PIR A UniProt CONSORTIUM MEMBER
Protein Information Resource

iPTMnet: Integrative Bioinformatics for Knowledge Discovery of PTM Networks

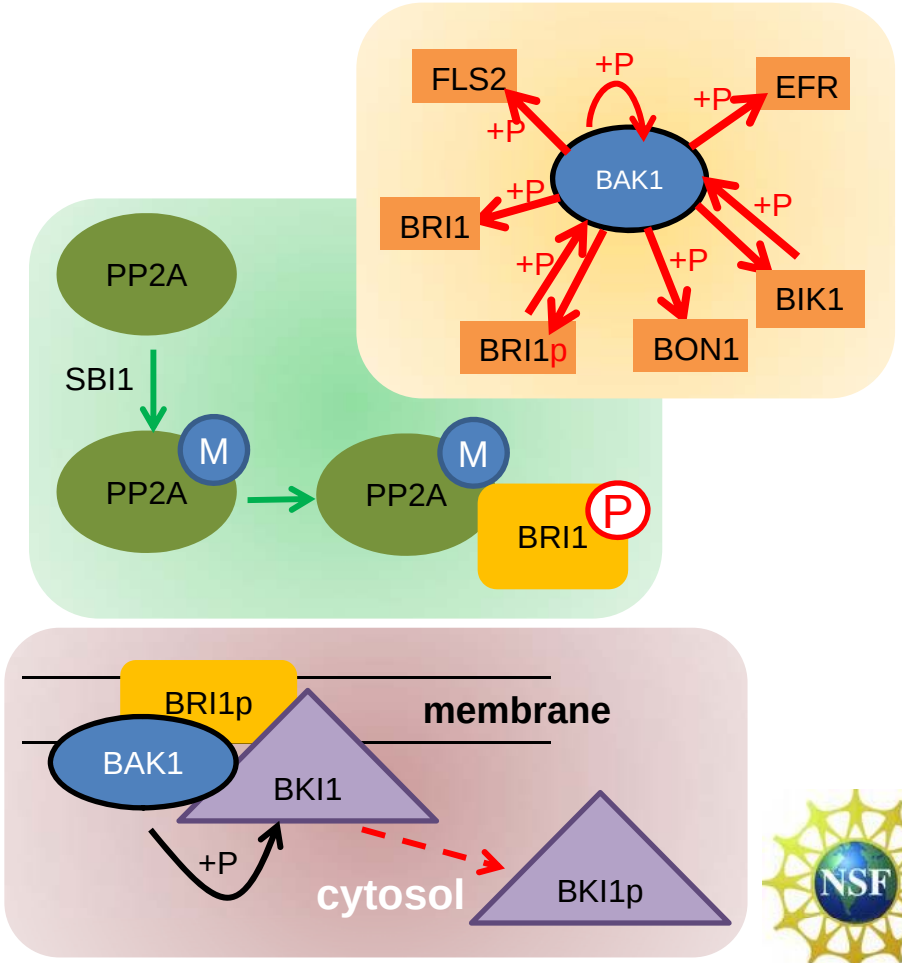
PIR Knowledge Mining Tools for Proteins, Complexes and PTMs
January 15, 2012

Cathy H. Wu, Ph.D.
Director, Protein Information Resource (PIR)
Edward G. Jefferson Chair and Director
Center for Bioinformatics & Computational Biology, University of Delaware
Professor of Biochemistry & Molecular Biology, Georgetown University

Scientific Goals

iPTMnet for Knowledge Discovery of PTM Networks

- Capture PTM enzyme-substrate relationships
- Connect major PTMs that work in concert in specific biological processes: phosphorylation, glycosylation, acetylation, methylation, myristoylation, ubiquitination
- Link relevant biological contexts to individual PTM forms
- Connect PTMs across taxons: orthologous PTMs



ABI Development: Integrative Bioinformatics for Knowledge Discovery of PTM Networks (http://www.nsf.gov/awardsearch/showAward.do?AwardNumber=1062520)

From Data to Knowledge Map

Literature-based Experimental Data

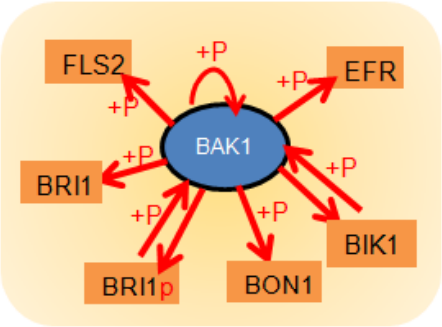
PTM Network

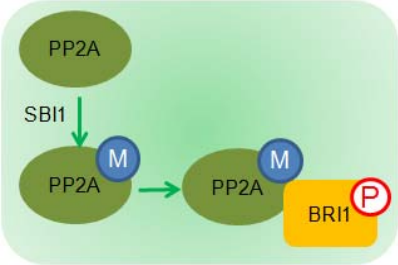
Substrate	Kinase	Site	PMID
BAK1	BRI1		21593986
	BAK1	Y-610	21350342
	BAK1	Y-610	21623975
	BIK1-pT237	T-237	20018686

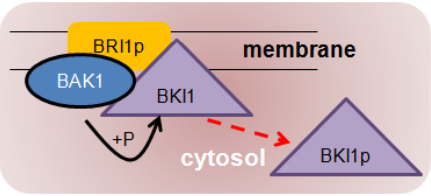
Kinase	Substrate	Site	PMID
BAK1	BRI1		12150928
	BIK1	T-237	20404519
	BON1		21623975
	BIR1		21623975
	FLS2		21593986
	EFR		21593986

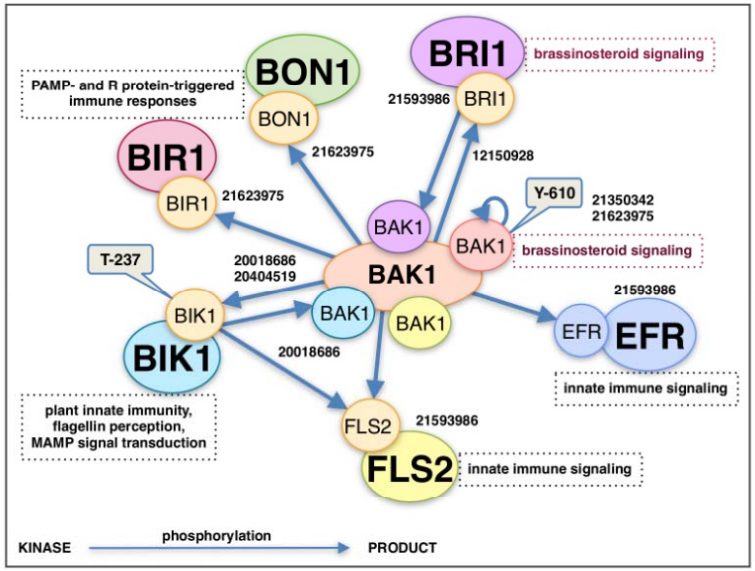
Substrate	Methylase	Methylated protein interacts with	PMID
PP2A	SBI1	BRI1-P	21558554

