

Protein-Protein Interactions

Introduction

May 2019

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

Molecular Interactions

Why do we care ?

UniProtKB - P04637 (P53_HUMAN)

Display

Entry: Protein | **Cellular tumor antigen p53**
Gene | **TP53**
Organism | **Homo sapiens (Human)**

Status: Reviewed - Annotation score: ***** - Experimental evidence at protein level¹

Function¹:

Act as tumor suppressor gene product, induces growth arrest and apoptosis depending on the phase of growth cycle and cell type. Involved in cell cycle regulation as a trans-activator that can to negatively regulate cell division by controlling a set of genes required for this process. One of the activated genes is an inhibitor of cyclin-dependent kinase. Apoptosis induction seems to be mediated either by stimulation of BAX and FAS antigen expression, or by repression of Bcl-2 expression. In cooperation with mitochondrial PPIF is involved in activating oxidative stress-induced necrosis; the function is largely independent of transcription. Induces the transcription of long intergenic non-coding RNA p21 (lncRNA-p21) and lncRNA-Mkhn1. LncRNA-p21 participates in TP53-dependent transcriptional repression leading to apoptosis and seems to have an effect on cell-cycle regulation.

Implicated in Notch signaling cross-over. Prevents CDK7 kinase activity when associated to CAK complex in response to DNA damage, thus stopping cell cycle progression. Isoform 1 transcript is an alternative transcript containing two in-frame start coders. Isoform 4 suppresses transcription activity and impairs growth suppression mediated by isoform 1. Isoform 7 inhibits isoform 1-mediated apoptosis. Regulates the circadian clock by repressing CLOCK-ARNTL/BMAL1-mediated transcriptional activation of PER2 (PubMed:240511492).

Cofactor: Zn²⁺
 Note: Binds 1 zinc ion per subunit.

Sites

Feature key	Position(s)	Description	Actions	Graphical view	Length
Metal binding ¹	176	Zinc			1
Metal binding ¹	179	Zinc			1
Metal binding ¹	238	Zinc			1
Metal binding ¹	742	Zinc			1

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 ?

The screenshot shows the UniProtKB interface for protein P04637 (P53_HUMAN). The main content area displays the 'GO - Biological process' section, which lists numerous biological processes. A sidebar on the left contains various filter categories like 'Display', 'Entry', 'Publications', etc., with 'Function' selected. The list of processes includes: autophagy, base-excision repair, cell aging, cell cycle arrest, cell differentiation, cell population proliferation, cellular protein localization, cellular response to actinomycin D, cellular response to DNA damage stimulus, cellular response to drug, cellular response to heat shock, cellular response to glucose starvation, cellular response to hypoxia, cellular response to ionizing radiation, cellular response to UV, chromatin assembly, circadian behavior, cytokine-mediated signaling pathway, determination of adult lifespan, DNA damage response, DNA damage response, signal transduction by p53 class mediator resulting in cell cycle arrest, DNA damage response, signal transduction by p53 class mediator resulting in transcription of p21 class mediator, DNA strand renaturation, ER overload response, extrinsic apoptotic signaling pathway, intrinsic apoptotic signaling pathway by p53 class mediator, intrinsic apoptotic signaling pathway in response to DNA damage by p53 class mediator, mitotic G1 DNA damage checkpoint, mRNA transcription, multicellular organism development, and negative regulation of apoptotic process.

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 protein P04637 (P53_HUMAN). The main content area displays the 'GO - Molecular function' section, listing various molecular functions. A sidebar on the left contains various filter categories like 'Display', 'Entry', 'Publications', etc., with 'Function' selected. The list of molecular functions includes: ATP binding, chaperone binding, chromatin binding, coiled coil interaction, core promoter sequence-specific DNA binding, disordered domain specific binding, DNA binding, DNA-binding transcription factor activity, enzyme binding, histone acetyltransferase binding, histone deacetylase binding, identical protein binding, mRNA 3'-UTR binding, p53 binding, promoter-specific chromatin binding, protease binding, protein heterodimerization activity, protein kinase binding, protein N-terminal binding, protein phosphotyrosine binding, protein self-association, receptor tyrosine kinase binding, RNA polymerase II distal enhancer sequence-specific DNA binding, RNA polymerase II proximal promoter sequence-specific DNA binding, RNA polymerase II transcription factor binding, TFIID-class transcription factor complex binding, transcription factor binding, transcription regulatory region DNA binding, and ubiquitin protein ligase binding.

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 ?

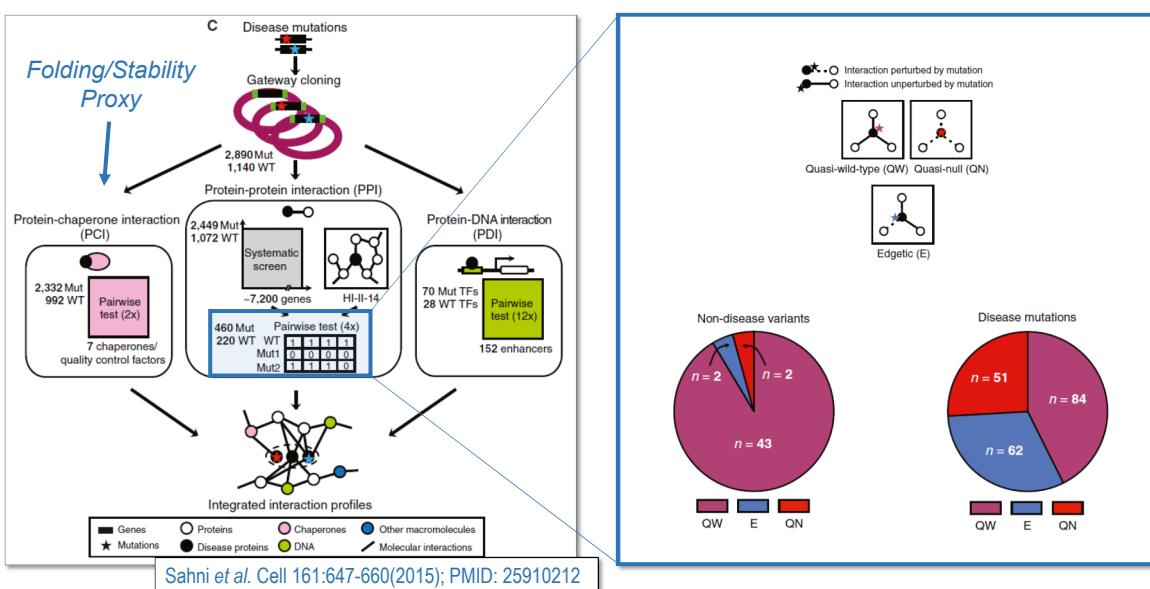
Function

- 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



Molecular Interactions

Why do we care ?

The screenshot shows the UniProtKB interface for protein P04637 (TP53_HUMAN). The left sidebar has 'Mutagenesis' selected under 'Feature key'. The main panel displays a table of mutations with columns for Feature key, Position(s), Description, Actions, Graphical view, and Length. Key mutations include:

- 15 S → A: Loss of interaction with PPP2RSC, PPP2CA AND PPP2R1A.
- 18 T → A: No effect on interaction with MDM2 and increase in protein levels after DNA damage. 1 Publication
- 20 S → A: Abolishes phosphorylation site. Abolishes increase in protein levels after DNA damage. 1 Publication
- 22 – 23 LW → QS: Loss of interaction with MDM2, leading to constitutively increased TP53 protein levels. 1 Publication
- 37 S → D: Abolishes phosphorylation by MAPKAPK5. 1 Publication
- 46 S → A: Abolishes phosphorylation by DYRK2 and HIPK2 and acetylation of K-382 by CREBBP. 3 Publications
- 46 Missing : Alters interaction with WWOX. 3 Publications
- 55 T → A: Blocks phosphorylation by TAF1. 1 Publication
- 183 S → A: Abolishes strongly phosphorylation. 1 Publication
- 183 S → E: Inhibits slightly its transcriptional activity. 1 Publication
- 248 R → S: Does not induce SNAI1 degradation. 1 Publication
- 269 S → E: Inhibits strongly its transcriptional activity. 1 Publication
- 269 S → E: Inhibits strongly its transcriptional activity. 1 Publication
- 284 T → E: Inhibits strongly its transcriptional activity. 1 Publication
- 291 – 292 KK → RR: Abolishes polyubiquitination by MKRN1. 1 Publication
- 319 K → A: Loss of nuclear localization; when associated with A-320 and A-321. 1 Publication
- 320 K → A: Loss of nuclear localization; when associated with A-319 and A-321. 1 Publication
- 321 K → A: Loss of nuclear localization; when associated with A-319 and A-320. 1 Publication
- 333 – 337 RGRR → GKGK: Reduced methylation by PRMT5. Reduced nuclear localization. Decreased binding to promoters of target genes. Reduced transnational activity. Decrease in cell cycle arrest. 1 Publication

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

In many cases
mutations modify
Molecular Interactions

Molecular Interactions

Why do we care ?

The screenshot shows the UniProtKB interface for protein P04637 (TP53_HUMAN). The left sidebar has 'Mutagenesis' selected under 'Feature key'. The main panel displays a table of mutations with columns for Feature key, Position(s), Description, Actions, Graphical view, and Length. A tooltip is shown over a mutation entry:

Manual assertion based on experiment in:
"Regulation of p53 activity by its interaction with homeodomain-interacting protein kinase-2."
Droge W., Will H., Schmitz M., Zentgraf H., Taya Y., Nat. Cell Biol. 4:1-10(2002) [PubMed] [Europe PMC]

Other mutations listed include:

- 15 S → A: Loss of interaction with PPP2RSC, PPP2CA AND PPP2R1A.
- 18 T → A: No effect on interaction with MDM2 and increase in protein levels after DNA damage. 1 Publication
- 20 S → A: Abolishes phosphorylation site. Abolishes increase in protein levels after DNA damage. 1 Publication
- 22 – 23 LW → QS: Loss of interaction with MDM2, leading to constitutively increased TP53 protein levels. 1 Publication
- 37 S → D: Abolishes phosphorylation by MAPKAPK5. 1 Publication
- 46 S → A: Abolishes phosphorylation by DYRK2 and HIPK2 and acetylation of K-382 by CREBBP. 3 Publications
- 46 Missing : Alters interaction with WWOX. 3 Publications
- 55 T → A: Blocks phosphorylation by TAF1. 1 Publication
- 183 S → A: Abolishes strongly phosphorylation. 1 Publication
- 183 S → E: Inhibits slightly its transcriptional activity. 1 Publication
- 248 R → S: Does not induce SNAI1 degradation. 1 Publication
- 269 S → E: Inhibits strongly its transcriptional activity. 1 Publication
- 269 S → E: Inhibits strongly its transcriptional activity. 1 Publication
- 284 T → E: Inhibits strongly its transcriptional activity. 1 Publication
- 291 – 292 KK → RR: Abolishes polyubiquitination by MKRN1. 1 Publication
- 319 K → A: Loss of nuclear localization; when associated with A-320 and A-321. 1 Publication
- 320 K → A: Loss of nuclear localization; when associated with A-319 and A-321. 1 Publication
- 321 K → A: Loss of nuclear localization; when associated with A-319 and A-320. 1 Publication
- 333 – 337 RGRR → GKGK: Reduced methylation by PRMT5. Reduced nuclear localization. Decreased binding to promoters of target genes. Reduced transnational activity. Decrease in cell cycle arrest. 1 Publication

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 we care ?

