Proteomics is a recent member of the ‘omics’ family that has gained rapid momentum at the turn of the century, particularly in the area of therapeutics. The proteomics repertoire of ‘tools’ promises a wide area of revolutionary hardware, processes, techniques and strategies that have potential application in various stages of the drug discovery process, from target identification and validation to lead optimization and, subsequently, to toxicological profiling in preclinical and clinical settings. Although there is no doubt that the successful completion of the human genome project in 2003 and the enormous genomic data being generated will help speed up the discovery of new drug targets, experts believe that information on the protein level is needed to bolster rapid drug development because proteins are ultimately the key ‘workhorses’ in our body. Further-more, comparative analysis of the information produced by genomic studies and information regarding protein expression has led to the conclusion that genetic messages often fail to correlate with protein abundance and, more importantly, protein function. The rule of ‘one gene–one polypeptide’ no longer holds true either because one gene can encode multiple forms of a protein owing to alternative splicing and post-translational modifications. In addition, the function of a protein is dictated by its three-dimensional (3D) structure, which may change when it interacts with another or with other components inside cellular tissues. Despite these intricacies, proteins are keys to unlocking the nanometer keyholes to understand biological functions and the way these functions vary between health and diseased states. The more keys we can find, the more doors we can open. In addition, most drug targets are proteins; therefore, it is important to focus drug discovery efforts at this level.

Because of the great excitement surrounding proteomics, there has been a noticeable emphasis for pharmaceutical and its biotechnological and academic partners to expend tremendous resources to refine and review their strategies to make drug development more efficient and successful. In recent years, businesses have also flocked to the thriving field of proteomics, with more than 100 companies now offering proteomics-related technologies, tools, data and services in the hope of gaining a competitive advantage and finding the best targets for successful drug candidates and diagnostic markers. According to the Cambridge Heathtech Institute’s (CHI) Drug Discovery and Development Deals Database (http://www.discoverydeals.com), the number of business deals clinched in proteomics saw an upward trend of 62%, with Incyte Genomics, Oxford GlycoSciences (OGS), Proteome Systems and Applied Biosystems making it to the top 10 dealmakers list in 2000–01. On top of that, proteomics companies have attracted more than $530 million in venture capital from 1999 to 2001 through a wide array of alliances and collaborations and through public offerings of stock.

**DRUG TARGET, EFFICACY, TOXICOLOGY AND BIOMARKER DISCOVERY FROM PROTEOMICS**

Proteomics has broad applications and it aims to: (i) catalogue and characterize the full protein complement in the genome; (ii) compare the levels of protein expression under different conditions (control vs pathological conditions); (iii) identify and characterize protein modifications, such as phosphorylation, acetylation, glycosylation etc.; (iv) identify protein localization and compartmentalization at a given time; and (v) understand how proteins interact with each other and with other biological molecules and chemicals in cellular tissues. However, these combined studies will not yield a...
comprehensive ‘proteome map’ of an organism because cells are in constant flux, and so is protein expression, in response to age, disease and trauma.

Proteomics is a promising approach in the identification of protein targets and the elucidation of fundamental cell and disease biology, biochemical processes and pathways involved in disease process and, thus, plays a significant role in drug development. Disease involves alterations in protein expression levels. Hence, by studying the change in the abundance of proteins within a cell over time or between different cellular states (normal vs diseased states), proteomics provide insights into the pathophysiological basis of protein target identification and validation for disease intervention and treatment. Specific biomarkers identified from proteomics may be used as protein signatures to screen new chemical entities for target organ toxicity in preclinical trials and later on in development during clinical trials, ensuring their usefulness in the diagnosis and prognosis of diseases. The use of proteomics to study toxicity at the protein level is called ‘pharmacoproteomics’. By holistically examining the entire protein profile of cellular tissues treated with drugs or drug candidates, proteomics establishes a comprehensive protein interaction map (‘interactomes’) related to disease pathways, hence optimizing the drug development process. Figure 1 summarizes some of the intervening stages in the drug discovery and development process that proteomics can play a major role in.

