Mass Spectrometry-based Parasitic Proteomics

Onrapak Reamtong

Department of Molecular Tropical Medicine and Genetics, Faculty of Tropical Medicine, Mahidol University, 420/6 Ratchawithi Road, Bangkok 10400, Thailand

Abstract

Proteomics is a technique that can generate profiles of whole proteins in a biological sample. Proteomics studies yield protein identification, post-translational modifications, as well as abundance information. In the general workflow, proteins from a biological sample are digested by protease, and separated by liquid chromatography; mass spectrometry is then used to determine their mass-to-charge ratio. The mass-spectrum pattern shows amino-acid sequences and post-translational modifications of peptide/protein ions. Data obtained from mass-spectrometric analysis are interpreted using bioinformatics. Proteomics is useful in several research areas, including parasitology, where it can potentially elucidate the pathology of parasitic illness. This paper reviews the principles of proteomics, as well as its key instruments and research applications in parasitology, including host-parasite interactions, vaccine development, drug-target identification, and diagnosis of parasitic diseases.

Keywords: proteomics, mass spectrometry, parasitic disease

Introduction

The core function of molecular biology is to describe the flow from genetics to the actual phenotypic appearance of living organisms [1]. DNA is genetic material that stores a vast amount of biological information. Only a limited number of viruses carry RNA genetic material. DNA can be transcribed to RNA, followed by protein synthesis from the mRNA template. While the entire human genome has been elucidated [2,3], the genome and relative mRNA are not sufficient for a complete understanding of the biological functions of organisms. To function, genes must be expressed to demonstrate their phenotypic appearance.

Within the genome, some genes have never been transcribed into mRNA, thus gene copy numbers do not always relate to the amount of RNA in cells. Some mRNA molecules have never been translated into proteins, and have different turnover rates. Accordingly, the amount of mRNA is not regularly related to the protein. The study of proteins has thus become an issue of great interest. This review provides an overview of parasitology-related protein research and explores the usefulness of this technology for expanding parasite knowledge.

The proteome is the entire set of proteins expressed by a genome. It is much more complex than a genome and a transcriptome, mainly due to its post-translational modifications. Proteomics is not static and it responds to the surrounding environment. The major goal of proteomics is to understand the protein complement of cells, including identification, modification,
Mass Spectrometry-based Parasitic Proteomics

quantification and localization of proteins. Several methods are currently available to achieve these goals. The state-of-the-art technique being used as part of current proteomic strategies is mass spectrometry (MS) [4]. Mass spectrometry offers higher throughput, and it remains the most sensitive technique for analysis of the proteome, replacing previous methods such as Edman degradation. Using general strategies of MS-based proteomics, proteins or complex protein mixtures are subjected to proteolytic cleavage, usually with trypsin, and the peptide products are then analyzed by MS, as shown in Fig 1.

In general, preparation of proteins/peptides for mass spectrometric analysis is one of the most crucial processes in proteomics research [5]. Employing the correct experimental model and careful sample preparation are both important in obtaining reliable results, and sample separation is a key step in reducing the complexity of the sample prior to mass spectrometric analysis. Common separation techniques are electrophoresis and liquid chromatography. Due to the complexity of the proteome, more than one separation technique is typically needed to increase the dynamic range of peptide or protein detection.

Two-dimensional gel electrophoresis (2D-gel) is a classical protein separation technique for proteomics. It separates proteins according to their isoelectric points (pl) in the first dimension, and according to their molecular weights in the second dimension. Separated spots are excised, in-gel digestion is performed, and analysis by mass spectrometry then takes place. This protocol provides information relating to protein identification and post-translational modifications. To obtain quantification information, two sets of 2D-gels are compared according to their intensities. Recently, 2-D fluorescence difference gel electrophoresis (2D-DIGE) has been developed to allow quantification of up to three separate samples in a single gel [6]. Samples are labeled pre-electrophoretically with fluorescent Cy dyes (Cy2, Cy3 and Cy5). Each protein profile pattern can be compared and indicates differences in protein expression. Limitations of 2D-gel for protein detection concern its reproducibility, poor representation of low-abundance proteins, highly acidic/basic proteins, proteins with extreme size or hydrophobicity, and co-migration of multiple proteins in a single band [7]. A huge number of proteins of interest thus cannot be identified and quantified using 2D-gel techniques.

