Published online 25 April 2008 Nucleic Acids Research, 2008, Vol. 36, Web Server issue W485–W490
doi:10.1093/nar/gkn196

ISPIDER Central: an integrated database web-server for proteomics

Jennifer A. Siepen[1], Khalid Belhajjame[2], Julian N. Selley[1], Suzanne M. Embury[2], Norman W. Paton[2], Carole A. Goble[2], Stephen G. Oliver[3], Robert Stevens[2], Lucas Zamboulis[4,5], Nigel Martin[4], Alexandra Poulovassillis[4], Philip Jones[6], Richard Côté[6], Henning Hermjakob[6], Melissa M. Pentony[7], David T. Jones[7], Christine A. Orengo[5] and Simon J. Hubbard[1,]*

[1]Faculty of Life Sciences, University of Manchester, M13 9PT, [2]School of Computer Science, Faculty of Engineering and Physical Sciences, University of Manchester, [3]Department of Biochemistry, University of Cambridge, Sanger Building, 80 Tennis Court Road, Cambridge CB2 1GA, [4]School of Computer Science and Information Systems, Birkbeck College, University of London, [5]Department of Biochemistry and Molecular Biology, University College London, Gower street, London, [6]EMBL Outstation European Bioinfomatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge and [7]Department of Computer Science, University College London, Gower street, London, UK

Received January 31, 2008; Revised March 14, 2008; Accepted April 3, 2008

ABSTRACT

Despite the growing volumes of proteomic data, integration of the underlying results remains problematic owing to differences in formats, data captured, protein accessions and services available from the individual repositories. To address this, we present the ISPIDER Central Proteomic Database search (http://www.ispider.manchester.ac.uk/cgi-bin/ProteomicSearch.pl), an integration service offering novel search capabilities over leading, mature, proteomic repositories including PRoteomics IDEntifications database (PRIDE), PepSeeker, PeptideAtlas and the Global Proteome Machine. It enables users to search for proteins and peptides that have been characterised in mass spectrometry-based proteomics experiments from different groups, stored in different databases, and view the collated results with specialist viewers/clients. In order to overcome limitations imposed by the great variability in protein accessions used by individual laboratories, the European Bioinformatics Institute's Protein Identifier Cross-Reference (PICR) service is used to resolve accessions from different sequence repositories. Custom-built clients allow users to view peptide/protein identifications in different contexts from multiple experiments and repositories, as well as integration with the Dasty2 client supporting any annotations available from Distributed Annotation System servers. Further information on the protein hits may also be added via external web services able to take a protein as input. This web server offers the first truly integrated access to proteomics repositories and provides a unique service to biologists interested in mass spectrometry-based proteomics.

INTRODUCTION

Proteomics is beginning to catch up with nucleotide-based microarray technology as a means to catalogue and quantify the gene products expressed in cells and tissues, with the added advantage that the proteome provides the functional molecular components of most biological systems—proteins are the principal effector molecules. Indeed, proteomics forms a key component in a range of experiments, from protein quantification in systems approaches to model and predict pathways (1) through to characterizing biomarkers for diagnostic purposes (2). Proteomics provides the opportunity to study the functional molecules of the cell directly, identifying and quantifying proteins expressed from the genome (3).

The ultimate goal in most proteomics experiments is therefore to confidently identify all the proteins contained within a sample. This is in turn usually based on observed peptide identifications derived from the collected tandem mass spectra and attendant database searches. Confidence in a particular identification may be derived simply from the score assigned by the identification tool, by some

*To whom correspondence should be addressed. Tel: +44 161 3068930; Fax: +44 161 2755082; Email: simon.hubbard@manchester.ac.uk measure of false positive rate, or by manual inspection (4). However, other groups worldwide may well have studied the same (species') proteome in the same or similar conditions, or may have studied a different developmental stage, and it would clearly be useful to determine what proteins and peptides other groups have identified. Furthermore, there is great value in the determination of 'proteotypic' peptides of a particular protein (5), i.e. those peptides that are repeatedly and consistently identified in a given protein over multiple experiments. The proteotypic peptides may be good 'markers' for a given protein or, more importantly, may be excellent candidates to produce as labelled internal standards for quantitation (5,6).

