



**Embargoed until Sunday, 26th August 2007, 18:00 BST**

## Minimum Information Standards - all for one and one for all

**Hinxton, 26th August 2007** - Three papers published by EMBL scientists and their collaborators will make it much easier to share and compare information from large-scale proteomics data. The papers are published in *Nature Biotechnology* on 8th and 26th August.

As the quantity of available biological information and the use of public data repositories increases, consistency in the information held in these databases is vital to allow full integration, exchange and comparison of their contents. As Europe's main provider of biological data, the EMBL-European Bioinformatics Institute (EBI) is involved in setting the precedent for reporting standards by applying these to its own data repositories such as ArrayExpress (microarray and gene expression data), IntAct (molecular interaction data) and PRIDE (protein identification linked to experimental evidence and publications).

The Minimum Information About a Proteomics Experiment (MIAPE) and the Minimum Information required for reporting a Molecular Interaction Experiment (MIMIX) guidelines propose the range of information to be recorded to document proteomics and molecular interaction data, respectively. The standards aim to reduce ambiguity and capture all the necessary information from an experiment to set the experimental results in both a biological and a methodological context, thereby providing a deeper level of understanding to others exploring the data. Henning Hermjakob from the EMBL-EBI, a co-author on both Perspectives papers published on 8th August, said "Through the community-wide uptake of agreed minimum reporting standards, we can all benefit from easier identification and use of information that is most relevant to our own areas of work. This is the next step in providing freely accessible data repositories of the highest possible quality."

The *Nature Biotechnology* Perspectives papers, published as open-source articles, outline the proposed reporting requirements for proteomics and molecular interaction experiments and discuss their implementation, impact and benefits. The later research paper, published in the same journal on 26th August, shows how implementation of these standards benefits not only the reporting researchers, but also the wider community through the development of more detailed and comprehensive information resources.

### Source Article

Taylor, C.F., et al. The Minimum Information About a Proteomics Experiment (MIAPE). *Nature Biotechnology*, 25, 887-893 (2007)

Orchard, S., et al. The Minimum Information required for reporting a Molecular Interaction Experiment (MIMIX). *Nature Biotechnology*, 25, 894-898 (2007)

Bantscheff, M., et al. A quantitative chemical proteomics approach reveals novel modes of action of clinical ABL kinase inhibitors. *Nature Biotechnology*, 26 August 2007

### Setting the standards

Both sets of reporting requirements, MIAPE and MIMIX, have been shaped by input from the scientific community and developed to minimise the burden on individual researchers. The MIAPE reporting guidelines are being developed as a range of individual modules, which can be combined as necessary to cover an entire experimental workflow – from study design to statistical data analysis. As the first finalised module, the MIMIX guidelines implement the general MIAPE principles for describing molecular interactions and also present recommendations on data deposition prior to publication.

### Standards in action

Both sets of reporting requirements were put into practice by Bantscheff et al. in the reporting of a large-scale approach to profile the interaction of protein kinases with small inhibitory molecules. The Cellzome researchers used this method to validate the action of three drugs sharing a particular kinase target and also to identify novel drug targets. The binding information shed light on the drug binding specificities and downstream effects on signalling pathways. The MIAPE and MIMIX-compliant mass spectrometry and interaction data were entered into the EBI-hosted PRIDE, IntAct, and ChEBI databases, and accession numbers are included in the research publication to provide direct access to the proteomic information, mass spectra and molecular interaction data. The interconnected EBI data resources serve not only to hold the direct experimental results, but also set them within a wider biological context, for example, by linking the identified kinases to known functions in the UniProt database.

The quantitative profiling method used has potential application in drug discovery and in gaining a greater understanding of drug action in patients. The systematic recording of such information in public repositories and adherence to reporting standards ensures that maximum use can be made of this progression in knowledge, offering benefits to the scientific community, and in the case of drug discovery, development and healthcare, there are clearly benefits to be had for society too. ●

### Contact:

Louisa Wright PhD, EMBL-EBI Scientific Outreach Officer, Hinxton, UK, Tel: +44 1223 494665, [www.ebi.ac.uk](http://www.ebi.ac.uk), [louisa@ebi.ac.uk](mailto:louisa@ebi.ac.uk)

Anna-Lynn Wegener, EMBL Press Officer, Heidelberg, Germany, Tel: +49 6221 387 452, [www.embl.org](http://www.embl.org), [wegener@embl.de](mailto:wegener@embl.de)

**About EMBL:**

The European Molecular Biology Laboratory is a basic research institute funded by public research monies from 19 member states (Austria, Belgium, Croatia, Denmark, Finland, France, Germany, Greece, Iceland, Ireland, Israel, Italy, the Netherlands, Norway, Portugal, Spain, Sweden, Switzerland and the United Kingdom). Research at EMBL is conducted by approximately 80 independent groups covering the spectrum of molecular biology. The Laboratory has five units: the main Laboratory in Heidelberg, and Outstations in Hinxton (the European Bioinformatics Institute), Grenoble, Hamburg, and Monterotondo near Rome. The cornerstones of EMBL's mission are: to perform basic research in molecular biology; to train scientists, students and visitors at all levels; to offer vital services to scientists in the member states; to develop new instruments and methods in the life sciences and to actively engage in technology transfer activities. EMBL's International PhD Programme has a student body of about 170. The Laboratory also sponsors an active Science and Society programme. Visitors from the press and public are welcome.

