

Curation and exchange of kinetic models of biochemical pathways

Nicolas Le Novère EMBL-EBI, United-Kingdom

BioPAX F2F meeting, Tokyo, November 18th 2005









∰-™ Systems Biology Markup Language - Diagram of EPSP_AChDynValue_SBML2.xml - Mozilla 🗖 🛂 🔀 No, it isn't В React0 React5 BLA Reactl React6 React3 React9 BLL ALReact2 React10 React14 React4 React7 ALL D \mathbf{IL} Reactll React8 React15 React12 -III: http://sbml.org/documents

BIOMODELS.NET

Home | Contacts | FAQs | Database | MIRIAM | SBO | Forums | News

The Next Step After Standard Formats

For computational modeling to become more widely used in biological research, researchers must be able to exchange and share their results. The development and broad acceptance of common model representation formats such as **SBML** is a crucial step in that direction, allowing researchers to exchange and build upon each other's work with greater ease and accuracy.

The BioModels net project is another step: an international effort to (1) define agreed-upon standards for model curation, (2) define agreed-upon vocabularies for annotating models with connections to biological data resources, and (3) provide a free, centralized, publicly-accessible database of annotated, computational models in SBML and other structured formats.

Helping to Define Community Standards

To facilitate assembling useful collections of quantitative models of biological phenomena, it is crucial to establish standards for the vocabularies used in model annotations as well as criteria for minimum quality levels of those models. The BioModels.net project aims to bring together a community of interested researchers to address these issues. We are working towards defining these standards through white papers and process definitions. All of the products of our efforts are open and freely available through this site.

Standards and Processes Developed Hand-in-Hand with a New Database

The database component of BioModels net is especially designed for working with annotated computational models: each model is carefully reviewed and augmented by human annotators on the BioModels net team to add metadata linking the model elements to other biological databases and resources. The **BioModels database at the EBI** system goes far beyond other collections of models by being a *true* database, featuring browsing, cross-referencing, searching, and (coming soon) facilities for visualization, exporting models in different formats, and remote API access.

Search

RSS / Atom

-III:



Minimum Information Requested In the Annotation of biochemical Models (MIRIAM)

Le Novère N.*, Finney A.*, Hucka M., Bhalla U., Campagne F., Collado-Vides J., Crampin E., Halstead M., Klipp E., Mendes P., Nielsen P., Sauro H., Shapiro B., Snoep J.L., Spence H.D., Wanner B.L.

Nature Biotechnology (2005), 23: in the press





Reference correspondence

- The model must be encoded in a public, standardized, machinereadable format (SBML, CellML, GENESIS ...)
- The model must comply with the standard in which it is encoded!
- The model must be clearly related to a single reference description. If a model is composed from different parts, there should still be a description of the derived/combined model.
- The encoded model structure must reflect the biological processes listed in the reference description.
- The model must be instantiated in a simulation: All quantitative attributes have to be defined, including initial conditions.
- When instantiated, the model must be able to reproduce all results given in the reference description within an epsilon (algorithms, round-up errors)





Attribution annotation

- The model has to be named.
- A citation of the reference description must be joined (complete citation, unique identifier, unambigous URL). The citation should permit to identify the *authors* of the model.
- · The name and contact of model *creators* must be joined.
- The date and time of creation and last modification should be specified. An history is useful but not required.
- The model should be linked to a precise statement about the terms of distribution. MIRIAM does <u>not</u> require "freedom of use" or "no cost".



EMBL-EBI

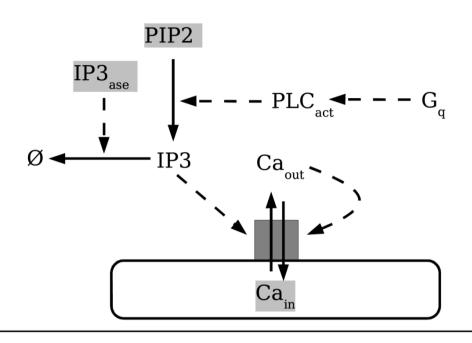
External resource annotation

- The annotation must permit to unambiguously relate a piece of knowledge to a model constituent.
- The referenced information should be described using a triplet {data-type, identifier, qualifier}
 - The data-type should be written as a Unique Resource Identifier (URI). Either a URL (webpage) or a URN (e.g. LSID). Not necessarily a physical location.
 - The identifier is analysed by the software within the framework of the datatype.
 - Data-type and Identifier can be combined in a single URI http://www.myResource.org/#myIdentifier urn:lsid:myResource.org:myIdentifier
 - Qualifiers (optional) should refine the link between the model constitutent and the piece of knowledge: "has a", "is version of", "is homolog to" etc.
- The community will have to agree on a set of standard valid URIs.





Model example



$$k_1 = k_2 = k_3 = 1 \, s^{-1}$$

$$Km_1 = 10^{-7} M, Km_2 = 10^{-8}, Km_3 = 2.10^{-6} M$$

$$K_A = 10^{-11}, m = 4, n = 3, \alpha = 0.001$$

$$\frac{d[Ca_{out}]}{dt} = \frac{k_1[IP3R] * ([Ca_{in}] - [Ca_{out}])}{Km_1 + |[Ca_{in}] - [Ca_{out}]|} * \frac{[IP3]^m}{K_A + [IP3]^m}$$

$$\frac{d[IP3]}{dt} = \frac{k_2[PLC_{act}] * [PIP2]}{Km_2 + [PIP2]} - \frac{k_3[IP3_{ase}] * [IP3]}{Km_3 + [IP3]}$$
$$\frac{d[PLC_{act}]}{dt} = \frac{[G_q]^n}{\alpha + [G_q]^n} * [PLC_{tot}]$$

