

The minimum information required for reporting a molecular interaction experiment (MIMIx)

Sandra Orchard¹, Lukasz Salwinski², Samuel Kerrien¹, Luisa Montecchi-Palazzi¹, Matthias Oesterheld³, Volker Stümpflen³, Arnaud Ceol⁴, Andrew Chatr-aryamontri⁴, John Armstrong⁵, Peter Woollard⁵, John J Salama⁶, Susan Moore^{6,7}, Jérôme Wojcik⁸, Gary D Bader⁹, Marc Vidal¹⁰, Michael E Cusick¹⁰, Mark Gerstein¹¹, Anne-Claude Gavin¹², Giulio Superti-Furga¹³, Jack Greenblatt⁹, Joel Bader¹⁴, Peter Uetz¹⁵, Mike Tyers¹⁶, Pierre Legrain¹⁷, Stan Fields¹⁸, Nicola Mulder¹⁹, Michael Gilson²⁰, Michael Niepmann²¹, Lyle Burgoon²², Javier De Las Rivas²³, Carlos Prieto²³, Victoria M Perreau²⁴, Chris Hogue⁶, Hans-Werner Mewes³, Rolf Apweiler¹, Ioannis Xenarios⁸, David Eisenberg², Gianni Cesareni⁴ & Henning Hermjakob¹

A wealth of molecular interaction data is available in the literature, ranging from large-scale datasets to a single interaction confirmed by several different techniques. These data are all too often reported either as free text or in tables of variable format, and are often missing key pieces of information essential for a full understanding of the experiment. Here we propose MIMIx, the minimum information required for reporting a molecular interaction experiment. Adherence to these reporting guidelines will result in publications of increased clarity and usefulness to the scientific community and will support the rapid, systematic capture of molecular interaction data in public databases, thereby improving access to valuable interaction data.

Deciphering the molecular mechanisms of cell function relies to a large extent on tracing the multitude of interactions between the numerous components of living cells, and between these molecules and any entity or compound of interest to the scientist, such as pharmaceutical agents or environmental contaminants. Molecular interactions may be direct, with two molecules in contact with each other, or the molecules may be in the same affinity complex, purifying together without a physical interaction between them. Several public databases strive to capture the ever increasing amount of published molecular interaction data, which are generated by a broad range of biophysical, biochemical, genetic or predictive methods. During the process of manual curation, the raw data are extracted from a published paper or from a submitted manuscript and systematically transferred into a database.

Initially, interaction databases such as BIND¹ and DIP² worked in isolation and according to their own internal standards and data formats. Because no one database can achieve complete coverage of all known molecular interactions, the user may need to download and combine datasets from two or more databases to answer a specific question. Until recently, this could not be done without first transforming the data into a common format, using a different parser for each database. In 2004, however, several major databases jointly published a community-standard data model for the representation and exchange of protein interaction data³. This data model, developed by members of the Molecular Interaction (MI) group of the Proteomics Standards Initiative (PSI), a work group of the Human Proteome Organization (HUPO)⁴, has already been adopted by major public interaction databases. Data sets can be downloaded from many of these databases in PSI-MI extensible markup language (XML) interchange format and further analyzed using a number of PSI-MI compatible tools, such as Cytoscape⁵, ProViz⁶ and PIMWalker⁷.

Building on the PSI-MI standard, several public interaction databases have formed the International Molecular Interaction Exchange consortium (IMEx; <http://imex.sf.net>). The consortium, originally founded by BIND¹, DIP², IntAct⁸, MINT⁹ and MPact (MIPS)¹⁰, has started to share the curation load and aims to regularly interchange data curated to the same common standards, in a manner similar to the well established pattern followed by the nucleotide sequence databases. However, the consortium's goal of achieving as near complete