The screenshot shows the UniProtKB entry for P04637 (P53_HUMAN). The left sidebar has 'Display' set to 'None'. Under 'GO - Molecular function', several annotations are listed, each with a source link. A yellow box highlights a specific annotation: "A novel cofactor for p300 that regulates the p53 response." This annotation is linked to a publication by Shikama N., Lee C.-W., France S., Delavaine L., Lyon J., Krstic-Demonacos M., La Thangue N.B. in Mol. Cell 4:365-379(1999) [PubMed] [Europe PMC] [Abstract].

- ATP binding * Source: UniProtKB
- chromatine binding * Source: UniProtKB
- chromatin binding * Source: UniProtKB
- copper ion binding * Source: UniProtKB
- core promoter sequence-specific DNA binding * Source: CAFA
- disordered domain specific binding * Source: CAFA
- DNA binding * Source: UniProtKB
- DNA-binding transcription activator activity, RNA polymerase II-specific * Source: ARUK-UCL
- DNA-binding transcription factor activity * Source: UniProtKB
- DNA-binding transcription factor activity, RNA polymerase II-specific * Source: UniProtKB
- DNAbinding
- histone acetyltransferase binding * Source: UniProtKB
- histone deacetylase binding
- identical protein binding * Source: UniProtKB
- mRNA 3'-UTR binding * Source: UniProtKB
- p53 binding * Source: CAFA
- promoter-specific chromatin binding
- protease binding * Source: UniProtKB
- protein heterodimerization activity
- protein kinase binding * Source: UniProtKB
- protein N-terminal binding
- protein phosphatase 2A binding * Source: UniProtKB
- protein phosphatase binding * Source: UniProtKB
- protein self-association * Source: AtpBase
- receptor tyrosine kinase binding * Source: BHF-UCL
- RNA polymerase II distal enhancer sequence-specific DNA binding * Source: UniProtKB
- RNA polymerase II proximal promoter sequence-specific DNA binding * Source: ParkinsonsUCL
- RNA polymerase II transcription factor binding * Source: BHF-UCL
- TFIID-class transcription factor complex binding * Source: ParkinsonsUCL
- transcription factor binding * Source: UniProtKB
- transcription regulatory region DNA binding * Source: BHF-UCL
- ubiquitin protein ligase binding * 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

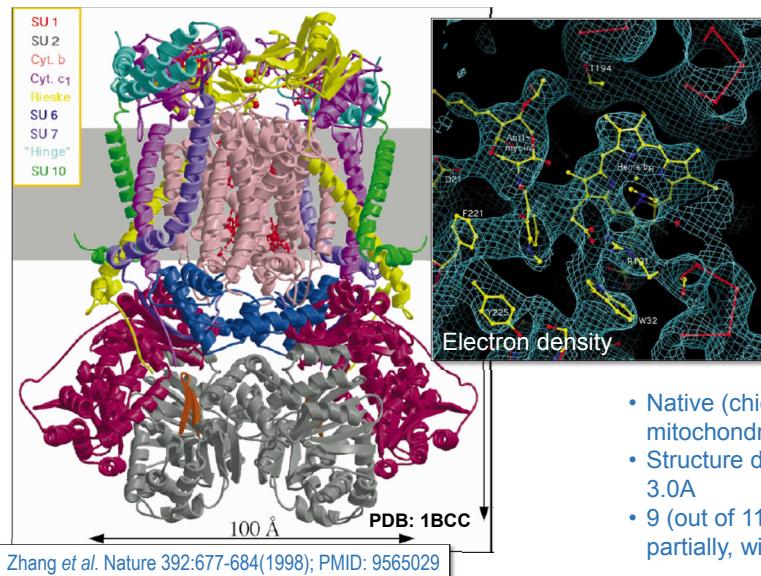
- Computer-readable description of function
- Three complementary sections
 - Biological process
 - Molecular function
 - Cellular localization
- Each GO term is backed up by evidence
 - Experimental evidence
 - Sequence similarity
 - Curator's inference
 -

Experimental Methods

Interaction Experiments: Caveats

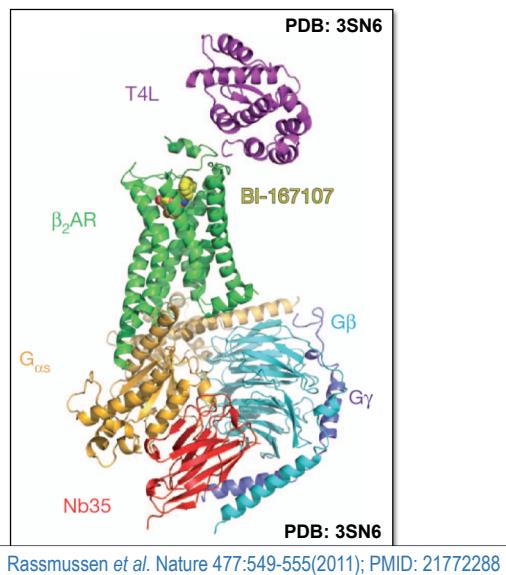
- Protein interactions are determined by 3D structure of a folded protein chain
 - Experimental conditions matter (ionic strength, pH, temperature, cofactors, etc)
 - Possible cellular compartment and tissue specificity
- Protein interactions might be modulated by post-translational modifications
 -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

Interaction Experiments: Examples



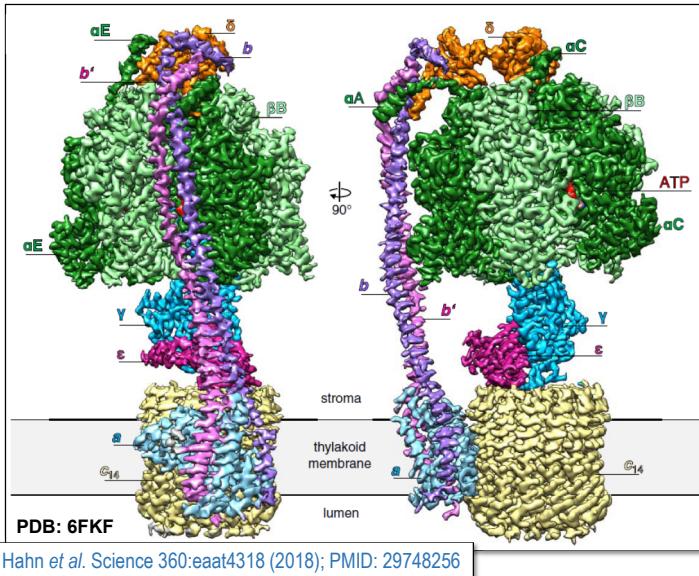
- Native (chicken) complex purified from heart mitochondria
- Structure determined by X-ray crystallography to 3.0 Å
- 9 (out of 11) subunits identified/localized, at least partially, within the structure

Interaction Experiments: Examples



- Fusion of a fragment of human β_2 -AR with lysozyme fusion expressed in insect (Sf9) cells
- G α s (bovine), G β_1 (rat, His-tagged) and G γ_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 3.2 Å