Today, considerable effort is being devoted to the development of new tools and methodologies, complementary to the existing strategies, for high-throughput analysis of the total protein repertoire of biological samples. These new, highly sensitive and high-throughput technologies are poised to unravel the mystery of proteins and add a new dimension to large-scale drug discovery. In this regard, the myriad of proteomic approaches currently on the market and areas of ongoing research could broadly be categorized into three basic categories: profiling approach, functional approach and structural approach.

Profiling approach

Traditional proteomics approaches have focused on profiling strategies, which sought to decipher the proteins that are differentially expressed in any given disease condition and/or after drug treatment. The profiling approach tends to provide information about the identity and, in some cases, abundance of proteins and it represents one of the primary approaches for target identification and validation. Examples of technologies that fall under the umbrella term of profiling/expression proteomics include protein separation methods, such as two-dimensional gel electrophoresis (2DGE), chromatographic techniques, isotope-coded affinity tag (ICAT) and protein chips/microarrays, as well as the use of protein identification methods via mass spectrometry (MS). In our laboratories, we have applied this approach to identify potential biomarkers that may be useful for disease classification and prognostication or as early predictive markers of drug responsiveness/resistance. One example of our work includes the identification of potential biomarkers of aggressive breast cancer by comparing the protein profiles of cancer biopsies from patients classified differentially based on their lymph node status and size (see Fig. 2). In many instances, we have applied several of the profiling technologies mentioned above together, namely 2DGE, tandem liquid chromatography (LC) followed by MS identification and analysis of peptides and protein chip profiling, to obtain a wider screen of potential markers because each profiling technology comes with its own advantages and limitations. The strengths and limits of
Fig. 2 Identifying the differences in protein profile in breast cancer biopsies between patients with cancers classified as having negative lymph node status and a size of less than 2 cm and those with a positive lymph node status, regional metastasis and cancer that is > 5 cm in size. This is done by comparing the two-dimensional gel electrophoresis protein profile of (a) biopsies from lymph node-negative status patients with (b) biopsies from patients showing regional metastasis (N Jiang et al., unpubl. data, 2005).

each technology are discussed further below. One should not view these as competing, but rather as complimentary, technologies, which would allow a deeper probe into different portions of the proteome.

Functional approach

The functional approach focuses on delineating protein–protein or protein–DNA/RNA interactions, enzymatic activity, downstream signalling pathways and post-translational modifications. Clearly, it would be of great advantage to be able to screen the function of protein, because it is now possible to assay the differential expression and distribution of proteins. The successful identification of a therapeutic target and its interacting drug or drug-like compound does not constitute an effective treatment for the diseases under investigation. Most proteins do not work in isolation and they are often involved in one or more cellular pathways. Thus, it is important to know how the multiple pathways that maintain the biology and physiology of the cell behave and vary in both normal and pathological states, as well as in control and treated conditions. Construction of a comprehensive cellular protein network map will facilitate the prediction of intervention points in a disease pathway and elicit new drug-associated parameters following drug treatment. Functional approaches are relatively diverse and the most established is the yeast two hybrid system, in addition to other proteomic strategies, such as phage display technology, chromophore-assisted laser inactivation and fluorophore-assisted laser inactivation and activity-based profiling.

Structural approach

One of the crucial components for the functional understanding of uncharacterized proteins is the elucidation of their 3D structure. Often, newly sequenced proteins share unrecognized structural and functional homology to known proteins in the complete absence of obvious sequence homology. Increasing numbers of biotechnology and pharmaceutical companies are turning heads to structure-based drug design, which promises reduced cost and time associated with traditional drug-discovery methods. For the pharmaceutical and biotech industries, structural information of proteins on an industrialized research scale has several substantial impacts in drug discovery: (i) the ability to ascribe functions to novel proteins, thereby revealing potential new drug targets; (ii) the ability to validate targets based on homology to known proteins that bind to specific chemical drug entities; (iii) optimizing hit-to-lead compounds and, subsequently, to new drug discovery; and (iv) prediction of structure and function based on sequence information by using structure–prediction algorithms. Although structural proteomics initiatives are generating protein structural data at an unprecedented rate, our current knowledge of the protein 3D structural space is still very limited.