¹European Molecular Biology Laboratory (EMBL) – European Bioinformatics Institute, Wellcome Trust Genome Campus, Cambridge, UK. ²UCLA–US Department of Energy Institute for Genomics & Proteomics, University of California, Los Angeles, California, USA. ³Institute for Bioinformatics, Forschungszentrum für Umwelt und Gesundheit – National Research Center for Environment and Health, Neuherberg, Germany. ⁴Department of Molecular Biology, University of Rome Tor Vergata, Rome, Italy. ⁵GlaxoSmithKline R&D, Stevenage, UK. ⁶Blueprint Initiative, Samuel Lunenfeld Research Institute, Ontario, Canada. ⁷National University of Singapore, Clinical Research Centre, Singapore. ⁸Merck Serono International S.A., Geneva, Switzerland. ⁹Banting and Best Department of Medical Research, University of Toronto, Ontario, Canada. ¹⁰Center for Cancer Systems Biology and Department of Cancer Biology, Dana-Farber Cancer Institute and Department of Genetics, Harvard Medical School, Boston, Massachusetts, USA. ¹¹Molecular Biophysics and Biochemistry Department, Yale University, New Haven, Connecticut, USA. ¹²EMBL Heidelberg, Germany. ¹³CeMM Center for Molecular Medicine of the Austrian Academy of Sciences, Vienna, Austria. ¹⁴Department of Biomedical Engineering, Johns Hopkins University, Baltimore, Maryland, USA. ¹⁵Institute of Toxicology and Genetics, Leopoldshafen, Forschungszentrum Karlsruhe, Germany. ¹⁶Samuel Lunenfeld Research Institute, Mount Sinai Hospital, Toronto, Canada. ¹⁷Commissariat à l'Energie Atomique, Institut de Biologie et de Technologie de Saclay, Gif sur Yvette, France. ¹⁸Howard Hughes Medical Institute Department of Genome Sciences & Medicine, University of Washington, Seattle, Washington, USA. ¹⁹Institute for Infectious Disease and Molecular Medicine, University of Cape Town, South Africa. ²⁰University of Maryland Biotechnology Institute, Rockville, Maryland, USA. ²¹University of Giessen, Germany. ²²Toxicogenomic Informatics and Solutions, Lansing, Michigan, USA. ²³Cancer Research Center (Centro de Investigación del Cáncer, University of Salamanca and Consejo Superior de Investigaciones Científicas), Salamanca, Spain. ²⁴Centre for Neuroscience, University of Melbourne, Victoria, Australia. Correspondence should be addressed to S.O. (orchard@ebi.ac.uk).

Box 1 The MIMIx checklist

Example data are taken from ref. 21. This example may be seen as a completed checklist in **Supplementary Note 2** online and as a MIMIx-compatible submission in Excel, XML and HTML format at <http://imex.sf.net>. The full paper, annotated to the richer IMEx standard, can be viewed in the IntAct database (<http://www.ebi.ac.uk/intact>) using accession numbers EBI-958406, EBI-958452, EBI-958498 and EBI-959602.

- **Submission**
 - A submission should contain the essential administrative information:
 - Contact email
 - Publication title
 - First author
 - Publication identifier (if manuscript is not yet submitted, authors, or the recipient, may substitute an internal tracker here)
- **Experiment**
 - Each experimental setup should be described separately, with the following parameters:
 - **Host system**
 - The host organism in which the interaction took place, identified by the NCBI taxonomy identifier
 - Example: Yeast (TaxID:4932) Further specification of cell line or tissue is recommended
 - **Interaction detection method**
 - The method by which the interaction was detected
 - Root term MI:0001
 - Example: two hybrid (MI:0018)
 - **Participant identification method**
 - The method by which the interaction participants were determined
 - Root term MI:0002
 - Example: nucleotide sequence (MI:0078)
- **Interaction**
 - **Participant list**
 - The list of all molecules participating in the interaction. The list can contain any number of elements. Each molecule should be characterized by:
 - **Database**
 - Root term: MI:0444
 - Example: UniProt (MI:0486)
 - **Accession number from that database**
 - Example: P48551
 - **Version number (optional)**
 - **Name**
 - The common name of the molecule used in the manuscript
 - Example: IFN- α R β L
 - **The species of origin for the molecule identified by NCBI taxonomy identifier**
 - Example: 9606
 - **Biological role**
 - The biological role of the molecule in the interaction
 - Root term: MI:0500
 - Example: neutral component (MI:497)
 - **Experimental role**
 - The experimental role of the molecule in the interaction
 - Root term: MI:0495
 - Example: bait (MI:0496)
 - **Confidence (optional)**
 - A confidence value attributed to the interaction
 - If a confidence value has been assigned, the confidence attribution system must be described in the manuscript. Ideally, the raw data for the confidence assignment should be available.