Developments in the use of mass spectrometric techniques over the last few decades have led to high speed, high sensitivity, high mass accuracy, and high resolution instrumentation, in turn enhancing the application of nano-liquid chromatography coupled with mass spectrometry. To increase the dynamic range of detection, two-dimensional liquid chromatography (2D-LC) is introduced to separate proteins or peptides before tandem mass spectrometry is used. Generally, 2D-LC is a combination of strong cation exchange chromatography and reverse phase HPLC separation. This technique is a non-gel based approach and the results show a high number of protein identifications, including low-abundance proteins [8]. In addition, 2D-LC provides protein identification, post-translational modification, and quantification information, in a single run. 2D-LC coupled with tandem mass spectrometry

![Fig 1 Overview of MS-based proteomics.](image-url)
demonstrates the power of separation and detection, and remains less time-consuming also. It is thus becoming a valuable experimental technique in contemporary protein research.

Mass spectrometry

Mass spectrometry stands as the most useful analytical technique, and the most comprehensive tool in large-scale proteomics [9]. A mass spectrometer consists of three main components: ionization source, mass analyzer, and detector. In general, the ionization source converts sample molecules into gas-phase ions, which are then separated by mass analyzer. In the mass analyzer, ions are influenced by electrical and/or magnetic fields, in order to separate them according to their mass-to-charge ratio (m/z). Separated ions are then detected and recorded as a mass spectrum. A mass spectrum displays both m/z values and ion signal intensities.

In order to analyze peptides and proteins, the analyte molecules require ‘soft’ ionization techniques, such as electrospray ionization (ESI) or matrix-assisted laser desorption/ionization (MALDI). Soft ionization techniques generate intact molecular peptide and protein ion species [10,11]. Different types of mass analyzers are available for separating peptide/protein ions in the gas phase, including quadrupole mass filter [12], time-of-flight [13], and quadrupole ion trap [12] instruments. To further improve quality of peptide/protein detection, tandem mass spectrometry (MS/MS) was developed. This involves multiple steps of mass spectrometry, including ion selection and fragmentation, between the different stages of analysis. In order to perform peptide fragmentation, collision-induced dissociation (CID) is commonly used to fragment ions in the gas phase. Collisions between peptide ions and inert gas result in dissociation of the peptide backbone, producing fragmented ions. Fragmentation patterns are used to determine the peptide sequence [14].

Tandem mass spectrometry currently enables analysis of protein identification, protein quantification, and post-translational modification identification, for example, phosphorylation [15] and acetylation [16]. Mass spectrometry provides significant information and allows for a better understanding of the integrative function of molecular pathways. To identify proteins, all ions from a mass spectrum are measured as calculated masses, and according to their mass-to-charge ratios. An experimental mass list is thus able to be generated. A single LC run may contain more than a thousand mass spectra, producing a huge number of mass lists. A central database containing known protein sequences, or a translated genome, is theoretically feasible, to digest all proteins into peptides and so that the absolute masses of each peptide can be calculated. Mass lists obtained from the available database are called ‘theoretical mass lists’. For identification, an experimental mass list is compared to the theoretical mass lists of known peptides. Identical mass lists are identified as known peptides from the database. Such identification is analyzed statistically to determine the best match.

