formulation of model equations by solving structural and parametric inverse problems optimally aligning the model complexity with the level of detail in the relevant experimental data.

To check the uniqueness of mathematical models in MAMMOTH database we have compared model variables (e.g., metabolites, enzymes) with model variables from other open access resources, such as “Sabio-RK” (38), “Biomodels.net” (39), CellML Model Repository (26) and “JWS online” (24). For this purpose we have used either APIs of the corresponding services or the manual data extraction (JWS online). PubMed IDs were used in the intersection search between articles that used for the model development or that contained experimental data for the model parameters fitting. We found two articles, which are also used in Sabio-RK to describe kinetic parameters (Pubmed: 1659321, 10074342; MAMMOTH links are [GN0000093; MM0000129] and [GN0000185; MM0000194], respectively).

To compare with data in *Biomodels.net* repository, we used the two-level procedure as follows. First, we used Pubmed ID for articles, which describe models or experimental data exploited for model adaptation. Then we compared database contents using Enzyme Commission numbers (EC number). At the first comparison level we have found three reference intersections only, which were articles with description of experimental data: 1) This case is PubMed ID:19561621 (40), where several metabolites concentration and enzymatic activities data for the *E.coli* cell were used for the model of central metabolic pathway in the Biomodels.net (BIOMD0000000244) (41) under the investigation of bacterial cell adaptation to environment conditions. But the same data was used for another model in the MAMMOTH (GN0000141) to study the mechanism of the enzymatic reaction, which is catalyzed by NADH dependent nitrate reductase (EC:1.7.1.4); 2) This case is PubMed ID:9202467 (42) where *cAMP* concentration in dependence on glucose concentration was used for reconstruction and adaptation of central cell metabolic pathway model in Biomodels.net (BIOMD0000000051) (17) while the same data was used for exploration *guaA* gene expression regulation in the MAMMOTH (GN0000236); 3) the last case is PubMed ID:2200929 (43) which is the review article dedicated to metabolite and enzymes concentrations in cell of the *E. coli*, yeast, *Dictyostelium discoideum* and others. The data were used for model reconstruction of enzymatic reaction, catalyzed by phosphofructokinase in the pancreatic beta-cells in the Biomodels.net (BIOMD0000000225, BIOMD0000000236) (44). At the same time the quantitative data were used for exploration of N-carbamoyl-L-aspartate biosynthesis in the MAMMOTH model (GN0000183). Thus, the first step of the comparative analysis has shown that there are only three intersections between Biomodels.net and MaMoNT databases on experimental data, which were used for reconstruction of different biological processes.

The comparative analysis performed with the EC numbers at the second step gave us 11 general cases. The majority of the cases in Biomodels.net are associated with another organisms (mouse, human) (45), another cell type (46) or associated with the pipelines/techniques for the automation of model reconstruction process (47) (the experimental data were not mentioned in them).

Intersections in references on articles between MAMMOTH database and both JWS online and CellML systems were not found. Eventually, we have presented approximately 100 mathematical models of BMS elementary subsystems which were adapted to experimental data and were not published before in articles (besides of own reviewed abstracts in the conference proceedings) (Figure. 2).

### **MAMMOTH software architecture**

The MAMMOTH has been developed as Vaadin (vaadin.com) application with PostgreSQL database. It provides an access to the repository of elementary subsystems models by means of search

for name of substances involved in the inquired subsystem, IDs of available kinetic models describing the process and IDs of the corresponding parameter sets. The database has been developed using the hybrid approach between relation and XML ones. The relational database provides fast access to models but the BMS subsystems are better to keep as XML, due to its benefits of the structural organization. Combination of approaches gave us the flexible database structure and simple data search requests (Figure 3).

The data storage system enables us to utilize additional features. Each model in the database has been developed on basis of the general entities vocabulary. It makes possible to compare entities in the different models by identifiers directly. Using this property, the MAMMOTH provides some functionality for manipulation with models. To follow the “building blocks” strategy for complex models reconstruction we have implemented a method of template models generation reconstructed by the automatic composition from a subset of elementary models which are presented in the database. These assembled models can be exported in such well known formats as SBML, MATHEMATICA and Pajek (48) for further analysis in other tools.

#### *Access and search routine*

The web-interface of the MAMMOTH enables the user to search for models by names and types (gene, RNA, protein and substance) of entities (Figure 1a). As a result of the search the system generates a list of structural models, the components of which have a match in at least one of synonyms. When the user selects a structural model from the resulting list, a window opens containing lists of substrates, products, regulators of the elementary subsystems with relevant links to other database as well as a set of mathematical models. Rate reaction functions are visualized as mathematical formulas. Each model has an information field with the model description and links to relevant publications as well as a list of parameter sets in the section describing the mathematical model. A several sets of parameter values can be stored for the individual model in the database.

