Proteomics & Bioinformatics Part II

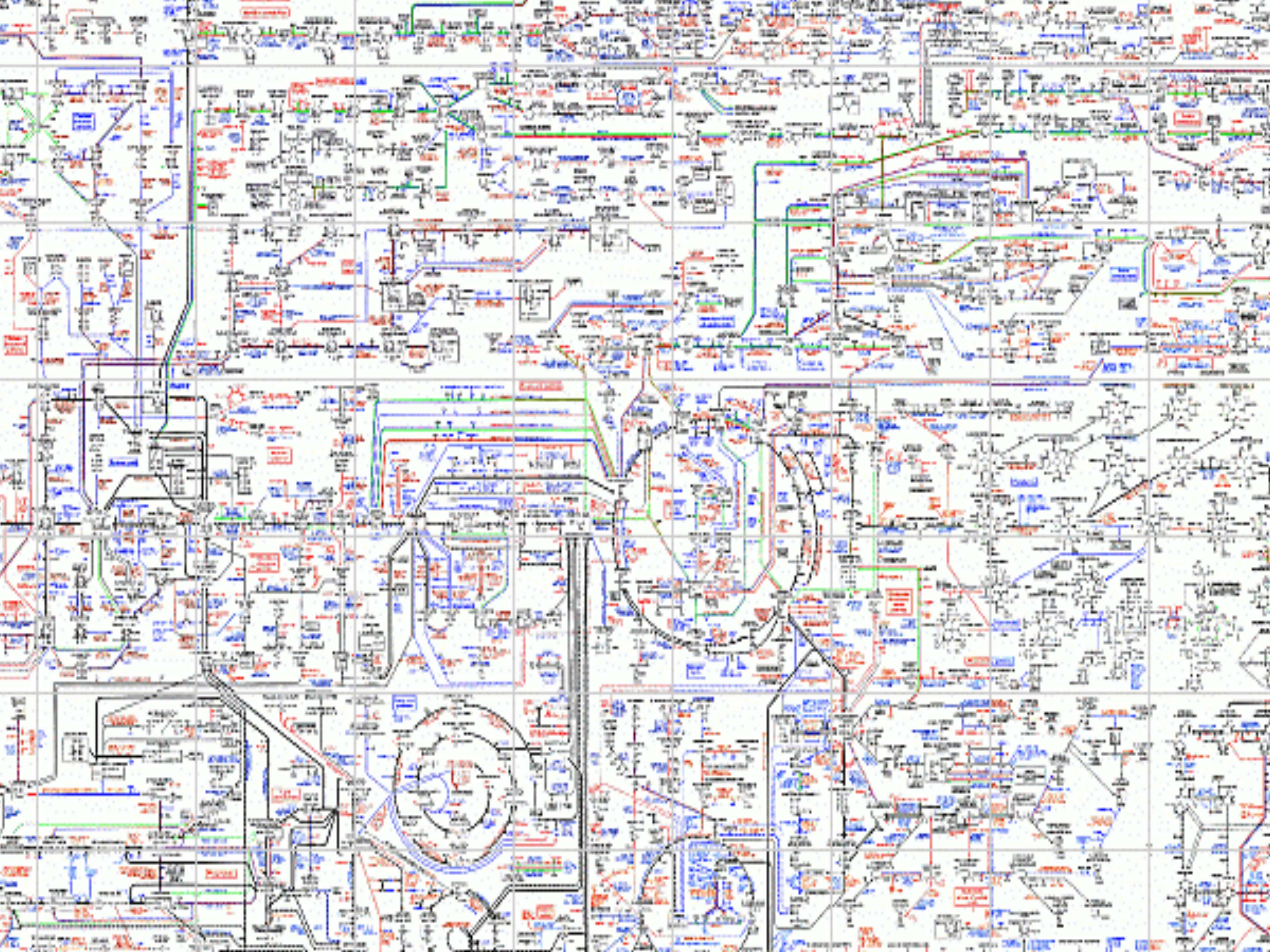
David Wishart
3-41 Athabasca Hall
david.wishart@ualberta.ca

3 Kinds of Proteomics*

- Structural Proteomics
  - High throughput X-ray Crystallography/Modelling
  - High throughput NMR Spectroscopy/Modelling
- Expressional or Analytical Proteomics
  - Electrophoresis, Protein Chips, DNA Chips, 2D-HPLC
  - Mass Spectrometry, Microsequencing
- Functional or Interaction Proteomics
  - HT Functional Assays, Ligand Chips
  - Yeast 2-hybrid, Deletion Analysis, Motif Analysis

Historically...

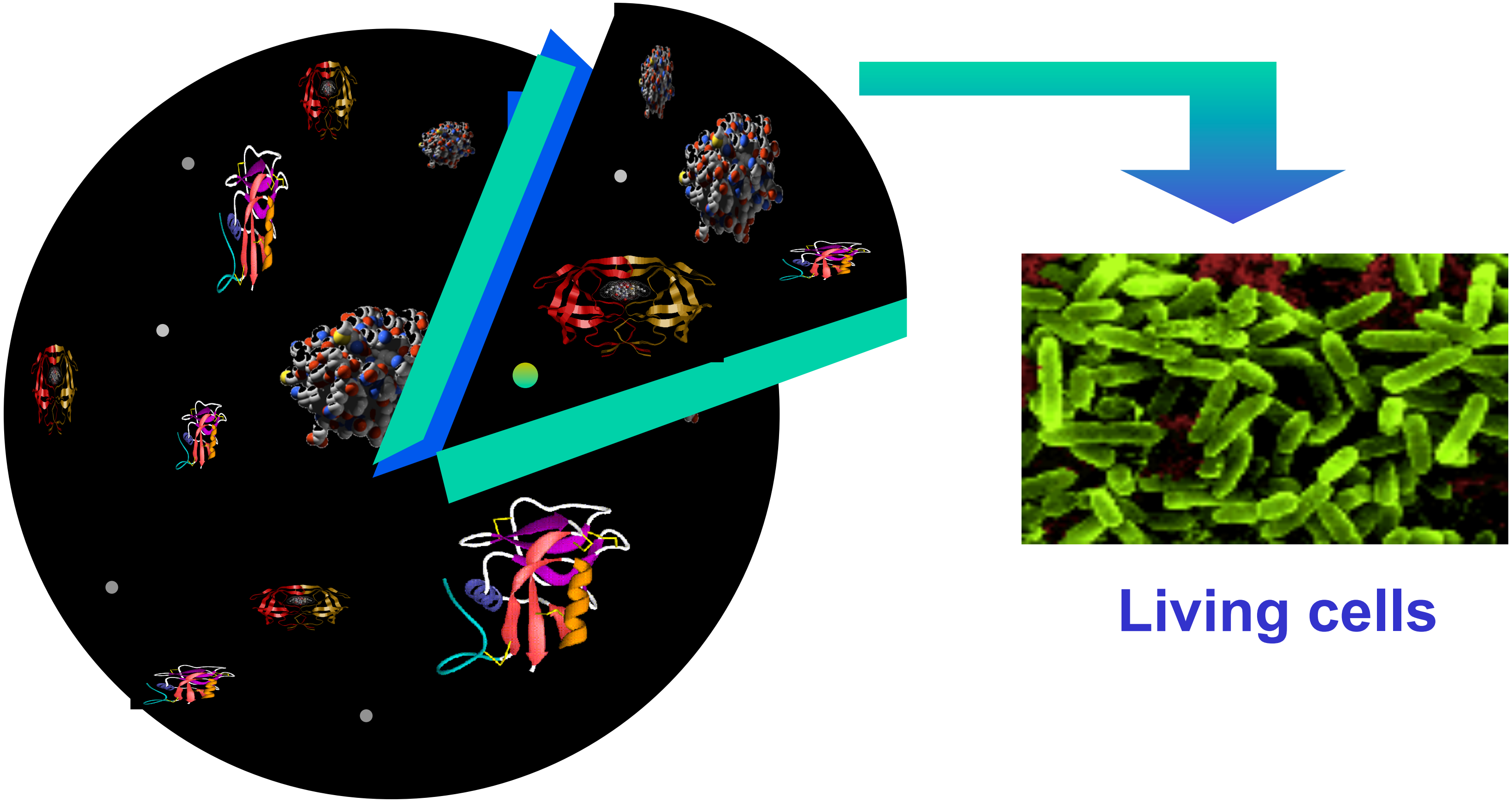
- Most of the past 100 years of biochemistry has focused on the analysis of small molecules (i.e. metabolism and metabolic pathways)
- These studies have revealed much about the processes and pathways for about 400 metabolites which can be summarized with this...



More Recently...

- Molecular biologists and biochemists have focused on the analysis of larger molecules (proteins and genes) which are much more complex and much more numerous
- These studies have primarily focused on identifying and cataloging these molecules (Human Genome Project)

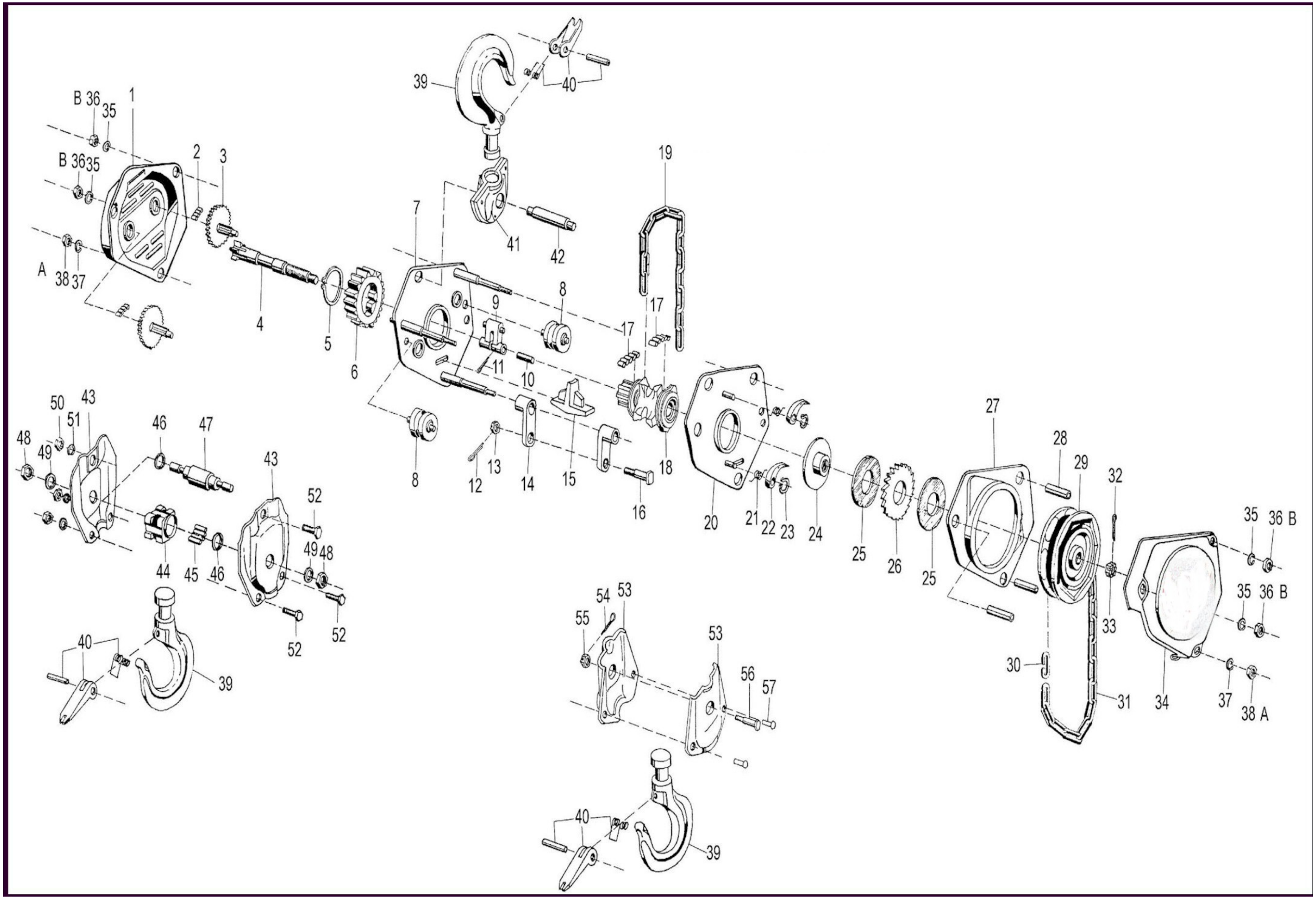
Nature's Parts Warehouse



The protein universe

Living cells

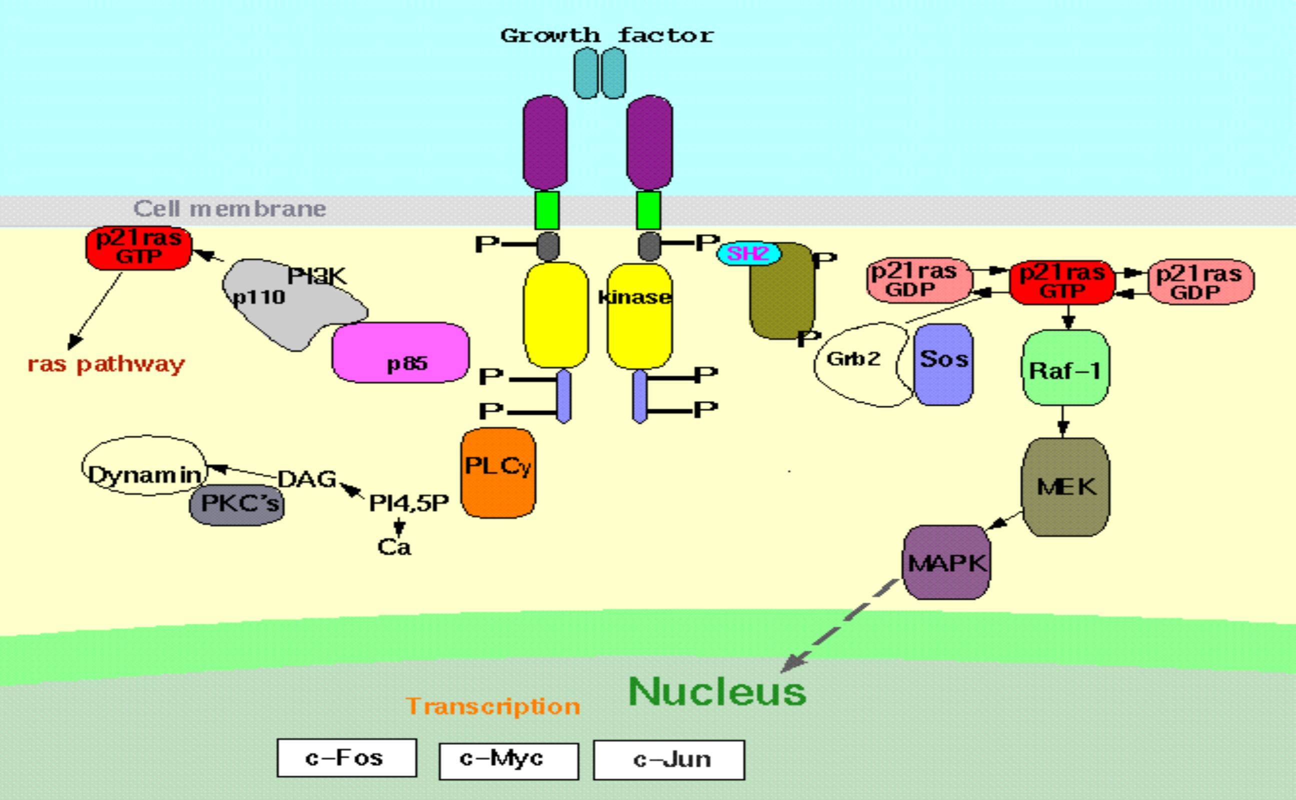
The Protein Parts List



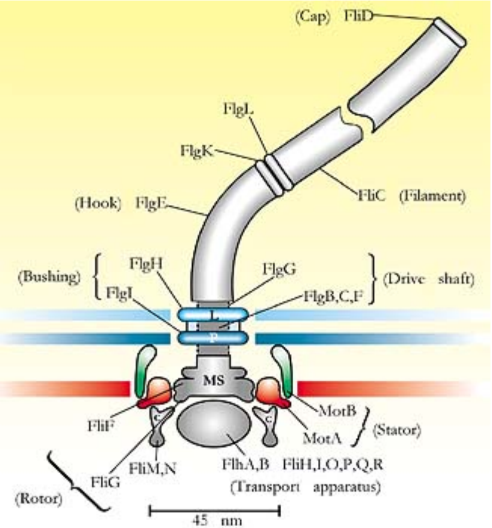
However...

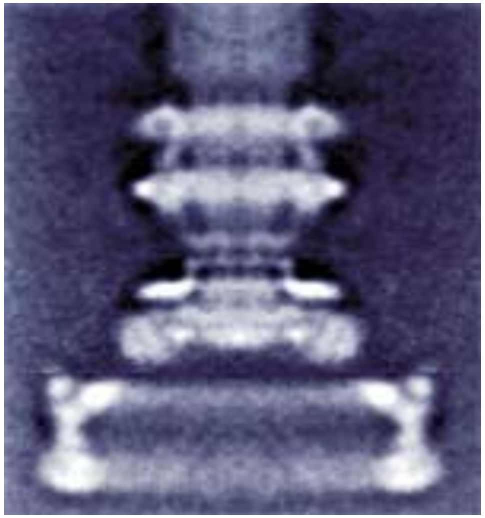
- This cataloging (which consumes most of bioinformatics) has been derogatively referred to as “stamp collecting”
- Having a collection of parts and names doesn’t tell you how to put something together or how things connect -- ***this is biology***

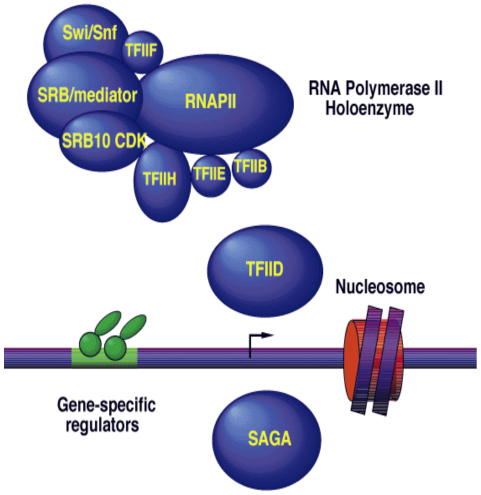
Remember: *Proteins Interact**

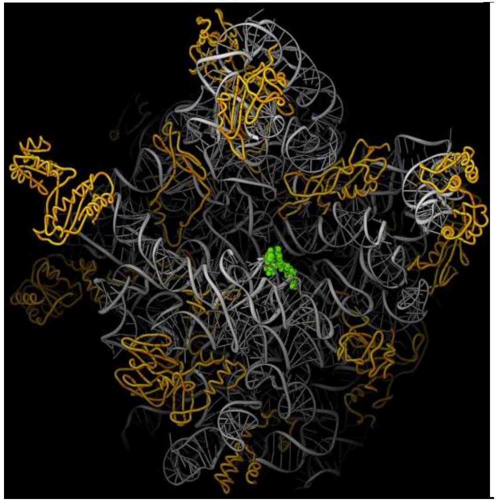


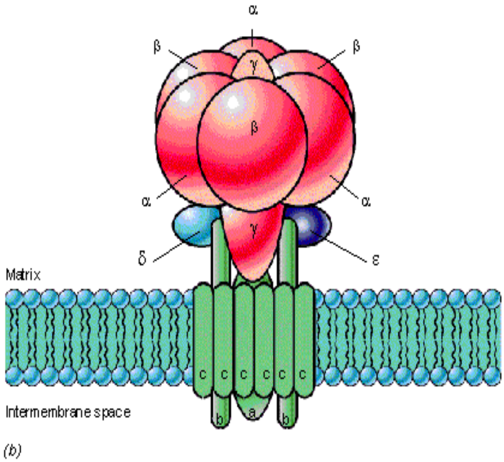
Proteins *Assemble**

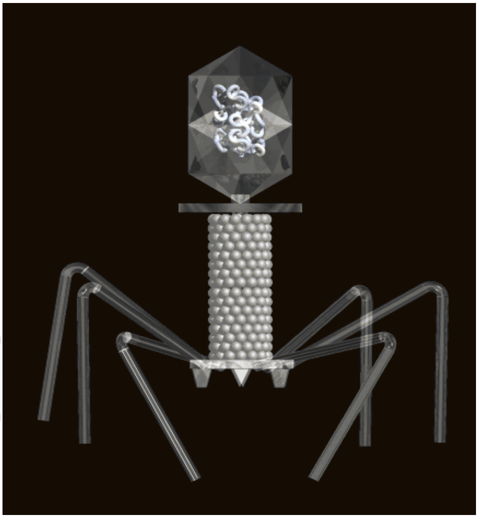










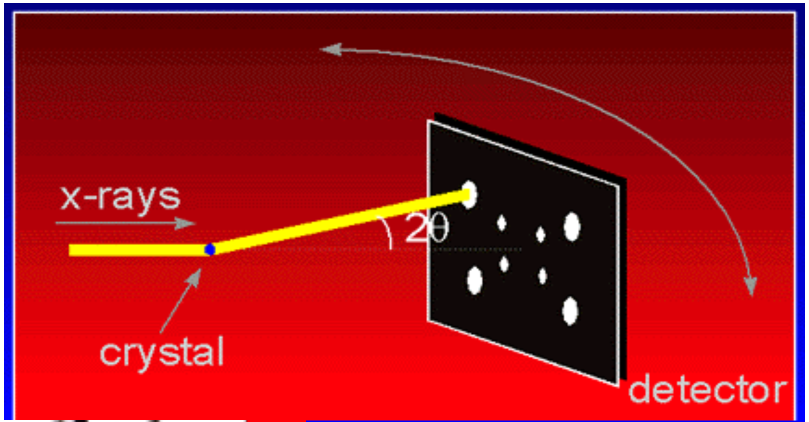


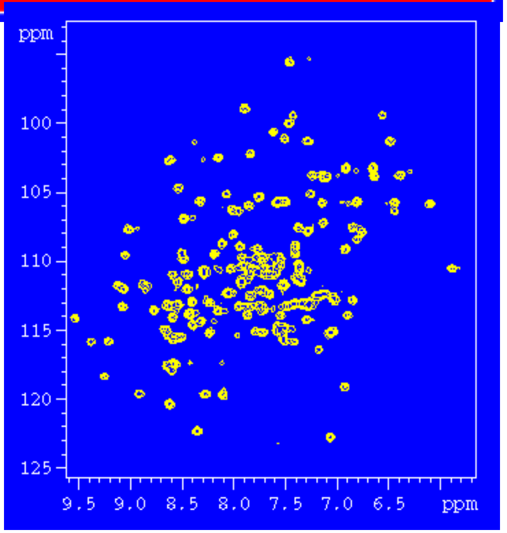
For the Past 10 Years...

- **Scientists have increasingly focused on "signal transduction" and transient protein interactions**
- **New techniques have been developed which reveal which proteins and which parts of proteins are important for interaction**
- **The hope is to get something like this..**

Protein Interaction Tools and Techniques - Experimental Methods

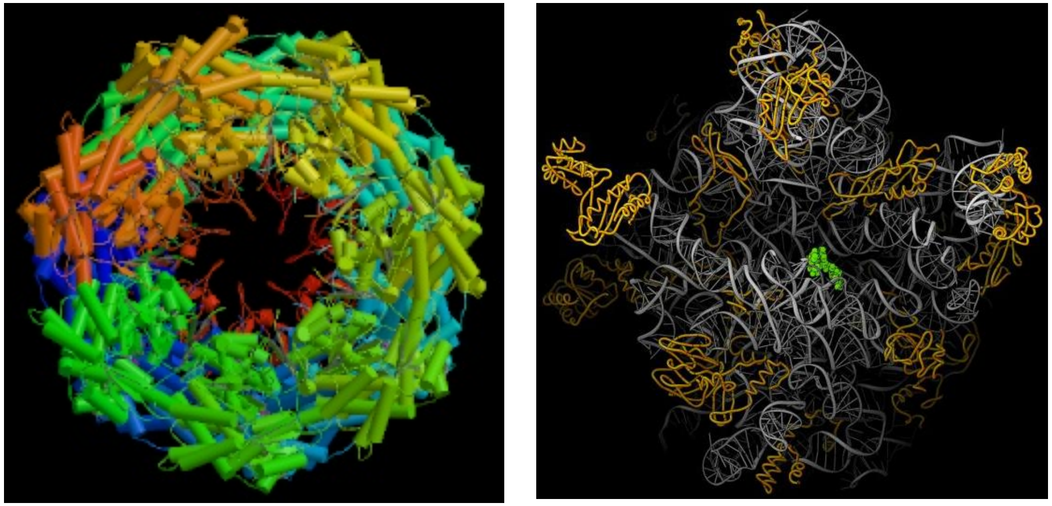
3D Structure Determination*



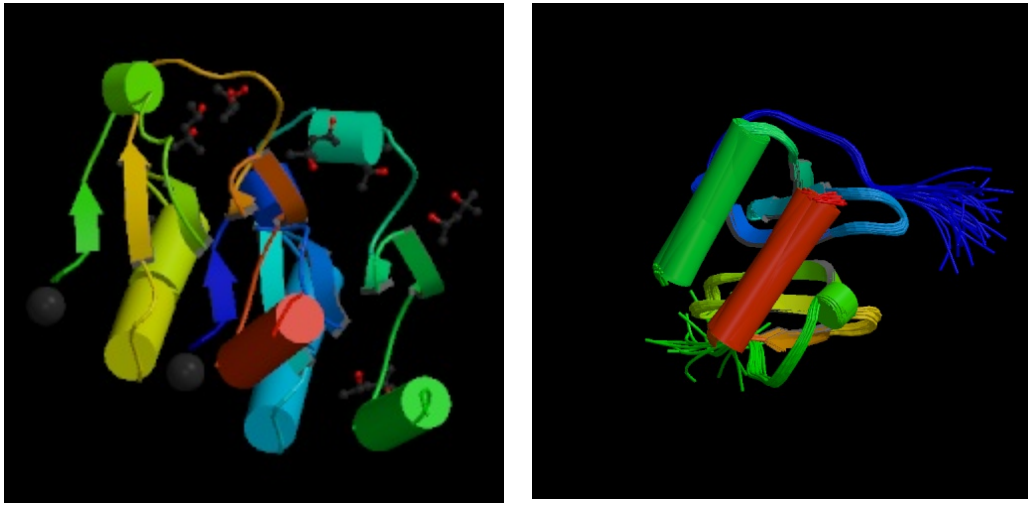


- X-ray crystallography
  - grow crystal
  - collect diffract. data
  - calculate e- density
  - trace chain
- NMR spectroscopy
  - label protein
  - collect NMR spectra
  - assign spectra & NOEs
  - calculate structure using distance geom.

