Thesis on

"Identification and Characterization of Biomarkers using NMR based Metabolomics - Implication to Disease Diagnosis and Treatment Monitoring"

Submitted for the Degree of

Doctor of Philosophy

in Biotechnology

By

ATUL

Co- Supervisor

Dr. Dinesh Kumar

Assistant Professor

Centre of Biomedical Research,

SGPGIMS Campus,

Lucknow, (U.P.) India

Supervisor

Dr. Anand Prakash

Assistant Professor

Department of Biotechnology,

B.B.A. University,

Lucknow, (U.P.) India

Submitted to

School of Biosciences and Biotechnology

Department of Biotechnology

Babasaheb Bhimrao Ambedkar University

Lucknow-226025, UP, INDIA

2017

Abstract

Increased prevalence of lifestyle related diseases is an impending threat to public health and calls for effective prevention and treatment strategies, as the exact molecular mechanisms behind these are unclear and are difficult to explore.

The existing clinical and pathological tools for diseases are insufficient for accurate response prediction, or for personalized treatment. Presently, there is a compelling need for identification and development of new biological markers to achieve a new era of predictive, preventive, targeted medicine and personalized therapy.

The concentrations of metabolites in the cell alters in response to a variation in enzyme concentrations and activities. Hence metabolites level changes much more rapidly than do the levels of enzymes or RNA. Thus, measurements of the concentrations of metabolites provide a good measure of cell metabolism and hence identify the response of a cell or tissue to genetic or environmental changes, much more rapidly and precisely than do changes in proteome or transcriptome. Altered cellular metabolism is a major factor in the pathogenesis of many diseases like cancer, cardiovascular disease, microbial infection, etc. and has become a major area of biomedical research.

Metabolomics is the "systematic study of the unique chemical fingerprints that specific cellular processes leave behind," the study of their small-molecule metabolite profiles, endogenous or exogenous. The metabolome represents the collection of all metabolites present in a biological sample i.e. cell, tissue, organ or organism, which are the end products of cellular processes. Thus, while mRNA gene expression data and proteomic analyses do not tell the whole story of what might be happening in a cell, metabolic profiling provides an instantaneous snapshot of the physiology of that cell. One of the biggest challenges posed towards systems biology and functional genomics is to integrate proteomic, transcriptomic, and metabolomic information to give a complete picture of living organisms.

Metabolomics -a newborn cousin to genomics and proteomics- is an analytical approach to metabolism that involves quantitative and comparative analysis of concentration profiles of low molecular weight metabolites and their intermediates in affected biological systems (typically urine, blood plasma/serum, cell lysates, or tissue extracts). Genomics, transcriptomics and proteomics analysis complemented with metabolomics information, offers the potential to understand the whole biological system including health or disease processes operating in that system –so-called systems biology approach. With its ability to discover disease-related biomarkers and underlying biochemical processes, today, metabolomics is used virtually in all aspects of biomedical research aiming to improve the understanding of the health and disease processes. The whole paradigm is based on the fact that

a pathophysiological condition or therapeutic intervention results in a distinct and characteristic change in the biochemical composition profiles of biofluids and metabolomics aims to identify these changes. The biochemical changes -that correlate with a disease (or disease type/grade) and treatment response- then allow the clinical researchers to improve diagnosis and treatment of disease including, early disease detection, monitoring response to treatment and patient stratification for treatment. The molecular biomarkers validated on wide-range of human populations form the basis for new clinical diagnostic assays.

The most information-rich techniques currently employed in metabolomics studies are Mass Spectrometry (MS) and Nuclear Magnetic Resonance (NMR) spectroscopy. Among these two techniques, MS is highly sensitive, with detection limits in the picogram range, however, for MS-based metabolomics studies, it requires a separation step (liquid chromatography/gas chromatography) before the MS detection and has the possibility of losing some constituents. Further, this requirement dictates that each sample will require hours per MS analysis. The overall throughput is further hampered by many unsolved problems such as: (a) non-uniform detection caused by variable ionization efficiency, (b) lack of well-established and standardized methods or procedures (as it requires optimization of separation conditions each time), (c) lack of a universal database due to different ionisation mode and (d) the difficulties still met in the identification of novel/unknown metabolites produced in response to a given therapeutic intervention. Although less sensitive (with detection limits in the low micromolar to submicromolar range) than MS, NMR is generally appreciated for metabolic analysis because of its versatility for analysing metabolites in the liquid state (serum, urine, and so on), in intact tissues (for example, tumors) or in vivo (e.g. brain, liver, kidney, and so on). The most important strengths of NMR spectroscopy are reproducibility, the ability to quantitate compounds in unmodified complex biological mixtures and the ability to identify unknown metabolites.

