Cellular Microbiology (2005) **7**(8), 1061–1076 doi:10.1111/j.1462-5822.2005.00554.x

Microreview

Proteomics meets microbiology: technical advances in the global mapping of protein expression and function

Carolyn I. Phillips[1] and Matthew Bogyo[1,2]
Departments of [1]Microbiology and Immunology and [2]Pathology, Stanford University School of Medicine, 300 Pasteur Dr., Stanford, CA 94305-5324, USA.

Summary

The availability of complete genome sequences for a large number of pathogenic organisms has opened the door for large-scale proteomic studies to dissect both protein expression/regulation and function. This review highlights key proteomic methods including two-dimensional gel electrophoresis, reference mapping, protein expression profiling and recent advances in gel-free separation techniques that have made a significant impact on the resolution of complex proteomes. In addition, we highlight recent developments in the field of chemical proteomics, a branch of proteomics aimed at functionally profiling a proteome. These techniques include the development of activity-based probes and activity-based protein profiling methods as well as the use of synthetic small molecule libraries to screen for pharmacological tools to perturb basic biological processes. This review will focus on the applications of these technologies to the field of microbiology.

Introduction

With the completion of over 1000 prokaryotic and eukaryotic genome sequences and hundreds of sequencing projects currently underway (Celestino *et al.*, 2004), researchers are now faced with the daunting task of assigning function to the thousands of predicted gene products. While microarray technologies (Shalon *et al.*, 1996) provide invaluable information concerning the levels of global mRNA expression, it is clear that in many cases mRNA transcripts do not directly correlate with protein expression (Anderson and Seilhamer, 1997; Gygi *et al.*, 1999a). Therefore, the field of proteomics is challenged with the task of providing both quantitative and functional data to further complement genomics efforts.

Proteomics is traditionally defined as the analysis of a complete set of proteins (proteome) of a given cell or organism. This field has primarily been limited to bulk identification and/or analysis of protein abundance. The most cost-effective and simplest way to carry out such an analysis makes use of two-dimensional gel electrophoresis coupled to mass spectroscopy-based sequencing methods (2DE-MS; Figs 1 and 2). 2DE-MS has been used extensively in the field of bacteriology to characterize and compare proteomes of pathogenic bacteria (Langen *et al.*, 2000; Vandahl *et al.*, 2001; Schmidt *et al.*, 2004), and has recently found applications in the field of parasitology (Rout and Field, 2001; Drummelsmith *et al.*, 2003; Parodi-Talice *et al.*, 2004) as more genome sequences of protozoan parasites become available. Recent advances in the use of multidimensional liquid chromatography-based separations (Fig. 3) have dramatically improved the scope and depth of proteomic analysis (Graves and Haystead, 2002) and have been applied to mapping out proteomes of critical pathogenic organisms such as *Plasmodium falciparum* (Lasonder *et al.*, 2002).

While classical proteomic techniques provide methods to determine the identity and abundance of proteins within a given proteome, an understanding of protein function can only be inferred from changes in expression. In the case of enzymes, activity is often tightly regulated by a series of post-translational controls and therefore may not be directly correlated with protein abundance. To address the need to profile global patterns of enzyme activity, a new set of proteomic techniques has been developed that make use of small molecule chemical probes that provide an indirect readout of enzyme activity (Fig. 4). Furthermore, several recent large-scale screening efforts have been used to search for novel pharmacological tools to assess global protein function in pathogenic organisms such as *Toxoplasma gondii* (Carey *et al.*, 2004) (Fig. 6). In this review, we describe advances in the fields of classical and chemical proteomics and discuss current and future applications to the study of microbiology.

Received 15 March, 2005; Accepted 2 May, 2005. For correspondence. E-mail mbogyo@stanford.edu; Tel. (+1) 650 725 4132; Fax (+1) 650 725 7424.

© 2005 Blackwell Publishing Ltd

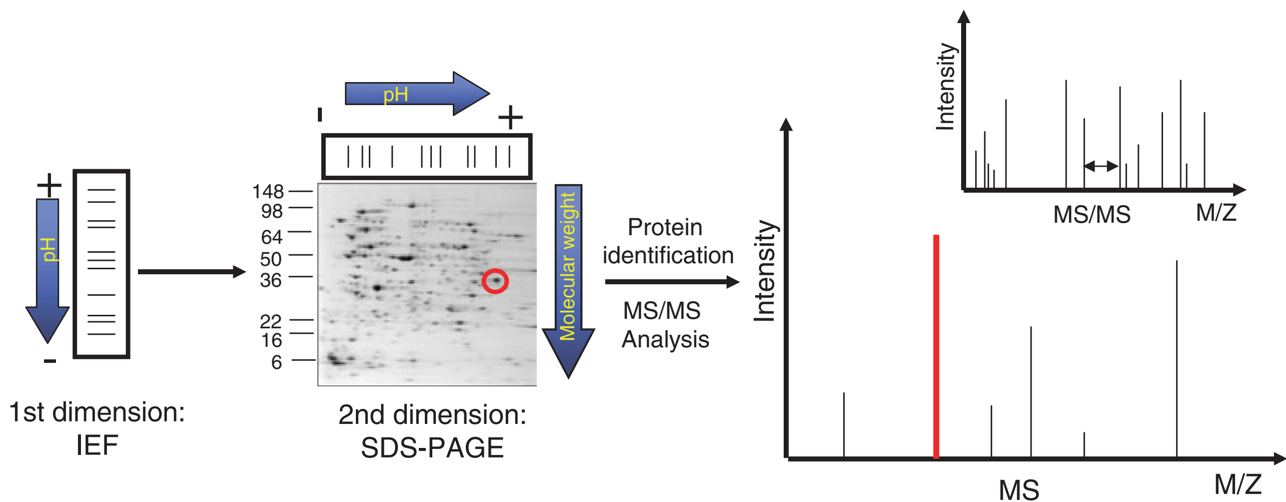


Fig. 1. Classical proteomics methods – 2DE-MS. Proteins are separated based on net charge by isoelectric focusing (IEF) in the first dimension. During this phase, proteins migrate towards the charged electrode through the gel strip containing a specific pH gradient until they reach their isoelectric point (pI), at which point they carry no net charge. Following isoelectric focusing, proteins are further separated in the second dimension based on their molecular weight using standard SDS-PAGE. Gels are then stained to identify protein spots and individual spots (shown with red circle) are excised from the gel, trypsin-digested and extracted peptides are sequenced using tandem mass-spectrometry. In the first MS analysis, peptides are ionized and resolved on the basis of mass to charge ratio (MS). From the resulting mass spectrum, parent ions (red peak) are selected for MS/MS analysis. These ions are passed through a collision cell where they undergo fragmentation along the peptide backbone to produce a predictable pattern of fragments (MS/MS). Differences in fragment ion masses in the MS/MS spectra (see double-headed arrow) can be used to determine the sequence of peptides from the target protein. MS/MS fragment profiles can be used to directly search databases for predicted peptide matches.

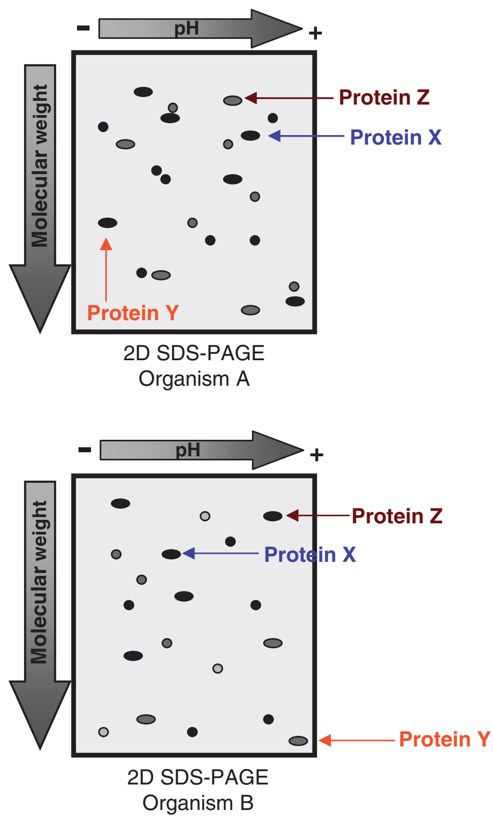
Classical proteomics: variations on a theme

The field of proteomics was established over a decade ago with the development of facile mass spectrometry-based protein sequencing methods coupled with methods to resolve protein mixtures using a variety of gel-based and chromatographic techniques. The critical parameter that has dictated the success or failure of this field is the ability to resolve individual proteins in complex mixtures. One of the most effective ways to separate proteins in a complex mixture is by the use of two-dimensional (2D) gel-based electrophoretic separations (Pandey and Mann, 2000) (Fig. 1). Using these methods proteins are separated according to net charge in the first dimension (isoelectric focusing, IEF) and by molecular weight in the second dimension (SDS-PAGE) (O'Farrell, 1975). 2DE has the ability to resolve a large number of proteins including those with post-translational modifications (which often cause a change in charge and/or apparent molecular weight) as well as unique forms of proteins that result from differential mRNA splicing or proteolysis (Anderson and Anderson, 1998; Cordwell *et al.*, 2001). Following separation on 2D gels, proteins are stained to allow visualization, excised as gel pieces and digested in-gel with trypsin. The resulting tryptic peptides can then be extracted from the gel slice and sequenced using mass-spectrometry-based sequencing methods followed by identification of proteins by database searching. In general, 2DE-MS methods are used for two primary purposes: reference mapping and protein expression profiling (Cordwell *et al.*, 2001) (Fig. 2).

Reference mapping

The goal of reference mapping is to define the identity and location on an index map of as many proteins of an organism as possible (Fig. 2) (Cordwell *et al.*, 2001). This index map then provides a point of reference for experiments aimed at studying the response of a given organism to various external stimuli. Reference mapping has been extensively used in the field of bacteriology and reference maps have been completed for a number of important human pathogens (Langen *et al.*, 2000; Vandahl *et al.*, 2001; Cho *et al.*, 2002; Cohen *et al.*, 2002; Drummelsmith *et al.*, 2003; Parodi-Talice *et al.*, 2004). One of the most comprehensive reference maps completed to date is that of *Haemophilus influenzae* covering nearly 33% of its open reading frame (Bumann *et al.*, 2001). The success of reference mapping of important pathogenic bacteria is a direct result of the completion of sequencing of more than 172 prokaryotic genomes coupled with more than 500 sequencing projects currently underway (Celestino *et al.*, 2004).

