

# BioPAX - Biological Pathway Data Exchange Format Tutorial

Nadia Anwar  
**[www.biopax.org](http://www.biopax.org)**  
BioPAX Workshop November 2009

# BioPAX Supporting Groups

## Current Participants

- Memorial Sloan-Kettering Cancer Center: E.Demir, M. Cary, C. Sander
- University of Toronto: G. Bader
- SRI Bioinformatics Research Group: P. Karp, S. Paley, J. Pick
- Bilkent University: U. Dogrusoz
- Université Libre de Bruxelles: C. Lemmer
- CBRC Japan: K. Fukuda
- Dana Farber Cancer Institute: J. Zucker
- Millennium: J. Rees, A. Ruttenberg
- Cold Spring Harbor/EBI: G. Wu, M. Gillespie, P. D'Eustachio, I. Vastrik, L. Stein
- BioPathways Consortium: J. Luciano, E. Neumann, A. Regev, V. Schachter
- Argonne National Laboratory: N. Maltsev, E. Marland, M. Syed
- Harvard: F. Gibbons
- AstraZeneca: E. Pichler
- BIOBASE: E. Wingender, F. Schacherer
- NCI: M. Aladjem, C. Schaefer
- Università di Milano Bicocca, Pasteur, Rennes: A. Splendiani
- Vassar College: K. Dahlquist
- Columbia: A. Rzhetsky

## Collaborating Organizations

- Proteomics Standards Initiative (PSI)
- Systems Biology Markup Language (SBML)
- CellML
- Chemical Markup Language (CML)

## Databases

- BioCyc, WIT, KEGG, BIND, PharmGKB, aMAZE, INOH, Transpath, Reactome, PATIKA, eMIM, NCI PID, CellMap

## Wouldn't be possible without

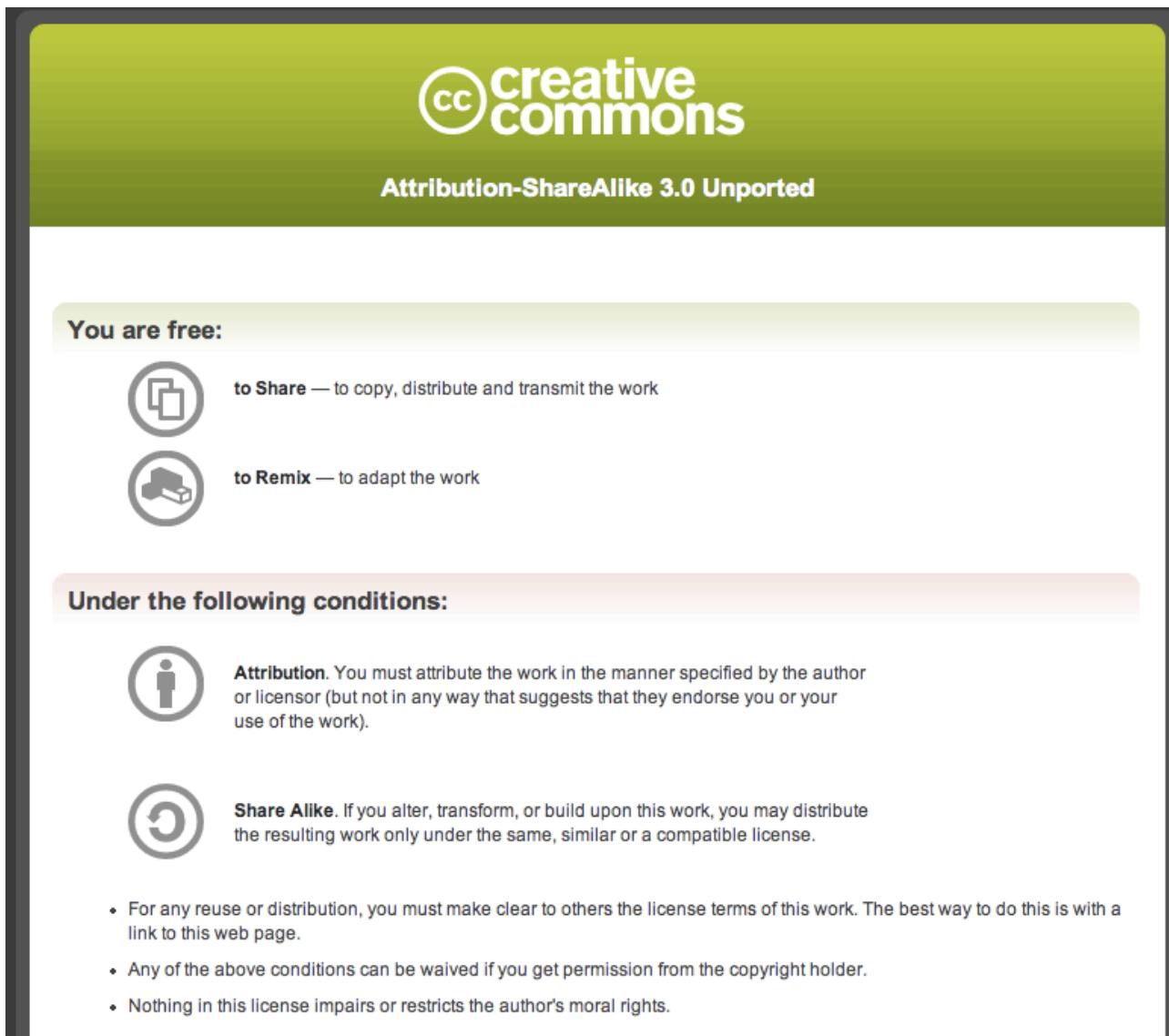
Gene Ontology  
Protégé, U.Manchester, Stanford

## Grants/Support

- Department of Energy (Workshop)
- caBIG



<http://creativecommons.org/licenses/by-sa/3.0/>



The image shows a screenshot of the Creative Commons Attribution-ShareAlike 3.0 Unported license page. The header features the "creative commons" logo with the "cc" icon. Below the logo, the text "Attribution-ShareAlike 3.0 Unported" is displayed. The page is divided into sections: "You are free:" and "Under the following conditions:". The "You are free:" section includes icons for "to Share" (copy, distribute, transmit) and "to Remix" (adapt). The "Under the following conditions:" section includes icons for "Attribution" (person icon) and "Share Alike" (circular arrow icon). Below these icons, detailed text explains the requirements: Attribution (must attribute the work in the manner specified by the author or licensor) and Share Alike (if you alter, transform, or build upon this work, you may distribute the resulting work only under the same, similar or a compatible license). A bulleted list at the bottom provides additional guidance: For any reuse or distribution, you must make clear to others the license terms of this work; Any of the above conditions can be waived if you get permission from the copyright holder; Nothing in this license impairs or restricts the author's moral rights.

**You are free:**

-  **to Share** — to copy, distribute and transmit the work
-  **to Remix** — to adapt the work

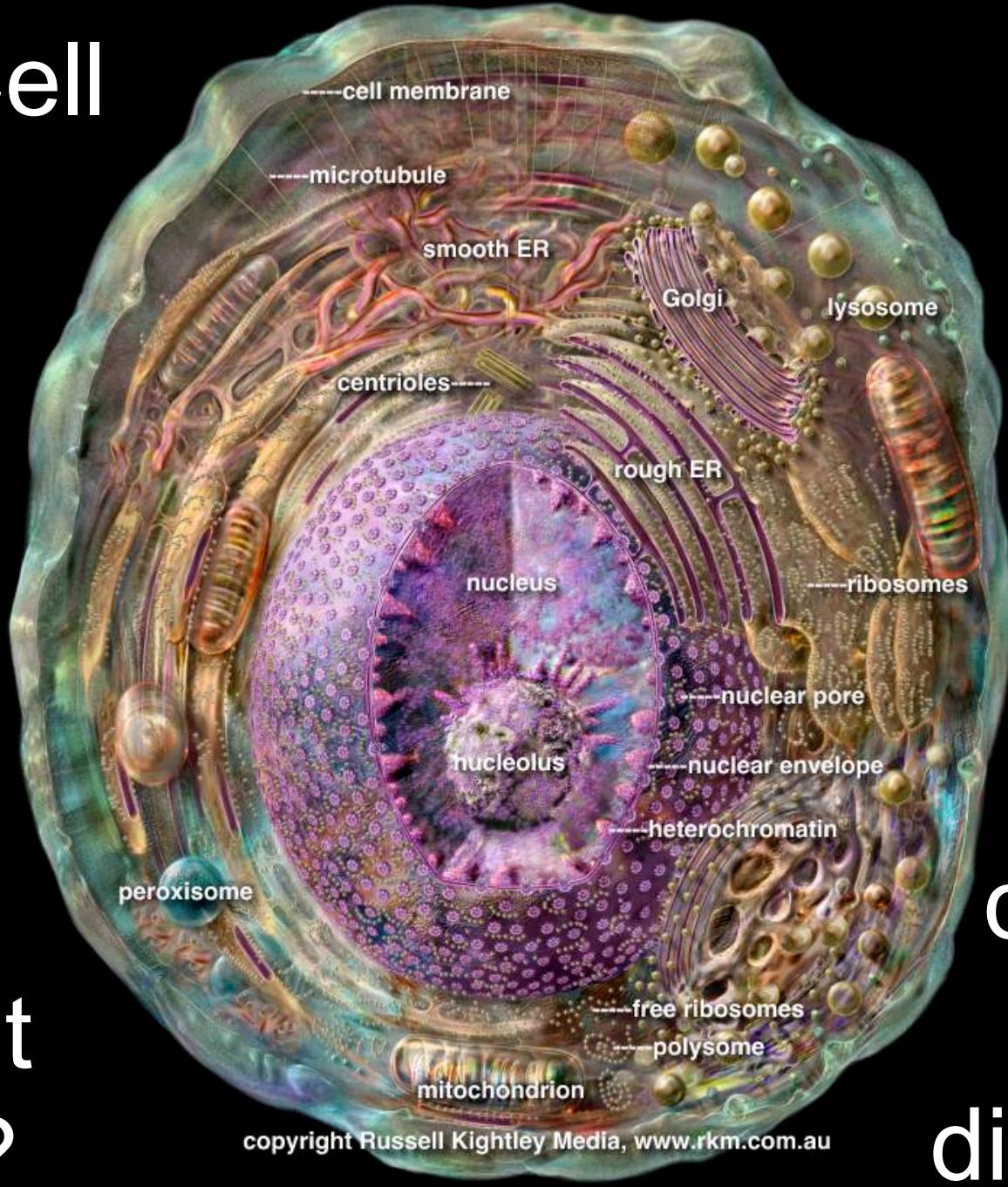
**Under the following conditions:**

-  **Attribution.** You must attribute the work in the manner specified by the author or licensor (but not in any way that suggests that they endorse you or your use of the work).
-  **Share Alike.** If you alter, transform, or build upon this work, you may distribute the resulting work only under the same, similar or a compatible license.
  - For any reuse or distribution, you must make clear to others the license terms of this work. The best way to do this is with a link to this web page.
  - Any of the above conditions can be waived if you get permission from the copyright holder.
  - Nothing in this license impairs or restricts the author's moral rights.

Will be made available from biopax.org wiki

# The Cell

How  
does it  
work?



How  
does it  
fail in  
disease?

