



Pathway Databases and Standards

Huaiyu Mi, Ph.D.

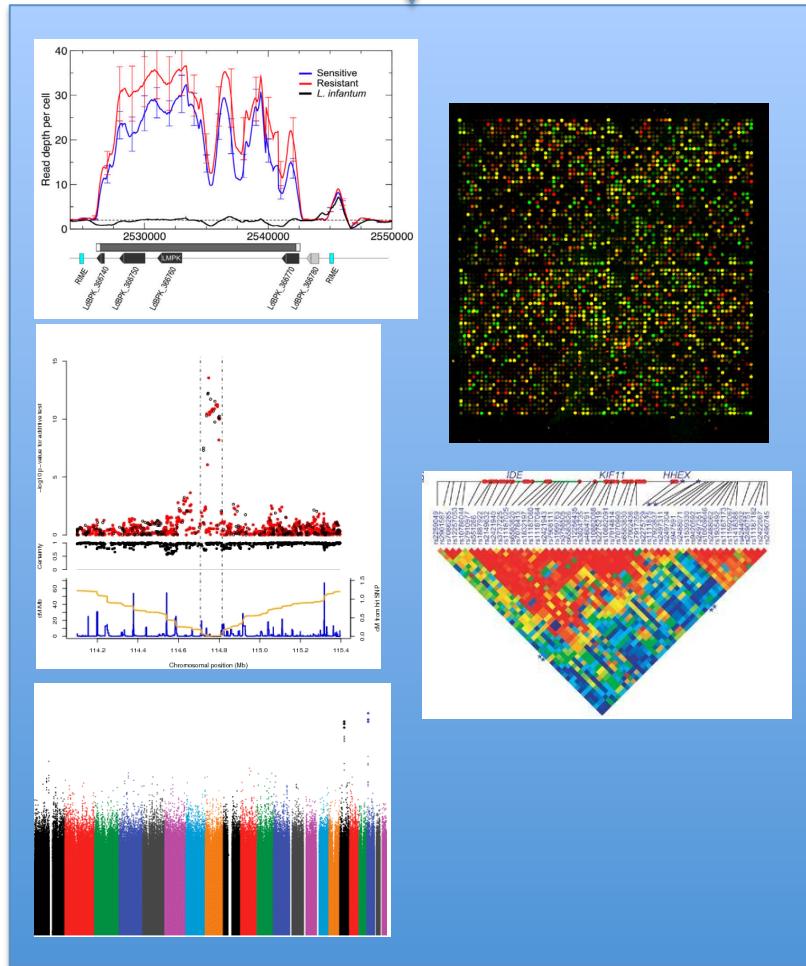
Division of Bioinformatics

Department of Preventive Medicine

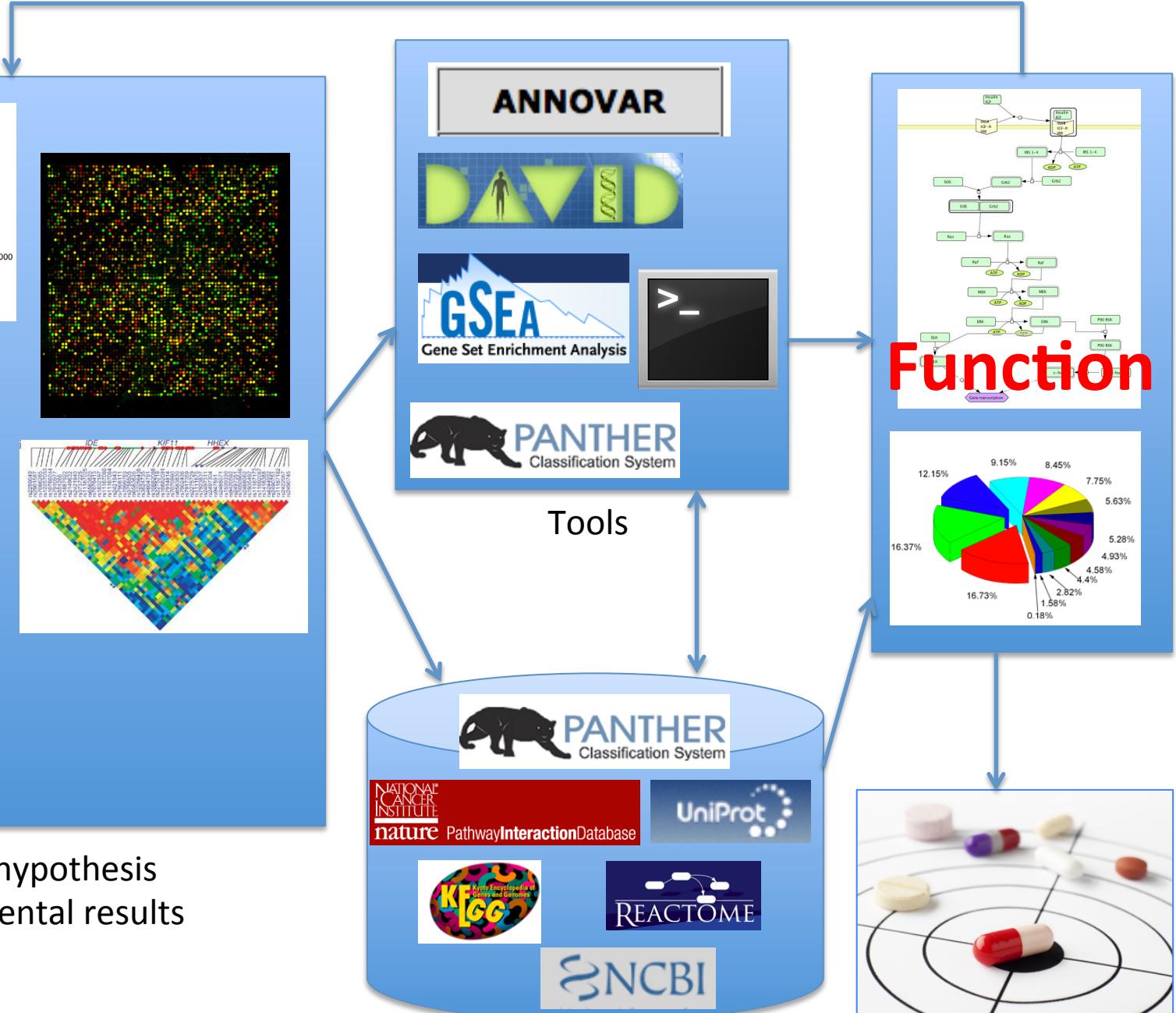
Keck School of Medicine

University of Southern California

April 9, 2013 NIMBioS Tutorial, Knoxville, Tennessee



Biological hypothesis
& experimental results



Databases

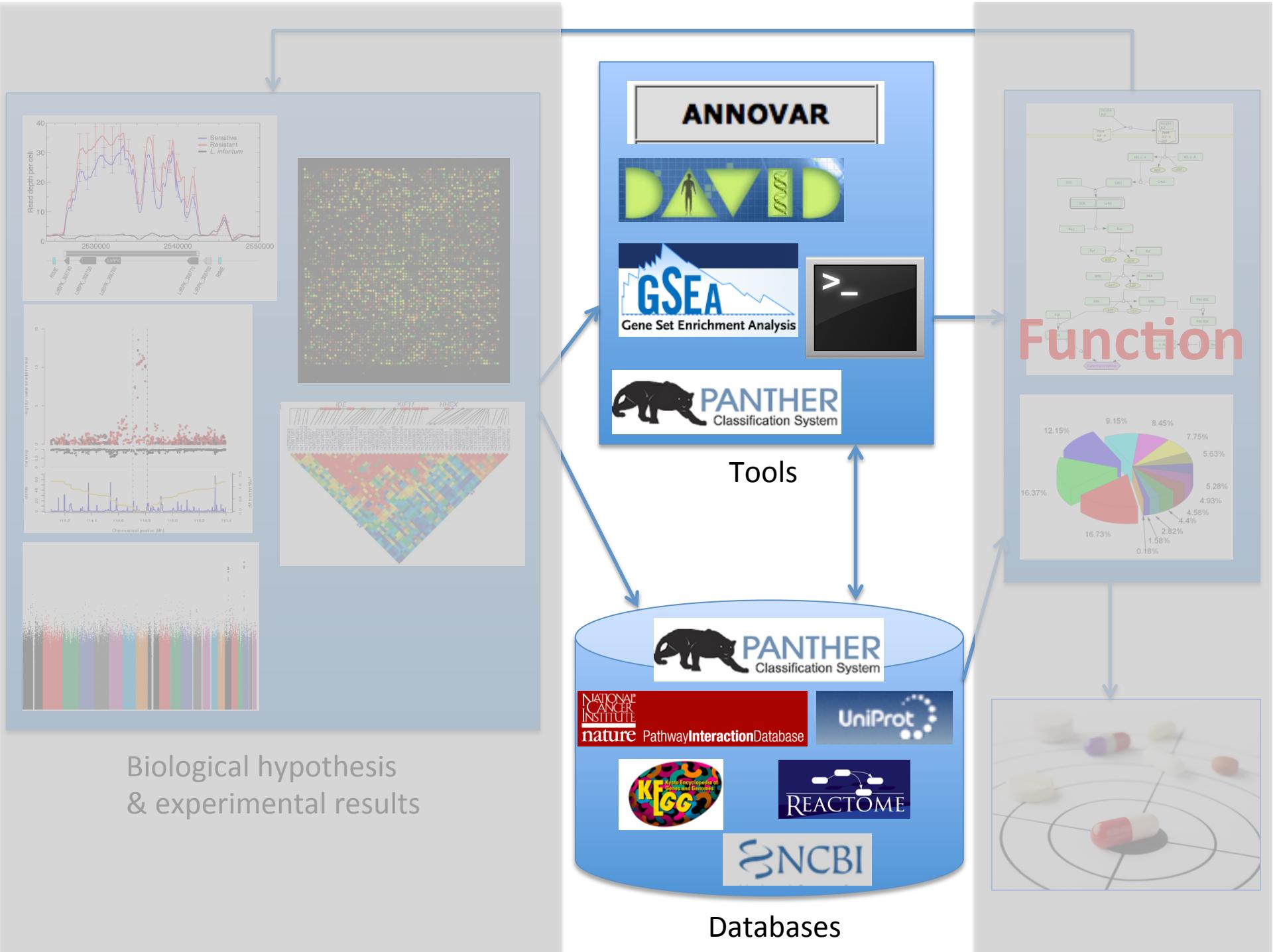
Systems Approach in Biological Research

- High throughput technologies have generated an unprecedented amount of data.
- A systems approach is essential.
 - Storage/representation
 - Modeling/simulation
- Computers and software will greatly facilitate the process.
- Scientists must interpret the data

Systems Approach – Two Key Elements

A formal data representation of pathway knowledge – pathway database

A standard data structure that both computer and human can read – standards and ontologies



Outline

- Introduction of pathway databases
- Standards used in pathway databases
 - History of ontology
 - Ontology for biology
 - Pathway standards
 - SBML
 - BioPAX
 - SBGN

Outline

- **Introduction of pathway databases**
- Standards used in pathway databases
 - History of ontology
 - Ontology for biology
 - Pathway standards
 - SBML
 - BioPAX
 - SBGN

EcoCyc – the first pathway database

Proc Int Conf Intell Syst Mol Biol. 1993;1:207-15.

Representations of metabolic knowledge.

Karp PD, Riley M.

SRI International, Menlo Park, CA 94025, USA.

Abstract

Construction of electronic repositories of metabolic information is an increasingly active area of research. Detailed knowledge of a complex biological domain requires finely honed representations. We survey the representations used for several metabolic databases, including Eco-Cyc, and reach the following conclusions. Representations of metabolism must distinguish enzyme classes from individual enzymes, because there is not a one-to-one mapping from enzymes to the reactions they catalyze. Individual enzymes must be represented explicitly as pairs, encoding their subunit structure. The species variation of metabolism must be represented. So must the specificity of enzymes, which may be treated in several ways.

32-39 Nucleic Acids Research, 1996, Vol. 24, No. 1

© 1996 Oxford University Press

EcoCyc: an encyclopedia of *Escherichia coli* genes and metabolism

Peter D. Karp*, Monica Riley¹, Suzanne M. Paley and Alida Pellegrini-Toole¹

Artificial Intelligence Center, SRI International, 333 Ravenswood Avenue, Menlo Park, CA 94025, USA

¹Marine Biological Laboratory, Woods Hole, MA 02543, USA

Received August 31, 1995; Revised and Accepted October 17, 1995

ABSTRACT

The encyclopedia of *Escherichia coli* genes and metabolism (EcoCyc) is a database that combines information about the genome and the intermediary metabolism of *E. coli*. It describes 2034 genes, 306 enzymes encoded by these genes, 580 metabolic reactions that occur in *E. coli* and the organization of these reactions into 100 metabolic pathways. The EcoCyc graphical user interface allows query and exploration of the EcoCyc database using visualization tools such as genomic map browsers and automatic layouts of metabolic pathways. EcoCyc spans the space from sequence to function to allow investigation of an unusually broad range of questions. EcoCyc can be thought of as both an electronic review article, because of its copious references to the primary literature, and as an *in silico* model of *E. coli* that can be probed and analyzed through computational means.

containing citations to) the primary literature. In this respect its aim differs from that of databases such as GenBank, because GenBank is designed as a repository of primary observations (an electronic mirror of the primary literature). Another difference between EcoCyc and databases such as GenBank is that EcoCyc describes several classes of biological objects (such as proteins, genes and pathways), whereas GenBank describes only nucleic acid sequences. EcoCyc is therefore an electronic reference source on *E. coli*. However, EcoCyc is also designed to facilitate complex computations on genomic and metabolic data and to provide an *in silico* model of *E. coli* that can be probed and analyzed through computational means.

Problems that might be addressed using EcoCyc include the following.

Genomic investigations

Coupled with sequence databases EcoCyc could be used to perform function-based retrieval of DNA or protein sequences.

Representations of Metabolic Knowledge: Pathways

Peter D. Karp and Suzanne M. Paley

Artificial Intelligence Center, SRI International

333 Ravenswood Ave., EJ229

Menlo Park, CA 94025

{pkarp,paley}@ai.sri.com

Abstract

The automatic generation of drawings of metabolic pathways is a challenging problem that depends intimately on exactly what information has been recorded for each pathway, and on how that information is encoded. The chief contributions of the paper are a minimized representation for biochemical pathways called the *predecessor list*, and inference procedures for converting this list into a *pathway-graph* representation. The conversion between the two representations can be formulated as both a constraint-satisfaction problem and a logical inference problem, involving directions to reactions, and to determine which are the main chemical compounds in a pathway. The conversion between the two representations can be formulated as both a constraint-satisfaction problem and a logical inference problem, involving directions to reactions, and to determine which are the main chemical compounds in a pathway.

a number of metabolic pathways, each consisting of a collection of bioreactions. The EcoCyc GUI is capable of generating displays of each of these types of objects, such as a reaction, an enzyme, or a metabolic pathway. The automatic generation of drawings of metabolic pathways is a challenging problem that depends intimately on exactly what information is recorded for each pathway, and on how that information is encoded. Our representation of metabolic pathways — such as the TCA cycle, glycolysis, and tryptophan biosynthesis — facilitates both knowledge acquisition of pathways, and automatic pathway drawing.

The chief contributions of this paper are a minimized representation for biochemical pathways called the *predecessor list*, and inference procedures for converting the predecessor list into a *pathway-graph* representation that can serve as input to a pathway-drawing algorithm. The predecessor list has several advantages over the pathway graph, including its compactness and its lack of redundancy. The conversion between the two representations can be formulated as both a constraint-satisfaction problem and a logical inference problem, whose goal is to assign directions to reactions, and to determine which are the main chemical compounds in a pathway.

Karp and Riley (1993) Proc Int Conf Intell Syst Mol Biol. 1:207-15

Karp and Paley (1994) Proc Int Conf Intell Syst Mol Biol. 2:203-11

Karp et. al. (1996) Nucleic Acid Res. 24:32-39

[!\[\]\(9c4f697052545ae4fab36076e03db94f_img.jpg\) Add to group](#)

Escherichia coli K-12 substr. MG1655 Pathway: glycolysis I

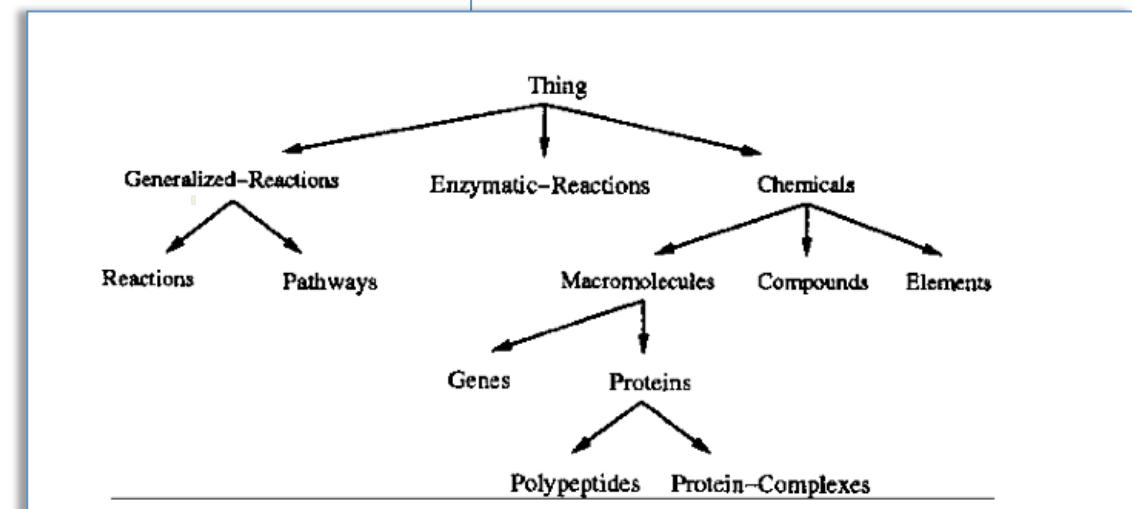
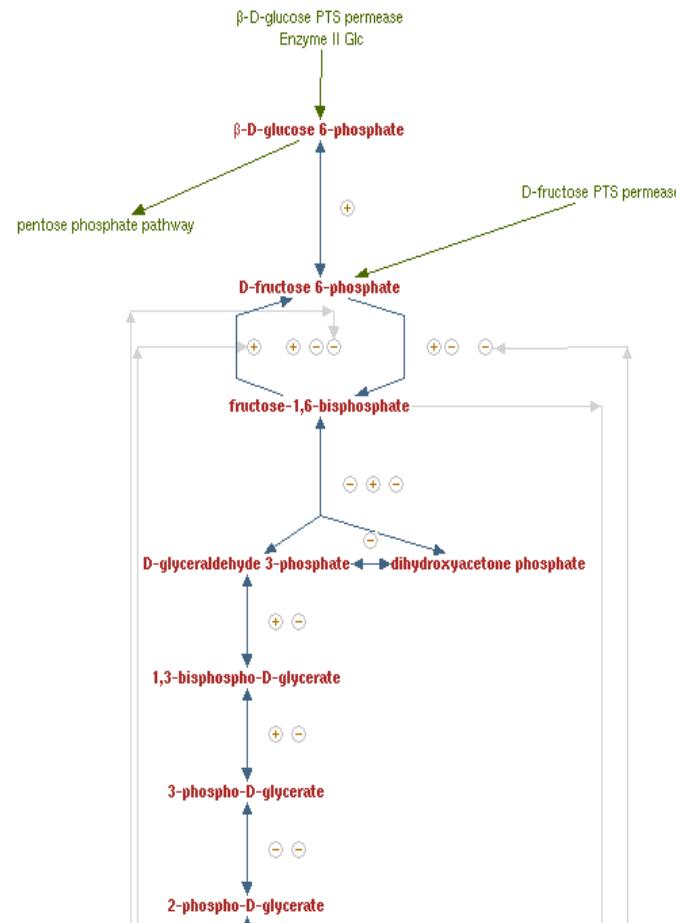
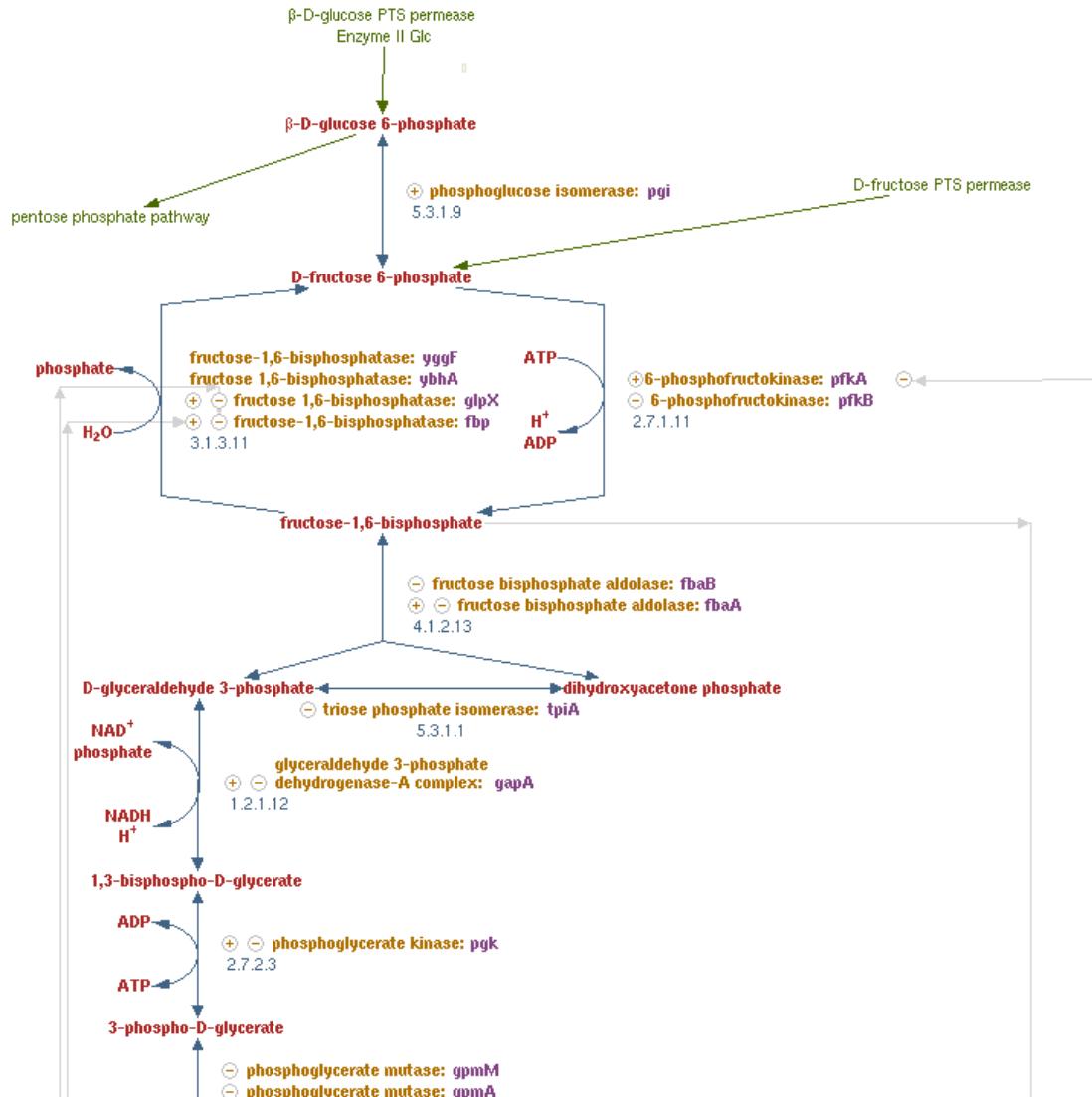

[More Detail](#) [Less Detail](#) [Species Comparison](#)


Figure 4. Karp et. al. 1996.

