

BioPAX

A Data Exchange Format for Biological Pathways

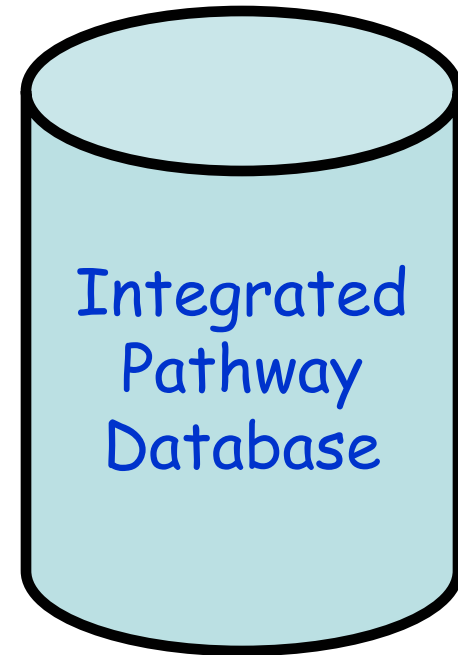
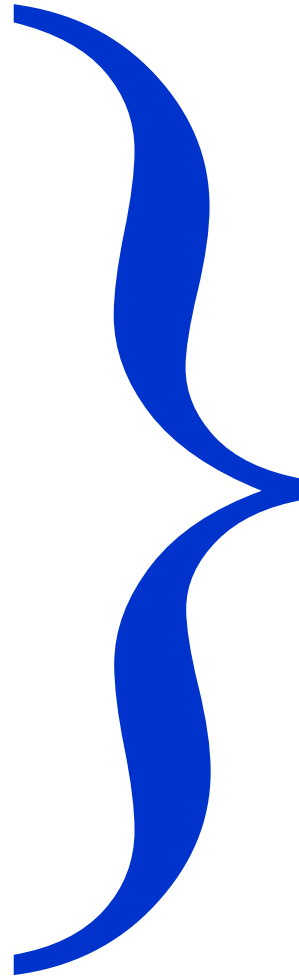
BioPAX Group
www.biopax.org

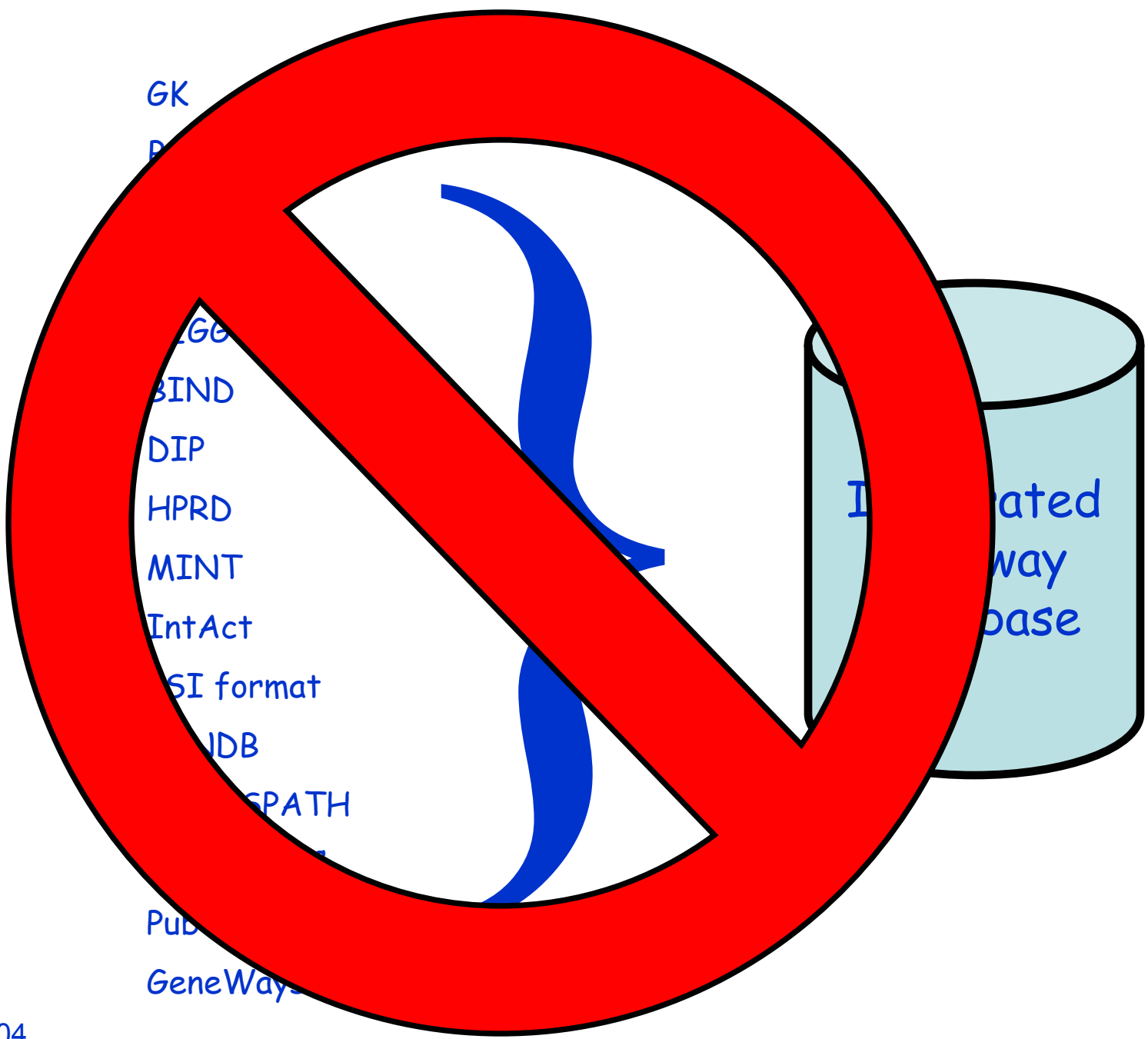
Pacific Symposium on Biocomputing
January 10, 2004
Mauna Lani, Hawaii
USA

