# Metabolomic profiling of *Momordica Charantia* L. using $^{1}$H NMR spectroscopy and multivariate statistical analysis

**Ankit**

*A dissertation submitted for the partial fulfilment of*
*BS-MS dual degree in Science*


**Indian Institute of Science Education and Research Mohali**

**April 2018**

# Certificate of Examination

This is to certify that the dissertation titled "Metabolomic profiling of *Momordica Charantia* L. using $^{1}$H NMR spectroscopy and multivariate statistical analysis" submitted by Mr. Ankit (Reg. No. MS13077) for the partial fulfilment of BS-MS dual degree programme of the Institute, has been examined by the thesis committee duly appointed by the Institute. The committee finds the work done by the candidate satisfactory and recommends that the report be accepted.

Dr. Samir Kumar Biswas  Dr. Sandeep K. Goyal  Dr. Kavita Dorai
(Supervisor)

Dated: April 20, 2018

# Declaration

The work presented in this dissertation has been carried out by me under the guidance of Dr. Kavita Dorai at the Indian Institute of Science Education and Research Mohali.

This work has not been submitted in part or in full for a degree, a diploma, or a fellowship to any other university or institute. Whenever contributions of others are involved, every effort is made to indicate this clearly, with due acknowledgement of collaborative research and discussions. This thesis is a bonafide record of original work done by me and all sources listed within have been detailed in the bibliography.

**Ankit**
(Candidate)
Date: April 20, 2018

In my capacity as the supervisor of the candidate's project work, I certify that the above statements by the candidate are true to the best of my knowledge.

**Dr. Kavita Dorai**
Associate Professor
Department of Physical Sciences
IISER Mohali
Date: April 20, 2018

# Acknowledgements

First of all, I would like to thank my supervisor, Dr. Kavita Dorai who gave me the opportunity to perform this Project. I would like to thank my committee members, Dr. Kavita Dorai (Supervisor), Dr. Samir Kumar Biswas, and Dr. Sandeep K. Goyal for their review and contribution to my thesis. I am very grateful to have such a wonderful committee for taking out the time to read my thesis and giving valuable suggestions.

I would then like to thank Prof. Debi P Sarkar, Director, IISER Mohali and Prof. N. Sathyamurthy, former Director, IISER Mohali, for providing all the help needed for my research.

I am obliged to the faculty of the Department of Physical Sciences of IISER Mohali whose direction and guidance has been very fruitful during the Lectures and Lab hours. My sincere thanks to all the faculty members of IISER Mohali for various direct and indirect interactions all through My BS-MS. I would also like to thank Department of Science and Technology (DST) for providing me INSPIRE Scholarship for Higher Education (SHE).

I would like to thank my doctoral fellow guides Sumit Mishra and Navdeep Gogna, for everything. Special thanks to all NMR Lab members Without their help, I would not have been able to complete my research. They are the most understanding and helpful people I have ever met and I am forever grateful.

I am highly obliged to the NMR facility at IISER, Mohali for providing me everything needed for my experiments. For the completion of this Project work, various experiments were performed on the 600 MHz spectrometer and on the 400 MHz spectrometer.

Finally i would like to thank my Parents, Brother and Sister for their love and support.

Ankit

IISER MOHALI

# List of Figures

# Notation

| | |
|---|---|
| $^{1}H$ | Proton |
| $^{13}C$ | 13 Carbon |
| $\delta$ | Chemical Shift |
| $\mu s$ | micro seconds |
| $\mu L$ | micro liter |
| MC | *Momordica Charantia* L. |
| NMR | Nuclear Magnetic Resonance |
| 1D | 1 Dimensional |
| 2D | 2 Dimensional |
| COSY | Correlations Spectroscopy |
| TOCSY | Total correlation spectroscopy |
| HSQC | Heteronuclear Single Quantum Coherence |
| HMBC | Heteronuclear Multiple Bond Correlation |
| HMQC | Heteronuclear Multiple Quantum Correlation |
| JRES | J-resolved spectroscopy |

# Abstract

The work presented in this thesis project shows the application of NMR spectroscopy in the area of metabolomics. This study signifies the applications of NMR-based metabolomics in the area of plant science and the thesis work mainly focuses on the utility of NMR-based metabolomics for metabolite profiling of different sections viz. Skin, Seed and the Flesh of *Momordica Charantia* (MC) or Bitter gourd or Karela in Hindi Fruit. profiled for metabolites helpful in treating type 2 Diabetes Mellitus.

The chapters describe the basic principle of NMR spectroscopy and metabolomics, as applied to metabolic studies. The study enlists the various pre-treatment and pre-processing steps required for conversion of a raw NMR data into the data suitable to do multivariate analysis namely PCA and OPLS-DA.

The results of present work demonstrated that the combined use of multivariate statistical techniques with NMR spectroscopy or Mass spectrometry is a feasible approach to discriminate different parts of (MC) fruit and finding relative concentration of secondary metabolites present in each part. Moreover, the sensitivity of this analytical tool allowed the identification of medicinally significant plant secondary metabolites, paving the way for further investigations.

# Contents

# Chapter 1

# Introduction

## 1.1 NMR Spectroscopy

NMR Spectroscopy is a unique spectroscopic tool having applications in various field Like Medicine, Food Technology, Structural Biology, Physics etc. NMR technique is very rich in history as there been recognised with seven Nobel prizes in Physics, Chemistry and medicine, starting from the beginning of NMR which is 1952 Nobel prize was given to Felix Bloch and E.M. Purcell who independently discovered the phenomenon of NMR Spectroscopy and R. R. Ernst got Nobel prize in 1991 for discovering and developing two dimensional NMR methods. Among all the other spectroscopy techniques NMR is unique because it is a non-invasive/non-destructive method which means if you take a sample of your compound and record your data you Will get back the sample intact nothing happens to your sample. In NMR spectroscopy you can actually study each and every atom selectively or looked at them in a very sensitive manner which is one of the significant advantages of this technique in comparison to others. As this is a quantitative technique so you can actually quantify the amount of sample present in your study, which is very useful in many applications like the Metabolomics where we looked at the quantity of sample in a given mixture where we like to know the presence of different amount of metabolites present. we can study all the states of matter viz. solids, liquids, and gases which is also a good advantage of NMR spectroscopy.

### 1.1.1 What is Spectroscopy

Spectroscopy methods involves the study of the interaction of radiation with matter. So in spectroscopy, we take a electromagnetic radiation or wave and shine it on the sample which consists of matter or molecule. This molecule will then respond to this incoming light which we write as h$\nu$ which is the energy of one photon. where h is the Planck's constant and $\nu$ is the frequency of the wave. when this matter absorbed the energy part of this photon then some energy absorbed and rest is radiated back or scattered. Now the energy which is shown upon this matter will only be absorbed, If the difference in the energy level matches the energy we are supplying.

## 1.2 NMR Phenomenon

Nuclei having non-zero spin are considered to be NMR active, as these nuclei have non-zero total angular momentum, the phenomenon of nuclear magnetic resonance can be observed for them. Now a nucleus with spin I has a nuclear magnetic moment associated with it given by:

$$\mu = \gamma \hbar I \tag{1.1}$$

Where $\gamma$ is called the *gyromagnetic ratio or magnetogyric ratio* and $\hbar I$ represents the spin magnetization, which is a vector operator where I is a dimensional operator representing the total angular momentum of the nuclear spin.
Now this $\mu$ magnetic dipole shown above interacts with the external magnetic field and the energy of interaction is given as:

$$E = -\mu.B \tag{1.2}$$

Now spliting $\mu$ into three components $\mu_x, \mu_y$ and $\mu_z$ and $B$ into $B_x, B_y$ and $B_z$. Applying magnetic field in z direction eliminate x and y components which simplifies above equation as:

$$E = -\vec{\mu_z}\vec{B_z} \tag{1.3}$$

From (1.1) writing for $\mu$ the above equation further simplifies to:

$$E = -\gamma I_z B_0 \tag{1.4}$$

where $B_0$ is the external magnetic field apllied in the z direction. From Quantum Mechanics if you have a spin $I$ then it has posibility of two $2I+1$ spin "states" which are energy levels. For example suppose $I = 1/2$ which is the case for $^1H$ it becomes 2 so that means hydrogen nucleus which is a proton has now 2 spin states given by either $+1/2$ or $-1/2$.

From (1.4) $I_z$ is quantized and can only take $2I+1$ values. For $I = 1/2$; $I_z = +1/2\hbar$ and $-1/2\hbar$

Now this energy of interaction has two values depending on spin $I_z$ is either $+1/2$ or $-1/2$. The system is thus split into two states given as:

$$E_{1/2} = -1/2\hbar B_0 \quad ; \quad E_{-1/2} = 1/2\hbar B_0 \tag{1.5}$$

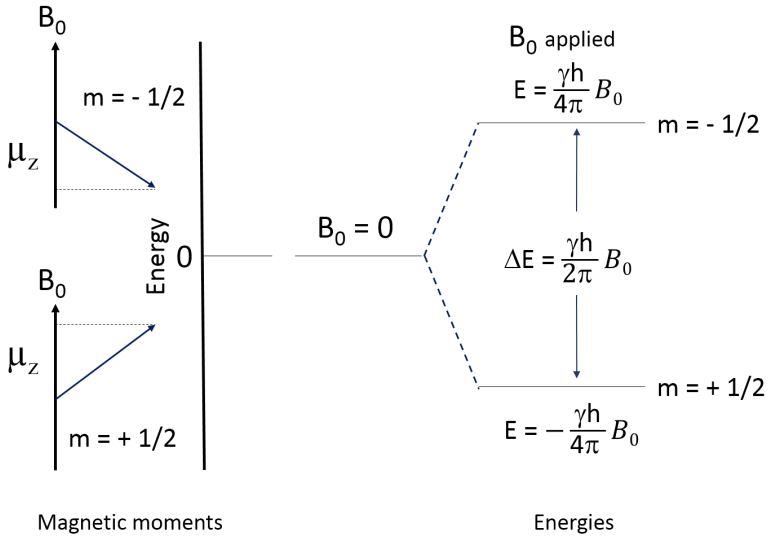


Figure 1.1: Energy levels in magnetic field

Because of the interaction of this spin 1/2 nucleus with the magnetic field we ended up with two energy levels $\alpha$ (lower state) and $\beta$ (exited state) which is characterized by azimuthal quantum number m = 1/ 2 (spin up) and m = - 1/ 2 (spin down). Now taking molecules from the ground state to excited state the energy gap should be exactly equal to the difference in energy levels given by:

$$\Delta E = \gamma\hbar B_0 \tag{1.6}$$

using

$$\Delta E = h\nu \tag{1.7}$$

or

$$\Delta E = h\omega/2\pi \tag{1.8}$$

Where $\omega = \gamma B_0$ is the Larmor frequency of the molecule and given in radio frequency range.

This is the basic idea behind the terminology **Resonance** (which results in energy transfer) in NMR.

In a sample if there are N number of molecules are present each containing a given nucleus I then the molecules will be partitioned into the two states as:

$$N = N_\alpha + N_\beta \tag{1.9}$$

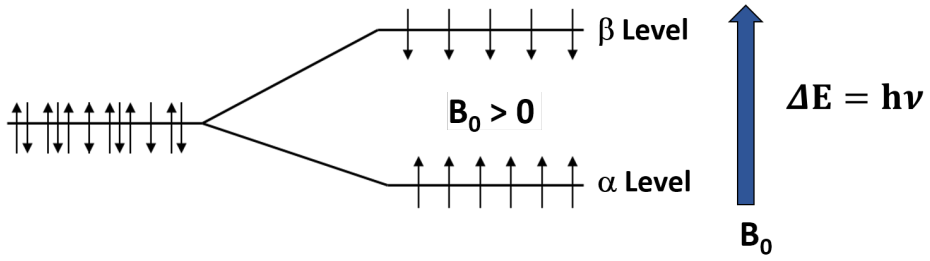


Figure 1.2: Quantum energy levels

According to **Boltzman law**, the population ratio is governed as:

$$N_\beta/N_\alpha = e^{-\Delta E/kT} \tag{1.10}$$

where k is the Boltzmann constant and T is temperature.

In NMR spectroscopy $\Delta E = \gamma\hbar B_0$ giving:

$$N_\beta/N_\alpha = e^{-\gamma\hbar B_0/kT} \tag{1.11}$$

To increase the sensitivity (increased $\Delta E$) of NMR I have to basicaly rely on three parameters

1. Going to higher $\gamma$
2. Increasing the magnetic field ($B_0$)
3. Lowering the temperature T.

Sensitivity can also be increased by increasing the concentration of the sample.

## 1.3 The Vector Model

The vector model of NMR spectroscopy is a visual way of comprehending the results of delay times and pulses in NMR experiments. After placing the sample into the magnet, the spin begin precessing about the external magnetic field, When the RF field is absent we usually take magnetic field in the z-direction. The spins present in the sample are all precessing out of phase with their magnetic moments pointing in random directions. Now If the magnetic moments were all to point in random directions, then the small magnetic field that each generates will cancel one another out and there will be no net effect. At equilibrium the magnetic moments (commonly represented as $\mu$) are not aligned randomly but in such a way that when their contributions are all added up then there is a net magnetic field along the direction of the applied field $B_0$. This is called the bulk magnetization of the sample.

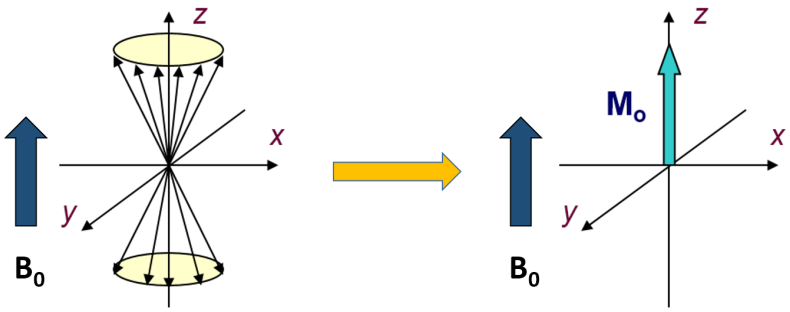


Figure 1.3: Bulk Magnetization

### 1.3.1 Larmor precession

When a Radiofrequency pulse is applied Let's say in x-direction the tip of the magnetization vector get tilted away from the z-axis, making an angle $\beta$ to that axis. Once tilted away from the z-axis we find is that the magnetization vector rotates about the direction of the magnetic field sweeping out a cone with a constant angle. The vector is said to precesses about the field and this particular motion is called *Larmor precession*.

