Proteomic databases and software on the web

Jérôme Wojcik and Vincent Schächter

Date received (in revised form): 16th May 2000

Jérôme Wojcik and **Vincent Schächter** are respectively Project Leader and Head of Research in bioinformatics at Hybrigenics. They hold PhDs in computational biology and computer science from the University of Paris. Hybrigenics is a biotechnology/bioinformatics company specialising in functional genomics and proteomics applied to the discovery of novel validated drug targets and corresponding lead compounds. Its proprietary data and technologies revolve around the construction, analysis and exploitation of protein interaction maps.

Abstract

In the wake of sequencing projects, protein function analysis is evolving fast, from the careful design of assays that address specific questions to 'large-scale' proteomics technologies that yield proteome-wide maps of protein expression or interaction. As these new technologies depend heavily on information storage, representation and analysis, existing databases and software tools are being adapted, while new resources are emerging. This paper describes the proteomics databases and software available through the World-Wide Web, focusing on their present use and applicability. As the resource situation is highly transitory, trends and probable evolutions are discussed whenever applicable.

Keywords: *functional proteomics, 2D PAGE, protein–protein interactions, protein linkage maps, proteomic databases*

INTRODUCTION

The field of proteomics deals with the global analysis of the 'proteome',[1] the collection of proteins produced by an organism, a tissue or a cell type.

The recent evolution of high-throughput techniques has positioned proteomics as the post-genomic discipline of choice to identify protein function 'in context'. Indeed, it has now become clear that knowledge of the genomic sequence is merely a first step towards prediction of the behaviour of gene products. However, proteins, not genes, sustain function. First, protein expression can be regulated post-transcriptionally, in which case protein expression does not correlate exactly to gene expression.[2] Second, protein function(s) and role(s) are further controlled by post-translational modifications, turn-over, dynamic behaviours or subcellular localisation. Furthermore, protein function may be characterised at different levels: for example, a molecular interaction is typically but one component of one or several processes/pathways. Proteomics, encompassing genetic and environmental factors, appears to be able to fill the gap between genomics and understanding of cellular behaviour.

The term 'proteomics' can be seen as referring to two distinct fields: 'classical' proteomics (identification of proteins, differential protein expression) and 'functional' proteomics (identification of protein interactions). Underlying technologies and analysis software differ radically, although strong bioinformatics post-processing is required in both cases.

The current proteomic databases and software available on the Internet are reviewed. The aim is neither to provide an exhaustive list of resources, nor to describe proteomic techniques precisely, but rather to provide a snapshot of the (fast-evolving) situation in terms of available resources and to show how an experimental biologist – not a proteomics specialist – can use the available data to gather clues about the function of his/her proteins of interest.

CLASSICAL PROTEOMICS

2D PAGE and mass spectrometry

High-resolution 2D gel sodium dodecyl sulphate (SDS)–polyacrylamide gel electrophoresis (2D PAGE) is the prevailing technique for analysis of the whole protein expression profile of a given cell type or organism, under a set of specific conditions. Typically, proteins

Jérôme Wojcik,
Hybrigenics SA, 180 av. Daumesnil, Paris 75012, France

Tel: +33 1 70912900
Fax: +33 1 70912948
E-mail: jwojcik@hybrigenics.fr

in a sample are separated according to both isoelectric point (pI) and molecular weight (Mw), by a combination of isoelectric focusing and electrophoresis respectively. Spots on the gel are then excised for further identification using mass spectrometry techniques coupled to protein sequence database searching (see below and reference 3). Image analysis software completes the process by accelerating and facilitating spot quantification and comparison.

2D PAGE technology

2D PAGE technology can be used to separate complex protein mixtures into their individual components, to compare expression profiles of sample pairs (normal versus transformed cells, cells at different stages of growth, etc.), and to determine the global protein response of a cell or tissue to a given set of conditions. In short, it provides global information about relative protein abundance, post-translational modifications or coregulation, without requiring prior knowledge of the studied proteome. Limitations include difficulty in detecting rare proteins, and the lack of maturity of the first bioinformatics post-processing stages, ie automated quantitative analysis and gel comparison software.