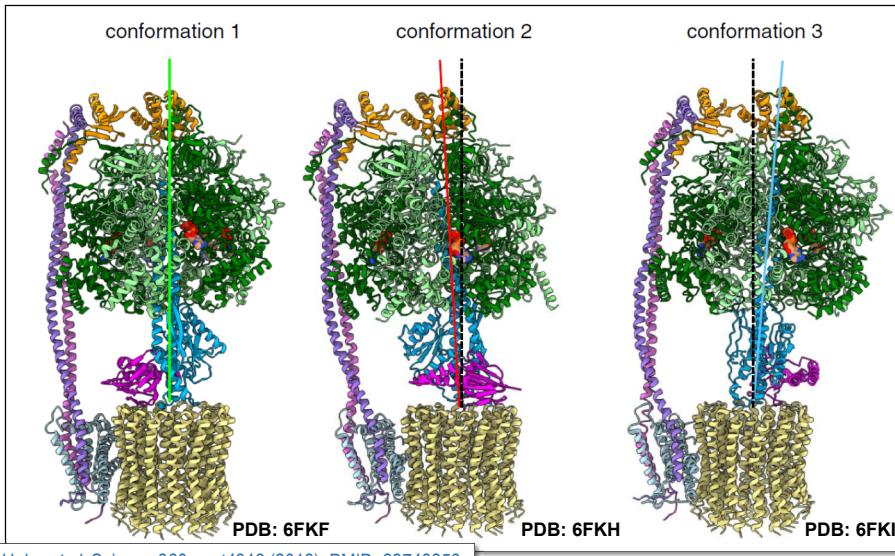
Interaction Experiments: Examples



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

- Native expression (spinach)
- Cryo-EM structure determination

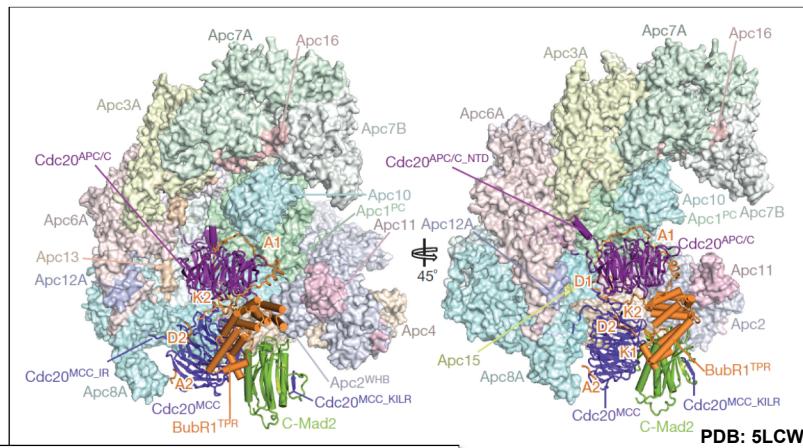
Interaction Experiments: Examples



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

- Conformational states

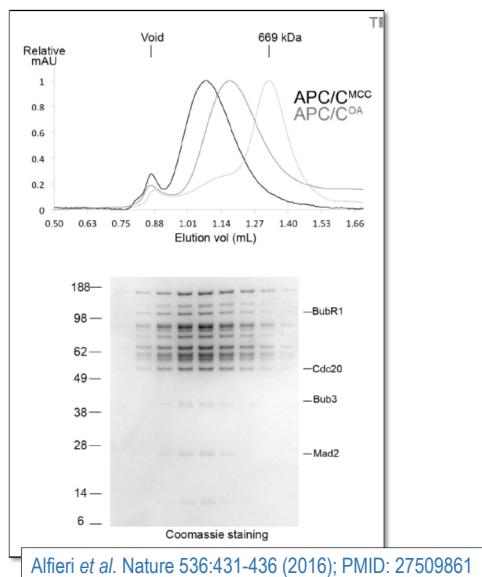
Interaction Experiments: Examples



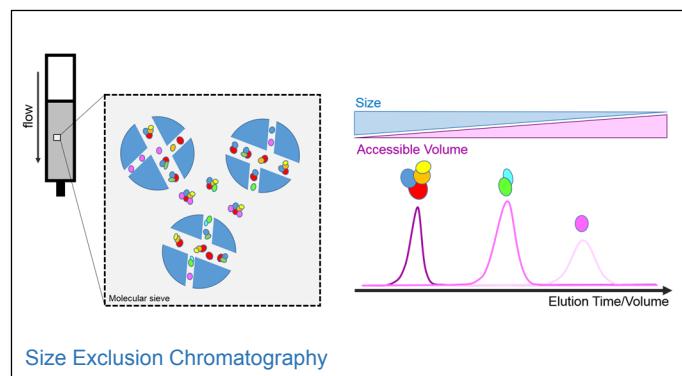
- Heterologous expression (human proteins/insect cell expression)
 - *In vitro* complex reconstitution
 - Cryo-EM structure determination
 - Alternative conformational states

Alfieri et al. Nature 536:431-436 (2016); PMID: 27509861

Interaction Experiments: Examples

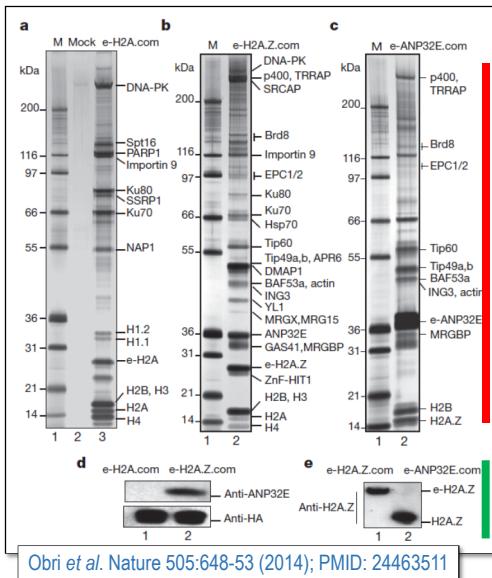


- Heterologous protein expression (insect cells)
 - *In vitro* complex reconstitution
 - Size exclusion chromatography of the final purified complex



Alfieri et al. Nature 536:431-436 (2016); PMID: 27509861

Interaction Experiments: Examples

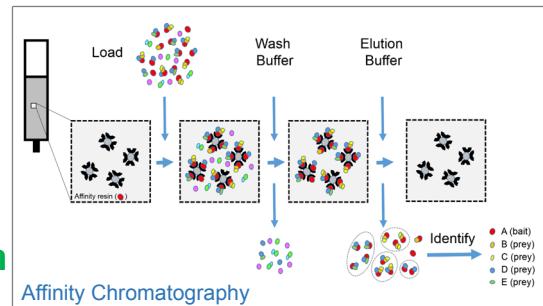


Obri et al. Nature 505:648-53 (2014); PMID: 24463511

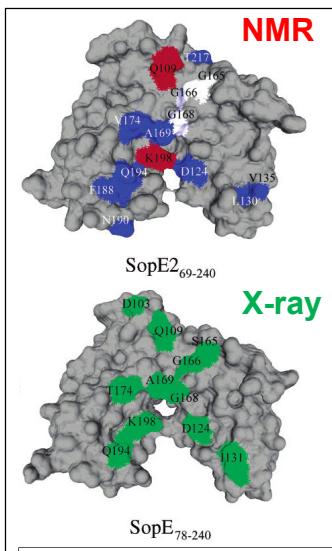
- Overexpression (human) of doubly-tagged (FLAG & HA) human H2A, H2A.Z & ANP32E proteins
- Tandem affinity purification of interacting partners
- Mass spectrometric and/or Western blot protein identification

MS

Western

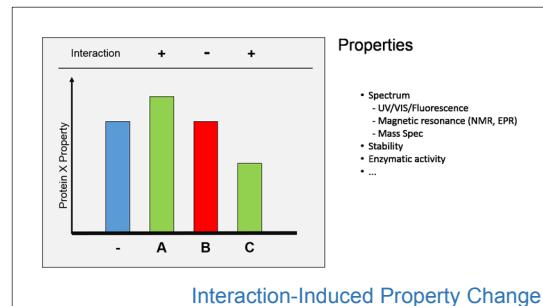


Interaction Experiments: Examples

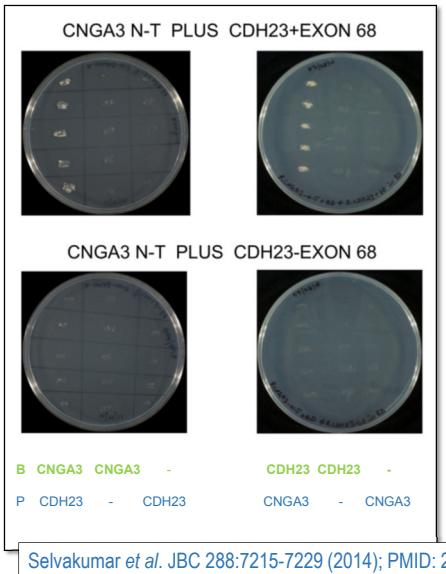


Williams et al. Biochem. 43:11990-1008 (2004); PMID: 15379540

- Heterologous (E. coli) expression of *Salmonella* (SopE2) and human (Cdc42) protein fragments
- *In vitro* complex formation by purified proteins
- Monitor perturbation of SopE2 NMR spectra by Cdc42 protein

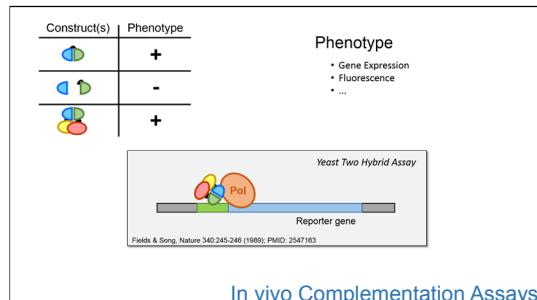


Interaction Experiments: Examples



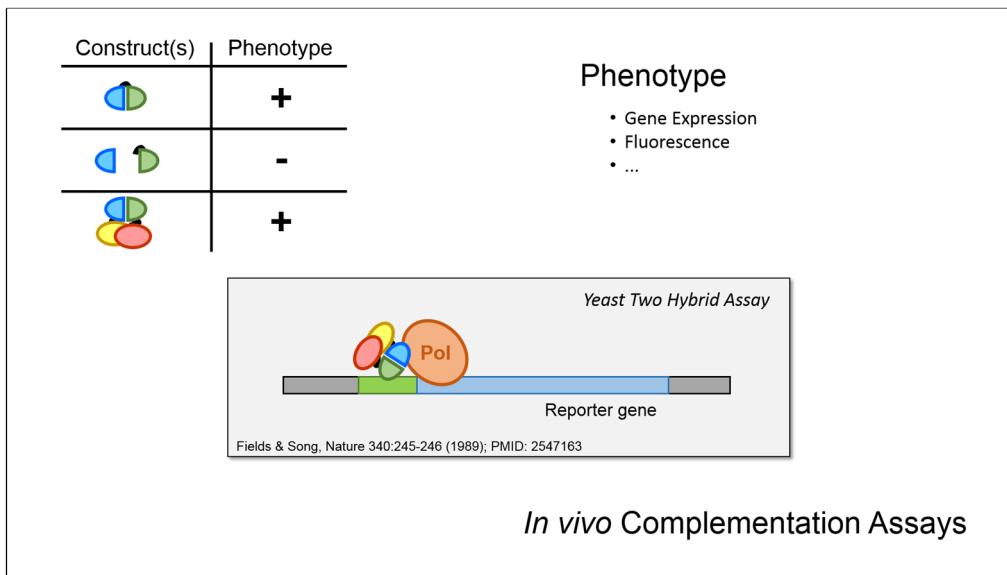
Selvakumar et al. JBC 288:7215-7229 (2014); PMID: 23329832

- 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

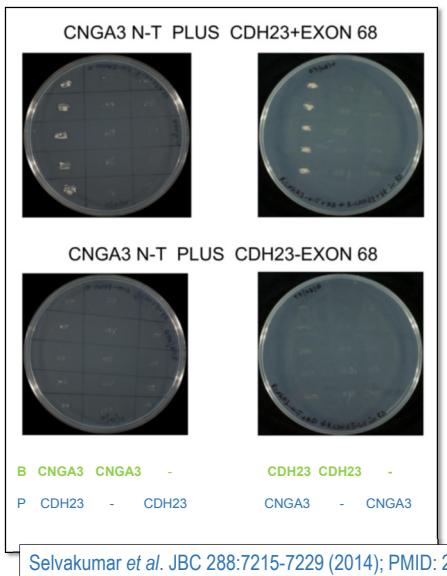


In vivo Complementation Assays

Interaction Experiments

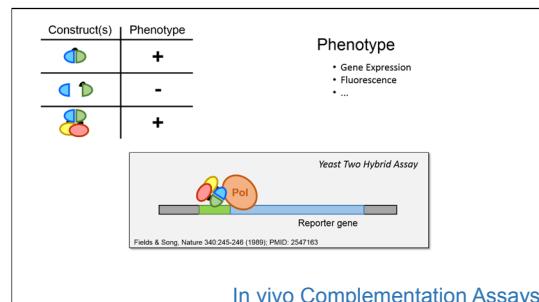


Interaction Experiments: Examples



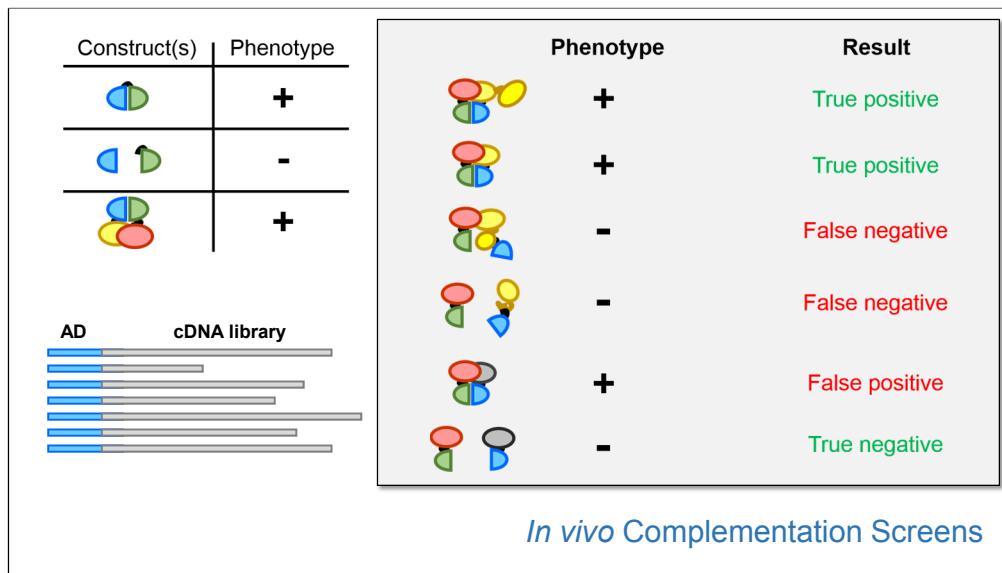
Selvakumar et al. JBC 288:7215-7229 (2014); PMID: 23329832

