Investigation of archived accurate mass data in mass spectrometry.

Alamri, Mesfer Mohamed

Swansea University
Prifysgol Abertawe

INVESTIGATION OF ARCHIVED ACCURATE MASS DATA IN MASS SPECTROMETRY

MESFER MOHAMED ALAMRI

A thesis submitted in fulfilment of the requirements
for the degree of M.Phil.

2010

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 – 1346

SUMMERY

This thesis presents an investigation of mass measurement accuracy in mass spectrometry using archived accurate mass measurement data obtained on different types of mass spectrometry instruments. A magnetic sector was used as the "gold standard" analyzer and other data obtained on electrospray ionization fitted with Orbitrap and a MALDI time-of-flight instrument. More than 4500 accurate masses were obtained from the EPSRC National Mass Spectrometry Service Centre at Swansea University using the above instrument types. Different ionization methods are available; each has its own characteristics. In this study the ionization techniques used are chemical ionization (CI), electron ionization (EI), electrospray ionization (ESI), fast-atom bombardment (FAB) and matrix-assisted laser desorption/ionization (MALDI). These mass analyzers have different attributes, including the mass range, resolution and mass accuracy. Mass measurement accuracies (MMAs) are reported for a number of preferred techniques. Statistical calculations were made using various software pack-ages such as SPSS, AccMass (house software), Origin, Easyfit and Excel add-ins. The data contains view outliers which can significantly affect the accuracy and precision of the data as well as affecting the underlying statistical distribution predicted.

DECLARATION

This work has not previously been accepted in substance for any degree and is not being concurrently submitted in candidature for any degree.

Signed..... (Candidate)

Date.....

STATEMENT 2

This thesis is the result of my investigations of archived accurate mass spectrometry, the data obtained from mass spectrometry unit at Swansea University as suggested and recommended by my supervisor. Other sources are acknowledged by footnotes giving explicit references. A bibliography is appended.

Signed..... (Candidate)

Date.....

STATEMENT 3

I hereby give consent for my thesis, if accepted, to be available for photocopying and for inter-library loans after expiry of a bar on access approved by Swansea University.

Signed..... (Candidate)

Date.....

Acknowledgment

I would like to thank Prof. Gareth Brenton for his supervision and guidance throughout my study for this programme. I would also like to show my appreciation for the help of Dr. Christopher Williams, Dr. Ruth Godfery, Dr. Mark Wyatt and all staff at EPSRC centre and for their answering my questions about the data throughout my study.

I would like to acknowledge the support received from Directorate of General Security forces and the General Directorate of Criminal Evidence for providing a scholarship. Special appreciation to Colonel, Mohamed Sultan for his assistance and advices.

Finally, I would like to thank my family for their support and offer special appreciation to my wife for her care and provision of a comfortable place for me during the period of my study

Table of Figures

Chapter 1

Table of Figures

Chapter 1

Chapter 2

Table of Figures

Table of Figures

Chapter 3

List of Tables

Chapter 1

Chapter 2

List of Tables

Chapter 3

List of Tables

$[M+H]^+$	*Protonated molecule*
aaMMA	*Average absolute mass measurement accuracy*
AME	*Average mass error*
AP	*Atmospheric pressure*
APCI	*Atmospheric pressure chemical ionisation*
Bzv	*Magnetic sector force*
Chi^2	*Chi-squared*
CI	*Chemical ionisation*
CID	*Collision-induced dissociation*
D/I	*Desorption / ionisation*
Da	*Dalton*
EI	*Electron ionization*
ESI	*Electrospray ionisation*
FAB	*Fast-atom bombardment*
FT	*Fourier transform*
FT-ICR	*Fourier transform-ion cyclotron resonance*
GC	*Gas chromatography*
ISTD	*Internal standard*
K-S	*Kolmogorov-Smirnov*
LC	*High-performance liquid chromatography*
LOD	*Limit of detection*
LSIMS	*Liquid secondary ion mass spectrometry*
m/z	*Mass-to-charge ratio*

MAD	*The median of the absolute deviation of median*
MALDI	*Matrix-assisted laser desorption /ionisation*
mDa	*Milli Dalton*
MMA	*Mass measurement accuracy*
MMAs	*Mass measurement accuracies*
MS	*Mass spectrometry*
MS/MS	*Tandem mass spectrometry*
oa-TOF	*Orthogonal acceleration time-of-flight*
PFK	*Perfluorokerosene*
Q	*Dixon's test*
QIT	*Quadrupole ion trap*
Q-Q plot	*Quantile Quantile*
QTOF	*Quadrupole time-of-flight mass spectrometer*
R	*Rosner's test*
RF	*Radio frequency*
RMS	*Root mean square*
R_s	*Resolution*
SD	*Standard deviation*
SPSS	*Statistical Packages for the Social Science*
T	*Grubbs' test*
TOF-MS	*Time-of-flight mass spectrometry*

1. MASS SPECTROSCOPY AND MASS MEASUREMENT

1.1 Introduction

Mass spectrometry is a microanalytical technique that can be used to detect and determine the amount of a given analysis or to determine the elemental composition and some aspects of the molecular structure of analyte. Unique features of mass spectrometry include its capability to determine the nominal mass and the fragments of an analyte. It also has the capacity to generate considerable structural information for small unit quantities of analyte. Mass spectrometry does not determine the mass directly, it determines the mass-to-charge ratio (*m/z*) of ions. The fundamental requirement of mass spectrometry is that the ions are in the gas phase before they can be separated according to their *m/z* value through the chosen mass analyzer.

Historically, mass spectrometric techniques were developed from the late 19th Century and early part of the 20th Century as a result of the efforts and success of pioneers in this field (Francis William Aston, John Joseph Thomson, Heinvich Elizabeth Mattauch and Jeffery Dempster). One of the major contributions, not only in this field, but generally to chemistry and physics during the period 1900 – 1960 was the accurate determination of the masses of the elements and the identification of the isotopes of

the elements[1]. John Joseph Thomson (British physicist) was a pioneer of the methods that resolved the isotopes of many elements. He observed a release of positive ions and neutral atoms from a solid surface induced by ion bombardment[2], whereas Dempster and some other pioneers contributed to advances in instrument design and discovered many more of the isotopes of the naturally occurring elements. Recently, mass measurement accuracy (MMA) and precision were being recognized as vital for the determination of elemental formulae of unknown molecules. Determination of the mass of an ion with sufficient accuracy leads to the elemental formula. This theory was developed by pioneers such as John Beynon and John Waldron during the (1950s). Beynon built a single-focusing sector instrument and applied his early work to organic mass spectrometry; he reported the use of accurate mass measurements to determine the elemental composition of ions[3]. In time high accuracy verification of mass at high resolution was achieved on double-focusing magnetic sector instruments[4-6]. Since this time the field of mass spectrometry has grown tremendously, due in a large part to the continued development of more powerful ionization techniques, ion dissociation techniques and mass analyser technologies. Many applications for which accurate mass measurement is required include identification of the components of mixtures in different fields. Before 1970, only volatiles with significant vapour pressure could be analyte using electron ionization (EI) and later by chemical ionization (CI), but the non-volatiles were not amenable to analysis. Fast-atom bombardment (FAB) was the first practical and widely accepted technique for desorption/ionisation (D/I) which required nanomoles of analyte to produce clear mass

spectra. Recently electrospray ionisation (ESI) and matrix–assisted laser desorption/ionization (MALDI) eclipsed FAB because they require only picomoles of analyte for analysis and ESI can be easily coupled to liquid chromatography.

Many reviews highlighted the recent rapid growth in the use of mass spectrometry for the determination of accurate mass for small or medium molecules. Small molecules ($\leq$ 300 Da) give excellent results with high precision and accuracy. Today, most instruments used for accurate mass are capable of achieving precision of 5 ppm or better.

1.1.1 Background

Mass spectrometry is not limited to analyses of organic molecules; it can be used for the detection of any element that can be analysed. The concept of mass spectrometry is to transfer the molecules of the analyte to gas-phase ions, separate them according to their mass-to-charge ratio (*m/z*) using electric field (or magnetic field) in an evacuated volume and counting the number of ions. This occurs with a computer system to control the analysis process and presenting and storing the data. All mass spectrometers use a vacuum system to control the pressure that is required for the analysis process. The vacuum must be well-distributed in the regions of instruments, and must be sufficient enough not to impede the ions, low pressure minimises ion molecule reactions in the flight path. Various analysers can be used as options for the analyst depending on what analysis is undertaken and the results expected from the instrument and what is the aim of the analyst. Modern instruments may involve more

than one analyser as an alternative facility to use in one machine. A description of instrument types is next and the discussion of the accuracy and precision will be discussed in the following chapters using a range of accurate mass-measurement data obtained on different types of instruments and by different ionisation methods.

Accurate mass measurement of small molecules is used to determine elemental formulae by identifying possible candidate formulae, for a given set of elements, over a small range of masses very close to the measured mass. Exact mass measurement at high resolution is an important tool in analytical chemistry to help confirm the structure of novel compounds prepared by the synthetic chemist, and is also used in pharmaceutical industry to confirm the expected empirical formula of a product[7]. When a low molecular weight sample is analysed using relatively low resolution mass spectrometry, it is common to use the integer value masses (H=1, C=12, N=14 and O=16). However, this method is rarely used with high molecular weight (peptide, proteins) as cumulative errors arise and the integer mass becomes less than the actual nominal mass. Mass resolution and mass accuracy are the primary considerations for determining whether an instrument is suitable for a given analysis. Accurate mass of an ion is calculated by summation of the exact mass between the theoretical value of the mass of an ion and the observed mass on a mass spectrometer. Each type of mass analyzers has unique characteristics that will contribute some uncertainty to the accurate mass measurement. Mass analysers can therefore be chosen depending on the purpose of analysis for a molecule, with some changes on specific parameters of the

analyser to reach a higher resolution, *i.e.* sector mass spectrometers adjust resolution by varying the width of slits.

1.1.2 Instrumental methods

Exact mass has been extensively used since the advent of high resolution mass spectrometry to determine the formula of ions by using a number of ionization techniques, such as atmospheric pressure chemical ionization, electron ionization, thermospray ionization, chemical ionization and fast-atom bombardment ionization.

In scanning mass spectrometry, electromagnetic fields separate the ions according to their mass-to-charge ratio (*m/z*) using different field strengths or field frequencies. Mixtures of ions have different mass-to-charge ratio and different abundance. The ions pass through a slit that is used to choose which ion reaches the detector in scanning MS, and the different of *m/z* sent as a signal to be recorded as a function of time, and presented in the format of *m/z* spectrum. The analyser and detector, and sometimes the ionisation source, are kept under high vacuum to give freedom of motion for ions to travel through the instrument with minimal air molecules collisions.

1.1.3 Ionisation methods

Mass spectrometry is an analytical method used to identify chemical substances by ionising the sample components, focusing the ion yield into a beam, then separating them according to their *m/z* ratios. Most instruments have four basic components: i) an ion source which converts the sample into a beam of charged particles, ii) a

mechanism for focusing it into a narrower beam and accelerating the ion beam through a voltage drop, iii) a mass analyzer that separates the beam into its components, and iv) a detector that can observe or collect the separated beams. Very early designs of mass spectrometers, such as thermal ionization mass spectrometers and electron ionization (gas source) mass spectrometers, vary in the type of ion source employed.

A sector mass spectrometer works by using magnetic and electric fields to exert forces on ions under vacuum to separate ions according to their *m/z* ratios. Therefore compounds must be ionised and then analysed by mass analyser mass spectrometry. The process of molecular ionisation occurs easily for gaseous or volatile samples and uses electron ionization or chemical ionisation. Some other samples decompose upon heating, and need other ionization methods. When there is a need to analyse this kind of samples we have to use desorption or desolvation methods.

The most commonly used ionisation methods for the majority of pharmaceutical and biochemical analyses are electrospray ionisation (ESI) and matrix-assisted laser desorption/ ionization (MALDI). Both ionisation methods are suitable for positive and negative ion modes, depending on the proton affinity of the sample. Ionisation is the process in which charged ions are generated from the analyte.

In EI ion mode, one electron is removed from the atom of analyte to form a radical cation $M^{+\bullet}$. When an electron is removed from an atom or molecule during ionization, enough excess energy is supplied to normally ionize the molecules, *i.e* breaking

bonds. Ionisation can be induced by using high energy radiation (sufficient energy) such as that used in electron ionization (EI), chemical ionisation (CI), photoionization or high electric field. The choice of ionisation method depends on the nature of the compound (sample) that we have and the target of analysis.

1.1.4 Sample separation

There are different types of mass analysers which have different features, including mass range, mass accuracy and resolving power. Mass analysers include time-of-flight (TOF), magnetic sector, Quadrupole; Fourier transforms (FT), Orbitrap and quadrupole ion traps, which are generally coupled to MALDI, electrospray ionisations and other atmospheric pressure ionization techniques. MALDI is generally only coupled to TOF, whereas EI and CI are commonly coupled to magnetic sector, quadruple and sometimes to TOF or hybrid analyzer.

Tandem mass spectrometry is performed on hybrid instruments and may include any of the above although ion traps and quadrupole are preferred. After ion production, precursor ions can be dissociated to product ions using energy imparted as radio frequency waveform and/or through collisions with molecules of neutral gas; this process is called collision-induced dissociation (CID), and is widely used in MS/MS experiments[8].

1.1.5 Mass measurement

Accurate mass measurements are a means to determine the chemical composition of complex organic mixtures compounds such as protein, enzymes and petroleum crude oil which contain many molecules with diverse elemental compositions C, H, N, O, and S[9]. This determination requires a certain degree of accuracy and precision in order to limit the number of possible elemental formulae. A mass spectrum can be annotated with its nominal masses or accurate masses, to an appropriate number of significant figures. The unified (atomic) mass unit (u), or Dalton (Da), is a unit of mass used to express atomic and molecular masses. It is defined as one-twelfth of the mass of an unbound atom of carbon-12 (^{12}C) at rest and in its ground state. The unified mass unit is recommended by IUPAC[10]. In this study the Da is used to indicate mass as it is widely recognised by mass spectrometrists. Nominal mass of an ion or molecule, for a given empirical formula is calculated using the mass of the most abundant isotope of each element, rounded to the nearest integer value and is equivalent to the sum of the mass numbers of its constituent atoms[11].

Accurate mass (measured quantity) is obtained by summing up the masses of individual isotopes masses of the atoms in the molecule. It can help to determine the elemental composition of a compound, but it does not give information about the structure. Historically, accurate mass measurement has been carried out using magnetic sector mass spectrometers, which tend to be large, expensive and require a highly-trained operator to deliver accurate data. Accurate mass measurement can be difficult to achieve and depends on the application and choice of instrument.

Orthogonal acceleration time-of-flight (oa-TOF) mass spectrometry simplified accurate mass for MS/MS measurement, when it was first introduced in the 1990s[12-14]. It provides accurate mass measurement to within a few part per million (ppm) of the molecule's true mass, also called exact mass. TOF instruments are used in a linear fashion or aided by electrostatic grids and lenses in a device called a reflectron. When using a reflectron the mass resolution is increased without losing sensitivity or the need to increase the size of the flight path. The quadrupole time-of-flight (QTOF) has a high mass accuracy and high resolution as much as 10 times higher than quadrupole instruments[15,16].

The difference between the measured value (accurate mass) and the exact mass is expressed as the accuracy of the accurate mass measurement on a mass spectrometer. When the measured value is close to the true value, this means the degree of precision is very good. TOF analysers can now reach resolutions between 5,000 and 20,000 FWHM. Comparable to the traditional sector instruments, the procedure to measure exact mass differs slightly from that of magnetic sector field mass spectrometers. Usually worse mean mass measurement accuracy is achieved by TOF rather than sectors, e.g. using peak matching mode, but similar values when using magnetic scanning modes[17]. The precision can be obtained through the repeatability of measurement. A random error usually causes the measurement to fall on either side of the average experimental measurement. When a set of mass measurements of an ion species lie close together, this leads to high precision. Multiple replicate experiments made under similar conditions are called repeatability. Repeated experiments made

several times in different circumstances or by different operators or instruments are called reproducibility. Both terminologies may involve some errors when the experiments are made. The effect of errors will affect the accuracy and the precision depending on the magnitude of the errors. If the measured mass gets close to the true mass the accuracy will improve towards zero.

1.2 Ionisation source

There are many types of ion sources and each has its unique design. Some are operated under vacuum and some at atmospheric pressure. One common feature is that the source will have electrical voltages applied to one or many electrodes so that ions formed can be accelerated and directed into the mass spectrometer. EI and CI are used for vapour and gases samples. CI uses a reagent ion to react with the analyte molecules, during collisions in the source, to form ions by either a proton or hydride transfer. In FAB techniques the material is mixed with non-volatile (matrix) and bombarded under vacuum with a high energy beam of atoms. Other sources often used with liquid and solid biological samples include electrospray ionization (ESI)[18] and matrix-assisted laser desorption/ionization (MALDI)[19,20].

1.2.1 Electron ionisation (EI)

Electron Ionisation (historically known as electron impact) is one of most important and widely used routine analysis (of volatile molecules) in organic mass spectrometry

because electron ionisation usually generates numerous fragment ions. The gas phase reactions in electron ionisation are:

$$M + e^- \longrightarrow M^{+\bullet} + 2e^- \qquad (1.1)$$

Where (M) is the analyte being ionized, e^- the electron, $M^{+\bullet}$ is the resulting radical cation. EI is often performed by exposing a sample to 70 eV electrons from a hot metal filament and is referred to as "hard ionisation technique" (see figure 1.1). Ionisation occurs when the energy of the electron interacting with the molecule of interest is greater than the energy contained in its bonds. Excess energy breaks bonds and fragmentation occurs.

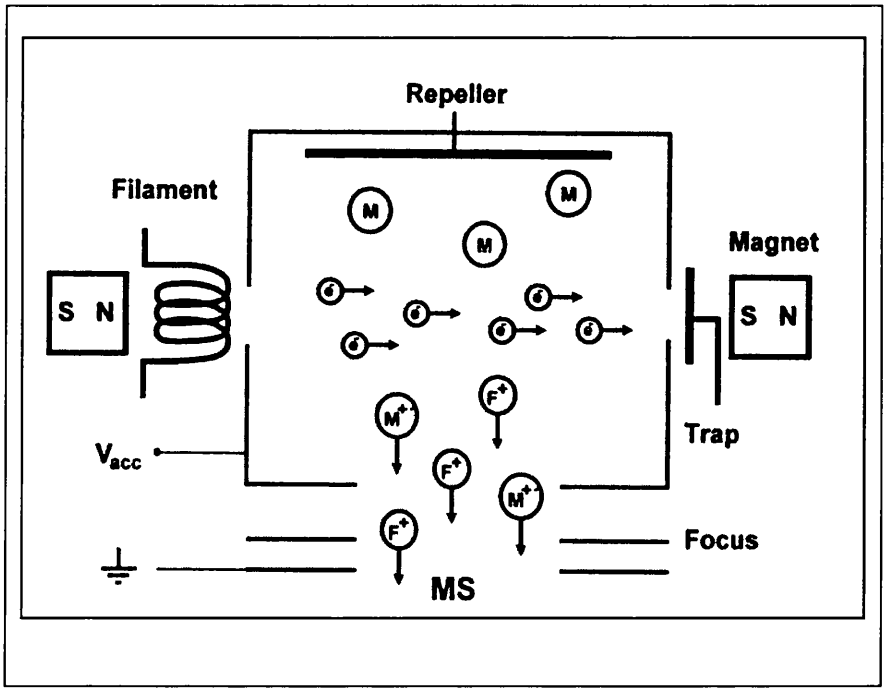

Figure 1.1 *Ionisation process, hot filament electron gun, ion source and molecule particles through the ionisation process in Electron Ionisation[21].*

Ionisation does not occur by electron capture, which is highly dependent upon molecular structure. Instead, EI produces positive ions by ejection of a valence electron from the analyte molecule

The fragment ion species can be predicted and by identifying them, the molecular structure of the analyte can be inferred and can be used in data base confirmation of the mass spectral pattern by matching unknown spectra with those residing in the data base.

1.2.2 Chemical ionisation (CI)

Chemical ionisation uses the same ion source device as electron ionization (EI). It's especially useful when no molecular ion is observed in the EI mass spectrum, or in the case of confirming the structure or the value of mass-to-charge (*m/z*) of the molecular ion. The purpose of ionisation is to form a sufficiently large number of reagent ions that make collisions during the reaction time of reactants in the ion source. Sample ions are formed by interaction of reagent ion species and analyte molecules to form $[M+H]^+$. This process is called an ion molecule reaction. CI can be used in positive and negative modes depending on the polarity of the source voltage and detector. The important issue is to set up the instrument with parameters that are appropriate, and the suitable mode that can be used for a given analyte species. The most important parameter is the pressure of reagent gas within the source volume; it must be higher than that used in EI ($\approx$ 0.1-1.0 torr) to avoid direct ionisation EI. The ionisation of

sample molecules happens through the transfer of a proton between the reagent ions and sample molecules. The reaction can be written as follow:

$$RH^+ + M \longrightarrow [M + H]^+ + R \qquad (1.2)$$

For example, EI of reagent gas to form ions: (mechanism of methane fragmentation):

$$CH_4 + e^- \xrightarrow{yields} CH_4^{+\bullet} \xrightarrow{yields} CH_3^+, CH_2^{+\bullet}, CH^+ \qquad (1.3)$$

Reaction of reagent gas ions to form adducts:

$$CH_4^{+\bullet} + CH_4 \xrightarrow{yields} CH_5^+ + CH_3 \qquad (1.4)$$

$$CH_3^+ + CH_4 \xrightarrow{yields} C_2H_5^+ + H_2 \qquad (1.5)$$

Reaction of reagent gas ions with analyte molecules (Analyte protonated):

$$CH_5^+ + M \xrightarrow{yields} [M + H]^+ + CH_4 \qquad (1.6)$$

$$C_2H_5^+ + M \xrightarrow{yields} [M + H]^+ + C_2H_4 \qquad (1.7)$$

$$C_2H_5^+ + M \xrightarrow{yields} [C_2H_5 + M]^+ \qquad (1.8)$$

In reaction (1.6) CH_5+ is the protonated reagent gas, M represented the analyte, while $[M+H]^+$ the protonated molecule and R the reagent gas. Overall the yield will depend on the enthalpy of the given reaction. A molecule is ionised via ion molecule collision between the molecule and protonated ion. The result is shown in figure 1.2. CI mass spectrum describes methane ion molecule products such as, CH_5^+ =17, the $C_2H_5^+$ =29 while 41 is $C_3H_5^+$ ion, these are all stable species formed in CI source containing methane.

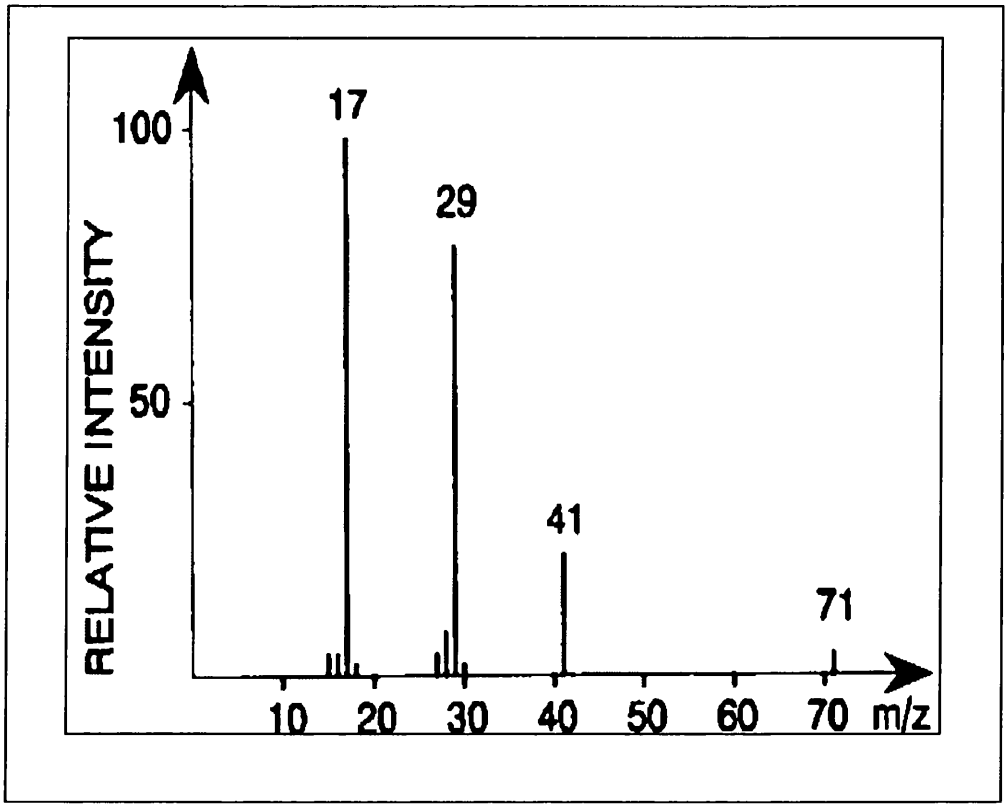

Figure 1. 2 Methane CI mass spectrum, presented [21]

1.2.3 Fast-atom bombardment (FAB) liquid secondary ion mass spectrometry (LSIMS)

In 1910, Thomson observed the release of positive ions from solid surfaces induced by ion bombardment. In 1949, Herzog and Vieh Bock achieved the first prototype experiment on solid SIMS by developing a high vacuum pump technique. Later experiments extracted ions from the source and accelerated them to the sample, thereby producing the first SIMS. The first secondary ion MS instrument was constructed under a NASA contract in 1960 to analyse and identify a sample that was extracted from moon rocks[22,23]. Bombardment of the sample surface with a primary ion beam followed by mass spectrometry analysis of the emitted secondary ions formed secondary ion mass spectrometry (SIMS). This technique is used to analyse solid surfaces by sputtering the solid surface with a focused ion beam and analysing the secondary ions. SIMS is the most widely used surface analysis technique. Secondary ions are emitted from material surfaces and used to identify the chemical composition of the material; small fragments from the sample are mass analysed to determine the elemental composition of the analyte. SIMS is a technique used for surface analysis. Primary ions are used to bombard the surface and by the collision of atoms primary ions will gain energy and transfer to the target atoms. Usually the energies of primary ions are in the keV range. Therefore the energies that yield from direct collision between primary ions and atoms are higher than bond energies[32]. Liquid secondary mass spectrometry (LSIMS) emerged in 1981 by Michael Barber. The sample is dissolved in a liquid matrix, such as glycerol, the surface of the liquid is

bombarded by xenon atoms to produce negative ions $[M-H]^-$ or positive ions $[M+H]^+$ which can be measured by mass spectrometry. This process is very soft and sensitive. Sensitivity depends on the level of the sample in the first 2μm of surface layer, and the hydrophobicity. The technique is now substituted by modern ionisation techniques, such as atmospheric pressure ionisation (API).

1.2.4 Atmospheric pressure chemical ionisation (APCI)

APCI is a form of chemical ionisation which takes place at atmospheric pressure. It can be used for the analysis of compounds of moderate polarity, such as some steroids. In effect, APCI is widely used for the ionisation of small molecules, such as drugs and their metabolites, pesticides, lipids and steroids derivatives. It is a soft ionisation technique, but unlike ESI, APCI usually produces some degree of fragmentation that is useful for structural characterisation. In APCI the samples are sprayed into a heated ionisation source.

APCI consists of a capillary through which the LC effluent passes and then through the nebuliser and surrounded by heated gas, a corona discharge needle. Background O_2 and N_2 molecules in the atmospheric source are ionised and further react with solvent molecules in the gas phase to form ions that will protonate (or deprotonate) the analyte molecules. Ions are extracted into the mass analyser by a set of differentially-pumped skimmer stages. Many ESI instruments can easily be switched to APCI mode as the basic setup of the atmospheric pressure-to-vacuum interface remains unchanged.

1.2.5 Electrospray ionisation (ESI)

Electrospray ionisation is one of the atmospheric pressure ionisation (API) techniques. The first electrospray experiments were carried out by Chapman in the 1930s[24]. However, practical work of ESI-MS appears in the 1960s when multiply charged ions were observed. In 1968, electrospray ionisation ESI atmospheric pressure was developed by Malcolm Dole and his colleagues[25]. In 1984, Yamashita and Fenn[26] developed the basis for modern ESI as it is seen today. The principle of the electrospray ionisation is the transfer of analyte species, present in solution into the gas phase as isolated entities. The technique is well-suited to the analysis of large molecules ranging from ~100 Da to more than 1,000,000 Da in molecular mass. In general, ESI is found to be flow rate independent but concentration dependent.

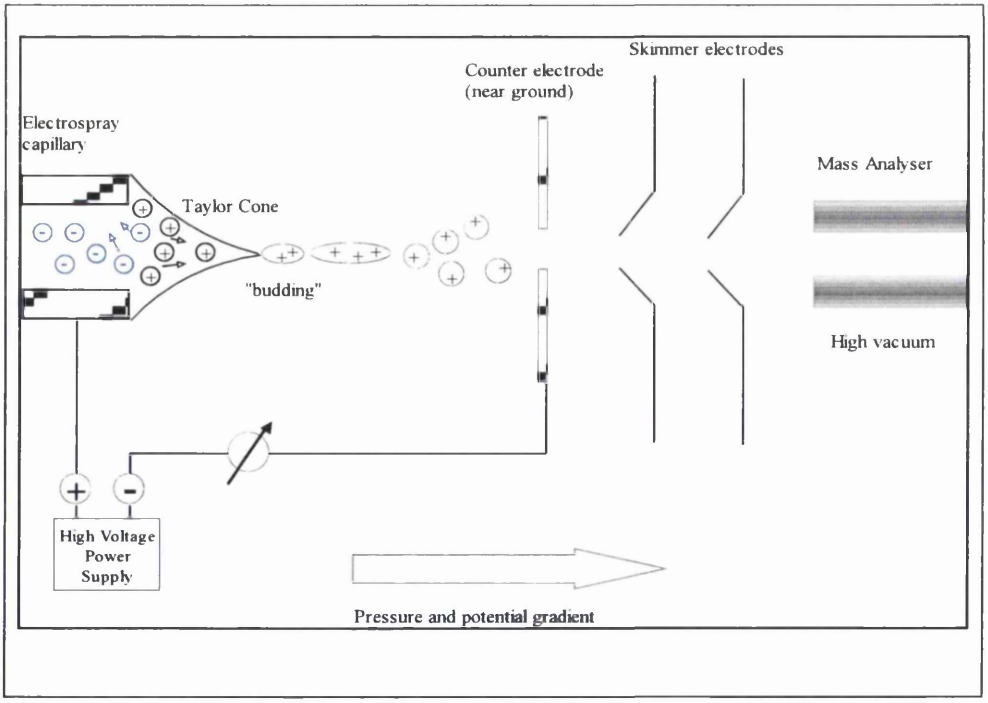

Figure 1. 3 Schematic of an electrospray interface.

During the transition of the ions, the droplet is reduced in size by the evaporation of the solvent or by Coulomb explosion, and pumped through a narrow stainless steel capillary (usually 75-150 micrometer) at flow rates from 1μL/min and 1 ml. Anelectrical potential is applied to the capillary tip (3 or 4 kV)[21]. The technique is used for the characterisation of small organic compounds such as peptides, metabolites and metal complexes[21], and is also extremely useful for the analysis of large, non-volatile, chargeable molecule such as proteins and nucleic acid polymers[27]. The difference to FAB is that the solution is composed of volatile solvent and the ionic molecule is at very low concentration (10^{-6} -10^{-3} M).

ESI is a soft ionisation technique for mass spectrometry, as it has the ability to produce intact molecular ion species. The transfer of ions from the condensed phase to that of an isolated gas phase ion starts at atmospheric pressure and leads to the high vacuum of mass analyser. The charged droplets are assisted by a warm flow of nitrogen gas known as drying gas which is passed across the ionisation source. Figure 1.4 shows the formation of gaseous analyte ions from charged droplets in the electrospray process.

Nowadays, ESI is the leading technique of the group of atmospheric pressure ionisation (API) methods as it is more readily coupled to chromatographic techniques such as liquid chromatography to form a hyphenated technique of LC/MS. Nano-electrospray ionisation (nano-ESI/MS) is a miniaturised solution form of electrospray, and used small amounts (sub-millilitre per-minute, 0.5 to few μL) of the sample

dissolved in a suitable volatile solvent at low concentration (pmol μL-1)[29,30]. Also, the flow rate is very low.

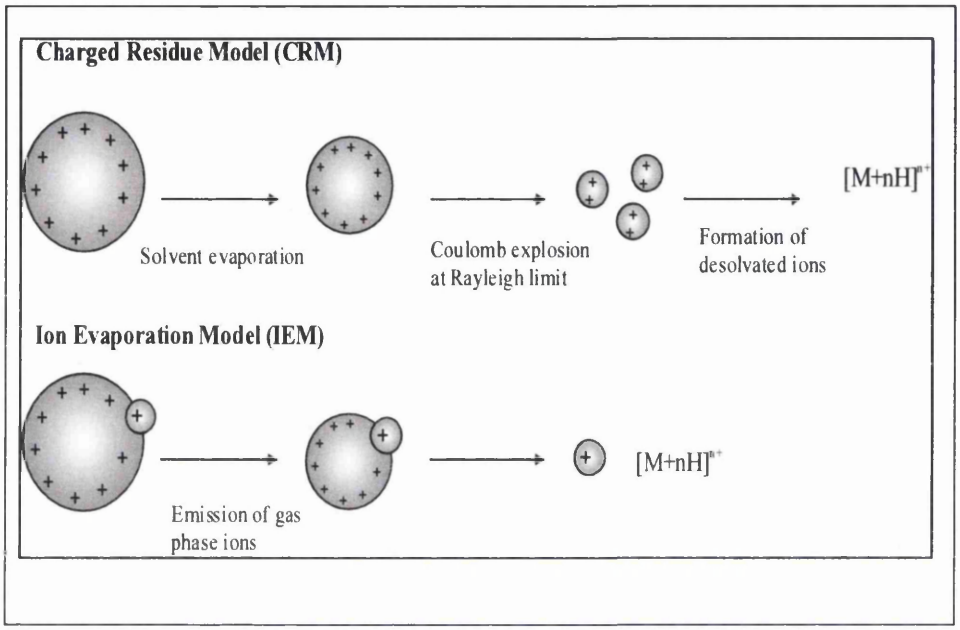

Figure 1. 4 Charge droplet formation (gaseous ion) in the electrospray process.

A common application of this method is for protein mixture analysis identifying the amino acid sequence to illustrate the structural characteristic of analyte molecules. ESI and nano-spray ionisation are very sensitive. This sensitivity is affected by the presence of any additives such as non-volatile buffers, which must be avoided. To catalyse protonation of sample molecules, acid must be added (formic acid or acetic acid) in positive mode (usually peptides and proteins) or ammonia in negative mode.

ESI and MALDI are the most commonly employed ionisation methods, which allow the widespread biological and biomedical application of mass spectrometry.

1.2.6 Matrix-assisted laser desorption/ionisation (MALDI)

MALDI is based on the bombardment of sample molecules with an intense laser light to bring about sample ionisation. The sample is pre-mixed with a highly absorbing matrix compound for the most consistent and reliable results and a low concentration-of-sample to matrix ratio usually work best. The laser is fired, the energy arriving at the sample/matrix surface optimised, and data accumulated until an *m/z* spectrum of reasonable intensity has been amassed. A time-of-flight analyser separates ions according to their mass-to-charge (*m/z*) ratios by measuring the time it takes for ions to travel through a field-free region known as the flight, or drift tube. The heavier ions are slower than the lighter ones. The *m/z* scale of the mass spectrometer is calibrated with a known sample that can either be analysed independently (external calibration) or pre-mixed with the sample and matrix (internal calibration).

MALDI is also a "soft" ionisation method and so results predominantly in the generation of singly charged molecular-related ions regardless of the molecular mass hence the spectra are relatively easy to interpret. Fragmentation of the sample ions does not usually occur. In positive ionisation mode the molecular ion species ($M+H^+$) are usually dominant, although they can be accompanied by salt adducts.

Positive ionisation is used in general for protein and peptide analyses. In negative ionisation mode, the deprotonated molecules $[M-H]^-$ are usually the most abundant species, accompanied by some salt adducts and possibly traces of dimeric or doubly charged ions.

Matrix-assisted laser desorption/ionisation is a technique that uses laser ablation of a sample-coated target to vaporise molecules for injection into a mass spectrometer. Attempts to use laser beams for ionisation in mass spectrometry began in the early 1960s, and continued during the 1970s, when pioneers worked on improving the use of lasers in the ionisation of organic samples. However, it was not until the late 1980s that MALDI was introduced as a technique for analysing peptides and proteins in relatively complex samples[20].

MALDI is a desorption /ionisation method and thus shares some similarities with FAB and LSIMS ionisation. The main differences between these techniques are the nature of the primary beam and the matrices used for the analysis. The primary beam in MALDI is a laser (photon) beam not a particle beam, and the analyte is crystallised with the matrix. The most common matrices used in UV MALDI experiments include nicotinic acid, benzoic acid and cinnamic acid derivatives, dithranol and azobenzoic acid derivatives[14]. Unlike ESI, MALDI produces mostly singly-charged ions without inducing the fragmentation of the fragile analyte. Although MALDI can be used on an ion trap, it cannot detect ions above 2,000 Da. MALDI is well suited to be coupled to time-of-flight (TOF), as both are "pulsed" techniques.

MALDI-TOF mass spectrometry is one of the latest and most fascinating developments in the analysis of organic compounds. Originally developed for the analysis of biomolecules, it has developed into one of the most powerful techniques for the characterisation of synthetic polymers[31-33]. Today, lasers are routinely used for

the production of stable ions from non-volatile, polar, thermally labile, and high molecular weight organic materials. Recent high-resolution TOF measurements on peptides and proteins, ionised by matrix-assisted laser desorption/ionisation (MALDI), are presented and discussed in terms of basic concepts important to accurate mass measurements. The laser desorption and ionisation technique is part of a wider array of soft ionization techniques available in mass spectrometry in which energy is transferred to the condensed phase under various conditions. Chemical ionisation (CI), field desorption (FD) and plasma desorption (PD), as well as fast-atom bombardment (FAB) make use of a focused atom beam for desorption and ionisation. Secondary ion mass spectrometry (SIMS) uses a focused beam of ions; these are only a few examples. The mechanisms of the MALDI process have some similarity to ESI ionisation, but also are still under investigation. The important part of this mechanism is that the matrix molecules are present in excess to analyte molecules and become electronically excited by the UV laser. This excess vibrational energy is enough to melt the crystal and transfer it to the vacuum along with analyte molecules where ionisation proceeds via a member of possible mechanisms e.g. protonation.

MALDI ionisation is, in general, more sensitive than ESI[34] and is not affected by impurities to such a high extent; it is still the most convenient method in analysing in mass region, and is routinely used for the analysis of proteomics samples. However, the mass resolution in the high mass region is too poor to give valuable information for biomolecules identification[35]. Another advantage of MALDI is that, coupled with a TOF analyser, biomolecules up to 500,000 Da can be investigated. The mass range

can even be stretched further to 1–2 mDa, but detection of these "ultrahigh" masses requires sensitive and specific detectors. Disadvantages of MALDI include low salt tolerance, even though it is generally less critical than for ESI.

1.3 Mass analysers

In many mass spectrometers, sample is ionised in the ion source at atmospheric pressure and the ions generated are introduced through a series of various pumped regions into the mass analyser at pressure equal to 10-4 Pa or less, except for quadrupole ion trap (10-1 Pa)[20]. Recently, many literature papers were published on the use of atmospheric pressure ionisation (API) techniques, electrospray (ES) and (MALDI), as ion sources for mass spectrometric analyses of molecular elemental species. Both are soft ionisation techniques which produce gas-phase analyte ions directly from pre-existing ions in solution. They have been used extensively, often coupled to LC, for analyses of large organic molecules such as proteins, peptides and other bio molecules by MS and tandem mass spectrometry (MS/MS). A chemical species is defined as a specific form of a chemical element, such as molecular or complex structure or oxidation state. In addition, there are many new surface ionisation techniques introduced in recent years based on which ions are formed adjacent to and on the surface of the material that assist the analyte. Fast-atom bombardment (FAB), and liquid secondary ion mass spectrometry (LSIMS) are also still used over a wide range of analysis of large organic compounds.

1.3.1 Magnetic sector mass spectrometry

As described in the introduction of this chapter, it is important to mention the work of the mass spectrometry pioneers such as Aston and Dempster, who have revolutionised the use of mass spectrometry. In a magnetic sector instrument ions of mass (m) and charge (z) leave the ion source with high velocity (v), (not all ions exactly have the same energy or the same velocity, therefore the resolution will be limited). The ion then passes toward the collector through a magnetic sector and deflects in a circular trajectory in the magnetic field. Their behaviour can be studied by the combined effects of magnetic and electric fields[21]. Radius (r), magnetic field (B) and electrical potential (V) are essential factors in this process. The relationship between the centrifugal force (mv2/r) and magnetic sector force (Bzv) is

$$\frac{mv^2}{r} = Bzv \qquad (1.9)$$

The kinetic energy ½ mv^2 is equal

equation of accelerating is given by;

$$zV = \frac{1}{2}mv^2 \qquad (1.10)$$

and by compensating in the two equations:

$$\frac{m}{z} = \frac{B^2r^2}{2v} \qquad (1.11)$$

The formula shows that the radius of curvature (r) for the ion in the mass spectrometer depends on the value of *m/z* as long as they have the same kinetic energy. Therefore the magnetic field is used to separate a monoenergetic ion beam into different masses. When a molecular ion of mass (m_1) carries a single positive charge it may decompose to form a fragment ion of mass (m_2) and neutral fragment ($m_3=m_1-m_2$). If there is no kinetic energy of separation of the fragments, the ion (m_2) and the neutral fragment will continue along the direction of motion of (m_1) with unchanged velocity. A fragment reaction may be written as follows:

$$m_1 \longrightarrow m_2 + m_3 \qquad (1.12)$$

Even though magnetic analyzers are now considered outdared and not widely used as it was in the previous decades, sectors have played an important role in establishing mass spectra fragmentation rules

1.3.2 Quadrupole mass spectrometry

A quadrupole consists of four (hyper-bolic cross-section) parallel circular metal rods, which are part of a mass spectrometer that is responsible for filtering ions based on *m/z* value. Each two opposite rods are connected together electrically. Mass separation is accomplished using a combination of direct current (dc) and radio frequency (RF) signal. A small voltage is applied to start ion motion before entry into the quadrupole device[28]. Only ions of certain *m/z* will reach the detector for given dc and rf voltages and induce a detection signal while other ions will collide with the rods. A mass

spectrum is obtained by monitoring the ions passing through the quadrupole filter as the voltage on the rods is varied.

1.3.3 Time-of-flight (TOF)

The principle of TOF is the simplest form of mass spectrometry[36]. It is a pulsed technique from which the initial pulse defines the start time of the ion, and it is used for accurate mass measurement for empirical formulae determination[37].

Time–of–flight (TOF) mass spectrometry is probably the simplest method of mass measurement to conceptualise, which identifies the elemental compounds by subjecting the sample of ions to strong electrical acceleration field, and by measuring the time that it takes for each ion to reach a detector while travelling over a known distance. The history of this technique started with Stephen's concept in 1946 at the American Physical Society[1,38]. His concept was to accelerate all ions to the same kinetic energy. Thus an accelerated ion would have the same kinetic energy as any other ion (provided they had the same charge), with the velocity depending solely on *m/z*. This concept was demonstrated by Cameron and Eggers with the analysis of mercury vapour[39]. In 1955, Wiley and McLaren introduced a TOF focusing scheme that improved mass resolution by correcting energy distributions of the ions[40]. Time-of-flight has become widely used over the past decade as an essential instrument in analytical chemistry, especially when coupled with MALDI and ESI ionisation methods.

Ions of different *m/z* value have different velocities and are consequently separated into clusters (groups) in space. The mechanism of this process begins by generating the ions in the ion source chamber that is formed as a packet over a short time. Separation of ions is based on their different flight times that are needed to transit the ion source to the detector through a field free region[25] and by using suitable voltages. All ion packets are accelerated into the mass analyser (flight tube) at a constant kinetic energy.

Alfred Benninghoven, at the University of Muenster was one of the pioneers in this field, and completed a TOF-SIMS in 1979 by using a Poshenreider mass analyser[25]. In the 1980s Kore personnel were involved in developing and selling numerous surface analysis instruments, based around time-of-flight secondary ion mass spectrometer (SIMS). In 1995 Kore also developed a novel table top SIMS (solid sample surface contamination analyzer). TOF analysers were fit for purpose in some quantitation applications covering both small and large molecule analyte.

In the time-of-flight analyzer, ions are rapidly accelerated toward a "field-free" drift region which is called the "flight tube", and their separation is achieved by measuring the difference in the time needed to travel from the ion source to the detector called "time of flight".

Ions of a given *m/z*-ratio, are accelerated to the same final kinetic energy (zeV) across a drift region (D), will have velocities (V) and times-of-flight (t). The relationship between these parameters is summarised in this equation:

$$K = \frac{mv^2}{2} = qV \qquad (1.13)$$

$$q = ze \qquad (1.14)$$

$$t = \frac{D}{v} \qquad (1.15)$$

$$t = D\sqrt{\frac{m}{2zeV}} \qquad (1.16)$$

There is an inverse relationship between velocity and the square root of *m/z*, so the lighter ions have a higher velocity than the heavier ones and will consequently reach the detector sooner.

Linear field reflectron is a technique that allows ions which have great kinetic energy to penetrate deeper into the reflectron than ions with smaller kinetic energies.

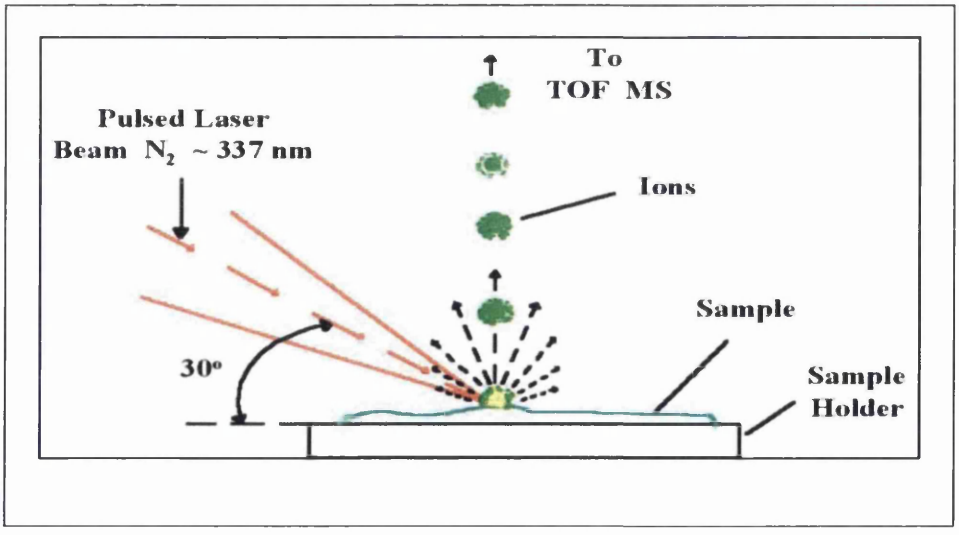

Figure 1. 5 MALDI-TOF MS Sample Ionization process.

The ion beam is folded onto itself and the ions are sent to the detector. In the case where ions have varying kinetic energies, the reflectron will decrease the spread in the ion flight times thus improving the resolution of the time-of-flight mass spectrometer.

Both linear and reflectron TOF devices are broadly used for biomolecule measurements depending on the signal level and resolution requirements[25]. The mass resolution for a MALDI linear TOF mass spectrometer can reach 2,000, while the mass resolution of a reflectron TOF mass spectrometer can be higher than 10,000. However, a reflectron TOF mass spectrometer is difficult to apply to molecules with *m/z* more than 20,000 Da if these ions are not multiply charged[25].

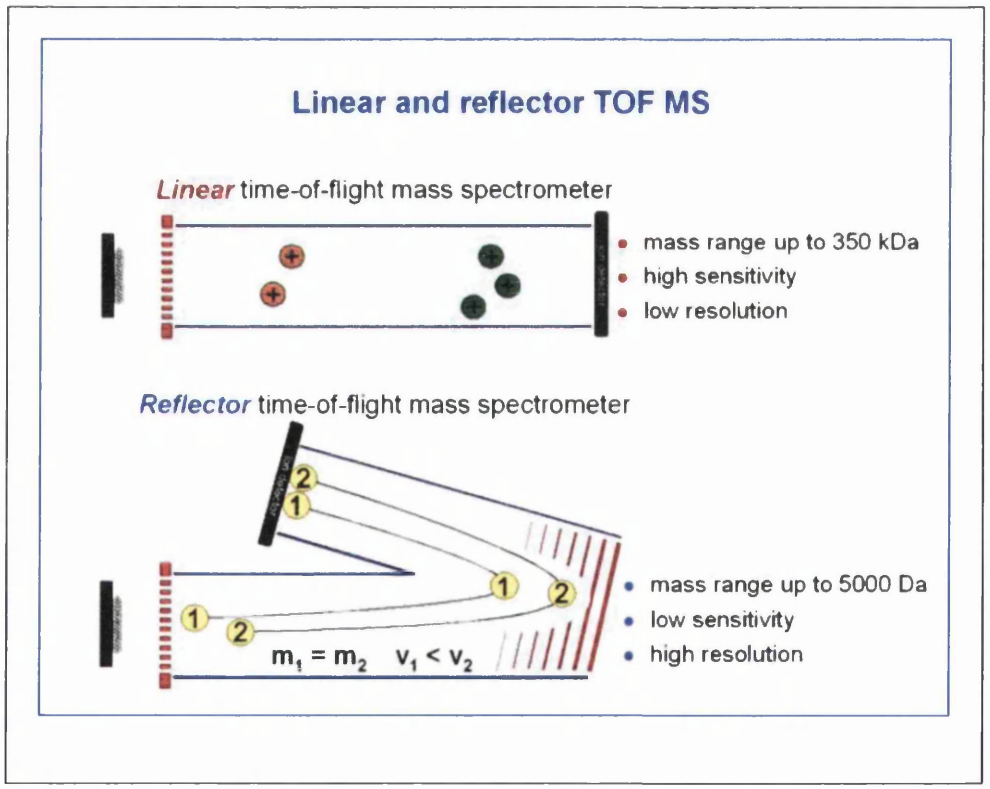

Figure 1. 6 *Time-of-flight mass spectrometry (taken from anagnostec site anagnostec@anagnostec.de*

1.3.4 Fourier–transform ion cyclotron resonance (FT-ICR)

Fourier–transform ion cyclotron resonance is a type of mass analyser (or mass spectrometer) for determining the mass-to-charge ratio (*m/z*) of ions based on the cyclotron frequency of the ions in a fixed magnetic field[41]. The theory of cyclotron motion was developed in the early 1930s by Lawrence. In the 1960s Llewellyn and Baldeschwieler[42,43] introduced a different detection method to be used with cyclotron mass spectrometry, when they detected ion resonant power absorption instead of counting ions on a collector plate. However, the major innovation came in 1974, when Comisarow and Marshall[44] made the technique more similar to nuclear magnetic resonance (NMR) by introducing the Fourier transform (FT) algorithm, which allows many different molecules to be analysed faster in a shorter time.