mass spectrometry

Since reaching high levels of sensitivity, automation and throughput for massive protein analysis, mass spectrometry has become one of the key technologies in the proteomics field.[4] Analysing femtomoles of protein materials after excision from 2D gels and proteolytic digestion is now routine using MALDI/TOF (Matrix-Assisted Laser Desorption/Ionisation/Time-Of-Flight)-based peptide mass fingerprinting. A protein can be rapidly identified by peptide mass mapping if its mass fingerprint fits one of the fingerprints computed from a known sequence. When there is no immediate fit (eg the protein sequence is unknown), a combination of mass map information and of sequence information from additional tandem mass spectrometry experiments (sequence tagging) can be used to search sequence databases, reconstruct the whole sequence and identify the protein.[5,6]

Finally, mass spectrometry coupled with high-performance liquid chromatography techniques[7] and/or combined with biochemical techniques (immunoprecipitation) allows fast identification of proteins present in complex biological mixtures, and may be used to study protein–protein interaction, to locate and identify single protein or protein complexes from a subcellular fraction.

Resources available on the Internet

2D PAGE databases

Table 1 presents a list of the major 2D gel databases available on the Internet. Although standardised and unified 2D gel 'meta-databases' (see review,[8] or National Cancer Institute's ((NCI's)) 2DWG[9,16]) or index sites (WORLD-2DPAGE[11]) are emerging, Internet resources are still mostly scattered and heterogeneous. Additional documentation, general-purpose information and pointers can be found at various URLs.[12–14] A web site dedicated to mass spectrometry[15] will provide to the interested reader much up-to-date information on this very fast-evolving proteomics field.

Specific 2D PAGE databases typically feature a bank of 2D gel images from a specific tissue or organism; identified spots are highlighted and varying amounts of gel annotation are provided. Most structured databases provide clickable map functionality as well: clicking on a spot leads the user to a protein 'reference card', featuring various experimental information, annotations and cross-references to classical protein databases. Many databases also provide protein query or image manipulation tools.

Finally, 2DHunt[16] is a specialised search engine for 2D PAGE resources on the web.[17]

Gel image analysis may range from very basic to fairly complex, depending

Table 1: Online classical proteomics resources

Database	URL	Organisms/organelles/phenotypes	Tools
Aarhus 2D PAGE database	http://biobase.dk/cgi-bin/celis	Human (skin biology, bladder cancer, fibroblasts), mouse (kidney)	*abcd*
Aberdeen 2D PAGE	http://www.abdn.ac.uk/~mmb023/2dhome.htm	*Haemophilus influenzae*	*ab(c)*
Argone protein mapping group	http://www.anl.gov/BIO/PMG/	Human (breast), mouse (liver), *Pyrococcus furiosus*	*a*
Cyano2Dbase	http://www.kazusa.or.jp/cyano/cyano2D/	*Synechosystis* cyanobacteria	*a(c)*
ES cell-2D PAGE	http://www.dur.ac.uk/~dbl0nh1/2DPAGE/	Embryonal stem cells	–
Harefield HSC 2D PAGE	http://www.harefield.nthames.nhs.uk/nhli/protein/	Human (heart, endothelial cells), rat (heart), dog (heart)	*abcd*
Human colon carcinoma 2D PAGE	http://www.ludwig.edu.au/jpsl/JP2DDAT.html	Human (colon)	–
Indiana University 2D PAGE	http://iupucbio1.iupui.edu/frankw/molan.htm	Rat (kidney, liver, brain, testis, plasma), mouse (liver, plasma), human (liver, plasma), bovine (testis)	*ab(c)*
Maize Genome database	http://moulon.moulon.inra.fr	Maize	*abd*
Maritime pine 2D PAGE	http://www.pierroton.inra.fr/genetics/2D/	Maritime pine	*ab(c)*
Max-Planck Institut 2D PAGE	http://www.mpiib-berlin.mpg.de/2D-PAGE/	*Borrelia garinii*, *Mycobacterium bovis BCG*, *Mycobacterium tuberculosis*, Jurkat T- cells	*abcd*
MDC Heart-2D PAGE	http://www.mdc-berlin.de/~emu/heart/	Human (heart)	*abc(d)*
Parasite Host Cell Interaction 2D PAGE	http://www.gram.au.dk	HeLa and IFN regulation	*abc(d)*
Plant Plasma Membrane Database	http://sphinx.rug.ac.be:8080/ppmdb/index.html	*Arabidopsis thaliana*	*abc(d)*
Rat Heart-2D PAGE	http://gelmatching.inf.fu-berlin.de/~pleiss/2d/	Rat (heart)	*abc(d)*
Rat Serum Protein study group	http://linux.farma.unimi.it/homeframed.html	Rat (serum)	*a(c)*
SWISS-2D PAGE	http://www.expasy.ch/ch2d	Human, mouse, yeast, *Escherichia coli*, *Dictyostelium discoideum*	*abc(d)*
SIENA-2D PAGE	http://www.bio-mol.unisi.it/2d/2d.html	Human (breast ductal carcinoma), *Chlamydia trachomatis*, *Caenorhabditis elegans*	*abc(d)*
SSI-2D PAGE	http://www.ssi.dk/en/forskning/tbimmun/tbhjemme.htm	*M. tuberculosis*	*ac*
Sub2D	http://microbio2.biologie.uni-greifswald.de:8880/sub2d/pub/sub2d.ask	*Bacillus subtilis*	*abc*
TMIG 2D PAGE	http://proteome.tmig.or.jp/2D/	Human (age-related proteome)	*abcd*
ToothPrint	http://bioc111.otago.ac.nz:8001/tooth/home.htm	Rat (dental tissues)	*abc*
UCSF 2D PAGE	http://rafael.ucsf.edu/2DPAGEhome.html	A375 cell line	–
Université Paris 13 2D PAGE	http://www-smbh.univ-paris13.fr/lbtp/Biochemistry/biochimie/bque.htm	Human (hematopoietic and *bc* lymphoid cell lines)	
2DWGDB (WebGel)	http://www-lmmb.ncifcrf.gov/2dwgDB	**(Meta-database: compilation of various gels from other databases)**	*abcd*
WU Inner Ear database	http://oto.wustl.edu/thc/innerear2d.htm	Human, mouse, pig, chick (inner ear)	*abc*
Yeast 2D PAGE	http://yeast-2dpage.gmm.gu.se/	Different yeast species	*(c)*
Yeast Protein Map (YPM)	http://www.ibgc.u-bordeaux2.fr/YPM/	*Saccaromyces cerevisiae*	*ac*