Concept of ontology was already used in Ecocyc back in 1993

[Add to group](#)
Escherichia coli K-12 substr. MG1655 Pathway: glycolysis I

[More Detail](#) [Less Detail](#) [Species Comparison](#)


- Controlled vocabulary
- Entity
- Reaction/relationship
- Graphical representation
- Evidence/literature

Pathguide

 » the pathway resource list**Navigation**

[Protein-Protein Interactions](#)
[Metabolic Pathways](#)
[Signaling Pathways](#)
[Pathway Diagrams](#)
[Transcription Factors / Gene Regulatory Networks](#)

[Protein-Compound Interactions](#)

[Genetic Interaction Networks](#)

[Protein Sequence Focused](#)

[Other](#)

Search

[Organisms](#)

All

[Availability](#)

All

[Standards](#)

All

[Reset](#)

[Search](#)

Analysis

[Statistics](#)

[Database Interactions](#)

Contact

[Comments, Questions, Suggestions are](#)

Complete Listing of All Pathguide Resources

Pathguide contains information about 325 biological pathway related resources and molecular interaction related resources. Click on a link to go to the resource home page or 'Details' for a description page. Databases that are free and those supporting BioPAX, CellML, PSI-MI or SBML standards are respectively indicated.

If you know of a pathway resource that is not listed here, or have other questions or comments, please [send us an e-mail](#).

News**New visual navigation**

May.2010
Click the 'Database interactions' link on the left menu to access.

Major update

All resources were recently reviewed and many new ones

Protein-Protein Interactions

Database Name (Order: alphabetically | by web popularity

[Full Record](#) [Availability](#) [Standards](#)

3DID - 3D interacting domains	Details	Free
ADAN - Prediction of protein-protein interaction of modular domains	Details	Free
AllFuse - Functional Associations of Proteins in Complete Genomes	Details	X
aMAZE - Protein Function and Biochemical Pathways Project	Details	Free
APID - Agile Protein Interaction DataAnalyzer	Details	Free
ASEdb - Alanine Scanning Energetics Database	Details	Free
ASPD - Artificial Selected Proteins/Peptides Database	Details	Free
AtPID - Arabidopsis thaliana Protein Interactome Database	Details	Free
BID - Binding Interface Database	Details	Free
BIND - Biomolecular Interaction Network Database	Details	Free
BioGRID - Biological General Repository for Interaction Datasets	Details	
BRITE - Biomolecular Relations in Information Transmission and Expression	Details	Free
CA1Neuron - Pathways of the hippocampal CA1 neuron	Details	Free
Cancer Cell Map - The Cancer Cell Map	Details	Free
CellCircuits - CellCircuits	Details	Free

Bader (2006) *Nucleic Acid Res.* 34:D504-6

Over 300 pathway databases today- according to Pathguide

Pathway databases are diverse

Pathway databases are serving a diverse user community

- Knowledge base
 - Metabolic – Ecocyc, KEGG
 - Signaling – Reactome, PANTHER
 - Protein interaction – BioGRID
 - Domain specific – based on specific domain of knowledge, e.g., Rice Kinase Database, PDZBase
- Targeted users
 - For wet lab scientists – focus on gene or protein sequences and pathway knowledge
 - For systems biologists – focus on kinetics – usually are referred to as “models”, e.g., Biomodel database.
- Relationship types
 - Reactions or processes – Ecocyc, Reactome, PANTHER
 - Interaction – BioGRID, Interactome
 - Activity flow – STKE

Different tools => different interfaces & languages





Standards



Outline

- Introduction of pathway databases
- Standards used in pathway databases
 - History of ontology
 - Ontology for biology
 - Pathway standards
 - SBML
 - BioPAX
 - SBGN

What is an Ontology?

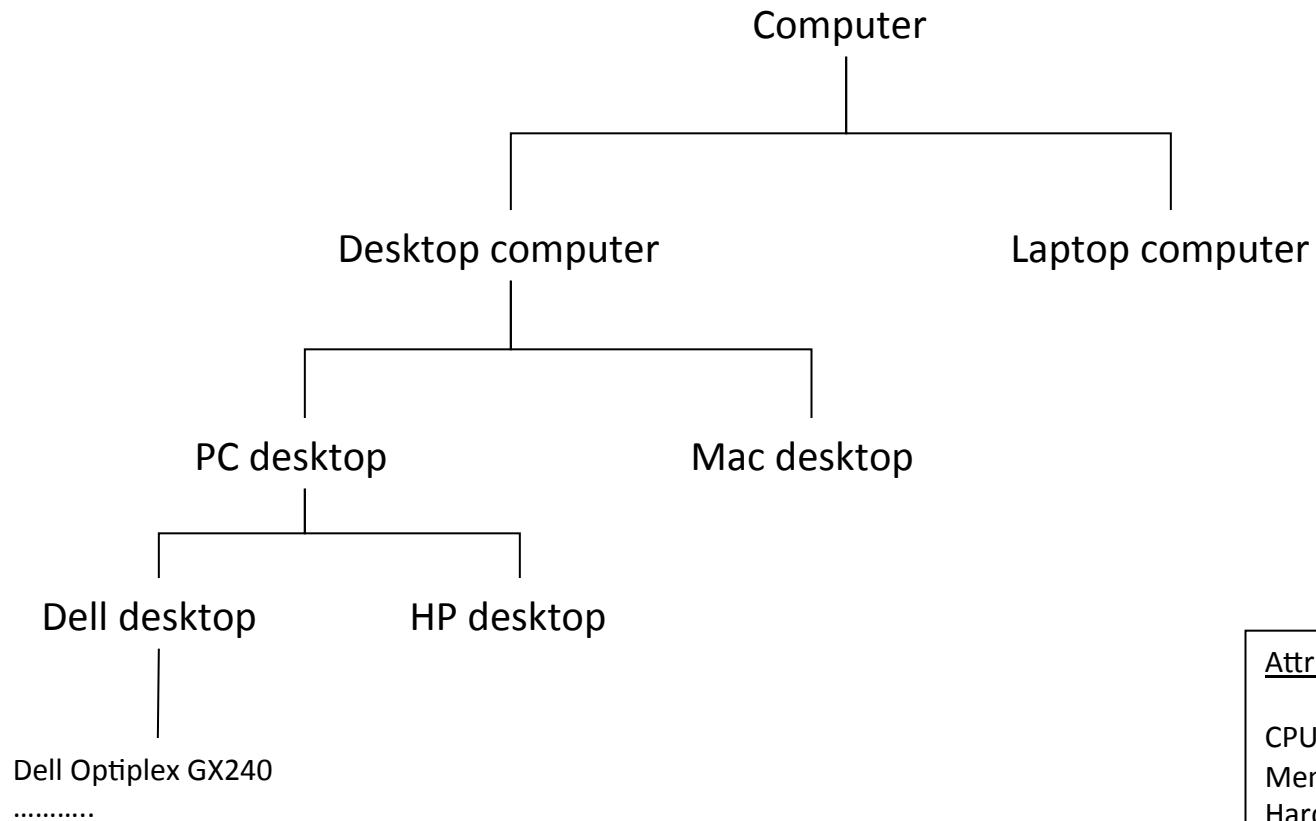
- Ontology is originated from ancient Greek philosophy.
- It is a study of being and of existence, and their basic categories and the relationships between them.
- It is a formal structuring of knowledge.

Smith B. (2003) Ontology in L. Floridi (ed.), *Blackwell Guide to the Philosophy of Computing and Information*, Oxford: Blackwell, 155

What is an Ontology? cont.

- Contemporary ontology was first used in computer science and information science
- It was used as a data model to represent a set of concepts within a domain and the relationships among those concepts.
- It is also used as a form of knowledge representation in a number of fields
 - artificial intelligence
 - semantic Web
 - software engineering
 - biomedical informatics
 - library science
- The contemporary ontology differs from the classical philosophical ontology in that it is expressed in a machine-readable format, and that it assesses in terms of usefulness rather than truth.

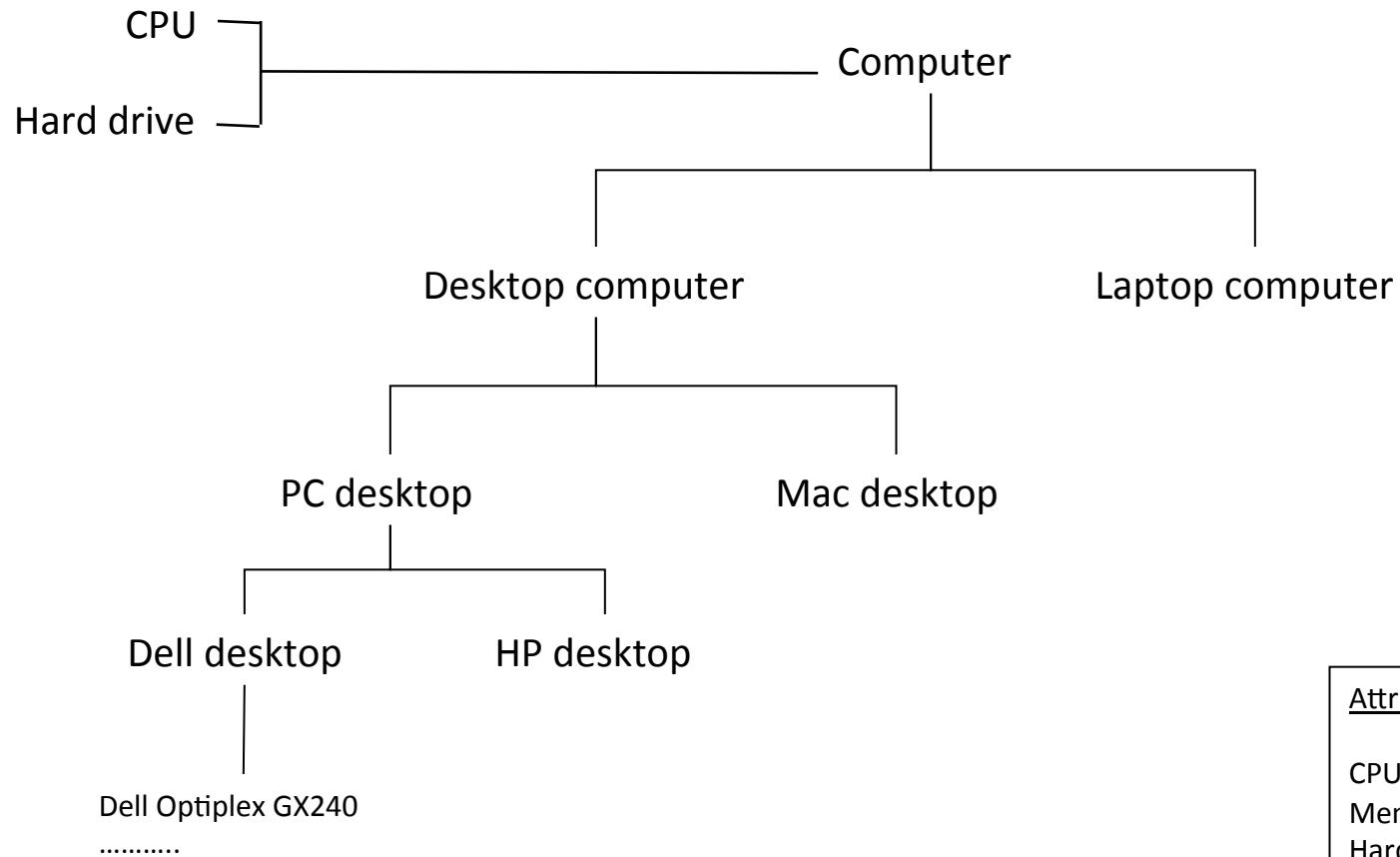
Example



Attributes:

CPU: Pentium 4 2GH
Memory: 1GB
Hard drive: 100GB

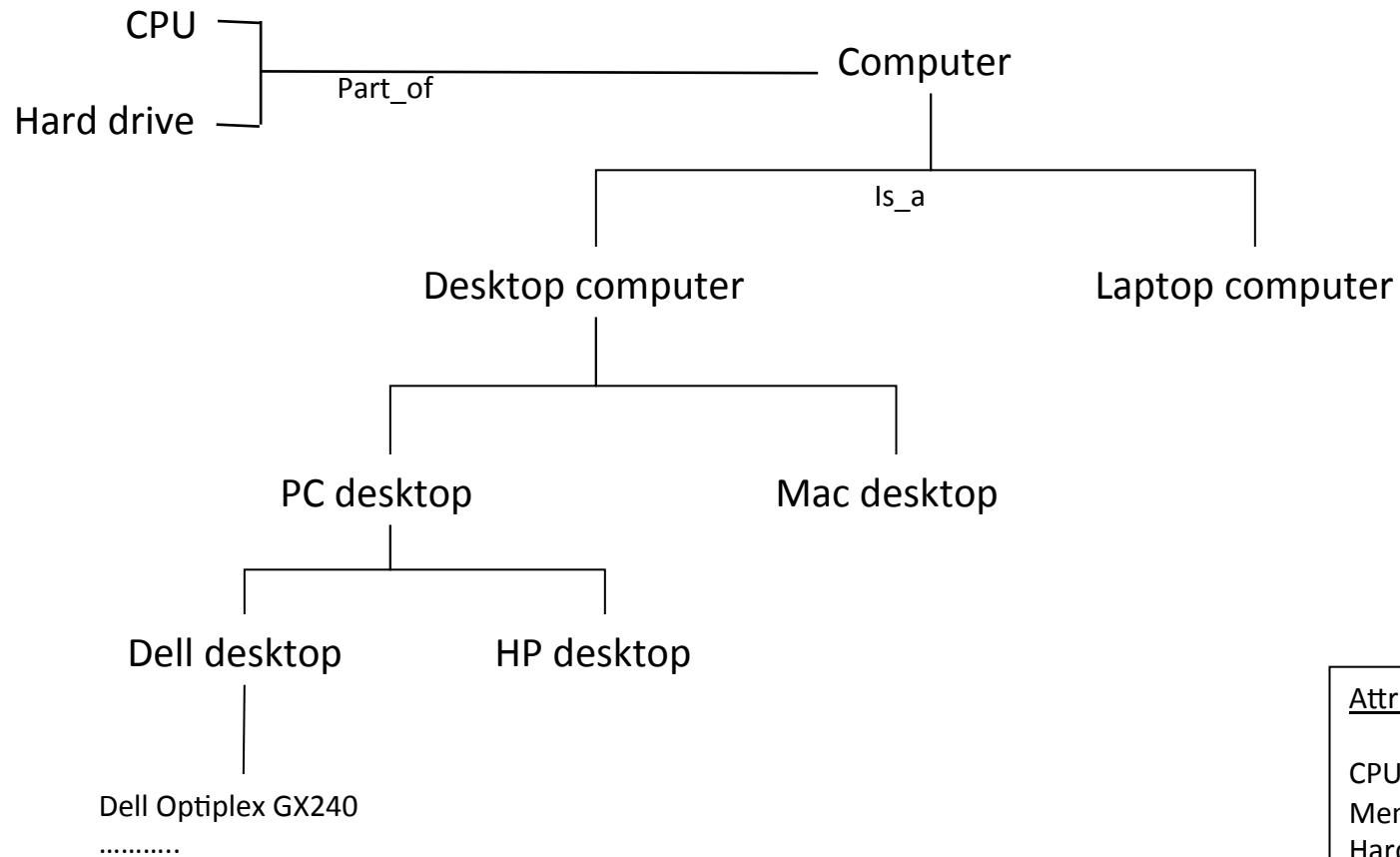
Example



Attributes:

CPU: Pentium 4 2GH
Memory: 1GB
Hard drive: 100GB

Example



Attributes:

CPU: Pentium 4 2GH
Memory: 1GB
Hard drive: 100GB

Outline

- Introduction of pathway databases
- Standards used in pathway databases
 - History of ontology
 - Ontology for biology
 - Pathway standards
 - SBML
 - BioPAX
 - SBGN

Ontology in Biological Knowledge -Gene Ontology

- The first ontology that was designed as a formal representation of biological knowledge
- Three knowledge domains:
 - molecular function
 - biological process
 - cellular component.

The screenshot shows the homepage of the Gene Ontology website. The header features the text "the Gene Ontology" and a search bar with the placeholder "gene or protein name". A sidebar on the left contains links to "Open menus", "Home" (which is highlighted in blue), "FAQ", "Downloads", "Tools", "Documentation", "About GO", "Projects", and "Contact GO". Below the sidebar is a navigation bar with icons for Home, About, Contact, Help, and Log In. The main content area has a heading "Welcome to the Gene Ontology website!". It describes the project's aim of standardizing gene and gene product attributes across species and databases, mentioning a controlled vocabulary of terms, gene product annotation data from GO Consortium members, and tools for access and processing. It encourages community input and contact. A section titled "Search the Gene Ontology Database" includes a search input field, a "GO!" button, and radio buttons for "gene or protein name" and "GO term or ID". At the bottom, it notes that AmiGO is the official GO browser and search engine, with a link to "Browse the Gene Ontology with AmiGO".

Hexokinase

Molecular function: GO:0004396 : hexokinase activity

Biological process: GO:0006096 : glycolysis

Cellular component: GO:0005829 : cytosol

Ashburner, M., et al. (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet.* 25: p. 25-9.

The Gene Ontology Consortium (2012) The Gene Ontology: enhancements for 2011. *Nucleic Acids Res.* 40:D559

- all : all [377130 gene products]
 - + GO:0008150 : biological_process [272439 gene products]
 - + GO:0009987 : cellular process [155181 gene products]
 - + GO:0044237 : cellular metabolic process [102936 gene products]
 - + GO:0044262 : cellular carbohydrate metabolic process [5325 gene products]
 - **GO:0044275 : cellular carbohydrate catabolic process [1748 gene products]**
 - + GO:0019405 : alditol catabolic process [66 gene products]
 - + GO:0046176 : aldonic acid catabolic process [3 gene products]
 - GO:0051692 : cellular oligosaccharide catabolic process [4 gene products]
 - GO:0044247 : cellular polysaccharide catabolic process [157 gene products]
 - GO:0046352 : disaccharide catabolic process [80 gene products]
 - GO:0019391 : glucuronoside catabolic process [1 gene product]
 - **GO:0006096 : glycolysis [1063 gene products]**
 - + GO:0016139 : glycoside catabolic process [72 gene products]
 - + GO:0006098 : pentose-phosphate shunt [367 gene products]
 - GO:0043471 : regulation of cellular carbohydrate catabolic process [71 gene products]
 - + GO:0006258 : UDP-glucose catabolic process [1 gene product]
 - GO:0044248 : cellular catabolic process [11927 gene products]
 - + GO:0044265 : cellular macromolecule catabolic process [6698 gene products]
 - **GO:0044275 : cellular carbohydrate catabolic process [1748 gene products]**
 - + GO:0019405 : alditol catabolic process [66 gene products]
 - + GO:0046176 : aldonic acid catabolic process [3 gene products]
 - GO:0051692 : cellular oligosaccharide catabolic process [4 gene products]
 - GO:0044247 : cellular polysaccharide catabolic process [157 gene products]
 - GO:0046352 : disaccharide catabolic process [80 gene products]
 - GO:0019391 : glucuronoside catabolic process [1 gene product]
 - **GO:0006096 : glycolysis [1063 gene products]**
 - + GO:0016139 : glycoside catabolic process [72 gene products]

GO biological process viewed in AmiGO

The Open Biomedical Ontologies (OBO) Foundry



The Open Biological and Biomedical Ontologies

Ontologies

Resources

Participate

About

The OBO Foundry is a collaborative experiment involving developers of science-based ontologies who are establishing a set of principles for ontology development with the goal of creating a suite of orthogonal interoperable reference ontologies in the biomedical domain. The groups developing ontologies who have expressed an interest in this goal are listed below, followed by other relevant efforts in this domain.

In addition to a listing of OBO ontologies, this site also provides a statement of the OBO Foundry principles, discussion fora, technical infrastructure, and other services to facilitate ontology development. We welcome feedback and encourage participation.

Click any column header to sort the table by that column. The s link to the term request trackers for the listed ontologies.

OBO Foundry ontologies

Title	Domain	Prefix	File	Last changed
Biological process	biological process	GO	go.obo	
Cellular component	anatomy	GO	go.obo	
Chemical entities of biological interest	biochemistry	CHEBI	chebi.obo	
Molecular function	biological function	GO	go.obo	
Phenotypic quality	phenotype	PATO	quality.obo	
PRotein Ontology (PRO)	proteins	PR	pro.obo	
Xenopus anatomy and development	anatomy	XAO	xenopus_anatomy_edit.obo	
Zebrafish anatomy and development	anatomy	ZFA	zebrafish_anatomy.obo	2013/02/07

OBO Foundry candidate ontologies and other ontologies of interest

Title	Domain	Prefix	File	Last changed
Adverse Event Reporting Ontology	health	AERO	aero.owl	
Amphibian gross anatomy	anatomy	AAO	AAO_v2_edit.obo	
Amphibian taxonomy	anatomy	ATO	amphibian_taxonomy.obo	
Anatomical Entity Ontology	anatomy	AEO	aea.obo	2012/06/01

To establish a set of principles for ontology development with the goal of creating a suite of orthogonal interoperable reference ontologies in the biomedical domain.

Basic Formal Ontology (BFO)

 **Basic Formal Ontology (BFO)** 

[Home](#) | [Overview](#) | [Manual](#) | [Related Material](#) | [Users](#) | [IFOMIS](#)

Home

BFO grows out of a philosophical orientation which overlaps with that of [DOLCE](#) and [SUMO](#). Unlike these, however, it is narrowly focused on the task of providing a genuine upper ontology which can be used in support of domain ontologies developed for scientific research, as for example in biomedicine within the framework of the [OBO Foundry](#). Thus *BFO* does not contain physical, chemical, biological or other terms which would properly fall within the special sciences domains.

The theory behind *BFO* has been developed and formulated by [Barry Smith](#) and [Pierre Grenon](#) in a series of publications. Its development has been carried out under the auspices of the project *Forms of Life* sponsored by the Volkswagen Foundation.

News

- CTG's MedMaP Technology uses open source and commercial ontologies to enable rapid analysis of biomedical data. MedMaP is a BFO-driven Rich Internet Application (RIA), where medical information pertaining to disease risks, patient registries, blood chemistries, patient education/compliance and value-based outcomes can be ontologically classified and presented in a variety of visual formats on the basis of user preferences (medical specialties, areas of research, disease states, etc.). (For further details of the AMIA presentation of this technology see [here](#))
- ChEBI chemistry ontology development (based on BFO) funded by BBSRC ([more](#))

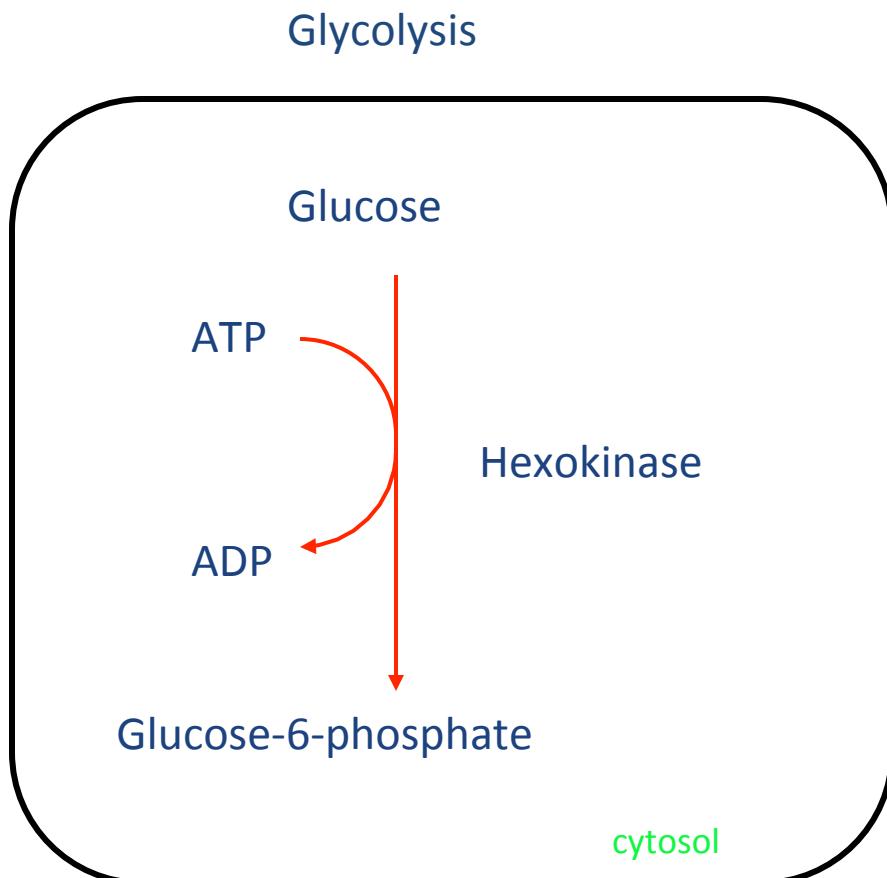
Implementations

(Current Version: 1.1.1)

- [OWL](#) - description logic based, created by [Holger Stenzhorn](#) with large contributions from [Andrew Spear](#) and others (namespace: <http://www.ifomis.org/bfo/1.1#>) ([HTML view](#))
- [Isabelle](#) - first-order logic based, created by [Thomas Bittner](#)
- [OBO](#) - created by [Chris Mungall](#)
- [Prover9/KIF](#) - BFO merged with RO, created by [Chris Mungall](#) (see also [here](#))

Arp R. and Smith B (2008) Function, Role, and Disposition in Basic Formal Ontology. *Nature Precedings*: hdl:10101/npre.2008.1941.1

Biological Pathway Representations



- Pathway data are usually represented by three major classes
 - Molecules
 - Reactions
 - Location
- GO Biological Process ontology does not capture
 - all the dynamic relationships in the pathways
 - quantitative and kinetic data

Pathway Standards Emerged

- Systems Biology Markup Language (SBML)
- Systems Biology Graphical Notation (SBGN)
- BioPAX - A **BIO**logical **P**Athways **eX**change language
- CellML
- More

- Hucka M et. al. (2004) Evolving a lingua franca and associated software infrastructure for computational systems biology: the Systems Biology Markup Language (SBML) project. *Syst. Biol.* 1:41-53.
- Demir E. et. al. (2010) BioPAX – A Community Standard for Pathway Data Sharing. *Nature Biotechnol.*, 28:935–942
- Le Novère N et al. (2009). The Systems Biology Graphical Notation. *Nat Biotechnol.*, 27 (8): 735–41
- Gurny A et. al. (2008) CellML and associated tools and techniques. *Philos Transact A Math Phys Eng Sci.* 366:3017-43.