Introduction

- BioPAX is a community-based effort conceived at ISMB '01; born at ISMB '02
- BioPAX = Biopathway Exchange Language
- Provide a consistent data exchange format to make it easier for database users (e.g. tool developers, DB curators, researchers) to integrate of pathway data from multiple sources:
 - Metabolic pathways
 - Signal transduction
 - Protein-protein interactions
 - Gene regulation

GK
BioCyc
WIT
aMAZE
KEGG
BIND
DIP
HPRD
MINT
IntAct
PSI format
CSNDB
TRANSPATH
TRANSFAC
PubGene
GeneWays





GK

BIND

DIP

HPRD

MINT

IntAct

PSI format

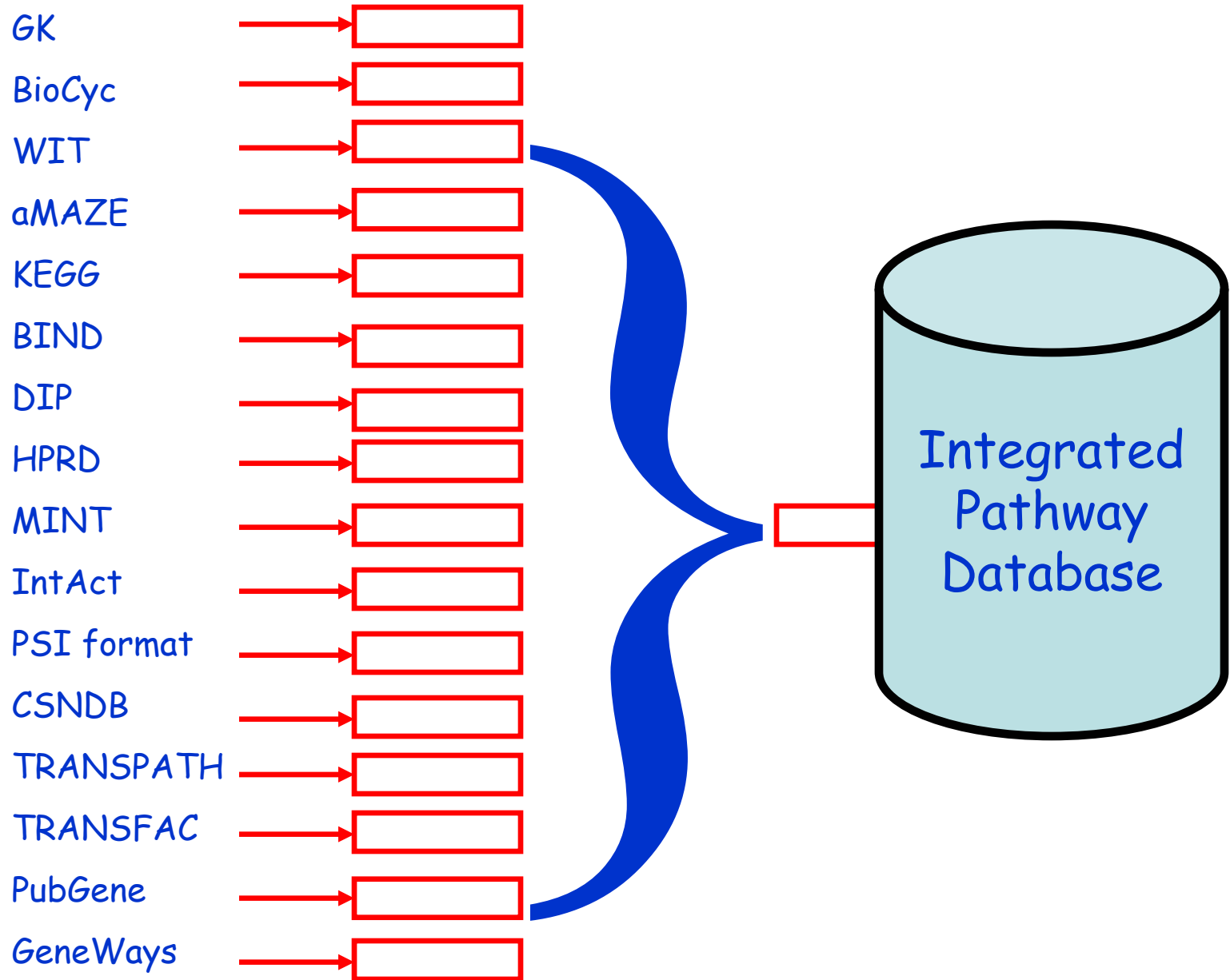
KEGG

SPATH

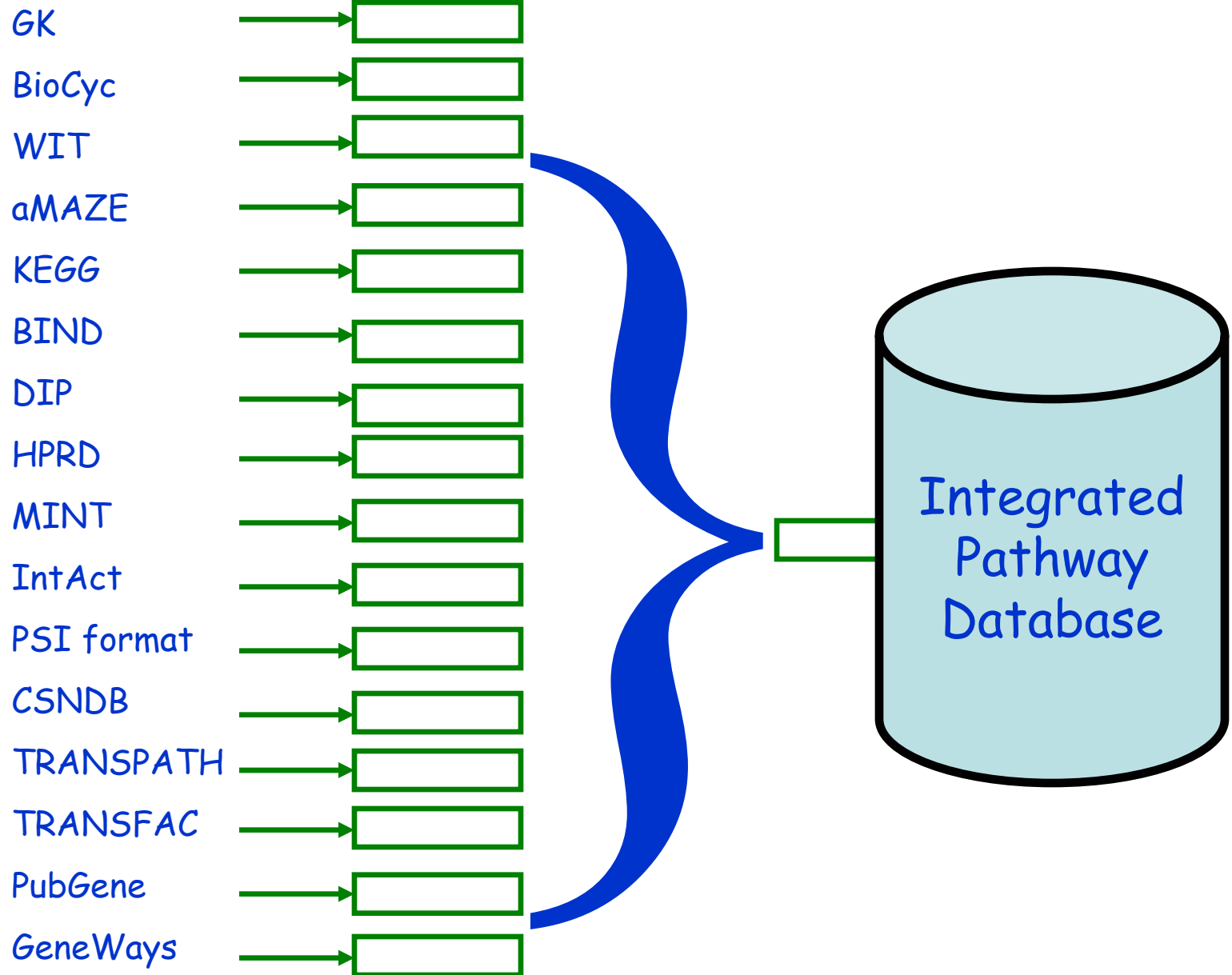
Pub

GeneWays

Integrated pathway database



BioPAX



Practical: Use Cases

- Joint learning through multiple types of data

	Metabolic	PPI	ST	GR
Met	Met-Met	PPI-Met	ST-Met	GR-Met
Met, PPI	Met-PPI-Met	Met-PPI-PPI	Met-ST-PPI	Met-PPI-GR
Met, PPI, ST	Met-ST	PPI-ST	ST-ST	GR-ST
Met, PPI, ST, GR	Met-GR	PPI-GR	ST-GR	GR-GR

It is powerful to have to have all these data in the same format when you want to integrate them

- Build a centralized public pathway DB
- Share data between existing DBs
- Distribute proprietary data from a commercial enterprise

Level 1

Physical Entities	Small molecules Proteins RNA Complexes
Interactions	Biochemical Reactions Enzyme Catalysis Transport Catalysis Complex Assembly
Data Source Compatibility	GenomeKnowledgebase (GK) BioCyc WIT aMAZE KEGG

Level 2

Physical Entities	(same as Level 1)
Interactions	Binding Interactions

Data Source Compatibility	BIND DIP HPRD MINT IntAct PSI format
------------------------------	---

Level 3

Physical Entities	Gene DNA
Interactions	Signaling Pathways Genetic interactions Gene regulation
Data Source Compatibility	CSNDB TRANSPATH TRANSFAC

Future Levels

Physical Entities	Environmental effects Cells Cell-compartments Photons
Interactions	Abstract set relationships - co-occurrence in: - a pathway - literature abstracts - a cell compartment - similar function - etc.
Data Source Compatibility	PubGene GeneWays

