IntAct Annotation Rules (February 2006).

Index

1. IntAct data model

2. Editors
2.1. IntAct AC
2.2. Short Label
2.3. Full Name
2.4. Editor Search
2.5. Locked entry
2.6. Submit, Save&Continue, clone, cancel, delete
2.6.1. Submit
2.6.2. Save&Continue
2.6.3. Clone
2.6.4. Cancel
2.6.5. Delete
2.6.6. Modifying existing entries

3. IntAct data release

4. Editor-Experiment
4.1. Experiment AC
4.2. Experiment Short Label
4.3. Experiment Full Name
4.4. Host Organism
4.4.1. In vitro interactions
4.4.2. In vivo interactions
4.5. Interaction detection
4.5.1. Additional Controlled Vocabulary terms
4.6. Participant detection
4.6.1 Additional Controlled Vocabulary terms
4.7. Interactions from Experiment page of editor
4.8. Experiment Annotation
4.8.1. antibodies
4.8.2. caution

4.8.3. comment
4.8.4. complex-properties
4.8.5. confidence-mapping
4.8.6. contact-comment
4.8.7. contact-email
4.8.8. copyright
4.8.9. data-processing
4.8.10. dataset
4.8.11. exp-modification
4.8.12. figure-legend
4.8.13. library
4.8.14. negative
4.8.15. on-hold
4.8.16. remark-internal
4.8.17. submitted
4.8.18. uniprot-cc-note
4.8.19. uniprot-dr-export
4.9. crossreferences and crossreference qualifiers
4.9.1. primary-reference
4.9.2. method
4.9.3. see-also
4.9.4. target-species

5. Editor-Interaction
5.1. Interactor AC
5.2. Short Label
5.3. Full Name
5.4. Kd
5.5. Interaction Type
5.6. Organism
5.7. Experiments
5.8. Proteins
5.8.1. 'Short Label', 'SP AC', 'organism' and 'Intact AC'
5.8.2. Protein accession numbers
5.8.2.1. How to find a protein accession number
5.8.2.2. Problems mapping accession numbers
5.8.2.3. Incomplete protein sequences/ protein without UniProtKB accession number
5.8.2.4. UniProtKB entries that contain multiple protein molecules
5.8.2.5. Natural peptides
5.8.2.6. Bi-cistronic expression
5.8.2.7. Synthetic peptides
5.8.2.8. Identification of proteins by sequence similarity
5.8.2.9. Proteins of unknown species
5.9. Protein roles
5.9.1. bait/prey
5.9.2. enzyme/enzyme target

5.9.3. neutral component
5.9.4. self
5.9.5. electron acceptor/electron donor
5.9.6. inhibited/inhibitor
5.9.7. acceptor fluorophore/donor fluorophore
5.9.8. unspecified
5.9.9. complex – made obsolete
5.10. ExpressedIn
5.10.1. Heterologous protein expression
5.10.2. *In vitro* transcribed and translated proteins
5.10.3. Synthetic peptides
5.11. Stoichiometry
5.12. Interaction annotation
5.12.1. agonist
5.12.2. antagonist
5.12.3. author-confidence
5.12.4. caution
5.12.5. comment
5.12.6. complex-properties
5.12.7. delivery-method
5.12.8. disease
5.12.9. expression-level
5.12.10. figure-legend
5.12.11. function
5.12.12. inhibition
5.12.13. isoform-comment
5.12.14. kinetics
5.12.15. negative
5.12.16. pathway
5.12.17. remark-internal
5.12.18. resulting ptm
5.12.19. sample-source
5.12.20. stimulation
5.12.21. uniprot-cc-note
5.12.22. url
5.12.23. 3d-structure
5.12.24. 3d-r-factor
5.12.25. 3d-resolution
5.13. Crossreferences and crossreference qualifiers
5.13.1. identity
5.13.2. see-also
5.13.3. GO crossreferences
5.13.3.1. component
5.13.3.2. process
5.13.3.3. function

6. Editor-Feature
6.1. The 6 fields
6.1.1. Short Label
6.1.2. Full name
6.1.3. Feature type
6.1.4. Feature Identification
6.1.5. Ranges
6.1.6. Crossreferences
6.2. Various cases
6.2.1. Molecule is tagged
6.2.2. Molecule is radio-labelled
6.2.3. Molecule is post-translationally modified (PTM)
6.2.4. Molecule is a fragment, truncation or deletion construct
6.2.5. Molecule contains a mutation
6.2.6. Molecule contains a Hotspot
6.2.7. Mutation editor
6.2.8. Mutation which affect the rate of an interaction
6.2.9. Residue mapping and linking
6.3. Linking unlinking and deleting features

7. Editor-Biosource
7.1. BioSource/Organism
7.1.1. Short Label
7.1.2. Full Name
7.1.3. Crossreference
7.2. BioSource/Tissue type
7.3. BioSource/CellType
7.3.1. Editor-CVCellType
7.3.1.1. Biosource with cell line
7.3.1.2. Cell lines without CABRI reference but with an ATCC number
7.3.1.3. Cell lines with a PubMed reference
7.3.1.4. Non-Standard cell lines
7.3.1.5. Reclassified cell lines
7.4. BioSource Compartment

8. Editor-Protein
8.1. Short label
8.2. Full Name
8.3. Source
8.3.1. Chemical synthesis
8.3.2. Chimeric protein
8.3.3. UniProtKB peptide chains
8.4. Sequence
8.5. Annotation
8.5.1. caution
8.5.2. comment

8.5.3. function
8.5.4. isoform-comment
8.5.5. copyright
8.5.6. remark-internal
8.5.7. submitted
8.5.8. on-hold
8.5.9. no-uniprot-update
8.6. Crossreference

9. Special Cases
9.1. Annotation of protein complexes
9.1.1. Complex definition
9.1.2. Partially characterised complexes
9.1.3. Organism
9.1.4. Annotation of protein complexes using GO terms
9.1.5. Additional literature references
9.2. Importing interactions from other databases
9.3. Cell surface protein-protein interactions
9.4. Interaction detected by Far Western
9.5. DNA-protein interactions
9.6. Data not shown

10. Curation checking procedure

11. Talisman

12. PubMed Lookup

13. Bibliography

1. IntAct data model

The IntAct model has:
one publication with one to many experiments,
one experiment with one to many interactions,
one interaction with one to many interactors (proteins/peptides participating in the interaction) and
one protein with none to many features (features describe the component of the interaction- e.g. region of protein used, PTM present on the protein involved in the interaction).

2. Editors

IntAct data entry is made using editors. There are separate editors for recording Experiments, Interactions, BioSource, Crossreference Databases, Annotation Topics, and

Non-UniProtKB Interactors. The editor to be used should be selected from the drop down list. The editor is available for internal curation only and needs a login and password.

There are 2 database instances: ZPRO and DOO3 both are accessible from:

http://www3.ebi.ac.uk/%7Esp/intern/projects/intact/local/doc/

- The production editor is used to add and edit entries in production database (ZPRO).
- The test editor uses the test database (D003). The test editor is used to test new updates of the editor before putting them on production editor. The test editor is also used for training new IntAct curators. If you are not sure about your entries, you can try entering these on the test editor first. Try not to add and delete too many times on the production editor as it will use up the EBI-numbers.

Unique IntAct AC and Short Label are mandatory in the Editor form. All the mandatory fields are green on each of the Editor forms. Common features of the editor forms are:

2.1. IntAct AC

Every IntAct object for example experiment, interaction, interactor, feature, annotation, biosource, CV object etc. gets an IntAct accession (AC) number when it is created. On being created, this entry can be visualised on the database and will have to be physically deleted. All objects in an editor page are allocated an Intact AC number on pressing 'save & continue' or 'submit' button. On pressing 'save & continue' you remain in the same entry after saving whereas you save and exit the specific editor page on submitting the entry. The Intact AC is hyperlinked to the IntAct Search from the editor pages.
An IntAct accession numbers may be cross-referenced on an interaction with a Reference Qualifier see-also when these are related interactions. In the case of negative experiments and interactions it is necessary to have the IntAct accession numbers cross-referenced with the Reference Qualifier 'see-also' (section 4.8.12). IntAct accession will be cross-referenced with the cross reference qualifier 'intact-secondary' when an experiment interaction or interactor has been on the public site but has subsequently been assigned a different accession number. This will ensure that the correct record/s is/are pulled up using both the former and the more current accession numbers.

2.2. Short Label

Short Labels are used for experiments, interactions and features. These use lowercase letters, numbers and special characters such as underscore (_), hyphen (-) and space. All characters that are not in [a-z] are replaced by _ (underscore). There is a character limit of 20 on the Short Label. The Short Label is linked to the editor page for that entry.

2.3. Full Name

The character limit on the full name is 250 characters and there is no restriction on the characters used - standard keyboard characters are allowed.

2.4. Editor Search

A search can be performed within the editor using either the Short Label or the IntAct AC number (EBI-number); wild card (*) can be used in the search. Choosing a term in the drop down list of controlled vocabulary used in IntAct and clicking on the heading will give you search results for the controlled vocabulary.

2.5. Locked entry

When a curator is editing an entry, it is displayed as 'locked' with the 'LOGIN' for the curator using the entry displayed in orange. Sometimes you may have locked it by accident. To unlock it, go back to the entry and press 'Cancel' or 'Submit' if you have made any changes.

2.6. Submit, Save & Continue, Clone, Cancel, Delete

2.6.1 Submit:

When making a fresh entry the Submit button commits an entry and gives it an IntAct AC. When modifying an entry the Submit button commits the changes made. On pressing the Submit button you get leave the entry.

2.6.2. Save & Continue:

This acts similar to the submit button but keeps you within the entry. This is useful when you may need to wait on finding out particulars of the entry.

2.6.3. Clone:

This is used to clone an entry. On cloning an entry the next available Short Label number is automatically assigned to the entry. The cloned entry does not have any interactions connected to it. You will need to check at least the Interaction Detection, Participant Detection, and Host-Organism on this entry. A cloned interaction will not have an experiment attached to it. An incremental Short Label will be automatically generated from the gene names of the cloned entry. Features attached to interactor are also cloned. An entry in its entirety should not be duplicated in IntAct. When you press the clone button the changes made to the parent entry are submitted and you go into the cloned entry. You cannot Clone Features and proteins made in protein editor.

2.6.4. Cancel:

This button gets you out of the entry without saving the changes made.

2.6.5. Delete:

This button is used to delete an entry. Once an entry has been made it has to be physically deleted from the database if not required.

2.6.6. Modifying existing entries

An entry made in the database can be changed by using the editors and then pressing the 'submit' button after the changes have been made. If an entry has been made public and the changes made are minor these can be done on the record itself. The IntAct objects that can be modified in this manner are experiments, interactions, interactors and components. If the changes made involve the transfer of contents from an entry to a new IntAct object with the original IntAct object being deleted then the accession of the old IntAct object will be crossreferenced in the new IntAct object using the Reference Qualifier - intact-secondary.The IntAct objects which may be modified in this manner are experiments, interactions and interactors.

3. IntAct data release

All interactions from each publication should be incorporated in their entirety before release of the data on the public server. Before curating a paper please check its status using the PubMed Lookup tool http://www3.ebi.ac.uk/%7Esp/intern/projects/intact/local/doc/. IntAct release is done at the beginning of each month. If you have not finished the entries within the publication please annotate the experiments within the publication 'on-hold' (section 4.8.12), when the e-mail about the release is circulated. Data from 'on-hold' and unchecked entries will not be released.

During the release, the data on the production database is first transferred onto the test server. Hence, while the release is in progress, you should not use the test editor. This data is then transferred to the public database (IWEB). As the back-up is on test server, the production database can be used again by curators before the release is completed. E-mail notification of the availability of the production editor database and test editor database will be sent.

4. Editor-Experiment.

This form records the experimental method used to demonstrate the interaction. For a given publication, where many interactions have been demonstrated using the same experimental method, a generic experiment is created for that interaction set. This means that one experiment may have many interactions linked with it.

Note: The converse is not allowed i.e. one interaction may not belong to multiple experiments.
Experiments that demonstrate that an interaction does not occur can be entered in IntAct under special conditions. See 'Annotation Topics 'Negative' section 4.8.11 below.

Case I: When two different interactions are shown using a single experimental technique and the primary reference is the same, these should be linked to a single experiment.
Example: experiment EBI-74060 has two interactions associated with it: EBI-74066 and EBI-74062.

Case 2: The same interaction is demonstrated by two different experimental techniques. This must be entered as two experiments and two interactions as shown below.

Interaction AC. no.	Experiment AC. no.	Bait UniProtKB AC	Prey UniProtKB AC	Interaction detection method
EBI-78422	EBI-78402	P05067	Q9UQF2	two hybrid
EBI-78430	EBI-80340	P05067	Q9UQF2	pull-down

Case 3: If the same interaction using the same experimental technique has been shown in two different papers these should be entered as two different experiments differing in the primary crossreference ID linked to different interactions.

An experiment may be created by entering a pubmed ID into the 'Pubmed ID' field and clicking 'Auto-complete'. The 'Short Label', 'Full Name', 'Annotations' -'author-list' and 'contact-email' and 'Crossreferences' - database 'pubmed' Primary ID as stated in the Pubmed ID field and reference qualifier 'primary-reference' should autofill to the rules stated below. Any of these which are missing should be entered manually.