Applications in parasitology research

Parasites are organisms that survive by having close relationships with other organisms. Normally, parasites are smaller and reproduce at a faster rate than their hosts. Some parasites cause disease directly. Disease-causing parasites are classified into three main categories: protozoa, helminths and ectoparasites. A protozoa is a unicellular eukaryotic organism. Many human diseases are caused by protozoal infections, such as malaria (*Plasmodium* spp), amoebiasis (*Entamoeba histolytica*), giardiasis (*Giardia lamblia*), toxoplasmosis (*Toxoplasma gondii*), cryptosporidiosis (*Cryptosporidium* spp), trichomoniasis (*Trichomonas vaginalis*), Chagas disease (*Trypanosoma cruzi*), leishmaniasis (*Leishmania* spp) and sleeping sickness (*Trypanosoma brucei*). Helminths are worm-like organisms, *eg* cestodes, nematodes and trematodes, which normally live in the digestive tract. Ectoparasites usually live on the surface of the host, *eg* mites. Parasitic disease problems include inefficient control, lack of vaccine, and difficult diagnosis and treatment. All of these
problems lead to considerable death and economic loss in both developing and developed countries. In this review, we discuss the role of proteomics in the investigation of host-parasite interactions, disease diagnosis, antiparasitic drug targets, and vaccine development for parasitic disease. The discipline of proteomics may improve the control of parasitic illness as well as overall public health.

**Diagnosis**

Common methods for the diagnosis of parasitic diseases are based on old-fashioned technologies like microscopy. The microscopic detection of parasites is performed on different types of specimen, e.g., blood smears, tissue, and feces. However, sample preparation is both time-consuming and laborious work. Moreover, efficiency of diagnosis depends on the abilities of laboratory technicians. After the recent emergence of proteomics-based approaches, biomarker discovery has continued to develop and is now being applied in the diagnosis of parasitic diseases. Protein biomarkers reveal the existence and biological state of a particular organism.

Chagas disease is a consequence of *T. cruzi* infection; this protozoan is transmitted by triatomid bugs. Serologic screen testing is currently used as a standard diagnostic method. However, up to 50% of true Chagas disease cases are missed by routine screening [17]. Thus, the proteomics approach was introduced to improve diagnostic accuracy for this illness. Serum protein profiles of patients are compared to those of uninfected controls. After the recent emergence of proteomics-based approaches, biomarker discovery has continued to develop and is now being applied in the diagnosis of parasitic diseases. Protein biomarkers reveal the existence and biological state of a particular organism.

Chagas disease is a consequence of *T. cruzi* infection; this protozoan is transmitted by triatomid bugs. Serologic screen testing is currently used as a standard diagnostic method. However, up to 50% of true Chagas disease cases are missed by routine screening [17]. Thus, the proteomics approach was introduced to improve diagnostic accuracy for this illness. Serum protein profiles of patients are compared to those of uninfected controls, using protein array mass spectrometry. As a result, several candidate biomarkers have been identified in Chagas disease sera, e.g., MIP1 alpha, C3a anaphylatoxin, truncated forms of fibronectin, apolipoprotein A1, and C3. Antipeptide of fibronectin demonstrated 90% specificity for Chagas disease by Western blot analysis [18]. This result provided valuable information in the development of a rapid diagnosis test for Chagas disease.

Infection by protozoa of the genus *Plasmodium* causes human malaria. *P. falciparum* is the most virulent species. The standard method to diagnose malaria infection is visualization of the parasite in blood smear under light microscope. Commercial rapid diagnostic test kits are available, such as histidine-rich protein 2, lactate dehydrogenase, and aldolase-based test [19]. However, additional markers are required in order to improve accuracy and effectiveness. Proteomic strategies were applied to analyze the plasma proteome of severe malaria, mild malaria, and convalescent controls in order to discover new candidate biomarkers. Protein concentration, hypoxanthine phosphoribosyltransferase (pHRPT), phosphoglycerate mutase (pPGM), lactate dehydrogenase (pLDH), and fructose-bisphosphate aldolase (pFBPA) were found to be greater in patients with *P. falciparum* than in controls. On the back of these findings, pHRPT has been introduced as a new biomarker for acute *P. falciparum* infection [20].