On the other hand, there are currently only 20 genome sequencing projects underway for protozoan parasites (http://www.ncbi.nlm.nih.gov/genomes/static/EG_T.html) with only one complete genome sequenced, that of *P.*



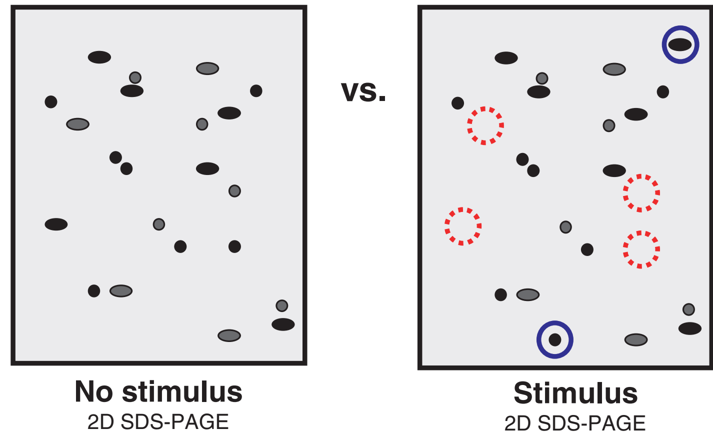


Fig. 2. Classical proteomics methods – reference mapping. Reference mapping (left) involves identification of all resolvable proteins on a 2D gel. The main goal is to determine both the identity and the location of proteins (in this example proteins X, Y and Z) in individual organisms. These reference maps can then be used as the basis for protein expression profiling (PEP; right). PEP compares either differences in proteomes of different organisms or differences in protein expression of a single organism under different stimuli (right). This allows visualization of both the downregulation (dotted red circles) and upregulation (blue circles) of proteins under different conditions.

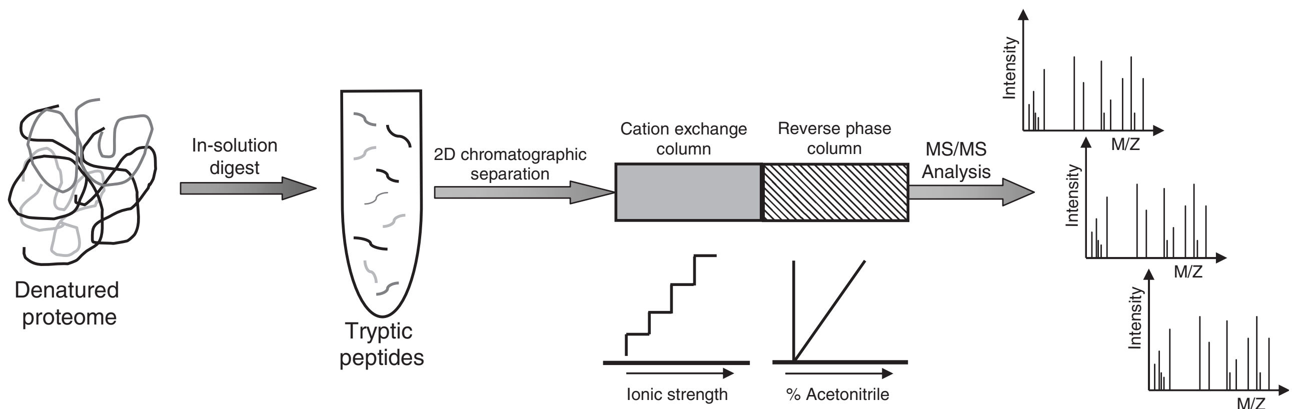


Fig. 3. Classical proteomics methods – gel-free methods. Complex protein mixtures are denatured and digested in solution with trypsin. Peptides are subsequently separated by two-dimensional liquid chromatography and directly analysed by mass spectrometry. Peptides are first separated via an ion exchange column where peptides are eluted by stepwise increases in salt concentrations. Eluted proteins bind to an in-line reverse-phase C18 column and are eluted based on increasing hydrophobicity. Resulting MS/MS spectra, obtained over the lifetime of the chromatographic separation, are compiled and used to identify proteins in the complex starting sample.

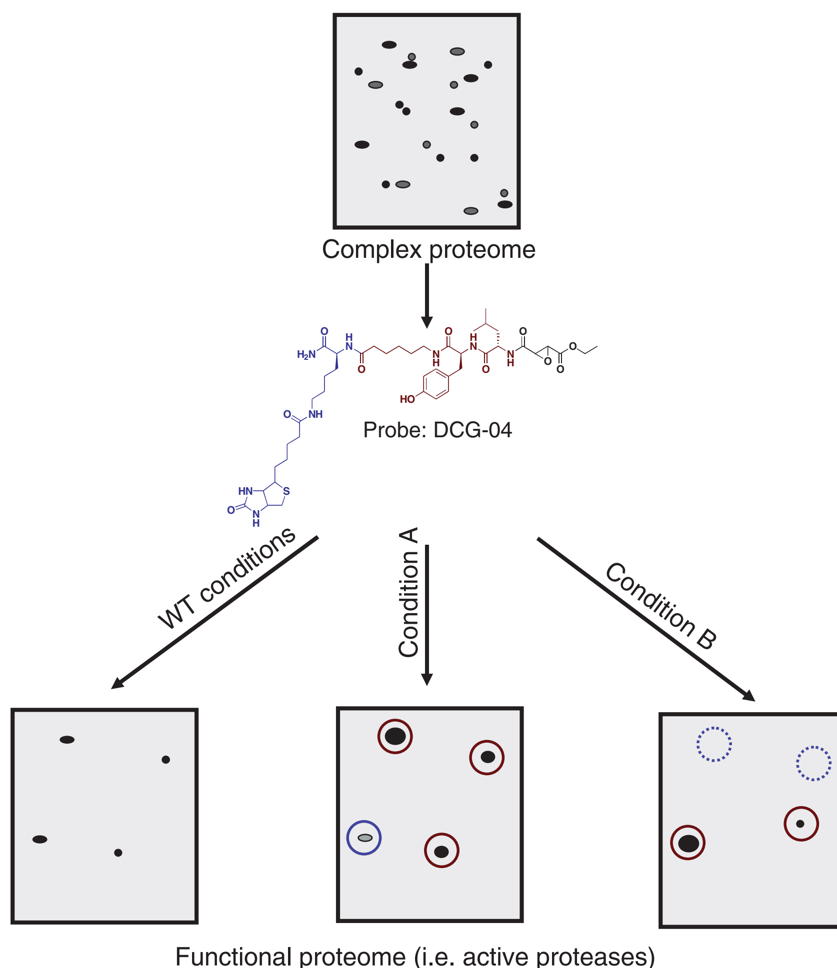


Fig. 4. Activity-based profiling of enzyme targets in complex proteomes. The use of activity-based probes significantly simplifies a complex proteome allowing the analysis of a specific set of protein activities (also referred to as the functional proteome). A complex proteome is labelled with an ABP (in this example, the general cysteine protease probe, DCG-04; Greenbaum *et al*., 2000). Proteins are separated by 2DE (or 1DE) and only active enzymes visualized by the probe tag (biotin, fluorophore or radiolabel). Proteomes can be probed under varying conditions/stimuli in order to determine changes in protein activity. Increases in spot (or band) intensity (red circles) correlate with increased protein activity. A decrease in spot intensity (blue circle) or the disappearance of spots (dotted blue circles) correlates with a decrease and disappearance of protein activity respectively. Spot intensities can be directly quantified and compared with spot intensities under native conditions to determine relative changes in protein activity.

falciparum (Gardner *et al*., 2002). This lack of available sequence information has proven to be a bottleneck in the characterization of parasitic proteomes. Recently, large-scale efforts have been made to characterize the complete proteomes of several important parasites including *P. falciparum* (Lasonder *et al*., 2002), *Trypanosoma cruzii* (Parodi-Talice *et al*., 2004), *Trypanosoma brucei* (van Deursen *et al*., 2003) and *Leishmania* sp. (Drummelsmith *et al*., 2003). As many of the genomes remain incomplete, identification of proteins resolved by 2DE often relies on incomplete genomes and close sequence homology between related species. In the case of *Leishmania mexicana*, identification of 47 proteins of the 2000 mapped by 2DE relied heavily on the genome sequence of the closely related *Leishmania major* (Nugent *et al*., 2004). With more attention focused on the global health importance of parasitic diseases, it is likely that significant headway will be made in the area of proteome mapping in the near future.

Protein expression profiling

Reference mapping provides the basis for identifying proteins of interest in pathogenic microorganisms through protein expression profiling (PEP; Fig. 2) (Cordwell *et al*., 2001). Proteins from an organism exposed to different stimuli are resolved on separate 2D gels and the gels are subsequently analysed using computer software to identify changes in protein expression in the two samples. PEP has been used for identification of proteins that are regulated in antibiotic resistant pathogens (McAtee *et al*., 2001; Drummelsmith *et al*., 2003), comparison of protein expression as a result of external stimuli (Guina *et al*., 2003; Kan *et al*., 2004; Nirmalan *et al*., 2004), mapping responses to antibiotics (Evers *et al*., 2001; Gmuender *et al*., 2001; Singh *et al*., 2001) and identification of life cycle/stage-specific proteins of protozoan parasites (Carucci *et al*., 2002; El Fakhry *et al*., 2002; Bente *et al*., 2003; Gongora *et al*., 2003; Curwen *et al*., 2004; Paba *et al*., 2004a).

The evolutionary success of bacteria is highly dependent on their ability to adapt to multiple environmental stresses including oxygen and nutrient supply, pH and naturally occurring antibiotics (Cordwell *et al*., 2001). Moreover, many synthetic antibiotics mimic naturally occurring

stresses (i.e. oxazolidinones simulate starvation by inhibiting protein synthesis; Mukhtar and Wright, 2005). To this end, understanding proteome changes in response to different external stimuli can identify various regulatory mechanisms as well as potential targets for the development of novel drug therapies. For example, *Vibrio cholera* causes severe diarrhoeal disease in humans resulting from bacterial colonization of the small intestine and release of cholera toxin. *V. cholerae* has two predominant habitats: the natural environment (highly aerobic) and the human small intestine (highly anaerobic). A comparative 2DE-MS analysis of *V. cholera* grown in aerobic versus anaerobic conditions was recently performed to identify proteins differentially expressed between the two environments (Kan *et al*., 2004). Approximately 244 proteins were found to be differentially regulated, of which several were identified as motility factors suggesting that the bacteria use a controlled change in motility during colonization, multiplication and spread within the human host. This study provides the starting point for future studies aimed at understanding *V. cholera* pathogenesis and vaccine development and introduces some potentially important new antibiotic targets.