Quaternary Structure

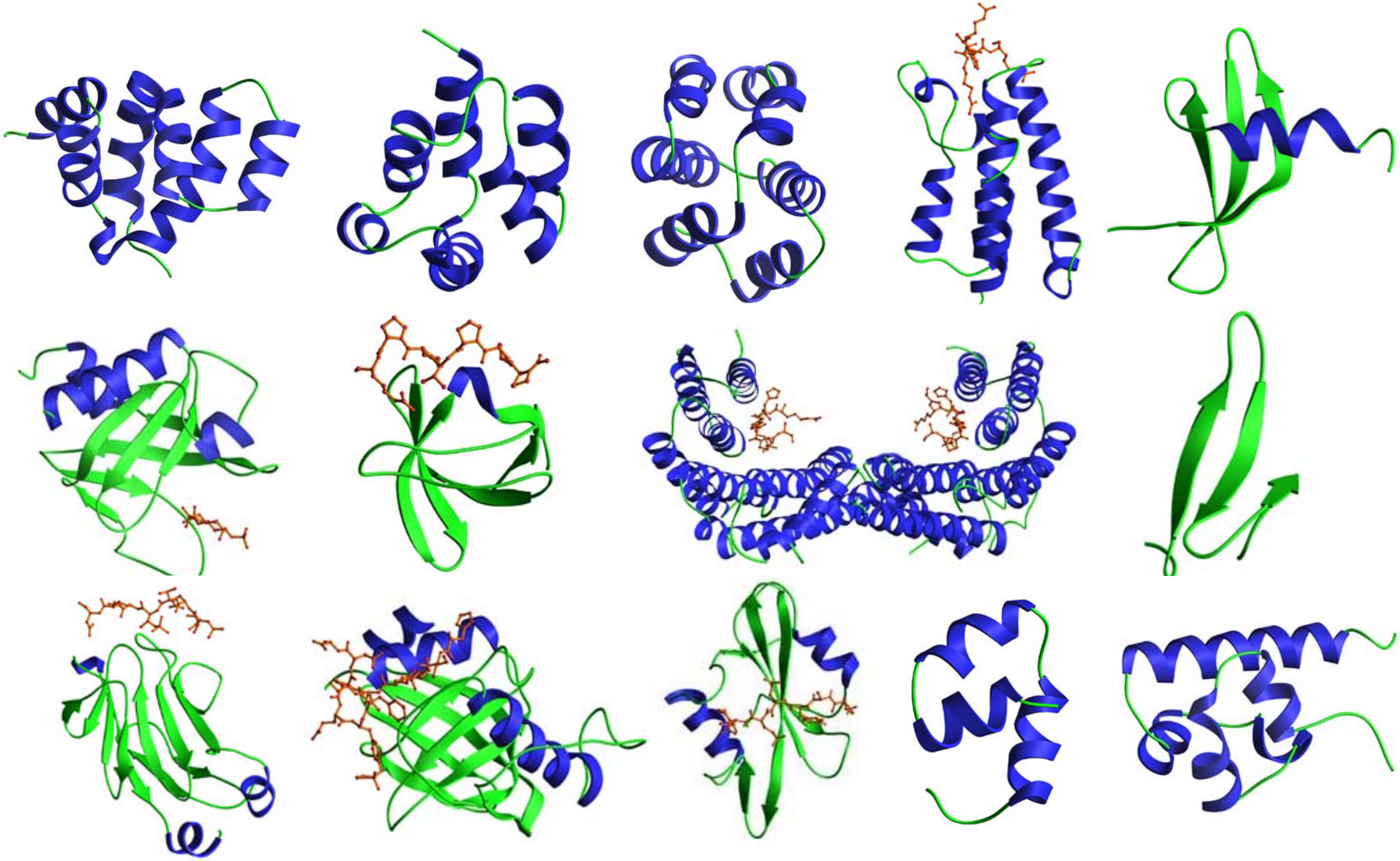


Some interactions are real



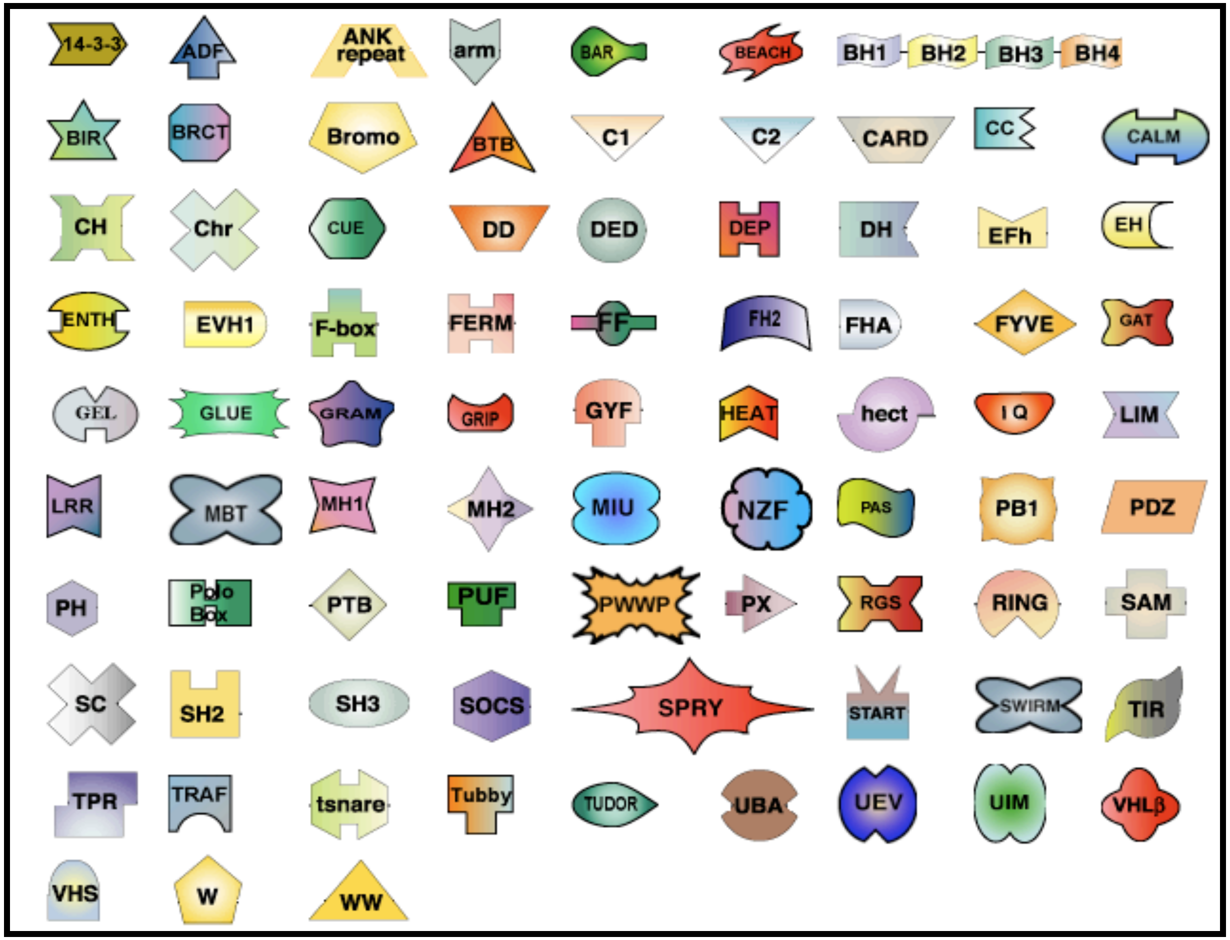
Others are not

Protein Interaction Domains*



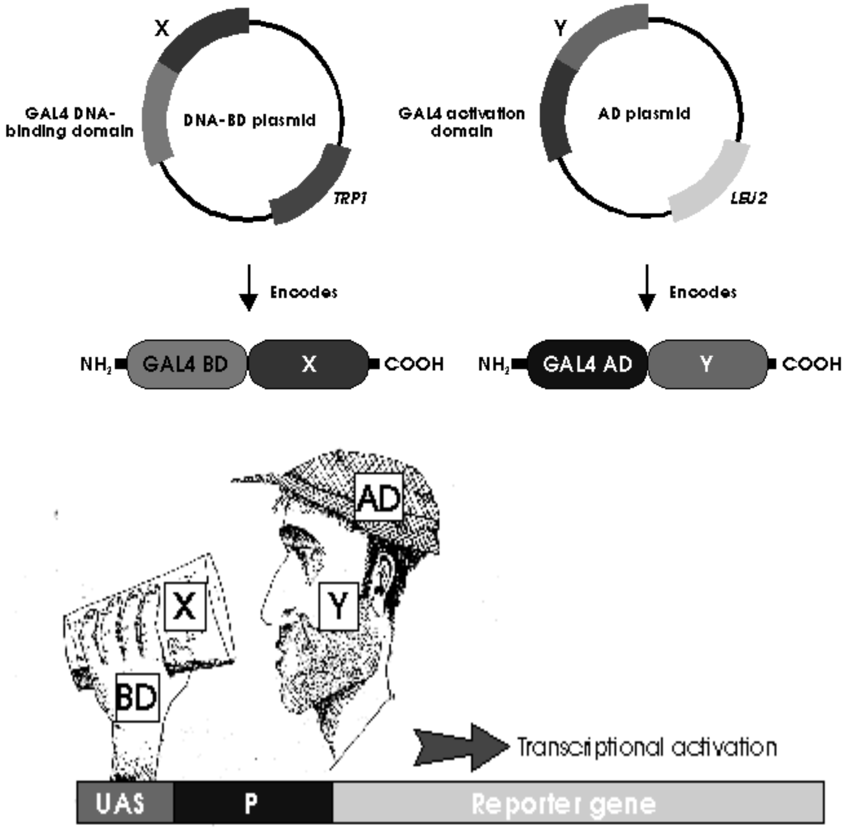
http://pawsonlab.mshri.on.ca/ *82 domains*

Protein Interaction Domains



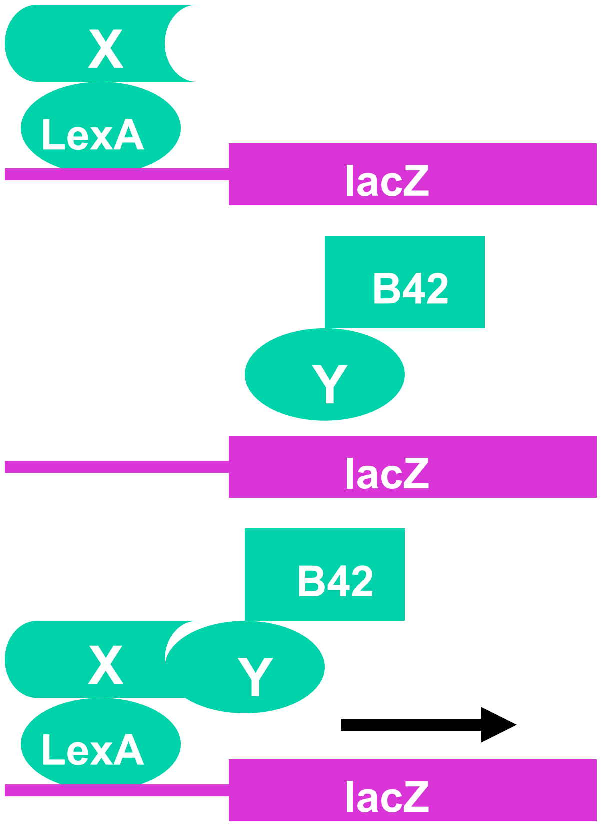
http://pawsonlab.mshri.on.ca/

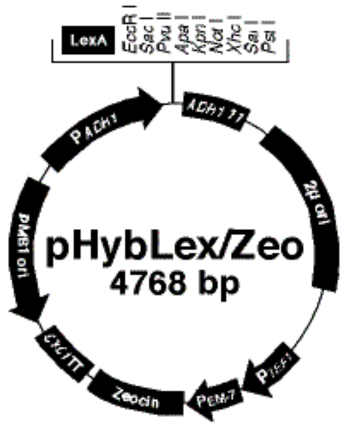
Yeast Two-Hybrid Analysis*

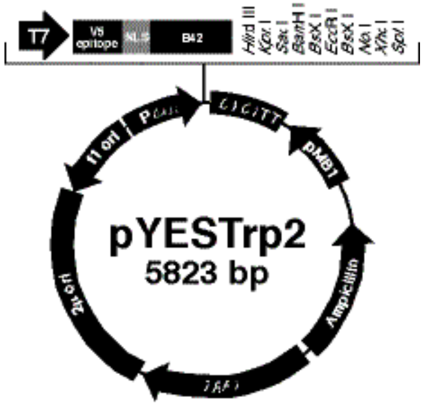


- Yeast two-hybrid experiments yield information on protein protein interactions
- GAL4 Binding Domain
- GAL4 Activation Domain
- X and Y are two proteins of interest
- If X & Y interact then reporter gene is expressed

Invitrogen Yeast 2-Hybrid







Example of 2-Hybrid Analysis*

- Uetz P. et al., "*A Comprehensive Analysis of Protein-Protein Interactions in Saccharomyces cerevisiae*" Nature 403:623-627 (2000)
- High Throughput Yeast 2 Hybrid Analysis
- 957 putative interactions
- 1004 of 6000 predicted proteins involved

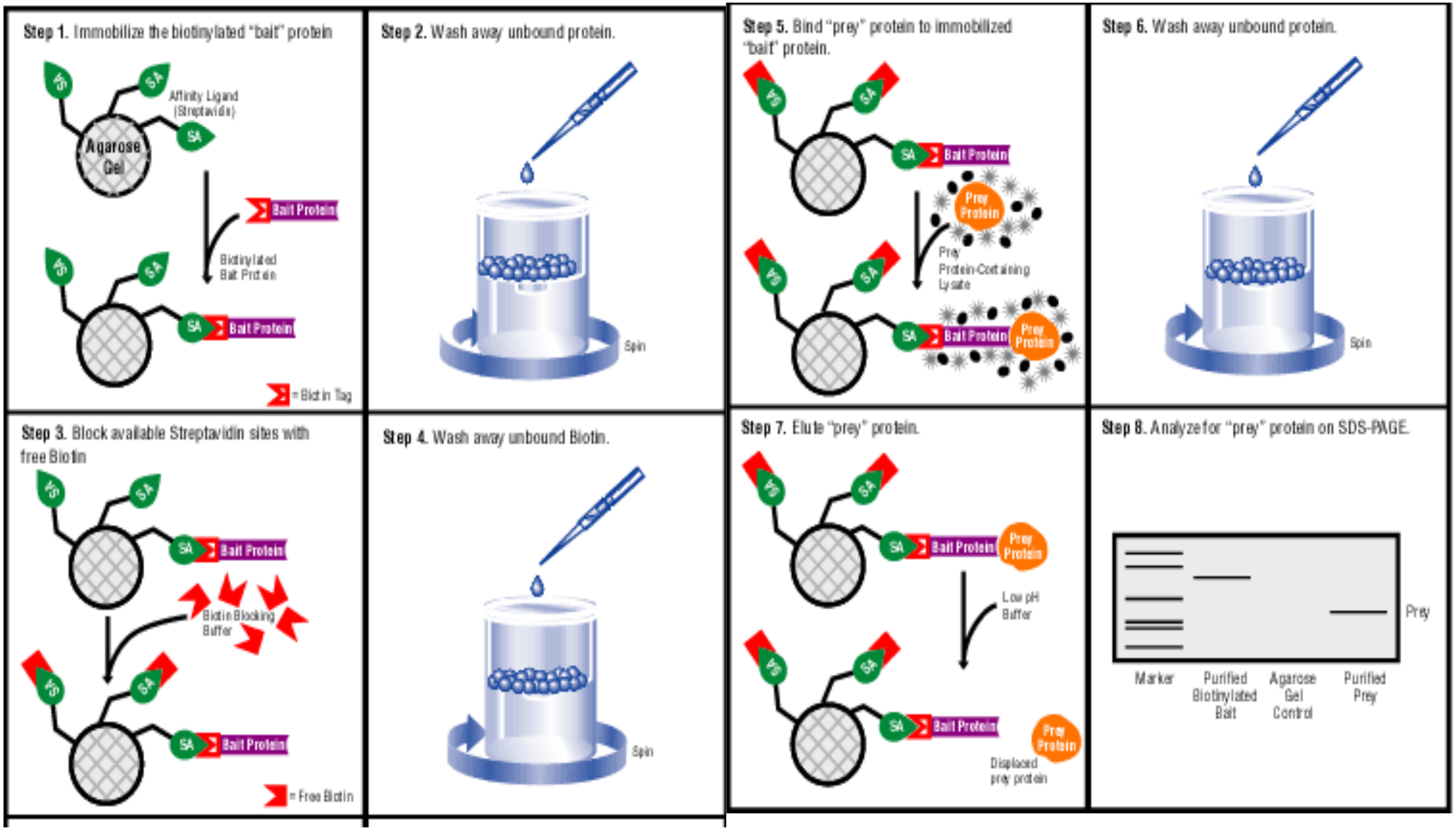
Example of 2-Hybrid Analysis

- Rain JC. et al., “*The protein-protein interaction map of Helicobacter pylori*” Nature 409:211-215 (2001)
- High Throughput Yeast 2 Hybrid Analysis
- 261 H. pylori proteins scanned against genome
- >1200 putative interactions identified
- Connects >45% of the H. pylori proteome

Another Way?*

- Ho Y, Gruhler A, et al. *Systematic identification of protein complexes in Saccharomyces cerevisiae by mass spectrometry*. Nature 415:180-183 (2002)
- High Throughput Mass Spectral Protein Complex Identification (HMS-PCI)
- 10% of yeast proteins used as “bait”
- 3617 associated proteins identified
- 3 fold higher sensitivity than yeast 2-hybrid

Affinity Pull-down*



Step 1. Immobilize the biotinylated "bait" protein

Step 2. Wash away unbound protein.

Step 3. Block available Streptavidin sites with free Biotin

Step 4. Wash away unbound Biotin.

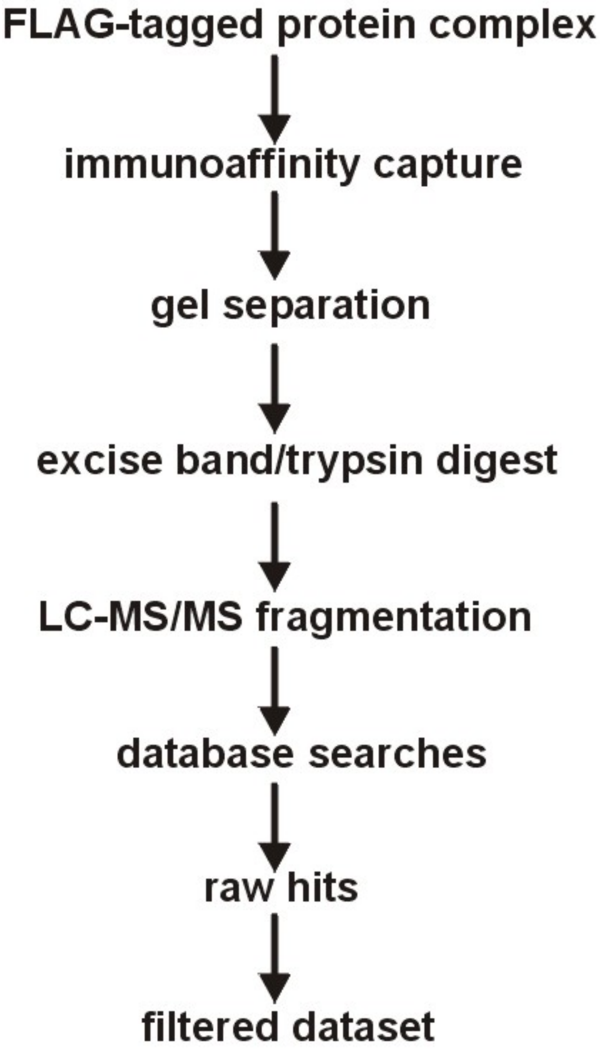
Step 5. Bind "prey" protein to immobilized "bait" protein.

Step 6. Wash away unbound protein.

Step 7. Elute "prey" protein.

Step 8. Analyze for "prey" protein on SDS-PAGE.

HMS-PCI*



FLAG-tagged protein complex

immunoaffinity capture

gel separation

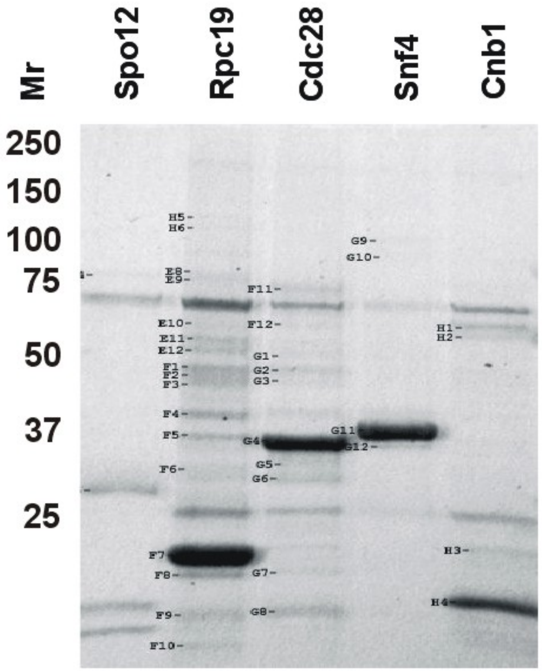
excise band/trypsin digest

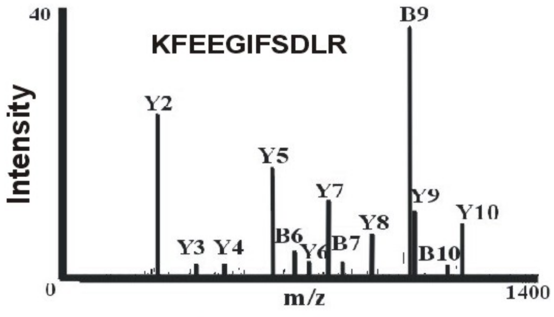
LC-MS/MS fragmentation

database searches

raw hits

filtered dataset





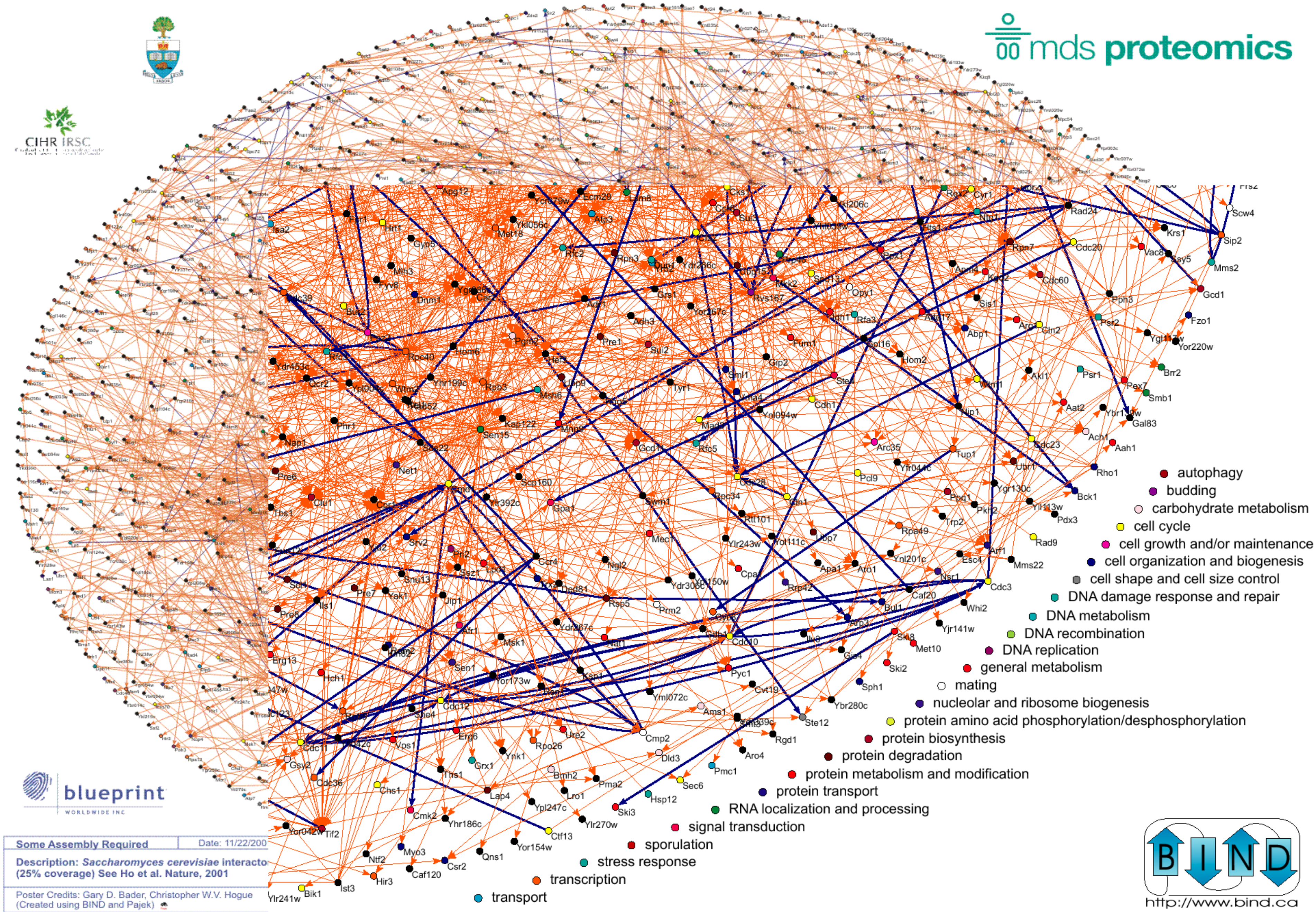
ms/ms fragmentation pattern

MOUNT
SINAI
HOSPITAL

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

mds proteomics

CIHR IRSC



● autophagy
● budding
○ carbohydrate metabolism
○ cell cycle
● cell growth and/or maintenance
● cell organization and biogenesis
● cell shape and cell size control
● DNA damage response and repair
● DNA metabolism
● DNA recombination
● DNA replication
● general metabolism
○ mating
● nucleolar and ribosome biogenesis
○ protein amino acid phosphorylation/desphosphorylation
● protein biosynthesis
● protein degradation
● protein metabolism and modification
● protein transport
● RNA localization and processing
● signal transduction
● sporulation
● stress response
● transcription
● transport

blueprint
WORLDWIDE INC

Some Assembly Required	Date: 11/22/200

Description: *Saccharomyces cerevisiae* interacto (25% coverage) See Ho et al. Nature, 2001

