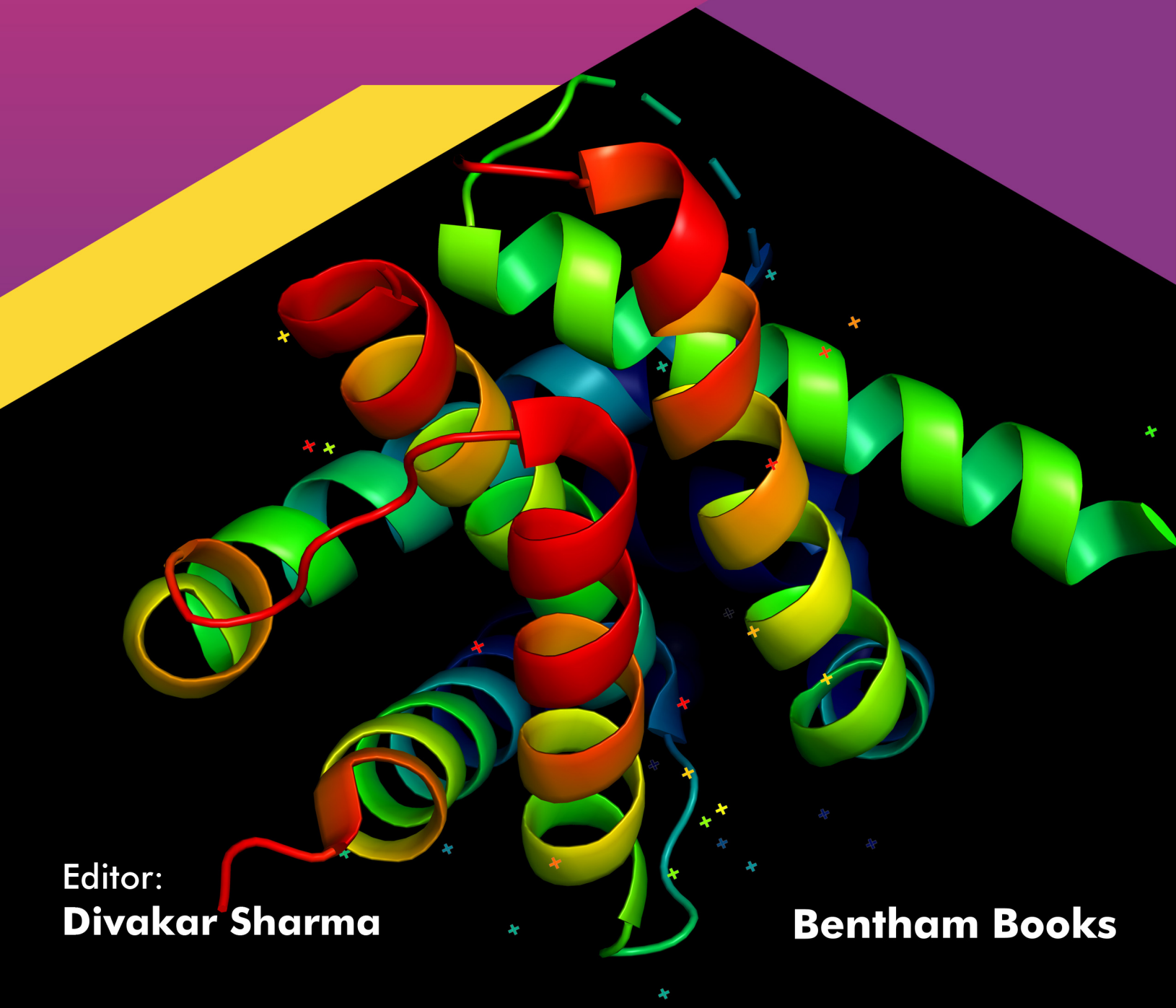
MICROBIAL PROTEOMICS:
DEVELOPMENT IN TECHNOLOGIES AND APPLICATIONS



Editor:
Divakar Sharma

Bentham Books

Current and Future Developments in Proteomics

(Volume 1)

Microbial Proteomics: Development in Technologies and Applications

Edited by

Divakar Sharma

Kusuma School of Biological Sciences
Indian Institute of Technology, Delhi
India

Current and Future Developments in Proteomics

Volume # 1

Microbial Proteomics: Development in Technologies and Applications

Editor: Divakar Sharma

ISSN (Online): 2737-5234

ISSN (Print): 2737-5242

ISBN (Online): 978-981-14-9141-2

ISBN (Print): 978-981-14-9142-9

ISBN (Paperback): 978-981-14-9140-5

need for a court order if at any point you breach any terms of this License Agreement. In no event will any delay or failure by Bentham Science Publishers in enforcing your compliance with this License Agreement constitute a waiver of any of its rights.

3. You acknowledge that you have read this License Agreement, and agree to be bound by its terms and conditions. To the extent that any other terms and conditions presented on any website of Bentham Science Publishers conflict with, or are inconsistent with, the terms and conditions set out in this License Agreement, you acknowledge that the terms and conditions set out in this License Agreement shall prevail.

Bentham Science Publishers Pte. Ltd.
80 Robinson Road #02-00
Singapore 068898
Singapore
Email: subscriptions@benthamscience.net

CONTENTS

FOREWORD

Dr. Divakar Sharma is intending to deliver a very useful work in the form of E-book **Volume I** entitled **"Microbial Proteomics: Development in Technologies and Applications"**, in the E-book series **"Current and Future Developments in Proteomics"**, a collection of chapters on recent development on microbial proteomics related to technologies and diseases. The editor has recruited a vast array of proteomics and microbial specialists from India to write the different chapters that constitute this volume. The chapters are structured in such a way that the reader may easily find recent developments about proteomic technologies and its application to combat the microbes related to current issues, especially on antimicrobial resistance and *M.tuberculosis*. I am sure proteomics and microbiological researchers will find this compilation useful and enjoyable.

I wish editor and authors' good luck for their contributions in delivering this very timely book.

Sanjeeva Srivastava
Group Head of Proteomics Laboratory
Indian Institute of Technology, Bombay
India

PREFACE

We are honored to contribute to the information and updates on current and future developments in proteomics for young researchers around the world. We have attempted to distill the current knowledge of proteomics technologies and development with special reference to microbe's related diseases. This volume of e-book includes various contributors from India which have been affiliated to the well known Indian Institutes.

It is divided into three volumes: Volume I, Microbial Proteomics: Development in technologies and applications; Volume II, and Volume III, will be proposed in the near future. This eBook contains full-color, high quality images of the most frequent technologies and applications with a brief and comprehensive review of microbe related diseases. Each chapter includes its novelty towards the development of proteomics technologies and applications. The format is concise, well organized, and didactic, without being exhaustive. I hope and expect that this volume will facilitate in providing updated basic and specific information to young researchers.

ACKNOWLEDGEMENTS

I would like to express my gratitude to the Indian proteomics experts (Dr. Deepa Bisht, Prof. Sanjeeva Srivastava, Dr. Suman Thakur, and Prof. Bishwajit Kundu), they provided me a positive attitude to think out of the box and sustainable support with English composition and edition. I also acknowledge Dr. Srikanth Tripathy, Dr. Krishnamurthy Venkatesan, Dr. Nirmala Deo, Dr. Shripad A. Patil, and Prof. Asad U. Khan for their positive support. I would also like to express my sincere thanks to my friends and colleagues without whose contribution; it would not have been possible to complete this project. I also want to thank the staff of Bentham Science for their help and support and giving me this opportunity to publish this eBook. I also thanks the ICMR and SERB for providing funds and fellowships for my research.

Divakar Sharma
Kusuma School of Biological Sciences
Indian Institute of Technology, Delhi
India

About the Editor

Presently, Dr. Divakar Sharma is a Principal Project Scientist at the Indian Institute of Technology, Delhi, India. He has completed his M.S. degree in Biotechnology from Chatrapati Sahu Ji Maharaj University Kanpur, India and a Ph.D. degree in Biotechnology from the Jiwaji University, Gwalior, India, as well as ICMR-National JALMA Institute for Leprosy and Other Mycobacterial Diseases, Agra, India. He has completed Two-Post Doctoral Fellowships research at ICMR-National JALMA Institute for Leprosy and Other Mycobacterial Diseases, Agra, India (2015-2017) and Aligarh Muslim University, Aligarh (2017-2019). After completion of the Post Doctoral Fellowships, he joined the Indian Institute of Technology, Delhi, India as a Principal Project Scientist. He is an active Proteomics Researcher in the Antimicrobial Resistance (AMR) for the last 12 years. His research focuses on drug resistance exploration of *M.tuberculosis* and opportunistic pathogens through proteomics and bioinformatics based approaches. His contribution to this field includes more than 50 peer-reviewed articles and book chapters of international repute. He is the editorial board member for various visible journals of international repute affiliated to the well-known publishers like Bentham Science, BMC, Frontiers, and many more. He is also an expert's panelist member of several international journals. Apart from that, he is a life member of various societies in India like Proteomics Society of India, Indian Society for Mass Spectrometry, and Association of Microbiologist of India.

DEDICATION

This volume of the Ebook series is especially dedicated to Dr. Deepa Bisht, Ph.D., Scientist E, at ICMR-National JALMA Institute for Leprosy and Other Mycobacterial Diseases, Agra, India. She always inspires me to work hard, try to do my best, and think out of the box. She guided me during my doctoral and post doctoral research; and has been my mentor and treats me in a friendly manner. She is a recognized Indian scientist in the area of M. tuberculosis proteomics research.

