

## **A Streamlined Approach to High-Throughput Proteomics**

Andrew N. Stephens<sup>‡</sup>, Patricia Quach<sup>‡</sup> and Elizabeth J. Harry<sup>¶§</sup>

<sup>‡</sup>Department of Molecular and Microbial Biosciences  
University of Sydney, NSW 2006, Australia

<sup>¶</sup>Institute for the Biotechnology of Infectious Diseases  
University of Technology, Sydney, NSW 2007, Australia

<sup>§</sup>To whom correspondence should be addressed:

Ph. 61 2 9514 4173

Fax 61 2 9514 4201

Email: [Liz.Harry@uts.edu.au](mailto:Liz.Harry@uts.edu.au)

## **Summary**

Proteomics has rapidly become an important tool for life science research, allowing the integrated analysis of global protein expression from a single experiment. To accommodate the complexity and dynamic nature of any proteome, researchers must use a combination of disparate protein biochemistry techniques, often a highly involved and time-consuming process. Whilst highly sophisticated individual technologies for each step in studying a proteome are available, true high-throughput proteomics that provides a high degree of reproducibility and sensitivity has been difficult to achieve. The development of high throughput proteomic platforms, encompassing all aspects of proteome analysis and integrated with genomics and bioinformatics technology, therefore represents a crucial step for the advancement of proteomics research. ProteomIQ™ (Proteome Systems, Sydney, Australia) is the first fully integrated, start-to-finish proteomics platform to enter the market. Sample preparation and tracking, centralized data acquisition and instrument control, and direct interfacing with genomics and bioinformatics databases are combined into a single suite of integrated hardware and software tools, facilitating high reproducibility and rapid turnaround times. This review will highlight some features of ProteomIQ, with particular emphasis on the analysis of proteins separated by two-dimensional polyacrylamide gel electrophoresis.

**Running Title**

A Streamlined Approach to High-Throughput Proteomics

**Key words**

bioinformatics; gel electrophoresis; isoelectric focusing; laboratory information management system; mass spectrometry; Proteome Systems; proteomics

## Introduction

In the wake of large scale genome sequencing projects, life science research has experienced a paradigm shift from the reductionist, single-protein biochemical approach to a more global view in the analysis of biological systems. Accordingly the field of proteomics, the study of the entire protein complement expressed by the genome [1], has experienced explosive growth and has become a revolutionary tool with promising applications across the biological sciences. Proteomics involves the systematic study of the identity, distribution, abundance and modifications of large sets of proteins from biological samples such as cells, tissues or fluids. Proteins are typically extracted, separated and analyzed using a diverse array of protein biochemistry technologies. By comparing the proteome of samples derived from normal and disease states or from differing metabolic conditions, for example, phenotypic or functional changes can be correlated with their underlying biochemical or pathological events. A major advantage of this global approach is that many changes can be considered in an integrated manner from a single experiment. Not surprisingly, there has been intense interest in the application of proteomics to foster better understanding of disease, identify new biomarkers to improve diagnosis, detection and prognosis, and to accelerate the development of new and existing therapeutics [2-5].

The proteome of any organism is highly variable and dynamic in nature, since protein expression and modification respond to both internal and external stimuli and are dependent on cell type, growth conditions, environmental influences or disease status. The analysis of an individual's proteome therefore generates large amounts of information, leading to the requirement for a highly coordinated approach to data collection, storage and analysis. Initiatives such as the Human Proteome Organization's (HUPO) Proteomics Standards Initiative (PSI) (<http://psidev.sourceforge.net/>) and the Seattle Proteome Center's Institute for Systems Biology (ISB) (<http://www.proteomecenter.org/>) are good examples of the effort to “develop, apply and disseminate” new technologies aimed at standardization of the approaches used to generate proteomic data. The often high throughput nature of proteomic studies and the large datasets generated from such analyses also requires that rigorous comparison across statistically significant numbers of samples can be performed with a high degree of

reproducibility and sensitivity. Automation of as many operations as possible during the generation of a proteome can minimize variability and significantly enhance the consistency and reproducibility of the data generated. Companies such as Amersham Biosciences [101] and BioRad [102] now market purpose-built, automated proteomics equipment for processes such as IEF, gel electrophoresis or spot cutting.

Despite technological advances, equipment for proteomic analyses is typically stand-alone and lacks integration- a situation which can contribute significantly to the complexity of results and bottlenecks in downstream analysis. The development of high throughput proteomics technology platforms, encompassing all aspects of proteome analysis and integrated with genomics and bioinformatics technology, represents a crucial direction for the advancement of proteomics research. The major challenge for any proteomics platform is to establish technologies which provide reliable, efficient and reproducible data acquisition and analysis procedures that are amenable to high throughput. ProteomIQ<sup>TM</sup> (Proteome Systems, Sydney, Australia) [103] is a suite of proteomics technologies comprising instruments, consumables and software, integrated through an information management and analysis system, and was the first fully integrated proteomics platform to enter the market [6]. Since two-dimensional polyacrylamide gel electrophoresis (2DE) remains the most commonly used technology for proteomics, we will focus on the use of ProteomIQ in the generation of proteomic data from 2DE arrays.

### **Dealing with the diversity of the proteome**

The highly variable and dynamic nature of protein expression and modification presents a number of unique challenges to the proteomics researcher, since any particular organism will have many permutations of its proteome dependent on cell type, growth conditions, environmental influences or disease status- all of which may influence the protein expression profile of a cell at any particular time. Whilst knowledge of the complete genomic content of humans and a number of pathogenic and biologically relevant organisms has become part of an indispensable platform for mining proteomic data, the DNA expression profile of a static genome does not reflect the additional dynamics of transcriptional and translational controls, and post-translational

modifications of proteins. For example, sequencing of the human genome has revealed between 20000 to 25000 protein-coding genes[7, 8], not significantly different to the 25498 protein-coding genes found in organisms such as *Arabidopsis thaliana* [9]. The additional complexity of the proteome can arise through splice and translation isoforms, and post-translational processing and modifications such as glycosylation, phosphorylation or acetylation [10-13]– in some cases, up to 1000 potential species of a single protein, based on transcriptional isoforms, have been described [14]. Taking such modifications into account, the estimated total number of protein species will therefore be orders of magnitude higher than the number of genes observed in any individual genome [15]. The occurrence and significance of such modifications, induced by intra- or extra-cellular stimuli or environmental responses, all contribute to the biological role of proteins *in vivo* and can neither be predicted nor understood from the genome. Proteins also display a wide dynamic range of expression levels; for example, 5-6 orders of magnitude in yeast [16], and by at least 12 orders of magnitude in human blood serum [17, 18]. Every sample therefore has the potential to produce large numbers of complex proteomic permutations, characteristic of specific cell types, growth conditions or disease states.

To accommodate the complexity of the expressed proteome, a typical 2DE proteomics workflow encompasses a number of sequential steps (fig 1). The first of these involves the isolation of proteins from a sample using a suitable cocktail of detergents and denaturants. This must be done under conditions which are amenable to subsequent manipulations in the proteomics process. In even the simplest proteome, proteins often differ in abundance, size, pI and solubility, and thus sample fractionation prior to further analysis is often undertaken at this point to reduce sample complexity. Next, the isolated proteins are separated (generally in triplicate) by 2DE and the gels typically Coomassie blue, silver or SyproRuby<sup>TM</sup> (Molecular Probes) [104] stained. Image acquisition and analysis of the stained polyacrylamide gels is performed to identify differences in the profile of expressed proteins, and those identified as potentially interesting are flagged for further analysis. These proteins are then excised from the polyacrylamide gel, enzymatically digested and the purified peptides submitted for MS analysis. The identity

of the protein/s analyzed may then be determined by peptide mass fingerprinting (PMF) or peptide fragmentation fingerprinting (PFF).