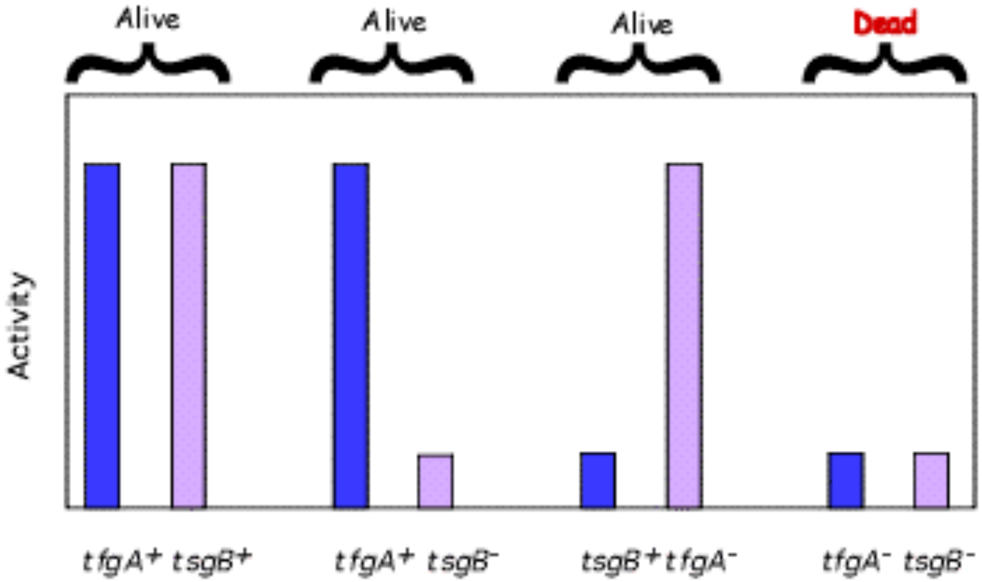
Poster Credits: Gary D. Bader, Christopher W.V. Hogue
(Created using BIND and Pajek)

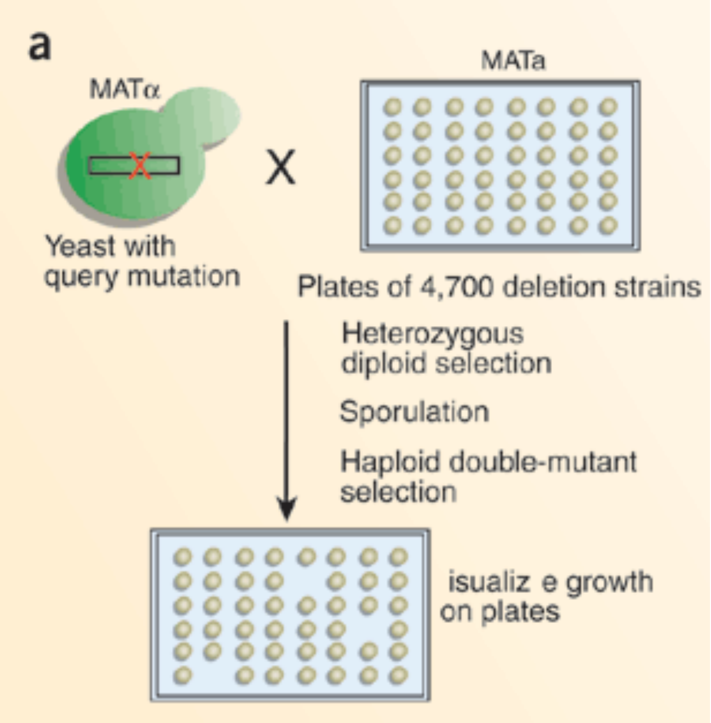
BIND
http://www.bind.ca

Synthetic Genetic Interactions*

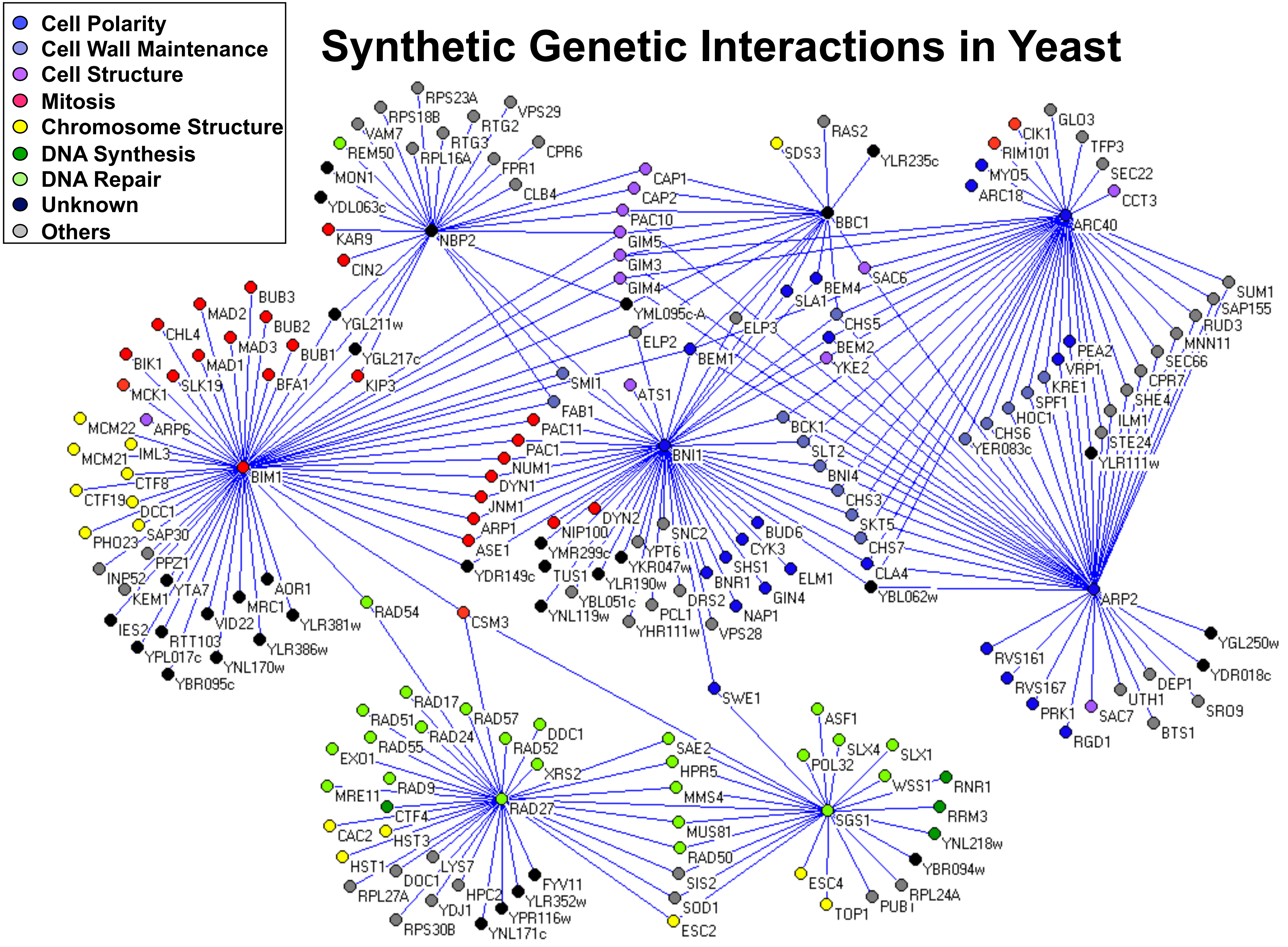
- Two mutations are synthetically lethal if cells with either of the single mutations are viable but cells with both mutations are non-viable
- Two types of synthetic lethal genetic interactions (lethal, slow growth)
- Mate two mutants without phenotypes to get a daughter cell with a phenotype
- Genetic interactions provide functional data on protein interactions or redundant genes
- About 23% of known SLs (1295 - YPD+MIPS) are known protein interactions in yeast

Synthetic Lethality*



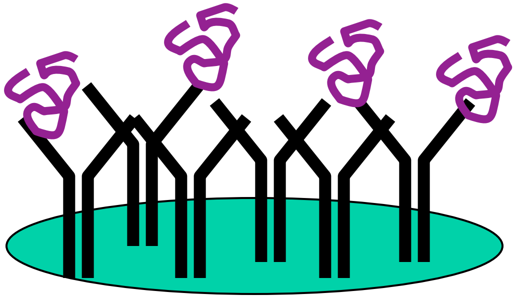


Synthetic Genetic Interactions in Yeast

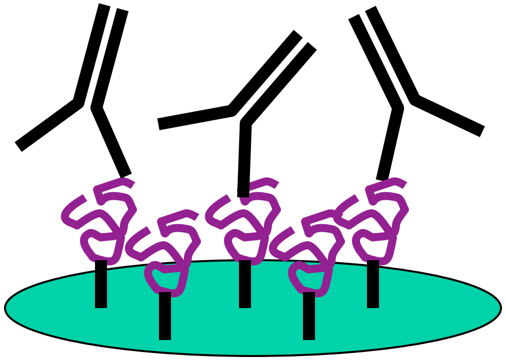


Protein Chips*

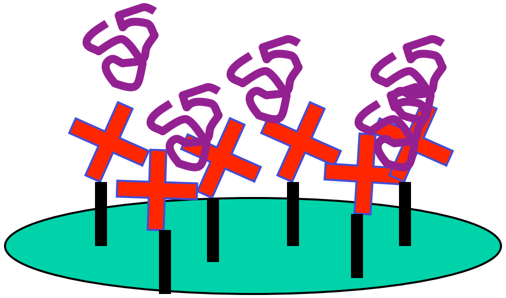
Antibody Array



Antigen Array



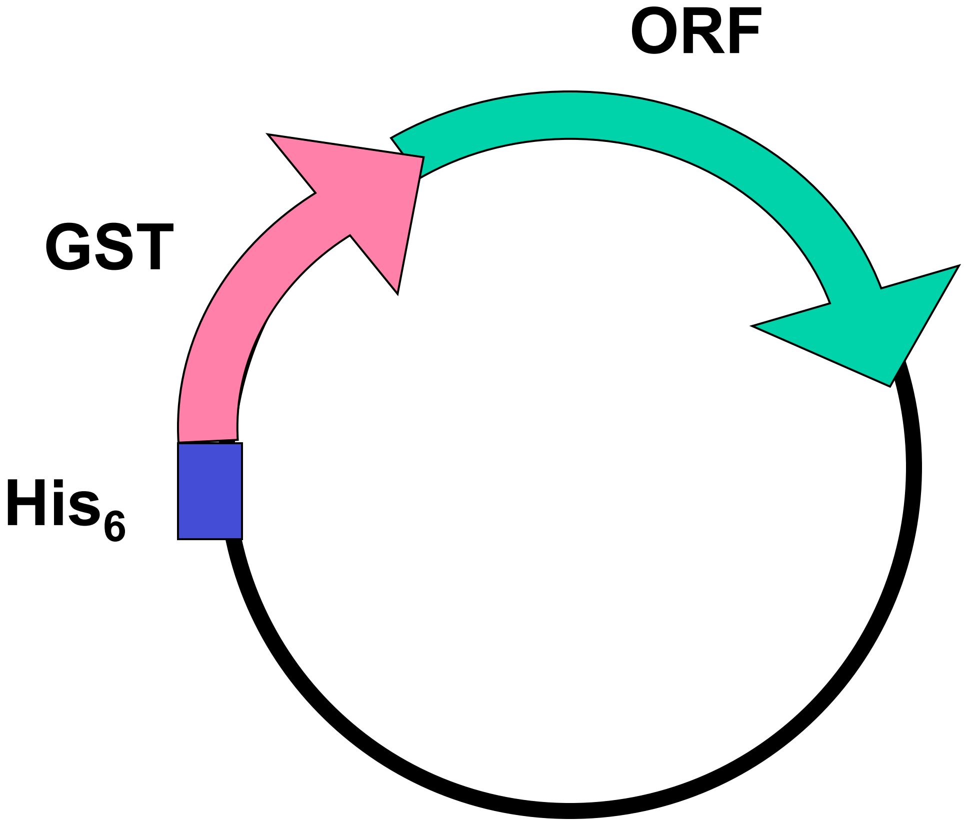
Ligand Array

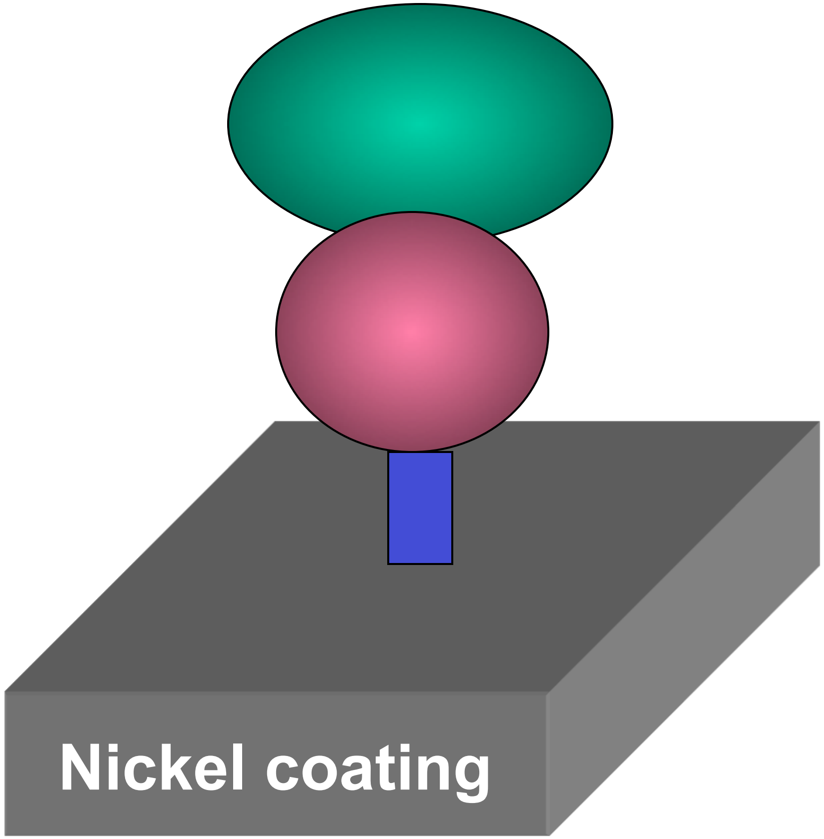


Detection by: SELDI MS, fluorescence, SPR, electrochemical, radioactivity, microcantelever

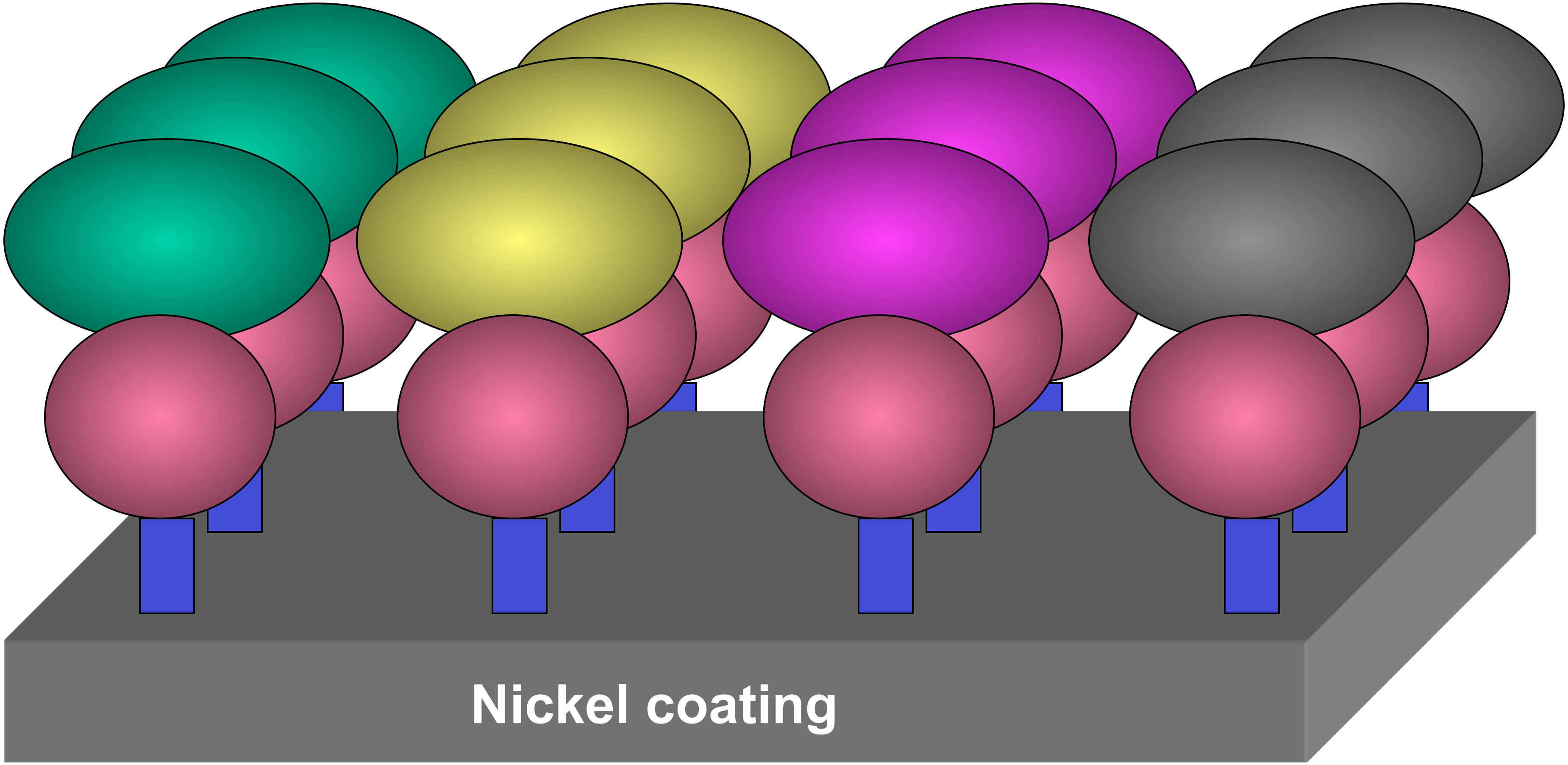
Protein (Antigen) Chips

H Zhu, J Klemic, S Chang, P Bertone, A Casamayor, K Klemic, D Smith, M Gerstein, M Reed, & M Snyder (2000).**Analysis of yeast protein kinases using protein chips**. Nature Genetics 26: 283-289

