

# **Systems Biology and Biomolecular Interaction Data**

**CBW Bioinformatics Workshop  
February 26th, Vancouver**

**Ian Donaldson  
Blueprint Initiative**



# Blueprint

- A multi-year research program that will develop, operate and maintain a free and publicly accessible biomolecular interaction database called BIND (Biomolecular Interaction Network Database)



# Blueprint

- Blueprint North America
- Based in Toronto
- A confirmed 3 Year Work Program
- A secured Cdn \$29 million budget and all required funds – from government and private partners
- A 74 person workforce at scale up in Year 3  
(40 curators/24 programmers/10 administrators)
- Will index 80,000 published and directly received interactions into BIND over three years



# Blueprint

- Blueprint Asia
- Based in Singapore
- 5 Year Work Program
- S\$23 million budget (CDN \$ 20 M)
- 37 person workforce  
(28 curators/5 programmers/4 administrators)
- Indexing 60,000 interactions into BIND over 5 years
- Start up in Q2/2004

# About this talk

- Why interaction data are important.
- A quick tour of BIND.
- Methods used to generate interaction data.
- High-throughput interaction data and representation.

# A general definition of life

- A life form is defined by the following properties:
- It is distinct from its physical surroundings.
- It uses (changes) parts of its physical surroundings.
- It responds to changes in its physical surroundings.
- It is able to change the way that it responds.
- It reproduces itself.

# The end goal of biology is to discover how life works.

- How do you discover how something “works”?
- Observe it.
- Poke it.
- Take it apart.
- Put it back together again.
  
- Biology is a collection of methods that allow you to do these things.

# What are the parts

- DNA
- RNA
- Proteins
- Small molecules
- Complexes

# How do the parts work?

- A biomolecule's function can be defined by the things that it interacts with and the new (or altered) molecules that result from that interaction.
- Like this...

# Biomolecular function

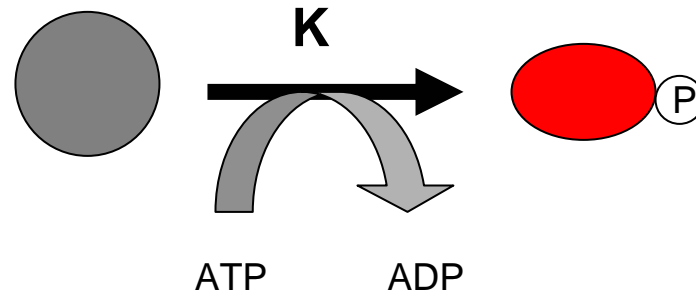


- This is a generalization of how a biochemist might represent the function of enzymes.

# Biomolecular function

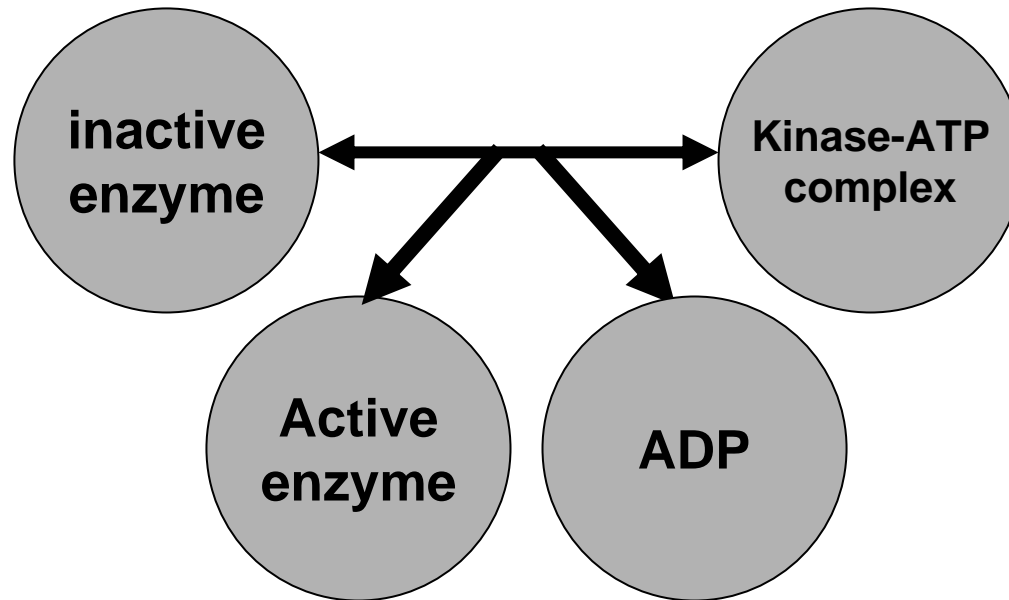


kinase-ATP complex + inactive-enzyme  $\rightleftharpoons$  Kinase + ADP + active enzyme



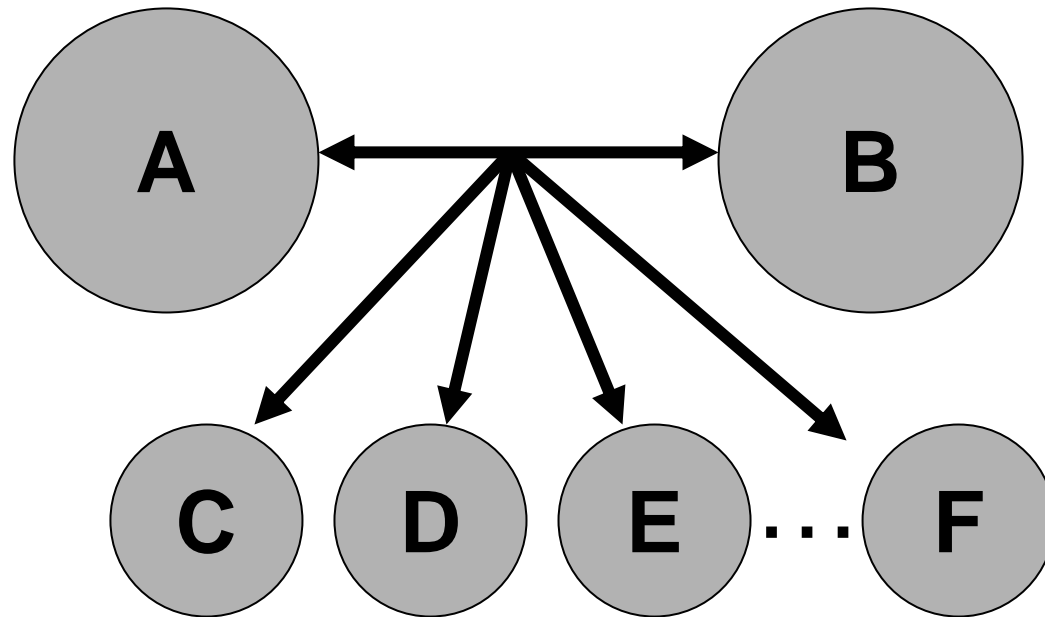
- Here is an example of the generalization represented two different ways.

# Biomolecular function



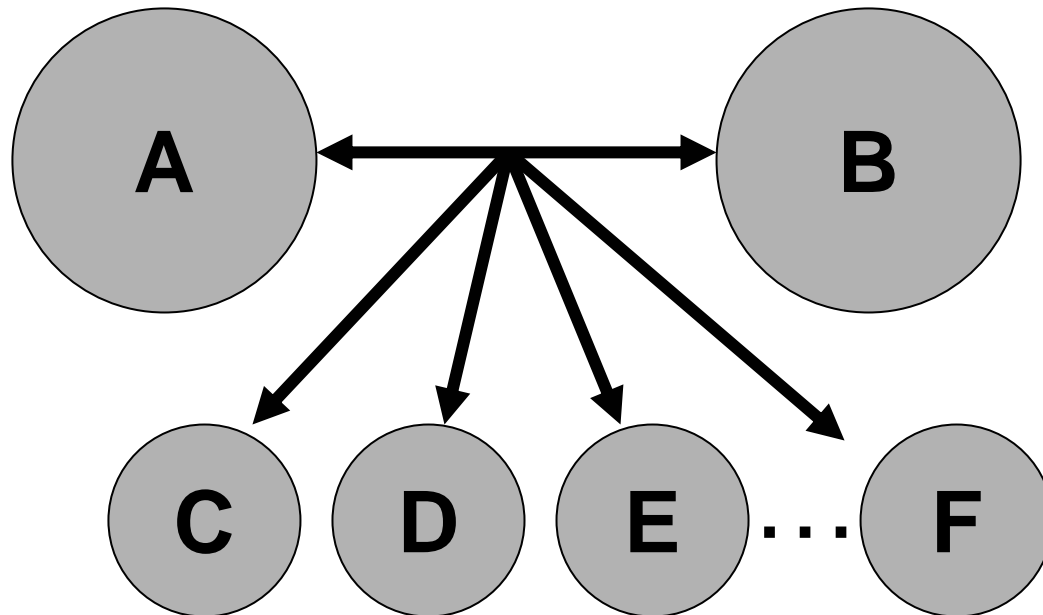
- This is another representation.

# Biomolecular function



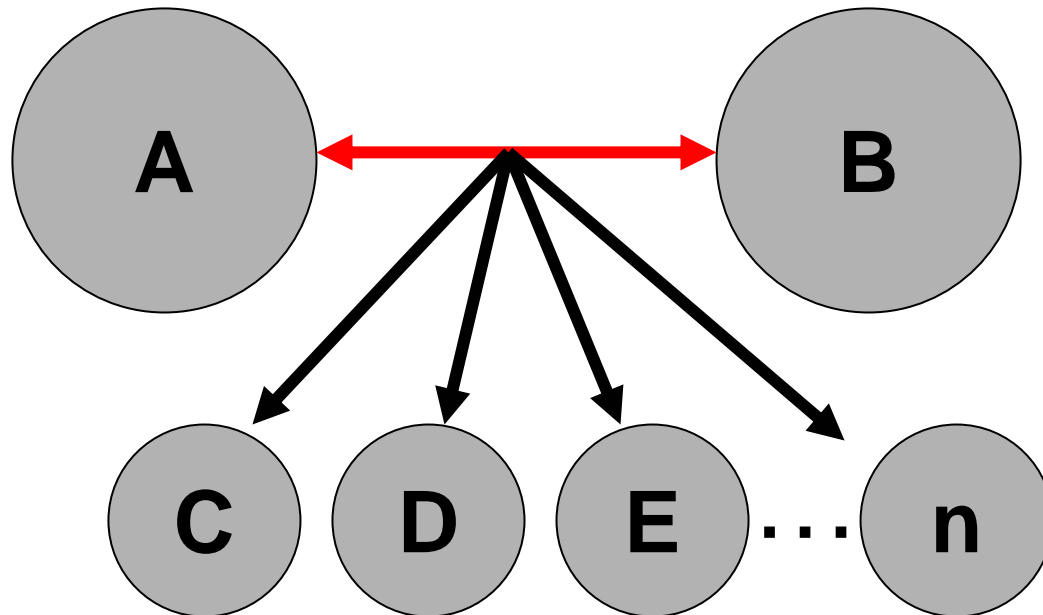
- This is a generalization of the representation.

# Biomolecular function



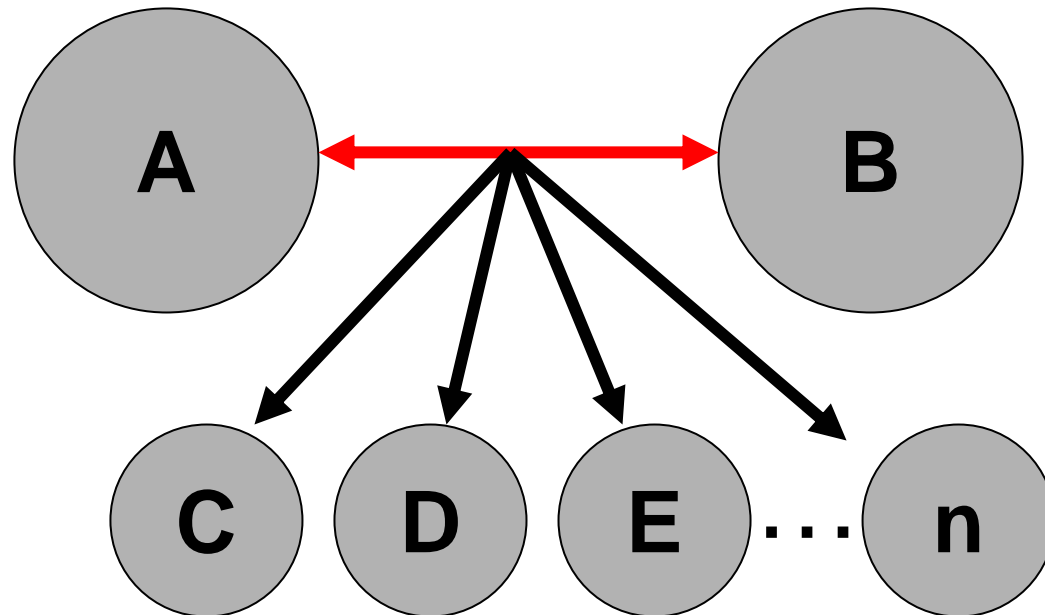
- A biomolecules function can be defined by the things that it interacts with and the new (or altered) molecules that result from that interaction.