If the magnetic field strength is $B_0$, then the frequency of the Larmor precession is $\omega_0$ (in rad $s^{-1}$) given by:

$$\omega_0 = -\gamma B_0 \tag{1.12}$$

Or if we want frequency in Hertz, it is given by:

$$\nu_0 = -\frac{1}{2}\gamma B_0 \tag{1.13}$$

Where $\gamma$ is the *gyromagnetic ratio.*

### In Rotating Frame

When viewed in the Rotating Frame the Larmor precession of the magnetization will appear to be at $(\omega_0 - \omega_{(rot.frame)})$. where $\omega_0$ is the fixed frame precession frequency and $\omega_{(rot.frame)}$ is the rotational frame frequency.
This difference in frequency is called the offset and is given the symbol $\Omega$:

$$\Omega = \omega_0 - \omega_{(rot.frame)} \tag{1.14}$$

Using the relationship between the magnetic field and the precession frequency:

$$\omega = -\gamma B \tag{1.15}$$

We can compute the apparent magnetic field in the rotating frame, from the apparent Larmor frequency $\Omega$:

$$\Omega = -\gamma \Delta B \tag{1.16}$$

the apparent magnetic field, given by $\Delta B$ can be given by:

$$\Delta B = -\frac{\Omega}{\gamma} \tag{1.17}$$

This apparent magnetic field in the rotating frame is usually called the reduced field, $\Delta B$.

## 1.4 NMR Relaxation

In the excited system the magnetization has tendency to return to its equilibrium position and size this behavior of system is known as **Relaxation**. The NMR signal relaxes back to ground state through two distinct processes, one is Transverse Relaxation and other is Longitudinal Relaxation.

### 1.4.1 The Spin-Lattice or Longitudinal Relaxation ($T_1$)

The amount of time a collection of nuclie will take to reach Boltzmann distribution after immersing in an external field is controlled by $T_1$. The process is governed by loosing energy in system to lattice as heat, which results in a minimal temperature increase in sample.

It's a time dependent exponential decay process of magnetization along the Z axis given by:

$$M_z = M_0(1 - e^{\frac{-t}{T_1}}) \tag{1.18}$$

### 1.4.2 The spin-spin or transverse Relaxation ($T_2$)

The magnetization loss in xy direction is known as transverse relaxation. In this process the nuclei starts to lose their phase coherence and return to a random arrangement around the z axis. The process is goverened by exchanging energy between nucleus excited and low energy state.

The spin-spin magnetization will decay as:

$$M_x = M_y = M_0 e^{\frac{-t}{T_2}} \tag{1.19}$$

## 1.5 Chemical Shift

The difference between the resonance frequency of the nucleus and a standard, relative to the standard is known as **Chemical Shift** denoted by $\delta$. For example, in the case of proton NMR the reference compound is TMS.

The chemical shift of the line is given by:

$$\delta = \frac{\nu - \nu_{TMS}}{\nu_{TMS}} \tag{1.20}$$

where $\nu$ (in Hz) is the the frequency of the line we are interested in and $\nu_{TMS}$ (in Hz) is the frequency of the line from TMS.

Chemical shift in *parts per million* or ppm is given by:

$$\delta_{ppm} = \frac{\nu - \nu_{TMS}}{\nu_{TMS}} * 10^6 \tag{1.21}$$

### $B_0-$ Electron Interactions

Material placed in a magnetic field magnetized to some degree and gives rise to extra magnetic field that modifies the external static magnetic field.
The local magnetic field result in:

$$B = B_0(1 - \sigma) \tag{1.22}$$

where $\sigma$ is the shielding constant.
The formal correction for chemical shielding is given by the **Zeeman Hamiltonian** as:

$$H_{zeeman} = -\gamma\vec{I}(1 - \tilde{\sigma})\vec{B} \tag{1.23}$$

## 1.6 J-Coupling

The indirect interaction between nucleus which arises from hyperfine interactions between the nuclei and local electrons is called spin-spin coupling or J-coupling. It is a scalar coupling
The J-coupling Hamiltonian is given as:

$$H_J = 2\pi J\vec{I}.\vec{S} \tag{1.24}$$

where J is the coupling constant and $\vec{I}.\vec{S}$ represents the product operator.

## 1.7 Dipolar Coupling

In Dipolar Coupling the Dipole fields from nearby spins interacts i.e they are coupled.
The nuclear Dipolar Coupling Hamiltonian is given as:

$$H_{dipole} = -\frac{\mu_0\gamma_I\gamma_S}{4\pi r^3}(\vec{I}.\vec{S} - \frac{3}{r^2}(\vec{I}.\vec{r})(\vec{S}.\vec{r})) \tag{1.25}$$

Where $\vec{r}$ vector from spin I to spin S.

## 1.8 Metabolomics

The "Omics" sciences is aimed at identifying the entire set of biomolecules which include genes (genomics), mRNA (transcriptomics), proteins (proteomics) and metabolites (metabolomics) contained in a biological system in a non-targeted and non-biased manner [2]. Metabolomics is considered as the end point of "Omics cascade" [3].
Metabolomics was first defined by Fiehn in 2001 as quantitative and qualitative analysis of all metabolites of the biological system. The metabolomics study is basically focused on all set of low molecular weight metabolites typically in the range of <1500 Da involving various organic molecules such as Carbohydrates, fatty acids, essential Amino Acids, proteins, organic acids etc. Majority of scintific studies in metabolomics use biofluids or cell extracts. Studies involving cell or tissue extracts such as those on tumour cells and tissues, plant cells, and tissues have also been done. Metabolomics study can be divided into two groups "targeted" analyses in which physico-chemical characteristics of the metabolites are known and an exhaustive separation of them is usually required for the quantification. and other is "non-targeted" analysis in which significant metabolites prior to the analysis are unknown.
Finally depending on specific metabolomics studies the analysis and data manipulation of metabolites in a complex multivariate data set can be visualized using different chemometric and bioinformatic methods.

# Chapter 2

# Two Dimensional NMR Spectroscopy

## 2.1 Introduction

One Dimensional NMR Spectra shows intensity vs frequency plots; In the case of two dimensional (2D) spectroscopy intensity is plotted as a function of two frequencies (usually represented as F1 and F2) i.e, in 2D NMR spectra each peak has two frequency coordinates as well as amplitude showing a correlation between two quantities. Figure 2.1 shows a spectrum of a hypothetical molecule containing just two protons, B and X, coupled to each other. In a 1D spectrum multiplets are formed because of scalar coupling, in the case of 2D spectra multiplets consists of an array of individual peaks giving a rectangular or square outline [1]. such array of peaks can be shown in the schematic COSY spectrum shown below:

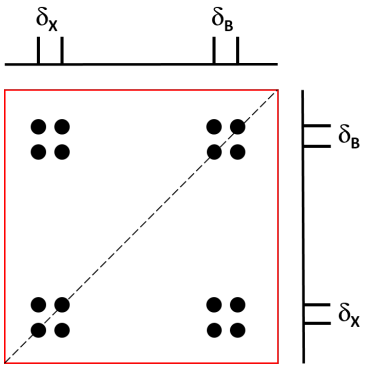


Figure 2.1: COSY Spectrum for two coupled spins, B and X

In Figure 2.1 there are two diagonal-peak multiplets centred at $F_1 = F_2 = \delta_B$ and $F_1 = F_2 = \delta_X$. The first cross-peak multiplet centred at $F_1 = \delta_B, F_2 = \delta_X$ and second cross-peak multiplet is centred at $F_1 = \delta_X, F_2 = \delta_B$. the schematics shows a scalar coupling between the two protons at shifts $\delta_B$ and $\delta_X$.

## 2.2 General scheme for 2D NMR

A general 2D pulse sequence consists of a preparation period frequently a $90^0$ pulse which generates some kind of coherence which is going to evolve during the following period. we then have the evolution period for a variable time $t_1$ and we are allowed to whatever coherence we generated during the preparation period to evolve during $t_1$, the important point here is that we do not make any observation during this time. then at the end, we have the mixing period which is the crucial part because it causes the magnetization to move from one spin to another that moves the multiple coherence into an observable coherence. And finally we have a normal detection period which is the second running time given by $t_2$ during this time we really do record the signal that has to be an observable signal as a Free Induction decay or FID [1]. Figure 2.2 below shows a general scheme for a one-dimensional spectroscopy.


Figure 2.2: A general 2D Pulse Sequence

### 2.2.1 Recording of a 2D NMR spectra

It starts with setting time $t_1$ to zero which means the preparation mixing periods are written on top of each other and we record an FID as a function of $t_2$ which digitized by a computer and gets stored separately. then we do a first increment in $t_1$ by a fixed amount $\Delta_1$ followed by a second increment $\Delta_2$ and so on. every value of $t_1$ requires a separate experiment. Thus recording a 2D experiment data set involves repetition for increasing values of $t_1$ and recording an FID as a function of $t_2$ for each value of $t_1$.

## 2.3 Two Dimensional Experiments

The basic process involved in a 2D experiments involving magnetization transfer from one spin to another via scalar coupling is given by;

$$I_{1X} \xrightarrow{Coupling} \underset{Spin1}{2I_{1Y}I_{2Z}} \xrightarrow{90^0(X)} \underset{Spin2}{2I_{1Z}I_{2Y}} \tag{2.1}$$

### 2.3.1 COrrelation SpectroscopY (COSY)

The pulse sequence for this experiment is given by:

$$90^0_X \rightarrow t_1 \rightarrow 90^0_X \rightarrow aquire(t_2) \tag{2.2}$$

Here the preparation period is a $90^0$ pulse the mixing period is just a another $90^0$ pulse in the X direction. For a two coupled spin system starting with a equilibrium magnetization on spin 1 $I_{1Z}$ the first $90^0$ pulse takes that to $-I_{1Y}$ and the offset for $t_1$ is given as:

$$-I_{1Y} \xrightarrow{\Omega_1 t_1 I_{1Z}} -cos(\Omega_1 t_1)I_{1Y} + sin(\Omega_1 t_1)I_{1X} \tag{2.3}$$

and then we allowed coupling to evolve as:

$$-cos(\Omega_1 t_1)I_{1Y} \xrightarrow{2\pi J_{12} t_1 I_{1Z} I_{2Z}} -cos(\pi J_{12} t_1)cos(\Omega_1 t_1)I_{1Y} + sin(\pi J_{12} t_1)cos(\Omega_1 t_1)2I_{1X}I_{2Z} \tag{2.4}$$

$$sin(\Omega_1 t_1)I_{1X} \xrightarrow{2\pi J_{12} t_1 I_{1Z} I_{2Z}} cos(\pi J_{12} t_1)sin(\Omega_1 t_1)I_{1X} + sin(\pi J_{12} t_1)sin(\Omega_1 t_1)2I_{1Y}I_{2Z} \tag{2.5}$$

After we allowed coupling and the offset to act, then we need to apply the final pulse in the sequence to both spins giving us four terms as:

$$\begin{aligned}
-cos(\pi J_{12} t_1)cos(\Omega_1 t_1)I_{1Y} &\xrightarrow{(\pi/2)(I_{1X}+I_{2X})} -cos(\pi J_{12} t_1)cos(\Omega_1 t_1)I_{1Z} \\
sin(\pi J_{12} t_1)cos(\Omega_1 t_1)2I_{1X}I_{2Z} &\xrightarrow{(\pi/2)(I_{1X}+I_{2X})} -sin(\pi J_{12} t_1)cos(\Omega_1 t_1)2I_{1X}I_{2Y} \\
cos(\pi J_{12} t_1)sin(\Omega_1 t_1)I_{1X} &\xrightarrow{(\pi/2)(I_{1X}+I_{2X})} cos(\pi J_{12} t_1)sin(\Omega_1 t_1)I_{1X} \\
sin(\pi J_{12} t_1)sin(\Omega_1 t_1)2I_{1Y}I_{2Z} &\xrightarrow{(\pi/2)(I_{1X}+I_{2X})} -sin(\pi J_{12} t_1)sin(\Omega_1 t_1)2I_{1Z}I_{2Y}
\end{aligned} \tag{2.6}$$

Now taking the observable term in above Eq. 2.6 which is the third term (given in red); observable on spin 1 (because it comes as $I_{1X}$) and modulated in $t_1$ at frequnecy $\Omega_1$ then this term gives rise to a diagonal peak at $(\omega_1, \omega_2) = (\Omega_1, \Omega_1)$.

Fourth term (given in green) is observable on spin 2 and gives rise to two different peaks also modulated in $t_1$ at frequnecy $\Omega_1$ then this term gives rise to a cross peak at $(\omega_1, \omega_2) = (\Omega_1, \Omega_2)$ [1].

The first and second terms are unobservable.

Similarly, in the case of equilibrium magnetization on spin 2 which is $I_{2Z}$ we end up with two more terms given by:

$$\begin{aligned} &cos(\pi J_{12} t_1) sin(\Omega_2 t_1) I_{2X} \\ &-sin(\pi J_{12} t_1) sin(\Omega_2 t_1) 2 I_{1Y} I_{2Z} \end{aligned} \tag{2.7}$$

Here the fifth term (given in blue) is observable on spin 2 and is modulated in $t_1$ at frequency $\Omega_2$ gives rise to a diagonal peak at $(\omega_1, \omega_2) = (\Omega_2, \Omega_2)$.

And the sixth term (given in magenta) is observable on spin 1 and is modulated in $t_1$ at frequency $\Omega_2$ gives rise to a cross peak at $(\omega_1, \omega_2) = (\Omega_2, \Omega_1)$ [1].

In schematic terms the COSY spectra will look like Figure 2.3 given below:

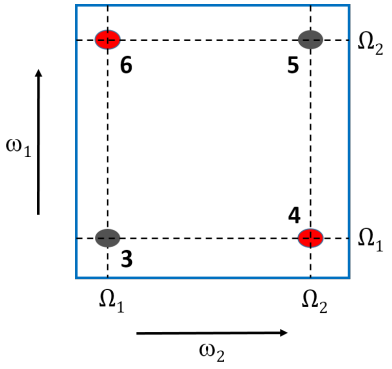


Figure 2.3: Schematic of A COSY spectrum

## 2.4 Heteronuclear correlation spectra

In these spectrums, we will see cross peaks indicating a coherence transfer through a heteronuclear coupling. Typically these correlation spectra involve $^{13}C$ and $^{1}H$ chemical shifts. And these spectra are also termed as shift correlation maps or shift correlation spectra.