- 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

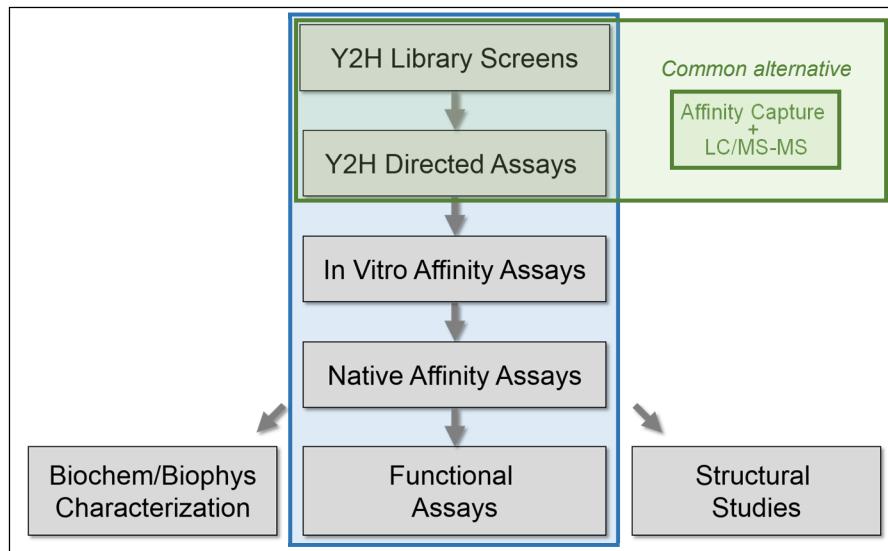


In vivo Complementation Assays

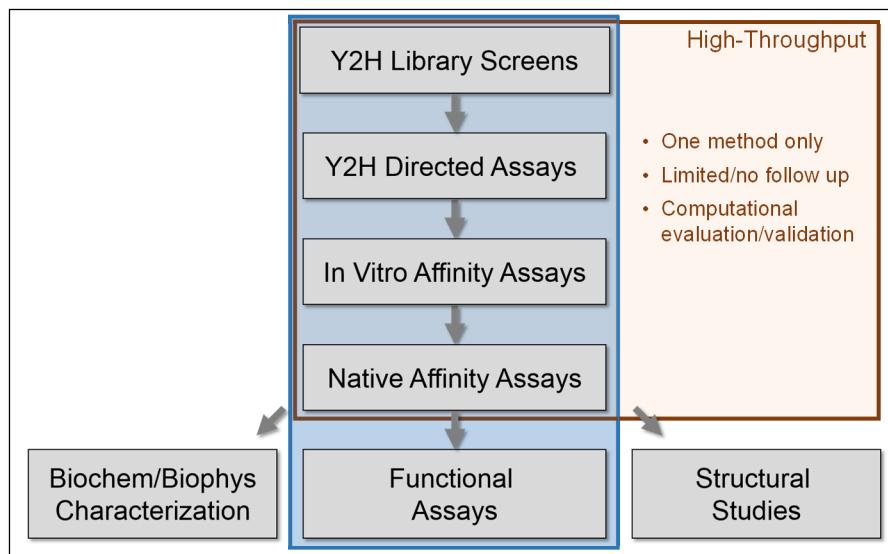
Interaction Experiments



Interaction Experiments: Typical Workflows



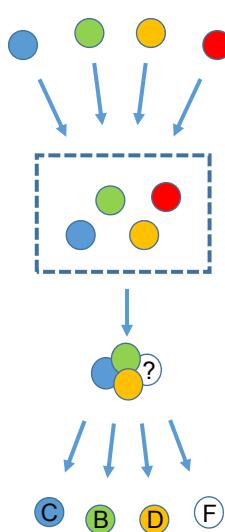
Interaction Experiments: Typical Workflows



Interaction Database Records

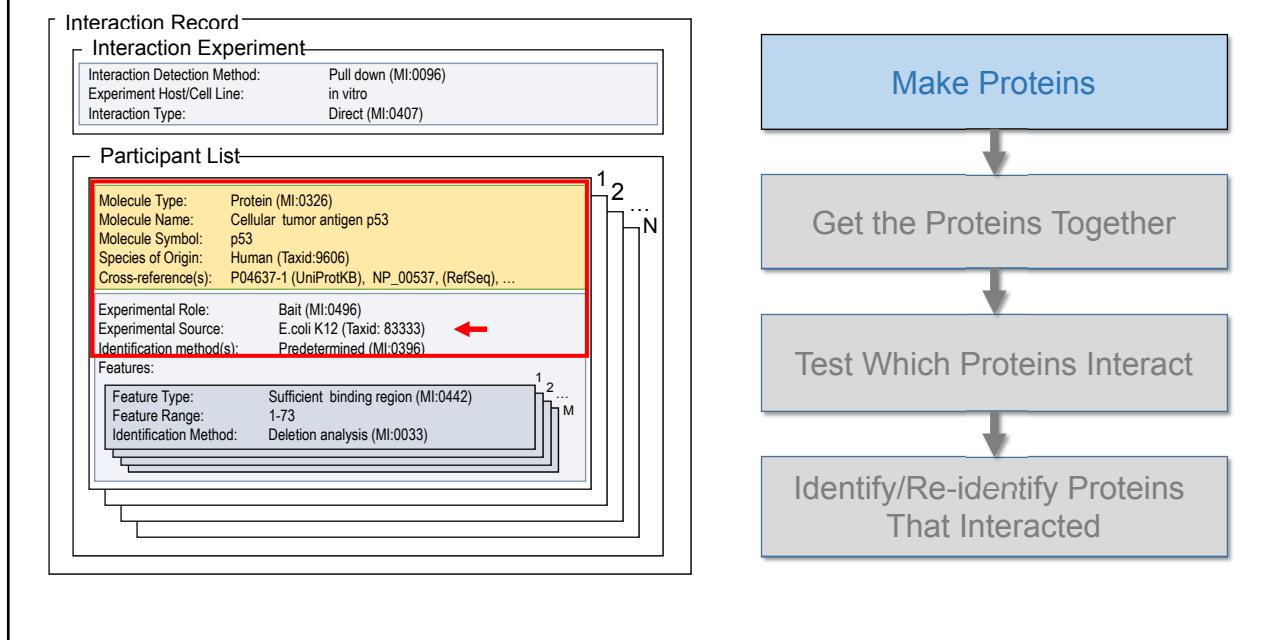
Content

Interaction Experiment Flow

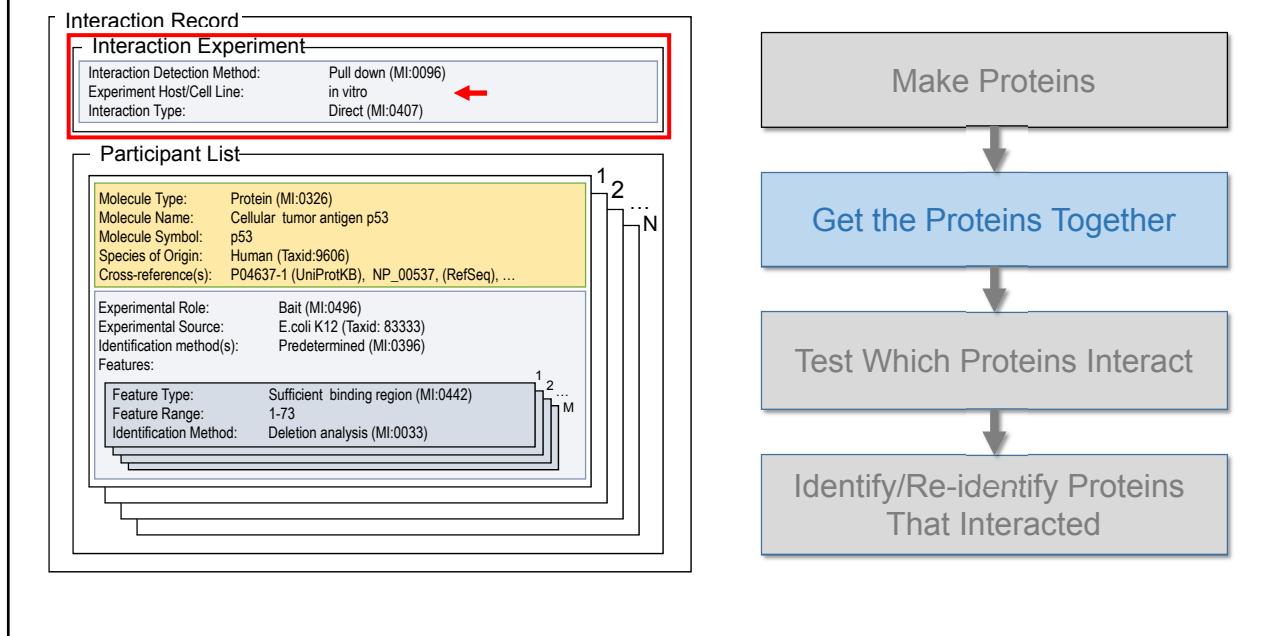


- **Make proteins**
 - Where: native vs heterologous host vs in vitro translation vs chemical synthesis
 - How much: native level vs overexpressed
 - Modifications: isoforms, fragments, mutations, PTMs present/absent
- **Get them together**
 - Where: native organism/cell type/tissue/compartment vs something 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 Experiment Record



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)	1 2 ... N
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)	1 2 ... M
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 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)	1 2 ... N
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)	1 2 ... M
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)	

