Standards in Proteomics
Protein-Protein Interactions

Christopher W.V. Hogue, Ph.D.
Scientist, Samuel Lunenfeld Research Inst. Mount Sinai Hospital
Associate Professor, Dept. of Biochemistry, U. of Toronto
Outline

- OECD Declaration – Data Access/Archiving
- About Blueprint and BIND
- On Standards, Components and Systems
- Reverse Salients
- Communities
- User Behavior
- OntoGlyphs - Visualizing Complex Information
- BIND PICKS – Scoring Yeast Interactions
- New BIND Record Types
DECLARATION ON ACCESS TO RESEARCH DATA FROM PUBLIC FUNDING

adopted on 30 January 2004 in Paris

The governments (1) of Australia, Austria, Belgium, Canada, China, the Czech Republic, Denmark, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Israel, Italy, Japan, Korea, Luxembourg, Mexico, the Netherlands, New Zealand, Norway, Poland, Portugal, the Russian Federation, the Slovak Republic, the Republic of South Africa, Spain, Sweden, Switzerland, Turkey, the United Kingdom, and the United States
Declare their commitment to

- Work towards the establishment of access regimes for digital research data from public funding in accordance with the following objectives and principles:
  - Openness
  - Transparency
  - Legal Conformity
  - Formal Responsibility
  - Professionalism
  - IP Protection
  - Interoperability
  - Quality and Security
  - Efficiency
  - Accountability
Q13. Archive? Form? Where?

- OECD Declaration in Canada was disseminated top-down to the granting agencies in Nov from the government.
- Mandate to sort out the problem of archiving all scientific research data.
- Timeline in 5 years – inter-agency coordination required.
A tea cup in a rainstorm...

- 2000 elemental observations (facts) about molecular assembly (interactions) published in the literature every month.

- By 1965 standards - 10 textbooks full of figures, descriptions, mechanisms every year

- Estimate 200,000 facts sitting in the literature on library shelves, not validated, not useable.

- Blueprint’s mandate is to capture this information in a machine readable database called BIND.
The Blueprint Initiative

- Develop, curate and maintain the Biomolecular Interaction Network Database (BIND) and related tools
- Carry out bioinformatics research in support of our vision.
BIND stores molecular interaction data...

**BIND Interaction Types**

- **Protein - Protein**: 54%
- **Protein - DNA**: 25%
- **Protein - Not Specified**: 12%
- **Other**: 9%
- **Protein - RNA**: 1%
- **Protein - Small Molecule**: 1%
- **Small Molecule - Gene**: 1%
- **Gene - Gene**: 4%

---

[Image: Mount Sinai Hospital, Blueprint Logo]
Interaction Experimental Evidence Captured

- Affinity Chromatography: 8%
- Three Dimensional Structure: 20%
- Two Hybrid Test: 38%
- Cross Linking: 25%
- SGA: 8%
- Other: 1%
- Not Specified: 6%
- Gel Filtration Chromatography: 14%
- Immunostaining: 9%
- Fluorescence Anisotropy: 6%
- Elisa: 6%
- Equilibrium Dialysis: 16%
- Microarray: 14%
- Light Scattering: 11%
- Resonance Energy Transfer: 9%
- Not Specified: 6%
- Other: 8%
- Electron Microscopy: 2%
- Competition Binding: 1%
- Gel Retardation Assays: 1%
- Gradient Sedimentation: 1%
- Colocalization: 1%
Standards in Proteomics?

- **Software Systems Components (OSI Layers...)**
  - Human Interfaces
  - Application Programming Interfaces
  - Communications Protocols
  - Content Structure
  - Database (ODBC/JDBC compliant MySQL)
  - Document Structures (XML)
  - Architectures (Compatible orchestration of the above)
  - Platforms (Runs the above: Windows, Linux, Unix)
What is in a “Standard”
A Historical Perspective

- Standards emerge from successful systems.
- Which one is “the standard” The light bulb – or the electrical grid?
- Lamps were the original killer app.
  - (bye-bye candles, gas lamps, oil lamps)
  - Other Apps: Motors, Heaters, Toasters
  - Unexpected Apps: radio, TV, transformers, computers, rechargables
- Entire “systems” become standards via ad-hoc and popular use – snowball effect.
Emergence and evolution of technological systems...

- Systems emerge across broad frontiers
  - Lots of small inventions are responsible for emerging technologies.