### ***Sample preparation and pre-fractionation***

Of primary importance in any proteomic investigation is the preparation of samples in an appropriate format for presentation to 2DE. The protein composition of any sample can be extremely varied- relative protein abundances, charge, size, pI or solubility, for example- and successful preparation requires an intimate knowledge of the sample and a clearly defined strategy to target proteins of interest. Typically, a sample is collected and often subjected to an “enrichment” stage prior to protein isolation- for example, the sub-cellular isolation of organelles or plasma membranes, or membrane stripping protocols designed to remove peripheral proteins [19]. Care must be taken to ensure consistency during the collection and enrichment phase, prior to entering the proteomics factory, as elements such as sample collection methods or storage may affect the quality and protein composition of the sample [20]. Proteins are then solubilized, denatured, reduced and alkylated in an appropriate cocktail of nonionic or zwitterionic surfactants, chaotropes and salts that are compatible with first dimension isoelectric focusing (IEF) [21]. The composition of the cocktail depends on the proteins being targeted- for example, membrane or cytoplasmic proteins- and no one solution is ideal for all samples [21]. Standardized methodologies for particular sample types are, therefore, extremely important to ensure reproducibility and consistency when preparing samples for 2DE. A number of commercial kits, specifically tailored to particular sample types and protein sub-classes, are available for optimized sample preparation. Companies including BioRad [102] (ReadyPrep<sup>TM</sup>), Pierce [105] and Sigma [106] all market kits for protein isolation and fractionation. Consumables in the ProteomIQ platform (table 1) include extraction kits for the preparation of general or membrane proteins, and also for the depletion of albumin from serum and plasma.

Pre-fractionation prior to IEF is an important part of sample preparation, reducing sample complexity and increasing the protein load which may be separated on narrow-pH gradient gels by IEF [22]. This theoretically allows increased resolution, enrichment of low abundance proteins, specific targeting of proteins with particular pI values or

elimination of proteins from a sample (e.g. serum albumin, which accounts for ~60% of protein present in human plasma) and also allows tailoring of isoelectric fractions to specifically match narrow range IEF strips [23]. A number of preparative pre-fractionation devices, their uses and limitations have been recently reviewed [23, 24]. ProteomIQ includes a multi-compartment electrolyzer (MCE) which fractionates proteins according to their isoelectric points (fig 2a). A series of liquid-filled chambers are separated by semi-permeable membranes, each of which carries an immobilon-containing gel of a particular pH value [22]. When a potential difference is applied across the MCE, a discontinuous pI gradient is established where the pI of each membrane delimits the pI range of proteins focusing between them- in effect, a series of isoelectric “traps”. Generally, three fractions are produced (alkaline, neutral and acidic) although any combination of membranes may be used. This allows targeting of particular sub-proteomic fractions such as alkaline proteins, which are often under-represented in proteomic analyses, and the efficacy of isoelectric prefractionation in 2DE studies has been demonstrated for proteins from *Escherichia coli* and human plasma [22], and also for low abundance and membrane proteins from the yeast *Saccharomyces cerevisiae* [25]. A similar strategy is employed by the BioRad Rotofor [102], which relies on the use of ampholytes rather than membranes during the isoelectric fractionation. This approach has the significant advantage that protein losses due to precipitation onto the isoelectric membranes do not occur. Unlike Rotofor, ProteomIQ’s MCE is modular and designed specifically for use in a bench-top electrophoresis device (IsoelectrIQ or ElectrophoretIQ- see below), requiring no external power supply or vacuum. Using MCE to fractionate samples prior to IEF, the authors have had great success in displaying proteins from traditionally “difficult” samples [26, 27] such as the alkaline pH 6.5-11.0 fraction of the gram positive bacterium *Bacillus subtilis* (fig 2b).

### ***Protein Separation and Image Analysis***

The most commonly used technique for protein separation remains 2DE, probably because of its economy, accessibility to most laboratories and a wealth of previous experience to draw upon [5]. In the first dimension, proteins are focused in an IPG strip. Commercially available pre-cast IPG’s are highly reliable and offer a wide range of broad



or narrow pH gradients. Several companies market pre-cast IPG's and focusing hardware. Amersham Biosciences [101] and BioRad [102] market the Ettan IPGphor and Protean IEF Cell, respectively, whilst ProteomIQ uses the IsoelectrIQ (fig 3a). In each case, the instruments are equipped with cooled platforms, on-board protocols and variable voltage, current and temperature parameters. IsoelectrIQ has greater throughput (up to 36 IEF strips focused at once, via 3 independent power supplies, versus 12 or 24 respectively) and also offers preparative pre-fractionation via MCE.

Following IEF the proteins are separated in the second dimension by sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE). Second dimension polyacrylamide gels, when combined with first dimension IEF, still offer a number of advantages over non-gel systems in that they are able to separate, to homogeneity, hundreds or even thousands of proteins in a single step, can resolve intact protein isoforms, and are not overtly biased against any particular type of proteins (although, specialized gels and buffer systems might be required to deal with particular classes of proteins- e.g. acidic, basic, or hydrophobic proteins) [5, 28]. Resolution can also be varied according to the size of the gel, and so-called “zoom” gels, where IEF is performed in narrow pH range IPG's, may be used to achieve higher proteins loads in discrete pH ranges [5]. The major requirement for this step is that the polyacrylamide gels are prepared with a high level of consistency. Pre-cast polyacrylamide gels that give highly reproducible results are now available (e.g. Biorad [102], Gradipore [107]). BioRad's Mini PROTEAN minigels can be completed in around 40 minutes, whilst the large format (26x20cm) gels used in the Ettan DALT gel system take around 6 hours. ProteomIQ uses a pre-cast, tris-acetate GelChip<sup>TM</sup> mini-gel format (10x15cm) designed for use in the ElectrophoretIQ<sup>TM</sup> (fig 3b), with run times typically around 90 minutes for 8 gels. ElectrophoretIQ also is able to perform isoelectric focusing, MCE or Western blotting; as for IsoelectrIQ, any combination of equipment modules (fig 3c) may be run simultaneously from the separately controlled power supplies.

After separation of proteins by 2DE, the polyacrylamide gel is typically stained (although alternate labeling technologies, such as isotopic or fluorescent labeling, may be used- see [29] for a recent review) and *in silico* image analysis performed to identify differences in the pattern of protein expression. Image analysis remains as a bottleneck in

most proteomic analyses, and no single package has the capacity to perform image analysis without user intervention [13]. ProteomIQ uses an AlphaImager™ 3300 (Alpha Innotech) [108], to capture 12-bit images to a resolution of 40 microns using a cooled CCD camera; however, any generic image capture platform may be used- (for example, the BioRad VersaDoc 4000 has similar capabilities). High resolution images can also be captured using laser-based scanners such as the Amersham Typhoon [101]. These are particularly useful if double fluorescent staining, such as with the Cy-3 and Cy-3 dyes, are employed.

Following image capture, image analysis software performs spot detection, gel matching, spot quantitation and reporting for images derived from the scanned sets of gels. Numerous programs including PDQuest (BioRad) [102], Ettan LWS (Amersham Biosciences) [101], Melanie (GeneBio; also named ImageMaster 2D Platinum, from Amersham Biosciences and GeneBio) [109], Proteomweaver (Definiens) [110] or Progenesis (Nonlinear Dynamics) [111] are available for this. ProteomIQ interfaces directly, via application program interface (API), with either Progenesis [111] or ImageIQ™, an image analysis program developed in-house which performs analyses on all 2DE gels except those using DIGE [101] technology. First, filters are applied to parameters such as background subtraction and spot detection thresholds and spots are detected in the scanned image. Anchor points for gel matching are chosen manually, and a composite virtual gel with normalized spot volumes is created from the replicate image sets. The ImageIQ program includes vectoring and color overlays for observation of differentially expressed spots, and matching of composites can be performed in the same manner. Reporting includes integrated statistical analyses for each match. The performance of some of these image analysis packages have been recently compared and contrasted in two publications [30, 31].

### ***Protein digestion and peptide analysis***

Spot cutting, washing and digestion of protein spots from gels can be demanding, and are often a major bottleneck in proteomic investigations. Using robotic platforms for spot excision from gels and subsequent liquid delivery steps minimizes handling and avoids contamination or the introduction of inconsistencies. Typically, these platforms

consist of an image capture platform linked to spot cutting and liquid delivery systems. The major requirement here is for precise spot cutting to minimize contamination, along with complete automation and rapid turnaround times to facilitate high throughput. Several companies produce commercially available benchtop equipment covering one or more of the spot cutting, liquid delivery and MALDI target deposition processes. The ProteomeWorks Spot Cutter [102] cuts 200 spots/hour, using a digital image captured by an on-board 12-bit camera and illumination source (UV or visible light). Amersham Biosciences [106] markets the Ettan Spot Picker, Digester and Spotter, a series of instruments which excise protein spots from gels and follow an automated procedure from destaining through to deposition of samples onto a MALDI target plate. Markers on the gel backing are used to align the gel on the Spot Picker in accordance with a digital image (“blind cutting”), and spots are cut (96 spots/30 mins) and deposited into microtitre plate wells. These plates are transferred manually to the digester, where the gel plugs are washed, dried and digested. On completion of digestion, the plates are manually transferred to the Spotter and deposited to a MALDI target plate (384 spots/3 hours) for further analysis. At each step, “plug and play” routines are used to track the samples between spot cutting and deposition on the MALDI target.