Following the explosion in sequenced genomes, proteomics is also experiencing a rapid growth in the amount of data that is produced and as a result there are a number of different proteomic data repositories. However, unlike sequence data, mature data standards for proteomics and related data are not yet available [although they are under development through the Proteomics Standards Initiative (PSI) (7)]. The main data repositories include PRIDE (Proteomics Identifications database) (8), GPM database (9), PeptideAtlas (5), GAPPdb (Genome Annotating Proteome Pipeline database) (10), PepSeeker (11), Tranche (12), PEDRO (13) and the OPD (Open Proteomic Database) (14), for a comprehensive review see Mead *et al.* (15). Many of these data repositories are now well established and continue to be populated with high quality mass spectrometric data and protein/peptide identifications. This increasing wealth of data is a goldmine of information that, when correctly collated, could be used to enhance our understanding of biological function through the proteins identified. Moreover, the mass spectrometry and proteomics experimental processes themselves could benefit from information contained in the proteome repositories. This could assist the validation of individual proteome experiments. For instance, one or more of these data repositories could be mined independently to determine all the different peptides observed for a particular protein in different experiments. Presently, this is not straightforward, as individual repositories contain results in different formats; collating these data into a common structure is clearly useful, but can be challenging and time consuming.

One can also imagine many typical data-driven tasks in the sequence domain having direct parallels with post-genomic fields such as proteomics. Bioscientists may wish to query and download data sets relating to their gene/protein(s) of interest and make direct comparisons between them, looking for similarities or differences. Furthermore, the ability to link different proteomics data with other biological features such as functional and structural annotations may prove useful when trying to understand why specific proteins and peptides are observed (or not) in different proteomics experiments. It should also permit an assessment of the impact of a given protein on a system, particularly as quantitative proteomic techniques evolve.

As part of the collection of services contained within the ISPIDER Central framework, we describe here a web server which enables scientists to carry out these integrative *in silico* experiments in proteomics. Users are able to query multiple proteomic data repositories and view the results using specialist viewers that ease the interpretation and browsing of the data, moreover, the viewers enable the user to invoke biological web services to add value to the identifications. We believe these services will be of great benefit to the proteomics and wider biological community wishing to see which proteins and peptides have been identified across the major proteomics repositories.

ISPIDER TOOLS AND RESOURCES

ISPIDER Central provides a range of services for integrative data analysis in mass spectrometry-based proteomics, and the systems architecture for the multiple database querying, data visualization and added-value web services is shown in Figure 1. This figure illustrates the overall project philosophy, where different proteomics repositories are queried, the results are integrated into a common format (*spidyXML*) and a variety of clients and views are available to examine the data. The features comprising this overall architecture are described below.

Proteomic data retrieval

The primary query engine is written using a simple Perl script whereby four of the main public proteomic data repositories, namely PRIDE (8), PepSeeker (11), the GPM database (9) and PeptideAtlas (16) can be queried by protein accession and/or peptide sequence. The query may also be limited by peptide or protein score, taken from the member repositories (when available). This approach makes use of the LWP (Library for World Wide Web in Perl) module, which enables information to be extracted directly via the HTTP protocol from external websites. Using this approach, queries over these repositories are

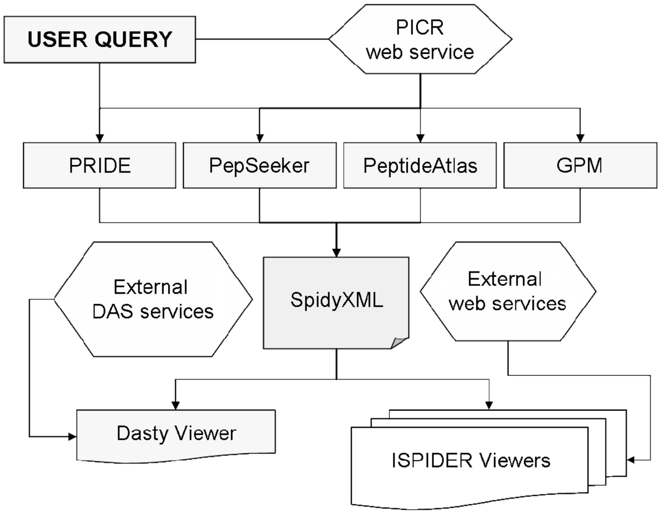


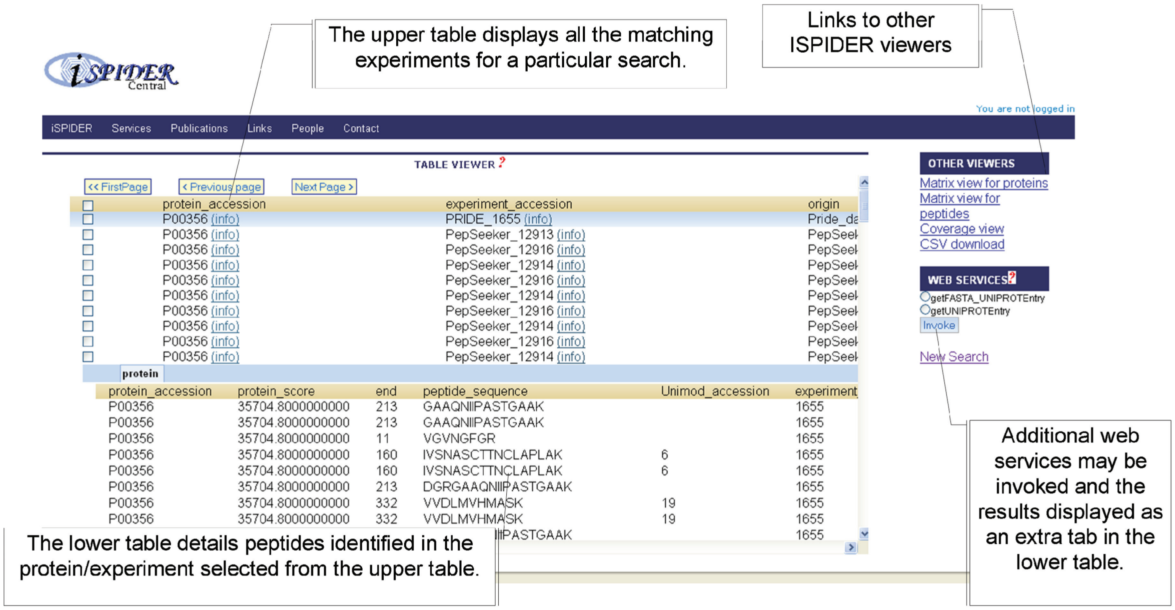
Figure 1. Flow chart of ISPIDER Central general architecture. The integration of the different results from independent queries over member repositories is achieved via *spidyXML*. This in turn may be processed by internal ISPIDER viewers or external clients. Objects in *spidyXML* may also be passed to external web services.

relatively straightforward since PeptideAtlas supports retrieval of results in an XML format from a simple URL, and both the PepSeeker and PRIDE repositories have a BioMart (17) interface which enables elegant XML-based querying and retrieval of data. The GPM database was slightly more complex; although, after stepwise LWP queries have been performed, the desired X!Tandem (18) XML output could be retrieved and the appropriate data extracted. Since all queries are performed on the fly, the latter method is slower than the direct XML queries described above. Users may select which of the repositories to search over via a set of checkboxes and, once complete, the outputs from the selected repositories are combined into a common internal format, *spidyXML* (see below).

A potential problem for this type of querying over a variety of databases containing protein data is that the same biological entity can be represented by different protein accession numbers or identifiers depending on the underlying sequence database that was searched in order to make the identifications. To overcome this, the ISPIDER web server provides the opportunity to first resolve accessions via the PICR (Protein Identifier Cross-Reference) web service (19) from the European Bioinformatics Institute (EBI), which translates between protein sequence identifier name spaces for a wide range of different databases. As a result, if selected, the ISPIDER Central resource can use the PICR service to determine all the possible matching accession numbers in a search for a single protein accession. The following queries over the member repositories are then executed for all the matching accession codes. Selecting this option is powerful, but can slow down searches substantially for some protein accessions which resolve into many pseudonyms.

In addition to the basic querying, the ISPIDER Central resource provides the means to add further information on to proteins retrieved from the searches. Users can thus discover more about the potential function of the identified proteins within the organism. This final service is integrated with other services through the ISPIDER interactive tabular viewer (see results section) and enables the integration of different web services with query results. The web service details are stored in a central registry (*ispiderRegistry*) which may be updated as new web services emerge. Currently, the registry contains two example services to retrieve corresponding UniProt accessions and protein sequences, although this list may be readily extended by adding protein-based web services. This exploits the common XML format delivered from the integrated ISPIDER database queries, as mentioned above, called *spidyXML*. This is a standard XML format that describes parent–child relationships between protein and peptide data retrieved from the repositories. The different data types described in *spidyXML* are consistent with the features described in the *ispiderRegistry*, a MySQL database containing details of all the web services available, including all their inputs and outputs, recorded as ISPIDER data types. For any parent attribute in *spidyXML* (and displayed in the top table in the viewer), the data type of that attribute is used to determine all the external web services from the *ispiderRegistry* that require this data type as input. Any suitable services are displayed in a menu to the right of the screen.

For example, in Figure 2, the parent attribute in the corresponding *spidyXML* is a protein accession, the ISPIDER data type for this is '*protein_accession*' and is hence displayed in the menu. To the right of the table in the figure are available web services that require a protein accession as input. By the selection of one or more protein accession numbers in the interactive table a user may simply invoke an available web service. The results of any web service are shown as an additional tab in the lower table.



The upper table displays all the matching experiments for a particular search.

Links to other ISPIDER viewers

	protein_accession	experiment_accession	origin
□	P00356 (info)	PRIDE_1655 (info)	Pride_da
□	P00356 (info)	PepSeeker_12913 (info)	PepSeek
□	P00356 (info)	PepSeeker_12916 (info)	PepSeek
□	P00356 (info)	PepSeeker_12914 (info)	PepSeek
□	P00356 (info)	PepSeeker_12916 (info)	PepSeek
□	P00356 (info)	PepSeeker_12914 (info)	PepSeek
□	P00356 (info)	PepSeeker_12916 (info)	PepSeek
□	P00356 (info)	PepSeeker_12914 (info)	PepSeek
□	P00356 (info)	PepSeeker_12916 (info)	PepSeek
□	P00356 (info)	PepSeeker_12914 (info)	PepSeek

protein_accession	protein_score	end	peptide_sequence	Unimod_accession	experiment
P00356	35704.8000000000	213	GAAQNIIPASTGAAK		1655
P00356	35704.8000000000	213	GAAQNIIPASTGAAK		1655
P00356	35704.8000000000	11	VGVNGFGR		1655
P00356	35704.8000000000	160	IVSNASCTTNCLAPLAK	6	1655
P00356	35704.8000000000	160	IVSNASCTTNCLAPLAK	6	1655
P00356	35704.8000000000	213	DGRGAAQNIIPASTGAAK		1655
P00356	35704.8000000000	332	VVDLMVHMASK	19	1655
P00356	35704.8000000000	332	VVDLMVHMASK	19	1655
			IIPASTGAAK		1655

The lower table details peptides identified in the protein/experiment selected from the upper table.

Additional web services may be invoked and the results displayed as an extra tab in the lower table.

Figure 2. The ISPIDER Interactive tabular view. Example results are shown for a protein accession-based query over multiple repositories.

THE ISPIDER SUITE OF VIEWERS

Results delivered in *spidyXML* may be examined by the user using a variety of clients and views, each with a slightly different focus, showing peptide and protein identifications for experiments from the repositories queried.

Nested view/Tabular view

This is the default page displayed to the end-user and is potentially the most comprehensive. The data is split into two interconnected tables whereby the top table, shown in Figure 2, contains a summary of the results and by clicking on each 'parent' the 'children' are shown in the lower table. What constitutes the 'parent' and 'children' is determined by the user at the query stage, and then refers directly to the *spidyXML*. For example, the parent can be a protein and the children a set of one or more peptides identified in that protein. Alternatively, the 'parent' can be an experiment from a proteomics data repository and the 'children' all the proteins or peptides identified in that experiment. This viewer itself is based on the Google toolkit (20), an open source java development framework, which facilitates the writing of AJAX (Asynchronous Javascript and XML) applications and dynamic HTML. The power of this framework is demonstrated in the 'value-added' application whereby for each parent various web services present in the *ispiderRegistry* can be employed to gather related information and is displayed in additional tabs in the lower table.

Matrix view

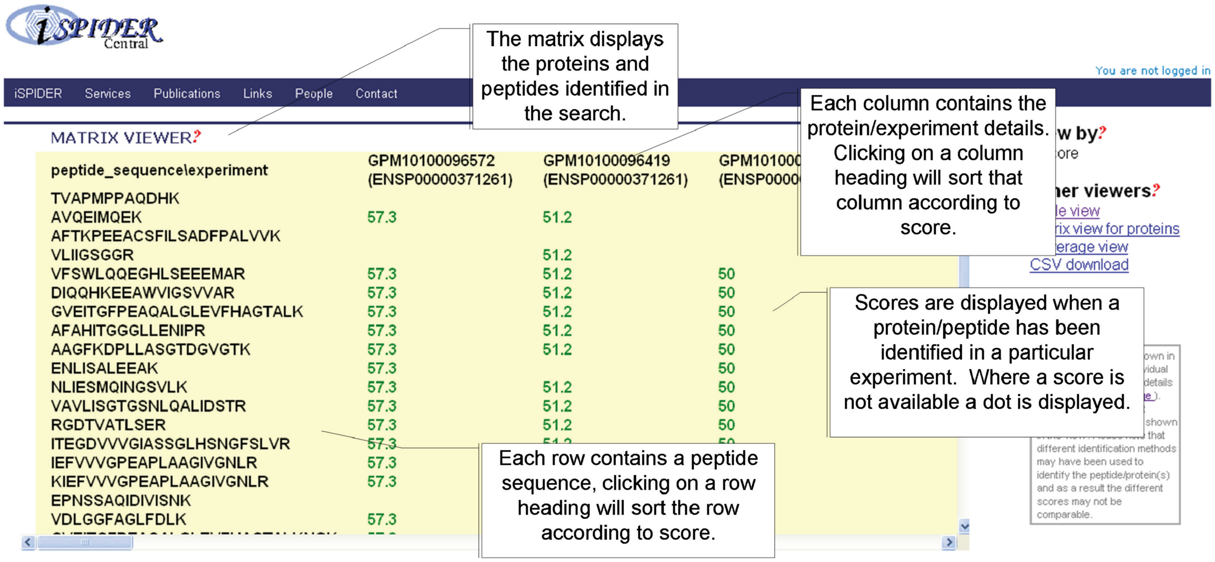
This view has also been developed with the Google toolkit (20) and allows a 2D 'matrix' type view with the 'parents' as column headings across the top and 'children' as rows along the left hand side (Figure 3). This view has many potential uses, depending on the query and parent–child relationship, not least the ability to look for the peptides/proteins which have been observed across multiple experiments. In addition, each attribute in the matrix can be coloured as a 'good' or 'bad' score according to a quality definition in the *spidyXML*, where available. Presently, protein/peptide score information is rather patchy in some of the repositories, although it is expected that standards being developed by the PSI (e.g. analysisXML) will address this. The Matrix view is also an interactive viewer, whereby clicking on a row/column heading will sort the row/column according to the score.

Coverage view

This view is written using the BioPerl (21) graphics module and provides a simple graphical view for each protein in the result set with the positions of all the matching peptide sequence. This is currently only available for those results where the start and end coordinates for each peptide are available.

Dasty2

Dasty2 (22) is a web client for visualizing protein sequence feature information using DAS (Distributed Annotation System) (23). The client establishes connections to the DAS reference server to retrieve sequence information and to one or more DAS annotation servers to retrieve feature information. Dasty2 then merges the information from all of the different servers and displays them in a web browser as a highly interactive graphical viewer. The Dasty2 client has been integrated into ISPIDER Central, using a Perl script to mimic a DAS annotation server, producing the appropriate XML from the *spidyXML* for a specific protein. This view, shown in Figure 4, enables the user to examine the peptide identifications retrieved for a given protein in the context of the large collection of functional and structural features associated with the parent protein obtained via Dasty2. This includes such relevant features



The matrix displays the proteins and peptides identified in the search.

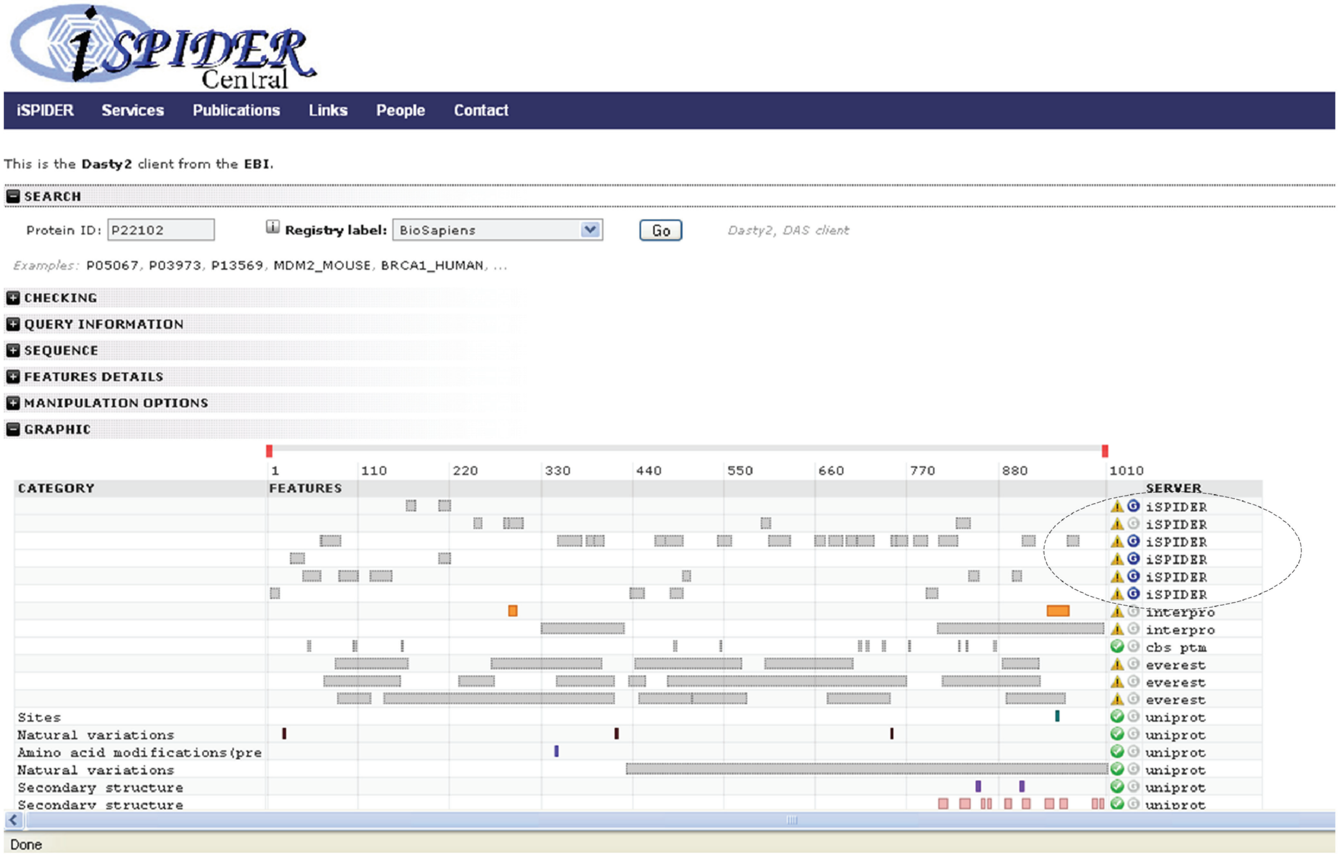
Each column contains the protein/experiment details. Clicking on a column heading will sort that column according to score.

Scores are displayed when a protein/peptide has been identified in a particular experiment. Where a score is not available a dot is displayed.

Each row contains a peptide sequence, clicking on a row heading will sort the row according to score.

peptide_sequence\experiment	GPM10100096572 (ENSP00000371261)	GPM10100096419 (ENSP00000371261)	GPM101000 (ENSP0000
TVAPMPPAQDHK			
AVQEIMQEK	57.3	51.2	
AFTKPEEACSFILSADFPALVVK			
VLIIGSGGR		51.2	
VFSWLQQEGHLSEEEMAR	57.3	51.2	50
DIQQHKEEAWVIGSVVAR	57.3	51.2	50
GVEITGFPEAQALGLEVFHAGTALK	57.3	51.2	50
AFAHITGGGLLENIPR	57.3	51.2	50
AAGFKDPLLASGTDGVGTK	57.3	51.2	50
ENLISALEEAK	57.3		50
NLIESMQINGSVLK	57.3	51.2	50
VAVLISGTGSNLQALIDSTR	57.3	51.2	50
RGDTVATLSER	57.3	51.2	50
ITEGDVVVGIASSGLHSNGFSLVR	57.3	51.2	50
IEFVVVGPEAPLAAGIVGNLR	57.3		
KIEFVVVGPEAPLAAGIVGNLR	57.3		
EPNSSAQIDIVISNK			
VDLGGFAGLFDLK	57.3		

Figure 3. The ISPIDER Interactive matrix view. An example is shown displaying peptide identifications cross-referenced against their parent proteins from a variety of experiments taken from repositories.



This is the **Dasty2** client from the **EBI**.

Protein ID: P22102 **Registry label:** BioSapiens

Examples: P05067, P03973, P13569, MDM2_MOUSE, BRCA1_HUMAN, ...

CATEGORY	FEATURES	SERVER
		iSPIDER
		iSPIDER
		iSPIDER
		iSPIDER
		iSPIDER
		iSPIDER
		interpro
		interpro
		cbs ptm
		everest
		everest
		everest
Sites		uniprot
Natural variations		uniprot
Amino acid modifications(pre		uniprot
Natural variations		uniprot
Secondary structure		uniprot
Secondary structure		uniprot

Figure 4. Dasty2 client displaying the ISPIDER search results. Results from an ISPIDER query may be linked out to the Dasty2 client, which shows peptide identifications in line with other annotations.

as secondary structure, disorder prediction (24) and post-translational modifications.

Comma Separated Variable download

Finally, a link is also provided to download a simple comma-delimited text file of the results which can be uploaded into external software for analysis by the user.

USAGE

This site is intended for use by biologists and proteomic researchers, wishing to compare/contrast protein and peptide identifications from different laboratories and in different experimental conditions. Similarly, we believe this will be beneficial to researchers looking to design internal standards for quantitation (25,26), e.g. proteotypic peptides which are commonly observed in the same protein across multiple experiments (5). The ISPIDER web server is able to capture data relating to individual experiments from different proteomic databases, which may include (depending on the database searched) proteins, peptides, genome loci, search engines used to make the identifications and the underlying database(s) searched, as well as various scores and confidence values associated with the identification. In addition, value may be added to these identifications through the use of external web services invoked from ISPIDER and also through the use of Dasty which graphically aligns the protein of interest with any identified peptides and other protein information from available DAS servers, such as PRINTS and PROSITE domains, phosphorylation sites and secondary structure information.

CONCLUSIONS

The ISPIDER Central resource reported here illustrates how integrative bioinformatics techniques can be applied to merge different repositories, and demonstrates a viable platform for data integration in proteomics. The resource itself represents one output in a larger data integration project for proteomics, exploring complementary technologies for the same purpose. These include a virtual integrated resource combining data from PRIDE (8), PEDRO (13) and PepSeeker (11), using the AutoMed heterogeneous data integration toolkit (27). Development of web services underpins much of this, either for integration into the existing framework or via proteome-specific workflows using myGrid (28,29). We aim to add these services to the ISPIDER portal, as well as provide a reanalysis pipeline for protein/peptide identification, to improve computation services for proteome biologists. The resource is accessible from http://www.ispider.manchester.ac.uk/cgi-bin/ProteomicSearch.pl.

ACKNOWLEDGEMENTS

The authors thank the Biotechnology and Biological Sciences Research Council (BBSRC) for support, in the form of the ISPIDER grant to all (refs, BBSB17239, BBSB17220, BBSB17204), technical support from Dr Nick Gresham and Eric W Deutsch of the ISB Seattle for assistance with PeptideAtlas. Funding to pay the Open Access publication charges for this article was provided by JISC.

Conflict of interest statement. None declared.

REFERENCES

1. Blagoev,B. and Mann,M. (2006) Quantitative proteomics to study mitogen-activated protein kinases. *Methods*, **40**, 243–250.
2. Bharti,A., Ma,P.C. and Salgia,R. (2007) Biomarker discovery in lung cancer—promises and challenges of clinical proteomics. *Mass Spectrom Rev.*, **26**, 451–466.
3. Desiere,F., Deutsch,E.W., Nesvizhskii,A.I., Mallick,P., King,N.L., Eng,J.K., Aderem,A., Boyle,R., Brunner,E., Donohoe,S. *et al.* (2005) Integration with the human genome of peptide sequences obtained by high-throughput mass spectrometry. *Genome Biol.*, **6**, R9.
4. Matthiesen,R. (2007) Methods, algorithms and tools in computational proteomics: a practical point of view. *Proteomics*, **7**, 2815–2832.
5. Mallick,P., Schirle,M., Chen,S.S., Flory,M.R., Lee,H., Martin,D., Raught,B., Schmitt,R., Werner,T., Kuster,B. *et al.* (2007) Computational prediction of proteotypic peptides for quantitative proteomics. *Nat. Biotechnol.*, **25**, 125–131.
6. Craig,R., Cortens,J.P. and Beavis,R.C. (2005) The use of proteotypic peptide libraries for protein identification. *Rapid Commun. Mass Spectrom.*, **19**, 1844–1850.
7. Taylor,C.F., Hermjakob,H., Julian,R.K., Garavelli,J.S., Aebersold,R. and Apweiler,R. (2006) The work of the Human Proteome Organisation's Proteomics Standards Initiative (HUPO PSI). *OMICS*, **10**, 145–151.
8. Jones,P., Cote,R.G., Martens,L., Quinn,A.F., Taylor,C.F., Derache,W., Hermjakob,H. and Apweiler,R. (2006) PRIDE: a public repository of protein and peptide identifications for the proteomics community. *Nucleic Acids Res.*, **34**, D659–D663.
9. Craig,R., Cortens,J.P. and Beavis,R.C. (2004) Open source system for analyzing, validating, and storing protein identification data. *J. Proteome Res.*, **3**, 1234–1242.
10. Shadforth,I., Xu,W.B., Crowther,D. and Bessant,C. (2006) GAPP: a fully automated software for the confident identification of human peptides from tandem mass spectra. *J. Proteome Res.*, **5**, 2849–2852.
11. McLaughlin,T., Siepen,J.A., Selley,J., Lynch,J.A., Lau,K.W., Yin,H.J., Gaskell,S.J. and Hubbard,S.J. (2006) PepSeeker: a database of proteome peptide identifications for investigating fragmentation patterns. *Nucleic Acids Res.*, **34**, D649–D654.
12. Falkner,J.A., Falkner,J.W. and Andrews,P.C. (2007) ProteomeCommons.org IO Framework: reading and writing multiple proteomics data formats. *Bioinformatics*, **23**, 262–263.
13. Garwood,K., McLaughlin,T., Garwood,C., Joens,S., Morrison,N., Taylor,C.F., Carroll,K., Evans,C., Whetton,A.D., Hart,S. *et al.* (2004) PEDRo: a database for storing, searching and disseminating experimental proteomics data. *BMC Genomics*, **5**, 68.
14. Prince,J.T., Carlson,M.W., Wang,R., Lu,P. and Marcotte,E.M. (2004) The need for a public proteomics repository. *Nat. Biotechnol.*, **22**, 471–472.
15. Mead,J., Shadforth,I. and Bessant,C. (2007) Public proteomic MS repositories and pipelines: available tools and biological applications. *Proteomics*, **7**, 2769–2786.
16. Pedrioli,P.G.A., Eng,J.K., Hubley,R., Vogelzang,M., Deutsch,E.W., Raught,B., Pratt,B., Nilsson,E., Angeletti,R.H., Apweiler,R. *et al.* (2004) A common open representation of mass spectrometry data and its application to proteomics research. *Nat. Biotechnol.*, **22**, 1459–1466.
17. Durinck,S., Moreau,Y., Kasprzyk,A., Davis,S., De Moor,B., Brazma,A. and Huber,W. (2005) BioMart and Bioconductor: a powerful link between biological databases and microarray data analysis. *Bioinformatics*, **21**, 3439–3440.
18. Craig,R. and Beavis,R.C. (2004) TANDEM: matching proteins with tandem mass spectra. *Bioinformatics*, **20**, 1466–1467.
19. Cote,R.G., Jones,P., Martens,L., Kerrien,S., Reisinger,F., Lin,Q., Leinonen,R., Apweiler,R. and Hermjakob,H. (2007) The protein identifier cross-referencing (PICR) service: reconciling protein identifiers across multiple souce databases. *BMC Bioinformatics*, **8**, 401–416.
20. Goth,G. (2007) The Google Web Toolkit shines a light on Ajax frameworks. *Ieee Software*, **24**, 94–98.
21. Stajich,J.E., Block,D., Boulez,K., Brenner,S.E., Chervitz,S.A., Dagdigian,C., Fuellen,G., Gilbert,J.G.R., Korf,I., Lapp,H. *et al.* (2002) The bioperl toolkit: Perl modules for the life sciences. *Genome Research*, **12**, 1611–1618.
22. Jones,P., Vinod,N., Down,T., Hackmann,A., Kahari,A., Kretschmann,E., Quinn,A., Wieser,D., Hermjakob,H. and Apweiler,R. (2005) Dasty and UniProt DAS: a perfect pair for protein feature visualization. *Bioinformatics*, **21**, 3198–3199.
23. Dowell,R.D., Jokerst,R.M., Day,A., Eddy,S.R. and Stein,L. (2001) The distributed annotation system. *BMC Bioinformatics*, **2**, 7.
24. Ward,J.J., Sodhi,J.S., McGuffin,L.J., Buxton,B.F. and Jones,D.T. (2004) Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. *J. Mol. Biol.*, **337**, 635–645.
25. Gerber,S.A., Rush,J., Stemman,O., Kirschner,M.W. and Gygi,S.P. (2003) Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS. *Proc. Natl Acad. Sci. USA*, **100**, 6940–6945.
26. Beynon,R.J., Doherty,M.K., Pratt,J.M. and Gaskell,S.J. (2005) Multiplexed absolute quantification in proteomics using artificial QCAT proteins of concatenated signature peptides. *Nat. Methods*, **2**, 587–589.
27. Zamboulis,L., Fan,H., Belhajjame,K., Siepen,J., Jones,A., Martin,N., Poulovassilis,A., Hubbard,S., Embury,S. and Paton,N. (2006) Data access and integration in the ISPIDER proteomics grid. In *Data Integration in the Life Sciences*. Springer, Hinxton, UK.
28. Stevens,R.D., Tipney,H.J., Wroe,C.J., Oinn,T.M., Senger,M., Lord,P.W., Goble,C.A., Brass,A. and Tassabehji,M. (2004) Exploring Williams-Beuren syndrome using myGrid. *Bioinformatics*, **20 (Suppl. 1)**, i303–i310.
29. Zamboulis,L., Martin,N. and Poulovassilis,A. (2007) Bioinformatics service reconciliation by heterogeneous schema transformation. In *Data Integration in the Life Sciences*. Springer, Philadelphia, PA, USA.