### 2.4.1 Heteronuclear Single-Quantum Correlation (HSQC)

The pulse sequence for HSQC experiment is given by Figure 2.4 below

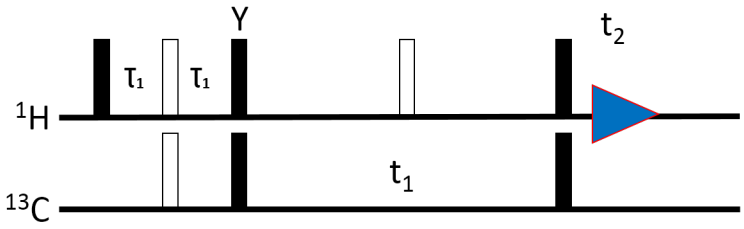


Figure 2.4: HSQC pulse sequence

In the figure above the filled rectangles represent $90^0$ pulse, open rectangles represent the $180^0$ pulse and the solid blue triangle represent the FID.
The first spin echo at the end of first period we got a in-phase term $I_Y$ and an anti-phase term $2I_XS_Z$.

$$cos(2\pi J_{IS}\tau_1)I_Y - sin(2\pi J_{IS}\tau_1)2I_XS_Z \tag{2.8}$$

Now the anti-phase term is transferred by two pulses in next period to S spin and comes out as

$$-sin(2\pi J_{IS}\tau_1)2I_XS_Z \xrightarrow{(\pi/2)I_y} sin(2\pi J_{IS}\tau_1)2I_ZS_Z \xrightarrow{(\pi/2)S_X} -sin(2\pi J_{IS}\tau_1)2I_ZS_Y \tag{2.9}$$

Then we have evolution period $t_1$ and the coupling will be refocused and the only thing which is acting is S spin of Z therefore;

$$\begin{aligned} -sin(2\pi J_{IS}\tau_1)2I_ZS_Y \xrightarrow{\Omega_S t_1 S_Z} & -cos(\Omega_S t_1)sin(2\pi J_{IS}\tau_1)2I_ZS_Y \\ & +sin(\Omega_S t_1)sin(2\pi J_{IS}\tau_1)2I_ZS_X \end{aligned} \tag{2.10}$$

Now just to compelete the calculation we need to apply the effect of $\pi$ pulse as:

$$\xrightarrow{\pi I_X} cos(\Omega_S t_1)sin(2\pi J_{IS}\tau_1)2I_ZS_Y - sin(\Omega_S t_1)sin(2\pi J_{IS}\tau_1)2I_ZS_X \tag{2.11}$$

And finally we get back to the I spin:

$$-cos(\Omega_S t_1)sin(2\pi J_{IS}\tau_1)2I_YS_Z \tag{2.12}$$

Which is our observable term and the spectrum will look like Figure 2.5 given below.

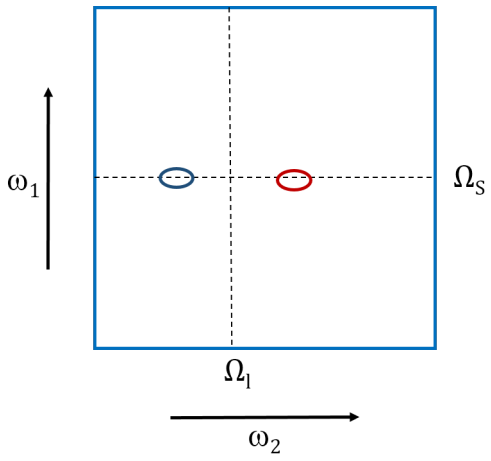


Figure 2.5: Schematic of A HSQC Spectra

### 2.4.2 Heteronuclear Multiple-Quantum Correlation (HMQC)

HMQC is an alternative to HSQC experiment the pulse sequence for this is given below

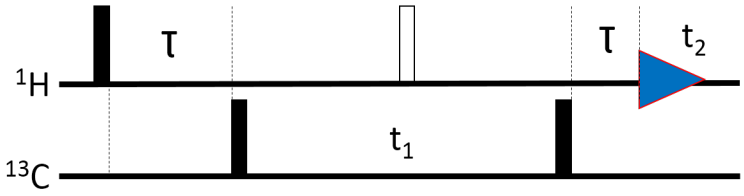


Figure 2.6: HMQC pulse sequence

The filled rectangles represent $90^0$ pulse, open rectangles represent the $180^0$ pulse and the solid blue triangle represent the FID.

At the end of firt delay $\tau$ the system will evolve under the influence of scalar couplig given by:

$$-cos(\pi J_{IS}\tau)I_Y + sin(\pi J_{IS}\tau)2I_X S_Z \tag{2.13}$$

Now only the $2I_X S_Z$ term will be transffered to multiple quantum (MQ) ignoring the $I_Y$ term. then the MQ term evolves under the S spin offset to give

$$-sin(\pi J_{IS}\tau)2I_XS_Y \xrightarrow{\Omega_S t_1 S_Z} -cos(\Omega_S t_1)sin(\pi J_{IS}\tau)2I_XS_Y \\ +sin(\Omega_S t_1)sin(\pi J_{IS}\tau)2I_XS_X \tag{2.14}$$

Then the final pulse only transfers the $2I_XS_Y$ to be observed by giving

$$-cos(\Omega_S t_1)sin(\pi J_{IS}\tau)2I_XS_Z \tag{2.15}$$

which further evolve under the coupling giving two terms

$$-cos(\Omega_S t_1)sin(\pi J_{IS}\tau)2I_XS_Z \xrightarrow{2\pi J_{IS}\tau I_Z I_S} -cos(\pi J_{IS}\tau)cos(\Omega_S t_1)sin(\pi J_{IS}\tau)2I_XS_Z \\ -sin(\pi J_{IS}\tau)cos(\Omega_S t_1)sin(\pi J_{IS}\tau)I_Y \tag{2.16}$$

So at the end when we observe $t_2$ it is only the in-phase term in $I_Y$ which is observable.

$$-[sin(\pi J_{IS}\tau)]^2 cos(\Omega_S t_1)I_Y \tag{2.17}$$

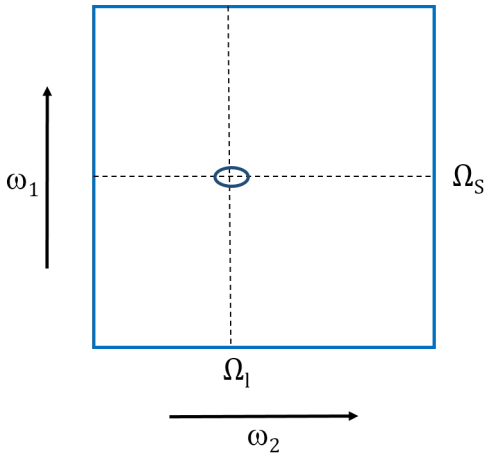


Figure 2.7: Schematic of A HMQC Spectra

## 2.5 TOtal Correlation SpectroscopY (TOCSY)

TOCSY or total correlation spectroscopy is a homonucleus experiment just like COSY. In TOCSY we will see a correlation between spins which are either coupled or have an unbroken chain of couplings between them. It is useful in picking extended spin systems so it provides more useful and complimentary information than COSY.

The TOCSY pulse sequence is given by;

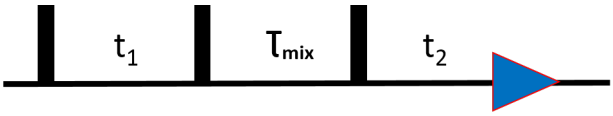


Figure 2.8: TOCSY pulse sequence

It is similar to COSY we start with $90^0$ $t_1$ then the mixing period $\tau_{mix}$ consist of a special pulse sequence which creates an isotropic mixing a special type of coherence transfer between coupled spins. at the start of isotropic mixing we only have the Z-magnetization which also remains after the mixing ends which is done by phase cycling [1].

The effect of isotropic mixing on Z-magnetization is given by;

$$\begin{aligned} I_{1Z} \rightarrow \frac{1}{2}[1 + cos(2\pi J_{12}\tau_{mix})]I_{1Z} + \frac{1}{2}[1 - cos(2\pi J_{12}\tau_{mix})]I_{2Z} \\ -sin(2\pi J_{12}\tau_{mix})\frac{1}{2}(2I_{1Y}I_{2X} - 2I_{1X}I_{2Y}) \end{aligned} \tag{2.18}$$

# Chapter 3

# Multivariate Analysis

In metabolomics experiments, we measure a large number of variables in the order of hundreds or even thousands to find the metabolites. So here we use data analysis techniques such as Chemometric analysis (used in metabolomics), Where the measurements can be arranged in complex data matrix or tables. where each row constitutes an observation, and the columns represent the variables measured. Multivariate Analysis (MVA) is extensively used in spectrometry techniques like NMR and Mass Spectrometry (MS) to extract the information by modelling these complex data which is not possible using univariate methods. Based on scientific study chemometric analysis can be performed using either supervised or unsupervised analysis. Both methods are briefly described in following sections:

## 3.1 Principal Component Analysis (PCA)

PCA is an unsupervised chemometric method which simplifies the complexity of higher dimensional data by transforming it to lower dimensions giving summaries of features. It retains the variation in data as much as possible. PCA projects the original variables or data to lower dimensions called principal components (PCs), with the aim of finding the best summary of the variables using a limited number of PCs. The first principal component (PC1) is the direction of maximum variance and the second component (PC2) orthogonal to PC1 explains the next maximum amount of variance possible, not accounted by PC1. In PCA the input data is a matrix (A) of multivariate data, with items(n) in rows and variates(p) in columns it summarizes

the many variates (p) into smaller set of variables (k).

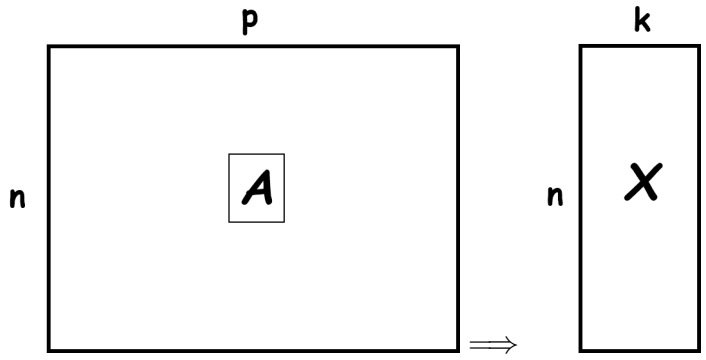


Figure 3.1: Original matrix A reduced to lower dimensional new matrix X

The decomposition of the (A) matrix by PCA can be represent by Score (T), loading (P) and residuals (E) matrices describing all the variation in (A) as:

$$A = TP^T + E \tag{3.1}$$

The graphical representation of the PCA score plot gives information about the clustering and presence of any outlier which can be eliminated before further analysis. In score plot, the observations which are close to each other shows the similar multivariate profile, while other observations which are far from each other shows dissimilar profile an illustrative example for the same is given below.

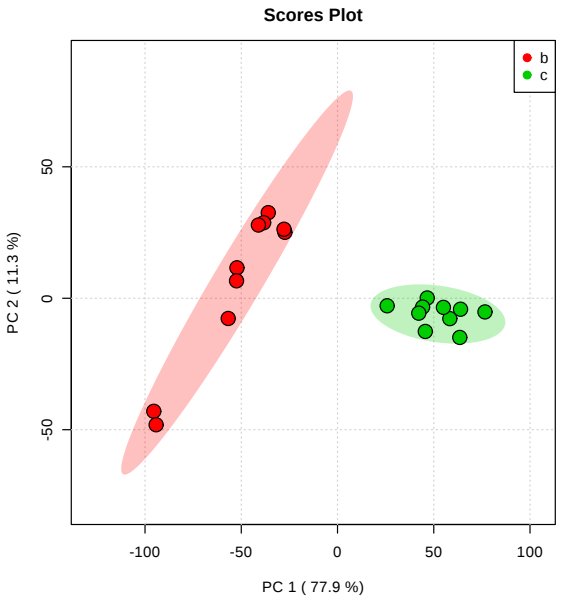


Figure 3.2: PCA score plot between two groups b and c showing separation

## 3.2 Partial least squares (PLS) Discriminant Analysis

PLS regression method is a supervised learning method. This method is used to evaluate the relationship between two matrices A and B. where matrix A contains the spectral intensity values and matrix B contains quantitative values (dependent variables). PLS method can also be used for class discrimination (when the matrix B contains qualitative information) known as PLS-DA which is one of the most useful methods used in a chemometric analysis.

The decomposition of matrix A and B using PLS model is given by:

$$A = TP^T + E \tag{3.2}$$

$$B = UQ^T + F \tag{3.3}$$

where T and U are projection matrices of A and B respectively, P and Q represents the loadings and residual matrices are given by E and F. PLS method maximizes the variance between the variables A and B and it gets negatively affected by the systematic variation in the two matrices.

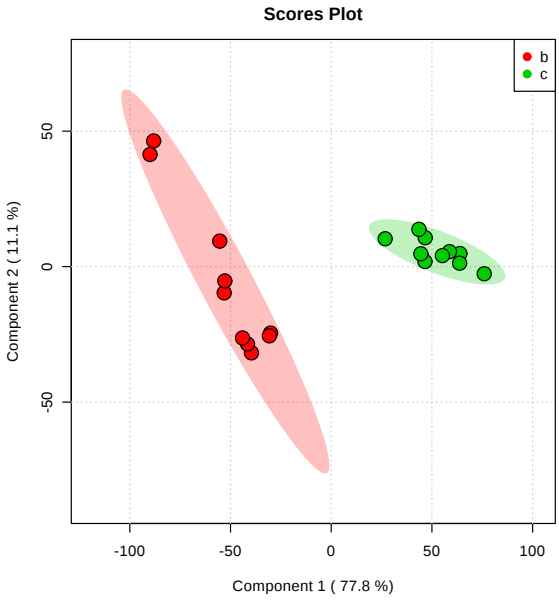


Figure 3.3: PLS-DA score plot between two groups b and c showing separation

## 3.3 Orthogonal Projections to Latent Structures (OPLS) Discriminant Analysis

OPLS-DA method is a better or modified version of PLS model. The main objective of OPLS method is to seperate variation in A into two components, in which one is linearly related to B and other is orthogonal or unrelated to B. The two components are given by; B-predictive ($T_P P_P^T$) and B-orthogonal ($T_0 P_0^T$). only B-predictive variation is used for modeling of B:

$$A = T_P P_P^T + T_0 P_0^T + E \tag{3.4}$$

$$B = UQ^T + F \tag{3.5}$$

Compared to PLS method OPLS has better visualization and interpretation in group separation and seems to be a better version of PLS-DA method.

