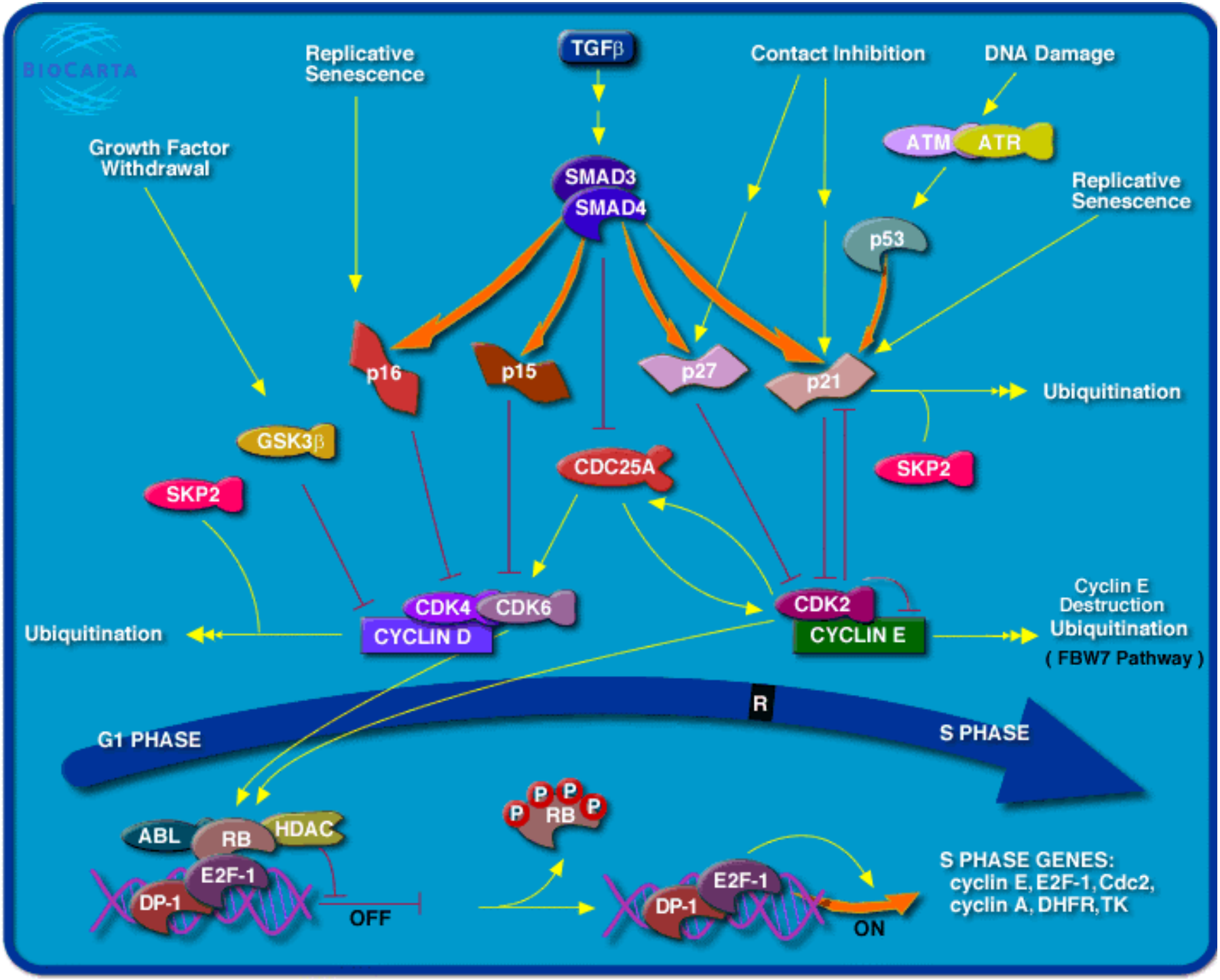
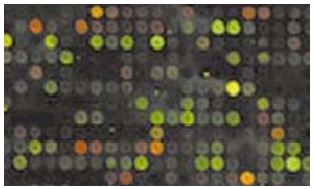
BioPAX
A Data Exchange Ontology for Biological Pathways

BioPAX Workgroup

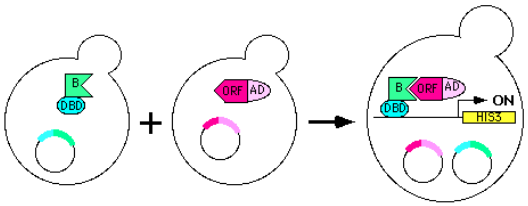
www.biopax.org


BIOCARTA
Replicative Senescence
TGFβ
Contact Inhibition
DNA Damage
Growth Factor Withdrawal
ATM
ATR
SMAD3
SMAD4
Replicative Senescence
p53
p16
p15
p27
p21
Ubiquitination
GSK3β
CDC25A
SKP2
SKP2
CDK4
CDK6
CYCLIN D
CDK2
CYCLIN E
Ubiquitination
Cyclin E Destruction
Ubiquitination
(FBW7 Pathway)
R
G1 PHASE
S PHASE
ABL
RB
HDAC
E2F-1
DP-1
OFF
P
RB
E2F-1
DP-1
ON
S PHASE GENES:
cyclin E, E2F-1, Cdc2,
cyclin A, DHFR, TK