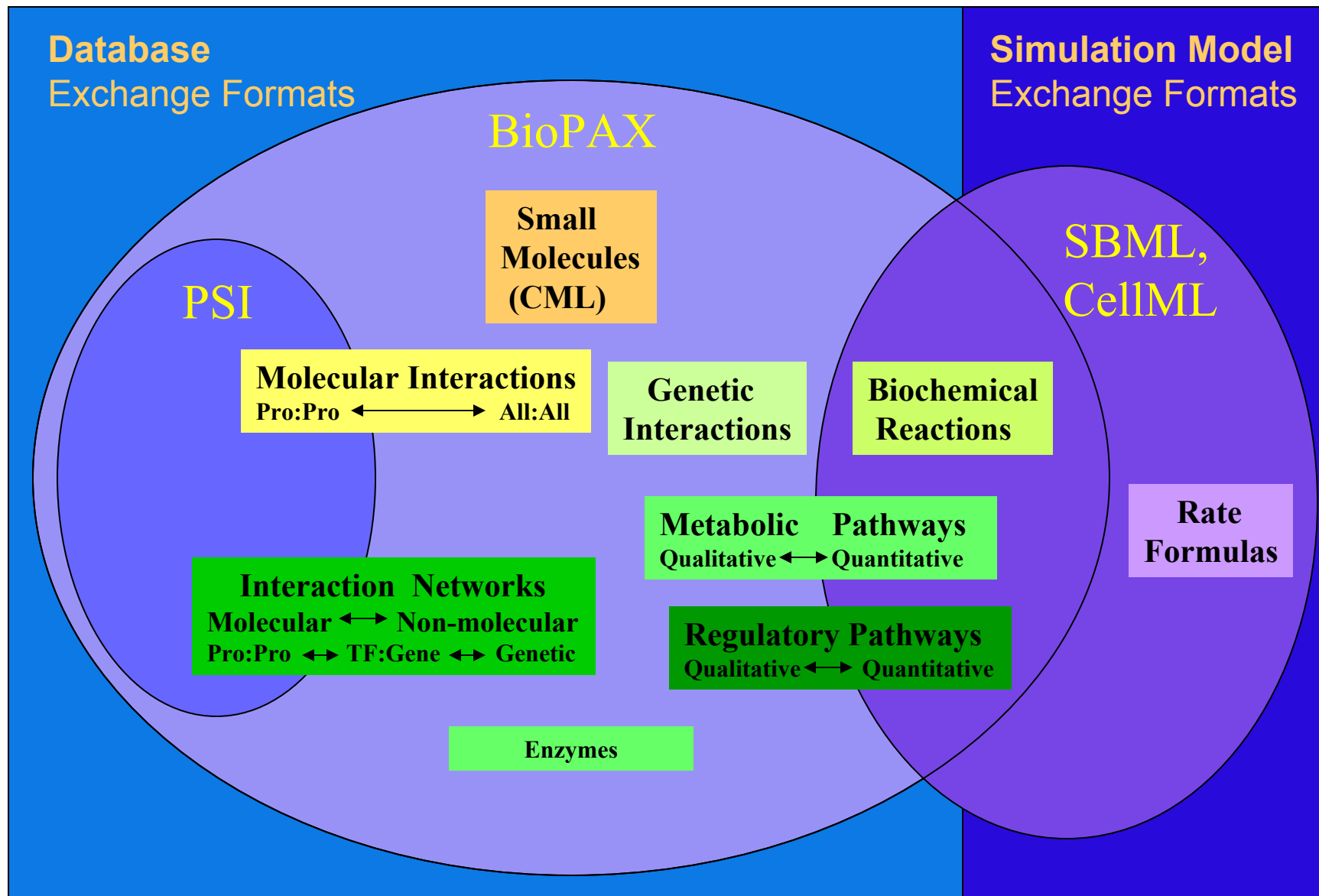
Current Status

- ✓ Initial meeting Nov. 2002
- ✓ Version 0.5 of Level 1 released Sep. 2003
- Translating records from major DBs to BioPAX

Milestones		
Point People	Target Date	Task Description
Karp	February 1, 2004	GKB OWL output capability. Investigate which OWL version, how to validate the OWL output, etc.
Shah / Bader	February 15, 2004	Specification document for BioPAX -- similar to W3C style
Bader	March 15, 2004	Complete v1.0 of BioPAX ontology based on feedback from example work
Maltsev	March 15, 2004	Write converter from WIT to BioPAX
Karp	March 15, 2003	Write converter from BioCyc to BioPAX
Shah	April 1, 2004	Writer converter from BioPAX to their software system to visualize a BioPAX pathway
Karp	April 1, 2004	Write converter from BioPAX to BioCyc (tentative)
Bader	April 1, 2004	Writer converter from BioPAX into CPATH (tentative)
Shah / Bader	May 1, 2004	Specification document for BioPAX -- similar to W3C style
	May 1, 2004	Release BioPAX

Release BioPAX May 2004

Exchange Formats in the Pathway Data Space



Design Goals

- **Encapsulation:** An entire pathway in one record
- **Compatible:** Use existing standards wherever possible
- **Computable:** From file reading to logical inference
 - OWL (Ontology Web Language)
 - Fast
 - Complete: all conclusions are guaranteed to be computed
 - Decidable: all computations will finish in finite time (with OWL Lite, short amount of time.

Requirements Specification

- Accommodate existing database representations: BioCyc, BIND, WIT, aMAZE, KEGG, etc.
 - Compatible as a superset of representations
- Support different pathway types:
 - Metabolic pathways
 - Signaling pathways
 - Protein-protein interactions
 - Gene regulatory pathways
- OWL- suppose data rep in BioPAX, if has XML schema can validate it as valid XML document