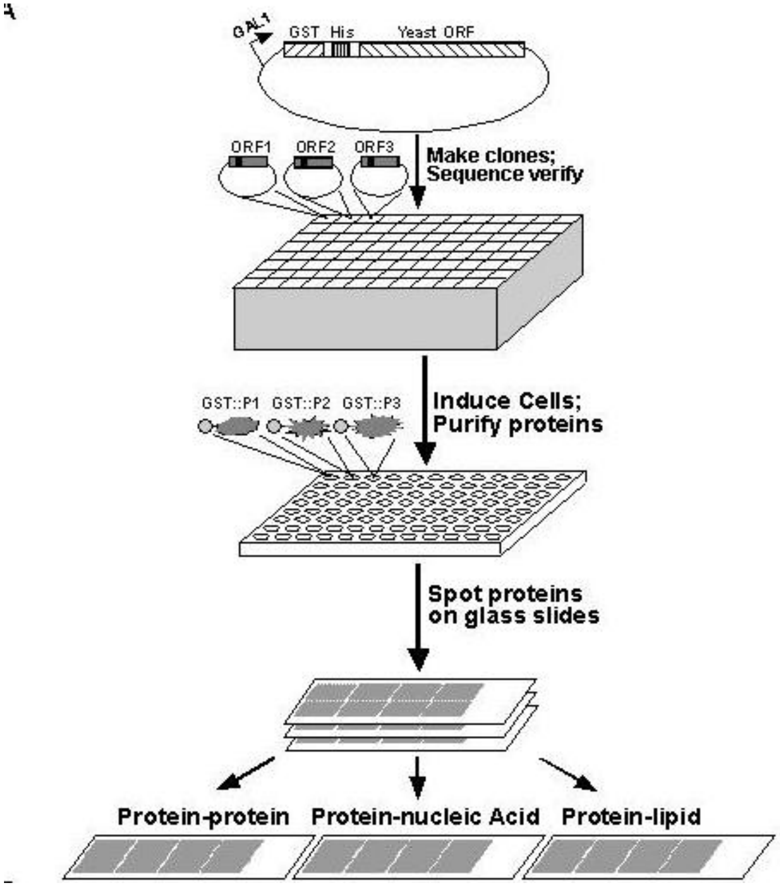




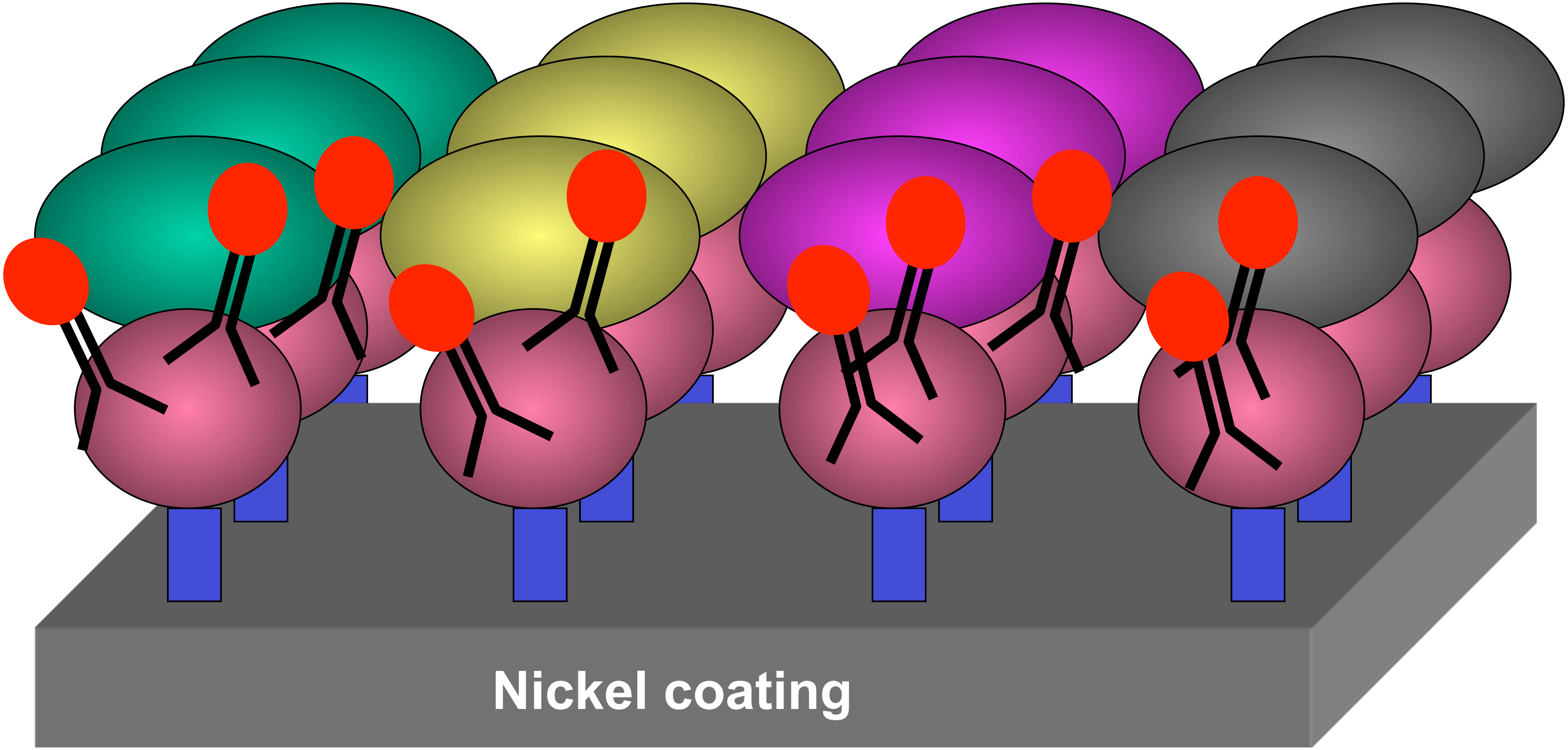
Protein (Antigen) Chips



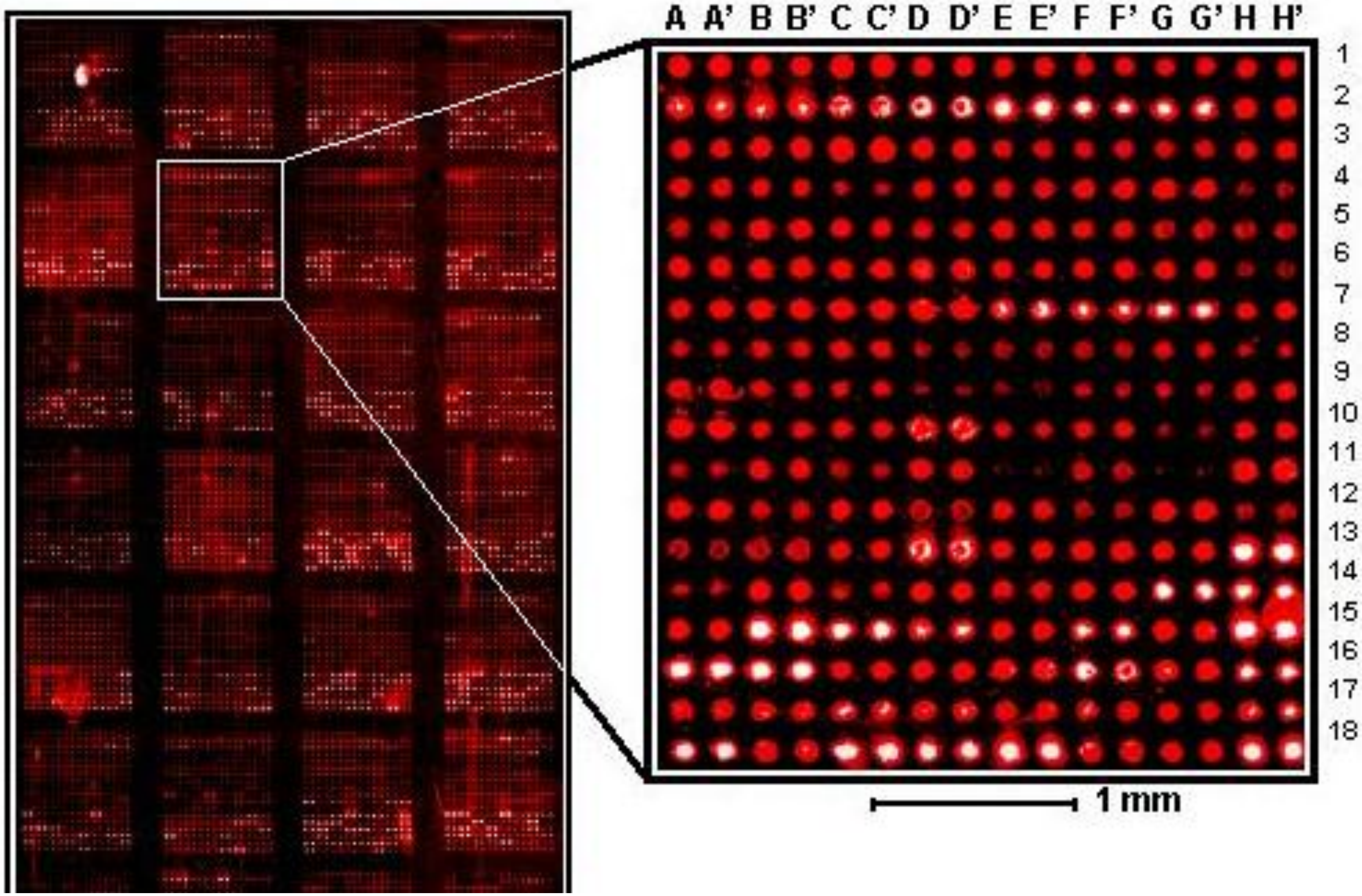
Arraying Process



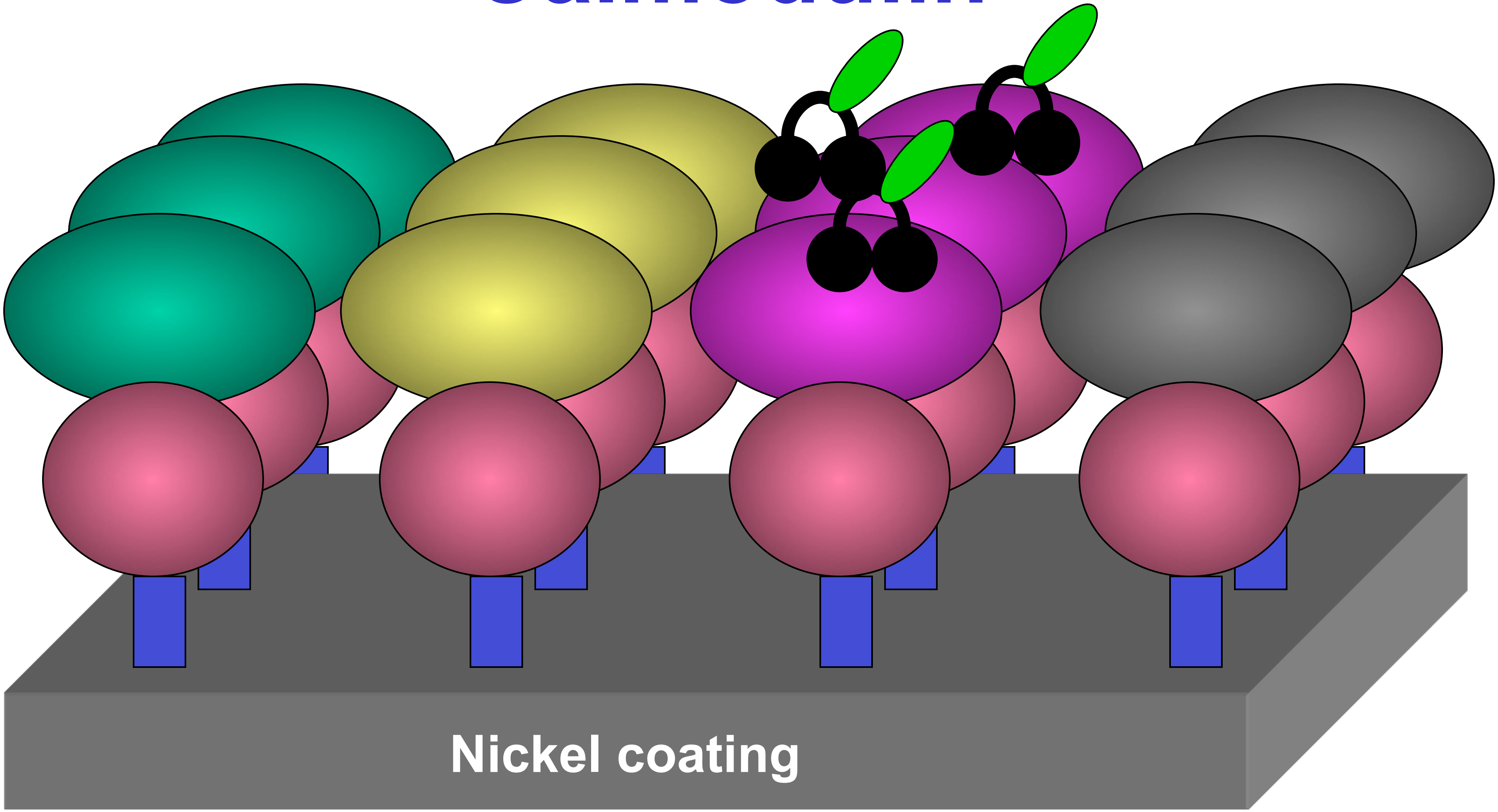
Probe with anti-GST Mab



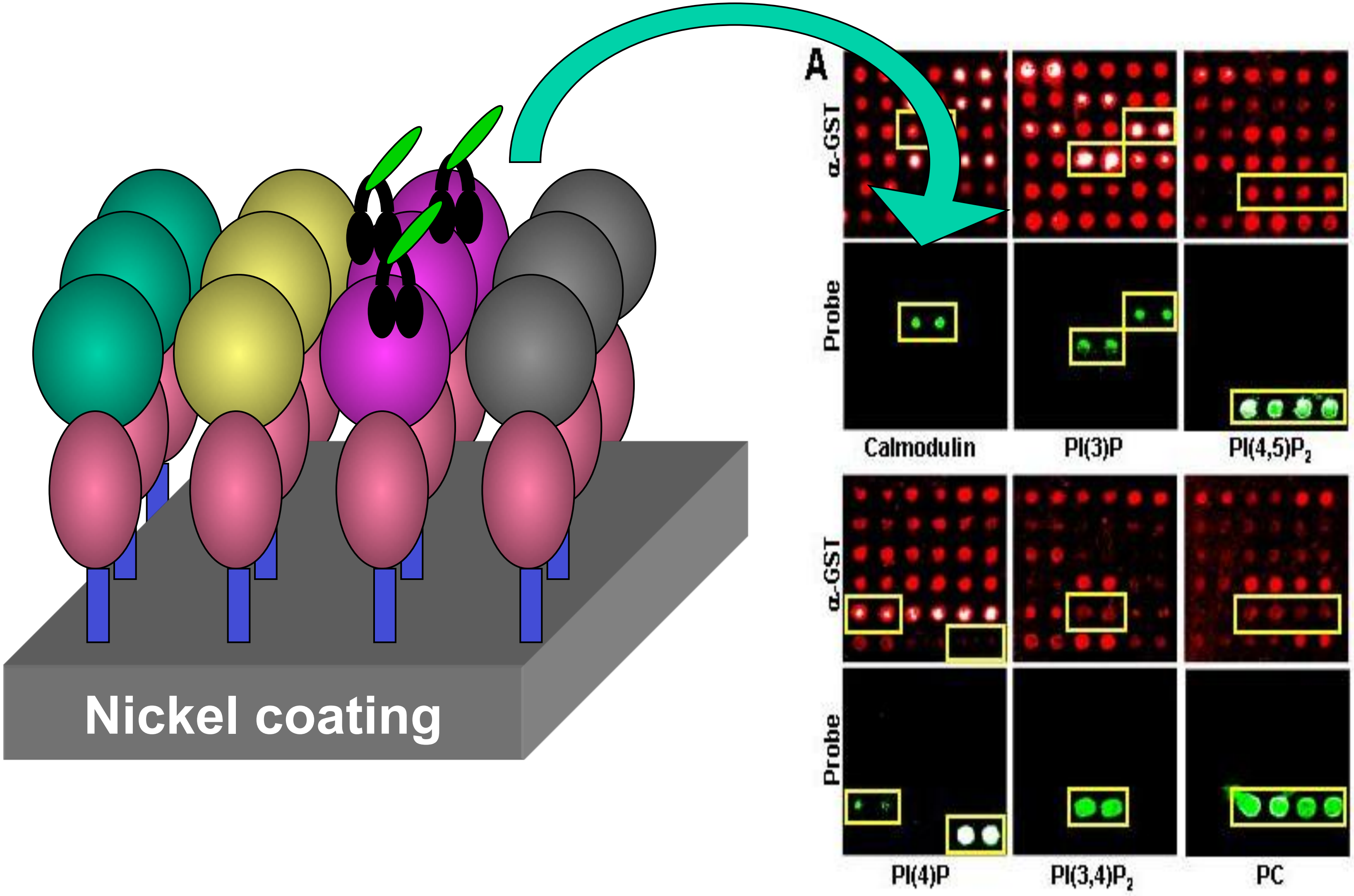
Anti-GST Probe



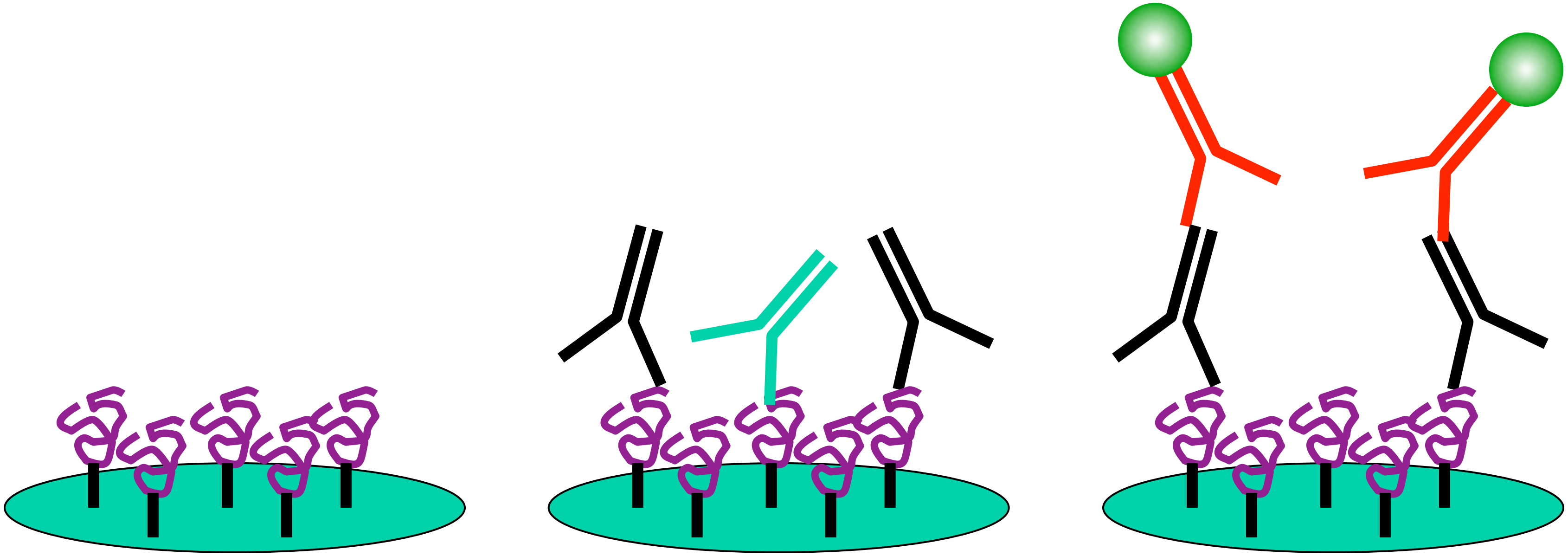
Probe with Cy3-labeled Calmodulin



“Functional” Protein Array*

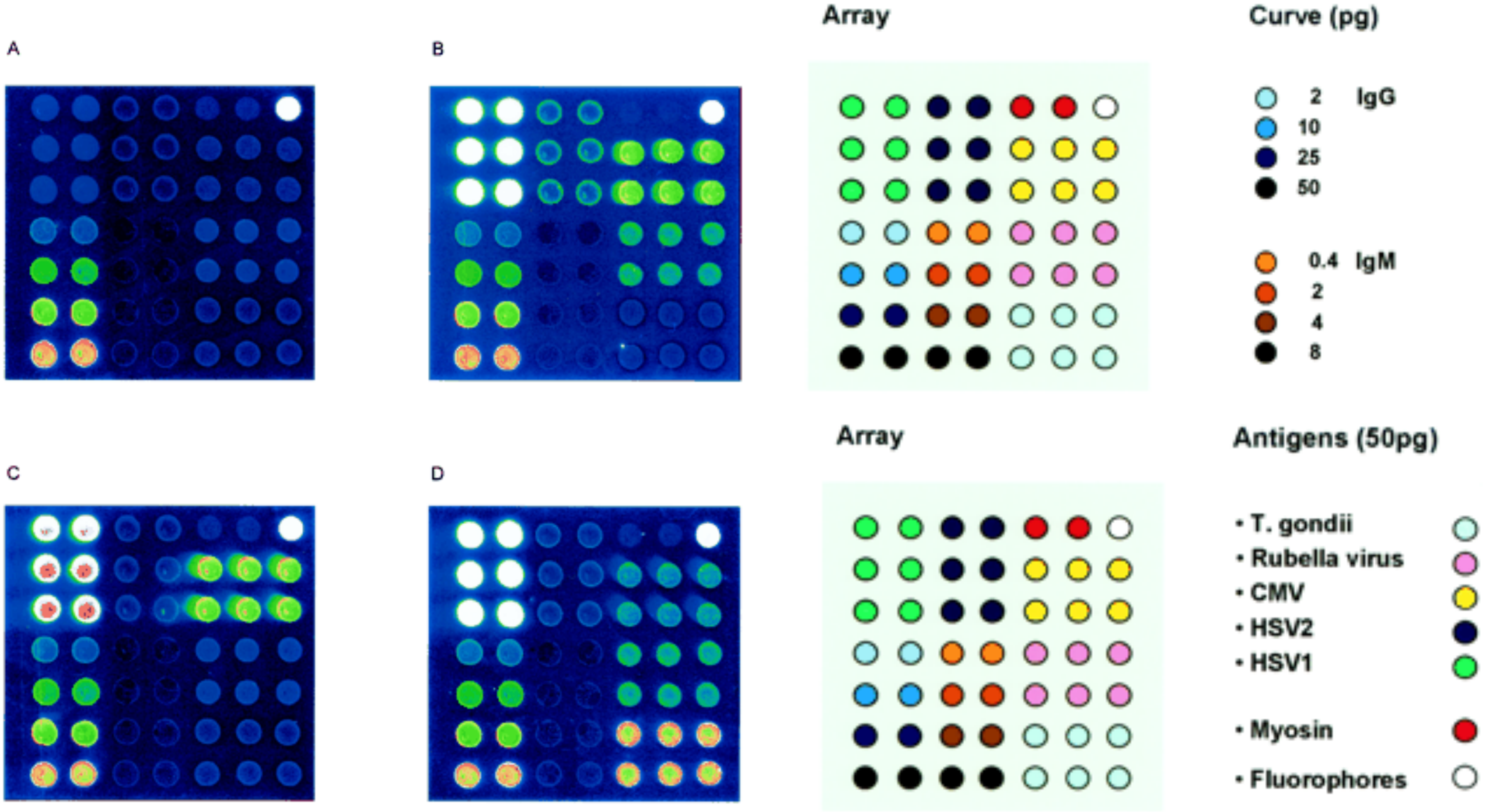


Antigen Array (ELISA Chip)*



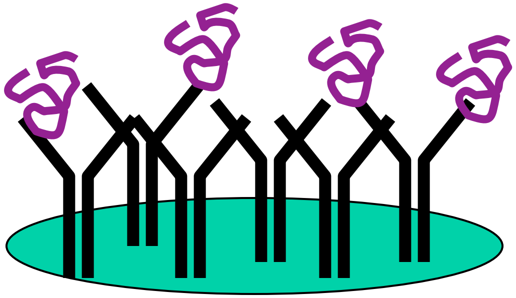
Mezzasoma et al. Clinical Chem. 48:121 (2002)

Diagnostic Antigen Array

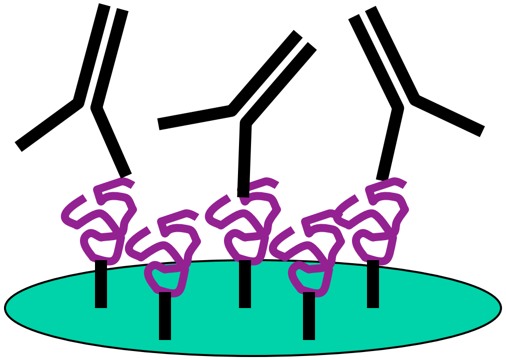


Protein Chips

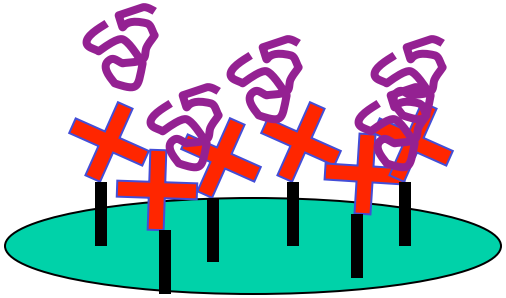
Antibody Array



Antigen Array



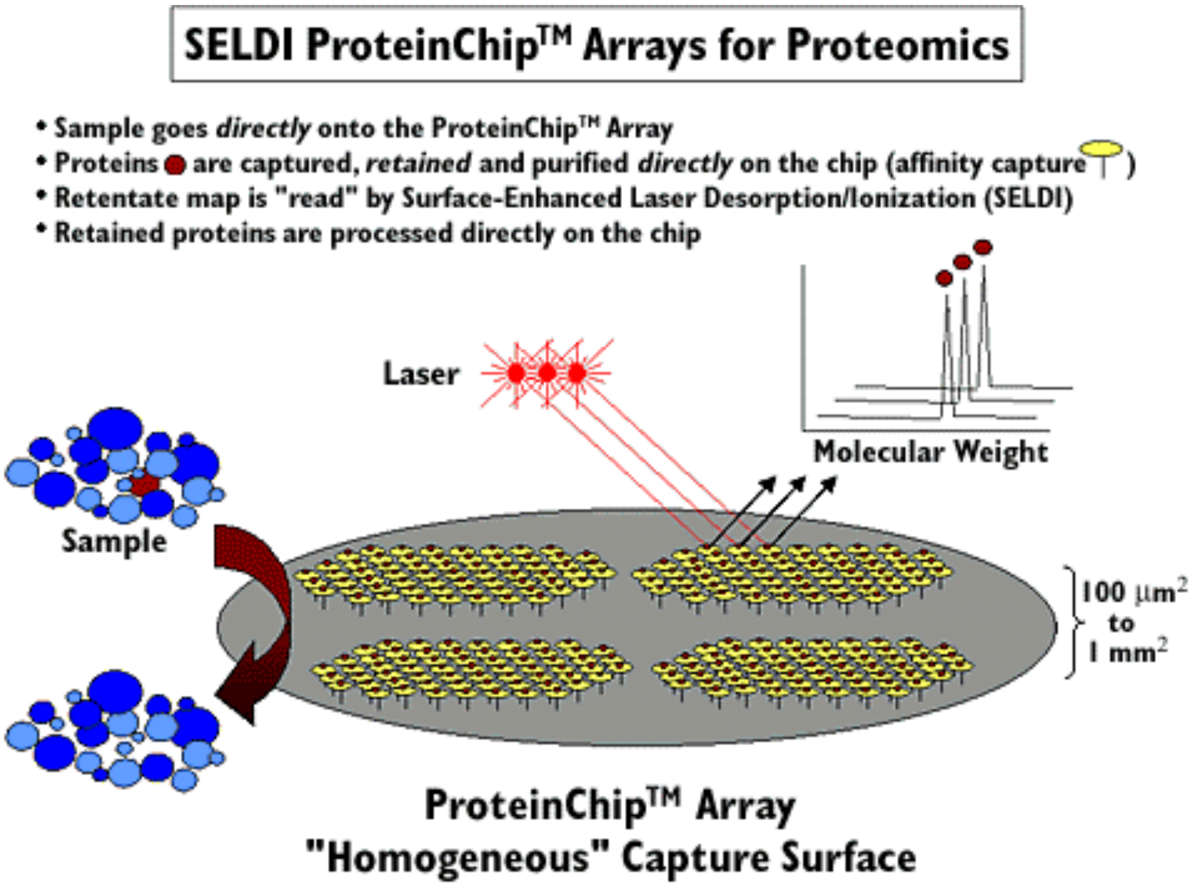
Ligand Array



Ciphergen “Ligand” Chips*

- Hydrophobic (C_8) Arrays
- Hydrophilic (SiO_2) Arrays
- Anion exchange Arrays
- Cation exchange Arrays
- Immobilized Metal Affinity (NTA-nitroloacetic acid) Arrays
- Epoxy Surface (amine and thiol binding) Arrays

Ciphergen (BioRad) ProteinChip*

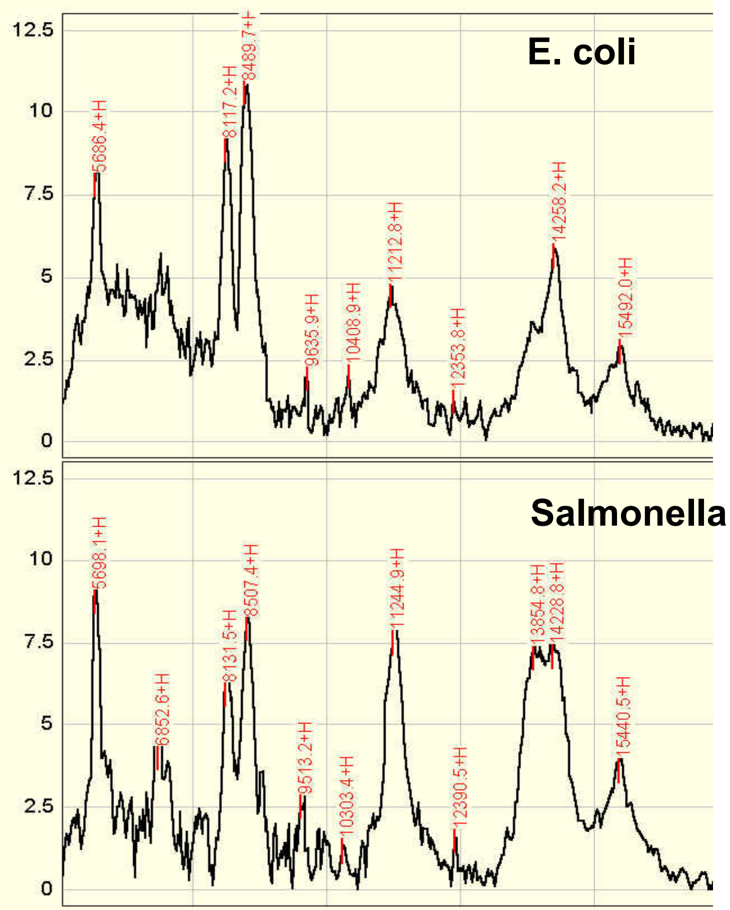


SELDI ProteinChip^{TM} Arrays for Proteomics

- Sample goes *directly* onto the ProteinChip^{TM} Array
- Proteins ● are captured, *retained* and purified *directly* on the chip (affinity capture)
- Retentate map is "read" by Surface-Enhanced Laser Desorption/Ionization (SELDI)
- Retained proteins are processed directly on the chip

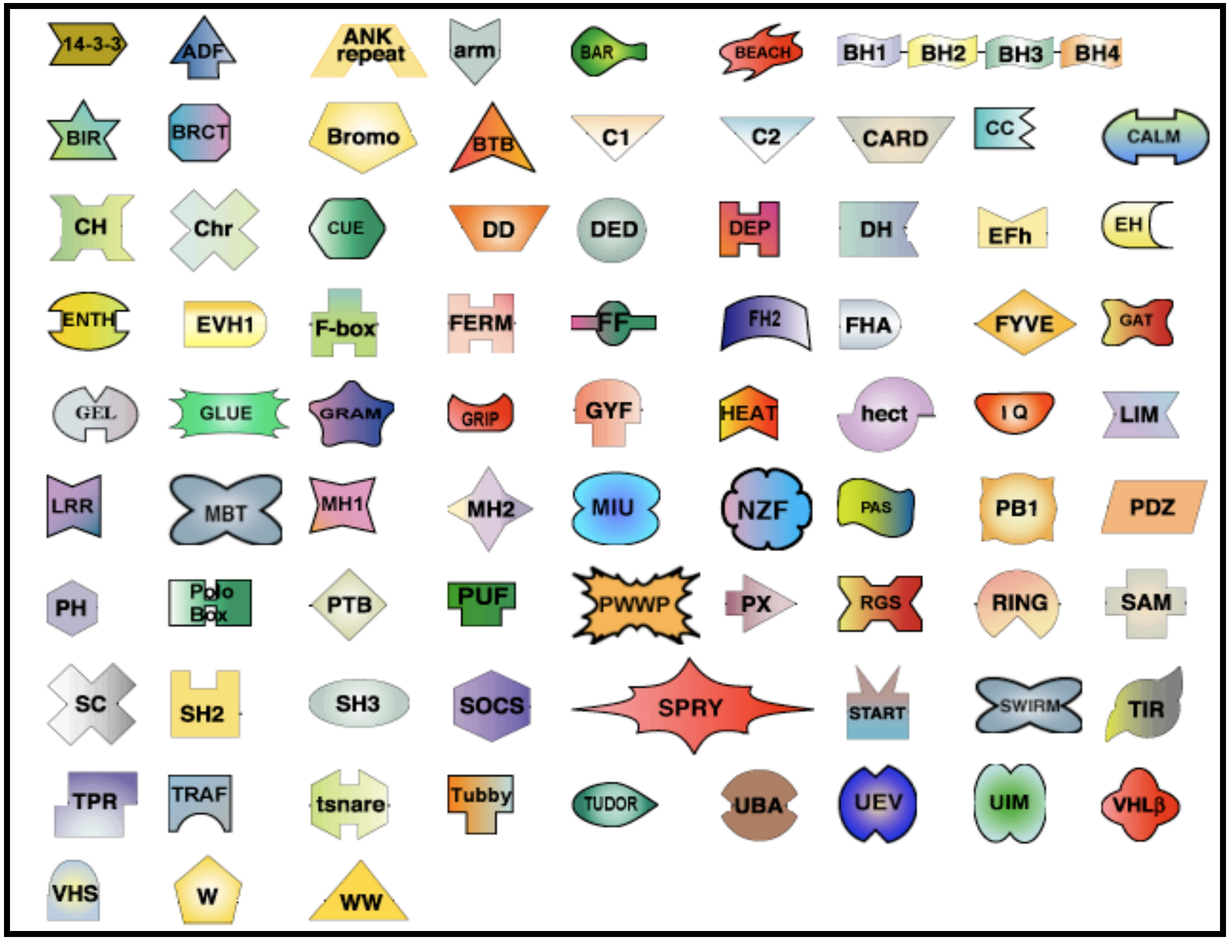
ProteinChip^{TM} Array
"Homogeneous" Capture Surface

