



Henning Hermjakob

Dipl. Inf (MSc) in Bioinformatics, University of Bielefeld, 1996. Research Assistant at the National Centre for Biotechnology (GBF), Braunschweig, Transfac Database team. At EMBL-EBI since 1997.

Proteomics Services

DESCRIPTION OF SERVICES

The Proteomics Services team develops tools and resources for the representation, deposition, distribution and analysis of proteomics and related data. We follow an open-source, open-data approach: all resources we develop are freely available. The team is a major contributor to the Proteomics Standards Initiative (PSI) of the international Human Proteome Organization (HUPO). We provide reference implementations for the PSI community standards, in particular the PRIDE proteomics identifications database and the IntAct molecular interaction database. We provide the Reactome pathway database in collaboration with New York University and the Ontario Institute for Cancer Research. In the context of the EU Virtual Physiological Human project, we contribute to the development of an interoperability framework that bridges physiology and molecular biology.

As a result of long-term engagement with the proteomics community, journal editors and funding organisations, proteomics data deposition in PSI-compliant data resources such as IntAct and PRIDE is becoming a strongly recommended part of the publishing process. This has resulted in a rapid increase in the data content of our resources. In addition, the Proteomics curation teams ensure consistency and appropriate annotation of all data, whether from direct depositions or literature curation, to provide the community with high-quality reference datasets.

We also contribute to the development of data integration technologies using the Distributed Annotation System (DAS) and web services across a range of European projects, including Apo-Sys, LipidomicNet, SLING, ENFIN and ProteomeBinders.

SUMMARY OF PROGRESS

- In the context of the HUPO Proteomics Standards Initiative (PSI), contributed to a series of community standard documents on gel electrophoresis (Hoogland et al, 2010, Gibson et al, 2010), protein-affinity reagents (Bourbeillon et al, 2010, Gloriam et al, 2010) and mass spectrometry (Martens et al, 2010);
- Co-authored a series of publications with third parties, focussing on use of PSD data in external data analysis and analysis tools (Perreau et al., 2010, Antonov et al., 2010, Lee et al., 2010);
- Oversaw the growth of the Proteomics Identifications Database (PRIDE) to more than 100 million mass spectra;
- Started the FP7 project RICORDO (coordinated by PST), which aims to integrate models and ontologies related to medical physiology and human anatomy – also contributed substantially to the EU-funded Virtual Physiological Human (VPH) project.

MAJOR ACHIEVEMENTS

The PSI Molecular Interactions workgroup collaborates with several key molecular interaction data providers to synchronise their curation efforts and provide non-redundant datasets curated to common standards. InAct is an active member of the International Molecular Exchange consortium (IMEX), which started full production mode and released a common website in February 2010. Other members of the consortium include DIP (University of California Los Angeles, USA), MINT (University of Rome, Italy), MatrixDB (University of Lyon, France), Molecular Connections Inc. and MPIDB (J. Craig Venter Institute, USA).

Based on the PSI molecular interaction standards, we developed the PSI Common Query Interface (PSICQUIC), a common query API for molecular interaction databases. PSICQUIC was released in 2010, providing access to more than 15 million binary interaction evidences from 16 different sources, including protein–protein interactions, protein–small molecule interactions, and simplified pathway data (Aranda et al., submitted).

IMEX and PSICQUIC are supported by an EU grant, PSIMEX, which the Proteomics Services Team coordinates. With the PSI molecular interaction formats well established and widely used in the community, the IMEX consortium in production mode and the widespread adoption of the PSICQUIC interface, we have achieved the core target of the PSIMEX grant: the global coordination and integration of major molecular interaction data resources.

In collaboration with major proteomics data providers (e.g. PeptideAtlas, Peptidome, UniProt, University of Ghent, University of Liverpool, ETH Zurich, University of Michigan, Wiley-VCH) we developed a concept for regular proteomics data exchange between key repositories. The resulting ProteomeXchange EU grant is under negotiation (June 2010).

Following a comprehensive review in 2009, we completely redeveloped the Reactome web site. The new Reactome (in beta testing as of June 2010) provides interactive pathway diagrams for all Reactome pathways, improved pathway-analysis tools and close integration with molecular interaction data via the PSICQUIC interface.