#### *Models integration routine*

The certain set of parameters can be selected by clicking the «select model» button which in turn allows the user to transfer mathematical model in the section «selected models» in upper right corner of the application window (Figure 1d). The user compiles a list of elementary models by the scheme described above. Additionally, the graph view of whole complex model is available too (Figure 4).

#### *Model export*

After completion of the selection procedure the user can reconstruct a mathematical model from elementary models by selecting the appropriate format and clicking the «generate» button. Models are automatically assembled by the formation of global rates of component concentration (model variable) changes. The global rate represents a sum of the rates of the model variable changes in the selected

elementary models. The MAMMOTH supports export of the BMS elementary subsystem models or models bunch, integrated into the complex one model, to the well-known standards: SBML, Mathematica and Pajek (48) (Figure 4). Due to the fact that each model in the database was adapted to the experimental data, further investigation will take less time to solve the parameters fitting (inverse) problem for the exported integrative model. The models bunch integrated into the one model could be a «growth point» in the process of complex models development and fitting or it can be just a part for more extended model(s) being developed in any other tool. We erase the support model information during the export to ensure non-conflicts with the selected standard.

## DISCUSSION

The hierarchical organization of biomolecular systems and elaborated “building blocks” approach enable systems biologists to iteratively develop integrative models by means of consistent composition of elementary models (Figure 5) (13,49–54). The elementary subsystems in the MAMMOTH database can be used in mathematical models of different biological systems.

However, a purely mathematical approach can be more often found to the study of conserved regulatory circuits - “functional networks”, which represent the minimal models of subsystems sufficient to perform a given function. These functional networks, aggregated in the database, (37) are extremely useful in the BMS model reconstruction process. These networks can be used to build preliminary “scratch” models, but the models obtained are rather mathematical abstractions. They describe interactions of biological entities, their regulatory effects that lead to the appropriate dynamics in the selected parameters constraints. But the model dynamics is sensitive to the variation of parameters values (30,55). Only the use of a physiologically appropriate parameters of the model allows researchers to reproduce the experimentally observed dynamics of the considered living system (30,55).

As an example of the “building blocks” approach, the most cited *Biomodels.net* database (56) was extended by means of 140 000 models of BMS subsystems models automatically generated by *Path2models* tool (57). The generation was based on structural and functional descriptions which were extracted from *KEGG* database (58). Meanwhile, less than 600 mathematical models in the *Biomodels.net* database are associated with papers and curated by experts. It means that together with the databases of “functional networks” abstract models (37,51,57,59–61) the databases of adapted models describing the dynamical patterns of biological subsystems on different hierarchical levels should be developed. In the future, it will allow us to replace the template models on fitted ones in the repositories and will give the opportunity for a full-scale reconstruction of models for different types of cells, tissues or even whole organisms. Karr and coauthors have already implemented this philosophy in the whole genome *Mycoplasma genitalium* model (4). It led to the development of the

*WholeCellKB* (<http://wholecellkb.stanford.edu>) database which contains the description of functional cell subsystems.

## CONCLUSIONS

We have presented the MAMMOTH database of the kinetic models where the elementary models (enzymatic reaction or the gene expression regulatory process) and integrated complex models (metabolic pathways) can be exported in the well-known standards for model analysis: SBML, Mathematica and Pajek (48). Today the database contains more than 100 BMS elementary subsystems mathematical models of the enzymatic reactions describing *de novo* nucleotide biosynthesis, aerobic/anaerobic respiration and nitrate/nitrite utilization in *Escherichia coli*, which have not been presented in public BMS mathematical models resources.

Each model in the database is fitted to experimental data (data was taken from over 140 sources) and manually curated by experts in systems biology. The developed elementary models have been used for reconstruction of a several more complex BMS models (62–68).

The MAMMOTH repository has a friendly user interface providing the manual models query and the REST API that can be used for programmatic data access and the integration with external software tools. Currently, there are a few tools which give the opportunity to generate the template mathematical models on basis of the structural description of simulated biochemical systems (57,69–74). We strongly believe that the developed database carrying the fitted mathematical models of BMS elementary subsystems will impact on and improve the reliability of the model automatic generation process.

## AVAILABILITY

The MAMMOTH is freely available for the academic use: <http://mammoth.biomodsgroup.ru>.  
REST API: <http://mammoth.biomodsgroup.ru/api>

## LIST OF ABBREVIATIONS

DDE – Delay Differential Equation.