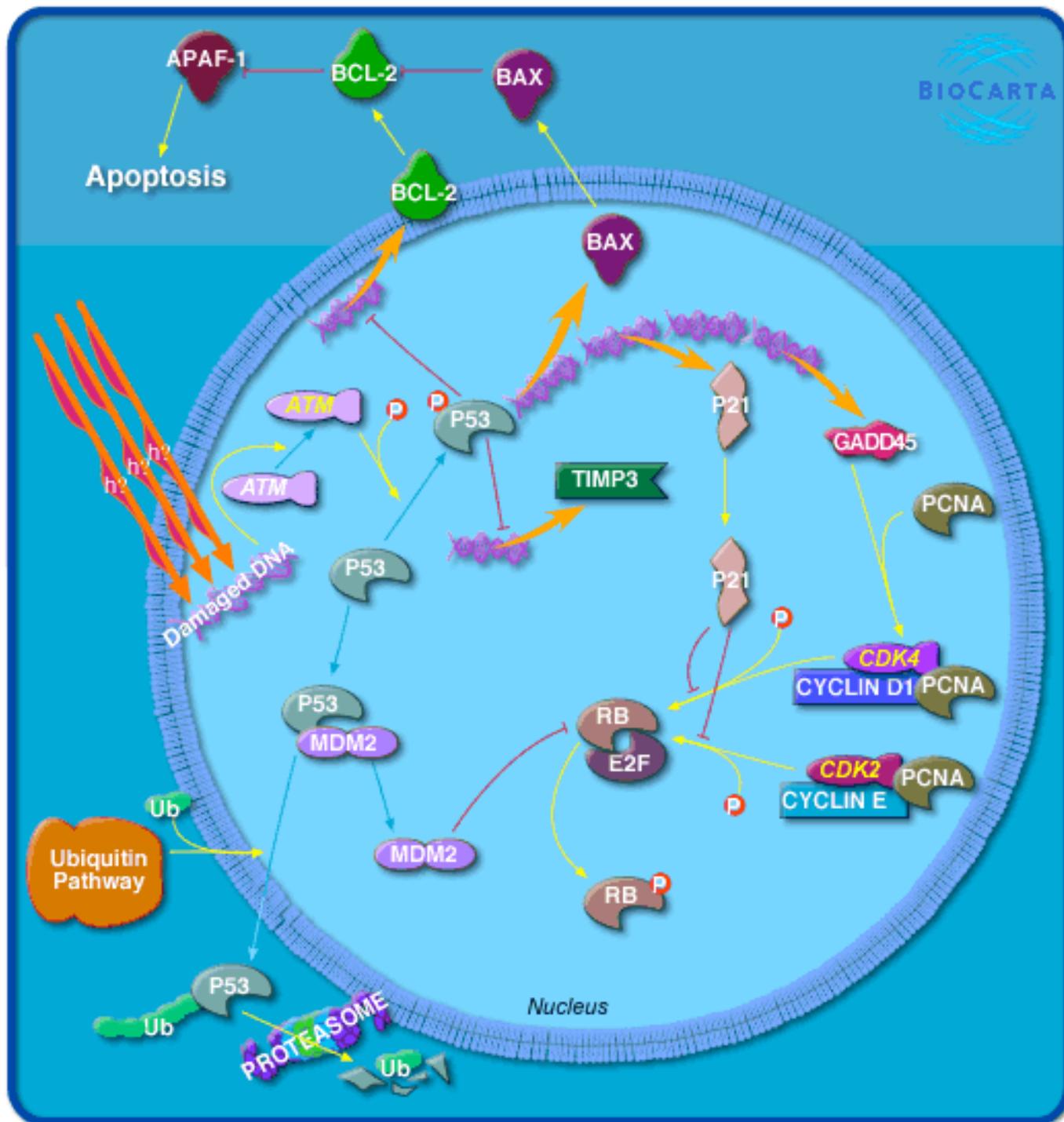
# Pathways

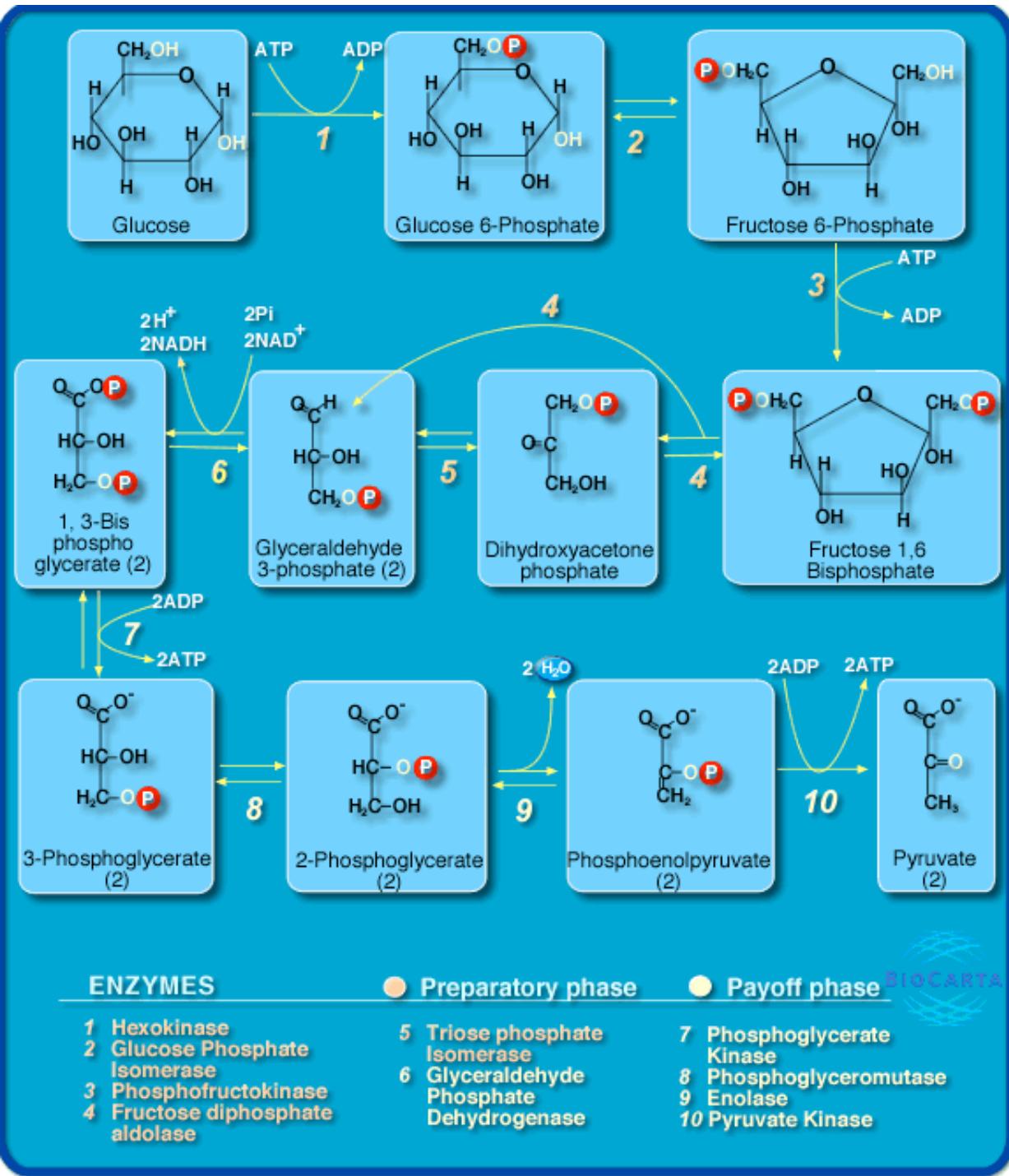
- Pathways are biological processes
- But, not really pathways → networks
- Metabolic, signaling, regulatory and genetic
- Define gene function at many different levels
- Biologists have found useful to group together for organizational, historic, biophysical or other reasons

Figure S3: View of the entire HMs-PCI Data set.  
Thick blue lines represent literature-derived interactions from ProBiND+MEPs in the HMs-PCI dataset.  
Thin grey lines represent integrated modelled interactions.

Note: generally out of cell context

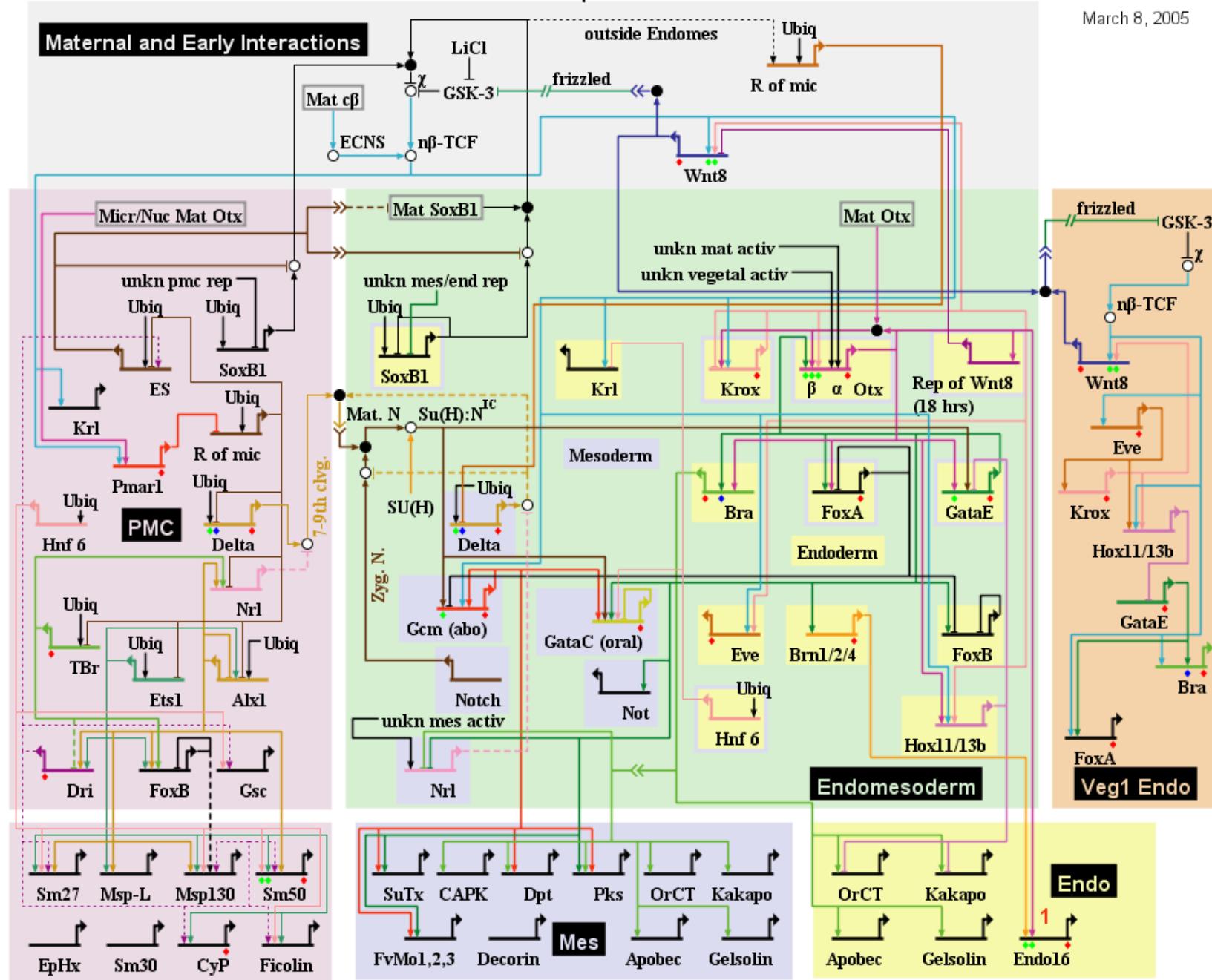




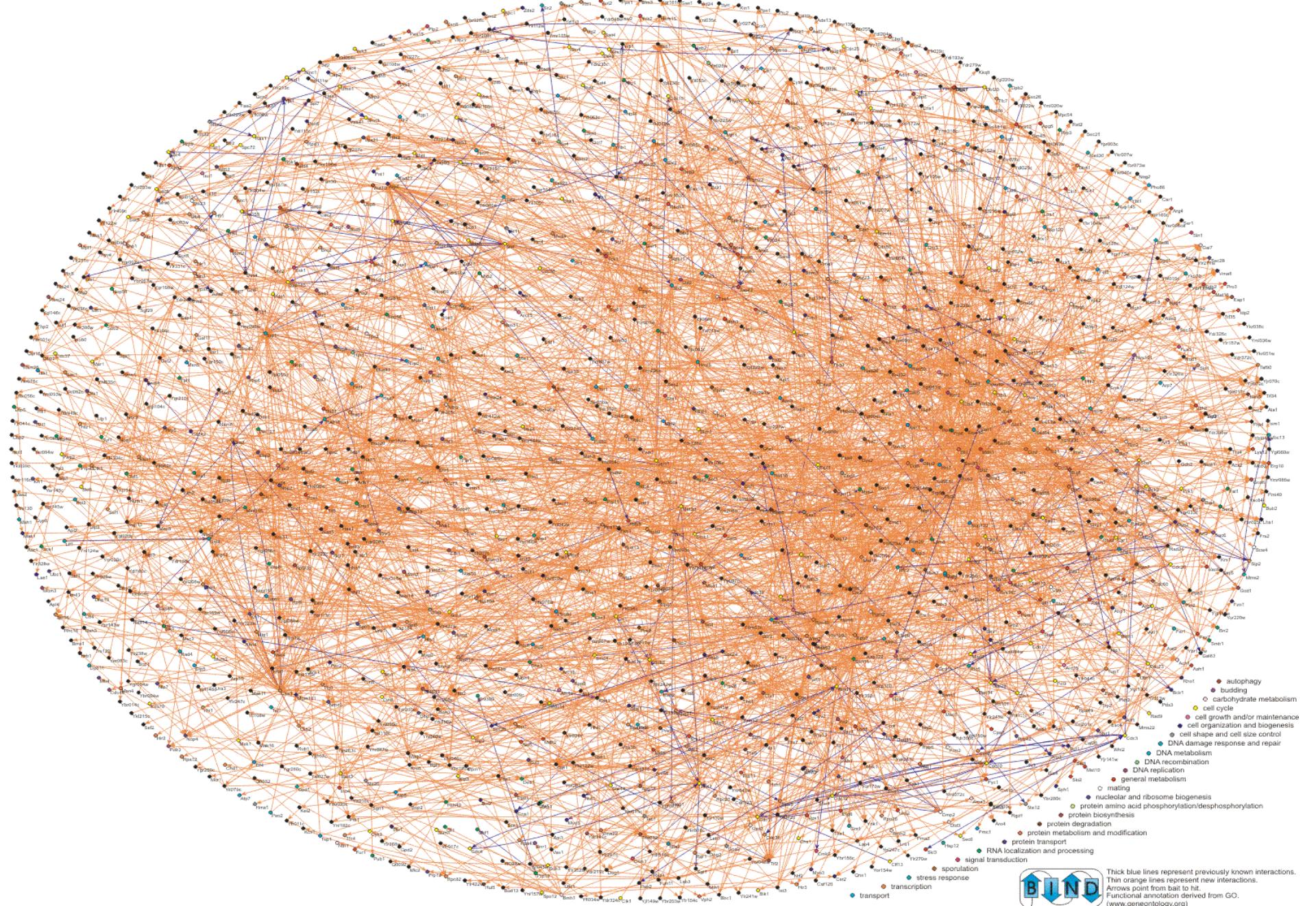


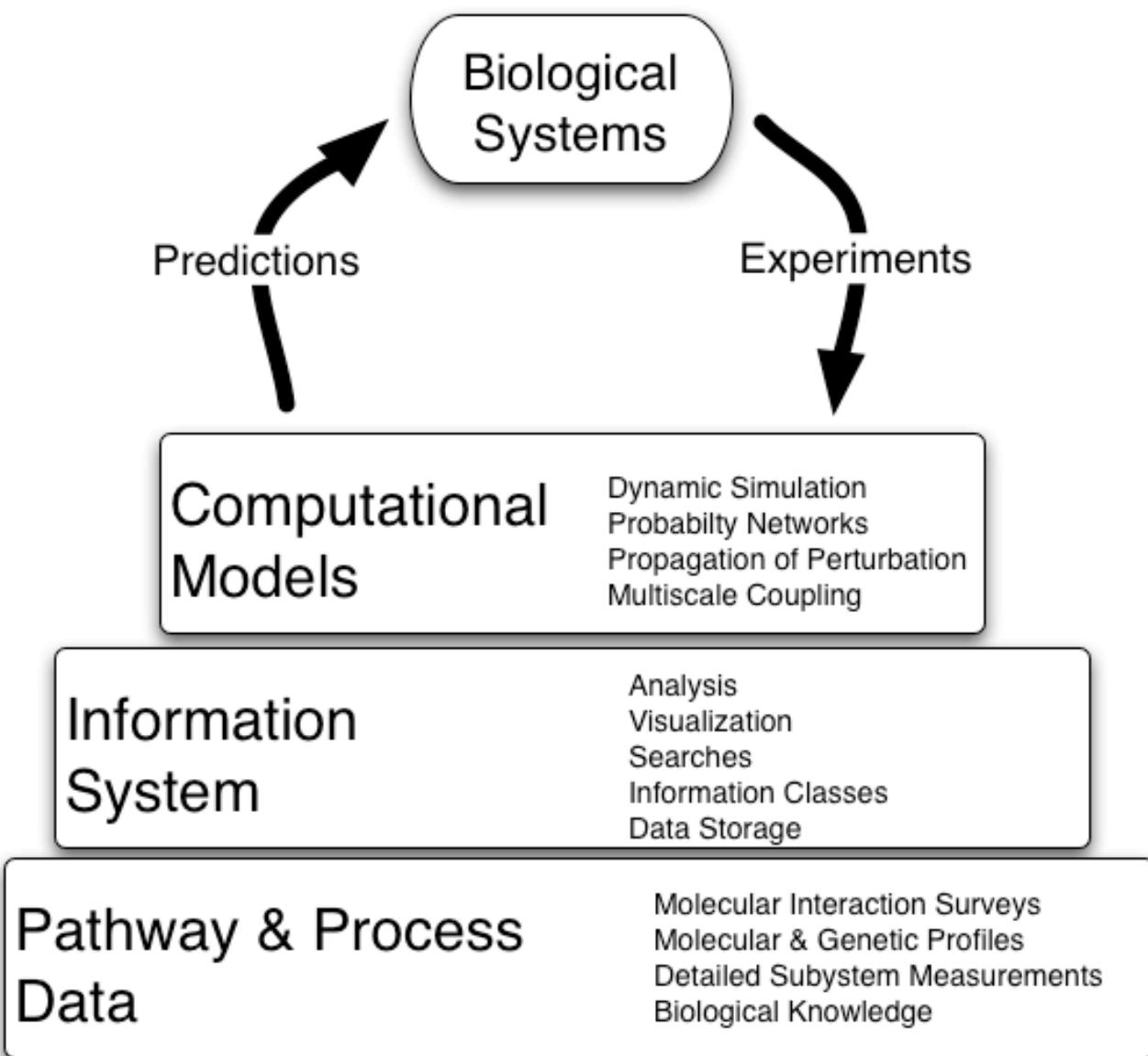
## Endomesoderm Specification to 30 Hours

March 8, 2005



# Systematic identification of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry

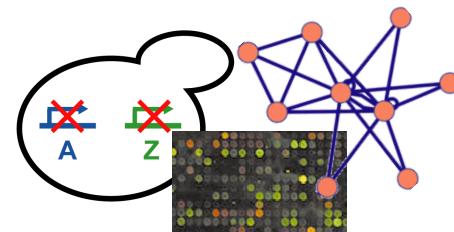
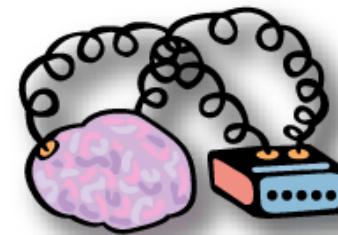
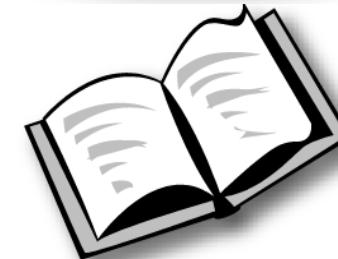
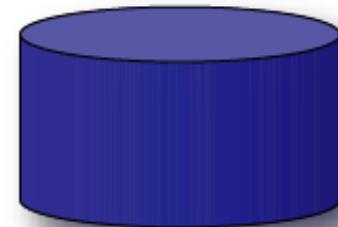