Fundamental aspects of FT-ICR can be understood from very simple idealised models. Initially, ion cyclotron frequency, ion cyclotron motion, velocity, and energy as a function of ion mass, ion charge, and the strength of magnetic field follow directly from the motion of ion in especially uniform static magnetic field[44]. Fourier transform ion cyclotron resonance mass spectrometry (FT-ICR MS) provides the highest mass resolution and mass accuracy of present MS technologies and the most powerful technique in proteomics research[45].

Charged particles in a fixed magnetic field move at a frequency related to *m/z* values. Ions produced by electron ionisation, laser ionisation, electrospray ionisation, or some other means are stored inside an ICR analyser cell that is situated in the homogeneous

region of a high field magnet, and are trapped in the fixed magnetic field under current potential (electric trapping plate). When the frequency of the applied sine wave signal is the same as the cyclotron frequency of the ions, the excitation process will be established and the ions will accelerate steadily by electric field perpendicular to both magnetic field (in oscillating motion) and velocity. Ions then will transfer in packets and give a signal as an image current on a pair of plates which is called the free induction decay (FID). Confinement of ions by the application of a three-dimensional axial quadrupole electric field shifts the ion cyclotron frequency, whereas excitation and detection remain basically linear[46]. Fourier cyclotron mass spectrometry is a unique method in that increased sensitivity and resolution regarding to increase the time, the ICR-MS, when operated in FT-ICR mode can increase resolution beyond 100,000 without energy filtering. Nowadays, most FT-ICR mass spectrometers are equipped with a hexapole accumulation region preceding the ion transfer optics. After their formation (outside the cell) and external accumulation, ions are transported along the field lines into the magnet.

1.3.5 Orbitrap mass spectrometer

An Orbitrap is a type of mass spectrometer developed by Alexander Makarov. It consists of an outer barrel-like electrode and a coaxial inner spindle-like electrode that form an electrostatic field with quadro-logarithmic potential distribution[47,48]. The Orbitrap is an electrostatic ion trap that uses Fourier transform to obtain mass spectra. The principle of ions trapping in electrostatic field was described by Makarov[47].

Orbital trapping has been studied experimentally and developed dramatically since that time. The concept of Orbitrap in a 3 dimensional electrostatic field was presented in elegantly revised of a new type of analyser and developed by Alexander Makarov, et al.[49] when they used ion axial oscillated and image current detection for HPLC mass analyzer. Typically, the Orbitrap instrument consists of three independent ion trapping devices, two works as frequency ion traps (LTQ & c-trap) whereas the Orbitrap is a coaxial spindle like an electrode that forms the electrostatic field[47,48].

1.4 Comparison of mass analyser properties

Analytical instruments in general have variations in capabilities as a result of their individual design and intended purpose. This is also true for mass spectrometers whose performance depend on its mass analyser. Some mass analysers separate ions in space, others by time. All measure the *m/z* ratio, not the mass, by subjecting the ions to constant, pulse, or periodically time- varying electric and /or magnetic fields. If an ion has multiple charges, the *m/z* ratio will be significantly less than its relative molecule was. In recent years, the mass analyser has been developed and the design of modern analysers have changed, offering high throughput, higher mass accuracy, broader mass range, increased sensitivity and the ability to give structural information. Scan speeds differ from type to type, some instruments needs seconds to make a full mass scan others less, and this depends on mass analyser design. A TOF analyser, for example, completes analysis in milliseconds. Table 1.1 and 1.2 illustrate a general

comparison of mass analysers typically used. These values vary with instrument manufacturer.

Table 1. 1 Attributes and typical range of specifications of commercial mass spectrometer instruments.

Mass analyzer type	Measures	Mass-to-charge (*m/z*) range	Mass resolution at 1000 u[1,2]	MMA (ppm)[3]	Dynamic range[4] (Upper limit)	Sensitivity moles
1.Magnetic Sector	Momentum-to-charge	10^4	10^5	2-5	10^7	$>10^{-10}$
2.Quadrupol	Path stability	10^4	10^3	~100	10^5	$\sim10^{-15}$
3. Ion trap (IT)	Frequency	10^3 - 10^4	10^3 - 10^4	~100	10^5	$\sim10^{-15}$
4. (TOF)	Flight time	10^6	10^3 - 10^4	5 – 10^5	10^2 - 10^5	10^{-15} - 10^{-13}
5. Orbitrap	Frequency	10^4	$5*10^5$	1-3	10^4	10^{-15} - 10^{-13}
6. (FTICR)	Cyclotron frequency	10^5	10^6	1 - 2	10^4	10^{-19} - 10^{-15}

1 u - IUPAC unit of mass

2 Mass resolutions is usually defined at half peak height except for the magnetic sector instrument where the definition is 10% valley.

3 ppm – parts per million resolution.

4 Typical best values; the wide range of values for TOF depends on the method for ion signal detection. Time-to-digital (TDC) detectors have lower dynamic range then analogue-to-digital based detectors. Hybrid TOF detectors can achieve up to 7 orders of magnitude dynamic range. For a typical acquisition of 1 sec. over a chromatographic peak the dynamic range would be 102 – 103, for the three most important techniques of IT, TOF and FTICR.

5 Stability of mass measurement accuracy can be poor with a TOF and some are very susceptible to temperature fluctuations. Internal calibration is essential.

Table 1.2 The main mass analyzers used in mass spectrometry and features of note.

Mass analyzer type	Suitable for Accurate mass	Robustness	Cost	Complexity	Speed
1.Magnetic sector	Yes	-	-	-	-
2.Quadrupole	No	Very	Low	Low	Fast
3.Iontrap (IT)	No	Very	Medium	Low-medium	Fast
4. (TOF)	Yes	Moderate	Medium - high	Low-medium	Very fast
5. Orbitrap	Yes	Very	Medium - high	Low-medium	Fast
6.Fourier-transform ion cyclotron resonance (FTICR)	Yes	Moderate	High	High	Moderate

**Some researchers have shown how quadruples can be used for accurate mass work but such systems have not been commonly adopted.*

1.5 Elemental formula determination and accurate mass measurement

Accurate mass measurement is used to determine the elemental formula for molecules and fragment ions to investigate the presence of known compounds or to assist in the identification of analyte molecules. In determining the elemental composition for an unknown compound from an exact mass measurement, it is important to know the molecule under study, and try to answer our questions; what is the size of this

molecule? Is it small molecule or large molecule? (Protein, enzyme, synthetic polymers) and what is the functional group? Answering will give us knowledge about the molecule and can help us to reach our goal and apply suitable methods to determine its elemental composition. It is difficult to determine the elemental formula for unknown molecules and this is more difficult if it's bigger than 500 Dalton (Da) because of the number of possible formulae might exist and this increases rapidly with decreasing mass. It is possible to reduce the size of a molecule by identifying the mass of functional groups. We can then subtract these from the measured mass to reduce the number of candidates.

Comparing precision from instrument: millimass unit, measurement error (ppm), and resolution according to the accurate mass best practice guide of the VIMMS programme, an initiative that form part of the UK national measurement system, most instruments that are used for accurate mass measurement are capable of achieving a precision better than 5 ppm. Accurate mass can be obtained by using magnetic sectors or FTICR, Orbitrap, TOF and QTOF instruments which have a high accuracy and sufficient resolving power.

1.5.1 Isotopes and atom properties

Isotopes are species of the same element that have the same number of protons and electrons but different mass numbers because they contain a different number of neutrons. The chemical properties of the isotopes of an element are the same but there may be differences in their physical properties due to the different number of

neutrons. A number of elements have distinctive isotope patterns consequently, if these elements are contained in the analyte, then these pattern will appear through the analysis process. Chlorine is a good example of distinctive stable isotopes patterns (^{35}Cl:^{37}Cl in the ratio 3:1) making the presence of chlorine in a sample obvious. Other common elements which have easily distinguishable isotope patterns include bromine (^{79}Br:^{81}Br in the ratio 1:1)[50]. Other long-lived common stable isotopes in nature is (^{12}C,^{13}C,^{16}O,^{18}O). Observation of the $^{12}C/^{13}C$ ratio can give a good indication of the number of carbon atoms in the sample according to the NBS standard. Naturally occurring carbon contains about ≈1.1% of ^{13}C isotope peak with abundance ~1.1% of the ^{12}C peak. So an ion containing 50 carbon atoms will have a ^{13}C isotopes peak with abundance ~ 55% of ^{12}C peak.

1.5.2 Nitrogen rule

A compound containing common elements and an odd number of nitrogen atoms will have an odd molecular weight. If the molecule ion is an even number it must have an even number or zero nitrogen atoms. This rule is used on an integer mass not with accurate mass.

1.5.3 Rings and double bounds

Organic mixture compounds involve a lot of elements which give complex mass spectra because the elements that compose are not isotopically pure. Because of valences of the elements involved, the total number of rings and double bounds in a molecule of formula $C_xH_yN_zO_n$ will be equal to x - 0.5y + 0.5z + 1. [McLafferty][51].

For an even electron ion the true value will be ending by 0.5. Example pyridine (C_5H_5N:) ring plus double bond = 5 - 2.5 + 0.5 +1 = 4 this value represent the ring and the three double bonds of pyridine. For example benzyl odd–electron, (C_7H_5O:) ring plus double bonds = 7 – 2.5 +1 = 5.5 for example C_6H_5CO+ Benzoyl (even electron) , the 5.5 calculated for the benzoyl ion represents the ring, the three double bound of benzene and double bound of the carbonyl group.

1.5.4 Elemental composition determination

The elemental composition of an ion containing C, H, O, N, P, or S atoms with a mass up to 150 Da can be determined from its exact mass when the error in the mass measurement is small [52,53] (measured to within few part per million (ppm)), most compounds of regulatory interest yield ions heavier than 150 Da. In this case, the elemental composition is determined from the fragment ions, which when put together will give the molecular ions, in other words the elemental composition of the analyte is inferred from its parts. Typically, accurate mass measurement is less-ambiguous up to 300 Da[54] due to the small number of potential molecular compositions, however with increasing *m/z* value, the number of possible formulae increases rapidly with its mass, and then the identification of the molecule is increasingly difficult.

The environmental Protection Agency developed ion composition elucidation (ICE) to determine the unique composition of ions with masses up to 600 Da, containing C, H, O, N, P, or S atoms depending on atomic masses, valence of element, and relative abundances of higher isotopes[55]. Many applications were applied throughout the last

century for which accurate mass measurement is required. As mentioned before, in 1927 Aston achieved an accuracy of one part in 10,000 (=100 ppm) using magnetic sector field mass spectrometry with resolving power of 600.

In 1959 John Beynon[4] published a report including the mass value of an ion which could be measured with sufficient accuracy and its elemental formulae found. Various techniques are used to achieve high accuracy in magnetic sector field mass spectrometry such as peak matching, accelerator voltage scanning and magnetic scanning. Peak matching can produce very accurate results close to the theoretical value (1-3 ppm). In magnetic sector mass spectrometry it is usual to vary the magnetic field strength in order to bring ions of given *m/z* into focus. The relationship between mass and accelerating voltage at constant magnetic field is inversely proportional. This voltage is switched between two values to bring ions on to the detector: one value brings reference ion to the detector, and another brings the analyte; the ratio between the two values gives the ion mass. In the accelerating voltage scanning technique, the magnetic field must be held constant.

The magnet sets the initial *m/z* value just below the mass range of interest, then the ions to be analysed will pass through the instrument together with some reference ions. The magnetic scanning technique scans all ion masses, fast scanning, and provides good quality data. Determining elemental composition from an exact mass measurement for a molecule, which is essentially entirely unknown, becomes difficult. In this case for the purpose of elemental composition determination, usually exact

mass measurements are performed on molecules below 500 Da with a required mass accuracy at or better than 5 mDa. The instrument was used for high resolution accurate mass of organic compounds using peak matching and magnetic scanning and with increasing mass the number of elemental combinations increases exponentially.

Exact mass traditionally has been obtained by FAB magnetic-sector instruments[54,56] and more recently by FTMS, MALDI TOF and quadrupole time-of-flight type instruments[54-57]. Time-of-flight mass spectrometer (TOF-MS) is an attractive instrument due to its potentially unlimited *m/z* range and high speed acquisition capabilities. The reflectron, delayed extraction, and orthogonal acceleration greatly increase resolving power and consequently affect mass accuracy measurement. Reflectron technology compensates for kinetic energy distribution, increases resolution, and increases the certainty of mass assignment. The coupling of time-of-flight mass spectrometry TOF-MS with electrospray ionisation ESI was achieved in the last decade[58,59].

1.5.5 Examples of elemental composition determination

The molecular weight of a compound can be measured with a high degree of accuracy. Subject to the accuracy obtained, it may be possible to assign the elemental composition to a compound based solely on the molecular weight measurement.

Table 1.3 shows an example of elemental composition determination for ion with mass measured as 309.11287 thought to be protonated molecule species $[M+H]^+$. Calculation for these compounds were measured by electrospray ionisation using

magnetic sector double focusing mass spectrometer whose precision has been measured to be 0.81 ppm (1 standard deviation) for ESI positive mode.

Table 1.3 Elemental composition determination for ion as an example applied by ESI + magnetic sector.

Rank	Mass (*m/z*)	Δmi (mDa)	Δmi (ppm)	RDB	Elemental Composition
1	309.1128	0.05	0.16	7.5	$^{12}C_{12}\,H_{17}\,O_2\,N_6\,^{32}S$
2	309.1128	0.11	0.36	2.5	$^{12}C_{11}\,H_{23}\,O_4\,N_2\,P_2$
3	309.1126	0.23	0.74	-1.0	$^{12}C_5\,H1_9\,O_{10}\,N_5$
4	309.1131	-0.26	-0.84	3.0	$^{12}C_8\,H_{20}\,O_2\,N_7\,P\,^{32}S$
5	309.1124	0.42	1.36	7.0	$^{12}C_{15}\,H_{20}\,O_4\,N\,P$
6	309.1135	-0.60	-1.94	16.5	$^{12}C_{20}\,H_{13}\,N_4$
7	309.1121	0.73	2.36	11.5	$^{12}C_{19}\,H_{17}\,O_4$

The number of choices can be reduced by applying the ring and double bond rule. There is non-zero possibility that the correct elemental composition may lie outside of any arbitrarily defined limit. The choice of limiting the listed elemental composition to ±3 standard deviations (2.43 ppm) is a reasonably rigorous option.

1.6 Research aims

The main objective of this study is to evaluate the relative capabilities of different mass spectrometry instruments, to investigate the effect of ionization techniques on the accuracy of instruments for different modes of operation and to study the distribution behaviour over the mass scale. The other target is to study the appropriate statistical methodology to evaluate the accuracy and the precision of instruments.

1.7 References

1. Purcell EM, Torrey HC, Pound RV. *Physical Review. American Physical Society.*1946; **69**: 674.

2. Thomson J. *Rays of positive electricity. Phil. Mag.* 1910; **20**: 752-767.

3. Beynon JH. *Qualitative Analysis of Organic Compounds by Mass Spectrometry. Nature.* 1954; **174**: 735-737.

4. Beynon JH. *High Resolution Mass Spectrometry of Organic Materials. Adv.Mass Spectrom.* 1959: 328-354.

5. Beynon JH. *Thirty Years of Mass Spectrometry: A Personal Perspective. Appl. Spectrosc.* 1979; **33**: 339-345.

6. Beynon JH. *Mass Spectrometry and Its Applications to Organic Chemistry, Amsterdam. Elsevier.* 1960.

7. Roboz J, Holland JF, McDowall MA, Hillmer MJ, Aplin R. *Accurate mass measurement in continuous flow fast atom bombardment quadrupole mass spectrometry. Rapid Communications in Mass Spectrometry.* 1988; **2**: 64-66.

8. Graham RC. *Special feature: Historical. Collision-induced dissociation: Readings and commentary. Journal of Mass Spectrometry.* 1995; **30**: 1215-1221.

9. Marshall AG, Rodgers RP. *Petroleomics: The Next Grand Challenge for Chemical Analysis. Accounts of Chemical Research.* 2003; **37**: 53-59.

10. IUPAC. *unified atomic mass unit. Compendium of Chemical Terminology-Gold Book.* 2 ed; 2006.

11. IUPAC. *Standard Definitions of Terms Relating to Mass Spectrometry-Provisional Recommendations. IUPAC.org.* 2006.

12. Guilhaus M. *Special feature: Tutorial. Principles and instrumentation in time-of-flight mass spectrometry. Physical and instrumental concepts. Journal of Mass Spectrometry.* 1995; **30**: 1519-1532.

13. Guilhaus M, Mlynski V, Selby D. *Perfect Timing: Time-of-flight Mass Spectrometry. Rapid Communications in Mass Spectrometry.* 1997; **11**: 951-962.

14. March RE. *An Introduction to Quadrupole Ion Trap Mass Spectrometry. Journal of Mass Spectrometry.* 1997; **32**: 351-369.

15. Ferrer I, Thurman EM. *Accurate mass mesurments with ortogonal axis Time-of-Flight mass spectrometry. Liquid Chromatography Time-of-flight Mass Spectrometry.* **173** of Chemical Analysis: John Wiley and Sons; 2009. 280

16. Griffiths WJ, Jonsson AP, Liu S, Rai DK, Wang Y. *Electrospray and tandem mass spectrometry in biochemistry. Biochem. J.* 2001; **355**: 545-561.

17. Hernandez F, Ibanez M, Sancho JV, Pozo OJ. *Comparison of Different Mass Spectrometric Techniques Combined with Liquid Chromatography for Confirmation of Pesticides in Environmental Water Based on the Use of Identification Points. Analytical Chemistry.* 2004; **76**: 4349-4357.

18. Fenn J, Mann M, Meng C, Wong S, Whitehouse C. *Electrospray ionization for mass spectrometry of large biomolecules. Science.* 1989; **246**: 64-71.

19. Karas M, Bachmann D, Hillenkamp F. *Influence of the wavelength in high-irradiance ultraviolet laser desorption mass spectrometry of organic molecules. Analytical Chemistry.* 1985; **57**: 2935-2939.

20. Tanaka K, Waki H, Ido Y, Akita S, Yoshida Y, Yoshida T, Matsuo T. *Protein and polymer analyses up to m/z 100 000 by laser ionization time-of-flight mass spectrometry. Rapid Communications in Mass Spectrometry.* 1988; **2**: 151-153.

21. Brenton AG. Mass spectrometry and combined system, Literature note. Swansea; 2007.

22. Lesney M. *Spectacular Spectrometry: The corporate-led evolution of MS produced an irreplaceable tool.*

23. *The Online Encyclopaedia and Dictionary Secondary ionisation In Mass spectrometry, FactArchive.com*, 2005.

24. Chapman S. *Carrier Mobility Spectra of Spray Electrified Liquids. Physical Review.* 1937; **52**: 184.

25. Dole ML, Hines RL. *Molecular Beams of Macroions. Journal of Chemical Physics.* 1968; **49**.

26. Yamashita M, Fenn JB. *Electrospray Ion Source. Journal of Chemical Physics.* 1984; **88**: 4451–4459.

27. Amad MH, Cech NB, Jackson GS, Enke CG. *Importance of gas-phase proton affinities in determining the electrospray ionization response for analytes and solvents. Journal of Mass Spectrometry.* 2000; **35**: 784-789.

28. Wilm M, Mann M. *Analytical Properties of the Nanoelectrospray Ion Source. Analytical Chemistry.* 1996; **68**: 1-8.

29. Wahl JH, Goodlett DR, Udseth HR, Smith RD. *Attomole level capillary electrophoresis-mass spectrometric protein analysis. Analytical Chemistry.* 1992; **64**: 3194-3196.

30. Wilm M, Shevchenko A, Houthaeve T, Breit S, Schweigerer L, Fotsis T, Mann M. *Femtomole sequencing of proteins from polyacrylamide gels by nano-electrospray mass spectrometry. Nature.* 1996; **379**: 466-469.

31. Montaudo G, Samperi F, Montaudo MS. *Characterization of synthetic polymers by MALDI-MS. Progress in Polymer Science.* 2006; **31**: 277-357.

32. Zhu H, Yalcin T, Li L. *Analysis of the Accuracy of Determining Average Molecular Weights of Narrow Polydispersity Polymers by Matrix-Assisted Laser Desorption Ionization Time-of-Flight Mass Spectrometry. Journal of the American Society for Mass Spectrometry.* 1998; **9**: 275-281.

33. Kuang JW, Odom WR. *Characterizing Synthetic Polymers by MALDI MS. American Chemical Society.* 1998: 456 A-461 A.

34. Schiller J, Süß R, Arnhold J, Fuchs B, Leßig J, Müller M, Petkovic M, Spalteholz H, Zschörnig O, Arnold K. *Matrix-assisted laser desorption and ionization time-of-flight (MALDI-TOF) mass spectrometry in lipid and phospholipid research. Progress in Lipid Research.* 2004; **43**: 449-488.

35. Chen CH. *Review of a current role of mass spectrometry for proteome research. Anal Chim Acta.* 2008; **624**: 16-36.

36. Wiley WC, McLaren IH. *Time-of-Flight Mass Spectrometer with Improved Resolution. Bendix Corporation.* 1955; **26**.

37. Boesl U, Bämann C, Käsmeier R. *Time-of-flight mass analyser for anion mass spectrometry and anion photoelectron spectroscopy. International Journal of Mass Spectrometry.* 2001; **206**: 231-244.

38. Stephens WE. *A Pulsed Mass Spectrometer with Time Dispersion. Phys.* 1946; **69**: 691.

39. Cameron AE, Eggers DF. *An Ion Velocitron. Rev. Sci. instrum.* 1948; **19**: 605-609.

40. Wiley WC, McLaren IH. *Review of Scientific Instruments. Bendix Corporation.* 1955; **26**: 1150.

41. Marshall AG, Hendrickson CL, Jackson GS. *Fourier transform ion cyclotron resonance mass spectrometry: A primer. Mass Spectrometry Reviews.* 1998; **17**: 1-35.

42. Baldeschwieler JD. *Ion cyclotron resonance spectroscopy. American Association for the Advancement of Science.* 1968; **159**: 263-273

43. Wobschal D. *Ion cyclotron resonance spectrometer. American Association for the Advancement of Science.* 1965; **36**: 466-475.

44. Comisarow MB, Marshall AG. *Fourier transform ion cyclotron resonance spectroscopy. Chemical Physics Letters.* 1974; **25**: 282-283.

45. Fernandez FM, Wysocki VH, Futrell JH, Laskin J. *Protein Identification via Surface-Induced Dissociation in an FT-ICR Mass Spectrometer and a Patchwork Sequencing Approach. Journal of the American Society for Mass Spectrometry.* 2006; **17**: 700-709.

46. Brown LS, Gabrielse G. *Geonium theory: Physics of a single electron or ion in a Penning trap. Reviews of Modern Physics.* 1986; **58**: 233.

47. Makarov A. *Electrostatic Axially Harmonic Orbital Trapping: A High-Performance Technique of Mass Analysis. Analytical Chemistry.* 2000; **72**: 1156-1162.

48. Hu Q, Noll RJ, Li H, Makarov A, Hardman M, Cooks RG. *The Orbitrap: a new mass spectrometer. Journal of Mass Spectrometry.* 2005; **40**: 430-443.

49. Makarov A. *Mass spectrometer. U.S. Patent.* 1999; **5**: 886,346.

50. Bollenbacher AF, Guenther PR, Keeling CD, Stewart EF, Wahlen M, Whorf TP. *Calibration Methodology for the Scripps 13C/12C and 180/160 Stable Isotope Program*, 1992-1996. La Jolla, CA 92093-0244: *Scripps Institution of Oceanography*; 2000.

51. McLafferty FW, Turecek F. *Interpretation of mass spectra,* University scince books, US ; 4th.ed, California, 1993.

52. Mas S, Perez R, Martinez-Pinna R, Egido J, Vivanco F. *Cluster TOF-SIMS imaging: A new light for in situ metabolomics. proteomics.* 2008; **8**: 3735-3745.

53. Guan S, Marshall AG. *Ion traps for Fourier transform ion cyclotron resonance mass spectrometry: principles and design of geometric and electric configurations. International Journal of Mass Spectrometry and Ion Processes.* 1995; **146-147**: 261-296.

54. Bristow AWT, Webb KS. *Intercomparison study on accurate mass measurement of small molecules in mass spectrometry. J Am Soc Mass Spectrom.* 2003; **14**: 1086-98.

55. Grange AH, Osemwengie LI, Brilis GM, Sovocool GW. *Ion Composition Elucidation (ICE): an Investigative Tool for Characterization and Identification of Compounds of Regulatory Importance. Environmental Forensics.* 2001; **2**: 61-74.

56. Stroh JG, Petucci CJ, Brecker SJ, Huang N, Lau JM. *Automated Sub-ppm Mass Accuracy on an ESI-TOF for Use with Drug Discovery Compound Libraries. Journal of the American Society for Mass Spectrometry.* 2007; **18**: 1612-1616.

57. Thurman EM, Ferrer I, Zweigenbaum JA. *High-Resolution and Accurate Mass Analysis of Xenobiotics in Food. Analytical Chemistry.* 2006; **78**: 6703-6708.

58. Potterat O, Wagner K, Haag H. *Liquid chromatography-electrospray time-of-flight mass spectrometry for on-line accurate mass determination and identification of cyclodepsipeptides in a crude extract of the fungus Metarrhizium anisopliae. Journal of Chromatography A.* 2000; **872**: 85-90.

59. Chernushevich IV, Ens W, Standing KG. *In Electrospray Ionization Mass Spectrometry: Fundamentals, Instrumentation & Applications*, New York, 1997.

2. ANALYSIS OF HISTORICAL ACCURATE MASS DATA OVER A RANGE OF IONIZATION AND INSTRUMENTATION

2.1 Introduction to accurate mass measurement

Mass spectrometers are used to separate different mass species and isotopes, and measure the abundance of mass and their isotopes when used as analytical instruments in fields such as chemistry, biology and medicine. All experimental measurements are subject to some errors and uncertainty. This is true for scientific measurements where errors can arise in a number of different ways, and a balanced approach to interpretation of the results is required. Statistical methods can provide an approach that permits us to give a quantitative estimation about the accuracy and precision so that sound conclusions are drawn. In general, any statistical evaluation of quantitative data involves a mathematical model for a theoretical distribution of the experimental measurements that reflects the precision of measurements[1]. The distribution of experimental data can also be compared by statistical means to the theoretical model distribution.

Mass spectrometry is a very sensitive and powerful analytical technique in which ionized sample molecules are presented, according to the value of their mass-to-charge ratios (*m/z*) by using the application of electric or magnetic fields. The measured *m/z* ratio depends on the chemical composition of the molecules and the number and sign of the charges physical properties. The *m/z* pattern then formed reflects the structure of the intact compound. Mass-to-charge ratio of the ions present will be plotted against their abundance and a mass spectrum obtained. Accurate mass measurement has been accepted as a method for confirmation of the identity of substances and for identification of the elemental composition of specific ions[2]. Confident elemental composition assignment is essential for the characterisation of small molecules as in drugs design and for natural products[3]. For the identification of unknowns it is important to get the best mass accuracy available, because the number of possible elemental compositions increases dramatically with increased uncertainty in the mass accuracy. Most commercially-available mass spectrometers achieve best mass accuracy (< 3 ppm) only with internal calibration. Accurate mass measurement of compounds is usually based on the comparison of the instrumental response for the unknown substance with an appropriate calibrant. Modern accurate mass measurements performed using high resolution instrument such as TOF, FT-ICR mass spectrometers, often utilise an automated programme to generate a list of elemental formulae.

2.1.1 Accuracy and precision of different types of mass spectrometers

Accurate mass measurements (AMM) are frequently used to determine the elemental composition of molecules and fragment ions. With the advent of high resolution mass spectrometers, accurate mass measurement at low part per million (ppm) by LC/MS analysis can be achieved and allows for the possible elemental composition determination of unknown compounds.

Figure 2.1 defines accuracy and precision, by analogy to the grouping of arrows in a target (how close measured arrows are to the target centre, and precision depend on how closely two or more arrows agree with one another). Accurate and precise mass measurement will increase the certainty of identification of molecular ion formula. Each type of mass analyser has unique parameters that will contribute to the overall uncertainty of accuracy of mass measurement. It is important to understand the degree of accuracy that is required for the *m/z* of the ion that is measured; with increasing mass-to-charge ratios (*m/z*) the number of formulae under test will increase until the clear result becomes difficult to obtain.

Accurate mass measurements on low-molecular weight organic molecules have been performed using high-resolution double-focusing magnetic sector instruments. Fourier transforms ion cyclotron resonance FT-ICR, and time-of-flight (TOF) instruments are also used for this application. Magnetic sector instruments can achieve a resolution of about 100,000 (using the 10% valley definition), while an FT-ICR mass spectrometer

is capable of achieving a resolution of ≥ 200,000 (full-width at maximum height, FWHM) for electrospray[4].

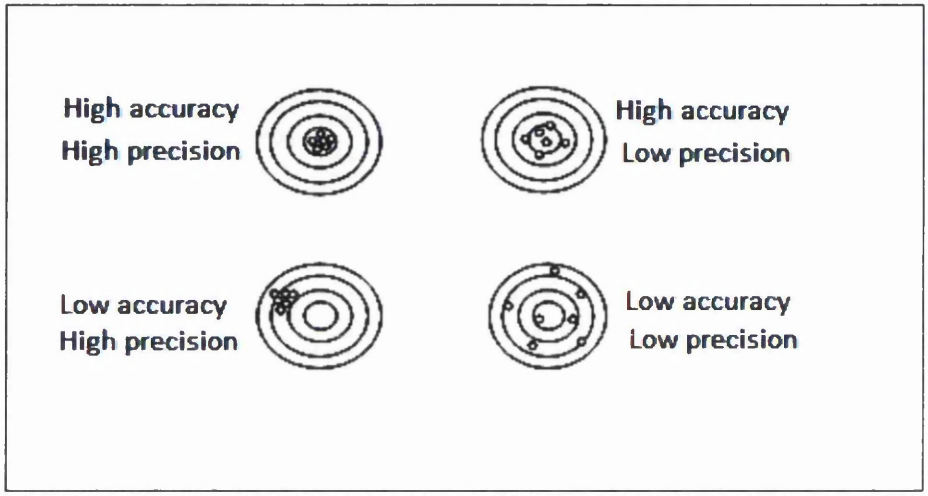

Figure 2. 1 Schematic representation of accuracy and precision levels of measurements.

Repeated measurements of the accurate mass of a molecular ion lead to an improvement in accuracy by calculating the arithmetic mean of the readings obtained. The mean should be reduced with increasing the number of readings, it will never be zero, but will reduce as ($\frac{1}{\sqrt{n}}$).

2.1.2 Calibration of the mass scale

In a mass spectrometer, ions are moved from one location to another as a beam. This beam has a tendency to shift, but it can be controlled by using lenses that are fixed at various stages, including the entrance and the exit of the mass analyser. As mentioned

in the introduction, mass spectrometry does not measure mass directly. For example, a sector mass spectrometer measures the magnetic field strength[5]; time-of-flight, the time an ion takes to fly between two locations; a quadruple ion trap measures mass by the stability of an ion's trajectory; and a Fourier-transform mass spectrometer measures the frequency of an ion in an ion trap. The mass is derived by measurement of the above physical parameters which are converted into mass using appropriate mathematical formulae. There will be a range on the achievable mass scale, with a lower limit and an upper limit. These will depend on the physical parameters of the instrument or can be chosen, or a mass scale may in fact be confined by the ionisation technique. Thus there is a need to connect physical measurements to a mass scale.

Once a mass scale is defined, we need to consider the accuracy and precision of that mass scale. There are two types of mass scale: nominal or accurate. A nominal mass scale is one where the mass is measured to the nearest integer, in practice the measured mass is usually recorded to the first decimal place. An accurate mass scale is where the mass can be determined to higher accuracy generally three or four or even more decimal places. The accurate mass scale can allow the elemental formula of an ion to be determined. If there is sufficient mass resolution, then high resolution accurate mass can reveal rich structure along the mass scale, for example, in the analysis of complex samples such as a crude oil extract. There may be dozens of chemical species at each nominal mass along the whole mass scale. Determining the "mass scale" is an essential part of mass spectrometry and requires calibration. In

order to calibrate the mass scale, two methods are normally used: external and internal calibration.

2.1.2.1 External calibration of the mass scale

External calibration is when a calibration standard is acquired on the mass spectrometer prior to the analysis of a sample. For many techniques an external calibration may be run once a week, once a month or once every six months[6]. The longer the time between calibrations, the less accurate the mass scale becomes, and for many techniques this method is used to establish a nominal mass scale. For accurate mass measurement external calibration can be used but is often carried out immediately prior to the mass measurement. Depending on the technique an external calibration will have a lifetime after which the mass scale will have drifted sufficiently to render it inaccurate. For time-of-flight technology the thermal expansion of the flight tube severely affects the mass scale and can be as much as 100 ppm per degree centigrade (°C) change in temperature. In this case, an external mass calibration may only last minutes. On the other hand, Fourier-transform ICR produces a stable accurate mass scale over a long period of time. This is due to the very high stability of the super conducting magnetic field (when a super conducting magnet is employed). For an FTICR, the magnetic mass scale can persist over weeks or even months because the instrument relies on measuring an ion's frequency; and measurement of frequency is inherently very precise, as modern electronics can measure such fundamental physical quantities to a very high accuracy.

2.1.2.2 Internal calibration of the mass scale

Internal calibration occurs when a known amount of a reference compound, different from the target analyte, is added to the unknown target analyte. Usually a reference (standard) molecule is chosen as close as possible in chemical and physical properties to that of the molecule that will be measured.

Internal calibration is defined when the internal standard is present at the same time as the analyte and therefore the mass spectrum contains both analyte ions and calibration ions. The mass spectrometer measures the mass of both the reference ions and analyte ions together, and as the mass of the reference is well established the computer then corrects the mass scale for each individual spectrum recorded. This methodology is intrinsically more robust and is the preferred method for high accuracy mass measurement, particularly for techniques where drift of the mass scale is high, for example in time-of-flight instruments.

2.1.3 Calibration standards

These are selected chemical compounds which produce useful and numerous reference ions over the chosen mass scale of interest. This will depend on the application and will depend crucially on the ionisation technique chosen and ionisation mode (positive or negative). Historically, mass calibration standards were chosen for electron ionization in positive ionisation mode, with early experiments standard in the mid 1940s when oil samples were first being analysed. In the 1950s and 1960s when organic mass spectrometry was becoming a mature technology,

electron ionization was the dominant ionisation technique of the few ionisation techniques that existed at the time. It was only in the mid-1960s and early 1970s that other ionisation techniques were developed. It is useful to discuss calibration standards therefore in terms of ionisation techniques and this will be done in the subsequent section.

2.1.3.1 Electron ionisation (EI) standards

A series of fluorinated organic compounds were originally synthesised for use in calibration of mass spectrometers, the most commonly used calibration standards for positive EI-MS are perfluorinated compounds[7]. These compounds, which consist of a mixture of hydrocarbons, are relatively volatile and give rise to several prominent peaks across the *m/z* range, that have exact masses almost equal to integral values in range *m/z* 50 to higher than 650 that are useful for calibration.

The most common references used with EI are perfluorokerosene (PFK) and perfluorotributylamine (PFTBA)[8,9]. PFK is useful in both positive and negative ion modes and has numerous peaks, often as closely spaced as 31 mass units; PFK has intense ions at low mass, but falls off rapidly above a few hundred mass units and is very weak at $\geq$ 1000 mass units. PFTBA has higher sensitively above a few hundred mass units with notable ions at 502 and 614 and is commonly used in EI with a mass range of up to 650. Fishman, et al.[8] developed a new fluorinated compound mixture for calibration for double-focusing mass spectrometers in the mass range 100-3000 Da in positive electron ionisation (EI) mode, which consist of a series of seven

fluorinated organic compounds (fluorinated silyl alkyl amines mixture) and have a mass range from 2237- 2417 Da. Wellington laboratories[10] have also introduced an additional compound for the fluorinated compounds. They introduced of mass-labelled perfluoro-1-octanesulfonate (PFOS) reference standard, sodium perfluoro-1-[$^{13}C_8$] octanesulfonate (M8PFOS) to provide their laboratories with a better selection of PFOS surrogates for use in their analysis. M8PFOS used as extraction standard and MPFOS as a recovery or clean up standard.

Fluorinated standards are extremely useful as the fluorine atom has a negative mass defect, *i.e.* its nominal mass is 19, but its accurate mass is 18.9984032. Thus calibration peaks will appear at slightly lower masses than the nominal mass, when compared to the organic sample being measured, which is hydrogen rich and therefore appears at a mass slightly higher than the nominal mass[11]. Therefore fluorinated standards are easily separated from the ions of organic compounds and distinct from the sample at high mass resolution. This is particularly useful for the analysis of complex sample mixtures which have a large number of components at each nominal mass. The EI spectra of these compounds show a series of intense singly charged ion peaks with a good distribution over the entire mass range[8].

2.1.3.2 Chemical ionisation (CI) standards

There are two common implementations of chemical ionisation (CI) sources which differ in the pressure at which they are operated. The source pressure depends on the type of instruments, for example, lower pressure is used with instruments designed for

electron ionization ionisation, whereas another type is used with instruments designed for atmospheric pressure ionisation (API) in conjunction with LC[7]. As mentioned before, the first type uses PFK to calibrate the mass scale in the instrument, as it generally has EI facility. There are not so many well-known specialist standards for CI. However a common standard that is used for CI GC/MS has been fatty acid methyl esters (FAMES), also for EI and CI commonly the following calibration references; PFTBA or HTBA in range 0-600Da and PFK in range up to 900 Da are used to calibrate the mass scale and Ultramark can be used for high resolution[12,13].

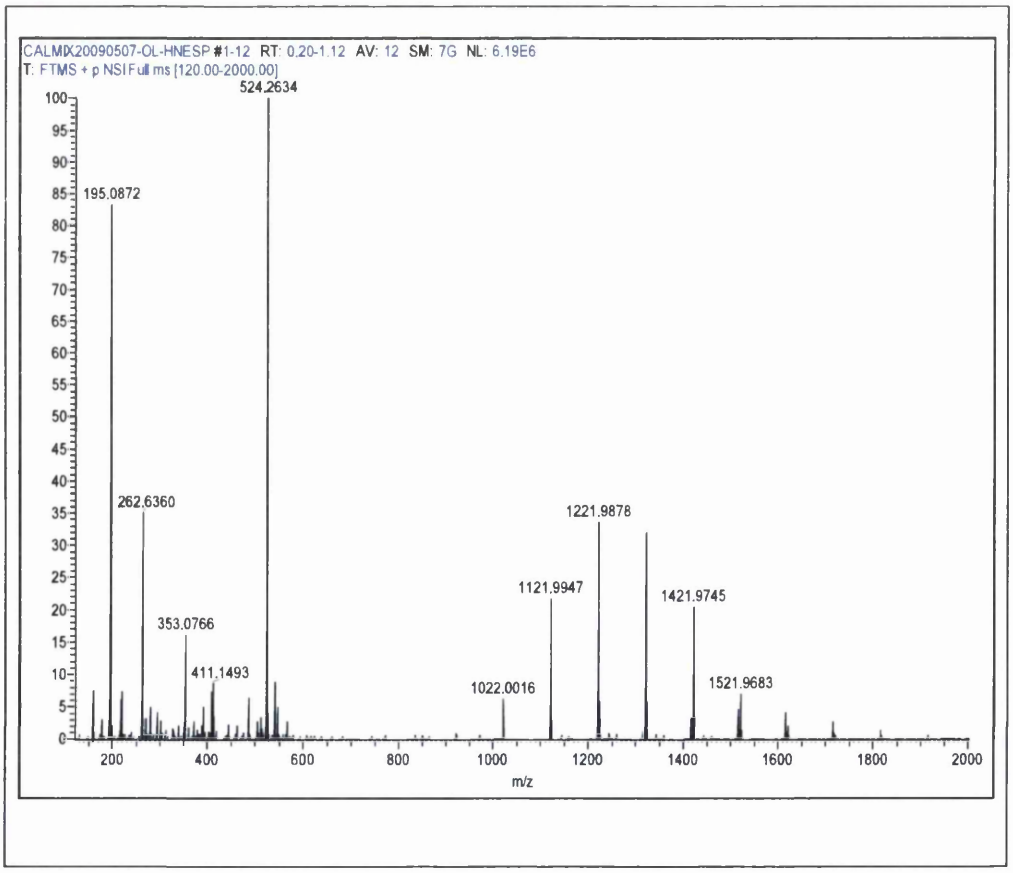

Figure 2. 2 Orbitrap CalMix mass spectrum used as an accurate mass standard in positive ESI.[14].

2.1.3.3 Electrospray ionisation (ESI) standards

A mass spectrometer can be calibrated for electrospray ionisation (ESI) by using a number of different techniques. Calibration is sometimes achieved using full scan mass spectra (MS mode) or collision-induced fragmentation (CID) spectra (MS/MS mode). There is a number of calibration standards used depending on the application and the mass range required. For ESI mass measurements, cesium tridecafluoroheptanone, (up to 10,000 Da), poly (ethylene glycol) (PEG), poly (ethylene glycol monomethyl ether) (PEGMME) and polypropylene glycol (PPG) (50-2,000 Da) are common calibrants used with ESI[13]. In proteomics one might use protein standards such as Angiotensin I (1295.6775 Da), Angiotensin II (1045.5345 Da), Ubiquitin (8565.87) and BSA (66429.8 Da)15. Thermofisher scientific have a very useful set of calibration mixtures. One covers the mass range 50 to 2,000 Da and it contains caffeine (195.0872), a peptide with the sequence MRFA (main isotope 524.265) and a compound called Ultramark (1022-1822 Da).

In ESI negative ion mode, the usual reference compound is sodium trifluoroacetate (NaTFA) which is injected in conjunction with the analyte; NaTFA is encouraged to form cluster ions to cover the required mass range. Here is an example for some typical ions formed by the ESI source;

Positive ion mode: $[M+H]^+$, $[M+nH]^{n+}$, $[M+NH_4]^+$, $[M+Na]^+$; $[2M+H]^+$, $[2M+Na]^+$.

Negative ion mode: $[M-H]^-$, $[M+Cl]^-$.

2.1.3.4 Liquid secondary ionisation mass spectrometry (LSIMS)

Fast-atom bombardment (FAB), liquid secondary ion MS (LSIMS) is usually performed on older instruments, such as magnetic sector (MAT95 or MAT900) in both negative and positive ionisation mode with a mass range up to 3,000 Da. The determination of accurate mass below 300 Da is not commonly done by FAB, for it is usually employed as a high mass technique. The chemical composition and accurate masses of some commonly used reference compounds were reported. Cesium iodide (CsI) with a calibrating range (130-30,000 Da), provides suitable negative ions, but it's not generally used for high resolution accurate mass measurement because the peaks are spaced quite far apart[12,16] and may not be the ideal standard when using sector mass analysers. Negatively charged glycerol matrix peaks if close enough to the unknown peak are frequently used as a peak matching standards in the low mass range (< 1000 Da). Nitrobenzyl alcohol (NOBA) sodium and potassium salts are sometimes added and glycerol gives peaks with mass spacings of 92 or additions of 92. Ultramark is used as a useful calibrant for exact mass measurements in the range 700-1,900 Da for positive and negative ion fast-atom bombardment high resolution mass spectrometry.

2.1.3.5 Matrix-assisted laser desorption/ionisation (MALDI)

Several reference compounds are used as reference mass markers, for example, α-CHCA matrix Angiotensin I[13] and Angiotensin II. Most measurements of binding

affinity of albumin for long chain fatty acids are based on heptane-water partition. For negative ionisation mode, oligonucleotides and oligosaccharides can be used.

2.1.4 Protocol for mass measurement

In order to achieve high accuracy and precision, we have to know a number of key points before starting: tuning, peak shape, ion abundance, resolving power and calibration. For sector instrument, resolving power is sometimes defined in terms of the overlap (or 'valley') between two peaks. Thus for two peaks of equal height, masses m_1 and m_2, when there is overlap between the two peaks to a stated percentage of either peak height (10% is recommended), then the resolving power is defined as $m_1/(m_1-m_2)$. The percentage overlap (or 'valley') concerned must always be stated. We can distinguish mass 50 from the next integer mass which is 51 or distinguish mass 1,000 from 1,001; its definition is commonly used when quadrupole or ion trap mass spectrometers are used. Magnetic sector mass spectrometer defines resolution as: $R = m/\Delta m$ where m are the mass number that is observed and Δm the difference between two close masses and the shape of peak in magnetic sector roughly gives Gaussian peak. Figure 2.3 below shows that peaks are usually defined to be separated to 5% or 10% valleys.

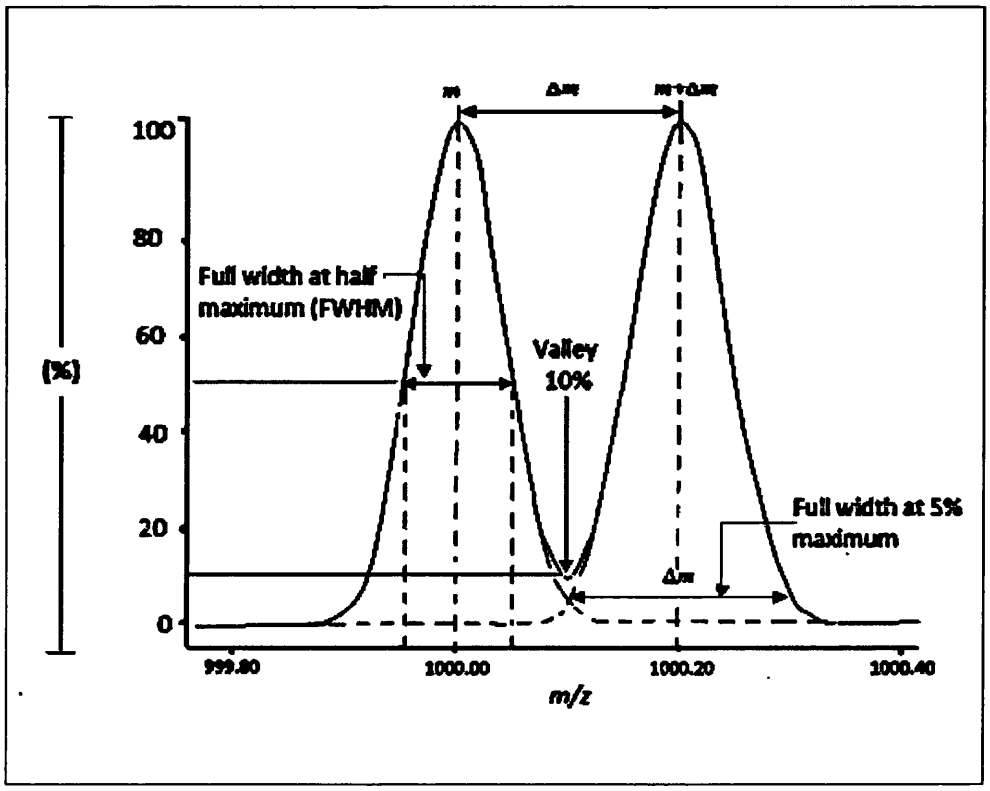

Figure 2. 3 *Intensity definition of resolving power (R.P) the peak resolved 5%, resolved 10% in the middle*[24].

2.1.5 Magnetic sector mass spectrometry

There are a variety of possible scanning methods used in mass spectrometry. The most commonly used to obtain accurate mass measurement on a magnetic sector are peak matching, narrow mass range and magnetic scanning.

2.1.5.1 Peak matching

Peak matching is one of the most accurate methods used to assign the mass of an ion to a value typically a few ppm, but sometimes <1 ppm from the calculated exact mass. This technique is usually performed on a high resolution magnetic based instrument,

but instead of adjusting the magnetic field strength it involves rapidly switching the accelerating voltage, V, and the electric sector voltage, E. Peak matching compares the mass of the unknown analyte to reference ion mass. The process involves matching their respective peak shapes so that at a given ratio of accelerating voltage these peaks overlap, *i.e.* peak match on a visual display unit or computer screen. Software is now used to measure the voltage ratio for "peak matching". From this ratio, the unknown mass can be calculated as follows. If m1 is the unknown mass, V_1 the accelerating voltage to transmit m_1, m_2 the reference mass, V_2 accelerating voltage to transmit m_2 where:

$$m_1 = m_2 * \frac{V_1}{V_2} \qquad (2.1)$$

$$m_1 / m_2 = \frac{V_1}{V_2} \qquad (2.2)$$

Instruments are operated at high resolution mode (typically $m/\Delta m \geq 10000$) for peak matching. The mass to be measured should be close to the reference ion mass. The peak matching voltages are controlled by a very accurate electrical circuit which allows the ratio of the unknown to the known mass to be determined to very high accuracy, typically parts per million. Traditionally the peak matching voltages were switched manually by using a resistive circuit, comprising very accurate resistors which are switched to create different reference voltages. Nowadays, a computer program does this automatically and also controls other circuitry to shape the peaks,

compare the peaks and adjust the sensitivity so that a very accurate match is made. For electron ionization PFK might be used. For example, an analyte at mass *m/z* 386 (cholesterol) can be compared to any of the reference ions in PFK at 369, 381, 400, 393, 419.

2.1.5.2 Narrow voltage scan

In the narrow voltage scanning technique (known as accelerating voltage scan or dynamic voltage scan)[17], a very narrow mass range is accessed by varying the accelerating voltage (V) and electric sector voltage (E) together, while the magnetic field is held constant at the *m/z* of the unknown or reference ion. The ratio between the accelerating voltage and the electric sector voltage V/E must be constant to ensure the ions are transmitted efficiently. While the magnetic field is constant, the narrow voltage scan and the electric sector voltage are changed automatically under computer control. This method was used by Bristow et al. in their studies to measure the accurate mass in different laboratories[18]. This can be very accurate, and is usually used over 0.1% of the mass scale.

2.1.5.3 Magnet scanning

In this technique the accelerating voltage V and electric sector voltage E are kept constant, the ratio V/E is kept in constant. The magnetic sector is scanned over a chosen *m/z* range. This method is used to determine all ions in the selected mass range and a mass spectrum is obtained.