Controlled vocabularies are an essential part of the characterization of a molecular interaction in PSI-MI format. Elements of these controlled vocabularies are referred to as MI:xxxx above. The complete controlled vocabularies can be accessed at <http://www.psidesv.info/index.php?q=node/31> or interactively at <http://www.ebi.ac.uk/ontology-lookup/browse.do?ontName=MI> (ref. 18).

coverage as possible of interaction data in the literature is greatly hindered by inconsistencies and missing information in published papers. The absence of key pieces of information can lead both to misinterpretation of the paper by scientists and to a time-consuming, error-prone attempt to derive the missing information by a database curation team. Often, the reason for such information deficits is simply the lack of a community consensus on what information is required to appropriately describe a molecular interaction.

To address this issue, we have developed MIMIx as a basis for discussion. MIMIx represents a compromise between the depth of information necessary to describe all relevant aspects of an interaction experiment and the reporting burden placed on scientists who generate data. Its purpose is to ensure that the bench scientist has a checklist (**Box 1**) of the information to be supplied when describing experimental molecular interaction data in a journal article, displaying data on a website or depositing data directly into a public database (**Box 2**).

A MIMIx-compliant dataset is not intended to allow an interaction experiment to be reproduced from a database record but to enable database users to quickly assess and focus on data relevant to them and then link to the source publications for the full experimental context. On the other end of the complexity scale, the PSI-MI XML interchange format, which is adopted by all IMEx partners, provides for a much richer representation of a molecular interaction experiment than that required by MIMIx. IMEx partners also welcome data submissions that use the full complexity of the PSI-MI format.

Molecules

The single greatest source of data loss in transferring interaction data into a database is the use of ambiguous molecule identifiers, such as gene names. According to anecdotal estimates from database curators, as much as 70% of overall curation time is spent mapping molecule identifiers unambiguously to well characterized database entries. For example, a paper may not indicate both the gene name and the species from which the gene originated. This information is implicit in the molecule identifiers generated by the major databases. The description 'lck cloned in a mammalian expression vector' gives no indication as to whether the protein source is human, mouse, bovine or rat. 'Human p56^{lck} protein' gives information about the species but not about the splice isoform, whereas both species and sequence are provided by the accession numbers UniProtKB¹¹ P06239

and RefSeq¹² NP_005347, and P06239-1 gives a full description of a specific isoform. ‘Human PI3-kinase p85 subunit’ may appear to be a unique reference, but does it refer to the alpha subunit (P27986) or the beta subunit (O00459), which are two distinct gene products? Such errors will almost certainly result in the paper in question not being added to a curated dataset and may also mislead the reader regarding the actual construction of the experiment. Similarly, it is important for authors to state whether an interaction described in one organism was modeled from an interaction detected between similar molecules in a related organism; for example, an interaction between a rat and a human protein being used to infer a human-human protein interaction. The constructs used, including the organism of origin of the sequence and the splice variant, should be clearly described.

We therefore request that all molecules be identified by a database accession number from a public database. For proteins, UniProt or RefSeq are strongly recommended; for genes, Ensembl¹³ or Entrez Gene¹⁴; for chemical entities, PubChem¹⁴ or ChEBI¹⁵. Nucleotide sequence database accession numbers (DDBJ, EMBL or GenBank, <http://www.insdc.org>) identify specific transcripts and give additional information as to the source and the class of nucleic acid under investigation. Where a molecule description is not available from these databases, identifiers from other public databases, such as model-organism databases, may be used. For a full list of recommended databases, please refer to the relevant section of the PSI-MI controlled vocabulary (see below), which also provides unified names for these resources.

