2018

# Fast Isotope Ratio Mass Spectrometry (FIRMS): A Tandem Mass Spectrometry Technique for the Rapid and Semi-Comprehensive Evaluation of Isotope Ratios

Fredrick Ochieng
*South Dakota State University*

# FAST ISOTOPE RATIO MASS SPECTROMETRY (FIRMS): A TANDEM MASS SPECTROMETRY TECHNIQUE FOR THE RAPID AND SEMI-COMPREHENSIVE EVALUATION OF ISOTOPE RATIOS

BY

FREDRICK OCHIENG

A dissertation submitted in partial fulfillment of the requirements for the

Doctor of Philosophy

Major in Chemistry

South Dakota State University

2018

FAST ISOTOPE RATIO MASS SPECTROMETRY (FIRMS): A TANDEM MASS SPECTROMETRY TECHNIQUE FOR THE RAPID AND SEMI-COMPREHENSIVE EVALUATION OF ISOTOPE RATIOS

FREDRICK OCHIENG

This dissertation is approved as a creditable and independent investigation by a candidate for the Doctor of Philosophy in Chemistry degree and is acceptable for meeting the dissertation requirements for this degree. Acceptance of this dissertation does not imply that the conclusions reached by the candidate are necessarily the conclusions of the major department.

Brian A. Logue, Ph.D.
Dissertation Advisor Date

Douglas Raynie, Ph.D.
Head, Department of Chemistry & Biochemistry Date

Dean, Graduate School Date

I would like to dedicate this dissertation to the memory of my beloved mother, Jenipher Awino Ombul. Despite the adversities and challenges she faced raising her seven children, she stood strong every day and provided for us. I am forever grateful for your unconditional love and support.

This dissertation would not have been possible without the love and constant support of my wife, Dzigbordi Akpoblu. You have encouraged me to stay focused on my goals during the most difficult times in this journey. Thank you for not only standing with me with unconditional love but also for supporting me throughout my work to achieve my goals and dreams. This work and many more to come are devoted to you.

I would like to thank my siblings, my in-laws, family, and friends especially Daniel and Sharice Otieno for the support and joy you all bring in my life every day, and for cheering me up during difficult times in graduate school. Each of you hold an important place in my heart. Thank you for your motivation, and without your companionship and word of encouragement I would not be where I am today.

# ACKNOWLEDGEMENT

This work would not have been possible without the exceptional guidance and support of my advisor, Professor. Brian A. Logue. You have played the most significant role in shaping my success in graduate school. My sincere gratitude and most heartfelt appreciation are all dedicated to you. Thank you for trusting me with this project, believing in me and constantly reminding me of my potential.

I am grateful to all of those with whom I have had the pleasure to work with in this project. I am especially indebted to Dr. Paul J. Hinker, Assistant Professor of Math and Computer Science at South Dakota School of Mines and Technology for playing the important role in the development of FIRMS software. I would also like to recognize Zachery Crandall of South Dakota School of Mines and Technology for his work in the development of FIRMS software. Thank you to Dennis Adjei-baah my long-term friend and software engineer for his support, advise and ideas on computer science programming. Also, thank you to William Simon of Princeton University for his contribution to FIRMS through REU summer program.

Thank you to the members of LARGE group for the constant inspiration, unwavering support, and guidance. I value and treasure you all because you are my inspiration and strength in my everyday work in the lab.

I would like to thank my committee members, the Department of Chemistry and Biochemistry and Department of Agronomy, Horticulture, and Plant Science at South Dakota State University, and all the faculty and staff for their assistance and encouragement. My honest gratitude also goes to the institutions and agencies that have funded my research and made the completion of my degree a possibility.

TABLE OF CONTENTS

# ABBREVIATIONS

ADT: Abstract Data Type

API: Active pharmaceutical ingredients

$CDCl_3$: Deuterated chloroform

CD-3: Counter Detection Device

CE: Collision energy

CEO: Chief executive officer

CF-IRMS: Continuous-flow Isotope Ratio Mass Spectrometry

CGI: Common Gate-way Interface

CID: Collision-induced dissociation

CML: Chemical Markup Language

$^{13}C$ NMR: Carbon-13 Nuclear Magnetic Resonance

$^{13}C$ SNIF-NMR: Carbon-13 Site-specific Natural Fractionation Nuclear Magnetic Resonance

CRDS: Cavity-Ring Down Spectroscopy

CRS6: Controlled random search algorithm

CTO: Chief technology officer

DCM: Dichloromethane

DIMPA: Diisopropyl methyl phosphonic acid

DLSI: Direct liquid sample introduction

DM: Decision maker

DP: Declustering potential

EA: Evolutionary Algorithms

EA-IRMS: Elemental-Analyzer Isotope Ratio Mass Spectrometry

EAS: Electrostatic analyzer

EI: Electron impact

EMPA: Ethyl methyl phosphonic acid

EP: Evolution Programs

FCIS: Fast Chemical Identification System

FIRMS: Fast Isotope Ratio Mass Spectrometry

GA: Genetic Algorithms

GC: Gas chromatography

GC-IRMS: Gas-Chromatography Isotope Ratio Mass Spectrometry

GC-C-IRMS: Gas Chromatography Combustion Isotope Ratio Mass Spectrometry

GPHF: Global Pharma Health Fund

GRG: Generalized Reduced Gradient

GUI: Graphical user interface

HCl: Hydrochloric acid

$^{1}$H NMR:  Proton Nuclear Magnetic Resonance

HPLC: High Pressure Liquid Chromatography

$^{2}$H SNIF-NMR: Proton Site-specific Natural Fractionation Nuclear Magnetic Resonance

HTML: Hypertext Markup Language

ILAS: Infrared Laser Absorption Spectroscopy

IMPA: Isopropyl methyl phosphonic acid

IR: Isotope ratio

IRMS: Isotope Ratio Mass Spectrometry

IRs: Isotope ratios

LC: Liquid chromatography

LIFO: Last-In-First-Out

LTM: Long-term memory

MOO: Multi objective optimization

MOF: Multi objective function

MRM: Multiple reaction monitoring

MS: Mass Spectrometry

MTM: Medium term memory

NAD: Nandrolone

NQR: Nuclear Quadrupole Resonance

NIR: Near-infrared

NMR: Nuclear Magnetic Resonance

NC: No coffee

NNW: No news watch

NSF: Nonsmooth functions

NW: News watch

PAHs: Polycyclic aromatic hydrocarbons

P-XRD: Powder X-Ray Diffraction

QCL: Quantum Cascade Laser (QCL)

RS: Ranking sum

SNIF-NMR: Site-specific Natural Fractionation Nuclear Magnetic Resonance

SOAP: Single Object Access Protocol

SOF: Single objective function

SOO: Single objective optimization

SPME: Solid Phase Extraction

SSFFC: Substandard, spurious, falsely labeled, falsified and counterfeit

SS-NMR: Solid-state Nuclear Magnetic Resonance

SS: Scatters Search

STM: Short-term memory

TC/EA-IRMS: High Temperature Thermal Conversion Elemental Analyzer Isotope Ratio Mass Spectrometry

TILDAS: Tunable Infrared Laser Direct Absorption Spectroscopy

TLC: Thin-layer chromatography

TS: Tabu Search

U.S. EPA: United States Environmental Protection Agency

WHO: World Health Organization

WSM: Weighted sum method

XML: Extensible Markup Language

XRD: X-Ray Diffraction

LIST OF FIGURES

# LIST OF TABLES

# ABSTRACT

## FAST ISOTOPE RATIO MASS SPECTROMETRY (FIRMS): A TANDEM MASS SPECTROMETRY TECHNIQUE FOR THE RAPID AND SEMI-COMPREHENSIVE EVALUATION OF ISOTOPE RATIOS

FREDRICK OCHIENG

2018

Isotope ratios (IRs) are a measure of the variation in abundance of light isotopes versus heavy isotopes of an element, e.g. $^{1}H/^{2}H$,$^{12}C/^{13}C$, and $^{14}N/^{15}N$. Because IRs change as a molecule undergoes certain chemical and physical processes, they represent unique fingerprints that can be used to trace the source of a chemical compound. Current techniques for determination of IRs each have major limitations, such as loss of structural information, vulnerability to contamination, high cost, large sample size requirements, low precision, limited applicability, and lengthy analysis. Fast Isotope Ratio Mass Spectrometry (FIRMS) was developed as a next-generation analytical technique for robust measurement of IRs using tandem mass spectrometric data. Based on the FIRMS nonlinear mathematical model, the difference between predicted and experimental tandem mass spectrometric data was minimized by modifying isotopic fractional abundances of the atoms involved. FIRMS was used to calculate the isotope abundances of molecules of 17 compounds. FIRMS produced excellent precision, with IRs standard deviations of less than

1‰, for $\delta^{13}C$ compared with IRMS in a fraction of the time. FIRMS also produced excellent accuracy for $\delta^{13}C$ values with an absolute error of less than 1.6‰ compared to IRMS. It was also successfully coupled with LCMSMS to analyze $\delta^{13}C$ for DIMPA, caffeine and acetylsalicylic acid with standard deviation of less than 1‰ and accuracies within 1‰ compared to IRMS. The bivariate plot of FIRMS versus IRMS showed an excellent $R^2$ of 0.992 with a slope of 1, validating the new technique as an accurate isotope ratio measurement technique. FIRMS offers several advantages over other methods of IR determination, and may ultimately become the preferred method for determination of IRs. The disciplines that will benefit from development of FIRMS include: 1) forensic science, 2) environmental chemistry, 3) geology, 4) geochemistry, 5) cosmochemistry, 6) food sciences, 7) earth science, and 8) pharmaceutical science. FIRMS will make IR analysis more widely available, fast and less expensive for scientists to conduct research in these fields, as well as train new graduate students for future scientific research and development in this area.

# CHAPTER 1: Introduction

## 1.1 Motivation and Objective

Counterfeiting is a threat to every aspect of the world economy, including innovation, technology, and industrial relations, such as production and manufacturing. Commonly used materials and products that require incredible input in research, development and marketing are frequently targeted.[1] Counterfeiting has grown into enterprise primarily because of weak regulation in many countries. Research shows that the regulations target the supply of fake goods and do not address the high consumer demand for cheap and affordable products. [2 3]

Thus, tools and mechanisms must be put in place to fight the supply of counterfeited products. For example, developing a fast and reliable technology capable of tracing the source of counterfeit products made from organic materials can be used to determine the authenticity of goods. Tracing organic products is possible because chemical compounds are composed of atoms, most of which have stable isotopic forms.[4-6] Using isotope tracing methods [7-9] and carbon dating[10] products can be authenticated.

Isotope ratios (IRs) are used to trace the history of compounds and a measure of the ratio abundance of light isotopes versus heavy isotopes of an element, such as $^{1}H/^{2}H$, $^{12}C/^{13}C$, $^{14}N/^{15}N$, $^{35}Cl/^{37}Cl$ and $^{79}Br/^{81}Br$. These ratios are largely quite consistent, with the original isotope frequencies of each terrestrial element fixed upon the formation of the earth.[11] But, experimental observations show changes in relative isotope abundances occur over time due to natural phenomena, [11] including radioactivity,[12-13] anthropogenic activities,[14] cosmic ray interaction,[15-17] and fractionation.[11] Radioactive decay affects IRs by reducing the abundance of heavier isotopes, such as $^{13}C$, $^{34}S$ and $^{81}Br$[12]. Cosmic rays,

which occasionally enter the earth's magnetic field, may result in neutron capture, thereby changing the isotope ratios of certain atoms. [17-19] For example, $^{36}Cl$ formation occurs when $^{35}Cl$ captures neutrons produced in muon capture and muon-induced photodisintegration reactions. [20] Anthropogenic activities, such as industrial nuclear reactions and mining affects IRs by enriching some isotopes.[21] Fractionation is a process where different isotopes of an element produce slightly different chemical and physical properties, resulting in changes in isotope ratios as molecules undergo certain chemical and physical processes.[22] For example, in biological processes, some factors, such as genetics, affect the fractionation process as plants utilize $CO_2$ for photosynthesis. Plant types follow different photosynthesis pathways and processes that result in unique IRs. These ratios from each plant food source become integrated into an animal's biology based on their diets. Table 1.1 shows the photosynthetic cycle and $\delta^{13}C$ values for different plant types (i.e., $\delta$ is a measure of $^{13}C/^{12}C$ compared to a reference material and represented as a "per mil" or parts per thousand, denoted as ‰).

**Table 1.1.** Photosynthesis cycle and $\delta^{13}C$ values of different plants

| **Plant type** | **Photosynthesis cycle** | **Examples** | **$\delta^{13}C$ values range (‰)** |
|---|---|---|---|
| $C_3$ (dicotyledonous) | Calvin-Benson | Wheat and beans | -23 to -22[23] |
| $C_4$ (monocotyledonous) | Hatch-Slack | Millet and corn | -8 to -20 [23-24] |
| $C_3/C_4$ | Benson and Hatch-Slack | Pineapples and orchid flower | -10 to -34[11] |

The change in IRs as a molecule undergoes certain chemical and physical processes represents unique fingerprints that can be used to determine the history of a

chemical compound. For example, animals from North America are distinguished from European animals because of different $\delta^{13}C$ values. European meat products have lower levels of $\delta^{13}C$ because they feed on $C_3$ plants, while North Americans have higher $\delta^{13}C$ values because they feed on $C_4$ plants.[11]

Although IRs are extremely important in a number of areas, it continues to be very difficult and costly to determine IRs from currently available techniques. Thus, the objective of this study was to develop a laboratory-based analytical technique, known as Fast Isotope Ratio Mass Spectrometry (FIRMS), as a next-generation analysis method. FIRMS uses tandem mass spectrometry data and a nonlinear mathematical model describing the probability of compound fragmentation to extract IRs of locations of individual atoms or groups of atoms within a molecule. It is a rapid, accurate, sensitive, robust, and more comprehensive method that evaluates IRs with promise to definitively identify the source/history of the chemical components of the material.

**1.2 Importance of isotope ratios (IRs)**

Isotope ratios (IRs) have remained a significant area of research interest for many decades as documented by the number of scientific publications. Applications of IRs have grown over time in areas that include forensic science, food and drugs analysis, diet, biochemistry and metabolism, doping investigations, and environmental pollution.[11] Figure 1.1 shows the number of publications based on isotope research work from the year 1950 to 2017. The figure shows that the number of publications with IRs as a subject of interest greatly increased over the past seven decades, especially from 1950-1990.[11]

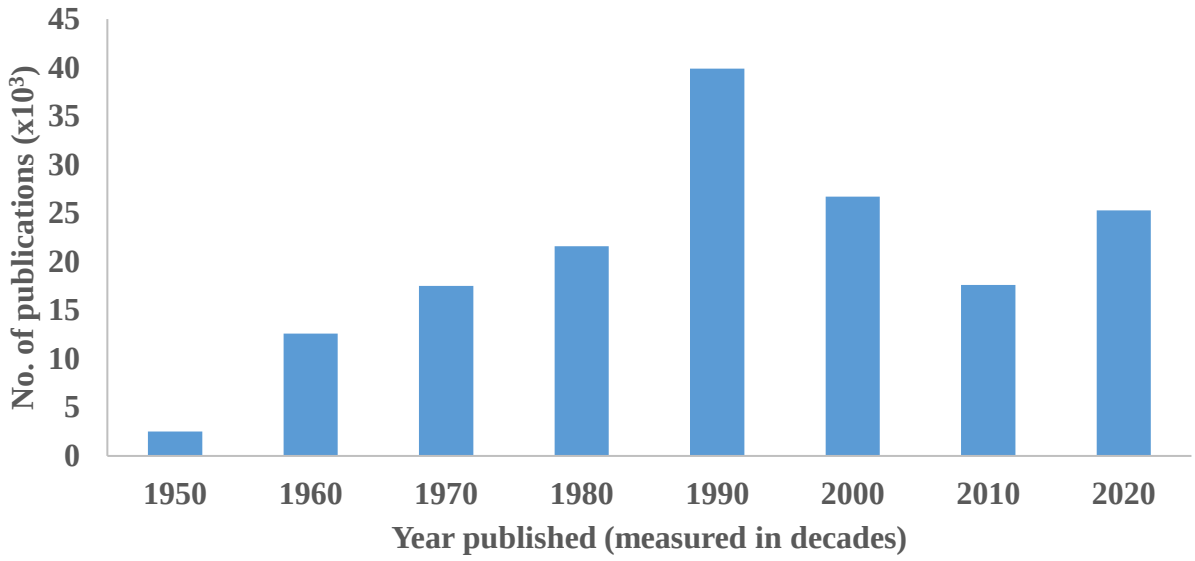



**Figure 1.1.** Shows the number of publications per decades for IRs research from the year 1950 to 2017 obtained from Google Scholar on May 4, 2017.

In forensic science, IRs have been used to identify the source of counterfeit materials by linking the evidence to a particular geographical place or a person.[25] In the past, these materials have included: explosives, ignitable liquids,[25] illicit drugs,[25-26] prehistoric evidence,[27] and nuclear arms.[28] For food, beverage, and water safety IRs are used to identify the origin of products,[29-31] confirm if components indicated by food and beverage manufacturer agree with the actual content of the product, and to analyze pollutants in drinking water.[11] IRs are also used to establish the dietary pattern, metabolic class, and movement of domestic animals such as cattle.[11] In suspected athlete doping activities, IRs are used to analyze performance enhancing compounds that are thought to be synthetic. For example, gas-chromatography IRMS (GC-IRMS) has been used to analyze carbon isotope ratios of nandrolone (NAD). NAD is an endogenous steroid hormone metabolite used by some athletes, or in racehorses to enhance performance.[32] Similarly, elemental-analyzer IRMS (EA-IRMS) has been used to distinguish between

endogenous and exogenous concentrations of synthetic steroid 19-norandrosterone, a metabolite of 19-nortestosterone by using isotopically labeled standards to compare isotopic carbon ratios and determine the origin of urinary norandrosterone.[33] Environmental pollution agencies (e.g., U.S. EPA) employ GC-IRMS methods for pollution control and health safety to analyze carbon isotopes of pollutants, such as polycyclic aromatic hydrocarbons (PAHs), to identify the source of pollutants.[11, 34]

In recent years, IRs have been expanded to include identification of counterfeit drugs. Counterfeiting medicine has led to a significant number of fatalities. According to World Health Organization (WHO), substandard, spurious, falsely labeled, falsified and counterfeit (SSFFC) products have continued to undermine the treatment of deadly diseases such as malaria and tuberculosis. WHO has reported that counterfeit drugs include over 920 medical products, such as therapeutics, and generic medicines, antimalarials and antibiotics as the most counterfeited drugs globally.[35]

Counterfeit drugs disproportionately exist in illegal markets (unregulated pharmacies, clinics, and hospitals) and online. Their production is done under unhygienic conditions by unqualified workers. These false medicines might contain nonmalignant ingredients such as flour, sugar, and pollen, instead of active pharmaceutical ingredients (API); may contain lower doses of the API than the authentic drug; may contain components, such as aspirin that are capable of feigning treatment; or an entirely illegal API that can be poisonous.[35]

Falsifying products is a practice that has thrived because of the limitations of current technologies as well as weak laws. Given the complexity and effects of the problem, it will require a number of complementary technologies to combat, including a

technology that allows rapid and robust determination of IRs in raw materials and products made from natural compounds. Organic compounds are ubiquitous in essentially all products that could be counterfeited. Additionally, this technology could be extended to analyze organic pollutants and biological samples collected from crime scenes for forensic analysis and other applications. If the technology developed can determine the source of a chemical compound, it will give organizations such as law enforcement the ability to investigate and arrest criminal entities involved in manufacturing and selling of illegal medicine, food and other consumer products.

### 1.3 Effects of counterfeit medicines

#### 1.3.1 Public health

Countries with a weak regulatory system for imported and locally produced medicine often experience public health crises because of SSFFC products. In such countries, a safe and reliable quality medicine supply for health care is almost non-existent.[36] Thus, these countries experience a high level of untreated disease, deaths and treatment failures based on counterfeit medicines.[37]

A significant number of fatalities have been recorded over the years as a result of SSFFC medicine. For example, in 2008 and 2009, 84 Nigerian children died from acute kidney failure due to presence of diethylene glycol (an active ingredient in antifreeze) in a counterfeited teething syrup.[38-39] The same diethylene glycol was sold to a European company as pharmaceutical-grade glycerin by a Chinese chemical manufacturer. It killed 219 people from acute kidney failure in Panama in 2007.[40] In the Panama case, the confirmed deaths could be a mere fraction of total mortality given that 60,000 bottles of

cough syrup were contaminated, as reported by the Panama Ministry of Health and WHO.[41]

Another example of the effect of falsified medicine is the rise of untreated diseases that progress over time, eventually leading to deaths. There is evidence of such cases documented from parts of the world where it is too costly to have the technology to verify and authenticate or evaluate the quality of medicine. For example, in Ghana researchers found that 89% of the uterotonic drugs oxytocin and ergometrine did not meet the required standards by British Pharmacopoeia specification, and 2% of the medicines were expired. In this case, only 9% of the drugs were found effective in curing the intended diseases.[42] Hemorrhage, a treatable condition using uterotonic medicine, is the main cause of a high maternal mortality ratio in Ghana estimated to be 350 per 100000 live births.[43-44] That means hemorrhage accounts for an estimated 17 to 22% of maternal deaths in Ghana. [45]

Treatment of chronic diseases such as diabetes has increasingly become a serious challenge to health institutions because cheap counterfeit drugs have flooded the health industry affecting health quality. For example, in China, a fake version of the diabetes drug, glyburide, was found to have six times the pharmacopeial standard dose and resulted in the death of two patients.[46] Fifty percent of popular brands of the oral diabetes drug metformin failed several pharmacopeial tests of bioequivalence, thus directly or indirectly affecting 80% of 347 million diabetic patients who live in developing countries.[47-48]

Antibiotics, which are the most important medication for infectious diseases globally, are also targeted by counterfeiters. Substandard antibiotics account for about

45% of the 771 counterfeit cases documented and reported by WHO.[49] Therefore, a large number of antibiotics, such as penicillin and amoxicillin commonly used for treating bacterial infections, including scarlet fever, respiratory and urinary tract infections, and pneumonia, are counterfeit.[37] It is worth noting that bacterial Pneumonia, which is caused by the pathogen *Streptococcus pneumoniae*, accounts for 12% child deaths globally.[50]

Lastly, counterfeit drugs can also lead to catastrophic treatment failures, such as bacteria, and virus drug resistance, based on incorrect API concentration. Bacteria resistance is problematic to health research plans, health systems, health care providers and people seeking medical care. Such resistance can jeopardize decades of research and development of medicine. There are already such cases reported by healthcare providers and health institutions. For example, many tuberculosis drugs have been reported to be ineffective for treatment because *Mycobacterium tuberculosis*, the bacterium causing tuberculosis, has gradually become drug resistant.[51] This bacterium accounts for nearly 6% of global drug-resistant infections.[52] Often, these infections are complex and challenging to treat, and are responsible for multiple deaths around the globe. [53-59]

### 1.3.2 Economic and social impact of counterfeit drugs

The cost of manufacturing drugs, from research to production, is currently estimated to be more than $1.3 billion.[60] Figure 1.2 shows the estimated increase in drug development cost since 1975. Since most drug research is tax funded by citizens, the cost is often transferred to consumers. For many families in middle-income and poor economic status, increasing drug prices create financial difficulties. In many cases, prescription costs are second after food and house expenses in many countries.[61] When

bacteria and viruses develop drug resistance, it renders drugs ineffective that cost billions of dollars to produce. This drives the cost of developing drugs higher.[62-64]

The high cost of developing medicine creates a situation where these treatments become too expensive to afford, and also drives health insurance costs higher.[62] This can lead to corruption in health institutions, in which only those who can afford it have access to the best medicines.[61] Organized crime and terrorist activities can also be funded by counterfeit medicines.[65]. All these cases are not good for the economy and have a negative impact on the social welfare of communities.

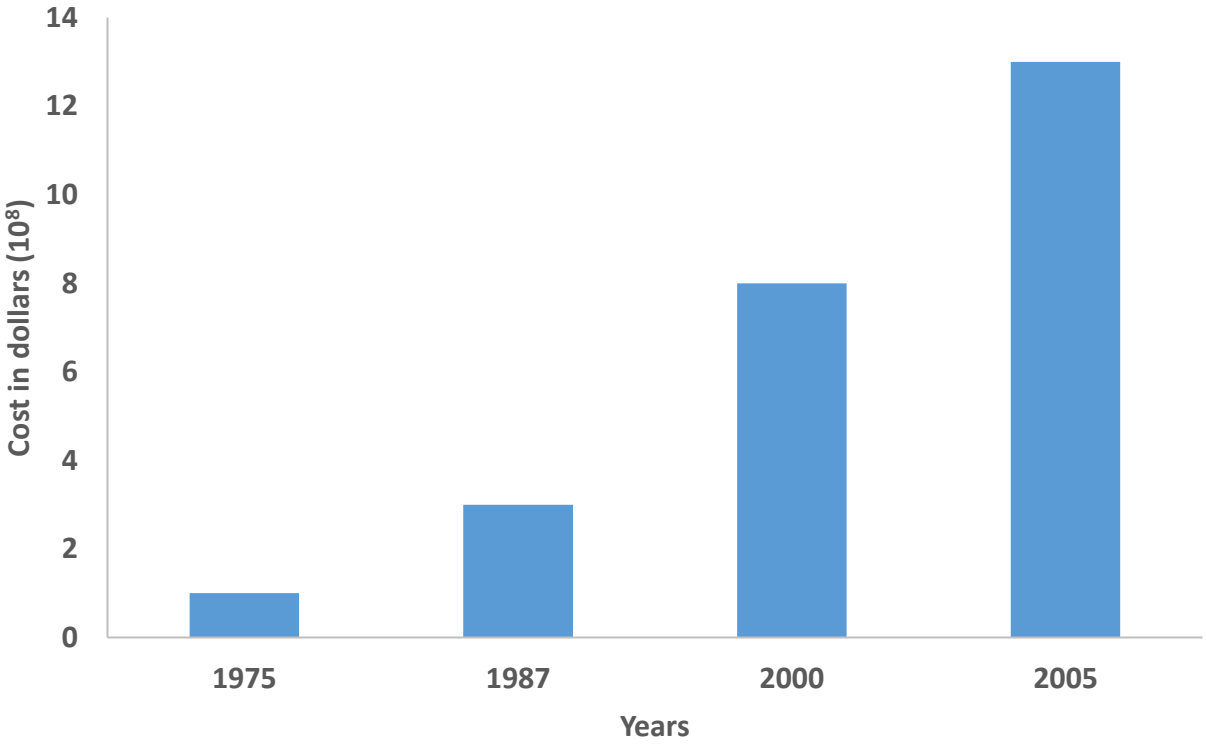



**Figure 1.2**. The increase in the cost of drug development since 1975.

Overall, counterfeit pharmaceuticals can lead to mass poisoning, disease progression, drug resistance, and increase in deaths resulting from chronic diseases, high

drug costs, loss of confidence in health systems, health care providers, medicine and national and international regulatory agencies, and funding of criminal and corrupt enterprises.

**1.4 Overview of Techniques for Detecting SSFFC**

The impact of SSFFC on health care, social life and economy of many countries around the world is severe. Thus, there is a critical need to develop technologies to examine and analyze counterfeit products. Visual inspection can sometimes identify counterfeit goods, such as currencies, relevant documents, antiques, and medicines.[65] The conventional optical technologies available for detecting SSFFC from the WHO Checklist,[66] are multidimensional atomic force microscopy,[67] GPHF-MiniLab[68-69] and Counter Detection Device (CD-3).[70-71] The WHO checklist is freely available for healthcare workers including pharmacy and laboratory technicians with minimum training requirements.[66] Atomic microscopy authenticates products by reading the molecular watermarks imprinted on products. It is expensive, requires a climate-controlled environment and a well-trained scientist to operate and maintain it. [67] Visual inspection in many cases is a first step in authenticating materials, and other additional advanced technologies are required to identify SSFFC because many counterfeiters have already developed technology that can evade hi-tech visual inspections.[64] The advanced technologies needed in addition to visual inspection include portable technologies and analytical laboratory techniques.

Large instruments cannot be utilized in certain areas, such as remote locations, ports, and pharmacies. Thus, mobile devices for detecting counterfeit products are employed. They include Counterfeit Detection Device (CD-3),[70-71] Nuclear Quadrupole

Resonance (NQR),[72-73] Raman,[74-75] near-infrared (NIR) light spectroscopy,[76-77] chromatography techniques,[78] Fast Chemical Identification System (FCIS),[79] and GPHF-MiniLab.[68-69]

These techniques have several advantages because they are cheap technologies (e.g., chromatography techniques that include TLC Speedy Apparatus, and CD-3) except Raman and NIR portable devices. In all cases, they do not require sample preparation, they are portable, and relatively fast in analyzing samples. Additionally, they do not need extra supplies and call for minimal training time. However, CD-3, Raman, and NIR portable devices have a limitation because they depend on the use of reference libraries of pharmaceuticals to identify counterfeit products.[80] Dependence on reference libraries causes problems especially if malicious individuals have access to the database. [80]They are also incapable of analyzing complex matrices that are uniquely identical such as enantiomer mixtures, or the same compounds produced from two different sources. Thus, complex forms and analysis of the same compounds from various sources require advanced laboratory-based analytical techniques for IR analysis discussed in Chapter 2.

# CHAPTER 2: Currently available IR analysis methods

## 2.1 Analytical laboratory technology for IRs analysis

There are several techniques available for evaluating IRs. Isotope Ratio Mass Spectrometry (IRMS),[11] Nuclear Magnetic Resonance (NMR) Spectroscopy,[81]Infrared Laser Absorption Spectroscopy (ILAS) and Cavity-Ring Down Spectroscopy (CRDS). [82-83] CRDS, IRMS and NMR are the most commonly used for counterfeit detection.

### *2.1.1 Infrared Laser Absorption Spectroscopy (ILAS)*

ILAS was developed as an alternative to the established IRMS for the measurement of IRs of atmospheric $CO_2$. It uses the mid-IR spectral region that allows measurements of IRs with high-sensitivity due to rotational-vibrational bands of strong and important vibrations that affect $CO_2$ in the region. It also uses fast scan rate to acquire spectra that increases sensitivity; for example, 1400 spectra / s at 1.4kHz for IRs calculations.[84-90]

It has an advantage over IRMS because mass isotopologues are eliminated based on distinctive absorptions by of trace gas molecules. ILAS is also cheap, consumes less power and benchtop space, and does not require high purifications and highly purified gases. However, it is not efficient for batch analysis in a laboratory set up except for medical detection and diagnosis of infections such as *H. pylori* through human breath. But it requires large amounts of $CaCO_3$ for that type of analysis which is unfeasible in most earth science research.[87, 90-91]

ILAS was recently improved using a more sensitive tunable infrared laser direct absorption spectroscopy (TILDAS) shown in Figure 2.1. It uses a quantum cascade laser (QCL) at 2310 cm-1 to measure absorption spectrum of $CO_2$ isotopologues (i.e., $^{12}C^{16}O_2$,

$^{12}C^{18}O^{16}O$, $^{12}C^{17}O^{16}O$ and $^{13}C^{16}O_2$) shown in Figure 2.2. The miniaturized absorption cell is vacuum tight, and allows $CO_2$ to flow into a pre-evacuated cell and then be separated at a low pressure and detected by a thermoelectrically cooled photovoltaic detector. For each isotopologue the output is calculated using nonlinear least-squares fitting of the absorption spectrum and produces a peak area that is used to find a ratio with other peaks. The calibration is done against $CO_2$ derived from standard carbonates with known stable IRs.[92]

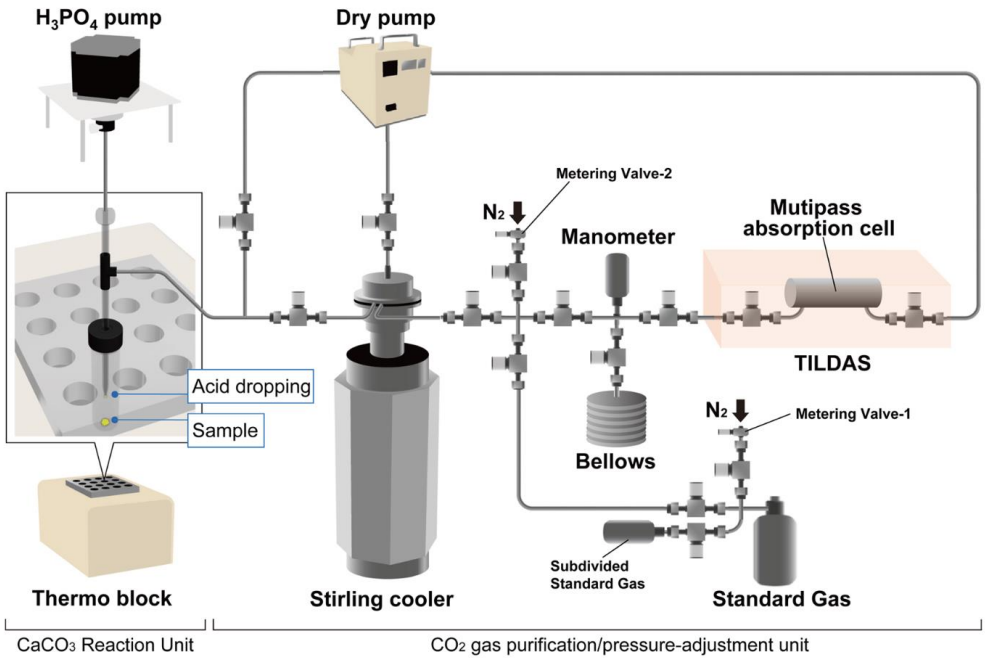



**Figure 2.1**. A schematic diagram of isotope ratio measurement system. It consists of a $CaCO_3$ reaction unit, $CO_2$ purification unit, and tunable infrared laser absorption spectroscopy (TILDAS) system. TILDA, shown highlighted in orange has multipass cell and two valves.[92]

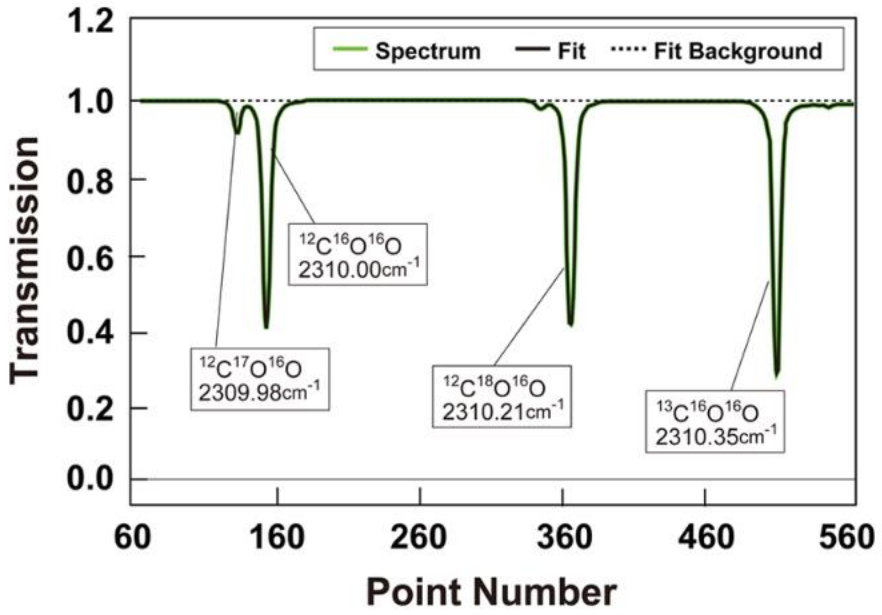
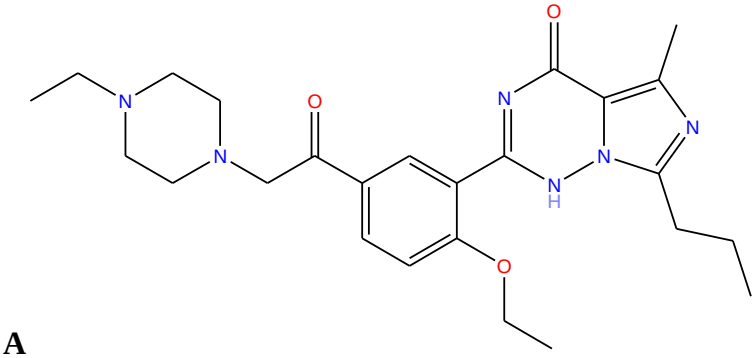



**Figure 2.2.** $CO_2$ absorption spectrum acquired by TILDAS technique with a 36 m absorption path at 10 Torr (1333 Pa) total pressure. The measured spectrum is marked green with a fitting marked as black. The black dotted line represents the fit background. The horizontal axis shows the quantum cascade laser (QCL) that scans the laser frequency range by means of 800 channels at 1.4kHz.[92]

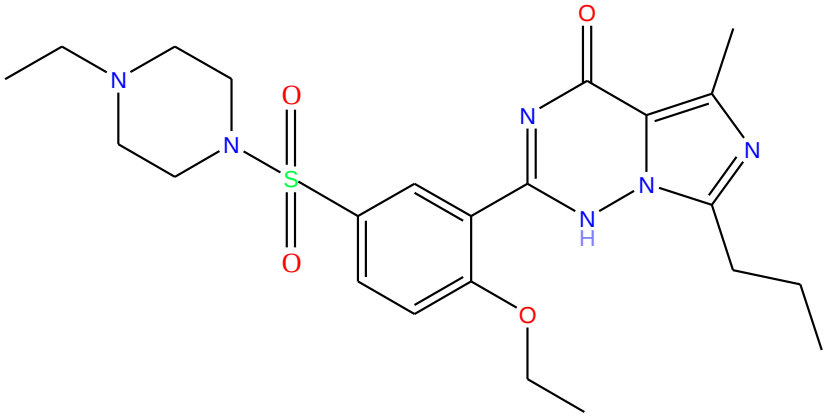
### *2.1.2 Nuclear Magnetic Resonance (NMR) Spectroscopy*

NMR methods have several advantages that are fundamental to their applications. They are non-destructive.[93] They can be faster than chromatographic techniques.[94] SS-NMR spectroscopy is capable of analyzing both active pharmaceutical ingredients (API) [95-96] and the excipients,[97-99] and is well-suited for the study of polymorphism[100] because it can differentiate conformational polymorphism from stereochemical differences.[96] When compared to powder X-ray diffraction (P-XRD), NMR is capable of analyzing chemical compounds in the amorphous state[101-102] of different patterns of generic drugs in the spectrum[81 103] thus widely applied in the analysis of counterfeit e.g., analysis of acetyl vardenafil (i.e., API in viagra) and vardenafil shown in Figure 2.3. But NMR does not analyze IRs and requires a large concentration of API, because it produces signals with

low intensity that might be suppressed or overlapped by signals from excipients[96, 104 93] Thus, it is hard to use NMR in cases where impure or limited material is available.