This table displays a list of the major databases dedicated to 2D PAGE, with their covered organisms. A few general-purpose databases including 2D PAGE data are listed as well. Available tools: *a* clickable gel; *b* querying tools; *c* annotated proteins (onto the gel or with cross-references); *d* gel image analysis/manipulation tools such as zoom in/out or mark/unmark spot according to protein properties.

2D PAGE analysis tools

on how much automation is required. Several commercial 2D gel image analysis software packages are available. They integrate a number of useful tools for the 2D PAGE specialist: functionalities include display, analysis and comparison of gel images, as well as determination, quantification and normalisation of spots (Melanie, Geneva Bioinformatics;[18] PDQuest, Bio-Rad; Phoretix advanced, etc.). Non-specialists can resort to Flicker,[19,20] a free web tool for comparing images from different Internet sources

Flicker

(including your own web or ftp server) on your browser. Given two gel image URLs, Flicker loads the images and displays them in your web browser. They can be enhanced in various ways (spatial warping, pseudo 3-dimensional, image sharpening, etc.), while regions of interest can be 'landmarked' with several corresponding points in each gel image. One gel image is then warped to the geometry of the other, and the two resulting images are compared visually in a third window (the 'flicker' window): as the two gels are rapidly alternated ('flickered'), the user can slide one gel past the other to visually align corresponding spots by matching local morphology.

Mass spectrometry data can be quickly and accurately analysed by commercial suites of software tools such as ProteinLynx from Micromass[21] or ProteinProspector from the UCSF.[22] In addition, there are numerous free tools to help protein identification.[23]

Using 2D PAGE databases

choosing a 2D PAGE database

First, suitable database resources must be selected. Unless the organism of interest is very specific and referenced only once in a database in Table 1, we recommend using WORLD-2DPAGE or better yet, the 2DWG meta-database as a starting point. These sites will help locate the gel(s) of interest available on the web and, in most cases, the latter will provide all the useful tools to analyse them. If this first search fails, one can use the 2DHunt software or manually browse the resources listed in Table 1 to try locating the data. In the latter case, as gel databases differ mostly in scope – organisms covered, number of gels – and by the set of available tools, we suggest starting the search with the database providing the richest set of tools (see Table 1, 'Tools' column).