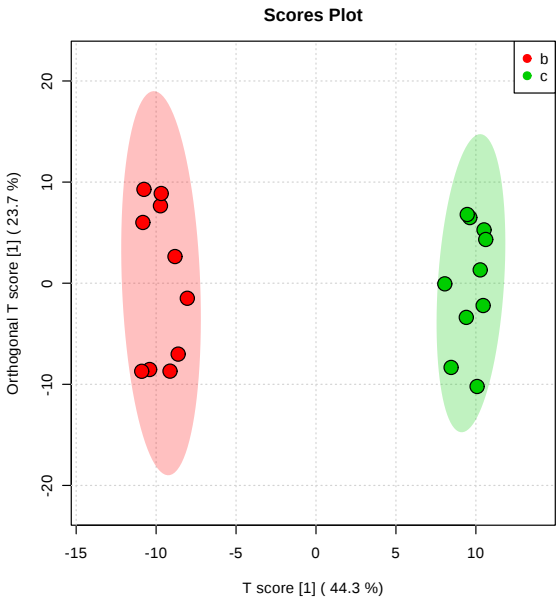


Figure 3.4: OPLS-DA score plot between two groups b and c showing separation

# Chapter 4

# NMR based Metabolomic profiling of *Momordica Charantia L.*

## 4.1 NMR based Metabolomics

Nuclear Magnetic Resonance (**NMR**) spectroscopy, next to Mass Spectrometry (**MS**), is one of the main metabolomics analytical techniques. Developments in NMR spectroscopy have enabled the identification and quantitative measurement of the many metabolites in a single sample of biofluids in a non-targeted and non-destructive manner i.e, a sample can be used again [4]. Main advantages of using NMR spectroscopy includes minimal sample preparation, it allows simultaneous detection of metabolites, it can be utilized *in vivo*.

Also in an NMR spectrum, the signals are proportional to their molar concentrations which enable a direct comparison of concentrations of all possible compounds, without the need for calibration curves of individual compound. or in other words, it shows the real molar levels of metabolites present in the sample. In addition to that NMR is very useful in structure eludication [5]. The basic procedure used in NMR based metabolomic study is given as:

(1) **Sample collection and Preparation**.

(2) **NMR Data Acquisition**.

(3) **Data Pre-processing and Pre-treatment**.

(4) **Data Analysis**.

(5) **Biomarker Identification**.

### 4.1.1 Metabolomic analysis of plants using NMR spectroscopy

NMR based metabolomics has many applications in the field of plant sciences and Natural products. The main objective of metabolomics is to provide a list of all metabolites present in the sample both quantitatively and qualitatively, which provide a clear idea about the sample (for example plant leaves, fruit or stem) under certain geographical Conditions. Because of a very large population of metabolites present in the plant sample it is very difficult to analysis each and every metabolite in a single experiment, here NMR comes as a very suitable method to carry out such study because it allows the simultaneous detection of diverse secondary metabolites groups along with primary metabolites [5].

Metabolic profiling of plant tissue extracts applied to different plant research area such as pharmacological studies of medicinal plants and in new drug discoveries has been reviewed in many recent articles. One such phytomedicinal plant of significance is *Momordica Charantia* or bitter melon, which has been reported to possess potential hypoglycemic activity (reducing blood sugar level), antioxidant, anticancer, anti inflammatory potency indicating a broad range of bioactive compounds [9]. Many studies have been conducted on the effectiveness of dried bitter melon fruit in diabetic animals and in type 2 diabetic human subjects [12].

The aim of this thesis is to identify the primary and secondary metabolites present in different parts of *Momordica Charantia* fruit i,e. skin, flesh, and seed. we are able to identify different bioactive metabolites present in skin, flesh and seeds of bitter melon using NMR spectroscopy with statistical analysis and found phytomedicinal metabolites in different (MC) fruit parts.

## 4.2 Sample Preparation

Organic M. charantia fruits were bought from the market. Ten replicates were analyzed for each sample type: skin, flesh and seeds. All three Samples were prepared by a similar method, by washing them with distilled water to remove all debris or dust and storing at $-80^0$ C for two consecutive days before experiments, followed by freeze-drying and grinding into fine powder and stored at $-20^0$ C prior to NMR experiments. NMR samples were prepared by taking 25 mg of dried skin, flesh and

seed powder per sample. Each sample was mixed with extraction solvent containing 420 $\mu$ L of methanol d4 (making up 70 % of the extraction solvent) and 0.95 mg of $Na_2HPO_4$ (30 mM) + 1 mg of TMSP in 180 $\mu$ L of $D_2$ O (making up 30 % of the extraction solvent). Each sample was mixed thoroughly with 600$\mu$L of extraction solvent and then centrifuged at 5000 rpm for 5 min and 500$\mu$L of supernatant was filtered out and used to prepare the NMR sample [6].

## 4.3 1D and 2D NMR Experiments

Metabolite fingerprinting of all three samples skin, flesh, and seeds was performed using 1D and 2D NMR spectroscopic methods. NMR spectra were recorded on a Bruker Biospin 600 MHz Avance-III spectrometer equipped with a 5 mm QXI probe. methanol-d4 was used as an internal lock before signal acquisition. then 1D proton spectra were acquired using a water suppressed Car–Purcell–Meiboom–Gill (CPMG) spin-echo pulse sequence.

For data acquisition, the 1D spectra were phase-corrected and baseline-corrected and referenced to added sodium-3-trimethylsilylpropionate (TMSP) as a standard resonating at 0.00 ppm. For metabolite identification, two-dimensional NMR spectra were also recorded, including $^1H$–$^1H$ correlation spectroscopy (COSY), total correlation spectroscopy (TOCSY), and $^1H$–$^{13}C$ heteronuclear coherence spectroscopy (HSQC, HMQC).

The Same set of experiments were recorded for all three samples and under similar experimental conditions.

## 4.4 Data Analysis

Metabolite fingerprinting for the bitter melon fruit parts was evaluated by checking identified metabolite peaks with standard NMR metabolite databases available online on Human Metabolome Database (HMDB)(`http://www.hmdb.ca/`) [6].

### 4.4.1 Multivariate statistical analysis

Principal component analysis (PCA) and Orthogonal partial least squares discriminant analysis (OPLS-DA) were performed using both Metaboanalyst software and using SIMCA software while t-test and ANOVA analysis were performed using the Metaboanalyst software. Before statistical analysis, binning of NMR data from -0.2 to 10 ppm was performed in 1D NMR spectra using MestReNova software. Spectral intensities were binned into integrated regions of equal width (0.04 ppm) corresponding to the region of $\delta$ -0.02-10.00 ppm. PCA and OPLSDA with Pareto scaling and mean-centring were performed. The data was first analyzed by PCA, which enabled to detect any outlier located outside the 95% confidence region. Outliers were hence removed prior to OPLS-DA analysis as they can have detrimental effect on data analysis. OPLS-DA S-plot data and t-test(p value<0.01) were used to find the significant metabolites contributing to differences between models.

### 4.4.2 Quantitative Analysis

Ratio method is used to perform the quantification of metabolites, as signal intensity is directly proportional to the molar concentration of metabolites in the $^1H$ NMR spectrum. TMSP was used as an internal standard and quantification of significant metabolites were performed as:

$$m(M) = \frac{m(S) \times I(M) \times MW(M) \times N(S)}{I(S) \times MW(S) \times N(M)} \tag{4.1}$$

where m(M) is the unknown mass of the metabolite, m(S) is the known mass of internal standard i.e. TMSP, I(M) is the total peak integral of the metabolite, I(S) is the peak integral of the standard, MW(M) is the molecular weight of the identified metabolite, MW(S) is the molecular weight of the standard, N(S) is the number of protons of the standard responsible for its peak integral, N(M) is the number of protons of the metabolite responsible for its peak integral. The mass calculated m(M) of the identified secondary metabolites is given by %w/w [6].

### 4.4.3 UPLC-ESI-MS analysis

Ultraperformance liquid chromatography coupled with electrospray ionization tandem mass spectrometry (UPLC-ESI-MS) analysis is consist of UPLC (Waters Acquity

Class I) and Mass spectrometry (Waters Synapt G2-S) system [6]. For separation reverse phase chromatography was used and for mobile phases LC/MS grade with 0.1 micron membrane filtered solvents were used.

For gradient elution, mobile phase consisted of two solvents, A(0.1% formic acid in $H_2O$) and B(0.1% formic acid in $CH_3CN$/MeOH 1:1; v/v) was used. Also, 300 $\mu$L of injection volume was used with flow rate of 0.2 mL/min and Negative ion mode was used with source temperature at $150^0$C, desolvation temperature $350^0$C, conegas flow 61 l/h, cone voltage 40 eV, capillary voltage 2 kV and desolvation gas flow 625 l/h. for recording the spectra ESI in negative mode was used between m/z 50 and 1200. The peaks were identified by comparing their retention times and mass spectrum with those of standard compounds. All observed masses were corrected using the standard mass of Leucine Enkephalon (554.26 Da in negative mode) [6].

# Chapter 5

# Results and Discussion

## 5.1 Metabolite fingerprinting using 1D and 2D $^1$H NMR spectra

A detailed analysis of the metabolites profile of *Momordica Charantia* L. skin, flesh and seeds was performed using both 1D and 2D NMR spectra. Figure 5.1 shows the representative 1D $^1H$ NMR spectrum of *Momordica Charantia* L. Flesh, recorded at 600 MHz. The NMR peak assignments of the Flesh NMR spectrum are summarized in Table 5.1 with a wide range of metabolites, including Lipids, Amino Acids, Carbohydrates, Organic acids and Phenolic compounds. The analysis also suggested the presence of several organic acids including Formic acid, Fumaric acid, Malic acid and Succinic acid. Amino acids present in MC flesh include Aspartic Acid, Threonine, Glutamic Acid, Glutamine, Proline, Glycine, Alanine, Valine, Cysteine, Methionine, Isoleucine, Leucine, Phenylalanine, Lysine, Arginine, and (GABA) in the low to mid frequency region and Tyrosine, Histidine and Tryptophan in the high-frequency region (above 6.0). Lipids were observed at 0.86(t, J = 6.9), 1.22(m), 1.66(m), and 2.31(t, J = 7.52)ppm. The presence of unsaturated fatty acids was suggested by resonances at 2.03(m) and 5.35(m) ppm. Sugars were identified by peaks at 5.40 ppm of sucrose. others sugars include alpha-glucose at 5.16(d, J = 3.8) and beta-glucose at 4.55(d, J = 7.9). Phenolic compounds and flavonoids were also identified in spectra which synthesize secondary metabolites in plants. The presence of all identified metabolites found in 1D $^1H$ spectra were further confirmed by 2D experiments. Figure 5.2-5.6

Shows 1D spectrums of epicatechin, quercitine, luteolin, kaemferol, and chlorogenic acid. Topspin 3.5 pl7 software was used to magnify the spectral regions.

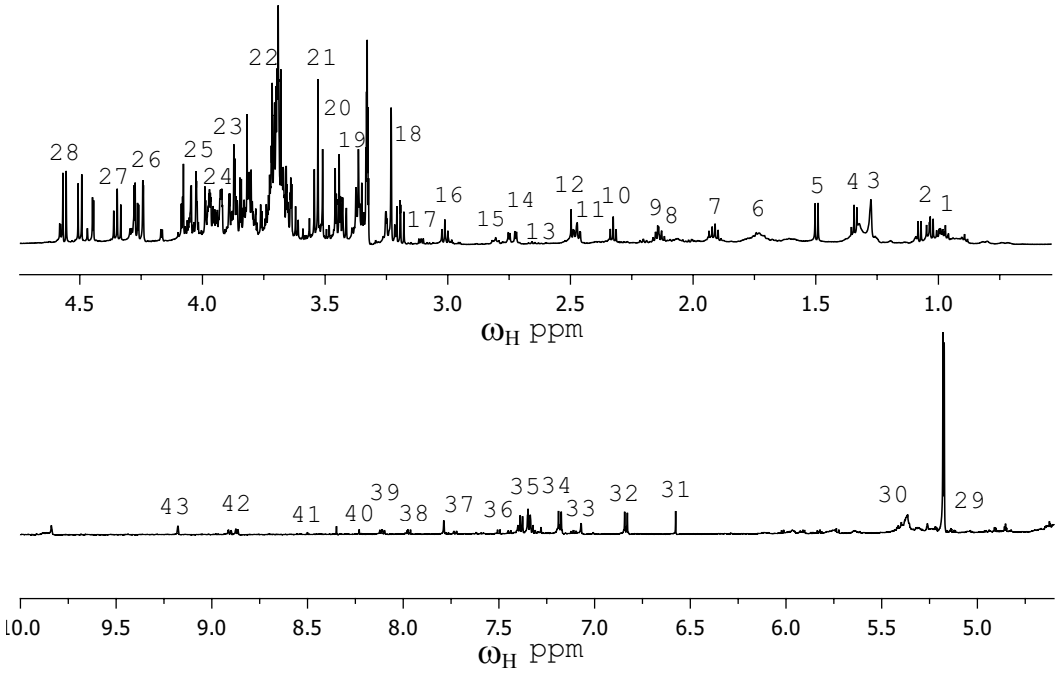


Figure 5.1: 1D $^{1}H$ NMR spectrum of *Momordica Charantia* L. Flesh, recorded at 600 MHz, showing specific resonances of metabolites identified. Peaks numbering: 1, beta-sitosterol (phytosterol); 2, valine; 3, lipid (CH2)n; 4, lactate; 5, alanine; 6, lipid (-CH2-CH2-COOH); 7, leucine; 8, chlorogenic acid; 9,glutamine/glutamic acid; 10, GABA; 11, epicatechin; 12, succinic acid; 13, methionine; 14, malic acid; 15, aspartic acid; 16, lysine; 17, cysteine; 18, choline; 19, inositol; 20, stigmasterol; 21, glycine; 22, arginine; 23, syringic acid; 24, threonine; 25, proline; 26, chlorogenic acid; 27, epicatechin; 28, catechin; 29, alpha-glucose; 30, sucrose; 31, luteolin; 32, gentisic acid; 33, gallic acid; 34, trans-ferulic acid; 35, protocatechuic acid; 36, vanillic acid; 37, histidine; 38, benzoic acid; 39, kaempferol; 40, adenine; 41, formic acid; 42, trigonelline; 43, trigonelline.

