# 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***


BABASAHEB BHIMRAO AMBEDKAR UNIVERSITY
LUCKNOW
प्रज्ञा शील करुणा
ESTABLISHED 1996


***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**

**The universe bows to Lord Shiva.**

**I bow to Lord Shiva.**

# *Dedicated to,*

***My family,***

***Gyanwati Rawat and Saheb Das Rawat, sister Pratima, brother Amit, and everyone who stood by me through all walks of my life***

BABASAHEB BHIMRAO AMBEDKAR UNIVERSITY
LUCKNOW
प्रज्ञा शील करुणा
ESTABLISHED 1996

# बबासाहेब भीमराव अम्बेडकर विश्वविद्यालय
(केन्द्रीय विश्वविद्यालय)
विद्या विहार, रायबरेली रोड, लखनऊ–226025

# BABASAHEB BHIMRAO AMBEDKAR UNIVERSITY
**(A Central University)**
**Vidya Vihar, Rae Bareli Road, Lucknow-226025**

# CERTIFICATE

This is to certify that the thesis titled **"Identification and Characterization of Biomarkers using NMR based Metabolomics - Implication to Disease Diagnosis and Treatment Monitoring"** submitted by **Atul** is an original research work and has not been previously submitted in part or full for the award of any other degree or diploma to this or other university.

The thesis submitted to Babasaheb Bhimrao Ambedkar University Lucknow satisfies all the requirements stipulated in the Doctor of Philosophy (Ph.D.) regulations – 1999 as amended in 2010 and it is fit for submission and evaluation for the award of the degree of Doctor of Philosophy of the University.

Date:

**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

**Head of Department**

# CANDIDATE'S DECLARATION

I hereby declare that thesis entitled **"Identification and Characterization of Biomarkers using NMR based Metabolomics - Implication to Disease Diagnosis and Treatment Monitoring"** is an authentic research work carried out be me under the supervision of **Dr. Anand Prakash**, Assistant Professor, Department of Biotechnology, Babasaheb Bhimrao Ambedkar University, Lucknow and the Co-Supervision of **Dr. Dinesh Kumar**, Assistant Professor, Centre of Biomedical Research, Lucknow. The research work is original, and no part of this work has been submitted for any other degree or diploma.

All the above given information is true to the best of my knowledge.

Date:

Submitted by

**Atul**
Enrolment No. : 005/12
Department of Biotechnology,
Babasaheb Bhimrao Ambedkar University,
Lucknow, U.P., India

BABASAHEB BHIMRAO AMBEDKAR UNIVERSITY

# बाबासाहेब भीमराव अम्बेडकर विश्वविद्यालय

(केन्द्रीय विश्वविद्यालय)

विद्या विहार, रायबरेली रोड, लखनऊ–226025

# BABASAHEB BHIMRAO AMBEDKAR UNIVERSITY

**(A Central University)**

**Vidya Vihar, Raebareli Road, Lucknow-226025**

Letter No.-...87...../COE/BBAU/2015 Dated: ...14/05/15.....

## Ph.D. Course Work Certificate

This is to certify that **Mr. Atul**, Enrollment No. 005/12 Ph.D. Research Scholar, Department of Biotechnology of this University has successfully completed his Ph.D. Course work in the examination held during November, 2012.

(A.K. Maurya)
**Deputy Registrar (Exam.)**

# Table of Contents

# Abstract

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 like urine, blood-plasma/serum, cell lysates, or tissue extracts). While vast progress in the fields of genomics, transcriptomics, and proteomics has occurred, additional evidence of biological end points of human diseases is highly desired for disease diagnosis, treatment monitoring and for understanding the pathogenesis of many diseases. Metabolomics -because of its ability to reveal altered metabolism produced in response to a disease or its therapeutic intervention- has huge potential to assess the pharmacology and toxicology of therapeutic interventions as well. Thus over the past several years, metabolomics has risen in prominence and is gradually becoming a mutually complementary technique to genomics, transcriptomics, and proteomics for disease diagnosis and therapeutic evaluation.

Nuclear Magnetic Resonance (NMR) spectroscopy coupled with multivariate data analysis is currently the technique of choice for metabolomics studies owing to its unbiased, non-destructive nature, reproducible and minimal sample preparation requirement. The approach commonly known as NMR based metabolomics has exclusively been used in my Ph.D. research work. The generalized NMR-based metabolomics approach involves (i) metabolite identification in biofluids, (ii) spectral data processing and (iii) identification of altered metabolic profiles following multivariate and univariate statistical methods, and (iv) pathway analysis and functional interpretation.

In this thesis NMR-based metabolomics coupled with pattern recognition methods has been exclusively used to address some distinct problems of clinical and pharmaceutical relevance as enlisted below.

1) NMR-based metabolic analysis of serum for understanding the pathophysiology of Acute myocardial infarction (AMI) during initial hours of its occurrence.
2) Rat serum analysis has been performed to explore the toxicity mechanism of anti-tubercular drug pyrazinamide.
3) NMR analysis of peritoneal dialysis effluent has been performed to differentiate bacterial or fungal peritonitis.

The results of this thesis add knowledge and are expected to impact the field in the near future to provide more understanding of the complex role of small metabolites and metabolic products on biological pathways of human metabolism.

## Thesis Objectives

Metabolomics is a promising approach to reveal altered metabolism induced by a disease or its therapeutic intervention. Therefore, it is gradually becoming a mutually complementary technique to genomics, transcriptomics, and proteomics both for identifying disease-specific biomarkers and evaluating the efficacy and safety of pharmaceutical products. NMR coupled with the multivariate analysis is currently the method of choice for rapid metabolomics analysis owing to its unbiased, non-destructive nature and minimal sample preparation requirement. NMR-based metabolomics has been extensively and exclusively used in this thesis work to identify the distinctive metabolic signatures of some critical pathologies under the framework of following objectives.

1) **To explore the biochemical basis of critical human diseases (like Acute Myocardial Infarction, Prostate Cancer etc.) through detailed and comprehensive metabolic profiling of biological samples affected in such diseased states.**

2) **To identify and validate disease-specific molecular biomarkers to aid early disease detection, disease diagnosis, and treatment monitoring using *invitro* or *invivo* approaches.**

3) **To develop novel NMR based methods for rapid identification of microbial infections and their classification and finally to explore their utility in the diagnosis and treatment of infectious human diseases like Urinary tract infection, blood stream infections, respiratory tract infections, lung infections, etc.**

***"If I were doing a PhD, I'd be doing it in Metabolomics."***

***- James Watson***

## Thesis Structure

NMR spectroscopy in combination with multivariate data analysis has emerged as a powerful analytical technique allowing quantitative and qualitative analysis of metabolic profiles in a range of biological systems including blood serum/plasma, urine, etc. The same has been employed during my Ph.D. work to identify distinctive metabolic signatures of some critical human pathologies such as infectious peritonitis, acute myocardial infarction, and liver toxicity. The overall thesis work of my Ph.D. has been arranged into following seven chapters:

**Chapter 1** serves as an introductory chapter to the field of metabolomics, and its fast-growing implication in diverse areas of research.

**Chapter 2** includes the brief history and basic principles of NMR spectroscopy, along with the experiments mainly useful in NMR-based metabolomics.

**Chapter 3** describes the various aspects of the chemometric analysis of NMR data with its primary focus on data processing, multivariate data analysis, its validation, and interpretation.

**Chapter 4** describes the metabolic perturbations in initial hours of the acute myocardial infarction in serum samples compared to healthy individuals.

**Chapter 5** covers the utility of NMR-based metabolomics to evaluate the toxicity potential of antitubercular drug pyrazinamide.

**Chapter 6** explores the diagnostic capability of NMR-based metabolomics in differentiating infectious peritonitis.

**Chapter 7** contains the conclusion and future perspectives of the thesis.

# List of Publication's

## *Article's*

**10. Metabolomics approach discriminates toxicity index of Pyrazinamide and its metabolic products: Pyrazinoic acid and 5-Hydroxy Pyrazinoic acid.**
**Rawat A,** Chaturvedi S, Singh A, Guleria A, Dubey D, Keshari A, Raj V, Rai A, Prakash A, Kumar U, Kumar D, Saha S,
Human & experimental toxicology. 2017 Mar

**9. 1H NMR-based serum metabolomics reveals erythromycin-induced liver toxicity in albino Wistar rats.**
**Rawat A**, Dubey D, Guleria A, Kumar U, Keshari A, Chaturvedi S, Prakash A, Saha S, Kumar D.
J Pharm Bioallied Sci. 2016 Oct-Dec; 8(4): 327–334. doi: 10.4103/0975-7406.199339

**8. NMR based serum metabolomics reveals a distinctive signature in patients with Lupus Nephritis.**
Guleria A, Pratap A, Dubey D, **Rawat A**, Chaurasia S, Sukesh E, Phatak S, Ajmani S, Kumar U, Khetrapal CL, Bacon P, Misra R, Kumar D.
Sci Rep. 2016 Oct 14;6:35309. doi: 10.1038/srep35309.

**7. NMR-based urinary profiling of lactulose/mannitol ratio used to assess the altered intestinal permeability in acute on chronic liver failure (ACLF) patients.**
Kumar D, Pandey G, Bansal D, **Rawat A**, Kumar U, Dubey D, Guleria A, Saraswat VA.
Magn Reson Chem. 2016 Sep 14. doi: 10.1002/mrc.4525.

**6. Serum Metabolic Disturbances Hailing in Initial hours of Acute Myocardial Infarction elucidated by NMR based Metabolomics.**
**Rawat A**, Srivastava R K, Dubey D, Guleria A, Singh S, Prakash A, Modi D R, Khetrapal C. L. , Tiwari S and Dinesh Kumar.
Curr Metabolomics. 2016; 4(2). doi: 10.2174/2213235X04666160809123143.

**5. Isolated flavonoids from Ficus racemosa stem bark possess antidiabetic, hypolipidemic and protective effects in albino Wistar rats.**
Keshari AK, Kumar G, Kushwaha PS, Bhardwaj M, Kumar P, **Rawat A**, Kumar D, Prakash A, Ghosh B, Saha S.
J Ethnopharmacol. 2016 Apr 2;181:252-62. doi: 10.1016/j.jep.2016.02.004.

**4. Antiproliferative effect of isolated isoquinoline alkaloid from Mucuna pruriens seeds in hepatic carcinoma cells.**
Kumar P, **Rawat A**, Keshari AK, Singh AK, Maity S, De A, Samanta A, Saha S.
Nat Prod Res. 2016;30(4):460-3. doi: 10.1080/14786419.2015.1020489.

**3. NMR-Based Serum Metabolomics Discriminates Takayasu Arteritis from Healthy Individuals: A Proof-of-Principle Study.**
Guleria A, Misra DP, **Rawat A**, Dubey D, Khetrapal CL, Bacon P, Misra R, Kumar D.
J Proteome Res. 2015 Aug 7;14(8):3372-81. doi: 10.1021/acs.jproteome.5b00422.

**2. Metabolite characterisation in peritoneal dialysis effluent using high-resolution (1)H and (1)H-(13) C NMR spectroscopy.**
Guleria A, Bajpai NK, **Rawat A**, Khetrapal CL, Prasad N, Kumar D.
Magn Reson Chem. 2014 Sep;52(9):475-9. doi: 10.1002/mrc.4094.

**1. NMR investigations of structural and dynamics features of natively unstructured drug peptide - salmon calcitonin: implication to rational design of potent sCT analogs.**
**Rawat A**, Kumar D.
J Pept Sci. 2013 Jan;19(1):33-45. doi: 10.1002/psc.2471.

***Book Chapter's***

**1. NMR Based Metabolomics: An Emerging Tool for Therapeutic Evaluation of Traditional Herbal Medicines.**
Dinesh Kumar, **Atul Rawat**, Durgesh Dubey, Umesh Kumar, Amit K Keshari, Sudipta Saha and Anupam Guleria.
SM-eBOOK: Nuclear Magnetic Resonance Spectroscopy (2016), June, 1-18.

# Acknowledgment

First and foremost, I would like to thank all my teachers from all walks of my life, for making the individual who I am today. I would like to express my sincere gratitude to my supervisor **Dr. Anand Prakash** and co-supervisor **Dr. Dinesh Kumar** and for considering me as their Ph.D. student, and for their constant belief in me, giving the privilege and pleasure of working with the very best, they continue to be a steady source of inspiration. During this course of time, I have had the pleasure to meet numerous exceptional people and most importantly learn much about science, myself and life in science.

Coming from a biotechnology background, understanding NMR was more than a task. I will always be thankful to my supervisor Dr. Dinesh who has always been very patient in teaching me the various fundamentals of NMR repeatedly. His supervision of the NMR data acquisition and provision of scripts for data analysis were of paramount importance for the success of this work in my hand.

The work presented in this thesis was possible due to NMR facility present at the Center of Biomedical Research (CBMR) and the animal facility at the Babasaheb Bhimrao Ambedkar University (BBAU). I would like to thank **Professor C.L. Khetrapal** and **Professor Ganesh Pandey**, director CBMR for providing the excellent research facilities, and access to NMR round the clock.

Thanks' to all my co-authors and colleagues at CBMR and BBAU for creating such a brilliant environment and helping me get through the difficult times, entertainments and the caring they provided. Sincere thanks for their willingness to give feedbacks on my writing for their contributions and critical discussions, your enthusiasm, inspiration, support and valuable advice are sincerely appreciated.

I would like to thank all the patients who consented to take part in all my studies and animals sacrificed without them this work would have been just on paper without the facts, figures, and results described as such in my thesis.

I am grateful to CSIR, New Delhi for providing me the financial assistance during this study.

Sincerely,
Atul

# *Chapter 1*

## *NMR-based Metabolomics: Introduction and Review of Literature*



**"There is no adequate defense, except stupidity, against the impact of a new idea." - Percy Williams Bridgman**

The study of a biological system as a whole rather than in parts i.e. in a holistic manner is progressively being considered as an essential requisite to provide qualitative and quantitative descriptions of the emergent properties of the complete system. To address the problem of complexity of biological systems and gain insight into the interactions of the systems' constituents (systems biology) the role of metabolomics is impeccable[1]. Systems biology is a phrase that envelopes the various '-omics' technologies, of which metabolomics is just one. Studies in systems biology aim to incorporate the structures and functions of different levels of encoded information of the organism. Systems biology carries out the studies focused on the complex interactions of system components underlining the complete system as opposed to the individual constituent entities independently. Perturbations to mammalian systems (diet, disease, drugs, and environment) are multi-factorial, and the study of small parts of the system is insufficient to comprehend the complete phenotypic changes induced. Although information stored in the genes in the form of DNA sequence, it is made available only through the cellular machinery that can decipher this sequence and can translate it into structure and function. The biological systems in systems biology are studied by systematically altering them (biologically, genetically, or chemically) and then monitoring the response of gene, protein, and informational pathways, integrating this information; and finally, devising mathematical models that depict the organization of the system and its response to individual perturbations[2].

Metabolomics is one functional level instrument being employed to investigate the complex interactions of metabolites with other metabolites but also the regulative role metabolites provide through interaction with genes, transcripts, and proteins. Metabolomics is a critically important technique in a systems biology framework contemporary to other traditional "omics" techniques. Metabolites are crucial components of the biological system and highly informative about its functional state, due to their closeness to functional endpoints and the organism's phenotypes[3]. The applications of metabolomics and studying the metabolome of not only mammalian systems but also other biological systems is fast becoming the approach of choice across a broad range of sciences including systems biology, drug discovery, molecular and cell biology, along with other medical and agricultural sciences. Comparative metabolomics platforms are evolving into novel technologies for monitoring health–disease continuum, drug efficacy and toxicity and dietary effects on mammalian health. Emerging efficient multidisciplinary metabolomics techniques with biotechnology will incredibly benefit both basic and applied medical research.

Technological advancements, new analytical and bioinformatics technologies and methods are continually being created or optimized, altogether expanding the cross-disciplinary capabilities of this new biology, and are the driving force behind advancements in scientific knowledge. A varied array of platforms are available for metabolome analysis. Which includes; high-performance liquid chromatography (HPLC), gas chromatography–mass spectrometry (GC–

MS), liquid chromatography–mass spectrometry (LC–MS), capillary electrophoresis–mass spectrometry (CE–MS), Fourier transform ion cyclotron mass spectrometry (FT-ICR-MS), and nuclear magnetic resonance (NMR). The present ongoing advances in the two analytical platforms of mass spectrometry (MS) and nuclear magnetic resonance (NMR) spectroscopy have driven forward the field of metabolomics and are the tools most widely being used. There is always a payoff between different technologies regarding sensitivity, high throughput, robustness, quantitation analysis, and suitability for specific chemical classes of metabolites. Nevertheless, a cautiously chosen analytical method can be an excellent initial strategy for gaining the first impression of a metabolic profile that can ultimately be used to identify key biochemical leads to further or more focused studies[4]. Many platforms available for high-throughput screening of metabolites, NMR spectroscopy has emerged as a fundamental tool for understanding metabolic processes in biological systems. Despite the presence of a variety of approaches to study the metabolism in a biological system, a new and fast growing approach Metabolomics has preceded over all the existing techniques.

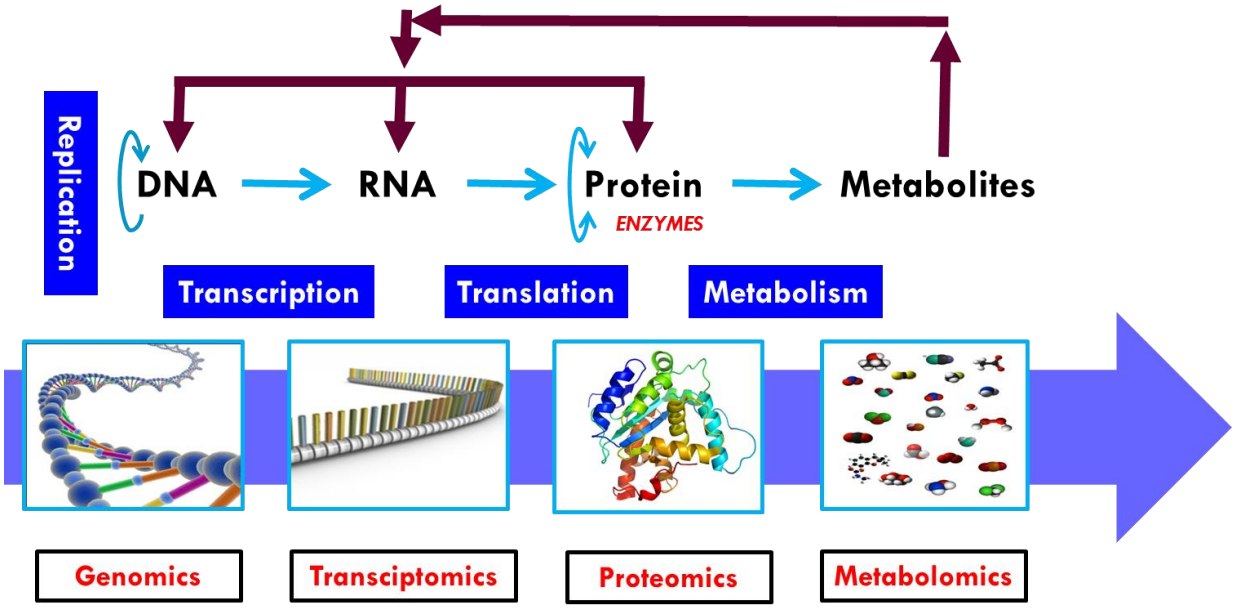



**Figure 1.1:** The "omics" cascade.

## 1.1 Metabolomics, Metabolome, and Metabolites:

Metabolomics can be defined as "the quantitative measurement of the multiparametric metabolic response of living systems to pathophysiological stimuli or genetic modification"[5]- Jeremy K. Nicholson.

Different jargon for the definition of metabolic approaches has been used by various metabolomics research areas some of which can be found in **Table 1.1**.

| | |
|---|---|
| **Metabolomics** | Unbiased identification and quantification of all the metabolites present in a biological system. |
| **Metabolome** | Complete set of low-molecular-weight metabolites present in a biological sample (i.e., biofluid, organism, bacterial community). |
| **Metabolite target analysis** | Qualitative and quantitative analysis of one or a few metabolites related to a specific metabolic reaction. |
| **Metabolite profiling** | Metabolic profiling Identification and quantification of a selected number of pre-defined metabolites, generally related to a specific metabolic pathway(s). |
| **Metabolic footprint** | Analysis of the metabolites secreted/excreted by an organism; it may include environmental and growth substances. Does not rely on the measurement of intracellular metabolites but rather, on monitoring those that are secreted or fail to be taken up by a cell or tissue. |
| **Metabolic fingerprinting** | Unbiased, high-throughput, rapid, global analysis of samples to provide sample classification. The analysis is oriented towards defining clinically relevant differences rather than identifying all the molecules present in a sample. |
| **Metabonomics** | Evaluation of tissues and biological fluids for changes in endogenous metabolite levels that result from disease or therapeutic treatments. |
| **Metabolite flux analysis** | Also known as fluxomics. Labeled metabolites are fed into a biosystem, and the destination of the label is assessed, usually in a time-dependent manner. |

**Table 1.1**: Definitions and terms used in metabolomics. Adapted from[6-9].

Metabolomics is the comprehensive, qualitative, and quantitative study of all the small molecules in an organism[10]. Biological systems comprise of organelles, cells, tissues, organs, and organisms; that can be divided into four basic biochemical components – genes, transcripts, proteins, and metabolites, which serve as building blocks and information repositories for the biological systems[11]. The biological system functions through complex interactions of these constituents. The low molecular weight molecules (metabolites), which make up the 'metabolome,'

are deduced from the interaction of the genome with its environment and are not simply the end products of gene expression, but also form part of the metabolic status of a biological system.

***Metabolomics*** **-** Is the comprehensive study of naturally occurring low molecular weight metabolites, in biological samples such as tissues, cells, biofluids (e.g. urine, serum, cerebrospinal fluid (CSF), saliva, bile, and synovial fluid, etc.)[12]. These low molecular weight metabolites/small molecules are <1500 Da (molecular weight 50 – 1500 Daltons) and are the constituents of biofluids[12,13], tissues, and cells which provide insights of the perturbation to an organism caused by diseases, aging or external stimuli (e.g. drug treatment)[14]. It is most relevant to phenotypes as compared with other "Omics" studies. The study of metabolomics excludes polymers of amino acids and sugars. Metabolomics instead focuses on intermediate metabolites which are important constituents of macromolecular structures and serve as regulatory molecules of metabolism as signaling molecules or secondary metabolites[15]. Metabolomics used in combination with physiological, biochemical assays and genomics, transcriptomics or proteomics, could lead to a profound understanding of an organism's organization and function. Research fields of metabolomics include sample preparation and extraction for: metabolic profiling, disease biomarker discovery, drug toxicity study, bioinformatics study (e.g. application of classification and prediction models) and data processing approaches related to specific analytical techniques such as NMR.

***Metabolome*** **-** The quantitative complement of metabolites in a biological system is defined as the metabolome[10,16]. The complexity and size of the metabolome are dependent on the organism and sample type (blood, urine, CSF or tissue for example). Metabolomes can be classified according to their origin. Endometabolomes are related to intracellular metabolism, exometabolomes (alternatively referred to as the metabolic footprint or secretome) refer to extracellular metabolomes[17]. In mammalian system, the metabolome can be described by the sample type and include serum or plasma, urine, CSF, breath, tears, saliva, fecal and a variety of tissues. One metabolome can be interrelated with another metabolome. For example, serum and urine are biofluids incorporating the metabolic composition of several tissues and organs which are related to multiple biological and physiological processes. This is beneficial when investigating these biofluids as they are comparatively easy to procure and provide a metabolic snapshot of the mammalian system as a whole. The study of the metabolome has numerous advantages, whether applied individually or in combination with other biochemical analysis[18]. The metabolome is downstream of other biochemical species with biochemical information conventionally viewed as flowing from genome to transcriptome to proteome to metabolome.The metabolome is an indirect measure of the biological phenotype, an indicator of both genetic and environmental (diet, drug, lifestyle) perturbations. The metabolome is extremely dynamic in nature, the rate of synthesis or consumption of metabolites is measured in seconds compared to

turnover in the proteome and transcriptome which are typically measured in minutes to hours. Also, the interaction of human and gut microflora, metabolome plays a significant role in the health–disease status, including the cross-talk between these separate metabolomes.

***Metabolites*** - are in a unique position, as they are the building blocks for all other biochemical species and structures including proteins (amino acids), genes and transcripts (nucleotides), and cell walls[19]. Metabolites are low molecular weight organic and inorganic chemicals which are the reactants, intermediates or products of enzyme-mediated biochemical reactions. They are functionally different to peptides, proteins, transcripts and genes though the exact division is often blurred. Macromolecules like DNA and RNA are synthesized from nucleotides, regulated by proteins; all of which are made up off metabolites, some of which also have significant roles in cellular energy processes. The human body contains approximately 3000-5000 detectable metabolites, a sizable fraction of which have been identified[20] **Figure 1.2**. It is expected the number of metabolites is lower than the number of all genes and protein in a cell. At present, it is challenging to determine the exact number of metabolites and also other cell products such as transcripts and proteins at a given time in a specific cell because of the limitations of analytical techniques. Metabolites and their relationship with other metabolites and biochemical species are currently the major focus of metabolomic investigations to understand biological function/phenotype. Metabolites are involved in many other biochemical processes directly or indirectly related to their synthesis or consumption and also in the regulation of metabolism[21,22]. Metabolism is regulated to ensure adequate biomass and energy production along with other requirements for growth and life. Metabolic homeostasis provides a constant chemical environment within a biological system maintained by regulation of metabolism and other processes. Upregulation or downregulation of the metabolite in the system can be self-regulated by increasing or decreasing the activity of enzymes responsible for the various kinds of reactions through different modification. Dysregulation of these regulatory processes can result in disease onset or progression[23]. The integrative diversity of metabolites provides wide ranges of physicochemical properties including molecular weight, hydrophobicity/hydrophilicity, acidity/basicity and boiling point. Covalent modification of proteins such as phosphorylation, acetylation, ubiquitination, sumoylation and through transcription factors offer regulation and control over metabolism across multiple organs, and recently, riboswitches (the interaction of RNA with metabolites) have been shown to modulate gene expression[24]. However, vital metabolic reactions and pathways required for energy, growth, and nutrient supply and are conserved across numerous organisms/life forms (for instance, the pathways of glycolysis and the citric acid cycle). This is a measure of the robustness of metabolic networks, often discussed in the evolution of metabolic networks[25].

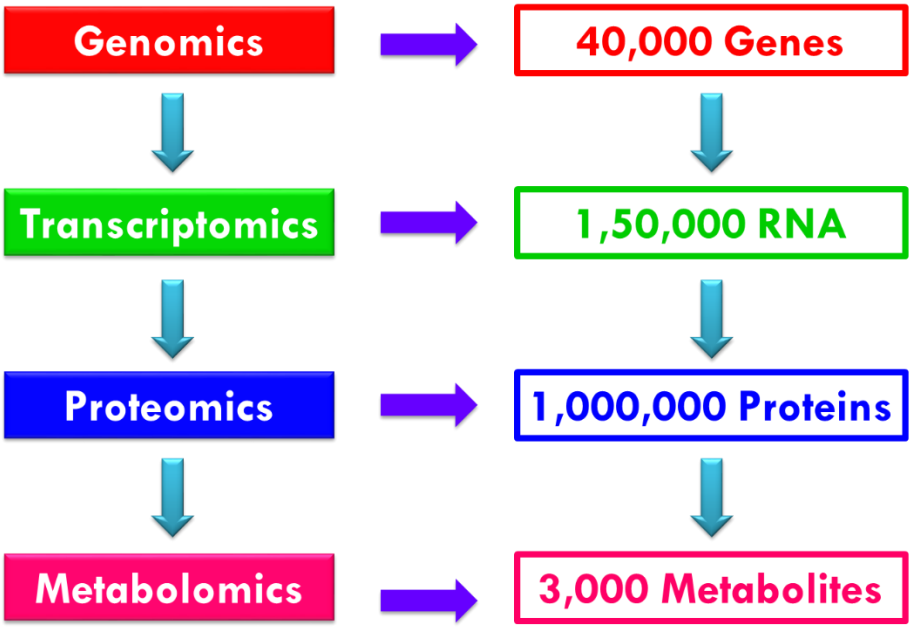



**Figure 1.2:** The relationship between diversity and complexity between the "omics" technologies. DNA, RNA, Protein and Metabolites.

## 1.2 Approaches for Metabolomics

The metabolomics experiment provides unique challenges to fulfill the goal of improving the current status of biological information related to the metabolome and more generally functional genomics[6,26]. A typical metabolomics research flow starts with the design of an experiment plan then proceeds to sample collection and sample preparation followed by data acquisition using analytical instrumentation, and finally data processing and data interpretation using a variety of statistical techniques[27,28]. For a metabolomics study, experimental design up to the biostatistics should be optimized and appropriate for their intended purpose. The ultimate goal of a metabolomics experiment is to identify and quantify all of the metabolites in a cell or tissue in a given state at a given point in time[7]. However, it is currently impossible to do so simultaneously in a single, high-throughput platform because no single extraction technique or analytical instrument can isolate and detect every metabolite within a biological sample[29,30]. These difficulties are further amplified by issues such as a human error in sample preparation and extraction, sample storage, and instrument reproducibility[31]. Additionally, metabolomics is challenging due to the structural diversity in chemical structure, size, abundance and reactivity of the collection of metabolites in any biological samples.

Analytical approaches for metabolomics can be categorized broadly into two distinct groups: targeted or untargeted[32]. This is dependent on whether the methodology implemented designed to quantify some specific metabolites (targeted) or to measure a larger set of

metabolites restricted only by the sensitivity and applicability of the analytical platform(s) and data processing employed (untargeted). These approaches can further be subdivided as metabolic profiling, using an untargeted approach or metabolite identification and quantitation using a targeted approach[30] as seen in **Figure 1.3**.

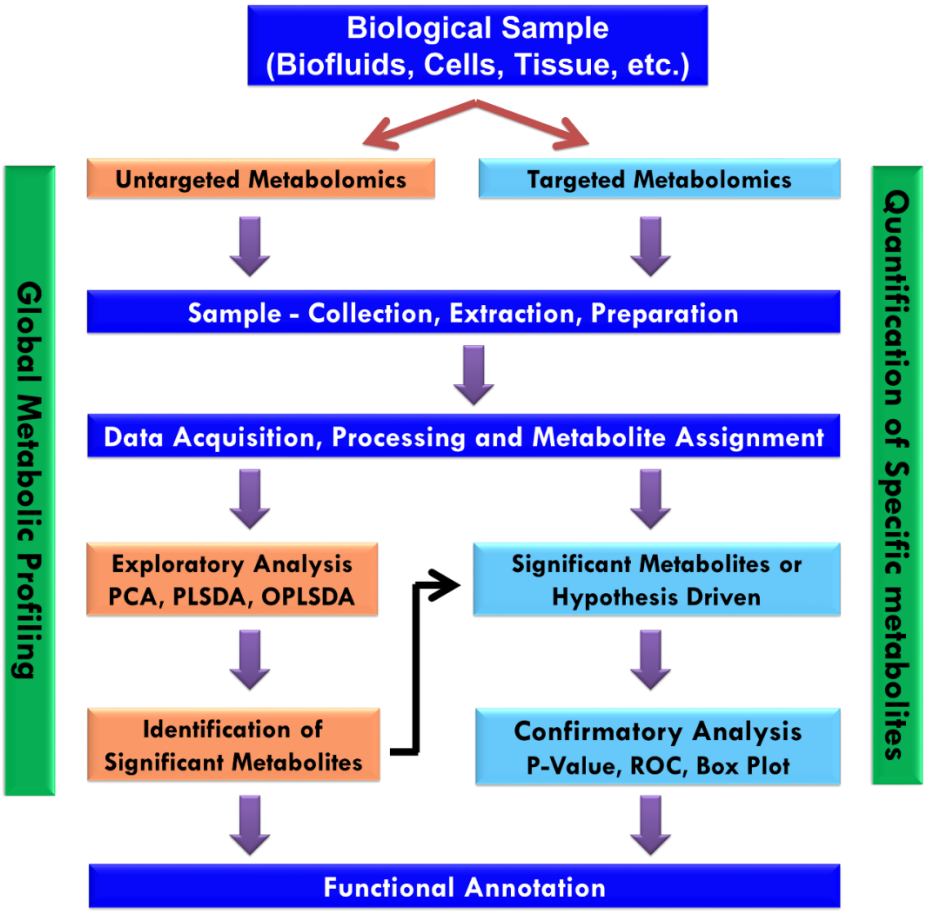



**Figure 1.3:** Workflow is illustrating both untargeted and targeted metabolomics approaches.

**Targeted metabolomics** is commonly driven by a specific biochemical question or hypothesis that motivates the investigation of one or more related pathways of interest[33]. These studies involve the use of biochemical and analytical tools for the quantification of known metabolites of biological interest. Targeted metabolomics studies can be effective for pharmacokinetic studies of drug metabolism as well as for measuring the influence of therapeutics or genetic modifications on a specific enzyme[34].

**Untargeted metabolomics** is global in its scope, usually hypothesis-free, and has the aim of simultaneously measuring as many metabolites as possible from biological samples[35]. It is used for the identification of metabolic pathways that are altered following distresses of biological systems[36]. Since untargeted metabolomics does not require prior knowledge, it can be utilized to identify novel metabolic biomarkers of disease and drug efficacy as well analyzing the global metabolic profile of the whole system. Both targeted and untargeted metabolomics reveals the

expected behavior of known metabolites, but only untargeted metabolomics allows the detection of concurrent effects between variables which cannot be observed at an individual level[11].

## 1.3 NMR-based Metabolomics

Metabolomics is the discipline where a change in the metabolites level (endogenous and exogenous) are assessed, identified and quantified in different biological samples[37]. Metabolomics is the terminal view of the biological system which provides the additional information about understanding the metabolic pathways and downstream processes. It is a well-known fact that in a biological system, the induced pathophysiological perturbations result in disturbances in the ratios and concentrations, binding, or fluxes of various endogenous biochemicals[38]. In many cases, drugs exert their toxic effects by interacting directly with genetic material or by inducing the synthesis of drug metabolizing enzymes, which generate toxic products. These can be due to direct chemical reaction or by binding of key enzymes or nucleic acids that control metabolism. The biofluid composition/profile is modulated by the function of the cells responsible for its manufacture and secretion. This biochemical profile may be altered when organ damage occurs due to toxicity, disease or any other cellular insult. Metabolites are in dynamic equilibrium with those inside cells and tissues in body fluids, so the abnormality in any cellular process in tissues of the whole organism will be revealed in altered biofluid compositions[39,40]. Consequently, these toxic effects will influence the functioning of the entire organism. These alterations can be very minute owing to the body's homeostasis mechanism. Hence we need an excellent analytical technique to detect all these alterations. Metabolomics utilizes classical analytical techniques to probe chemical fingerprints and concentration levels of metabolites in the investigated samples. NMR spectroscopy is the most common analytical techniques in metabolomics, employed to investigate disease processes and biological mechanisms through metabolic profiling. NMR spectroscopy appears to be the system of choice for investigating the abnormal body fluid compositions, as a broad range of metabolites can be detected simultaneously with minimal sample preparation and "without prejudice," and hence many cellular biochemical processes can be probed[31,41-43]. Along with biofluids, the abnormal metabolite profiles in intact tissue, tissue extracts or cell suspensions and extracts can also be screened efficiently. This makes it possible to apply NMR-based Metabolomics in many streams such as disease diagnosis, treatment monitoring, drug toxicity, xenobiotic studies, environmental sciences, plant metabolism, plant pathologies, and so on[44].

Henceforth, in recent times NMR has become an indispensable tool for metabolomics and system biology studies. The use of NMR spectroscopy for the analyses of biofluids can be reviewed by a significant increase in the number of NMR-based metabolomics studies. NMR spectroscopy has the advantages of being relatively robust across many samples (no part of the

sample becomes contaminated during the process), requires minimal or no sample preparation, and it is non-destructive and can usually be implemented in a non-invasive manner. Also relatively fast where many hundreds of samples can be acquired every week depending on the choice of experiment and time (with spectra acquired within a few minutes). This reduces the financial costs per sample to acceptable levels and significantly lower than for proteome and transcriptome. The high reproducibility of NMR-based techniques gives the method numerous benefits over other analytical techniques in large-scale and long-term metabolomic studies, such as epidemiological studies. Henceforth, it is being used to study a broad range of diseases, through the examination of biofluids, including blood plasma/serum, urine, saliva, CSF, bile, and semen, as well as cell and tissue extracts and intact tissue biopsies[45] representative spectra are shown in **Figure 1.4**. NMR analysis of biofluids gives a snapshot of the biochemical status of a living organism. Such biochemical information can reflect the modes and severity of disease, organ dysfunction. The multinuclear capabilities of NMR provide various means to observe different chemicals. The metabolomics work has exclusively used $^{1}H$ NMR for analyses of biofluids, but other (isotopically labeled) nuclei like $^{13}C$, $^{31}P$, $^{15}N$, $^{19}F$, and $^{2}H$ may provide additional information about various metabolite pools and to gain useful information about the flow of metabolites through metabolic pathways[46,47]. This leads to the employment of different NMR techniques in a wide range of metabolomics analyses.

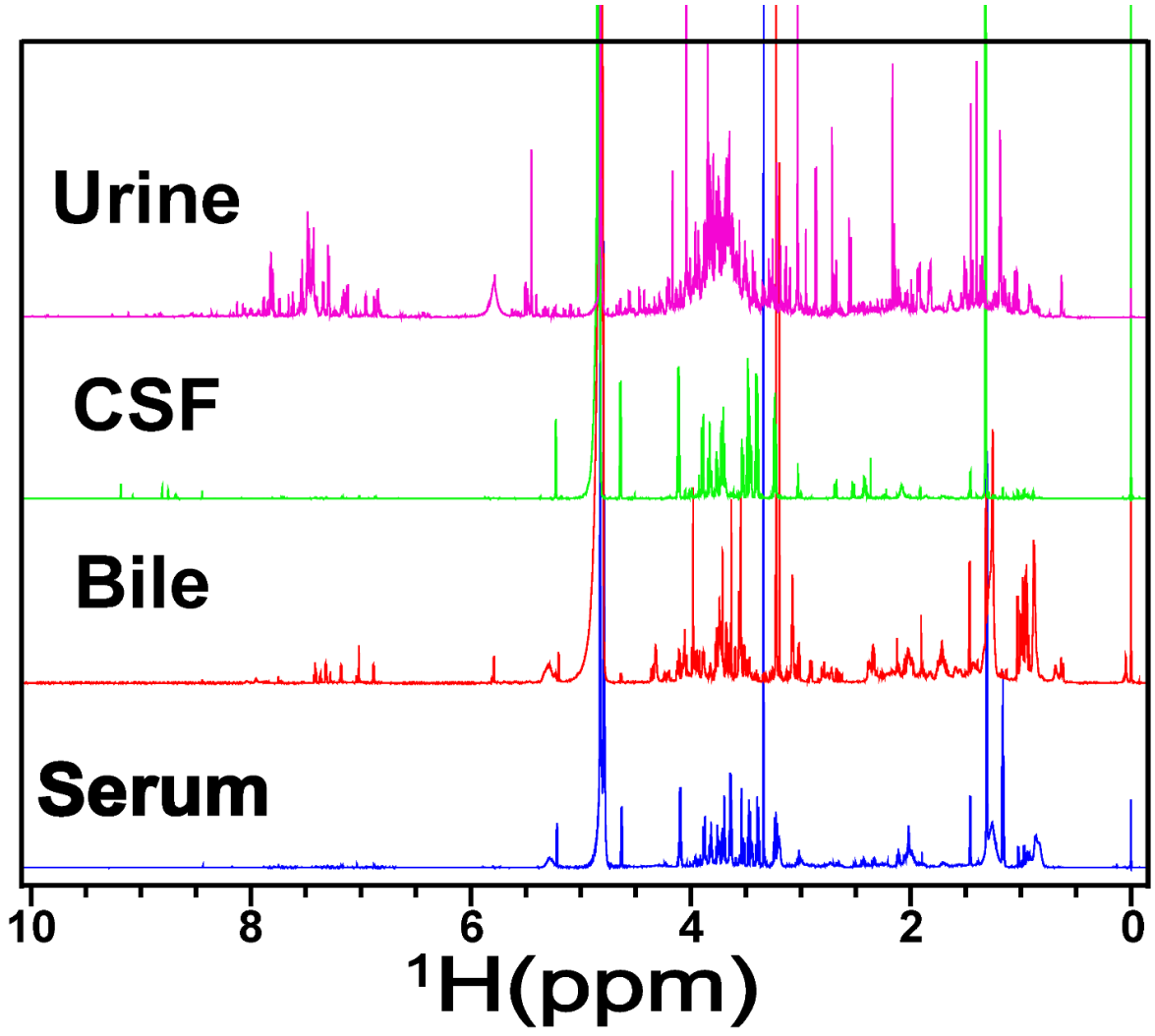


**Figure 1.4:** Example 1D $^{1}H$ spectra at 800 MHz are displayed for human samples of serum, bile, CSF, and urine.

NMR-based metabolomics studies have following steps involved in general as shown in **Figure 1.5**. The metabolomic analysis starts with the acquisition of samples. Samples may be derived from any biofluid, tissue/cell lysate suitably processed to be cell-free, including urines, plasma/serum, cerebrospinal fluid, fecal extracts and synovial fluid[48]; and pH should be made constant throughout the samples, via the addition of buffers[49]. Alternatively, cell extraction must be performed in a suitable solvent, such as methanol with chloroform extraction to remove protein components. Consistent processing and buffering are crucial for the success of metabolomic analysis due to the potential confounding factors that may result from variations in pH during acquisition – including shifted peaks and chemical reaction resulting in loss or gain of metabolites. Acquisition of samples results in spectra containing numerous peaks, representing proton resonance in NMR. However, the NMR data in itself is incomplete to provide any information, unless coupled with Univariate and Multivariate Statistical Analysis/Patter Recognition Analysis. The primary goal of the analysis is the same – using changes in peaks, and therefore metabolites, to describe the variations between sample groups. The large size of the produced datasets, and often numerous samples, requires the use of statistical analyses to simplify and distil the meaning/information from the resulting data. The statistical analysis of metabolic NMR data involves both unsupervised (exploratory) and supervised (e.g. classification/regression) analysis, with the aim of identifying and characterizing significant differences between classes resulting from the presence of external or internal stimuli to the organism[50]. Classical methods (employed in metabolomic analysis across technologies) include Principle Components Analysis (PCA) and Partial Least Squares (PLS) regression, which aid interpretation of the data by projecting onto lower dimension spaces corresponding to either the highest variance or highest covariance between the scores of the data and the response variable. Typically the analysis workflow will include firstly unsupervised clustering such as PCA and then be followed by a supervised method (PLS-DA, OPLS-DA) to enhance this separation and understand which variables are responsible for the differences between the datasets.

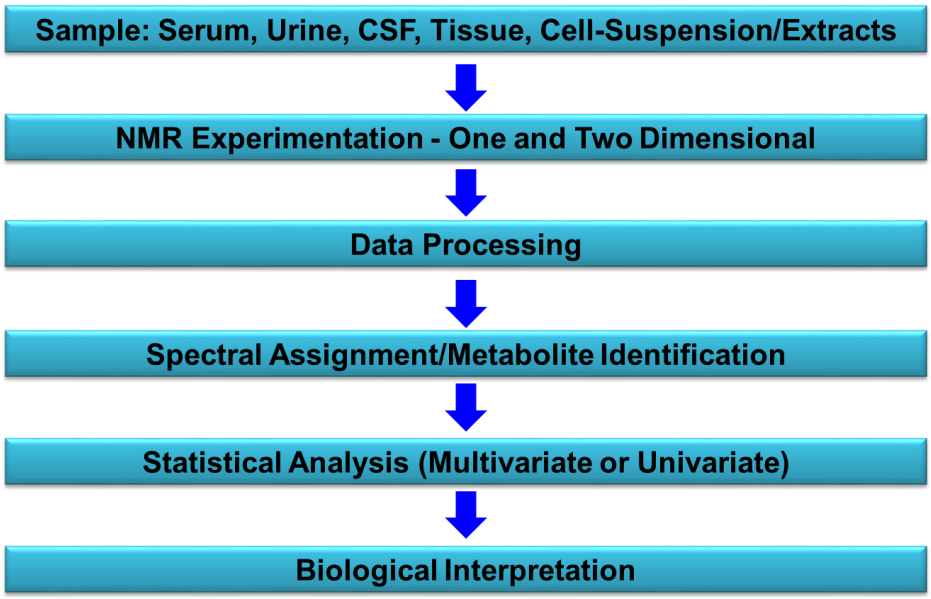



**Figure 1.5:** General steps involved in routine NMR-based Metabolomics approach.

## 1.4 Applications of NMR-based Metabolomics

*"Metabolomics can be applied to advance research across nearly every research area."*

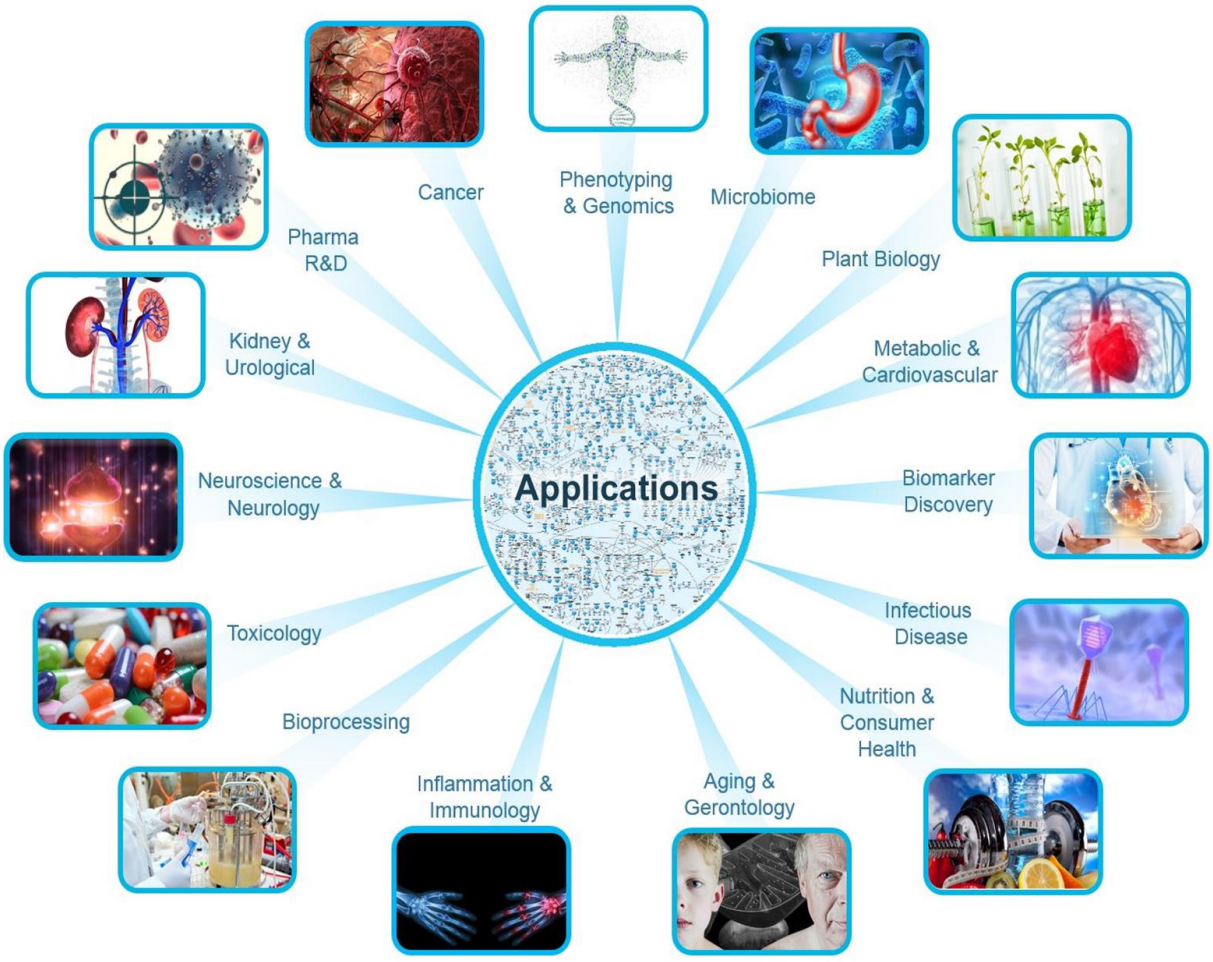


**Figure 1.6:** Depicting application of metabolomics in various field of biological sciences, adapted from slideshare.net/MargaretMegEasonMAJD/metabolomics-the-next-generation-of-biochemistry.

Metabolomics is a dynamic and emerging area of research, similar to proteomics, transcriptomics, and genomics in the comprehensive understanding of biological systems. It is specifically valuable for functional genomic studies in which metabolism is believed to be deranged. Metabolomics gives a snapshot of the metabolic dynamics that reflect the response of living systems to both pathophysiological stimuli and genetic modification. As this approach makes possible the examination of interactions between an organism and its diet or environment, it is particularly useful for identifying biomarkers of disease processes that involve the environment. Applications of metabolomics include disease diagnosis, monitoring the effects of medical interventions including drugs, detection of adulteration of food, and analysis of biochemical pathways and their perturbations resulting from mutations, aging, diet, exercise, or lifestyle **Figure 1.6**.

**Disease diagnosis approach:**

In past few years, NMR-based metabolomics is being extensively used in the human disease diagnosis[51]. There are various reports on metabolomics for disease diagnosis using biofluids such as inborn metabolism errors, drug overdoses, and toxicity inducing factors and so on. The data also highlight pathways and biological compounds that are disrupted at early stages of the diseases, to help elucidate target compounds and the pathophysiology of the considered diseases for early prognosis and diagnosis using noninvasive samples. Recently serum analysis has been carried out in various kinds of cancers like breast, hepatic cells[52], lung, ovarian cancers[53], colorectal, kidney/renal[54], pancreatic, oral, etc.. The CSF sample investigation using NMR-based metabolomics has been carried out and proved to be useful in differentiating the control subjects from those with meningitis and various kinds of bacterial, viral and fungal infections. Along with biofluids, intact tissue samples can also be studied using High-Resolution Magic Angle Spinning (HRMAS) NMR techniques. The clinical assays focused on non-targeted or targeted metabolomics and metabolite profiling as well as predictive modeling based on clinical trials[55].

**Pharmacometabolomics:**

Pharmacometabolomics is a fast-emerging discipline that involves the direct measurement of metabolites in an individual's body fluids to understand mechanisms of drug or xenobiotic action (for example, efficacy or toxicity) and to categorize novel biomarkers of the cellular response to drug intervention[56]. This approach further allows the identification of the metabolic pathways implicated in individual variations in response to drug treatment[57]. Further studies will address the current state of pharmacometabolomics, its applications, and its future potential in drug

development. Pharmacometabolomics is defined as a prognostic or predictive methodology, in contrast to metabolomics which is a diagnostic method.

**Pre-clinical drug trials:**

The selection of robust candidate drugs for final development based upon minimization of the occurrence of adverse drug effects is one of the most important aims of pharmaceutical R&D. In preclinical and clinical drug development, the applications of metabolomics have a substantial scope in the pharmaceutical industry, practically at each step from drug discovery to clinical development. These include determination of drug target, potential safety and efficacy biomarkers and mechanisms of drug action, the validation of experimental preclinical models against human disease profiles, and the discovery of clinical safety and effectiveness of biomarkers[58].

**Plant metabolomics:**

Along with human/animal metabolomics, NMR-based metabolomics also plays an imperative role in plant system[59]. Despite a lot number of metabolites remain to be identified, metabolomics has contributed significantly to the understanding of plant physiology and biology from the view of small chemical molecules that reflect the end point of biological activities. In past decades NMR methods have also been used for studying the effect of stress conditions, exposures to metals and herbicides, growth conditions and so on[60]. NMR-based plant metabolomics is a crucial tool for analyzing metabolic pathways in plant systems by using stable isotopes (such as $^{13}C$, $^{15}N$) for kinetic studies of various intermediates formed during growth. Based on the current knowledge on the genetic and biochemical mechanisms underlying plant growth, development, and stress responses, focusing further on the contributions of metabolomics to practical applications in crop quality improvement and food safety assessment, as well as plant metabolic engineering, leading to better crop yield and quality.

## 1.5 Metabolomics as a Tool for Discovery of Biomarkers

"a characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or biological response to a therapeutic intervention"[61].

"any substance, structure, or process that can be measured in the body or its products and influence or predict the incidence of outcome or disease."

– BIOMARKER Definition.

The use of biomarkers in basic and clinical research as well as in clinical practice has become so commonplace that their presence as primary endpoints in clinical trials is now accepted almost

without question. In the case of specific biomarkers that have been well categorized and repeatedly shown to correctly predict relevant clinical outcomes across a variety of treatments and populations, this use is entirely justified and appropriate. In many cases, however, the "validity" of biomarkers is assumed where, in fact, it should continue to be evaluated and reevaluated. Biomarkers play a critical role in improving the drug development process as well as in the larger biomedical research enterprise. Understanding the association between measurable biological processes and clinical outcomes is vital to expanding the array of managements for all diseases, and for deepening our understanding of normal, healthy physiology. A major practical application of comprehensive metabolomics profiling has been in the discovery of biomarkers which will provide early detection, monitoring the progression of diseases and new clinically relevant diagnostic targets or to find novel metabolite pathways to give pieces of information relating to the general homeostasis a living being. Any metabolite which is considered to be a biomarker should be qualified by a rigorous, multiphase process which has unambiguously demonstrated its diagnostic and prognostic value. The first phase of biomarker qualification, the discovery stage, is primarily concerned with identifying compounds which are either up or down-regulated or simply absent between a standard/control sample group and a treated/distressed sample group. It is this stage where comprehensive metabolomics has its greatest utility as a discovery tool. Compounds which appear to be significantly different between these groups are commonly referred to as a differentiating marker. At this stage, the full chemical identity of the marker may or may not be known, but it needs to be complemented by a unique set of measurement features and to be verified on multiple metabolic platforms. The next phase of discovery involves the full elucidation of the chemical identity of the differentiating marker followed by the analytical validation of the marker by an acceptable analytical method (targeted metabolomics). At this stage, a differentiating marker becomes a candidate biomarker[62]. The last phase of biomarker qualification involves the study of how the biomarker relates to biological mechanisms impacted by the event in question and the study of how clearly the candidate biomarker is indicative of a specific pathophysiology. In this last phase, it is necessary to assess the potential for false positive and false negative results[63]; this stage also encompasses pathway discovery[64]. Again, despite being in its infancy, comprehensive metabolomics approaches have been successfully used to discover novel biomarkers and pathways. Some of these biomarkers are validated and are available for use in the clinic.

**Reference's**


1. Ideker, T.; Galitski, T.; Hood, L. A new approach to decoding life: systems biology. *Annual review of genomics and human genetics* **2001**, *2* (1), 343-372.

2. Smolinska, A.; Blanchet, L.; Buydens, L. M.; Wijmenga, S. S. NMR and pattern recognition methods in metabolomics: from data acquisition to biomarker discovery: a review. *Analytica chimica acta* **2012**, *750*, 82-97.

3. Hirai, M. Y.; Klein, M.; Fujikawa, Y.; Yano, M.; Goodenowe, D. B.; Yamazaki, Y.; Kanaya, S.; Nakamura, Y.; Kitayama, M.; Suzuki, H. Elucidation of gene-to-gene and metabolite-to-gene networks in Arabidopsis by integration of metabolomics and transcriptomics. *Journal of Biological Chemistry* **2005**, *280* (27), 25590-25595.

4. Yoon, D.; Lee, M.; Kim, S.; Kim, S. Applications of NMR spectroscopy based metabolomics: a review. *Journal of the Korean Magnetic Resonance Society* **2013**, *17* (1), 1-10.

5. Nicholson, J. K.; Lindon, J. C.; Holmes, E. 'Metabonomics': understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data. *Xenobiotica* **1999**, *29* (11), 1181-1189.

6. Vaidyanathan, S.; Dunn, W. B.; Harrigan, G. G.; Kell, D. B. G. Metabolomics by numbers: acquiring and understanding global metabolite data. **2004**.

7. Goodacre, R.; Vaidyanathan, S.; Dunn, W. B.; Harrigan, G. G.; Kell, D. B. Metabolomics by numbers: acquiring and understanding global metabolite data. *Trends in biotechnology* **2004**, *22* (5), 245-252.

8. Barderas, M. G.; Laborde, C. M.; Posada, M.; de la Cuesta, F.; Zubiri, I.; Vivanco, F.; varez-Llamas, G. Metabolomic profiling for identification of novel potential biomarkers in cardiovascular diseases. *BioMed Research International* **2011**, *2011*.

9. Barba, I.; Garcia-Dorado, D. *Metabolomics in Cardiovascular Disease: Towards Clinical Application;* INTECH Open Access Publisher: 2012.

10. Oliver, S. G.; Winson, M. K.; Kell, D. B.; Baganz, F. Systematic functional analysis of the yeast genome. *Trends in biotechnology* **1998**, *16* (9), 373-378.

11. Dunn, W. B.; Broadhurst, D. I.; Atherton, H. J.; Goodacre, R.; Griffin, J. L. Systems level studies of mammalian metabolomes: the roles of mass spectrometry and nuclear magnetic resonance spectroscopy. *Chemical Society Reviews* **2011**, *40* (1), 387-426.

12. Bino, R. J.; Hall, R. D.; Fiehn, O.; Kopka, J.; Saito, K.; Draper, J.; Nikolau, B. J.; Mendes, P.; Roessner-Tunali, U.; Beale, M. H. Potential of metabolomics as a functional genomics tool. *Trends in plant science* **2004**, *9* (9), 418-425.

13. Koal, T.; Deigner, H. P. Challenges in mass spectrometry based targeted metabolomics. *Current molecular medicine* **2010**, *10* (2), 216-226.

14. Feng, Q.; Liu, Z.; Zhong, S.; Li, R.; Xia, H.; Jie, Z.; Wen, B.; Chen, X.; Yan, W.; Fan, Y. Integrated metabolomics and metagenomics analysis of plasma and urine identified microbial metabolites associated with coronary heart disease. *Scientific reports* **2016**, *6*.

15. Barchet, G. A brief overview of metabolomics: What it means, how it is measured, and its utilization. *The Science Creative Quarterly* **2013**, *8*.

16. Tweeddale, H.; Notley-McRobb, L.; Ferenci, T. Effect of slow growth on metabolism of Escherichia coli, as revealed by global metabolite pool (GÇ£metabolomeGÇ¥) analysis. *Journal of bacteriology* **1998**, *180* (19), 5109-5116.

17. Winder, C. L.; Dunn, W. B. Fit-for-purpose quenching and extraction protocols for metabolic profiling of yeast using chromatography-mass spectrometry platforms. *Yeast Systems Biology: Methods and Protocols* **2011**, 225-238.

18. Pearson, H. Meet the human metabolome. *Nature* **2007**, *446* (7131), 8.

19. Vander Heiden, M. G.; Cantley, L. C.; Thompson, C. B. Understanding the Warburg effect: the metabolic requirements of cell proliferation. *science* **2009**, *324* (5930), 1029-1033.

20. Wishart, D. S.; Tzur, D.; Knox, C. HMDB: the Human Metabolome Database. Nucleic Acids Res. *Database* **2007**, (D521-6).

21. Ishii, N.; Nakahigashi, K.; Baba, T.; Robert, M.; Soga, T.; Kanai, A.; Hirasawa, T.; Naba, M.; Hirai, K.; Hoque, A. Multiple high-throughput analyses monitor the response of E. coli to perturbations. *science* **2007**, *316* (5824), 593-597.

22. Dunn, W. B.; Broadhurst, D. I.; Atherton, H. J.; Goodacre, R.; Griffin, J. L. Systems level studies of mammalian metabolomes: the roles of mass spectrometry and nuclear magnetic resonance spectroscopy. *Chemical Society Reviews* **2011**, *40* (1), 387-426.

23. Muoio, D. M.; Newgard, C. B. Molecular and metabolic mechanisms of insulin resistance and +¦-cell failure in type 2 diabetes. *Nature reviews Molecular cell biology* **2008**, *9* (3), 193-205.

24. Henkin, T. M. Riboswitch RNAs: using RNA to sense cellular metabolism. *Genes & development* **2008**, *22* (24), 3383-3390.

25. Handorf, T.; Ebenh+¦h, O.; Heinrich, R. Expanding metabolic networks: scopes of compounds, robustness, and evolution. *Journal of molecular evolution* **2005**, *61* (4), 498-512.

26. Fiehn, O. Metabolomics-the link between genotypes and phenotypes. *Plant molecular biology* **2002**, *48* (1-2), 155-171.

27. Ogura, T.; Sakamoto, Y. *Application of Metabolomics Techniques using LC/MS and GC/MS Profiling Analysis of Green Tea Leaves*; Tech. Rep. 10, Lifescience, Tokyo: 07.

28. Leon Z, G.-C. J. D. M. L. A. Mammalian cell metabolomics: experimental design and sample preparation. *Electrophoresis* **2013**, *34* (19), 2762-2775.

29. Teusink, B.; van Enckevort, F. H.; Francke, C.; Wiersma, A.; Wegkamp, A.; Smid, E. J.; Siezen, R. J. In silico reconstruction of the metabolic pathways of Lactobacillus plantarum: comparing predictions of nutrient requirements with those from growth experiments. *Applied and environmental microbiology* **2005**, *71* (11), 7253-7262.

30. Bais, P. Bioinformatics methods for metabolomics based biomarker detection in functional genomics studies. **2011**.

31. Dunn, W. B.; Ellis, D. I. Metabolomics: current analytical platforms and methodologies. *TrAC Trends in Analytical Chemistry* **2005**, *24* (4), 285-294.

32. Boccard, J.; Veuthey, J.; Rudaz, S. Knowledge discovery in metabolomics: an overview of MS data handling. *Journal of separation science* **2010**, *33* (3), 290-304.

33. Nicholson, J. K.; Connelly, J.; Lindon, J. C.; Holmes, E. Metabonomics: a platform for studying drug toxicity and gene function. *Nature reviews Drug discovery* **2002**, *1* (2), 153-161.

34. Commisso, M.; Strazzer, P.; Toffali, K.; Stocchero, M.; Guzzo, F. Untargeted metabolomics: an emerging approach to determine the composition of herbal products. *Computational and structural biotechnology journal* **2013**, *4* (5), 1-7.

35. Hamdalla, M. Computational Methods in Metabolomics. **2014**.

36. Patti, G. J.; Yanes, O.; Siuzdak, G. Innovation: Metabolomics: the apogee of the omics trilogy. *Nature reviews Molecular cell biology* **2012**, *13* (4), 263-269.

37. Lindon, J. C.; Holmes, E.; Nicholson, J. K. Metabonomics techniques and applications to pharmaceutical research & development. *Pharmaceutical research* **2006**, *23* (6), 1075-1088.

38. Nicholson, J. K.; Lindon, J. C. Systems biology: metabonomics. *Nature* **2008**, *455* (7216), 1054-1056.

39. Reo, N. V. NMR-based metabolomics. *Drug and chemical toxicology* **2002**, *25* (4), 375-382.

40. Nicholson, J. K.; Connelly, J.; Lindon, J. C.; Holmes, E. Metabonomics: a platform for studying drug toxicity and gene function. *Nature reviews Drug discovery* **2002**, *1* (2), 153-161.

41. Gowda, G. N.; Zhang, S.; Gu, H.; Asiago, V.; Shanaiah, N.; Raftery, D. Metabolomics-based methods for early disease diagnostics. *Expert review of molecular diagnostics* **2008**, *8* (5), 617-633.

42. Hegeman, A. D. Plant metabolomicsGÇömeeting the analytical challenges of comprehensive metabolite analysis. *Briefings in functional genomics* **2010**, *9* (2), 139-148.

43. Griffin, J. L.; Bollard, M. E. Metabonomics: its potential as a tool in toxicology for safety assessment and data integration. *Current drug metabolism* **2004**, *5* (5), 389-398.

44. Beckonert, O.; Keun, H. C.; Ebbels, T. M.; Bundy, J.; Holmes, E.; Lindon, J. C.; Nicholson, J. K. Metabolic profiling, metabolomic and metabonomic procedures for NMR spectroscopy of urine, plasma, serum and tissue extracts. *Nature protocols* **2007**, *2* (11), 2692-2703.

45. Emwas, A. H.; Salek, R. M.; Griffin, J. L.; Merzaban, J. NMR-based metabolomics in human disease diagnosis: applications, limitations, and recommendations. *Metabolomics* **2013**, *9* (5), 1048-1072.

46. Schicho, R.; Nazyrova, A.; Shaykhutdinov, R.; Duggan, G.; Vogel, H. J.; Storr, M. Quantitative metabolomic profiling of serum and urine in DSS-induced ulcerative colitis of mice by 1H NMR spectroscopy. *Journal of proteome research* **2010**, *9* (12), 6265-6273.

47. Sinclair, A. J.; Viant, M. R.; Wallace, G. R.; Ball, A. K.; Burdon, M. A.; Walker, E. A.; Stewart, P. M.; Young, S. P.; Rauz, S. NMR-Based Metabolomic Analysis of Cerebrospinal Fluid and Serum in Neuro-Ophthalmological and Neurological Diseases-A Diagnostic Tool? *Investigative Ophthalmology & Visual Science* **2009**, *50* (13), 4036.

48. Beckonert, O.; Keun, H. C.; Ebbels, T. M.; Bundy, J.; Holmes, E.; Lindon, J. C.; Nicholson, J. K. Metabolic profiling, metabolomic and metabonomic procedures for NMR spectroscopy of urine, plasma, serum and tissue extracts. *Nature protocols* **2007**, *2* (11), 2692-2703.

49. Worley, B.; Powers, R. Multivariate analysis in metabolomics. *Current Metabolomics* **2013**, *1* (1), 92-107.

50. Lindon, J. C.; Nicholson, J. K.; Holmes, E. *The handbook of metabonomics and metabolomics;* Elsevier: 2011.

51. Medina, S.; Dominguez-Perles, R.; Gil, J. I.; Ferreres, F.; Gil-Izquierdo, A. Metabolomics and the diagnosis of human diseases-a guide to the markers and pathophysiological pathways affected. *Current medicinal chemistry* **2014**, *21* (7), 823-848.