$$[Ca_{in}] = [IP3R] = [PLC_{tot}] = [PIP2] = [IP3_{ase}] = 0.001 M$$

$$[G_q] = 0.01 M, [Ca_{out}] = [IP3] = [PLC_{act}] = 0 M$$



creators	Joe User (juser@eden.com), Anne Other (aother@eden.com)						
creation date	01 January 0000						
last modification	31 May 2005						
Constituent	Data Type	Identifier	Qualifier	Meaning			
model	http://www.pubmed.gov/ http://www.ncbi.nlm.nih.gov/Taxonomy/ http://www.geneontology.org/ http://www.geneontology.org/ http://www.genome.jp/kegg/pathway/ http://www.genome.jp/kegg/pathway/	0000000 9606 GO:0007204 GO:0051279 hsa04020 hsa04070	IsVersionOf IsVersionOf IsPartOf IsPartOf	Homo sapiens positive regulation of cytosolic [ca²+] regulation of release of sequestered ca²+ into cytop Calcium signaling pathway - H sapiens Phosphatidylinositol signaling system - H sapiens			
compartment ER	http://www.geneontology.org/	GO:0005790		smooth endoplasmic reticulum			
reactant Cain	http://www.ebi.ac.uk/chebi/	CHEBI:29108		calcium(2+)			
cytoplasm	http://www.geneontology.org/	GO:0005737		cytoplasm			
reactant Caout	http://www.ebi.ac.uk/chebi/	CHEBI:29108		calcium(2+)			
reactant IP3	http://www.ebi.ac.uk/chebi/	CHEBI:16595		1D-myo-inositol 1,4,5-tris(dihydrogen phosphate)			
reactant PIP2	http://www.ebi.ac.uk/chebi/	CHEBI:18348		1-phosphatidyl-1D-myo-inositol 4,5-bisphosphate			
reactant IP3R	http://www.uniprot.org/ http://www.uniprot.org/ http://www.uniprot.org/	Q14643 Q14571 Q14573	HasVersion HasVersion HasVersion	Inositol 1,4,5-trisphosphate receptor type 1 Inositol 1,4,5-trisphosphate receptor type 2 Inositol 1,4,5-trisphosphate receptor type 3			
reactant PLC_{act}	http://www.uniprot.org/	Q9NQ66	IsVersionOf	PIP2 phosphodiesterase β1			
reactant PLC _{tot}	http://www.uniprot.org/	Q9NQ66		PIP2 phosphodiesterase β1			
reactant IP3 _{ase}	http://www.uniprot.org/	Q14642		Type I inositol-1,4,5-trisphosphate 5-phosphatase			
reactant $\mathbf{G}_{\mathbf{q}}$	http://www.uniprot.org/	Q6NT27		Guanine nucleotide binding protein Gq			
reaction Carelease	http://www.geneontology.org/ http://www.geneontology.org/	GO:0005220 GO:0008095	IsVersionOf	IP3-sensitive calcium-release channel activity IP3 receptor activity			
reaction IP3 _{production}	http://www.geneontology.org/ http://www.ebi.ac.uk/intenz/	GO:0004435 3.1.4.11	IsVersionOf IsVersionOf	phosphoinositide phospholipase C activity phosphoinositide phospholipase C			
reaction IP3 _{degradation}	http://www.ebi.ac.uk/intenz/	3.1.3.56	IsVersionOf	inositol-polyphosphate 5-phosphatase			
reaction PLC _{activation}	http://www.geneontology.org/	GO:0007200		G-protein signaling coupled to IP3 2nd messenger			



BioModels Database: A Free, Centralized Database of Curated, Published, Quantitative Kinetic Models of Biochemical and Cellular Systems

Le Novère N., Bornstein B., Broicher A., Courtot M., Donizelli M., Dharuri H., Li L., Sauro H., Schilstra M., Shapiro B., Snoep J.L., Hucka M. Nucleic Acids Research, (2006), 34: in the press



BioModels Database

- Store and serve quantitative models of biomedical interest
- BioModels Database goes further than MIRIAM: only models described in the peer-reviewed scientific literature.
- Models are curated: computer software check the syntax, while human curators check the semantics.
- Models are simulated to check the reference correspondence
- Model components are annotated, to improve identification and retrieval.
- Models are accepted in several format, and served in several others.
- Aims to be the Swiss-Prot of quantitative modelling.





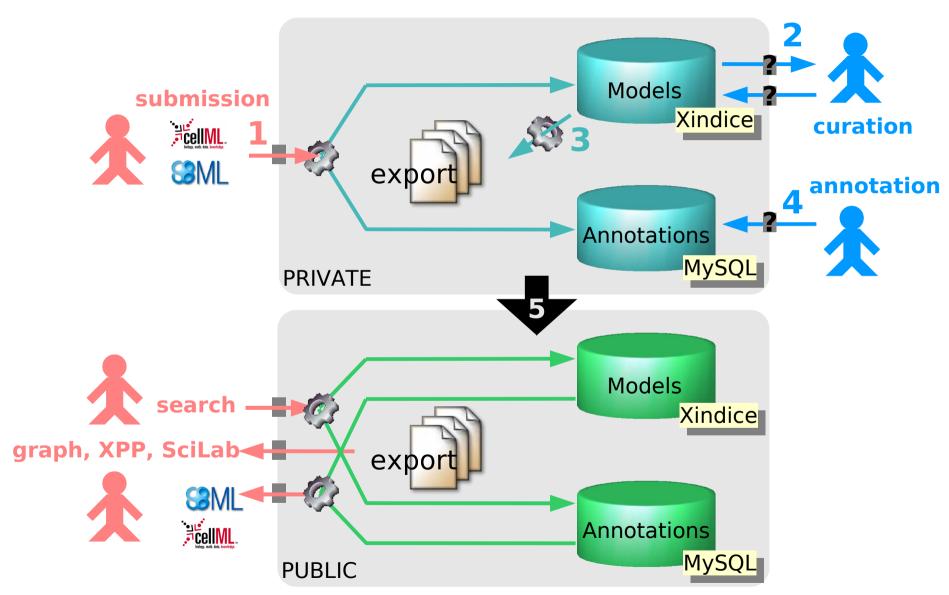
Where are the models coming from

- I) Repositories
- SBML repository
- · JWS Online
- E-Cell Developer Network
- II) Individuals
- Members of the SBML community (developers+modelers)
- Authors
- III) Journals (NPG/EMBO MSB; cf JWS)
- IV) BioModels Database curators





Structure of BioModels Database







Detailed pipeline



- 1) Non-authenticated submission: SBML or CellML
 - Annotation in the model (<notes>, <annotation>)
 - Annotation entered through submission form
- 2) [Transformation into SBML Level 2 Version 1]
- 3) XML schema validation
- 4) Consistency check
- 5) Generation of export formats (SBML, CellML, SciLab, XPP, dot)
- 6) Authenticated semantic curation (correspondence with paper)
- 7) 3 to 5 again
- 8) Authenticated annotation
 - SBML annotation in the RDF format proposed by Le Novère and Finney is recognised and reused!