In addition to understanding microbial response to external stimuli, PEP has proven invaluable in identifying novel drug targets of protozoan parasites and understanding mechanisms of drug resistance. *Leishmania* sp. have been used extensively in the field of parasite proteomics due to the availability of a near-complete *L. major* genome and the ease with which it can be cultured and manipulated (Drummelsmith *et al*., 2003). *Leishmania donovani* causes a variety of diseases in humans, several of which can be fatal. *L. donovani* is a vector-borne parasite that cycles between the definitive fly host (promastigote stage) and the mammalian host (amastigote stage). Sixty-seven proteins were identified by PEP to be differentially expressed between the two stages (Bente *et al*., 2003). Of these, 41 matched proteins of known function in the *L. donovani* database and could be separated into five functionally distinct categories: (i) stress response proteins (i.e. HSP60, HSP70), (ii) cytoskeleton and cell membrane associating proteins (i.e. α- and β-tubulin, ArfGap, enolase), (iii) proteins involved in energy metabolism and phosphorylation (i.e. aldolase, F1-ATPase, arginine kinase), (iv) cell cycle and proliferation-associated proteins (i.e. Ran-binding protein, negative regulator of cdc42p, replication factor A) and (v) proteins involved in amino acid metabolism (i.e. alanine transaminase). One of the major limitations of this study and similar studies in other pathogenic organisms is the identification of mainly cytosolic proteins, thereby excluding potentially important extracellular membrane proteins involved in host–pathogen interactions. However, these studies do shed some light on pathways used by the parasite to productively infect a host organism.

Limitations of 2DE-based proteomics

Bottlenecks in proteome mapping occur as a result of the inability of 2D gels to adequately resolve several classes of proteins including very large or small proteins, membrane-associated proteins, extremely hydrophobic, acidic or basic proteins and proteins found in low abundance within the cell (Corthals *et al*., 2000; Gygi *et al*., 2000a; Graves and Haystead, 2002). Even with advances in protein staining techniques including the use of fluorescent dyes with increased sensitivity (Lopez *et al*., 2000; Berggren *et al*., 2002) and the introduction of specialized pH gradients allowing for increased resolution (Bjellqvist *et al*., 1993), a high percentage of proteins within an organism cannot be mapped and identified using 2DE-MS. In addition to technical difficulties in resolving proteins, 2DE gels are also extremely labour intensive requiring several separations for comparative analyses. Furthermore, 2D gels methods are not easily adapted to automation (Gygi *et al*., 2000a; Graves and Haystead, 2002).

Technological advances in proteomic methods

Differential gel electrophoresis (DIGE)

The introduction of differential gel electrophoresis (DIGE) (Tonge *et al*., 2001) has addressed some of the pitfalls and limitations of classical protein expression profiling. This technique makes use of fluorescent dyes that covalently modify all proteins in a given proteomic sample. Using two fluorescent dyes with non-overlapping excitation and emission profiles it is possible to label two distinct proteomes that can then be compared on a single gel by mixing the samples. This method allows comparative studies using a single gel separation and solves issues associated with lack of run-to-run reproducibility of 2D gels (Tonge *et al*., 2001; Gharbi *et al*., 2002). This technique has been successfully applied to the study of protein processing events involved in invasion of host cells by the protozoan parasite *T. gondii* (Zhou *et al*., 2004). The protein complex MIC2/M2AP is proteolytically processed at several points along the secretory pathway and at the parasite surface culminating in a final cleavage event at the parasite surface (Carruthers *et al*., 2000). This key processing step releasing the complex from the parasite surface is accomplished by an as of yet unidentified protease, MPP1. In an effort to identify substrates of a second unrelated surface shedding activity termed MPP2, a 2D-DIGE method was used to compare parasites treated or not treated with the tripeptide aldehyde Ac–Leu–Leu–norLeucinal (ALLN) protease inhibitor (Zhou *et al*., 2004). This study confirmed that in addition to processing MIC2 and M2AP, MPP2 also appears to be responsible for processing MIC4 and SUB1.

Simplification of the proteome: LC/MS/MS and MuDPIT

The current trend in proteome mapping is to simplify the proteome from a complete proteome to a partial or enriched proteome, thus resolving some of the limitations in resolution and identification of proteins in highly complex samples (discussed below). This subproteome is often a fraction contained within a specific subcellular organelle(s) of interest. For example, membrane and extracellular proteins are likely to be of interest for drug development strategies as these proteins are often directly involved in host–pathogen interactions (Cordwell *et al*., 2001). Additionally, to address some of the limitations of protein resolution of 2D gels, several gel-free techniques have been developed. These include direct liquid chromatography separation of tryptic peptides coupled to tandem mass spectrometry (LC/MS/MS) (McCormack *et al*., 1997) and multidimensional liquid chromatography coupled to MS/MS analysis (MuDPIT) (Washburn *et al*., 2001; Wolters *et al*., 2001). While these are powerful methods there are some limitations of LC-based proteomic methodologies including the need for highly advanced instrumentation that many groups may not be able to access directly, and the limited ability to detect protein subspecies resulting from differential splicing or post-translational modification. To circumvent these limitations, several studies have used subcellular fractionation followed by simple one-dimensional (1D) LC/MS/MS to identify proteins of a specific organelle or subproteome (Jefferies *et al*., 2001; Bromley *et al*., 2003; Blonder *et al*., 2004; Sam-Yellowe *et al*., 2004; Tjalsma *et al*., 2004; Calvo *et al*., 2005) in an effort to identify new targets for drug development.

In one of the most comprehensive analyses of the cell membrane proteome of *Pseudomonas aeruginosa* to date, 786 cell membrane-associated proteins were identified (Blonder *et al*., 2004). Significantly, these studies identified a potentially novel penicillin amidase as well as the mexB transporter, both of which are involved in antibiotic resistance and the former of which may prove to be a target for the development of novel drug therapies. Furthermore, many parasites express proteins on the surface of host cells in order to control immune responses and cell adhesion. These proteins may also prove to be relevant targets for vaccine and drug development.

Direct LC/MS/MS methods have also been used to identify proteins expressed at the surface of erythrocytes infected by *P. falciparum* (Florens *et al*., 2004). These studies identified two novel proteins, PIESP1 and PIESP2, conserved across *Plasmodium* species that are likely to be candidates for drug development due to their broad expression across various *Plasmodium* isolates.

While a somewhat challenging technique, MuDPIT has proven useful in extending the fraction of the proteome capable of being analysed (Fig. 3). The use of two dimensional (cation exchange and reverse phase) chromatographic separation coupled to MS/MS analysis provides additional resolution for proteome mapping (Washburn *et al*., 2001). A large-scale effort to identify the proteomes associated with the four primary life cycle stages of *P. falciparum* (sporozoites, merozoites, trophozoites and gametocytes) succeeded in mapping over 2400 proteins. These proteome maps identified sets of proteins that were differentially expressed between the four stages. Interestingly, nearly 350 hypothetical integral membrane proteins and cytoplasmic proteins with signal sequences were identified. These are of great interest to the research community as they represent candidate targets for novel drug development (Florens *et al*., 2002).

MuDPIT has also been used to identify proteomes of subcellular organelles. Many of the Apicomplexan parasites have specialized secretory organelles including micronemes, rhoptries and dense granules that contain the majority of their proteins involved in host–pathogen interaction (Dubremetz *et al*., 1998). One such study focused on identifying the protein reportoire of the *Plasmodium* merozoite rhoptry organelle (Sam-Yellowe *et al*., 2004). Rhoptries are involved in erythrocyte invasion by the merozoite form of the parasite; however, relatively few rhoptry proteins have been identified. Following subcellular fractionation to isolate rhoptry organelles, bioinformatic and proteome analysis identified 36 proteins including known rhoptry proteins, proteases and proteins involved in lipid metabolism. These studies have shown that by making the proteome map less complex it is possible to identify low abundance and membrane-associated proteins and to correlate previously unidentified proteins with likely subcellular localizations thus facilitating their functional characterizations. While this is a powerful method for identification of subproteomes, the relevance of the results are dependent on the quality of fractionation protocols and controls used as organelle markers.

Quantitative proteomics

While 2DE-MS has been used for qualitative and quantitative proteome analysis it is often difficult to obtain direct and accurate quantitative information from gel-based methods. Direct quantitative analysis of proteins separated by 2DE requires a high degree of reproducibility, something that is often difficult to achieve due to the parallel nature of 2DE. In addition, proteins can only be quantified if their abundance falls within the dynamic range of the protein staining/detection method used and only spots representing one protein species can be considered for quantitative analysis (Adam *et al*., 2002a; Schmidt *et al*., 2004). Combined, these limitations have

hindered direct quantitative analysis of protein abundance by classical proteomic methods.

Isotope-coded affinity tagging (ICAT) reagents were recently developed for the simultaneous accurate quantification and identification of proteins within a complex proteome (Gygi *et al.*, 1999b). ICAT reagents consist of a cysteine reactive moiety, a biotin affinity tag for isolation and purification of tagged peptides and a linker region consisting of either eight hydrogens (light tag) or eight deuteriums (heavy tag) (Gygi *et al.*, 1999b; Moseley, 2001; Adam *et al.*, 2002a). In a typical experiment, two complex proteomes of interest are labelled with the light and heavy ICAT reagents, respectively, mixed and then subjected to in-solution tryptic digestion followed by LC/MS/MS. Peptides are quantified in the MS mode based on relative signal intensities of identical peptides labelled with either the heavy or light form of the ICAT reagent. This approach offers several advantages over 2DE techniques by circumventing the need for gel-based separation (Moseley, 2001; Gygi *et al.*, 2002; Schmidt *et al.*, 2004). ICAT reagents have been used successfully to quantify proteome changes in developmental stages and responses to external stimuli of several bacterial and parasitic pathogens. One such study compared the trypomastigote and amastigote stages of *T. cruzi* (Paba *et al.*, 2004b). As expected, proteins involved in structural changes were found to be differentially expressed between stages. In particular, the loss of PAR (paraflagellar rod) proteins was observed in amasitogotes. Furthermore, cruzain, the major cysteine protease activity in *T. cruzi*, showed higher expression in amastigotes. Cruzain has been shown to be involved in mediating critical calcium-signalling in host cells involved in invasion.

Isotope-coded affinity tagging reagents have also been used to analyse the proteome of *Mycobacterium tuberculosis* (Schmidt *et al.*, 2004). Comparing virulent and attenuated strains, these studies identified three proteins (Rv0223c, Rv1513 and Rv0570) that were unique to the virulent strain representing potential candidate vaccine and therapeutic targets. Moreover, of clinical relevance, proteins encoded by two additional genes (Rv2935 and Rv2940c) previously implicated in virulence of *M. tuberculosis* were found to be upregulated in the attenuated strain. These differential expression patterns in the two strains highlight the unique regulatory mechanisms that are in use.