BMS – Biomolecular Systems.

MAMMOTH – MATHematical Models of bioMOlecular sysTEms.

ODE – Ordinary Differential Equations.

PDE – Partial Differential Equations.

## COMPETING INTERESTS

authors declare that they have no competing interests.



## AUTHORS' CONTRIBUTIONS

FK has developed the MAMMOTH software including database and GUI levels. VT has developed the MAMMOTH GUI system. IA, NR, SL and AR have developed series of mathematical models, performed their analysis, checking and inserting it into the database. VL and TK performed analysis of the models, VL and AR supervised the development and implementation of the models. FK, IA, SL and AR wrote the manuscript with editing by TK and VL. All authors read and approved the final manuscript.

## FUNDING

Budget project No. 0324-2015-0003 and the RFBR grant No. 15-07-03879. AVR was supported by grants P50 GM076547 and P41 GM109824 from the National Institutes of Health.

## REFERENCES

1. Kitano, H. (2002) Systems biology: a brief overview., *Science*, **295**, 1662–4.
2. Westerhoff, H. V and Palsson, B. O. (2004) The evolution of molecular biology into systems biology, *Nat. Biotechnol.*, **22**, 1249–1252.
3. Najafi, A., Bidkhorji, G., Bozorgmehr, J. H., et al. (2014) Genome scale modeling in systems biology: algorithms and resources., *Curr. Genomics*, **15**, 130–59.
4. Karr, J. R., Sanghvi, J. C., Macklin, D. N., et al. (2012) A whole-cell computational model predicts phenotype from genotype., *Cell*, **150**, 389–401.
5. Akberdin, I. R., Kazantsev, F. V., Ermak, T. V., et al. (2013) In Silico Cell: Challenges and Perspectives, *Math. Biol. Bioinf.*, **8**, 295–315.
6. Hübner, K., Sahle, S. and Kummer, U. (2011) Applications and trends in systems biology in biochemistry., *FEBS J.*, **278**, 2767–2857.
7. Mast, F. D., Ratushny, a. V. and Aitchison, J. D. (2014) Systems cell biology, *J. Cell Biol.*, **206**, 695–706.
8. Csikász-Nagy, A. (2009) Computational systems biology of the cell cycle., *Brief. Bioinform.*, **10**, 424–434.
9. Demin, O. and Goryanin, I. *Kinetic Modelling in Systems Biology*; 2008.
10. Mironova, V. V., Omelyanchuk, N. A., Yosiphon, G., et al. (2010) A plausible mechanism for auxin patterning along the developing root., *BMC Syst. Biol.*, **4**, 98–117.
11. Sanchez-Osorio, I., Ramos, F., Mayorga, P., et al. (2014) Foundations for modeling the dynamics of gene regulatory networks: a multilevel-perspective review., *J. Bioinform. Comput. Biol.*, **12**, 1330003.



12. Mori, H. (2004) From the Sequence to Cell Modeling: Comprehensive Functional Genomics in Escherichia coli, *J. Biochem. Mol. Biol.*, **37**, 83–92.
13. Demidenko, G. V. and Likhoshvai, V. A. (2005) On Differential Equations with Retarded Argument, *Sib. Math. J.*, **46**, 417–430.
14. Likhoshvai, V. A., Matushkin, I. G., Ratushny, A. V, et al. (2001) A generalized chemical-kinetic method for modeling gene networks, *Mol. biol.*, **35**, 1072–1079.
15. Danziger, S. A., Ratushny, A. V., Smith, J. J., et al. (2014) Molecular mechanisms of system responses to novel stimuli are predictable from public data., *Nucleic Acids Res.*, **42**, 1442–1460.
16. Feist, A. M., Henry, C. S., Reed, J. L., et al. (2007) A genome-scale metabolic reconstruction for Escherichia coli K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information., *Mol. Syst. Biol.*, **3**.
17. Chassagnole, C., Noisommit-Rizzi, N., Schmid, J. W., et al. (2002) Dynamic modeling of the central carbon metabolism of Escherichia coli, *Biotechnol. Bioeng.*, **79**, 53–73.
18. Eungdamrong, N. J. and Iyengar, R. (2004) Computational approaches for modeling regulatory cellular networks., *Trends Cell Biol.*, **14**, 661–669.
19. Chen, W. W., Niepel, M. and Sorger, P. K. (2010) Classic and contemporary approaches to modeling biochemical reactions., *Genes Dev.*, **24**, 1861–1875.
20. Lashin, S. A., Suslov, V. V. and Matushkin, Y. G. (2010) Comparative modeling of coevolution in communities of unicellular organisms: adaptability and biodiversity, *J. Bioinform. Comput. Biol.*, **08**, 627–643.
21. Likhoshvai, V. A. and Ratushny, A. V (2007) Generalized Hill function method for modeling molecular processes, *J. Bioinform. Comput. Biol.*, **05**, 521–531.
22. Ocone, A., Millar, A. J. and Sanguinetti, G. (2013) Hybrid regulatory models: a statistically tractable approach to model regulatory network dynamics., *Bioinformatics*, **29**, 910–6.
23. Shapiro, B. E. and Mjolsness, E. (2015) Pycellerator: An arrow-based reaction-like modelling language for biological simulations, *Bioinformatics*, **9**, btv596.
24. Olivier, B. G. and Snoep, J. L. (2004) Web-based kinetic modelling using JWS Online., *Bioinformatics*, **20**, 2143–2144.
25. Funahashi, A., Matsuoka, Y., Jouraku, A., et al. (2008) CellDesigner 3.5: A versatile modeling tool for biochemical networks, *Proc. IEEE*, **96**, 1254–1265.
26. Lloyd, C. M., Lawson, J. R., Hunter, P. J., et al. (2008) The CellML Model Repository., *Bioinformatics*, **24**, 2122–3.
27. Chelliah, V., Juty, N., Ajmera, I., et al. (2014) BioModels: ten-year anniversary., *Nucleic*

