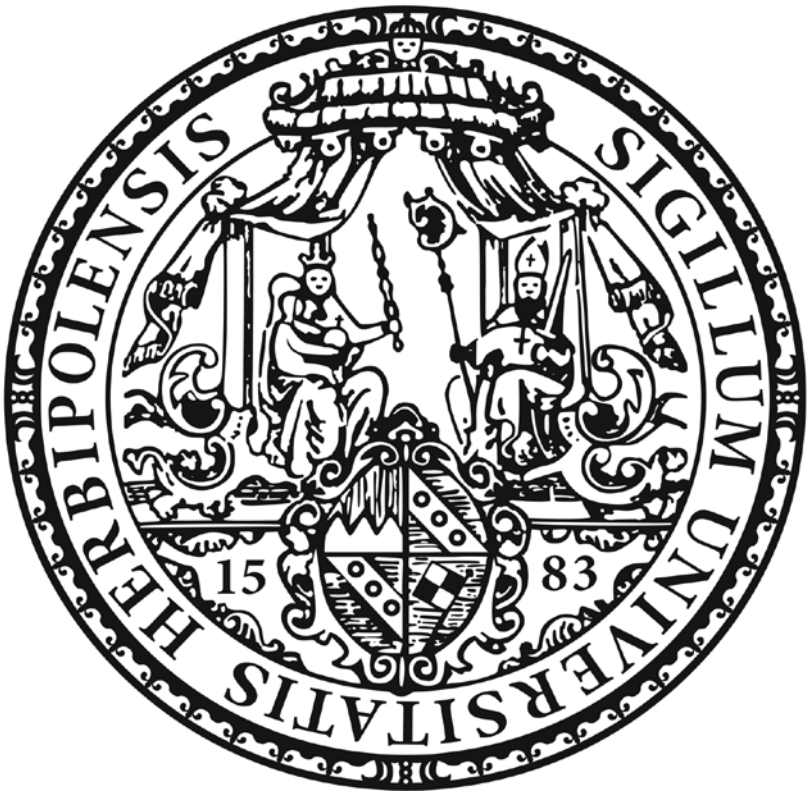
Bioinformatics Software for Metabolic and Health Care Data Management



Dissertation zur Erlangung des
naturwissenschaftlichen Doktorgrades
der Julius-Maximilians-Universität Würzburg

vorgelegt von
Zeeshan Ahmed
aus Gujrat, PK.
Würzburg 2012

Eingereicht am: …**02 October 2012**...

bei der Fakultät für Biologie der Julius-Maximilians-Universität Würzburg

Mitglieder der Promotionskommission:

Vorsitzende: Prof. Dr. Wolfgang Rössler
1. Gutachter: Prof. Dr. Thomas Dandekar
2. Gutachter: PD. Dr. Wolfgang Eisenreich

Tag des öffentlichen Promotionskolloquiums: … **20 December 2012**, 11:00hr …

Doktorurkunde ausgehändigt am: …….....

AFFIDAVIT / EIDESSTATTLICHE ERKLÄRUNG

English:

I hereby confirm that my thesis entitled “Bioinformatics Software for Metabolic and Health Care Data Management” is the result of my own work. I did not receive any help or support from commercial consultants. All sources and / or material applied are listed and specified in the thesis.

Furthermore, I confirm that this thesis has not yet been submitted as part of another examination process neither in identical nor in similar form.
Additionally, other than this degree, I have not applied or will attempt to apply for another degree or qualification in relation to this thesis.

Würzburg, Signature:

Place, Date (Zeeshan Ahmed)

Wuerzburg, 02 October 2012.

German:

Hiermit erkläre ich an Eides statt, die Dissertation “ Bioinformatik-Software für Stoffwechsel-und Health Care Data Management” eigenständig, d.h. insbesondere selbstständig und ohne Hilfe eines kommerziellen Promotionsberaters, angefertigt und keine anderen als die von mir angegebenen Quellen und Hilfsmittel verwendet zu haben.

Ich erkläre außerdem, dass dir Dissertation weder in gleicher noch in ähnlicher Form bereits in einem anderen Prüfungsverfahren vorgelegen hat.
Zusätzlich habe oder werde ich nicht versuchen neben diesem Abschluss einen weiteren Abschluss oder Qualifikation mit dieser Doktorarbeit zu erweben.

Würzburg, Unterschrift:

Ort, Datum (Zeeshan Ahmed)

Würzburg, 02 October 2012

SUMMARY

Computer Science approaches (software, database, management systems) are powerful tools to boost research. Here they are applied to metabolic modelling in infections as well as health care management. Starting from a comparative analysis this thesis shows own steps and examples towards improvement in metabolic modelling software and health data management.

In section 2, new experimental data on metabolites and enzymes induce high interest in metabolic modelling including metabolic flux calculations. Data analysis of metabolites, calculation of metabolic fluxes, pathways and their condition-specific strengths is now possible by an advantageous combination of specific software. How can available software for metabolic modelling be improved from a computational point of view? A number of available and well established software solutions are first discussed individually. This includes information on software origin, capabilities, development and used methodology. Performance information is obtained for the compared software using provided example data sets. A feature based comparison shows limitations and advantages of the compared software for specific tasks in metabolic modeling. Often found limitations include third party software dependence, no comprehensive database management and no standard format for data input and output. Graphical visualization can be improved for complex data visualization and at the web based graphical interface. Other areas for development are platform independency, product line architecture, data standardization, open source movement and new methodologies. The comparison shows clearly space for further software application development including steps towards an optimal user friendly graphical user interface, platform independence, database management system and third party independence especially in the case of desktop applications. The found limitations are not limited to the software compared and are of course also actively tackled in some of the most recent developments. Other improvements should aim at generality and standard data input formats, improved visualization of not only the input data set but also analyzed results. We hope, with the implementation of these suggestions, metabolic software applications will become more professional, cheap, reliable and attractive for the user. Nevertheless, keeping these inherent limitations in mind, we are confident that the tools compared can be recommended for metabolic modeling for instance to model metabolic fluxes in bacteria or metabolic data analysis and studies in infection biology.

In Section 3 and 4 introduces own software developed for metabolic modelling. We implement mathematical algorithms for the prediction of amino acid fluxes in bacteria. For this, relative intensity values, natural and relative abundance values from labeled metabolic substance have to be calculated based on experimental NMR data and spectra. To give technical reader a complete idea about the program structure behind these software applications (LS-MIDA and Isotopo), we have presented major implemented UML designs i.e. Use Case, Data Flow, Internal Work Flow, System Sequence and Component Diagrams. We have also described GUIs in brief. During the software development process from LS-MIDA to Isotopo we involved some mathematical algorithms e.g. Brauman's least square algorithm, binomial theorem to analyse isotopologue distribution, different matrix calculations, linear and multiple regressions. Using the new software applications, isotopologue patterns can be easily calculated from MS data, visualized and prepared for metabolic flux modelling. This allows numerous applications for metabolic modelling in infetion biology. We made an effort to provide easy-to-use software for users of isotopologue profiling and metabolic flux analysis. Isotopologue measurements are rapidly translated into metabolic flux prediction applying the Isotopo software with good application potential for microbiology and biotechnology.

Research in section 5 proposes the embedding of a new approach towards the advancement of healthcare informatics. Debating the aspects of global information sharing and bringing together researchers and clinical practitioners in a common ICT platform an analysis has been conducted in the field of Product Data Management (PDM) and its importance has been elaborated in detail in this section, starting from an existing knowledge base established in the department of bioinformatics on cardiovascular disease and molecular cascades in the platelet, the "platelet web knowledgebase". After scrutinizing existing PDM systems and their architecture we propose a new concept of adopting PDM in the field of clinical data management (using bioinformatics) by providing a new way of transition. We have presented our proposed conceptual and implementation architecture for the implementation of a new Advanced Product Data Analysis and Management (ADAM) System in cardiovascular disease following earlier mentioned PDM components, mandatory functional requirements and development guidelines according to the goals and scope of this research using mentioned tools and technologies. This new proposition unites molecular and clinical data and can contribute to the innovation objectives in a variety of ways and constitutes a significant change in existing approaches by transiting PDM concepts supporting new kind of clinical heath care management system implementation which by its structure significantly enhances the use of its generated knowledge content on molecular cascades involved in cardiovascular disease.

ZUSAMMENFASSUNG

Informatik Ansätze (Software, Datenbank, Management-Systeme) sind wichtige Werkzeuge für die Forschung in der Biologie. Ausgehend von einer vergleichenden Analyse zeigt diese Arbeit eigene Schritte und Beispiele zur Verbesserung von metabolischer Modellierungs-Software und Gesundheit Datenmanagementsystemen auf.

Neue experimentelle Daten über Metaboliten und Enzyme führen zu hohem Interesse an metabolischen Modellierungen einschließlich Stoffwechselflusses Berechnungen. In Kapitel 2 zeigen wir, das die Datenanalyse von Metaboliten, die Berechnung der Stoffflüsse und Wege sowie die spezifischen Softwarestärken nur durch eine vorteilhafte Kombination voll ausgeschöpft werden. Wie kann Software zur metabolischen Modellierung von einer informatischen Sicht her verbessert werden? Eine Anzahl von verfügbaren und gut etablierten Softwareansätzen wird zunächst einzeln diskutiert. Dazu gehören Informationen über Software-Herkunft, Fähigkeiten, Entwicklung und verwendeten Methodik einschließlich Testdatensätzen und Modellen. Ein Vergleich zeigt, merkmalsbasierte Einschränkungen und Vorteile der verglichenen Software für spezifische Aufgaben in der metabolischen Modellierung. Häufige Einschränkungen der verglichenen Software sind ihre Abhängigkeit von Drittanbietern, kein umfassendes Datenbank-Management und kein Standard-Format für Dateneingabe und -ausgabe. Die grafische Visualisierung für komplexe Visualisierungen von Daten und die Web-basierte grafische Benutzeroberfläche kann oft noch verbessert werden. Andere Bereiche für weitere Entwicklung sind Plattformunabhängigkeit, Produktlinien-Architektur, Daten-Standardisierung, die Open-Source-Bewegung und neue Algorithmen und Methoden. Der Vergleich zeigt deutlich Möglichkeiten für weitere Entwicklung von Softwareanwendungen auf, einschließlich Schritten in Richtung einer optimalen, benutzerfreundlichen grafischen Benutzeroberfläche, Plattform-Unabhängigkeit, Datenbank-Management-System und Unabhängigkeit von weiterer software, vor allem im Falle von Desktop-Anwendungen. Die gefundenen Einschränkungen sind von allgemeiner Bedeutung für bioinformatische Modellierungssoftware einschließlich jüngster Entwicklungen.

Weitere Verbesserungen betreffen standardisierte Formate und eine, verbesserte Visualisierung von Eingabedatensatz und analysierten Ergebnissen. Wir hoffen, dass mit der Umsetzung dieser Vorschläge metabolische Software-Anwendungen professioneller werden, billiger, zuverlässiger und attraktiver für den Anwender. Trotz dieser inhärenten Einschränkungen im Hinterkopf sind wir zuversichtlich und

zeigen auch, dass die verglichenen Werkzeuge mannigfaltige Aspekte der metabolischen Modellierung etwa von Stoffwechselflüssen in Bakterien bei Infektionen abdecken und die verglichene software für solche Aufgaben empfohlen werden kann.

In Kapitel 3 und 4 befassen sich mit der Etablierung eigener Software-Anwendungen zur metabolischen Modellierung. Es werden mathematische Algorithmen zur Vorhersage der relativen Intensitätswerte implementiert, um Aminosäureflüsse in Bakterien richtig vorherzusagen. Natürliche und relative Häufigkeiten werden aus den NMR-Daten für einzelne Metabolite berechnet . Das informatische Design der Software-Anwendungen (LS-MIDA und Isotopo) wird untersucht einschließlich UML-Schemata, Nutzeranforderungen, Datenfluss und interner Arbeitsablauf im Programm, Computerystem und seinen Komponenten. Auch die graphische Benutzeroberfläche (GUI) wird kurz beschrieben und erläutert. Bei der Software-Entwicklung nutzten wir verschiedene mathematische Algorithmen wie Braumans Methode der kleinsten Quadrate, das Binomialtheorem zur Ermittlung der Isotopologenverteilung, verschiedene Matritzenberechnungen, lineare und multiple Regressionen. Mit den neuen Software-Anwendungen können Isotopolog-Muster leicht von MS-Daten, berechnet und visualisiert werden. Dies erlaubt zahlreiche Anwendungen in der metabolischen Modellierung, etwa in der Infektionsbiologie. Wir waren bestrebt, eine möglichst einfach zu bedienende Software für Nutzer und Anwender von Isotopolog-Profilen und Stoffwechselflussanalysen zu etablieren. Isotopolog-Messungen können so schnell in Stoffwechselfluss-Vorhersagen mit gutem Anwendungspotenzial für Mikrobiologie und Biotechnologie umgesetzt werden.

In Kapitel 5 geht es um einen neuen Ansatz für die Weiterentwicklung von Datenbankmanagementsystemen im Gesundheitswesen. Wir diskutieren den modernen, globalen Informationsaustausch und die Zusammenführung von Forschern und Klinikern in einer gemeinsamen computer gestützten Plattform das Datenmanagement von molekularen und Patientendaten bei kardiovaskulären Krankheiten, ausgehend von Thrombozyten und einer am Lehrstuhl bereits etablierten Wissensdatenbank („Platelet Web Knowledgebase“) über Blutplättchen, Signalkaskaden und pharmakologischen Interventionsmöglichkeiten. Um zu einem innovativem, neuen Ansatz zu gelangen, werden Techniken des Product Data Management (PDM) und seine Bedeutung im Detail betrachtet. Wir stellen zunächst einige bestehende PDM-Systeme und deren Architekturen vor. Wir schlagen ein neues bioinformatisches Konzept für PDM im Bereich des klinischen Daten-Management vor, mit einer neuen Art der Integration molekularer und klinischer Daten. Die vorgeschlagene Architektur erlaubt damit die

Implementierung eines neuen Advanced Product Data Analysis and Management (ADAM) Systems für kardiovaskuläre Krankheiten unter Einbeziehung molekularer Kaskaden des Krankheitsgeschehens.

ACKNOWLEDGEMENT

As a believer, I would like to express my sincere gratitude to the God for blessing me with the potential, strength and ability to work on this doctoral research, development and thesis writing.

As a doctoral student, I would like to pay my heartiest regards to my supervisors Prof. Dr. Thomas Dandekar and PD. Dr. Wolfgang Eisenreich for their generous support, detailed instruction, kind encouragement, persistent support throughout my research and believe in me that I can do Doctoral research in the field of Bioinformatics. Being grateful to my supervisors, I would like to say, I wish every supervisor will be like them. Whether I work more for the short or the long period of time with you (supervisors) but I will always respect you from the bottom of my heart.

As a funded researcher, I recognize that this research would not have been possible without the financial assistance and support of German Research Foundation (DFG) SPP 1316 and SFB Transregio 34/Z1.

As a colleague, I would like to thank to all people from the Department of Bioinformatics, Biocenter, University of Wuerzburg and people (especially Dr. Eva Elyert and Dr. Claudia Huber) from the Department of Biochemistry, Technical University of Munich, Germany for their support.

I would also like to thank my family for the support they provided me through my entire life and in particular, I must acknowledge my mother “Dr. Mrs. Mussarat Saeeda” without whose love and encouragement, I would not have finished this thesis.

I would like to give special recognition to my son “Jibrael Zeeshan” (born 06.10.2010) for his love, especially during the time when I was bit upset and frustrated, with his smiles he changed my mood, gave me strength and cheering my life.

In the end would like to give special thanks to my wife “Mrs.Saman Zeeshan”, for her love, care and most of all introducing the field of Bioinformatics to me and helped me in realizing its importance, concepts and complexities.

I dedicate this doctoral thesis to my beloved family

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1 Introduction

1.1 Motivation and Scope

A branch of science, originally conceptualized in middle ages by Greeks and Muslim physicians then modernized in 19th Century by European scientists i.e. Karl Friedrich Burdach, Gottfired Reinhold Treviran and Jean Baptiste, renowned as Biology, dealing with living organisms including their structure, functions, growth and processes of classification, formation and mutation. Continuous research and vital progress in respective field has categorized it in to various sub disciplines on the basis of the subject and the methods i.e. molecular biology, genetics, biochemistry, cellular biology, physiology and ecology.

The major subject of our research belongs to the field of molecular biology, but as we are dealing with the complex interactions of molecules of living organisms and interested in understanding the interactions between cells (Cell; the combination of many interconnected pathways that demonstrate multifaceted regulation) belonging to the various cellular systems, the overlapped involvement of cellular biology, genetics and biochemistry is also there.

As the result of massive previously done and ongoing research in to the field of molecular biology, almost ten different techniques have been introduced and involved in molecular biology .i.e. Polymerase chain reaction (PCR), Gel electrophoresis, Macromolecule blotting and probing, Southern blotting, Northern blot, Western blottin, Eastern blotting, Deoxyribonucleic acid (DNA) microarray, Allele specific oligonucleotide (ASO), Antiquated technologies. Here, in our research we are not directly dealing with any of these types but trying to find better analytical approaches with the involvement of mathematics and informatics to analyze obtained results from performed experiments using some of all available methodologies.

Progressive research and development into the field of chemistry (and related e.g. Biochemistry) investigating living organisms is in a period which contains practical assurance of dealing with the complex pathways consisting of known living system comprising component, amino acids, vitamins, and minerals.

Metabolite is the composition of enzyme catalyzed (change in rate of a chemical reaction due to the participation of a substance) reactions occur naturally within cells. A metabolite is the byproduct of metabolism, pathway consisting of defined small but rapidly changing chemicals in between. Metabolites

and their pathways (meaningful apparently infinite series of chemical reaction occurring within living cells) are central for pathophysiological adaptation of cells.