# Biomolecular function



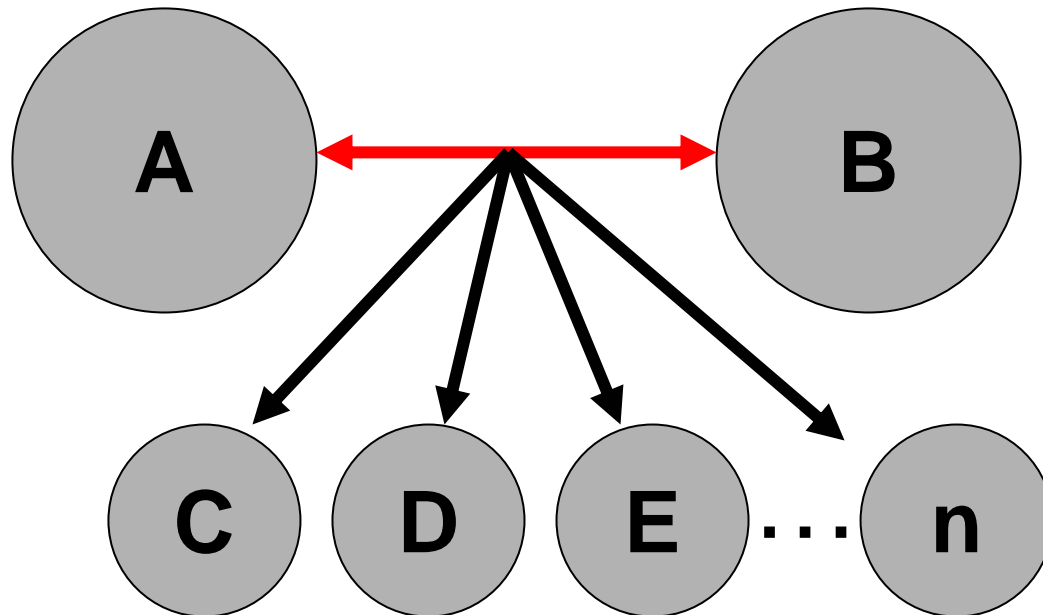
- This representation makes it easy to focus on the interaction part.

# Biomolecular function



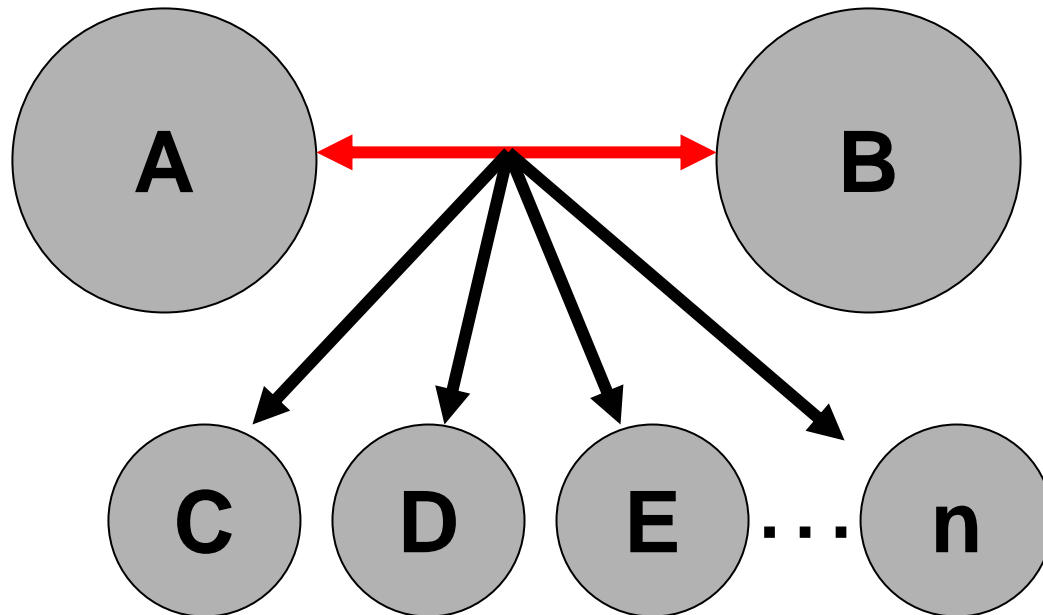
- This also happens to represent the BIND data model.

# Biomolecular function



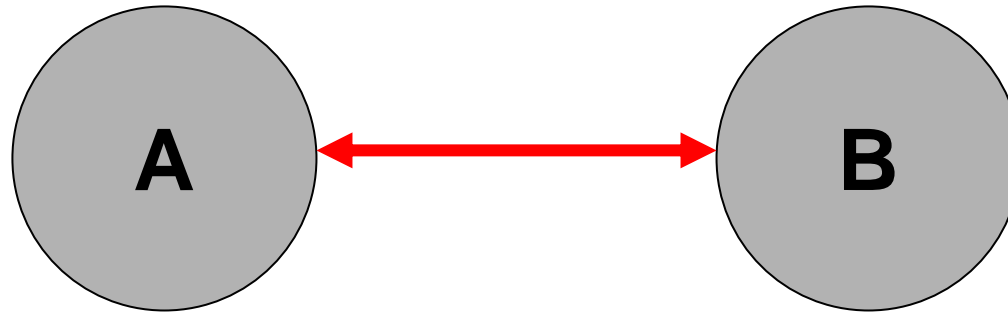
- A **data model** is just a way of organizing your observations...more later.

# Biomolecular function



- BIND stands for the **B**iomolecular **I**nteraction **N**etwork **D**atabase.

# A simple BIND record



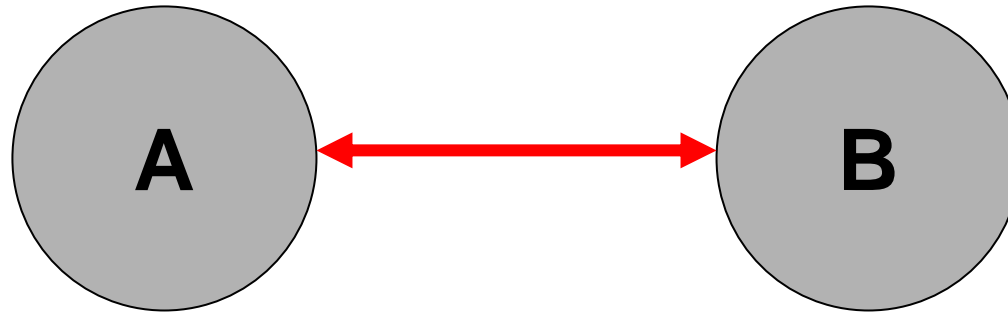
1. Short label for A
3. Molecule type for A
5. Database reference for A
7. Where A comes from

2. Short label for B
4. Molecule type for B
6. Database reference for B
8. Where B comes from

9. Publication reference

- The **minimal** BIND record has 9 pieces of information.

# An example BIND record



1. INAD  
3. Protein  
5. GenBank GI 3641615  
7. GenBank Taxonomy ID 7227

2. TRP  
4. Protein  
6. GenBank GI 7301861  
8. GenBank Taxonomy ID 7227

9. PubMed ID 8630257

- You can view this record in BIND

# <http://blueprint.org>

- Click on BIND in the right hand menu
- Enter 188 (the BIND record number) in the blue SEARCH box
- Click on the “Full BIND Record” link.
- More about Blueprint and searching BIND later.

Bioinformatics.ca		Contact Us	Help	Search	Mount Sinai	
Home	BIND	SeqHound	Protein Folding	Downloads	Exhibitions	News
About	Products	Services	Technical Support	Publications	People	Jobs

- Blueprint Home
- >BIND
- About
- Submit/Edit Records
- Search
- PreBIND
- BIND BLAST
- Curation
- BIND Development
- Downloads
- Database Stats
- Help
- FAQ
- Tutorials
- News
- Related Databases
- Credits
- Publications
- Contact BIND



**Recent News**

2004/02/11  
[New dataset imported into BIND 1285 Interaction Records](#)

2004/02/06  
[Release of BIND Version 2.5 Supporting Genetic Interactions 4054 Interaction Records](#)

2004/02/06  
[New High Throughput Genetic Interaction Dataset from Science in BIND 4054 Interaction Records](#)

2003/11/12  
[Blueprint Initiative First to Deliver CuraGen Corporation Fruit Fly Proteomics Map Data](#) [Query Guide](#)

BIND is an expanding database of biomolecular interaction, pathway and complex information. All information is stored in BIND database records that are freely available through a web interface that allows users to query, view, and submit records.

To access the database or learn more about BIND, use the left navigation column.

**LAUNCH SERVICE**

- In a new browser window
- In this browser window

**SEARCH**

188

- [Advanced Search](#)
- [Field Specific Search](#)
- [BIND Blast](#)
- [PreBIND](#)

**STATISTICS**

Interactions	53751
Complexes	1473
Pathways	8

**DOCUMENTATION**

- [Submissions Guide](#)
- [Curation Manuals](#)
- [Tutorials](#)

**LICENSE**

- Freely Available Under [GNU General Public License \(GPL\)](#)

**Data Manager**  
Menu

Version 2.5

[About](#) [Help](#)

[Search](#) [Browse](#)

[BIND Statistics](#)

[Administration](#)

[Print Page](#)

**Accession Query**

Search by:   
 Integer ID:

**Results for BIND Accession ID(s): 188**

Interaction 188
[Drosophila melanogaster](#)
[Full BIND Record](#)
 Launch Viewer:

Molecule	Description	Molecular Function	Cellular Component	Biological Process	Experiment(s)	Links
INAD	adaptor protein; inactivation-no-afterpotential D; contains 5 PDZ domains	<ul style="list-style-type: none"> <li><a href="#">structural molecule activity</a></li> </ul>	<ul style="list-style-type: none"> <li><a href="#">inaD signaling complex</a></li> </ul>	<ul style="list-style-type: none"> <li><a href="#">intracellular signaling cascade</a></li> <li><a href="#">deactivation of rhodopsin mediated signaling</a></li> </ul>	<ul style="list-style-type: none"> <li>Affinity Chromatography</li> <li>Immunoprecipitation</li> <li>Other (See description)</li> </ul>	NCBI SeqHound 1 BIND Complex 2 Abstracts [Pubmed] [Other BIND data]
TRP	transient receptor potential, subunit of store-operated calcium channel. Aliases: DIP1	<ul style="list-style-type: none"> <li><a href="#">light-activated voltage-gated calcium channel activity</a></li> </ul>	<ul style="list-style-type: none"> <li><a href="#">light-activated voltage-gated calcium channel complex</a></li> <li><a href="#">rhabdomere</a></li> </ul>	<ul style="list-style-type: none"> <li><a href="#">calcium ion transport</a></li> <li><a href="#">light-induced release of calcium from internal store</a></li> </ul>		NCBI SeqHound

Comments and suggestions to: [info@bind.ca](mailto:info@bind.ca)



**Data Manager**  
Menu  
Version 1.8

[About](#) [Help](#)  
[Search](#) [Browse](#)

Interactions

- [Add](#)
- [Change](#)

Pathways

- [Add](#)
- [Change](#)

Molecular Complexes

- [Add](#)
- [Change](#)

---

[BIND Statistics](#)

---

[Administration](#)

# Interaction

Interaction ID: 188

→ 9 minimal pieces of information

Accession date: Jan 31, 2000

Description: INAD binds TRP, a store-operated calcium channel in Drosophila phototransduction

[View record update history](#)

## Molecule A

**INAD**

Description: adaptor protein (inactivation-no-afterpotential D); contains 5 PDZ domains

Molecule Type: Protein

GI: 3641615 [more information on this protein in - \(NCBI\) \(SEQHOUND\) \(BIND\) \(FAST\)](#)

Molecule origin: Organismal

Organism: [Drosophila melanogaster](#)

## Molecule B

**TRP**

Description: transient receptor potential, subunit of store-operated calcium channel. Aliases: DIP1

Molecule Type: Protein

GI: 7301861 [more information on this protein in - \(NCBI\)](#)

Molecule origin: Organismal

Organism: [Drosophila melanogaster](#)

[Visualize Interaction!](#)

View other information?		
Main Info	Publications	AS
Cellular Place	Experimental Condition	Conse
N/A	Experimental Conditions	
Binding Sites	Chemical action	Ch
Binding Sites	N/A	

Comments and suggestions to: < [bind@mshri.on.ca](mailto:bind@mshri.on.ca) >

Search and Retrieve a BIND record... - Microsoft Internet Explorer

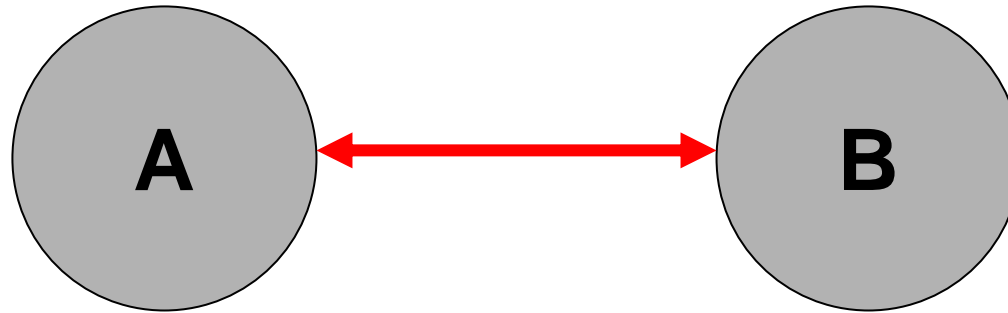
### BIND Publication links for Interaction 188

- [PubMed ID 8630257](#) Opinion: **Supports**  
Description: Regulation of the TRP Ca<sup>2+</sup> channel by INAD in Drosophila photoreceptors
- [PubMed ID 9230432](#) Opinion: **Supports**  
Description: A multivalent PDZ-domain protein assembles signaling complexes in a G-protein-coupled cascade

Comments and suggestions to: < [bind@mshri.on.ca](mailto:bind@mshri.on.ca) >

Internet

# A curated BIND record



1. Short label for A
3. Molecule type for A
5. Database reference for A
7. Where A comes from

2. Short label for B
4. Molecule type for B
6. Database reference for B
8. Where B comes from

9. Publication reference

- The **curated** BIND record may have many more pieces of information....