2.1.5.4 Full scan– Orbitrap mass spectra

The Orbitrap operates on the principle of injecting ions into storage trap of radio frequency (RF) electric field (no magnetic field). It is based on a new type of mass analyser invented by Makarov[19]. The ions are detected with a resolving power of more than 60,000 (FWHM). It employs a Fourier transform of the time domain trapping signal and the spectrum is similar to the method of Fourier transform ion cyclotron resonance mass spectrometer (FT-ICR). The LTQ-Orbitrap instrument consists of three independent ion trapping devices, to work as frequencies ion traps (LTQ & c-trap) whereas the Orbitrap is a coaxial spindle-like electrode that forms the electrostatic ion trap. The ions cycle around the central electrode with circular motion and move back and forth along the z-axis of the central electrode. By sensing the ion oscillation in z-axis, a time signal is built up and the mass spectrum is obtained similar to FT-ICR. The Orbitrap has a high resolving power of up to 100,000 and a high mass accuracy (1-2 ppm) (< 2 ppm using an internal standards and < 5 ppm) with external calibration over a reasonable range of intensities[20]. However, mass accuracy is based on the stability of the electric field, which is easier to achieve than a stable magnetic field and results in excellent stability of the device with minimal requirement to recalibrate often. The Orbitrap mass analyzer is usually fitted with electrospray ionisation (ESI) source. Ions are transferred from the ESI source through three stages of differential pumping using RF guides quadrupole lenses. Orbitrap mass spectrometry has three mass ranges: low mass range 15-200 for LTQ ion trap analyzer only, "normal" 50-2,000 and "high" 100-4,000 for ion trap and FTMS. The mass

resolution is chosen for the several options available; resolution settings of 7,500, 15,000, 30,000, 60,000, and 100,000 are available. Usually the LTQ Orbitrap is externally calibrated each day using caffeine, MRFA (meta-arg-phe-ala), and Ultramark 1621 in positive ionisation mode. For negative ionisation mode the instrument is externally calibrated using caffeine, MRFA, Ultramark 1621, sodium dodecyl sulphate and sodium taurocholate. The LTQ Orbitrap XL also has the ability to apply internal calibration using one or more ions (lock mass).

2.2 Statistical principles for accurate mass measurements

Experimentally-measured masses inevitably carry errors or uncertainties, for any cause may occur together with the analytical process. The total error of general measurement is often given as either a relative or absolute value, and the units scale used for these measurements are parts per million (ppm) scales for relative errors and milli Dalton (mDa) for absolute errors. The errors in mass measurement compute the difference between theoretical and observed masses, which should be close to zero, but because there might be some systematic errors, the difference sometimes will be quite biased from zero. The analyte and reference peaks must have sufficient intensity and be free of interference from other masses. Hence to evaluate the capability of instrument mass measurement accuracy, one should calculate the magnitude of the mass error e.g. the root mean square (RMS) error.

The mass measurement error of a single reading can be calculated by defining the theoretical mass (ma) and measured (observed) mass (mi) by using either of these equations:

$$\Delta m_i = m_i - m_a \qquad \text{In Da} \quad (2.3)$$

$$\Delta m_i = \frac{m_i - m_a}{m_a} \times 10^6 \qquad \text{In ppm} \quad (2.4)$$

Accuracy in mass spectrometry refers to how close the mass measurements are to the expected results (m_a). The average mass is used to assess the accuracy and is calculated as shown in equation 2.5, where N is the number of measurements. Average masses of n measurement.

$$\overline{m_i} = \frac{\sum_i (m_i)}{n} \qquad (2.5)$$

The average mass error (AME) is given by

$$\overline{\Delta m_i} = \frac{\sum_i (\Delta m_i)}{n} \qquad (2.6)$$

There are several ways in which the magnitude of mass errors is quoted; the statistical mass accuracy is estimated from statistical distribution of mass error. For example, the root mean square (RMS) deviation of a particular mass that may be reported by a manufacturer as a measure for the instrument accuracy. If the

experimental data form a Normal distribution then the standard deviation can be quoted and its meaning is well understood. However the term "average mass error" (AME) is commonly used. As the average mass error may be confused with mass measurement accuracy, it is better to state if absolute values of errors are used to create an absolute value of average mass error. The literature cites the terms "average mass error (absolute value)" (AAME) or absolute exact mass errors (AEME) or average absolute measurement accuracy (aaMMA).

The average absolute mass accuracy is a parameter that is easily calculated, and one found to be quite informative of quality of measurement. The difficulty with statistical mass accuracy may occur if the mass error does not follow the bell shape, as expected, for a Normal distribution. This will affect data dispersion and result in inaccurate value which may give other kind of distribution. For example if the data follow a Lorentzian distribution instead of a Normal distribution, the dispersion of the mass error may be wider and will have very extended wings. If any single measurement obtained was inaccurate, the statistical error will be infinite depend on the outcome of $\sigma/ (n-1)^{1/2}$.

The average absolute mass measurement accuracy (aaMMA) can be calculated by:

$$aaMMA = \frac{\sum (|m_i|)}{n} \qquad (2.7)$$

and the root mean square mass error:

$$RMS = \sqrt{\frac{\sum_i (\Delta m_i)^2}{n}} \qquad (2.8)$$

Although standard deviation gives a measure of the spread of individual results about the mean value, it does not indicate the shape of the distribution. The distribution of repeated measurement can be drawn as a histogram, and should form a "Normal distribution". If there are no systematic errors, then the mean of the all measurements (population), referred to as μ, is the true value. The mathematical model usually used is the Normal or Gaussian distribution[21] which is shown by equation 2.9.

$$y = \frac{1}{\sigma\sqrt{2\pi}} e\left\{-(x-\mu)^2 / 2\sigma^2\right\} \qquad (2.9)$$

Figure 2.4 shows the distribution of large numbers of data. The shape is symmetrical about the mean value μ and the greater the value of σ, the greater the value of the spread. Standard deviation is a useful measure of precision for analytical measurements and is given by this equation (2.10):

$$SD = \sqrt{\frac{\sum_i \left(\Delta m_i - \overline{\Delta m_i}\right)^2}{n-1}} \qquad (2.10)$$

Precision refers to the distribution or spread of individual results, and when plotted as a histogram should form a Normal distribution. It is a measure of the dispersion of a population of data set or probability of distribution. A low standard deviation for a data set indicates that the data points tend to be very close together therefore giving

high precision. A large standard deviation indicates that the data are spread out over a wider range of values, consequently low precision is obtained. Standard deviation is also used to measure the confidence interval for statistical conclusions. The mean of the Normal distribution is often referred to as mu (μ) and sigma (σ) for the standard deviation. The mean can assume negative or positive values while the SD always has a positive value.

The central limit theorem in Normal distribution tells us that the sums of random variables are approximately normally distributed if the numbers of observations are large. For measurements which have only random errors, the standard deviation means that 68% of the measured values are within 1σ from the mean, 95% are within 2σ of the mean and 99.7% are within 3σ from the mean[21]. As shown in figure 2.4, the exact proportion of values lie within any interval (*i.e.* ±1σ, ±2σ, ±3σ) and can be found from statistical tables, provided that the values are first standardised so as to give the values of standard normal variable

$$z = (x - \mu)/\sigma \qquad (2.11)$$

Where x is the reported value obtained, μ is the mean value and σ is standard deviation. Z-Scores are a simple method that gives each participant a normal performance score for bias from the mean; this test was adopted as a standard for ISO/IUPAC.[22].

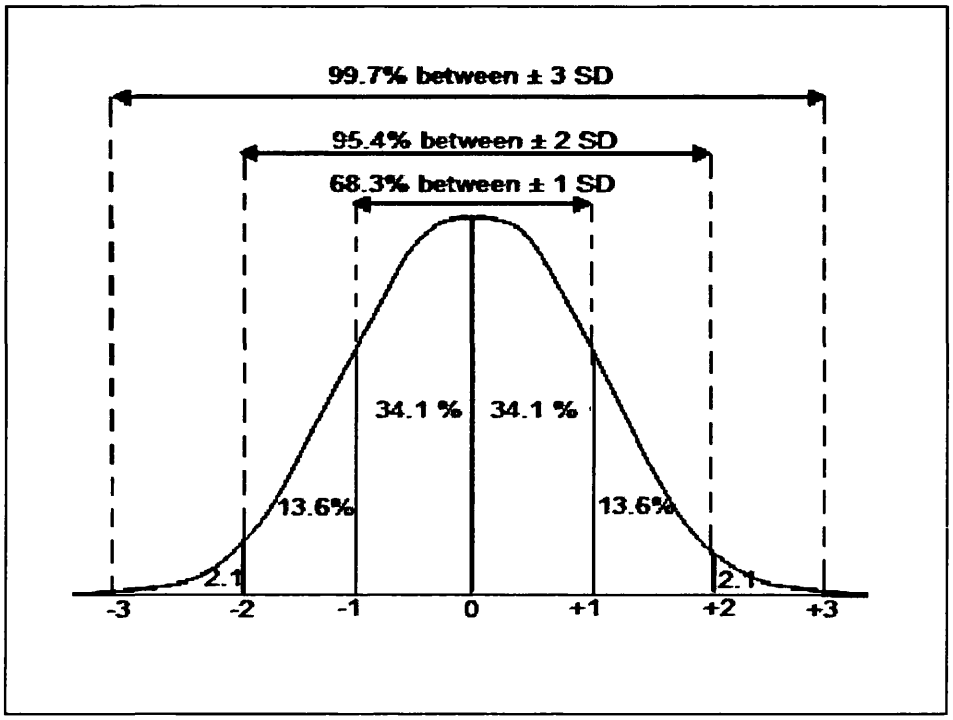

Figure 2. 4 *Properties of the Normal distribution (spread of standard deviation in both side of the mean: (i) approximately 68.2% of values lie within ±1σ of the mean, (ii) approximately 95.4% lie within ±2σ of the mean; (iii) about 99.7% lie within ±3σ of the mean.*

Confidence limits for data sets give a range within which the data set is expected to lie. We usually used 95% confidence interval of the mean

$$\Delta m = m_a \pm t\left(\frac{s}{\sqrt{n}}\right) \qquad (2.\ 12)$$

Such, t =1.96 when data set more than 60 (p=0.05), for n = ∞ at 95% confidence level, and presented the standard error of the mean that measure the closeness of the sample mean to the population mean. Spread of data can be measured graphically by using the Box and Whisker plot (figure 2.5) to indicate the central tendency and the spread

of the data. It may be drawn vertically or in a horizontal shape. The tail (end) of the graph indicates the limit of the data, while the Box indicates the data cumulated in range of 2σ, and the central vertical line is the mean. Sometimes the central line indicates the median and the Box encloses the upper and lower quartiles.

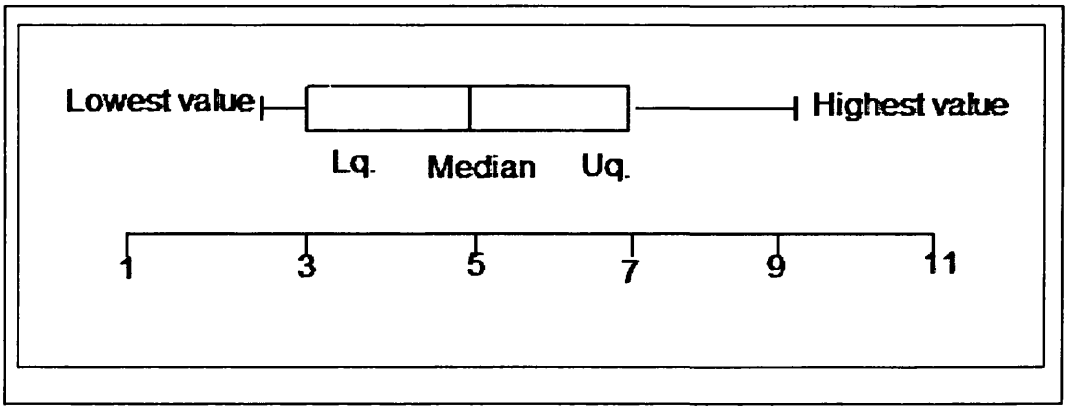

Figure 2. 5 *Simple plan for Box and Whisker plot on a numerical scale as well as a graphical presentation of the five -number summary (where Lq: lower quartile, Uq; Upper quartile) Whisker plot as an example for data set measurements graphically.*

2.3 Issues in quoting accurate mass data

The comparison of accurate mass measurements for small molecule analyses (molecular weight, 100-1400 Da) across a broad range of mass spectrometers was carried out. Each mass spectrometer has different capabilities and optimum methodology to record accurate mass measurements.

In a recent study, Muddiman, et al.[23] carried out a comparison between a double-focusing magnetic sector and a TOF analyser in obtaining accurate mass measurements. They reported that the TOF analyser provided better mass accuracy

than the magnetic sector across the mass range of 200- 1,300 Da, with an average absolute MMA value of 1.1 ppm compared with 3 ppm for the magnetic sector. Their conclusions however, seem rather misleading as their comparison appears to be incomplete. Not only have they compared two mass analysers, but they have also compared two different ionization sources, where they had used a FAB ionisation source with the sector instrument and an ESI source with the TOF analyzer. In addition to that, they acquired the data differently, *i.e.* used internal calibration for the TOF data while using an external calibrant for the FAB data. The comparison seems, therefore, statistically biased in favour of the TOF analyser even though the results did not appear to support these conclusions. At the end of their report, they focused on the high throughput of the TOF instrument in comparison to the magnetic sector, while concluding that the mass accuracies of the two analysers were comparable.

Muddiman's study stimulated our interest as it gave results differing from other published papers, as shown by Bristow et al. investigation[24]. The mass spectrometry unit already has historical data which exceed 4,500 accurate mass measurements recorded in EPSRC- National Mass Spectrometry Service Centre at Swansea University[14]. The data collected over a heterogeneous set of chemical samples were donated for this study. We have analysed these historical data sets, applying a wide range of statistical methods to determine accuracy and precision and the nature of the histogram distributions for these data, over a range of ionisation methods and instrumentation.

The data was obtained with consideration of the accuracy of measurement, how well the position of the peak representing the sample ions can be determined on the mass scale. This depends on the number of ions in the peak, and the method used to define the peak position. The other consideration is the presence of interfering peaks at the same nominal mass as the analyte ion. Mass measurement was performed using instruments capable of high resolution, typically magnetic sector, Orbitrap and TOF instruments. Calibrant reference ions were chosen to cover the range of samples and internal calibration was applied for the magnetic sector while external calibration was applied for the Orbitrap. This provides the mass of ions consistently when determining the *m/z* values of samples and measured as unknowns.

2.4 Overview of historical sets of accurate mass data over a range of instruments and ionisation technique

The basic principle to achieve a good analytical measurement is that the measurement should satisfy certain requirements of the analyst. In mass spectrometry reliable mass measurements are subjected to standard criteria that are used in most good laboratories. Methodology, resolution, accuracy and precision are the most important factors to produce reliable data. In a mass spectrum, we can see noise arises as random contributions and often due to various causes not under the control of the analyst, and leads to random errors which can be analysed by standard statistical methods. If systematic errors are involved then other methodology is required. Examination of large numbers of mass peaks obtained after calibrating the mass scale

gives an idea of the different errors we can expect. The determination of molecular mass tells us that the measured mass should be close or equal to the exact mass; the difference gives us a perspective of mass accuracy and mass precision. Comparing the theoretical mass (expected mass) with measured mass (observed) provides us with the accuracy of our techniques. To determine mass accuracy for any mass spectrometric technique, sufficient numbers of mass measurements should be collected, and the mean calculated and compared to the theoretical means to assess accuracy. In this chapter historical data sets have been analysed by applying a wide range of statistical methods to determine accuracy and precision and the nature of the histogram distributions for this data, over a range of ionisation methods and instrumentation. All data set were obtained by the EPSRC National Mass Spectrometry Service Centre[25] at Swansea University which offers a national accurate mass service.

These data records are in excess of 4,500 accurate mass measurements, over a heterogeneous set of chemical samples, and were donated for analysis in this study. The data was obtained using a number of ionisation techniques on sector mass spectrometers, time-of-flight (TOF) and Orbitrap. The main accurate mass instruments used in the EPSRC NMSSC are: i) MAT900 double focusing instrument, ii) MAT95 double-focusing instrument and iii) Orbitrap. The Finnigan MAT900 XLT high resolution double-focusing mass spectrometer (Thermo Fisher Scientific Gmbh, Bremen, Germany) has a mass range of 5,000 Daltons at full accelerating voltage; maximum resolution 60,000 (10% valley definition), and the MAT95XP high resolution double-focusing mass spectrometer has a capability of mass range 3,500

Daltons at full accelerating voltage; maximum resolution 60,000 (10% valley definition). Both instruments are fitted with electrospray (ESI), electron ionization (EI), chemical ionization (CI) and liquid SIMS (LSIMS or FAB) sources, with accurate mass measurements made by peak matching. Full-scan accurate mass measurements were also obtained for matrix-assisted laser desorption/ionisation (MALDI) on a Voyager STR (Applied Biosystems, Framingham, USA) and ESI using an LTQ-Orbitrap (Thermo Fisher Scientific). However, data that was obtained in positive ion mode for each instrument is a large enough sample size ($n > 57$) to evaluate the overall performance of all the instruments. A range of accurate mass standards have been used with these instruments (~200-1,400 Da), depending on ionisation technique and polarity. For peak matching and MALDI, an internal standard was always employed, whereas the Orbitrap generally employed an external calibration, as its calibration holds over a long period.

Generally, data interpretation has become more complex. The combination of complexity of techniques and application sometimes lead to poor quality mass spectrometric data and may lead to wrong conclusions being recorded. Standard protocols required all instruments to be subjected to daily calibration processes, using internal calibration for magnetic sector and external calibration with Orbitrap. The mass spectra of the calibration standards were acquired first and used for creating the calibration file using the mass spectrometer operating software. The accurate mass errors obtained were less than ±0.3 mDa.

Statistical tools were applied to identify whether each data set was parametric or non-parametric using methodologies recommended by the Analytical Methods Committee of the Royal Society of Chemistry[26-28]. Accmass (in-house software), SPSS and Excel add-ins were used. All data were subjected to the most sensitive effective methods for outlier detection. Dixon's test was applied for small sample sizes[29,30] (data less than 25), while Grubbs and Rosner's tests were used for large data sets[31]. The objective of denoting outliers is to identify observations that are inappropriate for representation in the population, from which the sample is drawn, so that they may be discounted,

Table 2. 1 Descriptive statistics for all data in ppm unit, Such SD Standard deviation, aaMMA: Absolute average mass measurement accuracy, RMS: Root mean square error.

Method	No. of pts	AMM	Mean	aaMMA	RMS	SD(σ_{n-1})
CI+ MAT95	71	352.1216	-0.1079	0.7592	0.928	0.9282
EI+ MAT95	964	402.4126	-0.1527	0.7014	0.871	0.8579
ESI+ MAT95	408	373.0933	0.0745	0.672	0.8083	0.8059
FAB+ MAT95	58	823.3390	-0.032	0.6044	0.7195	0.7251
ESI+ MAT900	2317	394.6231	0.0316	0.7118	0.9079	0.9076
ESI- Orbitrap	65	498.5575	0.1958	0.999	1.2674	1.2619
ESI+ Orbitrap	784	445.4300	0.1399	0.9781	1.2081	1.2007
MALDI+ TOF	195	726.9505	-0.4290	2.7193	3.4259	3.4077

Table 2. 2 Statistical summary result in (ppm) after Outliers were eliminated

Method	No of pts	AMM	Mean	aaMMA	RMS	SD (σ_{n-1})
CI+ MAT95	70	353.4216	-0.0766	0.7372	0.8933	0.8965
EI+ MAT95	961	402.7141	-0.1518	0.6918	0.8449	0.8316
ESI+ MAT95	408	373.0933	0.0745	0.672	0.8083	0.8059
FAB+MAT95	58	823.3390	-0.032	0.6044	0.7195	0.7251
ESI+ MAT900	2313	395.6145	0.0261	0.6942	0.8516	0.8514
ESI- Orbitrap	65	498.5575	0.1958	0.999	1.2674	1.2619
ESI+ Orbitrap	781	445.9416	0.1382	0.9642	1.1814	1.1744
MALDI+ TOF	190	726.9505	-0.3436	2.5279	3.0393	3.0198

eliminated or even recorded in the wrong way. Each data set was carefully proofed for errors and corrected, as minor transcriptional errors in data were found to significantly distort the predicted accuracy and precision of the analysed data. Histogram distribution plots for each data set were analysed to determine if they were "normally" distributed using goodness of fit tests[32,33]. The data were also subjected to some other graphical test like Box and Whisker plot and Q-Q plot. The accuracy and precision of each treated data set was analysed allowing a detailed comparison between different ionisation techniques and instrument types.

Table 2. 3 Shows the descriptive statistical analysis and Kolmogorov Smirnov value (K-S) for all data obtained by different methods on magnetic sector, Orbitrap and TOF mass spectrometry instruments. (The data presented on mDa scale including the outliers).

Method	No. of pts.	AMM	Mean	aaMMA+	RMS±	SD (σn-1)*	K-S
CI+ MAT95	71	352.1216	-0.0521	0.2577	0.3267	0.3268	0.1462
EI+ MAT95	964	402.4126	-0.0442	0.2602	0.3319	0.3292	0.1353
ESI+ MAT95	408	373.0933	0.0218	0.2419	0.3010	0.3006	0.1219
FAB+ MAT95	58	823.3390	-0.0052	0.5259	0.6919	0.6980	0.1607
ESI+ MAT900	2317	394.6231	0.0165	0.2816	0.3913	0.3910	0.0709
ESI- Orbitrap	65	498.5575	-0.0646	0.48	0.6373	0.6390	0.0898
ESI+ Orbitrap	784	445.4300	0.0298	0.4107	0.5414	0.5409	0.1103
MALDI+ TOF	195	728.8283	-0.3344	2.0133	2.6821	2.66112	0.3511

**STDEV: Standard deviation*
+aaMMA: Absolute average mass measurement accuracy
±RMS: Root mean square error, Kol: Kolmogorov exp.

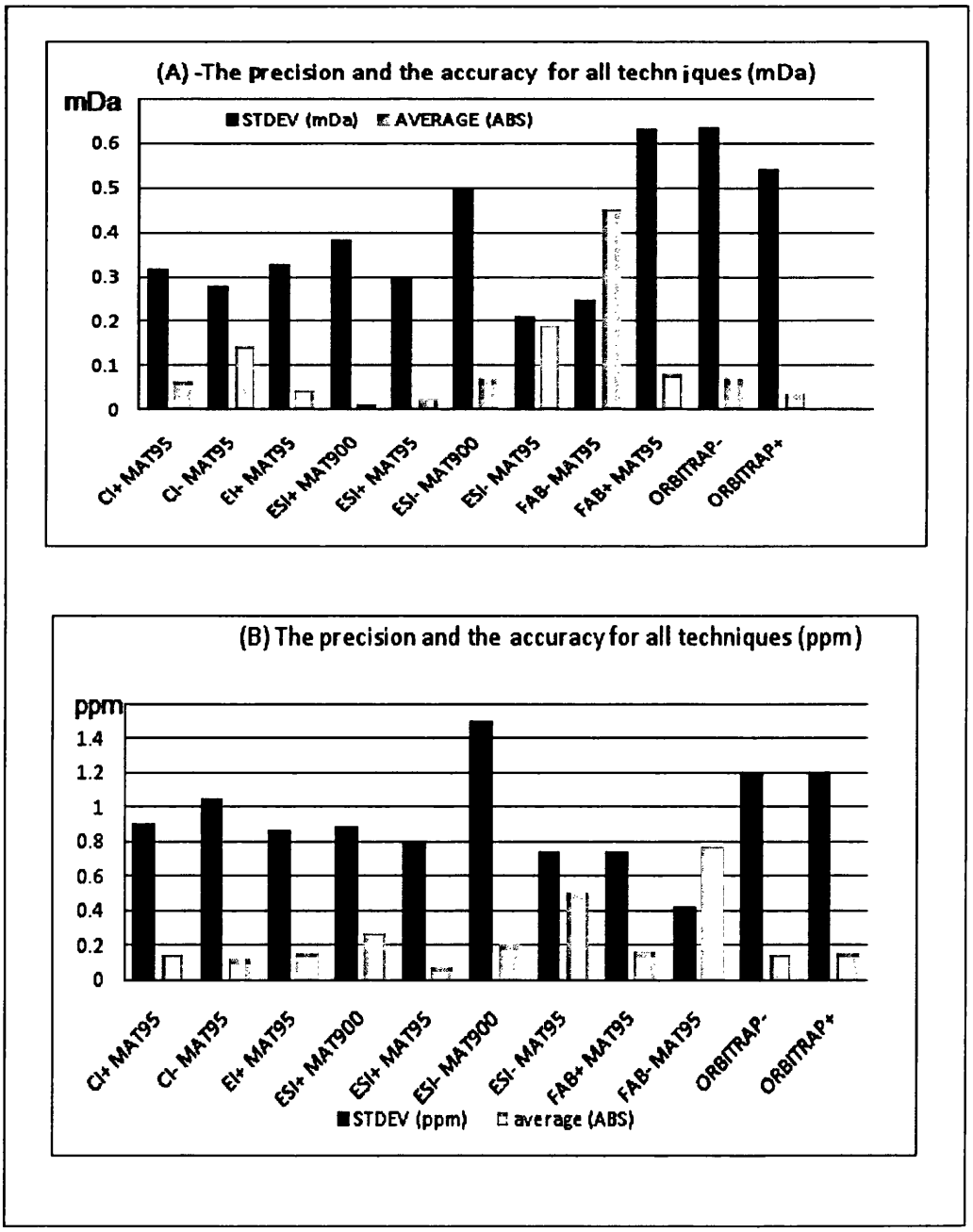

Figure 2. 6 *Two histograms (A,B) illustrating the precision (SD) and the accuracy (the average) in mDa and ppm, respectively. It can be seen that the high accuracy and the high precision is generally very good for these data sets and close to zero. Some methods particularly, when low numbers of data are taken, have larger errors.*

2.4.1 Sector mass spectrometer (MAT900, MAT95)

The mass spectrometry measurements were carried out using double focusing sector field MS MAT900 and MAT95 of the company Thermo Fisher (Bremen). Samples were mass measured and the data presented by measuring the mass-to-charge ratio of the ions produced, the ions presented as signal then detected using system software.

2.4.1.1 Electron ionization (EI) data

The mass measurements were carried out using a Finnigan MAT95 double focusing magnetic sector mass spectrometer, by the direct inlet method with electron ionization ionisation, and produced high quality data. Masses were measured at mass resolution set at 4,000 or greater (10% valley definition), EI source temperature was set at 130°C to 220°C; electron energy 70eV; Sample inlet: Liquid cooled (LC) and desorption probes (DEI) were available for analysis by EI of liquid or solid samples, including insoluble, air and moisture sensitive material. The solid's probe temperature range is 20 to 300°C and was particularly useful for samples of low vapour pressure.

Statistical calculations were made by MS Excel (Microsoft Corporation, Redmond, WA, USA), SPSS software ver. 16.0.1 (SPSS Inc, Chicago, IL, USA) and with R (DAnTE, 2008, Battelle Memorial Institute) and software written in-house (AccMass). The SPSS and AccMass packages allow further statistical tests of the data e.g. normality tests of the histograms as a quality check of the data sets, by Kolmogorov and Smirnov methodology. Mass measurement data recorded in the log

books were based on one measurement (not an average). Only TOF data was an average of measurements. All data for mass measurement of the molecular ($M^{+\bullet}$) or $[M+H]^+$ ion, at masses up to 1,500 Da, were recorded. The difference between exact mass and the observed mass is given as an error in two units; milli Dalton (mDa) and parts per million (ppm). The distribution of data is visualised by histograms, *i.e.* plotting the frequency of the data within mass error intervals (classes) in ppm, throughout the whole data range observed, and histograms generated using AccMass software. Figure 2.8 shows clearly that a large number of the mass errors (964) on EI+ MAT95 give a clear picture of the ppm error distribution around the mean. All ppm errors recorded fall in the interval ±3 standard deviation (SD), the observed range for data presented is (-3.5 to 2.16 ppm). The data has more than ±3SD may be considered as outliers. The calculated accurate mass values were sorted in 0.2 mDa wide bins over the mass range 88-1,300 Da and the distribution of accurate mass error was presented using a graphical test. The descriptive statistical analysis of these data was calculated and is shown in table 2.2. The mean and standard deviation were calculated to be -0.1527 ppm and 0.8579 ppm, respectively.

The absolute average mass measurement accuracy (aaMMA) and root square mean square (RMS) are 0.7014 and 0.871 ppm, respectively. The Normal distribution or Gaussian curve is bell shaped and symmetrically centred on the mean. Asymmetry of the data set can be measured by defining some statistical characterization. Skewness is a measure of symmetry or more precisely the lack of symmetry[34]. A distribution or a

data set is symmetric if it looks the same to the right and left of the central value in accordance to the skewness for Normal distribution. The shape obtained is quite distorted on the negative side (skewness 0.154), and is clearly visible from the distribution graph with quite a shift in mean (see figure 2.9). Negative values for the skewness indicate data that are skewed left and positive values for the skewness indicate data that are skewed right. The probability of measured mass errors decreases with the distance from the mean and can be calculated with a probability function (equation 2.9). The hypothesis regarding the distribution form is rejected based on the Kolmogorov Smirnov (K-S) test which gave 0.1353 for the normality greater than K-S critical value which is 0.0438. This gives that EI+ data set distribution is not normally distributed. The value of the 2SD gave the degree of dispersion of data; ±2 SD interval covers 95% of all data, which report a high precision.

964 mass measurements were subjected to outlier statistical tests using SPSS and AccMass software. Only three outliers were identified by the Box and Whisker method, (see figure 2.10). These outliers did not affect the results very much; this became obvious after eliminating the outliers and reapplying the statistical methodology again. This is because of the minority of outliers and the closeness of outliers to neighbour points.

A Gaussian peak shape was typically assumed, even though the instrument gave mass errors deviating significantly from a Gaussian peak shape. As the outliers were determined and these values omitted, the error range was restricted between ±2 ppm,

and the majority of measurements were close to the mean, 0.15 ppm, indicating high accuracy, with precision equal to 0.807 ppm. Data covered the mass range 88-1,300 Da. This range was divided into 12 sub-ranges (100-200, 200-300, 300-400, up to 1,300 Da, as shown in table 2.4) and again the statistical calculations were performed. The best accuracy was recorded at mass range from 600 to 800 Da (figure 2.7), which involve the majority of data fallen in this interval.

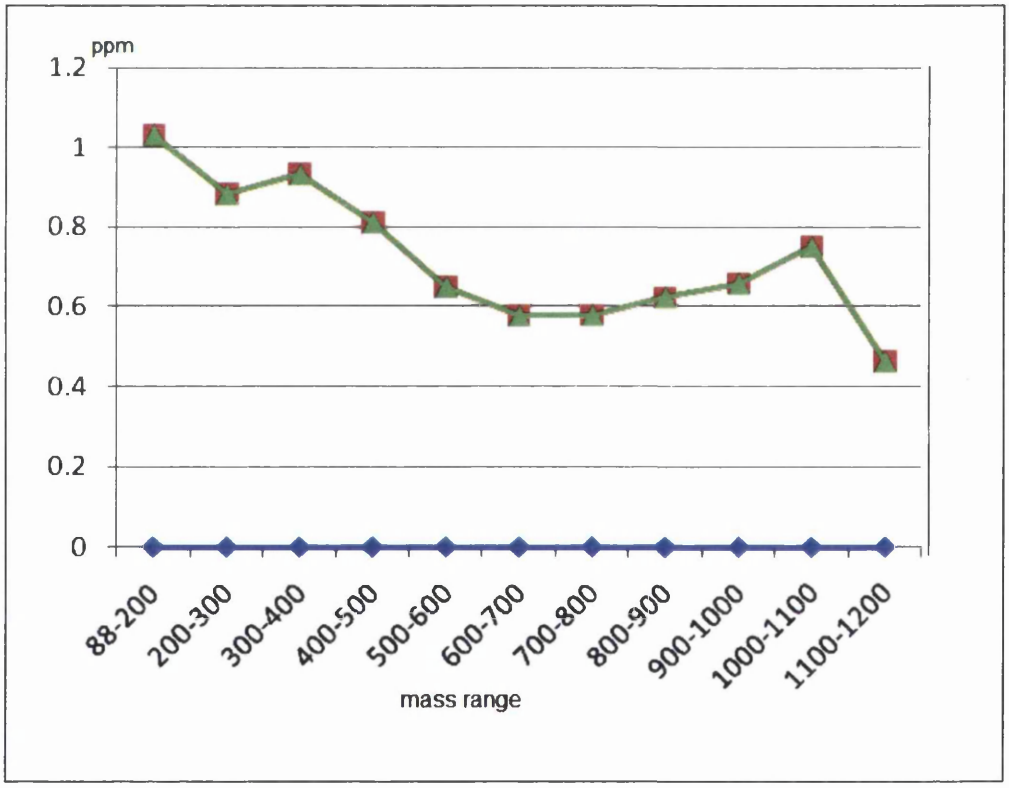

Figure 2. 7 *Shows the accuracy trend for EI+ MAT95 magnetic sector on the mass range 88-1200 Da, the best accuracy presented for the range 600-800 Da.*

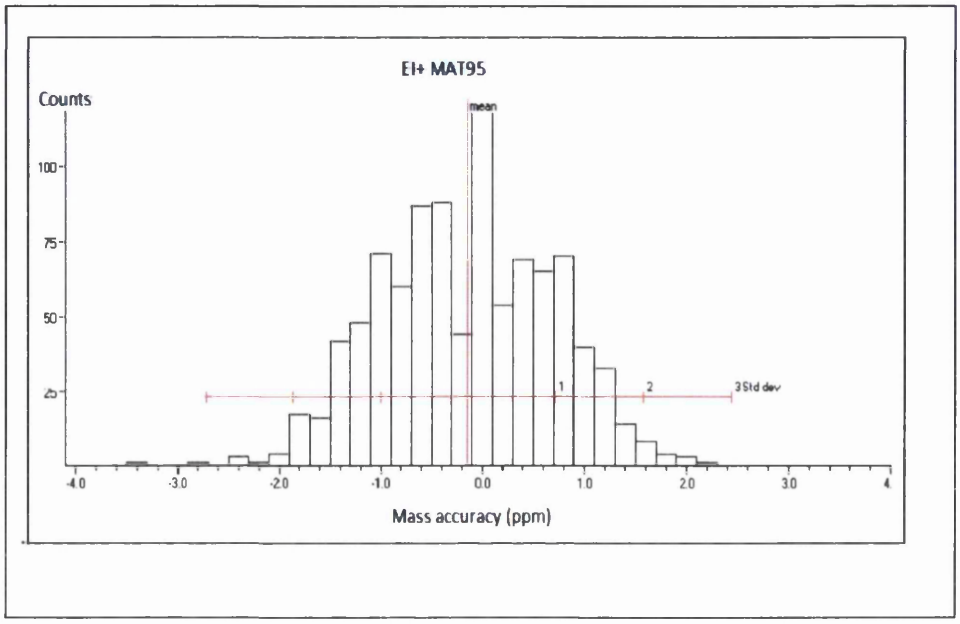

Figure 2. 8 *Bar graph (Histogram) describing the (964) measured points in 0.2 ppm bin width using AccMass software, the mean is shifted to the left side and the standard deviation is measured on the horizontal line (68%, 95% and 99%).*

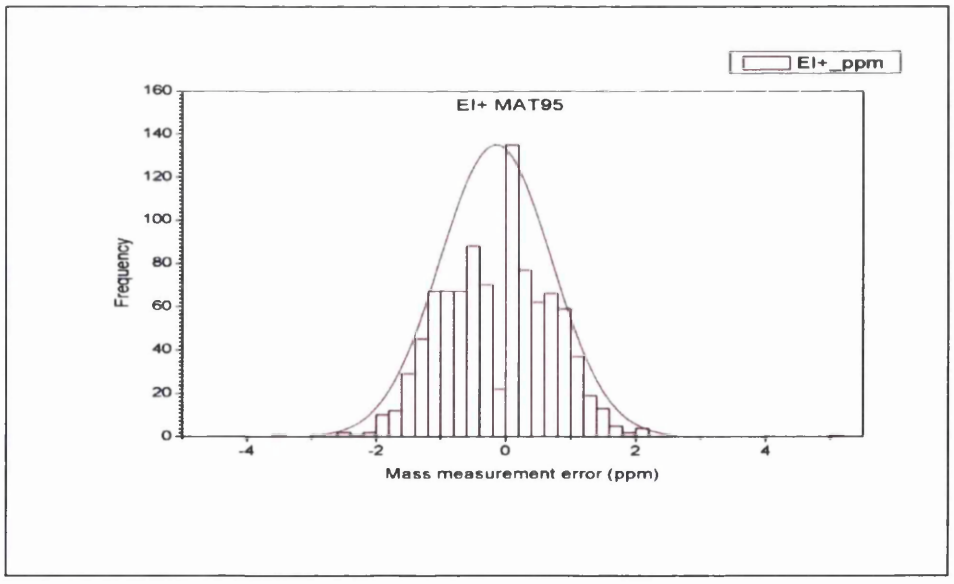

Figure 2. 9 *Mass measurement error distributed around the central value for EI+ MAT95. The shape is shifted to the negative value.*

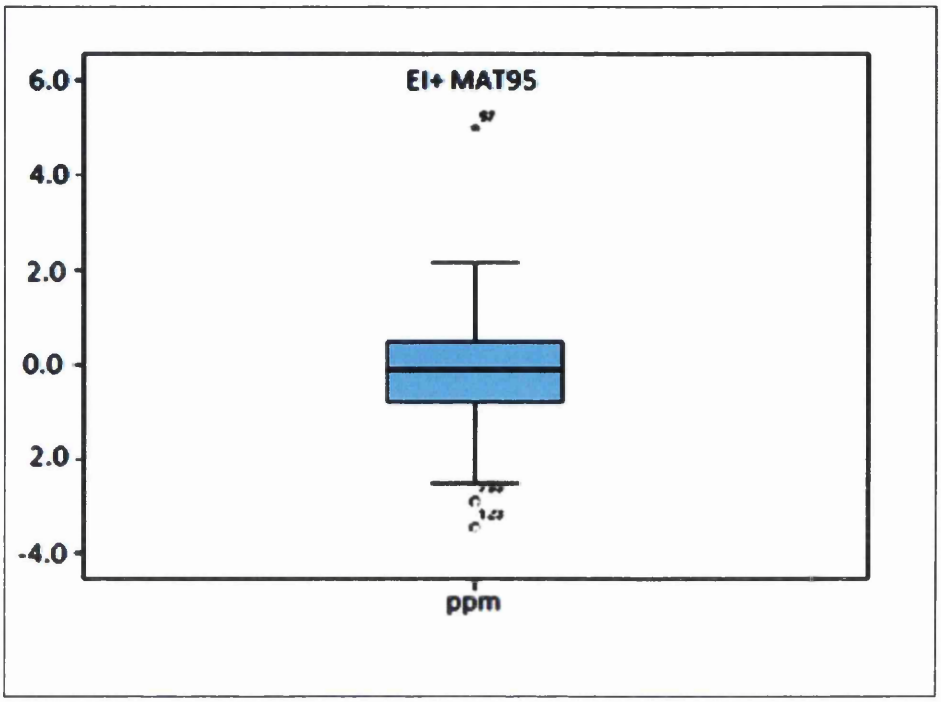

Figure 2. 10 *Box plot for 964 set points by EI+ MAT95 on ppm scale, 3 outlier were detected as shown (SPSS software).*

There is a small difference between the values of the descriptive statistics when outliers are included and omitted (refer to table 2.2 & 2.3 respectively). For example, EI+ technique has a standard deviation 0.858 ppm and aaMMA 0.701 ppm with outliers, whereas after omitting outliers the values for SD and the mean are 0.832 and 0.692 ppm respectively. This elucidates the slight effect of outliers on the precision and accuracy. It is a small adjustment, but we cannot neglect this value. This illustrates the importance of finding outliers and calculating accuracy and precision with and without outliers included. This will be discussed in more detail in the next chapter.

Table 2. 4 EI+ MAT95 data set (without outlier) separated into 12 mass range Mass errors in ppm.

Mass range	No.	AM	mean	σn-1	RMS	aaMMA	SEM	CL95%	CL90%
88-200	117	158.6072	-0.4241	0.9453	1.0324	0.8229	0.0874	± 0.1713	± 0.1433
200-300	260	249.9544	-0.2095	0.8598	0.8833	0.1808	0.0533	" ± 0.1045	0.0874
300-400	206	344.0145	-0.1042	0.9274	0.9333	0.789	0.0648	± 0.1270	± 0.1062
400-500	125	453.967	-0.1221	0.8058	0.8118	0.6822	0.0721	± 0.1413	± 0.1182
500-600	85	548.09	-0.0108	0.6546	0.6508	0.5288	0.071	± 0.1392	± 0.1164
600-700	71	648.2826	0.0024	0.5821	0.5779	0.5025	0.0691	± 0.1388	± 0.1160
700-800	44	753.2736	0.1227	0.5729	0.5795	0.4619	0.0864	± 0.1736	± 0.1451
800-900	21	844.8645	-0.0896	0.6324	0.6236	0.5492	0.138	± 0.2884	± 0.2374
900-1000	17	942.1055	-0.1469	0.6618	0.6586	0.5832	0.1605	± 0.3430	± 0.2809
1000-1100	10	1047.691	-0.4237	0.6535	0.7509	0.6574	0.2067	± 0.4671	± 0.3782
1100-1200	4	1146.3382	-0.1821	0.492	0.4634	0.3122	0.246	± 0.7823	± 0.5781
1200-1300	1	1270.2001							
ALL	961	402.7141	-0.1518	0.8316	0.8449	0.6918	0.0268	-0.1518 ± 0.0526	± 0.0440
				Kolmogorov = 0.1399					

2.4.1.2 Chemical ionisation (CI) data sets

CI positive ion mode (MAT95): Methane or ammonia was used as a CI gas. The preferred mass reference compound is PFTBA and used as in EI mode. PFTBA may be used with ammonia, but polyethylene glycol (PEG) is more usual, as PFTBA does not ionise well with ammonia. If PEG is used, it is placed on the solids probe with the analyte and the two species are encouraged to vaporise simultaneously.

Accurate mass measurements were taken by chemical ionisation technique (CI+) on the Finnigan MAT95 instrument, at (5 kV) accelerating voltage for different samples, using the internal mass calibration method. Mass errors were plotted as a histogram in figure 2.12 and show a distribution of 71 CI+ accurate mass over the mass range 169-998 Da. Statistical methodology was applied; the average mass is 352.1216 Da. The average mass error -0.1079 ppm and the precision (at 1 SD) is 0.9282ppm. Only one outlier was found, the 70 results (after eliminating outliers) gives a mass accuracy -0.0766 ppm (-0.0443 mDa) and a SD 0.8965 ppm (0.3224 mDa), while the aaMMA = 0.7372 ppm refer to table 2.2 & 2.3. The histogram seems to have a "Normal distribution". This is confirmed by the normality test [experimental K-S test 0.1462 while critical K-S 0.1614]. This result also is confirmed by the linearity test (Q-Q plot) for the CI+ data set (figure 2.14). A Q-Q plot indicates whether two data sets come from populations with a common distribution. The points should fall approximately along this reference line. More about this method will be discussed in detail in the next chapter. Here CI+ MAT95 data follows the linear line, except one

point, which was determined as an outlier, (see figure 2.13). This range of mass error measurements was divided into 14 sections, each 100 Da wide, (see table 2.5). The mass accuracy for each mass range 100 Da wide is observed to change linearly as the mass increases.

As shown in figure 2.11 the best accuracy was recorded at mass range 400-500, and the worst at lower masses 100-200. Overall the accuracy obtained here, based on the trend line, continued to increase until 500 Da then reduced at 600 Da, as get no reason of this observation can be made.

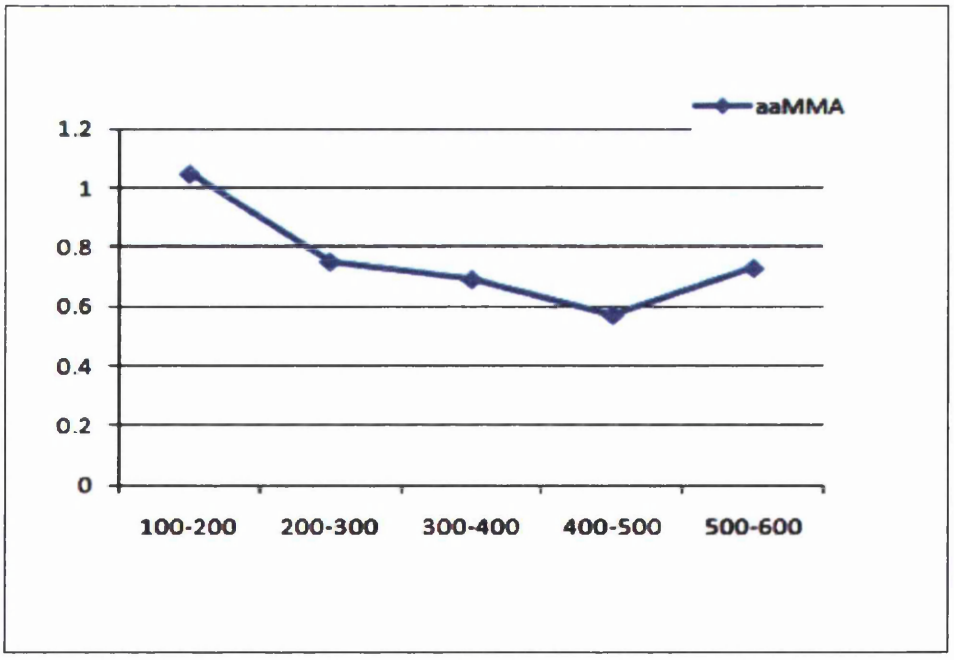

Figure 2. 11 *Average mass measurement accuracy (aaMMA) for CI+ MAT95 for 100 Da mass ranges over the whole range of masses experimentally measured.*

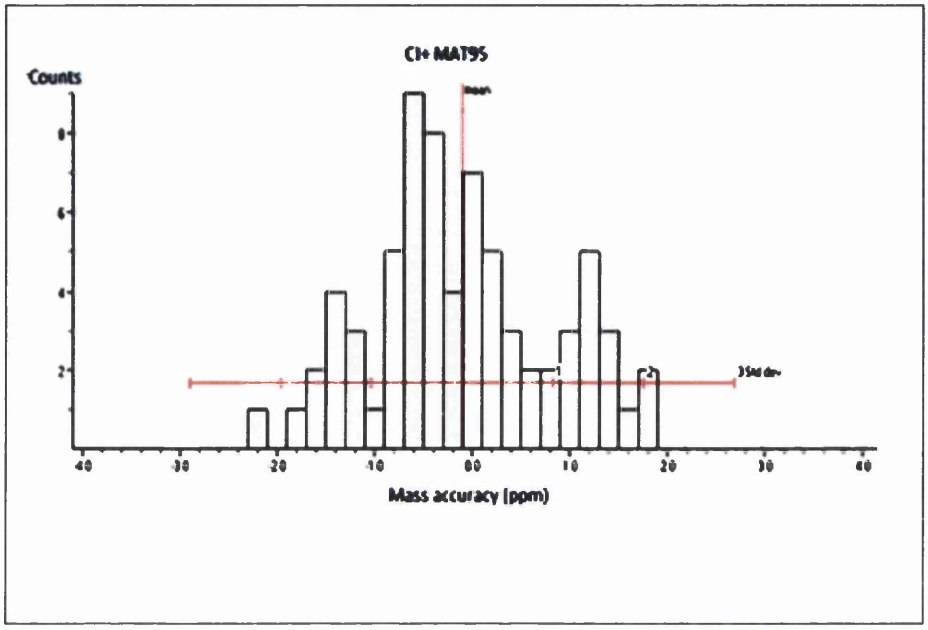

Figure 2. 12 *Mass accuracy and standard deviation measurements for CI+ MAT95 data set in 0.2 ppm ions plotted width using Acc mass software.*

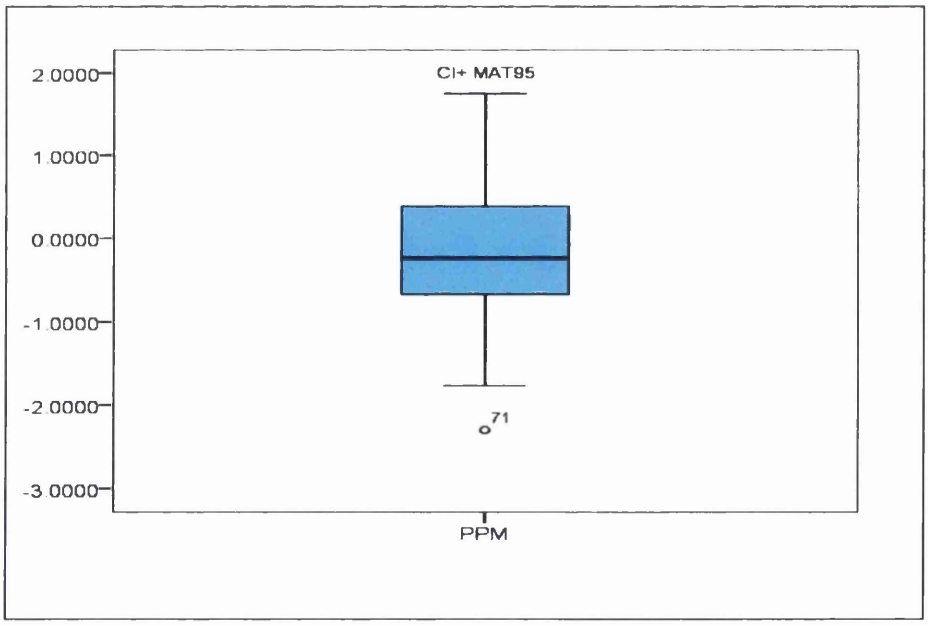

Figure 2. 13 *Box plot for 71 mass measured by CI+ MAT95, to examine the outliers.*

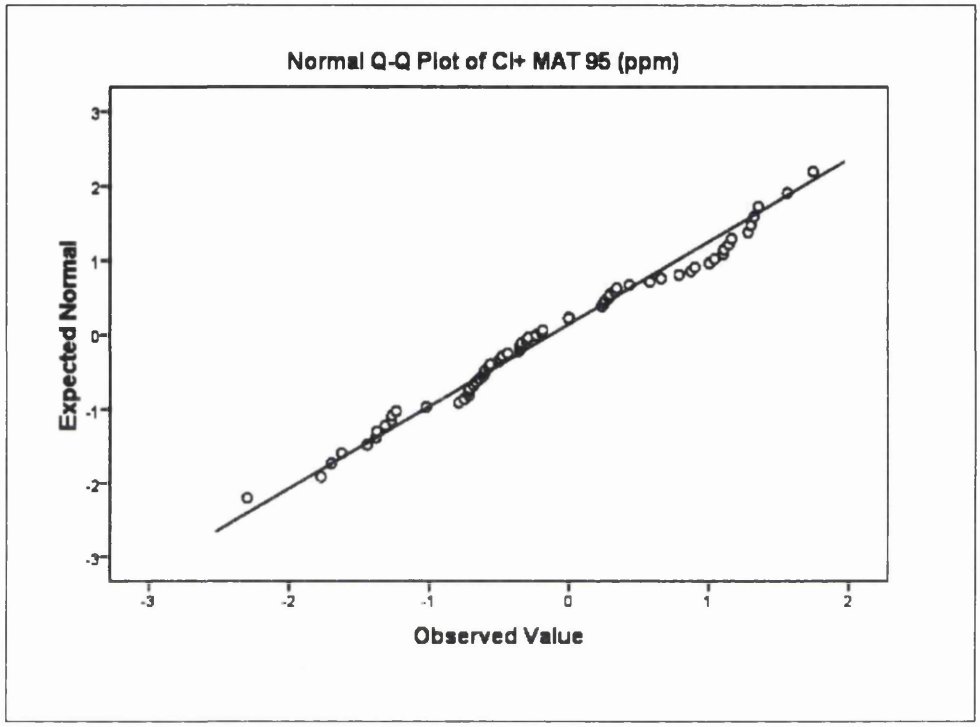

Figure 2. 14 Normal Q-Q plot for 71 points of CI+ MAT95 data sets.