High Throughput Experimental Methods

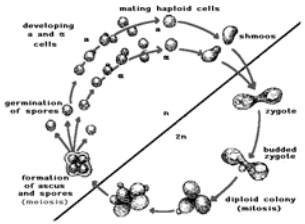


Microarray



Two-Hybrid

Mass Spectrometry



Genetics

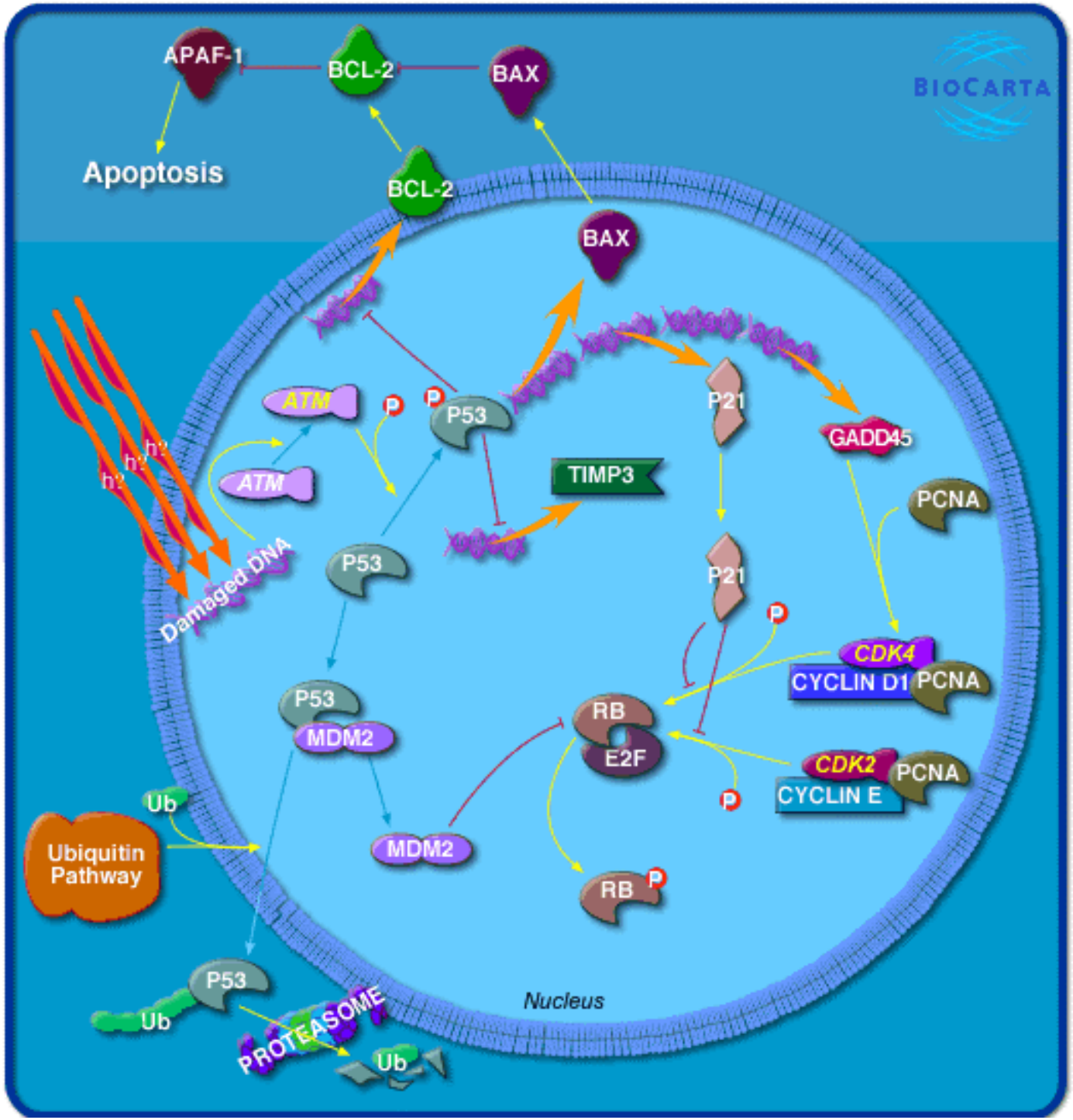
Expression, Interaction Data, Function, Protein modifications