Pathway information for systems biology, Cary MP, Bader GD, Sander C, FEBS Lett. 2005 Mar 21;579(8):1815-20

# Pathway Information

- Databases
  - Fully electronic
  - Easily computer readable
- Literature
  - Increasingly electronic
  - Human readable
- Biologist's brains
  - Richest data source
  - Limited bandwidth access
- Experiments
  - Basis for models



# Pathway Databases

<p><b>Metabolic Pathways</b></p> <p>amAZE - Protein Function and Biochemical Pathways Project          BioCyc - BioCyc Knowledge Library          BioModels - BioModels Database          BioSilico - BioSilico</p> <p>BRENDA - Comprehensive Enzyme Information System          CellML Repository - CellML Model Repository          EcoCyc - Encyclopedia of C. elegans Genes and Metabolism          EMP - Enzymes and Metabolic Pathways Database          ENZYME - Enzyme nomenclature database          GeneNet - Genetic Networks          GenMAPP - Gene MicroArray Pathway Profiler          GOLDb - Genomics of Lipid-associated Disorders          Indigo - Gene Neighborhoods and Codon Usage          Infrat - Integrated relational Enzyme database          KEGG - Kyoto Encyclopedia of Genes and Genomes          LIGAND - Database of Chemical Compounds and Reactions in Biological Pathways          Malaria - Malaria Parasite Metabolic Pathways          MetaCore - MetaCore pathway database          MetaCyc - Metabolic Pathway Database          MIPS CYGD - MIPS Comprehensive Yeast Genome Database          Monod - Monod's Notebook and Database          MPB - Metabolic Pathways of Biochemistry          MRAD - Metabolic Reaction Analysis Database          NetBioChem - Medical Biochemistry Resource          Nicobion Minimaps - IUBMB-Nicobion Minimaps          PathArt - Pathway Articulator          PathDB - Pathways Database          PATKA - Pathway Analysis Tool for Integration and Knowledge Acquisition          PDS - Pathways Database System          PharmGKB - The Pharmacogenetics and Pharmacogenomics Knowledge Base          ProcessDB - ProcessDB          Reactome - Reactome KnowledgeBase          SBML Model Repository - SBML Model Repository          SoyBase - USDA-ARS Soybean Genetics and Genome Database          TECR - Thermodynamics of Enzyme-Catalyzed Reactions          UMBB - Biocatalysis/Biodegradation Database          WIT - What Is There?</p> <p><b>Protein-Protein Interactions</b></p> <p>3DID - 3D interacting domains          ABCdb - Archaea and Bacteria ABC transporter database          ACF5 - Alliance for Cellular Signaling Molecule Pages Database          AFifuse - Functional Associations of Proteins in Complete Genomes          aMaze - Alarne Scanning Energistics Database          ASPD - Artificial Selected Proteins/Peptides Database          BD - Binding Interface Database          BIND - Biomolecular Interaction Network Database          BindingDB - The Binding Database</p>	<p>BRITE - Biomolecular Relations in Information Transmission and Expression          CTDB - Calmodulin Target Database          DDB - Database of Domain Interactions and Bindings          DIP - Database of Interacting Proteins          Doodle - Database of oligomerization domains from lambda experiments          DopaNet - DopaNet          DRC - Database of Ribosome Crosslinks          DSM - Dynamic Signaling Maps          FIMM - Functional Molecular Immunology          FlyNet - Flylets          FusionDB - Prokaryote Gene Fusion Events          GPCR-PD - G protein-coupled receptors protein database          GRID - General Repository for Interaction Datasets          GroEL_PPI - Proteins that interact with GroEL, and factors that affect their release          HIV-1 at NCBI - HIV-1, Human Protein Interaction Database at NCBI          HIVMD - HIV Molecular Immunology Database          HP-DPI - Helicobacter pylori Database of Protein Interactions          HPID - Human Protein Interaction Database          HPRD - Human Protein Reference Database          HSV1_PPI - Protein-Protein Interactions Table for Human herpesvirus HumanPSD - Human Proteome Survey Database          ICB5 - Inter-Chain Beta-Sheets          IntAct - IntAct          INTERACT - INTERACT protein-protein interaction database          InterDom - Database of Interacting Domains          Interfaces - DATASET OF PROTEIN-PROTEIN INTERFACES          Interolog - Interolog/Regulog Database          JenPep - JenPep Peptide binding database          KDBI - Kinetic Data of Bio-molecular Interactions database          KinaseDB - Kinase Pathway Database          MHKEP - Database of MHC binding peptides          MINT - Molecular Interaction Database          MIPS CYGD - MIPS Comprehensive Yeast Genome Database          MPID - MHC-Peptide Interaction Database          MycoPathPD - Human Fungal Pathogens Proteome Database          NetPro - Molecular Connections NetPro          oTFID - Object Oriented Transcription Factors Database          oTFIDH - The Online Predicted Human Interaction Database          PathCat - Curagen Pathcalling          PathDB - Pathways Database          PDZbase - PDZ domain protein-protein interaction database          Phospho-ELM - Post-translational phosphorylation database          PhosphoSite - Database of phosphorylation sites          PhosphoSite - Cell Signaling Technology's PhosphoSite Database          PIBASE - PIBASE          PIMdb - Drosophila Protein Interaction Map Database          PINwriter - Protein Interaction Map - Hybrigenics          PINdb - Proteins Interacting in the Nucleus database          POINT - Prediction of Interactome          PombePD - Schizosaccharomyces pombe Proteome Database</p>	<p>PPID - Protein-Protein Interaction Database          PPI (FANTOM) - RIKEN FANTOM Protein Protein Interaction Viewer          POS - Protein Quaternary Structure database          Predictome - Predictome          ProChart - Pro-Chart database of signal transduction pathway information          ProLinks - ProLinks          ProMesh - ProMesh Protein-Protein Interaction Database          ProN - Protein-protein Interaction Database</p> <p>CSNDb - Cell Signaling Networks Database          DOQCS - Database of Quantitative Cellular Signaling          DSM - Dynamic Signaling Maps          eMIM - Electronic Molecular Interaction Map          GenNet - Genetic Networks          GenMAP - Gene MicroArray Pathway Profiler          GOLDb - Genomics of Lipid-associated Disorders          GON - Genomic Object Net          INOH - Integrating Network Objects with Hierarchies</p>	<p>STCDB - Signal Transduction Classification Database          STKE - Signal Transduction Knowledge Environment          TRAPATH - Signal Transduction Browser          TRMP - Therapeutic-relevant Multiple Pathways Database          TRRD - Transcription Regulatory Regions Database          XPD - Phosphorylation Site Database</p>	<p>CTD - Comparative Toxicogenomics Database          HetPDB_Navi - HetPDB Navi          Kibanc - Kibanc          Ligand Depot - Ligand Depot          MDB - Metalprotein Database          NRR - Nuclear Receptor Resource          OROB - Olfactory Receptor Database          PDB-Ligand - PDB-Ligand          PDSP - Psychoactive Drug Screening Program Ki Database          PLD - Protein Ligand Database          Relibase - Protein-ligand database query tool          RedNet - RedNet          TTD - Therapeutic Target Database</p>
<p><b>Pathway Diagrams</b></p> <p>BBD - Biological Biochemical Image Database          BioCarta - BioCarta Pathway Diagrams          BWP - Boehringer Mannheim Biochemical Pathways Wall Chart</p>	<p>DSM - Dynamic Signaling Maps          eMIM - Electronic Molecular Interaction Map</p>	<p>HPRD - Human Protein Reference Database          INOH - Integrating Network Objects with Hierarchies          KEGG - Kyoto Encyclopedia of Genes and Genomes          KMMI - Kohl Molecular Interaction Maps          Malaria - Malaria Parasite Metabolic Pathways</p>	<p>MIPS CYGD - MIPS Comprehensive Yeast Genome Database          MPB - Metabolic Pathways of Biochemistry          NetBioChem - Medical Biochemistry Resource          Nicobion Minimaps - IUBMB-Nicobion Minimaps          PharmGKB - The Pharmacogenetics and Pharmacogenomics Knowledge Base          PID - CMAP Pathway Interaction Database          SPAD - Signaling Pathway Database          TRMP - Therapeutic-relevant Multiple Pathways Database</p>	<p><b>Genetic Interaction Networks</b></p> <p>BIND - Biomolecular Interaction Network Database          GeneNet - Genetic Networks          GenePath - GenePath</p>
<p><b>Protein Sequence Focused</b></p> <p>ASC - Active Sequences Collection          Inbase - The Inbase Database          MEROPS - MEROPS Peptidase Database          NRR - Nuclear Receptor Resource          OROB - Olfactory Receptor Database          PANTHER - PANTHER</p>	<p>PhosphoSite - Database of phosphorylation sites</p>	<p>RELEASE - Restriction Enzyme Database          SCOPEC - Database of protein catalytic domains          Senta - Senta Signal Transduction Database          TGBB - Tumor Gene Database</p>	<p>COMPEL - Database on Composite Regulatory Elements          DBTFBS - Database of <i>Bacillus subtilis</i> Promoters and Transcription Factors</p>	<p><b>Transcription Factors / Genetic Regulatory Networks</b></p> <p>DPInteract - DNA-Protein Interactions Database          GeNet - GeNet Networks Database          HoPro - HoPro Pro          Interolog - Interolog/Regulog Database          JASPAR - JASPAR Transcription Factor Binding Profile Database          MAPPER - MAPPER</p>
<p><b>Other</b></p> <p>SELEX_D8 - Randomized DNA/RNA sequence database          AARSD8 - Aminocycl-RNA Synthetase Database          MedGene - MedGene</p>	<p>ooFD - Object Oriented Transcription Factors Database</p>	<p>RegulonDB - Database on Transcriptional Regulation and Genome Organization          SGD - The Promoter Database of <i>Saccharomyces cerevisiae</i></p>	<p>TRACTOR_DB - TRACTOR_DB          TRANSAC - Transcription Factor Database          TRED - Transcriptional Regulatory Element Database          TRRD - Transcription Regulatory Regions Database</p>	<p><b>Protein-Compound Interactions</b></p> <p>CLUB - Computed Ligand Binding Energy</p>

## 310 Pathway Databases!

- Arguably the most accessible data source, but...

- Varied formats, representation, coverage

- Pathway data extremely difficult to combine and use

Pathguide Pathway Resource List (<http://www.pathguide.org>)

# Pathguide

the pathway resource list

HOME PROTEAS CETO MIRKO

## 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](#)
- [Availability](#)
- [Standards](#)

## Statistics

- [Analyze Pathguide](#)

## Contact

- [Comments, Questions, Suggestions are Always Welcome!](#)

## Complete Listing of All Pathguide Resources

Pathguide contains information about 310 biological pathway 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

**Major update**  
All resources were recently reviewed and many new ones were added

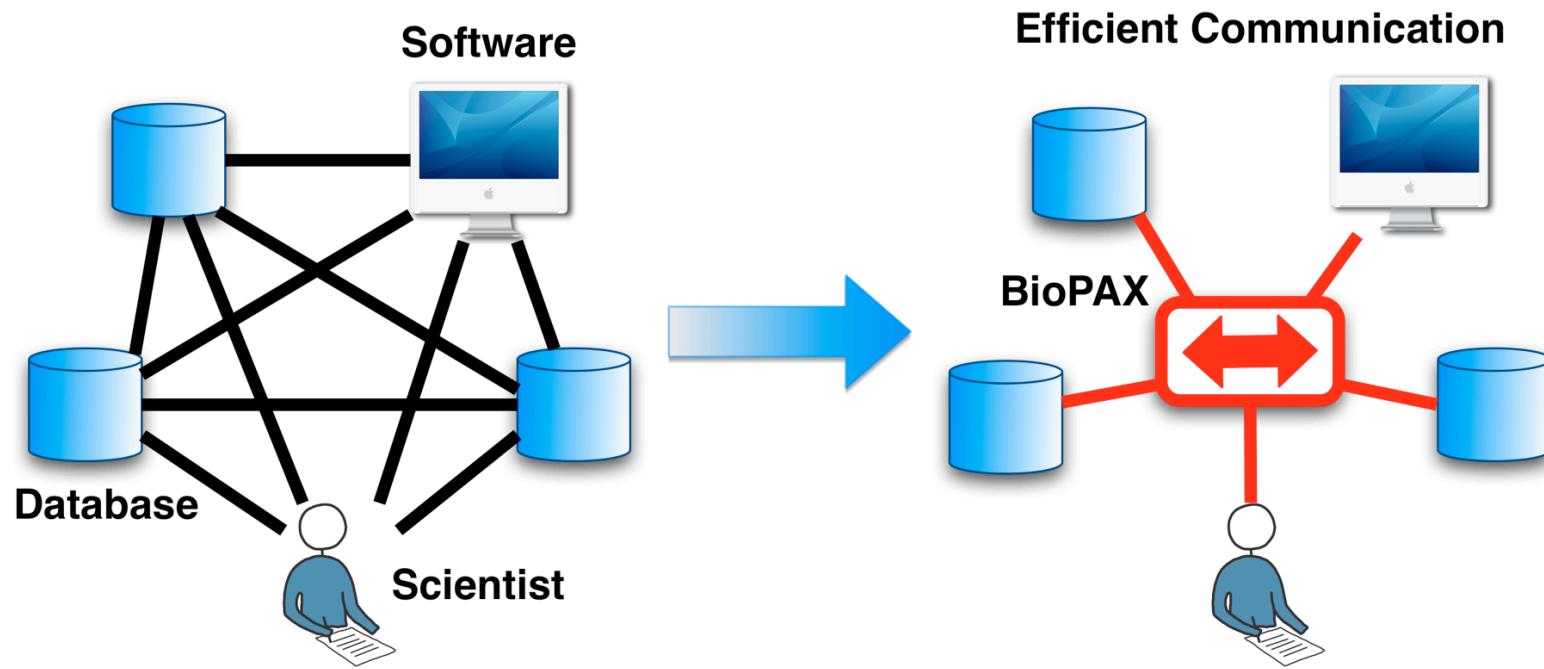
**Get the Stats**  
Detailed Pathguide resource statistics now available

**Pathguide Published**  
Please cite the [Pathguide Publication](#)

## Protein-Protein Interactions

Database Name (Order: alphabetically   by web popularity ⓘ)	Full Record	Availability	Standards
3DID - 3D interacting domains	<a href="#">Details</a>	<a href="#">Free</a>	
ADAN - Prediction of protein-protein interaction of modular domains	<a href="#">Details</a>	<a href="#">Free</a>	
AllFuse - Functional Associations of Proteins in Complete Genomes	<a href="#">Details</a>		
aMAZE - Protein Function and Biochemical Pathways Project	<a href="#">Details</a>	<a href="#">Free</a>	
APID - Agile Protein Interaction DataAnalyzer	<a href="#">Details</a>	<a href="#">Free</a>	
ASEdb - Alanine Scanning Energetics Database	<a href="#">Details</a>	<a href="#">Free</a>	
ASPD - Artificial Selected Proteins/Peptides Database	<a href="#">Details</a>	<a href="#">Free</a>	
BID - Binding Interface Database	<a href="#">Details</a>	<a href="#">Free</a>	
BIND - Biomolecular Interaction Network Database	<a href="#">Details</a>	<a href="#">Free</a>	
BioGRID - Biological General Repository for Interaction Datasets	<a href="#">Details</a>		
BRITE - Biomolecular Relations in Information Transmission and Expression	<a href="#">Details</a>	<a href="#">Free</a>	
CA1Neuron - Pathways of the hippocampal CA1 neuron	<a href="#">Details</a>	<a href="#">Free</a>	
Cancer Cell Map - The Cancer Cell Map	<a href="#">Details</a>	<a href="#">Free</a>	
CellCircuits - CellCircuits	<a href="#">Details</a>	<a href="#">Free</a>	
CPDB - ConsensusPathDB	<a href="#">Details</a>		