**A**




**B**

**Figure 2.3**. The structure of acetyl vardenafil (A) and vardenafil (B).

Unlike general NMR used for structural analysis of compounds, Site-specific Natural Isotope Fractionation NMR (SNIF-NMR) is used to determine hydrogen ($^{2}$H SNIF-NMR) and carbon ($^{13}$C SNIF-NMR) isotope ratios of organic compounds with excellent accuracy. SNIF-NMR has the major advantage of measuring the isotope ratios at individual locations in a molecule without converting compounds to gaseous products through combustion.[81] For example, SNIF $^{13}$C NMR employs the method of 180°

quantitative adiabatic composite refocusing $^{13}C$ NMR pulse sequence for site-specific isotopic measurements by determining $^{13}C/^{12}C$ site-specific isotope ratios.[105-107]

SNIF-NMR is built on the principle of natural isotopic fractionation and is regularly used to authenticate food. For SNIF-NMR function the molecules used must be purified or pure. The method involves, for example, the fermentation of juice, quantitative extraction of alcohol (e.g., ethanol) by the process of distillation, preparation of samples for NRM analysis, NRM acquisition and interpretation of results.[108-109]

For the interpretation of the site-specific isotope ratio, $R_i$, of a given molecular site, $i$, is defined as $(D/H)_i = D_i/H_i$ where $D_i$ is the number of deuterium atoms at site $i$ and represents the number of isotopomers mono-deuterated at site $i$. $H_i$, is the total number of hydrogen atoms of type $i$, which is equivalent to $P_i N_H$ where $P_i$ is the number of equal positions at site $i$ and $N_H$ the number of fully protonated molecules.[110]

Another important parameter that is used to quantify experimental data is $R_{i/j}$also defined as $R_{i/j} = F_j S_i/S_j$ where S is the intensity of signal $i$ or $j$ (i.e., the area of the NMR signal). $R_{i/j}$ represents the number of heavy isotopes at site $i$ in circumstances where site $j$is randomly given its known stoichiometric number of atoms, $F_j$. Using this information, SNIF can access directly and simultaneously the hydrogen isotope ratios, (D/H) relative to different positions, $i$, in a given molecule.[110]

Overall, NMR methods developed to analyze $^{13}C/^{12}C$ and $^{2}H/^{1}H$ isotope ratios for specific atomic sites in organic compounds and to study isotope ratio patterns of complex organic molecules have numerous applications.[106, 111-114] They are used to analyze and authenticate food such as wine, sugar, fruit juices and olive oil.[115] It has also been employed in determining the geographic origin of heroin and cocaine in illegal drug trade

investigations.[116] However, technical limitations of NMR include relatively low precision (~1 per mil), large samples size requirement (mg or greater), long integration times, and low sensitivity.[114]

### *2.1.3 Cavity Ring-Down Spectroscopy (CRDS)*

CRDS is an efficient technique for determining $^{13}C/^{12}C$[117] ,$^{16}O/^{17}O$ and $^{2}H/^{1}H$[118] IRs based on the relative absorption lines of carbon dioxide, water vapor[118-119] and methane.[118] It operates by tuning a laser source to the spectrophotometric absorption lines of interest and measuring the decay rate of the light intensity, [120] which allows determination of the relative abundance of isotopes of interest.

CRDS has the advantages of accurate quantitative measurements, speed, and simplicity. However, it has several limitations, since it suffers from low selectivity and low sensitivity. Combustion of compounds is necessary for CRDS, leading to loss of structural information which allows for determination of the aggregate isotopes for one element at a time. It is prone to inconsistencies because of the delicate mirrored surfaces necessary to perform the technique. CRDS is commonly used to determine IRs of only carbon, hydrogen, and oxygen. [120] Despite the fact that nitrogen and sulfur isotopes have been analyzed via optical techniques they are rarely analyzed using CRDS for most organic compounds. [118-119]

### *2.1.4 Isotope Ratios Mass Spectrometry (IRMS)*

#### *2.1.4.1 Background*

Forensic, geological and environmental scientists widely apply IRMS for tracing origin/history of materials. Many IRMS applications have emerged in the past four decades, notably study of the stable isotopes at the natural abundance. The concept of

IRMS can be traced back to 1918 by A. J. Dempster.[121-122] Alfred Nier built the original version of the IRMS in 1936 that featured a 180° magnetic sector field analyzer. It was used to determine significant isotopic data[122-123] in wood, oil and other biogenic materials for $^{13}C$ depletion. Later in 1947, Nier introduced a modern day IRMS that featured multiple collectors and a magnetic sector mass filter.[122-124] IRMS technique became a standard instrument for the geosciences and other related fields almost immediately after World War II. By 1950, IRMS had been used in studies of systematic variation of $^{13}C/^{12}C$, and fundamental depletion of $^{13}C$ in organic materials.[122, 124] Modern IRMSs (schematic diagram shown in Figure 2.4) retain most of the original features, but a novel interface, called continuous flow interface, was developed derived from the principle of viscous flow of carrier gas continuously spiking bands of sample gas into the ion source. The continuous flow method enabled online coupling of different analytical techniques with IRMS. With this coupling ability, sample signals are transient and appear as peaks. A special algorithm has since been created to calculate isotope ratios from signal/peaks accurately.[122, 125] The continuous flow interface has also eliminated analytical errors due to increased pressure in the ion source.[122, 126]

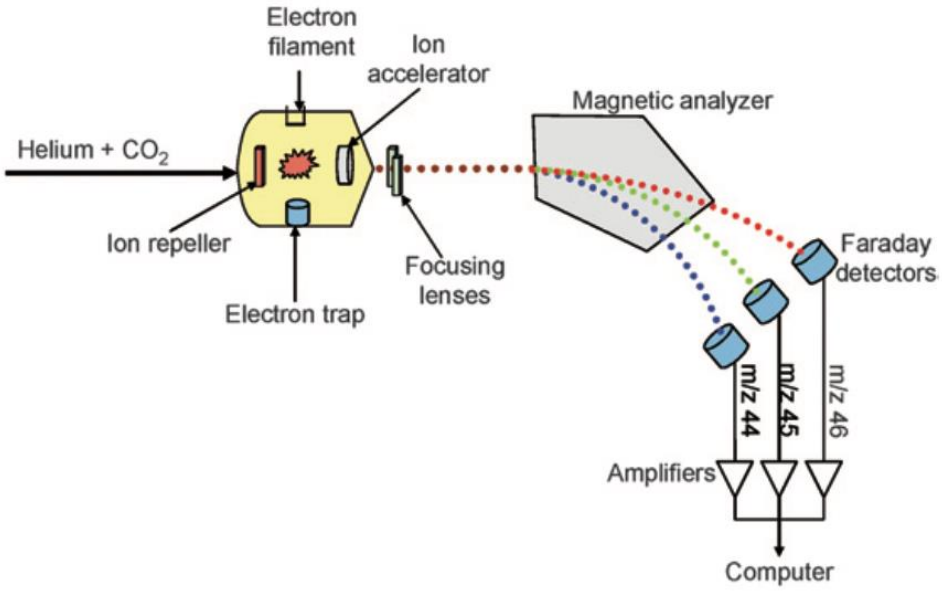



**Figure 2.4**. Schematic diagram of modern-day IRMS.[11]

The first interfaced continuous-flow isotope ratio mass spectrometry (CF-IRMS) unit was introduced by Preston and McMillan when they coupled an elemental analyzer to an IRMS via a variable leak.[127] Among other improvements was the ability of IRMS to analyze $^{15}N$ labeled compounds, introduced in 1978.[128] Around 1984, IRMS was integrated with gas chromatography (GC) via a combustion unit for GC-C-IRMS. This eventually allowed analysis of stable isotope ratios at natural abundance.[122, 129] Currently, two types of IRMS are available, the Elemental Analyzer (EA-IRMS), used for analyzing C and N, and the High-Temperature Thermal Conversion EA-IRMS (TC-EA-IRMS) used for analyzing H and O isotopes.[11]

*2.1.4.2 EA-IRMS and TC-EA-IRMS*

In EA-IRMS, shown in Figure 2.5, a sample is initially subjected to elemental analysis by being combusted at 900°C to produce $N_2$, NO, $NO_2$, $H_2O$ and $CO_2$ in the presence of an oxidant, such as copper or chromium oxide. A reduction process follows in a stream of helium before determination of the relative abundance of isotopes in the sample via IRMS. On the other hand, the analysis process in TC-EA-IRMS shown in Figure 2.6 involves the thermal conversion of sample in a reactor at elevated temperatures (usually > 1000°C) in the presence of reactive C in a continuous flow of He.[130]

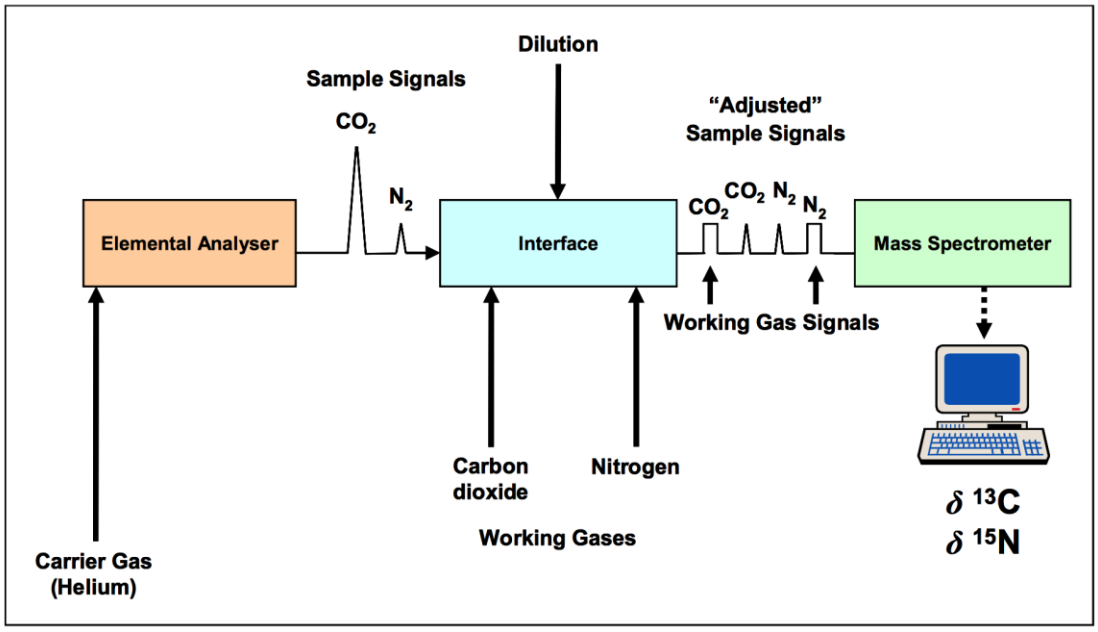



**Figure 2.5**. Schematic diagram of an EA-IRMS for the analysis of $\delta^{13}C$ and $\delta^{15}N$.[131]

In both IRMS techniques, a sample is quantitatively converted into a gas (with approximately identical molecular weight) and chromatographically separated by a GC before introduction into the ion source of the IRMS. The gaseous sample undergoes ionization by producing a highly charged gaseous species, [132] for example, $CO_2^+$, $N_2^+$, and $H_2O^+$ ions. [133-135] The ions produced travel in the form of ion beams accelerated to,

e.g., 10 keV relative to the ground passing through an electrostatic analyzer (ESA) before entering the magnetic field. The ion beam then goes through the rotation and focusing quadrupole lenses before inflowing to the magnetic sector electromagnet controlled by feedback from a field probe. After the ion beam exits the magnet, they pass through a second focusing quadrupole before reaching the detector array.[114] When the ions reach the mass analyzer, they encounter a multi-collector array containing eight pairs of detectors, each pair consisting of an electron multiplier and Faraday cup adjacent to one another. The electric

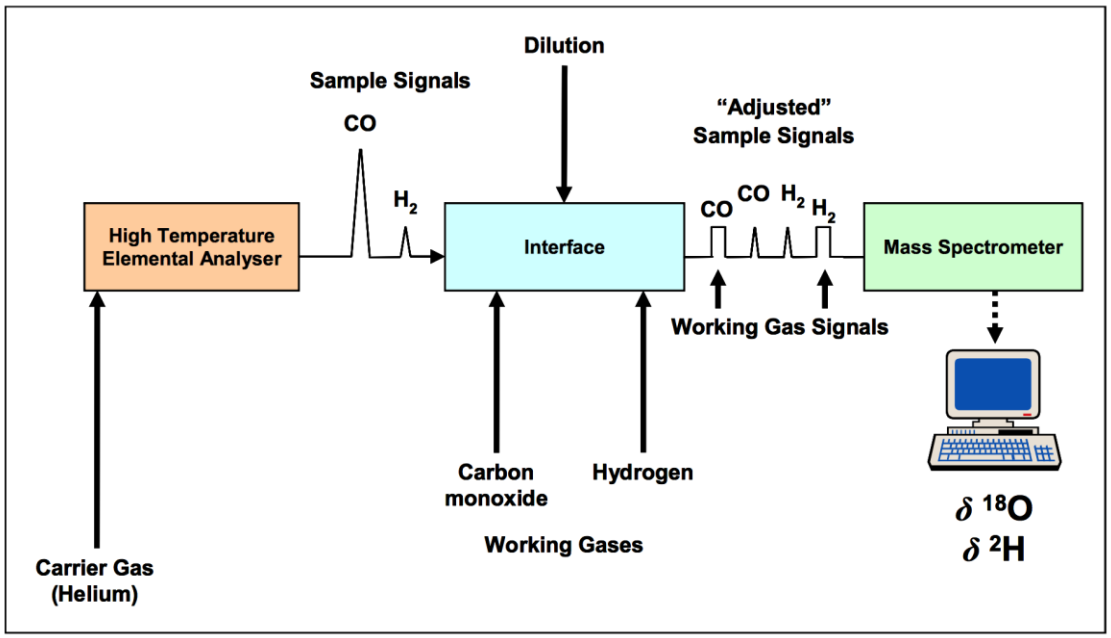



**Figure 2.6**. Schematic diagram of TC/EA-IRMS for the analysis of $\delta^2H$ and $\delta^{18}O$.[131]

currents produced by ion beam are registered in Faraday cups and recorded as signals.[114] Then, IRs are measured by monitoring ion current intensity at selected m/z ratios. For example, $^{18}O/^{16}O$ is found by monitoring m/z 46 and 44, which are the molecular masses

of $[^{12}C^{16}O^{18}O]^+$ and $[^{12}C^{16}O^{16}O]^+$ ion, respectively.[130] The IR is reported as δ (parts per mil) calculated using Equation 1.[11, 114]

$$\partial = (\frac{R_{sample}}{R_{standard}} - 1)\ x\ 1000 \quad (1)$$

where δ is the value of the ratio of sample and standard measured, $R_{sample}$ is the ratio of the isotope of samples, and $R_{standard}$ is the ratio of the isotope of the standard.[11]

*2.1.4.3 Advantages and disadvantages of IRMS*

In general, IRMS is considered the most accurate, precise and reliable method for determining IRs. It is exceptional for analysis of total IRs of specific elements in compounds. Additionally, it has several significant strengths: 1) its Nier-type electron impact (EI) ion source can generate excellent currents (nA to μA) and stable ion beams; 2) its multiplier-collector arrays allow very precise measurements of small relative differences in isotopic compositions; 3) the dual viscous-bleed inlet system used on many gas source IRMS allows fast comparison of sample and standard under identical conditions, producing excellent accuracy.[114, 136-139]

Overall, IRMS is exceptional for determining aggregate IRs of compounds. However, the current IRMSs also have limitations that affect the extent to which they measure IRs. IRMS analytes must be gases mostly of low molecular weight with a significantly high vapor pressure at standard room temperature for delivery to the EI ion source with a carrier gas. IRMSs detect ions using only Faraday cups, which are good for linearity but have significant background currents (~fA) which increases detection limits.[128, 140-141] IRMSs also have low mass resolving power (~200), significantly lower than most complex thermal, plasma and secondary ion mass spectrometers (~10,000) (measured as mass M/ΔM; ΔM is width which is equivalent to a mass unit of the ion

beam). The low resolution means that IRMSs are incapable of differentiating analyte ions from isobaric interferences, including contaminant gases. For example, $H^{16}O$, and $^{13}CH_4$ at 17 amu, or $^{12}C^{16}O_2{}^2H^+$ and $^{13}C^{16}O_2H^+$ at 45 amu, and the isotopologues of the same gas that share cardinal mass e.g., $^{13}CH_4$ and $^{12}CH_3D$; a phenomenon defined as isotopic isobars. Thus, IRMSs are mostly used to measure specific analytes that can be easily purified, with less significant isobaric interferences and ubiquitous vacuum contaminants with a simple stoichiometry that ion-correction schemes can deconvolve, at least for some of their isotopic isobars.[114] Compounds analyzed by IRMS must be highly purified, otherwise impurities from solvents, and other contaminants might affect IRs obtained. For example, during isolation of compounds from plants or drugs, achieving adequate purification (i.e., close to 100%) is very difficult. IRMS is a very sophisticated technique that is difficult to implement, especially for certain elements. For example, Br and Cl IR measurements are quite involved, since they require conversion of Br and Cl compounds to other products, such as HCl[142] which requires additional sample preparation methods, such as SPME.[143] Also, IRMS methods lead to loss of all structural information[11] such that only aggregate isotope ratios are available at the end of analysis due to combustion of compounds to produce gases, such as $CO_2$, $N_2$, $H_2O$ or $NO_2$. Lastly, IRMS is limited to analysis of one element at a time.

Collectively, the existing IRs analysis techniques and methods have played important roles towards the advancement of isotope research and studies. However, given their limitations, there is a critical need for a new method for IRs analysis which is fast, sensitive, selective, inexpensive and relatively easy to implement. It should be able to determine IRs for locations of individual atoms or groups of atoms simultaneously and

have the ability to determine IRs for all elements simultaneously. The method should have the ability to easily analyze isotopes that are difficult, such as Br and Cl.

# CHAPTER 3: Fast Isotope Ratios Mass Spectrometry (FIRMS)

## 3.1 Introduction to FIRMS

We developed FIRMS as an alternative method for IR analysis due to the limitations of IRMS. It is a novel, unique and potentially transformative technology that accurately and precisely predicts isotope ratios for elements in a compound in a fraction of the time of other methods. It calculates separate isotope ratios for individual atom locations or portions of the molecule. FIRMS has the potential to also calculate IRs that are difficult to determine (e.g., S, Cl, and Br), and to determine the IRs of all elements in a molecule simultaneously. It is the first IR analysis technique to use experimental data obtained via various tandem MS techniques to determine IRs and is self-reproducing. Unlike IRMS, which relies on high-resolution MS with a magnetic sector as the primary mass analyzer, FIRMS can be coupled with multiple types of mass spectrometers.

The tandem MS technologies amenable for FIRMS include ion-trap, quadrupole ion-trap, triple quadrupole and quadrupole time-of-flight. The experimental data collected via these techniques are analyzed using a non-linear mathematical model describing the probability of compound fragmentation to extract IRs of locations of individual atoms or group of atoms within the molecule. Mathematically, the model is of the form $c(\overrightarrow{x}) = |f(\vec{x})_{Calculated} - f(\vec{x})_{Experimental}|$: Where $\overrightarrow{c(x)}: \mathbb{R}^n \rightarrow \mathbb{R}$ are composite or multi-objective functions (MOF) mapped on a vector space (given n is an integer) with $\boldsymbol{\Omega} \subseteq \mathbb{R}^n$ as the feasible region, subject to constraints. The isotope abundance fractions are the variables, $f_{Calculated}$ is a polynomial function of $\mathbf{n^{th}}$ order mapped in a vector space, and $f_{Experimental}$ is a constant calculated from the experimental data. FIRMS's non-linear solver determines isotope ratios through an optimization process with minimization of

$\overrightarrow{\boldsymbol{c}(\boldsymbol{x})}$, a composite or sum of several single objective functions (SOF) multiplied by weighting coefficients.

### 3.2 Development of FIRMS

#### 3.2.1 Isotope Ratios (IRs)

The evaluation of IRs can help probe the history and source of molecules. For example, the fixed average global isotope abundances of tertbutyl chloride ($C_4H_9Cl$) are 98.93% $^{12}C$, and 1.07% $^{13}C$.[144] However, the isotope ratio of each element of a group of related tertbutyl chloride molecules is unique and can act like a fingerprint.[11] Moreover, the IR of each atom location can reveal further information to determine the history of a group of tertbutyl chloride molecules. The IR information can be obtained from mass spectra using various tandem MS techniques.

#### 3.2.2 Transformation of mass spectra for FIRMS

Fragmentation of a molecule during mass spectrometric analysis produces signals that correlate to the number of charged molecules of a particular mass-to-charge ratio. Thus, mass spectral data for a given fragment can be associated to isotope ratios. For example, the typical isotope ratio of $^{37}Cl/^{35}Cl$, 0.35, is commonly used to evaluate mass fragments for the presence of chlorine. Figure 3.1A shows a simplified mass spectrum of tertbutyl chloride. The inset shows a close up of the m/z ratios associated with the isotopes of the molecular ion (i.e., 77, 78 and 79 m/z). The ratio of the abundance of the 79/77 m/z ions, approximately 0.35, supports the presence of chlorine in the molecular ion. When evaluating the entire mass spectrum of a molecule, the ratio between separate fragments is not solely due to differences in isotope ratios. It is also due to a number of factors, including differences in fragmentation efficiency and sensitivity of the detector

for individual fragments. Because most of these factors cannot be accurately estimated, isotope ratios cannot be calculated by this type of mass spectrum alone (i.e., the mass spectrum produced from fragmentation and mass filtering once each, referred to as a precursor ion scan when using tandem mass spectrometry). Fortunately, tandem mass spectrometry offers the ability to fragment and mass filter molecules twice, which allows the ability to gather information as to the fragmentation efficiency and the sensitivity of fragments relative to each other under a particular set of conditions. When the fragmentation and sensitivity are consistent, the product scans of isotopes (e.g., 77, 78, and 79 m/z; Figures 3.1B-D, respectively) should only differ based on the isotopic makeup of the molecular ion or fragment. Therefore, a series of mathematical equations can be derived that relates the relative ion abundances to the isotope ratios. This is done using the product ion scan of a molecular ion or a fragment (e.g., 77 m/z in Figure 3.1B) as a "template" and relating the differences in the heavier isomer fragmentation to the isotope ratios (e.g., 78 and 79 m/z; Figure 3.1C and 3.1D, respectively).

The mathematical equations that define the FIRMS technique are derived by first transforming the product ions of the base mass spectrum into fractions of the total abundance via Equation 2,

$$\alpha_i = \frac{\gamma_i}{T_M}, \tag{2}$$

where $\alpha_i$ is the frequency of occurrence of the fragment of interest $i$, $\gamma_i$ is the abundance of fragment $i$, and $T_M$ is the sum of abundances of all fragments for the mass of interest, M. For example, when evaluating 77 m/z (Figure 3.1B) for tertbutyl chloride, $\alpha_{39}$ is calculated by Equation 3,

$$\alpha_{39} = \frac{\gamma_{39}}{T_{77}}\,. \tag{3}$$

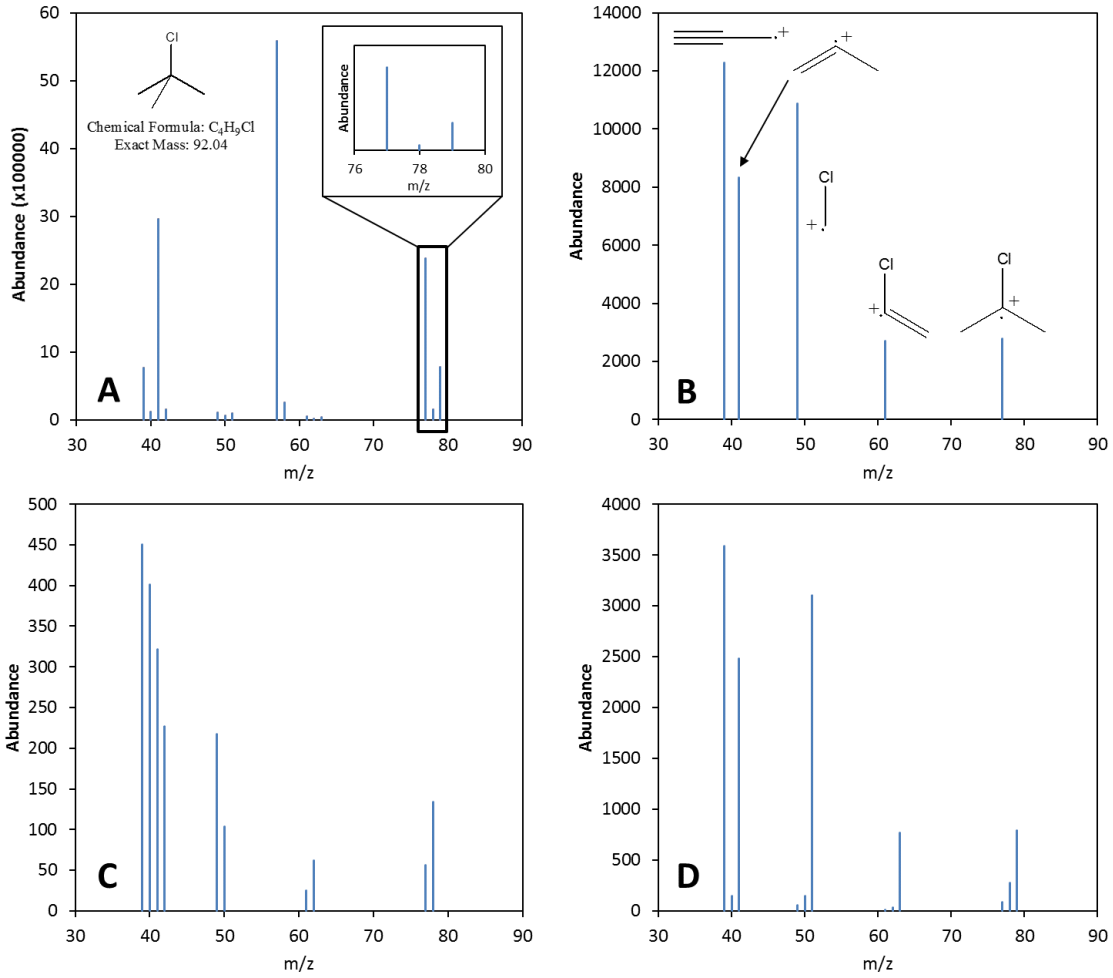



**Figure 3.1.** Simplified tandem mass spectral data for t-butyl chloride (2-chloro-2-methylpropane) obtained from a GC-MS-MS. A) Precursor ion scan (Q3-scan) of t-butyl chloride with the molecular ion region magnified in the inset. B) Product ion scan of m/z = 77. C) Product ion scan of m/z = 78. D) Product ion scan of m/z = 79. The differences between the product ion scan of the base mass (m/z = 77; B) and the heavier stable isotopes (m/z = 78 and 79; C and D, respectively) can be attributed to the isotope ratios of the atoms involved. The difference between the precursor ion scan (A) and the product ion scan of 77 (B) is due to the differences in ionization processes responsible for fragmentation.

The $\alpha$ values are the "template" to which all the heavier isotope fragmentations are related. The same transformation process is then performed on the heavier stable isotope product ion scans, with a mass of $m$ and the frequency of occurrence now represented by $\mu_i$, as shown in Equation 4.

$$\mu_{i,m} = \frac{\gamma_{i,m}}{T_m} \ . \quad (4)$$

### 3.2.3 Isotope ratio calculation

Fragmentation patterns observed in mass spectra occur when a molecule fragments into smaller pieces. The fragmentation patterns of each heavier stable isotope (i.e., the $\mu$s) can be related by the base mass (M) fragmentation pattern (i.e., the $\alpha$s) and the isotope ratios of the atoms which make up the heavier isotopes. Therefore, $\mu_{i,m}$ can also be written in terms of Equation 5, with each isotope ratio product series calculated by Equation 6.

$$\mu_{i,m} = \frac{\alpha_{i,M} \sum_{j=1}^{x} \sigma_{j,i,m}}{\sum_{k=1}^{y} \sigma_{k,m}} \quad (5)$$

$$\sigma_{j,i,m} = \prod_{e=1}^{E} \prod_{a=1}^{A} f_{e,a} \quad (6)$$

where $\sigma_{j,i,m}$ is the isotope ratio product series function corresponding to the precursor ion of interest ($m$), where $x$ is the number of possible molecular ions or fragments, with mass m, potentially contributing to the mass fragment of interest ($i$). The $\sigma_{k,m}$ is the isotope ratio product series function corresponding to all possible isotopes with mass $m$, where $y$ is the number of possible molecular ions/fragments with the m/z of the precursor ion of interest ($m$). $f_{e,a}$ is the isotope ratio of an individual atom, a, of element, e, in the fragment of interest ($i$), fragmented from the precursor ion of interest ($m$), with E as the total number of elements and A as the total number of atoms of element e associated with the precursor ion of interest ($m$).

Equations 5 and 6 are original and can be used to predict the intensities of the product ions from the stable isotopes of interest, typically M+1, M+2, etc. For example, the frequency of occurrence for fragments 39 and 40 m/z in Figure 1C can be written as

Equation 7 and 8, respectively, and the isotope ratio product series function for $^{13}C^{12}C_2{}^{1}H_6{}^{35}Cl$, which contributes to $\mu_{40,78}$, is represented by Equation 8.

$$\mu_{39,78} = \frac{\alpha_{39,77}\left[3\left(\sigma_{12_{C_3}{}^{2}H^{1}H_5{}^{35}Cl}\right)\right]}{3\left(\sigma_{13_C{}^{12}C_2{}^{1}H_6{}^{35}Cl}\right) + 6\left(\sigma_{12_{C_3}{}^{2}H^{1}H_5{}^{35}Cl}\right)} \tag{7}$$

$$\mu_{40,78} = \frac{\alpha_{39,77}\left[3\left(\sigma_{13_C{}^{12}C_2{}^{1}H_6{}^{35}Cl}\right) + 3\left(\sigma_{12_{C_3}{}^{2}H^{1}H_5{}^{35}Cl}\right)\right]}{3\left(\sigma_{13_C{}^{12}C_2{}^{1}H_6{}^{35}Cl}\right) + 6\left(\sigma_{12_{C_3}{}^{2}H^{1}H_5{}^{35}Cl}\right)} \tag{8}$$

$$\sigma_{13_C{}^{12}C_2{}^{1}H_6{}^{35}Cl,40,78} = f_{13_C} f_{12_C}^2 f_{1_H}^6 f_{35_{Cl}} \tag{9}$$

The coefficients 3 and 6 in Equations 7 and 8 represent the number of possible isotopic combinations of atoms involved which would lead to a fragment with a mass of 78 m/z and product ion of 39 m/z. Equations 7 and 8 are actually a simplification of the FIRMS model in which the isotope ratios of each element are calculated in aggregate (i.e., only one $^{13}C/^{12}C$ ratio is calculated for the entire molecule). In the full FIRMS model, the isotope ratio of each atom location is calculated individually. This is important for accuracy of the calculated isotope ratios since, in some situations, only specific atoms may contribute to certain fragments.

For very simple molecules, explicit solutions may be found if another independent equation is available. Fortunately, another equation that relates the total abundance of the base mass to the abundance of the heavier isotope fragment or molecular ion can be written as Equation 10, with $T_{c,m}$ and $T_{c,M}$ experimentally available by adding all the abundances of fragments in the mass spectrum of $m$ and $M$ m/z, respectively.

$$R_{m/M} = \frac{T_{c,m}}{T_{c,M}} = \frac{\sum_{k=1}^{y} \sigma_{k,m}}{\sum_{l=1}^{z} \sigma_{l,M}} \tag{10}$$

where $R_{m/M}$ is the ratio of the sum of $\sigma$s that make up the masses $m$ and $M$, and $z$ is the number of possible molecular ions or fragments with the m/z of the base mass of interest ($M$).

### 3.2.4 FIRMS IRs theorem

It is important that the general proposition of IR calculations by FIRMS be reasonably proved with a logical demonstration that they will predict correct IRs. If the predicted IRs model is accurate, FIRMS functions will accurately calculate isotope abundances with very good accuracy. The proof of FIRMS IRs theorem must follow the logical true statement that says:

*For the isotope ratio calculations, the sum of FIRMS μs must be one, because the sum of experimental μ(s) is one.*

To test this statement, let us consider Equation 4, which if true, mass the sum of all experimental μs must be one. Thus, the sum of calculated μs from the FIRMS model in Equation 4 must also be one. Consider three fragments 39, 41 and 61 m/z for tert-butyl chloride in Figure 3.1B as the only fragments used to calculate μs and IRs. The paired data (i.e., the product ions m and m+1 m/z) for calculating μs will be 39 and 40 m/z, 41 and 42 m/z, and 61 and 62 m/z (Figure 3.1C). The logical proof is evaluated using 5 steps below as in Table 3.1.

**Table 3.1.** Calculation of μ(s) using Equations 5 for tertbutyl chloride

| μ(s) | Equation |
|---|---|
| $\mu_{39,78}$ | $\frac{\alpha_{39,77}\left[3\left(\sigma_{12_{C_3}2_H1_{H_5}35_{Cl}}\right)\right]}{3\left(\sigma_{13_C12_{C_2}1_{H_6}35_{Cl}}\right)+6\left(\sigma_{12_{C_3}2_H1_{H_5}35_{Cl}}\right)}$ |
| $\mu_{40,78}$ | $\frac{\alpha_{39,77}\left[3\left(\sigma_{13_C12_{C_2}1_{H_6}35_{Cl}}\right)+3\left(\sigma_{12_{C_3}2_H1_{H_5}35_{Cl}}\right)\right]}{3\left(\sigma_{13_C12_{C_2}1_{H_6}35_{Cl}}\right)+6\left(\sigma_{12_{C_3}2_H1_{H_5}35_{Cl}}\right)}$ |
| $\mu_{41,78}$ | $\frac{\alpha_{41,77}\left(\sigma_{12_{C_3}2_H1_{H_5}35_{Cl}}\right)}{3\left(\sigma_{13_C12_{C_2}1_{H_6}35_{Cl}}\right)+6\left(\sigma_{12_{C_3}2_H1_{H_5}35_{Cl}}\right)}$ |
| $\mu_{42,78}$ | $\frac{\alpha_{41,77}\left[3\left(\sigma_{13_C12_{C_2}1_{H_6}35_{Cl}}\right)+5\left(\sigma_{12_{C_3}2_H1_{H_5}35_{Cl}}\right)\right]}{3\left(\sigma_{13_C12_{C_2}1_{H_6}35_{Cl}}\right)+6\left(\sigma_{12_{C_3}2_H1_{H_5}35_{Cl}}\right)}$ |
| $\mu_{61,78}$ | $\frac{\alpha_{61,77}\left[\left(\sigma_{13_C12_{C_2}1_{H_6}35_{Cl}}\right)+3\left(\sigma_{12_{C_3}2_H1_{H_5}35_{Cl}}\right)\right]}{3\left(\sigma_{13_C12_{C_2}1_{H_6}35_{Cl}}\right)+6\left(\sigma_{12_{C_3}2_H1_{H_5}35_{Cl}}\right)}$ |
| $\mu_{62,78}$ | $\frac{\alpha_{61,77}\left[2\left(\sigma_{13_C12_{C_2}1_{H_6}35_{Cl}}\right)+3\left(\sigma_{12_{C_3}2_H1_{H_5}35_{Cl}}\right)\right]}{3\left(\sigma_{13_C12_{C_2}1_{H_6}35_{Cl}}\right)+6\left(\sigma_{12_{C_3}2_H1_{H_5}35_{Cl}}\right)}$ |

Step 1: Calculate μs.