*Acids Res.*, 1–7.

28. King, Z. A., Lu, J., Dräger, A., et al. (2015) BiGG Models: A platform for integrating, standardizing and sharing genome-scale models, *Nucleic Acids Res.*, gkv1049.
29. Stanford, N. J., Wolstencroft, K., Golebiewski, M., et al. (2015) The evolution of standards and data management practices in systems biology, *Mol. Syst. Biol.*, **11**, 851–851.
30. Kirk, P. D. W., Babbie, A. C. and Stumpf, M. P. H. (2015) Systems biology (un)certainties, *Science (80-. )*, **350**, 386–388.
31. Scharm, M., Wolkenhauer, O. and Waltemath, D. (2014) An algorithm to detect and communicate the differences in computational models describing biological systems.
32. Henkel, R., Wolkenhauer, O. and Waltemath, D. (2015) Combining computational models, semantic annotations and simulation experiments in a graph database, *Database*, **2015**, bau130.
33. Palsson, B. (2000) The challenges of in silico biology Moving from a reductionist paradigm to one that views cells as systems will necessitate, **18**, 1147–1150.
34. Guido, N. J., Wang, X., Adalsteinsson, D., et al. (2006) A bottom-up approach to gene regulation., *Nature*, **439**, 856–60.
35. Bruggeman, F. J. and Westerhoff, H. V (2007) The nature of systems biology., *Trends Microbiol.*, **15**, 45–50.
36. Lopez, C. F., Muhlich, J. L., Bachman, J. A., et al. (2014) Programming biological models in Python using PySB, *Mol. Syst. Biol.*, **9**, 646–646.
37. Kurata, H., Maeda, K., Onaka, T., et al. (2014) BioFNet: biological functional network database for analysis and synthesis of biological systems., *Brief. Bioinform.*, **15**, 699–709.
38. Wittig, U., Golebiewski, M., Kania, R., et al. (2006) SABIO-RK□: Integration and Curation of Reaction Kinetics Data, 94–103.
39. Le Novère, N., Bornstein, B., Broicher, A., et al. (2006) BioModels Database: a free, centralized database of curated, published, quantitative kinetic models of biochemical and cellular systems., *Nucleic Acids Res.*, **34**, D689–D691.
40. Bennett, B., Kimball, E. and Gao, M. (2009) Absolute metabolite concentrations and implied enzyme active site occupancy in Escherichia coli, *Nat. Chem Biol*, **5**, 593–599.
41. Kotte, O., Zaugg, J. B. and Heinemann, M. (2010) Bacterial adaptation through distributed sensing of metabolic fluxes., *Mol. Syst. Biol.*, **6**, 355.
42. Notley-mcrob, L., Death, A. and Ferenci, T. (1997) The relationship between external glucose concentration and cAMP levels inside, *Microbiology*, **143**, 1909–1918.
43. Albe, K. R., Butler, M. H. and Wright, B. E. (1990) Cellular concentrations of enzymes