Isotope-coded affinity tagging reagents have undergone several improvements (Smolka *et al.*, 2001; 2002; Patel and Perrin, 2004; van Swieten *et al.*, 2004) since their emergence in the late 1990s including the development of fluorescent (visible) ICAT reagents (Bottari *et al.*, 2004; Lu *et al.*, 2004), cleavable ICAT reagents (Hansen *et al.*, 2003; Li *et al.*, 2003; Yi *et al.*, 2005) and incorporation of different reactive moieties (Moseley, 2001). However, the current available ICAT reagents suffer from several weaknesses (Gygi *et al.*, 2000b; Hamdan and Righetti, 2002; Smolka *et al.*, 2002). The primary limitation of ICAT reagents is the requirement of a free cysteine on the protein being labelled. It is estimated that as much as 20% of any given proteome lacks a free cysteine residue making these proteins transparent to ICAT analysis. In addition, ICAT methods do not provide information about post-translational modifications such as phosphorylation. Moreover, as highlighted in the analysis of the *M. tuberculosis* proteome, ICAT methods show preference for high-molecular-weight proteins while 2DE-MS shows a preference for low-molecular-weight proteins and cysteine-free proteins. Of the hundreds of proteins identified by ICAT LC/MS/MS and 2DE-MS only 27 were common to both data sets (Schmidt *et al.*, 2004). This highlights the importance of using complementary analysis methods to maximize proteome coverage.

Chemical proteomics: it's all about activity!

While classical proteomic techniques have found useful applications in the field of microbiology over the past decade, these methods only provide information regarding identity and abundance of proteins. Additionally most classical proteomic techniques do not allow direct assessment of post-translational regulation by factors such as protein–protein and/or protein–inhibitor interactions so they often fail to provide an understanding of global protein function (Gygi *et al.*, 1999a). In particular, many enzymatic proteins such as proteases are synthesized as precursor (zymogen) forms that must be post-translationally activated resulting in activity profiles that are often poorly correlated with overall abundance. In order to address this limitation in classical proteomic methods, the field of chemical proteomics has recently been established as a means to profile global patterns of enzyme activity through the use of synthetic small molecules known as activity-based probes (ABPs) (Cravatt and Sorensen, 2000; Greenbaum *et al.*, 2002a). These chemical reagents can be designed to target a distinct subset of enzymatic targets and facilitate identification as well as activity-based protein profiling (ABPP) of enzymes in complex proteomes. These methods have proven to be a powerful addition to the proteomics toolbox, however, as with the use of any kind of small molecule tool there exists the possibility of off-target effects that are difficult to resolve. Generally, these issues can be addressed with detailed follow-up studies using classical genetics and biochemistry and by using multiple classes of probes that modify the same enzyme target.

The central component of this method is the ABP which consists of three basic structural elements: a reactive group (or 'warhead'), a linker/specificity region and a tag

region (Jeffery and Bogyo, 2003). The warhead group serves as the key functional element that directs covalent, activity-dependant modification of a target enzyme or enzyme family. The linker region serves as a scaffold for attachment of functional groups that direct the selectivity of the probe and the tag is used to visualize the resulting modified target enzymes. A number of different warheads, linkers and tags have been used and are the focus of several other extensive reviews (Adam *et al*., 2002a; Jeffery and Bogyo, 2003; Berger *et al*., 2004; Speers and Cravatt, 2004). Two basic approaches have been taken in the design of ABPs: a directed approach aimed at designing probes to target a specific enzyme or class of enzymes and a non-directed approach aimed at characterizing enzyme families for which selective active-site reagents have not yet been identified. In both cases, the ABP provides a facile means to simplify a complex proteome into a 'functional' proteome thus allowing a focused look at regulation of a small set of enzymatic targets (Fig. 4).

The design and synthesis of activity-based probes

Directed approach

The design and synthesis of directed ABPs has capitalized on the wealth of information resulting from medicinal chemistry efforts aimed at developing selective inhibitors for a number of enzyme families. Thus, it is possible to use well-established chemistries to target a number of enzymes. Often, creation of a chemical probe requires no more than simply tagging a covalent inhibitor whose selectivity and reactivity have already been defined. In other cases, extensive synthetic chemistry efforts are required to develop probes with the necessary reactivity, selectivity and labelling characteristics. To date a range of directed ABPs have been developed to target serine hydrolases (Liu *et al*., 1999; Kidd *et al*., 2001), caspase family cysteine proteases (Thornberry *et al*., 1994; Faleiro *et al*., 1997; Winssinger *et al*., 2002), papain family cysteine proteases (Bogyo *et al*., 2000; Greenbaum *et al*., 2000), metalloproteases (Chan *et al*., 2004; Saghatelian *et al*., 2004), phosphatases (Lo *et al*., 2002), the proteasome (Bogyo *et al*., 1997; Kessler *et al*., 2001; Nazif and Bogyo, 2001) and the ubiquitin-specific proteases (Borodovsky *et al*., 2002; 2005; Hemelaar *et al*., 2004; Ovaa *et al*., 2004). These reagents have been applied to a number of important biological studies that have led to a greater understanding of protein function on a global scale. Some examples of these applications are discussed below.

Non-directed approach

While directed ABPs have proven useful for the study of several key enzyme families, they are limited on a broader scale by the fact that selective affinity labels have not been identified for a large number of proteins/protein families. The purpose of a non-directed approach is to expand the number of targets that can be profiled using ABPs (Adam *et al*., 2001; Speers and Cravatt, 2004). Using combinatorial approaches it is possible to begin to generate diverse sets of potential probes that can then be applied to profiling experiments in complex proteomic samples.

One effort to identify novel chemical proteomic tools made use of a library of probes containing a sulphonate ester reactive moiety (Adam *et al*., 2002b). This warhead was selected because of its moderate reactivity and ease of synthesis. The primary functional groups attached to the warhead were varied using a combinatorial approach. Proteins labelled by these probes in an activity-dependant manner could be easily identified by affinity isolation followed by mass spectrometry-based sequencing. Initial studies identified six mechanistically distinct enzyme families including epoxide hydrolases, aldehyde dehydrogenases, thiolases, NAD/NADP-dependant oxidoreductases and enoyal coA hydratases that were not targeted with other directed ABPs.

To address some of the limitations in structural diversity of the sulphonate ester-based ABPs, a second study developed ABPs containing an α-chloroacetamide (α-CA) reactive group linked to a range of dipeptide scaffolds (Barglow and Cravatt, 2004). Again, the α-CA functional group was chosen as a warhead because of its relatively mild reactivity and because carbon electrophiles have been shown to have no strong bias towards a particular active-site functional group. Therefore, a set of diverse protein families could be labelled by these non-directed probes. Profiling of complex proteomes derived from tissue samples and cell lines produced distinct reactivity profiles and showed little overlap with the sulphonate ester library. Thus, the two libraries target unique subsets of the proteome.

Applications for activity-based probes

The successful development of a number of novel ABP families has led to the parallel development of a range of applications for these probes. Several strategies aimed at identifying novel candidate drug targets (target discovery) and chemical inhibitors of these targets (inhibitor discovery) have been introduced.

Target discovery

With the rise of drug resistance, there is an urgent need to identify microbial targets for the development of novel drug therapies and vaccines. There is a growing trend to shift from identifying compounds that simply kill a microbe

to identifying compounds that target and inactivate a factor specifically involved in pathogenesis (Brotz-Oesterhelt *et al.*, 2005). The rationale for such a shift in focus is that by targeting non-essential proteins required for pathogenesis but not basic survival the microbe will be less likely to develop resistance to the drug. To this end, there is a need for methods for rapid target identification and characterization.

Activity-based probes can be used to profile the activities of several protein targets in a single experiment without the need to generate reagent quantities of recombinant enzymes. In an example of the use of ABPs to study pathogens, a general cysteine protease probe, DCG-04 (Greenbaum *et al.*, 2002b), was used to profile cysteine protease activities in the human malarial parasite, *P. falciparum.* Initial profiling efforts identified four proteases including falcipain-1, -2, -3 and a cathepsin C-like protease. Falcipain-1 was the first cysteine protease cloned from *P. falciparum*; however, biochemical and functional analyses had been limited due to difficulties in the recombinant expression of this protease (Rosenthal and Nelson, 1992). A falcipain-1-specific inhibitor, YA29-Eps(S,S), was identified by screening a set of epoxide-based positional scanning libraries (PSLs, discussed further below) in crude parasite extracts (Greenbaum *et al.*, 2002b). This compound was then used in parasite culture models to propose a potential role for falcipain-1 in erythrocyte invasion by merozoites, making it an ideal target for future development of novel anti-malarial drug therapies.

Activity-based probes have also been used to characterize the major cysteine proteases of *T. brucei* (Caffrey *et al.*, 2001). An ^{125}I-tagged version of the covalent inhibitor LHVS-PhOH was used to identify rhodesain as the predominant cysteine protease activity in *T. brucei*. Additionally, a series of irreversible cysteine protease inhibitors including peptide vinyl sulphones, chloromethylketones, diazomethylketones and fluoromethylketones were tested against the blood form of *T. brucei in vitro* (Troeberg *et al.*, 1999). Biotinylated ABP versions of those found to be anti-trypanosidal were used to identify the likely protease targets of the respective inhibitors. At least one of these target proteases, trypanopain-Tb, is a potential target for development of novel drug therapies. Additionally, a series of peptide vinyl sulphone ABPs were used to extensively characterize the proteasome of *T. brucei* (Wang *et al.*, 2003). These studies highlighted differences in the substrate and inhibitor specificities of the parasite enzyme relative to the human enzyme that could potentially be used to create novel anti-parasite agents.

Inhibitor discovery

One of the most useful applications for ABPs is in screening large sets of inhibitor libraries against multiple enzyme targets within a crude proteome (Greenbaum *et al.*, 2002a). Screening in crude cell extracts allows for rapid identification of lead compounds for endogenously expressed protease targets with minimal effort and no *a priori* knowledge of the inhibitor targets. PSLs of small molecule inhibitors have been developed in an effort to identify specific inhibitors of the papain family of cysteine proteases (Greenbaum *et al.*, 2002a,c). These PSLs were created by fixing one residue of the inhibitor while varying the other positions as a mixture of all the natural amino acids. Thus, information regarding the overall specificity of each residue could then be combined to generate selective inhibitors. Screening of the libraries was accomplished using an indirect competition assay in which a crude proteome is treated with a mixture library and the potency and selectivity determined by labelling of residual active enzymes with a general ABP (Fig. 5). The ability to inhibit active-site labelling of each target in the proteome is measured as a per cent competition value relative to an untreated control. PSLs have been used to identify inhibitors of rat liver cathepsins (Greenbaum *et al.*, 2002c) and of *P. falciparum* cysteine proteases (Greenbaum *et al.*, 2002b). Furthermore, similar approaches can be used to screen diverse libraries of individual inhibitors (Leung *et al.*, 2003).