Step 2: Determine sum of μs given by $\mu_{39,78}+\mu_{40,78}+\mu_{41,78}+\mu_{42,78}+\mu_{61,78}+\mu_{62,78}$ and also calculated as follows:

$$\frac{\alpha_{39,77}\left[3\left(\sigma_{12_{C_3}2_H1_{H_5}35_{Cl}}\right)\right]}{3\left(\sigma_{13_C12_{C_2}1_{H_6}35_{Cl}}\right)+6\left(\sigma_{12_{C_3}2_H1_{H_5}35_{Cl}}\right)}+\frac{\alpha_{39,77}\left[3\left(\sigma_{13_C12_{C_2}1_{H_6}35_{Cl}}\right)+3\left(\sigma_{12_{C_3}2_H1_{H_5}35_{Cl}}\right)\right]}{3\left(\sigma_{13_C12_{C_2}1_{H_6}35_{Cl}}\right)+6\left(\sigma_{12_{C_3}2_H1_{H_5}35_{Cl}}\right)}+\frac{\alpha_{41,77}\left(\sigma_{12_{C_3}2_H1_{H_5}35_{Cl}}\right)}{3\left(\sigma_{13_C12_{C_2}1_{H_6}35_{Cl}}\right)+6\left(\sigma_{12_{C_3}2_H1_{H_5}35_{Cl}}\right)}$$
$$+\frac{\alpha_{41,77}\left[3\left(\sigma_{13_C12_{C_2}1_{H_6}35_{Cl}}\right)+5\left(\sigma_{12_{C_3}2_H1_{H_5}35_{Cl}}\right)\right]}{3\left(\sigma_{13_C12_{C_2}1_{H_6}35_{Cl}}\right)+6\left(\sigma_{12_{C_3}2_H1_{H_5}35_{Cl}}\right)}+\frac{\alpha_{61,77}\left[\left(\sigma_{13_C12_{C_2}1_{H_6}35_{Cl}}\right)+3\left(\sigma_{12_{C_3}2_H1_{H_5}35_{Cl}}\right)\right]}{3\left(\sigma_{13_C12_{C_2}1_{H_6}35_{Cl}}\right)+6\left(\sigma_{12_{C_3}2_H1_{H_5}35_{Cl}}\right)}+\frac{\alpha_{61,77}\left[2\left(\sigma_{13_C12_{C_2}1_{H_6}35_{Cl}}\right)+3\left(\sigma_{12_{C_3}2_H1_{H_5}35_{Cl}}\right)\right]}{3\left(\sigma_{13_C12_{C_2}1_{H_6}35_{Cl}}\right)+6\left(\sigma_{12_{C_3}2_H1_{H_5}35_{Cl}}\right)}$$

Step 3: Simplify $\mu_{39,78}+\mu_{40,78}+\mu_{41,78}+\mu_{42,78}+\mu_{61,78}+\mu_{62,78}$

$$\frac{3\left(\sigma_{13_C12_{C_2}1_{H_6}35_{Cl}}\right)\{\alpha_{39,77}+\alpha_{41,77}+\alpha_{61,77}\}+6\left(\sigma_{12_{C_3}2_H1_{H_5}35_{Cl}}\right)\{\alpha_{39,77}+\alpha_{41,77}+\alpha_{61,77}\}}{3\left(\sigma_{13_C12_{C_2}1_{H_6}35_{Cl}}\right)+6\left(\sigma_{12_{C_3}2_H1_{H_5}35_{Cl}}\right)}$$

Step 4: Given that $\alpha_{39,77}+\alpha_{41,77}+\alpha_{61,77}=1$, simplify step 3 further to the following fractions.

$$\frac{3\left(\sigma_{13_C12_{C_2}1_{H_6}35_{Cl}}\right)*1+6\left(\sigma_{12_{C_3}2_H1_{H_5}35_{Cl}}\right)*1}{3\left(\sigma_{13_C12_{C_2}1_{H_6}35_{Cl}}\right)+6\left(\sigma_{12_{C_3}2_H1_{H_5}35_{Cl}}\right)}$$

$$\frac{3\left(\sigma_{13_C12_{C_2}1_{H_6}35_{Cl}}\right)+6\left(\sigma_{12_{C_3}2_H1_{H_5}35_{Cl}}\right)}{3\left(\sigma_{13_C12_{C_2}1_{H_6}35_{Cl}}\right)+6\left(\sigma_{12_{C_3}2_H1_{H_5}35_{Cl}}\right)}$$

Step 5: Simplify the numerator and denominator, which gives a solution of 1.

The calculations in Table 3.1 show that the model predicted by FIRMS isotope ratios calculations is accurate for tert-butyl chloride. For all FIRMS molecules, the theorem stands to be true as the sum of μs are found to be one for the paired isotope data for precursor ion M+1 and precursor ions M+2. However, the M+2 μs must include all the three product ions m, m+1, and m+2 m/z.

### 3.2.5 Analytical solvers

As alluded to earlier, the calculation of IRs using FIRMS method relies on the ability of nonlinear solvers to optimize solutions from composite MOFs for isotope frequencies, a process known as a multi-objective optimization (MOO). The MOF, $c(\overrightarrow{x})=|f(\vec{x})_{Calculated}-f(\vec{x})_{Experimental}|$ is the sum of single objective functions (SOF) multiplied by weighting factors. The calculated abundance functions described by

Equations 5 and 6 are $n^{th}$ order nonlinear polynomial functions that are also differentiable. The terms and variables of these equations depend on the number of atoms present in a fragment. The number of equations computed depends on the number of fragments considered for a particular FIRMS model. In all circumstances, the models created by FIRMS are complex because there are at least two atoms of the same elements, such as carbon and hydrogen that creates largely complex polynomials. Thus, solving for the isotope abundances of IRs of each atom location in molecules that have more than two atoms is analytically difficult, and can take an unpredictable amount of time without finding solutions.

However, the numerical solving approach offers relief for the analysis of the complex FIRMS models. Because FIRMS models are described by a complex $n^{th}$ order polynomial and differential functions, numerical solvers offer an incredible advantage over analytical solving because they are fast, accurate, and may be the only available method to solve the complex nonlinear problem produced by FIRMS.

In the numerical solver approach, the calculated and experimental $\mu_i s$ (Equation 5 and 6) are compared via a MOF, $c\ (\overrightarrow{x})$. The numerical solver minimizes the objective function by evaluating the fractional abundances of all atoms in a molecule that gives satisfactory solutions. The solutions are comprised of fractional abundance values of the atoms present in the compound.

### 3.2.6 Numerical Solvers

There are three categories of analytical solvers, namely simulation, optimization, and data mining that perform computations and numerical modeling. A computer model such as a simulation model is a theoretical or a simplified representation of a real-world

complex system that can be conceptualized. It is designed to show important features and characteristics of the complex system which is being studied and analyzed. It can provide meaningful understanding, predictions, and solutions to the problems investigated in the real system. A significant percentage of a model often represents the actual system.[145] Optimization involves finding the maxima or minima of mathematical functions sometimes subjected to constraints.[146] Simulation optimization automatically finds the best solution of one or several variables that can be controlled. Optimization allows researchers to make better choices when all data are available. Simulation enables researchers to understand the possible results of decisions they do not have control over, considering a range of possible outcomes that are uncontrollable in any model.[145-146] Data mining, on the other hand, is a process of obtaining implicit, formerly unknown and possibly important patterns and relationships from data.[147]

### 3.2.7 Simulation optimization solvers

For FIRMS analysis, coupling simulation and optimization solver is valuable because simulation replicates the structure of a molecule and the fragments that result from subjecting a compound through MS analysis, while optimization minimizes FIRMS MOF to determine the actual values of isotope frequencies subject to constraints. This process is known as the multi-objective optimization (MOO).

#### 3.2.7.1 Overview of the MOO formulation

Solving a MOO problem constitutes one major challenge i.e., finding a solution that is acceptable to all MOF. The MOF solution can be calculated through single objective optimization (SOO) (i.e., optimization of a composite objective function of all the MOF) or multiobjective optimization (MOO).

A general single-objective optimization problem is described as:

*minimizing (or maximizing)* $\boldsymbol{f(x)}$ *subject to* $\boldsymbol{g_i(x) \leq 0, i = \{1, \ldots, m\}}$ *and* $\boldsymbol{h_j(x) = 0, j = \{1, \ldots, p\}\ x \in \Omega}$. *A solution minimizes or maximizes the scalar* $\boldsymbol{f(x)}$ *given* $\boldsymbol{x}$ *is a n-dimensional decision variable vector* $\boldsymbol{x = (x_1, \ldots, x_n)}$ *in a space* Ω.

Notice that $g_i(x) \leq 0$ and $h_j(x) = 0$ represents constraints that must be fulfilled while performing optimization. Ω is a space that contains all possible x values that can satisfy the function $f(x)$ and constraints, where $x$ can also be a continuous or discrete vector and $f(x)$ can as well be continuous or discrete.[148]

The multiobjective optimization problem (MOO) is defined as:

"*a vector of decision variables which satisfies constraints and optimizes a vector function whose elements represents the objective functions. These functions form a mathematical description of performance criteria which are usually in conflict with each other. Hence, the term "optimize" means finding such a solution which would give the values of all the objective functions acceptable to the decision maker."*[148]

The mathematical formulation of MOO is

$$\left.\begin{array}{c} Minimize\ \boldsymbol{f_i(x)} \\ subject\ to\ \boldsymbol{x \in \Omega} = \left[\boldsymbol{x} \in \mathbb{R}^{\boldsymbol{n}}, g_j(\boldsymbol{x}) \geq 0, j = 1, 2, \ldots, m\right] \end{array}\right\} \quad (11)$$

where

$$\boldsymbol{x} = [\boldsymbol{x_1}, \boldsymbol{x_2}, \ldots, \boldsymbol{x_n}]^{\boldsymbol{T}}$$

$$\boldsymbol{f_i(x)} = [\boldsymbol{f_1(x)}, \boldsymbol{f_2(x)}, \ldots, \boldsymbol{f_k(x)}],$$

k is the number of objective functions, Ω is the decision space, $x \in \Omega$ is the decision vector, m is the number of constraints, and $\Omega \subseteq \mathbb{R}^n$ is the feasible region.[149-150]

In the SOO problem, the optimum solution minimizes the composite or SOF $f_i(x)$ subject to constraints. However, defining a minimum point for MOO at which all components of the objective function vector $\boldsymbol{f_i(x)}$ are simultaneously minimized is practically implausible because a "utopia" point that can achieve that goal is rarely possible. Thus, a solution for Equation 11 is always unclear, because there is not a single unique point that can minimize all MOF simultaneously. This unique problem, therefore, gives rise to the search for optimality. It is different from solving a scalar optimization problem and requires finding optimum solutions to the vector MOO. In reality, there are infinitely many optimal solutions. An optimal solution exists where there are no other feasible solutions that can improve the value of a SOF without worsening the values of other objective functions. Therefore, there is not one unique solution; however, there are a set of pareto solutions found through the use of Pareto Optimality Theory. The theory argues that $x^*$ is Pareto optimal if there exists no feasible vector $x$ which will reduce some criterion without increasing simultaneously at least another criterion for minimization process.[148] Thus, Pareto optimal solutions have the characteristics that are required to find solutions for the vector MOO problem.[149-152]

Pareto optimality was first proposed by Edgeworth and Pareto and is defined as follows,[153]

**(1)**. A vector $\boldsymbol{u} = (\boldsymbol{u_1}, \boldsymbol{u_2}, \ldots, \boldsymbol{u_m})$ is described as dominant over another **vector** $\boldsymbol{v} = (\boldsymbol{v_1}, \ldots, \boldsymbol{v_m})$, represented as $\boldsymbol{u} < \boldsymbol{v}$, **if** $\boldsymbol{f} \forall_{\boldsymbol{i}} \in \{\boldsymbol{1}, \ldots, \boldsymbol{m}\}, \boldsymbol{u_i} \leq \boldsymbol{v_i}$ and $\boldsymbol{u} \neq \boldsymbol{v}$.

**(2)**. Mathematically, a decision vector $\boldsymbol{x^*} \in \boldsymbol{\Omega}$ is Pareto optimal if there is no another existing $\boldsymbol{x} \in \boldsymbol{\Omega}$ such that $\boldsymbol{f_i(x)} \leq \boldsymbol{f_i(x^*)}\ \forall_{\boldsymbol{i}} = \boldsymbol{1}, \ldots, \boldsymbol{k}$ and $\boldsymbol{f_j(x)} < \boldsymbol{f_j(x^*)}$ for at least one value of $j$.

The definition can be interpreted to mean that Pareto solutions exist only if a solution point cannot be moved from one point to the other without negatively affecting other objective functions. A solution point is also a weak optimal point if it is impossible to move the point and improve all objective functions simultaneously.[150] The vector of the objective function is only Pareto optimal if the corresponding decision vector x is Pareto optimal.[149-150]

A solution for a specific MOO formulation is sufficient for Pareto optimality, because it captures all SOF; thus, the method that establishes the solution is significant. But, it also undermines the objectivity of SOO because the formulation that provides the necessary conditions for a Pareto optimal point is a solution for that specific formulation and cannot be applied on a single SOO. In some cases, the solution for a specific formulation might not be Pareto optimal, or with sufficient conditions, it will always be Pareto optimal that does not capture all of the Pareto optimal points.[150] Therefore, decisions are often made to allow one of the problem solving processes, either MOO or SOO, depending on the objective of the decision maker. At no point can the two MOO or SOO be allowed to find a solution for the same problem.

There are infinite Pareto optimal solutions for the MOO problem. Thus, user preference for specific or various objective functions is valuable in order to find a single solution. Given methods that include *a priori articulation of preferences*, users can specify preferences before running an optimization algorithm that determines a single solution suitable to the user preferences. There are other cases where users do not specify prior preferences but select a single solution from a representation of the Pareto optimal set, a process known as *a posteriori articulation of preferences*.[150] However, for FIRMS

analysis, a weighted sum method for a priori articulation preference is used. Weighted sum method (WSM) allows the MOO problem to be cast as a single-objective mathematical optimization problem.[149-150]

Using the WSM to solve the problem represented by Equation 11 involves selecting scalar weights $w_i$ and minimizing the general composite reformulated Equation 11 as the objective function represented as:

$$\mathbf{min} \sum_{i=1}^{k} \boldsymbol{w_i f_i(x)} \quad (12)$$

$$\boldsymbol{s.t} \quad \boldsymbol{x \epsilon \Omega},$$

where $\boldsymbol{w_i \geq 0, \forall_i = 1, \ldots, k}$ and $\sum_{i=1}^{k} \boldsymbol{w_i} = \mathbf{1}$ under convexity assumptions.[149] All the positive weights are assumed to minimize Equation 12 and provides the necessary conditions for Pareto optimality which means that the minimum of Equation 12 is always Pareto optimal.[154-155]

**3.2.7.2 Normalization in the WSM**

Ideally, the scalar weights of each SOF are assigned by the decision maker (DM) or user based on the deep-rooted knowledge of the problem at hand. But because different SOF can have different magnitude, the normalization of the SOF is necessary to find a Pareto optimal solution consistent with the scalar weights assigned by the FIRMS user. Thus, the scalar weights are calculated as:

$$\boldsymbol{w_i = u_i \theta_i}, \quad (13)$$

where $\boldsymbol{u_i}$ are the weights assigned by the user and $\boldsymbol{\theta_i}$are the normalization factors.[149]

There are many ways of determining normalization factors, such as:

- normalization by the degree of the objective function at the initial point $x_0$, where $\boldsymbol{\theta_i = \frac{1}{f_i(x_0)}}$;

- normalization using the minimum of the objective functions, $\boldsymbol{\theta_i} = \frac{\mathbf{1}}{\boldsymbol{f_i(x^{[i]})}}$ where $\boldsymbol{x}^{[i]}$ solves $\mathbf{min}_{\boldsymbol{x}}\{\boldsymbol{f_i(x)}\ \boldsymbol{x}: \boldsymbol{x}\ \in\ \boldsymbol{\Omega}\ \}$;
- normalization by the difference between optimal function values evaluated in the Nadir and Utopia points which gives the length of the intervals where the ideal objective function varies within the Pareto optimal set.[149]

The first two normalization schemas are ineffective and impractical because the initial point, $x_0$, does not provide accurate representation of the function behavior at optimality. $f_i(x_0)$ equal to zero cannot be used, and using optimal solutions to individual objective functions can distort the scaling because optimal values are not associated with the geometry of the Pareto set.[149]

In the third normalization scheme, it is easy to determine the ranges of the objective Pareto optimal set which provides incredible information for the solution process. The initial step is to determine $\boldsymbol{z_i^*} = \boldsymbol{f_i}\left(\boldsymbol{x}^{[i]}\right) \in\ \boldsymbol{\mathcal{R}}$ values of the ideal objective vector $\boldsymbol{z}^*\ \in\ \boldsymbol{\mathcal{R}^k}$ by minimizing the individual objective function subject to user-defined original constraints, i.e.,[149]

$$\boldsymbol{z_i^*} = \boldsymbol{f_i}(\boldsymbol{x}^{[i]}) \textbf{ where } \boldsymbol{z_i^*} = \boldsymbol{arg}\ \mathbf{min}_{\boldsymbol{x}}\{\boldsymbol{f_i}(\boldsymbol{x})\ :\boldsymbol{x}\ \in\ \boldsymbol{\Omega}\}. \tag{14}$$

Two ideal objective vectors $\boldsymbol{z^U} = \boldsymbol{z}^*$ known as the *Utopia point* which provides the lower bounds, and *Nadir point* $\boldsymbol{z^N}$ components that define the upper bounds can be defined for the Pareto optimal set using the ideal vector $\boldsymbol{z}^*\ \in\ \boldsymbol{\mathcal{R}^k}$. The Nadir point can be formulated as:

$$\boldsymbol{z_i^N} = \mathbf{max}_{\mathbf{1}\ \leq \boldsymbol{j}\ \leq \boldsymbol{k}}\left(\boldsymbol{f_i}\left(\boldsymbol{x}^{[j]}\right)\right), \forall_{\boldsymbol{i}} = \mathbf{1}, \ldots, \boldsymbol{k}.^{149} \tag{15}$$

Using the difference of optimal function for normalization the values for $\theta_i$ can be formulated as,

$$\boldsymbol{\theta_i} = \frac{\mathbf{1}}{\boldsymbol{z_i^N - z_i^U}}.^{149} \quad (16)$$

The normalization schema using Nadir and Utopia points provides the best normalization solution because the objective functions are normalized by the true intervals of their variations within the Pareto optimal set. Thus, all objective functions normalized are bounded by

$$\mathbf{0} \leq \frac{\boldsymbol{f_i(x) - z_i^U}}{\boldsymbol{z_i^N - z_i^U}} \leq \mathbf{1}, \quad (17)$$

which gives equal magnitude to each objective function.[149]

In practice, Nadir and Utopia points can be estimated without finding solutions for any optimization problem, rather by evaluating random sampling points x subjected to user-defined constraints. For FIRMS, the points were calculated as maximum and minimum of each objective function evaluated at the upper and lower boundaries defined by the user.[149]

The sub weights, $u_i$, are assigned by the user; for example the Ranking Sum (RS) method. In this method, the sub weights are the individual ranks normalized by dividing the sum of the ranks. The RS function is formulated as follows:[156]

$$\boldsymbol{u_i} = \frac{\boldsymbol{k - r_i + 1}}{\sum_{\boldsymbol{j=1}}^{\boldsymbol{k}} \boldsymbol{k - r_j + 1}} \quad (18)$$

where k is the total number of objective functions, $r_i$ is the rank of the i$^{th}$ objective function and $i^{th} = 1, 2, \ldots, n$.[156]

Equation 18 can be further be written as:

$$\mu_i = \frac{2(k+1-r_i)}{k(k+1)}. \quad (19)$$

The WSM has a major weakness for linear objective functions and linear problems that use simplex-type methods to find solutions for weighted sum functions. Because, for any linear function, the simplex method evaluates the vertex to find solutions and ignores the weights that can yield solutions in the interior of the linear face of Pareto optimal set. This problem is known to have optimality issues, and the vertex solutions on the Pareto optimal set are not practically valuable solutions. An optimal solution might also experience enormous jumps when there is a slight change in weights but will not experience the same jumps when there are broad changes in weights, a problem known as sensitivity issues. However, the probability of encountering the same problems with non-linear functions is very low. Because, for example, for polynomial objective functions, the efficient frontier of the Pareto optimal is non-linear. And one of the ways to overcome linearity of Pareto face is to convert the objective functions into polynomials; for example, converting them to quadratics by squaring the functions.[149]

Fortunately, FIRMS functions are not prone to WSM problems since they are all polynomials (nonlinear). Moreover, FIRMS functions are further classified as non-linear ($n^{th}$ order polynomials and multi-variable differential functions), non-convex and non-smooth functions. This classification was very significant in selecting the simulation optimization algorithms for MOO problems. Thus, understanding nonlinear, nonconvex and nonsmooth functions, and how to apply this knowledge to FIRMS, is fundamentally important.

### 3.2.8 N$^{th}$ order polynomial and multi-variable differential functions

Polynomials include $f(x)$ functions, such as quadratic, cubic, quartic and functions with nonnegative integer powers of x. A polynomial of a degree n is a function of the general form[157]

$$f(x) = a_n x^n + a_{n-1} x^{n-1} + \ldots + a_2 x^2 + a_1 x + a_0 \quad (20)$$

where the a(s) are real numbers also known as coefficients of the polynomial.

From the general formula, an example function of n = 4, such as modified Ferrari equation, can be formulated as, [158-159]

$$f(x) = x^4 + x^3 - 6x^2 - 6x + 36 \quad (21)$$

The function in Equation 19 is also known as a quartic polynomial, [157] and has a graph shown below.

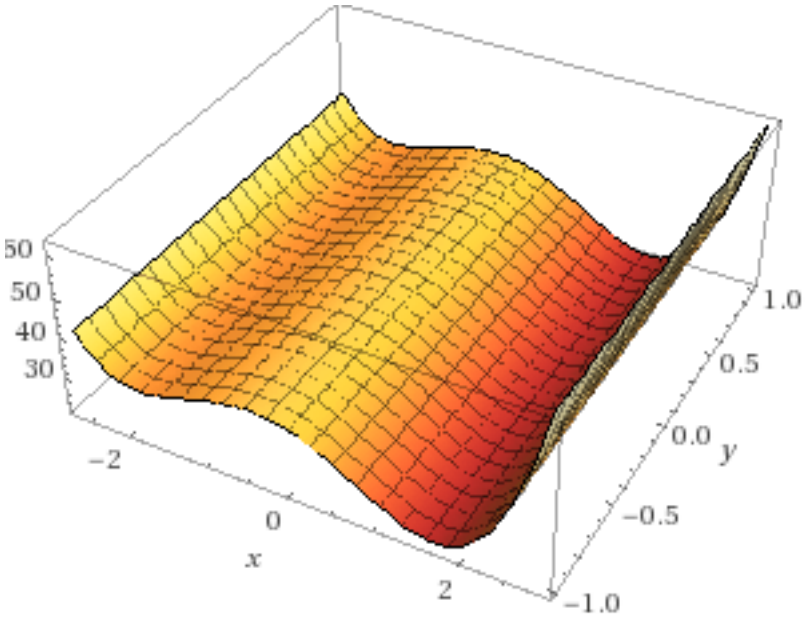



**Figure 3.2**. 3D plot of a 4$^{th}$ degree polynomial function represented by Equation 21.

A differential function contains term(s) involving derivatives of the dependent variable, y with respect to independent variable, x. The solutions for differential

equations are functions. For example, Equation 22 shows a simple differential function.[160-163]

$$\frac{dy}{dx} = 4x^3 + 3x^2 - 12x - 6 \quad (22)$$

The functions are characteristics of the relationship between a continuous variable and its rate of change. Differential equations are widely used in applications such as FIRMS for Chemistry, Physics, Biology, Economics, Engineering, and other areas of natural science[160-163]. The functions represented by the FIRMS model are both polynomials and differential because there is at least a 2nd degree term in all equations.

### 3.2.9 Isotope combinations and data structure algorithms

FIRMS Equations 5 and 6 are both $n^{th}$ degree polynomials and are differentiable. When expanded a specific example, such as tertbutyl chloride, the functions become Equations 7, 8 and 9. These equations represent complex polynomial functions with several independent variables. The degree of the polynomials depends on the number of atoms of individual elements (e.g., C, O, and H), n, where n = 1, 2, 3, 4, etc. Thus, unlike general polynomial function described in Equations 11 and 12 above, the terms and coefficients of FIRMS functions are independent of each other and only dependent on the atoms present in that term. The coefficients of each term also depend on the probability of occurrence of product ion in the molecular fragmentation pattern. For example, given Equation 8, the numerator has two terms, each of which have a coefficient of 3, meaning that the fragment obtained by changing an isotope of carbon or hydrogen for m/z 40 can only occur three times for both elements. Meanwhile, in the denominator, the coefficient of the first term is 3, while for the second term it is 6. For the precursor ion in question, there are 3 accepted molecule combinations that can be formed by changing an isotope of

carbon, while there are 6 possible molecule combinations that can be formed by replacing a hydrogen atom with a deuterium atom.

Determining the possible combinations of atoms which make up a particular fragment leading to a polynomial function, can be complicated. Using the general laws of permutations, it can be shown that the number of possible combinations increases exponentially, depending on the number of atoms in a molecule. For example, potential combinations of tertbutyl chloride molecules with the chemical formula $C_4H_9Cl$ that can occur by changing an isotope by mass of 1 is equivalent to $2^{13}$ (i.e., the number of carbon and hydrogen atoms as the power) which gives 16384 possible combinations. However, the number of utilized combinations is 3 for carbon and 6 for hydrogen. Therefore, producing $2^{13}$ combinations and filtering out unwanted combinations to form FIRMS polynomial equations is a complex task that requires specialized data structure algorithms.

There are many different types of data structures, in which data are arranged sequentially. Data structures have logical starts, and logical ends, with elements embedded in them, and are used to store and organize data. The choice of data structure depends on several factors, such as what needs to be stored, the cost of frequently performed operations, computer memory usage, and the ease of implementation.[164]

We developed two types of data structures algorithms for FIRMS, tree (for non-linear functions) and stack (for linear functions), and their performance was tested with FIRMS models. Even though FIRMS functions are non-linear, the linear stack data structure easily produced the combinations of atoms for the molecular fragments of interest. A tree is a type of data structure often used to represent hierarchal data. For

example, in a company, sits the chief executive officer (CEO) at the top, with two deputies (i.e., president and chief technology officer (CTO)) who also have deputies called managers. Below the managers are the supervisors who oversee workers. The company, therefore, can be described by a tree. A logical model of a tree can be represented such that the top hierarchy is the root, and from the root there are branches (inverted tree diagram). In computational modeling, the tree is a non-linear collection of entities called nodes linked together to simulate a hierarchy. In each node, beginning with the root, there can be different types of data (integer, character, or string) stored in the node and may contain some reference to some other nodes, also known as "children." In the tree diagram, the only direction used is from the top (root) to the bottom, or from ancestors to descendants (grandchildren). A tree can be a recursive data structure if it contains a distinguished root and subtrees in which the arrangement is such that root of the tree contains links to root of the subtrees. A tree with N nodes can have only N-1 links or edges represented by an arrow. The depth of *x* in a tree is defined as the length of the path from the root to node *x*, or the number of edges in the path from root to *x*. The height of a node entry is defined as the number of edges in the longest path from *x* to a leaf node or height of a root node. A leaf node is a node without linked nodes compared to other nodes. The height and depth of a tree are two different properties and may or may not be the same. A tree data structure may be binary, with nodes that have a maximum of two edges.[165-166] Figure 3.3 shows an example of tree (binary) data structure for $BrO_3$.

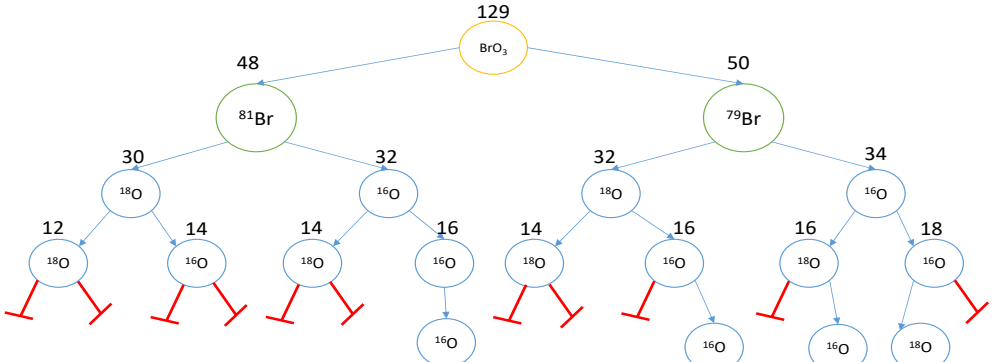


**Figure 3.3**. A tree diagram of $BrO_3$ with tree depth four.

The root of the tree ($BrO_3$ with a molecular mass (M) of 129 m/z) represents a molecular ion mass of M+2. From the root, each node is occupied by atoms in the molecule. The first level (parents) shows $^{81}Br$ picking up $^{18}O$ (left child) and $^{16}O$ (right child), or $^{79}Br$ (right branch doing the same. The second level in the tree shows the molecule picking up $^{18}O$ and $^{16}O$ for each of the first offspring (first children). The integers over the nodes represent the remaining mass after picking an atom, or a "child." For example, starting with 129 m/z at the root and following the left branch with $^{81}Br$, the mass remaining is 48 (i.e., 129-81). While on the other side, the right path gives a remainder of 50 (i.e., 129-79). These processes continue from the first to the third oxygen. Notably, in the tree diagram, some paths discontinue before making the target molecule $BrO_3$. Also, some paths in the diagram have all four atoms; however, there is still mass remaining. Therefore, the combination is also not valid. For the two scenarios, the combinations are pruned. This tree diagram has a tree depth of 4. The pruning can be carried further, as shown in Figure 3.4.

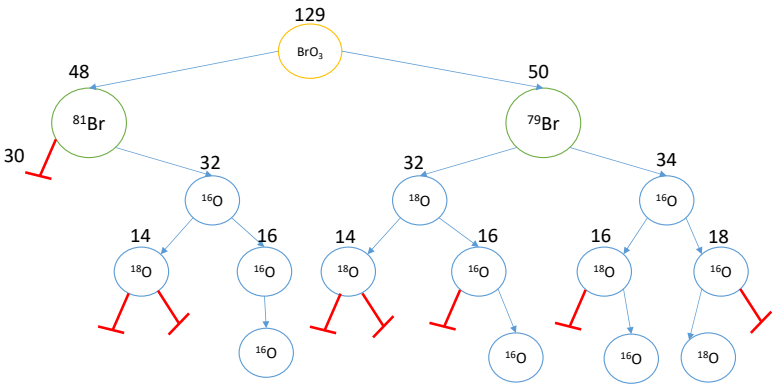



**Figure 3.4**. Represents a pruning process of a tree diagram of $BrO_3$ with tree depth four

Other examples of tree diagrams for FIRMS include dichloromethane (DCM) (Figure 3.5), and folic acid (Figure 3.6). For DCM, shown in Figure 3.5, with $CH_2Cl_2$, there are 3 elements, each with 2 isotopes giving a total of 6 atoms. This is tree depth of 6 which gives combinations of $2^6$, or 64.

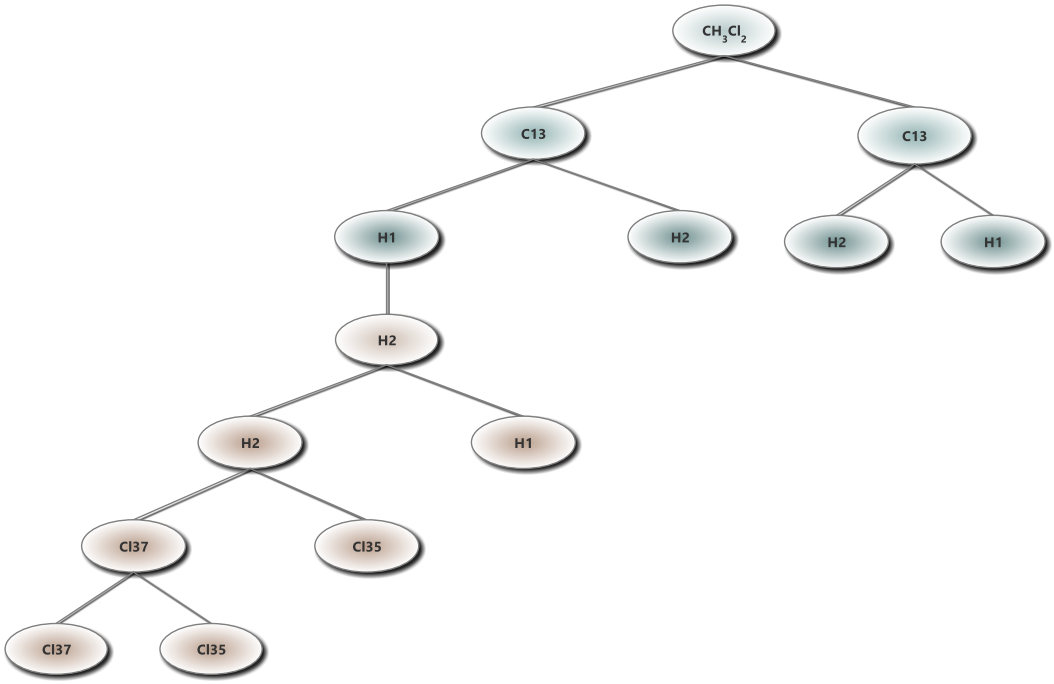



**Figure 3.5**. Represents a tree diagram of dichloromethane (DCM) with a tree depth 6.

The tree data structure for folic acid ($C_{19}H_{19}N_7O_6$) is shown in Figure 3.6. In folic acid molecule, there are 4 elements. Thus, this is a tree depth of 51 which gives molecular combinations of $2^{51}$ equivalent to $2.25 \times 10^{15}$

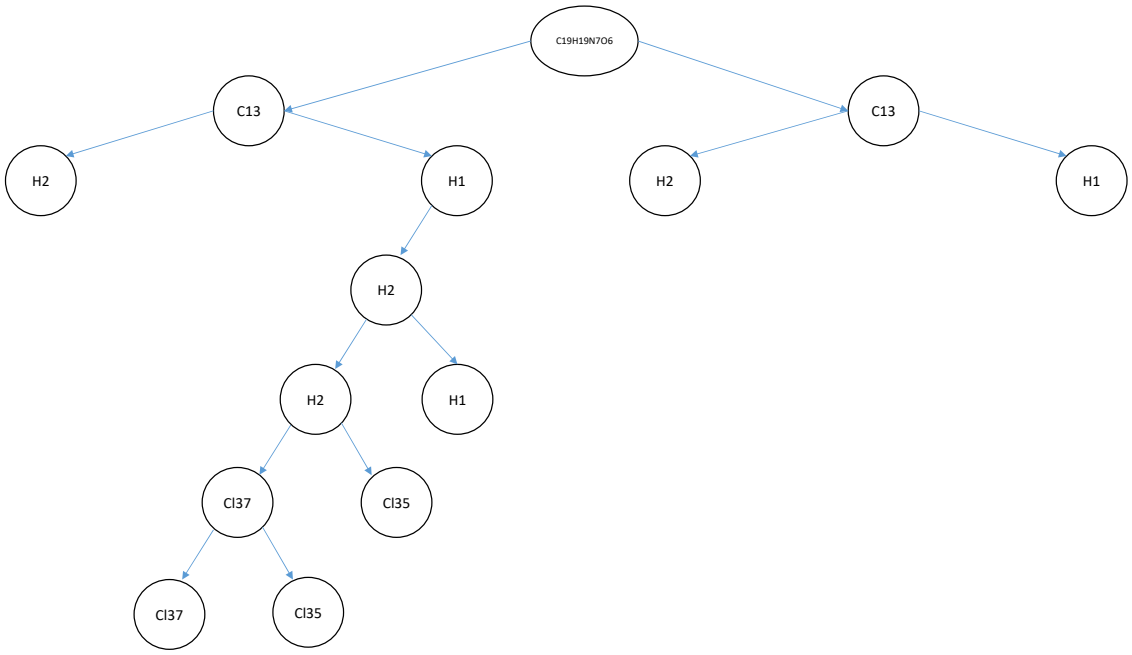



**Figure 3.6**. Represents a tree diagram of folic acid with a tree depth 51.

In the last tree diagram example, sorting the relevant molecule combinations is a tedious process, and can lead to a high cost in computation time, memory use and the difficulty of implementation. Thus, another data structure type is desired which does not have limitations of the tree-type data structure. In this case, a linear data structure, known as stack, was implemented for FIRMS models.

Stacks are an Abstract Data Type (ADT) that can be logically constructed as a linear structure that represents a real physical structure, in which insertion and deletion can be implemented, such as with a deck of cards or a pile of plates. In real applications, the stack structure allows operations to take place at one end only. For example, plates or cards can only be removed or added at the top of the pile. This idea is also known as Last-In-First-Out (LIFO). The argument is that the element added/inserted last, is

accessed first. Any process that would allow operations from the bottom is eliminated. The syntax for insertion is known as a push operation, and removing elements is known as a pop operation. Stack can be implemented using Array, Structure, Pointer, and Linked List. A stack can be a fixed size or can have the ability to resize which is an important feature for FIRMS models.