**Data Manager Menu**  
Version 2.5

[About](#) [Help](#)  
[Search](#) [Browse](#)

---

[BIND Statistics](#)

---

[Administration](#)

**Molecule B**

**BACH1**

**Aliases:** BRIP1

Description: BRCA1-associated C-terminal helicase-1; BRCA1 Interacting Protein 1; member of the DEAH helicase family, contains 7 helicase-specific motifs conserved among members of the DEAH family, helicase domain contains a nuclear localization signal; ubiquitous expression; mutation may be involved in familial breast cancer, OMIM ID: 605882

Molecule Type: Protein

GI: 14042978 **Use This GI to search -** [\(NCBI\)](#) [\(SEQHOUND\)](#) [\(BIND\)](#) [\(Protein Domains\)](#)

Molecule origin: Organismal

Organism: [Homo sapiens](#)

**GO Annotation**

**Molecular Function**

- [DNA binding](#)
- [ATP dependent DNA helicase activity](#)
- [ATP binding](#)
- [molecular function unknown](#)

**Cellular Component**

- [nucleus](#)
- [cellular component unknown](#)

**Biological Process**

- [nucleotide-excision repair](#)

[Cellular Place](#)

← GO annotation

**Click below to view the interaction annotation**

Main Info	Publications	ASN.1	XML
Cellular Place	Experimental Evidence	Conserved Sequence	
Cellular Place	Experimental Evidence	N/A	
Binding Sites	Chemical action	Chemical State	
Binding Sites	Chemical Action	N/A	

← Other data about the interaction

Comments and suggestions to: [info@bind.ca](mailto:info@bind.ca)

**View other information?**

Main Info	Publications	ASN.1	XML
Cellular Place	Experimental Condition	Conserved Sequence	
Cellular Place	Experimental Conditions	N/A	
Binding Sites	Chemical action	Chemical State	
Binding Sites	Chemical Action	N/A	

View other information?	
Main Info	Publications
Cellular Place	Experimental Condition
Cellular Place	Experimental Conditions
Binding Sites	Chemical action
Binding Sites	Chemical Action

Search and Retrieve a BIND record... - Microsoft Internet Explorer

3.

- Internal Condition ID (ICID): 2
- General Experimental Conditions: *In vitro*
- Experimental System: Affinity Chromatography
- Description: An interaction between BRCA1 and BACH1 was demonstrated by GST pull-down assay. [35S]-methionine labeled in vitro transcribed and translated BACH1 was isolated using purified BRCT immobilized on glutathione beads. Eluates were resolved by SDS-PAGE and detected by autoradiography. Fig. 4. The mutants BACH1[888-1249] and BACH1[888-1063] were also isolated by BRCT. Fig 4.

**Specific Binding Sites attached to this experimental evidence data:**

- [Sequence Location ID \(SLID\) 0](#) (Molecule A)
- [Sequence Location ID \(SLID\) 0](#) (Molecule B)

**Molecule A Experimental Form: Molecule**

**GST-BRCT**  
 Description: Glutathione-S-transferase fusion of carboxy terminal region of BRCA1 [residues 1529-1863] containing the BRCT repeat motif labelled with [32P]  
 Molecule Type: Not specified.  
 Molecule origin: Not Specified

**Molecule B Experimental Form: Molecule**

**BACH1[888-1063]**  
 Description: part of carboxy terminal region of BACH1  
 Molecule Type: Not specified.  
 Molecule origin: Not Specified

**Publications attached to this experimental evidence data:**

- [PubMed ID 11301010](#) Opinion: **Supports**  
 Description: BACH1, a Novel Helicase-like Protein, Interacts Directly with BRCA1 and Contributes to Its DNA Repair Function

Internet

### View other information?

Main Info	Publications	ASN.1	XML
Cellular Place	Experimental Conditions	Classical Science	
Cellular Place	Experimental Conditions		
Binding Sites	Chemical Analysis		
Binding Sites	Chemical Analysis		

**Interaction Location for BIND ID 12575**

- BIND Place ID (BPID): 0
  - Description: MCF7 cells were immunostained with anti-BRCA1 and anti-BACH1 antibodies. Colocalization of BRCA1 and BACH1 in punctate nuclear bodies was observed. Fig 3B, see ICID 1. Near complete colocalization was detected in late S-G2 synchronous cell populations.
  - **General Cellular Place of Interaction:**
    - **Place:** Nucleus - General
    - **General place description:** punctate bodies within nucleus

Close

Comments and suggestions to: < [info@bind.ca](mailto:info@bind.ca) >

View other information?	
Main Info	Publications
Cellular Place	Experimental Conditions
Cellular Place	Experimental Conditions
Binding Sites	Chemical action
Binding Sites	Chemical Action

Search and Retrieve a BIND record... - Microsoft Internet Explorer

## Binding Sites for BIND ID 12575

**General:**

No list of binding site pairings has been defined

### Binding Sites on BRCA1

- Sequence Location ID (SLID): 0
- Binding Site: 1529-1863
- Description: The wt GST-BRCT (1529-1863) binds to BACH1 species. One mutation [P1749R] led to greatly reduced binding and another [M1775R] completely abolished binding of BACH1. Fig 1B, 1C, data not shown, p. 149.
- **Publications associated with this binding site:**
  1. [PubMed ID 11301010](#) Opinion: **Supports**  
Description: BACH1, a Novel Helicase-like Protein, Interacts Directly with BRCA1 and Contributes to Its DNA Repair Function.

### Binding Sites on BACH1

- Sequence Location ID (SLID): 0
- Binding Site: 888-1063
- Description: A discrete region of BACH1, C-terminal to the helicase domain and spanning residues 888-1063, is sufficient for BRCA1 binding. Fig 4.
- **Publications associated with this binding site:**
  1. [PubMed ID 11301010](#) Opinion: **Supports**  
Description: BACH1, a Novel Helicase-like Protein, Interacts Directly with BRCA1 and Contributes to Its DNA Repair Function

Close

---

Comments and suggestions to: < [info@bind.ca](mailto:info@bind.ca) >

Done Internet

# Curation of BIND records

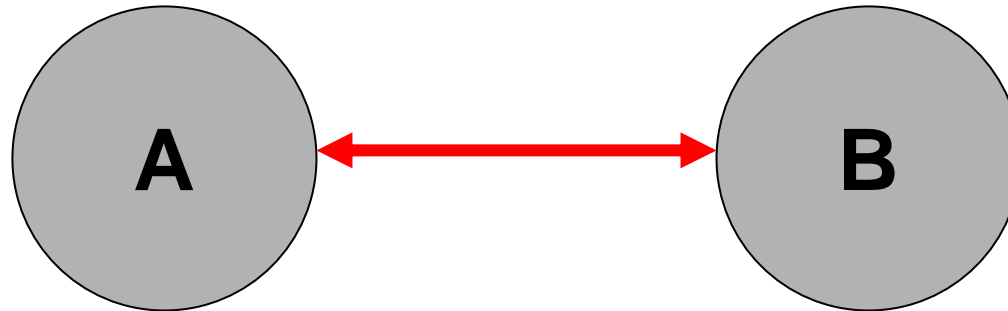
- A lot more information on the use of the BIND data structure can be found in the BIND curator's manual:

[http://www.blueprint.org/bind/curation/bind\\_about\\_curation.html](http://www.blueprint.org/bind/curation/bind_about_curation.html)

- The complete BIND data structure can be found at:

<ftp://ftp.blueprint.org/pub/BIND/spec/>

# BIND records are observations



1. Short label for A
3. Molecule type for A
5. Database reference for A
7. Where A comes from

2. Short label for B
4. Molecule type for B
6. Database reference for B
8. Where B comes from

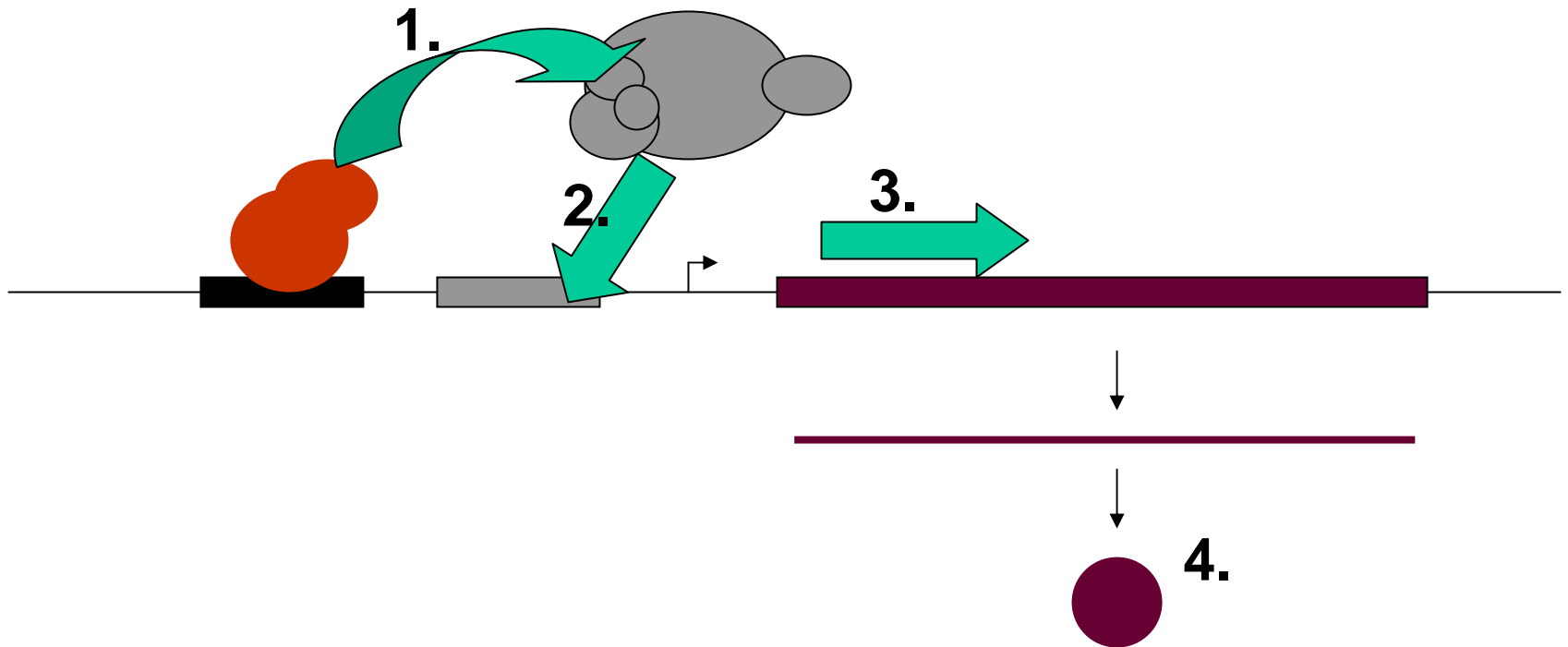
9. Publication reference

- All BIND records will have a publication reference and most will specifically list a method(s) used to demonstrate the interaction.

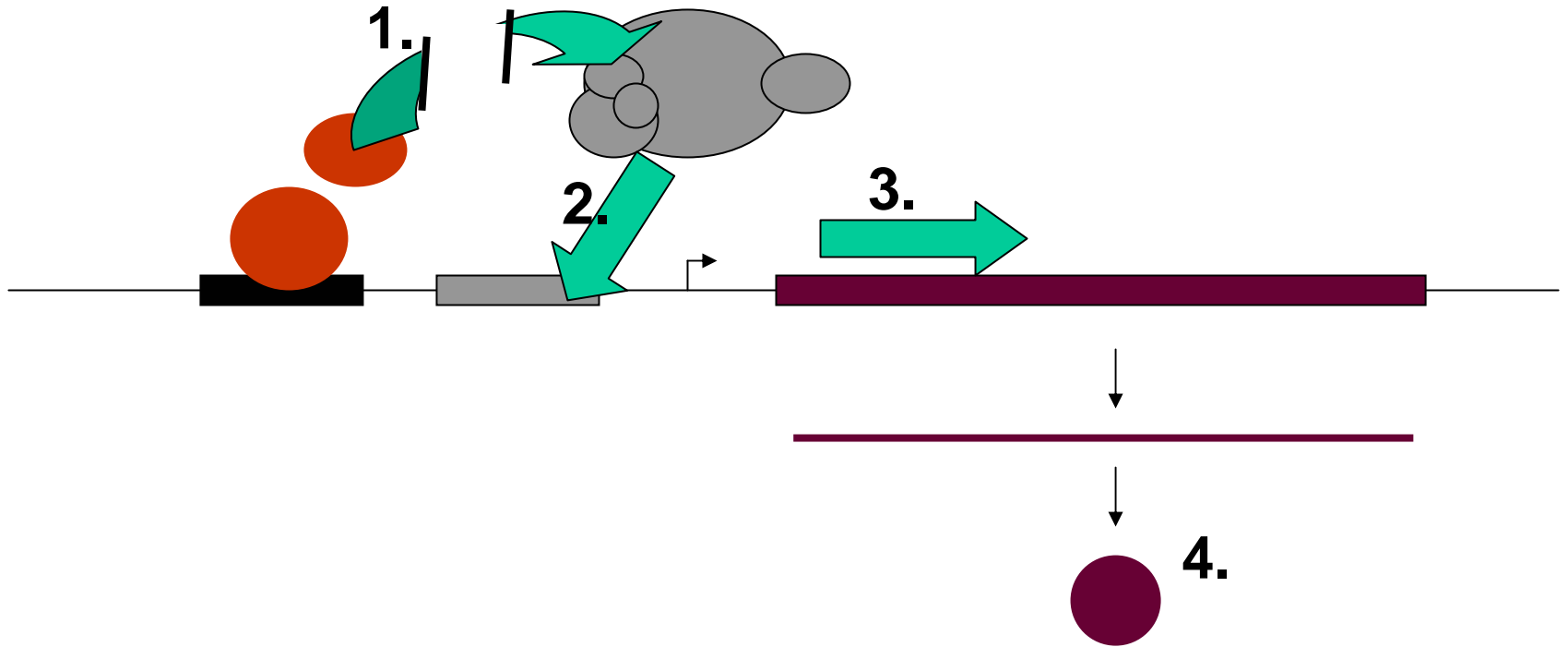
# Methods used to detect interactions.