Substrate	Phosphorylated residue	Impact	PMID
BKI1	tyrosine	cytosol	21289069



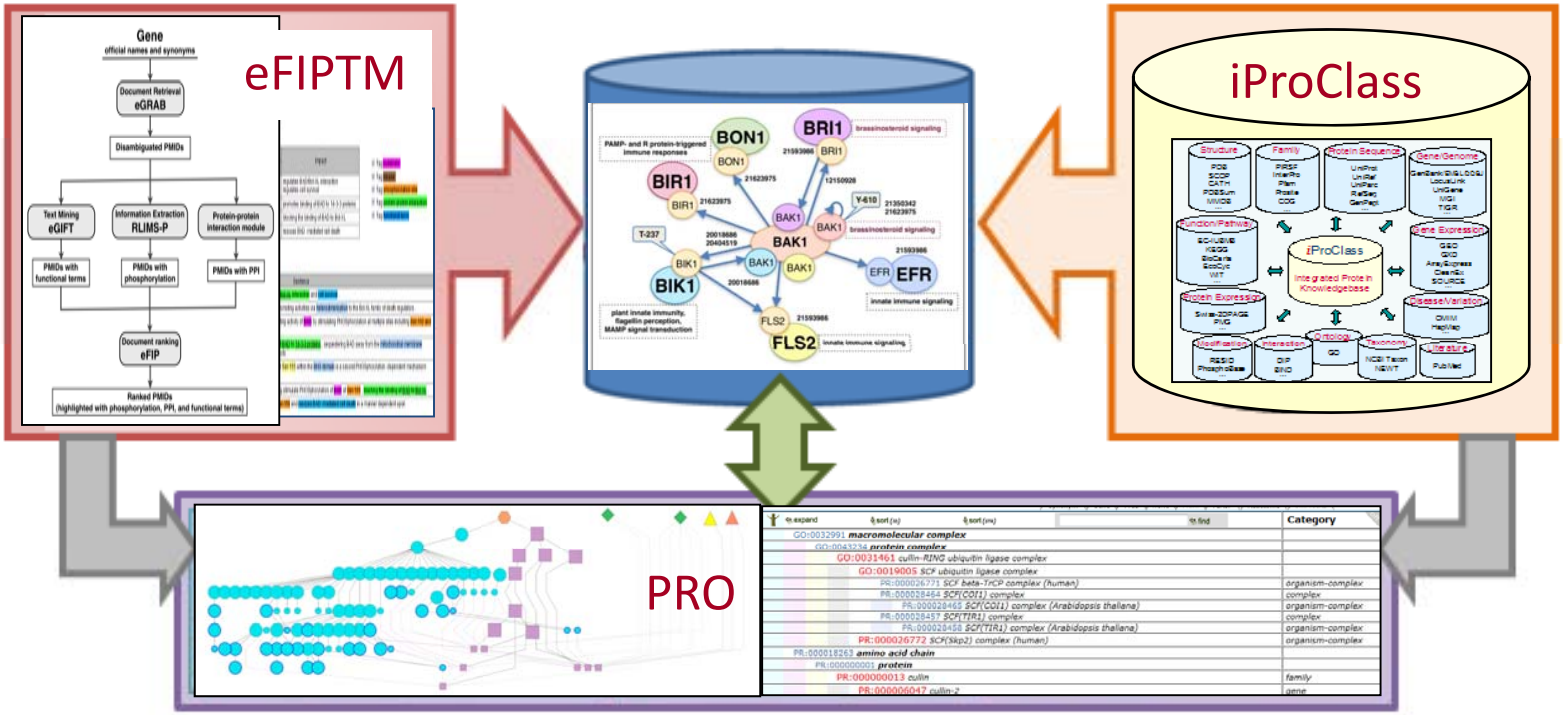




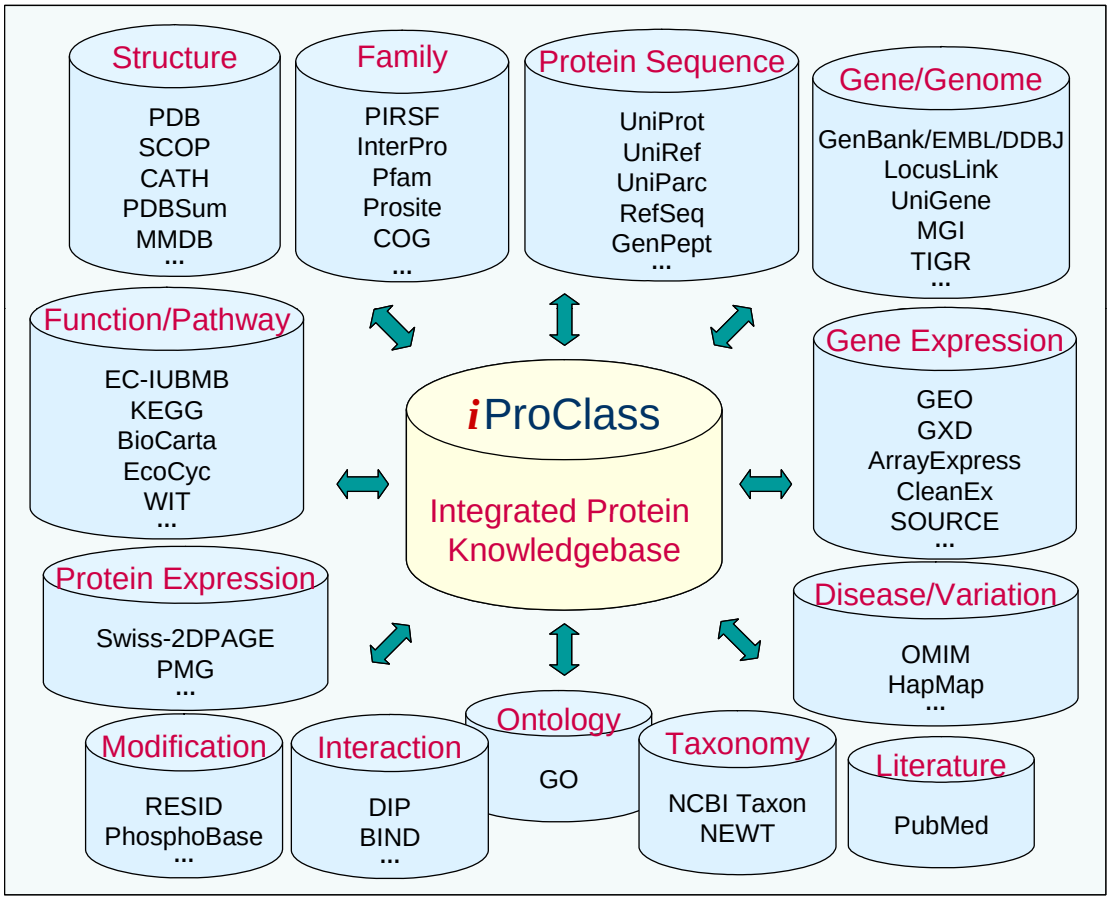


Bioinformatics Framework

- Database for molecular and omics data integration: iProClass
- Text mining system for knowledge extraction from scientific literature: eFIPTM
- Ontologies for knowledge representation of specific protein PTM forms: PRO
- Web portal linking data and analysis/visualization tools for scientific queries (http://proteininformationresource.org/iPTMnet)



iProClass Database



- Data integration from >120 molecular/omics databases
- Additional PTM and plant-specific databases to be integrated
  - Phosphorylation and PTM: e.g., PhosPhAt and P3DB
  - Plant interactome and PPI: e.g., AtPIN, AtPID, plant interactome network map
  - Plant pathways: e.g., reference (PlantCyc) & species-specific (e.g. AraCyc)
  - Subcellular localization: e.g., SUBA

The iProClass integrated database for protein functional analysis
Wu CH, Huang H, Nikolskaya A, Hu Z, Barker WC. (2004)
Computational Biology & Chemistry 28, 87-96. [PMID:15022647]

RLIMS-P: Rule-based Literature Mining System for Phosphrylation (http://www.proteininformationresource.org/pirwww/iprolink/rlimsp.shtml)

Extract phosphorylation objects (substrate, kinase, site) from PubMed

2. Report: Full annotation with evidence tagging and PMID mapping to UniProt
3. Name mapping searches BioThesaurus

HOME / iProLINK / RLIMS-P

RLIMS-P Search Form

Enter PubMed IDs (PMIDs) delimited by "," or space, e.g., 2108025, 16436437...