THE PROTEOMICS REPERTOIRE OF TOOLS FOR THE DRUG-DISCOVERY PROCESS

As noted earlier, proteomics encompasses a wide range of technologies. Thus, we have identified several key areas as pivotal points in this burgeoning field: 2DGE, LC, MS, ICAT and protein biochips. A point to note, however, is that unlike the traditional 2DGE system, which is gel-based, the later technologies (LC, MS, ICAT and protein biochips) are, in essence, gel-free systems. An overview of these existing as well as emerging technologies, and the way they contribute to the drug development process, is given in the following sections.

Two-dimensional gel electrophoresis

Two-dimensional gel electrophoresis is the oldest approach and the most powerful protein separation method available today, providing parallel investigation of up to 10 000 proteins from a single gel. In 2DGE, samples are separated according to their charge (isoelectric focusing; IEF) in the first dimension, followed by size (molecular weight) using sodium dodecyl sulphate–polyacrylamide gel electrophoresis in the second dimension. Two types of IEF techniques are currently used: (i) the immobilized pH gradient (IPG) technique; and (ii) non-equilibrium pH gradient gel electrophoresis (NEPHGE). Two-dimensional gel electrophoresis is useful in identifying disease-specific proteins and in biomarker discovery by comparing the protein expression profiles between normal and diseased samples.
Interestingly, the 2DGE approach also has great strength in identifying protein modifications, such as post-translational modification, alternative splicing or proteolytic digestion, as a result of disease condition or drug treatment. However, 2DGE has been deemed as an old dying art by some due to limitations in examining proteins with extreme pH, molecular weight ranges, hydrophobicity or low abundance levels in addition to its narrow dynamic range of stain and low throughput. Despite these drawbacks, 2DGE remains the ‘gold-standard’ in proteomics discovery tools and recent advances have been made in this approach to improve handling, throughput and reproducibility. The big trend for 2DGE across the board is definitely automation. Many commercial and academic laboratories have made advances in automating the technology and other process improvements. Key players, such as OGS, have parallelized the running of many gels and OGS was the first group to have developed a rigid form of gel that is bound to a non-interfering glass support to improve gel handling. In addition, Amersham Biosciences has automated the process of spot-picking for further analysis using MS and, in 2001, Amersham and Nonlinear Dynamics launched the global distribution of Nonlinear’s Progenesis 2D-gel image analysis software.

Chromatographic techniques

An alternative approach to 2DGE is LC, a non-gel-based protein separation technique. Single- and multidimensional LC can be directly interfaced with the MS, enabling automated analysis of large amounts of data for subsequent protein identification. Shotgun proteomics, which resembles shotgun genome sequencing, is facilitated by the use of multidimensional protein identification technology (MudPIT). Multidimensional protein identification technology incorporates multidimensional high-pressure liquid chromatography (HPLC), tandem mass spectrometry (MS/MS) and database searching by the SEQUEST algorithm (http://www.thermo.com). This technology shows advantages over gel-based techniques in terms of speed, sensitivity, scope of analysis and dynamic range. Shotgun proteomics has come to preeminence only recently and relatively few descriptions of its used in human therapeutic intervention and diagnostics have been mentioned. Nonetheless, this approach has been used in the analysis of body fluids for the detection and development of novel biomarkers for human diseases. In our laboratories, we have applied this approach to the analysis of human sweat, tears and saliva, in relation to different disease states and pathological conditions. Protein components in human sweat, for example, have been evaluated in the past several decades by traditional gel electrophoresis. Among the major components identified previously were proteases, lysozymes and immunoglobulins. In our investigations, sweat samples from healthy volunteers were collected, dialysed and concentrated. The protein components were then digested to polypeptides by trypsin and analysed by two-dimensional LC tandem mass spectrometry. We found more than 300 different types of proteins in human sweat with two or more confident peptides matches and these included the immunoglobulins (IgG; IgE), dermcidin precursor (antimicrobial peptides), lysozymes, protease inhibitors and other proteins that may play important innate immune functions or antimicrobial roles (S Gopinath et al., unpubl. data, 2005). In the near future, we would expect this approach to become more widely used by scientists in academia and industries. In fact, Celera has adopted this methodology for proteome-wide analyses.