- and their substrates., *J. Theor. Biol.*, **143**, 163–195.
44. Westermark, P. O. and Lansner, A. (2003) A model of phosphofructokinase and glycolytic oscillations in the pancreatic beta-cell., *Biophys. J.*, **85**, 126–139.
  45. Curto, R., Voit, E. O., Sorribas, a, et al. (1998) Mathematical models of purine metabolism in man., *Math. Biosci.*, **151**, 1–49.
  46. Jiang, N., Cox, R. D. and Hancock, J. M. (2007) A kinetic core model of the glucose-stimulated insulin secretion network of pancreatic beta-cells, *Mamm. Genome*, **18**, 508–520.
  47. Smallbone, K. and Mendes, P. (2013) Large-Scale Metabolic Models: From Reconstruction to Differential Equations, *Ind. Biotechnol.*, **9**, 179–184.
  48. de Nooy, W., Mrvar, A. and Batagelj, V. *Exploratory Social Network Analysis with Pajek*; 2nd Editio.; Cambridge University Press: New York, USA, 2012.
  49. Likhoshvai, V. A., Kogai, V. V. and Fadeev, S. I. In *BGRS2006 proceedings*; 2006; pp. 391–404.
  50. Alon, U. (2003) Biological networks: the tinkerer as an engineer., *Science*, **301**, 1866–1867.
  51. Milo, R., Shen-Orr, S., Itzkovitz, S., et al. (2002) Network motifs: simple building blocks of complex networks., *Science*, **298**, 824–827.
  52. Danino, T., Mondragón-Palomino, O., Tsimring, L., et al. (2010) A synchronized quorum of genetic clocks., *Nature*, **463**, 326–330.
  53. Ratushny, A. V., Ramsey, S. A. and Aitchison, J. D. In *Methods in Molecular Biology*; 2011; Vol. 781, pp. 415–433.
  54. Palsson, B. *Systems Biology*; Cambridge University Press, 2015.
  55. Lim, W. A., Lee, C. M. and Tang, C. (2013) Design principles of regulatory networks: searching for the molecular algorithms of the cell., *Mol. Cell*, **49**, 202–212.
  56. Chelliah, V., Juty, N., Ajmera, I., et al. (2014) BioModels: ten-year anniversary., *Nucleic Acids Res.*, **43**, 542–548.
  57. Büchel, F., Rodriguez, N., Swainston, N., et al. (2013) Path2Models: large-scale generation of computational models from biochemical pathway maps., *BMC Syst. Biol.*, **7**, 116.
  58. Kanehisa, M. (2000) KEGG: Kyoto Encyclopedia of Genes and Genomes, *Nucleic Acids Res.*, **28**, 27–30.
  59. Pieper, U., Eswar, N., Davis, F. P., et al. (2006) MODBASE: a database of annotated comparative protein structure models and associated resources., *Nucleic Acids Res.*, **34**, D291–D295.

60. Shen-Orr, S. S., Milo, R., Mangan, S., et al. (2002) Network motifs in the transcriptional regulation network of *Escherichia coli*., *Nat. Genet.*, **31**, 64–68.
61. Fraser, J. S., Gross, J. D. and Krogan, N. J. (2013) From systems to structure: bridging networks and mechanism., *Mol. Cell*, **49**, 222–31.
62. Likhoshvai, V. A., Khlebodarova, T. M., Ree, M. T., et al. (2010) Metabolic engineering in silico, *Appl. Biochem. Microbiol.*, **46**, 671–687.
63. Oshchepkova-Nedosekina, E. A. and Likhoshvai, V. A. (2007) A mathematical model for the adenylosuccinate synthetase reaction involved in purine biosynthesis., *Theor. Biol. Med. Model.*, **4**, 11.
64. Nedosekina, E. A. and Ratushny, A. V. (2006) Regulation of pyrimidine biosynthesis in *Escherichia coli*: gene network reconstruction and mathematical modeling, *Proc. fifth Int. Conf. Bioinforma. genome Regul. Struct.*, **2**, 35–39.
65. Nedosekina, E. A. (2006) Mathematical modeling of regulation of *Escherichia coli* purine biosynthesis pathway enzymatic reactions, *Proc. fifth Int. Conf. Bioinforma. genome Regul. Struct.*, **2**, 60–67.
66. Khlebodarova, T. M., Lashin, S. A. and Apasieva, N. V. (2006) Gene network reconstruction and mathematical modeling of *E. coli* respiration: regulation of *f0f1*-atp synthase by metal ions, *Proc. fifth Int. Conf. Bioinforma. genome Regul. Struct.*, **2**, 55–59.
67. Smirnova, O. G. and Ratushny, A. V. (2006) Gene network reconstruction and mathematical modeling of salvage pathways: regulation of adenine phosphoribosyltransferase activity by structurally similar substrates, *Proc. fifth Int. Conf. Bioinforma. genome Regul. Struct.*, **2**, 73–77.
68. Khlebodarova, T. M., Ree, N. A. and Likhoshvai, V. A. (2016) On the control mechanisms of the nitrite level in *Escherichia coli* cells: the mathematical model, *BMC Microbiol.*, **16**, 15–30.
69. König, M., Dräger, A. and Holzhütter, H.-G. (2012) CySBML: a Cytoscape plugin for SBML., *Bioinformatics*, **28**, 2402–3.
70. Dräger, A., Hassis, N., Supper, J., et al. (2008) SBMLsqueezer: a CellDesigner plug-in to generate kinetic rate equations for biochemical networks., *BMC Syst. Biol.*, **2**, 39.
71. Dräger, A., Zielinski, D. C., Keller, R., et al. (2015) SBMLsqueezer 2: context-sensitive creation of kinetic equations in biochemical networks, *BMC Syst. Biol.*, **9**, 68.
72. Hu, B., Matthew Fricke, G., Faeder, J. R., et al. (2009) GetBonNie for building, analyzing and sharing rule-based models., *Bioinformatics*, **25**, 1457–1460.
73. Faeder, J. R., Blinov, M. L. and Hlavacek, W. S. (2009) Rule-based modeling of biochemical systems with BioNetGen., *Methods Mol. Biol.*, **500**, 113–167.