Direct competition screening offers several advantages over classical substrate-based kinetic assays. First, it bypasses the need for reagent quantities of recombinant proteins and allows screening of uncharacterized targets in complex proteomes. Furthermore, it allows the analysis of multiple enzymes in a single assay without the need to optimize conditions for each target (Adam *et al.*, 2002a; Greenbaum *et al.*, 2002a; Speers and Cravatt, 2004). In addition, PSLs enable affinity fingerprinting, which can be used to create a database of reference affinity patterns for known enzymes. Such a database can then be used to classify unknown proteases from complex proteomes by comparison of their affinity fingerprints (Greenbaum *et al.*, 2002c).

Chemical genetics: small molecules as tools for functional analysis

For many years the pharmaceutical industry has made use of high-throughput screens for inhibitors in the discovery of novel therapeutic agents for the treatment of disease and infections (Crews and Splittgerber, 1999). Typically, a diverse small molecule library is screened against a single well-characterized target. Recently, this strategy of high-throughput screening has been adapted to the discovery of small molecules that can be used as tools for dissecting protein function, a field often referred to as chemical genetics (Ward *et al.*, 2002). Chemical

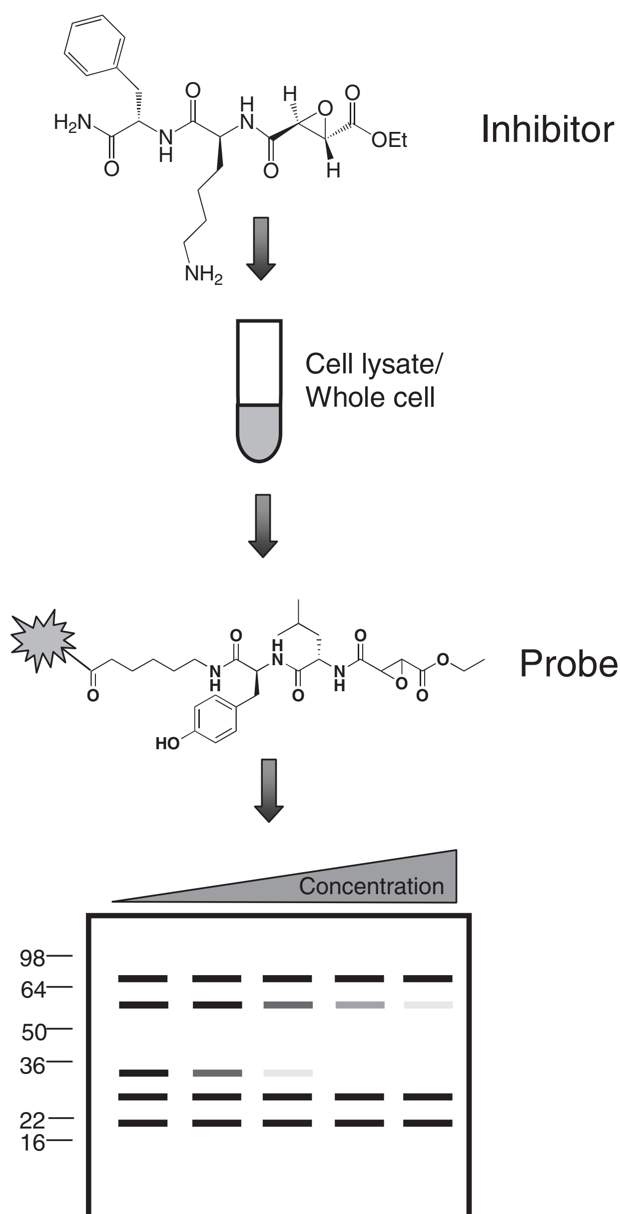


Fig. 5. Application of activity-based probes (ABPs) for indirect competition studies with small molecules. ABPs are useful tools for determining selectivity and potency of small molecule inhibitors. This approach has the advantage of being able to analyse several related proteins in a single gel in their native condition. Cell lysates or whole cells are pre-treated with a range of concentrations of a small molecule inhibitor. The sample is then incubated with a tagged ABP and active proteases are separated by SDS-PAGE. Band intensities are subsequently quantified to determine the degree of inhibition of active-site labelling by the inhibitor. This is measured as a per cent competition with respect to an untreated control. Decreases in band intensities are correlated with more potent inhibition by the small molecule inhibitor. In addition, selectivity of the inhibitor relative to closely related members of the same enzyme family can be determined in the same experiment.

genetics is akin to classical genetics in that there are two predominant approaches: forward and reverse genetic strategies (Fig. 6). Forward chemical genetic (or phenotype-based) screens involve screening small molecule libraries for phenotypic effects in a biological assay of interest followed by subsequent identification of targets responsible for the observed phenotype. In a reverse chemical genetic (or target-based) screen, small molecule libraries are screened against a single protein target to identify selective inhibitors that can be used in pharmacological studies to assign functional roles to the protein of interest (Ward *et al*., 2002; Khersonsky and Chang, 2004).

These chemical genetics strategies offer several advantages over classical genetic and biochemical techniques. First, forward chemical genetic strategies identify compounds capable of perturbing key biological processes without knowledge of protein targets. Additionally, the use of small molecule inhibitors bypasses the need to genetically disrupt (knock down or knock out) genes of interest (Ward *et al*., 2002). The process of gene disruption can often be tedious and time-consuming and results can be difficult to interpret in cases where there are genes that perform similar functions and/or enable the organism to develop alternate survival strategies. Moreover, several organisms such as *P. falciparum* remain difficult to manipulate genetically (de Bruin *et al*., 1992) and essential genes in haploid organisms such as *T. gondii* (Boothroyd *et al*., 1997; Black and Boothroyd, 2000) cannot be studied using gene disruption techniques. To this end, chemical genetics offers an alternate strategy for studying protein function in a highly controlled manner.

This technique has been employed successfully in several mammalian systems to identify compounds that block protein–protein interactions important in cell signalling (Roehrl *et al*., 2004), cytokinesis (Eggert *et al*., 2004), endocytic trafficking (Nieland *et al*., 2004), mitosis and cell migration (Yarrow *et al*., 2003a,b) and DNA binding (Shipps *et al*., 1997). While chemical genetics has found widespread use in mammalian systems, it has only recently been employed in the field of microbiology. The majority of chemical genetic screens performed in microbial systems have focused on identifying small molecules that show general cytotoxic effects on microbes or that inhibit growth (Ward *et al*., 2002). More recently, the first high-throughput screen for small molecules capable of perturbing biological processes of the intracellular protozoan parasite, *T. gondii*, was performed (Carey *et al*., 2004). In this study, a high-throughput microscopy-based assay was developed to screen 12 160 structurally diverse small molecules. This study identified several inhibitors that perturbed various aspects of the invasion process including motility, secretion and conoid extension. These newly identified small molecules are likely to be

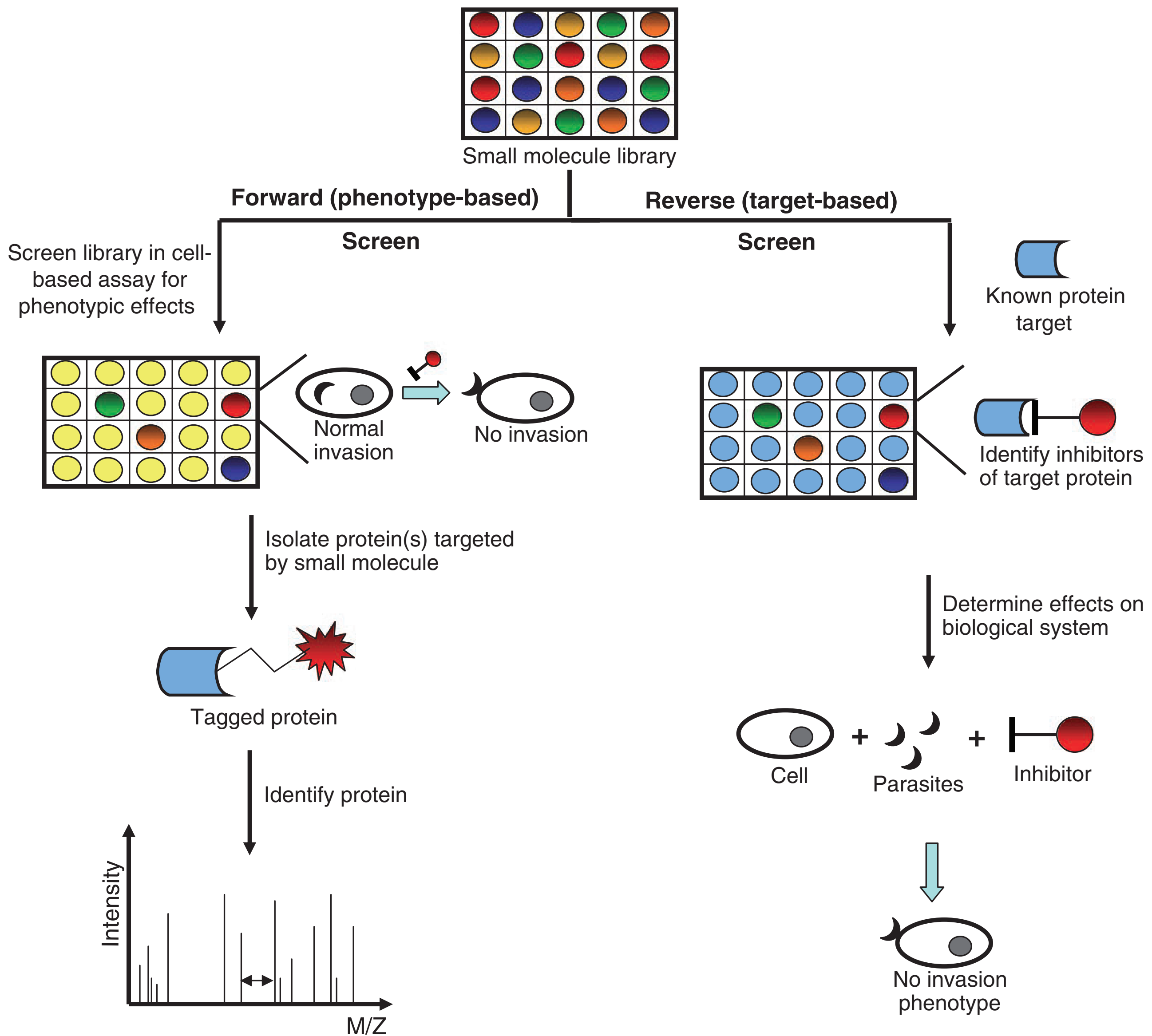


Fig. 6. Chemical genetic screens. This schematic illustrates both forward and reverse chemical genetic screens for inhibitors that perturb invasion of host cells by an obligate intracellular parasite (represented by the half moon shape). Forward (phenotype-based) screen. In a forward screen, a biological process of interest (in this case invasion) is developed into a relatively high-throughput assay that can be performed in the presence of a library of small molecules. For positive hits that block host cell invasion, tagged versions of the inhibitors are synthesized and used to isolate the protein(s) responsible for the phenotype. The isolated proteins are then identified using mass spectrometry-based sequencing techniques. Reverse (target-based) screens. In reverse screens, a protein of interest with unknown function is treated with small molecule libraries to identify compounds with potent inhibitory activity. Positive hits that show high selectivity for the desired target relative to other related enzymes are then used in cell-based assays to determine the effect of target inhibition on the biological system. In this example, treatment of parasites with the inhibitor in the presence of host cells prevents parasite invasion of host cells.

useful tools in dissecting the complex process of host cell invasion.