2.4.1.3 Electrospray ionisation (ESI) data set

Accurate mass measurements were taken by ESI+ on the Finnigan MAT95 and MAT900 double focusing mass spectrometers, operating at 5 kV accelerating voltage, and on an LTQ-Orbitrap fitted with an Advion TriVersa NanoMate (Advion spray voltage was 1.45 kV). The Orbitrap extraction plate was typically set at +10 to 40 V and the heated transfer tube set to 200°C. Low molecular weight grade of chemical material, ammonium acetate (NH4Ac), polyethylenamine (PEI), nitrobenzyl alcohol (NOBA), HPLC grade solvent, were employed. Samples for ESI were dissolved in 100 µl of dichloromethane (DCM) or another solution if DCM is inappropriate.

Table 2. 5 Statistical calculation for CI+ MAT95 data set over mass range (169:998 Da), the table illustrates the accuracy and precision for every 100 range Da of masses.

Mass range	no. of pts	AM	mean	σ(n-1)	RMS	aaMMA	SEM	CL-95%	CL-90%
100-200	7	169.9657	0.3891	1.2035	1.1803	1.0485	0.4549	± 1.1145	± 0.8825
200-300	22	252.4006	-0.2655	0.8905	0.9096	0.7514	0.18898	± 0.3968	± 0.3265
300-400	20	352.177	0.2019	0.8526	0.8552	0.6918	0.1906	± 0.3994	± 0.3289
400-500	11	440.7041	-0.1324	0.6815	0.6631	0.5717	0.2055	± 0.4582	± 0.3719
500-600	7	532.3629	-0.4129	0.8346	0.8761	0.7291	0.3155	± 0.7729	± 0.612
600-700	1	678.252	***	***	***	***	***	***	***
700-800	1	702.3201	***	***	***	***	***	***	***
800-900	0		***	***	***	***	***	***	***
900-1000	1	998.5413	***	***	***	***	***	***	***
ALL	70	353.4216	-0.0766	0.8965	0.8933	0.7372	0.1071	± 0.2154	± 0.1800

Kolmogorov = 0.1442

For the Orbitrap in ESI+ mode, a standard solution mix was used for calibration, which consisted of caffeine, methionine-arginine-phenylalanine-alanine (MRFA) acetate salt and Ultramark.

Data obtained on the MAT900 using ESI+ ionisation consisted of 2,317 measurements covering a mass range 100-1,500 Da. The accuracy and precision obtained was excellent at ±1 mDa. An average mass error equal to 0.0316 ppm (0.0165 mDa) was achieved, and a standard deviation of 0.9076 ppm (0.391 mDa). Box and Whisker plot gave nine outliers over the whole data set. Only two outliers were recorded as extreme outliers which is more than 3*IQR and falls outside the fences, while the other seven, called mild outliers, are close to the main data and fall between 1.5*IQR and 3*IQR, (see figure 2.16). The star (*) indicates the extreme outlier and the circle (o) indicates the mild outlier. This indicates that the measurements afforded excellent accuracy, with mean = 0.0261 ppm and SD= 0.8514 ppm. Absolute average mass measurement accuracy (aaMMA) = 0.6942 and RMS=0.8516 ppm, and indicates again the high accuracy obtained for this technique.

Figure 2.15 is a histogram plot of the 2,317 masses on MAT900-ESI+, the distribution assumes a bell shape distribution, typical for "normally distributed data". The plot was generated using AccMass and Easyfit software and the data were tested by Kolmogorov Smirnov (K-S) methodology. The critical points of K-S test were presented from statistical tables. The test is defined by the greatest vertical distance at any point between the two cumulative probability distributions. There are nine outliers

recorded by a Box plot. The distribution is non-Normal distribution. This is clear from K-S test, the largest difference between Normal distribution and the ESI+ data sets distribution (D) is 0.0709. The 5% critical value for n= 2317 is 0.0283. Therefore the conclusion from these that the K-S test rejects the normality hypothesis and the ESI+ MAT900 data set is not evenly distributed. The outliers detected are defined according to Q-Q plot and Box Whisker plots (see figure 2.16 and 2.17), both present 9 outliers. It can be verified that even after elimination of the 9th highest results, which are suspected of being outlier, the data not fit a Normal distribution as also the D-value is 0.07295 which is greater than the critical value.

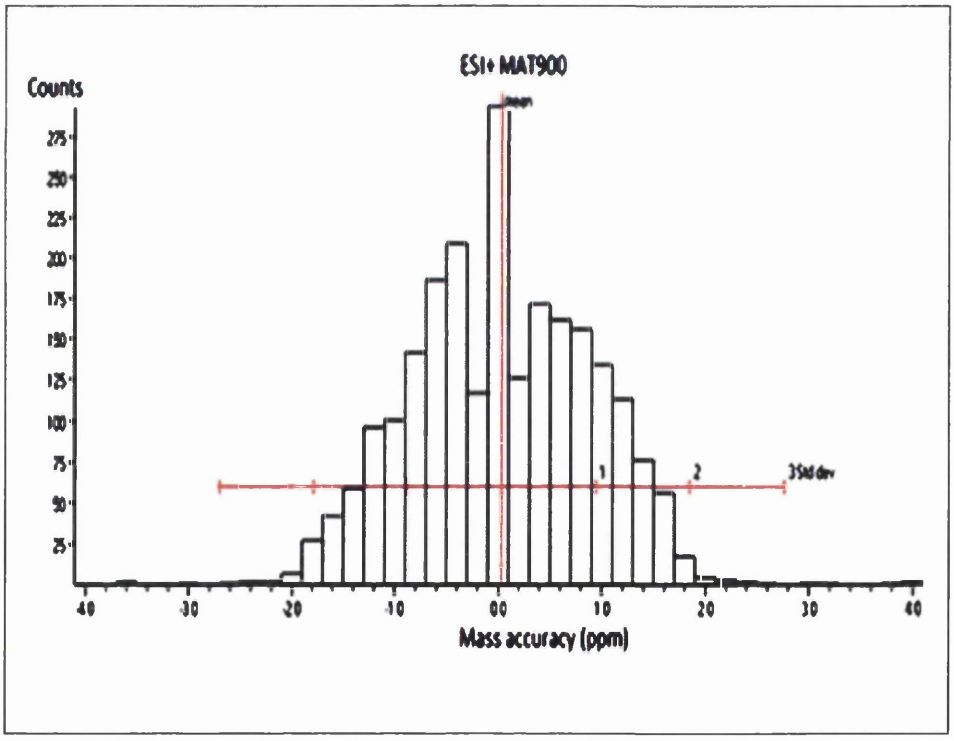

Figure 2. 15 Histogram of ESI+ MAT900 data (n=2317) which approximates to a Gaussian distribution

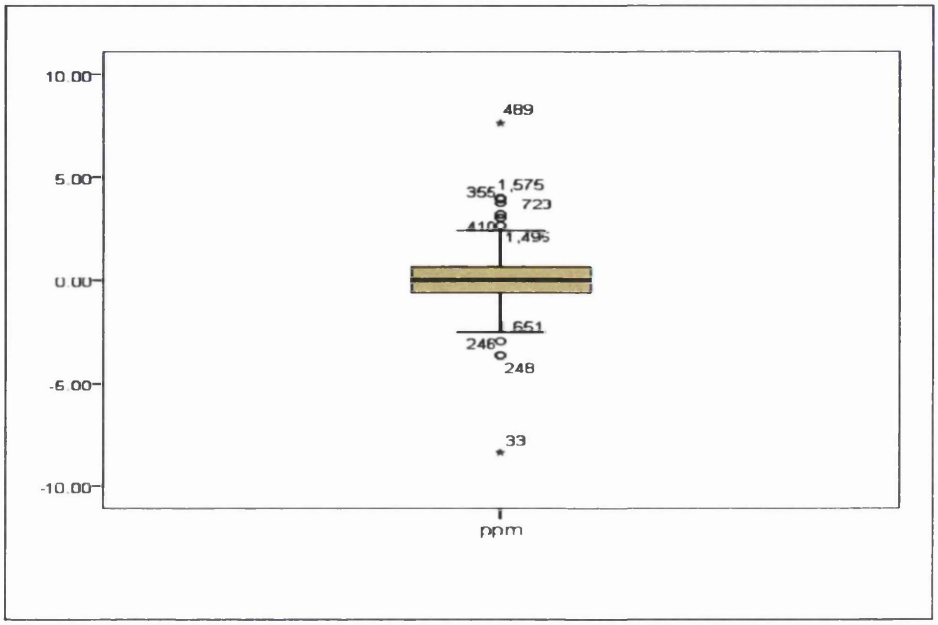

Figure 2. 16 *Box and Whisker plot for the ESI+ MAT900 data set where 10 outliers were detected.*

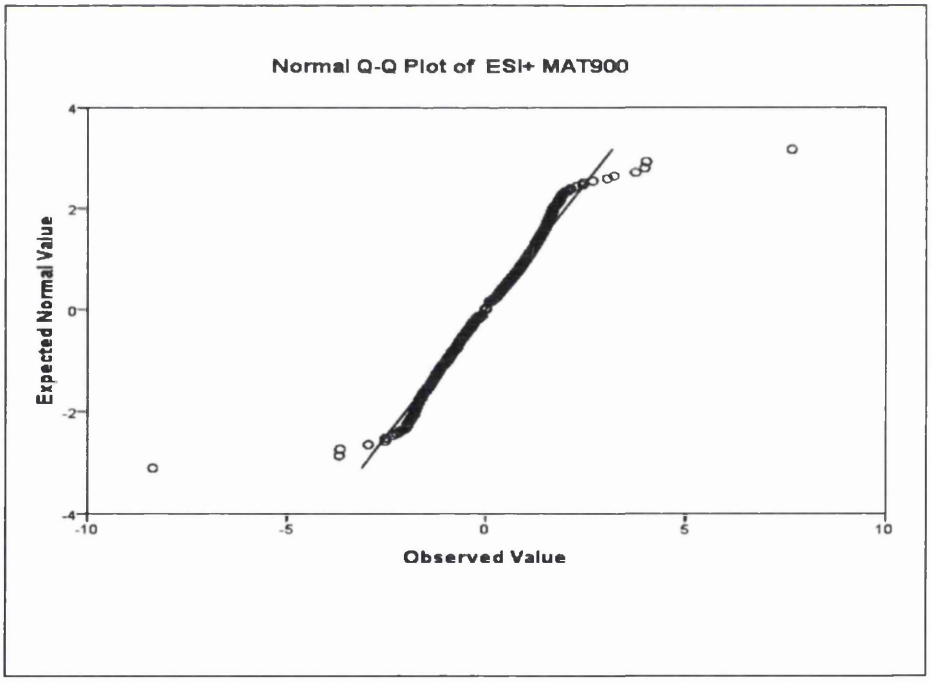

Figure 2. 17 *Q-Q Plot for data set of ESI+ MAT900 outliers were detected which appear far from the line (10 points).*

The ESI+ data consist of 408 accurate mass data, obtained on MAT95 over mass range 116 to 1088 Da, the average mass is 373.0933 Da. The histogram and the distribution of ESI+ accurate mass data is shown on figure 2.18. The data points do not have a Normal distribution, based on K-S test, which gave D-values 0.1219, and the critical value 0.06723. No outliers were detected, (see Box and Whisker plot shown in Figure 2.19). The average mass error measured is 0.0745 ppm (0.0218 mDa), and standard deviation SD= 0.8059 ppm (0.3006 mDa). These results suggest excellent accuracy and precision of the measurements. Estimated magnitudes of MMAs are aaMMA = 0.672 ppm; and RMS = 0.808 ppm.

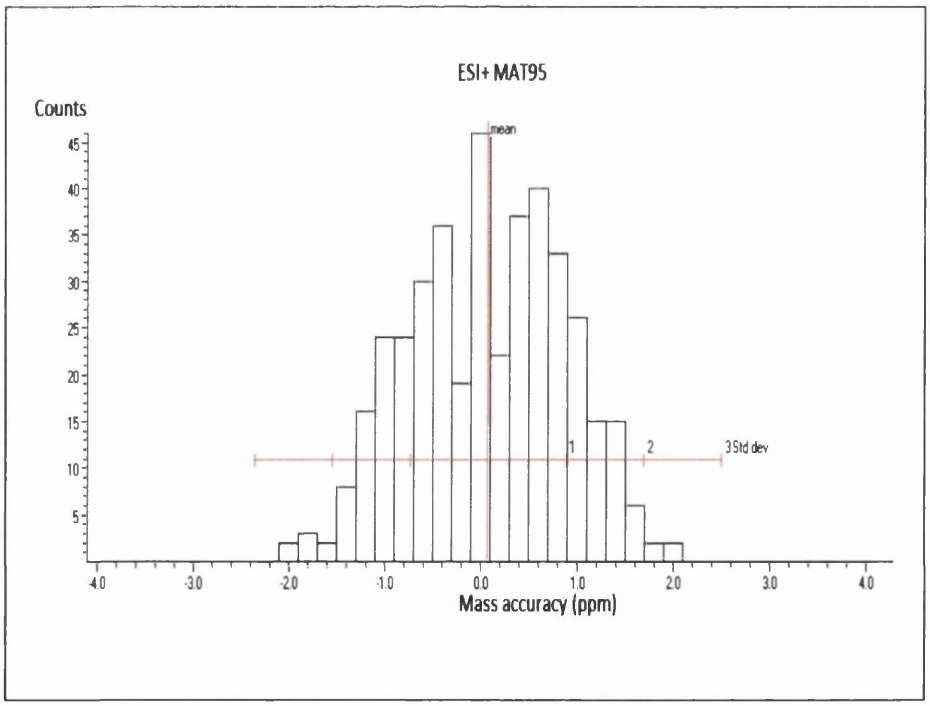

Figure 2. 18 Bar graph shown the distribution of ESI+ MAT95 data set.

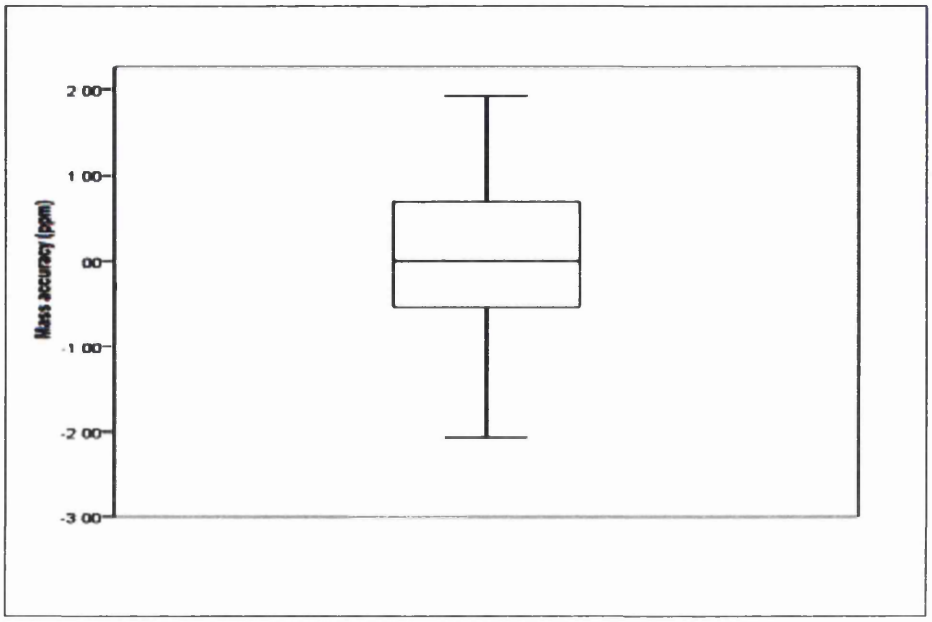

Figure 2. 19 Box Whisker plot for data set of ESI positive MAT95 with no outliers detected in this technique.

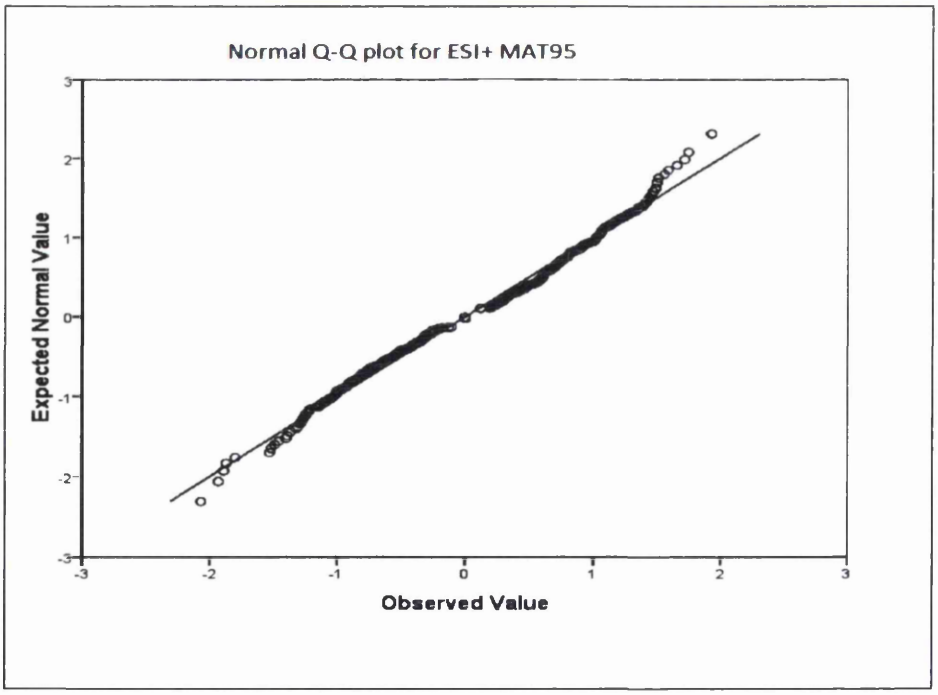

Figure 2. 20 Q-Q Plot for data set of ESI+MAT95, no outlier was detected

Table 2. 6 Electrospray data set in positive mode over MAT900 instrument (2313 mass measurements)

Mass range	No. of pts	AM	Mean	σ(n-1)	RMS	aaMMA	SEM	CL-95%	CL-90%
100-200	167	165.2534	-0.0282	0.9205	0.9182	0.738	0.0712	± 0.1396	± 0.1168
200-300	590	256.0224	0.0366	0.8312	0.8313	0.6715	0.0342	± 0.0671	± 0.0561
300-400	678	344.8853	0.0116	0.8503	0.8497	0.7024	0.0327	± 0.0640	± 0.0536
400-500	410	447.0488	0.0651	0.8233	0.8249	0.6799	0.0407	± 0.0797	± 0.0667
500-600	175	543.5029	-0.1252	0.8138	0.821	0.6407	0.0615	± 0.1206	± 0.1009
600-700	127	645.0656	0.0752	0.8602	0.8601	0.7036	0.0763	± 0.1496	± 0.1252
700-800	73	740.7722	0.2885	0.7606	0.8086	0.6592	0.089	± 0.1789	± 0.1496
800-900	48	846.2945	0.0727	1.0609	1.0523	0.8966	0.1531	± 0.3078	± 0.2572
900-1000	18	950.1609	-0.194	0.8693	0.8667	0.6802	0.2049	± 0.4323	± 0.3565
1000-1100	14	1030.324	-0.2892	0.9319	0.9434	0.7451	0.249	± 0.5379	±0.4408
1100-1200	7	1154.3347	-0.3413	1.4475	1.3829	1.3489	0.5471	± 1.3404	± 1.0614
1200-1300	3	1268.1675	1.2591	0.4716	1.3166	1.2591	0.2723	± 1.1709	± 0.7951
1300-1400	2	1352.0637	0.6703	0.2355	0.6907	0.6703	0.1665	± 2.1167	± 1.0509
1400-1500	1	1432.4477							
ALL	2313	395.6145	0.0261	0.8514	0.8516	0.694	0.0177	± 0.0347	± 0.0290
Kolmogorov = 0.0729									

2.4.1.4 Fast-atom bombardment (FAB) data

Accurate mass data was acquired on the Finnigan MAT95 mass spectrometer equipped with a FAB source over the mass range 331-1388 Da, (n= 58 results) using peak matching. The average mass obtained was 823.3390 Da, the average mass error measured was -0.032 ppm (-0.0052 mDa), and the standard deviation was 0.7251 ppm (0.698 mDa). These parameters are indicative of the accuracy and precision obtained, respectively. The aaMMA is 0.604 ppm and root mean square RMS = 0.720 ppm. These results indicate that high accuracy was achieved; the aaMMA obtained was better than the 3 ppm achieved by Muddiman and his group on the JEOL JMS-HX110HF double- focusing magnetic sector[23]; and data obtained using peak matching are more accurate than dynamic voltage scanning. These results are inconsistent with the findings of Muddiman, et al. which showed 3 ppm mass accuracy obtained using dynamic voltage scanning.

The histogram in figure 2.21 shows that the distribution of the 58 mass errors obtained on FAB is roughly symmetrical about the mean, with a small deviation of the mean towards a negative mean error. No outliers were found, this is clear from the graphical test shown in figure 2.22 and figure 2.23.

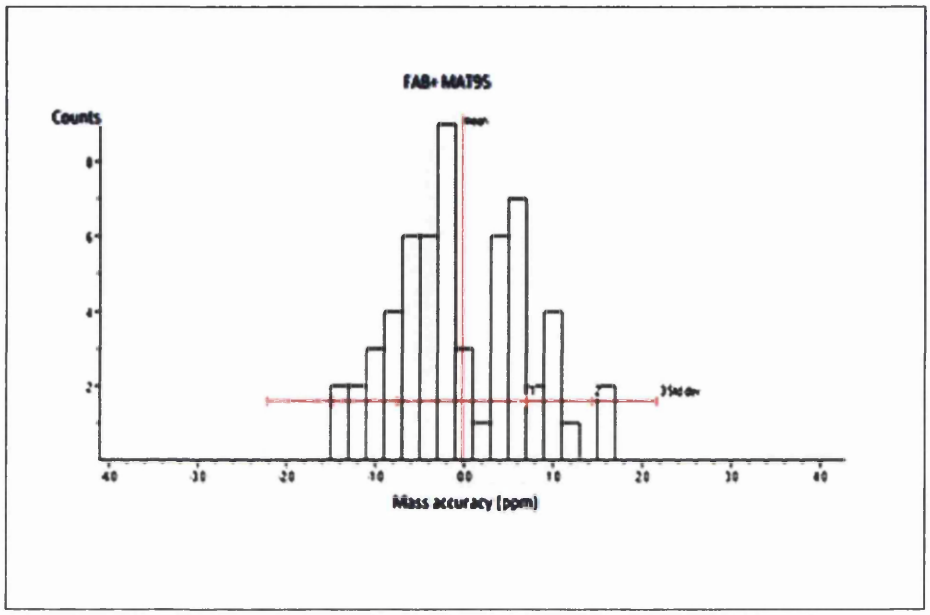

Figure 2. 21 *Histogram of mass measurements accuracy (ppm) for 58 individual masses taken on MAT95 sector mass spectrometer.*

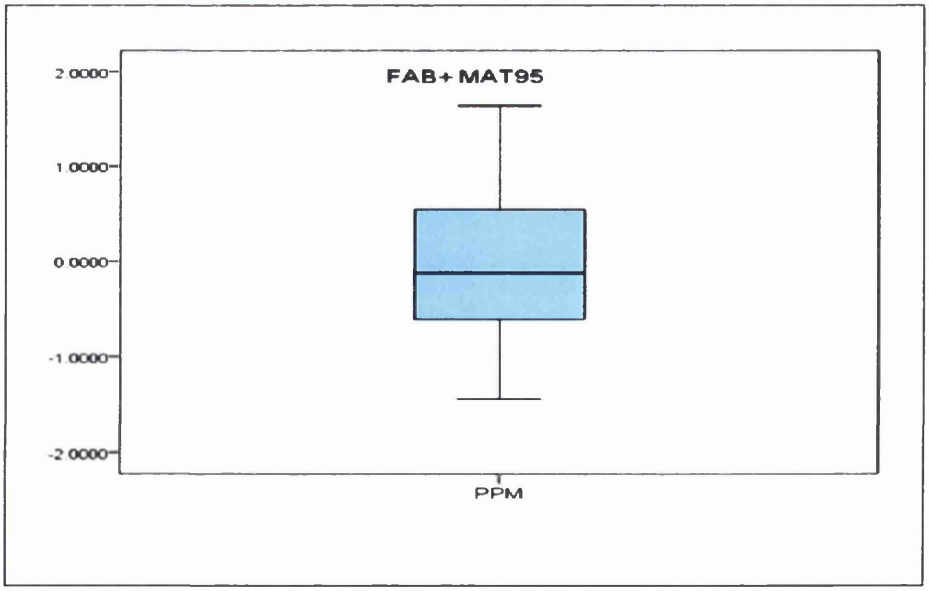

Figure 2. 22 *Box and Whisker plot for testing for outliers in FAB+ MAT95 measurements, there appear to be no outliers.*

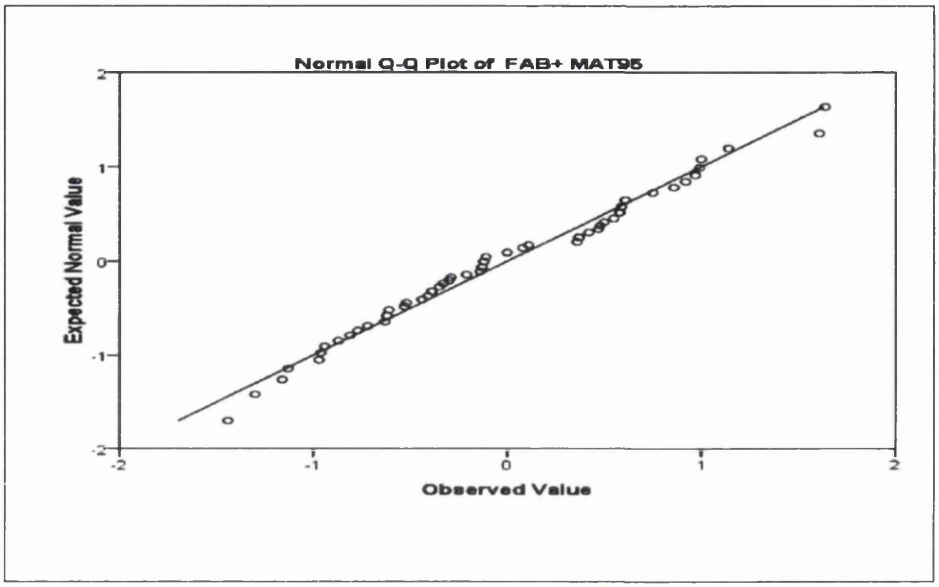

Figure 2. 23 *Q-Q plot for FAB+ MAT95 for outlier test masses measured and outlier was detected.*

2.4.2 Orbitrap mass spectrometer

The Orbitrap is a high mass accuracy instrument. Recent literature papers mention its accuracy is less than 2 ppm for protein and peptide mixtures[35,36], it can reliably deliver mass accuracy below 1 ppm[37,38]. The Orbitrap mass analyser is a feasible alternative to FT-ICR as a high resolution and accurate mass detector for separation techniques. The mass accuracy can be further improved when using internal calibration; employing known background ions from ambient air has proven to be very useful. The use of such lock masses allows the possibility of achieving reliable sub-ppm mass accuracy[20].

2.4.2.1 ESI+ Orbitrap

More recent ESI+ work on a sample set of 784 masses, all singly charged, over a range of molecular weights from 131 to 1,566 Da, was selected and employed in an Orbitrap mass spectrometer fitted with an Advion Triversa injection system as the ESI interface. The Advion system greatly improves the quality of measurements due to greater reproducibility of nano-electrospray from the Advion chip and good sensitivity. It provides zero carry-over, which is essential in the EPSRC National Mass Spectrometry Service Centre, where a very diverse range of samples are submitted. Orbitrap full scan operation is very efficient for high throughput accurate mass analysis, but with slightly larger estimates of the magnitude of MMA: statistical analysis for Orbitrap data was applied and the mean obtained was 0.0139 ppm, aaMMA = 0.978 ppm; SD = 1.201 ppm and RMS =1.208 ppm. The histogram of this data was found to be of slightly inferior quality to the MAT95's data, as assessed from SPSS statistics, (figure 2.23), and it would be expected to improve if internal calibration was employed on the Orbitrap.

As described earlier, a Box and Whisker plot was used to determine the outliers; only four outlier measurements were found in this data set, as can be seen from figure 2.24. Omitting the outliers did not affect the statistical result; mass accuracy obtained after removing the outliers is 0.1382 ppm.

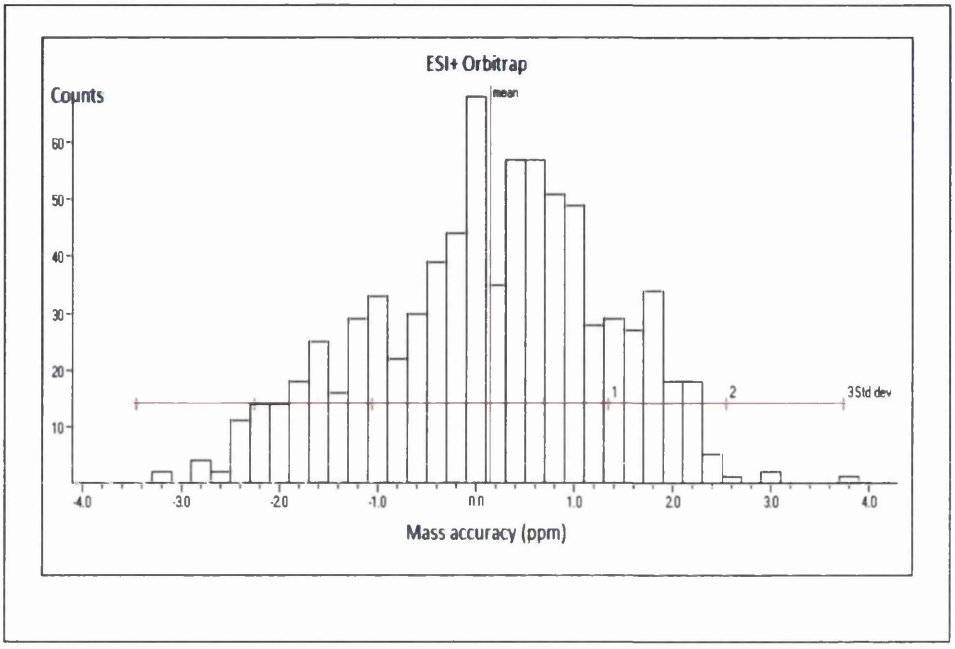

Figure 2. 24 *Histogram and the distribution of 784 mass measured for ESI+ Orbitrap.*

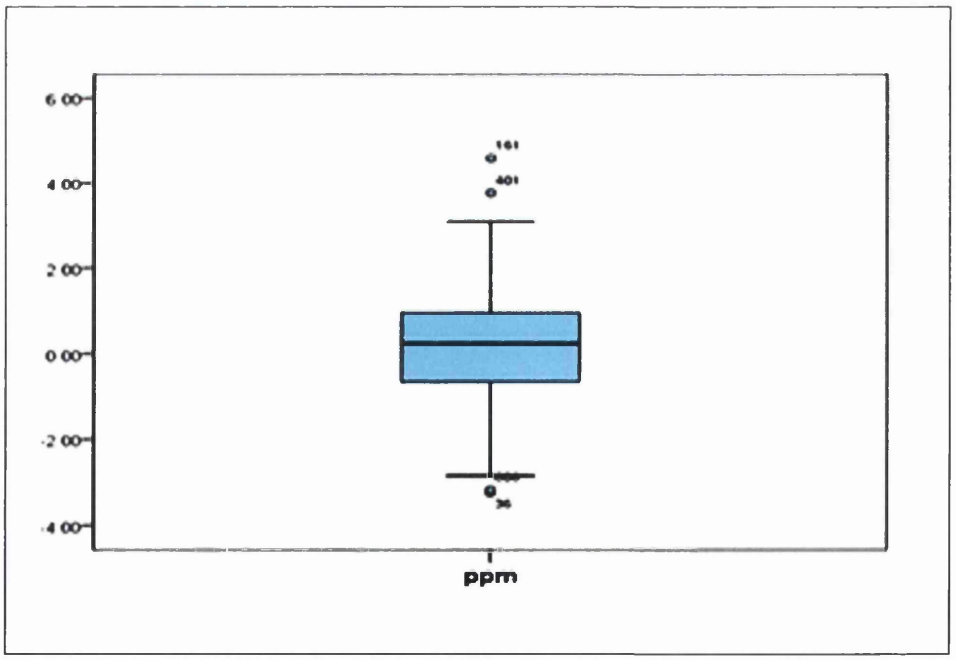

Figure 2. 25 *Box Whisker test for ESI+ Orbitrap data set. Four outliers appear outside of the Box.*

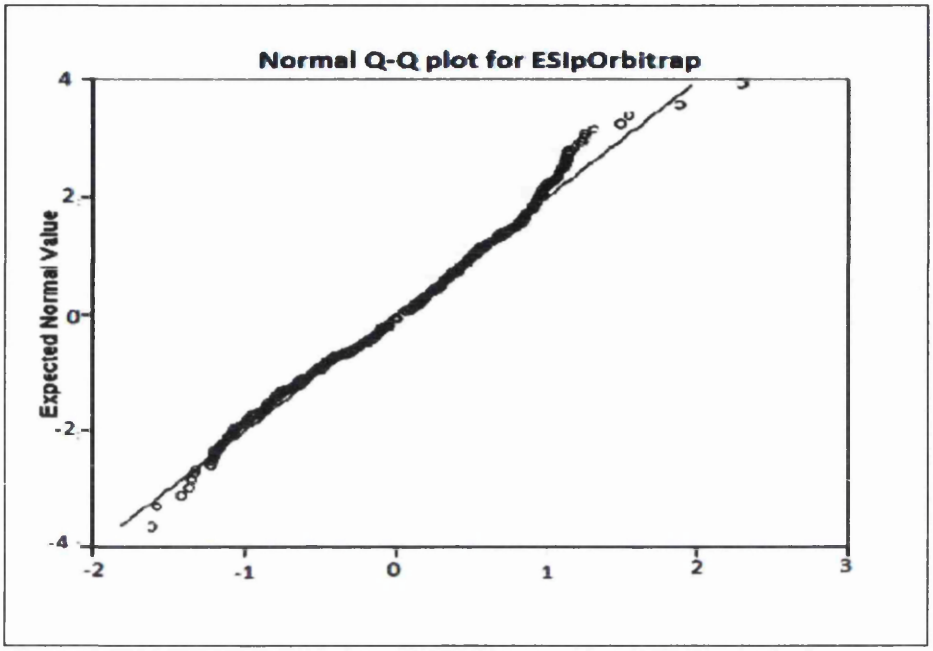

Figure 2. 26 *Normal Q-Q plot test for ESI+ Orbitrap data sets, four outliers deviate from the straight line.*

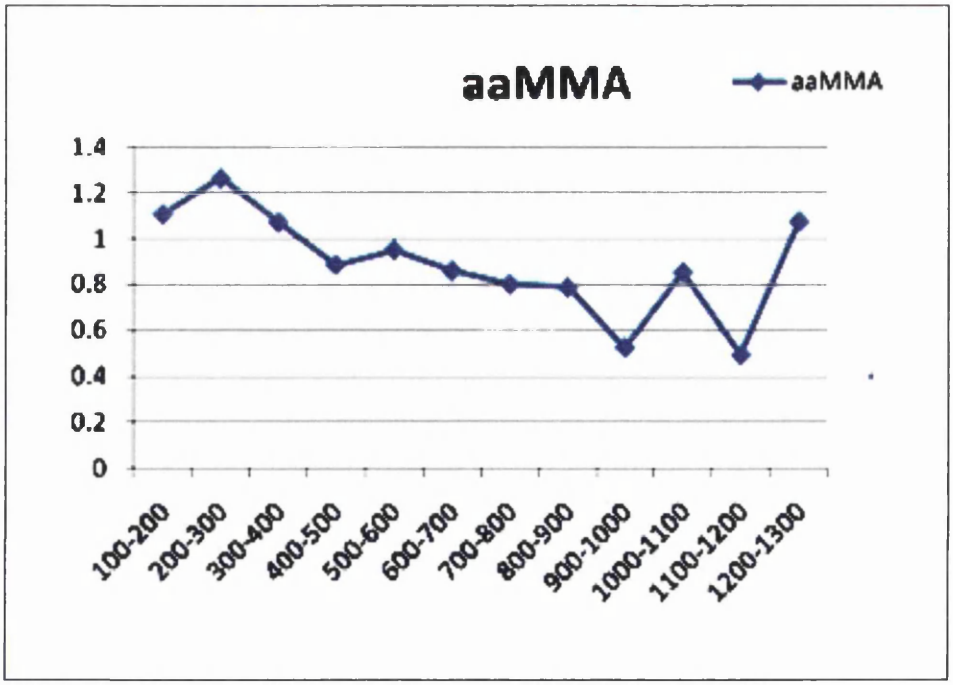

Figure 2. 27 *Mass accuracy against mass range for ESI+ Orbitrap.*

A Q-Q plot of ESI+ Orbitrap data set versus a Normal distribution is illustrated in figure 2.25. The outliers are evident at the high end and low end of the range. The majority of data follows the normal line with skewness to the high end. The outliers distort the linearity of data sets. This seems consistent with the data histogram which deviates to the right side. The K-S test confirms this result by rejecting the normal hypothesis. The D-value is 0.1103 and the critical value is 0.0486. The accuracy of the ESI+ data set is achieved at 2 ppm which is similar to recent studies[37,38]. The trend line for the accuracy against the mass errors displayed the rising of accuracy with increasing mass range from 200 to 900 then the accuracy become worse at the upper range (see figure 2.27). The precision of Orbitrap data set is recorded between 1 and 1.2 ppm for mass range 100 -500 Da. It then improves to be less than 1 ppm at range 500-800 Da, and 0.6 ppm at 900 Da, with unstable trend over mass range more than 900 Da.

2.4.2.2 ESI-Orbitrap

Another 65 masses measured in ESI negative ion mode were obtained on the Orbitrap, which was fitted with an Advion Triversa injection system as the ESI interface. Normality was examined over mass range 146 to 947 Da, with average mass 498.5575 Da. Figure 2.28 shows the distribution of 65 accurate masses obtained for singly charged ions on the Orbitrap mass spectrometer by ESI negative mode. The summary statistics of 65 individual masses gave average absolute mass measurement

accuracy: aaMMA= 0.999 ppm, root mean square RMS = 1.267 ppm and a standard deviation SD= 1.2619 ppm (0.639 mDa).

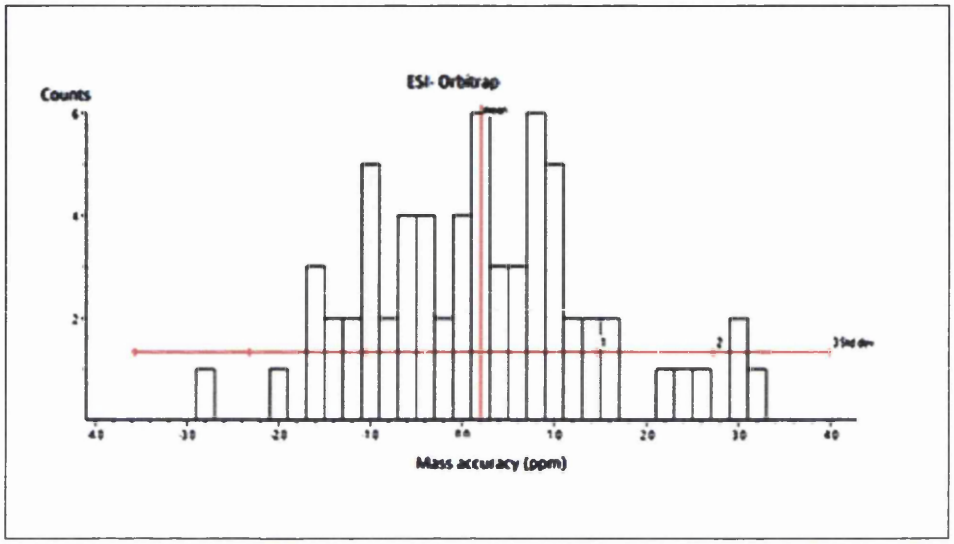

Figure 2. 28 ESI negative mode- Orbitrap data histogram, 65 mass measured, the distribution curves seems normal.

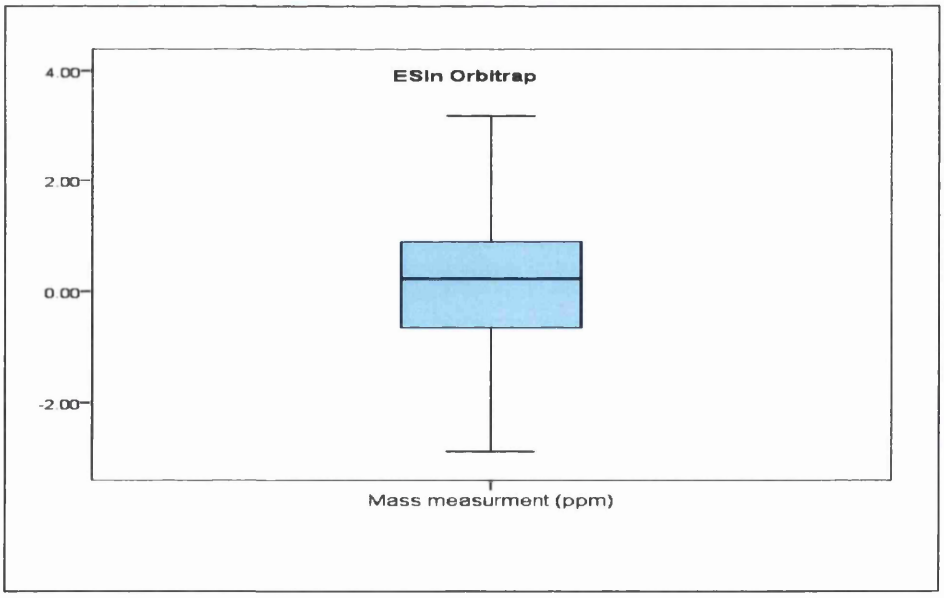

Figure 2. 29 Box plot for ESI- Orbitrap (n=65) and there are no outliers.

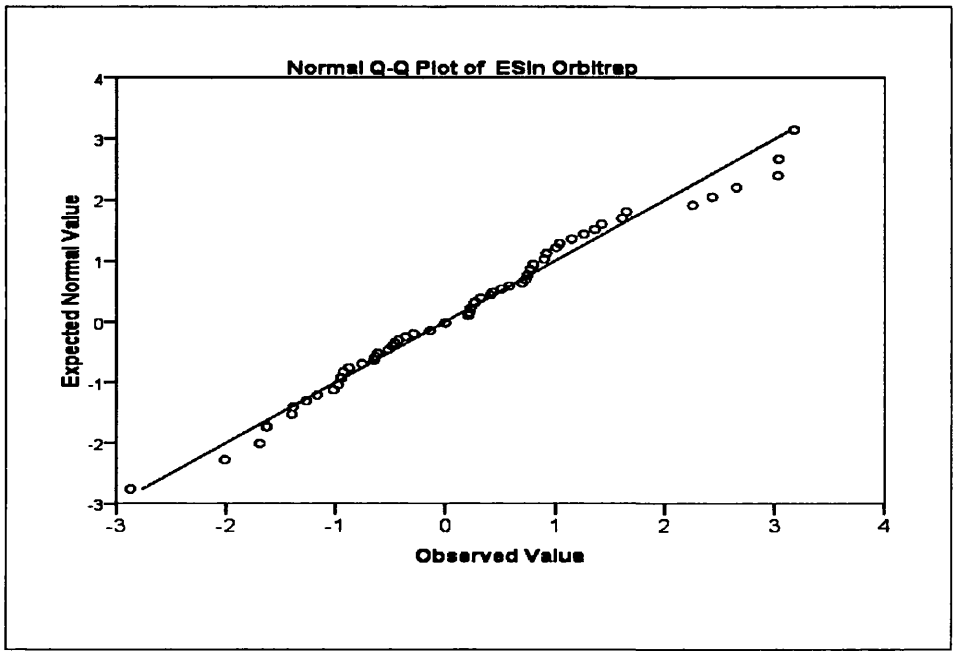

Figure 2. 30 *Q-Q plot for ESI- Orbitrap (n=65.) Data fits close to the line with departure for some measurement from the reference line at the higher and lower end.*

2.4.3 Time-of-flight (TOF)

All time-of-flight mass analysers rely on very low pressures for good performance, because in order to discriminate ions at high resolution on the basis of their flight times, collisions must be eliminated. TOF has no theoretical upper limit to mass range[39]. MALDI-TOF instruments typically extend well beyond 200,000 *m/z* ; oaTOF applications rarely require >30,000 *m/z*; mass resolution of the best reflectron TOF instruments is now beyond 30,000, but 10,000 is a more typical performance. Resolution is severely compromised in linear TOF mode, and resolution does not exceed 103. Electrospray orthogonal time-of-flight mass spectrometers (oa-TOF/ESI

MS) are capable of generating mass data of sufficient accuracy and precision to establish the elemental composition of an analyte[40].

2.4.3.1 MALDI-TOF

MALDI-TOF mass measurements were acquired with a voyager DE-STR instrument (Applied Biosystems, Framingham, CT,USA). Data was acquired using positive ion reflectron mode. Delay time was set to 150 or 100 ns, acceleration voltage =20 kV, and grid voltage = 65.5%. Every acquisition was an accumulation of 2 x 25 laser shots, and the laser power was optimised for each analysis. 195 masses were obtained on time-of-flight (TOF) using the matrix-assisted laser desorption /ionization source, for small molecules at a mass range 246-1158 Da, generally observed as positive radical ions as an aprotic matrix was usually employed giving $M^{+\bullet}$ and sometimes $[M+H]^+$ species. The mean obtained was -0.4290 ppm with absolute average mass measurement accuracy, aaMMA=2.7193, standard deviation, SD=3.4077, RMS = 3.4259 ppm. Figure 2.31 shows the distribution of 195 mass measurements for TOF plotted as a histogram, which assumes a broad distribution. The Kolmogorov Simonov test gives a result of 0.3511, which is greater than the critical value 0.1687. This indicates the data are not normally distributed.

The K-S and Chi^2 tests are the most powerful tests since most statistical tests are based on the assumption that the data follow a Normal distribution, both are used for normality test, but K-S is the most applicable, If we compared with Chi^2 test, it does

not dependant on the underlying cumulative distribution function being tested, another is that it is an exact test. The Chi^2 goodness-of-fit test depends on an adequate sample size for the approximations to be valid[41]. Box and Whisker plot analysis was applied to this data and gave five outliers, (see figure 2.32). After eliminating the outliers, the mean value obtained is -0.3436 and the standard deviation SD = 3.0198 ppm, aaMMA= 2.528, RMS= 3.039 ppm. The linearity of data was also tested by a Q-Q plot, and five points drifted from the reference line, (see figure 2.33).

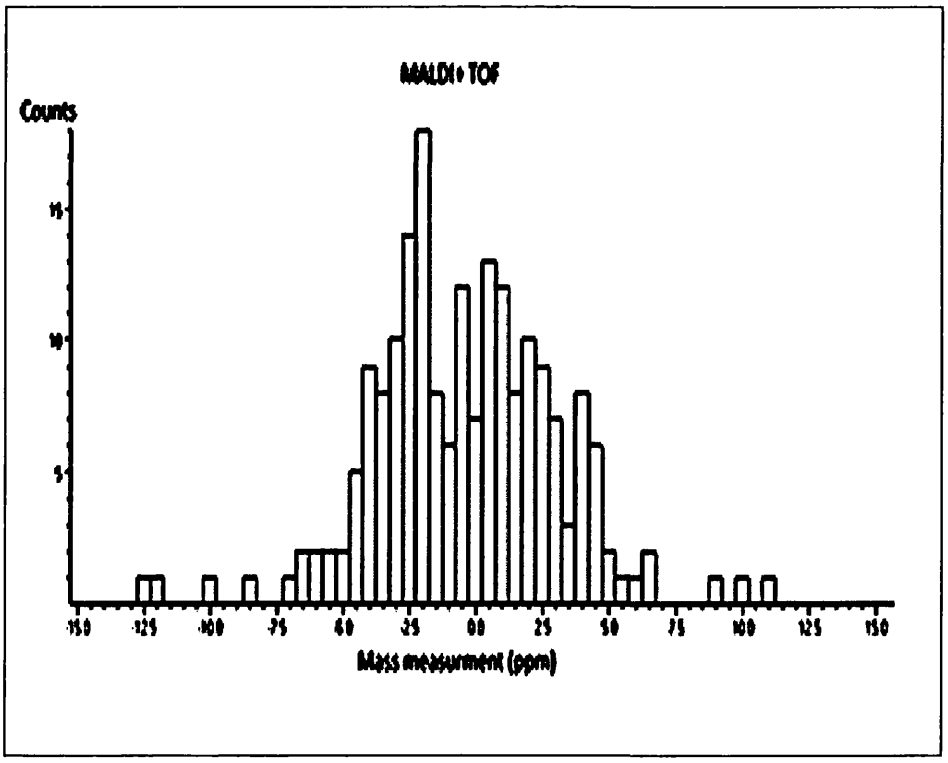

Figure 2. 31 *MALDI-TOFMS histogram illustrating the distribution of 189 set points as a bar graph, several outliers shown, and the majority of data located in 1σ.*

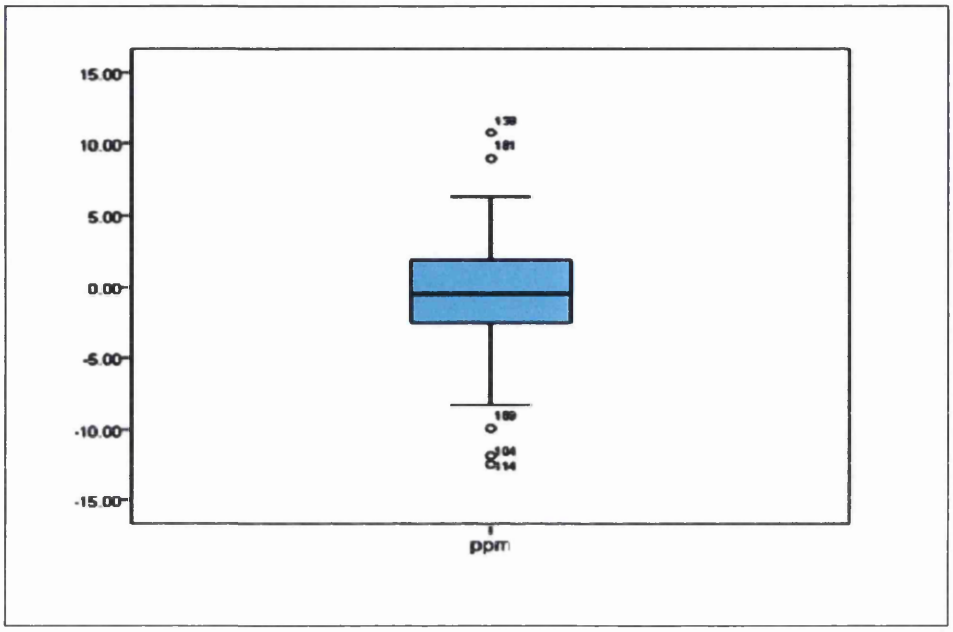

Figure 2. 32 *Box and Whisker plot of 195 set points for MALDI+ TOF techniques, and 5 outliers are present over a wide range on mass scale.*

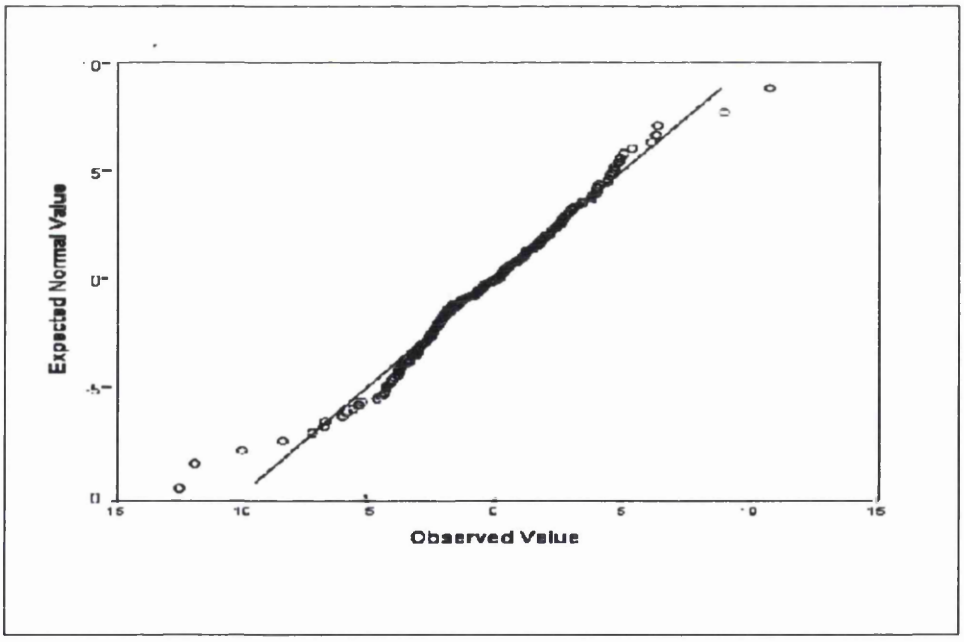

Figure 2. 33 *Shows the normal Q-Q plot for 195 set point of MALDI+ TOF techniques and 5 outliers detected.*

2.5 Comparisons of statistical data as a function of mass

Accurate mass measurements are carried out using a variety of mass spectrometers. Currently, time-of-flight mass spectrometers provide accuracy within 10 ppm[42]. Orbitrap mass measurement accuracies have been reported to be between 2 and 5ppm[43]. Fourier-transform ion cyclotron resonance (FTICR) mass spectrometry, developed by Marshall, et al. currently provides the best mass resolution and mass accuracy (<1 ppm) of all types of mass analysers[44]. Mass measurement accuracy (MMA) at the sub part-per-million (ppm) level, using internal calibration on FAB magnetic sector using peak matching was achieved in this study and sub ppm was achieved using external calibration, on the ESI Orbitrap full scan.