Outline

- Introduction of pathway databases
- Standards used in pathway databases
 - History of ontology
 - Ontology for biology
 - Pathway standards
 - SBML (http://sbml.org/Main_Page)
 - slides kindly provided by Dr. Hucka
 - BioPAX
 - SBGN

The first formal standard

Format for representing computational models of biological processes

- Data structures + usage principles + serialization to XML

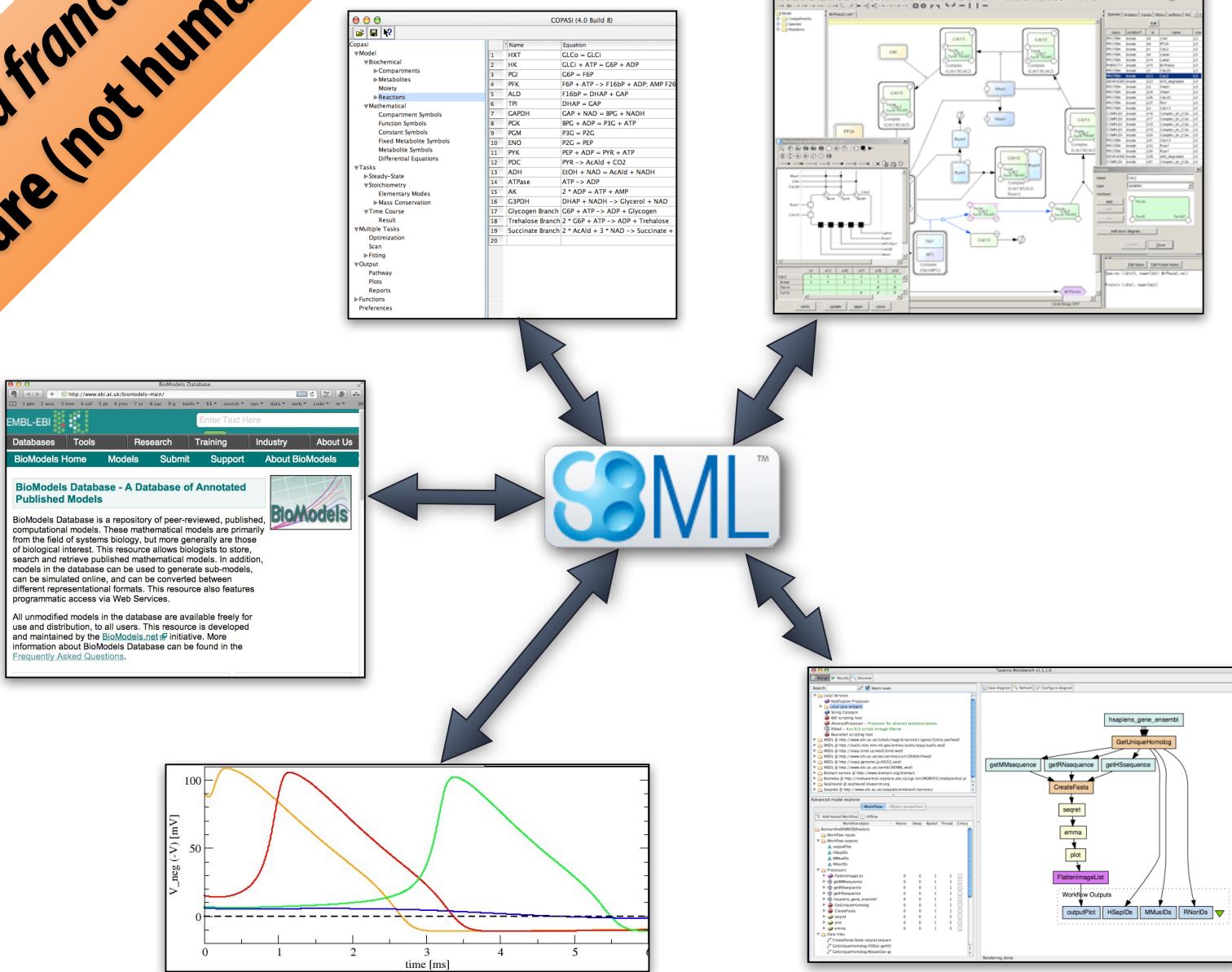
Neutral with respect to modeling framework

- E.g., ODE, stochastic systems, etc.

Development started in 2000, with first specification distributed in 2001

- XML was still relatively new, RDF even more so

A lingua franca for software (not humans)



Basic SBML concepts are fairly simple

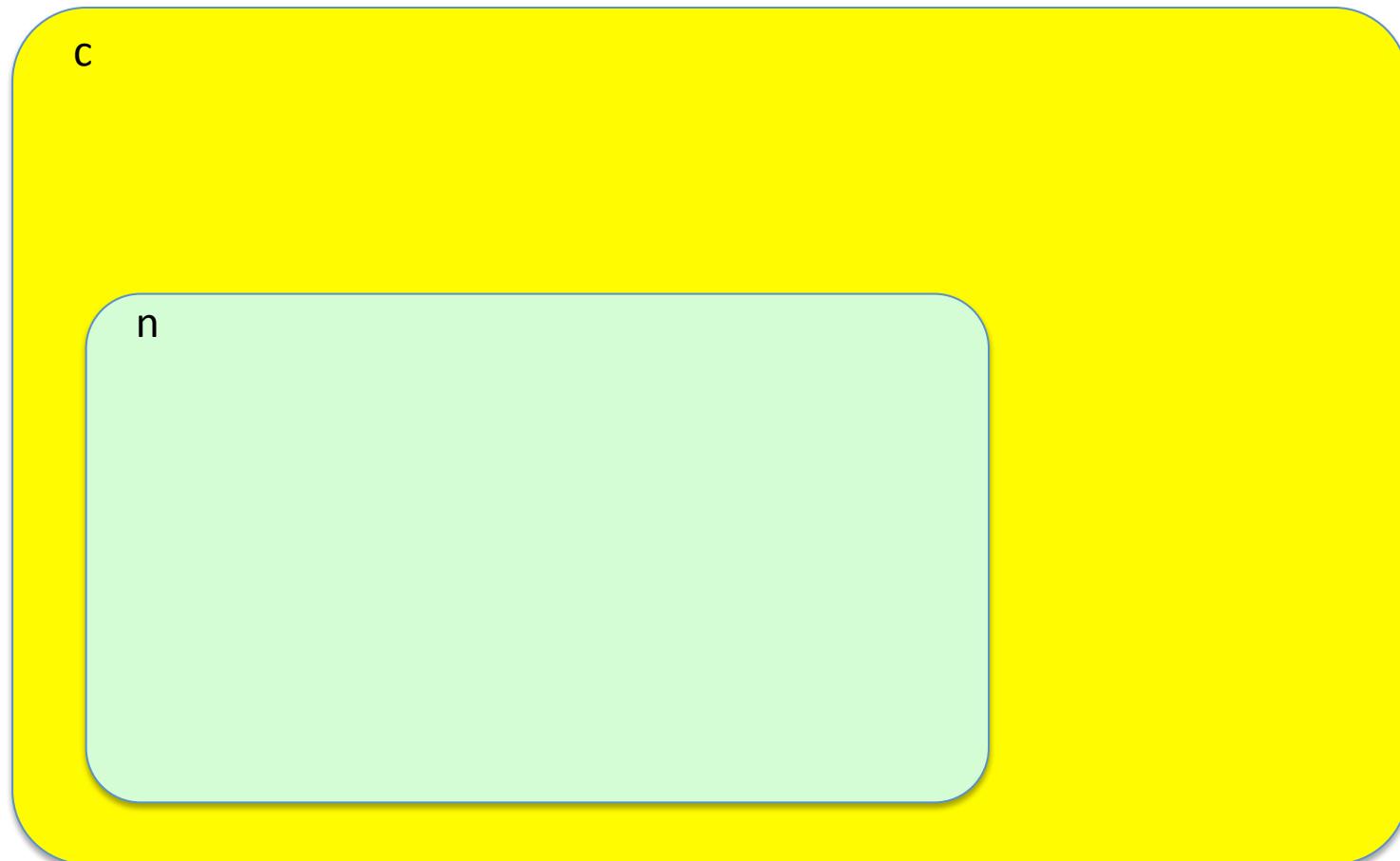
The **process** is central

- Called a “reaction” in SBML
- Participants are pools of entities (**species**)

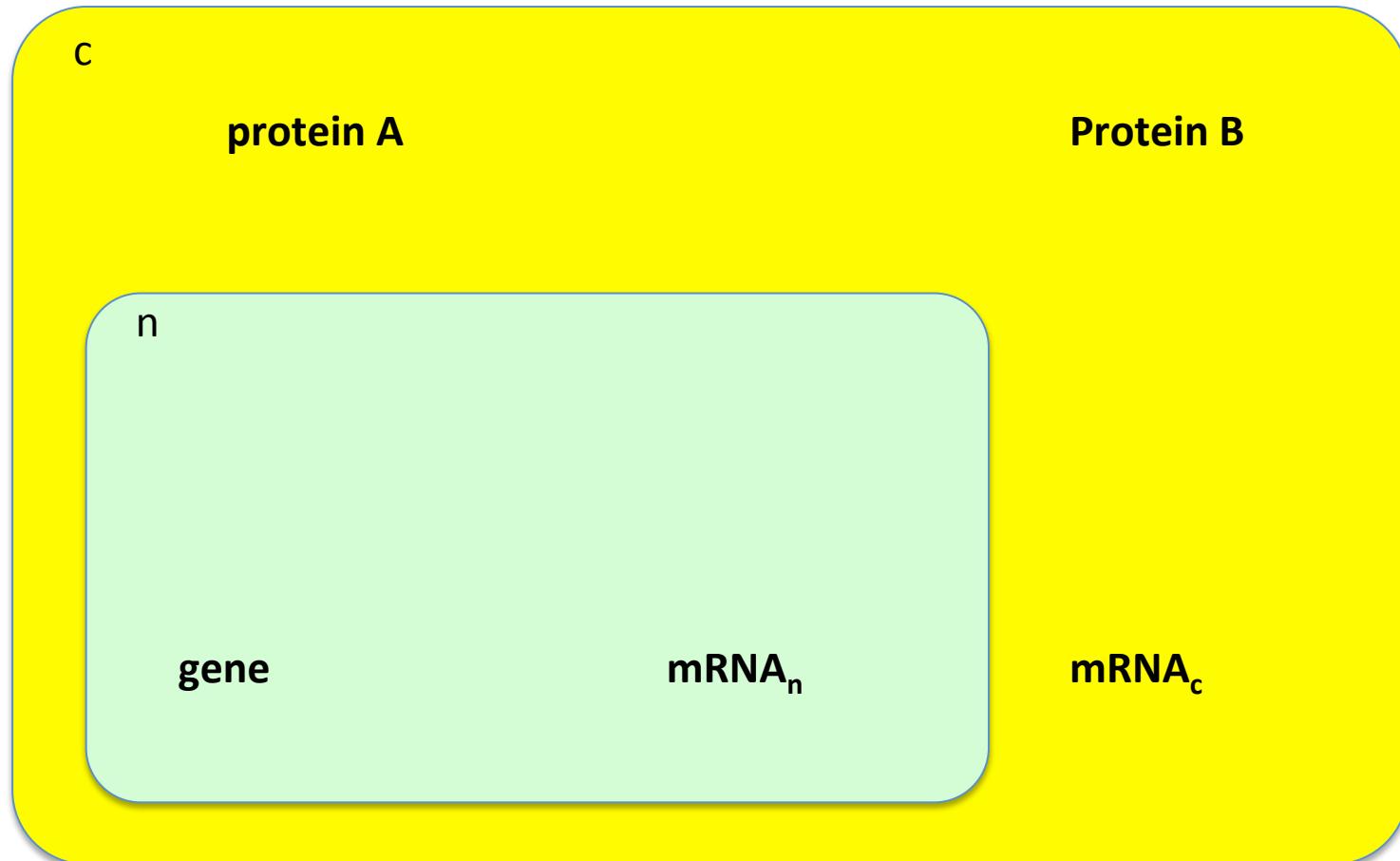
Models can further include:

- Other constants & variables
- Compartments
- Explicit math
- Discontinuous events
- Unit definitions
- Annotations

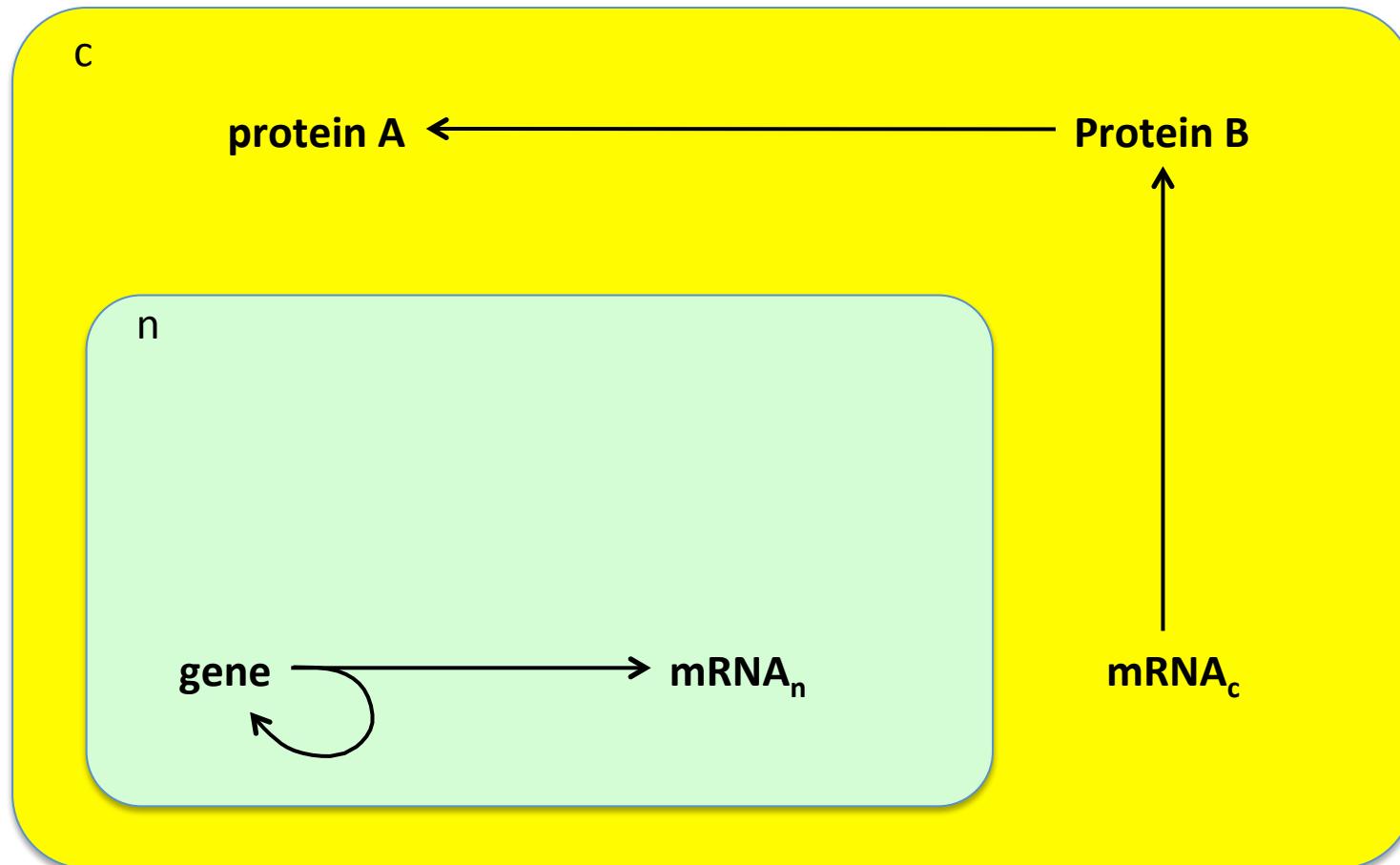
Well-stirred compartments



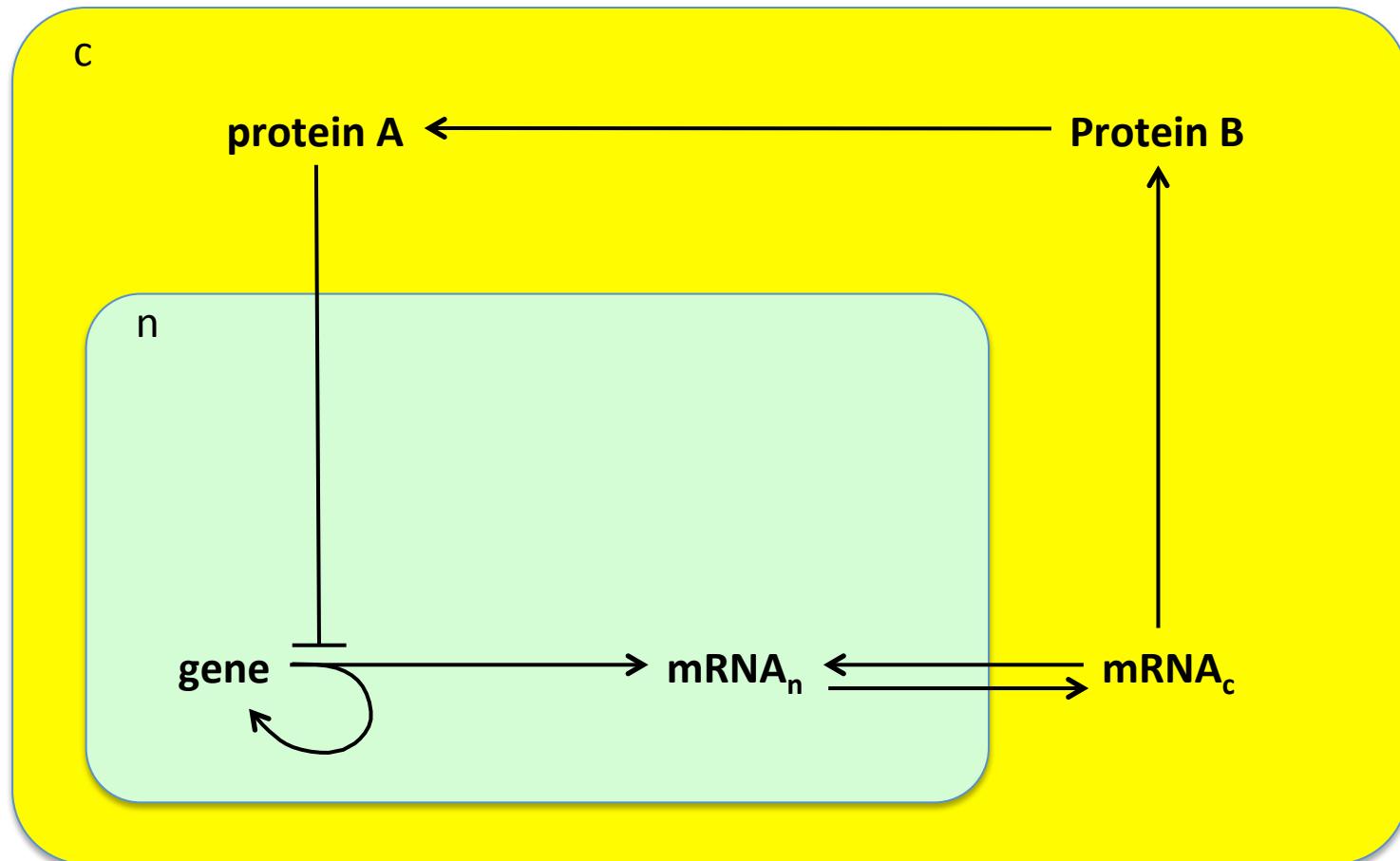
Species pools are located in compartments



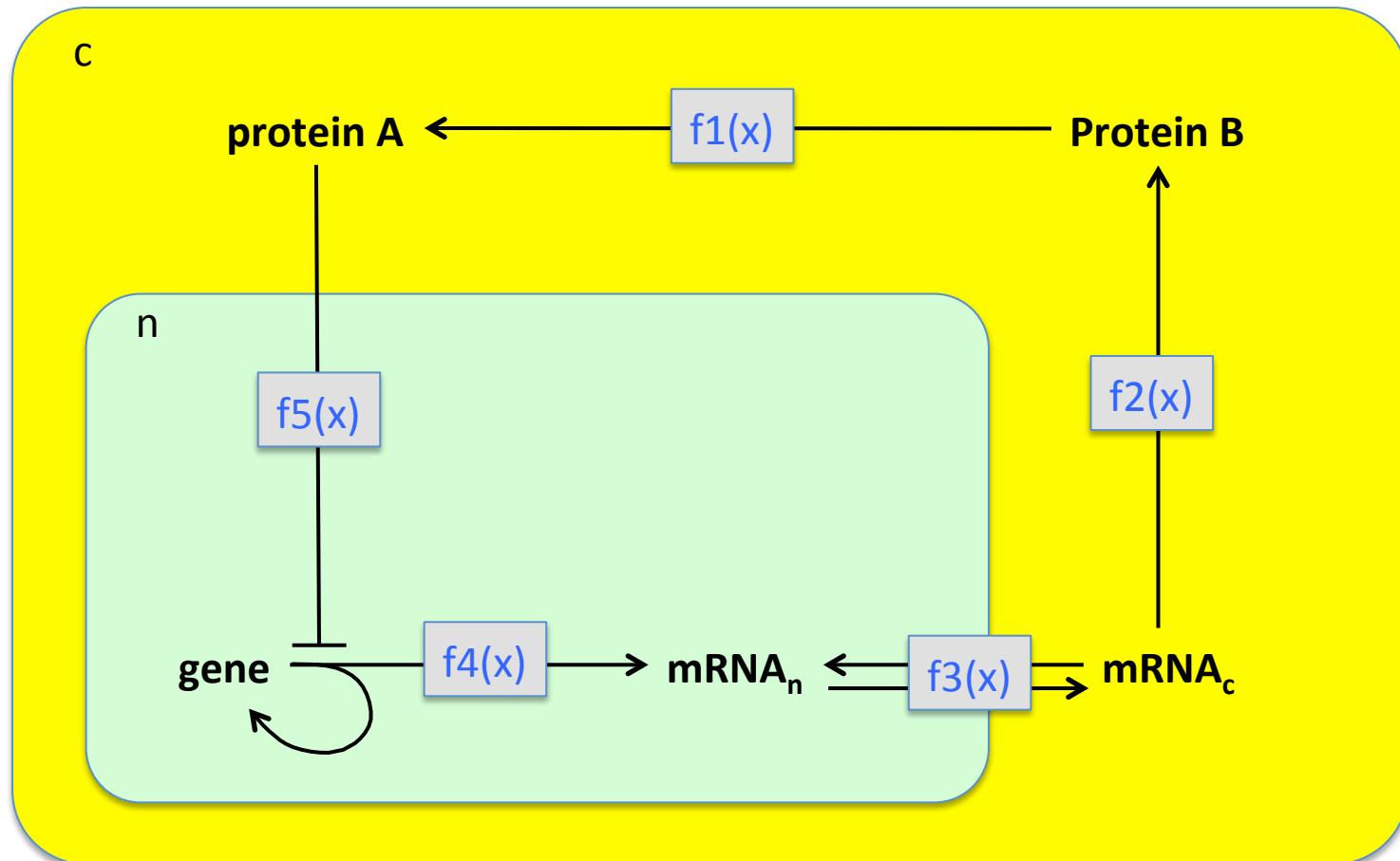
Reactions can involve any species anywhere



