

W22. Protein-Protein Interactions

February 2020

Lukasz Salwinski
lukasz@mbi.ucla.edu
Boyer Hall 205

Workshop Overview

Day I

Introduction *presentation*

- Experimental Methods
 - Experimental strategies overview
 - Experiment uncertainties
 - Interaction types
 - Ambiguous protein identification
- Database Record Contents
 - Record structure
 - Controlled vocabularies
 - Record transformations
- Protein Interaction Resources
 - Primary Databases
 - Integrators

Day II

Interaction Data Processing *hands on*

- Database Record Formats
 - PSI-MI Standards
 - BioGRID
 - Cytoscape/SIF
- Parsing XML Files
 - Python lxml toolkit
- PSI-MI MIF Format
 - Record structure
 - Compact vs expanded form
- SIF (Cytoscape) Format
 - MIF to SIF conversion

NOTE

Working knowledge of Python required. The day can be skipped by non-programmers.

Day III

Cytoscape *hands on*

- Interaction Networks
 - Record structure
 - Compact vs expanded form
- Cytoscape Overview
 - Installation
 - Basic features
 - Apps/Plugins overview
- Loading Interaction Data
 - Local files
 - Public resources
- Basic Network Visualization
 - Expression data
 - Functional enrichment

Protein-Protein Interactions

Introduction

May 2019

Lukasz Salwinski
lukasz@mbi.ucla.edu
Boyer Hall 205

Molecular Interactions

Why do we care ?

The screenshot shows the UniProtKB entry for P04637 (TP53_HUMAN). The main content area displays the following information:

- Protein:** Cellular tumor antigen p53
- Gene:** TP53
- Organism:** Homo sapiens (Human)
- Status:** Reviewed - Annotation score: 5/5 - Experimental evidence at protein levelⁱ

The sidebar on the left contains a list of filters, many of which are checked:

- Function
- Names & Taxonomy
- Subcellular location
- Pathology & Biotech
- PTM / Processing
- Expression
- Interaction
- Structure
- Family & Domains
- Sequences (8+)
- Similar proteins
- Cross-references
- Entry information
- Miscellaneous

The main content area has a highlighted section titled "Functionⁱ" containing a detailed description of TP53's role as a tumor suppressor. It mentions its involvement in cell cycle regulation, apoptosis induction, and its interaction with other genes like BAX and FAS.

Below the function section is a "Cofactorⁱ" section for Zn²⁺, noting it binds 1 zinc ion per subunit. There is also a "Sites" section showing metal binding sites at positions 176, 179, 238, and 242, each associated with Zinc.

A blue box at the bottom right contains the URL: <https://www.uniprot.org/uniprot/P04637>

Function

- What does a protein do in a cell ?
- What happens when a protein is missing ?
- What happens when a protein is altered ?

Molecular Interactions

Why do we care ?

UniProtKB - P04637 (P53_HUMAN)

GO - Biological processⁱ

Display

Entry

Publications

Feature viewer

Feature table

None

Function

Names & Taxonomy

Subcellular location

Pathology & Biotech

PTM / Processing

Expression

Interaction

Structure

Family & Domains

Sequences (8+)

Similar proteins

Cross-references

Entry information

Miscellaneous

Autophagy ✓ Source: CAFA ✓

base-excision repair ✓ Source: UniProtKB ✓

cell aging ✓ Source: UniProtKB ✓

cell cycle arrest ✓ Source: BHF-UCL ✓

cell differentiation ✓ Source: UniProtKB ✓

cell population proliferation ✓ Source: UniProtKB ✓

cellular protein localization ✓ Source: UniProtKB ✓

cellular response to actinomycin D ✓ Source: CAFA ✓

cellular response to DNA damage stimulus ✓ Source: UniProtKB ✓

cellular response to drug ✓ Source: UniProtKB ✓

cellular response to gamma radiation ✓ Source: CAFA ✓

cellular response to glucose starvation ✓ Source: UniProtKB ✓

cellular response to hypoxia ✓ Source: UniProtKB ✓

cellular response to ionizing radiation ✓ Source: BHF-UCL ✓

cellular response to UV ✓ Source: CAFA ✓

chromatin assembly ✓ Source: UniProtKB ✓

circadian behavior ✓ Source: UniProtKB ✓

cytokine-mediated signaling pathway ✓ Source: Reactome ✓

determination of adult lifespan ✓ Source: BHF-UCL ✓

DNA damage response, signal transduction by p53 class mediator ✓ Source: BHF-UCL ✓

DNA damage response, signal transduction by p53 class mediator resulting in cell cycle arrest ✓ Source: CAFA ✓

DNA damage response, signal transduction by p53 class mediator resulting in transcription of p21 class mediator ✓ Source: CAFA ✓

DNA strand renaturation ✓ Source: UniProtKB ✓

entrainment of circadian clock by photoperiod ✓ Source: UniProtKB ✓

ER overload response ✓ Source: MGI ✓

intrinsic apoptotic signaling pathway ✓ Source: HGNC ✓

intrinsic apoptotic signaling pathway by p53 class mediator ✓ Source: UniProtKB ✓

intrinsic apoptotic signaling pathway in response to DNA damage by p53 class mediator ✓ Source: UniProtKB ✓

mitotic G1 DNA damage checkpoint ✓ Source: BHF-UCL ✓

mRNA transcription ✓ Source: UniProtKB ✓

multicellular organism development ✓ Source: UniProtKB ✓

negative regulation of apoptotic process ✓ Source: UniProtKB ✓

Function

- What does a protein do in a cell ?
- What happens when a protein is missing ?
- What happens when a protein is altered ?

Gene Ontology Annotation

- Computer-readable description of function
- Three complementary sections
 - Biological process
 - Molecular function
 - Cellular localization

Molecular Interactions

Why do we care ?

The screenshot shows the UniProtKB interface for the protein P04637 (P53_HUMAN). The left sidebar has a 'Display' tab selected, showing various filter options like 'Entry', 'Publications', 'Feature viewer', and 'Feature table'. Under 'Function', several checkboxes are checked, including 'Function', 'Names & Taxonomy', 'Subcellular location', 'Pathology & Biotech', 'PTM / Processing', 'Expression', 'Interaction', 'Structure', 'Family & Domains', 'Sequences (8+)', 'Similar proteins', 'Cross-references', 'Entry information', and 'Miscellaneous'. The main content area shows a list of molecular functions with their sources: ATP binding (UniProtKB), chaperone binding (UniProtKB), chromatin binding (UniProtKB), copper ion binding (UniProtKB), core promoter sequence-specific DNA binding (CAFA), disordered domain specific binding (CAFA), DNA binding (UniProtKB), DNA-binding transcription activator activity, RNA polymerase II-specific (ARUK-UCL), DNA-binding transcription factor activity (UniProtKB), DNA-binding transcription factor activity, RNA polymerase II-specific (UniProtKB), enzyme binding (UniProtKB), histone acetyltransferase binding (UniProtKB), histone deacetylase binding (CAFA), identical protein binding (CAFA), mRNA 3'-UTR binding (CAFA), p53 binding (CAFA), promoter-specific chromatin binding (UniProtKB), protease binding (UniProtKB), protein heterodimerization activity (UniProtKB), protein kinase binding (UniProtKB), protein N-terminus binding (UniProtKB), protein phosphatase 2A binding (UniProtKB), protein phosphatase binding (UniProtKB), protein self-association (AgBase), receptor tyrosine kinase binding (BHF-UCL), RNA polymerase II distal enhancer sequence-specific DNA binding (UniProtKB), RNA polymerase II proximal promoter sequence-specific DNA binding (ParkinsonsUK-UCL), RNA polymerase II transcription factor binding (BHF-UCL), TFIID-class transcription factor complex binding (ParkinsonsUK-UCL), transcription factor binding (UniProtKB), transcription regulatory region DNA binding (BHF-UCL), and ubiquitin protein ligase binding (UniProtKB).

Function

- What does a protein do in a cell ?
- What happens when a protein is missing ?
- What happens when a protein is altered ?

Gene Ontology Annotation

- Computer-readable description of function
- Three complementary sections
 - Biological process
 - Molecular function
 - Cellular localization

Protein Function
is executed through
Molecular Interactions

Molecular Interactions

Why do we care ?

Feature key	Position(s)	Description	Actions	Graphical view	Length
Mutagenesis ⁱ	15	S → A: Loss of interaction with PPP2R5C, PPP2CA AND PPP2R1A. ↳ 1 Publication			1
Mutagenesis ⁱ	18	T → A: No effect on interaction with MDM2 and increase in protein levels after DNA damage. ↳ 1 Publication			1
Mutagenesis ⁱ	20	S → A: Abolishes phosphorylation site. Abolishes increase in protein levels after DNA damage. ↳ 1 Publication			1
Mutagenesis ⁱ	20	S → D: Constitutively increased TP53 protein levels. ↳ 1 Publication			1
Mutagenesis ⁱ	22 – 23	LW → QS: Loss of interaction with MDM2, leading to constitutively increased TP53 protein levels. ↳ 1 Publication			2
Mutagenesis ⁱ	37	S → D: Abolishes phosphorylation by MAPKAPK5. ↳ 1 Publication			1
Mutagenesis ⁱ	46	S → A: Abolishes phosphorylation by DYRK2 and HIPK2 and acetylation of K-382 by CREBBP. ↳ 3 Publications			1
Mutagenesis ⁱ	46	Missing : Alters interaction with WWOX. ↳ 3 Publications			1
Mutagenesis ⁱ	55	T → A: Blocks phosphorylation by TAF1. ↳ 1 Publication			1
Mutagenesis ⁱ	183	S → A: Abolishes strongly phosphorylation. ↳ 1 Publication			1
Mutagenesis ⁱ	183	S → E: Inhibits slightly its transcriptional activity. ↳ 1 Publication			1
Mutagenesis ⁱ	240	R → S: Does not induce SNAI1 degradation. ↳ 1 Publication			1
Mutagenesis ⁱ	269	S → A: Abolishes phosphorylation. ↳ 1 Publication			1
Mutagenesis ⁱ	269	S → E: Inhibits strongly its transcriptional activity. ↳ 1 Publication			1
Mutagenesis ⁱ	284	T → E: Inhibits strongly its transcriptional activity.			1
Mutagenesis ⁱ	291 – 292	KK → RR: Abolishes polyubiquitination by MKRN1. ↳ 1 Publication			2
Mutagenesis ⁱ	319	K → A: Loss of nuclear localization; when associated with A-320 and A-321. ↳ 1 Publication			1
Mutagenesis ⁱ	320	K → A: Loss of nuclear localization; when associated with A-319 and A-321. ↳ 1 Publication			1
Mutagenesis ⁱ	321	K → A: Loss of nuclear localization; when associated with A-319 and A-320. ↳ 1 Publication			1
Mutagenesis ⁱ	333 – 337	RGRER → KGKEK: Reduced methylation by PRMT5. Reduced nuclear localization. Decreased binding to promoters of target genes. Reduced transcriptional activity. Decrease in cell cycle arrest. ↳ 1 Publication			5

Function

- What does a protein do in a cell ?
- What happens when a protein is missing ?
- What happens when a protein is altered ?

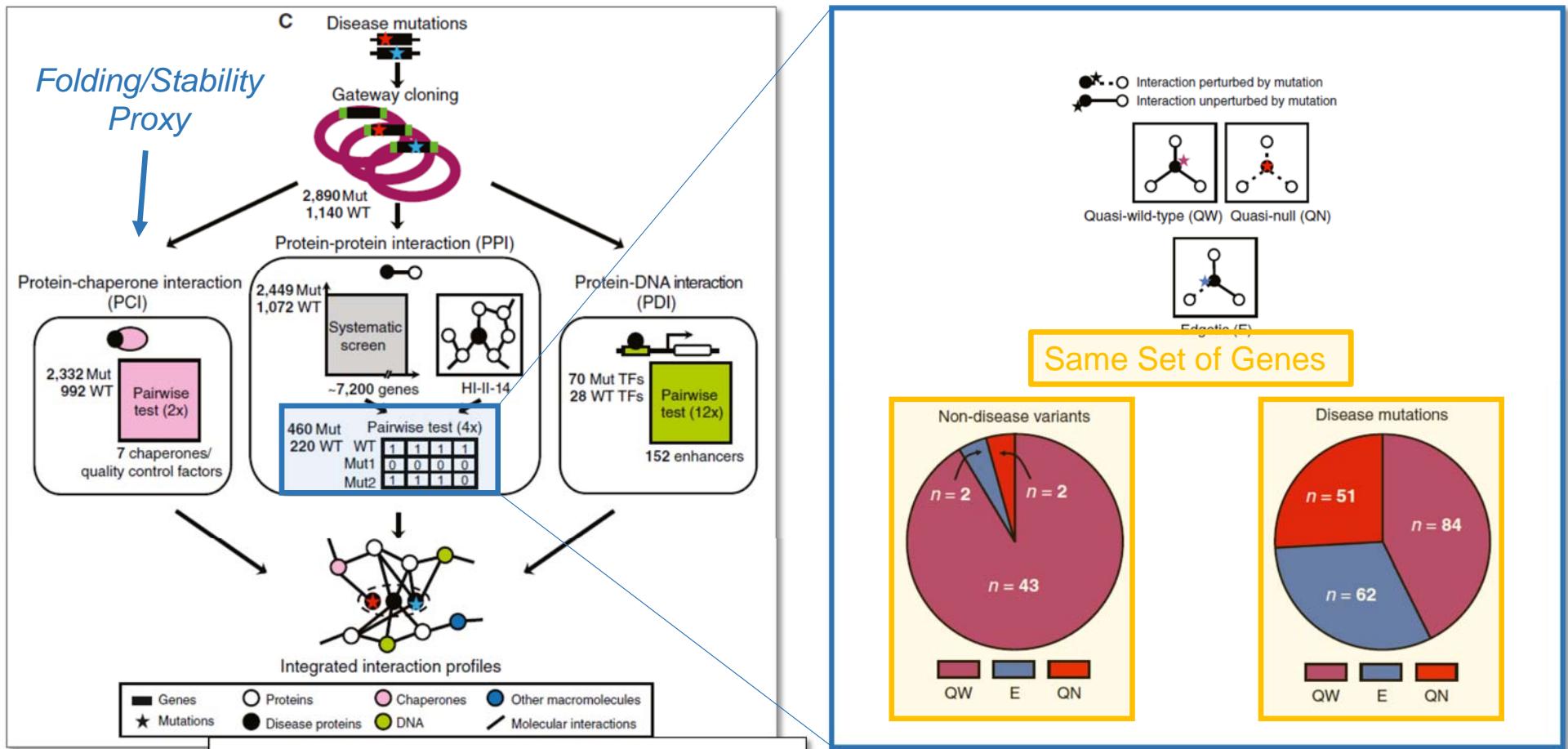
Gene Ontology Annotation

- Computer-readable description of function
- Three complementary sections
 - Biological process
 - Molecular function
 - Cellular localization