The stack data structure was more advantageous for FIRMS applications, and thus, replaced the tree data structure for determining isotope combinations that could produce a particular fragment mass. It gives FIRMS the ability to create a combination of atoms to produce a molecule and add it into a stack, test if it meets the requirements and pop out unwanted combinations. Compared to the tree method, this method consumed less time, required a relatively small amount of memory, and was easy to implement.

Within the FIRMS software, a computer algorithm was built with a function that was comprised of a string (p) variable list and integers (massDiff) to represent the atomic mass difference between the base mass M and precursor ion M+1. For example, if p = $C_3$ and massDiff = 1 the two strings of carbon atoms produced by the functions are $^{12}C^{12}C^{12}C$ and $^{13}C^{12}C^{12}C$ as shown in Figure 3.7. This idea can be expanded to a molecule like DCM ($CH_3Cl_2$) where the function is called for each element C, H, and Cl to produce a string of isotopes and the mass difference of interest.

When the individual element lists are created (i.e., C, H, H, H, Cl, Cl) by the isotope combination algorithm they are pushed on the stack creating the initial state of the first element in the list as shown in Figure 3.7. Each element list produced in the stack also has a multiplier embedded in them. The product of the multiplier for the elements in a string becomes the coefficients of sigma in Equations 5 through 9. When all the atoms

of the first elements have been added to the second element in the list the last one is added and inserted into the second stack as shown in Figure 3.8 by the arrows. The process continues until all elements have been combined and added to the stack as shown in Figure 3.9. The stack size depends on the number of molecule combinations and can be N = 1, 2, 3, … k where k is an integer. In the final step, the completed molecular formula is double checked to make sure it is accurately equal to the required mass difference, as shown in Figure 3.10. Any undesired molecular formula is eliminated from the corrected list. At this point the algorithm has enough molecular combinations with accurate isotopes. These combinations are then passed to Equations 5 and 6 which further produces Equations 7-9 for isotope abundance calculations through minimization of the objective function.

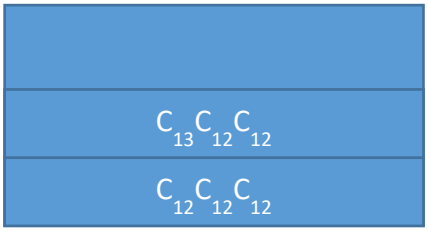



**Figure 3.7.** A list of string showing combinations of atom for carbons in stack data structure.

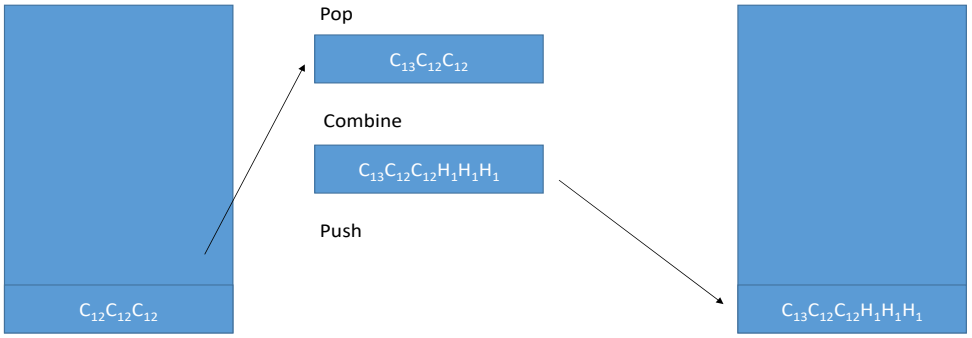



**Figure 3.8.** A string of carbon atoms combining with hydrogen atoms in stack data structures with molecular formula $C_3H_3$

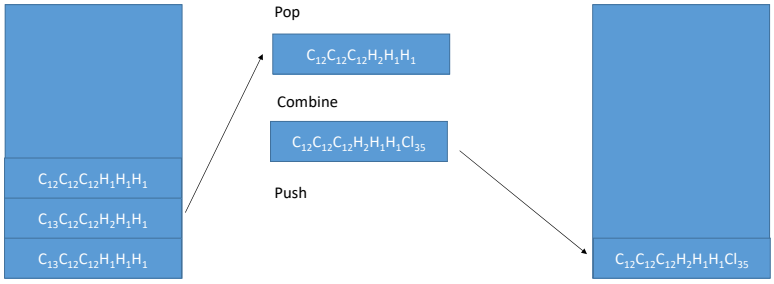



**Figure 3.9.** A string of combined carbon, hydrogen and chlorine atoms in a stack data structure with molecular formula $C_3H_3Cl$

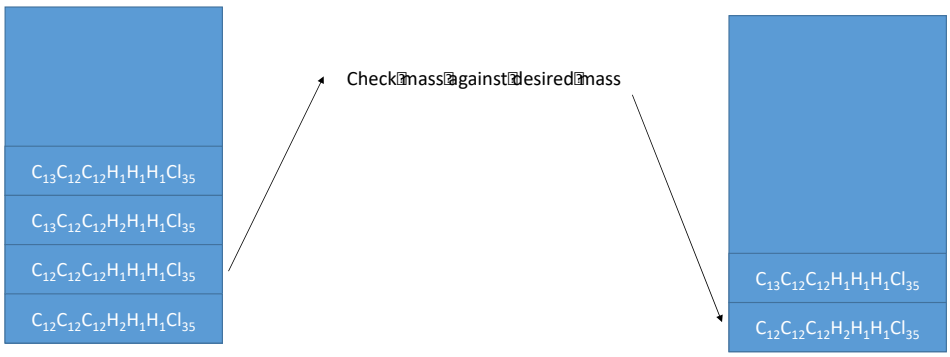


**Figure 3.10.** A stack of combined atoms in arranged on top of each other like plates stack together. The right side are the molecular formulas that have been crossed checked for accuracy

Another example of the implementation of stack data structure is shown using a $BrO_3$ molecule. A list is created {Br, O, O, O} for all the atoms in the molecule. Then an atom string is built up which comprises stable isotopes of each atom. For this example, $^{79}Br$ and $^{81}Br$ are put together on the stack as the first atoms of two possible molecules as shown in Figure 3.11. The two stable isotopes are popped off the stack and then added to the stable isotopes of ($^{16}O$, $^{17}O$, $^{18}O$), as shown in Figure 3.12. The process (i.e., brute force) continues until all the stable isotopes of oxygen are added as shown in Figure 3.13. The brute force method proceeds and eventually produces molecular formulas for all possible atom combinations. The molecular formulas are cross checked again to eliminate the unwanted combinations. For a mass difference of 2 (e.g., base mass M+2), the process can be optimized by keeping a running count of the additional mass as shown in Figure 3.14. The process continues until the full molecular formula is reached, and all unwanted combinations that exceed the mass difference are eliminated, as shown in Figure 3.15. Then the accepted isotope combinations are used to calculate fractions based on isotope frequency (probability).

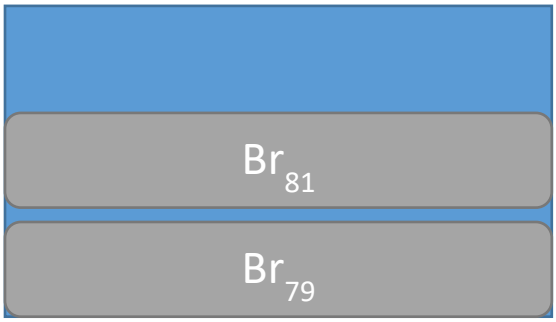



**Figure 3.11**. Using $^{81}Br$ and $^{79}Br$ atoms to describe stack structures for single atoms.

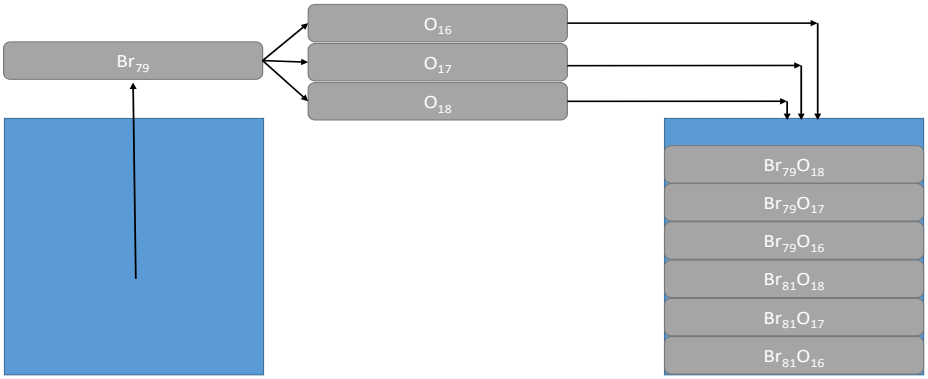



**Figure 3.12**. A stack structure of BrO molecules describing stack diagram with two atoms.

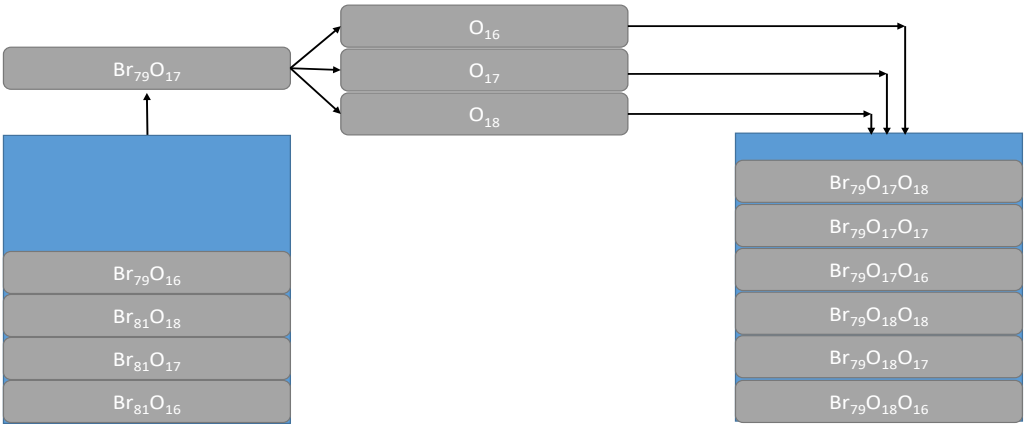
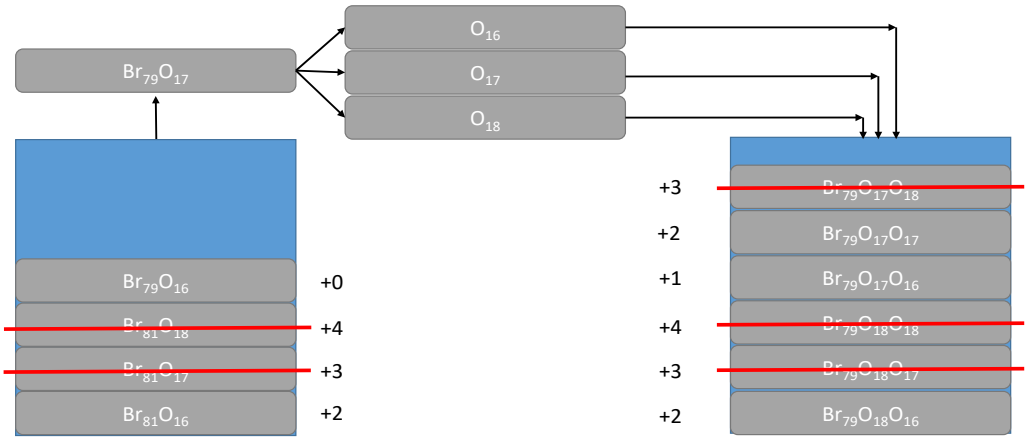



**Figure 3.13**. A stack structure of $BrO_3$ molecules describing stack diagram of four atoms.



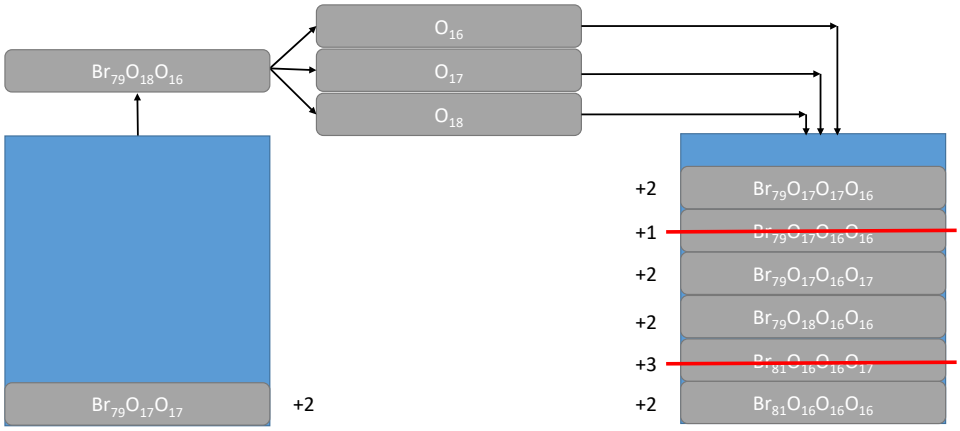



**Figure 3.15.** A stack structure of $BrO_3$ molecules describing stack diagram with four atoms and showing the elimination of unwanted structures.

Calculating fractions for isotope combinations produced by stack data structures can be described as events and involves two methods of addition and multiplication. For single events, the simple mathematical rule of multiplication is employed. However, if two events occur randomly with the same outcome but following two different paths, the probability calculations become complicated. For example, in isotope combinations, two different fragments might have the same mass, but different molecular ion formulas. The same outcome, which is the calculated mass, must be determined in two different ways to represent different paths, e.g., $^{13}C$ or $^{2}H$ isotopes that contributes the same change in mass by one unit. Both isotopes give a change of molecular mass by 1, but the two paths leading to the mass difference of 1 must be determined independently. In such circumstance, the probability tree diagram can be invoked which follows the rules of multiplication of probabilities along the branches (path) and addition of the paths to

determine the probabilities of independent events following a different path but with the same outcome. For example, the tree diagram in Figure 3.15, shows two independent events that would likely lead to the same outcome using independent paths. The likelihood that Hawi (H) will watch news (NW) when she wakes up can follow two paths, i.e., Hawi (H) wakes up, drinks coffee (C) and watches news (NW) or Hawi wakes up, does not drinks coffee (NC) and watches news (NW). The probability that Hawi will watch news can be then calculated as the sum of P (C n NW) and P (NC n NW) which results in a value of 0.5 given that there are equal chances of the first two events occurring, and where P is the probability/likelihood of the event happening. In Figure 3.16, the same probability result of 0.5 can be arrived at using the sum of probabilities P (C n NNW) and P (NC n NNW). But that cannot be allowed as a solution, since none will result in Hawi watching the news. And such solutions are thus eliminated from the overall results using computer algorithms in applications that involve complex analysis such as FIRMS analysis.

The same knowledge of probability tree diagram is applied in FIRMS to determine the probability that a particular fragment of M+1 will have a specific combination of atoms that will produce acceptable results and filter out unwanted combinations with similar solutions. After adverse combinations are eliminated, the multiplication and addition rules are applied. Since the frequency (isotope abundance) of occurrence is known, the probability of atom combinations that produces the mass M+1 can be calculated as multiples of the frequencies raised to the power of the number of atoms as shown in Equations 5-9. The probabilities of possible combinations that produce the mass M+1 are then added together (Figure 3.16).

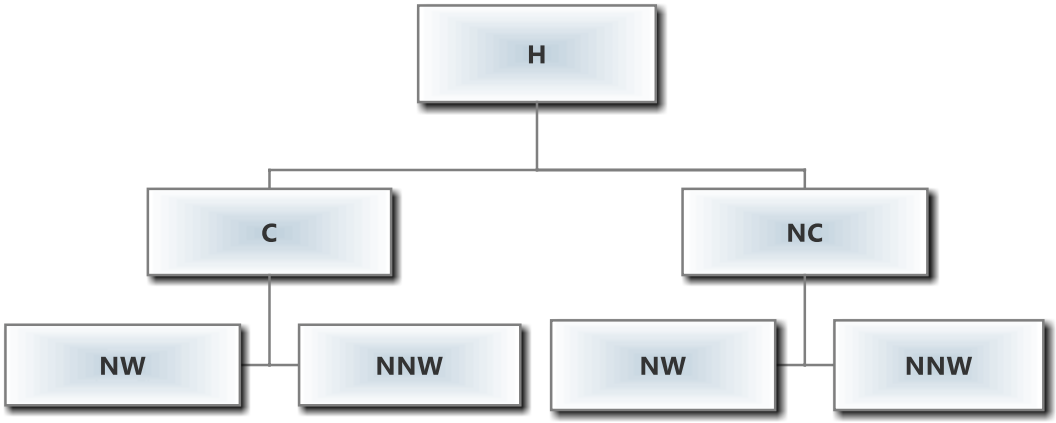



**Figure 3.16**. A flow chart of a probability tree diagram showing the path of events.

The isotope combinations produced by the computer algorithm are used to generate Equations 5-9, which make up the MOF. As described earlier, these functions have the characteristics of nonlinear functions. Finding a solution for MOF described as MOO problems requires unique algorithms. Existing computer algorithms for finding solutions to MOO problems are built for specific types of functions, such as nonlinear, nonconvex, and nonsmooth functions (NSF), linear, convex and smooth functions, and nonlinear, convex and smooth functions. Thus, knowledge about the convexity and smoothness of FIRMS functions is significantly important to select the best solution algorithm.

### 3.2.10 Nonconvex, constrained and non-smooth functions (NSF)

A function is defined as nonconvex if it has multiple minima.[167] Often, optimization problems become numerically complex to solve when neither the objective functions nor constraints have a convexity structure. Constraints are conditions of an optimization problem that must be satisfied by the optimum solutions.[168] A function, *f*

geometrically described as *f*: $R^n \rightarrow R$ is convex if *dom f* is a convex set and $f(\theta x + (1-\theta) y) \leq \theta f(x) + (1-\theta) f(y)$ for all $x, y \in dom\ f$, $0 \leq \theta \leq 1$, as shown in Figure 3.17.[169]

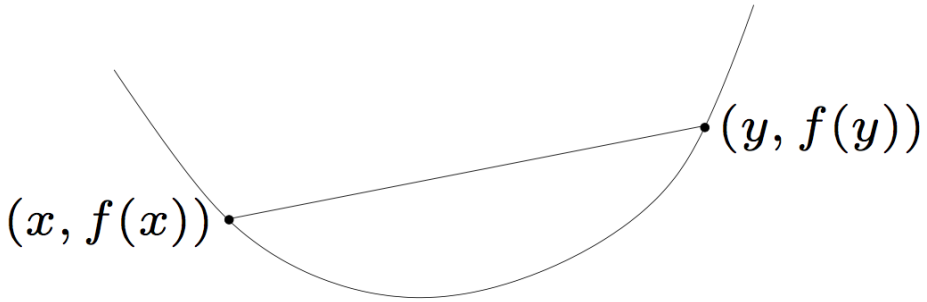



**Figure 3.17**. A graph describing a geometrical convex function f.

*f is* concave if *–f is* convex. *f* is strictly convex if dom *f* is convex and $f(\theta x + (1 - \theta) y) < \theta f(x) + (1-\theta) f(y)$ for $x, y \in$ dom $f$, $x \neq y$, $0 < \theta < 1$.[169]

A function is described as non-convex if it is neither convex nor concave, and has wave-like shapes, i.e., troughs and crests. Non-convex functions are often assumed to be nonsmooth (a function without continuous derivatives within a certain domain) with several feasible regions, and many locally optimal points within each region described by a set of boundaries. Since non-convex functions are nonsmooth and discontinuous, the derivatives are impossible to compute. Hence, the function cannot be described as increasing or decreasing (i.e., a possible solution does not provide enough information about a given point that can give enough information).

Of course, the general explanation of a nonconvex and nonsmooth function cannot explicitly explain the convex properties of FIRMS models. FIRMS functions are complex and require a different approach to characterize them as nonconvex and

nonsmooth. The procedure involves using vector and matrices in a given search space subject to constraints to explain the nonconvex properties of FIRMS functions.

Consider the objective function $c(x)$: $\mathbb{R}^n \rightarrow \mathbb{R}$ subject to constraints described by vector $g_j$: $\mathbb{R}^n \rightarrow \mathbb{R}$ for some integer *(n)* and isotope frequency *(i)*. The function *c(x)* is also of the form

$$c(\vec{x}) = |f(\vec{x})_{calculated} - f(\vec{x})_{experimental}| \quad (23)$$

for the vector $\vec{x}$ with isotope abundance frequencies as the variables, $f$ is a polynomial mapping from $R^n \rightarrow R$, and $f(\vec{x})_{experimental}$ is a constant calculated from the experimental data. Both the objective functions and polynomial functions are continuous with infinite derivatives, with the exception where $f(\vec{x})_{calculated} = f(\vec{x})_{experimental}$.

The objective function mapping from $\mathbb{R}^n \rightarrow \mathbb{R}$, which is continuous with 1st and 2nd derivatives can only be convex if it has a positive definite Hessian matrix.[170] The polynomial function $f(\vec{x})$ in Equation 15 is either smooth or nonsmooth and constrained to the convex n-cell defined by the isotope abundance frequency in $g_j$: $\mathbb{R}^n \rightarrow \mathbb{R}$.[171-172] Using tert-butyl chloride and the polynomial $f(\vec{x})$, Equation 7 can be formulated as Equation 24.

$$f(x, y, z) = 6\mu x^3 y^5 z - 9\mu x^3 y^6 z + 3\mu x^2 y^6 z - 3\alpha x^3 y^6 z - 3\alpha x^3 y^5 z \quad (24)$$

where vector $\vec{x}$ = *(x, y, z). x, y,* and *z* represent the isotope frequencies for carbon, hydrogen, and chlorine respectively, and μ and α are empirical constants. The Hessian matrix of Equation 16 can be formulated as

$$\mathrm{H} = \begin{bmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x\,\partial y} & \frac{\partial^2 f}{\partial x \partial z} \\ \frac{\partial^2 f}{\partial y\,\partial x} & \frac{\partial^2 f}{\partial y^2} & \frac{\partial^2 f}{\partial y \partial z} \\ \frac{\partial^2 f}{\partial z\,\partial \mathrm{x}} & \frac{\partial^2 f}{\partial z \partial y} & \frac{\partial^2 f}{\partial z^2} \end{bmatrix} \tag{25}$$

Where

$$\frac{\partial^2 f}{\partial x^2} = 36\mu x y^5 z - 54\mu x y^6 z + 6\mu y^6 z - 18\alpha x y^6 z - 18\alpha x y^5 z \tag{26}$$

$$\frac{\partial^2 f}{\partial y\,\partial x} = 90\mu x^2 y^4 z - 162\mu x^2 y^5 z + 36\mu x y^5 z - 54\alpha x^2 y^5 z - 45\alpha x^2 y^4 z \tag{27}$$

$$\frac{\partial^2 f}{\partial z\,\partial \mathrm{x}} = 18\mu x^2 y^5 - 27\mu x^2 y^6 + 6\mu x y^6 - 9\alpha x^2 y^6 - 9\alpha x^2 y^5 \tag{28}$$

$$\frac{\partial^2 f}{\partial y^2} = 120\mu x^3 y^3 z - 270\mu x^3 y^4 z + 90\mu x^2 y^4 z - 90\alpha x^3 y^4 z - 60\alpha x^3 y^3 z \tag{29}$$

$$\frac{\partial^2 f}{\partial x\,\partial y} = 90\mu x^2 y^4 z - 162\mu x^2 y^5 z + 36\mu x y^5 z - 54\alpha x^2 y^5 z - 45\alpha x^2 y^4 z \tag{30}$$

$$\frac{\partial^2 f}{\partial z \partial y} = 30\mu x^3 y^4 - 54\mu x^3 y^5 + 18\mu x^2 y^5 - 18\alpha x^3 y^5 - 15\alpha x^3 y^4 \tag{31}$$

$$\frac{\partial^2 f}{\partial z^2} = 0 \tag{32}$$

$$\frac{\partial^2 f}{\partial x \partial z} = 18\mu x^2 y^5 - 27\mu x^2 y^6 + 6\mu x y^6 - 9\alpha x^2 y^6 - 9\alpha x^2 y^5 \tag{33}$$

$$\frac{\partial^2 f}{\partial y \partial z} = 30\mu x^3 y^4 - 54\mu x^3 y^5 + 18\mu x^2 y^5 - 18\alpha x^3 y^5 - 15\alpha x^3 y^4 \tag{34}$$

But $f(\vec{x})$ is not Hermitian since it is not equal to its conjugate transpose, i.e.,

$$A^H \neq \bar{A}^T \tag{35}$$

where $A$ is an $n$ x $n$ matrix, $A^T$ denotes the transpose of the matrix $A$, $\bar{A}$ denotes the conjugate matrix, and $A^H$ symbolizes the Hermitian which arises from the combined operation of taking the transpose and complex conjugate of complex matrices.[173-174] Additionally, the second leading minor of the matrix in Equation 25 is negative which

means the matrix is not positive definite. The additive inverse of the matrix also shows leading minor and second-leading minor which results to nonnegative semidefinite. These descriptions do not meet the criteria described below:

let A = $(a_{ij})$ be an n x n symmetric matrix and $D_k$ the determinant for k = 1…, n be the leading principal minors:

$$D_k = \begin{bmatrix} a_{11} & a_{12} \dots & a_{1k} \\ a_{21} & a_{22} \dots & a_{2k} \\ \dots & \dots & \dots \\ a_{k1} & a_{k2} \dots & a_{kk} \end{bmatrix} \quad (36)$$

Then, A is positive definite if and only if $D_k > 0$ for k 1…, n. A is negative definite if and only if $(-1)^k D_k > 0$ for k = 1…, n (i.e., if and only if the leading principal minors alternate in sign beginning with $D_1$). A is positive semidefinite only if all its principal minors are nonnegative and, A is negative semidefinite only if its k$^{th}$ order principal minors are nonpositive for k odd and nonnegative for k even.[175-176]

FIRMS equations were tested against the principles described above to demonstrate that the matrix in Equation 25 cannot be convex or concave on the n$^{th}$ cell it is constrained. Thus, the FIRMS polynomial function *f(x)* in Equations 6 and 7 are nonconvex, hence, all FIRMS functions are nonconvex.

Additionally, the nonconvex and nonsmooth property of FIRMS functions were tested using computer algorithms in Analytical Solver SDK Platform acquired from Frontline Solvers. FIRMS functions showed the presence of many local extrema within the constrained boundaries, further evidence of nonconvex and nonsmooth functions that require nonlinear and nonconvex optimization algorithms. Such algorithms perform a random sampling of all possible solutions to determine the best solutions. These methods are nondeterministic or stochastic, which means that they might give different results

with the negligible difference in every analysis. They are entirely dependent on the points sampled.[148]

### 3.2.11 Solver algorithms

There are few known solvers for analyzing nonconvex and nonsmooth nonlinear models which have the characteristics of multiple feasible regions, multiple local optima within given regions, discontinuity, nonsmooth behavior, and lack of general information about the slope or derivatives in the direction of increasing or decreasing function. There are currently three algorithms for solving nonconvex and non-smooth complex problems such as FIRMS functions. They are Evolutionary/Genetic Algorithm, Tabu Search and Scatter Search Algorithms.

#### 3.2.11.1 Evolutionary/Genetic Algorithms (EA/GA)

Mathematical problems have three types of solutions: exact answers or roots, optimal, or extrema. In the case of optimal and extrema, the best solution is often determined relative to the problem being analyzed, the solution method, and the accepted tolerance values. Optimization is the process of altering the input of mathematical methods to find the minimum or maximum output. The input is comprised of variables, the mathematical method referred to as an objective function, and the output is the solution of an optimized objective function.[177]

In a real physical system where complex nonlinear problems exist and entropy plays a significant role, it is quite difficult to determine the best solution. Quantum theory implies that there are infinite dimensions that are different and correspond to decisions made. A small change in quantum systems can result in complex and random results. Many systems in the universe appear to replicate quantum systems, or are relatively less

complex. Thus, the high number of complex systems has demanded the development of sophisticated solvers. Optimization algorithms were designed as primary tools required to solve highly complex problems encountered when analyzing nature, and designing or predicting results.[177] As shown in Figure 3.18, there are six categories of optimization algorithms.

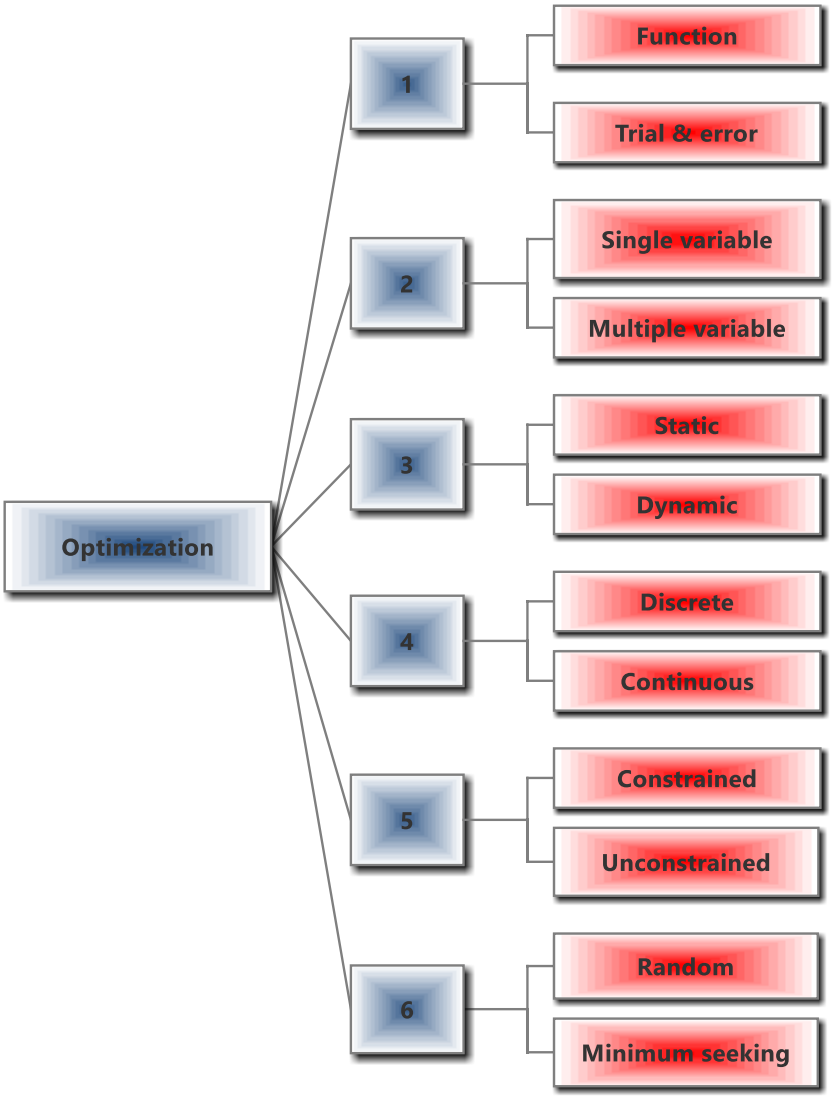



**Figure 3.18**. Examples of the optimization algorithms.[177]

Finding a solution to a problem can be seen as searching through a defined or non-defined space for possible solutions. Since the main objective is finding "the best" solution, the task thus bears the characteristics of optimization. For relatively small search spaces, there are classical methods that are suited for such problems, whereas for larger spaces, distinctive artificial techniques are used, such as EAs/GAs [177-178]

Evolutionary/Genetic algorithms (EA/GA) represent a popular meta-heuristic optimization and search technique guided by the principles of genetic evolution and natural selection.[159-160] This method was invented by John Holland over a decade between the 1960s and 1970s. It was later advanced and introduced to research by Holland's student, David Goldberg, when he solved a complex problem involving the control of gas-pipeline transmission for his dissertation research.[177]

The multi-objective EA has appealed to many researchers because it possesses many readily apparent advantages. EA can optimize with continuous, or discrete variables and functions without derivative information. The algorithm performs a simultaneous search from a broad sampling of the objective function while handling a large number of variables. EA can couple with parallel computers and optimize variables with extremely complex objective functions. It provides a list of optimum variables, performs optimization with encoded variables, addresses discontinuity, can solve nondifferential and nonconvex problems with multiple peaks, and can acquire the Pareto front in analysis.[151, 177]

There are two approaches to solving MOO problems. One involves the use of a single composite function or optimizing one objective function while the rest are considered a constraint set. The second optimization approach is to determine the

complete Pareto optimal solution set or a representative subset. As discussed previously, Pareto optimal solutions are nondominated with respect to each other. The single composite function with weighted sum method was applied in the optimization of FIRMS models.[179-180]

In EA optimization, a solution vector $x \in \Omega$ is known as an individual or a *chromosome.* They are of discrete units called *genes*. Genes control single or multiple features of the chromosome. In the original implementation of EA, Holland assumed that genes were represented by binary digits. However, modern EAs have varied gene types. Usually, a chromosome relates to a specific solution $\boldsymbol{x}$ in the feasible space. As a result, it requires mapping mechanisms between the feasible solution space and the chromosomes which is a process known as encoding.[178-179]

A collection of chromosomes is known as the *population* which is initiated through a random process. During optimization, the search evolves, as does the population including the fitness functions. The two operators used to generate new solutions in the EA are the *crossover* (the most important operator) and *mutation*. During crossover two chromosomes known as *parents* combine to form new chromosomes or *offspring*. These parents are selected among the existing chromosome population with good fitness in order for the offspring to inherit good genes that make them fitted parents. Through iterative application of the crossover operator, the chromosomes with good genes survive in the population, which eventually leads to convergence of the optimum solution.[178-179]

The function of the mutation operator is to introduce random changes into the features of chromosomes. Its implementation at gene level is very minute and depends on

the size of the chromosomes. Mutation is fundamentally significant in GA because it reintroduces genetic diversity back into the population which helps the algorithm overcome local optima.[178-179]

Similar to biological processes, the reproduction comprises a selection of chromosomes for the following generation. The fitness of individuals in a population determines their probability of surviving through generations. The selection process depends on how fitness functions are applied. Popular selection procedures include proportional selection, and ranking.[179]

The GA algorithm used for the FIRMS model optimization was built by Frontline Solvers utilizing Evolution Programs (EP) referenced in Michalewicz.[178] It uses a floating-point representation. The mutation methods are implemented as uniform, non-uniform (time-dependent), and bounds. The crossover methods include quadratic interpolation following controlled random search algorithm (CRS6)[181-182] and total and partial arithmetic crossover (convex combination). However, in general, all GAs/EAs methods follow the process described by:[179]

*Step 1*

Set **t = 1**, Produce **N** solutions randomly to form the first population, $\boldsymbol{P_1}$ and evaluate the fitness of solutions in $\boldsymbol{P_1}$.

*Step 2*

*Crossover*: Produce an offspring population $\boldsymbol{S_1}$ through the following process:

2.1. Select two solutions **x** and **y** from population $\mathbf{P_1}$ based on fitness values.

2.2. Generate offspring and add to $\mathbf{S_1}$ using crossover operator.

*Step 3*

*Mutation*: Mutate individual solution $\boldsymbol{x} \in \boldsymbol{S_1}$ using a predefined mutation rate.

*Step 4*

*Fitness assignment*: Assess and assign a fitness value to individual solutions of $\boldsymbol{x} \in \boldsymbol{S_1}$based on its objective function value and infeasibility.

*Step 5*

*Selection*: Choose N solutions from $\boldsymbol{S_1}$ based on their fitness and replicated them to $\boldsymbol{P_{t+1}}$.

*Step 6*

If the termination criteria is met, the process is stopped and the search returns to the existing population otherwise $\boldsymbol{t = t + 1}$ and Step 2 is repeated.

**3.2.11.2 Tabu search and Scatter search algorithms**

There are other meta-heuristic optimization methods other than GAs/EAs that have been shown as very effective in finding solutions for complex physical problems. For example, they are effective for systems with nonlinear objective functions or constraints with the possibility of a large number of discrete variables or nonsmooth, discontinuity and presence of uncertainty in important data. Thus, to solve such complex systems, researchers have trusted other metaheuristic methods, such as Tabu Search (TS) and Scatters Search algorithms (SS).[180, 183-184]

**3.2.11.2.1 Tabu search algorithms**

Tabu search was developed as a computational algorithm for solving combinatorial optimization problems with a wide range of applications, such as graph theory, matroid settings, and general pure mixed integer programming problems. It can

adapt and apply other existing methods (e.g., linear programming algorithms and specialized heuristics) which allows it to overcome the limitations of local optimality.[185] It was originally developed for combinatorial procedures applied in nonlinear problems[186] and also different groups of problems that include scheduling and computer channel balancing, cluster analysis, and space planning.[187-190]

Caballero et al.[191] developed a novel TS for combinatorial optimization problems with two approaches: 1) a local Tabu Search Algorithm that generated a list of locally Pareto-equivalent solutions (non-dominant solutions with respect to the neighboring solutions); 2) a procedure that strengthened solutions with a path-relinking strategy.[191] Path relinking strategies were developed for use in TS and SS algorithms, Greedy Randomized Adaptive procedures and discrete optimization field.[180, 192-194] The strategy follows the underlying principle that the direction of the path in a design space can be better used to guide the search in that direction than a random one, with the hope of finding good points. The direction (or path) can be determined by use of initial point and a reference point. The reference is the best-known solution in the search space. For example, given two points $\boldsymbol{x}$ and $\boldsymbol{y}$ (reference point) of a vector, the search space between the two points is a linear combination of the vector variables in the space, and as such the retained information about the reference point can lead to good solutions points.[180, 192] Jaeggi et al.[180, 195] developed the first multi-objective TS variant that is widely used today for unconstrained functions and constrained optimization problem.