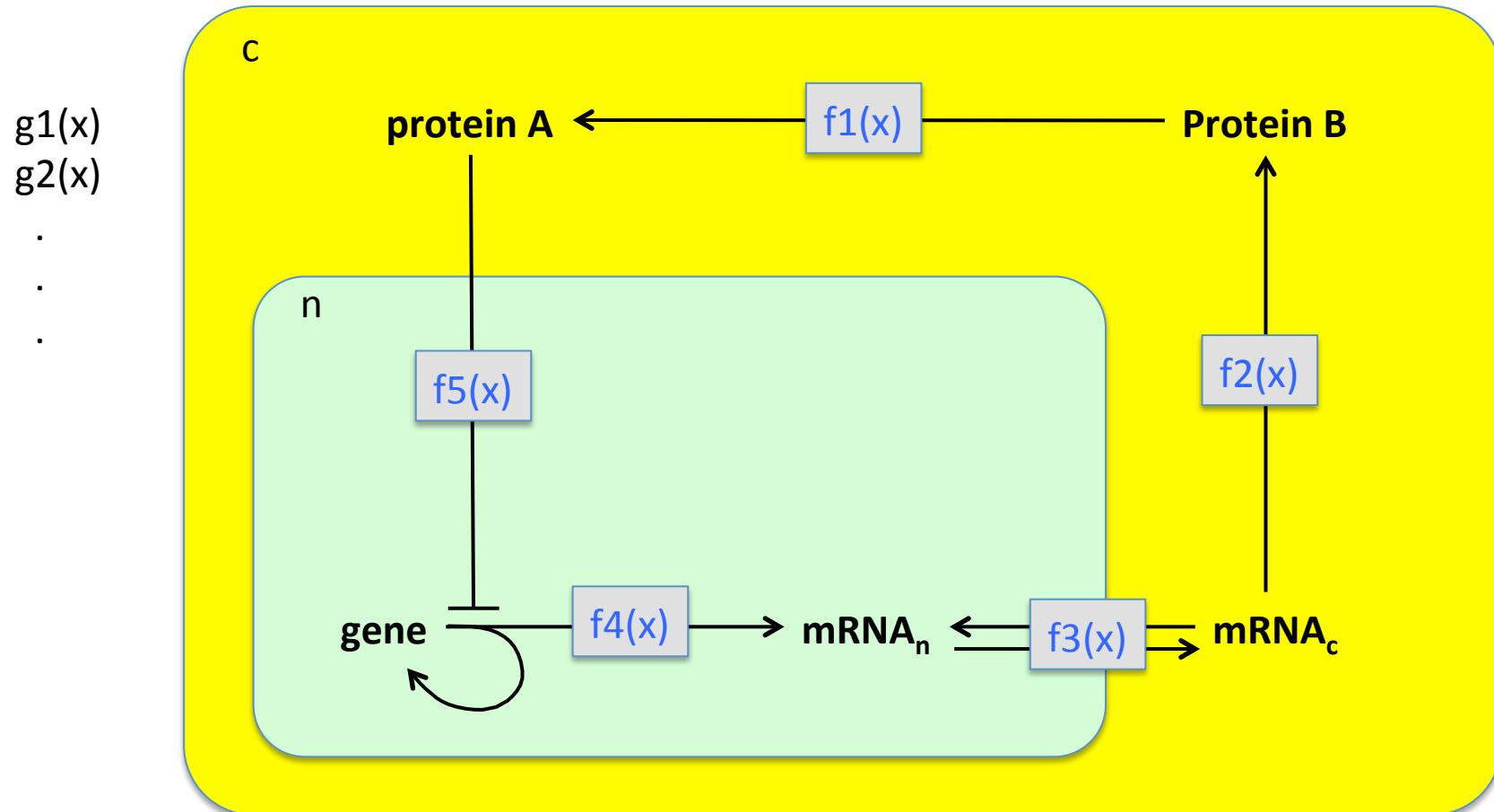
Reactions can cross compartment boundaries



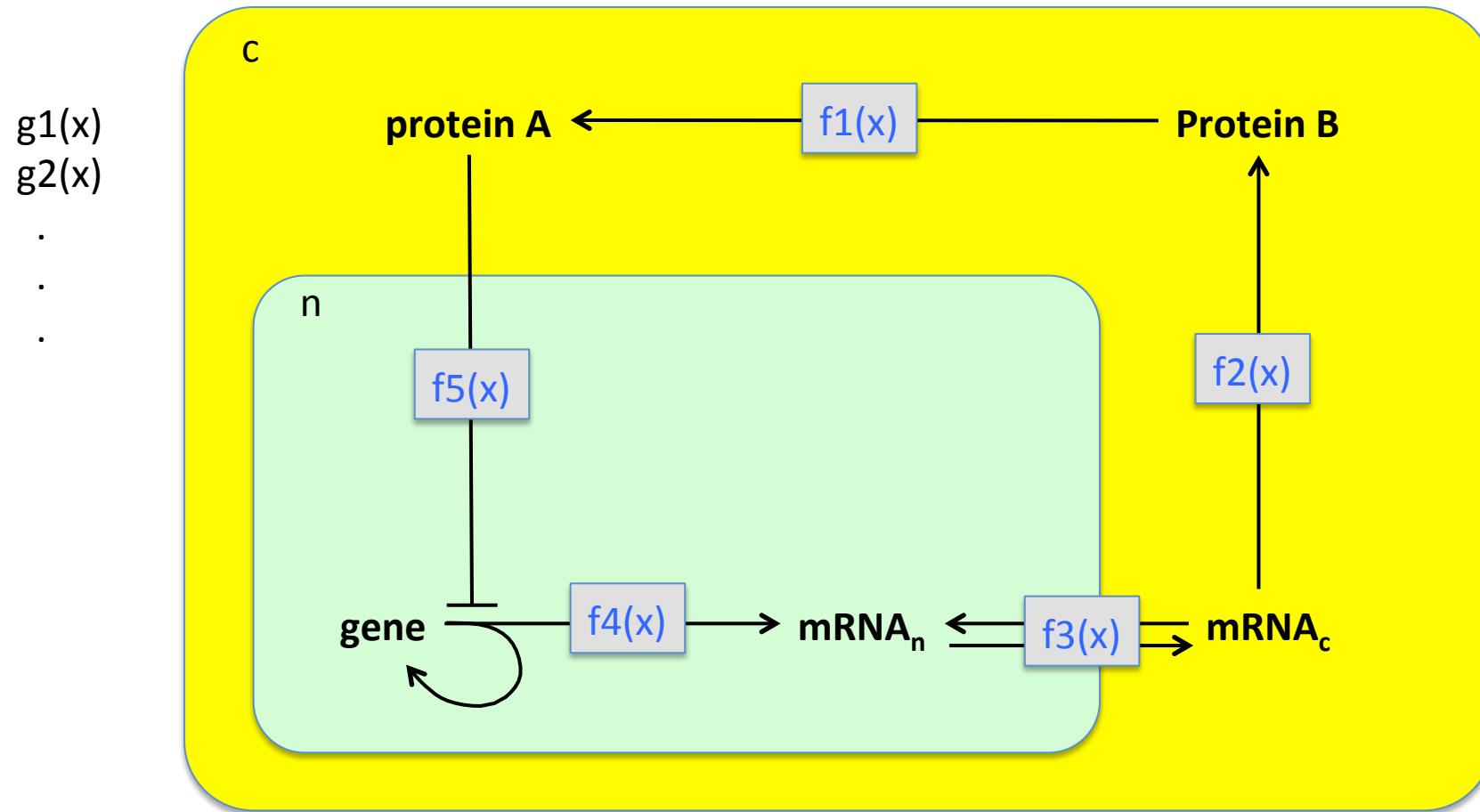
Reaction/process rates can be (almost) arbitrary formulas



“Rules”: equations expressing relationships in addition to reaction sys.



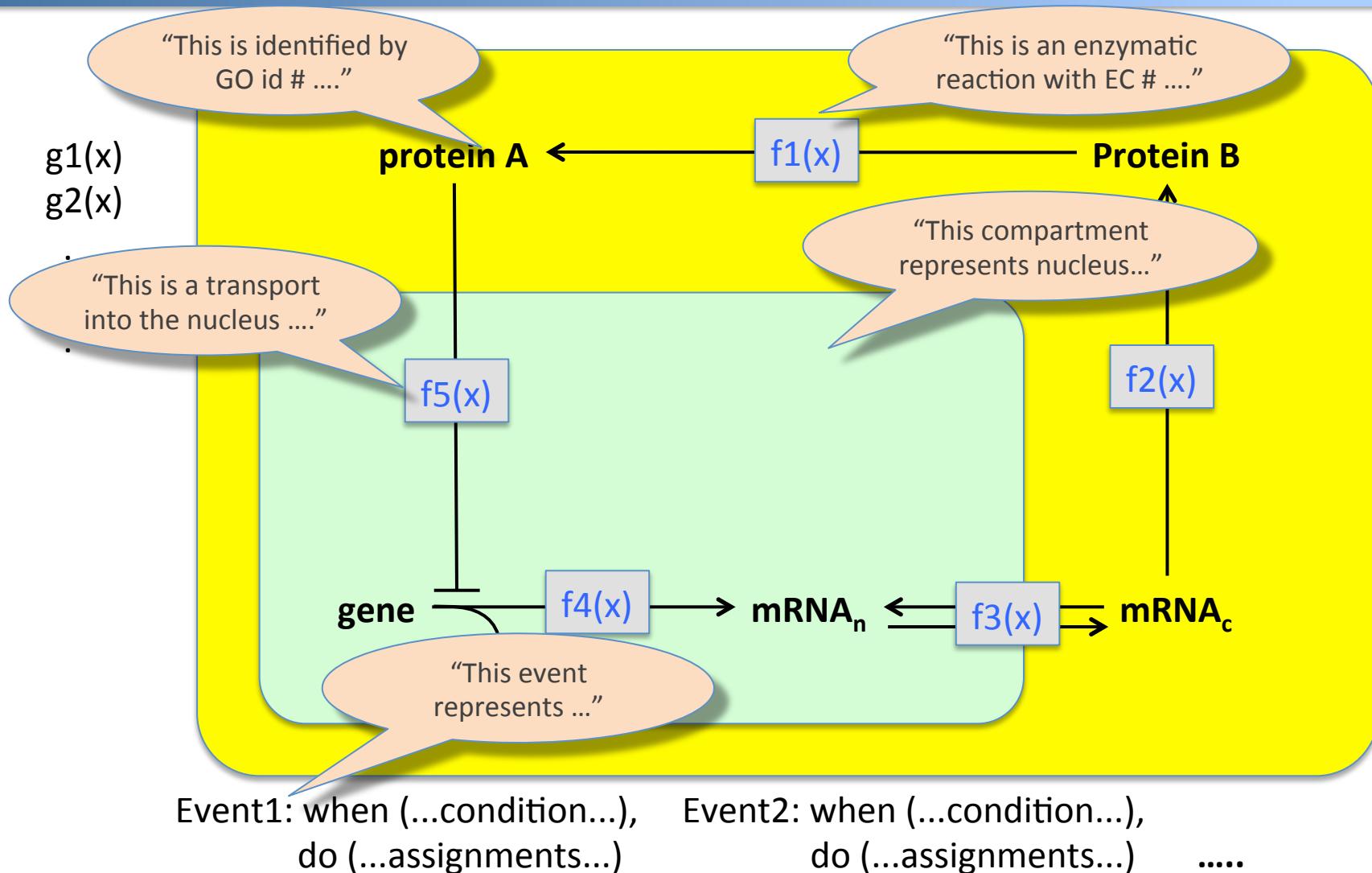
“Events”: discontinuous actions triggered by system conditions



Event1: when (...condition...),
do (...assignments...)

Event2: when (...condition...),
do (...assignments...)

Annotations: machine-readable semantics and links to other resources



Scope of SBML encompassing many types of models

Today: spatially homogeneous models

- Metabolic network models
- Signaling pathway models
- Conductance-based models
- Neural models
- Pharmacokinetic/dynamics models
- Infectious diseases



Parent pages: [SBML.org](#)

Community

SBML continues to improve and evolve thanks to the involvement of a vibrant and active community of users, developers and researchers. You can get involved too! Follow the links below to find out more.



Over 250 software packages today support SBML. Visit the [SBML Software Guide](#) for a list of different software systems and their features.



The [SBML community wiki](#) contains recent material and working notes from people in the SBML community.



The mailing lists are where questions are asked and the latest developments discussed. The [forums area](#) provides a web interface to the lists, and you can also subscribe to any lists you find interesting.



People sometimes contribute simple utilities and scripts for working with SBML. We gather them in a special [Contributed Programs](#) area of the community wiki.



Regular events are such a big part of the SBML community that there is a [separate area](#) of this site devoted to them. Check there to find out about events coming up.



Announcements by the SBML community also have their own area on this site—visit the [News](#) page. (Don't forget to [send us](#) your SBML-related announcements!)

You may also want to look at the [News](#) section to learn about recent announcements of interest to SBML users and developers worldwide.

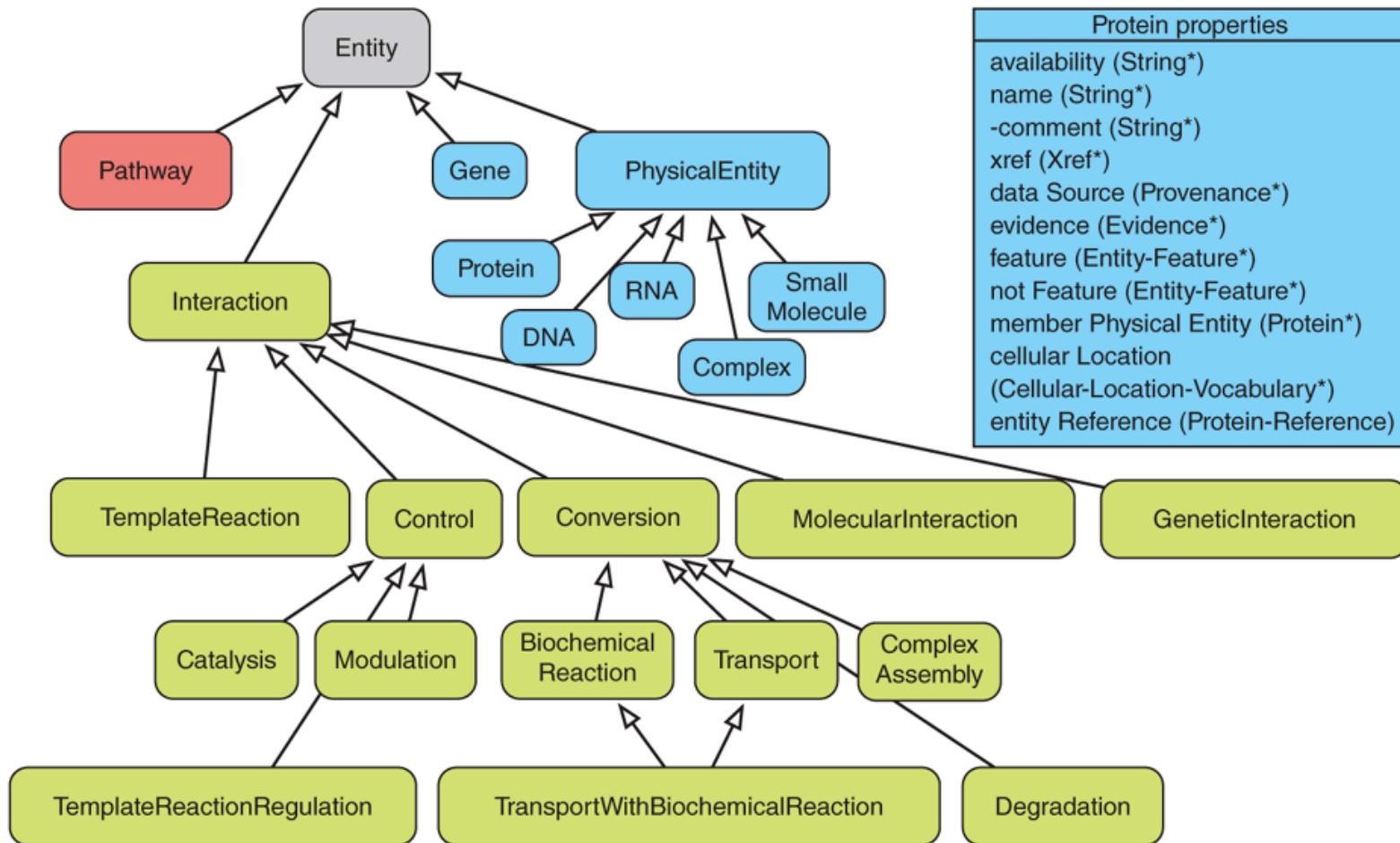
Outline

- Introduction of pathway databases
- Standards used in pathway databases
 - History of ontology
 - Ontology for biology
 - Pathway standards
 - SBML
 - BioPAX (www.biopax.org)
 - slides kindly provided by Dr. Demir
 - SBGN

- A standard exchange format to represent biological pathways at the molecular and cellular level.
- Major use is to facilitate the exchange of pathway data.
- In owl/RDF format

Demir E. et. al. (2010) BioPAX – A Community Standard for Pathway Data Sharing. *Nature Biotechnol.*, 28:935–942

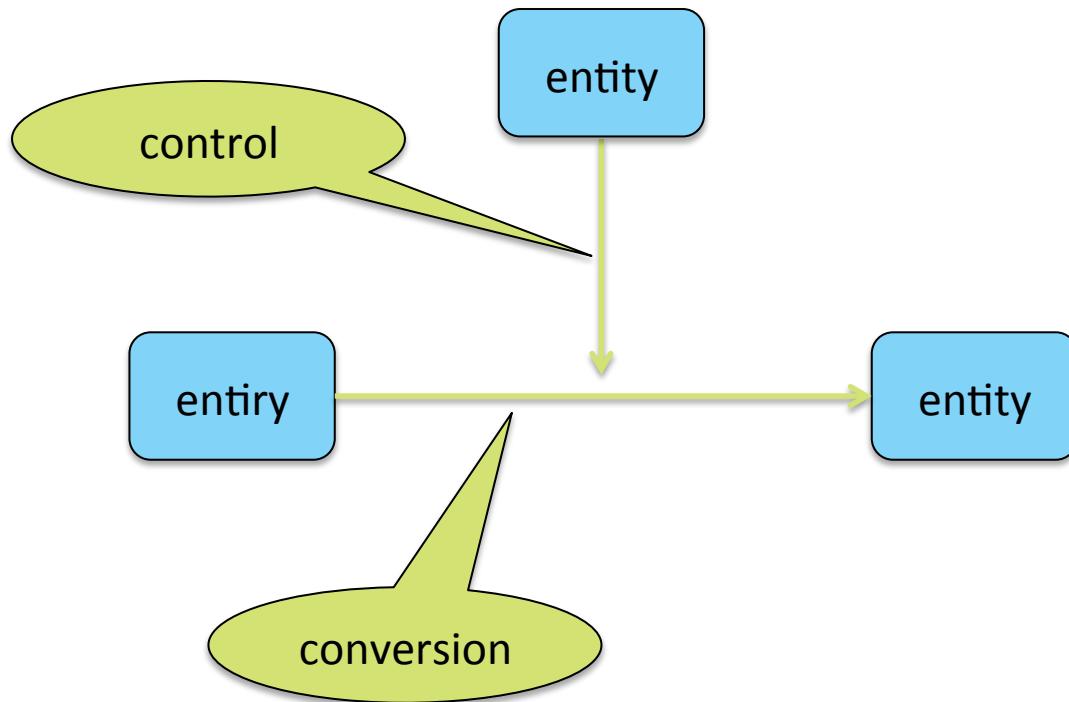
BioPAX ontology structure



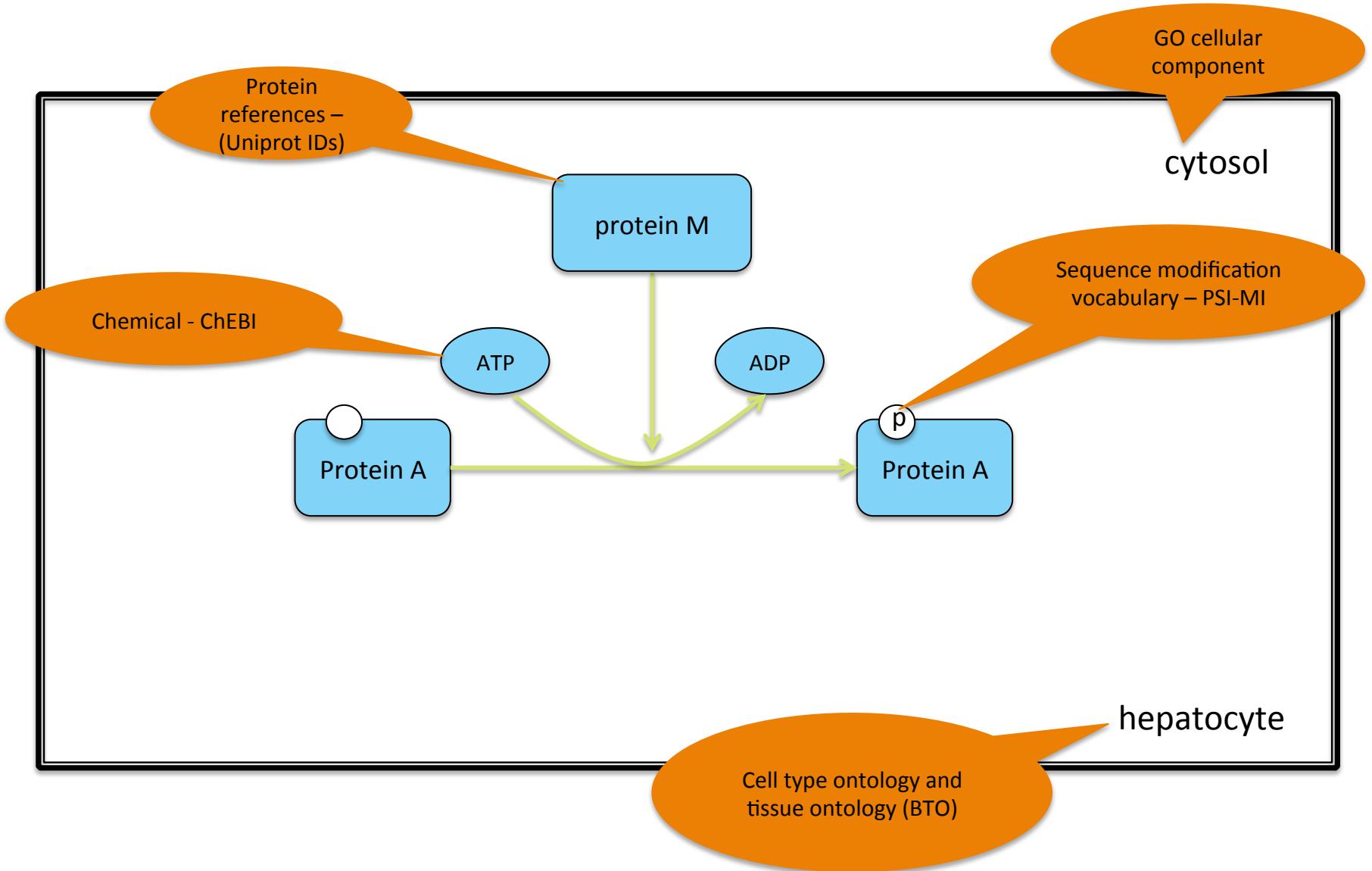
Classes, shown as boxes and arrows, represent inheritance relationships. The three main types of classes in BioPAX are Pathway (red), Interaction (green) and PhysicalEntity and Gene (blue). For brevity, the properties of the Protein class only are shown as an example at the top right. Asterisks indicate that multiple values for the property are allowed. Refer to BioPAX documentation at <http://www.biopax.org/> for full details of all classes and properties.