Recent methodological advances allow direct measurements of metabolites. Metabolic modeling elucidates in silico all possible flux pathways (flux balance analysis) as well as predicts actual fluxes under a given situation. There is a growing need for bioinformatics to process such data. We compare regarding calculation and analysis of flux pathways the programs C13, Classical and Dynamic FBA, MetaTool, FiatFlux, BioOPT, Rematch, VANTED and YANASquare. A feature based comparison includes limitations and aspirations for specific tasks in metabolic modeling. The above software differs with specific advantages for the specific problems solved. Often found limitations include third party software required to execute, no comprehensive database management systems available, no standard format for data input and output. The graphical user interface can be improved further as well as data visualization for analysis of experimental data (for details please see section 2).

Synthesizing metabolite (chemical compound produced by metabolic reaction) and balancing it, is the most essential mechanism for the cellular existence. It is really needed to prevent the development of complex and dangerous diseases e.g. homeostatic balance is a key method to stop the growth of cancer. There are molecular relationships between cell survival, cell death and cell cycle (Maddika et al., 2007). Numerous molecular pathways have been attributed which are regulating the cell survival, and death as well. One of the basic problems is analysis of intracellular fluxes which is caused by the presence of cyclic metabolic pathways in the cell metabolism (Bonarius et al., 1998). It is quite difficult to quantify the fluxes in such pathways exclusively by measuring extracellular metabolic rates and the biomass composition (Vallino and Stephanopoulos, 1990).

The knowledge of metabolic pathways and fluxes is important to understand the adaptation of organisms to their biotic and abiotic environment. Some computational methods have been published which are proposed to measure fluxes of underdetermined metabolic networks e.g. (Bonarius et al., 1996), (Van Gulik and Heijnen, 1995), (Savinell and Palsson, 1992) and (Fell and Small, 1986). Flux in cyclic pathways can be determined by isotopic tracer techniques e.g. tracing the metabolic outcome by carbon labeling, using Nuclear Magnetic Resonance (NMR) or Gas Chromatography- Mass Spectrometry (GC-MS) techniques to biological systems (Shulman et al., 1979). For more than seven decades many attempts have been made to measure metabolic fluxes using isotopes, but still currently a comprehensive, general

approach is still missing, as it is difficult to deconvolute the full isotopic composition in the metabolic network under physiological relevant conditions (Hellerstein and Neese, 1992). Moreover, generally, two isotopic methodologies were proposed and widely used for quantitative polymer analysis i.e. first by quantifiably incorporating labeled precursor into cell (Zilversmit et al., 1943) and second by determining isotopic dilution (decay) (Steele, 1959), but both methodologies experienced the problem of determination of a mixture of polymers.

The specific distribution of stable isotope labeled precursors into metabolic products can be taken as fingerprints of the metabolic events and dynamics through the metabolic networks. There are already some good software applications available for metabolic flux and pathways analysis but still useful open source software are required that can easily and rapidly calculates mass spectra of labeled metabolites, derivatives and their fragments global isotope excess and isotopomer distribution. Narrowing focus of research towards this problem, an open source software application "Least Square Mass Isotopomers Analyzer" (LS-MIDA) is presented that processes experimental mass spectrometry data on the basis of metabolite information such as the number of atoms in the compounds, mass to charge ratio (m/e or m/z) values of the compounds and fragments under study, and the experimental relative intensities (or abundance) of the labeled compounds, in comparison to those of the unlabelled molecules.

The software uses Brauman's least square method of linear regression. As a result, global isotope enrichments and fractional molar abundances of each isotopomer in the compounds and fragments under study are obtained and displayed. The new software provides an open source platform that easily and rapidly converts experimental Mass Spectrometry (MS) patterns of labeled metabolites into isotopomer patterns that are the basis for subsequent observation-driven analysis of pathways and fluxes, as well as for model-driven metabolic flux calculations.

During this research we have investigated different bacterial systems and analyzed bacterial data obtained from Staph-aurous, Listeria but our presented study revolves around the data obtained using Salmonella, a gram-negative bacteria which causes gastroenteritis and enteric fever. One of the major reasons for choosing Salmonella, nevertheless, it is an alive threat to the public health but with a fruitful model system for the study of fundamental mechanisms of bacterial pathogenesis (Ohl and Miller, 2001). The effectiveness of the LS-MIDA is extensively tested using metabolites based isotopic experimental data obtained from Salmonella labeling experiments, resulting with estimated natural abundances, relative

abundances and percentage of relative abundances per m/z values and drawn spectrums (for details please see section 3).

The obtained results were correct and according to the implemented least square method but further continuing the research for more optimistic results, this thesis also presents a user-friendly platform for mass isotopomer distribution analysis, another technique that enables the determination of metabolic fluxes on the basis of labeling experiments using ^{13}C-enriched precursors. A new software application named 'Isotopo' was developed with facile data management and robustness to quantify the populations of isotopomers in mixtures of ^{13}C-labelled amino acids. Isotopo is an application with the ability of analyzing quantitative mass spectrometry for isotopologue mixtures of compounds (e.g. amino acids) to derive metabolic fluxes. Isotopo processes experimental isotopomer data i.e. metabolite information, mass to charge ratio (m/z) values, relative intensities of labeled and unlabelled compounds, and the number of carbon atoms in the fragments. It is using partial implementation of least square method of linear regression based on the experimental data elements, Isotopo estimates mass values (M_{-0}, M_{-1}, $M_{maximum}$) and predicts relative intensities with respect to the used mass to charge ratios, natural abundances, relative abundances and fractional molar abundances of each fragment derived from the compound under study. This includes data sets with three actual intensity values against one mass to charge ratio value and affords absolute global enrichments in conjunction with both natural and relative abundances for the underlying isotopologue. Using the new software application, isotopologue patterns can be easily calculated from MS data, visualized and prepared for metabolic flux modelling in an effort to provide easy-to-use software for users of isotopologue profiling and metabolic flux analysis (for details please see section 4).

Data analyzer is the most important module of any software application developed for experimental data analysis but this does not kill the importance of experimental data management. If data will not be properly maintained and available for analysis, then it will be really difficult to perform computational or measurement analysis, especially in case of heavy data. To overcome the problems of maintaining heavy amount of experimental isotopomers data, new data management systems were also needed. This research thesis also proposes a new database structural schema for global data standardization and management of experimental isotopomers data with a local third party independent data management system (for details please see section 4).

Enhancing the capabilities of extensive data management this research proposes a new approach with intelligent information retrieval mechanism embedding with a flexible graphical user interface. This research also support the medical community with the proposition and implementation of an innovative initiative, addressing major challenges of providing optimal management of a certain diagnosis, working rapidly under emergency conditions, protecting personal data, coupling individual patient data with general repositories, allowing therapy monitoring, analyzing individual variations with the and incorporation of golden standard therapy guidelines. The concepts of Product Data Management (PDM) transit the field of Bioinformatics for the implementation of an advanced clinical data analysis and management system. This research proposes layered conceptual and implementation architectures along with tools and technologies for the development of a prototype system (for details please see section 5).

This thesis report is organized as follows; starting from a more general overview, section1 presents targeted problems and challenges (section 1.2), highlights metabolic flux analysis (section 1.3), mass isotopomers distribution analysis (section 1.4) and health care data management (section 1.4). Going into the details, section 2 presents software review research conducted towards metabolite flux modeling tools analysis and feature comparison. Section3 presents research conducted solely towards the field of mass isotopomer distribution analysis, focusing on living problems of the respective field a new software application “LS-MIDA” is proposed, designed, implemented and evaluated. Expanding the scope of research for more optimistic results, section 4 presents another newly developed software application “Isotopo” towards the identification of the quantity of population of labeled isotopomers for resolving the exact rate of synthesized fractions present in the mixture and metabolic experimental data management. Section 5 supports medical community with the proposition of an innovative initiative, addressing major challenges of providing optimal management of a certain diagnosis, working rapidly under emergency conditions, protecting personal data, coupling individual patient data with general repositories, allowing therapy monitoring, analyzing individual variations with the and incorporation of golden standard therapy guidelines.

1.2 Problems and Challenges

Where in past the results of research in the field of Biology based on enormous findings lead to a heavy amount of success there at the same time this high rate increases the expectations for good results in future as well. Now days sufficient availability of biological data is here but major problem is to analyze

and understand it in possible shortest time and conclude with best optimum results. To overcome these deficiencies an extended field of biology has been introduced i.e. Bioinformatics, with the inclusion of mathematics and informatics. Bioinformatics has attempted to explore several areas e.g. metabolic network analysis and (re)construction, automation of genome annotation, protein structure determination etc., and provided values to the field of biology but still lots of areas need to be targeted and improved.

Metabolic flux analysis is an analysis technique for metabolic modeling and flux balance analysis, allows important insights into metabolism and adaptations of different organisms (Eisenreich et al. 2010), (Eylert et al 2010). Metabolic modeling is a complex and broad field including a number of specific tasks requiring dedicated software development offering a multitude of different solutions to the user. Flux balance analysis is a method to analyze metabolism, using mathematical methods applied to biological system. In the research one of the challenges is to conclude with some of the good software applications developed for metabolic flux prediction, analysis and pathway analysis, a software review research is conducted. Furthermore describe metabolite flux modeling tools considering also isotopologue data. Perform software evaluations process, a feature based comparison and conducted reviewed research results including identified limitations and aspirations.

The major objective of this research is to study metabolic isotope to quantify the fraction of metabolites of interest in the mixture typically by tracing isotopes. Mass Isotopomer Distribution Analysis (MIDA) measures in mixtures of polymers (e.g. lipids, carbohydrates, and proteins) by quantifying relative abundances of molecular species with mass spectrometry (Hellerstein et al., 1999). Estimating mass isotopomers distribution from spectral data is an extension of the quantitative mass spectrometric method to a multi component mixture analysis (Lee et al., 1991). Mass spectrometric analysis describes relative abundances quantitatively based on combinatorial probabilities. It is essential to identify the quantity of the labeled isotopomer population to resolve the exact amount present in the mixture. The exact identification of the number of labeled isotopomers in a mixed population of molecules is a mathematical challenge. Different calculation algorithms have already been proposed and published for MIDA with overlapping solutions in successive iterations (Korzekwa et al., 1990). These generate identical results. A formal mathematical algorithm generates an appropriate set of linear simultaneous equations and finds their solutions, Brauman's Least Square Algorithm (Brauman, 1966). For MIDA this allows calculation

of natural isotope abundances and relative isotope abundances. Two current and incompletely resolved issues are thus addressed in this research i.e. MIDA and experimental isotopomer data management.

There is no such efficient solution available for extensive isotopologue data management and manipulation. Using our experience in relational data management system development, we are looking forward to take this challenge and design a new data management system for experimental isotopologue data.

Considering the problem of steadily increasing cardiovascular diseases in an aging population despite generally high health standards, and looking at the need for a new health care management system; patient life cycle management system. One bioinformatical solution regarding platelet proteome, phosphoproteome, kinome, transcriptome and interactome is produced by Prof. Dandekar's group (Functional genomics and systems Biology, Department of Bioinformatics, Biocenter at the University of Wuerzburg Germany) i.e. PlateletWeb (Boyanova et al., 2012), a knowledge base providing comprehensive protein-protein interaction data. Further enhancing the scope of our research, using our previous experiences in product data management together with concepts of relational data management system development, we are objecting to propose a new clinical data management system capable of managing key cardiovascular diseases entities (stroke, thrombosis, failing heart, diabetic angiopathies) as well as individual case histories.

In this research we address the challenge for extensive experimental data management for high quality healthcare and ageing to provide a new comprehensive solution towards Personal Health Systems Implementation for Patient Guidance Services (PGS), safety healthcare record information reuse (with disease management) with integrated solution for the information infrastructure of larger repositories. The subject Health is aligned with the fundamental objectives of medicine related research to improve the health of living beings and increase the competitiveness of health related services, socio economic dimensions of health care and global clinical data sharing. Our focus is on one specific aspect of Bioinformatical analysis: the molecular processes involved in cardiovascular diseases. The aim of this research is to enable living beings to live longer, independently, in good health by increasing the average number of healthy life years to improve the sustainability and efficiency of our social and healthcare

systems. Furthermore the scope is up to the proposition of creating innovative products and services (global health care management system) for international markets.

1.3 Metabolic Flux Analysis (MFA)

Metabolites and their pathways are central for pathophysiological adaptation of cells. Recent methodological advances allow direct measurements of metabolites. Metabolic modeling elucidates in silico all possible flux pathways (flux balance analysis) as well as predicts actual fluxes under a given situation. There is a growing need for bioinformatics to process such data.

Metabolic modeling is a broad field including a number of specific tasks. We describe metabolite flux modeling tools with a focus on isotopologue data. Textbooks give representations of central pathways such as the glycolysis. However, bioinformatics allows formally calculating and establishing how many pathways are there for a given set of enzymes. Most well known for this task are flux balance analyses methods which try to balance all metabolites within a cellular system by a suitable combination of enzymes: For each so called internal metabolite the same amount is consumed as is produced.

The stable and balanced metabolite flux achieved by such a combination of enzymes is called a flux mode. In practice, already many such flux modes are possible for a moderate sized set of enzymes (30-50) and it can be further distinguished whether the enzyme combination cannot be reduced or split further without losing the ability to balance all used internal metabolites, which would then be an elementary mode (Papin et al., 2004). Furthermore, a specific set of such elementary modes is sufficient to reproduce all other modes by linear combination of the modes. Such a set is called a convex basis of the flux system and it can be shown (Schuster et al., 2000) that these modes are extreme situations of the systems, so this is called extreme pathway analysis.

Modifications and extensions regarding metabolic flux analysis have continuously been proposed (e.g. Kaleta et al. 2009; Rahman et al. 2006). Despite limitations, in particular the combinatorial explosion for large-scale systems, FBA has become a standard to model metabolite fluxes in different systems with software packages such as METATOOL (Schuster and Schuster, 1993), Classical and Dynamic FBA (Mahadevan. et al. 2002) etc. Another problem is to map fluxes to actual observed experimental data. For this solutions such as YANA (Schwarz R. et al., 2005) or Rematch (Esa Pitk¨anen. et al., 2008) etc. exist

which fit flux distributions to measured metabolites or protein or gene expression data. Furthermore, there is the specific problem of fitting fluxes to measured isotopologue data.

Regarding processing of isotopologue data, a number of software packages have come up such as FiatFlux and C13. We will not examine software treating GC-MS data to estimate e.g. nucleotide concentrations or software for lipidomics and the large amount of solutions available to estimate proteome data (Nahnsen et al. 2011). The list of enzymes involved in a metabolic model is often established from genome data involving annotation or biochemical pathway data such as KEGG database (Kanehisa, 2002). Furthermore, there are various annotation tools to identify enzymes from the genome sequence. Software we developed for such tasks includes the KEGGbrowser (Schwarz et al., 2007) to rapidly establish enzyme networks, GENOVA (Liang et al., 2009) and InGENO (Liang and Dandekar, 2006) for genome annotation and comparisons, as well as JANE to integrate transcriptome data (Liang et al., 2009).

However, many groups working in genome-based bioinformatics developed similar types of tools of their own for such annotation tasks. Hence, the problem is important but too complex for a fair comparison between all available tools. It will not discussed here in detail for the shear amount of different annotation tools developed by various groups involved in major genome annotation efforts including quite advanced user interfaces (MAGPIE (Gaasterland and Sensen, 1996), GenDB (Meyer et al. 2003), GENEDB (Flora et al., 2012), PEDANT (Dmitrij et al., 2003)), large-scale interactome searching tools (e.g. STRING (Szklarczyk et al., 2011) and important large-scale repositories (e.g. EBI, NCBI) including various tools for domain annotation SMART (Ponting et al., 1999; Letunic et al., 2002), COG (Roman et al., 2000), Daileon (Perillo et al., 2009) and E2D (Lee et al., 2007).

This research gives a “best mode” on how to combine different software to achieve quantitative metabolic flux analysis including visualization. For this different possibilities and recommended optimal solution have to be compared. However, focus is only on the combination which worked best in our hands, with the current development and depending on experience alternative combinations maybe more attractive or powerful. Moreover, different software tested from a computer scientist’s view and point out some general limitations in the tested software. In summary the reader gets a useful tool-kit at hand

including recommendations, tasks and capabilities as well as knowledge on basic limitations and suggestions for further development.

1.4 Mass Isotopomer Distribution Analysis (MIDA)

Metabolism is central for all cellular processes including adaptation of organisms to their respective life style and conditions. Triggered by the presence and activity of metabolic enzymes and the metabolite fluxes through pathways, cellular reactions constitute a highly dynamic network that can be rapidly and efficiently modulated in response to environmental changes. A number of theoretical techniques have been established to predict metabolic fluxes (Dauner et al., 2001), (Nicola et al., 2005).

In contrast, only few methods allow to directly determine metabolic fluxes, one of which is based on in vivo experiments using stable isotope labeled precursors, such as ^{13}C-glucose or $^{13}CO_2$. The transfer of label to the metabolic network and the specific isotope distribution in metabolic products can then be taken as evidence of metabolic pathways and fluxes during the experimental period. However, robust technology is required to quantitatively determine the isotopomer abundances in multiple metabolites. Specifically, experimental intensities of mass traces (typically of silylated derivatives metabolites and fragments thereof) have to be converted into relative and molar isotopomer abundances.

Isotopologues are species of a compound that differ only in their isotopic composition (Brenninkmeijer et al., 2003). The term isotopomer is a contraction of ‘isotopic isomer‘, grouping isotopologues into those molecules which contain the same number of a specific isotope (e.g. ^{13}C) at different positions. As an example, 64 stable carbon isotopologues exist for glucose. Out of these, six species constitute isotopomers with one ^{13}C-atom at position 1, 2, 3, 4, 5, or 6. In natural compounds, i.e. obtained from the natural environment, the natural abundance is the consequence of the natural isotope abundance (i.e. ca. 1.1 % for ^{13}C at a given carbon position) that is diverted through the complete population of isotopologues due to statistical reasons. In contrast, artificial isotopologue abundances are observed in labeling experiments where isotope-enriched precursors (e.g. ^{13}C-labelled) are supplied to the organism under study. This results in the enrichment of specific isotopologues, i.e. on top of the natural abundances, in the metabolic products. The deconvolution of mass intensities yielding isotopomer enrichment is the key task of the software described in this manuscript. Notably, mass intensities provide information of the abundances of isotopologues harbouring a specific number of the isotope, i.e. one, two,

three etc. ^{13}C-atoms, and therefore, the enrichment of Isotopomeric groups (isotopomer distribution) are obtained. Since metabolic pathways lead to specific isotopomer distributions, the latter values can be used to identify and to quantify the relative contributions of metabolic routes from the labeled precursor to the products observed by MS.