ProteomIQ includes the completely automated Xcise robotic cutting platform (fig 4), for these processes. Images are acquired using an on-board 300dpi, 24 bit color flatbed scanner. Typically gels are Coomassie or silver-stained prior to scanning, although blind cutting may be performed in a manner similar to that used by the Ettan Spot Cutter (in which case, image acquisition at this stage is not required). Spot cutting (150 spots/hour), liquid delivery and protein digestion, and MALDI target deposition are performed with similar efficiency to the Amersham Biosciences instruments and similar sample tracking facilities. A major advantage of the Xcise, however, is that it is a single, enclosed, bench top instrument and no user intervention is required- this not only saves lab space and decreases throughput time, but also eliminates any potential problems which might arise during manual sample transfer between robotic instruments.

An interesting inclusion in the ProteomIQ platform is the ChIP<sup>TM</sup> (Chemical Inkjet Printer- fig 5a), the result of collaboration between Proteome Systems [103], Shimadzu Biotech [112] and MicroFab Technologies [113]. This instrument is designed

to process samples that have been transferred onto a PVDF membrane, essentially forming a protein array for each experiment. ChIP works on the principle that a 2DE gel, blotted onto a membrane, represents a unique protein macroarray [32]. Unlike protein microarrays, which are constructed from a finite set of known proteins [33], such a macroarray theoretically includes both known and unknown protein isoforms complete with co- and post-translational modifications. Protein analysis using the ChIP begins with immobilization of proteins onto a PVDF membrane (i.e. conventional blotting). Following image capture and selection of protein bands of interest, a piezoelectric microdispensing system is used to print nanolitre volumes of reagents (drop size 55µm) onto the membrane's surface (fig 5b and c). Dispensing is non-contact, allowing multiple dispensations onto a single protein spot without contamination. By dispensing minute amounts of primary and secondary antibodies, blocking solution and developing reagents precisely onto the protein band of interest (fig 5c), the ChIP has successfully been used to perform “nanoWestern” blotting. This strategy allows detection of multiple antigens from a single blot (fig 5d) or multiple chemistries to be dispensed onto a single protein spot. Dispensing reagents for MS analysis has also been achieved, allowing identification of proteins by MALDI-TOF directly from the membrane [32].

This approach offers some significant advantages. By bringing the chemistry to the protein, multiple liquid handling steps where protein can be lost are bypassed. The use of reagents such as antibodies, antigen, enzyme, MALDI matrix or membrane is also significantly reduced, resulting in lower costs for 2DE investigations. The reduced requirement for reagents is also important where samples are difficult to obtain or are in very limited supply (for example, clinical isolates). Using the ChIP, the detection of proteins by peptide mass fingerprinting (PMF) has been demonstrated in the 5-15 fmol range, and sensitivity of western blotting is comparable to conventional methods [32]. Second, liquid is dispensed onto only a portion of the protein spot to be analyzed, allowing multiple chemistries (and thus multiple experiments) to be performed on a single spot. This might include the use of multiple enzymes prior to MALDI-TOF analysis to increase peptide coverage by PMF, dispensation of multiple chemistries to examine protein activity or composition, or detection of multiple proteins from a single lane by western blotting. Also, because the proteins embedded on the PVDF membrane

are derived from a real sample, post-translational modifications can be characterized as has been demonstrated for the glycosylation of human plasma proteins [32]. Third, the PVDF membrane may be stored, allowing further testing of proteins from an experiment when new protocols or chemistries become available, new targets are identified or new ideas occur. The potential use of this technology as a diagnostic tool has already been demonstrated [32], and the ChIP promises to become an extremely important adjunct to current proteomics applications.

### ***Protein Identification, Informatics and Validation***

The final step in the 2DE proteomic workflow is the identification of protein spots by MS analysis. Whilst a number of different types of MS procedures are now available [34, 35], MALDI-TOF and ESI-TOF remain the two most commonly employed. MALDI-TOF is particularly well suited to proteomic analyses, as it is easily automatable and so is amenable to high throughput. It also has the advantages of high sensitivity and excellent resolution, and therefore is most commonly employed in high throughput proteomic analyses for protein identification [12]. The combination of MALDI-TOF and 2DE also offers some advantages over non-gel based proteomics approaches. Typically, LC-MS systems separate peptides in complex mixtures, providing identification of many more proteins than is possible from a 2DE system. However, identification is typically based on only one or a few peptides, and the intact protein is never observed [28]. This prevents additional protein coverage and makes identification of protein isoforms difficult [28]. The combination of MALDI-TOF and 2DE allows identification of tryptic peptides derived from the entire isolated protein, typically providing good coverage. Since the intact protein is the starting material it is possible to examine individual modifications on peptides, providing information on the biological function and status of the protein [28]. ProteomIQ is marketed with the Axima-CFR<sup>TM</sup> Plus MALDI-TOF MS from Shimadzu Biotech [112], which is capable of automated spectra acquisition and provides resolution at 60ppm. The Axima MALDI-TOF MS also performs post source decay (PSD), which can provide sequence information from peptides and increase the confidence of identifications obtained by PMF. ProteomIQ also supports and integrates mass spectrometers from other vendors, and a generic mass spectrometer interface

included in the platform allows information from other instruments, such as the Bruker Autoflex (Bruker) [114] or the Micromass Q-TOF (Waters Corporation) [115], to be captured and stored as part of a proteomics experiment.

Following collection of MS data the spectra must be processed and compared to theoretical peptide masses, generated from a genomic or protein database, in order to identify the gene encoding the isolated protein. ProteomIQ contains tools for the processing and analysis of MALDI-TOF MS data, as well as post-source decay (PSD) and tandem mass spectrometry (MS/MS), and includes all bridging software and algorithms necessary for peptides and protein identification. Monoisotopic mass spectra peaks are harvested automatically using the Peak Harvester tool, which first processes spectra and then applies Poisson modeling to determine which peak in an isotopically resolved group represents the monoisotopic peptide mass [36]. Following peak harvesting, reliable protein identification can depend on several factors including the quality of MS data produced, the number of masses or mass distribution of masses submitted for query- the key problem in identification remains that peptide masses are generally not unique, and random matching of a peptide to a database sequence may produce a false positive [37]. In ProteomIQ, the harvested peptide masses are submitted to the IonIQ search engine, which uses a modified MOWSE scoring system [38] to assign a measure of probability to protein identifications, for PMF. Queries are submitted to SWISS PROT, TrEMBL (Swiss-Prot, TrEMBL, and PIR database information is now contained in UniProt; <http://www.ebi.uniprot.org>) or proprietary databases, and potential protein identifications and probability scores are returned along with links to the annotated gel(s) and integrated MS data viewers. Where peptide identification is ambiguous, further peptide fragmentation by PSD or MS/MS may be undertaken and ProteomIQ also includes search engines for analysis of these data types. Comparisons between several different scoring algorithms, including MS-FIT and MASCOT (which use a MOWSE scoring system), ProFound and SEQUEST have been recently reviewed [37]. A significant improvement in the pipeline would be the integration of several search engines to be used simultaneously for automated protein identification from MS data; thus far, algorithms implementing this kind of search are unavailable.

Protein identification by MS analysis ultimately requires experimental validation, especially where proteins match the database records for “hypothetical proteins” or “putative function”. The nature of the validation process depends greatly on the kind of study undertaken, and the transition from putative protein identifications to specific biological or functional analyses is a limitation for many studies. The LIMS component of BioinformatIQ accommodates a number of sample preparation, electrophoretic or other handling procedures and allows uploading and association of third party data with any particular experiment. In this way, all subsequent validation information for the study may be stored in the central database.

### **Data Management, Integration and Quality Control**

The major challenge facing proteomics resides in integration of the complex processes and management of the diverse data types generated during analysis of a proteome. The Scierra™ Proteomics LWS and Ettan Spot Handling System [101], the ProteomeWorks™ System [102], ProteomIQ [103] and the Bruker ProteinScape [114] are platforms which integrate the processes of 2D gel generation, image analysis and protein identification. In any project of a size sufficient to require automation, tasks such as sample tracking, linking samples to gels, or linking protein spots to database entries rapidly become overwhelming and software tools to manage these processes are of paramount importance. BioinformatIQ™ is the web-based software platform that fully integrates and manages all hardware, data acquisition and analysis in ProteomIQ. The software is hosted on a scalable 64-bit IBM pSeries server or on Linux-based Intel servers, and functions both as a centralized laboratory information management system (LIMS) and a suite of data analysis modules. The LIMS component of BioinformatIQ is an IBM DB2 relational database, with a web-based interface accessible from any desktop PC connected to the network or across the internet. It enables remote monitoring and external control over experiments and allows all information concerning sample management, experimental methods and results to be captured and stored in the centralized database. The LIMS is scalable according to need, and accommodates both standard and customized experimental workflows for both 2DE and non-gel proteomics but does not have a worksheet generation capacity. Data capture from all hardware in

ProteomIQ is automatic; the system also supports data import from third party instruments (e.g. plate readers), but does not manage the hardware in this case. The Amersham Biosciences platform also provides limited third party support. A commercial platform allowing integration of either an alternate provider's or home-built tools, and data sharing in a common format, would be particularly advantageous, but none of the currently available platforms accommodate this degree of flexibility yet.