Make Proteins

Get the Proteins Together

Test Which Proteins Interact

Identify/Re-identify Proteins That Interacted

Interaction Experiment Record

Interaction Record

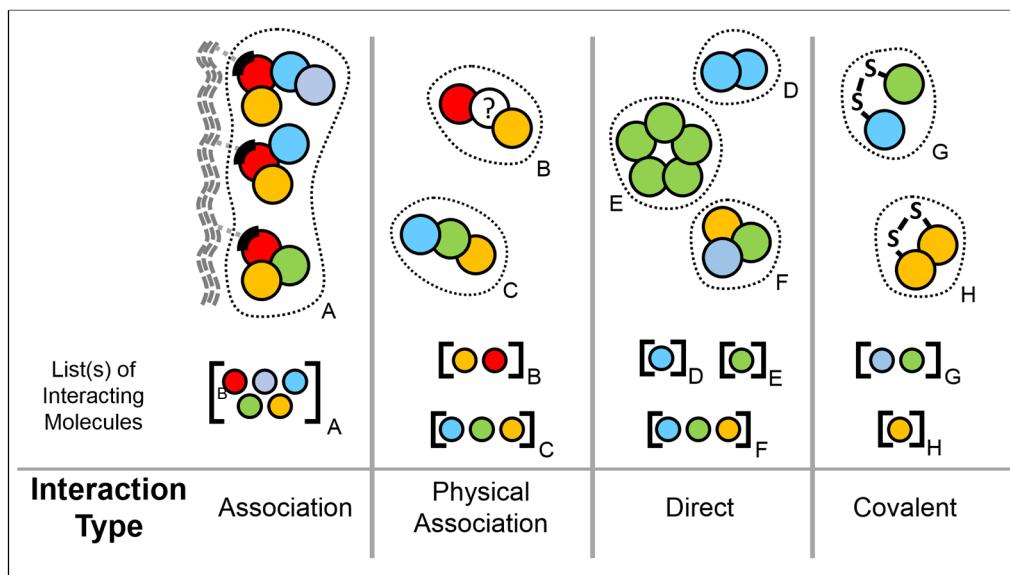
- Interaction Experiment

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

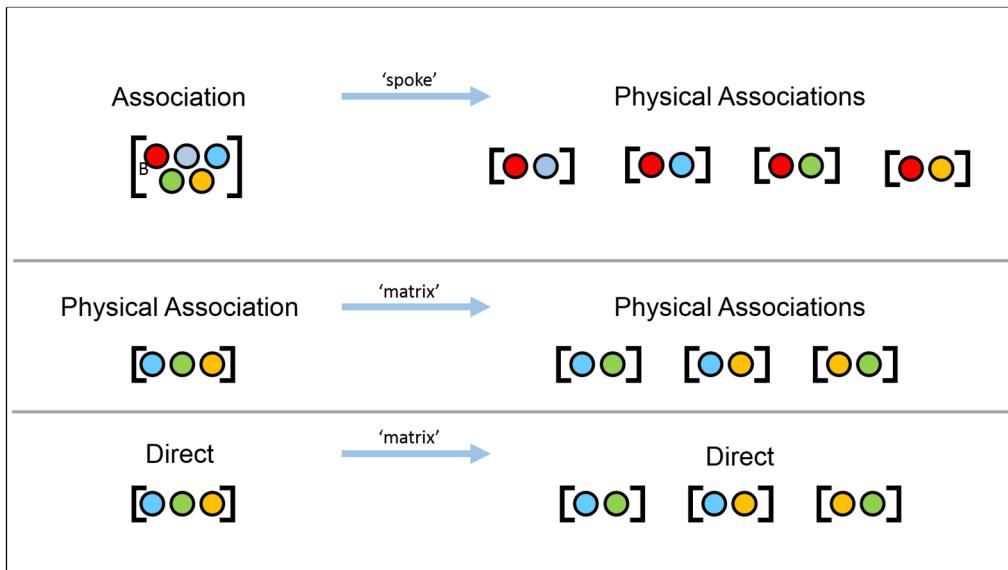
Molecule Type:	Protein (MI:0326)
Molecule Name:	Cellular tumor antigen p53
Molecule Symbol:	p53
Species of Origin:	Human (Taxid:9606)
Cross-reference(s):	P04637-1 (UniProtKB), NP_00537, (RefSeq), ...
Experimental Role:	Bait (MI:0496)
Experimental Source:	E.coli K12 (Taxid: 83333)
Identification method(s):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

Make Proteins
↓
Get the Proteins Together
↓
Test Which Proteins Interact
↓
Identify/Re-identify Proteins That Interacted

Interaction Types



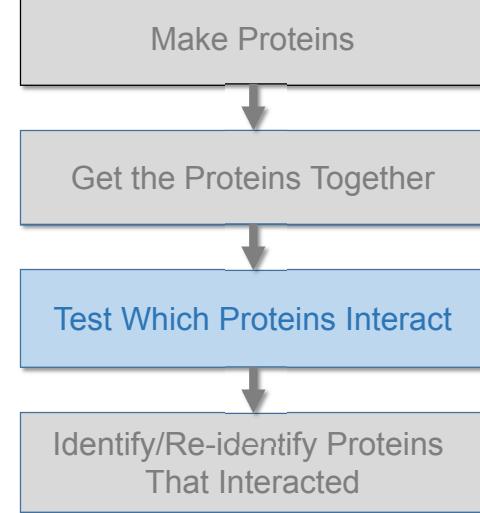
Binary Expansion



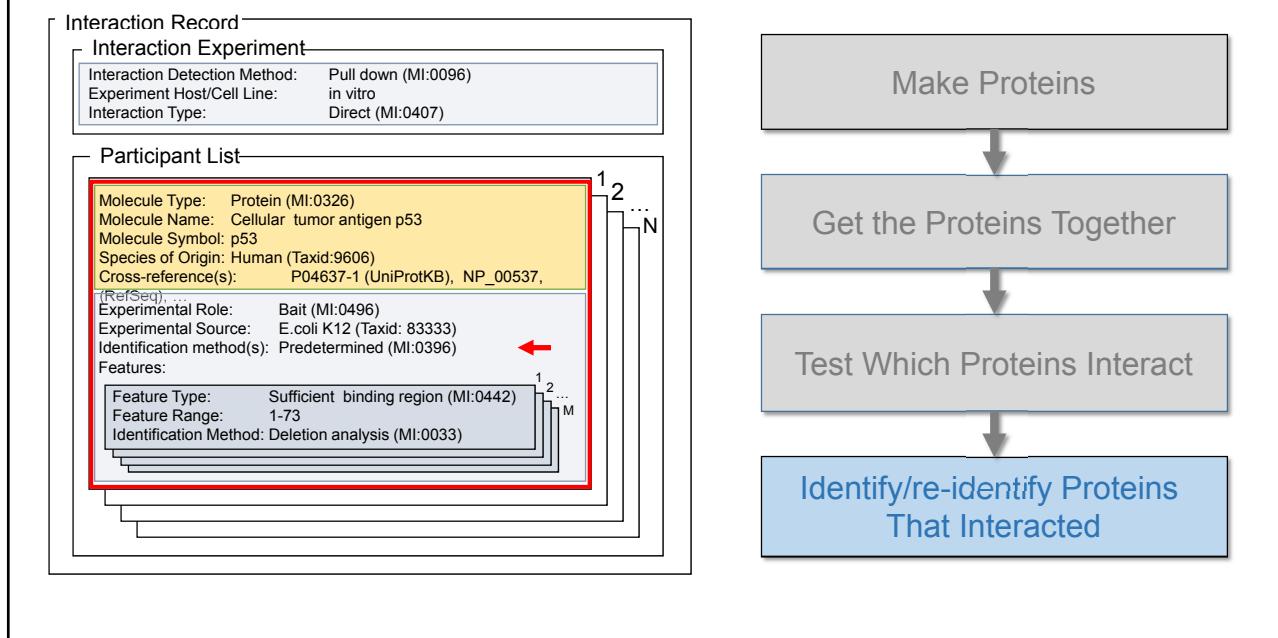
Interaction Experiment Record

Interaction Record

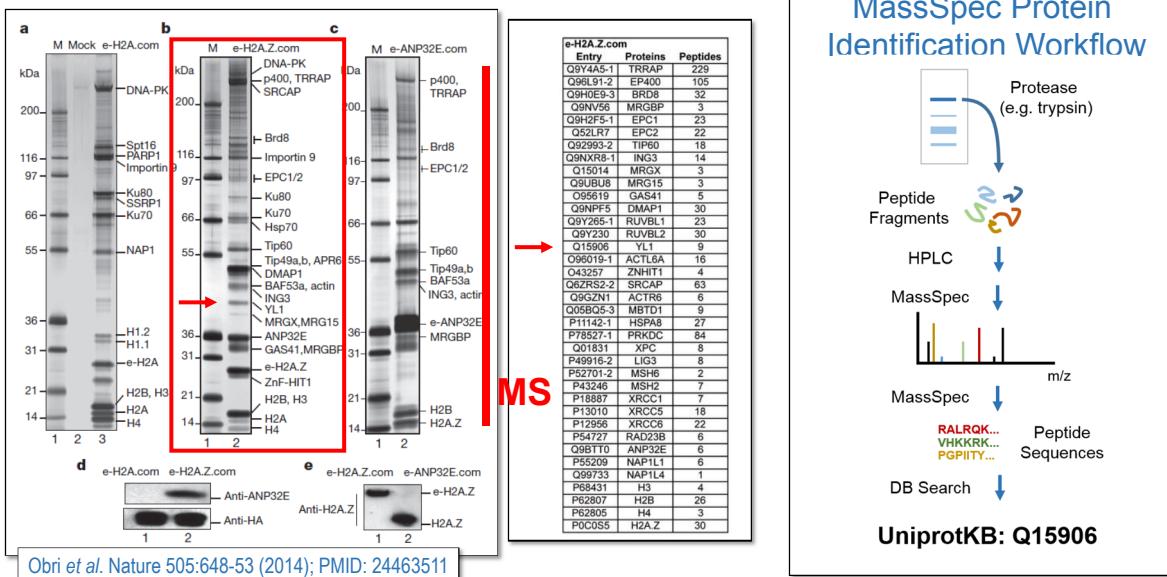
- Interaction Experiment
 - Interaction Detection Method: Pull down (MI:0096)
 - Experiment Host/Cell Line: *in vitro*
 - Interaction Type: Direct (MI:0407)
- Participant List
 - Molecule Type: Protein (MI:0326)
 - Molecule Name: Cellular tumor antigen p53
 - Molecule Symbol: p53
 - Species of Origin: 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)



Interaction Experiment Record



Protein Identification



UniprotKB: Q15906

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

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

Length: 364
Mass (Da): 40,594
Last modified: November 1, 1996 - v1
Checksum: 0AE9B0B62B2BCA4A