Protein mutations
may modulate
molecular interactions

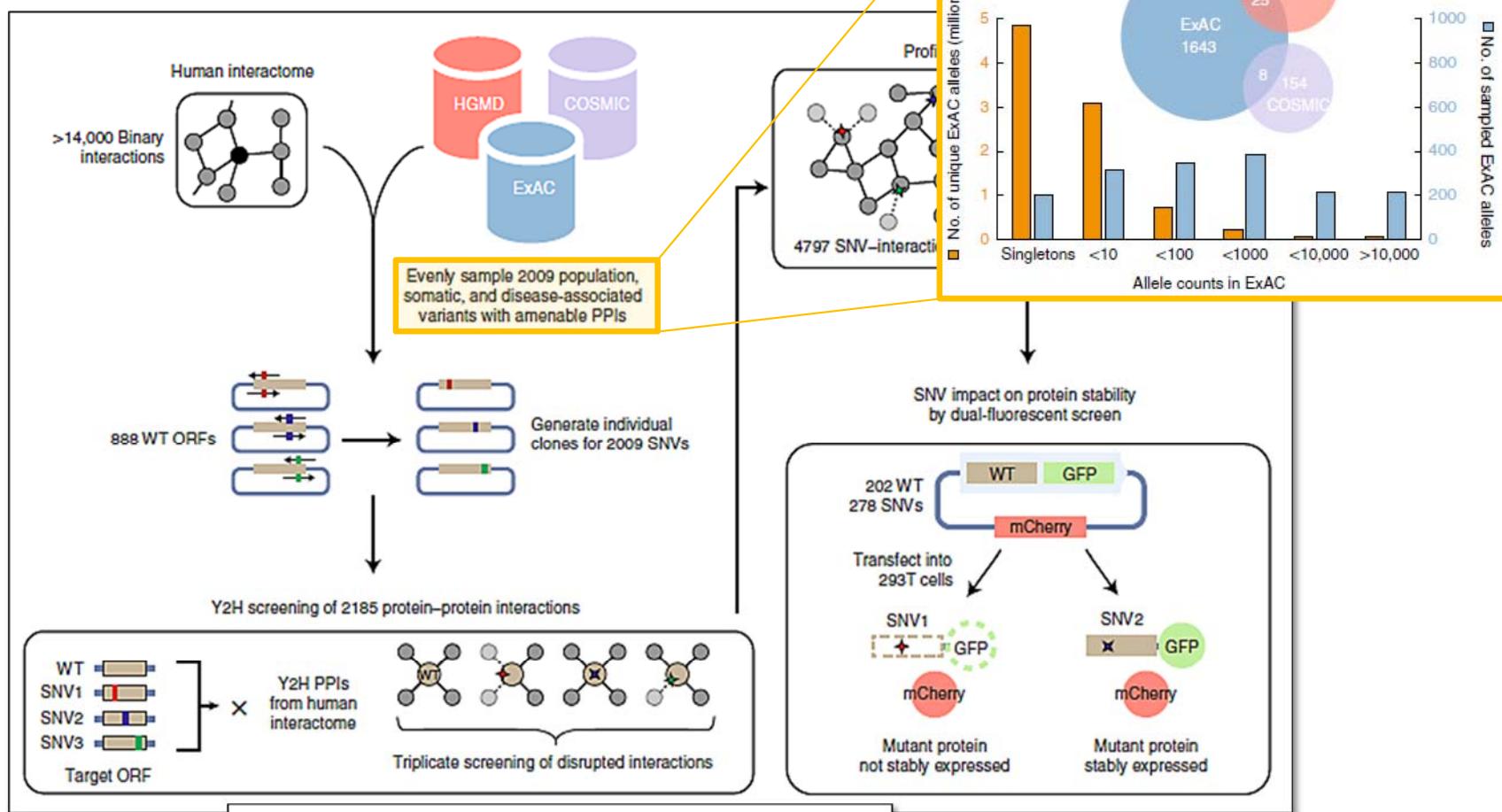
Molecular Interactions

Mutation/Disease Correlation



Molecular Interactions

Mutation/Disease Correlation

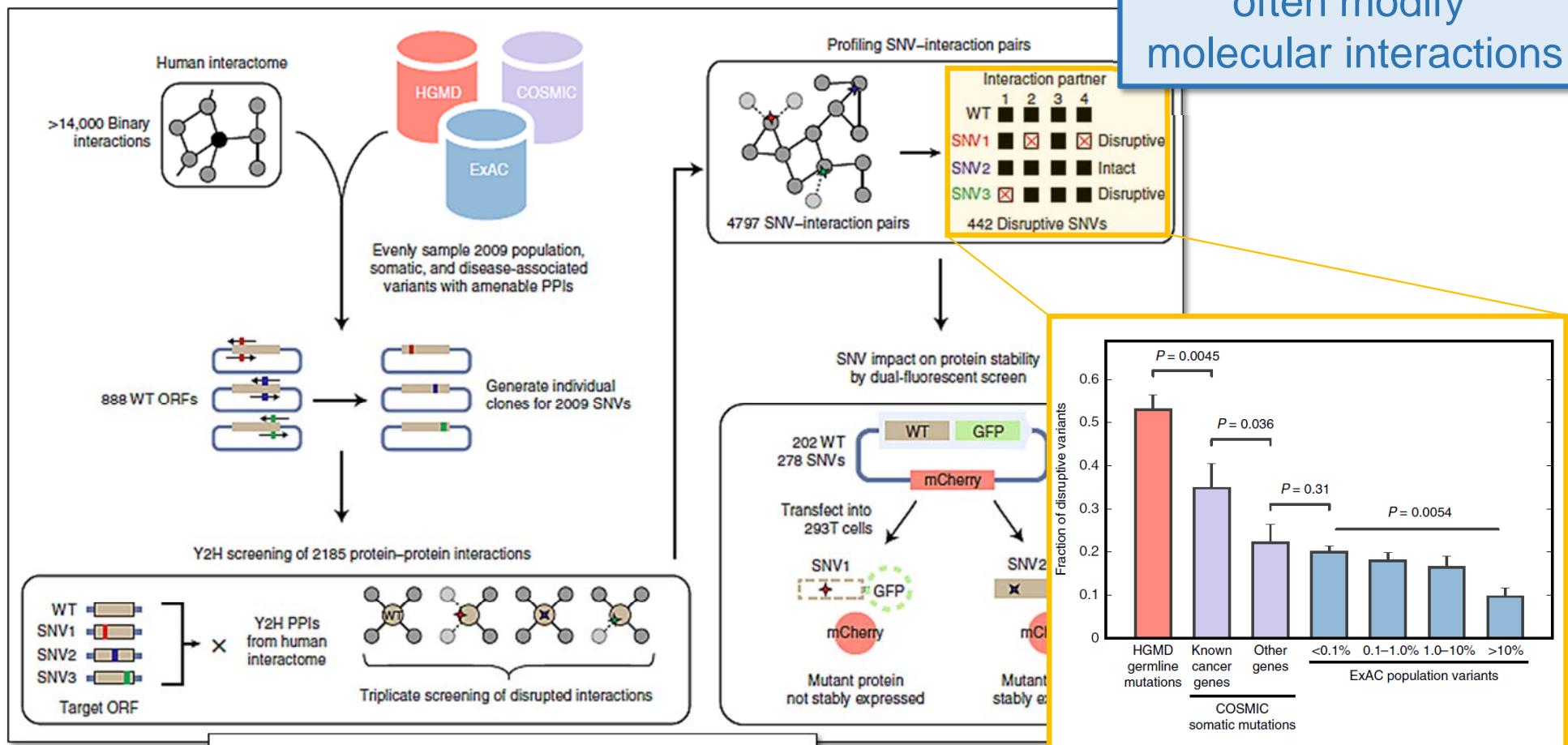


Fragoza et al. *Nat Com* 10:4141(2019); PMID:31515488

Molecular Interactions

Mutation/Disease Correlation

Disease mutations often modify molecular interactions



Fragoza et al. *Nat Com* 10:4141(2019); PMID:31515488

Molecular Interactions

Why do we care ?

- Protein interactions reflect protein function

To know how protein works means to know when and with what other molecules it interacts with

- Disruption of protein interactions often correlate with disease

Knowing perturbation of the protein function in disease helps to understand its mechanism and might provide insights into potential therapies

Experimental Methods

Molecular Interactions

Why do they come from?

The screenshot shows the UniProtKB interface for protein P04637 (P53_HUMAN). The left sidebar displays various annotation categories like Function, Names & Taxonomy, Subcellular location, etc. The main content area shows a table of mutations. A tooltip is overlaid on the row at position 18, which describes a mutation where T → A: No effect on interaction with MDM2 and increase in protein levels after DNA damage. The tooltip provides a detailed citation from Hofmann et al. (2002) and states that it is cited for interaction with HIPK2.

Feature key	Position(s)	Description	Actions	Graphical view	Length
Mutagenesis ⁱ	15	S → A: Loss of interaction with PPP2R5C, PPP2CA AND PPP2R1A. <small>1 Publication</small>			1
Mutagenesis ⁱ	18	T → A: No effect on interaction with MDM2 and increase in protein levels after DNA damage. <small>1 Publication</small>			1
Mutagenesis ⁱ	20	S → A: Abolishes phosphorylation site. Abolishes increase in protein levels after DNA damage. <small>1 Publication</small>			1
Mutagenesis ⁱ	20	S → D: Constitutively increased TP53 protein levels. <small>1 Publication</small>			1
Mutagenesis ⁱ	22 – 23	LW → QS: Loss of interaction with MDM2, leading to constitutively increased TP53 protein levels. <small>1 Publication</small>			2
Mutagenesis ⁱ	37	S → D: Abolishes phosphorylation by MAPKAPK5. <small>1 Publication</small>			1
Mutagenesis ⁱ	46	S → A: Abolishes phosphorylation by DYRK2 and HIPK2 and acetylation of K-382 by CREBBP. <small>3 Publications</small>			1
Mutagenesis ⁱ	46	N → D: Abolishes phosphorylation by DYRK2 and HIPK2 and acetylation of K-382 by CREBBP. <small>3 Publications</small>			1
Mutagenesis ⁱ	55	T → S: Inhibits transcriptional activity. <small>1 Publication</small>			1
Mutagenesis ⁱ	183	"Regulation of p53 activity by its interaction with homeodomain-interacting protein kinase-2." Hofmann T.G., Moeller A., Sirma H., Zentgraf H., Taya Y., Droege W., Will H., Schmitz M.L. Nat. Cell Biol. 4:1-10[2002] [PubMed] [Europe PMC]			1
Mutagenesis ⁱ	183	[Abstract]			1
Mutagenesis ⁱ	240	Cited for: INTERACTION WITH HIPK2, PHOSPHORYLATION AT SER-46, MUTAGENESIS OF SER-46 AND LYS-382.			1
Mutagenesis ⁱ	269	T → E: Inhibits strongly its transcriptional activity.			1
Mutagenesis ⁱ	284	KK → RR: Abolishes polyubiquitination by MKRN1. <small>1 Publication</small>			2
Mutagenesis ⁱ	291 – 292	K → A: Loss of nuclear localization; when associated with A-320 and A-321. <small>1 Publication</small>			1
Mutagenesis ⁱ	319	K → A: Loss of nuclear localization; when associated with A-319 and A-321. <small>1 Publication</small>			1
Mutagenesis ⁱ	320	K → A: Loss of nuclear localization; when associated with A-319 and A-321. <small>1 Publication</small>			1
Mutagenesis ⁱ	321	K → A: Loss of nuclear localization; when associated with A-319 and A-320. <small>1 Publication</small>			1
Mutagenesis ⁱ	333 – 337	RGRER → KGKEK: Reduced methylation by PRMT5. Reduced nuclear localization. Decreased binding to promoters of target genes. Reduced transcriptional activity. Decrease in cell cycle arrest. <small>1 Publication</small>			5

Function

- What does a protein do in a cell ?
- What happens when a protein is missing ?
- What happens when a protein is altered ?

Gene Ontology Annotation

- Computer-readable description of function
- Three complementary sections
 - Biological process
 - Molecular function
 - Cellular localization
- Each GO term is backed up by evidence

Molecular Interactions

Why do they come from?

The screenshot shows the UniProtKB entry for P04637 (P53_HUMAN). The left sidebar has 'Function' selected. The main content lists GO-Molecular function annotations, many of which have evidence codes (e.g., CAFA, ARUK-UCL) and sources (UniProtKB or CAFA). A yellow box highlights an 'Inferred from physical interaction' section with the following text:

"A novel cofactor for p300 that regulates the p53 response."
Shikama N., Lee C.-W., France S., Delavaine L., Lyon J., Krstic-Demonacos M., La Thangue N.B.
Mol. Cell 4:365-376(1999) [PubMed] [Europe PMC]
[Abstract]

Function

- What does a protein do in a cell ?
- What happens when a protein is missing ?
- What happens when a protein is altered ?

Gene Ontology Annotation

- Computer-readable description of function
- Three complementary sections
 - Biological process
 - Molecular function
 - Cellular localization
- Each GO term is backed up by evidence
 - Experimental evidence
 - Physical interaction
 - Enzymatic assay
 - Expression level
 -
 - Sequence similarity
 - Curator's inference
 -

Interaction Experiments: Caveats

- Protein interactions are determined by 3D structure of a folded protein chain
 - Experimental conditions matter (ionic strength, pH, temperature, cofactors, etc)
 - Possible cellular compartment and tissue specificity
- Protein interactions might be modulated by post-translational modifications
 - Phosphorylation, methylation, etc. might matter
- Protein interactions might be transient
 - Life-cycle, cell-cycle, stimuli response, etc. might modify interaction patterns

Interaction Experiments: Strategies

- Structure determination

- X-Ray/Electron/Neutron diffraction
 - Cryo-EM
 - NMR

- Copurification

- Chromatography (size exclusion, affinity, ion exchange, reversed phase....)
 - Native electrophoresis
 - Co-sedimentation

- Solid Phase Assays

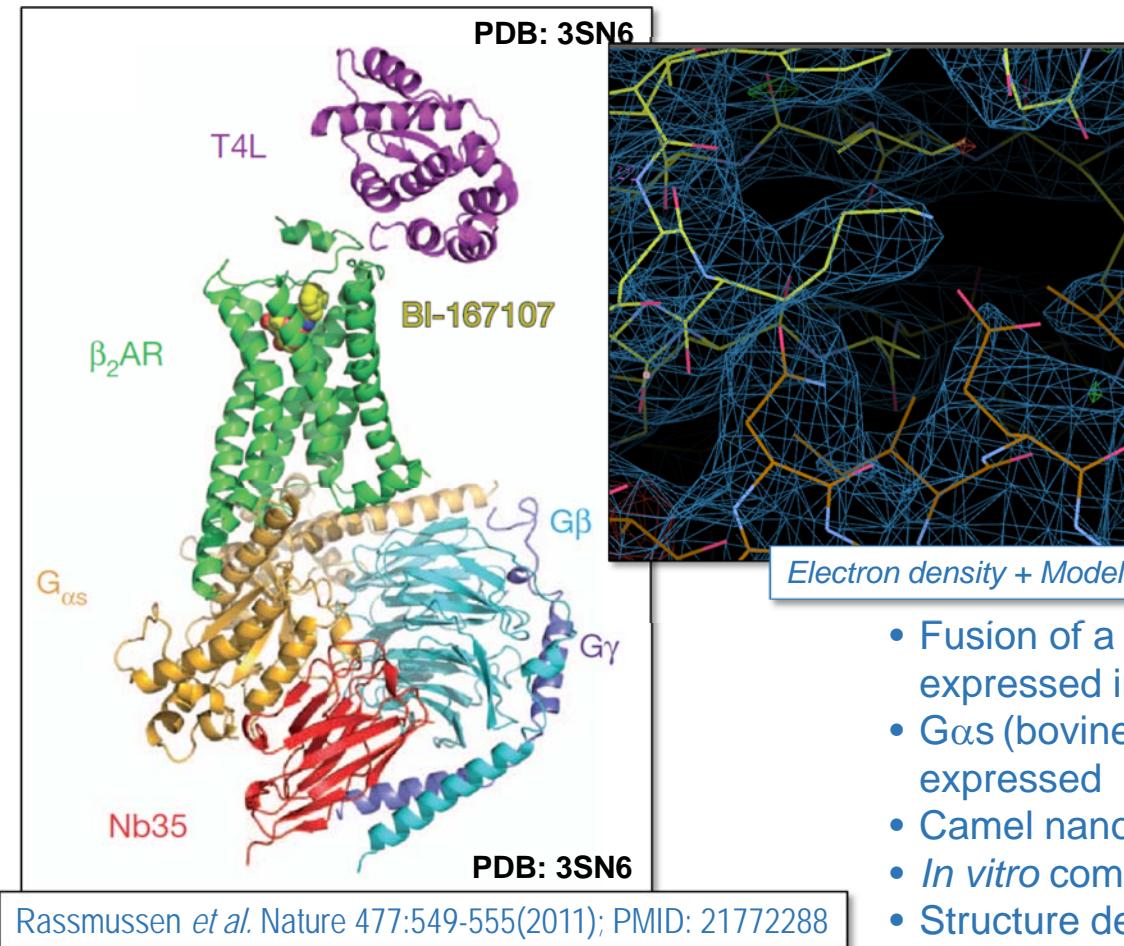
- ELISA (enzyme linked immunosorption assay)
 - Membrane-binding/far western assays
 - SPR (surface plasmon resonance) and like
 - ...

- Biochemical/biophysical property perturbation

- Spectroscopy-based methods (UV/Vis, Fluorescence, NMR, Mass Spec, etc)
 - Thermochemistry (calorimetry – ITC, DSC; thermal stability)
 - Co-sedimentation
 - ...

Interaction Experiments: Methods

X-Ray Crystallography



Advantages

- Detailed, atom-level description of the interacting molecules

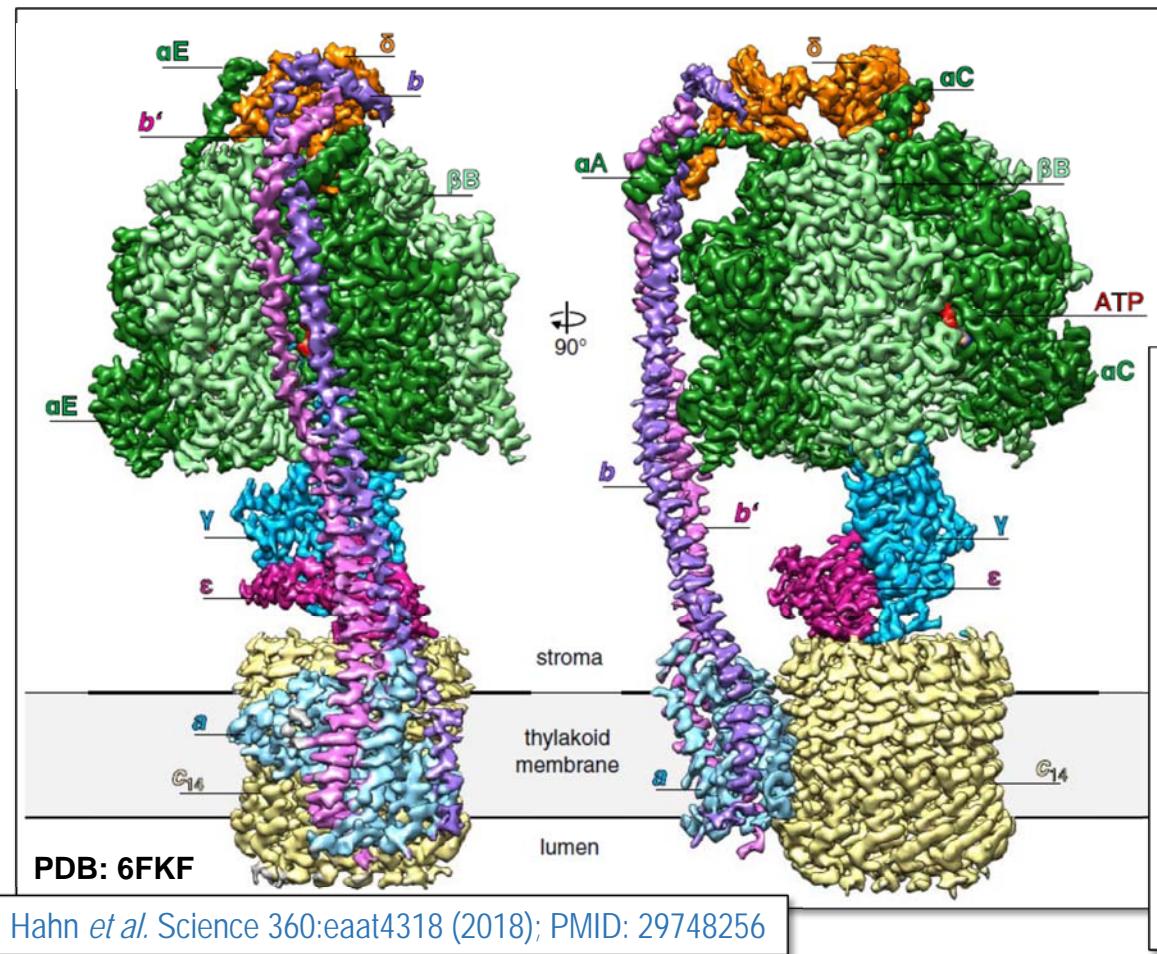
Disadvantages

- Expensive/slow/might not work at all
- Only one static image at a time

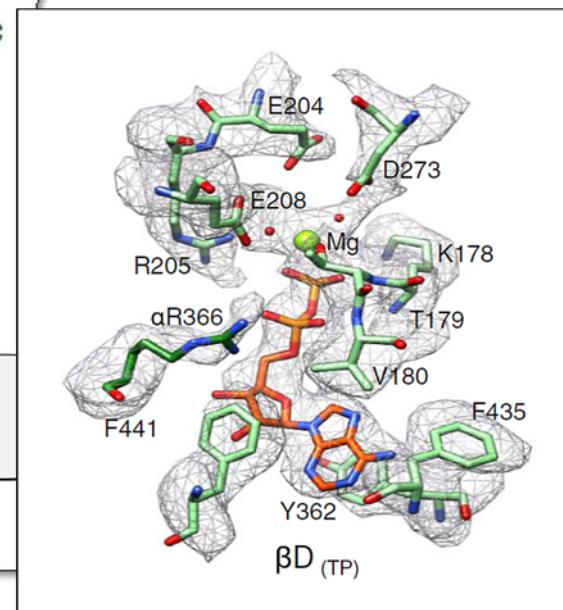
- Fusion of a fragment of human β_2 -AR with lysozyme fusion expressed in insect (Sf9) cells
- $G_{\alpha s}$ (bovine), $G_{\beta 1}$ (rat, His-tagged) and $G_{\gamma 2}$ (bovine) co-expressed in insect (HiFive) cells
- Camel nanobody expressed in E. coli
- *In vitro* complex reconstitution
- Structure determined by X-ray crystallography to 3.2A

Interaction Experiments: Methods

Cryo-EM (cryogenic electron microscopy)