Peptide/Protein Profile



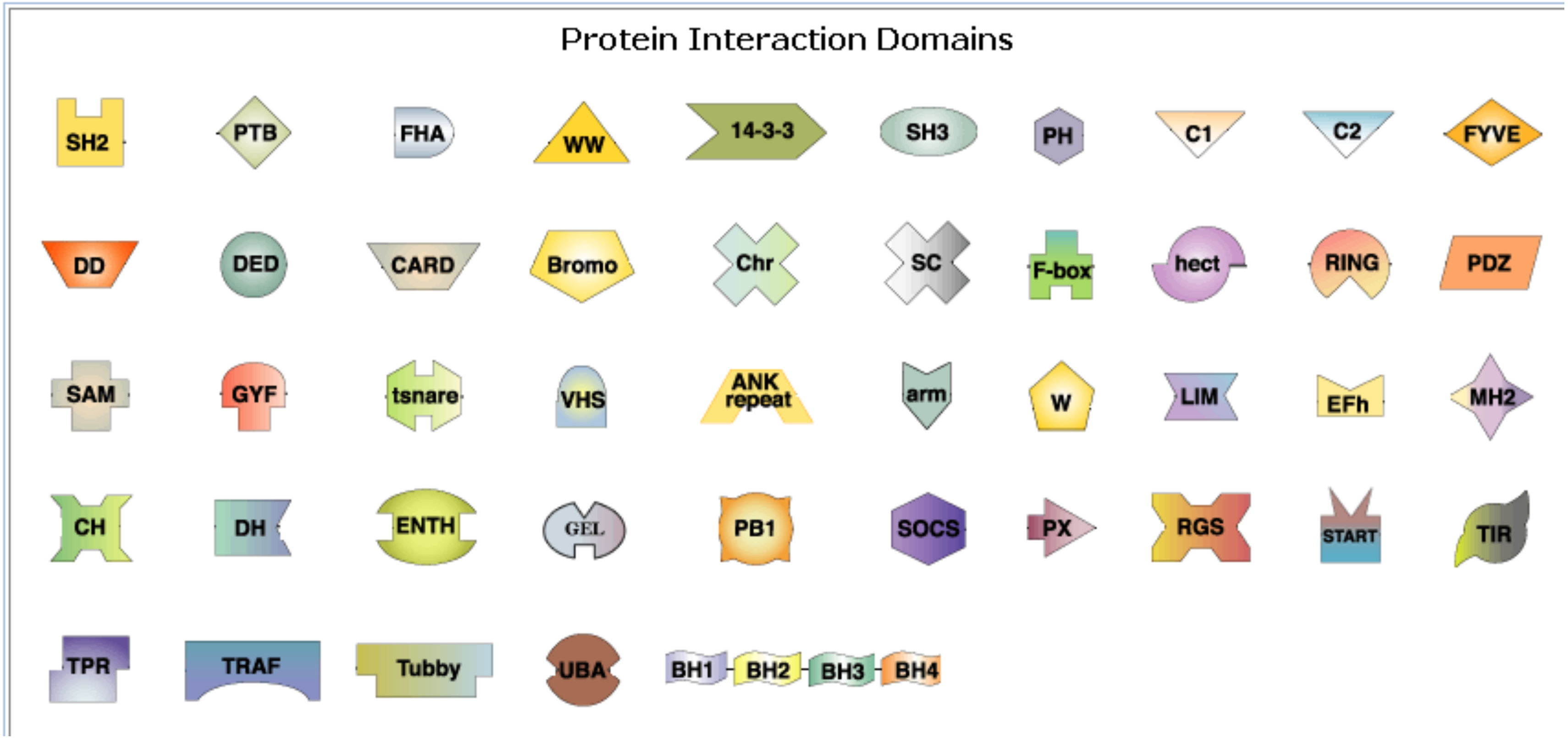
Protein Interaction Tools and Techniques - Computational Methods

Sequence Searching Against Known Domains*



http://pawsonlab.mshri.on.ca/

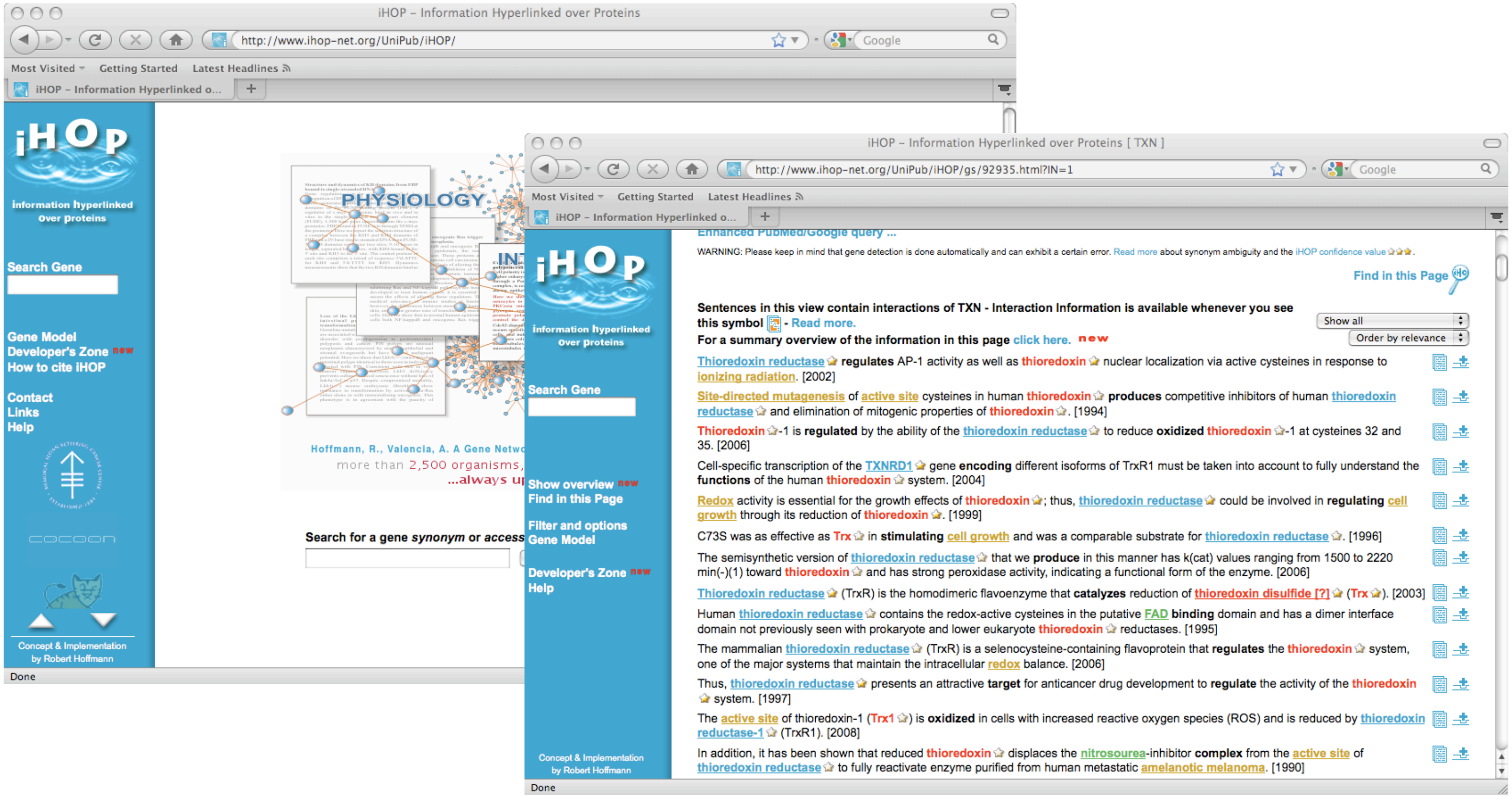
Motif Searching Using Known Motifs


Protein Interaction Domains

Text Mining*

- Searching Medline or Pubmed for words or word combinations
- “X binds to Y”; “X interacts with Y”; “X associates with Y” etc. etc.
- Requires a list of known gene names or protein names for a given organism (a protein/gene thesaurus)

iHOP (Information hyperlinked over proteins)



WARNING: Please keep in mind that gene detection is done automatically and can exhibit a certain error. Read more about synonym ambiguity and the iHOP confidence value.

Sentences in this view contain interactions of TXN - Interaction Information is available whenever you see this symbol - Read more.

For a summary overview of the information in this page click here. new

Thioredoxin reductase **regulates** AP-1 activity as well as thioredoxin nuclear localization via active cysteines in response to ionizing radiation. [2002]

Site-directed mutagenesis of active site cysteines in human thioredoxin **produces** competitive inhibitors of human thioredoxin reductase and elimination of mitogenic properties of thioredoxin. [1994]

Thioredoxin-1 is **regulated** by the ability of the thioredoxin reductase to reduce **oxidized** thioredoxin-1 at cysteines 32 and 35. [2006]

Cell-specific transcription of the TXNRD1 gene **encoding** different isoforms of TrxR1 must be taken into account to fully understand the **functions** of the human thioredoxin system. [2004]

Redox activity is essential for the growth effects of thioredoxin; thus, thioredoxin reductase could be involved in **regulating** cell growth through its reduction of thioredoxin. [1999]

C73S was as effective as Trx in **stimulating** cell growth and was a comparable substrate for thioredoxin reductase. [1996]

The semisynthetic version of thioredoxin reductase that we **produce** in this manner has k(cat) values ranging from 1500 to 2220 min(-)(1) toward thioredoxin and has strong peroxidase activity, indicating a functional form of the enzyme. [2006]

Thioredoxin reductase (TrxR) is the homodimeric flavoenzyme that **catalyzes** reduction of thioredoxin disulfide [?] (Trx). [2003]

Human thioredoxin reductase contains the redox-active cysteines in the putative FAD **binding** domain and has a dimer interface domain not previously seen with prokaryote and lower eukaryote thioredoxin reductases. [1995]

The mammalian thioredoxin reductase (TrxR) is a selenocysteine-containing flavoprotein that **regulates** the thioredoxin system, one of the major systems that maintain the intracellular redox balance. [2006]

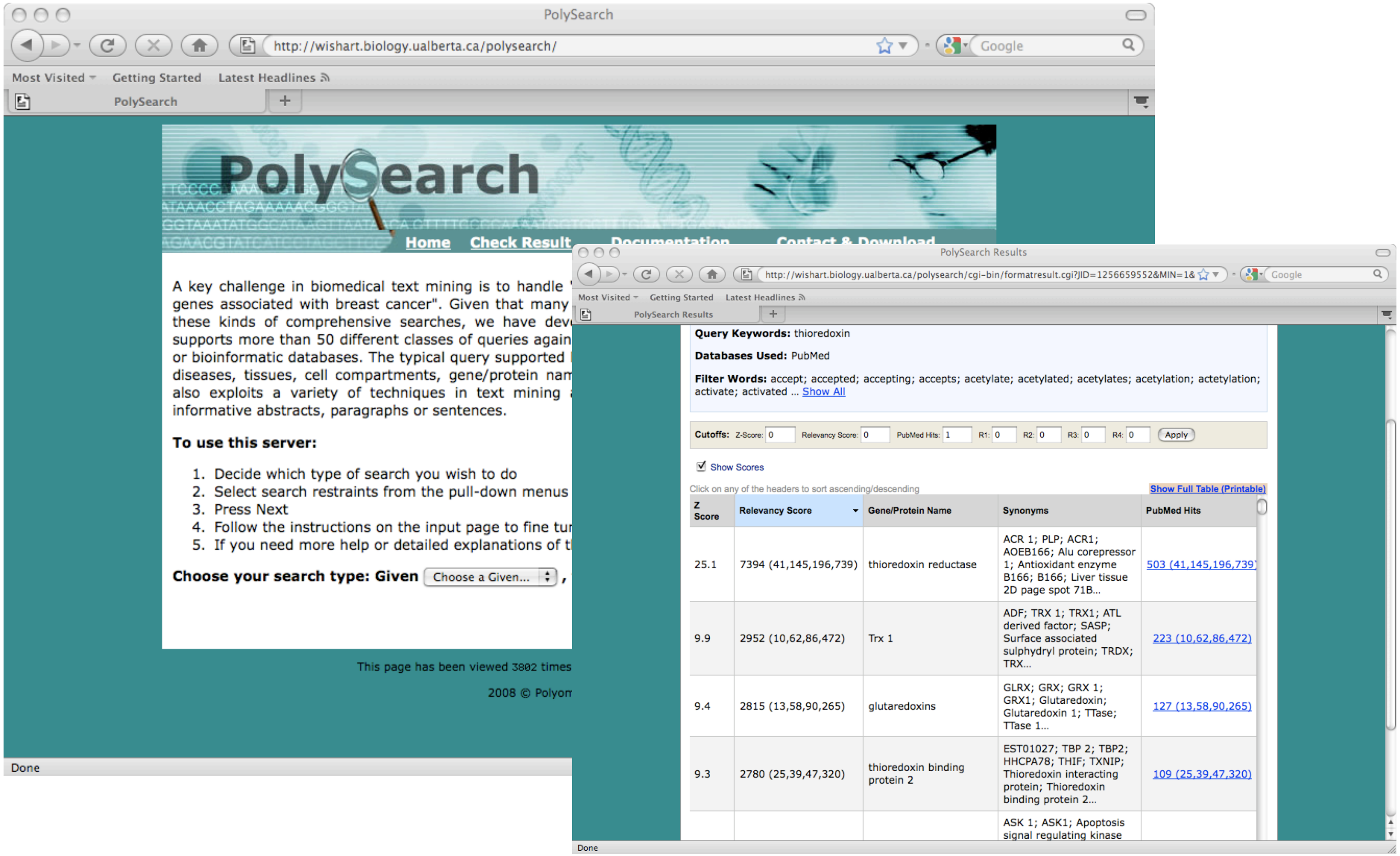
Thus, thioredoxin reductase presents an attractive **target** for anticancer drug development to **regulate** the activity of the thioredoxin system. [1997]

The active site of thioredoxin-1 (Trx1) is **oxidized** in cells with increased reactive oxygen species (ROS) and is reduced by thioredoxin reductase-1 (TrxR1). [2008]

In addition, it has been shown that reduced thioredoxin displaces the nitrosourea-inhibitor **complex** from the active site of thioredoxin reductase to fully reactivate enzyme purified from human metastatic amelanotic melanoma. [1990]

http://www.ihop-net.org/UniPub/iHOP/

PolySearch*



A key challenge in biomedical text mining is to handle
genes associated with breast cancer". Given that many
these kinds of comprehensive searches, we have dev
supports more than 50 different classes of queries agai
or bioinformatic databases. The typical query supported
diseases, tissues, cell compartments, gene/protein nam
also exploits a variety of techniques in text mining
informative abstracts, paragraphs or sentences.

To use this server:

1. Decide which type of search you wish to do
2. Select search restraints from the pull-down menus
3. Press Next
4. Follow the instructions on the input page to fine tu
5. If you need more help or detailed explanations of t

Choose your search type: Given Choose a Given... ,

This page has been viewed 3802 times

2008 © Polyom

Query Keywords: thioredoxin

Databases Used: PubMed

Filter Words: accept; accepted; accepting; accepts; acetylate; acetylated; acetylates; acetylation; actetylation; activate; activated ... Show All

Cutoffs: Z-Score: 0 Relevancy Score: 0 PubMed Hits: 1 R1: 0 R2: 0 R3: 0 R4: 0

☑ Show Scores

Click on any of the headers to sort ascending/descending

Z Score	Relevancy Score	Gene/Protein Name	Synonyms	PubMed Hits
25.1	7394 (41,145,196,739)	thioredoxin reductase	ACR 1; PLP; ACR1; AOEB166; Alu corepressor 1; Antioxidant enzyme B166; B166; Liver tissue 2D page spot 71B...	503 (41,145,196,739)
9.9	2952 (10,62,86,472)	Trx 1	ADF; TRX 1; TRX1; ATL derived factor; SASP; Surface associated sulphydryl protein; TRDX; TRX...	223 (10,62,86,472)
9.4	2815 (13,58,90,265)	glutaredoxins	GLRX; GRX; GRX 1; GRX1; Glutaredoxin; Glutaredoxin 1; TTase; TTase 1...	127 (13,58,90,265)
9.3	2780 (25,39,47,320)	thioredoxin binding protein 2	EST01027; TBP 2; TBP2; HHCPA78; THIF; TXNIP; Thioredoxin interacting protein; Thioredoxin binding protein 2...	109 (25,39,47,320)
			ASK 1; ASK1; Apoptosis signal regulating kinase	

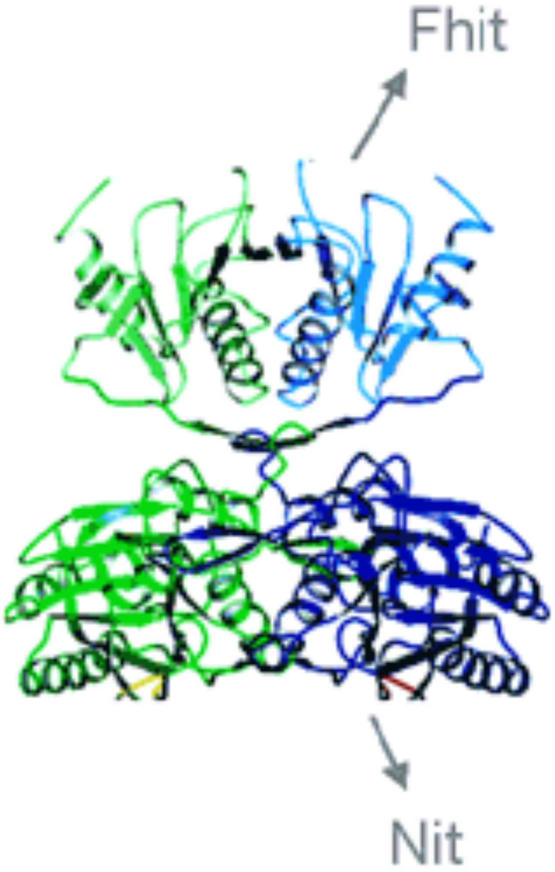
http://wishart.biology.ualberta..ca/polysearch

Rosetta Stone Method

Monomeric proteins that are fused in other organisms tend to be functionally related and physically interacting.

For example, using the Rosetta Stone™ method, it was found that human Nit and Fhit proteins are:

- fused in invertebrates
- form a heterocomplex in mammals

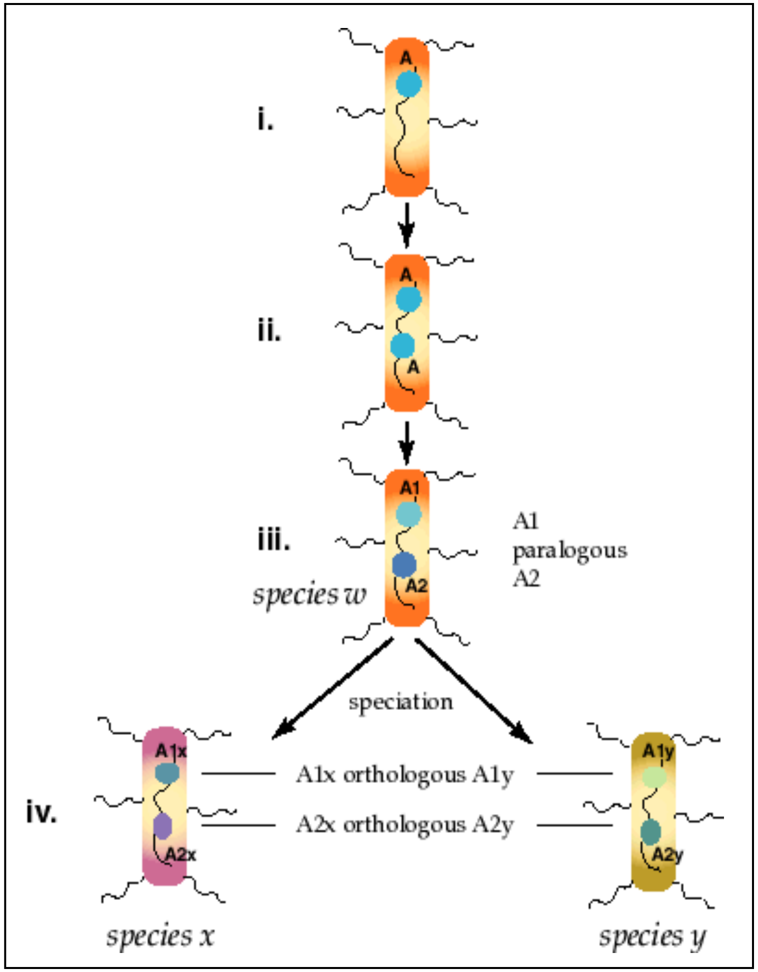


Interologs, Homologs, Paralogs*...

- **Homolog**
  - Common Ancestors
  - Common 3D Structure
  - Common Active Sites
- **Ortholog**
  - Derived from Speciation
- **Paralog**
  - Derived from Duplication

YM2

- **Interolog**
  - Protein-Protein Interaction



Finding Interologs*

- If A and B interact in organism X, then if organism Y has a homolog of A (A') and a homolog of B (B') then A' and B' should interact too!
- Makes use of BLAST searches against entire proteome of well-studied organisms (yeast, E. coli)
- Requires list of known interacting partners

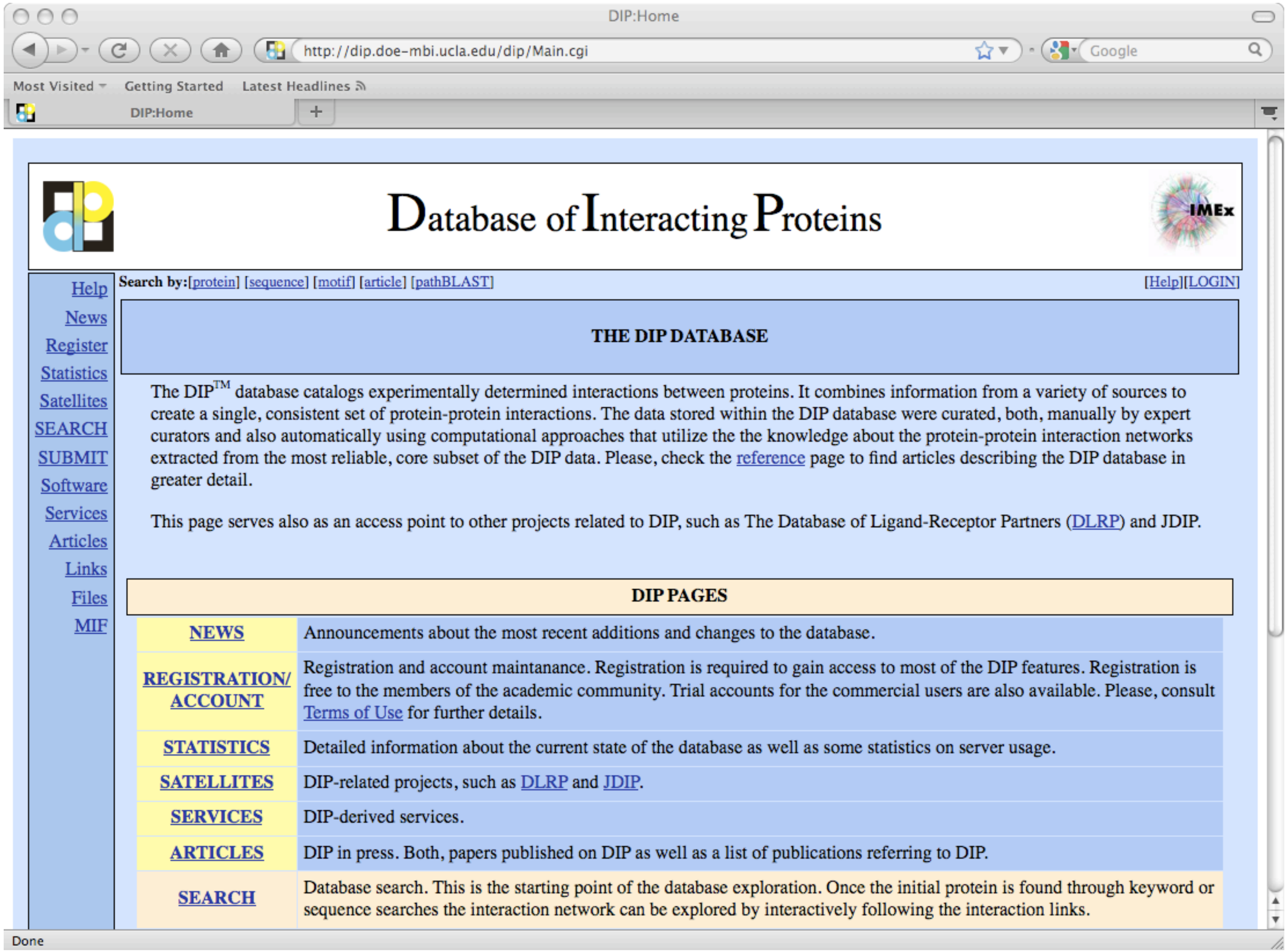
A Flood of Data

- High throughput techniques are leading to more and more data on protein interactions
- This is where bioinformatics can play a key role
- Some suggest that this is the “future” for bioinformatics

Interaction Databases

- DIP
  - http://dip.doe-mbi.ucla.edu/dip/Main.cgi
- MINT
  - http://mint.bio.uniroma2.it/mint/
- String
  - http://string.embl.de/
- IntAct
  - http://www.ebi.ac.uk/intact/main.xhtml

DIP Database of Interacting Proteins



Database of Interacting Proteins

Search by:[protein] [sequence] [motif] [article] [pathBLAST]

THE DIP DATABASE

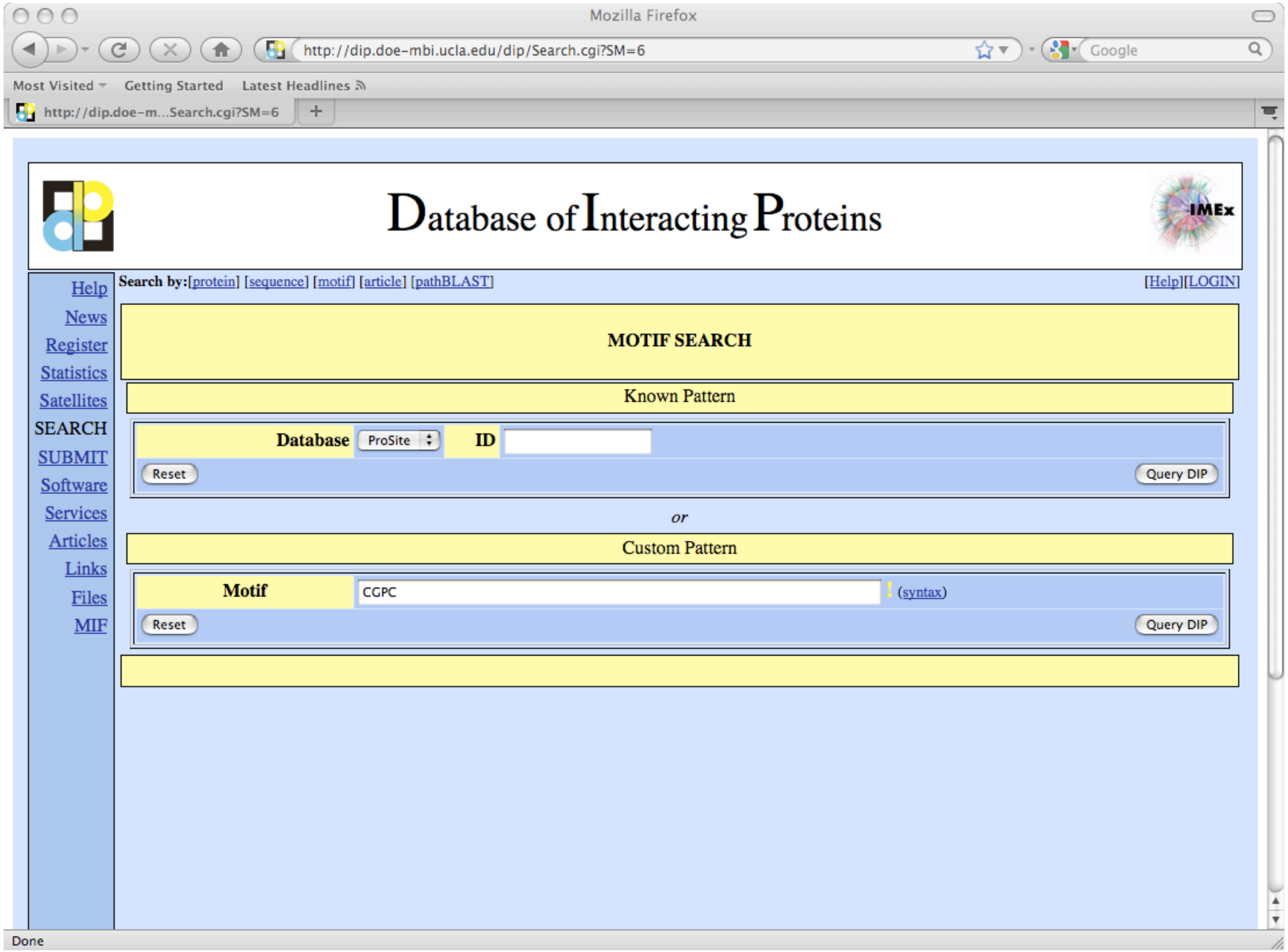
The DIP^{TM} database catalogs experimentally determined interactions between proteins. It combines information from a variety of sources to create a single, consistent set of protein-protein interactions. The data stored within the DIP database were curated, both, manually by expert curators and also automatically using computational approaches that utilize the the knowledge about the protein-protein interaction networks extracted from the most reliable, core subset of the DIP data. Please, check the reference page to find articles describing the DIP database in greater detail.