Compared to other analytical and biochemical analysis methods used for metabolomics studies, NMR offers several clear advantages: (a) it is applicable to a variety of biological and clinical samples, tissue extracts, and even cell lysates, (b) it is rapid, quantitative, and offers the potential for high-throughput (i.e. analysis of >100 samples/day is attainable), (c) it is least-destructive (i.e. the prepared sample can be used in multiple consecutive NMR experiments, or after the NMR experiments are completed, the sample can be analyzed by other analytical techniques), (d) it is unbiased (i.e. all protonated metabolites present In a biological mixture are detectable irrespective of their physical properties) (e) Peaks in the NMR spectra can be consistently assigned to the particular metabolite, based on their chemical shifts and multiplet patterns and (f) lastly it requires virtually no sample preparation and provides highly reproducible results. These are the reasons that NMR has become the method of choice for studying metabolic alterations associated with distinct human pathologies and also gaining tremendous

popularity in pharmaco-metabolomics studies as well. Today, the sensitivity of NMR is also not a major issue; even nanogram detection limits are possible with appropriate instrumentation and novel pulse methodology. Further, the recent technological advances in cryogenically cooled probe technology, the miniaturizing of sample probe head (i.e. with microvolume probe it is now possible to analyze samples in 10 µl solution), higher field-strength superconducting magnets and high-performance radio-frequency coils have increased the sensitivity of NMR spectrometers by a factor of ~ 3–4.

There are two main approaches to perform NMR based metabolomics studies, that depend on whether the metabolites to be detected are known (Targeted Metabolomics), or they are unknown (Untargeted Metabolomics). Conducting a targeted or an untargeted metabolomic study is a critical issue for sample preparation, for the choice of the experimental setup and the statistical approach used. The basic underlying difference between both types of study is that the main objective in Untargeted Metabolomics is to detect the highest possible number of metabolites in the samples, where as Targeted Metabolomics is merely focused on detecting specific metabolites of interest or hypothesis driven.

The main objective in Metabolomics is to study all the metabolites within a biological sample in an unbiased manner. The metabolomics studies involve a number of specimens together for the analysis. For metabolomics, well-planned experiment design and execution are required along with univariate and multivariate statistical analysis is needed for pattern recognition and interpretation of metabolomics data. Overall, the NMR spectroscopy coupled with multivariate statistical analysis allows the identification of metabolic disturbances and specific metabolic pathways associated with the disease, but it also permits the identification of metabolic signatures which have their potential diagnostic and prognostic implications for clinical management of the disease.

The NMR-based metabolomics approach has already been helpful in identification of metabolic markers for neurological, cardiovascular disease, cancer, etc.. The detection of disease biomarkers plays a critical role not only in early disease diagnosis or risk prediction but also in classification and disease progression or assessment prognosis and treatment response. Establishment of these biomarkers in routine clinical use has the potential to provide insight into the pathogenesis of disease states and discover diagnostic markers for therapeutic targets. Thus, the metabolic biomarkers have the power to increase the overall survival and quality of patient life in, addition to saving huge expenses for the society.

The primary objective of the research presented in this thesis was to evaluate the use of high-resolution NMR spectroscopy together with multivariate analysis based metabolomics for identifying and characterizing potential biomarkers of the health-disease continuum. This thesis consists of original research work in which the applicability of NMR Metabolomics in identifying biomarkers of disease dependent changes has been explored. Metabolomic studies were performed on a broad range of

subjects ranging from healthy volunteers to patients with advanced stage of the disease along with animal models. The research objectives undertaken has been discussed briefly.

Objective 1: Acute myocardial infarction (AMI) also known as heart attack- is a serious medical condition and the devastating outcome is the sudden death of the patient within first few hours from the onset of symptoms. It is also the leading cause of cardiac-related hospitalizations and disabilities worldwide. The rapid detection of physiological transformations associated with AMI coupled with instant treatment to reset these changes and subsequent monitoring the response to treatment can greatly decrease the complications or mortalities associated with it. Therefore, there is an immense interest in the identification of reliable biomarkers which could improve the diagnosis and clinical management of AMI in emergency settings. In this context, we applied NMR-based metabolomics approach to investigate the altered serum metabolic patterns associated with AMI and sought to identify specific metabolites that could lead to rapid diagnosis of AMI and its timely management in emergency settings. Metabolic profiles of sera obtained from 42 AMI patients and 38 age/sex matched normal controls were analyzed using high-resolution 1D ^{1}H CPMG and diffusion edited NMR spectra coupled with multivariate statistical analysis. The analysis revealed significantly altered serum metabolome of AMI patients compared to normal controls cohorts. The up-regulated metabolites in AMI condition include arginine, glycine, tyrosine, phenylalanine, glucose, creatine, creatinine, lactate, N-acetyl glycoproteins and phospholipids, while the levels of amino acids (such as valine, alanine, glutamate, glutamine, threonine, and methionine), citrate, acetone, choline, glycerophosphocholine, trimethylamine-N-oxide and lipids/fatty acids were decreased. Receiver operating curve characteristics (ROC) confirmed the robustness and validity of these metabolic markers. The observed metabolic perturbations were associated with aberrant amino acid metabolism, affected protein biosynthesis, and profoundly dampened glycolysis, TCA cycle, and fat metabolism; alluding to dyslipidemia, systemic inflammation, heightened oxidative stress, necrosis and reduced energy biogenesis from all sources. We foresee that the serum metabolic patterns identified in this work would guide further studies aiming to identify biomarkers for rapid diagnosis of AMI and to define proper treatment for individual AMI patients.