One of the major limitations of forward chemical genetic screens of small molecules lies in the inability to rapidly identify targets of the small molecules. Most small molecule libraries consist of compounds whose mechanism of action remains unknown. Furthermore, there is often no way to predict the likely target family or class based on the structure of the small molecule showing interesting biological activity. One potential solution to this problem is to use focused libraries of small molecules designed to target a defined enzyme family. Additionally, by focusing this library further to contain only covalent inhibitors it becomes possible to rapidly identify target enzymes by direct affinity isolation methods. With continued advances in these areas, chemical genetics will provide an efficient

means of identifying tools for studying protein function and activity in a highly controlled, temporal manner.

Conclusions and future challenges

The ever increasing number of completed sequences for important human pathogens will lead to a similar rise in demand for new methods to facilitate identification and functional analysis of the gene products. Here we review a myriad of classical proteomic approaches that have been used to make significant headway in mapping and analysing the proteomes of several important pathogenic bacteria and parasites. In addition, the advent of multidimensional chromatography methods and chemical proteomic tools has increased the quality of information obtained by proteomic analysis. The field of chemical proteomics will likely hold a unique niche in the study of pathogenic organisms, in particular in haploid organisms such as the protozoan parasites *T. gondii* and *P. falciparum* as it provides a means of temporally regulating enzyme function thereby allowing the study of essential proteins that cannot be knocked out by standard genetic techniques.

Although the field of chemical proteomics has enjoyed several years of intense technological development, there are still several challenges that lie ahead. ABPs have only been developed for a handful of enzyme classes covering only a small portion of the proteome. Future efforts aimed at developing probes that target a wider range of both enzyme and protein targets will greatly increase the fraction of the proteome accessible to ABPP. This will require the synthesis of more structurally diverse candidate probes. In addition, current methods of identifying enzymes targeted by ABPs suffer from some of the same limitations faced by classical proteomics separation methods. Future efforts must focus on coupling ABPP with gel-free separation and identification technologies such as LC/MS/MS and capillary electrophoresis-MS/MS.

Finally, chemical genetic screens to identify small molecule inhibitors of protein function suffer from a lack of highly robust cell-based assays. Future efforts directed at establishing reproducible assays will greatly advance the field of chemical genetics and aid in the discovery and design of novel, highly selective inhibitors for perturbing biological function of target proteins. In addition, the use of more focused libraries of direct irreversible inhibitors may increase the amount of useful data obtained from a screening experiment. In conclusion, continued advances in both the development of diverse ABPs as well as in the technological aspects of both classical and chemical proteomics will provide a more comprehensive understanding of complex proteomes in the field of microbiology.

Acknowledgements

This work was funded by an NIH National Technology Center for Networks and Pathway Grant U54-RR020843 (to M.B.) and funding from a Searle Scholar Fellowship (to M.B.) as well as funding from an NIH Cellular and Molecular Biology Training Program Grant 5T32GM07276 (to C.I.P.).

References

Adam, G.C., Cravatt, B.F., and Sorensen, E.J. (2001) Profiling the specific reactivity of the proteome with non-directed activity-based probes. *Chem Biol* **8:** 81–95.

Adam, G.C., Sorensen, E.J., and Cravatt, B.F. (2002a) Chemical strategies for functional proteomics. *Mol Cell Proteomics* **1:** 781–790.

Adam, G.C., Sorensen, E.J., and Cravatt, B.F. (2002b) Proteomic profiling of mechanistically distinct enzyme classes using a common chemotype. *Nat Biotechnol* **20:** 805–809.

Anderson, L., and Seilhamer, J. (1997) A comparison of selected mRNA and protein abundances in human liver. *Electrophoresis* **18:** 533–537.

Anderson, N.L., and Anderson, N.G. (1998) Proteome and proteomics: new technologies, new concepts, and new words. *Electrophoresis* **19:** 1853–1861.

Barglow, K.T., and Cravatt, B.F. (2004) Discovering disease-associated enzymes by proteome reactivity profiling. *Chem Biol* **11:** 1523–1531.

Bente, M., Harder, S., Wiesgigl, M., Heukeshoven, J., Gelhaus, C., Krause, E., *et al.* (2003) Developmentally induced changes of the proteome in the protozoan parasite *Leishmania donovani*. *Proteomics* **3:** 1811–1829.

Berger, A.B., Vitorino, P.M., and Bogyo, M. (2004) Activity-based protein profiling: applications to biomarker discovery, *in vivo* imaging and drug discovery. *Am J Pharmacogenomics* **4:** 371–381.

Berggren, K.N., Schulenberg, B., Lopez, M.F., Steinberg, T.H., Bogdanova, A., Smejkal, G., *et al.* (2002) An improved formulation of SYPRO Ruby protein gel stain: comparison with the original formulation and with a ruthenium II tris (bathophenanthroline disulfonate) formulation. *Proteomics* **2:** 486–498.

Bjellqvist, B., Pasquali, C., Ravier, F., Sanchez, J.C., and Hochstrasser, D. (1993) A nonlinear wide-range immobilized pH gradient for two-dimensional electrophoresis and its definition in a relevant pH scale. *Electrophoresis* **14:** 1357–1365.

Black, M.W., and Boothroyd, J.C. (2000) Lytic cycle of *Toxoplasma gondii*. *Microbiol Mol Biol Rev* **64:** 607–623.

Blonder, J., Goshe, M.B., Xiao, W., Camp, D.G., 2nd, Wingerd, M., Davis, R.W., and Smith, R.D. (2004) Global analysis of the membrane subproteome of *Pseudomonas aeruginosa* using liquid chromatography-tandem mass spectrometry. *J Proteome Res* **3:** 434–444.

Bogyo, M., McMaster, J.S., Gaczynska, M., Tortorella, D., Goldberg, A.L., and Ploegh, H. (1997) Covalent modification of the active site threonine of proteasomal beta subunits and the *Escherichia coli* homolog HslV by a new class of inhibitors. *Proc Natl Acad Sci USA* **94:** 6629–6634.

Bogyo, M., Verhelst, S., Bellingard-Dubouchaud, V., Toba, S., and Greenbaum, D. (2000) Selective targeting of lysosomal cysteine proteases with radiolabeled electrophilic substrate analogs. *Chem Biol* **7:** 27–38.

Boothroyd, J.C., Black, M., Bonnefoy, S., Hehl, A., Knoll, L.J., Manger, I.D., *et al.* (1997) Genetic and biochemical analysis of development in *Toxoplasma gondii*. *Philos Trans R Soc Lond B Biol Sci* **352:** 1347–1354.

Borodovsky, A., Ovaa, H., Kolli, N., Gan-Erdene, T., Wilkinson, K.D., Ploegh, H.L., and Kessler, B.M. (2002) Chemistry-based functional proteomics reveals novel members of the deubiquitinating enzyme family. *Chem Biol* **9:** 1149–1159.

Borodovsky, A., Ovaa, H., Meester, W.J., Venanzi, E.S., Bogyo, M.S., Hekking, B.G., *et al.* (2005) Small-molecule inhibitors and probes for ubiquitin- and ubiquitin-like-specific proteases. *Chembiochem* **6:** 287–291.

Bottari, P., Aebersold, R., Turecek, F., and Gelb, M.H. (2004) Design and synthesis of visible isotope-coded affinity tags for the absolute quantification of specific proteins in complex mixtures. *Bioconjug Chem* **15:** 380–388.

Bromley, E., Leeds, N., Clark, J., McGregor, E., Ward, M., Dunn, M.J., and Tomley, F. (2003) Defining the protein repertoire of microneme secretory organelles in the apicomplexan parasite *Eimeria tenella*. *Proteomics* **3:** 1553–1561.

Brotz-Oesterhelt, H., Bandow, J.E., and Labischinski, H. (2005) Bacterial proteomics and its role in antibacterial drug discovery. *Mass Spectrom Rev* **24:** 549–565.

de Bruin, D., Lanzer, M., and Ravetch, J.V. (1992) Characterization of yeast artificial chromosomes from *Plasmodium falciparum*: construction of a stable, representative library and cloning of telomeric DNA fragments. *Genomics* **14:** 332–339.

Bumann, D., Meyer, T.F., and Jungblut, P.R. (2001) Proteome analysis of the common human pathogen *Helicobacter pylori*. *Proteomics* **1:** 473–479.

Caffrey, C.R., Hansell, E., Lucas, K.D., Brinen, L.S., Alvarez Hernandez, A., Cheng, J., *et al.* (2001) Active site mapping, biochemical properties and subcellular localization of rhodesain, the major cysteine protease of *Trypanosoma brucei* rhodesiense. *Mol Biochem Parasitol* **118:** 61–73.

Calvo, E., Pucciarelli, M.G., Bierne, H., Cossart, P., Pablo Albar, J., and Garcia-Del Portillo, F. (2005) Analysis of the *Listeria* cell wall proteome by two-dimensional nanoliquid chromatography coupled to mass spectrometry. *Proteomics* **5:** 433–443.

Carey, K.L., Westwood, N.J., Mitchison, T.J., and Ward, G.E. (2004) A small-molecule approach to studying invasive mechanisms of *Toxoplasma gondii*. *Proc Natl Acad Sci USA* **101:** 7433–7438.

Carruthers, V.B., Sherman, G.D., and Sibley, L.D. (2000) The *Toxoplasma* adhesive protein MIC2 is proteolytically processed at multiple sites by two parasite-derived proteases. *J Biol Chem* **275:** 14346–14353.

Carucci, D.J., Yates, J.R., III, and Florens, L. (2002) Exploring the proteome of *Plasmodium*. *Int J Parasitol* **32:** 1539–1542.

Celestino, P.B., de Carvalho, L.R., de Freitas, L.M., Dorella, F.A., Martins, N.F., Pacheco, L.G., *et al.* (2004) Update of microbial genome programs for bacteria and archaea. *Genet Mol Res* **3:** 421–431.

Chan, E.W., Chattopadhaya, S., Panicker, R.C., Huang, X., and Yao, S.Q. (2004) Developing photoactive affinity probes for proteomic profiling: hydroxamate-based probes for metalloproteases. *J Am Chem Soc* **126:** 14435–14446.

Cho, M.J., Jeon, B.S., Park, J.W., Jung, T.S., Song, J.Y., Lee, W.K., *et al.* (2002) Identifying the major proteome components of *Helicobacter pylori* strain 26695. *Electrophoresis* **23:** 1161–1173.