This page serves also as an access point to other projects related to DIP, such as The Database of Ligand-Receptor Partners (DLRP) and JDIP.

DIP PAGES	
NEWS	Announcements about the most recent additions and changes to the database.
REGISTRATION/ ACCOUNT	Registration and account maintanance. Registration is required to gain access to most of the DIP features. Registration is free to the members of the academic community. Trial accounts for the commercial users are also available. Please, consult Terms of Use for further details.
STATISTICS	Detailed information about the current state of the database as well as some statistics on server usage.
SATELLITES	DIP-related projects, such as DLRP and JDIP.
SERVICES	DIP-derived services.
ARTICLES	DIP in press. Both, papers published on DIP as well as a list of publications referring to DIP.
SEARCH	Database search. This is the starting point of the database exploration. Once the initial protein is found through keyword or sequence searches the interaction network can be explored by interactively following the interaction links.

http://dip.doe-mbi.ucla.edu/dip/Main.cgi

DIP Query Page



Mozilla Firefox

http://dip.doe-mbi.ucla.edu/dip/Search.cgi?SM=6

Database of Interacting Proteins

Search by:[protein] [sequence] [motif] [article] [pathBLAST]

[Help][LOGIN]

MOTIF SEARCH

Known Pattern

Database ProSite **ID**

or

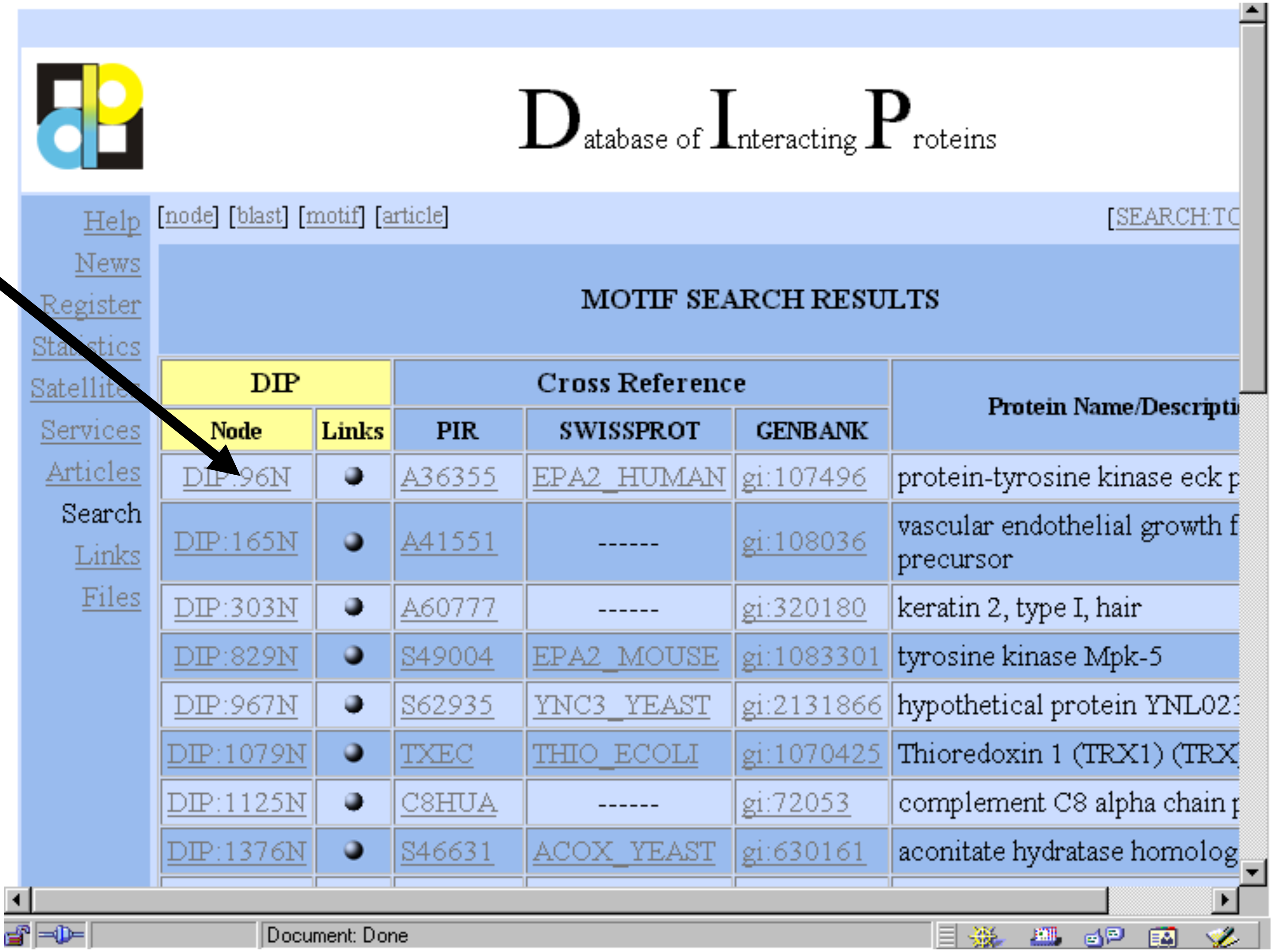
Custom Pattern

Motif CGPC (syntax)

Done

DIP Results Page

click



Database of Interacting Proteins

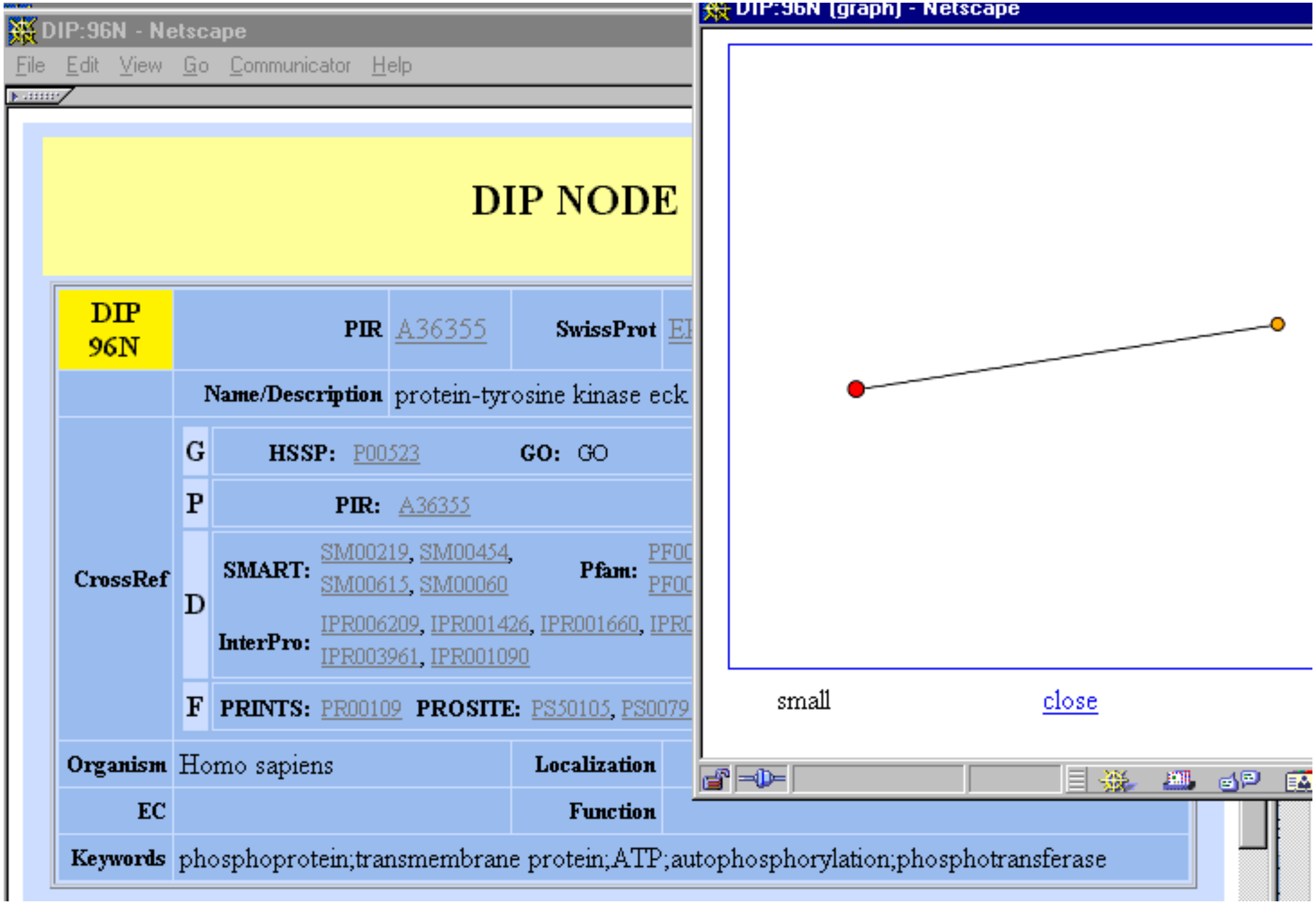
[node] [blast] [motif] [article] [SEARCH:TO

MOTIF SEARCH RESULTS

DIP		Cross Reference			Protein Name/Descripti
Node	Links	PIR	SWISSPROT	GENBANK	
DIP:96N	●	A36355	EPA2_HUMAN	gi:107496	protein-tyrosine kinase eck p
DIP:165N	●	A41551	------	gi:108036	vascular endothelial growth f precursor
DIP:303N	●	A60777	------	gi:320180	keratin 2, type I, hair
DIP:829N	●	S49004	EPA2_MOUSE	gi:1083301	tyrosine kinase Mpk-5
DIP:967N	●	S62935	YNC3_YEAST	gi:2131866	hypothetical protein YNL02
DIP:1079N	●	TXEC	THIO_ECOLI	gi:1070425	Thioredoxin 1 (TRX1) (TRX
DIP:1125N	●	C8HUA	------	gi:72053	complement C8 alpha chain p
DIP:1376N	●	S46631	ACOX_YEAST	gi:630161	aconitate hydratase homolog

Document: Done

DIP Results Page



DIP NODE

DIP 96N

PIR A36355 **SwissProt** EI

Name/Description protein-tyrosine kinase eck

CrossRef

G **HSSP:** P00523 **GO:** GO

P **PIR:** A36355

D **SMART:** SM00219, SM00454, SM00615, SM00060 **Pfam:** PF0(
PF0(

InterPro: IPR006209, IPR001426, IPR001660, IPR(
IPR003961, IPR001090

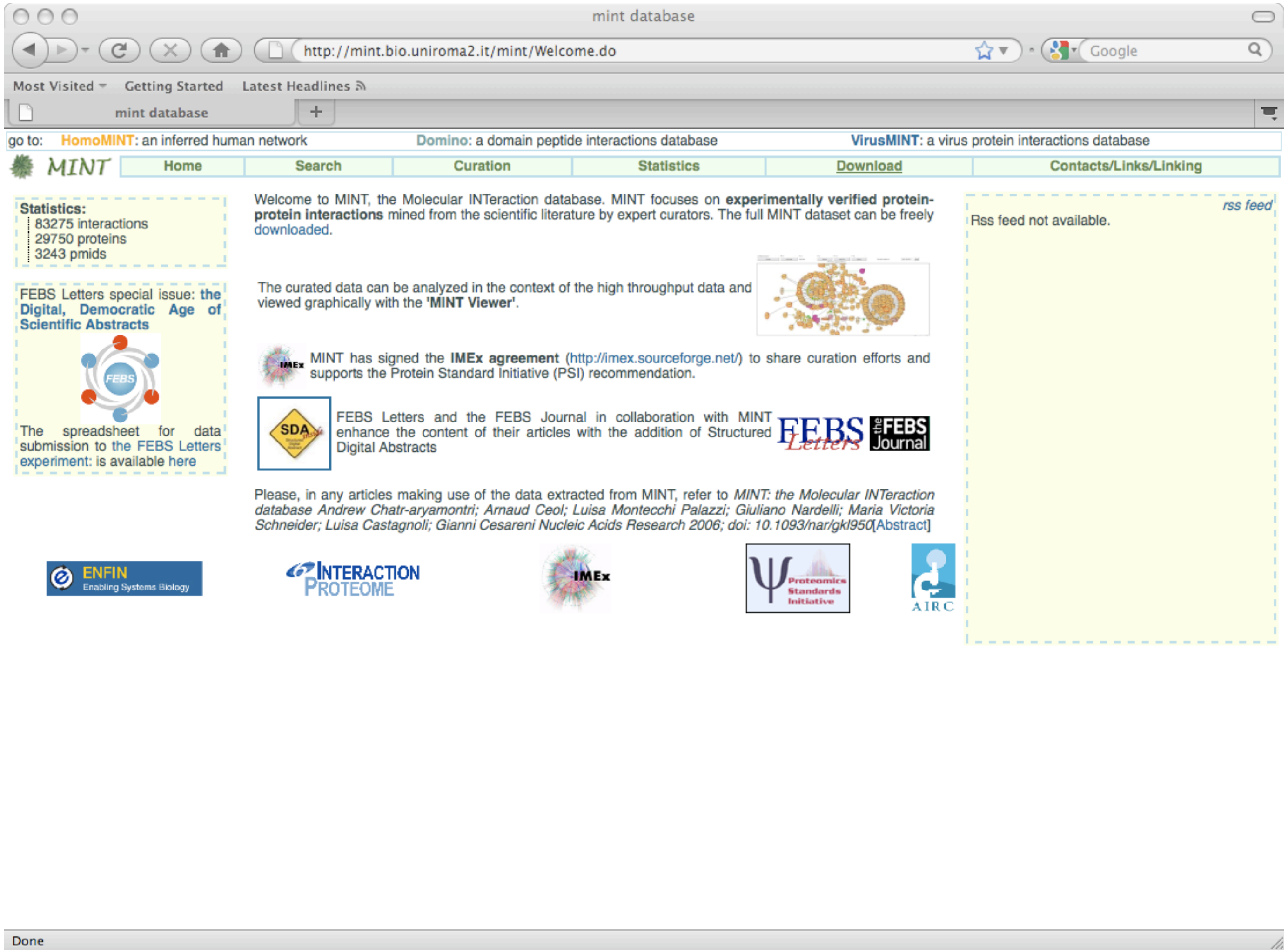
F **PRINTS:** PR00109 **PROSITE:** PS50105, PS0079

Organism Homo sapiens **Localization**

EC **Function**

Keywords phosphoprotein;transmembrane protein;ATP;autophosphorylation;phosphotransferase

MINT Molecular Interaction Database



go to: HomoMINT: an inferred human network Domino: a domain peptide interactions database VirusMINT: a virus protein interactions database

Statistics:
83275 interactions
29750 proteins
3243 pmids

FEBS Letters special issue: **the Digital, Democratic Age of Scientific Abstracts**

The spreadsheet for data submission to the FEBS Letters experiment: is available here

Welcome to MINT, the Molecular INTeraction database. MINT focuses on **experimentally verified protein-protein interactions** mined from the scientific literature by expert curators. The full MINT dataset can be freely downloaded.

The curated data can be analyzed in the context of the high throughput data and viewed graphically with the **'MINT Viewer'**.

MINT has signed the **IMEx agreement** (http://imex.sourceforge.net/) to share curation efforts and supports the Protein Standard Initiative (PSI) recommendation.

FEBS Letters and the FEBS Journal in collaboration with MINT enhance the content of their articles with the addition of Structured Digital Abstracts

Please, in any articles making use of the data extracted from MINT, refer to *MINT: the Molecular INTeraction database Andrew Chatr-aryamontri; Arnaud Ceol; Luisa Montecchi Palazzi; Giuliano Nardelli; Maria Victoria Schneider; Luisa Castagnoli; Gianni Cesareni Nucleic Acids Research 2006; doi: 10.1093/nar/gkl950*[Abstract]

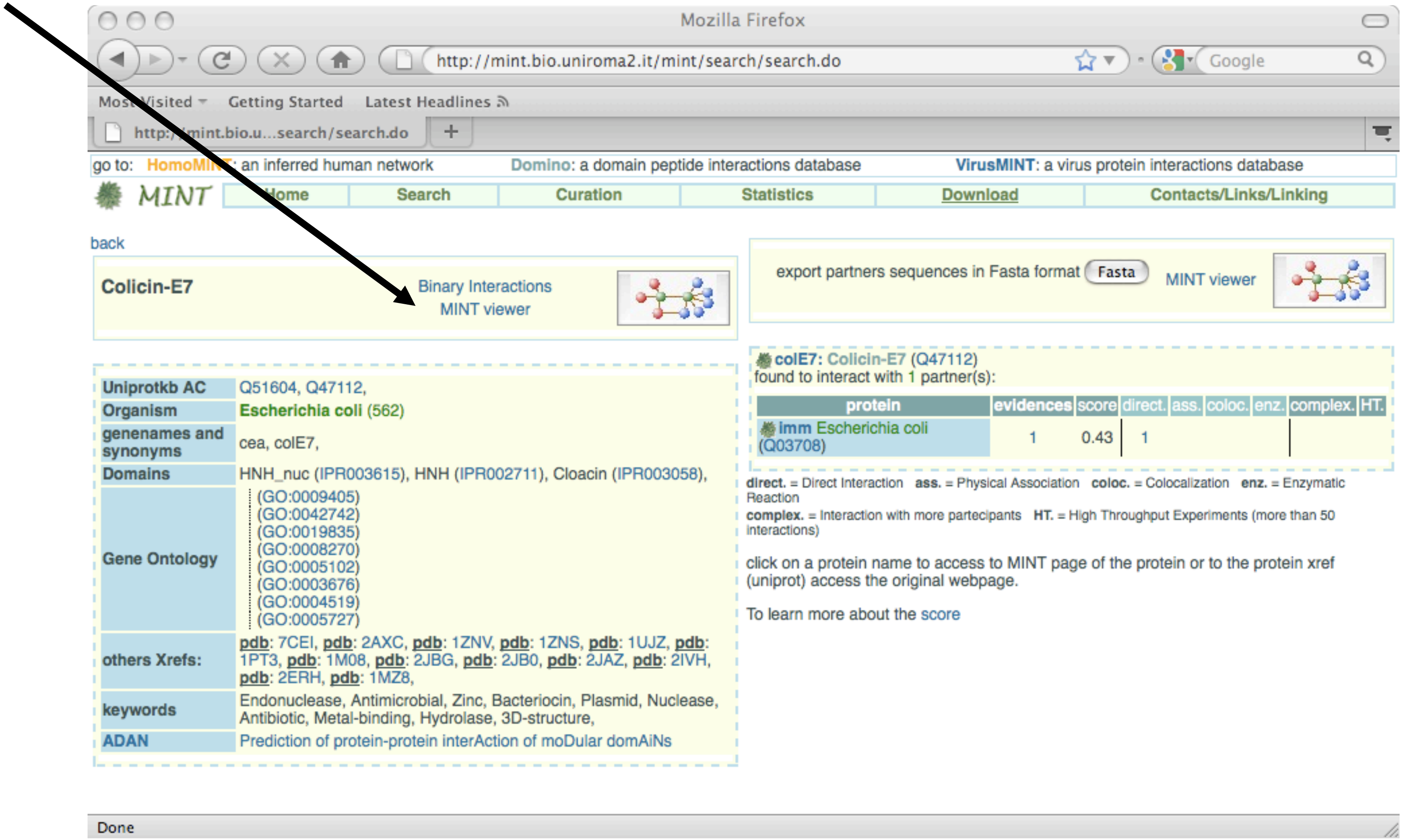
rss feed

Rss feed not available.

http://mint.bio.uniroma2.it/mint/

MINT Results

click



go to: HomoMINT: an inferred human network   Domino: a domain peptide interactions database   VirusMINT: a virus protein interactions database

MINT | Home | Search | Curation | Statistics | Download | Contacts/Links/Linking

back

Colicin-E7

Binary Interactions
MINT viewer

export partners sequences in Fasta format Fasta   MINT viewer

Uniprotkb AC	Q51604, Q47112,
Organism	Escherichia coli (562)
genenames and synonyms	cea, colE7,
Domains	HNH_nuc (IPR003615), HNH (IPR002711), Cloacin (IPR003058),
Gene Ontology	(GO:0009405) (GO:0042742) (GO:0019835) (GO:0008270) (GO:0005102) (GO:0003676) (GO:0004519) (GO:0005727)
others Xrefs:	pdb: 7CEI, pdb: 2AXC, pdb: 1ZNV, pdb: 1ZNS, pdb: 1UJZ, pdb: 1PT3, pdb: 1M08, pdb: 2JBG, pdb: 2JB0, pdb: 2JAZ, pdb: 2IVH, pdb: 2ERH, pdb: 1MZ8,
keywords	Endonuclease, Antimicrobial, Zinc, Bacteriocin, Plasmid, Nuclease, Antibiotic, Metal-binding, Hydrolase, 3D-structure,
ADAN	Prediction of protein-protein interAction of moDular domAiNs

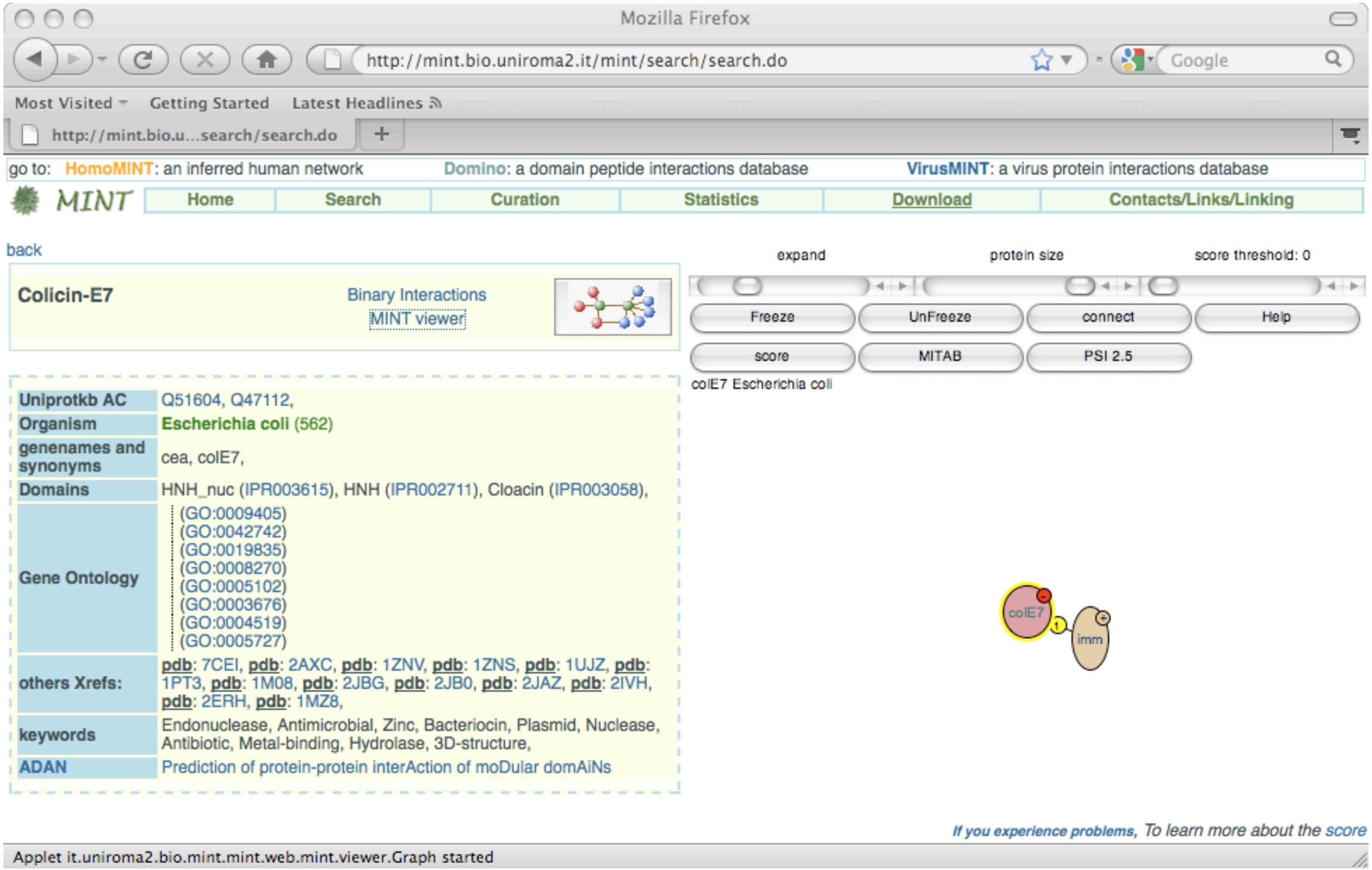
colE7: Colicin-E7 (Q47112)
found to interact with 1 partner(s):

protein	evidences	score	direct.	ass.	coloc.	enz.	complex.	HT.
imm Escherichia coli (Q03708)	1	0.43	1					

direct. = Direct Interaction **ass.** = Physical Association **coloc.** = Colocalization **enz.** = Enzymatic Reaction
complex. = Interaction with more partecipants **HT.** = High Throughput Experiments (more than 50 interactions)

click on a protein name to access to MINT page of the protein or to the protein xref (uniprot) access the original webpage.