52. Zhang, T.; Wu, X.; Yin, M.; Fan, L.; Zhang, H.; Zhao, F.; Zhang, W.; Ke, C.; Zhang, G.; Hou, Y. Discrimination between malignant and benign ovarian tumors by plasma metabolomic profiling using ultra performance liquid chromatography/mass spectrometry. *Clinica Chimica Acta* **2012**, *413* (9), 861-868.

53. Zira, A.; Kostidis, S.; Theocharis, S.; Sigala, F.; Engelsen, S. B.; Andreadou, I.; Mikros, E. 1 H NMR-based metabonomics approach in a rat model of acute liver injury and regeneration induced by CCl 4 administration. *Toxicology* **2013**, *303*, 115-124.

54. Kim, H. J.; Yoon, Y. R. Pharmacometabolomics: Current Applications and Future Perspectives. *Translational and Clinical Pharmacology* **2014**, *22* (1), 8-10.

55. Huang, Q.; Tan, Y.; Yin, P.; Ye, G.; Gao, P.; Lu, X.; Wang, H.; Xu, G. Metabolic characterization of hepatocellular carcinoma using nontargeted tissue metabolomics. *Cancer research* **2013**, *73* (16), 4992-5002.

56. Yang, Z.; Marotta, F. Pharmacometabolomics in drug discovery & development: Applications and challenges. *Metabolomics: Open Access* **2012**, *2012*.

57. Kumar, B.; Prakash, A.; Ruhela, R. K.; Medhi, B. Potential of metabolomics in preclinical and clinical drug development. *Pharmacological Reports* **2014**, *66* (6), 956-963.

58. Hall, R.; Beale, M.; Fiehn, O.; Hardy, N.; Sumner, L.; Bino, R. Plant metabolomics the missing link in functional genomics strategies. *The Plant Cell* **2002**, *14* (7), 1437-1440.

59. Hong, J.; Yang, L.; Zhang, D.; Shi, J. Plant Metabolomics: An Indispensable System Biology Tool for Plant Science. *International journal of molecular sciences* **2016**, *17* (6), 767.

60. Colburn, W. A.; DeGruttola, V. G.; DeMets, D. L.; Downing, G. J.; Hoth, D. F.; Oates, J. A.; Peck, C. C.; Schooley, R. T.; Spilker, B. A.; Woodcock, J. Biomarkers and surrogate endpoints: Preferred definitions and conceptual framework. Biomarkers Definitions Working Group. *Clinical Pharmacol & Therapeutics* **2001**, *69*, 89-95.

61. World Health Organization ( . International Programme on Chemical Safety. Biomarkers in Risk Assessment: Validity and Validation, 2001. 2015. Ref Type: Generic

62. Koulman, A.; Lane, G. A.; Harrison, S. J.; Volmer, D. A. From differentiating metabolites to biomarkers. *Analytical and bioanalytical chemistry* **2009**, *394* (3), 663-670.

63. Goodsaid, F. M.; Frueh, F. W.; Mattes, W. Strategic paths for biomarker qualification. *Toxicology* **2008**, *245* (3), 219-223.

64. Weckwerth, W.; Fiehn, O. Can we discover novel pathways using metabolomic analysis? *Current Opinion in Biotechnology* **2002**, *13* (2), 156-160.

# *Chapter 2*

## *NMR Spectroscopy and Metabolomics Data Acquisition*

**2.1 A Brief History of NMR**

**2.2 Timeline of Noble Prizes in NMR Spectroscopy**

**2.3 NMR Spectroscopy - General Background**

**2.4 Experiments Used in Metabolomics Studies**

**"Technology is dominated by two types of people: those who understand what they do not manage, and those who manage what they do not understand." - Putt's Law**

## A Brief History of NMR Spectroscopy:

Nuclear Magnetic Resonance (NMR) is one of the most powerful analytical technique presently available to the scientific community. The dominance of NMR over other spectroscopic techniques resides in the vast array of experiments made possible by an almost unlimited choice of radio frequency pulse sequence that may be used to perturb and then observe nuclear spins. The method is specific for isotopes, but nearly every element in the periodic table provides a suitable magnetic nucleus. With the advancement of science, it became apparent that the NMR technique could be applied in many fields, and concurrent developments occurred in physics, chemistry, biology, and other disciplines.

The concept of NMR sprouted after Pauli's prediction of nuclear spin in 1924. Later on, Gorter (1936) attempted to detect the first NMR in bulk matter at 20 MHz in LiF and Al-K alum, which did not worked at low temperature. Isidor Rabi, an American physicist who was awarded the Nobel Prize in Physics in 1944 for his invention of the atomic and molecular beam magnetic resonance method of observing atomic spectra, came across the NMR experiment in the late 1930's but considered it to be an artifact of his apparatus and disregarded its importance.

The phenomenon of Nuclear Magnetic Resonance in bulk material was first observed by two different groups of scientists in different labs, simultaneously in 1945. One of the groups was from Harvard University, including Purcell, Torrey, and Pound[1] whereas the other group was from Stanford University including Bloch, Hansen, and Packard[2]. Bloch and Purcell were jointly awarded the Nobel Prize in Physics for their remarkable discovery in 1952. A few years back, the work of Hendrik Antoon Lorentz and Pieter Zeeman on the influence of magnetism on radiation (the Zeeman Effect)[3] laid the basic foundation for the NMR phenomenon. They also received the Noble Prize in 1902 (in Physics). The discovery of the magnetic moment of the nuclei (proton), one of the most important parameters for NMR led to another Noble prize in 1943 which was given to Prof. Otto Stern[4]. Scientist I.I. Rabi was awarded the Noble prize in 1944 for the discovery of resonance phenomenon[5,6]. After the discovery of NMR in 1945, the next few years were crucial as several of the related parameters such as chemical shift, relaxation, Hahn spin echo, coupling constant, etc. were discovered[7].

Advancement in the technology led to the development of commercial NMR spectrometers, first of which was made in 1952 by Varian Associates[8]. In the coming years with other technological advances, superconducting magnets were invented which led to the development of very high field NMR spectrometers. One of the most important discoveries in NMR spectroscopy was the introduction of Fourier Transformation NMR, i.e., FT-NMR by R.R. Ernst and W.A. Anderson in 1956 which enhances the sensitivity of NMR experiments greatly[9,10]. In later years, the development of Imaging techniques (MRI) was another breakthrough due to its importance in

healthcare, which was proposed by Lauterbur & Mansfield[11,12] and led to another Noble Prize in medicine. Other important experimental techniques such as two-dimensional (2D), heteronuclear 2D NMR, three-dimensional (3D) and 4D NMR methods were developed. These higher dimensional methods proved to be very useful for the structure elucidation of proteins at the atomic level in aqueous environment[13].

NMR played a major role in physics, chemistry, and biology for the development of theories and provided an experimental platform for spectroscopic and quantum mechanical theories. However, the use of NMR in medicine /biology commenced during 1960. Since then NMR spectroscopy has, turn out to be one of the most powerful analytical techniques for elucidating the structure of metabolites in biological samples, proteins, organic molecules, peptides, natural products, RNAs, DNAs. NMR spectroscopy has been used in various disciplines of scientific research as well as in industries and the health care sector. It has been applied in pharmaceutical research and development, structural biology, quantum computing, food industries, molecular dynamics studies of biopolymers, organic synthesis and so on. In time, it has become a multidisciplinary field as it can be applied to different streams, from pharmacy to health care to mathematics, physics, biology, chemical sciences, etc.. Recently it has become one of the best techniques for metabolomics studies which are extensively used in disease diagnosis, toxicological studies, and pharmaceuticals along with many other fields.


**Figure 2.1:** 800 MHz NMR at Centre of Biomedical Research, Lucknow.

## 2.1 Timeline of Noble Prizes in NMR Spectroscopy:

### *Hendrik Antoon Lorentz and Pieter Zeeman*

**The Nobel Prize in Physics 1902**

**Prize motivation:** *"in recognition of the extraordinary service they rendered by their researches into the influence of magnetism upon radiation phenomena."*

### *Otto Stern*

**The Nobel Prize in Physics 1943**

**Prize motivation:** *"for his contribution to the development of the molecular ray method and his discovery of the magnetic moment of the proton."*

### *Isidor Isaac Rabi*

**The Nobel Prize in Physics 1944**

**Prize motivation:** *"for his resonance method for recording the magnetic properties of atomic nuclei."*

### *Felix Bloch and Edward Mills Purcell*

**The Nobel Prize in Physics 1952**

**Prize motivation:** *"for their development of new methods for nuclear magnetic precision measurements and discoveries in connection therewith."*

### *Richard R. Ernst*

**The Nobel Prize in Chemistry 1991**

**Prize motivation:** *"for his contribution to the development of new methodology of high resolution nuclear magnetic resonance (NMR) spectroscopy."*

### *Kurt Wuthrich*

**The Nobel Prize in Chemistry 2002**

**Prize motivation:** *"for his development of nuclear magnetic resonance spectroscopy for determining the three-dimensional structure of biological macromolecules in solution."*

### *Paul C. Lauterbur and Sir Peter Mansfield*

**The Nobel Prize in Medicine 2003**

**Prize motivation:** *"for their discoveries concerning magnetic resonance imaging."*

*A list of Noble Prizes in NMR.*

## 2.2 NMR Spectroscopy - General Background:

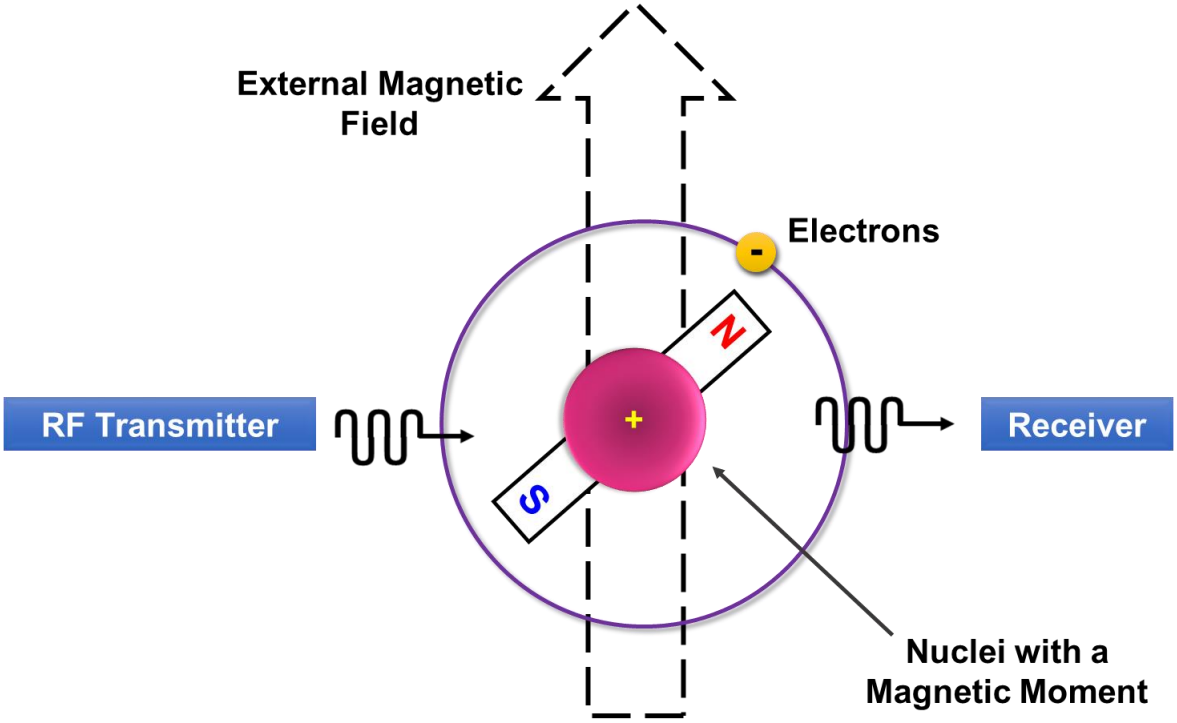



**Figure 2.2:** The nuclear magnetic resonance (NMR) phenomenon.

The phenomenon that is studied by NMR occurs when a static magnetic field is applied to the nuclei of certain atoms ($^{1}$**H**, $^{13}$**C**, $^{31}$**P**, $^{19}$**F**, etc.) **Table 2.1**, followed by a second oscillating field. Any nuclei with an odd number of protons or neutrons can be measured using NMR, though $^{1}$H and $^{13}$C are the most common one. These nuclei exhibit specific rotating electrical charges or spin properties. There are three steps to get NMR signal.

***First step:***

**Polarization** entails interaction between the set of spin properties and a static magnetic field. This results in a balanced or equilibrium state.

***Second step:***

**Resonance** implies exciting radio frequency (RF) field supplies energy to the spin system, perturbing the previously mentioned equilibrium state. The RF field is applied in short pulses, simultaneously exiting the spin system into transitions between energy states.

***Third step:***

**Relaxation** occurs immediately after the RF pulse when the spin system returns to the original balanced state. As the relaxation advances, a current is induced in an RF coil, which picks up all the transitions nuclei simultaneously. This current or exponentially decaying signal is known as free induction decay (FID), which is then Fourier transformed, resulting in a spectrum.

| Isotope | Spin | Natural abundance (%) |
|---|---|---|
| $^{1}H$ | ½ | 99.99 |
| $^{2}D$ | 1 | 0.02 |
| $^{13}C$ | ½ | 1.07 |
| $^{15}N$ | ½ | 0.37 |
| $^{19}F$ | ½ | 100.00 |
| $^{31}P$ | ½ | 100.00 |
| $^{23}Na$ | 1 ½ | 100.00 |
| $^{67}Zn$ | 2 ½ | 4.10 |

**Table 2.1:** NMR active nuclei and their properties.

The magnetic properties of the atomic nucleus form the basis of nuclear magnetic resonance spectroscopy. Most of the nuclei possess angular momentum 'P' which is also responsible for the presence of a magnetic moment 'μ' in the nuclei. According to the quantum theory, angular momentum and nuclear magnetic moment are quantized. The two quantities are related as the following expression: $\mu = \gamma P$.

The proportionality factor 'γ' is a constant for each nuclide and is called the "Magnetogyric ratio" or "Gyromagnetic ratio." For most of the nuclei, angular momentum and nuclear magnetic moment vectors are parallel however for a few cases such as $^{15}N$; these are antiparallel. The angular momentum and magnetic moment are related to the 'nuclear spin' I as the following expression:

$$\boldsymbol{P} = \sqrt{\boldsymbol{I}(\boldsymbol{I}+\mathbf{1})} \times \hbar$$

$$\boldsymbol{\mu} = \boldsymbol{\gamma}\sqrt{\boldsymbol{I}(\boldsymbol{I}+\mathbf{1})} \times \hbar$$

Here, $\hbar = h/2\pi$, where 'h' is the Planck's constant. The spin quantum number can have values greater than or equal to zero, and multiples of ½. When I=0, the mass, and atomic numbers are both even, and the nuclear spin is not observable with NMR spectroscopy.

Luckily, the key organic chemical elements have at least one nucleus that does possess an observable nuclear spin. These are the elements that can be of use in NMR spectroscopy. The reason why the nuclear spin is fundamental to the NMR phenomenon is that the spinning nucleus possesses an angular momentum which gives rise to an associated magnetic moment. When placed in an external magnetic field, such as the NMR instrument, the magnetic moments align themselves relative to the field in a discrete number of orientations, because the energy states are quantized. For a spin of the magnetic quantum number I, there exist 2I+1 possible spin states, which means that for a spin-½ nucleus (a nucleus with an odd mass number) such as the proton,

there exists two possible spin states denoted –½ and +½. Another example, deuterium with I=1, has three spin states, denoted -1, 0, and +1 and so on. The spin -½ nucleus can be thought of having an orientation parallel to the external magnetic field (**α**-state), or an orientation antiparallel to the magnetic field (**β**-state). Since the α-state is usually lower in energy, there are a bit more nuclei in that state. The energy difference between the two spin levels is rather small and the population difference between the, **α**- and **β**- states can be calculated with the Boltzmann distribution. The population difference between the states depends on the strength of the external magnetic field and temperature. The drawback is that only the population difference between the **α**- and **β**- states can be detected by means of NMR spectroscopy and this is what makes NMR spectroscopy insensitive relative to other techniques such as infrared or ultraviolet spectroscopy.

The magnetic moment of the spin can be stimulated with short duration radio frequency energy, called a pulse. Because the energy states are quantized, the pulse induces changes in the populations: **α→β** and **β→α** of which **α→β** is, the more likely one; this meaning that if the pulse is repeated many times, the spin populations even out and the system is said to saturate. After the pulse is applied, the emission signal of the energy stored in the nucleus (called the FID, free induction decay) can be detected. As the nuclei emit energy and relax, the population states return to their normal state and can be restimulated. The FID, known as the time-domain, can be converted to the frequency-domain spectrum by a process known as Fourier transformation (FT). The time-domain-format must be converted because it is much harder to interpret than the frequency-domain where one can explicitly see the signals arising from nuclei. The domains of time and frequency are functions of intensity versus time, and intensity versus frequency, respectively. There is a significant relation between the frequency of the NMR signal and the chemical structure of a molecule; this can be used to identify specific signals and can only be done in the frequency domain.

## Important NMR parameters:

**Relaxation**

During an NMR experiment, nuclear spins are promoted from a lower energy spin state to a higher energy spin state by absorption of energy. This energy dissipates into the surrounding environment until the system is once again in its equilibrium state. This process of shedding excess energy is called as relaxation. The relaxation process affects the line width and intensities. There are two major relaxation processes:

1. Spin-lattice (longitudinal) relaxation
2. Spin-Spin (transverse) relaxation

**Chemical shift:**

In terms of NMR, nuclei of different elements will presses with different larmour frequencies and give signals at different frequencies in a particular magnetic field due to their different

gyromagnetic ratios. Similarly, nuclei of the same type can resonate at different frequencies giving signals at different frequencies. This condition can occur if the local magnetic field experienced by a nucleus is slightly different from that of another similar nucleus. These variations in the local magnetic fields can be explained on the basis of electromagnetic theory. According to this, when a molecule is placed in an external magnetic field Bo, it induces electron currents in the plane perpendicular to the applied magnetic field. These induced currents then produce a smaller magnetic field which opposes the applied field and partially cancels the applied field, thus shielding the nucleus. The induced opposing field is about a million times smaller than the external magnetic field, so the magnetic field perceived by the nucleus will be very slightly different from the applied field. The effect of this small shielding magnetic field varies with the electron density at a particular nucleus. So these chemically different nuclei shields to a different extent and therefore resonate at different frequencies. The induced magnetic fields are million times smaller than the external magnetic field, and the differences in resonance frequencies for two different nuclei will be of the order of several hertz (for proton nuclei). The relative positions of two signals in the NMR spectrum can be measured with greater accuracy than the absolute frequencies. So a reference signal is chosen, and the difference between the position of the signal of interest and that of the reference is measured which is termed the chemical shift[14]. If chemical shift is measured in Hz, then it would be magnetic field dependent, so to remove this dependence, chemical shift is expressed in parts per million (ppm) and defined as where the difference between the resonance frequency of the reference and the sample $\boldsymbol{n_{ref} - n_{sample}}$ measured in hertz (e.g., 75 Hz) divided by the spectrometer's operating frequency (e.g. 500 MHz) gives the chemical shift (e.g., 0.15 ppm). Typical ranges in chemical shifts for signals emanating from biochemically important samples are $^{1}$H, 15 ppm; $^{13}$C, 250 ppm; $^{15}$N, 400 ppm; and $^{31}$P, 35 ppm[14-16].

**J-coupling:**

If two spins A and B share a common electronic orbital and are separated by less than five chemical bonds, there is an interaction that is mediated through the spin-orbit coupling of the nuclear spins with the electronic system. Due to this interaction, the magnetic field seen by these spins varies, and this variation is orientation dependent. As a consequence, the signals of spins A and B split and appear as doublets with the intensity ratio of 1:1. The difference between these two peaks of doublets (measured in Hz) is a direct measure of the electronic overlap which consequently tells about the topology and the nature of the chemical bonds linking both these spins. The difference between these peaks is termed as the scalar coupling constant and denoted as J. The j-coupling is magnetic field independent. There exist two types of scalar coupling[17]: homonuclear and hetero-nuclear between two similar and dissimilar nuclei respectively. While providing the scalar coupling value number of bonds separating two spins is also provided, so 3JHC refers to a three-bond coupling measured for a $^{1}$H nucleus separated from a $^{13}$C nucleus by three bonds. Spins having scalar coupling to more than one spin give rise to very complicated but

informative coupling pattern. On the basis of these patterns different kinds of spin systems has been deduced, such as (i) AX spin system (consisting of two spins sharing one coupling which gives two doublet signals (1:1 ratio) of equal intensity in NMR spectrum) and (ii) AX2 spin system (consisting of three spins: spin A corresponding to one nucleus (e.g. CH) whereas spin X, the magnetically equivalent spin, corresponds to two nuclei (e.g. CH2) having identical chemical shifts and coupling topology to spin A. The spectrum of the AX2 system consists of one triplet signal (1:2:1ratio of the three components) with intensity 1 for spin A, and one doublet signal with intensity 2 for spins X2). The scalar coupling can be weak if the coupling constants between two spins are much smaller than the differences of their chemical shifts (both in Hz) and strong if the coupling constant is equal to or greater than chemical shift difference.

### Signal Intensity

The quantitative information of the concentration of chemical compounds can also be obtained from the NMR spectrum by calculating the spectrum intensity or the area under the peak (the integral of a peak is considered to be the area under that peak). The area under the peak is proportional to the number of nuclei that contribute to that chemical compound. If there are multiplets in a spectrum, the whole group of peaks should be integrated. Just like the chemical shifts and indirect spin-spin couplings, the signal intensities are also important for the structure determination. The signal intensities will help in the quantification of mixtures. However, the effect of spin-lattice relaxation (T1) and spin-spin relaxation (T2) could also influence the signal intensities. These relaxation effects could be used for selective detection or suppression of specific signals.

## 2.3 Experiments Used in Metabolomics Studies:

NMR spectroscopy has played a pivotal role in our understanding of metabolism and metabolic processes past many decades. Its extraordinary capacity to resolve thousands of peaks in complex metabolite mixtures such as blood and urine led NMR to be the technology of choice to initially develop the field of metabolomics. The technology behind NMR spectroscopy is quite advanced and most technological, or equipment advances over the past decade have been relatively incremental. The developments in instrumentation, such as higher magnetic field magnets[18], cryogenically cooled probes[19], and microprobes[20,21], are major advancements that have enhanced the resolution and sensitivity of the method. Another recent technological development of note is called dynamic nuclear polarization (DNP). However, it is not possible to use all kinds of pulse sequences for metabolomics studies as these samples contain relatively very high concentrations of water proton than those of metabolite protons. For samples with very low metabolite concentration suppression of water, signals become paramount and up to 100 times suppression of water signal is recommended. Although various types of NMR experiments are readily available, simple one-dimensional (1D) proton ($^1$H) NMR spectroscopy has become the

standard data collection technique for most NMR-based metabolomics studies. As a data collection method, it is fast, simple, highly reproducible and readily yields spectra that contain hundreds or even thousands of peaks. However, the main limitation of 1D $^{1}H$ NMR spectroscopy is the fact that it generates spectra with many overlapping signals. These are due to the limited spectral dispersion of $^{1}H$ chemical shifts, which can lead to uncertainties in the assignment and quantification of many metabolites. Other methods, such as one-dimensional carbon ($^{13}C$) NMR spectroscopy, exhibit much greater spectral dispersion compared with $^{1}H$ NMR spectroscopy. As a result, $^{13}C$ NMR has been proposed as an alternative method that can help overcome the spectral overlap problem of $^{1}H$ NMR[22]. However, the low natural abundance of $^{13}C$ nuclei (1.1 %) and the low gyromagnetic ratio contribute to the relatively low sensitivity of this nucleus and hinder the widespread use of $^{13}C$ in NMR-based metabolomics applications. In the subsequent sections, various one-dimensional experiments for proton ($^{1}H$) and X-nuclei ($^{13}C$, $^{15}N$, $^{31}P$, etc.), solvent suppression techniques and different two-dimensional experiments are described.

**Single pulse experiment:**

This experiment is the simple 1D experiment providing chemical shift and coupling constant information. This experiment consists of a delay and 90° hard pulse which is followed by acquisition time. One pulse experiments cannot be used for most of the native biofluids samples due to the presence of high water content. So a very intense peak of water makes it impossible to see any other metabolites, hence the need of suppressing this intense water signal using different suppression techniques.

**Water suppression techniques:**

In the past few years, various suppression techniques have been developed, which includes presaturation techniques and gradient pulse based techniques. Most widely used suppression technique is the presaturation technique. In this method, water signal is saturated by irradiating it during the inter-scan delays. This presaturation part can be incorporated in other experiments like a 1D version of NOESY, in which the continuous irradiation of water signal is done during the inter-scan delay and mixing time. In this experiment, mixing time has a very good effect on the flat baseline, thus enhancing the interpretability of the spectra. In gradient pulse based methods, a gradient pulse is used for presaturation. Here the gradient pulse creates an inhomogeneous magnetic field along the z-axis. Due to this gradient, the magnetization in transverse plane will be dephased. One such pulse sequence is WET[23] (water suppression enhanced through T1 effect) in which water signal is suppressed by applying a series of weak frequency selective pulses interleaved with gradient pulses. Another very useful pulse sequence for water suppression is WATERGATE[24] in which two frequency-selective 90° pulses are applied along with gradient pulses during the delay-180°-delay sequence. When using WATERGATE sequence, water suppression is very good. However the signals adjacent to water resonance are also affected due

to these selective pulses. So using these water suppression techniques, decent looking NMR spectra can be obtained which show the less intense water resonance and enhanced metabolites signals.

**NOESY**

Stands for nuclear overhauser enhanced spectroscopy and the sequence is based on nuclear overhauser principle that small alteration in the electron spin populations would produce a large change in the nuclear spin polarization. The pulse sequence comprises of three alternative $90^0$ pulses with relaxation delay (RD) of 2.0 seconds and an acquisition time of 2.5 seconds. NOESYPR sequence is used to acquire urine spectrum, which enhances the detection of metabolites from urine samples. An NOESY spectra of the human urine sample is presented in **Figure 2.3**.

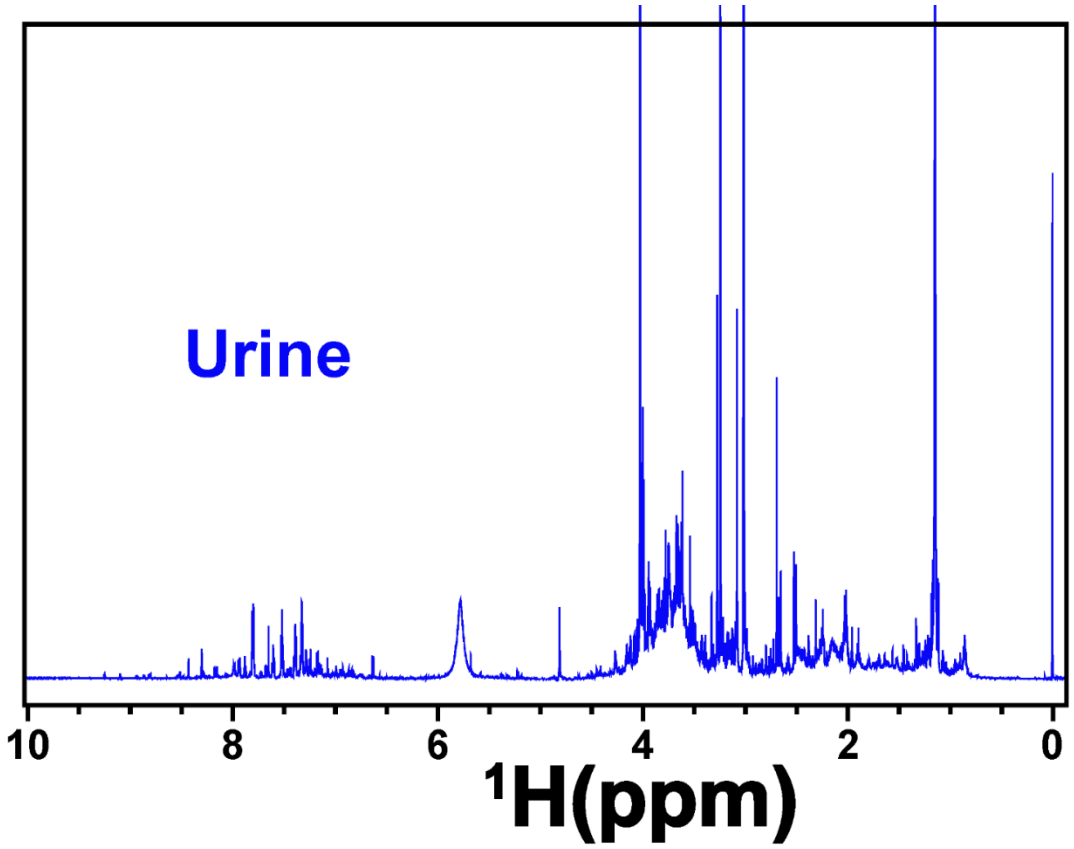



**Figure 2.3:** $^1$H-NMR spectrum of the human urine sample.

**Relaxation editing technique (CPMG):**

Macromolecules like proteins, lipids, etc. with very high molecular weight and a high degree of rigidity show fast transverse relaxation. These molecules also cause magnetic field inhomogeneity in samples. This property has been exploited very nicely in metabolomics studies where signals from these macromolecules can be filtered/ removed as they are very broad. For this purpose, a pulse sequence called CPMG, named after its inventors Carr Purcell Meiboom and Gill has been developed for removing this inhomogeneity and measuring transverse relaxation by applying an 180° pulse before acquisition[25,26]. In metabolomics, CPMG pulse sequence is used for the analysis of blood serum, seminal fluids, etc. in which several macromolecules are present, and their signals need to be suppressed. The CPMG pulse sequence is also very useful for HR-MAS studies of animal tissues as well as plant tissues. Water suppression techniques can also be incorporated into

CPMG sequence, making it very robust and NMR spectra obtained with this combination have an improved receiver gain and excellent resolution for very low molecular weight metabolites[15]. A CPMG spectrum of the human serum and urine sample is presented in **Figure 2.4**.

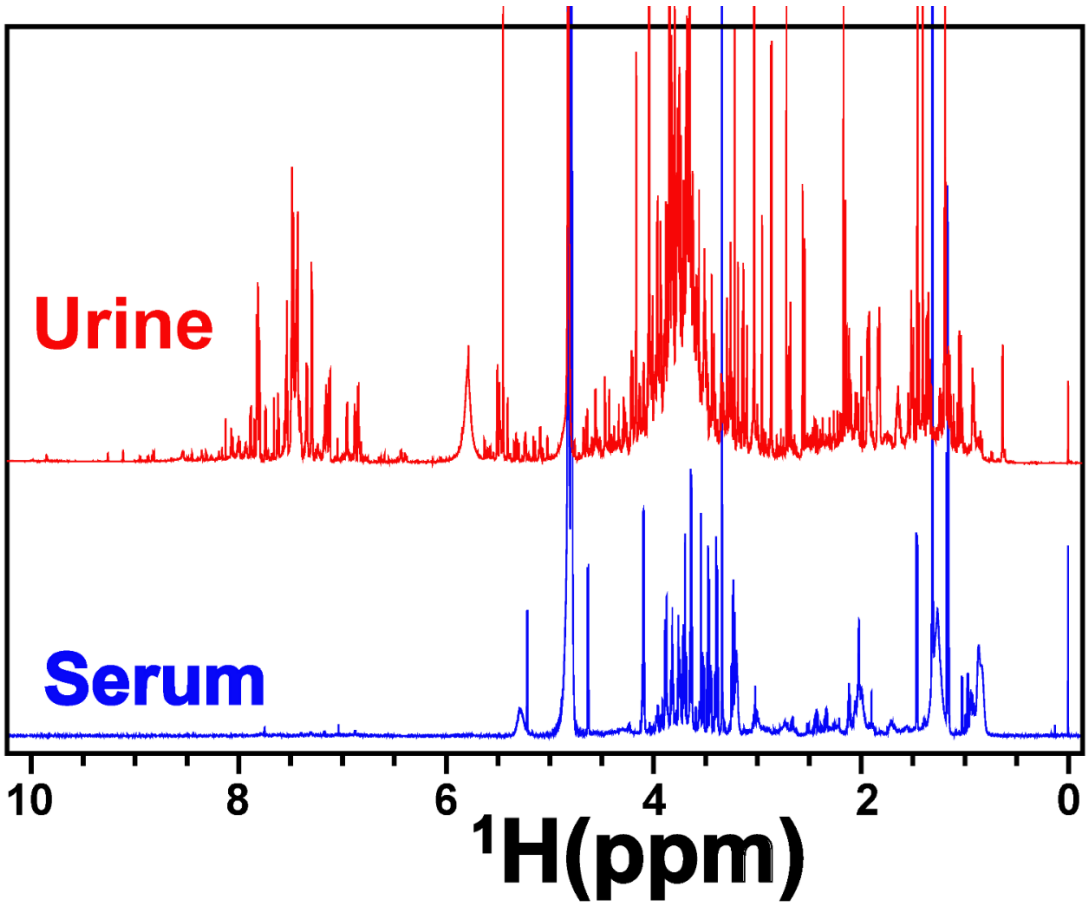



**Figure 2.4:** $^{1}$H-NMR spectrum of human serum and urine sample.

**Diffusion Edited:**

Diffusion editing aims to simplify overcrowded 1D NMR spectra by utilizing the differences in molecular diffusion coefficients either alone or by using a combination of diffusion and relaxation parameters. Therefore, diffusion editing methods are commonly used for the same type of purpose than relaxation editing method. These experiments are usually based on two standard types of pulse sequences, the spin echo diffusion sequence, and the stimulated echo sequence; both of these are based on pulsed field gradients (PFGs). In complex biofluids, the combination of both diffusion and relaxation editing brings about considerable spectral simplification leading to an easier resonance assignment and quantification process. It is known that the separation of slowly diffusing species (generally large macromolecules) can be separated from fast diffusing ones (small metabolites) by diffusion editing. The diffusion properties are most commonly used to select macromolecular signals. However, by combining diffusion and relaxation editing simultaneously, one can, for example, remove the peaks of the smallest, fastest diffusing components (e.g. solvents such as $H_2O$) and at the same time remove peaks from the fastest relaxing macromolecular components, thus resulting in a spectrum displaying small metabolites

and lipids. A diffusion edited spectra of the human serum and bile sample is presented in **Figure 2.5**.

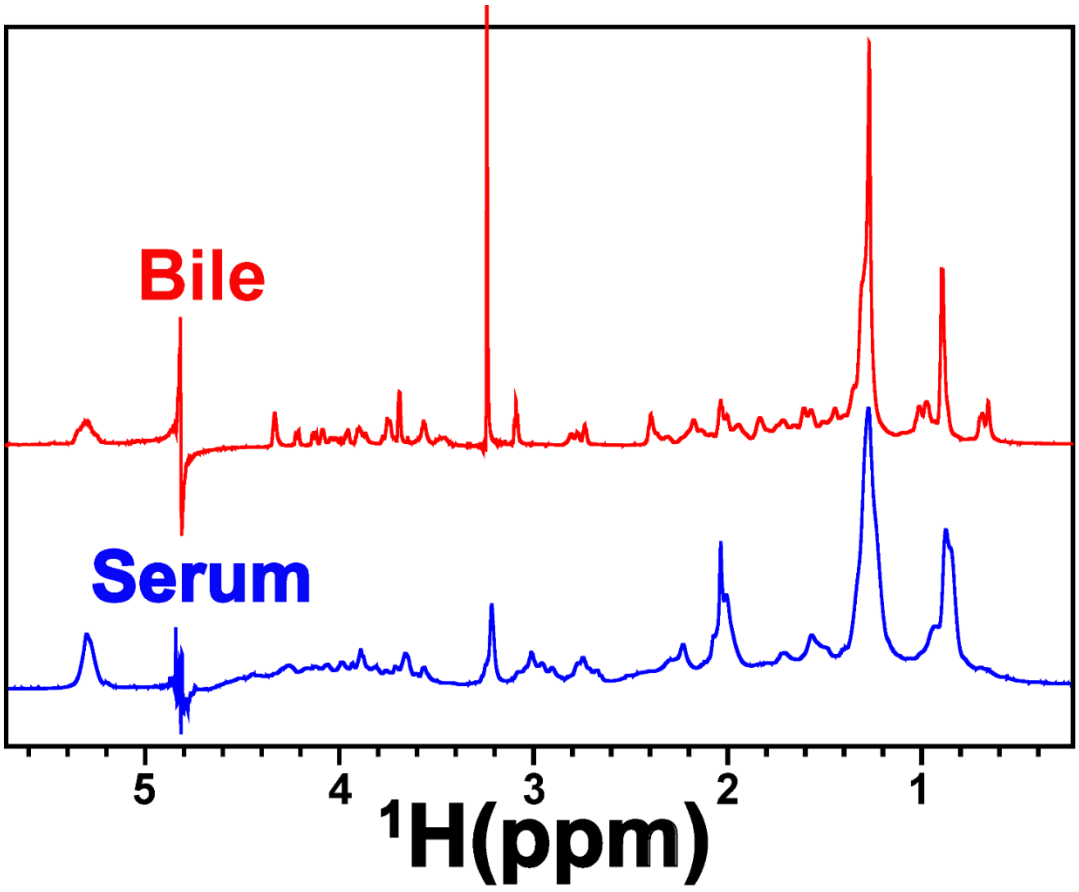



**Figure 2.5:** ¹H-NMR spectrum of human serum and bile sample.

**Two-dimensional NMR:**

One-dimensional NMR is sufficient for the spectral assignment of metabolomics spectra, however, in cases where there are severe overlaps in several NMR signals, it becomes tough to assign unambiguously. In these cases, two-dimensional NMR becomes essential[13,14]. The 2D NMR methods were developed in the mid-1970s, in which a series of the 1D experiment is recorded by systemically varying one experimental parameter. In general, an inter-pulse delay between two pulses is the commonly varying parameter which is called as t1. During t1 period, a coherent magnetic state begins to evolve, which can be monitored by stepwise incrementation. If the evolving parameter is chemical shift precession, then the measured parameter in 2D NMR will be the chemical shift of spins involved. So the evolving spin state will be transferred to the 2nd spin by the end of the t1 period, this magnetization transfer is then detected as an NMR signal evolving under t2. After Fourier transformation in both dimensions, a 2D spectrum is obtained where the spectral amplitude is plotted as the function of two frequency dimension. The cross-peak in 2D spectrum represents the pair of spins between which the magnetization transferred has occurred. When both interacting spins are of the same type such as proton 2D experiment is termed as homonuclear otherwise, it is called heteronuclear. Though these experiments are very useful regarding information recovery, they cannot be used for a large array of samples due to

very long acquisition time. Some of the important 2D experiments are described in the following section.

**Homonuclear 2D NMR experiments:**

**J-resolved:**

The JRES or J-resolved spectroscopy is the simplest 2D experiment in which the J-coupling information is separated into the second dimension from the chemical shift. From this experiment, a chemical shift spectrum can be obtained as a tilted projection of full 2D matrix. Here all signals appear as singlet's which is placed at the mid of their respective multiplets in 1D projection spectra. This reduces the spectral complexity greatly. This experiment has been used in many metabolomics studies for assigning the metabolites in biofluids spectra[15,27-30]. A JRES spectrum of the human urine sample is presented in **Figure 2.6**.

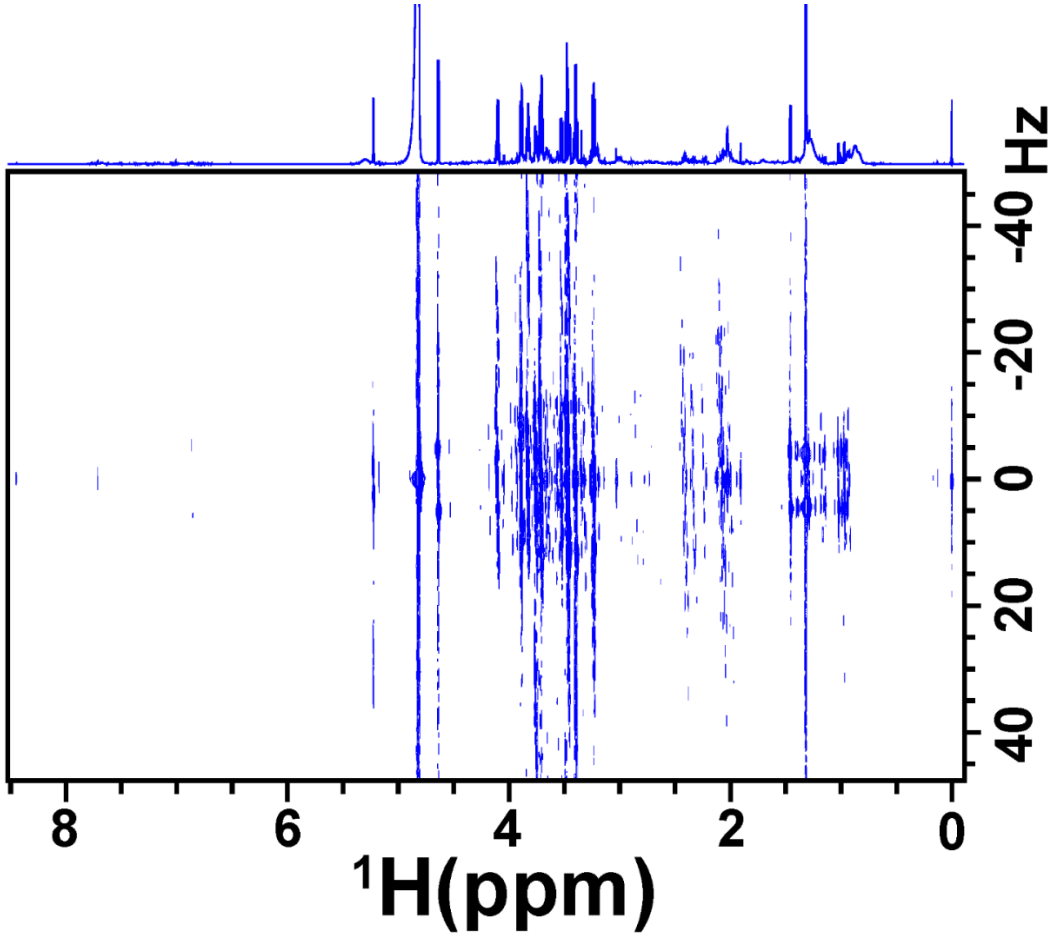



**Figure 2.6:** ¹H-JRES NMR spectrum of the human urine sample.

**Correlation Spectroscopy (COSY and DQF-COSY):**

These are the basic experiments in 2D NMR spectroscopy. In this method cross-peaks are obtained between 2 spins connected through homonuclear J-coupling over 3-5 bonds, providing

the information about spin system topology. In metabolomics studies, Double Quantum Filtered COSY (DQF-COSY) is preferred over COSY as the resolution of diagonals is enhanced due to the suppression of intense signals coming from singlets. Biofluids contain several metabolites giving only singlets so DQF-COSY gives good resolution and should be used for such samples[15,31]. A COSY spectrum of the human serum sample is presented in **Figure 2.7**.

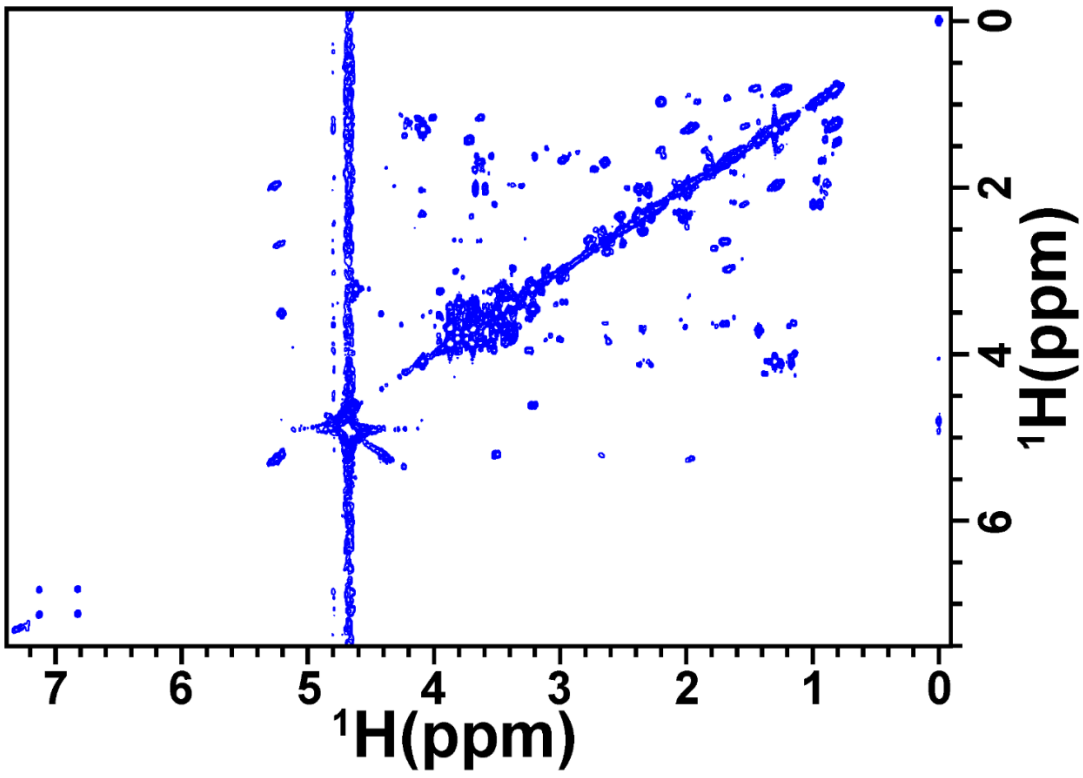



**Figure 2.7:** $^1$H-$^1$H COSY NMR spectrum of the human serum sample.

**Total correlation spectroscopy (TOCSY):**

The TOCSY experiment is used to identify all the spin system from one metabolite which are J-coupled. This is done by simultaneously transferring the magnetization over many spins. So the signal appearing here are not only from J-coupled spins but also from the spins that are not directly connected through H-X-H bonds. In this experiment, mixing time plays a very vital role as during this period only the magnetization is transferred along the J-coupled network. In metabolomics studies most widely used mixing pulse sequence is MLEV which is incorporated in between t1 evolution and signal acquisition. Optimum mixing time duration for metabolomics studies is 80-300 ms. The TOCSY experiments have been applied very frequently for confirming the spectral assignment in various metabolomics applications. A TOCSY spectrum of the human serum sample is presented in **Figure 2.8**.

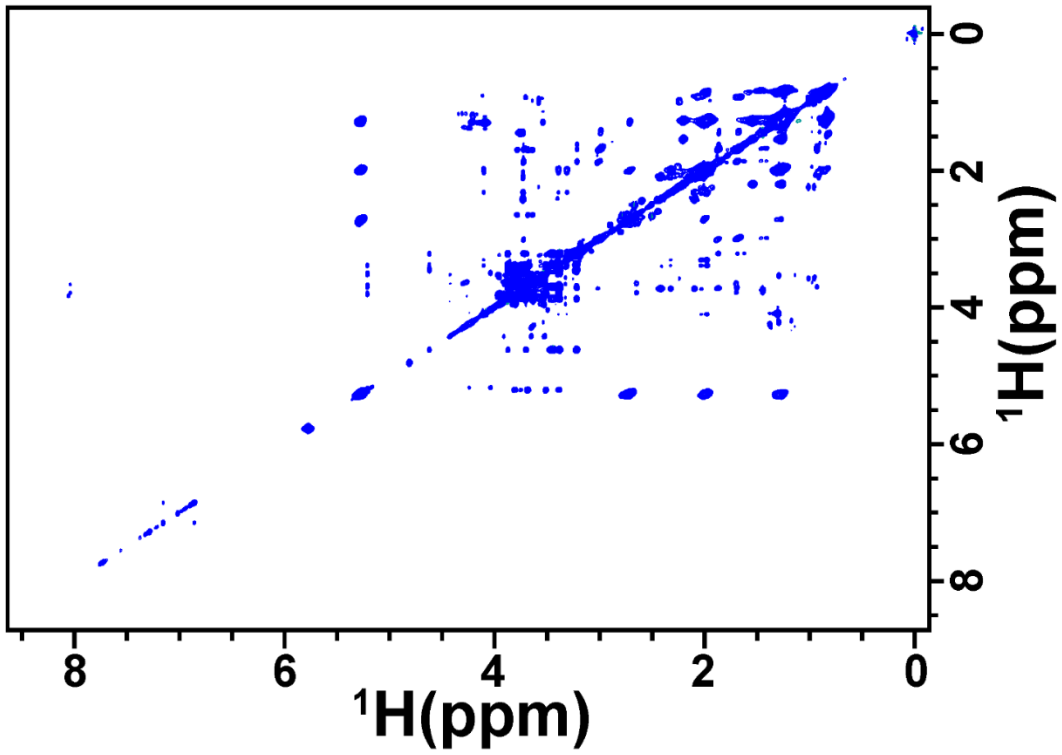



**Figure 2.8:** $^{1}$H-$^{1}$H TOCSY NMR spectrum of the human serum sample.

**Heteronuclear 2D NMR experiments:**

In these types of experiments spins, studies are of a dissimilar kind. One nucleus here is a proton whereas another one can be any of X nuclei such as $^{13}$C, $^{15}$N, etc.. In metabolomics mostly 13C nuclei is studied. For observing $^{13}$C instead of direct, indirect detection (through J-coupled 1H spins) is preferred. In this detection technique, sensitivity is enhanced very much due to the very high gyromagnetic ratio of $^{1}$H and chemical shifts of $^{13}$C and J-coupled $^{1}$H are correlated which plays a crucial role in heteronuclear correlation[15]. Heteronuclear 2D experiments used in metabolomics are described in the following section.

**Heteronuclear Single Quantum Coherence (HSQC):**

As stated, that the chemical shift of $^{13}$C is correlated to directly bonded $^{1}$H chemical shift through one bond coupling constant so this experiment provides these two coupled nuclei. The $^{1}$H-$^{13}$C HSQC is one of the most important and frequently used experiments in metabolomics for establishing proton and carbon correlations for metabolites in biofluids thus aiding the assignment of metabolites[15,16]. Due to highly dispersed chemical shift range of $^{13}$C and very high sensitivity, it can be used in routine metabolomics practices. A typical HSQC spectrum is shown in **Figure 2.9**.

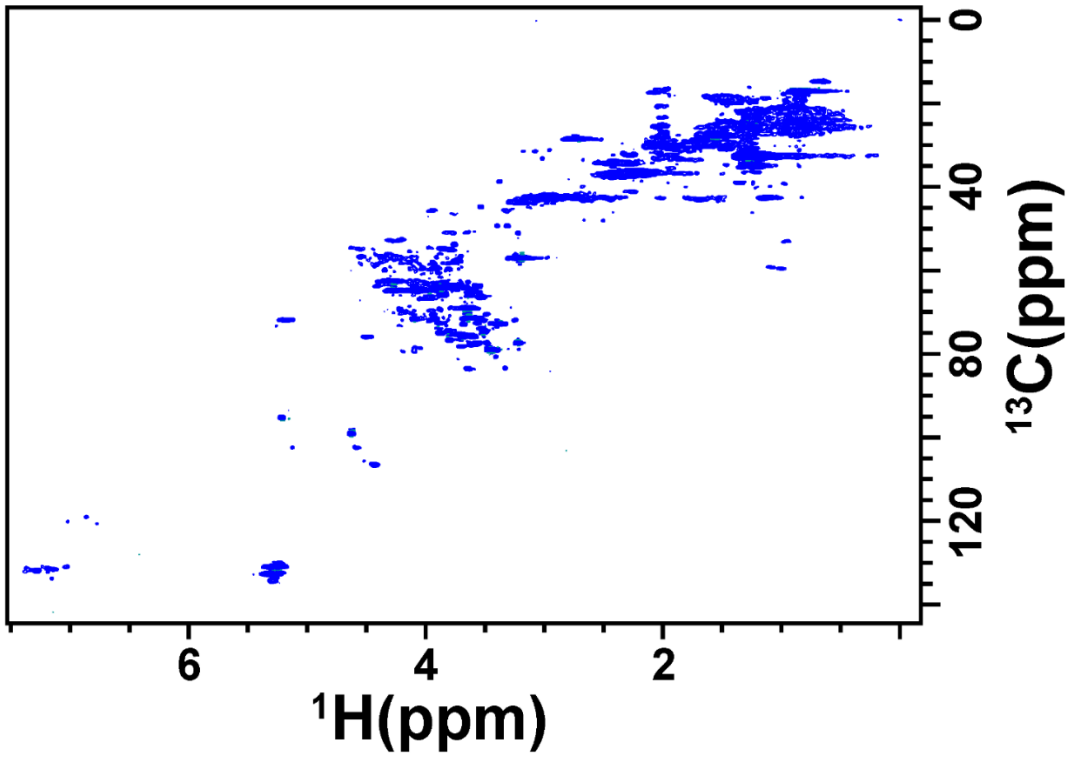



**Figure 2.9:** $^{1}$H-$^{13}$CNMR spectrum of serum from a healthy control.

## Conclusion

Metabolomics represents a major and rapidly evolving field. The development of NMR experiments allows for accurate measurement of many metabolites in different biofluids. The versatility of NMR has enabled this analytical technique to make valuable contributions to biomarker discovery in metabolomics field of biofluids. The application of NMR to the study of different systems is increasing with each passing day.

**Reference's**


1. Purcell, E. M.; Torrey, H. C.; Pound, R. V. Resonance absorption by nuclear magnetic moments in a solid. Physical review 1946, 69 (1-2), 37.

2. Bloch, F. Nuclear induction. Physical review 1946, 70 (7-8), 460.

3. Zeeman, P. XXXII. On the influence of magnetism on the nature of the light emitted by a substance. The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science 1897, 43 (262), 226-239.

4. Estermann, I.; Simpson, O. C.; Stern, O. The magnetic moment of the proton. Physical review 1937, 52 (6), 535.

5. Rabi, I. I.; Millman, S.; Kusch, P.; Zacharias, J. R. The Magnetic Moments of Li 6 3, Li 7 3 and F 19 9. Physical review 1938, 53 (6), 495.

6. Rabi, I. I.; Millman, S.; Kusch, P.; Zacharias, J. R. The Molecular Beam Resonance Method for Measuring Nuclear Magnetic Moments. The Magnetic Moments of Li 6 3, Li 7 3 and F 19 9. Physical review 1939, 55 (6), 526.

7. Freeman, R. A short history of NMR. Chemistry of Heterocyclic Compounds 1995, 31 (9), 1004-1005.

8. Emsley, J. W.; Feeney, J. Forty years of progress in nuclear magnetic resonance spectroscopy. Progress in Nuclear Magnetic Resonance Spectroscopy 2007, 50 (4), 179-198.

9. Ernst, R. R.; Anderson, W. A. Sensitivity enhancement in magnetic resonance. II. Investigation of intermediate passage conditions. Review of scientific instruments 1965, 36 (12), 1696-1706.

10. Ernst, R. R.; Anderson, W. A. Erratum: Sensitivity Enhancement in Magnetic Resonance. II. Investigation of Intermediate Passage Conditions. Review of scientific instruments 1966, 37 (9), 1275.

11. Lauterbur, P. C. Image formation by induced local interactions: examples employing nuclear magnetic resonance. 1973.

12. Mansfield, P.; Grannell, P. K. NMR'diffraction'in solids? Journal of Physics C: solid state physics 1973, 6 (22), L422.

13. Wuthrich, K. NMR of proteins and nucleic acids; Wiley: 1986.

14. Gunter, H. NMR spectroscopy: basic principles, concepts, and applications in chemistry. NMR spectroscopy: basic principles, concepts, and applications in chemistry 1995.

15. Beckonert, O.; Keun, H. C.; Ebbels, T. M.; Bundy, J.; Holmes, E.; Lindon, J. C.; Nicholson, J. K. Metabolic profiling, metabolomic and metabonomic procedures for NMR spectroscopy of urine, plasma, serum and tissue extracts. Nature protocols 2007, 2 (11), 2692-2703.

16. Kalinowski, H. O.; Berger, S.; Braun, S. Carbon-13 NMR spectroscopy. 1988.

17. Kup-ìe, Æ.; Freeman, R. Compensation for spinGÇôspin coupling effects during adiabatic pulses. Journal of Magnetic Resonance 1997, 127 (1), 36-48.

18. Gruetter, R.; Weisdorf, S. A.; Rajanayagan, V.; Terpstra, M.; Merkle, H.; Truwit, C. L.; Garwood, M.; Nyberg, S. L.; Ugurbil, K. +. Resolution Improvements inin Vivo 1 H NMR Spectra with Increased Magnetic Field Strength. Journal of Magnetic Resonance 1998, 135 (1), 260-264.

19. Keun, H. C.; Beckonert, O.; Griffin, J. L.; Richter, C.; Moskau, D.; Lindon, J. C.; Nicholson, J. K. Cryogenic probe 13C NMR spectroscopy of urine for metabonomic studies. Analytical chemistry 2002, 74 (17), 4588-4593.

20. Sukumaran, D. K.; Garcia, E.; Hua, J.; Tabaczynski, W.; Odunsi, K.; Andrews, C.; Szyperski, T. Standard operating procedure for metabonomics studies of blood serum and plasma samples using a 1HGÇÉNMR microGÇÉflow probe. Magnetic Resonance in Chemistry 2009, 47 (S1), S81-S85.

21. Grimes, J. H.; OGÇÖConnell, T. M. The application of micro-coil NMR probe technology to metabolomics of urine and serum. Journal of biomolecular NMR 2011, 49 (3-4), 297-305.

22. Rai, R. K.; Sinha, N. Fast and Accurate Quantitative Metabolic Profiling of Body Fluids by Nonlinear Sampling of 1HGÇô13C Two-Dimensional Nuclear Magnetic Resonance Spectroscopy. Analytical chemistry 2012, 84 (22), 10005-10011.

23. Smallcombe, S. H.; Patt, S. L.; Keifer, P. A. WET solvent suppression and its applications to LC NMR and high-resolution NMR spectroscopy. Journal of magnetic resonance, Series A 1995, 117 (2), 295-303.

24. Liu, M.; Mao, X. a.; Ye, C.; Huang, H.; Nicholson, J. K.; Lindon, J. C. Improved WATERGATE pulse sequences for solvent suppression in NMR spectroscopy. Journal of Magnetic Resonance 1998, 132 (1), 125-129.

25. Carr, H. Y.; Purcell, E. M. Effects of diffusion on free precession in nuclear magnetic resonance experiments. Physical review 1954, 94 (3), 630.

26. Meiboom, S.; Gill, D. Modified spinGÇÉecho method for measuring nuclear relaxation times. Review of scientific instruments 1958, 29 (8), 688-691.

27. Lindon, J. C.; Nicholson, J. K.; Everett, J. R. NMR spectroscopy of biofluids. Annual reports on NMR spectroscopy 1999, 38, 1-88.

28. Ludwig, C.; Easton, J. M.; Lodi, A.; Tiziani, S.; Manzoor, S. E.; Southam, A. D.; Byrne, J. J.; Bishop, L. M.; He, S.; Arvanitis, T. N. Birmingham Metabolite Library: a publicly accessible database of 1-D 1H and 2-D 1H J-resolved NMR spectra of authentic metabolite standards (BML-NMR). Metabolomics 2012, 8 (1), 8-18.

29. Parsons, H. M.; Ludwig, C.; Gunther, U. L.; Viant, M. R. Improved classification accuracy in 1-and 2-dimensional NMR metabolomics data using the variance stabilising generalised logarithm transformation. BMC bioinformatics 2007, 8 (1), 234.

30. Viant, M. R.; Lyeth, B. G.; Miller, M. G.; Berman, R. F. An NMR metabolomic investigation of early metabolic disturbances following traumatic brain injury in a mammalian model. NMR in Biomedicine 2005, 18 (8), 507-516.

31. Ernst, R. R.; Bodenhausen, G.; Wokaun, A. Principles of nuclear magnetic resonance in one and two dimensions; 14 ed.; Clarendon Press Oxford: 1987.

# *Chapter 3*

## *Methodology: From Scratch to Biomarker Discovery*



**"Data does not equal information; information does not equal knowledge; and, most importantly of all, knowledge does not equal wisdom. We have oceans of data, rivers of information, small puddles of knowledge, and the odd drop of wisdom"- Henry Nix**

## 3.1: Workflow for Metabolomics:

The primary objective in Metabolomics is to study the metabolism using either quantitatively or semi-quantitative approaches. In metabolomic studies, it is followed a well-established pipeline going from experimental design and statistical data analysis to biological interpretation of the results. **Figure 3.1** depicts the stages of the typical workflow for metabolomics studies.

***Step 1:*** **Defining the aim/rationale of the study:**

Formulating the purpose and objectives of the metabolomics study is the first step in all metabolomics study. There should be several questions to be answered and taken into consideration in both the design of the study as well as in the evaluation of the outcome. The questions may be of the nature of like what is already known? What additional method or information is required? What experimentations are needed and how to optimize them? These questions should be answered. Otherwise, the study will be pointless.

***Step 2:*** **Selecting the objectives:**

A clear objective is critical to identifying the ideal metabolomics approach for the study. The objective selection needs to span the experimental domain in a balanced and systematic manner. To achieve this, objects should be characterized with measured and observed descriptors. Selection needed setting up specific inclusion and exclusion criteria for the study. Projection based method and multivariate design are two main object selection methods.

***Step 3:*** **Sample preparation and characterization:**

The metabolic analysis should be global, quantitative, robust, reproducible, accurate, and interpretable. Moreover, there should be minimum experimental and biological variation. The physicochemical diversity of metabolites, including amino acids, fatty acids, carbohydrates, and organic acids, raises problems for extraction and working procedures in different analytical techniques. So it becomes imperative to design experiments, very wisely as it represents to be a major strategy for systematically investigating the responsible factors and optimize the experimental protocols.

***Step 4:*** **Evaluation of the collected data:**

It is very important and critical step in the protocol of chemometric analysis of NMR metabolomics data as any unwanted parameter can drastically change the final result/models. Variability in NMR peak shifts causes problems for statistical modeling as the variables in the data table define the same property for all samples and if there is shifting result will not be entirely correct. So to avoid this problem peaks should be aligned properly (using any of the available peak alignment methods). Bucketing methods can also be used for such purpose.

Other parameters that can alter the results in the multivariate method, especially projection based methods, are the scaling of variables that changes the length of each axis in the K-dimensional space[1,2]. The main objective of scaling is to reduce the noise in the data and thereby enhance the information content and quality. Pareto scaling (divides each variable by the square root of its standard deviation), is recommended for metabolomics data. Principle component analysis (PCA) and partial least square (PLS)[3] based methods are very useful in getting a good overview of the multivariate profile. The scatter plots of first two score vectors give the information about the homogeneity of data as well as information related to all the group's presents, no of outliers present and any trend if present in the data[1,4]. NMR-based metabolomics studies have following steps involved[5] as shown in **Figure 3.1**:

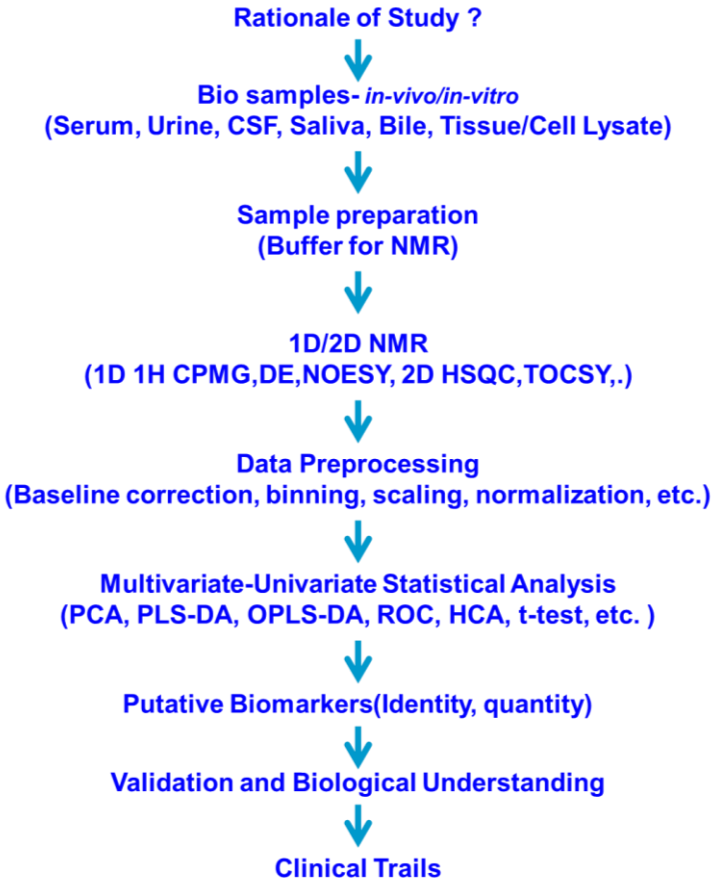



**Figure 3.1:** Schematic representation illustrating the NMR-based metabolomics workflow.

## 3.2: Sample, -Collection, Extraction, and Preparation:

Metabolomics studies are performed on biological fluids. Samples includes urine, serum, plasma, whole blood, cerebrospinal fluids (CSF), bile[6] (liver or gallbladder), pericardial fluid (PCF), ascetic fluid, seminal fluid[7], amniotic fluid[8], synovial fluid, saliva[9], lung aspirates, digestive fluid, tissue samples (extracts & intact), cell lysate, bacterial cultures, plant (intact or extract). However, urine, blood plasma or serum are the most commonly used biofluids in metabolomics

studies, because they contain hundreds to thousands of metabolites and they both can be acquired in relatively non-invasive manner[10-12].

To achieve a "real-time" assessment of homeostasis, all the chemical reactions must be stopped as soon as possible after collection of the sample to prevent unwanted degradation or alteration of metabolite levels. The most common means of achieving metabolic arrest is to immediately flash freeze samples in liquid nitrogen or a Dry-Ice bath or to immediately extract/process the samples so as to remove enzymes and other proteins from catalyzing any further metabolic reactions[13]. Depending on the choice of analytical method to be used, extraction can be performed in a variety of ways; however, a key requirement is that whichever extraction method is used, it should be capable of solubilizing all metabolites while excluding as much of the proteins and other higher molecular mass components as possible. Recovery standards are generally added before any extraction process begins. Once extracted, internal standards may also be added, if any chemical derivatization protocols may be carried out to prepare the sample for analysis. The biological samples or biofluids used in metabolomics studies should be collected under strict conditions and must always be stored at -80 °C, If analysis cannot be performed immediately after collection of samples. Detailed procedures to collect, store and measure biofluids such as blood or serum, refer[10,14].