4.1. Experiment AC

AC number is assigned to the experiment when you click the 'save and continue' or 'submit' button. This is automatically assigned a unique number. Clicking on this number in the editor form or index page will connect to the search. Search will show you the entry as it will appear on the public database. It does take about 15 minutes for the latest changes in the entry to appear.

4.2 Experiment Short Label

This is a mandatory field.
An experiment Short Label has the following format:
[first author last name]-[year of publication][optional char]-[integer]
Overall character limit is 20, hence, character distribution will be [11]-[4][1]-[2].
All characters that are not in [a-z] are replaced by _ (underscore).
The Short Label is lowercase.
The author names may need to be truncated to accommodate the character limit.
When an author has published more than one paper (different pubmed ID) in the **same year**, enter year a, b, c etc. as illustrated below. A a, b, c, d ... after the year is added based on the chronological order in which the data was entered in IntAct.

e.g. stargell-2003-1
stargell-2003a-1
stebbins_bo-1999-1 (author name is stebbins-boaz)

Note: Two different authors who have the same name cannot be distinguished between.

The experiment number or final integer will be added based on the chronological order of entering the data in IntAct.
If more than one primary-references are available on an IntAct experiment, we use the latest publication date, e.g. ito-2001-1.
EPUB: While entering the data an EPUB year may be used, however, once the article is published and gets a pubmed ID, the publication date is used to update entry. This will be done automatically. Example: trinczek-2003-1 is the title entered by the curator (epub year - 2003) this will be updated to trinczek-2004-1 (publication year - 2004).

4.3 Experiment Full Name

The full name is free text and the title of the publication being curated. There is a character limit of 250 characters on the Full Name after which the title will be truncated. In case the data is not from publication or yet to be published the Full Name will be filled in by the curator – make up a title which pertains to the topic of the paper.

4.4 Host organism

This is a mandatory field.
The Host Organism is the organism the experiment has been performed in.
The Host Organism (along with tissue or cell-line if relevant) is chosen from a drop down list of organisms. Please see the section 7 on Editor-BioSource for how to create the drop down list of 'Host organism'.
The 'Host Organism' is *in vitro* when the interaction has not been shown to occur *in vivo*.

4.4.1 *In vitro* interactions

'Host Organism' is *in vitro* when one of the interacting proteins is extracellular or has an extracellular domain that participates in the interaction and /or the interaction is shown to occur in a tube (or equivalent).

Example:
EBI-458044,

Note: Proteins are considered to be extracellular when the part of the protein taking part in the interaction is outside the cell membrane.
For a definition of extracellular see GO:0005576 .

The components of the interaction may be purified or semi-purified and each component could come from any of the sources below.

i) Purified proteins: The proteins may be expressed in a heterologous system example: E. coli, Baculovirus, yeast, insect, mammalian or/and plant cells and purified from these or from its natural sources.
ii) Fractionated proteins: Proteins are in a fractionated cell extract from cell lines, embryos or tissues. Example: cell extract, nuclear extract, vesicle fraction, Golgi membranes, membrane fraction.
ii) Proteins are from more than one cell lysate.
iv) Secreted protein
v) *In vitro* transcribed and translated protein.
vi) Synthetic peptide
vii) Purified peptide

Examples of ***in vitro*** interactions:

1. Interaction between a cell-surface receptor and a labelled (radio-labelled, biotinylated) purified protein ligand. This could be demonstrated example: by *in vitro* cross linking [2] See examples below.

2. Two cells each of which express cell-surface molecules and are demonstrated to interact example: by FACS. See examples below.

3. One cell expressing two or more cell-surface molecules which are demonstrated to interact (Example: by cross-linking).

Details of heterologous protein expression for each protein must be entered in the 'ExpressedIn' section 5.10 on the interaction page. Information whether the protein is purified or transfected and overexpressed should be stored in annotation topic 'exp-modification'.

4.4.2 *In vivo* interactions

All of the interacting proteins must be expressed **in a cell** or cell membrane; an exception would be micro-injected proteins.
The proteins may be modified (tags, promoters etc).
The cell may be subsequently lysed to enable access of an immuno-precipitating antibody or other detection agent.
The names of cell lines and tissue types used when the interaction is demonstrated to occur *in vivo* should be added in the experiment under "host organism" from the drop down list. To create a 'Host Organism' or 'BioSource' please refer to section 7 on Editor-BioSource. If the protein has been overexpressed this information should be stored in annotation topic 'comment'.

Example:
Human proteins A and B are expressed in Cos-7 cells and shown to interact by co-immunoprecipitation.

Experiment: Host Organism: Cos-7 cells (cerae-cos_7)
Note: In this experiment the host organism is *C. aethiops* (African green monkey) as the proteins were expressed in Cos cells and the complex was formed in Cos cells.

Example: EBI-458060

Examples of '*in vivo*' interactions:

1. One cell expressing two or more transmembrane proteins and the intracellular domains or transmembrane regions are demonstrated to interact (Example: by *in vivo* cross-linking [3] using a cross-linker that crosses the cell membrane).

2. Detection of protein-protein interactions in live cells using techniques such as FRET (fluroscence resonance energy transfer) or BRET bioluminescence resonance energy transfer [4].
3. Coimmunoprecipitation of proteins from an extract.

4.5. Interaction Detection

This is a mandatory field.
You choose the experimental method used to detect the interaction from a drop down list.
Example: co-immunoprecipitation, gel filtration, two-hybrid.
You can see the definitions used for the term by selecting the term in the box and clicking on the title (Interaction Detection). The search should give you details for the interaction detection technique.

Cases showing co-localisation by imaging techniques should be entered only if supported by other interaction detection techniques, since, these by themselves may not indicate physical interaction. In this case the 'InteractionType' chosen on the Interaction-Editor would be 'colocalization'. Instances where the interactors co-occur in the same subcellular compartment or distribution in various regions of the cell can also be annotated using the interaction type : colocalization. Example: EBI-398042

See controlled vocabulary (CV).
ftp://ftp.ebi.ac.uk/pub/databases/intact/current/cv/

Examples of experiments in the database containing some of the interaction detection techniques:

Method	**Experiment no.**
molecular sieving	EBI-530172
fps	EBI-371686
anti tag coimmunoprecipitation (in vitro)	EBI-349776
anti bait coimmunoprecipitation (in vivo)	EBI-472888

cosedimentation	EBI-457884
two-hybrid	EBI-1296
far-western	EBI-75286
fluorescence spectr	EBI-476810
facs	EBI-523390
pull down (pull down - in vitro)	EBI-457587
pull down (pull down - in vivo)	EBI-456973
protein crosslink	EBI-604584
itc	EBI-456897
spa	EBI-490047
x-ray	EBI-457207
fret	EBI-456897
nmr	EBI-458375
two hybrid	EBI-1296
dhfr reconstruction	EBI-448165
beta galactosidase	EBI-65752
yeast display	EBI-296471
phage display	EBI-444056
peptide array	EBI-371696
protein array	EBI-375112
fluroscence imaging	EBI-82041
spr	EBI-456914
chromatography	EBI-73621

Please use as specific a term as possible from the hierarchy of the controlled vocabulary. Example: If coimmunprecipitation was carried out using anti bait or tag antibodies, anti bait coimmunoprecipitation and anti tag coimmunoprecipitation should be chosen over coimmunoprecipitation respectively.

4.5.1. Additional controlled vocabulary terms

Controlled vocabularies evolve depending on the demand for terms and inability to find a suitable term to describe the experiments described by authors. Please forward request for controlled vocabulary terms to https://sourceforge.net/projects/psidev/. In cases where the controlled vocabulary term is pending give a description of the method used by the authors under annotation topic ‘exp-modification’ and a ‘remark-internal’- ‘Revisit when CV in place’. Under ‘Interaction Detection’ select ‘experimental’ from the drop down list.

4.6. Participant detection

See the Participant Detection controlled vocabulary**.**
ftp://ftp.ebi.ac.uk/pub/databases/intact/current/cv/
This is the method by which the participants have been identified. Example: protein sequencing, nucleotide sequencing, radiolabelled protein.
'Participant detection' describes the method used for detecting the participants of an interaction and may be used during or after the interaction has occurred. All neutral participants, the preys where roles of components are bait and prey, enzyme-targets where roles of the components are enzyme and enzyme-target etc.
Example:
In the case of interaction detection by yeast two hybrid:
If the yeast cells have been transfected with specific constructs and these have not been resequenced after scoring for interaction use the participant detection method 'predetermined'.
If a specific construct has been resequenced after scoring for positive interaction, use 'nucleotide sequence identification' or appropriate child term as 'participant detection'.
For library screening the participant detection is usually 'nucleotide sequence identification' or an appropriate child term.
If a combination of techniques has been used, select the technique which gives the most information and describe the other techniques, and the proteins identified by these, in the annotation topic 'exp-modification'.

4.6.1 Additional Controlled Vocabulary terms

Example: Electrophoretic methods- SDS PAGE, non-denaturing gel electrophoresis (heterodimer experiments), band shift (two proteins shown to interact where the interaction is DNA-binding dependent).
This will be dealt with using the following remark until the electrophoretic method CVs are agreed and in place.
Remark: Revisit when Electrophoretic Method in place.

If you need a new participant-detection term forward request for new controlled vocabulary terms to https://sourceforge.net/projects/psidev/. In cases where the controlled vocabulary term is pending give a description of the method used by the authors under annotation topic 'exp-modification' and a 'remark-internal'- 'Revisit when CV in place.'

4.7 Interactions from Experiment page of editor

On linking the experiment to interactions, the interactions become visible on the Editor-Experiment under 'Interactions' and can then be accessed for editing and deleting from the experiment page. The deletion of the interaction from this page however will only result in unlinking of the experiment and interaction. You will need to go to the Editor-Interaction to delete the interaction from database. You can add the unlinked interaction to a different experiment through the Editor-Interaction page. Alternatively you can also search an interaction and add it to an experiment from the Editor-Experiment page.

4.8. Experiment Annotation

Annotations on the Editor-Experiment page relate to the experimental conditions only. Annotations related to individual proteins should be entered on the Editor-Interaction pages. Each topic has a free text box associated with it and additional information may be included in the free text box. The following annotation topics may be used while curating an experiment:

4.8.1. antibodies:

This annotation will list the specific antibodies used in the interaction detection or participant detection methods used. This topic is annotated on the experiment.

4.8.2. caution:

This is used for warning about possible errors in experiments or for specifying grounds for confusion. This can also be used in experiment were the author has expressed misgivings about a technique while comparing with another described in the same or different paper. Some examples are:

Experiment AC	Annotation Description
EBI-1363	Non-specific anti-phosphoTyr used to identify binding partners
EBI-80856	Cross-reacting human antibodies used to precipitate proteins from thymus extracts.
EBI-526206	There appears to be a discrepancy between the results described within the text (page 762) and figure 5 of this report, concerning interaction between F box proteins and Cullin proteins. The data relating to these protein-protein interactions was therefore not included in this entry.

4.8.3. comment:

'comment' is free text; this annotation topic can be used to describe additional information which cannot fit under other annotation topics. It is desirable that the comments are restricted to as few as possible and are complete sentences.

4.8.4. complex-properties:

Information on complexes being annotated in interaction can be entered here. This topic is available for annotation on Editor-Interaction page as well and the interaction page annotation topic should be preferentially used. Some examples of annotation are:

Experiment AC	Annotation Description
EBI-73668	Two complexes of different composition detected in this experiment,

	one complex containing TBP and one without.
EBI-73751	The RSC complex comprises 15 subunits. Only three of them were identified in this paper.

4.8.5. confidence-mapping:

This annotation topic gives a description of the method/s used by the author for confidence mapping the interactions attached to this experiment.
In large-scale experiments where the authors of the paper have assigned confidence values to the interactions, the experiment must contain an explanation of the author's definition of the confidence values. This must be written in the experiment annotation topic 'confidence-mapping'.
In addition, the author confidence values that are suitable for export to UniProtKB must be added in the experiment annotation topic 'uniprot-dr-export'
Example: In experiment EBI-332598 the authors have assigned three confidence values to the interactions: Core-1, Core-2 and Non-core. The authors have designated the Core-1 and Core-2 as being high confidence interactions, hence only the interactions with the confidence values Core-1 and Core-2 should be exported to UniProtKB.
The annotation of this experiment is as follows:

Experiment AC	**Topic**	**Description**
EBI-332598	confidence-mapping	The authors have assigned the following confidence values to the interactions: Core-1: found at least three times independently and the AD-Y junction is in frame. Core-2: found less than three times independently, retest positive, AD-Y junction is in frame. Non-core: all other interactions.
EBI-332598	uniprot-dr-export	CORE_1
EBI-332598	uniprot-dr-export	CORE_2

4.8.6. contact-comment:

The original address given in the publication was john@doe.com. This address is now invalid and has been replaced by the one given in contact-email. The original address is stored here. This topic can also be used to indicate that the contact email is absent from the publication.
This annotation topic is not viewed on the public database.

4.8.7. contact-email:

This annotation topic carries the valid e-mail address to contact the data producer. Usually it is the contact given in the publication. If the e-mail address is not given in the publication leave the field blank and enter contact-comment: e-mail not available from

publication. When there are many corresponding authors the contact emails for correspondence can be added in the same line as a comma separated list without blanks. This list can then be used directly in the 'To:' line of the E-mail.

4.8.8. copyright:

Individual experiments or interactions might have specific copyright statements attached to them. A copyright statement on experiment level applies to all interactions which are part of the experiment. Copyright statements attached to individual interactions override the statements inherited from the experiment. Example:

EBI-449125 and EBI-449126 have annotation
copyright: These interactions are the sole property of HYBRIGENICS, and shall not be used for any business or commercial purposes without the prior written license from HYBRIGENICS (http://www.hybrigenics.com)

4.8.9. data-processing:

This annotation topic is used to describe the steps in processing of large scale data. Information about the original number of interactions described by the authors in the paper and the corresponding number in IntAct should also be stored here. These should be used with large data sets of 100 or more proteins, and where finding the protein accession numbers is an involved process and the protein ID used may be doubtful.

4.8.10. dataset:

This annotation topic is used to link various publications pertaining to a topic of interest. E.g. Cancer. Please send the topic you consider of interest to krobbe@ebi.ac.uk she will include it as a dataset and send back the annotation to be added under the annotation topic.

4.8.11. exp-modification:

'exp-modification' is used to describe the modification of experimental method for interaction or participant detection used by the authors if the protocol differs from the one explained in CV. These may affect the confidence level of the interactions. Information about the libraries used in the experiment is also recorded here. Some examples:

Experiment AC	**Annotation Description**
EBI-79857	murine E9.5-10.5 day c DNA library used.
EBI-80470	IL7 were crosslinked to PHA-activated peripheral blood leukocytes which were affinity labeled with radio-labeled IL7.
EBI-74033	Purified RSC complex was immunodepleted using anti-rsc3 antibodies.
EBI-296917	KIAA derived from human brain used, where all the proteins are

	transmembrane proteins with sizes larger than 1000 AA. Cytoplasmic regions used as baits. The prey proteins were sonicated to obtain various domains. Interactions were verified by exchanging the cDNAs used for baits and preys in the initial screening.

4.8.12. figure-legend:

Use this annotation topic to refer to the Figure Legend(s) associated with this specific experiment, as they appear in the paper being annotated. It is essential that this be added when available.

Examples:
figure-legend: 1
figure-legend: 2a and 2b
figure-legend: 2 and 4c

4.8.13. library:

The information about the library which was scanned to obtain the interacting protein clone is recorded under this annotation topic.
Example: mouse brain cDNA library,
mouse H2Kb-tsA58 myotube cDNA library. EBI-781687

4.8.14. negative:

This is used for annotation of experiments that demonstrate that an interaction does not occur. The annotation is done as shown below:

Negative: This experiment relates to a negative interaction.

Rules for annotation of negative interactions.

1. The interaction must have supporting positive interaction data in the same paper before any negative information can be entered.

2. Only experimental data in the same paper or by the same scientific group that demonstrates the negative interaction is acceptable: do not enter negative interactions based on comments in the paper or 'data not shown' since there is no evidence for this shown in the paper.

3. Post-translational modifications inducing negative interaction should be entered as annotation to the positive interactions as this is really supporting evidence for the positive interaction.

4. Protein isoforms that do not interact can be entered as negative interactions as long as there is complementary positive interaction data for other isoforms.

Data entry for negative interactions.

1) Create an experiment for the positive interaction, and then create the positive interaction.
2) Create a separate experiment for the negative interaction. Add the following annotation:
Annotation topic: negative: 'This experiment relates to a negative interaction.'
3) Create the negative interaction and link it to the experimental method. Add the following annotation:
Annotation topic: negative: 'The interaction does not occur under these experimental conditions.'
4) The experiments demonstrating positive and negative interactions **from the same paper** must be crossreferenced to each other as shown below.
In the crossreference of the positive experiment add IntAct crossreference to negative interaction with a crossreference qualifier see-also. The experiment demonstrating the negative interaction should be cross-referenced to that demonstrating the positive interactions in the same way.

Note: The experiment that is linked to a negative interaction must never be linked to a positive interaction even if the experimental method is the same. Data would be entered as in the example case 1 below so that Harwood-2004-2 and its interaction could be separated from the rest of the experiments/interactions in this paper. Otherwise, if the data were entered as in case 2, there could be orphan interactions in the database when the negative interactions and their associated experiments are separated from the positive ones. Example:

In Battle 2003: PMID 12809483
Positive interactions shown by yeast two hybrids.
Experiment Ac: EBI-296688
Interaction Ac: EBI-296713
: EBI-296720
: EBI-296734
crossreferences

Database	Primary Id	Reference Qualifier
intact	EBI-353694	See-also

Negative interactions were shown by co-immunoprecipitation.
Experiment Ac: EBI-353694
Interaction Ac: EBI-353709
: EBI-353714
: EBI-353721
crossreferences

Database	Primary Id	Reference Qualifier
intact	EBI-296688	See-also

4.8.15. on-hold:

Any experiment that is annotated with this topic will not be released to the public. This annotation is used to exclude submitted data which has not been published and integration of the data into IntAct is 'in progress'. This annotation should only be used when really necessary. The database should not become heavily populated with incomplete entries. The 'on-hold' entries can be tracked using the talisman tool (section 10).
Note: Subsequent to release of the pubmed ID or the completion of the entries from a pubmed ID; the 'on-hold' annotation **must** be deleted. Experiments **will not** be released automatically after the date has passed.
This annotation topic is not viewed on the public database.
This system gradually will be extended into a full submission management system.

4.8.16. remark-internal:

'remark-internal' is for internal use only and cannot be viewed by public. Experimental method remarks are based on the temporary annotation fixes for use until the method CVs are agreed. Please delete the remark-internal when you have made the update and/or the remark is no longer essential. Some of the 'remark-internal' used are:

Experiment AC	Annotation Description
EBI-371696	revisit when ligand binding in place.
EBI-529609	More to add to this entry! (12/4/05)
EBI-620802	Revisit when complex editor in place.

4.8.17. submitted:

This refers to the data submitted by the author which is not yet in the public domain.
This annotation topic is not viewed on the public database.

4.8.18. uniprot-cc-note:

This annotation topic is not viewed on the public database. This contains text which is exported into UniProtKB as a "note" in the CC INTERACTION block.

4.8.19. uniprot-dr-export:

Using this annotation topic determines whether the interactions of an experiment should be exported to UniProtKB DR lines. This can be according to a setting in the CvInteraction controlled vocabulary, or by an explicit setting for an experiment.
If this topic is set to value set by "author-confidence", Example: "high", then this means only interactions with that author-confidence are to be exported to UniProtKB.
Refer to the example for confidence-mapping.
This annotation topic is not viewed on the public database.

4.9. Crossreferences and crossreference qualifiers

Database:	**Qualifier**
pubmed:	primary reference
pubmed:	method
intact:	see-also

4.9.1 primary reference:
Primary ID is the primary source of data and is usually the PubMed ID of the paper being curated. This is only added on the Editor-Experiment page. The primary experimental reference should be added in the Crossreference field with the qualifier "primary reference". When the data has been submitted prepublication and a PubMed ID can only be assigned at a later date 'pubmed' should be the primary-reference and the primary ID column should be filled in with 'to be assigned'. You should have only one PMID primary reference.

4.9.2. method:

The crossreference qualifier "method" should be used with database as 'pubmed' and primary ID as the PubMed number, where there is a relevant paper describing the experimental technique used. It should not be used for information about the interactions.

4.9.3. see-also:

This is used while referring to an IntAct database crossreference to an experiment used to show a negative interaction. See section 4.8.11.

4.9.4. **target-species:**

This cross reference is added automatically and relates to the organism/s of the protein/s involved in the interaction. This cross reference does not need to be changed by curators.

5. Editor-Interaction

This form records the information about an interaction and the various factors that have an effect on the interaction. Each interaction is linked to a single experiment. Many interactions may be linked to a single experiment if they are all demonstrated using the same experimental techniques. However the interaction between identical proteins needs to be entered multiple times if the experimental technique demonstrating it is different. Please DO NOT enter protein interactions used as controls.
Interactions can be entered as binary interactions or n-nary interactions depending on the interaction detection and the participant identification method used. All the interactions should be entered in the database if possible. This may however not be possible where the proteins involved cannot be identified. This information should be entered in the annotation 'comments' of the interaction if only part of the interaction can be entered. If the unidentifiable interactors are part of a large scale interaction a description of the

method used to identify the proteins should be included in the experiment under annotation topic 'data-processing'. Both these should be accompanied by a annotation 'remark-internal' 'Revisit when identifiers available'.
Authors confidence in the interactions are entered using the annotation topics confidence-mapping and author-confidence as described here. For a n-nary interaction/s if the author has expressed more confidence in some interactors as being part of the interaction the proteins with lower interaction confidence as expressed by author should be included in the annotation topic 'caution'.

5.1. Interaction AC

An IntAct accession number is assigned to the following in an interaction: interaction, interactors, features, crossreferences and annotations. Following submission the interaction entries can be searched using the IntAct accession number.

5.2. Short Label

Usual rules for Short Label apply. Short Labels are lower case and should not exceed 20 characters in length.

5.2.1. If there is a numbering system, which has been developed and used by the author of the paper to denote the interactions within it, then this numbering system should be used for the Short Label.

5.2.2. If there is no numbering system developed by the author the Short Label should be derived from the gene names of two of the proteins taking part in the interaction**.** The gene name used is taken from the UniProtKB entry (GN line). If there are more than two proteins in the interaction, only two gene names should be chosen. If there is a bait protein, the gene name for the bait should be taken and one of the prey gene names should be used. Example: EBI-457906. If there is more than one gene name in the UniProtKB GN line, the first name is chosen.
Format: gn1-gn2-1

5.2.3. If the same interaction is shown by more than one method or if the first two gene names of an interaction involving multiple proteins create a duplicate Short Label, then the format becomes as follows:
gn1-gn2-1
gn1-gn2-2 or
gn2-gn1-1

5.2.4. If there is no gene name then the first part of the UniProtKB/Swiss-Prot entry name (ID line) is taken Example: for ACSL4_HUMAN, take ACSL4.
Format: gn1-acsl4-1

5.2.5. For a UniProtKB/TrEMBL entry where entry name or ID line is of the form Q6KZY6_PICTO, the accession number is used.

Example: gn1- q6kzy6-1

5.2.6. Short Labels and incorporation of special characters
The GN line of UniProtKB entries may contain special characters, use underscore to substitute for the character. Any special character is replaced by '_' for all species.
Some examples of gene names from UniProtKB and generated Short Label:
GN CG32767 {EI2} or CG5086 {EI2} or CG5113 {EI2}.
First name from the GN line used: cg32767-cg15119.
BETA-SPEC becomes beta_spec
EG:95B7.3 becomes eg_95b7_3
Su(Var)2-10 becomes su_var_2_10

5.2.7 Short Labels for self interactions
In case you have a self interaction (section 5.9.4), example autophosphorylation there is only one protein molecule involved in the interaction and the Short Label in this case will have the following format:
gn1-1
gn1-2

5.3 Full Name

The Full name for Interaction is an indication of the type of interaction and may include information about the participating interactors.

5.4. Kd

Kd values may be entered as a float or by a float followed by the letter e, followed by optional +/-, followed by integer. Example: EBI-445553

Examples :
Enter for 0.000000001 or 1.10^{-9} enter 1E-9
Enter for 0.02 or 2.10^{-2} enter 1E-2
Enter for 1300 or $1.3.10^{3}$ enter 1.3E+3

Additional information about the kinetics will be added in annotation topic 'kinetics'. In cases where the mutants have an effect on the kinetics and decrease or increase the rate of the reactions the mutants should be annotated in the features for the protein and the full name should indicate that the effect is seen on kinetics of interaction. The actual kinetic values will be included under the annotation topic 'kinetics'. The unit used should always be Moles (M).

5.5 Interaction type:

The interaction type defines the type of interaction Example: physical, enzymatic, etc.
See controlled vocabulary.
ftp://ftp.ebi.ac.uk/pub/databases/intact/current/cv/

Please refer to section 4.5 for additional CV terms.

5.6 Organism

The organism is non mandatory. The organism is important on the Editor-Interaction page only in the cases where the interaction defines a complex and could be used as an interactor in another interaction.

5.7 Experiments

Only one experiment should be added per interaction. The experiment can be unlinked or added here. The interaction will not be saved unless an experiment has been linked to it. On cloning an interaction the experiment is not cloned and needs to be added in. Search parameters in this box are the Intact AC or the Short Label for the Experiment. Wild card (*) search is permitted. When you delete an interaction from the experiment page it will get unlinked from the experiment and will be orphan. You have to either attach it to a different experiment or delete it.

5.8 Proteins

5.8.1 'Short Label', 'SP AC', 'organism' and 'Intact AC'

The short label, SP AC, Gene Name and Organism are imported directly from the UniProtKB entry. To pull up a UniProtKB entry you would need the UniProtKB ID, or the UniProtKB AC. An Intact AC number is assigned to every interactor be it a protein or an interaction. The Intact AC number can also be used to get a protein entry.

In the case of a complex the IntAct Interaction accession number for the complex can be used as an interactor while defining its interaction with proteins (this has not been implemented yet). Please read section 9.1 on annotating complexes.

5.8.2 Protein accession numbers

5.8.2.1 How to find a protein accession number

1. Try searching the following databases with the accession number, identification number, gene name or any identifier given in the paper. You could limit the search to a given organism.
a) UniProtKB via SRS or UniProtKB
b) UniParc via SRS or UniProt.

2. If you have GI number, Ref-Seq identifier, model database ID, EMBL ID, PIR ID or ECOGENE ID you could try using the PIR mapping service.
http://www.pir.uniprot.org/search/batch.shtml

3. If you have a GI number, then try searching NCBI Entrez to try to get a protein accession number. A sequence should be available if it is still live. GI numbers are being archived in UniParc, hence it is worth looking for the GI identifiers in UniParc.

4. If you are looking for a model organism accession number, try the model organism databases .
These databases usually have excellent archiving services, which should mean that you would be able to find the sequence or an accession number.

5. If a sequence is available but there is no live accession number then use MPsrch to search UniProtKB, this will give sequences and accession numbers of exact or near matches to your query.

6. If a published sequence cannot be traced to a UniProtKB entry, deposit the sequence information at
http://www.ebi.ac.uk/swissprot/Submissions/spin/index.jsp
Alternatively you could contact Ruth Eberhardt (eber@ebi.ac.uk)

7. If you have an IPI number, try searching the following databases.
a) IPI
b) UniParc
OR
c) Ask Paul Kersey (pkersey@ebi.ac.uk) for help. It helps to know both the accession number and which release of IPI the authors were working with (you may need to e-mail the authors for this information).

8. If you can not link an identifying number to a sequence then search using the gene or protein name in
a) UniProtKB
b) UniParc
c) Locus Link (covers limited number of species)
d) EMBL/DDBJ/Genbank
e) NCBI Entrez
f) PubMed and look for other synonyms
SRS can be used to search these databases individually or simultaneously.

9. If all else fails you could try a web search with any information you have.

5.8.2.2 Problems mapping accession numbers

If you cannot map a protein identifier given in a paper to a UniProtKB ID:
Enter the protein in via the Protein Editor and use the IntAct Ac to pull up the protein into the interaction. The protein editor should be used only if you cannot identify the UniProtKB AC number using all the above methods.
If the experiment is small scale, whatever information you have about the protein can be entered in the comments of the interaction:

Example: Protein with GIXXXX was also shown to interact but no UniProtKB ID could be mapped for this protein.
If the protein belongs to a large scale experiment the data should be maintained at P:\pro4\intact\local\data\pending under the experiment Short Label. In this case it is also important that you record the number of interactions in publication and number entered in intact and steps taken to identify the protein in annotation topic 'data-processing'.

5.8.2.3 Incomplete protein sequences/proteins without UniProtKB accession number.

If a partial protein sequence is given in the paper but a corresponding sequence cannot be found in UniProtKB contact Paul Browne (browne@ebi.ac.uk) and ask him to generate a UniProtKB entry from this for you.
If you find a cDNA sequence in EMBL but no corresponding translation, contact Michele Magrane (magrane@ebi.ac.uk) and an entry to be made. In both cases, provide the paper giving the evidence for the existence of the protein and any other information you feel may be of use.
If the protein sequence is generated in your own lab, but not yet published, it can be submitted directly to UniProtKB using SPIN, and an accession number will be sent to you.

SPIN: http://www.ebi.ac.uk/swissprot/Submissions/spin/index.jsp

If the protein is only in UniParc, contact Claire O'Donovan (Claire.ODonovan@ebi.ac.uk).

This incomplete protein or peptide sequence can be entered using the Protein Editor. This should also be accompanied by a remark-internal 'Update when UniProt AC available'.

5.8.2.4. UniProtKB entries that contain multiple protein molecules

All molecules in swissprot have the Feature 'chain' with a non-redundant id assigned to it. This id gives the exact sequence present in the molecule and will be linked ot the sequence. The splice forms, naturally occurring peptides, processed peptides and poly proteins all have chain ids and these can be used in the case of natural peptides and polyproteins to identify the molecule sequence.
These molecules will be entered using the protein-editor. Individual molecules will be constructed in the protein editor using information from the UniProtKB entry and any other available information. The molecule will be assigned an Intact AC number and the UniProtKB accession number-chain ID will be cross-referenced as the identity. Examples: Bioactive peptides cleaved from full length parent molecules; viral polyproteins. These molecules will have a distinct Feature ID in annotated UniProtKB/Swiss-Prot entries. These should be manually created and annotated to the limit of available identifiers, as follows:
- Short Label: UniProtKBID-FTID (lower case)
- annotation: isoform-comment if interesting

- xref: database – uniprotkb, primaryid – UniProtID-FTID, reference-qualifier - identity

Example: 1. NFKB1
NFKB1 (KBF1) is in UniProtKB as accession number P19838

NF-kB1/p105	P19838	full length
NF-kB1/p50	P19838	from N term to end of glycine rich region range 1-433 has identifier P19838-PRO_0000030311

The processed form (the p50 subunit) of NFKB1 does not have a separate identifier from the full-length protein in UniProtKB. The p50 subunit was used as the bait in TAP tagging experiments by Bouwmeester *et al.* [5]

Example: 2. Polyproteins
Polyproteins are proteins that are synthesized as a single polypeptide and then cleaved into several distinct proteins.
The Polyproteins should be entered using the Protein Editor with a crossreference to the UniProtKB AC and FTID. Example: EBI-709578

5.8.2.5 Natural Peptides

The interaction may involve a natural peptide. This is annotated using the feature chain identifier as mentioned above.
Example:

Experiment	Interaction	Interactor
EBI-79357	EBI-79361	EBI-821760

5.8.2.6 Bi-cistronic expression

Interaction of two tagged proteins bi-cistronically expressed in *E. coli*.
Interaction of these proteins was shown by co-purification on anti-tag resin. Enter this as follows:
Participant detection is ‘nucleotide sequence’, or if sequence of the participating proteins was already known or ‘predetermined'. The interaction of the two proteins forming the heterodimer is entered as one interaction: any subsequent interaction of another protein with the heterodimer is entered as another interaction.

5.8.2.7 Synthetic Peptides

When a chemically synthesised peptide is a component of an interaction, add the UniProtKB AC number of the protein to which it relates, then add ‘chemical synthesis’ under ‘ExpressedIn’ and add annotation under ‘comment’ that the component is a peptide

rather than a protein.
Example: EBI-77665.

5.8.2.8. Identification of proteins by sequence similarity

When a similarity search (Example: MPsrch) does not give an exact match to the protein you are looking for, check any possible candidate proteins with the following people:

Sandra Orchard: orchard@ebi.ac.uk
Eleanor Whitfield: eleanor@ebi.ac.uk
Gill Fraser: fraser@ebi.ac.uk
Michele Magrane: magrane@ebi.ac.uk
Kati Liaho: kati@ebi.ac.uk
Bernd Roechert: Bernd.Roechert@isb-sib.ch

5.8.2.9. Proteins of unknown species

Examples:
a) Cases where the species of the protein is known but there is no database entry, as the gene has not been cloned for the protein in question. These should be entered via the protein editor.

b) The paper does not state the species of the protein.
These cannot be entered unless we can be certain of the origin of the protein used in the experiment in the paper and find a database entry.
If there is no response from the authors of the paper then we will have to disregard these data.

5.9. Protein roles

The Protein roles are split into biological role and experimental role in PSI2.5. Both of these will need to be specified at a future date. For now use the role most accurately describing the role of the interactor.
See ftp://ftp.ebi.ac.uk/pub/databases/intact/current/cv/

5.9.1. bait/prey

Bait must always have one or many prey(s)**.** If there is an interaction with prey(s) there must be bait.
The bait is in general the protein which is used to fish out a protein. This may be immobilised except for far western. The roles bait and prey are associated with Y2H, CoIP, TAP, Pull downs etc.

5.9.2 enzyme/enzyme target

An **enzyme** must always have an **enzyme target** (and vice–versa).

The agent is the enzyme which modifies the protein (example: kinase) and the target is the modified protein.

5.9.3. neutral component

Role neutral component is used where the role of the protein cannot be classified as bait/prey or enzyme/ enzyme target. Role neutral component is often observed with stoichiometry determination or oligomerisation studies.
Unmodified identical proteins may be shown to self-associate using techniques such as density gradient sedimentation, non-denaturing gel electrophoresis, mass spec, crystallography, and gel filtration (size exclusion chromatography). These will be entered as neutral component components. The number of molecules interacting will be indicated by stoichiometry (section 5.11)

5.9.4 self

Only proteins that interact intra-molecularly should be given the role 'self' (Example: autophosphorylation or disulphide bond).

Where the same protein has been modified in different ways, (Example: tagged with different tags) and these molecules are shown to interact, these are considered as non-identical subunits. They should be entered as separate proteins with the role neutral component or bait/prey as appropriate to the experiment and the tags used described as features of the protein.

5.9.5. electron acceptor/electron donor

These will be used where the interaction involves an exchange of electrons.

5.9.6. inhibited/inhibitor

This is used where the interaction detection technique indicated interaction between the inhibitor and inhibited molecule.

5.9.7. acceptor fluorophore/donor fluorophore

This pair of roles is used in conjunction with a FRET experiment where the interacting proteins are coupled to the acceptor or donor fluorophore molecules. The donor molecule will be coupled to a fluorophore which gives off an emission on excitation, which is then recognised by the acceptor fluorophore. Some of the common donor acceptor pairs are CFP/YFP, BFP/GFP, GFP/Rhodamine and FITC/Cy3. This method detects interaction by demonstrating proximity of 1-10 nm. In the case of BRET and HTRF assays there is an acceptor fluorophore and the protein coupled to the molecule giving the fluorescence is assigned the role unspecified.

5.9.8. unspecified

This will be used when the biological role of the interactor is difficult to ascertain example in an ITC experiment. Example: EBI-456897
NOTE: If the roles are not apparent from the paper please read back references to the constructs. If these are also not clear contact the author for explanation.
This is a biological role however, till a solution has been found it is also used as a experimental role for cases where the interaction method is known, interactors are published as a non redundant set of interactors. Example: EBI-711122

5.9.9 complex

This role has been made obsolete. Please see section 9.1 on annotation of complexes.
This field should be chosen in specific cases when the proteins are subunits forming a complex with a specific activity. This field should also be used while annotating curated complexes. Use roles bait/prey or neutral component as appropriate while annotating interactions as much as possible.

5.10. ExpressedIn

'ExpressedIn' refers to the source of the protein when it is not expressed in the same system as that in which the experiment was carried out. This is the case for '*in vitro*' interaction. Leaving the field on --select-- indicates an endogenous protein.
The protein may come from any of the following sources:

5.10.1 Heterologous protein expression

Example: Protein A is a human protein expressed in baculovirus/sf9 cells, then purified and mixed with Protein B which is a nuclear extract from HeLa cells (endogenous protein)

Experiment Host organism: *in vitro*
Interaction: Protein A 'ExpressedIn' Sf9 cells
(enter spofr-sf_9 -*Spodoptera frugiperda* insect cells)

No information about the expression of protein B is added, as this is an endogenous HeLa cell protein.

If however a tagged protein from human source is over-expressed in Hela cell line this will have the expressedin as human-hela. The entry will have annotations: delivery method and expression level. The information that HeLa cells were used could be stored under experiment under the annotation topic exp-modification.

5.10.2 *In vitro* transcribed and translated proteins

In vitro transcription and translation : this is entered in the interactions page as '*in vitro*' in the 'ExpressedIn' box.

5.10.3 Synthetic peptides

Please see section 5.8.2.7. The ExpressedIn is 'synthetic peptide'.

5.11. Stoichiometry

The default stoichiometry is set to 1, these needs to be modified. In cases where the authors have determined the number of molecules of each protein interacting please enter the values here. If stoichiometry is 1:1 please enter this in remark-internal.

5.12. Interaction Annotations

5.12.1. agonist:

Add to CVtopic when one or both interactors are activated by specific agonist. Example: EBI-448589. An agonist is defined as external cell receptor or an upstream molecule to stimulate a downstream interaction, potentially by modification of one or more of the interactors. For example cells are treated with the agonist IL-2 causing initiation of a kinase cascade and phosphorylation of cellular proteins, leading to interaction formation.

Interaction AC	Description
EBI-624110	Insulin - pretreatment for 5 minutes stimulates reaction
EBI-617942	Erythropoietin. Interaction peaks 1 minute after addition and decreases to basal levels by 60mins.
EBI-446460	cells expressing EGFR have to be EGF-stimulated for interaction
EBI-472895	Interaction occurs in response to hypoxic stimulation.

5.12.2. antagonist:

Add to CVtopic when one or both interactors are inactivated by specific antagonist. Example: EBI-529081. Opposes the effect of a natural ligand by binding at the same receptor or alters the state of an upstream molecule to inhibit a downstream interaction.

Interaction AC	Description
EBI-365437	Stimulation of the cells with TNFalpha negatively affects this interaction.
EBI-493544	Antiprogestin treatment.
EBI-529081	Treatment of L929 cells with Interleukin-15 and tumor necrosis factor (alpha) as described in the paper abolished these interactions
EBI-624105	Insulin - pretreatment for 5 minutes inhibits reaction

5.12.3. author-confidence:

This is the confidence value assigned by the author to the interaction. This could be numerical based on a statistical method or as set by author. Please see section 4.8.4 for the method used for confidence-mapping. Some examples are:

Interaction AC	Description
EBI-196485	High (can similarly have low)
EBI-196485	1.00 (and other statistically derived values, description of method used in confidence-mapping)
EBI-375806	ito-core
EBI-470750	c_leu =3 and c_lacZ = 4

5.12.4. caution:

This is a warning about errors/ grounds for confusion example: significant differences between the sequences of the protein described in the paper and the UniProtKB entry for the protein. Some examples are:

Interaction AC	Description
EBI-602746	The TC21 peptide used does not match 100% the entry, it has been mutated to make it more soluble: KKKSKTKCVIF
EBI-602921	The authors note that attempts to coimmunoprecipitate ETA2 and Cullin1 from extracts of arabidopsis seedlings using the respective antibodies did not result in detection of the reciprocal protein (data not shown in paper).
EBI-539614	Residue numbering for SAFB1 as given in the paper - this is stated as being the C-terminal but sequence actually extends beyond 874 by a further 57 residues
EBI-523545	Other papers have reported an inability to detect an interaction between Daxx and Ask-1. See Michelson,J.S. The Daxx enigma. Apoptosis 2000;5:217-220.

5.12.5. comment:

'comment' is free text, which can be used to describe additional information about the interaction and the proteins involved in it. Comment can be added to state the small molecules detected in the complex, use the ChEBI nomenclature for naming the molecules and cross-reference it. They will be entered as component later. Add the remark 'revisit when small molecule editor is in place'. Please see section on 'remark-internal' for comments accompanied by remarks.

Interaction AC	Description
EBI-1416	Interaction inhibited the growth of yeast
EBI-371775	Binding prevents homodimerization and activation of ERN1
EBI-626867	LdpA interacts with CikA independently of clock proteins as shown in

	kaiA, kaiB or kaiC mutant strains.
EBI-618534	The interactions are specific to the GTP form of Ypt7.

5.12.6. complex-properties:

This field should be used to describe properties and activities of the complex being annotated. This annotation topic can also be used to store information about the complex until a suitable GO term is available for the complex. See section 9.1 for annotating complex.

Interaction AC	**Description**
EBI-626046	The presence of PIF3 is essential for formation of this complex.
EBI-626652	Based on their migration, four NDH-1 complexes were denoted : NDH-1L (large), NDH-1M (medium size), NDH-1S1 (small 1) and NDH-1S2 (small 2).
EBI-626652	Complex formed is about 350 kDa
EBI-626757	The inner ring consists of either 6 or 7 copies of IsiA, whereas the outer ring is formed by 7-10 copies of IsiA.

5.12.7. delivery-method:

Delivery method refers to the mode of inclusion of the interactor or its template or the precursor in the system where interaction was detected. This will be used as follows:
Interactor accession number followed by the delivery method ie. Transformation, electroporation, calcium treatment, heatshock, infection, microinjection, transfection, recombination.
Structured text is used in this annotation and the format should be accession number of the interactor space and one of the above CV terms. This will be relocated appropriately – to the interactor when possible.
Example: EBI-781689
Q91WZ8-1 transformation
O70585 transformation

5.12.8. disease:

Add to the annotation CVTopic as a specific comment where:
a) An interaction is shown to have a disease association
b) Mutations in any or all of the interacting proteins have a disease association
Example: FAD mutation in APP.
Add the crossreference to OMIM database as a ‘see also’ if possible.

Interaction AC	**Description**
EBI-608157	prostate cancer
EBI-297057	Disruption of this interaction by mutations may cause congenital heart

	defect (CHD).
EBI-524382	The mutation E303Q is closed to the mutation found in TORSION DYSTONIA 1 disease which is often a deletion of E303 (glutamic acid deletion). Both mutated proteins are mislocalized in the nuclear envelop lumen instead of the endoplasmic reticulum.
EBI-371460	Level of ATRX-DAXX complex is significantly less in an ATRX syndrome patient cell line.

5.12.9. expression-level:
This annotation will be used to determine the expression level of the engineered construct. Strict guidelines will be followed in the way the free text box is used. The usage will be as follows: The accession umber of the protein followed by space and a choice of over-expressed, under-expressed or endogenous level as shown below:
P12345 over-expressed level
P12345 under-expressed level
P12345 endogenous level

5.12.10. figure-legend:

This is the figure number in the paper where the interaction was shown. This should be preferentially added to the experiment. It is however essential that this be added when available.

5.12.11. function:

Use this topic when the function of the interaction has been determined by the interaction and the function cannot be assigned using a GO function term. Suggest a Go term along with all ancillary evidence to the GO tea from the sourceforge pages GO sourceforge: (http://sourceforge.net/tracker/?func=add&group_id=36855&atid=605890 and update once the term has been assigned.

5.12.12. inhibition:

Use this topic when the interaction being annotated is disrupted by biological entity or by modification of interactors. This suggests more immediate effect on interaction. The downstream effects will be covered by 'antagonist'.
Example:

Interaction AC	**Description**
EBI-447320	This interaction was disrupted by chetomin.
EBI-447720	Inhibition by autophosphorylation of IRAK-1 at undefined position(s).
EBI-620781	By the C-terminal 99 amino acids of the protein
EBI-490284	Competitive inhibition of this interaction is observed when GST-complexes are incubated with anti-cd44 antibodies.

5.12.13. isoform-comment:

This topic is used to add comments about isoforms. Example:

Interaction AC	Description
EBI-198400	Nuclear localization signal at positions 335-340
EBI-202571	Has been shown to exist only in D.simulans so far
EBI-208761	Can partially substitute for the cytoplasmic enzyme activity. Isoform D is produced by alternative initiation at Met-106 of isoform B
EBI-208766	Unable to compensate for the loss of the mitochondrial enzyme activity

5.12.14. kinetics:

Use this topic when a mutation induces change in the kinetics of interaction, and information on the new Kd or another numeric values are available (see 5.2.5). For Example: "Mutation Asp234Glu caused the Kd to become 0.0003M." kd of the parent molecule should be added to the kd field.

Interaction AC	Description
EBI-519455	Addition of p21-CIP1(P38936) alone increased affinity by 35 fold (2.10E-4 M) by reducing the rate of dissociation
EBI-524451	FIST protein which had a truncated kinase domain (142- 433) displayed reduced binding affinity for Daxx in this study.
EBI-591030	The affinity between TGN38 cytoplasmic tail and the mutant Y350A adaptin mu2 mutant is kd = 7.55 10-7 M
EBI-528854	Kd values were obtained by fitting a four site model.

5.12.15. negative:

This can be added for information on the entries made for negative interactions. The filter is placed on the 'negative' annotation in Editor-Experiment page. The IntAct Ac of the negative interaction should be cross-referenced with qualifier 'see-also' on this interaction. For examples see EBI-489624 and EBI-489729

5.12.16. pathway:

This refers to the metabolic pathway that the interaction/complex is involved in.
Note: where possible, GO annotation through cross-reference is preferable to free text. If the pathway involves a Reactome Complex please also enter the Reactome ID under database 'reactome-complex'.

5.12.17. remark-internal:

'remark-internal' is not for public use. Interaction remarks are temporary annotation fixes for use until other tools are available. Some of the remark-internal, which are used on the interaction page, are:

Requires updating in UniProtKB
Revisit when Complex Table in place
Revisit when uncharacterized participant in place

Use of comments and remarks
a) Comments about the UniProtKB entry should be accompanied by the remark 'Requires updating in UniProtKB'.
b) Comments about well-characterized complexes should be added here with the remark 'Revisit when Complex Table in place'

5.12.18. resulting-ptm:

This annotation topic is used to record the effect of an interaction on the PTM on its interactors. The PTM could be added or removed from molecules as a consequence of the interaction. The PTM if present on the interacting molecule should be added as feature of the interactor.

5.12.19. sample-process:

This annotation topic is used to indicate the origins of the interactor. This may be any one of the following: cdna library, cell lysate, subcellular preparation, fixed cell, living cell, purified, homogenous and partially purified.
Structured text is used in this annotation and the format should be accession number of the interactor space and one of the above CV terms. This will be relocated appropriately – to the interactor when possible.
Example: EBI-781689
Q91WZ8-1 cdna library

5.12.20. stimulation:

Use this topic when the interaction being annotated is stimulated by a modification in the protein or small molecule. The interaction itself can occur in the absence of the stimulator. This suggests more immediate effect on interaction. The downstream effects will be covered by 'agonist'. Example: EBI-74820 , EBI-465428

Interaction AC	Description
EBI-595254	Neddylation of CUL1 enhanced SKP2 binding
EBI-515302	Cross-linking of GPIV (Q9UIF2) in platelets increased associated FcR gamma chain (P30273).
EBI-495266	Interaction increased by DNA damage dependent phosphorylation of histone H2A.X on Ser-139.
EBI-465471	AlF4- is added to block the complex between GAP/G protein, here it enables to detect the P115/Galpha13 interaction. Same experiments were conducted without AlF4-, but with the GTPase-deficient G alpha 13 mutant Gln226Leu.

5.12.21. uniprot-cc-note:

This annotation is added to specify the text which is exported into UniProtKB as a "note" in the CC INTERACTION block. This is not exported for public viewing.

5.12.22. url:

The url of the database storing the protein-protein interaction data. Please note that the interaction Short Label used by the database should be crossreferenced to as 'see also'.

Interaction AC	Description
EBI-308790	http://www.kazusa.or.jp/tech-cgi/tablelist.ppi.cgi

5.12.23. 3d-structure:

This annotation topic should be used in conjunction with crystallographic or structure data and should be used to comment on the structure of the 3D-complex.

Interaction AC	Description
EBI-446431	Presence of COA-S-ACETYL TRYPTAMINE (COT) in the crystal
EBI-449105	The natural inhibitor Brefeldin A (BFA), GDP and Magnesium are bound to the complex
EBI-541124	MDM2 has a deep hydrophobic cleft on which the p53 peptide binds as an amphipathic alpha- helix. The interface relies on the steric complementarity between the MDM2 cleft and the hydrophobic face of the p53 helix. A triad of p53 amino acids - Phe19, Trp23, and Leu26 - insert deep into the MDM2 cleft.
EBI-603513	The non hydrolysable GTP analogue GMPPPCP and Mg are bound to Cdc42

5.12.24. 3d-r-factor:

This annotation topic should be used in conjunction with crystallographic data to denote the r-factor of the structure.

Interaction AC	Description
EBI-297231	working 19.1%, free 20.6%
EBI-539447	working 25.0%, free 29.6%

5.12.25. 3d-resolution:

This annotation topic should be used in conjunction with crystallographic data to denote the resolution of the structure. Units in deg angstrom (^{o}A) should be used; the text field however does not allow the entry of the superscript hence enter it as shown below.

Interaction AC	Description
EBI-449117	1.46
EBI-602717	2.40A

5.13. Crossreferences and crossreference qualifiers

Databases:	Qualifiers
pubmed:	see-also
go :	component, process, function, translocation, translocation start, translocation end
pdb:	identity, see-also
intact:	see-also, intact-secondary
huge:	identity
reactome-complex:	identity, see-also
omim:	see-also

5.13.1 identity:

This crossreference qualifier should be used to define the identity of the interaction. Example: For a PDB entry the primary reference is the PDB ID if the PubMed ID in the PDB entry is the same as that of the paper being curated. In this case the database is 'pdb', the ID is 'PDB entry name' and the crossreference qualifier is 'identity'.

Example: EBI-297231

When a paper describes a full or partial reactome-complex it should be cross referenced as follows: database is 'reactome-complex', the ID is the reactome identifiers the crossreference type is 'primary reference' and the crossreference qualifier is 'identity'.

Note: The two examples differ in that for the PDB entry the structure has been described by the authors in the paper being curated and they provide the crossreference to the PDB entry while in the case of reactome-complex the interaction forms an integral part of the reactome entry. You may have more than one reactome crossreference to an interaction.

5.13.2 see also:

This crossreference qualifier should be used where an interaction has been shown by two or more experimental approaches or there is a significant crossreference which gives added information about the interaction and cannot be cross referred in any other way.
Some instances where this may be used are where there is another IntAct entry describing the interaction; the authors have quoted another study for the same interaction from the

same or a different lab., which corroborates the interaction or negates it; an interaction has been described in a PDB structure with interaction shown *in vivo* by other techniques. Example: EBI-297231

5.13.3 GO crossreferences:

There may be additional references or GO terms on the interaction level**.**
The cross reference qualifier is directly picked up from the GO tables, hence, it is only necessary to choose the database and primary ID for the GO terms.

5.13.3.1 component:

Interaction shown by co-localisation (classical, electronic or confocal microscopy…) should have a cross-reference to a GO component term/s where possible.

5.13.3.2 process:

Interactions which form a part of a process should be crossreferenced using GO molecular process term/s where possible.

5.13.3.3 function:

When an interaction decribes a function it should be crossreferenced using the GO:molecular function term/s where possible.

GO crossreferences: molecular function, molecular component and molecular process

Protein complexes can be annotated at the interaction level using crossreferences. See below for annotation of protein complexes.

Note: References relating to the **protein** should be added on the Protein level using the Feature editor (e.g. InterPro Cross-Reference); see 6.1.6.

6. Editor- Feature:

This editor is available from the interaction page by pressing the 'Add Features' button on the protein you are annotating. You will have to define the role of the protein and save the protein before using the Feature Editor.
This editor is used to add the features onto the proteins. Features describe attributes of the interacting molecule. The attributes of the interacting molecule are determined based on the interaction type chosen for the interaction. This in turn depends on the interaction detection method used. Example: If you define the 'InteractionDetection' method as being a 'kinase assay' the 'InteractionType' would be 'phosphorylation' and the mutations affecting phosphorylation would be added as features of the interactor. This does not necessarily mean that the physical interaction of the interactors is affected. If the physical interaction is not hindered this would be indicated in the paper by another

interaction detection method and the information should be stored in the comments section of the interaction showing effect of mutation on phosphorylation.
On submitting the features these will be visible on the interaction page.
In the Interaction-Editor under Proteins: light blue title bar corresponds to table heading for features.

6.1 The 6 fields

6.1.1 Short Label

This is a mandatory field but the Short Label does not need to be unique to every feature entry so there is no need for numbering. There is a character limit of 20, all characters have to be lower case and no special characters can be used ('-', '_' and multiple spaces are allowed). This has to be filled by curators for all entries except in the case of 'mutation decreasing' or 'mutation increasing' where it will be automatically generated if you use the mutation editor. The Short Label should describe the feature. The end user views the features as an auto generated sentence with the following format: 'Feature Type' 'Short Label' of 'UniProtKB protein ID line' detected by 'Feature Identification'. This should be taken into consideration while generating the Short Label of the feature.

6.1.2 Full Name

This field is non-mandatory and will be left empty in most cases. The only places where it will be filled in are in the cases of PTM and Hotspots. Please see explanation in the specific cases given below. This field does not show up in the external search currently.

6.1.3 Feature Type

This is a mandatory field and you need to pick out the appropriate type from the drop down list. This is part of the controlled vocabulary.

Please see the following link for hierarchy and definitions of controlled vocabularies.
ftp://ftp.ebi.ac.uk/pub/databases/intact/current/cv/

6.1.4 Feature Identification

This is a non-mandatory field. The appropriate identification method has to be chosen from the drop down list, from a controlled vocabulary. However, feature identification method may not be applicable for all features.

6.1.5 Ranges

The range is mandatory. These are used to keep track of the feature positions in the protein sequence. There are two boxes which need to be filled in 'From Range' and 'To Range'. Values include

- integers for the beginning (From Range) and end of feature (To Range)

- Fuzzy type denoted as integer followed by two dots followed by integer. Example: region around amino acid 10 involved will be entered as follows- From Range: (start)10..(end)10, to range (start)10..(end)10. It will then be displayed as 10..10-10..10 on interaction page.
- Fuzzy type between two regions: Example-from region around amino acid 10 and to region around amino acid 40 will be entered as follows- From Range: (start)10..(end)10, to range (start)40..(end)40. It will then be displayed as 10..10-40..40 on interaction page.
- Fuzzy type between two ranges: Example-from region between amino acids 10-13 and to region between amino acids 40-45 will be entered as follows- From Range: (start)10..(end)13, to range (start)40..(end)45. It will then be displayed as 12..15-40..45 on interaction page.
- From Range: n ,To Range; n, displayed as 'n-n' for N-terminus of protein
- From Range: c, To Range: c, displayed as 'c-c' for C-terminus of protein
- From Range: ?, To Range: ?, displayed as '?-?' for undetermined
- Region with a less than and more than boundary can be displayed as '<integer' for less than and '>integer' for more than.

The integer range, fuzzy type range, less than and greater than can be used in combinations to define the range. Example ranges can be defined as follows:
From: >112 To: 234
From:<112 To: 234..256
From:10..10 To: >56

In the case where the feature is on a single amino acid, for Example: phosphorylation on serine 10 of the protein - you would enter 10 in both the 'From' and 'To' boxes.
If there is intramolecular crosslinking or a disulfide bond between any two amino acids then the two positions are entered in to the 'To' and "From" boxes and the "Link" set to True
For Example:, a disulfide bond exists between amino acid 3 and 56, then you would enter '3' in 'From' box, '56' in 'To' box and set the 'Link' to true.
When Features added on the proteins entered using a protein editor the ranges should correspond to the position in the sequence entered

6.1.6 Crossreferences

In cases where the feature can be assigned an InterPro crossreference or a PDB crossreference, you may do so using the drop down list for the databases and the IPR number or PDB entry name as primary reference. This is a non mandatory field. InterPro references can be found for domains (Example: IPR000001 for kringle), repeats (Example: IPR000357 for heat repeat), binding site (Example: IPR000205 for NAD-binding site), active site. PDB cross-references should only be added if different from that already given on the Edit Interaction page, for Example: a paper contains a second structure giving a more detailed view of the binding feature in question. Both these references are added with crossreference qualifier 'identity'. Some examples of various feature types:

Feature type	Interaction AC	Feature AC	Short Label	Range	Feature detection	Xref qualifier
binding site	EBI-518195	EBI-518200	sh2 domain	149-257	deletion analysis	interpro identity
mutation decreasing	EBI-490049	EBI-490092	ser54phe	54-54	-	omim see-also
phosphothreonine	EBI-593877	EBI-593886	thr735	735-735	-	-
acetylated						
35s radiolabelled	EBI-518545	EBI-518556	region	?-?	-	-
flag tagged	EBI-630171	EBI-630177	n-terminus	n-n	-	-
gfp tagged	EBI-602285	EBI-602288	region	?-?	-	-
tagged molecule	EBI-529065	EBI-529068	polyoma epitope	?-?	-	-
mutation decreasing	EBI-525187	EBI-525205	pro253ala-pro256ala	256-256 253-253	mutation analysis	-

Note: binding site has two children 'required to bind' and 'sufficient to bind'; if amino acids are sequentially mutated to alanine the feature detection method is alanine scanning.

6.2 Various cases

The Feature Editor should be used in the following cases:

6.2.1 Molecule is tagged

Do not enter feature type 'tagged protein' or children of this term where the Interaction detection method on the linked experiment is 'protein complementation assay' or a child term.

The common tags include HIS-, TAP-, HA-, FLAG-, GST, GFP etc. This is divided into fusion proteins and tagged molecule. Fusion proteins will include the GFP and GST tags from the above examples.

Short Label - in these cases is used to describe the position of the tag i.e. 'c-terminus' or 'n-terminus' where position is known, 'region' when unknown. If a novel tag is used, i.e. one not present in the current controlled vocabularies, include the name of the tag Example: 'xyz-tag n-terminus' having annotated at the level of 'tagged-protein' as FeatureType. Examples of short labels are n-terminus and region.

'FeatureType' – will be a description of the tag Example: 'flag-tagged'. If the tag is not part of the existing controlled vocabulary, annotate to 'tagged-protein' and describe the tag used in the Short Label.

Range - These tags may be introduced at the 'C-terminus' or the 'N-terminus' of the protein (generally molecules are n-terminally tagged). In cases where this information is provided this information can be filled in the ranges field as 'From Range'- 'c' and 'To Range' – '**c**' or 'From Range' 'n' and 'To Range'- '**n**' respectively. Use '**?**' if unknown.
If the protein has multiple tags, they should be entered as separated features.
Examples:
EBI-448740
EBI-465463
EBI-464885

6.2.2 Molecule is radio-labelled

Short Label – Use the region of radiolabelling Example: 'region' or 'n-terminus'

Feature Type – choose radiolabel used from the drop down list example 131i radiolabel or 35s radiolabel

Range - The actual position of the radio-labelling, if specifically known, can be described in range – use From '**?**' To '**?**' if the position is unknown or the protein was globally labelled.

Example: Protein expressed in bacteria grown in 35S growth medium.
EBI-465724

6.2.3 Molecule is post-translationally modified (PTM)

Each individual PTM goes in as a separate feature.
For Example: Phosphorylation of Serine 10

Short Label – use to describe the PTM i.e. 'ser10'. For PTMs where the site is unknown Short Label will be 'region'.

Full Name - The full name of the PTM will be '(possible)' or '(required)'. '(possible)' will be used where the authors have used a protein with a PTM but have not shown that the interaction depends on the PTM. For Example: phosphorylated Protein A has been shown to interact with Protein B but no-one has shown that the dephosphorylated form does not. Do not list all "possible" PTMs present in a UniProtKB entry – only those discussed in the paper. Only in cases where the authors have shown in the paper being annotated that the PTM is necessary for the interaction should '(required)' be used.

Feature Type – select the appropriate method Example: 'o3-phosphoserine'. Feature identification should also be entered, if described in the paper.
EBI-458277 (feature EBI-458290)

Range – Range will be the position of the PTM, if unknown it will be '? - ?'.

Note – if a PTM is produced as a RESULT of the interaction that you are annotating, this is **not** a feature of the molecules but a "Resulting PTM" which is an annotation topic.
EBI-458277 (resulting PTM)
For Example:
Protein B is phosphorylated on Ser-99 by Ser/Thr protein kinase A
The interaction between Protein kinase A and Protein B can be annotated as an enzyme/substrate interaction, but add the resulting serine-99 phosphorylation under the Annotation Topic "Resulting PTM" not as a feature.

6.2.4 Molecule is a fragment, truncation or deletion construct

If a paper describes a series of deletion constructs, annotate to the shortest region that shows interaction. Only positive interactions should be described. These should be described as necessary or sufficient where possible.
Example: EBI-77202, EBI-308946

For Example: Protein A (length 50aa) binds to Protein B. A deletion mutant of Protein A is constructed from regions 1-20 and 30-50 and this fails to bind. Annotate the fragment 21-29 as the 'binding site' with feature detection as 'deletion analysis'.
Example: EBI-465428

Short Label of the feature – This may be described just as 'region' or the Short Label may be used to give a more detailed description or features within the fragment Example: 'sh3 domain', 'heat repeat', 'nad'(for NAD-binding site), where the fragment contains the domain, repeat or binding site respectively.
For C-terminal fragments, the Short Label will be 'c-terminal'
For N-terminal fragments, the Short Label will be 'n-terminal'
For the cytoplasmic region, the Short Label will be 'cytoplasmic region'
For transmembrane fragments, Short Label will be 'transmembrane region'
Example: EBI-77516
Where an InterPro domain, repeat or binding site has been defined the Short Label will be derived from the names of these domains (in minuscule to be recorded). Example: sh3 domain', 'heat repeat', 'nad' (for NAD-binding site) – use the InterPro short name if the full name exceeds 20 characters.

FeatureType – 'binding site' (this does not imply all domains within a fragment are necessarily involved in binding). Children of this term – required to bind and sufficient to bind should be used where possible.

Range – Range defines the fragment of protein. This may be specific (2-10), fuzzy (2..2-10..10), 'c-c' for C-terminal fragment, 'n-n' for N-terminal fragment, and '?-?' where it has not been defined.
When fragment boundaries are not known but the domain(s) is/are known, the boundaries for the domain found in UniProtKB entry should be used with fuzzy type with an Feature Feature detection method being 'inferred by curator'. (Example: From Range:597..597, To Range:638..638). There is a link to the UniProtKB entry by clicking on the protein Short Label. EBI-353537.

Crossreference – If this binding site has a domain, repeat, well defined binding site or active site relevant to the interaction, this should be added in as a crossreference with the database 'interpro' primary id 'InterPro accession number' and crossreference qualifier 'identity'. EBI-457906

6.2.5 Molecule contains a Mutation

If the interacting protein/fragment used contains a mutation, it should be annotated in 'feature' as 'mutation'.
If the experiment was used both with wild type fragment and mutated fragment:

- if the mutation impairs interaction, it should be entered as mutation decreasing see 6.2.6
- if the mutation does not change the interaction: no need to enter this mutation
- if the mutation changes the kinetics, it should be entered as mutation with the

Full Name: Alters kinetics of interaction. See section 6.2.8. The feature type mutation increasing or decreasing will be used to indicate the effect of mutation.

Example:

Short Label: lys235thr
Full name: blank.
Feature type: mutation
Range**:** 'From' 235 'To' 235

When there are several mutations in the same construct, hyphens are added in Short Label: lys235thr-ser283thr. Each mutant construct should go in as an independent feature. The mutation/hotspot editor can facilitate annotation by automatically writing ranges and Short Label using the mutation editor. See .6.7.

6.2.6 Molecule contains a mutation decreasing (synonym: hotspot)

For the purpose of protein-protein interaction detection a Hotspot is one or two adjacent residues which are required for the interaction, though these may not necessarily be the interacting residues. These may be defined by mutation or deletion analysis. These mutations or deletions may significantly negatively modulate or prevent the interaction from occurring. These should be annotated on a positive entry with feature type – mutation decreasing.

If a deletion is 3 amino acids or more it should be annotated as
FeatureType - binding site
FeatureIdentification – deletion analysis

If mutation studies cover 3 or more amino acids these should be annotated as
FeatureType – binding site
FeatureIdentification – mutation analysis

For example, Protein A fails to interact with Protein B when Ser-10 of protein A is mutated to alanine. This should be annotated as an A-B interaction with Protein A containing a mutation decreasing feature on Ser-10

Short Label – should describe the hotspot Example: ser10ala.
EBI-458224
In an alanine scanning experiment, where a series of residues are sequentially mutated to alanine, use the Short Label to describe the residues Example: 'ser glu gly prop gly' if 5 or less residues are mutated. If more than 5, use the Short Label 'region'. In both cases, describe the mutated range in Range. EBI-448734

FeatureType – 'mutation decreasing' or 'required to bind' depending on if a mutation or region is delineated as abolishing the interaction.

Feature Identification - 'mutation analysis,' 'alanine scanning' or 'deletion analysis' as appropriate
Each mutant construct should go in as an independent feature. The mutation/hotspot editor can facilitate annotation by automatically writing ranges and Short Label.

If there is a specific region which has been studied by mutation analysis the region is annotated as binding site and the specific mutations within this region are annotated using mutation decreasing or mutation increasing. The interacting regions proven by this mutation analysis may then be linked if necessary.

6.2.7 Mutation Editor

Each mutant construct should go in as an independent feature. In cases where multiple mutant constructs have been used to analyse the amino acids involved in the interaction, these should each go in as independent features. To facilitate this, there is a button on the left top corner of the Feature Editor which toggles the editor between the 'Mutation Editor' and normal editor. The mutation Editor can be used to facilitate mutation and hotspot annotation. The Mutation Editor only has 'submit' and 'cancel' buttons. The information about the mutations will be put in the full name and this will automatically be translated into different features, the Short Labels of the form 'lys234ala' will be generated from the full name.

The specifications for the Mutation Editor are as follows:

A modified feature editor to incorporate the hotspots:
Full Name must consist of a description of the mutants to a specific format which may include:
a. '&' - differentiates between range for a feature on the same construct
b. '|' (pipe) -separates features on different constructs

For Example:
Full name: lys235thr & ser283thr | lys5632thr (spaces can be omitted)
defines two constructs.

Feature 1:
Derived from lys235thr & ser283thr
Short Label: lys235thr-ser283thr.
Full name: blank.
Feature type: hotspot
Range: 235-235
283-283

Feature 2:
Derived from lys5632thr
Short Label: lys5632thr
Full name: blank.
Feature type: hotspot
Range: 5632.

Note:
- By pressing 'submit', you go back to interaction page with the new feature added
- If you change your mind, you can go back to normal feature editor by pressing the top left button, be aware that it will be an empty one.

6.2.8 Mutation which effects rate of an interaction

In cases where the mutation has an effect on the kinetics of the interaction but does not disrupt the interaction, the information is entered as a mutation of the parent molecule. The full name will have information on the effect of the mutation.
For Example: Mutation at D234E of Protein A changes the kinetic parameters of the interaction.
The feature under protein A will have

Short Label: asp234glu,
Feature type: mutation,
Range: 'From' 234 'To' 234
Full name: Alters kinetics of interaction

When the values for kinetics are given, the Kd of the parent proteins should be entered into the Kd box on the Interaction Editor. Changes in Kd due to the mutation or any other numerical values should be added in as annotation, under the Annotation Topic 'Kinetics' For Example:
"Mutation Asp234Glu caused the Kd to become 0.0003M."

6.2.9 Residue mapping and linking

A paper may show specific residues within two proteins interact, or a residue on Protein A interacts with a domain on Protein B. (Example: shown with structural studies like NMR or X-Ray).
For example, residue A234 on protein A has been shown to interact with residue T456 of protein B
Protein A will have feature Short Label 'ala234', feature type 'binding site' range '234-234'
Protein B will have feature Short Label **'**thr456**'**, feature type 'binding site' range '456-456'.

6.3 Linking unlinking and deleting features

These functions are available on the Editor-Interaction page. Once you have saved the features for individual proteins, two features should be selected by clicking on the boxes on the left of the features. Selected features are then linked by clicking on "Link Features" button.
Some examples where Features should be linked are:

- one amino-acid of one of the protein is shown to interact with a amino-acid of the interactor
- Phosphorylated amino acid is shown to interact with a SH2 domain
- Authors have shown that 2 specific domains interact. (If it is only the fragment used without precision on the real interacting domains, no link is needed).
- The minimal interacting ranges for the interacting proteins have been specified.
- When PTM (s) on proteins are important or necessary for the interaction these should be linked to the interacting region.

Only two features can be linked at a time. If Feature A on Protein A binds to both Feature B and Feature C on Protein B, you need to create Feature A twice, and separately link to Feature B and Feature C.

To unlink features, select the two features to be unlinked and unlink with the "Unlink" button.
To delete features, select the Feature and use the "Delete Feature" button.

The features on an interactor in an interaction may be linked to each other. Example: Interaction: EBI-519830, Features EBI-540445 and EBI-540447 are linked and interacting.

7. Editor-BioSource

The BioSource is used for:

- host organism of the experiment
- organism of the interaction
- expressed-In (Biosource used for expressing the protein)
- Protein-Organism: the Organism from which the protein originates, this is picked up from the UniProtKB entry.

This may be any of the following:
a) An organism (tissue/cell line unspecified) Example: *Homo sapiens*
b) A tissue type, which includes cell suspensions derived from tissues.
c) A cell line
N.B. in the IntAct editor 'Cell type' denotes 'Cell line'. This will be changed in the editor in the future.

7.1. BioSource/Organism

The main source for organism information is http://www.ebi.ac.uk/newt/
You can use this server to retrieve the NCBI Tax Id of the organism, and enter it in NCBI Tax Id field in the IntAct editor. This will then auto-complete the entry by assigning the Short Label and 'crossreference'. The 'crossreference' qualifier in this case is 'identity'.

7.1.1. Short Label:

The Short Label uses the UniProtKB organism identifier. Short Label is created by using the "auto-complete" button. This is derived from the genus and species of the organism by choosing the first three letters from the genus and first two letters from the species and fusing them. Example: Cercopithecus aethiops has a Short Label cerae. This nomenclature is followed for UniProtKB entry curation. However some of the model organisms do not follow this rule.
In cases where it is difficult to assign a child tax id you may end up with shortlabel followed by hyphen and number
Example: EBI-360031, EBI-625798 and EBI-763562, EBI-763617
Some standard species codes which do not follow the above rule are:
human
mouse
rabit
yeast

7.1.2. Full name:

Full name for the organism should be the genus and species followed by the common name, if any, in brackets. Example: Short Label canfa will have the full name *Canis familiaris* (dog).

7.1.3. Crossreference:

This should be database-NEWT with the ID being NEWT ID and the 'crossreference' qualifier set to 'identity'. The Secondary ID should also be the Short Label. The only exceptions to this are '*in vitro*' and 'chemical synthesis'. These do not have a 'crossreference'.

7.2. BioSource / Tissue type.

Add a new tissue type in CVTissue using **Editor-CvTissue**.

Take the tissue name from UniProtKB tissue list:
F:\SPROT\tisslist.txt or http://ca.expasy.org/cgi-bin/lists?tisslist.txt)

Select the tissue required. Make an entry in Editor-CVTissue by adding
Short Label: tissue name
Full Name: Description of the tissue if required.

Then make a new entry in the Editor-BioSource for the organism with this tissue type:
Make an entry for the organism as explained above.
Edit the BioSource Short Label by adding the tissue Short Label using a dash. If necessary, truncate the name so that the entire name does not exceed 20 characters.
Examples:
human-liver
canfa-brain-cortex
bovin-colon-smooth-m

When specific cell types are used e.g. embryonic fibroblast check if this is part of the tisslist.txt mentioned above. If it is part of the list as in this case please use the nomenclature used therein. Embryo becomes a foetus when organogenesis sets in and the dpc (days post coitum) of the organism should be considered when using these terms. When the specific cell type is not part of the list use the closest term from the tisslist.txt and add the detailed description in the 'exp-modification' annotation on experiment.

7.3. BioSource/ CellType.

Add a new cell line in CVCellType as follows:
1. Get the information on the cell line from the CABRI catalogue:
http://www.cabri.org/

This can be searched in external SRS under Biological Resources Catalogues (bottom of the page) from the following link:
http://srs.ebi.ac.uk/srsbin/cgi-bin/wgetz?-page+top

The categories below can be used in IntAct.

Animal and Human
DSMZ_MUTZ
ECACC_CELL
ICLC

Bacteria
CABI_BACT

Yeast
CABI_YEAST

Plant
DSMZ_PLANT_CELL

In the case of a specific bacterial strain, take the strain name from UniProtKB species list: F:\SPROT\speclist.txt or http://ca.expasy.org/cgi-bin/lists?speclist.txt)

7.3.1 Editor-CVCellType

Short Label: Use the cell line name from the CABRI database as the Short Label.
Full Name: The full name must have the organism and cell line name with a concise description of the cell line.
Annotation Topic: Description: This should be used for involved description of the cell lines if necessary.
Crossreference: The full cell line identifier must be added to the 'crossreferences' and the crossreference qualifier in this case is 'identity'. Example: ECACC 84121901 and CABRI must be added under database.

If a cell type is derived from a tissue this is a primary cell culture and the entry is made by entering the tissue type and then creating the Biosource using the Biosource editor. Where a further differentiation of the tissue is done this should be entered using the Tissue-editor if possible.

7.3.1.1 Biosource with cell line:

Create a new BioSource for the organism as described above. Do NOT edit an existing BioSource of the appropriate species.

Short Label: To the edit the BioSource Short Label by adding the cell line Short Label using a dash. If necessary, truncate the name so that the entire name does not exceed 20 characters.
Select the Cell line required from the CVCellType list.

Crossreference the NEWT reference for organism and CABRI reference for Cell line as identity.

Full Name The BioSource full name should be entered by repeating the genus/species name for cell lines and tissue types and then adding the tissue name or the derivation of the cell line.
Example: Homo sapiens spleen
Homo sapiens HeLa: epitheloid cervix carcinoma cells

Note: Start the species name with upper case, unless there's a good reason not to. Do not add a full-stop at the end of the full name.

Examples:
mouse-3t3
human-u-937

7.3.1.2 Cell lines without a CABRI reference but with an ATCC number:

These should be entered as above
annotation topic : 'url' the search URL of ATCC 'http://www.lgcpromochem.com/atcc/'should be entered.
No Crossreference is present in this entry.

7.3.1.3 Cell Lines with a PubMed reference:

Under the rare conditions that the Cell Lines are cross-referenced only in a publication, the cell lines should be entered as above with a 'crossreference' to the PubMed ID as a 'primary reference' qualifier.

7.3.1.4 Non-standard

Invent a long and short name and send this plus description of cell line to the IntAct curator's mailing list
intact-curators@ebi.ac.uk, then a decision can be made and the information can be integrated into the database and the curation rules. Please regard this as the last resort.

7.3.1.5 Re-classified cell lines

If a cell line has been re-classified or proved to be contaminated it must be annotated to the corrected BioSource.
Example: In EBI-367374 (PMID:15109305) the authors state that KB cells were used in the experiment. Subsequently these cells have been re-classified as a HeLa subclone (See definition in CABRI database). Therefore the BioSource for this experiment must be entered as HeLa cells and a note in the annotation under 'CAUTION' to explain the discrepancy.

7.4. Where BioSource is a compartment

The organism and tissue to which the compartment belongs is entered in the Host organism on the experiment page. The compartment of interaction itself is cross-referenced using the GO: Cellular Component on the interaction. Translocation of the interacting proteins between two cellular components are also covered using the GO translocation terms. Where appropriate translocation terms are not available use the translocation start and translocation end using the GO terms for the to and from locations. Please request the term on GO sourceforge site and replace the translocation start and end with this term when available.
Example:
GO:0007184: SMAD protein nuclear translocation
GO:0019066: viral translocation

If the interaction exist in multiple components, the GO ID for the cellular components should be included in the cross-references.

8. Editor-Protein

This form is used to record information about a peptide or protein when any of the following conditions apply:

- The protein is synthetic and does not have a biological equivalent with a UniProtKB accession number assigned to it.

- The protein is a chimeric protein which cannot be described adequately within the feature editor. Such as a chimeric protein made up of two (or more) peptides which interacts with other proteins or peptides.

- The protein does not have an available sequence at present. Such as a protein recognised by an antibody specific for a homologous protein.

- The peptide originates from an mRNA but does not have a UniProtKB accession number, e.g. a peptide originating from a polypeptide.

8.1 Short Label.

The short label should wherever possible use the protein name of a homologous protein (ortholog or paralog). In cases where this is not possible, for example the peptide is synthetic with no biological counterpart having a UniProtKB accession number, the identifier used by the author of the paper should be used as the short label. As a last resort an identifier consisting of the word pep and the first 17 amino acid single letter code characters should be used. In the latter case the short label should take the form:

pep your amino acid sequence (lowercase).

The short label should be no more than 20 characters in length and be in lowercase. Numbers, letters and the following special characters are permitted (hyphen (-),

underscore (_) and space). Any other special character should be replaced by an underscore (_).

8.2 Full Name.

The full name is entered as free text and should be the full name of the protein homologue or name used by the author. The full name section can also contain information relevant to the protein or peptide but this information should be kept as concise as possible and further information could be entered as a comment in the annotation section.

8.3 Source

The source of the protein should be selected from the drop down list. If the source organism of the peptide or protein is not within the drop down list the Bio-source editor can be used to create an entry for it. If the protein or peptide has multiple biological sources each fragment (range and source) should be described in the annotation section as a comment.

8.3.1 Chemical synthesis.

The peptide has been synthesised through a series of physical and chemical manipulations usually involving one or more chemical reactions *in vitro*.

8.3.2 Chimeric protein

The peptide or protein has been produced by a splicing together of two or more complete or partial DNA or protein sequences to produce a chimeric or mosaic protein.

8.3.3 UniProtKB peptide chains

For proteins or peptides that are annotated in a UniProtKB entry under Features PEPTIDE. These should be manually created and annotated to the limit of available identifiers, as follows:
- Short Label: UniProtKBID-FTID (lower case)
- annotation: isoform-comment if interesting
- xref: database – uniprotkb, primaryid – UniProtID-FTID, reference-qualifier - identity

8.4 Sequence

As long as the protein meets at least one of the criteria set out in section 5.8, any translated sequence (EST, cDNA, synthetic, chimeric and so on) can be entered into the sequence section. The sequence should be entered using single letter code for amino acids in uppercase.

The protein sequence should be added in raw format. Any sequences which are over 20 amino acids in length should first be analysed by InterProScan to determine if further information can be obtained about the protein sequence. This information can be entered as a comment, for example a crossreference to InterPro via an InterPro accession number.

Any sequence error(s) should be noted under the caution term in the annotation section (as a sequencing error or polymorphism)

Note: If a protein is entered via the Protein Editor and has no accompanying sequence, numeric feature ranges should not be used to describe domains, PTMs etc. even if given in the paper. Non-numeric characters "n", "c" and "?" may be used, numeric information should be stored in a Comment until a sequence becomes available. A cross-reference to InterPro may be added.

For example a protein may have no sequence either in UniProtKB or written out in the relevant paper, but the authors may state they have sequenced the protein and that the binding site is an SH2 domain. In such a case, the feature range will be "?-?" and the database cross reference will be IPR000980.

8.5 Annotation

The annotation topic is added as free text and while there is no upper limit regarding length of text it should be as informative yet as concise as is possible.

8.5.1. caution:

This is a statement providing a warning about errors and or grounds for confusion; for example possible sequencing errors in the sequence provided by the authors. This topic is added on interaction

8.5.2. comment:

Comments are entered as free text. This section provides additional useful information concerning the protein or peptide which cannot be accommodated under the other annotation topics. This topic is added on interaction.

8.5.3 function:

This topic is added on interaction.

8.5.4 isoform-comment: This topic is added on interaction. If this is a new isoform – submit entry to SPIN and update when UniProtKB Ac available (remark-internal).

8.5.5 copyright:

This topic provides information relevant to copyright statements attached to peptide or proteins (e.g. patents). This topic is added on interaction.

8.5.6 remark-internal:

This topic records any internal remark for the IntAct database team and will normally not be shown to the public. This topic is added on interaction.

8.5.7 submitted:

This topic is added on interaction.

8.5.8 on-hold:

This topic is added on interaction.

8.5.9 no-uniprot-update:

This annotation topic is added on all entries made through the protein editor. This also means that there will be no update of the protein using the UniProtKB entry AC. The range and sequence specific features will be as entered with respect to the sequence entered in the protein editor.

8.6 Cross-reference

Cross references should be added to provide additional information about the peptide or protein. In addition to those cross reference sources already listed in the drop down list, the following databases should be cross referenced, ENSEMBL (e.g. protein without an assigned UniProtKB AC number) , UniParc (e.g. deleted protein), EMBL (e.g. cDNA with no corresponding translated sequence) and model organism databases as they become available in future versions of the editor.

9. Special cases:

9.1. Annotation of protein complexes.

A complex that can be annotated from a pubmed reference should be annotated as a normal interaction with use of annotation topic complex-properties to describe the function and characteristics of the complex. Only information that describes the composition of a biological stable and functional complex should be added to IntAct database as a complex. Data that describes proteins interacting with a complex must not be annotated yet. The IntAct data model supports entry of these data but this has not been implemented yet.
GO will handle the nomenclature of protein complexes while IntAct will provide experimental data, which demonstrates the purification of specific complexes.

9.1.1. Complex definition:

A stable set (i.e. two or more) of interacting proteins can form a complex. It should be demonstrated that these co-purify, and should be shown to exist as a functional unit *in vivo*.
Note: Complexes are dynamic therefore their component parts may vary. They must be defined in context using the annotation topic complex-properties.

Example: Sanders et al (PMID 12138208) purified two different forms of the TFIID complex, with and without TBP, from *S. cerevisiae*.

See Experiment EBI-73668 and interactions EBI-73685 and EBI-73670

9.1.2. Partially characterised complexes.

When a complex is purified but only a few of the subunits are identified; only the identified subunits can be entered into the database. An annotation under complex-properties may be added to indicate the total number of subunits identified by the experimental method used. These can be crossreferenced using database: reactome-complex.

9.1.3 Organism

It is important to add the organism on the interaction if you think the interaction forms a complex. This will then be used as the organism of origin of the complex. If the complex has been purified *in vivo* the organism of origin of the proteins will be the organism of the interaction. If proteins from multiple sources have been used the organism of interaction should be *in vitro*.

9.1.4. Annotation of protein complexes using GO terms.

GO terms **relating to the complex** can be added to the interaction cross-references from any of the three GO catagories:

Molecular Function
Molecular function GO terms can be added if the function of the complex is proven within the paper.
Examples:
a) GO:0005154 if the complex binds to epidermal growth factor receptor.
b) GO:0008094 if the complex has DNA-dependent ATPase activity

Cellular Component.
Complex name
1. If an appropriate name for a complex already exists in GO, use this name in the Interaction Full Name and cross reference to the GO entry in the database cross-reference section of IntAct.

2. If there is no name for the complex in GO choose something appropriate, preferably from the literature, and add this complex name to the annotation topic complex-properties. Add the Remark "Revisit when GO term available" to your entry. Send a request for the complex name to GO using GO SourceForge

GO SourceForge: [https://sourceforge.net/tracker/?group_id=36855]

The GO sourceforge request should be submitted using the heading "Complex nomenclature – xxx" where xxx is your suggested name. The request must include the IntAct interaction accession number, a detailed description of the complex and information regarding its function if available. Add supporting information to your request such as PubMed ID's of relevant papers and any other useful experimental information you have read. Example: the complex may have been shown to comprise alternative subunits under differing conditions. In cases like this, the GO team may add children to the parent entry.

Sub-cellular location of complexes.
Use terms from the GO component ontology for sub-cellular location where it is known. If the location of the complex is inferred in the paper from the location of its participants, add a comment in the annotation field. Note: Sub-cellular location is inferred rather than proven.

Remark: When there is a GO component ontology term referring to a known, well-characterised complex, this GO term may have proteins, which are not components of the complex mapped to it, but their location of function is at the complex.

Biological Process.
Use terms from the GO biological process ontology that describe any process that the complex is involved in Example: apotosis. In the interaction cross-references add 'GO' to the 'database field, and then add the appropriate GO ID. Click 'add cross-reference', then the relevant GO ontology will be generated automatically in the 'Reference Qualifier'.

9.1.5. Additional Literature References

References containing additional information about the complexes may be added with the appropriate reference qualifier, for Example Function. Consider adding this information as an additional experiment if evidence of interaction/complex formation is given in the paper.

9.2 Importing interaction data from other databases.

If an interaction is imported from another database (Example: KIAA and Riken) put the accession number or reference number **for that interaction** from the source database in the interaction annotation cross-references. **Do not just add the URL of the source database.** Use the parent database short label for the interactions.

9.3. Cell surface protein-protein interactions

These include receptor binding interactions, cell surface proteins and interaction of extracellular domains.
a) receptor-protein ligand interactions where the ligand is extracellular are classed as '*in vitro*'.
b) interactions in which the interacting proteins are membrane-bound and the interaction is through their extracellular domains (Example: by FACS analysis) are classed as '*in vitro*.' The environment in which the interaction is taking place is outside the normal cellular environment i.e. in a tube in a buffer solution.

Example:

a) Demonstration of the interaction of a receptor with a protein ligand by cross-linking and co-immunoprecipitation:

In Noguchi *et al*. EBI-80473, Fig 1C, the IL-7 receptor was radio-affinity labelled in Cos-7 cells, proteins were cross-linked with a homo-bifunctional reagent before cell lysis, then an anti-IL-7R antibody was used to co-immunoprecipitate IL-7R and IL-7.

The experiment should be entered as follows:
Host organism: *in vitro*
Interaction detection: anti bait coip (MI:0006)
Participant detection: predetermined (MI:0396)
Further information about the experimental method can be added as below:
exp-modification : Proteins were cross-linked before cell-lysis.

The information about the radiolabel tag should be added as a feature of the radiolabelled protein.

Remark: In this case the interaction detection could be entered as 'cross-linking by homo-bi-functional reagent (MI:0031)' or as 'anti bait coip (MI:0006)'. This judgement must be made on a case by case basis by the curator.

The interaction information should be entered as follows:
Interaction type: aggregation
Proteins:
IL-7R (P16871)
Role : bait
Expressed in cerae-cos_7

Prey: IL-7 (P13232)
Role : prey
Expressed in cerae-cos_7

Feature type: radiolabelled, Short Label: region, range: ?-?.

b) Demonstration of interaction using cross-linking and 2D gels.

Noguchi *et al.,* EBI-368460 also demonstrated the IL7-IL7R interaction using cross-linking and 2D IEF gels in cos-7 cells. IL7 was radio-labelled and cross-linked to cos-7 cells expressing IL7R. Cell lysates were analysed on 2D gels.

Experiment:
Host organism: *in vitro*
Interaction detection: cross-linking (MI:0030)
Participant detection: predetermined (MI:0396)

Feature for radiolabelled proteins is also added.

Further information about the experimental method can be added as below:
Comment: samples were run on 2D IEF gels after cross-linking and cell lysis.

In this case the interaction detection is cross-linking. The samples were boiled in detergent (therefore denatured) before 2D IEF electrophoresis therefore the interaction is shown by cross-linking and not by electrophoresis.

9.4. Interaction detection by Far-Western.

If a Far-Western was used to detect an interaction, the protein probe is the bait and the protein interacting with it is the prey.

Example: EBI-593412
A purified complex was used in a Far-Western with a radioactively labelled protein as the probe. The protein that was detected in the complex as interacting with the bait by Far-Western was identified by Western blot.
This was entered into IntAct as follows:
Interaction detection: Far-Western
Participant detection: Western

In this case the bait was made by *in vitro* transcription and translation i.e. 'predetermined' this is entered as a comment.

9.5. DNA-protein interactions.

Do not enter DNA-protein interactions when one protein is shown to bind directly to nucleic acid. This type of interaction will may be dealt when DNA and small molecules is implemented. At the moment IntAct holds protein-protein interaction data and test case nucleic acid – protein interactions.
However, techniques that involve DNA-binding but demonstrate the interaction of two or more proteins can be entered Example: supershift assays, fluorescence anisotropy.

Example: Experiment EBI-371686 and Interaction EBI-371689.

In this experiment, purified TBP was bound to a TATA box sequence, and then purified Mot1 was added. The interaction between Mot1 and TBP was detected by the change in DNA anisotropy when Mot1 was added to TBP pre-bound to DNA.

9.6. Data not shown

Comment: In this experiment the authors claim that Rsc2 is also co-immunodepleted but the data is not shown.
These data should not be entered in the database as interactions. A comment as shown in the Example: above can be added so that it is clear that the authors have claimed but not demonstrated an interaction.
Caution: In this experiment the authors claim that the interaction could not be demonstrated by alternative method.
If the authors have only mentioned the data not being shown by alternative method this goes in a caution, not as a negative interaction.

10. Curation checking procedure

The curation checking is carried out in two stages. The first stage is using automated sanity checks to check that the rudimentary information present in the entry follows the basic curation rules. A second stage of checking involves a senior curator checking the entries for their content with respect to the article being curated. If experiments and linked interactions of an article are curated as per rules and are satisfactory then the experiments get an annotation: accepted. If there is a problem with the curation of the experiment or the interaction the experiment will get an annotation: to-be-reveiwed. The curator will be explained why the experiment was not accepted and this may be further discussed or changed. Once the changes are made the entry is checked again and an annotation of accepted or to-be-reveiwed may be again added to it. In the case of problematic entries two or more senior curators will discuss the specific example.

11. Talisman

http://www3.ebi.ac.uk/%7Esp/intern/projects/intact/local/doc/

This is an internal tool which can query the production database. There are a list of queries that you can perform:

Number of interactions
Number of experiments by curator (Incorrect due to bulk database updates by programmers)
Number of interactions by curator (Incorrect due to bulk database updates by programmers)
All experiments with remarks
All interaction with remarks
Proteins with Interactions by species
All Experiments with this PubMed id

All Interactions with this bit of annotation (can be slow)
All Experiments being annotated: 'on hold'

Some interactions/experiment you added may not be displayed as the ‘userstamp’ corresponds to the last user saving the entry (by pressing ‘submit’ or ‘save and continue’ with editor or by editing in SQL).

You can also write your own SQL queries:

12. PubMed Lookup

http://www3.ebi.ac.uk/%7Esp/intern/projects/intact/local/doc/

This tool helps track the PubMed ID’s that are ‘done’, ‘in progress’, ‘incomplete’ or ‘dumped as irrelevant’. It is essential that you check the paper you are about to curate by performing the ‘Fetch’ option on the PubMed ID. If the paper is already curated you will get back the experiment Short Label and the Intact AC of the experiments from this paper; If this paper is being curated by another curator this can also be checked; if this paper has been deemed unsuitable at an earlier date this will also show up. Please do not curate the paper if any of the above show up. There is also an option to suggest papers that you think should have an entry in IntAct.

13. Bibliography

1. Noguchi, M., *et al*., Interleukin-2 receptor gamma chain: a functional component of the interleukin-7 receptor. Science, 1993. 262(5141): p. 1877-80.
2. Sarkar, F.H. and S.L. Gupta, Receptors for human gamma interferon: binding and cross-linking of 125I-labeled recombinant human gamma interferon to receptors on WISH cells. Proc Natl Acad Sci U S A, 1984. 81(16): p. 5160-4.
3. de Gunzburg, J., R. Riehl, and R.A. Weinberg, Identification of a protein associated with p21ras by chemical crosslinking. Proc Natl Acad Sci U S A, 1989. 86(11): p. 4007-11.
4. Subramanian, C., *et al*., In Vivo Detection of Protein-Protein Interaction in Plant Cells Using BRET. Methods Mol Biol, 2004. 284: p. 271-86.
5. Bouwmeester, T., *et al*., A physical and functional map of the human TNF-alpha/NF-kappa B signal transduction pathway. Nat Cell Biol, 2004. 6(2): p. 97-105.