Implementation of BioPAX

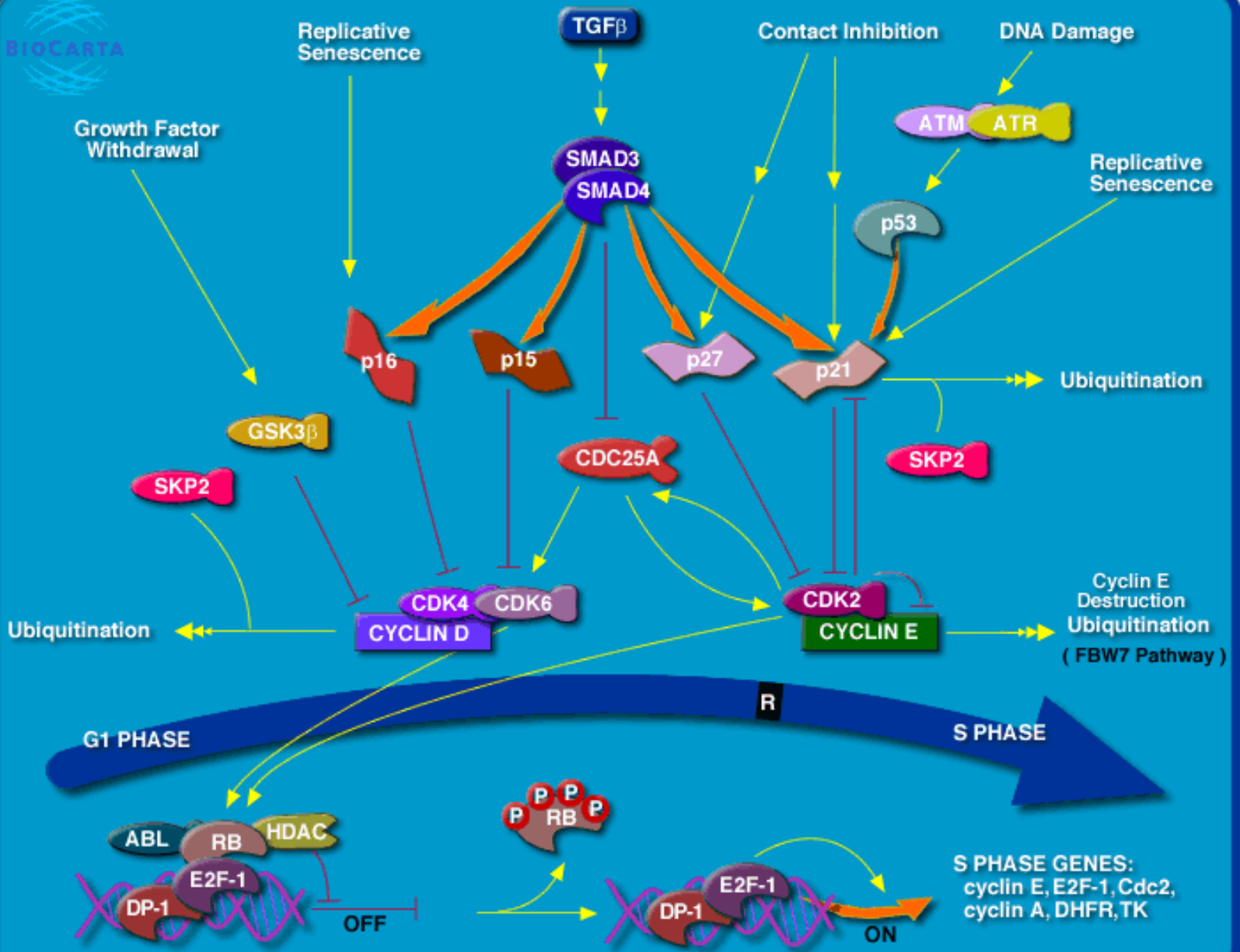
- Implemented using OWL language
- OWL is
 - Ontology Web Language
 - XML based
 - W3C standard www.W3C.org
- Example of a BioPAX Class and Instance in OWL

Example - Class def in OWL

```
<owl:Class rdf:ID="protein">
  <rdfs:subClassOf>
    <owl:Class rdf:about="#physicalEntity"/>
  </rdfs:subClassOf>
  <rdfs:comment
    rdf:datatype="http://www.w3.org/2001/XMLSchema#string">
    A protein (e.g. The EGFR protein sequence.  See Swiss-Prot
    for more examples.)
  </rdfs:comment>
</owl:Class>
```

Example - Instance in OWL

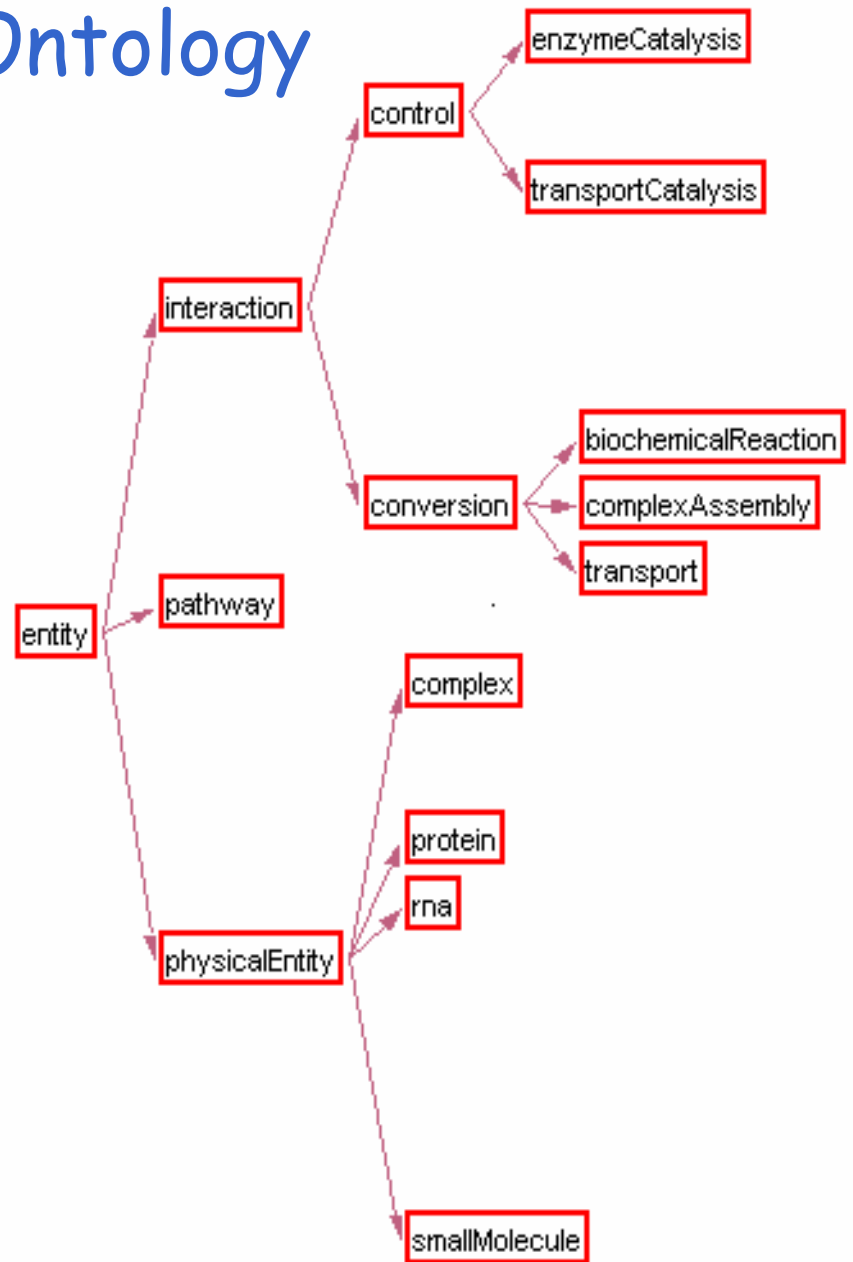
```
<bpX:protein rdf:ID="biopax-L1v0.5_Instance_42">
  <bpX:NAMES>
    <bpX:namesType rdf:ID="biopax-L1v0.5_Instance_43">
      <bpX:SHORTLABEL>phosphoglucose isomerase</bpX:SHORTLABEL>
    </bpX:namesType>
  </bpX:NAMES>
</bpX:protein>
```