74. Rodriguez, N., Pettit, J.-B., Dalle Pezze, P., et al. (2016) The systems biology format converter, *BMC Bioinformatics*, **17**, 154–161.

## Figures

Figure 1. The MAMMOTH interface. (a) The query box. (b) The biochemical preview of the BMS elementary subsystem with the information about participants and model itself. (c) The mathematical model description of the selected subsystem. There are mathematical equation that you can download in different standards, model parameters and their values. (d) The area of available parameters value sets. The «select model» button is for moving of the selected model with certain parameters values into the «selected models» list. Each part has the «information» tab that contains additional information for selected entity as well as links to external sources.

Figure 2. The quantity of mathematical models that are in open access and validated by experts.

Figure 3. The MAMMOTH database scheme.

Figure 4. The interface for selected models. The “Generate” button activates the export of the complex model into the selected standard: SBML, Mathematica or Pajek.

Figure 5. The scheme of model reconstruction process in terms of “building blocks”.

## Structural motif search

Name to search

Search

All objects

scheme\_name

CTP + a reduced flavodoxin <=> pCTP + an O<sub>2</sub>dCTP + H<sub>2</sub>O -> NH<sub>2</sub> + dUTP

-&gt; NRA

UTP + a reduced flavodoxin <=> dUTP + H<sub>2</sub>Ocaasp + H+ <=> dcaasp + H<sub>2</sub>O

## Structural model description: [GF000184]

caasp + H+ <=> dcaasp + H<sub>2</sub>O

Model consists of:

	Alias	Short name	Function	SD
	s1	caasp	substrate	520000000
	s2	H+	substrate	520000000
	p1	dcaasp	product	520000000
	p2	H <sub>2</sub> O	product	520000000
	r1	PyC	actuator	520000000

Information (press to expand)

## Mathematical models

MM000193

$$\frac{[CTP]^2}{([caasp][CTP] + k_{cat}[CTP])} \cdot \frac{1}{(K_m caasp + [caasp])}$$

Data:

Name	Value	Units	Description
k <sub>dis</sub>	0.001	mM	Dissociation constant
k <sub>cat</sub>	200	1/sec	Catalytic constant
K <sub>m</sub> caasp	0.47	mM	Michaelis constant with respect to carbamoyl aspartate