2108025, 2261989, 16436437, 16436420, 16438659

Submit Reset

When input exceeds 100 PMIDs, the process may take up to 1 to 2 minutes. Sample output

1. Summary table: PMIDs & top-ranking annotation

1

RLIMS-P Summary Table

Abstract(s) with protein phosphorylation. One top hit is displayed for each abstract. Go to Text Evider

PubMed ID	Protein Kinase	Phosphorylated Protein (Substrate)	Phosphorylati
2108025	phosphorylase kinase	rabbit skeletal muscle phosphorylase kinase	serine972;serin
2261989	partially purified EF-2 kinase	Elongation factor 2 (EF-2)	Thr56;Thr58;Th
16436437	PknB	PBPA	Thr362;Thr437,

Retrieve Full Reports

Abstract(s) with no protein phosphorylation found. PubMed ID: 16436420, 16438659

2

PubMed Information

PubMed ID	Publication Date	Authors	Journal
2108025	1990/Mar/10	Meyer HE , Meyer GF , Dirks H , Heilmeyer LM Jr	Eur J Biochem

PMID Mapping to UniProtKB Advanced Search

Protein AC/ID	Protein Name	Organism Name	PIRSF ID	Matched Fields
P18688/KPB1_RABIT ProClass UniProtKB/Swiss-Prot	Phosphorylase b kinase alpha regulatory chain, skeletal muscle isoform (Phosphorylase kinase alpha M subunit)	Oryctolagus cuniculus (Rabbit)	PIRSF015844	PubMed ID=>2108025

Name Mapping to UniProtKB

Protein name from abstract: rabbit skeletal muscle phosphorylase kinase

User-specified protein name:

3 Name Mapping

Annotation

No.	Protein Kinase	Phosphorylated Protein (Substrate)	Phosphorylation Site
1	phosphorylase kinase	rabbit skeletal muscle phosphorylase kinase	serine 972; serine 1007; serine 985
2	the cyclic-AMP -dependent protein kinase	rabbit skeletal muscle phosphorylase kinase	Serine 1018
3	N/A	rabbit skeletal muscle phosphorylase kinase	serine 1018; serine 1023; serine 1020

Text Evidence

TI - Localization of **PHOSphoserine residues** in **the alpha subunit** of **rabbit skeletal muscle phosphorylase kinase**.

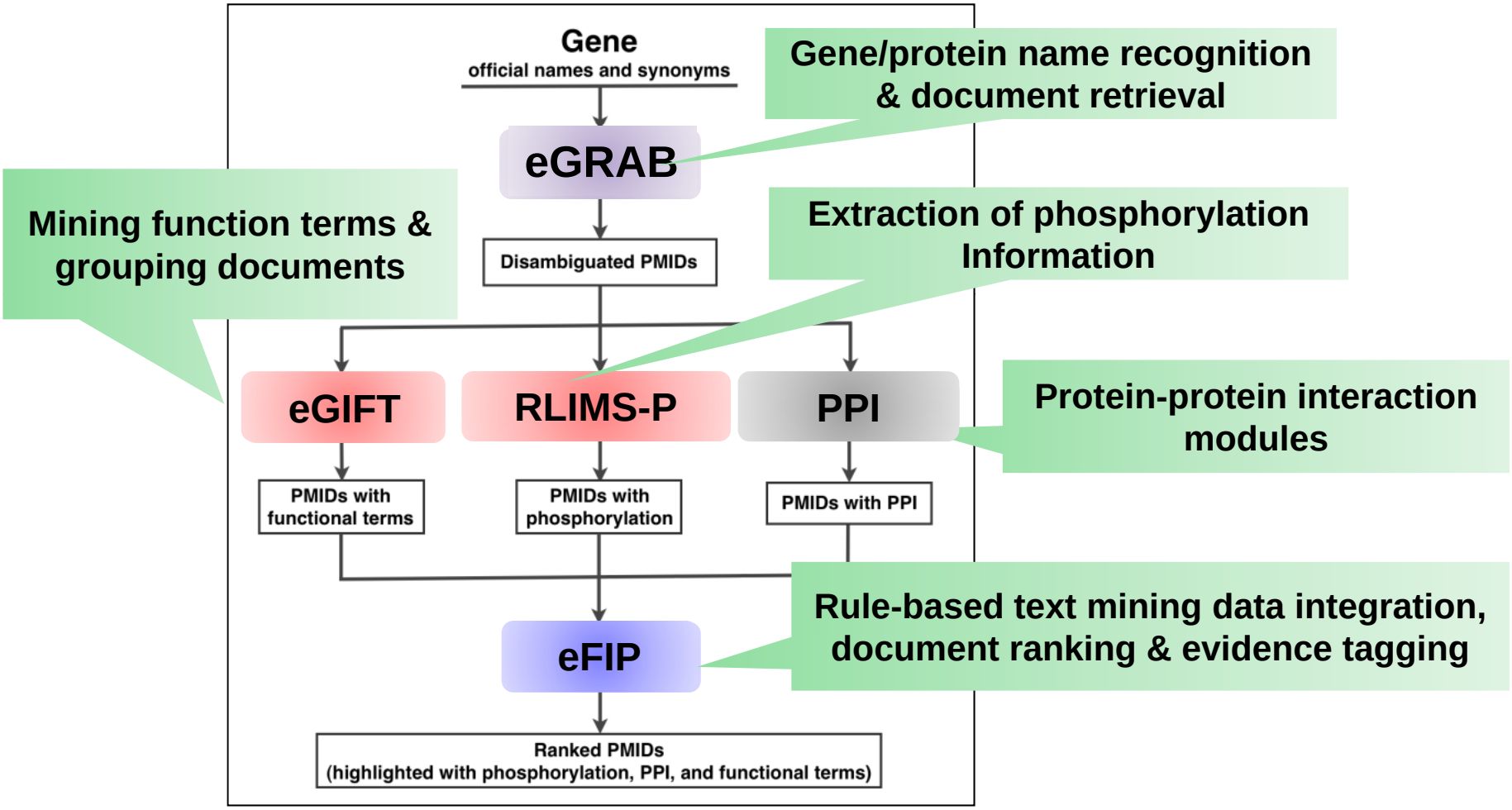
AB - The alpha subunit of skeletal muscle phosphorylase kinase, as isolated, carries phosphate at **the serine residues 1018, 1020 and 1023**. Employing the S-ethyl-cysteine method, these residues are found to be phosphorylaTED partially, i.e.differently PHOSphorylated species exist in muscle. **Serine 1018** is a site which can be PHOSphorylated by **the cyclic-AMP -dependent protein kinase**. **The serine residues 972, 985 and 1007** are PHOSphorylated by **phosphorylase kinase** itself when its activity is stimulated by micromolar concentrations of Ca2. These phosphorylaTION sites are not identical to those found to be PHOSphorylated already in the enzyme as prepared from freshly excised muscle. A multiphosphorylaTION loop uniquely present in this but not in the homologous beta subunit contains **all the PHOSphoserine residues** so far identified in the alpha subunit.