The objective of metabolic isotope studies is to quantify the fraction of metabolites of interest in the mixture typically by tracing isotopes (Lee et al., 1991, Nicola et al. 2005). Isotopologues are species of a compound that differ only in their isotopic composition (Brenninkmeijer et al., 2003). Isotopic labeling allows the quantification of any compound, as it uses a known amount of stable isotopically substituted analog to a sample. During this technique the primary ion for each isotopic species is monitored and correct ion intensities are measured overlapping with other isotopic species (Korzekwa et al., 1990).

So far, three different methods are available for positional isotopomer determination, i.e. NMR, MS analysis of a sufficient number of metabolite fragments, and multiple reaction monitoring (MRM). These methods can provide orthogonal information and can thus be combined using our software to improve positional isotopomer determination (Christensen and Nielsen, 1999). The well established instrument used for MS analysis is GC-MS.

MS is a systematic technique to measure the mass-to-charge ratio values of charged particles). It converts individual molecules into ions to direct them in magnetic fields using Mass Spectrometer. During GC-MS, at first a mixture of compounds is inserted into the GC to vaporize using a heated chamber to separate compounds for MS analysis, by travelling into GC column. A chromatogram is drawn, representing each compound with its peak. All Mass Spectrometers consists of three main sections: Ionizer, Ion Analyzer and Detector. Electron impact ionization is performed during Ionizer by providing gas chromatograph using a high-energy electron beam to collect molecular ions and fragments. Ion Analyzer accelerates obtained molecular ions and fragments by maneuvering the charged particles using mass spectrometer, to eliminate uncharged molecular ions and fragments. The job of the Detector is to generate an electronic signal at ion hit. During this process mass analyzer classifies the ions with respect to the mass to charge ratio values and detector extracts the abundance values of each mass to charge ratio value.

Mass isotopomer distribution analysis (MIDA) measures in mixtures of polymers (e.g. lipids, carbohydrates, and proteins) by quantifying relative abundances of molecular species with mass spectrometry (Hellerstein and Neese, 1999). Estimating mass isotopomers distribution from spectral data is an extension of the quantitative mass spectrometric method to a multi component mixture analysis (Lee et al., 1991).

Mass spectrometric analysis describes relative abundances quantitatively based on combinatorial probabilities. It is essential to identify the quantity of the labeled isotopomer population to resolve the exact amount present in the mixture. The exact identification of the number of labeled isotopomers in a mixed population of molecules is a mathematical challenge. Different calculation algorithms have already been proposed and published for MIDA with overlapping solutions in successive iterations (Korzekwa et al., 1990). These generate identical results. A formal mathematical algorithm generates an appropriate set of linear simultaneous equations and finds their solutions, Brauman's Least Square Algorithm (Brauman, 1966). For MIDA this allows calculation of natural isotope abundances and relative isotope abundances.

Currently, only commercial software or user-specific approaches are available for the conversion of mass intensities (provided by the specific software implemented to the mass spectrometer) to the relative and molar isotopomer enrichments such as tandem mass spectrometric data computing for positional isotopomer distributions (Rantanen et al., 2002), measurements of mass distributions by mass spectrometry (Winden et al., 2002), isotopomer analysis using GC-MS (Lee et al., 1991) and GC-MS analysis for isotopomer balancing (Dauner et al., 2000).

Two current and incompletely resolved issues are thus addressed here: Mass Isotopomers Distribution Analysis (MIDA) and Experimental Data Management. MIDA is delivered starting from labeled substance based experimental data elements e.g. mass to charge ratio, number of fragments, actually observed relative abundances during experimentation and mass values. Spectrometry analysis examines each peak of a spectrum of a given mass to find contributions of each molecular species. Furthermore, our experimental data management system provides efficient file based data manipulation and management standardizing experimental data format and decreasing extra labor efforts. We provide an updated perspective on a technique that provides a fundamental solution to the problem of measuring relative isotopomers abundances.

Two newly different software applications: Least Square Mass Isotopomers Distribution Analyzer (LS-MIDA) and Isotopo are proposed, developed and evaluated during this research. These both have with the ability of performing quantitative mass spectrometry readily to mixtures of materials labeled with stable isotopes that can be very important for both biomedicine and biochemistry. LS-MIDA is developed to estimate mass isotopomer distributions from spectral data, by analyzing each peak of given mass using Brauman´s algorithm for accurate estimation of natural and relative abundances. However Isotopo is an application with the extended ability of analyzing quantitative mass spectrometry for isotopologue mixtures of compounds (e.g. amino acids) to derive metabolic fluxes. Isotopo processes experimental isotopomer data i.e. metabolite information, mass to charge ratio (m/z) values, relative intensities of labelled and unlabelled compounds, and the number of carbon atoms in the fragments. It is using Brauman's least square method of linear regression based on the experimental data elements, Isotopo estimates mass values (M-0, M-1, M maximum) and predicts relative intensities with respect to the used mass to charge ratios, natural abundances, relative abundances and fractional molar abundances of each fragment derived from the compound under study. This includes data sets with three actual intensity values against one mass to charge ratio value and affords absolute global enrichments in conjunction with both natural and relative abundances for the underlying isotopologue.

The presented methods are useful ways to examine quantitative mass spectral data by analyzing the contributions to each peak of provided mass. Taking advantages of existing research (mathematical or algorithmic) in the field of mass isotopomers distribution analysis, we adopted the least square method and implemented it in the form of a software application to provide a comprehensive way for experimental biological (isotopomers) data standardization, management and manipulation. The reliability, mathematical originality and importance of the least square method have already been discussed in detail over the years by J. I. Brauman and W.N. Paul Lee. Isotopologue measurements are rapidly translated into metabolic flux prediction applying the Isotopo software with good application potential for microbiology and biotechnology.

1.5 Health Care Data Management (HCDM)

Cardiovascular diseases are steadily increasing in an aging European population despite generally high health standards. Together with increasing efforts in individualized medicine and direct burden of care for

aged, multi-morbid patients, new, highly effective solutions for prevention and treatment of cardiovascular diseases in aging populations are needed.

To achieve this we start from a scientific backbone of integrated genomics, proteomics and clinical laboratory data on key molecular cascades in thrombosis and hemostasis together with detailed data on individual genetic and post-genomic variation. Clinical data on key cardiovascular diseases entities (stroke, thrombosis, failing heart, diabetic angiopathies) as well as individual case histories are systematically collected and connected on the next level of investigation providing a systems biological view on disease states, changes and individual clinical variation. However, our project concept is to go beyond this and establish, integrating these two levels for better treatment of individual patients, an intelligent individual health data management (PHS, Personal Health System) providing knowledge, prevention and optimal advice and care for individual patients.

An innovative solution is achieved using PDM techniques as framework on all levels considered, integrating both molecular data on pathophysiological cascades involved in thrombosis and haemostasis with laboratory and clinical data and case histories of key cardiovascular disease entities to manage individual patients, diagnoses and treatments for high quality healthcare in an aging population. PDM allows Patient Life Cycle Management (PLCM), integrating patient information (throughout the whole project time as well as individual case histories) from multiple sources and locations with active patient participation via ubiquitous access. To meet our goals, a data management approach provides secure and global data management (e.g. digital health records, digital images etc.), data modelling and simulations as well as visualization with intelligent, multi role based, flexible and self learning graphical user interface. We will analyze multi parametric data and individual variations, for therapy monitoring, optimization and incorporating gold standard therapy guidelines. PDM techniques allow optimal management of diagnoses, work rapidly even under emergency conditions, integrate social, organizational, ethical and legal aspects, protect personal data and couple individual patient data with general repositories. Our system will be ready for the new European Union (EU) health care system.

Clinics and university hospitals consist of different departments located in different places e.g. cities, countries etc. Every clinical place has its own setup, regulations, staff and patient data. There is no doubt that the treatment (practitioners, surgeons etc) departments are the important sections and expected to play a vital role in the progress of a clinic by treating patients but at the same time the department taking

care of patient data consisting of its personal and medical records (reports) cannot be ignored. In the past, there were no such systems available to store, track and manage patient data consisting of its personal and medical records (reports). This doesn't mean that there was no system for data management; there were some systems to store the information about clinics (or hospitals), personnel involved in different operations, financial details and patient data but there was no such comprehensive system to manage and share data globally.

Most of the available health care management systems consist of four different but integrated sections: Patient, Hospital/Clinical Staff, Business Management and Attached Institute. "Patient" section is divided into three subsections: Treatment Process, Provided Facilities and Data Records. "Treatment" is responsible for patient's registration at the time of arrival, consultation of patient with doctors, admission discharge transfer (ADT) decision, nursing and operation (according to need). "Provided Facilities" is about to help patient with the provision of medicine from pharmacy, radiology (on doctor's recommendation), clinical tests (e.g. blood etc) from laboratories, providing diet and laundry with housekeeping facilities (in case if patient is hospitalized). "Data Records" contain the documentation about patient's billing, medicinal reports and order communications. "Hospital/Clinical Staff" section is based on physicians, surgeons, nurses, managerial and operational people. "Business Management" is an independent section responsible for business related activities e.g. hospital/clinic information system, human resource management, payroll management, account/finance management, inventory management, income analysis, report management and revenue cycle management. "Attached Institute" is divided into medical and research institutes.

Each section of the health care management system has independent data and management system to maintain it. The job of existing data management systems is limited up to the data storing, managing and sharing to limited number of people. Main data (associated to each section) are divided into two categories: private and public. In most of the cases private data is personal (patient, clinic and business) and not shared outside the boundaries unless it is really needed. In contrast public data is available at open access to people (e.g. notifications, warnings, guidelines and advertisements etc.).

The existing health care community (Europe or even Worldwide) doesn't have such health care data management solution which could offer a life cycle management system capable of providing the treatment life cycle data management of the patient (caused with any kind of disease), patient life cycle

management maintaining individual patient data history, help in analyzing the individual and accumulative changes in a person's life over time caused by aging, hospital life cycle management and global information (only public) sharing among medicinal and related communities.

2 MFA Software Review

Various software applications have been developed for pathway analysis (Dandekar et al., 2012b) e.g., BioMet Toolbox (Cvijovic. et al., 2010), CellNetAnalyzer (formerly FluxAnalyzer) (Steffen and Axel, 2002), COBRA Toolbox (Becker et al., 2007), COPASI (Stefan et al., 2006), Fluxor (Emanuele et al., 2008), Jarnac (Sauro, 2000), Pathway Analyser (Oehm et al., 2008), MMT (Hurlebaus et al., 2002), SBTOOLBOX2 (Henning and Mats, 2005), TinkerCell (Deepak et al., 2009), WebCell (Dong-Yup et al., 2006) SCAMP (Herbert, 1993) and iMAT (Zur et al., 2010). All of these can be used for certain aspects of pathway analysis considering also regulation, involved interaction networks as well as kinetics. However, the following software will now be discussed in more detail and compared as it presents a powerful combination for metabolic modeling starting from raw data (e.g. isotopologue measurements) and ending up with a description of flux distributions. However, this is a dynamic field and use and preferences (including our own) for software depends a lot on how often this particular software was already used, operating system preferences, the scientific question and purpose in mind and of course the type of data analyzed. Hence, we offer for each step of analysis two alternatives.

We discuss and analyze also build-in limitations of the software. However, overall, the software presented is considered by us to be quite useful and is recommended to the reader for the specific purposes analyzed. Our selection covers C13 (Wiechert et al. 1997), Classical and Dynamic FBA (Mahadevan. et al. 2002), MetaTool (Schuster and Schuster, 1993), BioOPT (Cvijovic. et al. 2010), FiatFlux (Zamboni. et al. 2005) (Zamboni. 2007), Rematch (Esa Pitkänen. et al. 2008), VANTED (Junker, B.H., et al. 2006) and YANASquare (Schwarz R, et al. 2007). We thus discuss different openly (freely) available software (desktop and web based) applications providing solutions to necessary and required steps of metabolic flux analysis: First there is the construction of the metabolic network (KEGGconverter (Konstantinos, 2009), Metannogen (Christoph et al. (2007), KEGGbrowser (Schwarz et al., 2007) and then a ^{13}C-constrained flux analysis (C13 (Wiechert et al. 1997), Openflux (Lake-Ee et al. 2009), ReMatch(Esa Pitk¨anen. et al., 2008)), flux balance analysis (Metatool (Schuster and Schuster, 1993), BioOpt (Cvijovic. et al. 2010) or Classical and dynamical FBA (Mahadevan. et al. 2002)), or a fit to data (FiatFlux (Zamboni et al., 2007), YANAsquare (Schwarz R, et al. 2007), YANAvergence (Liang et al. 2011)) can be performed and then can in detail be analysed and visualised (VANTED (Junker, B.H., et al. 2006), YANAsquare (Schwarz R, et al. 2007)). The tests and data provided by the different applications can be used more or less by the others and involved tasks such as reconstructing glycolysis

and pentose phosphate pathway (Schuster et al., 2000), map isotopologue data on these pathways (Eisenreich et al., 2006), calculate elementary modes and resulting fluxes according to the data and visualize these (Liang et al., 2011).

Applications generally aim for biological insights for instance mutations in any of the above pathways with effect of survival in macrophages such as mutations in the aldolase gene in Listeria (Schauer et al., 2010). However, in the details each package provides different advantages and limitations and is involved in different steps of this analysis chain. This makes it difficult to give a fair comparison choosing a specific test data set and applying it to all. Hence, we focus in the following more on their technical performance regarding the comparison.

2.1 Software Descriptions

2.1.1 C13

C13 is a software application developed in MATLAB using the Wiechert and de Graaf framework (Wiechert et al. 1997) for metabolic flux analysis by stationary carbon isotope labeling on fractional enrichment data. C13 takes the composition of measurements of extracellular rates, fractional enrichment data and a metabolite as input, to process with. Extracellular rates are substrate uptake rates, product formation rates and drain fluxes into biomass. Fractional enrichment data consider here the labeling state of the substrates. The metabolic model consists of biochemical conversions, transport processes and the outcome of carbon atoms throughout the metabolic network.

C13 performs metabolic flux analysis by implementing the concept of isotopic labeling (of carbon atoms), analyses fluxes and fractional labeling and obtains the required results by iterative error minimization. The procedure consists of four major steps.

- Estimate fluxes satisfying stoichiometric constraints.
- Resolve limited enrichments by isotope balances around carbon atoms.
- Compute deviation between fluxes and between fractional labeling.

- Keep repeating first three steps until the resultant information based on total deviation is under a certain threshold.

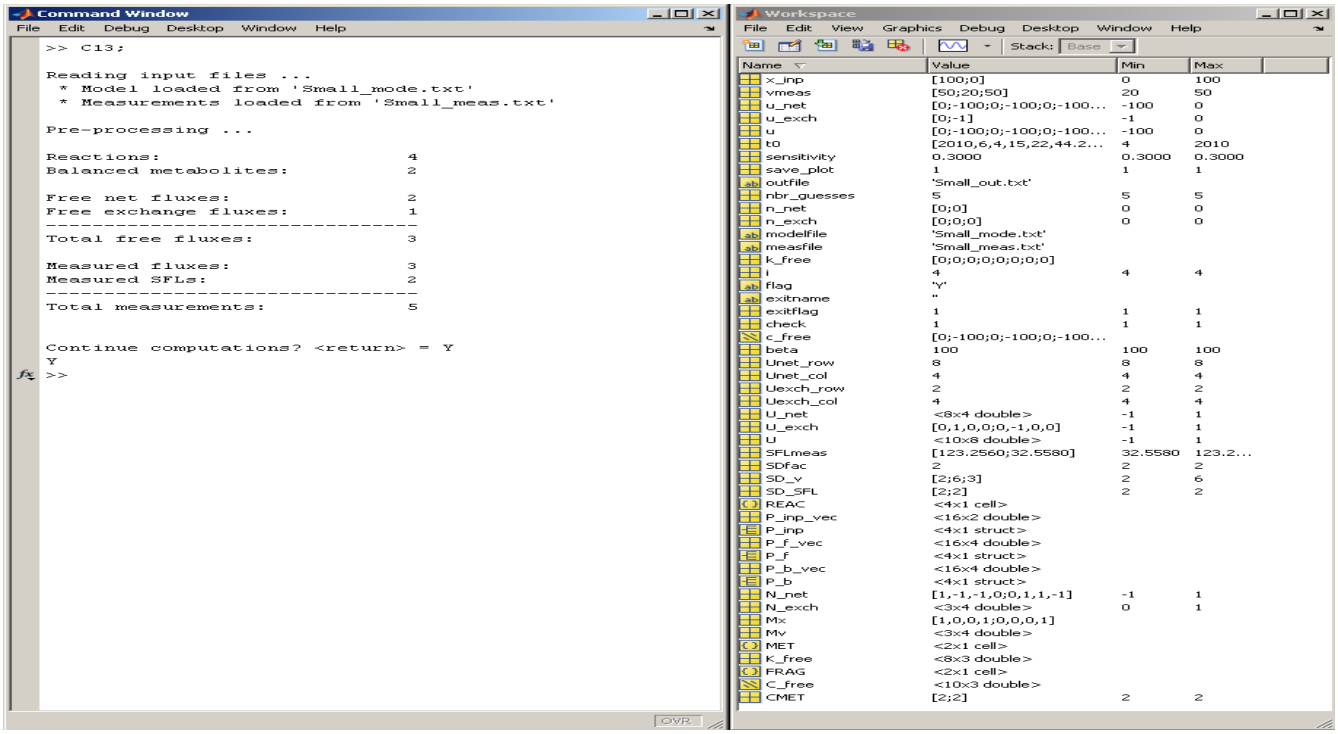


Figure 1: C13 Software.