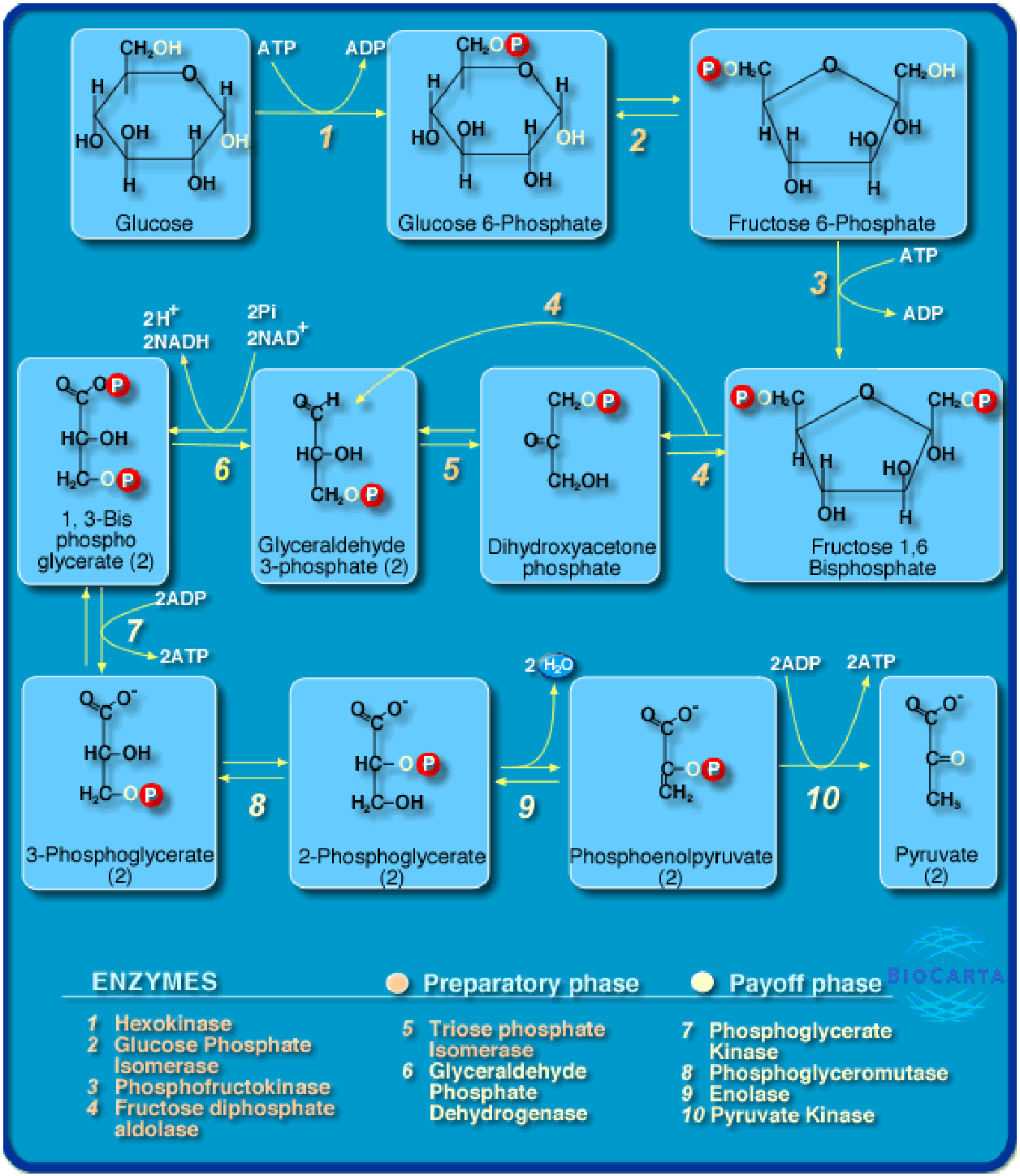
Existing Literature

PubMed

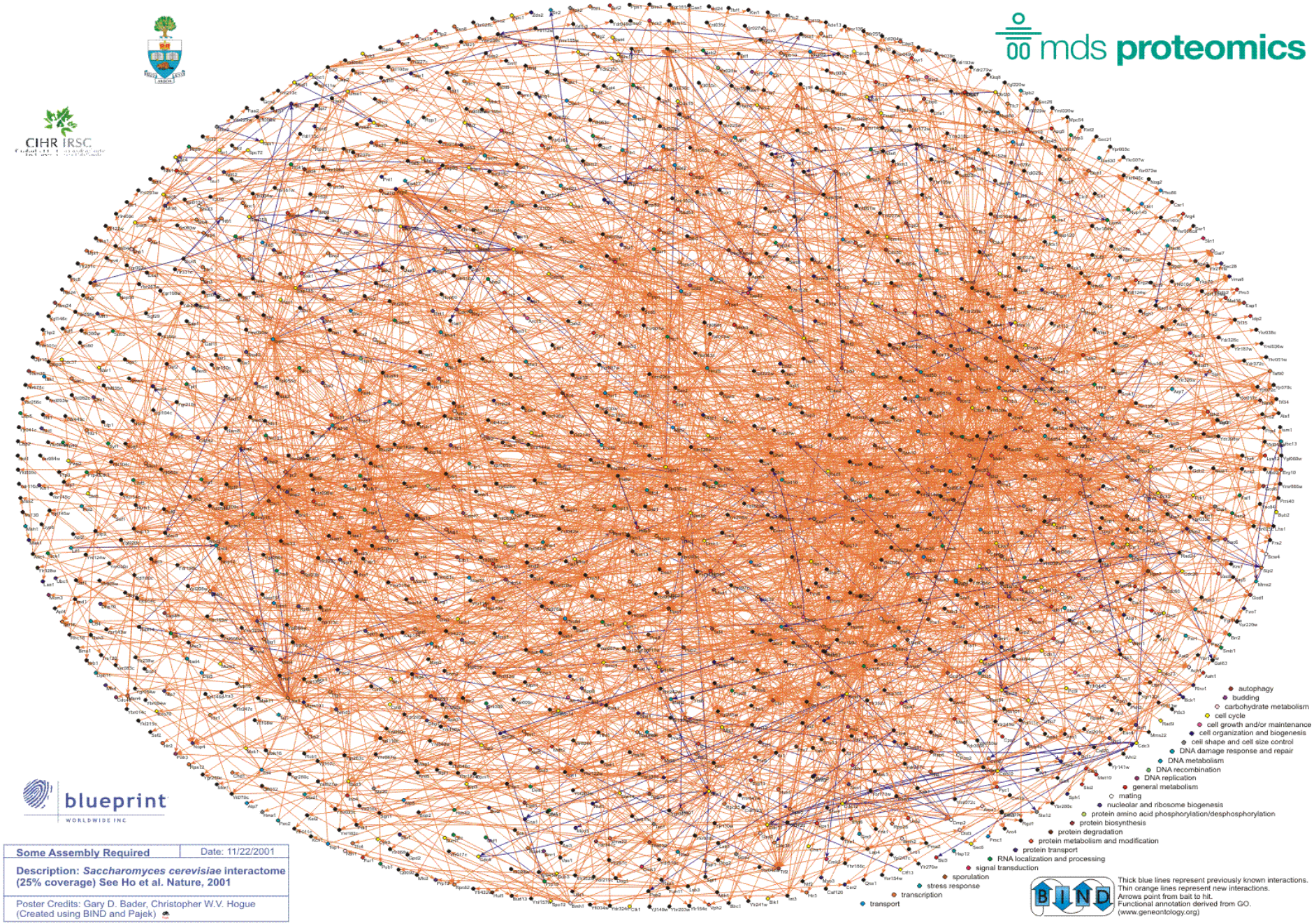
Multiple Pathway Databases

Integration Nightmare!


APAF-1
BCL-2
BAX
BIOCARTA
Apoptosis
BCL-2
BAX
ATM
P
P
P53
P21
GADD45
ATM
TIMP3
PCNA
h?
h?
h?
Damaged DNA
P53
P21
P
CDK4
CYCLIN D1
PCNA
P53
MDM2
RB
E2F
CDK2
PCNA
CYCLIN E
P
Ub
Ubiquitin Pathway
MDM2
RB
P
P53
Ub
PROTEASOME
Ub
Nucleus


Glucose
ATP
ADP
1
Glucose 6-Phosphate
2
Fructose 6-Phosphate
3
ATP
ADP
4
2H+
2NADH
2Pi
2NAD+
1, 3-Bis phospho glycerate (2)
6
Glyceraldehyde 3-phosphate (2)
5
Dihydroxyacetone phosphate
4
Fructose 1,6 Bisphosphate
2ADP
7
2ATP
3-Phosphoglycerate (2)
8
2-Phosphoglycerate (2)
2 H2O
9
Phosphoenolpyruvate (2)
2ADP
2ATP
10
Pyruvate (2)
ENZYMES
Preparatory phase
Payoff phase
BIOCARTA
1 Hexokinase
2 Glucose Phosphate Isomerase
3 Phosphofructokinase
4 Fructose diphosphate aldolase
5 Triose phosphate Isomerase
6 Glyceraldehyde Phosphate Dehydrogenase
7 Phosphoglycerate Kinase
8 Phosphoglyceromutase
9 Enolase
10 Pyruvate Kinase

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



Thick blue lines represent previously known interactions.
Thin orange lines represent new interactions.
Arrows point from bait to hit.
Functional annotation derived from GO.
(www.geneontology.org)

Some Assembly Required	Date: 11/22/2001
Description: ***Saccharomyces cerevisiae*** **interactome (25% coverage) See Ho et al. Nature, 2001**	
Poster Credits: Gary D. Bader, Christopher W.V. Hogue (Created using BIND and Pajek)	