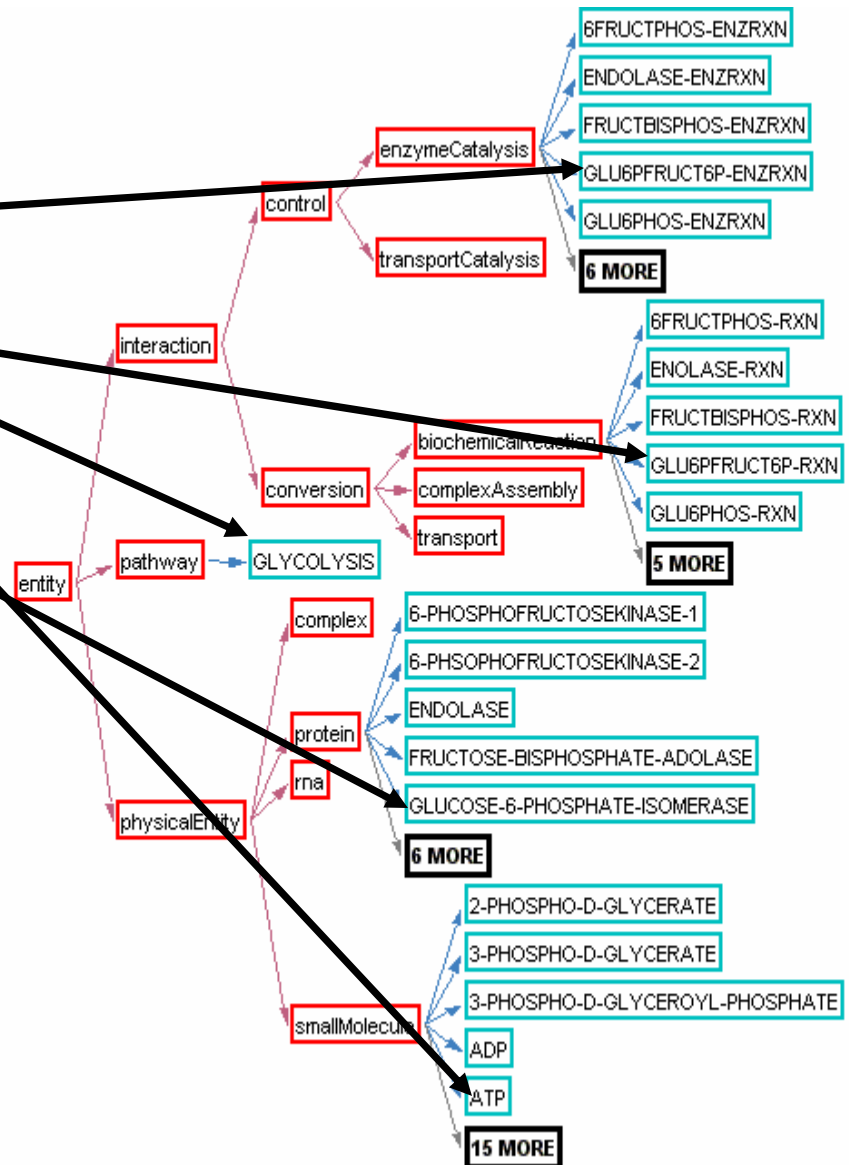
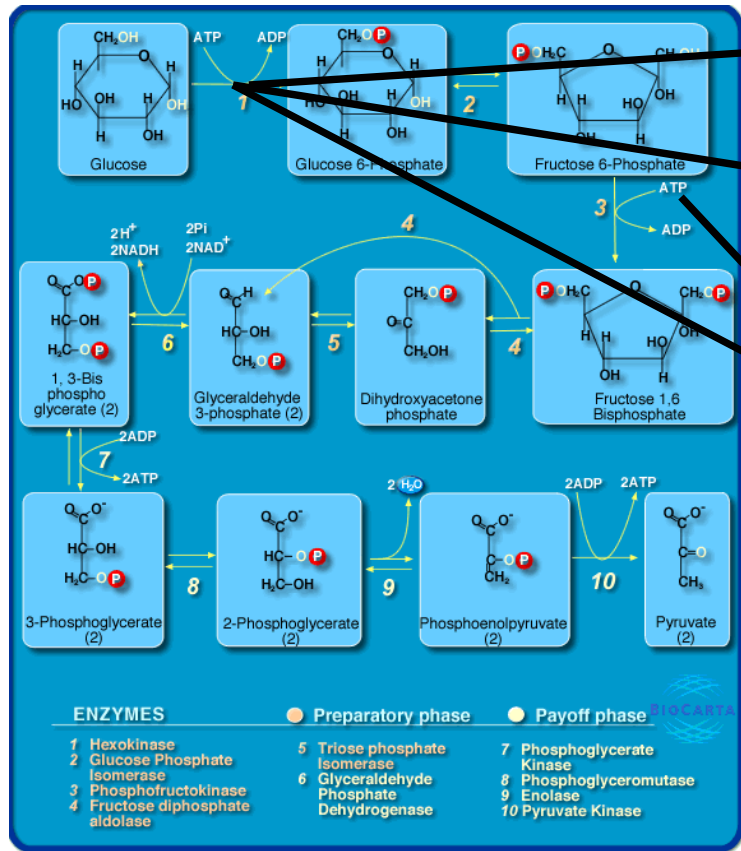