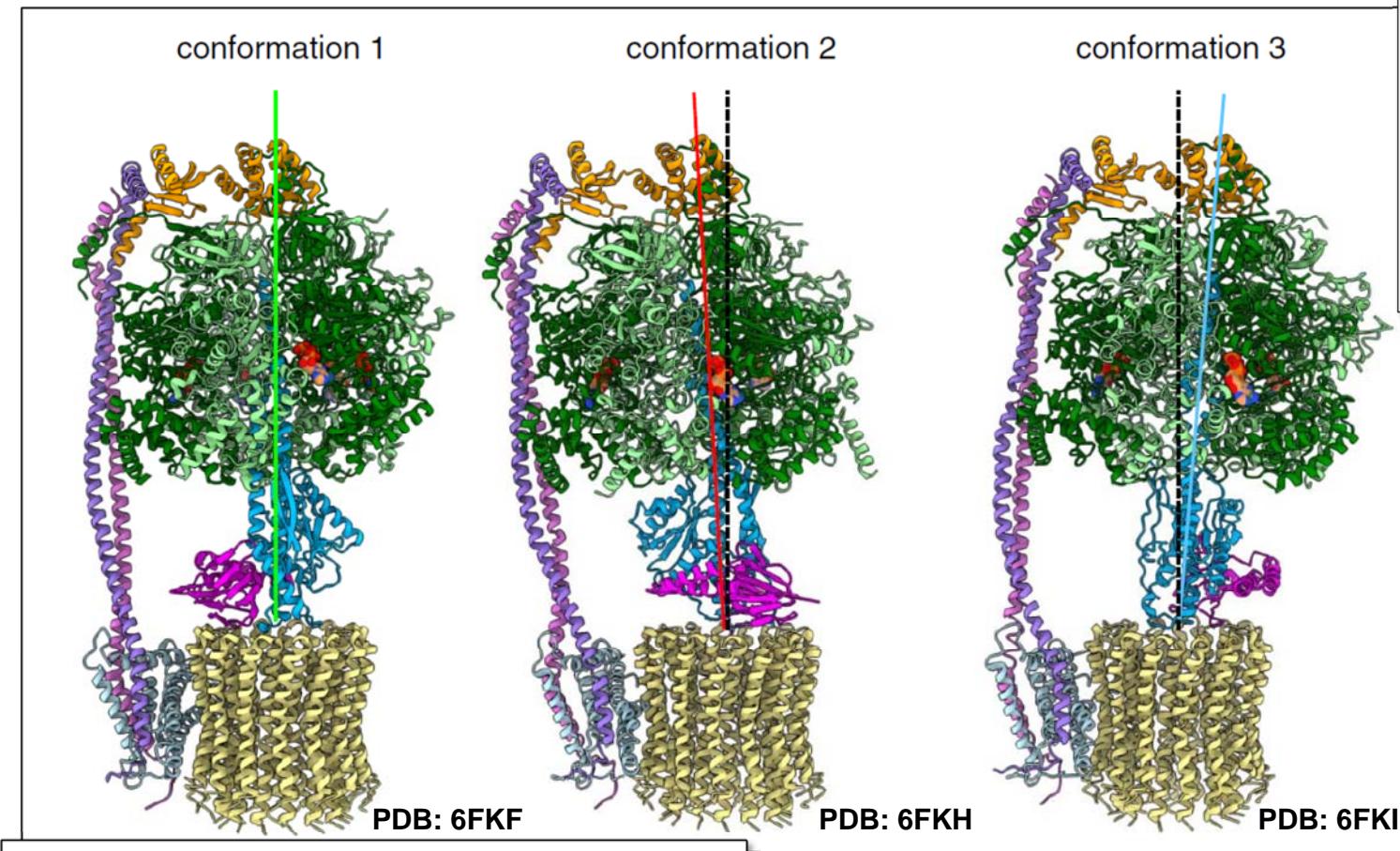


- Native expression (spinach)
- Cryo-EM structure determination
- 2.9-3.8 Å resolution

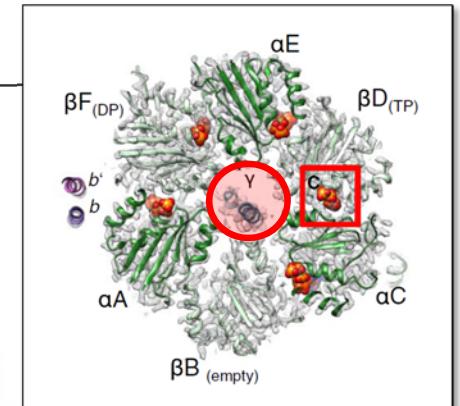


Interaction Experiments: Methods

Cryo-EM (cryogenic electron microscopy)



Hahn *et al.* Science 360:eaat4318 (2018); PMID: 29748256



- Conformational states provide insight into mechanism of ATP synthesis

Interaction Experiments: Strategies

- Structure determination

- X-Ray/Electron/Neutron diffraction
 - Cryo-EM
 - NMR

- Copurification

- Chromatography (size exclusion, affinity, ion exchange, reversed phase....)
 - Native electrophoresis
 - Co-sedimentation

- Solid Phase Assays

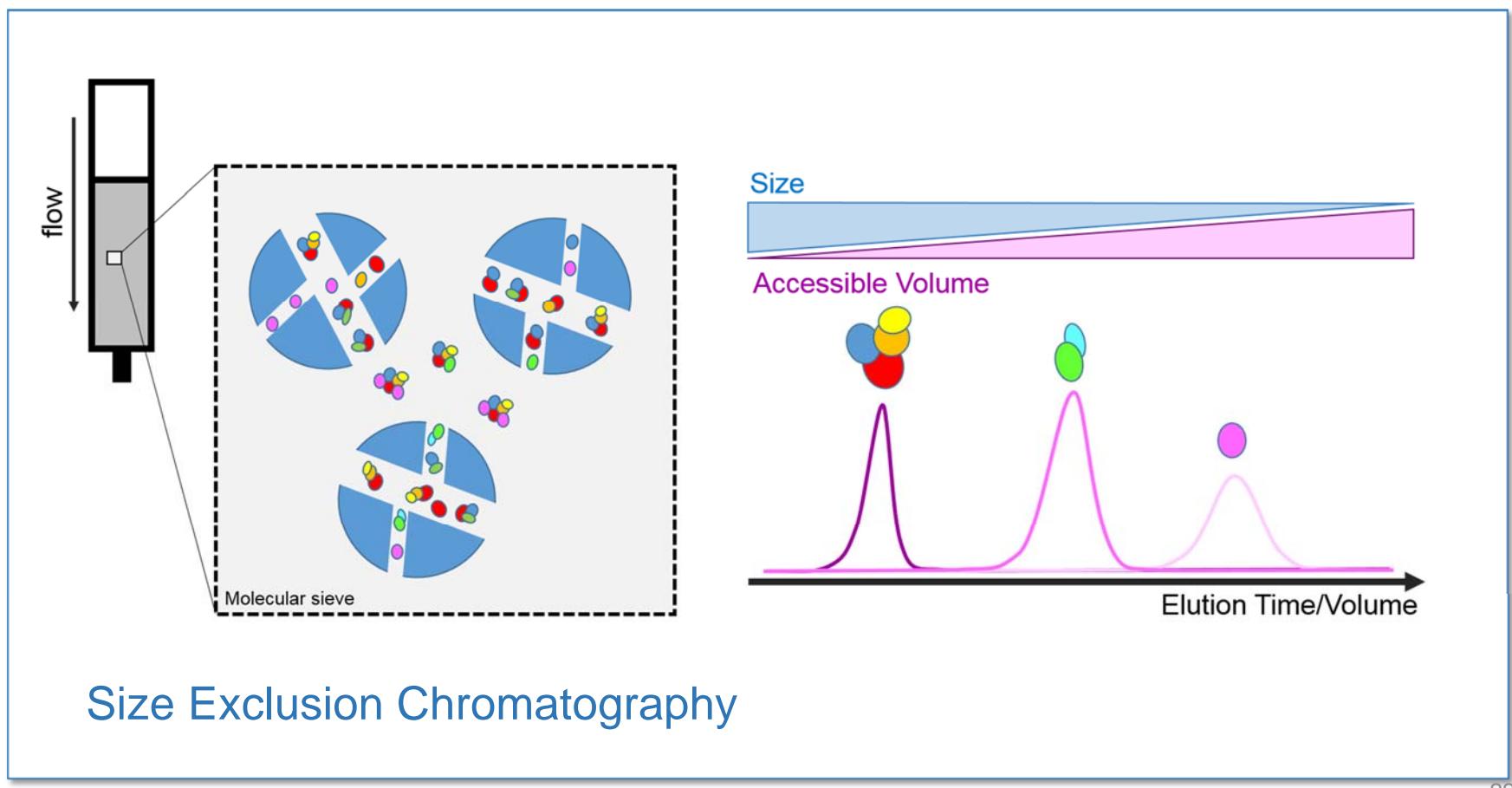
- ELISA (enzyme linked immunosorption assay)
 - Membrane-binding/far western assays
 - SPR (surface plasmon resonance) and like
 - ...

- Biochemical/biophysical property perturbation

- Spectroscopy-based methods (UV/Vis, Fluorescence, NMR, Mass Spec, etc)
 - Thermochemistry (calorimetry – ITC, DSC; thermal stability)
 - Co-sedimentation
 - ...

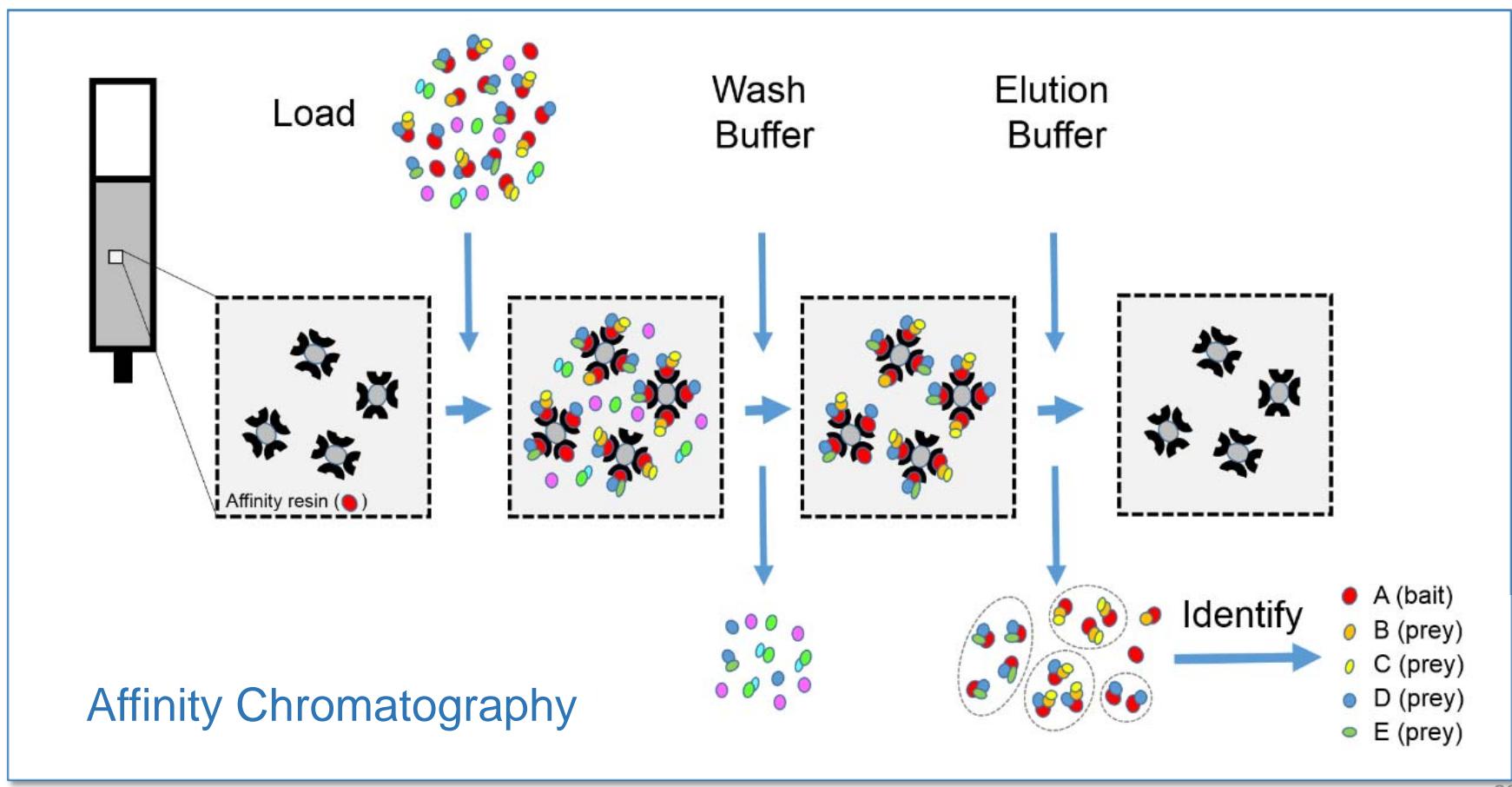
Interaction Experiments: Methods

Chromatography/Co-purification

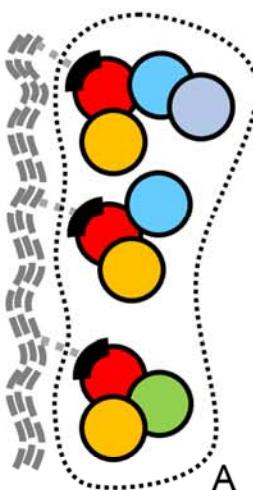
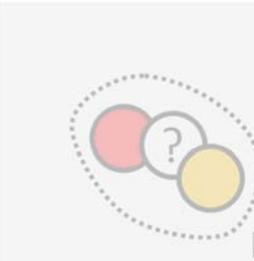
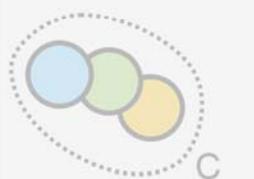


Interaction Experiments: Methods

Chromatography/Co-purification

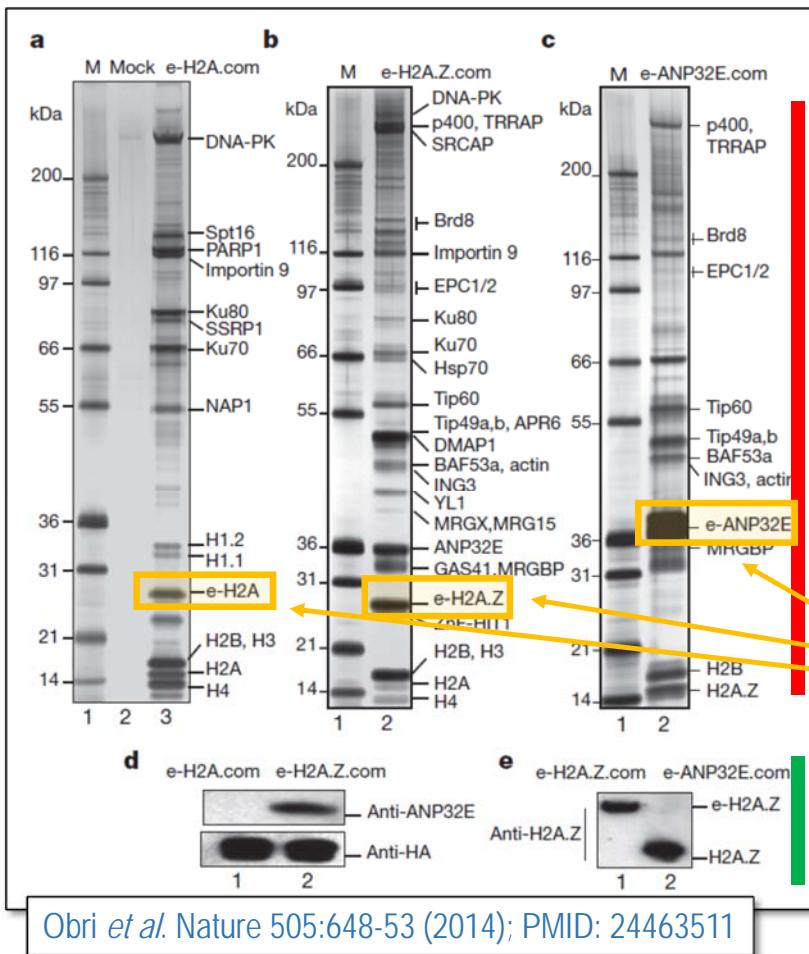


Interaction Types

	 	  	 	
Interaction Type	Association	Physical Association	Direct	Covalent
List(s) of Interacting Molecules	$[B]$ $[B, C]$ $[A]$	$[B, C]$ $[B, D]$ $[C]$	$[D]$ $[E]$ $[F]$	$[G]$ $[H]$

Interaction Experiments: Methods

Chromatography/Copurification



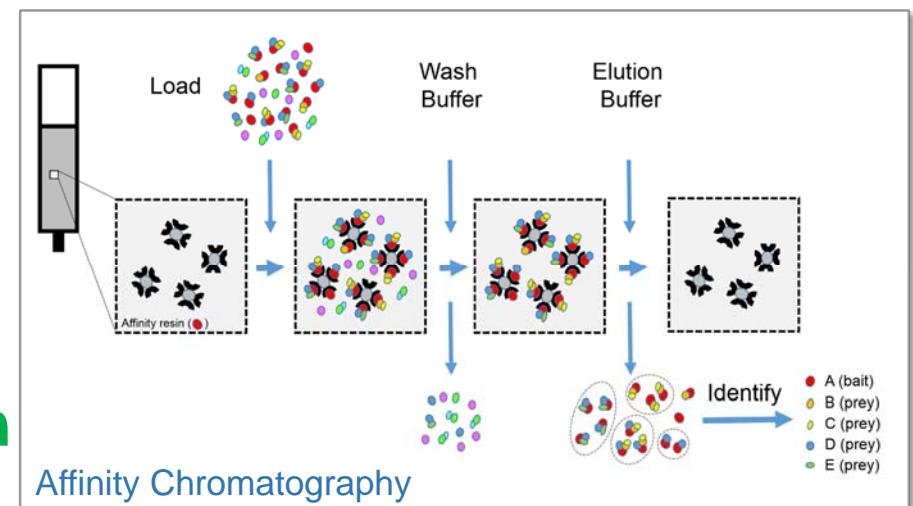
Obri *et al.* Nature 505:648-53 (2014); PMID: 24463511

- Overexpression (human) of doubly-tagged (FLAG & HA) human H2A, H2A.Z & ANP32E proteins
- Tandem affinity purification of interacting partners
- Mass spectrometric and/or Western blot protein identification