To learn more about the score

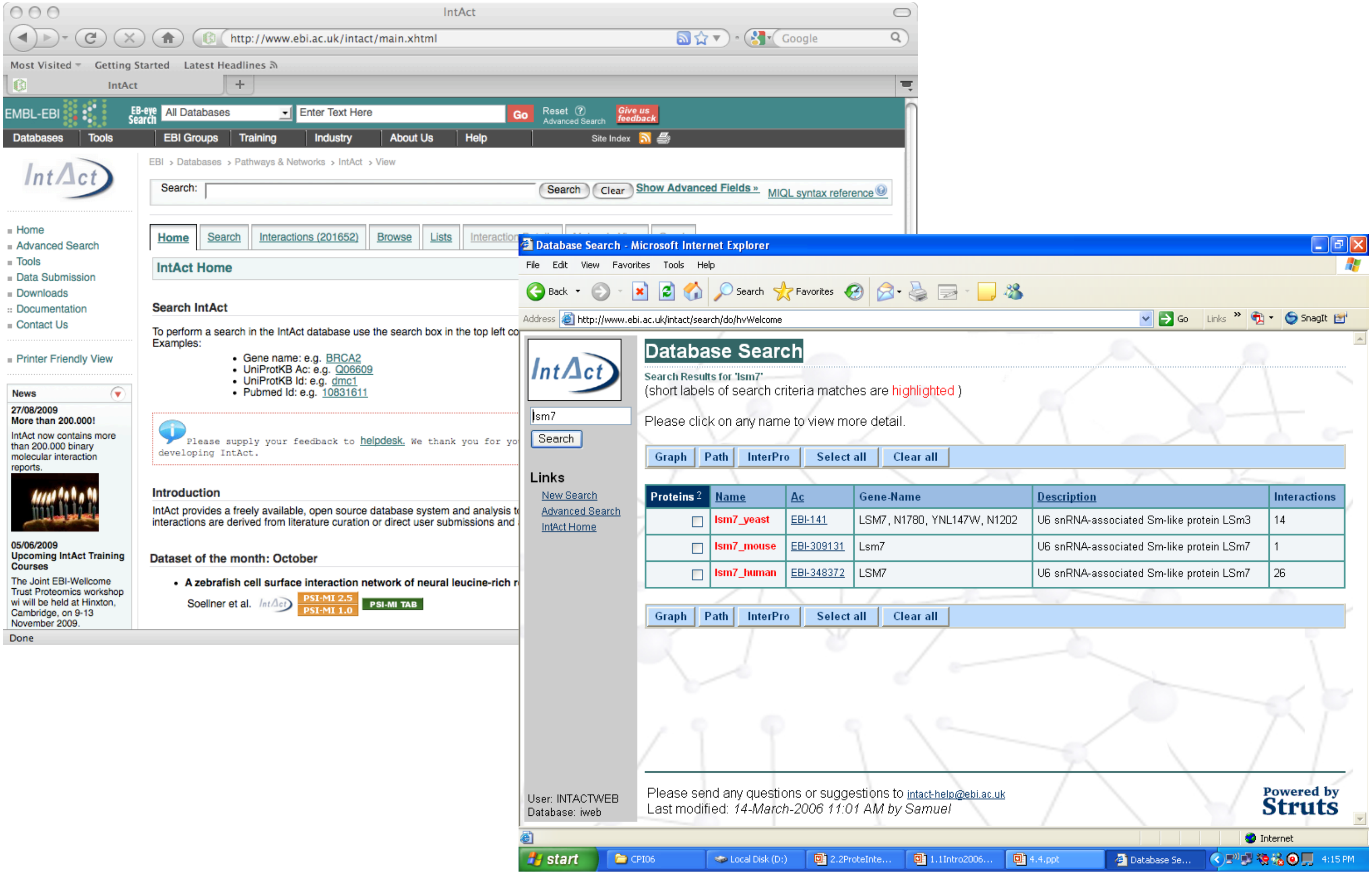


Colicin-E7

Uniprotkb AC	Q51604, Q47112,
Organism	**Escherichia coli** (562)
genenames and synonyms	cea, colE7,
Domains	HNH_nuc (IPR003615), HNH (IPR002711), Cloacin (IPR003058),
Gene Ontology	(GO:0009405) (GO:0042742) (GO:0019835) (GO:0008270) (GO:0005102) (GO:0003676) (GO:0004519) (GO:0005727)
others Xrefs:	**pdb**: 7CEI, **pdb**: 2AXC, **pdb**: 1ZNV, **pdb**: 1ZNS, **pdb**: 1UJZ, **pdb**: 1PT3, **pdb**: 1M08, **pdb**: 2JBG, **pdb**: 2JB0, **pdb**: 2JAZ, **pdb**: 2IVH, **pdb**: 2ERH, **pdb**: 1MZ8,
keywords	Endonuclease, Antimicrobial, Zinc, Bacteriocin, Plasmid, Nuclease, Antibiotic, Metal-binding, Hydrolase, 3D-structure,
ADAN	Prediction of protein-protein interAction of moDular domAiNs

Applet it.uniroma2.bio.mint.mint.web.mint.viewer.Graph started

IntAct*



EBI > Databases > Pathways & Networks > IntAct > View

27/08/2009
More than 200.000!
IntAct now contains more than 200.000 binary molecular interaction reports.

05/06/2009
Upcoming IntAct Training Courses
The Joint EBI-Wellcome Trust Proteomics workshop wi will be held at Hinxton, Cambridge, on 9-13 November 2009.

IntAct Home

Search IntAct

To perform a search in the IntAct database use the search box in the top left co
Examples:

- Gene name: e.g. BRCA2
- UniProtKB Ac: e.g. Q06609
- UniProtKB Id: e.g. dmc1
- Pubmed Id: e.g. 10831611

Please supply your feedback to helpdesk. We thank you for yo
developing IntAct.

Introduction

IntAct provides a freely available, open source database system and analysis t
interactions are derived from literature curation or direct user submissions and

Dataset of the month: October

- **A zebrafish cell surface interaction network of neural leucine-rich r**

Soellner et al.

Database Search

Search Results for 'lsm7'
(short labels of search criteria matches are highlighted)

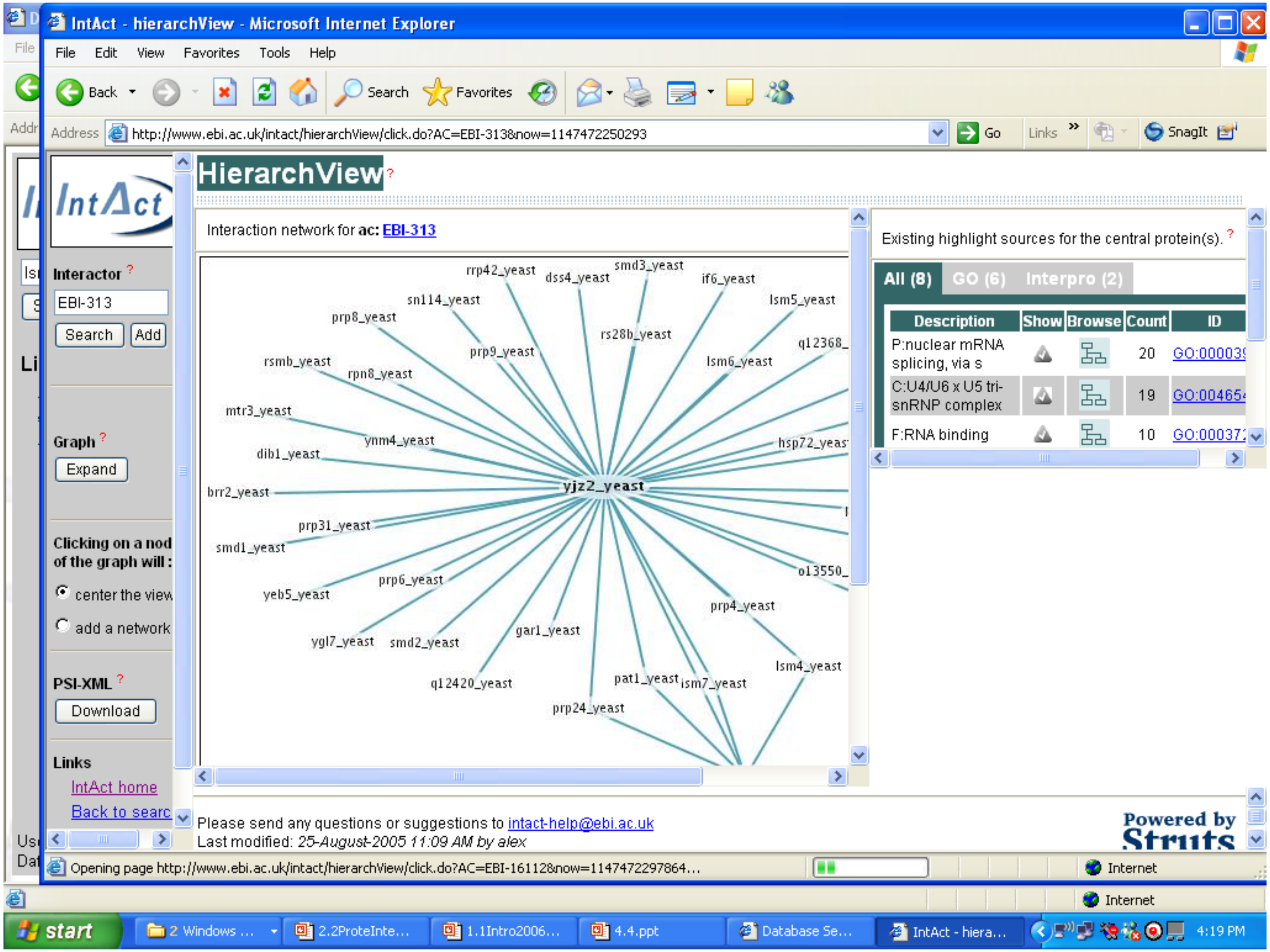
Please click on any name to view more detail.

Proteins 2	Name	Ac	Gene-Name	Description	Interactions
☐	lsm7_yeast	EBI-141	LSM7, N1780, YNL147W, N1202	U6 snRNA-associated Sm-like protein LSm3	14
☐	lsm7_mouse	EBI-309131	Lsm7	U6 snRNA-associated Sm-like protein LSm7	1
☐	lsm7_human	EBI-348372	LSM7	U6 snRNA-associated Sm-like protein LSm7	26

User: INTACTWEB
Database: iweb

Please send any questions or suggestions to intact-help@ebi.ac.uk
Last modified: *14-March-2006 11:01 AM by Samuel*

IntAct



http://www.ebi.ac.uk/intact/hierarchView/click.do?AC=EBI-313&now=1147472250293

Interactor: EBI-313

Interaction network for **ac:** EBI-313

Clicking on a nod of the graph will :

● center the view

○ add a network

Existing highlight sources for the central protein(s).

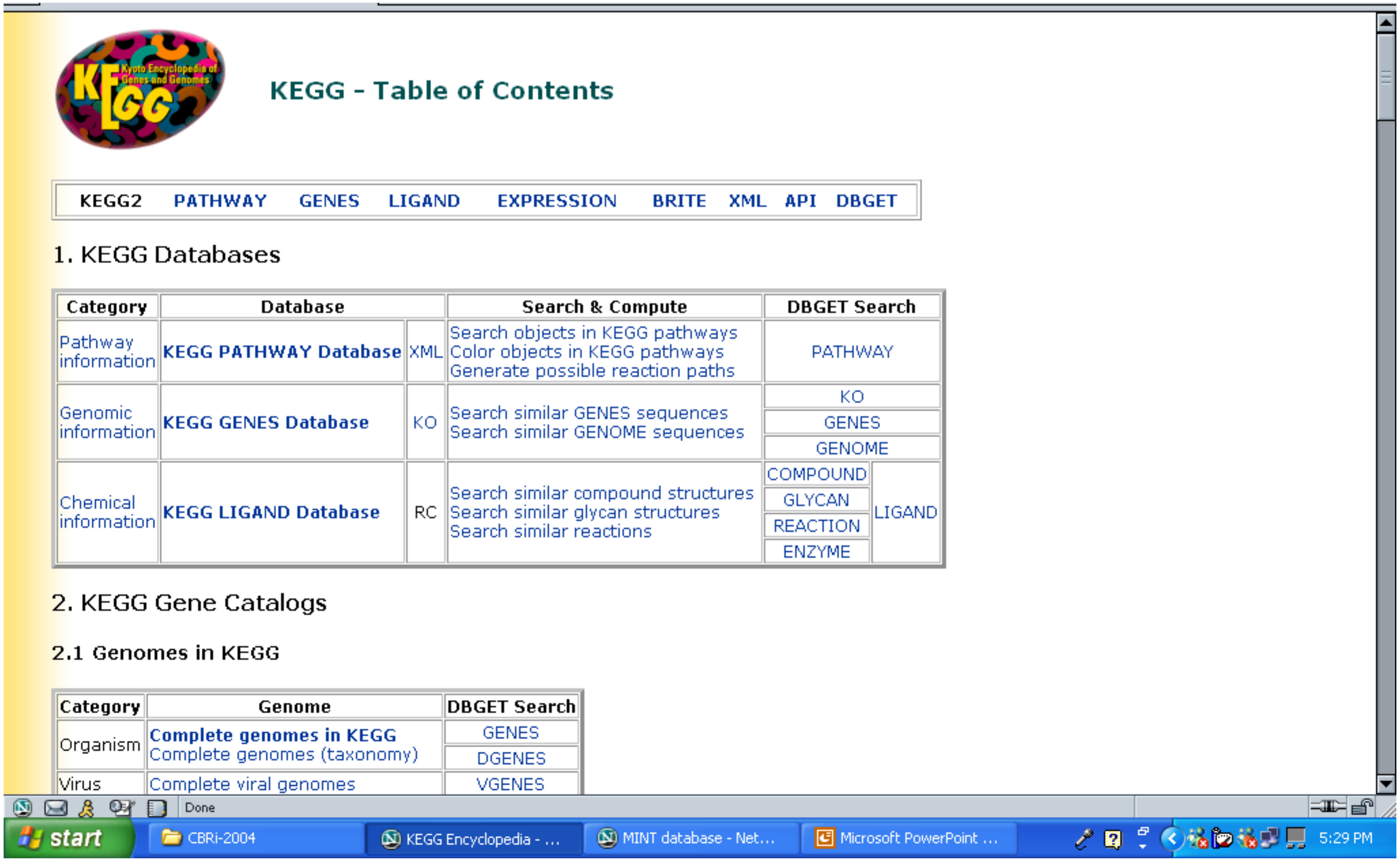
Description	Show	Browse	Count	ID
P:nuclear mRNA splicing, via s			20	GO:00003
C:U4/U6 x U5 tri-snRNP complex			19	GO:00465
F:RNA binding			10	GO:00037

Please send any questions or suggestions to intact-help@ebi.ac.uk

Last modified: *25-August-2005 11:09 AM by alex*

Opening page http://www.ebi.ac.uk/intact/hierarchView/click.do?AC=EBI-16112&now=1147472297864...

KEGG Kyoto Encyclopedia of Genes and Genomes*



KEGG - Table of Contents

KEGG2 PATHWAY GENES LIGAND EXPRESSION BRITE XML API DBGET

1. KEGG Databases

Category	Database		Search & Compute	DBGET Search	
Pathway information	**KEGG PATHWAY Database**	XML	Search objects in KEGG pathways Color objects in KEGG pathways Generate possible reaction paths	PATHWAY	
Genomic information	**KEGG GENES Database**	KO	Search similar GENES sequences Search similar GENOME sequences	KO GENES GENOME	
Chemical information	**KEGG LIGAND Database**	RC	Search similar compound structures Search similar glycan structures Search similar reactions	COMPOUND GLYCAN REACTION ENZYME	LIGAND

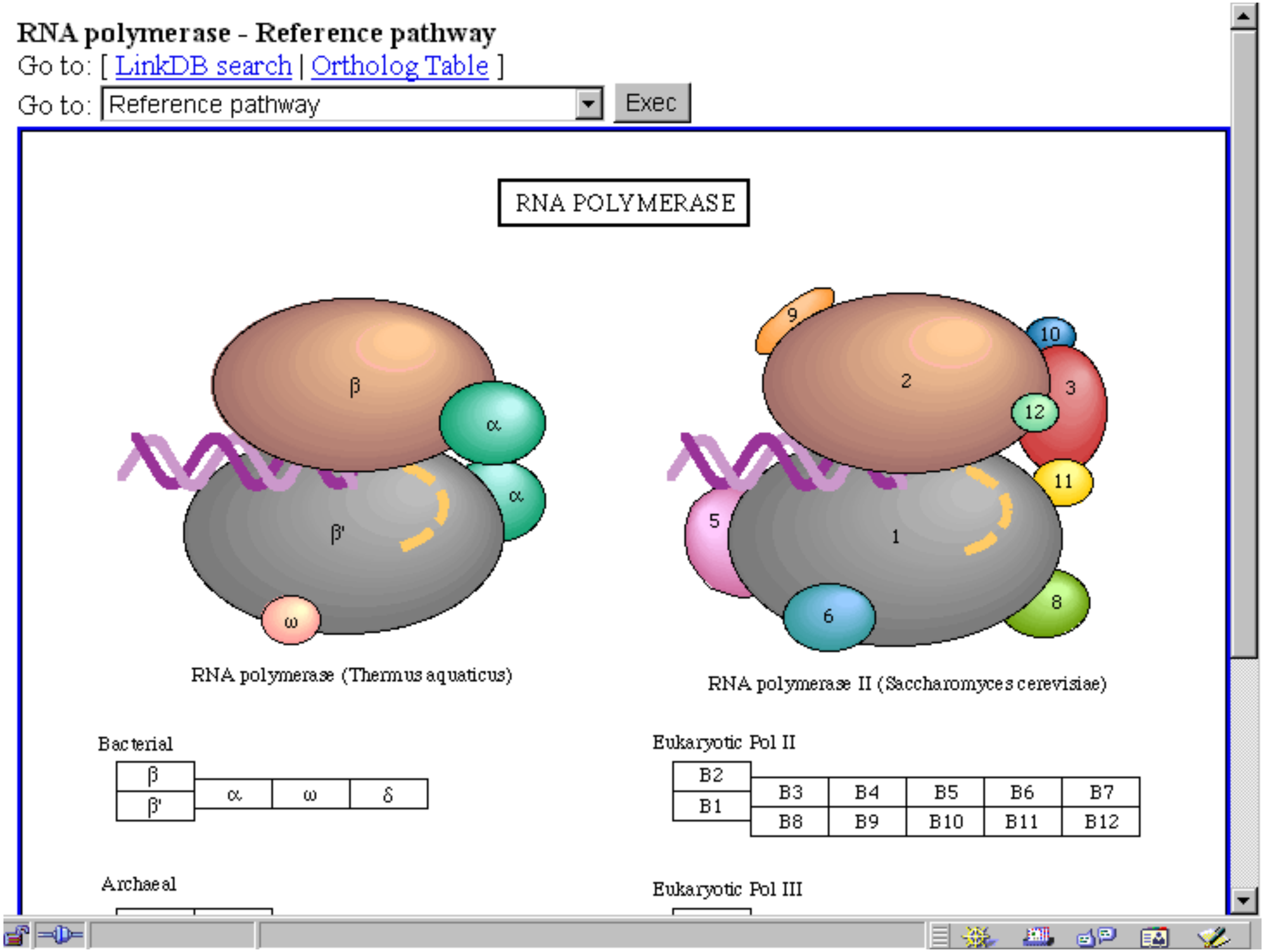
2. KEGG Gene Catalogs

2.1 Genomes in KEGG

Category	Genome	DBGET Search
Organism	**Complete genomes in KEGG** Complete genomes (taxonomy)	GENES DGENES
Virus	Complete viral genomes	VGENES

http://www.genome.ad.jp/kegg/kegg2.html

KEGG



RNA polymerase - Reference pathway

Go to: [LinkDB search | Ortholog Table]

Go to: Reference pathway

RNA POLYMERASE

RNA polymerase (Thermus aquaticus)

RNA polymerase II (Saccharomyces cerevisiae)

Bacterial

β	α	ω	δ
β'			

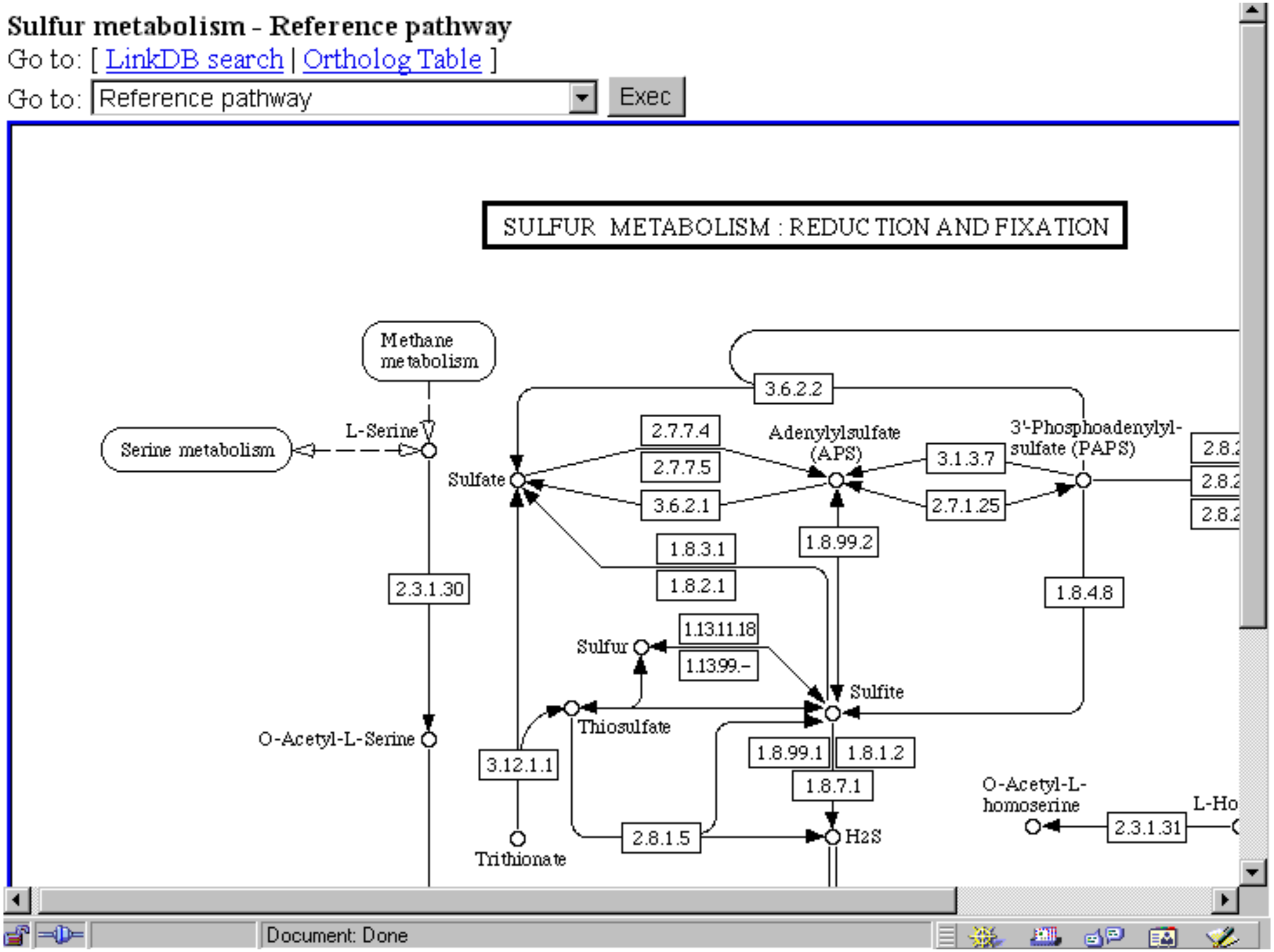
Eukaryotic Pol II

B2	B3	B4	B5	B6	B7
B1	B8	B9	B10	B11	B12

Archaeal

Eukaryotic Pol III

KEGG

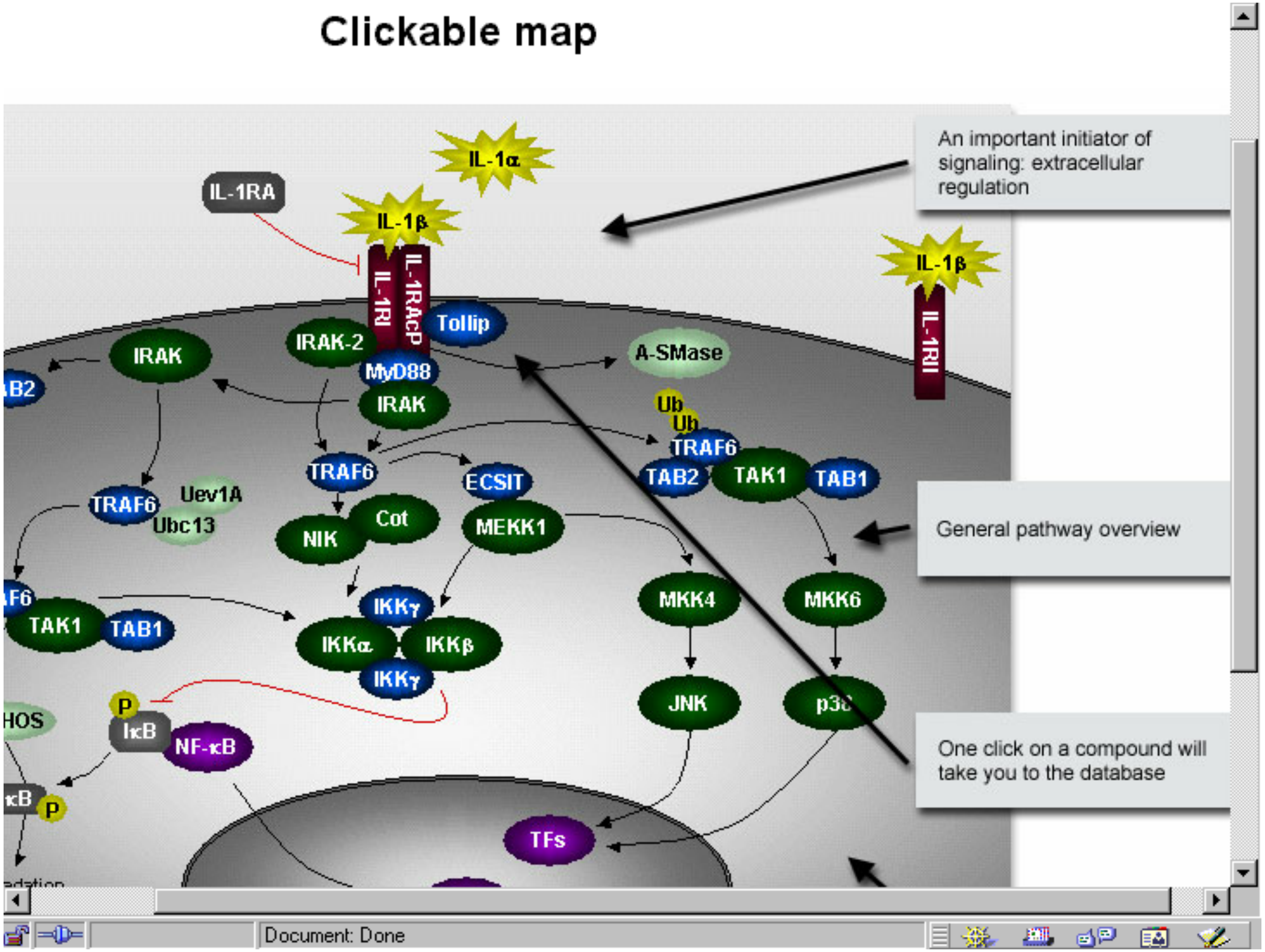


Sulfur metabolism - Reference pathway

Go to: [LinkDB search | Ortholog Table]