Tag Protein Kinase Tag Phosphorylated Protein Tag Phosphorylation Site

An online literature mining tool for protein phosphorylation

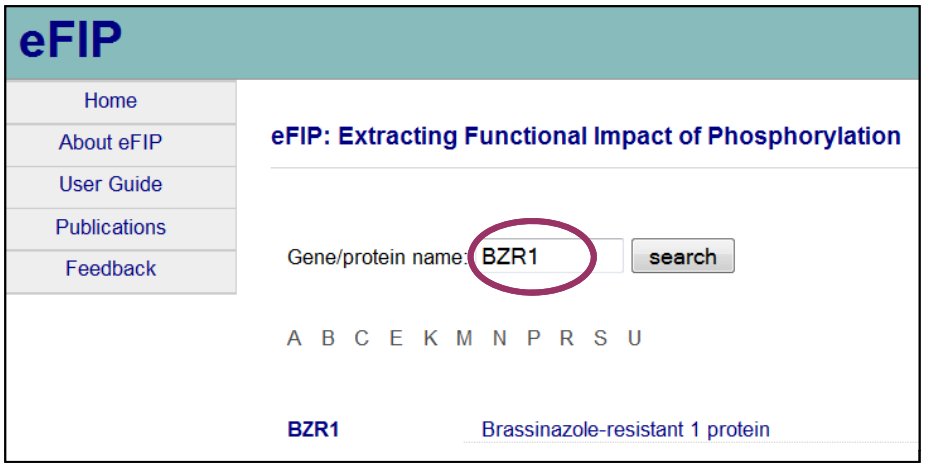
Yuan X, Hu ZZ, Wu HT, Torii M, Narayanaswamy M, Ravikumar KE, Vijay-Shanker K, Wu CH. (2006) *Bioinformatics* 22, 1668-1669 [PMID:16644790]

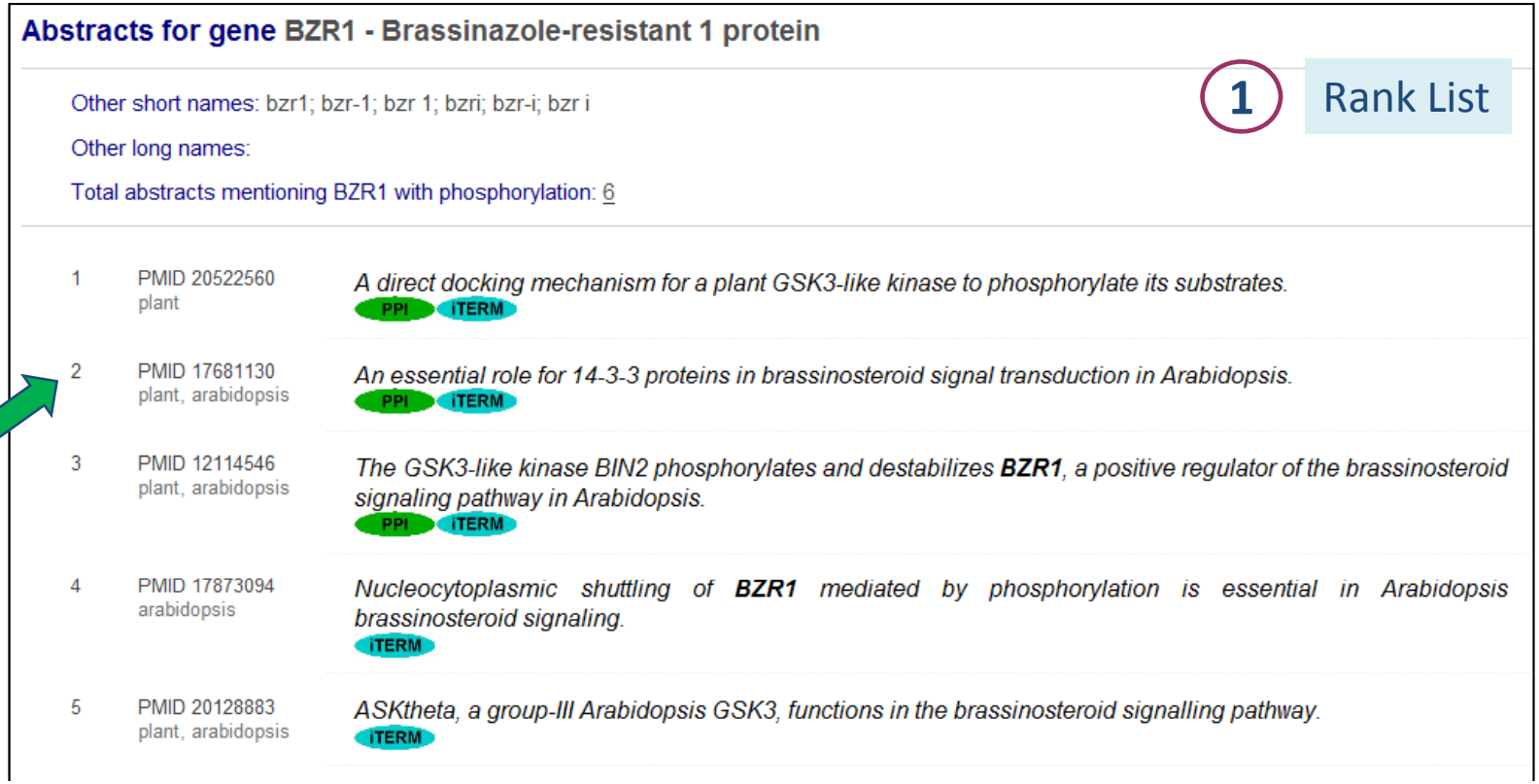
eFIP: an integrated system for mining Functional Impact of Phosphorylation from literature



eFIP: a tool for mining functional impact of phosphorylation from literature
Arighi CN, Siu AY, Tudor CO, Nchoutmboube JA, Wu CH, Shanker VK. (2011)
Methods in Molecular Biology 694, 63-75 [PMID:21082428]

eFIP to Mine Phosphorylation Network





eFIP to Mine Phosphorylation Network

Go to PubMed **PMID 17681130 for gene BIN2 - Shaggy-related protein kinase eta**

Other short names: bin2; bin-2; bin 2; binii; bin-ii; bin ii

Other long names: shaggy-related protein kinase eta

2 Summary Table

Authors: Gampala SS, Kim TW, He JX, Tang W, Deng Z, Bai MY, Guan S, Lalonde S, Sun Y, Gendron JM, Chen H, Shibagaki N, Ferl RJ, Ehrhardt D, K, Burlingame AL, Wang ZY

Species: plant, arabidopsis

Sentence	Phosphorylation			Impact to protein-protein interactions	
	Substrate	Site	Kinase	How	What
5	BZR1/BZR2	---	BIN2	also promotes	binding to the 14-3-3 proteins
6	BZR1	---	---	abolish	14-3-3 binding

Sentence	Phosphorylation			iTerms	
	Substrate	Site	Kinase	How	What
5	BZR1/BZR2	---	BIN2	inhibits	DNA binding
6,7	BZR1	---	---	increases	cytoplasmic localization

Toggle highlights: ☑ Substrate ☑ Kinase ☑ Phosphorylation site ☑ Protein-protein interaction ☑ iTerm

3 Evidence Tagging

Title 1 An essential role for 14-3-3 proteins in brassinosteroid signal transduction in Arabidopsis .