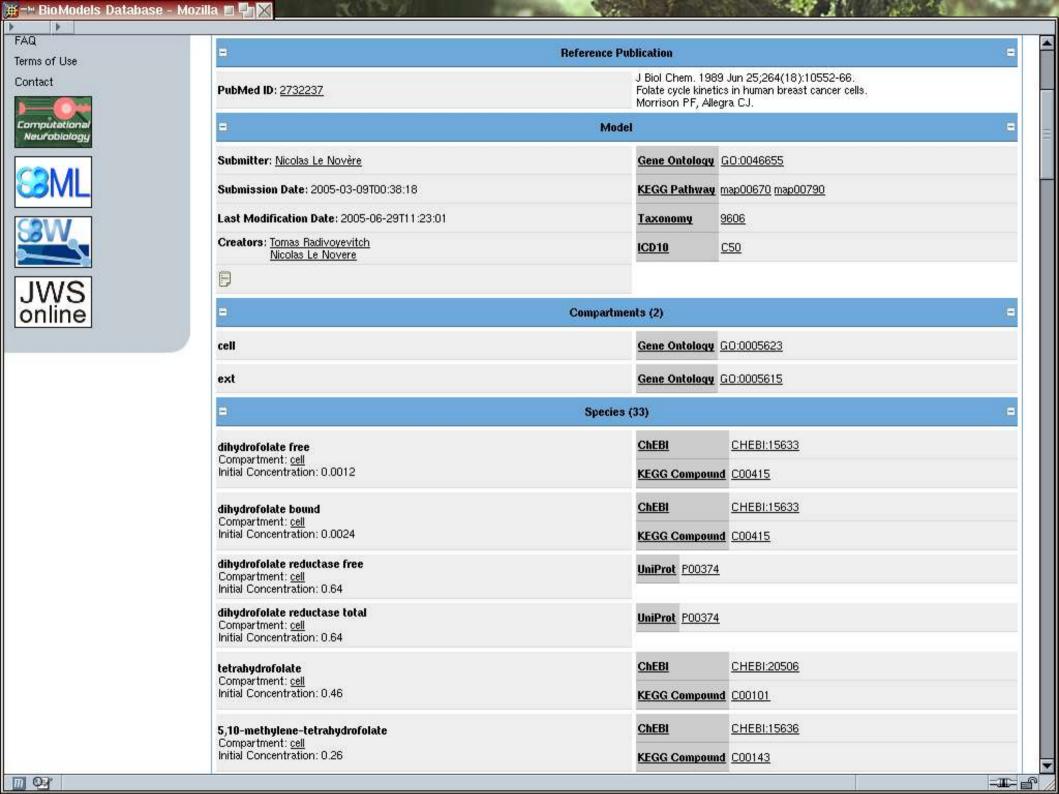


Model annotation

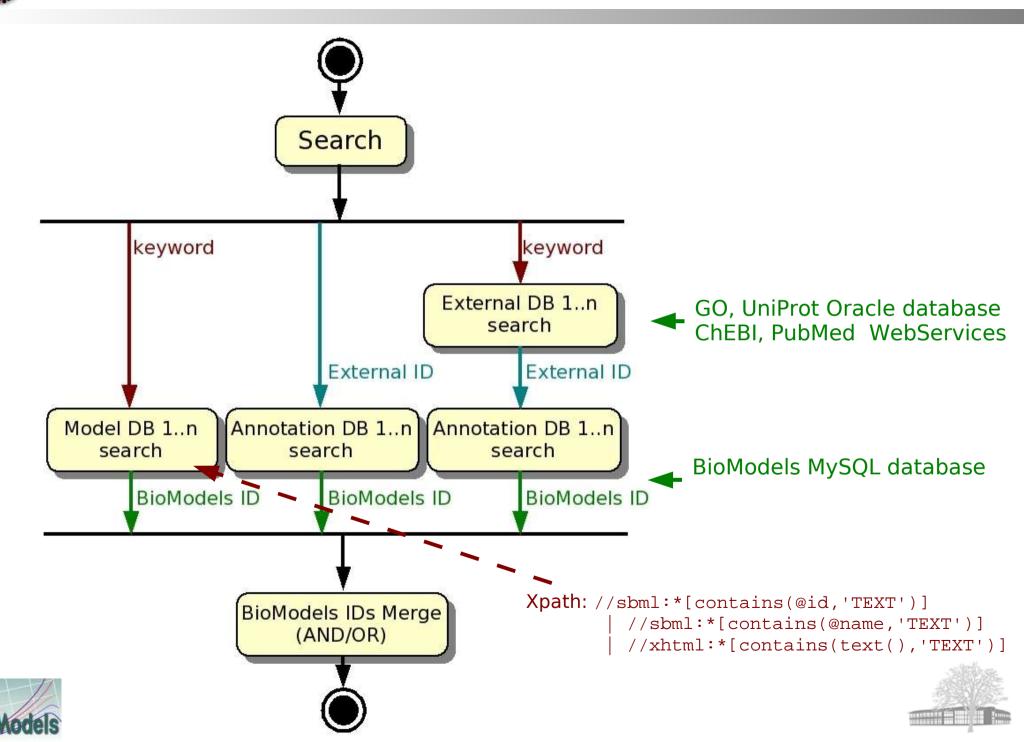
model	Submitter, Creators, publication, Gene Ontology, ICD10, KEGG pathway, OMIM, Reactome, Taxonomy		
compartment	Gene Ontology		
species	BIND complex, ChEBI, Ensembl, Gene Ontology, InterPro, KEGG compound, OMIM, Taxonomy, UniProt		
rule	BIND interaction, EC code, Gene Ontology, IntAct, KEGG pathway, KEGG Reaction, Reactome		
reaction	BIND interaction, EC code, Gene Ontology, IntAct, KEGG pathway, KEGG Reaction, Reactome		
event	BIND interaction, EC code, Gene Ontology, IntAct, KEGG pathway, KEGG Reaction, Reactome		

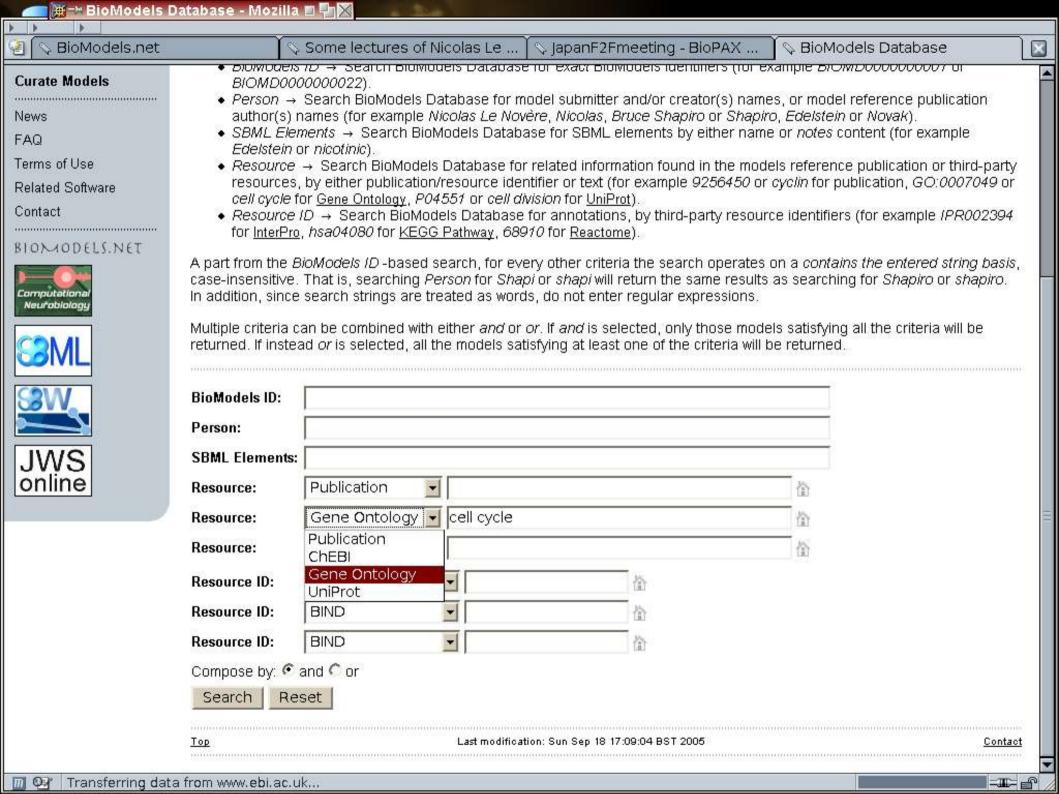






Search strategy





EBI Home

About EBI

Groups Services

Toolbox

Databases

Downloads

Submissions

BIOMODELS DATABASE

BIOMODELS.NET

......

BioModels Database

···· Browse

.... Search

Submit Your Model

Curators Sign In

......

FAQ

Terms of Use

Contact









Search Models

The search returned 6 models.