- Portions of the frontier that are held back become the focus of intense innovation
  - Called a “reverse salient” by students of technology
  - An inadequately functioning or accessible component in a complex system of components
  - Opportunities for invention and replacement
Reverse Salient – AC/DC Example

- Edison’s DC standard lit up Wall Street in Manhattan
- High-level buy-in for DC.
- AC was too complicated, could kill a person!
- Edison’s DC system only worked over short-range.
- This flaw is the “reverse salient”.

- Westinghouse/Stanley/Tesla saw the flaw in this standard
- AC technology raced to fill the gap
- Light bulbs work with both AC or DC.
- Motors required re-invention
- E.S. Rogers “batteryless” radio

Result: Cars, Battery based devices emerged with DC.
Reverse Salient – AC/DC Example

- Edison’s DC standard lit up Wall Street in Manhattan
- High-level buy-in for DC.
- AC was too complicated, could kill a person!
- Edison’s DC system only worked over short-range.
- This flaw is the “reverse salient”.

- Westinghouse/Stanley/Tesla saw the flaw in this standard
- AC technology raced to fill the gap
- Light bulbs work with both AC or DC.
- Motors required re-invention
- E.S. Rogers “batteryless” radio

- Result: Cars, Battery based devices emerged with DC.
- Result: The electrical Grid emerged with AC.

NOT A WINNER-TAKE-ALL RESULT!
Reverse Salient Attitudes

What holds us back?

- Oversights (didn’t think of that!).
- Shortsightedness (won’t ever need that!).
- Inability (can’t do it!)
- Stubbornness (won’t do it!)
- Prescriptivism (do it like this!)
- Nationalism, Continentalism, Colonialism
  - (because that’s the way we do it here!)
  - 110 vs 220
Q1. Communities – who matters?


- Database Communities
  - IMEX (BIND/DIP/INTACT/MINT/MIPS)
  - BioPAX (pathway databases)
  - SBML (>70 software systems collaborating)
  - Cytoscape (collaborating interface developers)
  - NCBI/Blueprint (architecture)
  - Model Organism Databases (GMOD architecture)

- Journals and Editors
- Scientific Societies (HUPO)
- Member and Non-member Scientists
Q7. Integrate/Synergize

- Identify the communities.
- Recognize that communities are disjoint.
- Require funded efforts to show their efforts to collaborate with and integrate across the spectrum of identified communities.
Q3? Standard Improvement?

- Service all communities effectively with a whole system.
- Drive innovation more through use.
- Gain and effectively incorporate user critique.
- Understand user needs, behaviors.
User Behavior

- The problem of too much choice.
  - (M. Lepper @Stanford and S. Iyengar @Columbia)

- Two tables in a supermarket:
  - 24 jars of jam vs 6 jars of jam.
  - 3% vs 30%

- Choice frustration.

- Can't Debug - This jam is a little bitter compared to the other 26?
- A whole lot of bad jam that nobody wants to buy…
User Behavior

- The problem of too much choice.
  - (M. Lepper @Stanford and S. Iyengar @Columbia)
- Two tables in a supermarket:
  - 24 jars of jam vs 6 jars of jam.
  - 3% vs 30%
- Choice frustration.

- Leads to incrementalism
- Essential user criticism is withdrawn.
  - Can’t Debug - This jam is a little bitter compared to
    - the other 6?
    - the other 26?
  - A whole lot of bad jam that nobody wants to buy…
Q4? Main Problems?

- Standards Fatigue
  - Interactions/Pathways since NIH meeting in Nov 1999. Efforts are still not integrated (PSI/IMEX and BIOPAX).

- Data Standards are not an effective goal to achieve results in a timely way

- Information Systems are better goals.

- Wet Lab Scientists are busy people who are (excuse me) trying to write papers.

- Ongoing wishful thinking about latest new technology (the semantic web will fix everything!)
Q8. Proteomic Dictionary (semantics)

Isn’t that what Gene Ontology is – a collection of terms about proteins?

>17,000 terms – ½ the size of a pocket dictionary.

Structured data curation vs term-tagging. Both are required.
OntoGlyphs

- A graphical language
- Derived from Gene Ontology annotation
- The most-used terms/categories
- Simpler – Fewer Choices
- Summarize Long, long long long long long long long long long long lists of data results – like query “cancer”
## Blueprint

### ONTOGLYPHS

Symbols to display protein attributes schematically. A collection of glyphs consisting of 34 functional, 25 binding and 24 location categories populated with Gene Ontology (GO) terms.