- A great deal of interaction data in BIND originates from high-throughput experiments designed to detect interactions between proteins.
- The most common methods are:
  - Two-hybrid assay
  - Affinity purification

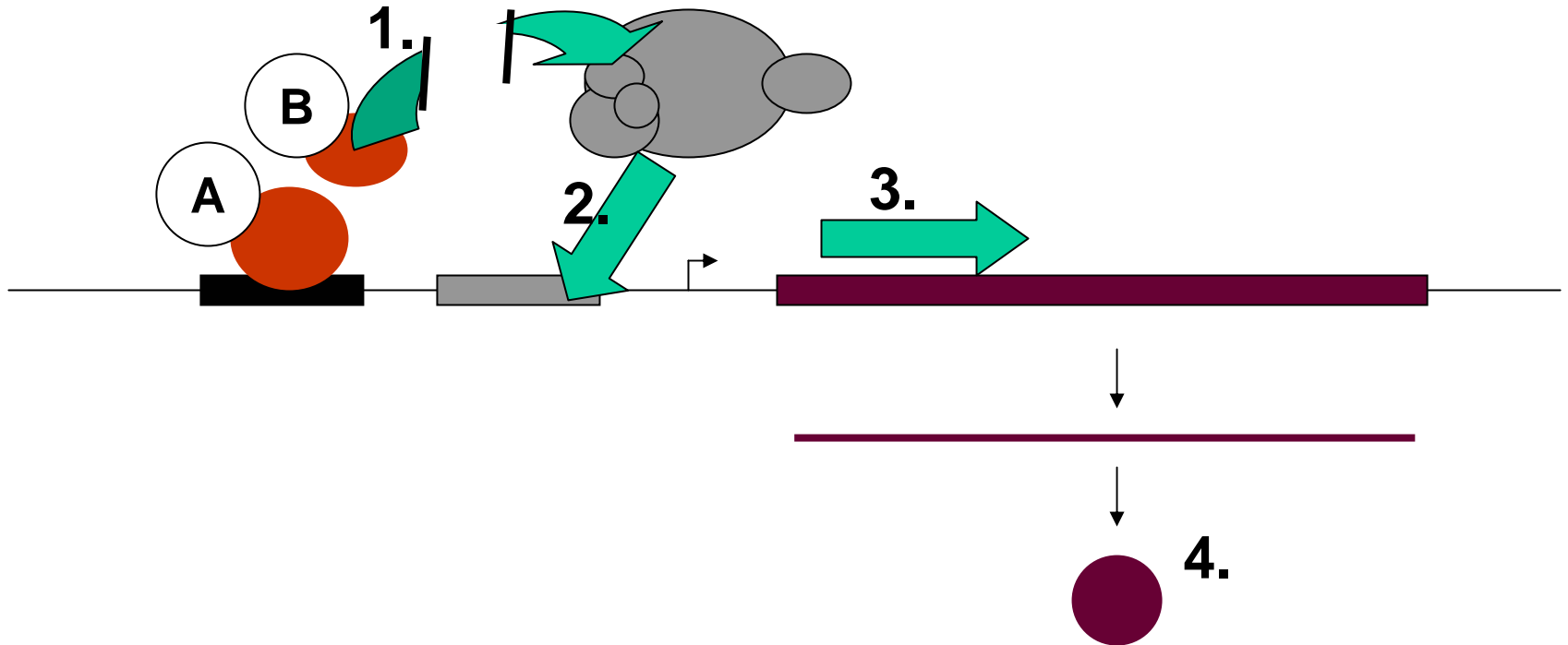
# Two-hybrid assay



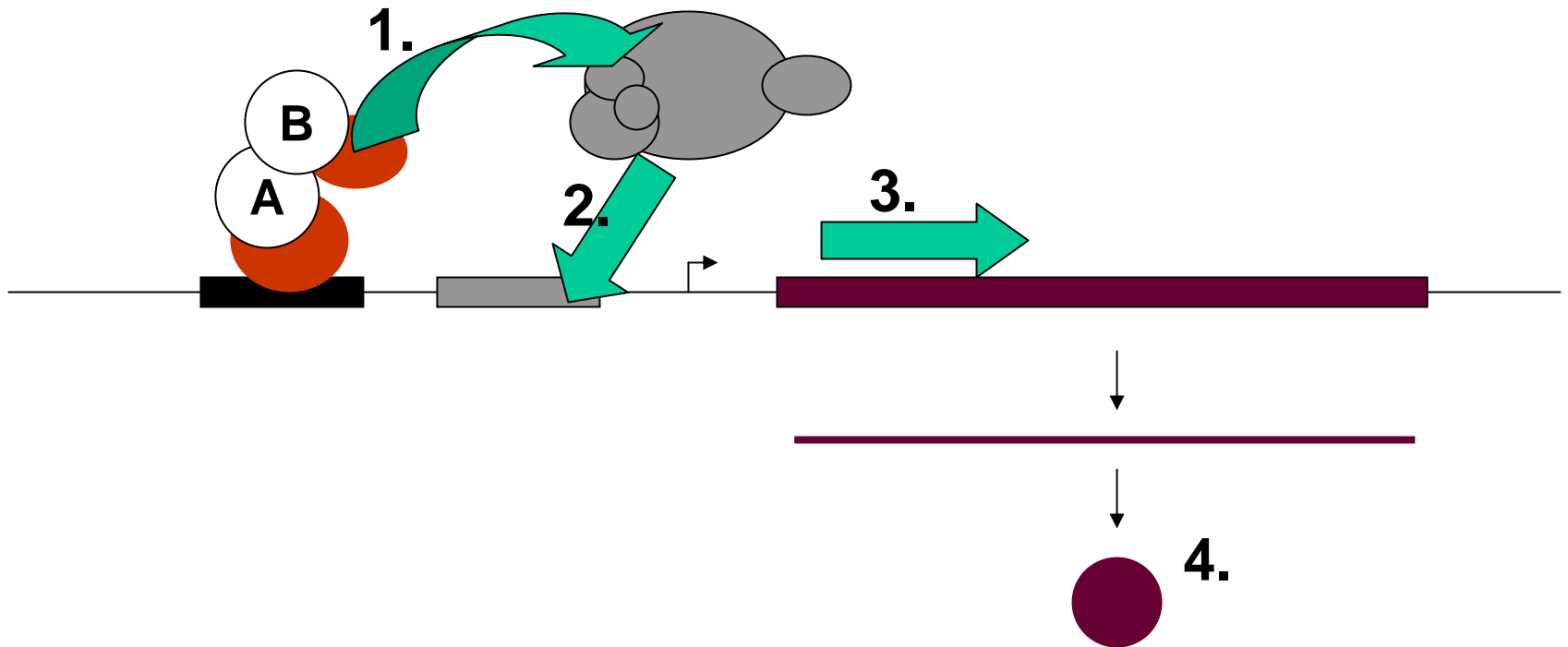
# Two-hybrid assay



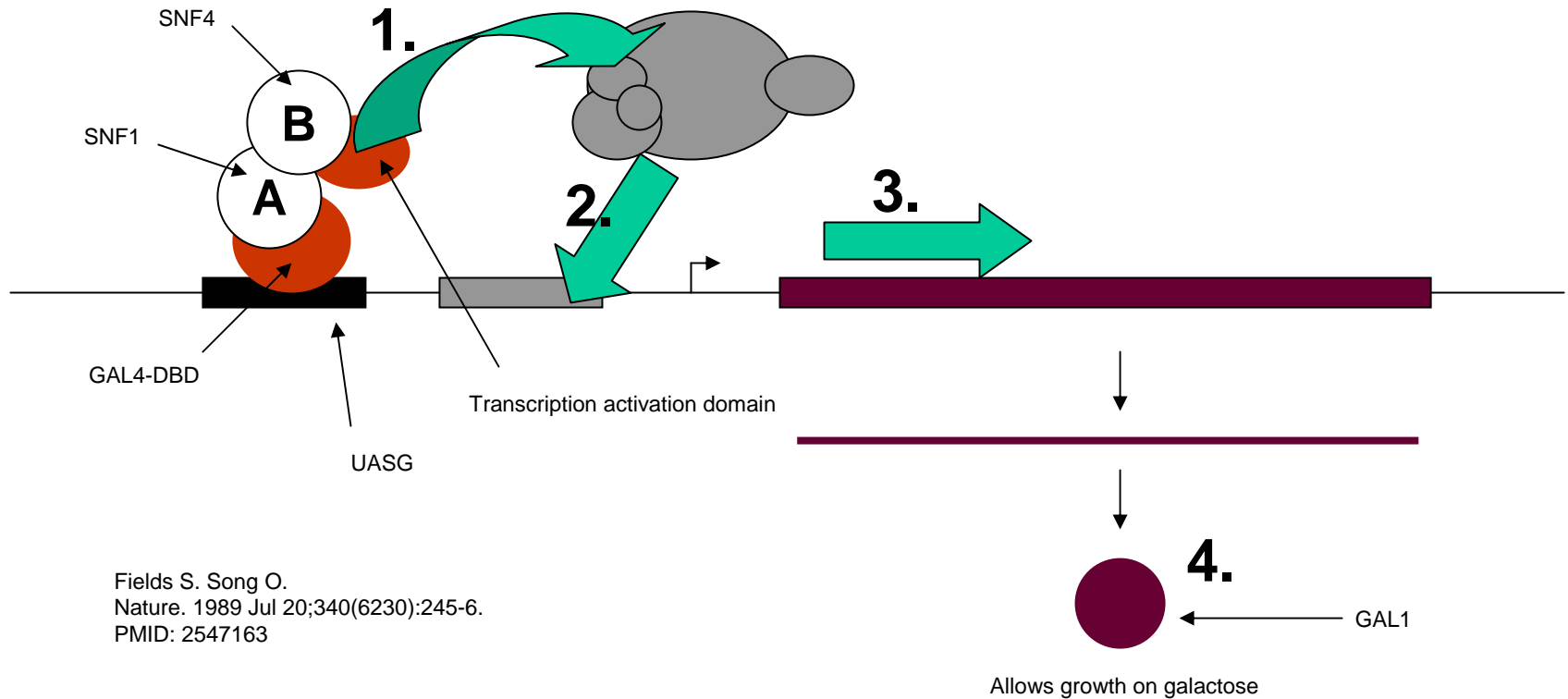
# Two-hybrid assay



# Two-hybrid assay

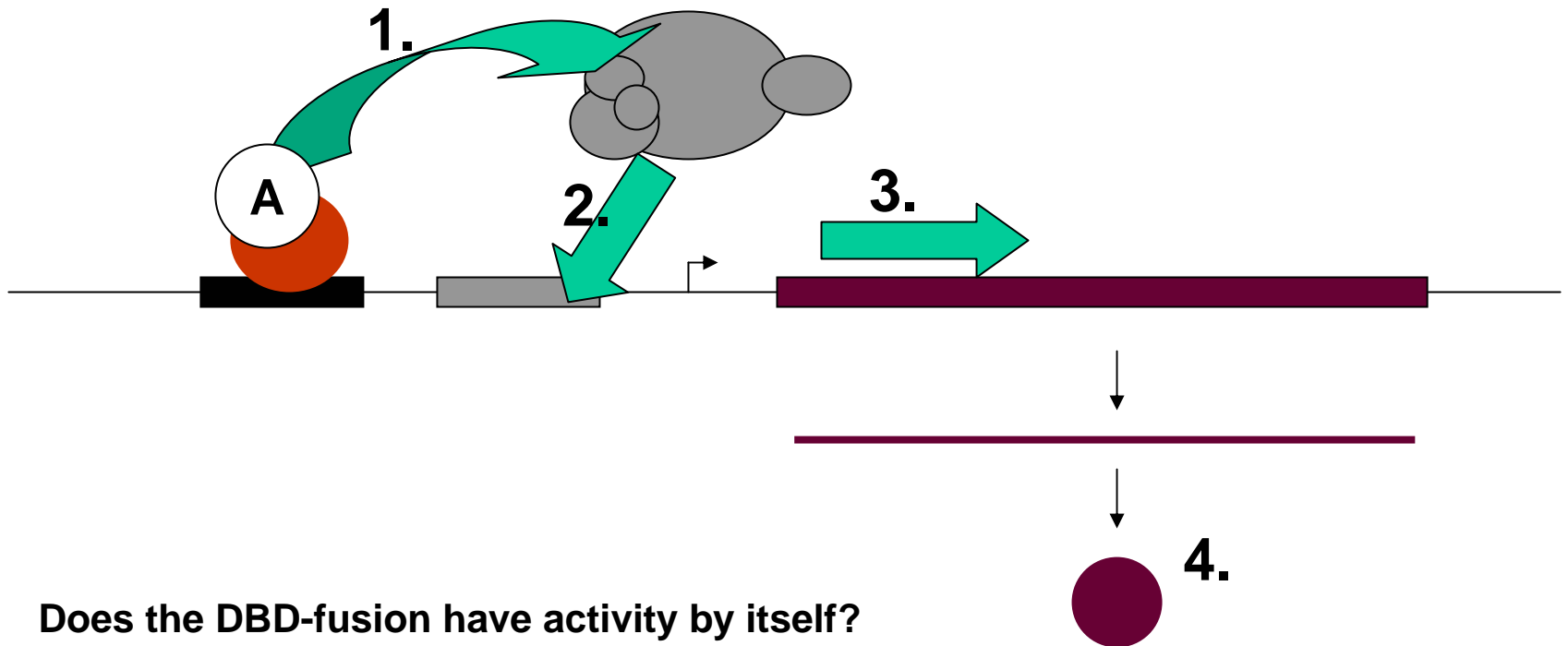


# Two-hybrid assay

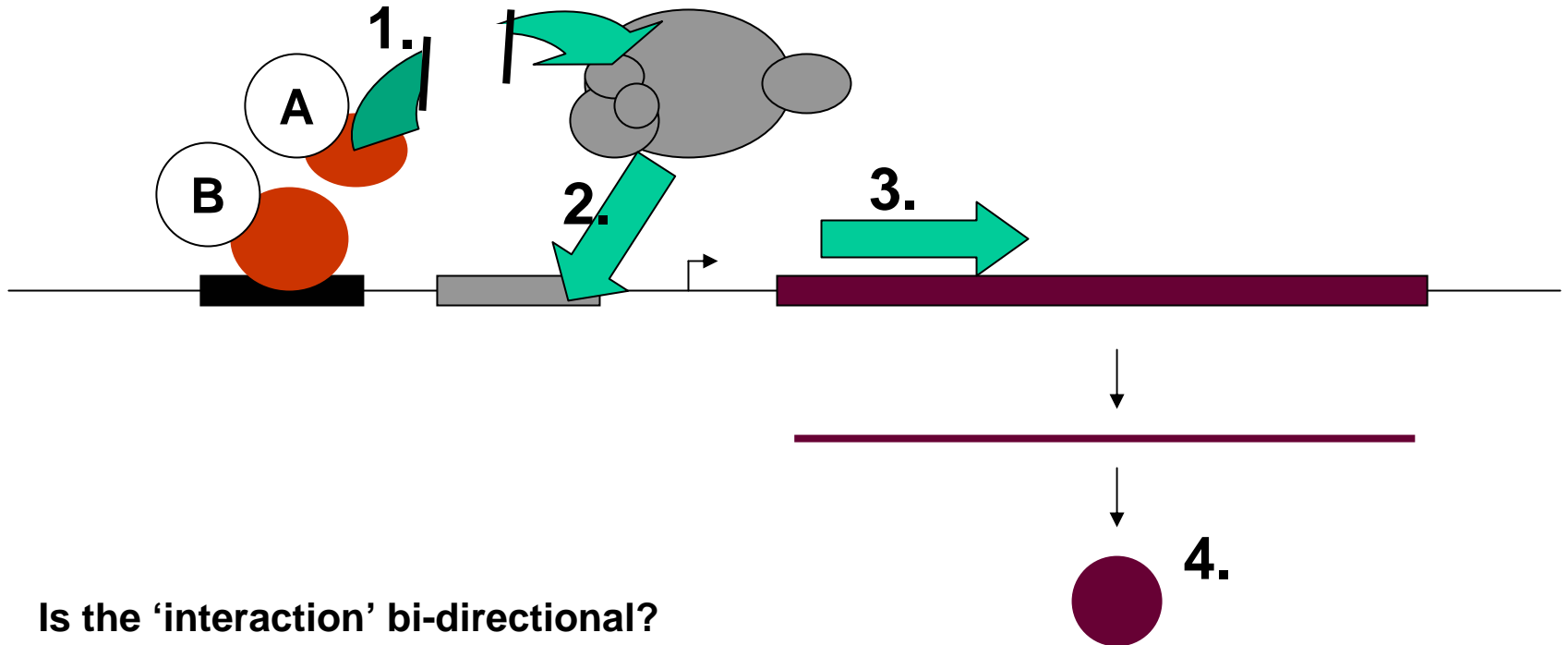


Fields S. Song O.  
Nature. 1989 Jul 20;340(6230):245-6.  
PMID: 2547163

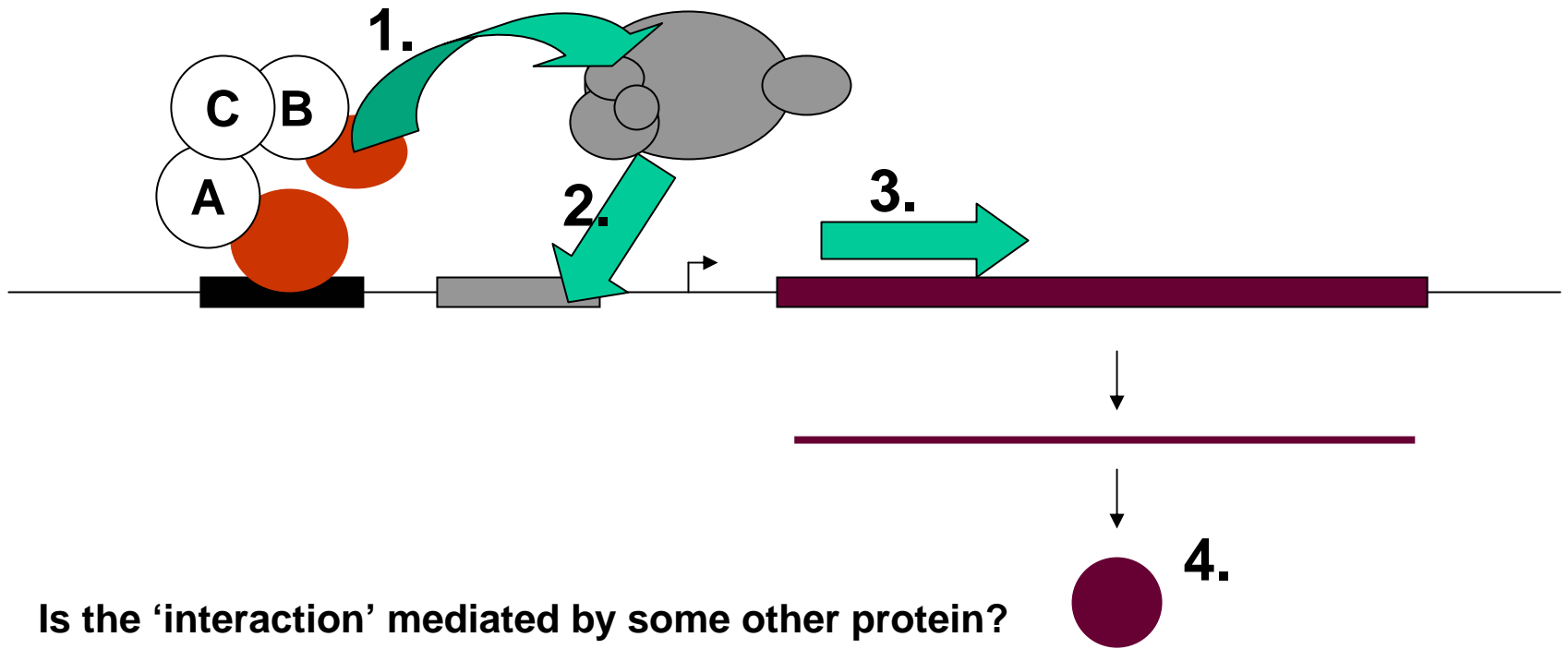