**Urine** is another common biofluid of choice for metabolomics studies chiefly due to the fact that sample collection is least invasive compared to blood. This creates advantages especially in toxicology or drug follow-up studies where several samples (e.g. pre-dose, after dose) are needed. Hence effects prior and post dosing can be monitored effectively. Urine samples can also be pooled, thus averaging variability (exercise, diurnal variation, etc.), which is important since variability is a problem in some sample cases[15]. Urine contains more metabolites than blood or any other biofluid, which in turn creates more overlapping signals in an NMR spectrum, and while urine is one of the simplest fluids in physicochemical terms, the need to maintain homeostasis results in it being one of the most complex in composition[16]. For urine samples, the addition of sodium azide is required to control bacterial growth.

**Blood Samples** commonly blood is collected by venipuncture into standard vials containing either ethylene diamine tetra acetate (EDTA) or lithium heparin as anti-coagulant to give plasma (e.g., BD Vacutainer, Li-heparin), for serum leave it on ice to coagulate or at room temperature (37$^0$C) (e.g., BD Vacutainer, no additive) and, centrifuge samples at 1,600*g* for 15 min at 4°C[17]. Preferably blood samples should be dealt as cold as possible to preserve the metabolic snapshot in the most optimum way. As when processing the blood sample for collection, the enzymatic metabolite degradation may be pronounced. Avoid EDTA, citrate, and other added stabilizers since they give additional signals in the NMR spectra. This is due to the formation of complexes between EDTA and ions $Ca^{2+}$ and $Mg^{2+}$ which are present in plasma[18]. Also, collection tubes should be avoided, which use gel to separate blood cells from plasma. The time before

separation of blood cells should ideally not exceed for more than 30 minutes. Store the supernatant in a separate microcentrifuge tube with proper cataloging at −80 °C until analyzed.

Proteins can be removed before NMR measurements by either organic solvent precipitation or by ultra-filtration using a 3kDa molecular weight cut-off filter or as per the user's requirements. However, a recent comparison of organic solvent precipitation with ultrafiltration has demonstrated that ultra-filtration was superior for metabolic NMR measurement. Before usage, the filter has to be cleaned from glycerol, which is present in many commercially available filters, by centrifuging it with water or phosphate buffer. The effect of protein removal on the $^1$H-NMR spectrum of blood serum is shown in **Figure 3.2**.

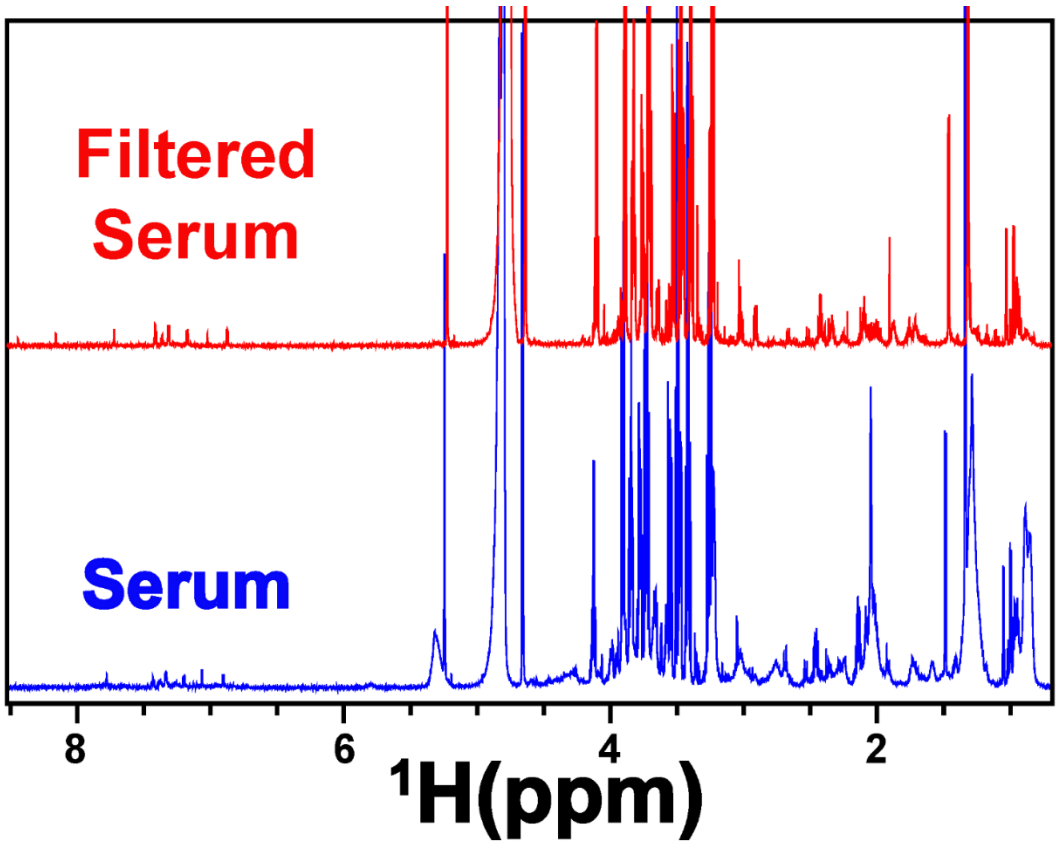



**Figure 3.2:** 1D $^1$H NMR spectra of ultra-filtered serum (3 kDa Cutoff filter) and normal serum.

The pH of samples has a substantial influence on the chemical shifts observed in the NMR spectrum. Therefore, it is important to control the pH of the biofluid sample. A very popular and fast method to adjust pH is the addition of a phosphate buffer stock solution made up in $D_2O$ at pH 7.0 or 7.4. Plasma and serum can be measured directly with minimal sample preparation, however buffering will provide homogeneity throughout the samples. In $^1$H NMR analysis of urine samples, special concerns have to be focused on minimizing the chemical shift due to the difference in pH between the samples, and a buffer should be applied to the sample. Typically, a phosphate buffer in $D_2O$ and with TSP (internal/external) as a reference compound is used.

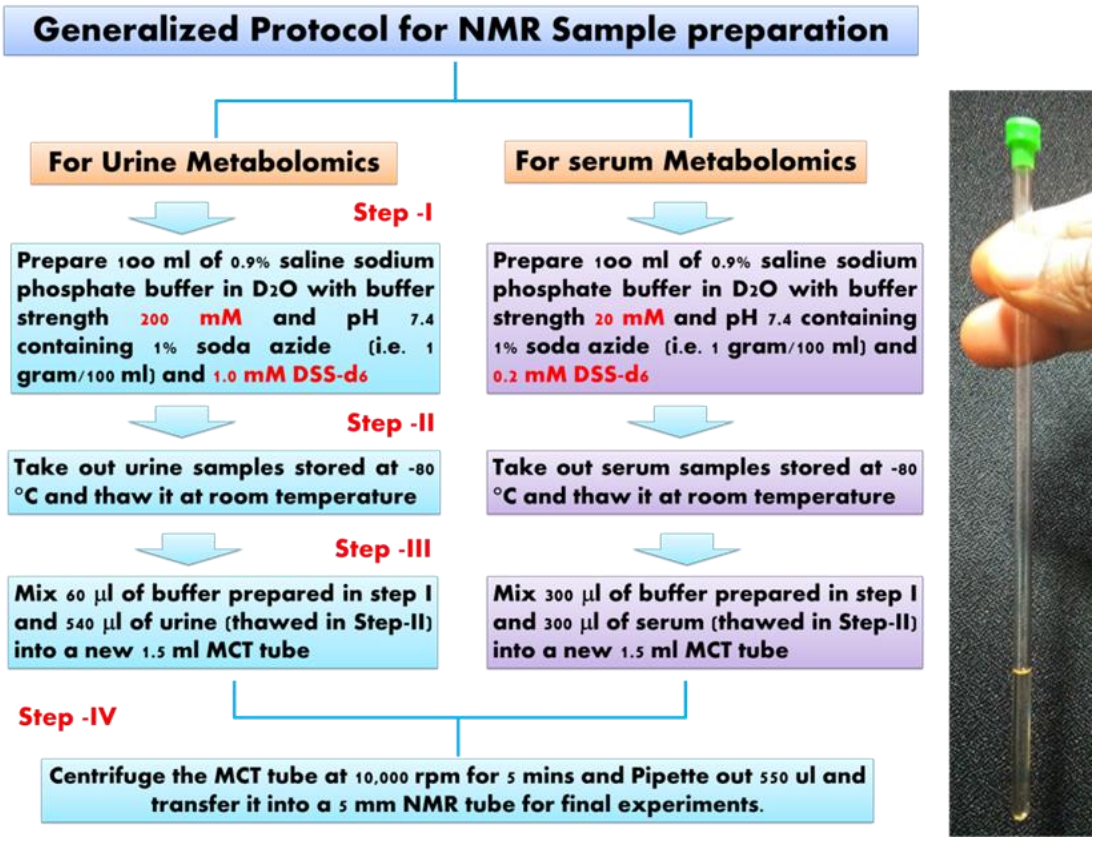



**Figure 3.3:** Flowchart showing the generalized protocol for preparing the NMR samples for serum and urine based metabolomics studies. The protocol assumes that in each case, more than 300 µl of serum and more than 540 µl of urine is obtained.

Usually, 5-mm diameter NMR tubes are used for NMR measurements. There are also much smaller NMR tubes with a diameter of 1-3 mm or a micro scale SHIGEMI tube in which a reduced sample size is compensated by solid glass beneath the level of the sample.

## 3.3: NMR Experiments and Data Acquisition:

A number of one and two-dimensional NMR methods are currently used for metabolomics applications[19]. However, the detection limit for biofluid NMR spectroscopy depends on many factors such as field strength, the number of protons contributing to a resonance and a region of the spectrum where the resonance is observed. In general, the detection limit is in the low micromolar range in the less crowded regions of the spectrum. There are two experimental issues in NMR of biofluids. The first one is connected to accurate solvent suppression. The second experimental issue is associated with the distinction between small molecular weight metabolites (typically <1500Da) and macromolecules. Currently, to overcome these problems, the simple one pulse sequence and one-dimensional nuclear Overhauser enhancement spectroscopy (NOESY) sequence with water suppression are the most commonly used NMR methods for metabolomics applications. Water suppression in the one-pulse experiment depends more critically on good shimming. On the contrary, 1D NOESY is more robust and provides a flatter baseline under

similar conditions[20]. Various pulse sequences are available all of which are designed to effectively suppress the high-intensity water signal leaving the metabolites signals intact[21,22]. A commonly used method for suppressing the broad signals from large molecules (such as in tissue or serum samples) is the Carr–Purcell–Mieboom–Gill (CPMG) sequence[23]. CPMG as special pulse sequence is used to remove broad proteins signals if they have not been taken away before NMR measurement. This sequence is robust and is widely utilized in some studies to date. In contrast, the so-called "diffusion edited" NMR experiment may be employed for observing signals from large molecules such as lipids[24]. The 1D selective TOCSY experiment has been successfully implemented in metabolomics studies to detect metabolites quantitatively even if they are found at concentrations 10–100 times below those of the major components[25]. This approach has been shown to be highly useful for detecting targeted metabolites in biological samples[26].

2D NMR methods are extremely helpful in reducing the spectral complexity and obtaining connectivity between the nuclei to make assignments and identification of metabolites. However, these methods have not been widely used in metabolomics to date because of their increased acquisition time, data size, and complexity in data analysis. Nevertheless, a small but growing number of papers report using 2D approaches in metabolomics studies[27-29]. The most commonly used include 2D–J-spectroscopy, correlation spectroscopy (COSY), total correlation spectroscopy (TOCSY), heteronuclear single quantum coherence (HSQC) spectroscopy and heteronuclear multiple bond correlation (HMBC) experiments. 2D J-resolved (JRES) spectroscopy is attractive for metabolomics studies[30] because it can lead to a substantial simplification of the spectra. One drawback of this method is that the integral of signals is strongly influenced by $T_2$ relaxation during generally the long $T_1$ evolution period and hence only a relative quantification of the concentration of metabolites is possible. The analysis of complex mixtures such as urine, serum or other bio-fluids has improved with fast 2D ($^1$H-$^{13}$C) heteronuclear experiments to yield spectra with good signal-to-noise ratios. The method has been applied to identify patients with inborn errors of metabolism, cancer, and cardiovascular disease, etc..

## 3.4 Post-acquisition processing of NMR spectra:

Data processing is performed on NMR spectra after the Fourier transform of the FID. Depending on the precision and reliability of the desired parameters (i.e. chemical shift, coupling constant) that are extracted from the spectra, it allows improving the quality of the spectrum. This section aims at providing a summary of the most important NMR data processing tools:

***Phase correction*** **-** Phase is a property of the NMR Peak in the Spectrum. In NMR spectroscopy we seek our peaks to be positive, narrow at the base and as symmetric as possible. Phase correction is the process of mixing the real and imaginary signals in the complex spectrum that is obtained

after Fourier transformation of the free induction decay. The phase of the spectrum is typically corrected such that a properly phased spectrum is similar to an absorption spectrum. Phasing the spectrum is a routine procedure and involves zero-order and first-order phase corrections. Zero-order correction is frequency independent and is the same for all lines across the spectrum. Whereas the first-order correction is frequency dependent and applies a phase change, whose amount increases linearly with the distance to the reference signal.

***Baseline correction*** **-** Baseline distortion can be a problem in NMR, as of spectra without a flat baseline gives wrong integrals, while in spectra with baseline roll small signals may not be recognized. Baseline distortions affect not only the statistical analysis but also the quantification of the metabolites. Baseline distortion is mainly caused by the corruption of the first few data points in FID. These corrupted data points add low-frequency modulations in the Fourier-transformed spectrum and thus form the distorted baseline. These distortions can be corrected in many different ways. Usually, an automated baseline correction is applied. Currently, the most popular methods are based on fitting, baseline correction models the baseline by fitting it to polynomial, sine or exponential functions[31,32], robust estimation procedure implemented in Chenomx NMR Suite[33], locally weighted scatterplot smoothing (Lowess) fit[34]. Asymmetric Least Squares[35] is a method that uses a different constraint for baseline correction. It tries to estimate the baseline by fitting a regression curve to a spectrum using a penalized least square approach. The baseline can also be corrected by using B-splines, B-splines with Penalization (i.e. P-splines)[36] or by applying mixture models[37]. The horizontal baseline is a crucial point to achieve, for accurate spectral signal quantification.

***Zero-filling*** **-** In NMR spectroscopy the need for shorter acquisition times lead to the loss of digital resolution. Zero-filling is the procedure of increasing the digital resolution of the NMR spectra by increasing the FID data points. It is applied just before performing Fourier transformation, by adding zeros to the end of the FID. Since the zeros contain no new information, there is no new information added to the spectrum. It is a method of artificially increasing the acquisition time after the data collection without adding any new information or true resolution to the spectrum. It improves the line shapes and the appearance of the spectrum by resolving very small couplings in multiple structures.

***Apodization*** **-** is a digital filtering manipulation of the NMR spectra which consists in multiplying the FID with different functions (such as Lorentzian or simple exponential decay, Gaussian, or Sine-bell function) which can be chosen either to enhance the sensitivity or resolution in the final resulting spectrum. For instance, exponential multiplication leads to line broadening and a reduction in noise, while other trigonometric functions produce a narrowing of the spectral line while an increase in the noise.

Zero filling and apodization are two data processing techniques that are used to enhance the resolution and the signal to noise ratio (S/N) of the NMR spectrum. The resolution of the spectrum refers to how clearly defined two adjacent peaks are, this is typically expressed in Hz. A smaller number means that the peaks are narrower and that the resolution is higher. The S/N is expressed as a ratio of the peak height over the standard deviation of the background signal.

## 3.5: Spectral Assignment/Metabolite Identification:

In NMR-based metabolomics studies, metabolite identification also known as the spectral assignment is very crucial as all the further analysis will depend on, upon this step. 1D $^{1}H$ NMR spectra are routinely acquired due to the high NMR sensitivity of the hydrogen nucleus. However, the 1D spectra of samples containing a complex mixture of metabolites can be congested due to the resonance appeared at the same chemical shift, making the resonance assignment a challenge. 2D homonuclear/heteronuclear correlation and J-resolved (JRES), HSQC, experiments can be applied to alleviate the congestion of 1D spectrum. Spectral resonance assignment is usually performed by searching chemical shift lists, in metabolic databases (HMDB, BMRB, etc.), and software's like Chenomx[33], MetaboID, and published literature along with the analysis of specific J-coupling patterns of the metabolite spin systems. Usually, confident resonance assignment is achieved by combining the assignment of 1D $^{1}H$ spectra along with the knowledge obtained from the 2D experiments of selected samples. The basic methods for identification of metabolites in biofluids are described in detail in a subsequent section.

***Assignment based on literature:***

All biofluids contain thousands of metabolites, and almost all biofluids have been once studied by using NMR-based metabolomics techniques. These reports give information related to the identification of these metabolites in the form of patterns[38,39]. There are plethora of reports related to metabolic identification and spectral assignment in biofluids[40] or pathological samples like urine[41], bile, blood plasma/serum[42], cerebrospinal fluids[43] (CSF), pericardial fluid, amniotic fluid[44], seminal fluids[45], saliva[46], follicular fluids, various abscesses like liver abscess, brain abscess, then normal and cancerous tissues like brain, breast, kidney, liver, gallbladder, pancreas, ovary, head and neck, prostate etc.. So, on the basis of the published reports metabolites present in the different biofluids can be identified on the basis of chemical shift information, coupling constant and their respective coupling patterns.

***Biological NMR Databases:***

Apart from previously reported literature, there are several databases that provide very valuable information related to metabolite identification and spectral assignment of several

biofluids like urine, serum, a tissue/cell extracts, plant extracts, etc.. These databases contain spectral information as well as raw data for thousands of metabolites including 1D and 2D NMR experimental data. These data can be downloaded and use to compare with the biofluids spectra. Link to some of the extensive databases is given in **Table 3.1**.

| Data Bank | Full Form | Web Link |
|---|---|---|
| **BMRB** [47] | Biological Magnetic Resonance Bank | http://www.bmrb.wisc.edu/metabolomics/ |
| **HMDB** [48] | The Human Metabolome Database | http://www.hmdb.ca/ |
| **MMCD** [49] | Madison Metabolomics Consortium Database | http://mmcd.nmrfam.wisc.edu/) |
| **BML-NMR** [50] | Birmingham Metabolite Library | http://www.bml-nmr.org/ |
| **LIPIDBANK** [51] | The LipidBank | http://lipidbank.jp/ |
| **MDL** [52] | NMR Metabolomics Database of Linkoping | http://www.liu.se/hu/mdl/main/ |
| **PRIME** | Platform for RIKEN Metabolomics | http://prime.psc.riken.jp/ |

**Table 3.1**: A list of Biological NMR databases.

***Spiking experiments:***

This is one of the most precise methods for metabolite identification, and even it can be considered as a confirmatory method for the assignment. In this method, first of all, the spectra of a sample are recorded, and then one other spectrum is recorded for the same sample with the addition of suspected compound. If the signal intensity of all peaks from the compound increased, it proves the presence of that compound in the sample. Spiking experiments are very useful in metabolomics studies as sometimes the metabolite concentration is so low that it cannot be identified even using 2D methods. In those cases after spiking with suspected compound followed by recording 1D and 2D spectra, these metabolites can be assigned unambiguously.

***Using standard sample spectra:***

When NMR spectra of biofluids are compared with the NMR spectra of the standard compound, they can easily be assigned. For this comparison, NMR spectra of standard compounds can be recorded using physiological parameters similar to biofluids like similar solvent system (here it should be water as all biofluids are aqueous in nature), pH and concentration. Then these spectra will serve as control spectra for comparison. Some databases provide similar kind of information for a lot of metabolites in the form of raw NMR data ($^{1}H$, $^{13}C$, DEPT, COSY, TOCSY, HSQC, and HMBC). These data can be downloaded from the sites and then processed and are used for the comparison[53].

## 3.6: Quantitative Analysis of Metabolites:

After the metabolites are identified and assigned unambiguously, quantitative analysis is the next step in metabolomics studies. History of quantitative analysis in NMR is as old as NMR itself and has been used in many studies. Similarly, it is an integral part of metabolomics studies as stated in the definition of metabolomics itself. So it becomes crucial to quantify the metabolites of interest to draw any conclusive results. A real essential advantage of NMR is absolute quantification[54]. The linear response of NMR experiment is the key benefit of NMR over other analytical methods. The signal intensities observed in NMR spectrum are directly proportional to the concentration (i.e. molar amount) of that nucleus in the sample[55]. In order to obtain an absolute concentration of metabolites detected by NMR usually, the addition of internal standards is used. The reference compound used for concentration reference as well as for chemical shift ($\delta$=0.00) is usually the sodium salt of 3-trimethylsilylpropionic acid-d4 (TSP-d4) with deuterated methylene groups. Other references standards are 2,2-dimethyl-2-silapentane-5-sulfonate sodium salt (DSS) or for organic solvent tetramethylsilane (TMS). Another internal reference standard, 4,4-dimethyl-4-silapentane-1-ammonium trifluoroacetate has been also proposed as promising universal for metabolic profiling[56]. Another approach includes a synthetic electronic reference signal, Electronic Reference To Access *Invivo* Concentrations (ERETIC), for quantification purpose[57]. A similar approach, QUANTification by Artificial Signal, has been recently introduced[58]. In these methods, an internal standard is claimed to be not required. The standard way of quantifying compounds in NMR spectrum is integrating the different NMR resonances and comparing them to the area of the internal standard. Another possibility is spectral deconvolution, available for instance in MestReNova[59], where global spectral deconvolution is applied automatically to whole NMR spectrum. In the Chenomx NMR Suite software the spectral signatures (singlets, doublets, triplets, etc.), *i*.e. the peak shapes, of a compound from an internal database of reference spectra is fitted to the experimental NMR spectrum[60]. In contrast, peak integration is very sensitive to baseline distortions. Moreover even slightly overlapping resonances cannot be reliably quantified. Peak-shape fitting, like in Chenomx, is not affected by baseline distortions and still efficient when some of the resonances and part of a resonance overlap with that of another compound, the peak shape can still be fitted with reasonable accuracy and the concentration of the compound reliably determined.

The number of observed metabolites in biofluids largely depends on the magnetic field strength of NMR spectrometer. Therefore, working at the highest available magnetic field is recommended.

**Theory:**

The basic concept behind the quantitation using NMR is that the signal intensity in NMR spectrum is directly proportional to the no of nuclei giving that signal and can be given as:

$$I_x \infty N_x \text{ or } I_x = K_{spec}\, N_x$$

Here k spec is the spectrometer constant which will be same for all the resonances in the spectrum. Various methods have been developed for the quantitative analysis of compounds/metabolites in NMR spectra and are discussed in subsequent section.

**Methods:**

**Relative quantitation:**

It is one of the easiest methods for obtaining quantitative information of metabolites. Molar ratios of two metabolites can be computed as:

$$\frac{n_A}{n_B} = \frac{I_A N_B}{I_B N_A}$$

moreover, from these, the fractions of metabolites can be calculated as:

$$\frac{n_A}{\sum_{i=1}^{m} n_i} = \frac{I_A N_A}{\sum_{i=1}^{m} I_i/N_i} 100\%$$

Here 's' represents the no of compounds/metabolites in the mixture.

**Absolute quantitation:**

For absolute quantitation of metabolites, there are two methods. In one of the methods, when impurities or all components present can be assigned correctly and measured quantitatively, then their amount/ quantity will be just a difference of 100%. This method fails when there are overlaps in the spectral region, and some impurities/components do not have $^1H$ nuclei. In another method of absolute quantitation, an internal reference compound with known concentration is added to the sample, and then the concentration of component can be computed using the following expression.

$$C_x = \frac{I_x}{I_{ref}} . \frac{N_{ref}}{N_x} . \frac{M_x}{M_{ref}} . \frac{W_{ref}}{W} . C_{ref} .$$

Where C, I, N, M, and W represents the concentration, spectral intensity, no of nuclei, molar masses, and weight respectively for component (***x***) and reference.

**Standard addition method:**

This method is another very important analytical method for quantitative analysis of metabolites in NMR spectra. In this method, the standard compound is added (generally two additions) to the

sample and then the solution is diluted prior to spectrum acquisition. Spectra are recorded for original sample and for each addition. A calibration curve is plotted between the integral intensity and concentration of the standard compound. By extrapolating the plot to zero addition, the concentration of unknown compound can be calculated.

**Calibration curve method:**

In this method, a calibration curve is prepared using the standard solution (with known concentration) of the metabolite that is present in the sample. Standard solutions are prepared in varied concentration that will cover the whole range of concentration in the sample. Then NMR spectra are recorded for sample and all standard solutions using similar parameters. From the concentration of standard solutions and integral intensities, a calibration curve is prepared for that metabolite. For determination of concentration, usually regression method is used, and regression equation is derived from which unknown concentration can be determined.

## 3.7: Pre-processing of NMR data for Chemometric analysis:

In NMR-based metabolomics, NMR spectra must be transformed into a suitable format prior to multivariate data analysis. The most common pre-processing techniques used for this purpose are alignment, exclusion, binning, normalization, transformation, and scaling.

***Alignment:***

One of the most significant challenges in the comparative analysis of NMR metabolic profiles is the occurrence of shifts between peaks across different spectra. It originates when, for different samples (spectra), the same molecule displays peaks at different chemical shift positions (on the horizontal axis). Shifts can be due to instrumental factors, changes in the pH and temperature, changes in salt concentration, overall dilution and relative concentration of specific ions. All these parameters influence peaks shifts, although not all peaks are affected to the same extent. Some of these effects can be (partially) avoided by using adjusted sample preparation protocols, for example by buffering samples to avoid pH-induced chemical shifts. Proper alignment of spectral peaks is, therefore, a crucial pre-processing step prior to downstream quantitative analysis.

Therefore an essential step in pre-processing is to adjust the peaks shifts, i.e. alignment or warping. NMR spectra are usually first aligned by spectral referencing. This simple, global method for peaks alignment sets the internal reference signal (e.g. TMS, DSS, TSP, etc.) of each spectrum to 0 ppm. These can be corrected post-processing by various automatic peak alignment procedures as depicted in **Table 3.2**. In addition, correct spectral alignment/shift calibration is required for reliable metabolite identification[61].

| Short Name | Full Name | Technique | Target Function | Original Applied Data |
|---|---|---|---|---|
| PLF [62] | Partial Linear Fit | Segmentation model by consecutive peaks distances less than window size D | Sum of squared differences in intensity | 1D $^1$H NMR |
| COW [63] | Correlation Optimized Warping | Dynamic programming | Pearson correlation coefficient | 1D $^1$H NMR |
| PAGA [64] | Peak alignment by genetic algorithm | Genetic Algorithm | Pearson correlation coefficient | 1D $^1$H NMR |
| PARS [65] | Peak alignment using Reduced Set | Breadth-first search (BFS), Dynamic Programming (DP), complexity reduced dynamic programming (crDP) | Euclidean distances | 1D $^1$H NMR |
| PABS [66] | Peak alignment by Beam search | Beam search algorithm | Pearson correlation coefficient | 1D $^1$H NMR |
| PAPCA [67] | Peak alignment by PCA | Principle Component Analysis | CORREL | 1D $^1$H NMR |
| FW [68] | Fuzzy Warping | Fuzzy logic for matching most intense peaks | Maximize fuzzy membership Gaussian function | 1D $^1$H NMR |
| GFHT [69] | Generalized Fuzzy Hought Transform | Hough transform | Hough score | 1D $^1$H NMR |
| RSPA [70] | Recursive segment-wise peak alignment | Recursive segmentation model | FFT cross-correlation | 1D $^1$H NMR |
| PCANS [71] | Progressive Consensus Alignment of NMR Spectra | Segmentation model+Dynamic programming + progressive consensus alignment | Scoring by similarity between peaks calculated by height, half height and position of peaks | 1D $^1$H NMR |
| BAA [72] | Bayesian approaches for alignment | Bayesian modeling | Bayesian estimation | 1D $^1$H NMR |
| Icoshift [73] | interval correlation shifting | Segmentation model by equal size segments or manually selecting segments | FFT cross-correlation | 1D $^1$H NMR |
| CluPA [74] | hierarchial Cluster-based Peak Alignment | Segmentation model by hierarchical clustering | FFT cross-correlation | 1D $^1$H NMR |

**Table 3.2:** A List of different alignment methods and their features. Adapted from[75,76].

***Exclusion (Eliminating Uninformative Peaks):***

After phase correction, baseline correction, and resonance alignment, next crucial pre-processing step is the exclusion of unwanted spectral regions i.e. regions devoid of any metabolite[77-79]. In routine practice, the spectral width of NMR spectrum is wider enough for the chemical shifts of all

the metabolites present in the sample. Broad spectral width has its own advantage in pre-processing steps like phase correction and baseline correction, but in data analysis steps spectral regions devoid of metabolites (endogenous metabolites) are not useful. These regions are also more sensitive and error prone to several types of spectral artifacts, so it is better to exclude these regions. So in normal practice, spectral region outside 0.2 ppm to 10 ppm in 1H (proton) NMR spectrum is excluded[79,80].

Apart from regions (where there are no metabolite resonances), there are some other parts of the spectrum that should also be excluded like broad resonances from the solvent water. In most of the water pre-saturation experiments, water is suppressed considerably, yet some water signal remains in the spectrum that dominates the 4.6 ppm – 5.0 ppm which can make the analysis of metabolite signal that comes in this region almost impossible. The signal of urea, which is very close to water signal, is very intense in urine spectrum and appears as a very broad signal ranging from 5.4 ppm to 6.0 ppm is also excluded for chemometric analysis[80-83]. Exclusion of urea is commendable due to the fact that it cannot be used for quantitative analysis and chemometric analysis as it has the exchangeable proton that exchange protons with water hence its signal intensity varies with the quality of water suppression and can affect the model output. In the case of other biofluids used in metabolomics such as plasma/serum spectral region between 0.20 ppm to 4.60 ppm and 5.05 ppm to 10 ppm is used for multivariate analysis. Similar is the case of other biofluids, where mostly water resonance (4.6 ppm – 5.0 ppm) is excluded, and remaining spectral part is used for analysis[78].

***Binning:***

Due to several physiochemical properties of the sample which includes pH change, change in salt concentration, dilution in the sample, relative concentration, there is peak shifting for some metabolites for different extents[77]. This peak shifting can hamper automated data analysis and peak assignments and make it difficult for interpretation of classification and separation of classes. Peak shifting causes distortion in peak shapes of loading plots of different decomposition and discrimination methods (PCA, PLS, and OPLS)[79] which definitely need thorough inspection and manual interpretation. So it becomes very important to reduce the peak shifting. Peak shifting reduction can be done either by taking care of some sample preparation steps or by mathematical approaches. Sample preparation steps that are important to minimize shifting include the addition of buffers to the biofluids[14,38,40,77]. “Binning or bucketing" involves summing peaks over spectral segments and is fast and easy to implement. Commonly, bins of 0.04 ppm are used for urine samples, which seems to correct for small shifts that may be present for peaks[84,85]. Bins of different sizes can be used. However, some studies use bins as small as 0.005 ppm for tissue and plasma/serum samples[86]. Summing over groups of peaks simplifies the data, but information is lost in the process. Also, inaccuracies can arise since unrelated peaks can be placed in the same bin and peaks can be split between different bins[87]. Some try to get around

this problem by combining bins known to contain the same or related substance[86,88]. This method is used in classification studies, but since peak locations are combined, biomarker identification is not possible using such a method[84]. With modern computing power and mathematical approach, with other more sensitive methods such as spectral alignment are being used for binning. Mathematical approaches can further be divided into two groups, one of these is the reduction of spectral resolution by binning or bucketing procedure, and another one is the alignment of resonances by use of statistical algorithms. The Binning approach is much widely used approach and is of four types as described.

**Equidistance binning:**

In this method, whole spectral region is integrated into the smaller spectral region known as bins or buckets. Equidistance bins of 0.04 ppm normally used in metabolomics, but bins with smaller regions can also be obtained. In this approach, the whole spectrum is divided into an equally spaced spectral region of 0.04 ppm or less. This method has non-flexible boundaries, so if a peak falls in between two bins, this shifting will significantly affect the data analysis. When this type of data is subjected to PCA or PLS analysis, shifted peaks will result in dispersive peaks in loading plots[89].

**Point-wise bucketing:**

Point-wise bucketing is just a special case of rectangular bucketing. This time, each data point is taken as a bucket. Sweep width and calibration of the involved spectra are taken into account. Point-wise bucketing would preferably be applied for broad line spectra such as in vivo spectra.

**Non-equidistance binning/variable sized bucketing:**

To overcome the effects or drawbacks of Equidistance binning, an advance binning approach has been developed which is termed as non-equidistance or variable sized binning. This method prevents the peaks being cut by the non-flexible boundaries of bins. Here boundaries of bins are adjusted in such a way that one complete peak will be inside the boundaries of that bin. So bin width will depend upon the area/width of the peak shape and on the shift width of that peak. In this approach all peaks remain intact; there is no cutting of peaks by the boundaries of bins. Along with these advantages, there are some disadvantages too in this approach. With variable bucketing/ binning bin width varies significantly and data points from different bins will affect data analysis in different ways. The data point in a very small bin will have higher influence than the data point of the bigger bin.

**Advanced Bucketing:**

It is a newer technique for bucketing which is internally based on picked peaks. One of the most important advantages of this technique is that the peaks existing in only one spectrum are stored

in columns which have only zero entries otherwise. Their significance is raised accordingly. The advanced bucketing only creates columns if needed and reduces the table sizes.

### *Normalization*

Normalization is also one of the crucial pre-processing methods for metabolomics studies. Normalization methods aim to remove unwanted sample-to-sample variation[90]. This includes correction for the overall concentrations of samples, which influences metabolite dilution[90,91]. Samples can display greatly varying concentrations of metabolites from subject to subject. A large part of these subject-to-subject variations is similar across the spectrum for each subject. Thus, all metabolites can be scaled based on some common measure, in order to obtain comparable samples, regardless of variations in concentration in the raw spectra[92]. For removing/reducing unwanted biases (so that the targeted metabolite signals can be interpreted unambiguously), normalization methods can be categorized into two groups: (a) Methods for removal of sample-sample variation, and (b) method that works on adjusting the variance of different metabolites. In common practices, normalization tends to scale the whole spectra in such a way that after normalization all the spectra have similar overall concentration. There are several methods[93] in both the categories that are used in normalization and are described briefly in **Table 3.3**.

| Method | Features |
|---|---|
| Total of sum normalization or Integral normalization[91] or Constant sum normalisation[94] | The integrals of spectra are a function of metabolite concentration |
| Probabilistic quotient normalization (PQN)[95] | Integral normalization of each spectrum, followed by the calculation of a reference spectrum such as a median spectrum. Finally, all variables of the test spectrum are divided by the median quotient. |
| Creatinine normalisation[96] | Normalization using the area under the creatinine peak as a reference |
| Group aggregating normalisation[97] | The samples are normalized so that they aggregate close to their group centers. |
| Histogram matching normalisation[98] | A time series, image, or higher dimension scalar data is modified such that its histogram matches that of another (reference) dataset. |

**Table 3.3:** A List of different normalization methods generally used and their features.

### *Transformation*

Transformation is another data pre-processing method that causes the nonlinear conversion in metabolomics data[99,100]. Transformation of data converts it to normalized form having a constant variance, hence making it more suitable for multivariate analysis. One of the principal purposes

of transformation is that it corrects the heteroscedasticity of the metabolomics data so that the multiplicative relations are converted into additive relations. This also causes the skewed distribution to be much more symmetric, hence improving the model. Transformations are known to have pseudo-scaling effects as they reduce the larger values of the dataset into a comparatively smaller data set, thus reducing the difference between large and small values. Different transformation methods are available in all chemometric software, of which log, generalized log (glog) and power transformation are very important for metabolomics studies as given.

**Log transformation:**

Log transformation is the most commonly used method of transformation which completely removes the heteroscedasticity in the data given that standard deviation is a constant value. Following are the expressions for log transformation of data:

$$\tilde{x}_{ij} = \log_{10}(x_{ij})$$

The major drawback of the log transformation is that it fails when there are zero values in the data set. For metabolites with low concentration, the standard deviation is relatively large which can be problematic when using log transformation[101].

**Generalized logarithm transformation or glog:**

When metabolomics data have very different variance values, it becomes very difficult to obtain good multivariate analysis results and good prediction models. In that case, generalized form of log transformation, also known as 'glog' transformation can be used which makes the variance somewhat constant so that analysis can be done easily[102]. So this method stabilizes the variance by use of the following expression:

$$\tilde{x}_{ij}(\lambda) = \ln\left(x_{ij}^2 + \sqrt{(x_{ij}^2 + \lambda)}\right)$$

Here $\tilde{x}_{ij}$ is the transformed variable/ intensity, $x_{ij}$ is the original variable/intensity and $\lambda$ is the transformation parameter. Transformation parameter $\lambda$ can be obtained by different methods such as maximum likelihood method using various replicates of samples. This transformation method gives very prominent results for biomarker identification as well as classification.

The modification of glog is the extended glog[103] which effectively suppress the noise since improving the discrimination between sample classes can help to identify metabolic biomarkers. The extended glog transformation is given by the equation

$$\tilde{x}_{ij}(\lambda) = \ln\left((x_{ij}\text{-}x_0) + \sqrt{(x_{ij}\text{-}x_0)^2 + \lambda)}\right)$$

Where $x_0$ shifts the glog function to suppress noise and $x_0$ is dependent on the choice of λ.

**Power transformation:**

Power transformation method is slightly better method than log transformation method as it overcomes the above-mentioned problems of that. It also has a positive effect on heteroscedasticity. Its transformation pattern is similar to log transformation as shown by given expressions:

$$\tilde{x}_{ij} = \sqrt{(x_{ij})}$$

This method has all the properties of log method except that it cannot convert multiplicative relations into additive relations which are very important part in metabolomics studies[100,104].

***Scaling:***

The intensity of metabolites can range over orders of magnitude. Furthermore, metabolites with high intensity will often have high variation and thus the greatest effect on the analysis. Scaling is done to prevent selection of the metabolites with the largest intensity as significant[91]. Scaling methods are aimed at adjusting the variance of the different metabolites. These include variable scaling and variance stabilization approaches[90]. Mean-centring is technically not a scaling method but is described in this section, since it is a pre-processing step applied per variable across samples. The different scaling methods according to the category, before discussing each method in more detail.

1) Variable scaling methods divide each variable by a scaling factor determined individually for each variable. Variable scaling can be divided into two subclasses[104], namely methods that use

A measure of data dispersion as a scaling factor

- **Auto-scaling** is also called unit variance (uv) scaling, and the standard deviation of the data is used as a scaling factor. In short, the data are first mean-centred across spectra (i.e. by feature), then divided by the standard deviation of each feature. After Auto scaling, all features in the dataset are considered equally important, but the effect of noise will be increased[90]. By Auto scaling, unit variance is attained. Therefore data are then analyzed based on the correlations instead of covariances[91]. Between-sample variation, caused by different dilution of samples, is not removed by Auto scaling. In urine samples, dilution can be due to differences in fluid intake.

$$\tilde{x}_{ij} = \frac{x_{ij} - \bar{x}_j}{S_j}$$

- **Pareto scaling:** Pareto scaling is similar to auto scaling, but uses the square root of the standard deviation as a scaling factor, instead of the standard deviation[105]. The scaling effect of Pareto scaling is not as strong as for Auto scaling, i.e. after Pareto scaling the data remain closer in value to the original data. Pareto scaling is less likely to inflate noisy background data and to diminish the importance of large fold changes compared to small ones. Huge fold changes may, nevertheless, still display a dominating effect.

Pareto scaling is popular in biomarker identification. Specifically, in metabolomics, the aim is to identify metabolites that behave differently in two populations, and the interest is focused on high-intensity variables[106].

$$\tilde{x}_{ij} = \frac{x_{ij} - \bar{x}_j}{\sqrt{S_j}}$$

- **Vast scaling:** Vast scaling is an extension of Auto scaling[107]. Vast is an acronym for variance stability scaling[104]. The method concentrates on metabolites with small variations, i.e. metabolites that are stable.

$$\tilde{x}_{ij} = \frac{(x_{ij} - \bar{x}_j)}{S_j} \times \frac{\bar{x}_j}{S_j}$$

- **Range scaling:** In Range scaling, the range of each metabolite is used as the scaling factor[108]. Range scaling is sensitive to outliers. For spectral data, the natural lower bound is zero and in this way Range scaling only considers the maximum[106].

$$\tilde{x}_{ij} = \frac{x_{ij} - \bar{x}_j}{\left(\max(x_j) - \min(x_j)\right)}$$

- **Level scaling:** The mean is used as a scaling factor in Level scaling. Level scaling is relevant when large relative changes are of interest[91].

$$\tilde{x}_{ij} = \frac{x_{ij} - \bar{x}_j}{\bar{x}_j}$$

2) Variance stabilization methods which reduce heteroscedasticity

- **Variance Stabilisation Normalisation:** Variance Stabilisation Normalisation (VSN) is a set of non-linear transformations that aim to keep the variance constant over the entire data range[103,109]. VSN addresses the problem of the non-constant coefficient of variation by using the inverse hyperbolic sine. The data are returned on a generalized logarithm (glog) scale to base-2.

For large values, this transformation approaches the logarithm, thus removing heteroscedasticity. For small values, it approaches the linear transformation, and the variance remains unchanged[90].

**Mean centering (mean-centered only):**

Mean-centring is applied per variable across samples[93]. Technically it is not a scaling method; it is applied as a pre-processing step or as part of statistical analysis. In the centering method, all the fluctuations in metabolite concentration are converted toward zero instead of around the mean, thus regulating for differences between high-intensity and low-intensity chemical compounds (metabolite concentrations). So, all the differences in the offset between high and low abundant metabolites are adjusted. That is why centering method used to focus on the fluctuating part of the data leaving the relevant variation between the samples for analysis. Mean-centring does not remove heteroscedasticity. The method is often used in combination with other scaling methods[91].

$$\tilde{x}_{ij} = x_{ij} - \bar{x}_j$$

Where in all the above equation, $\tilde{x}_{ij}$ represent the intensity value for the $i^{th}$ spectrum at position $j$ on the chemical shift axis, where $i = 1 \ldots N$ and $j = 1 \ldots n$. Then, at each position $i$ on the chemical shift axis, the estimated mean and standard deviation (across spectra) are respectively

$$\bar{x}_j = \frac{1}{N} \sum_{i=1}^{N} x_{ij} \quad \text{and} \quad S_j = \frac{1}{N} \frac{\sum_{i=1}^{N} (x_{ij} - \bar{x}_j)^2}{N-1}$$

## 3.8: Online Computational Tools and Software's for Chemometric Analysis

Some free software packages and online computational tools are available for the processing and analysis of metabolomics data, as given in **Table 3.4**. Several statistical tools listed were designed for NMR data analyses but might also be useful for MS data analyses. Among the web tools available MetaboAnalyst is one of the most widely used online tools for metabolomics data analysis.

| Name | Main Application | Available | Web Link | User Interface |
|---|---|---|---|---|
| **MetaboAnalyst-3** [110] | Univariate, Multivariate, Pathway Analysis, etc. | Free | http://www.metaboanalyst.ca/ | Web Server |
| **Multibase** | Multivariate Analysis | Free | www.numericaldynamics.com/ | Desktop |
| **PAST** [111] | Multivariate Analysis | Free | http://folk.uio.no/ohammer/past/ | Desktop |
| **MUMA** | Univariate and Multivariate Analysis | Free | http://www.r-project.org | Desktop |
| **SIMCA** | Multivariate Analysis | License Required | http://umetrics.com/products/simca | Desktop |
| **Unscrambler** | Multivariate Analysis | License Required | http://www.camo.com/ | Desktop |
| **MVAPACK** | Multivariate Analysis | Free | http://bionmr.unl.edu/mvapack-form.php | Desktop |

**Table 3.4:** A list of commonly used multivariate statistical software's for metabolomics data analysis

## 3.9: Chemometric Analysis: Multivariate, Univariate

Metabolomics data obtained from NMR spectroscopy typically contains thousands of variables from each sample; the variables attained are usually highly correlated. The multidimensionality of this type of data is difficult to comprehend and visualize and invoke for analytical techniques which can extract the relevant information. Chemometric methods are here an obvious choice due to their ability to decompose complex multivariate data into simpler and potentially interpretable structure. Depending on the aim of the analysis, unsupervised or supervised methods may be applied and assist in e.g. obtaining an overview of data, in variable selection, in group classification or to relate the data set to a reference value for construction of prediction models.

### Multivariate Analysis

The key element in the metabolomics workflow is the data analysis, which involves using several mathematical and statistical tools aimed to maximize the information that can be extracted from complex chemical data consisting of many identified and unidentified components. To study the impact of all metabolites on the outcome of a measurement (e.g. disease and normal), due to the

presence of hundreds or even thousands of metabolites signals arising from the biofluids. The inherent variability in each sample makes the use of competent and robust data mining techniques a pre-requisite for maximizing the recovery of relevant metabolic signatures or biomarkers. Data mining techniques are mainly based on multivariate statistical approaches which have been extensively described in the literature[112] and briefly in the later sections. Depending on the multivariate method, the statistics utilized in metabolomics can be divided into non-supervised and supervised techniques. Non-supervised approaches such as Principal Component Analysis (PCA)[113] are applied to explore the overall statistical variance with the goal of clustering the samples, according to the different metabolic features, and detecting outliers. Supervised statistical techniques, such as Partial Least Square Discriminant Analysis (PLS-DA)[114] and methods combining further data filtering, are often needed in metabolomics to detect class-related biomarkers. Additionally, the multivariate model can be used to determine the metabolites contributing to the result as well as predicting the result[115].

### *Unsupervised analysis*

The unsupervised analysis is the application of statistical models without the prior knowledge of the sample classification labels, and it is usually the first step in data pattern exploration. Principal component analysis (PCA) is representative of the unsupervised method and is commonly applied to reduce the dimensionality and examine the structure of the data set. Scores plot is generated to assess the clustering of different samples, with the corresponding loadings plot demonstrating the variables accounting for the most variation in the specified principal component. In addition, the cluster analysis, which is useful for the visualization of subgroups of multivariate data by partitioning methods or hierarchical clustering, is also widely used in metabolomics data analysis.

## Principal component analysis (PCA):

PCA is a commonly used unsupervised technique for multivariate data exploration (visualization and interpretation). It is a mathematical procedure that transforms a number of possibly correlated variables into a (smaller) number of uncorrelated variables called principal components[116]. The principal components (PCs) are linear combinations of the original input descriptors with appropriate weighting coefficients, such that the first PC contains the greatest amount of variance in the data and subsequent PCs contain as much of the variability in the data as possible. By plotting only the first two or three PCs, the original N-dimensional data are effectively compressed into two or three manageable dimensions. Therefore, PCA is only a coordinate system transformation where new coordinates are selected so that each of them explains the maximum amount of remaining variance in the data. The loadings are the components of the PC in the variable space, while the score is the eigenvalue of each sample

point within the PC. The resulting PCs can be used to visualize any clustering patterns associated with a metabolic response. In typical experimental studies, the data dimensions are extremely high because hundreds of variables can be observed in one NMR spectrum, so PCA is very useful in sample interpretation. PCA has frequently been applied in the evaluation of metabolomic data and should be the method of choice for obtaining an overview, find clusters, and to identify outliers, **Figure 3.4**.

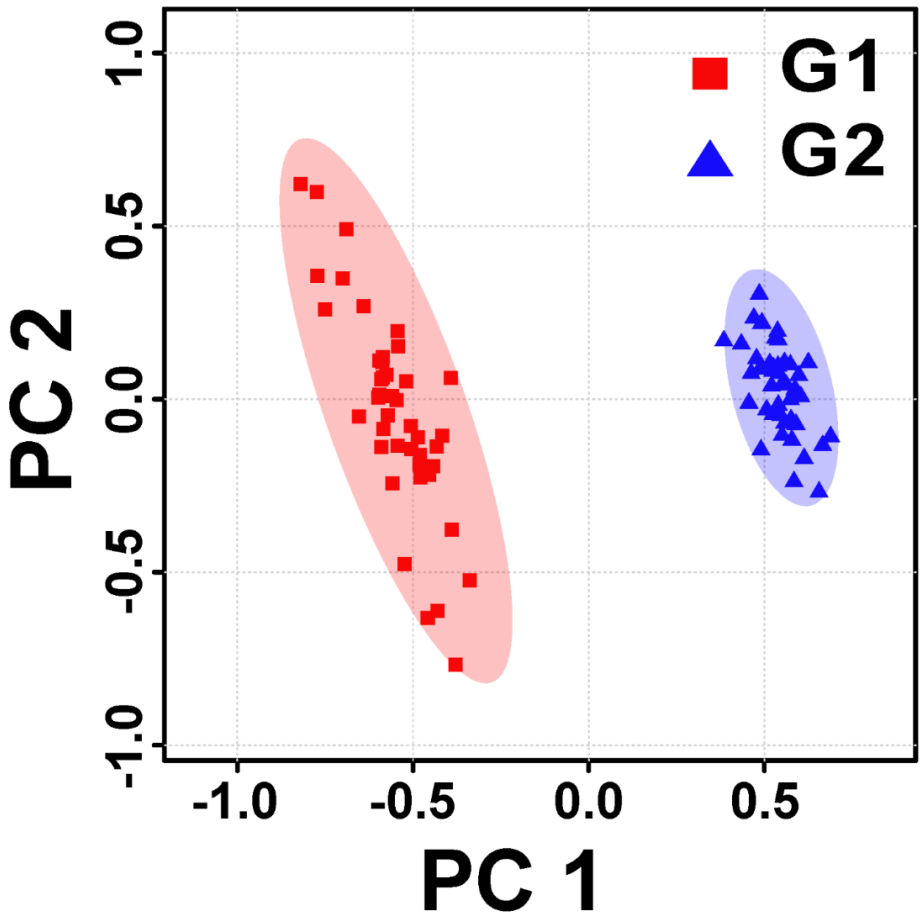



**Figure 3.4:** 2D PCA score plot.

### Cluster Analysis:

Clustering is used to group samples with similar metabolite profiles together. Clustering organizes the data into groups such that objects that are in a cluster are more closely related to each other than with the objects from the other clusters. To measure proximity many distance functions, e.g. Euclidian distance, Pearson distance, etc. are used. Good clustering results in the compact clusters that are also distant from other clusters. Some of the traditional clustering algorithms include hierarchical (HCA), k-means and self-organizing maps (SOM).

### Hierarchical Cluster Analysis (HCA):

Hierarchical clustering is a method of cluster analysis which seeks to build a hierarchy of clusters. The HCA algorithm connects objects to form clusters based on their distance. The inputs required

are similarity measures or data from which similarities can be computed using different distance functions. The main property of HCA is to highlight grouping of samples on the basis of similarities or distances (dissimilarities) with the general idea that objects are more related to nearby objects than to objects farther away. The results of hierarchical clustering are presented in a dendrogram, in which the y-axis marks the distance among clusters, while the objects are placed along the x-axis, **Figure 3.5**.

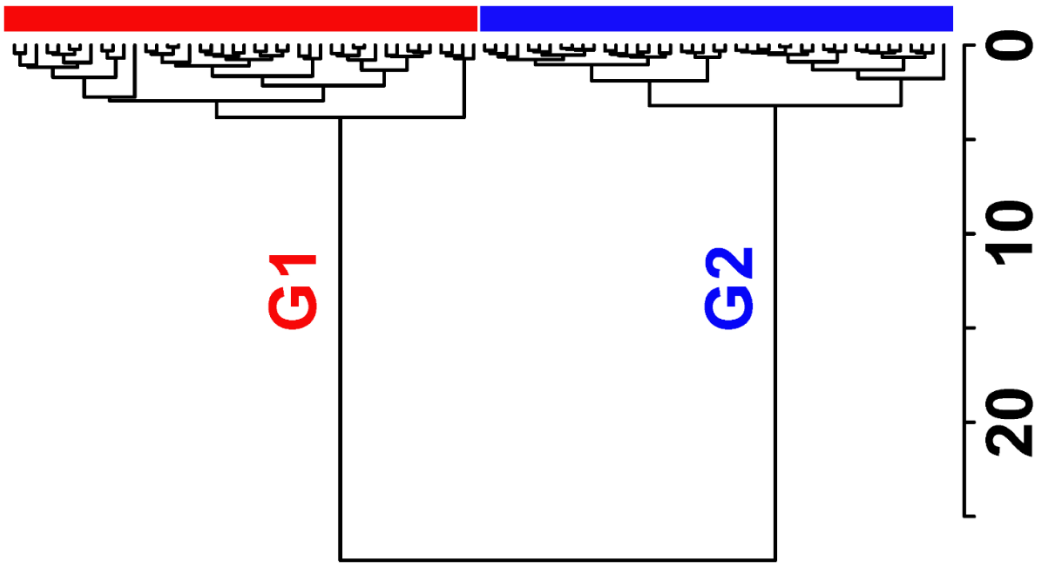



**Figure 3.5:** HCA dendrogram based on NMR data, showing similarities between samples. Samples falling within the same group are indicated with the same colour.

### Heat Map:

Heat maps are useful visualization tools to provide a quick overview of how peaks are regulated in the experimental samples. Heat maps allow users to easily visualize changing patterns in metabolite concentrations across samples and experimental conditions. In contrast to the scores plots, heat maps display the actual data values using carefully chosen colour gradients. To display only statistically significant peaks based on p-value and VIP score cut-offs. Clustered heat maps are another very popular visualization tool, **Figure 3.6**.

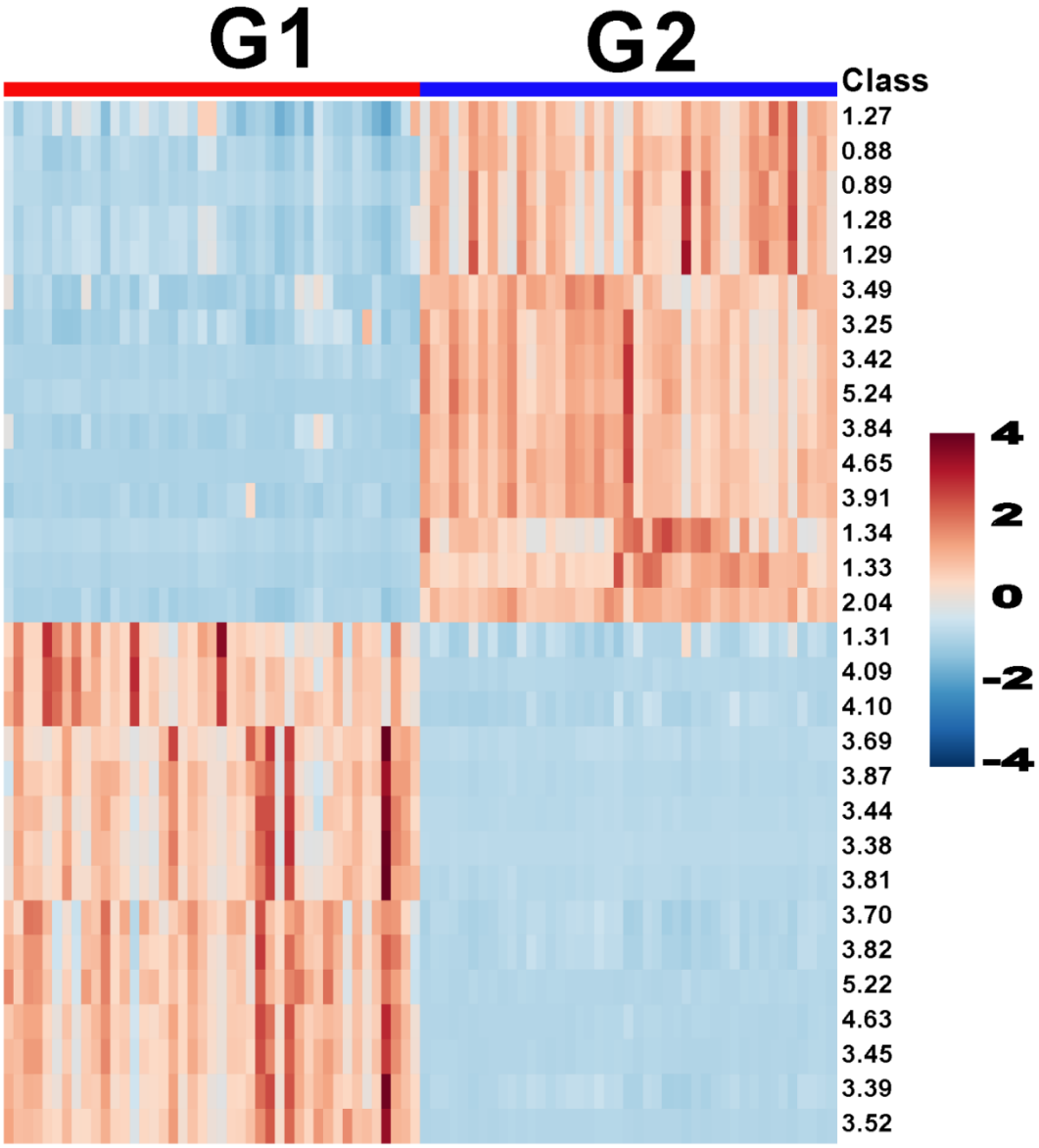



**Figure 3.6:** Heat Map visualization based on VIP scores. Red and Cyan represent elevated and depleted levels of metabolites respectively.

### *Supervised analysis*

In the supervised analysis, information of sample class labels (e.g. disease and control) are also utilized in building the statistic models. One commonly used supervised analysis is partial least squares discriminant analysis (PLS-DA) which maximizes the covariance between predictor variables (metabolite intensities from NMR measurement) and the response variables (e.g. the classes of each sample). Namely, PLS-DA finds components (e.g. latent variables) in the predictor matrix that best predict the response variables. It uses variable importance to projection (VIP) scores to demonstrate the contribution of each variable to the model, with metabolites VIP scores $> 1$ considered important in classification. If class separation is not observed in scores plot of PLS-DA model, orthogonal partial least squares discriminant analysis (OPLS-DA) can be performed. The OPLS-DA is a modification of PLS-DA model, in which the systematic variation of data that is not related to the response variable (e.g. sample class labels) is removed. It presents similar

prediction ability to PLS-DA and demonstrates improved model interpretability. Diagnostic parameters such as the number of misclassifications, cross-validation parameters: R2, Q2, permutation statistics, and the Area under the Curve (AUC) of a Receiver Operating Characteristic (ROC) analysis are commonly used to indicate the model performance.

## Partial Least Squares Discriminant Analysis (PLS-DA):

PLS-DA is a supervised method used to improve the separation between different classes of observations. This method is based on PLS regression; PLS regression is a modeling technique that takes into account collinearity in the data. Instead of using strongly dependent predictor – variables it calculates a new set of LVs (LV, the equivalents to principal components of PCA) that have a reduced dimensionality. For the PLS-DA analysis, the data analysis algorithm is provided with information on which sample belongs to which class. It then rotates the principal component in a way that a maximum separation between the classes is obtained. The method is closely related to (PLS-1) which finds the relationship between predictor variables and dependent variables by building models, one for each response. PLS-DA (also referred to as PLS-2) instead of building separate models for each response, can deal with multiple responses simultaneously via an application on the unfolded class matrix and transforming the class vector into a matrix of zeros and ones. However, because of this additional information fed into the computer algorithm, the resulting model has to be verified by a cross-validation procedures. In the end, a discriminant analysis is performed which gives an estimate how much the different groups overlap and which can be used when using the established model to analyze unknown samples to ascertain the group of the unknown sample. In general, the separation between the classes is better when using PLS-DA compared to PCA. However, one needs to be cautious to not over-fitting the data and not introducing a bias by wrong class assignments. Quite often the first two latent variables contribute to the maximum class separation, **Figure 3.7**.

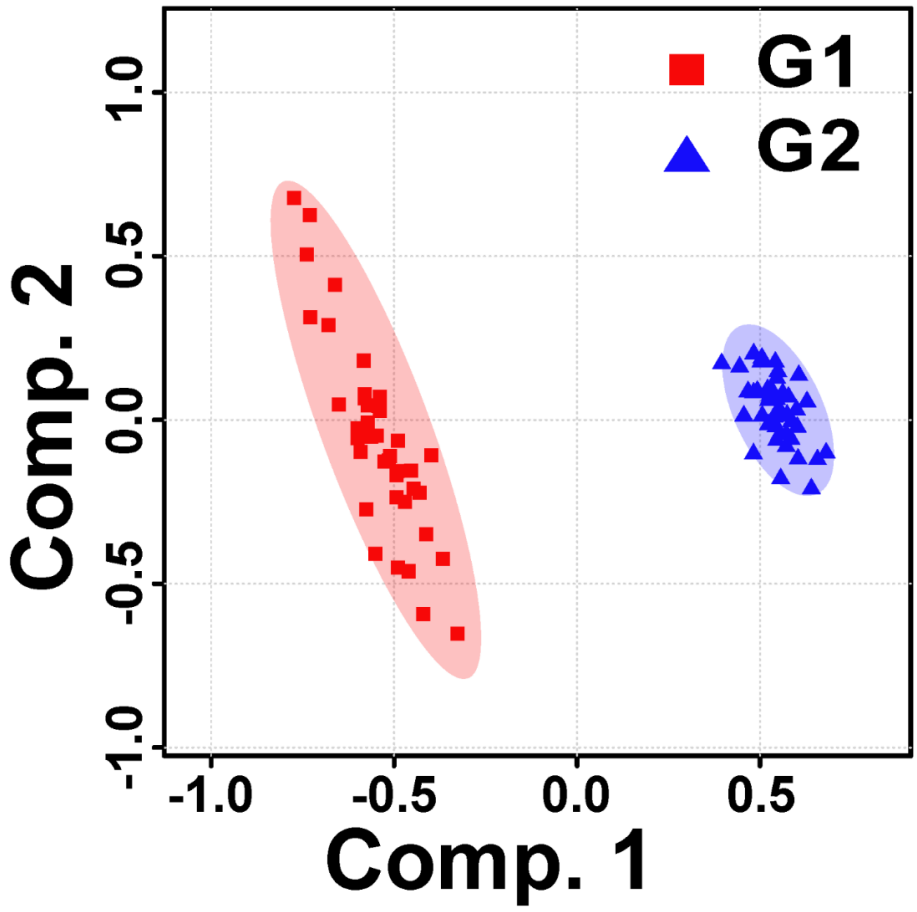



**Figure 3.7:** 2D PLS-DA score plot. Showing the class separation of the PLS-DA model, between G1 (Red squares) and HC (Blue triangles).

### Orthogonal Partial Least Squares Discriminant Analysis (OPLS-DA):

OPLS-DA is an improvement of the PLSDA method to discriminate two or more groups (classes) using multivariate data. In OPLS-DA a regression model is calculated between the multivariate data and a response variable that only contains class information. The advantage of OPLS-DA compared to PLS-DA is that a single component is used as a predictor for the class, while the other components describe the variation orthogonal to the first predictive component. In this way, OPLS-DA is an extension combining the strengths of PLS-DA and soft independent modeling of class analogy classification (SIMCA). Regarding results prediction, OPLS-DA is identical to PLS-DA modeling. The primary benefit of OPLS-DA lies in the ease of interpretation, especially in the multi-class case. This is achieved by the separate modeling of predictive and class-related variation in the x-matrix through the identification of y-orthogonal variation. Although there no predictive performance advantage of OPLS-DA over PLS-DA, its interpretation is much more superior, **Figure 3.8**.

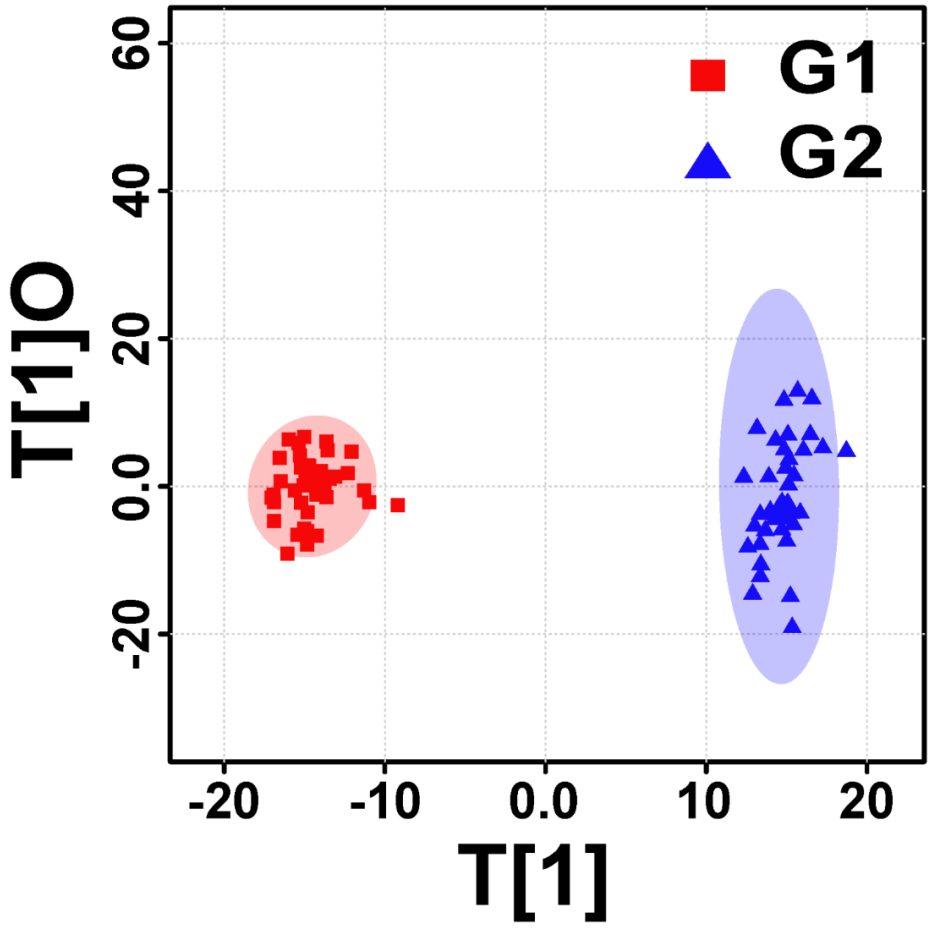



**Figure 3.8:** 2D OPLS-DA score plot. Showing the class separation of the OPLS-DA model, between G1 (Red squares) and HC (Blue triangles).