gel analysis

functional annotation from 2D gels

WebGel's 2DWG is a meta-database of gel images available on the web, compiled using a combination of review of results from web searching and submissions by web database authors.[9] The database is organised as a spreadsheet table, with each gel image represented by a row. Gel records can be retrieved by typing a set of keywords related to the tissue or fluid type, or choosing from a list of tissue types. The search can be restricted to entries featuring map images (ie an active or passive spot map identifies proteins on the raw gel image) or raw gel images. Data associated with a gel entry in 2DWG includes tissue, species, cell-line, image URL, database URL, organisation URL, image properties and map URL if it exists, as well as details on the gel protocol such as the gel type, pH range or detection method. Most specialised databases roughly provide the same type of information.

Once one or several sources are selected, a number of analyses can be performed, ranging from the straightforward to the fairly complex. We list here a short description of the possible analyses (for a detailed example, see reference 24). Assuming the starting point of the search is a human protein P of interest, we list below a number typical basic 'functional' questions that a non-specialised biologist may try to answer using 2D gel data.

- Identification of potential subcellular and/or tissue locations, by retrieving all the gels where P is expressed and looking at corresponding experimental conditions.

- Detection of phenotype-associated proteins by comparing a specific-condition gel to a wild-type gel. If P is expressed differently in a wild-type tissue and in the same tissue under another conditions, P is probably directly or indirectly involved in the phenotype difference (for examples, see references 24 and 25). The same approach can be followed to detect changes between tissues at different development stages, to assess genetic variability between proteins from

related species (for review, see reference 26), or to identify parasite proteins by comparing expression in wild-type and infected host cells.[27]

- Identification of post-translational modifications: a protein P subject to such transformations will be represented by several spots with different pI/Mw couples. Moreover, mass spectrometry techniques exist that precisely identify the post-translational modifications.[28] These modifications are annotated in the SWISS-PROT database, as well as in some 2D PAGE databases.

functional annotations from 2D gels

- Detection of co-regulated proteins: if P is expressed in several tissues or in the same tissue under different conditions, comparing the different expression 'fingerprints' can help identify other proteins with the same expression pattern as P. However, these comparisons (performable for instance with Flicker) are sometimes too delicate to operate because of the physical non-reproducibility of the gels. Once a co-regulated protein has been identified, it can be used as a new probe to scan classical protein and 2D PAGE databases in an iterative process.

functional proteomics

- Determination of the relative abundance of different proteins (or of different protein variants) if the corresponding experimental data on spot intensity are available.

- Establishment of genetic distances: 2D gels can sometimes be helpful in establishing such distances between P and its homologous proteins in closely related species. These distances can be used to extract phylogenetic relationships (for review, see reference 26).

In summary, 2D PAGE analysis helps to determine protein function, mostly by providing sets of proteins correlated to the initial protein from which functional hints can be taken. However, the number of possible 'neighbour' proteins is often significant and additional clues are required to narrow the search.

Future improvements in classical proteomics include optimisation of sample preparation and detection methods,[29] experimental protocol standardisation and automation,[30] database format harmonisation[5] and better bioinformatic post-processing and/or gel distortion control.

For now, the principal limitation to the development and scale-up of 2D PAGE is probably the gel comparison process, resulting in a data interpretation and analysis bottleneck. Until the output images of different 2D PAGE experiments – yielding different gel shapes – can be reliably compared, tedious and time-intensive manual analysis remains necessary. Research in this field is active (see for example reference 31) and significant progress on image analysis issues may yield qualitative as well as quantitative changes in the biological information that can be extracted from 2D gel data.