2.5.1 ESI+ magnetic sector instruments

Modern mass spectrometry provides high sensitivity and rapid methods for analysis of high mass molecules, especially with the development of ionisation methods such as electrospray ionisation. Mass analysers have also made dramatic developments over the past decades, with improvements in high performance such as time-of-flight (TOF)[45], electrospray ionisation (ESI), FTICR MS, and Orbitrap. ESI mass spectra have been measured on magnetic sector double-focusing mass spectrometers for different organic molecules. Figure 2.34 displays a plot of the aaMMAs achieved as a function of *m/z* for the magnetic sector instruments. As shown in the figure the instruments fitted with an ESI source give results of superior mass accuracy compared

to the other analyzer investigated in this study. The average absolute mass measurement accuracy (aaMMA) achieve an excellent result over mass range 100-800 Da which is ~0.65 ppm, then recorded about 0.8 ppm at mass range 900 - 1,100 Da. The accuracy got worse over mass range 1100 Da despite the excellent result achieved at mass range 1,300-1,400 Da; this is possibly due to the sample size.

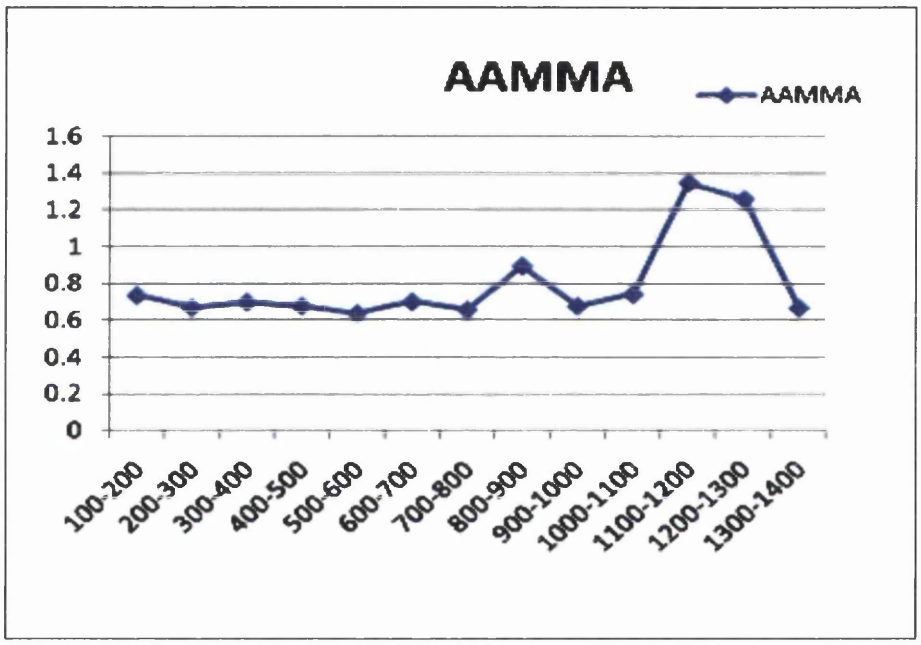

Figure 2. 34 Average absolute mass measurement accuracy (aaMMA) for ESI on a magnetic sector instrument.

The accuracy on ESI+ magnetic sector instrument was centred near zero. The Box and Whisker plot in figure 2.19 illustrates a statistical approach to the data achieved. The figure represented the mean and 95% confidence interval of the mean, (the central tendency and the spread of the data), the line extended from both sides of the figure (the upper and lower) indicate to parametric percentile range (limit of the data), while

the Box indicates the data points cumulated in range of 2σ, and the central line is the mean. The median of the data is represented by the line crossing the Box. As displayed by 95% confidence interval, the outliers observed with TOF data were the result of high or low ion abundances, which appear in the large mass measurement error.

Accurate mass measurements are no longer restricted to double focusing magnetic sector instruments and are now carried out using a variety of mass spectrometers. In recent studies, the FT mass spectrometer offers the highest mass resolution and mass accuracy of all kind of mass spectrometers. TOF, low resolution quadrupole, triple quadrupole, and quadrupole ion trap mass spectrometers have also used for accurate mass measurements[19].

In conclusion the most precise and accurate mass measurements were obtained with magnetic sector instruments with individual mass errors always < 1 ppm, The FAB MAT95 system has the greatest accuracy and precision of accurate mass measurements whilst the MAT95 CI+ records less accuracy but all less than 1 ppm. although these imprison are incompletes as it was not possible to compare each analyzer for all ionization modeles.

2.5.2 ESI+ Orbitrap

The data obtained on the ESI+ Orbitrap is useful to both the novice and experienced user for simple comparison of the capability of each instrument type. Table 2.2

illustrates descriptive statistics for all data sets obtained by different methods over different types of instruments. In the field of Orbitrap, slightly large values for MMA, (aaMMA = 0.978 ppm, with precision = 1.20 ppm), compared to magnetic sector.

Mass accuracy for most of the data is close to the mean with slight deviation. However, the Orbitrap has good resolution and mass accuracy performance close to that of sector instruments. The histogram of the Orbitrap data was found to be of statistically different from the poorer quality MAT95's distribution, as assessed from SPSS statistics this would be expected to improve if internal calibration was employed on the Orbitrap.

2.6 Discussion of results

The primary focus of this study was to investigate and compare mass measurement accuracy for different types of mass spectrometer. Accurate mass measurements were acquired using a range of instruments, including MAT900 XLT and MAT95, fitted with EI, CI, FAB, and ESI sources, with accurate mass measurement controlled by computer peak matching.

Full-scan accurate mass measurements were also obtained for ESI using the LTQ Orbitrap. The K-S test was applied to assess whether the distribution is normal. For EI data sets obtained in positive mode on the magnetic sector K-S = 0.1353, and by

referring to the critical value, which is 0.0438, the test rejects the normality hypothesis.

The aaMMA is a parameter that is easily calculated and one that we have found to be quite informative in terms of the quality of the measurement. A problem with the statistics so far is that mass error may not follow the expected Normal distribution. If mass errors were normally distributed, three standard deviations would capture 68%.

Table 2. 7 *Summary of descriptive statistics of the accurate mass data sets. No = no outliers present in data set, MMA= Mass Measurement Accuracy and aaMMA= Absolute Average Mass Measurement Accuracy.*

Technique	Mass Measurement (ppm)							
	With Outliers				Without outliers			
	No pts	MMA	aaMMA	precision	No pts	MMA	aaMMA	precision
CI+ MAT95	71	-0.108	0.759	0.928	70	-0.076	0.737	0.896
EI+ MAT95	964	-0.152	0.701	0.857	961	-0.151	0.691	0.831
ESI+ MAT95	408	0.074	0.672	0.805	408	No	No	No
FAB+ MAT95	58	-0.03	0.604	0.725	58	No	No	No
ESI+ MAT900	2317	0.031	0.711	0.907	2308	0.026	0.694	0.851
ESI-ORBITRAP	65	0.195	0.999	1.261	65	No	No	No
ESI+ORBITRAP	784	0.139	0.978	1.200	780	0.138	0.964	1.174
MALDI_TOF	195	-0.42	2.719	3.408	190	-0.343	2.527	3.019

of data, 95.5% and 99.7%, respectively. In addition, the Box plot identifies the outliers that are present in the data. Ten outliers were detected over 2,317 set points for the ESI magnetic sector. Four outliers presented for the Orbitrap and five outliers on time-of-flight. On the other hand, no outliers were observed in the FAB+ and ESI+ MAT95 on the magnetic sector instruments.

Orbitrap data gives slightly skewed with an inferior quality histogram. There is a wide spread range on mass scale of TOF (23.18 ppm), but the majority of data falls in the interval ±5ppm. Finally, high accurate mass measurements were achieved on magnetic sector using four different methods. Also the difference in accuracy between magnetic sector peak matching and Orbitrap full-scan was poorer but < 1ppm worse.

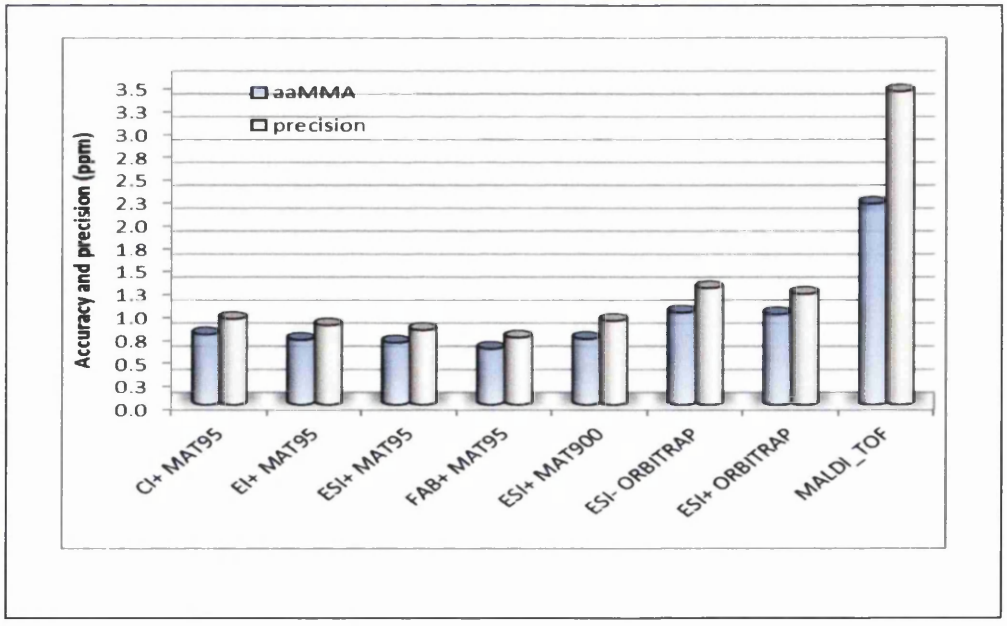

Figure 2. 35 Comparison of Absolute Average Mass Accuracy (aaMMA) and the precision of accurate mass data acquired for each technique.

The excellent precision where all data (even with outliers included) are measured with RMS and precision of ~0.8 ppm (and < 1.2 ppm) regardless of mass range and instrumentation, (see table 2.7 and figure 2.35). (Both RMS and SD generate similar values; the SD was quoted for statistical consistency for precision of the data). Overall the MAT95-FAB system has the greatest accuracy and precision for accurate mass measurements, whilst the Orbitrap –ESI (external calibration) is less, and the MALDI-TOF has the worst. A comparative trend is observed for the accuracy of mass measurement when describe as an absolute value (aaMMA) but is always much smaller in magnitude. Deferent between magnetic sector and Orbitrap in accuracy and precision is due to varying techniques for calibrating each instrument.

Both magnetic sector and Orbitrap technique presented here capable of mass measurements (single measurement) with in an error of 3 ppm of the theoretical mass. Within the scope of the data available the most precise and accurate mass measurements were obtained with sector technology where estimates of the magnitude of accuracy all lie in a range of 0.60-1.0 ppm of the theoretical mass.

2.7 Conclusions

The study summarises the data obtained on three types of instruments; magnetic sector, Orbitrap and time-of-flight (TOF), using different ionisation methods to evaluate the accuracy and precision over the three types of instruments. One can conclude that the protocols used to record accurate mass measurements of small

molecules were appropriate, as good quality data has been obtained after examination by extensive statistical tests. The study involved the accurate mass measurement of mono-isotopic ions from different compound samples, over mass range of 100 to 1,200 Da with no interfering ions present from the sample. At this stage, the data evaluation has been based solely on the mean mass measurement accuracy. Magnetic sector mass spectrometers have traditionally been used to record accurate mass measurements by electron ionisation (EI), chemical ionisation (CI), electrospray ionisation (ESI) and fast-atom bombardment (FAB) of the common mass measurement techniques. Peak matching produced the most accurate results, with mean mass measurements <1 ppm. The results clearly reflect the accepted capability of this technique.

For the Orbitrap, using ESI ionisation methods, the data reported mean mass measurement accuracy of ≤ 1.2 ppm. This result is comparable to that published in the most recent papers, and superior to that recorded using external calibration methodology. The results present the expectation of the capability of Orbitrap for accurate mass measurements.

The capabilities of TOF mass spectrometers have only recently been enhanced to allow accurate mass measurements to be recorded. To summarise the performance of the instruments, magnetic sector field mass spectrometers used in peak matching mode give the higher accurate mass. Orbitrap and TOF instruments generally achieve mean mass measurement accuracy between 1.2 and 3 ppm. The precision of the data

from each instrument and ionisation method was computed in this study. Magnetic instruments tend to give mass distributions which are essentially Gaussian in shape. One paper was published about this comparison[46] and other paper in preparation.

Sack, et al.[47] proposed a method for statistical evaluation of accurate mass measurement quality for double-focusing magnetic sector mass spectrometers. Such statistical understanding is important, as it will reduce the uncertainty when selecting the elemental formulae that result from accurate mass measurement. Another investigation involving a comparison of accurate mass measurement of small molecules was published by Bristow, et al.[24].

Many studies have investigated mass accuracy and confirmed the capability of each type of mass spectrometer to achieve high accuracy. Mass spectrometers of the same type often produce equivalent mean mass measurement accuracy, but with significant differences in precision. The key points for best practice will then be established from this comparison for each type of mass spectrometer and accurate mass measurement technique. The study summarises the data from those various methods to evaluate the accuracy of instrument types mentioned over mass range $\leq$ 1500 Da.

The protocol used is to record accurate mass measurements of small molecules. Magnetic sector field MS has been used over different ion source EI, CI, FAB, and ESI, all employed peak matching technique and reported mean mass measurement $\leq$ 1 ppm, and there was no significant difference obtained. Peak matching techniques

clearly produced the most accurate results, precise and accurate mass measurements with this technology, where estimates of the magnitude of accuracy, all below 1 ppm. The results clearly reflected the accepted capability of sector technology. The statistical analysis for all data was performed using SPSS statistical software, AccMass and DAnET software. The normality of the distribution of the variable measured by Kolmogorov- Smirnov, Gaussian distribution were obtained for ESI magnetic sector, while slightly lower quality for Orbitrap data sets was observed. Orbitrap full scan operation is sufficient for high throughput accurate mass analysis, but with slightly larger estimates of magnitude of MMA. Sub–ppm mass measurement accuracy has been demonstrated for historical data on a magnetic sector instrument using internal calibration, while Orbitrap gave a value for ESI+ data of 1.20 ppm using an external calibration method. Internal calibration is typically at least twice as accurate as external calibration[48], so improvement of accuracy in the Orbitrap would be expected if an internal calibration was used[49]. The effect of ionisation techniques on the MAT95 is not very significant. FAB+ mass measurement accuracies were typically 0.2 ppm smaller than ESI+ and both less than 1 ppm. It is not clear why different ionization techniques have different affects or the value of MA, however it can be speculated that this is due to; i) the physical proportion of the source; *i.e* energy and spatial spread of ion beam; ii) the mechanical tolerance of each sources of each source affecting reproducibility over a large set of disjoint measurements and iii) different cleanliness and aging behaviour as each source is used.

2.8 References

1. Boyd RK, Basic C, Bethem RA. *Trace Quantitative Analysis by Mass Spectrometry* John Wiley & Sons Ltd, 2008.

2. Beynon JH. *Qualitative Analysis of Organic Compounds by Mass Spectrometry. Nature.* 1954; **174**: 735-737.

3. Zhang L-K, Rempel D, Pramanik BN, Gross ML. *Accurate mass measurements by Fourier transform mass spectrometry. Mass Spectrometry Reviews.* 2005; **24**: 286-309.

4. Schmid DG, Grosche P, Bandel H, Jung G. *FTICR-mass spectrometry for high-resolution analysis in combinatorial chemistry. Biotechnology and Bioengineering.* 2000; **71**: 149-161.

5. Smith RM. *Understanding mass spectra: a basic approach* (2). Wiley, New Jersey, 2004.

6. Huguet C, Hopmans EC, Febo-Ayala W, Thompson DH, Sinninghe Damsté JS, Schouten S. *An improved method to determine the absolute abundance of glycerol dibiphytanyl glycerol tetraether lipids. Organic Geochemistry.* 2006; **37**: 1036-1041.

7. Watson JT, Sparkman OD. *Introduction to Mass Spectrometry: Instrumentation, Applications & Strategies for Data Interpretation* (4th). Wiley, 2007.

8. Fishman VN, Linclau B, Curran DP, Somayajula KV. *Tris (perfluoroalkylethyl)silyl alkyl amines as calibration standards for electron ionization mass spectrometry in the mass range of 100-3000 Da. Journal of the American Society for Mass Spectrometry.* 2001; **12**: 1050-1054.

9. Hagenhoff B, Benninghoven A, Barthel H, Zoller W. *Supercritical fluid chromatography and time-of-flight secondary ion mass spectrometry of poly(dimethylsiloxane) oligomers in the mass range 1000-10000 Da. Analytical Chemistry.* 1991; **63**: 2466-2469.

10. *New mass-labelled (PFOS) reference standard. Environment International: Wellington Laboratories,* 2009.

11. Wieser ME. *Pure Applied Chemistry. Atomic weights of the elements 2005 (IUPAC Technical Report).* 2006; **78**: 2051-2066.

12. Moini M. *Ultramark 1621 as a calibration/reference compound for mass spectrometry. II. Positive- and negative-ion electrospray ionization. Rapid Communications in Mass Spectrometry.* 1994; **8**: 711-714.

13. Dass C. *Fundamentals of contemporary mass spectrometry* Wiley Interscience, 2007.

14. National Mass Spectrometry Service Centre. In: EPSRC, Swansea; 2008.

15. Jeannot MA, Zheng J, Li L. *Observation of sodium gel-induced protein modifications in dodecylsulfate polyacrylamide gel electrophoresis and its implications for accurate molecular weight determination of gel-separated proteins by matrix-assisted laser desorption ionization time-of-flight mass spectrometry. Journal of the American Society for Mass Spectrometry.* 1999; **10**: 512-520.

16. Jiang L, Moini M. *Fomblin as a reference compound for negative ion fast atom bombardment high-resolution mass spectrometry in the mass range of 180-3000 u. Biological Mass Spectrometry.* 1993; **22**: 170-175.

17. Bristow AWT. *Accurate mass measurement for the determination of elemental formula - A tutorial. Mass Spectrometry Reviews.* 2006; **25**: 99-111.

18. Bristow AWT, Constantine J, Harrison M, Cavoit F. *Performance optimisation of a new-generation orthogonal-acceleration quadrupole-time-of-flight mass spectrometer. Rapid Communications in Mass Spectrometry.* 2008; **22**: 1213-1222.

19. Hu Q, Noll RJ, Li H, Makarov A, Hardman M, Cooks RG. *The Orbitrap: a new mass spectrometer. Journal of Mass Spectrometry.* 2005; **40**: 430-443.

20. Olsen JV, de Godoy LMF, Li G, Macek B, Mortensen P, Pesch R, Makarov A, Lange O, Horning S, Mann M. *Parts per Million Mass Accuracy on an Orbitrap Mass Spectrometer via Lock Mass Injection into a C-trap. Mol Cell Proteomics.* 2005; **4**: 2010-2021.

21. Miller J, Miller JC. *Statistics and Chemometrics for Analytical Chemistry* (5th). Prentice Hall, 2005.

22. ISO. *Proficiency testing by interlaboratory comparisons. ISO/IEC.* 1996: Guide 43-1, ISO, Geneva.

23. Bereman MS, Lyndon MM, Dixon RB, Muddiman DC. *Mass measurement accuracy comparisons between a double-focusing magnetic sector and a time-*

of-flight mass analyzer. Rapid Communications in Mass Spectrometry. 2008; **22**: 1563-1566.

24. Bristow AWT, Webb KS. *Intercomparison study on accurate mass measurement of small molecules in mass spectrometry. J Am Soc Mass Spectrom.* 2003; **14**: 1086-98.

25. *The National Mass Spectrometry Service Centre at Swansea: EPSRC, editor;* 2008.

26. Committee AM. *Robust statistics: a method of coping with outliers. Royal Society of Chemistry.* 2001.

27. Committee AM. *Analyst.* 1989; **114**: 1489.

28. Committee AM. *Analyst.* 1989; **114**: 1693.

29. Dixon WJ. *Analysis of extreme values. Ann. Math. Statistics.* 1950; **21**: 488--506.

30. Dixon WJ. *Processing Data for Outliers Biometrics, J.* 1953; **9**: 74-89.

31. Barnett V, Lewis T. *Outliers in statistical data. Wiley.* 1985: 365.

32. Snedecor GW, Cochran WG. *Statistical Methods. Iowa State University Pres.* 1989.

33. Kolmogorov AN, Fomin SV. *Elements of the Theory of Functions and Functional Analysis. Nauka, Moscow.* 1989.

34. Joanes DN, Gill CA. *Comparing measures of sample skewness and kurtosis. Journal of the Royal Statistical Society (Series D): The Statistician.* 1998; **47**: 183-189.

35. Yates JR, Cociorva D, Liao L, Zabrouskov V. *Performance of a Linear Ion Trap-Orbitrap Hybrid for Peptide Analysis. Analytical Chemistry.* 2005; **78**: 493-500.

36. Macek B, Waanders LF, Olsen JV, Mann M. *Top-down Protein Sequencing and MS3 on a Hybrid Linear Quadrupole Ion Trap-Orbitrap Mass Spectrometer. Mol Cell Proteomics.* 2006; **5**: 949-958.

37. Makarov A, Denisov E, Lange O, Horning S. *Dynamic Range of Mass Accuracy in LTQ Orbitrap Hybrid Mass Spectrometer. Journal of the American Society for Mass Spectrometry.* 2006; **17**: 977-982.

38. Cox J, Mann M. *Computational Principles of Determining and Improving Mass Precision and Accuracy for Proteome Measurements in an Orbitrap. Journal of the American Society for Mass Spectrometry*. 2009; **20**: 1477-1485.

39. Handerson W, McIndoe JS. *Mass Spectrometry of Inorganic, Coordination and Organometallic Compounds* John Wiley & Sons, 2005.

40. Blom KF. *Estimating the Precision of Exact Mass Measurements on an Orthogonal Time-of-Flight Mass Spectrometer. Analytical Chemistry*. 2001; **73**: 715-719.

41. Engeneering statistics handbook. e-Handbook of Statistical Methods: NIST/SEMATECH; 2003.

42. Wyatt MF, Stein BK, Brenton AG. *Investigation into Accurate Mass Capability of Matrix-Assisted Laser Desorption/Ionization Time-of-Flight Mass Spectrometry, with Respect to Radical Ion Species. Journal of the American Society for Mass Spectrometry*. 2006; **17**: 672-675.

43. Makarov A, Denisov E, Lange O. *Performance Evaluation of a High-field Orbitrap Mass Analyzer. Journal of the American Society for Mass Spectrometry*. 2009; **20**: 1391-1396.

44. Shi SDH, Drader JJ, Hendrickson CL, Marshall AG. *Fourier transform ion cyclotron resonance mass spectrometry in a high homogeneity 25 tesla resistive magnet. Journal of the American Society for Mass Spectrometry*. 1999; **10**: 265-268.

45. Loo JA, Muenster H. *Magnetic sector-ion trap mass spectrometry with electrospray ionization for high sensitivity peptide sequencing. Rapid Communications in Mass Spectrometry*. 1999; **13**: 54-60.

46. Brenton AG, Godfrey AR, Alamri M, Stein BK, Williams CM, Hunter AP, Wyatt MF. *Analysis of large historical accurate mass data sets on sector mass spectrometers. Rapid Communications in Mass Spectrometry*. 2009; **23**: 3484-3487.

47. Sack TM, Lapp RL, Gross ML, Kimble BJ. *A method for the statistical evaluation of accurate mass measurement quality. International Journal of Mass Spectrometry and Ion Processes*. 1984; **61**: 191-210.

48. Marshall AG, Hendrickson CL. *High-Resolution Mass Spectrometers. Annual Review of Analytical Chemistry*. 2008; **1**: 579-599.

49. Scheltema RA, Kamleh A, Wildridge D, Ebikeme C, Watson DG, Barrett MP, Jansen RC, Breitling R. *Increasing the mass accuracy of high-resolution LC-MS data using background ions - a case study on the LTQ-Orbitrap. PROTEOMICS.* 2008; **8**: 4647-4656.

3. CHARACTERIZATION OF ACCURATE MASS MEASUREMENTS (AMM) DATA SETS: OUTLIER AND DISTRIBUTION

3.1 Background

The measurement of data obtained by experimental methods is a very important operation for many analysts. Data should be presented accurately and consistently in relation to each other as required or at critical points in an experiment. Questions on whether all members of a data set are accurate or whether an outlier is present need to be answered in most experimental cases. Outliers can be caused by incorrect measurements, data entry errors or arises from a different measurement which may have a different population to that of the rest of the data. This chapter will discuss methods that are useful for looking at data as a body of information. Methods will be used to discuss whether all individual observations in a data set can be considered a member of that population, and then can be treated by appropriate statistical tools to describe the data. Data from accurate mass measurement obtained on mass

spectrometers are subjected to many statistical methods to identify outliers and describe the distribution of the data around the mean. Outlier detection methods can be divided between parametric and non-parametric methods depending on the data model. The major statistical reasons for detecting outliers is they effect description of that data and the calculated mean could be in error as a result. This causes much concern in small sets of data. Also the outliers can greatly influence the calculated value for the standard deviation, again especially for a small data set. These reasons are very important, especially if the range is used for the estimation of standard deviation. There are many data types which could be involved in outlier detection analysis, such as binary variables, nominal and ordinal types. However, mass measurement data sets follow the numerical data types.

Recently, a few studies on outlier detection have been conducted for large data sets[1]. The main ideas in this chapter are to discuss outliers and the distribution of accurate mass measurements, over the mass scale, and based on the size of data and the number of outliers. The techniques that are generally useful for detection of outlier are based on appropriate and recommended software and will be described below.

3.2 Statistical tools used to examine data sets

Following on from the work presented in chapter 2, further statistical analysis was undertaken to ascertain and classify outliers and the type of "distribution" the AMM data sets form. Three software methods were used for these purposes, which involve different statistical methods and are described below. The comparison of data sets for

all techniques was performed using; basic descriptive analysis as a preparatory step to the normality test and the identification of outliers. Various statistical tools and methods were employed for the detection of outliers and classification of the data distribution[2-11]. Some of these methods will be discussed and compared and the suitable methods established for the analysis of the accurate mass data sets under investigation in this thesis.

3.2.1 Excel and Excel add-ins

Excel is probably the most commonly used spreadsheet for PCs (Microsoft Corp, Redmond, WA, USA). In addition there are a variety of add-ins package and tools which include collections of statistical functions, and Excel has its own Data Analysis Toolpak which was employed in the early stage of the investigation. Excel 2007 software was used to prepare an electronic copy of all data sets obtained from different sources in order to perform statistical calculation using Excel itself or any other software. A spreadsheet was prepared and all data sets were inputted into the Excel spreadsheet. Conversion of data set from Excel format to any other required format could be achieved with Excel using its export function.

Basic data analysis tasks were studied initially with Excel software as a reasonable alternative to using more sophisticated statistical packages for the same task. Initially the use of Excel seemed adequate, providing good results. Excel disadvantages were that it is time consuming, especially for large data set. This restricted analysis and it was difficult to conduct analysis of more than one column at a time, making it

inconvenient to do the same analysis on multi columns. Most of Excel's statistical procedures are part of the data analysis tool pack that appears on the tool menu. This package includes statistics tests for the mean, standard deviation correlation, t-tests and others.

The other feature of Excel is its ability to add specialist software add-ins. An add-ins is a small program that incorporates new functions into Excel. Once installed, the add-in loads automatically every time Excel starts up. Through the study data was also subjected to various tests *i.e.* Dixon's Test, Huber's test, Grubbs' test and Rosner's test, by manually coding formulae into Excel cells. Data was studied by manual (and repetitive) application of those tests, and for Huber's test an iterative procedure had to be created to achieve converge, which was very slow to complete. These manual methods are not suitable for calculation, and consequently a search for convenient software was made. Small software add-ins for Excel were identified and used, which were applied for a particular statistical analysis that was not available as standard in Excel. The add-ins investigated were XL stat, Sigma XL (sigma-XL, Toronto, Canada), Analyse-it (Analyse-it, software, Ltd, version 2, Leeds, United Kingdom) and Robstat.Xl (RSC Analytical Method Committee)[12]. All were found to be suitable software for this study and used to various extents described below.

The Robstat.Xl software is an add-in tool published and reproduced by the analytical methods committee of the Royal Society of Chemistry[12] as a robust statistical method that includes the following functions described in two papers[12,13]. Median Absolute

Deviation (MAD), the derived MADe estimate of standard deviation, SMAD (which returns the mean absolute deviation if MAD = 0 the A15 estimate of the mean, the H15 estimates of mean and standard deviation are included in the Robust Reproducibility calculation. Many problems were encountered in the process of evaluating these software packages including unusual and unpredicted bugs in some of the software packages, and export/import functions truncated or improperly rounding. Sometimes wrong values were found when applying Rosner's test in data organisation and even when Excel was used. The literature recommends using packages, such as SPSS (version 16.0, SPSS Inc., Chicago, IL, USA), or Minitab (Minitab, State College, Pennsylvania, USA).

3.2.2 SPSS and Minitab software

SPSS (originally, Statistical Package for the Social Sciences) is most often used in social science fields. It has an enormous number of statistical features and is widely used by researchers to perform quantitative analysis. The statistical data analysis discusses the uses in assumptions of regression, which include the removing of outliers from data set, the examination of linearity and constant variance. SPSS is a powerful software program. It requires a predesigned data sheet to be organised and the type of variables determined prior to performing any statistical test. Variables fit into two main types, categorical and numerical, and statistic tests are selected accordingly. The software allows assigning specific description to every category within the categorical variable, apart from its input digits used in the data sheet.

SPSS was used mainly for descriptive analysis of data sets for this study. In general, particular descriptions were used, such as the number of samples, mean, median, mode, standard deviation and normality. These parameters were also used to compare all the techniques with each other and similar techniques. Minitab-15 (evaluation) was also used in the analysis simply to recheck the results obtained by SPSS. The comparison considered the accuracy of the results, as well as the ease with which the interface could be used for bigger data sets. SPSS is supported by the university and was used in the work.

3.2.3 AccMass (in-house software)

AccMass statically software is in-house software built in 2009 used to record the mass differences (errors) between the theoretical accurate mass and the observed accurate mass. The software is very simple and the difference (error) is reported in three types of units, Dalton (Da), milli Dalton (mDa), and ppm units (Figure 3.1). AccMass has options to choose preferred units and mass range. A wide mass range for a data set can be applied to imported data from Excel (or by cut and paste from Excel). The number of decimal places can be defined. Descriptive statistics is in software as well as other statistical tests (formal and informal) such as Dixon's and the Kolmogorov Smirnov normality test.

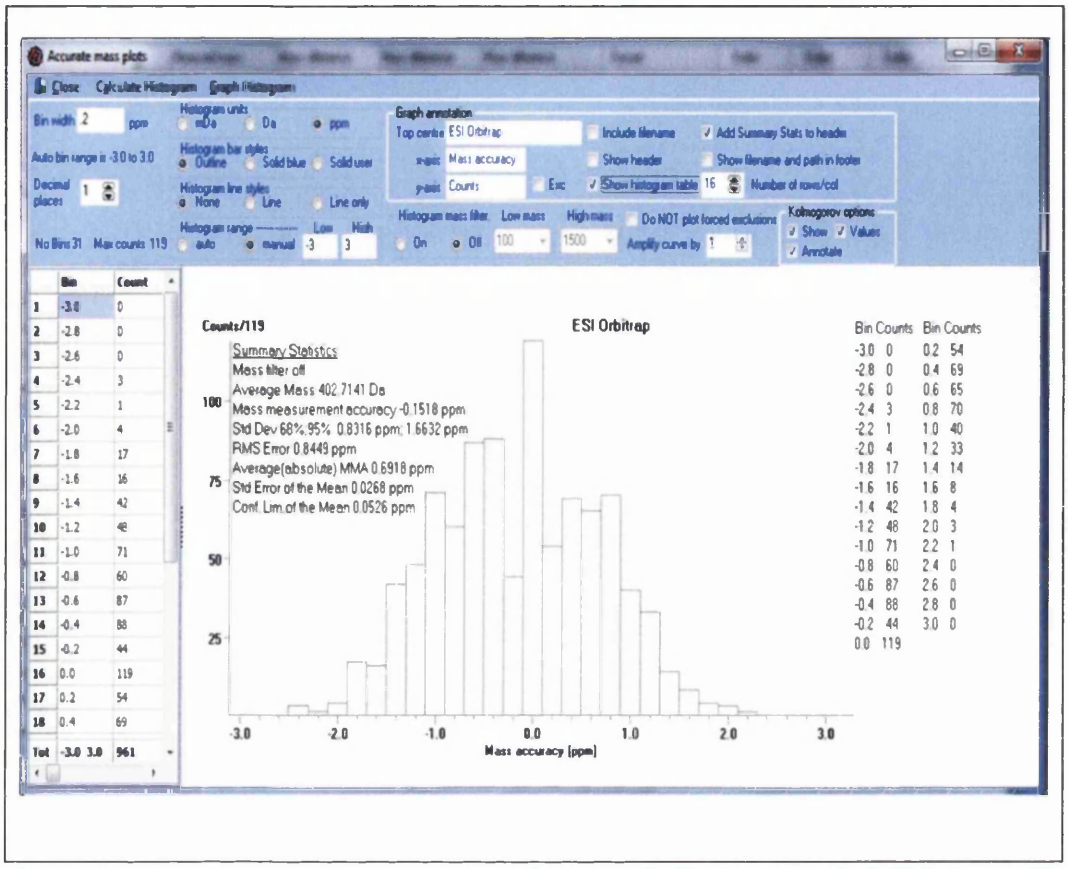

Figure 3. 1 Shows data treatment by AccMass software.

3.3 Methodologies employed to characterise data sets

The first stage in any analysis is to explore the data collected. Descriptive statistics give an overview of the data and enable the researcher to quickly understand and describe a potentially large data set. It is a common test used throughout data analysis to illustrate and characterise the data set under study *i.e* the mean, the median, standard deviation, RMS error, confidence limits of the mean as well as, other features.

These criteria come to show the distribution of data around the mean "parametric analysis". The mode and median can also be used to identify the centre of a data set and describe "non-parametric analysis"[14]. If these parameters are equal to the mean, this indicates a Normal distribution was presented. The mean value is very important in the determination of the accuracy of a measurement and reflects the accuracy of the technique and instruments used. Testing the normality of data set is sought in this study to allow a better understanding of data set characteristics prior to performing any type of outlier identification. The skewness of the data set may be used to indicate the locations of outliers. Even when the data set is normally distributed, the distribution shape provides an indication about the precision of instrumental techniques.

An outlier is defined as an observation that deviates so much from other observations as to arouse suspicion that it was generated by a different mechanism[13,15,16]. Barnett and Lewis[17] indicate that outliers are observations that appear to be inconsistent with other observations in the data set. Outliers can provide useful information about the process and are indicated by the shift from the mean or median value. Handling of outliers can be identified by means of graphical or visual inspection[18]; the visual inspection alone cannot always identify an outlier and may lead to mislabelling good data as outliers. Various approaches are used for outlier detection depending on the application and the number of measurements in the data set. In this study Dixon's, Huber's, Grubb's and Rosner's tests were applied to our data to investigate and compare the ability of these outliers' tests.

Grubbs' and Dixon's tests are used to detect single outliers[19,20]. Dixon's test is based on a ratio involving the range over which data lies, and it is flexible enough to test small samples. It is one of the simpler and easier implemented tests that are usually described in textbooks for analytical chemistry. Miller and Miller[21] mention that in large sets where more than one suspected outlier will sometimes occur, there are techniques and other outlier tests that can be used for large data sets. M-estimators appear to be one of the more appropriate choices given the underlying Gaussian distribution of errors produced[22]. It was used to distinguish the outliers and the form of the distribution by applying SPSS software under the term of outliers test. This robust method was published as add-in software and recommended by the RSC analytical method committee[12]. M-estimators involves iterative procedures designed to find a robust estimate for the first moment and can be extended to estimate the second one successive data points as suspect outliers on the scale[23]. It is possible to get the result of M-estimators using SPSS software. The data was also analysed for Normality by Kolmogorov Smirnov test.

The most common tests used for outlier detection and the useful test for large data sets are the Box and Whisker plot[24]. It is a graphical representation of the data. This test was chosen as a convenient test for this data. The graph represent of the lower quartile (Q_1) (25th percentile) and upper quartile (Q_3) (75th percentile) along with the median is defined as the 50th percentile of the data. The interquartile is the distance between Q_1 and Q_3 (Q_3-Q_1), (see figure 3.2). The upper and lower "fences" are set at 1.5 times

the interquartile range and any observation outside of these fences is considered to be an outlier, even when data are not normally distributed.

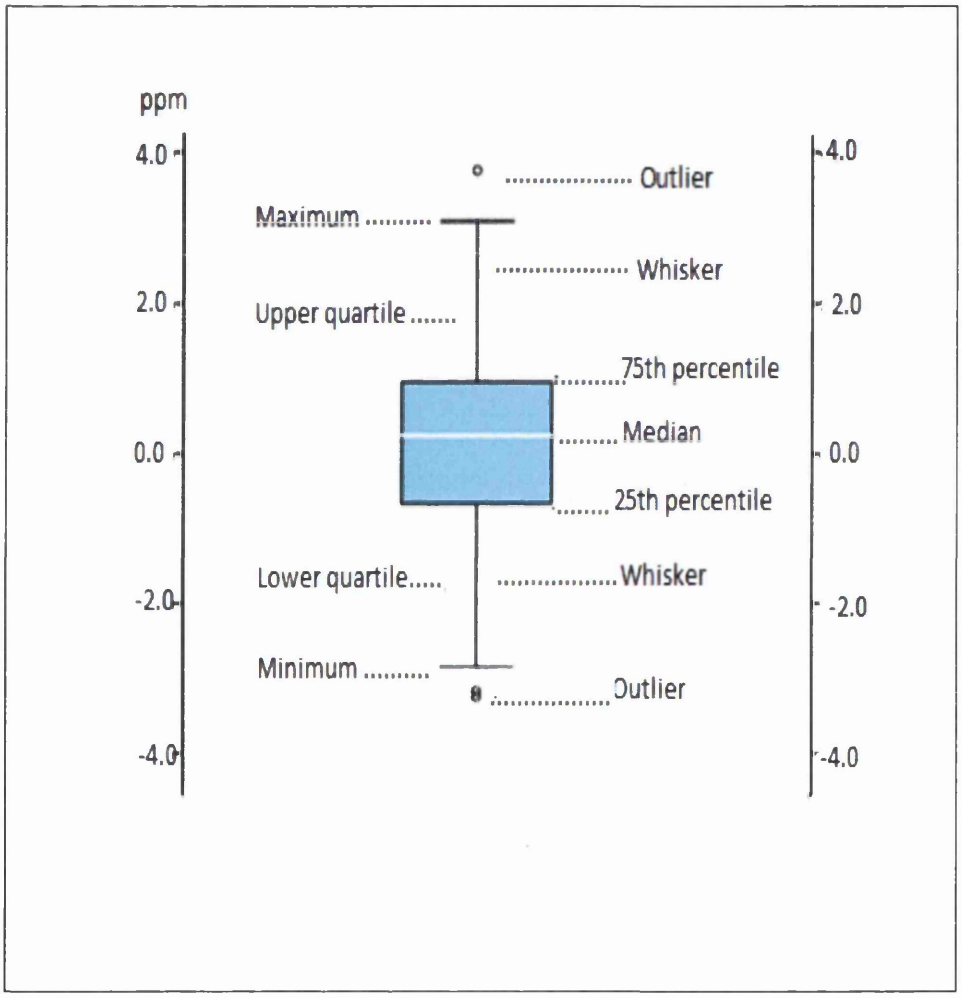

Figure 3. 2 *Box and Whisker plot displaying a statistical symbol summary of data set using only five points (Median, quartiles, range, and extreme values (outliers)). Lower quartile is Q1 and the upper quartile is Q2, then the difference (Q2 - Q1) is called the interquartile range or IQ. The inner fences equal Q1± 1.5*IQ, and the outer fences Q1±3*IQ.(drown by SPSS).*

3.3.1 Parametric and non-parametric methods

One of the first steps in the treatment of data is to calculate the mean, median, and the standard deviation. If the mean is used as a measurement for the data with the standard deviation, both parameters indicate any shift in the data spread. A large, standard deviation value indicates that points are spread out far from each other. A small value reflects tightness of the data and the points lie close together. However, before the above comparison is made, it is required to assume that the data is normally distributed. A plot of the histogram is assumed to be a bell-shaped curve. When the data behaves in this way it's possible to apply the Normal distribution. For example, the mean plus or minus one standard deviation contain about 67% of the data, the mean plus two standard deviation contains 95%, and the mean plus three standard deviation contain 99.7% of data.

A parametric test is a test that requires a parametric assumption, such as normality, whilst a non-parametric test does not rely on a parametric assumption like normality. Data is considered parametric if it conforms to the following criteria: normality, equal variance and independent. Consequently, tests that do not make assumptions about the population distribution are referred to as non-parametric tests. All commonly used non-parametric tests rank the outcome variable from low to high and then analyse the ranks. Non-parametric tests can be selected if the population is clearly not Gaussian manner, or some values of the data are too high or too low to use, even if the population is a Gaussian distribution, it is impossible to analyse such data by a parametric test since we do not know all of the values.

Kolmogorov Smirnov test can be used to test whether the distribution of the data set differs significantly from a Gaussian distribution. The non-parametric statistical methods are less powerful because they use less information in the calculation (only ordinal positions), while the parametric methods use information about the mean and the deviation from the mean. SPSS software has been used to check the skew and the Kurtosis measure for relatively Normal distribution.

3.3.2 Accuracy and precision of all data set

Accuracy and precision are ways of describing data sets and methods that measure or estimate measurements. In these cases there are some parameters that are needed to be known as a true value that can be compared to values to measure the accuracy and/or the precision of the comparison. As with all empirical measurements, a certain amount of error exists between the measured and true values. In mass spectrometry, accuracy and precision are very important information for the analyst. The term accurate mass and exact mass are commonly used for the measured and calculated masses, respectively. The difference between the measured value (accurate mass) and the true mass (exact mass) is the accuracy of the "mass measurement accuracy" (see figure 3.3). The accuracy of measurement in the mass spectrometry field reflects the presence of systematic errors. This means the accuracy refers to how close mass measurements are to expected results. For example, consider a molecular ion $M^{+\bullet}$ that has exact mass assumed to be 400.0000 Da, and the experimental result of the

instrument. The error exists between the measured and the exact mass and could be affected by random noise in the spectrum or systematic error.

To investigate the accuracy and precision, over ~4,500 accurate mass measurements obtained on three instruments types with different ionisation techniques. These obtained by statistical descriptive analysis and graphical tests. Data measurements for every technique were plotted separately as histograms, and allow investigation of the distribution shape. For example, the FAB data set, drawn as a histogram is as expected (from the central limit theorem), and the acquired data is normally distributed based on the result obtained by Kolmogorov Smirnov test (K-S). The statistical K-S value is 0.1607 and experimental K-S value 0.1786. The mean occurs in the middle of the distribution and represents the best estimate of mass accuracy in all measurement data. The standard deviation defines the width of the distribution, describing how much variation occurs between measurements. In this case, two errors occur; first, the mean is not equal to zero as it is in Normal distribution, so it is shifted from the true value. The amount of this shift is (-0.032 ppm) to the left side which indicates the accuracy. The second error is the individual measurements may not agree well with each other, this is represented by the width of the distribution. This error is called the precision of mass measurements and is expressed by quoting the standard deviation (SD).

Consider a measurement that has good accuracy, but poor precision; the histogram is centred over the true value, but is very broad. Although the measurements are correct

as a group, each individual reading is a poor measure of the true value. This situation is said to have poor repeatability; measurements taken in succession don't agree well. Poor precision is a result of large random errors. This is the name given to errors that change each time the measurement is repeated. Averaging several measurements should improve the accuracy. In short, precision is a measure of random noise. Accurate mass measurement for assessments of molecular formula is non-ambiguous up to 300 Da due to the small number of putative molecular formula when a limited of member of elements are known to be involved (C, H, N, O)[25].

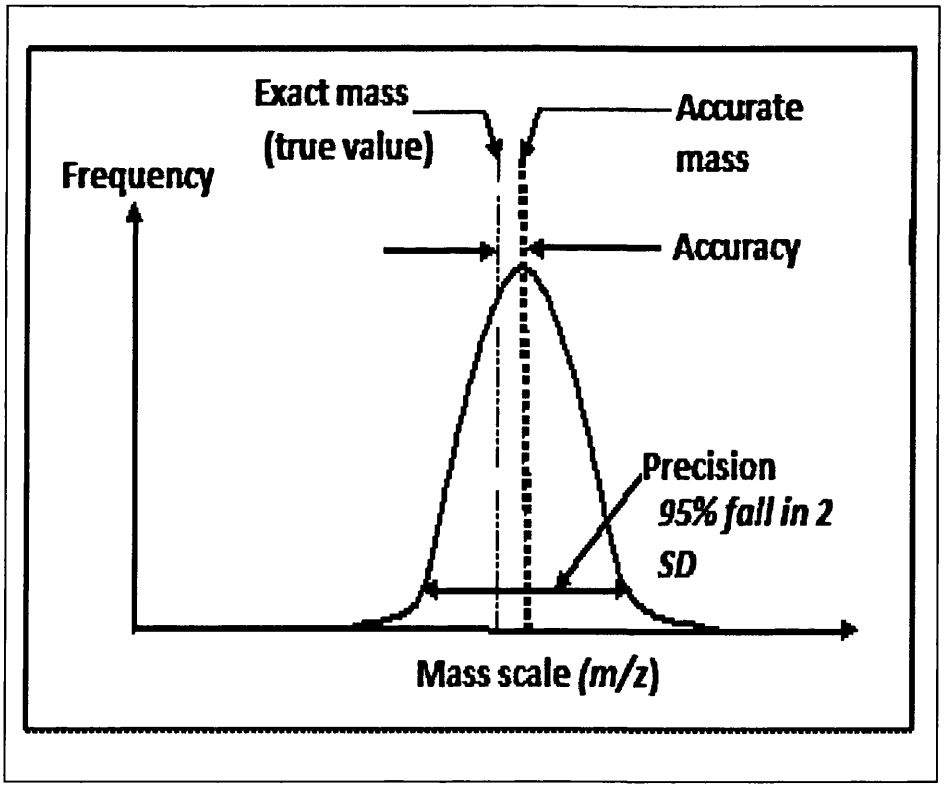

Figure 3. 3 *Definitions of accuracy and precision. Accuracy is the difference between the true value and the mean of the underlying process that generates the data. Precision is the spread of the values, specified by the standard deviation.*

A variety of techniques established more than forty years ago to achieve high accuracy used double-focusing magnetic sector mass spectrometers. Bristow, et al.[25] have investigated mass accuracy for magnetic sectors using peak matching, dynamic voltage scanning and dynamic magnetic scanning and the results are described in detailed. More recent work on accurate mass is not restricted to magnetic sector, but it is carried out using a variety of mass spectrometers. The highest mass accuracy and mass resolution among mass spectrometers is afforded by Fourier transform mass spectrometry (FTMS)[26]. FTMS is routinely used for accurate mass with resolving power >106 and mass accuracy < 1 ppm. The Orbitrap has a high mass accuracy (1-2 ppm), a high resolving power (up to 200,000) and reasonably good mass range (5,000). Time-of-flight (TOF) instruments are widely used and readily coupled with chromatography, but can suffer from mass drift due to temperature change and from lower mass accuracy, usually about ± 5 ppm, when compared to the other techniques mentioned above.

In this investigation accurate mass measurement were taken in CI+, EI+, ESI+ and FAB on a MAT95, and further ESI+ on a MAT900 magnetic sector instrument. CI+ on a magnetic sector, and after omission of outliers. The aaMMA is 0.7372 ppm with a standard deviation (SD) = 0.896 ppm and an RMS = 0.893 ppm, over a mass range (151-998 Da) for 70 mass measurements.