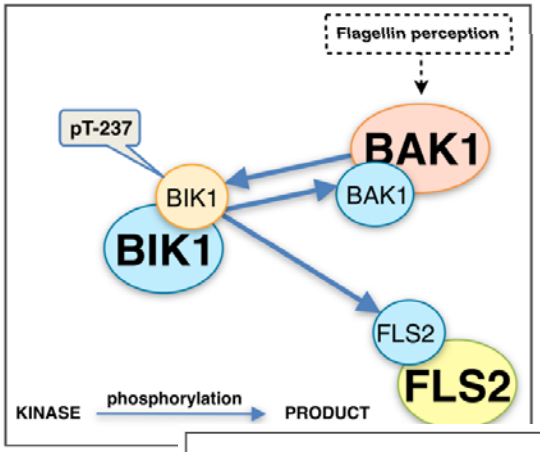
Abstract 2 Brassinosteroids (BRs) are essential hormones for plant growth and development .

3 BRs regulate gene expression by inducing dePHOSphorylation of two key transcription factors , BZR1 and BZR2/BES1 , through a signal transduction pathway that involves cell -surface receptors (BRI1 and BAK1) and a GSK3 kinase (BIN2) .

4 How BR -regulated PHOSphorylation controls the activities of BZR1/BZR2 is not fully understood .

5 Here , we show that BIN2 -catalyzed PHOSphorylation of BZR1/BZR2 not only inhibits DNA binding , but also promotes binding to the 14-3-3 proteins .

Phosphorylation Network



Sentence	Phosphorylation			Impact to protein-protein interactions	
	Substrate	Site	Kinase	How	What
5,6	cytoplasmic kinase BIK1	---	---	---	BIK1 associates with FLS2 and BAK1

Sentence	Phosphorylation			iTerms	
	Substrate	Site	Kinase	How	What
8	BIK1	The flagellin phosphorylaTION site Thr(237)	---	---	flagellin perception

Toggle highlights: ☑ Substrate ☑ Kinase ☑ Phosphorylation site ☑ Protein-protein interaction ☑ iTerm

Title

1 A receptor -like cytoplasmic kinase , BIK1 , associates with a flagellin receptor complex to initiate plant innate immunity .

2 Plants and animals rely on innate immunity to prevent infections by detection of microbe-associated molecular patterns (MAMPs) through pattern-recognition receptors (PRRs) .

3 The plant PRR FLS2 , a leucine-rich repeat receptor kinase , recognizes bacterial flagellin and initiates immune signaling by association with another leucine-rich repeat-receptor -like kinase , BAK1 .

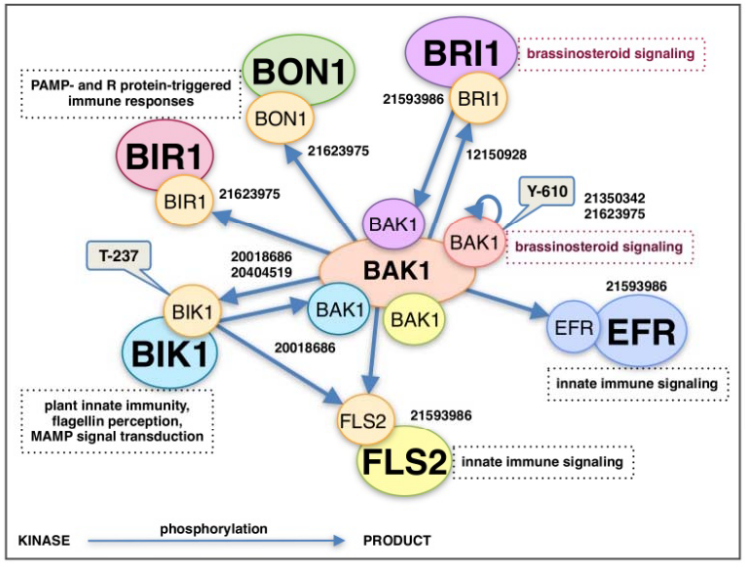
4 It remains unknown how the FLS2/BAK1 receptor complex activates intracellular signaling cascades .

5 Here we identified the receptor -like cytoplasmic kinase BIK1 that is rapidly PHOSphorylated upon flagellin perception , depending on both FLS2 and BAK1 .

6 BIK1 associates with FLS2 and BAK1 in vivo and in vitro .

7 BIK1 is phosphorylaTED by BAK1 and BIK1 also directly PHOSphorylates BAK1 and FLS2 in vitro .

8 The flagellin phosphorylaTION site Thr(237) of BIK1 is required for its phosphorylaTION on BAK1 and FLS2 , suggesting that BIK1 is likely first phosphorylaTED upon flagellin perception and subsequently transPHOSphorylates FLS2/BAK1 to propagate flagellin signaling .

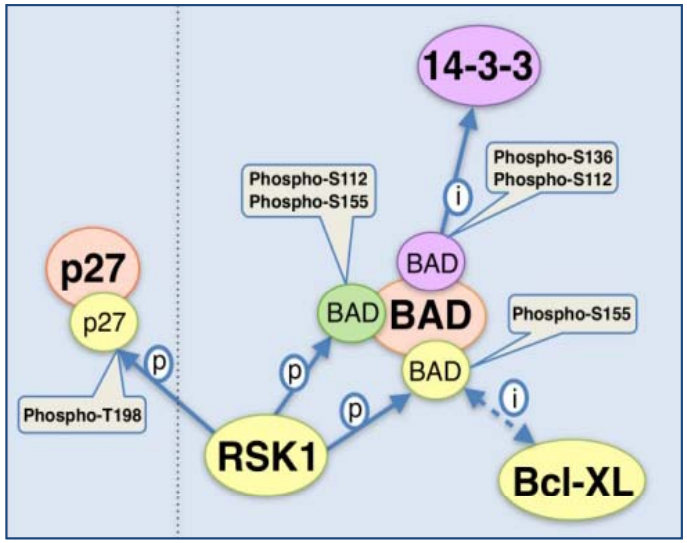


eFIP (phosphorylation)