FUNCTIONAL PROTEOMICS

Protein interaction maps: towards a new approach of function

For methodological reasons – partly because of the way functional annotations are propagated using sequence homology – 'function' has long been treated as a non-contextual attribute of a given protein.[32] Practically speaking, databases represent function by a keyword or set of keywords, coming at best from a list or a fixed functional hierarchy. Clearly, however, this is an oversimplification that does not convey the distinction between 'functional levels' (such as 'molecular' versus 'cellular' function), and does not accurately translate the fact that the same protein or protein domain can participate in several distinct cellular processes (eg metabolic or signal

transduction pathways). Recent advances in high-throughput technologies and bioinformatics analysis open the way to more accurate definitions of the function of a protein, that take its context within the cell into account. In particular, the aim of functional proteomics is to describe the function of a given protein from the global pattern of its molecular interactions.

protein interaction map databases

interactome

high throughput two hybrid

protein interaction map analysis tools

Several techniques are available to explore the 'interactome' (defined as the pattern of interactions of a proteome[33]). Classical chemical techniques (affinity chromatography, coimmunoprecipitation or cross-linking) are accurate and provide high-quality results, but are often hard to apply on a large scale. In contrast, two-hybrid strategies can be applied at high throughput and offer the first real functional proteomics tool to examine the whole interactome.[34–37] In return, they require some minimal initial knowledge of the proteome to choose a starting point, high-quality control standards to avoid false negatives, and bioinformatics post-processing to filter out false positives.

A number of computational techniques for interaction prediction, based on sequence data and full-genome comparisons[38,39] or on bibliographical data,[40] are starting to emerge as well. So far, all of them show theoretical limitations, and false positives as well as false negatives abound: additional information is thus required to reach conclusive evidence on both the existence and the nature of a given predicted interaction. Combination of different prediction techniques may be a promising approach to reducing this uncertainty.[41]

Resources available on the Internet

As significant amounts of protein interaction data started accumulating very recently, web resources are still scarce. Earlier interaction databases merely provide a basic display of the alphabetical interaction list (with annotations or cross-references to other protein databases) and some basic query tools. Among these, the Database of Interacting Proteins, DIP,[42,43] includes interactions from a variety of different organisms, entered by curators on the basis of journal publications. Data on protein–protein interactions in yeast can be found at MIPS,[44] while the FlyNets interaction list[33,45] combines known protein–DNA, protein–RNA and protein–protein interactions in *Drosophila melanogaster*.

Newer databases tend to better structure the interactome model in order to offer real navigation tools. Interact[46,47] is based on a simple object-oriented model and displays the yeast interactome in 3D (requires downloading of a VRML viewer) but provides no real research tool. FlyNets is also developing a graphical display on-line tool.[33] For now, only commercial software, featuring protein interaction map navigation tools built on relational databases, allows actual exploration of the interactome. The three main software currently available are the PIMrider,[48] (Hybrigenics), the CuraGen data analysis software[49] (CuraGen) and Myriad Genetics' ProNet.[50] All allow local exploration of the interaction map (travel from neighbour to neighbour, by collapsing and expanding the interaction net around a node) and display of annotations and cross-references for each protein node. The CuraGen software also features the possibility of comparing protein interaction maps of different organisms. In addition, the PIMrider lets the user search for pathways between two proteins and provides a reliability score for each displayed interaction, as well as specific interaction domains in a pair of partners.

Riding the protein interaction maps

What would be the typical use of an interaction map navigation software to help functional annotation of proteins?

functional annotation from protein interaction maps

Assuming P is an unknown protein of interest, the first step is to try P as an entry point to the interactome, through a basic query to the protein interaction map database:

- If P is identified in a given protein interaction map, the navigator will display all of its neighbours:
  - their annotations (if any) can then be checked;
  - the search can be expanded by selecting neighbours of neighbours, and reiterating the process until a coherent sub-map is extracted;
  - functional annotations common to proteins of this subset can then be searched, and tentatively assigned to P;
  - the interaction pattern of the subset can be mapped and compared with existing signal transduction or metabolic pathways, helping infer the function of proteins other than P;
  - if data on the domain of a given interaction are provided as well, functional annotations in domain databases may be checked for more accurate clues.

- If P is not yet present in any existing interaction map, the user may scan protein interaction maps using sequence homology, find the nearest orthologous protein and apply the previous protocol.

future proteomics

functional proteomics quality requirements

Although the main choice criterion at the present stage in the development of high-throughput functional proteomics is the availability of the interaction data on a given organism, data quality and reliability rank as close seconds. As many of the methods used to identify protein–protein interactions are prone to false positives as well as false negatives, caution should be exercised in interpreting interactions provided without a clear explanation of the underlying protocol and of the (preferably quantitative) procedure used to extract putative interactions from raw experimental data.