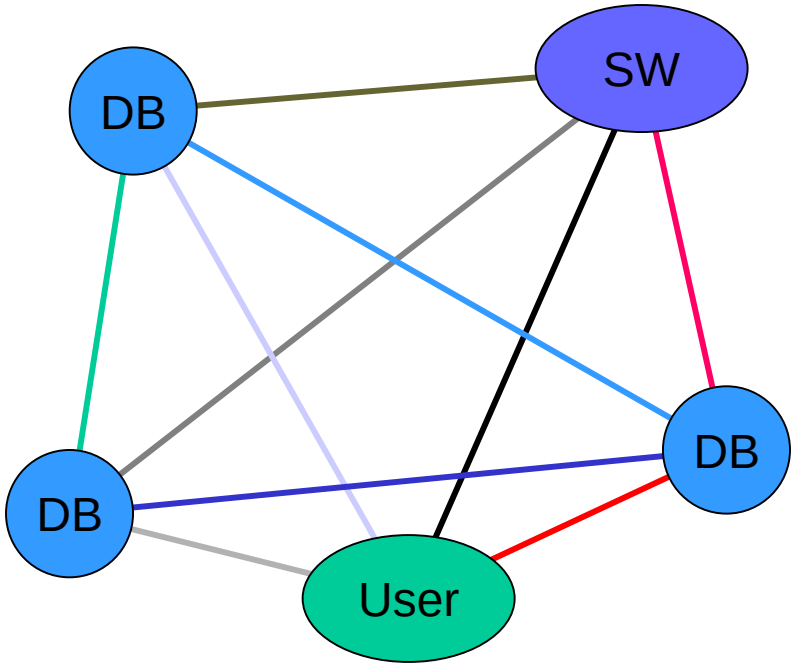
BioPAX

- BioPAX = Biological Pathway Exchange
- Data exchange format for pathway data
- Include support for these pathway types:
  - Metabolic pathways
  - Signaling pathways
  - Protein-protein interactions
  - Genetic regulatory pathways
- Accommodate representations used in existing databases such as BioCyc, BIND, WIT, aMAZE, KEGG, etc.