*Opisthorchis viverrini*, common name liver fluke, is a trematode parasite; it infects humans causing opisthorchiasis. Opisthorchiasis remains a major health problem in Southeast Asia, including Thailand. Diagnosis of *O. viverrini* infection is difficult due to absence of clinical symptoms. Generally, infection is detected only after chronic infection using ultrasound imaging [21]. To improve diagnosis, biomarkers were investigated using proteomics technology. The plasma protein profile of *O. viverrini*-infected hamsters was separated using sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE). Gel bands from SDS-PAGE were subjected to tryptic digestion, then analyzed by LC-MS/MS. Fibronectin significantly increased after 1 month of infection, and displayed a tendency to increase further. In addition, fibronectin increased significantly in the plasma of *O. viverrini*-infected patients as compared to healthy subjects, using Western blot analysis. Therefore, fibronectin may be a novel diagnostic marker we can use in the near future [22]. Mass spectrometry-based proteomics facilitate the discovery of biomarkers for parasitic disease diagnosis. Biomarkers can be either proteins of the parasite itself, or host proteins responding to infection.
Vaccine development

Vaccines are still unavailable for several parasitic diseases, and those which are available are not effective among all groups of patients. Immunoproteomics is key for discovering new antigens in vaccine development. This technique combines proteomics with the immunoreaction in order to identify potential vaccine candidates. Generally, protein lysate of a parasite is separated by 2D-gel electrophoresis, then blotted onto a membrane. The 2D-blot is probed with infected or immunized host serum. This technique enables the discovery of novel proteins involving stimulation of the host immune system. Schistosomiasis is caused by infection of trematodes belonging to the genus *Schistosoma*. Schistosomiasis remains a significant public health problem in tropical areas due to ineffective transmission control [23]. To improve transmission control efficiency, development of a vaccine against schistosomiasis is required in order to assist current chemotherapy strategies. The immunoproteomics approach has been applied to identify those *Schistosoma japonicum* antigens that are recognized by IgE and IgG antibodies, using plasma from chronically infected patients. Reactive 2D-spots were identified by LC-MS/MS, revealing 18 antigens. Zinc finger and RanBP2-type domain-containing protein are examples of interacting proteins with IgE and IgG4. In addition, ubiquitin-conjugating enzyme and cytosolic II 5'-nucleotidase were recognized by IgG1, IgG3 and IgG4. This study demonstrates the potential of immunoproteomics in identifying major immunoreactive antigens of *S. japonicum*, yet further information about novel antigens is required to continue *S. japonicum* vaccine development [24].

Infection by *L. donovani* and *L. infantum chagasi* causes visceral leishmaniasis – the most severe form of the disease. Indeed, deaths from leishmaniasis are the second largest in the world, after malaria [25]. Unfortunately, the life cycle of this parasite has led to production of antigenic diversity, so development of vaccine against visceral leishmaniasis has become more and more urgent. Immunoproteomics has thus been employed to track certain antigenic proteins for the purposes of developing vaccine. Protein lysate of *L. infantum* promastigotes (mid-log) was separated by 2D-gel, and Western blot analysis was performed with rabbit hyper-immune serum raised against *L. infantum* promastigote. In-gel digestion was performed on the reactive spots, prior to MALDI-TOF and MALDI-TOF/TOF analysis. Propionyl carboxilasa, ATPase beta subunit, transketolase, proteasome subunit, succinyl-diaminopimelate desuccinylase, tubulin alpha chain, heat shock protein 70, as well as several proteins of unknown function have been identified and considered as potential vaccine candidates [26]. *L. donovani*, another causative parasite of leishmaniasis, was investigated for immunogens using immunoproteomics. Parasites isolated from patients in India were analyzed using 2D-blot on acquired sera. Six antigens were identified by mass spectrometry, including heat shock 70-related protein 1, leishmanolysin, eukaryotic initiation factor 4A, and elongation factor 2 [27]. Through employing immunoproteomics strategies, these studies have demonstrated the potential of vaccine candidates of two species in the *Leishmania* genus.