Mass spectrometry

Mass spectrometry is a method of choice for analytical characterization of potential drug molecules and protein identification. Protein structural information, such as peptide molecular weight, amino acid sequence composition, type and location of post-translational modification, could be obtained by MS analysis. Two most widely used MS are: (i) matrix-assisted laser desorption ionization time-of-flight mass spectrometry (MALDI-TOF-MS), which generates ions from solid phase samples (crystallization of samples within a matrix) and measures their peptide mass in a flight tube; and (ii) electrospray ionization mass spectrometry (ESI-MS), which generates ions from liquid samples (fine sprays of highly charged droplets from a strong electric field) and measures their mass using either quadrupole or time of flight detector. Mass spectrometry analyses of proteolytic digests provide insights to important protein attributes for protein identification. Two techniques are generally used. The first technique is peptide mass fingerprinting, which compares the masses obtained from spectrometric techniques against theoretical fingerprints from non-redundant protein databases, whereas the second, MS/MS, subjects selected tryptic digest derived from parent protein to secondary fragmentation to obtain a ‘peptide sequence tag’ that is compared against protein databases for protein identification. Recently, there have been marked improvements in MS performances, including higher sensitivity, resolution and mass accuracy, owing to improved instrumental design. Newcomers, such as ion trap, quadruple time-of-flight (Q-TOF) and fourier transform ion cyclotron resonance (FTICR) MS are increasingly being used for protein biomarker discovery to gauge disease progression and drug action. Affinity based MS methods can be used to study ligand–target interaction for lead compound discovery and optimization. Industry leaders, such as Applied Biosystems and Bruker Daltonics, are now offering TOF/TOF MS instruments; ProteomeWorks System (a global alliance of Micromass-Waters and Bio-Rad Laboratories) are advancing their latest entries, including M@LDI HT, Q-TOF micro and Q-TOF Ultima, and Thermo Electron Corporation has just introduced the Finnigan LXQ linear ion trap mass spectrometer, which enables rapid compound detection, confident structural identification and fast screening for drugs, their metabolites and degradation products. In short, MS-based techniques will continue to flourish and come to the forefront in drug discovery process.

Isotope-coded affinity tag

Isotope-coded affinity tag is a chemical modification strategy that enables the rapid and accurate quantification of protein activity and concurrent sequence identification of proteins in a complex mixture. The ICAT approach involves selective conjugation of cysteine thiol groups with two identical reagents isotopically different in mass (one light and one heavy linker) and biotinylated affinity tags, followed by proteolytic digestion and quantitative analysis of the peptide-conjugates by LC/MS. The isotopic difference allows relative abundance measurement of the same peptide ions from two different samples (normal vs diseased or control vs treated) by comparing signal intensities in MS mode and subsequent MS/MS peptide sequencing results for protein identification in altered expression levels. Isotope-coded affinity tag has tremendous advantages in the analysis of membrane proteins and low abundant proteins, which are likely to be promising targets for therapeutic and diagnostic
development. The approach also offers high-throughput analysis and is highly automated and reproducible. However, a major drawback of the earlier versions of the technology is that the protein must contain cysteine residues in its structure. Because of this requirement, many important proteins, including those with post-translational modifications, are overlooked. Several groups have since described alternative labelling strategies that target lysine and tryptophan residues or peptide N- or C-termini. These include a non-isotopic variation called the mass-coded abundance tagging (MCAT) strategy developed in 2002.22 In this technique, O-methyl isourea is used to guanidinate the C-terminal lysine residues of peptides in one sample, whereas the other sample remains unmodified; the samples of each are combined and the relative abundance of each can be determined by mass spectrometry.

The ICAT technology has been commercialized by Applied Biosystems (a business unit of Applera).1 They have now improved the system to include other isotopic labelling variants, for example, a set of four stable isotopic labels released under the brand name iTRAQ, that can be used for multiplexed protein profiling up to four different samples. These reagents use labels that bind covalently to lysine side chains and N-termini of any peptide and a reporter group specific to each tag. The reporter groups generate fragment ions that appear in the low-mass region of the spectrum where other fragment ions are not generally found and, thus, the protein abundance of samples differentially labelled with each reagent can be compared by looking at the ratio of the peak areas of each reporter group.