Figure 4. of Demir et. al. (2010)

BioPAX data – different meanings of interaction

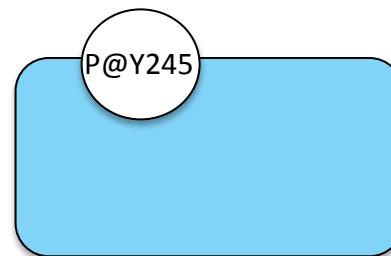


BioPAX – link to external resources and ontologies

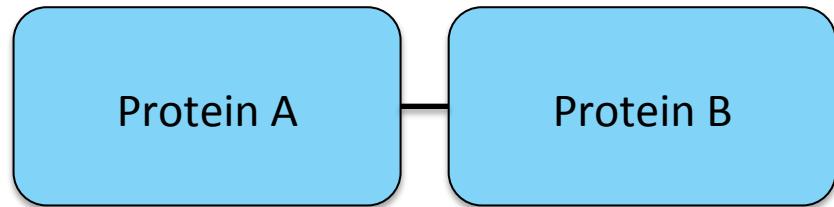


BioPAX captures more detailed molecular features

Protein modification feature



Binding feature



BioPAX - glycolysis viewed in protégé

The screenshot shows the Protégé interface with the following panels:

- Class hierarchy (Protein)**: Shows the class hierarchy under the Protein category. Root classes include Thing, Entity, Interaction, PhysicalEntity, UtilityClass, and various subclasses like Gene, Control, Conversion, BiochemicalReaction, ComplexAssembly, Degradation, Transport, GeneticInteraction, MolecularInteraction, TemplateReaction, Pathway, and Protein.
- Members list: _Hexokinase_s11_a9**: Shows a list of members related to the protein Hexokinase_s11_a9. The list includes Aldolase_s15_a19, Enolase_s30_a33, Glyceraldehyde_3_phosphate_deh, Hexokinase_s11_a9 (selected), Phosphofructokinase_1_s13_a16, Phosphoglucose_isomerase_s12, Phosphoglycerate_kinase_s24_a2, Phosphoglyceromutase_s29_a31, Pyruvate_kinase_s35_a36, and Triosephosphate_isomerase_s16.
- Usage: _Hexokinase_s11_a9**: Shows the usage of the protein Hexokinase_s11_a9. It lists 7 uses, primarily involving catalysis in reactions like Hexokinase_r5m1_r5 and Hexokinase_r5m1_r5 controller.
- Property assertions: _Hexokinase_s11_a9**: Shows the property assertions for the protein. It includes:
 - entityReference** _Hexokinase_PROTEIN
 - Data property assertions:
 - displayName "Hexokinase"^^string
 - standardName "Hexokinase"^^string
 - comment "SPECIES_TYPE=PROTEIN"^^string
 - comment "ACTIVATION=FALSE"^^string
 - Negative object property assertions
 - Negative data property assertions

BioPAX : Biological Pathway eXchange

BioPAX is a standard language for the exchange, visualization and analysis of biological pathways. It supports data exchange between different pathway databases and reduces the complexity of interchange between data sources. BioPAX is an accepted standard format for pathway data. It is an open and collaborative effort by the community of researchers, software developers, and institutions. BioPAX is defined in [OWL DL](#) and is represented in the RDF/XML format. [BioPAX Paper](#) was published in Nature Biotechnology in 2010.

Software package to
create BioPAX file

Coordination

- [The BioPAX Forum](#) (BioPAX Discussion Google Group).
- [BioPAX Level3 Announcement](#)
- [BioPAX Specification and Documentation](#)
- [BioPAX Workgroups](#)
- [BioPAX Proposals](#)
- Next Workshop: [Harmony 2013](#)
- Last Workshop: [COMBINE 2012 Toronto, ON](#)
- [FAQ](#)
- [Archive](#)
 - [Previous Presentations and Publications](#)
 - [Past meetings](#)
 - [Level 3 BioPAX Workgroup Coordination](#)
 - [biopax-discuss](#)
 - [biopax-boston](#)
 - [Original Wiki Archive](#)

BioPAX paper in Nature
Biotechnology

+ Special Section



_computation
TOLOGY

Research Article mir et al., FREE

bioRxiv preprint doi: https://doi.org/10.1101/2010.02.10.193375; this version posted February 10, 2010. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a [CC-BY-NC-ND 4.0 International license](#).

BioPAX validator

bioRxiv preprint doi: https://doi.org/10.1101/2010.02.10.193375; this version posted February 10, 2010. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a [CC-BY-NC-ND 4.0 International license](#).

Recent Wiki Edits

- [Main Page](#)
- [Paxtools](#)
- [Main Page](#)
- [Data](#)
- [Presentations and Publications](#)
- [Specification](#)
- [Main Page](#)



Search and visualize public biological pathway information. Single point of access.
[\[more...\]](#)

[Home](#) | [Data Sources](#) | [Download](#) | [FAQ](#) | [Web Service](#) | [About](#)

Send us your [feedback](#). Sign up for Pathway Commons [announcements](#). [RSS Feed](#)

Search Pathway Commons:

[Find Pathways](#)

[Find Molecules](#)

[Search](#)

For example, if you enter: [BRCA1](#), you will **get back the list of pathways** containing the keyword "BRCA1", and the list of pathways that contain the BRCA1 gene.

Current filter settings: All Organisms, All Data Sources. [Set filters](#).

What's New:

- **NEW!** Oct 27, 2011:
 - BioGRID data set (September 25, 2011 Version 3.1.81).
 - IntAct data set (September 29, 2011 Version 3.1.17288).
 - Nature Pathway Interaction data set (October 12, 2011).
 - Reactome data set (September 20, 2011 Version 38).
- June 24, 2011:
 - BioGRID data set (May 1, 2011 Version 3.1.76).
 - HumanCyc data set (June 8, 2011 Version 15.1).

Using Pathway Commons:

Biologists: Browse and search pathways across multiple valuable public pathway databases.

Computational biologists: Download an integrated set of pathways in BioPAX format for global analysis.

Software developers: Build software on top of Pathway Commons using our [web service API](#). Download and install the [cPath software](#) to create a local mirror.

Current Data Sources:

Pathway Commons currently contains the following data sources ([batch download](#)):

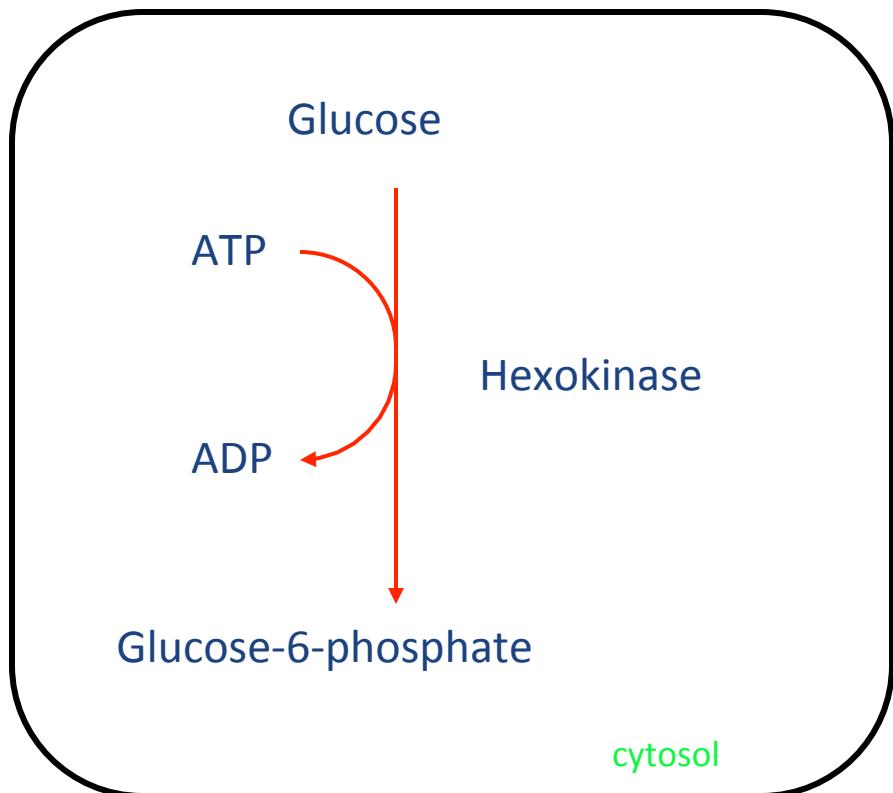
Pathway Commons – A repository for pathway data in validated BioPAX format

Outline

- Introduction of pathway databases
- Standards used in pathway databases
 - History of ontology
 - Ontology for biology
 - Pathway standards
 - SBML
 - BioPAX
 - SBGN (www.sbgn.org)

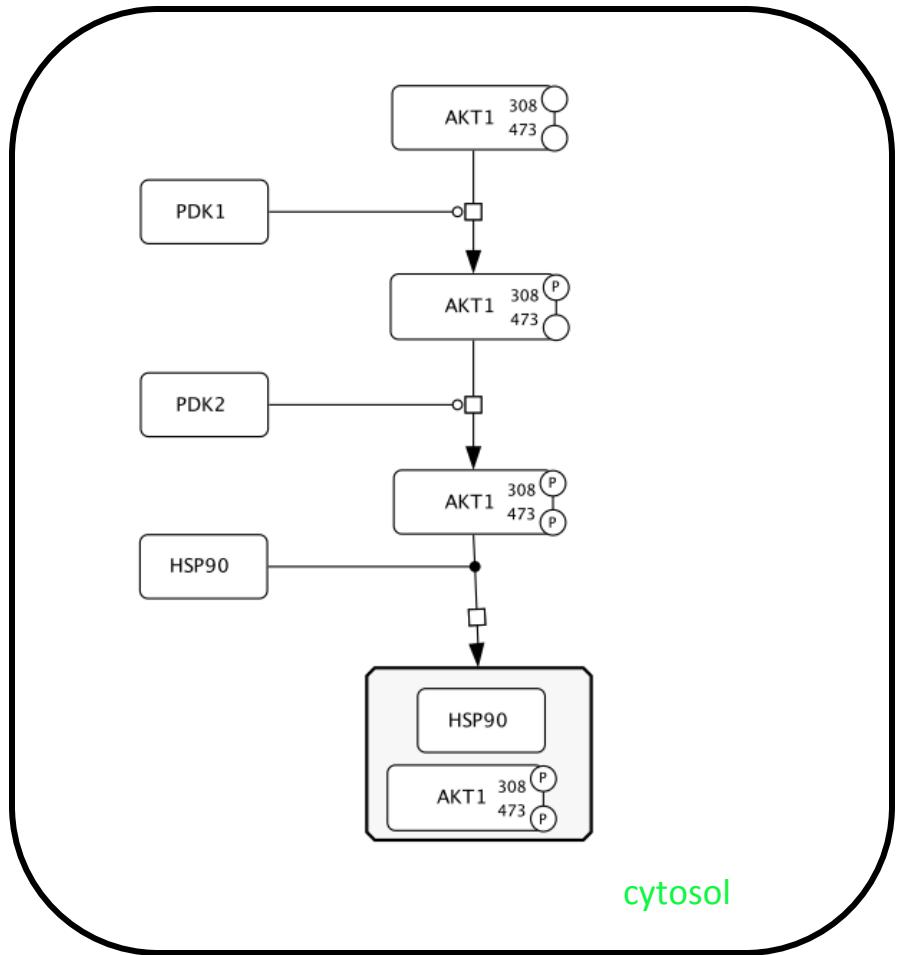
Pathway Network Diagram

– An easy way to read biological knowledge



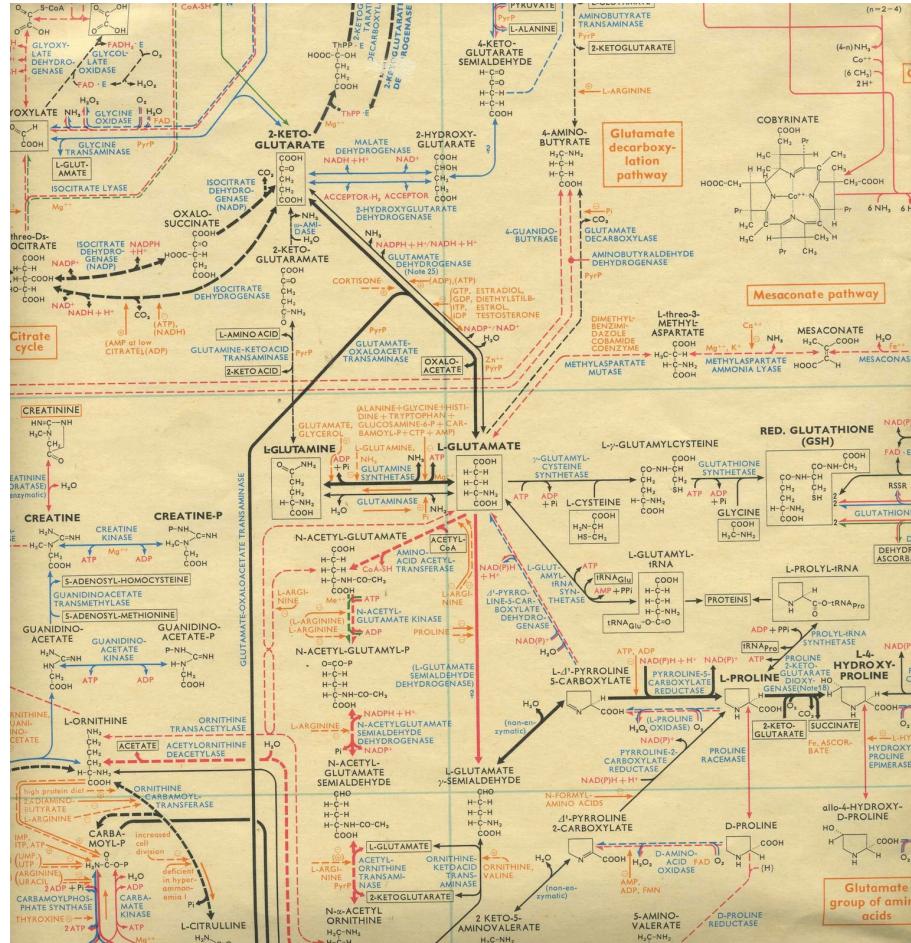
- Glucose is converted to glucose -6-phosphate with the consumption of ATP and production of ADP. The process is catalyzed by hexokinase. The reaction occurs in the cytosol of a cell.

Pathway Network Diagram – An easy way to read biological knowledge



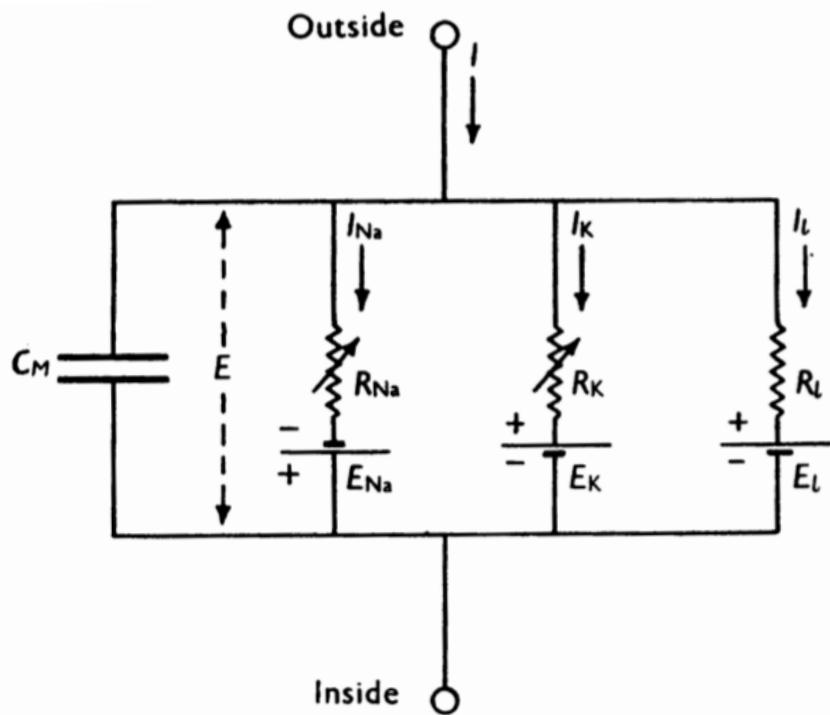
AKT1 is phosphorylated at residue 308 by PDK1. The phosphorylated AKT is then phosphorylated at residue of 473 by PDK2. The second phosphorylation reaction does not happen until the first residue (308) is phosphorylated. When both sites are phosphorylate. All the reactions occur in cytosol.

Pathway diagram has been used a long time ago



A metabolic pathway diagram

From the wall chart of *Biochemical Pathways* created by Gerhard Michal (1968)



Electrical circuit diagram
representing cell membrane.

From Hodgkin A.L. and Huxley A.F. (1952) A quantitative description of membrane current and its application to conduction and excitation in nerve. *J. Physiol.* 117:500-544

Standardized symbols are important



Most English
Speaking country



Quebec



Iran



China



Israel



Singapore



Norway



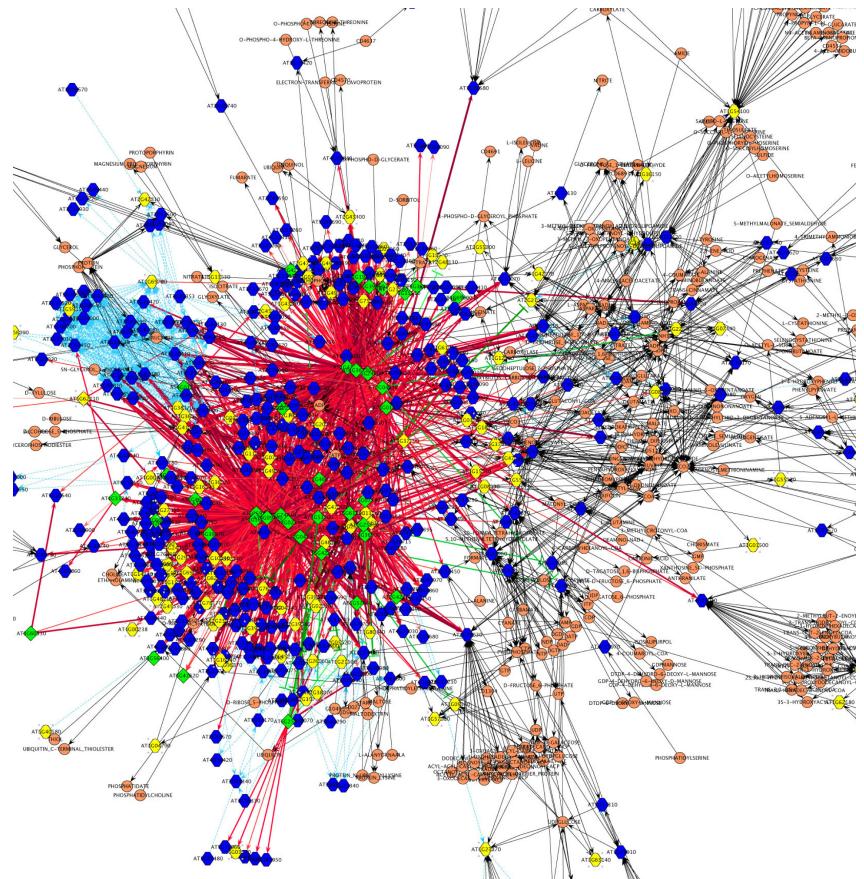
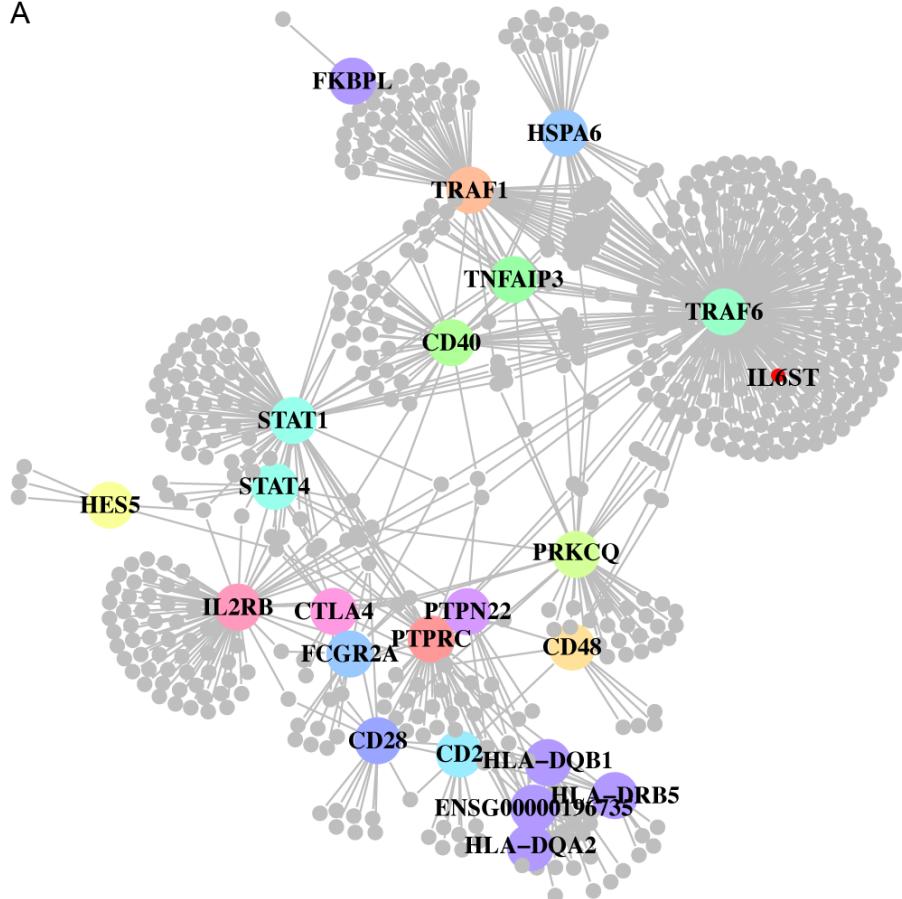
Poland