#### *3.2.11.2.2 Algorithm*

The TS is sequential and iterative, in which the search starts at a chosen point and the algorithm picks a new point in the search space that becomes the next current point.

The recently selected points are stored in the short-term memory (STM) and the search is not allowed to return to the values. Optimal points are stored in medium term memory (MTM) and are used for strengthening and directing the search on areas of the search space with known good objective function values. The long-term memory (LTM) keeps the record of regions of the search space explored and is used to strengthen and focus the search in regions less explored. This is done by dividing each design variable into $\boldsymbol{n}$ regions and keeping the count of evaluated solutions in that region. After a successful addition to MTM, the process is reset using a counter until the counter reaches the user-specified value, and then the solution to the problem is produced.[180]

Mathematically, TS optimization can be applied in minimization via the following steps.

Using Equation 10, one trial solution is selected in the space $\boldsymbol{x} \in \boldsymbol{\Omega}$. A subset $\boldsymbol{\Omega}(\mathbf{s})$ of $\boldsymbol{\Omega}$ is also created in the form: $\boldsymbol{\Omega}(\boldsymbol{s}) \rightarrow \boldsymbol{\Omega}$. The space associated with $\boldsymbol{x} \in \boldsymbol{\Omega}$ is then defined in the subset $\boldsymbol{S}(\boldsymbol{x})$ that consists of the moves $\boldsymbol{s} \in \boldsymbol{S}$ that are applied for the **x** values. Those elements in $\boldsymbol{s}$ are obtained by a non-Markovian function that uses information stored from search process extended to $\boldsymbol{t}$ iterations in the past, where $\boldsymbol{t}$ can be a constant or variable depending on the optimization problem. The values in $\boldsymbol{s}$ are well guided by a set of tabu conditions, such as linear inequalities or logical relationships which are expressed indirectly in the form of a trial solution $\boldsymbol{x}$; for example, $\boldsymbol{s}(\boldsymbol{x}) = \{\boldsymbol{s} \in \boldsymbol{S}: \boldsymbol{s}(\boldsymbol{x})\}$ that may or may not defy the set tabu conditions. A simple tabu search process is described in the simple example below for the function in Equation 10.[185]

*Step 1*

Select an initial $\boldsymbol{x} \in \boldsymbol{\Omega}$ and let $\boldsymbol{x}^* := \boldsymbol{x}$. Set the iteration counter $\boldsymbol{t} = \boldsymbol{0}$ and begin with **S** empty.

*Step 2*

If **S(x) – s** is empty, go to Step 4.

Otherwise, set $\boldsymbol{k}: = \boldsymbol{k} + \boldsymbol{1}$ and select $\boldsymbol{s_k} \in \boldsymbol{S}(\boldsymbol{x}) - \boldsymbol{s}$ such that $\boldsymbol{s_k}(\boldsymbol{x})$ = **OPTIMUM(s(x))**:

$\boldsymbol{s} \in \boldsymbol{S}(\boldsymbol{x}) - \boldsymbol{s}$.

*Step 3*

Let $\boldsymbol{x}: = \boldsymbol{s_k}(\boldsymbol{x})$. If $\boldsymbol{f}(\boldsymbol{x}) < \boldsymbol{f}(\boldsymbol{x}*)$, where x* denotes the best solution currently found, let **x*: = x**.

*Step 4*

If a selected number of iterations has elapsed either in total or since **x*** was last upgraded, or if $\boldsymbol{S}(\boldsymbol{x}) - \boldsymbol{s} = \varnothing$ (where $\varnothing$ is an arbitrary value) upon reaching this step directly from Step 2, stop.

Otherwise, update s (as subsequently identified) and return to step 2.[185]

In the above Tabu optimization: 1) $\boldsymbol{s}$ provides the constraints values, therefore the solution found depends on values of $\boldsymbol{s}$ and the update process in Step 4; 2) there is no reference to the condition of local optimum with the exception of areas where the local optimum implicitly provides a better solution from previous values, and 3) the best move is chosen at every step using the criteria in step 2 above.[185]

Tabu search demonstrates its superiority over other classical optimization methods through solving a problem by use of best available values that do not lead to inferior solutions. Also, local optimum values do not provide an obstacle, therefore enhancing the speed at which the local minimum might be considered and abandoned if necessary. [185] The tabu search algorithm is recommended in many circumstances where other traditional solvers cannot optimize complex nonlinear functions and find the best global minimum solution.

***3.2.11.2.3 Scatter Search Algorithms***

Glover[194] developed the scatter search method for finding starting solutions and trial solutions of a function. It uses initialized values that are purposely generated to account for the behavior of the function in numerous parts of the solution search space. It works by orienting its explored values relative to a set of reference points. The reference points might consist of good solutions found by prior problem-solving or evaluation of functions.[186] The scatter search performs the evaluation of a function through the following procedure:

1. A starting set of solution vectors are generated by heuristic processes intended for the evaluation of the problem under analysis. Some best vectors of the selected subset are established as reference solutions.
2. New points are created consisting of linear combinations of subset of the existing reference solutions. These linear combinations are chosen to produce points both in and outside the convex regions covered by the reference solutions. They are also modified by a general rounding method to produce integer values for integer constrained vector components.

3. The best solutions generated in Step 2 are extracted as a group and then used as initial points for a new application of the heuristic processes in Step 1. The process is repeated until the required iteration limit is reached. An example of the method is shown in the Figure 3.19 below.[186]

In Figure 15, the points labeled 1 through 16 are the designated central points for a given subregion of the simplex A, B and C. Point B is the main central point of A, B, and C. The points A, B C are not always used as the original points of reference. Any coordinate point such as 16, 11 and 13 could be established as reference points. For scatter search the choice relies on the distribution of the original points that are relative to each other and also their feasibility. In order to ensure the desirable limits for larger dimensions, the points produced as shown in the diagram are examined in a sequence ascending order, or an alternative sequence that depends on the gradient of the objective function contour. The generation of points continue until a desired point is reached as shown in Figure 3.19, and sometimes less points might be considered.[186]

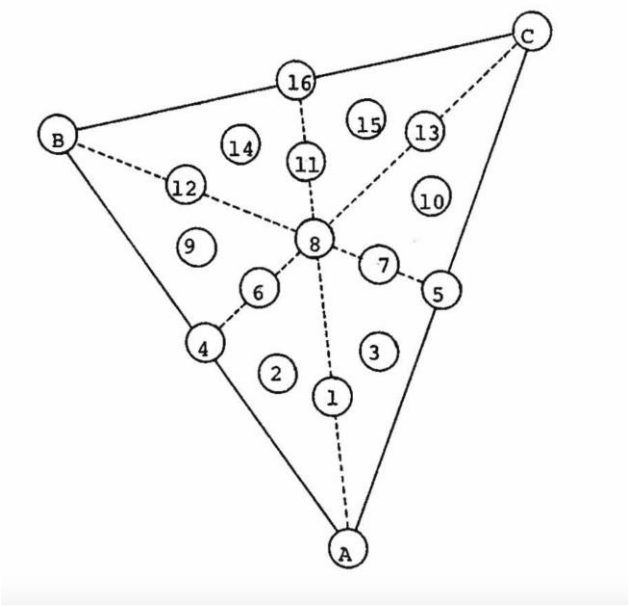


**Figure 3.19**. Generated vectors of scatter search process.

SS has three important features namely: 1) the linear combinations are built with the objective of producing weighted centers of selected subregions, hence allowing for nonconvex combinations that focus the centers into regions outside the original reference solutions. The dispersion pattern produced by such centers and their external focus are applicable for mixed integer optimization; 2) the process for selecting specific subsets of a solution are built to make use of clustering, hence allowing different types of important variation by producing new solutions within and across the clusters; 3) they support heuristics that are able to begin from infeasible solutions, therefore removing any restriction that selected solutions for re-applying the heuristic method must be possible.[186]

The SS and TA algorithm provided by the Frontline Solvers were built as solver engines using the work of Ugray et al.[196] GAs/ EAs, TA and SS algorithms obtained from the Frontline Solvers were found suitable for analysis of IRs using FIRMS methods. They were integrated into the FIRMS graphical user interface (GUI) software.

## 3.3 FIRMS Software

### 3.3.1 Introduction

The software is a major component of the FIRMS method and the main computational tool used in the analysis of tandem MS data for FIRMS compounds. The models created by FIRMS using the information from molecular fragmentation data of compounds obtained via tandem mass spectrometry were impossible to solve analytically and too complex to solve numerically without the help of advanced computation. The complexity of the models and functions increased from simple molecules to complex molecules with long chain branches, rings, functional groups and in some cases,

rearrangement, bond formation and occurrence of protonation, deprotonation . Moreover, determining the combination of atoms that produce a particular fragment of molecular ion was extremely difficult for molecules with more than two atoms and elements. The complexity of FIRMS models and functions were further amplified by the molecules that fragment from a molecular ion mass M+2. Finding solutions for the FIRMS models and functions were therefore impossible by hand and necessitated the use of FIRMS software that is integrated with simulation optimization solver.

### 3.3.2 C# programming language

There are dozens of computer programming languages in existence. For example, Alice, APL, COBOL, Pascal, Eiffel, FORTRAN, Haskell, Mercury, ML, Mondrain, Oberon, Perl, Python, RPG, Scheme, Smalltalk, C, C++, Visual F# and C# among many. We chose to build the FIRMS software in C#[197-198] language for the reasons discussed below.

C# was originally created from scratch for .NET, which is a platform in which computer programs run and was designed as a new programming paradigm. It is not an operating system, like Windows platforms, but it is a layer between the operating system and other applications. Thus, it offers a platform for developing and running computer code that is easy and efficient to use with a set of development tools. Some of the standard features of .NET are:[197]

- *Multilanguage independence*: It supports the development of computer applications using a number of programming languages i.e., Visual Basic, C# and C++ and also third-party languages that compile to the established Microsoft Intermediate Language (IL). This can allow parts of an applications to be

developed in different languages like Visual Basic and C# then translated to a common IL.

- *New Framework base classes*: It has a class library, which provides an expansive collection of more than 2500 reusable classes available to the .NET languages.
- *Dynamic Web pages and Web services*: .NET can deploy applications over intranets or internet using ASP.NET, ADO.NET and XML which allow access to relational databases, and import and export data through internet web.
- *Scalable component development*: .NET supports object-oriented programming and development that is faster and easier in terms of performance. It allows the creation of components of codes as individual entities, and can be kept independently, reused or combined with other applications. These components can be created in different languages.

For these reasons, the .NET platform was used for FIRMS. Thus, it was very convenient to use C# programming language since it was created and designed for the .NET platform. The large number of classes included as part of .NET could, therefore, be used and new classes easily designed in C# language. As a result, it would reduce the algorithms and programming codes needed for FIRMS software. Moreover, C# coupled together with .NET Framework classes, offered an opportunity to incorporate and use emerging Web standards, for example, Hypertext Markup Language (HTML), Extensible Markup Language (XML), and Simple Object Access Protocol (SOAP) which makes it easy to download updates for FIRMS software via internet. Additionally, the Visual Studio IDE and the .NET Framework provided a simple platform to create graphical user

interfaces (GUI) which is an essential component of FIRMS software. C# also offers excellent data handling speed to which C and C++ programmers are used to. C# also enables programmers to build packages available over the internet using many standard features available only in Java, Common Gate-way Interface (CGI), or PERL.[197] Thus, users who would like to couple their applications with FIRMS would find it easy to do so without having to convert their apps algorithm into C#.

### 3.3.3 FIRMS software components

The FIRMS software has the following major components: GUI, structural modeling algorithm, combinatorial data structure algorithm, IRs calculation algorithm, import data and compound structures algorithms, export data algorithms, and simulation optimization solver algorithms.

#### 3.3.3.1 Structural modeling algorithm

The molecular structures of organic compounds analyzed by FIRMS are imported from PubChem. The process involves extracting molecular structures in canonical SMILES format. SMILES is chemical notation language designed for converting an arbitrarily selected description of a chemical structure to one distinctive notation. The notation is unique in order to preserve the identifier of the specific structure for database and other computer applications.[199]

SMILES chemical notations involve a method known as CANGEN which combines two separate algorithms, CANON, and GENES. In the first step, CANON labels a molecular structure with canonical labels and treats it as a graph with nodes (atoms) and edges (bonds). Each atom acquires a unique numerical label based on its topology. In the second step, GENES generates the distinctive SMILES notation as a tree

diagram of the molecular graph. The GENES choose the starting atoms, and builds and makes decisions in reference to the canonical labels as required. These two processes designate a unique SMILES notation for every chemical structure irrespective of the existence of many possible equivalent descriptions of the structures.[199] An example of SMILES is shown below using the structure of vardenafil.


**Figure 3.20**. Conversion of vardenafil into SMILES.

The SMILES format of the molecule is then converted to a 2D Chemical Markup Language (CML) format using the ***Open Babel*** utility. CML was developed by Murray-Rust et al.,[200] as an operational system for managing complex chemical content in interoperating XML-based markup languages. It incorporates the CML 1.0 and a mechanism that displays CML molecules within a standard web browser. It has significant advantages of data recovery capabilities and flexibility over other existing computer HTML/plugin solutions.[200-202]

Open Babel is an open-source chemical toolbox written in computer codes based on languages of chemical data. It provides a platform for multiple chemical file formats with additional utilities of 2D depiction, filtering, batch conversion, and substructure and similarity searching. Researchers and developers use it as a programming library for

handling chemical data in disciplines including drug design, organic chemistry, materials science and computational chemistry.[203]

After the process that involves SMILES, CML and Open Babel, the formatted 2D CML information obtained from Open Babel is saved in the FIRMS files directory and can be imported into the FIRMS GUI as a new XML file. Once a molecular ions structure is imported in FIRMS GUI, the user can select different atoms within the structure to associate with fragments observed in the mass spectrum of a molecule. The user also has the ability to manipulate the structure to account for rearrangements, multiple bond formation, and protonation or deprotonation within a molecular ion. When users are satisfied with the structures created, they continue to the next step of producing a model that represents the structure.

#### 3.3.3.2 Combinatorial data structure algorithm

The structure selected by the user, which is associated with a fragment in the mass spectrum is linked together with the stack data structure algorithm previously discussed (Section 3.2.9) to create all accepted combinations of atoms that produce the fragment mass observed in the mass spectra. Once these models are created, they are linked to tandem mass spec data for further analysis and calculation.

#### 3.3.3.3 Import data and structure algorithm

The FIRMS software allows the user to import data extracted from MS data analysis software and link the fragments with the structural models created in the GUI. Specifically, the structural fragments drawn by the user are assigned to specific m/z values from the mass spectral data.

#### 3.3.3.4 IRs calculations, export, and numerical solver algorithm

Once the computer models are created and mass spectra are imported into the software, the user sets constraints and tolerances and directs the software to initiate the numerical solver. At the end of optimization, the FIRMS software displays the results in the form of isotope frequencies, their standard deviations, and calculated δ values. The information obtained by the user can be then be used to characterize the compound in order to draw a conclusion.

### 3.4 Significance of FIRMS

While other IR analysis techniques have been excellent at accurately determining IRs, FIRMS was developed in order to compensate for other limitations of these techniques. The first important role that FIRMS would play as a new method of IRs analysis is the ability to investigate the structure of a molecule under analysis. This is a current problem with IRMS and CRDS, in which combustion is necessary to produce a gas that is then ionized and measured against standard gases. The standard gases include $N_2$, $CO_2$, $SO_2$, and $H_2O$. Due to loss of structure, it is extremely difficult to identify compounds in situations where unknown compounds are analyzed. Also, in a mixture of several molecules, converting each component into a gas generates problems, because it is impossible to associate simple gas molecules produced with a specific compound in the mixture. With FIRMS, it is easy to evaluate the IRs of individual molecules of a mixture. This is because mass spectral techniques have excellent resolution and sensitivity that can be used to identify a molecule. Furthermore, FIRMS relies on mass fragments of precursor ions (product ions) to analyze IRs. Every molecule fragments differently in a mass spectrometer, which makes identification easier in many cases. On the other hand, FIRMS can also be directly coupled with chromatography methods that will separate

components of mixtures. Moreover, co-eluting components would theoretically not interfere with the IR analysis of each.

Secondly, FIRMS reduces the possibility of contamination. NRM, CRDS, and IRMS are prone to contamination during sample preparation, and introduction. Standard gases might also be contaminated. FIRMS eliminates this problem by targeting a particular mass to charge, and uses both precursor and product ions.

Third, most IR analysis techniques produce aggregate isotope ratios of the atoms present in a molecule. For example, by converting a molecule to $CO_2$, both IRMS and CRDS analyze the aggregated value $\delta\ ^{13}C/^{12}C$ as one number despite the fact that most organic compounds have more than one carbon present. FIRMS has the ability to analyze each atom or group of atoms separately to determine IRs at specific locations in a molecule.

### 3.5. Conclusion

FIRMS, a nonlinear mathematical model was developed with the potential to analyze the isotope ratio of organic compounds. The theorem was proved to be true and can model a structure of any organic compound by utilizing FIRMS data structure algorithms. Theoretically it has been shown that Evolutionary/Genetic, Tabu and Scatter search algorithms are the relevant methods of minimizing the objective functions of FIRMS based on full characterization and analysis of FIRMS functions. However, Evolutionary (a non-gradient based) algorithm will be used in developing FIRMS and will be compared to gradient based Generalized Reduced Gradient (GRG) methods for validation. Additionally, computations and data structures proved to be valuable in putting together the FIRMS graphical user interface in C# programming language.

# Chapter 4: Analysis of Evolutionary/Genetic Algorithms (EA/GA), Generalized Reduced Gradient (GRG) & Multi-start GRG

## 4.1. Introduction

Problems with multiple objectives naturally arise in many disciplines. Calculating solutions to these problems is difficult for scientists to this day. Although there are many techniques that may solve these problems, the complexity of the mathematical analysis involved has influenced the development of alternative approaches. Additionally, some MOO problems are so complex that the use of standard multiobjective techniques does not provide adequate results. Complexity can manifest as a large search space, uncertainty, noise, and discontinuity of Pareto curves. The application of EA/GAs to find solutions for MOO problems has been of interest to many researchers. Mainly because of their population-based nature, EAs allow the production of several elements of the Pareto optimal set in a single analysis. Thus, given the nature and complexity of FIRMS functions, three alternative approaches to solving MOO problems were evaluated for their performance in solving the FIRMS MOO problem. EA/GA, a non-gradient algorithm, GRG (gradient based), and Multistart GRG (gradient based) algorithms were compared to evaluate the most appropriate numerical optimization algorithm. As discussed in Chapter 3, FIRMS functions should require non-gradient based algorithm, and GRG was included to test this hypothesis.

## 4.2 EA/GA versus GRG

Frontline Systems Inc. recommends Multistart GRG and Evolutionary algorithms for optimization of non-convex and nonsmooth functions, such as FIRMS equations. GRG (non-gradient) is ineffective at analyzing non-convex problems because the

solutions found are local extrema rather than global extrema. However, Multistart GRG is capable of evaluating many different locations and initial values within a domain and examines the best solution. But both GRG and Multistart GRG require derivatives for continuous functions. FIRMS equations of the form $c(\vec{x}) = |f(\vec{x})_{calculated} - f(\vec{x})_{experimental}$ are not differentiable where $c(\vec{x}) = f(\vec{x})_{experimental}$ the solution point. This is the reason why FIRMS IRs theory argues for non-gradient based optimization methods, such as Tabu search, Scatter search and EA/GA. To test this hypothesis, a non-gradient based algorithm was compared with the performance of gradient-based algorithms. Thus, a study was initiated to compare the accuracies of GRG and Multistart GRG, gradient-based methods, to the EA/GA non-gradient based method using several organic compounds.

**4.3 Evaluation of EA/GA and GRG algorithms using organic compounds**

Several trials were conducted on EMPA, p-toluenesulfonic acid monohydrate and 5-sulfosalicylic acid dihydrate. The isotope abundance variables were held constant for each compound (e.g., $^{12}C$, $^{13}C$, $^{1}H$, $^{2}H$, $^{16}O$, $^{17}O$ and $^{31}P$ for EMPA). Different settings of constraints were used based on linear inequalities with varying settings for each type of solver. For example, in the Evolutionary solver, the mutation rate was varied between 0.075 and 0.2 while GRG solver was alternated between the Multistart GRG and standard GRG.

The data used to evaluate the two algorithms were calculated using Equations 5 and 6 with total abundances for the base mass (M) experimentally determined, and μ values calculated based on Equation 7-10, given an α calculated using Equations 2 and 3. The solver minimized the objective value calculated from these data to give δ values. δ

values calculated by GRG and EA were then compared to the real δ calculated from the known isotope abundances of elements. The average, standard deviation, minimum, lower quartile, median, upper quartile, maximum, and range of the sum of absolute differences from true δ values (IRs), and of the sum of absolute differences between the μ values after optimization were analyzed.

Comparison of the results produced by GRG and Evolutionary algorithms showed that the latter outperformed the former. As shown in Table 4.1, the average sum of absolute difference between the real δ values and those calculated by EA shows that EA results were significantly lower, meaning EA results were more accurate than GRG. Similar results were observed when evaluating standard deviations, minimum, maximum, lower and upper quartile, and median. Moreover, the range of sum of absolute differences of δ observed were greater for GRG than EA, which means that GRG δ values were significantly far from the real solution.

**Table 4.1.** The difference of sum of absolute δ values.

| | **Sum of absolute δ differences** | | | | |
|---|---|---|---|---|---|
| **Compound** | **Algorithm** | **Average** | **Standard Deviation** | **Minimum** | **Maximum** |
| EMPA | EA | 170.7 | 159.4 | 9.5 | 397.9 |
| | GRG | 218.8 | 180.3 | 56.7 | 928.3 |
| p-Toluenesulfonic acid monohydrate | EA | 356.0 | 87.5 | 257.3 | 511.2 |
| | GRG | 394.3 | 159.1 | 200.5 | 541.5 |
| 5-sulfosalicylic acid dihydrate | EA | 543.6 | 134.8 | 300.5 | 669.0 |
| | GRG | 556.7 | 149.9 | 283.1 | 669.2 |

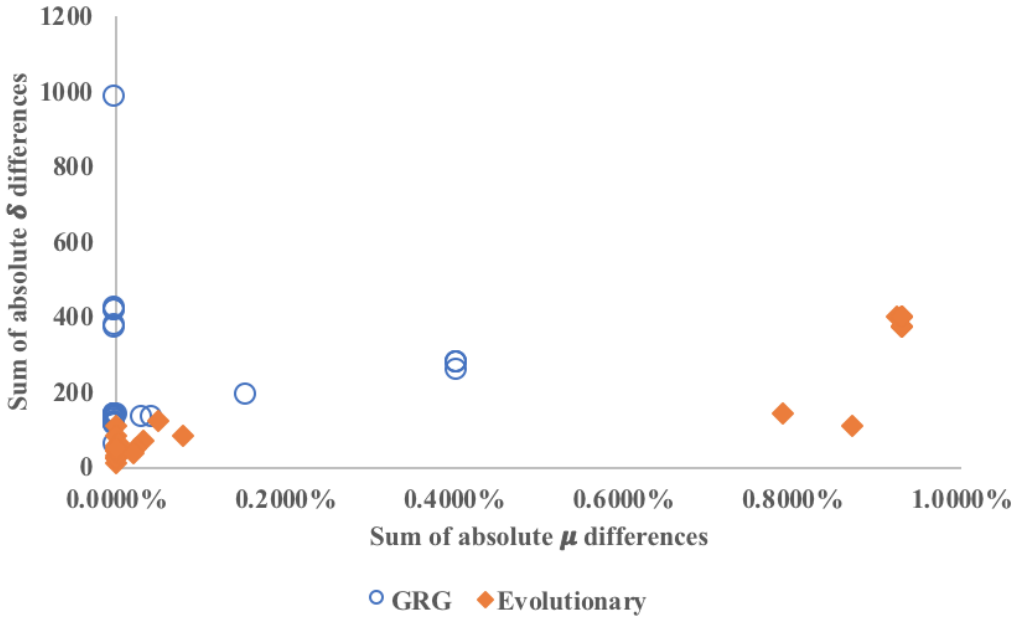



**Figure 4.1.** A plot of sum of absolute difference of δ values (y axis) and μ values (x axis) for EMPA. The average sum of absolute difference between the real δ values and those calculated by EA shows that EA results were significantly lower meaning EA results were more accurate than GRG.

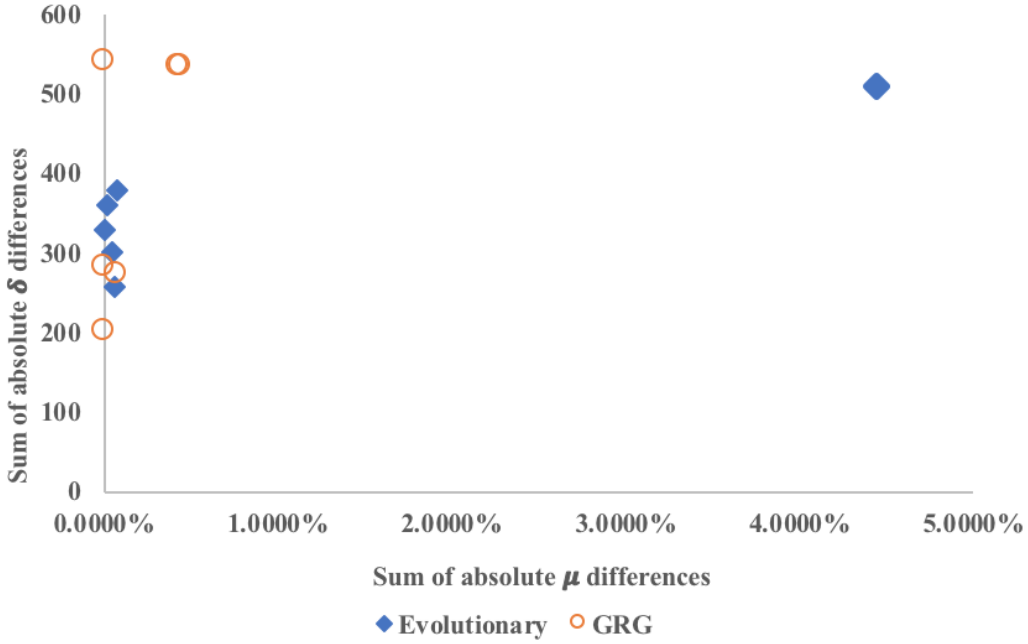



**Figure 4.2.** A plot of sum of absolute difference of δ values (y axis) and μ values (x axis) for p-Toluene sulfonic acid monohydrate. The average sum of absolute difference

between the real δ values and those calculated by EA shows that EA results were significantly lower meaning EA results were more accurate than GRG.

However, finding accurate solutions did not mean that EA minimized the objective function better than GRG algorithms. On the contrary, as shown in Figures 4.1 and 4.2, for the same analysis GRG methods minimized the objective functions to the lower values than EA algorithms. The small difference between calculated and experiment μ suggested that GRG algorithms are sometimes better in terms of optimization performance.

Figure 4.3 shows the δ differences of different trials, indicating that the EA and GRG algorithms appear to have similar or parallel characteristics, with EA performing better than GRG. For each adjacent pair of GRG data, the first is regular GRG and the second is Multistart GRG. And for an adjacent pair of EA, the first is Evolutionary with mutation rate 0.075, and the second is Evolutionary with mutation rate 0.2. This figure shows that individual configuration and changing different parameters may improve or give undesired results. For example, for trial 4, the mutation rate of 0.2 produced δ solutions with a larger error compared to mutation rate of 0.075.

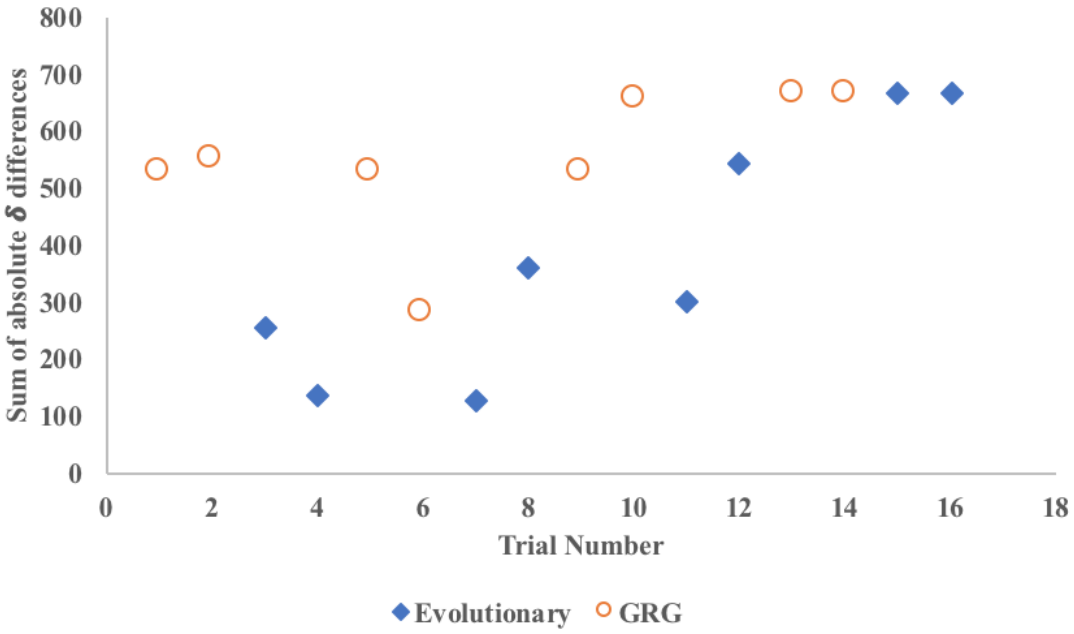



**Figure 4.3.** The δ differences for 5-sulfosalicylic acid dihydrate against the trial numbers. Each pair of data have different parameters being tested. It shows that individual configuration and changing different parameters may improve or give undesired results.

### 4.4. Conclusion

Overall, the solutions found using the Evolutionary algorithm were more accurate compared to solutions produced by GRG algorithms. However, some of the GRG minima for δ were smaller than the minima calculated by EA/GA. However, Figure 4.3 also shows that EA/GA solutions were more accurate compared to GRG. The reason for this observation is because EA/GA tends to find solutions corresponding to more gradual declines that revolve around the true solution. On the other hand, GRG appears to find steep and narrow troughs that may exist far from the true solution. Thus, EA/GA was selected as the most appropriate algorithm for optimization of FIRMS objective functions

# Chapter 5: Materials and Method

## 5.1 Reagents and Standards

Ammonium hydroxide, methanol, ethanol, isopropyl alcohol, acetonitrile and acetone (HPLC-grade or higher), formic acid, o-phosphoric acid hydrochloric acid, and acetylsalicylic acid were purchased from Thermo Fisher Scientific (Ward Hill, MA, USA). Ethyl methylphosphonic acid (EMPA) was acquired from Cerilliant (Round Rock, TX, USA). 3-chloro-2-nitrobenzoic acid, 2,3-dicyanohydroquinone, L-histidine monochloride, L-tyrosine, o-tert-butyl-L-tyrosine, eugenol, caffeine, 3,4-dibromo benzaldehyde, 3-phenylpropionic acid, 4-nitrobenzaldehyde, p-toluene sulfonic acid monohydrate, tert-butyl chloride, and caffeine were acquired from Sigma-Aldrich (St. Louis, MO, USA). Diisopropyl methylphosphonate (DIMPA) and p-toluene sulfonyl chloride were purchased from Acros Organics (Morris Plains, NJ, USA). 5-chloro-2-methylbenzoic acid was purchased from both Sigma-Aldrich (St. Louis, MO, USA) and Alfa Caesar (Tewksbury, MA, USA).

## 5.2 Sample preparation

Compounds were dissolved in various solvents, as shown in Table 5, to produce 5 mM stock solutions. The solutions were further diluted to create standards with concentrations between 60 and 100 μM for analysis using DLSI-MS-MS, and LC-MS-MS.

In general, FIRMS analysis included four major steps: 1) the analysis of a compound using MS to identify fragments and transitions of interests; 2) the collection of MS-MS data using DLSI and/or chromatography; 3) converting the MS-MS data to the FIRMS format, including selection of fragments, boundaries, and constraints, and 4) calculating $\delta^{13}C$ using FIRMS software.

**Table 5.1.** Analytes and solvents used to create the desired concentration standard.

| # | Compound | Solvent | Conc (μM) |
|---|---|---|---|
| 1 | Acetylsalicylic Acid | Ethanol | 100 |
| 2 | Caffeine | Acetone | 100 |
| 3 | 5-chloro-2-methylbenzoic acid[1] | Methanol | 100 |
| 4 | 5-chloro-2-methylbenzoic acid[2] | Methanol | 100 |
| 5 | 3,4-dibromo benzaldehyde | Methanol | 100 |
| 6 | 2,3-dicyano hydroquinone | Methanol | 100 |
| 7 | Diisopropyl methyl phosphonic acid[3] | Water | 60 |
| 8 | Diisopropyl methyl phosphonic acid[4] | Water | 60 |
| 9 | Ethyl methyl phosphonic acid | Water | 60 |
| 10 | Eugenol | Methanol | 60 |
| 11 | L-Histidine | Water | 100 |
| 12 | 4-nitrobenzaldehyde | Methanol | 100 |
| 13 | 3-Phenylpropionic acid | Methanol | 70 |
| 14 | o-tert-butyl-L-Tyrosine | Aqueous Formic Acid (1 mM) | 70 |
| 15 | L-tyrosine | Aqueous Formic Acid (1 mM) | 100 |
| 16 | p-Toluene sulfonyl chloride | Methanol | 100 |
| 17 | p-toluene sulfonic acid | Methanol | 80 |

[1]China, [2]India, [3]France, [4]Great Britain

### 5.3 Direct liquid sample introduction analysis (DLSI)

The standards prepared according to Table 5.2 were analyzed using an AB Sciex Q-trap 5500 MS-MS (Applied Biosystems, Foster City, CA, USA) DLSI to acquire precursor and product ions scans. The standards were introduced using a 1 mL Luer Lock Gastight Syringe (Hamilton Robotics, Reno, NV, USA). The declustering potential (DP) and collision energy (CE) were optimized for product ions (fragments) of interest. A multiple reaction monitoring (MRM) method was developed using the optimized parameters and used to detect MRM product ions in either negative or positive polarity, depending on the compound being analyzed. For example, MRM transitions for L-histidine monochloride are listed in Table 5.2. MRMs were monitored consecutively for 100 ms

each for a total of 10 minutes at a flow rate of 10 μL/min for five different days. At the end of each sample analysis, product ion abundances of each transition for the compounds were extracted as text files for analysis using the FIRMS software.

**Table 5.2.** MRM transitions for L-histidine monochloride including declustering potential (DP) and collision energy (CE) measured in volts (V).

| Precursor Ion(Da) | Product ion (Da) | DP(volts) | CE(volts) |
|---|---|---|---|
| 156.0 | 56.0 | 79.83 | 44.91 |
| 156.0 | 66.1 | 91.74 | 58.37 |
| 156.0 | 83.0 | 46.61 | 34.49 |
| 156.0 | 93.1 | 75.56 | 30.41 |
| 156.0 | 110.0 | 18.49 | 18.15 |
| 157.0 | 56.1 | 26.46 | 54.74 |
| 157.0 | 57.0 | 80.62 | 48.90 |
| 157.0 | 66.1 | 19.34 | 57.36 |
| 157.0 | 67.0 | 59.12 | 53.56 |
| 157.0 | 83.1 | 40.31 | 36.27 |
| 157.0 | 84.0 | 58.27 | 36.79 |
| 157.0 | 93.0 | 17.55 | 37.06 |
| 157.0 | 94.1 | 20.34 | 31.02 |
| 157.0 | 110.1 | 66.81 | 17.81 |
| 157.0 | 111.1 | 60.12 | 17.51 |

### 5.4 LC-MS-MS analysis

Liquid chromatography-tandem mass spectrometry (HPLC-MS-MS) experiments were conducted for five different (sometime consecutive) days to evaluate the applicability of FIRMS when performing separation via LC. A Shimadzu HPLC (LC-20AD, Shimadzu Corp., Kyoto, Japan) was coupled to the AB Sciex Q-Trap 5500 MS. MRM parameters, as described in Section 5.3, were used. Caffeine, acetylsalicylic acid and, DIMPA were analyzed via LC-MS-MS. MRM transitions are listed in Tables 5.3 - 5.6.