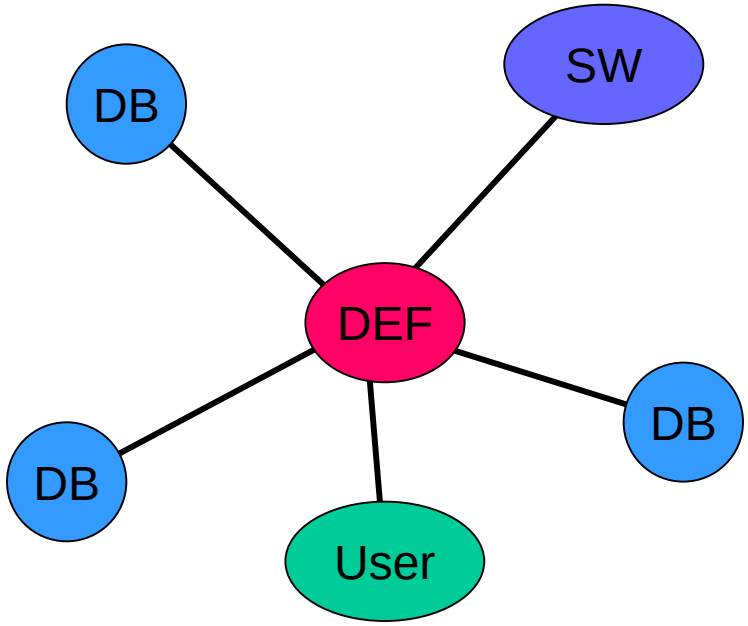
BioPAX Motivation

Common good

Before DEF

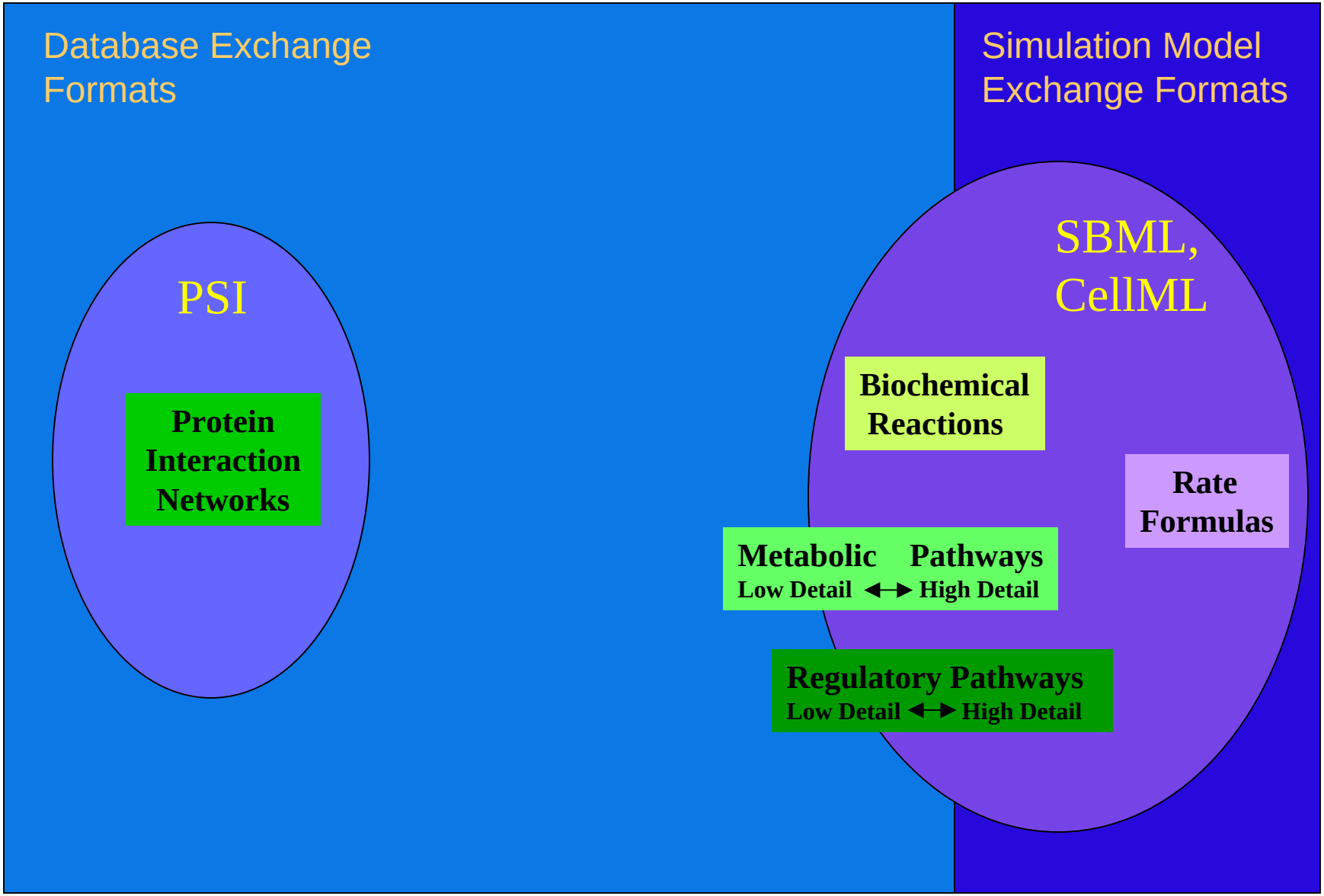


After DEF

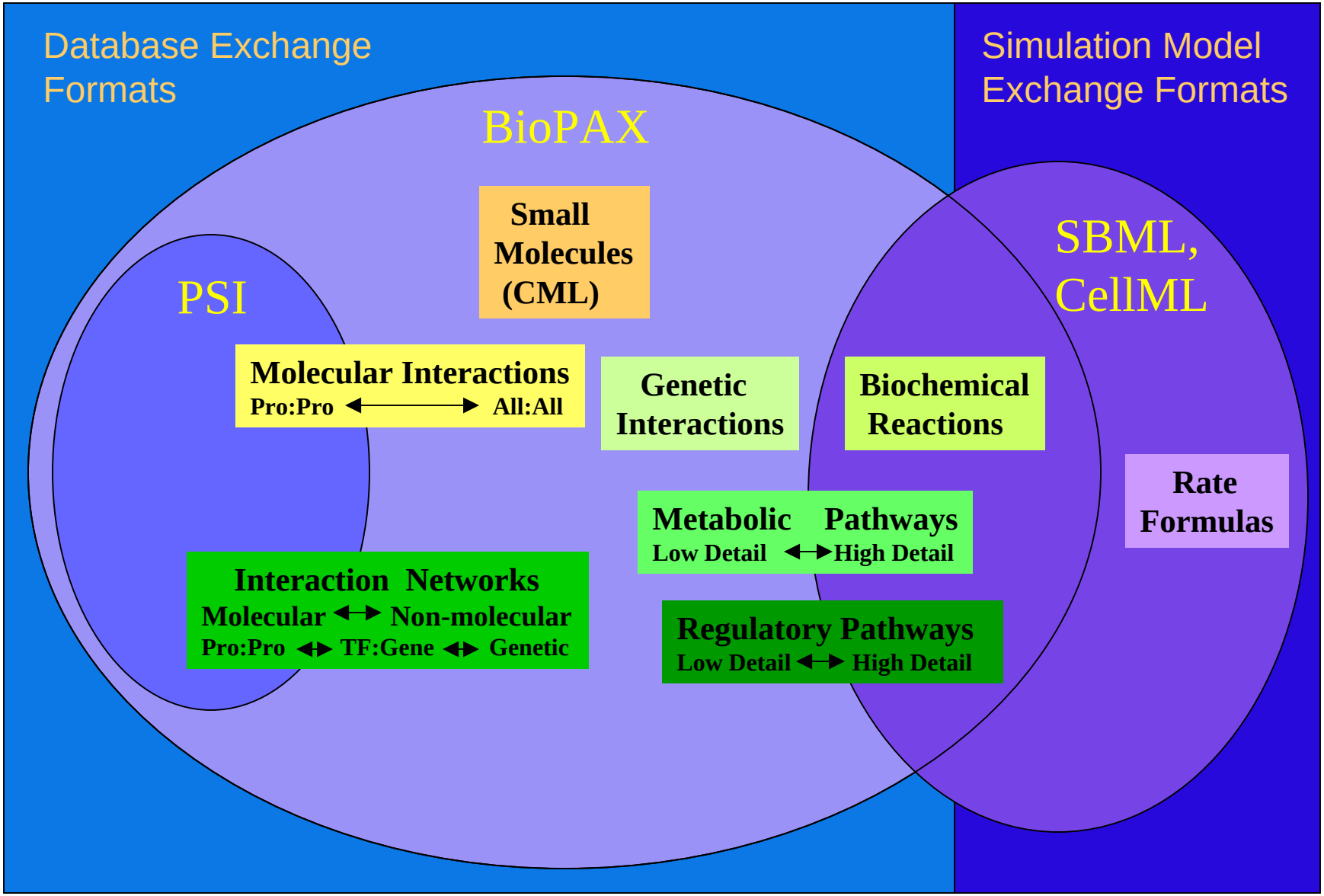


Promotes collaboration (big science), accessibility

Exchange Formats in the Pathway Data Space



Exchange Formats in the Pathway Data Space



Desired Features

- **Extensible**: Specific classes of data in BioPAX have been marked as extensible to allow addition of new types of data in the future
- **Encapsulation**: An entire pathway can be encapsulated in a single BioPAX record
- **Compatible**: BioPAX will try to use existing standards for encoding biological pathway related information wherever possible
- **Flexible**: Different preferred representations of pathway data can be described using BioPAX
- **Computable**: From file reading to logical inference