BLAST ▾ [GO](#)

```

10   20   30   40   50
MSLAGGRAPK KTAGNRLSGL LEAEEEDDEFY QTTYGQFTEE SGDDEYQQDQ
 60   70   80   90   100
SDTEDEVDSQ FQIEGEDEPS SGDGEAEPEPK KRRVVVTKAYK EPLKSLSRKK
110  120  130  140  150
VNTPAGSSQK AREEKFALLPEL ELQDGGSDDSK KSMRSQTAEH TRQTFLVQE
160  170  180  190  200
RQQQSERRSK PHECERFLQEE ELLEREAKITKE ELMRSLSLETY ERLEACKEQ
210  220  230  240  250
VRKKEKRCGP K IITTHSTVTFV LVGEPGPKKEE VNUIEDLQPA PSVSVLTHA
260  270  280  290  300
GTGPNPFAK CSETFITFSD DATTEENFPQ GRFPKVVFRE VCFVTHRFAL
310  320  330  340  350
YRDPTVTPY ATARAFKIIIR EAKKYKTIATN GLPPTIASLG PGFFFFPEPLP
360
GSGFRALRKQK 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

Note: No experimental confirmation available.
[Show ▾](#)

Computationally mapped potential isoform sequences¹

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
<input checked="" type="checkbox"/> A0A1W2PPT2	A0A1W2PPT2_HUMAN	Vacuolar protein sorting-associated...	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 'canonical' sequence. All positional information in this entry refers to it. This is also the sequence that appears in the downloadable versions of the entry.

Length: 364
Mass (Da): 40,594
Last modified: November 1, 1996 - v1
Checksum: 0AE9B0B62B2BCA4A

BLAST ▾ [GO](#)

```

10   20   30   40   50
MSLAGGRAPK KTAGNRLSGL LEAEEEDDEFY QTTYGQFTEE SGDDEYQQDQ
 60   70   80   90   100
SDTEDEVDSQ FQIEGEDEPS SGDGEAEPEPK KRRVVVTKAYK EPLKSLSRKK
110  120  130  140  150
VNTPAGSSQK AREEKFALLPEL ELQDGGSDDSK KSMRSQTAEH TRQTFLVQE
160  170  180  190  200
RQQQSERRSK PHECERFLQEE ELLEREAKITKE ELMRSLSLETY ERLEACKEQ
210  220  230  240  250
VRKKEKRCGP K IITTHSTVTFV LVGEPGPKKEE VNUIEDLQPA PSVSVLTHA
260  270  280  290  300
GTGPNPFAK CSETFITFSD DATTEENFPQ GRFPKVVFRE VCFVTHRFAL
310  320  330  340  350
YRDPTVTPY ATARAFKIIIR EAKKYKTIATN GLPPTIASLG PGFFFFPEPLP
360
GSGFRALRKQK 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

Note: No experimental confirmation available.
[Show ▾](#)

Computationally mapped potential isoform sequences¹

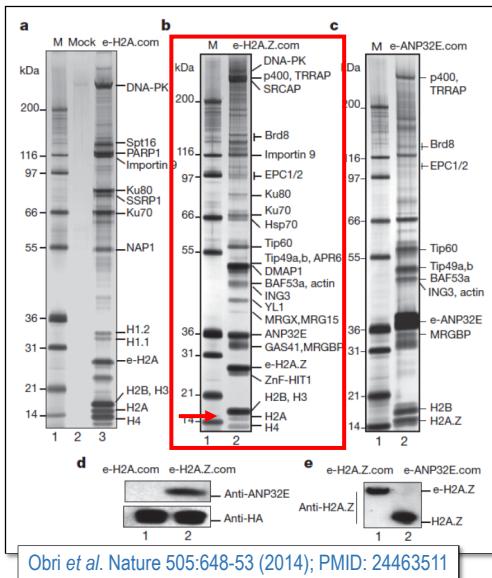
There is 1 potential isoform mapped to this entry. [BLAST](#) [Align](#) [Show all](#) [Add to basket](#)

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

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:
 EMBL: X83549 Genomic DNA Translation: CAA06539.1
 GenBank: X57138 Genomic DNA Translation: CAA40417.1
 DDBJ: L19778 mRNA Translation: AAC24466.1
AV131987 Genomic DNA Translation: AAC59966.1
AV131989 Genomic DNA Translation: AAN59970.1
AV131991 Genomic DNA Translation: AAN59972.1
AV131992 Genomic DNA Translation: AAN59973.1
AV131993 Genomic DNA Translation: AAN59974.1
Z98744 Genomic DNA No translation available.
AL009179 Genomic DNA translation available.
AL021807 Genomic DNA No translation available.
BC016200 mRNA Translation: AA06539.1
BC106300 mRNA Translation: AA06539.1
BC104199 mRNA Translation: AA04200.1
BC104190 mRNA Translation: AA04199.1
BC105120 mRNA Translation: AA05130.1
BC112072 mRNA Translation: AA112073.1
BC112254 mRNA Translation: AA112255.1
BC112256 mRNA Translation: AA112257.1

CCDS^j: CDD54619.1
CDD54626.1
CDD54632.1
CDD54634.1
CDD54639.1

PIR^k: B56624 HSUUA1

RefSeq^l: 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^m: Hs.134999
Hs.233568
Hs.51011
Hs.534035
Hs.734715

Gene duplication !!!

PICR: Protein Identifier Cross Reference

R.I.P.
As of May 2019

PICR Protein Identifier Cross-Reference

[Home](#) [User Guide](#) [Implementation](#) [Webservice](#) [RESTful](#) [Contact Us](#)

Input Data

Accessions Exact Sequence Sequence Similarity search [BLAST]

Browsing: No file selected.

Enter one or more valid protein accessions (one per line) and click on submit. Alternatively, enter one or more protein sequences in FASTA format and select "Protein Sequence" as input type. You may also upload a file that contains either protein identifiers (one per line) or protein sequences in FASTA format. The file must be less than 2MB in size.

Click here to load example data to try out PICR: [Load example data](#)

Output Parameters

View results as:
 Simple HTML Detailed HTML CSV XLS

Input Parameters

Limit by species:
All species

Return only active mappings

Mapping Databases

Include mappings to the following databases (if available)

UniProt 'best guess'	<input checked="" type="checkbox"/>	<input type="checkbox"/>
SwissProt	<input checked="" type="checkbox"/>	<input type="checkbox"/>
TiEMBL	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Ensembl Genomes	<input checked="" type="checkbox"/>	<input type="checkbox"/>
EPO	<input checked="" type="checkbox"/>	<input type="checkbox"/>
H Inv	<input checked="" type="checkbox"/>	<input type="checkbox"/>
PDB	<input checked="" type="checkbox"/>	<input type="checkbox"/>
PRF	<input checked="" type="checkbox"/>	<input type="checkbox"/>
SGD	<input checked="" type="checkbox"/>	<input type="checkbox"/>
TROME	<input checked="" type="checkbox"/>	<input type="checkbox"/>
UniParc	<input checked="" type="checkbox"/>	<input type="checkbox"/>
USPTO	<input checked="" type="checkbox"/>	<input type="checkbox"/>

URL: <http://www.ebi.ac.uk/Tools/picr/>

- Hosted by EBI (Hinxton, UK)
- Cross-references between close to a 100 of databases (UniProt, Refseq, Ensembl, model organisms, PDB, ...)
- Stable 'over time' and over 'interface'
- Access in a batch mode and as a service (SOAP and RESTful)
- Results presented in either tabular or HTML/XML format

UniProtKB: Identifier Mapping

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

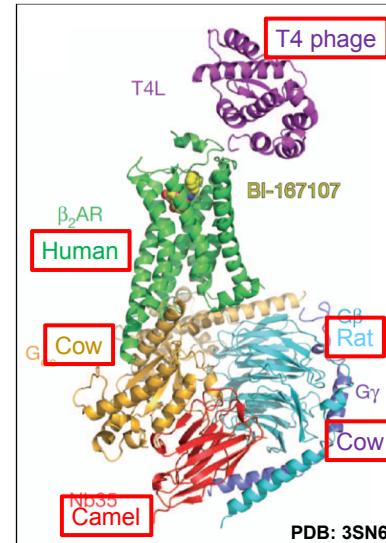
- Hosted by SIB (Geneva, Switzerland)
- Cross-references between UniProtKB records and about 80 databases (Refseq, Ensembl, model organisms, PDB, ...)
- Stable ‘over time’ and over ‘interface’
- Access Results presented as 2 column table
- in a batch mode and as a service (RESTful)

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:

- Identifier provided by the authors
- Histone tails (as separate entities)

Polyproteins

- Viral proteins
- Ubiquitin

Protein Identification: Polyproteins

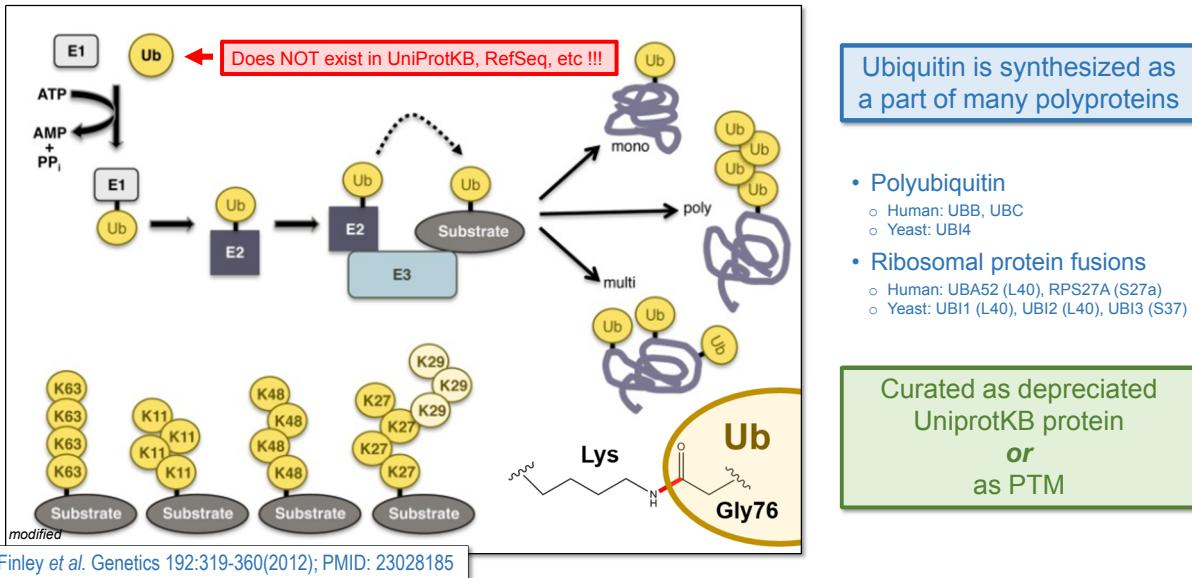
UniprotKB: P04585

PTM / Processing

Molecule processing

Feature key	Position(s)	Description	Actions	Graphical view	Length
Initiator methionine ⁱ		Removed; by host			
Chain ⁱ (PRO_000023620)	2 – 1435	Gag-Pol polyprotein			1434
Chain ⁱ (PRO_0000042439)	2 – 132	Matrix protein p17			131
Chain ⁱ (PRO_0000042440)	133 – 363	Capsid protein p24			
Peptide ⁱ (PRO_0000042441)	364 – 377	Spacer peptide 1			
Chain ⁱ (PRO_0000042442)	378 – 432	Nucleocapsid protein p7			55
Peptide ⁱ (PRO_0000246716)	433 – 440	Transframe peptide			8
Chain ⁱ (PRO_0000042443)	441 – 488	p6-pol			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			288