CONCLUSION AND PERSPECTIVES

Proteomics aims at determining the nature and quantity of proteins present in biological samples, at identifying interactions between proteins, and ultimately at understanding protein function. We have reviewed the main proteomics resources available on the web: 2D PAGE databases are numerous and scattered, and exhibit important variability in quality as well as coverage, whereas dedicated functional proteomics resources are only starting to emerge, enabling exploitation of the first wave of proteome-wide interaction data.

Databases and software resources are bound to evolve fast, however; first following and then probably shaping the current scale-up of the underlying proteome data production technologies. As sequencing projects reach or approach completion, proteomics is increasingly targeted as the next massive wave of collaborative research effort (see for example the Human Proteomics Initiative site[51]). Indeed, it will become increasingly difficult for single groups to support large-scale proteomics projects or maintain corresponding databases. Future proteomics projects are therefore increasingly likely to result from collaboration between several academic and private organisations, built around the sharing of experimental data and post-analysis results. For these to succeed, establishing experimental quality-control standards, setting benchmarks and deciding on common data representations are critical issues, on a par with process automation and the specific technological improvements already mentioned.[5,52] Depending on progress on these issues, the integration of heterogeneous data and development of higher-level tools capable of querying them meaningfully will be considered.

multiple integrated resources

For the time being, the databases mentioned in this review can best be exploited in conjunction with other resources related to protein function:[53] general-purpose protein databases such as SWISS-PROT[54] or 3D structure databases such as the PDB,[55] but also domain or profile/family databases such as PROSITE,[56] Pfam,[57] PRINTS[58] or ProDom[59] (now merged in the InterPro database[60]) and numerous specialised software tools (see for example reference 23), purposes ranging from protein identification and characterisation to similarity searches or structure determination. As proteomics databases or exploration software mature, they may integrate some of these resources so as to provide a more homogeneous working environment for functional analysis.

protein-linkage maps

Both 2D PAGE and two-hybrid technologies, by allowing the construction of proteome-wide 'protein-linkage maps',[61] enable a change in perspective: the 'function' of a given protein may be conceived and represented as a context-dependent pattern of expression and interaction, rather than as an intrinsic property of each individual molecule (for review, see reference 32). For instance, the global analysis of expression patterns of subsets of proteins that correlate with a given cell state could lead to the assignment of 'proteomic signatures'[62] that may be used to link genotype and phenotype. One step further lies the integration of protein linkage maps with the information available on signal transduction pathways and metabolic pathways, which may lead to the understanding, identification and assignment of 'higher-level' functions to proteins.[32,63,64] There is no doubt that the continuing development of this shift in perspective depends heavily on the availability of software platforms that integrate the relevant heterogeneous data types in biologically relevant ways.

Acknowledgements
We thank Dr Whiteside and Dr Legrain for their numerous comments on this review. We are also grateful to Dr Zeev Smilanski for his valuable insights on the future of 2D PAGE analysis and to Dr Collura for his precious contribution on mass spectrometry.

References

1. Wilkins, M. R., *et al.* (1996), 'Progress with proteome projects: why all proteins expressed by a genome should be identified and how to do it', *Biotechnol. Genet. Eng. Rev.*, Vol. 13, pp. 19–50.
2. Gygi, S. P., *et al.* (1999), 'Correlation between protein and mRNA abundance in yeast', *Mol. Cellular Biology*, Vol. 19(3), pp. 1720–1730.
3. Shevchenko, A., *et al.* (1996), 'Linking genome and proteome by mass spectrometry: large-scale identification of yeast proteins from two dimensional gels', *Proc. Natl. Acad. Sci. USA*, 1996. Vol. 93(25), pp. 14440–14445.
4. Lottspeich, F. (1999), 'Proteome analysis: a pathway to the functional analysis of proteins', *Angew. Chem. Int. Ed. Engl.*, Vol. 38(17), pp. 2476–2492.
5. Quadroni, M. and James, P. (1999), 'Proteomics and automation', *Electrophoresis*, Vol. 20(4–5), pp. 664–677.
6. Yates, J. R., 3rd, (1998), 'Mass spectrometry and the age of the proteome', *J. Mass Spectrometry*, Vol. 33(1), pp. 1–19.
7. http://www.micromass.co.uk
8. Oh, J. M., Hanash, S. M. and Teichroew, D. (1999), 'Mining protein data from two-dimensional gels: tools for systematic post-planned analyses', *Electrophoresis*, Vol. 20(4–5), pp. 766–774.
9. Lemkin, P. F. *et al.* (1999), 'Exploratory data analysis groupware for qualitative and quantitative electrophoretic gel analysis over the Internet-WebGel', *Electrophoresis*, Vol. 20(18), pp. 3492–3507.
10. http://www.lmmb.ncifcrf.gov/2dwgDB
11. http://www.expasy.ch/ch2d/2d-index.html
12. Phoretix International, http://www.phoretix.com
13. GELS2D at Info-biogen, http://www.infobiogen.fr/services/deambulum/english/d64.html
14. NCI, http://www_lmmb.ncifcrf.gov/EP/table2Ddatabases.html