Figure Legend 1: C13 software. On the left the console view (command window) for C13 is shown. Rapidly input data (measurements) are read in; C13 needs a model of the flux distribution and then calculates net fluxes according to the data. On the right you see the workspace, this gives an overview on all files managed as well as some kinetic data and ranges of the metabolite data.

C13 takes input data based on two different files: model and measurements files (Figure 1). “Model” is a text input file based on a metabolic model consisting of the following input elements: reactions, bounds, reversibility, constraints and substrate uptake rate. “Reactions” is the list of the reactions that compose a metabolic network and the respective carbon atom transitions. Two bounds i.e. lower and upper, net and exchange fluxes can be defined. Enzyme reaction reversibility or irreversibility has to be provided. If the reaction is irreversible the net flux may be constrained to be within a positive range of values and the exchange flux should be set to zero. In contrast, if the reaction is highly reversible the exchange flux needs to be set to 1. Furthermore, if the reaction is reversible then net fluxes can be positive, zero or negative and with no constraints on the exchange flux. Constraints are defined, in case of absence of any

constraints, default values with defined ranges (-500 to +500 for net fluxes and 0 to 1 for exchange fluxes) are considered. Substrate uptake rate is needed to be set considering the net flux (equal to a specific value, normalized to 100). “Measurement” is also a textual input file on the measurements to sum fractional labeling and fluxes.

To get started using C13, one follows the guide line by the authors and the provided example Small_mode.txt and Small_meas.txt (model and measurement files). The input file Small_mode.txt consists of the following input elements: Reactions, Labeled Metabolites and External Metabolites, where as the input file Small_meas.txt contains the information about Fragments, SFL and Flux. Processing results into the output is presented in Figure 1 and describes useful and practical flux estimates according to the input ^{13}C-labeling data. This includes the information about number of reactions, balanced metabolites, free net fluxes, free exchange fluxes, total free fluxes, measured fluxes, measured SFLs and total measurements.

Limits of the software were observable under some input conditions. This included incomplete output presentation, either regarding the output image or that the output text file was empty.

2.1.2 FBA

Flux balance analysis (FBA) is firstly a general constraint-based method for calculating flux distributions (Wiback SJ et al, 2004; Schilling CH et al. 1999) that is powerful with far reaching implications e.g. for biotechnology (Dugar and Stephanopoulos, 2011). Furthermore, it can be useful to optimize certain objective functions e.g. a biomass equation in a metabolic network as well as a number of other biotechnological, biological and biomedical applications (Schäuble et al., 2011).

“Classical and Dynamic Flux Balance Analysis” is a software package (linear programming based) developed to study metabolic systems at steady state, by writing a mass balance for each metabolite Xi of a network over time dt (dynamic mass balance- Eq.1 comparing reactions V for synthesis syn, degradation deg taking usage for other processes use and transformations trans into account) and defining a series of constraints (stoichiometric constraints and the capacity constraints) on the metabolic network.

$$dXi / dt = Vsyn - Vdeg - (Vuse \pm Vtrans) \qquad \text{(Eq. 1)}$$

Classical and Dynamic FBA balances metabolic fluxes (reactions) around each node (metabolite) to constrain the metabolic network by modeling (Mahadevan. et al. 2002). Classical and Dynamic FBA can be recommended, to deal with the lack of kinetic information and have to study feasible and optimal metabolic flux distributions. The input file for FBA consists of metabolic reactions and metabolic pathways to make the stoichiometric matrix, setup 'b' vector and defining an objective function and bounds on metabolic fluxes. CFBA performs the following tasks:

- Reading the stoichiometric matrix
- Providing a good solution regarding the objective function and the optimal value, for all metabolic fluxes.
- Analyzing a dual file to extract shadow prices for all the metabolites.
- Reconstructing metabolic network by systemic mass balance and reaction capacity constraints.
- Drawing the metabolic (network) map, capable of displaying the metabolites with zero, negative, and positive shadow prices in different colors, allows user to display the reactions that have a metabolic flux equal to zero in different colors based on the value of the reduced cost (negative, zero, or positive).
- Adjusting objective function of the metabolic network at run time.
- Performing robustness analysis by analyzing robustness (property of metabolism allowed for fluxes in each metabolic reaction) without altering the objective function, adjusting parameters and drawing robustness diagram.
- Phenotypes are mapped on the metabolic genotype starting from the stoichiometric matrix and graphically showing differences in flux values for the different genotypes and modified reactions, the so called Phase Plane.
- Creating Isoclines (regions) consisting of the same value for the objective function or a certain flux.

- Calculating and comparing flux distributions of two different reaction files.

For easy use of Classical Flux Balance Analysis (CFBA), the authors provided example input data (consisting of metabolic reactions associated with the Escherichia coli (E.coli) genotype and metabolic pathways consisting of glycolysis, pentose phosphate pathway, citric acid cycle (TCA) cycle and respiration) and the resultant output is presented in Figure 2. The analysis consists of three steps (Figure 3; Edwards et al. 2001):

- Collection of input requires metabolic reactions and pathway data for CFBA and metabolic network reconstruction including annotation of the genome sequence.

- Using either metabolic pathway analysis or Phenotype phase plane (PhPP) analysis the feasible steady-state metabolic flux distributions are determined.

- A third analysis step considers and analyses the value of the objective function for specific values of fluxes.

CFBA predicts the pathway flux without explicit consideration of the regulation. It constructs a flux model using all the metabolic reactions and metabolites, starting with the genome sequence annotation to identify all the metabolic enzymes. Then CFBA characterizes identified reactants and products for enzymatic reactions of known enzymes by constructing iterative models. After this, dynamic mass balance is applied in CFBA using metabolite data and reactions to estimate a flux distribution for each reaction. CFBA estimates flux in steady state operation of the metabolic network, causing the need of determination of additional constraints by calculating a value for all fluxes in the network.

Limitations for CFBA include:

1. CFBA is not capable of determining exact metabolic flux values but ranges of allowable flux values.

2. Optimal solutions are calculated but the actual cell behavior is a mixture of these solutions.

3. CFBA does not yield a unique solution for the flux distribution.

4. The optimal flux distribution is only meaningful when interpreted in terms of the flux constraints on transport fluxes.

5. CFBA is not able to predict the metabolite concentrations.

6. CFBA may sometimes incorrectly predict pathway behaviour, e.g. in the example the reutilization of acetate.

7. FBA does not incorporate (but also does not require) characterized kinetic expressions.

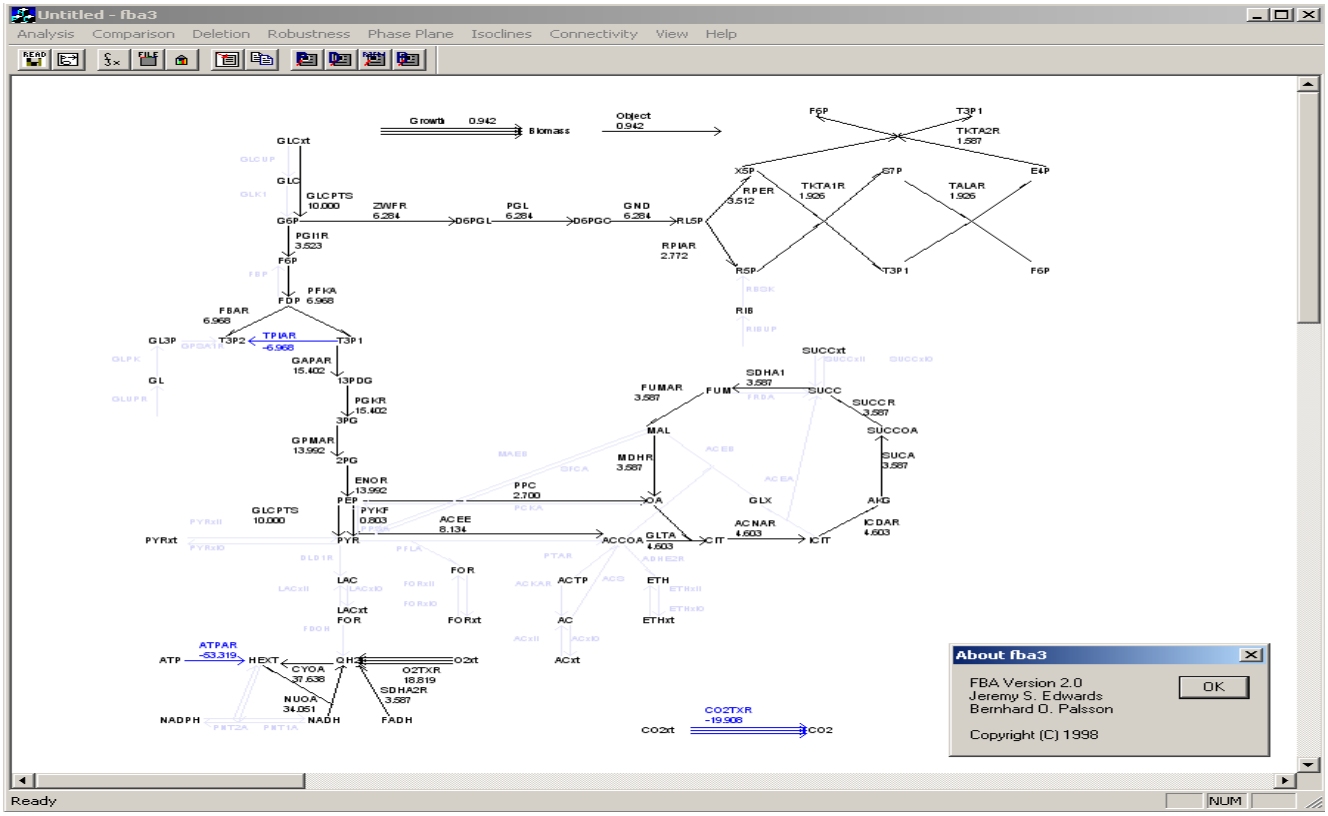


Figure 2: Flux Balance Software.

Figure Legend 2: Flux Balance Analyzer. Shown here is the pathway view of "flux balance analyzer" (FBA, version 2.0), we see the TCA (right) and the PPP (on top, right), the branching of glycolysis (left). FBA calculates flux values for all reactions. Different pathways are calculated and given with graphical output including enzyme names, metabolite fluxes, branching pattern, cofactors.

A new improved version of the existing software (Classical FBA) is developed i.e. Dynamic Flux Balance (DFBA), capable of predicting metabolite concentrations and reutilization of acetate and allowing the incorporation of well characterized kinetic expression (Figure 2). DFBA is also able to provide metabolic engineering the possibility to make strategies to design network. DFBA is a quantitative analysis tool for studying the dynamic reprogramming of metabolic networks, based on two mathematical approaches i.e. Static Optimization Approach (SOA) and Dynamic Optimization Approach (DOA) (Mahadevan. et al. 2002). The static optimization approach (SOA – Eq.2) divides the batch time into N time intervals and tries to solve instantaneous optimization problem at the beginning of each interval, with the application of repeated linear programming to obtain flux distribution at a particular time instant.

$$\text{MAX} \sum \text{wivi}(t) \qquad \text{(Eq.2)}$$

For comparison, the Dynamic Optimization Approach (DOA – Eq.3) tries to solve the optimization problem by considering the entire time at once to obtain time profiles of fluxes and metabolite levels, using non linear programming:

$$Dz/dt = AvX,\ dX/dt = \mu X,\ = \sum \text{wivi} \qquad \text{(Eq.3)}$$

The dynamic FBA software application was not easy to test, but yields more effective results according to its documentation. In summary, CFBA is already a powerful package for FBA, extensions are available, but the specific purpose for which FBA is required. CFBA excels when complex constraints such as a biomass equation have to be account in addition, in the following METATOOL (Schuster and Schuster, 1993) offers simplicity, ease of use and additional information on the modes such as subsets (list of enzymes appearing always together) and network connectivity, whereas BioOPT (Cvijovic. et al. 2010) advantageously takes constraints in the form of shadow prices into account and besides the network calculation predicts effects of gene knockouts with ease (Figure 3).

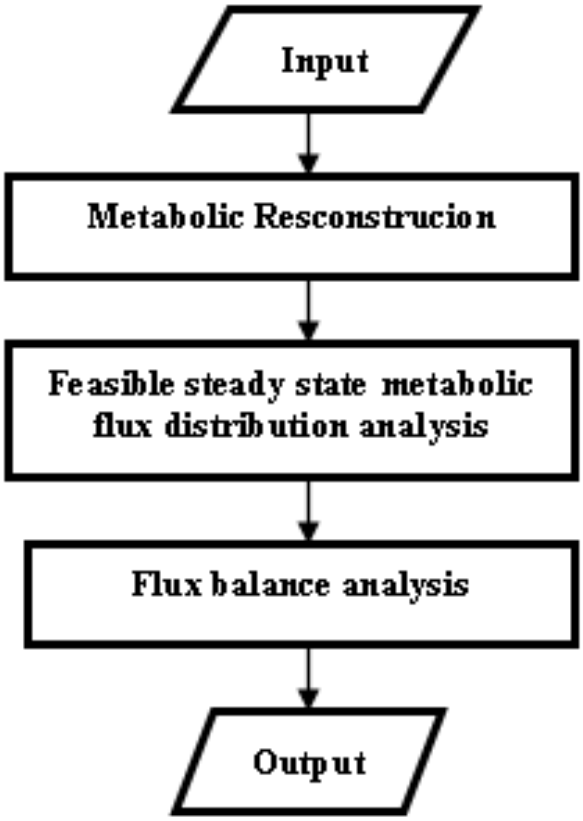


Figure 3: Flux Balance Analysis Procedure.

2.1.3 Metatool

METATOOL was originally programmed by Thomas Pfeiffer from 1998 to 2000 using the C Programming Language for elementary flux mode calculation using an algorithm published by (Schuster and Schuster, 1993). These early versions of METATOOL (Figure 4) are simple stand alone applications capable of taking the input and providing an output in a simple text file format. The algorithm processing speed needed to be revised and the quality of the results was also supposed to be improved. Moreover the input and output file (data) formats of METATOOL were also not based on any standard data format.

For different METATOOL versions, elementary flux mode calculation can be compared in biochemical networks using provided test examples consisting of irreversible enzymes, internal metabolites, external metabolites and reaction equations. One considers the tricarboxylic acid cycle, glyoxylate shunt and adjacent reactions of amino acid synthesis in E. coli. After taking the input, the program gives the numbers of internal metabolites and reactions, parses the reaction equations and translates them into a

stoichiometric matrix (by including stoichiometric coefficients of the internal metabolites in all the reaction equations) and identifies the kernel or nullspace (subspace of all flux vectors). The resultant information is presented in Figure 4. Furthermore, during experimentation, some minor observations were noticed i.e. processing time differs in each execution when even the input file was same all the time, the noted processing time difference is with on average difference of 1 second and 50 milliseconds, this difference can be possible might be due the availability of free memory space and processor speed of used machine (computer) at processing time.

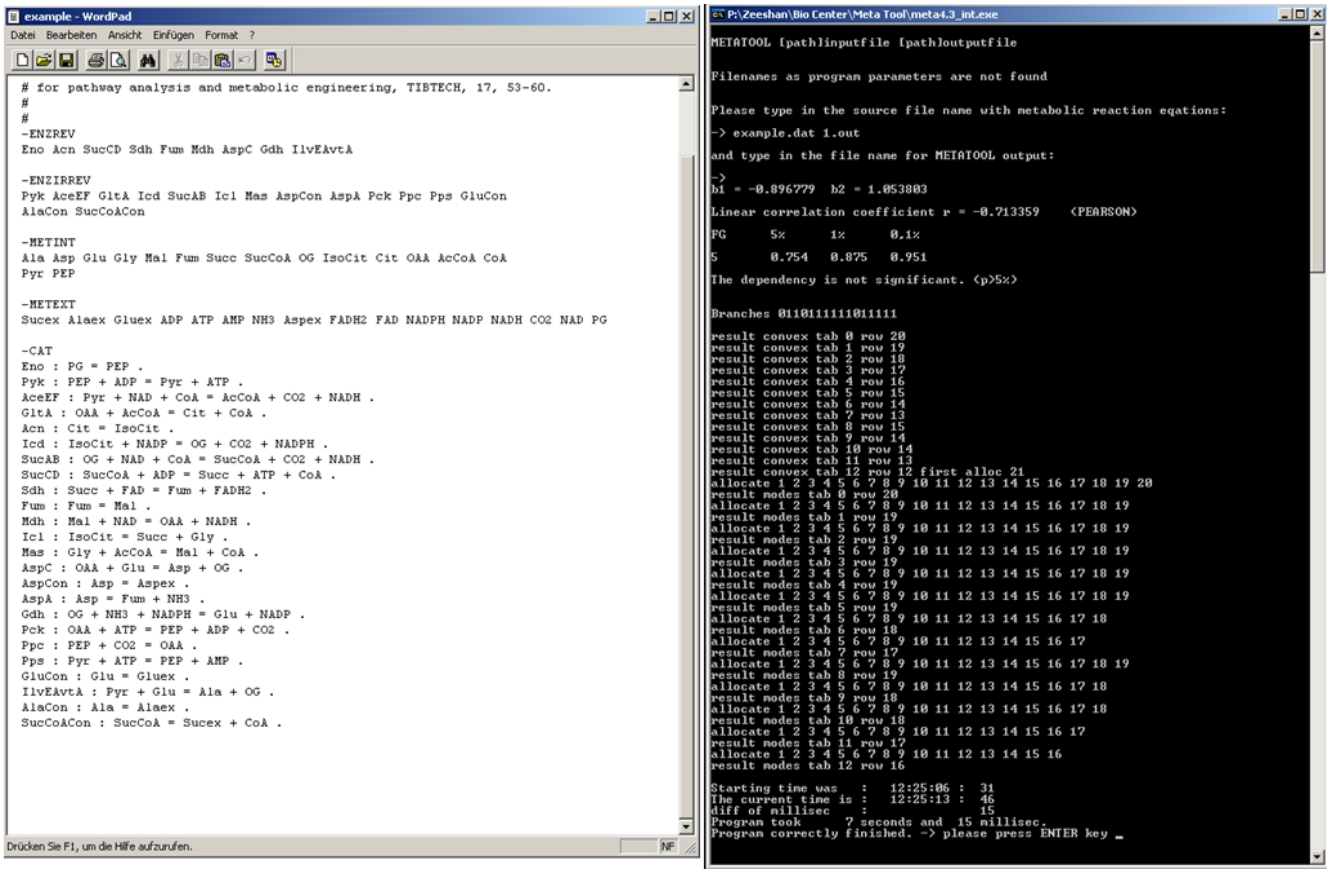


Figure 4: METATOOL 4.3 Software.

Figure Legend. 4: The graphical user interface of version 4.3 (integer executable) is shown on a linux operating system. The left shows an example input file (citric acid cycle and lower glycolysis). Standard terms for the metatool algorithm to calculate elementary modes concern irreversible and reversible enzyme reactions (left, top), internal metabolites (to be balanced), external metabolites (not to be balanced as source or drain metabolites) as well as the list of involved biochemical reactions. Abbreviations denote the different metabolites and enzymes*. On the right the start of the calculations of metabolite are shown (black background window). The stochiometric matrix with all enzyme reactions and internal metabolites is solved by combinations of enzymes which balance all internal metabolites they use or consume to obtain balanced flux modes which

are elementary (can not be decomposed any further without hurting the balancing condition). The calculations terminate after about 7 seconds. Hitting the enter key will then allow further analysis and inspection of the output file which includes all elementary modes as well as the transformations achieved by the calculated pathways in terms of external metabolites converted into other external metabolites as products. Metatool is constantly developed further and versions for LINUX and windows platforms are available.

Targeting the objective of providing efficient elementary flux mode calculations with the application of a new and fast algorithm for scripted input files from shared libraries, a new platform independent version of METATOOL (5.0) has been developed, capable of providing fast elementary flux mode calculations, tackling larger reaction systems, integrating with other third party tools and able to run using GNU Octave and MATLAB environments. Along with all new features METATOOL 5.0 is also capable of processing input files of previous METATOOL versions.

Flux mode calculations are in addition covered with increasing performance by a number of related software. To cite one example, the Null-Space algorithm proposed by Urbanczik and Wagner calculates elementary fluxes of chemical reaction systems based on stoichiometry matrices 20 times faster than the previously used algorithm by Schuster (Urbanczik et al. 2005). Correspondingly, this is used in the developed version of METATOOL (5.0) (Axel et al. 2006). It also internally distinguishes stoichiometric matrices and performs exact representation of numbers using integer data format. Moreover METATOOL (5.0) is also capable of computing structural invariants like conservation relations, enzyme subsets and fits a power law to the connectivity distribution of metabolites.

METATOOL communities are facing two known and explicitly mentioned challenges i.e.

- Parsing of METATOOL input files, to make them useable between DOS and UNIX based systems, currently uses third party utilities unix2dos and dos2unix. This can pose a challenge for feeding in the last line of the input file.

- Determination of the math program can fail, leading to different consecutive errors because Octave and MATLAB are not 100% compatible with each other.

Regarding quality of the output, METATOOL 5.1 with its provided example data set can be compared to METATOOL 4's (Figure 5). Besides the elementary modes, the convex basis and resulting conversions of external metabolites the program gives node frequencies, linear correlation coefficient, significant dependency value, stoichiometric matrix, enzymes and reactions.

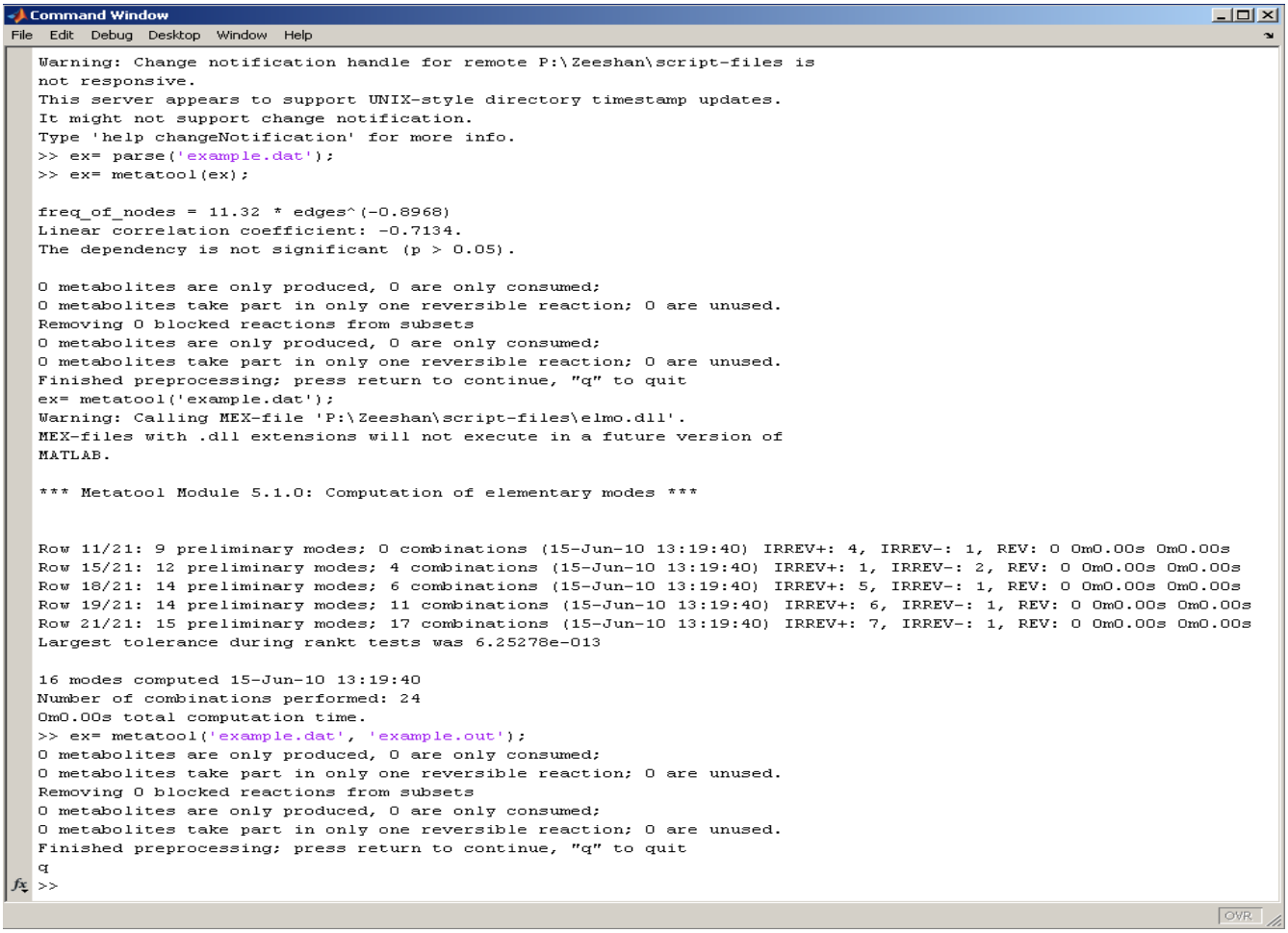


Figure 5: METATOOL 5.0 Software.

For comparison, METATOOL 4.3 shows no difference regarding linear correlation coefficients and dependency significant value ($p > 0.05$) but gives a less explanatory output regarding the stoichiometric matrix, enzymes and reactions. METATOOL focuses on elementary mode calculations and performs here well. However, for larger metabolic networks there is the challenge of combinatorial explosion yielding very large numbers of modes to calculate. A number of further developments change their strategy to avoid this.

Thus the Palsson group first changed algorithm and calculation so that already modes are calculated one by one before the complete list of enzyme combinations is established. This allows sampling over the full network, the so-called uniform sampling approach (Barrett et al., 2009; Price et al., 2004). Furthermore, the Schuster group recently established the concept of calculating modes only in a subnetwork and extending them over the rest of the network (effective flux modes). As these algorithms do not really calculate the elementary modes completely, it is difficult to compare them in a fair way, however, these are certainly highly advantageous in large networks where direct calculation of all elementary modes is no longer possible. To cite another advantageous development, (Centler et al. 2010) calculate flux modes consecutively in a parallelized, highly efficient routine.

2.1.4 BioOPT

BioOpt is developed in C++ and uses integer linear programming principles to perform flux balance analysis (Figure 6). Using as input reactions, constraints, external metabolites and an objective function, it calculates all internal mass balance fluxes, reduced costs and shadow prices depending on the constraints and objectives defined by the user. BioOpt´s linear programming identifies the best set of gene deletions for a given objective function value, implements an exhaustive combinatorial search for gene deletions and includes a basic sensitivity analysis (Cvijovic et al. 2010).

To operate, BioOpt takes a composition of reactions, constraints, external metabolites and an objective function as input, the resulting output is presented in Figure 7. Reactions are the set of reactions of the model, constraints give reaction bounds (lower / upper), external metabolites are the metabolites with no mass balance and two kinds of objectives are offered i.e. MAXIMIZE and MINIMIZE. The resultant information consists of the information about calculated objective value, reactions, internal fluxes, reduced cost, metabolites and shadow price.

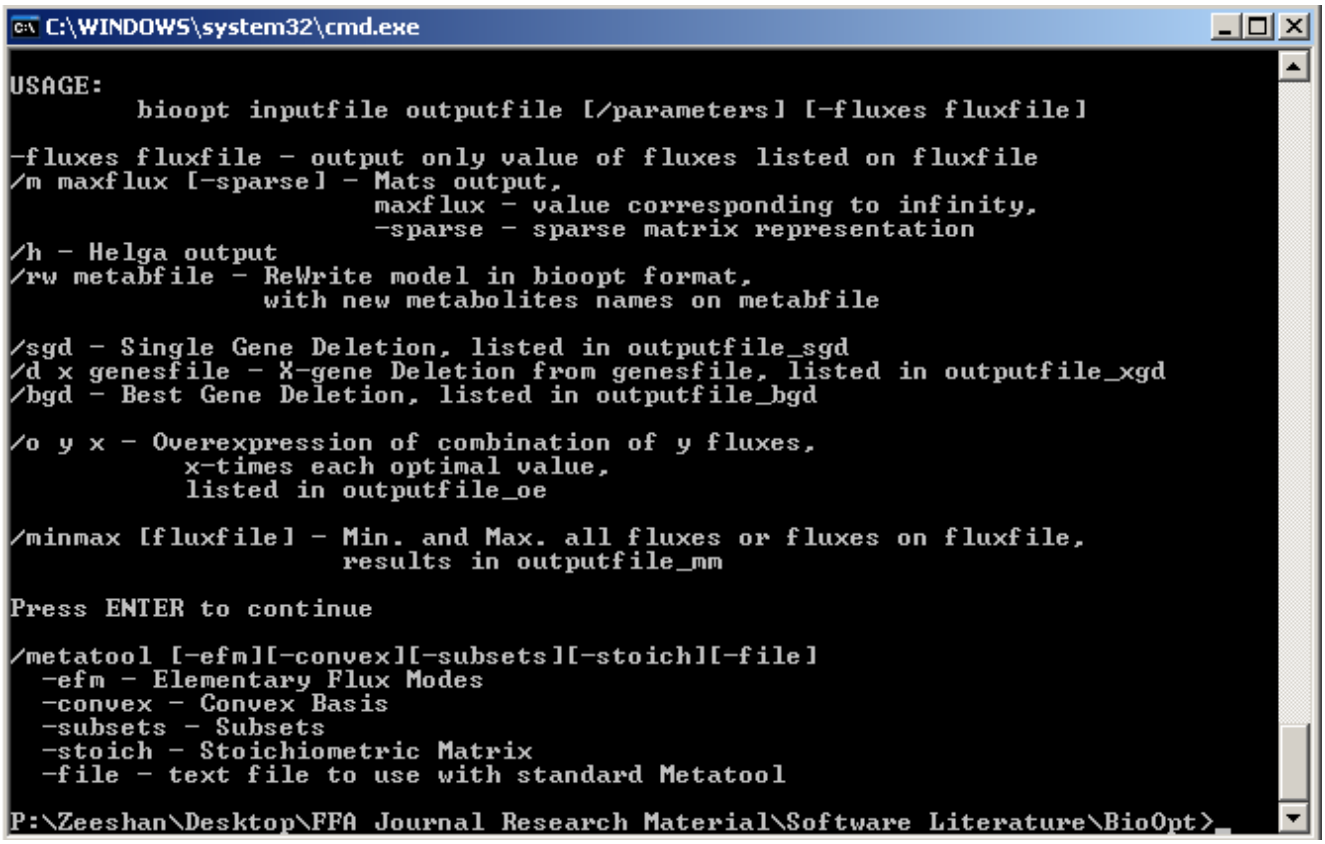


Figure 6: BioOPT Application.

Figure Legend 6: Command line interface uses windows operating system. BioOPT´s standard format corresponds with metatool, calculates also elementary flux modes, convex basis (a smallest set of elementary modes describing all metabolic states). Options include subsets (enzymes linked always together in metabolic flow) as well as incorporates optimizing conditions for the calculated flows (option minmax, middle), calculates gene deletion results (option sgd), suggests optimal gene deletions (bgd option, middle) as well as overexpression of certain pathways (/o x y option, middle).

Furthermore BioOpt is able to use as a third party tool METATOOL 4.3, to compute the null space matrix, elementary modes and other structural properties of biochemical reaction networks (convex basis, subsets and stoichiometric matrix), by directly passing the input to METATOOL.

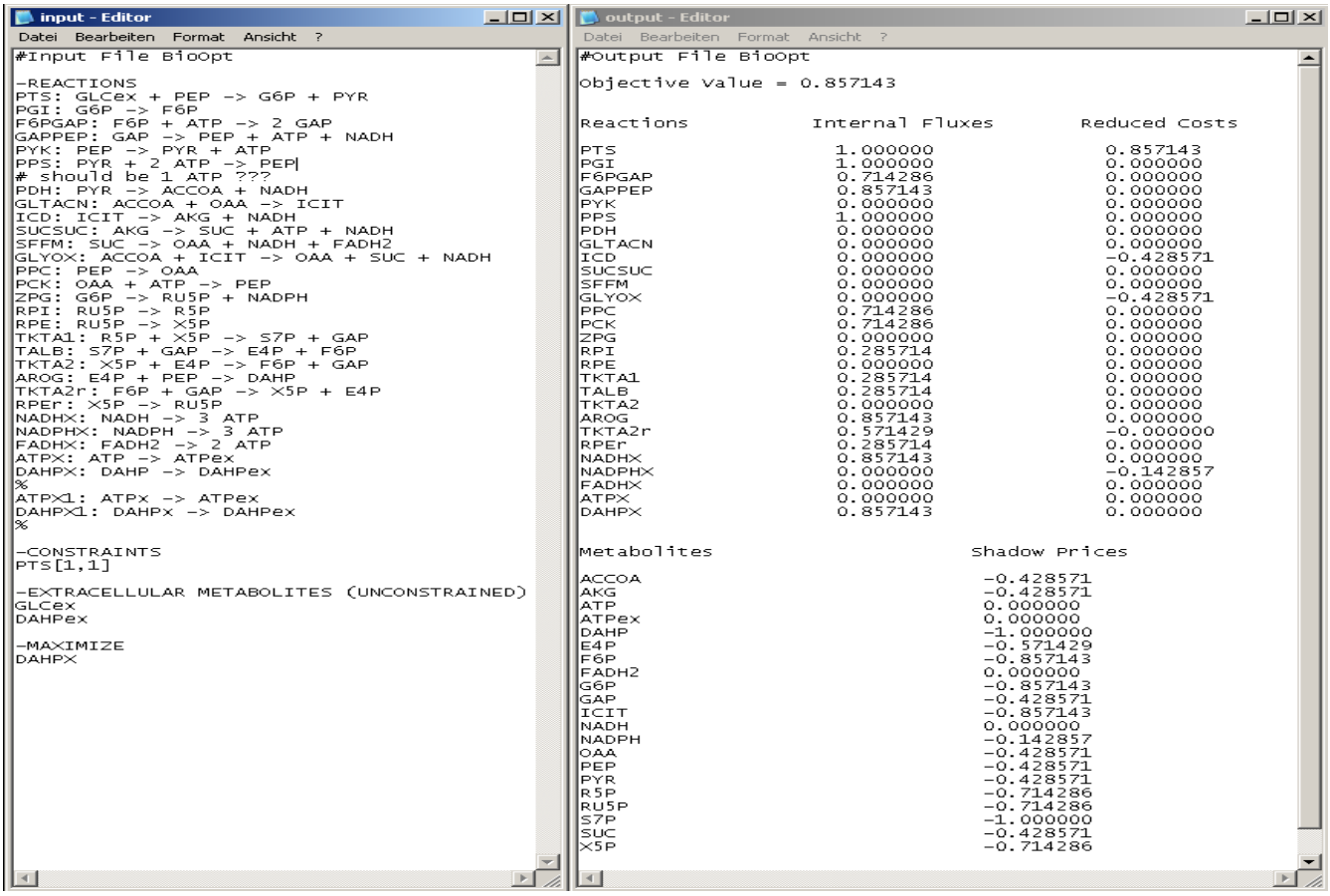


Figure 7: BioOPT Input and Output.

Figure Legend 7: Input format (right) is similar to Metatool, arrows indicate reactions. Processing of additional constraints (here a Phosphotransferase System (PTS), lower part) as well as maximization of metabolites are possible (DAHPX, given at the bottom). Internal fluxes (right window) are calculated, additional constraints and maximization of a metabolite are taken care of using costs and shadow prices.

2.1.5 Fiat Flux

FiatFlux is a MATLAB based software application developed for flux ratio analysis, strives to be user friendly and was one of the first publicly available software for flux ratio analysis (Zamboni et al., 2007) in labeling mixtures, including various substrates.

FiatFlux is divided into two different and independent modules i.e., analytical metabolic flux ratio analysis and ^{13}C-constrained flux analysis (Wiechert 2002), to take advantage in computing metabolic flux ratios exclusively from MS data in the RATIO module and to estimate net carbon fluxes within a

comprehensive model of metabolite balances from measured extracellular fluxes, previously determined flux ratios, and biomass requirements (Zamboni et al. 2005).