**Table 5.3.** LCMSMS MRM transitions for caffeine including declustering potential (DP) and collision energy (CE) measured in volts (V).

| Precursor Ion | Product Ion | DP(V) | CE(V) |
|---|---|---|---|
| 194.6 | 69.1 | 178.97 | 36.29 |
| 194.6 | 83.1 | 208.00 | 37.50 |
| 194.6 | 110.1 | 168.12 | 33.58 |
| 194.6 | 138.1 | 49.910 | 33.30 |
| 195.6 | 69.1 | 166.05 | 26.59 |
| 195.6 | 70.1 | 88.870 | 44.46 |
| 195.6 | 83.0 | 114.51 | 42.85 |
| 195.6 | 84.1 | 111.51 | 57.76 |
| 195.6 | 110.1 | 70.930 | 32.21 |
| 195.6 | 111.1 | 122.97 | 43.45 |
| 195.6 | 138.1 | 27.190 | 26.33 |
| 195.6 | 139.1 | 100.56 | 26.69 |

**Table 5.4.** LCMSMS MRM transitions for acetylsalicylic acid including declustering potential (DP) and collision energy (CE) measured in volts (V).

| Precursor Ion | Product Ion | DP(V) | CE(V) |
|---|---|---|---|
| 179.0 | 59.0 | -57.72 | -14.22 |
| 179.0 | 93.0 | -40.49 | -31.42 |
| 180.0 | 59.0 | -52.77 | -39.56 |
| 180.0 | 60.0 | -69.90 | -30.03 |
| 180.0 | 93.0 | -20.25 | -32.03 |
| 180.0 | 94.0 | -28.66 | -31.44 |

The LC separation was carried out using a C8 (ZORBAX Eclipse XDB-C8, 5 µM 4.6 x 150 mm) reversed-phase column (Agilent, USA) with analytical guard column C8 (ZORBAX 4.6 x12.5 mm 5-Micron). The analytes were eluted isocratically at a flow rate of 0.25 mL/min with 25.0%, 25.0%, 50.0% and 0.1% (v/v) isopropyl alcohol, acetonitrile, water and formic acid[204] respectively. The DIMPA LC method generally followed the method of Appel et al.[205] using transitions in Table 5.5 shown below. The isotope abundance data were extracted by analyzing chromatographic peaks between time intervals

of peak elution above noise and summing the abundances within the time range. The abundances were analyzed using FIRMS software in order to calculate isotope ratio values.

**Table 5.5**. LCMSMS MRM transitions for DIMPA originating from France including declustering potential (DP) and collision energy (CE) measured in volts (V).

| Precursor Ion | Product Ion | DP(V) | CE(V) |
|---|---|---|---|
| 181.1 | 79.0 | 36.65 | 46.18 |
| 181.1 | 97.0 | 67.71 | 25.35 |
| 182.1 | 79.0 | 43.51 | 44.40 |
| 182.1 | 80.0 | 77.66 | 44.03 |
| 182.1 | 97.0 | 35.70 | 20.60 |
| 182.1 | 98.0 | 43.27 | 18.48 |

**Table 5.6.** LCMSMS MRM transition for DIMPA originating from Great Britain including declustering potential (DP) and collision energy (CE) measured in volts (V).

| Precursor Ion | Product Ion | DP(V) | CE(V) |
|---|---|---|---|
| 181.1 | 79.0 | 30.46 | 47.00 |
| 181.1 | 97.0 | 111.24 | 18.73 |
| 182.1 | 79.7 | 112.78 | 50.05 |
| 182.1 | 80.0 | 86.48 | 56.43 |
| 182.1 | 96.9 | 4.600 | 12.62 |
| 182.1 | 97.9 | 4.840 | 11.55 |

### 5.5 IRMS analysis

Pure compounds of interest (Table 5.1) were analyzed via IRMS. Solid compounds were ground through a 2-mm screen with a centrifugal grinding mill (Retsch ZM-1, Brinkmann Instrument Co., Westbury, NY, USA). Samples (3 mg each) were analyzed for total C and $^{13}C$ isotope abundance ($^{13}C$ relative to total $^{12}C + ^{13}C$) on a 20-20 Europa isotope ratio mass spectrometer (Europa Scientific Ltd., Crewe, Cheshire, UK). Compounds which are liquid at room temperature were mixed with 3 mg Chromosorb W and analyzed for total C and $^{13}C$ isotope abundance. A standard sample was evaluated after every eight

samples. The standard used for solid samples was bleached all-purpose wheat flower purchased from a local market source with standard values previously verified through multiple testing labs. The standard used for liquid samples, Fish tissue, was obtained from the Natural Resource Management department at South Dakota State University, and was previously verified through multiple testing labs.

**5.6 FIRMS software**

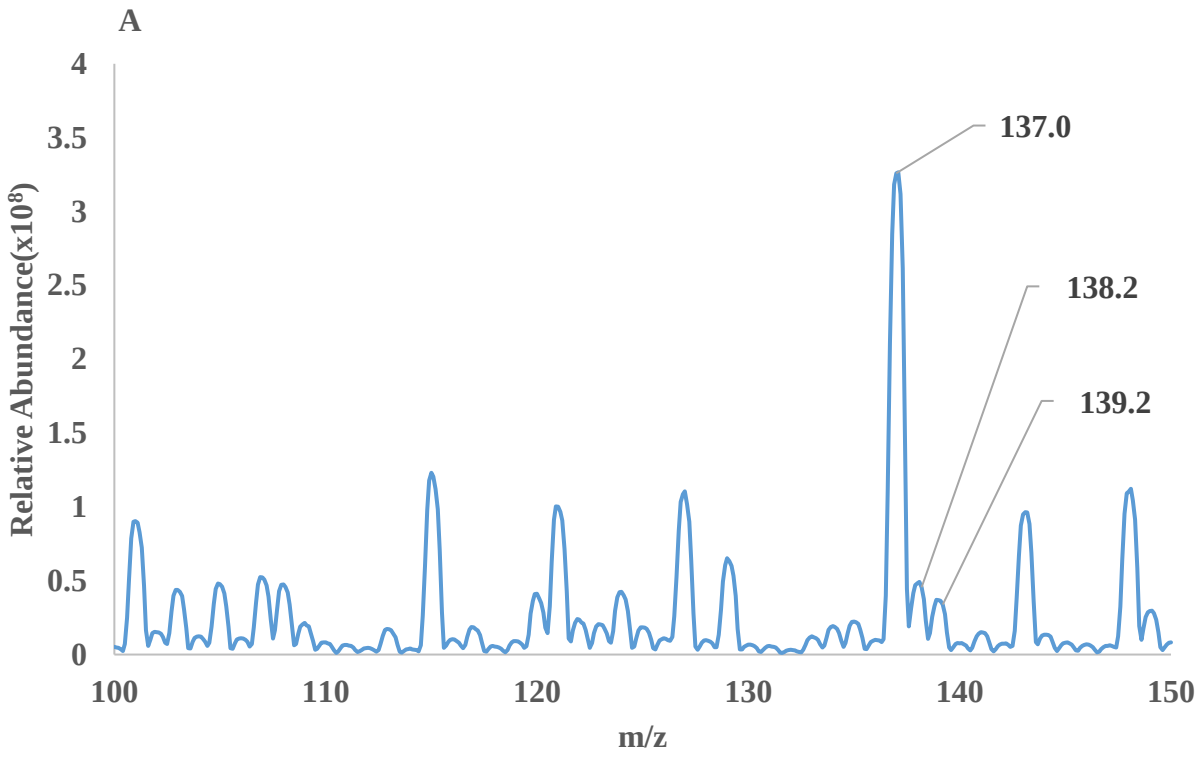
Genetic Algorithm (GA), Tabu Search (TS) and Scatter Search (SS) algorithms were acquired from Frontline Systems Inc. (Incline Village, NV USA) and used to minimize the objective function (i.e., differences between the calculated and experimental μ values (Equation 21) produced from the tandem MS, LC-MS-MS and GC-MS-MS data). The minimization was performed twenty times on the data collected for each day. For each data set, the $\delta^{13}C$ was calculated by averaging twenty calculated $\delta^{13}C$ values produced by FIRMS algorithms. A $\delta^{13}C$ value was only included on the average when the EA algorithm produced a "converged" solution and the objective function value was lower than the original objective value function. The average $\delta^{13}C$ value was calculated from isotope abundance of all five data sets. The isotope abundances for the multi-objective optimization (MOO) problems were reported as $\partial^{13}C$ measured in ‰ (parts per mill).

# Chapter 6: Results and Discussion

## 6.1 Mass Spectrometry

### Mass Spectra and Molecular Ion Fragmentation for FIRMS

Multiple compounds (N = 17) were evaluated both by FIRMS and IRMS to quantify the performance of FIRMS compared to the current "gold standard" technique for measuring isotope ratios (i.e., IRMS). The relative abundances of tandem MS product ions were analyzed using the FIRMS software. Selected MS spectra of these compounds, including IMPA and p-Toluenesulfonic acid monohydrate and their identified fragments are discussed below. They include precursor ion and product ion scans for molecular ions of mass M and M+1.



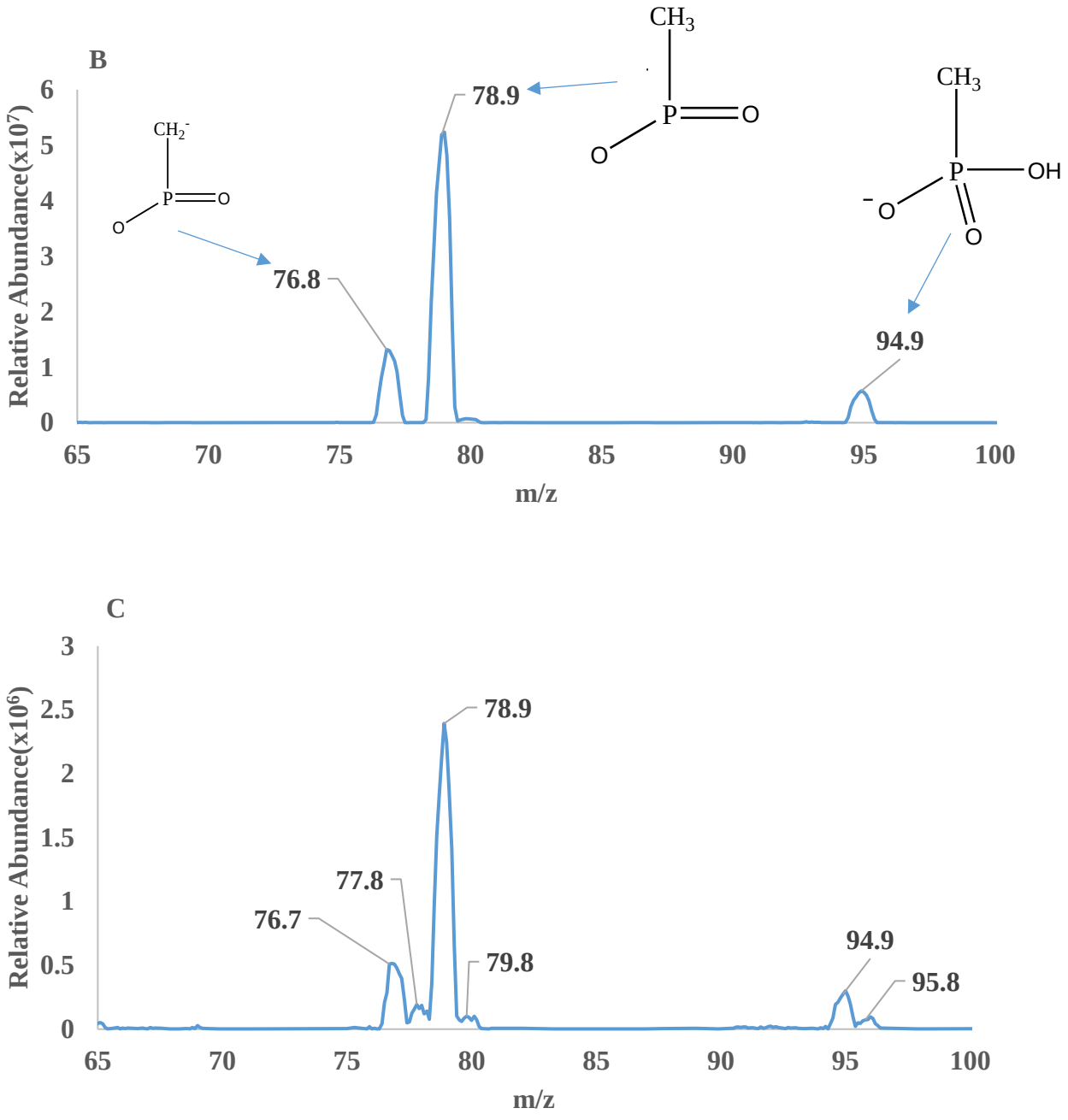
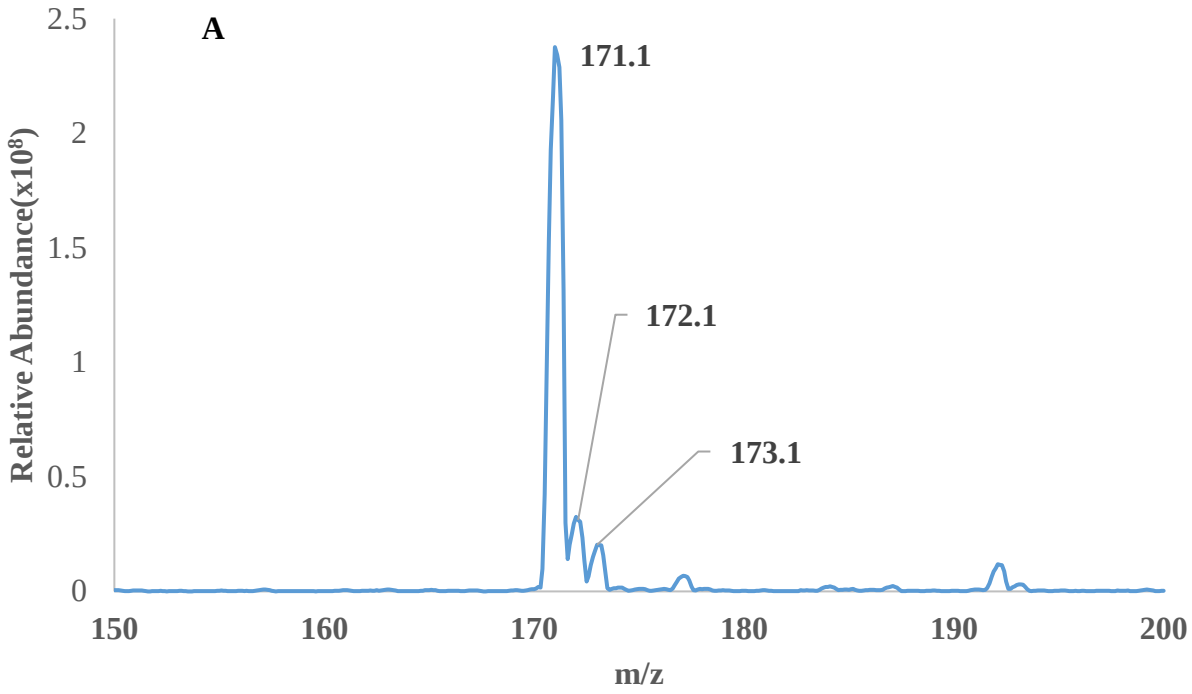
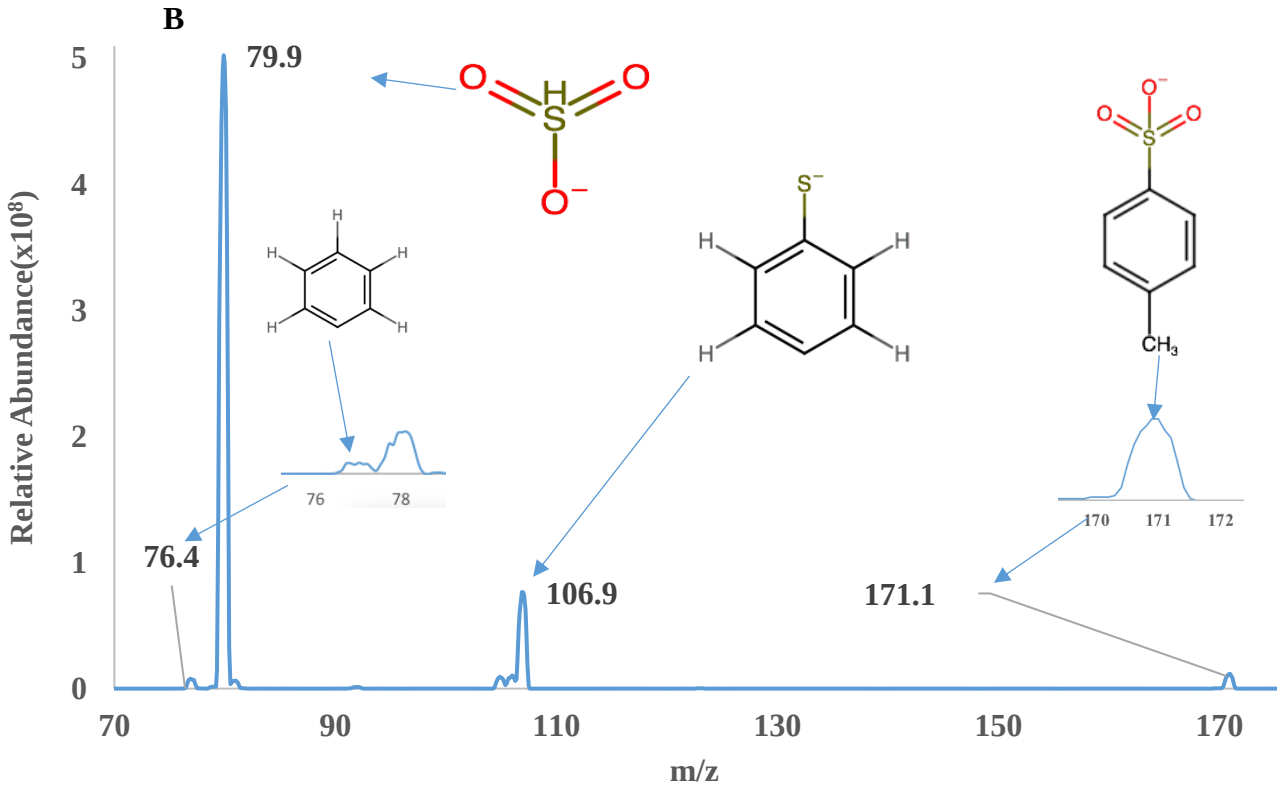



**Figure 6.1**. Simplified tandem mass spectral data for isopropyl methylphosphonic acid (IMPA). (A) Precursor ion scan (Q1- scan) of IMPA showing the molecular ion. (B) Product ion scan of m/z = 137. (C) Product ion scan of m/z = 138.1. The differences between the product ion scan of the base mass (m/z = 137) and the heavier stable isotope (m/z = 138.1) can be attributed to the isotope ratios of the atoms involved.


A
171.1
172.1
173.1
Relative Abundance(x10⁸)
2.5
2
1.5
1
0.5
0
150
160
170
180
190
200
m/z


B
79.9
76.4
106.9
171.1
Relative Abundance(x10⁸)
5
4
3
2
1
0
70
90
110
130
150
170
m/z
76
78
170
171
172
CH₃

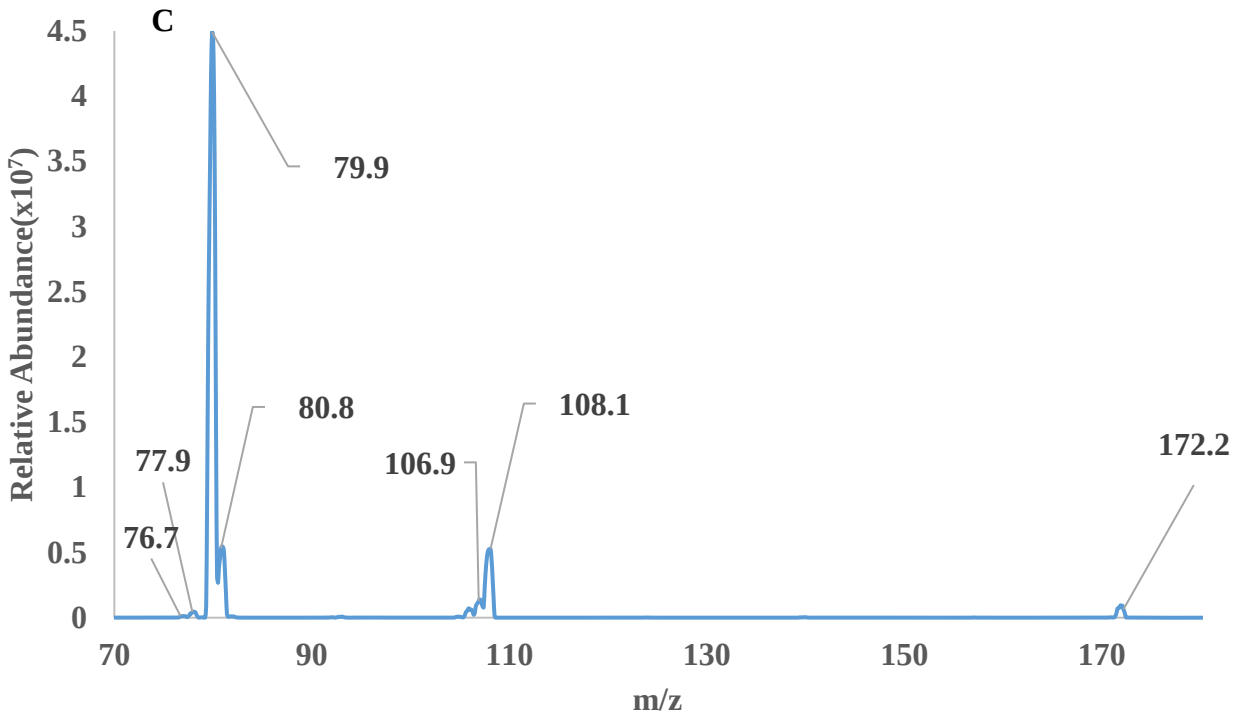



**Figure 6.2**. Simplified tandem mass spectral data for p-Toluenesulfonic acid monohydrate. (A) Precursor ion scan (Q1-scan) of p-Toluenesulfonic acid monohydrate. (B) Product ion scan of m/z = 171.1. (C) Product ion scan of m/z = 172.1. The differences between the product ion scan of the base mass (m/z = 171.1) and the heavier stable isotope (m/z = 172.1) can be attributed to the isotope ratios of the atoms involved.

The fragmentation of precursor ions during FIRMS analysis was achieved through collision-induced dissociation (CID) in the MS after initial ionization of precursor ions. These ions were either negatively charged or positively charged. For example, IMPA, EMPA and acetaminophen have molecular weights of 138.103 g/mol, 124.076 g/mol and 151.165 g/mol, respectively. Their molecular formulas are $C_4H_{11}O_3P$, $C_3H_9O_3P$, and $C_8H_9NO_2$, respectively. However, deprotonation occurs during ionization or before ionization, thus forming precursor ions of 137.0, 123.0 and 150.2 m/z with formula of $C_4H_{10}O_3P^-$ and $C_3H_8O_3P^-$, and $C_8H_8NO_2^-$, respectively. DIMPA has a molecular weight of 180.184g/mol, and molecular formula of and $C_7H_{17}O_3P$. It forms a precursor ion of 181.1

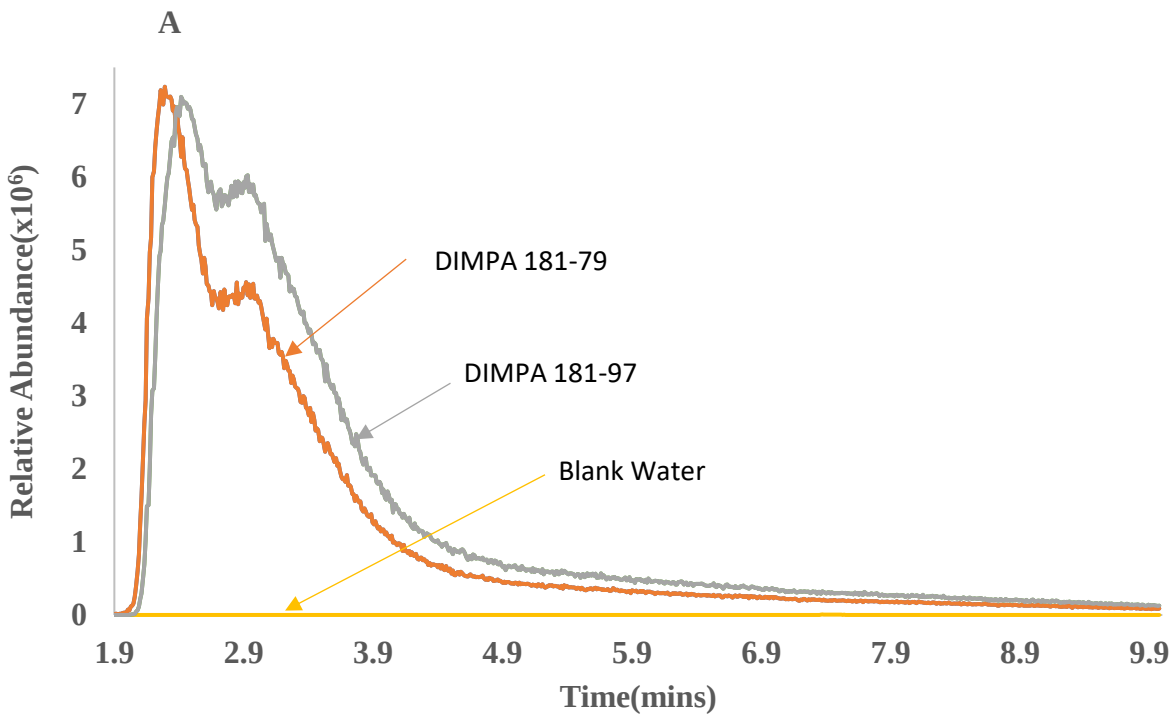
m/z which means that DIMPA undergoes protonation to form a molecular ion $C_7H_{18}O_3P^+$. Other compounds, such as L-histidine chloride, and p-Toluenesulfonic acid monohydrate lost larger fragments during precursor ion filtration, losing chlorine and water, respectively.

Some of the products ions, such as caffeine ions, were produced by rearrangement of molecular ions. The rearrangements of compounds generally included conversion of single bonds to double or triple bonds, conversion of straight chain carbons to a ring or opening of a ring to form straight chain carbons.

Given that the abundances correlate with isotope ratios and isotope frequencies as mentioned in Chapter 3, the MS data were used to calculate the unknown isotope values shown in Section 6.4. The FIRMS software was used to calculate the isotope abundance values $\delta^{13}C$ of compounds measured in parts per mil based on complex mathematical formulation and numerical minimization. As shown in Section 6.3, FIRMS data were accurate with excellent precision and accuracy (measured as an absolute error), < 1‰ and < 1.6‰ compared to IRMS values, respectively. This precision is within the generally accepted isotope abundance precision range of ≤ 1‰ for $\delta^{13}C$.[117] The $R^2$ (coefficient of determination) shown by a bivariate plot of IRMS versus FIRMS technique of the compounds was 0.992 with a slope of 1, as shown in Figure 6.6. FIRMS showed both good precision and accuracy versus the "gold standard" of IRMS, validating it as an accurate method for measurement of isotope ratios.

## 6.2 FIRMS LCMSMS

Three compounds, acetylsalicylic acid, caffeine and DIMPA were analyzed using liquid chromatography. The data were collected over a period of 5 days in order to evaluate day-to-day variability of the isotope ratios calculated by FIRMS. Figures 6.3, 6.4, and 6.5 show single day triplicate liquid chromatograms of DIMPA, acetylsalicylic acid and caffeine respectively.



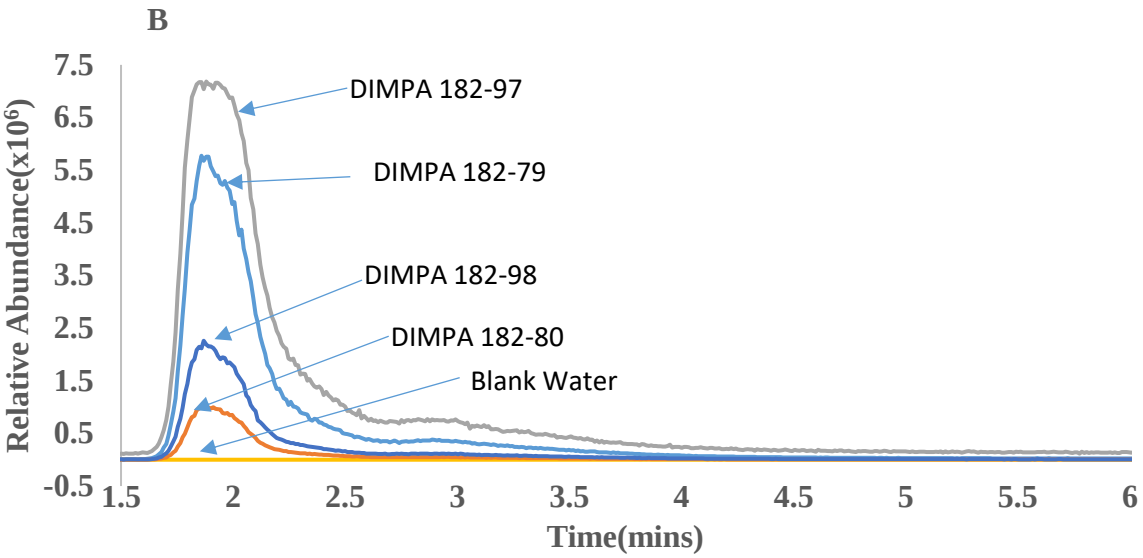
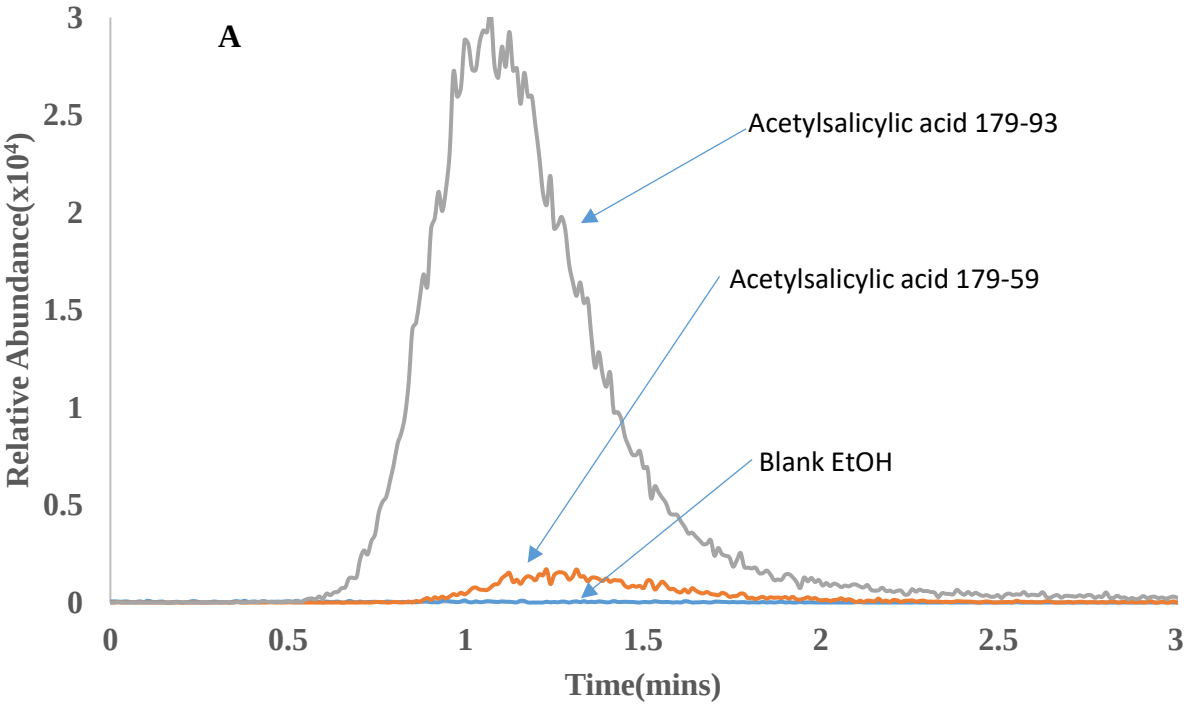



**Figure 6.3**. A) Liquid chromatograms of DIMPA originally from France for 181 m/z transitions. B) DIMPA chromatograms for 182 m/z transitions.



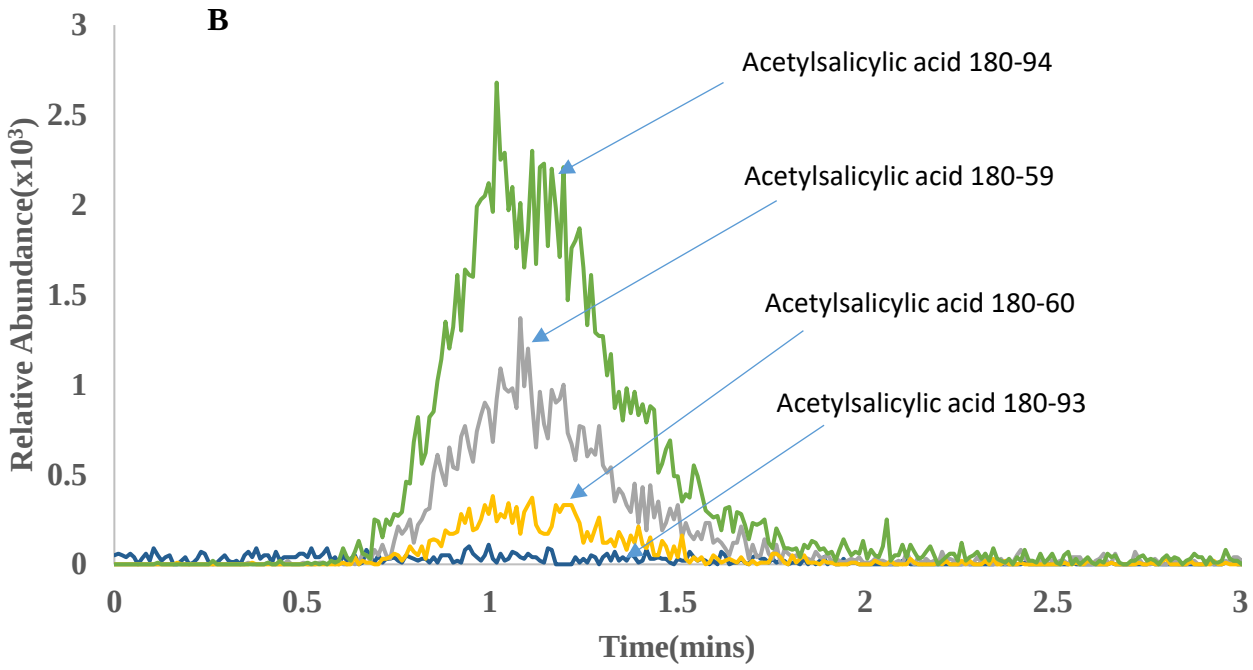
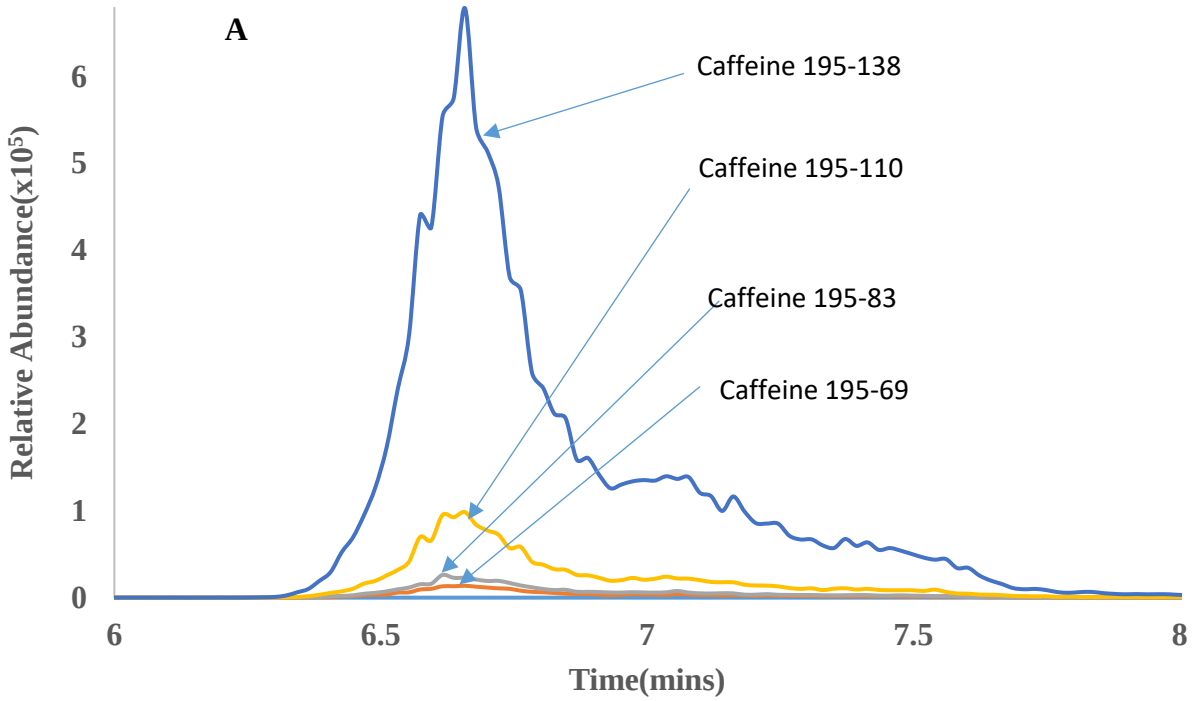



**Figure 6.4**. A) Liquid chromatograms of acetylsalicylic acid for 179 m/z transitions. B) Acetylsalicylic acid chromatograms for 180 m/z transitions.