USA and
Canada

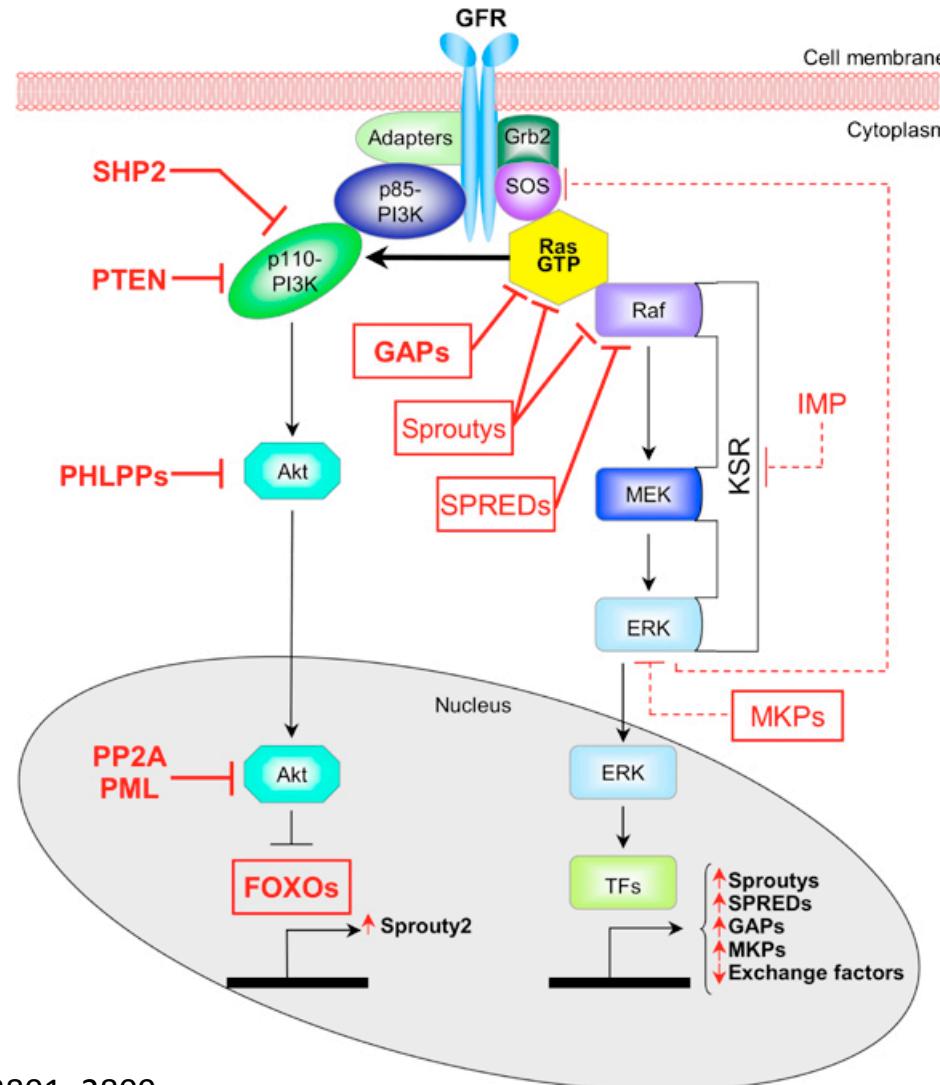
Diagrams generated by computer

A



EJ Rossin, *PLoS Genetics*, 2011

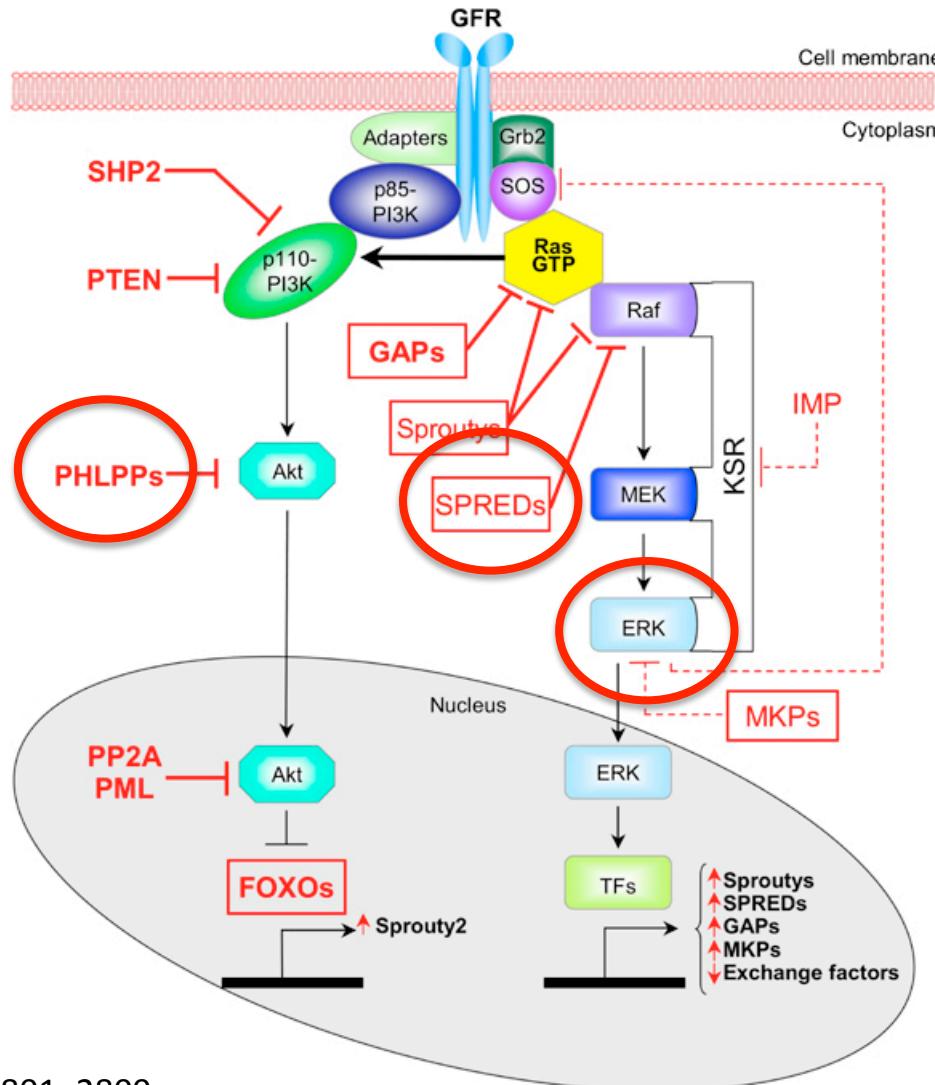
Diagram generated by biologist



Review

Oncogene (2008) 27, 2801–2809

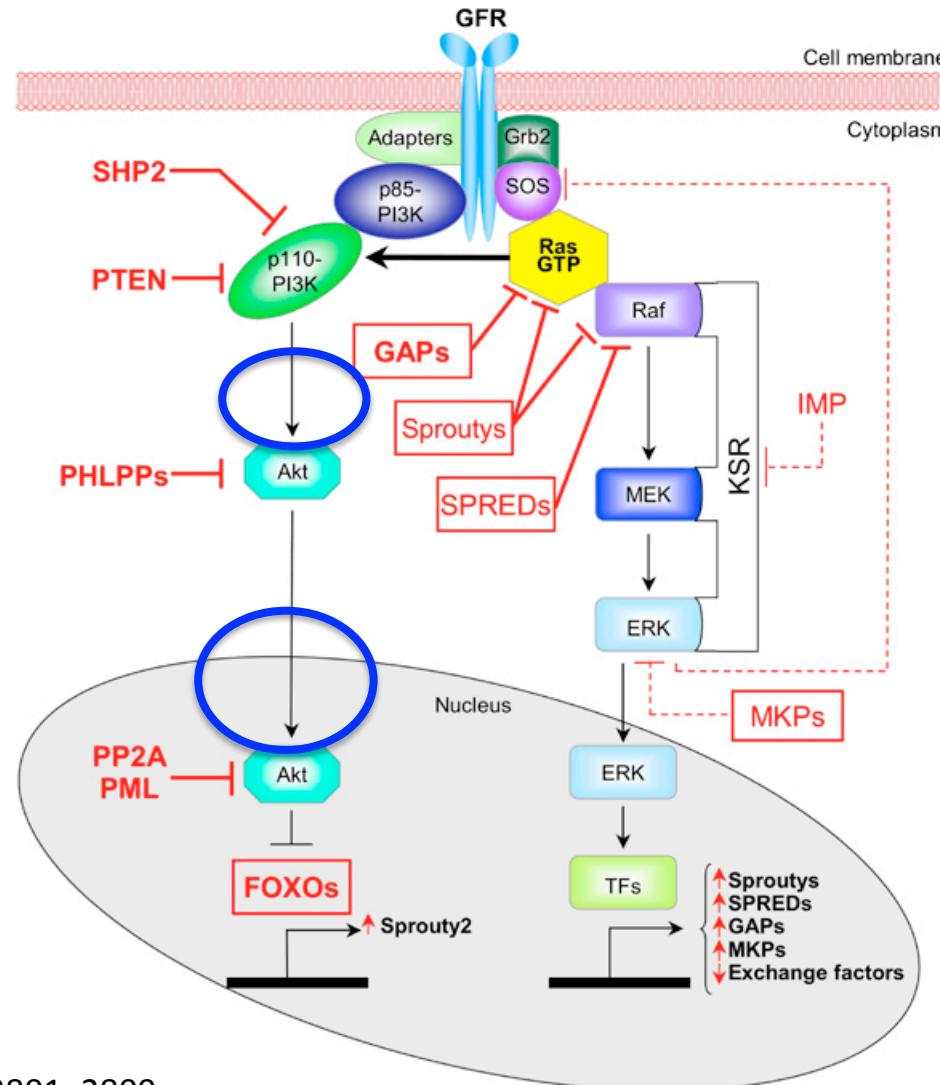
Diagram generated by biologist



Review

Oncogene (2008) 27, 2801–2809

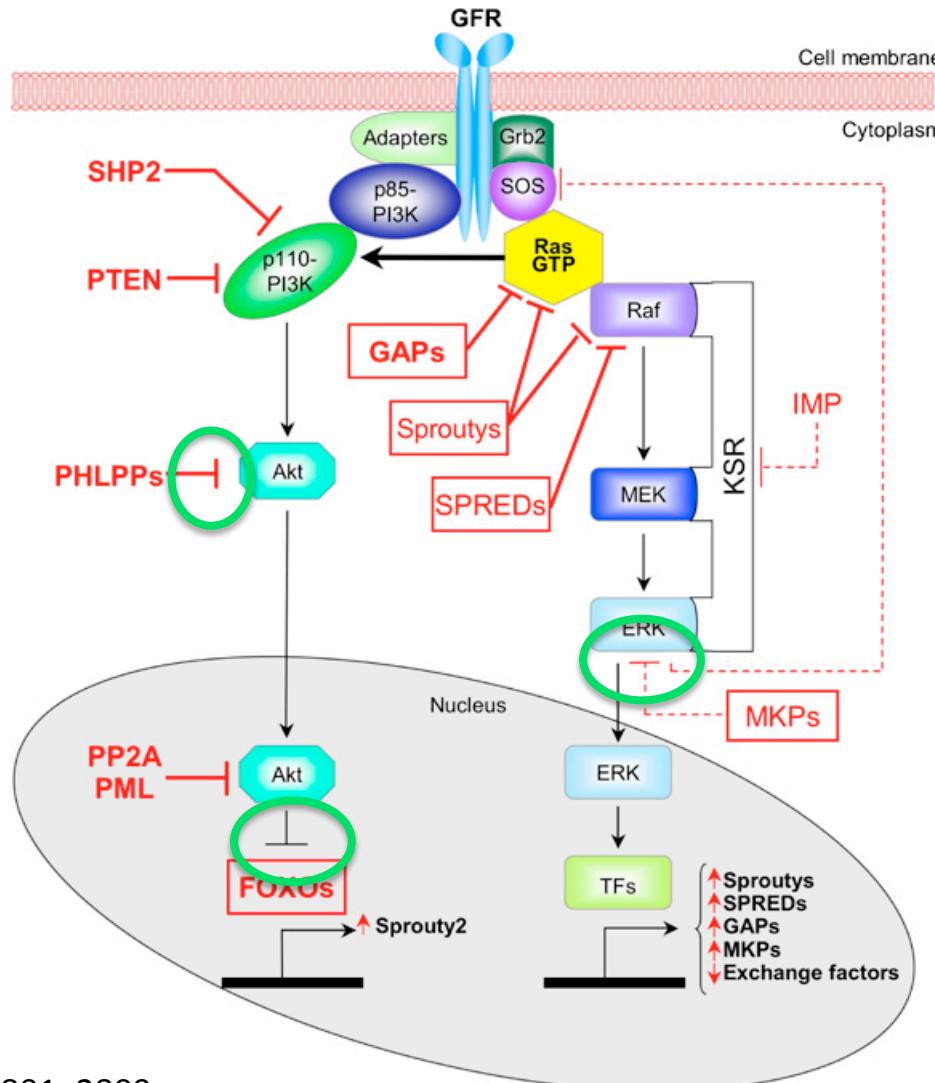
Diagram generated by biologist



Review

Oncogene (2008) 27, 2801–2809

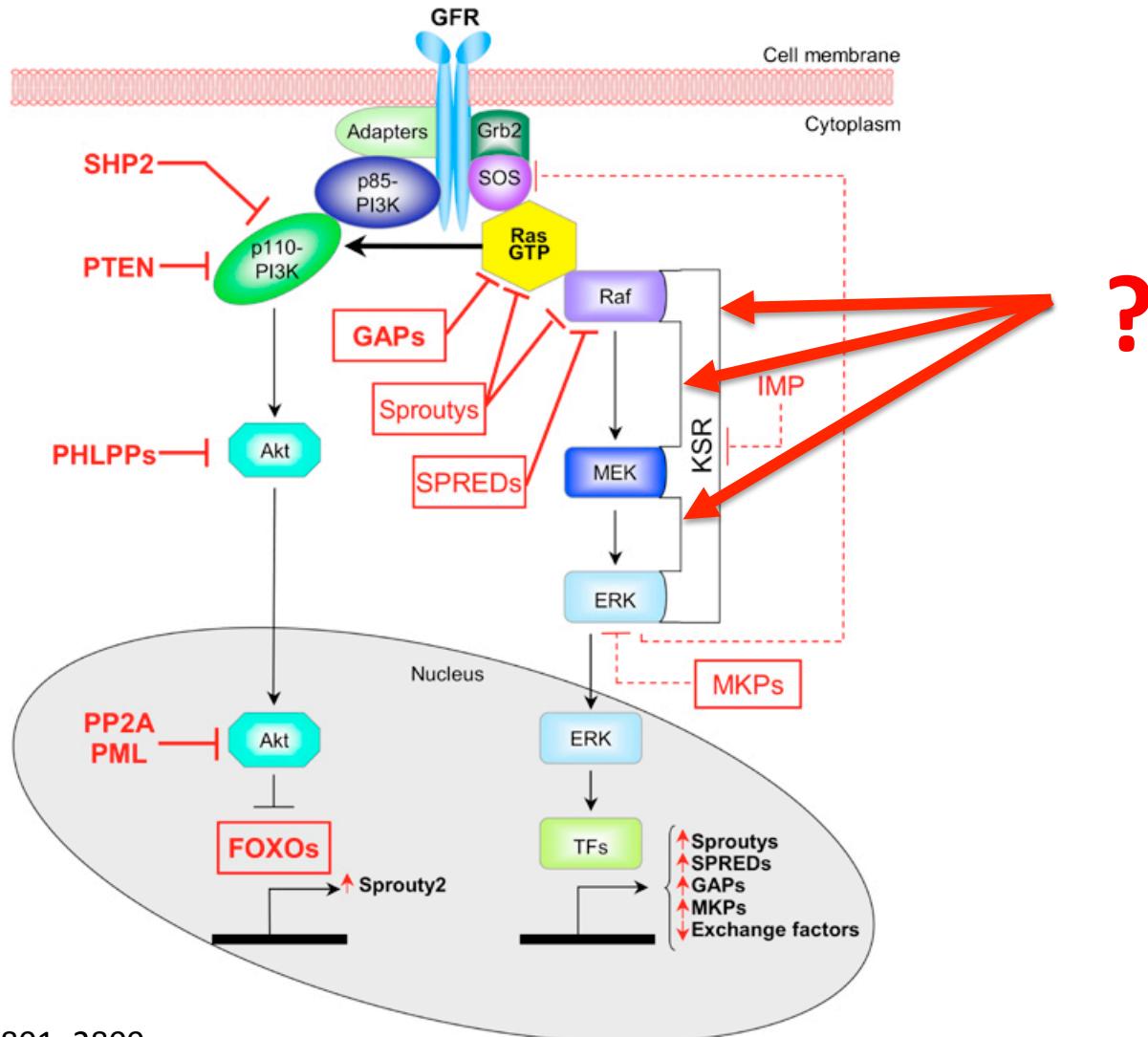
Diagram generated by biologist



Review

Oncogene (2008) 27, 2801–2809

Diagram generated by biologist



Review

Oncogene (2008) 27, 2801–2809

History of SBGN

- The SBGN effort was initiated by Dr. Hiroaki Kitano.
- The inaugural SBGN workshop was held in Tokyo in February 2006.



What is SBGN?

- An unambiguous way to graphically describe and interpret biochemical and cellular events
- Limited number of controlled symbols
Re-use existing symbols
 - + Smooth and easier learning curve
- Can represent logical or mechanistic models, biochemical pathways, at different levels of granularity
- Detailed technical specification, precise data-models and growing software support
- Developed over almost four years by a diverse community

The Systems Biology Graphical Notation

Nicolas Le Novère¹, Michael Hucka², Huaiyu Mi³, Stuart Moodie⁴, Falk Schreiber^{5,6}, Anatoly Sorokin⁷, Emek Demir⁸, Katja Wegner⁹, Mirit I Aladjem¹⁰, Sarala M Wimalaratne¹¹, Frank T Bergman¹², Ralph Gauges¹³, Peter Ghazal^{4,14}, Hideya Kawaji¹⁵, Lu Li¹, Yukiko Matsuoka¹⁶, Alice Villéger^{17,18}, Sarah E Boyd¹⁹, Laurence Calzone²⁰, Melanie Courtot²¹, Ugur Dogrusoz²², Tom C Freeman^{14,23}, Akira Funahashi²⁴, Samik Ghosh¹⁶, Akiya Jouraku²⁴, Sohyoung Kim¹⁰, Fedor Kolpakov^{25,26}, Augustin Luna¹⁰, Sven Sahle¹³, Esther Schmidt¹, Steven Watterson^{4,22}, Guanming Wu²⁷, Igor Goryanin⁴, Douglas B Kell^{18,28}, Chris Sander⁸, Herbert Sauro¹², Jacky L Snoep²⁹, Kurt Kohn¹⁰ & Hiroaki Kitano^{16,30,31}

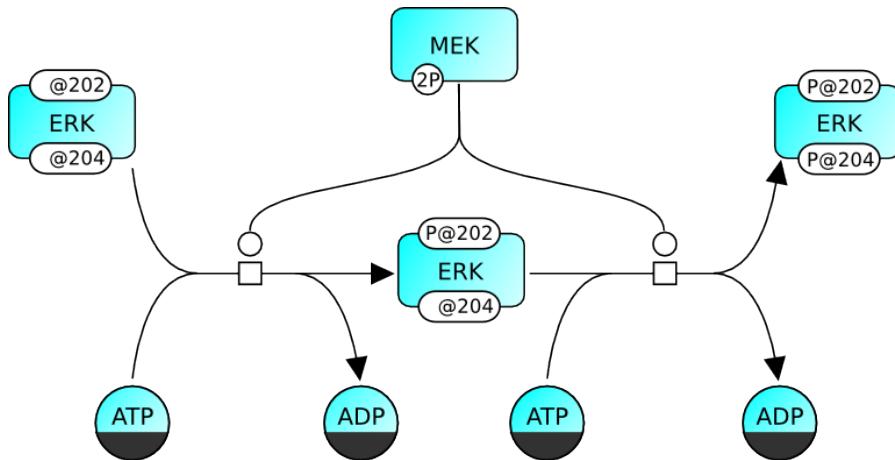
¹EMBL European Bioinformatics Institute, Hinxton, UK. ²Engineering and Applied Science, California Institute of Technology, Pasadena, California, USA. ³SRI International, Menlo Park, California, USA. ⁴Centre for Systems Biology at Edinburgh, University of Edinburgh, Edinburgh, UK. ⁵Leibniz Institute of Plant Genetics and Crop Plant Research, Gatersleben, Germany. ⁶Institute of Computer Science, University of Halle, Halle, Germany. ⁷School of Informatics, University of Edinburgh, Edinburgh, UK. ⁸Memorial Sloan Kettering Cancer Center - Computational Biology Center, New York, NY, USA. ⁹Science and Technology Research Institute, University of Hertfordshire, Hatfield, UK. ¹⁰National Cancer Institute, Bethesda, Maryland, USA. ¹¹Auckland Bioengineering Institute, University of Auckland, Auckland, New Zealand. ¹²Department of Bioengineering, University of Washington, Seattle, Washington, USA. ¹³BIOQUANT, University of Heidelberg, Heidelberg, Germany. ¹⁴Division of Pathway Medicine, University of Edinburgh Medical School, Edinburgh, UK. ¹⁵Riken OMICS Science Center, Yokohama City, Kanagawa, Japan. ¹⁶The Systems Biology Institute, Tokyo, Japan. ¹⁷School of Computer Science, University of Manchester, Manchester, UK. ¹⁸Manchester Interdisciplinary Biocentre, Manchester, UK. ¹⁹Clayton School of Information Technology, Faculty of Information Technology, Monash University, Melbourne, Victoria, Australia. ²⁰U900 INSERM, Paris Mines Tech, Institut Curie, Paris, France. ²¹Terry Fox Laboratory, British Columbia Cancer Research Center, Vancouver, British Columbia, Canada. ²²Bilkent Center for Bioinformatics, Bilkent University, Ankara, Turkey. ²³The Roslin Institute, University of Edinburgh, Midlothian, UK. ²⁴Department of Biosciences and Informatics, Keio University, Hiyoshi, Kouhoku-ku, Yokohama, Japan. ²⁵Institute of Systems Biology, Novosibirsk, Russia. ²⁶Design Technological Institute of Digital Techniques SB RAS, Novosibirsk, Russia. ²⁷Ontario Institute for Cancer Research, Toronto, Ontario, Canada. ²⁸School of Chemistry, University of Manchester, Manchester, UK. ²⁹Department of Biochemistry, Stellenbosch University, Matieland, South Africa. ³⁰Sony Computer Science Laboratories, Tokyo, Japan. ³¹Okinawa Institute of Science and Technology, Okinawa, Japan. Correspondence should be addressed to N.L.N. (lenov@ebi.ac.uk).