# Some Two-hybrid caveats



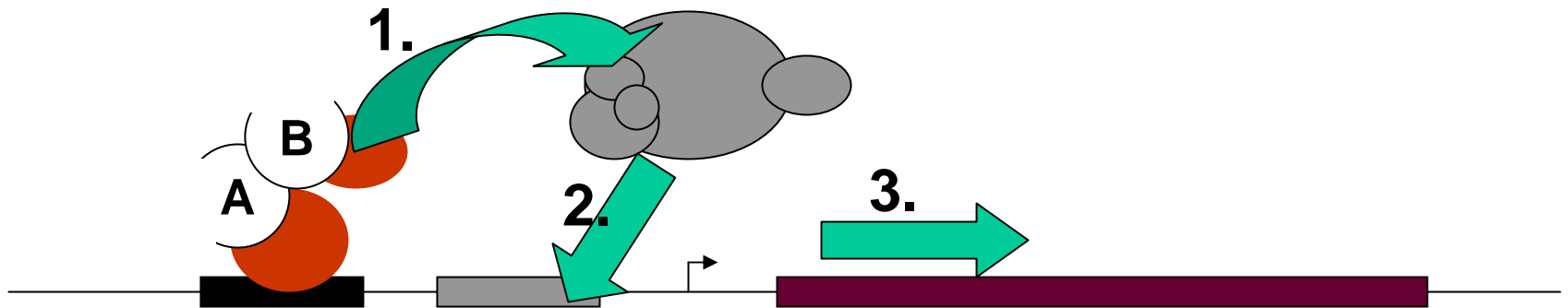
# Some Two-hybrid caveats



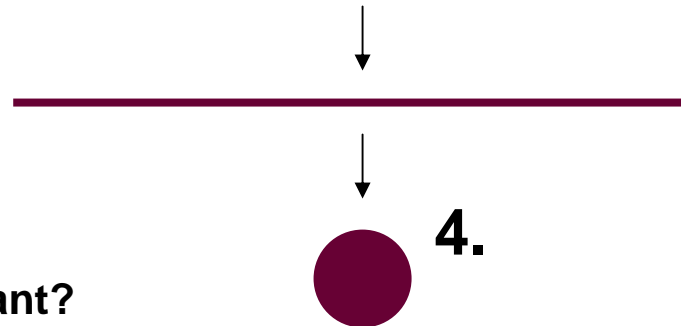
# Some Two-hybrid caveats



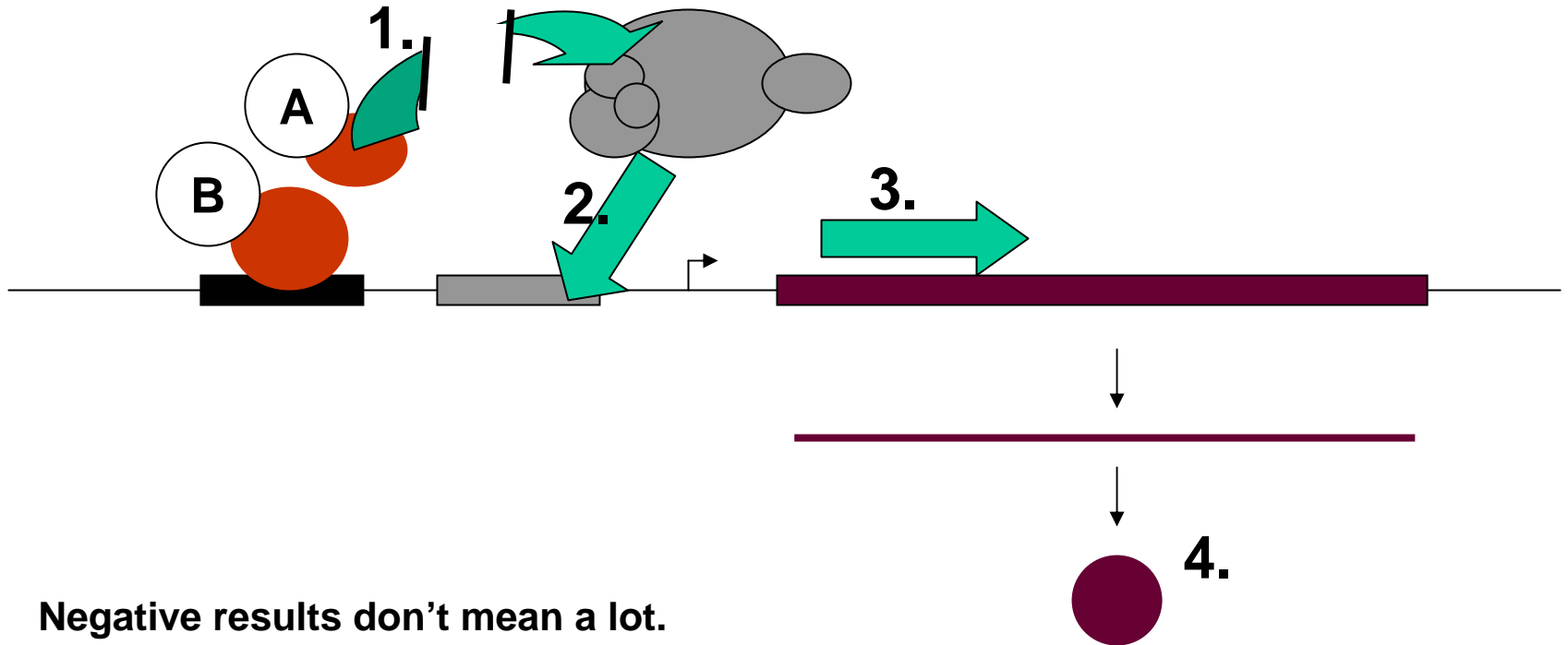
# Some Two-hybrid questions



- Are the proteins expressed?
- Are they over-expressed?
- Are they in-frame?
- Are the interacting domains defined?
- Was the observation reproducible?
- Was the strength of interaction significant?
- Was another method used to back-up the conclusion?
- Are the two proteins from the same compartment?

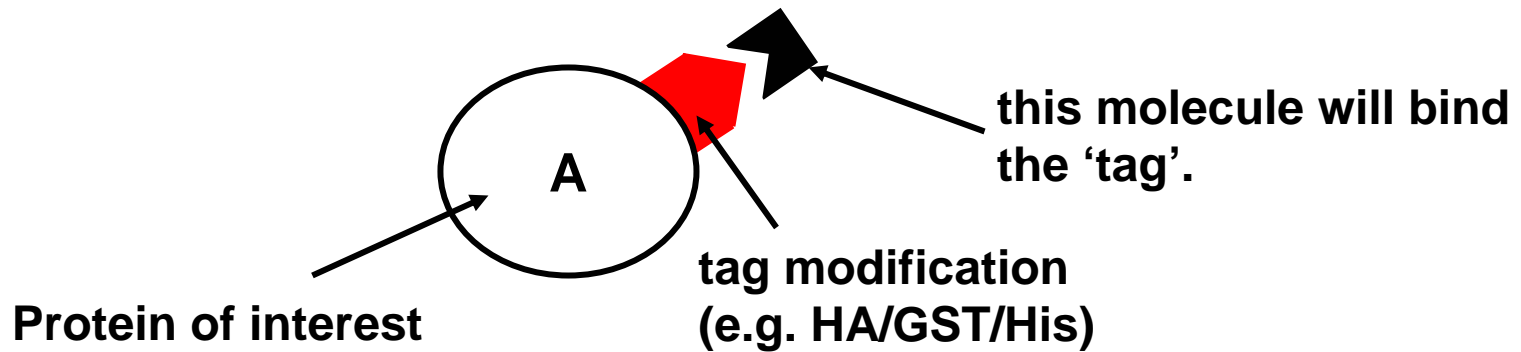


# Two-hybrid assay

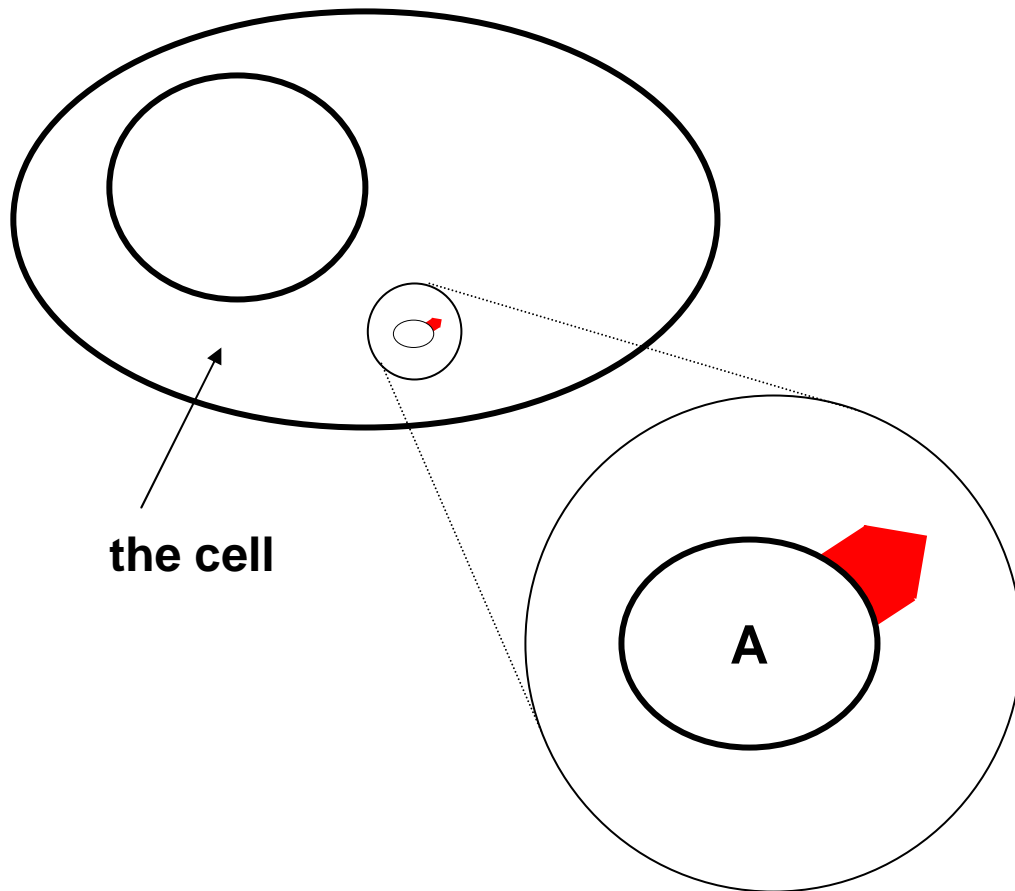


Negative results don't mean a lot.