Table 5.1 Metabolites present in $^{1}H$ NMR and 2D NMR spectra of bitter melon Flesh, with chemical shift values given in ppm and the corresponding multiplicity and scalar coupling J values (in Hz).

| Metabolites | Chemical Shift (multiplicity,J) |
|---|---|
| Phytosterol | 2.02(m),5.31(m) |
| **Lipids** | |
| Term. Methyl Group | 0.86(t,6.9) |
| $(CH_2)n$ | 1.22(m) |
| $CH_2CH_2COOH$(C3) | 1.66(m) |
| Allylic protons of unsaturated fatty acids | 2.03(m) |
| $CH_2COOH$(C2) | 2.31(t,7.52) |
| Olefinic proton of unsaturated fatty acid | 5.35(m) |
| **Amino Acids** | |
| Aspartic Acid | 2.79(dd, J=17.6, 3.75), 3.88(dd, J=8.5, 3.5) |
| Threonine | 3.825(dd, J=5.74, 3.24) |
| Glutamic Acid | 2.07(m), 2.37(m) |
| Glutamine | 2.15(m),2.48(m) |
| Proline | 4.07(dd, J=8.56, 6.4) |
| Glycine | 3.52(s) |
| Alanine | 1.46(d,7.2) |
| Valine | 1.02(d,7.19) |
| Cysteine | 3.98(dd),3.03(m) |
| Methionine | 2.62(t, J=7.49),2.13(m) |

| | |
|---|---|
| Isoleucine | 1.01(d, J=7.01),0.96(t,J=7.62) |
| Leucine | 1.74(m) |
| Tyrosine | 7.16(m),6.81(m) |
| Phenylalanine | 3.24 |
| Histidine | 3.19(dd,J=14.5,7.9),7.77(d,1.04) |
| Lysine | 3.02(t) |
| Arginine | 3.73(t,J=6.5) |
| Tryptophan | 7.31(t,J=7.5),7.48(d,J=7.76) |
| γ-Amino-Butyrate(GABA) | 2.31(t,J=7.11),3.0(dd,J=8.9,6.3) |
| **Carbohydrates** | |
| Alpha glucose | 5.16(d,J=3.8) |
| Beta glucose | 4.55(d,7.9) |
| Sucrose | 5.40(d,J=3.8) |
| **Organic Acids** | |
| Formic Acid | 8.49(s) |
| Fumaric Acid | 6.56(s) |
| Malic acid | 2.72(dd,J=16,4.10) |
| Succinic Acid | 2.59(s) |
| **Phenolics** | |
| Gallic Acid | 7.06(s) |
| Protocatechuic acid | 7.398(d) |
| Gentisic acid | 6.82(d,8.58) |
| catechin | 4.55(d,7.90) |

| | |
|---|---|
| Vanillic acid | 7.43(dd),3.90(s) |
| Chlorogenic Acid | 7.17(d,1.8),4.25(d,2.58),3.89(dd),2.13(m),2.02(m) |
| Syringic acid | 7.03(s),3.81(s) |
| Epicatechin | 4.28(m),2.47(m),4.28(m) |
| trans-Ferulic acid | 7.10(dd) |
| Benzoic acid | 7.34(dd),7.84(d) |
| trans-Cinnamic acid | 7.77(m),7.34(m) |
| Kaempferol | 8.10(m) |
| **Others** | |
| Adenine | 8.2(s) |
| Choline | 3.22(s) |
| Inositol | 4.01(t,J=2.8),3.6(t,9.7),3.44(dd,J=9.7,2.74),3.22(t,J=9.3) |
| Luteolin | 6.55(s),7.38(m) |
| Quercetin | 6.82(d,8.57) |