39 authors, 31 affiliations

SBGN is being developed by a large user community

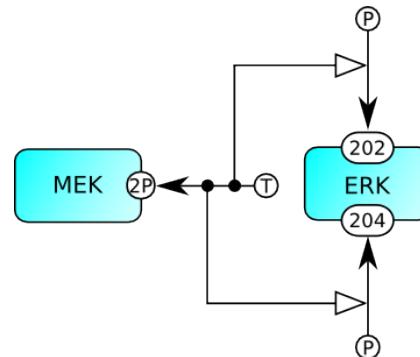
Graph trinity: three languages in one notation

Process Descriptions



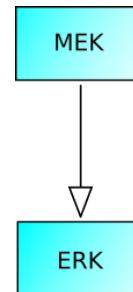
- Unambiguous
- Mechanistic
- Sequential
- Combinatorial explosion

Entity Relationships



- Unambiguous
- Mechanistic
- Non-sequential
- Independence of relationships

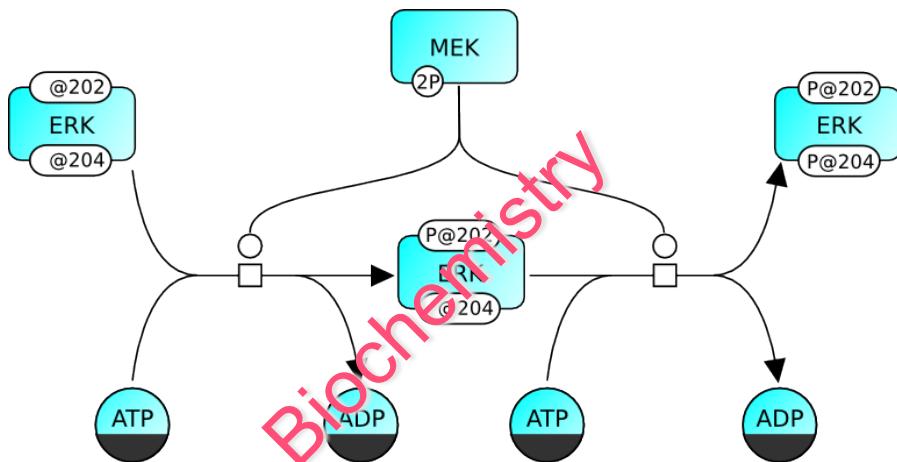
Activity Flows



- Ambiguous
- Conceptual
- Sequential

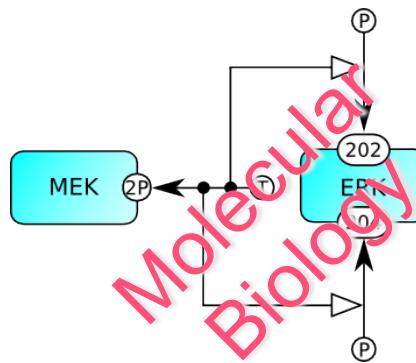
Graph trinity: three languages in one notation

Process Descriptions



- Unambiguous
- Mechanistic
- Sequential
- Combinatorial explosion

Entity Relationships



- Unambiguous
- Mechanistic
- Non-sequential
- Independence of relationships

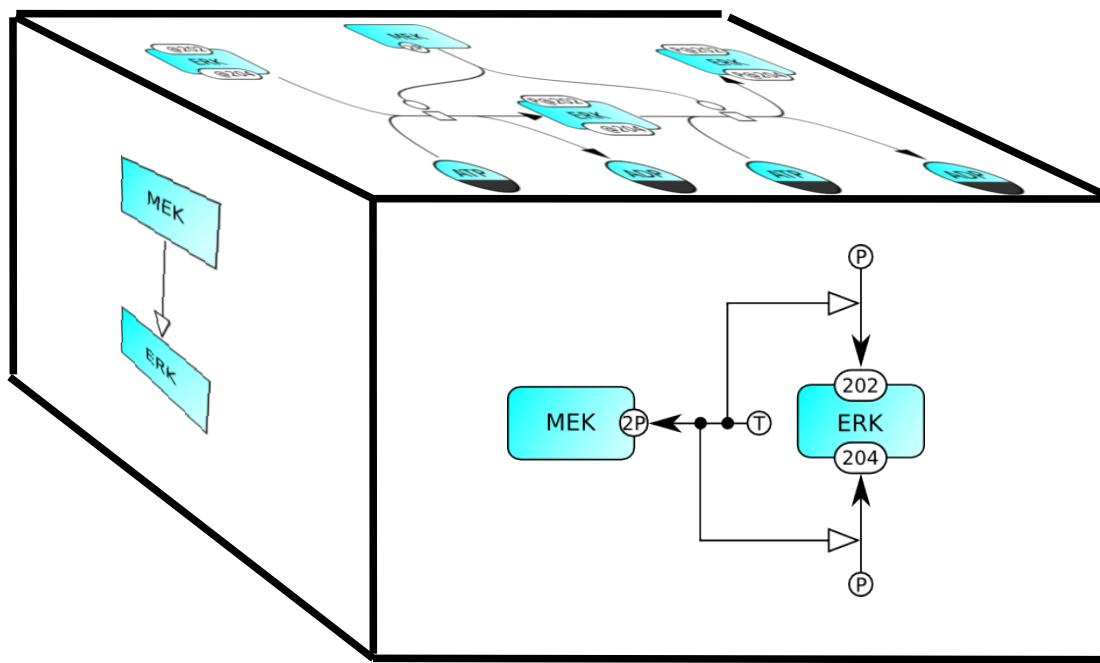
Activity Flows



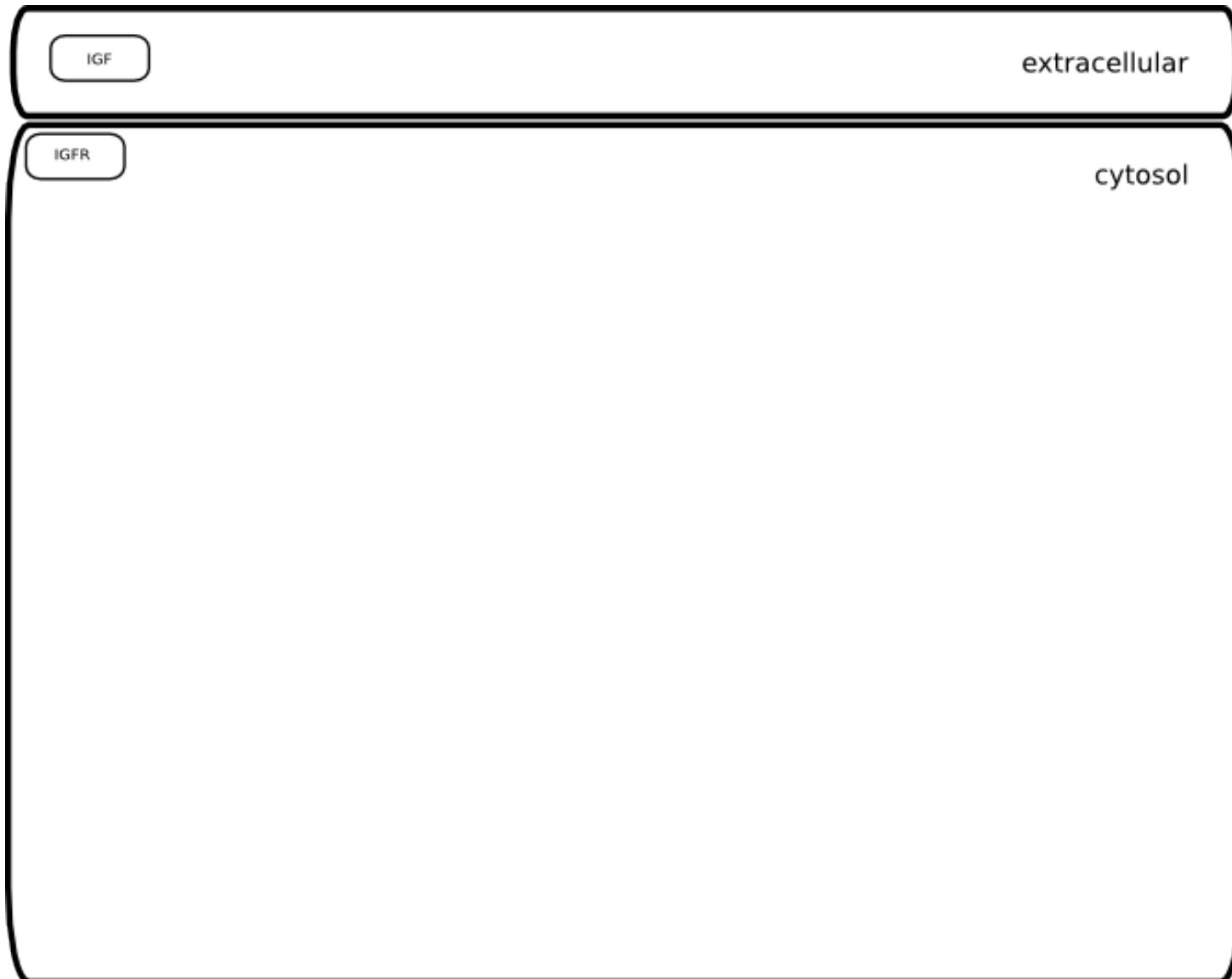
- Ambiguous
- Conceptual
- Sequential

Biochemistry *Molecular Biology* *Physiology genetics*

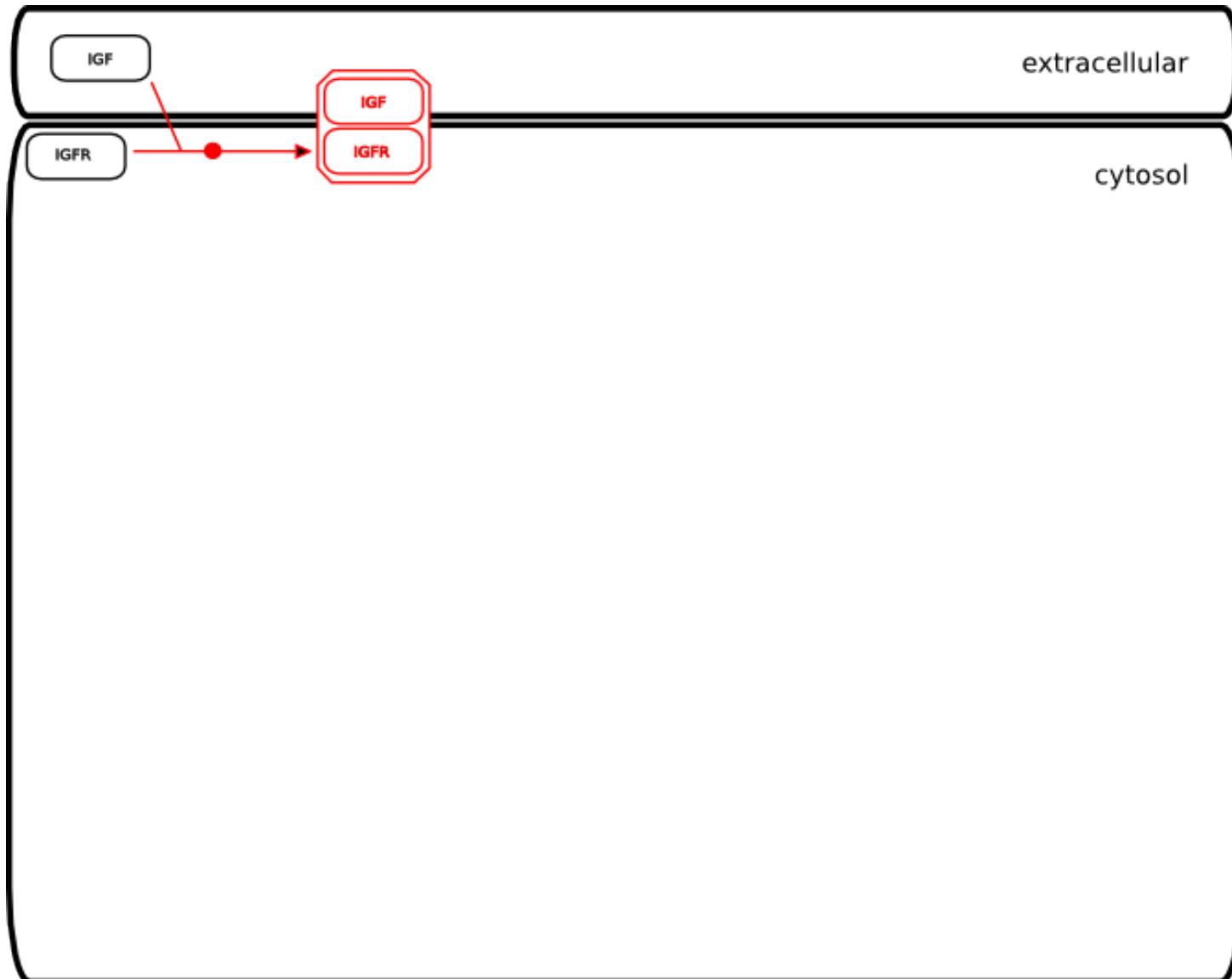
Orthogonal projection of biology



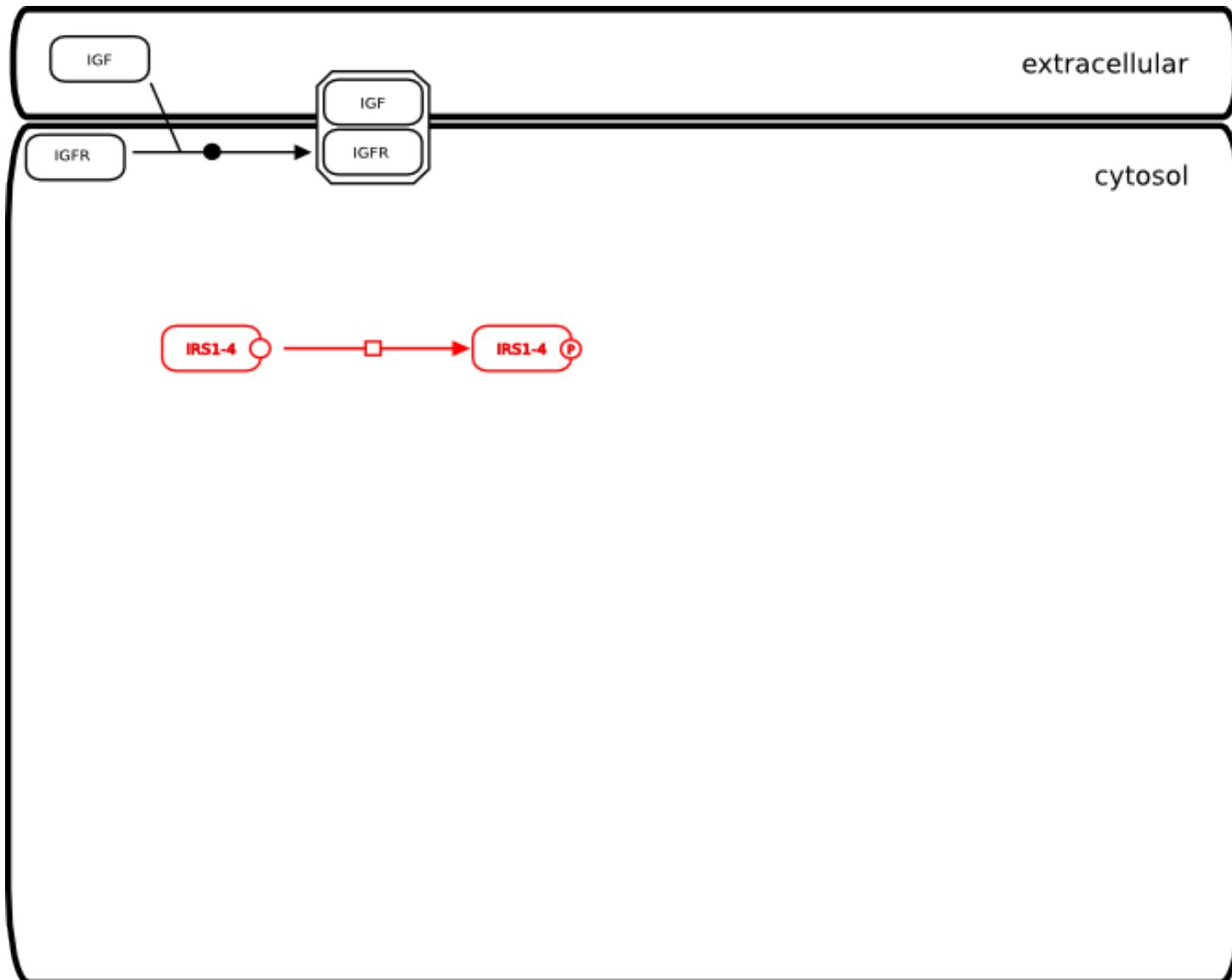
Process Description map



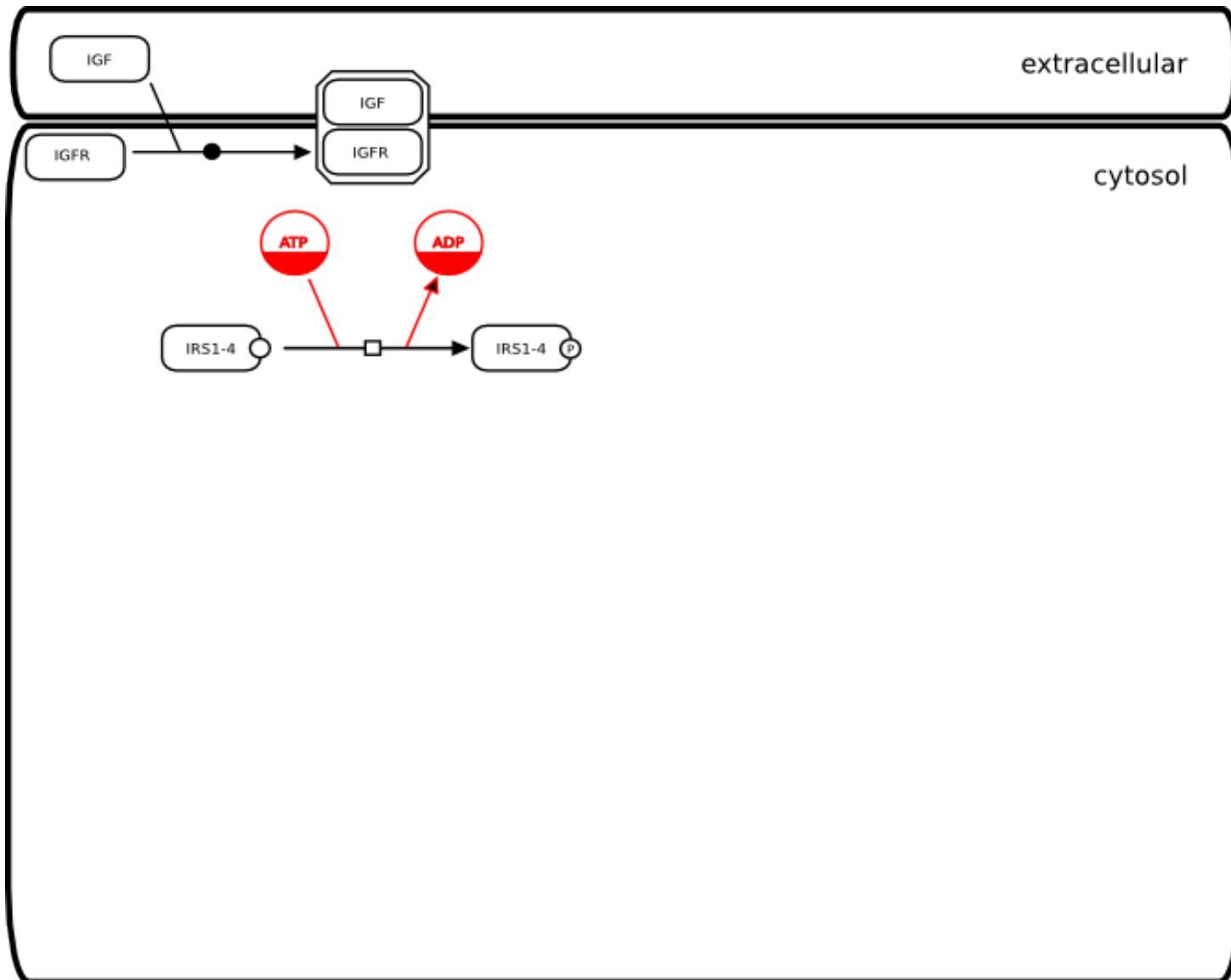
Process Description map



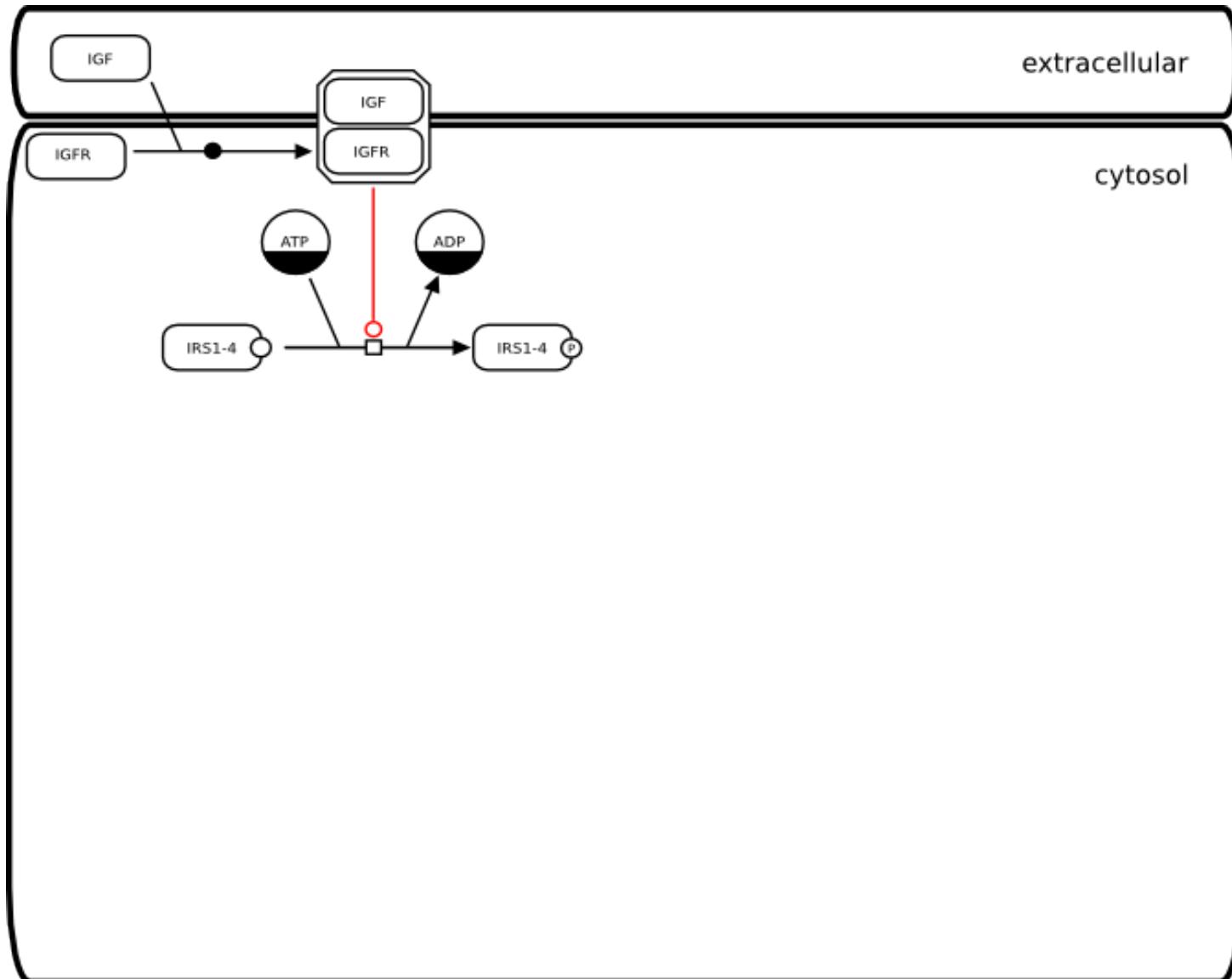
Process Description map



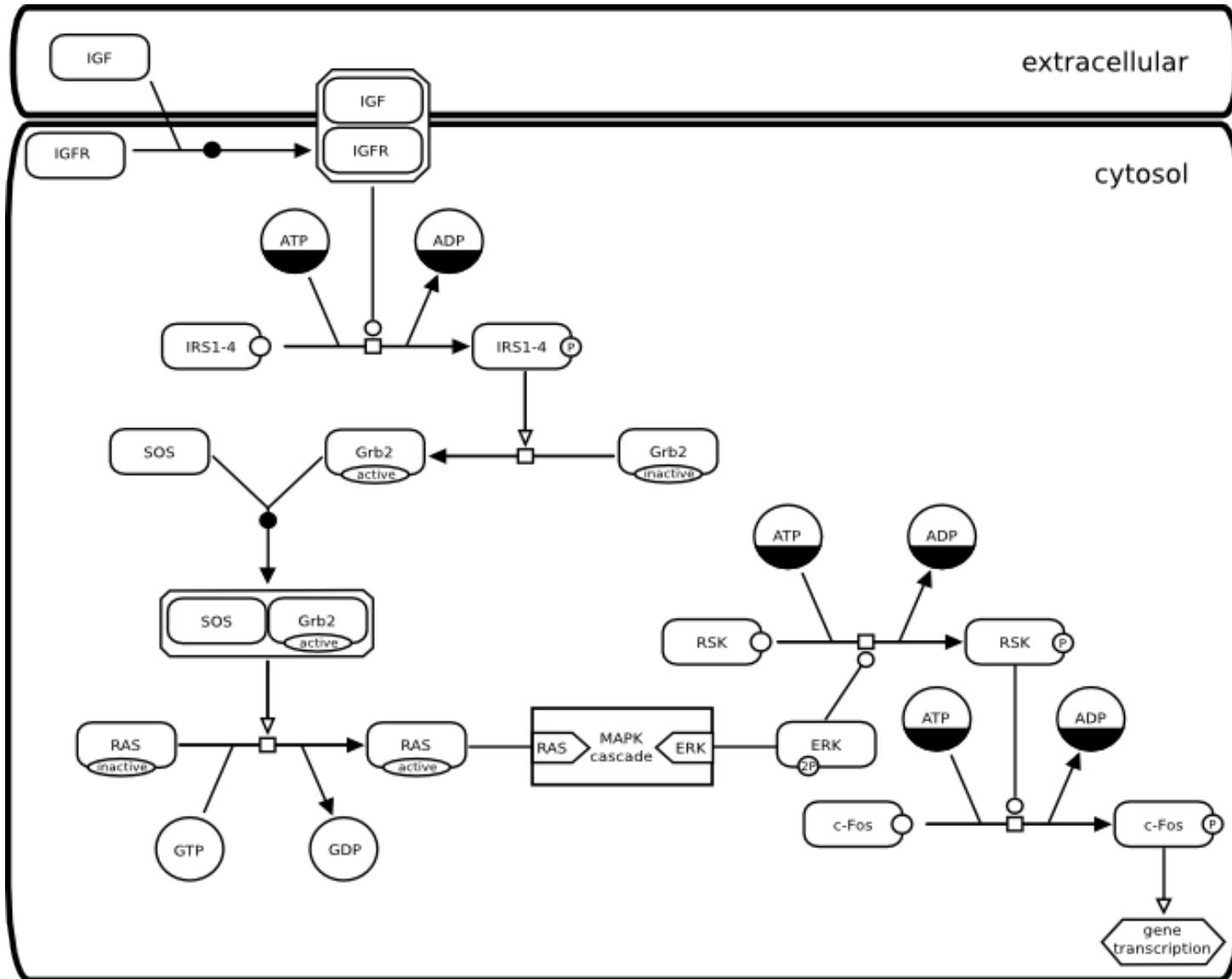
Process Description map



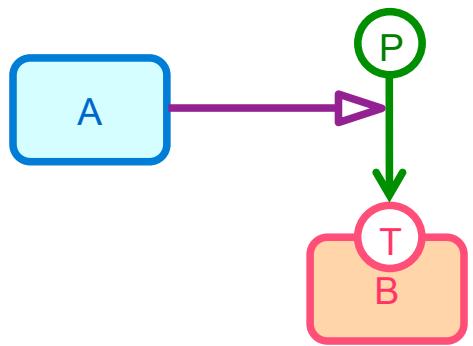
Process Description map



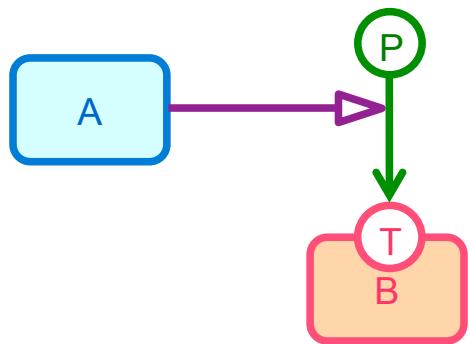
Process Description map



Entity Relationships can be viewed as rules

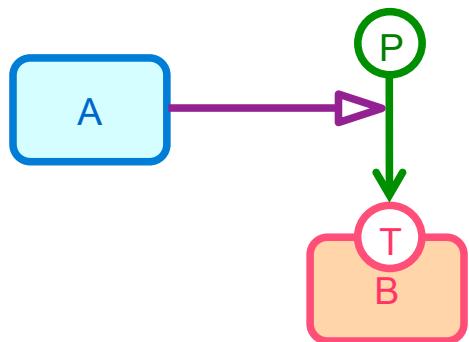


Entity Relationships can be viewed as rules



If A exists, the assignment of the value P to the state variable T of B is increased