An annotated protein or nucleic acid sequence may vary with time as the original submitters update their coding sequence prediction programs, frameshifts are identified, and correction or resequencing is undertaken. This may invalidate the mapping of specific sequence positions; for example, those where deletion mutants or binding domains

are described. We therefore request the addition of version numbers, either of the molecule (for example, P06239.5) or of the database, to the MIMIx record.

Although the identification of molecules by accession number is precise, it may be unwieldy to refer to ‘UniProt:P06239.5’ instead of ‘lck’ in the text of a paper. To satisfy the need for both precision and readability, we recommend that the accession number and the molecule name used in the text be associated either in the submitted database record or at least at the first occurrence in the paper (for example, “...lck (UniProt: P06239.5)...”).

A key element in the description of an interaction experiment is the role a molecule has in the interaction. MIMIx requests the classification of the molecule role in two ways: the biological role, for example, enzyme or enzyme target; and the experimental role, for example, bait or prey. For both of these, the PSI-MI standard defines a comprehensive controlled vocabulary, ensuring that the same term, rather than synonyms or alternative spellings, is used throughout a paper and that the interpretation of the meaning of that term remains fixed. A list of controlled vocabulary terms that describe the various methods used to detect molecular interactions, current as *Nature Biotechnology* went to press, is available in **Supplementary Note 1** online. This list undergoes continual revision as technologies evolve; the latest version is available at <http://www.ebi.ac.uk/ontology-lookup/browse.do?ontName=MI>.

Finally, it should be noted that databases describe the canonical form of a molecule. The actual participant in a molecular interaction may have been altered, either naturally by the cell (e.g., by cleavage of a bioactive peptide from a precursor protein) or by engineering (e.g., by addition of a tag or creation of a deletion mutant). Terms to describe the ‘participant’ in an interaction as a derivative of the ‘molecule’ in the database entry are available in the PSI-MI controlled vocabularies.

Box 2 Frequently asked questions about the MIMIx guidelines

To clarify when researchers should use MIMIx and some frequently asked questions about the guidelines, we provide select questions and answers below:

When should I use MIMIx?

- When describing any molecular interaction within a paper that you are writing
- When preparing your data for submission to any MIMIx-compliant database

How should I describe my experiment in a way that is clear to the reader?

- Use the appropriate terms from the PSI-MI controlled vocabulary; if you cannot find your technique listed, request a new term
- Add any additional detail to Materials and Methods as usual, or in a free-text description in a database submission

Why do I need to use accession numbers; surely a gene name is enough?

- Gene names often change with time, and their use may lead to ambiguities in the data
- Gene names give no indication as to originating species, and this important information is often missing from papers

What is meant by ‘participant role’?

- In many experiments, a specific protein will be modified or specifically targeted by antibodies and then used to capture its interacting partners; this is a bait-prey relationship
- We can also infer that proteins are physically interacting if they have a biological relationship, such as enzyme–enzyme target
- In many cases, for example, cosedimentation, no such relationships can be assigned, and these interactions are described as ‘neutral’ in MIMIx terminology

Why should I submit my results to a MIMIx-compatible database; surely writing a paper is enough?