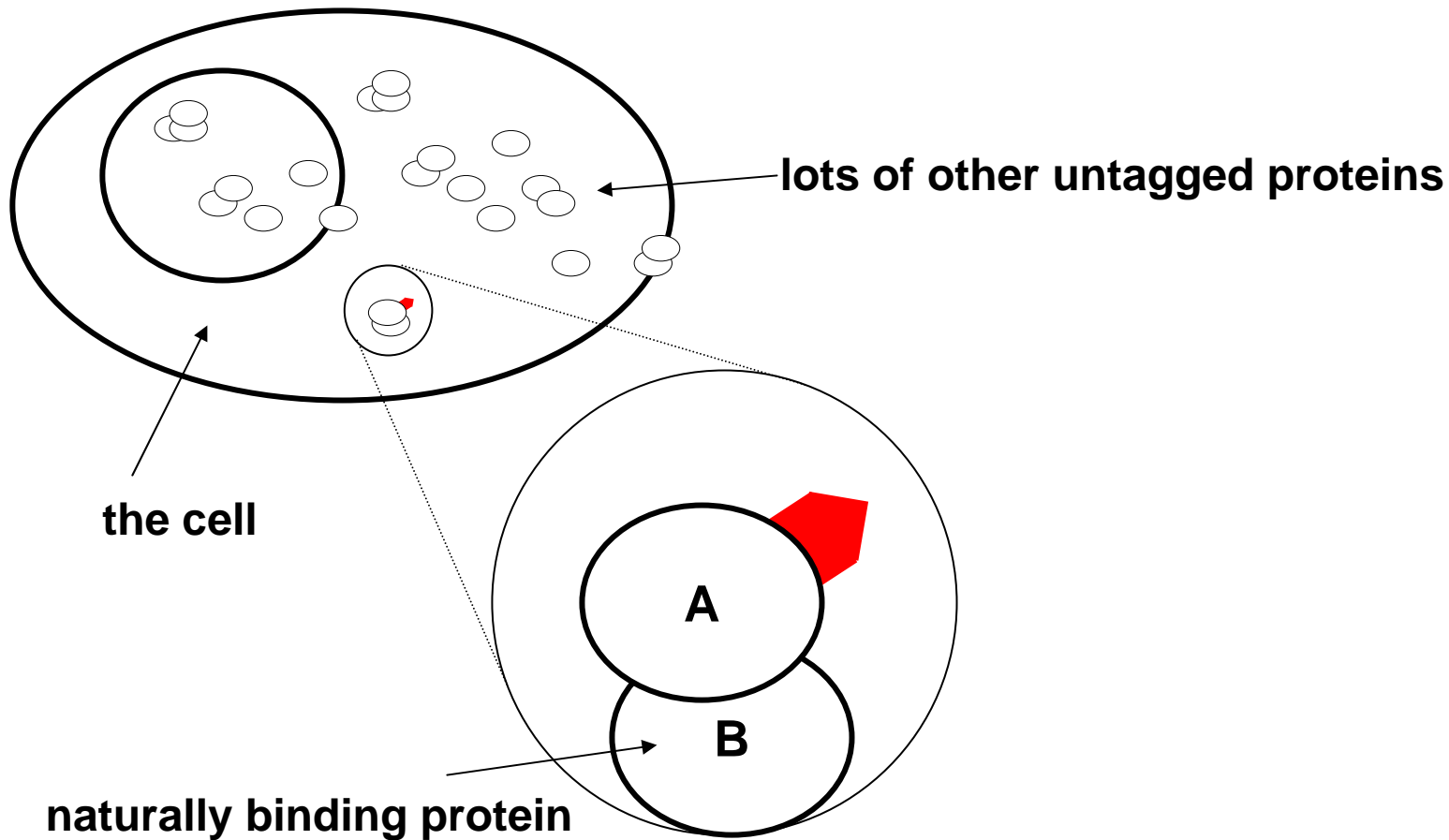
# Affinity purification



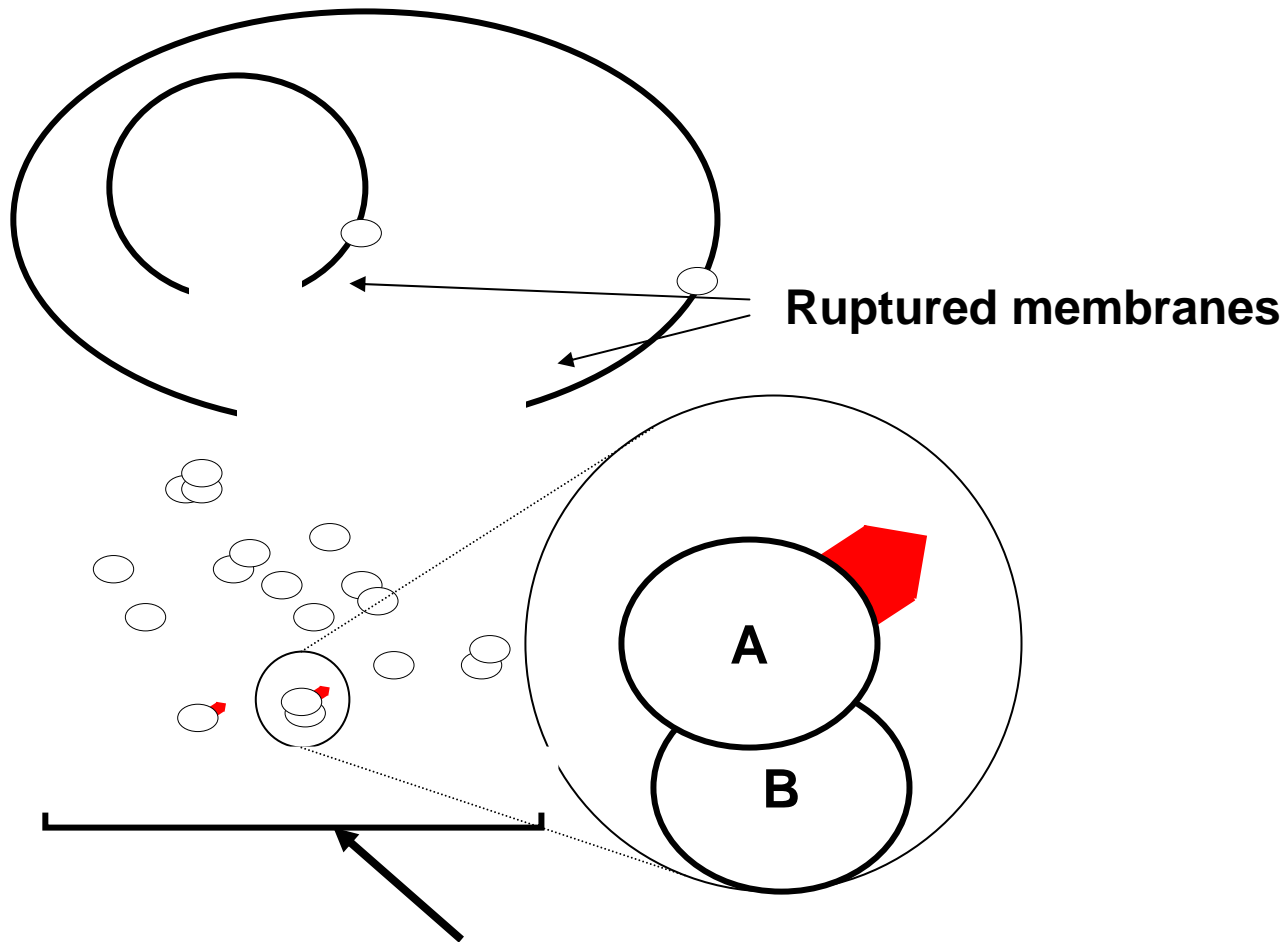
# Affinity purification



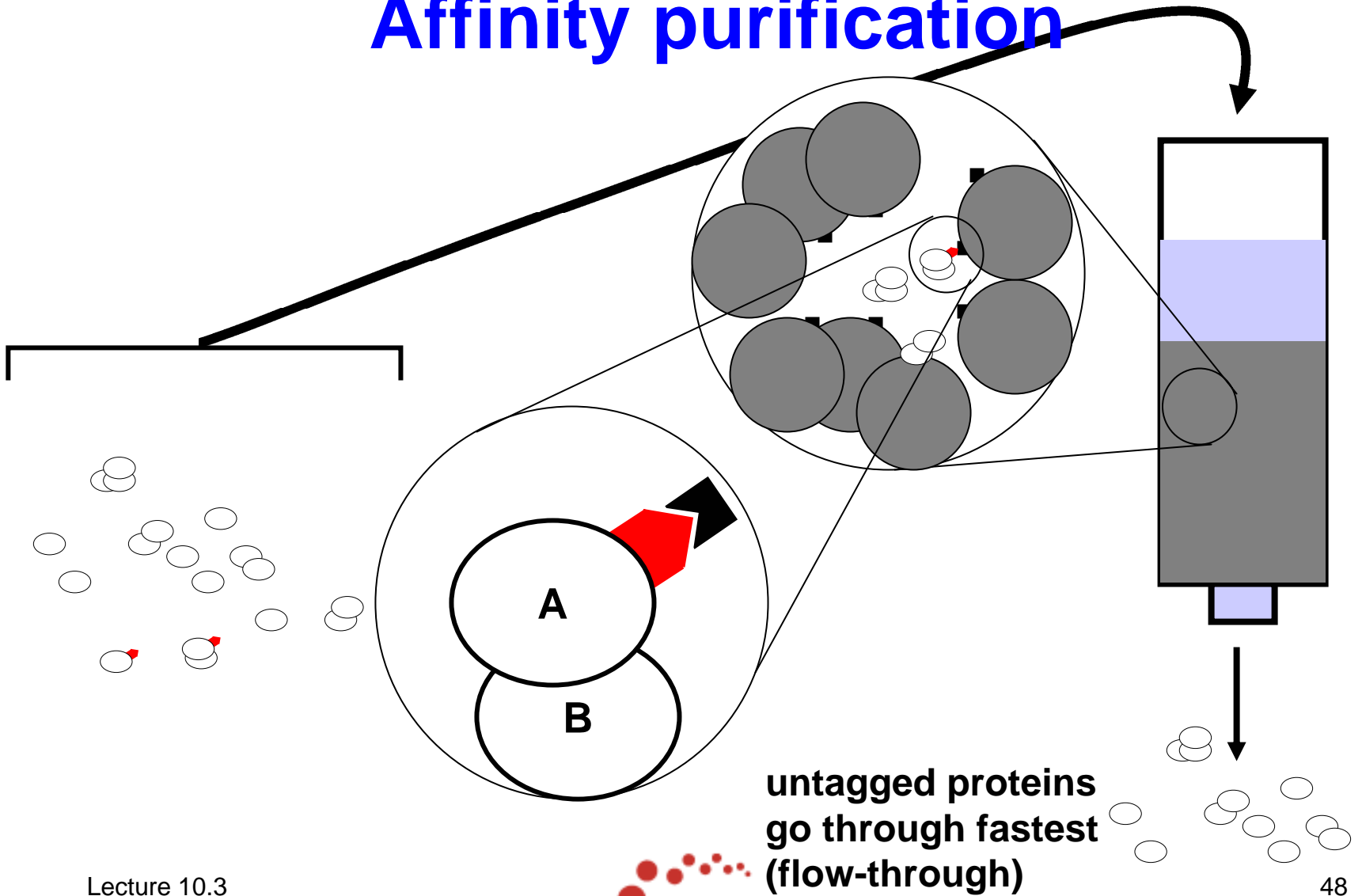
# Affinity purification



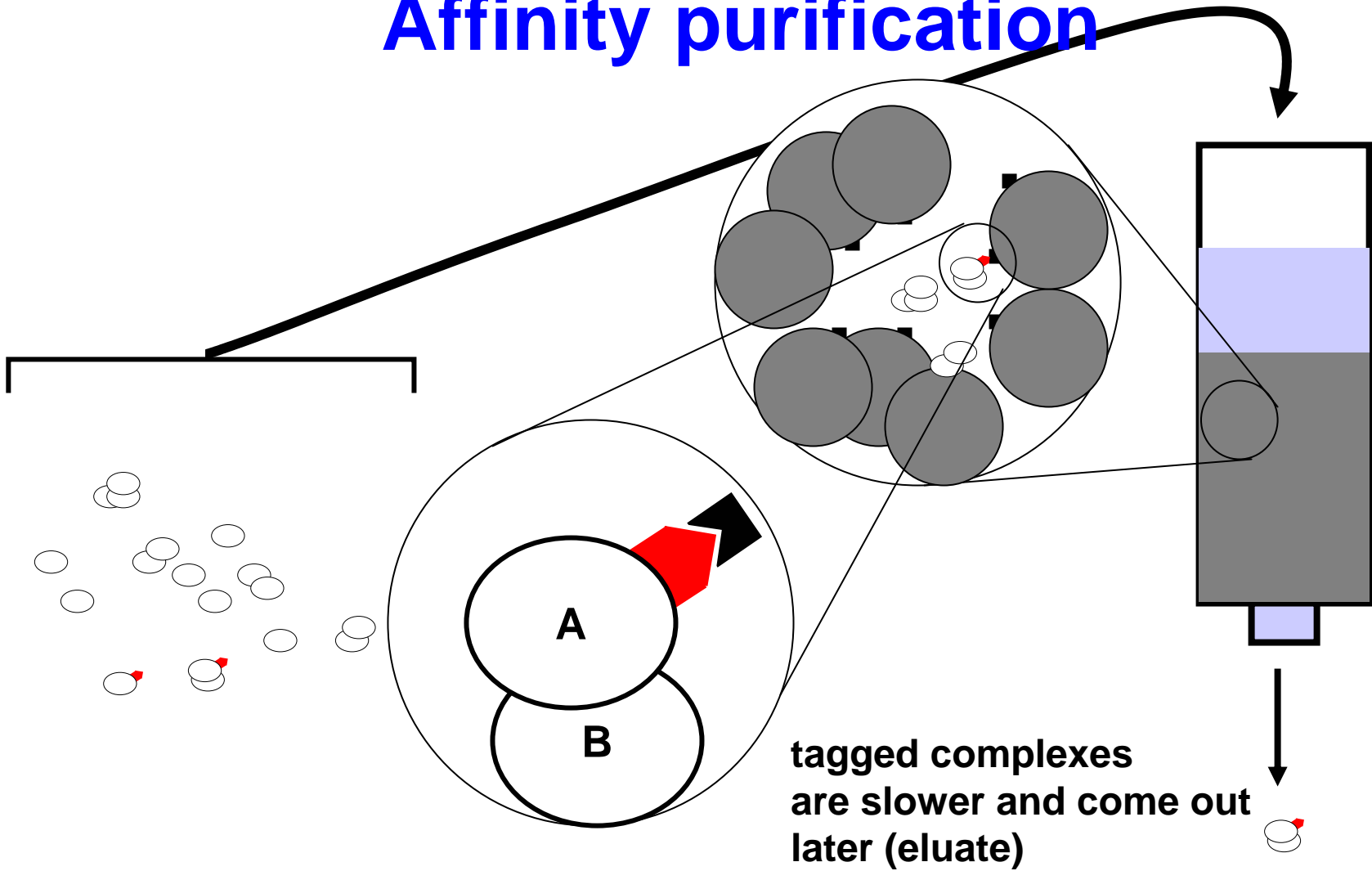
# Affinity purification



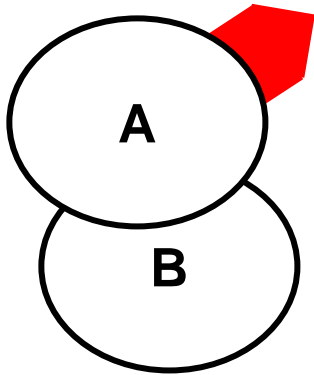
# Affinity purification



# Affinity purification



# Some affinity purification questions



**Is the bait protein expressed and in frame?**

**Is the bait protein observed?**

**Is the bait protein over-expressed?**

**Are the interacting domains defined?**

**Was the observation reproducible?**

**Was the interactor found in the background?**

**Was the strength of interaction significant?**

**Was the interaction saturable?**

**Was the interactor stoichiometric with the bait protein?**

**Was another method used to back-up the conclusion?**

**Was tandem-affinity purification (TAP) used?**

**Was the interaction shown using an extract or a purified protein?**

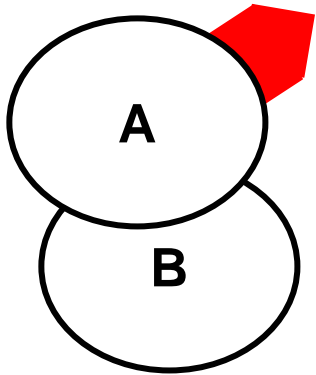