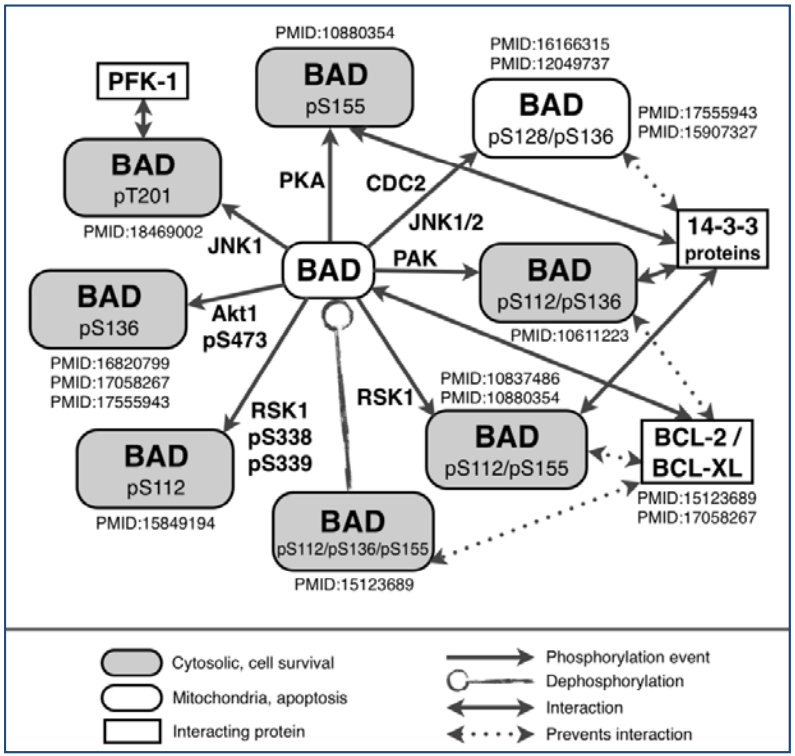
=> **eFIPTM** (extending to other PTMs)

Discovery from Literature Mining

- Distinct phosphorylated forms of a protein may have different interacting proteins, leading to different subcellular locations, functions and pathways
- Literature mining connects the impact (cytosolic vs mitochondria; apoptosis vs cell survival) to different BAD forms, and, through kinases, links BAD to pathways.
- PTM network of enzyme-substrate relationships and protein-protein interactions mined from the scientific literature => iPTMnet with rich relations
- Knowledge gap: PTM-specific annotation not captured in most sequence, model organism, or PPI databases.



p = phosphorylation
i = protien-protein interaction



Protein Ontology (PRO)

- PRO in OBO (Open Biological and Biomedical Ontologies)
  - Reference Ontology for Proteins
  - Capture knowledge about proteins to model biology

- One of the first set of OBO Foundry ontologies

OBO Foundry ontologies

Title	Domain	Prefix	
Biological process	biological process	GO	gene_ontolo
Cellular component	anatomy	GO	gene_ontolo
Chemical entities of biological interest	biochemistry	CHEBI	chebi.obo
Molecular function	biological function	GO	gene_ontolo
Phenotypic quality	phenotype	PATO	quality.obo
PRotein Ontology (PRO)	proteins	PR	pro.obo
Xenopus anatomy and development	anatomy	XAO	xenopus_an
Zebrafish anatomy and development	anatomy	ZFA	zebrafish_ar

- Three sub-ontologies
  - Ontology for Protein Evolution (ProEvo): Captures protein classes reflecting evolutionary relatedness of proteins
  - Ontology for Protein Forms (ProForm): Captures different protein forms of a given gene locus from genetic variations, alternative splicing, proteolytic cleavage, **PTMs**
  - Ontology for Protein Complexes (ProComp): Captures distinct complexes as they exist in different species and defines complexes through component proteins

The Protein Ontology: a structured representation of protein forms and complexes
Natale DA, Arighi CN, Barker WC, Blake JA, Bult CJ, Caudy M, Drabkin HJ, D'Eustachio P, Evsikov AV, Huang H, Nchoutmboube J, Roberts NV, Smith B, Zhang J, Wu CH. (2011)
Nucleic Acids Res. 39, D539-545 [PMID:20935045]

NIGMS

PRO Hierarchical View

66 shown of 28321 records

PMID | Taxon | PANTHER | EcoCyc | Definition | Synonym | Gene | MGI | HGNC | Pfam | PIRSF | Reactome | UniProtKB