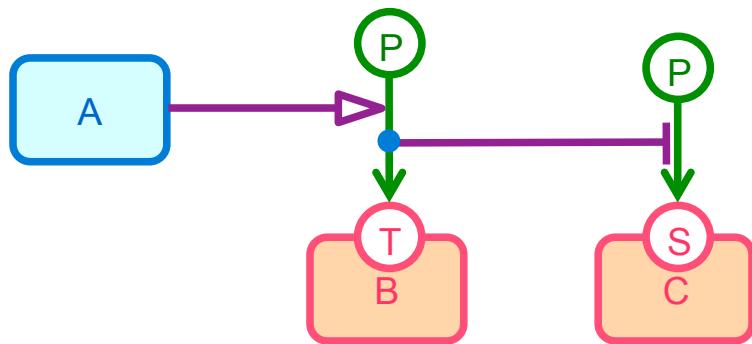
Entity Relationships can be viewed as rules



If A exists, the assignment of the value P to the state variable T of B is increased

(A stimulates the phosphorylation of B on the threonine)

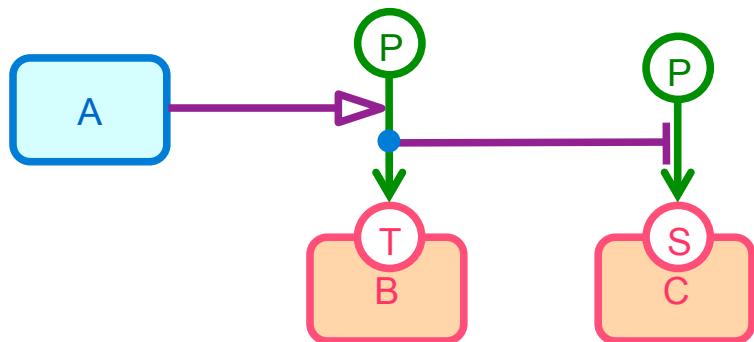
Entity Relationships can be viewed as rules



If A exists, the assignment of the value P to the state variable T of B is increased

If P is assigned to the state variable T of B, the assignment of the value P to the state variable S of C is decreased

Entity Relationships can be viewed as rules



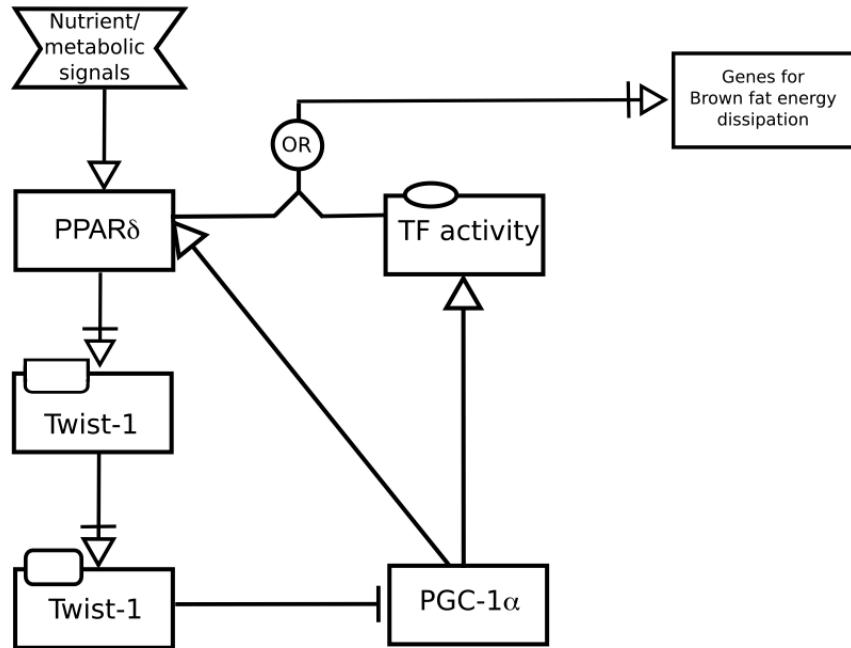
If A exists, the assignment of the value P to the state variable T of B is increased

If P is assigned to the state variable T of B, the assignment of the value P to the state variable S of C is decreased

(Phosphorylated B inhibits the phosphorylation of C on the serine.)

Example of Activity Flow Map

-PPAR δ regulation in brown fat metabolism



Pan et. al., (2009) *Cell*, 137:73-86

Twist-1 Is a PPAR δ -Inducible, Negative-Feedback Regulator of PGC-1 α in Brown Fat Metabolism

Dongning Pan,^{1,2} Masaki Fujimoto,^{1,2} Andrea Lopes,¹ and Yong-Xu Wang^{1,*}

¹Program in Gene Function and Expression and Program in Molecular Medicine, University of Massachusetts Medical School,

364 Plantation Street, Worcester, MA 01605, USA

²These authors contributed equally to this work

*Correspondence: yongxu.wang@umassmed.edu

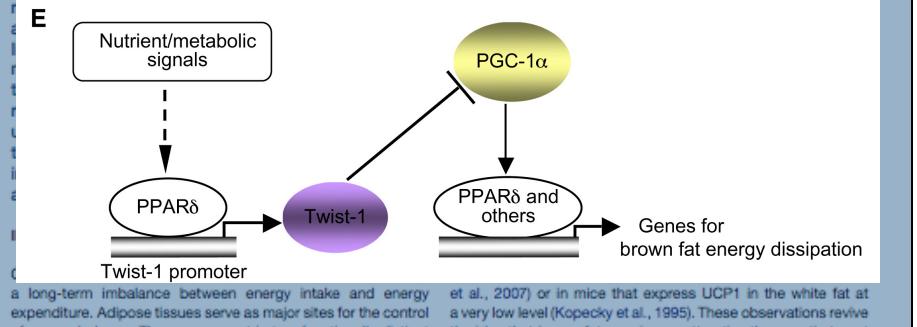
DOI 10.1016/j.cell.2009.01.051

SUMMARY

Brown fat is specialized for energy expenditure, a process that is principally controlled by the transcriptional coactivator PGC-1 α . Here, we describe a molecular network important for PGC-1 α function and brown fat metabolism. We find that twist-1 is selectively expressed in adipose tissue, interacts with PGC-1 α , and is recruited to the promoters of PGC-1 α 's target genes to suppress mitochondrial metabolism and uncoupling. *In vivo*, transgenic mice expressing twist-1 in the adipose tissue are prone to high-fat-diet-induced obesity, whereas twist-1 heterozygous knockout mice are obesity

the mitochondrial proton gradient from ATP production. Given the fundamental importance of adipose tissues in the maintenance of systematic energy homeostasis, their functions must be tightly regulated.

As a heat-generating organ, brown fat plays a key part in the regulation of energy balance and obesity, as evidenced in rodent studies. For instance, either ablation of brown fat through expression of toxic transgene or knockout of UCP1 leads to high susceptibility to diet-induced obesity (Kontani et al., 2005; Lowell et al., 1993), whereas increase of UCP1 expression protects animals against diet-induced obesity (Kopecky et al., 1995). However, human adults, unlike rodents and human neonates, do not possess discrete brown fat depots, and brown fat cells are dispersed within white fat, casting doubt on whether human brown fat cells are of physiological and/or pharmacological



a long-term imbalance between energy intake and energy expenditure. Adipose tissues serve as major sites for the control of energy balance. They are present in two functionally distinct types: white fat and brown fat. White fat stores excess energy in the form of triglycerides and releases them in times of energy need. By contrast, brown fat is specialized for energy expenditure by dissipating energy as heat, a process termed as adaptive thermogenesis (Cannon and Nedergaard, 2004; Lowell and Spiegelman, 2000). The unique metabolic property of brown fat is due to its high mitochondrial density and fuel oxidation capacity, and to its exclusive expression of uncoupling protein-1 (UCP1) in the inner mitochondrial membrane, which uncouples

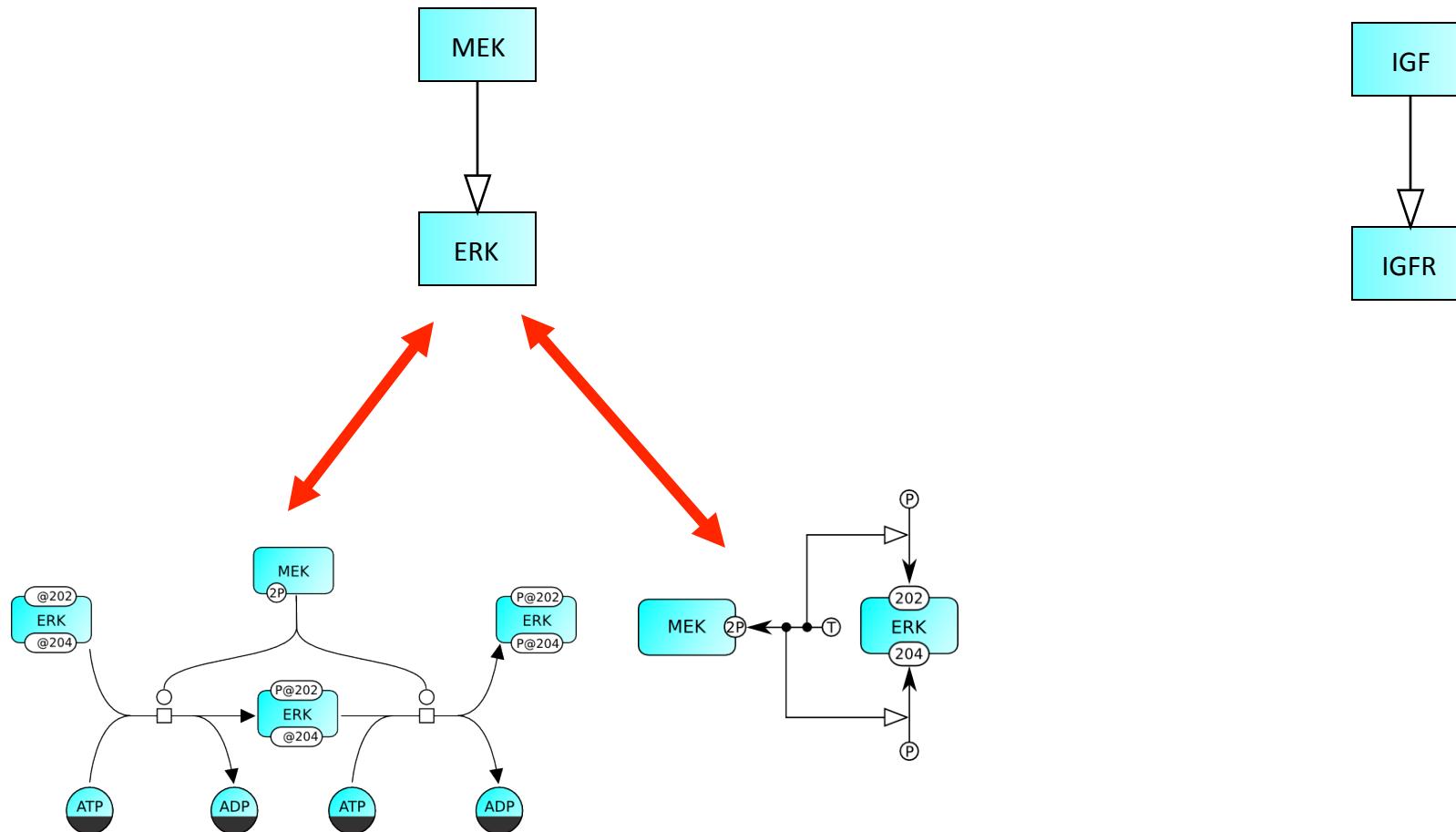
et al., 2007) or in mice that express UCP1 in the white fat at a very low level (Kopecky et al., 1995). These observations revive the idea that brown fat remains an attractive therapeutic target tissue for obesity and associated diseases. Clearly, there is a strong need to understand the molecular basis underlying brown fat metabolism.

A central regulator in brown fat thermogenesis is the transcriptional coactivator PGC-1 α (reviewed in Lin et al., 2005). PGC-1 α is predominantly expressed in the brown fat, and its expression is highly influenced by nutritional and environmental cues. Both overexpression and loss-of-function studies demonstrate that PGC-1 α regulates the entire program of thermogenesis

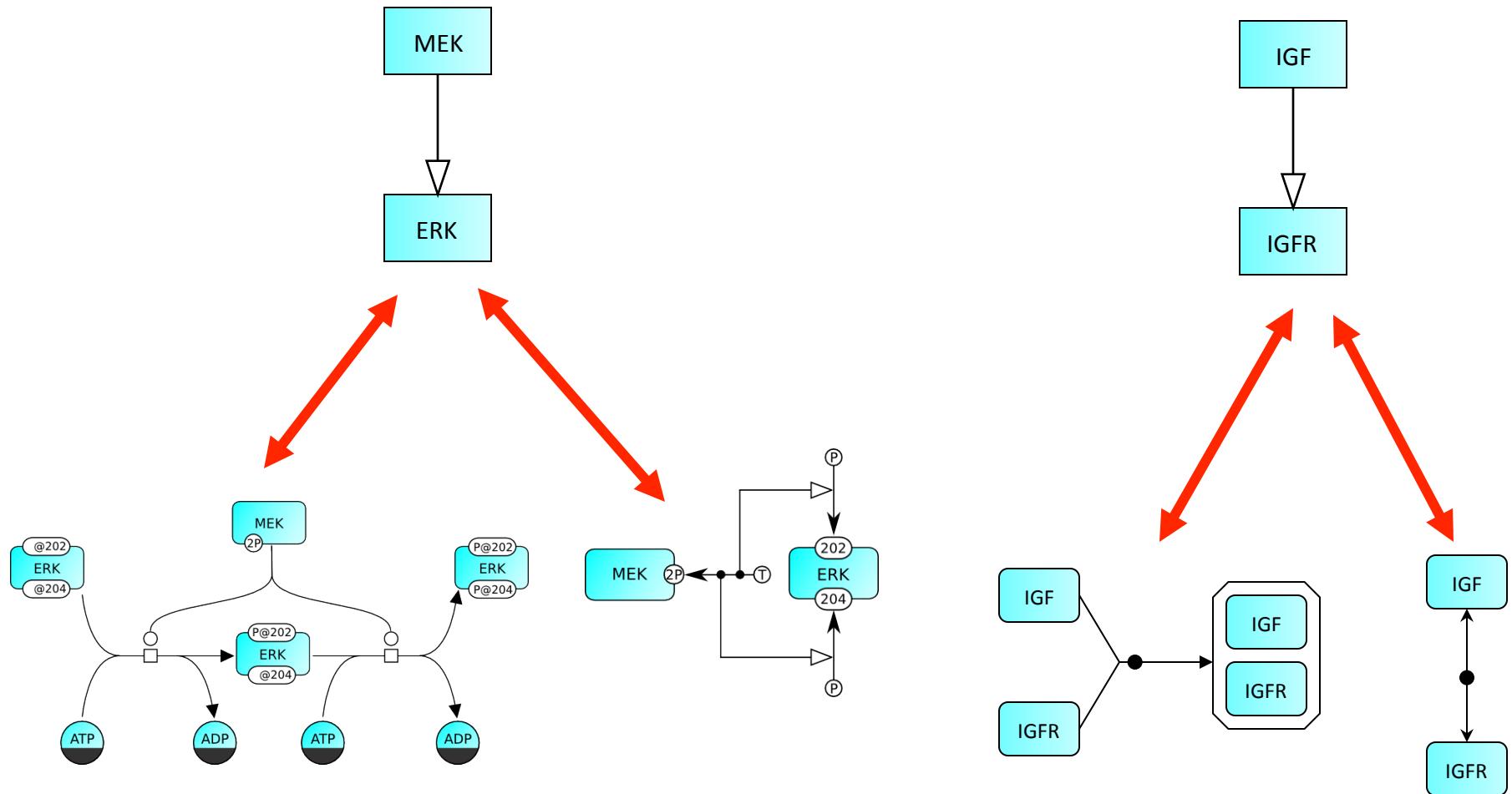
Activity Flow map is ambiguous



Activity Flow map is ambiguous



Activity Flow map is ambiguous



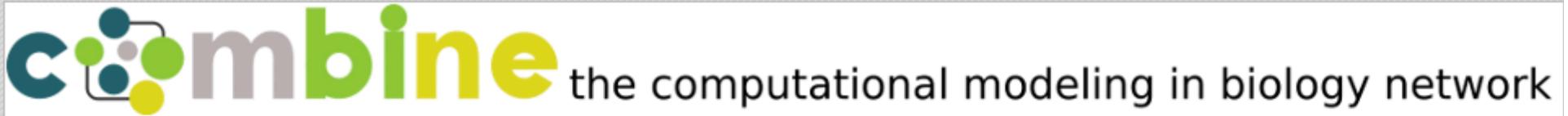


BioPAX

- **Systems biology markup language.**
 - Machine readable format for pathway networks exchange.
 - User community: **modeling and simulation.**
 - Ability to handle numbers and equations.
 - Simple data structure; nodes and edges.
- **Systems biology graphical notation**
 - Controlled graphical notations, layout syntax and semantics.
 - Unambiguous graphical representation of biological pathways for **both human and machine.**
- **Biological pathway exchange format.**
 - Captures more detailed pathway data for signaling pathways, molecular and genetic interactions and gene regulation network.
 - User community: **pathway database, biologists**

Different tools => different interfaces & languages





HARMONY 2013

Standards

Events

Documents

About

Search

COMBINE

- Standards
- Events
- Documents
- About
- Help
- Sign-in
- Home

Home

The 'COmputational Modeling in BIology' NEtwork (COMBINE) is an initiative to coordinate the development of the various community [standards and formats](#) for computational models, initially in Systems Biology and related fields. By doing so, it is expected that the federated projects will develop a set of interoperable and non-overlapping standards covering all aspects of modeling in biology.

Building on the experience of mature projects, which already have stable specifications, software support, user-base and community governance, COMBINE will help foster or support fledgling efforts aimed at filling gaps or new needs. As those efforts mature, they may become part of the [core set of COMBINE standards](#).

One of the initial activities of COMBINE is to coordinate the organization of scientific and technical [events](#) common to several standards.

To receive announcements from COMBINE, subscribe to combine-announce@mbine.org (Note that the main list of each of the [COMBINE standards](#) is already subscriber).

To discuss the goals, organization and operation of COMBINE, subscribe to combine-discuss@mbine.org.

To report issues about the co.mbine.org website, send a mail to combine-support@mbine.org.

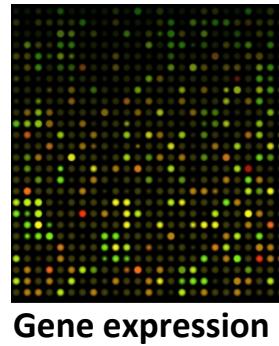
COmputational Modeling in BIology NEtwork

- Coordinate
- Interoperability
- Non-overlapping
- Software development

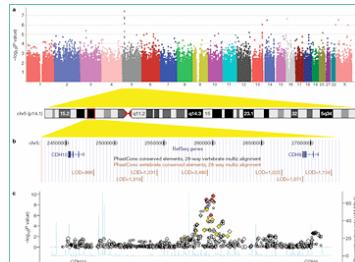


Attendees at COMBINE, Edinburgh, 2010

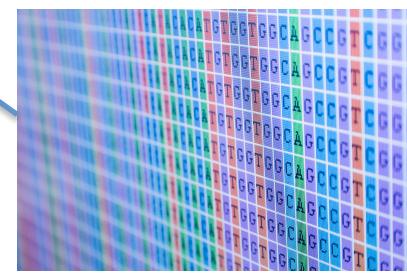
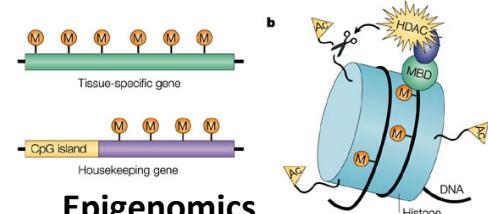
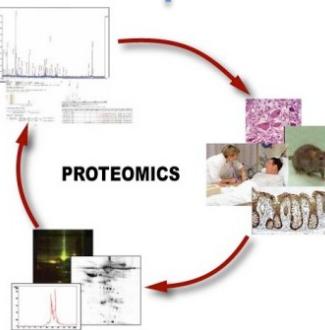
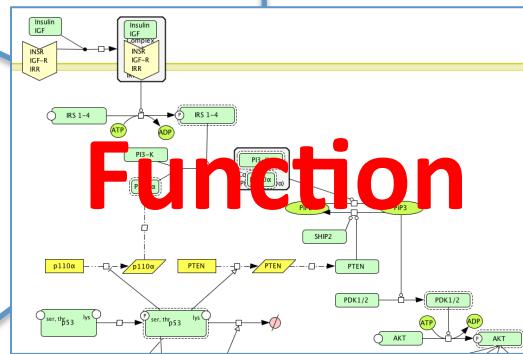
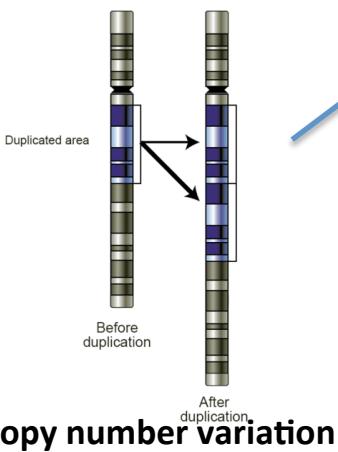
The success is dependent on your involvement



Gene expression



GWAS



Next generation sequencing

To be continued