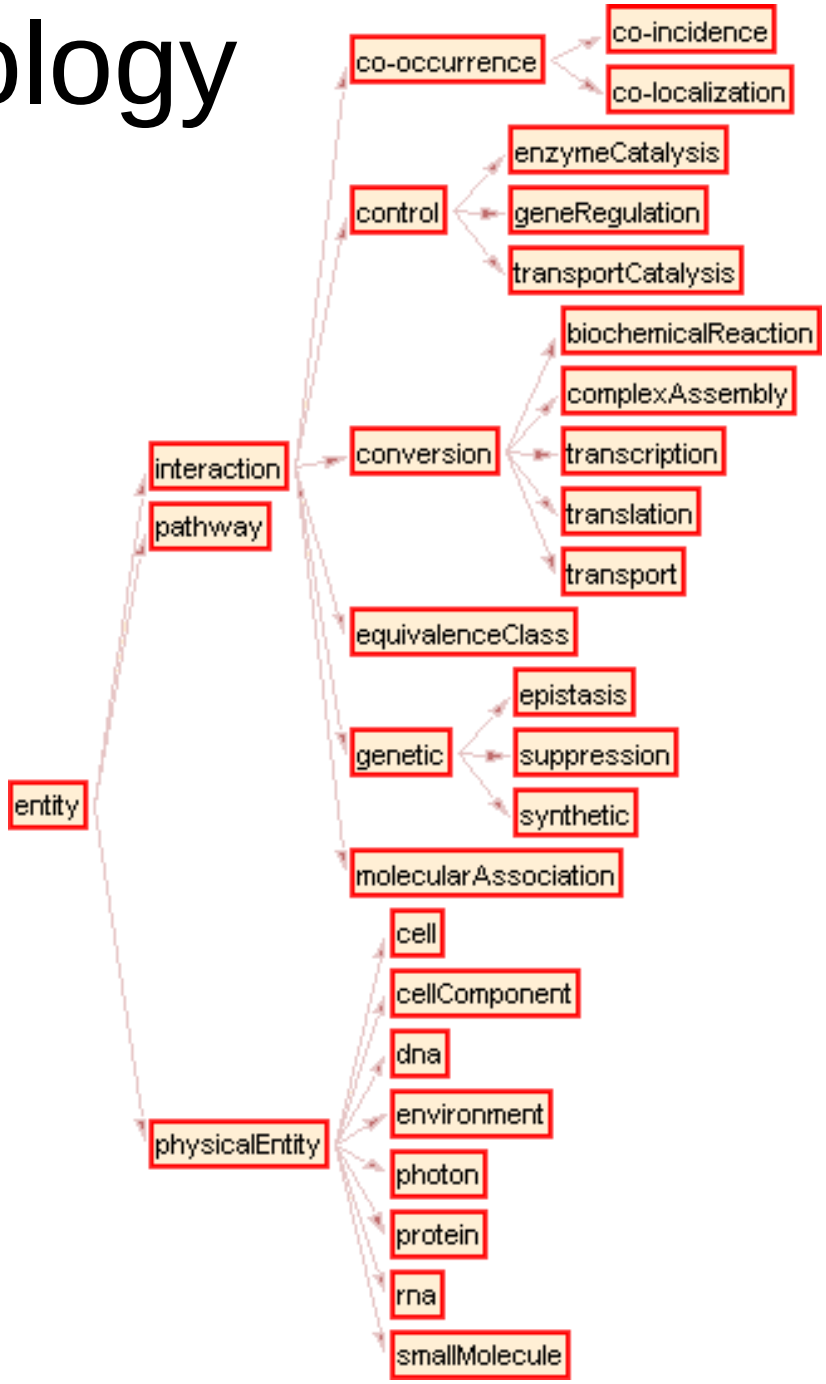
Ontology Development Process

- Developed Conceptual Framework
  - Based upon existing DB schemas
    - aMAZE, BIND, EcoCyc, WIT, KEGG, etc.
  - Allows wide range of detail, multiple levels of abstraction
- Constructed an ontology
  - GKB, Protégé → OWL
  - Practical leveled approach (~Postscript, SBML)
  - Discussion items tracked using SourceForge
  - Support OWL and XML Schema

BioPAX Ontology

- Released Level 1 v0.5 on Sep.24.2003

- Released structure of class hierarchy:

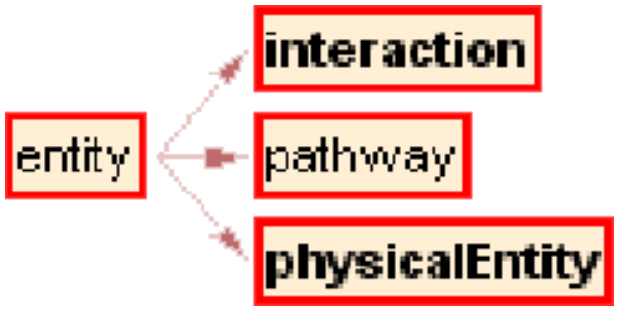


BioPAX Ontology : Root

- Root class: Entity
  - Any concept that we will refer to as a discrete unit when describing the biology of pathways.
  - Does not include metadata
    - E.g. “DB source”, “PubMed ID”, “Experimental technique”, etc.

BioPAX Ontology : Root

- Entity Subclass: Physical Entity
  - A building block of simple interactions
  - E.g. Small molecules, Proteins, DNA, RNA
- Entity Subclass: Interaction
  - A set of entities and some relationship between them
  - E.g. Reactions, Molecular Associations, Catalyses
- Entity Subclass: Pathway
  - A set of interactions
  - E.g. Glycolysis, MAPK, Apoptosis