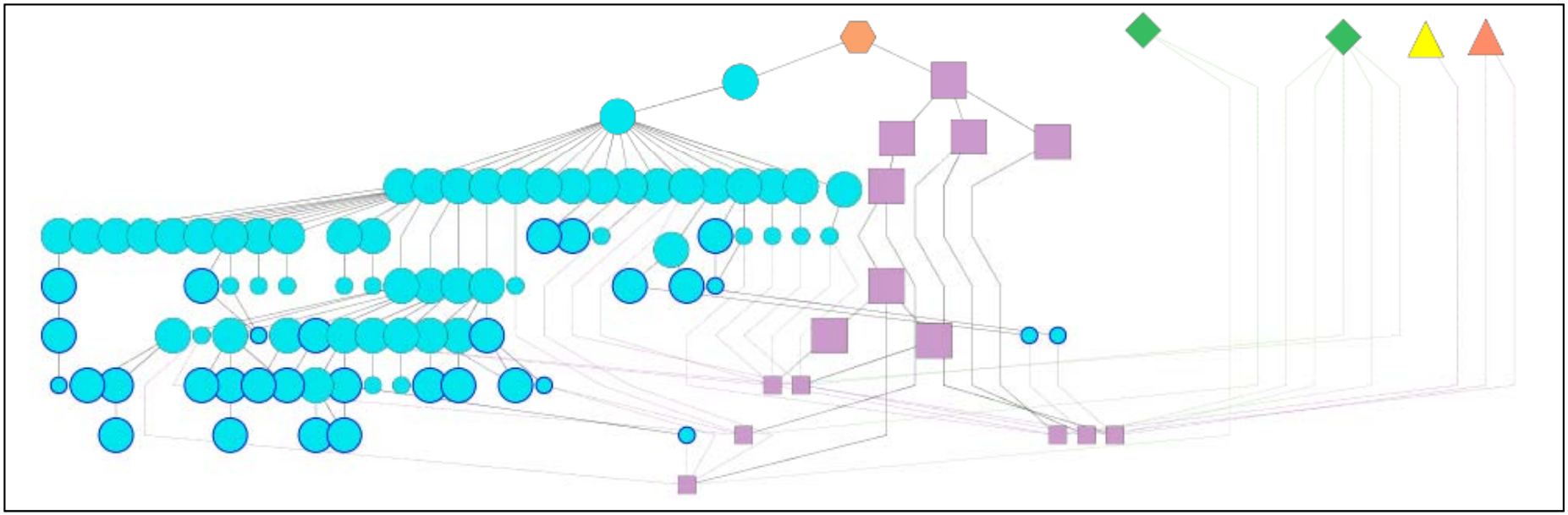
expand | sort (ID) | sort (STR) | find

	Category
GO:0032991 ***macromolecular complex***	
GO:0043234 ***protein complex***	
GO:0031461 *cullin-RING ubiquitin ligase complex*	
GO:0019005 *SCF ubiquitin ligase complex*	
PR:000026771 *SCF beta-TrCP complex (human)*	*organism-complex*
PR:000028464 *SCF(COI1) complex*	*complex*
PR:000028465 *SCF(COI1) complex (Arabidopsis thaliana)*	*organism-complex*
PR:000028457 *SCF(TIR1) complex*	*complex*
PR:000028458 *SCF(TIR1) complex (Arabidopsis thaliana)*	*organism-complex*
PR:000026772 *SCF(Skp2) complex (human)*	*organism-complex*
PR:000018263 ***amino acid chain***	
PR:000000001 ***protein***	
PR:000000013 *cullin*	*family*
PR:000006047 *cullin-2*	*gene*
PR:000006048 *cullin-3*	*gene*
PR:000006049 *cullin-4A*	*gene*
PR:000006050 *cullin-4B*	*gene*
PR:000006051 *cullin-5*	*gene*
PR:000000092 *fungal/metazoan cullin-1*	*gene*
PR:000026757 *cullin-1 (human)*	*organism-gene*
PR:000000267 *cullin-1 isoform 1*	*sequence*
PR:000000427 *cullin-1 isoform 1 neddylated form*	*modification*
PR:000000542 *cullin-1 isoform 1 neddylated 1*	*modification*
PR:000002507 *cullin-1 isoform 1 unmodified form*	*modification*
PR:000027962 *plant cullin-1*	*gene*
PR:000027963 *cullin-1 (Arabidopsis thaliana)*	*organism-gene*
PR:000027976 *plant cullin-1 rubylated form*	*modification*
PR:000027977 *cullin-1 rubylated (Arabidopsis thaliana)*	*organism-modification*
PR:000028501 *plant cullin-2*	*gene*
PR:000028494 *cullin-2 (Arabidopsis thaliana)*	*organism-gene*
PR:000028495 *plant cullin-3A*	*gene*
PR:000028496 *cullin-3A (Arabidopsis thaliana)*	*organism-gene*
PR:000028497 *plant cullin-3B*	*gene*
PR:000028498 *cullin-3B (Arabidopsis thaliana)*	*organism-gene*

PRO Cytoscape Network View

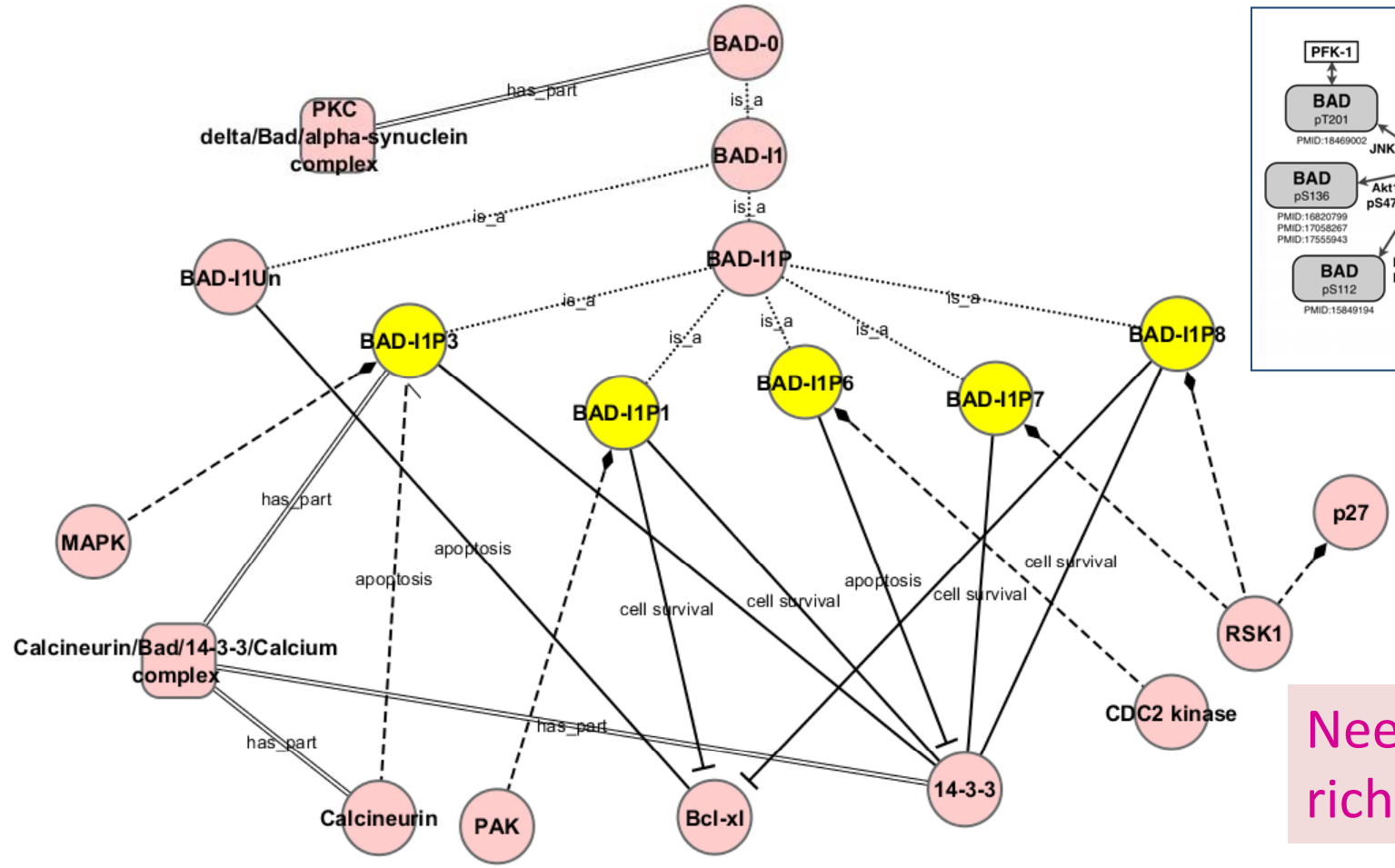
Connecting protein forms and complexes with annotation => Modeling biology

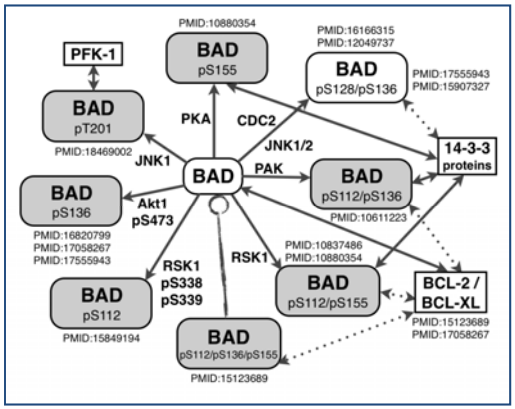
- Show full set of relationships between terms, including those from other ontologies
- Traverse hierarchical (parent-child) relationships
- Find all components of a complex, including modified forms
- Find all complexes that a protein is constituent of = first (complex) neighbors of a protein
- Find all complexes that bind small molecules = those that have relationship to ChEBI



complex
Organism-complex
protein
modified protein form
Organism-protein
ChEBI
Taxon