MS

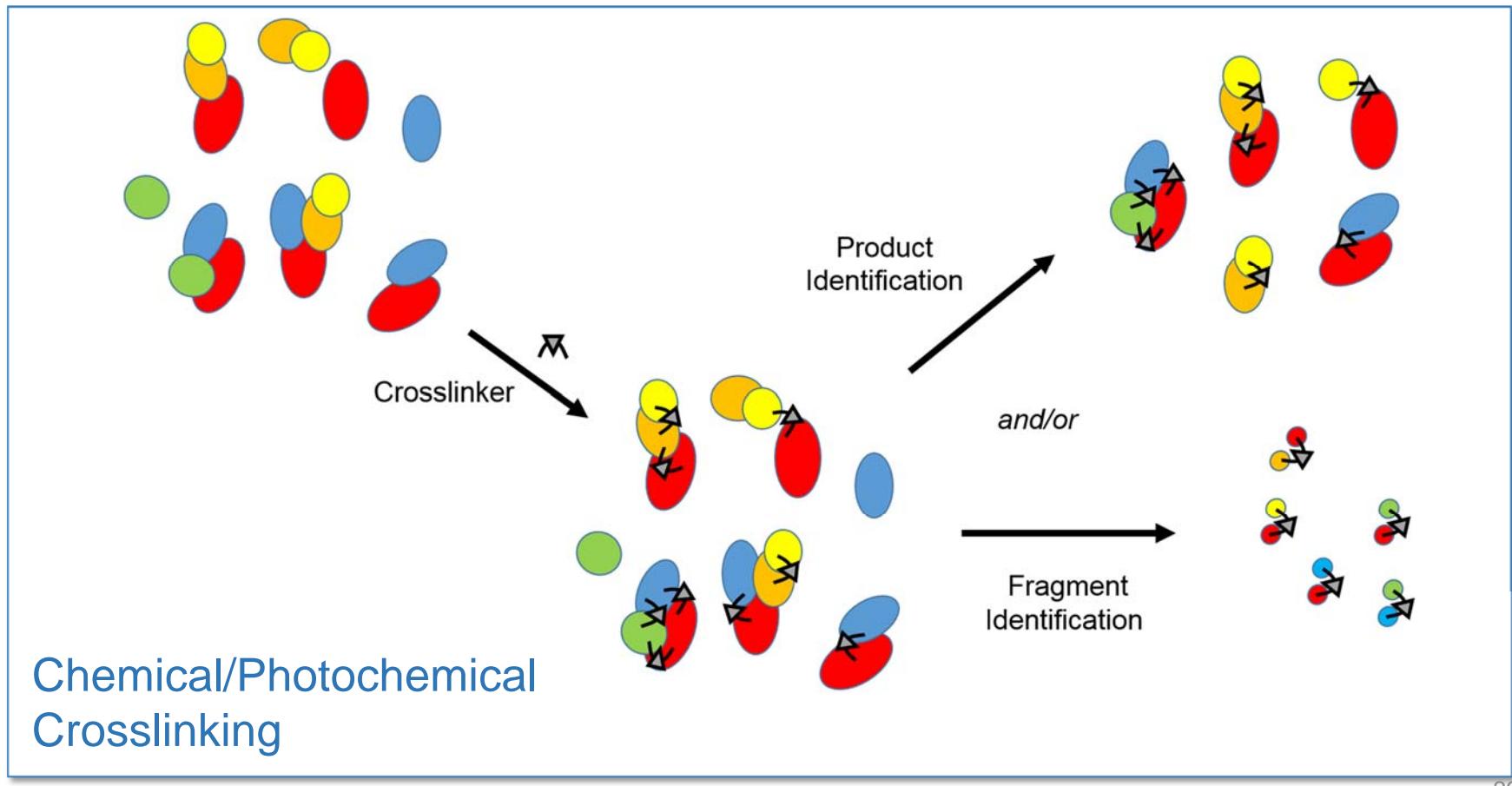
BAITS

Western



Interaction Experiments: Methods

When interacting complexes are not stable enough...



Interaction Experiments: Strategies

- Structure determination

- Copurification

- Optional crosslinking

- Solid Phase Assays

- ELISA (enzyme linked immunosorption assay)
 - Membrane-binding/far western assays
 - SPR (surface plasmon resonance) and like
 - ...

- Biochemical/biophysical property perturbation

- Spectroscopy-based methods (UV/Vis, Fluorescence, NMR, Mass Spec, etc)
 - Thermochemistry (calorimetry – ITC, DSC; thermal stability)
 - Co-sedimentation
 - ...

- Complementation (Bio-) Assays

- Yeast Two Hybrid
 -

Advantages

- Cheap/Fast
- Can be used to identify interacting regions/residues
- Can detect weak interactions...

Disadvantages

- ... and also innocent bystanders

Advantages

- Detailed, atom-level description of the interacting molecules

Disadvantages

- Expensive/slow/might not work at all
- Only one static image at a time

Advantages

- Cheap/Fast
- Amenable to high-throughput approaches

Disadvantages

- Often only partial information:
 - Missing components
 - No stoichiometry
 - No contacts
- Might miss weak/transient interactions

Advantages

- May provide very detailed information about one particular interaction
- May be customized to address one particular question about interaction

Disadvantages

- Might be slow and/or not amenable to high throughput approaches
- Might require specialized equipment

Interaction Experiments

Where to learn more ?

Structural methods

Crystallography and Structure Determination Core Facility

<https://www.doe-mbi.ucla.edu/x-ray-crystallography-core-technology-center/>

Crystallization Core Facility

<https://www.doe-mbi.ucla.edu/crystallization/>

Biosciences NMR Core Facility

<https://www.doe-mbi.ucla.edu/ucla-doe-biosciences-nmr-core-facility/>

M230B/M230D: Winter quarter (annually)

- principles of X-ray crystallography, cryo-EM and NMR
- Lecture (M230B) and hands-on lab (M230D)
- Non-registered participants OK

230D Syllabus: <https://people.mbi.ucla.edu/sawaya/m230d>



Chem 257 – Winter Quarter – 2020
Macromolecular Interactions
Tuesday and Thursday (1:00-1:50 p.m.)
3069 Young Hall

Instructors:
Jose Rodriguez, Emil Reisler and Martin Phillips

In this course, you will examine macromolecular interactions and the parameters that describe them. The course focuses on various modern experimental techniques that measure or map these interactions, including sedimentation, light scattering, titration calorimetry, and fluorescence methods. Expert guests will instruct on applications of mass spectrometry, fluorescence imaging, EPR, cryoEM, Atomic Force Microscopy, molecular modeling and molecular docking to measure, map, and quantify macromolecular interactions. The use of these methods will be practiced or demonstrated in a lab component held 2-4 hrs/week, in a small (4-5) group setting.

For more information please contact:
Jose Rodriguez (jrodriguez@mbi.ucla.edu) or Emil Reisler (reisler@mbi.ucla.edu)



Interaction Experiments

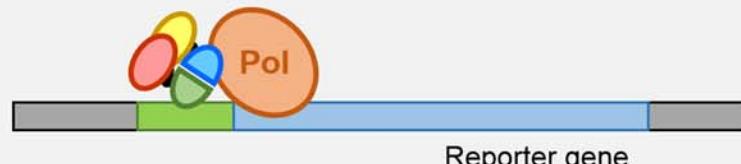
Complementation Assays

Construct(s)	Phenotype
	+
	-
	+

Phenotype

- Gene Expression
- Fluorescence
- ...

Yeast Two Hybrid Assay

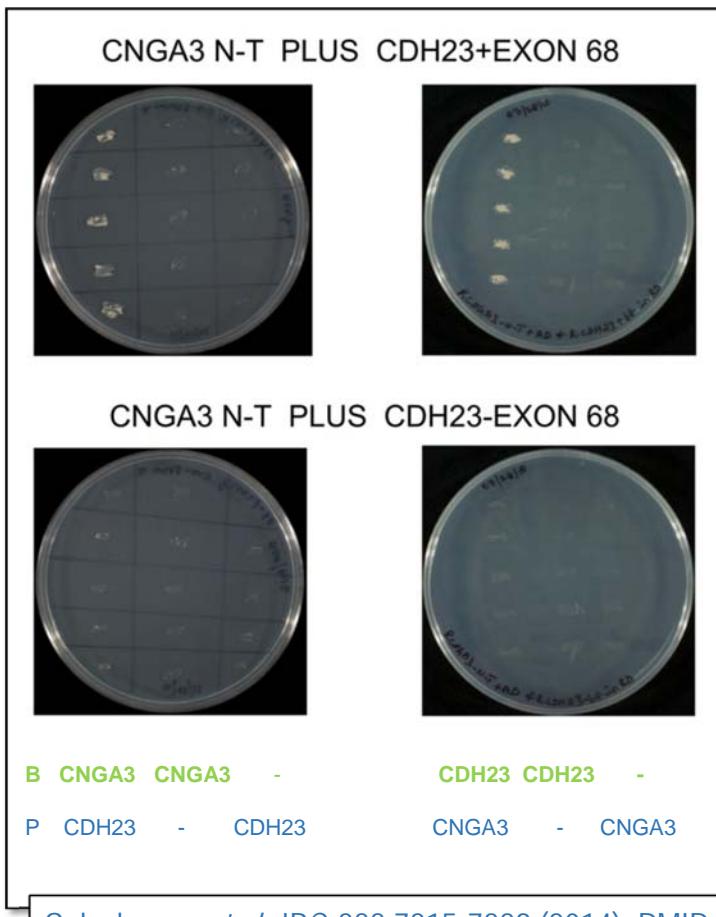


Fields & Song, Nature 340:245-246 (1989); PMID: 2547163

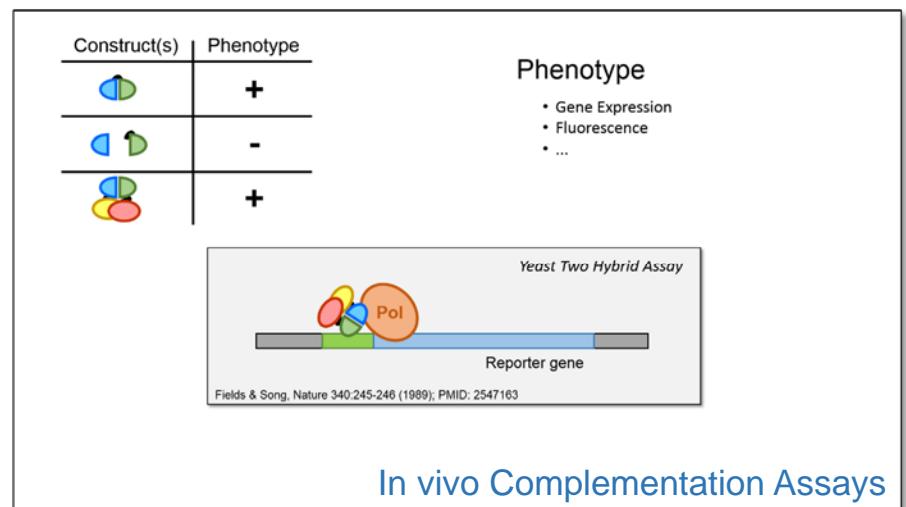
In vivo Complementation Assays

Interaction Experiments: Methods

Complementation Assays

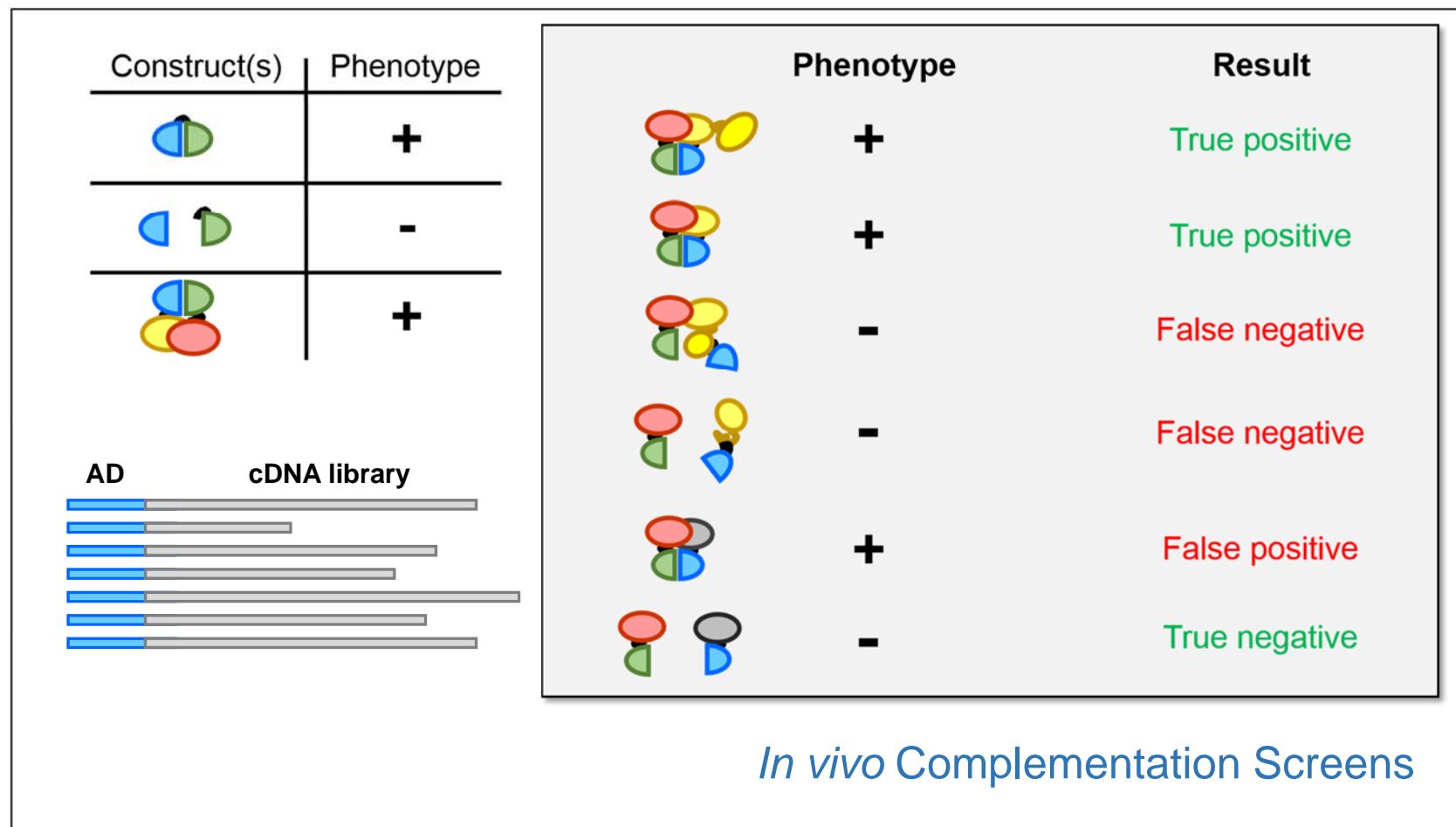


- Overexpression (yeast) of rat CNGA3 (fragment) & CDH23 (+/- exon 68) proteins fused, respectively to DNA and activation domains of Gal4 (transcription factor)
- Yeast strain (Y187) with Ura3 marker under control of Gal4 promoter
- Selection on minimal/-uracil medium