Referring to table 2.3 in chapter 2, the EI positive mode, with 961 measurement, has an average mass AMM= 402.714 Da, aaMMA = 0.691 ppm, SD= 0.831 and RMS =

0.844 ppm. For ESI+ on the MAT95 aaMMA obtained is 0.672 ppm, SD = 0.805, and RMS 0.808 ppm. Whilst FAB gives the highest accuracy, aaMMA = 0.604 ppm, SD = 0.725, and RMS= 0.719 ppm. On the other hand ESI+ MAT900 gives aaMMA = 0.694 ppm, SD= 0.851 and RMS also give 0.851 ppm. Overall, when comparing ionisation methods, FAB were found to give the highest accuracy and highest precision, while CI+ MAT95 and EI were joint worst were the least effective methods.

Orbitrap data obtained in ESI positive mode using full scan were 781 mass measurements accuracy over the mass range 131-1566 Da and have an average mass measurement (AMM= 445.941 Da). Statistical analysis gives aaMMA = 0.964 ppm, SD = 1.174 and RMS = 1.181 ppm. In ESI negative mode 65 mass measurements were acquired with AMM = 498.5575 Da, the aaMMA is 0.999 ppm, RMS = 1.267 ppm.

MALDI time-of-flight, presented 194 mass measurements taken over mass range (246-1158 Da) absolute average mass measured accuracy was aaMMA = 2.917 ppm, SD = 3.841 and RMS = 3.836 ppm.

It is important to note that internal calibration was applied with magnetic sector peak matching, while external calibration used with Orbitrap full scan mode. The effect of ionisation techniques on the MAT95 is not significant and all methods give accuracy of less than 1ppm.

3.3.3 Comparison of the data distributions to the Normal distribution

Statistically, three types of probability distributions are common for random data, binomial, Poisson and Normal distribution. The Binomial and Poisson are used with discrete random variables, and the normal for continuous random variables. For accurate mass data, as a continuous variable, its probability distribution is usually expressed as a formula that can be used to find the probability that the variable will fall in a specified interval. The graph of the Normal distribution can be described by two factors, which are the mean and the standard deviation. The mean of the distribution determines the location of the centre of the graph defines the accuracy. The standard deviation determines the width of the graph, and indicates the precision of the data set. All data were subjected Kolmogorov Smirnov test which called (K-S test). This allows us to determine whether the results are consistent with the expected outcome. Some data involved one or more of outlier which may present some skewness of the data distribution. Discarding outliers from data adjusted the distribution curve and gives an improved distribution and best results among the three instruments.

3.3.3.1 Chi-squared test (Chi^2)

Chi-square is an interesting non-parametric test that allows us to determine whether the observed data distribution is similar to the expected distribution[27,28]. It can be used to test that the obtained data are distributed normally or not. This type of Chi^2 test is used for goodness of fit, while there is another type Chi^2 test used for independence.

As mention above, the test for goodness of fit compared the frequency distribution of accurate mass measurement observed to the frequency of Normal distribution that was predicted by the null hypothesis (H_0), but it seems not the appropriate test. The quantity Chi^2 in the simplest case is computed according to the equation[21]:

$$\chi^2 = \sum_{i=1} \frac{(O-E)^2}{E} \qquad (3.1)$$

Where O is the observed value and E indicates the estimated value. The hypothesis is that there is no difference between the expected and the observed frequencies. Large values of Chi^2, greater than the critical value (obtained by software or on statistical table), mean the hypothesis is rejected. The critical value for this test can be read from statistical books (statistical table).

The data is organised according to their values in increasing order, starting from the lower value, then the frequency is calculated. Chi^2 must be a positive value as given in equation 3.1 and it will not be symmetrical rather being skewed, and the tail, of course, will be on the right side. The hypothesis test will be on the tail, with rejected region on the right side if it's above the critical value. Since the calculation involved using statistics to test for normality was relatively complicated it cannot be used for accurate mass measurements and the K-S test will be used.

3.3.3.2 Kolmogorov-Smirnov test (K-S)

Kolmogorov–Smirnov test is another method of goodness of fit tests that is used to determine whether two data sets differ significantly. K-S test is a non-parametric statistic, and one of the attractive features of this test is that the distribution of K-S test does not depend on the assumption about the distribution. It is used for comparing two empirical distributions, defined by the largest absolute distance between the two cumulative distribution functions as a measure of disagreement. The K-S test is based on the maximum distance between these two curves or an assumption that the ideal curve is the Normal distribution..

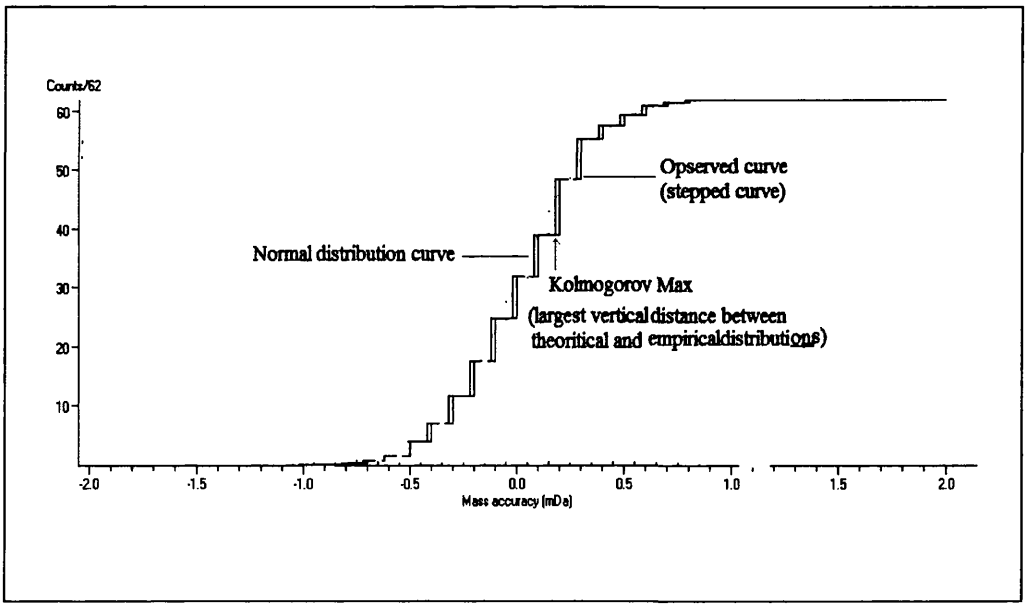

Figure 3. 3 *Representative a plot of the empirical distribution function for ESI+ MAT95 data set with a normal cumulative distribution function. I) the stepped data is the cumulative experimental data, these are two stepped plot each for different mean. II) the ideal curve, the continues dashed line, is the cumulative curve representing the Normal distribution, also called Erf function.*

This means that if the observed data depart substantially from the expected distribution; the two functions will be widely separated over part of the diagram. If the data is closely in accord with the expected distribution, the two functions will never be very far apart[21]

The hypotheses regarding the distribution form is rejected if the K-S value is greater than the critical value obtained from table or software. This means if the test is not significant ($P < 0.05$) it indicates that the sample is not significantly different from the Normal distribution. If the difference is significant ($P > 0.05$) then the distribution is significantly different from the Normal distribution. With a large sample it is easy to get significant result from small deviation from the normality.

3.3.3.3 Application of outlier tests to the historical data sets collected

In this chapter the selected outlier methods are applied for all data which are obtained on the three types of mass spectrometry instrument. This was performed by SPSS statistical software (version 16.0), DAnET, and AccMass as well as Microsoft Excel and add-ins tools. In type one of instruments, mass measurement errors were obtained on magnetic sector MS using deferent ionisation methods and its distribution is narrow, but its differ from one to another based on normality test. EI+ MAT95, ESI+ MAT95, ESI+ Orbitrap, ESI+ MAT900 and MALDI-TOF do not follow the Normal distribution based on K-S test even after omitted outliers, (see table 3.1).

Table 3. 1 Statistical and critical values for Kolmogorov-Smirnov test over all instruments. The cell (normal) indicate to the result of the test if it's normal (yes) or not normal (no). The critical value were calculated at 5% level, (p=0.95).

		Kolmogorov-Smirnov values			
Method	**N**	**Incl. outlier**	**Excl. outlier**	**critical value**	**Normal**
CI+MAT95	71	0.1462	0.1442	0.1614	Yes
EI+MAT95	964	0.1353	0.1399	0.0438	No
ESI-Orb.	65	0.0898	0.0898	0.1687	Yes
ESI+ MAT95	408	0.1219	0.1219	0.0673	No
ESI+ MAT900	2317	0.0709	0.07295	0.0283	No
ESI+ORB.	784	0.1103	0.1124	0.0486	No
FAB+MAT95	58	0.1607	0.1607	0.1786	Yes
TOF	195	0.3511	0.295	0.0974	No

Type 2 of instruments, mass errors were obtained on Orbitrap using electrospray (ESI) method in both positive and negative mode, and its distribution is mildly skewed. On type 3 data sets obtained by time-of-flight (TOF) using matrix-assisted laser desorption ionisation (MALDI), in positive mode. The TOF distribution has a highly skewed and the broader distribution of errors among the three instrument types, because of most observation have slightly large value and five extreme outliers, (see figure 3.4). The modified data sets exclude outliers to see the effect of outliers on the distribution.

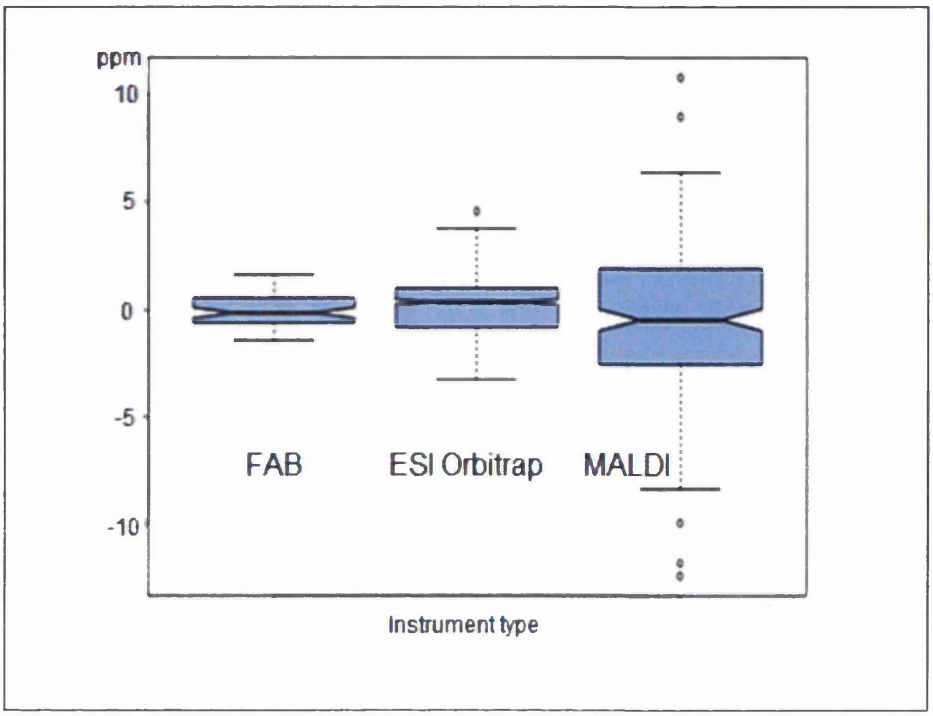

Figure 3. 4 *Comparison of MMAs distribution between three instruments types; magnetic sector, Orbitrap and time-of-flight*

Figure 3.4 shows Box plot comparison between FAB sector, ESI Orbitrap and MALDI-TOF presented by DAnET package as a powerful statistical programme. The graph presented the outliers and comparing the distribution around its mean. There are no far outliers, and just one was presented on ESI+ Orbitrap, five outlier shown MALDI-TOF, and no outlier presented on FAB sector instrument. Visual comparison is enhanced and from the empirical distribution of data, one can easily notes that the MALDI-TOF is the wider distribution and the larger outliers whilst FAB sector is the narrowest distribution among the three types of instruments.

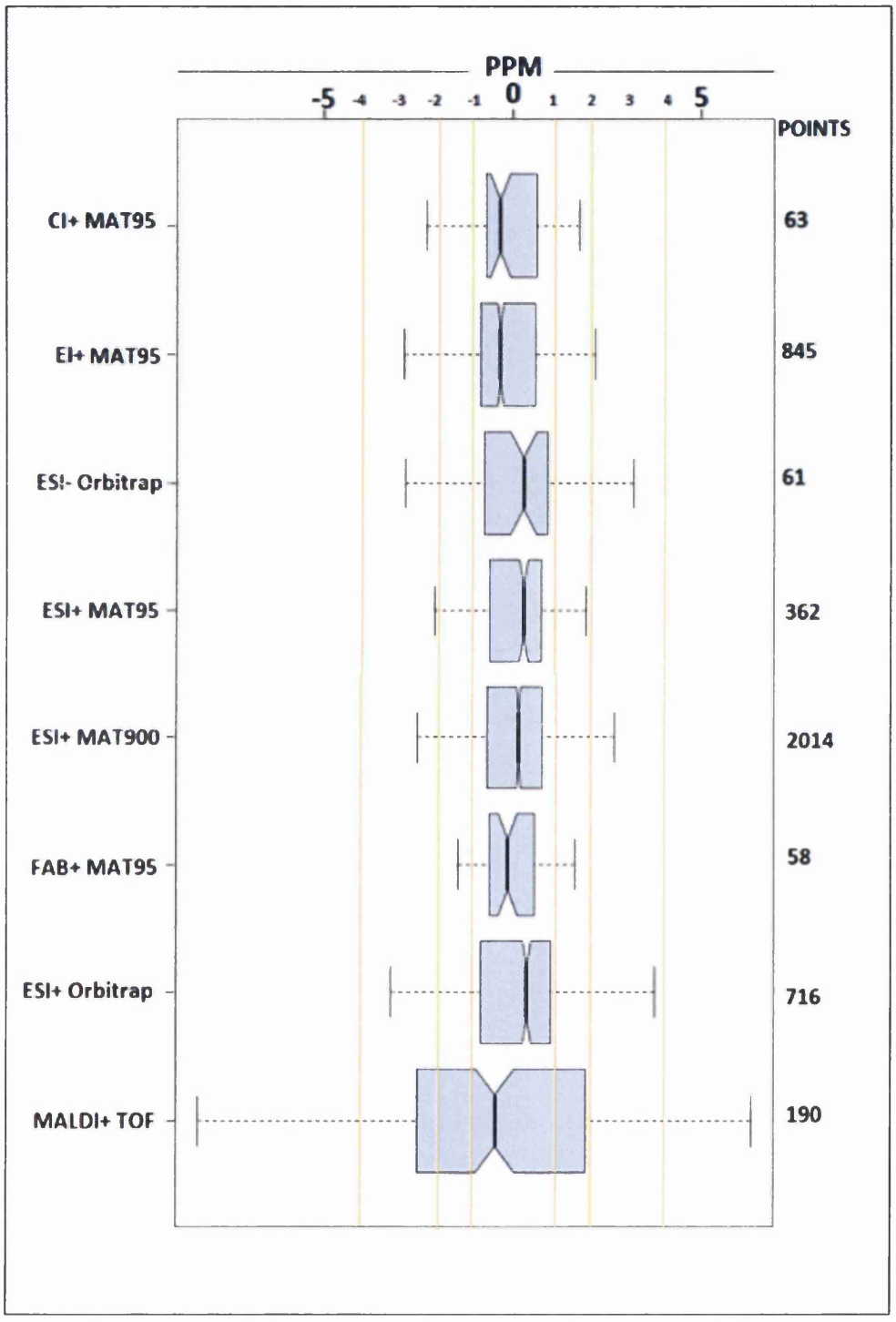

Figure 3. 5 *Represent the graphical test for each data set after omitting the outliers, number of mass measurements are shown on the right side, MMAs measured in ppm.*

Here the Box and Whisker plot demonstrates the distribution of all data sets after omitting outliers. As shown in figure 3.5, the majority of data falls in the interval 1 to -1 ppm (-1< errors < 1 ppm) over all techniques. The Boxes (interquartile range) measure the statistical dispersion of data. It displays the median, the 25^{th} and 75^{th} percentile as lower and upper edges of Box surrounding the median. The Box length represents the interquartile range of the measurement. The outlier is shown below which falls between 1.5 and three Box lengths from the edge of the Box, while the extreme values are usually more than three Box lengths. The Box plot allows a preview to be viewed about the data distribution. The median relates to the central value and the length of the Box indicates to precision (range) of the data sets. If the median is not in the middle, the preview will be skewed.

Over all, the Box plot is less affected by the extreme value than standard deviation and the interval value of SD method becomes much narrow after exclusion of outliers.

3.4 Data pre-treatment outlier identification

Data pre-treatment is necessary to ensure that the data used in modelling, monitoring and control activities provides an accurate representation[29]. Repeated measurements of the accurate masses molecular ions ($M^{+\bullet}$) or the protected molecules ion species e.g. $[M+H]^+$ leads to an improvement in the accuracy of mass measurements by taking the average of all measurements. Typically, if the average mass is close to the expected mass, this expresses the accuracy of the instrument.

3.4.1 Origin and identification of outliers

There are various origins of outliers. Human error often produces unintentional outliers. This may arise due to one of the following causes: the measurements are recorded on the sheet incorrectly[30], entered into the computer incorrectly; the measurements come from a different population due to a sampling error, instrument error, or the measurement represented is not measured accurately due to a measurement problem with the instrument. An observation is named an outlier when the measurement is very different from the others in a data set, or sometimes the single value seems normal but does not fit with the data structure.

3.4.2 Mean and standard deviation

The measurement of accurate molecular mass by mass spectrometry is an important step in the identification process of small molecules over a range of applications. In order to rely on the measured values, it is important to know the performance of the mass spectrometer for accurate mass measurement. Mean and standard deviation indicates the accuracy and precision of mass spectrometric data, respectively. The mean of the population distribution is indicated by μ and is the statistician's jargon for the average value of measurments[31]. The closeness of the mean to the theoretical mass reflects the accuracy of the technique used. This means that most of observed data is around the theoretical mass. Therefore, this parameter is very important in identifying the accuracy of compared techniques, as in this study, and is the mean mass accuracy (MMA);

$$MMA = \frac{\sum_i (m_i)}{n} \tag{3.4}$$

mi is the mass measurement over readings.

Definitions of the statistical values calculated to measure the magnitude of the mass measurement accuracy are:

Average absolute mass measurement accuracy (aaMMA)

$$aaMMA = \frac{\sum_i (|\Delta m_i|)}{n} \tag{3.5}$$

The standard deviation of the population is donated by σ in calculated as

$$SD = \sqrt{\frac{\sum_i (m_i - \bar{m})^2}{n-1}} \tag{3.6}$$

The expression, $|m_i - \bar{m}|$, describes how far the i^{th} sample deviate from the mean.

To support statistical analysis of mass errors 4856 mass measurements made over different techniques were analysed. In this study, the mean mass accuracies and standard deviations were determined for every technique. Excellent mass accuracy was obtained on magnetic sector instruments using internal calibration (sub ppm) with standard deviation SD=0.604 ppm and < 0.32 ppm for ESI, while the Orbitrap gave a MMA for ESI+ of 0.1399 ppm, using an external calibration method with precision at

1 standard deviation =1.20 ppm. On the other hand MALDI-TOF gave a mean measured accuracy MMA = 0.469 ppm and was the worst mass accuracy found between the three types of instruments.

3.5 Tests employed to identify outliers

Published literature contains a number of hypothesis tests for outlier. Some of these tests are especially for single outliers with others that can be employed for multiple outlier detection. Some are valid for large sample sizes and others are specific for small sample sizes.

3.5.1 Dixon's test

Dixon's Q-test is one of the simpler tests and more popular than other more versatile outlier tests. The test allows us to examine if one (and only one) observation from a small set of replicate observations. Dixon's test assumes that the data is a normal (Gaussian) distribution. The test is passed on the calculation of experimental Q value, also known as R_{10} in Dixon's notation, which is defined as the ratio of the distance of suspected value from its nearest neighbour by the range of the values.

$$Q = \frac{x_2 - x_1}{x_n - x_1} \tag{3.7}$$

If n is the sample size, the corresponding n values are arranged in ascending order: x_1, $< x_2$, x_3..... $< x_n$. Then for testing the smallest value (x_1) or the large value (x_n) the following equation is used respectively:

$$Q = \frac{x_n - x_{n-1}}{x_n - x_1} \tag{3.8}$$

If the value obtained exceeds the tabulated critical Q-test, designated as the critical value for a given confidence level *i.e* 95%, then the suspect value can be rejected with a probability of 5% or less ($P \leq 0.05$). Finally the test is valid only for smallest sample sizes (up to 25 points) as mentioned in the literature; consequently it's not useful for our data.

3.5.2 Grubbs' test

Statisticians have several ways to detect outliers. One of the most useful and popular test is Grubbs' test. This test is particularly easy to follow. It can be used to detect outliers in a small or large data set, either creating a new variable (equal to 1 if the observation is an outlier and 0 if otherwise) or dropping outliers out of the data set. Grubbs test is also known as the maximum normalised residual test. Grubbs test detects one outlier at each iteration. This outlier is eliminated from the data set and the test repeated iteratively until no outlier is detected. This test is based on the assumption of a normally-distributed population. It can be applied for sets of data that consist of three measuring up to ~150 measurements. The first step is to determine the outliers by this method and to quantify how far a putative outlier is from another data point. The mean and standard deviation for the full data set must be calculated and used to compute if the upper and lower data values are outliers, and are estimated of

the corresponding population parameters μ and σ, respectively. In the Grubbs' test a standardised parameter, z, is employed and is defined:

$$z = \frac{x - \bar{x}}{SD} \tag{3.9}$$

Where x: data value, and x: mean, SD; the standard deviation, the hypothesis for this test follows:

H_0: All sample points are from the same population

H_A: The most extreme sample point in the sample is unlikely to have come from the normal population from which the remainder of the sample points were drawn. Rejection of the null hypothesis suggests that the extreme sample value tested is outliers.

3.5.3 Rosner's test (R)

Rosner's test is another statistical test that can be used to detect multiple outliers. The test can be used when the number of data points is 25 or more. It identifies outliers that are both high and low; it is always two-tailed[32]. Data must be ranked in ascending order and the mean and standard deviation determined at each iteration. This test is applicable to a Normal distribution. The procedure entails removing from the data set, the observation, x, that is farthest from the mean, it is an iterative approach for multiple outliers, and then a test statistic, R, is calculated. If all points from the same

Normal distribution then the null hypothesis is H_0 or H_A. If all points are not from the sample population it indicates the presence of outlier. The calculation method presented computes the mean $\overline{x_1}$ and the standard deviation SD_1 then the sample x_1 that deviates from the mean is identified. Rosner's test computes R as follows:

$$R_1 = \frac{|x_1 - \overline{x_1}|}{SD_1} \tag{3.10}$$

x_1 can be the largest or the smallest value of the sample.

The sample x_1 is removed from the data next the mean $\overline{x}_2$ and SD_2 is computed from the remaining data then R_2 is calculated, for the points of the data subsequent values of R are computed as follows for the m^{th} iteration:

$$R_m = \frac{|x_m - \overline{x_m}|}{SD_m} \tag{3.11}$$

If the resulting value for the testing R_m is less than the critical value obtained from statistical tables, then it's not outliers, and if it's more than critical value it is an outlier.

3.5.4 Box and Whisker plot

The Box and Whisker plot (or Box plot) is a robust statistical method because it's more resistant to the presence of outliers. The method was described in chapter 2 and under the term of methodology in this chapter. Here we need to compare the histogram and the Box plot distribution, as it's used throughout our investigations. The histogram, as looking at a statistical distribution, is more intuitive than looking to the Box plot. Comparing the Box plot against the theoretical histogram for Normal

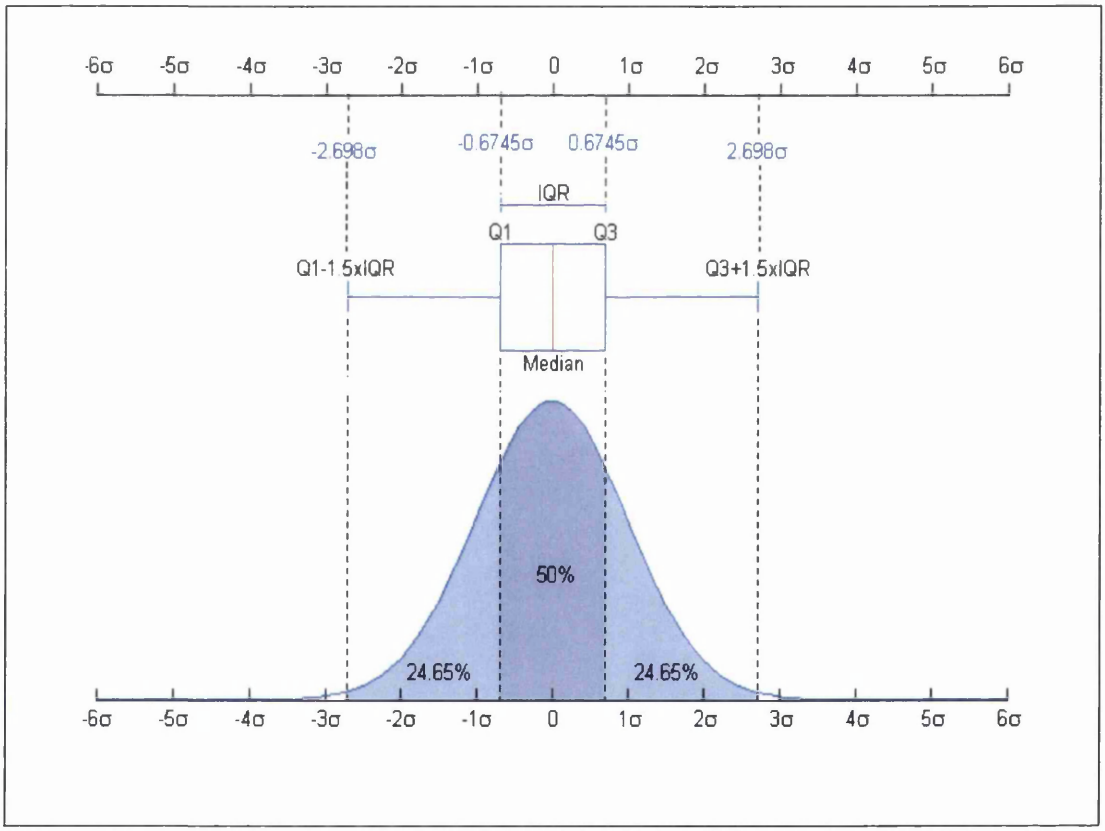

Figure 3. 6 *Correspondence the two rules of outlier detection for Gaussian distribution. Box plot and approvability density function of Normal distribution, the data restricted between ±2.698σ for Box plot, whilst histogram spread the data depending on sigma value. The graph is downloaded from the Wikipedia website[51].*

distribution is the best way to understand how Box plot work. The Box plots are uniform in their use of the Box: the bottom and top of the Box are always the 25th and 75th percentile (the lower and upper quartiles respectively), and the band near the middle of the Box is always the 50th percentile (the median), (see figure 3.6). The measurement more than Q_3 +1.5 IQR or less than Q_1-1.5 IQR it will consider to be an outlier.

3.5.5 Quantile Quantile (Q-Q) plot

Q-Q is a probability plot, displayed as a graphical method for comparing two probability distributions by plotting their quantiles against each other. The test is commonly used to compare a data set to a theoretical model, or expected quantile from a Normal distribution[33]. If x and y are identically distributed variables, then the plot of x-quantiles versus y quantiles will, of course, be a straight line configuration with slope 1 pointing towards the origin[34]. The points in the Q-Q plot will roughly lie on a line. If the two distributions, being compared, are identical, The Q-Q plot follows the 45° line y = x. If the two distributions agree, after linearly transforming the points, they should fall approximately along the reference line, but not necessarily the line y=x. The greater the departure from the reference line, the greater the evidence for the two data sets come from populations with different distributions. A Q-Q plot of the x and y samples is then just a scatter plot of qv (probability (p)) versus qx (p) for various p, corresponding to any abscissa value q and the graph give a visual

perspective for the location, scatter and skewness of data sets compared to test distribution.

3.6 Discussion of treated data

Typically all data was subjected to tests, such as Dixon's, Grubbs' or Rosner's tests to identify outliers at a particular confidence level. This procedure is not necessarily straightforward. Firstly, variance can sometimes be misleading to use if two or more outliers are present. Secondly, it has to be decided whether to exclude the outlier during the calculation for further statistics. Another procedure is a graphical test, which is the best way for data exploration. As usual, the first step of outlier detection is to plot the data using one or more of these graphical methods, such as a histogram, Box plot or Q-Q plot. Small amounts of data were tested manually assuming a Normal distribution, and tested if the extreme target value is an outlier of the distribution or tested if it deviates from the central values. Some of the tests are for a single outlier and other for multi-outliers. For small data sets Dixon's test is a good method as it's a recommended test in many studies. It does not require any distribution assumptions, but it is not useful for large data sets (use for only <25). Grubbs's test was also applied manually using Microsoft Excel software. The test is easy to use, but it is time consuming, and both tests are used for only one outlier. Rosner's test was applied manually, using Microsoft Excel and through the analysis process there were some errors. Despite repeated analysis the same errors still appeared and we assume that it is a bug in the Excel software.

SPSS and DAnTE statistics software involve packages of statistical tools for data treatment and analysis. Both these techniques have "Box plot" and "Q-Q plot" for testing normality, distribution and outliers. They present information that shows how the data distribution forms and assesses the degree of skewness, and how many outliers are involved.

3.6.1 Outliers

Observed data often contain outliers that have unusually large or small values that seem different when compared with other data in the set. It may be considered as an error or noise that is distinct from the main body of the data, and are incompatible with the rest of the data. These values may be genuine observations from individual measurements with very extreme levels of the variable. However, they may also result from typing errors or the incorrect choice of units, and so any suspicious values should be checked[35]. In this study any point that is more than three standard deviations will be considered an outlier. It was noted earlier in this chapter that the mean or the median value is the central tendency of the distribution. The outliers lie very far outside the central part of the distribution which are high in value and can inflate the mean, or are very low in value that may deflate the mean[36,37]. The three standard deviation definition for an outlier was primarily chosen since there is no determined point which robust estimators start to reduce the influence of outlier on the estimated parameters. However, outlier tests only start to reject outliers once the first defined threshold point of an outlier has been observed. Since it is generally

considered that any point greater than three standard deviations is an outlier, and it is likely that robust methods have significantly reduced the influence of residuals of the magnitude, then the three standard deviation definition is chosen. Detection methods are available for use in the univariate case, but methods for identifying outliers in multivariate data are limited38. Also some methods are used only for one outlier, and different outlier tests do not always yield the same results. In this study there are large data sets. Here the multivariate methods have been used which are recommended by many studies.

Dixon's test is one of the most popular tests for the detection of outliers because it is very easy to calculate. It depends on a comparison of the difference between the suspect value and the close value with in a defined overall range. This test is not used in this study because it's used only for small data sets, generally less than 25, but not for more than 100 data points. The recommended test used for large data sets is Grubbs's test, which described at the beginning of this chapter. The test is repeatedly performed (iteratively) until no more extreme values are detected; if the outlying observations are significant at the 5% level it's then discarded. The maximum normalized deviation test described by Grubb is based on the calculation of:

$$T = \left(x_i - \bar{x}\right)/s \qquad (3.12)$$

Where T is Grubbs statistic, x_i is the suspected outlier (highest or lowest), the sample mean, and s the sample standard deviation. This test is used for a single outlier. Rosner's test is another recommended test for multiple outliers. To use Rosner's test,

an upper limit m must be specified on the number of potential outliers present. The test is valid for $n \geq 25$ values; the test can be calculated based on equation (3.11).

Graphically, normal probability data will form a straight line, as a Q-Q plot. Data that cannot be fitted to the line can be considered as representative of another distribution. Several of the hypothesis tests assume that the data is normally distributed. As mentioned in chapter 2, the Normal distribution is completely characterised by the mean and standard deviation[39]. This graphical Q-Q test is used to compare data sets to particular probability distributions or to compare with another data set. The idea is that if two population distributions are exactly the same, then they have the same quintiles (percentiles), so the plot of the quintiles will fall on the 0-1 line[40]. The Box and Whisker plot (simply called Box plot) is a well-known, simple display of the five-number summary (lower extreme, lower quartile, median, upper quartile and upper extreme)[41]. It's most suitable for exploring both symmetric and skewed quantitative data. The Box length represents the inter-quartile range of the scores.

There are two principal approaches for outlier management. One is outlier accommodation, which is characterised by the development of a variety of statistical estimations or testing procedures which are robust against, or relatively unaffected by, outliers[42]. The approach that we followed is the second one, which is characterised by identifying outliers and deciding whether they should be retained or rejected.

All data obtained over each technique was sorted in ascending order, the minimum and the maximum observations are the two extreme values, respectively. CI+ data sets

obtained on the magnetic sector MAT95 is presented, the mass difference is listed in the table 3.2.

Table 3. 2 Mass difference (in ppm) between the exact mass and observed mass for chemical ionization positive mode (CI+) techniques on magnetic sector (MAT95).

0.000	-1.441	-1.234	-0.354	-1.266	0.000	-0.505	0.901	-1.313
1.323	0.000	0.000	-0.347	-0.305	-0.562	0.252	-0.664	-0.186
-0.613	0.000	-0.787	0.342	-0.599	1.102	-0.713	-0.438	1.107
1.142	-1.769	0.787	-0.680	0.000	-1.373	-0.475	-0.639	-0.351
-1.695	1.746	1.162	-1.020	0.000	1.354	0.237	1.279	-1.622
1.562	0.873	-2.298	-0.340	0.294	0.260	-0.712	-1.378	0.285
1.004	0.435	-0.746	0.660	-0.290	1.301	-0.233	-0.195	-0.601
-0.492	-1.265	-0.709	0.329	0.579	1.041	0.000	-0.573	

Mean = -0.1079, Max = 1.746, Min = -2.298, aaMMA = 0.7592
Root mean square (RMS) =0.928 Standard deviation = 0.928 ppm.

The summary statistics of the 71 individual masses (*m/z* 169-998 Da) consists of a minimum, maximum, mean; median and standard deviation were computed. The Box plot clearly shows only one outlier on CI+ magnetic sector (0.014 % of observation as an outlier). No significant difference in the summary statistics was found, with or without the outlier. The extreme value for the whole data set is -2.298 which is defined as an outlier, but after outliers omitted, the extreme value is -1.769 ppm, excluding the outlier. As -2.298 was the omitted outlier. The accuracy obtained (the absolute average mass measurement accuracy) aaMMA = 0.7372 ppm while with

outlier, it was 0.7592 ppm slight difference of 0.022 ppm. The difference in RMS value with and without the outlier is 0.035 ppm. This indicates that, because there is only one outlier and it is very close to the next value of samples. The outlier does not give a significant effect on the result of CI+ with magnetic sector data sets. Overall, the majority of data had sub-ppm mass accuracy.

Rosner's test was applied to CI+MAT95 data set and the 15 extreme values were chosen to present the result of this test which shown in table 3.4 below.

Table 3. 3 Tests applied showing a comparison of outlier's tests.

Test of outlier	**Total measurement**	**no of outlier**
Dixon's test ***(3 ≤ n ≤ 25)***	71	not used
Grubbs test ***(n > 3)***	71	1
Rosner's test ***(n > 25)***	71	1
2-Standard Deviation	71	1
Box and Whisker plot	71	1

The discussion presented here is to look behind the application of outlier tests and the distribution of data set around the mean of data sets. This is a part of the statistical analysis and interpretation. Referring to the results discussed in the previous chapter, it is one of the aims of this study to identify and interpret the data distribution and discuss the behaviour of outliers in context of the underlying data distribution.

Each data set is plotted separately using graphical tools and subjected to outlier tests. Each one has a different experimental distribution and behaviour according to the skewness and sample size of the data.

Table 3. 4 Rosner's test of the highest mass measurement errors obtained by CI+ MS (n=71). The calculated mean and SD use all the original value.

ppm	Mean	SD	Rosner's value (R)
2.298	0.734366	0.531081	2.944248
1.769	0.712029	0.500194	2.113122
1.746	0.69671	0.487038	2.15443
1.695	0.681279	0.473363	2.141531
1.622	0.666149	0.46007	2.077619
1.562	0.651667	0.447942	2.032258
1.441	0.637662	0.436623	1.839892
1.378	0.625109	0.428091	1.758718
1.373	0.613159	0.42063	1.806436
1.354	0.600903	0.412568	1.825387
1.323	0.588557	0.404279	1.816671
1.313	0.576317	0.396128	1.859711
1.301	0.563831	0.387437	1.902684
1.279	0.551121	0.37821	1.924539
1.266	0.538351	0.368741	1.973333

. The application of standard statistical tests to mass spectrometry often assumes the data is modelled by the Normal distribution, depending on the use of such data and the application that may be needed to confirm and test the distribution using a suitable test, such as K-S. The commonly used and simplest classical approach to screen for outliers is the use of the standard deviation method (2SD, 3SD), and Box plot method.

SD method is quite reasonable (as mentioned above) when the data distribution is symmetric, such as Normal distribution 2 SD Method = mean ± 2 SD, 3SD Method= mean ± 3SD, where the mean is the sample mean and SD is the sample standard deviation. The observation of mass measurement data two or three standard deviations outside these intervals may be considered as outliers. This method is not necessary true, as the mean and standard deviation may be highly affected by extreme values. Also conference interval will be inflated as the data increases in skewness. Also z-score test that should follow the Normal distribution will not be useful for these causes.

Here we have a small data as example to see how the extreme outliers affect the mean, the standard deviation and z- score test; [1.10, 1.30, 1.60, 1.80, 1.90, 2.00, 2.10, 2.20, 2.30, 2.40, 2.50, 2.80, 11, and 12]. The last two measurements (11, 12) seem outliers. By applying SD test and z-score test found that for the 14 points, mean = 3.31, and SD =3.50, the 2SD interval 3.31 ± 2*3.5= (-3.69, 10.31), and for 3SD is (-7.4, 13.81). Observation 11 and 12 are identified as outliers using a 2SD method, while both 11, 12 measurements lie in the interval of 3 SD.

Computation of the z-score depends on the mean and SD. If the observation (x) follows a Normal distribution with a mean μ and standard deviation σ, then the z follows the standard Normal distribution with mean = 0 and SD = 1, and z-score that gives > 3 in absolute value and is considered an outlier. Statistically, the maximum possible z-score depends on the sample size and it can be calculated as $[(n-1)/\sqrt{n}]$.

Table 3.4 illustrate the calculation and masking problem of z-score method using the previous example data set. As the masking problem happened when one outlier masked the second outlier in the same side, the second outlier can be consider as an outlier only by itself, but not of the presence of the first outlier. Masking occurs when a cluster of outlying observations skews the mean and the covariance estimates toward it, and the resulting distance of the outlying point from the mean is small

In case A the value 11 and 12 are outlier but there is no observation exceed 3, and in case B exclude the most extreme value 12, among example data , 11 is consider an outlier. This is referred to the inflation of the standard deviation of z-score. To avoid this effect of extreme value, the median and the median of the absolute deviation of median (MAD) are used instead of the mean and SD of the sample, respectively. Then the z-score method can be applied as a robust method as follow:

$$MAD = median\left\{\left|x_i - \tilde{x}\right|\right\}$$

where $\tilde{x}$ is the sample median.

The Box plot which was described earlier was developed by Tukey (1977) and is a very helpful method since it makes no distribution assumptions dependent on the mean and standard deviation[43].

Tukey defined Q_1-1.5*IQR) and Q_3+1.5*IQR) as the "inner fences", Q_1-(3*IQR) and Q_3+(3*IQR) as outlier fences, the observations between an inner fence and its nearby

outer fence as "outside", and anything beyond the outer fences as "far out"44. These outside observations are called outliers.

Table 3. 5 Computation and masking problem of the z-score

Case A, (mean =3.31, SD=3.50)			Case B,(mean=2.64, SD=2.55)	
i	Xi	Z	Xi	Z
1	1.10	0.632	1.10	0.605
2	1.30	0.575	1.30	0.527
3	1.60	0.48	1.60	0.409
4	1.80	0.432	1.80	0.331
5	1.80	0.432	1.80	0.331
6	1.90	0.403	1.90	0.292
7	2.00	0.375	2.00	0.253
8	2.10	0.346	2.10	0.214
9	2.20	0.318	2.20	0.174
10	2.30	0.289	2.30	0.135
11	2.50	0.232	2.50	0.057
12	2.80	0.146	2.80	0.0602
13	11.00	2.194	11.00	3.273
14	12.00	2.480	*	*

For the previous example, the graphical method test (Box and Whisker plot) is applied and the values 11 and 12 are presented as outliers in this method (figure 3.7 and table 3.5). If using Box plot method the outliers are not affect the mean or SD. Tukey's method may be applicable for symmetrical and skewed data; the more skewed the data the more observations may be detected as outliers[45].

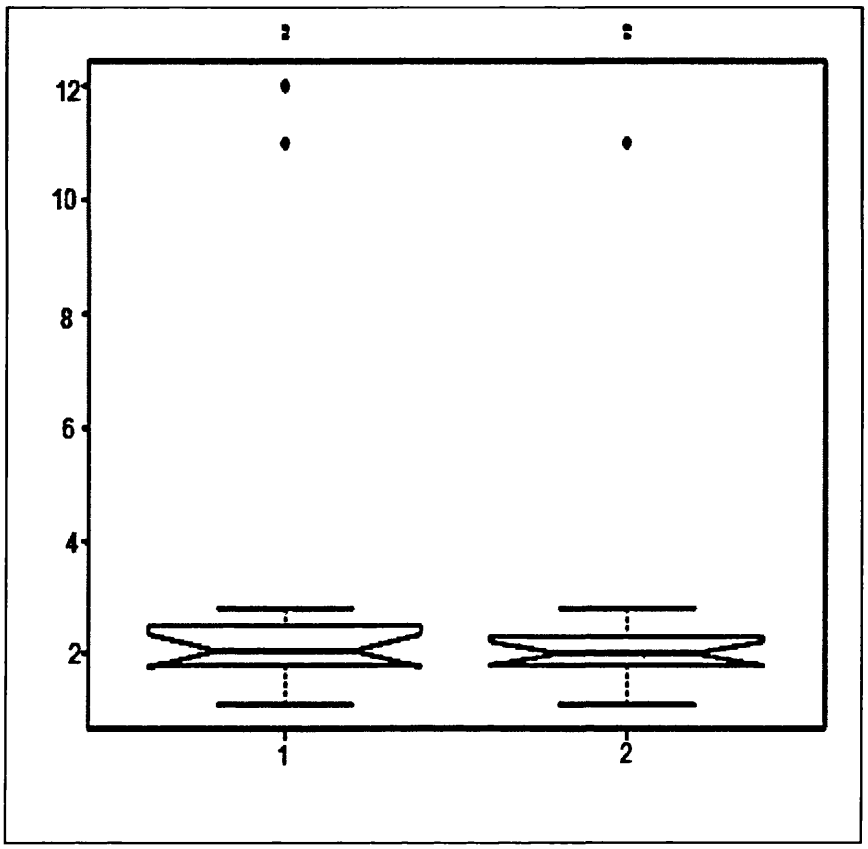

Figure 3. 7 *Box and Whisker plot for the example data set. The extreme outliers do not affect the distribution of data when using a box and Whisker plot.*

Overall, from the result of applications above, it is obvious that the Box and Whisker plot is the more appropriate statistical method for the large data set for many reasons. First, the Box and Whisker plot method classifies more observations as outliers than any other method. Secondly, it is less affected by the extreme value than the SD method, and is not affected by masking problems or skewed data. For these reasons the Box and Whisker plots is selected to identify outliers in this study and as a recommended method in many published papers.

Table 3. 6 Illustration of the confidence level ($\bar{x}$±2SD) for all data sets; the ppm errors outside of this interval will be considered as an outlier.

Method	No. of pts	Mean	SD	2SD	$\bar{x}$-2 D	$\bar{x}$+2SD
CI+ MAT95	71	-0.1079	0.9282	1.8564	-1.9643	1.7485
EI+ MAT95	964	-0.1527	0.8579	1.7158	-1.8685	1.5631
ESI+ MAT95	408	0.0745	0.8059	1.6118	-1.5373	1.6863
ESI+MAT900	2317	0.0316	0.9076	1.8152	-1.7836	1.8468
FAB+MAT95	58	-0.032	0.7251	1.4502	-1.4822	1.4182
ESI+ Orbitrap	784	0.14	1.201	2.402	-2.262	2.542
ESI- Orbitrap	66	0.195	1.262	2.524	-2.329	2.719
MALDI+ TOF	195	-0.429	3.408	6.816	-7.245	6.387

As shown in table 3.6, the confidence level for 71 points for CI+ MAT95 data is restricted to the interval (-1.96 to 3.71ppm) and only one point falls outside of this interval which is less than -1.96, see figure 3.8.

Box and Whisker plot is applied to all data sets; these data were discussed in chapter 2 as mass errors. The Q-Q plot is also applied to CI+ data sets and the points align well along a straight line. The lone outlier falls at the upper end of the straight line and it seems that the data confirm the impression from the graph that the data is not normally distributed.

Table 3. 7 Grubbs test for CI+ MAT95 data set; here the five extreme values are chosen from both sides of CI+ data set to see how many outliers can be detected.

	CI+ ppm			Grubbs
	-2.298			2.403
The lowest	-1.769			1.816
five values	-1.695			1.734
	-1.622			1.653
	-1.441			1.452
	.	mean=	-0.133	.
	.	SD=	0.901	.
	1.301			1.592
The highest	1.323			1.616
five values	1.354			1.650
	1.562			1.881
	1.746			2.085

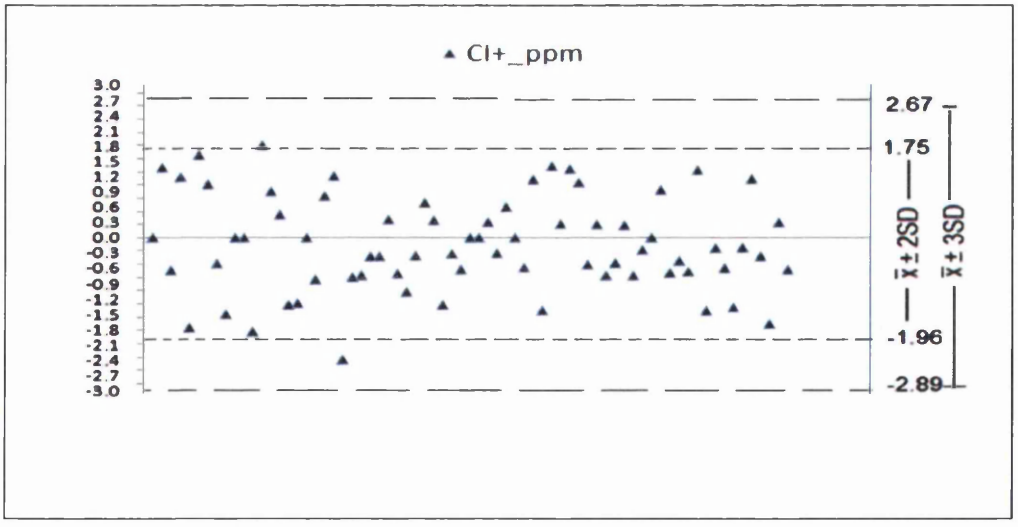

Figure 3. 8 Diagram showing mass error for CI+ MAT95. The data points falling below -1.96 or more than 1.75 they can be considered as outliers. We can see only one point less than -1.96 and no points are above 1.74ppm, and data points appear within 3SD.

Table 3. 8 Grubbs test for EI+ MAT95 data set; here we chose the five extreme values from both sides to see how many outliers can be detected.

	EI+ (ppm)			Grubbs
	-3.450	mean=	-0.153	3.843
The lowest	-2.890	SD=	0.858	3.190
five values	-2.500			2.735
	-2.480			2.712
	-2.400			2.619
	.			.
	.			.
	2.000			2.509
The highest	2.020			2.533
five values	2.060			2.579
	2.160			2.696
	5.000			6.006

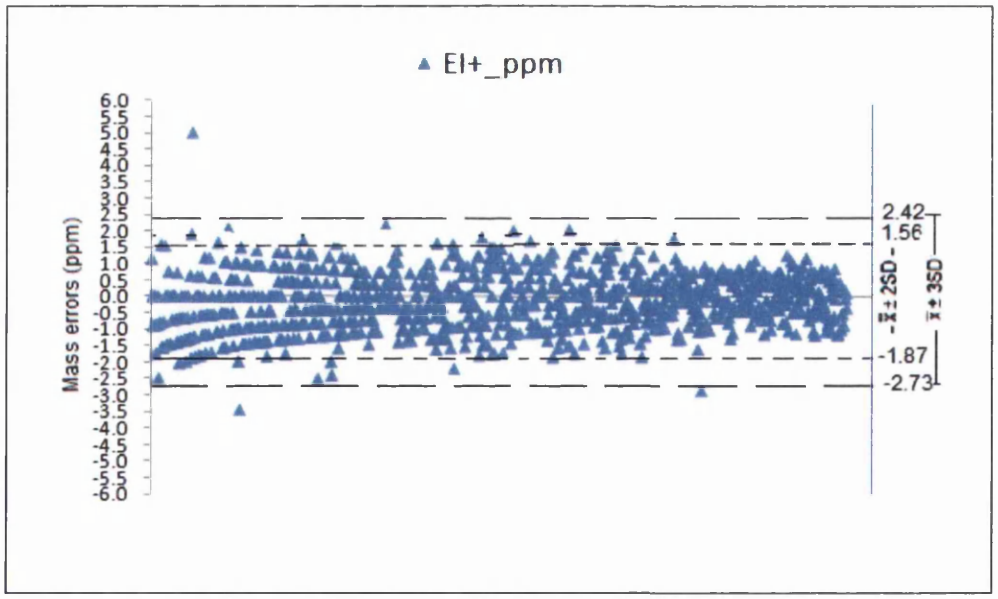

Figure 3. 9 Diagram shows the mass error for EI+ MAT95. The points falling below -1.87 or more than 1.56 ppm can be considered as outliers. There are several points over 1.56 ppm and we can see many points less than -1.87 ppm, and 3 outliers with 3SD.

Table 3.9 Grubbs test for ESI+ MAT95 data set; here we chose the five extreme values from both sides to see how many outliers can be detected.