Protein Identification: Ubiquitin



Interaction Experiment Record

Interaction Record

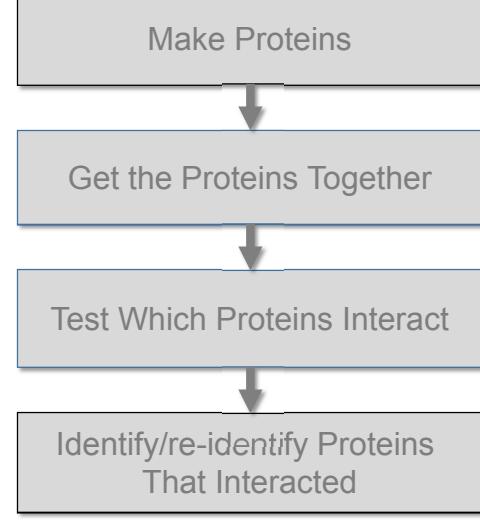
Interaction Experiment

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)



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

The screenshot shows a controlled vocabulary interface for molecular interactions. On the left, a tree view lists various feature types and experimental features. On the right, a detailed view of a specific term is shown:

Term info

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

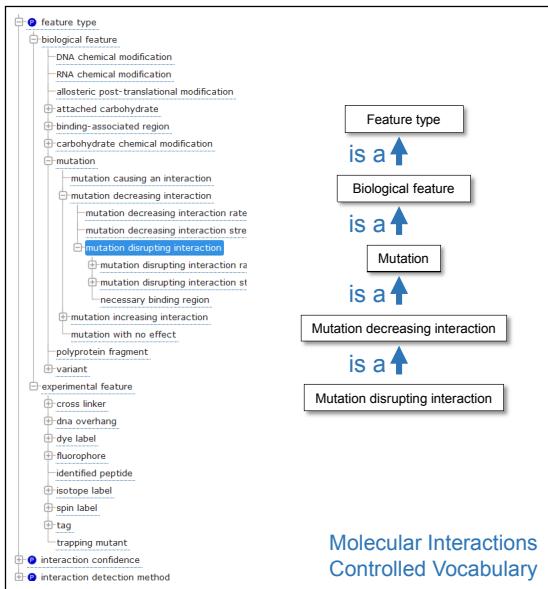
id: MI:0573

Molecular Interactions Controlled Vocabulary

Molecular Interactions (MI) Controlled Vocabulary

- 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

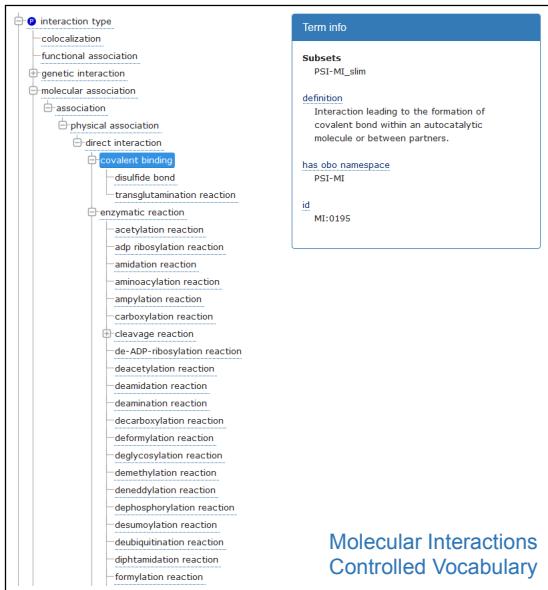
Controlled Vocabularies



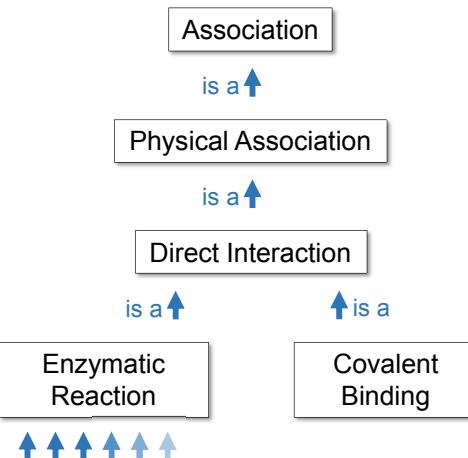
Molecular Interactions (MI) Controlled Vocabulary

- 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'

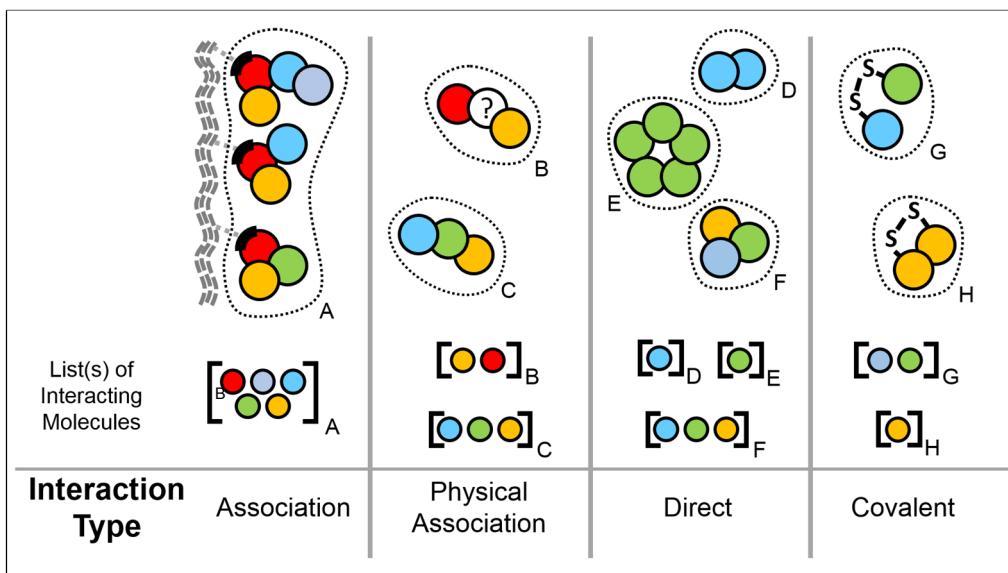
Controlled Vocabularies



Interaction Types



Interaction Types



Controlled Vocabularies

Molecular Interactions Controlled Vocabulary

- interaction type
 - cotocalization
 - functional association
 - genetic interaction
 - molecular association
 - association
 - physical association
 - direct interaction
 - covalent binding
 - disulfide bond
 - transglutamination reaction
 - enzymatic reaction
 - acetylation reaction
 - adp ribosylation reaction
 - amidation reaction
 - aminoacylation reaction
 - amplification reaction
 - carboxylation reaction
 - cleavage reaction
 - de-ADP-ribosylation reaction
 - deacetylation reaction
 - deamidation reaction
 - deamination reaction
 - decarboxylation reaction
 - deformylation reaction
 - deglycosylation reaction
 - demethylation reaction
 - deneddylation reaction
 - dephosphorylation reaction
 - desumoylation reaction
 - deubiquitination reaction
 - diphthamidation reaction
 - formylation reaction

Interaction Types

Association



Physical Association



Direct Interaction



Enzymatic Reaction



Covalent Binding



Controlled Vocabularies

Molecular Interactions Controlled Vocabulary

Term info

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 (MI) Controlled Vocabulary

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

Ontology Lookup Service

Welcome to the EMBL-EBI Ontology Lookup Service.

Search OLS: Examples: diabetes, GO_0098743

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, view their metadata 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, Zemina and Webulous services. Oxo provides cross-ontology mappings between terms from different ontologies. Zemina is a service to analyse mapping data to ontologies in OLS and Webulous is a tool for building ontologies from spreadsheets.

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

Data Content: Updated 06 Mar 2019 11:11

- o 224 ontologies
- o 5,415,430 terms
- o 21,348 properties
- o 480,657 individuals

Tweets by @EBIOLS

EBSIOT OLS
SERVICE UPDATE: OLS is having a few issues at the moment. Dec 6, 2018

EBSIOT OLS
SERVICE UPDATE: OLS is back to normal.

Controlled Vocabularies

Protein Modifications (MOD) Controlled Vocabulary

Term info

database cross reference

- o Formabs: (C20 H 38 N 12 O 16 P 2)
- o DiffKvp: (783.54)
- o MassAvg: (920.68)
- o Origin: (V)
- o MassMono: (920.200396)
- o DiffFormula: (C27 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(τα)-[βαλφα-FAD]-histidine

Description (full name): UniMod Flavin adenine dinucleotide

Systematic name from RESID: (S)-2-amino-3-[βαλφα riboflavin 5'-trihydrogen diphosphate] 5'-5'-ester with adenosineimidazol-4-ylpropanoic acid

Short label curated by PSI-MOD: NTBaFADHis

Protein feature description from UniProtKB: MOD_RES_Tele-βαλφα-FAD histidine

Alternate name from RESID: βαλφα-(N(ερσιν)-histidyl)FAD

Protein Modifications (MOD) Controlled Vocabulary

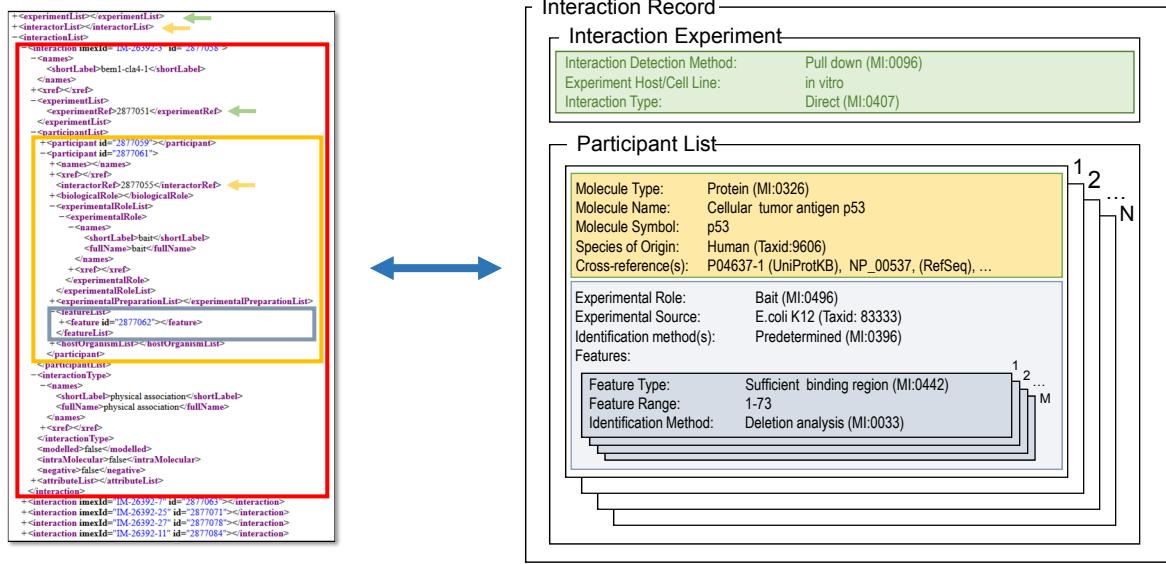
URL: <https://www.ebi.ac.uk/ols/ontologies/mod>

- 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
 - ...