### Functional Categories

<table>
<thead>
<tr>
<th>Category</th>
<th>Glyph</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alcohol &amp; aromatic compound metabolism/transport</td>
<td><img src="image" alt="Glyph" /></td>
</tr>
<tr>
<td>Behaviour</td>
<td><img src="image" alt="Glyph" /></td>
</tr>
<tr>
<td>Binding activity</td>
<td><img src="image" alt="Glyph" /></td>
</tr>
<tr>
<td>Biomembranes: organization and basic functions</td>
<td><img src="image" alt="Glyph" /></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Category</th>
<th>Glyph</th>
</tr>
</thead>
<tbody>
<tr>
<td>Carbohydrate metabolism/transport</td>
<td><img src="image" alt="Glyph" /></td>
</tr>
<tr>
<td>Cell communication</td>
<td><img src="image" alt="Glyph" /></td>
</tr>
<tr>
<td>Cell motility and structural activity</td>
<td><img src="image" alt="Glyph" /></td>
</tr>
<tr>
<td>Cell multiplication</td>
<td><img src="image" alt="Glyph" /></td>
</tr>
<tr>
<td>Cell physiology</td>
<td><img src="image" alt="Glyph" /></td>
</tr>
<tr>
<td>Cellular organization/biogenesis</td>
<td><img src="image" alt="Glyph" /></td>
</tr>
<tr>
<td>Coenzymes vitamin &amp; hormone metabolism/transport</td>
<td><img src="image" alt="Glyph" /></td>
</tr>
<tr>
<td>Death</td>
<td><img src="image" alt="Glyph" /></td>
</tr>
<tr>
<td>Defence/immune response</td>
<td><img src="image" alt="Glyph" /></td>
</tr>
<tr>
<td>Development</td>
<td><img src="image" alt="Glyph" /></td>
</tr>
<tr>
<td>DNA metabolism</td>
<td><img src="image" alt="Glyph" /></td>
</tr>
<tr>
<td>Energy production/conversion</td>
<td><img src="image" alt="Glyph" /></td>
</tr>
<tr>
<td>General metabolism</td>
<td><img src="image" alt="Glyph" /></td>
</tr>
<tr>
<td>General physiological processes</td>
<td><img src="image" alt="Glyph" /></td>
</tr>
<tr>
<td>General transport</td>
<td><img src="image" alt="Glyph" /></td>
</tr>
<tr>
<td>Ion transport</td>
<td><img src="image" alt="Glyph" /></td>
</tr>
<tr>
<td>Lipid metabolism/transport</td>
<td><img src="image" alt="Glyph" /></td>
</tr>
<tr>
<td>Organic &amp; amino acids &amp; derivatives &amp; amine metabolism/transport</td>
<td><img src="image" alt="Glyph" /></td>
</tr>
<tr>
<td>Pathogenesis &amp; toxin activity</td>
<td><img src="image" alt="Glyph" /></td>
</tr>
<tr>
<td>Phosphorus metabolism</td>
<td><img src="image" alt="Glyph" /></td>
</tr>
<tr>
<td>Protein degradation</td>
<td><img src="image" alt="Glyph" /></td>
</tr>
<tr>
<td>Protein metabolism</td>
<td><img src="image" alt="Glyph" /></td>
</tr>
<tr>
<td>Protein modification</td>
<td><img src="image" alt="Glyph" /></td>
</tr>
<tr>
<td>Protein Synthesis, Processing, and Regulation</td>
<td><img src="image" alt="Glyph" /></td>
</tr>
<tr>
<td>Protein transport</td>
<td><img src="image" alt="Glyph" /></td>
</tr>
<tr>
<td>Response to stress endogenous or exogenous stimulus</td>
<td><img src="image" alt="Glyph" /></td>
</tr>
<tr>
<td>RNA &amp; other nucleic acids metabolism and transport</td>
<td><img src="image" alt="Glyph" /></td>
</tr>
<tr>
<td>Signal transduction</td>
<td><img src="image" alt="Glyph" /></td>
</tr>
<tr>
<td>Transcription</td>
<td><img src="image" alt="Glyph" /></td>
</tr>
<tr>
<td>Viral life cycle</td>
<td><img src="image" alt="Glyph" /></td>
</tr>
</tbody>
</table>
### BINDING CATEGORIES