New Search

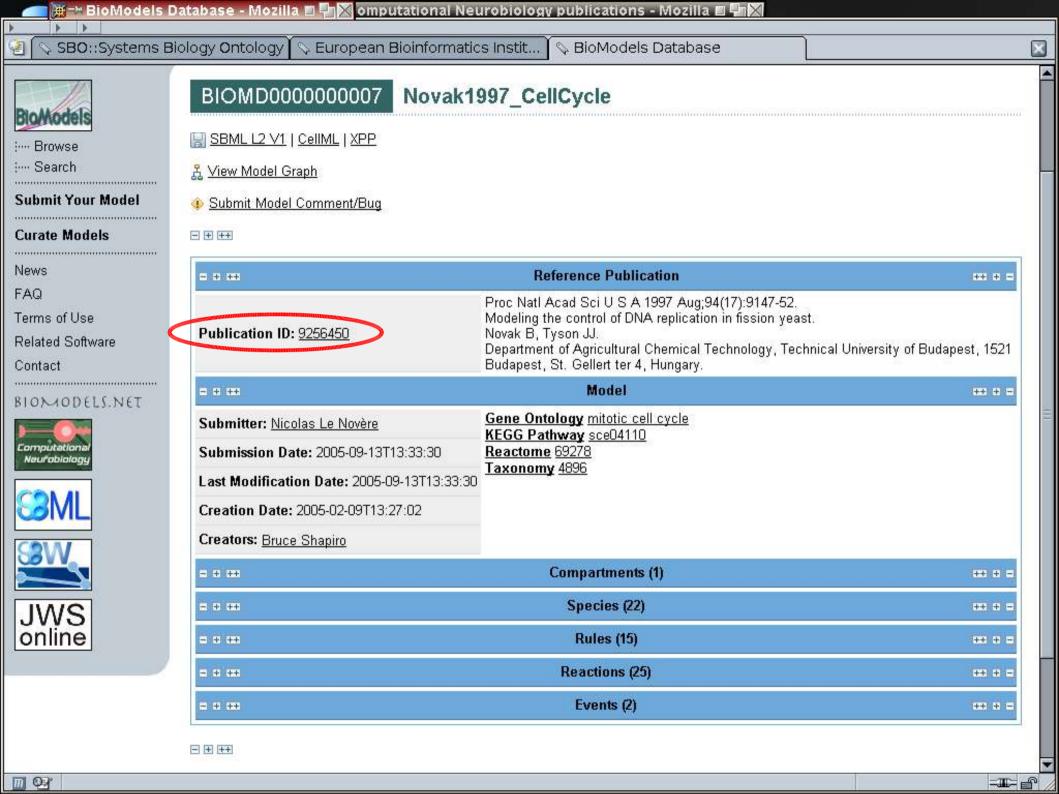
BioModels ID ▽	Name 🛆	PubMed Id △	<u>Last Modified</u> △	
BIOMD000000003	Goldbeter1991_MinMitOscil	<u>1833774</u>	2005-06-27T16:53:31	SBML
BIOMD000000004	Goldbeter1991_MinMitOscil_ExplInact	<u>1833774</u>	2005-06-27T16:53:42	SBML
BIOMD000000005	Tyson1991_CellCycle_6var	<u>1831270</u>	2005-06-27T16:54:42	SBML
BIOMD0000000006	Tyson1991_CellCycle_2var	<u>1831270</u>	2005-06-27T16:54:55	SBML
BIOMD0000000007	Novak1997_CellCycle	<u>9256450</u>	2005-06-27T16:56:33	SBML
BIOMD000000008	Gardner1998_CellCycle_Goldbeter	9826676	2005-06-27T16:56:45	SBML