	ESI+MAT95 (ppm)			Grubbs
	-2.070	mean=	0.074	2.660
The lowest	-1.930	SD=	0.806	2.486
five values	-1.890			2.437
	-1.870			2.412
	-1.800			2.325
	.			.
	.			.
	1.660			1.968
The highest	1.720			2.042
five values	1.750			2.079
	1.930			2.303
	1.930			2.303

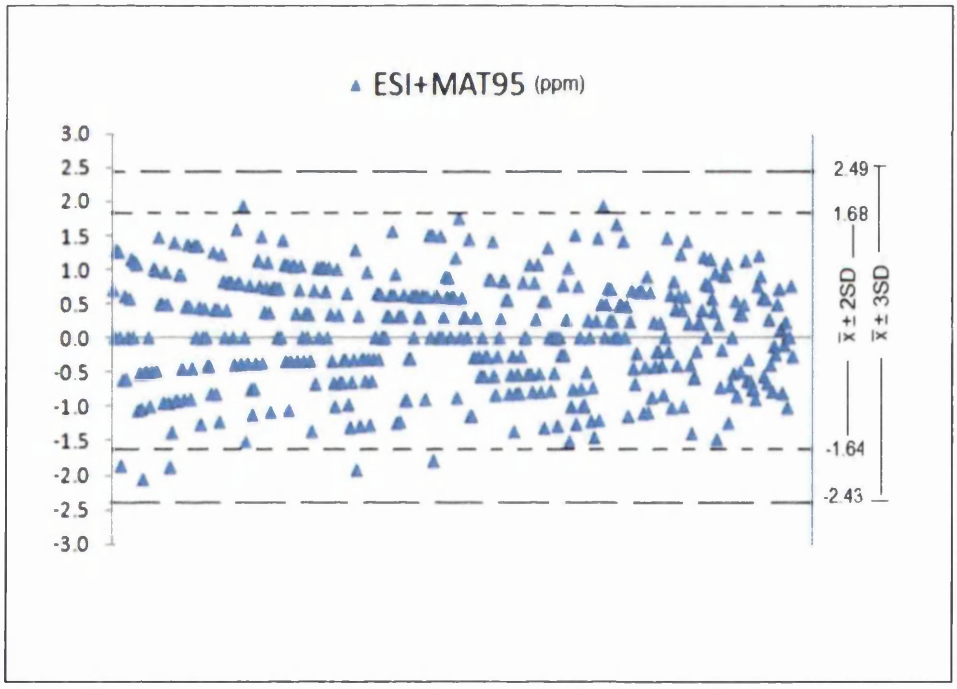

Figure 3. 10 Diagram showing mass error for ESI+MAT95. The points falling below -1.64 or more than 1.68 can be considered as outliers. There are two points over 1.68 whilst we can see five points fallen less than -1.64 ppm, and no outliers with 3SD.

Table 3. 10 Grubbs test for ESI+ MAT900 data set; here we chose the five extreme values from both sides to see how many outliers can be detected.

	ESI+MAT900 (ppm)			Grubbs
	-8.370	mean=	0.031	9.252
The lowest	-3.670	SD=	0.908	4.076
five values	-3.650			4.054
	-2.940			3.272
	-2.510			2.798
	.			.
	.			.
	3.210			3.501
The highest	3.750			4.096
five values	3.980			4.349
	4.020			4.393
	7.660			8.402

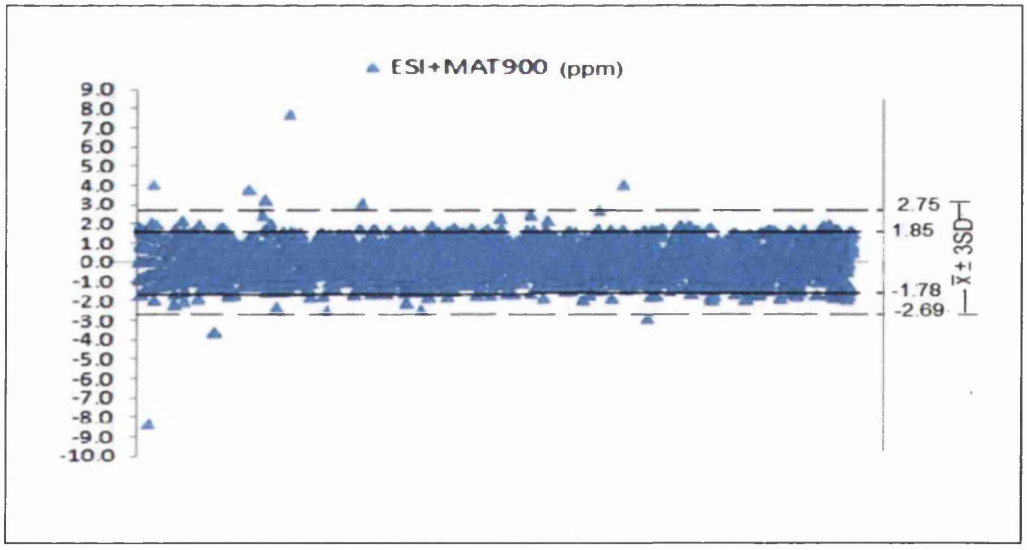

Figure 3. 11 Diagram showing the mass error for ESI+MAT900. Any point fall below -1.78 or more than 1.85 can be considered as outlier. There are many points over 1.85 and also many points less than -1.78 ppm, and 9 outliers outside 3SD.

Table 3. 11 Grubbs test for FAB+ MAT95 data set; here we chose the five extreme values from both sides to see how many outliers can be detected.

	FAB (ppm)			Grubbs
	-1.440	mean=	-0.033	1.943
The lowest	-1.300	SD=	0.724	1.750
five values	-1.160			1.557
	-1.130			1.515
	-0.970			1.294
	.			.
	.			.
	0.990			1.413
The highest	1.000			1.427
five values	1.140			1.620
	1.610			2.269
	1.640			2.311

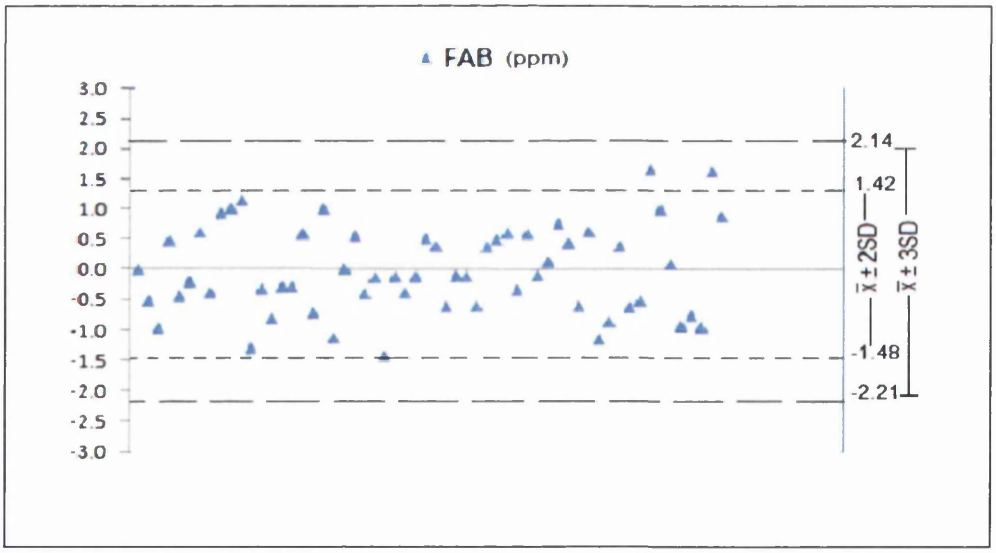

Figure 3. 12 *Diagram showing mass error for FAB+MAT95. The points falling below -1.48 or more than 1.42 ppm can be considered as outliers. There is no point less than -1.48 whilst there are two points above 1,42 ppm and no outlier detected with 3SD. No outliers outside of 3SD.*

Table 3. 12 Grubbs test for ESI+ Orbitrap data set, here we chose the five extreme values from both sides to see how many outliers can be detected.

	ESI+ Orb. (ppm)			Grubbs
	-3.240	mean=	0.140	2.814
The lowest	-3.170	SD=	1.201	2.756
five values	-2.840			2.481
	-2.740			2.398
	-2.710			2.373
	.			.
	.			.
	2.610			2.057
The highest	2.970			2.356
five values	3.090			2.456
	3.760			3.014
	4.580			3.697

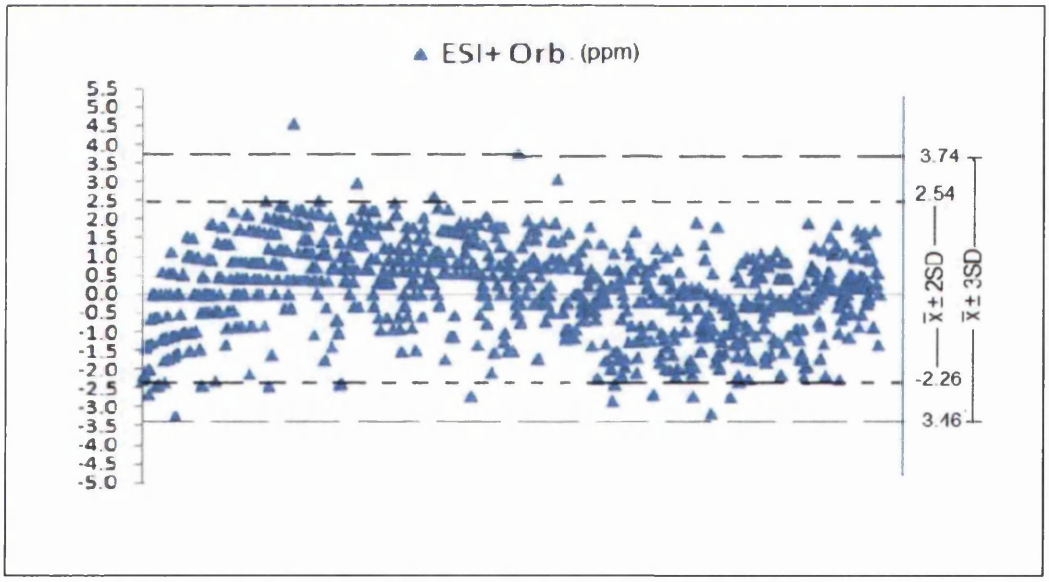

Figure 3. 13 Diagram showing mass error for ESI+ Orbitrap. The points falling below -2.26 or more than 2.54 can be considered as outliers. As shown there are 4 points over 2.54 and we can see more than 7 points less than -2.26 ppm. Only one outlier detected outside of 3SD

Table 3. 13 Grubbs test for ESI- Orbitrap data sets. Here we chose the five extreme values from both sides to see how many outliers can be detected.

	ESI-Orb (ppm)			Grubbs
	-2.870	mean=	0.195	2.423
The lowest	-2.010	SD=	1.262	1.743
five values	-1.690			1.490
	-1.630			1.443
	-1.630			1.443
	.			.
	.			.
	2.430			1.767
The highest	2.650			1.941
five values	3.040			2.249
	3.050			2.257
	3.180			2.360

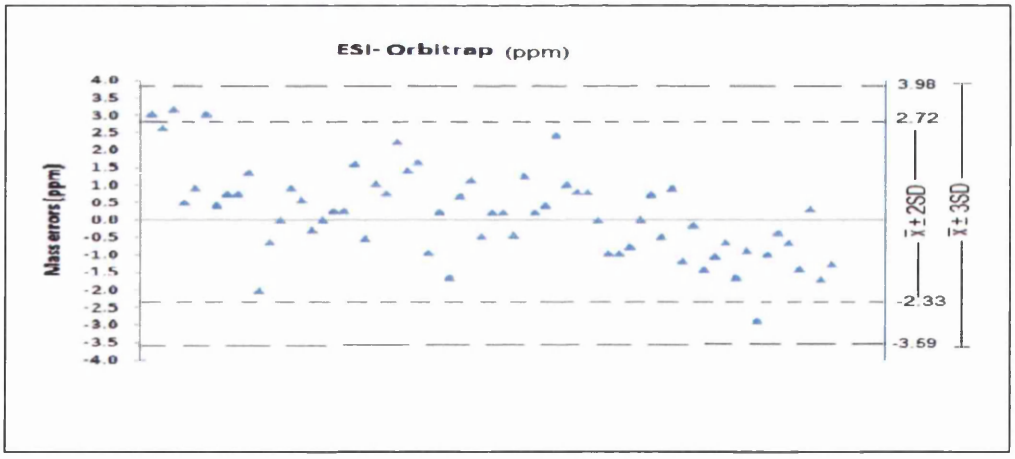

Figure 3. 14 Diagram showing mass error for ESI- Orbitrap. The points falling below -2.33 or more than 2.72 can be considered as outliers. As shown there are three points over 2.72 and we can see only one point fallen less than -2.33 ppm. No outliers are outside of 3SD.

Table 3. 14 Grubbs test for MALDI+ TOF data set; here we chose the five extreme values from both sides to see how many outliers could be detected.

	MALDI TOF (ppm)			Grubbs
	-12.420	Mean=	-0.429	3.518
The lowest	-11.810	SD=	3.408	3.339
five values	-9.930			2.788
	-8.330			2.318
	-7.180			1.981
	.			.
	.			.
	6.130			1.925
The highest	6.280			1.969
five values	6.340			1.986
	8.960			2.755
	10.760			3.283

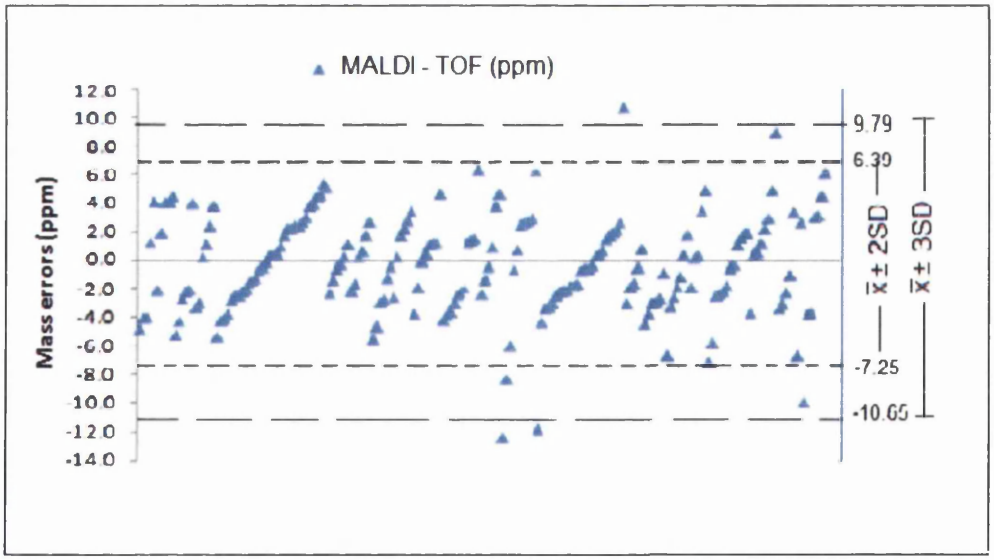

Figure 3. 15 Diagram showing mass error for ESI+ Orbitrap. The points falling below -7.25 or more than 6.39 can be considered as outliers. As shown there are two points over 6.39 three points less than -7.25 ppm, and three outliers outside of 3SD.

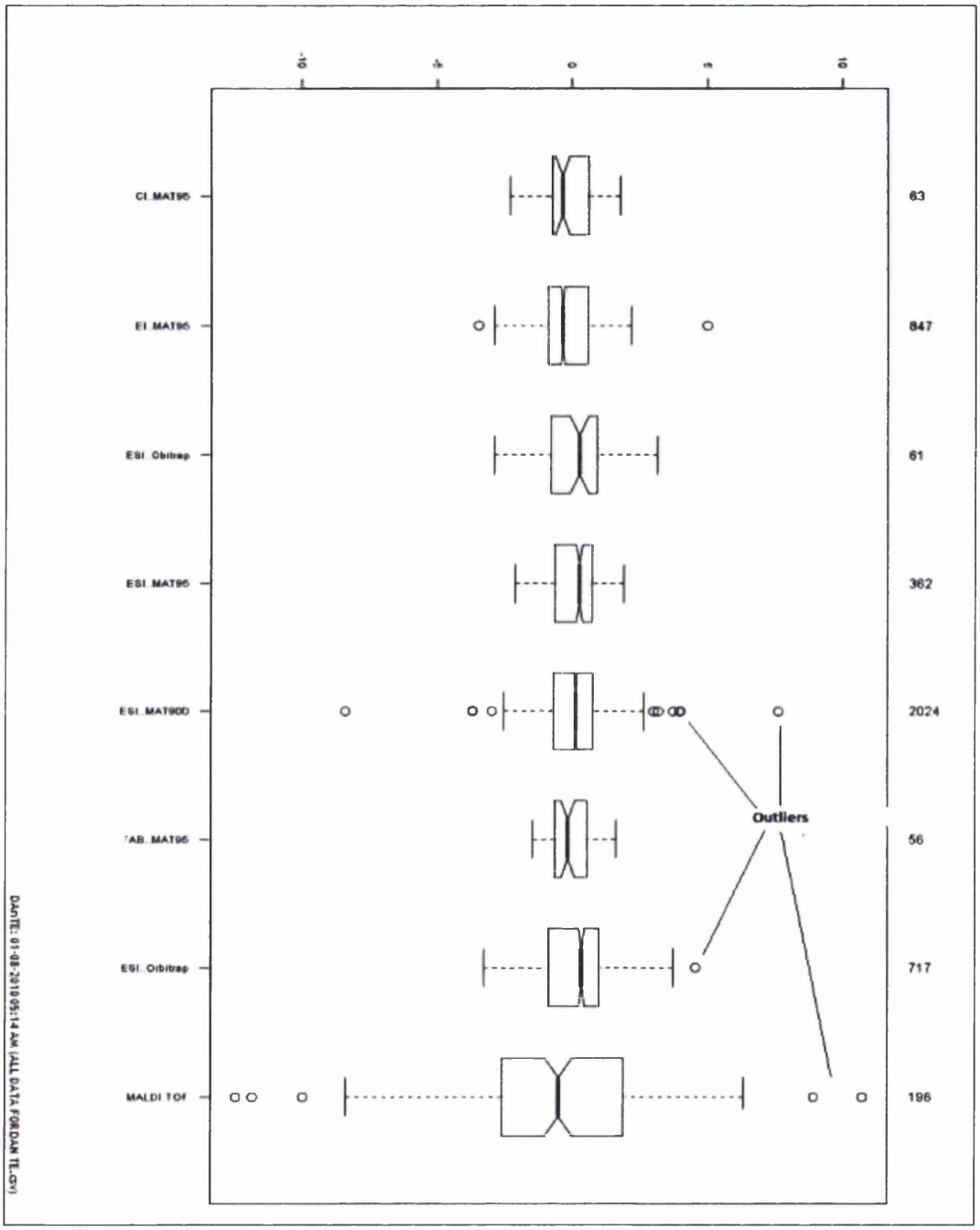

Figure 3. 16 *Box and Whisker plots (described in figure 3.3) of MMAs achieved in these studies for the different MS instruments. The figure displays the number of data points used to generate each Box and Whiskers plot. (DAnET. Software). Outliers are indicated as (o).*

Table 3.15 Illustration of skewness and Kolmogorov practical values for data sets obtained on magnetic sector instruments.

Method	n	Range	Mean	MSE	SD	Skewness	K-S
CI+ MAT95	71	4.050	-0.108	0.106	0.928	0.053	0.1462
EI+ MAT95	964	8.450	-0.152	0.027	0.857	0.154	0.1353
ESI+ MAT95	408	4.000	0.074	0.039	0.805	-0.121	0.1219
ESI+MAT900	2317	16.030	0.031	0.018	0.907	-0.047	0.0709
FAB+MAT95	58	3.080	-0.032	0.095	0.725	0.239	0.1607

Table 3. 16 Illustrate the skewness, range and Kolmogorov for data sets which obtained on Orbitrap and time-of-flight instruments.

Method	N	Range	Mean	Mean Error	SD	Skewness	K-S
ESI- Orbitrap	65	6.050	0.195	0.156	1.26	0.298	0.0898
ESI+Orbitrap	784	7.820	0.140	0.043	1.20	-0.187	0.1103
MALDI TOF	195	23.18	-0.429	2.745	3.407	3.468	0.3511

All methods described in this chapter (SD, z-score, Grubbs, Box and Whisker plot and Q-Q plot) in the case of detecting outliers in mass measurements can be applicable if data has a Normal distribution without possible masking problem or gap between the extreme mass error and the mean. If the data set has a Normal distribution with

possible masking or a big difference between the majority of data points and the extreme value of mass error, z-score and SD method would not be preferred or appropriate to use since these methods are highly sensitive to extreme values. Box and Whisker plot and the median rule may be suitable when data set is skewed; however, the Box plot takes into account the skewness of data distribution. (See tables 3.15 and 3.16 which include the skewed value for each technique).

The conclusion from the results of Grubbs test, 2SD, 3SD and Box plot which are illustrated in table 3.7 to table 3.14 and in figure 3.8 to figure 3.15 give different numbers of outliers, as shown in table 3.16.

Table 3.17 Number of outliers obtained by different tests over all data sets.

		Number of outliers			
Method	**N**	**2SD**	**3SD**	**Grubbs**	**Box plot**
CI+ MAT95	71	1	0	1	1
EI+ MAT95	964	14	3	3	2
ESI+ MAT95	408	7	0	1	0
ESI+MAT900	2317	53	9	9	9
FAB+MAT95	58	0	0	0	0
ESI+ Orbitrap	784	16	1	3	1
ESI- Orbitrap	66	4	0	0	0
MALDI+ TOF	195	6	3	5	5

3.6.2 Data distribution

The data obtained and analysed on time-of-flight (TOF) and Orbitrap are less precise than the results obtained on the magnetic sector. This data is presented in figure 3.16 the shape of the distribution for TOF and Orbitrap techniques gave wider distributions than those obtained by the magnetic sectors. For example, the magnetic sector FAB data distribution has 58 mass measurements and gives an accuracy and precision better than both Orbitrap and time-of-flight. Despite the sample size for FAB being smaller than the sample size of the Orbitrap and TOF (195 points), and referring to histograms; diagrams of 2SD and Q-Q plots, it is clear that FAB data are closer together and give a very narrow interval (±1.5 ppm). Whilst Orbitrap data are more dispersed and lie in an interval between -4 and 4ppm. On the other hand, TOF data are spread over a much wider interval (-10 to 10ppm). These results are also illustrated by another graphical test which is the Box and Whisker plot (figure 3.16) and give a clear presentation of the distribution for all techniques; the variance for each technique and outliers can be seen in this figure.

The presence of a large set to form a distribution gives better estimates of the accuracy and precision of data set. For example; ESI+MAT900 has 2,317 mass measurements gave excellent accuracy, with a mean = 0.0316 ppm, SD = 0.9 ppm and a median value very close to zero. The symmetrical shape around the mean and corresponding data curve to the Normal distribution is assumed from the shape, but this is also subjected to the K-S test.

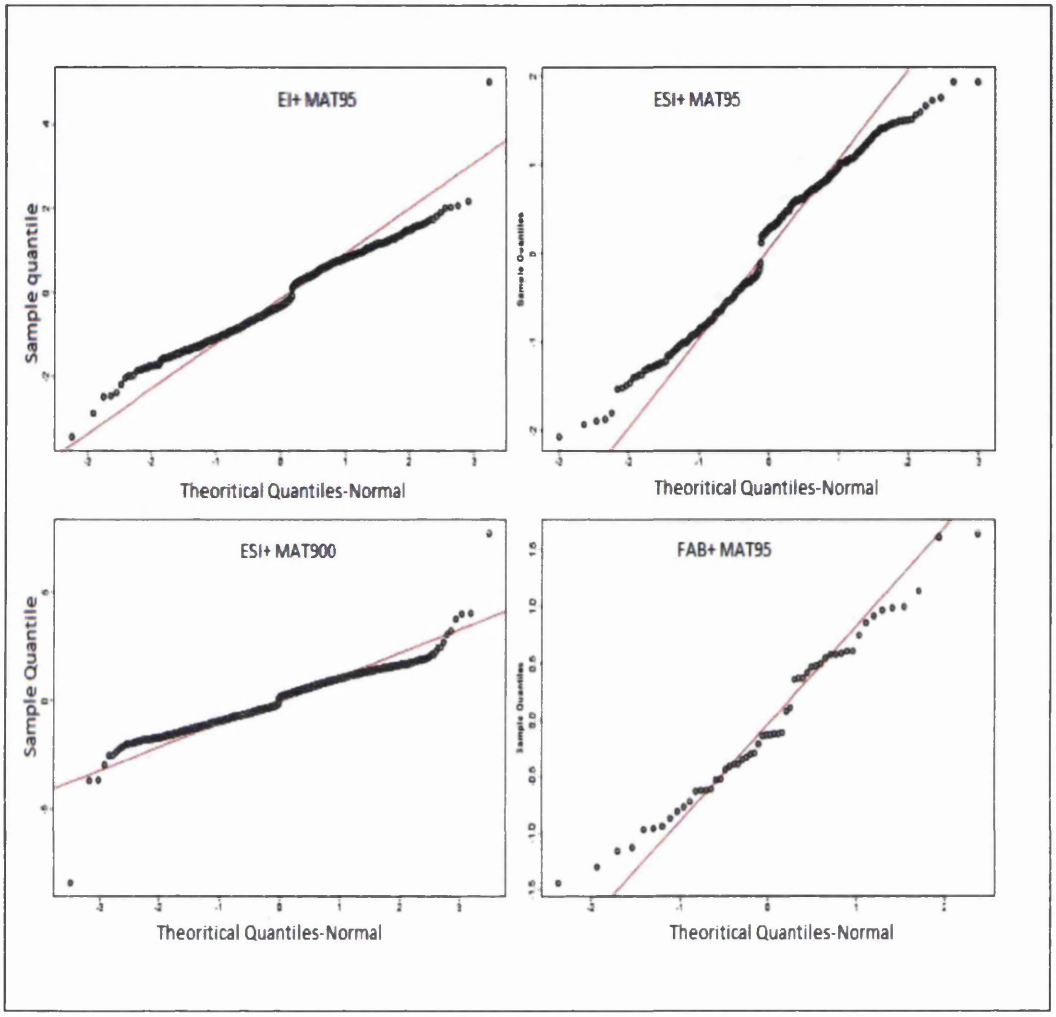

Figure 3. 17 *Q-Q plot shows the data distribution around the z-score (the straight line). The individual points far away from the other points indicate outliers.*

An interesting characteristic of the ESI+ shape is the tail rapid toward zero, much narrower than others, especially with the Orbitrap and TOF. This is also clear from the standard deviation value obtained which controls the width of a normal curve.

Tukey's recommendations in connection to Box and Whisker plots, values greater than 3 IQRs from the upper or lower edges of the Box is referred to as an extreme

outlier. Values that lie between 1.5 and 3 IQRs from the upper and lower fence of the Box plot are minor outliers46. The distribution presented by this method also illustrates the distribution of the data sets obtained.

Eight Box and Whisker plots are drawn for the data sets and displayed in figure (3.18). The plots illustrate the spread of MMAs achieved over eight techniques using three types of mass spectrometer instruments. The distribution curve shape is presented. The median of the data is presented by the line inside the Box, while the notched line from both sides represents 95% of the parametric percentile range. As displayed by 95% the confidence interval of means, the magnetic sector (EI, CI, ESI+, and FAB) demonstrated much higher accuracy and precision than the Orbitrap and TOF. The outliers that displayed as circle (o) upper and lower the shapes, the number of points used to construct each plot. The effect of ionisation techniques on the MAT95 is not very significant, FAB+ mass measurement accuracies were typically 0.2 ppm smaller than ESI+ and both less than 1 ppm. Nine outliers presented on sector MAT900 represent about 0.0038 % of the total points, so do not give significant effect on the distribution. This refers to the sample size of data which is 2317. The MMA obtained on ESI Orbitrap was 1.2 ppm using external calibration. Internal calibration is typically at least twice as accurate as external calibration[47]. An Orbitrap detector can reliably deliver mass accuracy below 1ppm, if so; improvement of accuracy in Orbitrap would be expected if internal calibration were used[48].The normal probability plot (or the normal test plot) is a graphical technique used to determine or investigate whether or not data is normally distributed. If the data fits the

line it can be assumed that the data is normally distributed. If it is not fitted, or the data is skewed, it may not be Normal distribution and data may need some transformation process. The normal probability plot (figure 3.17) shows a strongly linear pattern for ESI+ MAT95, the straight line represents the z-score of corresponding sorted columns ranks in unit Normal distribution. There are only two points deviate from the line fit to the points on the probability plot.

Table 3.18 Confidence level ($\bar{x}$± 3SD) for all data sets; the ppm error outside of this interval will be considered as an outlier.

Method	No. of pts.	Mean	SD σn-1	3SD	$\bar{x} - 3SD$	$\bar{x} + 3SD$
CI+ MAT95	71	-0.1079	0.9282	2.7846	-2.8925	2.6767
EI+ MAT95	964	-0.1527	0.8579	2.5737	-2.7264	2.421
ESI+ MAT95	408	0.0745	0.8059	2.4177	-2.3432	2.4922
ESI+MAT900	2317	0.0316	0.9076	2.7228	-2.6912	2.7544
FAB+MAT95	58	-0.032	0.7251	2.1753	-2.2073	2.1433
ESI+ Orbitrap	784	0.14	1.201	3.603	-3.463	3.743
ESI- Orbitrap	66	0.195	1.262	3.786	-3.591	3.981
MALDI+ TOF	195	-0.429	3.408	10.224	-10.653	9.795

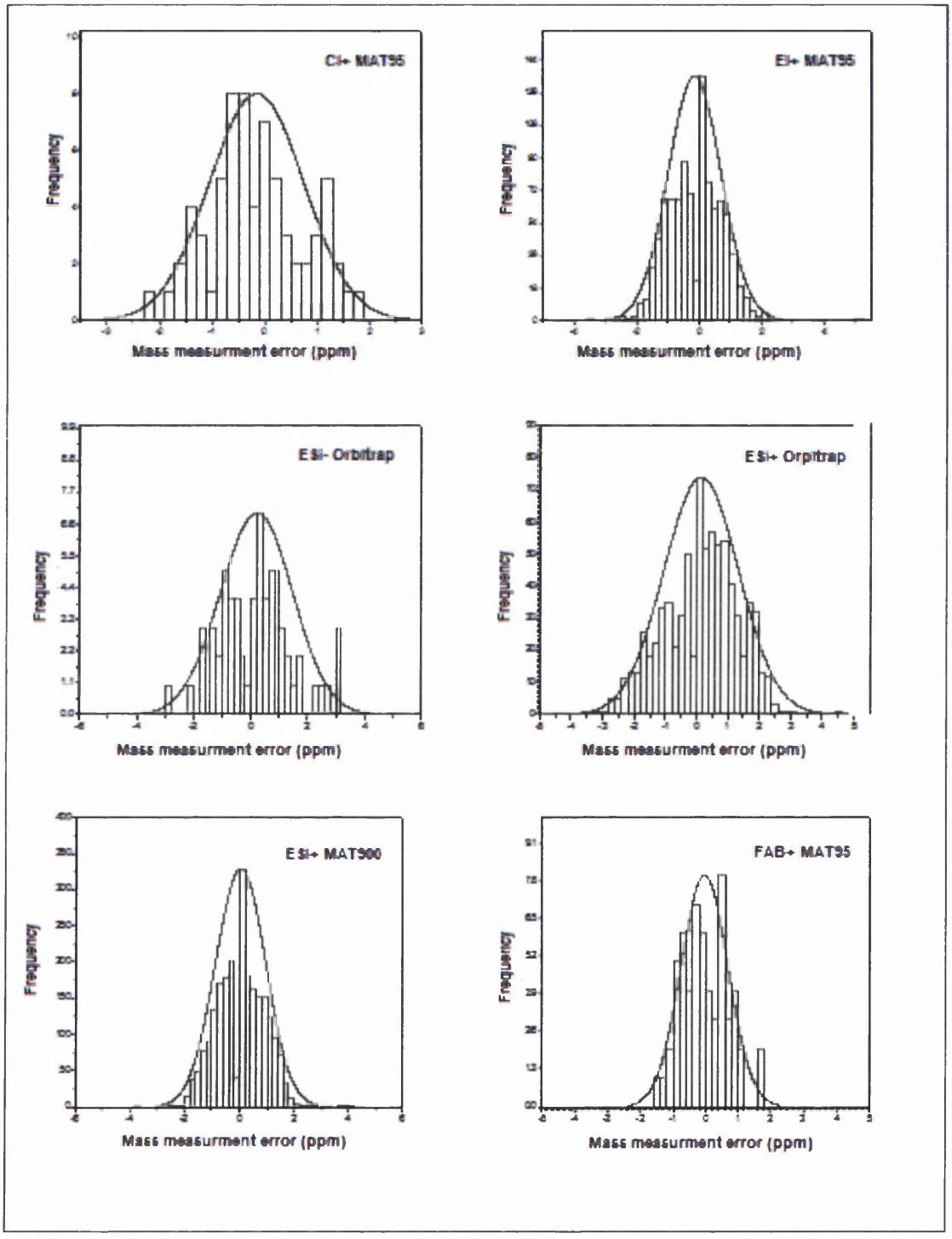

Figure 3.18 *Mass error distributions for all techniques using Origen software (CI+, EI+, ESI-Orbitrap, ESI+ Orbitrap, ESI+ MAT900 and FAB+ MAT95).*

As Box and Whisker plots are used to determine outliers it's also used to show data distribution, especially with data that are close to being normally distributed as many distributions in the science field do not follow Normal distribution. FAB and Orbitrap data give the highest skewness among all techniques investigated here. In such cases the distributions are frequently closer to a lognormal distribution than the normal one[49]. The outliers and data distribution change according to the skewness of the data in SD method (mean ± 2SD, mean ± 3SD) and Box plot method, and the standard Normal distribution is used for data sets whose skewness is equal to zero. Here the z-score is not usually considered because the criteria to define outliers depend on Normal distribution. Most intervals in these outlier methods (table 3.18) are effective under Normal distribution. For example 2SD, 3SD and 1.5 methods, the expected observations outside the interval are 5%, 0.3% and 0.7%, respectively as the confidence level are 95%, 99.7% and 99.3%, respectively. Table 3.19 shows the left; right and the total number of outliers identified which effect the distribution.

Overall the 2SD method classifies more observations as outliers than other methods. The 3SD and Box plot method (3IQR) are almost identical in results, with less outlier when the Box and Whisker plot is used. Also the skewness in all data had shown in table 3.17, biases to the low mass side (left). This is clear from the number of outliers which fall to the lower mass side. The interval of 3SD is much narrower than the interval of the Box plot.

Table 3.19 Measure the upper and lower Interval and total present number of outlier according to the three outlier methods.

	METHOD	No	Interval	Left	%	Right	%	Total (%)
EI+MAT95	2SD	964	(-1.87 , 1.56)	8	0.83	6	0.62	1.45
	3SD	964	(-2.73 , 2.42)	2	0.21	1	0.10	0.31
	Box plot	964		1	0.10	1	0.10	0.21
ESI+MAT95	2SD	408	(-1.54 , 1.69)	5	1.23	3	0.74	1.96
	3SD	408	(-2.34 , 2.49)	0	0.00	0	0.00	0.00
	Box plot	408		0	0.00	0	0.00	0.00
ESI+ Orbitrap	2SD	784	(-2.26 , 2.54)	15	1.91	4	0.51	2.42
	3SD	784	(-3.46 , 3.74)	0	0.00	1	0.13	0.13
	Box plot	784		0	0.00	1	0.13	0.13
MALDI TOF	2SD	195	(-7.24 , 6.39)	4	2.05	2	1.03	3.08
	3SD	195	(-10.65 , 9.79)	2	1.03	1	0.51	1.54
	Box plot	195		3	1.54	2	1.03	2.56

All the defined outlier methods in this study can be applied if data set is a Normal distribution without possible masking or a large gap between the extreme value and the majority of data. But it is possible to have a masking problem or large gap between the majority of data and the extreme value, the z-score and the SD may be not appropriate to use because these methods are highly sensitive to extreme values. Box and Whisker plot may be more suitable than others when a data set is skewed.

Analysis of the K-S test can provide confidence intervals proportional to the probability that the fitted distribution could have produced the observed data. Critical values as shown in table 3.1 are determined for the required confidence level. The critical value for K-S test is calculated at p=0.95 by $1.63/\sqrt{n}$. K-S test is compared the experimental distribution to the Normal distribution which is named Gaussian after Carl Friedrich Gauss introduced it in 1809 as a way of rationalizing the method of least squares. The statistics was set up that the cumulative frequency distribution was based on a Normal distribution. The experimental cumulative frequency plot is made in two ways:

a) Assuming the mean is identical to zero,

b) Assuming the mean is the experimentally determined mean.

In both cases the experimental standard deviation is used. As can be seen from table 3.1 where the experimental value of K-S is less than the critical value then the experimental data is adequately modelled by the Normal distribution. This is the case for the following data sets e.g. CI+ MAT95 at 0.95, n=71 the critical value is 0.1614 whilst the experimental value gives 0.1462. On the other hand, there are cases where clearly the experimental K-S value is much larger than the critical value indicating the Normal distribution does not adequately describe the distribution of experimental measurements. This is the case for the following data sets e.g. TOF has an experimental value 0.3511 (at 0.95 of possible value and n= 195) and the critical

value for this data is 0.0974. This indicates the magnitude of the difference between the experimental distribution and the theoretical distribution.

There are some general trends to be observed that the distribution for Orbitrap data is taily to low mass. For the Orbitrap this can be explained on the basis of a number of ions in the ion trap and earlier Orbitraps show this behaviour[50]. Thus the experimental distribution might be expected to have a lower mass tail. Another general trend is the number of data points when more data points are available for distributions then divergence from a Normal distribution is seen. This may be a consequence that the underlying distribution is in fact not normal, however this investigation has not gone that deep to fully investigate what other mathematical distributions might describe the data best.

3.7 Conclusions

Throughout this thesis a comparison of a large set of accurate mass data, obtained on magnetic sector, Orbitrap and time-of-flight using different ionization techniques, was studied. The results showed that magnetic sector technology still has a role to play in the confirmation of the elemental formulae of compounds prepared by synthetic chemists. Furthermore, statistical test of an accurate mass data was analysed using a number of statistical methods with extensive application to mass spectrometry data, and characterised in terms of mean and standard deviation. The results showed that outliers hardly affect these statistics and require appropriate methods of detection.

Therefore, they have been removed from the data sets to improve the statistical description of the underlying data.

Data sets were subjected to various types of tests such as; Dixon's, Grubbs' and Rosner's, to identify outliers at an acceptable level (95%). These procedures are not necessarily straightforward due to two factors. Firstly, variance can sometimes be misleading to use when two outliers or more are present. Secondly, it has to be decided either include or exclude outliers during the calculation of further statistics. Historical data sets of three different mass spectrometry techniques including large data sets have been examined. The most easily and recommended tests are the histogram, the Box and Whisker plot and the Q-Q plot. In all outlier tests used in this study, it was found that no method is perfect to identify all outliers at all situations in a single outlier or multi–outlier scenarios. Grubbs test is a quite good to identify a single outlier in a normal sample, whereas Box and Whisker plot is the best method of outlier identification in a large data set. This is due to that Box and Whisker plot depends on the median of the data and not on the mean. The results showed that in a magnetic sector using Box and Whisker plot test, FAB has no outliers, whilst TOF data has the maximum outliers. The mathematical theory (e.g., the central limit theorem) has proven that Normal distribution–based hypothesis test can be performed when a large number of samples are taken. This is due the fact that in a normal function based test, if the size of samples is large enough, the shape of the sampling distribution approaches Normal shape even if the distribution in question is not Normal. However, in larger samplings of data, some data points will be further away

from the sample mean than that is deemed reasonable. This can be due to incidental systematic error, a wrongly assumed model for the probability distributions, or some observations that may be far from the centre of the data. Outliers can, therefore, indicate faulty data. In large samples, however, a small number of outliers are expected and this may distort the distribution. Outliers contribute to the divergence from the Normal distribution as well as the skewness which characterises the degree of the asymmetry of the distribution around its mean.

Kolmogorov-Smirnov methodology was used to test if the distributions of mass measurement data were represented by the Normal distribution. For smaller data sets, there was agreement that distributions were nearly Normal, whereas for much larger data sets on the sector instruments, the Kolmogorov-Smirnov experimental statistic was always larger than the critical value resulting in divergence.

Orbitrap full scan operation is an efficient approach for high-throughput accurate mass analysis, but with slightly higher estimates of the magnitude of MMA: aaMMA=0.978 ppm; precision SD=1.2 ppm; and RMS=1.21 ppm. The histogram of this data was a bit lower than the MAT95 data, as assessed from SPSS statistics, and would be expected to improve when internal calibration is employed on the Orbitrap.

An extensive data over a range of ionization techniques is derived by estimating the magnitude of mass measurement accuracies, *i.e.* aaMMA or RMS or even from one standard deviation. Sub-ppm mass measurement accuracy has been demonstrated for historical data on a magnetic sector instrument using internal calibration, while

Orbitrap gave a value of 1.2 ppm for ESI+ data using an external calibration method. Improvements in mass measurement accuracy on the Orbitrap would be expected if internal calibration were used[52,53], however, this is not our standard operating procedure for the Orbitrap.

The effect of ionization technique on the MAT95 is not significant; FAB+ mass measurement accuracies were typically 0.2ppm smaller than those associated with ESI+, and all were less than 1 ppm. Overall, it is not clear why different ionization techniques have different effects on the value of MA, however it can be speculated that this is due to; i) the physical properties of the source; *i.e* energy and spatial spread of the ion beam, ii) the mechanical tolerance of each sources of affecting reproducibility over a large set of disjoint measurements and iii) different cleanliness and aging behaviour as each source is used.

3.8 References

1. Aggarwal CC, Yu SP. *An effective and efficient algorithm for high-dimensional outlier detection. The VLDB Journal.* 2005; 14: 211–221.

2. Urfer W, Grzegorczyk M, Jung K. *Statistics for Proteomics: A Review of Tools for Analyzing Experimental Data. Proteomics.* 2006; 6: 48-55.

3. Burns M, Nixon G, Foy C, Harris N. *Standardisation of data from real-time quantitative PCR methods - evaluation of outliers and comparison of calibration curves. BMC Biotechnology.* 2005; 5: 31.

4. Walczak, B, Massart DL. *Robust principal components regression as a detection tool for outliers. Chemometrics and Intelligent Laboratory Systems,* 1995. 27: p. 41-54.

5. Pell RJ. *Multiple outlier detection for multivariate calibration using robust statistical techniques. Chemometrics and Intelligent Laboratory Systems. 2000; 52: 87-104.*

6. Schulz-Trieglaff O, Machtejevas E, Reinert K, Schluter H, Thiemann J, Unger K. *Statistical quality assessment and outlier detection for liquid chromatography-mass spectrometry experiments. BioData Mining. 2009; 2: 4.*

7. Rosario P, Martínez J, Silván J. *Comparison of different statistical methods for evaluation of proficiency test data. Accreditation and Quality Assurance: Journal for Quality, Comparability and Reliability in Chemical Measurement.* 2008; 13: 493-499.

8. Egan WJ, Morgan SL. *Outlier Detection in Multivariate Analytical Chemical Data. Analytical Chemistry.* 1998; 70: 2372-2379.

9. Solberg HE, Lahti A. *Detection of Outliers in Reference Distributions: Performance of Horn's Algorithm. Clin Chem. 2005; 51: 2326-2332.*

10. Committee AM. Robust statistics*: a method of coping with outliers. Royal Society of Chemistry.* 2001.

11. Cho H, Kim Y-j, Jung HJ, Lee S-W, Lee JW. Outlier: *An R package for outlier detection using quantile regression on mass spectrometry data. Bioinformatics.* 2008; 24: 882-884.

12. Committee AM. Analytical Methods Committee, *Robust Statistics, Parts 1 and 2. Analyst.* 1989; 114: 1693.

13. Committee AM. *Report on the experiment test of recommendation for the conduct and interpretation of cooperative trials. Analyst.* 1989; 114: 1489.

14. Vekey K, Telekes A, Vertes A. *Medical Applications of Mass Spectrometry (1st edn).* pp.309-343. *Elsevier Science,* 2008.

15. Hawkins D. *Identification of Outliers. P188. Chapman and Hall, London.* 1980.

16. Johnson RA, Wichern DW. *Applied multivariate statistical analysis (5 edn). Prentice-Hall, Inc., Pearson, USA, 2001.*

17. Barnett V, Lewis T. *Outliers in Statistical Data. 3rd edition. J. Wiley & Sons* 1994, 1985: 365.

18. Hund E, Massart DL, Smeyers-Verbeke J. *Robust regression and outlier detection in the evaluation of robustness tests with different experimental designs. Analytica Chimica Acta.* 2002; 463: 53-73.

19. Dixon WJ. *Analysis of Extreme Values. Institute of Mathematical Statistics. Stat.*1950; 21: 488-506.

20. Thode HC. *Testing for normality. Volume 164. New York: Marcel Dekker;* 2002. p 479.

21. Miller J, Miller JC. *Statistics and Chemometrics for Analytical Chemistry (5th edn). Prentice Hall, UK,* 2005.

22. Chang L-C, Jones DK, Pierpaoli C. *RESTORE: Robust estimation of tensors by outlier rejection. Magnetic Resonance in Medicine.* 2005; 53: 1088-1095.

23. Galai D, Kedar-Levy H, Schreiber BZ. *Seasonality in outliers of daily stock returns. International Review of Financial Analysis.* 2008; 17: 784-792.

24. Nikolić-ðorić EK, Čobanović K, Lozanov-Crvenković Z. Nikolic-Doric. *Statistical Graphics and Experimental Data. Serbia. 2006.*

25. Bristow AWT, Webb KS. *Intercomparison study on accurate mass measurement of small molecules in mass spectrometry. J Am Soc Mass Spectrom.* 2003; 14: 1086-98.

26. Marshall AG, Hendrickson CL, Jackson GS. *Mass Spectrometry Reviews: Fourier transform ion cyclotron resonance mass spectrometry: A primer. 1998; 17: 1-35.*

27. Donnelly RA. *The Complete Idiot's Guide to Statistics (2 edn). Maria Butkr-knight,* 2004.

28. Salkind NJ. *Statistics for People Who (Think They) Hate Statistics: Excel 2007 Edition (2 edn). SAGE Puplications, Inc., 2009.*

29. Cinar A, Parulekar SJ, Undey C. *Batch fermentation: modeling, monitoring, and control. New York: Mercel Dekker, Inc;* 2003.

30. Laurikkala j, Juhola M, Kentala E. *Informal identification of outliers in medical data. Tempera;* 2000.

31. Smith WS. *The Scientist and Engineer's Guide to Digital Signal Processing. California: San Diego;* 1997.

32. McCuen RH. *Modeling hydrologic change: statistical methods. CRC press LLC,* 2003.

33. Thode HC. *Testing for normality. New York: Marcel Dekker; 2002; 8. p 479.*

34. Wilk MB, Gnanadesikan R. Biometrika. *Probability Plotting Methods for the Analysis of Data.* 1968; 55: 1-17.

35. Petrie A, Sabin C. *Medical Statistics at Glance. 2nd ed: Blackwelll Publishing Ltd.;* 2005.

36. Dunn SD. *Statistics and Data Analysis For Behavioral Sciences (1st edn). McGraw-Hill,* 2000.

37. Meier PC, Zünd RE. *Statistical Methods in Analytical Chemistry. Wiley,* 2000.

38. Kirk AJ, McCuen RH. *Outlier Detection in Multivariate Hydrologic Data. Journal of Hydrologic Engineering.* 2008; 13: 641-646.

39. Massart DL, Vandeginste GM. *Handbook of chemometrics and qualimetrics. 2 ed. Volume 20 A: Elsevier Science B.V.;* 1997.

40. Millard SP, Neerchal NK. *Environmental statistics with S-Plus. CRC applied environmental statistics series;* 2001. p 830

41. Jorma L, Martti J, Erna K. *Informal Identification of Outliers in Medical Data. IOS Press. Amsterdam.* 2000.

42. Berthold M, Hand DJ. *Intelligent Data Analysis. 2nd ed: Springer, Ramoni and Sebastiani, New Yourk,* 1999; p 514.

43. High R. *Dealing with outliers: How to maintain your data's integrity. Available from: cc.uoregon.edu/cnews/spring 2000/outlier.html (accessed Sep. 2005).*

44. Hoaglin DC, Mosteller F, Tukey JW. *Fundamentals of exploratoty analysis of variance. In: Hoaglin DC, editor. Canada: Wiley-Interscience; 1991.*

45. Vanderviere E, Huber M. *An adjusted Box plot for skewed distributions. Compstat, section: Graphics. 2004.*

46. Ramsey JO, Wiley D. *Applied Psychological Measurement. Book Reviews : Exploratory Data Analysis, John W. Tukey. Addison-Wesley, 1978; 2: 151-155.*

47. Marshall AG, Hendrickson CL. *Annual Review of Analytical Chemistry. High-Resolution Mass Spectrometers.* 2008; 1: 579-599.

48. Olsen JV, de Godoy LMF, Li G, Macek B, Mortensen P, Pesch R, Makarov A, Lange O, Horning S, Mann M. Mol *Parts per Million Mass Accuracy on an Orbitrap Mass Spectrometer via Lock Mass Injection into a C-trap. Cell Proteomics.* 2005; 4: 2010-2021.

49. Iglewicz B, Hoaglin D. *How to detect and handle outliers. ASQC Quality Press. Volume 16; 1933.*

50. Makarov A, E. Denisov, and O. Lange, *Performance Evaluation of a High-field Orbitrap Mass Analyzer. Journal of the American Society for Mass Spectrometry,* 2009. 20(8): p. 1391-1396.

51 Box plot [online], *2010.http://en.wikipedia.org/wiki/Boxplot.[Accessed March 2010].*

52 Bristow AWT. *Accurate mass measurement for the determination of elemental formula - A tutorial. Mass Spectrometry Reviews.* 2006; **25**: 99-111.

53 Cox J, Mann M. *Computational Principles of Determining and Improving Mass Precision and Accuracy for Proteome Measurements in an Orbitrap. Journal of the American Society for Mass Spectrometry.* 2009; **20**: 1477-1485.