<table>
<thead>
<tr>
<th>Category</th>
<th>Glyph</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adenyl nucleotide binding</td>
<td>![Glyph]</td>
</tr>
<tr>
<td>Antigen binding</td>
<td>![Glyph]</td>
</tr>
<tr>
<td>ATP binding</td>
<td>![Glyph]</td>
</tr>
<tr>
<td>Calcium ion binding</td>
<td>![Glyph]</td>
</tr>
<tr>
<td>Calmodulin binding</td>
<td>![Glyph]</td>
</tr>
<tr>
<td>Carbohydrate binding</td>
<td>![Glyph]</td>
</tr>
<tr>
<td>Cytokine binding</td>
<td>![Glyph]</td>
</tr>
<tr>
<td>Cytoskeletal protein binding</td>
<td>![Glyph]</td>
</tr>
<tr>
<td>Damage DNA binding</td>
<td>![Glyph]</td>
</tr>
<tr>
<td>DNA binding</td>
<td>![Glyph]</td>
</tr>
<tr>
<td>Double stranded DNA binding</td>
<td>![Glyph]</td>
</tr>
<tr>
<td>Guanyl nucleotide binding</td>
<td>![Glyph]</td>
</tr>
<tr>
<td>Lipid binding</td>
<td>![Glyph]</td>
</tr>
<tr>
<td>Metal ion binding</td>
<td>![Glyph]</td>
</tr>
<tr>
<td>mRNA binding</td>
<td>![Glyph]</td>
</tr>
<tr>
<td>Nucleic acid binding</td>
<td>![Glyph]</td>
</tr>
<tr>
<td>Nucleotide binding</td>
<td>![Glyph]</td>
</tr>
</tbody>
</table>

### SUB-CELLULAR LOCALIZATION CATEGORIES

<table>
<thead>
<tr>
<th>Category</th>
<th>Glyph</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oxygen binding</td>
<td>![Glyph]</td>
</tr>
<tr>
<td>Protein binding</td>
<td>![Glyph]</td>
</tr>
<tr>
<td>Purine nucleotide binding</td>
<td>![Glyph]</td>
</tr>
<tr>
<td>Receptor binding</td>
<td>![Glyph]</td>
</tr>
<tr>
<td>RNA binding</td>
<td>![Glyph]</td>
</tr>
<tr>
<td>Single stranded DNA binding</td>
<td>![Glyph]</td>
</tr>
<tr>
<td>Transcription factor binding</td>
<td>![Glyph]</td>
</tr>
<tr>
<td>Transition metal ions binding</td>
<td>![Glyph]</td>
</tr>
<tr>
<td>Actin cytoskeleton</td>
<td>![Glyph]</td>
</tr>
<tr>
<td>Axon or dendrite</td>
<td>![Glyph]</td>
</tr>
<tr>
<td>Biological membrane</td>
<td>![Glyph]</td>
</tr>
<tr>
<td>Cell periphery</td>
<td>![Glyph]</td>
</tr>
<tr>
<td>Cytoplasm</td>
<td>![Glyph]</td>
</tr>
<tr>
<td>Cytoplasmic vesicle</td>
<td>![Glyph]</td>
</tr>
<tr>
<td>Endoplasmic reticulum</td>
<td>![Glyph]</td>
</tr>
<tr>
<td>Endosome</td>
<td>![Glyph]</td>
</tr>
<tr>
<td>Extracellular /cell surface</td>
<td>![Glyph]</td>
</tr>
<tr>
<td>Flagellum /cilium</td>
<td>![Glyph]</td>
</tr>
<tr>
<td>Golgi apparatus</td>
<td>![Glyph]</td>
</tr>
<tr>
<td>Lipid particle</td>
<td>![Glyph]</td>
</tr>
<tr>
<td>Microtubule cytoskeleton</td>
<td>![Glyph]</td>
</tr>
<tr>
<td>Mitochondrion</td>
<td>![Glyph]</td>
</tr>
<tr>
<td>Nuclear periphery</td>
<td>![Glyph]</td>
</tr>
<tr>
<td>Nucleolus</td>
<td>![Glyph]</td>
</tr>
<tr>
<td>Nucleus</td>
<td>![Glyph]</td>
</tr>
<tr>
<td>Peroxisome</td>
<td>![Glyph]</td>
</tr>
<tr>
<td>Plastid</td>
<td>![Glyph]</td>
</tr>
<tr>
<td>Protoplasm</td>
<td>![Glyph]</td>
</tr>
<tr>
<td>Ribonucleoprotein complex</td>
<td>![Glyph]</td>
</tr>
<tr>
<td>Site of polarized growth.</td>
<td>![Glyph]</td>
</tr>
<tr>
<td>Vacuole</td>
<td>![Glyph]</td>
</tr>
<tr>
<td>Virion provirus.</td>
<td>![Glyph]</td>
</tr>
</tbody>
</table>
Results - Cancer query – interactions only – restricted to BIND-metazoa + MGI divisions