- Data deposition ensures that your data are available, in a downloadable format, to the entire interactome community
- It increases the visibility and readership (and thus potentially citations) of your paper
- It results in your data being available through many other routes, such as the UniProtKB database
- The database can help you provide data to the journal, give you accession numbers and ensure long-term storage of your data

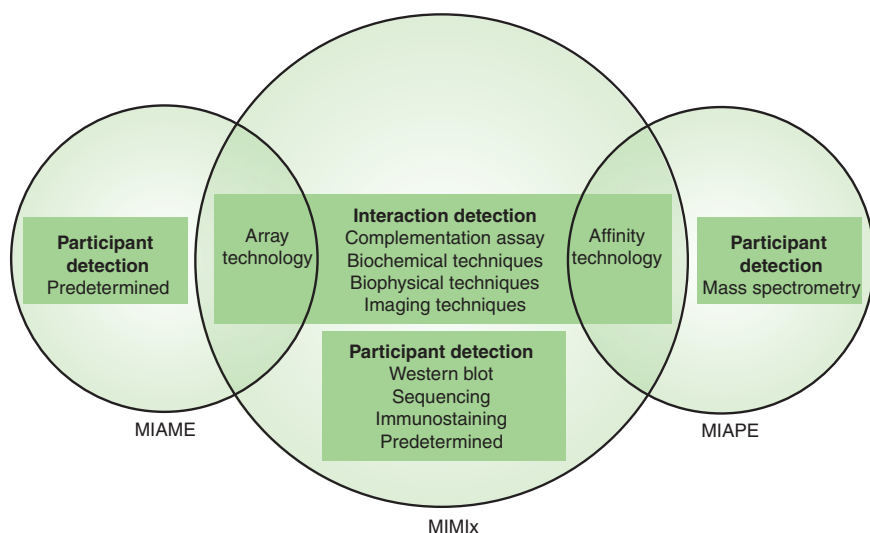


Figure 1 The relationship of MIMix to two guidelines that may be relevant to molecular interaction studies, MIAPE¹⁷ and MIAPE¹⁶/MIAPE-MS. Almost all interaction data may be described using MIMix; however, MIAPE provides guidelines for describing a microarray experiment, and MIAPE allows the submitter to supply details of the peptides and underlying spectra when mass spectrometry has been used to identify protein participants.

Experiment

The MIMix experiment description implements the core requirements of PSI's "minimum information about a proteomics experiment" (MIAPE) guidelines¹⁶ (see p. 887) and aims to capture the aspects of an interaction experiment that are necessary to classify and critically assess the results and the interpretation of the results. It is likely to be further refined in the future as other technology-specific MIAPE modules evolve. The attributes we consider essential at present are as follows:

The 'host organism' describes the system in which the interactions were detected. The host organism should be described by a National Center for Biotechnology Information (NCBI; Bethesda, Maryland, USA) taxonomy identifier and should contain further specification, such as cell-line or tissue descriptors. When the experiment was performed *in vitro*, this should be described as free text.

The 'interaction detection method' describes the method by which the interaction was determined (for example, tandem affinity purification (MI:0676)).

The 'participant detection method' names the experimental procedure for the detection of the molecules participating in the interaction (for example, peptide mass fingerprinting (MI:0082)).

Beyond these essential requirements, we recommend that authors provide additional detail on molecule sources, sample preparation and further relevant experimental parameters using the detailed controlled vocabularies provided by the PSI-MI standard.

Interaction

The PSI-MI standard provides a formal frame for a detailed description of an interaction, including both qualitative parameters, such as details of mutations, and quantitative parameters, such as dissociation constants. However, these data are often not available, and, thus, MIMix requires only one element for the description of an interaction: the list of molecules participating in it, characterized as above. If a quality assessment was carried out, the confidence value assigned to the interaction and the confidence attribution system must also be included in the manuscript. Particularly in large-scale experiments, interactions are

usually assigned a quality score, which might be derived from data collected in the experiment itself or from additional data outside the experiment. Inclusion of interaction data in public databases requires that this reliability score be easily accessible. Ideally, not only the score but also the raw data used to derive the score should be reported so that users can perform alternative quality assessments.

Relationships to other biological standards

The MIMix guidelines have been developed in close collaboration with related standards bodies within both the HUPO-PSI and the wider community, in consultation with contributors to the MIAPE microarray standards¹⁷. MIMix is one of a series of modules developed within the framework of the MIAPE guidelines¹⁶. When an interaction experiment encompasses experimental data that are more fully described by other modules, authors should refer to the relevant guidelines when preparing their data for submission to a journal or a database (Fig. 1). For example, identification of prey proteins in a tandem affinity purification (TAP) pulldown by mass spectrometry should

be described according to the MIAPE-MS guidelines (C.F. Taylor *et al.*, unpublished).

Similarly, the HUPO-PSI and the Microarray Gene Expression Data (MGED) consortium are jointly working to provide guidelines for the annotation of array data that will ensure a smooth path for the annotation and submission of such data. Overarching all of these standards is the "functional genomics experiment" (FuGE) model¹⁸, which can be used to provide protocols and data flow models should the user wish to annotate such detail. All these guidelines are being managed through a central repository of standards, as described in the "minimum information for biological and biomedical investigations" (MIBBI)¹⁹, to ensure that they are complementary and nonoverlapping.

Data deposition

Curators of the main molecular interaction databases work to collect and archive data from journal publications. Although a systematic reporting of published interaction data according to the above guidelines would greatly increase the efficiency of the curation task, literature curation after publication is only a second-best option. We therefore recommend that all reported interaction data be deposited in a publicly available molecular interaction database before publication.

Data deposition has benefits for all parties involved. The databases will be able to work more efficiently and will have more direct access to the data producer to resolve unclear issues. The scientific community will benefit from more, and more precise, information in the databases, as database records can be checked directly by the data producer. Journals and data producers will benefit from consistently formatted database records, which can be included in the supplementary material of a publication. Accession numbers issued by a database and included in the journal publication will allow direct access to the data in the database and a quick connection to related data in the database, such as other records on the same molecules. Finally, data producers and journals will gain exposure for the publication through cross-references from the database records.

IMEx databases offer several options for data deposition (<http://imex.sf.net/deposition.html>). The submission of fully formatted PSI-MI

XML files is recommended for large-scale data producers, who usually have the data available in in-house databases anyway. For smaller-scale experiments, a preformatted Microsoft Excel spreadsheet file is available, with instructions on how to complete it. In addition to technical systems, such as the Ontology Look-up Service (OLS) browser²⁰ and a system for the automatic validation of PSI-MI XML files (<http://www.ebi.ac.uk/intact/validator>), database curation teams provide assistance in all stages of the data deposition process, for example, in the correct use of the detailed controlled vocabularies used to characterize an interaction. We particularly encourage early contacts with database curation teams, to embed appropriate data collection protocols into the experiment-planning stage.

In addition to the biological data, each data deposition must be accompanied by the minimal administrative data, namely contact email, publication title, first author and the publication identifier, usually a PubMed or Digital Object (<http://www.doi.org>) identifier. In the prepublication stage, a journal-specific identifier can be used to provide a unique identification of the manuscript accompanying the data deposition; before manuscript submission, the authors may use their own in-house identifier.

To optimize the use of public resources, IMEx partners have developed common curation guidelines and have agreed to synchronize their curation work and exchange all user-submitted data so as to build up a network of stable, well coordinated molecular interaction databases freely accessible to the community. Although accession numbers for deposited interactions will be issued within five working days of the provision of all necessary data, deposited data will be released only upon publication of the associated manuscript or at the request of the data provider.

Conclusion

The MIMIx guidelines presented here will not be static. They will evolve based on community requirements in the context of a rapidly developing science. This document has been assembled by a large number of experts and subjected to public review both on the PSI website and through *Nature Biotechnology* community review. At all stages, we have discussed input and fed it back into the document. The MIMIx guidelines, PSI-MI XML interchange format and the corresponding controlled vocabularies are all maintained and updated through the PSI-MI workgroup using mailing lists, issue trackers and annual workshops. If you wish to make specific comments on the MIMIx guidelines, please use the issue tracker at <http://www.psidev.info/index.php?q=node/279> or, for a wider involvement, refer to the mailing lists at <http://www.psidev.info/>.

Note: Supplementary information is available on the *Nature Biotechnology* website.

COMPETING INTERESTS STATEMENT

The authors declare no competing financial interests.

Published online at <http://www.nature.com/naturebiotechnology>

Reprints and permissions information is available online at <http://npg.nature.com/reprintsandpermissions/>

1. Bader, G.D., Betel, D. & Hogue, C.W. BIND: the Biomolecular Interaction Network Database. *Nucleic Acids Res.* **31**, 248–250 (2003).
2. Salwinski, L. *et al.* The Database of Interacting Proteins: 2004 update. *Nucleic Acids Res.* **32** (Database issue), D449–D451 (2004).
3. Hermjakob, H. *et al.* The HUPPO-PSI's molecular interaction format—a community standard for the representation of protein interaction data. *Nat. Biotechnol.* **22**, 177–183 (2004).
4. Orchard, S. *et al.* Autumn 2005 Workshop of the Human Proteome Organisation Proteomics Standards Initiative (HUPPO-PSI) Geneva, 4–6 September 2005. *Proteomics* **6**, 738–741 (2006).
5. Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003).
6. Iragne, F., Nikolski, M., Mathieu, B., Auber, D. & Sherman, D. ProViz: protein interaction visualization and exploration. *Bioinformatics* **21**, 272–274 (2005).
7. Meil, A., Durand, P. & Wojcik, J. PIMWalker: visualising protein interaction networks using the HUPPO PSI molecular interaction format. *Appl. Bioinformatics* **4**, 137–139 (2005).
8. Kerrien, S. *et al.* IntAct—open source resource for molecular interaction data. *Nucleic Acids Res.* **35** (Database issue), D561–D565 (2007).
9. Chatr-aryamontri, A. *et al.* MINT: the Molecular Interaction database. *Nucleic Acids Res.* **35** (Database issue), D572–D574 (2007).
10. Pagel, P. *et al.* The MIPS mammalian protein-protein interaction database. *Bioinformatics* **21**, 832–834 (2005).
11. The UniProt Consortium. The Universal Protein Resource (UniProt). *Nucleic Acids Res.* **35** (Database issue), D193–D197 (2007).
12. Pruitt, K.D., Tatusova, T. & Maglott, D.R. NCBI Reference Sequence project: update and current status. *Nucleic Acids Res.* **31**, 34–37 (2003).
13. Hubbard, T.J. *et al.* Ensembl 2007. *Nucleic Acids Res.* **35** (Database issue), D610–D617 (2007).
14. Wheeler, L. *et al.* Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* **34** (Database issue), D173–D180 (2006).
15. de Matos, P. *et al.* ChEBI – Chemical Entities of Biological Interest, Database Summary Paper 646, *Nucleic Acids Res.* <<http://www3.oup.co.uk/nar/database/summary/646>> (2006).
16. Taylor, C.F. *et al.* The minimum information about a proteomics experiment (MIAPE). *Nat. Biotechnol.* **25**, 887–893 (2007).
17. Brazma, A. *et al.* Minimum information about a microarray experiment (MIAME)—toward standards for microarray data. *Nat. Genet.* **29**, 365–371 (2001).
18. Jones, A.R., Pizarro, A., Spellman, P., Miller, M. & FuGE Working Group. FuGE: Functional Genomics Experiment object model. *OMICS* **10**, 179–184 (2006).
19. Taylor, C.F. *et al.* Promoting coherent minimum reporting requirements for biological and biomedical investigations: the MIBBI project. *Nat. Biotechnol.* (in the press).
20. Cote, R.G., Jones, P., Apweiler, R. & Hermjakob, H. The Ontology Lookup Service, a lightweight cross-platform tool for controlled vocabulary queries. *BMC Bioinformatics* [online] **7**, 97 (2006).
21. Croze, E. *et al.* Receptor for activated C-kinase (RACK-1), a WD motif-containing protein, specifically associates with the human type I IFN receptor. *J. Immunol.* **165**, 5127–5132 (2000).