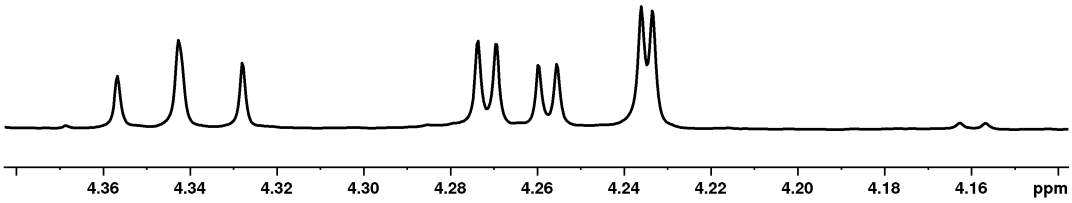


Figure 5.2: 1D $^1H$ NMR spectrum of epicatechin

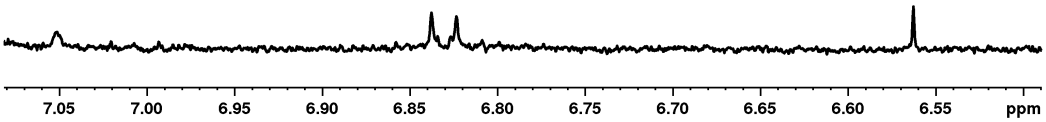


Figure 5.3: 1D $^1H$ NMR spectrum of quercitine

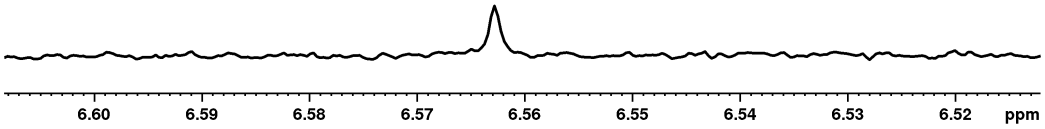


Figure 5.4: 1D $^1H$ NMR spectrum of luteolin

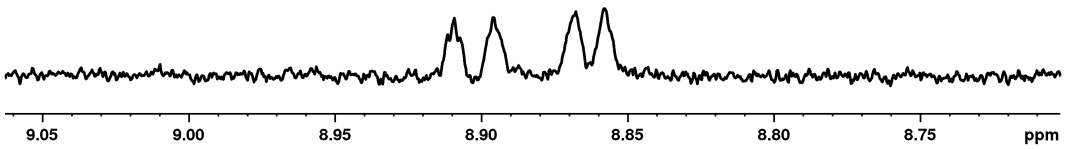


Figure 5.5: 1D $^1H$ NMR spectrum of kaempferol

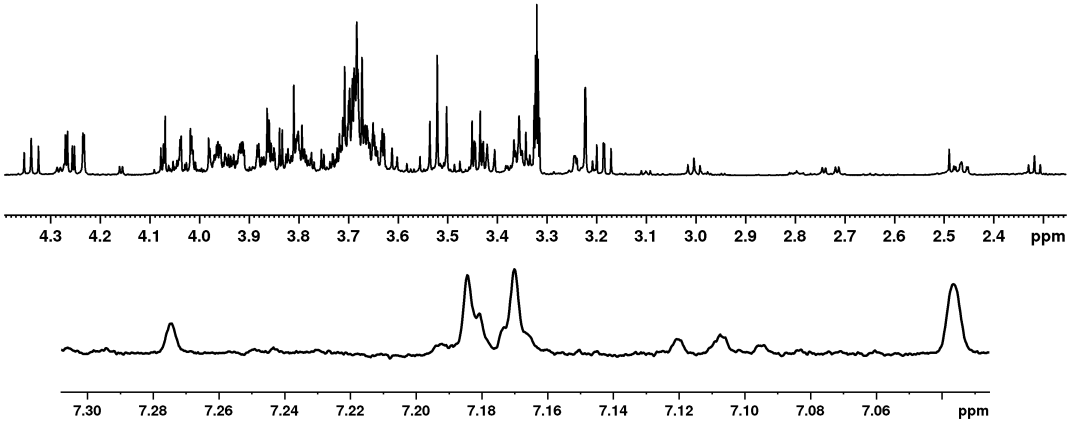


Figure 5.6: 1D $^1H$ NMR spectrum of Chlorogenic Acid

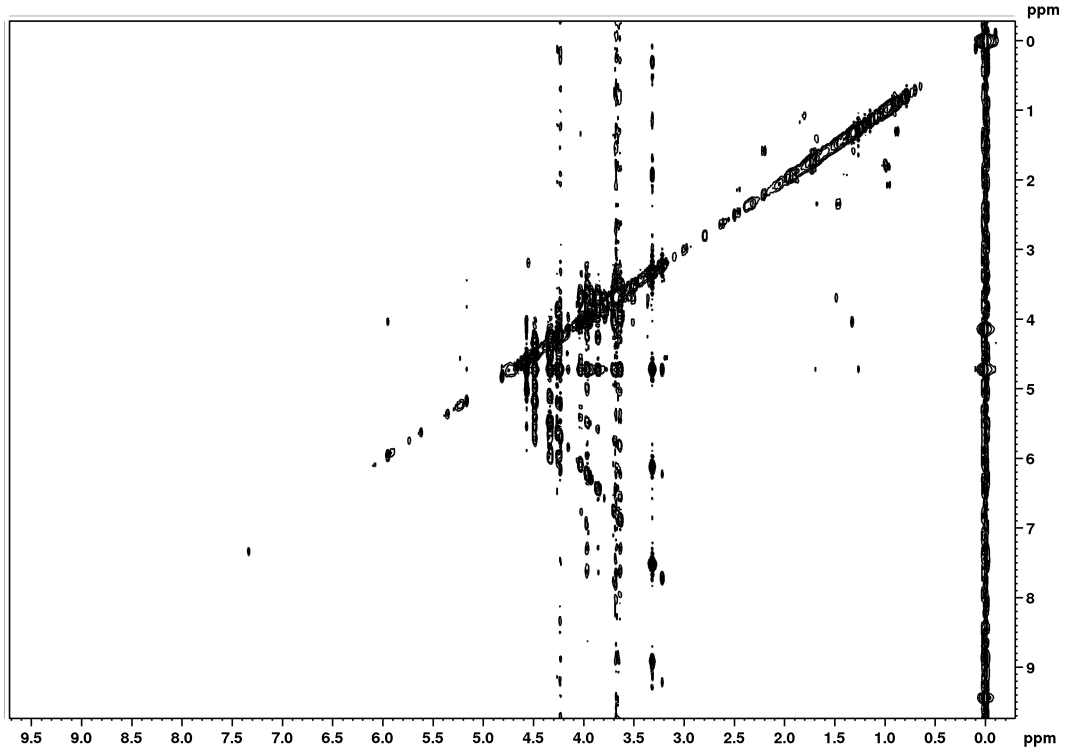


Figure 5.7: 2D COSY NMR spectrum of *M. Charantia* L. skin recorded at 600 MHz

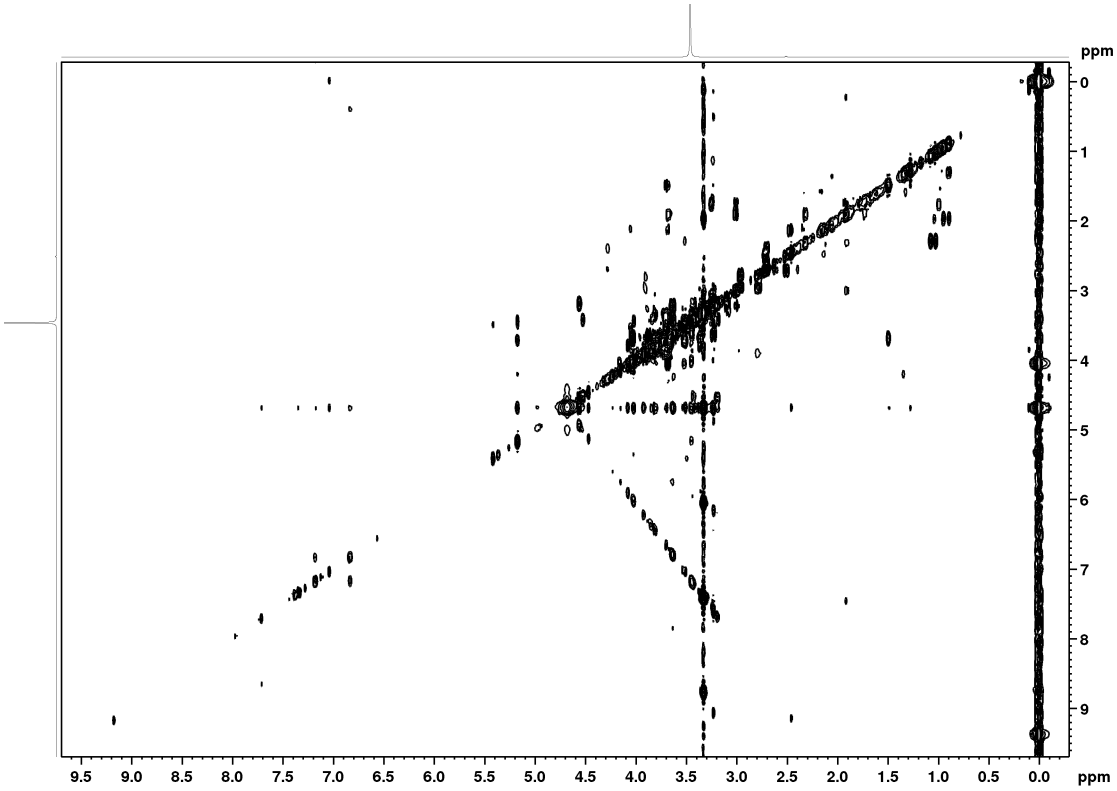


Figure 5.8: 2D COSY NMR spectrum of *M. Charantia* L. seed recorded at 600 MHz

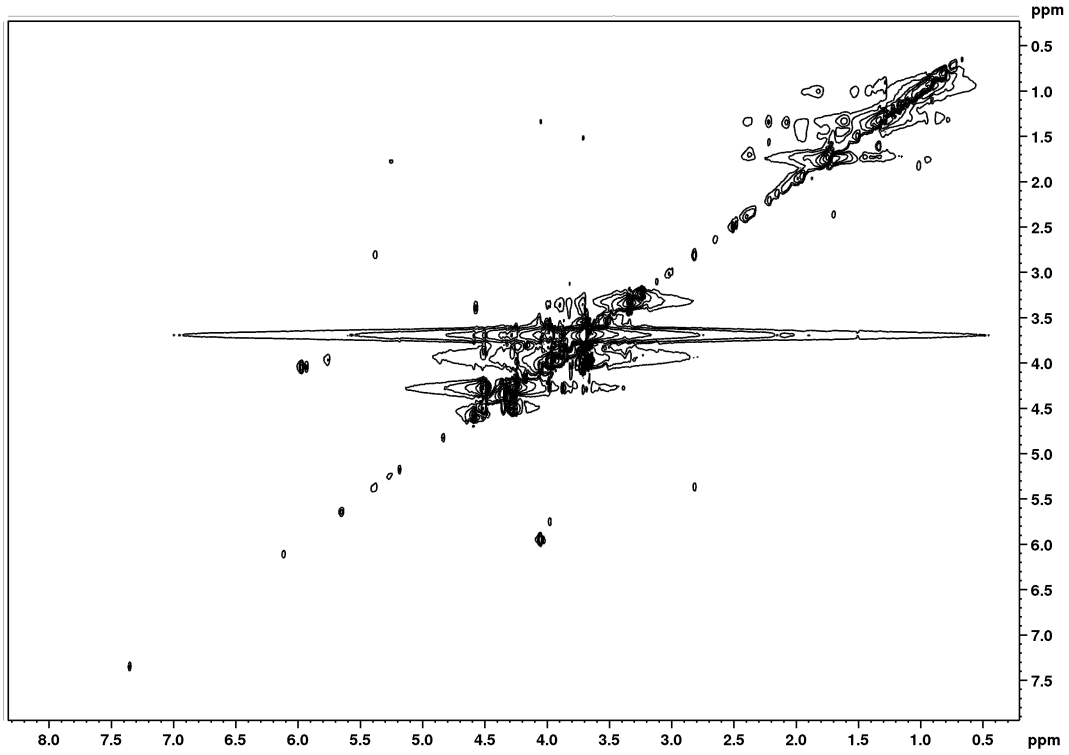


Figure 5.9: 2D TOCSY NMR spectrum of *M. Charantia* L. skin recorded at 600 MHz

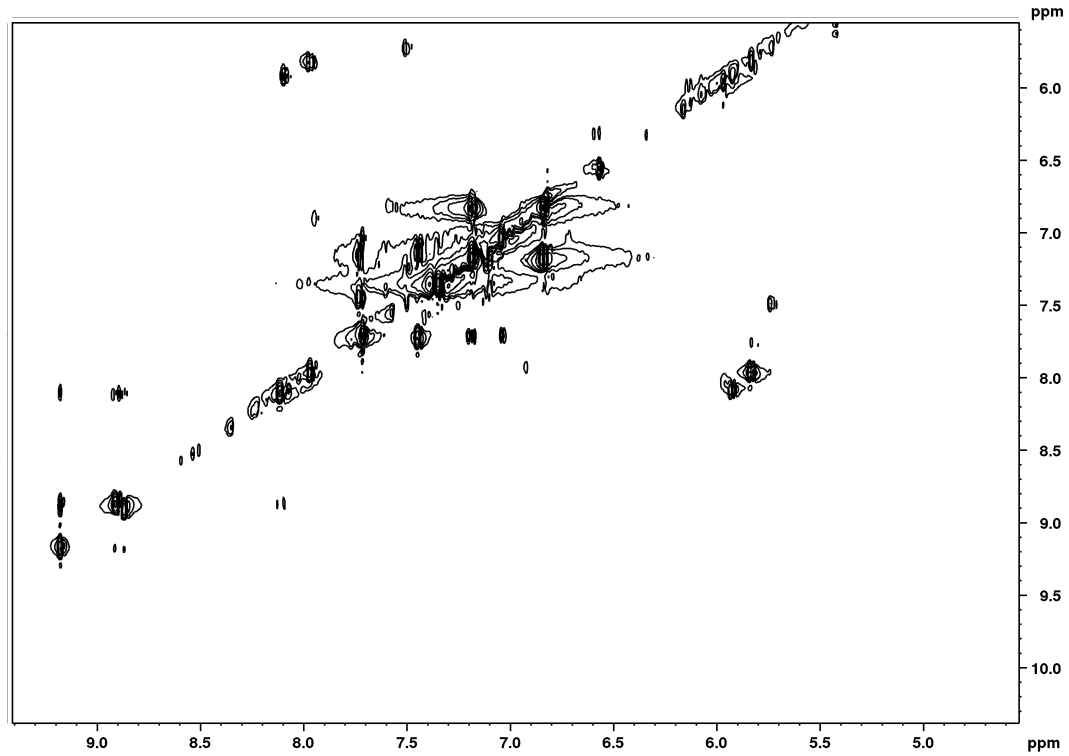


Figure 5.10: 2D TOCSY NMR spectrum of *M. Charantia* L. seed recorded at 600 MHz

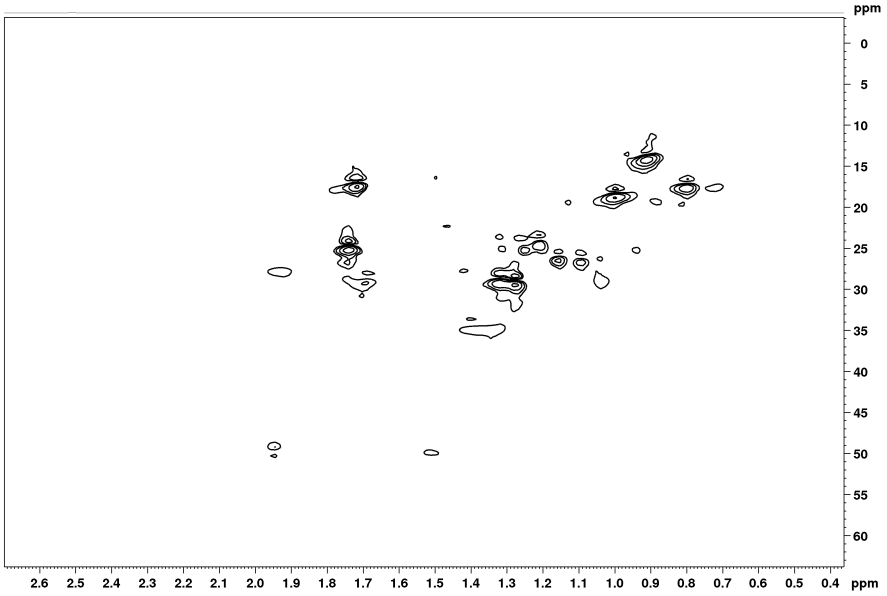


Figure 5.11: 2D HMQC NMR spectrum of M. charantia seed recorded at 600 MHz showing peaks in lipids and amino acid region. Along X axis proton chemical shift ($\omega_H$ ppm) is represented and along Y axis carbon chemical shift ($\omega_C$ ppm) is represented.

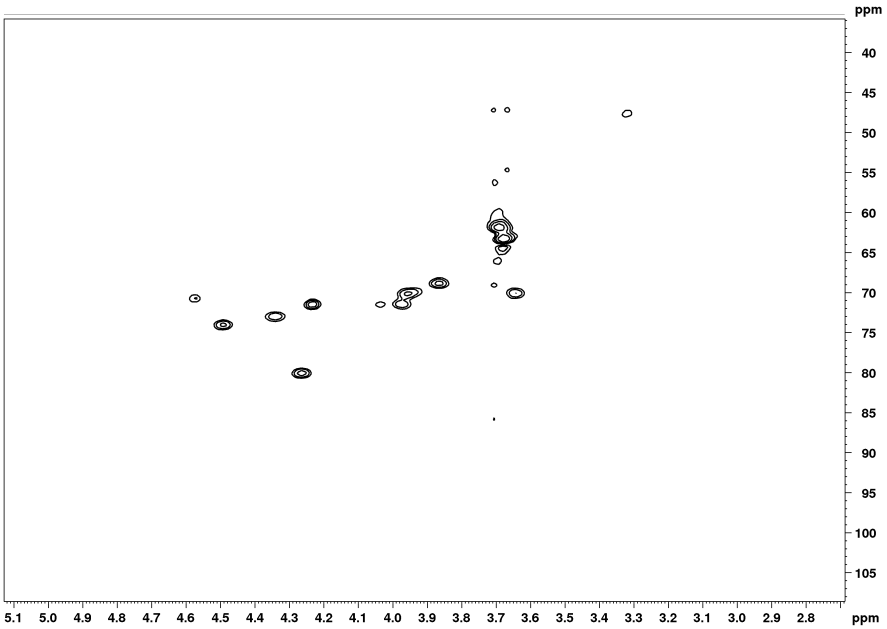


Figure 5.12: 2D HMQC NMR spectrum of M. charantia seed recorded at 600 MHz showing peaks in carbohydrates and lipids region. Along X axis proton chemical shift ($\omega_H$ ppm) is represented and along Y axis carbon chemical shift ($\omega_C$ ppm) is represented.

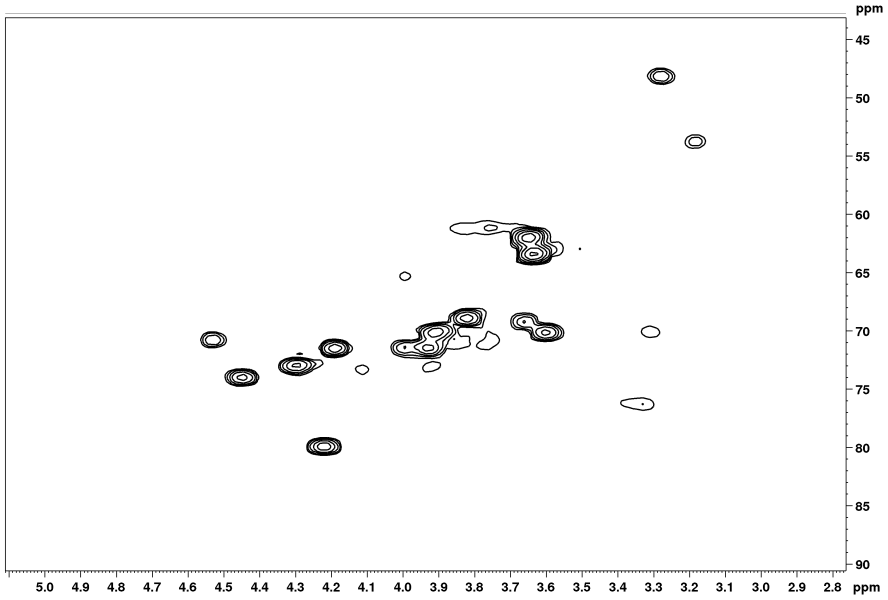


Figure 5.13: 2D HSQC NMR spectrum of M. charantia seed recorded at 600 MHz showing peaks in carbohydrates region. Along X axis proton chemical shift ($\omega_H$ ppm) is represented and along Y axis carbon chemical shift ($\omega_C$ ppm) is represented.

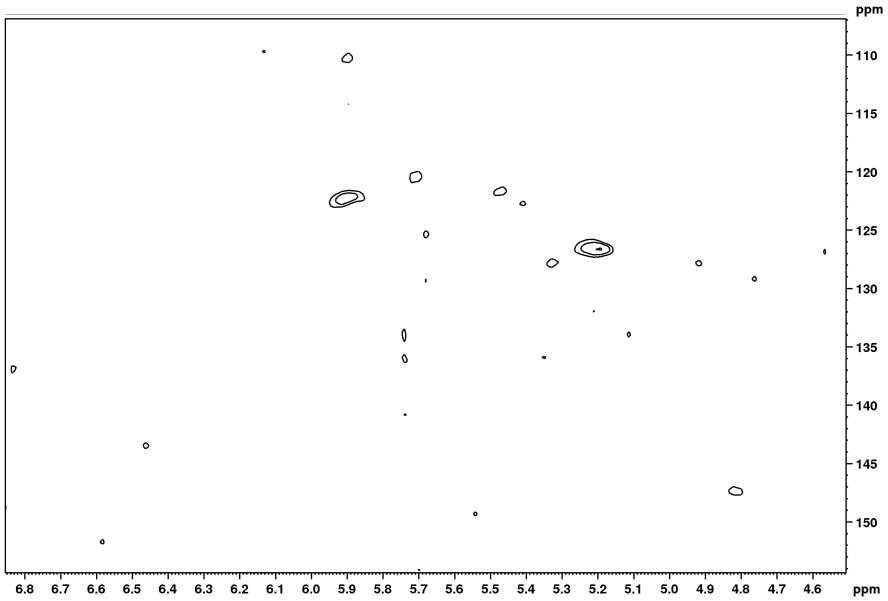


Figure 5.14: 2D HSQC NMR spectrum of M. charantia seed recorded at 600 MHz showing peaks of secondary metabolites. Along X axis proton chemical shift ($\omega_H$ ppm) is represented and along Y axis carbon chemical shift ($\omega_C$ ppm) is represented.

## 5.2 Statistical Analysis

The Multivariate statistical analysis was performed for the chemical classification of Metabolites present in *Momordica Charantia* L. was performed by PCA and OPLS-DA statistical methods. For this, we imported the Data file into SIMCA 14.1, Pareto scaled and first analyzed using the (PCA) to look for possible outliers, groups, similarities and other patterns in the data. Figure 5.15 shows the PCA score plot for skin, flesh, And seeds, with component 1 explaining (70.1)% of the variation and component 2 explaining (16.6)% of the variation, showing a clear separation between these three groups. With $R^2 = 0.939$ and $Q^2 = 0.866$, the PCA model shows good predictability. The data were further analyzed using the method of Orthogonal Projections to Latent Structures (OPLS-DA) discriminant analysis Figure 5.16, which maximizes the covariance between the NMR peak intensities and measured data. Showing inherent difference in groups.

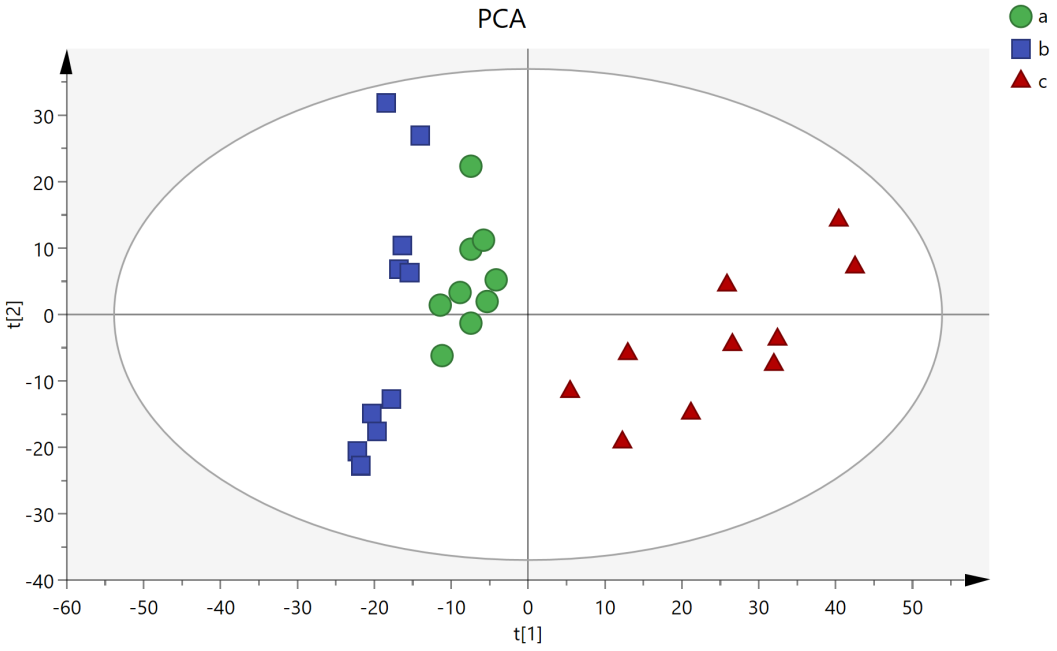


Figure 5.15: Score plots of principal component analysis (PCA) showing separation for Momordica charantia L. by different shapes and colours, Blue Square, seeds; Green circles; flesh and red triangles, skin

.