### Model Validation:

Validation is an essential step to guarantee the reliability of any developed statistical models[117]. It is also essential from the biological point of view. Properly validated statistical models give confidence to the findings (i.e. relevant metabolites)[118]. This is significant if the results from the statistical analysis are used later, for instance in clinical application. If there are many more variables than samples it is possible to find by chance a perfect model that fits data. Therefore it is mandatory to check the model for its predictive ability. There are several options for validating a model for predictive performance. To assess the predictive ability of the multivariate models, methods such as cross-validation (CV), permutation or bootstrap are usually conducted. Cross-validation involves separating the data into a training set and a test set. The training set is used to build the classification models (e.g. PLS-DA), and the resulting model is used to predict the classes of the test set[119]. Depending on how the full data is partitioned into the two data sets, leave-one-out CV (LOOCV), Monte Carlo CV (MCCV) and 10-fold CV are commonly applied. A permutation test can assess whether the classification based on true sample class is significantly better than classification based on randomly assigned sample class. Moreover, bootstrap is a

resampling method used for model validation. It generates a new dataset the same size as the original by sampling with replacement from the original dataset. This new dataset is used to build the prediction model, and the validation is applied on the original data set. For example, the most common method PLS-DA over Fits models to data where completely random variables have excellent class separation. The categorical variable Y indicates class membership, meaning that values of - 1 and 1 represent the healthy and the disease under study, respectively. Nevertheless, due to the properties of regression models, the prediction is not inevitably the exact value. The challenges of PLS-DA are classification procedure and precise estimation of the quality of the obtained models and thereby differences between two classes. To verify the quality of the obtained discrimination models, several tools have been developed, among them cross validation. Regression model validation determines that if the relationships between the variables, obtained from regression analysis, truly describe the data. In the present work, an out-of-sample evaluation was considered. Cross-validation is widely applied in chemometric as the use of this tool is indispensable to validate data from supervised tests such as PLS-DA, and if not mentioned, it is simply referred as the leave-one-out cross-validation (LOO-CV). LOO-CV is a practical and reliable fashion to test the significance of the developed models. Briefly, LOO-CV uses one observation for validation purposes while the remaining observations are considered the training data and the process are repeated until each observation is used once for validation. LOO-CV has the disadvantage of being computationally expensive and often causes overfitting, and on average, gave an underestimation of the true predictive error. MCCV can avoid an unnecessary large model and therefore decreases the risk of over-fitting for the calibration model. MCCV splits the data into a learning set or a test set, and the model developed on the learning set and the error evaluated in the test set. The test set estimates are averaged over the learning testing random splits, and each case only appears in the learning set or the test set, but not in both. MCCV substantially reduces the variance of the split-sample error estimate. MCCV yields the statistical model classification rate, sensitivity, and specificity. The classification rate indicates the samples correctly identified, the sensitivity yields the percentage of positives that are correctly identified and specificity measures the percentage of negatives to be true negatives. Receiver operating characteristics (ROC) curves plot sensitivity versus (1-specificity). This graphical illustration gives the proportion of true positives against the false positives. Also, the values of $R^2$ could be used to assess the degree of the model to the data in spite not being a cross-validation evaluation. However, if $R^2$ is much higher than $Q^2$ obtained in MCCV it possibly indicates model overfitting, **Figure 3.9**.

Validation can be also performed using a completely new set of samples, coming from the independent new experiment. This kind of validation is the ultimate one because it allows one to test the outcomes of analysis on a different population. Unfortunately, such validation is rarely applied.

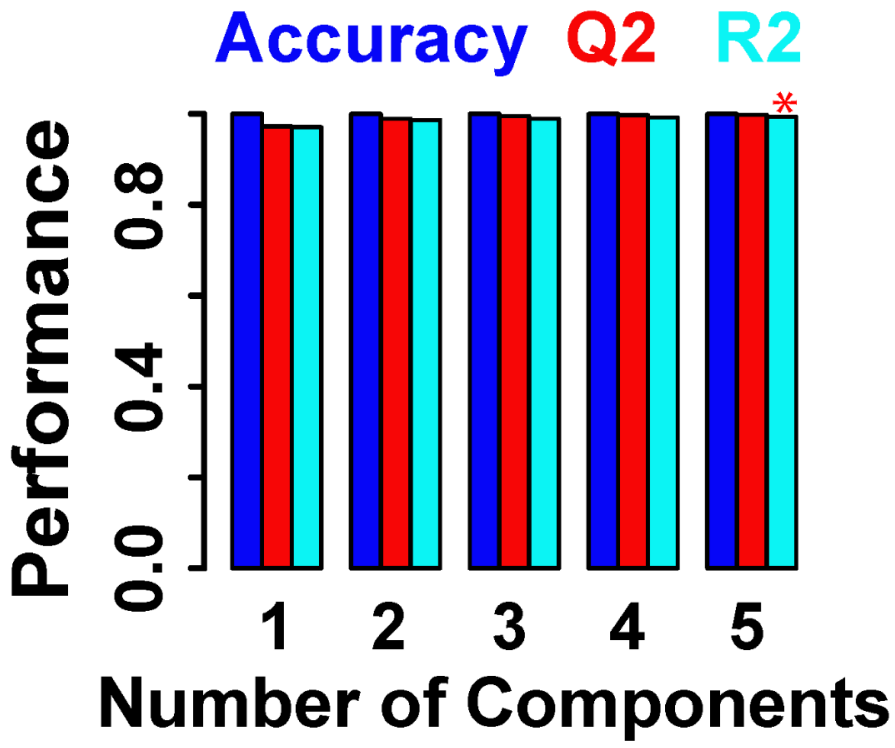



**Figure 3.9:** Statistical validation of the PLS-DA model, using 10-fold CV as represented by Accuracy, $R^2$ and $Q^2$ Values.

### Receiver Operating Characteristic Curve (ROC):

The ROC curve is a tool for biomarker analysis which is normally used in clinical applications. A ROC curve is a graphical plot of the sensitivity vs. 1-specificity. The sensitivity is calculated by true positives / (true positives + false negatives) and specificity is calculated by true negative / (true negative + false positive). The best possible prediction method yields a point in the upper left corner of the ROC space, representing 100% sensitivity and 100% specificity. A completely random guess creates a point along a diagonal line (the so-called line of no-discrimination from the left bottom to the top right corner). Visual inspection of the scores plot and the area under the ROC curve was used to determine the strength of a model. The area under the curve (AUC) represents the accuracy of the test and is used to estimate the degree of difference between groups such as excellent (0.9-1.0), good (0.8-0.9) and fair (0.7-0.8)[120]. ROC has been used in many metabolomics studies for potential biomarker validation but is not commonly used in the early stages of a study where the discovery of potential biomarkers is the focus. However, since ROC is widely used in clinical studies, the application of ROC curve in metabolomics is useful, especially when assessing the potential value of a newly discovered biomarker, **Figure 3.10**.

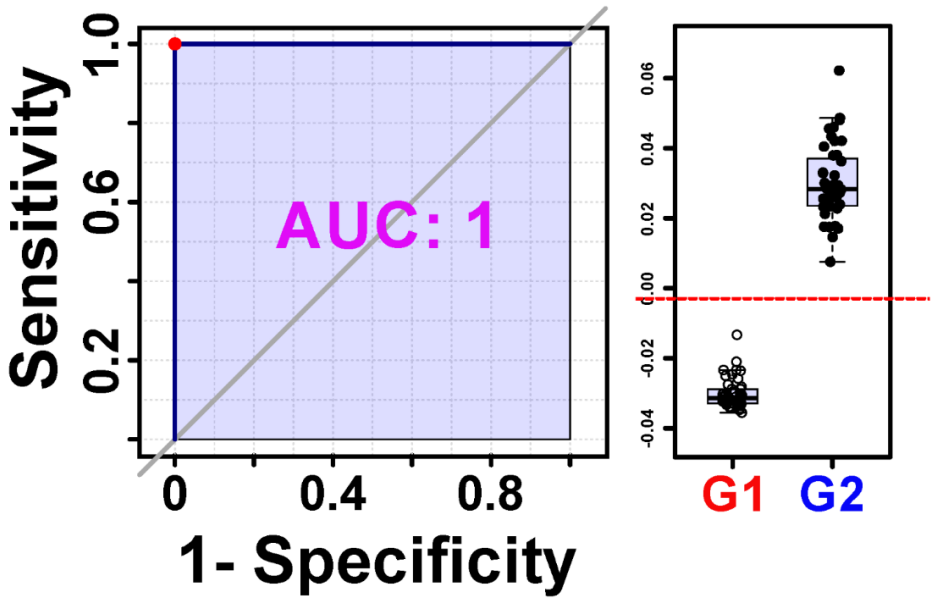



**Figure 3.10:** Represents the Receiver operating characteristic (ROC) curves of the metabolites discriminating G1 from G2 along with their respective box plots. AUC indicates the area under the curve and the dot refer to the cutoff value maximizing sensitivity and specificity for the given samples. For each box plot, boxes denote interquartile ranges; lines denote medians, and whiskers denote 10$^{th}$ and 90$^{th}$ percentiles.

## Univariate Analysis

The univariate statistical analysis encompasses the broad range of traditional statistical methods in which only one predictor variable is considered at a time. In the context of metabolomics data analysis, univariate analysis is often used in the first stages of research for descriptive purposes, where individual metabolites are viewed (and sometimes modeled) singly, or also as a confirmatory tool following multivariate analysis. Though, it is supplemented by more advanced multivariate statistical methods. The focus of the present work is the application of multivariate techniques for the analysis of high-dimensional data. Univariate statistical significant test (or hypothesis testing) can be divided into two groups, parametric and non-parametric, based on the variable distribution (**Table 3.5**). However, some selected univariate analysis methods (specifically, parametric and non-parametric equivalents of t-tests are used for comparative purposes.

| Comparison | Parametric | Non-parametric |
|---|---|---|
| | Compare Means | Compare Medians |
| Compare 2 unpaired groups | Unpaired Student's t-test | Mann-Whitney |
| Compare 2 paired groups | Paired Student's t-test | Wilcoxon signed-rank |
| Compared more than 2 unmatched group | One-way ANOVA | Kruskal-Wallis |
| Compared more than 2 matched group | Two-way ANOVA | Friedman |

**Table 3.5:** A list of various parametric and non-parametric univariate statistical tests.

## Pathway analysis in disease biomarker discovery:

Metabolic pathway analysis is essential for the understanding of cellular processes of specific diseases, providing insight into the development of treatment methods. After the identification of potential metabolite biomarkers, the particular pathway can be assessed using databases or searching for literature that contains pathway analysis of certain metabolite and disease. The enzymes controlling the metabolite biomarker levels in the cell play a critical role in the development of the disease. Thus, studies focusing on the knockdown of a certain enzyme in the cell can be applied to test its impact on the biomarker level and promote understanding of mechanisms associated with the specific disease.

| Pathway Databases | | Pathway Analysis Tools | |
|---|---|---|---|
| **Name** | **Link** | **Tool** | **Link** |
| **HMDB** | www.hmdb.ca | **PathVisio** | http://www.pathvisio.org/ |
| **KEGG** | www.genome.jp/kegg/ | **MetaCore** | https://portal.genego.com/ |
| **Reactome** | www.reactome.org | **Ingenuity** | http://www.ingenuity.com/ |
| **WikiPathways** | www.wikipathways.org | **Cytoscape** | http://www.cytoscape.org/ |

**Table 3.6:** A list of various pathway databases and pathway analysis tools.

## Biological Interpretation

Biological interpretation is the next crucial step in metabolomics to draw insights and conclusions about a disease/stressor. The task of understanding each metabolite is confounded by the fact that metabolites can be involved in multiple pathways and that the type of sample would change the interpretation accordingly. The output from a successful statistical analysis in any metabolomics study is usually a long list of features (or metabolites) that have changed significantly or showed interesting patterns of coordinated change under the different conditions. Obtaining this kind of list is usually not the end point of one's analysis; rather, it is the starting point for data interpretation. Over the past decade, many approaches to computer-assisted data interpretation have been explored and tested. Among them, group-based significance tests and pathway analysis methods have gained widespread acceptance among many researchers involved in "omics" data analysis[121]. These two approaches allow the incorporation of pre-existing biological knowledge into the data analysis process, which greatly facilitates the data interpretation process. As a general rule, functional analysis can only be applied when we know the compound identities and their concentrations. Consequently, data produced from standard quantitative metabolomics approaches (i.e., concentration tables) can be directly used for functional analysis. On the other hand, for chemometric (non-targeted) approaches, significant

features (i.e., peaks, spectral bins) must be identified first, which must be further characterized to produce a list of compound names that can then be used for functional interpretation[122].

**Reference's**


1. Eriksson, L.; Andersson, P. L.; Johansson, E.; Tysklind, M. Megavariate analysis of environmental QSAR data. Part IGÇôA basic framework founded on principal component analysis (PCA), partial least squares (PLS), and statistical molecular design (SMD). *Molecular diversity* **2006**, *10* (2), 169-186.

2. John, J. K. N.; Lindon, C.; Holmes, E. The Handbook of Metabonomics and Metabolomics. 2007. Elsevier, Amsterdam, The Netherlands. Ref Type: Generic

3. Cramer III, R. D. Partial least squares (PLS): its strengths and limitations. *Perspectives in Drug Discovery and Design* **1993**, *1* (2), 269-278.

4. Parvin, C. A. An Introduction to Multivariate Statistical Analysis, TW Anderson. Hoboken, NJ: John Wiley & Sons, 2003, 742 pp., $99.95, hardcover. ISBN 0-471-36091-0. *Clinical Chemistry* **2004**, *50* (5), 981-982.

5. Gowda, G. N.; Zhang, S.; Gu, H.; Asiago, V.; Shanaiah, N.; Raftery, D. Metabolomics-based methods for early disease diagnostics. *Expert review of molecular diagnostics* **2008**, *8* (5), 617-633.

6. Gowda, G. N.; Shanaiah, N.; Cooper, A.; Maluccio, M.; Raftery, D. Bile acids conjugation in human bile is not random: new insights from 1H-NMR spectroscopy at 800 MHz. *Lipids* **2009**, *44* (6), 527-535.

7. Maher, A. D.; Cloarec, O.; Patki, P.; Craggs, M.; Holmes, E.; Lindon, J. C.; Nicholson, J. K. Dynamic biochemical information recovery in spontaneous human seminal fluid reactions via 1H NMR kinetic statistical total correlation spectroscopy. *Analytical chemistry* **2008**, *81* (1), 288-295.

8. Graca, G.; Duarte, I. F.; Goodfellow, B. J.; Barros, A. S.; Carreira, I. M.; Couceiro, A. B.; Spraul, M.; Gil, A. M. Potential of NMR spectroscopy for the study of human amniotic fluid. *Analytical chemistry* **2007**, *79* (21), 8367-8375.

9. Bertram, H. C.; Eggers, N.; Eller, N. Potential of human saliva for nuclear magnetic resonance-based metabolomics and for health-related biomarker identification. *Analytical chemistry* **2009**, *81* (21), 9188-9193.

10. Beckonert, O.; Keun, H. C.; Ebbels, T. M.; Bundy, J.; Holmes, E.; Lindon, J. C.; Nicholson, J. K. Metabolic profiling, metabolomic and metabonomic procedures for NMR spectroscopy of urine, plasma, serum and tissue extracts. *Nature protocols* **2007**, *2* (11), 2692-2703.

11. Lindon, J. C.; Nicholson, J. K.; Holmes, E.; Everett, J. R. Metabonomics: metabolic processes studied by NMR spectroscopy of biofluids. *Concepts in Magnetic Resonance* **2000**, *12* (5), 289-320.

12. Khan, A. R.; Rana, P.; Tyagi, R.; Kumar, I. P.; Devi, M. M.; Javed, S.; Tripathi, R. P.; Khushu, S. NMR spectroscopy based metabolic profiling of urine and serum for investigation of physiological perturbations during radiation sickness. *Metabolomics* **2011**, *7* (4), 583-592.

13. Fiehn, O. MetabolomicsGÇôthe link between genotypes and phenotypes. *Plant molecular biology* **2002**, *48* (1-2), 155-171.

14. Lindon, J. C.; Nicholson, J. K. Analytical technologies for metabonomics and metabolomics, and multi-omic information recovery. *TrAC Trends in Analytical Chemistry* **2008**, *27* (3), 194-204.

15. Lenz, E. M.; Wilson, I. D. Analytical strategies in metabonomics. *Journal of proteome research* **2007**, *6* (2), 443-458.

16. Holmes, E.; Foxall, P. J.; Spraul, M.; Farrant, R. D.; Nicholson, J. K.; Lindon, J. C. 750 MHz 1 H NMR spectroscopy characterisation of the complex metabolic pattern of urine from patients with inborn errors of metabolism: 2-hydroxyglutaric aciduria and maple syrup urine disease. *Journal of pharmaceutical and biomedical analysis* **1997**, *15* (11), 1647-1659.

17. Beckonert, O.; Keun, H. C.; Ebbels, T. M.; Bundy, J.; Holmes, E.; Lindon, J. C.; Nicholson, J. K. Metabolic profiling, metabolomic and metabonomic procedures for NMR spectroscopy of urine, plasma, serum and tissue extracts. *Nature protocols* **2007**, *2* (11), 2692-2703.

18. Ebbels, T. M.; Keun, H. C.; Beckonert, O. P.; Bollard, M. E.; Lindon, J. C.; Holmes, E.; Nicholson, J. K. Prediction and classification of drug toxicity using probabilistic modeling of temporal metabolic data: the consortium on metabonomic toxicology screening approach. *Journal of proteome research* **2007**, *6* (11), 4407-4422.

19. Beckonert, O.; Keun, H. C.; Ebbels, T. M.; Bundy, J.; Holmes, E.; Lindon, J. C.; Nicholson, J. K. Metabolic profiling, metabolomic and metabonomic procedures for NMR spectroscopy of urine, plasma, serum and tissue extracts. *Nature protocols* **2007**, *2* (11), 2692-2703.

20. Mckay, R. T. How the 1DGÇÉNOESY suppresses solvent signal in metabonomics NMR spectroscopy: An examination of the pulse sequence components and evolution. *Concepts in Magnetic Resonance Part A* **2011**, *38* (5), 197-220.

21. Mo, H.; Raftery, D. Pre-SAT180, a simple and effective method for residual water suppression. *Journal of Magnetic Resonance* **2008**, *190* (1), 1-6.

22. Viant, M. R. Improved methods for the acquisition and interpretation of NMR metabolomic data. *Biochemical and biophysical research communications* **2003**, *310* (3), 943-948.

23. Meiboom, S.; Gill, D. Modified spinGÇÉecho method for measuring nuclear relaxation times. *Review of scientific instruments* **1958**, *29* (8), 688-691.

24. Wu, D. H.; Chen, A. D.; Johnson, C. S. An improved diffusion-ordered spectroscopy experiment incorporating bipolar-gradient pulses. *Journal of magnetic resonance, Series A* **1995**, *115* (2), 260-264.

25. Sandusky, P.; Raftery, D. Use of selective TOCSY NMR experiments for quantifying minor components in complex mixtures: application to the metabonomics of amino acids in honey. *Analytical chemistry* **2005**, *77* (8), 2455-2463.

26. Sandusky, P.; Raftery, D. Use of semiselective TOCSY and the pearson correlation for the metabonomic analysis of biofluid mixtures: application to urine. *Analytical chemistry* **2005**, *77* (23), 7717-7723.

27. Dumas, M. E.; Canlet, C. +.; Andr+¬, F.; Vercauteren, J.; Paris, A. Metabonomic assessment of physiological disruptions using 1H-13C HMBC-NMR spectroscopy combined with pattern recognition procedures performed on filtered variables. *Analytical chemistry* **2002**, *74* (10), 2261-2273.

28. Dumas, M. E.; Canlet, C.; Vercauteren, J.; Andre, F.; Paris, A. Homeostatic signature of anabolic steroids in cattle using 1H-13C HMBC NMR metabonomics. *Journal of proteome research* **2005**, *4* (5), 1493-1502.

29. Xi, Y.; de Ropp, J. S.; Viant, M. R.; Woodruff, D. L.; Yu, P. Automated screening for metabolites in complex mixtures using 2D COSY NMR spectroscopy. *Metabolomics* **2006**, *2* (4), 221-233.

30. Viant, M. R. Improved methods for the acquisition and interpretation of NMR metabolomic data. *Biochemical and biophysical research communications* **2003**, *310* (3), 943-948.

31. Chang, D.; Banack, C. D.; Shah, S. L. Robust baseline correction algorithm for signal dense NMR spectra. *Journal of Magnetic Resonance* **2007**, *187* (2), 288-292.

32. Gan, F.; Ruan, G.; Mo, J. Baseline correction by improved iterative polynomial fitting with automatic threshold. *Chemometrics and Intelligent Laboratory Systems* **2006**, *82* (1), 59-65.

33. Chenomx, N. M. R. Suite 7.0.(2010). *Chenomx Inc.*

34. Xi, Y.; Rocke, D. M. Baseline correction for NMR spectroscopic metabolomics data analysis. *BMC bioinformatics* **2008**, *9* (1), 1.

35. Eilers, P. H. A perfect smoother. *Analytical chemistry* **2003**, *75* (14), 3631-3636.

36. Eilers, P. H.; Marx, B. D. Flexible smoothing with B-splines and penalties. *Statistical science* **1996**, 89-102.

37. de Rooi, J. J.; Eilers, P. H. Mixture models for baseline estimation. *Chemometrics and Intelligent Laboratory Systems* **2012**, *117*, 56-60.

38. Lindon, J. C.; Nicholson, J. K.; Everett, J. R. NMR spectroscopy of biofluids. *Annual reports on NMR spectroscopy* **1999**, *38*, 1-88.

39. Bell, J. D.; Brown, J. C.; Sadler, P. J. NMR studies of body fluids. *NMR in Biomedicine* **1989**, *2* (5GÇÉ6), 246-256.

40. Nicholson, J. K.; Wilson, I. D. High resolution proton magnetic resonance spectroscopy of biological fluids. *Progress in Nuclear Magnetic Resonance Spectroscopy* **1989**, *21* (4-5), 449-501.

41. Griffin, J. L.; Nicholls, A. W.; Keun, H. C.; Mortishire-Smith, R. J.; Nicholson, J. K.; Kuehn, T. Metabolic profiling of rodent biological fluids via 1 H NMR spectroscopy using a 1 mm microlitre probe. *Analyst* **2002**, *127* (5), 582-584.

42. Nicholson, J. K.; Foxall, P. J.; Spraul, M.; Farrant, R. D.; Lindon, J. C. 750 MHz 1H and 1H-13C NMR spectroscopy of human blood plasma. *Analytical chemistry* **1995**, *67* (5), 793-811.

43. Lutz, N. W.; Maillet, S.; Nicoli, F.; Viout, P.; Cozzone, P. J. Further assignment of resonances in 1H NMR spectra of cerebrospinal fluid (CSF). *FEBS letters* **1998**, *425* (2), 345-351.

44. Gra+°a, G.; Duarte, I. F.; Barros, A. n. S.; Goodfellow, B. J.; Diaz, S. ü.; Carreira, I. M.; Couceiro, A. B.; Galhano, E.; Gil, A. M. 1H NMR based metabonomics of human amniotic fluid for the metabolic characterization of fetus malformations. *Journal of proteome research* **2009**, *8* (8), 4144-4150.

45. Beckonert, O.; Keun, H. C.; Ebbels, T. M.; Bundy, J.; Holmes, E.; Lindon, J. C.; Nicholson, J. K. Metabolic profiling, metabolomic and metabonomic procedures for NMR spectroscopy of urine, plasma, serum and tissue extracts. *Nature protocols* **2007**, *2* (11), 2692-2703.

46. Walsh, M. C.; Brennan, L.; Malthouse, J. P.; Roche, H. M.; Gibney, M. J. Effect of acute dietary standardization on the urinary, plasma, and salivary metabolomic profiles of healthy humans. *The American journal of clinical nutrition* **2006**, *84* (3), 531-539.

47. Markley, J. L.; Ulrich, E. L.; Berman, H. M.; Henrick, K.; Nakamura, H.; Akutsu, H. BioMagResBank (BMRB) as a partner in the Worldwide Protein Data Bank (wwPDB): new policies affecting biomolecular NMR depositions. *Journal of biomolecular NMR* **2008**, *40* (3), 153-155.

48. Wishart, D. S.; Jewison, T.; Guo, A. C.; Wilson, M.; Knox, C.; Liu, Y.; Djoumbou, Y.; Mandal, R.; Aziat, F.; Dong, E. HMDB 3.0-the human metabolome database in 2013. *Nucleic acids research* **2012**, gks1065.

49. Cui, Q.; Lewis, I. A.; Hegeman, A. D.; Anderson, M. E.; Li, J.; Schulte, C. F.; Westler, W. M.; Eghbalnia, H. R.; Sussman, M. R.; Markley, J. L. Metabolite identification via the madison metabolomics consortium database. *Nature biotechnology* **2008**, *26* (2), 162-164.

50. Ludwig, C.; Easton, J. M.; Lodi, A.; Tiziani, S.; Manzoor, S. E.; Southam, A. D.; Byrne, J. J.; Bishop, L. M.; He, S.; Arvanitis, T. N. Birmingham Metabolite Library: a publicly accessible database of 1-D 1H and 2-D 1H J-resolved NMR spectra of authentic metabolite standards (BML-NMR). *Metabolomics* **2012**, *8* (1), 8-18.

51. Watanabe, K.; Yasugi, E.; Oshima, M. How to Search the Glycolipid Data in" LIPID~ B~ A~ N~ K for Web", the Newly Developed Lipid Database in Japan. *Trends in Glycoscience and Glycotechnology* **2000**, *12* (65), 175-184.

52. Lundberg, P.; Vogel, T.; Malusek, A.; Lundquist, P. O.; Cohen, L.; hlqvist Leinhard, O. MDLGÇôthe magnetic resonance metabolomics database. 2005.

53. Berman, H.; Henrick, K.; Nakamura, H.; Markley, J. L. The worldwide Protein Data Bank (wwPDB): ensuring a single, uniform archive of PDB data. *Nucleic acids research* **2007**, *35* (suppl 1), D301-D303.

54. Pauli, G. F.; Jaki, B. U.; Lankin, D. C. Quantitative 1H NMR: Development and Potential of a Method for Natural Products Analysis -°. *Journal of natural products* **2005**, *68* (1), 133-149.

55. Lane, S.; Boughtflower, B.; Mutton, I.; Paterson, C.; Farrant, D.; Taylor, N.; Blaxill, Z.; Carmody, C.; Borman, P. Toward single-calibrant quantification in HPLC. A comparison of three detection strategies: evaporative light scattering, chemiluminescent nitrogen, and proton NMR. *Analytical chemistry* **2005**, *77* (14), 4354-4365.

56. Alum, M. F.; Shaw, P. A.; Sweatman, B. C.; Ubhi, B. K.; Haselden, J. N.; Connor, S. C. 4, 4-Dimethyl-4-silapentane-1-ammonium trifluoroacetate (DSA), a promising universal internal standard for NMR-based metabolic profiling studies of biofluids, including blood plasma and serum. *Metabolomics* **2008**, *4* (2), 122-127.

57. Akoka, S.; Barantin, L.; Trierweiler, M. Concentration measurement by proton NMR using the ERETIC method. *Analytical chemistry* **1999**, *71* (13), 2554-2557.

58. Farrant, R. D.; Hollerton, J. C.; Lynn, S. M.; Provera, S.; Sidebottom, P. J.; Upton, R. J. NMR quantification using an artificial signal. *Magnetic Resonance in Chemistry* **2010**, *48* (10), 753-762.

59. Willcott, M. R. MestRe Nova. *Journal of the American Chemical Society* **2009**, *131* (36), 13180.

60. Weljie, A. M.; Newton, J.; Mercier, P.; Carlson, E.; Slupsky, C. M. Targeted profiling: quantitative analysis of 1H NMR metabolomics data. *Analytical chemistry* **2006**, *78* (13), 4430-4442.

61. Roberts, M. J.; Schirra, H.; Lavin, M. F.; Gardiner, R. A. NMR-based metabolomics: global analysis of metabolites to address problems in prostate cancer. **2014**.

62. Vogels, J. T. W. E.; Tas, A. C.; Venekamp, J.; Van Der Greef, J. Partial linear fit: A new NMR spectroscopy preprocessing tool for pattern recognition applications. *Journal of Chemometrics* **1996**, *10* (5GÇÉ6), 425-438.

63. Nielsen, N. P. V.; Carstensen, J. M.; Smedsgaard, J. +. Aligning of single and multiple wavelength chromatographic profiles for chemometric data analysis using correlation optimised warping. *Journal of Chromatography A* **1998**, *805* (1), 17-35.

64. Forshed, J.; Schuppe-Koistinen, I.; Jacobsson, S. P. Peak alignment of NMR signals by means of a genetic algorithm. *Analytica chimica acta* **2003**, *487* (2), 189-199.

65. Torgrip, R. J.; +àberg, M.; Karlberg, B.; Jacobsson, S. P. Peak alignment using reduced set mapping. *Journal of Chemometrics* **2003**, *17* (11), 573-582.

66. Lee, G. C.; Woodruff, D. L. Beam search for peak alignment of NMR signals. *Analytica chimica acta* **2004**, *513* (2), 413-416.

67. Stoyanova, R.; Nicholls, A. W.; Nicholson, J. K.; Lindon, J. C.; Brown, T. R. Automatic alignment of individual peaks in large high-resolution spectral data sets. *Journal of Magnetic Resonance* **2004**, *170* (2), 329-335.

68. Wu, W.; Daszykowski, M.; Walczak, B.; Sweatman, B. C.; Connor, S. C.; Haselden, J. N.; Crowther, D. J.; Gill, R. W.; Lutz, M. W. Peak alignment of urine NMR spectra using fuzzy warping. *Journal of chemical information and modeling* **2006**, *46* (2), 863-875.

69. Csenki, L.; Alm, E.; Torgrip, R. J.; +àberg, K. M.; Nord, L. I.; Schuppe-Koistinen, I.; Lindberg, J. Proof of principle of a generalized fuzzy Hough transform approach to peak alignment of one-dimensional 1H NMR data. *Analytical and bioanalytical chemistry* **2007**, *389* (3), 875-885.

70. Veselkov, K. A.; Lindon, J. C.; Ebbels, T. M.; Crockford, D.; Volynkin, V. V.; Holmes, E.; Davies, D. B.; Nicholson, J. K. Recursive segment-wise peak alignment of biological 1H NMR spectra for improved metabolic biomarker recovery. *Analytical chemistry* **2008**, *81* (1), 56-66.

71. Staab, J. M.; O'Connell, T. M.; Gomez, S. M. Enhancing metabolomic data analysis with Progressive Consensus Alignment of NMR Spectra (PCANS). *BMC bioinformatics* **2010**, *11* (1), 1.

72. Kim, S. B.; Wang, Z.; Hiremath, B. A Bayesian approach for the alignment of high-resolution NMR spectra. *Annals of Operations Research* **2010**, *174* (1), 19-32.

73. Savorani, F.; Tomasi, G.; Engelsen, S. B. icoshift: A versatile tool for the rapid alignment of 1D NMR spectra. *Journal of Magnetic Resonance* **2010**, *202* (2), 190-202.

74. Vu, T. N.; Valkenborg, D.; Smets, K.; Verwaest, K. A.; Dommisse, R.; Lemi+¿re, F.; Verschoren, A.; Goethals, B.; Laukens, K. An integrated workflow for robust alignment and simplified quantitative analysis of NMR spectrometry data. *BMC bioinformatics* **2011**, *12* (1), 405.

75. Vu, T. N.; Laukens, K. Getting your peaks in line: a review of alignment methods for NMR spectral data. *Metabolites* **2013**, *3* (2), 259-276.

76. Giskeodegard, G. F.; Bloemberg, T. G.; Postma, G.; Sitter, B.; Tessem, M. B.; Gribbestad, I. S.; Bathen, T. F.; Buydens, L. M. Alignment of high resolution magic angle spinning magnetic resonance spectra using warping methods. *Analytica chimica acta* **2010**, *683* (1), 1-11.

77. Lindon, J. C.; Nicholson, J. K.; Holmes, E.; Keun, H. C.; Craig, A.; Pearce, J. T.; Bruce, S. J.; Hardy, N.; Sansone, S. A.; Antti, H. Summary recommendations for standardization and reporting of metabolic analyses. *Nature biotechnology* **2005**, *23* (7), 833.

78. John, J. K. N.; Lindon, C.; Holmes, E. The Handbook of Metabonomics and Metabolomics. 2007. Elsevier, Amsterdam, The Netherlands. Ref Type: Generic

79. Holmes, E.; Nicholls, A. W.; Lindon, J. C.; Connor, S. C.; Connelly, J. C.; Haselden, J. N.; Damment, S. J.; Spraul, M.; Neidig, P.; Nicholson, J. K. Chemometric models for toxicity classification based on NMR spectra of biofluids. *Chemical research in toxicology* **2000**, *13* (6), 471-478.

80. Howells, S. L.; Maxwell, R. J.; Peet, A. C.; Griffiths, J. R. An investigation of tumor 1H nuclear magnetic resonance spectra by the application of chemometric techniques. *Magnetic resonance in medicine* **1992**, *28* (2), 214-236.

81. Messana, I.; Forni, F.; Ferrari, F.; Rossi, C.; Giardina, B.; Zuppi, C. Proton nuclear magnetic resonance spectral profiles of urine in type II diabetic patients. *Clinical Chemistry* **1998**, *44* (7), 1529-1534.

82. Lundina, T. A.; Knubovets, T. L.; Sedov, K. R.; Markova, S. A.; Sibeldin, L. A. Variability of kidney tubular interstitial distortions in glomerulonephritis as measured by 1H-NMR urinalysis. *Clinica chimica acta* **1993**, *214* (2), 165-173.

83. Lindon, J. C.; Holmes, E.; Nicholson, J. K. Pattern recognition methods and applications in biomedical magnetic resonance. *Progress in Nuclear Magnetic Resonance Spectroscopy* **2001**, *39* (1), 1-40.

84. Craig, A.; Cloarec, O.; Holmes, E.; Nicholson, J. K.; Lindon, J. C. Scaling and normalization effects in NMR spectroscopic metabonomic data sets. *Analytical chemistry* **2006**, *78* (7), 2262-2267.

85. Odunsi, K.; Wollman, R. M.; Ambrosone, C. B.; Hutson, A.; McCann, S. E.; Tammela, J.; Geisler, J. P.; Miller, G.; Sellers, T.; Cliby, W. Detection of epithelial ovarian cancer using 1HGÇÉNMRGÇÉbased metabonomics. *International journal of cancer* **2005**, *113* (5), 782-788.

86. Viant, M. R.; Lyeth, B. G.; Miller, M. G.; Berman, R. F. An NMR metabolomic investigation of early metabolic disturbances following traumatic brain injury in a mammalian model. *NMR in Biomedicine* **2005**, *18* (8), 507-516.

87. Alm, E.; Torgrip, R. J.; +àberg, K. M.; Schuppe-Koistinen, I.; Lindberg, J. A solution to the 1D NMR alignment problem using an extended generalized fuzzy Hough transform and mode support. *Analytical and bioanalytical chemistry* **2009**, *395* (1), 213-223.

88. Keun, H. C.; Ebbels, T. M.; Antti, H.; Bollard, M. E.; Beckonert, O.; Schlotterbeck, G. +.; Senn, H.; Niederhauser, U.; Holmes, E.; Lindon, J. C. Analytical reproducibility in 1H NMR-based metabonomic urinalysis. *Chemical research in toxicology* **2002**, *15* (11), 1380-1386.

89. Cloarec, O.; Dumas, M. E.; Trygg, J.; Craig, A.; Barton, R. H.; Lindon, J. C.; Nicholson, J. K.; Holmes, E. Evaluation of the orthogonal projection on latent structure model limitations caused by chemical shift variability and improved visualization of biomarker changes in 1H NMR spectroscopic metabonomic studies. *Analytical chemistry* **2005**, *77* (2), 517-526.

90. Kohl, S. M.; Klein, M. S.; Hochrein, J.; Oefner, P. J.; Spang, R.; Gronwald, W. State-of-the art data normalization methods improve NMR-based metabolomic analysis. *Metabolomics* **2012**, *8* (1), 146-160.

91. Smolinska, A.; Blanchet, L.; Buydens, L. M.; Wijmenga, S. S. NMR and pattern recognition methods in metabolomics: from data acquisition to biomarker discovery: a review. *Analytica chimica acta* **2012**, *750*, 82-97.

92. Liland, K. H. Multivariate methods in metabolomicsGÇôfrom pre-processing to dimension reduction and statistical analysis. *TrAC Trends in Analytical Chemistry* **2011**, *30* (6), 827-841.

93. Liland, K. H.; Alm++y, T.; Mevik, B. H. Optimal choice of baseline correction for multivariate calibration of spectra. *Applied spectroscopy* **2010**, *64* (9), 1007-1016.

94. Craig, A.; Cloarec, O.; Holmes, E.; Nicholson, J. K.; Lindon, J. C. Scaling and normalization effects in NMR spectroscopic metabonomic data sets. *Analytical chemistry* **2006**, *78* (7), 2262-2267.

95. Dieterle, F.; Ross, A.; Schlotterbeck, G. +.; Senn, H. Probabilistic quotient normalization as robust method to account for dilution of complex biological mixtures. Application in 1H NMR metabonomics. *Analytical chemistry* **2006**, *78* (13), 4281-4290.

96. Holmes, E.; Foxall, P. J. D.; Nicholson, J. K.; Neild, G. H.; Brown, S. M.; Beddell, C. R.; Sweatman, B. C.; Rahr, E.; Lindon, J. C.; Spraul, M. Automatic data reduction and pattern recognition methods for analysis of 1 H nuclear magnetic resonance spectra of human urine from normal and pathological states. *Analytical biochemistry* **1994**, *220* (2), 284-296.

97. Dong, J.; Cheng, K. K.; Xu, J.; Chen, Z.; Griffin, J. L. Group aggregating normalization method for the preprocessing of NMR-based metabolomic data. *Chemometrics and Intelligent Laboratory Systems* **2011**, *108* (2), 123-132.

98. Torgrip, R. J. O.; +àberg, K. M.; Alm, E.; Schuppe-Koistinen, I.; Lindberg, J. A note on normalization of biofluid 1D 1H-NMR data. *Metabolomics* **2008**, *4* (2), 114-121.

99. Kvalheim, O. M.; Brakstad, F.; Liang, Y. Preprocessing of analytical profiles in the presence of homoscedastic or heteroscedastic noise. *Analytical chemistry* **1994**, *66* (1), 43-51.

100. Sokal, R. R.; Rohlf, F. J. Assumptions of analysis of variance. *Biometry: The principles and practice of statistics in biological research* **1995**, 392-450.

101. Kvalheim, O. M.; Brakstad, F.; Liang, Y. Preprocessing of analytical profiles in the presence of homoscedastic or heteroscedastic noise. *Analytical chemistry* **1994**, *66* (1), 43-51.

102. Purohit, P. V.; Rocke, D. M.; Viant, M. R.; Woodruff, D. L. Discrimination models using variance-stabilizing transformation of metabolomic NMR data. *Omics: a journal of integrative biology* **2004**, *8* (2), 118-130.

103. Parsons, H. M.; Ludwig, C.; G++nther, U. L.; Viant, M. R. Improved classification accuracy in 1-and 2-dimensional NMR metabolomics data using the variance stabilising generalised logarithm transformation. *BMC bioinformatics* **2007**, *8* (1), 234.

104. van den Berg, R. A.; Hoefsloot, H. C.; Westerhuis, J. A.; Smilde, A. K.; van der Werf, M. t. J. Centering, scaling, and transformations: improving the biological information content of metabolomics data. *BMC genomics* **2006**, *7* (1), 1.

105. Wold, S. PLS for multivariate linear modeling. *Chemometric methods in molecular design* **1995**, *2*, 195.

106. Wehrens, R. *Chemometrics with R: multivariate data analysis in the natural sciences and life sciences;* Springer Science & Business Media: 2011.

107. Keun, H. C.; Ebbels, T. M.; Antti, H.; Bollard, M. E.; Beckonert, O.; Holmes, E.; Lindon, J. C.; Nicholson, J. K. Improved analysis of multivariate data by variable stability scaling: application to NMR-based metabolic profiling. *Analytica chimica acta* **2003**, *490* (1), 265-276.

108. Smilde, A. K.; van der Werf, M. t. J.; Bijlsma, S.; van der Werff-van der Vat, B. J.; Jellema, R. H. Fusion of mass spectrometry-based metabolomics data. *Analytical chemistry* **2005**, *77* (20), 6729-6736.

109. Huber, W.; Von Heydebreck, A.; S++ltmann, H.; Poustka, A.; Vingron, M. Variance stabilization applied to microarray data calibration and to the quantification of differential expression. *Bioinformatics* **2002**, *18* (suppl 1), S96-S104.

110. Xia, J.; Sinelnikov, I. V.; Han, B.; Wishart, D. S. MetaboAnalyst 3.0GÇömaking metabolomics more meaningful. *Nucleic acids research* **2015**, *43* (W1), W251-W257.

111. Hammer, O.; Harper, D. A. T.; Ryan, P. D. Paleontological Statistics Software: Package for Education and Data Analysis. *Palaeontologia Electronica* **2001**.

112. Worley, B.; Powers, R. Multivariate analysis in metabolomics. *Current Metabolomics* **2013**, *1* (1), 92-107.

113. Wold, S.; Esbensen, K.; Geladi, P. Principal component analysis. *Chemometrics and Intelligent Laboratory Systems* **1987**, *2* (1-3), 37-52.

114. Barker, M.; Rayens, W. Partial least squares for discrimination. *Journal of Chemometrics* **2003**, *17* (3), 166-173.

115. Worley, B.; Halouska, S.; Powers, R. Utilities for quantifying separation in PCA/PLS-DA scores plots. *Analytical biochemistry* **2013**, *433* (2), 102-104.

116. Wold, S. Chemometrics, why, what and where to next? *Journal of pharmaceutical and biomedical analysis* **1991**, *9* (8), 589-596.

117. Brereton, R. G. Consequences of sample size, variable selection, and model validation and optimisation, for predicting classification ability from analytical data. *TrAC Trends in Analytical Chemistry* **2006**, *25* (11), 1103-1111.

118. Daszykowski, M.; Wu, W.; Nicholls, A. W.; Ball, R. J.; Czekaj, T.; Walczak, B. Identifying potential biomarkers in LC-MS data. *Journal of Chemometrics* **2007**, *21* (7), 292-302.

119. Westerhuis, J. A.; Hoefsloot, H. C.; Smit, S.; Vis, D. J.; Smilde, A. K.; van Velzen, E. J.; van Duijnhoven, J. P.; van Dorsten, F. A. Assessment of PLSDA cross validation. *Metabolomics* **2008**, *4* (1), 81-89.

120. Xia, J.; Broadhurst, D. I.; Wilson, M.; Wishart, D. S. Translational biomarker discovery in clinical metabolomics: an introductory tutorial. *Metabolomics* **2013**, *9* (2), 280-299.

121. Xia, J.; Wishart, D. S. Metabolomic data processing, analysis, and interpretation using MetaboAnalyst. *Current protocols in bioinformatics* **2011**, 14-10.

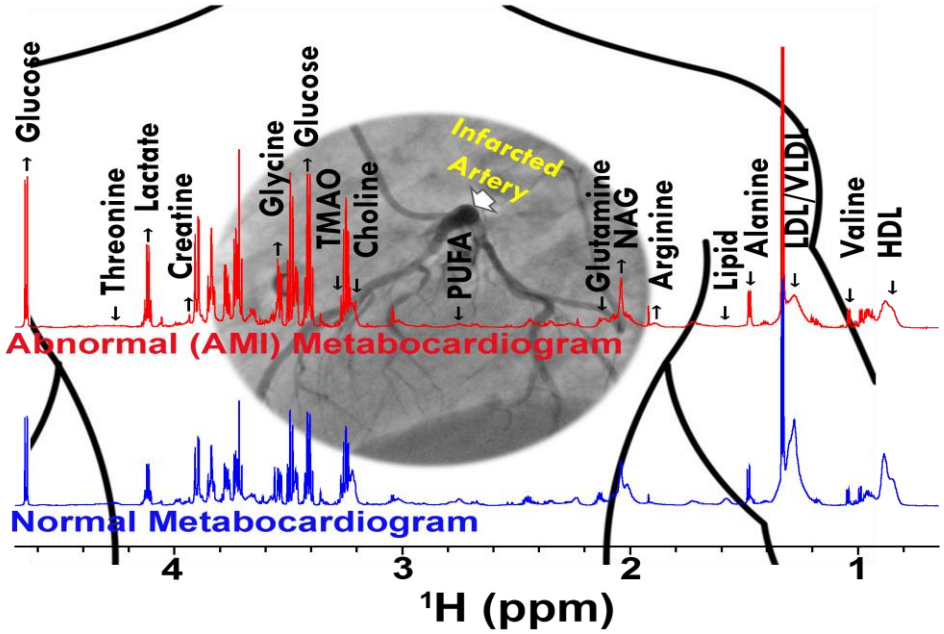
122. Xia, J.; Wishart, D. S. Using MetaboAnalyst 3.0 for Comprehensive Metabolomics Data Analysis. *Current protocols in bioinformatics* **2016**, 14-10.

# Chapter 4

## Serum Metabolic Signatures Hailing in Initial hours of Acute Myocardial Infarction elucidated by NMR-based Metabolomics


### Abstract:

Acute myocardial infarction (AMI) is a serious medical emergency and leading cause of cardiac-related deaths within first few hours from the onset of symptoms. The rapid detection of physiological transformations associated with AMI coupled with instant treatment to reset these changes and monitoring response to treatment can greatly decrease the mortality and morbidity of patients. In this study, metabolic profiling of sera obtained from 42 AMI patients (immediately after the myocardial infarction) and 38 age/sex matched normal controls were performed using high-resolution 1D $^{1}$H CPMG and diffusion edited NMR spectra. Multivariate and univariate statistical analysis were carried out to identify the AMI specific metabolic disturbances present in the first few hours of AMI and, therefore, the perturbed biochemical pathways in this condition. Our results revealed significant metabolic perturbations in AMI compared to control 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, glycero-phosphocholine, trimethylamine-N-oxide and lipids/fatty acids were decreased. Receiver operating curve characteristics (ROC) confirmed the robustness and validity of these metabolic markers. The resulted metabolic profiling provided new insights into the metabolic processes involved in acute myocardial infarction. Further, we foresee that these changes would aid rapid clinical evaluation of myocardial infarction in emergency and its timely management.

## Introduction:

Acute myocardial infarction (AMI) -also known as heart attack- is a catastrophic event 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.[1] The majority of AMI-related deaths and disability episodes are the result of inappropriate diagnosis or delayed treatment given to such patients presenting to the emergency departments with sudden acute chest pain.[2,3] Statistics has shown that death victims usually die in less than 6 hours after AMI occurs and the current diagnostic approaches are not sensitive enough to make such an early and rapid diagnosis of AMI.[4] Traditionally, the predictive diagnosis of AMI is based on 3 criteria[5]: (a) severe chest pain, (b) characteristic changes it produces on the electrocardiogram (ECG), imaging and morphological studies and (c) rise and fall of serological biomarkers such as troponin I and T, creatine kinase (CK), an isoenzyme of creatine kinase (CK-MB), myoglobin glutamic oxaloacetate transaminase (GOT), and lactate dehydrogenase (LDH).[6,7] However, these traditional indicators are not very successful in timely diagnosis of AMI owing to the procedural complexity and low sensitivity.[6,8] Further, the aforementioned serological markers lack sufficient specificity and therefore, are limited to provide a reliable diagnosis of AMI or predicting its risk. Particularly, the predictive diagnosis of AMI in patients with normal levels of serological biomarkers and electrocardiographic results and therefore suspected for acute myocardial ischemia (clinically manifested in the majority of AMI patients) is the most challenging task; the doctors face in emergency departments.[9] Of course, myocardial imaging is the gold standard to confirm the myocardial infarction. However, such an expensive facility is not widely available in all the hospitals, and even though if it is available, it is restricted in this acute setting due to lack of 24 hours accessibility,[10] involves shifting the patients from emergency to radiology department and the long time required to perform the imaging scans and subsequent evaluation by the radiologist. To date, several noninvasive biochemical assays are available in clinical settings which are used to predict risk factors and severity of cardiovascular diseases, such as lipid and lipoprotein properties, inflammation, and oxidative stress. However so far, no reliable diagnostic and prognostic biomarkers are available in clinics for rapid diagnosis of AMI or predicting its risk index, while unnecessary admissions to cardiac care units may waste financial and medical resources.[6] Therefore, there is an immense interest to improve the clinical diagnosis and management of AMI patients. To this end, we applied NMR-based metabolomics approach to investigate the altered metabolic profiles in sera obtained from AMI patients and sought to identify metabolic disturbances associated with myocardial infarctions to unravel the associated physiological effects and thus improve the treatment accordingly. These metabolic signatures perhaps conceived as surrogate markers of AMI for rapid diagnosis to evaluate metabolic modifications and monitoring after treatment has been initiated.

Metabolomics allows rapid identification of metabolic perturbations in biological systems and is routinely used to evaluate systemic responses to any subtle pathophysiological stimuli or stress.[11,12] NMR coupled with multivariate analysis is currently the technique of choice for rapid screening of metabolic alterations.[12-16] The approach has been demonstrated to differentiate ischemic from non-ischemic individuals under controlled conditions of ischemia.[16] Serum metabolic profiling in animal models of myocardial infarction have also demonstrated abnormalities of systemic metabolism;[17] however to date, there is no report focused on studying the changes in metabolism that occur in patients with AMI. We hypothesized that patient with AMI and hospitalized in an emergency have acute changes in their metabolism that can be identified using $^{1}$H-NMR based serum metabolomics. To prove this hypothesis, we studied the metabolic profiles of sera derived from AMI patients hospitalized in critical care immediately after the spontaneous acute chest pain and compared with those of normal controls using multivariate statistical approach. The metabolic alterations also provided useful insights into the disease processes trailing the AMI attack. Overall, the study establishes the metabolic patterns of AMI and the information can be used to aid its clinical management under emergency settings.

## Materials and Method:

### *Recruitment of Subjects:*

The study protocol was approved by the institutional research and ethical committee, King George's Medical University, Lucknow, Uttar Pradesh, India (Reference Number: **57 E.C.M. IIB/P7**). The serum samples used in this study were obtained from AMI patients, admitted after acute chest pain in the emergency department of King George's Medical University (KGMU), Lucknow. The enrollment of subjects was carried out according to the norms of World Medical Association (WMA) Declaration of Helsinki. Blood samples were collected from the subjects presenting with acute chest pain admitted to the intensive care unit within first few hours (6-8 hrs), were only recruited in the study. A written informed consent was obtained from the guardians/kin of the patients. General and relevant clinical details were collected for all the patients such as gender, age, diabetes, hypertension, smoking and alcohol usages in a custom-designed questionnaire. Recruited patients in the AMI group were those with the established diagnosis of AMI based on clinical, electrocardiographic and elevated serological markers. Electrocardiograms were performed upon hospital admission, after the initial treatment. Patients were excluded with at least one of the following conditions: (a) the patients with cardio embolic stroke, cerebral venous sinus thrombosis, CNS vasculitis and hemorrhage due to trauma, tumor, vascular malformation and coagulopathy, (b) subjects having bacterial and viral infections, inflammatory diseases, thyroid, liver or kidney diseases and suffering with any kind of cancer, and (c) pregnant women. The normal controls were recruited from the random population after

taking an informed consent. In each case, it was confirmed that the normal controls are normotensive with no cardiovascular abnormalities and satisfying the above exclusion criteria.

### *Sample Preparation:*

In each case, the 3.0 ml of blood sample was drawn and processed to extract the serum as per the established protocol.[18] The extracted serum was transferred into a sterile 1.5 ml microcentrifuge tube (MCT) and stored at −80 °C immediately after the processing until the NMR experiments were performed. All serum samples were thawed and centrifuged at 10,000 rpm for 5 minutes to remove precipitates just before acquiring the NMR data. A total 400 µl of sample was used in 5 mm NMR tubes (Wilmad Glass, USA) for data acquisition:200 µl of serum, final volume adjusted by adding 200 µl of 0.9% saline sodium phosphate buffer[19] of strength 20 mM and pH 7.4 prepared in $D_2O$ and adding a co-axial insert containing the known concentration of TSP (Sodium salt of 3-trimethylsilyl-(2,2,3,3-d4)-propionic acid) i.e. 0.1mM was used as external standard reference to aid metabolite quantification for NMR experiment. Deuterium oxide ($D_2O$; as a co-solvent and to provide a deuterium field/frequency lock) and the sodium salt of trimethylsilyl propionic acid-d4 (TSP) used for NMR experiments were purchased from Sigma-Aldrich (Rhode Island, USA).

### NMR Measurements:

All NMR spectra were recorded at 298 K on Bruker Biospin Avance-III 800 MHz NMR spectrometer operating at a proton frequency of 800.21 MHz, equipped with CryoProbe and an actively shielded gradient unit with a maximum gradient strength output of 53 G/cm. The NMR data were processed in Topspin-2.1 (Bruker NMR data Processing Software). For each serum sample, two types of 1D $^1$H NMR spectra were recorded: (a) transverse relaxation-edited CPMG (Carr–Purcell–Meiboom–Gill) spectra[20] and (b) diffusion-edited bipolar pulse pair longitudinal eddy current delay (BPP-LED) spectra.[21] The 1D $^1$H CPMG NMR spectra were recorded using the standard Bruker's pulse program library sequence(cpmgpr1d) with pre-saturation of the water peak through irradiating it continuously during the recycle delay (RD) of 5 sec. Each spectrum consisted of the accumulation of 128 scans and lasted for approximately 15 minutes. To remove broad signals from triglycerides, proteins, cholesterols, and phospholipids, a total spin–spin relaxation time of 60 ms (n=300 and $2\tau$=200µs) was applied. Each FID (free induction decay) was zero filled and Fourier-transformed to 64 K data points following manual phase and baseline-correction using Bruker NMR data Processing Software Topspin-v2.1. A line broadening factor of 0.3 Hz and a sine–bell apodization function was applied to FIDs before Fourier Transformation. After FT, the chemical shifts were referenced internally to methyl peak of lactate (at $\delta$=1.33 ppm). To obtain spectra with signals only from lipids or lipoproteins, the diffusion-edited 1D$^1$H NMR spectra were recorded using the standard Bruker's pulse program library

parameters (ledbpgp2s1d i.e. 1-dimensional longitudinal eddy current bipolar gradient pulse presaturation with 2 stimulated echoes). The spectra were measured using sine shaped gradient pulses of strength 30% and a duration of 1.5 ms followed by a delay of 200 µs to allow for the decay of eddy current, a diffusion time of 120 ms and an eddy current decay time of 5 ms. The relaxation delay was 4 seconds, and water peak irradiation was applied during the recycle delay and the delay after the first BPP. A line broadening factor of 1 Hz was applied to FIDs before Fourier Transformation. Spectra were acquired with 128 scans, then zero filled and Fourier-transformed to 64 K data points. All recorded spectra were, visually inspected for acceptability and subjected to multivariate statistical analysis to identify the altered metabolic patterns.

## Identification of Metabolite Peaks:

Chemical shifts were identified and assigned as far as possible, by comparing them with the chemical shifts available with the open access software program ***MetaboMiner***[22] with tolerances of 0.05 ppm ($^{1}H$) and 0.1 ppm ($^{13}C$). The metabolite peaks in one-dimensional $^{1}H$ CPMG NMR spectra were identified and assigned as far as possible, by comparing them with the chemical shifts available with the software **Chenomx** (NMR Suite, v8.1, Chenomx Inc., Edmonton, Canada). The assigned resonances of the metabolite peaks were validated using: (a) previously reported NMR assignments of metabolites, data obtained from BMRB database (Biological Magnetic Resonance Data Bank) and HMDB (The Human Metabolome Database)[23-26] and (b) Assigned resonances in two-dimensional spectra. For unambiguous assignment of various peaks in these spectra, two-dimensional (2D) $^{1}H$–$^{1}H$ TOCSY (Total Correlation Spectroscopy) and $^{1}H$-$^{13}C$ HSQC (Heteronuclear Single Quantum correlation) NMR spectra were also acquired at 298 K for some of the serum samples using the parameters as described previously.[27] The assignments of peaks from lipid moieties were obtained based on previous literature reports.[25,26]

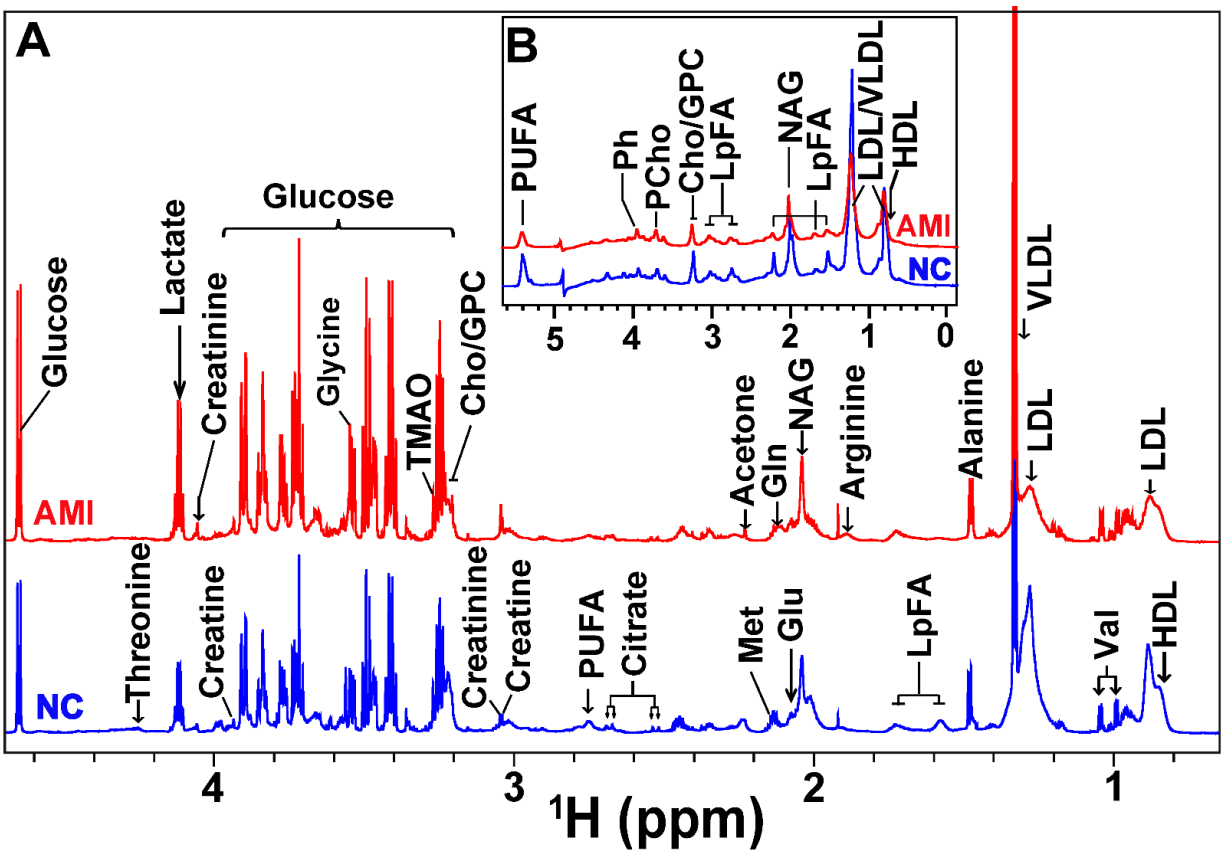



**Figure 1:** Stack plot of cumulative 1D $^{1}$H CPMG NMR spectra (ranging from δ0.7–δ4.6 ppm) **(A)** and diffusion-edited NMR spectra (ranging from δ0.0–δ5.6 ppm) **(B)** obtained for sera of AMI patients (red) and normal controls (blue). The metabolite peaks of discriminatory potential assigned. Diffusion edited spectra were measured for exclusive profiling of lipid and lipoprotein metabolites through attenuating the peaks from fast-diffusing, low molecular weight metabolites. Key acronyms are: HDL; high density lipoproteins; LDL: low density lipoproteins; VLDL: very-low-density lipoproteins; Met: Methionine; NAG: N-acetyl glycoproteins, Glu: Glutamate; Gln: Glutamine; Cho: Choline-$N^{+}(CH_3)_3$; GPC: glycerophosphocholine; PCho: Phosphatidylcholine, Ph: phospholipid; LpFA: Lipid-bound fatty acid chains; PUFA: polyunsaturated fatty acids; TMAO: trimethyl-amine oxide; Gly: glycine.

## Data Reduction:

Before multivariate data analysis, all the NMR spectra were manually phased and baseline corrected. The CPMG (δ0.5−8.5 ppm) and diffusion edited (δ0.5–6.0 ppm) spectra were binned and automatically integrated using AMIX package (Version 3.8.7, Bruker, BioSpin). The region δ (4.7-5.5) distorted due to water suppression and lipid dominated regions δ(0.5-0.9), δ(1.95-2.06) and δ (1.2- 1.39) ppm were excluded from the CPMG data set to avoid the effects of imperfect water suppression and lipid attenuation; whereas from the diffusion edited spectra, only the region δ(4.7-5.0) distorted due to residual water signal was removed. Finally, the selected regions were reduced to spectral bins of δ0.01 ppm. Subsequently, the spectral bins were integrated and normalized to the sum of all integral regions for each spectrum to

compensate for the differences in concentration of metabolites among individual serum samples. The resultant datasets were finally used for multivariate analysis using the open access web-based metabolomics data processing tool, named **MetaboAnalyst.**[28]

## Multivariate Pattern recognition Analysis:

The most commonly used pattern recognition analysis tools for dealing with multivariate data in metabolomics studies of biofluids are principal component analysis (PCA) and partial least-squares discriminant analysis (PLS-DA)[29]. These pattern recognition tools reduce hundreds of variables into few linearly uncorrelated components, which facilitate us to interpret a vast amount of data easily and rapidly[30]. The scores plots display discrimination patterns, whereas loading plots reveal the variables that are responsible for influencing the discrimination pattern. Of two, the PCA is the unsupervised approach and provides an initial overview of the grouping trend (i.e. intrinsic clustering) and outliers within the data set, whereas, the PLS-DA is a supervised approach (requires class information) and is used to classify/discriminate the two or more groups[30]. To evaluate the CPMG and DE discrimination patterns between AMI and normal control groups, we used the statistical analysis module of **MetaboAnalyst**[28] (a freely available, user-friendly, web-based analytical platform for high-throughput metabolomics studies from the University of Alberta, Canada[29]). The binned spectral data both from CPMG and DE experiments were Pareto scaled and then, subjected to unsupervised PCA analysis for an initial overview and identifying the outliers. The outliers were excluded from both the data sets and the resulted data sets were then subjected to supervised PLS-DA to differentiate the groups and further to identify the marker metabolites that can differentiate the AMI group from the control group. PLS-DA tends to-over fit the data and therefore before using this as a diagnostic model needs to be rigorously validated to see whether the separation is statistically significant or is due to random noise. To avoid the over-fitting of the PLS-DA model, 10-fold cross-validation algorithm was used to evaluate 100% classification accuracy based on top 5 latent variables. The goodness of model and the model robustness were assessed by the cross-validation parameters, $R^2$ and $Q^2$, respectively. $R^2$ is the fraction of variance explained by a component, and cross validation of this component provides $Q^2$, which describes the fraction of the total variation predicted by a component. The value of $Q^2$ ranges from 0 to 1 and typically a $Q^2$ value greater than 0.5 is considered a good model, and those with $Q^2$ values over 0.7 are robust. Interpretation of PLS-DA model was based on the score plot, regression coefficients and the variable importance in the projection plot (VIP). The coefficient importance is based on the weighted sum of PLS-regression scores; whereas, the VIP score represents a weighted sum of squares of the PLS loadings and indicates the importance of the variable to the whole model and the corresponding coefficient values attribute its discriminatory potential. Generally, the variables (or metabolite peaks) with high VIP and coefficient scores indicate that it is important for class discrimination. The robustness of the PLS-DA model for discriminating the AMI from control cohorts was further verified using receiver

operating characteristic (ROC) analysis. Unsupervised hierarchical clustering in the form of heat map was used to assess, how similar or different the AMI samples are compared to normal control samples based on their metabolite profiles. The boxplot representation (evaluated through univariate analysis) was, used to visualize the variation in the levels of significantly altered metabolites in AMI patients identified in the multivariate analysis.

### Pathway Analysis

A list of serum metabolites found significantly altered in AMI patients (identified through PLS-DA analysis) was subjected to pathway analysis and Metabolite Set Enrichment analysis (MSEA) which is the metabolomics version of Gene Set Enrichment Analysis (GSEA) approach (available in MetaboAnalyst[28,31]). The MSEA function in MetaboAnalyst enables identification of altered metabolic pathways from its extensive HMDB-derived collection of more than 71 pathways and metabolite libraries.[28] The lipid and membrane metabolites such as LDL, VLDL, HDL, and N-acetyl glycoprotein are not recognized by the program; thus were not included in this analysis. The final list of altered metabolites was uploaded and analyzed by Over-Representation Analysis (ORA) in MetaboAnalyst.[28] One-tailed *p*-values are provided after adjusting for multiple testing. The output of this program will mark a metabolic pathway as significant if the input list contains significantly more compounds involved in that pathway than would be expected by random chance. The matched pathways are shown according to their p-values from the pathway enrichment analysis (vertical axis or y-axis, the intensity of color) and pathway impact values from pathway topology analysis (horizontal axis or x-axis, the size of circle), with the most impacted pathways colored in red.

## Results

### Clinical and demographic Details:

The clinical and demographic characteristics of the subjects are summarized in **Table 1**. Based on the inclusion and exclusion criteria, a total 80 subjects were involved in this study, 42 forming the disease group (AMI) and 38 forming the normal control group. The mean age of the AMI and NC groups was 49.83±14.35 and 49.05±3.26 years, respectively. The angiograms of patients with AMI were suggestive of significant stenoses (>50% luminal diameter stenosis). The serum samples of patients those diagnosed specifically with AMI were finally used in this study to profile the metabolic differences compared to normal control (NC).