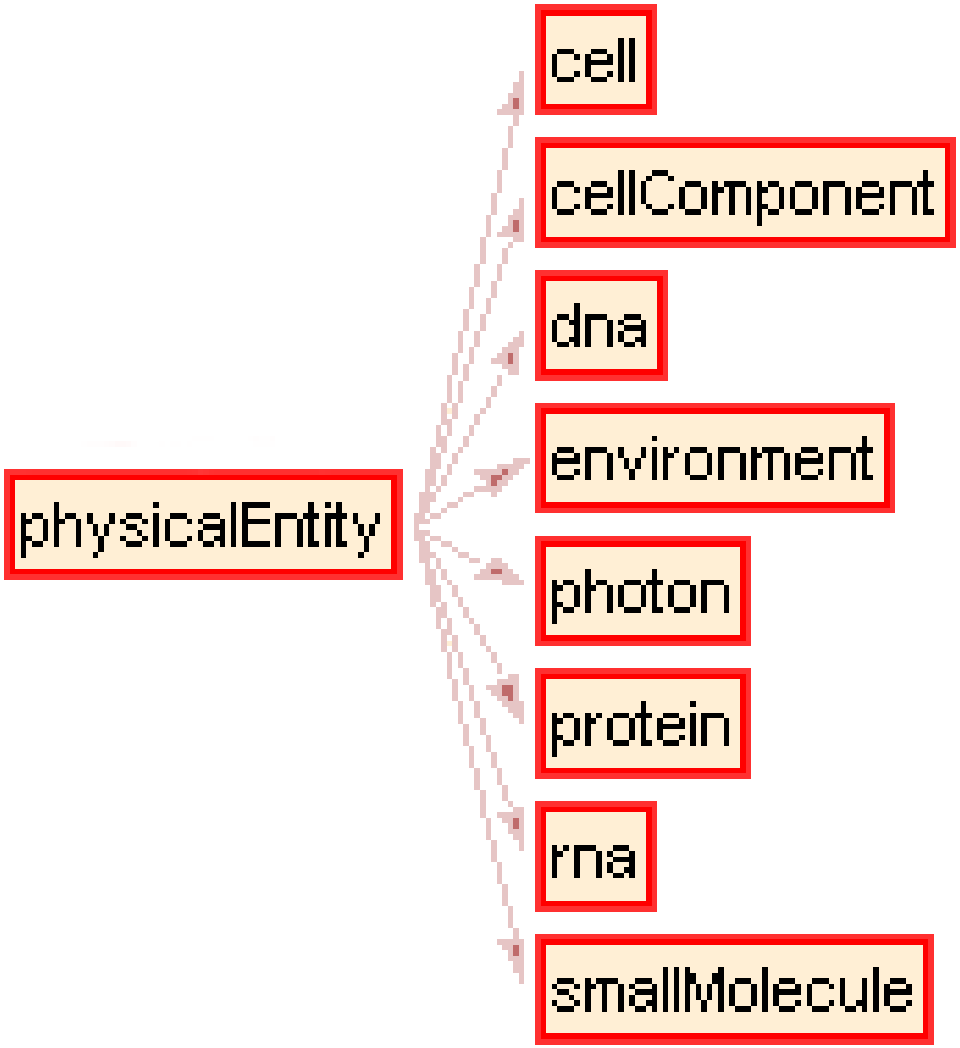


Subclass

Container

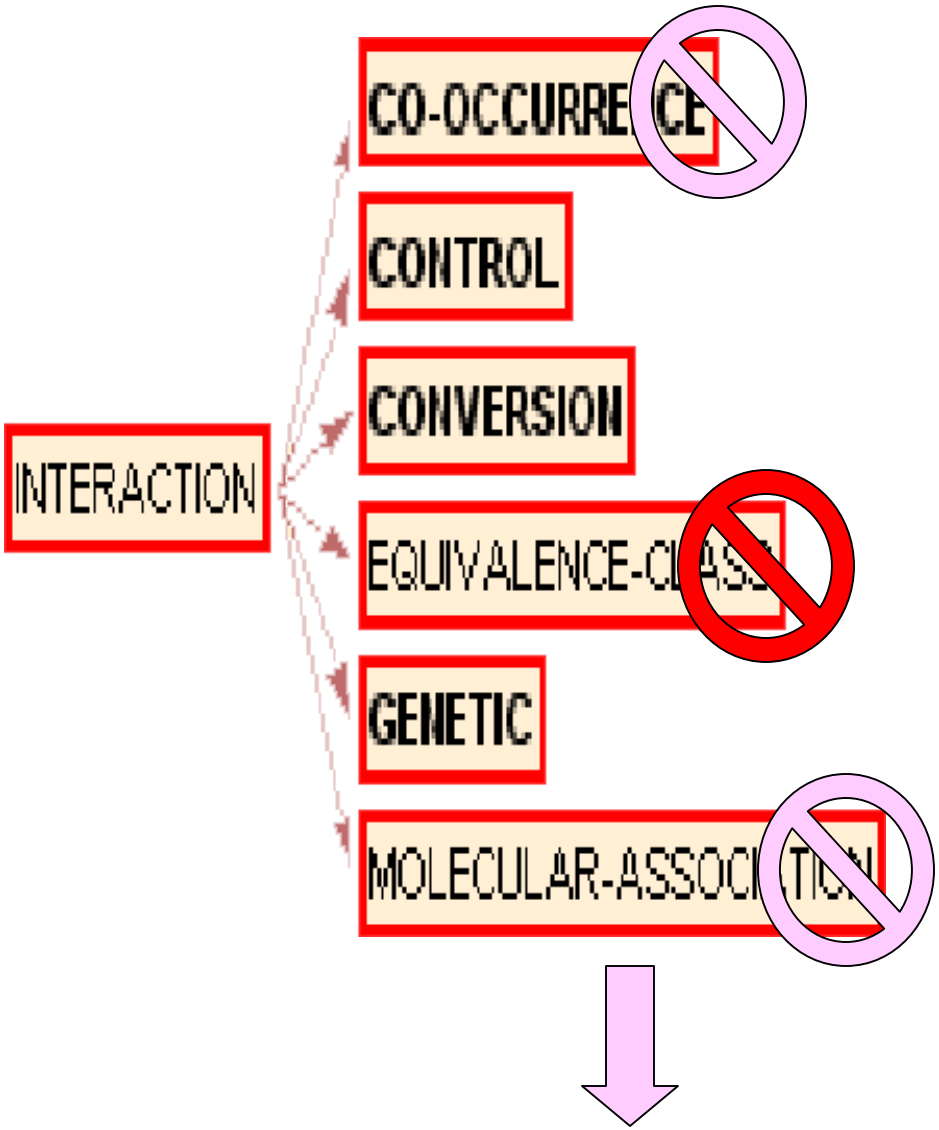
BioPAX Ontology: Parts

- **Cell**
- **Cell Component**
- **DNA**
- **Environment**
- **Photon**
- **Protein**
- **RNA**
- **Small Molecule**