15. http://base-peak.wiley.com

16. http://www.expasy.ch/ch2d/2DHunt/

17. Hoogland, C. *et a*l. (1999), 'Two-dimensional electrophoresis resources available from ExPASy', *Electrophoresis*, Vol. 20(18):, pp. 3568–3571.

18. Melanie, Geneva Bioinformatics, http://www.expasy.ch/melanie

19. Lemkin, P. F. and Thornwall, G. (1999), 'Flicker image comparison of 2-D gel images for putative protein identification using the 2DWG meta-database', *Mol. Biotechnol.*, Vol. 12(2), pp. 159–172.

20. Flicker, http://www-lecb.ncifcrf.gov/flicker

21. Micromass, http://www.micromass.co.uk/biotech1c.htm

22. UCSF, http://prospector.ucsf.edu

23. http://www.expasy.ch/tools/

24. Celis, J. E. *et al*. (1998), 'Human and mouse proteomic databases: novel resources in the protein universe', *FEBS Lett*, Vol. 430(1–2), pp. 64–72.

25. Jungblut, P. R. *et al*. (1999), 'Proteomics in human disease: cancer, heart and infectious diseases', *Electrophoresis*, Vol. 20(10), pp. 2100–2110.

26. Thiellement, H. *et al*. (1999), 'Proteomics for genetic and physiological studies in plants', *Electrophoresis*, Vol. 20(10), pp. 2013–2026.

27. Rabilloud, T. *et al*. (1999), 'Analysis of membrane proteins by two-dimensional electrophoresis: comparison of the proteins extracted from normal or Plasmodium falciparum-infected erythrocyte ghosts', *Electrophoresis*, Vol. 20(18), pp. 3603–3610.

28. Wilkins, M. R. *et al*. (1999), 'High-throughput mass spectrometric discovery of protein post-translational modifications', *J. Mol. Bio.*, Vol. 289(3), pp. 645–657.

29. Celis, J. E. and Gromov, P. (1999), '2D protein electrophoresis: can it be perfected?', *Current Opinion Biotechnology*, Vol. 10(1), pp. 16–21.

30. Walsh, B. J., Molloy, M. P. and Williams, K.L. (1998), 'The Australian Proteome Analysis Facility (APAF): assembling large scale proteomics through integration and automation', *Electrophoresis*, Vol. 19(11), pp. 1883–1890.

31. http://www.cgen.com/science/proteomics.htm

32. Bork, P. *et al*. (1998), 'Predicting function: from genes to genomes and back', *J. Mol. Biol.*, Vol. 283(4), pp. 707–725.

33. Sanchez, C. *et al*. (1999), 'Grasping at molecular interactions and genetic networks in *Drosophila melanogaster* using FlyNets, an Internet database', *Nucleic Acids Res.*, Vol. 27(1), pp. 89–94.

34. Rain, J. C. *et al*. (2000), 'Proteome-wide protein interaction map of the bacterial pathogen *Helicobacter pylori*', submitted.

35. Ito, T. *et al*. (2000), 'Toward a protein-protein interaction map of the budding yeast: A comprehensive system to examine two-hybrid interactions in all possible combinations between the yeast proteins [In Process Citation]', *Proc. Natl. Acad. Sci. USA*, Vol. 97(3), pp. 1143–1147.

36. Walhout, A. J. *et al*. (2000), 'Protein interaction mapping in *C. elegans* using proteins involved in vulval development [see comments]', *Science*, Vol. 287(5450), pp. 116–122.