**Table 1:** Clinical and demographic characteristics of AMI patient and control cohorts.

| Variables/Parameters | | Case (n=42) | Normal Control (n=38) |
|---|---|---|---|
| Mean Age (Range) | | 49.83±14.35 (range ) | 46.05±3.26 (range ) |
| Sex: M/F | | 29/13 | 26/12 |
| Cardiac risk factors | Smoking | 25 (59.52%) | 10 (26.3) |
| | Diabetes | 21 (50.00%) | 2 (10.00%) |
| | Hypertension | 25 (59.52%) | 16 (5.00%) |
| | Previous history of CAD | 1 (2.38%) | 0% |
| | Family history of CAD | 16 (38.09) | 5 (13.00%) |
| Heart rate | | 114.61±44.32 | -- |
| Body mass index (#BMI) | | 29.16±3.87 | -- |
| In hospital complication after acute myocardial infarction | Postinfarction angina | 11 (26.19%) | |
| | Reinfarction | 1 (2.38%) | -- |
| | Complete heart blockage (CHB) | 14 (33.33 %) | -- |
| | Sever arrhythmias | 6 (14.28 %) | -- |
| | Other complications | 3 (7.14%) | -- |
| | No complications | 7 (16.66%) | -- |
| Culprit artery identified by percutaneous coronary intervention (PCI) | LAD | 17 (54.83 %) | -- |
| | CRX | 5 (16.12 %) | -- |
| | RCA | 8 (25.81 %) | -- |
| | Others | 1 (3.22 %) | -- |

LAD- Left anterior descending artery, CRX- Circumflex artery, RCA- Right coronary artery, CAD- coronary artery disease. #Patients having BMI ≥25 kg/m$^2$ considered as obese.

## Metabolic alterations in AMI:

In the present study, we discriminated 42 AMI patients from 38 age, and sex matched normal controls, and established the serum metabolic patterns of AMI. The study subjects were recruited judiciously to minimize the differences due to confounding variables as evident from **Table 1**. The representative 1D $^1$H CPMG and diffusion edited NMR spectra of serum samples (one from each group) are shown in **Figure 1A and 1B** respectively. The visual comparison of the $^1$H-NMR spectra failed to identify any major differences between AMI patients and control groups. Therefore, the NMR spectra were subjected to multivariate data analysis to identify AMI induced serum metabolic changes. First, the $^1$H CPMG and diffusion edited NMR data were analyzed using unsupervised PCA method, for initial grouping trends and class separation, as shown in **Figure 2**. Further, we performed supervised clustering method PLS-DA to investigate subtle metabolic

differences among the groups. The parameters used to assess the quality of each model, including explained variation $R^2$ and the predictive capability, $Q^2$, are displayed in their respective score-plots in **Figure 2**. The model quality parameters $R^2$ and $Q^2$, were significantly higher ($R^2$, $Q^2>0.5$), indicating that the PLS-DA models (constructed from CPMG and diffusion-edited spectra) possessed satisfactory fit with good predictive power. Further, the model validation and significance of class discrimination were assessed using permutation test statistics and ROC analysis as depicted in **Figure 3**. The two-dimensional PLS-DA score plots derived from 1D $^1H$ CPMG and diffusion edited NMR spectra **(Figure 2)** showed that the AMI and control groups are well clustered and separated from each other indicating that the biochemical composition profiles of serum metabolites in AMI patients are significantly different from normal controls, small metabolites relatively more perturbed. The metabolites responsible for the discrimination of two cohorts were identified using the VIP and coefficient scores. Overall, we identified 30 metabolites significantly perturbed in sera of AMI patients (VIP score ≥1 and coefficient score ≥30) as enlisted in **Table 2**.

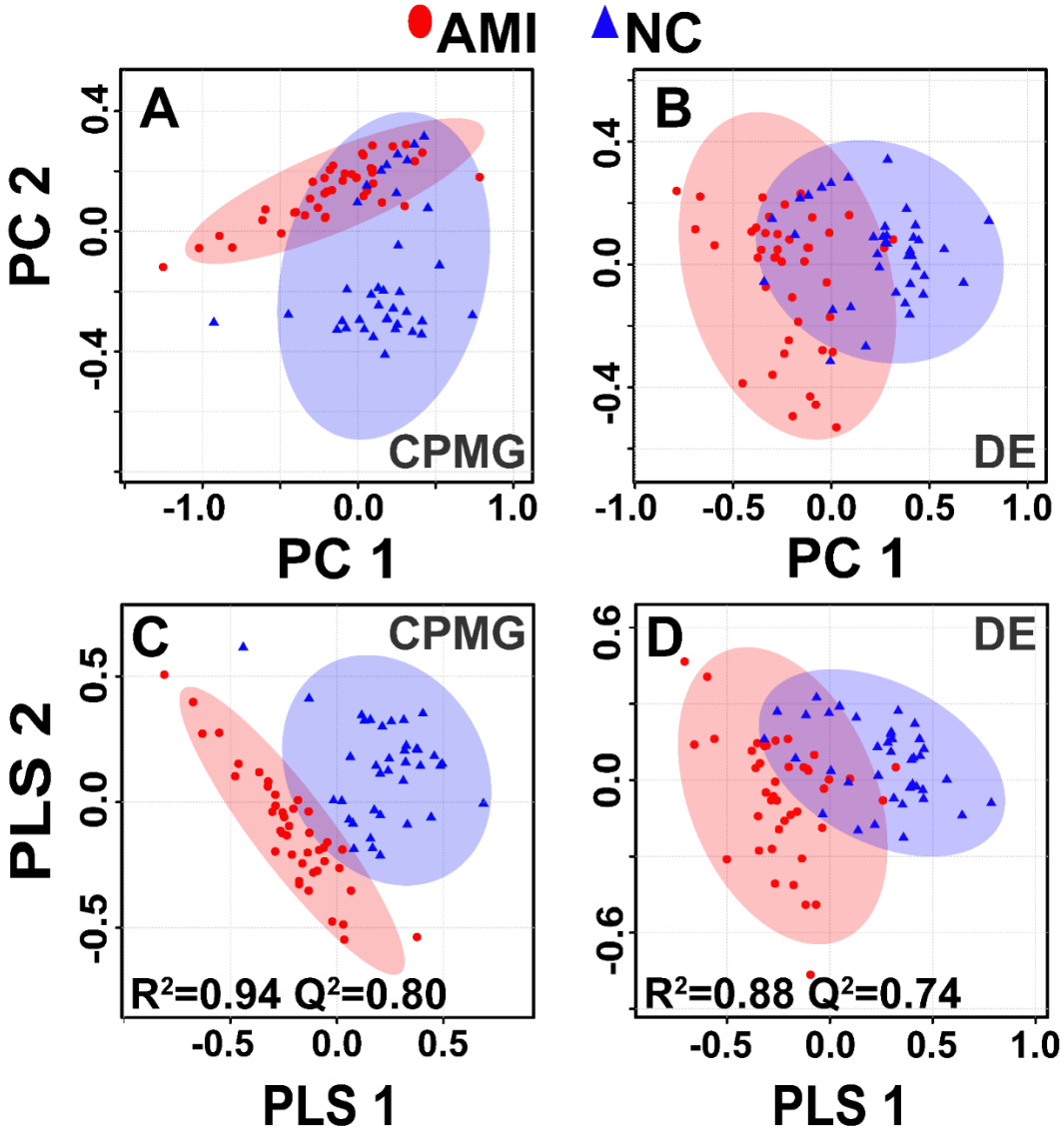


**Figure 2:** The two-dimensional PCA **(A-C)** and PLS-DA **(B-D)** score plots derived, respectively, from CPMG and diffusion edited (DE) spectra showing clear statistical separation between AMI (represented by red circles) and normal control (NC) samples (represented by blue triangles). Each circle and triangle represent one subject. The validation parameters ($R^2$ and $Q^2$) corresponding to each PLS-DA model are also displayed in their respective score plots. The semi-transparent red and blue ovals represent the 95% confidence interval.

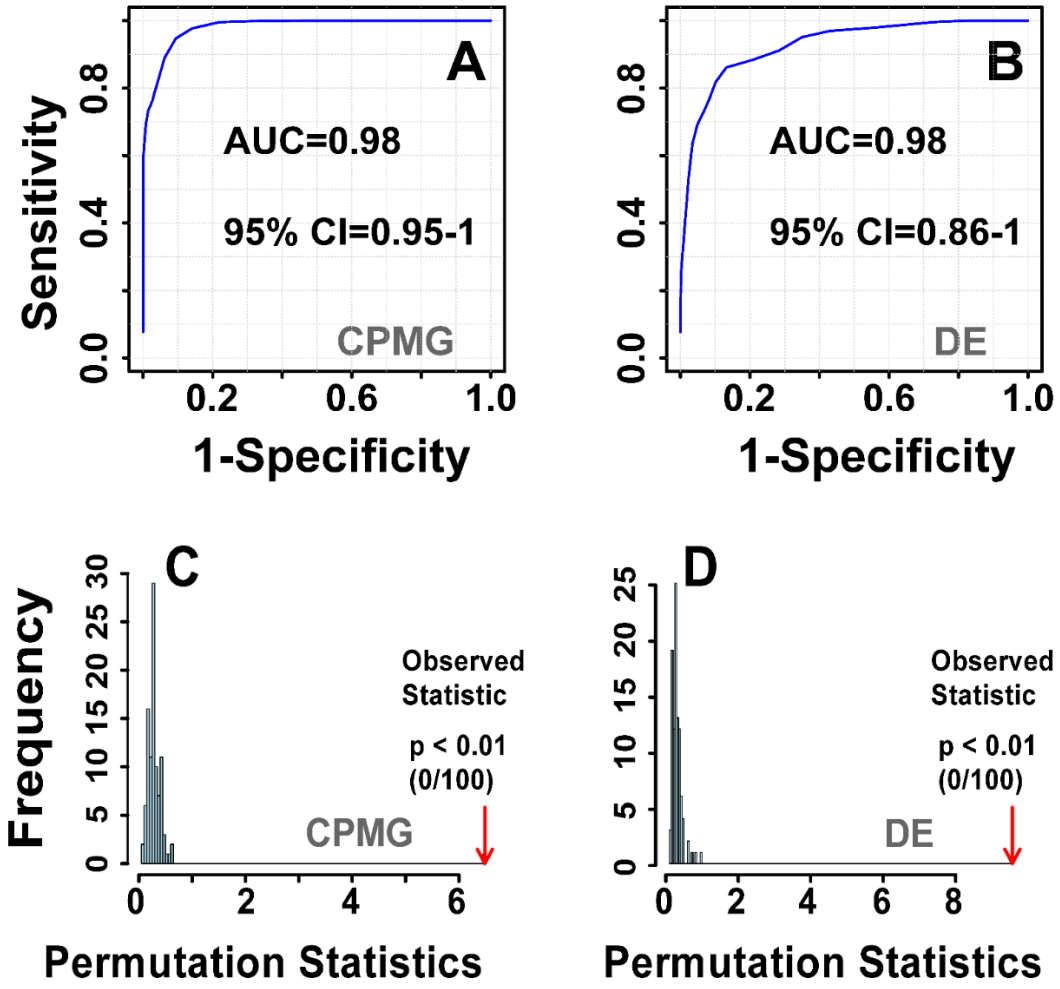



**Figure 3:** The diagnostic potential of each model as shown in **(A)** and **(B)** obtained from ROC analysis and (**C**) and **(D)** permutation analysis.

**Table 2:** Details of the metabolites best describing the variation in AMI with respect to the normal control (NC) group. The arrows up and down represent increased or decreased levels of metabolites respectively.

| Serial Number | NMR Spectra | Metabolite (δ, ppm) | VIP Score | Serum Level ↑/↓ | AUROC |
|---|---|---|---|---|---|
| 1 | CPMG | Valine (1.045) | $1.97^{\Psi}$ | ↓*** | 0.82 |
| 2 | CPMG | Alanine (1.485) | $3.01^{\Psi\Psi\Psi}$ | ↓*** | 0.80 |
| 3 | CPMG | LpFA (1.575) | $3.62^{\Psi\Psi\Psi}$ | ↓*** | 0.94 |
| 4 | CPMG | Arginine (1.885) | $1.78^{\Psi\Psi}$ | ↑*** | 0.84 |
| 5 | CPMG | Glutamate (2.065) | $2.38^{\Psi\Psi}$ | ↓*** | 0.87 |
| 6 | CPMG | Glutamine (2.135) | $3.25^{\Psi\Psi\Psi}$ | ↓*** | 0.79 |
| 7 | CPMG | Methionine (2.145) | $2.96^{\Psi\Psi}$ | ↓*** | 0.81 |
| 8 | CPMG | Acetone (2.225) | $2.23^{\Psi\Psi}$ | ↓*** | 0.88 |
| 9 | CPMG | Citrate# (2.515) | $1.69^{\Psi\Psi\Psi}$ | ↓*** | 0.94 |
| 10 | CPMG | PUFA (2.755) | $3.01^{\Psi\Psi}$ | ↓*** | 0.95 |
| 11 | CPMG | Choline (3.215) | $5.12^{\Psi\Psi\Psi}$ | ↓*** | 0.90 |
| 12 | CPMG | GPC (3.225) | $3.40^{\Psi\Psi}$ | ↓*** | 0.62 |
| 13 | CPMG | TMAO (3.275) | $2.46^{\Psi\Psi}$ | ↓*** | 0.89 |
| 14 | CPMG | Glycine (3.555) | $2.37^{\Psi\Psi\Psi}$ | ↑*** | 0.74 |
| 15 | CPMG | Glucose (3.725) | $4.81^{\Psi\Psi\Psi}$ | ↑*** | 0.87 |
| 16 | CPMG | Creatine (3.935) | $1.42^{\Psi\Psi}$ | ↑*** | 0.76 |
| 17 | CPMG | Creatinine (4.055) | $1.30^{\Psi\Psi}$ | ↑*** | 0.67 |
| 18 | CPMG | Lactate (4.115) | $1.78^{\Psi\Psi}$ | ↑ | 0.62 |
| 19 | CPMG | Threonine (4.255) | $1.74^{\Psi}$ | ↓*** | 0.90 |
| 20 | CPMG | Tyrosine (7.175) | $1.14^{\Psi}$ | ↑*** | 0.84 |
| 21 | CPMG | Phenylalanine (7.305) | $1.24^{\Psi}$ | ↑*** | 0.95 |
| 22 | DE | HDL (0.855) | $2.17^{\Psi\Psi}$ | ↓*** | 0.89 |
| 23 | DE | LDL (1.245) | $3.28^{\Psi\Psi\Psi}$ | ↓** | 0.90 |
| 24 | DE | LDL/VLDL (1.275) | $5.22^{\Psi\Psi\Psi}$ | ↓** | 0.88 |
| 25 | DE | LpFA(1.575) | 1.396 | ↓*** | 0.92 |
| 26 | DE | NAG (2.055) | $2.77^{\Psi\Psi}$ | ↑*** | 0.90 |
| 27 | DE | Choline (3.215) | $1.78^{\Psi\Psi}$ | ↓*** | 0.87 |
| 28 | DE | Phosphatidyl choline (3.675) | $1.95^{\Psi}$ | ↑*** | 0.93 |
| 29 | DE | Ph (3.915) | 1.76 | ↑*** | 0.93 |
| 30 | DE | PUFAs (5.295) | $1.94^{\Psi}$ | ↓*** | 0.89 |

**Note:** The discriminatory metabolite entities were identified based on the criterion if their VIP ≥1.0 and validated further using p-value and AUROC analysis. *represents a statistically significant difference (p-value <0.05), ** represents p-value <0.001, *** p<0.0001), and $^{\Psi}$ highlights coefficient score (since the number of variables used in this study are more than 100, we used a more relaxed coefficient cutoff of around 30% ($^{\Psi}$, $^{\Psi\Psi}$, and $^{\Psi\Psi\Psi}$ represent cut-offs, respectively, 30-40%, 40-60% and more than 60%). The abbreviations PUFA: Polyunsaturated Fatty Acids, GPC: Glycerophosphocholine; TMAO: trimethyl-amine oxide; HDL: High-density lipoprotein; LDL: low-density lipoprotein; VLDL: very low-density lipoprotein, NAG: N-acetyl signal from glycoproteins; Ph: Phospholipid, LpFA: Lipid bound fatty acid chains.

These discriminatory metabolite entities were used to construct the heatmaps, commonly used for unsupervised clustering (**Figure 4**) which clearly shows that AMI group is visually distinguishable from normal control group based on these significant metabolites (up-regulated and down-regulated metabolites are shown in red and cyan color, respectively). The combination of altered metabolites can provide an indication of metabolic pathways more likely associated with the metabolic alterations induced after AMI. Therefore, we performed pathway analysis and metabolite set enrichment analysis (MSEA) in MetaboAnalyst[29] to establish which pathways are affected in AMI patients. The resulted summary of pathway analysis is shown in **Figure 5**. Mainly, five metabolic pathways of importance (protein biosynthesis, amino-acid metabolism, glucose-energy metabolism, lipid metabolism and choline metabolism) were, found to be disturbed in acute myocardial infarction. Although MetaboAnalyst pathway analysis tool is extremely valuable, to do an overall analysis of metabolic data, the metabolic pathways associated with the identified combinations of metabolites are biased owing to limited metabolites identified by NMR in the serum.

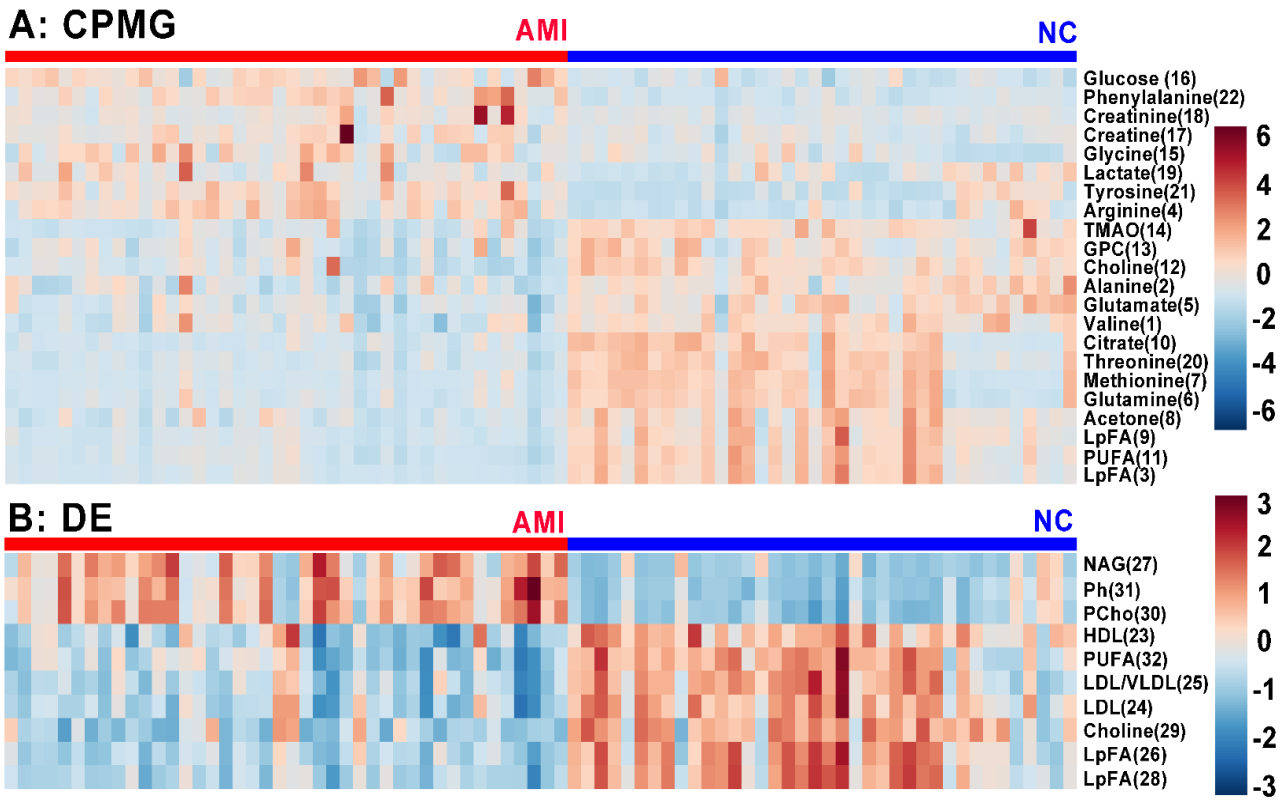



**Figure 4:** Heat maps showing z-scores of identified 30 statistically significant metabolite entities altered in AMI patients compared to normal controls. The numbers used within bracket correspond to the metabolite identity as shown in **Table 2.** Here, **(A)** and **(B)** are the heat maps derived from discriminatory metabolite entities identified, respectively, in CPMG and diffusion edited (DE) spectra. The red and cyan here signify, respectively, elevation and reduction in metabolite concentration in AMI patients. The color key indicates the metabolite expression level, values (fold change): dark blue: lowest; dark red: highest.

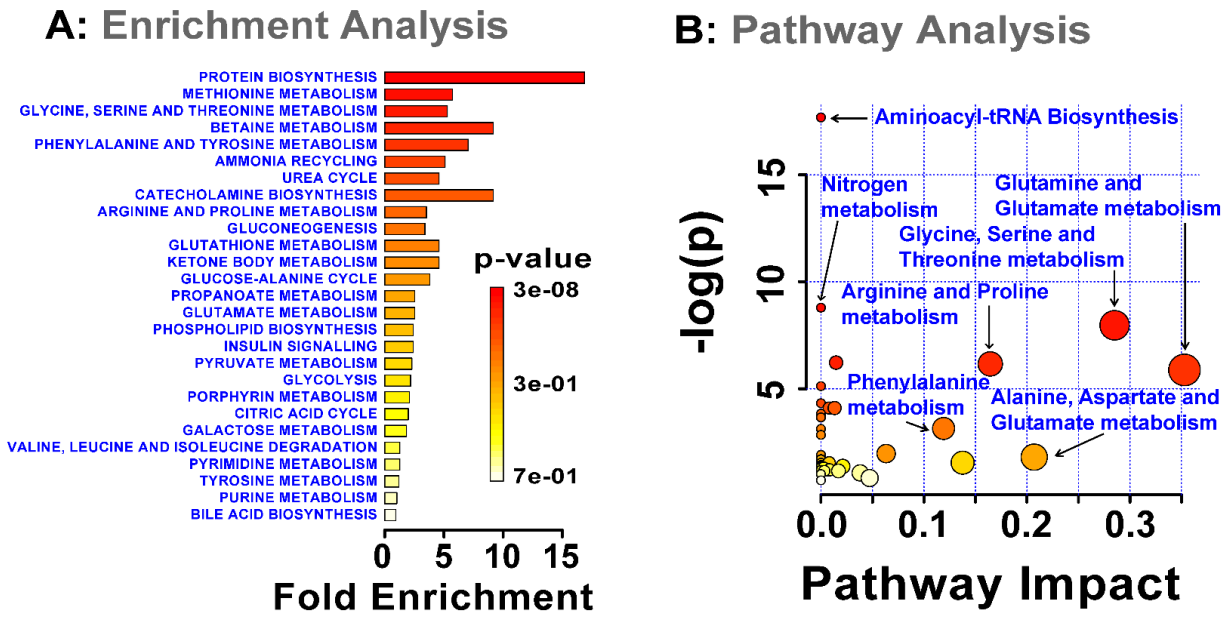



**Figure 5: (A)** Summary of metabolic pathway enrichment analysis performed in MetaboAnalyst (Version 3.0, URL: http://www.metaboanalyst.ca) using metabolite entities found significantly altered in AMI patients compared to normal control (NC) (as shown in **Table 2** excluding the lipids and Glycoproteins). **(B)** The list of these metabolites was further used as an input for pathway impact analysis in Metaboanalyst which is based on the Over Representation Analysis (ORA) algorithm and implemented using the hypergeometric test to evaluate over representation of a particular metabolite set; provided were fold-enrichment values and one-tailed p-values corrected for multiple testing. The most significant *p*-values are in the red while the least significant are in yellow and white. Pathway analysis is showing altered metabolic pathways.

To evaluate clinical utility, the quantitative difference in the metabolite concentration was assessed using a non-parametric Mann–Whitney test, and all the 30 metabolite entities were found to be statistically significant (with p-value $< 0.05$). The discriminatory metabolites were mainly related to lipid, amino acid, glucose, and energy metabolism. Further receiver's operating characteristic (ROC) curves analysis was performed for the significant metabolite markers to evaluate their predictive power or diagnostic accuracy. The area under the ROC curve gives of discriminatory ability (0.5=no discrimination; 1=perfect discrimination). The largest and smallest resulting AUC values range from 0.95 to 0.62 **(Table 2)** which indicated that these metabolites could be potential biomarkers for clinical evaluation and surveillance (through parallel assessment of a wide range of metabolites) of such patients. Representative ROC curves for some of the serum metabolites significantly altered in AMI patients and their respective box plot (drawn from the Univariate analysis) are shown in **Figure 6-7.**

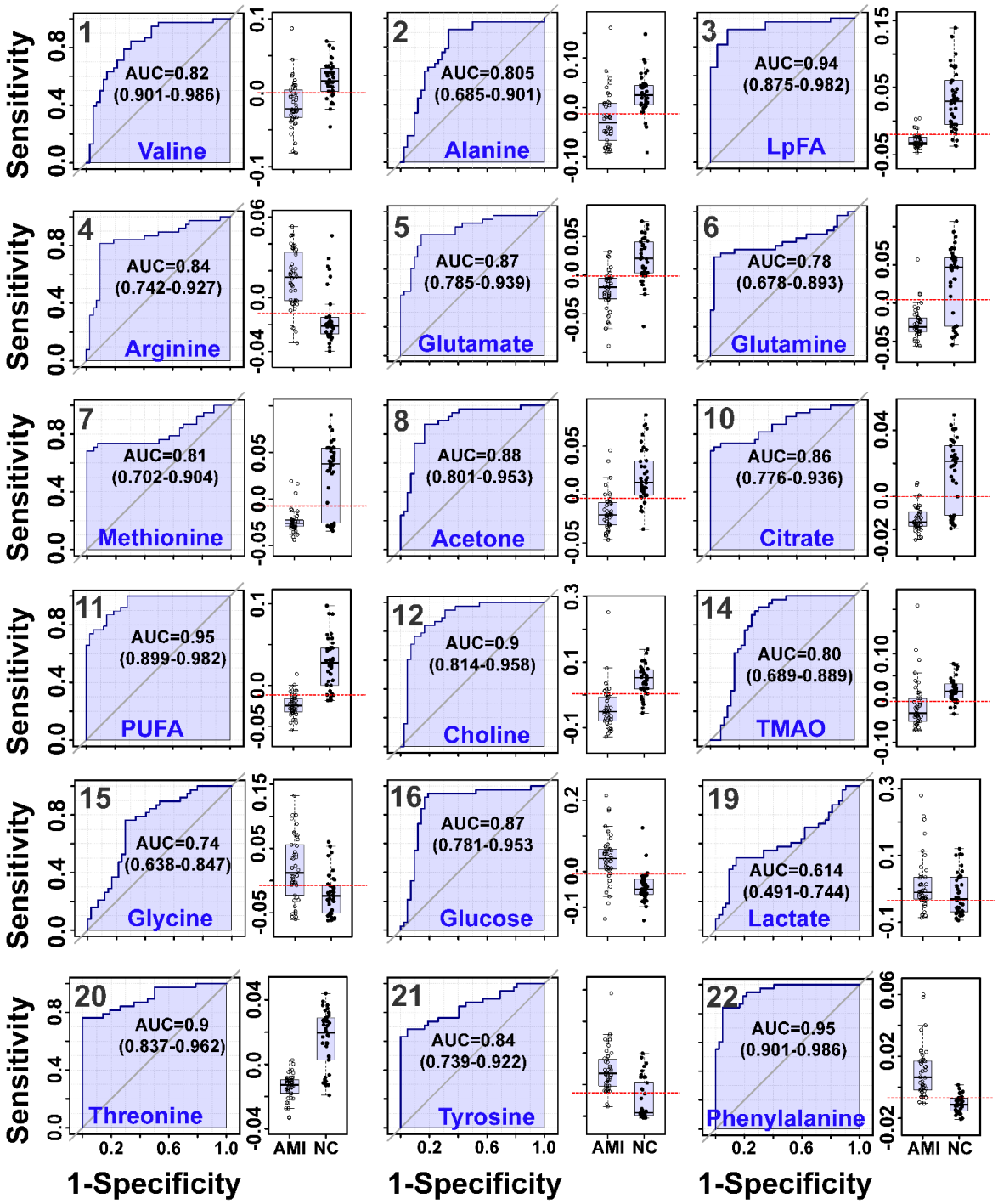



**Figure 6:** Receiver operating characteristic (ROC) curves of the metabolites discriminating AMI patients from normal controls (as listed in **Table 2**) along with their respective box plots derived from CPMG NMR data set. For each box plot, boxes denote interquartile ranges, lines denote medians, and whiskers denote 10$^{th}$ and 90$^{th}$ percentiles. The ROC plot of each metabolite entity contains the AUC (i.e. area under the ROC curve) value highlighting its discriminant potential.

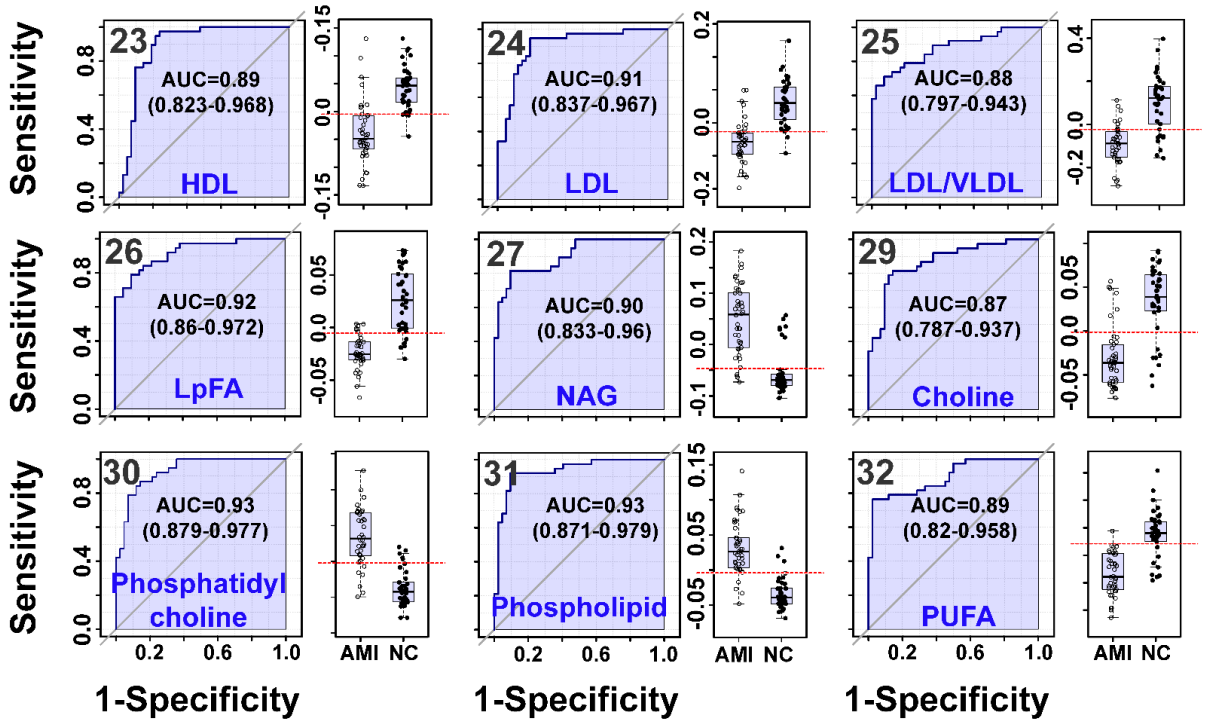



**Figure 7:** Receiver operating characteristic (ROC) curves of the metabolites discriminating AMI patients from normal controls (as listed in **Table 2**) along with their respective box plots derived from Diffusion edited NMR data set. For each box plot, boxes denote interquartile ranges, lines denote medians, and whiskers denote 10$^{th}$ and 90$^{th}$ percentiles. The ROC plot of each metabolite entity contains the AUC (i.e. area under the ROC curve) value highlighting its discriminant potential.

## Discussion

Acute myocardial infarction (AMI) is the irreversible death (necrosis) of heart muscle secondary to prolonged lack of oxygen supply (ischemia). The blood serum metabolome –which is directly influenced by the normal/abnormal functioning of the cardiovascular system- therefore, can provide a wealth of information about early hour's pathophysiology of AMI. Further, the information may help to define an appropriate treatment during an emergency to minimize the post-attack mortalities or disabilities. Such clinical implications derived our interest to study the serum metabolomics of AMI patients for identifying the metabolic patterns specific to AMI. The study clearly demonstrates the sera of AMI patients are present with significant metabolic alterations compared to normal controls. The potential discriminatory metabolites altered in the sera of AMI patients are enlisted in **Table 2.** The metabolic pathways that may be more relevant in the context of AMI attack are shown in **Figure 5.** The perturbed metabolites include: (a) intermediates of tricarboxylic acid (TCA) cycle and products of glycolysis and energy metabolism (such as glucose, lactate, citrate, creatine, etc.), (b) amino acids (alanine, valine, glycine, glutamate, glutamine, methionine, phenylalanine and tyrosine) and (c) molecules related to lipid and membrane metabolism like lipoproteins, polyunsaturated fatty acids, choline/GPC,

phospholipids and N-acetyl-glycoproteins (NAGs). The majority of these metabolites are intermediates of glucose-energy, amino acid metabolism and lipid metabolism as depicted schematically in **Figure 8.** The implications of these metabolic alterations in the pathophysiology of AMI with previous concordant and discordant reports, if any, have been discussed below in detail.

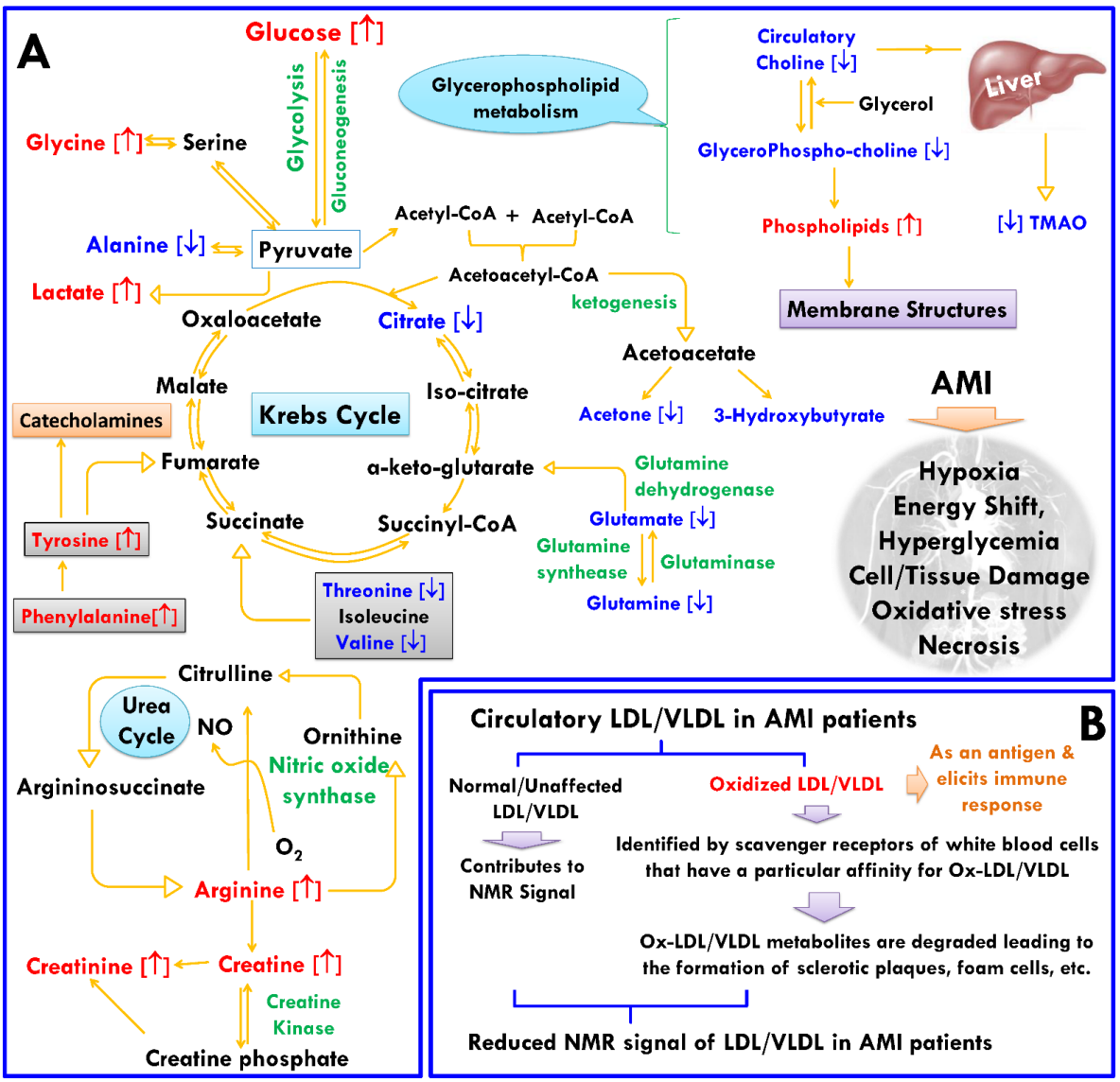



**Figure 8:** Altered metabolic pathways for the most relevant metabolites perturbed in AMI patients with respect to normal controls. The metabolites which were found to be elevated and decreased in AMI patients compared to normal controls are shown in red and blue colored texts, respectively. Whereas, the metabolites which were significantly elevated and decreased in AMI patients compared to normal controls are highlighted here using upward (in red) and downward (in blue) arrows, respectively.

**Shift in Energy Metabolism:**

A general comparison of AMI patients with NC subjects showed a significant increase in glucose and a concomitant decrease in fatty acid levels in the sera of AMI patients. Since more than 90 % of the cardiac ATP is synthesized by mitochondrial oxidative phosphorylation of glucose and fatty acids (FAs) during normal functioning of the heart,[32,33] these metabolic alterations clearly suggest aerobic glycolysis or TCA cycle activity is compromised in these patients. These changes might be related to (a) dampened glycolytic activity and reduced myocardial uptake of glucose which in turn contribute to myocardial insulin resistance (an important clinical manifestation related to AMI)[34] and (b) preferential myocardial uptake and oxidation of FAs during early hours of AMI.[35] These results further support the earlier studies claiming that insulinotropic metabolic adjuvants -which increase myocardial glucose uptake- ameliorate the conditions like heart failure.[34] Our data also showed decreased serum levels of citrate and increased levels of lactate in the sera of AMI patients. The similar higher serum levels of glucose and lactate and concomitant lower serum levels of fatty acids have also been seen in the sera of patients with myocardial ischemia.[9] The key feature of both these medical conditions is hypoxia because of prolonged lack of oxygenated blood supply and thus producing conditions like limited aerobic glycolysis. AMI also demonstrates an acute systemic inflammatory response,[36] therefore, triggering a hypercatabolic state and an increased energy demand.[37] However, to meet the energy requirements under hypoxic conditions and simultaneously to maintain the physiological homoeostasis (a) there is benign switching from aerobic to anaerobic metabolism and (b) an increased reliance on alternative energy substrates preferably amino acids.[38] Which can be oxidized anaerobically with lower contribution to acidosis. In addition, the low serum concentrations of amino acids have also been shown to be closely related to protein-energy wasting, inflammation, and oxidative-stress.[27,39] Consistent with such reports, our results also indicated decreased serum levels of several amino acids in AMI patients (such as alanine, valine, glutamate, glutamine, and methionine) suggesting aberrant amino acid metabolism and protein synthesis to regulate biological functions (**Figure 5**). Our data also suggested the increased serum levels of creatine and creatinine and decreased serum levels of TMAO in AMI patients. A clinical biochemistry report on 483 AMI patients[40] has demonstrated elevated creatinine in 22% of these patients and majority of them were having renal dysfunction suggesting that serum creatinine is not a reliable metabolic signature of AMI, rather, elevated creatine –which is a key intermediate in energy metabolism- might be related to increased energy demand in AMI.[41] While, TMAO is a ubiquitous natural osmoprotectant[42] and in an hypoxic environment may participate in energy conversion through acting as an external electron acceptor (chemiosmotic mechanism of energy production)[43]. The decreased serum levels of TMAO, therefore, might be related to its intracellular accumulation to counteract biochemical stress including energy crisis under conditions of relative oxygen deficit. Taken together, the elevated serum levels of glucose, lactate, and creatine and a concomitant decrease in the serum levels of citrate, TMAO, and several glucogenic

amino acids **(Table 2)**, provided many indications of energy shift and metabolic reprogramming essential for maintaining physiological homeostasis (osmotic, pH, and ionic balance).

**Oxidative Stress: The potential hallmark of AMI**

Oxidative stress after AMI episodes is mainly attributed to hypoxia[44,45] and hyperglycemia.[46,47] A feature common to all cell types that are damaged by oxidative stress or hyperglycemia is an increased production of reactive oxygen species[45,48,49] which lead to oxidative modification of proteins and lipoproteins. These modified species function aberrantly and often lead to immune reactions and inflammations.[50,51] The increased serum levels of N-acetyl glycoproteins is the primary response to counter-act this effect and maintain physiological homeostasis owing to their anti-inflammatory and antioxidant properties.[27,39,52] However to ameliorate the deleterious effect of oxidized proteins, it requires another set of proteins, which could eradicate or detoxify these oxidized proteins. The decreased serum levels of amino acids might be closely related to this phenomenon. However, contrary to this, our data have shown the significantly elevated levels of glycine, arginine, tyrosine, and phenylalanine in the sera of AMI patients compared to normal controls. The similar higher serum levels of glycine, tyrosine, and phenylalanine have also been seen in controlled swine, and human models of acute myocardial ischemia.[9] Phenylalanine is a substrate for tyrosine and tyrosine is a key precursor for biosynthesis of catecholamines, which regulate immune and inflammatory responses.[53] It is also well-established that inflammatory states often cause significant increases in serum phenylalanine.[54] Therefore, the observed increase in the levels of phenylalanine and tyrosine could be part of a compensatory response during initial pain-related burst of catecholamine discharge and afterward to regulate inflammatory or autoimmune responses.[47] On the other hand, glycine is an essential component of the vital antioxidant, glutathione[55,56] and is well-documented cytoprotective metabolite owing to its involvement in antioxidative reactions, purine synthesis, and collagen formation.[55] Collagen formation is the primary response to repair an injury[57] and plays an important role in wound healing.[58,59] Glycine has also been reported to be elevated in heart failure patients[60] and glycine treatment reduces the infarct size by 21% in ischemia-reperfusion injury rats compared to vehicle-treated AMI rats.[61] The increased serum levels of glycine, therefore, might be related to inhibit the proteolysis of skeletal muscle proteins and to augment the collagen synthesis for repairing the infarcted myocardium. Arginine is known for its anti-oxidant properties and is a key precursor molecule for nitric oxide (NO), citrulline, ornithine, urea, and creatine, which makes it an important constituent of many significant biological pathways related to coronary circulation, myocardial contractility, inflammation, etc.[62] Arginine-NO metabolism is a determining factor in the normal functioning of the heart during hypoxia.[63] The adaptive enhancement of NO synthesis is known to improve vascular functions (i.e. vasodilation) and makes the protection more robust and sustained through further activating or increasing the expression of other protective factors, including heat shock proteins, antioxidants, and prostaglandins.[63-65] A previous study on acute myocardial

infacrtion[65] proposed that the elevated serum levels of arginine metabolites are closely related to increased coronary circulation, myocardial contractility, and inflammation; the mechanisms crucial to repair the infarcted myocardium. Taken together, the elevated serum levels of glycine, arginine, tyrosine, and phenylalanine -a potential metabolic hallmark discovered in this study- could be the useful infarction markers for diagnosis or prognosis of AMI.

Recent clinical biochemistry based studies have shown the decreased serum levels of LDL and HDL lipoproteins soon after the AMI attack.[66-68] The decreased serum levels of lipid metabolites have also been reported in a previous metabolomics study on myocardial ischemia.[9] Consistent with these reports; our results also indicated the decreased levels of circulatory LDL and HDL in AMI patients compared to normal controls.[69] The possible reason for decreased serum levels of LDL could be attributed to the excessive peroxidation of LDL to Oxidized-LDL (OxLDL) which is aberrantly involved in inflammatory processes through the formation of higher molecular weight complexes with distinct inflammatory mediators (as well documented previously[70-72]). The oxidized LDL particles also play a pivotal role in atherogenesis, and this might be one of the factors contributing to systemic atherosclerotic inflammation following acute myocardial infarction.[73,74] Inflammation and oxidative stress may lead to the alteration of lipid metabolism.[27,75] Additionally, our data also suggested the increased serum levels of phospholipids and the decreased serum levels of membrane metabolites choline and glycerophosphocholine (GPC), which are the major constituents of biological membranes. Possibly, the stress response of cell membrane damage activates the membrane metabolism **(Figure 8A)** leading to the production of phospholipids. The produced phospholipids are then used for repairing the membranes structures damaged by oxidative and inflammatory processes. Overall, the lipid metabolic profiles in sera of AMI patients suggested dysregulation of lipid metabolism in the background of acute oxidative stress.

**Elevated serum Levels of N-Acetyl Glycoproteins indicated systemic Inflammation:**

N-Acetyl glycoproteins (NAGs) such as N-acetylglucosamine and N-acetylneuramic acid are secreted in response to tissue or cell membrane damage and act as inflammatory mediator.[52,76,77] As such, these are acute phase proteins and exhibit remarkable anti-inflammatory and anti-oxidant properties. The increased serum levels of N-acetyl glycoproteins were likely to reflect a systemic inflammation in AMI patients. Our data also suggested the decreased levels of poly-unsaturated fatty acids (PUFAs) in the sera of AMI patients. PUFAs are key mediators and regulators of inflammation[78] which is a natural defense mechanism to repair the damaged tissue sites and helps to restore the homeostasis.[79,80] Possibly, the death of cardiomyocytes or other cells (in an hypoxia and hyperglycemic environment) triggers an inflammatory response and increased production of proinflammatory cytokines.[81-83] Therefore, we hypothesize that the increased serum levels of NAGs might be related to suppress the post-infarction inflammation and oxidative stress,

whereas the decreased serum level of PUFAs may indicate their augmented utilization to regulate inflammatory and auto-immunity responses.

## Conclusion

The study established serum metabolic patterns of AMI in initial hours (i.e. within 6-8 hours from the onset of clinical symptoms and provided valuable insights into the disease processes operating in AMI patients. Overall, the AMI sera was characterized by (a) the increased levels of glucose, glycine, arginine, creatine, creatinine, tyrosine, phenylalanine and N-acetyl glycoproteins (NAG), and (b) the decreased levels of amino acids (like alanine, valine, glutamate, glutamine, methionine, etc) and membrane metabolites including choline, glycerophosphocholine and polyunsaturated fatty acids. These metabolite perturbations were associated with profoundly dampened glycolysis (or TCA cycle activity), dyslipidemia, and aberrant amino acid metabolism alluding to hypoxia/hyperglycemia-induced oxidative stress, cell/tissue damage, and systemic inflammation.

The application of metabolic analysis to improve the clinical diagnosis and prognosis of cardiovascular diseases is an emerging field[84]. Further, appropriate treatment in addition to accurate diagnosis is equally important to minimize the mortalities or disabilities associated with devastating cardiovascular events including acute myocardial infarction (AMI). The present study underscores the potential of serum metabolic profiling to evolve as a surrogate method for improved diagnosis/prognosis of myocardial infarction and its better clinical management of the patient, in initial hours of admission.

**Reference's**


1. White, H. D.; Chew, D. P. Acute myocardial infarction. *The Lancet* **2008**, *372* (9638), 570-584.

2. Makikallio, T. H.; Barthel, P.; Schneider, R.; Bauer, A.; Tapanainen, J. M.; Tulppo, M. P.; Schmidt, G.; Huikuri, H. V. Prediction of sudden cardiac death after acute myocardial infarction: role of Holter monitoring in the modern treatment era. *European heart journal* **2005**, *26* (8), 762-769.

3. Huikuri, H. V.; Tapanainen, J. M.; Lindgren, K.; Raatikainen, P.; Makikallio, T. H.; Airaksinen, K. J.; Myerburg, R. J. Prediction of sudden cardiac death after myocardial infarction in the beta-blocking era. *Journal of the American College of Cardiology* **2003**, *42* (4), 652-658.

4. Virmani, R.; Burke, A. P.; Farb, A. Sudden cardiac death. *Cardiovascular pathology* **2001**, *10* (5), 211-218.

5. Hall, A. S.; Barth, J. H. Universal definition of myocardial infarction. *Heart* **2009**, *95* (3), 247-249.

6. Zhang, G.; Zhou, B.; Zheng, Y.; Feng, K.; Rao, L.; Zhang, J.; Xin, J.; Zhang, B.; Zhang, L. Time course proteomic profile of rat acute myocardial infarction by SELDI-TOF MS analysis. *International journal of cardiology* **2009**, *131* (2), 225-233.

7. Rosenblat, J.; Zhang, A.; Fear, T. Biomarkers of myocardial infarction: past, present and future. *www. uwomj. com* **2012**, *81* (1), 23.

8. Melanson, S. F.; Tanasijevic, M. J. Laboratory diagnosis of acute myocardial injury. *Cardiovasc. Pathol.* **2005**, *14* (3), 156-161.

9. Bodi, V.; Sanchis, J.; Morales, J. M.; Marrachelli, V. G.; Nunez, J.; Forteza, M. J.; Chaustre, F.; Gomez, C.; Mainar, L.; Minana, G.; Rumiz, E.; Husser, O.; Noguera, I.; Diaz, A.; Moratal, D.; Carratala, A.; Bosch, X.; Llacer, A.; Chorro, F. J.; Vina, J. R.; Monleon, D. Metabolomic profile of human myocardial ischemia by nuclear magnetic resonance spectroscopy of peripheral blood serum: a translational study based on transient coronary occlusion models. *J Am. Coll. Cardiol.* **2012**, *59* (18), 1629-1641.

10. Pandey, R.; Gupta, N. K.; Wander, G. S. Diagnosis of acute myocardial infarction. *J Assoc Physicians India* **2011**, *59* (Suppl), 8-13.

11. Gowda, G. N.; Zhang, S.; Gu, H.; Asiago, V.; Shanaiah, N.; Raftery, D. Metabolomics-based methods for early disease diagnostics. *Expert Rev Mol Diagn.* **2008**, *8*, 617.

12. Nicholson, J. K.; Lindon, J. C. Systems biology: Metabonomics. *Nature* **2008**, *455* (7216), 1054-1056.

13. Serkova, N. J.; Niemann, C. U. Pattern recognition and biomarker validation using quantitative 1H-NMR-based metabolomics. *Expert Rev Mol Diagn.* **2006**, *6*, 717.

14. Viant, M. R.; Rosenblum, E. S.; Tjeerdema, R. S. NMR-based metabolomics: a powerful approach for characterizing the effects of environmental stressors on organism health. *Environmental science & technology* **2003**, *37* (21), 4982-4989.

15. Calderon-Santiago, M.; Priego-Capote, F.; Galache-Osuna, J. G.; Luque de Castro, M. D. Metabolomic discrimination between patients with stable angina, non-ST elevation myocardial infarction, and acute myocardial infarct. *Electrophoresis* **2013**, *34* (19), 2827-2835.

16. Bodi, V.; Marrachelli, V. G.; Husser, O.; Chorro, F. J.; Vi+¦a, J. R.; Monleon, D. Metabolomics in the diagnosis of acute myocardial ischemia. *Journal of cardiovascular translational research* **2013**, *6* (5), 808-815.

17. Barba, I.; Jaimez-Auguets, E.; Rodriguez-Sinovas, A.; Garcia-Dorado, D. 1H NMR-based metabolomic identification of at-risk areas after myocardial infarction in swine. *Magn Reson Mater Phy* **2007**, *20* (5-6), 265-271.

18. Beckonert, O.; Keun, H. C.; Ebbels, T. M.; Bundy, J.; Holmes, E.; Lindon, J. C.; Nicholson, J. K. Metabolic profiling, metabolomic and metabonomic procedures for NMR spectroscopy of urine, plasma, serum and tissue extracts. *Nature protocols* **2007**, *2* (11), 2692-2703.

19. Dona, A. C.; Jime¦ünez, B.; Scha¦êfer, H.; Humpfer, E.; Spraul, M.; Lewis, M. R.; Pearce, J. T.; Holmes, E.; Lindon, J. C.; Nicholson, J. K. Precision high-throughput proton NMR spectroscopy of human urine, serum, and plasma for large-scale metabolic phenotyping. *Analytical chemistry* **2014**, *86* (19), 9887-9894.

20. Tang, H.; Wang, Y.; Nicholson, J. K.; Lindon, J. C. Use of relaxation-edited one-dimensional and two dimensional nuclear magnetic resonance spectroscopy to improve detection of small metabolites in blood plasma. *Analytical biochemistry* **2004**, *325* (2), 260-272.

21. Wu, D. H.; Chen, A. D.; Johnson, C. S. An improved diffusion-ordered spectroscopy experiment incorporating bipolar-gradient pulses. *Journal of magnetic resonance, Series A* **1995**, *115* (2), 260-264.

22. Xia, J.; Bjorndahl, T. C.; Tang, P.; Wishart, D. S. MetaboMiner--semi-automated identification of metabolites from 2D NMR spectra of complex biofluids. *BMC Bioinformatics* **2008**, *9*, 507.

23. Wishart, D. S.; Jewison, T.; Guo, A. C.; Wilson, M.; Knox, C.; Liu, Y.; Djoumbou, Y.; Mandal, R.; Aziat, F.; Dong, E.; Bouatra, S.; Sinelnikov, I.; Arndt, D.; Xia, J.; Liu, P.; Yallou, F.; Bjorndahl, T.; Perez-Pineiro, R.; Eisner, R.; Allen, F.; Neveu, V.; Greiner, R.; Scalbert, A. HMDB 3.0--The Human Metabolome Database in 2013. *Nucleic Acids Res 41* (Database issue).

24. Nicholson, J. K.; Foxall, P. J.; Spraul, M.; Farrant, R. D.; Lindon, J. C. 750 MHz 1H and 1H-13C NMR spectroscopy of human blood plasma. *Analytical chemistry* **1995**, *67* (5), 793-811.

25. Guleria, A.; Bajpai, N. K.; Rawat, A.; Khetrapal, C. L.; Prasad, N.; Kumar, D. Metabolite characterisation in peritoneal dialysis effluent using high-resolution (1) H and (1) H-(13) C NMR spectroscopy. *Magn. Reson. Chem* **2014**, *52* (9), 475-479.

26. Liu, M.; Tang, H.; Nicholson, J. K.; Lindon, J. C. Use of 1H NMR-determined diffusion coefficients to characterize lipoprotein fractions in human blood plasma. *Magnetic Resonance in Chemistry* **2002**, *40* (13).

27. Guleria, A.; Misra, D. P.; Rawat, A.; Dubey, D.; Khetrapal, C. L.; Bacon, P.; Misra, R.; Kumar, D. NMR-Based Serum Metabolomics Discriminates Takayasu Arteritis from Healthy Individuals: A Proof-of-Principle Study. *J Proteome Res* **2015**, *14* (8), 3372.

28. Xia, J.; Sinelnikov, I. V.; Han, B.; Wishart, D. S. MetaboAnalyst 3.0-making metabolomics more meaningful. *Nucleic Acids Res* **2015**, *43 (W1)*, W251.

29. Xia, J.; Psychogios, N.; Young, N.; Wishart, D. S. MetaboAnalyst: a web server for metabolomic data analysis and interpretation. *Nucleic acids research* **2009**, *37* (suppl 2), W652-W660.

30. Worley, B.; Powers, R. Multivariate analysis in metabolomics. *Current Metabolomics* **2013**, *1* (1), 92-107.

31. Xia, J.; Wishart, D. S. MSEA: a web-based tool to identify biologically meaningful patterns in quantitative metabolomic data. *Nucleic acids research* **2010**, *38* (suppl 2), W71-W77.

32. Opie, L. H. Metabolism of free fatty acids, glucose and catecholamines in acute myocardial infarction: relation to myocardial ischemia and infarct size. *The American journal of cardiology* **1975**, *36* (7), 938-953.

33. Neely, J. R.; Rovetto, M. a.; Oram, J. F. Myocardial utilization of carbohydrate and lipids. *Progress in cardiovascular diseases* **1972**, *15* (3), 289-329.

34. Nikolaidis, L. A.; Elahi, D.; Hentosz, T.; Doverspike, A.; Huerbin, R.; Zourelias, L.; Stolarski, C.; Shen, Y. t.; Shannon, R. P. Recombinant glucagon-like peptide-1 increases myocardial glucose uptake and improves left ventricular performance in conscious dogs with pacing-induced dilated cardiomyopathy. *Circulation* **2004**, *110* (8), 955-961.

35. Mazumder, P. K.; O'Neill, B. T.; Roberts, M. W.; Buchanan, J.; Yun, U. J.; Cooksey, R. C.; Boudina, S.; Abel, E. D. Impaired cardiac efficiency and increased fatty acid oxidation in insulin-resistant ob/ob mouse hearts. *Diabetes* **2004**, *53* (9), 2366-2374.

36. Blancke, F.; Claeys, M. J.; Jorens, P.; Vermeiren, G.; Bosmans, J.; Wuyts, F. L.; Vrints, C. J. Systemic inflammation and reperfusion injury in patients with acute myocardial infarction. *Mediators of inflammation* **2005**, *2005* (6), 385-389.

37. Frankenfield, D. Energy expenditure and protein requirements after traumatic injury. *Nutrition in Clinical Practice* **2006**, *21* (5), 430-437.

38. Drake, K. J.; Sidorov, V. Y.; McGuinness, O. P.; Wasserman, D. H.; Wikswo, J. P. Amino acids as metabolic substrates during cardiac ischemia. *Experimental Biology and Medicine* **2012**, *237* (12), 1369-1378.

39. Ouyang, X.; Dai, Y.; Wen, J. L.; Wang, L. X. (1)H NMR-based metabolomic study of metabolic profiling for systemic lupus erythematosus. *Lupus* **2011**, *20* (13), 1411.

40. Walsh, C. R.; O'Donnell, C. J.; Camargo Jr, C. A.; Giugliano, R. P.; Lloyd-Jones, D. M. Elevated serum creatinine is associated with 1-year mortality after acute myocardial infarction. *American heart journal* **2002**, *144* (6), 1003-1011.

41. Taegtmeyer, H.; Ingwall, J. S. Creatine--a dispensable metabolite? *Circ Res* **2013**, *112* (6), 878.

42. Zerbst-Boroffka, I.; Kamaltynow, R. M.; Harjes, S.; Kinne-Saffran, E.; Gross, J. +. TMAO and other organic osmolytes in the muscles of amphipods (Crustacea) from shallow and deep water of Lake Baikal. *Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology* **2005**, *142* (1), 58-64.

43. StrØm; Olafsen, J. A.; LARSEN, H. E. L. G. Trimethylamine oxide: a terminal electron acceptor in anaerobic respiration of bacteria. *Microbiology* **1979**, *112* (2), 315-320.

44. Sedlakova, E.; Racz, O.; Lovasova, E.; Beoaeka, R.; Kurpas, M.; Chmelarova, A.; Sedlak, J.; Studenean, M. Markers of oxidative stress in acute myocardial infarction treated by percutaneous coronary intervention. *cent. eur. j. med* **2009**, *4* (1), 26-31.

45. Filippo, C. D.; Cuzzocrea, S.; Rossi, F.; Marfella, R.; D'Amico, M. Oxidative stress as the leading cause of acute myocardial infarction in diabetics. *Cardiovascular drug reviews* **2006**, *24* (2), 77-87.

46. Esposito, K.; Nappo, F.; Marfella, R.; Giugliano, G.; Giugliano, F.; Ciotola, M.; Quagliaro, L.; Ceriello, A.; Giugliano, D. Inflammatory cytokine concentrations are acutely increased by hyperglycemia in humans role of oxidative stress. *Circulation* **2002**, *106* (16), 2067-2072.

47. Opie, L. H. Metabolic management of acute myocardial infarction comes to the fore and extends beyond control of hyperglycemia. *Circulation* **2008**, *117* (17), 2172-2177.

48. Brownlee, M. The pathobiology of diabetic complications a unifying mechanism. *Diabetes* **2005**, *54* (6), 1615-1625.

49. Nishikawa, T.; Edelstein, D.; Du, X. L.; Yamagishi, S. i.; Matsumura, T.; Kaneda, Y.; Yorek, M. A.; Beebe, D.; Oates, P. J.; Hammes, H. P. Normalizing mitochondrial superoxide production blocks three pathways of hyperglycaemic damage. *Nature* **2000**, *404* (6779), 787-790.

50. Sohal, R. S. Role of oxidative stress and protein oxidation in the aging process. *Free Radical Biology and Medicine* **2002**, *33* (1), 37-44.

51. Keller, J. N. Interplay between oxidative damage, protein synthesis, and protein degradation in Alzheimer's disease. *BioMed Research International* **2006**, *2006*, 12129.

52. Sun, L.; Hu, W.; Liu, Q.; Hao, Q.; Sun, B.; Zhang, Q.; Mao, S.; Qiao, J.; Yan, X. Metabonomics reveals plasma metabolic changes and inflammatory marker in polycystic ovary syndrome patients. *Journal of proteome research* **2012**, *11* (5), 2937-2946.

53. Flierl, M. A.; Rittirsch, D.; Huber-Lang, M.; Sarma, J. V.; Ward, P. A. Catecholamines--Crafty Weapons in the Inflammatory Arsenal of Immune/Inflammatory Cells or Opening Pandora's Box ? *Molecular medicine* **2008**, *14*, 195.

54. Wannemacher, R. W.; Klainer, A. S.; Dinterman, R. E.; Beisel, W. R. The significance and mechanism of an increased serum phenylalanine-tyrosine ratio during infection. *The American journal of clinical nutrition* **1976**, *29* (9), 997-1006.

55. Hasegawa, S.; Ichiyama, T.; Sonaka, I.; Ohsaki, A.; Okada, S.; Wakiguchi, H.; Kudo, K.; Kittaka, S.; Hara, M.; Furukawa, S. Cysteine, histidine and glycine exhibit anti-inflammatory effects in human coronary arterial endothelial cells. *Clinical & Experimental Immunology* **2012**, *167* (2), 269-274.

56. Matilla, B.; Mauriz, J. L.; Culebras, J. M.; Gonzalez-Gallego, J.; Gonz+ílez, P. Glycine: a cell-protecting anti-oxidant nutrient. *Nutricion hospitalaria* **2001**, *17* (1), 2-9.

57. Barbul, A. Proline precursors to sustain mammalian collagen synthesis. *The Journal of nutrition* **2008**, *138* (10), 2021S-2024S.

58. Albina, J. E.; Abate, J. A.; Mastrofrancesco, B. Role of ornithine as a proline precursor in healing wounds. *Journal of Surgical Research* **1993**, *55* (1), 97-102.

59. Ponrasu, T.; Jamuna, S.; Mathew, A.; Madhukumar, K. N.; Ganeshkumar, M.; Iyappan, K.; Suguna, L. Efficacy of l-proline administration on the early responses during cutaneous wound healing in rats. *Amino acids* **2013**, *45* (1), 179-189.

60. Deidda, M.; Piras, C.; Dessalvi, C. C.; Locci, E.; Barberini, L.; Torri, F.; Ascedu, F.; Atzori, L.; Mercuro, G. Metabolomic approach to profile functional and metabolic changes in heart failure. *Journal of translational medicine* **2015**, *13* (1), 297.

61. Zhong, X.; Li, X.; Qian, L.; Xu, Y.; Lu, Y.; Zhang, J.; Li, N.; Zhu, X.; Ben, J.; Yang, Q. Glycine attenuates myocardial ischemiaGÇôreperfusion injury by inhibiting myocardial apoptosis in rats. *Journal of biomedical research* **2012**, *26* (5), 346-354.

62. Boger, R. H. The pharmacodynamics of L-arginine. *The Journal of nutrition* **2007**, *137* (6), 1650S-1655S.

63. Manukhina, E. B.; Downey, H. F.; Mallet, R. T. Role of nitric oxide in cardiovascular adaptation to intermittent hypoxia. *Experimental Biology and Medicine* **2006**, *231* (4), 343-365.

64. Schulman, S. P.; Becker, L. C.; Kass, D. A.; Champion, H. C.; Terrin, M. L.; Forman, S.; Ernst, K. V.; Kelemen, M. D.; Townsend, S. N.; Capriotti, A. L-arginine therapy in acute myocardial infarction: the Vascular Interaction With Age in Myocardial Infarction (VINTAGE MI) randomized clinical trial. *Jama* **2006**, *295* (1), 58-64.

65. Nicholls, S. J.; Wang, Z.; Koeth, R.; Levison, B.; DelFraino, B.; Dzavik, V.; Griffith, O. W.; Hathaway, D.; Panza, J. A.; Nissen, S. E. Metabolic profiling of arginine and nitric oxide pathways predicts hemodynamic abnormalities and mortality in patients with cardiogenic shock after acute myocardial infarction. *Circulation* **2007**, *116* (20), 2315-2324.

66. Rott, D.; Klempfner, R.; Goldenberg, I.; Leibowitz, D. Cholesterol Levels Decrease soon after Acute Myocardial Infarction. *The Israel Medical Association journal: IMAJ* **2015**, *17* (6), 370-373.

67. Fresco, C.; Maggioni, A. P.; Signorini, S.; Merlini, P. A.; Mocarelli, P.; Fabbri, G.; Lucci, D.; Tubaro, M.; Gattone, M.; Schweiger, C. Variations in lipoprotein levels after myocardial infarction and unstable angina: the LATIN trial. *Italian heart journal: official journal of the Italian Federation of Cardiology* **2002**, *3* (10), 587-592.

68. Shrivastava, A. K.; Singh, H. V.; Raizada, A.; Singh, S. K. Serial measurement of lipid profile and inflammatory markers in patients with acute myocardial infarction. *EXCLI journal* **2015**, *14*, 517.

69. Cho, K. H.; Shin, D. G.; Baek, S. H.; Kim, J. R. Myocardial infarction patients show altered lipoprotein properties and functions when compared with stable angina pectoris patients. *Exp Mol Med* **2009**, *41*, 67-76.