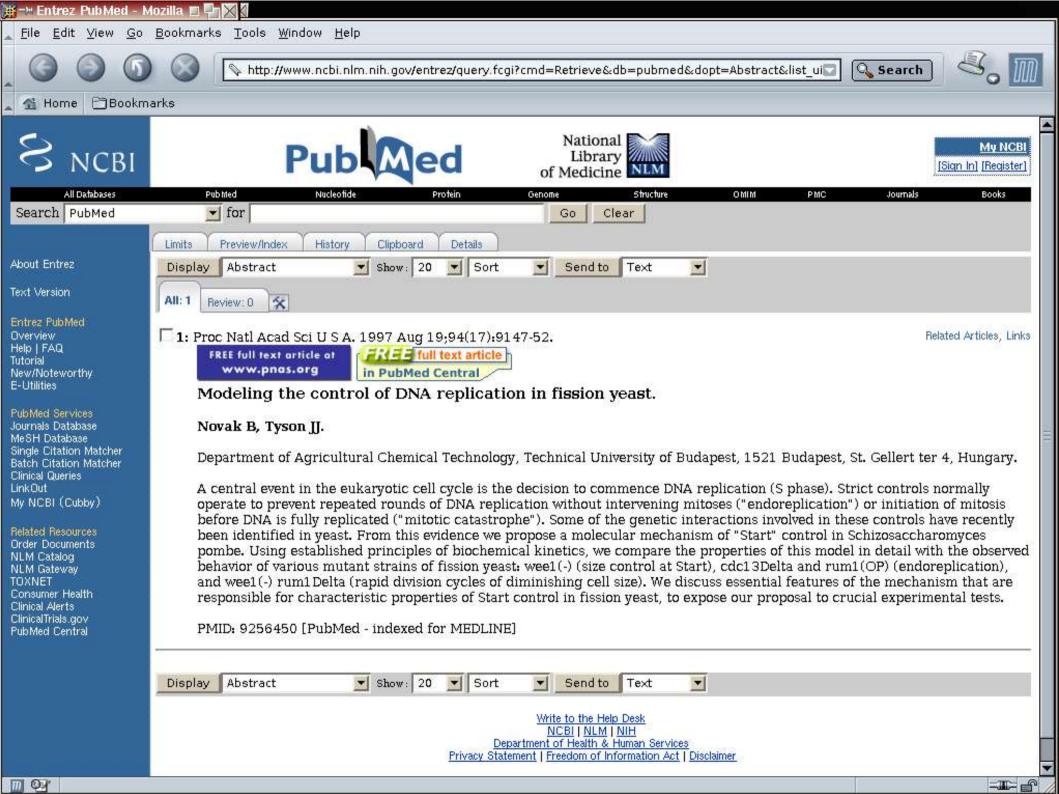
New Search

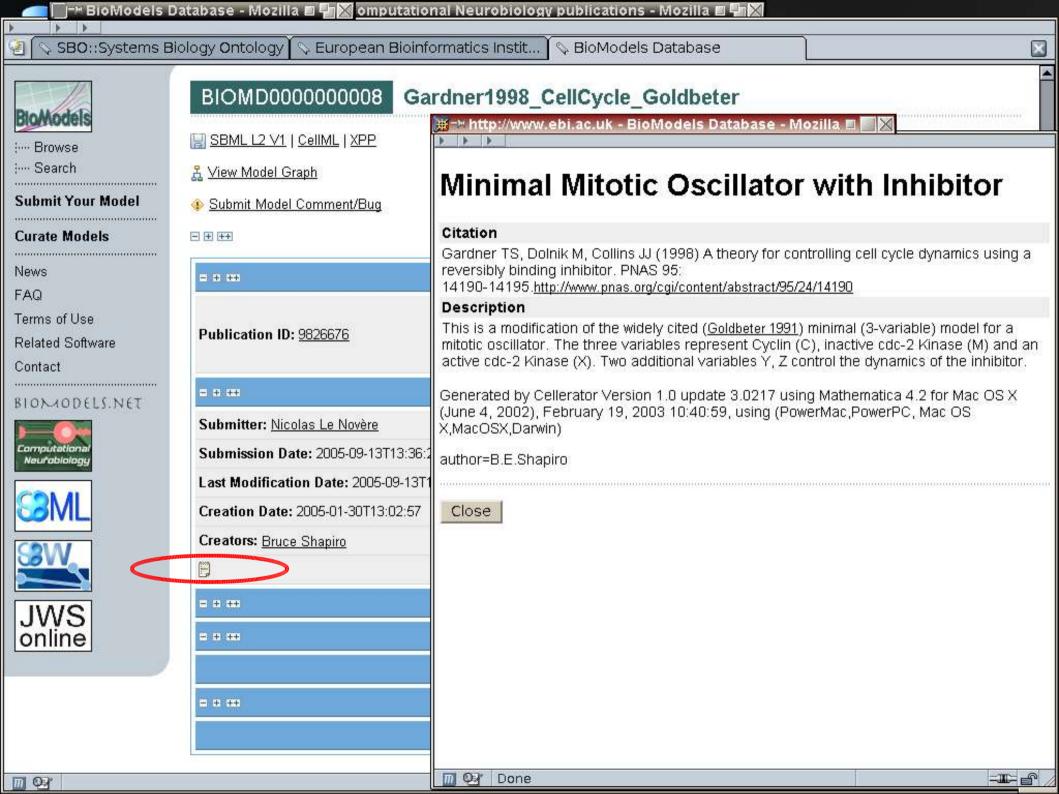
Top Last modification: Thu Apr 7 12:31:47 BST 2005