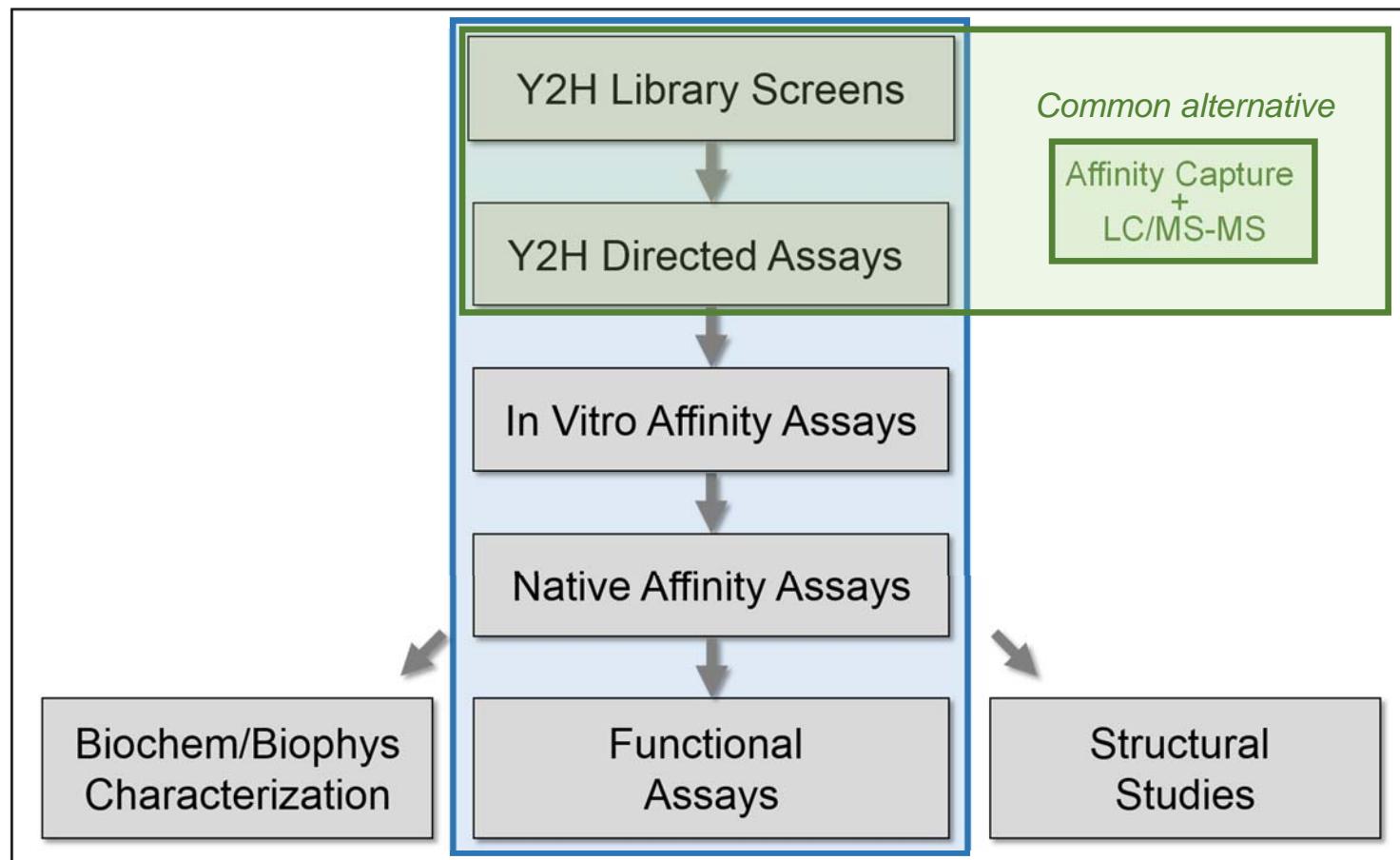


Interaction Experiments

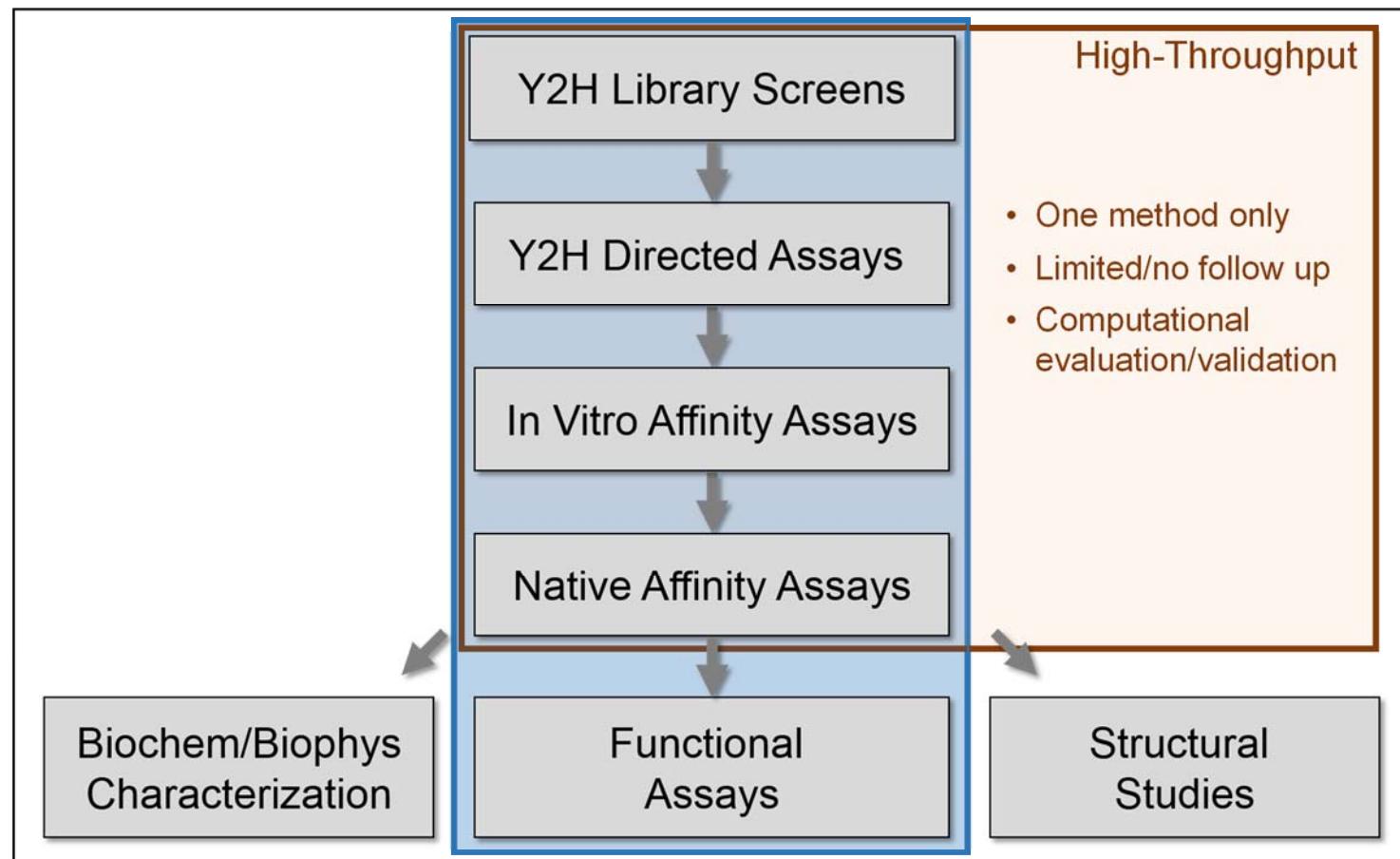
Complementation Assays



Interaction Experiments: Typical Workflows



Interaction Experiments: Typical Workflows

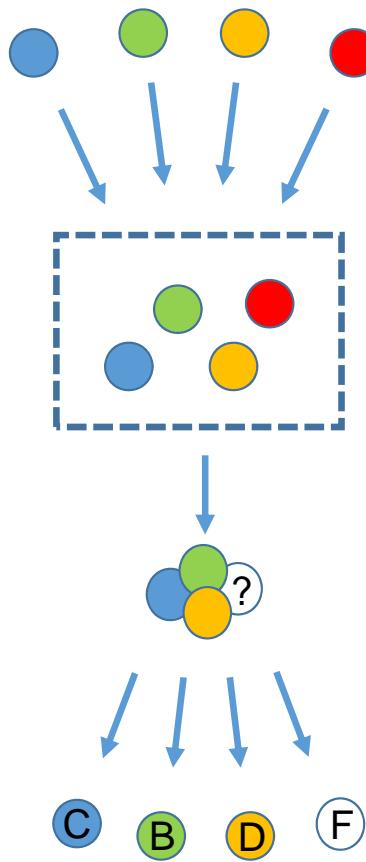


Interaction Experiments

- Different experimental approaches may provide different (and often complementary) information about molecular interactions
- Many experimental approaches provide only partial information about molecular interactions that may result in experimental ambiguities of protein identities and interaction type

Interaction Database Records

Interaction Experiment Flow



- **Make proteins**
 - Where: native vs heterologous host vs in vitro translation vs chemical synthesis
 - How much: native level vs overexpressed
 - Modifications: isoforms, fragments, mutations, PTMs present/absent
- **Get them together**
 - Where: native organism/cell type/tissue/compartment vs something else
 - When: cell cycle phase/cell state
- **Test which ones interact**
 - Diverse methods can be used to determine that that proteins interact
 - Information that can be inferred from each experiment depends on the method and experimental setup
- **Identify proteins that interact**
 - Identity of some proteins might be known a priori (eg purified, cloned/ tagged bait, etc)
 - Identity and/or state of some proteins might be ambiguous (eg unknown splice form, PTMs)
 - Some molecules participating in the interaction might remain unidentified

Interaction Experiment Record

Interaction Record

Interaction Experiment

Interaction Detection Method: Pull down (MI:0096)
Experiment Host/Cell Line: in vitro
Interaction Type: Direct (MI:0407)

Participant List

Molecule Type: Protein (MI:0326)
Molecule Name: Cellular tumor antigen p53
Molecule Symbol: p53
Species of Origin: Human (Taxid:9606)
Cross-reference(s): P04637-1 (UniProtKB), NP_00537, (RefSeq), ...

Experimental Role: Bait (MI:0496)
Experimental Source: E.coli K12 (Taxid: 83333) 
Identification method(s): Predetermined (MI:0396)

Features:

Feature Type: Sufficient binding region (MI:0442)
Feature Range: 1-73
Identification Method: Deletion analysis (MI:0033)

1 2 ... N

1 2 ... M

Make Proteins

Get the Proteins Together

Test Which Proteins Interact

Identify/Re-identify Proteins
That Interacted

Interaction Experiment Record

Interaction Record

Interaction Experiment

Interaction Detection Method: Pull down (MI:0096)
Experiment Host/Cell Line: in vitro
Interaction Type: Direct (MI:0407)



Participant List

Molecule Type: Protein (MI:0326)
Molecule Name: Cellular tumor antigen p53
Molecule Symbol: p53
Species of Origin: Human (Taxid:9606)
Cross-reference(s): P04637-1 (UniProtKB), NP_00537, (RefSeq), ...

Experimental Role: Bait (MI:0496)
Experimental Source: E.coli K12 (Taxid: 83333)
Identification method(s): Predetermined (MI:0396)

Features:

Feature Type: Sufficient binding region (MI:0442)
Feature Range: 1-73
Identification Method: Deletion analysis (MI:0033)

1 2 ... N

1 2 ... M

Make Proteins

Get the Proteins Together

Test Which Proteins Interact

Identify/Re-identify Proteins
That Interacted

Interaction Experiment Record

Interaction Record

Interaction Experiment

Interaction Detection Method: Pull down (MI:0096)
Experiment Host/Cell Line: *in vitro*
Interaction Type: Direct (MI:0407)



Participant List

Molecule Type: Protein (MI:0326)
Molecule Name: Cellular tumor antigen p53
Molecule Symbol: p53
Species of Origin: Human (Taxid:9606)
Cross-reference(s): P04637-1 (UniProtKB), NP_00537, (RefSeq), ...

Experimental Role: Bait (MI:0496)
Experimental Source: E.coli K12 (Taxid: 83333)
Identification method(s): Predetermined (MI:0396)

Features:

Feature Type: Sufficient binding region (MI:0442)
Feature Range: 1-73
Identification Method: Deletion analysis (MI:0033)

1 2 ... N

1 2 ... M

Protein Expression	Interaction Host
	<i>S. cerevisiae</i>
	<i>E. coli</i>
	<i>in vitro</i>
	<i>in vitro</i>

Interaction Experiment Record

Interaction Record

Interaction Experiment

Interaction Detection Method: Pull down (MI:0096) ←
Experiment Host/Cell Line: in vitro
Interaction Type: Direct (MI:0407)

Participant List

Molecule Type: Protein (MI:0326)
Molecule Name: Cellular tumor antigen p53
Molecule Symbol: p53
Species of Origin: Human (Taxid:9606)
Cross-reference(s): P04637-1 (UniProtKB), NP_00537, (RefSeq), ...

Experimental Role: Bait (MI:0496)
Experimental Source: E.coli K12 (Taxid: 83333)
Identification method(s): Predetermined (MI:0396)

Features:

Feature Type: Sufficient binding region (MI:0442)
Feature Range: 1-73
Identification Method: Deletion analysis (MI:0033)

1 2 ... N

1 2 ... M

Make Proteins

Get the Proteins Together

Test Which Proteins Interact

Identify/Re-identify Proteins
That Interacted

Interaction Experiment Record

Interaction Record

Interaction Experiment

Interaction Detection Method: Pull down (MI:0096)
Experiment Host/Cell Line: in vitro
Interaction Type: Direct (MI:0407)



Participant List

Molecule Type: Protein (MI:0326)
Molecule Name: Cellular tumor antigen p53
Molecule Symbol: p53
Species of Origin: Human (Taxid:9606)
Cross-reference(s): P04637-1 (UniProtKB), NP_00537, (RefSeq), ...

Experimental Role: Bait (MI:0496)
Experimental Source: E.coli K12 (Taxid: 83333)
Identification method(s): Predetermined (MI:0396)

Features:

Feature Type: Sufficient binding region (MI:0442)
Feature Range: 1-73
Identification Method: Deletion analysis (MI:0033)

1 2 ... N

1 2 ... M

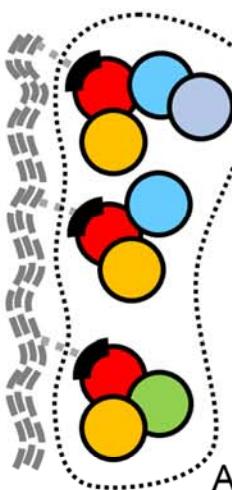
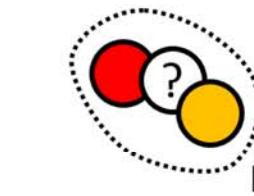
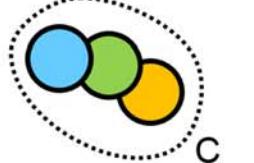
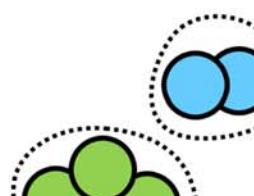
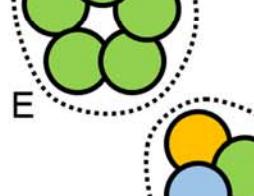
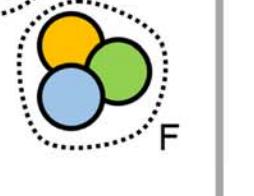
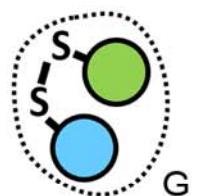
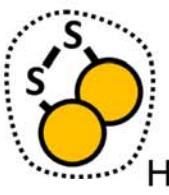
Make Proteins

Get the Proteins Together

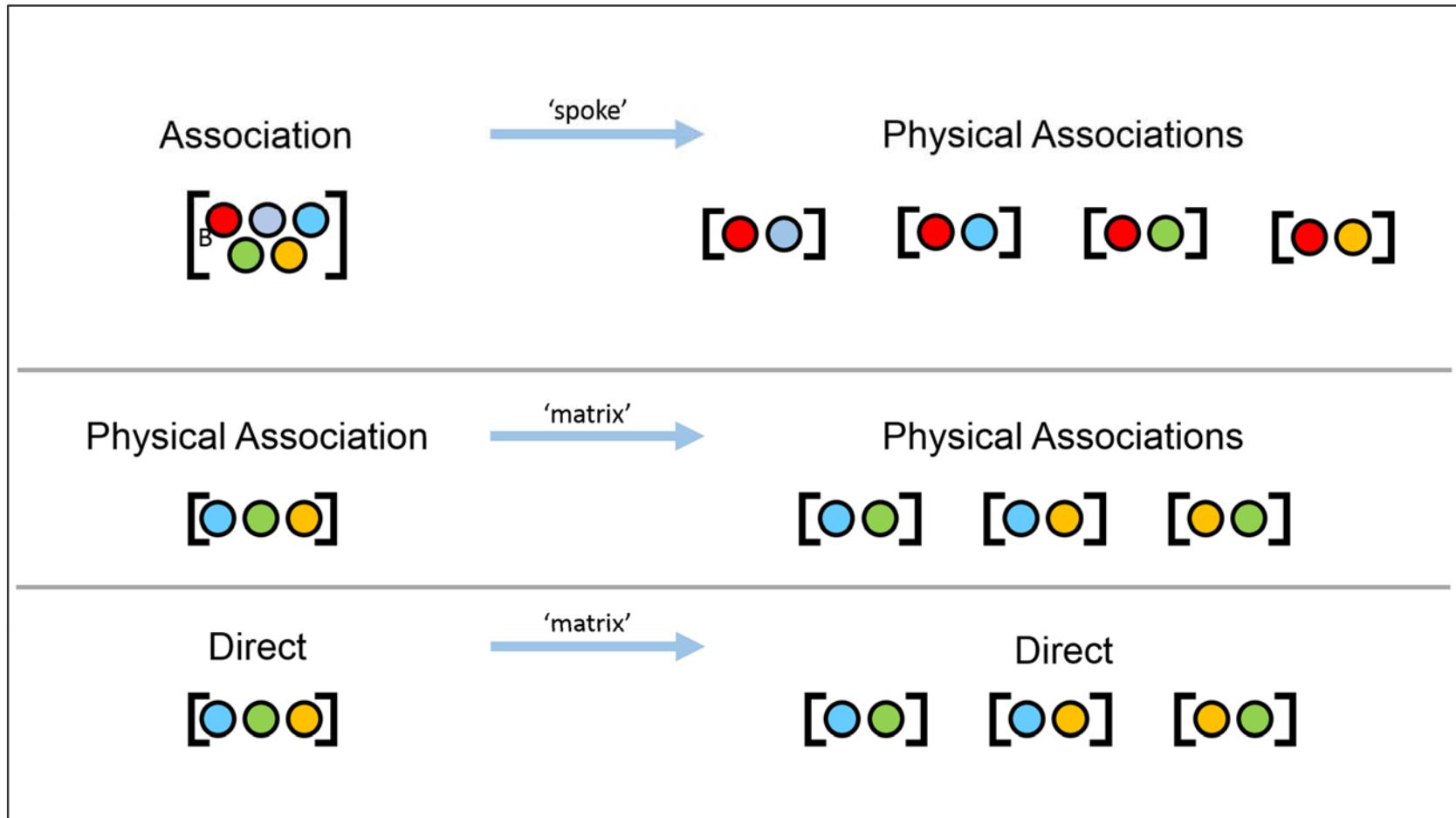
Test Which Proteins Interact

Identify/Re-identify Proteins
That Interacted

Interaction Types

	 $[\text{ } \text{ } \text{ }]_B$	 $[\text{ } \text{ } \text{ }]_C$	 $[\text{ } \text{ }]_D$	 $[\text{ } \text{ } \text{ }]_E$	 $[\text{ } \text{ } \text{ }]_F$	 $[\text{ } \text{ }]_G$	 $[\text{ } \text{ }]_H$
Interaction Type	Association	Physical Association	Direct	Covalent			

Binary Expansion



Interaction Experiment Record

Interaction Record

Interaction Experiment

Interaction Detection Method: Pull down (MI:0096)
Experiment Host/Cell Line: in vitro
Interaction Type: Direct (MI:0407)



Participant List

Molecule Type: Protein (MI:0326)
Molecule Name: Cellular tumor antigen p53
Molecule Symbol: p53
Species of Origin: Human (Taxid:9606)
Cross-reference(s): P04637-1 (UniProtKB), NP_00537, (RefSeq), ...