70. Holvoet, P.; Mertens, A.; Verhamme, P.; Bogaerts, K.; Beyens, G.; Verhaeghe, R.; Collen, D. +.; Muls, E.; Van de Werf, F. Circulating oxidized LDL is a useful marker for identifying patients with coronary artery disease. *Arteriosclerosis, thrombosis, and vascular biology* **2001**, *21* (5), 844-848.

71. Tripathi, P.; Misra, M. K.; Pandey, S. Role of l-Arginine on Dyslipidemic Conditions of Acute Myocardial Infarction Patients. *Indian Journal of Clinical Biochemistry* **2012**, *27* (3), 296-299.

72. Ballantyne, F. C.; Clark, R. S.; Simpson, H. S.; Ballantyne, D. High density and low density lipoprotein subfractions in survivors of myocardial infarction and in control subjects. *Metabolism* **1982**, *31* (5), 433-437.

73. Joshi, N. V.; Toor, I.; Shah, A. S. V.; Carruthers, K.; Vesey, A. T.; Alam, S. R.; Sills, A.; Hoo, T. Y.; Melville, A. J.; Langlands, S. P.; Jenkins, W. S. A.; Uren, N. G.; Mills, N. L.; Fletcher, A. M.; van Beek, E. J. R.; Rudd, J. H. F.; Fox, K. A. A.; Dweck, M. R.; Newby, D. E. Systemic Atherosclerotic Inflammation Following Acute Myocardial Infarction: Myocardial Infarction Begets Myocardial Infarction. *Journal of the American Heart Association* **2015**, *4* (9), e001956.

74. Wang, H.; Eitzman, D. T. Acute myocardial infarction leads to acceleration of atherosclerosis. *Atherosclerosis* **2013**, *229* (1), 18-22.

75. Khovidhunkit, W.; Kim, M. S.; Memon, R. A.; Shigenaga, J. K.; Moser, A. H.; Feingold, K. R.; Grunfeld, C. Effects of infection and inflammation on lipid and lipoprotein metabolism: mechanisms and consequences to the host. *The Journal of Lipid Research* **2004**, *45* (7), 1169-1196.

76. Bell, J. D.; Brown, J. C.; Nicholson, J. K.; Sadler, P. J. Assignment of resonances for 'acute-phase' glycoproteins in high resolution proton NMR spectra of human blood plasma. *FEBS Lett.* **1987**, *215* (2), 311-315.

77. Arnold, J. N.; Saldova, R.; Hamid, U. M.; Rudd, P. M. Evaluation of the serum N-linked glycome for the diagnosis of cancer and chronic inflammation. *Proteomics* **2008**, *8* (16), 3284-3293.

78. Calder, P. C. Polyunsaturated fatty acids and inflammatory processes: new twists in an old tale. *Biochimie* **2009**, *91* (6), 791-795.

79. Salmon, J. A.; Higgs, G. A. Prostaglandins and leukotrienes as inflammatory mediators. *British Medical Bulletin* **1987**, *43* (2), 285-296.

80. Marinetti, G. V. *Disorders of lipid metabolism;* Springer Science & Business Media: Plenum Press, New York, 1990.

81. Frangogiannis, N. G.; Smith, C. W.; Entman, M. L. The inflammatory response in myocardial infarction. *Cardiovascular research* **2002**, *53* (1), 31-47.

82. Berton, G.; Cordiano, R.; Palmieri, R.; Pianca, S.; Pagliara, V.; Palatini, P. C-reactive protein in acute myocardial infarction: association with heart failure. *American heart journal* **2003**, *145* (6), 1094-1101.

83. Saxena, A.; Russo, I.; Frangogiannis, N. G. Inflammation as a therapeutic target in myocardial infarction: learning from past failures to meet future challenges. *Translational Research* **2016**, *164*, 152.

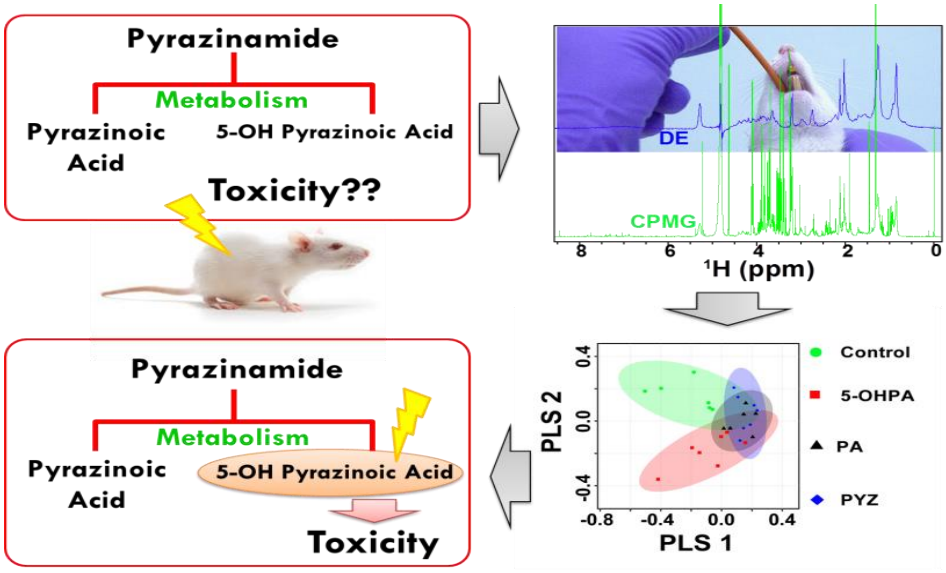
84. Barderas, M. G.; Laborde, C. M.; Posada, M.; de la Cuesta, F.; Zubiri, I.; Vivanco, F.; varez-Llamas, G. Metabolomic profiling for identification of novel potential biomarkers in cardiovascular diseases. *BioMed Research International* **2011**, *2011*, Article ID 790132.

# Chapter 5

## Understanding the toxicity mechanism of Pyrazinamide as revealed by NMR based serum metabolomics and Biochemical analysis


### Abstract

Pyrazinamide (PYZ) -an essential component of primary drug regimen used for the treatment and management of multidrug-resistant or latent tuberculosis- is well known for its hepatoxicity. However, the mechanism of pyrazinamide-induced hepatotoxicity is still unknown to researchers. Studies have shown that the drug is metabolized in the liver to pyrazinoic acid (PA) and 5-hydroxy pyrazinoic acid (5-OHPA) which individually may cause different degrees of hepatotoxicity. To evaluate this hypothesis, PYZ, PA and 5-OHPA were dosed to albino Wistar rats orally (respectively, at 250, 125, and 125 mg/kg for 28 days). Compared to normal rats, PZA and its metabolic products decreased the weights of dosed rats and induced liver injury and a status of oxidative stress as assessed by combined histopathological and biochemical analysis. Compared to normal controls, the biochemical and morphological changes were more aberrant in PA and 5-OHPA dosed rats with respect to those dosed with PYZ. Finally, the serum metabolic profiles of rats dosed with PYZ, PA, and 5-OHPA were measured and compared with those of normal control rats. With respect to normal control rats, the rats dosed with PYZ and 5-OHPA showed most aberrant metabolic perturbations in their sera compared to those dosed with PA. Altogether, the study suggests that PYZ induced hepatotoxicity might be associated with its metabolized products, where 5-OHPA contributes to a higher degree of its overall toxicity than PA.

## Introduction

Drug induced hepatotoxicity is a crucial health care issue and one of the leading causes of morbidity and mortality around the world[1]. Individuals suffering from drug induced liver injury exhibit a wide range of manifestations clinically, biochemically and histologically including acute liver failure with severe encephalopathy, acute hepatitis with or without jaundice, etc[2]. Therefore, it is necessary to assess drug-induced hepatotoxicity for the design of safer and better therapeutic agents. Within this framework, the recent study was performed to evaluate the mechanism of toxicity of PYZ using albino Wistar rats.

Pyrazinamide (PYZ), an amide derivative of the pyrazine-2-carboxylic acid, is an essential component of the first line drug for the treatment and management of multidrug resistant or latent tuberculosis[3]. The drug is usually given in combination with isoniazid, rifampicin, mainly as an antibiotic course for a period of 6-months to tuberculosis patients as a standard treatment regimen[4]. Hepatotoxicity is the most serious complication arising from the first line treatment of tuberculosis (TB)[5, 6]. Previous studies unveiled that anti-TB drug-induced hepatotoxicity was more pronounced with PYZ than any other tubercular medication[7]. PYZ is reported to possess various side effects like skin rashes, malaise, dysosmia, anemia, anorexia, gastrointestinal upset, arthralgia's, photosensitivity and hypersensitivity reactions[8]. However, the major side effect associated to PYZ is hepatotoxicity which can be severe to lethal if not diagnosed properly in time. The mechanism of hepatic injury by PYZ is not known, but the drug is extensively metabolized by cytochrome 2E1, 3A4 in the liver to pyrazinoic acid (PA) and 5-hydroxy pyrazinoic acid (5-OHPA) and injury may be caused either by the drug itself or its metabolites[9]. A recent investigation suggested that both these metabolites were toxic to normal human hepatocytes (Hep-G2 cells) in vitro[10, 11]. No further in vivo investigations had been yet performed to evaluate whether PYZ or its metabolites (PA and 5-OHPA) cause the hepatotoxicity. To get the answer, PYZ and its main metabolites (PA and 5-OHPA) were administered orally (two doses a day, respectively, at 250 and 125 mg/kg of body weight) for 28 days to albino Wistar rats. First, we measured the various liver injury and oxidative stress related parameters in liver and serum to evaluate the comparative effect of PZA and its metabolites. Reduced glutathione (GSH), catalase (CAT), superoxide dismutase (SOD), protein carbonyl (PC), malonaldehyde, (MDA), bilirubin and biliverdin in liver tissues and aspartate aminotransferase (AST) and alanine aminotransferase (ALT) in serum were measured to evaluate the comparative effect of PYZ and its metabolites. Scanning electron microscopic (SEM) and histopathological studies of the liver tissues were also performed to evaluate the morphological changes. In a separate experiment, we also measured the tissue deposition of these analytes using high-performance liquid chromatography (HPLC). Like PYZ, the PA and 5-OHPA increased liver enzyme concentrations in serum and decreased CAT, SOD, GSH in liver tissues and caused aberrant morphological changes of liver tissue as well compared to controls. Likewise, MDA, PC formation and conjugated bilirubin

and biliverdin deposition in liver were also increased in the similar experiment. Next, we employed NMR based metabolomics approach in combination with multivariate statistical analysis to evaluate the toxicity potential of PYZ and its metabolic products. Particularly, we performed serum metabolomics analysis to identify the biochemical perturbations induced by the parent drug PYZ and its metabolic product PA and 5-OHPA. To the best of our knowledge, in the present study, we report for the first time using NMR-based metabolomics that metabolites of PYZ might be responsible for its in vivo hepatotoxicity.

## Materials and Method

### Animals

Male albino Wistar rats (100 to 120 g, of the same age group) were used for this experiment, and the prior protocol approval was taken from Institutional Animal Ethical Committee (Approval No. SDCOP&VS/AH/CPCSE01/017/R3). The animals were housed under standard laboratory conditions of temperature (25±1°C) with a light/dark cycle of 12 hours with free access to commercial pellet diet and water ad libitium. Animals were acclimatized to laboratory conditions for one week before the experiment.

### Experimental Design

All animals were randomly divided into 5 groups of 6 animals each. Drugs were suspended in 0.25% carboxy methyl cellulose (CMC), subjected to 28 days treatment orally. As reported previously, the PYZ administered at a dose of 250 mg/kg body weight causes substantial hepatotoxicity. It might be assumed that 50% of the PYZ is converted to PA and 5-OHPA after metabolism. Thus we also included the half dose (125 mg/kg body weight) for PA and 5-OHPA in the present study. The procedure was adopted from the literature by Zhang et al., (2013)[12] where it is shown that PYZ produced toxicity potential at 250 mg/kg dose for 28 days in the liver of albino Wistar rats. Overall, the rats groups were divided as follows: Group I (normal control): 0.25% CMC (2 mL/kg), group II: PYZ (250 mg/kg), group III: PA (125 mg/kg), and group IV: 5-OHPA (125 mg/kg). After 28 days, rats were sacrificed, and blood samples were collected by heart puncture in sterile centrifuge tubes and kept at 37°C for 30 minutes, centrifuged at 2000 rpm for 10 minutes, the pale yellow color supernatant was collected in a sterile micro centrifuge tube. The serum obtained was frozen immediately at −80°C prior to NMR spectroscopic analysis. Further, livers were dissected out and rinsed with ice cold saline and stored at -80°C for further studies.

## Biochemical estimations

### Serum aspartate aminotransferase (AST) and alanine aminotransferase (ALT)

Liver function biomarkers like AST and ALT were estimated in serum using a commercially available kit from Transia Biomedicals Ltd., Baddi, Himachal Pradesh, India[13]. According to manufacturer's protocol, 100 µl of serum samples added 1 ml of working reagent and measured the absorbance at 340 nm at regular intervals of 1 min for 3 minutes.

$$\text{Activity of ALT or AST (U/mL)} = \Delta A340/\text{min} \times 1768$$

### Tissue bilirubin and biliverdin

#### *Tissue bilirubin*

Bilirubin in liver was measured as per the following procedure published earlier in the literature with slight modifications[14]. All the tissue samples were thawed and homogenized in phosphate buffer saline (8.0g sodium chloride, 0.2g potassium chloride, 0.2g potassium dihydrogen phosphate, 1.15g disodium hydrogen phosphate, 0.372g ethylenediaminetetraacetic acid disodium salt, pH 7.4). 500 µL of tissue homogenate (10%) was added to 2.0 mL of 1.5% butylatedhydroxy toluene in acetone:ethanol (1:1) in an eppendorf tube. Simultaneously, the fresh diazo reagent was prepared by mixing 300 µL of 10% sodium nitrite and 8.0 mL of 2M p-toluene sulfonic acid, then combining 4.0 mL of this mixture with 2.0 mL of 2.1% p-iodoaniline in glacial acetic acid, kept at room temperature for 2.0 min. Then this solution was diluted with distilled water (10 mL) and 200 µL of 1.5M ammonium sulfamate. This working diazo reagent was kept on ice for 5min, and 500 µL was added to each sample homogenates. Diazo blank reagent was freshly prepared by combining 2 mL of p-toluene sulfonic acid and 5.0 mL of 10% ascorbic acid, followed by addition of 2.1% p-iodoaniline in glacial acetic acid and 2.0 mL of n-butyl acetate, mixed and used immediately. Finally, all the tubes were incubated for 1 hour on ice in the dark. After incubation freshly prepared 3.0 mL of 1% ascorbic acid in 0.1 M sodium chloride was added to each vial. All the vials were shaken vigorously, kept for 1.0 min and centrifuged at 2400 rpm for 10 min. The absorbance of the upper organic phase was taken at 530 nM wavelength. The content of bilirubin was calculated as follows:

$$A_{530}\text{sample} - A_{530}\,\text{sample blank} = \Delta A_{530}\text{Test}$$

#### *Tissue biliverdin*

Estimation of biliverdin was performed as per the method prescribed in the previous literature with slight modifications[14]. Tissue samples were homogenized in phosphate buffer saline as per described in the previous section. 500 µL of tissue homogenate (10%) was combined with 500 µL of 10 M glacial acetic acid, 400 µL of 40 mM ascorbic acid, 500 µL of double distilled water and 100 µL of 200 mM barbituric acid. Samples were incubated in a water bath at 95°C

in the dark and then samples were extracted with butanol, vortexed and centrifuged. The upper organic layer was carefully removed and extracted with 2.5 mL 2M sodium hydroxide. The absorbance of the upper layer was taken at 535 nM wavelength. The content of biliverdin was calculated as follows:

$$A_{535}\text{sample} - A_{535}\text{ sample blank} = \Delta A_{535}\text{Test}$$

**Determination of oxidative stress parameters**

Malonaldehyde (MDA)[15], Protein carbonyl (PC)[16], and thiobarbituric acid reactive substances (TBARS) were evaluated in liver. In pancreatic tissue, other oxidative parameters like tissue catalase (CAT)[15], reduced glutathione (GSH), and superoxide dismutase (SOD)[13] levels were measured in the similar experiment. The total protein content of each sample was measured using the Bradford reagent, and bovine serum albumin (BSA) was used as a standard.

***Tissue malonaldehyde (MDA)***

MDA assay was performed which had been published previously. 1.0 mL of 10% (w/v) tissue homogenate, 0.5 mL of 30% trichloroacetic acid and 0.5 mL of 0.8% thiobarbituric acid were taken together in a falcon tube and covered with aluminum foil. Then, the tubes were kept in a shaking water bath for 30 min at 80°C. Later, it was cooled for 15 min and centrifuged at 3000 rpm for 15 min. Absorbance was recorded spectrophotometrically at 540 nM against blank in which tissue sample is absent. The amount of MDA present in a sample was calculated according to the following equation:

nM of MDA/µg of protein = (V × OD at 540 nM) / (0.56 × protein concentration), where, V is final volume of the test solution.

***Tissue protein carbonyl (PC)***

PC assay was performed as per the method prescribed in the previous literature with slight modifications. 10% tissue homogenate was prepared in distilled water. 150 µL of tissue homogenate was taken in eppendorf tube and precipitated by adding 500 µL of 10% trichloracetic acid. Then the tubes were centrifuges at 13,000 rpm for 2 min, and the supernatant was discarded. Later, the cell pellets were incubated with 500 µL of 0.2% 2,4-dinitrophenylhydrazine with constant vortexing at every 5 min interval for 1 h. After that, the supernatant was removed, and cell pellets were washed with 500 µL ethanol:ethyl acetate (1:1) solution three times. At last, pellets were dissolved in 600 µL Guanidine Hydrochloride (6M), and absorbance was measured at 360 nM. Blanks solution was prepared in the similar procedure where cells were absent. The PC content was calculated as follows:

$$\text{PC (µg/mg of protein)} = (A_{360}\text{sample} - A_{360}\text{ sample blank})/\text{ mg of protein}$$

***Tissue glutathione (GSH)***

GSH assay was performed as per the method prescribed in the previous literature with slight modifications. 0.2 mL of 10% (w/v) tissue homogenate was taken in eppendorf tube and 1.8 mL distilled water added to it. Simultaneously, we prepared precipitating solution by mixing 1.67 g of glacial metaphosphoric acid, 0.2 g ethylenediaminetetraacetic acid disodium salt and 30 g sodium chloride in 100 mL distilled water. This precipitating solution was added to the above mixture. The mixture was then allowed to stand for 5 min and filtered. To 2 ml of filtrate, 1.0 mL of 0.4% w/v 5,5'-dithio-bis-2-nitrobenzoic acid and 8.0 mL of 0.3 M phosphate solution were added and centrifuged at 13000 rpm for 1 min. A blank was prepared in the similar procedure where tissue sample was absent. Then, the optical density (OD) was measured at 412 nM. The total protein content of each sample was measured using the Bradford reagent, and bovine serum albumin (BSA) was used as a standard. The tissue GSH content was calculated as follows:

GSH (µM/µg of protein) = (310.4 × $E_i$ × OD at 412 nm) / µg of protein, where $E_i$ is the correction factor (0.542).

***Tissue superoxide dismutase (SOD)***

Determination of SOD in the test samples was performed as per the method prescribed in the previous literature with slight modifications. 100 µL of 10% cytosolic supernatant was prepared with tris–hydrochloric acid buffer (pH=8.5) and final volume were adjusted up to 3.0 mL with the same buffer. Finally, 25 µL of pyrogallol was added, and change in absorbance was recorded at 420 nM at the one-minute interval for 3 minutes. Blank was prepared in which tissue sample was absent. One unit of SOD is described as the amount of enzyme required causing 50% inhibition of pyrogallol auto-oxidation per 3 mLof assay mixture and is given by the formula:

$$\text{Unit of SOD / µg or protein} = [100 \times [(A-B)/ (A\times 50)]]/\text{µg of protein}$$

Where A = Change in absorbance per minute in control and B = Change in absorbance per minute in the test sample.

***Tissue catalase (CAT)***

CAT enzyme estimation was performed as per procedure which was described previously[2]. 10% (w/v) tissue homogenate was prepared in 50mM phosphate buffer and centrifuged at 10000 rpm for 20 minutes. 50 µL of supernatant was added to a tube containing 2.95 mL of 19 mM solution of hydrogen peroxide ($H_2O_2$) prepared in potassium phosphate buffer. The disappearance of $H_2O_2$ was monitored at 1 min interval for 3 mins at 240 nM. CAT activity was calculated as follows:

nM of $H_2O_2$/min/µg of protein = (ΔA/min × volume of assay) / (0.0719 × volume of sample × µg of protein)

**Histopathology and SEM analyses**

Histopathological studies were also performed to find out the morphological changes of liver cells after PYZ, PA AND 5-OHPA administration[17]. Liver tissue from each group was assessed for their morphological changes using eosin and haematoxylin staining. The tissues were preserved in 10% formalin overnight. Next day, the cells were again superseded by 70% isopropanol overnight. Later, the tissues were exposed to isopropanol at various concentrations (70, 90 and 100%) and dehydrated by 100% xylene. The tissue samples were then embedded in bees wax and 5 µM sections were prepared by using microtome. Then, the tissues were succeeded by haematoxylin and eosin staining and observed under a microscope (magnification 40 X).

For SEM analysis, liver tissue samples were collected (2-4 mm) and fixed in 2.5% glutaraldehyde for 2-6 h at 4°C for primary fixation. Then, the samples were washed with 0.1 M phosphate buffer for 15 min at 4°C. After that, 1% osmium tetroxide was used as a post-fixation for 2 h at 4°C. Again, the samples were washed in 0.1 M phosphate buffer for 3 times at 15 min interval and kept at 4°C. Later, these samples were dehydrated with acetone at various concentrations (30, 50, 70, 90, 95 and 100%). After that, all specimens were air dried at room temperature and critical point drying (31.5°C at 1100 psi). Finally, samples were mounted on to the aluminum stubs with adhesive tape and observed for the morphological changes using scanning electron microscope (JEOL JSM-6490LV).

**$^1$H NMR based Metabolomics method**

**Sample preparation**

250 µL of serum was mixed with 250 µL 0.9% saline sodium-phosphate buffer (20 mM, pH 7.4) in D2O. The samples were then centrifuged at 10,000 rpm for 5 min at 4°C to remove any precipitates, before acquiring the NMR data. A total 400 µL of the supernatant was used in 5 mm NMR tubes (Wilmad Glass, USA) for data acquisition with a co-axial insert containing 0.1% TSP (Sodium salt of 3-trimethylsilyl-(2,2,3,3-d4)-propionic acid) as an external standard reference to aid metabolite quantification via NMR experiment. Deuterium oxide (D2O; as a co-solvent and to provide a deuterium field/frequency lock) and the sodium salt of trimethylsilylpropionic acid-d4 (TSP) were purchased from Sigma-Aldrich (Rhode Island, USA).

**NMR measurements**

NMR spectra were recorded at Bruker Biospin Avance-III 800 MHz NMR spectrometer, running at a proton frequency of 800.21 MHz. The NMR instrument was equipped with

CryoProbe with shielded maximum gradient-strength output of 53 G/cm. The raw NMR data were acquired on Topspin-V2.1 (Bruker NMR data Processing Software). For each serum sample, transverse relaxation-edited CPMG (Carr–Purcell–Meiboom–Gill) NMR spectra were acquired using the standard Bruker's pulse program library sequence (cpmgpr1d) with pre-saturation of the water peak through irradiating it continuously during the recycle delay (RD) of 5 sec. Each spectrum consisted of the accumulation of 128 scans and lasted for approximately 15 minutes. A total spin–spin relaxation time of 60 ms (n=300 and $2\tau=200\delta s$) was applied to remove broad signals from triglycerides, proteins, cholesterols, and phospholipids. Each FID (free induction decay) was zero filled and Fourier-transformed to 64 K data points following manual phase and baseline-correction using Bruker NMR data Processing Software Topspin-V3.0. A line broadening factor of 0.3 Hz and a sine–bell apodization function was employed to FIDs before Fourier Transformation. After FT, the chemical shifts were referenced internally to methyl peak of lactate (at $\delta=1.33$ ppm). To obtain spectra with signals only from lipids or lipoproteins, the Diffusion Edited (DE) 1D $^1$H NMR spectra were recorded using the standard Bruker's pulse program library parameters (ledbpgp2s1d i.e. 1-dimensional longitudinal eddy current bipolar gradient pulse presaturation with 2 stimulated echoes). The spectra were measured using sine shaped gradient pulses of strength 30% and a duration of 1.5ms followed by a delay of 200µs to allow for the decay of eddy current, a diffusion time of 120ms and an eddy current decay time of 5ms. The relaxation delay was 4 seconds, and water peak irradiation was applied during the recycle delay and the delay after the first BPP. A line broadening factor of 1 Hz was applied to FIDs before Fourier Transformation. The FIDs were processed using exponential line broadening of 1.0 Hz; spectra were recorded with 128 scans and zero-filled to 64K points before Fourier transformation. All recorded spectra were, visually monitored for acceptability and subjected to multivariate statistical analysis to discriminate the altered metabolic patterns.

**Spectral assignment**

For the assignment of various peaks in the 1D $^1$H CPMG and Diffusion-edited NMR spectra, chemical shifts were identified and assigned by comparing them with the chemical shifts available with the software Chenomx 8.1 (Chenomx Inc., Edmonton, Canada). The remaining peaks in the CPMG $^1$H NMR spectra were assigned by adopting previously reported NMR assignments of metabolites[18,19], data obtained from BMRB database (Biological Magnetic Resonance Data Bank) and HMDB (The Human Metabolome Database)[20] for Diffusion-edited $^1$H NMR spectra assignment was done using previously reported assignments of metabolites in literature[21,22].

**Multivariate Statistical analysis**

Before multivariate data analysis, all the NMR spectra were manually phased, and baseline corrected using TopSpin 3.0 (Bruker NMR data Processing Software). For multivariate

analysis, the CPMG δ(8.5-0.5) ppm and DE δ(6.0 to 0.5) ppm spectra were binned and integrated automatically using AMIX package (Version 3.8.7, Bruker, BioSpin). To be noted here is that the $^{1}H$ CPMG spectra contain signals both from low molecular weight (MW) metabolites and lipid metabolites. The quantitative difference of lipid signals may understate the discriminatory importance of metabolites present in less abundant; therefore by excluding them from the analysis allows better quantitative comparison of low MW metabolites and surmounts their discriminatory importance as well. Therefore, excluding the lipid signals from the data matrix has been employed here to explore the metabolic alterations for the metabolites other than lipids/fatty-acids. The spectral regions corresponding to water and lipids excluded from the CPMG data set were: δ(5.505–4.705), δ(3.675–3.575), δ(3.35–3.33), δ(3.225–3.055), δ(2.705–2.655), δ(2.565–2.525), δ(1.405–1.075), and δ(0.9–0.5), whereas for DE spectra, the spectral region δ(5.0-4.7) ppm distorted due to water was excluded. Finally, the selected regions were reduced to spectral bins of δ(0.01) ppm and δ(0.05) ppm, respectively for CPMG and DE spectra and each spectral bin is further normalized using the total spectral intensity to eliminate the dilution effect among samples and to give the same total integration value for each spectrum.

The binned data both from CPMG and DE experiments were subjected to multivariate data analysis using web-based tools available with open access server MetaboAnalyst (Version 3, from the University of Alberta, Canada)[23,24]. For each data set, NMR variables were Pareto scaled and subsequently, subjected to unsupervised principal component analysis, (PCA) for an initial overview of the grouping trend, which displays the internal structure of datasets in an unbiased way and diminishes the dimensionality of data (i.e. intrinsic clustering) and outliers within the data set (**Figure 2**). Next, the data were modeled with the supervised method of Partial Least Squares Discriminant Analysis (PLS-DA) to reveal class separations between the groups and to further identify the metabolites responsible for class separation. The PLS-DA models were cross-validated by a permutation analysis (100 times), and the resulting cross-validation parameters $R^2$ and $Q^2$ were used to assess the quality of the PLS-DA models i.e. the goodness-of-fit parameter by $R^2$ (also referred to as explained variance) and the goodness of prediction parameter by $Q^2$ (or the predictive capability of the model). The PLS-DA model was further used to identify the metabolites responsible for the discrimination based on their higher values of variable importance on projection scores (i.e. VIPs)[25] and exhibiting statistical significance as evaluated based on 0.05 level of probability i.e. p-value $<0.05$ (calculated using Mann-Whitney test for pairwise comparisons). The VIP score represents a weighted sum of squares of the PLS loadings and takes into account the amount of explained Y-variation in each dimension to measure the impact of each metabolite in the model. Generally, metabolites with high-impact have VIP values higher than one. In this study, the VIP score cut-off value $\geq 2$ for CPMG and $\geq 1$ for diffusion edited data were used for discriminatory significance. The boxplot representation (evaluated through

univariate analysis) was used to visualize the variation in the levels of significantly altered metabolites, identified in the multivariate analysis.

## Results

### Biochemical Parameters

#### Serum aspartate aminotransferase (AST) and alanine aminotransferase (ALT)

According to **Table 1**, both ALT and AST levels in serum were increased in all treatment groups with respect to control. It was observed that the levels of these enzymes were dramatically increased (twice than normal control) after oral administration of PYZ and it metabolites. Both enzyme levels were more increased for PA and 5-OHPA than PYZ.

#### Liver bilirubin and biliverdin

Measurement of conjugated bilirubin and biliverdin was the important parameter for hepatotoxicity. We observed that both bilirubin and biliverdin levels were increased in toxicant groups. As shown in **Table 1**, bilirubin level was increased three times for both PA (~ 85 ng/dL) and 5-OHPA (~ 116 ng/dL) than normal control (~ 34 ng/dL). PYZ demonstrated moderate result though slightly higher than normal. A similar trend was observed for biliverdin where we found that biliverdin level also slightly increased for both PA and 5-OHPA than normal control.

#### Determination of oxidative stress parameters

We also observed various oxidative stress parameters in the liver to evaluate the toxicity potential of PYZ and its metabolites. Various oxidative stress parameters like SOD, CAT, GSH, TBARS and PC in liver were also measured in the similar experiment. We observed that there was a dramatic reduction of reduced GSH in both PA (~0.56 µM) and OHPA (~2.02 µM) than normal control (~11 µM). There was a slight reduction of GSH level for PYZ (~7.75 µM) (**Table 1**).

A similar trend was observed for SOD, where we also found that SOD level decreased to 20-30% both for PA and 5-OHPA as compared to the normal group. 5-OHPA revealed the highest toxicity than other toxicant. Similar observation was observed for CAT assay where we found that this enzyme activity was lower for both positive control and treated groups than normal control (**Table 1**).

Separately, we measured tissue MDA and PC formation to evaluate the oxidative stress caused by toxicant. The MDA formation was double for both PA (~119 nM) and 5-OHPA (~101 nM) than normal (~51 nM). Again, we observed that PC formation was higher for both PYZ metabolites (~ 0.19 and 0.17 µM for PA and OHPA) than normal control (~0.04 µM) (**Table 1**).

**Table 1:** Biochemical parameters determined to evaluate the liver toxicity effects produced by PYZ, PA, and 5-OHPA after oral administration for 28 days.

| # | Biochemical Parameters | Control | PYZ | PA | 5-OHPA |
|---|---|---|---|---|---|
| 1. | ALT (U/L) | 58.93 ± 2.70 | 94.29 ± 5.02** | 143.79 ± 7.70* | 117.28 ± 5.70** |
| 2. | AST (U/L) | 26.52 ± 1.77 | 51.86 ± 4.70*** | 58.93 ± 6.04* | 65.39 ± 7.70* |
| 3. | Bilirubin (ng/µg of protein) | 34.59±6.31 | 42.20±3.49 | 85.31±10.05*** | 116.09±11.08*** |
| 4. | Bilverdin (ng/µg of protein) | 14.00±3.99 | 18.86±3.98 | 23.41±2.23*** | 20.16±1.81 |
| 5. | SOD (U/µg of protein) | 8.33 ± 0.006 | 3.24 ± 0.03*** | 3.15 ± 0.05*** | 2.21 ± 0.04*** |
| 6. | CAT(nM of H2O2 /min/µg of protein) | 6.67 ± 1.59 | 1.30 ± 0.24*** | 1.26 ± 0.33*** | 1.12 ± 0.28*** |
| 7. | Reduced GSH (µM/µg of protein) | 11.37 ± 0.08 | 7.75 ± 0.62** | 0.56 ± 0.15*** | 2.02 ± 0.26*** |
| 8. | PC (µM/µg of protein) | 0.04 ± 0.009 | 0.15 ± 0.02*** | 0.19 ± 0.004*** | 0.17 ± 0.01*** |
| 9. | MDA (nM/µg of protein) | 51.81 ± 5.69 | 81.09 ± 4.84** | 119.91 ± 6.97** | 101.82 ± 5.60** |
| 10. | Liver tissue concentration (ng/µg of protein) | - | 42.09±5.06 | 59.54±13.60 | 146.77±11.22 |

Data represented as mean±SD (n=6). Statistically significant differences were observed between control and test groups [one way-ANOVA followed by Bonferroni multiple comparison tests (*p<0.05, **p<0.01, ***p<0.001)]. Parameters in rows 5-9 represent the oxidative stress parameters in the liver after oral administration for 28 days.

**Histopathology and SEM analysis of liver**

According to **Figure 1**(A), we observed kupffer (K) cells with the normal architecture of nucleus. In both positive toxic control and treated samples, there was the presence of degenerated nucleus (dN) in kupffer cells (K) and also ruptured kupffer cells (RC) **Figure 1**(B-D). However, the RC was present more in metabolites treated rats as compared to PYZ. The representative SEM images are shown in **Figure 1**(A'-D'). Consistent with histopathological analysis, SEM analysis also revealed that the rats dosed with PA and 5-OHPA exhibit more aberrant lesions in their liver than PYZ.

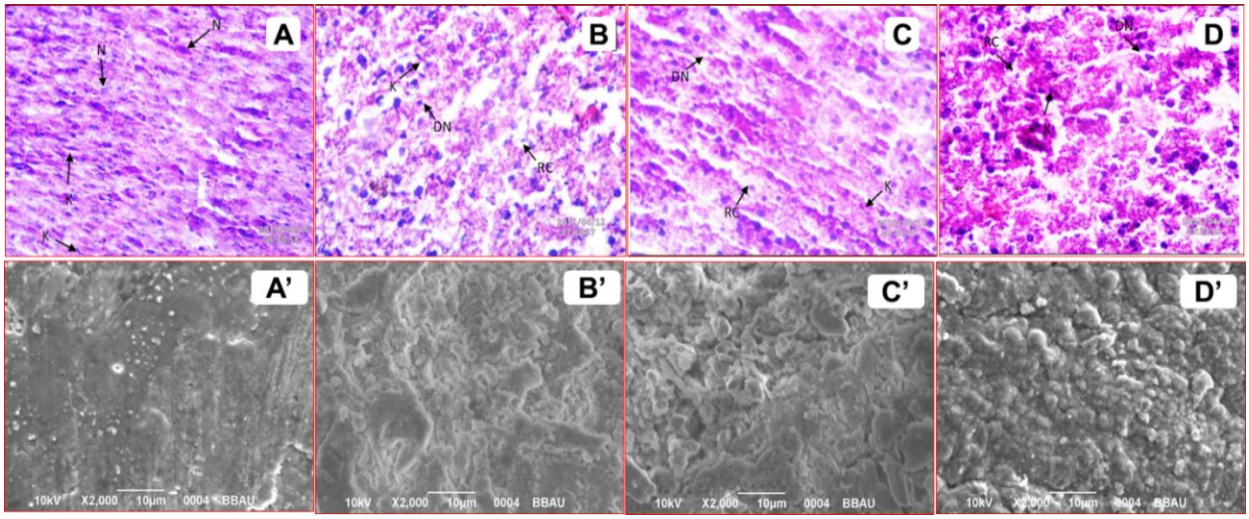



**Figure 1:** (**A-D**) Histopathology and (**A'-D'**) SEM analysis of liver: (**A/A'**) Control, (**B/B'**) PYZ, (**C/C'**) PA and (**D/D'**) 5-OHPA. Lesions were found in PYZ, PA and 5-OHPA groups with respect to control [Nucleus (N), Kupffer cell (K), degenerated nucleus (DN), and ruptured hepatic cells (RC)]. We observed prominent DN and RC in PYZ (toxic control) and 5-OHPA groups; these features were less prominent in PA group and absent in normal control group. As evident from SEM analysis, the hepatic lesions are increasingly prominent in PYZ, PA, and 5-OHPA groups with respect to control.

**Liver tissue deposition through HPLC**

A linear regression performed over a range of 1 to 250 ng/mL yielded a correlation coefficient (r2) of > 0.9 for all three analytes. The accuracy of the assay was found to be within 87-106%, and recovery of the samples was 67-85%. The retention time for PYZ, PA and 5-OHPA were 2.00, 2.87 and 1.88 min, respectively. As depicted in **Table 1**, the liver tissue deposition was higher for 5-OHPA (~146 ng/mL) than PA (~59 ng/mL) and PYZ (~ 42 ng/mL).

**Serum metabolomics to assess the biochemical effects of pyrazinamide and its metabolic products:**

The representative 1D $^{1}$H CPMG NMR spectra of rat serum samples with the assigned resonances of relevant metabolites are shown in **Figure 2**. The NMR spectra showed signals mainly from lipids/lipoproteins (e.g. low-density lipoprotein (LDL), very low-density lipoprotein (VLDL), polyunsaturated fatty acids (PUFAs) etc., and amino acids (e.g. alanine, valine, lysine, leucine, isoleucine, phenylalanine, histidine, tyrosine, glutamine, glutamate and proline etc.). Other identified metabolites were glucose, choline, creatine, creatinine, pyruvate, acetoacetate, acetate, citrate, lactate, N-acetyl and O-acetyl glycoproteins (NAG, OAG).

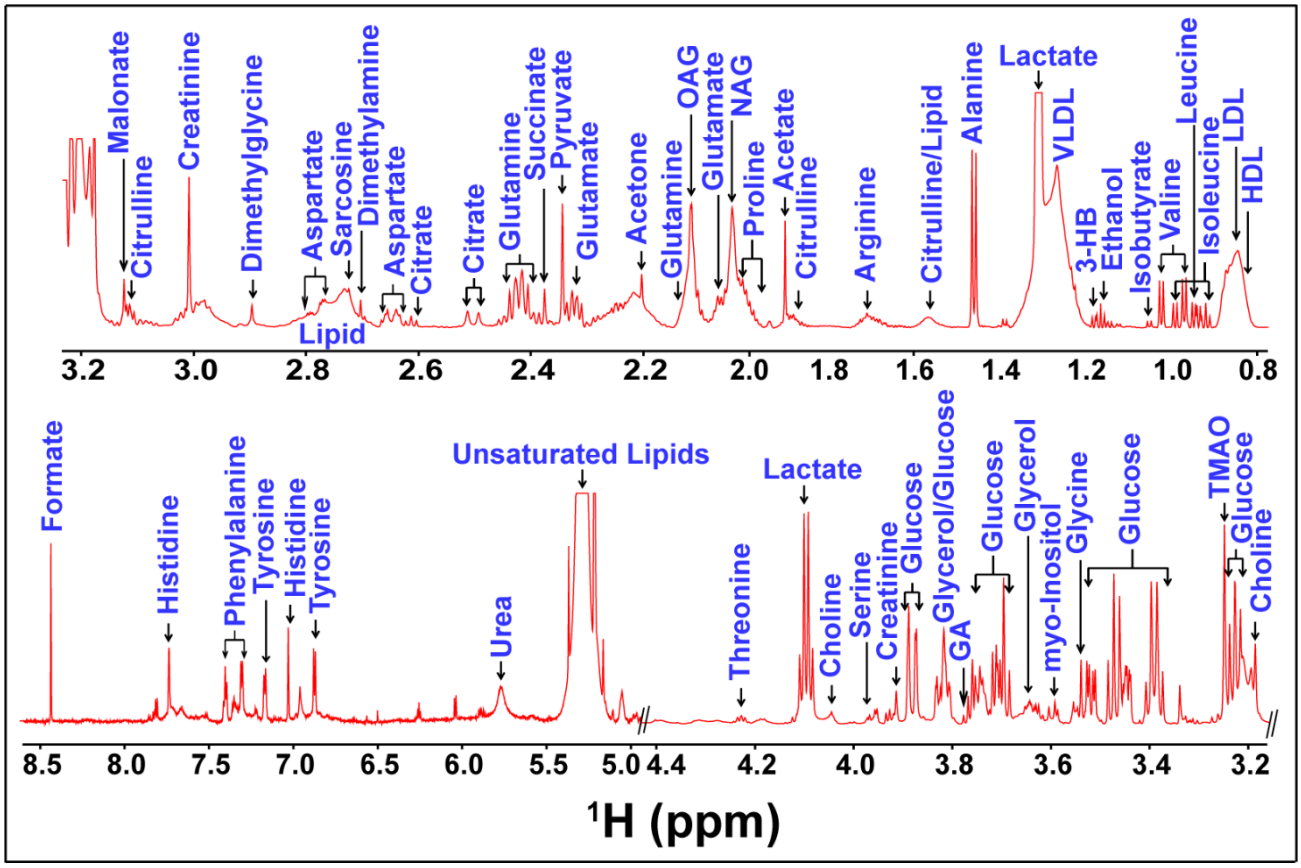



**Figure 2:** The representative 1D $^{1}$H CPMG NMR spectra of rat serum. The peaks annotated in the figure show the assignments of serum metabolites. The abbreviations used are LDL/VLDL: Low/very-low-density lipoproteins; HDL: high-density lipoproteins; PUFA: polyunsaturated fatty acids; NAG: N-acetyl glycoproteins; OAG: O-acetyl glycoprotein; GA: Guanidinoacetate; TMAO: Trimenthylamine-N-oxide.

**Multivariate data analysis**

The PCA score plots revealed the differences and separation among the groups (**Figure 3**). A supervised PLS-DA model was further used to discover the difference among groups. The applied PLS-DA to the 1D $^{1}$H NMR data was helpful to determine the extent of differences between the groups, as shown in **Figure 4**, derived from the CPMG spectra (**Figure 4A**) and DE spectra (**Figure 4E**). For determining the differences among the different classes, the pair-wise PLS-DA analysis was also performed with respect to the normal control group and the resulted 2D score plots highlighting the metabolic differences induced by the treatments: PYZ, PA and 5-OHPA, derived from the CPMG spectra (**Figure 4B-D**) and DE spectra (**Figure 4F-H**). As evident from the group separation in 2D PLS-DA score plots, the small serum metabolites are discriminating the groups more compared to lipid metabolites. As evaluated based on $R^2$ and $Q^2$ values, the discriminatory models were more robust for PYZ and 5-OHPA groups with respect to the controls than between PA and NC groups. In other words, the PYZ (dosed at 250 mg/kg) and 5-OHPA (dosed at 125 mg/kg) toxicity potential was higher than the PA as PA group is lying more close to normal control group than the PYZ and 5-OHPA groups when compared with respect to the

normal controls (**Figure 4**). To further evaluate the toxic effects of aforementioned treatments, the PA and 5-OHPA groups were separately compared to PYZ positive toxic control group, as shown in **Figure 5**. Pair-wise PCA and PLS-DA model analysis revealed that the PYZ and 5-OHPA overlapped to form an intermittent cluster whereas PA formed an independent discriminant cluster further suggesting that the PYZ and 5-OHPA are causing the same degree of toxicity. This is in concordant to our assumption as well where we have assumed that PYZ is metabolized in the liver into PA and 5-OHPA. If 250 mg dose of PYZ is metabolized into 125 mg 5-OHPA and 125 mg PA, and if 5-OHPA is the main toxic component, its toxicity potential at 125 mg/kg dose should reach to that of PYZ at 250 mg/kg dose. As reported previously, the PA is also metabolized to 5-OHPA by the enzyme xanthine oxidase. Therefore, the observed toxicity of PA at 125 mg/kg dose (as inferred in SEM and histopathological analysis) might also be partly because of 5-OHPA (a common metabolite of PYZ/PA metabolism). The observed metabolic response as evaluated based on PLS-DA discriminant model parameters ($R^2$, $Q^2$, and VIP scores), see **Figure 6, and 7**, further supports this hypothesis that the 125 mg/kg dose of 5-OHPA is almost producing the same effect as produced by 250 mg/kg dose of PYZ (i.e. toxic control group here) suggesting that the toxicity of PYZ could be mainly because of its metabolite 5-OHPA. The heat map analysis **Figure 8**, further corroborated this conjecture where the up-regulated and down-regulated metabolite entities are visually following the same pattern in 5-OHPA sample PYZ, whereas these are distinctly different from the normal control and PA groups.

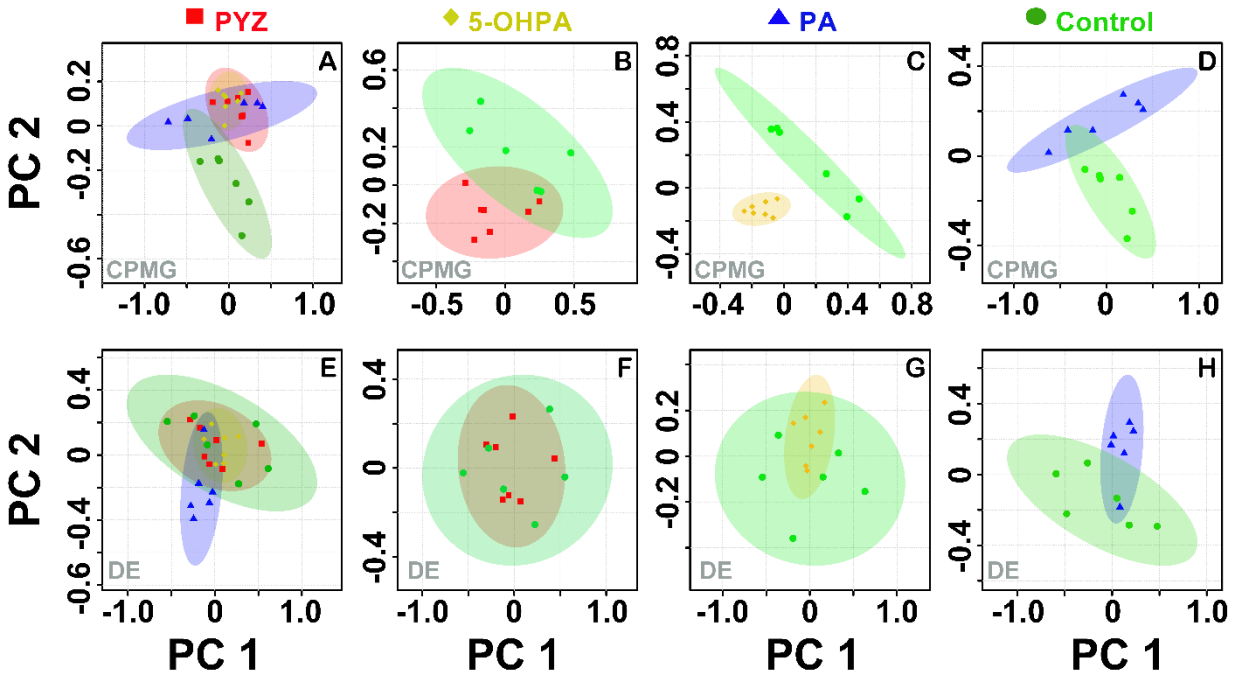



**Figure 3:** 2D PCA scores plots derived from 1D $^1$H (**A-D**) CPMG and (**E-H**) DE, NMR spectra. The various groups compared are well evident from the figure. (**A**) and (**E**) represent the PLS-DA analysis involving all the four groups, whereas (**B-D**) and (**F-H**) represent the paired analysis. The shaded areas are the 95% confidence regions of each treatment as depicted by their respective colors.

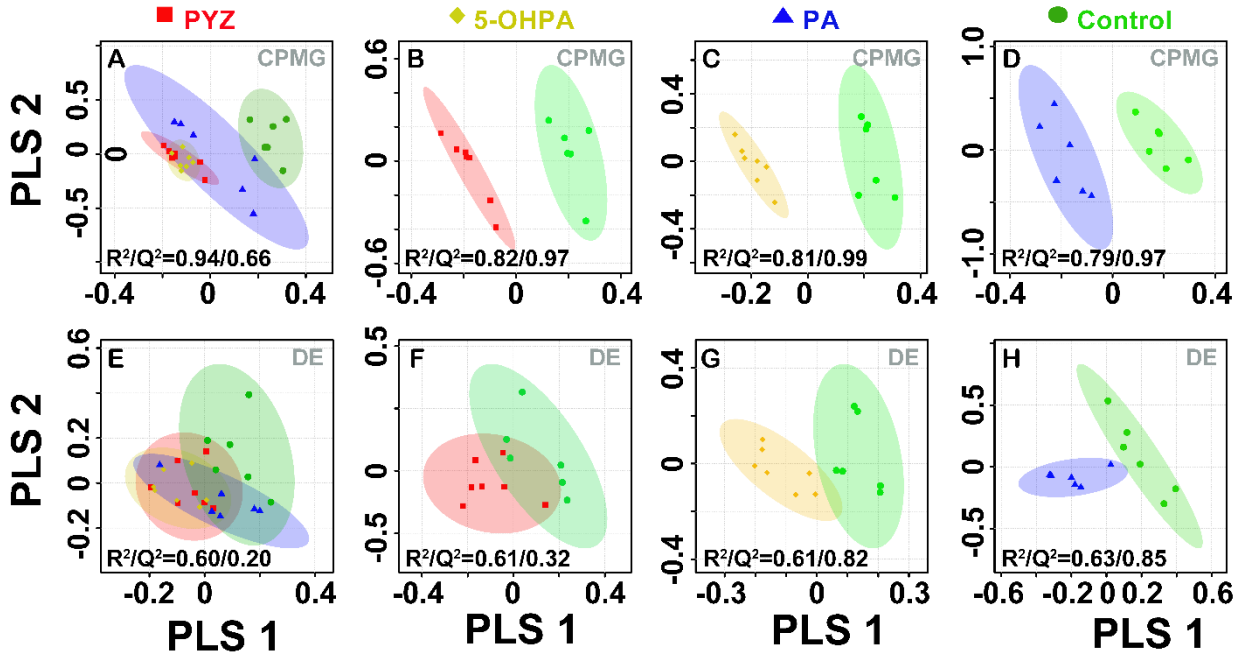



**Figure 4:** 2D PLS-DA scores plots derived from 1D $^{1}$H **(A-D)** CPMG and **(E-H)** DE, NMR spectra. The various groups compared are well evident from the figure. **(A)** and **(E)** represent the PLS-DA analysis involving all the four groups, whereas **(B-D)** and **(F-H)** represent the paired analysis. The shaded areas are the 95% confidence regions of each treatment as depicted by their respective colors.

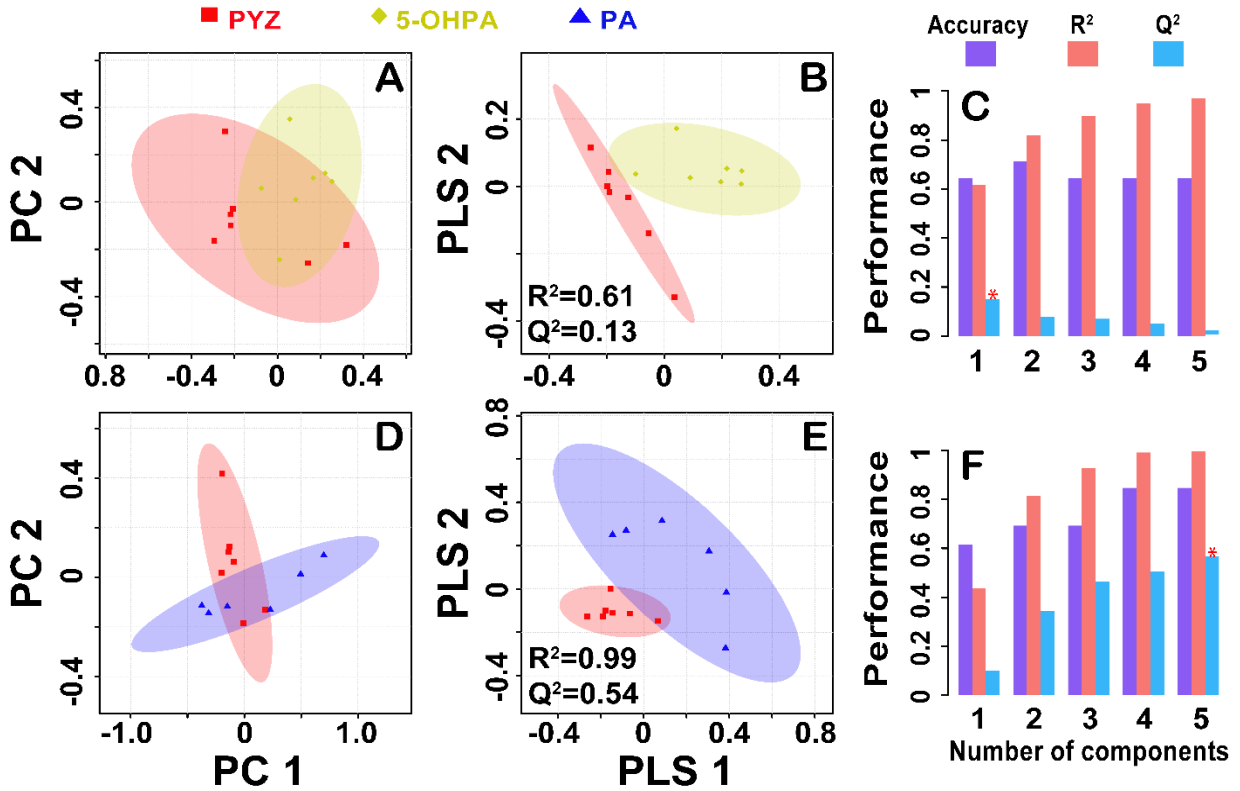



**Figure 5:** 2D scores plots derived from the PCA **(A,D)** and PLS-DA **(B,E)** analysis of 1D $^{1}$H CPMG NMR spectra. The treatment groups compared are well evident from the figure: **(A,B)** 5-OHPA vs. PYZ and **(D,E)** PA vs. PYZ. The shaded areas are the 95% confidence regions of each treatment as depicted by their respective colors. **(C,F)** Bar plots showing the three performance measures (prediction accuracy, $R^2$ and $Q^2$) using different number of components.

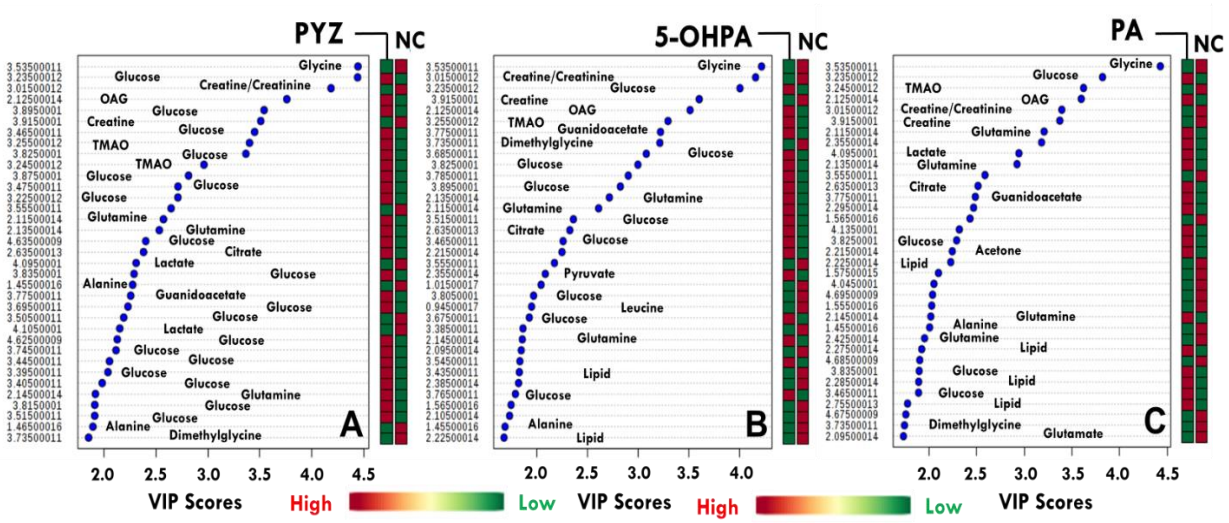



**Figure 6:** The potential biomarker metabolite entities identified from pair-wise PLS-DA analysis of 1D $^{1}$H CPMG spectra: **(A)** PYZ vs. NC, **(B)** 5-OHPA vs. NC and **(C)** PA vs. NC. The metabolites are listed in decreasing order of VIP score to highlight their discriminatory potential.

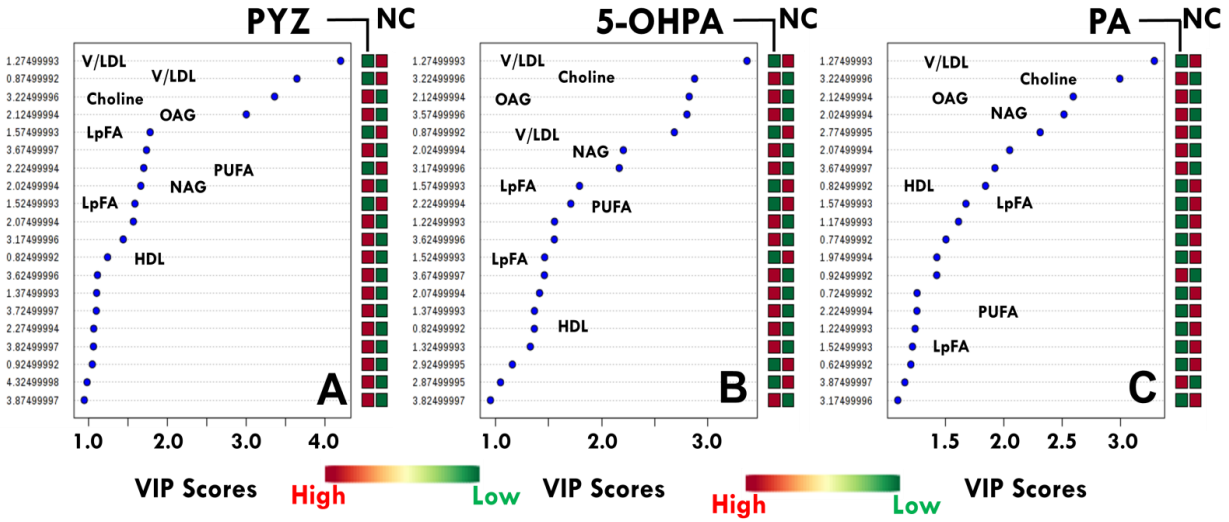



**Figure 7:** The potential biomarker metabolite entities identified from pair-wise PLS-DA analysis of 1D $^{1}$H diffusion edited NMR spectra: **(A)** PYZ vs. NC, **(B)** 5-OHPA vs. NC and **(C)** PA vs. NC. The metabolites are listed in decreasing order of VIP score to highlight their discriminatory potential.

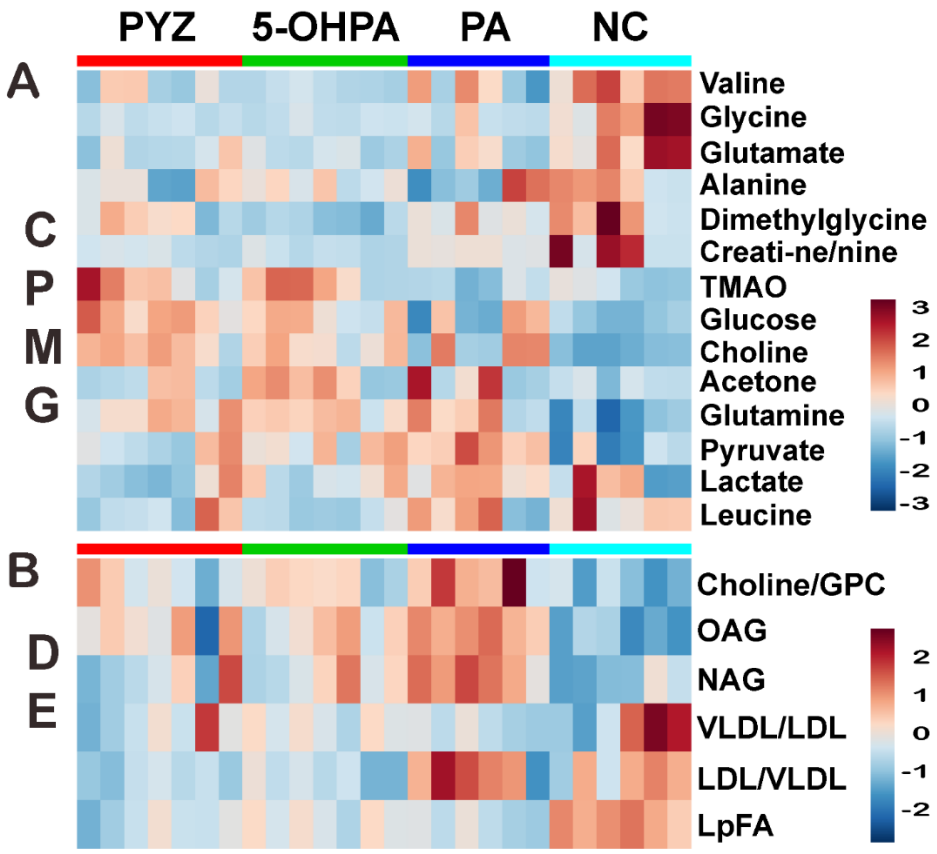



**Figure 8:** Heat maps of statistically significant metabolite entities as shown in **Table 2**. Here, (**A**) and (**B**) are the heat maps derived from discriminatory metabolite entities identified, respectively, in CPMG and diffusion edited (DE) spectra. The red and cyan here signify, respectively, elevation and reduction in metabolite concentration.

**Metabolic perturbations induced by PYZ, PA, and 5-OHPA treatments**

In order to evaluate the metabolic differences, the significantly altered metabolite entities were identified based on discriminant (PLS-DA) analysis of serum samples of normal control (NC) and PYZ dosed rats (PYZ). The perturbed metabolite entities responsible for class separation are mainly related to energy and lipid metabolism (**Table 2**). Compared to normal control rats, 19 serum metabolic markers were identified significantly perturbed in PYZ group as listed in **Table 2**. The serum levels of membrane metabolites (Choline/GPC), N-acetyl glycoproteins (NAG), O-acetyl glycoproteins (OAG), were found to be significantly elevated in PYZ dosed rats, whereas the serum levels of lipid-bound fatty acids (LpFA) and lipoproteins (LDL/VLDL) were decreased. These serum metabolic perturbations were relatively more predominant in rats dosed with 5-OHPA and PA compared to PYZ group (**Figure 9**, **Table 2**). Other serum perturbations involved increased levels of glucose, glutamine, pyruvate, citrate, and TMAO, whereas the serum levels of lactate, glutamate, alanine, creatine/creatinine, and branched chain amino-acids (leucine, valine, etc.) were significantly decreased in the PYZ group.

**Table 2:** Details of metabolites best describing the variation between PYZ, PA, and 5-OHPA administered group with respect to control group. The up (↑) and down (↓) arrows represent, respectively, increased and decreased metabolite levels.

| S.No | $^1$H (ppm) | Metabolite | PYZ vs NC | 5-OHPA vs NC | PA vs NC |
|---|---|---|---|---|---|
| 1 | **0.945** | Leucine | ↓↓ | ↓↓* | - |
| 2 | **1.015** | Valine | ↓↓* | ↓↓* | ↓* |
| 3 | **1.455** | Alanine | ↓↓ | ↓↓ | ↓ |
| 4 | **2.095** | Glutamate | ↓↓* | ↓↓* | ↓↓* |
| 5 | **2.145** | Glutamine | ↑↑ * | ↑↑ * | ↑↑ * |
| 6 | **2.215** | Acetone | -- | ↑ * | ↑ |
| 7 | **2.355** | Pyruvate | ↑↑ | ↑↑ * | ↑↑↑ * |
| 8 | **2.635** | Citrate | ↑↑ * | ↑↑ * | ↑↑ * |
| 9 | **3.015** | Creatine/Creatinine | ↓↓* | ↓↓* | ↓ |
| 10 | **3.255** | TMAO | ↑↑ * | ↑↑ * | - |
| 11 | **3.895,3.235** | Glucose | ↑↑↑ * | ↑↑ * | ↑ * |
| 12 | **3.535** | Glycine | ↓↓↓* | ↓↓* | ↓* |
| 13 | **3.735** | Dimethylglycine | ↓↓ | ↓↓↓* | ↓↓ |
| 14 | **4.105** | Lactate | ↓↓↓ | ↓↓ | ↑ |
| 15 | **0.875,1.275** | LDL/VLDL | ↓↓* | ↓↓* | ↑ |
| 16 | **1.575** | LpFA | ↓↓ | ↓↓* | ↓↓↓* |
| 17 | **2.025** | NAG | ↑ | ↑↑ * | ↑↑↑ * |
| 18 | **2.125** | OAG | ↑↑ * | ↑↑ * | ↑↑↑ * |
| 19 | **3.225** | Choline/GPC | ↑↑ * | ↑↑ * | ↑↑↑ * |

**Note:** * represents, p-Value <0.05; x - represents metabolite not significant in the group; The metabolites 1-15 belong to CPMG, and 15-19 belong to diffusion edited, 1D 1H NMR spectra. For visualization interpretation, single (↑,↓) double (↑↑,↓↓), and triple (↑↑↑,↓↓↓) arrows are used to represent high, higher and highest change in the mean value of metabolite concentration (as seen in their respective box-plots, **Figure 9**).