Objective 2: Pyrazinamide (PYZ), is an essential component of first-line drug regimen used for the treatment and management of multidrug-resistant or latent tuberculosis. Most of the first-line drugs used for the treatment of tuberculosis are found to be hepatotoxic, which is a serious complication associated with the success of the therapy. Past studies uncovered that the hepatotoxicity potential was more with PYZ than any other tubercular medication. However, the mechanism of pyrazinamide-induced hepatotoxicity is still unknown to the researcher. Studies have shown that the drug is metabolized in the liver into pyrazinoic acid (PA) and 5-hydroxy pyrazinoic acid (5-OHPA); therefore, it is legitimate to consider that hepatotoxicity may be caused either by the parent drug PYZ or its metabolized products.

To evaluate this hypothesis, we performed NMR-based serum metabolomics using albino Wistar rats. The serum samples were obtained from control rats and those treated with PYZ, PA, and 5-OHPA. The metabolic profiles of these serum samples obtained using high-resolution 1D ^{1}H CPMG and diffusion edited NMR spectra were compared using multivariate statistical analysis to identify the drug-induced metabolic disturbances. The multivariate and pathway analysis revealed that the metabolic profile of rats treated with 5-OHPA was significantly more perturbed from the other two groups viz. PYZ and PA with respect to control. The significant metabolic perturbations evident in rats treated with 5-OHPA were characterized by an increased level of glucose, N- and O- acetyl glycoproteins along with the decreased level of lipids (HDL, LDL, and VLDL), creatine/creatinine, and glucogenic amino acids. The metabolic perturbations suggested energy deficit, oxidative stress, inflammation, and muscle degradation due to hypolipidemia associated with liver injury. As evident from our results along with the histopathological and SEM studies the compound 5-OHPA induced most significant changes in the biochemical composition profiles in rat sera. NMR-based metabolomics enables the rapid and accurate measurements of many metabolites in a single run than by using any routine biochemical methods including detailed analysis of lipoproteins and many other metabolic parameters. The present study has implication in human studies, patient stratification, decision making, and personalized medicine, as the drug-induced liver injury is a major complication associated with many treatment procedures in critical care and routine clinical management. It could be speculated that metabolomics might be helpful to improve the management and the decision-making process in patients with liver dysfunction or liver transplant susceptible to the drug-induced toxicity.

Objective 3: Rapid detection of microbial pathogens in clinical samples is crucial for directing appropriate antimicrobial therapy and improving patient care and associated outcomes. Every hour the appropriate treatment is delayed in patients in critical care has been shown to increase mortality by ~10-15 %. Therefore, earlier detection of the pathogenic microorganism has the potential to benefit patient care significantly. In this context, an efficient ^{1}H NMR-based method –named here as "Add Antibiotic to Detect by NMR" or "**AADNMR**"- has been proposed and utilized for rapid identification of bacterial/mycobacterial infection in PD effluent samples. However, the method has general applicability for rapid identification of bacterial/mycobacterial infection in a suspected clinical/biological sample. The method is based on the fact that the ring methylene protons of cyclic fatty acids –constituting the cell membrane of several species of bacteria and mycobacteria- resonate specifically between -0.40 and 0.68 ppm regions of the ^{1}H NMR spectrum. These cyclic fatty acids are rarely found in the eukaryotic cell membranes. Therefore, the signals from cyclic ring moiety of these fatty acids can be used as markers (a) for the identification of bacterial and mycobacterial infections and (b) for differential diagnosis of bacterial and fungal infections. However, these microbial fatty acids when present in the membrane are not readily detectable by NMR owing to their fast T2

relaxation. Nonetheless, the problem can easily be circumvented if these fatty acids become suspended in solution. This has been achieved by abolishing the membrane integrity using broad-spectrum antibiotics (including ampicillin). The suspended fatty acids are then detected by NMR to probe the infection. Thus, the method derives its name “Add Antibiotic to Detect by NMR or **<AADNMR>**.” The method has been tested here using both Gram +ve and Gram -ve bacterial strains, and finally, the utility of the method is demonstrated for discriminating bacterial and fungal infections. The present method will aid in timely diagnosis and differentiate between an infectious organism and strains.

The present thesis addresses the usefulness of NMR-based metabolomic techniques in capturing molecular signatures of the disease, drug metabolism, etc. and hence prediction of the metabolomic profiles. The particular advantages of the approach are its rapidity and reproducibility with little or no added technical resources to existing in vivo studies. New technological and methodological advances in combination with rapid generation of metabolomics databases and automation will further improve the information that can be obtained using this approach and will make a significant impact in pharmaceutical development and mechanistic exploration. With the ever increasing metabolomics applications and the type of applications for which metabolomics can be used. Thus, the application of metabolomics to both basic and clinical research will be imperative in the near future. In particular; metabolomics now plays an increasingly important role in functional genomics, disease diagnosis, toxicogenomics, nutrition and nutrigenomics, systems biology, environmental monitoring and even cellular imaging.