iPTM Network View





Need to capture rich relations

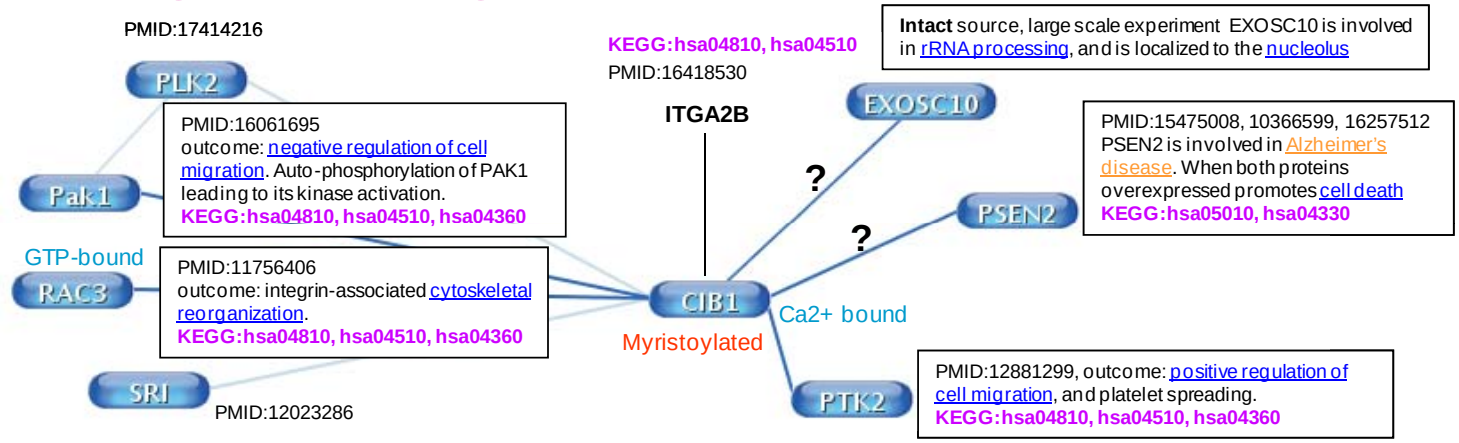
Data Panel

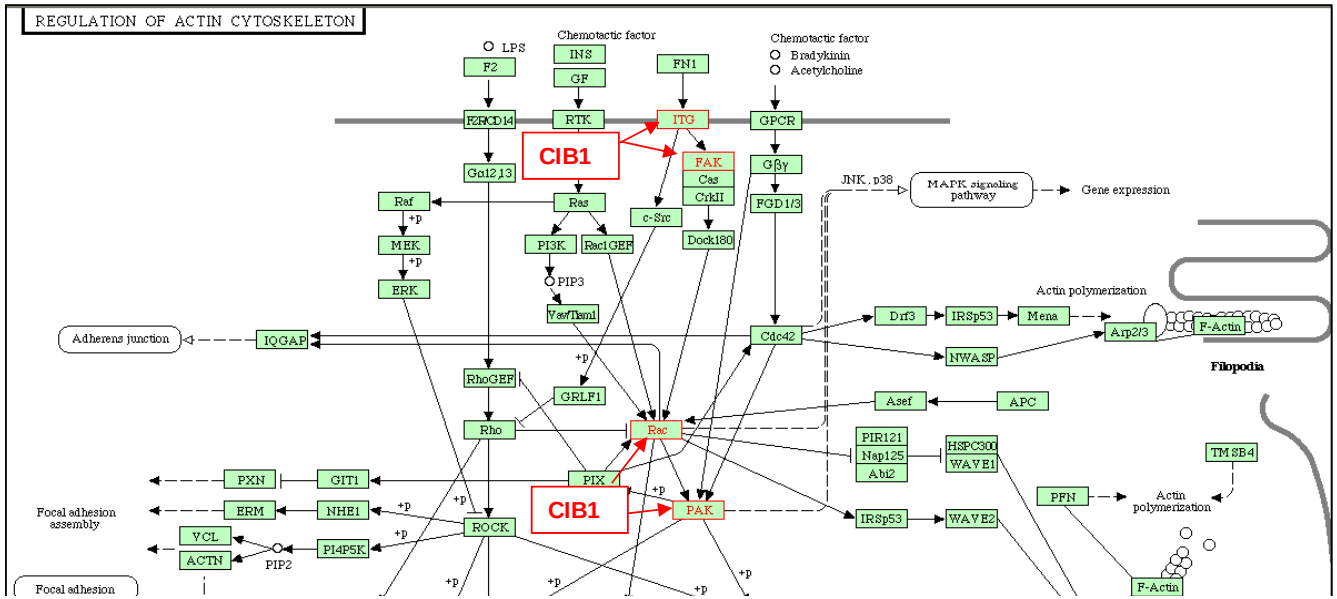
ID	proID	protein name	Modification1	relationship1	Biologicalprocess
BAD-I1P7	PR:000025849	Bcl2 antagonist of cell death isoform 1 phosphorylated 7	UniProtKB:Q61337-1,Ser-112/Ser-155	participates_in	GO:0043066; negative regulation of apoptosis
BAD-I1P1	PR:000003085	Bcl2 antagonist of cell death isoform 1 phosphorylated 1	UniProtKB:Q61337-1, Ser-112/Ser-136	participates_in	GO:0043066; negative regulation of apoptosis
BAD-I1P6	PR:000003269	Bcl2 antagonist of cell death isoform 1 phosphorylated 6	UniProtKB:Q61337-1, Ser128/Ser-136	participates_in	GO:0006915; apoptosis
BAD-I1P8	PR:000025850	Bcl2 antagonist of cell death isoform 1 phosphorylated 8	UniProtKB:Q61337-1, Ser-155	participates_in	GO:0043066; negative regulation of apoptosis
BAD-I1P3	PR:000003087	Bcl2 antagonist of cell death isoform 1 phosphorylated 3	UniProtKB:O35147-1,Ser-112	participates_in	GO:0043066; negative regulation of apoptosis

Exploration of Protein Network

How a single molecule like CIB1 is able to regulate many diverse pathways and what is the consequence of this regulation to the overall cellular behavior

=> Data Mining + Text Mining





iPTMnet Use Cases

iPTMnet is designed to assist plant biologists for knowledge extraction and discovery of proteins undergoing various PTMs. It may provide answers for questions such as:

- What PTMs of a protein of interest are known?
- What specific enzymes (kinases, methylases, etc.) are known to modify a given protein?
- What substrates are known for a given PTM enzyme?
- What interacting partners are known for each PTM form of a given protein?
- What protein modifications and enzymes are known in a given signaling pathway?
- In addition, iPTMnet can be used to generate training sets for PTM site prediction methods.

Contribute to iPTMnet

To capture expert knowledge and enrich the PTM network/pathway map for proteins of interest

- Validating text mining results

  In the text mining result page, use check boxes to indicate correct or wrong annotations as appropriate. In this way we can present the validated result to the public after editorial checking.

- Input user annotations

  Use the Race-PRO interface to add information about a protein and its modification(s) based on literature evidence. The contribution will be reflected in the Protein Ontology (PRO) with appropriate source attribution, and linked from iPTMnet.

PIR A UniProt CONSORTIUM MEMBER
Protein Information Resource

Acknowledgement

Cathy Wu (PI) at PIR/UD-CBCB

Cecilia Arighi at PIR/UD-CBCB
leading Curation Group
Qinghua Wang
Silvia Jimenez (Merck, Geneva)

Vijay Shaker (Co-PI) at UD-CIS
leading Text Mining Group
C. Oana Tudor
Manabu Torii

Jung-Youn Lee (Co-PI) at UD-PLSC
leading Validation Group
Jake Kairis

New Graduate RA and Postdoc positions are available!

NSF ABI #1062520

CENTER FOR BIOINFORMATICS & COMPUTATIONAL BIOLOGY

DELAWARE BIOTECHNOLOGY INSTITUTE

UNIVERSITY OF DELAWARE