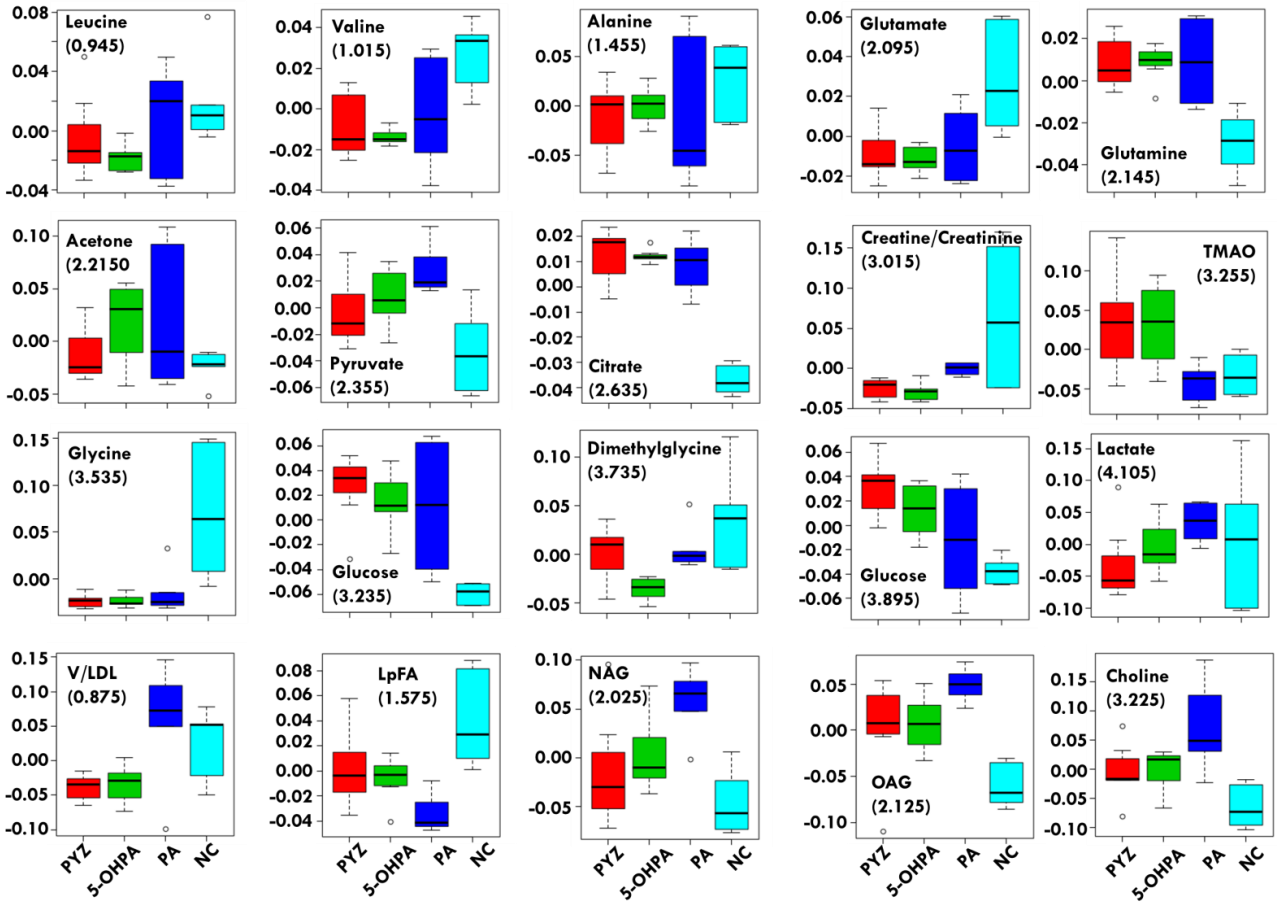



**Figure 9:** Box-whisker plots of metabolites that were significantly perturbed across the groups derived from CPMG spectra.

### Pathway Analysis

The metabolites significantly altered in the treatment samples compared to the control samples identified through PLS-DA analysis with a good VIP scores were used to determine the affected metabolic pathways using pathway analysis module, which is a combination of enrichment analysis and pathway topology analysis inbuilt in the MetaboAnalyst[23]. The lipid and membrane metabolites such as LDL, VLDL, HDL, lipids, NAG, OAG were not recognized by the program, thus were not included in this analysis. The final list of altered metabolites was uploaded and analyzed by Over-Representation Analysis (ORA) in MetaboAnalyst. The rat (Rattus norvegicus) pathway library, the hypergeometric test, and the out-degree centrality algorithms were employed for pathway enrichment analysis and pathway topology analysis. The pathway analysis module provided a fit coefficient (p) from pathway enrichment analysis and an impact factor from pathway topology analysis for each analyzed pathway. While the metabolite list employed (**Table 2**) is rather limited and provides only two or three metabolite hits for each pathway, the mapping does allow a ranking of the relative importance and identification of different possibilities. The output of this program will mark a metabolic pathway as significant if significantly more compounds involved in that pathway are present in the input list than would be

expected by random chance. All matched pathways are shown according to their p-values from the pathway enrichment analysis (vertical axis or y-axis, the intensity of color) and pathway impact values from pathway topology analysis (horizontal axis or x-axis, the size of circle), with the most impacted pathways colored in red as shown in **Figure 10**.

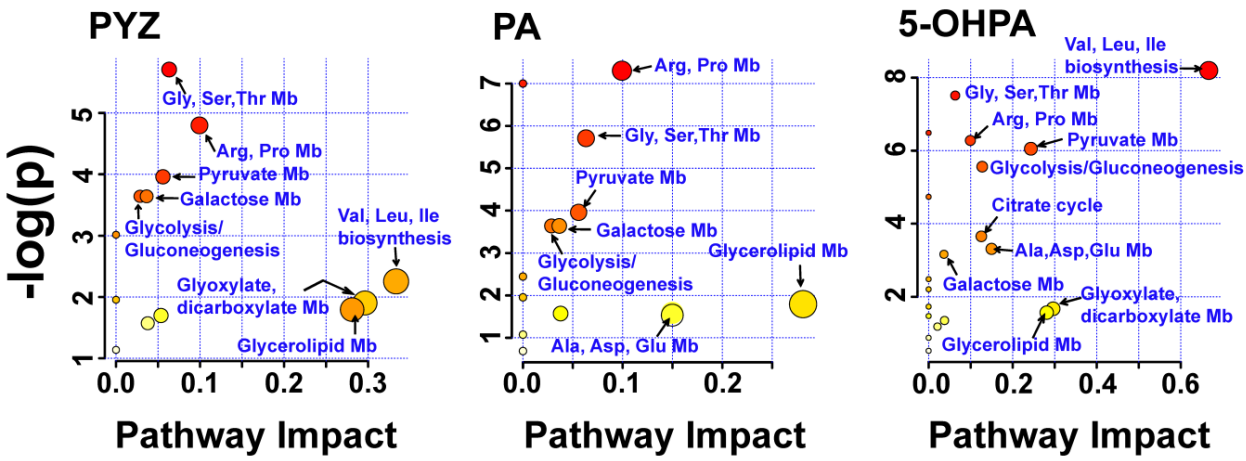



**Figure 10:** Identification of the perturbed metabolic pathways by overrepresentation analysis (ORA) using the significantly altered metabolites identified by PLSDA VIP score. The analysis was done by using a pathway library restricted to Rattus norvegicus, and p-values for ORA stand for hypergeometric test. Test p-value (vertical axis, the intensity of color) and impact factor (horizontal axis, the size of circle). ***Abbreviations used:*** Mb: Metabolism, Val: Valine, Leu: Leucine, Ile: Isoleucine, Asp: Aspartate, Ala: Alanine, Arg: Arginine, Ser: Serine, Thr: Threonine, Pro: Proline.

## Discussion

**Histopathological parameters revealed PYZ induced hepatotoxicity mechanism:**

The liver plays a prominent role in the lipid metabolism, dictating lipoprotein production, export, and clearance to and from the circulation[26]. The mechanism of hepatotoxicity induced by PYZ is still unknown to the researcher. A recent research[11] assumed that both PA and 5-OHPA were the main metabolites which might be responsible for PYZ induced hepatotoxicity. They observed the toxicity in vitro cell line (HepG2 cells) and found that both metabolites were more toxic than parent PYZ[11]. No further investigation had not been yet performed to evaluate whether PYZ is toxic or it's metabolites in vivo. To evaluate the mechanism of hepatotoxicity, we estimated serum AST and ALT levels of different groups. Our results collectively suggested that all the enzyme levels were higher for PA and 5-OHPA than PYZ and control. Both AST and ALT are the key enzymes of the liver which are extracted out during liver damage[27, 28]. The increase of these enzyme levels in serum during metabolites treatment indicated the hepatic damage caused by PA and 5-OHPA (**Table 1**). This was an indirect indication of liver damage. Therefore, we performed various biochemical estimation of the liver to evaluate the mechanism of toxicity. The PA and 5-

OHPA treated rats depleted more GSH than PYZ and control groups (**Table 1**). GSH is a tripeptide which is most abundant in all tissues including liver. GSH has a major role in the oxidation-reduction process, resulting in the formation of disulfide glutathione[16] during oxidative damage. Reduction of GSH level by these two metabolites is an indication of oxidative stress-induced liver damage.

To prove the oxidative damage induced hepatotoxicity, the tissue MDA levels were also performed. Oxidation of lipids is another important parameter to measure the oxidative stress in living body. Form the result; it was observed that tissue MDA level was higher for PA and 5-OHPA treated groups. In order to the relationship between oxidative stress and hepatotoxicity, PC assay was performed where a higher amount of PC was formed among metabolites treated groups. The carboxyl group of protein becomes oxidized due to the formation of reactive oxygen species[29] and converted to PC which is an important marker for oxidative stress. As evident, PA and 5-OHPA treated groups formed more PC than normal control. Both PC and MDA assay signified that oxidative damage occurred in liver cells during metabolites treatment.

For further measurement of oxidative stress, we measured both CAT and SOD enzyme levels in liver. The enzyme CAT is also most abundant in the liver which catalyzes the conversion of $H_2O_2$ to corresponding oxygen and water. This enzyme action is reduced due to the presence of peroxides and reactive oxygen species[30]. $H_2O_2$ levels were measured, and values were compared between various treated groups. The increase in concentration of $H_2O_2$ in PA and 5-OHPA treated rats depicted that there were less amount of CAT enzyme available in the tissue to decompose the $H_2O_2$ with respect to control and other treated groups (**Table 1**). This assay indirectly indicated that oral administration of PA and 5-OHPA reduced the level of CAT enzyme in the liver. Separately, the estimation of SOD levels in liver was performed. SOD is a free radical scavenging enzyme which neutralizes superoxide free radical in normal physiological situations[31]. Again, SOD levels were also decreased among the metabolites treated rats, but this enzyme level became normal in PYZ treated rats (**Table 1**). Therefore, we concluded that both metabolites reduced these enzyme levels during oxidative damage in the liver which was less prominent during PYZ administration.

Bilirubin and biliverdin are the liver pigment whose concentrations were increased during liver damage. PYZ inhibited uptake and excretion of bilirubin in a dose-dependent manner and increased hyperbilirubinemia, due to blockade of uptake of plasma membrane of the hepatocytes. Bilirubin caused damage of liver cells via bleaching action[32]. When we administered PYZ and its metabolites, we observed that both metabolites increased the concentration of conjugated bilirubin in the liver. PA and 5-OHPA treated rats increased this concentration twice and thrice than normal control, respectively (**Table 1**), produced hyperbilirubinemia and liver damage through bleaching action.

For direct evidence of liver damage, histopathology and SEM analyses of liver were performed (**Figure 1**). The results obtained from histopathological studies reflected the toxicity caused by PA and 5-OHPA and showed a remarkable damage of the architecture of liver cells. The present detected necrosis of the hepatic cells of rats treated with metabolites might be attributed to the formation of free radicals, which acted as a stimulator of lipid peroxidation and protein oxidation and simultaneously destruction of the cell membrane (**Figure 1**). This action was less prominent in PYZ treated rats. SEM analysis of liver tissue also supported our hypothesis, and we observed lesions in PA and 5-OHPA treated rats which were also less prominent in PYZ treated groups.

Finally, we determined liver tissue deposition studies via single oral administration of PYZ, PA and 5-OHPA separately to albino Wistar rats and quantified using HPLC. The highest tissue deposition was observed for 5-OHPA>PA>PYZ (**Table 1**). This assay signified that both metabolites were more deposited in the liver and responsible for toxicity. The possible reason for the higher degree of liver tissue deposition and associated hepatotoxicity with 5-OHPA could be its predominance existence in the keto form which imparts lipophilic nature to the metabolite and thus complications associated with PYZ. To confirm this hypothesis, future studies are required which will compare the toxicity of PYZ/5-OHPA with 5-methyl-PYZ and 5-fluoro-PA.

**Serum Metabolic changes useful insights into PYZ induced hepatotoxicity:**

Our investigation identified 19 metabolites that were perturbed in the treatment group as compared to control (**Table 2**). Consistent with previous NMR based metabolomics studies involving rat serum[33], the serum levels of choline/GPC, NAG and OAG were found to be elevated indicating an inflammatory mechanism prevailing in PYZ dosed rats and so in the case of rats dosed with PYZ metabolites. The depleted level of the total lipid is caused by the severity of the liver disease and has been reported in a number of studies associated with liver diseases such as Nonalcoholic fatty liver disease (NAFLD), alcoholic liver disease, hepatitis C, hepatitis B, cholestatic liver disease and cirrhosis, etc[34]. Consistent with these reports, in our NMR study, we also found significantly decreased serum levels of apolipoproteins (LDL/VLDL) in PYZ, PA and 5-OHPA groups compared to control group, whereas choline was found to be elevated in the sera. Similar lipid dysregulation has also been found to be associated with a number of liver diseases and in cancer suggesting an altered lipid metabolism[35, 36]. Cell toxicity may involve inflammation and oxidative stress further leading to lipid peroxidation[28]. The function of LDL/VLDL is to deliver cholesterol to cells, where it is used in the membrane synthesis. Choline, an important constituent of the cell membrane and phospholipid metabolism, are a breakdown product of phosphatidylcholine. Therefore, the increase in serum choline and phosphocholine concentrations, together with a decreased serum level of lipids and lipoproteins, demonstrate the liver toxicity is related to significant disruption of the cell membrane and the increased lipid utilization in the synthesis of membrane metabolites (Choline/GPC) to sustain the membrane repair.

A general comparison of the control group with PYZ, PA and 5-OHPA groups also showed a significant increase in glucose accompanied by lactate, suggesting disturbed glucose metabolism with dampened aerobic glycolytic activity in liver toxicity. With respect to normal controls, the most significant changes in energy metabolites were observed in PYZ and 5-OHPA compared to PA. Hyperglycemia results in oxidative stress, insulin resistance, glucose intolerance and diabetes[37]. Oxidative stress results in structural and functional abnormalities affecting multiple metabolic pathways. During oxidative stress the glucose metabolism shifts from aerobic to anaerobic pathway to meet its energy demands. However, the process being less efficient cannot keep up with the production and utilization of adenosine triphosphate (ATP), the intracellular concentration of ATP drops and the depletion of cytosolic glucose. Hence the metabolism switches from glycogenolysis to gluconeogenesis. Consistent with these reports, our results also indicated an increased level of tricarboxylic acid (TCA) cycle intermediates i.e. citrate and pyruvate in the sera of PYZ and 5-OHPA group. Citrate serves as a regulator of glycolysis and gluconeogenesis and thus regulates blood glucose levels[38]. Increased levels of citrate have also been reported in cirrhosis and hepatocellular carcinoma, attributed to a reduced capacity for citrate clearance in cirrhotic patients[39, 40], and also involved in the metabolic pathway of de novo lipid biosynthesis[41]. The relatively high levels of pyruvate in the serum might be attributed to energy metabolism alterations of the TCA cycle and glycolysis. Along with this, ketone bodies (acetone and acetate) were also found to be elevated. The high levels of ketone bodies reported in serum could be due to increased energy requirements. Liver plays a cardinal role in maintaining the serum glucose levels, excess glucose is converted into glycogen or fatty acids which are further processed and utilized by the body[42, 43]. Due to impaired liver function, liver's efficacy in removing the excess glucose from the serum is dampened[44]. Alterations in glucose homeostasis have been associated with increased severity of liver disease and an elevated risk of liver carcinoma. The up regulation of glucose in serum is also accompanied by a down regulation of glycerol that can be converted to glucose in the liver and provides energy for cellular metabolism, which suggested that the rates of glycogenolysis and glycolysis increased because of inhibited lipid metabolism in these animals. The cumulative effects of all these pathologies seem to be frequently associated with alcoholic, NAFLD, cirrhosis and hepatitis C virus (HCV)[45, 46]. Serum creatine and creatinine levels was also significantly reduced. Creatine is synthesized primarily in the kidney, pancreas, and liver; a great deal of creatine present in the body is in the phosphorylated form of phosphocreatine in the muscles and act as an instant source of energy. The discrepancy in creatine and creatinine level accords well with the interplay between liver and renal dysfunction in liver disease. Moreover, creatine has antioxidant properties[47], and its diminished levels could be pertained to oxidative stress. Along with this, guanidoacetate (GA) which is a major constituent in the synthesis of creatine was also found to be downregulated[48]. GA is readily converted to creatine in rat hepatocytes in collaboration with the kidney enzymatically[49]. The drug-induced toxicity to liver enzymes might be responsible for the decrease level of GA and creatine.

Amino acids (AA) are the key metabolites involved in a number of chemical and physiological processes and play an impeccable role in catabolism and anabolism[50]. Not only to this but they are also equally important for protein biosynthesis as well as biosynthesis of several biogenic amines essential for survival in conditions of acute stress[51]. The disturbances of amino acids reflect cellular needs for a higher turnover of structural proteins in maintaining energy homeostasis under hypoxic conditions. Under oxidative stress, the metabolism switches from aerobic to an anaerobic pathway to carry out the vital cellular processes. The lipids and carbohydrates cannot be utilized as an energy source due to the limited supply of oxygenated blood in a hypoxic environment, the reliance on alternative energy substrates (amino acids) increases which can be oxidized anaerobically with a lower contribution to acidosis. Our results were consistent with these studies that leucine, valine, glutamate, proline, and alanine, were markedly decreased in 5-OHPA treated group. The levels of glucogenic amino and ketogenic amino acids (glutamate, alanine, valine and leucine) were found to be dampened. It is an important part for excretion of the nitrogenous waste and a key energy metabolite to be metabolized during oxidative stress condition. It is considered to be biologically very important for maintenance and promotion of cell function. The demands of the body for glutamate and glutamine are enormous during critical illness, and their blood concentration may rise and fall during the recovery phase[52]. Consistent with this, as reported in a number of studies, serum glutamine, and glutamate ratio may be a useful method to estimate the pathophysiological state of patients with liver diseases[53]. Decreased level of proline suggests altered collagen metabolism as proline is the end product of collagen metabolism and excess amino acids are stored in collagen; during oxidative stress, collagen is released and used to generate energy in the time of crisis. Proline metabolism is of particular importance in nutrient stress, as it is interchangeably converted into glutamate and glutamine[54]. The low serum levels of amino acids have also been shown to be closely related to protein-energy wasting, inflammation, and oxidative stress. The up and down regulated metabolites, thus suggest perturbed glycolysis and lipid metabolism in liver disease consistent with a number of studies[55].

Using $^{1}H$ NMR diffusion edited data from serum, the NAG, and OAG were found to be elevated in the spectra of serum from the treatment group compared with the control group. The O-acetylated carbohydrate-bound protein resonance found in rat blood serum, are alternate "acute-phase" glycoproteins in animal models of human inflammatory conditions[56]. Acetyl-glycoproteins (both N and O) are acute phase proteins, acting as inflammatory mediators and could be a response to tissue damage, and thus, the increased concentrations of serum NAG and OAG were likely to reflect an inflammatory response. Elevated levels of NAG and OAG in blood serum of oxidatively stressed animals are consistent with previous investigations of the metabolic response to stress[57]. The concentrations of serum "acute phase" NAG are markedly elevated in a range of abnormal clinical conditions, including inflammatory disease, cancer, and certain liver diseases[58]. The present observations provide supportive evidence for systemic inflammation

associated with liver injury, in agreement with the actions of other serum inflammatory markers. Thus, increased levels of NAG and OAG are likely to indicate hyperglycemia-induced inflammation in our study.

**Biomedical Relevance:**

An individual's susceptibility to drug toxicity can be predicted from the metabolomic analysis of urine or serum samples of the individual before drug exposure. One of the major areas of research in pharmaceutical drug discovery is directed towards the identification of biomarkers of drug toxicity, efficacy, and selectivity that can be used in pre-clinical and clinical studies of drug[59]. There is a substantial need to identify and develop new diagnostic and prognostic biomarkers that can precisely anticipate toxicity in the preclinical development of chemical entities early in the drug development process. In this regard, the NMR based metabolomics – involving bio-fluids collected through minimally invasive procedures- is currently the technique of choice for rapid screening of biochemical effects induced in response to a drug. The identified metabolic patterns may be used as early biomarkers of toxicity as well[60]. Therefore, we believe that the metabolic disturbances as evaluated here related to drug-induced liver injury will be of potential biomedical relevance to guide the future clinical studies aiming to evaluate the efficacy and safety of new drug candidates.

Liver, being the principal site for the metabolism of the drugs, makes it primarily targeted sites of drug toxicity. The drug itself or its metabolic products can be toxic making the organ susceptible to injury and may play an important role in the development of anti-TB drug-induced liver injury (DILI). There are no conventional road maps for the treatment and management of tuberculosis in correlation to the severity of liver disease. Identification of risk factors associated with DILI is important; as no specific treatment exists for the prevention or treatment of hepatotoxicity, and also no specific early diagnostic biomarkers exists. However, the available once like microRNAs, cytokeratin-18 (CK18), and high mobility group box protein 1 (HMGB-1) have not yet been approved for mundane clinical use[61]. A high risk of multidrug-resistant TB is another major issue due to prolonged and interrupted treatment[62]. The frequency of hepatotoxicity is increased in patients with liver dysfunction, frequently leading to severe liver failure culminating in death or liver transplantation. Henceforth, there has an utter need for novel diagnostic and prognostic biomarkers of early liver injury, for treatment monitoring and overall management of the drug therapy especially in an endemic area for TB and liver disease[63]. The study revealed that the rats dosed with 5-OHPA showed most severe changes in liver histopathology as well as most prominent and aberrant metabolic changes similar to PYZ with respect to PA and normal control rats. The major serum metabolic changes observed in rats dosed with 5-OHPA were: increased levels of glucose, pyruvate, citrate, glutamine, N- and O- acetyl glycoproteins and ketone bodies along with the decreased level of lipids, 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.

## Conclusions

The current study reveals that the drug PYZ itself is not responsible for the hepatotoxicity; however, two of its metabolic product PA and 5-OH-PA are the major culprit for the drug-induced liver injury. Off which 5-OHPA is found to be the most toxic metabolic product of the drug PYZ, based on biochemical estimation, histopathological and SEM studies. The metabolic product is found to induce liver injury due to oxidative stress and inflammation. Metabolomics studies also supported that the 5-OHPA is more toxic than the parent drug compound. The present study has implication in human studies; patient stratification, decision making, and personalized medicine. As drug-induced liver injury is a major complication associated with many treatment procedures in critical care and routine clinical management. NMR 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. 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.

Based on various observations, we believe that the metabolic products of PYZ –i.e. PA and 5-OHPA might be responsible for free radical generation and followed by oxidative stress induced liver damage and of which 5-OHPA may be the main culprit for PYZ induced hepatotoxicity. Therefore, the final conclusion was that PYZ metabolites might be responsible for hepatotoxicity.

**Reference's**


1. Tostmann, A.; Boeree, M. J.; Aarnoutse, R. E.; De Lange, W.; Van Der Ven, A. J. A. M.; Dekhuijzen, R., Antituberculosis drugâ€• induced hepatotoxicity: concise upâ€• toâ€• date review. *Journal of gastroenterology and hepatology* **2008**, 23, (2), 192-202.

2. De Groot, L. J.; Beck-Peccoz, P.; Chrousos, G.; Dungan, K.; Grossman, A.; Hershman, J. M.; Koch, C.; McLachlan, R.; New, M.; Rebar, R., Metabolic Syndrome--Endotext. **2000**.

3. Wu, J.-W.; Shih, H.-H.; Wang, S.-C.; Tsai, T.-H., Determination and pharmacokinetic profile of pyrazinamide in rat blood, brain and bile using microdialysis coupled with high-performance liquid chromatography and verified by tandem mass spectrometry. *Analytica chimica acta* **2004**, 522, (2), 231-239.

4. Espinal, M. A.; Kim, S. J.; Suarez, P. G.; Kam, K. M.; Khomenko, A. G.; Migliori, G. B.; BaÃ©z, J.; Kochi, A.; Dye, C.; Raviglione, M. C., Standard short-course chemotherapy for drug-resistant tuberculosis: treatment outcomes in 6 countries. *Jama* **2000**, 283, (19), 2537-2545.

5. Girling, D. J., The hepatic toxicity of antituberculosis regimens containing isoniazid, rifampicin and pyrazinamide. *Tubercle* **1977**, 59, (1), 13-32.

6. Durand, F. o.; Jebrak, G.; Pessayre, D.; Fournier, M.; Bernuau, J., Hepatotoxicity of antitubercular treatments. *Drug Safety* **1996**, 15, (6), 394-405.

7. Tostmann, A.; Boeree, M. J.; Peters, W. H. M.; Roelofs, H. M. J.; Aarnoutse, R. E.; van der Ven, A. J. A. M.; Dekhuijzen, P. N. R., Isoniazid and its toxic metabolite hydrazine induce in vitro pyrazinamide toxicity. *International journal of antimicrobial agents* **2008**, 31, (6), 577-580.

8. Awofeso, N., Anti-tuberculosis medication side-effects constitute major factor for poor adherence to tuberculosis treatment. *Bulletin of the World Health Organization* **2008**, 86, (3), B-D.

9. Lacroix, C.; Tranvouez, J. L.; Phan, H. T.; Duwoos, H.; Lafont, O., Pharmacokinetics of pyrazinamide and its metabolites in patients with hepatic cirrhotic insufficiency. *Arzneimittel-Forschung* **1990**, 40, (1), 76-79.

10. Iruzubieta, P.; Arias-Loste, M. T.; Barbier-Torres, L. a.; Martinez-Chantar, M. L.; Crespo, J., The need for biomarkers in diagnosis and prognosis of drug-induced liver disease: does metabolomics have any role? *BioMed research international* 2015.

11. Shih, T.-Y.; Pai, C.-Y.; Yang, P.; Chang, W.-L.; Wang, N.-C.; Hu, O. Y.-P., A novel mechanism underlies the hepatotoxicity of pyrazinamide. *Antimicrobial agents and chemotherapy* 57, (4), 1685-1690.

12. Zhang, Y.; Jiang, Z.; Su, Y.; Chen, M.; Li, F.; Liu, L.; Sun, L.; Wang, Y.; Zhang, S.; Zhang, L., Gene expression profiling reveals potential key pathways involved in pyrazinamideâ€• mediated hepatotoxicity in Wistar rats. *Journal of Applied Toxicology* 33, (8), 807-819.

13. Kushwaha, P. S.; Raj, V.; Singh, A. K.; Keshari, A. K.; Saraf, S. A.; Mandal, S. C.; Yadav, R. K.; Saha, S., Antidiabetic effects of isolated sterols from Ficus racemosa leaves. *RSC Advances* 5, (44), 35230-35237.

14. Makos, B. K.; Youson, J. H., Tissue levels of bilirubin and biliverdin in the sea lamprey, Petromyzon marinus L., before and after biliary atresia. *Comparative Biochemistry and Physiology Part A: Physiology* **1988**, 91, (4), 701-710.

15. Lodhi, R. L.; Maity, S.; Kumar, P.; Saraf, S. A.; Kaithwas, G.; Saha, S., Evaluation of mechanism of hepatotoxicity of leflunomide using albino wistar rats. *African Journal of Pharmacy and Pharmacology* 7, (24), 1625-1631.

16. Saha, S.; Chan, D. S. Z.; Lee, C. Y.; Wong, W.; New, L. S.; Chui, W. K.; Yap, C. W.; Chan, E. C. Y.; Ho, H. K., Pyrrolidinediones reduce the toxicity of thiazolidinediones and modify their anti-diabetic and anti-cancer properties. *European journal of pharmacology* 697, (1), 13-23.

17. Ong, M. M. K.; Latchoumycandane, C.; Boelsterli, U. A., Troglitazone-induced hepatic necrosis in an animal model of silent genetic mitochondrial abnormalities. *Toxicological Sciences* **2007**, 97, (1), 205-213.

18. Guleria, A.; Bajpai, N. K.; Rawat, A.; Khetrapal, C. L.; Prasad, N.; Kumar, D., Metabolite characterisation in peritoneal dialysis effluent using highâ€• resolution 1H and 1Hâ€“13C NMR spectroscopy. *Magnetic Resonance in Chemistry* 52, (9), 475-479.

19. Nicholson, J. K.; Foxall, P. J. D.; Spraul, M.; Farrant, R. D.; Lindon, J. C., 750 MHz 1H and 1H-13C NMR spectroscopy of human blood plasma. *Analytical chemistry* **1995**, 67, (5), 793-811.

20. Wishart, D. S.; Jewison, T.; Guo, A. C.; Wilson, M.; Knox, C.; Liu, Y.; Djoumbou, Y.; Mandal, R.; Aziat, F.; Dong, E., HMDB 3.0â€”the human metabolome database in 2013. *Nucleic acids research*, gks1065.

21. Liu, M.; Tang, H.; Nicholson, J. K.; Lindon, J. C., Use of 1H NMRâ€• determined diffusion coefficients to characterize lipoprotein fractions in human blood plasma. *Magnetic Resonance in Chemistry* **2002**, 40, (13).

22. Malik, A.; Sharma, U.; Lakshmy, R.; Narang, R.; Jagannathan, N. R., Biochemical characterization of blood plasma of coronary artery disease patients by in vitro high-resolution proton NMR spectroscopy. *Journal of biosciences* 40, (1), 31-39.

23. Xia, J.; Sinelnikov, I. V.; Han, B.; Wishart, D. S., MetaboAnalyst 3.0â€”making metabolomics more meaningful. *Nucleic acids research* 43, (W1), W251-W257.

24. Xia, J.; Wishart, D. S., Web-based inference of biological patterns, functions and pathways from metabolomic data using MetaboAnalyst. *Nature protocols* 6, (6), 743-760.

25. Chong, I.-G.; Jun, C.-H., Performance of some variable selection methods when multicollinearity is present. *Chemometrics and Intelligent Laboratory Systems* **2005**, 78, (1), 103-112.

26. Patterson, A. D.; Maurhofer, O.; BeyoÄŸlu, D.; Lanz, C.; Krausz, K. W.; Pabst, T.; Gonzalez, F. J.; Dufour, J.-F. o.; Idle, J. R., Aberrant lipid metabolism in hepatocellular carcinoma revealed by plasma metabolomics and lipid profiling. *Cancer research* 71, (21), 6590-6600.

27. Zhao, H.; Si, Z.-H.; Li, M.-H.; Jiang, L.; Fu, Y.-H.; Xing, Y.-X.; Hong, W.; Ruan, L.-Y.; Li, P.-M.; Wang, J.-S., Pyrazinamide-induced hepatotoxicity and gender differences in rats as revealed by a 1 H NMR based metabolomics approach. *Toxicology Research* 6, (1), 17-29.

28. Wei, S.; Suryani, Y.; Gowda, G. A.; Skill, N.; Maluccio, M.; Raftery, D., Differentiating hepatocellular carcinoma from hepatitis C using metabolite profiling. *Metabolites* 2, (4), 701-716.

29. Suzuki, Y. J.; Carini, M.; Butterfield, D. A., Protein carbonylation. *Antioxidants & redox signaling* 12, (3), 323-325.

30. Baudrimont, I.; Ahouandjivo, R.; Creppy, E. E., Prevention of lipid peroxidation induced by ochratoxin A in Vero cells in culture by several agents. *Chemico-biological interactions* **1997**, 104, (1), 29-40.

31. Karaman, A.; Iraz, M.; Kirimlioglu, H.; Karadag, N.; Tas, E.; Fadillioglu, E., Hepatic damage in biliary-obstructed rats is ameliorated by leflunomide treatment. *Pediatric surgery international* **2006**, 22, (9), 701.

32. Rohatagi, S.; Carrothers, T. J.; Kshirsagar, S.; Khariton, T.; Lee, J.; Salazar, D., Evaluation of Population Pharmacokinetics and Exposureâ€• Response Relationship With Coadministration of Amlodipine Besylate and Olmesartan Medoxomil. *The Journal of Clinical Pharmacology* **2008**, 48, (7), 823-836.

33. Beckwith-Hall, B. M.; Thompson, N. A.; Nicholson, J. K.; Lindon, J. C.; Holmes, E., A metabonomic investigation of hepatotoxicity using diffusion-edited 1 H NMR spectroscopy of blood serum. *Analyst* **2003**, 128, (7), 814-818.

34. MohammadReza, G.; AliAkbar, R.; Abbas, H.; Mehrdad, N.; AbbasAli, H., The relationship between lipid profile and severity of liver damage in cirrhotic patients. *Hepatitis monthly* 2010, (4, Autumn), 285-288.

35. Garcia-Compean, D.; Jaquez-Quintana, J. O.; Gonzalez-Gonzalez, J. A.; Maldonado-Garza, H., Liver cirrhosis and diabetes: risk factors, pathophysiology, clinical implications and management. *World J Gastroenterol* **2009**, 15, (3), 280-288.

36. Zhang, X.; Xu, L.; Shen, J.; Cao, B.; Cheng, T.; Zhao, T.; Liu, X.; Zhang, H., Metabolic signatures of esophageal cancer: NMR-based metabolomics and UHPLC-based focused metabolomics of blood serum. *Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease* 1832, (8), 1207-1216.

37. Ganapathy, V.; Fei, Y., Biological role and therapeutic potential of NaCT, a sodium-coupled transporter for citrate in mammals. *PROCEEDINGS-INDIAN NATIONAL SCIENCE ACADEMY PART B* **2005**, 71, (2), 83.

38. Kramer, L.; Bauer, E.; Joukhadar, C.; Strobl, W.; Gendo, A.; Madl, C.; Gangl, A., Citrate pharmacokinetics and metabolism in cirrhotic and noncirrhotic critically ill patients. *Critical care medicine* **2003**, 31, (10), 2450-2455.

39. Costello, L. C.; Franklin, R. B., A review of the important central role of altered citrate metabolism during the process of stem cell differentiation. *Journal of regenerative medicine & tissue engineering* 2.

40. Zheng, Y.; Xu, Z.; Zhu, Q.; Liu, J.; Qian, J.; You, H.; Gu, Y.; Hao, C.; Jiao, Z.; Ding, F., Citrate pharmacokinetics in critically ill patients with acute kidney injury. *PloS one* 8, (6), e65992.

41. Barthel, A.; Schmoll, D., Novel concepts in insulin regulation of hepatic gluconeogenesis. *American Journal of Physiology-Endocrinology And Metabolism* **2003**, 285, (4), E685-E692.

42. Levinthal, G. N.; Tavill, A. S., Liver disease and diabetes mellitus. *Clinical diabetes* **1999**, 17, (2), 73.

43. Postic, C.; Dentin, R.; Girard, J., Role of the liver in the control of carbohydrate and lipid homeostasis. *Diabetes & metabolism* **2004**, 30, (5), 398-408.

44. CichoÅ¼-Lach, H.; Michalak, A., Oxidative stress as a crucial factor in liver diseases. *World journal of gastroenterology: WJG* 20, (25), 8082.

45. Custro, N.; Carroccio, A.; Ganci, A.; Scafidi, V.; Campagna, P.; Di Prima, L.; Montalto, G., Glycemic homeostasis in chronic viral hepatitis and liver cirrhosis. *Diabetes & metabolism* **2001**, 27, (4 Pt 1), 476-481.

46. Paris, D.; Melck, D.; Stocchero, M.; Dâ€™Apolito, O.; Calemma, R.; Castello, G.; Izzo, F.; Palmieri, G.; Corso, G.; Motta, A., Monitoring liver alterations during hepatic tumorigenesis by NMR profiling and pattern recognition. *Metabolomics* 6, (3), 405-416.

47. Sestili, P.; Martinelli, C.; Colombo, E.; Barbieri, E.; Potenza, L.; Sartini, S.; Fimognari, C., Creatine as an antioxidant. *Amino acids* 40, (5), 1385-1396.

48. da Silva, R. P.; Nissim, I.; Brosnan, M. E.; Brosnan, J. T., Creatine synthesis: hepatic metabolism of guanidinoacetate and creatine in the rat in vitro and in vivo. *American Journal of Physiology-Endocrinology And Metabolism* **2009**, 296, (2), E256-E261.

49. Wu, G., Amino acids: metabolism, functions, and nutrition. *Amino acids* **2009**, 37, (1), 1-17.

50. Springer, J.; Ruth, P.; Beuerlein, K.; Palus, S.; Schipp, R.; Westermann, B., Distribution and function of biogenic amines in the heart of Nautilus pompilius L.(Cephalopoda, Tetrabranchiata). *Journal of molecular histology* **2005**, 36, (5), 345-353.

51. Kovacevic, Z.; McGivan, J. D., Mitochondrial metabolism of glutamine and glutamate and its physiological significance. *Physiological Reviews* **1983**, 63, (2), 547-605.

52. Tominaga, T.; Suzuki, H.; Mizuno, H.; Kouno, M.; Suzuki, M.; Kato, Y.; Sato, A.; Okabe, K.; Miyashita, M., Clinical significance of measuring plasma concentrations of glutamine and glutamate in alcoholic liver diseases. *Alcohol and Alcoholism* **1993**, 28, (Supplement 1A), 103-109.

53. Southam, A. D.; Easton, J. M.; Stentiford, G. D.; Ludwig, C.; Arvanitis, T. N.; Viant, M. R., Metabolic changes in flatfish hepatic tumours revealed by NMR-based metabolomics and metabolic correlation networks. *Journal of proteome research* **2008**, 7, (12), 5277-5285.

54. Kalhan, S. C.; Guo, L.; Edmison, J.; Dasarathy, S.; McCullough, A. J.; Hanson, R. W.; Milburn, M., Plasma metabolomic profile in nonalcoholic fatty liver disease. *Metabolism* 60, (3), 404-413.

55. Giordano, D.; Corrado, F.; Santamaria, A.; Quattrone, S.; Pintaudi, B.; Di Benedetto, A.; D'Anna, R., Effects of myo-inositol supplementation in postmenopausal women with metabolic syndrome: a perspective, randomized, placebo-controlled study. *Menopause* 18, (1), 102-104.

56. Wang, Y.; Holmes, E.; Tang, H.; Lindon, J. C.; Sprenger, N.; Turini, M. E.; Bergonzelli, G.; Fay, L. B.; Kochhar, S.; Nicholson, J. K., Experimental metabonomic model of dietary variation and stress interactions. *Journal of proteome research* **2006**, 5, (7), 1535-1542.

57. Afzali, B.; Bakri, R. S.; Bharma-Ariza, P.; Lumb, P. J.; Dalton, N.; Turner, N. C.; Wierzbicki, A. S.; Crook, M. A.; Goldsmith, D. J., Raised plasma total sialic acid levels are markers of cardiovascular disease in renal dialysis patients. *Journal of nephrology* **2003**, 16, (4), 540-545.

58. Baumann, H.; Jahreis, G. P.; Gaines, K. C., Synthesis and Regulation of Acute Phase Plasma in Primary Cultures of Mouse Hepatocytes. *The Journal of cell biology* **1983**, 97, 866-876.

59. Kaddurah-Daouk, R.; Kristal, B. S.; Weinshilboum, R. M., Metabolomics: a global biochemical approach to drug response and disease. *Annu. Rev. Pharmacol. Toxicol.* **2008**, 48, 653-683.

60. Xia, J.; Psychogios, N.; Young, N.; Wishart, D. S., MetaboAnalyst: a web server for metabolomic data analysis and interpretation. *Nucleic acids research* **2009**, 37, (suppl 2), W652-W660.

61. Weiler, S.; Merz, M.; Kullak-Ublick, G. A., Drug-induced liver injury: the dawn of biomarkers? *F1000prime reports 7*.

62. Aia, P.; Kal, M.; Lavu, E.; John, L. N.; Johnson, K.; Coulter, C.; Ershova, J.; Tosas, O.; Zignol, M.; Ahmadova, S., The burden of drug-resistant tuberculosis in Papua New Guinea: Results of a large population-based survey. *PloS one* 11, (3), e0149806.

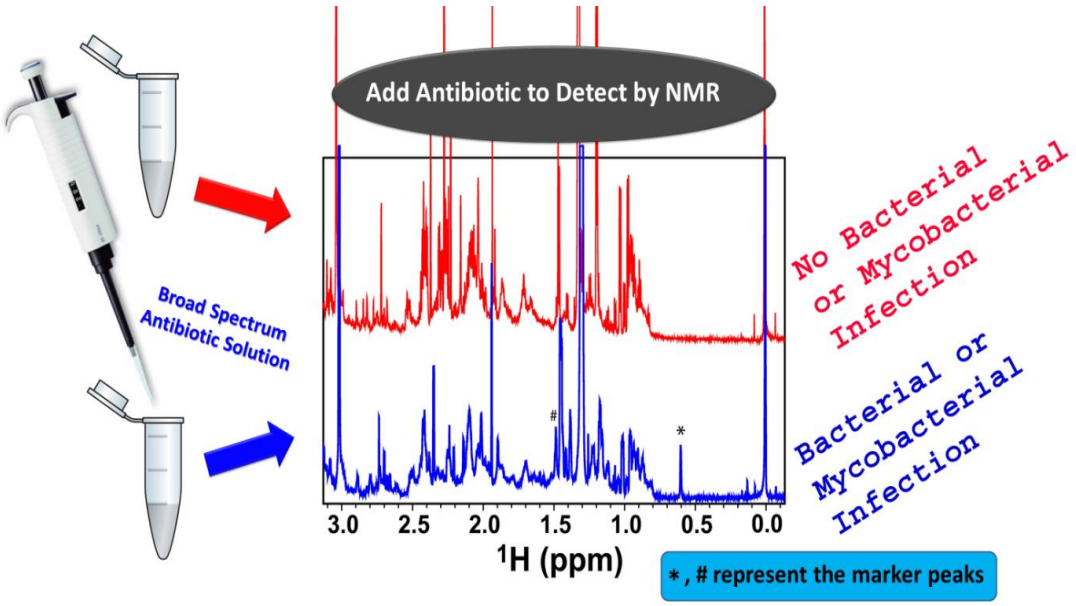
63. Dubey, A.; Rangarajan, A.; Pal, D.; Atreya, H. S., Chemical Shifts to Metabolic Pathways: Identifying Metabolic Pathways Directly from a Single 2D NMR Spectrum. *Analytical chemistry* 87, (24), 12197-12205.

# Chapter 6

## AADNMR: A Rapid Identification of Bacterial/Mycobacterial Infection for Diagnosis and Treatment Monitoring of Infectious Peritonitis


### Abstract:

An efficient method is reported 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 region 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 special microbial fatty acids when present inside the membrane are not easily detectable by NMR owing to their fast $T_2$ relaxation. Nonetheless, the problem can easily be circumvented if these fatty acids become suspended in solution. This has been achieved here by abolishing the membrane integrity using broad spectrum antibiotics. The suspended fatty acids are then detected by NMR to probe the infection. Therefore, the method has been given the 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 method is demonstrated for making distinction between bacterial and fungal infections during the diagnosis of infectious peritonitis –a life-threatening complication associated with prolonged peritoneal dialysis.

## Introduction:

Rapid detection of microbial pathogens in clinical samples is crucial for directing appropriate antimicrobial therapy and improving patient care and associated outcomes[1]. Every hour the appropriate treatment is delayed in patients in critical care has been shown to increase mortality by ~10-15%[2-6]; therefore, earlier detection of the pathogenic microorganism has the potential to greatly benefit patient care. Infectious peritonitis is one of the most serious complications associated with prolonged peritoneal dialysis (PD) therapy[3,7] –a technique used for treating the patients with end-stage renal failure (ESRF). Severe and prolonged infectious peritonitis is the major cause of mortality in PD patients[8] or permanent malfunctioning of peritoneal membrane and switching to haemodialysis. Therefore, considerable attention has been paid on prevention and treatment of PD-related infections[3,7]. According to treatment procedure, before the patient is discharged from the hospital, he is given an in hospital training (~10-14 days) on the continuous ambulatory peritoneal dialysis (CAPD) and is instructed to put intra-peritoneal antibiotics (generally a combination against gram positive and gram negative bacteria) as soon he notices cloudy effluent with any of these symptoms like abdominal pain, vomiting and/or fever. The condition may also be associated with low ultra-filtration. However in case if the cloudy effluent does not clear in three days (after instilling initial antibiotics) along with subsiding of symptoms, the patient is instructed to report to the treating physician for further investigations (i.e. total and differential leucocyte count with bacterial and fungal culture of effluent). The initial treatments of bacterial and fungal peritonitis differ -i.e. in bacterial peritonitis, intravenous antibiotics are added in addition to intra-peritoneal and wait for a response, whereas in fungal peritonitis initial treatment is to remove the CAPD catheter and patient is put on antifungals. The fungal peritonitis is more serious (than the bacterial peritonitis) and it leads to death of the patient in approximately 25% or more of the episodes[3]. Therefore, there is an urgent need for a rapid method to differentiate bacterial and fungal peritonitis and not to wait for 48-72 hours for culture reports to come as this may affect the prognosis of PD patient.

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. The method exploits the inherent difference present in the fatty acid composition of microbial and eukaryotic cell membranes and provides unambiguous information about the bacterial/mycobacterial infection very rapidly. The method has been tested here using both gram +ve and gram -ve bacterial strains, and finally, the utility of method has been demonstrated for the diagnosis of infectious peritonitis –a serious complication associated with prolonged peritoneal dialysis. The method helps the physician in two ways: (a) if it confirms the presence of bacterial/mycobacterial infection, the PD patient is continued on broad-spectrum antibacterials and (b) if it rules out the possibility of having

bacterial/mycobacterial infection –which then possibly indicates the presence of fungal infection- the patient is immediately referred for the antifungal treatment. Overall, the method has its great implication to start timely treatment of infected PD patients.

## Materials and Method:

### Chemicals:

Deuterium oxide ($D_2O$) and sodium salt of trimethylsilylpropionic acid-$d_4$ (TSP) used for NMR spectroscopy were purchased from Sigma-Aldrich (Rhode Island, USA). Ampicillin –a broad spectrum antibacterial effective against both Gram +ve and Gram –ve bacteria– used here for *in vitro* NMR studies was also purchased from Sigma-Aldrich (Rhode Island, USA).

### Bacterial Cell Culture:

The proposed **<AADNMR>** method was first tested on representative Gram-negative and Gram-positive bacterial strains named, respectively, *Escherichia coli* [ATCC 25922] and *Staphylococcus aureus* [ATCC 3160]. Both the strains were cultured in 25 mL of Luria-Bertani (LB) medium and were kept in a shaker incubator under identical conditions (37 °C, 200 rpm). Growth in these cultures was monitored by taking the absorbance (optical density, $OD_{600nm}$) on a Thermo Spectronic UV spectrophotometer at intervals of 1 h. In each case, two aliquots of 1.0 mL culture were collected in the exponential phase (OD600 = 0.6); one aliquot was used as a control (representing live cell suspension), and the other aliquot was treated with ampicillin (50 µl of 10 mg/ml stock solution). Both the culture aliquots were again kept in the shaker incubator for about 2 hours under identical conditions (37 °C, 300 rpm). Finally, each culture aliquot was centrifuged at 12,000 rpm for 5 minutes to remove all cell debris and other contaminants. The supernatant part was decanted and stored at -20 °C until the $^1H$ NMR experiments were performed.

### Clinical Samples:

***Infected Urine Samples:*** To ensure its general utility, the method has first been validated on a variety of infected urine samples obtained from UTI (Urinary Tract Infection) patients admitted in the Urology wards of Sanjay Gandhi Post Graduate Institute of Medical Sciences Lucknow. Inclusion criteria include urge to urinate frequently and need to urinate at night, painful burning sensation when urinating, discomfort or pressure in the lower abdomen, cloudy appearance and strong smell in urine, fever (typically lasting more than 2 days), impaired immune systems, or a history of relapsing or recurring UTIs, pain in the flank (pain that runs along the back at about waist level), vomiting and nausea. Each urine sample was divided into two parts of 1.0 mL each; one part was used as a control (representing live bacterial infection) and the other part was treated with ampicillin (50 µl of 10 mg/ml stock solution). Both the parts were incubated for

about 1 hour under identical conditions (37 °C, 300 rpm). Finally, each part was centrifuged at 12,000 rpm for 5 minutes to remove all cell debris and other contaminants. The supernatant part was decanted and stored at -20 °C until the $^{1}$H NMR experiments were performed.

***Infected PD Effluent Samples:*** For the clinical utility of the method, PD effluent samples (32 episodes) were obtained from 20 PD patients (n=20) admitted in the Nephrology wards of Sanjay Gandhi Post Graduate Institute of Medical Sciences, Lucknow. The study protocol was approved by the Hospital's Research Committee. All the selected PD patients were instilled with dialysate solution (Dianeal, 2.5 %) intraperitoneally containing dextrose. Among all the PD patients involved in this study (n=20), there was suspicion of having infectious peritonitis in 12 patients based on clinical symptoms and cloudy PD effluent (confirmed as per the guidelines of International Society of Peritoneal Dialysis)[3,7]. The PD patients with suspicion of having infectious peritonitis were given broad spectrum antibiotics instilled directly into the dialysate solution. Depending upon the criticality of infection, the antibiotics like Cefazolin, Tobramycin, Vancomycin, and ceftazidime were used for intraperitoneal instillation following standard treatment procedures. In each case, PD effluent sample was collected after a 4 hour dwell time, and was frozen and stored at a temperature of -20 °C, within 1-2 hours until the NMR measurements were performed.

**$^{1}$H NMR Spectroscopy:**

High Resolution NMR spectra were recorded at 298 K on a Bruker Avance III 800 MHz spectrometer (equipped with Cryoprobe). Standard relaxation edited 1D $^{1}$H NMR spectra were acquired using the Carr–Purcell–Meiboom–Gill (CPMG) pulse sequence [-recycle delay−π/2−(τ−π−τ)$_n$-acquisition][9], with simple pre-saturation of the water peak, a total spin–spin relaxation time of 160 ms (n=400 and 2τ=400 μs), and a recycle delay (RD) of 5 sec. Each spectrum consisted of the accumulation of 64 scans and lasted for approximately 8 minutes. All the spectra were processed in Topspin-2.1 (Bruker NMR data Processing Software). Prior to Fourier Transformation (FT), the 1D $^{1}$H NMR data were zero-filled to 4096 data points and a sine–bell apodization function was applied.

To confirm the assignment of marker peak as reported earlier[10], two-dimensional (2D) $^{1}$H-$^{1}$H total correlation spectroscopy (TOCSY) and $^{1}$H-$^{13}$C heteronuclear single quantum coherence (HSQC) spectra were acquired for some of the samples (including both bacterial cell cultures as well as PD effluents). Two-dimensional $^{1}$H–$^{1}$H TOCSY spectra were acquired in the phase sensitive mode using time proportional phase incrementation (TPPI), and the DIPSI2 pulse sequence for the spin lock[11]. 2048 data points with 16 transients per increment and 400 increments were acquired with a spectral width of 12 ppm in both dimensions. The RD between successive pulse cycles was 2

s and the mixing time of the DIPSI2 spin lock was 80 ms. The FIDs were weighted using a sine–bell-squared function in both dimensions and zero filled to 2048 and 1024 data points, respectively, in the $F_1$ and $F_2$ dimensions prior to FT. $^{1}H$–$^{13}C$ HSQC spectra (phase sensitive with echo–antiecho) were recorded with inverse detection and $^{13}C$ decoupling during acquisition[11]. A RD of 2.2 s was used between pulses, and a refocusing delay equal to $1/(4*{}^{1}J_{C-H} = 1.75$ ms) was employed. 2048 data points with 64 scans per increment and 256 increments were acquired with spectral widths of 12 and 165 ppm in the $^{1}H$ and $^{13}C$ dimensions, respectively. The FIDs were weighted using a sine–bell-squared function in both dimensions and zero filled to 2048 and 1024 data points, respectively, in the $F_1$ and $F_2$ dimensions prior to FT.

**<AADNMR> Protocol:**

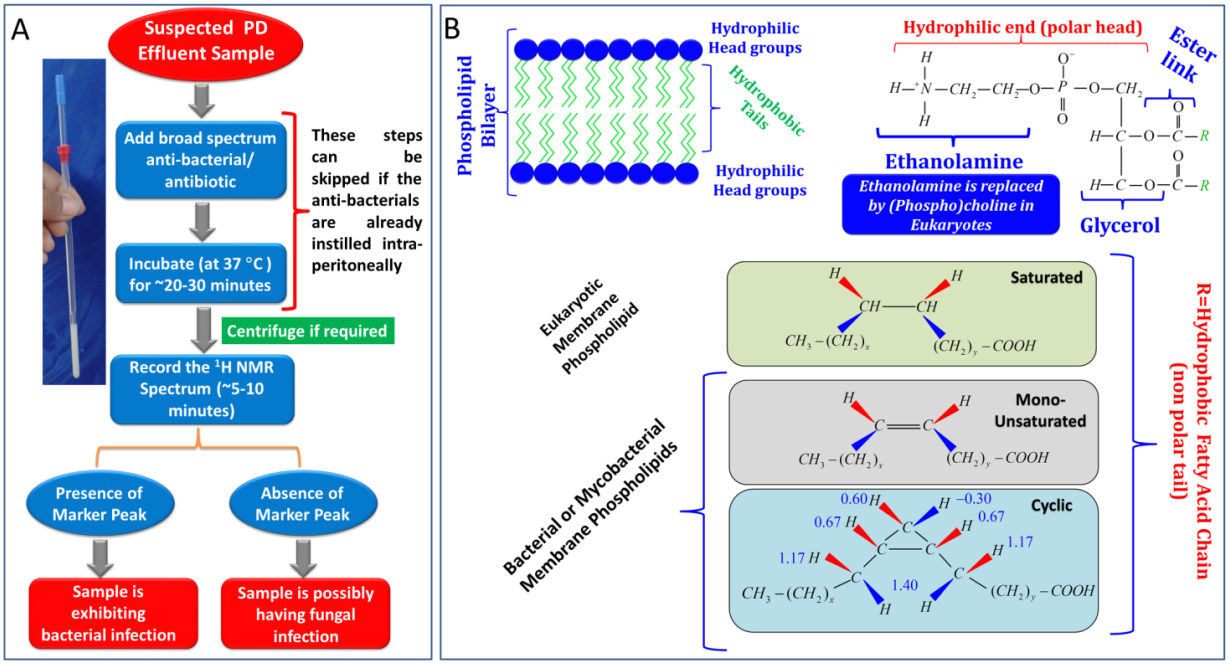



**Figure 1: (A)** Schematic showing the $^{1}H$ NMR based method –named here as ***"add antibiotic to detect by NMR"*** or "***AADNMR***"- for rapid identification of bacterial/mycobacterial infection in a biological/clinical sample. **(B)** Generalized structure of cell membrane with phospholipid bilayer (top right). The membranes of bacteria are structurally similar to the cell membranes of eukaryotes, except that bacterial/mycobacterial membranes consist of cyclic or monounsaturated fatty acids (rarely, polyunsaturated fatty acids) and do not normally contain sterols[12,13]. The numbers in the right bottom panel (representing the cyclopropyl fatty acid geometry) represent the $^{1}H$ NMR chemical shifts of ring methylene protons as reported earlier[14].

The **<AADNMR>** method for rapid identification of bacterial/mycobacterial infection has been illustrated in **Fig. 1A** and is based on the fact that several species of bacteria and mycobacteria contain cyclic fatty acids in their phospholipid bilayer membranes[15-20] which are rarely found in eukaryotic cell membranes except for some plant cells/organisms

(http://lipidlibrary.aocs.org/Lipids/fa_cycl/index.htm). **Fig. 1B** displays the generalized features of various such phospholipid bilayer membranes. Various NMR based studies on structural analysis of cyclic fatty acids of microbial origin (including mycolic acids frequently found in mycobacterial cell membranes and contain di-substituted cyclopropane rings[10]) have shown that the ring methylene protons of the cyclic fatty acids resonate between -0.40 and 0.68 ppm region of the $^1$H NMR spectrum[14,21,22]. After an extensive analysis of $^1$H NMR chemical shifts of various small metabolites (including those in human, bacteria, and fungi metabolomics NMR databases) we found that except for long-chain compounds containing cyclopropane/cyclopentane ring moiety, no small molecular weight compounds and metabolites resonate upfield to the 0.68 ppm region of the $^1$H NMR spectrum. Therefore the presence of a peak (peaks) between -0.40 and 0.68 ppm region of the $^1$H NMR spectrum may indicate identification of bacterial/mycobacterial infection. With this strategic aim, the present study was designed and signals from cyclic fatty acids were targeted. However when present within the membrane of a live bacterial/mycobacterial cell, these cyclic fatty acids are not easily detectable by NMR because of their very fast transverse relaxation (i.e. short $T_2$) owing to very long correlation time of the whole cell organism and therefore of associated membrane and its components. To produce signals from these fatty acids, the broad-spectrum antibiotics have been used, especially, the bactericidal antibiotics which disrupt the membrane (lipid bilayer) integrity of these bacterial cells (for details see these reference:[23,24]). The antibiotic induced bacterial cell membrane disruption brings these fatty acids into the sample solution for their facile detection by NMR. A point to be mentioned here that in the present study, we have found that the sample solution containing cyclic fatty acids of microbial origin (as produced here using antibiotic induced bacterial cell death) results into a strong $^1$H NMR signal resonating between 0.45 and 0.65 ppm. In the $^1$H-$^1$H TOCSY spectrum shown in **Figure 2**, this peak was found to be correlated to another peak centred about 1.5 ppm (through vicinal trans coupling). In accordance with previous literature[10,14,21], the signal between 0.45 and 0.65 ppm represents cumulative NMR signal from trans methylene protons of cyclopropane ring moiety (as per the assignment of cyclopropane ring reported earlier[14] and also depicted in **Figure 1**). This particular peak has been referred here as the marker peak and has been labeled by asterisk (*) in subsequent Figures. The signal centred about 1.5 ppm has been attributed to vicinal *trans* protons attached to the carbons adjacent to the ring and has been labeled by symbol hash (#) in some of the Figures.

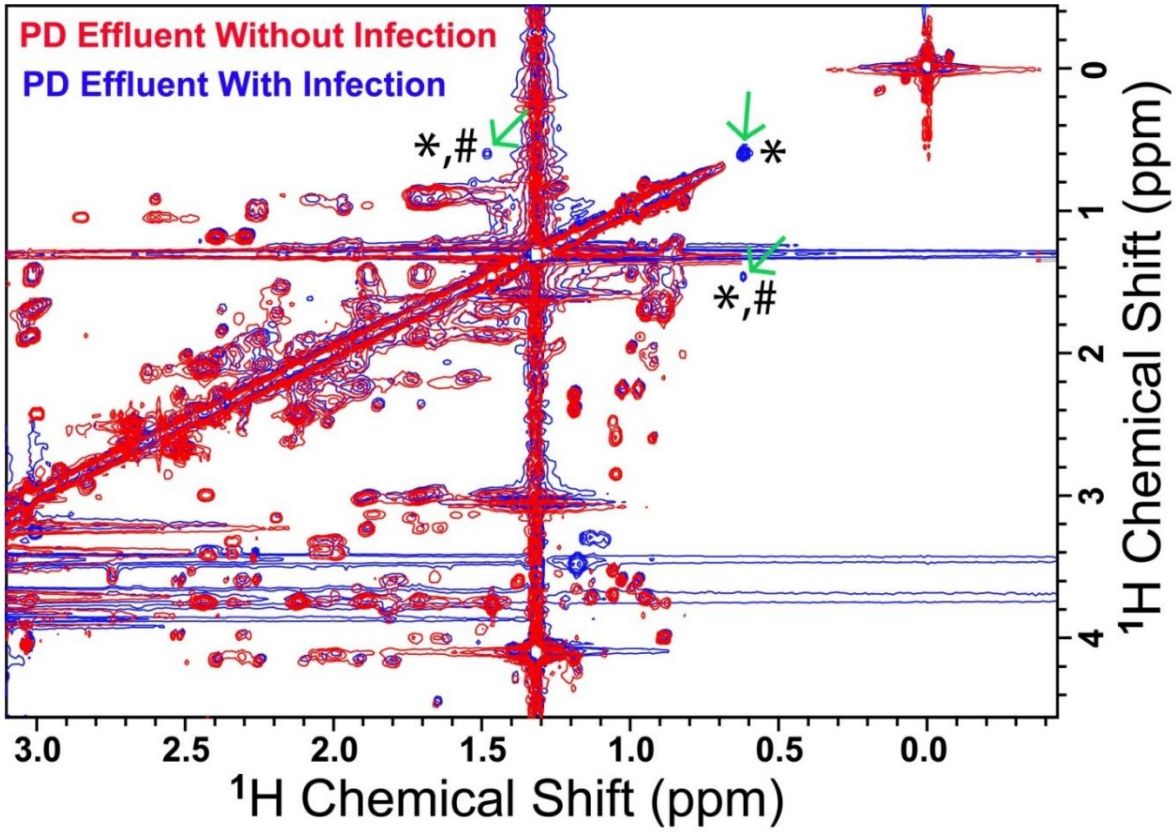



**Figure 2:** An overlay of the chemical shift matched regions of $^{1}$H-$^{1}$H TOCSY spectra of PD effluent samples obtained after 4 hours dwell time from suspected PD patients instilled with broad spectrum antibiotic intraperitoneally: one PD patient was having bacterial infection (blue) as evident from the presence of marker peak, whereas in other suspected case (red) the absence of marker peak simply rule out the possibility of any such infection. The correlation peak between marker peak (*) and peak at 1.5 ppm (#) confirms the presence of cyclic fatty acids of microbial origin in accordance with the previous reports (see the text).

Based on the facts and observations described above, the current **<AADNMR>** method has been designed. Briefly, the procedure includes adding a small amount of broad spectrum antibiotic solution (~1 mM final concentration) to the suspected clinical sample (if it is obtained without any antimicrobial treatment), incubating it for about 1 hour (at 37 °C and 200 rpm) and then recording a high resolution 1D $^{1}$H NMR spectrum. The spectrum is analyzed simply through visual inspection, and if the marker peak (as described above from 0.40 to 0.65 ppm region of the spectrum) is present, it simply confirms the presence of bacterial or mycobacterial infection. Typically a high resolution 1D $^{1}$H NMR spectrum can be acquired within a few minutes and since NMR requires minimal or no sample preparation or separation procedure, therefore for a clinical sample obtained after antimicrobial therapy has been started (like e.g. in the case of infectious peritonitis), the method can provide identification of microbial infection within about 20-30 minutes. The presence or absence of bacterial/mycobacterial infection further helps in differential diagnosis of bacterial and fungal infections and may aid physician treatment decisions. However, the conventional culture-based methods may produce false negative results in such situations.

## Results and Discussion:

### Validation of the Method through Experimental Testing:

The proposed **<AADNMR>** method was first tested on representative Gram-negative and Gram-positive bacterial strains named, respectively, *Escherichia coli* and *Staphylococcus aureus* (also an important pathogen in catheter exit-site infection in peritoneal dialysis[25]). **Figure 3A** and **3B** displays their respective 1D $^{1}$H NMR spectra before (blue) and after (red) treating with ampicillin. As evident from visual inspection, the marker peak (between 0.40 and 0.65 ppm region of the spectrum) is clearly visible in both the treated culture samples of Gram +ve and Gram –ve bacterial strains (prepared as described in the Materials and Method Section). To rule out the possibility that the observed marker peak is arising because of added antibiotic formulation, the 1D $^{1}$H NMR spectra of antibiotics frequently used in this study (like Ampicillin, Vancomycin, Azolin, etc.) were also recorded. The spectra are shown in the **Figure 4** and clearly, show that the marker peak is not at all present in any of the antibiotic formulation and is arising only because of the antibiotic induced bacterial cell lysis.

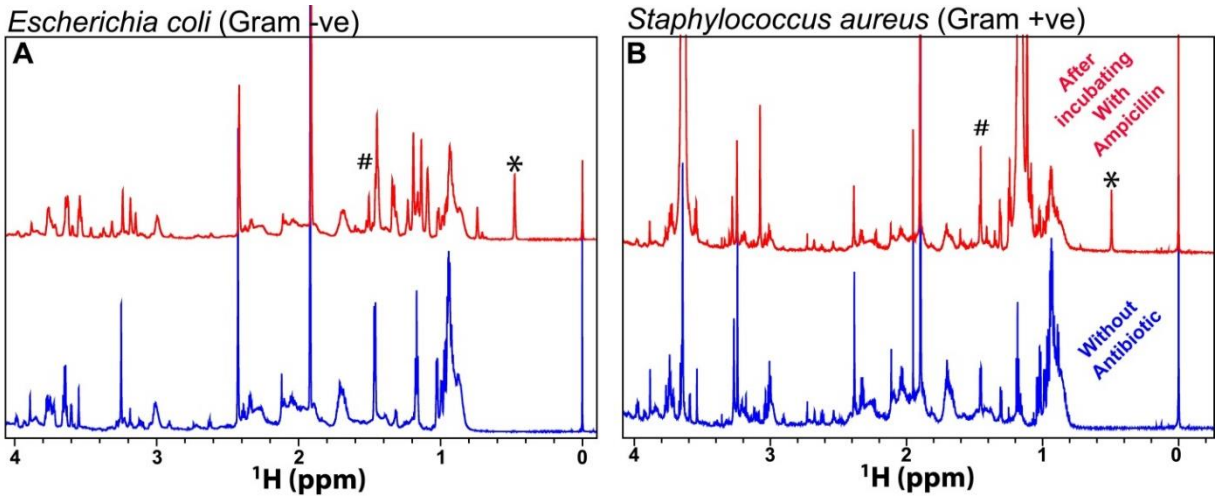



**Figure 3:** Representative one-dimensional (1D) $^{1}$H NMR spectra of extracellular metabolite solutions obtained, respectively, from ampicillin-treated (red) and untreated (blue) bacterial cell cultures (grown in the LB media at 37 °C till the optical density of culture at 600 nm reaches to ~0.6): **(A)** *Escherichia coli* [ATCC-25922] (as a representative of Gram –ve bacteria) and **(B)** *Staphylococcus aureus* [ATCC-3160] (as a representative of Gram +ve bacteria).