Protein Modifications Controlled Vocabulary

Interaction Database Records Formats

Interaction Record Formats PSI-MI XML (MIF) format



MIF/MI Definitions

HUPO translating the code of life [JOIN HUPO](#)

ABOUT HUPO

The Human Proteome Organization (HUPO) is an international scientific organization representing and promoting proteomics through international cooperation and collaborations by fostering the development of new technologies, techniques and training.

HUPO MISSION STATEMENT

To define and promote proteomics through international cooperation and collaborations by fostering the development of new technologies, techniques and training to better understand human disease.

Objectives

- Foster global collaboration in major proteomics projects by gathering leading international laboratories in life sciences, bioinformatics, mass spectrometry, systems biology, pathology, and medicine;
- Become the point of contact for proteomics research and commercialization activities worldwide;
- Support large-scale proteomics projects that are aimed at:
 - A mechanistic understanding of fundamental biological processes (often using model

PROTEOMICS STANDARDS INITIATIVE

Website: <http://www.psidev.info/>

Overview of Project

The HUPO Proteomics Standards Initiative (PSI) defines community standards for data representation in proteomics to facilitate data comparison, exchange and verification.

PSI Governance

Andy Jones, Chair
Eric Deutsch, Co-chair
Sandra Orchard, Co-chair

The main organizational unit of the Proteomics Standards Initiative is the work group. Currently, there are the following work groups:

- CompMS
- Early Career Researcher (ECR) Initiative
- Human Antibody Initiative
- Human Proteome Project
- Initiative on MultiOrganism Proteomes
- Pathology Pillar
- Proteomics Standards Initiative**
- Clinical Proteome Tumor Analysis Consortium (CPTAC)

Interaction Record Formats

PSI-MI XML (MIF) format

Good	Bad
<ul style="list-style-type: none"> Stable <ul style="list-style-type: none"> MIF 2.5 (2007) MIF 3.0 (2018; mostly backward-compatible) Database-neutral <ul style="list-style-type: none"> Developed by HUPO-PSI Widely used by data providers <ul style="list-style-type: none"> IMEx Consortium (DIP, IntAct, MINT,...) – native BioGRID – export Expressive enough to describe most of the interaction experiments <ul style="list-style-type: none"> Multi-protein interactions Protein features (PTMs, mutations, ...) Multiple experimental methods/protein Not limited to proteins <ul style="list-style-type: none"> Nucleic Acids Small Molecules 	<ul style="list-style-type: none"> Overly verbose <ul style="list-style-type: none"> Redundant open/close tags Several levels of nested elements But compresses quite well – 20x is not that rare Does not fit ‘Excel spreadsheet’ paradigm <ul style="list-style-type: none"> Not too surprising – interaction data is NOT tabular Limited set of good quality user-side tools <ul style="list-style-type: none"> Java JAMI is very versatile but complicated Limited support for reporting experiment ambiguities <ul style="list-style-type: none"> MIF 3.0 provides some support but curation lags No support for oligo-/poly-saccharides <ul style="list-style-type: none"> Limited by the current state of nomenclature No active curation (to my knowledge) XML is considered to be hard to work with

Interaction Record Formats

PSI-MI tab-delimited (MITAB) format

Good

- Easy to read into a spreadsheet
- Supported by third-party libraries

Bad

- Applicable only to binary interactions
 - Cannot handle multi-protein complexes
- Many columns can be multi-valued
 - Requires custom parsing routines
- Some information originally available in MIF format cannot be stored as MITAB
 - The format is lossy – is, essentially, impossible, to restore fully-featured MIF record from its MITAB representation
- Less stable than MIF
 - MITAB 2.5, 2.7, 2.8 (MIF 2.5 derivatives)
 - MITAB 3.0 (3.0 derivative)

Interaction Record Formats

BioGRID tab & complex tab formats

Good

- Easy to read into a spreadsheet

Bad

- Separate format for binary and multi-protein complexes
- Supports only protein-protein interactions
- Identifies proteins by gene identifiers
 - It works only for organisms that do not splice
- Provides less information than PSI-MI MIF
 - This is because BioGRID extracts less information about interactions that IMEx Consortium databases
- Uses simplified, non-standard CV terms
- See also MITAB deficiencies

Interaction Record Formats

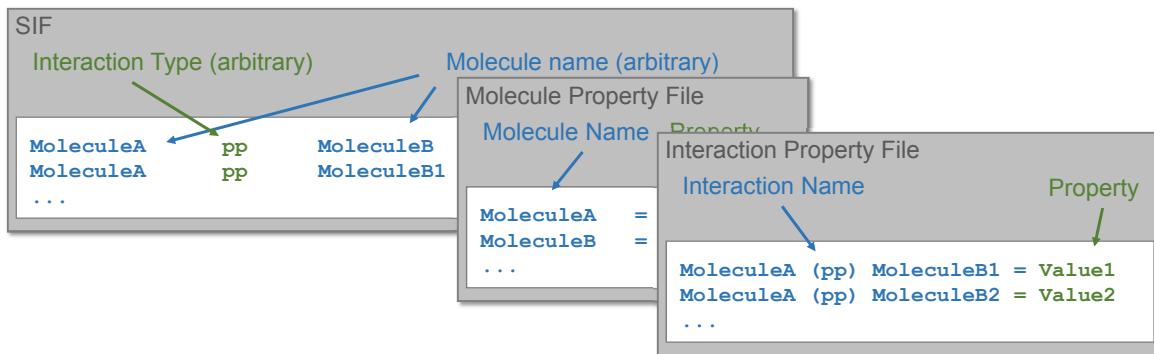
Cytoscape SIF format

Good

- Native Cytoscape format
- Simple/easy to prepare (spreadsheet will work)
- Not limited to biological data

Bad

- Only binary interactions
- Must be combined with information from additional files in order to provide more detailed information



Interaction Data Resources

Interaction Data Resources

Primary Data Providers

The screenshot shows the IMEx (International Molecular Exchange Consortium) website. 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 a "Use as input: UniProt Accs, Gene names, Publication IDs" field. The main content area has a section titled "IMEx data" which contains a bulleted list of data characteristics. To the right of this is a "Citing IMEx" section featuring a paper by Orchard, S., et al. at the International Molecular Exchange (IMEx) consortium. The bottom of the page features sections for "IMEx Partners" and "IMEx Observers", each listing logos for various databases and 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.unroma2.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>

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

Protein	Functional Partner	Description	Score
LAST1	Protein of protein LAST1 Actin assembly factor C-terminal RICA domain activates Arp2/3 complex-mediated nucleation of branched actin filaments.	0.993	
VRP1	Vesopin, proline-rich actin-associated protein; involved in cytoskeletal organization and cytokinesis; promotes actin nucleation.	0.988	
MYO5	ATPase superfamily, Myosin family (1279 aa).	0.963	

• Interaction resources

- STRING
- GeneMania
- HIPPIE
- HINT
- APID
- ...

• General resources

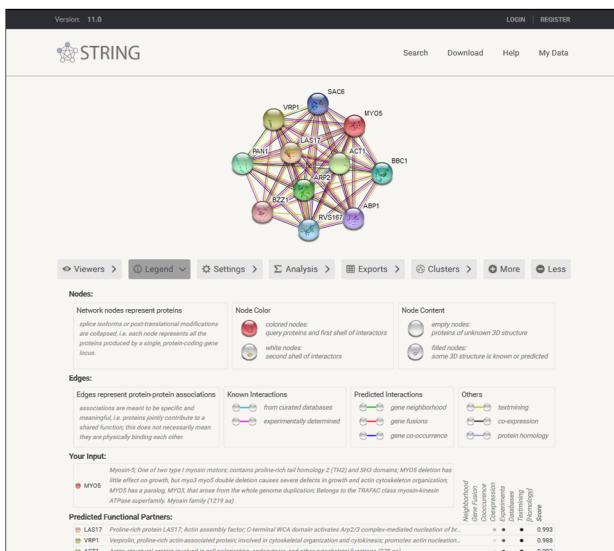
- UniprotKB
- Reactome
- ...

• 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)
 - Cooccurrence in genomes
 - Genome neighborhood
 - Gene fusion (Rosetta Stone)
 - Text Mining
 - Cooccurrence in publications
 - Molecular interactions
 - IMEx Consortium
 - BioGRID
 - ... and others
 - Pathway information
 - Reactome
 - KEGG
 - BioCyc
 - ... and others
 - Coexpression

Interaction Data Resources

PSI Common Query Interface (PSICQUIC) Servers

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

PSICQUIC Registry							
Name		Status	Interactions	Version	URLs	Description	Restricted
itRefIndex	●	2.338.337	2.13.14	SOAP: http://refindex.vib.be/webservices/psicquic REST: http://refindex.vib.be/webservices/current/search/ REST example			NO protein-protein imported bytaxa expansion evidence
BioGrid	●	1.515.281	1.3.14	SOAP: http://tyerrest.tyrellab.com:8803/psicquic/webservices/ REST: http://tyerrest.tyrellab.com:8803/psicquic/webservices/ current/search/ REST example			NO protein-protein imported curated evidence spoke expansion evidence
BindingDB	●	1.011.029	v1.3	SOAP: http://bindingdb.org/psicquic/webservices/ REST: http://bindingdb.org/psicquic/webservices/psicquic/ current/search/ REST example			NO smallmolecule: protein imported: curated evidence spoke expansion evidence observed
I2D	●	817.915	1.1.6	SOAP: http://ophid.utoronto.ca/psicquic/webservices/ REST: http://ophid.utoronto.ca/psicquic/webservices/current/ search/ REST example			NO protein-protein imported: curated evidence
IMEx	●	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 verified experimental interactions annotated by members of the IMEx Consortium		NO protein-protein smallmolecule: protein nucleicacid-protein imported: 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