The overall internal work flow of FiatFlux starts working with a textual input file (*.txt) consisting of reactions, ratios and biomass precursors. Reactions are unique identifiers, ratios are the equality constraints ('=','>','<') and biomass precursors is the list of growth rate dependent withdrawals of metabolites in µmol/gCDW. These parameters are input for FiatFlux – Netto to compute flux using measured extracellular rates and to estimate error using flux ratio from ^{13}C-labeling (Fischer et al. 2004). At first mass balances and flux ratios are identified and then an appropriate method is selected. Furthermore, depending upon the active set of constraints and reactions, the input system can be undetermined, determined and overly constrained. In case of undetermined, possible fluxes should be estimated and then the objective function should be maximized. If the system is determined, manual analytical skills are needed to have good results, and if the system is constrained based then it can be solved by best fit optimization.

FiatFlux is software platform with two main modules, Ratio and Netto. Ratio estimates error using Flux ratio from ^{13}C labeling where as Netto estimates flux using measured extracellular rates. FiatFlux is third party software dependent, as it requires NetCDF tool to prepare an input file ('*.cdf') based on Network Common Data Form (netDFC standard) from input text file, which then will be converted to FiatFlux Ratio input file format ('*.FF') for flux analysis using a MATLAB programmed module CDF2FF.

The input format for the FiatFlux Netto module is also different and requires additional data conversion for data input (from '*.txt' to '*.m') and processing using another MATLAB based external module CREATEMOD. The FiatFlux Ratio creates a matrix consisting of total ion count and searches for known compounds (based on their predefined fragmentation pattern) and extracts MDVα (mass isotopomer distribution vector) for every matched analyte.

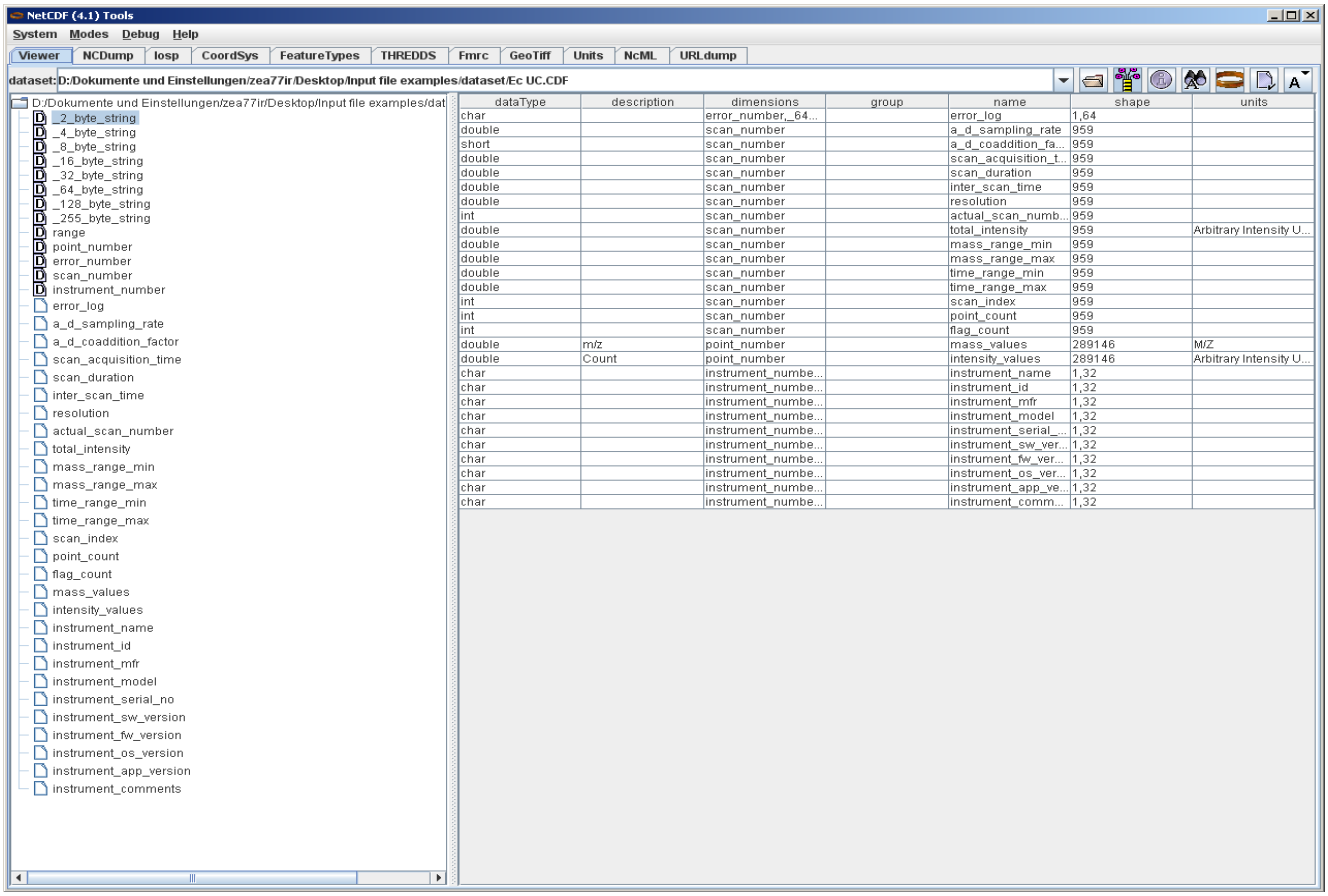


Figure 8: NetCDF input data conversion.

Figure Legend 8: NetCDF is a software tool to prepare an input file (‘*.cdf’) based on network Common Data Form (netDFC standard) from input text file, which then will be converted to FiatFlux Ratio input file format (‘*.FF’) for flux analysis using a matlab programmed module CDF2FF.

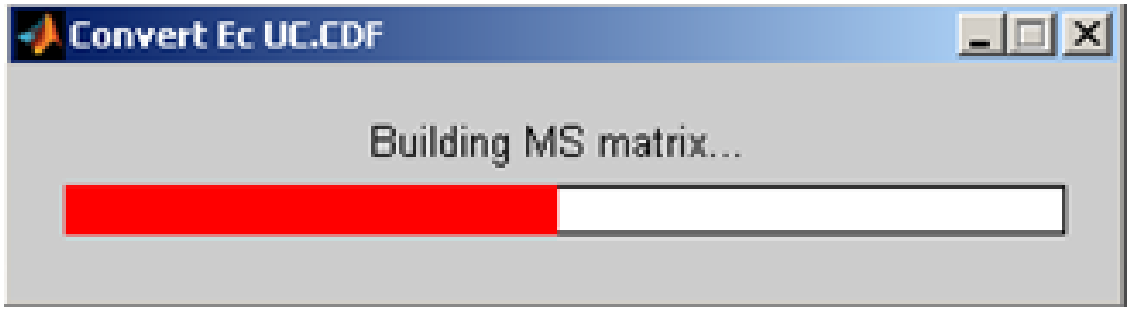


Figure 9: FiatFlux – Concerting CDF to FF.

Further MDVA (mass isotope distribution vector) is obtained from MDVα, which then is used to estimate MDVM (mass distribution of precursors) in central carbon metabolism. Furthermore, final flux ratios are also estimated from MDVM that should be equal to the substrate.

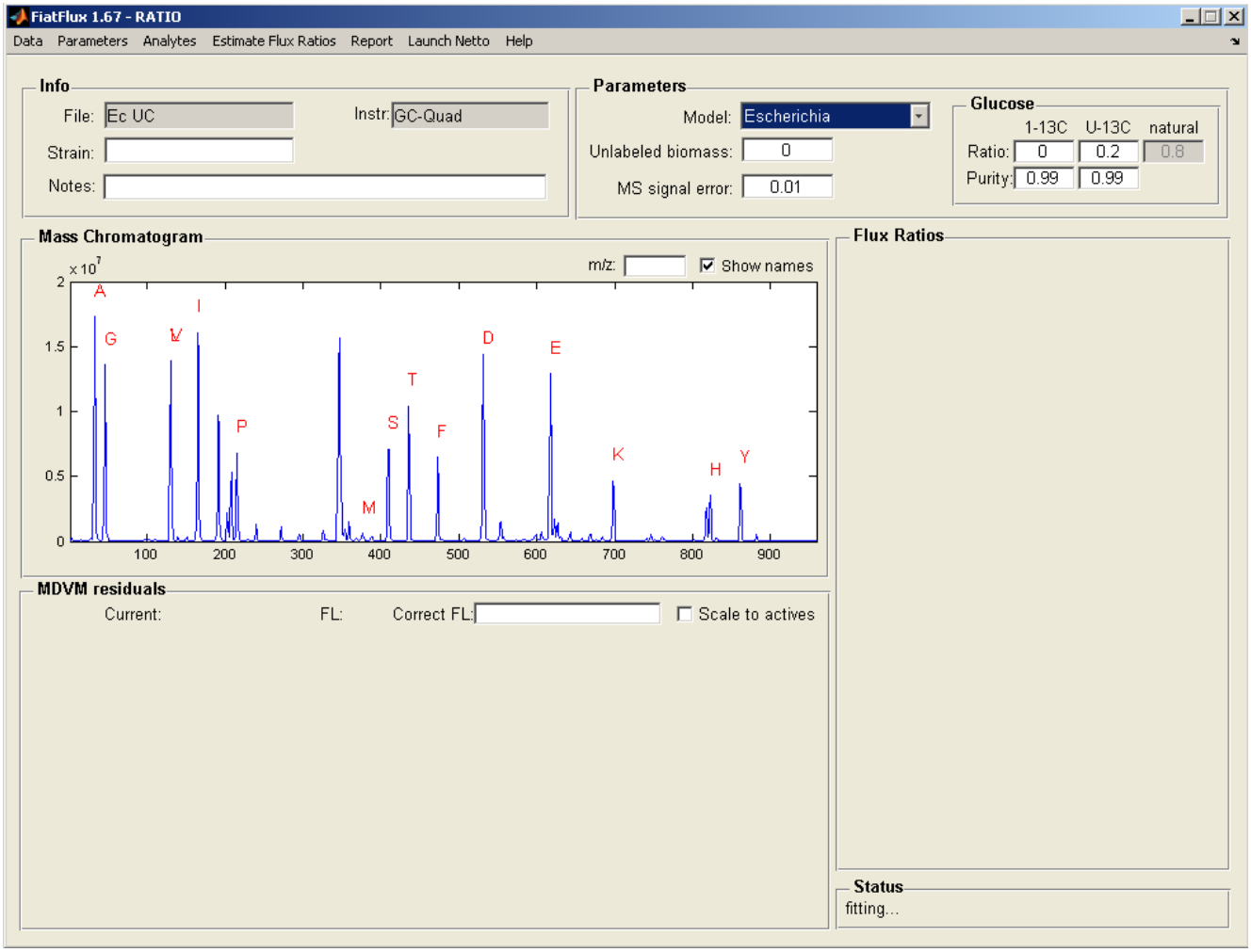


Figure 10: FiatFlux Ratio – Data analysis.

Figure Legend 10: Data analysis. FiatFlux is a Matlab based software application developed for flux ratio analysis. FiatFlux Ratio is one of the main modules of FiatFlux software application to estimate error using Flux ratio from 13C labeling. FiatFlux Ratio creates a matrix consisting of total ion count and searches for known compounds (based on their predefined fragmentation pattern) and extracts MDVα (mass isotopomer distribution vector) for every matched analyte. As shown in Fig. 6, it also draws a mass chromatogram of m/z values considering total ion counts for each scan.

CREATEMOD can be used manually as well as can also automatically be executed from FiatFlux ratio graphical user interface or called from the FiatFlux Netto interface. FiatFlux Netto estimates absolute net

fluxes using the reaction network and flux balance analysis from the stoichiometric model leading to ^{13}C-constrained flux balancing. FiatFlux Netto provides a platform to integrate metabolite balances and ^{13}C-constraints (equal or unequal). Furthermore, new metabolic models can also be manually constructed.

- To execute a textual data set based on reactions, ratios and biomass the following procedure is used by FiatFlux.
- Textual data is input to NetCDF, to convert it to common standard data form (Figure 8).
- Then output file "EC UC.cdf" is inputted to the CDF2FF (Figure 9).
- Resultant EC UC.ff is inputted to FiatFlux Ratio.
- To estimate extracellular flux, first a *.mod file is created and used in FiatFlux Netto.

The output consists of two different parts i.e. FiatFlux Ratio (Figure 10): $^{1\text{-}13}C$ ratio and purity, U-13C ratio and purity, model, unlabelled biomass, MS signal error, mass chromatogram, MDVM residual and flux ratios. FiatFlux Netto (Figure 11) consists of active and reversible reactions, metabolites, constraints and ratios, furthermore it also allows you to manually set extracellular fluxes.

Besides the known benefiting features in the available version of FiatFlux there are some unresolved issues:

1. Solution quality compared to other software applications (calculation is fine, the quality of presentation may be improved)
2. FiatFlux is claimed to be a user friendly software application, helpful for especially non-specialists, but at the same time data input procedure for FiatFlux is quite complex and third party tool dependent (which require at least basic knowledge of MATLAB and NetCDF), that even a specialist will have problems in only preparing an input file for flux analysis using FiatFlux.

3. FiatFlux is an open source on request and a license based freely available software application but at the same time it is quite expensive, as it requires the MATLAB software application as prerequisite.

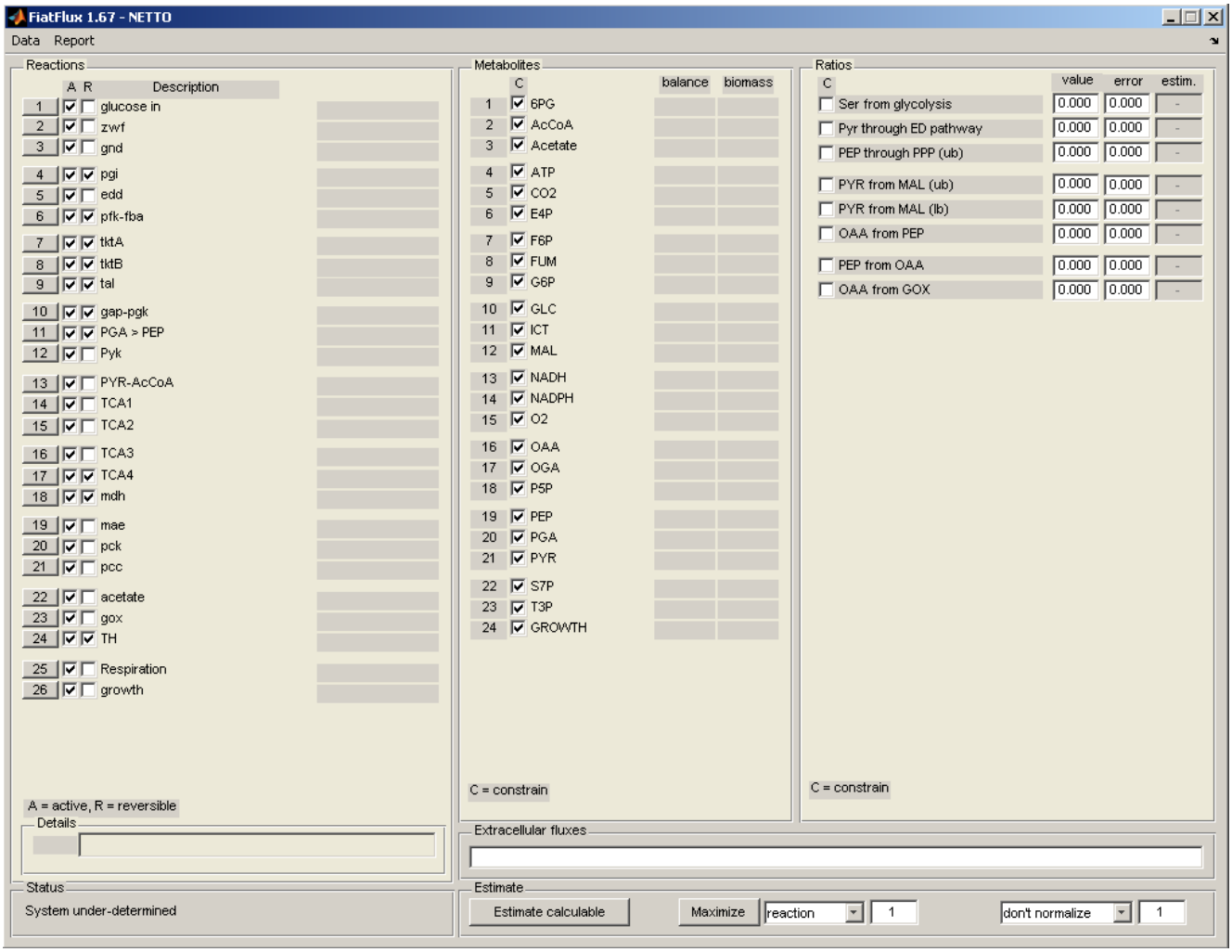


Figure 11: FiatFlux Netto – Data analysis.

Figure Legend 11: The Netto module estimates fluxes using measured extracellular rates from 13C isotopologue labeling and exploiting the underlying reaction network according to the stoichiometric model, so called 13C constrained flux balance. It provides a platform to integrate metabolite flux balance analysis and 13C-labeling based constraints.

There are some known issues, need to be resolved, explicitly presented by the authors (Zamboni et al. 2005) i.e.

a. In FiatFlux, user supervision is necessary only when MS signals are low, saturated, or overlapping. This affects the ion statistics of the corresponding fragment and results in relatively high residuals after inferring MDVM from the MDVA,

b. FiatFlux RATIO is not compatible with MS/MS product ion scans.

c. Only operating system Microsoft Windows is preferred because some minor problems were encountered using MATLAB's graphic user interface with Linux.

2.1.6 Rematch

Rematch is the first web-based tool capable of metabolic network model construction, store, sharing (by exporting into ^{13}C-FLUX and Systems Biology Markup Language (SBML) file formats) and integrating carbon mappings for ^{13}C-metabolic flux analysis.