Divakar Sharma
Kusuma School of Biological Sciences
Indian Institute of Technology, Delhi
India

List of Contributors

Abhishek Parashar	Animal Biotechnology Centre, National Dairy Research Institute, Karnal, India
Aditya Arya	Centre for bioinformatics, computational and systems biology, Pathfinder Research and Training Foundation, Greater Noida, Uttar Pradesh, India
Amit Singh	Department of Microbiology, All India Institute of Medical Sciences, Saket Nagar, Bhopal, India All India Institute of Medical Sciences, New Delhi, India
Anand Kumar Maurya	Department of Microbiology, All India Institute of Medical Sciences, Bhopal, India
Ankita Mathur	Mewar University, Gangrar, Chittorgarh, Udaipur, 312901, India
Anil Kumar Gupta	Department of Microbiology, All India Institute of Medical Sciences, Saket Nagar, Bhopal, India
Apoorva Narain	Department of Respiratory Medicine, King George Medical University, Lucknow, India
Arpana Sharma	Central University of Gujarat, Gandhinagar, India Mehsana Urban Institute of sciences, Department of Biotechnology, Ganpat University, Kherva, 384012, India
Arvind K. Verma	Animal Biotechnology Centre, National Dairy Research Institute, Karnal, India
Ashok K. Mohanty	Animal Biotechnology Centre, National Dairy Research Institute, Karnal, India
Chandrajeet Singh	Central University of Gujarat, Gandhinagar, India
Deepa Bisht	Department of Biochemistry, National JALMA Institute for Leprosy and other Mycobacterial Diseases, Tajganj, Agra 282004, India
Divakar Sharma	Kusuma School of Biological Sciences, Indian Institute of Technology, Delhi, India
Gopika Raval	Mehsana Urban Institute of sciences, Department of Biotechnology, Ganpat University, Kherva, 384012, India
Juhi Sharma	Department of Botany and Microbiology, St. Aloysius College, Sadar Cantt, Jabalpur, 482002, India
Kanchan Srivastava	Department of Respiratory Medicine, King George Medical University, Lucknow, India
Kruti Dave	Mehsana Urban Institute of sciences, Department of Biotechnology, Ganpat University, Kherva, 384012, India
M. Madhan Kumar	ICMR-National JALMA Institute for Leprosy and other Mycobacterial Diseases, Tajganj, Agra-282004, India
Monalisa Tiwari	Department of Biochemistry, Central University of Rajasthan, Ajmer-305817, India
Munna Lal Yadav	Animal Biotechnology Centre, National Dairy Research Institute, Karnal, India

Nirmala Deo	Department of Biochemistry, National JALMA Institute for Leprosy and other Mycobacterial Diseases, Tajganj, Agra 282004, India
Pranav Kumar Prabhakar	Department of Medical Laboratory Sciences, Lovely Professional University, Punjab-144411, India
Preeti Rawat	Animal Biotechnology Centre, National Dairy Research Institute, Karnal, India
Rikesh Kumar Dubey	Division of Microbiology, CSIR-Central Drug Research Institute, Lucknow, India
Sandeep Sharma	Department of Medical Laboratory Sciences, Lovely Professional University, Phagwara, Punjab-144411, India
Sarika Sharma	Department of Life Science, Arni University, Kathgarh, Indora, Kangra. H.P.-176401, India
Saroj Sharma	Department of Biochemistry, Central University of Rajasthan, Ajmer-305817, India
Sudarshan Kumar	Animal Biotechnology Centre, National Dairy Research Institute, Karnal, India
Surya Kant	Department of Respiratory Medicine, King George Medical University, Lucknow, India
Vishvanath Tiwari	Department of Biochemistry, Central University of Rajasthan, Ajmer-305817, India
Vivek DharDwivedi	Centre for bioinformatics, computational and systems biolog, Pathfinder Research and Training Foundation, Greater Noida, Uttar Pradesh, India
Vivek Kumar Gupta	ICMR-National JALMA Institute for Leprosy and other Mycobacterial Diseases, Tajganj, Agra-282004, India
Yogesh K. Dhuriya	Developmental Toxicology Laboratory, Systems Toxicology and Health Risk Assessment Group,CSIR-Indian Institute of Toxicology Research (CSIR-IITR) Vishvigyan Bhawan, 31, Mahatma Gandhi Marg Lucknow–226 001, India

CHAPTER 1

Microbial Biofilm and Drug Resistance: A Proteomic Approach

Sarika Sharma[1] and **Sandeep Sharma**[2,*]

[1] *Department of Life Sciences, Arni University, Kathgarh, Indora, Kangra. H.P.-176401, India*

[2] *Department of Medical Laboratory Sciences, Lovely Professional University, Phagwara, Punjab-144411, India*

Abstract: Bacterial Biofilms are densely packed microbial communities formed by the microorganism to escape from any external threat. These Biofilms are composed of a polysaccharidic matrix, which formed a slimy layer outside the cell wall and protected the microbes from any damage (both physical and chemical). Biofilm can be developed by microorganisms on any surface, including medical devices, oral cavity and other biomaterials. These biofilms are difficult to treat and required almost 10,000 more concentrations of antibiotics compared to planktonic microorganisms. These biofilms are the main hindrance in the treatment of hospital-acquired infections. Majority of the treatment against microorganisms fails due to these biofilms in a clinical setting. Combination of different antibiotics, natural molecules and other strategies is in use to combat these biofilms. Microorganisms inside the Biofilm trigger some genes by quorum sensing and affect the expression of several protein factors. The current chapter will focus on the use of the proteomic approach for better understanding the nature and role of Biofilm in microbial pathogenesis and lower the emergence of drug resistance in these microorganisms.

Keywords: Antibiotics, Biofilm, Drug resistance, Microorganism, Pathogenesis, Polysaccharidicmatrix, Proteomics.

INTRODUCTION

Microbes are ubiquitous in nature and always busy in executing tasks. Some are building products for the benefit of humankind, and some are creating problems for humans. Biofilms are defined as organized communities of bacterial cells that are enclosed in a self-constructed polymeric matrix and adhere to any surfaces, an interface or each other [1]. Bacterial cells in biofilms are much different from their planktonic counterparts or free-floating forms. Biofilms are the extremely preferred approach of bacterial growth to survive in adverse conditions [2]. In the

* **Corresponding author Sandeep Sharma:** Department of Medical Laboratory Sciences, Lovely Professional University, Phagwara, Punjab-144411, India; E-mails: sandeep4380@gmail.com and sandeep.23995@lpu.co.in.

Divakar Sharma (Ed.)

natural environment, biofilms are found almost everywhere, but the negative side of Biofilm lies within Biofilm associated infections in humans. Nowadays, biofilms are accountable for many human infections, and 65 to 80% of all human infections are linked with microbial biofilms [3]. These include wound infection, infection of the lower respiratory tract in cystic fibrosis, infection in kidney stones and implanted devices in orthopedic surgery. The Biofilm associated bacterial cells have a clinically significant property with their high antibiotic resistance level compared with their free-living (planktonic) counterparts [4]. The reason for increased drug resistance may be that the bacteria within the Biofilm grow more slowly than their planktonic formed or the matrix formed by sessile bacteria, preventing the entry of the drug into the cell, while transcriptomics and proteomics revealed that the gene expression in free-floating bacterial cells is much different from bacterial cells associated with Biofilm or sessile bacterial cells [5, 6]. There were reports published in the last few years that increase resistance to an antibiotic is directly related to the high level of protein expression involved in glycolysis, microbial metabolism, and synthesis of secondary metabolites [7]. Moreover, there are also some other biological pathways which are directly involved in cell to cell communications (LuxS-mediated quorum sensing, arginine metabolism, rhamnose biosynthesis). Few other proteins (pheromone and adhesion associated proteins) were found to be upregulated during the biofilm transit from planktonic stages [8]. The exact molecular mechanisms for increased drug resistance to Biofilm are not fully understood, although many studies have already been conducted for so many years. Advances in proteomics techniques during the past decade have facilitated an in-depth analysis of the possible mechanisms underpinning increased drug resistance within Biofilm. Proteomic studies are also to establish its role in better understanding the nature of bacterial biofilms. Biofilm proteomics is defined as to study the complete set of proteins of bacteria embedded in the Biofilm [9, 10]. In this chapter, various proteomics approaches related to the Biofilm, and their increased drug resistance to pathogens will be illustrated.

The Process of Biofilm Formation

Biofilm formation is an exceptionally sophisticated process in which free-floating microbial cells will be transformed into sessile form. The formation of Biofilm is a multistage process in which a series of events takes place as an adaptation of microbial cells to survive and divide in a broad range of harsh environmental conditions [11]. Biofilm formation is a multi-step process which includes attachment of free-floating microbial cells to a surface, the growth and aggregation of cells into microcolonies followed by growth into mature, structurally complex Biofilm (maturation), and the dispersal of detached bacterial cells as shown in Fig. (**1**) [12].

Attachment

The attachment of the bacterial cells to the surface is mediated by any one of them (like week van der Waals forces, electrostatic and hydrophilic interactions) or by a combination of all these interactions [13]. Pili and pilus-associated adhesins have also been reported to play an important role in adhesion and colonization of microbial cells to the surfaces [14, 15]. The microbial cells that initiate attachment are covered by only small amounts of exopolymeric material. While following the initial attachment to a surface, the bacteria have to maintain contact with the surface. Thus, the stage of attachment is followed by a phase during which the synthesis of bacterial exopolysaccharides (EPS) provides steady attachment by forming organic bridges between the cells and the surface [16].

Micro-Colony Formation

After the bacteria adhered to the surface, the bacterial cell multiplies within the embedded exopolysaccharide matrix, which lastly results in micro-colony formation. After the micro-colony formation stage of Biofilm, there is an upregulation of certain biofilm genes. It is observed that bacterial attachment can trigger the formation of an extracellular matrix [17]. Matrix formation is followed by water-filled channels formation for the transport of nutrients within the Biofilm. Researcher has proposed that these water channels are like a circulatory system, distributing different nutrients to and removing waste materials from the communities in the micro-colonies of the Biofilm [3, 18].

Detachment

The growth of the bacterial Biofilm is eventually restricted by the availability of nutrients accumulation of toxic by-products and other factors, including pH, oxygen perfusion, carbon source availability and osmolarity [19]. At some point, dispersing of biofilm cells occur either by the detachment of newly formed cells from growing cells or dispersion of biofilm aggregates. The cells may detach, and together with the progeny of other biofilm cells, may colonize other surfaces [20]. Enzymes such as polysaccharide lyases that degrade the extracellular polysaccharide matrix have been reported to play a role in biofilm dissolution in several organisms [21]. The cells which are dispersed from biofilm as a result of growth may return quickly to their normal planktonic phenotype.

Mechanism of Biofilm Resistance against Antimicrobials

Resistance to drugs and antibiotics is attained by microbes permanently or temporarily during the survival strategy of microbes in adverse circumstances [22]. Antibiotic resistance is nowadays a serious threat for humankind as the

CHAPTER 2

Past, Present, and Future of Gel-Based Microbial Proteomics

Munna Lal Yadav[1,*], Arvind K. Verma[1], Preeti Rawat[1], Abhishek Parashar[1], Divakar Sharma[2], Sudarshan Kumar[1] and **Ashok K. Mohanty[1,*]**

[1] *Animal Biotechnology Centre, National Dairy Research Institute, Karnal, India*

[2] *Kusuma School of Biological Sciences, Indian Institute of Technology, Delhi, India*

Abstract: Proteome deals with complete set of proteins, expressed by a genome of a cell present inside of an organism at a specific time period. It includes expression study, information of modifications in proteins, interactions with other protein biomolecules, *etc*. The goal of proteomics is to know new protein information, whether a protein is over expressed or under expressed in specific situation and overall gives information of its effect on an organism. Thousands of proteins can be examined at a time as compared to other methodologies. Many advanced technologies have been evolved to investigate proteome in depth and generate a huge amount of data. The most common high throughput techniques such as polyacrylamide & agarose gel electrophoresis and their advanced versions in combination with mass spectrometry are being used in modern proteomics. The study of microbial proteome data helps us to know how bacteria get resistant to particular drug and which biomolecules are involved in that process. Database also gives the opportunity to develop better drugs that target new places on bacterial surface or new drugs on the same target. The advancement of proteomics technique and their applications in microbial research has granted a new hope to explore disease biomarkers and the development of diagnostic assays. Microbial protein's benefits are extensively used in the agricultural sector. Proteomics profiling has a key role in disease identification in humans and animals. Thanks to protein information as new targets are identified and more safer and effective drugs are produced.

Keywords: 2D PAGE, Biomarker discovery, Dairy processing, DeCyder, DIGE, Gel electrophoresis, IEF, Microbial proteomics, SDD-AGE, SDS-PAGE, Proteome, Tricine.