FUTURE PLANS

After rapid development and achievement of major milestones in the molecular interaction domain, we now need to consolidate the achievements, selectively open the IMEX collaboration to new partners and develop advanced tools to take advantage of detailed IMEX curation and the integrative PSICQUIC interface. A major challenge is the complete redevelopment of the PRIDE database, necessary to cope with the rapid increase in data content but also to turn PRIDE from a publication-centric repository to a key source for protein expression information. Beyond the technical challenges of data quantity, the two major conceptual challenges are to capture the very diverse quantitative proteomics data and to develop quality criteria to enable the selective export of high confidence PRIDE data to other resources like UniProt, Reactome or integrative data analysis tools. We plan to intensify data integration within and beyond the projects of the Proteomics Services team, in particular using web services and the DAS. We will also continue to integrate Reactome pathways and IntAct molecular interactions, as well as integrating PRIDE and IntAct, to enable efficient data deposition and navigation between molecular interactions and underlying mass spectrometry data. We will continue our successful collaboration with all PSI partners, in particular with journals and editors, to encourage data producers to make their data available to the community through public databases by utilising community-supported standards.

SELECTED REFERENCES

- Martens, L., et al. (2010) mzML - a community standard for mass spectrometry data. *Mol. Cell. Proteomics* (in press). DOI: 10.1074/mcp.R110.000133.
- Perreau, V.M., et al. (2010) A domain level interaction network of amyloid precursor protein and Abeta of Alzheimer's disease. *Proteomics* 10, 2377-2395.
- Lee, K., et al. Mapping plant interactomes using literature curated and predicted protein-protein interaction data sets. *Plant Cell*. 22, 997-1005.
- Gloriam, D.E., et al. (2010) A community standard format for the representation of protein affinity reagents. *Mol Cell Proteomics* 9, 1-10.

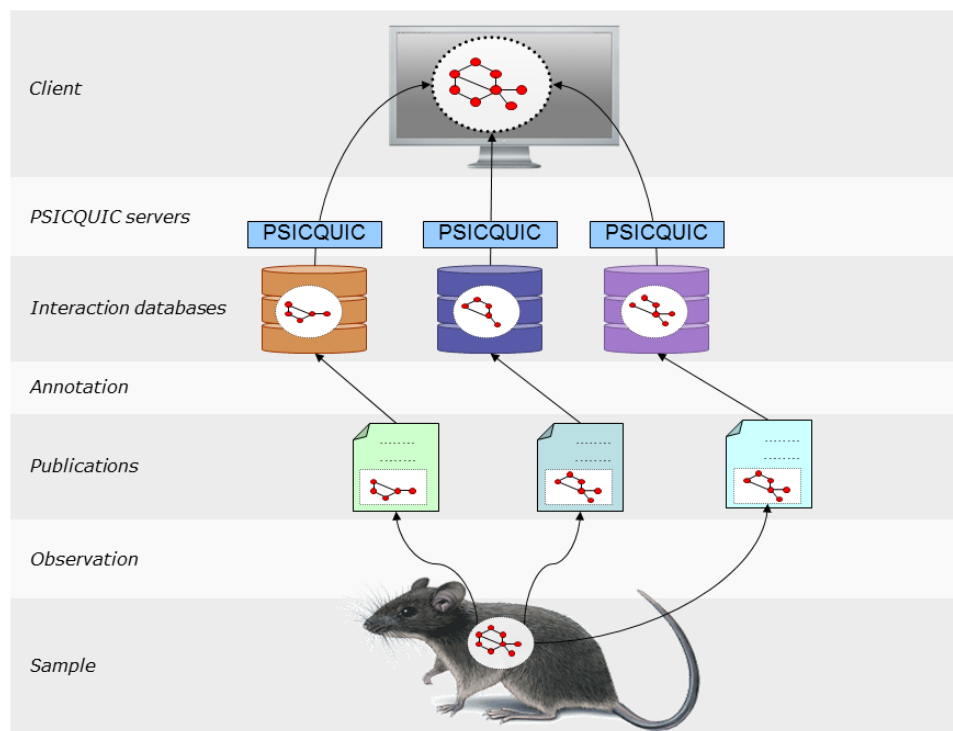


Figure. PSICQUIC and IMEX support a truly distributed provision of molecular interaction data: an experimental system is observed in multiple independent studies, resulting in multiple publications. Based on the collaboration in the International Molecular Exchange Consortium (IMEX), independent interaction databases curated these publications in work-sharing mode. Data is then released in PSI format through the PSICQUIC interface. A web client queries all PSICQUIC services and integrates the data on the client side.