Experimental Role: Bait (MI:0496)
Experimental Source: E.coli K12 (Taxid: 83333)
Identification method(s): Predetermined (MI:0396)

Features:

Feature Type: Sufficient binding region (MI:0442)
Feature Range: 1-73
Identification Method: Deletion analysis (MI:0033)

1 2 ... N

1 2 ... M

Make Proteins

Get the Proteins Together

Test Which Proteins Interact

Identify/Re-identify Proteins
That Interacted

Interaction Experiment Record

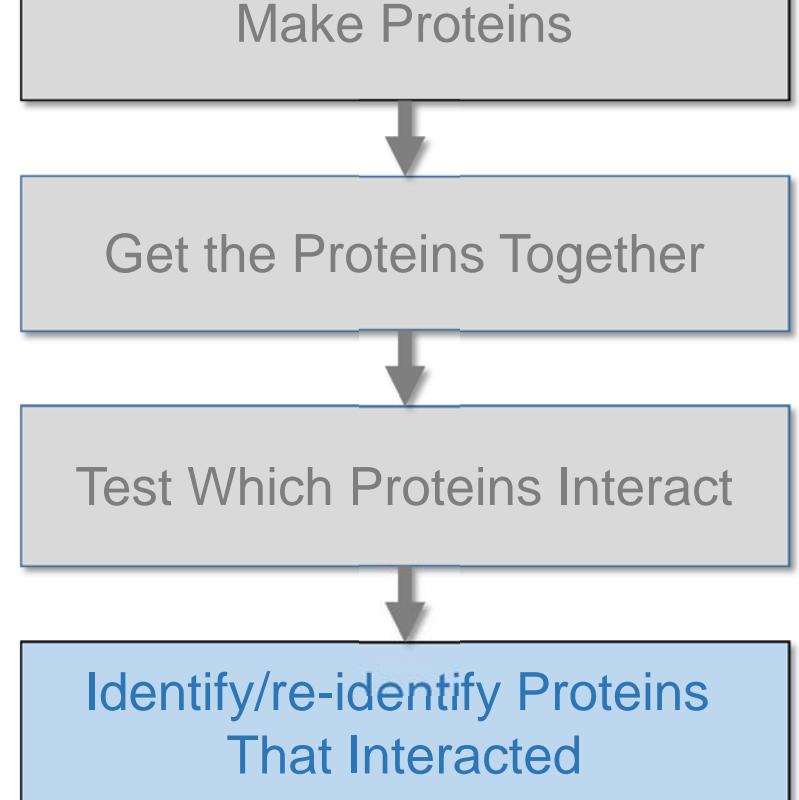
Interaction Record

Interaction Experiment

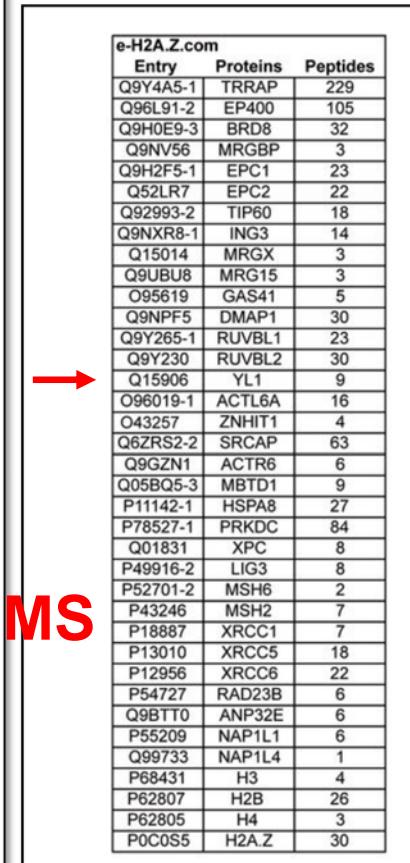
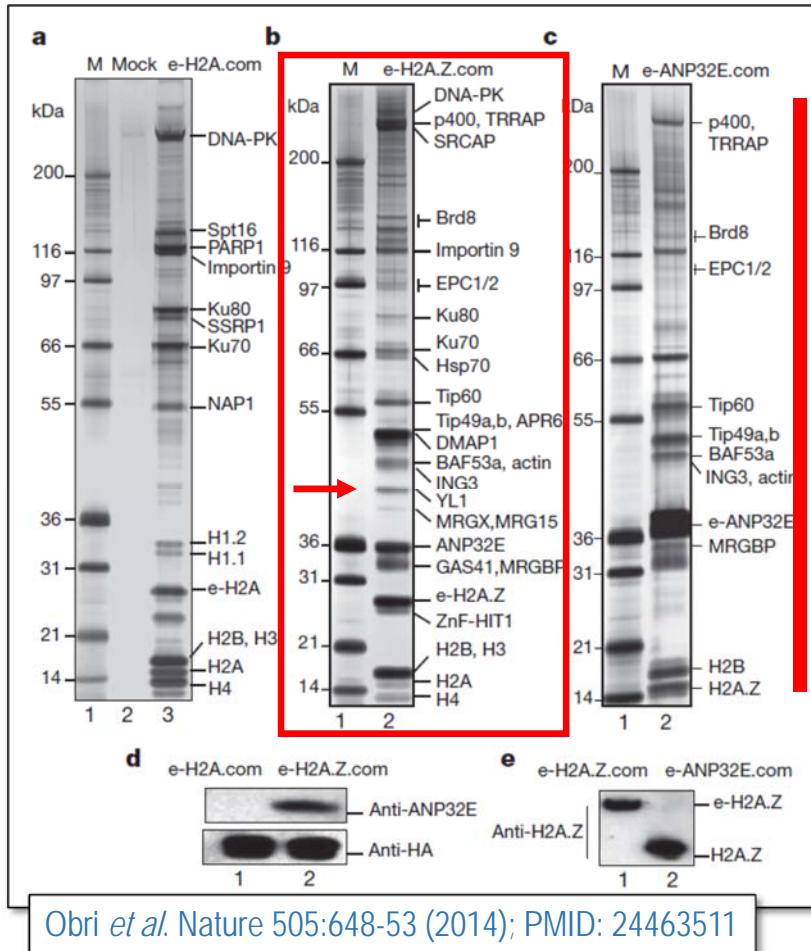
Interaction Detection Method: Pull down (MI:0096)
Experiment Host/Cell Line: in vitro
Interaction Type: Direct (MI:0407)

Participant List

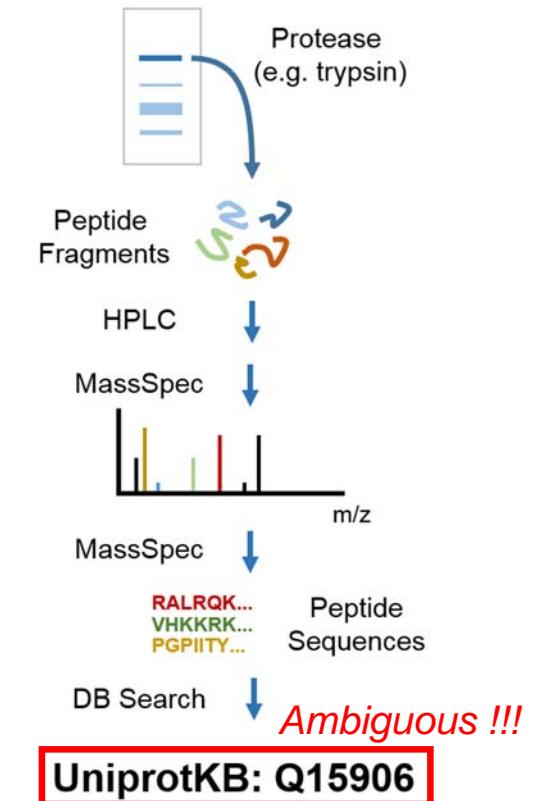
Molecule Type: Protein (MI:0326)
Molecule Name: Cellular tumor antigen p53
Molecule Symbol: p53
Species of Origin: Human (Taxid:9606)
Cross-reference(s): P04637-1 (UniProtKB), NP_00537,
(RefSeq), ...
Experimental Role: Bait (MI:0496)
Experimental Source: E.coli K12 (Taxid: 83333)
Identification method(s): Predetermined (MI:0396)
Features:
Feature Type: Sufficient binding region (MI:0442)
Feature Range: 1-73
Identification Method: Deletion analysis (MI:0033)



Protein Identification



MassSpec Protein Identification Workflow



UniprotKB: Q15906

Isoform 1 (identifier: [Q15906-1](#)) [UniParc] [FASTA](#) [Add to basket](#)

This isoform has been chosen as the canonicalⁱ sequence. All positional information in this entry refers to it. This is also the sequence that appears in the downloadable versions of the entry.

« Hide

10	20	30	40	50
MSLAGGRAPR	KTAGNRLSGL	LEAEEEDEFY	QTYGGFTEE	SGDDEYQQDQ
60	70	80	90	100
SDTEDEVDSDFDIDEDEPS	SDGEAEEPRR	KRRVVTKAYK	EPLKSLPRK	
110	120	130	140	150
VNTPAGSSQK	AREEKALLPL	ELQDDGSDSR	KSMRQSTAEH	TRQTFLRVQE
160	170	180	190	200
RQGQSRRKG	PHCERPLTQE	ELLREAKITE	EINLRSLETY	ERLEADKKQ
210	220	230	240	250
VHKKRKCPGP	IITYHSVTVP	LVEPEPGPKEE	NVDIEGLDPA	PSVSAALTPHA
260	270	280	290	300
GTGPVNPPAR	CSRPFITFSD	DATFEEWFPQ	GRPPKVVPRE	VCPVTHRPAL
310	320	330	340	350
YRDPTDIPY	ATARAFKII	EAYKKYITAH	GLPPTASALG	PGPPPPEPLP
360				
GSGPRALRQK	IVIK			

Isoform 2 (identifier: [Q15906-2](#)) [UniParc] [FASTA](#) [Add to basket](#)

The sequence of this isoform differs from the canonical sequence as follows:

236-236: G → GSLCFSLSFVLR

Show »

Computationally mapped potential isoform sequencesⁱ

There is 1 potential isoform mapped to this entry. [BLAST](#) [Align](#) [Show all](#) [Add to basket](#)

<input type="checkbox"/>	Entry	Entry name	Protein names	<input type="checkbox"/>	Gene names	Length	Annotation
<input type="checkbox"/>	A0A1W2PPT2	A0A1W2PPT2_HUMAN	Vacuolar protein sorting-associated...	<input type="checkbox"/>	VPS72	199	Annotation score: ●○○○○

Experimental Ambiguities

- In many cases multiple isoforms exist
- It is hard to unambiguously demonstrate which one is present/absent
- The problem is not unique to MS – similar issues arise when identifying endogenous proteins by Western blots

UniprotKB: Q15906

Isoform 1 (identifier: Q15906-1) [UniParc] [FASTA](#) [Add to basket](#)

This isoform has been chosen as the canonicalⁱ sequence. All positional information in this entry refers to it. This is also the sequence that appears in the downloadable versions of the entry.

« Hide

10	20	30	40	50
MSLAGGRAPR	KTAGNRLSGL	LEAEEEDEFY	QTYGGFTEE	SGDDEYQQGDQ
60	70	80	90	100
SDTEDEVDS	FDI	DEGDEPS	SDGEAEEPRR	KRRVVTKAYK
110	120	130	140	150
VNT	PAGSSQK	AEEKALLPL	ELQDDGSDSR	KSMRQSTAEH
160	170	180	190	200
RQGQSRRKG	PHCERPLTQE	ELLREAKITE	EINLRSLETY	ERLEADKKQ
210	220	230	240	250
VHKKRKCPGP	IITYHSVTVP	LVEGPGPKEE	NVDIEGLDPA	PSVSAITPA
260	270	280	290	300
GTGPVNPPAR	CSRTFITFS	DATFEEWFPQ	GRPPKVVPRE	VCPVTHRPAL
310	320	330	340	350
YRDPTDIPY	ATARAFKII	EAYKKYITAH	GLPPTASALG	PGPPPPEPLP
360				
GSGPRA	LQRK	IVIK		

Length: 364
Mass (Da): 40,594
Last modified: November 1, 1996 - v1
Checksum: i OAE890B62B2BCA4A

BLAST [GO](#)

Isoform 2 (identifier: Q15906-2) [UniParc] [FASTA](#) [Add to basket](#)

The sequence of this isoform differs from the canonical sequence as follows:

236-236: G → GSLCFSLSFVLR

Show »

Computationally mapped potential isoform sequencesⁱ

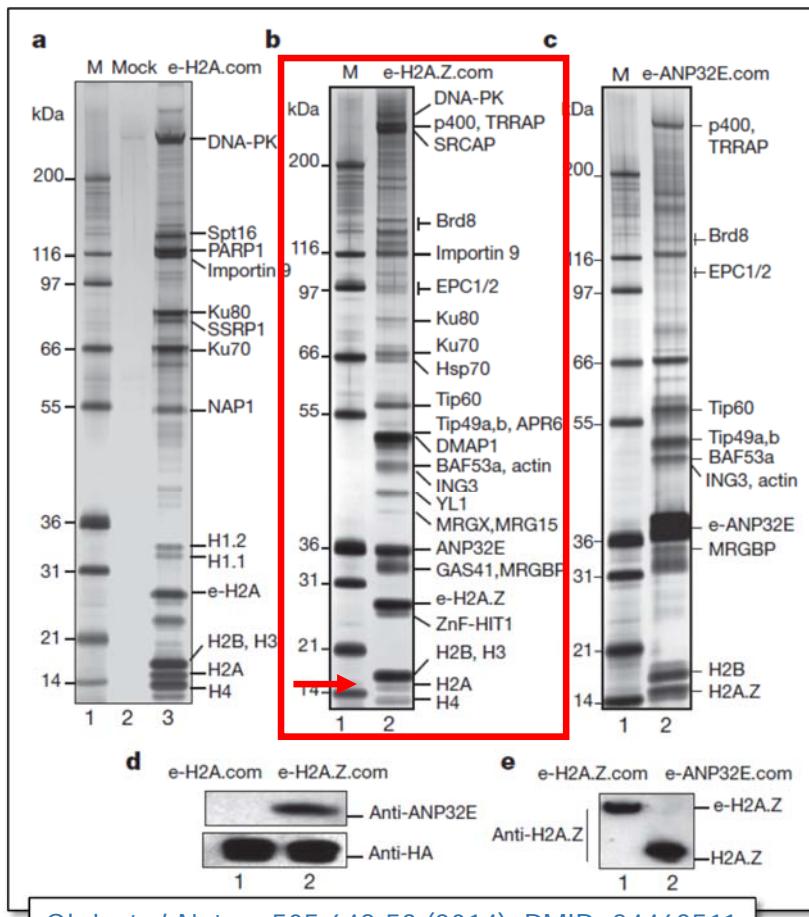
There is 1 potential isoform mapped to this entry. [BLAST](#) [Align](#) [Show all](#) [Add to basket](#)

Entry	Entry name	Protein names	Gene names	Length	Annotation
A0A1W2PPT2	A0A1W2PPT2_HUMAN	Vacuolar protein sorting-associated...	VPS72	199	Annotation score: 00000

Annotation Ambiguities

- In some cases there is no single identifier that covers all isoforms
- In some cases there are multiple identifiers that cover the same protein
- Database records change over time

Protein Identification



UniprotKB: P0C0S8

Cross-referencesⁱ

Sequence databases

Select the link destinations:	Z83742 Genomic DNA Translation: CAB06037.1 Z83739 Genomic DNA Translation: CAB06034.1 X83549 Genomic DNA Translation: CAA58539.1 X57138 Genomic DNA Translation: CAA40417.1 L19778 mRNA Translation: AAC24466.1 AY131987 Genomic DNA Translation: AAN59968.1 AY131989 Genomic DNA Translation: AAN59970.1 AY131991 Genomic DNA Translation: AAN59972.1 AY131992 Genomic DNA Translation: AAN59973.1 AY131993 Genomic DNA Translation: AAN59974.1 Z98744 Genomic DNA No translation available. AL009179 Genomic DNA No translation available. AL021807 Genomic DNA No translation available. BC016677 mRNA Translation: AAH16677.1 BC069306 mRNA Translation: AAH69306.1 BC104199 mRNA Translation: AAI04200.1 BC104198 mRNA Translation: AAI04199.1 BC105129 mRNA Translation: AAI05130.1 BC112072 mRNA Translation: AAI12073.1 BC112254 mRNA Translation: AAI12255.1 BC112256 mRNA Translation: AAI12257.1
CCDS ⁱ	CCDS4619.1 CCDS4626.1 CCDS4632.1 CCDS4634.1 CCDS4639.1

PIR ⁱ	BS6624 HSHUA1
RefSeq ⁱ	NP_003500.1, NM_003509.2 NP_003501.1, NM_003510.2 NP_003502.1, NM_003511.2 NP_003505.1, NM_003514.2 NP_066408.1, NM_021064.4

UniGene ⁱ	Hs.134999 Hs.233568 Hs.51011 Hs.534035 Hs.734717
----------------------	--

Gene duplication !!!

UniProtKB: Identifier Mapping

The screenshot shows the UniProtKB Identifier Mapping page. At the top, there is a navigation bar with links for BLAST, Align, Retrieve/ID mapping, Peptide search, Help, Advanced search, and a search bar. Below the navigation bar, the title "Programmatic access - Mapping database identifiers" is displayed. A note explains that users can use the "Retrieve/ID mapping" service programmatically by knowing database abbreviations. It provides examples for Perl, Python, Ruby, and Java. An alternative download link is provided for the data underlying the service. Below this, a "See also:" section links to the REST API documentation. A "Related terms" section lists programmatic access, program, script, wget, curl, web services, API, and uploadlists. The main content is a table titled "Category: UniProt" which maps database abbreviations to their names and directions. A second table titled "Category: Sequence databases" follows.

Name	Abbreviation	Direction
Category: UniProt		
UniProtKB AC/ID	ACC+ID	from
UniProtKB AC	ACC	both
UniProtKB ID	ID	both
UniParc	UPARC	both
UniRef50	NF50	both
UniRef90	NF90	both
UniRef100	NF100	both
Gene name	GENENAME	both
CRC64	CRC64	both
Category: Sequence databases		
EMBL/GenBank/DDBJ	EMBL_ID	both
EMBL/GenBank/DDBJ CDS	EMBL	both

URL: <https://www.uniprot.org/uploadlists/>

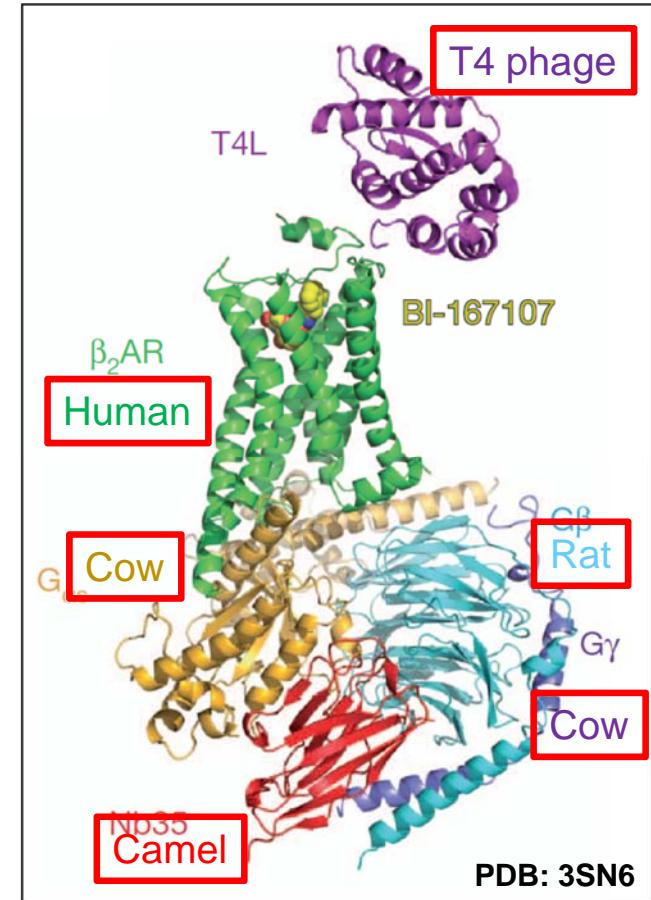
- Hosted by SIB (Geneva, Switzerland)
- Cross-references between UniProtKB records and about 80 databases (Refseq, Ensembl, model organisms, PDB, ...)
- Stable 'over time' and over 'interface'
- Access Results presented as 2 column table
- in a batch mode and as a service (RESTful)

When are two proteins ‘identical’ ?

Possible criteria:

- Identical sequence
- Identical species (strain ?) of origin
- Identical PTMs
- Identical localization (space & time)
- Identical gene
- Same orthologous group

Definition which proteins are treated as identical is project specific and likely different than the definition used by at least some authors of the interaction data.