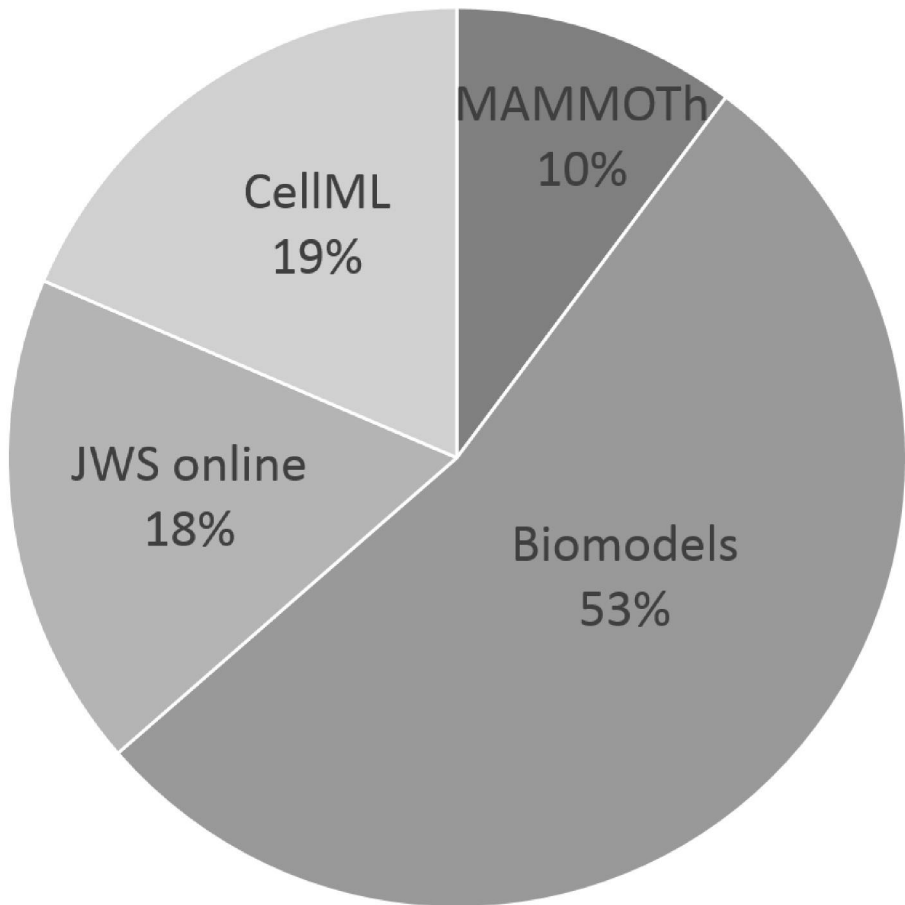
Information (press to expand)

Select model

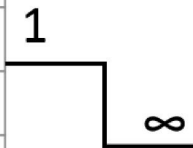
Parameters set

PS000183





Structural model	
	GN_ID
	Scheme "a + b -> c"
	Process type
XML	GN structure



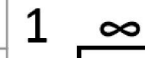
Mathematical model	
	MM_ID
	GN_ID
	Velocity equation
XML	Parameters



Parameter set	
	PS_ID
	MM_ID
XML	parameter values

Information	
	ID
	Source
	Text
XML	Article links
	Authors

Substance	
	ID
	Name
	Type (gene, protein, ...)
XML	Reference links



Substance synonym	
	ID
	Substance
	Synonym

Elementary subsystem	ID	Action
UTP + a reduced flavodoxin $\Leftrightarrow$ dUTP + H <sub>2</sub> O + an oxidized flavodoxin	GN0000030_MM0000114_PS0000	<a href="#">remove</a>
UMP + ATP $\Leftrightarrow$ UDP + ADP + H <sup>+</sup>	GN0000064_MM0000125_PS0000	<a href="#">remove</a>
dUDP + ATP $\rightarrow$ dUTP + ADP + H <sup>+</sup>	GN0000067_MM0000130_PS0000	<a href="#">remove</a>
CTP + a reduced flavodoxin $\Leftrightarrow$ dCTP + an oxidized flavodoxin	GN0000068_MM0000113_PS0000	<a href="#">remove</a>
UDP + ATP $\rightarrow$ UTP + ADP + H <sup>+</sup>	GN0000078_MM0000126_PS0000	<a href="#">remove</a>
dUDP + ATP $\rightarrow$ TTP + ADP + H <sup>+</sup>	GN0000079_MM0000132_PS0000	<a href="#">remove</a>
UTP + ATP + glu + H <sub>2</sub> O $\rightarrow$ CTP + ADP + Pi + glu + H <sup>+</sup>	GN0000081_MM0000170_PS0000	<a href="#">remove</a>
CTP + ADP + H <sup>+</sup> $\Leftrightarrow$ CDP + ATP	GN0000086_MM0000171_PS0000	<a href="#">remove</a>
dCDP + ATP $\rightarrow$ dCTP + ADP + H <sup>+</sup>	GN0000093_MM0000129_PS0000	<a href="#">remove</a>