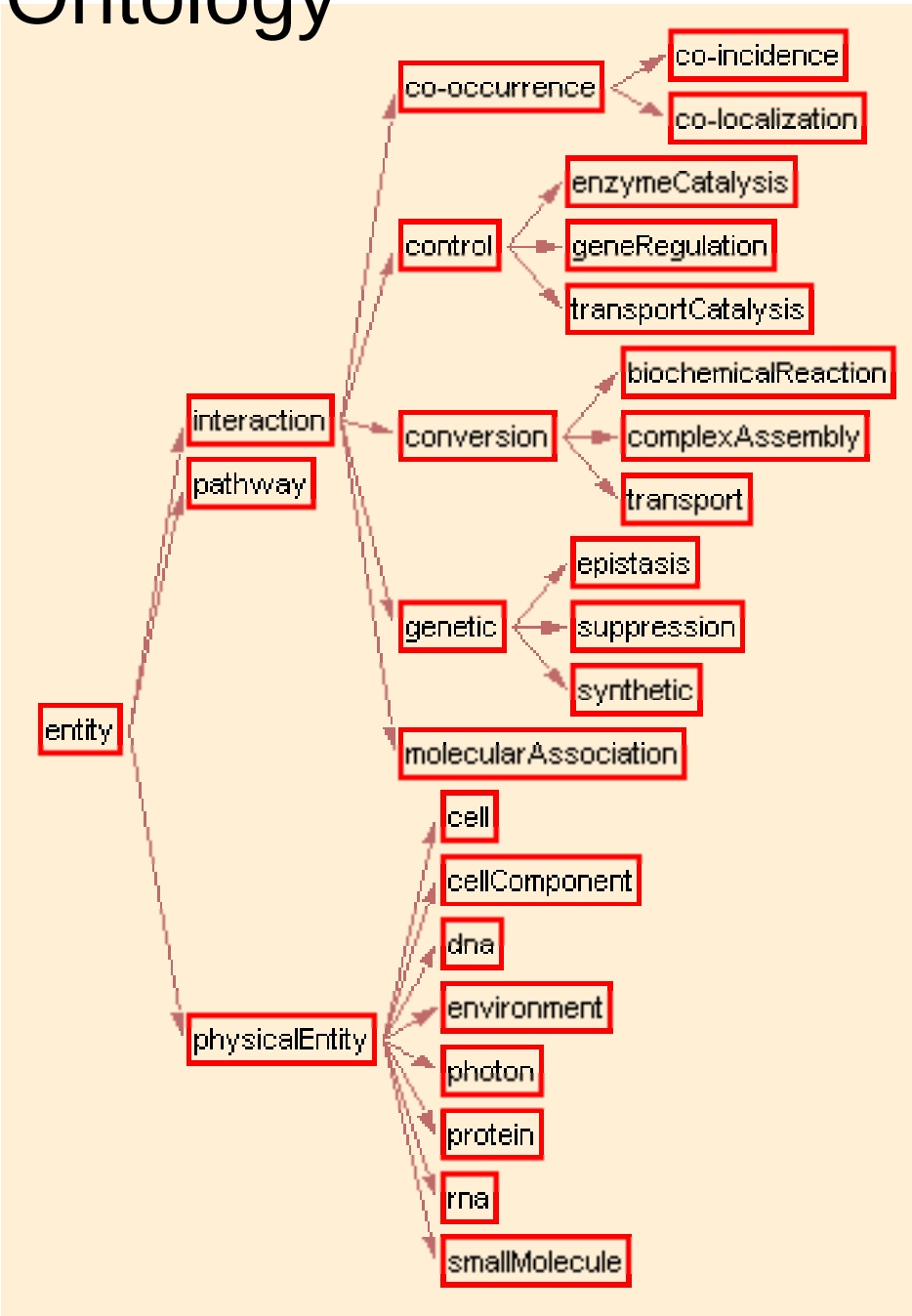
BioPAX Ontology: Interactions

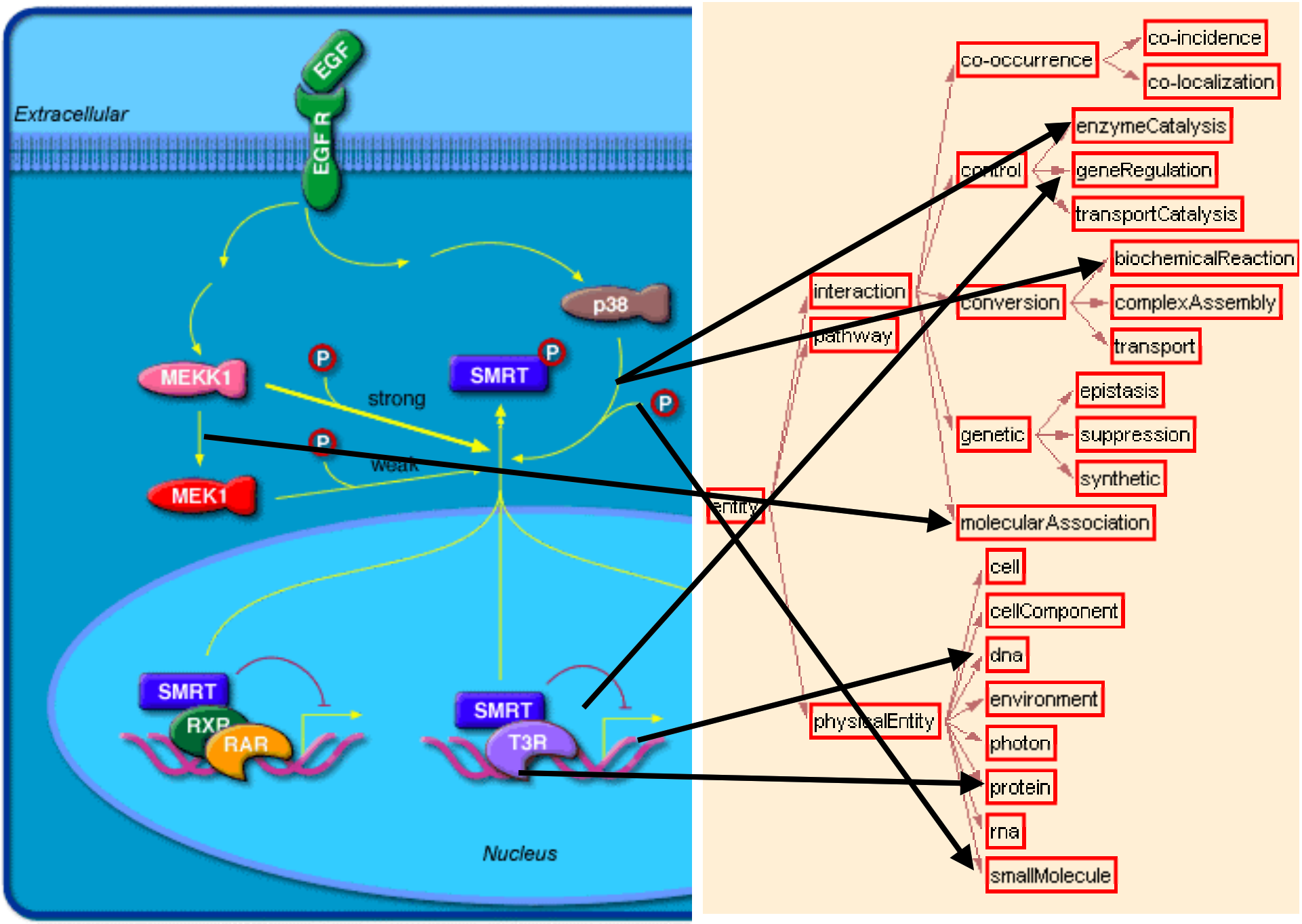
- **Control**
- **Conversion**
- **Molecular Association**
- **Co-occurrence**
- **Equivalence Class**
- **Genetic**



BioPAX Ontology

- Current structure of class hierarchy:




Extracellular
EGF
EGF R
p38
MEKK1
P
SMRT
strong
weak
MEK1
SMRT
RXR
RAR
SMRT
T3R
Nucleus
co-occurrence
co-incidence
co-localization
control
enzymeCatalysis
geneRegulation
transportCatalysis
interaction
pathway
conversion
biochemicalReaction
complexAssembly
transport
genetic
epistasis
suppression
synthetic
entity
molecularAssociation
physicalEntity
cell
cellComponent
dna
environment
photon
protein
rna
smallMolecule

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: Small Molecule

- Evaluated CML 2.0 as means for exchanging small molecules
  - No comprehensive small molecule DB exists
    - Need to transfer entire small molecule structure, not just DB x-ref
  - Proof of concept:
    1. EcoCyc small molecules → CML 2.0 file
    2. CML 2.0 file → Shah lab visualization program
       - No loss of information

BioPAX Subgroups

- State Representation:
  - Determining best mechanism to represent biological states
    - E.g. post-translational modification states of proteins, cell-cycle states
- Examples:
  - Gather examples to test and illustrate BioPAX

Current Status

- Initial meeting Nov. 2002
- Maintained conference call level
- Version 0.5 of Level 1 Ontology released Sep.24
  - Finishing slot definitions on Level 1 main-tree classes
  - Finishing class structure of side-trees
    - States, evidence, cellular localization
- Translating records from major DBs to BioPAX

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 Supporting Groups

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 Insitute: C. Hogue

Collaborating Organizations:

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

Databases

- BioCyc (www.biocyc.org)
- BIND (www.bind.ca)
- WIT (wit.mcs.anl.gov/WIT2)

Grants

- Department of Energy