1. Import a model or create a new model by specifying the substrate, product metabolites and their molar ratios or stoichiometric coefficients for each metabolic reaction.

2. To resolve conflicts between the nomenclature and augmented reactions with carbon mappings (if available) in the model and existing information in database, fully/semi automatic matching is performed amongst user given reactions and reactions stored in database.

3. Using BMV viewer visualize the metabolic network.

4. Export model in SBML, ^{13}C-FLUX or stoichiometric matrix formats.

5. Share exported model by declaring it public in ReMatch platform.

Following above stated five steps, the dataset “Blank_etal_2005_S.cerevisiae_on_glc_corrected” is first analyzed and visualized using Rematch as shown in Figures 12 and 13.

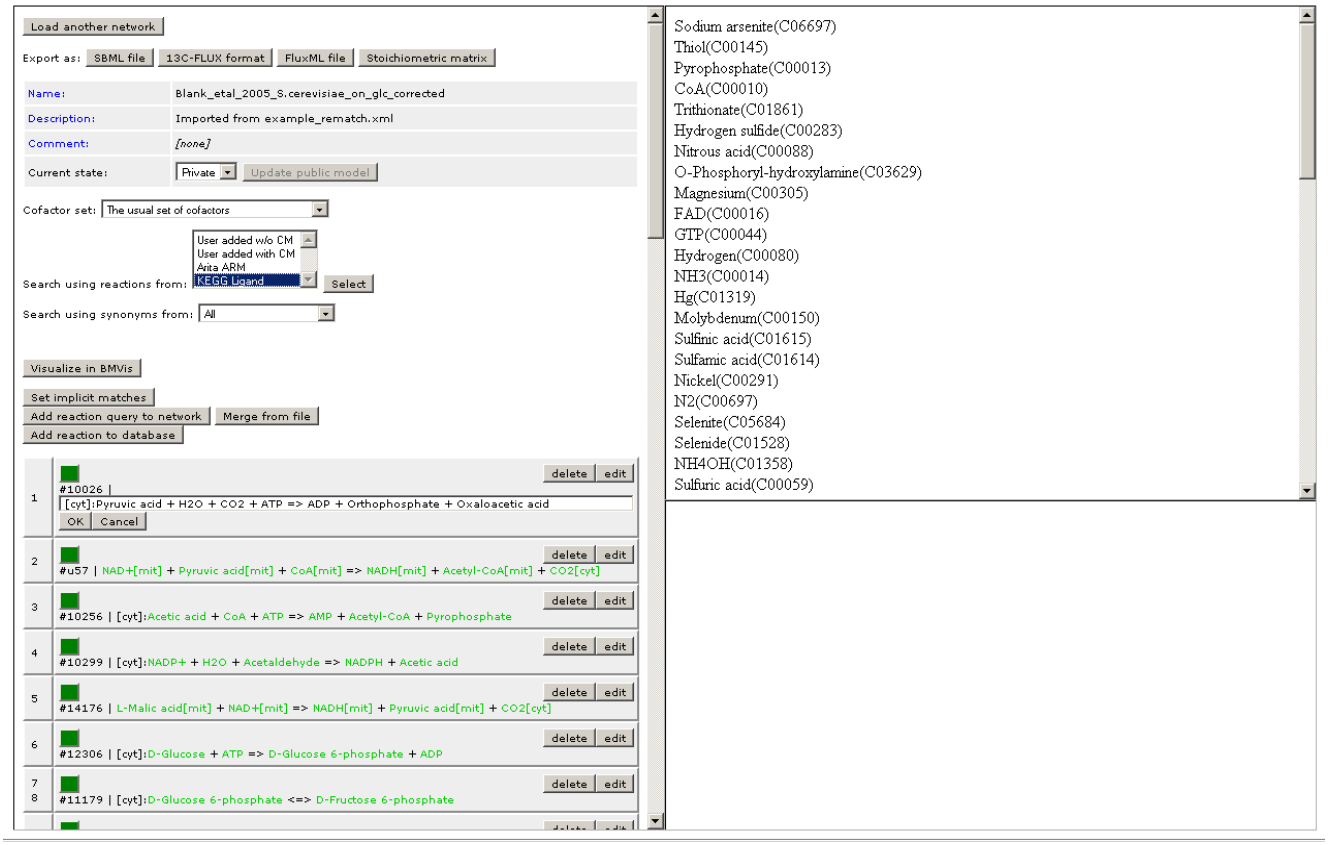


Figure 12: Data Analysis using Rematch.

Figure Legend 12: We show the standard example given for Rematch "Blank_etal_2005_S.cerevisiae_on_glc_corrected". Rematch is a web-based tool* capable of metabolic network model construction, storing, sharing and integrating carbon mappings for 13C metabolic flux analysis. It combines user developed models and several comprehensive metabolic data resources into a common repository for metabolic network models. It also allows combining user developed models from several comprehensive metabolic data resources into a common repository for metabolic network models. To provide better understanding of metabolic network models, Rematch is capable of generating stoichiometric matrix and visualizations. Main workflow of ReMatch consists of the following five steps: (Esa Pitk¨anen. et al. 2008)**.**

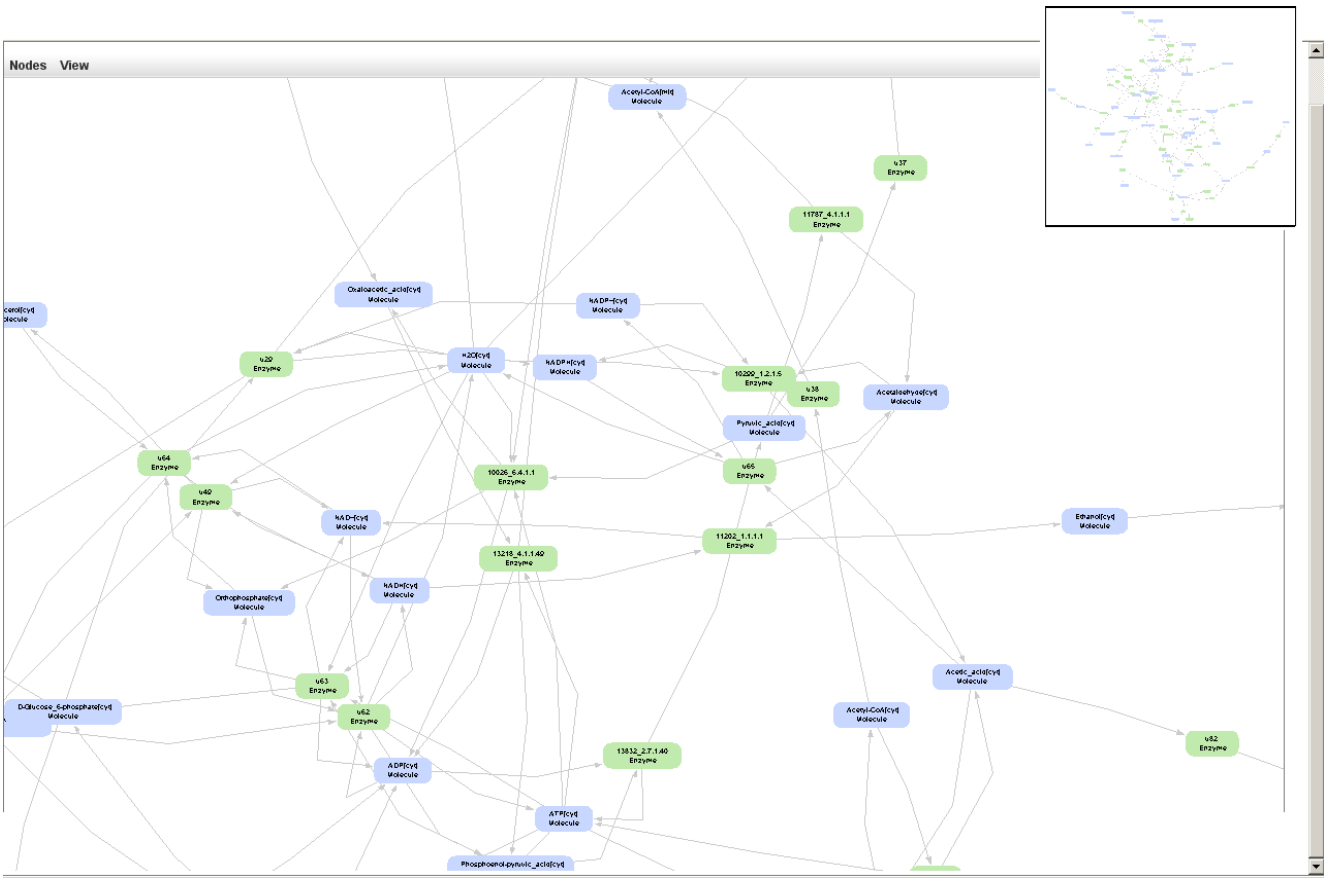


Figure 13: Rematch – Data Visualization (Network).

Figure Legend 13: Shown is the visualization by Rematch of a small example model (from phosphor-sugar metabolism) with enzymatic reactions (green nodes) and metabolites (blue nodes) in its viewer BMVis.

2.1.7 VANTED

VANTED is a platform independent Java-based software tool that provides a framework for visualization of experimental (-omics) data and statistical analysis (Junker et al. 2006). The user can integrate complex structured data sets and connect several values to one single network element by presenting them as e.g. line- or bar-charts. Supported input and output network formats are e.g. Graph Modelling Language (GML), SBML and Pajek.NET.

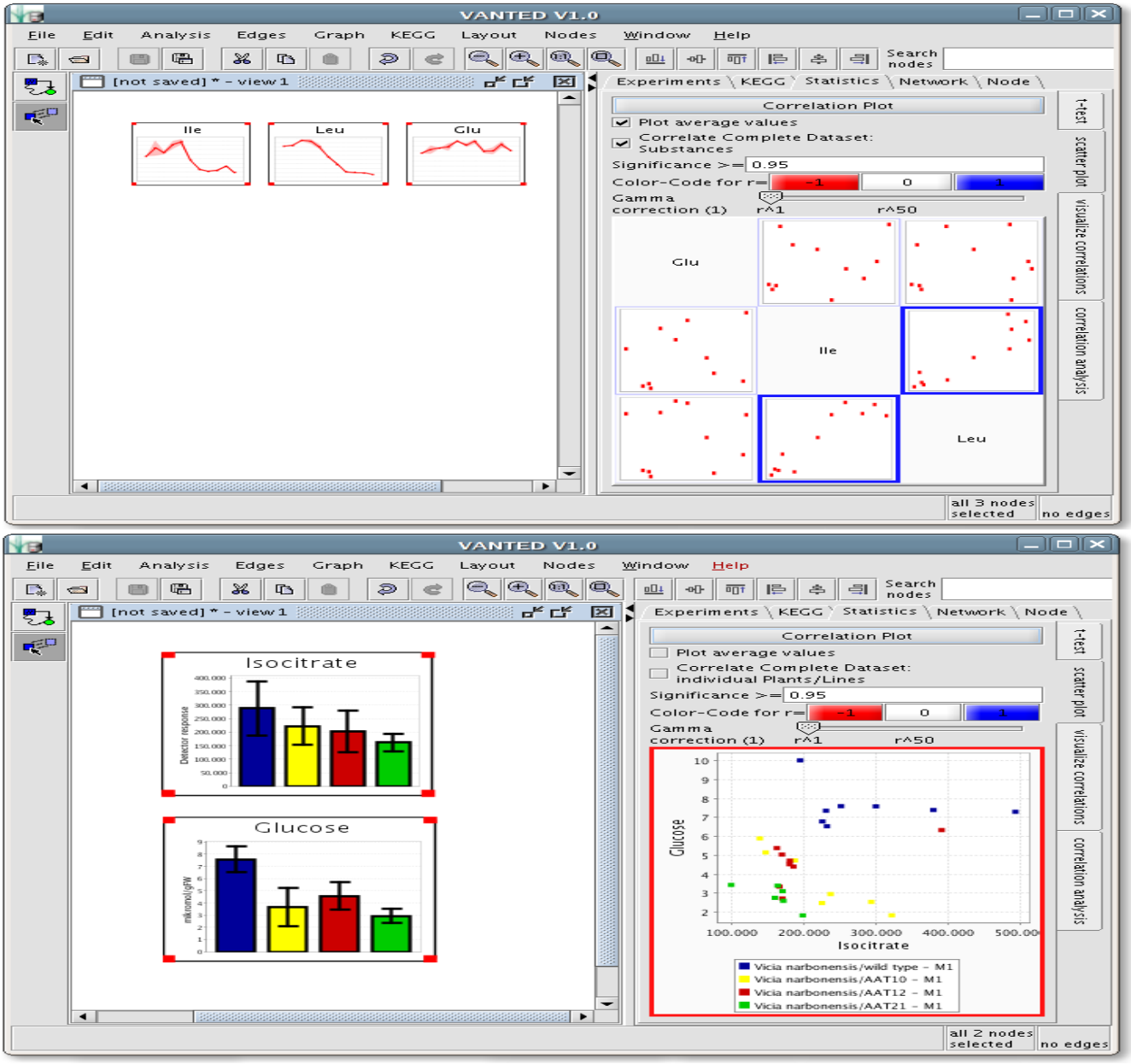


Figure 14: Network analysis using VANTED.

Figure Legend 14: Top: amino acids isoleucine, leucine and glutamic acid are compared as time series data (left) of a single mapped dataset. The right window analyzes the inter-correlation between the data for these branched-chain amino acids including filter options to identify high correlation/anticorrelation. Bottom: Scatter Plot created from two non-time series measurements of four different plant samples (blue, yellow, red, green) with detector intensity shown in upper insert window and resulting concentrations shown at the bottom window. Right: detailed correlation analysis of citrate and isocitrate for these samples See < http://vanted.ipk-gatersleben.de/index.php?file=doc119.html > for more information.

VANTED provides the visualization of flux distributions from various sources with the recently developed add-on FluxMap (Rohn et al., 2011). Basically, FluxMap visualizes flux distributions via different edge thicknesses and provides several parameters to adapt the appearance of a flux distribution as e.g. the global edge thickness, the arrowhead/-tail ratio, and the style of reactions nodes. Besides, quality of flux measurements (e.g. confidence) is illustrated via different edge colors, as shown in Figure 14.

Input of flux data is established via a structured Excel sheet template in which the user can allocate flux data to different conditions and time points. Thus, FluxMap allows switching through and comparing different flux distributions. As FluxMap is specialized on illustrating net fluxes we suggest to combine the visualization routine with the software YANAsquare in case when analyzing and visualizing data from elementary mode analysis.

2.1.8 YANAsquare

The Java-based software YANAsquare provides the elementary mode specific visualization of biological networks by e.g. distinguishing internal and external species with different node styles and colors, as shown in Figure 15. Likewise VANTED, it provides model set-up and modification with the choice to do a KEGG import for a rapid retrieval of biological networks. Input and output formats are further SBML or METATOOL. Internal elementary mode calculation is performed using METATOOL and resulting modes are presented in a tabular format as net reactions equations or enzyme equations. Data input of other elementary mode applications is not available up to now. For elementary mode visualization with efmtool we suggest e.g. BioOpt (Cvijovic et al. 2010). YANAsquare provides dynamic analysis and visualization of elementary modes by mapping gene expression data, proteome data sets or external metabolite data (Schwarz et al. 2007) on a biological network and calculate the corresponding elementary mode activity set.

Estimated enzyme activities are visualized. Furthermore, YANAsquare calculates flux distributions for each elementary mode activity set by a genetic algorithm and visualizes them by different edge thicknesses. The fit of the experimental data to a metabolic network is also an interesting challenge, for instance we recently compared the genetic algorithm with a steepest descent strategy and this is particular

efficient to fit data in larger metabolic networks (200-300 enzymes; Liang et al., 2011; software YANAvergence).

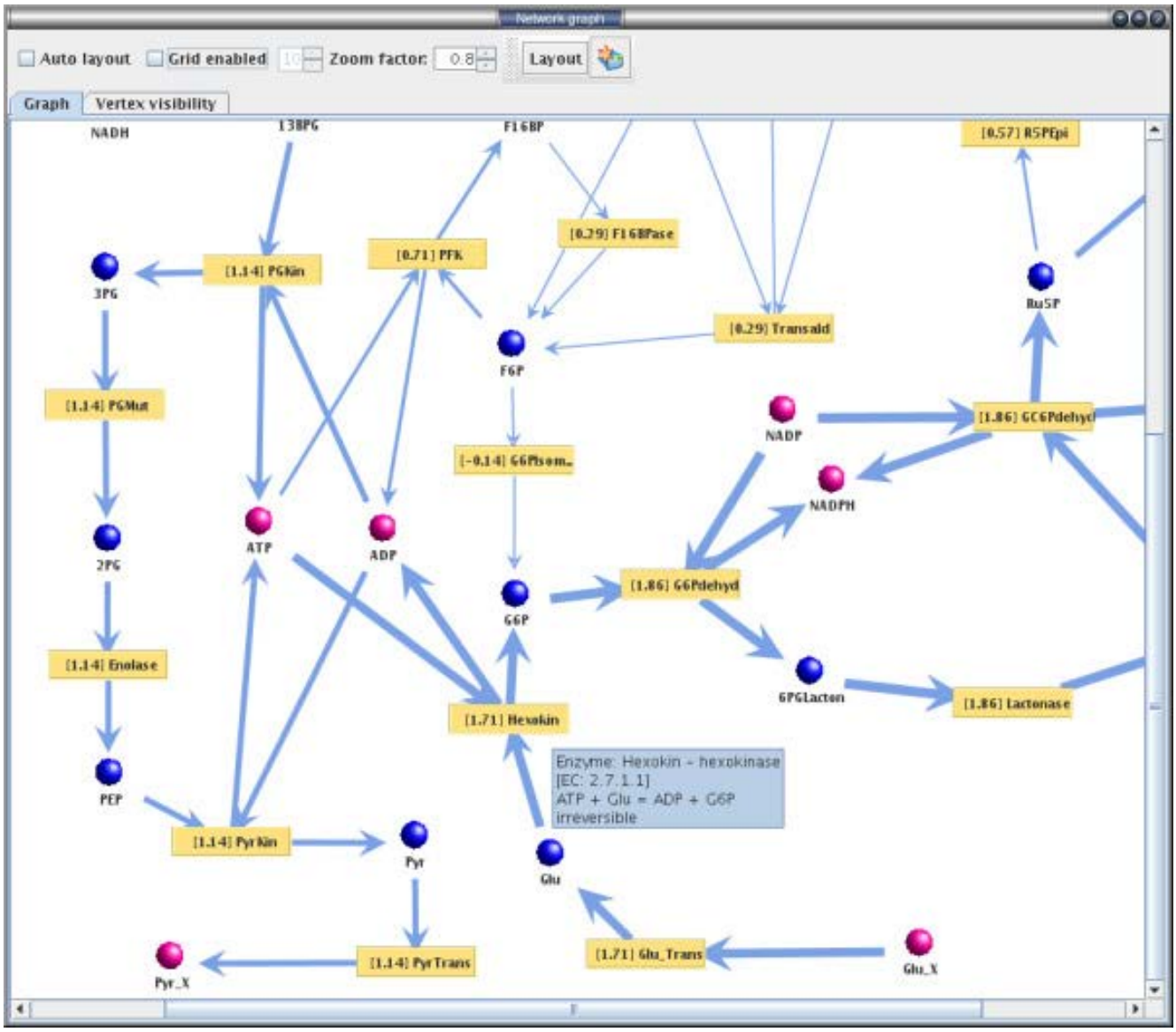


Figure 15: YANASquare graphical output.