Large amounts of data are generated from even a simple proteomic workflow, and thus data management is of paramount importance. The Ettan LWS [101], ProteinScape [114] and BioinformatIQ offer complete sample tracking, from initial sample acquisition, through the protein separation process to ultimate data mining and linking of individual spots on a gel with protein identification. In each case, bar-coding or transponder-based technology is initially used to assign unique sample identities and to manage and track sample fate. Information concerning each sample, including the biological source, method of isolation, transport, storage and so-on may be entered into a database (an Oracle™ database in the case of Ettan LWS). Unique identification of samples in this manner greatly simplifies the management of complex processes and reduces the potential for downstream confusion of samples. In BioinformatIQ, the user enters experimental procedures as they are performed (using either pre-defined or customized methods); the workflow for each individual experiment is stored in the database and displayed as a hierarchical tree, with links to each procedure and the associated data collected (fig 6a and b). In this way, BioinformatIQ also functions as an electronic lab book that allows real-time data acquisition and accurate, rapid sharing of methods and data across a network or secure web connection. Multiple and complex queries can be processed by the LIMS, and various report types for all stages of the workflow generated.

Integration between hardware and software is a key requirement for any automated proteomics platform, facilitating accurate data capture and analysis and avoiding bottlenecks. A unique feature of BioinformatIQ is direct interfacing with all instruments in the platform: all hardware components of ProteomIQ provide run information directly to BioinformatIQ, and this is automatically uploaded into the central database and associated with the appropriate experiment. This information may also be used to generate the next step in the proteomic workflow. For example, following image



analysis by ImageIQ protein spots of interest can automatically be uploaded to Xcise for spot cutting. BioinformatIQ and Xcise also interface with Progenesis [111] via defined file formats; information from other image analysis packages may also be uploaded and stored, but is not integrated to the same degree. Liquid handling data generated by Xcise is automatically uploaded into BioinformatIQ, which then generates positional information correlating each protein spot processed with the subsequent position of eluted peptide on a MALDI target plate. This information then drives automated acquisition of PMF data- the only user intervention required throughout these processes is the manual transfer of the MALDI target plate from Xcise to the MALDI-TOF instrument.

The generation of large amounts of data requires careful quality control, and the issues of technical and biological variation in quantitative protein expression profiling were recently assessed [39]. Many sources of variation in quantitative 2DE analyses are possible- for example, biological variation in protein levels between experiments, error introduced during sample preparation, protein staining/labeling or gel running, or quantitation errors during image analysis and background subtraction [40]. Information capture is therefore of paramount importance to address the statistical aspects of variance in an experiment [40]. In the ProteomIQ platform, information concerning all aspects of the 2DE process- from sample preparation, through 2DE separation conditions and ultimately to MS analysis- is captured and stored in the LIMS. Integrated platforms such as ProteomIQ, ProteinScape or Ettan LWS also limit procedural inconsistencies by minimizing user intervention. Whilst no instrument is faultless, centralized data acquisition, storage and automated analyses facilitated by the software package controlling the platform all contribute to maintaining the quality of data output.

### ***Non-gel based proteomics***

Whilst this review is focused on the generation of 2DE arrays, each of the BioRad, Amersham Biosciences and Proteome Systems platforms also includes an automated LC-MS/MS platform for the separation and resolution of peptides in complex mixtures. This is useful for large-scale analyses of proteins present in the proteome, which exist in at least one form at any particular time. Automated LC-MS systems have

been widely used in “shotgun” approaches to proteomics (for review, see [12]), generally presenting peptides for ESI-MS/MS analysis. ProteomIQ uses a LCQ™ Deca XP Plus [116] with an on-board LC system, and is integrated with ProteomIQ via BioinformatIQ. The use of LC-MS/MS can enhance the ability to resolve protein modifications, or provide additional protein coverage and thus more accurate protein identification and analysis of post-translational modifications. Non-gel based proteomics therefore forms an important complement to any high-throughput 2DE platform.

## Conclusion

Proteomics has emerged as a powerful tool for the global analysis of biological systems, and has clear potential for accelerating the discovery and development of new diagnostic and therapeutic molecules [2-5]. ProteomIQ is the first high-throughput, purpose-designed proteomics platform that successfully integrates existing and emerging proteomics technologies. Several studies have already demonstrated the versatility and efficacy of ProteomIQ for integrated proteomic analyses. These include the identification of low abundance and membrane proteins from the yeast *S. cerevisiae* [25], identification of novel biomarkers from sputum and serum samples of patients suffering from cystic fibrosis or tuberculosis, and the identification of protein glycosylation modifications on mucins isolated from the sputum of CF patients, which have potential as novel diagnostic or therapeutic molecules [41]. The application of multiple chemistries to single protein spots using the ChIP has also been established, and has enormous potential for high-throughput proteomics and development of new diagnostic approaches [32]. The authors are currently using ProteomIQ in a proteomic analysis of the cell cycle in bacteria, aimed at identification of novel antibiotic targets and new therapeutic agents.

The development of high throughput platforms that are purpose-designed represents a major advancement in proteomics technology. Integrated proteomics platforms will become more common- Amersham Biosciences [101] and ThermoElectron Corporation [116] have recently launched a similar integrated platform (Ettan Spot Handling Workstation and Scierra™ 2D-MS LWS software) and BioRad [102] market the ProteomeWorks™ System [6]. These platforms offer high-throughput screening integrated with the genomics and bioinformatics capabilities necessary for identification

of biomarkers, drug targets and drug leads. Large lists of proteomic profiles for cataloguing are of limited value, and more targeted experimental approaches merging systems biology and hypothesis-driven research will become routine. The ability to perform iterative, efficient systematic profiling of differentially perturbed systems is vital for advancing proteomics research.

### **Expert Opinion and Five Year View**

Several key areas are expected to develop proteomics research further in the immediate future. These include improved detection systems for 2DE based proteomics, increased use of non-gel based proteomics, and further developments in informatics for handling and analyzing the huge datasets generated by proteomics researchers.

2DE remains the most commonly employed technique for proteomics, but as yet it has proven to be limited particularly for the display of low abundance and membrane proteins [42, 43]. These classes of proteins require specialized protocols and preparation methods, based on their particular compositions. We have recently had success in isolating and displaying membrane proteins from the gram positive bacterium *B. subtilis*, based around development of new sample buffer compositions and sample treatment protocols (A.N. Stephens, S.K. Pedersen, E.J. Harry, unpublished results). Whilst sample preparation methods continue to make incremental improvements in protein solubility, new detection systems which are compatible with current 2DE separation and detection technologies will provide enhanced resolution of the proteome and still be economically attainable for most researchers. Advances have been made in protein labeling technologies, including fluorescent dyes, chemical reagents and radioisotopes (for review, see [29]). For example, multi-labeling techniques can significantly reduce the time taken for image analysis of 2DE arrays, and also reduce the number of gels which must be produced. Affinity purification using anti-phosphoprotein antibodies and immobilized metal ion affinity chromatography has been used extensively for phosphoprotein enrichment (reviewed in [24]). New strategies for the chemical derivitisation of phosphate groups on serine or threonine residues to incorporate a thiol group [44] or biotin tag [45] as an affinity handle have also been reported, as has a new fluorescent reagent (Pro-Q Diamond phosphoprotein dye [104]) specific for

phosphoproteins and suitable for their detection after separation by 2DE [46, 47]. Enhanced detection of proteins using existing instrumentation would be an enormous asset to researchers, enabling deeper proteomic analyses without the cost associated with the purchase of more sensitive equipment. Quantitative protein detection, whether via labeling technology or quantitative MS techniques, could eliminate the need for a reference gel and represents a major potential advancement for future proteomics research.