Go to: Reference pathway ▾ Exec

TRANSPATH



Clickable map

An important initiator of signaling: extracellular regulation

General pathway overview

One click on a compound will take you to the database

http://www.gene-regulation.com/pub/databases.html

BIOCARTA*

- www.biocarta.com
- Go to “Pathways”
- Web interactive links to many signalling pathways and other eukaryotic protein-protein interactions

PATHWAYS ▸ **All Pathways**

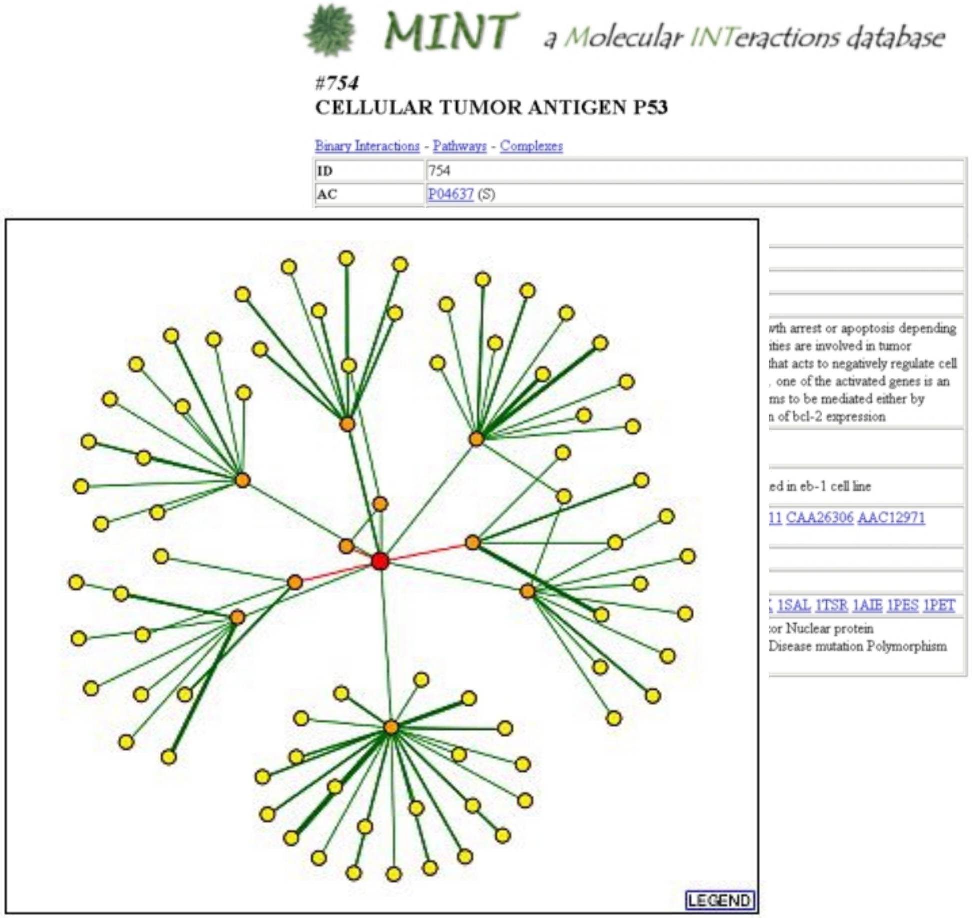
Observe how genes interact in dynamic graphical models. Our online maps depict molecular relationships from areas of active research. In an "open source" approach, this community-fed forum constantly integrates emerging proteomic information from the scientific community. It also catalogs and summarizes important resources providing information for over 120,000 genes from multiple species. Find both classical pathways as well as current suggestions for new pathways.

ALL PATHWAYS

A

- Acetylation and Deacetylation of RelA in The Nucleus H M
- Actions of Nitric Oxide in the Heart H
- Activation of cAMP-dependent protein kinase, PKA H M
- Activation of Csk by cAMP-dependent Protein Kinase Inhibits Signaling through the T Cell Receptor H M
- Activation of PKC through G protein coupled receptor H M
- Activation of Src by Protein-tyrosine phosphatase alpha H M
- Acute Myocardial Infarction H
- Adhesion and Diapedesis of Granulocytes H
- Adhesion and Diapedesis of Lymphocytes H
- Adhesion Molecules on Lymphocyte H M
- ADP-Ribosylation Factor H M
- Agrin in Postsynaptic Differentiation H M
- Ahr Signal Transduction Pathway H M
- AKAP95 role in mitosis and chromosome dynamics H M
- AKT Signaling Pathway H M
- ALK in cardiac myocytes H
- Alpha-synuclein and Parkin-mediated proteolysis in Parkinson's disease H M
- Alternative Complement Pathway H M
- Angiotensin II mediated activation of JNK Pathway via Pyk2 dependent signaling H M
- Angiotensin-converting enzyme 2 regulates heart function H M
- Anthrax Toxin Mechanism of Action H

Visualizing Interactions



MINT a Molecular INTeractions database

#754

CELLULAR TUMOR ANTIGEN P53

Binary Interactions - Pathways - Complexes

ID	754
AC	P04637 (S)

wth arrest or apoptosis depending
ities are involved in tumor
that acts to negatively regulate cell
. one of the activated genes is an
ms to be mediated either by
n of bcl-2 expression

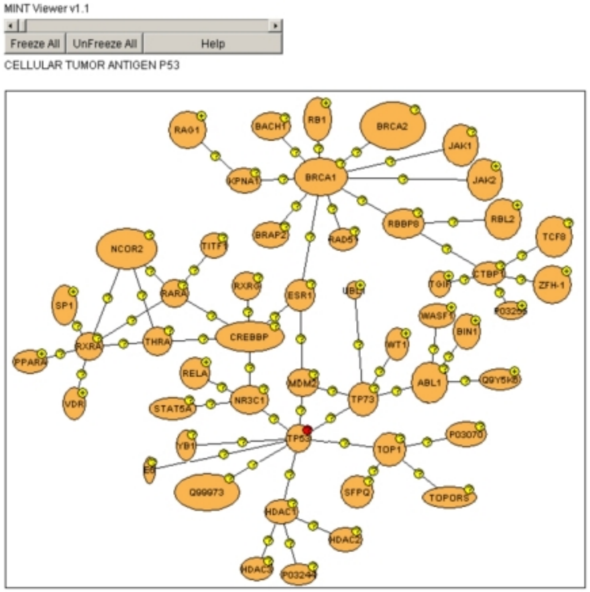
ed in eb-1 cell line

11 CAA26306 AAC12971

; 1SAL 1TSR 1AIE 1PES 1PET

tor Nuclear protein
Disease mutation Polymorphism

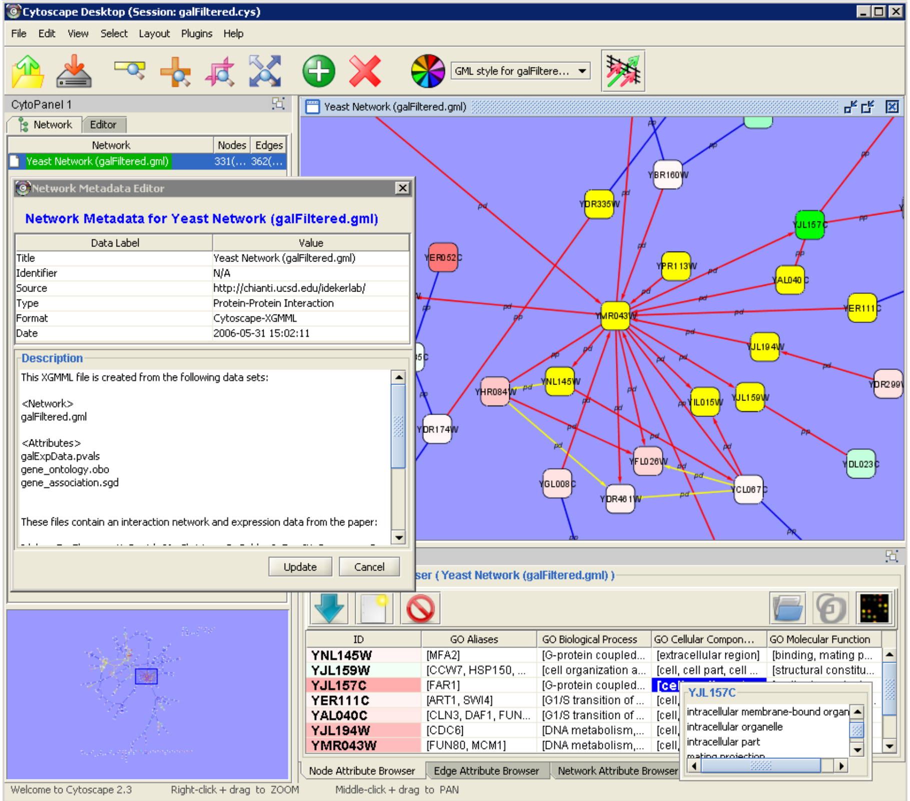
DIP



CELLULAR TUMOR ANTIGEN P53

MINT

Visualizing Interactions*



Network	Nodes	Edges
Yeast Network (galFiltered.gml)	331(...	362(...

Network Metadata for Yeast Network (galFiltered.gml)

Data Label	Value
Title	Yeast Network (galFiltered.gml)
Identifier	N/A
Source	http://chianti.ucsd.edu/idekerlab/
Type	Protein-Protein Interaction
Format	Cytoscape-XGMML
Date	2006-05-31 15:02:11

Description

This XGMML file is created from the following data sets:

<Network>
galFiltered.gml

<Attributes>
galExpData.pvals
gene_ontology.obo
gene_association.sgd

These files contain an interaction network and expression data from the paper:

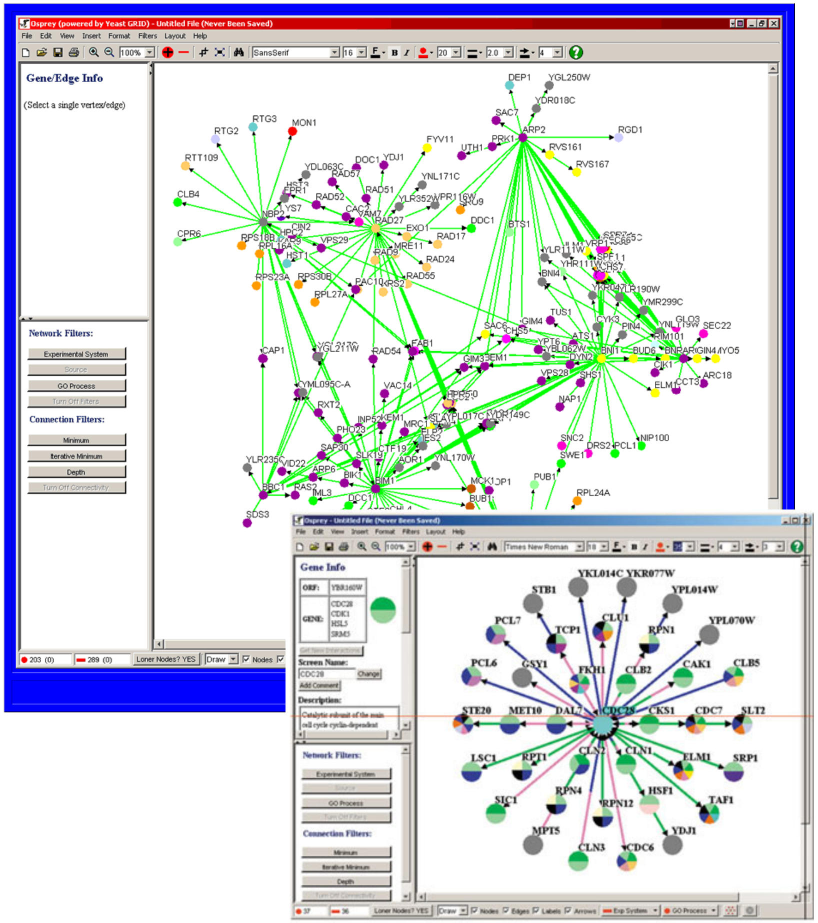
ser (Yeast Network (galFiltered.gml))

ID	GO Aliases	GO Biological Process	GO Cellular Compon...	GO Molecular Function
YNL145W	[MFA2]	[G-protein coupled...	[extracellular region]	[binding, mating p...
YJL159W	[CCW7, HSP150, ...	[cell organization a...	[cell, cell part, cell ...	[structural constitu...
YJL157C	[FAR1]	[G-protein coupled...	[ce	
YER111C	[ART1, SWI4]	[G1/S transition of ...	[cell,	
YAL040C	[CLN3, DAF1, FUN...	[G1/S transition of ...	[cell,	
YJL194W	[CDC6]	[DNA metabolism,...	[cell,	
YMR043W	[FUN80, MCM1]	[DNA metabolism,...	[cell,	

YJL157C

intracellular membrane-bound organ
intracellular organelle
intracellular part
mating projection

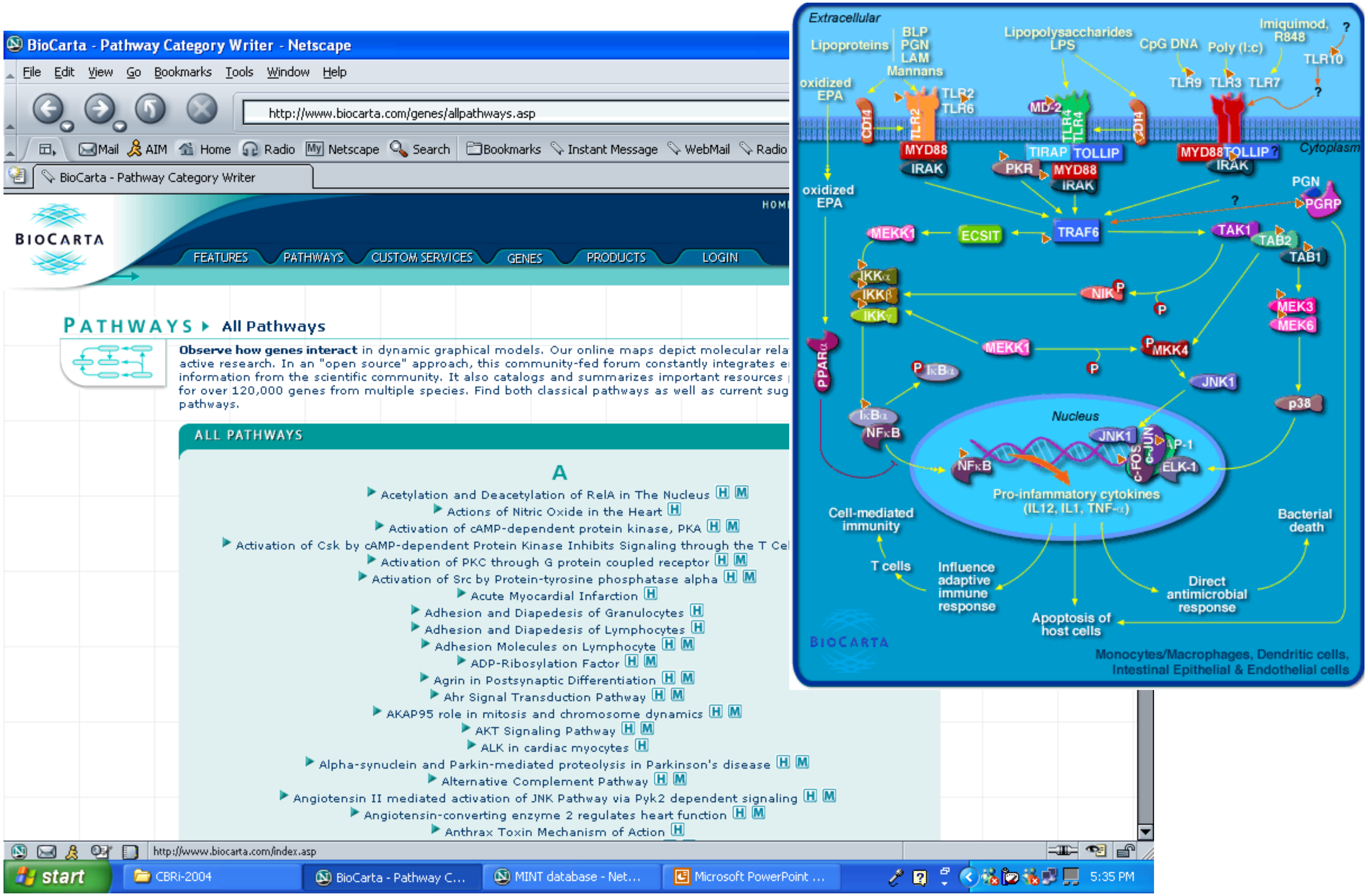
Welcome to Cytoscape 2.3 Right-click + drag to ZOOM Middle-click + drag to PAN



Cytoscape (www.cytoscape.org)

Osprey http://biodata.mshri.on.ca/osprey/servlet/Index

Pathway Visualization with BioCarta*



PATHWAYS ▸ **All Pathways**

Observe how genes interact in dynamic graphical models. Our online maps depict molecular rela active research. In an "open source" approach, this community-fed forum constantly integrates e information from the scientific community. It also catalogs and summarizes important resources for over 120,000 genes from multiple species. Find both classical pathways as well as current sug pathways.

ALL PATHWAYS

A

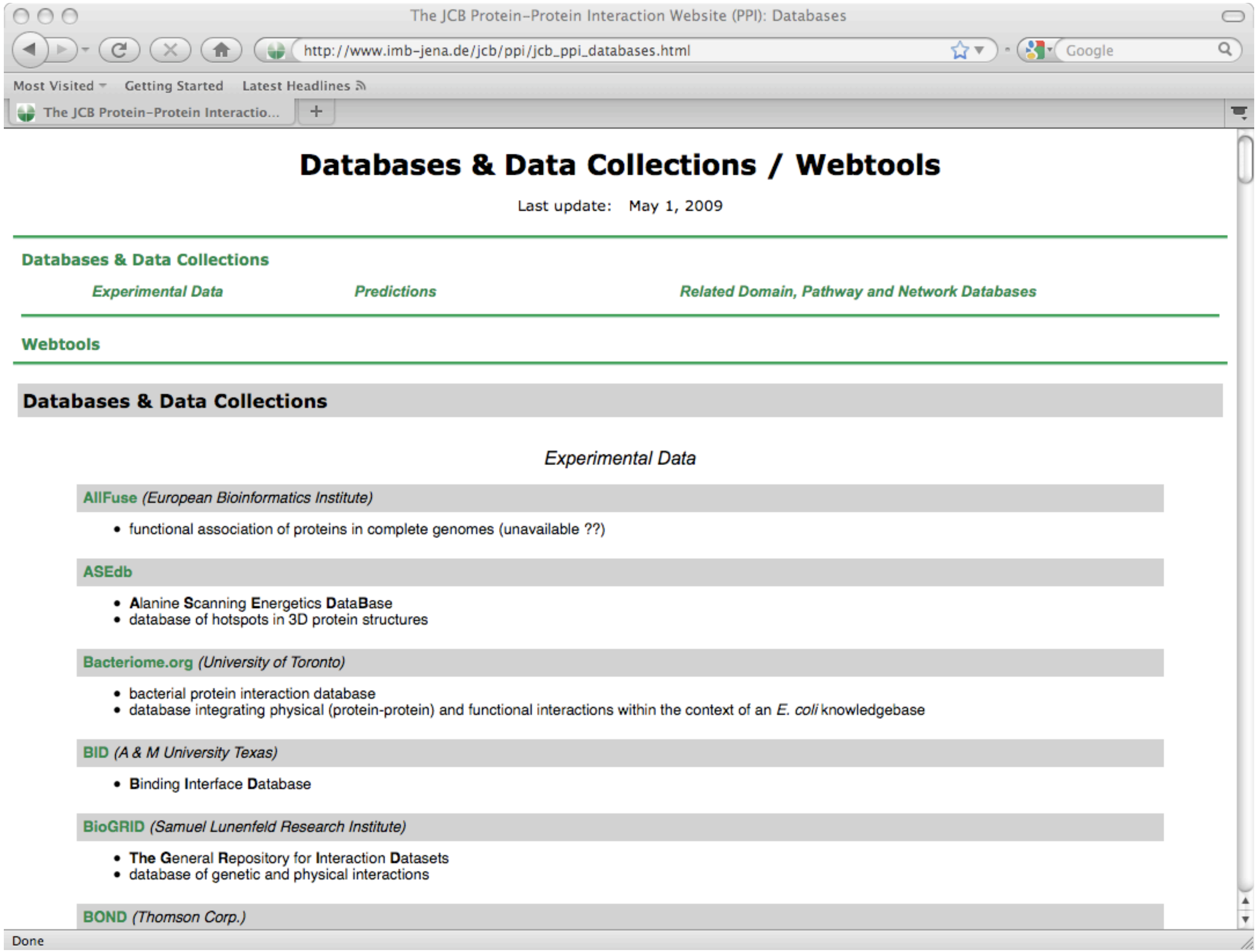
- ▶ Acetylation and Deacetylation of RelA in The Nucleus H M
- ▶ Actions of Nitric Oxide in the Heart H
- ▶ Activation of cAMP-dependent protein kinase, PKA H M
- ▶ Activation of Csk by cAMP-dependent Protein Kinase Inhibits Signaling through the T Cel
- ▶ Activation of PKC through G protein coupled receptor H M
- ▶ Activation of Src by Protein-tyrosine phosphatase alpha H M
- ▶ Acute Myocardial Infarction H
- ▶ Adhesion and Diapedesis of Granulocytes H
- ▶ Adhesion and Diapedesis of Lymphocytes H
- ▶ Adhesion Molecules on Lymphocyte H M
- ▶ ADP-Ribosylation Factor H M
- ▶ Agrin in Postsynaptic Differentiation H M
- ▶ Ahr Signal Transduction Pathway H M
- ▶ AKAP95 role in mitosis and chromosome dynamics H M
- ▶ AKT Signaling Pathway H M
- ▶ ALK in cardiac myocytes H
- ▶ Alpha-synuclein and Parkin-mediated proteolysis in Parkinson's disease H M
- ▶ Alternative Complement Pathway H M
- ▶ Angiotensin II mediated activation of JNK Pathway via Pyk2 dependent signaling H M
- ▶ Angiotensin-converting enzyme 2 regulates heart function H M
- ▶ Anthrax Toxin Mechanism of Action H

http://www.biocarta.com/genes/allpathways.asp

Pathway Database Comparison*

	KEGG	BioCyc	GenMAPP	Reactome	BioCarta	TransPATH
Organisms	181 (varied)	E.Coli, human (20 others)	Human, mouse, rat, fly, yeast	Human, rat, mouse, chicken, fugu, zebrafish	Human, mouse	Human, mouse
Pathway types	Metabolic, genetic, signaling, complexes	Metabolic, complexes	Metabolic, signaling, complexes	Metabolic, signaling, complexes	Metabolic, signaling, complexes	Signaling, genetic
Tools/ viewing	linked to from many	Pathway Tools	GenMAPP	PathView applets	none	Pathway Builder
Images	Static box flow diagrams	Detailed flow diagrams	Static box flow diagrams	“starry sky”	“Graphics rich” cell diagrams	Graphics rich cell diagrams
Download Formats	KGML XML SBML	BioPax SBML	MAPP format	SBML MySQL	Just images	Propietary XML files

Other Databases



Databases & Data Collections / Webtools

Last update: May 1, 2009

Databases & Data Collections

Experimental Data *Predictions* *Related Domain, Pathway and Network Databases*

Webtools

Databases & Data Collections

Experimental Data

AllFuse *(European Bioinformatics Institute)*

- functional association of proteins in complete genomes (unavailable ??)

ASEdb

- **A**lanine **S**canning **E**nergetics **D**ata**B**ase
- database of hotspots in 3D protein structures

Bacteriome.org *(University of Toronto)*

- bacterial protein interaction database
- database integrating physical (protein-protein) and functional interactions within the context of an *E. coli* knowledgebase

BID *(A & M University Texas)*

- **B**inding **I**nterface **D**atabase

BioGRID *(Samuel Lunenfeld Research Institute)*

- **The G**eneral **R**epository for **I**nteraction **D**atasets
- database of genetic and physical interactions

BOND *(Thomson Corp.)*

http://www.imb-jena.de/jcb/ppi/jcb_ppi_databases.html

Functional Proteomics

- Mixture of experimental and computational techniques
- Trying to reach a point where functions and interactions can be predicted and modelled
- The future of proteomics (and bioinformatics)

Final Exam

- Short answer to long answer format
- Bring calculators
- Typically one question from each of the lectures in the last ½ of the course
- Some questions/answers will involve recall
- Most questions require analysis or some thinking or explaining
- Dec. 13, 9:00 am - 2 hours not 3 hours
- This room, M-229

Typical Questions

- What is the correlation between protein expression and transcript expression? Provide three reasons to explain the difference
- Describe the algorithm or diagram a flow chart for XXXXX
- Explain the differences and similarities between functional proteomics and structural proteomics

Typical Questions

- Here is some YYYY data from some XXXX experiment – interpret it and explain what it means
- Explain the difference between the XXX algorithm and the YYY algorithm. Give some examples or provide an illustration
- Here are two small molecules, calculate their difference distance matrix, show calculations. What is the difference between the two?

Typical Questions

- Define normalization. Provide 3 examples. Show equations or algorithms
- What are the three different kinds of proteomics, compare and contrast
- Show the equations and explain the algorithm you would use to rotate, expand and translate this small molecule