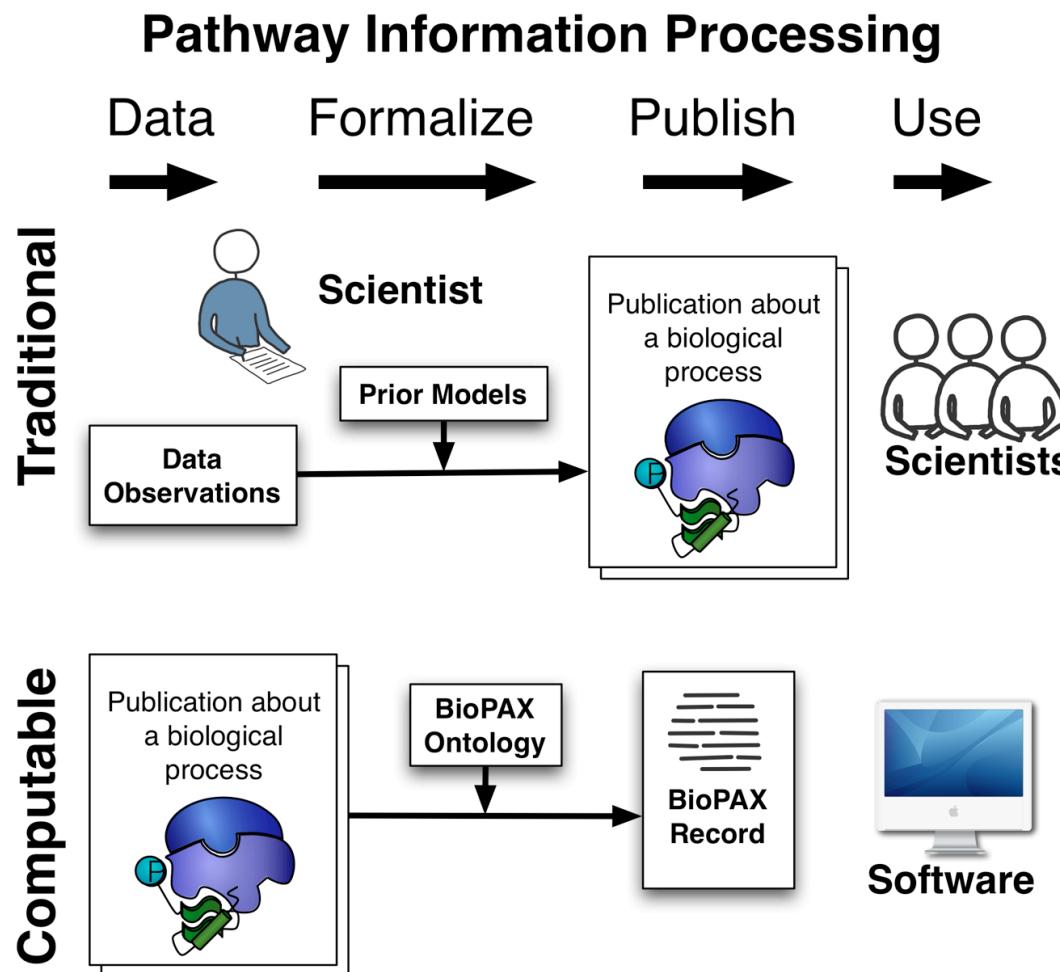
# Biological Pathway Exchange (BioPAX)



Before BioPAX  
>100s DBs and tools  
Tower of Babel

After BioPAX  
Unifying language

Reduces work, promotes collaboration, increases accessibility



# BioPAX Pathway Language

- Represent:
  - Metabolic pathways
  - Signaling pathways
  - Protein-protein, molecular interactions
  - Gene regulatory pathways
  - Genetic interactions
- Community effort: pathway databases distribute pathway information in standard format

# Ontologies: Components

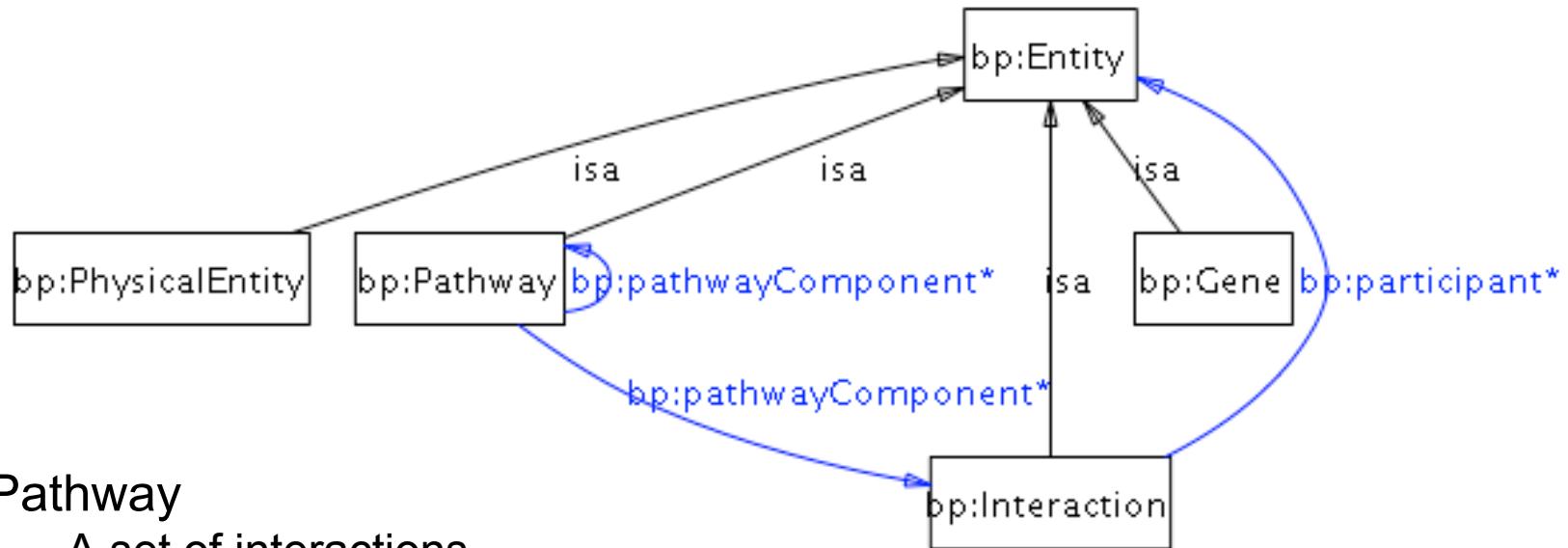
- Classes, relations & attributes, constraints, objects, values
- Classes (AKA “Concepts”, “Types”)
  - Arranged into a specialization hierarchy (AKA “Taxonomy”)
    - Parent-child relationships between classes
    - Class A is a parent of class B iff all instances of B are also instances of A
  - E.g. “Protein”, “RNA”, “Reaction”
- Relations & Properties (AKA “Slots”, “Attributes”, “Fields”)
  - Classes *have* properties, which may have values of specific types
  - Relationships: the value type is some other class in the ontology
    - E.g. “Substrate”, “Transporter”, “Participant”
  - Attributes: the value type is a simple data type
    - E.g. “Molecular Wt.”, “Sequence”, “ $\Delta G$ ”

From Peter Karp, “Ontologies: Definitions, Components, Subtypes”, SRI International, presentation available at <http://www.biopax.org>

# Ontologies: Components (cont)

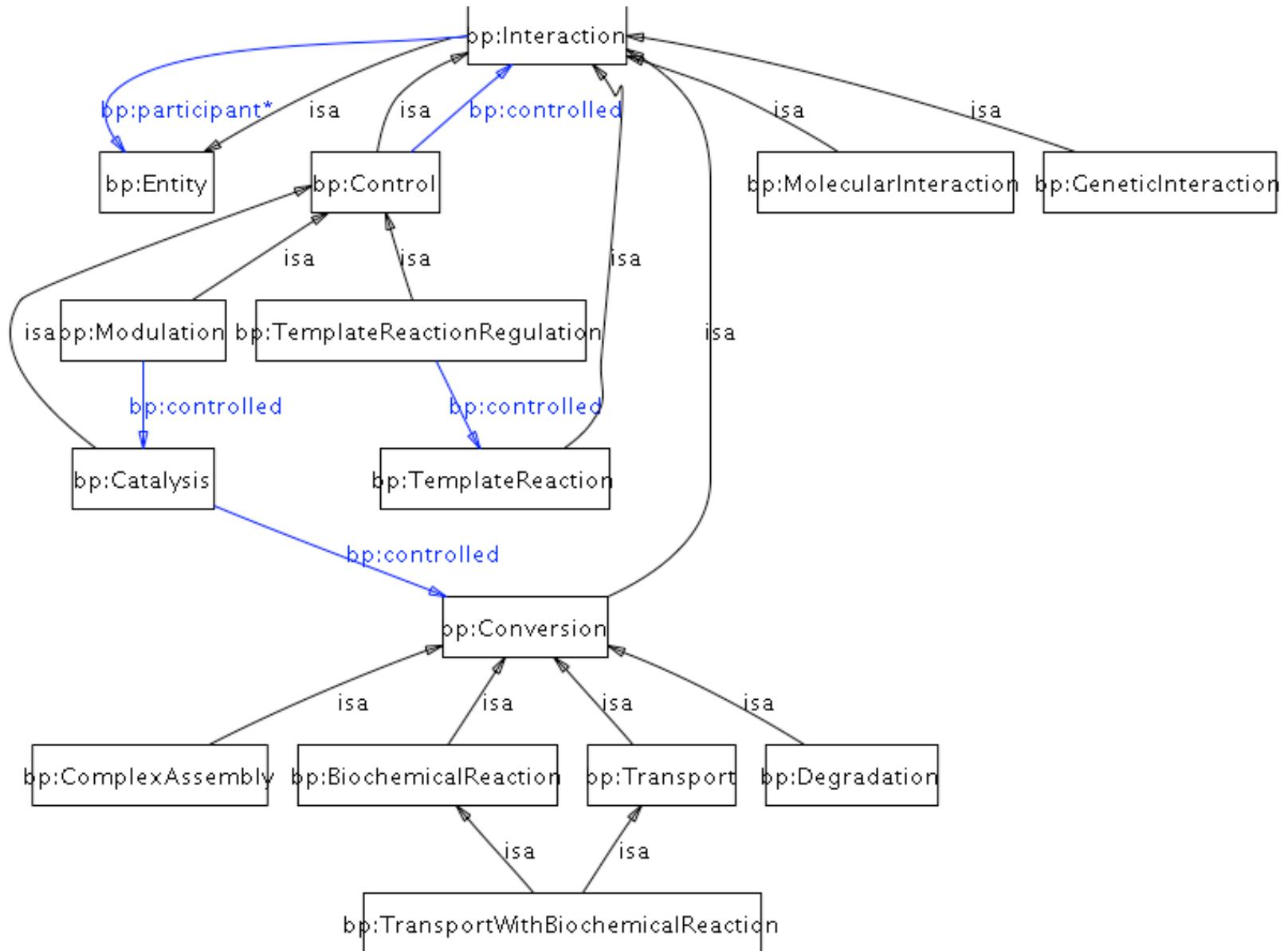
- Constraints
  - Define allowable values and connections within an ontology
  - E.g. “MOLECULAR\_WT must be a positive real number”
- Objects and Values
  - Objects are instances of classes
  - Values occupy the slots of those instances
  - Strictly speaking, an ontology with instances is a knowledge base
  - Beyond the scope of BioPAX workgroup, our users will create the instances of classes in the BioPAX ontology

# BioPAX Structure

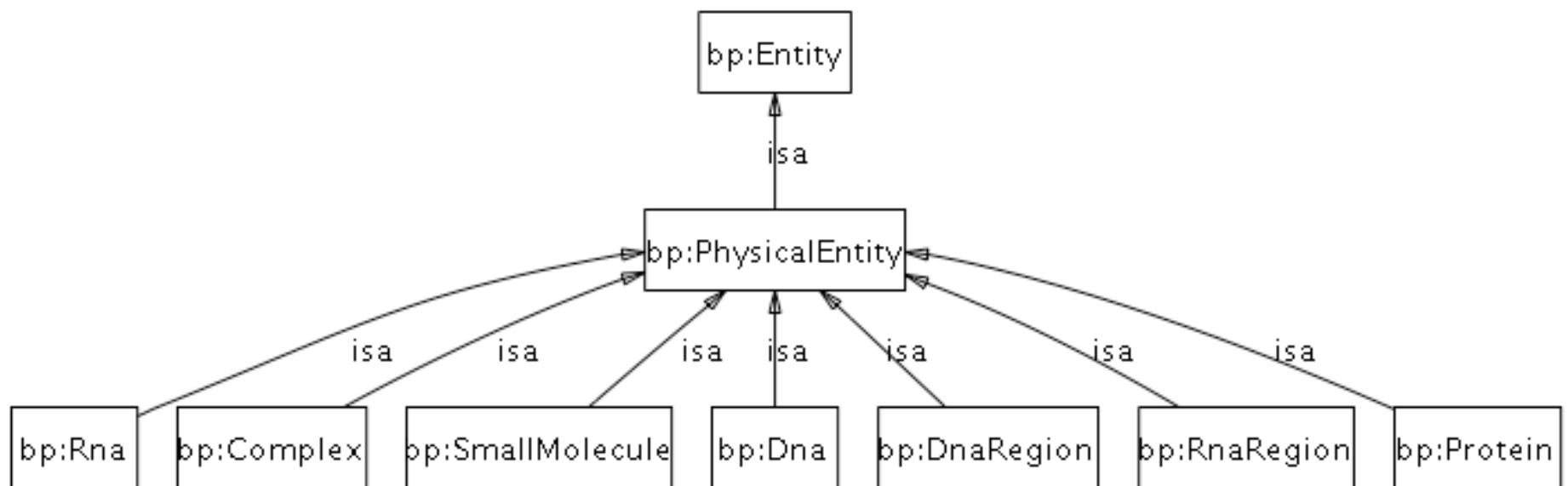


- Pathway
  - A set of interactions
  - E.g. Glycolysis, MAPK, Apoptosis
- Interaction
  - A basic relationship between a set of entities
  - E.g. Reaction, Molecular Association, Catalysis
- Physical Entity
  - A building block of simple interactions
  - E.g. Small molecule, Protein, DNA, RNA