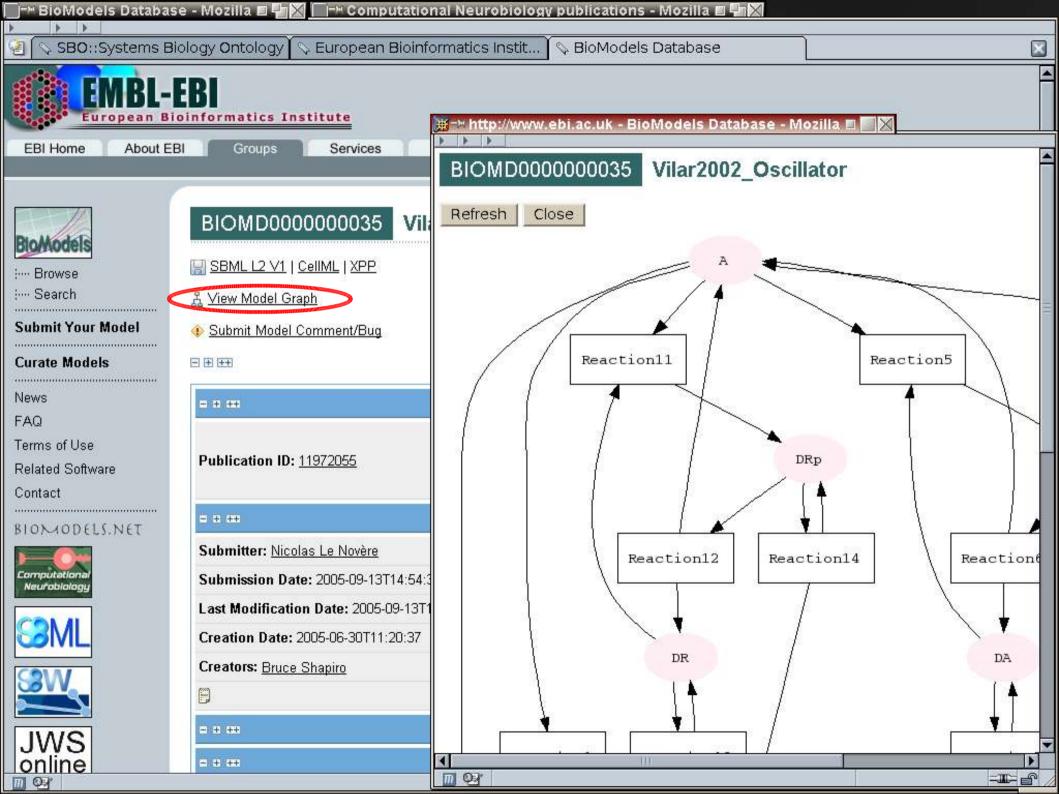


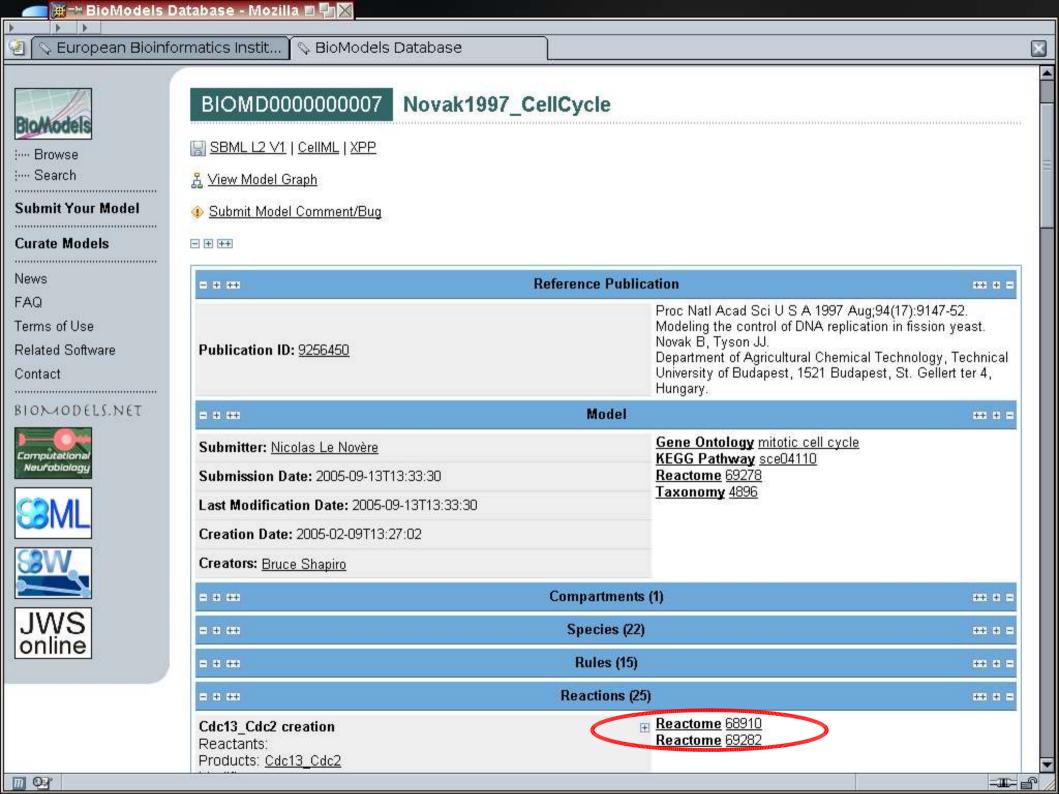


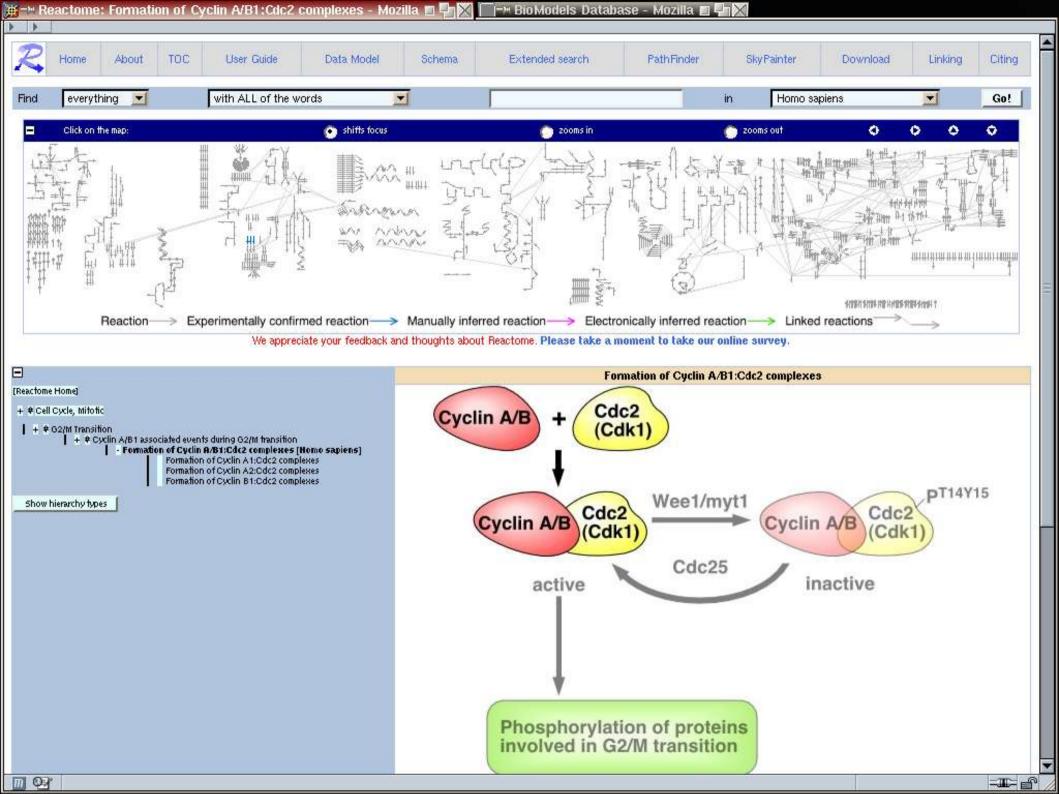


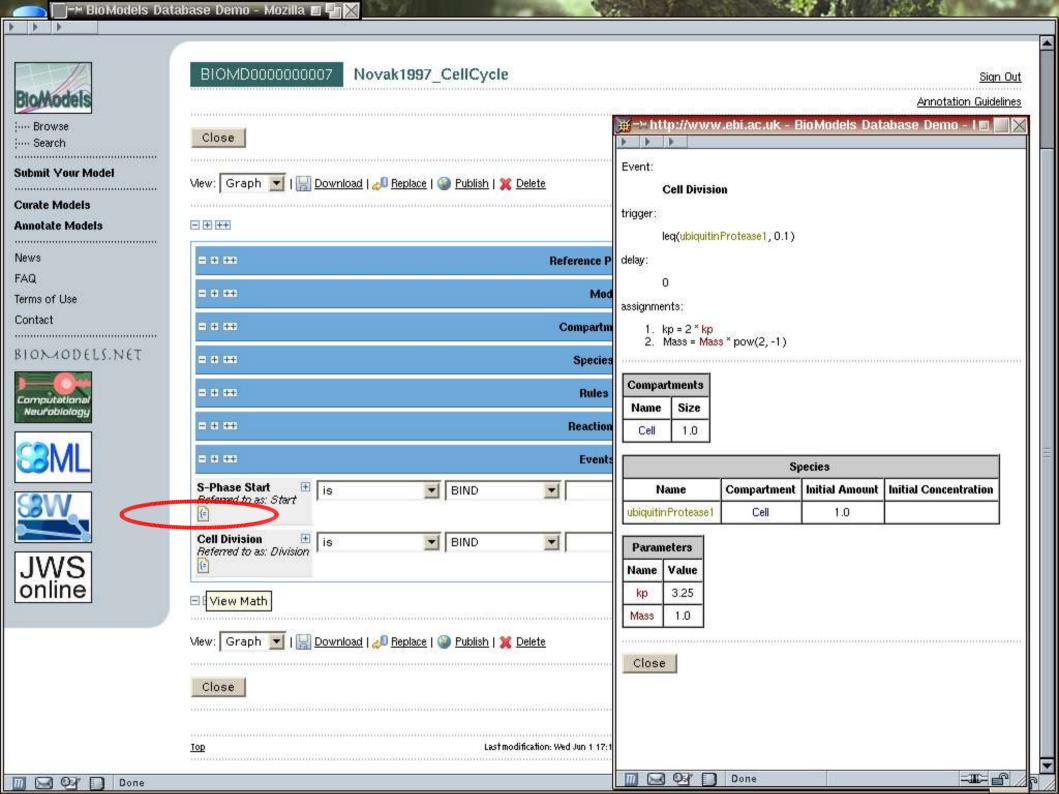


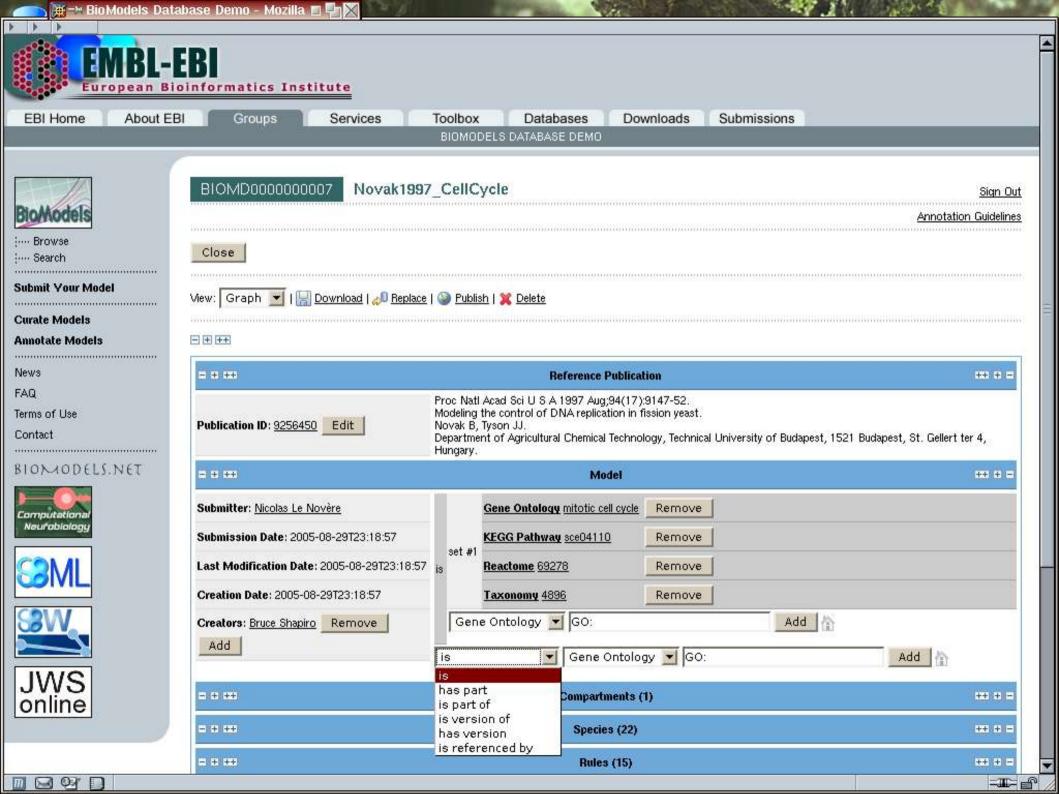


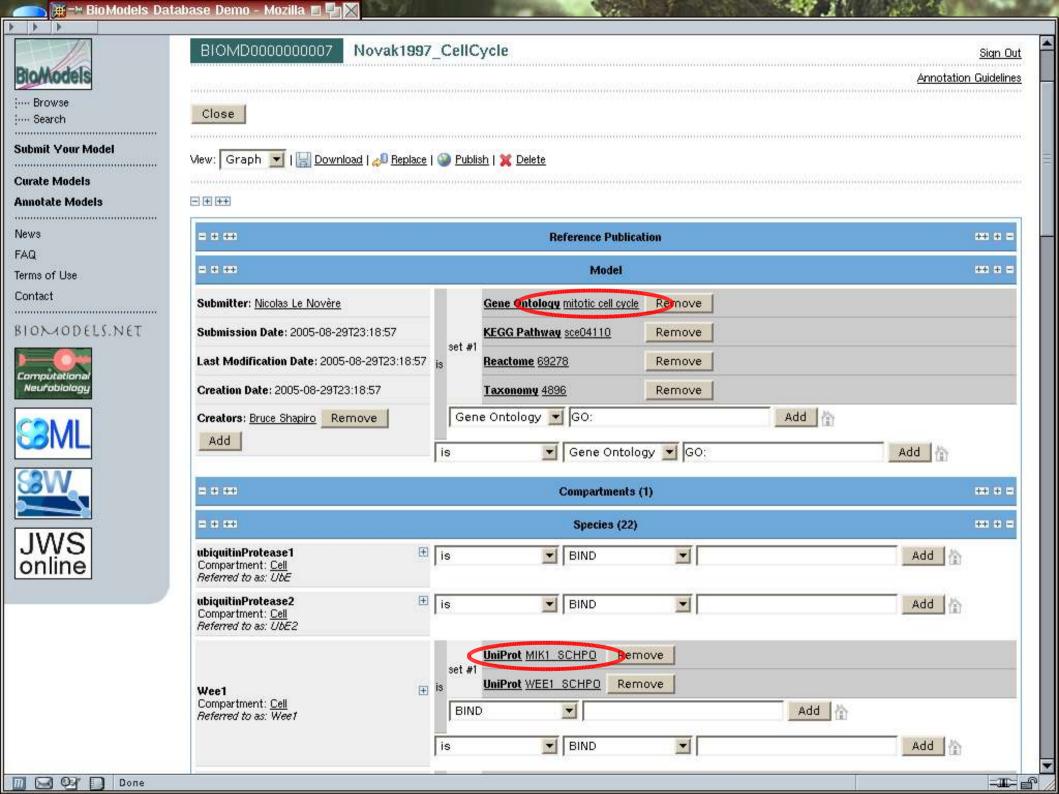


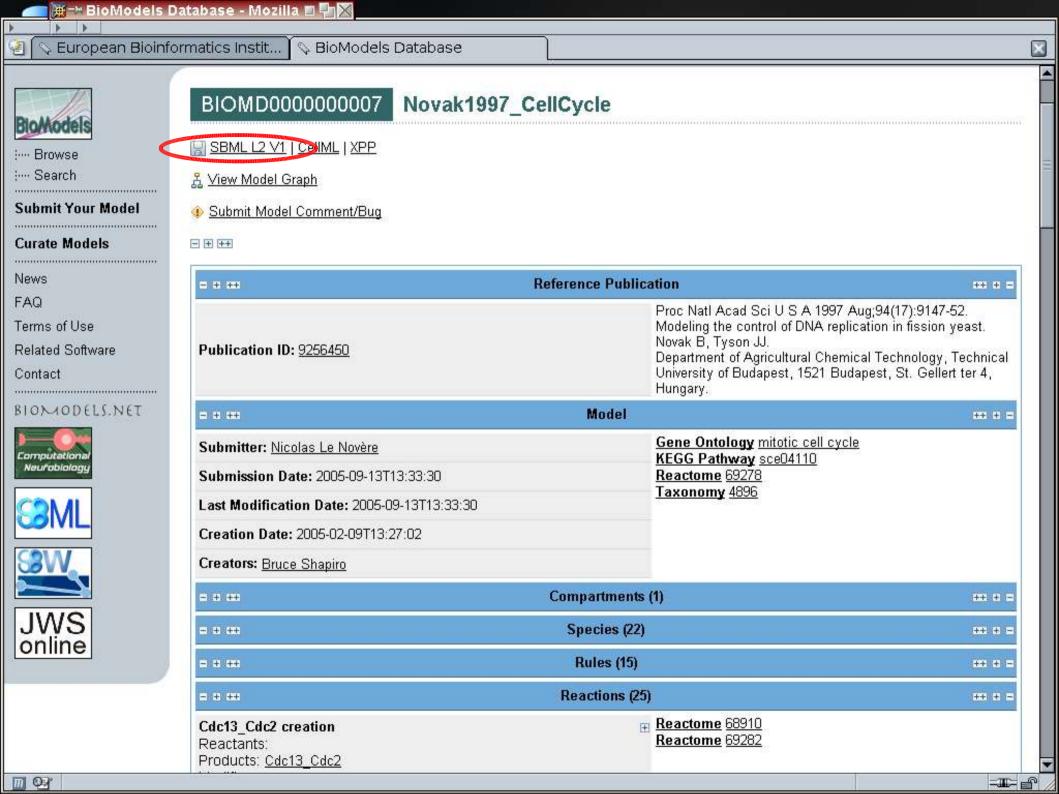


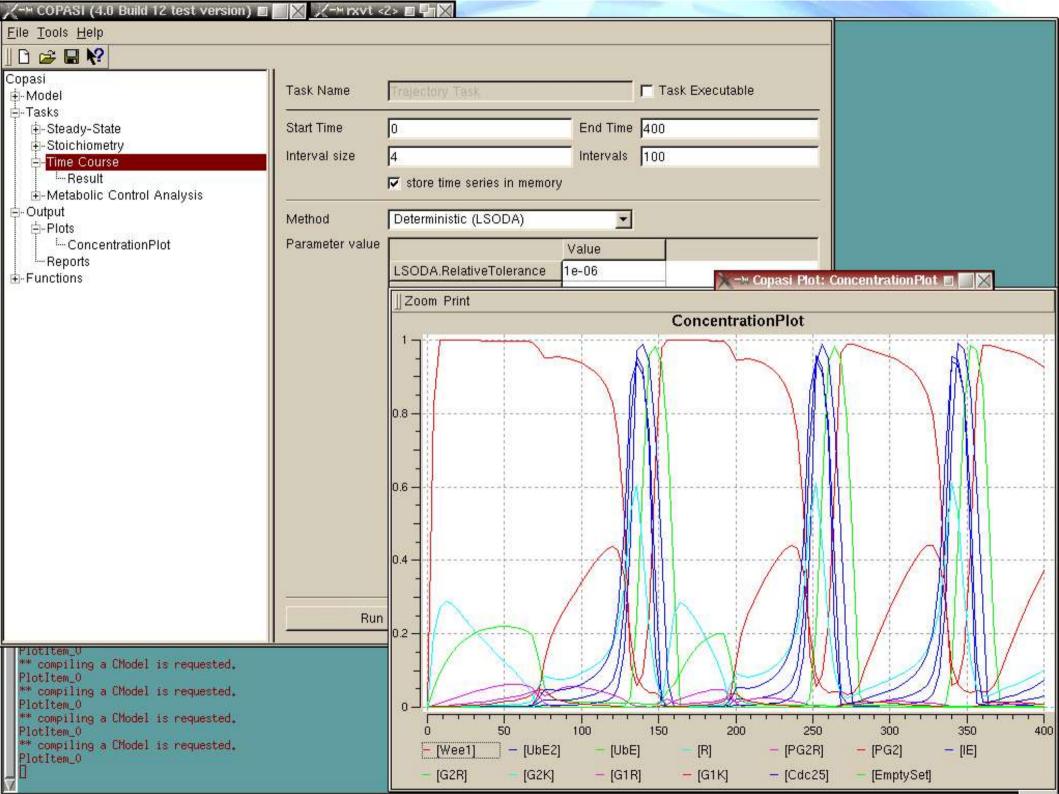














Steady increase

Release	# models	# reactions	#annotations	
11 April 2005	20	631	1084	
01 June 2005	30	736	1609	
28 July 2005	44	943	2373	
?? Decemberish	~75			

- · Still important legacy from JWS Online, DOQCS, CellML Repo
- Supported by Nature Publishing Group and BioMedCentral.
 MSB advices deposition, and forward supplementary material
- · Curation teams encode new models from literature







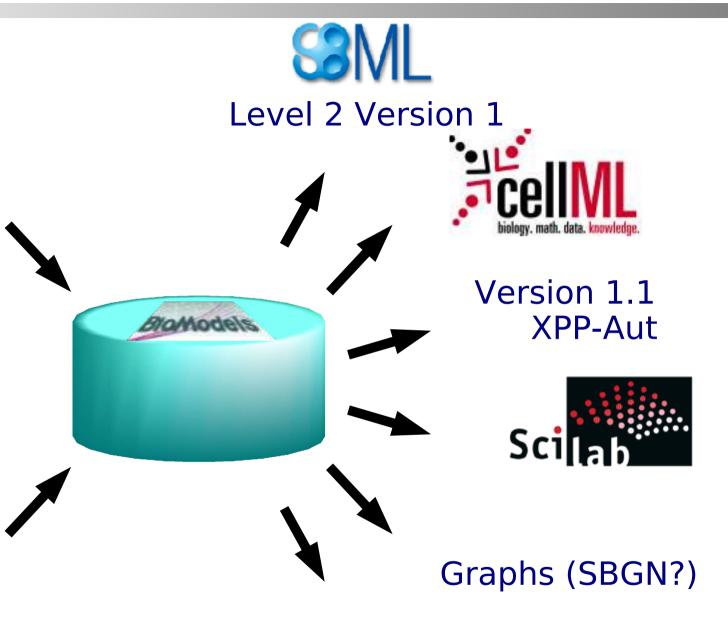




Level 1 Version 1 Level 1 Version 2 Level 2 Version 1