Apart from new labeling technologies, increased sensitivity of MS instrumentation will also be important in providing a greater depth of proteome characterization. Accelerator MS, where atoms are accelerated through a high voltage field to strip away outer valency electrons and the resulting positive species analyzed [48], can provide protein sequence information at the attomole level [35] and would potentially be extremely useful for low abundance proteins. Of course, complementary sensitivity in detection and a capacity to process low abundance proteins without high losses will be essential for increased MS sensitivity to be effectively employed. In addition, current MS techniques are still not adequate for the analysis of complex post-translational modifications, although much progress has been made [35].

Non-gel systems are gaining popularity and show great promise in the separation and global analysis of proteins [12, 20]. Techniques such as multi-dimensional protein identification technology (MudPIT), where peptides are separated based on their charge and hydrophobicity prior to submission for ESI-MS/MS, theoretically allows large numbers of proteins to be identified [3]. Isotope-coded affinity tagging (ICAT) [49] allows quantitative expression profiling in an LC-MS/MS system by using isotopically coded tags to specifically label proteins. A number of other labeling chemistries have also been explored (reviewed in [12]). Whilst LC-MS/MS remains slow compared to MALDI-TOF MS [50] it is easily automated and recent work has demonstrated the comprehensive analysis of both soluble and membrane proteins [51, 52]. Non-gel proteomic systems will therefore provide an important complement to 2DE in the future.

Protein microarrays have begun to appear and have enormous potential as an adjunct to 2DE as protein identification, quantitation and interaction analysis tools [53]. Protein arrays are typically constructed from a known subset of proteins, such as

antibodies or recombinantly expressed proteins [33]. However, this approach ignores the many protein isoforms generated by protein modifications or variations in transcriptional control. In addition, the limiting factor for analytical microarrays is the availability of capture molecules with sufficient specificity and affinity- for example, an antibody array may recognize multiple variants of the same protein, masking potentially important changes at the level of protein modification. Nevertheless, microarray technology offers the ability to assay biological function such as ligand binding properties, phosphorylation or protein-protein interactions and promises, amongst other things, to form the basis of an important clinical diagnostic tool. Zhu *et al.* have cloned as GST fusions ~5800 ORF's from the yeast *S. cerevisiae* to produce a functional microarray [54], and antibody microarrays are now commercially available from BioChip Technologies [117]. The ability to produce unique proteins arrays for any individual, as provided by the ChIP [103], is a particularly powerful development which will assuredly have far-reaching applications in the future.

A major area of interest in proteomics is the post-translational modification of proteins. Recently, Proteome Systems [103] launched Glycosuite<sup>TM</sup>, a new database of glycan structures and protein modifications which interfaces directly with BioinformatIQ [55, 56]. This software has already been employed in the proteomic identification of glycoprotein targets with a view to developing new therapeutics to combat lung disease in patients suffering from cystic fibrosis [41]. Glycosuite has also been used to examine glycan structures present on human plasma glycoproteins [32, 57]. The ability to analyze glycosylation modifications on proteins provides a powerful addition to the integrated package of proteomics research tools.

Systems biology relies on the quality of the databases it exploits, and informatics packages such as BioinformatIQ, designed to handle large and diverse datasets and directly integrate them with genomics and bioinformatics databases, are a major step in the right direction. Currently, our ability to collect data far outweighs our ability to analyze it in a meaningful way. To provide further insights into the functional organization of complex biological systems, integration of centralized proteomics databases with validation data, activity profiles, interaction maps and protein modifications will become standard. It has been proposed that a score might be assigned

to proteins in such a database, as a measure of confidence for any protein interactions observed [12]. Of course, the complexity and context-dependent nature of each component in a biological network make this a daunting task which will be some time away. An interesting proposition is the development of modeling algorithms for the integration of such information [58]. These might be used to predict interacting components of pathways, which may not be expected by typical biochemical techniques or knowledge-based approaches. Such an integrated genomics and proteomics approach has been applied to the galactose metabolic network in yeast [59]. By assimilating data using similar algorithms, it should become possible to produce computational models for individual organisms, where cellular behaviors are predicted prior to being experimentally tested.

### **Acknowledgements**

This work was supported by the Australian Research Council. We are grateful to Marc Wilkins for insightful suggestions and critical evaluation of this manuscript, to Dr Michael Hsu and Ms Emma Richards for providing data from ChIP experiments and to Proteome Systems for the use of their ProteomIQ platform.

## **Key Issues**

- proteomics research generates very large and diverse datasets, which require a coordinated approach to data collection and analysis
- high throughput proteomics requires integration for storing, manipulating and mining large amounts of data, and automation for consistent sample handling on a large scale
- the application of new, integrated start-to-finish automated proteomics platforms such as ProteomIQ simplifies proteomic analyses and facilitates rapid and efficient data collection, analysis and archiving
- increased detection sensitivity, new adjuncts to 2DE technology and more integrated informatics approaches will continue to drive proteomics research and aid in the identification of new biomarkers for diagnostic and therapeutic applications

## References

Papers of special note have been highlighted as:

\* of interest

\*\* of special interest

1. Wilkins, MR, Sanchez, JC, Gooley, AA, *et al.* Progress with proteome projects: why all proteins expressed by a genome should be identified and how to do it. *Biotechnol Genet Eng Rev* 13, 19-50 (1996).
2. Hanash, S. Disease proteomics. *Nature* 422, 226-32 (2003).  
**\* A good review of the use of proteomics in identification and development of new biomarkers**
3. Walgren, JL and Thompson, DC. Application of proteomic technologies in the drug development process. *Toxicol Lett* 149, 377-85 (2004).  
**\* An excellent summary of the processes involved in the identification and development of new therapeutics**
4. He, QY and Chiu, JF. Proteomics in biomarker discovery and drug development. *J Cell Biochem* 89, 868-86 (2003).  
**\*\* An excellent review of the application of different proteomic strategies to the identification of new biomarkers and drug targets**
5. Marko-Varga, G and Fehniger, TE. Proteomics and disease--the challenges for technology and discovery. *J Proteome Res* 3, 167-78 (2004).
6. Gershon, D. Proteomics technologies: probing the proteome. *Nature* 424, 581-7 (2003).
7. Pennisi, E. Human genome. A low number wins the GeneSweep Pool. *Science* 300, 1484 (2003).
8. Stein, LD. Human genome: end of the beginning. *Nature* 431, 915-6 (2004).
9. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408, 796-815 (2000).
10. Mann, M and Jensen, ON. Proteomic analysis of post-translational modifications. *Nat Biotechnol* 21, 255-61 (2003).



11. Aebersold, R and Goodlett, DR. Mass spectrometry in proteomics. *Chem Rev* 101, 269-95 (2001).
12. Patterson, SD and Aebersold, RH. Proteomics: the first decade and beyond. *Nat Genet* 33 Suppl, 311-23 (2003).
- \* **Good general review of proteomics**
13. Harry, JL, Wilkins, MR, Herbert, BR, *et al.* Proteomics: capacity versus utility. *Electrophoresis* 21, 1071-81 (2000).
14. Ullrich, B, Ushkaryov, YA, and Sudhof, TC. Cartography of neurexins: more than 1000 isoforms generated by alternative splicing and expressed in distinct subsets of neurons. *Neuron* 14, 497-507 (1995).
15. Miklos, GL and Maleszka, R. Protein functions and biological contexts. *Proteomics* 1, 169-78 (2001).
16. Gygi, SP, Rochon, Y, Franza, BR, and Aebersold, R. Correlation between protein and mRNA abundance in yeast. *Mol Cell Biol* 19, 1720-30 (1999).
17. Lescuyer, P, Hochstrasser, DF, and Sanchez, JC. Comprehensive proteome analysis by chromatographic protein prefractionation. *Electrophoresis* 25, 1125-35 (2004).
18. Anderson, NL, Polanski, M, Pieper, R, *et al.* The human plasma proteome: a nonredundant list developed by combination of four separate sources. *Mol Cell Proteomics* 3, 311-26 (2004).
19. Molloy, MP, Herbert, BR, Slade, MB, *et al.* Proteomic analysis of the Escherichia coli outer membrane. *Eur J Biochem* 267, 2871-81 (2000).
20. Ryan, TE and Patterson, SD. Proteomics: drug target discovery on an industrial scale. *Trends Biotechnol* 20, S45-51 (2002).
21. Herbert, B. Advances in protein solubilisation for two-dimensional electrophoresis. *Electrophoresis* 20, 660-3 (1999).
22. Herbert, B and Righetti, PG. A turning point in proteome analysis: sample prefractionation via multicompartiment electrolyzers with isoelectric membranes. *Electrophoresis* 21, 3639-48 (2000).
23. Righetti, PG, Castagna, A, Herbert, B, Reymond, F, and Rossier, JS. Prefractionation techniques in proteome analysis. *Proteomics* 3, 1397-407 (2003).