Protein Identification: Special cases

Somatic recombination

Not curated

- V(D)J recombination
 - B-cell receptors/antibodies
 - T-cell receptors
- class switching
 - antibodies

Gene Duplication

Curated if:

- Histones
 - Multiple gene families, many genes/family
 - <https://www.ncbi.nlm.nih.gov/research/HistoneDB2.0> (PMID:26989147)

Polyproteins

- Viral proteins
- Ubiquitin

Protein Identification: Polyproteins

UniprotKB: P04585

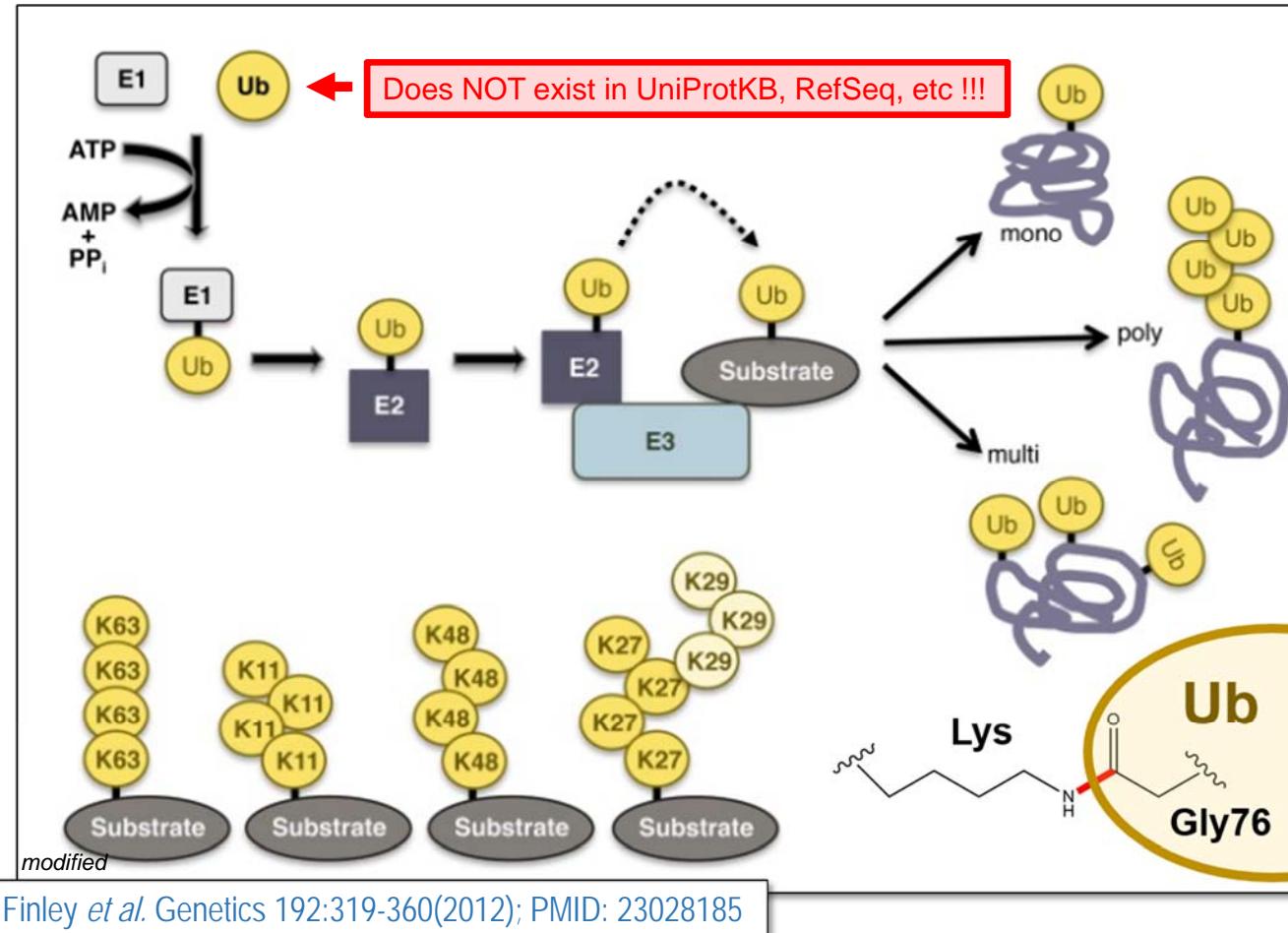
PTM / Processingⁱ

Molecule processing

Feature key	Position(s)	Description	Actions	Graphical view	Length
Initiator methionine ⁱ		Removed; by host By similarity			
Chain ⁱ (PRO_0000223620)	2 – 1435	Gag-Pol polyprotein			1434
Chain ⁱ (PRO_0000042439)	2 – 132	Matrix protein p17 By similarity			131
Chain ⁱ (PRO_0000042440)	133 – 363	Capsid protein p24 By similarity			
Peptide ⁱ (PRO_0000042441)	364 – 377	Spacer peptide 1 By similarity			
Chain ⁱ (PRO_0000042442)	378 – 432	Nucleocapsid protein p7 By similarity			55
Peptide ⁱ (PRO_0000246716)	433 – 440	Transframe peptide Sequence analysis			8
Chain ⁱ (PRO_0000042443)	441 – 488	p6-pol Sequence analysis			48
Chain ⁱ (PRO_0000038665)	489 – 587	Protease			99
Chain ⁱ (PRO_0000042444)	588 – 1147	Reverse transcriptase/ribonuclease H			560
Chain ⁱ (PRO_0000042445)	588 – 1027	p51 RT			440
Chain ⁱ (PRO_0000042446)	1028 – 1147	p15			120
Chain ⁱ (PRO_0000042447)	1148 – 1435	Integrase By similarity			288

P04585-PRO_0000042439

Protein Identification: Ubiquitin



Ubiquitin is synthesized as a part of many polyproteins

- Polyubiquitin
 - Human: UBB, UBC
 - Yeast: UBI4
- Ribosomal protein fusions
 - Human: UBA52 (L40), RPS27A (S27a)
 - Yeast: UBI1 (L40), UBI2 (L40), UBI3 (S37)

Curated as deprecated
UniprotKB protein
or
as PTM

Interaction Experiment Record

Interaction Record

Interaction Experiment

Interaction Detection Method: Pull down (MI:0096)
Experiment Host/Cell Line: in vitro
Interaction Type: Direct (MI:0407)

Participant List

Molecule Type: Protein (MI:0326)
Molecule Name: Cellular tumor antigen p53
Molecule Symbol: p53
Species of Origin: Human (Taxid:9606)
Cross-reference(s): P04637-1 (UniProtKB), NP_00537, (RefSeq), ...

Experimental Role: Bait (MI:0496)
Experimental Source: E.coli K12 (Taxid: 83333)
Identification method(s): Predetermined (MI:0396)
Features:

Feature Type: Sufficient binding region (MI:0442)
Feature Range: 1-73
Identification Method: Deletion analysis (MI:0033)

1
2
...
N

1
2
...
M

Make Proteins

Get the Proteins Together

Test Which Proteins Interact

Identify/re-identify Proteins
That Interacted

Interaction Experiment Record

Interaction Record

Interaction Experiment

Interaction Detection Method: Pull down (MI:0096)
Experiment Host/Cell Line: in vitro
Interaction Type: Direct (MI:0407)

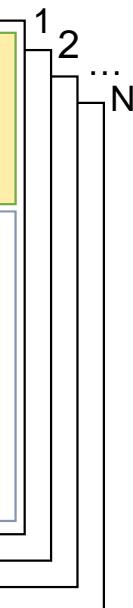
Participant List

Molecule Type: Protein (MI:0326)
Molecule Name: Cellular tumor antigen p53
Molecule Symbol: p53
Species of Origin: Human (Taxid:9606)
Cross-reference(s): P04637-1 (UniProtKB), NP_00537, (RefSeq), ...

Experimental Role: Bait (MI:0496)
Experimental Source: E.coli K12 (Taxid: 83333)
Identification method(s): Predetermined (MI:0396)

Features:

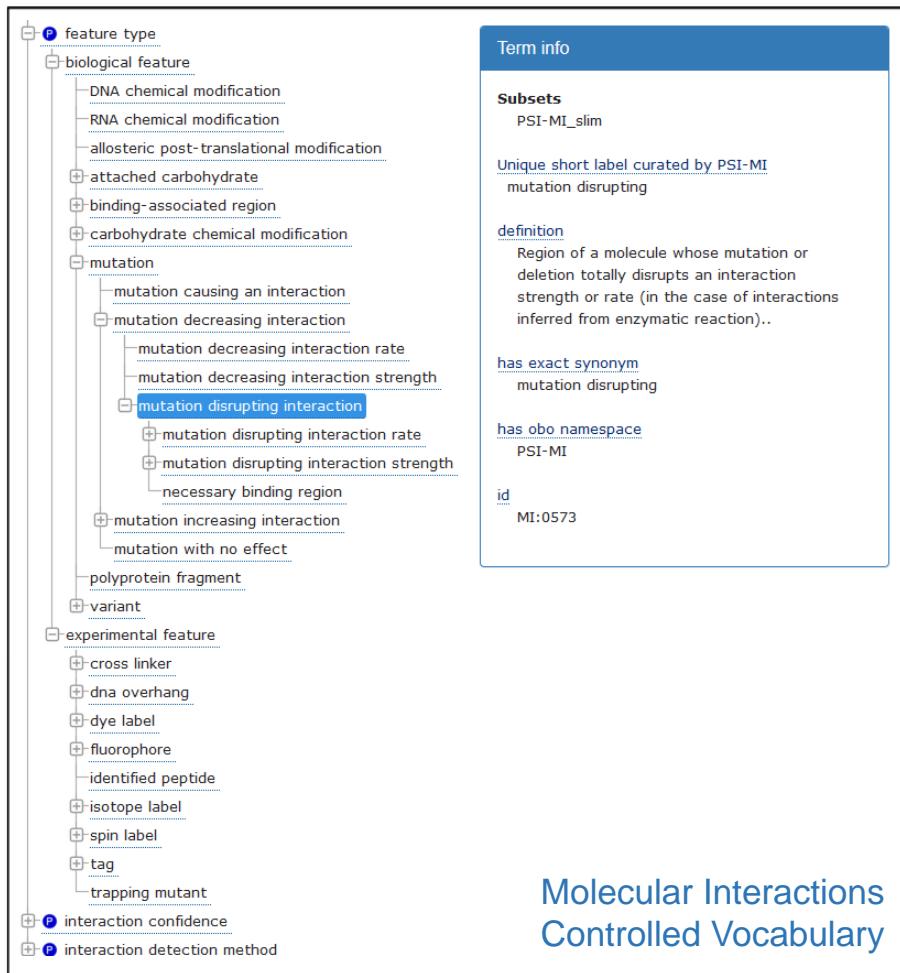
Feature Type: Sufficient binding region (MI:0442)
Feature Range: 1-73
Identification Method: Deletion analysis (MI:0033)



Protein Features

- Deviations from the default sequence
 - Mutations (substitutions, deletions, insertions)
 - Regions necessary/sufficient for binding
- Post-translational modifications (PTMs)
 - Phosphorylation, methylation, etc
 - **Ubiquitination sites**
- Experimental artifacts
 - Affinity tags (His, FLAG, TAP, GFP, ...)
 - Labels (biotin, isotope, fluorescent, spin labels,...)
 - Cross-linking sites
 - Proteolytic sites

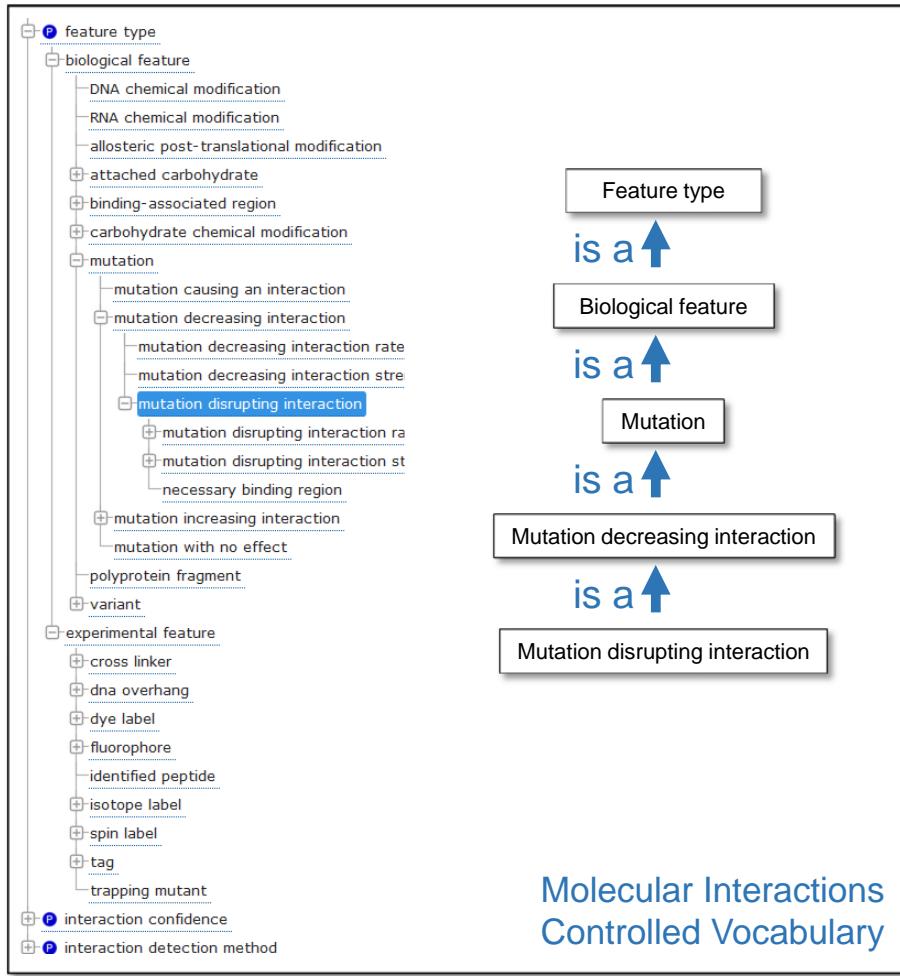
Controlled Vocabularies



Molecular Interactions (MI) Controlled Vocabulary

- Collection of terms, each with:
 - Unique identifier
 - Definition
 - Common aliases
- Include most of the terms used to annotate molecular interactions
- Provides relationships between terms
 - Most often of 'is a' type

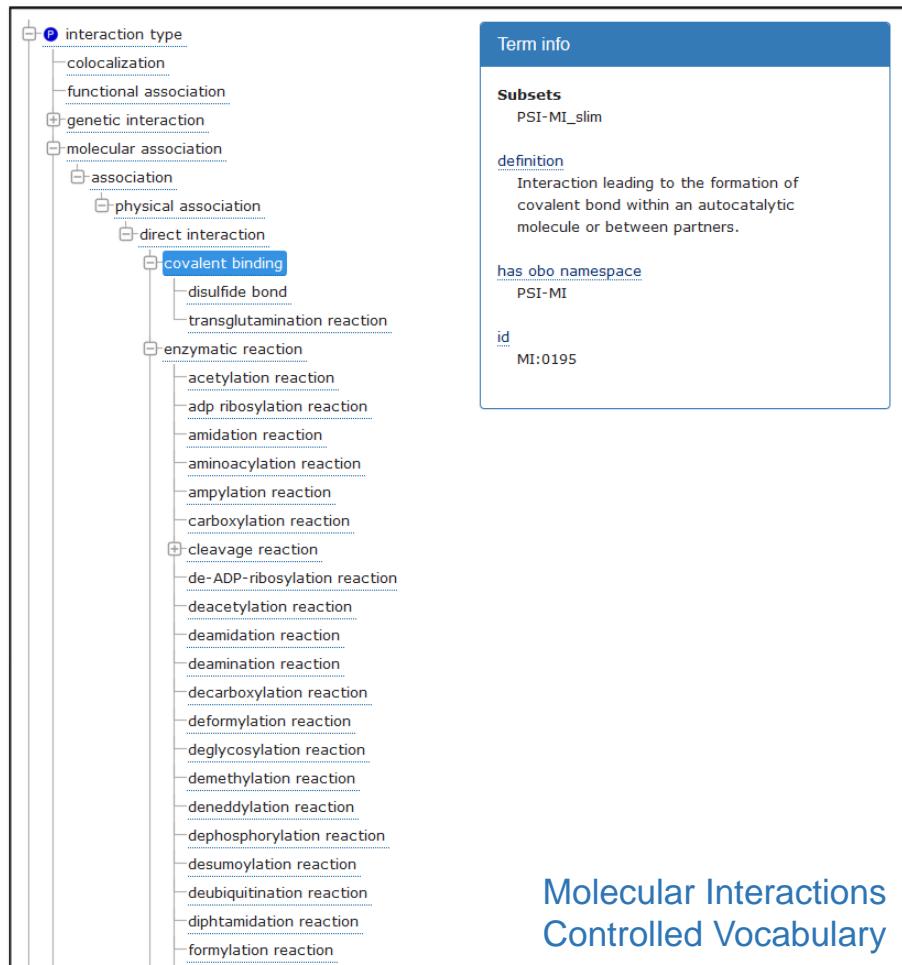
Controlled Vocabularies



Molecular Interactions (MI) Controlled Vocabulary

- Collection of terms, each with:
 - Unique identifier
 - Definition
 - Common aliases
- Include most of the terms used to annotate molecular interactions
- Provides relationships between terms
 - Most often of 'is a' type
 - So, formally, MI is 'ontology'

Controlled Vocabularies



Interaction Types

Association



Physical Association



Direct Interaction

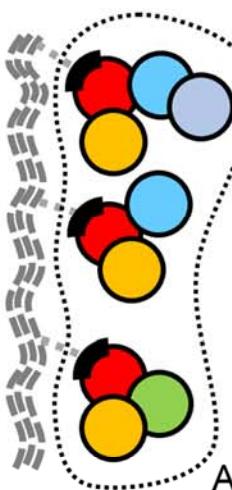
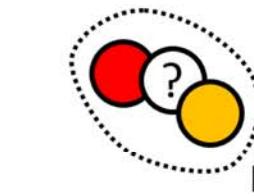
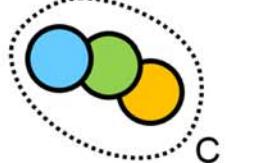
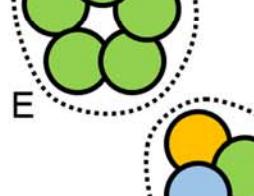
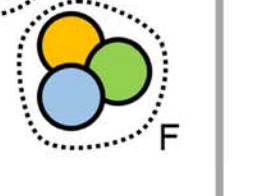
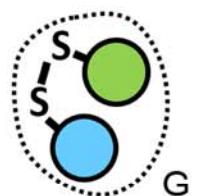


Enzymatic
Reaction



Covalent
Binding

Interaction Types

	 $[\text{ } \text{ }]_B$	 $[\text{ } \text{ } \text{ }]_C$	 $[\text{ } \text{ }]_D$	 $[\text{ } \text{ } \text{ }]_E$	 $[\text{ } \text{ } \text{ }]_F$	 $[\text{ } \text{ }]_G$	 $[\text{ } \text{ }]_H$
Interaction Type	Association	Physical Association	Direct	Covalent			

Controlled Vocabularies

The screenshot shows a controlled vocabulary interface for Molecular Interactions. On the left, a tree view displays various biological features and experimental features. A specific term, "mutation disrupting interaction", is selected and expanded. To the right, a detailed "Term info" panel provides the following information:

- Subsets:** PSI-MI_slim
- Unique short label curated by PSI-MI:** mutation disrupting
- definition:** Region of a molecule whose mutation or deletion totally disrupts an interaction strength or rate (in the case of interactions inferred from enzymatic reaction)..
- has exact synonym:** mutation disrupting
- has obo namespace:** PSI-MI
- id:** MI:0573

Molecular Interactions Controlled Vocabulary

Molecular Interactions (MI) Controlled Vocabulary

URL: <https://www.ebi.ac.uk/ols/ontologies/mi>

The screenshot shows the EMBL-EBI Ontology Lookup Service (OLS) homepage. The header includes the OLS logo, navigation links for Home, Ontologies, Documentation, and About, and a Contact Us link. The main content area features a welcome message, a search bar, and a sidebar with data statistics and tweets.