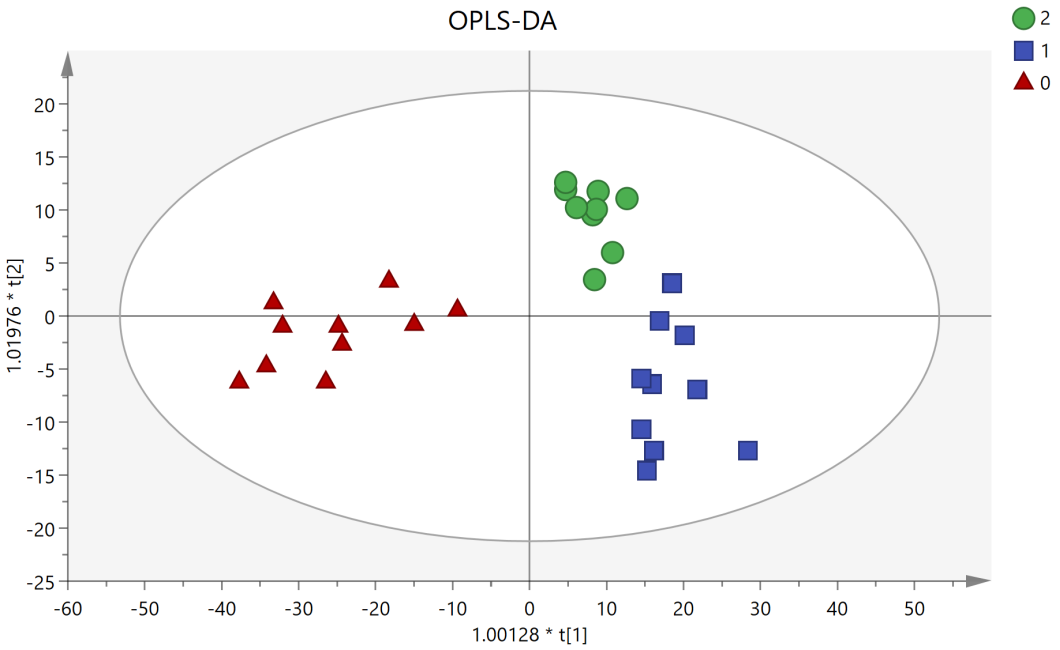


Figure 5.16: Score plots of Orthogonal Projections to Latent Structures discriminant analysis (OPLS-DA) showing separation for *Momordica charantia* L. by different shapes and colours, Blue Square, seeds; Green circles; flesh and red triangles, skin

.

We then performed OPLS-DA analysis for skin and seeds samples individually as shown in the figure 5.17. It can be seen from the plot that there is a clear separation between these two groups. figure 5.18 shows the Dendrogram or tree plot for all the three samples. As can be seen, seeds, and flesh are closely related to each other and separated from the skin. OPLS-DA loading plot was then used to find the significant metabolites responsible for the separation. The significant metabolites were identified from linear S-plot (p(corr)[1] $>$ 0.5) values. This was followed by a t-test with a threshold of p$<$0.01. Table 5.2 shows the list of 22 Significant metabolites which contributing significantly in group separation. Table 5.3 shows the relative concentrations of the metabolites identified which are useful in treating type 2 Diabetes Mellitus to differ significantly in the three groups i.e, skin, flesh and seeds.

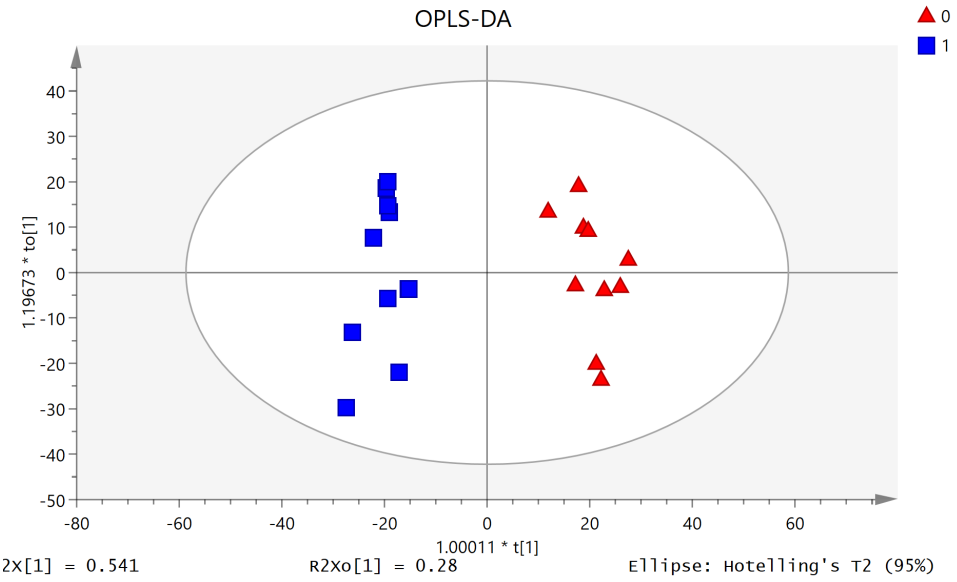


Figure 5.17: (OPLS-DA) plot further showing separation for *Momordica charantia* L. by different shapes and colours, Blue Square representing seeds and Red triangles representing skin

.

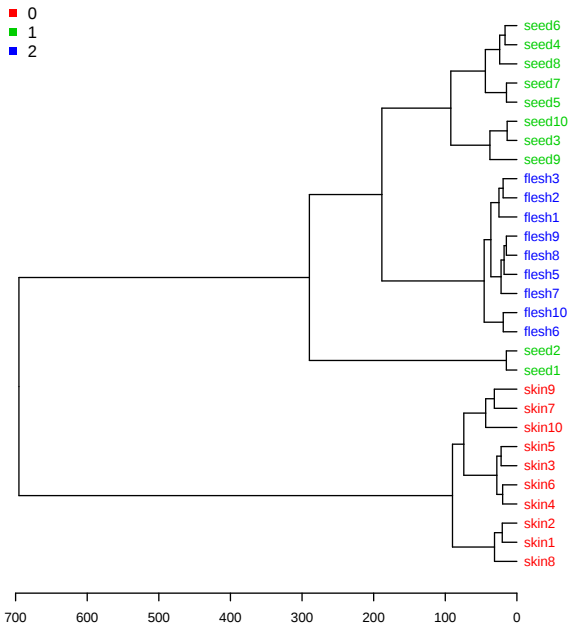


Figure 5.18: Dendogram showing differences between all the three samples(flesh, skin and seeds). Each sample consist of 10 replicates

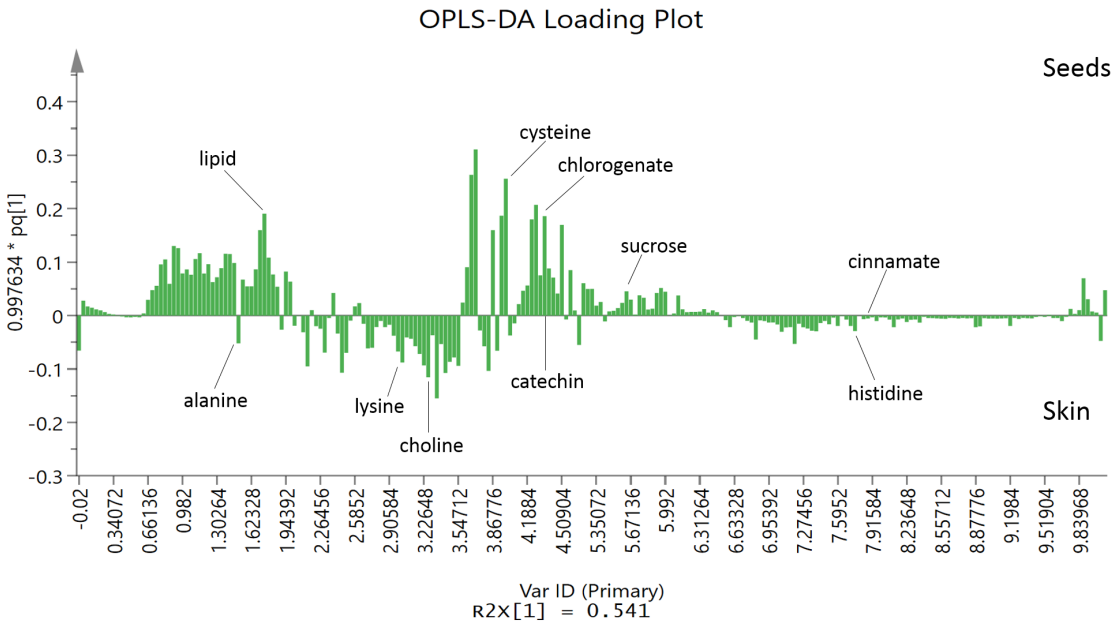


Figure 5.19: Loading plot showing metabolites marked below the baseline present in larger amounts and those marked abobe the baseline being present in lesser amounts in skin as compared to seeds.

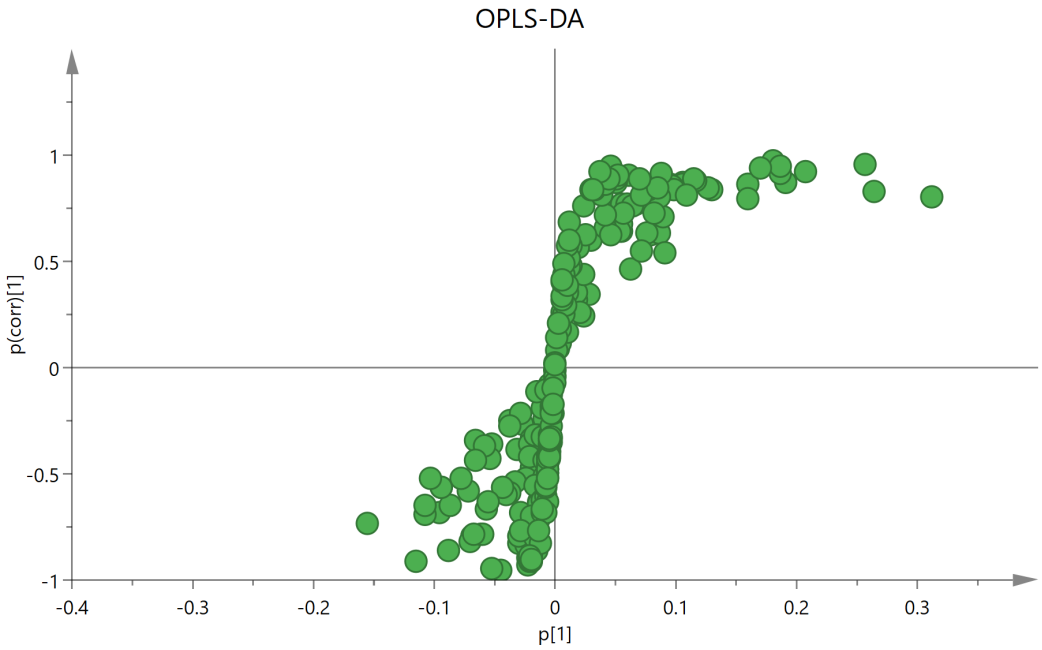


Figure 5.20: S-plot showing metabolites responsible for separation between the two groups.

Table 5.2 List of significant metabolites found in *Momordica Charantia* L. Statistical significance was confirmed by splot pcorr. Values (>0.5) and by t-test with $p<0.01$.

| Metabolites | Chemical Shift (multiplicity,J) |
|---|---|
| Phytosterol | 2.02(m),5.31(m) |
| Term. Methyl Group | 0.86(t,6.9) |
| $(CH_2)n$ | 1.22(m) |
| $CH_2CH_2COOH$(C3) | 1.66(m) |
| Threonine | 3.825(dd, J=5.74, 3.24) |
| Alanine | 1.46(d,7.2) |
| Cysteine | 3.98(dd),3.03(m) |
| Leucine | 1.74(m) |
| Lysine | 3.02(t) |
| Tryptophan | 7.31(t,J=7.5),7.48(d,J=7.76) |
| γ-Amino-Butyrate(GABA) | 2.31(t,J=7.11),3.0(dd,J=8.9,6.3) |
| Gentisic acid | 6.82(d,8.58) |
| Gallic Acid | 7.06(s) |
| Vanillic acid | 7.43(dd),3.90(s) |
| Chlorogenic Acid | 7.17(d,1.8),4.25(d,2.58),3.89(dd),2.13(m),2.02(m) |
| Epicatechin | 4.28(m),2.47(m),4.28(m) |
| trans-Cinnamic acid | 7.77(m),7.34(m) |
| Kaempferol | 8.10(m) |
| Adenine | 8.2(s) |
| Choline | 3.22(s) |
| Luteolin | 6.55(s),7.38(m) |
| Quercetin | 6.82(d,8.57) |

Table 5.3 Relative amounts (% w/w) of medicinally important in treating Type 2 metabolites present in skin, flesh and seeds part of *Momordica Charantia* L. Data are represented as mean ±SD and statistical significance was confirmed by a t-test.

| Metabolite | Peak(in ppm) | Quantity present in Skin | Quantity present in Flesh | Quantity present in Seeds |
|---|---|---|---|---|
| Epicatechin | 4.28 | 0.00 | 0.5495±0.2235 | 0.9166±0.4899 |
| Luteolin | 6.55 | 0.0067±0.0037 | 0.0339±0.0121 | 0.0456±0.0193 |
| Quercetin | 6.82 | 0.0373±0.0141 | 0.1432±0.0349 | 0.2047±0.0470 |
| Chlorogenic acid | 7.17 | 0.0277±0.0104 | 0.1138±0.0274 | 0.1492±0.0289 |
| Kaempferol | 8.10 | 0.00 | 0.0310±0.0066 | 0.0433±0.0100 |

## 5.3 UPLC-ESI-MS analysis

Figure 5.21 and 5.22 shows UPLC based chromatographic separation and ESI-MS spectrums of M. charantia L. skin respectively. In accordance with the previous NMR results that we got, M. charantia L. skin extracts showed the presence of phenolic and lipids region in the mass spectrum.