Single-Line OntoGlyphs now appear on BIND search results...
Any list of genes can be converted into 3 lines of ontoglyphs. BIND query, BLAST output, microarray data...

Search terms: cancer

<table>
<thead>
<tr>
<th>function</th>
<th>47 33 31 30 29 24 19 19 13 9 6 6 5 5 4 3 3 3 3 2 2 2 2 1 1 1 1 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>binding</td>
<td>56 46 28 22 17 15 8 7 6 5 3 3 2 2 2 1 1 1 1</td>
</tr>
<tr>
<td>location</td>
<td>98 36 14 11 10 8 5 5 3 3 3 3 2 1 1 1 1</td>
</tr>
</tbody>
</table>

MTMT SINAI HOSPITAL

Blueprint
Q6. Main Concerns to be Addressed

- The entire 300-year old publish-or-perish reward model.
  - Papers have to be stories, not observations (data)
  - No reward for novel, unique data from isolated experiments
  - No capture into print or into databases
Q12. Should or can journals enforce submission of data?

- BIND is working with 23 Journal titles currently including Science, NPG, Cell Press, NRC Press on pre and post-publication capture of interactions. We are working to extend these relationships with our IMEX partners.

- Yes, it can be done. The devil is in the details – it requires very focused and respectful consideration of the need of journal editors, and manuscript submitters. It needs trained curators.

- No scientist wants to have to make an XML document in addition to their paper. Training required is a large burden.
  - Tim Hughes - MIAME frustration – echoed by journal editors.

- BIND curators structure the data for the submitters, not just tag it with semantic terms. BIND data is intended to be computable in the long-term, justifying the effort.
Data Archiving vs Data Analysis

- Effort is not balanced
  - Data Generation >> Data Analysis
  - Data hoarding results...

- Archivists are expected to behave like
  - Librarians? (Archive the data)
  - Literary Critics? (Evaluate/Rank the data)
Many hits from yeast-two-hybrid data.
BIND Interaction

BIND Id: 5982

Interaction Description: No description provided

Publication: 1

PMB: 123456

Option: Support

Article Title: A comprehensive bio-computing analysis to explore the yeast protein interaction.

Authors:
- Ita T
- Chiba T
- Cohen R
- Yosef N
- Ishio M
- Nakai K

Affiliation: Division of Genome Biology, Cancer Research Institute, Kansazawa University, Kansazawa 920-0924, Japan.

Abstract: Protein-protein interactions play crucial roles in the execution of various biological...

Molecule A

Protein: Sua7

Description: Transcription factor, TFIIH component

GeneInfo ID: 6325343

Origin: Organism - Saccharomyces cerevisiae

GO Terms: 4 Molecular Function(s)
- Cellular Component(s)
- Biological Process(es)

Cellular Location: 1 cellular location(s)

Molecule B

Protein: Atf14

Description: ATP synthase subunit h

GeneInfo ID: 6325354

Origin: Organism - Saccharomyces cerevisiae

GO Terms: 2 Molecular Function(s)
- Cellular Component(s)
- Biological Process(es)

Cellular Location: 1 cellular location(s)
Too many...

Which molecules are co-localized with the atp14?
Invert
Voila
Only co-localized Proteins!
Scoring High Throughput Interactions

- **BIND PICKS** – Protein Interaction Confidence
  Kernel Scores – SVM Classifier for Yeast.