*Brugia malayi* is a nematode causing filariasis in humans, also known as elephantiasis. This parasitic disease remains a health problem in tropical countries due to lack of availability of vaccine for prevention, and no accurate diagnostic test. Immunoproteomics was applied in order to define those potential antigenic proteins of adult *B. malayi* recognized by IgM, IgG1 and IgG4 in patient’s sera. A total of 30 immunogens undergoing a reaction with IgM, IgG1 and IgG4 were analyzed by 2D-blot and Q-TOF mass spectrometry, including endoplasmin precursor, galectin and transketolase. Identified proteins remain an important resource for development of vaccine against filariasis, and for rapid diagnostic tests [28]. The use of ‘omics’ technologies have, in summary, proved repeatedly successful in novel antigen discovery.
**Drug targets**

As part of the drug discovery process, proteomics can help in finding new drug targets. Chemical proteomics represent one of the most direct approaches in screening for drug-protein interactions. *T. brucei* is the causative parasite of African sleeping sickness. Currently, treatments for African sleeping sickness are expensive, toxic, and are also difficult to administer [29]. Therefore, it is necessary to develop novel compounds for treating *T. brucei*. Protein kinases play an important role in the cell signaling of most organisms. Accordingly, kinases are an attractive target for drug discovery. The chemical-proteomics approach was introduced to profile targets for kinase inhibitors; the method is based on ‘kinobeads’, which is a mixture of kinase-inhibitors on the matrix. Kinobeads are incubated with *T. brucei* lysate. Specific targets bind to the beads, whereas non-targeted proteins remain unaffected. Bound proteins are analyzed by mass spectrometry, and the results show that over 50 *T. brucei* kinases are sensitive to typical kinase inhibitors. Identified kinases stand as potential drug targets, reacting specifically with different inhibitors [30].

Paromomycin is an aminoglycosidic antibiotic currently used to treat visceral leishmaniasis. A basic knowledge of paromomycin is required as its mode of action is still unclear, and resistance to paromomycin is also increasing. The proteomics approach has been employed to understand the resistance mechanism to this antibiotic. The proteomes of wild-type AG83 strain, as well as a paromomycin-resistant strain of *L. donovani*, have been elucidated by quantitative mass spectrometry. Resistant strains exhibited up-regulation expression of ribosomal proteins, glycolytic enzymes, and stress proteins. The mode of action of paromomycin is being investigated further by proteomics. In addition, paromomycin affinity pull-down assay followed by mass spectrometry has been used to identify specific drug-bound proteins. Paromomycin was attached to AminoLink Plus Coupling Resin and incubated with AG83 whole cell lysate. Bound proteins were loaded onto 10% SDS-PAGE gel and then visualized by SYPRO Ruby protein blot stain. The gel bands were excised, tryptic digested, and then analyzed by LC-MS/MS. Paraflagellar rod proteins and prohibitin have been shown to interact with paromomycin. This study has revealed significant data regarding the mode of action of paromomycin as well as the mechanism of paromomycin resistance in *L. donovani* [31].

*Fasciola hepatica* is the trematode liver fluke that causes fasciolosis. Triclabendazole has been an effective drug for the treatment of fasciolosis for over 20 years, however, triclabendazole-resistant liver flukes have now been uncovered in Europe and Australia [32]. In addition, the mode of action and biological target(s) of triclabendazole are still unclear. 2D-gel was used to compare soluble proteomes of triclabendazole-resistant and -susceptible isolates of *F. hepatica* in the presence and absence of triclabendazole sulfoxide. Glutathione transferase and a fatty acid binding protein were identified as potential drug targets. Indeed, binding of these proteins with triclabendazole has already been validated [33].

Some remarkable successes using proteomic technologies in the discovery and identification of new drug targets have been reviewed. This knowledge will be valuable for drug design and development in the treatment of parasitic diseases going forward.