Like any other discovery science method, the ICAT method is, by itself, not sufficient to address an entire biological process. Prefractionation methods to reduce the complexity of the generated peptides, ICAT technology in combination with some form of MS instrument, as well as improved bioinformatics tools for automated data collection and interpretation23 will undoubtedly provide us with a comprehensive map of protein networks and their biological functions in cells and tissues under conditions of health and disease.

**Protein chips**

Protein chips, protein biochips or protein microarrays are microproteomic technologies for studying protein interaction and function. Two basic formats of protein arrays have been generated, the forward phase array (FPA) and the reverse phase array (RPA).24 In the FPA, a biochip surface is arrayed with a specific substrate (‘bait’), such as antibodies, proteins, peptides, nucleic acids or other small molecules, and is used to select specific bound analyte to the bait molecule from a heterogeneous mixture of sample analytes. A large amount of patient sample is required for processing such an array. In the RPA, the whole repertoire of a patient’s protein representing

---

**Fig. 3** Use of rapid Ciphergen (http://www.ciphergen.com/) protein chip arrays to differentiate individuals with high levels of exposure to environmental aflatoxin from those with minimal or no exposure. (a) Peak map and (b) gel views of representative plasma samples from the two exposure groups, with the major differential peak highlighted. (c) Three-dimensional principle component analysis plot differentiating the plasma profiles of the high exposure group (pyramids) compared with the control group (diamonds; N Jiang et al., unpubl. data, 2005).
the diseased state of individual tissue cell populations is arrayed directly onto a spot on a biochip surface.24–26 Many patient specimens could be deposited onto the same biochip surface, with each substrate being probed with a distinct antibody. This technology uses only a very small amount of crude sample (e.g. cell lysates, urine or serum) from the patient.25,26 Protein chips are viewed as the most promising tools for proteome-wide analysis. This technology enables thousands of proteins to be analysed en masse (a more rapid profiling approach compared with 2DGE) and it enables screening for specific types of post-translational modifications. Thus, proteins may be designed to study receptor–ligand interactions, enzyme activity and downstream signalling events, protein–protein and protein–antibody interactions as well as to serve as the basis of a serodiagnostic chip.27 Commercial players, for instance Ciphergen Biosystems and Phylos, are developing a wide range of protein arrays. Ciphergen has developed ProteinChip Arrays, which use chromatographic surfaces derivatized with different affinity matrices to capture proteins of interest.28 These proteins are then purified and analysed through a surface-enhanced laser desorption/ionization time-of-flight (SELDI-TOF) mass spectrometer.29 Based on the comparisons of protein mass profiles from any two samples from different biological and pathological conditions, potential biomarkers or disease-related protein targets could be identified. We have used this rapid profiling chip array system for a variety of studies, each exploiting the key strengths and advantages of this technology: (i) it being amendable to analyse small and usually limited quantities of biological samples (0.5–500 µL); (ii) the ability to detect and evaluate proteins without the need for tagging, labelling or processing; (iii) the sensitivity of the system with the ability to detect fmol concentrations of proteins; and (iv) the rapidity in obtaining results, thus making it amendable for analysis even in large epidemiological settings. One such example is the use of the Ciphergen protein chip arrays to identify plasma markers in individuals exposed to high levels of environmental aflatoxin, which could lead to liver cancer (Fig. 3). Samples obtained from finger pricks and in an epidemiological setting were evaluated on the protein chip arrays. Our data show that the plasma protein profiles of individuals with high chronic environmental aflatoxin exposure could be differentiated from individuals coming from environments with minimal or low exposure.

LOOKING TOWARDS THE FUTURE

The development of proteomic techniques and technologies has introduced a new dimension into the drug-discovery arena. By elucidating the pathogenesis of human diseases, proteomics could aid in the discovery of patient-tailored molecular medicine (personalized medicine) of the future. Proteomics has set foot in every stage of the drug-discovery process, from target discovery to clinical trials, often providing new insights and interesting discoveries. Future improvements in this field promise to streamline drug development and to continually meet the quest for safer, more effective and cost-effective therapeutics. We are only at the beginning and it seems as if the sky is the limit.

REFERENCES