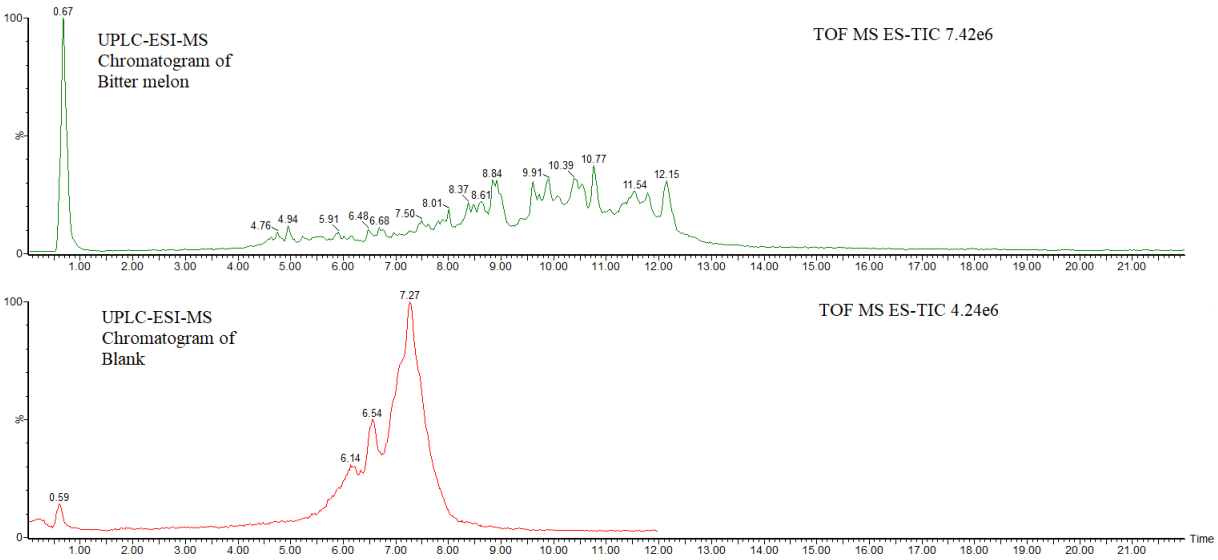


Figure 5.21: UPLC-ESI-MS chromatogram of M. *charantia* skin MeOH extract

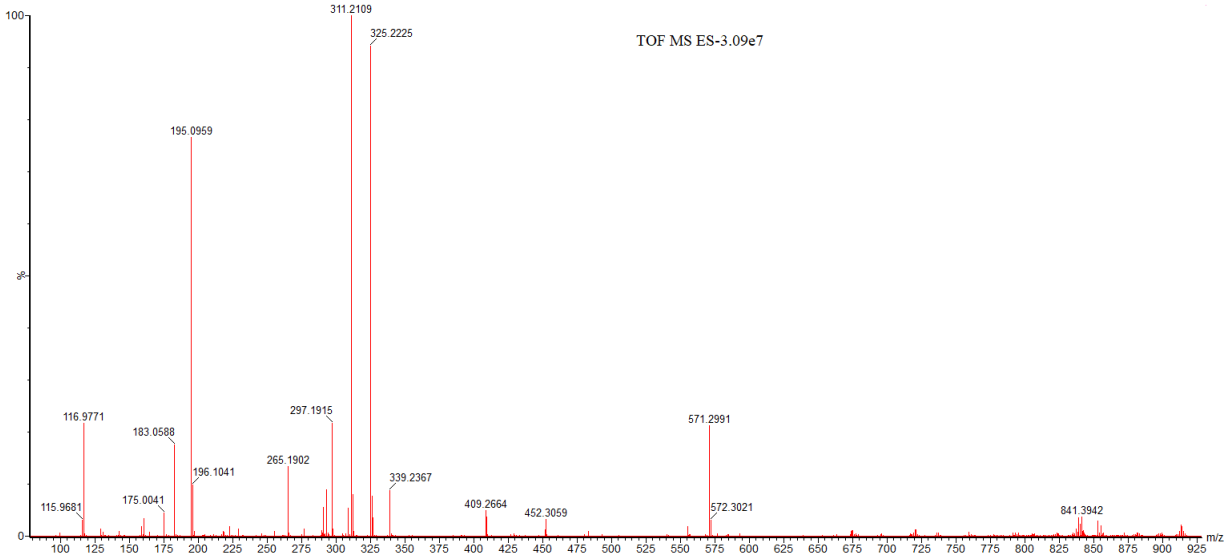


Figure 5.22: MS of M. *charantia* skin MeOH extract with peaks showing regions of phenolics and lipids.

## 5.4 Conclusions

*Momordica Charantia* L. (MC) has been reported with a large number of biologically active metabolites with a wide range of phytomedicinal properties. The presence of secondary metabolites such as flavonoids, phenolics, alkaloids and others confirms the medicinal significance of M. charantia especially, in treating diabetes. Table 5.3 shows the percentage of secondary metabolites (which are helpful in treating diabetes) present in flesh and seeds that contribute to statistically difference between them. As can be seen, Epicatechin is present in highest concentrations in both flesh and seeds but absent in skin, followed by Chlorogenic acid, quercetin, luteolin, and kaempferol. The number of metabolites present was higher in seeds as compared to flesh and least in skin part. We inferred from our analysis that M. Charantia skin does not have enough secondary metabolites (or present in very fewer amounts) of known medicinal importance. chlorogenate is known to have benificial effects on type 2 diabetes and Attenuates hypertension, endothelial dysfunction and vascular dystrophy [6]. Phenolic acids present in (MC) flesh and seeds were mainly gallic acid, gentisic acid, chlorogenic acid, vanillic acid and epicatechin which shows antioxidant activities and these natural phenolics can be used as a good source of antioxidants for application in food system [10]. Charantin composed of stigmasterol glucoside and $\beta$sitosterol glu-

coside found as multiplets at 0.88 ppm and 0.91 ppm respectively is one of the most important compounds present in the (MC) fruit, possess potential hypoglycemic activity and have been used for treating diabetes [12]. Polypeptide-p found in (MC) fruit is reported to be a very effective hypoglycemic agent, whose structure is similar to insulin which is helpful in treating Type 1 diabetes [11]. Several studies show many other properties of (MC) such as anti-inflamatory, anticancer and cholesterol lowering effects [16]. Flavonoids Luteolin, Kampferol and Quercitin present in (MC) fruit parts are known to have antioxidant properties [15].

In recent times many plant researchers are focused on identifying plant phytoconstituents for phytomedicines. The presence of biologically active metabolites in the 1D and 2D NMR spectra and in UPLC-ESI-MS spectra of M. *charantia* L. skin, flesh and seeds, confirms the medicinal significance of (MC) and its efficacy as a anti-diabetic.

# Appendix A

# The t-test

It is a statistical hypothesis test which has a null hypothesis ($H_0$) and follows t-distribution. If the test statistics is normally distributed and has a scaling term then t-test can be applied. Depending on test statistics t-test can be applied in following ways:

## Univariate t-test

In univariate or one sample t-test the null hypothesis is assumed to have the population mean equal to a specified value $\mu$.
The test statitics is given as:

$$t = \frac{\bar{X} - \mu}{s_{\bar{X}}} \tag{A.1}$$

where $\bar{X}$ is the sample mean and $s_{\bar{X}}$ is the spred of the sampling distribution of the mean given by standard error as:

$$s_{\bar{X}} = \frac{s}{\sqrt{n}} \tag{A.2}$$

where s is the sample standard deviation and n is the size of sample with n - 1 degrees of freedom.

## Two sample t-test

A two sample t-test assumes a normal distribution with equal sample size and equal variance. the null hypothesis in this case assumes that the two samples are taken

from population having equal means.

The test statistics in this case can be given as:

$$t = \frac{\bar{X_1} - \bar{X_2}}{s_{X_1X_2}\sqrt{\frac{2}{n}}} \tag{A.3}$$

where $s_{X_1X_2}$ is given by:

$$s_{X_1X_2} = \sqrt{\frac{s_{X_1}^2 + s_{X_2}^2}{2}} \tag{A.4}$$

Here $s_{X_1X_2}$ is the combined standard deviation of the two populations with $n_1 = n_2$ = n and $s_{X_1}^2 + s_{X_2}^2$ are the variances of two samples. the degree of freedom in this test is taken as 2n - 2.

## In the case of equal variance

The t-statistic in this case can be given as:

$$t = \frac{\bar{X_1} - \bar{X_2}}{s_{X_1X_2}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \tag{A.5}$$

where $s_{X_1X_2}$ is given as:

$$s_{X_1X_2} = \sqrt{\frac{(n_1 - 1)s_{X_1}^2 + (n_2 - 1)s_{X_2}^2}{n_1 + n_2 - 2}} \tag{A.6}$$

# Appendix B

# "p-Value"

The "p-Value" or the probability value is a measure of strength for a statistical model which helps in determining the significance of statistical test results.
In other words the formal defination of p-value can be given as:

The p-value is the probability of getting the observed value of the test statistic, or a value with even greater evidence against the null hypothesis ($H_0$), if the $H_0$ is true.

p-Value ranges between 0-1. and If the P-value $\leq \alpha$, then the null hypothesis ($H_0$) is rejected in favor of the alternative hypothesis ($H_A$) or ($H_1$). And, if the P-value is $\geq \alpha$, then the $H_0$ is not rejected.
Majority of authors refer to statistically significant as $P < 0.05$ or $P < 0.01$ and statistically highly significant as $P < 0.001$. Here $P < 0.001$ means less than one in a thousand chance of being wrong.

# Bibliography


[1] Keeler, J. *Understanding NMR Spectroscopy*; Wiley, 2002.

[2] Horgan R P and Kenny L C. SAC review 'Omic' technologies: genomics, transcriptomics, proteomics and metabolomics. *The Obstetrician & Gynaecologist*. 2011; 13: 189–195.

[3] B. D. Dettmer, K.; Hammock, "Metabolomics A New Exciting Field within the ' omics ' Sciences. *Environ. Health Perspect*, vol. 112, no. 7, pp. A396–397, 2004.

[4] Agnieszka Smolinska, Lionel Blanchet, Lutgarde M.C. Buydens, Sybren S. Wijmenga NMR and pattern recognition methods in metabolomics: *From data acquisition to biomarker discovery: A review*: Analytica Chimica Acta 750 (2012) 82-97.

[5] Hye Kyong Kim, Young Hae Choi, Robert Verpoorte, NMR-based metabolomic analysis of plants. *Nature Protocols*. 2010; 55: 36-549.

[6] N. Gogna, N. Hamid and K. Dorai, Metabolomic profiling of the phytomedicinal constituents of Carica papaya L. leaves and seeds by $^{1}$H NMR spectroscopy and multivariate statistical analysis. *Journal of Pharmaceutical and Biomedical Analysis*; 115 (2015) 74–85.

[7] N. Gogna, V. J. Singh, V. Sheeba and K. Dorai, NMR-based investigation of the Drosophila melanogaster metabolome under the influence of daily cycles of light and temperature. *Mol. BioSyst*; 2015; 11, 3305.

[8] N. Gogna, M. Krishna, A. M. Oommen and K. Dorai, Investigating correlations in the altered metabolic profiles of obese and diabetic subjects in a South Indian

Asian population using an NMR-based metabolomic approach. *Mol. BioSyst*; 2015; 11, 595.

[9] Sing P. Tan, Tuyen C. Kha, Sophie E. Parks & Paul D. Roach, Bitter melon (Momordica charantia L.) bioactive composition and health benefits: *A review, Food Reviews International*. (2016); 32:2, 181-202.

[10] R. Horax, N. Hettiarachchy, and S. Islam, Total Phenolic Contents and Phenolic Acid Constituents in 4 Varieties of Bitter Melons (Momordica charantia) and Antioxidant Activities of their Extracts: *JOURNAL OF FOOD SCIENCE*. 2005; Vol. 70, Nr. 4.

[11] Mahmoud Bahmani, Hannaneh Golshahi, Kourosh Saki, Mahmoud Rafieian-Kopaei, Bahram Delfan, Tahereh Mohammadi, Medicinal plants and secondary metabolites for diabetes mellitus control: *Asian Pacific Journal of Tropical Disease*. 2014; 4(Suppl 2): S687-S692.

[12] S. Desai, P. Tatke, Charantin: An important lead compound from Momordica charantia for the treatment of diabetes: *JPP*. 2015; 3(6): 163-166.

[13] P. Daniel, U. Supe, M.G.Roymon, A review on Phytochemical analysis of Momordica charantia: *IJAPBC*. 2014; Vol. 3(1), ISSN: 2277 - 4688.

[14] D. Sathish Kumar, K. Vamshi Sharathnath, P. Yogeswaran, A. Harani, K. Sudhakar, P. Sudha, David Banji, A Medicinal Potency of Momordica Charantia: *International Journal of Pharmaceutical Sciences Review and Research*. 2010; ISSN 0976 – 044X.

[15] M. Agarwal and R. Kamal, Studies on flavonoid production using in vitro cultures of Momordica Charantia L.: *IJB*. 2007; Vol 6, pp 277-279.

[16] Baby Joseph, D Jini, Antidiabetic effects of Momordica charantia (bitter melon) and its medicinal potency: *Asian Pac J Trop Dis*. 2013; 3(2): 93-102.

[17] T.B. NG, C.M. WONG, W.W. LI and H.W. YEUNG, INSULIN LIKE MOLECULES IN MOMORDICA CHARANTIA SEEDS: *Journal of Ethnopharmacology*. 1986; 15: 107-117.

[18] Xia, J. and Wishart, D.S, Using MetaboAnalyst 3.0 for Comprehensive Metabolomics Data Analysis: *Current Protocols in Bioinformatics*. (2016); 55:14.10.1-14.10.91.

[19] SIMCA-14.1 for Using Multivariate Analysis.