**About EBI:**

The European Bioinformatics Institute (EBI) is part of the European Molecular Biology Laboratory (EMBL) and is located on the Wellcome Trust Genome Campus in Hinxton near Cambridge (UK). The EBI grew out of EMBL's pioneering work in providing public biological databases to the research community. It hosts some of the world's most important collections of biological data, including DNA sequences (EMBL-Bank), protein sequences (UniProt), animal genomes (Ensembl), three-dimensional structures (the Macromolecular Structure Database), data from microarray experiments (ArrayExpress), protein-protein interactions (IntAct) and pathway information (Reactome). The EBI hosts several research groups and its scientists continually develop new tools for the bio-computing community.

**About IntAct:**

IntAct provides a freely available, open source database system and analysis tools for protein interaction data. All interactions are derived from literature curation or direct user submissions and are freely available. Currently, IntAct stores more than 150,000 binary molecular interactions, accessible at <http://www.ebi.ac.uk/intact>

**About PRIDE:**

The PRIDE PRoteomics IDentifications database is a centralised, standards compliant, public data repository for proteomics data. It has been developed to provide the proteomics community with a public repository for protein and peptide identifications together with the evidence supporting these identifications. PRIDE is accessible at <http://www.ebi.ac.uk/pride>

**About ChEBI:**

The Chemical Entities of Biological Interest (ChEBI) database is a freely available, high quality dictionary of chemical entities. ChEBI encompasses an ontological classification, whereby the relationships between molecular entities or classes of entities and their parents and/or children are specified. ChEBI uses nomenclature, symbolism and terminology endorsed by the International Union of Pure and Applied Chemistry (IUPAC) and the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology (NC-IUBMB). ChEBI is accessible at <http://www.ebi.ac.uk/chebi>

**About UniProt:**

UniProt (Universal Protein Resource) is the world's most comprehensive catalogue of information on proteins. It is a central repository of protein sequence and function created by joining the information contained in UniProtKB/Swiss-Prot, UniProtKB/TrEMBL, and PIR. The UniProt Consortium comprises the EBI, the Swiss Institute of Bioinformatics (SIB) and the Protein Information Resource (PIR). EBI, located at the Wellcome Trust Genome Campus in Hinxton, UK, hosts a large resource of bioinformatics databases and services. SIB, located in Geneva, Switzerland, is the founding centre of Swiss-Prot and maintains the ExPASy (Expert Protein Analysis System) servers that are a central resource for proteomics tools and databases. PIR, hosted by the Georgetown University Medical Center (GUMC) in Washington, DC, USA, is heir to the oldest protein sequence database, Margaret Dayhoff's Atlas of Protein Sequence and Structure published by the National Biomedical Research Foundation (NBRF) from 1965-1978. The primary mission of the consortium is to support biological research by maintaining a high quality database that serves as a stable, comprehensive, fully classified, richly and accurately annotated protein sequence knowledgebase, with extensive cross-references and querying interfaces freely accessible to the scientific community. UniProt will build upon the solid foundations laid by the consortium members over many years. UniProt is accessible from <http://www.ebi.ac.uk/uniprot>

**About Cellzome Inc.:**

Cellzome is a drug discovery company with world-class, proprietary proteomics technology for target and lead identification, and a drug discovery programme in Alzheimer's disease. The technology provides molecular understanding of disease and helps to identify the role of active compounds and their possible side effects, and thus provides insight into how potential drugs might behave in man.

Cellzome's strategy is to commercialise its technology through building its own pipeline of clinical products and through collaboration with leading pharmaceutical partners. To date it has signed such collaborations with Bayer HealthCare, Johnson & Johnson Pharmaceutical Research & Development (J&JPRD), a large pan-therapeutic collaboration with Novartis and, a drug discovery collaboration in Alzheimer's disease with Ortho-McNeil Pharmaceutical and J&JPRD. Cellzome's holding company is domiciled in the US and employs about 75 people across its two operating subsidiaries in Cambridge, UK and Heidelberg, Germany. To learn more about Cellzome, please visit the website: [www.cellzome.com](http://www.cellzome.com)

**Policy regarding use**

EMBL press releases may be freely reprinted and distributed via print and electronic media. Text, photographs & graphics are copyrighted by EMBL. They may be freely reprinted and distributed in conjunction with this news story, provided that proper attribution to authors, photographers and designers is made. High-resolution copies of the images can be downloaded from the EMBL web site: [www.embl.org](http://www.embl.org)