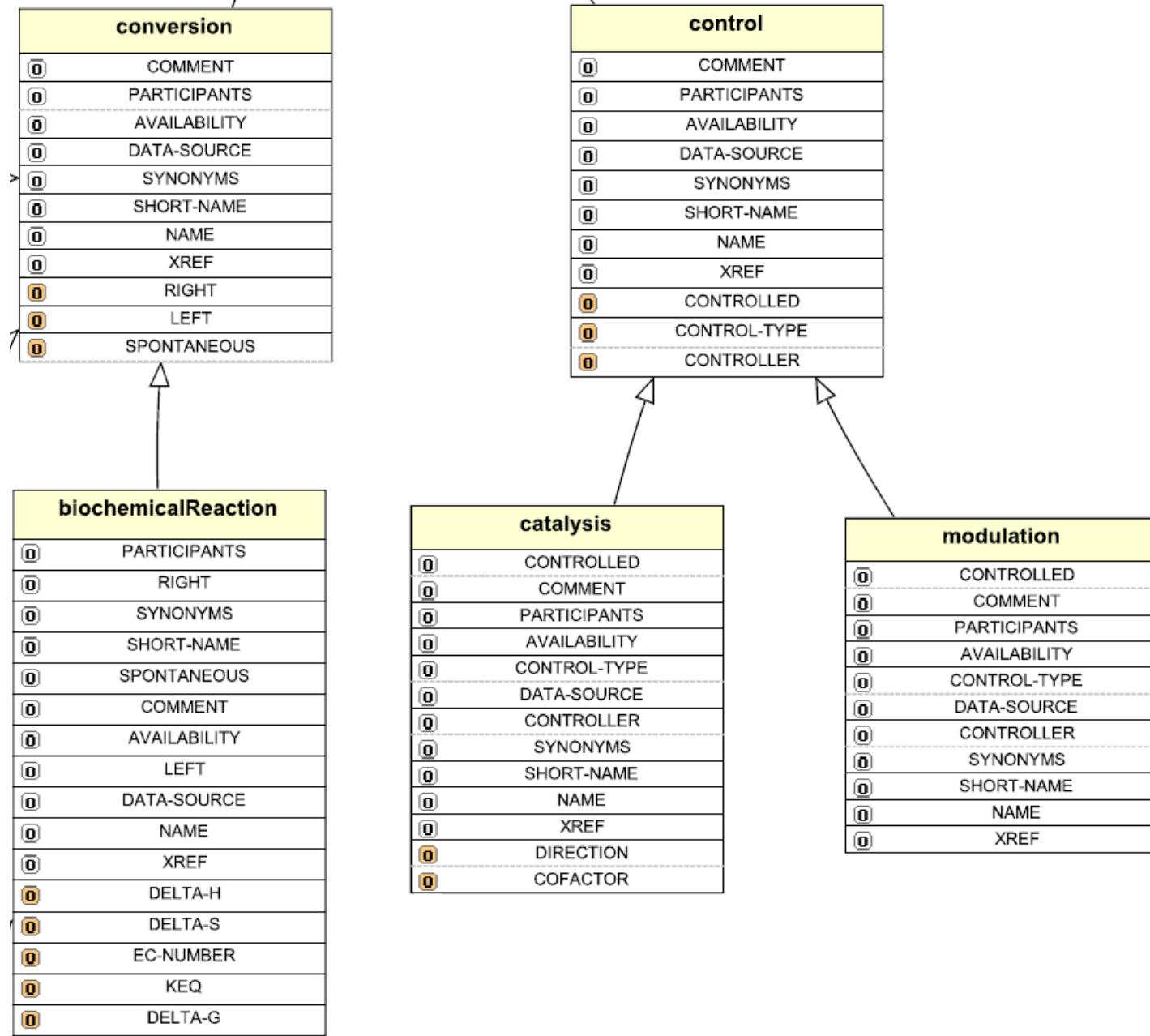
# BioPAX: Interactions



# BioPAX: Physical Entities

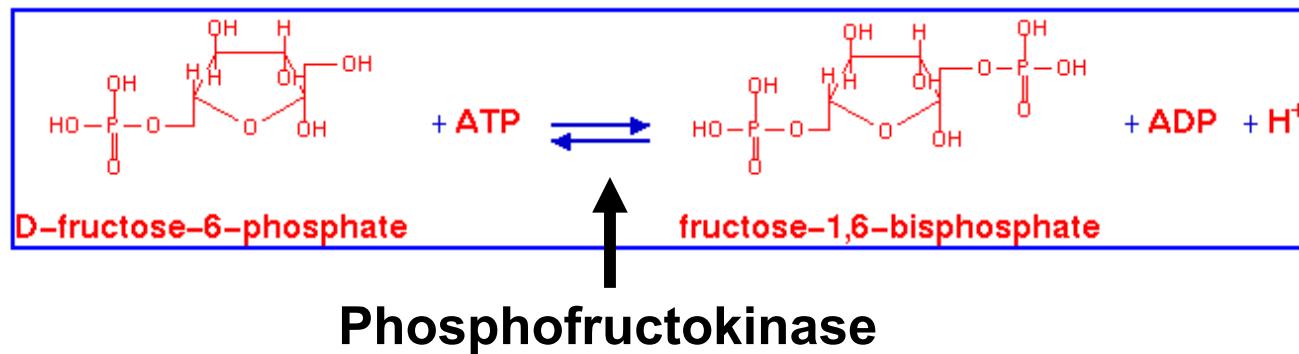


# BioPAX Ontology



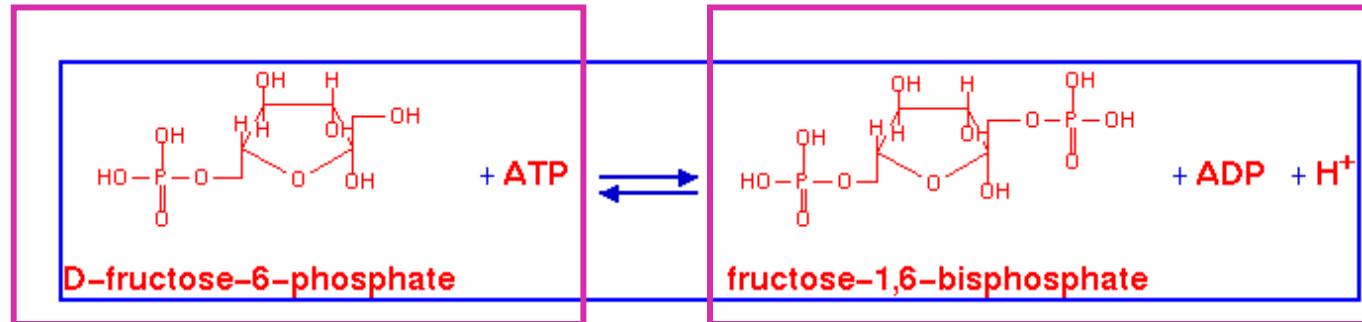
# XML Snippet

```
<bp:BiochemicalReaction rdf:ID="glucokinase">
  <bp:eCNumber rdf:datatype="&xsd:string">2.7.1.1 </bp:eCNumber>
  <bp:eCNumber rdf:datatype="&xsd:string">2.7.1.2 </bp:eCNumber>
  <bp:name rdf:datatype="&xsd:string"
    >glucose ATP phosphotransferase      </bp:name>
  <bp:name rdf:datatype="&xsd:string"
    >ATP:D-glucose 6-phosphotransferase </bp:name>
  <bp:conversionDirection rdf:datatype="&xsd:string">REVERSIBLE</bp:conversionDirection>
  <bp:standardName rdf:datatype="&xsd:string"
    >b-D-glu =&gt; b-D-glu-6-p</bp:standardName>
  <bp:displayName rdf:datatype="&xsd:string"
    >beta-D-glu + ATP =&gt; beta-D-glu-6-p + ADP</bp:displayName>
  <bp:spontaneous rdf:datatype="&xsd:boolean">true</bp:spontaneous>
  <bp:xref rdf:resource="#KEGG_R01786"/>
  <bp:dataSource rdf:resource="#KEGG"/>
  <bp:dataSource rdf:resource="#aMAZE"/>
  <bp:left rdf:resource="#ATP"/>
  <bp:left rdf:resource="#alpha-D-glucose"/>
  <bp:right rdf:resource="#ADP"/>
  <bp:right rdf:resource="#alpha-D-glucose_6-phosphate"/>
  <bp:participantStoichiometry rdf:resource="#Stoichiometry_43"/>
  <bp:participantStoichiometry rdf:resource="#Stoichiometry_37"/>
  <bp:participantStoichiometry rdf:resource="#Stoichiometry_52"/>
  <bp:participantStoichiometry rdf:resource="#Stoichiometry_49"/>
</bp:BiochemicalReaction>
<bp:Catalysis rdf:ID="glucokinase Converts alpha-D-glu to alpha-D-glu-6-p">
  <bp:comment rdf:datatype="&xsd:string"
    >The source of this data did not store catalyses of reactions as separate objects, so there
  <bp:catalysisDirection rdf:datatype="&xsd:string">LEFT-T0-RIGHT</bp:catalysisDirection>
  <bp:standardName rdf:datatype="&xsd:string"
    >GLK -&gt; (a-D-glu <=&gt; a-D-glu-6-p)</bp:standardName>
  <bp:controlType rdf:datatype="&xsd:string">ACTIVATION</bp:controlType>
  <bp:displayName rdf:datatype="&xsd:string"
```



## Biochemical Reaction Glycolysis Pathway

Source: BioCyc.org

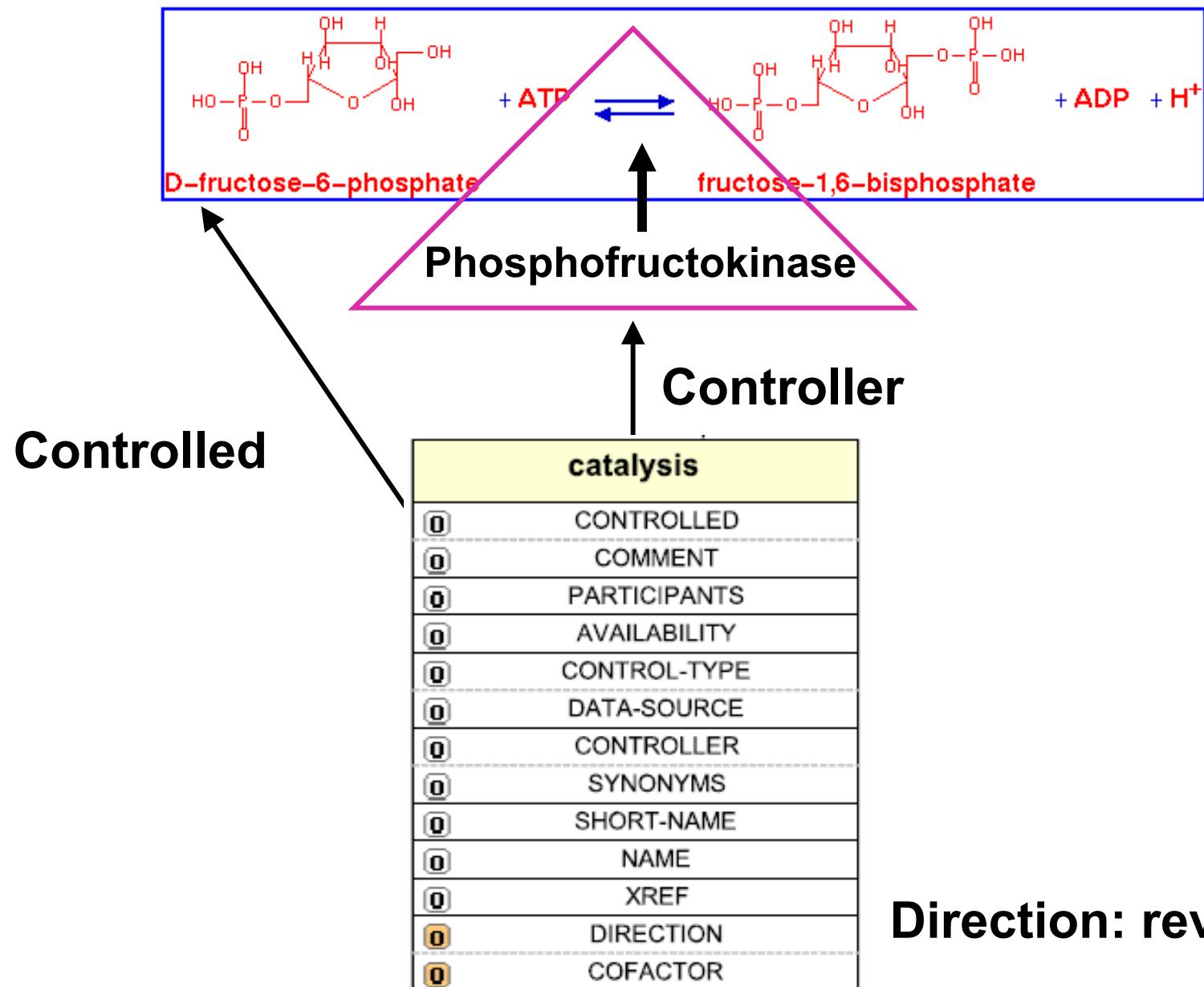


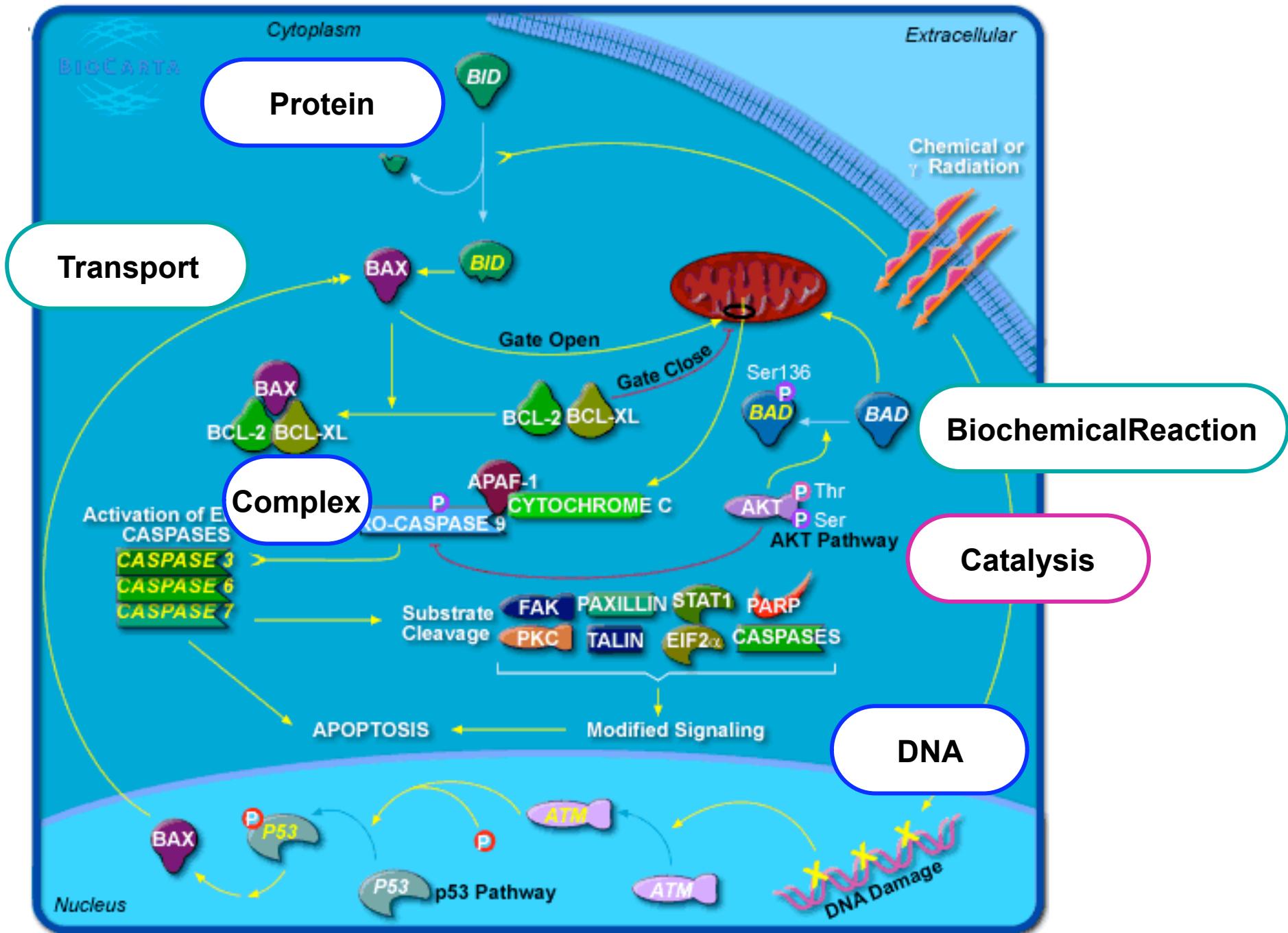
**Left**

**Right**

biochemicalReaction	
<input type="checkbox"/>	PARTICIPANTS
<input type="checkbox"/>	RIGHT
<input type="checkbox"/>	SYNONYMS
<input type="checkbox"/>	SHORT-NAME
<input type="checkbox"/>	SPONTANEOUS
<input type="checkbox"/>	COMMENT
<input type="checkbox"/>	AVAILABILITY
<input type="checkbox"/>	LEFT
<input type="checkbox"/>	DATA-SOURCE
<input type="checkbox"/>	NAME
<input type="checkbox"/>	XREF
<input checked="" type="checkbox"/>	DELTA-H
<input checked="" type="checkbox"/>	DELTA-S
<input checked="" type="checkbox"/>	EC-NUMBER
<input checked="" type="checkbox"/>	KEQ
<input checked="" type="checkbox"/>	DELTA-G

**EC # 2.7.1.11**





# Controlled Vocabularies (CVs)

- BioPAX uses existing CVs where available via ControlledVocabulary instances
  - Cellular location: Gene Ontology (GO) component
  - PSI-MI CVs for:
    - Protein post-translational modifications
    - Interaction detection experimental methods
    - Experimental form
  - PATO phenotypic quality ontology
  - Some database providers use their own CVs
    - E.g. BioCyc evidence codes
- More at the Ontology Lookup Service
  - <http://www.ebi.ac.uk/ontology-lookup/>

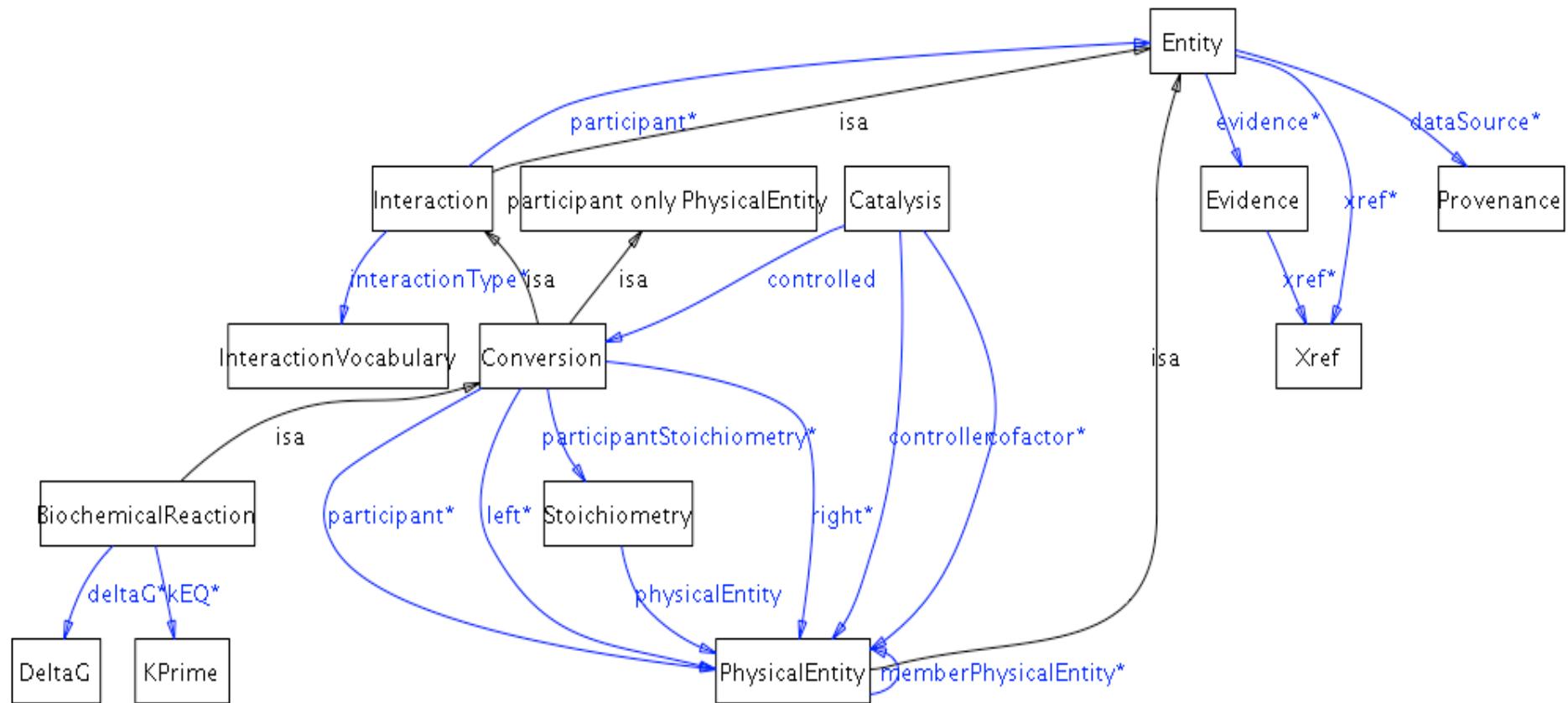
# BioPAX Level 3 (to be ratified)

- States and generics
  - E.g. phosphorylated P53, alcohols
- Gene regulation
  - E.g. Transcription regulation by transcription factors, translation regulation by miRNAs
- Genetic interactions
  - E.g. synthetic lethality, epistasis
- Better controlled vocabulary integration
  - More accessible to reasoners

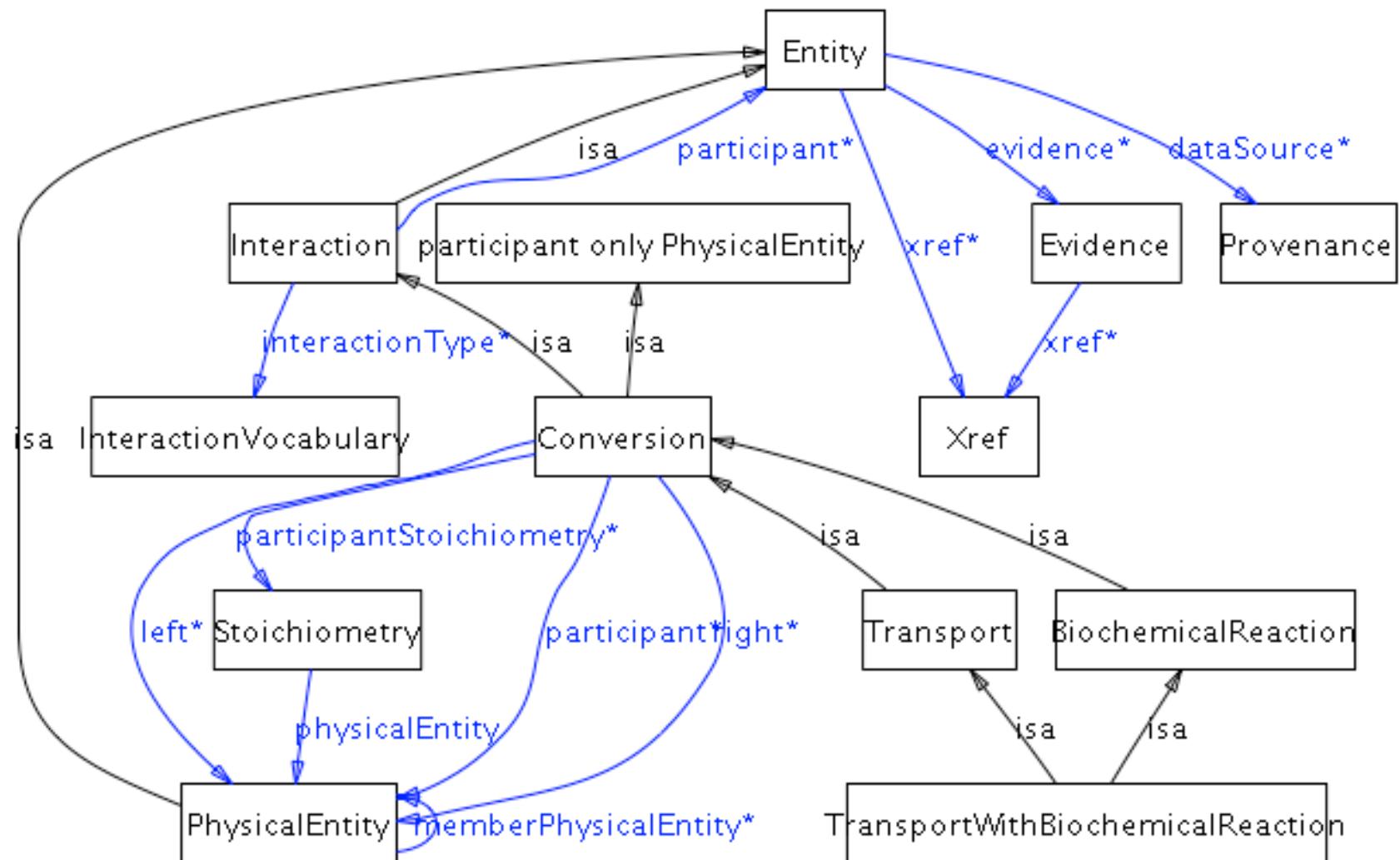
# Worked examples

- Metabolic pathway
  - EcoCyc Glycolysis (energy metabolism pathway)
- Protein-protein interaction
  - Proteomics, PSI-MI
- Signaling pathway step
  - Reactome CHK2-ATM
- Switch to Protégé
- Available from biopax.org
  - <http://www.biopax.org/Downloads/Level2v1.0/biopax-level2.zip>

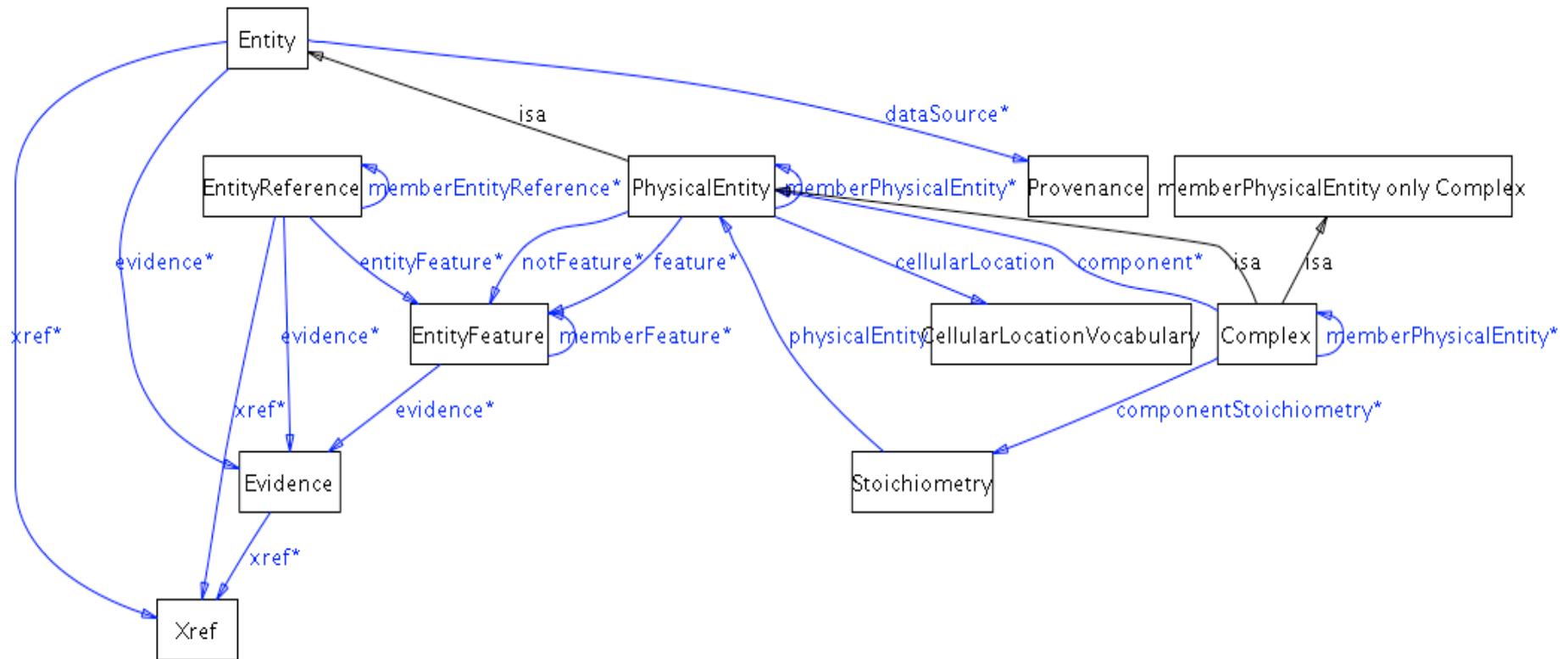
# Biochemical Reaction



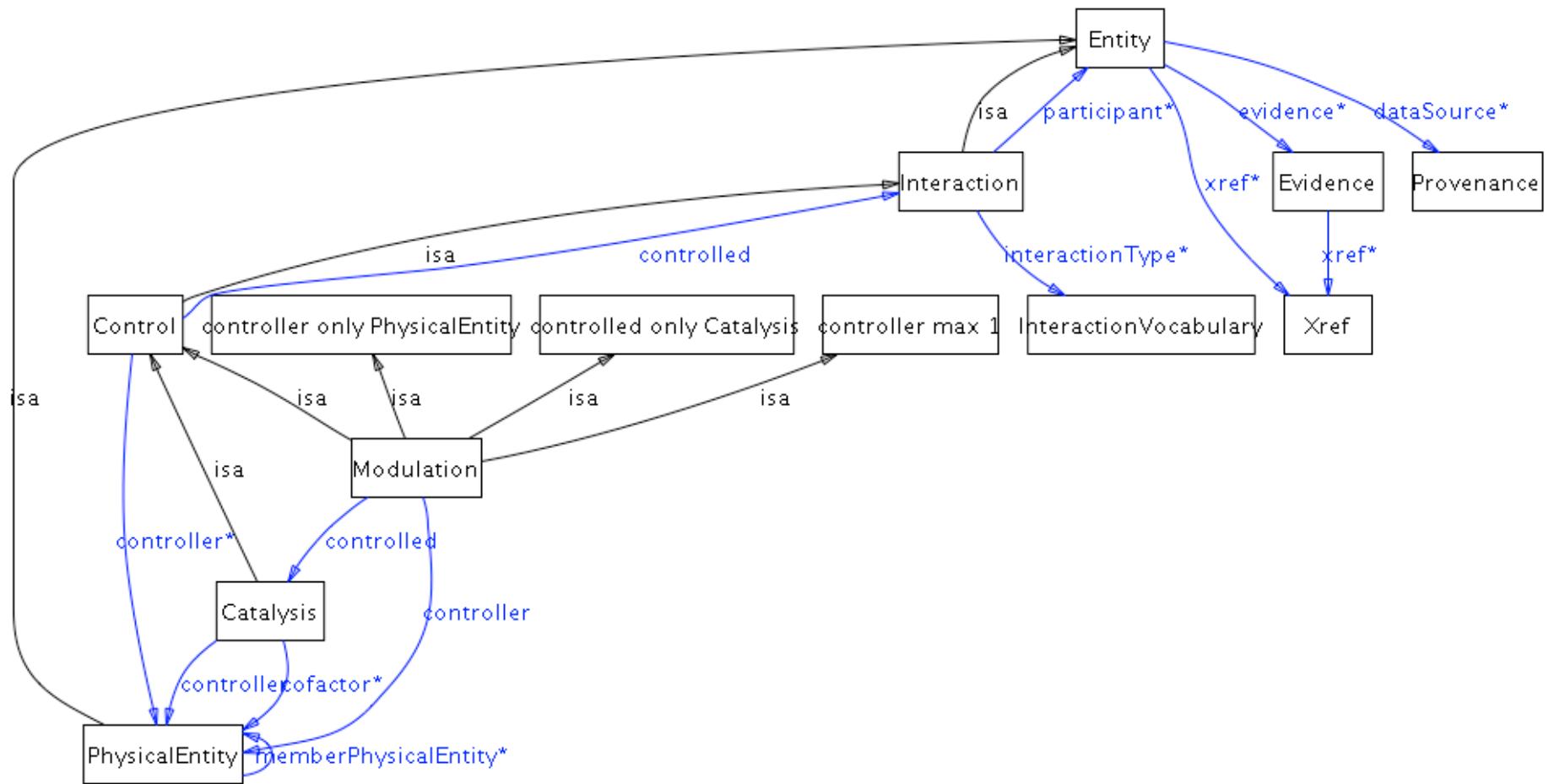
# Transport with Biochemical Reaction



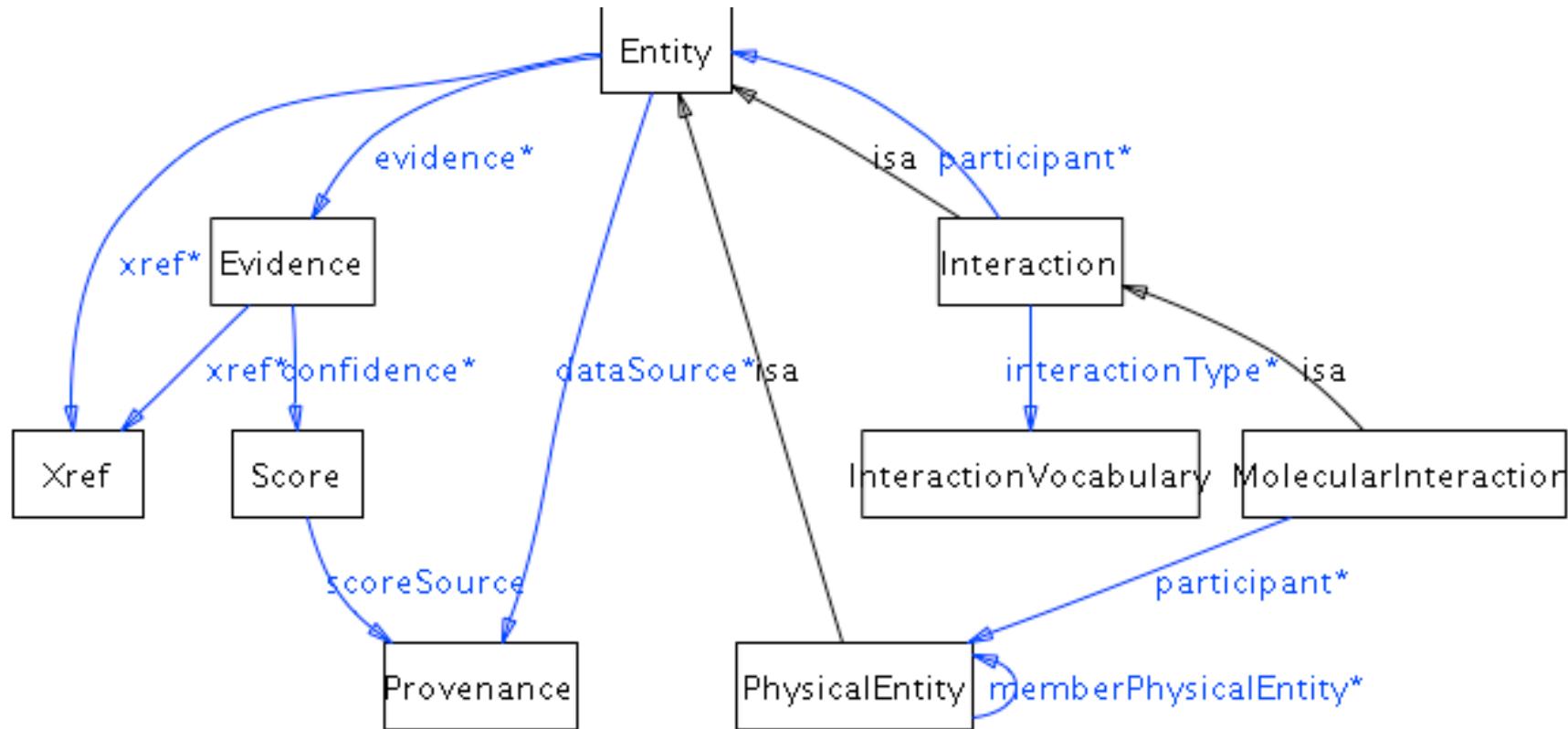
# Complex Assembly



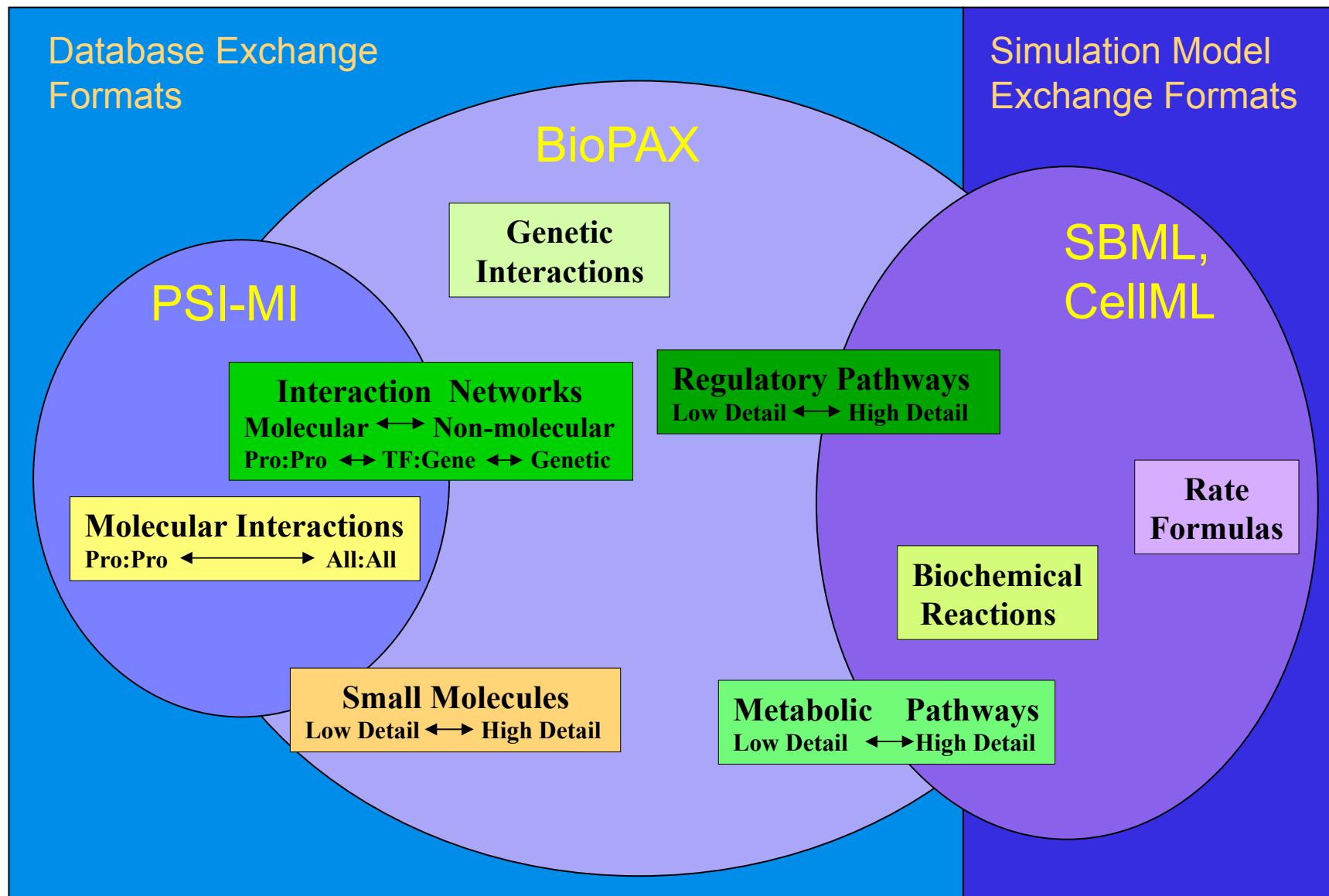
# Modulation



# Molecular Interaction

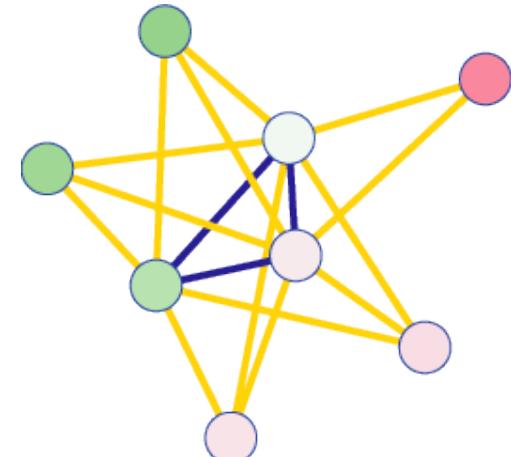
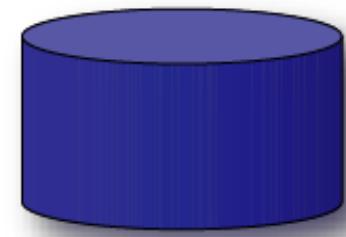


# Exchange Formats in the Pathway Data Space

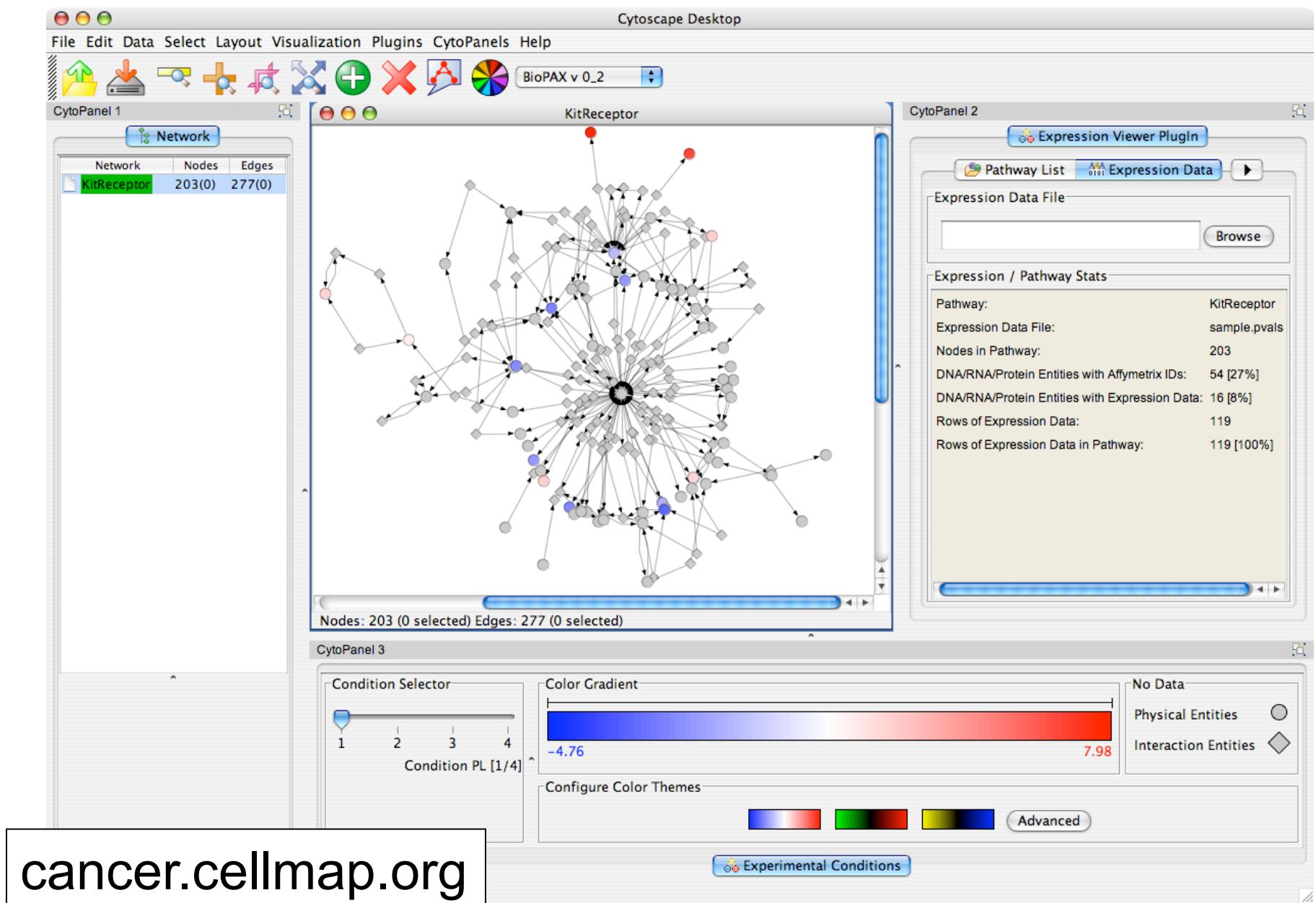


# Using BioPAX

- Databases
  - BioCyc (EcoCyc, MetaCyc, many pathway genome databases)
  - KEGG (KEGG, aMAZE, Sander)
  - MSKCC Cancer Pathway Resource
  - Reactome
  - PSI-MI (via converter)
  - Switch to Pathguide
- Tools
  - cPath, Cytoscape, GenMAPP, PATIKA, QPACA, VisANT
- caBIG

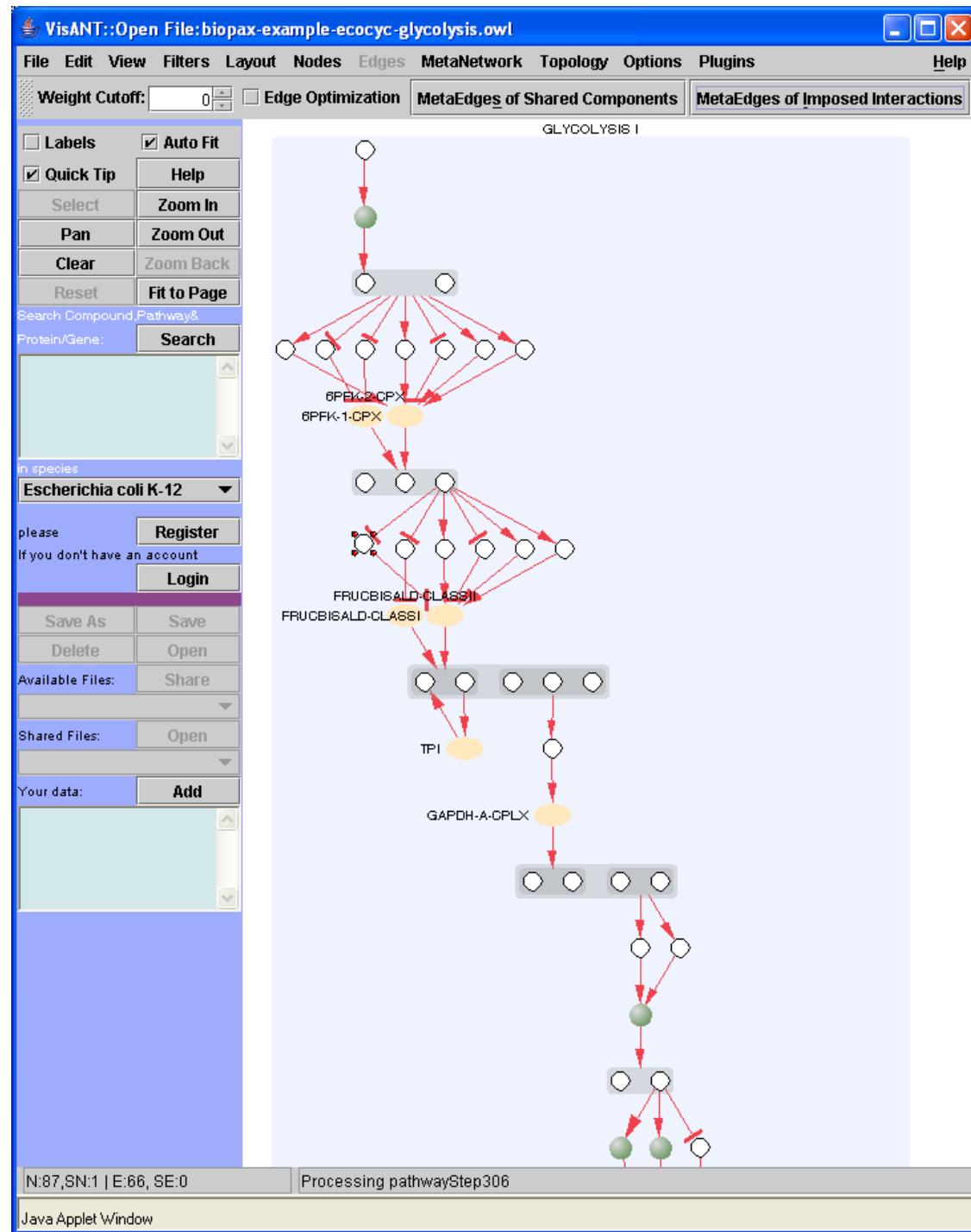


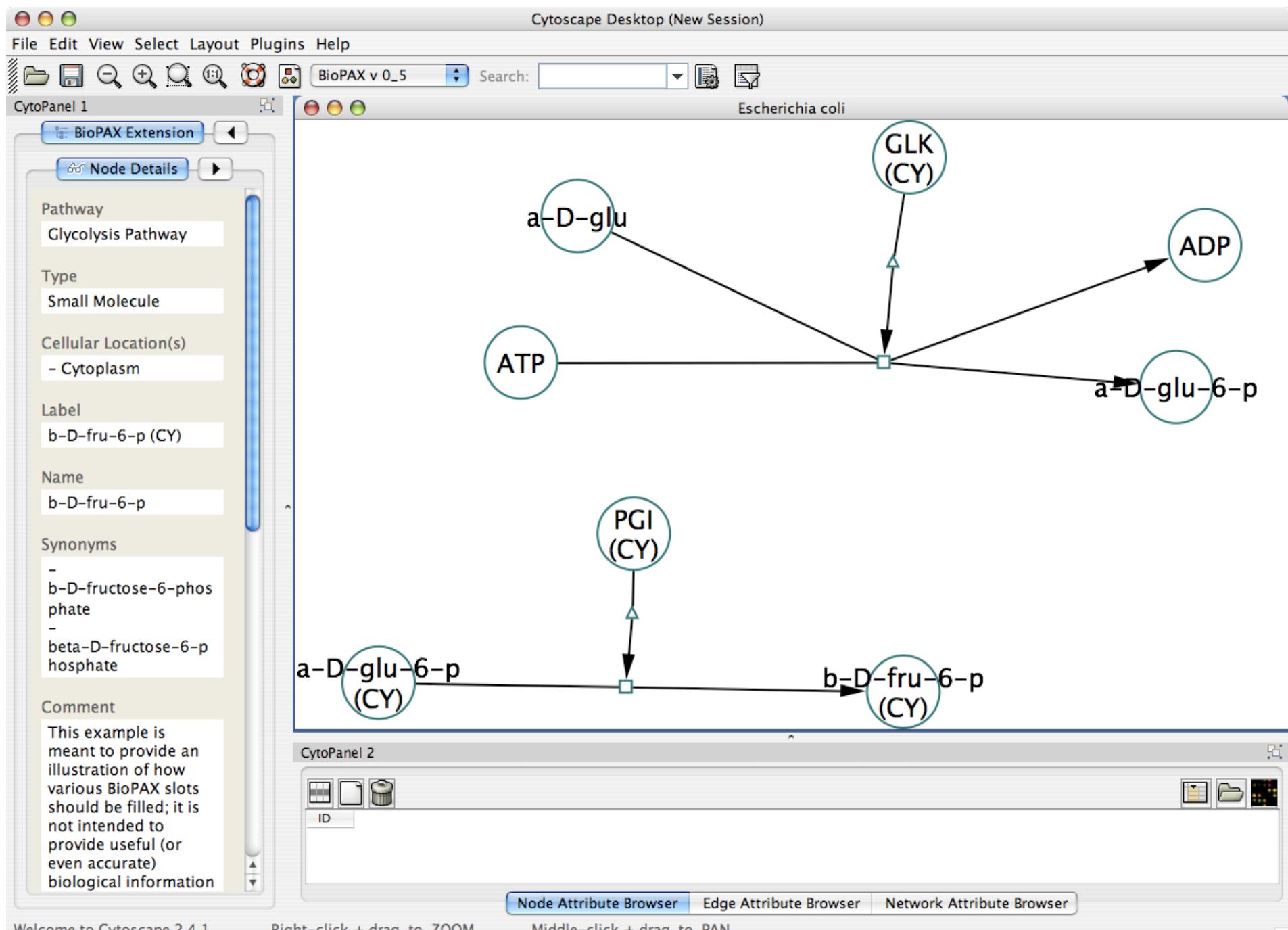
# The Cancer Cell Map



cancer.cellmap.org

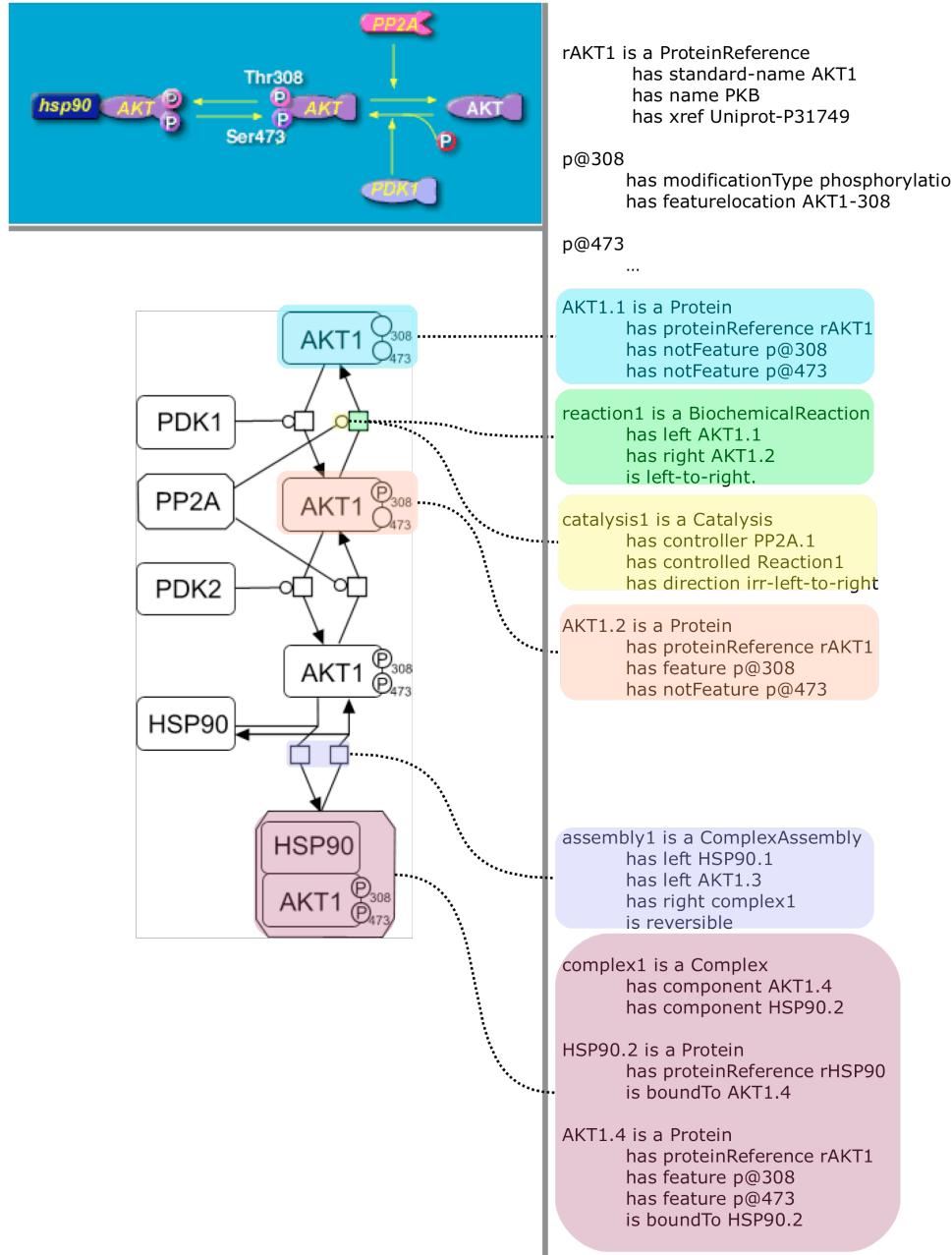
<http://visant.bu.edu/>





Ethan Cerami, MSKCC

# Systems Biology Graphical Notation



# Software Development

- PaxTools
  - Open source Java
  - Read/write BioPAX files (Level 1,2)
  - Object model in memory that can be populated and queried
  - Validation on create, read (under development by MSKCC, OHSU)
  - <http://biopax.cvs.sourceforge.net/biopax/Paxerve/>

# How to participate and contribute

- Visit [biopax.org](http://biopax.org) and join the discussion mailing list
  - [biopax-discuss@biopax.org](mailto:biopax-discuss@biopax.org)
- Make pathway data available in BioPAX
- Build software that supports BioPAX
- Contribute BioPAX worked examples, documentation and specification reviews
- Spread the word about BioPAX

# BioPAX Supporting Groups

## Current Participants

- Memorial Sloan-Kettering Cancer Center: E.Demir, M. Cary, C. Sander
- University of Toronto: G. Bader
- SRI Bioinformatics Research Group: P. Karp, S. Paley, J. Pick
- Bilkent University: U. Dogrusoz
- Université Libre de Bruxelles: C. Lemmer
- CBRC Japan: K. Fukuda
- Dana Farber Cancer Institute: J. Zucker
- Millennium: J. Rees, A. Ruttenberg
- Cold Spring Harbor/EBI: G. Wu, M. Gillespie, P. D'Eustachio, I. Vastrik, L. Stein
- BioPathways Consortium: J. Luciano, E. Neumann, A. Regev, V. Schachter
- Argonne National Laboratory: N. Maltsev, E. Marland, M. Syed
- Harvard: F. Gibbons
- AstraZeneca: E. Pichler
- BIOBASE: E. Wingender, F. Schacherer
- NCI: M. Aladjem, C. Schaefer
- Università di Milano Bicocca, Pasteur, Rennes: A. Splendiani
- Vassar College: K. Dahlquist
- Columbia: A. Rzhetsky

## Collaborating Organizations

- Proteomics Standards Initiative (PSI)
- Systems Biology Markup Language (SBML)
- CellML
- Chemical Markup Language (CML)

## Databases

- BioCyc, WIT, KEGG, BIND, PharmGKB, aMAZE, INOH, Transpath, Reactome, PATIKA, eMIM, NCI PID, CellMap

## Wouldn't be possible without

Gene Ontology  
Protégé, U.Manchester, Stanford

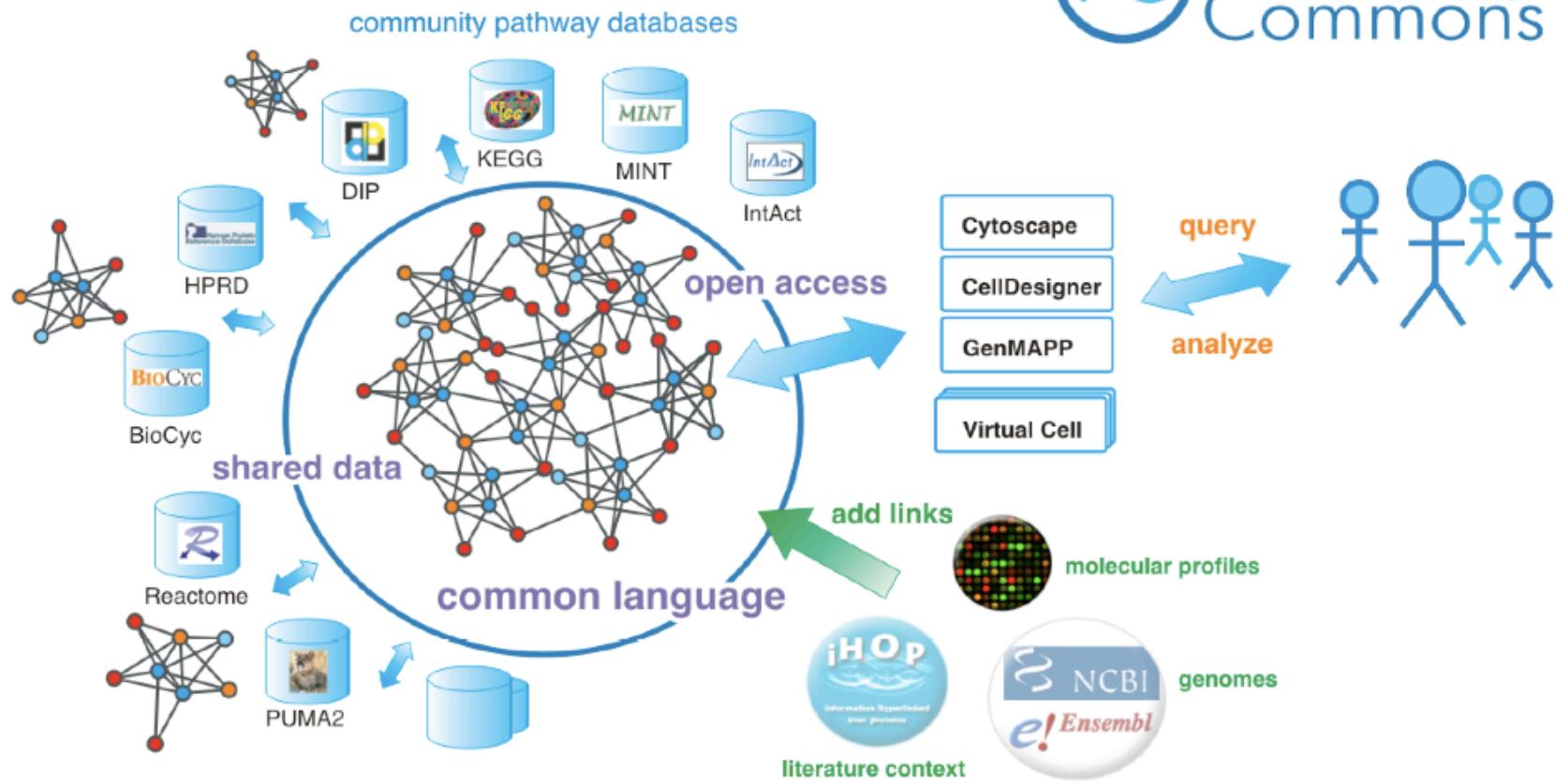
## Grants/Support

- Department of Energy (Workshop)
- caBIG



# Aim: Convenient Access to Pathway Information

<http://www.pathwaycommons.org>



Facilitate creation and communication of pathway data  
Aggregate pathway data in the public domain  
Provide easy access for pathway analysis

Long term: Converge to integrated cell map