Export models as:

theModel

SBML

-

Generate

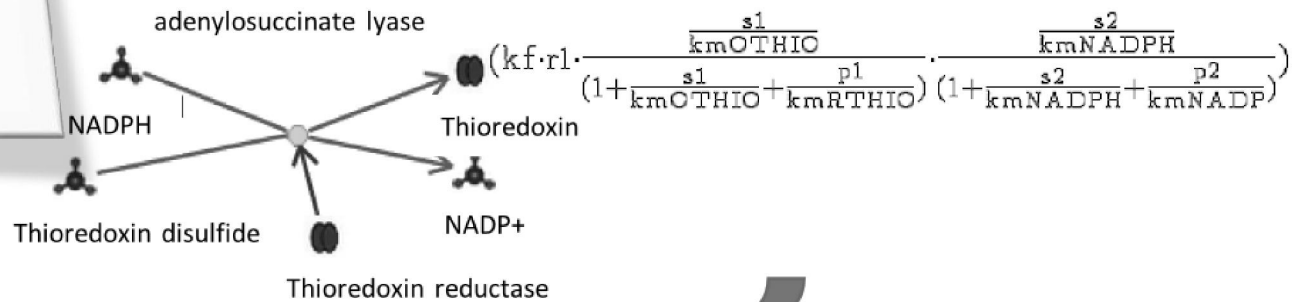
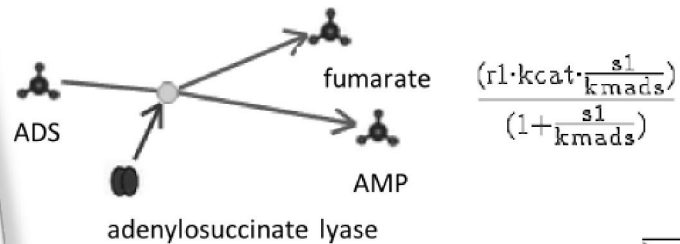
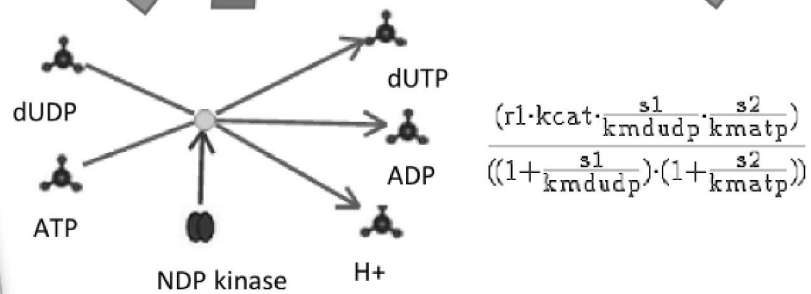
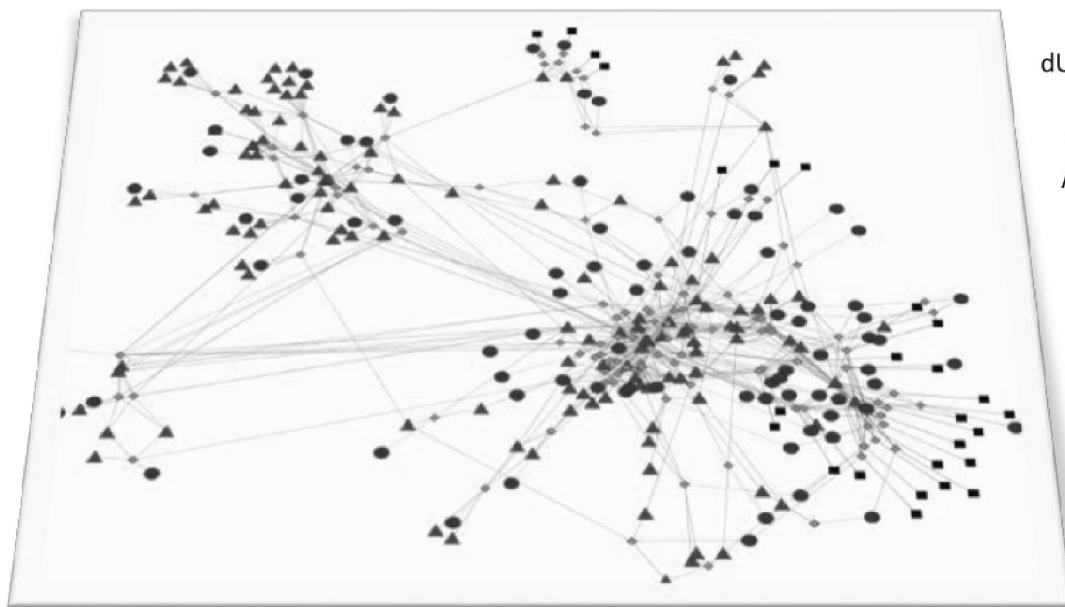
SBML

Mathematics

Pajek

decomposition

reconstruction



Model compilation