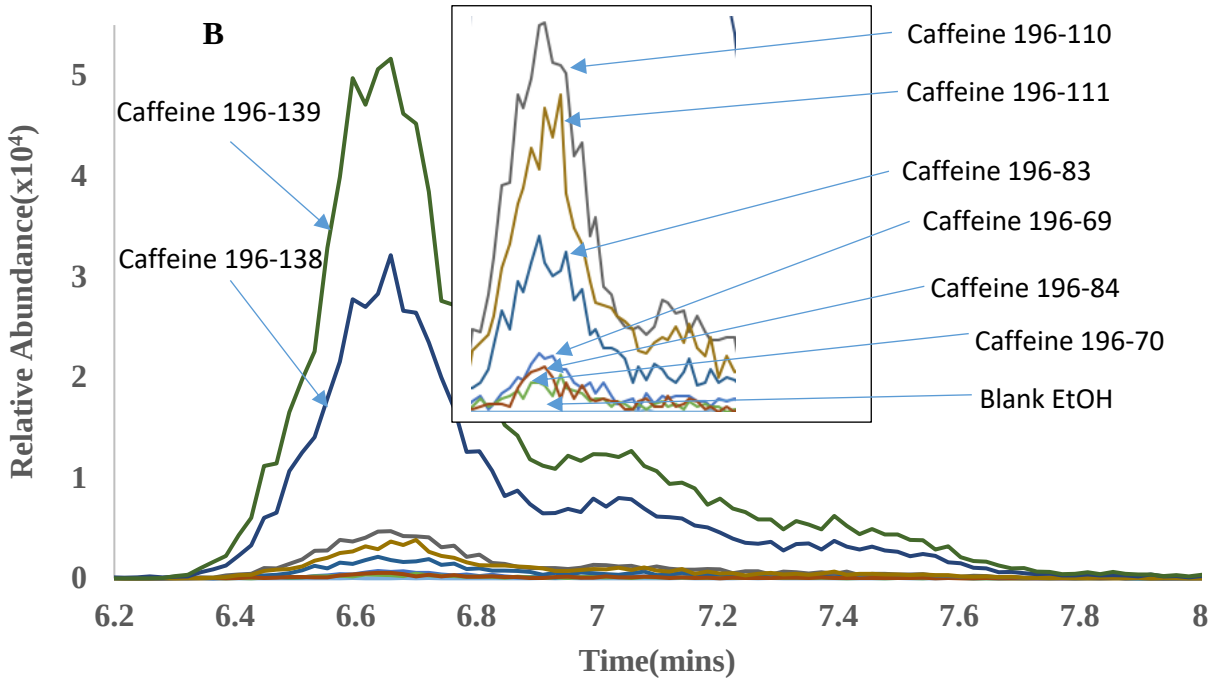



**Figure 6.5**. A) Liquid chromatograms of caffeine for 195 m/z transitions. B) Caffeine chromatograms for 180 m/z transitions.

The use of chromatography to separate complex samples prior to MS-MS analysis open up FIRMS to a much wider application space. Although FIRMS-LCMSMS is theoretically possible, the typical LC peak only allows data collected for a short period of time, compared to 10 minutes of DLSI method. Therefore, the error associated with LCMSMS may be larger than FIRMS DLSI-MSMS. For example, LC method for caffeine involved analysis of peak abundance between 6.5 and 7.5 minutes. This was compared to caffeine abundances data acquired via DLSI method for 10 minutes.

The DIMPA, acetylsalicylic acid and caffeine chromatograms used for FIRMS analysis were non-smooth and lack sharpness required for chromatographic research work. There was also presence of tailing observed. In general, this would be a concern, but for FIRMS we were more interested in the mean abundances of the peaks than the

quality of chromatograms. Thus, it was concluded that acetylsalicylic acid, caffeine and DIMPA chromatograms provided the required data despite the peak shapes. The isotope abundance values calculated using LC data were significantly accurate with good precision of ≤ 1‰, and accuracy of < 1‰ compared to IRMS values.

LC method results show that the overall accuracy was excellent compared to DLSI method of the same compounds. Although we hypothesized that we will observe lower accuracy compared to DLSI method the opposite was established with LC method. We had expected lower accuracy because of the fact that collecting data over 10 minutes gradually minimizes error by normalizing peak abundance values. Peaks analyzed over a short amount of time are likely to have larger errors. Thus, the LC method results would show slightly inaccurate data compared to IRMS and DLSI-FIRMS method. Because the precision and accuracy for $\delta^{13}C$ were ≤ 1‰, and because there was a small difference between LCMSMS-FIRMS results and IRMS, FIRMS can be coupled with LC methods to analyze a single compound or a mixture of compounds.

**6.3 Isotope ratio calculations by FIRMS**

All data collected using various sample methods, including DLSI, HPLC and GC, were analyzed using FIRMS. As shown in Table 6.1, the precision and accuracy of $\delta^{13}C$ measured by FIRMS method were < 1‰ and < 1.6‰ hence meeting the standard requirements. These precisions significantly matched IRMS values or showed better precisions in some circumstances.

FIRMS isotope abundance values were relatively accurate as compared to IRMS measurements without any standardization involved. A bivariate plot (Figure 6.6) of

FIRMS versus IRMS for a total of 17 compounds produces a coefficient of determination ($R^2$) of 0.992 and a slope of 1. The $R^2$ and slope of 1 demonstrated the validity of FIRMS.

**Table 6.1.** Results of FIRMS $\delta^{13}C$, standard deviations, accuracy (absolute error in ‰) and RSD

| # | Compound | IRMS $\delta^{13}C$ (‰) | $S_{IRMS}$ (‰) | FIRMS $\delta^{13}C$ (‰) | $S_{FIRMS}$ (‰) | Accuracy (‰) |
|---|---|---|---|---|---|---|
| 1 | Acetylsalicylic acid | -30.13 | 0.10 | -30.29 | 0.27 | -0.16 |
| 2 | Caffeine | -30.11 | 1.58 | -30.28 | 0.50 | -0.17 |
| 3 | 5-chloro-2-methylbenzoic acid[1] | -29.01 | 0.35 | -28.74 | 0.22 | +0.27 |
| 4 | 5-chloro-2-methylbenzoic acid[2] | -29.14 | 0.67 | -29.03 | 0.41 | +0.11 |
| 5 | 3,4-dibromo benzaldehyde | -27.20 | 0.23 | -27.83 | 0.65 | -0.63 |
| 6 | 2,3-dicyano hydroquinone | -36.11 | 0.07 | -36.26 | 0.73 | -0.15 |
| 7 | Diisopropyl methyl phosphonic acid[3] | -31.80 | 0.33 | -31.40 | 0.68 | +0.41 |
| 8 | Diisopropyl methyl phosphonic acid[4] | -31.06 | 0.07 | -31.35 | 0.59 | -0.30 |
| 9 | Ethyl methyl phosphonic acid | -33.57 | 0.10 | -32.80 | 0.27 | +0.77 |
| 10 | Eugenol | -31.82 | 0.56 | -30.86 | 0.34 | +0.96 |
| 11 | L-Histidine | -7.483 | 0.10 | -7.240 | 0.65 | +0.24 |
| 12 | 4-nitro benzaldehyde | -30.21 | 0.19 | -30.72 | 0.57 | -0.51 |
| 13 | 3-phenylpropionic acid | -27.27 | 0.05 | -27.95 | 0.44 | -0.68 |
| 14 | O-tert-butyl-L-tyrosine | -21.46 | 0.00 | -20.87 | 0.40 | +0.59 |
| 15 | L-tyrosine | -26.02 | 0.05 | -26.62 | 0.46 | -0.60 |
| 16 | p-toluene sulfonic acid | -26.45 | 0.02 | -27.50 | 0.69 | -1.05 |
| 17 | p-toluene sulfonyl chloride | -27.76 | 0.00 | -27.76 | 0.42 | 0.00 |

[1]China,[2]India,[3]France,[4]Great Britain

**Table 6.2.** Results of FIRMS-LCMSMS $\delta^{13}C$, standard deviations, accuracy (absolute error in ‰) and RSD

| # | Compound | LC-FIRMS $\delta^{13}C$ (‰) | Accuracy (‰) | $S_{FIRMS}$ (‰) |
|---|---|---|---|---|
| 1 | Acetylsalicylic Acid | -30.06 | -0.07 | 0.64 |
| 2 | Caffeine | -29.76 | +0.35 | 0.66 |
| 7 | Diisopropyl methyl phosphonic acid[1] | -31.37 | +0.43 | 0.85 |
| 8 | Diisopropyl methyl phosphonic acid[2] | -30.30 | -0.76 | 0.15 |

[1]France, [2]Great Britain

### 6.4 Isotope Ratios Mass Spectrometry (IRMS) analysis versus FIRMS

In order to validate FIRMS method for $\delta^{13}C$ calculations, several compounds were analyzed using IRMS. The average $\delta$ values, standard deviations and %RSD of those compounds are represented in the Table 5.5 below. These results showed good precisions of ≤ 1‰ with % RSD ≤ 6%.

**Table 6.3.** Results of IRMS $\partial^{13}C$, standard deviations, and RSD

| # | Compound | IRMS $\delta^{13}C$ (‰) | $S_{IRMS}$ (‰) |
|---|---|---|---|
| 1 | Acetylsalicylic acid | -30.13 | 0.10 |
| 2 | Caffeine | -30.11 | 1.58 |
| 3 | 5-chloro-2-methylbenzoic acid[1] | -29.01 | 0.35 |
| 4 | 5-chloro-2-methylbenzoic acid[2] | -29.14 | 0.67 |
| 5 | 3,4-dibromo benzaldehyde | -27.20 | 0.23 |
| 6 | 2,3-dicyano hydroquinone | -36.11 | 0.07 |
| 7 | Diisopropyl methyl phosphonic acid[3] | -31.80 | 0.33 |
| 8 | Diisopropyl methyl phosphonic acid[4] | -31.06 | 0.07 |
| 9 | Ethyl methyl phosphonic acid | -33.57 | 0.10 |
| 10 | Eugenol | -31.82 | 0.56 |
| 11 | L-Histidine | -7.483 | 0.10 |
| 12 | 4-nitro benzaldehyde | -30.21 | 0.19 |
| 13 | 3-phenylpropionic acid | -27.27 | 0.05 |
| 14 | O-tert-butyl-L-tyrosine | -21.46 | 0.00 |
| 15 | L-tyrosine | -26.02 | 0.05 |
| 16 | p-toluene sulfonic acid | -26.45 | 0.02 |
| 17 | p-toluene sulfonyl chloride | -27.76 | 0.00 |

[1]China,[2]India,[3]France,[4]Great Britain

FIRMS showed good accuracy compared to IRMS analysis of the same compounds. Figure 6.6 shows a linear relationship between FIRMS and IRMS analysis of the same compounds. The high $R^2$ value of 0.992 shows that 99% of the variance is accounted for data falling within the regression model indicating a good fit between two independent methods. Thus, the bivariate plot of IRMS versus FIRMS validates the FIRMS as an accurate new method for isotope ratio analysis.

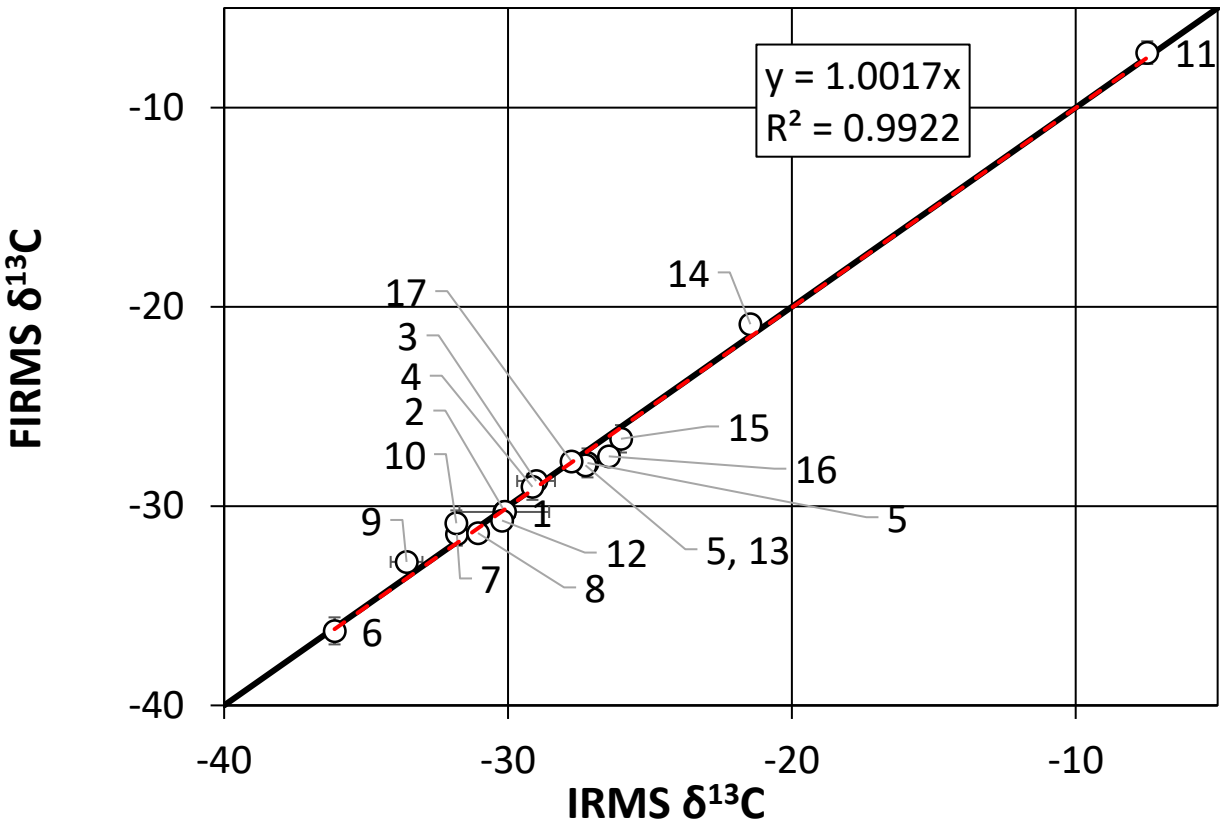



**Figure 6.6**. The bivariate plot of IRMS versus FIRMS showing the validation of the FIRMS as an accurate new method for isotope ratio analysis.

# Chapter 7: Broader impacts, Conclusion and Future Work

## 7.1. Broader impacts

FIRMS is an analytical technique that should expand IR research to any scientist with access to a tandem MS. Unlike the current approaches, where IR analysis is available to a relatively small number of facilities that can afford to purchase, staff, and maintain IR analysis instruments, the tandem MS instruments are more readily available. Moreover, FIRMS can be directly coupled to analytical separation techniques, such as LC and GC, to aid in determination of IRs of compounds from complex mixtures without the need for combustion. The disciplines that will benefit from development of FIRMS include: 1) forensic science, 2) environmental chemistry, 3) geology, 4) geochemistry, 5) cosmochemistry, 6) food sciences, 7) earth science, and 8) pharmaceutical science. FIRMS will make IR analysis more widely available and less expensive for scientists to conduct research in these fields, as well as train new graduate students for future scientific research and development in this area.

## 7.2. Conclusion

FIRMS was developed as a novel laboratory based analytical technique for the analysis of $\delta^{13}C$ and $\delta^{15}N$ isotope ratios with the potential for simultaneously analyzing other isotopes, such as $\delta^{2}H$, $\delta^{18}O$, $\delta^{34}S$, $\delta^{81}Br$, and $\delta^{37}Cl$, with further development. This technique also showed the ability to accurately calculate IRs when coupled with LC. FIRMS showed excellent precision and accuracy compared to IRMS, validating FIRMS as an accurate isotope ratio measurement technique. FIRMS is relatively economical, fast, able to evaluate IRs with a small amount of sample (i.e., μg), does not lead to loss of

structural information for organic compounds, is self-standardizing, and can simultaneously evaluate IRs for a mixture of compounds. Moreover, FIRMS uses MS-MS technology, which has the ability to detect nM concentrations and is more readily available than IRMS.

### 7.3 Future work

FIRMS is still a work in progress that requires more data collection and improved modeling. Future research should include analysis of isotope ratios of mixtures separated by LC and GC. It should also involve analysis of FIRMS functions as a multi-objective optimization problem rather than single objective optimization problem. FIRMS's full potential will only be realized with extension to the simultaneous analysis of $\delta^{2}H$ and other isotopes, such as $\delta^{2}H$, $\delta^{15}N$, $\delta^{18}O$, $\delta^{34}S$, $\delta^{81}Br$, and $\delta^{37}Cl$. Thus, future work should expand FIRMS analysis of other isotopes. Eventually, a comprehensive study of FIRMS applications to other disciplines including: 1) forensic science, 2) environmental chemistry, 3) geology, 4) geochemistry, 5) cosmochemistry, 6) food sciences, 7) earth science, and 8) pharmaceutical science should also be conducted.

## Appendix A

All MRM method tables for all compounds analyzed by FIRMS including the one that were not validated using IRMS.

**Table 1**. MRM method for p-toluene sulfonyl chloride including declustering potential (DP) and collision energy (CE) measured in volts (V).

| Precursor Ion | Product Ion | DP(V) | CE(V) |
|---|---|---|---|
| 190.2 | 65.0 | 64.75 | 68.21 |
| 190.2 | 77.1 | 238.4 | 87.44 |
| 190.2 | 91.1 | 277.9 | 36.54 |
| 190.2 | 144.2 | 264.36 | 27.52 |
| 190.2 | 160.1 | 226.51 | 38.77 |
| 191.1 | 65.0 | 14.73 | 55.56 |
| 191.1 | 65.9 | 74.16 | 60.20 |
| 191.1 | 77.1 | 79.53 | 61.73 |
| 191.1 | 78.1 | 167.29 | 70.66 |
| 191.1 | 143.9 | 228.52 | 24.98 |
| 191.1 | 145.1 | 195.97 | 31.54 |
| 191.1 | 160.1 | 48.44 | 20.22 |
| 191.1 | 161.1 | 210.34 | 12.72 |

**Table 2**. MRM method for Eugenol including declustering potential (DP) and collision energy (CE) measured in volts (V).

| Precursor Ion | Product Ion | DP(V) | CE(V) |
|---|---|---|---|
| 164.2 | 78.0 | 90.35 | 35.22 |
| 164.2 | 91.0 | 150.4 | 31.89 |
| 164.2 | 94.0 | 159.18 | 47.68 |
| 164.2 | 103.1 | 163.4 | 29.68 |
| 164.2 | 131.1 | 175.03 | 22.19 |
| 165.2 | 78.0 | 156.86 | 42.65 |
| 165.2 | 79.0 | 154.41 | 32.48 |
| 165.2 | 91.1 | 159.8 | 38.25 |
| 165.2 | 92.1 | 154.39 | 33.81 |
| 165.2 | 94.1 | 152.32 | 38.85 |
| 165.2 | 95.0 | 157.46 | 30.27 |
| 165.2 | 103.1 | 158.65 | 35.35 |
| 165.2 | 104.1 | 174.72 | 29.29 |
| 165.2 | 131.1 | 152.13 | 32.38 |
| 165.2 | 132.1 | 153.34 | 21.12 |

**Table 3.** MRM method for o-tert-butyl-l-tyrosine including declustering potential (DP) and collision energy (CE) measured in volts (V).

| Precursor Ion | Product Ion | DP(V) | CE(V) |
|---|---|---|---|
| 238.3 | 57.0 | 66.78 | 51.44 |
| 238.3 | 91.1 | 60.17 | 61.39 |
| 238.3 | 119.1 | 27.90 | 37.69 |
| 238.3 | 123.1 | 103.43 | 35.26 |
| 238.3 | 136.1 | 62.84 | 22.90 |
| 238.3 | 165.0 | 50.47 | 16.69 |
| 239.2 | 57.2 | 144.41 | 52.81 |
| 239.2 | 58.1 | 61.38 | 75.01 |
| 239.2 | 91.1 | 51.08 | 74.89 |
| 239.2 | 92.0 | 13.14 | 44.58 |
| 239.2 | 119.2 | 92.64 | 42.61 |
| 239.2 | 120.1 | 43.14 | 38.17 |
| 239.2 | 123.0 | 79.30 | 30.11 |
| 239.2 | 124.1 | 104.57 | 37.83 |
| 239.2 | 136.1 | 67.84 | 22.69 |
| 239.2 | 137.1 | 43.15 | 21.15 |
| 239.2 | 165.1 | 76.93 | 19.38 |
| 239.2 | 166.0 | 58.06 | 18.73 |

**Table 4.** MRM method for L-tyrosine including declustering potential (DP) and collision energy (CE) measured in volts (V).

| Precursor Ion | Product Ion | DP(V) | CE(V) |
|---|---|---|---|
| 182.1 | 91 | 35.56 | 40.22 |
| 182.1 | 125.8 | 157.37 | 24.64 |
| 182.1 | 153.7 | 115.37 | 7.98 |
| 182.1 | 164.9 | 54.84 | 13.71 |
| 183.1 | 91.0 | 66.62 | 40.45 |
| 183.1 | 92.1 | 37.69 | 41.74 |
| 183.1 | 125.9 | 14.19 | 9.49 |
| 183.1 | 126.9 | 87.78 | 9.25 |
| 183.1 | 153.8 | 178.88 | 63.46 |
| 183.1 | 155.2 | 177.09 | 13.34 |
| 183.1 | 165.1 | 81.86 | 14.03 |
| 183.1 | 166.0 | 98.39 | 14.26 |

**Table 5.** MRM method for 3-phenyl propionic acid including declustering potential (DP) and collision energy (CE) measured in volts (V).

| Precursor Ion | Product Ion | DP(V) | CE(V) |
|---|---|---|---|
| 150.0 | 91.1 | 83.56 | 31.89 |
| 151.2 | 91.1 | 60.23 | 26.07 |
| 151.2 | 92.1 | 118.09 | 32.53 |

**Table 6.** MRM method for p-toluene sulfonic acid including declustering potential (DP) and collision energy (CE) measured in volts (V).

| Precursor Ion | Product Ion | DP(V) | CE(V) |
|---|---|---|---|
| 171.0 | 77.1 | -57.80 | -47.07 |
| 171.0 | 79.9 | -41.81 | -25.87 |
| 171.0 | 106.9 | -37.64 | -26.59 |
| 172.0 | 77.1 | -107.68 | -37.19 |
| 172.0 | 78.0 | -110.78 | -31.93 |
| 172.0 | 79.9 | -100.41 | -42.31 |
| 172.0 | 80.9 | -84.30 | -39.86 |
| 172.0 | 107.0 | -13.53 | -30.07 |
| 172.0 | 108.1 | -81.30 | -27.07 |

**Table 7.** MRM method for 3, 4-dibromo benzaldehyde including declustering potential (DP) and collision energy (CE) measured in volts (V).

| Precursor Ion | Product Ion | DP(V) | CE(V) |
|---|---|---|---|
| 263.1 | 75.0 | 260.61 | 65.67 |
| 263.1 | 79.1 | 166.37 | 73.65 |
| 263.1 | 105.2 | 205.72 | 60.62 |
| 263.1 | 156.0 | 214.05 | 41.60 |
| 263.1 | 234.9 | 223.54 | 36.82 |
| 264.2 | 75.1 | 269.37 | 80.79 |
| 264.2 | 76.0 | 270.78 | 75.59 |
| 264.2 | 105.1 | 4.81 | 54.28 |
| 264.2 | 106.1 | 67.83 | 61.81 |
| 264.2 | 156.0 | 226.68 | 46.85 |
| 264.2 | 157.0 | 249.22 | 49.46 |
| 264.2 | 234.9 | 223.81 | 38.67 |
| 264.2 | 235.9 | 208.33 | 32.92 |

**Table 8.** MRM method for 2, 3-dicyano hydroquinone including declustering potential (DP) and collision energy (CE) measured in volts (V).

| Precursor Ion | Product Ion | DP(V) | CE(V) |
|---|---|---|---|
| 159.7 | 77.0 | -139.91 | -29.44 |
| 159.7 | 105.0 | -137.31 | -31.43 |
| 159.7 | 131.0 | -147.10 | -33.15 |
| 161.2 | 77.0 | -138.49 | -28.05 |
| 161.2 | 77.9 | -131.04 | -24.90 |
| 161.2 | 105.1 | -158.6 | -17.12 |
| 161.2 | 106.1 | -145.99 | -33.40 |
| 161.2 | 131.0 | -130.58 | -37.69 |
| 161.2 | 132.0 | -133.07 | -32.73 |

**Table 9.** MRM method for 4 – nitro benzaldehyde including declustering potential (DP) and collision energy (CE) measured in volts (V).

| Precursor Ion | Product Ion | DP(V) | CE(V) |
|---|---|---|---|
| 150.0 | 76.0 | -139.58 | -32.08 |
| 150.0 | 92.1 | -33.54 | -17.34 |
| 150.0 | 121.9 | -36.50 | -6.25 |
| 151.1 | 76.1 | -131.08 | -19.05 |
| 151.1 | 77.2 | -24.90 | -25.80 |
| 151.1 | 92.1 | -50.54 | -33.75 |
| 151.1 | 93.1 | -34.80 | -29.31 |
| 151.1 | 121.9 | -61.36 | -16.36 |
| 151.1 | 122.8 | -30.09 | -15.94 |

**Table 10.** MRM method for 5-chloro-2methyl benzoic acid originating from China including declustering potential (DP) and collision energy (CE) measured in volts (V).

| Precursor Ion | Product Ion | DP(V) | CE(V) |
|---|---|---|---|
| 171.1 | 91.0 | 23.68 | 26.07 |
| 171.1 | 125.1 | 91.38 | 34.26 |
| 171.1 | 141.3 | 254.36 | 52.34 |
| 171.1 | 153.1 | 21.33 | 26.80 |
| 172.1 | 91.0 | 235.06 | 48.73 |
| 172.1 | 92.1 | 69.39 | 34.71 |
| 172.1 | 125.1 | 98.35 | 19.10 |
| 172.1 | 125.9 | 51.98 | 26.49 |
| 172.1 | 141.0 | 276.6 | 86.63 |
| 172.1 | 141.9 | 224.73 | 64.70 |
| 172.1 | 153.3 | 139.26 | 27.33 |
| 172.1 | 154.1 | 76.43 | 16.69 |

**Table 11.** MRM method for 5-chloro-2methyl benzoic acid originating from India including declustering potential (DP) and collision energy (CE) measured in volts (V).

| Precursor Ion | Product Ion | DP(V) | CE(V) |
|---|---|---|---|
| 171.2 | 91.0 | 77.31 | 32.25 |
| 171.2 | 125.1 | 124.5 | 35.06 |
| 171.2 | 141.2 | 253.75 | 59.62 |
| 171.2 | 153.1 | 23.65 | 23.39 |
| 172.2 | 91.1 | 138.67 | 32.68 |
| 172.2 | 92.0 | 105.51 | 31.64 |
| 172.2 | 125.0 | 18.12 | 27.13 |
| 172.2 | 126.0 | 23.81 | 52.30 |
| 172.2 | 141.1 | 250.6 | 104.16 |
| 172.2 | 142.1 | 241.26 | 49,97 |
| 172.2 | 153.3 | 66.05 | 60.11 |
| 172.2 | 154.1 | 78.98 | 19.30 |

**Table 12.** MRM method for ethyl methyl phosphonic acid including declustering potential (DP) and collision energy (CE) measured in volts (V).

| Precursor Ion | Product Ion | DP(V) | CE(V) |
|---|---|---|---|
| 123.0 | 77.0 | -23.56 | -17.87 |
| 123.0 | 78.9 | -14.48 | -29.61 |
| 123.0 | 94.9 | -19.25 | -11.93 |
| 123.9 | 76.7 | -18.49 | -36.80 |
| 123.9 | 77.8 | -106.16 | -32.07 |
| 123.9 | 78.9 | -54.56 | -38.24 |
| 123.9 | 79.9 | -48.23 | -36.41 |
| 123.9 | 94.8 | -62.35 | -16.54 |
| 123.9 | 95.8 | -73.50 | -15.50 |

## References


1. Wee, C.-H.; Ta, S.-J.; Cheok, K.-H., Non-price determinants of intention to purchase counterfeit goods: An exploratory study. *International Marketing Review* **1995,** *12* (6), 19-46.
2. Norum, P. S.; Cuno, A., Analysis of the demand for counterfeit goods. *Journal of Fashion Marketing and Management: An International Journal* **2011,** *15* (1), 27-40.
3. Hoon Ang, S.; Sim Cheng, P.; Lim, E. A.; Kuan Tambyah, S., Spot the difference: consumer responses towards counterfeits. *Journal of consumer Marketing* **2001,** *18* (3), 219-235.
4. Rosman, K.; Taylor, P., Report of the IUPAC subcommittee for isotopic abundance mea surements. *Pure Appl. Chem* **1999,** *71*, 1593-1607.
5. Audi, G.; Wapstra, A., The 1995 update to the atomic mass evaluation. *Nuclear Physics A* **1995,** *595* (4), 409-480.
6. Audi, G.; Wapstra, A., The 1993 atomic mass evaluation:(I) Atomic mass table. *Nuclear Physics A* **1993,** *565* (1), 1-65.
7. Paces, J. B.; Wurster, F. C., Natural uranium and strontium isotope tracers of water sources and surface water–groundwater interactions in arid wetlands–Pahranagat Valley, Nevada, USA. *Journal of Hydrology* **2014,** *517*, 213-225.
8. Wiederhold, J. G., Metal stable isotope signatures as tracers in environmental geochemistry. *Environmental science & technology* **2015,** *49* (5), 2606-2624.
9. Kendall, C.; McDonnell, J. J., *Isotope tracers in catchment hydrology*. Elsevier: 2012.
10. Coleman, D. C., *Carbon isotope techniques*. Academic Press: 2012.
11. Muccio, Z.; Jackson, G. P., Isotope ratio mass spectrometry. *Analyst* **2009,** *134* (2), 213-222.
12. Dickin, A. P., *Radiogenic isotope geology*. Cambridge University Press: 2005.
13. Clark, I. D.; Fritz, P., *Environmental isotopes in hydrogeology*. CRC press: 1997.
14. Farquhar, G. D.; Ehleringer, J. R.; Hubick, K. T., Carbon isotope discrimination and photosynthesis. *Annual review of plant biology* **1989,** *40* (1), 503-537.
15. Masarik, J.; Beer, J., Simulation of particle fluxes and cosmogenic nuclide production in the Earth's atmosphere. *Journal of Geophysical Research: Atmospheres* **1999,** *104* (D10), 12099-12111.
16. Lal, D., In situ-produced cosmogenic isotopes in terrestrial rocks. *Annual Review of Earth and Planetary Sciences* **1988,** *16*, 355-388.
17. Webber, W.; Higbie, P.; McCracken, K., Production of the cosmogenic isotopes 3H, 7Be, 10Be, and 36Cl in the Earth's atmosphere by solar and galactic cosmic rays. *Journal of Geophysical Research: Space Physics* **2007,** *112* (A10).
18. Dunai, T. J., *Cosmogenic nuclides: principles, concepts and applications in the earth surface sciences*. Cambridge University Press: 2010.
19. Stone, J. O., Air pressure and cosmogenic isotope production. *Journal of Geophysical Research: Solid Earth* **2000,** *105* (B10), 23753-23759.

20. Stone, J.; Evans, J.; Fifield, L.; Allan, G.; Cresswell, R., Cosmogenic chlorine-36 production in calcite by muons. *Geochimica et Cosmochimica Acta* **1998,** *62* (3), 433-454.
21. Ta, T. T.; Le, S. H.; Trinh, H. Q.; Luu, T. N. M.; Trinh, A. D., Interpretation of anthropogenic impacts (agriculture and urbanization) on tropical deltaic river network through the spatio-temporal variation of stable (N, O) isotopes of NO− 3. *Isotopes in environmental and health studies* **2016**, 1-11.
22. Kendall, C.; Caldwell, E. A., Fundamentals of isotope geochemistry. *Isotope tracers in catchment hydrology* **1998**, 51-86.
23. Platzner, M.; Rinner, B.; Weiss, R., Parallel qualitative simulation. *Simulation Practice and Theory* **1997,** *5* (7), 623-638.
24. Tremblay, P.; Paquin, R., Improved detection of sugar addition to maple syrup using malic acid as internal standard and in 13C isotope ratio mass spectrometry (IRMS). *Journal of agricultural and food chemistry* **2007,** *55* (2), 197-203.
25. Benson, S.; Lennard, C.; Maynard, P.; Roux, C., Forensic applications of isotope ratio mass spectrometry—a review. *Forensic Science International* **2006,** *157* (1), 1-22.
26. Ehleringer, J. R.; Cooper, D. A.; Lott, M. J.; Cook, C. S., Geo-location of heroin and cocaine by stable isotope ratios. *Forensic Science International* **1999,** *106* (1), 27-35.
27. Price, T. D.; Burton, J. H.; Bentley, R. A., The characterization of biologically available strontium isotope ratios for the study of prehistoric migration. *Archaeometry* **2002,** *44* (1), 117-135.
28. Mayer, K.; Wallenius, M.; Ray, I., Nuclear forensics—a methodology providing clues on the origin of illicitly trafficked nuclear materials. *Analyst* **2005,** *130* (4), 433-441.
29. Drivelos, S. A.; Georgiou, C. A., Multi-element and multi-isotope-ratio analysis to determine the geographical origin of foods in the European Union. *TrAC Trends in Analytical Chemistry* **2012,** *40*, 38-51.
30. Voerkelius, S.; Lorenz, G. D.; Rummel, S.; Quétel, C. R.; Heiss, G.; Baxter, M.; Brach-Papa, C.; Deters-Itzelsberger, P.; Hoelzl, S.; Hoogewerff, J., Strontium isotopic signatures of natural mineral waters, the reference to a simple geological map and its potential for authentication of food. *Food chemistry* **2010,** *118* (4), 933-940.
31. Schellenberg, A.; Chmielus, S.; Schlicht, C.; Camin, F.; Perini, M.; Bontempo, L.; Heinrich, K.; Kelly, S. D.; Rossmann, A.; Thomas, F., Multielement stable isotope ratios (H, C, N, S) of honey from different European regions. *Food chemistry* **2010,** *121* (3), 770-777.
32. Yamada, M.; Kinoshita, K.; Kurosawa, M.; Saito, K.; Nakazawa, H., Analysis of exogenous nandrolone metabolite in horse urine by gas chromatography/combustion/carbon isotope ratio mass spectrometry. *Journal of pharmaceutical and biomedical analysis* **2007,** *45* (4), 654-658.
33. Hebestreit, M.; Flenker, U.; Fußhöller, G.; Geyer, H.; Güntner, U.; Mareck, U.; Piper, T.; Thevis, M.; Ayotte, C.; Schänzer, W., Determination of the origin of urinary norandrosterone traces by gas chromatography combustion isotope ratio mass spectrometry. *Analyst* **2006,** *131* (9), 1021-1026.