24. Stasyk, T and Huber, LA. Zooming in: fractionation strategies in proteomics. *Proteomics* 4, 3704-16 (2004).
25. Pedersen, SK, Harry, JL, Sebastian, L, *et al.* Unseen proteome: mining below the tip of the iceberg to find low abundance and membrane proteins. *J Proteome Res* 2, 303-11 (2003).
- \* Demonstration of the use of isoelectric sample pre-fractionation to reduce sample complexity and concentrate low abundance and membrane proteins**
26. Eymann, C, Dreisbach, A, Albrecht, D, *et al.* A comprehensive proteome map of growing *Bacillus subtilis* cells. *Proteomics* 4, 2849-76 (2004).
27. Ohlmeier, S, Scharf, C, and Hecker, M. Alkaline proteins of *Bacillus subtilis*: first steps towards a two-dimensional alkaline master gel. *Electrophoresis* 21, 3701-9 (2000).
28. Herbert, BR, Harry, JL, Packer, NH, *et al.* What place for polyacrylamide in proteomics? *Trends Biotechnol* 19, S3-9 (2001).
29. Hu, Y, Huang, X, Chen, GY, and Yao, SQ. Recent advances in gel-based proteome profiling techniques. *Mol Biotechnol* 28, 63-76 (2004).
- \*\* An excellent review of existing and emerging labeling technologies for gel-based proteomics**
30. Raman, B, Cheung, A, and Marten, MR. Quantitative comparison and evaluation of two commercially available, two-dimensional electrophoresis image analysis software packages, Z3 and Melanie. *Electrophoresis* 23, 2194-202 (2002).
31. Rosengren, AT, Salmi, JM, Aittokallio, T, *et al.* Comparison of PDQuest and Progenesis software packages in the analysis of two-dimensional electrophoresis gels. *Proteomics* 3, 1936-46 (2003).
32. Sloane, AJ, Duff, JL, Wilson, NL, *et al.* High throughput peptide mass fingerprinting and protein macroarray analysis using chemical printing strategies. *Mol Cell Proteomics* 1, 490-9 (2002).
- \*\* Demonstration of novel chemical printing technology in the generation and analysis of individual protein macroarrays**
33. Cahill, DJ. Protein and antibody arrays and their medical applications. *J Immunol Methods* 250, 81-91 (2001).

34. Binz, PA, Hochstrasser, DF, and Appel, RD. Mass spectrometry-based proteomics: current status and potential use in clinical chemistry. *Clin Chem Lab Med* 41, 1540-51 (2003).
- \* Good review of current mass spectrometry technology and typical approaches used in proteomics**
35. Griffin, TJ and Aebersold, R. Advances in proteome analysis by mass spectrometry. *J Biol Chem* 276, 45497-500 (2001).
- \* Discussion of mass spectrometry techniques and their application to proteomics**
36. Breen, EJ, Hopwood, FG, Williams, KL, and Wilkins, MR. Automatic poisson peak harvesting for high throughput protein identification. *Electrophoresis* 21, 2243-51 (2000).
37. Chamrad, DC, Korting, G, Stuhler, K, *et al.* Evaluation of algorithms for protein identification from sequence databases using mass spectrometry data. *Proteomics* 4, 619-28 (2004).
38. Pappin, DJ, Hojrup, P, and Bleasby, AJ. Rapid identification of proteins by peptide-mass fingerprinting. *Curr Biol* 3, 327-32 (1993).
39. Molloy, MP, Brzezinski, EE, Hang, J, McDowell, MT, and VanBogelen, RA. Overcoming technical variation and biological variation in quantitative proteomics. *Proteomics* 3, 1912-9 (2003).
40. Gustafsson, JS, Ceasar, R, Glasbey, CA, Blomberg, A, and Rudemo, M. Statistical exploration of variation in quantitative two-dimensional gel electrophoresis data. *Proteomics* 4, 3791-9 (2004).
41. Harcourt, R, Cole, RA, Harry, EJ, Lindner, RA, Pedersen, SK, Prasad, SS, Sebastian, LT, Schulz, BL, Sloane, AJ, Harry, JL. Proteomics: the new paradigm for biomarker and drug target discovery. *J Med Tech (Japan)* 47, 1429-1436 (2003).
42. Gygi, SP, Rist, B, and Aebersold, R. Measuring gene expression by quantitative proteome analysis. *Curr Opin Biotechnol* 11, 396-401 (2000).

43. Gygi, SP, Corthals, GL, Zhang, Y, Rochon, Y, and Aebersold, R. Evaluation of two-dimensional gel electrophoresis-based proteome analysis technology. *Proc Natl Acad Sci U S A* 97, 9390-5 (2000).
  44. McLachlin, DT and Chait, BT. Improved beta-elimination-based affinity purification strategy for enrichment of phosphopeptides. *Anal Chem* 75, 6826-36 (2003).
  45. Oda, Y, Nagasu, T, and Chait, BT. Enrichment analysis of phosphorylated proteins as a tool for probing the phosphoproteome. *Nat Biotechnol* 19, 379-82 (2001).
  46. Steinberg, TH, Agnew, BJ, Gee, KR, *et al.* Global quantitative phosphoprotein analysis using Multiplexed Proteomics technology. *Proteomics* 3, 1128-44 (2003).
  47. Martin, K, Steinberg, TH, Cooley, LA, *et al.* Quantitative analysis of protein phosphorylation status and protein kinase activity on microarrays using a novel fluorescent phosphorylation sensor dye. *Proteomics* 3, 1244-55 (2003).
  48. Garner, RC. Accelerator mass spectrometry in pharmaceutical research and development--a new ultrasensitive analytical method for isotope measurement. *Curr Drug Metab* 1, 205-13 (2000).
  49. Gygi, SP, Rist, B, Gerber, SA, *et al.* Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat Biotechnol* 17, 994-9 (1999).
  50. Hancock, WS, Wu, SL, Stanley, RR, and Gombocz, EA. Publishing large proteome datasets: scientific policy meets emerging technologies. *Trends Biotechnol* 20, S39-44 (2002).
  51. Wu, CC, MacCoss, MJ, Howell, KE, and Yates, JR, 3rd. A method for the comprehensive proteomic analysis of membrane proteins. *Nat Biotechnol* 21, 532-8 (2003).
- \*\*an elegant demonstration of the use of MudPIT technology to analyze membrane proteins in a complex solution**
52. Wu, CC and Yates, JR, 3rd. The application of mass spectrometry to membrane proteomics. *Nat Biotechnol* 21, 262-7 (2003).

53. Templin, MF, Stoll, D, Schwenk, JM, *et al.* Protein microarrays: promising tools for proteomic research. *Proteomics* 3, 2155-66 (2003).
54. Zhu, H, Bilgin, M, Bangham, R, *et al.* Global analysis of protein activities using proteome chips. *Science* 293, 2101-5 (2001).
55. Cooper, CA, Joshi, HJ, Harrison, MJ, Wilkins, MR, and Packer, NH. GlycoSuiteDB: a curated relational database of glycoprotein glycan structures and their biological sources. 2003 update. *Nucleic Acids Res* 31, 511-3 (2003).  
**\*\* A searchable database of glycoprotein glycan structures**
56. Cooper, CA, Harrison, MJ, Wilkins, MR, and Packer, NH. GlycoSuiteDB: a new curated relational database of glycoprotein glycan structures and their biological sources. *Nucleic Acids Res* 29, 332-5 (2001).
57. Wilson, NL, Schulz, BL, Karlsson, NG, and Packer, NH. Sequential analysis of N- and O-linked glycosylation of 2D-PAGE separated glycoproteins. *J Proteome Res* 1, 521-9 (2002).
58. Papin, J and Subramaniam, S. Bioinformatics and cellular signaling. *Curr Opin Biotechnol* 15, 78-81 (2004).
59. Ideker, T, Thorsson, V, Ranish, JA, *et al.* Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. *Science* 292, 929-34 (2001).