**Is the inverse interaction observable?**

**Are the two proteins from the same compartment?**

**Are the two proteins known to be involved in the same process?**

**Is the interactor likely to be physiologically significant?**

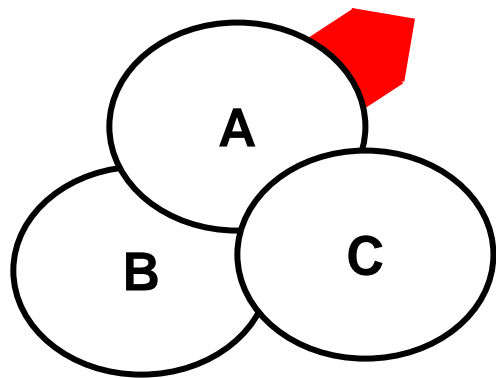
# Some affinity purification caveats



First and most importantly,  
this is only a representation of the observation.

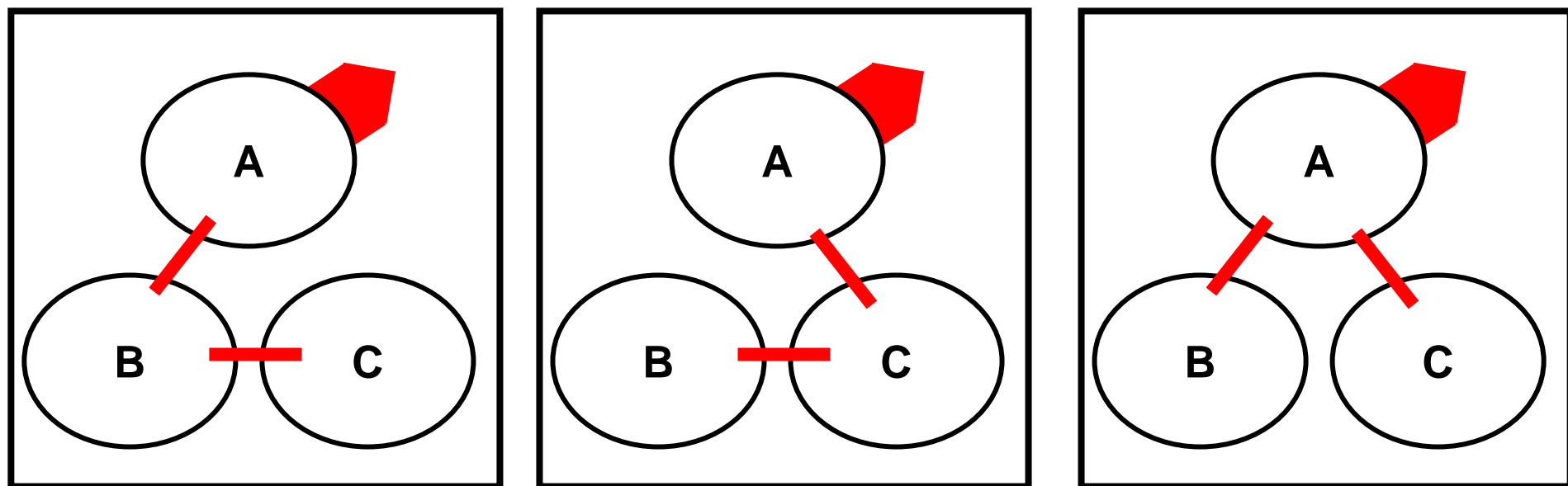
You can only tell what proteins are in the eluate;  
you can't tell how they are connected to one another.

If there is only one other protein present (B), then its likely that  
A and B are directly interacting.



But, what if I told you that two other proteins (B and C) were  
present along with A....

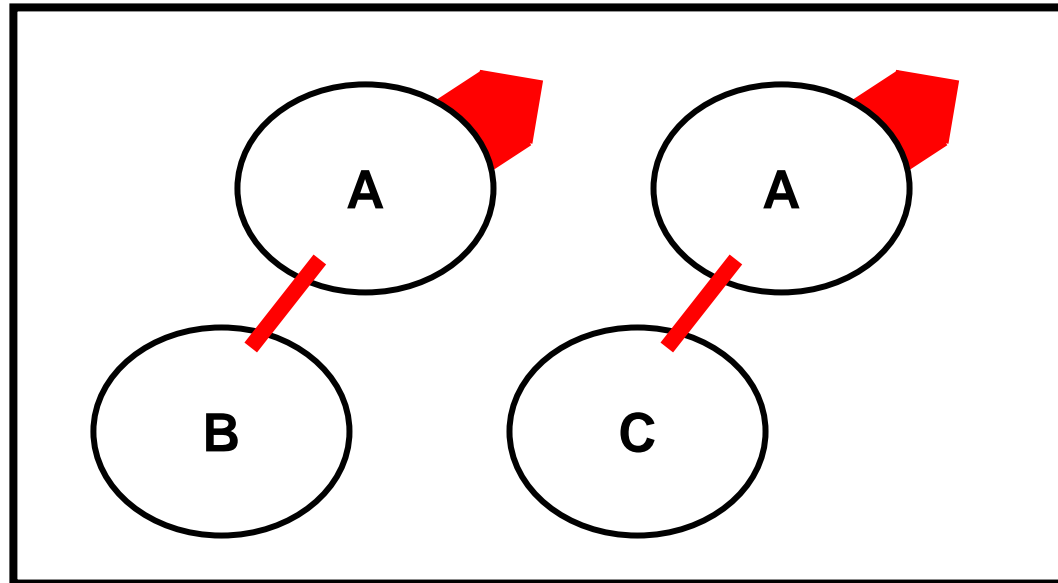
# Complexes with unknown topology



Which of these models is correct?

The complex described by this experimental result is said to have an **Unknown Topology**.