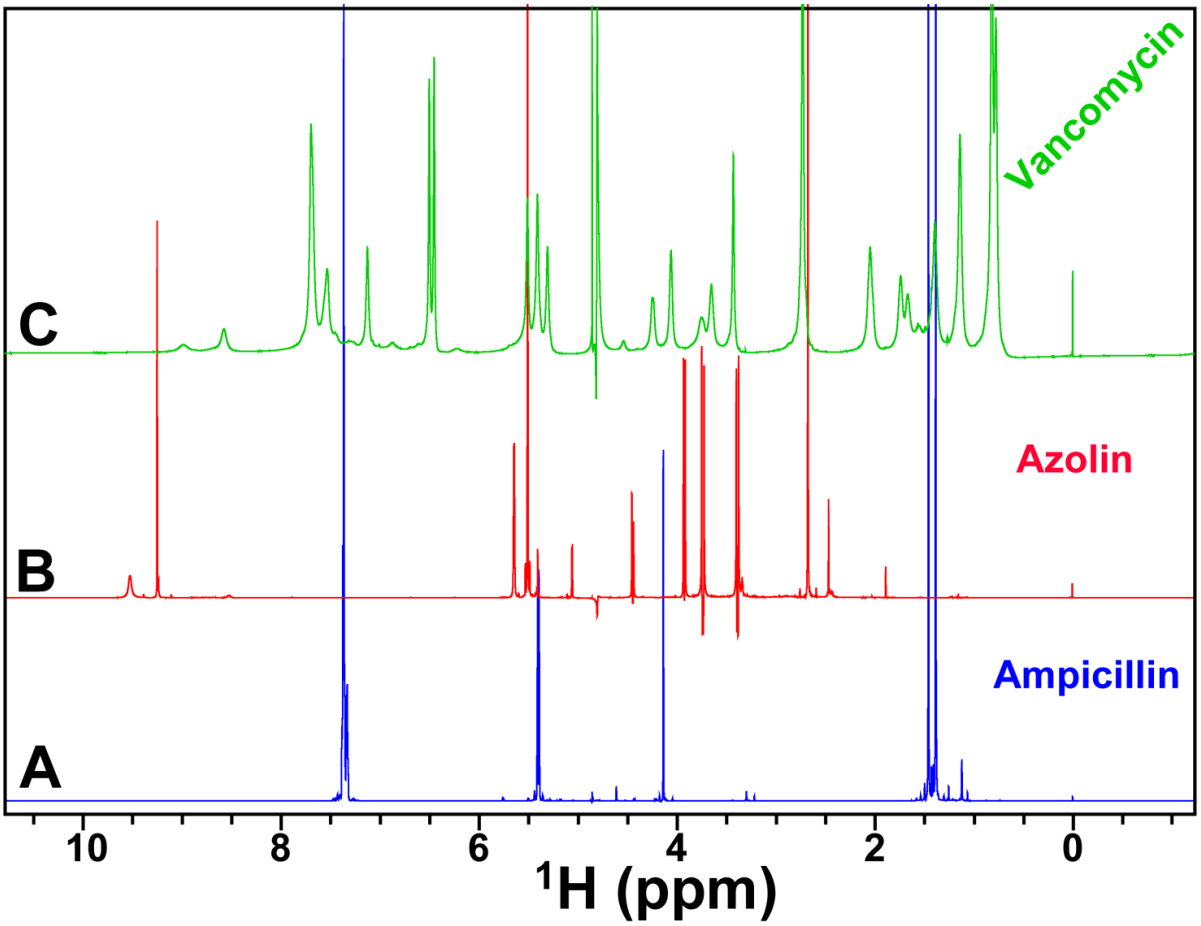



**Figure 4:** Stack plot of one-dimensional (1D) $^{1}$H NMR spectra of some of the antibiotics (antibacterials) most frequently used in the present study: **(A)** Ampicillin (effective against both Gram +ve and Gram –ve bacteria), **(B)** Azolin (effective against Gram +ve bacteria) and **(C)** Vancomycin (effective against Gram +ve bacteria). Spectra clearly reveal that there is no peak in the spectral region defined for the marker peak (i.e. between 0.40 and 0.68 ppm).

To further ensure that the marker peak is a reliable metabolic signature for identifying bacterial/mycobacterial infection, the method was further tested on urine samples obtained from Urinary tract infection (UTI) patients. **Figure 5** displays the 1D $^{1}$H NMR spectra of infected urine samples. The spectra in **Figure 5A-5E** represent the infected urine samples obtained without involving any antibiotic treatment, whereas the spectra in **Figure 5A'-5E'** represent the infected urine samples treated *ex vivo* with ampicillin (a broad spectrum antibiotic active against both Gram +ve and Gram –ve bacteria). As evident, the marker peak is present in all the ampicillin treated infected urine samples, whereas, it is totally absent in untreated infected urine samples.

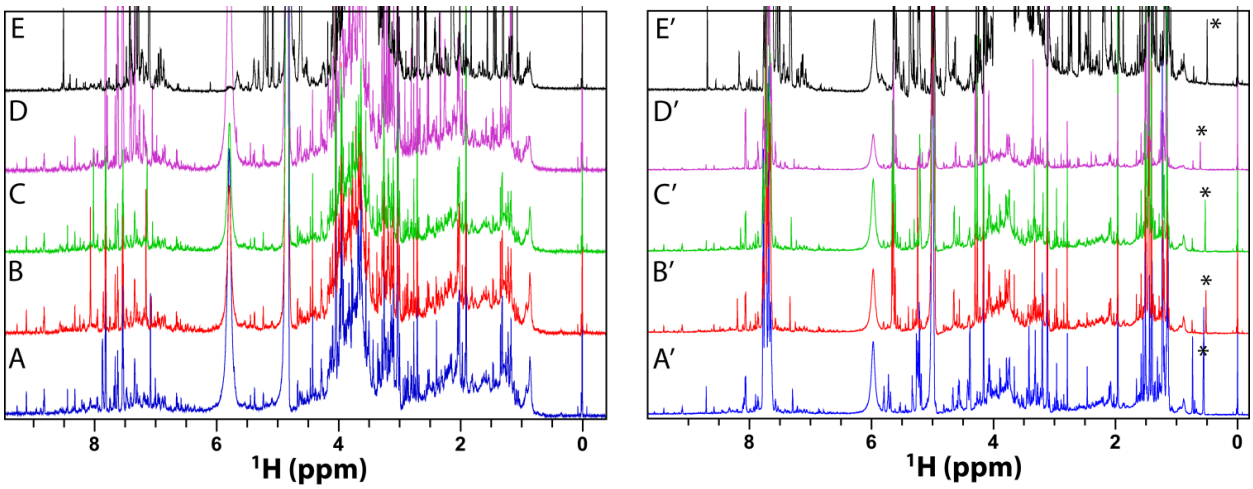



**Figure 5:** Stack plot of 1D $^{1}$H NMR CPMG spectra of urine samples of patients with gram (-ve) bacterial urinary tract infection (UTI) recorded at 800 MHz. **(A-E)** and **(A'-E')** represent, respectively, the control and ampicillin treated infected urine samples: **(A/A')** *Klebsiella pneumoniae*, **(B/B')** *Escherichia coli*, **(C/C')** *Pseudomonas aeruginosa*, **(D/D')** *Enterobacter aerogenes*, and **(E/E')** *Proteus mirabilis*.

To ensure the microbial origin of marker peak and its association with antibiotic treatment only, some additional experiments were also carried out. The various experimental results and observations are shown are summarized below in **(Figures 6-7)**:

**1.** To rule out the possibility that eukaryotic cells, when treated with antibiotics, may also produce the marker peak (because of some biochemical reaction), human embryonic kidney cells (HEK-293) were grown in Dulbecco's Modified Eagle's Medium (DMEM, Invitrogen) containing 10 % foetal bovine serum (FBS, Invitrogen) and 1% ampicillin and penicillin. 1 x $10^{6}$ cells were seeded in a T75 flask, when the cells reached 80% confluency, they were divided into two equal volumes: the first volume was directly used to record the 1D $^{1}$H CPMG NMR spectrum shown in **Figure 6A**, whereas the second volume was sonicated and centrifuged at 4 °C and 9000 rpm for 10 minutes, supernatant was collected and used to record the 1D $^{1}$H CPMG NMR spectrum shown in **Figure 6B**. As evident, the marker peak is absent in both the conditions suggesting that the marker peak is not resulting from a biochemical reaction between the antibiotic and the eukaryotic cell membrane.

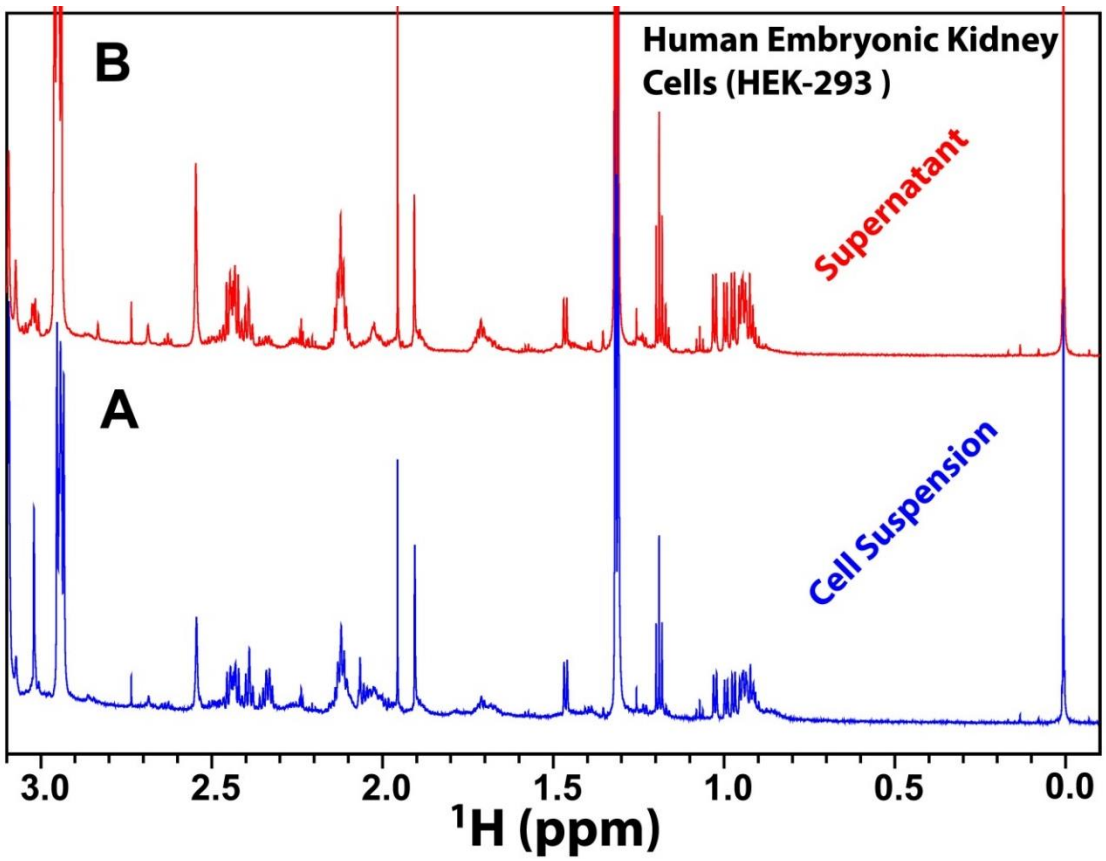



**Figure 6:** Stack plot of one-dimensional (1D) $^{1}$H NMR spectra of human embryonic kidney (HEK-293) cell line culture (grown in the presence of broad spectrum of anti-bacterials as described above): **(A)** cell suspension and **(B)** supernatant obtained after sonication (at 4 °C) and centrifugation at 6000 rpm for 10 minutes.

**2.** To see if the cyclic fatty acids can be made suspended in the solution using sonication and/or using lysozyme induced cell lysis, the bacterial (*E. coli*) cells in the exponential phase (OD@600nm=0.6) were transferred to four separate 2 ml centrifuge tubes (1 ml in each tube). One aliquot was directly used to record the 1D $^{1}$H CPMG NMR spectrum shown in **Fig 7A**, whereas the second aliquot was sonicated and centrifuged at 9000 rpm for 10 minutes, the supernatant was used to record the 1D $^{1}$H CPMG NMR spectrum shown in **Fig 7B**. Third aliquot was treated with lysozyme (15,000 units of longlife™ Lysozyme, G Biosciences), incubated at 37 °C for 1 hour and used to record the 1D $^{1}$H CPMG NMR spectrum shown in **Fig 7C**. Fourth aliquot was also treated with lysozyme (15,000 units of longlife™ Lysozyme, G Biosciences), incubated at 37 °C for 1 hour, sonicated and used to record the 1D $^{1}$H CPMG NMR spectrum shown in **Fig 7D**. As evident from the spectra **(Figure 7)**, the marker peak does not appear in any of these conditions suggesting that sonication, lysozyme treatment and even the combination of both these cell lysis treatments, all failed to suspend the membrane fatty acid components into the solution as required here to probe the infection.

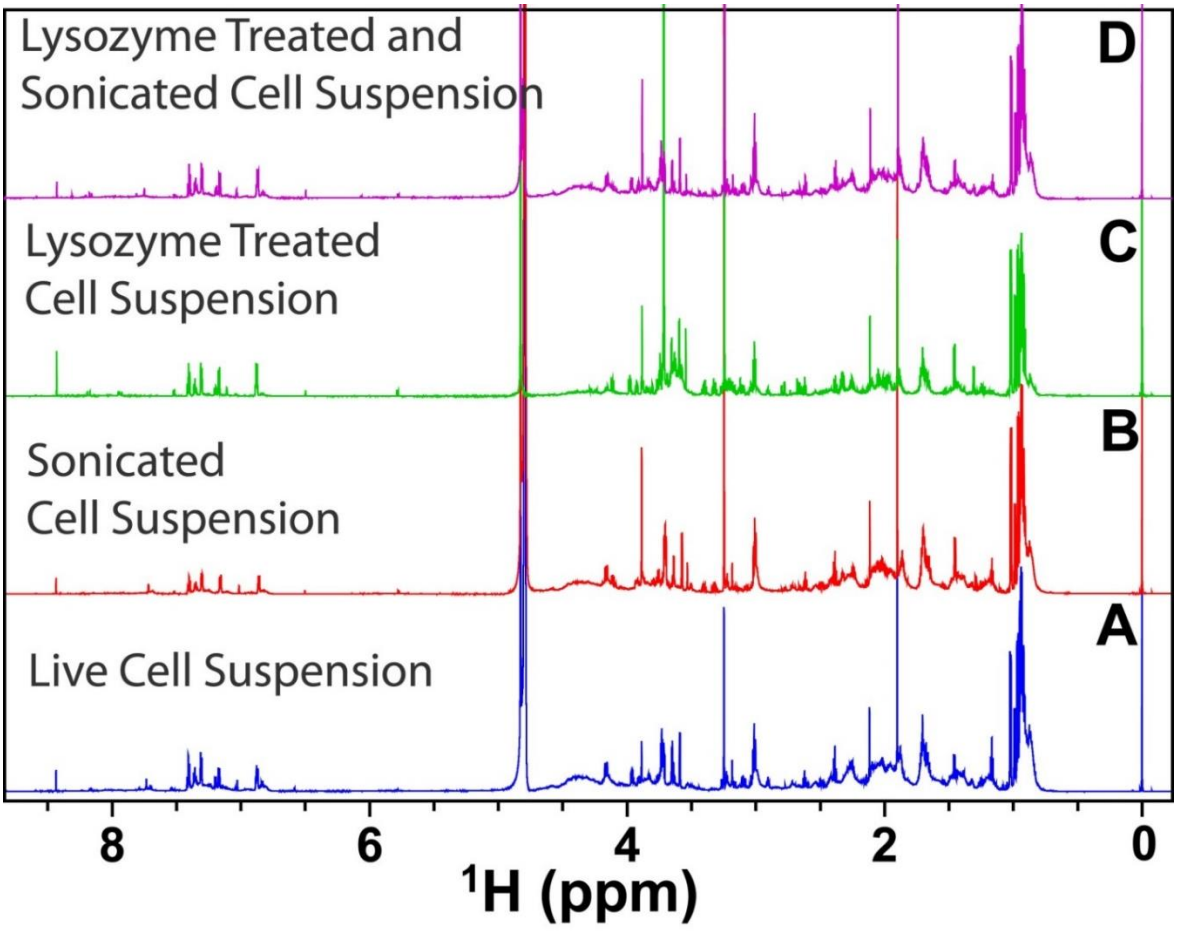



**Figure 7:** Stack plot of one-dimensional (1D) $^{1}$H CPMG NMR spectra of *Escherichia coli* bacterial culture (grown in standard LB media at 37 °C): **(A)** live cell suspension, **(B)** sonicated cell suspension, **(C)** cell suspension treated with lysozyme (at 37 °C for 1 hour) and **(D)** cell suspension treated with lysozyme (at 37 °C for 1 hour) followed by sonication. In each case, the cell suspension obtained in the exponential phase (when OD@600nm = 0.6) has been used.

**Clinical Utility in the Diagnosis of Infectious Peritonitis:**

Infectious peritonitis is well-known cause of mortality in PD patients[8]. Therefore, timely diagnosis of PD related infections is crucial to save the patient's life via early intervention. According to the treatment procedure described in the introduction part, (as per the guidelines of ISPD i.e., The International Society of Peritoneal Dialysis[3,7]), the PD effluent samples are mostly obtained after the antibiotic treatment has been started. Therefore the established methods [26] used for identification of infection may provide false negative results as they are all based on recovering and culturing of microorganisms from patient's body fluid. This prompted us to explore the $^{1}$H NMR spectroscopy in the diagnosis of PD effluent. A simple $^{1}$H NMR based method -as described in **Fig. 1-** has been proposed which enables the identification of bacterial/mycobacterial infection in a clinical or biological sample obtained after the antibiotic treatment therapy has been started as is the case with infectious peritonitis treatment procedure. The method –named here as **<AADNMR>**– can be used to rule in or rule out the presence of

bacterial/mycobacterial infection in a PD effluent sample, therefore, would greatly aids the physician's treatment decision. The method was applied to analyse 32 PD effluent samples (corresponding to 15 episodes of 8 normal PD patients and 17 episodes of 12 suspected PD patients having infectious peritonitis). The various clinical details are tabulated below:

**Table 1.** Clinical details of PD effluent

| | **PD effluent** | | | |
|---|---|---|---|---|
| | **No infection** | **Bacterial infection** | | **Fungal infection** |
| | | **Without Antibiotic** | **With Antibiotic** | |
| Number of Episodes (No. of Patients) | 15 (n=8) | 3 (n=3) | 12 (n=7) | 2 (n=2) |
| Sample turbidity | No | Yes | Yes | Yes |
| Bacterial Culture Test | -ve | Gram -ve (Bacillus) | -ve | -ve |
| Fungal Culture Test | -ve | -ve | -ve | +ve |

For all the 15 episodes of normal PD patients and 3 episodes of infectious peritonitis (not given the antibiotic treatment) the marker peak was not observed suggesting that there is either no infection (as in the case with former 15 episodes) or PD patients with infection are not given the intra-peritoneal antibiotic treatment (as in the case with latter 3 episodes). For remaining 14 episodes of 9 PD patients with suspicion of having infectious peritonitis and are given intraperitoneal antibiotic treatment, the ***marker peak*** was observed in 12 episodes indicating the presence of intraperitoneal infection (bacterial/mycobacterial) in those corresponding PD patients. However, for remaining two episodes with fair suspicion of infectious peritonitis, the negative <**AADNMR**> tests simply ruled out the possibility of bacterial/mycobacterial infection and hinted towards the possibility of having fungal infection. Later on, clinical microbiology laboratory tests confirmed that these fairly suspected PD patients were having fungal infection. The presence of ***marker peak*** also correlated well with the onset and course of infection in a patient with 2 episodes of bacterial peritonitis and with response to therapy. A typical demonstration of various experimental results is displayed in **Figure 8**.

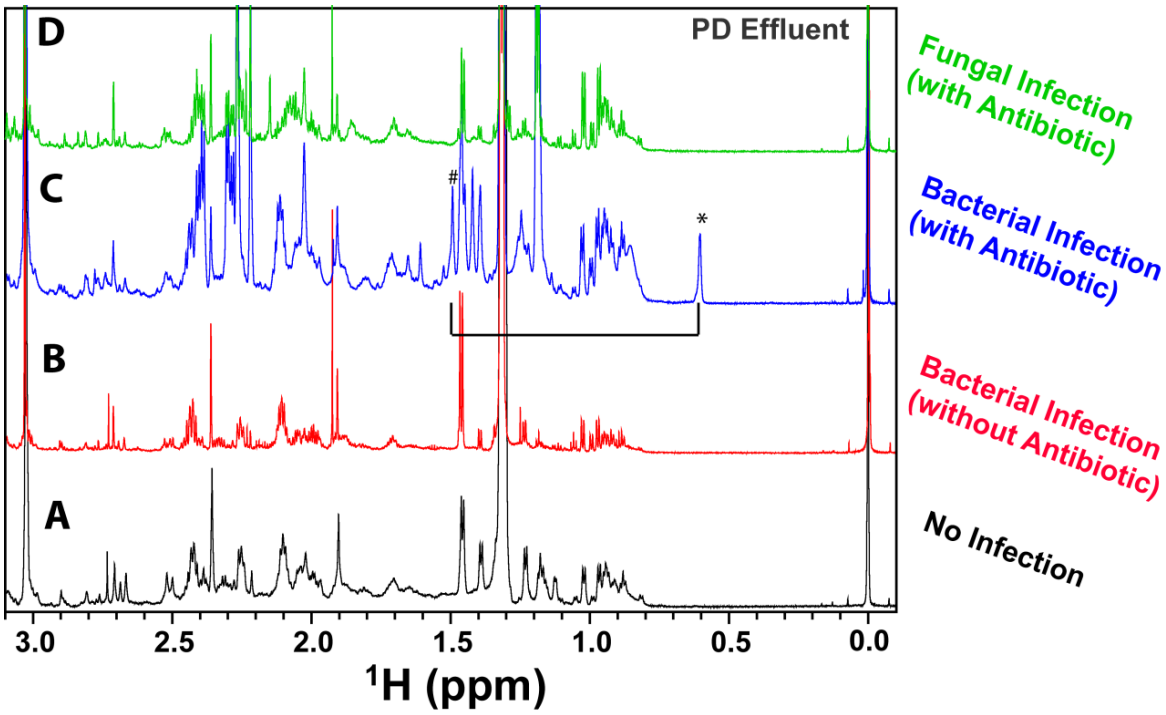



**Figure 8:** Stack plot of representative one-dimensional (1D) $^{1}$H NMR spectra of PD effluent (2.5 %, after 4 hour dwell time) from PD patients. **(A)** Represents the control PD effluent sample of a PD patient with no infection. **(B** and **C)** represent PD effluent samples from a PD patient having bacterial peritonitis and are obtained, respectively, before **(B)** and after **(C)** the intra-peritoneal antibiotic treatment has been started. In **(C)**, the presence of marker peak at ~0.61 ppm confirms the bacterial peritonitis which is otherwise difficult to predict through the routine cell culture based method if the patient has already been given the intraperitoneal antibiotic treatment. **(D)** Represents PD effluent sample from a PD patient having fungal peritonitis and given intraperitoneal antibiotic treatment. The absence of marker peak indirectly hints towards the fungal infection.

Compared to ampicillin treated bacterial cell cultures, where the marker peak was mostly centred about 0.5 ppm, in PD effluent samples the marker peak was slightly downfield shifted mostly centred about 0.63 ppm. These slight differences in the chemical shifts may be attributed to varying fatty acid composition of bacteria (particularly the position of the cyclopropane ring) which is markedly affected both quantitatively and qualitatively by the nature of the medium and by the conditions under which the culture is grown (more details on this phenomenon can be seen in these references [17,18,20,27]). To check this possibility, the bacterial (*Staphylococcus aureus*) cells were cultured directly into the PD dialysate solution (2.5 %) and treated with ampicillin. The experimental procedure and results are as described as follows: The PD effluent samples obtained from a PD patient having infectious peritonitis and given the intraperitoneal antibiotic treatment, the chemical shift of marker peak was centred about ~0.6 ppm compared to that in

the antibiotic treated bacterial cell culture where it is mostly centred about ~0.5 ppm. To check if the intraperitoneal conditions are involved in this chemical shift difference or Dianeal (Baxter U S, containing 2.5 % glucose) PD solution -instilled intraperitoneally in all the PD patients- is causing this chemical shift perturbation, bacterial (*Staphylococcus aureus*) cells were cultured directly into the Dianeal PD solution. Culture in the exponential phase (when OD @ 600 nm = 0.6) was transferred into two 2 ml centrifuge tubes (1.0 ml culture in each tube). One aliquot was used as a control (representing live cell suspension), and the other aliquot was treated with ampicillin (50 µl of 10 mg/ml stock solution). Both the culture aliquots were again kept in the shaker incubator for about 1 hour under identical conditions (37 °C, 200 rpm). Finally, each culture aliquot was centrifuged for 5 min at 4 °C and 12,000 rpm to remove all cell debris and other contaminants. The supernatant part was decanted and stored at -20 °C until the $^1$H NMR experiments were performed. As evident, in **Figure 9**. the marker peak -detected in the ampicillin treated culture sample- is slightly upfield shifted towards 0.5 ppm **(Fig. 9C)** compared to that detected in infected PD effluent sample obtained after intraperitoneal antibiotic treatment **(Fig. 8C)** suggesting that the varying environmental/culture conditions are modulating the synthetic chemistry of microbial cyclic fatty acids.

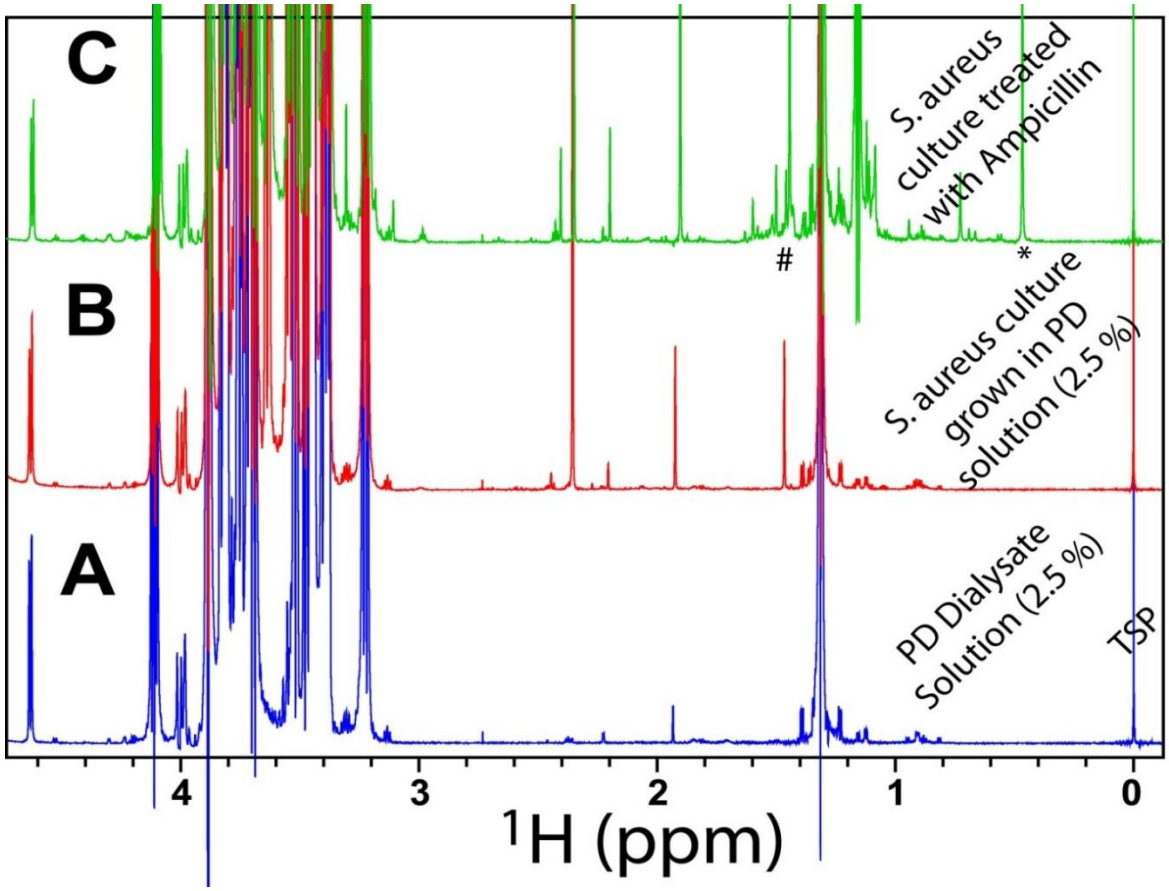



**Figure 9:** Stack plot of one-dimensional (1D) $^1$H CPMG NMR spectra of: **(A)** commercial 2.5 % PD dialysate solution and **(B, C)** bacterial cell culture (*Staphylococcus aureus*, grown in the PD solution): **(B)** without antibiotic treatment and **(C)** after incubating the bacterial cell culture with ampicillin at 37 °C for 1 hour. The marker peak in ampicillin treated sample was found to be

upfield shifted (close to 0.5 ppm) suggesting that the intraperitoneal conditions may be involved in modulating the microbial cyclic fatty acids qualitatively.

**Future Directions:**

The importance of NMR based metabolomics in diagnostics -e.g. in identifying biomarkers or defining pathological status- has been growing exponentially[9,28-32]. As reported and reviewed previously, the prolonged peritoneal dialysis (PD) therapy is often associated with several complications[3,7,25,33-37]. These complications –which could be either immediate or delayed- have broadly been classified into two categories based on their aetiologies: (a) ***infectious*** including bacterial peritonitis, tuberculous peritonitis, fungal peritonitis, and infections of the catheter exit site and tunnel and (b) ***non-infectious*** including catheter failure, peritoneum damage (caused by PD fluids rendering it unsuitable for adequate dialysis), hernias and encapsulating peritoneal sclerosis (EPS, a rare but serious complication)[33-35]. However so far, no NMR based metabolomics studies on body fluids and PD effluent has been carried out in the context of PD. The potential advantages of NMR and the persistent problems regarding the timely diagnosis of various complications associated with prolonged peritoneal dialysis, therefore, prompted us to explore $^{1}$H NMR spectroscopy as a diagnostic utility based on the analysis of PD effluent. The present study was mainly focused towards rapid identification of the onset of bacterial/mycobacterial infection, especially, in the context of infectious peritonitis. However, the study of changes in the metabolome of PD effluent -which defines the metabolites consumed from and secreted into the effluent from the tissue or cell in close proximity [36,38]- would be of great implication in predicting the complications associated with prolonged peritoneal dialysis and therefore to give timely therapeutic interventions. With this view, NMR based metabolomics study of PD effluent has been undertaken to ensure that the complications could be detected earlier and managed appropriately. The study is in progress, and its outcomes will be presented separately in future.

## Conclusion:

In conclusion, an efficient and simple $^{1}$H NMR based method has been reported here for rapid identification of bacterial/mycobacterial infection in a suspected clinical/biological sample. The utility of the method has been demonstrated to analyse the PD effluent samples obtained after the antibiotic treatment has been started and the whole exercise –including sample preparation, data collection, and data analysis/interpretation (simply through visual inspection)- can be completed within about 20-30 minutes. The method aid the physician's treatment decision in two ways: (a) if it confirms the presence of bacterial/mycobacterial infection, the PD patient is continued on broad-spectrum antibacterials and (b) if it rules out the possibility of having

bacterial/mycobacterial infection –which then possibly indicates the presence of fungal infection- the patient is immediately referred for the antifungal treatment. Further, the method can also be used for treatment monitoring and disease recovery in PD patients suffering from bacterial peritonitis. Overall, the method has its great implication to start timely treatment of infected PD patients to improve their surveillance.

The limitation of the method is that it is non-specific and does not provide any distinction between the infections caused by different bacterial strains. The other limitation of the method is that it may lead to false negative results when there is resistance in the bacterial strain against the antibiotic used for the cell lysis. However, the problem can be circumvented using a cocktail of broad spectrum antibiotics. Nevertheless, we foresee that the method because of its speed and simplicity will open up new avenues for a wide variety of clinical and biomedical applications as summarized below and depicted in **Figure 10**.

(a) The method can be used for treatment monitoring and recovery of infected patients in ICUs (like those with Urinary tract infection, respiratory tract infection, lung infection, etc.) where clinical samples (like urine, serum, sputum, nasal or pharyngeal swabs, etc.) are obtained after the antibiotic treatment has been started.

(b) Because of its ability to rule in or rule out the presence of bacterial/mycobacterial infection in a clinical sample, the method can be used for differential diagnosis of bacterial meningitis and ventriculitis[39,40].

(c) The method can also be employed for rapid diagnosis of bloodstream infections by bacteria (known as bacteraemia) which is a serious medical problem associated with significant morbidity and mortality[41-45].

(d) The method can be used to test the antibacterial/bactericidal activity of a newly developed drug or to check if its mode of action is similar to the common antibiotics. Likewise, the method can also be used to guide the treatment of patients in ICUs suffering from a bacterial/mycobacterial infection exhibiting multidrug-resistant (MDR). In such cases, the infected body fluid (blood, serum, urine, CSF, etc.) will be subjected for *ex vivo* treatment with empirical broad-spectrum antibiotics and the presence of marker peak will help to confirm if the antibiotic combination is effective or not.

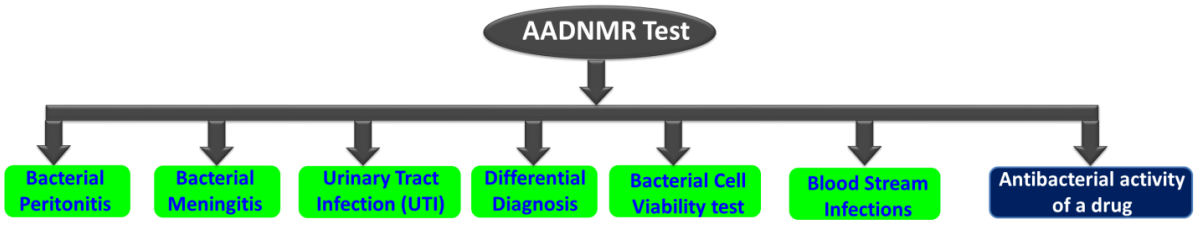



**Figure 10:** Summary of some of the potential medical and biomedical applications of the proposed **<AADNMR>** method.

**Reference's**


1. Afshari, A.; Schrenzel, J.; Ieven, M.; Harbarth, S. Bench-to-bedside review: Rapid molecular diagnostics for bloodstream infection - a new frontier? *Crit. Care* **2012**, *16* (3), 222.

2. Kumar, A.; Roberts, D.; Wood, K. E.; Light, B.; Parrillo, J. E.; Sharma, S.; Suppes, R.; Feinstein, D.; Zanotti, S.; Taiberg, L.; Gurka, D.; Kumar, A.; Cheang, M. Duration of hypotension before initiation of effective antimicrobial therapy is the critical determinant of survival in human septic shock. *Crit. Care Med.* **2006**, *34* (6), 1589-1596.

3. Li P T; Szeto C C; Piraino B; Bernardini J; Figueiredo A E; Gupta A; Johnson D W; Kuijper Ed J; Lye W C; Salzer W; Schaefer F; Struijk D G . ISPD Guidelines/Recommendations: Peritoneal Dialysis-Related Infections Recommendations (2010 Update). Peritoneal Dialysis International 30, 393-423. 2010. Ref Type: Journal (Full)

4. De Luca, G.; Suryapranata, H.; Ottervanger, J. P.; Antman, E. M. Time Delay to Treatment and Mortality in Primary Angioplasty for Acute Myocardial Infarction: Every Minute of Delay Counts. *Circulation* **2004**, *109* (10), 1223-1225.

5. Kang, C. I.; Kim, S. H.; Kim, H. B.; Park, S. W.; Choe, Y. J.; Oh, M. d.; Kim, E. C.; Choe, K. W. Pseudomonas aeruginosa Bacteremia: Risk Factors for Mortality and Influence of Delayed Receipt of Effective Antimicrobial Therapy on Clinical Outcome. *Clinical Infectious Diseases* **2003**, *37* (6), 745-751.

6. Morrell, M.; Fraser, V. J.; Kollef, M. H. Delaying the empiric treatment of candida bloodstream infection until positive blood culture results are obtained: a potential risk factor for hospital mortality. *Antimicrob. Agents. Chemother.* **2005**, *49* (9), 3640-3645.

7. Keane W F; Bailie G R; Boeschoten E; Gokal R; Golper T A; Holmes C J; Kawaguchi Y; Piraino B; Riella M; Vas S . ISPD Guidelines/Recommendations: Adult Peritoneal Dialysis-Related Peritonitis Treatment Recommendations (2000 Update). Peritoneal Dialysis International 20, 396-411. 2000. Ref Type: Journal (Full)

8. Perez, F. M.; Rodriguez-Carmona, A.; Garcia-Naveiro, R.; Rosales, M.; Villaverde, P.; Valdes, F. Peritonitis-related mortality in patients undergoing chronic peritoneal dialysis. *Perit. Dial. Int.* **2005**, *25* (3), 274-284.

9. Wishart, D. S. Quantitative metabolomics using NMR. Trends Anal.Chem. 27[3], 228-237. 2008. Ref Type: Journal (Full)

10. Mederos, L.; Valdivia, J. A.; Valero-Guillen, P. L. Analysis of the structure of mycolic acids of Mycobacterium simiae reveals a particular composition of alpha-mycolates in strain 'habana' TMC 5135, considered as immunogenic in tuberculosis and leprosy. *Microbiology.* **2007**, *153* (Pt 12), 4159-4165.

11. Cavanagh J; Fairbrother W J; Palmer A G; Skelton N J *Protein NMR Spectroscopy: Principles and Practice;* Elsevier Academic Press: San Diego, CL, 2006.

12. Kaneda T . Iso- and anteiso-fatty acids in bacteria: biosynthesis, function, and taxonomic significance. Microbiol.Rev. 55[2], 288-302. 1991. Ref Type: Journal (Full)

13. Suutari M; Laakso S . Microbial fatty acids and thermal adaptation. Crit.Rev.Microbiol. 20[4], 285-328. 1994. Ref Type: Journal (Full)

14. Knothe, G. NMR characterization of dihydrosterculic acid and its methyl ester. *Lipids* **2006**, *41* (4), 393-396.

15. Sebedio, J. L.; Grandgirard, A. Cyclic fatty acids: natural sources, formation during heat treatment, synthesis and biological properties. *Prog. Lipid. Res* **1989**, *28* (4), 303-336.

16. Grogan D W; Cronan J E . Cyclopropane ring formation in membrane lipids of bacteria. Microbiol.Mol.Biol.Rev. 61, 429-441. 1997. Ref Type: Journal (Full)

17. varez-Ordonez, A.; Fernandez, A.; Lopez, M.; Arenas, R.; Bernardo, A. Modifications in membrane fatty acid composition of Salmonella typhimurium in response to growth conditions and their effect on heat resistance. *Int. J Food. Microbiol* **2008**, *123* (3), 212-219.

18. varez-Ordonez, A.; Fernandez, A.; Lopez, M.; Bernardo, A. Relationship between membrane fatty acid composition and heat resistance of acid and cold stressed Salmonella senftenberg CECT 4384. *Food. Microbiol* **2009**, *26* (3), 347-353.

19. Satyanarayana T; Johri B N *Microbial Diversity: Current Perspectives and Potential Applications;* I.K. International Publishing House Pvt. Ltd., New Delhi: 2005.

20. Annous, B. A.; Kozempel, M. F.; Kurantz, M. J. Changes in membrane fatty acid composition of Pediococcus sp. strain NRRL B-2354 in response to growth conditions and its effect on thermal resistance. *Appl. Environ. Microbiol* **1999**, *65* (7), 2857-2862.

21. Marseglia, A.; Caligiani, A.; Comino, L.; Righi, F.; Quarantelli, A.; Palla, G. Cyclopropyl and omega-cyclohexyl fatty acids as quality markers of cow milk and cheese. *Food. Chem.* **2013**, *140* (4), 711-716.

22. Gunstone F D Cyclic Acids in The Lipid Handbook, 2nd edn. Gunstone F D, Harwood J L, Padley F B, Eds.; **Chapman and Hall, London.**: 1994; pp 13-14.

23. Kohanski, M. A.; Dwyer, D. J.; Collins, J. J. How antibiotics kill bacteria: from targets to networks. *Nat. Rev Microbiol* **2010**, *8* (6), 423-435.

24. Green, D. W. The bacterial cell wall as a source of antibacterial targets. *Expert. Opin. Ther. Targets.* **2002**, *6* (1), 1-19.

25. Kreft, B.; Ilic, S.; Ziebuhr, W.; Kahl, A.; Frei, U.; Sack, K.; Trautmann, M. Adherence of Staphylococcus aureus isolated in peritoneal dialysis-related exit-site infections to HEp-2 cells and silicone peritoneal catheter materials. *Nephrol. Dial. Transplant.* **1998**, *13* (12), 3160-3164.

26. Banoo S; Bell D; Bossuyt P; Herring A; Mabey D; Poole F; Smith P G; Sriram N; Wongsrichanalai C; Linke R; O'Brien R; Perkins M; Cunningham J; Matsoso P; Nathanson C M; Olliaro P; Peeling R W; Ramsay A . Evaluation of diagnostic tests for infectious diseases: general principles. Nature Reviews Microbiology , S16-S28. 2014. Ref Type: Journal (Full)

27. William M O'Leary . The Fatty Acids of Bacteria. Bacteriol Rev 26[4], 421-447. 1962. Ref Type: Journal (Full)

28. Beckonert, O.; Keun, H. C.; Ebbels, T. M.; Bundy, J.; Holmes, E.; Lindon, J. C.; Nicholson, J. K. Metabolic profiling, metabolomic and metabonomic procedures for NMR spectroscopy of urine, plasma, serum and tissue extracts. Nat.Protoc. 2[11], 2692-2703. 2007. Ref Type: Journal (Full)

29. Lindon, J. C.; Nicholson, J. K.; Holmes, E.; Everett, J. R. Metabonomics: Metabolic processes studied by NMR spectroscopy of biofluids. *Concepts Magn. Reson.* **2000**, *12* (5), 289-320.

30. Coen, M.; Holmes, E.; Lindon, J. C.; Nicholson, J. K. NMR-based metabolic profiling and metabonomic approaches to problems in molecular toxicology. *Chem Res. Toxicol.* **2008**, *21* (1), 9-27.

31. Nicholson, J. K.; Lindon, J. C.; Holmes, E. 'Metabonomics': understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data. *Xenobiotica* **1999**, *29* (11), 1181-1189.

32. Gowda, G. A.; Zhang, S.; Gu, H.; Asiago, V.; Shanaiah, N.; Raftery, D. Metabolomics-based methods for early disease diagnostics. *Expert. Rev Mol Diagn.* **2008**, *8* (5), 617-633.

33. Stuart, S.; Booth, T. C.; Cash, C. J.; Hameeduddin, A.; Goode, J. A.; Harvey, C.; Malhotra, A. Complications of continuous ambulatory peritoneal dialysis. *Radiographics.* **2009**, *29* (2), 441-460.

34. Cochran, S. T.; Do, H. M.; Ronaghi, A.; Nissenson, A. R.; Kadell, B. M. Complications of peritoneal dialysis: evaluation with CT peritoneography. *Radiographics.* **1997**, *17* (4), 869-878.

35. Goldstein, M.; Carrillo, M.; Ghai, S. Continuous ambulatory peritoneal dialysis-a guide to imaging appearances and complications. *Insights Imaging* **2013**, *4* (1), 85-92.

36. Dunn W B; Summers A; Brown M; Goodacre R; Lambie M; Johnson T; Wilkie M; Davies S; Topley N; Brenchley P . Proof-of-principle study to detect metabolic changes in peritoneal dialysis effluent in patients who develop encapsulating peritoneal sclerosis. Nephrol.Dial.Transplant. 27, 2502-2510. 2012. Ref Type: Journal (Full)

37. Sampimon, D. E.; Coester, A. M.; Struijk, D. G.; Krediet, R. T. The time course of peritoneal transport parameters in peritoneal dialysis patients who develop encapsulating peritoneal sclerosis. *Nephrol. Dial. Transplant.* **2011**, *26* (1), 291-298.

38. Kell, D. B.; Brown, M.; Davey, H. M.; Dunn, W. B.; Spasic, I.; Oliver, S. G. Metabolic footprinting and systems biology: the medium is the message. *Nat. Rev Microbiol* **2005**, *3* (7), *557-565.*

39. Subramanian, A.; Gupta, A.; Saxena, S.; Gupta, A.; Kumar, R.; Nigam, A.; Kumar, R.; Mandal, S. K.; Roy, R. Proton MR CSF analysis and a new software as predictors for the differentiation of meningitis in children. *NMR. Biomed.* **2005**, *18* (4), 213-225.

40. Coen, M.; O'Sullivan, M.; Bubb, W. A.; Kuchel, P. W.; Sorrell, T. Proton nuclear magnetic resonance-based metabonomics for rapid diagnosis of meningitis and ventriculitis. *Clin. Infect. Dis.* **2005**, *41* (11), 1582-1590.

41. Riedel, S.; Carroll, K. C. Blood cultures: key elements for best practices and future directions. *J Infect. Chemother.* **2010**, *16* (5), 301-316.

42. Micek, S. T.; Lloyd, A. E.; Ritchie, D. J.; Reichley, R. M.; Fraser, V. J.; Kollef, M. H. Pseudomonas aeruginosa bloodstream infection: importance of appropriate initial antimicrobial treatment. *Antimicrob. Agents. Chemother.* **2005**, *49* (4), 1306-1311.

43. Martin, G. S.; Mannino, D. M.; Eaton, S.; Moss, M. The epidemiology of sepsis in the United States from 1979 through 2000. *N. Engl. J Med.* **2003**, *348* (16), 1546-1554.

44. Maki, D. G.; Kluger, D. M.; Crnich, C. J. The risk of bloodstream infection in adults with different intravascular devices: a systematic review of 200 published prospective studies. *Mayo. Clin. Proc.* **2006**, *81* (9), 1159-1171.

45. Angus, D. C.; Wax, R. S. Epidemiology of sepsis: an update. *Crit. Care Med.* **2001**, *29* (7 Suppl), S109-S116.

# Chapter 7

## Conclusion and Future Prospects

7.1 Conclusion

7.2 Future Prospects

7.3 Future Work Plans

**"By three methods we may learn wisdom: First, by reflection, which is noblest; second, by imitation, which is easiest; and third by experience, which is the bitterest." - Confucius**

## 7.1: Conclusion

NMR-based metabolomics is a promising analytical tool for the characterization and identification of metabolites in complex mixtures and provides key information in diverse applications. Metabolites are the end products of gene expression and are the result of enzymatic and protein activity. Thus, metabolites are closer to represent the phenotype or a disease than either genetic or proteomics information. NMR provides a high-throughput and reproducible method of analysis that requires minimal sample preparation to visualize the metabolic profile of the study groups providing an instant profile of the metabolome. For as it produces -long-lasting database - greatly reproducible data and -accurate quantitation. The thesis chiefly focused on the application of the discriminatory power of NMR-based metabolomics coupled with multivariate statistical methods, on a varied group of samples comprising of human disease states, using in vitro and in vivo approaches with model organisms to evaluate the toxic potential or treatment efficacy of a drug and an efficient method for differentiating infectious peritonitis. A successful conclusion of such studies paves the way for new biomarkers and treatment strategies in the management of disease state. Once the metabolites involved in disease state are known, the occurrence of the disease can be pre-diagnosed or can be predicted on the basis of their metabolic profile. From the increasing amounts of publications in the field, it seems that NMR-based methods will continue to contribute to the area of metabolomics.

## 7.2: Future Prospects

Metabolomics is a fast-growing field that has immense potential to impact our understanding of the molecular mechanism of disease. Lifestyle, drugs, and the environment, immensely contribute to the human biocomplexity and identification of subtle metabolic variations associated with various diseases is a great challenge. A disease state commonly disturbs metabolism and leaves behind metabolic footprints that can be monitored and identified using metabolomics protocols. Metabolomics provides a global overview of such metabolic perturbations in biochemical pathways of complex diseases (cardiovascular, Alzheimer, autoimmune disease, etc.) and on drug treatment. Evaluating drug -effects and variation in drug response will provide new diagnostic, prognostic biomarkers leading to a more extensive and a precise classification of disease based on diverse etiologies, and biochemical perturbations. It should, however, be noted that this is still the learning phase and therefore the research at this stage is still in infancy, towards the development of a metabolic signature as a biomarker for a disease diagnosis or its treatment for their routine general use. Many confounding factors exist, such as the sample size, proper matching of patients and controls for age, gender, ethnic background and many other factors should be considered carefully because a change or difference in only one of these factors (e.g. a different ethnic background if several patients) can have pronounced effect on the final results. A close monitoring of diet and exercise and the possible effects of medication should also be considered along with other possible disease states. This could, in turn, streamline clinical trials

and improve outcomes. Longitudinal and demographic studies are required to confirm and expand on these initial findings, and furthermore, replication studies and blind cross validation studies are needed to validate the markers identified to attain the status of a universal biomarker. By combining the latest advancements in NMR methods and targeted metabolite profiling using sensitivity enhanced approaches, such as isotope labeling or others, a breakthrough in biomarker detection may be accomplished. While MS is more sensitive than NMR, it may not be the solution at this stage since there are problems with reproducibility arising from the chromatography of biofluids and with factors such as matrix effects and ion suppression. NMR, GC/LC-MS spectral libraries for metabolomics studies such as HMDB, Metlin and MassBank continue to evolve and expand for increased metabolome coverage, with technological advancements in NMR and MS -based metabolomics. As also, since metabolites do not exist alone but within a certain biological context, such as metabolic networks and pathways, integration of contextual information into identification can potentially reduce the ambiguity in metabolite identification. Integrating the metabolomics data with other "omics" approaches might be powerful ways to achieve these goals and enhance our understanding of the organisms as a whole. Therefore, in near future metabolomics is likely to play a pivotal role in different areas of disease and drug research.

## 7.3: Future Work Plans

- Enlarge the sample size to strengthen the results and applying the developed methodologies to other locations in the country or even abroad.
- Study the possibility of characterizing different types of cardiovascular diseases like angina pectoris, STEMI, etc. and studying the metabolic profiles of AMI patients at various time points after infraction with a fair sample size.
- Pharmacometabolomics: to evaluate the therapeutic efficacy and safety of pharmaceutical products to aid into rapid drug development and discovery.
- Behavior metabolomics: behavioral models of post-traumatic stress disorder (PTSD), fear memory consolidation and extinction, and the metabolic changes involved and does metabolic profiling can ameliorate such conditions.
- To improve the information content, NMR based metabolomics can be complemented with high throughput methodologies such as mass spectrometry which will vastly improve the amount of information obtained.

## List of Symbols and Abbreviations

| | |
|---|---|
| **$^{13}C$** | Carbon-13 |
| **$^{1}H$** | Single-proton Hydrogen |
| **ANOVA** | analysis of variance |
| **AUC** | Area under the Curve |
| **BMRB** | Biological Magnetic Resonance Data Bank |
| **C** | Carbon |
| **COSY** | Correlated spectroscopy |
| **CPMG** | Carr Purcell Meiboom Gill sequence |
| **CSF** | Cerebrospinal Fluids |
| **CV** | Cross validation |
| **D2O** | Deuterated water |
| **DA** | Discriminant Analysis |
| **DE** | Diffusion Edited |
| **F** | Fluorine |
| **FID** | Free induction decay |
| **H** | Hydrogen |
| **HMDB** | Human Metabolome Database |
| **HSQC** | Heteronuclear Single Quantum Coherence |
| **Hz** | Hertz |
| **KEGG** | Kyoto Encyclopedia of Genes and Genomes |
| **N** | Nitrogen |
| **NMR** | Nuclear Magnetic Resonance Spectroscopy |
| **NOESY** | Nuclear Overhauser Enhancement Spectroscopy |
| **OPLS** | Orthogonal PLS |
| **P** | Phosphorus |
| **PCA** | Principle Component Analysis |
| **PLS** | Projection to Latent Structures |
| **PLS-DA** | Partial least squares discriminant analysis |
| **ppm** | Parts per million |
| **ROC** | Receiver operating characteristic |

**S/N** Signal to Noise ratio

**TOCSY** Total correlation spectroscopy

**TSP** Trimethylsilyl 3-propionic acid, sodium salt

**WATER-GATE** WATER suppression by GrAdient Tailored Excitation

**α** Alpha

**β** Beta

**γ** Gamma

**δ** Delta

| **Amino acids** | **Symbols** | |
|---|---|---|
| Alanine | Ala | A |
| Cysteine | Cys | C |
| Aspartic acid | Asp | D |
| Glutamic acid | Glu | E |
| Phenylalanine | Phe | F |
| Glycine | Gly | G |
| Histidine | His | H |
| Isoleucine | Ile | I |
| Lysine | Lys | K |
| Leucine | Leu | L |
| Methionine | Met | M |
| Asparagine | Asn | N |
| Proline | Pro | P |
| Glutamine | Gln | Q |
| Arginine | Arg | R |
| Serine | Ser | S |
| Threonine | Thr | T |
| Valine | Val | V |
| Tryptophan | Trp | W |
| Tyrosine | Tyr | Y |

## Figures:

| | | | |
|---|---|---|---|
| | | triangles). | |
| | 3.9 | Statistical validation of the PLS-DA model, using 10-fold CV as represented by Accuracy, $R^2$ and $Q^2$ Values. | 69 |
| | 3.10 | Represents the Receiver operating characteristic (ROC) curves of the metabolites discriminating G1 from G2 along with their respective box plots. AUC indicates the area under the curve and the dot refer to the cutoff value maximizing sensitivity and specificity for the given samples. For each box plot, boxes denote interquartile ranges; lines denote medians, and whiskers denote $10^{th}$ and $90^{th}$ percentiles. | 70 |
| **4** | 1 | Stack plot of cumulative 1D $^1H$ CPMG NMR spectra (ranging from $\delta$0.7–$\delta$4.6 ppm) (A) and diffusion-edited NMR spectra (ranging from $\delta$0.0–$\delta$5.6 ppm) (B) obtained for sera of AMI patients (red) and normal controls (blue). The metabolite peaks of discriminatory potential assigned. Diffusion edited spectra were measured for exclusive profiling of lipid and lipoprotein metabolites through attenuating the peaks from fast-diffusing, low molecular weight metabolites. Key acronyms are: HDL; high density lipoproteins; LDL: low density lipoproteins; VLDL: very-low density lipoproteins; Met: Methionine; NAG: N-acetyl glycoproteins, Glu: Glutamate; Gln: Glutamine; Cho: Choline-$N^+(CH_3)_3$; GPC: glycerophosphocholine; PCho: Phosphatidylcholine, Ph: phospholipid; LpFA: Lipid bound fatty acid chains; PUFA: polyunsaturated fatty acids; TMAO: trimethyl-amine oxide; Gly: glycine. | 86 |
| | 2 | The two-dimensional PCA (A-C) and PLS-DA (B-D) score plots derived, respectively, from CPMG and diffusion edited (DE) spectra showing clear statistical separation between AMI (represented by red circles) and normal control (NC) samples (represented by blue triangles). Each circle and triangle represent one subject. The validation parameters ($R^2$ and $Q^2$) corresponding to each PLS-DA model are also displayed in their respective score plots. The semi-transparent red and blue ovals represent the 95% confidence interval. | 90 |
| | 3 | The diagnostic potential of each model as shown in (A) and (B) obtained from ROC analysis and (C) and (D) permutation analysis. | 91 |
| | 4 | Heat maps showing z-scores of identified 30 statistically significant metabolite entities altered in AMI patients compared to normal controls. The numbers used within bracket correspond to the metabolite identity as shown in Table 2. Here, (A) and (B) are the heat maps derived from discriminatory metabolite entities identified, respectively, in CPMG and diffusion edited (DE) spectra. The red and cyan here signify, respectively, elevation and reduction in metabolite concentration in AMI patients. The color key indicates the metabolite expression level, values (fold change): dark blue: lowest; dark red: highest. | 93 |

| | | | |
|---|---|---|---|
| | 5 | (A) Summary of metabolic pathway enrichment analysis performed in MetaboAnalyst (Version 3.0, URL: http://www.metaboanalyst.ca) using metabolite entities found significantly altered in AMI patients compared to normal control (NC) (as shown in Table 2 excluding the lipids and Glycoproteins). (B) The list of these metabolites was further used as an input for pathway impact analysis in Metaboanalyst which is based on the Over Representation Analysis (ORA) algorithm and implemented using the hypergeometric test to evaluate over representation of a particular metabolite set; provided were fold-enrichment values and one-tailed p-values corrected for multiple testing. The most significant p-values are in the red while the least significant are in yellow and white. Pathway analysis is showing altered metabolic pathways. | 94 |
| | 6 | Receiver operating characteristic (ROC) curves of the metabolites discriminating AMI patients from normal controls (as listed in Table 2) along with their respective box plots derived from CPMG NMR data set. For each box plot, boxes denote interquartile ranges, lines denote medians, and whiskers denote 10th and 90th percentiles. The ROC plot of each metabolite entity contains the AUC (i.e. area under the ROC curve) value highlighting its discriminant potential. | 95 |
| | 7 | Receiver operating characteristic (ROC) curves of the metabolites discrimnating AMI patients from normal controls (as listed in Table 2) alongwith their respective box plots derived from Diffusion edited NMR data set. For each box plot, boxes denote interquartile ranges, lines denote medians, and whiskers denote $10^{th}$ and $90^{th}$ percentiles. The ROC plot of each metabolite entity contains the AUC (i.e. area under the ROC curve) value highlighting its discriminant potential. | 96 |
| | 8 | Altered metabolic pathways for the most relevant metabolites perturbed in AMI patients with respect to normal controls. The metabolites which were found to be elevated and decreased in AMI patients compared to normal controls are shown in red and blue colored texts, respectively. Whereas, the metabolites which were significantly elevated and decreased in AMI patients compared to normal controls are highlighted here using upward (in red) and downward (in blue) arrows, respectively. | 97 |
| **5** | 1 | (A-D) Histopathology and (A'-D') SEM analysis of liver: (A/A') Control, (B/B') PYZ, (C/C') PA and (D/D') 5-OHPA. Lessions were found in PYZ, PA and 5-OHPA groups with respect to control [Nucleus (N), Kupffer cell (K), degenerated nucleus (DN), and ruptured hepatic cells (RC)]. We observed prominent DN and RC in PYZ (toxic control) and 5-OHPA groups; these features were less prominent in PA group and absent in normal control group. As evident from SEM analysis, the hepatic lessions are increasingly prominent in PYZ, PA and OHPA | 119 |

| | | | |
|---|---|---|---|
| | | groups with respect to control. | |
| | 2 | The representative 1D $^{1}$H CPMG NMR spectra of rat serum. The peaks annotated in the figure show the assignments of serum metabolites. The abbreviations used are LDL/VLDL: Low/very-low density lipoproteins; HDL: high density lipoproteins; PUFA: polyunsaturated fatty acids; NAG: N-acetyl glycoproteins; OAG: O-acetyl glycoprotein; GA: Guanadinoacetate; TMAO: Trimenthylamine-N-oxide. | 120 |
| | 3 | 2D PCA scores plots derived from 1D $^{1}$H (A-D) CPMG and (E-H) DE, NMR spectra. The various groups compared are well evident from the Figure. (A) and (E) represent the PLS-DA analysis involving all the four groups, whereas (B-D) and (F-H) represent the paired analysis. The shaded areas are the 95% confidence regions of each treatment as depicted by their respective colors. | 121 |
| | 4 | 2D PLS-DA scores plots derived from 1D $^{1}$H (A-D) CPMG and (E-H) DE, NMR spectra. The various groups compared are well evident from the Figure. (A) and (E) represent the PLS-DA analysis involving all the four groups, whereas (B-D) and (F-H) represent the paired analysis. The shaded areas are the 95% confidence regions of each treatment as depicted by their respective colors. | 122 |
| | 5 | 2D scores plots derived from the PCA (A,D) and PLS-DA (B,E) analysis of 1D $^{1}$H CPMG NMR spectra. The treatment groups compared are well evident from the Figure: (A,B) 5-OHPA vs. PYZ and (D,E) PA vs. PYZ. The shaded areas are the 95% confidence regions of each treatment as depicted by their respective colors. | 122 |
| | 6 | The potential biomarker metabolite entities identified from pair-wise PLS-DA analysis of 1D $^{1}$H CPMG spectra: (A) PYZ vs. NC, (B) 5-OHPA vs. NC and (C) PA vs. NC. The metabolites are listed in decreasing order of VIP score to highlight their discriminatory potential. | 123 |
| | 7 | The potential biomarker metabolite entities identified from pair-wise PLS-DA analysis of 1D $^{1}$H diffusion edited NMR spectra: (A) PYZ vs. NC, (B) 5-OHPA vs. NC and (C) PA vs. NC. The metabolites are listed in decreasing order of VIP score to highlight their discriminatory potential. | 123 |
| | 8 | Heat maps of statistically significant metabolite entities as shown in Table 2. Here, (A) and (B) are the heat maps derived from discriminatory metabolite entities identified, respectively, in CPMG and diffusion edited (DE) spectra. The red and cyan here signify, respectively, elevation and reduction in metabolite concentration. | 124 |
| | 9 | Box-whisker plots of metabolites that were significantly perturbed across the groups derived from CPMG spectra. | 126 |
| | 10 | Identification of the perturbed metabolic pathways by overrepresentation analysis (ORA) using the significantly | 127 |

| | | | |
|---|---|---|---|
| | | altered metabolites identified by PLSDA VIP score. The analysis was done by using a pathway library restricted to Rattus norvegicus, and p-values for ORA stand for the hypergeometric test. Test p-value (vertical axis, the intensity of colour) and impact factor (horizontal axis, the size of circle). *Abbreviations used:* Mb: Metabolism, Val: Valine, Leu: Leucine, Ile: Isoleucine, Asp: Aspartate, Ala: Alanine, Arg: Arginine, Ser: Serine, Thr: Threonine, Pro: Proline. | |
| **6** | 1 | (A) Schematic showing the $^{1}$H NMR based method –named here as "*add antibiotic to detect by NMR*" or "*AADNMR*"- for rapid identification of bacterial/mycobacterial infection in a biological/clinical sample. (B) Generalized structure of cell membrane with phospholipid bilayer (top right). The membranes of bacteria are structurally similar to the cell membranes of eukaryotes, except that bacterial/mycobacterial membranes consist of cyclic or monounsaturated fatty acids (rarely, polyunsaturated fatty acids) and do not normally contain sterols[12,13]. The numbers in the right bottom panel (representing the cyclopropyl fatty acid geometry) represent the $^{1}$H NMR chemical shifts of ring methylene protons as reported earlier[14]. | 142 |
| | 2 | An overlay of the chemical shift matched regions of $^{1}$H-$^{1}$H TOCSY spectra of PD effluent samples obtained after 4 hours dwell time from suspected PD patients instilled with broad spectrum antibiotic intraperitoneally: one PD patient was having bacterial infection (blue) as evident from the presence of marker peak, whereas in other suspected case (red) the absence of marker peak simply rule out the possibility of any such infection. The correlation peak between marker peak (*) and peak at 1.5 ppm (#) confirm the presence of cyclic fatty acids of microbial origin in accordance with the previous reports (see the text). | 144 |
| | 3 | Representative one-dimensional (1D) $^{1}$H NMR spectra of extracellular metabolite solutions obtained, respectively, from ampicillin treated (red) and untreated (blue) bacterial cell cultures (grown in the LB media at 37 °C till the optical density of culture at 600 nm reaches to ~0.6): (A) *Escherichia coli* [ATCC-25922] (as a representative of Gram –ve bacteria) and (B) *Staphylococcus aureus* [ATCC-3160] (as a representative of Gram +ve bacteria). | 145 |
| | 4 | Stack plot of one-dimensional (1D) $^{1}$H NMR spectra of some of the antibiotics (antibacterials) most frequently used in the present study: (A) Ampicillin (effective against both Gram +ve and Gram –ve bacteria), (B) Azolin (effective against Gram +ve bacteria) and (C) Vancomycin (effective against Gram +ve bacteria). Spectra clearly reveal that there is no peak in the spectral region defined by the marker peak (i.e. between 0.40 and 0.68 ppm). | 146 |
| | *5* | Stack plot of 1D $^{1}$H NMR CPMG spectra of urine samples of | 147 |

| | | | |
|---|---|---|---|
| | | patients with gram (-ve) bacterial urinary tract infection (UTI) recorded at 800 MHz. (A-E) and (A'-E') represent, respectively, the control and ampicillin treated infected urine samples: (A/A') *Klebsiella pneumoniae*, (B/B') *Escherichia coli*, (C/C') *Pseudomonas aeruginosa*, (D/D') *Enterobacter aerogenes*, and (E/E') *Proteus mirabilis*. | |
| | 6 | Stack plot of one-dimensional (1D) $^{1}$H NMR spectra of human embryonic kidney (HEK-293) cell line culture (grown in the presence of broad spectrum of anti-bacterials as described above): (A) cell suspension and (B) supernatant obtained after sonication (at 4 °C) and centrifugation at 6000 rpm for 10 minutes. | 148 |
| | 7 | Stack plot of one-dimensional (1D) $^{1}$H CPMG NMR spectra of *Escherichia coli* bacterial culture (grown in standard LB media at 37 °C): (A) live cell suspension, (B) sonicated cell suspension, (C) cell suspension treated with lysozyme (at 37 °C for 1 hour) and (D) cell suspension treated with lysozyme (at 37 °C for 1 hour) followed by sonication. In each case, the cell suspension obtained in the exponential phase (when OD@600nm = 0.6) has been used. | 149 |
| | 8 | Stack plot of representative one-dimensional (1D) $^{1}$H NMR spectra of PD effluent (2.5 %, after 4 hour dwell time) from PD patients. (A) Represents the control PD effluent sample of a PD patient with no infection. (B and C) represent PD effluent samples from a PD patient having bacterial peritonitis and are obtained, respectively, before (B) and after (C) the intra-peritoneal antibiotic treatment has been started. In (C), the presence of marker peak at ~0.61 ppm confirms the bacterial peritonitis which is otherwise difficult to predict through the routine cell culture based method if the patient has already been given the intraperitoneal antibiotic treatment. (D) Represents PD effluent sample from a PD patient having fungal peritonitis and given intraperitoneal antibiotic treatment. The absence of marker peak indirectly hints towards the fungal infection. | 151 |
| | 9 | Stack plot of one-dimensional (1D) $^{1}$H CPMG NMR spectra of: (A) commercial 2.5 % PD dialysate solution and (B, C) bacterial cell culture (*Staphylococcus aureus*, grown in the PD solution): (B) without antibiotic treatment and (C) after incubating the bacterial cell culture with ampicillin at 37 °C for 1 hour. The marker peak in ampicillin treated sample was found to be upfield shifted (close to 0.5 ppm) suggesting that the intraperitoneal conditions may be involved in modulating the microbial cyclic fatty acids qualitatively. | 152 |
| | 10 | Summary of some of the potential medical and biomedical applications of the proposed <AADNMR> method. | 155 |

## Tables: