IntAct Annotation Rules (January 2005).

Index

2.5 Interaction Type
2.6 Organism
2.7 Experiments
2.8 Proteins
2.8.1 ‘Short Label’, ‘SP AC’, ‘organism’ and ‘Intact AC’
2.8.2 protein accession numbers
2.8.2.1 How to find a protein accession number
2.8.2.2 Problems mapping accession numbers
2.8.2.3 Incomplete protein sequences/ protein without UniProt accession number
2.8.2.4 UniProt entries that contain multiple protein molecules
2.8.2.5 Natural peptides
2.8.2.6 Bi-cistronic expression
2.8.2.7 Synthetic molecules
2.8.2.8 Synthetic peptides
2.8.2.9 Identification of proteins by sequence similarity
2.8.2.10 Proteins of unknown species
2.9 Protein roles
2.9.1 bait/prey
2.9.2 agent/target
2.9.3 neutral
2.9.4 self
2.9.5 complex
2.9.6 unspecified
2.10 ExpressedIn
2.10.1 Heterologous protein expression
2.10.2 *In vitro* transcribed and translated proteins
2.10.3 Synthetic peptides
2.11 Stoichiometry
2.12 Interaction annotation
2.12.1 agonist
2.12.2 antagonist
2.12.3 author-confidence
2.12.4 caution
2.12.5 comment
2.12.6 disease
2.12.7 function
2.12.8 inhibition
2.12.9 stimulation
2.12.10 isoform-comment
2.12.11 kinetics
2.12.12 negative
2.12.13 pathway
2.12.14 remark-internal
2.12.15 resulting ptm
2.12.16 search-url
2.12.17 search-url-ascii

IntAct data model

The IntActmodel has one publication with one to many experiments, one experiment with one to many interactions, one interaction with one to many proteins and one protein with one to many features.

All interactions from each publication should be incorporated in their entirety before releasing the data on the public server. Before curating a paper please check its status using the PubMed Lookup tool (section 6) http://www3.ebi.ac.uk/%7Esp/intern/projects/intact/local/doc/. IntActrelease 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 1.4), when the e-mail is circulated about the release.

Editors:

IntAct data entry is made using editors. There are separate editors for recording Experiments and Interactions, creating a BioSource, creating Crossreference databases, and adding annotation topics. The editor to be used should be selected from the drop down list. The editor is available for internal curation and needs a login and password. A search can be performed within the editor using either the Short Label or the IntAct accession number (EBI-number); wild card (*) can be used in the search. Unique IntAct accession numbers and Short Label are mandatory in the Editor form. All the mandatory fields are colored green on each of the Editor forms.

There are 2 database instances: ZPRO and DOO2 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 (D002). The test editor is used to test new updates of the editor before putting them on production editor. Test editor is also used for training new IntAct curators. If you are not sure with 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.

During the release, the data on the production database is first transferred into the test server as a back-up. Thus, while the release is in progress, you should not use the test editor. The production 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-mails notification of the availability of the production editor database and test editor database will be sent.
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. In that case to unlock it to go back to the entry and press 'cancel' or 'submit' if you made any changes.

1. 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 1.8.1.5 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 o 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 Acc no	**Experiment Acc no**	**Bait UniProt acc**	**Prey UniProt acc**	**Interaction detection method**
EBI-78422	EBI-78402	P05067	Q9UQF2	Yeast two hybrid
EBI-78430	EBI-80340	P05067	Q9UQF2	Pull-down

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

1.1. Intact AC

Every IntAct object for example experiment, interaction, interactor, feature, annotation, biosource, CV object etc. gets an Intact ACcession 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 search from the editor pages.

1.2. Short Label

Remark on permitted characters for 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. If there are any other special characters these should be replaced by underscore. There is a character limit of 20 on the Short Label.

An experiment Short Label has the following format: first author surname-year-experiment number for that paper. The author names may need to be abbreviated to accommodate the 20 character limit. When an author has published more than one paper in the same year, enter year a, b, c etc. as illustrated below.

oehrl-1998-1
chan-2003-2
chan-2003-3
chan-2003**a**-1

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

1.3 Full Name

The full name is free text and is normally the title of the paper being curated.

1.4 Host organism

The Host Organism is the organism the experiment has been performed in. This information also includes the information about the tissue or cell line if available. The 'Host Organism' is *in vitro* when the interaction has not been shown to occur *in vivo*. Please see the section on Editor-BioSource for how to create the drop down list of 'Host organism' (section 4).