Cohen, A.M., Rumpel, K., Coombs, G.H., and Wastling, J.M. (2002) Characterisation of global protein expression by two-dimensional electrophoresis and mass spectrometry: proteomics of *Toxoplasma gondii*. *Int J Parasitol* **32:** 39–51.

Cordwell, S.J., Nouwens, A.S., and Walsh, B.J. (2001) Comparative proteomics of bacterial pathogens. *Proteomics* **1:** 461–472.

Corthals, G.L., Wasinger, V.C., Hochstrasser, D.F., and Sanchez, J.C. (2000) The dynamic range of protein expression: a challenge for proteomic research. *Electrophoresis* **21:** 1104–1115.

Cravatt, B.F., and Sorensen, E.J. (2000) Chemical strategies for the global analysis of protein function. *Curr Opin Chem Biol* **4:** 663–668.

Crews, C.M., and Splittgerber, U. (1999) Chemical genetics: exploring and controlling cellular processes with chemical probes. *Trends Biochem Sci* **24:** 317–320.

Curwen, R.S., Ashton, P.D., Johnston, D.A., and Wilson, R.A. (2004) The *Schistosoma mansoni* soluble proteome: a comparison across four life-cycle stages. *Mol Biochem Parasitol* **138:** 57–66.

van Deursen, F.J., Thornton, D.J., and Matthews, K.R. (2003) A reproducible protocol for analysis of the proteome of *Trypanosoma brucei* by 2-dimensional gel electrophoresis. *Mol Biochem Parasitol* **128:** 107–110.

Drummelsmith, J., Brochu, V., Girard, I., Messier, N., and Ouellette, M. (2003) Proteome mapping of the protozoan parasite leishmania and application to the study of drug targets and resistance mechanisms. *Mol Cell Proteomics* **2:** 146–155.

Dubremetz, J.F., Garcia-Reguet, N., Conseil, V., and Fourmaux, M.N. (1998) Apical organelles and host-cell invasion by Apicomplexa. *Int J Parasitol* **28:** 1007–1013.

Eggert, U.S., Kiger, A.A., Richter, C., Perlman, Z.E., Perrimon, N., Mitchison, T.J., and Field, C.M. (2004) Parallel chemical genetic and genome-wide RNAi screens identify cytokinesis inhibitors and targets. *PLoS Biol* **2:** e379.

El Fakhry, Y., Ouellette, M., and Papadopoulou, B. (2002) A proteomic approach to identify developmentally regulated proteins in *Leishmania infantum*. *Proteomics* **2:** 1007–1017.

Evers, S., Di Padova, K., Meyer, M., Langen, H., Fountoulakis, M., Keck, W., and Gray, C.P. (2001) Mechanism-related changes in the gene transcription and protein synthesis patterns of *Haemophilus influenzae* after treatment with transcriptional and translational inhibitors. *Proteomics* **1:** 522–544.

Faleiro, L., Kobayashi, R., Fearnhead, H., and Lazebnik, Y. (1997) Multiple species of CPP32 and Mch2 are the major active caspases present in apoptotic cells. *EMBO J* **16:** 2271–2281.

Florens, L., Washburn, M.P., Raine, J.D., Anthony, R.M., Grainger, M., Haynes, J.D., *et al.* (2002) A proteomic view of the *Plasmodium falciparum* life cycle. *Nature* **419:** 520–526.

Florens, L., Liu, X., Wang, Y., Yang, S., Schwartz, O., Peglar, M., *et al.* (2004) Proteomics approach reveals novel proteins on the surface of malaria-infected erythrocytes. *Mol Biochem Parasitol* **135:** 1–11.

Gardner, M.J., Hall, N., Fung, E., White, O., Berriman, M., Hyman, R.W., *et al.* (2002) Genome sequence of the

human malaria parasite *Plasmodium falciparum*. *Nature* **419:** 498–511.

Gharbi, S., Gaffney, P., Yang, A., Zvelebil, M.J., Cramer, R., Waterfield, M.D., and Timms, J.F. (2002) Evaluation of two-dimensional differential gel electrophoresis for proteomic expression analysis of a model breast cancer cell system. *Mol Cell Proteomics* **1:** 91–98.

Gmuender, H., Kuratli, K., Di Padova, K., Gray, C.P., Keck, W., and Evers, S. (2001) Gene expression changes triggered by exposure of *Haemophilus influenzae* to novobiocin or ciprofloxacin: combined transcription and translation analysis. *Genome Res* **11:** 28–42.

Gongora, R., Acestor, N., Quadroni, M., Fasel, N., Saravia, N.G., and Walker, J. (2003) Mapping the proteome of Leishmania Viannia parasites using two-dimensional polyacrylamide gel electrophoresis and associated technologies. *Biomedica* **23:** 153–160.

Graves, P.R., and Haystead, T.A. (2002) Molecular biologist's guide to proteomics. *Microbiol Mol Biol Rev* **66:** 39–63; table of contents.

Greenbaum, D., Medzihradszky, K.F., Burlingame, A., and Bogyo, M. (2000) Epoxide electrophiles as activity-dependent cysteine protease profiling and discovery tools. *Chem Biol* **7:** 569–581.

Greenbaum, D., Baruch, A., Hayrapetian, L., Darula, Z., Burlingame, A., Medzihradszky, K.F., and Bogyo, M. (2002a) Chemical approaches for functionally probing the proteome. *Mol Cell Proteomics* **1:** 60–68.

Greenbaum, D.C., Baruch, A., Grainger, M., Bozdech, Z., Medzihradszky, K.F., Engel, J., *et al.* (2002b) A role for the protease falcipain 1 in host cell invasion by the human malaria parasite. *Science* **298:** 2002–2006.

Greenbaum, D.C., Arnold, W.D., Lu, F., Hayrapetian, L., Baruch, A., Krumrine, J., *et al.* (2002c) Small molecule affinity fingerprinting. A tool for enzyme family subclassification, target identification, and inhibitor design. *Chem Biol* **9:** 1085–1094.

Guina, T., Wu, M., Miller, S.I., Purvine, S.O., Yi, E.C., Eng, J., *et al.* (2003) Proteomic analysis of *Pseudomonas aeruginosa* grown under magnesium limitation. *J Am Soc Mass Spectrom* **14:** 742–751.

Gygi, S.P., Rochon, Y., Franza, B.R., and Aebersold, R. (1999a) Correlation between protein and mRNA abundance in yeast. *Mol Cell Biol* **19:** 1720–1730.

Gygi, S.P., Rist, B., Gerber, S.A., Turecek, F., Gelb, M.H., and Aebersold, R. (1999b) Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat Biotechnol* **17:** 994–999.

Gygi, S.P., Corthals, G.L., Zhang, Y., Rochon, Y., and Aebersold, R. (2000a) Evaluation of two-dimensional gel electrophoresis-based proteome analysis technology. *Proc Natl Acad Sci USA* **97:** 9390–9395.

Gygi, S.P., Rist, B., and Aebersold, R. (2000b) Measuring gene expression by quantitative proteome analysis. *Curr Opin Biotechnol* **11:** 396–401.

Gygi, S.P., Rist, B., Griffin, T.J., Eng, J., and Aebersold, R. (2002) Proteome analysis of low-abundance proteins using multidimensional chromatography and isotope-coded affinity tags. *J Proteome Res* **1:** 47–54.

Hamdan, M., and Righetti, P.G. (2002) Modern strategies for protein quantification in proteome analysis: advantages and limitations. *Mass Spectrom Rev* **21:** 287–302.

Hansen, K.C., Schmitt-Ulms, G., Chalkley, R.J., Hirsch, J., Baldwin, M.A., and Burlingame, A.L. (2003) Mass spectrometric analysis of protein mixtures at low levels using cleavable 13C-Isotope-coded affinity tag and multidimensional chromatography. *Mol Cell Proteomics* **2:** 299–314.

Hemelaar, J., Galardy, P.J., Borodovsky, A., Kessler, B.M., Ploegh, H.L., and Ovaa, H. (2004) Chemistry-based functional proteomics: mechanism-based activity-profiling tools for ubiquitin and ubiquitin-like specific proteases. *J Proteome Res* **3:** 268–276.

Jefferies, J.R., Campbell, A.M., van Rossum, A.J., Barrett, J., and Brophy, P.M. (2001) Proteomic analysis of *Fasciola hepatica* excretory-secretory products. *Proteomics* **1:** 1128–1132.

Jeffery, D.A., and Bogyo, M. (2003) Chemical proteomics and its application to drug discovery. *Curr Opin Biotechnol* **14:** 87–95.

Kan, B., Habibi, H., Schmid, M., Liang, W., Wang, R., Wang, D., and Jungblut, P.R. (2004) Proteome comparison of *Vibrio cholerae* cultured in aerobic and anaerobic conditions. *Proteomics* **4:** 3061–3067.

Kessler, B.M., Tortorella, D., Altun, M., Kisselev, A.F., Fiebiger, E., Hekking, B.G., *et al.* (2001) Extended peptide-based inhibitors efficiently target the proteasome and reveal overlapping specificities of the catalytic beta-subunits. *Chem Biol* **8:** 913–929.

Khersonsky, S.M., and Chang, Y.T. (2004) Strategies for facilitated forward chemical genetics. *Chembiochem* **5:** 903–908.

Kidd, D., Liu, Y., and Cravatt, B.F. (2001) Profiling serine hydrolase activities in complex proteomes. *Biochemistry* **40:** 4005–4015.

Langen, H., Takacs, B., Evers, S., Berndt, P., Lahm, H.W., Wipf, B., *et al.* (2000) Two-dimensional map of the proteome of *Haemophilus influenzae*. *Electrophoresis* **21:** 411–429.

Lasonder, E., Ishihama, Y., Andersen, J.S., Vermunt, A.M., Pain, A., Sauerwein, R.W., *et al.* (2002) Analysis of the *Plasmodium falciparum* proteome by high-accuracy mass spectrometry. *Nature* **419:** 537–542.

Leung, D., Hardouin, C., Boger, D.L., and Cravatt, B.F. (2003) Discovering potent and selective reversible inhibitors of enzymes in complex proteomes. *Nat Biotechnol* **21:** 687–691.

Li, J., Steen, H., and Gygi, S.P. (2003) Protein profiling with cleavable isotope-coded affinity tag (cICAT) reagents: the yeast salinity stress response. *Mol Cell Proteomics* **2:** 1198–1204.

Liu, Y., Patricelli, M.P., and Cravatt, B.F. (1999) Activity-based protein profiling: the serine hydrolases. *Proc Natl Acad Sci USA* **96:** 14694–14699.