Version 1.0 Version 1.1

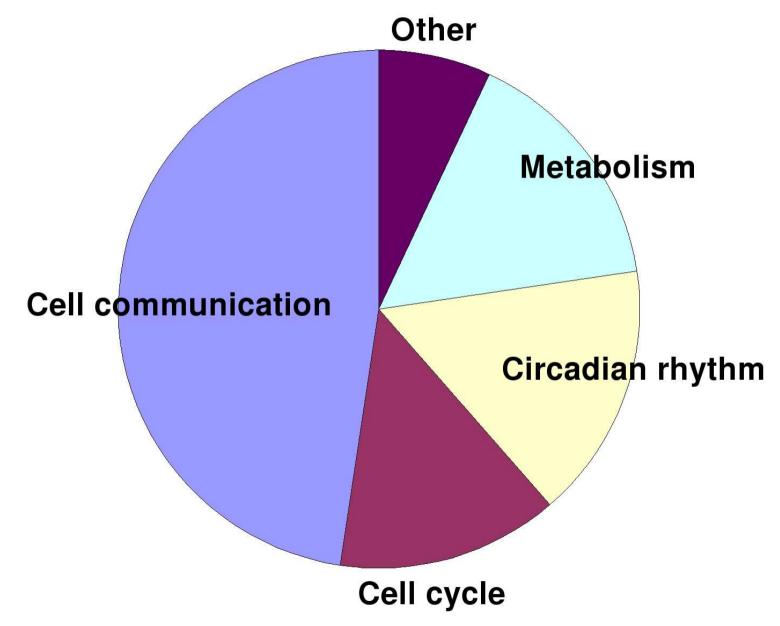


BioPAX?













The EBI team



Marco Donizelli, Chen Li: Tomcat/Xindice/Web interface





Melanie Courtot: MySQL



Lu Li: Graph, CellML, XPP, SciLab exports and curation



Curation



EMBL-EBI

An international collaboration

· EBI

- Nicolas Le Novère
- Marco Donizelli
- Mélanie Courtot
- Chen Li
- Lu Li
- Nicolas Rodriguez
- Alexander Broicher

SBML team

- Michael Hucka
- Andrew Finney
- Bruce Shapiro
- Benjamin Borstein
- Maria Schilstra
- Sarah Keating

Keck Graduate Institute

- Herbert Sauro
- Harish Dharuri

· Systems Biology Institute

- Hiroaki Kitano
- Akira Funahashi

· Stellenbosh University

Jacky Snoep

External contributors

- Samuel Bandara
- Upinder Bhalla
- Christoph Flamm
- Ryan Gutenkunst
- Adam Halasz
- Ken Lau
- Rainer Machné
- Marc Poolman
- Tomas Radivoyevitch
- Oleg Sokolsy
- Joanne Matthews
- Les Grivell
- Boris Kholodenko
- Tjeerd olde Scheper
- Birgit Schoeberl
- Paul Smolen
- Yukiko Matsuoka

Programs used for curation

- COPASI
- SBMLodeSolver
- Jarnac/JDesigner













European Bioinformatics Institute

British outstation of the European Molecular Biology Laboratory

- Data resources
 - Sequences, structures
 - Transcriptomics, Proteomics pathways, models
 - Controlled vocabularies and dictionaries
 - + ~200 other resources
- Research groups
 - Comparative genomics (Ouzounis)
 Structural Genomics (Thornton)
 Molecular Evolution (Goldman)
 - Text-Mining (Rebholz-Schumman)
 Computational Systems Biology (Le Novère)
 Statistical array analysis (Huber)
 Genomic analysis of regulatory systems (Luscombe)

Marie Curie Training site Fellowships: 3-6 months. Fully funded.