1.4.1 *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]

2. Two cells each expressing a cell-surface molecule which are demonstrated to interact example: by FACS.

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 'expressed in' section on the interaction page. See 2.5.2.

1.4.2 *In vivo* interactions
All of the interacting proteins must be expressed **in a cell** or cell membrane; 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 on Editor-BioSource (Section 4).

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 BRET (bioluminescent resonance energy transfer) [4].

1.5. Interaction Detection

The experimental method used to detect the interaction example: co-immunoprecipitation, gel filtration, yeast two-hybrid.

See controlled vocabulary.
http://www3.ebi.ac.uk/~sp/intern/projects/intact/local/data/controlledVocab/

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 'interaction detectio' from the drop down list.

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

Method	**Experiment no.**
co-immunoprecipitation (in vivo)	EBI-74784
anti tag coimmunoprecipitation (in vivo)	EBI-465449
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
co-localisation by immunostaining	EBI-82041
pull down (in vitro)	EBI-457795
gst pull down (in vitro)	EBI-457587
gst pull down (in vivo)	EBI-456973
Tandem affinity purification	EBI-464830

isothermal titration calorimetry	EBI-456897
molecular seiving	EBI-457487
x-ray crystallography	EBI-457207
fret	EBI-456897
nuclear magnetic resonance	EBI-458375
classical two hybrid	EBI-1296
triple hybrid: dehydrofolate reductase reconstruction	EBI-448165
beta galactosidase complementation	EBI-65752
yeast display	EBI-296471
phage display	EBI-444056
peptide array	EBI-371696
protein array	EBI-375112
colocalization by immunostaining	EBI-82041
surface plasmon resonance	EBI-456914

Please use as specific a term as possible from the hierarchy of the controlled vocabulary. Example if the interaction was detected by a GST pull down, the term 'gst pull down' should be used over 'pull down'.

1.6. Participant detection

This is the method by which the participants have been identified example: protein sequencing, nucleotide sequencing.
See the Participant Detection controlled vocabulary**.**
Additional controlled vocabulary terms are pending for various methods as shown below. See section 1.5 for request for new controlled vocabulary terms.
See controlled vocabulary.
http://www3.ebi.ac.uk/~sp/intern/projects/intact/local/data/controlledVocab/

1.6.1 Electrophoretic methods

Example: SDS PAGE, non-denaturing gel electrophoresis (heterodimer expts), 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.

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 'Participant Detection' select 'participant detectio' from the drop down list.

1.7 Interactions

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 need to go to the Editor-Interaction to delete the interaction. 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.

1.8. 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.

1.8.1 Annotation topics used while curating an experiment:

1.8.1.1. 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.

1.8.1.2. comment:

'comments' are free text, which can be used to describe additional information about the experimental method which cannot fit under other annotation topics. It is desirable that the comments are restricted to as few as possible.

1.8.1.3. confidence-mapping:

Description of author-confidence mappings in the experiment. If more than one system of confidence mapping is used by the author, these will be described here.
In large-scale experiments when 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 UniProt 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 annotator decided that only those interactions with the confidence values Core-1 and Core-2 should be exported to UniProt.
The annotation of this experiment therefore is as follows:
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. The corresponding 'uniprot-dr-export' is shown in section 1.8.1.9.

1.8.1.4. 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.

Examples: EBI-448095, EBI-457795, EBI-457377

1.8.1.5. negative

This is used as shown below for annotation of experiments that demonstrate that an interaction does not occur, 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 (See example cases below).

2. Only experimental data in the same paper 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 insufficient evidence. (Interaction data refuting a previously shown interaction by the same scientific group can be entered as negative interaction if it is a well controlled experiment.)

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:
Database: intact
Primary Id: The Intact AC number of the experiment demonstrating the negative interaction.
Reference 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.