BioPAX Ontology

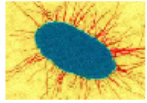
Current structure of
class hierarchy

Level 1 v0.9 (Dec. 2003)





Representing Metabolic Data in BioPAX

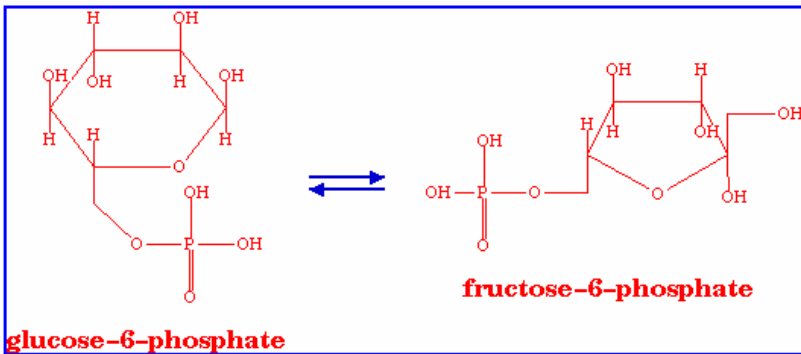


E. coli Reaction: 5.3.1.9

Superclasses: [5.3.1 -- INTERCONVERTING ALDOSES AND KETOSES](#)

[phosphoglucose isomerase](#): [pgi](#)

In pathway: [glycolysis](#), [gluconeogenesis](#)



ΔG° (kcal/mol): 0.4 [[1](#)]

Gene-Reaction Schematic: [?](#)



Unification Links: [ENZYME:5.3.1.9](#)



Biochemical Reaction	
ID	1
Full Name	Glucose-6-p to fructose-6-p
Left	<cml>glucose-6-phosphate</cml>
Right	<cml>fructose-6-phosphate</cml>
Delta G	0.4 kcal/mole
EC	5.3.1.9

EcoCyc: Reaction

1/10/2004

BioPAX: Biochemical Reaction

22

Representing Metabolic Data in BioPAX (cont 1)

Enzymatic reaction of: phosphoglucose isomerase

Synonyms: glucose-6-phosphate isomerase , D-glucose-6-phosphate-ketol-isomerase

[glucose-6-phosphate](#) \rightleftharpoons [fructose-6-phosphate](#)

The reaction direction shown, that is, $A + B \rightleftharpoons C + D$ versus $C + D \rightleftharpoons A + B$, is in accordance with the direction of the reaction within a pathway.

In pathways: [gluconeogenesis](#) , [glycolysis](#)

Comment: 2-deoxyglucose-6-p is a known inhibitor in mammalian systems. E.coli cells with mutated *pgi* gene apparently utilize glucose primarily by the pentose phosphate pathway and to a lesser extent by the Entner-Duodoroff pathway. [3]

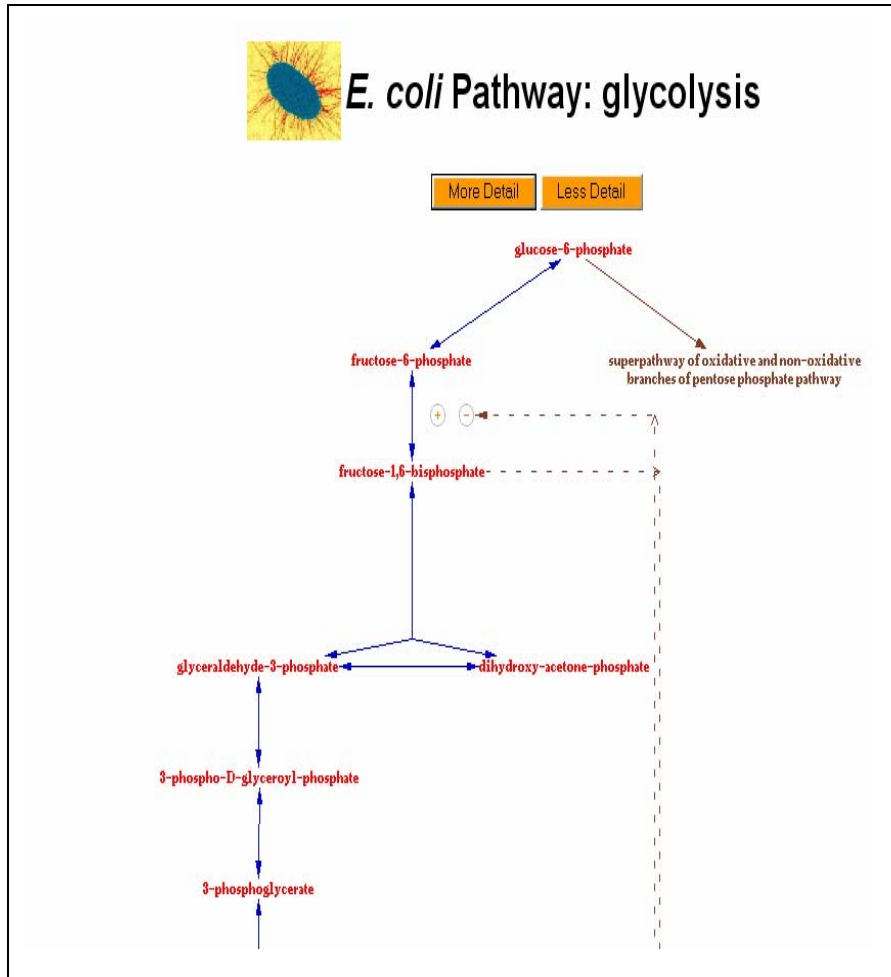
Citations: [3 , 4]

Inhibitors (neither competitive nor allosteric) [5]:



Catalysis	
ID	2
Name	Catalysis of glucose-6-p to fructose-6-p
Enzyme	glucose-6-phosphate isomerase
Reaction	BioPAX ID=1
Inhibitors	Low pH

Representing Metabolic Data in BioPAX (cont 2)



Pathway	
ID	10
Name	Glycolysis
Interactions	1. BioPAX ID=2 2. BioPAX ID=4 3. BioPAX ID=6 etc.