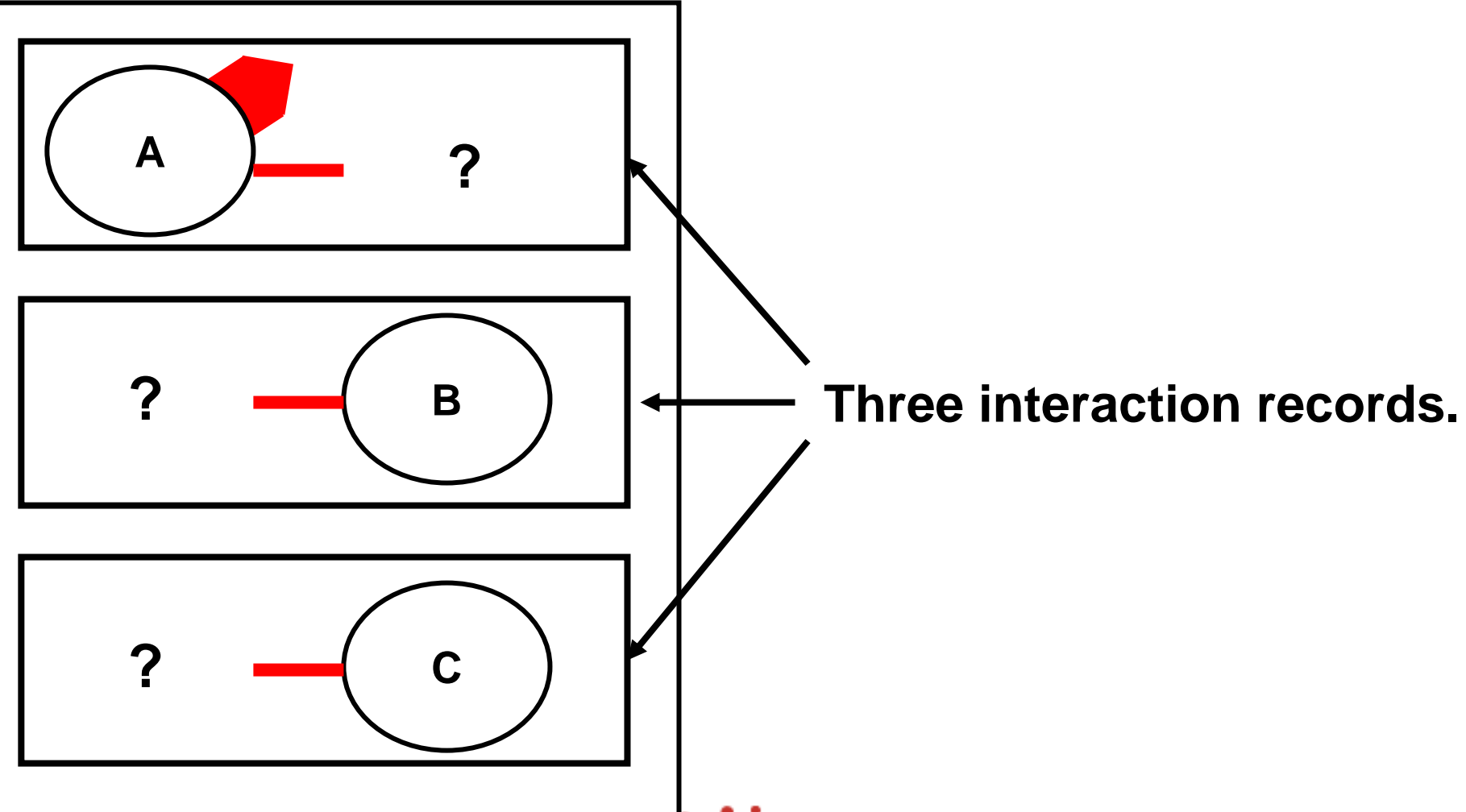
# Complexes with unknown stoichiometry



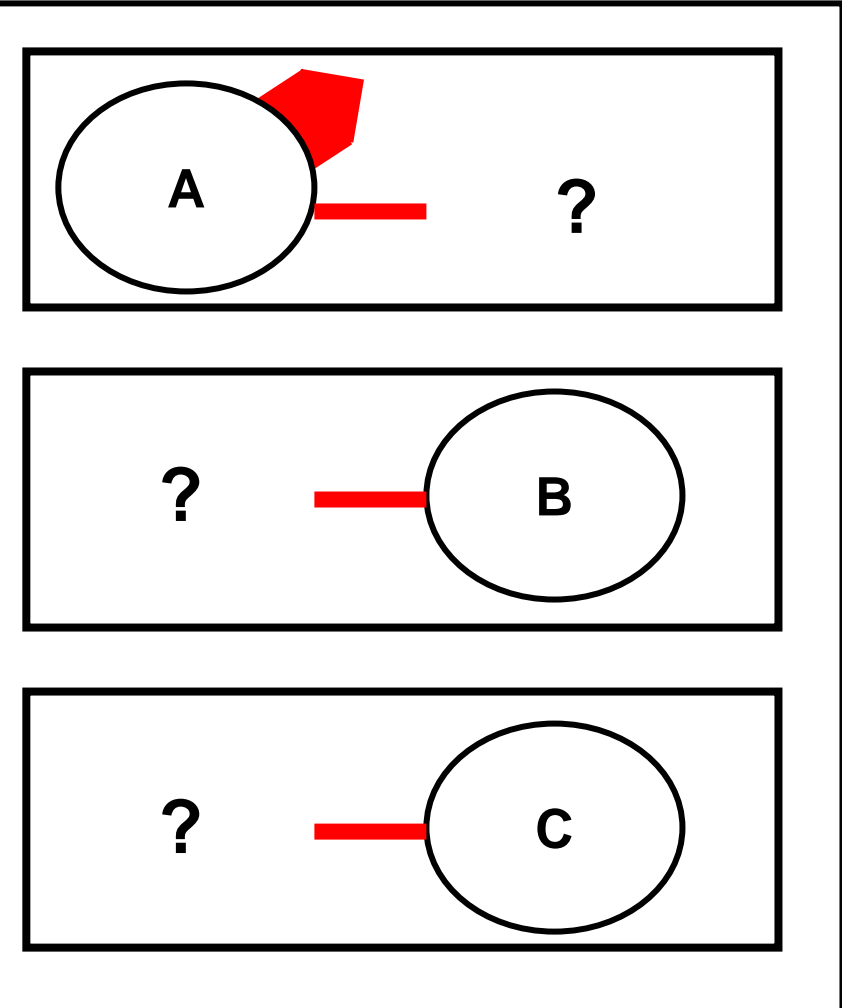
Here's another possibility?

The complex described by this experimental result is also said to have **Unknown Stoichiometry**.

# How complex data are stored in BIND.

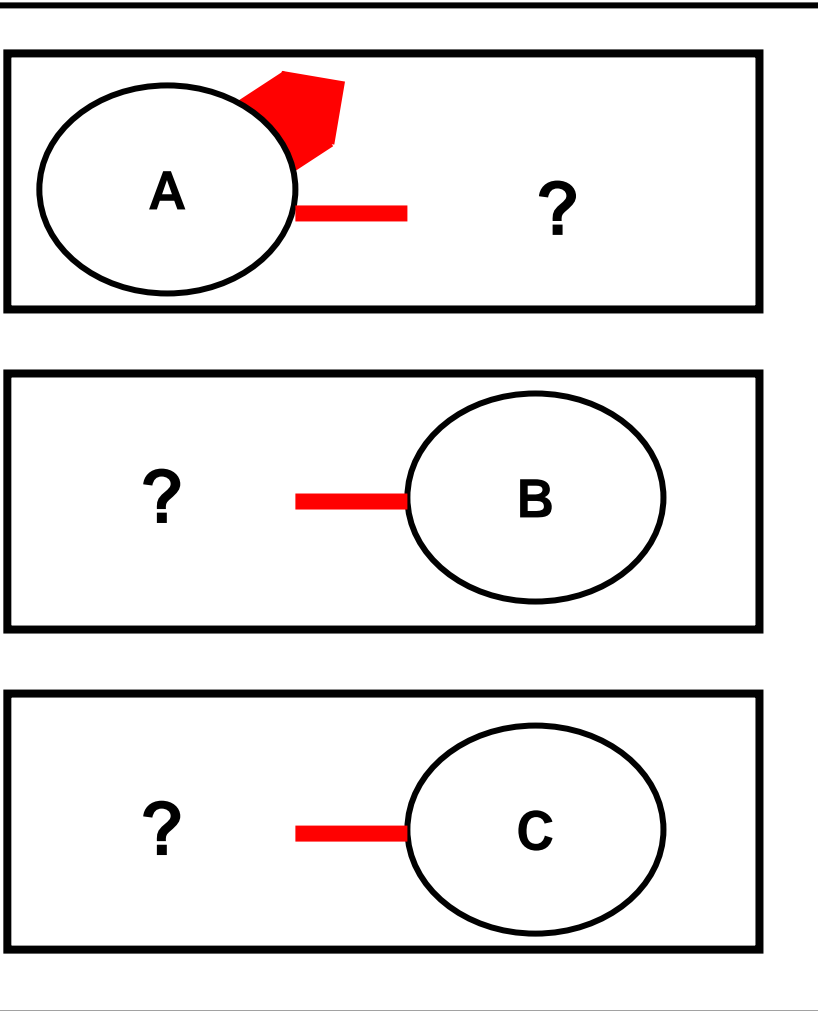


# How complex data are stored in BIND.



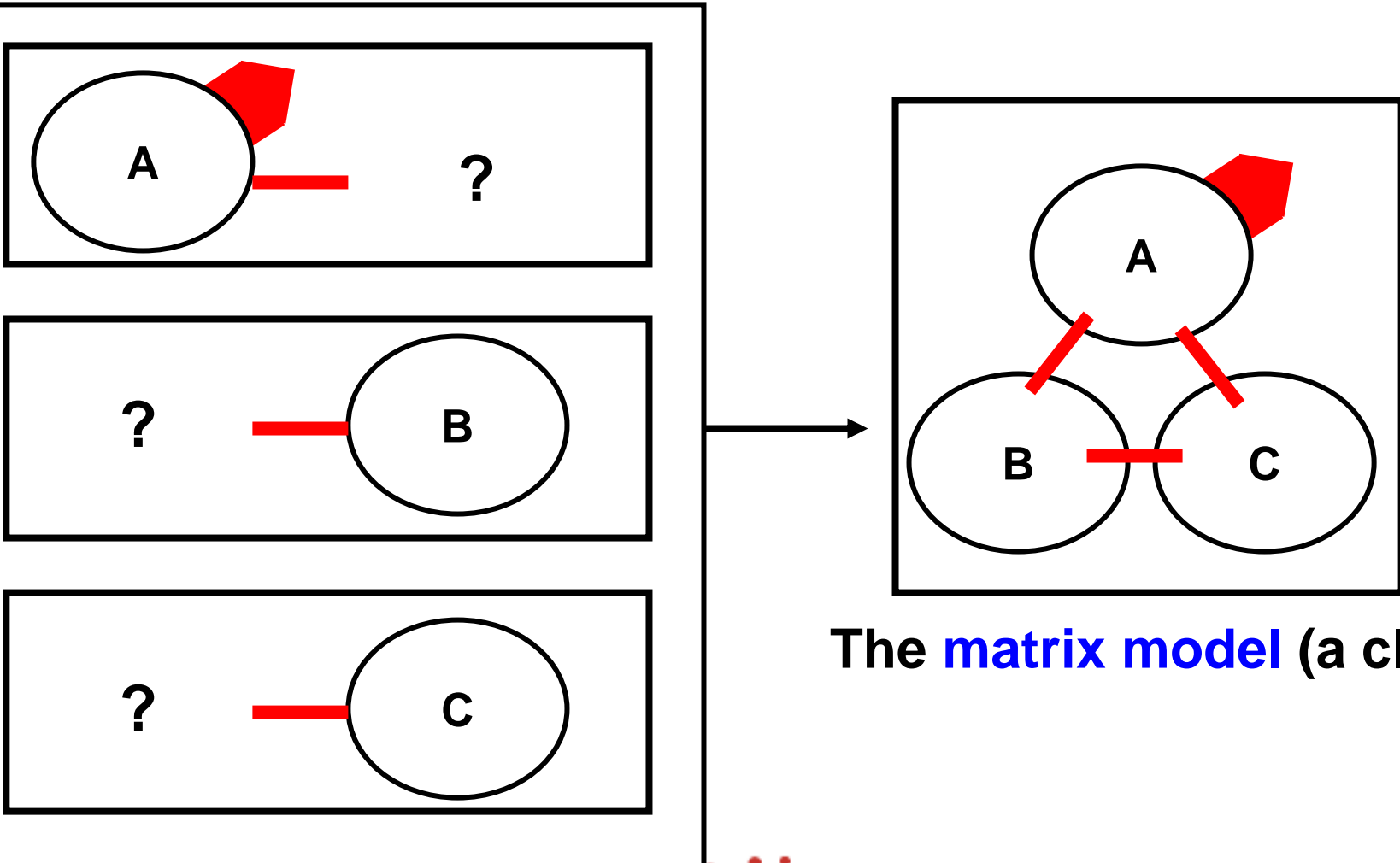
**A complex record in BIND is simply a collection of interaction records.**

# How complex data are stored in BIND.



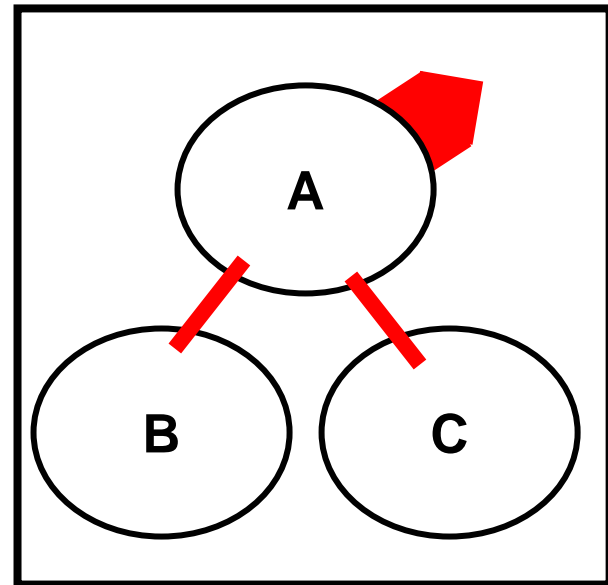
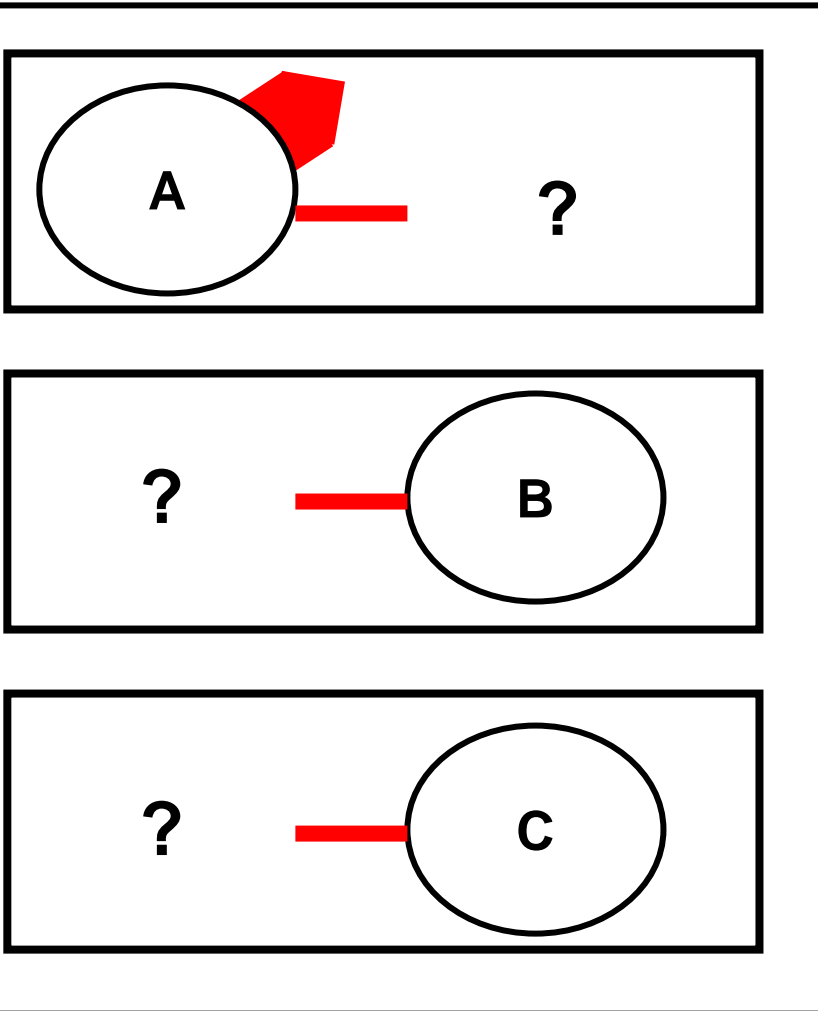
**A complex record in BIND is simply a collection of interaction records.**

# Alternate representations.



The **matrix model** (a clique).

# Alternate representations.



The **spoke model**.  
Which model you use  
Depends on what you are  
Doing with the data...  
maybe you don't care.

# High throughput data in BIND

- **Affinity purification:**  
Systematic identification of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry (2002). PMID: [11805837](#)
- **Two-hybrid:**  
A protein interaction map of *Drosophila Melanogaster* (2003). PMID: [14605208](#)
- **Two-hybrid and Affinity purification:**  
A map of the interactome network of the metazoan *C. Elegans* (2004). PMID: [14704431](#)
- Data from these examples can be retrieved from BIND using a PMID search.

# Use of high-throughput data

- Identifying members of a complex/pathway.
- Inferring function by association.
- Inferring interactions in other organisms.

# Other data in BIND

- Curated data
  - 21 curators
  - 150 interaction records per week
- MMDB BIND
  - Interactions found in the Molecular Modelling Database (NCBI's curated version of PDB)
  - Includes protein-small molecule interactions

# In the lab

- Tools for working with BIND data
- Field-specific text-searching
- Accession number searches
- BIND BLAST
- PreBIND
- The Interaction viewer
- FAST
- SeqHound