Case1
PMID:123456

Interacting proteins	Experimental method	Interaction Positive/negative	Experiment Short Label
A and B	Y2H	Positive	Harwood-2004-1
A and B	Co-IP	Negative	Harwood-2004-2
C and D	Co-IP	Positive	Harwood-2004-3

These data would not be entered as below:

Case 2

PMID:123456

Interacting proteins	Experimental method	Interaction Positive/negative	Experiment Short Label
A and B	Y2H	Positive	Harwood-2004-1
A and B	Co-IP	Negative	Harwood-2004-2
C and D	Co-IP	Positive	Harwood-2004-2

Case 3

PMID:123456

Interacting proteins	Experimental method	Interaction Positive/negative	Experiment Short Label
E and F	Y2H	Negative	Harwood-2004-4

There is no supporting positive interaction information for these data so this must not be entered.

For an example: of an entry containing complementary positive and negative interaction data in IntAct, see the following:

In Battle 2003: PMID 12809483
Positive interactions shown by yeast two hybrid.
Experiment Ac: EBI-296688
Interaction Ac: EBI-296713
Ac: EBI-296720
Ac: EBI-296734
Negative interactions shown by co-immunoprecipitation.

Experiment Ac: EBI-353694
Interaction Ac: EBI-353709

Ac: EBI-353714

Ac: EBI-353721

1.8.1.6. on-hold

Any experiment that is annotated with this topic will not be released to the public. This annotation also can be used to exclude 'experiments in progress' from publication, but this should be used only 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 5).
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 system gradually will be extended into a full submission management system.

1.8.1.7. 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. Some of the 'remark-internal' used in the experiment page are:

Revisit when Electrophoretic Method in place.

Revisit when Co-localization in place.

Revisit when CV in place.

1.8.1.8. submitted:

This refers to the data submitted by the author which is not yet in the public domain.

1.8.1.9. uniprot-dr-export:

Determines whether the interactions of an experiment should be exported to UniProt DR lines. This can be according to a setting in the CvInteraction controlled vocabulary, or by an explicit setting for an experiment. For experiments: If set in an experiment, the setting overrides any setting derived from the CvInteraction controlled vocabulary. 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 UniProt.

In the example on confidence-mapping
uniprot-dr-export will be CORE_1
uniprot-dr-export will be CORE_2

1.9. Crossreferences

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’.
"see-also" should be used where there is a relevant paper describing the experimental technique used. It should not be used for information about the interactions.

2. 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.

2.1. Intact AC

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

2.2. Short Label

Short Labels are lower case and should not exceed 20 characters in length.

2.2.1. 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.

2.2.2. 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 **t**aken from the UniProt entry) of the first two proteins in the interaction. If there are more than two proteins in the interaction, only two gene names should be chosen. If there is a bait, 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 UniProt GN line, the first name is taken.

Format: gn1-gn2-1

2.2.3. 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
gn1-gn2-3

2.2.4. If there is no gene name then the first part of the UniProt/Swiss-Prot entry name is taken Example: for HTL1_YEAST, take HTL1.
Format: gn1-htl1-1

2.2.5. For a TrEMBL entry where there is no UniProt entry name, (i.e. the ID line contains the accession number) the accession number is used.
Example: gn1- q03940-1

2.2.6. Short Labels and incorporation of special characters
In the GN line of Drosophila UniProt entries, CG numbers are often used instead of a gene name. In this case, use the CG number in the Short Label. Where there are characters in the UniProt GN line, which are not accepted in the IntAct short-label field, use underscore to substitute for the character. Any special character is replaced by '_' for all species.

Example: of UniProt GN line containing GN numbers:
GN CG32767 {EI2} or CG5086 {EI2} or CG5113 {EI2}.
Use the first number from the GN line in the Short Label Example: cg32767-cg15119.

Other Example:s of Drosophila GN lines and derived Short Labels:
BETA-SPEC becomes beta_spec
EG:95B7.3 becomes eg_95b7_3
Su(Var)2-10 becomes su_var_2_10

2.3 Full Name

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

2.4. 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 0.000000001 or 1e-9 for 1.10^{-9}
Enter 0.02 or 1e-2 for 2.10^{-2}
Enter 1300 or 1.3e+3 for $1.3.10^{3}$

Additional information about the kinetics will be added in annotation topics. 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' (Section 2.12.11).

2.5 Interaction type:

The interaction type defines the type of interaction Example: physical, enzymatic, etc.

See controlled vocabulary.
http://www3.ebi.ac.uk/~sp/intern/projects/intact/local/data/controlledVocab/
Please refer to section 1.5 on additional CV terms.

2.6 Organism

The organism is automatically filled in from the experiment page. The organism is important on the Editor-Interaction page only in the cases where the interaction defines a complex.

2.7 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.

2.8 Proteins

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

These are imported directly from the UniProt entry. To pull up the entry you would need the Short Label or the UniProt ID line or the UniProt accession number. 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.

The search for proteins can be done on UniProt ID line or UniProt accession number or Intact AC number. 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. Please read section (2.10.1) for annotating complexes.

2.8.2 Protein accession numbers

2.8.2.1 How to find a protein accession number

1.Try searching the following databases with the accession number/identification number
a) UniProt
b) UniParc

2. 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 will be archived in UniParc in future.

3. 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.

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

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

6. 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. He deals with the IPI database. 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).

7. If you can not link an identifying number to a sequence then search using the gene or protein name in

a) UniProt
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.

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

2.8.2.2 Problems mapping accession numbers

If you cannot map a protein identifier given in a paper to a UniProt ID:
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 UniProt 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.

2.8.2.3 Incomplete protein sequences/proteins without UniProt accession number.

If a partial protein sequence is given in the paper but a corresponding sequence cannot be found in UniProt contact Paul Browne (browne@ebi.ac.uk) and ask him to generate a UniProt 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 UniProt using SPIN, and an accession number will be sent to you.

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

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

2.8.2.4. UniProt entries that contain multiple protein molecules

These molecules will be dealt with when the protein-sequence editor becomes available. Individual molecules will be constructed in the protein editor using information from the UniProt entry and any other available information. The molecule will be assigned an Intact AC number and the UniProt accession number from which the molecule was derived will be crossreferenced as the parent molecule. Examples: Bioactive peptides cleaved from full length parent molecules; viral polyproteins.

Example: 1. NFKB1

NFKB1 (KBF1) is in UniProt 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

The processed form (the p50 subunit) of NFKB1 does not have a separate identifier from the full-length protein in UniProt. 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.
Example: UniProt entry P26664 contains multiple viral proteins including NS1, NS2 and NS4A.

EBI-465775

2.8.2.5 Natural Peptides
The interaction may involve a natural peptide. Enter the UniProt ID of the parent protein **and the ranges with the feature editor** then add more precise additional information about the peptide in the interaction comments.
Example: APP beta amyloid peptide see the Example: in IntAct below.

Experiment	Interaction	PMID
EBI-79357	EBI-79361	10681545

See also section 2.6 Synthetic peptides.

2.8.2.6 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 by sequence, as the DNA sequence of the participating proteins was already known or 'pre-determined'. 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.

2.8.2.7.Synthetic Molecules

2.8.2.8 Synthetic Peptides
When a chemically synthesised peptide is a component of an interaction, add the UniProt accession 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.
Also add the remark 'Revisit when protein sequence editor in place.' so that you can go back and add the details of the peptide later.

Example: EBI-77665.

2.8.2.9. 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

2.8.2.10. 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 cannot be dealt with at present.

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.

2.9. Protein roles

See
http://www3.ebi.ac.uk/~sp/intern/projects/intact/local/data/controlledVocab/CvComponentRole.def.html

2.9.1. bait/prey

Bait must always have one or many prey(s). If there is an interaction with prey(s) there must be a bait.
The bait is in general the protein which is used to fish out a protein. This may be immobilised except for far western (section 2.10.4). The roles bait and prey are associated with Y2H, CoIP, TAP, Pull downs etc.

2.9.2 agent/target

An **agent** must always have a **target** (and vice–versa).
The agent is the enzyme which modifies the protein (kinase..) and the target is the modified protein.

2.9.3. neutral

Neutral component is used where the role of the protein cannot be classified as bait/prey or agent/target. Neutral roles are often observed with stoichiometry determination and oligomerisation.
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 components. The number of molecules interacting will be indicated by stoichiometry (section 2.7)

2.9.4 self

Only proteins that interact intra-molecularly should be given the role 'self' (Example:: autophosphorylation, di-sulphide 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.

2.9.5 complex

This field should be chosen in specific cases when the proteins are subunits forming a complex with a specific activity. See 2.9 for annotating complex.

2.9.6 unspecified

The role is added when you do not know the role played by the interactors.

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 explaination.

2.10. 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:

2.10.1 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.

2.10.2 *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.

2.10.3 Synthetic peptides

Please see section 2.8.2.8. The ExpressedIn is 'synthetic peptide'.

2.11. 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.

2.12. Interaction Annotations

2.12.1. agonist:

Add to CVtopic when one or both interactors are activated by specific agonist. Example: EBI-448589.

2.12.2. antagonist:

Add to CVtopic when one or both interactors are inactivated by specific antagonist.

2.12.3. author-confidence:

This is the confidence value assigned by the author to the interaction. Please see section 1.8.1.3 for the method used for confidence-mapping.

2.12.4. caution:

This is a warning about errors/ grounds for confusion example: significant differences between the sequence of the protein described in the paper and the Uniprot entry for the protein.

2.12.5. 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.

2.12.6. 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.

2.12.7. 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. You should also suggest a term to GO (http://sourceforge.net/tracker/?func=add&group_id=36855&atid=605890) and update once the term has been assigned.

2.12.8 inhibition:

Use this topic when the interaction being annotated is disrupted by biological entity or by modification of interactors Example: EBI-447320 , EBI-447720 .

2.12.9. 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 Example: EBI-74820 , EBI-465428

2.12.10. isoform-comment:

This topic is used to add comments about isoforms.

2.12.11. 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 3.2.5). For Example: : "Mutation Asp234Glu caused the Kd to become 0.0003M." kd of the parent molecule should be added to the kd field.

2.12.12 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.

2.12.13. pathway:

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.

2.12.14. remark-internal

'remark-internal' is not for public use. Interaction remarks are temporary annotation fixes for use until other tools are available. The only remarks, which should be used in the interaction page, are:
Requires updating in UniProt
Revisit when Complex Table in place
Revisit when uncharacterized participant in place
Use of comments and remarks
a) Comments about the UniProt entry should be accompanied by the remark 'Revisit when Swiss-Prot demerges' and/or ' Requires updating in UniProt' as appropriate.
b) Comments about well-characterized complexes should be added here with the remark 'Revisit when Complex Table in place'

2.12.15. resulting-ptm:

This annotation topic is used to record the effect of the interaction on the PTM. The PTM could be added or removed from molecules as a result fo the interaction.

2.12.16. search-url:

This is a search url for the CVdatabases. This should be used when a protein interaction is stored at a url and a search for interactions can be made on the page.

2.12.17. search-url-ascii:

This retrives the search in ascii format.

2.12.19. url:

The url of the database storing the protein-protein interaction data. For more information see the section 2.10. Please note that the interaction Short Label used by the database should also be crossreferenced to as 'see also'.

2.13. **Crossreferences**

There may be additional references or GO terms on the interaction level**.** References relating to the **protein** should be added on the Protein level using the Feature editor (eg. InterPro Cross-Reference); see 3.1.6.
Interaction shown by co-localisation (classical, electronic or confocal microscopy…) should have a cross-reference to a GO component term whenever it is possible.
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.
Protein complexes can be annotated at the interaction level using crossreferences. See 2.10.1.3.
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.

2.14. Special cases:

2.14.1. Annotation of protein complexes.

Only information that describes the composition of a biological stable and functional complex should be added to IntAct. 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.
IntAct will provide Examples of experimental data demonstrating the purification of complexes.

2.14.1.1. Complex definition:

A stable set (i.e. two or more) of interacting proteins, which can be co-purified and that has been 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.
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

2.14.1.2. **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. A comment may be added to indicate the total number of subunits identified.

2.14.1.3. 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 or speculated upon 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 Interaction Full Name. Add the Remark "Revisit when GO term available" to your entry. Send a request for the complex name to GO using GO SourceForge using the heading "Complex
nomenclature – xxx" where xxx is your suggested name. Mention the EBI number in the GO SourceForge request, so that you can add the GO term to the right IntAct complex once it has been created.

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

Add supporting information to your request such as PMID'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'.

When a complex has been annotated with GO terms, send the EBI accession number to Sandra Orchard (orchard@ebi.ac.uk) who will collate them as Examples for the GO curators.

2.10.1.4. 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.

2.14.2 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.**

2.14.3. 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*. [1], 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: co-immunoprecipitation (MI:0019)
Participant detection: radio-labelled protein (MI:0222)
Further information about the experimental method can be added as below:
exp-modification : Proteins were cross-linked before cell-lysis.

Remark: In this case the interaction detection could be entered as ‘cross-linking by homo-bi-functional reagent (MI:0031)’ or as ‘co-immunoprecipitation (MI:0019)’. This decision must be made on a case by case basis by curator judgement.

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

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

Noguchi *et al.,* [1] 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: radio-labelled protein (MI:0222)

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.

2.14.4. 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:
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.

2.14.5. 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 with later in the database. At the moment IntAct only holds protein-protein interaction data.
However, techniques that involve DNA-binding but demonstrate the interaction of two or more proteins can be entered Example: supershift assays, fluorescence anisotropy.

See 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.

2.14.6. 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.

3. 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. On submitting the features these will be visible on the interaction page.
Proteins: light blue title bar corresponds to table heading for features

3.1 The 6 fields

3.1.1 Short Label

This is mandatory but not necessarily 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 "hotspot" where it will be automatically generated. The Short Label should describe the feature. The end user views the features as a autogenerated sentence with the following format: 'Feature Type' 'Short Label' of 'Uniprot protein ID line' detected by 'Feature Identification'. This should be taken into consideration

while generating the Short Label of the feature. Some examples of shortlabels (bold) are: 'GST tagged **region** of cdc42_human [c-c], tagged protein **S35** of cdc42_human [?-?]. Some more examples are also shown with specific cases.

3.1.2 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.

3.1.3 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.
http://www3.ebi.ac.uk/~sp/intern/projects/intact/local/data/controlledVocab/

3.1.4 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.

3.1.5 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 arround aminoacid 10 and to region arrond aminoacid 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 aminoacids 10-13 and to region between aminoacids 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

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.

3.1.6 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.

3.2 Various cases

The Feature Editor should be used in the following cases:

3.2.1 Molecule is tagged

The common tags include HIS-, TAP-, HA-, FLAG-, and GST-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.

'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

3.2.2 Molecule is radio-labelled

Radio-labelling may be regarded as a form of tag.
Short Label – Use to describe the radiolabel Example: '35s' or '32p'.
FeatureType - 'tagged-protein'

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

3.2.3 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 UniProt 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)

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.

3.2.4 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.
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'
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).

Range:
Range defines the fragment of protein. This may be specific (2-10), fuzzy (2..-10..), 'c-c' for C-terminal fragment, 'n-n' for N-terminal fragment, and '?-?' where it has not been defined.
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 primary cross reference with the InterPro accession number. EBI-457906
When fragment boundaries are not known but the domain(s) is/are known, the boundaries for the domain found in UniProt entry should be used with fuzzy type (Example: From Range:10.., To Range:56..). There is a link to the UniProt entry when clicking on the protein Short Label. EBI-353537

3.2.5 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 hotspot see 3.2.6
- 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 3.2.8.

Example::

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

When there are several mutation 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. See 3.2.7.

3.2. 6 Molecule contains a Hotspot

A Hotspot is a residue or short motif which appears to be required for the interaction, though may not necessarily be an interacting residue or motif. These are defined by mutational analysis. These mutations may prevent the interaction from occurring but should be annotated positively.

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 Hotspot 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 – hotspot
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. See 3.2.7.

3.2.7 Mutation and hotspot 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.

3.2.8 Mutation which effects rate of an interaction

In cases where the mutation has an effect on 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.”

3.2.9 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’.

3.3 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 interacting proteins have ranges 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.

4. 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 pickedup from the UniProt 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.

4.1. 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'.

4.1.1. Short Label:

The Short Lable uses the UniProt 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. However some of the model organisms do not follow this rule.
Some standard species codes which do not follow the above rule are:
human
mouse
rabit
yeast

4.1.2. 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 Cannis familiaris (dog).

4.1.3. 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'.

4.2. BioSource / Tissue type.

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

Take the tissue name from UniProt 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

4.3. 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

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' Example: ECACC 84121901 and CABRI must be added under database.
Enter the Reference Qualifier 'identity'.

4.3.1 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 Lable: 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

4.3.1.1 Cell lines without a CABRI reference but with a 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.

4.3.1.2 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.

4.3.1.3 Non-standard
Invent a long and short name and send this plus description of cell line to the IntAct curators 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.

4.3.1.4 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.

4.4. BioSource compartment

The compartment of interaction is cross-referenced using the GO:Cellular Component. Translocation of the interacting proteins between two cellular components are also covered using the GO translocation terms.
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.

5. 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:

6. 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 a option to suggest papers that you think should have an entry in IntAct.

7. 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.