<table>
<thead>
<tr>
<th>id</th>
<th>eigen0</th>
<th>eigen1</th>
<th>eigen2</th>
<th>eigen3</th>
<th>eigen4</th>
</tr>
</thead>
<tbody>
<tr>
<td>num_pos_homolog_interactions</td>
<td>-0.002262</td>
<td>-0.374565</td>
<td>0.002792</td>
<td>-0.080864</td>
<td>-0.104295</td>
</tr>
<tr>
<td>num_homolog_interactions</td>
<td>-0.130207</td>
<td>0.533672</td>
<td>-0.52235</td>
<td>-0.223424</td>
<td>-0.561946</td>
</tr>
<tr>
<td>num_publications</td>
<td>-0.449107</td>
<td>-0.242582</td>
<td>-0.092797</td>
<td>-0.221117</td>
<td>0.090698</td>
</tr>
<tr>
<td>num_prebind_interactions</td>
<td>-0.258069</td>
<td>0.153001</td>
<td>-0.404787</td>
<td>-0.29502</td>
<td>0.661245</td>
</tr>
<tr>
<td>num_common_func_go</td>
<td>-0.455811</td>
<td>-0.166435</td>
<td>0.19885</td>
<td>0.194212</td>
<td>-0.12717</td>
</tr>
<tr>
<td>num_common_proc_go</td>
<td>-0.468985</td>
<td>-0.297926</td>
<td>0.064311</td>
<td>0.020879</td>
<td>-0.124666</td>
</tr>
<tr>
<td>num_common_com_go</td>
<td>-0.488325</td>
<td>-0.297634</td>
<td>0.172517</td>
<td>-0.009326</td>
<td>-0.082158</td>
</tr>
<tr>
<td>ma_binding_dom</td>
<td>-0.063564</td>
<td>0.654886</td>
<td>0.726009</td>
<td>-0.235928</td>
<td>0.074127</td>
</tr>
<tr>
<td>dna_binding_dom</td>
<td>0.197966</td>
<td>0.339412</td>
<td>-0.152669</td>
<td>0.80965</td>
<td>0.154268</td>
</tr>
<tr>
<td>prox_func_go</td>
<td>0.841798</td>
<td>-0.095478</td>
<td>0.004688</td>
<td>0.017365</td>
<td>0.006152</td>
</tr>
<tr>
<td>prox_proc_go</td>
<td>0.837514</td>
<td>-0.100904</td>
<td>0.001824</td>
<td>0.012879</td>
<td>0.007085</td>
</tr>
<tr>
<td>prox_com_go</td>
<td>0.834984</td>
<td>-0.105447</td>
<td>0.001614</td>
<td>0.010693</td>
<td>0.006661</td>
</tr>
</tbody>
</table>

**Table 12:** Eigen vectors from PCA analysis

First five eigen vectors (principal components) and eigen values generated by the principal component analysis. Highlighted cells indicate variables that contribute most to the value of that principal component.
Figure 15: PCA

Positive subset (magenta) and negative subset (blue) were mapped onto to the first three principal components and graphed.
Skolnick Dataset

59% < 1

DIP core Dataset

93% > 0
30% < 1

MIPS Dataset

94% > 0
Von Mering Scores vs. BIND PICKS

- High – supported by more than one method.
- Low – Seen only once in one method.

Total = 65,731 (Matrix)
Score > 0 = 24%

Total = 2445
Score > 0 = 68%

Medium – 47%
Gerstein Scores vs. BIND PICKS

Total = 9897
Score > 0 = 57%

Total = 163
Score > 0 = 99%
Reliable “spoke” HTP data? 36% >1

Figure 18: Distribution of BIND PICKS scores for HTP yeast interaction subset. The percentage of interactions that were identified as true interactions (score >1) was 36%.
HTP spoke interactions
Above score = 5
RED – positives
from SVM training set
BIND 4.0 – New Record Types
Supporting Pathway/Proteomics Data

- 3 Types in use:
  - Interaction, Complex, Pathway

- Improvements to Complex, Pathway
  - EcoCyc, aMaze, PID, STKE

- 2 high-level types being added in v4.0
  - Reactant List
  - Assembly List
New BIND record types

- Reactant List: In-vitro reagents required to elicit activity (e.g. Mg, ATP, etc).
  - Arises from an active collaboration with pathway dbs, data analysis and mapping, and efforts to integrate their data into the BIND query system.

- Assembly List: Observed List of Proteins/Genes (non-ratiometric!)
  - Currently unarchived data.
  - Phosphoproteomics, localization (e.g. human nucleolar proteins), transcription factor target proteins, concentration/copy number experiments.
  - May be suitable for “GeneString” conversion.
Staff and Contributors