Lo, L.C., Pang, T.L., Kuo, C.H., Chiang, Y.L., Wang, H.Y., and Lin, J.J. (2002) Design and synthesis of class-selective activity probes for protein tyrosine phosphatases. *J Proteome Res* **1:** 35–40.

Lopez, M.F., Berggren, K., Chernokalskaya, E., Lazarev, A., Robinson, M., and Patton, W.F. (2000) A comparison of silver stain and SYPRO Ruby protein gel stain with respect to protein detection in two-dimensional gels and identification by peptide mass profiling. *Electrophoresis* **21:** 3673–3683.

Lu, Y., Bottari, P., Turecek, F., Aebersold, R., and Gelb, M.H. (2004) Absolute quantification of specific proteins in complex mixtures using visible isotope-coded affinity tags. *Anal Chem* **76:** 4104–4111.

McAtee, C.P., Hoffman, P.S., and Berg, D.E. (2001) Identifi-

cation of differentially regulated proteins in metronidozole resistant *Helicobacter pylori* by proteome techniques. *Proteomics* **1:** 516–521.
McCormack, A.L., Schieltz, D.M., Goode, B., Yang, S., Barnes, G., Drubin, D., and Yates, J.R., III (1997) Direct analysis and identification of proteins in mixtures by LC/MS/MS and database searching at the low-femtomole level. *Anal Chem* **69:** 767–776.
Moseley, M.A. (2001) Current trends in differential expression proteomics: isotopically coded tags. *Trends Biotechnol* **19:** S10–S16.
Mukhtar, T.A., and Wright, G.D. (2005) Streptogramins, oxazolidinones, and other inhibitors of bacterial protein synthesis. *Chem Rev* **105:** 529–542.
Nazif, T., and Bogyo, M. (2001) Global analysis of proteasomal substrate specificity using positional-scanning libraries of covalent inhibitors. *Proc Natl Acad Sci USA* **98:** 2967–2972.
Nieland, T.J., Feng, Y., Brown, J.X., Chuang, T.D., Buckett, P.D., Wang, J., *et al.* (2004) Chemical genetic screening identifies sulfonamides that raise organellar pH and interfere with membrane traffic. *Traffic* **5:** 478–492.
Nirmalan, N., Sims, P.F., and Hyde, J.E. (2004) Quantitative proteomics of the human malaria parasite *Plasmodium falciparum* and its application to studies of development and inhibition. *Mol Microbiol* **52:** 1187–1199.
Nugent, P.G., Karsani, S.A., Wait, R., Tempero, J., and Smith, D.F. (2004) Proteomic analysis of *Leishmania mexicana* differentiation. *Mol Biochem Parasitol* **136:** 51–62.
O'Farrell, P.H. (1975) High resolution two-dimensional electrophoresis of proteins. *J Biol Chem* **250:** 4007–4021.
Ovaa, H., Kessler, B.M., Rolen, U., Galardy, P.J., Ploegh, H.L., and Masucci, M.G. (2004) Activity-based ubiquitin-specific protease (USP) profiling of virus-infected and malignant human cells. *Proc Natl Acad Sci USA* **101:** 2253–2258.
Paba, J., Ricart, C.A., Fontes, W., Santana, J.M., Teixeira, A.R., Marchese, J., *et al.* (2004b) Proteomic analysis of *Trypanosoma cruzi* developmental stages using isotope-coded affinity tag reagents. *J Proteome Res* **3:** 517–524.
Paba, J., Santana, J.M., Teixeira, A.R., Fontes, W., Sousa, M.V., and Ricart, C.A. (2004a) Proteomic analysis of the human pathogen *Trypanosoma cruzi*. *Proteomics* **4:** 1052–1059.
Pandey, A., and Mann, M. (2000) Proteomics to study genes and genomes. *Nature* **405:** 837–846.
Parodi-Talice, A., Duran, R., Arrambide, N., Prieto, V., Pineyro, M.D., Pritsch, O., *et al.* (2004) Proteome analysis of the causative agent of Chagas disease: *Trypanosoma cruzi*. *Int J Parasitol* **34:** 881–886.
Patel, A., and Perrin, D.M. (2004) Efficient synthesis of isotopically pure isotope-coded affinity tagging reagents. *Bioconjug Chem* **15:** 224–230.
Roehrl, M.H., Wang, J.Y., and Wagner, G. (2004) Discovery of small-molecule inhibitors of the NFAT–calcineurin interaction by competitive high-throughput fluorescence polarization screening. *Biochemistry* **43:** 16067–16075.
Rosenthal, P.J., and Nelson, R.G. (1992) Isolation and characterization of a cysteine proteinase gene of *Plasmodium falciparum*. *Mol Biochem Parasitol* **51:** 143–152.
Rout, M.P., and Field, M.C. (2001) Isolation and characterization of subnuclear compartments from *Trypanosoma brucei*. Identification of a major repetitive nuclear lamina component. *J Biol Chem* **276:** 38261–38271.
Saghatelian, A., Jessani, N., Joseph, A., Humphrey, M., and Cravatt, B.F. (2004) Activity-based probes for the proteomic profiling of metalloproteases. *Proc Natl Acad Sci USA* **101:** 10000–10005.
Sam-Yellowe, T.Y., Florens, L., Wang, T., Raine, J.D., Carucci, D.J., Sinden, R., and Yates, J.R., III (2004) Proteome analysis of rhoptry-enriched fractions isolated from *Plasmodium* merozoites. *J Proteome Res* **3:** 995–1001.
Schmidt, F., Donahoe, S., Hagens, K., Mattow, J., Schaible, U.E., Kaufmann, S.H., *et al.* (2004) Complementary analysis of the *Mycobacterium tuberculosis* proteome by two-dimensional electrophoresis and isotope-coded affinity tag technology. *Mol Cell Proteomics* **3:** 24–42.
Shalon, D., Smith, S.J., and Brown, P.O. (1996) A DNA microarray system for analyzing complex DNA samples using two-color fluorescent probe hybridization. *Genome Res* **6:** 639–645.
Shipps, G.W., Jr, Pryor, K.E., Xian, J., Skyler, D.A., Davidson, E.H., and Rebek, J., Jr (1997) Synthesis and screening of small molecule libraries active in binding to DNA. *Proc Natl Acad Sci USA* **94:** 11833–11838.
Singh, V.K., Jayaswal, R.K., and Wilkinson, B.J. (2001) Cell wall-active antibiotic induced proteins of *Staphylococcus aureus* identified using a proteomic approach. *FEMS Microbiol Lett* **199:** 79–84.
Smolka, M.B., Zhou, H., Purkayastha, S., and Aebersold, R. (2001) Optimization of the isotope-coded affinity tag-labeling procedure for quantitative proteome analysis. *Anal Biochem* **297:** 25–31.
Smolka, M., Zhou, H., and Aebersold, R. (2002) Quantitative protein profiling using two-dimensional gel electrophoresis, isotope-coded affinity tag labeling, and mass spectrometry. *Mol Cell Proteomics* **1:** 19–29.
Speers, A.E., and Cravatt, B.F. (2004) Chemical strategies for activity-based proteomics. *Chembiochem* **5:** 41–47.
van Swieten, P.F., Maehr, R., van den Nieuwendijk, A.M., Kessler, B.M., Reich, M., Wong, C.S., *et al.* (2004) Development of an isotope-coded activity-based probe for the quantitative profiling of cysteine proteases. *Bioorg Med Chem Lett* **14:** 3131–3134.
Thornberry, N.A., Peterson, E.P., Zhao, J.J., Howard, A.D., Griffin, P.R., and Chapman, K.T. (1994) Inactivation of interleukin-1 beta converting enzyme by peptide (acyloxy) methyl ketones. *Biochemistry* **33:** 3934–3940.
Tjalsma, H., Antelmann, H., Jongbloed, J.D., Braun, P.G., Darmon, E., Dorenbos, R., *et al.* (2004) Proteomics of protein secretion by *Bacillus subtilis*: separating the 'secrets' of the secretome. *Microbiol Mol Biol Rev* **68:** 207–233.
Tonge, R., Shaw, J., Middleton, B., Rowlinson, R., Rayner, S., Young, J., *et al.* (2001) Validation and development of fluorescence two-dimensional differential gel electrophoresis proteomics technology. *Proteomics* **1:** 377–396.
Troeberg, L., Morty, R.E., Pike, R.N., Lonsdale-Eccles, J.D., Palmer, J.T., McKerrow, J.H., and Coetzer, T.H. (1999) Cysteine proteinase inhibitors kill cultured bloodstream forms of *Trypanosoma brucei brucei*. *Exp Parasitol* **91:** 349–355.
Vandahl, B.B., Birkelund, S., Demol, H., Hoorelbeke, B., Christiansen, G., Vandekerckhove, J., and Gevaert, K. (2001) Proteome analysis of the *Chlamydia pneumoniae* elementary body. *Electrophoresis* **22:** 1204–1223.
Wang, C.C., Bozdech, Z., Liu, C.L., Shipway, A., Backes, B.J., Harris, J.L., and Bogyo, M. (2003) Biochemical anal-

ysis of the 20 S proteasome of *Trypanosoma brucei*. *J Biol Chem* **278:** 15800–15808.

Ward, G.E., Carey, K.L., and Westwood, N.J. (2002) Using small molecules to study big questions in cellular microbiology. *Cell Microbiol* **4:** 471–482.

Washburn, M.P., Wolters, D., and Yates, J.R., III (2001) Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat Biotechnol* **19:** 242–247.

Winssinger, N., Ficarro, S., Schultz, P.G., and Harris, J.L. (2002) Profiling protein function with small molecule microarrays. *Proc Natl Acad Sci USA* **99:** 11139–11144.

Wolters, D.A., Washburn, M.P., and Yates, J.R., III (2001) An automated multidimensional protein identification technology for shotgun proteomics. *Anal Chem* **73:** 5683–5690.

Yarrow, J.C., Lechler, T., Li, R., and Mitchison, T.J. (2003a) Rapid de-localization of actin leading edge components with BDM treatment. *BMC Cell Biol* **4:** 5.

Yarrow, J.C., Feng, Y., Perlman, Z.E., Kirchhausen, T., and Mitchison, T.J. (2003b) Phenotypic screening of small molecule libraries by high throughput cell imaging. *Comb Chem High Throughput Screen* **6:** 279–286.

Yi, E.C., Li, X.J., Cooke, K., Lee, H., Raught, B., Page, A., *et al.* (2005) Increased quantitative proteome coverage with (13)C/(12)C-based, acid-cleavable isotope-coded affinity tag reagent and modified data acquisition scheme. *Proteomics* **5:** 380–387.

Zhou, X.W., Blackman, M.J., Howell, S.A., and Carruthers, V.B. (2004) Proteomic analysis of cleavage events reveals a dynamic two-step mechanism for proteolysis of a key parasite adhesive complex. *Mol Cell Proteomics* **3:** 565–576.