**Host-parasite interactions**

The proteomics approach has been introduced to explore host-parasite interactions for a better understanding of the fundamental biology of parasites. *Dirofilaria immitis* is a parasitic roundworm that causes filariasis. The common name for *D. immitis* is ‘heartworm’. However, it is known to reside in the lung arteries for extended periods. It is assumed this parasite interacts with the host’s intravascular system as a result. Serum levels of D-dimer – products of fibrinolysis – significantly increase during infection. Interactions between antigens of *D. immitis* and the fibrinolytic system of the host have been investigated using proteomics. 2D-blot of *D. immitis* antigens was incubated...
against human plasminogen, then detected by sheep anti-human plasminogen IgG. Several proteins, such as HSP60, actin-1/3, actin, actin4, transglutaminase, GAPDH, Ov87, LOAG 14743, galectin and P22U, were found to bind with the human plasminogen identified by MALDI-TOF/TOF. This study suggests that parasite antigens may interact with host plasminogen, regulating activation of the fibrinolytic system [34]. *T. gondii* is an intracellular parasitic protozoan, causing toxoplasmosis. It invades and multiplies within human macrophages. Proteomes of toxoplasma-infected macrophages and non-infected controls were examined, using 2D-gel and MALDI-TOF/TOF, in which differentially expressed proteins could be revealed. Expression of actin, enolase, calumenin, vimentin, plastin 2, annexin A1, cathepsin S, arginase-1, arachidonate 12-lipoxygenase, and aminoacylase-1 of the macrophages were affected by the infection. The result provides useful information for tracking the biology of the *T. gondii* pathogen [35].

*Taenia solium* is a cyclophyllid cestode. The common name of *T. solium* is the pork tapeworm. Humans are infected by ingesting food contaminated with feces containing eggs. Oncospheres of *T. solium* hatch in the intestine, invade the intestinal wall, then migrate to other tissues. Oncospheres can migrate to the central nervous system, causing cysticercosis, which is a major cause of seizures in humans. *T. solium* may remain in the host, while being attacked by the host immune system. Indeed, it appears to develop some mechanisms of defense against the host response. Generally, the defense mechanism of helminths against host immunity is the production of excretory-secretory proteins [36]. To understand the interactions between *T. solium* and its host, natural larvae were collected and cultured in vitro. The larval culture media were separated by 1D-gel. Each gel lane was cut into 48 identical pieces and subjected to gel digestion. All peptide mixtures were analyzed by LC-MS/MS. Thirty-two (32) proteins, which were likely from *T. solium*, were identified and used as diagnostic antigens. Moreover, these identified secretory proteins are believed to be involved in the host's immune reaction, such as reducing an inflammatory response and inhibiting complement formation [37].

**Proteomic technologies offer new opportunities in characterizing host-pathogen interactions.** A better understanding of these interactions helps to elucidate both pathogenicity and virulence, and may potentially result in development of pathogen-specific host biomarkers and more effective drug treatments. Although proteomics provides a great deal of valuable data, it also has some limitations. At present, a standardized gel matrix for setting up 2D-gel to reproduce aligned protein patterns is still unavailable. In addition, the database is swamped with redundant data of the most abundant and separable proteins, due to missed or inaccurate identification. To overcome these limitations, researchers are working on developing novel technologies to resolve existing issues.

**Conclusion**

Proteomics is the large-scale study of proteins, particularly for protein identification, quantification, post-translational modifications, and localization. Proteins are important parts of all living organisms because they are the main components in metabolic pathways. The development of mass spectrometry provides tools with a high potential for protein research. In parasitology, for instance, basic knowledge is required to understand parasite biology for more effective disease control. Proteomics is being used in parasite research and has already provided some valuable findings. Advanced research related to identification of biomarkers for parasitic diseases enables the potential of rapid diagnostic tests to be explored. The application of proteomics, leading to discovery of novel immunogens, and those proteins involved in stimulation of the host immune system, have been summarized and discussed. Identified antigens stand as potential vaccine candidates for protecting against parasitic diseases. Interactions between promising antiparasitic compounds and their target proteins have been studied successfully.
using mass spectrometry-based proteomics. An accurate definition of the drug-target interaction elucidates the drug mechanism, which is needed for drug design. Proteomics is also being used in host-parasite interaction studies. Current literature continues to facilitate an understanding of parasite pathogenicity and virulence, while mass spectrometry-based proteomics is starting to provide valuable information as part of the process of developing effective vaccines, drugs, and diagnostics for the treatment of parasitic illness.

References