34. Saber, D.; Mauro, D.; Sirivedhin, T., Environmental forensics investigation in sediments near a former manufactured gas plant site. *Environmental Forensics* **2006,** *7* (1), 65-75.
35. Organization, W. H. Substandard and falsified medical products. http://www.who.int/mediacentre/factsheets/fs275/en/ (accessed March 2017).
36. Ratanawijitrasin, S.; Wondemagegnehu, E., Effective drug regulation. *A multicountry study. Geneva: World Health Organization* **2002**.
37. Byarugaba, D., Antimicrobial resistance in developing countries and responsible risk factors. *International journal of antimicrobial agents* **2004,** *24* (2), 105-110.
38. Akuse, R. M.; Eke, F. U.; Ademola, A. D.; Fajolu, I. B.; Gbelee, H. O.; Ihejiahi, U.; Bugaje, M. A.; Anochie, I. C.; Asinobi, A. O.; Okafor, H. U., Diagnosing renal failure due to diethylene glycol in children in a resource-constrained setting. *Pediatric Nephrology* **2012,** *27* (6), 1021-1028.
39. Control, C. f. D.; Prevention, Fatal poisoning among young children from diethylene glycol-contaminated acetaminophen---Nigeria, 2008--2009. *MMWR: Morbidity and mortality weekly report* **2009,** *58* (48), 1345-1347.
40. Bogdanich, W.; Hooker, J., From China to Panama, a trail of poisoned medicine. *New York Times* **2007,** *6*.
41. Rentz, E. D.; Lewis, L.; Mujica, O. J.; Barr, D. B.; Schier, J. G.; Weerasekera, G.; Kuklenyik, P.; McGeehin, M.; Osterloh, J.; Wamsley, J., Outbreak of acute renal failure in Panama in 2006: a case-control study. *Bulletin of the World Health Organization* **2008,** *86* (10), 749-756.
42. Stanton, C.; Koski, A.; Cofie, P.; Mirzabagi, E.; Grady, B. L.; Brooke, S., Uterotonic drug quality: an assessment of the potency of injectable uterotonic drugs purchased by simulated clients in three districts in Ghana. *BMJ open* **2012,** *2* (3), e000431.
43. Unicef, At a glance: Ghana. 2012 ed.; Unicef: Online, 2012.
44. Asamoah, B. O.; Moussa, K. M.; Stafström, M.; Musinguzi, G., Distribution of causes of maternal mortality among different socio-demographic groups in Ghana; a descriptive study. *BMC Public Health* **2011,** *11* (1), 159.
45. Ganyaglo, G. Y.; Hill, W. C. In *A 6-year (2004-2009) review of maternal mortality at the Eastern Regional Hospital, Koforidua, Ghana*, Seminars in perinatology, Elsevier: 2012; pp 79-83.
46. Cheng, M. M., Is the drugstore safe? Counterfeit diabetes products on the shelves. *Journal of diabetes science and technology* **2009,** *3* (6), 1516-1520.
47. Olusola, A. M.; Adekoya, A. I.; Olanrewaju, O. J., Comparative Evaluation of Physicochemical Properties of Some Commercially Available Brands of Metformin Hcl Tablets in Lagos, Nigeria. **2012**.
48. Unachukwu, C.; Uchenna, D.; Young, E., Mortality among Diabetes In-Patients in Port-Harcourt, Nigeria. *African Journal of Endocrinology and Metabolism* **2008,** *7* (1), 1-4.
49. Wondemagegnehu, E., Counterfeit and substandard drugs in Myanmar and Viet Nam. *WHO Report. WHO/EDM/QSM* **1999,** *99*.
50. O'Brien, K. L.; Wolfson, L. J.; Watt, J. P.; Henkle, E.; Deloria-Knoll, M.; McCall, N.; Lee, E.; Mulholland, K.; Levine, O. S.; Cherian, T., Burden of disease caused by

Streptococcus pneumoniae in children younger than 5 years: global estimates. *The Lancet* **2009,** *374* (9693), 893-902.
51. Udwadia, Z. F., MDR, XDR, TDR tuberculosis: ominous progression. BMJ Publishing Group Ltd and British Thoracic Society: 2012.
52. Jain, A.; Mondal, R., Extensively drug-resistant tuberculosis: current challenges and threats. *FEMS Immunology & Medical Microbiology* **2008,** *53* (2), 145-150.
53. Okeke, I. N.; Laxminarayan, R.; Bhutta, Z. A.; Duse, A. G.; Jenkins, P.; O'Brien, T. F.; Pablos-Mendez, A.; Klugman, K. P., Antimicrobial resistance in developing countries. Part I: recent trends and current status. *The Lancet infectious diseases* **2005,** *5* (8), 481-493.
54. Okeke, I. N.; Klugman, K. P.; Bhutta, Z. A.; Duse, A. G.; Jenkins, P.; O'Brien, T. F.; Pablos-Mendez, A.; Laxminarayan, R., Antimicrobial resistance in developing countries. Part II: strategies for containment. *The Lancet infectious diseases* **2005,** *5* (9), 568-580.
55. Parasa, L. S.; Kumar, L. C. A.; Para, S.; Atluri, V. S. R.; Santhisree, K.; Kumar, P.; Setty, C., Epidemiological survey of methicillin resistant Staphylococcus aureus in the community and hospital, Gannavaram, Andhra Pradesh, South India. *Reviews in Infection* **2010,** *1* (2), 117-123.
56. Vincent, J.-L.; Rello, J.; Marshall, J.; Silva, E.; Anzueto, A.; Martin, C. D.; Moreno, R.; Lipman, J.; Gomersall, C.; Sakr, Y., International study of the prevalence and outcomes of infection in intensive care units. *Jama* **2009,** *302* (21), 2323-2329.
57. Guzmán-Blanco, M.; Mejía, C.; Isturiz, R.; Alvarez, C.; Bavestrello, L.; Gotuzzo, E.; Labarca, J.; Luna, C. M.; Rodríguez-Noriega, E.; Salles, M. J., Epidemiology of meticillin-resistant Staphylococcus aureus (MRSA) in Latin America. *International journal of antimicrobial agents* **2009,** *34* (4), 304-308.
58. Sahoo, K. C.; Tamhankar, A. J.; Johansson, E.; Lundborg, C. S., Antibiotic use, resistance development and environmental factors: a qualitative study among healthcare professionals in Orissa, India. *BMC Public Health* **2010,** *10* (1), 629.
59. Bartoloni, A.; Gotuzzo, E., Bacterial-resistant infections in resource-limited countries. In *Antimicrobial resistance in developing countries*, Springer: 2010; pp 199-231.
60. Kaitin, K. I., Deconstructing the drug development process: the new face of innovation. *Clinical pharmacology and therapeutics* **2010,** *87* (3), 356.
61. Cameron, A.; Ewen, M.; Ross-Degnan, D.; Ball, D.; Laing, R., Medicine prices, availability, and affordability in 36 developing and middle-income countries: a secondary analysis. *The lancet* **2009,** *373* (9659), 240-249.
62. DiMasi, J. A.; Hansen, R. W.; Grabowski, H. G., The price of innovation: new estimates of drug development costs. *Journal of health economics* **2003,** *22* (2), 151-185.
63. Whitty, C. J.; Chandler, C.; Ansah, E.; Leslie, T.; Staedke, S. G., Deployment of ACT antimalarials for treatment of malaria: challenges and opportunities. *Malaria Journal* **2008,** *7* (1), S7.
64. Buckley, G. J.; Gostin, L. O., *Countering the problem of falsified and substandard drugs*. National Academies Press: 2013.
65. Fernandez, F. M.; Hostetler, D.; Powell, K.; Kaur, H.; Green, M. D.; Mildenhall, D. C.; Newton, P. N., Poor quality drugs: grand challenges in high throughput detection,

countrywide sampling, and forensics in developing countries. *Analyst* **2011,** *136* (15), 3073-3082.
66. Organization, W. H. Tool for visual inspection. http://www.whpa.org/Toolkit_BeAware_Inspection.pdf (accessed January 29).
67. Lal, R.; Ramachandran, S.; Arnsdorf, M. F., Multidimensional atomic force microscopy: a versatile novel technology for nanopharmacology research. *The AAPS journal* **2010,** *12* (4), 716-728.
68. Jähnke, R. W.; Küsters, G.; Fleischer, K., Low-cost quality assurance of medicines using the GPHF-Minilab®. *Drug information journal* **2001,** *35* (3), 941-945.
69. JÄHNKE, R. W., Counterfeit medicines and the GPHF-Minilab for rapid drug quality verification. *Pharmazeutische Industrie* **2004,** *66* (10), 1187-1193.
70. Ranieri, N.; Tabernero, P.; Green, M. D.; Verbois, L.; Herrington, J.; Sampson, E.; Satzger, R. D.; Phonlavong, C.; Thao, K.; Newton, P. N., Evaluation of a new handheld instrument for the detection of counterfeit artesunate by visual fluorescence comparison. *The American journal of tropical medicine and hygiene* **2014,** *91* (5), 920-924.
71. Green, M. D.; Hostetler, D. M.; Nettey, H.; Swamidoss, I.; Ranieri, N.; Newton, P. N., Integration of novel low-cost colorimetric, laser photometric, and visual fluorescent techniques for rapid identification of falsified medicines in resource-poor areas: application to artemether–lumefantrine. *The American journal of tropical medicine and hygiene* **2015,** *92* (6 Suppl), 8-16.
72. Barras, J.; Murnane, D.; Althoefer, K.; Assi, S.; Rowe, M. D.; Poplett, I. J.; Kyriakidou, G.; Smith, J. A., Nitrogen-14 nuclear quadrupole resonance spectroscopy: A promising analytical methodology for medicines authentication and counterfeit antimalarial analysis. *Analytical chemistry* **2013,** *85* (5), 2746-2753.
73. Balchin, E.; Malcolme-Lawes, D. J.; Poplett, I. J.; Rowe, M. D.; Smith, J. A.; Pearce, G. E.; Wren, S. A., Potential of nuclear quadrupole resonance in pharmaceutical analysis. *Analytical Chemistry* **2005,** *77* (13), 3925-3930.
74. de Veij, M.; Deneckere, A.; Vandenabeele, P.; de Kaste, D.; Moens, L., Detection of counterfeit Viagra® with Raman spectroscopy. *Journal of pharmaceutical and biomedical analysis* **2008,** *46* (2), 303-309.
75. Ricci, C.; Eliasson, C.; Macleod, N. A.; Newton, P. N.; Matousek, P.; Kazarian, S. G., Characterization of genuine and fake artesunate anti-malarial tablets using Fourier transform infrared imaging and spatially offset Raman spectroscopy through blister packs. *Analytical and bioanalytical chemistry* **2007,** *389* (5), 1525.
76. Puchert, T.; Lochmann, D.; Menezes, J.; Reich, G., Near-infrared chemical imaging (NIR-CI) for counterfeit drug identification—a four-stage concept with a novel approach of data processing (Linear Image Signature). *Journal of pharmaceutical and biomedical analysis* **2010,** *51* (1), 138-145.
77. O'Brien, N. A.; Hulse, C. A.; Friedrich, D. M.; Van Milligen, F. J.; von Gunten, M. K.; Pfeifer, F.; Siesler, H. W. In *Miniature near-infrared (NIR) spectrometer engine for handheld applications*, SPIE Defense, Security, and Sensing, International Society for Optics and Photonics: 2012; pp 837404-837404-8.

78. Deconinck, E.; Sacré, P.-Y.; Courselle, P.; De Beer, J. O., Chromatography in the detection and characterization of illegal pharmaceutical preparations. *Journal of chromatographic science* **2013,** *51* (8), 791-806.
79. Hu, C.-Q.; Zou, W.-B.; Hu, W.-S.; Ma, X.-K.; Yang, M.-Z.; Zhou, S.-L.; Sheng, J.-F.; Li, Y.; Cheng, S.-H.; Xue, J., Establishment of a Fast Chemical Identification System for screening of counterfeit drugs of macrolide antibiotics. *Journal of pharmaceutical and biomedical analysis* **2006,** *40* (1), 68-74.
80. Kovacs, S.; Hawes, S. E.; Maley, S. N.; Mosites, E.; Wong, L.; Stergachis, A., Technologies for detecting falsified and substandard drugs in low and middle-income countries. *PLoS One* **2014,** *9* (3), e90601.
81. Remaud, G. S.; Bussy, U.; Lees, M.; Thomas, F.; Desmurs, J.-R.; Jamin, E.; Silvestre, V.; Akoka, S., NMR spectrometry isotopic fingerprinting: A tool for the manufacturer for tracking Active Pharmaceutical Ingredients from starting materials to final medicines. *European Journal of Pharmaceutical Sciences* **2013,** *48* (3), 464-473.
82. Scherer, J.; Paul, J.; O'keefe, A.; Saykally, R., Cavity ringdown laser absorption spectroscopy: history, development, and application to pulsed molecular beams. *Chemical reviews* **1997,** *97* (1), 25-52.
83. Romanini, D.; Kachanov, A.; Sadeghi, N.; Stoeckel, F., CW cavity ring down spectroscopy. *Chemical Physics Letters* **1997,** *264* (3), 316-322.
84. Bowling, D. R.; Sargent, S. D.; Tanner, B. D.; Ehleringer, J. R., Tunable diode laser absorption spectroscopy for stable isotope studies of ecosystem–atmosphere CO 2 exchange. *Agricultural and forest meteorology* **2003,** *118* (1), 1-19.
85. Jimenez, R.; Herndon, S.; Shorter, J. H.; Nelson, D. D.; McManus, J. B.; Zahniser, M. S. In *Atmospheric trace gas measurements using a dual quantum-cascade laser mid-infrared absorption spectrometer*, Novel In-Plane Semiconductor Lasers IV, International Society for Optics and Photonics: 2005; pp 318-332.
86. Mohn, J.; Werner, R.; Buchmann, B.; Emmenegger, L., High-precision δ13CO2 analysis by FTIR spectroscopy using a novel calibration strategy. *Journal of molecular structure* **2007,** *834*, 95-101.
87. Tuzson, B.; Mohn, J.; Zeeman, M.; Werner, R.; Eugster, W.; Zahniser, M.; Nelson, D.; McManus, J.; Emmenegger, L., High precision and continuous field measurements of δ 13 C and δ 18 O in carbon dioxide with a cryogen-free QCLAS. *Applied Physics B* **2008,** *92* (3), 451.
88. Tuzson, B.; Zeeman, M.; Zahniser, M.; Emmenegger, L., Quantum cascade laser based spectrometer for in situ stable carbon dioxide isotope measurements. *Infrared Physics & Technology* **2008,** *51* (3), 198-206.
89. Nelson, D. D.; McManus, J. B.; Herndon, S. C.; Zahniser, M. S.; Tuzson, B.; Emmenegger, L., New method for isotopic ratio measurements of atmospheric carbon dioxide using a 4.3 μm pulsed quantum cascade laser. *Applied Physics B* **2008,** *90* (2), 301-309.
90. van Geldern, R.; Nowak, M. E.; Zimmer, M.; Szizybalski, A.; Myrttinen, A.; Barth, J. A.; Jost, H.-J. r., Field-based stable isotope analysis of carbon dioxide by mid-infrared laser spectroscopy for carbon capture and storage monitoring. *Analytical chemistry* **2014,** *86* (24), 12191-12198.

91. Maity, A.; Banik, G. D.; Ghosh, C.; Som, S.; Chaudhuri, S.; Daschakraborty, S. B.; Ghosh, S.; Ghosh, B.; Raychaudhuri, A. K.; Pradhan, M., Residual gas analyzer mass spectrometry for human breath analysis: a new tool for the non-invasive diagnosis of Helicobacter pylori infection. *Journal of breath research* **2014,** *8* (1), 016005.
92. Sakai, S.; Matsuda, S.; Hikida, T.; Shimono, A.; McManus, J. B.; Zahniser, M.; Nelson, D.; Dettman, D. L.; Yang, D.; Ohkouchi, N., High-Precision Simultaneous 18O/16O, 13C/12C, and 17O/16O Analyses for Microgram Quantities of CaCO3 by Tunable Infrared Laser Absorption Spectroscopy. *Analytical chemistry* **2017,** *89* (21), 11846-11852.
93. Pisklak, D. M.; Zielińska-Pisklak, M.; Szeleszczuk, Ł.; Wawer, I., 13 C cross-polarization magic-angle spinning nuclear magnetic resonance analysis of the solid drug forms with low concentration of an active ingredient-propranolol case. *Journal of pharmaceutical and biomedical analysis* **2014,** *93*, 68-72.
94. Önal, A.; Tekkeli, S. E. K.; Önal, C., A review of the liquid chromatographic methods for the determination of biogenic amines in foods. *Food chemistry* **2013,** *138* (1), 509-515.
95. Sanchez, S.; Ziarelli, F.; Viel, S.; Delaurent, C.; Caldarelli, S., Improved solid-state NMR quantifications of active principles in pharmaceutical formulations. *Journal of pharmaceutical and biomedical analysis* **2008,** *47* (4), 683-687.
96. Wawer, I.; Pisklak, M.; Chilmonczyk, Z., 1 H, 13 C, 15 N NMR analysis of sildenafil base and citrate (Viagra) in solution, solid state and pharmaceutical dosage forms. *Journal of pharmaceutical and biomedical analysis* **2005,** *38* (5), 865-870.
97. Zielińska-pisklak, M.; Pisklak, D. M.; Wawer, I., Application of 13C CPMAS NMR for qualitative and quantitative characterization of carvedilol and its commercial formulations. *Journal of pharmaceutical sciences* **2012,** *101* (5), 1763-1772.
98. Zhang, J.; Tozuka, Y.; Uchiyama, H.; Higashi, K.; Moribe, K.; Takeuchi, H.; Yamamoto, K., NMR investigation of a novel excipient, α-glucosylhesperidin, as a suitable solubilizing agent for poorly water-soluble drugs. *Journal of pharmaceutical sciences* **2011,** *100* (10), 4421-4431.
99. Nep, E. I.; Conway, B. R., Characterization of grewia gum, a potential pharmaceutical excipient. **2010**.
100. Sanphui, P.; Kumar, S. S.; Nangia, A., Pharmaceutical cocrystals of niclosamide. *Crystal Growth & Design* **2012,** *12* (9), 4588-4599.
101. Brus, J.; Urbanova, M.; Sedenkova, I.; Brusova, H., New perspectives of 19 F MAS NMR in the characterization of amorphous forms of atorvastatin in dosage formulations. *International journal of pharmaceutics* **2011,** *409* (1), 62-74.
102. Miura, H.; Kanebako, M.; Shirai, H.; Nakao, H.; Inagi, T.; Terada, K., Stability of amorphous drug, 2-benzyl-5-(4-chlorophenyl)-6-[4-(methylthio) phenyl]-2H-pyridazin-3-one, in silica mesopores and measurement of its molecular mobility by solid-state 13 C NMR spectroscopy. *International journal of pharmaceutics* **2011,** *410* (1), 61-67.
103. Trefi, S.; Gilard, V.; Balayssac, S.; Malet-Martino, M.; Martino, R., Quality assessment of fluoxetine and fluvoxamine pharmaceutical formulations purchased in different countries or via the Internet by 19 F and 2D DOSY 1 H NMR. *Journal of pharmaceutical and biomedical analysis* **2008,** *46* (4), 707-722.

104. Brettmann, B.; Bell, E.; Myerson, A.; Trout, B., Solid-state NMR characterization of high-loading solid solutions of API and excipients formed by electrospinning. *Journal of pharmaceutical sciences* **2012,** *101* (4), 1538-1545.
105. Silvestre, V.; Mboula, V. M.; Jouitteau, C.; Akoka, S.; Robins, R. J.; Remaud, G. S., Isotopic 13 C NMR spectrometry to assess counterfeiting of active pharmaceutical ingredients: site-specific 13 C content of aspirin and paracetamol. *Journal of pharmaceutical and biomedical analysis* **2009,** *50* (3), 336-341.
106. Gilbert, A.; Silvestre, V.; Robins, R. J.; Remaud, G. r. S., Accurate quantitative isotopic 13C NMR spectroscopy for the determination of the intramolecular distribution of 13C in glucose at natural abundance. *Analytical chemistry* **2009,** *81* (21), 8978-8985.
107. Thibaudeau, C.; Remaud, G. r.; Silvestre, V.; Akoka, S., Performance evaluation of quantitative adiabatic 13C NMR pulse sequences for site-specific isotopic measurements. *Analytical chemistry* **2010,** *82* (13), 5582-5590.
108. Fruit, A., fruit products, Official Method 995.17: Beet Sugar in Fruit juices. *AOAC Official Methods of Analysis* **1996**.
109. Martin, G.; Martin, M., The site-specific natural isotope fractionation-NMR method applied to the study of wines. In *Wine Analysis*, Springer: 1988; pp 258-275.
110. Martin, G. J.; Akoka, S.; Martin, M. L., SNIF-NMR—Part 1: Principles. In *Modern Magnetic Resonance*, Springer: 2008; pp 1651-1658.
111. Martin, G.; Martin, M., Deuterium labelling at the natural abundance level as studied by high field quantitative 2H NMR. *Tetrahedron letters* **1981,** *22* (36), 3525-3528.
112. Robins, R. J.; Billault, I.; Duan, J.-R.; Guiet, S.; Pionnier, S.; Zhang, B.-L., Measurement of 2 H distribution in natural products by quantitative 2 H NMR: An approach to understanding metabolism and enzyme mechanism? *Phytochemistry Reviews* **2003,** *2* (1), 87-102.
113. Remaud, G. S.; Martin, Y.-L.; Martin, G. G.; Martin, G. J., Detection of sophisticated adulterations of natural vanilla flavors and extracts: application of the SNIF-NMR method to vanillin and p-hydroxybenzaldehyde. *Journal of agricultural and food chemistry* **1997,** *45* (3), 859-866.
114. Eiler, J. M.; Clog, M.; Magyar, P.; Piasecki, A.; Sessions, A.; Stolper, D.; Deerberg, M.; Schlueter, H.-J.; Schwieters, J., A high-resolution gas-source isotope ratio mass spectrometer. *International Journal of Mass Spectrometry* **2013,** *335*, 45-56.
115. Ogrinc, N.; Košir, I.; Spangenberg, J.; Kidrič, J., The application of NMR and MS methods for detection of adulteration of wine, fruit juices, and olive oil. A review. *Analytical and Bioanalytical Chemistry* **2003,** *376* (4), 424-430.
116. Hays, P. A.; Remaud, G. S.; Jamin, É.; Martin, Y.-L., Geographic origin determination of heroin and cocaine using site-specific isotopic ratio deuterium NMR. *Journal of Forensic Science* **2000,** *45* (3), 552-562.
117. Zare, R. N.; Kuramoto, D. S.; Haase, C.; Tan, S. M.; Crosson, E. R.; Saad, N. M., High-precision optical measurements of 13C/12C isotope ratios in organic compounds at natural abundance. *Proceedings of the National Academy of Sciences* **2009,** *106* (27), 10928-10932.

118. Kerstel, E., Isotope ratio infrared spectrometry. *Handbook of stable isotope analytical techniques* **2004,** *1*, 759-787.
119. Kerstel, E.; Gianfrani, L., Advances in laser-based isotope ratio measurements: selected applications. *Applied Physics B: Lasers and Optics* **2008,** *92* (3), 439-449.
120. Wheeler, M. D.; Newman, S. M.; Orr-Ewing, A. J.; Ashfold, M. N., Cavity ring-down spectroscopy. *Journal of the Chemical Society, Faraday Transactions* **1998,** *94* (3), 337-351.
121. Dempster, A., A new method of positive ray analysis. *Physical Review* **1918,** *11* (4), 316.
122. Flenker, U., Isotope ratio mass spectrometry–history and terminology in brief. *Drug testing and analysis* **2012,** *4* (12), 893-896.
123. Nier, A. O., A mass-spectrographic study of the isotopes of argon, potassium, rubidium, zinc and cadmium. *Physical Review* **1936,** *50* (11), 1041.
124. Nier, A. O.; Gulbransen, E. A., Variations in the relative abundance of the carbon isotopes. *Journal of the American Chemical Society* **1939,** *61* (3), 697-698.
125. Ricci, M. P.; Merritt, D. A.; Freeman, K. H.; Hayes, J., Acquisition and processing of data for isotope-ratio-monitoring mass spectrometry. *Organic geochemistry* **1994,** *21* (6-7), 561-571.
126. Craig, H., Isotopic standards for carbon and oxygen and correction factors for mass-spectrometric analysis of carbon dioxide. *Geochimica et cosmochimica acta* **1957,** *12* (1-2), 133-149.
127. Preston, T.; McMillan, D., Rapid sample throughput for biomedical stable isotope tracer studies. *Biological Mass Spectrometry* **1988,** *16* (1-12), 229-235.
128. Matthews, D.; Hayes, J., Isotope-ratio-monitoring gas chromatography-mass spectrometry. *Analytical Chemistry* **1978,** *50* (11), 1465-1473.
129. Barrie, A.; Bricout, J.; Koziet, J., Gas chromatography—stable isotope ratio analysis at natural abundance levels. *Biological Mass Spectrometry* **1984,** *11* (11), 583-588.
130. Accoe, F.; Berglund, M.; Geypens, B.; Taylor, P., Methods to reduce interference effects in thermal conversion elemental analyzer/continuous flow isotope ratio mass spectrometry δ18O measurements of nitrogen-containing compounds. *Rapid Communications in Mass Spectrometry* **2008,** *22* (14), 2280-2286.
131. Carter, J.; Barwick, V. *Good practice guide for isotope ratio mass spectrometry, FIRMS*; ISBN 978-0-948926-31-0: 2011.
132. Cody, R. B.; Tamura, J.; Musselman, B. D., Electrospray ionization/magnetic sector mass spectrometry: calibration, resolution, and accurate mass measurements. *Analytical Chemistry* **1992,** *64* (14), 1561-1570.
133. King, S. J.; Price, S. D., Electron ionization of CO 2. *International Journal of Mass Spectrometry* **2008,** *272* (2), 154-164.
134. Orient, O.; Strivastava, S., Electron impact ionisation of H2O, CO, CO2 and CH4. *Journal of Physics B: Atomic and Molecular Physics* **1987,** *20* (15), 3923.
135. Thomann, I.; Lock, R.; Sharma, V.; Gagnon, E.; Pratt, S. T.; Kapteyn, H. C.; Murnane, M. M.; Li, W., Direct measurement of the angular dependence of the single-

photon ionization of aligned N2 and CO2. *The Journal of Physical Chemistry A* **2008,** *112* (39), 9382-9386.
136. Nier, A. O., A redetermination of the relative abundances of the isotopes of carbon, nitrogen, oxygen, argon, and potassium. *Physical Review* **1950,** *77* (6), 789.
137. McKinney, C. R.; McCrea, J. M.; Epstein, S.; Allen, H.; Urey, H. C., Improvements in mass spectrometers for the measurement of small differences in isotope abundance ratios. *Review of Scientific Instruments* **1950,** *21* (8), 724-730.
138. Luz, B.; Barkan, E.; Bender, M. L.; Thiemens, M. H.; Boering, K. A., Triple-isotope composition of atmospheric oxygen as a tracer of biosphere productivity. *Nature* **1999,** *400* (6744), 547-550.
139. Severinghaus, J. P.; Grachev, A.; Luz, B.; Caillon, N., A method for precise measurement of argon 40/36 and krypton/argon ratios in trapped air in polar ice with applications to past firn thickness and abrupt climate change in Greenland and at Siple Dome, Antarctica. *Geochimica et Cosmochimica Acta* **2003,** *67* (3), 325-343.
140. Eiler, J. M.; Schauble, E., 18 O 13 C 16 O in Earth's atmosphere. *Geochimica et Cosmochimica Acta* **2004,** *68* (23), 4767-4777.
141. Eiler, J. M., "Clumped-isotope" geochemistry—The study of naturally-occurring, multiply-substituted isotopologues. *Earth and Planetary Science Letters* **2007,** *262* (3), 309-327.
142. Renpenning, J.; Hitzfeld, K. L.; Gilevska, T.; Nijenhuis, I.; Gehre, M.; Richnow, H.-H., Development and validation of an universal interface for compound-specific stable isotope analysis of chlorine (37Cl/35Cl) by GC-high-temperature conversion (HTC)-MS/IRMS. *Analytical chemistry* **2015,** *87* (5), 2832-2839.
143. Shouakar-Stash, O.; Drimmie, R. J.; Zhang, M.; Frape, S. K., Compound-specific chlorine isotope ratios of TCE, PCE and DCE isomers by direct injection using CF-IRMS. *Applied Geochemistry* **2006,** *21* (5), 766-781.
144. Rosman, K.; Taylor, P., Isotopic compositions of the elements 1997 (Technical Report). *Pure and Applied Chemistry* **1998,** *70* (1), 217-235.
145. Kellner, M. I.; Madachy, R. J.; Raffo, D. M., Software process simulation modeling: Why? what? how? *Journal of Systems and Software* **1999,** *46* (2), 91-105.
146. Govan, A. In *Introduction to optimization*, North Carolina State University, SAMSI NDHS, Undergraduate workshop, 2006.
147. Nguyen, H. S., Approximate boolean reasoning: foundations and applications in data mining. In *Transactions on rough sets V*, Springer: 2006; pp 334-506.
148. Coello, C. A. C.; Lamont, G. B.; Van Veldhuizen, D. A., *Evolutionary algorithms for solving multi-objective problems*. Springer: 2007; Vol. 5.
149. Grodzevich, O.; Romanko, O., Normalization and other topics in multi-objective optimization. **2006**.
150. Marler, R. T.; Arora, J. S., The weighted sum method for multi-objective optimization: new insights. *Structural and multidisciplinary optimization* **2010,** *41* (6), 853-862.
151. Yijie, S.; Gongzhang, S., Improved NSGA-II multi-objective genetic algorithm based on hybridization-encouraged mechanism. *Chinese Journal of Aeronautics* **2008,** *21* (6), 540-549.

152. Soland, R. M., Multicriteria optimization: A general characterization of efficient solutions. *Decision sciences* **1979,** *10* (1), 26-38.
153. Zhou, A.; Qu, B.-Y.; Li, H.; Zhao, S.-Z.; Suganthan, P. N.; Zhang, Q., Multiobjective evolutionary algorithms: A survey of the state of the art. *Swarm and Evolutionary Computation* **2011,** *1* (1), 32-49.
154. Zadeh, L., Optimality and non-scalar-valued performance criteria. *IEEE transactions on Automatic Control* **1963,** *8* (1), 59-60.
155. Goicoechea, A.; Hansen, D. R.; Duckstein, L., Multiobjective decision analysis with engineering and business applications. **1982**.
156. Stillwell, W. G.; Seaver, D. A.; Edwards, W., A comparison of weight approximation techniques in multiattribute utility decision making. *Organizational behavior and human performance* **1981,** *28* (1), 62-77.
157. Pólya, G.; Szegö, G., *Problems and Theorems in Analysis II: Theory of Functions. Zeros. Polynomials. Determinants. Number Theory. Geometry*. Springer Science & Business Media: 1997.
158. Boyer, C. B.; Merzbach, U. C., *A history of mathematics*. John Wiley & Sons: 2011.
159. Smith, D. E., *A source book in mathematics*. Courier Corporation: 2012.
160. Podlubny, I., *Fractional differential equations: an introduction to fractional derivatives, fractional differential equations, to methods of their solution and some of their applications*. Academic press: 1998; Vol. 198.
161. Miller, K. S.; Ross, B., An introduction to the fractional calculus and fractional differential equations. **1993**.
162. Hale, J. K.; Lunel, S. M. V., *Introduction to functional differential equations*. Springer Science & Business Media: 2013; Vol. 99.
163. Hirsch, M. W.; Smale, S.; Devaney, R. L., *Differential equations, dynamical systems, and an introduction to chaos*. Academic press: 2012.
164. Mehlhorn, K.; Sanders, P., *Algorithms and data structures: The basic toolbox*. Springer Science & Business Media: 2008.
165. Bentley, J. L., Multidimensional binary search trees used for associative searching. *Communications of the ACM* **1975,** *18* (9), 509-517.
166. Knott, G. D., A numbering system for binary trees. *Communications of the ACM* **1977,** *20* (2), 113-115.
167. Szu, H. H. In *Non-convex optimization*, Proceedings of the SPIE, 1986; pp 59-65.
168. Yu, W.; Lui, R., Dual methods for nonconvex spectrum optimization of multicarrier systems. *IEEE Transactions on Communications* **2006,** *54* (7), 1310-1322.
169. Clarke, F. H.; Ledyaev, Y. S.; Stern, R. J.; Wolenski, P. R., *Nonsmooth analysis and control theory*. Springer Science & Business Media: 2008; Vol. 178.
170. Shanno, D. F., Conditioning of quasi-Newton methods for function minimization. *Mathematics of computation* **1970,** *24* (111), 647-656.
171. Bertsekas, D. P.; Nedi, A.; Ozdaglar, A. E., Convex analysis and optimization. **2003**.
172. Borwein, J.; Lewis, A., Convex Analysis and Nonlinear Optimization. Number 3 in CMS Books in Mathematics. Springer-Verlag: 2000.

173. Strang, G., Linear Algebra and its Applications 3rd edition, Thomsom Learning. Inc: 1988.
174. Courant, R.; Hilbert, D., Methods of mathematical physics, vol. I, Interscience, New York, 1953. *Google Scholar* **1989**, 414-415.
175. Meyer, C. D., *Matrix analysis and applied linear algebra*. Siam: 2000; Vol. 71.
176. Prussing, J., The principal minor test for semidefinite matrices. *Journal of Guidance, Control, and Dynamics* **1986,** *9* (1), 121-122.
177. Haupt, R. L.; Haupt, S. E., *Practical genetic algorithms*. John Wiley & Sons: 2004.
178. Michalewicz, Z., *Genetic algorithms+ data structures= evolution programs*. Springer Science & Business Media: 2013.
179. Konak, A.; Coit, D. W.; Smith, A. E., Multi-objective optimization using genetic algorithms: A tutorial. *Reliability Engineering & System Safety* **2006,** *91* (9), 992-1007.
180. Jaeggi, D.; Parks, G. T.; Kipouros, T.; Clarkson, P. J., The development of a multi-objective Tabu Search algorithm for continuous optimisation problems. *European Journal of Operational Research* **2008,** *185* (3), 1192-1212.
181. Ali, M. M.; Törn, A.; Viitanen, S., A numerical comparison of some modified controlled random search algorithms. *Journal of Global Optimization* **1997,** *11* (4), 377-385.
182. Jónasson, K., *Applied Parallel and Scientific Computing: 10th International Conference, PARA 2010, Reykjavík, Iceland, June 6-9, 2010, Revised Selected Papers*. Springer Science & Business Media: 2012; Vol. 7134.
183. Pacheco, J.; Martí, R., Tabu search for a multi-objective routing problem. *Journal of the Operational Research Society* **2006,** *57* (1), 29-37.
184. Molina, J.; Laguna, M.; Martí, R.; Caballero, R., SSPMO: A scatter tabu search procedure for non-linear multiobjective optimization. *INFORMS Journal on Computing* **2007,** *19* (1), 91-100.
185. Glover, F., Tabu search—part I. *ORSA Journal on computing* **1989,** *1* (3), 190-206.
186. Glover, F., Heuristics for integer programming using surrogate constraints. *Decision Sciences* **1977,** *8* (1), 156-166.
187. Glover, F., Future paths for integer programming and links to artificial intelligence. *Computers & operations research* **1986,** *13* (5), 533-549.
188. Glover, F., Tabu search methods in artificial intelligence and operations research. ORSA. *Artificial Intelligence Newletter* **1987,** *1* (2).
189. Glover, F.; McMillan, C.; Novick, B., Interactive decision software and computer graphics for architectural and space planning. *Annals of operations research* **1985,** *5* (3), 557-573.
190. Glover, F.; McMillan, C., The general employee scheduling problem. An integration of MS and AI. *Computers & operations research* **1986,** *13* (5), 563-573.
191. Caballero, R.; Gandibleux, X.; Molina, J. *MOAMP: A generic multiobjective metaheuristic using an adaptive memory*; Technical report, University of Valenciennes, Valenciennes, France: 2004.
192. Aiex, R. M.; Binato, S.; Resende, M. G., Parallel GRASP with path-relinking for job shop scheduling. *Parallel Computing* **2003,** *29* (4), 393-430.

193. Glover, F.; Laguna, M., Tabu search (Vol. 22). Boston: Kluwer academic publishers: 1997.
194. Glover, F.; Laguna, M.; Martí, R., Fundamentals of scatter search and path relinking. *Control and cybernetics* **2000,** *29* (3), 653-684.
195. Jaeggi, D.; Asselin-Miller, C.; Parks, G.; Kipouros, T.; Bell, T.; Clarkson, J. In *Multi-objective parallel tabu search*, International Conference on Parallel Problem Solving from Nature, Springer: 2004; pp 732-741.
196. Ugray, Z.; Lasdon, L.; Plummer, J. C.; Glover, F.; Kelly, J.; Marti, R., A multistart scatter search heuristic for smooth NLP and MINLP problems. In *Metaheuristic Optimization via Memory and Evolution*, Springer: 2005; pp 25-57.
197. Doyle, B., *C# Programming: From Problem Analysis to Program Design*. Cengage Learning: 2013.
198. Malik, D. S., *C++ Programming: From problem analysis to program design*. Cengage Learning: 2014.
199. Weininger, D.; Weininger, A.; Weininger, J. L., SMILES. 2. Algorithm for generation of unique SMILES notation. *Journal of Chemical Information and Computer Sciences* **1989,** *29* (2), 97-101.
200. Murray-Rust, P.; Rzepa, H. S.; Wright, M., Development of chemical markup language (CML) as a system for handling complex chemical content. *New journal of chemistry* **2001,** *25* (4), 618-634.
201. Murray-Rust, P.; Rzepa, H. S., Chemical markup, XML, and the Worldwide Web. 1. Basic principles. *Journal of Chemical Information and Computer Sciences* **1999,** *39* (6), 928-942.
202. Murray-Rust, P.; Townsend, J. A.; Adams, S. E.; Phadungsukanan, W.; Thomas, J., The semantics of Chemical Markup Language (CML): dictionaries and conventions. *Journal of cheminformatics* **2011,** *3* (1), 43.
203. O'Boyle, N. M.; Banck, M.; James, C. A.; Morley, C.; Vandermeersch, T.; Hutchison, G. R., Open Babel: An open chemical toolbox. *Journal of cheminformatics* **2011,** *3* (1), 33.
204. Altun, M.; Ceyhan, T.; Kartal, M.; Atay, T.; Özdemir, N.; Cevheroğlu, Ş., LC method for the analysis of acetylsalicylic acid, caffeine and codeine phosphate in pharmaceutical preparations. *Journal of pharmaceutical and biomedical analysis* **2001,** *25* (1), 93-101.
205. Appel, A. S.; McDonough, J. H.; McMonagle, J. D.; Logue, B. A., Analysis of Nerve Agent Metabolites from Hair for Long-Term Verification of Nerve Agent Exposure. *Analytical chemistry* **2016,** *88* (12), 6523-6530.