EcoCyc: Pathway

BioPAX Class: Pathway

Representing Signal Transduction in BioPAX

Cell_Signaling : MAP-kinase -> NF-kB

Entry Takako
From_molecule [MAP-kinase](#)
To_molecule [NF-kB](#)
Effect activation
Reference [\[Malinin_1997\]](#)
Role [apoptosis](#)

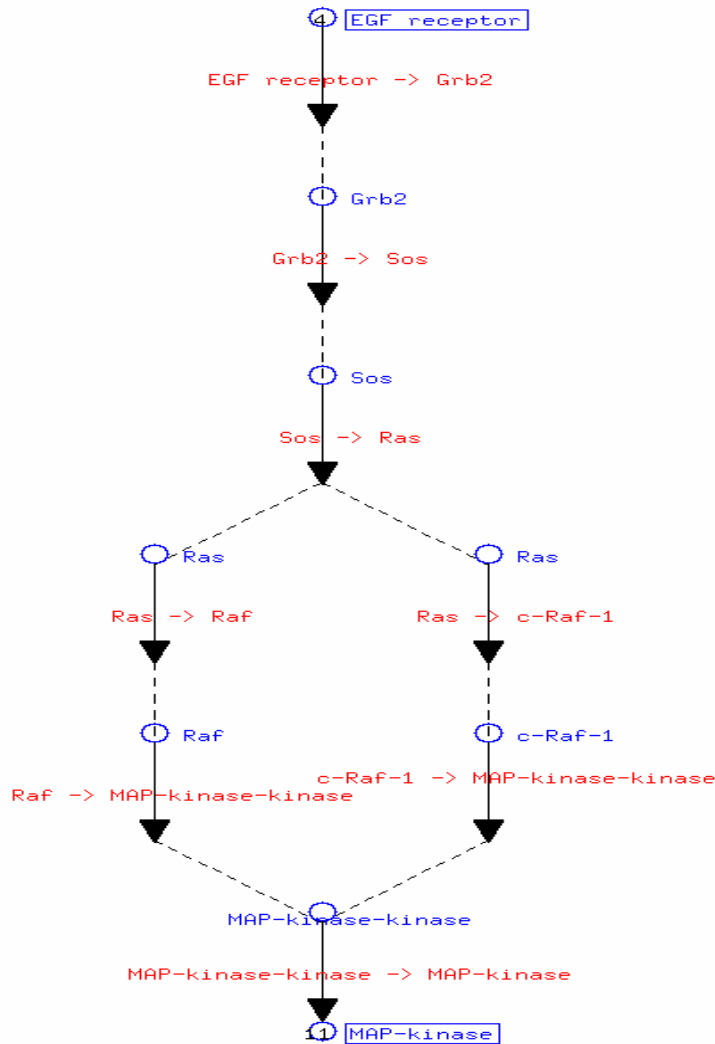
CSNDB Signaling Pathway Step



Reaction	
ID	20
Name	Activation of NF-kB
Substrate	NF-kB (inactive)
Product	NF-kB (active)

Enzyme Catalysis	
ID	21
Name	MAP-kinase activates NF-kB
Enzyme	MAP-kinase
Reaction	BioPAX ID=20

Representing Signal Transduction in BioPAX



Pathway	
ID	10
Name	MAPK
Interactions	1. BioPAX ID=21 2. BioPAX ID=23 3. BioPAX ID=25 etc.

CSNDB Pathway

Organizational Structure

- Small core group advancing standard
- Increased representation from mailing lists
- Bi-weekly conference calls, bi-monthly F2F
- Cost paid by participants with some support from DOE
- Special interests have subgroups
 - Core group member + outside experts
 - Tackle specific challenges

BioPAX Subgroups

- Created for multiple purposes:
 - Tackling specific conceptual problems
 - Developing spin-off projects
 - Small Molecule Database
 - Database of Pathway Resources
 - Gathering specific resources for core group
- Typically consist of:
 - Core group members (1-3)
 - Experts from external community (1-2)

How to Contribute

- Participate in email list discussions
 - sign up via web site:
<http://www.biopax.org>
- Participate in meetings and subgroups
- Make your data available in BioPAX format, when complete
- Promote BioPAX to colleagues

BioPAX Support

Groups

- Memorial Sloan-Kettering Cancer Center: C. Sander, J. Luciano, M. Cary, G. Bader
- SRI Bioinformatics Research Group: P. Karp, S. Paley, J. Pick
- University of Colorado Health Sciences Center: I. Shah
- Harvard Medical School: Aviv Regev
- BioPathways Consortium: J. Luciano, E. Neumann, V. Schachter
- Argonne National Laboratory: N. Maltsev
- Samuel Lunenfeld Research Institute: C. Hogue

Organizations:

- Proteomics Standards Initiative (PSI) (psidev.sf.net)
- Systems Biology Markup Language (SBML)
- CellML
- Chemical Markup Language (CML)

Databases

- GK (Genome Knowledge Base)
- BioCyc (www.biocyc.org)
- BIND (www.bind.ca)
- WIT (wit.mcs.anl.gov/WIT2)
- KEGG (www.genome.ad.jp/kegg)
- aMAZE

Grants

- Department of Energy