[*] **Corresponding author Munna Lal Yadav:** Animal Biotechnology Centre, National Dairy Research Institute, Karnal, India; Tel: 0184-2259542; E-mails: munna.biotek@gmail.com and ashokmohanty1@gmail.com

Divakar Sharma (Ed.)

INTRODUCTION

Proteomics is the branch of life science that deals with the study of protein expression, protein modifications and interactions between proteins. It involves the identification of proteins with a possible role, change in protein expression, modification such as glycosylation, phosphorylation and interaction with other molecules inside the cell [1 - 3]. Proteomics has major implications in understanding the physiological process better. This allows the researcher to identify various proteins that are expressed inside a cell, which may not be the same as the transcriptome profile [2]. Generation and analysis of large proteomics data are expected to show how cells vigorously respond to changes in their surroundings, and disease conditions [2, 4].

Microbial proteome helps us to know how microbes get resistant to a particular drug, their biochemical pathway involves. This information is vital in selecting/creating novel antibiotics or drugs for their destruction [5, 6]. Further, it gives new hope to explore disease biomarkers and development of diagnostic assays. On the other hand, profiling of good bacteria proteins is important for the agriculture sector for increasing yield of crops and dairy processing [5, 7 - 8]. Many advanced technologies have been evolved to investigate proteome in-depth and generate a huge amount of data. The most commonly high throughput techniques used are gel based methods: Polyacrylamide & agarose Gel Electrophoresis and their advanced versions (SDS-PAGE, Native PAGE, SDD-AGE, Tricine SDS-PAGE, 2D-PAGE, DIGE and mass spectrometry (MS). 2D-PAGE or with LC-MS (Liquid Chromatography–Mass Spectrometry) are used in comparative proteomics study such as difference profiling of protein from normal condition to disease condition [2].

PROTEIN SAMPLE PREPARATION FROM MICROBIAL SOURCE

Microbial cell wall or cell membranes are different from other mammalian cells or tissues. Bacterial cell wall is relatively rigid and tough. It is important to have an idea about the localization of proteins. As some of the proteins are secreted in medium so their presence needs tobe checked in medium instead of cell lysate fraction [9]. The detergents are considered good to break cell wall proteins. Lysozyme is another component of bacterial lysis buffer which breaks peptidoglycan bonds. Some non-ionic detergent solutions are commercially available that mimic the role of lysozyme. The combination of DNase I and Lysozyme gives good result in purifying large molecular weight proteins [4, 10].

For a comprehensive analysis of all intracellular proteins, the cells should be efficiently disrupted. During protein extraction method, it is important to maintain

temperature throughout cell lysis process and centrifugation. Proteases maybe liberated upon cell disruption. Cell lysis leads to the action of unregulated action of endogenous protease and phosphatase on the cellular protein [10, 11]. Over time the actual activity of the target protein in the study or whole cellular protein concentration is reduced.To minimize these effects and acquire the simplest possible protein yield after cell lysis, protease inhibitor cocktails are added to the lysis buffer [4, 12].

The protocols for proteome analysis depends upon various factors such as sample type, its physical and chemical properties, location of proteins inside the cell, availability of quantity and designing of an experiment [2, 5]. The methods are optimized at different steps for accurate quantification of proteins and comparison without loss of the actual sample. Sample must be prepared in such a way that there is minimum loss of proteins and various proteins are maintained in their true state of expression. The main challenges are the small sample size, difficulties in extraction, losses associated with protein precipitation and preservation of proteins without loss of low abundant proteins [2, 4 - 5, 13]. Further, for the solubilization of proteins, compatible buffers are required. In 2D gel electrophoresis urea is the choice of chemical for protein solubilization, however, in shotgun applications, most of the proteins are solubilized in ammonium bicarbonate containing buffers.

CLEAN-UP AND QUANTIFICATION OF PROTEINS

Protein sample clean up refers to the process of removing non-protein agents (detergents, salts and other molecules used in or generated during protein extraction or purification) that can hinder the results of downstream processes such as isoelectric focusing (IEF), 2D electrophoresis, adverse effects on protein function or stability, or interfere with downstream applications [4, 11]. Therefore, it may be necessary to remove or reduce these contaminants by different methods of protein precipitations (ammonium sulphate, acetone, trichloro acetic acid) or dialysis and desalting of the samples. Quantification of protein requires before proceeding to isolation, purification, and gel analysis [14 - 15]. Most protein estimation methods depend on the levels of tryptophan, tyrosine and other aromatic amino acids. There are many methods of protein estimation but Bradford assay and Bicinchoninic Acid (BCA) assay are most widely used.

GEL ELECTROPHORESIS OF PROTEINS

Gel Electrophoresis is a process where macromolecules like proteins and nucleic acids are moved on the gel in the presence of electric current. The rate of

CHAPTER 3

Liquid Chromatography-Mass Spectrometry (LCMS): An Advanced Tool for the Microbial Proteomics Analysis

Pranav Kumar Prabhakar*

Department of Medical Laboratory Sciences, Lovely Professional University, Punjab-144411, India

Abstract: Almost 20 years after the publication of the first microbial sequences, today's world is growing in the direction of the post-genomic era with the understanding of transcriptomics and proteomics, which offer insight into cellular physiology. As the function of protein is associated with the phenotypic characters of an organism, the power to exceed the entire protein network of a microbial world has a huge impact on microbiology. Nowadays, mass spectroscopy (MS) has been extensively used in microbiology for the identification, characterization, and serotyping. In recent times, the large genome size, its complexity, and their study is a big task for the microbiologist. It is interesting to see how scientists examine microbial proteomics and analyze them. Proteomics has a key task to carry out in endeavors to develop far-reaching cell maps of biochemical procedures happening inside explicit microorganisms at given spatial and worldly focuses. Here we are going to discuss the methodologies for the identification of bacteria up to species level with great accuracy with the use of proteomes of bacterial pure culture. This chapter will also discuss the sample preparation and identification of a specific strain of bacteria for the liquid chromatography-tandem mass spectrometry (LC-MS/MS) and the recent application of it.

Keywords: Bacteria, Chromatography, Electrospray, Isotope, Mass spectroscopy, Metabolic, Proteomics, Protein, Resistance, Transcriptomics.

INTRODUCTION

Several microbes grow in our surroundings and most of them are not visible with naked eyes. These microorganisms are extensively used for studies and research due to their simple cellular structure and less growth requirement. The genome of microbes is very small compared to most of the eukaryotes and hence easy to handle during the study, analyze, and sequencing. Proteomics has been efficiently

* **Corresponding author Pranav Kumar Prabhakar:** Department of Medical Laboratory Sciences, Lovely Professional University, Punjab-144411, India; Tel: 07696527883; E-mail: prabhakar.iitm@gmail.com

Divakar Sharma (Ed.)

used for both fundamental and applied research to analyze and study cellular metabolic processes. There are several tools available in proteomics to probe the gene expression in the case of bacteria in given specific environmental conditions. The microbial proteomics has been used for the study of hotspot areas of choice such as stress responses, an adaptation of microbes in extreme environmental conditions, the pathophysiology of microbes, metabolic engineering. The ability of proteomics to address essential issues in the microbial field is largely subject to the supported advancement of various proteomic advances, which separately show their ability in proteomic exploration either subjectively or/and quantitatively [1].

The active role of biomolecules in the maintenance and helping in better life has been studied since the initial days of biochemical and biomedical researches. To explain the significance of these biomolecules in our life, in 1938, Berzelius used the term protein for them, which comes from the Greek and means the first [2]. All the protein content of a cell specified by its destination, protein-protein interaction, protein-DNA (Deoxyribose nucleic acid) interaction, posttranslational modifications, and its turnout rate is called the proteome. In 1996, Wilkins and Williams used the term proteomics first time for the total protein content of a cell, tissue, or organism and it denotes, "PROTein complement of a genOME" [3].

Most of the physiological information of DNA is expressed in terms of protein and is characterized by proteome. The proteome of prokaryotes is simple and easy to study compared to eukaryotes. The proteins expressed by prokaryotic bacterial cells are responsible for its pathogenicity and the study of such proteins isvery challenging and difficult to study due to huge differences in their physical and chemical properties such as amount produced, size and nature of protein, solubility, hydrophobicity, and lipophilicity, regulations [4].

Proteomics is a very effective tool for the early diagnosis of a disease, medical prognosis, and also to monitor the development and growth of disease as well as in the drug development process. All the proteome of a cell or tissue at any given time and in any given condition is characterized in terms of the proteomics, which includes its expression, molecular structure, shape and sizes, physiology and function, interactions with protein, or DNA, modification and post-transcriptional modifications [5]. The proteome of a cell, tissue, or organism varies from time to time and tissue to tissue or also in case of any kind of stimulations [6]. The variation in the gene expression can be studied and analyzed by the transcriptome and proteome to differentiate between the two physiological states [6].

Proteomics is one of the most significant techniques to encompass the gene function, though compared to genomics, it is more complex [7]. The microarray

chip has been designed and developed for the large scale analysis of complete transcriptome [8]. Proteins of the cells are the functional component and their expression level is not only dependent on its mRNA (messenger ribose nucleic acid) level but also the translational regulatory system of the host cell and hence the proteomics is treated as one of the important data sets for analyzing the physiological status of a biological system [9]. There are several methodologies available for the proteome study. These techniques are classified into different subcategories such as conventional methods such as Chromatography, ELISA (Enzyme Linked Immuno Sorbent Assay), western blotting, advanced technologies (microarray-based techniques, gel-based assays, mass spectroscopy, sequencing), quantitative techniques, high throughput techniques, *etc*. The genome-based microarray is common, and the hybridization pattern of protein is completely different from the nucleic acid. The investigation of a huge number of proteins at a single time is very difficult as we require antibodies against each of the analyzed proteins and also the antibody binding pattern and process need to be optimized for each protein. And hence, protein microarrays are not a good choice for the complete proteome analysis process. Similar to the microarray, gel electrophoresis is also not a good choice to study gene expression as there is a limitation in the spotting, identification, and estimation and also not very effective for the mixture of proteins [10, 11]. Mass spectrometry (MS) has come with an efficient tool for the characterization of proteins from a complex mixture of protein [12]. There are several MS methods available for the investigation of the complete protein of cells/tissues. And they are "Surface-Enhanced Laser Desorption Ionization (SELDI)" [13], "Matrix-Assisted Laser Desorption Ionization (MALDI)" [14] "coupled with time-of-flight (TOF)" or other instruments, and "as chromatography MS (GC-MS)" or "liquid chromatography MS (LC-MS)." One of the limitations of SELDI and MALDI is that they do not induce online fractionation during MS analysis, and hence the protein complex mixture needs to be fractioned beforehand. MALDI is used in most of the cases for tissue imaging [15 - 17]. GS-MS or LC-MS has advantages over other MS tools that induce an online mode of separation for the protein test sample and hence commonly used for high-throughput proteomics.

None of these techniques are perfect as they are associated with some more drawbacks, although they also have advantages (Fig. **1**; Table **1**).

CHAPTER 4

Functional Annotation and Enrichment of Microbial Proteins Using Systems Biology: Tools and Applications

Aditya Arya* and **Vivek Dhar Dwivedi**

Centre for Bioinformatics, Computational and Systems Biology, Pathfinder Research and Training Foundation, Greater Noida, Uttar Pradesh, India

Abstract: Genomics and proteomics methods have witnessed a huge surge in the recent decade. Provided the pace of data accumulation from several advanced technologies such as next-generation sequencing, transcriptomics, ChipSeq and quantitative proteomics, the parallel growth in the functional annotation is relatively slow and needs continuous curation and analysis of existed data from repositories. Nevertheless, the standard procedures of functional annotations of proteins which were based on classical data sets, need to be revised in light of advanced and more comprehensive data. In fact, most omics technologies are now integrating and their merger can provide a much realistic and holistic picture of any biochemical and molecular scenario provided the data handling and curation is performed on cellular and molecular principles. A number of genomes and proteomes have been functionally annotated in the past; microbial genomes offer a better opportunity with less complex models. Nevertheless, microbes are among one of the most common causatives of pathogenic diseases and their diagnosis, treatment is limited by available information of proteins and their functions. Previously annotated hypothetical functions to proteins are likely to change in some cases therefore a number of research groups have attempted to re-annotate the microbial genomes with newer data-sets and new tools. Re-annotation of *Mycobacterium tuberculosis* was one such example. Besides pathogenic microbes, emerging trends in various useful microbes sequencing have shown a tremendous increase in information on human microflora and remain a highly prospective area in biology. Systems biology lies at the interphase of biology, mathematics and computational biology and involves a holistic approach to visualize a biological phenomenon. This chapter describes the basic principles involved in the functional annotation of hypothetical proteins in light of emerging datasets and tools. Besides the suggestions on improving standard pipelines, it also presents a summary of recently annotated microbial genomes and future prospects of involving systems biology in the functional annotation for improved quality and output.

* **Corresponding author Aditya Arya:** Centre for bioinformatics, computational and systems biology, Pathfinder Research and Training Foundation, Greater Noida, Uttar Pradesh, India; Tel: +91 9811886440;
E-mail: contact.adityarya@gmail.com

Divakar Sharma (Ed.)

Keywords: Annotations, Microbial proteins, Network enrichment, Systems biology.

INTRODUCTION

The study of microorganisms is important for various applications, such as understanding the diseases caused by them, exploration of their biotechnological potentials and understanding the fundamental processes of life. Nowadays, microorganisms are also being employed for resolving environmental issues and therefore, several non-cultivable forms of microbes are also studied. Among them, various microbes have industrial applications like the production of pharmacologically and commercially important metabolites, the use of microbes as biocatalysts in the food and other industries, and biotransformation of metabolites or pharmacologically important molecules. These applications require a continuous search of microbes with better traits or their improvements by available techniques. Among the classical approaches of improving the microbes is a selection from natural environments or random mutagenesis, while other methods include strain improvement using recombinant DNA technology, mutagenesis or genome editing. Almost all these methods have a prerequisite of genomics, proteomics or metabolomics data and its proper annotation. Advancements in these technologies have greatly advanced the field of microbiology and their industrial applications.

Genome is the complete representative set of genetic material present in an organism and most likely stores the genetic information required for various processes. A large number of genomes of microorganisms have been successfully sequenced and their genomics data is publicly available in repositories (www.nih/genomes). Although genomics information is excellent for the prediction of potential proteins, yet a significant part of the genome is also non-coding, which becomes predominant in higher eukaryotes [1]. Phylogenetics, prediction of putative proteins and identification are important objectives which can be achieved by the study of genomes. Unlike the genome, the proteome, a complete set of proteins in a cell or its compartment, is highly dynamic and remains an important tool for microbial characterization. Metabolome is still more dynamic than proteome and becomes an essential tool for the biotechnological assessment of microorganisms. Several technology upgradation and advancements have led to a prolific rise in the availability of genomics, proteomics and metabolomics data for microorganisms. However, lack of proper annotations and description of proteins in existing databases remains a potential lacuna in the characterization and evaluation of microorganisms and therefore, a number of tools have been developed and annotation remains a continuous process [2].

IMPORTANCE OF ANNOTATION AND ENRICHMENT

Annotation in simple terms is the assignment of a term or value to a variable or a parameter, while enrichment is the global annotation of several variables with a common type of parameter. A more organized definition is "DNA annotation or genome annotation is the process of identifying the locations of genes and all of the coding regions in a genome and determining what those genes do. An annotation (irrespective of the context) is a note added by way of explanation or commentary. Once a genome is sequenced, it needs to be annotated to make sense of it" [3]. For example, if a protein is assigned, its association with a disease will be called annotation and if the disease is assigned to a complete set or a sub-set of proteome within a biological network, it is called enrichment. Enrichment and annotation are highly useful in comparison of genomic and proteomic data and also add value to existing data. Besides this, newly identified genomes may contain some predicted genomic regions and some hypothetical proteins, which may be attributed to functions through the annotation process. An annotated dataset is particularly useful for downstream applications such as drug-discovery, determination of metabolic directions, flux studies and regulation of expression in specific pathophysiological conditions. In the case of microbes, the annotations provide valuable information while developing models for their commercial use and development of useful products. Moreover, previous annotations might help and speed up the genetically related annotation process. Over the past couple of decades, a number of bioinformatics methods have evolved and accumulated in public databases, such as gene ontology, which makes it possible to systematically dissect large gene lists in an attempt to assemble a summary of the most enriched and pertinent biology. Huang *et al*., in their review, described a number of high-throughput enrichment tools, including, but not limited to, Onto-Express, MAPPFinder, GoMiner, DAVID, EASE, GeneMerge and FuncAssociate, *etc*., independently developed during 2002 and 2003 as initial studies to address the challenge of functionally analyzing large gene lists. Since then, the enrichment analysis field has been very productive, resulting in more similar tools becoming publicly available [4].

EMERGING TRENDS IN GENOMICS

Since the advent of DNA sequencing in the early 1960s, the genomes of microorganisms begun to be sequenced and that led to the quest of human genome sequencing, which was published as an initial draft in the year 2001 [5 - 7] and later a complete draft in 2003. With this, the rapid proliferation of genome sequencing begun and also the technology improved to a great extent. Conventional Sanger's sequencing modified a lot over a few decades and the emergence of next-generation sequencing revolutionized the whole paradigm.

CHAPTER 5

Emerging Paradigm of Post-translational Modifications in Microbes: An Undefeatable Weapon

Arpana Sharma[1,2], Chandrajeet Singh[1], Gopika Raval[2], Kruti Dave[2], Ankita Mathur[3], Juhi Sharma[4] and **Divakar Sharma[5,*]**

[1] *Central University of Gujarat, Gandhinagar, India*

[2] *Mehsana Urban Institute of Sciences, Department of Biotechnology, Ganpat University, Kherva, 384012, India*

[3] *Mewar University, Gangrar, Chittorgarh, Udaipur, 312901, India*

[4] *Department of Botany and Microbiology, St. Aloysius College, Sadar Cantt, Jabalpur, 482002, India*

[5] *CRF, Mass Spectrometry Laboratory, Kusuma School of Biological Sciences (KSBS), Indian Institute of Technology, Delhi, 110016, India*

Abstract: Bacteria are simple organisms, optimized with basic but robust cell regulation mechanisms for efficient growth. Nonetheless, fitting in the description, they possess remarkable adaptive capacity with diverse survival strategies. Post-translational modifications (PTMs) provide a competitive edge to adapt bacteria with limiting, fluctuating nutrients and extremes of environment variables. PTM is one of the cellular processes in bacteria and helps them to adapt to a new environment for their survival. Several post-translation modifications regulate bacterial functions and provide strength to bacteria for surviving in adverse conditions. On-going investigations revealed many reversible or irreversible novel bacterial PTMs like the addition of simple group (acetylation, phosphorylation, methylation and hydroxylation) or composite molecules (AMPylation, ADP-ribosylation, glycosylation and isoprenylation), or small protein ubiquitin and modifying the side chain residues like (elimination and deamidation). PTMs are recognized as important players in directing cellular dynamics like cell metabolism, stress response, pathogenesis and virulence factors. Bacteria with several PTMs are capable of modulating the signaling pathways by destabilising the host cell defense machinery, protein-protein interactions, ultimately promoting their replication. Currently, many studies focus on the relationship between PTMs and antibiotic resistance, increasing bacterial tolerance to various antimicrobials. A paradoxical behaviour is that a single protein may be modified at one or variable positions in interspecies, but changes also exist in the same species. So, to characterize the multifaceted interactions of PTMs, it is still a challenge in metabolic engineering,

[*] **Corresponding author Divakar Sharma:** CRF, Mass Spectrometry Laboratory, Kusuma School of Biological Sciences (KSBS), Indian Institute of Technology, Delhi, 110016, India; Tel: +91-7906026680; E-mail: divakarsharma88@gmail.com

Divakar Sharma (Ed.)

synthetic biology, and medical sciences. Therefore, understanding of bacterial PTMs and PTMs directed host proteins modification will provide better insights into host-pathogen interaction. This chapter focuses on the roles of PTMs in nutrition sequestration and cellular response. Furthermore, we discuss the prospects and advances of proteomics tools in enhancing knowledge related to PTMs of human gut microbiota.

Keywords: Acetylation, Carboxylation, Glycosylation, Interactomics, Lipidation, Methylation, Microbial proteins, Nitrosylation, Phosphorylation, Post-translational modifications, Proteomic tools.

INTRODUCTION

Bacteria are simple, single-celled organisms that can thrive in extreme environmental conditions. They are classified as prokaryotic cells which have a simple internal structure, lack a nucleus and contain DNA in the form of a loosely warped fiber-like structure called “nucleoid”. The relationship of human-bacteria is complicated as some bacteria are useful while others are disease-causing. They have been seen in many associations such as symbiotic, parasitic or commensalism. In the symbiotic associations, natural selection favours the organisms to interact within species in a manner to favour them in extreme environments, which is absent in the individual species. The bacteria, which associate together as co-cultures, which have many advantages in the biotechnology industries, are also a very important part of gut microbiota, which are essential for the normal working of the host organisms [1]. Most of the mutualistic associations are harmless, but there are some which are harmful and cause the disease to the human species. Such bacteria are identified as pathogenic organisms and their symbiotic interactions cause infectious diseases and host damage [2, 3]. These pathogenic microorganisms adapt themselves by immune-compromising the host by suppressing innate immunity to survive in the host cells, subsequently posing a constant challenge to human wellbeing.

Hence, infection biology has been in use over the years to study the abnormal responses which are acquired by the pathogenic organisms inside the host cells. The bacteria have a specific infection cycle, which is identified [4]. Firstly they will introduce into the host tissue by a specific mode of transmission. Subsequently, they will adhere to the tissues and establish themselves by evading the host immune responses. Consequently, they will disseminate into the nearby tissues by secreting toxins. In the process, they will hijack the host molecular mechanisms and mimic the same for its growth.

Nowadays, there is an appearance of many multidrug resistance species, which has created a lot of curiosity in the infection biology fields. Also, with the

advancement of proteomic techniques, there is a lot of research going on to study the molecular signals during infections. The assertiveness of virulence and infection is due to the secretion of various proteins that may be secreted in the original form or may be modified [5].

It has been investigated that various post-translational modifications have been studied, which have been shown to increase the virulence and pathogenicity of bacteria. It is pivotal that these protein secretion systems must be studied and understood for the eradication of the pathogens. To curb the bacterial disease, it is important to target any of these proteins and their secretion systems [6].

Bacterial cells change their endogenous biochemical functions to adapt to the altering environmental conditions. They achieve these adjustments by successive post-translational modifications (PTMs), which alter the biological and chemical properties of proteins. Till date, approximately 200 PTMs have been established and illustrated by many novel proteomic approaches. These PTMs are very important as they confer important protein function properties, maintains its stability, modifies its activity to the changing environmental cues and helps in establishing interactions with other proteins [7 - 11].

The chapter will focus on the different types of post-translational systems involved in bacterial pathogenicity, their implication in providing virulence to pathogens, and how they help in establishing the host-pathogen interactions. The following sections will also elucidate the various proteomics tools and technologies available to study these post-translational modifications and what are the technical difficulties involved in studying such systems.

What are the General Characteristics of Bacteria and the Significance of the Study?

Bacteria are simple but robust cellular entities that have basic cellular machinery, which is better developed for speedy division and rapid growth. But, bacteria, despite being uncomplicated, acquire special properties of adaptations to survive in diverse extreme situations by modifying their size, morphology, shape development patterns and metabolism behaviour [12]. This robust cellular mechanism and amazing adaptive capacity have given emergence to certain pathogenic strains of bacteria. The flora which is present in healthy individuals is essential for accurate working but multiple pathogenic organisms can mutually coexist in the same host niche to produce unnatural responses [12].

These responses have to be studied in the co-existence of multiple microbial species conditions, which break the immune system of the host and elicit virulence conditions. Hence, bacterial co-culture growth analysis in an appropriate

CHAPTER 6

Pupylation: A Novel Proteolysis Pathway in Prokaryotes Functionally Reminiscent to Eukaryotic Ubiquitination

Yogesh K. Dhuriya[1] and **Divakar Sharma**[2,*]

[1] *Developmental Toxicology Laboratory, Systems Toxicology and Health Risk Assessment Group, CSIR-Indian Institute of Toxicology Research (CSIR-IITR), Vishvigyan Bhawan; 31, Mahatma Gandhi Marg Lucknow – 226 001, India*

[2] *CRF, Mass Spectrometry Laboratory, Kusuma School of Biological Sciences (KSBS), Indian Institute of Technology, Delhi (IIT-D)110016, India*

Abstract: Posttranslational modification of proteins is a prevalent method for the regulation of cells according to changes in the surrounding environment and diversifications. Pupylation, architecturally similar but not homologous to eukaryotic proteasomal degradation machinery, exists in a certain order of bacteria, especially actinobacteria. Pupylation supports the bacteria to survive under challenging environmental conditions like stress (physical or chemical) and nutrient starvation. Pupylation is also involved in iron homeostasis, which is necessary for cellular metabolism and the normal growth of bacteria. Pupylation is a posttranslational modification through which intrinsically disordered proteins are tagged for proteasomal degradation. Although this process is functionally reminiscent of ubiquitination in eukaryotes; it is carried out by a different set of enzymes in evolutionarily connected bacterial carboxylate-amine ligases. In this chapter, we will discuss the recent advances in the understanding of how proteins are tagged for proteasomal degradation in actinobacteria and its role in the survival of mycobacterium during pathogenesis in the host. Furthermore, we will examine the role of accessory factors associated with the proteasomal system in bacteria that function independently of proteolysis.

Keywords: Mpa (Mycobacterial proteasomal ATPase) and Dop (Deamidase of Pup), Mycobacterium, Proteasome, Pup, Pupylation.

INTRODUCTION

Regulated proteolysis is a fundamental process involved in posttranslational regulation especially to environmental changes, intracellular stress, and removal

* **Corresponding author Divakar Sharma:** CRF, Mass Spectrometry Laboratory, Kusuma School of Biological Sciences (KSBS), Indian Institute of Technology, Delhi (IIT-D)110016, India; Tel: +91-7906026680;
E-mail: divakarsharma88@gmail.com

Divakar Sharma (Ed.)

of disordered proteins. Removal and addition of functional groups on proteins increase the structural and functional diversification of proteins; these modifications alter the function, stability, localization, function, and regulation of proteins.The modification that targets the proteins towards proteasomal degradation affects the stability of proteins and has been extensively studied amongst protein modifications. Bacteria depend exclusively upon compartmentalized protease complex to degrade the proteins as they lack the sorting process in contrast to eukaryotes. In bacteria, the classical protease complex is Lon, Clp and the membrane attached Ftsh protease complex, the homologs of these proteases are also found in mitochondria and chloroplast. In addition to this, Mycobacteria and some other actinobacteria possess proteasome [1 - 3] which is not found in other bacteria. Interestingly, it is not essential under the mycobacterium culture condition [2, 4] though it persists in mycobacteria suggesting its important role in the survival of mycobacteria in the host in a specific environment [1, 2]. Pupylation, a posttranslational modification firstly observed in *Mycobacterium tuberculosis* and *Mycobacterium smegmatis* [5], suggests that prokaryotes also employ the macromolecular tags. Several lines of evidence demonstrate that modification of bacterial proteins with Pup occurs by a different pathway in contrast to eukaryotic ubiquitination [6]. Pupylation involves two homologous sequential events that involve different enzymology, Dop (deamidase of Pup), deamidation of C-terminal glutamine residue on Pup into glutamate being the first step, and then it attaches to target proteins through PafA (Proteasome accessory factorA). The various roles of pupylation have been identified for bacterial physiology is the degradation of a pupylated substrate; one of the most notable roles essential for virulence of *Mycobacterium tuberculosis*. Pupylation renders the proteins to proteasomal degradation, but not all Pup-conjugated proteins undergo this fate [7, 8]. Pupylation also regulates the activity of Mpa (Mycobacterial proteasomal ATPase) by rendering it functionally inactive [8, 9]. Further existence of depupylation in actinobacteria suggests broader role of pupylation in the bacteria [5, 10], showing that the signaling behind this process in future may help in a better understanding of the role of pupylation in the bacteria.

Bacterial Proteasome: An Evolutionary Precursor of Eukaryotic Proteasome

The proteasome is found in all three domains of living organisms that carry out the removal of damaged, unfolded and non-functional proteins. The existence of 20S CP (20S core particle) proteasome is exclusively found in actinobacteria, adopted by horizontal gene transfer during evolution [2, 4, 11]. Genes that encode this proteasome complex were not observed in standard culture conditions; hence evolution selected this proteasome for a specific condition [2, 12]. The first indication of the presence of bacterial proteasome comes from the study of

Frankia [13]. Genes (prcA and prcB) encoded to the proteasomal subunits were characterized by *Frankia, Steptomyces coelicolor* and *Rhodococcus erythropolis* [14 - 16]. The proteasomal degradation system in mycobacteria was firstly investigated in *Mycobacterium smegmatis* [4]. Lack of degradation activity in *Msm* ΔpcrB strain has been observed in *Mycobacterium smegmatis*, which suggests the presence of functional proteasome in mycobacterium. Further, the identification of prcA and prcB in *Mycobacterium tuberculosis* reveals that the proteome supports this notion [17]. Bacterial proteasome, the heptameric barrel-shaped structure, contains two rings of homo-heptameric of α (prcA) and two rings of β (prcB) subunits [3, 18, 19]. As similar to eukaryotes, the presence of threonine nucleophile in the β subunit of the bacterial proteasome is the cause of catalytic activity [20]. Gene *mpa* encodes an ATPase (equal to Regulatory particle of eukaryotic proteasome) found to co-localize with proteasome encoding gene (CP) in actinobacteria [21]. *Mycobacterium tuberculosis* lacking *mpa* gene is sensitive to nitrosative and oxidative stress. Similar phenotypes have been observed by the addition of proteasome inhibitor in mycobacterium culture [22]. In addition to this, a mutation in the mpa gene exhibits an altered level of some proteins in contrast to wild type strain, which provides evidence that the product of mpa gene plays a role in the proteasome degradation pathway [23]. ARC (AAA-ATPase ring-shaped complex) and Mpa perform a similar function as AAA-ATPase of a regulatory particle (RP) of eukaryotic proteasome while strong interaction between CPs and these proteins has not been observed [21, 24], suggesting that either it transiently interacts with these proteins or requires an accessory factor or co-factors.

Structure – *Mycobacterium* Proteasome

Structure of the *Mycobacterium* proteasome is just like archaeal CP and eukaryotic proteasome [2, 25], consisting of four stacked barrel-shaped structures, which is made up of two central β-rings flanked by alpha on both sides ($\alpha_7\beta_7$ β_7 α_7). *Mycobacterium* proteasome shows only 32% sequence identity with archaeal CP of *Thermoplasma* while 65% with bacterial CP of *Rhodococcus*. Despite this difference, the 3D structure of CP of these bacteria is superimposable with each other [26 - 28]. The proteolytic active sites have been found to associate with β-subunit of CPs [29], which synthesize as the N-terminal peptide; on auto-cleavage, it exposes the threonine acting as a nucleophile [30]. The proteolytic activity of *Mycobacterium* CP is somewhat different from the CPs of other actinobacterial species as Mycobacterium CP, which exhibits caspase and tryptic catalytic activity in contrast to archaeal CPs [2, 26]. Structural analysis of *Mycobacterium* CP disclosed that all proteolytic activity is performed by a single type of β subunit in contrast to eukaryotic CP [2, 26]. Alpha rings of bacterial CP form the pores on either end of central β-rings to prevent the undesired protein

CHAPTER 7

Overview and Challenges in Proteins Sample Preparation for 2-DGE in Bacterial Proteomics

Divakar Sharma[1,*], Juhi Sharma[2], Nirmala Deo[3] and **Deepa Bisht[3]**

[1] *CRF, Mass Spectrometry Laboratory, Kusuma School of Biological Sciences (KSBS), Indian Institute of Technology-Delhi, 110016, India*

[2] *Department of Botany and Microbiology, St. Aloysius College, Sadar Cantt, Jabalpur, 482002, India*

[3] *Department of Biochemistry, National JALMA Institute for Leprosy and other Mycobacterial Diseases, Tajganj, Agra, 282004, India*

Abstract: Sample preparation is the most crucial step in the proteome research of microbes. 2D gel electrophoresis (2-DGE) is a high throughput approach by which all proteins imprint on the gel in the form of spots or dots. Various protocols for the extraction of proteins exist in the literatures that are compatible with 2-DGE based proteome analysis from different microbes. Analysis of low abundance proteins, the solubility of proteins, and their resolution on the gel are some major issues observed with sample preparation and 2-DGE. To combat these issues, researchers have developed improved versions of the existing protocols using detergent/chaotropes for the enrichment of proteins during extraction along with the compatible chemical precipitation for better resolution of 2D gel pattern. In this chapter, we will discuss the overview, challenges encountered during the protein sample preparation, and their possible solutions in order to get a better 2-D gel of microbial proteins.

Keywords: 2-DGE, Detergents, Microbial proteins, Protein sample preparation.

INTRODUCTION

Protein sample preparation is one of the most crucial, yet problematic steps for high-quality resolution of proteins through two-dimensional gel electrophoresis (2-DGE) in microbial research. A lot of studies were conducted where high-throughput 2-DGE based approach was applied for the exploration of the various basics and clinical facts in different types of microbes [1 - 9]. Since the last twenty-five years, 2-DGE based approach has been used to analyze the whole microbial proteome as well as different fractions like a membrane, cytosolic, and

* **Corresponding author Divakar Sharma:** CRF, Mass Spectrometry Laboratory, Kusuma School of Biological Sciences (KSBS), Indian Institute of Technology-Delhi, 110016, India; Tel: +91-7906026680; E-mail: divakarsharma88@gmail.com

extracellular/secretory. Hence, 2-DGE has been considered the most accepted analytical method for resolving complex mixtures of proteins or expression proteomics [2 - 13]. Though various sample preparation protocols are available that are compatible with 2-DGE, researchers still keep on improving the strategies to get better 2D-gels [3, 14, 16]. In this chapter, we will discuss the overview and challenges faced during protein sample preparation and possible solutions to them.

Overview, Challenges and Probable Solutions

Sample preparation, which is the most crucial step for 2-DGE faces many problems and often encountered by the co-extraction of non-protein cellular components that can affect protein migration. While many methodologies have been described to alleviate these problems, defining optimal conditions for sample preparation from every new cell source is still somewhat a state-of-the-art process [14 - 16]. Proteins can be separated from non-protein contaminants by precipitation, or based on their characteristics, specific proteins can be enriched and then solubilized in sample buffer [3]. Because these techniques are often protein-specific or must be optimized for use, so a more universal approach for sample preparation is needed- one yielding a quality resolution of proteins in 2-D gels without laborious optimization and along with a higher tolerance for interfering non-protein components. Improved protein sample preparations have been used to reduce impurities as well to augment the low abundance proteins. Various types of selective protein extraction and protein precipitation with or without using detergents protocols were used to improve sample preparation for 2-DGE, which not only led to an increased number of spots but also increased resolution of the 2D gel patterns [3, 14 - 17].

Proteomics of lipophilic/membrane & associated proteins remains a major challenge in microbial research. Profiling of lipophilic proteins has been considered to be difficult within the bounds of conventional protocols for 2-DGE due to the limited solubility in aqueous buffer systems and relatively low abundance as compared to higher abundant cytoplasmic proteins. The nature of first dimension isoelectric focusing (IEF) requires solubilization of the proteins because they are subjected to an electric field in which they migrate to their isoelectric point. In addition to being highly hydrophobic, many integral membrane proteins tend to be very large and have transmembrane helices (TMH), which are major hurdles in the preparation of lipophilic/membrane protein samples for 2DGE [3, 16, 18].

Fractionation of cell extracts and subsequent analysis of these fractions remain a commonly used method for sample preparation. Several studies have reported the

extraction of lipophilic/membrane and membrane-associated proteins using ultracentrifugation to obtain purified cell wall and cell membrane fractions for analysis by 2-DGE coupled with mass spectrometry [3, 18]. Common for these studies are the pre-isolation of the membrane and cell wall fractions from bacteria and the application of different washing techniques before protein extraction by detergents. Proteins could also be separated from non-protein contaminants by precipitation or based on their characteristics, specific proteins can be enriched and then solubilized in sample buffer [3, 18]. These major issues can be overcome by firstly, selective extraction of membrane proteins into a specific extraction buffer that contains IEF compatible detergent. Second, maintaining protein solubility throughout loading onto IPG strips and then subsequent first dimension IEF separation.

Our group has developed an efficient and rapid method to enrich the lipophilic proteins from *Mycobacterium tuberculosis* for 2-DGE [16]. The use of nonionic detergent (Triton X-100) in sonication buffer during extraction enriched the solubilization of the lipophilic proteins. Further, enriched whole cell lysate was treated with Triton X-114 and subjected to direct phase separation of lipophilic proteins, without the need for pre-isolation of membranes. The optimized extraction buffer increased the solubility of lipophilic proteins and improved resolution on 2-DGE as compared to standard extraction buffer and therefore considered as an efficient and rapid protocol for lipophilic proteins extraction and separation on 2-DGE.

Although highly efficient membrane protein extraction methods are routinely carried out with a detergent (SDS) for one-dimensional PAGE, due to the charged head group, SDS is incompatible with IEF experiments. To overcome this problem, SDS solubilized samples are subjected to precipitation by organic solvent or acid for removal of SDS and lipids contaminants [14]. Despite these harsh treatments followed by precipitation and delipidation by solvent, there was improved protein recovery without any modification.

Isoelectric focusing with immobilized pH gradients leads to severe quantitative loss of proteins in the resulting 2-D gel, although the resolution is usually high. Due to poor solubility, membrane proteins have not been properly resolved by high-resolution 2-DGE. To improve the solubility of proteins, various denaturing solution detergents (both nonionic and zwitterionic) and chaotropes (urea and thiourea) were used to get better results [16 - 19]. Various studies showed a significant improvement in the extraction of mycobacterial membrane proteins by enriched differential centrifugation and alkaline treatment of crude membranes with sodium carbonate and urea [20, 21].

CHAPTER 8

Proteomics Based Biomarker Development Against Infections: An Overview

M. Madhan Kumar[1,*], Vivek Kumar Gupta[1] and **Divakar Sharma[2]**

[1] *ICMR-National JALMA Institute for Leprosy and other Mycobacterial Diseases, Tajganj, Agra-282004, India*

[2] *CRF, Kusuma School of Biological Sciences, Indian Institute of Technology, Delhi*

Abstract: Pathogens like viruses, bacteria, protozoa, fungi and helminths have caused and are causing diseases as well as the death of mankind since time immemorial. A proper understanding of the infection mechanisms will aid in designing efficient vaccines, diagnostics and drugs for efficient disease management. When compared to genes, proteins are involved in effector functions of the cell (of the pathogen as well as the host) and play a pivotal role in pathogen entry into host cells, its transmission and disease pathogenesis. Hence, the study of proteins will decipher the intricate communication networks involved in disease pathogenesis and will also elucidate their use as biomarkers in diagnostic and therapeutic scenarios. Thus, this chapter will highlight the proteomic and mass spectrometry-based approaches in comprehending the pathogen interactions with the host at the molecular level as well as biomarker development for use in diagnostics, prognostics and disease management.

Keywords: Biomarkers, Diagnostics, Mass spectrometry, Infectious diseases, Proteomics, Therapeutics.

INTRODUCTION

Proteins and Their Role in Infectious Diseases

Life is the Mode of Action of Proteins - Friedrich Engels [1]

Proteins form an integral part of life, whether in nutrition or in pathogen defence/pathogen attack. Proteins from the host in the form of cytokines, immunoglobulins, complement or other immune components (immune peptides like defensins, cathelicidins *etc.*), mount an attack against the pathogen and the pathogen, in turn, produces endotoxins or exotoxins to cause damage to the host. Moreover, pathogens also utilize the host cell components (like lipids, proteins,

[*] **Corresponding author M. Madhan Kumar:** ICMR-National JALMA Institute for Leprosy and other Mycobacterial Diseases, Tajganj, Agra-282004, India; Tel: +91-562-2331756; E-mail: madhanbiophile@gmail.com

Divakar Sharma (Ed.)

carbohydrates) for their survival and propagation. But, for successful survival, the pathogen must usurp the protein machinery of the host, or suppress the functioning of host proteins [2]. It is because, in the host, protein interactions play a pivotal role in cellular processes like transcription, translation, cell communication, cell adhesion, protein synthesis, cell signaling, *etc.* A disruption of this protein-protein interaction (PPI) network by the invading pathogen/establishment of new networks with host proteins by the pathogens play a major role in the progression of pathogenesis and the establishment of disease. Thus, a detailed study of PPI will unveil the pathological mechanisms underlying infection and will also aid in devising strategies for targeting the pathogen, help in disease diagnosis as well as prevention [3]. The protein-protein interactions can be direct (wherein the protein directly interacts with other protein) or can be indirect (the protein interacts with the other protein *via* intermediate molecules). A detailed study of this cross-talk between proteins (between host proteins or between host-pathogen proteins) in a cell has been termed as 'interactome'. The PPI networks have been studied in *Yersinia pestis, Bacillus anthracis, Franciscella tularensis* in order to define the proteins involved in pathogenesis and the associated mechanisms [4].

One of the challenges in interactome studies is the dynamic change in protein profile at different stages of the disease, which reflects the immune response of the host to the pathogen and also the expression of different virulent proteins by the pathogen during the course of disease [5]. Thus, the study of such interactions also takes into consideration the time as well as space for getting thorough knowledge about the interactomes (spatio-temporal analysis). Apart from the perspective of studying proteins in host-pathogen interactions, the study of proteins of the pathogens will help in discriminating between species and strains.

Apart from this aspect, functions and protein interactions are regulated by post-translational modification (PTM) of proteins. Such regulation by PTM on the host or pathogen proteins plays an important role in the infection progression and outcome. Regulation of host and pathogen PTM signalling deserves importance during infection as it impacts the changes in protein interactions, subcellular localization and, ultimately, the replication of pathogens. Among the PTMs, phosphorylation plays an important role as the pathogens are known to encode kinases, which can manipulate the host kinase signalling pathways. Thus, if proteomic technologies are applied efficiently in studying PTMs, they will help in unravelling pathogen-specific PTM signalling cascades, which can be targeted in therapeutics for a favourable outcome.

Although proteomics *per se* has contributed a lot to understanding proteins, integrating it with other molecular biology and biochemical methods has yielded

more information about infections. A combination of different 'omics' technologies like genomics, transcriptomics, complements the information obtained from proteomics and has made the understanding of infection a better way. Thus, multiple 'omics' technology approaches have given an enhanced insight into the process of pathogenesis in the case of infectious diseases.

Techniques Used for Studying Protein-protein Interactions (PPI)

The most common technique used for studying PPI for studying host-pathogen interactions is immunoaffinity purification coupled with mass spectrometry (IP-MS). This method has helped in identifying proteins involved in the interaction between pathogens and the host. It has also aided in elucidating cellular defense mechanisms of the host [6]. The method involves isolating the proteins by using antibodies raised against them. The antibodies are bound to sepharose or agarose beads in which the cell suspension containing the protein of interest is passed through, and the unbound proteins are washed off. Once the specific protein gets bound to the antibody, it is eluted by using competing epitope peptides for the antibody or by varying the pH or by the use of salts to dissociate the antigen-antibody complex. The dissociated protein is then subjected to mass spectrometry to identify the proteins. Thus, this technique can be useful for enriching proteins of low abundance, identifying unknown proteins in the complex. In Immunoprecipitation of protein complexes, the antibody (primary) is directed towards one of the proteins in the complex. Once the antibody is bound to the protein of interest, sepharose beads containing secondary antibodies are added to this mixture, wherein the secondary body attached to the beads bind to the primary antibodies bound to the protein complexes. This is centrifuged (washing) to precipitate the bead containing protein complexes from other proteins. The protein complexes are then eluted using the methods as said above, and then the proteins are subjected to mass spectrometry for identification to know about the interacting proteins in the complexes (Fig. **1**).

In certain circumstances, the proteins of interest are tagged with epitopes, and then antibodies against the epitopes are used for isolating the proteins (wherein no antibodies can be raised against a particular protein, epitope tagging is employed). The epitope tagging method is useful for studying proteins, which are very low or for studying novel proteins. After this process, the purified proteins are identified by the proteomic approach wherein the proteins are cleaved into peptides (by trypsin) and identified by mass spectrometric approaches.

CHAPTER 9

Microbial Metalloproteome: Approaches and Biomedical Application in Microbial Antibiotics Resistance

Saroj Sharma, Monalisa Tiwari and **Vishvanath Tiwari***

Department of Biochemistry, Central University of Rajasthan, Ajmer-305817, India

Abstract: Microbial metalloproteomics involves a detailed analysis of the proteins that have metals as an important part or known to bind to metals in biological samples. Recent updates showed that metalloproteome helps in understanding the role of different environments/conditions in the survival of microbes and is involved in microbial pathogenesis. Microbial metalloproteomics could also be used in understanding the resistance mechanisms of microbes. We have explored different metalloproteomic approaches such as inductively coupled plasma-Mass spectroscopy (ICP)-MS, X-ray absorption/fluorescence, radionuclide, and bioinformatics. We have also discussed the role of metalloproteins such as metallo-beta-lactamases in microbial drug resistance, the alternation of the microbial proteome in response to the metal, and their role in host-pathogen interactions. We have also surveyed different therapeutics targeting the microbial metalloproteins. Current advancements in the metalloproteome would help in understanding the mechanism better and the adequate role of metalloenzymes and proteins in conferring the drug resistance in microbes.

Keywords: Host-pathogen interaction, ICP-MS, Metalloproteome, Metallo-beta lactamases, Metalloproteomic approaches, Metal inducible proteome, Microbial drug resistance, Microbial therapeutics, X-ray absorption.

INTRODUCTION

Metalloproteome is a very common term used for a protein that contains metal ions. Metals form a very important part of our biological system, and their incorporation into the protein ensures the proper functioning of the enzyme. For example, Zinc containing proteins have an important role in the biological world, including fixation and synthesis of Carbon, and regulation of redox reactions [1]. Zinc is the most abundant part of the biological world as it represents 10% of the

* **Corresponding author Vishvanath Tiwari:** Department of Biochemistry, Central University of Rajasthan, Ajmer-305817, India; Tel: +91-850-300-2573; E-mail: vishvanath@curaj.ac.in

Divakar Sharma (Ed.)

entire proteome, whereas the cellular zinc content remains in submillimolar ranges [2]. Metalloproteins have various roles in the cell, like storage and transport of proteins, enzymes, and signal transduction [3]. Studies have projected that about 30% of protein candidates are known to have metal ions as cofactors [4]. This field has started gaining importance in the current scenario, such as the role of metalloproteins has also been documented in Alzheimer's disease [5] and it helps in predicting the function of metal ions in the molecular mechanism of diseases. Different metals are required for diverse biological functions (Table **1**). Metalloproteomics is an area that includes various approaches covering the expression of metalloproteins and their variations in a biological system [6]. The field of metalloproteomics offers novel insights into the basic biological processes. To analyze the potential of this field, it is very important to look forward to the methodological advances [7].

Table 1. Biological function of various metal ions that are found abundantly in the living system. (Chemistry of the Elements, Greenwood 2004).

Name of the Metal	The Biological Function of the Metal
Zn	A part of metalloenzymes like carbonic anhydrase, Calcium ion absorption, and structure maintenance
Cu	Oxygen transport, electron transfer, redox reaction catalysis
Ni	Hydrogenase, hydrolase, iron absorption
Co	Development and maintenance of blood vessels, skin, bones and joints, Redox catalysis
Fe	Redox catalysis, oxygen transport and electron transport
Mn	Have antioxidant property as it is a component of Superoxide dismutase, Nitrogen fixation
Mo	Nitrogen fixation, oxo transfer, electron transfer
Mg	Protein synthesis, muscle and nerve, oxidative phosphorylation, and glycolysis
Ca	Cell signaling, structure, and carrier of charge

EMERGENCE OF METALLOPROTEOMICS

Metals administered in our body unnaturally (through environmental contamination/ with drugs) are known to have various effects on the body. The varying coordination numbers, sites, and solvent accessibilities of these sites ensure the current placement of metals in terms of location, concentration, and time of delivery for human health [8]. One of the most common examples is zinc (Zn(II)), representing 10% of the proteome [2]. Zinc has an important function in various pathways, like fixation of carbon, biosynthesis, and redox regulation. Though Zinc is an essential metal, overloaded amounts of Zn(II) are toxic for the body because it competes for binding with the metalloproteins destined for Fe(II),

Mn(II) and Co(II) [3]. To overcome this problem, the cell over-expresses zinc-containing proteins to manage the cytosolic concentration of zinc in the free form [8]. This has led to a hypothesis that as soon as the zinc ion enters the cells, it changes the sites of presence, and in the precisely same manner, copper shifts its position concerning the binding efficiencies in the cytosol [9]. A complex network of interacting proteins called "interactomes" manages the zinc distribution in the sites that are known to compete with each other [10]. But the pathway of the zinc ions to reach their targets is still undiscovered. The uptake and efflux of zinc (zinc interactome) are clear [11]; beyond this point, the pathway of zinc transport is entirely unknown. For example, the secretory granules of pancreatic β-cells store insulin in the form of zinc-coordinated hexamers [12]. The abnormal insulin secretion due to the impairments of β-cells is the crucial factor for the development of glucose tolerance to the disease type-2 diabetes [10]. Another example is the entrapment of zinc in senile plaques during Alzheimer's [13]. The plaque formation takes place due to the aggregation of amyloid-beta peptides (Aβ), and Zinc facilitates the seeding by binding to Aβ [14]. Tracing the path utilised by Zinc to finds its way to Aβ can reveal information about the neural pathogenesis of Alzheimer's disease; hence suggesting new targets to develop drugs [10]. For a better understanding of the working of these metals, we need new techniques. Imaging is one of the reliable tools that provide information about the concentration and localization of various minerals in the cells. The development of selective metal-ion sensors with optical-fluorescence is one of the primary examples. Hence new fluorescent indicators have been synthesized by measuring the concentration of free Ca^{2+} in the cytosol [15]. Detectors for Cu, Zn, Fe are based on the usage of photochemical tools like fluorescent sensors with photocaged complexes to look into the details of homeostasis and signalling mechanism [16]. X-ray fluorescence microscopy (XFM), due to the deep penetrating activity of X-rays, shows marvellous trace element sensitivity [17]. The specimens can be observed in their natural state as there is no requirement of sectioning in XFM.

Among the techniques developed for the imaging of metals at the cellular level is the coupling of liquid chromatography with inductively-coupled-plasma mass-spectrometry (ICP-MS). ICP-MS has been used as a detector in chromatography and electrophoresis. It is supported by electrospray and Matrix-assisted laser desorption/ionisation MALDI MS, being one of the top analytical techniques [18]. Many other approaches are yet to be explored, and the current existing has its advantages and limitations.

CHAPTER 10

Proteomics of *Mycobacterium Tuberculosis*: An Overview

Anil Kumar Gupta[1,3], **Divakar Sharma**[2] and **Amit Singh**[1,3,*]

[1] *Department of Microbiology, All India Institute of Medical Sciences, Saket Nagar, Bhopal, India*

[2] *CRF, Kusuma School of Biological Sciences, Indian Institute of Technology, Delhi, India*

[3] *All India Institute of Medical Sciences, New Delhi, India*

Abstract: Tuberculosis (TB) is a universally prevalent disease caused by an aerobic, gram-positive bacterium *Mycobacterium tuberculosis* (*M. tuberculosis*). It has continued to pose a significant threat to human health. The emergence of multi-drug resistance (MDR) strains of *Mycobacterium tuberculosis* (*M. tuberculosis*) has further worsened the situation worldwide. Its genome has been primarily focusing on the past exploration of the molecular basis of disease. The genetic information or genes are transcribed into mRNA that is then processed, spliced, and translated into a single or multitude of proteins. Proteomics is the large-scale study of proteins, focusing on their structure and functions. For understanding the biology of any living organisms including humans, we need to decode the information encoded by proteins or its associated genes, as it reflects the true status of the cell. Although proteins play a major role in the whole life of the organism, its profile may vary from cell to cell due to spontaneous changes or biochemical interaction of microbial genome with environments. Even an immense amount of DNA/gene sequences data has been deposited by the scientific community in the databases, which is very useful in determining the virulencity, pathogenicity of the organisms. It simply contains a complete sequence of genomes that lacks its usefulness to illuminate biological function. The regulation process of a single cell involves complex mechanisms, including a multitude of metabolic and regulatory pathways for its survival. To date, no strict linear relationship has been documented between genes and proteome of the cell. The available technologies, *i.e.* microarray, identify a large number of differentially expressed genes very quickly. However, these have failed to identify the functional significance of the associated genes. In this chapter, we highlight prospects of the advances in the proteomics study that could be beneficial to mycobacterial research. Also, it will provide information to explore the possibility of protein-based biomarkers in the development of new diagnostic therapeutics for TB.

[*] **Corresponding author Amit Singh:** Department of Microbiology, All India Institute of Medical Sciences, Saket Nagar, Bhopal, India and All India Institute of Medical Sciences, New Delhi, India; Tel: 011-26549372; E-mail: amit4us@gmail.com

Divakar Sharma (Ed.)

Keywords: Biomarkers, Drug discovery, Drug targets, *M. tuberculosis*, Metabolic pathways, Multidrug-resistant TB (MDR-TB), Proteomics, Structural biology, Structural genomics, Vaccine, XDR-TB.

INTRODUCTION

The "genome" has been a primary focus of the past exploration of the molecular basis of disease. The genetic information coded in genes is first transcribed into mRNA, which is then further processed, spliced, and translated into a multitude of proteins by strict complex mechanisms. Proteomics is a large-scale study of proteins, particularly their structure and functions. This term is coined to make an analogy with genomics. To understand the biology of any living things present on the planet, we need to know about macromolecules such as proteins that are the functional unit of the cell. Thus, proteins reflect the true status of the cell, and its differential expression may rule out the health status of the organisms. The proteome profile varies from cell to cell due to spontaneous changes or biochemical interaction of microbial genome with environmental conditions. Proteins play a central role in the life of an organism. The regulation process of a single cell involves complex mechanisms, including a multitude of metabolic and regulatory pathways for its survival. Due to continuing development and improvement in the proteomics technology, difference-gel electrophoresis (DIGE), two-dimensional gel electrophoresis (2D-GE) and liquid chromatography linked to mass spectrometry (LCMS) are now able to detect multiple proteins from complex biological samples (blood, tissues) with higher sensitivity and specificity. Further development of mass spectrometer combined with a technique to remove interfering proteins, such as immunodepleting, will be able to detect even lower concentrations in the samples. Besides, the protein microarray system (chips) is developed as a matrix-support surface to facilitate the binding of selected proteins in matrices of mass spectrometry. These protein chips were developed to bind specific proteins found from the biological samples and identify them very quickly [1]. Proteomics, also known as "complementary to genomics", has been explored enormously for the development of new drug targets. The majority of the currently available drugs target specific proteins of infectious agents than nucleic acid.

The recent development in gel electrophoresis and mass spectrometry has significantly facilitated separation, purification, identification phosphorylation analysis of proteins [2].The transcript detection by mRNA profiling does not reflect the all regulatory process clearly due to post-translational processes altered the amount of active protein insight the cell. The problem could be solved effectively by proteome analysis of the cell. Proteome analysis not only identifies insight changes but also identify post-translational modifications, protein-protein

interactions, distribution of cellular and subcellular proteins, and progressive pattern of gene expression also. The utility of differential and/or functional analysis of protein profiling is to acquire hidden cellular information at the atomic level, which will enhance our understanding of the cellular pathways and their inter-relationship with cells/tissue. Through proteomics, it had already been uncovered and validated various potential drug targets against numerous diseases such as TB, HIV, cancer, *etc.*, some of them are available commercially. The current era of proteomics in TB is now beginning to inspect how proteomic technology can help the clinician in biomedical science.

By combining, the advancement of proteomic tools available to the researchers, the entire proteome can now begin to be unraveled the molecular basis of infectious disease; identify the novel biomarkers sets and potential drug targets also. The power of the proteomic approach has been most effectively used to date in the fields of cancer and TB research [2 - 9]. Celis and co-workers [10] have utilized 2-dimensional gel electrophoresis and MS analysis in cancer research to find differentially expressed protein between diseased and healthy tissue including normal urothelium *vs* squamous cell carcinoma (SCCs) that has defined some of the stages involved in the squamous differentiation of bladder transitional epithelium. High through output mass spectrometric analysis of human plasma/serum proteomics is emerging as a potent technique for the identification of different protein profiles in cancer patients. Utilizing a proteomic approach, Nishio *et al*. [11] identified noticeable differences in the phosphorylation status of specific nuclear proteins between drug-resistant and sensitive cell lines. Proteomic analysis was successfully applied to the development of erythromycin resistance in *Streptococcus pneumoniae* (S. pneumoniae) that revealed a significantly increased amount of Glyceraldehyde 3-phosphate dehydrogenase (GAPDH) in the M phenotype of drug-resistant bacterium. It was hypothesized that the GAPDH may provide reducing equivalents for the active efflux mechanism or may even be directly involved in the membrane transport mechanism, which regulates the erythromycin resistance in the M phenotype [12 - 14]. A new function of PstS, a subunit of the phosphate ABC transporter in *S. pneumoniae* penicillin-resistant strains, was revealed by proteomic analysis [15]. Xu and his colleagues [16] reported three antibiotic-resistant proteins of TolC, OmpC, and YhiU, together with the antibiotic resistance-related proteins of FimD (precursor), LamB, Tsx, YfiO, OmpW, and NlpB, which responded to tetracycline and ampicillin resistance in *E. coli* K-12 through proteomic technology.

Proteomics approaches have also been applied to the numerous communicable and non-communicable human diseases such as TB, HIV, Malaria, Leishmaniasis, Toxoplasmosis, cancer, diabetics, *etc.* for development of diagnostic/therapeutic regimens. Numerous potential drug targets were identified by performing

CHAPTER 11

Proteome Based Insights in Drug-Resistant *Mycobacterium tuberculosis*

Apoorva Narain[1], Surya Kant[1], Rikesh Kumar Dubey[2], Kanchan Srivastava[1] and **Anand Kumar Maurya[3,*]**

[1] *Department of Respiratory Medicine, King George Medical University, Lucknow, India*

[2] *Division of Microbiology, CSIR-Central Drug Research Institute, Lucknow, India*

[3] *Department of Microbiology, All India Institute of Medical Sciences, Bhopal, India*

Abstract: *Mycobacterium tuberculosis* (MTB) has been an exceptionally successful human pathogen over the centuries, infecting almost one-third of the global population. An exponential increase in tuberculosis (TB) cases, mainly by the drug-resistant (DR) strains of MTB, has created an urgent need for identifying and developing new antituberculosis drugs acting *via* novel mechanisms. The multi drug-resistant (MDR) TB and extensively drug-resistant (XDR) TB strains, accelerating through drug specific resistance amplification, are resistant to a majority of antibiotics used in the treatment and are challenging to remove from the host's system. Since proteins are the functional beings of the biological arrangement, they make promising drug targets for immunodiagnostics or therapeutics. To identify and characterize such novel proteins, which directly or indirectly regulatedrug resistance in mycobacteria, proteomics approaches could be successfully employed. Serological techniques like immunoassay have higher chances of rendering false positive or false negative results and hence could be rectified by using more sophisticated techniques like mass spectrometry. In the past two to three decades, proteomics-based approach has seen a pivotal rise. The application of proteomics-based approaches has helped to gain insights into MTB and its relevance to clinical science. They have aided in the identification and characterization of novel proteins. To have a better understanding of pathophysiology of MTB, proteome-based science could help simultaneously in the identification of proteins, which can be potential targets. Recent progress in the area of proteomics has opened up the doors to address many previously unanswered questions, with studies on DR-TB being no exception. The Beijing family of MTB forms an interesting candidate for proteomic analysis as it constitutes 13% of the global isolates and has higher chances of acquiring drug resistance. Proteomics can play an important role in the discovery of biomarkers for TB and other diseases. Also, it can aid in the development of effective vaccines as well as simple, rapid, and cost-effective tests for the diagnosis of TB, which are crucial for the management and control of the disease.

[*] **Corresponding author Anand Kumar Maurya:** Department of Microbiology, All India Institute of Medical Sciences, Bhopal, India; Tel: 0755 2672328; E-mail: anandmaurya1@gmail.com

Divakar Sharma (Ed.)

Keywords: Diagnosis, Drug resistance, Host-pathogen interaction, *Mycobacterium, tuberculosis*, Proteomics, Tuberculosis.

INTRODUCTION

Mycobacterium tuberculosis (MTB) is a facultative intracellular pathogen that has the ability to survive inside the macrophages of its host. Though tuberculosis (TB) has been known to have emerged around 70,000 years ago, it still remains a global issue. TB has a mortality of ~70%, if left untreated or improperly treated.

It is the leading cause of death by a single infectious agent and has killed about 1 billion people over two centuries. The impact of the disease is such that it has even slowed down the GDP of countries [1]. Robert Koch, in 1882, used the term tuberkulose to describe his discovery of the bacterium he called Tubercle bacillus. He was conferred with Nobel Prize in physiology/medicine for this landmark discovery in 1905. This discovery, in association with the discovery of tuberculin, the Bacillus-Calmette Guérin (BCG) vaccine, and anti-tuberculosis drugs in the years 1890, 1908, and 1943, respectively, paved a way to combat the deadly disease of tuberculosis. The main causal organism of TB in humans is *Mycobacterium tuberculosis*, though other *M.tuberculosis* complex (MTBC) members (*M.bovis, M.microti and M.africanum*) can also cause tuberculosis. MTB, being an intracellular pathogen, interacts with host's immune response to cause major immunopathological symptoms. To establish successful infection, MTB have adapted several strategies to survive intracellularly, such as inhibition of phagolysosomal fusion, maturation of phagosomes, drug efflux, and targeting mitochondria to disturb the balance of pro-apoptotic and anti-apoptotic factors. All these cellular events are primarily regulated by bacterial proteins. Nearly 40% of the proteins identified in bacteria are uncharacterized in MTB and approximately 50% of the proteins belong to this earlier mentioned category [2]. Proteins have a responsible astonishing range of biological functions and roles, such as structural proteins, enzymes, and transportation. To promote a healthy growth rate and intracellular survival, every protein in the cell interacts with the cellular environment. Every protein is expressed *via* a gene that forms an interface with all other cellular components to promote the growth and survival of a cell. Hypothetical or uncharacterized category of proteins could be unique to bacteria since they normally have no hits in other genomes, such as hypothetical proteins of the PE/PPE family in mycobacteria [3]. The term proteome was first introduced by Mark Wilkins in 1986 and the most noteworthy contribution was made by R. Aebersold *et al.* in the development of proteomics of mycobacteria. In the 20th century, top-down proteomics was primarily employed for MTB. It works on the principle of separating proteins on the basis of their physical and chemical properties, *i.e*., gel electrophoresis and gel filtration and ultimately identifying

them through mass spectrometric (MS) methods. This way, around 3% of the total proteome of MTB (100 mycobacterial proteins) has been identified. Another type of strategy is known as bottom-up proteomics, which works on the principle of isolating total protein from the sample, proteolytically cleaving it into peptides and then analyzing using high performance liquid chromatography, coupled with tandem mass spectrometry (HPLC- MS/MS) [4, 5].

DRUG RESISTANT TUBERCULOSIS

Mycobacterium spp. possesses intrinsic resistance against a varied range of antibiotics, mainly due to the presence of thick cell walls that are rich in mycolic acid and drug efflux pumps. Some antibiotics become inactive after they are cleaved or modified upon penetrating the cell membrane. The absence of horizontal gene transfer in MTB directs towards chromosomal aberrations being responsible for drug-resistant phenotype instead of resistant plasmid or transposons. Mycobacterial cell wall composition and a low number of porins heavily contribute to cell membranes less permeability to compounds. The lipid layer of the cell wall is linked to the peptidoglycan layer *via* arabinogalactan. Small hydrophilic compounds enter the extremely hydrophobic cell wall by water-filled porins [3, 6]. Drugs able to penetrate the thick cell wall of mycobacteria are enzymatically inactivated inside the bacteria. β-lactamases are one such group of enzymes. They target the antibiotics with β- lactam ring by hydrolyzing them. MTB protein, BlaC (β-lactamase), is localized in the periplasmic space either in free form or as a liposome. Clavulanate, a known inhibitor of β- lactamase, irreversibly inhibits the activity of BlaC. MTB β-lactamase has broad-spectrum substrate specificity and the upscaling of drug-resistant cases in TB has opened the doors of discussion for β-lactam antibiotics to be introduced in the drug regimen. Antibiotics are also rendered inactive by chemical modification, acetylation, or methylation [6, 7]. Enhanced intracellular survival protein (Eis) of MTB acetylates second-line TB drugs kanamycin A and capreomycin, thereby inactivating them in the process [6]. Efflux pumps (EPs) are the membrane-associated transporters that pump out a wide range of molecules from the bacteria system, including drugs, into the outside environment. Efflux systems in *M. tuberculosis* are essential for intracellular growth in macrophages and get activated by immune or drug pressures. Upon activation, they decrease the accumulation of drugs, thereby reducing the cytoplasmic concentration to sub-inhibitory levels. Bacterial efflux pumps have been divided into five super-families: resistance nodulation division (RND), multidrug and toxic compound extrusion (MATE) family, ATP-binding cassette (ABC), small multidrug resistance (SMR), and the major facilitator super-family (MFS) [2]. The average transcription and replication rate in mycobacteria slows down in the presence of drugs, which cause the reduced ribosomal proteins' expression and fluoroquino-

SUBJECT INDEX

M

N

P

S

T

V

X