Welcome to the EMBL-EBI Ontology Lookup Service.

Search OLS... Examples: diabetes, GO:0098743

Looking for a particular ontology?

Data Content
Updated 06 Mar 2019 11:11
o 224 ontologies
o 5,415,430 terms
o 21,348 properties
o 480,657 individuals

About OLS
The Ontology Lookup Service (OLS) is a repository for biomedical ontologies that aims to provide a single point of access to the latest ontology versions. You can browse the ontologies through the website as well as programmatically via the OLS API. OLS is developed and maintained by the Samples, Phenotypes and Ontologies Team (SPOT) at EMBL-EBI.

Related Tools
In addition to OLS the SPOT team also provides the OxO, Zooma and Webulous services. OxO provides cross-ontology mappings between terms from different ontologies. Zooma is a service to assist in mapping data to ontologies in OLS and Webulous is a tool for building ontologies from spreadsheets.

Contact Us
For feedback, enquiries or suggestion about OLS or to request a new ontology please contact ols-support @ ebi.ac.uk. For bugs or problems with the code or API please report on GitHub issue For announcements relating to OLS, such as new releases and new features sign up to the OLS announce mailing list

Tweets by @EBIOLS

EBISPORT OLS @EBIOLS
Replying to @EBIOLS
SERVICE UPDATE: OLS is back to normal
Dec 6, 2018

EBISPORT OLS @EBIOLS
SERVICE UPDATE: OLS is having a few issues at the

Controlled Vocabularies

The screenshot shows a controlled vocabulary interface for protein modifications. On the left, a tree view lists categories like 'protein modification' and 'modified L-alanine residue'. On the right, a detailed 'Term info' panel for 'modified L-histidine residue' is displayed, containing sections for 'database cross reference', 'Subsets', 'Alternate name from RESID', 'Description (full name) from UniMod', 'Systematic name from RESID', 'Short label curated by PSI-MOD', 'Protein feature description from UniProtKB', and 'Alternate name from RESID'. A large blue banner at the bottom reads 'Protein Modifications Controlled Vocabulary'.

Term info

database cross reference

- o Formula: (C 33 H 38 N 12 O 16 P 2)
- o DiffAvg: (783.54)
- o MassAvg: (920.68)
- o Origin: (H)
- o MassMono: (920.200396)
- o DiffFormula: (C 27 H 31 N 9 O 15 P 2)
- o Source: (natural)
- o DiffMono: (783.141485)
- o TermSpec: (none)

Subsets

PSI-MOD-slim

Alternate name from RESID

N(tau)-(8alpha-FAD)-histidine

Description (full name) from UniMod

Flavin adenine dinucleotide

Systematic name from RESID

(S)-2-amino-3-(1-(8alpha riboflavin 5'- (trihydrogen diphosphate) 5'->5'-ester with adenosine]imidazol-4-yl)propanoic acid

Short label curated by PSI-MOD

Nt8aFADHis

Protein feature description from UniProtKB

MOD_RES Tele-8alpha-FAD histidine

Alternate name from RESID

8alpha-(N(epsilon)-histidyl)FAD

Protein Modifications
Controlled Vocabulary

Protein Modifications (MOD) Controlled Vocabulary

URL: <https://www.ebi.ac.uk/ols/ontologies/mod>

- Covers a wide range of modifications:
 - Functional groups (e.g. phosphorylation, methylation)
 - Stereochemistry modifications (e.g. D- aminoacids)
 - Isotope variants
 - Modifications to protein backbone (eg. cleavage)
- Natural and man-made changes
 - Fluorescent labels
 - Isotope tags
 - ...

Interaction Records

Issues to be aware of...

- Varying level of detail

- Most of the experimental methods provide only partial information about interactions
 - Not all databases record and provide access to all information collected in a given experiment

- Possible ambiguities

- Identities of the proteins participating in a given interaction might be ambiguous
 - Experiments often do not reveal spatial relationships between interacting molecules

Interaction Data Resources

Interaction Data Resources

Primary Data Providers

The screenshot shows the IMEx website interface. At the top, there's a navigation bar with links for Home, About, Curation, Submit Your Data, and Contact us. Below the navigation is a search bar with placeholder text "Search the IMEx data resource" and "Use as input: UnProtKB Accs, Gene names, Publication Ids". The main content area has two sections: "IMEx data" and "Citing IMEx". The "IMEx data" section contains a bulleted list of features: "A non-redundant set of physical molecular interaction data from a broad taxonomic range of organisms.", "Expertly curated from direct submissions or peer-reviewed journals to a consistent high standard.", "Available in standard formats MITAB or PSI-MI XML 2.5.", and "Provided by a network of participating major public domain databases." The "Citing IMEx" section lists an article: "Protein interaction data curation: the International Molecular Exchange (IMEx) consortium Nat Methods 2012, 9, 345-350" by Orchard, S., et al. At the bottom, there are sections for "IMEx Partners" and "IMEx Observers", each displaying logos for various organizations.

- Detailed (aka 'IMEx-level' or 'deep') curation

- IntAct, EBI (European Bioinformatics Institute), EU
- DIP (Database of Interacting Proteins), UCLA
- MINT (Molecular INTeractions), U of Rome, Italy
- UniprotKB/SwissProt, SIB, Geneva, Switzerland

- Lightweight curation

- BioGRID
 - University of Montreal, Canada
 - Lunenfeld-Tannenbaum Research Institute, Toronto, Canada
 - Princeton University

- Defunct

- BIND
- MIPS/MPact
- HPRD
- MPIDB
- ...

Interaction Data Resources

IMEx Consortium

Participating databases

- Active curation
 - IntAct, EBI (European Bioinformatics Institute), Hinxton, UK
<https://www.ebi.ac.uk/intact>
 - DIP (Database of Interacting Proteins), UCLA
<https://dip.doe-mbi.ucla.edu>, <https://dip.mbi.ucla.edu>
 - MINT (Molecular INTeractions), University of Rome, Italy
<https://mint.bio.uniroma2.it/>
 - UniprotKB/SwissProt, SIB, Geneva, Switzerland
<https://www.uniprot.org/>
- Data redistribution
 - I2D (Interologous Interactions Database), Ontario Cancer Institute, Toronto, Canada
<http://ophid.utoronto.ca>
 - InnateDB, Simon Fraser University, Canada, University of British Columbia, Canada, & EMBL Australia
<https://www.innatedb.com>
 - MatrixDB , Claude Bernard University Lyon 1, France
<http://matrixdb.univ-lyon1.fr/>
 -

Interaction Data Resources

IMEx Consortium

Data availability

- Interactive
 - Individual IMEx databases
 - UniprotKB ('high quality' subset)
- Dataset downloads
 - Original (complete) IMEx records
<https://www.ebi.ac.uk/intact>
 - Legacy (pre-IMEx) DIP records (as part of Current/Full dataset)
<https://dip.mbi.ucla.edu/dip/page?id=download>

Interaction Data Resources

<https://thebiogrid.org>

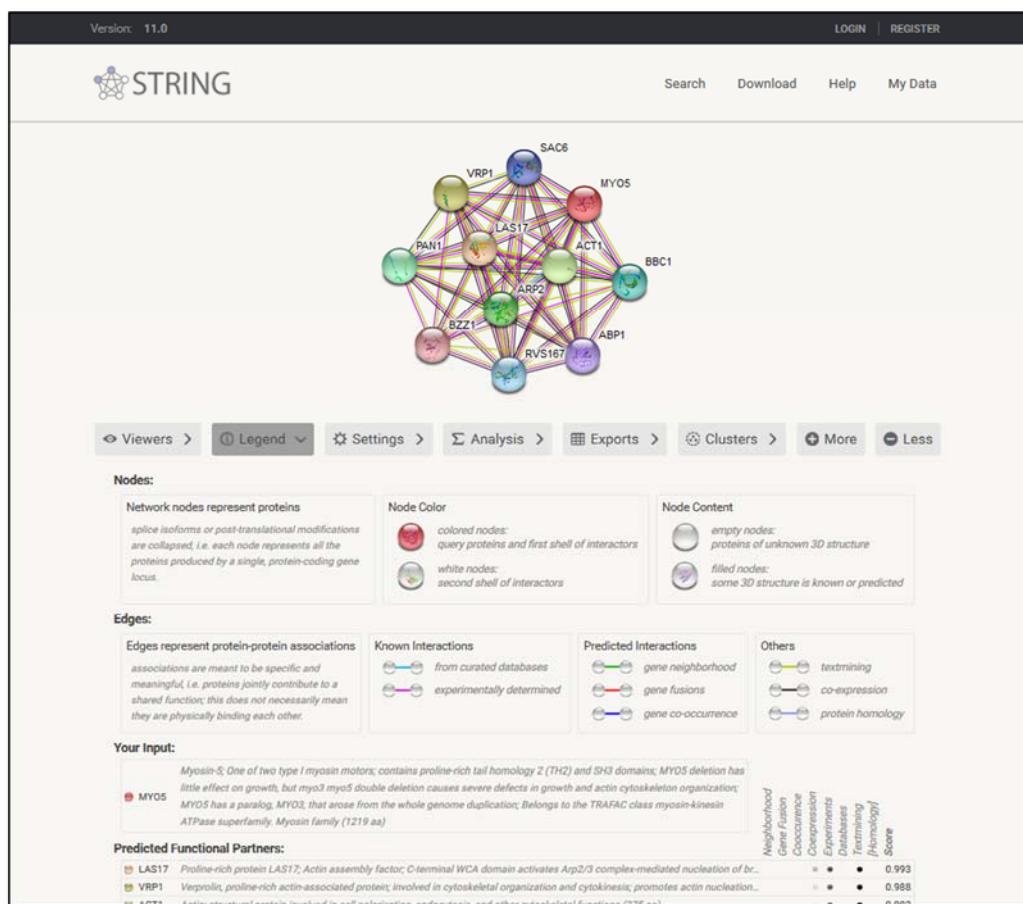
The screenshot shows the BioGRID 3.5 homepage. At the top, there's a navigation bar with links for home, help wiki, tools, contribute, stats, downloads, partners, and about us. A Twitter icon is also present. Below the navigation is a search bar with a dropdown menu set to 'By Identifier'. The main content area features a large search form with fields for 'Enter search terms here...' and 'All Organisms', along with a 'Submit Identifier Search' button. Below the search form are links for 'Advanced Search', 'Helpful Search Tips', and 'Featured Datasets'. To the left, there's a section titled 'Related Resources' with three items: 'BioGRID ORCS - An open repository of CRISPR screens', 'The Kinase and Phosphatase Interactome in *S. cerevisiae* Curation Project', and 'Gene Info eXtension (GIX)'. Each item has a brief description and a 'Learn more' link. To the right, there's a 'Partners' section with logos for NIH, ORIP, CIHR IRSC, Mount Sinai Hospital, Princeton University, and Université de Montréal. Below the partners is a 'Latest Updates' section with a 'Tweets by biogrid' link. At the bottom left, there's a 'BioGRID Statistics' button.

Key Features

- Lightweight curation (thus bigger than IMEx)
 - Pairwise interactions
 - No information on protein state
 - No splice form data
 - No features
 - No PTMs
- Additional interaction types
 - Genetic interactions
- Non-standard (tabular) data format
 - Simplified controlled vocabulary

Interaction Data Resources

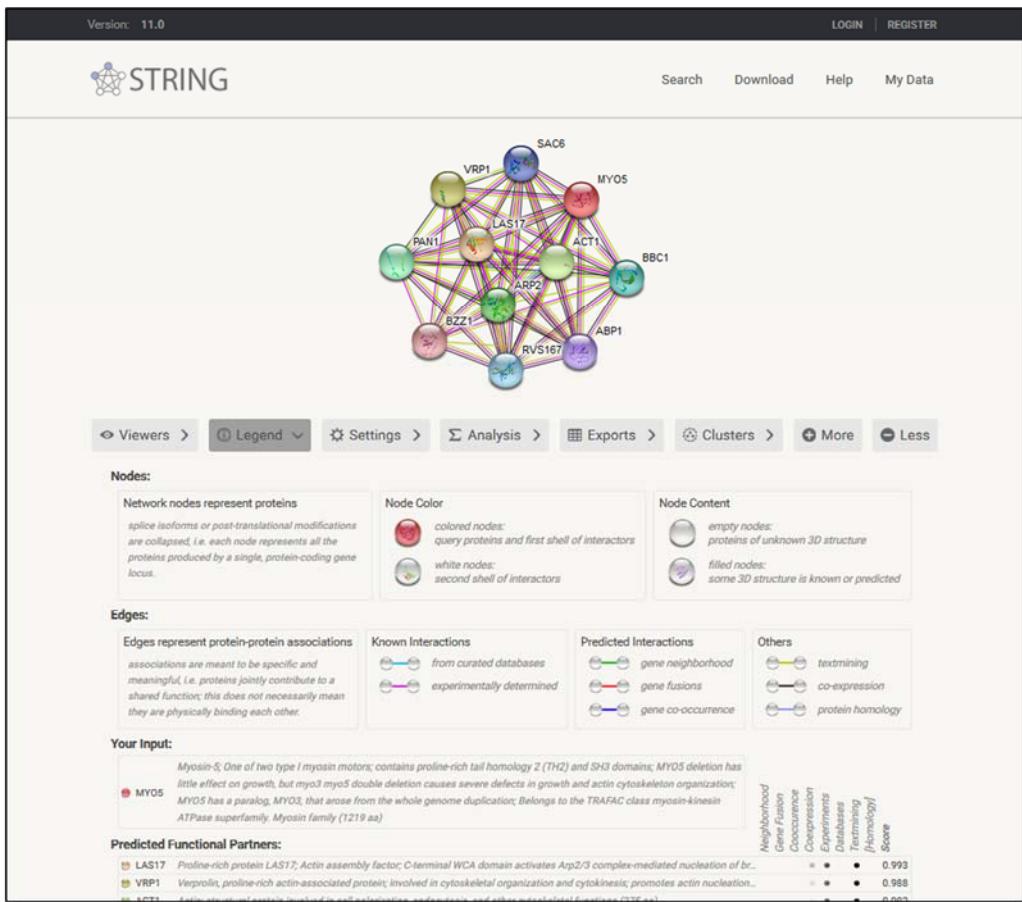
Integrators



- Interaction resources
 - STRING
 - GeneMania
 - HIPPIE
 - HINT
 - APID
 - ...
 - General resources
 - UniprotKB
 - Reactome
 - Defunct
 - iRef database/iRefIndex
 - MiMI database
 - ...

Interaction Data Resources

<https://string-db.org>



Key Features

- Identifies pairwise functional associations between genes

Operational definition: genes that function within the same pathway or module.

- Only gene level information
- Functional associations include:
 - Computational predictions (in house)
 - Co-occurrence in genomes
 - Genome neighborhood
 - Gene fusion (Rosetta Stone)
 - Text Mining
 - Co-occurrence in publications
 - Molecular interactions
 - IMEx Consortium
 - BioGRID
 - ... and others
 - Pathway information
 - Reactome
 - KEGG
 - BioCyc
 - ... and others
 - Co-expression

Interaction Data Resources

PSI Common Query Interface (PSICQUIC) Servers

Registry: <https://www.ebi.ac.uk/Tools/webservices/psicquic/registry/registry?action=STATUS>

The screenshot shows the PSICQUIC Registry interface. At the top, there's a navigation bar with links for Services, Research, Training, About us, and a search bar. Below the header, the title "PSICQUIC Registry" is displayed. Underneath, there's a sub-header "Registry" and a message stating "Total: 9,131,104 Interactions from 33 PSICQUIC Services, of which 10 are currently down." A "Filter:" input field is present. The main content is a table listing services with columns for Name, Status, Interactions, Version, URLs, Description, Restricted, and Tags. The table includes rows for iRefIndex, BioGrid, BindingDB, I2D, and IMEx. Each row provides details about the service, including its URL, version, and specific evidence types like "protein:protein internally-curated evidence".

Name	Status	Interactions	Version	URLs	Description	Restricted	Tags
iRefIndex	Green	2,338,337	1.3.14	SOAP: http://irefindex.vib.be/webservices/psicquic REST: http://irefindex.vib.be/webservices/current/search/ REST example		NO	protein:protein imported bipartite expansion evidence
BioGrid	Green	1,513,281	1.3.14	SOAP: http://tyersrest.tyerslab.com:8805/psicquic/webservices/psicquic REST: http://tyersrest.tyerslab.com:8805/psicquic/webservices/current/search/ REST example		NO	protein:protein internally-curated rapid curation spoke expansion evidence
BindingDB	Green	1,011,029	v1.3	SOAP: http://bindingdb.org/psicquic-vs REST: http://bindingdb.org/psicquic-vs/webservices/psicquic/current/search/ REST example		NO	smallmolecule: protein internally-curated evidence spoke expansion experimentally: observed
I2D	Red	817,915	1.1.6	SOAP: http://ophid.utoronto.ca/psicquic-vs/webservices/psicquic REST: http://ophid.utoronto.ca/psicquic-vs/webservices/current/search/ REST example		NO	protein:protein internally-curated evidence
IMEx	Green	717,696	1.3.14	SOAP: http://www.ebi.ac.uk/Tools/webservices/psicquic/imex/webservices/psicquic REST: http://www.ebi.ac.uk/Tools/webservices/psicquic/imex/webservices/current/search/	IMEx contains experimentally verified protein interaction records annotated by members of the IMEx Consortium	NO	protein:protein smallmolecule: protein nucleicacid:protein internally-curated

Programmatic access to interaction data

- SOAP & REST APIs
- Results returned as PSI-MI, MITAB, etc
- MIQL Queries
 - (brca1 or brca2) and species:human
 - (atpa_yeast or atpb_yeast) and type:"physical association"

References

- <https://psicquic.github.io/MiQLDefinition.html>
- Aranda *et al.* PSICQUIC and PSISCORE: accessing and scoring molecular interactions.
Nat Methods. 8:528-9 (2011) PMID:21716279
- Nucleic Acids Res. del-Toro *et al.* A new reference implementation of the PSICQUIC web service.
NAR 41:W601-606 (2013) PMID:2367134