## Websites

101. Amersham Biosciences  
<http://www.Amersham Biosciences.com>
102. BioRad  
<http://www.biorad.com>
103. Proteome Systems  
<http://www.proteomesystems.com>
104. Molecular Probes  
<http://www.probes.com>
105. Pierce  
<http://www.pierce.com>
106. Sigma  
<http://www.Sigma.com>
107. Gradipore  
<http://www.gradipore.com>
108. Alpha Innoetch  
<http://www.alphainnotech.com>
109. GeneBio  
<http://www.genebio.com>
110. Definiens  
<http://www.definiens.com>
111. Nonlinear Dynamics  
<http://www.nonlinear.com>
112. Shimadzu Biotech  
<http://www.shimadzu-biotech.net>
113. MicroFab Technologies  
<http://www.microfab.com>
114. Bruker  
<http://www.bruker.com>
115. Waters Corporation

- <http://www.waters.com/LifeScience>
116. ThermoElectron Corporation  
<http://www.Thermo.com>
117. BioChip  
<http://www.biochipnet.de>

**Affiliations**

Dr Elizabeth J. Harry, PhD

Associate Professor

Institute for the Biotechnology of Infectious Diseases

University of Technology, Sydney, NSW 2007

Australia

Email [Liz.Harry@uts.edu.au](mailto:Liz.Harry@uts.edu.au)

Dr Andrew N. Stephens

Postdoctoral Fellow

Department of Molecular and Microbial Biosciences

University of Sydney, NSW 2006

Australia

Email [aste9574@usyd.edu.au](mailto:aste9574@usyd.edu.au)

Ms Patricia Quach

Department of Molecular and Microbial Biosciences

University of Sydney, NSW 2006

Australia

Email [pqua0758@mail.usyd.edu.au](mailto:pqua0758@mail.usyd.edu.au)



**TABLE 1. CONSUMABLES REQUIRED IN THE PROTEOMIQ PLATFORM.**

| PROCESS / CONSUMABLES   | DESCRIPTION/COMMENTS  |
|---|---|
| <b><i>SAMPLE PREPARATION</i></b>                                    |   |
| ProteoPrep Extraction Kits  | Includes detergents, solubilizing solutions, reducing and alkylating reagents. Specific kits for general proteins, membrane proteins or universal extraction      |
| Montage Albumin Deplete Kit   | Rapid removal of albumin from serum and plasma  |
| <b><i>PRE-FRACTIONATION</i></b>                                     |   |
| ProteomIQ MCE Kit   | Includes sample solubilising solutions for either general or low solubility proteins, and immobilized pH membranes used for preparative isoelectric fractionation |
| <b><i>ISOELECTRIC FOCUSING</i></b>                                  |   |
| Immobilized pH gradient strips<br>OrangeG tracking dye<br>IPG wicks | Pre-cast, broad and narrow range IPG strips and other consumables for IEF   |
| Sample rehydration/equilibration trays<br>IEF tray                  | Disposable or reusable  |
| <b><i>SDS-PAGE</i></b>  |   |
| ProteomIQ GelChip   | Tris-acetate 6-15% gradient small format (10x15cm) gels for use in ElectrophoretIQ  |
| ProteomIQ running buffer, tris/tricine/SDS<br>Equilibration buffer  | Standard buffers for gel running and IEF strip equilibration  |
| <b><i>WESTERN BLOTTING</i></b>                                      |   |
| ProteomIQ Blotting kit  | Reagents, blotting paper and PVDF membrane for western blotting   |
| <b><i>PROTEIN STAIN</i></b>   |   |
| ProteomIQ Blue Gel Stain kit  | Colloidal Coomassie gel stain, MS-compatible  |
| ProteomIQ Blot Stain Kit, Direct Blue 71                            | Protein stain for use with PVDF membrane  |
| <b><i>PROTEIN PROCESSING</i></b>                                    |   |
| ProteomIQ Xcise In-Gel Digest Kit                                   | Contains reagents, MTP plates, ZipTips™ and plastic solution trays designed for use in Xcise for spot cutting and peptide purification. Trypsin not supplied.     |

## Figure legends

**Fig 1. A typical proteomics workflow on the ProteomIQ platform.** Samples are prepared and pre-fractionated if desired, and the workflow entered into the LIMS component of BioinformatIQ (right). Proteins are then separated by 2DE and image analysis of the stained gels identifies protein spots of interest. These proteins are excised from the gel, digested and the peptides identified by MS analysis. At each step, information concerning the experiment is uploaded into BioinformatIQ, and appropriate information for the next step is downloaded (arrows) into robotic equipment for generation of automated protocols.

**Fig 2. Prefractionation of proteins into discrete pH ranges increases resolution in 2DE.** (a) the MCE device used for pre-fractionation of proteins. The multi-chamber design allows a sample to be fractionated into up to five pH increments, depending on requirements. (b) Proteins isolated from the gram positive bacterium *Bacillus subtilis* were fractionated using an MCE device and the pH 6.5-11 fraction arrayed by 2DE. First dimension focusing was in precast IPG strips covering the pH 3-10 range, and second dimension SDS-PAGE was in 6-15% gradient GelChips<sup>TM</sup>. Proteins were visualized by staining with silver.

**Fig 3. Hardware components of the ProteomIQ platform for 2DE.** (a) the IsoelectrIQ, displayed with one MCE and two different sizes of IEF modules on board; (b) the ElectrophoretIQ, displayed with two SDS-PAGE modules on board; (c) individual modules for use in the IsoelectrIQ or ElectrophoretIQ. From left; IEF module, SDS-PAGE module, western blotting module.

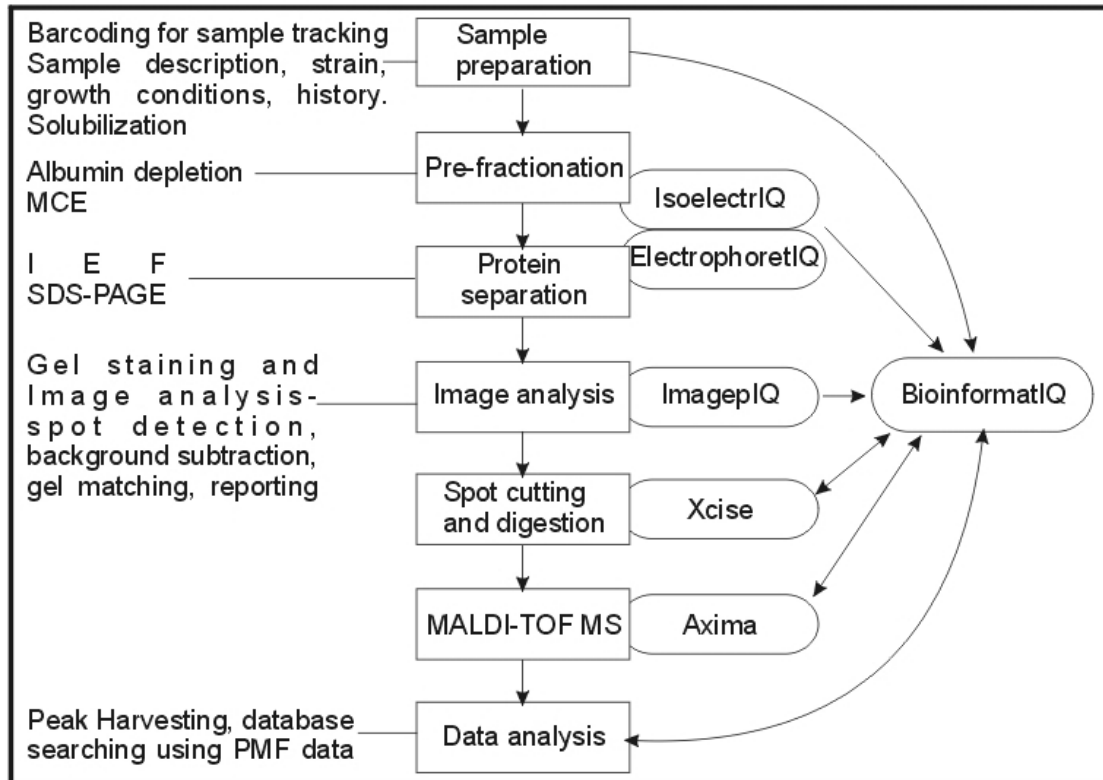
**Fig 4. The Xcise liquid handling robot and user interface.** Gels are placed into the enclosed gel platform, and the processes of spot cutting, gel destaining, protein digestion,

peptide purification and deposition onto a MALDI target plate are carried out automatically.

**Fig 5. The ChIP chemical inkjet printer.** (a) The ChIP. (b) A stained blot loaded under the print head of the ChIP. (c) Reagents are dispensed from a piezoelectric microjet over selected bands (represented as white lanes over the magnified bands of interest) onto the membrane surface. Primary and secondary antibody, blocking and washing solutions are all dispensed in this manner and the blot developed. (d) Comparison of conventional blotting versus ChIP nanowestern blotting. Cytoskeletal supernatant samples from mast cells were probed using antibodies as shown. Multiple experiments may be performed on a single blot or a single protein spot in this manner. From left; a stained protein blot from 1D SDS-PAGE gel; 3 conventional replicate western blots probed with anti-myosin, anti-vimentin or anti-actin antibodies; a single replicate western blot probed with all three antibodies using the ChIP.