4. APPENDICES

Mass different in ppm between the exact mass and observed mass for Chemical Ionization techniques on magnetic sector (MAT95)

MAT95 CI+ (71 points)

0.000	-1.441	-1.234	-0.354	-1.266	0.000	-0.505	0.901	-1.313
1.323	0.000	0.000	-0.347	-0.305	-0.562	0.252	-0.664	-0.186
-0.613	0.000	-0.787	0.342	-0.599	1.102	-0.713	-0.438	1.107
1.142	-1.769	0.787	-0.680	0.000	-1.373	-0.475	-0.639	-0.351
-1.695	1.746	1.162	-1.020	0.000	1.354	0.237	1.279	-1.622
1.562	0.873	-2.298	-0.340	0.294	0.260	-0.712	-1.378	0.285
1.004	0.435	-0.746	0.660	-0.290	1.301	-0.233	-0.195	-0.601
-0.492	-1.265	-0.709	0.329	0.579	1.041	0.000	-0.573	

MAT95 EI+ (964 points)

1.140	-1.320	-1.550	0.000	-1.300	0.400	0.370	-0.340	-1.280
0.000	-1.280	1.040	-0.930	1.300	1.590	-0.730	0.690	-0.960
-0.900	0.630	1.040	-1.390	0.860	-0.400	0.370	-1.360	0.000
-0.890	-1.890	-0.520	-1.390	0.000	-1.190	0.360	-1.360	-0.960
0.000	1.890	2.060	-0.930	-0.860	-1.590	-1.090	-1.020	0.320
0.000	5.000	0.000	-0.930	1.720	-0.790	0.000	-1.020	0.000
-1.750	-1.250	0.000	-1.390	-0.850	0.000	0.000	-1.020	-1.580
-0.870	-0.620	0.510	-1.830	0.850	-1.180	-1.090	-1.020	1.270
-0.870	-1.240	-1.020	-1.380	0.430	-0.780	0.730	0.340	-0.950
0.000	-1.850	0.510	-0.460	-0.850	1.180	-0.360	0.000	1.580
-2.480	0.000	-0.510	0.000	-0.420	-1.170	0.730	-0.680	0.310
-0.820	0.000	-1.520	-0.460	0.420	-1.170	-0.360	-0.680	-2.200
-0.810	0.610	-1.010	-0.460	0.850	-1.170	0.730	0.000	0.310
1.570	0.000	0.000	0.000	0.420	0.000	0.000	-0.340	0.310
-1.560	-1.810	1.010	-0.910	1.270	-0.390	0.000	-0.680	1.260
-1.560	0.000	0.500	0.000	1.270	0.770	0.000	-0.680	-0.940
-1.560	-0.600	1.000	0.910	-1.260	1.160	0.000	0.680	0.000
-0.780	-1.810	-0.990	0.910	0.000	0.000	0.360	-1.010	0.940
1.540	0.600	-1.980	0.900	-0.420	0.770	2.160	0.340	-0.940
0.000	-1.190	-0.980	0.000	1.260	-0.770	-0.360	-0.670	0.940
0.000	-0.590	-3.450	-0.900	-0.840	0.770	0.360	-0.670	-0.930
-1.490	0.000	1.480	0.000	0.000	0.770	0.360	-1.000	0.000
0.000	1.170	0.000	1.350	-0.830	-0.380	-0.360	-0.670	0.620
0.750	-1.170	0.000	0.450	-0.830	-1.140	0.000	0.670	0.930
0.000	-1.740	-1.470	-1.340	0.420	1.140	0.360	-0.330	0.000
-1.470	-0.580	0.000	0.000	0.830	0.000	-0.360	0.990	-0.310

-1.470	-0.580	0.980	-1.340	-0.420	-0.760	0.000	-0.330	-1.540
0.000	-0.580	-1.460	0.000	-2.500	0.380	1.060	0.330	-0.310
-0.730	-1.160	0.970	-1.330	0.000	-1.130	0.710	0.990	-1.230
0.000	1.160	-0.970	-0.880	0.000	-0.380	-0.350	0.330	0.920
0.000	-0.580	0.490	0.880	-0.410	0.380	0.350	-1.320	0.000
0.000	-0.580	0.970	0.440	-0.830	-0.380	1.060	0.000	-0.610
0.000	-1.720	0.000	-0.880	-1.230	0.750	0.710	0.650	-0.610
-0.710	0.000	-1.440	-1.760	0.820	0.750	-0.350	0.320	0.000
-1.430	0.560	-0.960	1.320	0.820	0.750	1.400	-1.300	-0.310
0.710	-0.560	-0.480	0.440	-0.410	-1.120	1.050	-0.650	-0.920
-1.390	0.560	-0.950	-1.320	-1.220	0.000	-1.400	-0.320	0.610
-0.690	0.560	-0.950	0.880	-0.410	-0.750	0.000	-1.620	0.910
-2.050	0.550	0.000	0.880	0.400	0.370	0.350	1.610	-0.910
-0.680	-1.100	0.000	-0.440	-0.400	0.740	-0.700	0.000	-0.300
-1.370	-1.100	-1.410	0.440	-1.210	-0.370	-1.050	-1.290	0.610
0.000	1.630	-1.410	-0.870	-0.400	0.740	-1.040	-1.290	-0.300
-0.680	0.000	0.940	-0.870	1.200	0.370	-0.690	0.000	0.300
-0.680	0.000	1.410	0.440	-0.400	0.000	-0.690	-1.290	1.210
-0.680	0.000	0.930	-1.300	-0.400	-0.370	-1.040	-0.640	-0.910
-1.340	-0.530	0.000	-0.430	-2.000	0.000	-0.350	-0.320	1.210
0.000	-0.530	-0.930	0.000	-2.400	-1.480	-0.350	-0.640	-1.800
-2.000	-1.580	-1.400	-0.870	-0.800	-0.740	-0.350	0.000	-1.200
-1.330	-1.040	0.930	-0.430	0.800	-0.370	-1.380	-0.960	-0.600
-0.660	-0.520	-0.470	0.430	-0.800	-0.740	-1.380	0.960	1.800
0.000	1.040	0.930	0.000	0.800	-1.100	-0.690	-1.280	-0.600
-1.490	-1.130	-0.790	0.480	-1.270	0.400	0.720	0.000	-0.140
-0.890	0.280	-0.260	-0.950	-0.420	0.000	0.900	0.000	0.710
0.300	-0.560	-1.570	-0.950	-0.630	-0.590	-0.890	0.480	0.140
0.600	0.000	0.260	0.470	-0.420	-0.590	0.350	0.630	-0.430
-0.890	-0.560	-1.040	-0.470	0.420	0.000	-0.880	-0.470	0.710
0.890	-0.280	-0.520	1.180	1.050	-0.590	0.180	-0.470	0.570
-1.480	-0.560	-0.260	1.170	-0.840	0.000	0.350	-0.160	0.000
-0.590	-0.560	-1.300	-0.470	0.210	0.580	0.180	0.160	-0.840
0.590	1.120	0.000	-0.920	1.050	-0.580	-1.050	-0.470	0.420
1.180	1.120	-0.520	-0.690	-0.630	-0.390	0.000	-0.160	0.000
1.480	-0.280	-0.520	0.230	0.420	1.730	0.700	-0.470	0.000
0.880	0.560	-0.770	0.000	0.620	-0.190	0.700	0.620	-0.270
0.000	0.280	0.000	0.460	0.410	1.150	0.350	0.930	-0.820
-1.760	0.560	-1.030	-1.380	1.450	0.380	0.350	0.150	1.230
1.170	1.680	0.510	-0.460	0.830	0.190	-0.520	-0.920	0.270
1.470	-1.120	-0.770	-0.680	-1.450	-1.330	0.520	-0.310	0.540
-1.760	-0.840	-1.530	0.910	0.830	-1.140	-0.690	0.000	-1.070
1.170	0.840	2.020	-1.580	-1.860	0.950	-0.170	0.310	0.270
-1.170	-0.280	1.010	0.680	0.210	0.760	0.000	0.460	-0.130
0.880	-1.110	-1.520	1.360	-0.410	-0.190	0.000	0.770	1.060
-0.580	0.830	-0.760	0.900	-0.210	-1.320	-0.680	1.080	0.270
-0.580	-0.280	0.760	0.000	0.210	0.380	0.680	-0.610	-0.530
1.170	0.000	0.250	0.680	-1.230	0.000	0.510	0.150	0.930
-0.870	-0.830	1.250	1.130	-0.820	-0.190	-0.510	-0.610	-0.530
-0.580	1.100	0.000	1.340	-0.410	0.380	0.680	0.300	0.260
1.450	0.000	-1.750	-0.670	0.410	0.750	0.510	-0.450	0.520
-0.290	-1.100	-0.500	-0.220	-0.200	0.000	0.840	-0.300	-1.170
0.870	1.100	1.240	-1.110	-0.610	0.750	-0.170	-0.750	-0.260

-1.740	-1.100	-0.740	1.110	-0.410	0.370	-0.340	0.600	0.130
0.000	0.550	-0.740	0.220	0.200	0.560	0.670	0.300	0.770
0.580	0.820	-0.490	0.000	-0.820	-0.930	0.500	-0.450	0.380
0.000	0.270	-0.990	-0.660	-0.410	0.190	-0.330	0.300	0.640
-0.580	0.820	0.740	1.550	0.200	0.000	0.670	0.740	-0.130
0.580	0.000	0.240	-1.100	1.220	-0.190	-0.170	-1.180	-0.510
-0.860	-1.090	0.000	0.000	-1.010	0.560	-0.830	0.590	0.000
0.000	-1.090	0.730	0.220	0.610	-0.180	-0.660	-0.740	0.890
-1.150	1.090	1.460	0.220	-0.200	-0.370	0.660	-0.150	0.000
0.290	0.820	-0.490	-1.520	-1.010	0.740	-1.310	0.590	-0.130
1.150	-1.350	-1.450	-0.650	0.200	-0.550	0.660	-0.590	0.250
-1.440	-0.540	0.000	-1.310	-0.400	-0.740	0.330	0.290	0.250
0.860	-1.080	-0.240	-1.740	-1.010	-0.370	-1.140	0.590	1.140
-1.140	-0.270	0.240	0.220	-1.010	0.180	-0.330	-0.740	0.130
2.000	1.070	-0.480	-0.430	-0.810	0.000	0.320	-1.030	0.380
-1.140	1.330	-0.720	0.870	-0.600	-1.640	0.320	-0.440	-0.500
-0.570	1.060	0.000	0.870	-0.400	-1.090	-0.490	-0.290	-0.750
-1.140	0.270	-0.240	0.650	0.800	0.180	0.650	-0.290	0.500
-0.280	-1.860	0.720	-1.070	0.200	-0.540	0.650	-0.140	0.740
0.850	1.320	0.720	-0.210	-0.200	-0.900	0.160	0.580	-0.620
0.850	0.530	-0.480	-0.210	0.800	-2.890	0.800	0.290	-0.610
0.280	-0.260	0.480	-1.060	-0.990	-1.080	0.800	0.580	0.120
0.280	-0.260	-0.480	-0.420	0.400	-0.180	-0.320	0.430	-1.090
0.720	0.580	0.800	0.550	0.430	0.310	0.200	0.390	-0.91
-1.200	0.120	0.680	-1.090	-0.750	-0.730	0.300	0.190	-0.900
-0.360	0.000	-0.230	0.760	0.000	-0.630	-1.190	-1.030	0.180
-0.590	-0.460	-0.570	-0.650	-1.060	0.820	-1.170	-0.560	-0.090
-0.240	0.570	-0.330	-0.870	0.630	-0.100	0.290	-0.550	0.080
								-0.390

Orbitrap ESI- (65 points)

3.050	0.750	0.000	1.420	0.230	0.800	-0.480	-0.880	-1.270
2.650	1.360	0.270	1.650	0.220	0.800	0.900	-2.870	
3.180	-2.010	0.270	-0.930	-0.430	0.000	-1.170	-0.980	
0.510	-0.620	1.610	0.230	1.260	-0.950	-0.140	-0.360	
0.920	0.000	-0.520	-1.630	0.210	-0.950	-1.400	-0.650	
3.040	0.920	1.040	0.700	0.410	-0.760	-1.020	-1.390	
0.430	0.580	0.770	1.150	2.430	0.000	-0.640	0.320	
0.750	-0.290	2.250	-0.460	1.010	0.730	-1.630	-1.690	

MAT95 FAB+ (58 points)

0.000	-0.390	-0.300	0.550	0.500	0.480	0.420	-0.530	1.610
-0.520	0.920	-0.290	-0.410	0.370	0.590	-0.620	1.640	0.860
-0.970	1.000	0.580	-0.140	-0.620	-0.350	0.610	0.970	
0.470	1.140	-0.720	-1.440	-0.120	0.580	-1.160	0.080	
-0.440	-1.300	0.990	-0.130	-0.120	-0.110	-0.870	-0.940	
-0.210	-0.330	-1.130	-0.390	-0.610	0.110	0.370	-0.770	
0.610	-0.810	0.000	-0.130	0.360	0.750	-0.630	-0.960	

ORBITRAP ESI+ (784 points)

-2.290	-1.010	1.730	1.170	0.700	0.310	0.290	-1.340	0.000
-2.170	1.000	0.870	0.390	1.740	0.620	0.290	1.880	1.480
-2.130	1.500	-0.870	0.000	-0.690	1.240	0.290	-0.530	1.980
-1.420	0.000	0.430	1.920	-0.690	0.310	0.290	0.000	0.740
-1.340	0.000	0.000	-0.380	-1.030	0.310	2.610	0.530	0.250
-2.000	-0.980	0.000	-0.380	1.370	-0.620	0.870	0.260	0.740
-2.660	-0.980	0.860	0.380	-2.400	1.850	1.160	1.320	1.720
-1.990	0.980	0.430	4.580	1.020	0.930	0.000	0.000	0.740
-0.660	0.980	2.150	0.380	-2.390	1.230	2.300	0.000	0.980
-1.310	0.000	2.140	1.890	0.680	0.310	1.150	2.090	-0.250
0.000	-1.460	1.720	2.250	0.340	0.620	-0.290	-0.780	1.470
0.000	-0.490	-2.140	1.870	0.670	-0.920	0.860	-0.260	1.710
0.000	-2.430	-0.850	1.120	1.010	0.620	1.720	1.040	0.730
-0.620	0.480	0.850	0.370	0.340	2.460	0.570	0.780	-1.710
-2.450	0.000	-0.850	1.860	0.670	2.150	1.710	1.820	-1.710
-0.610	-0.480	1.700	2.240	1.010	2.150	0.860	-2.080	0.490
-2.420	0.000	0.420	2.230	0.000	0.920	-1.710	0.520	-0.240
-2.410	0.000	0.850	0.370	1.340	-0.610	0.570	-1.550	0.000
0.000	0.950	0.850	1.480	2.010	1.220	1.700	-1.550	1.930
-1.200	-0.470	0.000	1.110	0.670	-0.920	2.270	1.550	0.720
0.600	0.000	1.690	1.100	0.330	0.910	1.700	1.800	0.720
-1.190	1.890	0.420	0.370	-0.330	-1.520	1.130	0.260	-0.720
-1.770	0.930	0.840	1.460	1.650	0.610	-1.130	1.800	0.720
-2.340	1.870	0.000	1.830	0.660	0.000	-1.130	0.770	-0.240
0.000	0.000	0.840	1.090	2.970	-0.300	1.680	1.280	0.000
-1.160	-0.460	0.830	2.170	1.320	1.510	-1.120	0.000	0.000
-0.570	-2.310	1.670	0.720	-0.330	1.810	-0.560	1.790	0.240
-1.720	-0.460	-0.830	1.080	1.310	-0.600	0.560	1.790	-0.480
0.570	1.380	2.490	0.360	1.960	1.200	0.840	-1.280	0.960
0.000	0.460	2.060	-1.080	2.270	0.600	1.670	-1.020	1.440
1.130	-0.460	0.410	0.360	1.300	0.300	1.670	0.510	1.440
-1.120	0.000	1.230	0.720	1.610	0.900	1.670	0.760	1.910
0.560	0.450	-2.450	0.000	2.260	-0.890	1.940	-1.520	1.910
-1.650	1.810	1.630	2.150	0.640	1.790	0.830	1.010	0.720
-1.100	1.800	1.630	2.500	1.610	0.000	0.830	0.510	3.090
-3.240	1.350	-1.620	2.140	1.610	0.300	0.550	0.760	0.950
0.000	1.340	0.810	0.360	1.920	-1.480	1.100	1.010	0.950
-1.580	-1.340	2.020	1.060	0.640	0.590	1.650	-1.510	0.950
0.000	-0.900	0.000	1.410	1.590	0.890	1.090	-1.510	1.430
0.000	-0.450	0.800	1.060	0.960	1.480	1.370	0.250	-0.470
-0.520	1.340	1.600	-1.760	2.220	0.890	1.910	0.000	-0.470
-1.040	1.340	0.400	-1.760	0.000	0.000	1.370	-0.250	0.710
0.520	0.440	1.190	-0.350	-0.320	1.180	1.370	-0.500	-1.180
0.520	0.880	1.980	1.050	2.220	-0.590	0.540	3.760	-0.940
-1.030	2.210	2.380	0.350	0.950	2.060	-2.710	0.500	-0.230
-1.020	-0.440	0.790	1.750	1.580	1.760	0.810	-0.250	-0.940
-1.020	0.000	1.190	2.100	0.320	0.290	0.810	-0.750	1.640

1.530	0.880	1.970	0.700	-0.630	1.460	1.620	1.250	-0.940
1.020	-0.880	2.350	-1.400	1.870	0.290	1.890	0.500	-1.170
1.010	0.440	-0.390	0.700	-0.940	0.580	1.350	0.750	0.930
-1.510	0.000	1.170	0.700	0.310	0.880	0.540	0.500	-1.170
0.930	-1.760	-1.040	-1.970	-0.920	-2.260	0.420	0.120	0.400
0.000	0.660	-0.210	0.000	-0.180	-0.960	-1.550	-0.730	0.200
-1.160	0.220	0.420	0.590	0.000	-0.320	-0.280	-1.710	1.000
-0.930	-1.970	0.420	0.000	-1.810	-1.760	0.850	-0.850	1.000
0.930	0.220	0.620	0.000	1.800	0.960	0.420	0.000	1.700
0.000	-2.840	0.620	-0.390	-1.620	0.640	0.410	1.200	0.190
0.690	1.090	0.210	0.200	-1.620	0.320	-1.100	-1.320	-0.870
0.000	-2.400	0.620	-0.980	-1.970	-0.320	-0.280	0.240	1.520
1.150	-1.520	0.620	0.000	-0.180	-1.110	-0.280	-2.150	0.480
-0.460	0.650	-1.230	0.000	-1.250	0.470	-0.140	0.240	-0.550
-0.230	-1.520	-0.820	-0.590	-0.360	0.780	-0.270	0.480	0.280
-0.680	-1.730	0.000	-1.950	-1.070	-0.310	0.000	0.600	0.460
-0.680	0.220	-2.670	0.000	-1.580	0.000	0.000	0.590	1.640
-0.230	0.430	0.820	-1.750	-0.350	1.080	-0.950	1.520	1.320
1.360	-0.650	-0.200	0.190	-0.530	0.460	0.000	0.120	0.960
-0.450	-2.140	1.230	-2.130	-0.170	-0.310	-0.950	1.050	-0.260
0.230	0.000	-0.200	-2.710	-0.170	-0.610	-1.220	1.050	-0.090
0.450	-0.210	-0.200	-1.930	-0.350	-0.910	-0.400	0.120	-0.170
1.810	-1.920	-1.830	0.000	-2.740	-1.670	-1.730	1.050	0.170
0.680	-1.710	-1.630	1.920	-1.020	-1.820	-1.600	1.280	0.910
-1.360	-0.430	-1.020	-0.380	-0.510	0.750	-0.930	1.860	0.990
-0.230	1.280	-0.200	-0.380	-0.850	0.900	-2.240	-0.350	-0.890
-0.230	-1.270	0.410	0.190	-1.350	-0.450	0.260	-1.030	-0.880
-0.450	0.420	-1.210	-0.960	-0.680	0.000	-1.940	-2.270	1.710
-0.670	1.060	-2.020	-1.520	0.500	-1.610	-0.130	-0.110	0.590
0.450	-0.840	1.010	-1.140	-1.670	-0.290	-1.910	-0.670	0.280
-2.230	0.630	-1.410	-0.760	-2.170	0.880	1.910	0.220	0.560
-0.220	0.630	0.400	1.320	-2.340	-1.170	-0.380	0.220	-1.350
-0.220	0.630	-2.010	0.940	0.330	-1.300	0.630	-0.540	0.000
0.000	1.680	-0.200	-0.750	0.830	-0.290	0.620	1.090	
1.340	-1.260	0.200	0.000	-1.160	1.150	0.750	-0.210	
0.220	-2.100	-0.400	-0.560	-0.330	-1.300	-0.120	0.000	
0.000	-1.680	-1.610	-0.930	-1.320	0.430	-0.250	0.320	
0.000	-0.210	-1.000	-0.560	-0.160	-1.290	1.240	-1.150	
-1.770	0.840	-1.190	-1.120	-0.820	-0.140	0.620	0.520	
-1.320	-0.840	-2.180	-3.170	0.980	-2.150	0.370	-0.510	
-0.220	-1.250	1.190	0.560	-1.630	0.850	-0.120	0.500	

MAT95 ESI+ (408 points)

1.720	-0.900	0.360	0.650	-0.900	0.830	-1.000	-0.420	-0.500
0.700	-0.450	-1.090	-0.980	0.600	0.550	-1.000	-0.840	0.330
0.000	1.350	0.720	-1.310	1.500	0.550	-0.500	0.000	0.480
1.280	0.000	0.720	-0.650	0.000	-0.830	0.250	1.460	1.130
1.260	1.340	0.720	-0.330	1.500	-0.550	0.000	-0.210	-0.640
0.000	0.440	0.000	1.290	-1.800	-0.270	-1.220	0.620	-0.320

-1.870	-1.270	0.720	-1.930	0.600	-1.370	-0.730	-1.030	-0.640
-0.620	0.000	0.000	0.320	0.600	-0.820	-1.460	0.820	-0.760
0.610	0.420	1.430	-1.290	0.000	-0.540	0.240	0.410	-0.760
-0.610	0.000	1.070	-0.320	1.490	-0.810	1.460	-0.410	-0.900
0.000	-0.420	-0.350	-0.320	0.000	-0.270	-1.210	0.610	0.750
0.570	-0.420	1.060	-0.640	0.300	0.810	0.480	1.220	1.200
1.150	-0.830	-1.060	0.960	0.890	0.000	1.930	0.400	0.890
0.000	1.250	-0.350	-0.320	0.590	0.000	0.000	-1.010	0.570
1.100	0.410	1.050	-1.280	0.880	-0.540	0.480	0.600	-0.570
1.070	-0.830	1.050	-0.640	0.000	1.070	0.720	1.410	0.560
-1.070	0.410	0.350	-0.320	0.590	-0.540	0.240	0.200	-0.670
-0.530	-1.230	-0.350	0.000	0.590	-0.800	0.710	-0.400	0.260
-1.050	1.220	0.700	0.640	1.170	0.800	0.240	-1.400	-0.400
-2.070	0.820	1.050	-0.320	-0.880	0.270	0.000	-0.600	-0.790
-0.520	0.000	0.000	0.630	1.750	1.070	1.660	-0.590	-0.130
-0.510	0.400	-0.350	0.000	0.000	-0.530	0.000	-0.200	-0.250
0.000	0.810	0.350	0.000	0.580	-0.800	0.470	0.200	0.480
-1.010	0.800	0.000	0.000	0.290	0.530	0.470	0.200	0.710
-0.500	0.800	0.340	0.310	0.290	-1.320	1.410	0.400	0.120
1.010	0.000	-0.340	0.310	0.000	0.530	0.000	1.180	-0.810
0.980	-0.400	-1.370	0.620	1.440	1.320	0.460	0.780	-0.110
-0.490	1.590	0.690	1.560	-1.150	0.000	-1.150	0.000	0.220
1.470	0.790	-0.680	0.620	-1.150	-0.780	0.230	0.760	-1.030
0.490	0.790	1.020	0.930	-0.290	0.000	0.690	1.150	0.000
0.480	-0.390	0.000	-1.250	0.290	-0.520	-0.450	0.570	0.750
-0.960	1.930	1.020	0.310	0.280	0.000	-0.680	0.380	-0.280
0.960	0.000	1.020	-1.230	-0.280	-1.290	-0.230	0.940	
0.480	-1.530	1.020	0.310	-0.570	0.000	0.670	-1.490	
-0.960	-0.380	0.680	0.620	0.000	-0.260	0.670	0.190	
-1.890	0.760	0.000	-0.920	-0.570	0.770	0.670	-0.730	
-1.390	-0.760	1.020	-0.920	-0.280	-0.260	-1.110	-0.180	
1.390	-1.130	0.340	-0.310	-0.280	0.260	-0.440	0.910	
-0.930	-0.750	-0.340	-0.310	0.850	1.020	0.890	0.900	
-0.930	-0.380	-0.670	0.610	0.850	-1.520	-1.100	1.080	
0.920	1.130	-1.010	0.000	1.410	-1.010	0.660	-1.250	
0.920	0.750	1.000	0.610	-0.560	-0.760	-0.880	-0.710	
-0.460	1.490	0.330	0.610	-0.840	1.510	0.220	0.000	
-0.910	-0.370	-0.660	0.300	-0.280	-1.260	-0.430	-0.530	
0.450	0.370	-0.660	0.300	0.000	0.750	-0.210	0.530	
1.360	0.730	-0.330	0.610	0.280	0.000	-0.210	-0.860	
0.450	1.100	-0.330	0.600	0.830	-0.750	0.210	0.340	

MAT900 ESI+ (2317 points)

-0.930	0.650	0.000	-1.030	0.000	-0.450	-0.870	0.830	1.610
0.000	0.000	0.570	-0.510	-0.480	0.000	0.000	-0.420	3.210
0.000	-0.650	-0.570	1.030	0.000	-0.450	-1.740	-1.250	0.800
0.000	0.000	-0.570	0.510	-1.430	-0.450	-0.430	0.420	0.400
1.800	-0.640	-0.570	1.020	0.960	0.000	0.000	0.000	0.000
0.000	1.290	0.560	0.510	0.000	0.450	-0.870	1.250	-1.600

-1.720	1.920	-0.560	0.000	-0.480	0.900	0.000	1.240	-0.800
0.850	1.270	0.560	-0.510	-0.950	1.350	0.000	0.830	-0.400
-0.840	-0.630	-1.120	-0.510	0.000	1.350	-0.870	1.240	0.000
1.670	-0.630	1.120	-0.510	0.000	0.900	0.000	0.410	0.000
0.000	-1.260	-2.220	-1.010	-0.470	0.450	0.860	0.410	0.400
-0.810	-0.620	-1.670	1.510	0.470	1.790	-0.860	0.000	1.190
0.800	-1.250	1.110	0.500	0.000	-0.900	0.000	-0.830	-0.400
1.570	-1.240	0.560	0.500	0.940	-0.450	0.430	-0.830	-1.190
0.000	-1.240	1.670	1.000	1.410	-0.450	0.850	0.000	-0.400
0.000	-0.620	0.550	-0.500	0.940	-1.340	-0.850	1.240	1.590
0.000	1.240	0.550	-1.000	-1.410	0.000	-0.850	0.410	-0.400
0.770	-1.230	-1.100	0.000	-1.410	0.900	0.850	1.240	0.000
0.000	0.620	-0.550	-0.500	-0.470	-0.900	-0.430	-0.410	1.980
-1.480	-1.230	-0.550	0.500	-1.410	0.450	-0.850	-1.240	-1.590
1.480	0.610	0.000	0.000	0.000	-0.450	0.000	0.000	-0.790
0.000	-0.610	0.550	-0.500	0.940	1.780	0.850	0.830	0.000
-1.470	-1.200	0.000	0.990	0.930	1.340	0.430	0.000	1.580
-0.730	-0.600	-0.550	1.490	1.400	-0.890	0.850	0.410	0.000
0.000	0.000	1.090	-0.490	0.000	0.890	0.000	-0.410	-1.190
-0.730	-0.600	0.540	-0.490	0.470	0.890	-1.270	-0.820	-0.400
-0.730	-1.200	1.090	0.990	0.000	0.000	0.000	0.000	-0.390
1.460	0.600	-0.540	1.480	0.930	0.440	-0.420	-0.820	-0.790
0.720	1.190	0.540	0.490	0.460	-0.440	0.420	-0.410	0.790
-0.710	0.590	-1.080	-0.990	-0.930	0.890	-0.420	0.820	0.390
0.000	0.000	-0.540	1.480	0.930	0.000	-0.420	0.820	0.390
-0.710	0.000	1.070	0.490	0.000	-1.330	-1.270	0.820	-1.570
-8.370	0.590	-0.530	0.000	-0.460	0.440	-1.260	0.820	-0.790
0.000	0.000	1.060	0.980	-0.930	-0.440	0.420	0.000	0.390
1.390	1.180	-0.530	1.470	0.000	0.000	0.840	-0.820	-0.790
-0.690	0.000	2.120	-0.490	0.000	-1.770	0.840	0.820	-0.790
-1.380	0.000	0.000	0.490	0.460	-1.760	0.000	0.000	0.390
-0.690	-1.170	-1.580	0.000	-0.460	1.320	0.840	0.810	-0.790
0.000	1.160	-0.530	0.490	0.920	0.000	0.420	0.000	-2.350
0.000	1.160	1.580	-0.480	-1.380	-0.440	-0.420	-1.220	-0.390
2.050	-0.580	-0.520	-1.930	-0.460	0.880	0.840	-0.410	0.000
1.370	0.000	0.000	-0.480	-3.670	-1.320	-1.260	0.000	-1.180
-0.680	0.000	-2.080	0.000	-1.380	-0.880	-0.840	2.430	0.000
1.360	-0.580	-1.040	1.920	-3.650	1.310	-0.840	0.000	1.180
0.000	0.000	0.000	0.000	0.460	-0.870	0.840	0.810	-0.390
-0.670	0.580	0.520	-0.480	0.000	0.870	-0.420	-0.810	0.000
4.020	-0.570	0.000	0.480	1.360	0.440	0.420	0.810	-1.560
0.000	0.570	-1.550	0.960	0.910	-0.440	0.000	-0.810	-0.390
-1.970	0.570	1.030	-1.440	-0.910	0.440	3.750	0.400	0.000
-1.310	-0.570	0.510	-0.960	0.450	0.440	-1.250	-0.400	0.000
0.650	-1.150	-0.510	0.960	-0.450	0.000	1.250	0.810	-0.780
-1.560	0.380	0.000	-1.070	1.390	1.690	1.000	0.330	-1.600
-1.170	0.750	0.730	-0.360	-0.690	0.000	0.000	-0.330	0.960
1.170	0.380	0.730	0.000	1.040	-1.020	0.330	0.000	-2.140
0.390	-1.130	-0.730	-0.360	0.690	-0.340	-0.670	0.650	-0.320
0.000	-1.130	1.090	-0.360	-0.350	0.680	-1.330	0.000	-0.320
1.170	-1.130	0.730	-0.710	1.040	0.000	0.670	1.300	1.280
-0.780	0.750	-0.730	1.420	0.000	0.680	-0.670	1.300	-0.960
-0.780	-0.750	-0.360	0.710	-0.690	0.680	-0.670	-0.320	0.960

1.550	0.380	-0.730	0.360	-1.040	3.040	-0.330	-0.650	0.320
-0.390	-0.380	0.360	0.000	1.040	1.010	-1.000	0.320	0.640
-1.550	1.130	-0.360	0.350	0.000	-1.350	-0.330	0.970	-0.960
0.770	-0.380	-0.360	1.060	0.000	0.680	-0.660	0.650	0.320
-0.770	1.500	0.730	-0.710	-0.690	-0.340	-0.330	0.000	-0.320
1.160	-1.120	0.360	0.000	0.350	-0.680	-1.660	1.300	-1.590
0.770	-1.120	0.000	0.000	1.040	0.340	0.660	0.320	-0.640
1.160	0.000	1.090	0.710	1.380	0.340	-1.000	-0.320	0.320
-0.390	-0.370	-0.360	-1.060	0.000	0.340	1.000	0.650	-0.320
0.000	0.750	1.090	0.000	0.690	0.680	-0.660	0.000	0.000
0.380	-1.120	0.000	-0.350	1.380	0.670	1.000	0.650	0.950
0.770	-0.370	-0.360	-0.710	-0.690	-1.350	0.330	-0.970	1.590
0.770	0.370	1.080	-1.410	1.030	1.010	-1.660	-0.970	-0.320
0.000	1.120	-0.720	-0.710	-1.030	-0.340	-0.660	-0.320	-1.270
0.380	-0.370	0.360	-0.710	0.000	-0.340	-0.330	1.290	1.270
-0.380	0.370	1.440	1.060	-0.340	-0.670	0.660	0.640	-1.590
-0.380	0.740	0.720	-1.410	-0.340	0.670	-0.990	-0.970	-1.270
0.000	0.740	0.720	0.350	0.690	-1.010	-0.990	1.290	0.000
-1.150	0.370	1.080	-0.350	-0.340	0.000	-0.660	1.290	0.000
0.380	0.740	1.440	-1.060	-0.340	-0.670	0.000	0.000	-1.590
0.380	-0.370	0.720	-0.700	0.000	0.000	-0.660	-0.960	0.320
7.660	0.740	0.720	1.060	0.340	-0.340	-0.660	0.640	-0.320
-1.530	0.000	1.440	-0.350	-0.690	1.010	-1.320	-0.960	-0.950
1.530	-1.480	0.360	-0.700	1.030	0.670	-0.660	-0.960	-0.630
0.760	-0.740	0.720	0.000	1.030	1.340	0.330	0.640	0.320
0.000	-1.110	0.000	0.700	-1.710	0.000	-1.650	-0.640	-1.590
0.380	0.370	1.080	1.050	0.000	-0.340	-0.660	-0.640	-0.950
0.000	-0.370	-0.360	-0.350	0.680	-1.670	-0.330	0.640	0.630
-0.760	0.000	-1.790	1.400	-0.340	0.330	0.990	0.640	0.630
0.000	0.740	0.000	1.050	-1.030	0.670	-0.660	-0.960	-1.270
0.760	0.000	0.720	1.400	0.680	1.340	0.330	-0.640	-0.950
1.140	0.000	-0.360	0.700	-0.680	0.670	-0.330	-0.320	-0.950
-1.140	-0.740	1.070	-0.350	-0.680	1.000	-0.330	-0.640	-1.260
-0.760	0.740	0.000	-0.350	0.000	0.000	-1.310	-0.960	0.950
-0.760	-1.470	0.720	-0.350	-0.340	0.330	-0.330	-1.280	0.630
0.380	0.000	0.360	-0.700	0.000	0.000	0.000	-0.320	-0.320
0.380	-1.840	0.360	1.040	0.680	1.670	0.650	0.000	1.260
1.510	-0.370	0.360	-0.350	-1.360	1.670	1.630	-1.600	-0.630
0.000	-0.730	-2.500	0.350	-0.340	-0.670	-1.300	-0.320	-1.580
1.140	-1.100	-0.360	-0.350	-1.360	1.000	0.330	0.320	0.630
0.380	0.000	0.000	0.350	-1.020	1.670	-0.330	-0.960	-1.260
0.760	-0.730	-0.710	-0.690	-0.680	0.000	-0.980	1.600	-2.510
0.380	-0.730	-1.430	0.350	0.000	-0.330	-1.300	0.320	0.940
0.630	-0.920	-0.610	-0.300	0.870	0.280	0.000	0.000	0.800
-0.310	0.610	1.820	-0.890	-0.290	2.260	-0.280	0.810	-0.270
-0.630	0.000	-1.510	0.590	1.160	1.690	1.110	-0.810	2.120
0.620	0.310	-0.610	-0.890	1.160	-0.850	-0.560	0.000	0.790
0.620	0.920	1.210	0.890	-0.580	-0.280	-1.110	0.000	1.060
-0.310	1.530	-0.600	0.300	1.160	0.560	-0.560	-0.270	-1.060
-0.620	0.310	0.000	-1.180	0.000	0.000	0.280	1.630	-0.530
0.310	-0.920	0.600	1.180	0.290	0.560	-1.110	0.000	0.790
0.620	-0.610	0.000	0.290	-0.290	0.560	-1.110	-0.540	-0.790
0.310	-0.310	0.910	-0.590	1.730	0.560	-0.830	1.080	1.320

-0.310	-0.920	-0.600	1.770	0.870	-1.130	0.550	0.000	-0.260
0.310	0.000	0.600	0.880	1.160	0.840	0.000	0.000	-0.260
-1.240	-0.610	1.210	-1.180	-0.580	0.560	0.550	0.810	-1.050
1.240	-0.920	-0.300	0.880	-0.860	0.560	-0.550	0.000	1.050
-0.620	-0.610	-1.210	0.290	-0.580	0.840	1.100	-0.810	0.520
-0.310	-0.910	1.510	1.180	-0.580	0.000	0.550	0.540	0.000
0.310	0.300	0.000	0.000	0.290	0.000	0.830	-1.350	-0.260
0.930	0.000	0.000	-0.290	0.000	-0.280	0.550	-0.810	-0.520
0.000	0.910	0.900	-0.290	-0.580	-0.280	0.280	-0.540	-0.520
-0.620	-1.220	-0.900	0.000	-1.730	0.840	-1.100	0.000	1.050
-0.620	-1.520	-0.900	0.880	0.860	0.560	0.280	-0.810	1.050
-1.860	0.000	0.900	-0.290	-0.860	-1.680	0.830	0.270	-0.260
0.930	-0.300	0.900	0.000	-0.860	-0.280	0.000	-0.270	-0.520
1.240	0.610	-0.600	0.590	1.150	1.120	1.100	1.080	-0.260
-0.620	0.610	0.000	1.170	0.290	0.000	-0.550	1.350	1.310
-0.620	-0.300	0.900	-0.590	0.570	-1.400	0.270	-1.070	-0.260
0.620	1.220	-0.600	-1.760	0.290	1.120	0.270	0.000	0.520
-0.930	0.000	1.500	0.000	-1.430	-0.840	0.000	-0.810	-0.260
1.240	1.520	0.600	-0.880	1.430	0.280	0.550	-0.270	-1.300
0.310	-0.610	0.900	0.880	-1.150	-0.280	0.550	1.340	0.000
1.860	-1.220	0.900	0.580	-0.860	-1.120	1.100	1.070	-1.040
-1.550	-0.300	0.900	-0.290	-0.570	1.120	0.820	0.810	-0.520
1.240	-1.820	-0.600	-0.880	0.290	-0.840	0.000	1.070	1.040
1.240	0.300	-0.600	0.880	0.290	0.000	-0.550	0.270	-0.260
-0.310	0.000	-0.300	-1.170	0.290	0.840	1.640	-0.540	-0.780
1.230	1.220	1.490	0.290	0.280	-0.840	1.370	1.070	0.780
-0.930	1.520	0.000	0.000	0.570	0.000	0.550	-0.530	0.260
-0.620	0.910	-0.590	-1.460	1.140	0.000	0.550	0.270	-0.520
-0.930	-0.910	-0.590	0.580	-0.570	0.000	0.000	1.340	0.000
-0.310	0.000	0.300	-1.170	-1.140	0.560	0.270	-1.340	-0.260
0.000	0.910	0.300	0.000	-1.140	0.000	1.360	-1.870	0.780
-0.310	1.220	1.190	-0.870	0.570	0.840	0.540	0.530	1.040
0.920	-0.610	-1.190	1.460	0.000	-1.670	-0.270	1.330	1.040
0.620	-0.910	0.590	0.870	-0.570	0.840	0.820	1.330	-0.260
-1.230	0.000	0.890	0.000	-0.850	0.000	-0.270	0.800	-1.040
-0.310	0.610	-0.590	0.580	1.140	-0.280	-0.540	1.600	-0.780
0.310	-0.910	0.000	0.000	0.000	0.280	0.540	0.530	0.260
0.000	1.220	-0.890	0.290	0.850	-0.560	2.450	0.270	-0.260
0.000	0.000	0.300	0.000	1.420	0.280	-1.630	0.530	-0.260
1.230	1.520	0.000	1.160	-0.280	-0.280	0.540	0.530	-0.520
-0.610	0.300	-0.890	0.290	-1.130	-0.840	1.360	1.590	0.260
-0.780	0.500	0.250	0.000	-0.470	0.680	-0.220	1.280	1.860
-1.290	-1.010	0.740	-0.480	-1.400	-0.460	0.000	0.000	0.410
-0.780	0.000	-0.980	-0.950	-0.700	0.910	-0.660	0.850	1.240
-1.290	-0.500	0.240	1.670	1.160	1.370	0.440	0.000	-0.210
0.520	-0.250	-1.220	1.190	-0.230	-0.230	-0.220	0.210	-1.240
-0.260	-1.000	0.730	1.420	0.230	0.910	1.550	-0.210	0.210
-1.550	-0.750	1.220	0.000	0.700	1.590	-0.220	0.210	0.210
0.770	-0.750	0.730	0.710	-0.930	-0.230	-0.440	-0.210	1.850
0.260	-0.500	-0.490	-1.900	0.000	1.590	-0.880	-0.210	0.210
0.770	0.500	-0.730	-1.190	-0.460	-0.450	0.440	-0.850	-1.650
1.290	-0.500	1.460	0.470	-0.930	0.230	-0.440	-0.420	0.210
0.520	-0.250	1.710	0.950	-1.160	0.230	1.760	-0.420	-0.410

0.000	-2.000	0.000	-0.470	0.930	0.450	1.320	1.270	-0.210
0.510	0.000	-1.460	0.950	-0.460	0.450	-0.660	0.420	1.230
0.510	0.500	0.490	-0.950	-0.690	0.230	0.220	0.630	-1.230
1.540	1.000	2.680	-0.240	-0.930	-0.450	0.000	-0.840	0.000
-0.260	-0.750	-0.970	0.000	0.230	-0.230	-0.660	-0.630	-1.230
-0.770	-0.750	0.970	0.950	0.920	0.680	0.220	0.630	1.230
1.540	1.240	1.220	-1.180	0.460	-2.940	0.220	0.000	0.620
0.510	0.750	-0.240	-0.710	-1.390	-0.680	-0.870	0.420	-0.820
-0.770	-0.250	-0.730	-0.710	1.150	0.680	0.000	-0.210	0.410
0.000	-0.500	0.000	1.420	-0.690	1.350	0.000	0.000	-1.020
1.530	-0.500	0.000	-0.710	0.460	-0.230	0.220	1.880	-0.200
-0.510	-0.990	0.970	0.240	-0.690	-1.350	0.440	-0.420	0.610
0.250	0.990	-0.490	-0.710	0.000	0.230	0.430	0.210	-0.820
-1.270	-1.240	-0.490	0.470	-0.230	-0.680	0.430	0.830	1.430
0.000	0.500	-1.450	0.240	0.460	-1.120	-1.090	-0.210	0.410
-1.520	0.000	1.210	0.940	-0.230	-1.790	-0.870	1.880	1.630
-0.760	-0.250	0.480	0.710	0.000	0.900	-0.870	-0.630	0.410
0.760	-0.740	0.240	0.240	-0.230	-0.220	0.220	0.000	-0.410
-1.010	0.740	0.720	1.180	1.380	0.220	-0.870	-0.420	0.200
1.010	0.990	1.200	0.470	0.230	1.560	-0.650	-0.210	1.020
-0.510	-1.730	-0.480	1.180	-0.460	1.560	-0.430	0.420	1.020
0.760	0.490	-0.720	0.240	1.380	-0.670	-0.430	-1.250	-1.220
0.510	0.490	0.720	1.180	-0.690	1.560	-0.430	0.830	-0.410
-0.510	-0.990	0.000	0.000	0.000	1.110	0.860	0.210	0.610
-0.760	0.980	0.480	0.230	0.920	0.890	0.220	-0.210	-0.810
0.250	0.490	0.000	-0.230	-1.150	-0.670	-0.220	0.210	-0.610
1.520	0.980	-0.720	-0.700	-1.150	-1.110	0.430	0.210	-0.200
0.760	0.000	-0.480	-0.230	0.000	-0.670	-1.510	0.410	0.000
-1.270	0.250	0.960	0.700	1.380	0.220	-0.430	1.660	0.610
0.000	0.980	0.480	0.230	0.460	0.000	-0.650	1.450	-0.400
0.510	0.490	0.000	-0.230	-0.460	-0.670	0.220	0.830	-0.610
-0.500	1.230	0.960	1.640	-0.690	-1.330	-0.430	-0.620	-0.200
0.500	-0.740	0.950	3.980	-0.230	1.550	-0.430	-0.620	-0.610
0.500	0.490	1.430	-0.940	0.920	-1.770	0.640	-0.410	1.010
0.000	0.980	0.720	-0.930	-1.140	1.550	-1.070	0.210	1.410
-0.250	1.470	0.000	-0.700	-0.910	-0.890	-1.290	0.620	0.000
-1.010	-0.250	0.720	0.230	-0.230	-1.110	0.860	1.030	1.210
-1.010	-1.230	-0.720	-1.170	0.460	-0.440	-0.430	0.410	-0.600
0.000	0.740	-0.240	0.470	-0.460	0.660	-1.290	0.620	0.000
0.400	-1.360	0.370	0.530	-0.990	0.620	0.150	0.400	0.730
0.200	0.580	0.370	0.000	-0.490	-0.150	-1.160	-0.130	0.120
-1.410	-1.170	1.480	0.180	1.640	0.460	0.870	1.070	0.000
0.800	0.000	0.740	-0.180	-0.980	-1.240	1.160	0.400	-0.360
-0.400	0.970	0.920	0.180	1.480	-1.080	0.860	0.270	-1.550
0.000	-0.390	-0.370	-0.700	-1.150	0.150	0.290	-0.270	1.420
0.600	-0.190	0.740	-1.390	0.330	-0.460	-0.720	0.530	0.120
-0.600	-0.780	1.110	0.170	-1.150	0.150	1.150	1.330	-0.470
-0.200	-0.190	0.920	-1.040	-0.330	-0.310	0.000	1.060	-0.700
-0.600	-0.390	0.550	-0.350	1.640	-0.150	0.140	-1.720	0.700
0.600	-0.970	-0.550	-0.690	-1.470	0.000	-1.570	0.130	-0.350
0.200	0.000	-1.100	1.210	0.330	0.770	-0.710	0.780	0.350
-1.000	-1.350	-0.550	-0.690	-0.810	-0.920	1.000	1.300	-1.160
0.600	0.380	0.000	-0.170	-0.160	0.610	-0.140	1.040	-1.850

0.800	-0.190	-0.180	-0.860	1.300	-0.310	1.280	-0.390	-1.270
-1.400	-0.380	0.370	-0.690	-0.490	-0.310	-0.850	0.130	0.110
-1.400	0.380	-0.180	0.510	-0.650	0.760	-0.570	1.300	1.250
-0.200	0.960	0.180	-0.510	0.970	-1.680	1.710	-1.040	1.020
1.790	1.150	1.090	-1.710	1.130	0.610	0.850	1.170	1.020
0.990	-1.140	-0.360	-1.700	1.130	0.150	0.280	1.300	0.790
-0.990	0.000	-0.730	0.000	-0.810	1.210	0.000	1.040	-0.340
0.200	-1.140	1.090	0.680	1.130	-0.910	-0.140	0.520	1.010
-1.790	-0.190	0.720	-0.850	0.640	1.660	0.850	0.130	1.680
-1.980	-0.380	0.000	0.170	0.800	-0.600	-0.850	0.390	1.680
-0.400	-0.190	0.000	-1.020	0.960	1.510	-0.980	-0.390	0.780
0.200	-0.380	-1.270	0.510	1.600	1.350	0.000	0.770	-1.790
-0.200	0.190	1.270	-1.700	-0.160	0.450	0.700	1.270	1.560
0.790	0.000	0.720	0.340	1.120	0.150	1.260	-0.510	-0.670
0.000	0.000	0.360	0.340	0.480	0.000	0.840	0.630	0.990
-0.390	-0.380	0.180	-1.520	0.160	1.200	0.280	1.880	0.220
-0.590	-0.570	0.180	1.510	-0.160	-0.150	-0.560	0.130	1.100
0.000	-0.940	0.900	-0.170	0.160	-0.890	0.970	0.120	0.110
0.980	0.000	-1.610	0.340	-0.640	1.190	-0.140	-0.250	-0.440
-1.770	-1.320	0.360	-0.500	0.800	-1.040	-0.140	1.370	-0.870
0.000	0.380	-0.540	0.170	0.000	-1.480	0.280	0.370	-1.470
0.200	0.000	-1.960	0.170	-0.640	1.480	0.000	-0.860	0.100
-1.370	0.560	-1.070	0.170	-0.790	-1.040	-0.550	-0.860	1.040
-0.200	-0.750	-0.530	0.500	1.430	-0.300	0.550	-0.610	-0.730
0.390	1.120	1.070	0.670	0.320	-0.440	-0.680	-0.740	-1.750
-0.980	1.120	1.250	0.170	0.160	-1.480	-0.140	0.980	0.000
-0.390	-0.190	0.000	0.330	-0.320	-0.150	-0.680	0.370	-0.310
0.000	1.490	0.180	0.000	1.100	-0.440	0.680	-0.490	-0.200
1.180	-0.560	0.000	1.490	-1.890	0.290	-0.550	1.460	0.400
0.200	1.310	-0.530	-1.160	0.470	-1.770	0.270	-0.850	0.400
0.000	-1.310	-0.710	0.660	0.630	0.000	0.550	-0.240	0.800
0.000	-0.370	1.420	1.480	-0.160	-0.590	-0.140	0.610	0.000
-1.170	-1.670	0.180	0.490	0.470	0.730	0.820	1.700	-1.090
-1.370	-0.370	1.240	-0.330	-0.470	-0.730	1.360	-1.700	-1.480
-0.590	0.000	-0.530	0.160	-0.160	0.150	0.000	-1.700	0.590
-0.190	-0.370	-0.180	0.160	-0.310	0.000	0.950	-0.970	
-1.170	-0.930	0.000	-0.490	0.310	0.880	-0.130	1.940	
-0.200	-1.940	0.380	-1.760	-1.300	1.280	1.490	1.470	
-0.490	-0.290	0.850	0.700	-1.380	0.720	0.840	-0.420	
-1.660	0.570	-0.090	-1.480	1.550	1.570	0.500	-1.680	

TOF MALDI+ (195 points)

-4.880	-5.360	2.370	-5.570	-3.620	2.780	1.440	0.350	1.190
-4.060	-4.260	2.370	-4.640	-3.020	2.920	1.710	1.760	2.160
-4.060	-4.100	2.680	-2.940	-2.420	6.280	1.970	-1.940	2.920
1.220	-3.780	3.000	-2.790	-2.270	-11.810	2.100	0.110	4.860
4.060	-2.840	3.780	-1.240	-1.960	-4.330	2.620	0.340	8.960
-2.110	-2.520	3.940	-0.460	1.210	-3.410	10.760	3.420	-3.400
1.900	-2.520	4.420	-2.630	1.360	-3.280	-3.100	4.910	-3.110

4.010	-2.210	4.570	0.150	1.510	-3.020	-1.940	-7.180	-2.330
4.010	-2.050	5.360	1.700	6.340	-2.490	-1.680	-5.810	-1.070
4.430	-1.580	5.050	2.160	-2.410	-2.230	-0.520	-2.510	3.300
-5.240	-1.420	-2.360	2.780	-1.350	-2.100	0.770	-2.390	-6.700
-4.280	-1.260	-1.420	3.400	-0.450	-2.100	-4.470	-2.170	2.590
-2.700	-0.790	-0.630	-3.780	0.900	-1.840	-3.760	-1.820	-9.930
-2.220	-0.470	-0.320	-1.960	3.760	-1.840	-3.060	-0.570	-3.800
-2.060	-0.160	0.160	-0.150	4.660	-1.710	-2.940	-0.340	-3.710
3.970	0.320	1.100	0.450	-12.420	-0.790	-2.700	1.020	3.020
-3.320	0.320	-2.190	1.060	-8.330	-0.660	-0.940	1.590	3.110
-3.010	0.470	-1.720	1.210	-5.990	-0.660	-6.670	1.930	4.490
0.160	0.950	0.160	1.210	-0.730	-0.390	-3.280	1.930	6.130
1.110	1.730	0.630	4.680	0.730	0.260	-2.580	-3.670	
2.370	2.210	1.720	-4.230	2.480	0.390	-1.870	0.430	
3.800	2.210	2.660	-3.930	2.630	0.660	-1.170	0.540	