37. Uetz, P. *et al*. (2000), 'A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*', *Nature*, Vol. 403, pp. 623–627.

38. Marcotte, E. M. *et al*. (1999), 'Detecting protein function and protein-protein interactions from genome sequences', *Science*, Vol. 285(5428), pp. 751–753.

39. Enright, A. J. *et al*. (1999), 'Protein interaction maps for complete genomes based on gene fusion events', *Nature*, Vol. 402(6757), pp. 86–90.

40. Blaschke, C. *et al*. (1999), 'Automatic extraction of biological information from scientific text: protein–protein interactions in 'Proc. 7th International Conference on Intelligent Systems for Molecular Biology', AAAI Press, Menlo Park, CA, pp. 60–67.

41. Marcotte, E. M. *et al*. (1999), 'A combined algorithm for genome-wide prediction of protein function [see comments]', *Nature*, Vol. 402(6757), pp. 83–86.

42. Xenarios, I. *et al*. (2000), 'DIP: the database of interacting proteins', *Nucleic Acids Res.*, Vol. 28(1), pp. 289–291.

43. Database of Interacting Proteins, http://dip.doe-mbi.ucla.edu/

44. MIPS, http://www.mips.biochem.mpg.de/

45. FlyNets, http://gifts.univ-mrs.fr/FlyNets/FlyNets_home_page.html

46. Eilbeck, K. *et al*. (1999), 'INTERACT: an object-oriented protein–protein interaction database, 'Proc. 7th International Conference of Intelligent Systems for Molecular Biology', AAAI Press, Menlo Park, CA, pp. 87–94.

47. Interact, http://bioinf.man.ac.uk/interactso.htm

48. PIMrider, Hybrigenics, http://pim.hybrigenics.com/

49. CuraGen, http://portal.curagen.com

50. ProNet, Myriad Genetics, http://pronet.doubletwist.com

51. Human Proteomics Initiative, http://www.expasy.ch/sprot/hpi

52. Cordwell, S. J. *et al.* (1999), 'The microbial proteome database – an automated laboratory catalogue for monitoring protein expression in bacteria', *Electrophoresis*, Vol. 20(18), pp. 3580–3588.

53. Wilkins, M. R. *et al.* (1997), 'Proteome database', in 'Proteome Research: New Frontiers in Functional Genomics', Wilikins, M.R. Ed. Springer-Verlag, Berlin Heidelberg. pp. 93–129.

54. Bairoch, A. and Apweiler, R. (2000), 'The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000', *Nucleic Acids Res.*, Vol. 28(1), pp. 45–48.

55. Sussman, J. L. *et al.* (1999), 'The protein data bank. Bridging the gap between the sequence and 3D structure world', *Genetica*, Vol. 106(1–2), pp. 149–158.

56. Hofmann, K. *et al.* (1999), 'The PROSITE database, its status in 1999', *Nucleic Acids Res.*, Vol. 27(1) pp. 215–219.

57. Bateman, A. *et al.* (2000), 'The Pfam protein families database', *Nucleic Acids Res.*, Vol. 28(1), pp. 263–266.

58. Attwood, T. K. *et al.* (1999), 'PRINTS prepares for the new millennium', *Nucleic Acids Res.*, Vol. 27(1), pp. 220–225.

59. Corpet, F., Gouzy, J. and Kahn, D. (1999), 'Recent improvements of the ProDom database of protein domain families', *Nucleic Acids Res.*, Vol. 27(1), pp. 263–267.

60. Interpro, http://www.ebi.ac.uk/interpro/

61. Legrain, P., Jestin, J.-L. and Schächter, V. (2000), 'From the analysis of protein complexes to proteome-wide linkage maps', *Current Opinions Biotechno*.

62. VanBogelen, R. A. *et al.* (1999), 'Diagnosis of cellular states of microbial organisms using proteomics', *Electrophoresis*, Vol. 20(11), pp. 2149–2159.

63. Klose, J. (1999), 'Genotypes and phenotypes', *Electrophoresis*, Vol. 20(4–5), pp. 643–652.

64. Williams, K. L. (1999), 'Genomes and proteomes: towards a multidimensional view of biology', *Electrophoresis*, Vol. 20 (4–5), pp. 678–688.