**Fig 6. The BioinformatIQ web interface.** The tree structure of the BioinformatIQ user interface, showing (a) organization of a current project, and (b) details of an experiment in progress.

**Fig 1**

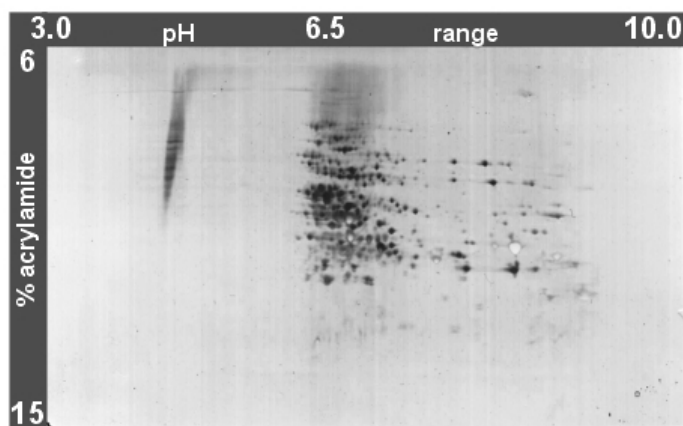


**Fig 2**

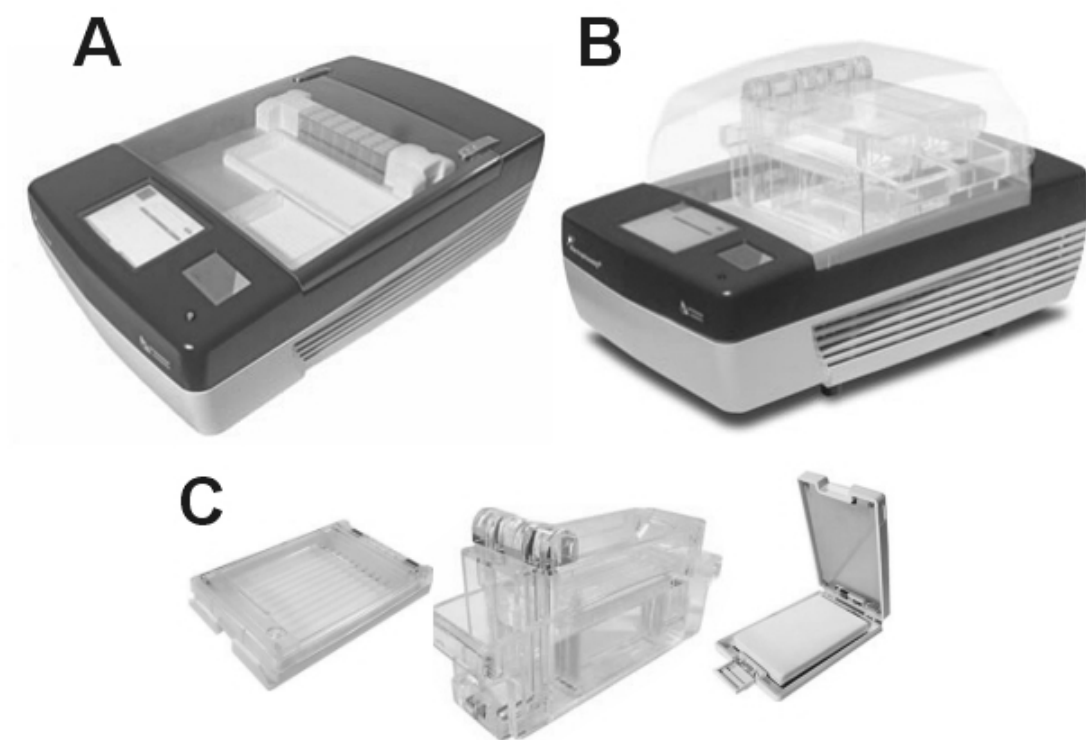
**A**



**B**



**Fig 3**



**Fig 4**



**Fig 5**

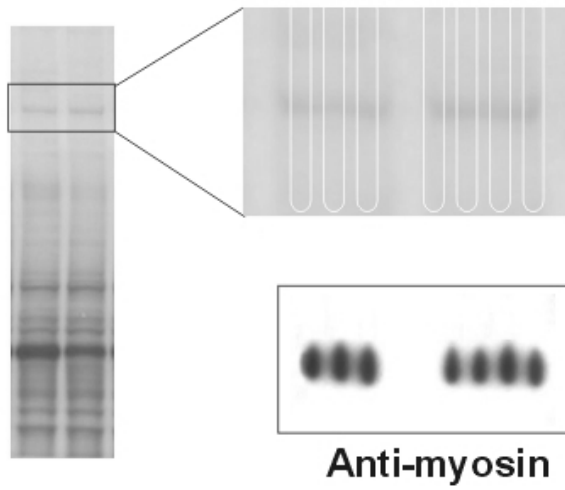
**A**



**B**



**C**





**D**

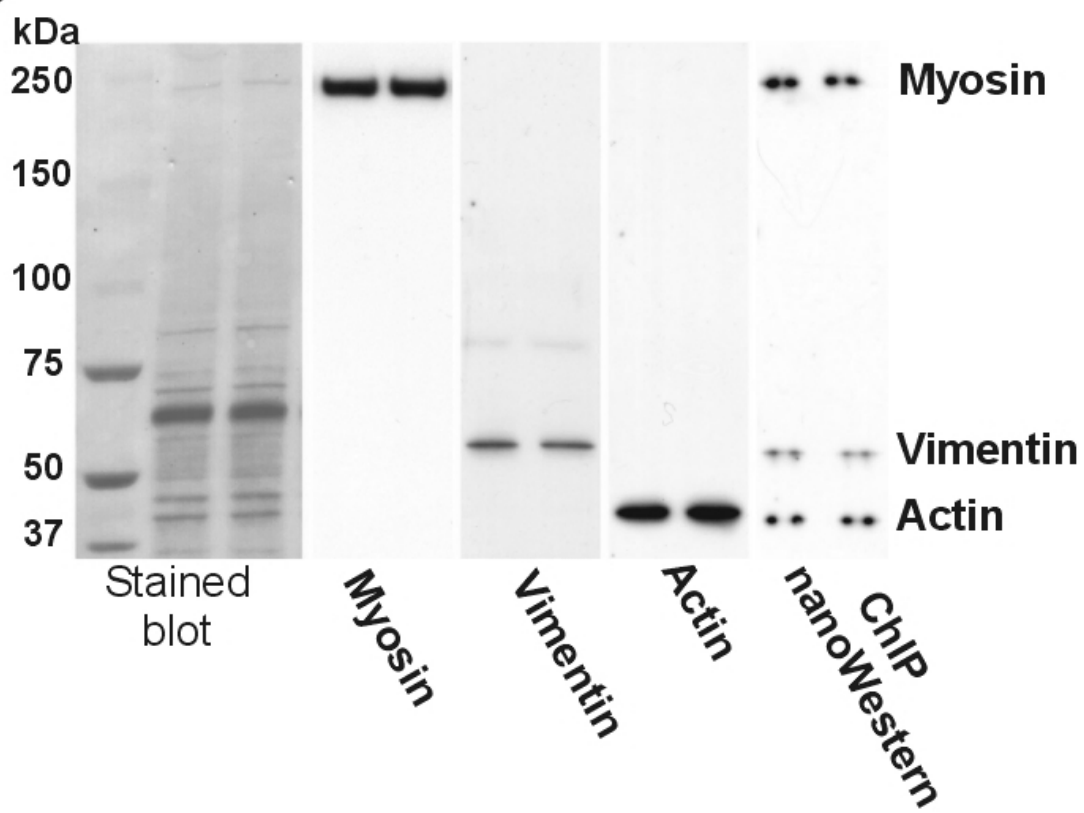
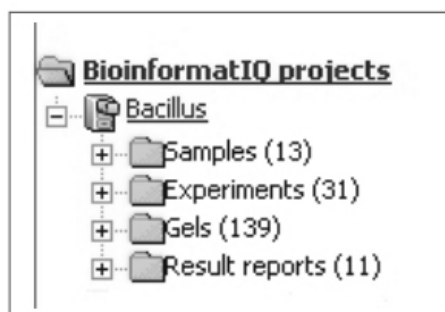


Fig 6

A



B