Figure Legend 15: YANAsquare visualizes both elementary flux modes as well as actual flux strenths according to measured data (different thickness of the blue arrows, their direction indicates direction of measured metabolite flow). Red balls: Source or drain metabolites. Blue: Internal metabolites. Small pop-up windows (light blue background) give details on the involved enzyme reactions. The insert shown can be zoomed in and out to look at other regions of the system analyzed. To create the graphical output, YANAsquare calculates first elementary modes and processes next actual data regarding metabolite concentrations, protein or gene expression of involved enzymes. There is a steepest-descent routine as well as a genetic

algorithm available to fit the theoretical flux distribution to the actual observed data with low error. This works well up till medium sized networks (up to 50 reactions). To calculated fluxdistributions with experimental data fitting points for lager networks (up to 300 reactions) use YANAvergence (Liang,C. et al. (2011) Staphylococcus aureus physiological growth limitations: insights from flux calculations built on proteomics and external metabolite data. Proteomics, 11, 1915-35).

If the wish is to illustrate flux distributions from elementary mode calculations in context with other -omics data or to compare them with other flux measurements we suggest to integrate the resulting flux distributions from YANAsquare in VANTED/FluxMap. Hereby, one can make use of the software advantages in an integrated visualization approach.

2.2 Feature based Comparison

To evaluate the software from a computer scientist's point of view, made a feature based comparison amongst the applications including type of application (desktop / web / database), third party dependence for execution, reusable application, user friendly GUI, platform independent, text based input format, standards of input format, text based output format and standards of output format, visual output presentation, evaluation standards, used methodology and ease of use (Table 1).

1. Desktop Application; to see whether it's a desktop application.

2. Web Application; to see whether it's a web application.

3. Database Application; to see whether it's a database application which maintains data and provide efficient data manipulation system.

4. Third party tool dependent for execution; to see whether the concerned application need some other application to execute independently to process input data sets or it needs some other third party software tool or technology to run.

5. Reusable application; to see whether the concerned application is flexible enough, so then, it can be used by any other application at back end.

6. User friendly GUI; to see whether the concerned application's graphical user interface is much friendly that a new user can easily adopts and use it.

7. Platform independent; to see whether the concerned application is platform independent and can be executed on any operating system.

8. Text based input (format); to see whether the concerned application takes data input in simple text format or any specific input is needed.

9. Any standard input format; to see whether the concerned application's input file format is based on any existing standard data format or not.

10. Text based output format; to see whether the concerned application produces output in text data format or not.

11. Any standard output format; to see whether the concerned application output data format is based on any existing standard data format or not. Furthermore, apart from its own output data format is it capable of also producing output in any other standard format.

12. Visual output presentation; to see whether the concerned application is capable of producing demonstrable data visualization of input or output data.

13. Successful evaluation; to check whether the claimed feature(s) and advantage(s) of concerned application is (are) successfully validated in experiment.

14. Used methodology (ies); to know which methodology is used in concerned application.

15. Well explained for use; to authenticate (with personal opinion) that the explicitly provided information about concerned application by authors is helpful in validating it.

Key Features / Software Applications	**C13**	**Metatool 4.3**	**Metatool 5.1**	**BioOpt**	**Fiat Flux**	**Rematch**	**Biolayout Express**	**Vanted**	**Yana Square**
Desktop Application	Yes (Matlab)	Yes	Yes (Matlab)	Yes (C++)	Yes	No	Yes	Yes	Yes

Web Application	No	No	No	No	No	Yes	No	No	No
Database Application	No	No	No	No	No	Yes	No	No	No
Third part tool depended for execution	yes (Matlab)	No	Yes (Matlab)	Yes (Meta tool 4.3)	Yes (Matlab)	No	No	Yes	Yes
Reusable application	No	Yes	No	Yes	No	No	No	No	Yes
User friendly GUI	No	No	No	No	No	Yes	Yes	Yes	Yes
Platform independent	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes
Text based input (format)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Any Standard Input format	No	No	Yes (SMBL)	No	Yes (CDF)	Yes (SMBL)	Yes (Owl & SIF)	Yes (GML, SBML Pajek.NET.l)	Yes (SBML)
Text based output (format)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Any Standard Output format	No	No	No	No	Yes (CSV & Excell)	Yes (SMBL)	No	Yes (Excel)	Yes (SBML)
Visual output presentation	No	No	No	No	Yes	Yes	Yes	Yes	Yes
Successful evaluation	No	yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Used Methodology (ies)	Isotopic labelling	S.Schuster Algo.	Null Space Algo.	Mass Balance Eq.	Mass Distribution & Isotopic Labelling	Carbon Mapping	Markov Clustering	t-test, scatter plot matrix, Correlation Analysis	Metatool 4.3. S.Schuster Algo.
Well explained for use	No	yes	yes	yes	No	Yes	Yes	Yes	Yes

Table 1: Metabolic Flux Analysis; Software Comparison.

No.	Description
1	During metabolic flux analysis it is tedious to map carbon atoms from substrates into products, as required in model (Esa Pitk¨anen. et al. 2008).
2	The absence of good quality carbon mappings in a systematic fashion in current metabolic model repositories costs more time resource in new models construction for ^{13}C metabolic flux analysis (J. L. Snoep. et al. 2002).
3	Manually adding carbon mappings to stoichiometric model is a complex, time consuming and error prone task because even small central carbon metabolism models contain hundreds to thousands of carbon mappings (Esa Pitk¨anen. et al. 2008).
4	Online repositories of SMBL models with carbon mapping information and new intelligent applications are needed for rapid metabolic network model construction, carbon mapping and information sharing (Esa Pitk¨anen. et al. 2008).
5	Although the cross linking and data compatibility between different metabolite databases like KEGG LIGAND (M. Kanehisa. et al. 2008) is improved but still due to the heterogeneous naming conventions of metabolites in different databases it is a time consuming and complex task to construct bio-chemical reactions based network model (Esa Pitk¨anen. et al. 2008).
6	Integration, analysis and best quality visualization of disparate data types consisting of raw data expressions and large graphs comprising of many thousands of nodes and edges, needed to improve biological systems understanding mechanism (A. Theocharidis. et al. 2009).
7	Accurate prediction of metabolite concentrations is not possible till now.
8	Lack of calculation of flux distribution (by not changing the actual cell behavior) is still not possible.
9	Quantitative analytical approach is needed to completely analyze a cell of living organism to provide information about its metabolism (Bailey, J.E. 2001)
10	(Re)Construction and simulation of the complex in-formation based cellular function (K.J. Kauffman. et al. 2003).
11	Need to improve the tendency to accommodate high diversity of biomolecules.
12	Today many modeling approaches have been developed to analyze flux by quantification of metabolic flux but still these are unable to resolve high complexity dataset with maximum prepurification.
13	It's also not possible to identify unambiguous metabolite and mass isotopomers.

Table 2: Observed limitations during MFA software review.

Based on the comparison presented in Table 1, we conclude with some drawbacks of existing software based solutions (Table 2).

- Most of the applications are desktop based applications and some of them require additional third party software applications to even execute themselves e.g. Fiat Flux, C13 and METATOOL 5.1 require MATLAB. This dependency makes these solutions infeasible and expensive by requiring an additional costly software application (MATLAB), furthermore also expects the user to have some additional skills to use it which again consumes extra time as well.

- One another drawback of these discussed solutions is, except Rematch all applications do not maintain their history by providing a database for data management and manipulation, which in return not only makes application ordinary but also requires for the user to manually maintains the history of data including input data sets and results.

- Most of the application's graphical user interface is not friendly, which makes an impression as they didn't follow the recommended design pattern for graphical user interface design.

- Most of the applications are not following any standard format for data input and output presentation, not only increases the burden for the user to first convert the data into respective software's input format but also makes it difficult for the potential user to compare it with the output of any other software application.

- Data visualization is also often not available, good examples are, however, ReMatch, YANAsquare and Fiatflux.

- One very important drawback found in some of the applications, is that, they are not very well explained e.g. FiatFlux, due to which it's sometimes very hard for a new user to understand the actual behavior of the software.

- Another aspect is the difference in parameters and input data sets between the considered applications, though often the same kind of analysis i.e. Flux Analysis, needs to be performed.

The comparison shows clearly space for further software application development including steps towards an optimal user friendly graphical user interface, platform independence, database management system and third party independence especially in case its desktop application. Other improvements could aim at generality and standard data input formats, improve visualization of not only input data set but also analyzed results. We hope, with the implementation of these suggestions, metabolic software applications will become more professional, cheap, reliable and attractive for the user.

2.3 Analysis

Based on the extensive comparison and including computational features (Table 3), we suggest the following best mode for a metabolic flux analysis including visualization of measured and predicted fluxes, pathways and metabolites (availability, weblinks are given in Table 4):

In most of the metabolic flux analysis tools a metabolic network is a prerequisite. For setting up a metabolic network model especially for complex and genome-scale networks we suggest using ReMatch providing various output formats, like SBML, ^{13}C -Flux format or a stochiometric matrix format. In contrast, the KEGGbrowser routine from the YANAsqare package (Schwarz et al., 2007) allows rapid (few minutes) set-up of pathways and metabolic maps according to information from KEGG database (including further data stored in the same format). For the concept of constraint-based modeling and most notably flux balance analysis different software tools like the COBRA toolbox, BioOpt and Classical and Dynamic FBA exist. Differences between these applications are mainly the possibility to perform additional calculations beyond FBA for example gene deletions studies implemented in BioOpt. If one intends to study a dynamic behavior of metabolic transient steady-states in addition to FBA then Classical and Dynamical FBA may be the best choice.

As flux balance analysis provides only one single flux distribution which may possibly be not the only solution for the biological system under studied conditions a deeper understanding can be achieved by studying the elementary flux modes. Here we recommend METATOOL for small to middle-size networks. In ^{13}C-constrained flux analysis the calculation of net fluxes and flux ratios depending on the experimental setup is applied. In ^{13}C based metabolic flux analysis a software combination of ^{13}C and FiatFlux may be useful to calculate absolute fluxes and ratios of fluxes with different methods and cross-validate results. For flux visualization we recommend using VANTED and its add-on FluxMap for the

visualization of net fluxes from various sources. In combination with YANAsquare, the visualization routine is extended by elementary mode visualization and analysis.

The VANTED framework allows to embed flux data in the systems wide analysis as this software supports visualization of various (-omics) experimental data types. YANAsquare provides as an analysis and visualization software for elementary mode calculation the dynamic analysis of elementary modes and the estimation of flux distributions for elementary mode activity sets. We propose to either analyze elementary modes in YANAsquare or integrate estimated flux distributions from YANAsquare into VANTED/FluxMap to embed them in the systems wide analysis. Taken together, the recommended software combinations allow complete analysis of metabolic flux starting from the scenario of all potential pathway fluxes up to dynamic visualization of different metabolite concentration changes.

Software	**Advantages**
C13	1. Estimate fluxes satisfying stoichiometric constraints 2. Resolve limited enrichments by isotope balances around carbon atoms 3. Computes deviation between fluxes and between fractional labeling
Classical and Dynamic FBA	Classical FBA 1. Reads stoichiometric matrix 2. Provides prima solution to objective function and optimal value, for all metabolic fluxes 3. Analyses dual file to extract shadow prices for all the metabolites 4. Reconstructs metabolic network by systemic mass balance and reaction capacity constraints 5. Draws metabolic (network) map 6. Displays the metabolites with zero, negative, and positive shadow prices in different colors 7. Display the reactions that have a metabolic flux equal to zero in different colors based on the value of the reduced cost 8. Adjusts objective function of the metabolic network at run time 9. Performs robustness analysis by analyzing robustness without altering the objective function, adjusting parameters and drawing robustness diagram 10. Creates two dimensional diagram based on the mapping of

		phenotypes of a metabolic genotype based on the flux value for genotype of two different reactions, so called Phase Plane
	11.	Creates Isoclines consisting on the same value for the objective function or a certain flux
	12.	Calculates and compares flux distributions of two different reactions (files).
	13.	Provides a graphical user interface for providing two dimensional network based visualization.
	Dynamic FBA	
	1.	Predicts metabolite concentrations and reutilization of acetate
	2.	Allows the incorporation well characterized kinetic expression.
	3.	Provides metabolic engineering for making strategies to design network.
Metatool	1.	Gives numbers of internal metabolites and reactions
	2.	Parses reaction equations and translates them into a stoichiometric matrix
	3.	Identifies kernel or null space
	4.	Provides fastest elementary flux mode calculations
	5.	Tackles larger reaction systems
	6.	Integrates with other third party tools
	7.	Runs using GNU Octave and Matlab environments
	8.	Capable of computing structural invariants like conservation relations, enzyme subsets and fits a power law to the connectivity distribution of metabolites.
Bio Opt	1.	Calculates all internal mass balance fluxes, reduced costs and shadow prices
	2.	Identifies best set of gene deletions for given objective function value
	3.	Implements exhaustive combinatorial search for combinations of gene deletions and over expression of fluxes and basic sensitivity analysis
	4.	Capable of using a third party tool (Metatool 4.3)
Fiat Fux	1.	User friendly and first publicly available software for flux ratio analysis
	2.	Computes metabolic flux ratios exclusively from MS data in the RATIO module
	3.	Estimates net carbon fluxes within a comprehensive model of

		metabolite balances from measured extracellular fluxes, previously determined flux ratios, and biomass requirements.
	4.	Estimates error using Flux ratio from ^{13}C labeling
	5.	Estimates flux using measured extracellular rates.
Rematch	1.	First web-based tool capable of metabolic network model construction, store and sharing
	2.	Integrates carbon mappings for ^{13}C metabolic flux analysis.
	3.	Allows combining user developed models from several comprehensive metabolic data resources into a common repository for metabolic network models.
	4.	Generates stoichiometric matrix and visualizations.
	5.	Resolves conflicts between the nomenclature and augmented reactions with carbon mappings (if available) in model and existing information in database, fully/semi automatic matching is performed amongst user given reactions and reactions stored in database.
	6.	Visualizes the metabolic network
	7.	Exports in use model in SBML, ^{13}C-FLUX or stoichiometric matrix formats
	8.	Shares exported model by declaring it public in ReMatch platform
Bio Layout Express 3D	1.	Imports data from various standard graph formats
	2.	Handles up to 500 to 1,000 expression arrays,
	3.	Cluster of graphs of up to 30,000 nodes and 2 – 3 million edges
	4.	Construct network graphs
	5.	Converts a 2D representation to a 3D graph
	6.	Calculates correlation matrix
	7.	View and navigate results in the 3D interface
	8.	Alters aesthetic characteristics of the 2D and 3D interfaces
	9.	Cluster Graph using the MCL, by defining the granularity of the clustering by setting high and low inflation values
	10.	Identifies genes of interest and export list of selected genes for further use
	11.	Mine selected genes for overrepresentation of classes, the number of calculations necessary to perform this task inevitably makes this process slow
	12.	Works with other data formats and edit networks using GraphML

		input file
	13.	Select nodes (by a number of ways) for the editing of their properties e.g., activation (A), inhibition (I), translocation (T) or logic nodes
	14.	Render pathways in 2D and 3D modes
VANTED	1.	A platform-independant Java-based software tool that provides a framework for visualization of experimental (-omics) data and statistical analysis.
	2.	Allows user to integrate complex structured data sets and connect several values to one single net-work element by presenting them as e.g. line- or bar-charts. Sup-ported input and output network formats are e.g. GML, SBML and Pajek.NET.
	3.	Provides the visualization of flux distributions from various sources with the recently developed add-on FluxMap.
YANA square	1.	The Java-based software provides the elementary mode specific visualization of biological networks by e.g. distinguishing internal and external species with different node styles and colors.
	2.	Provides model set-up and modification with the choice to do a KEGG import for a rapid retrieval of biological networks.
	3.	Takes input and produces output in SBML or Metatool format.
	4.	Performs Internal elementary calculation using Metatool
	5.	Calculates flux distributions for each elementary mode activity set by a genetic algorithm and visualizes them by different edge thicknesses.

Table 3: Observed advantages during MFA software review.

2.4 Results and Aspiration

Apart from observed and earlier discussed limitations, we have also concluded with some very important information, regarding the helpful contribution of the above software in many ways for metabolic flux analysis and visualization i.e. C13 can be used for metabolic flux analysis by stationary carbon isotope labeling on fractional enrichment data, Classical FBA can be used towards metabolic behavior analysis by writing a mass balance for each metabolite of a network, METATOOL can be used to calculate the elementary flux modes, BioOpt is available to perform flux balance analysis, FiatFlux can be used e.g. for flux ratio analysis, Rematch can be used for metabolic network model construction, store, sharing and integrating carbon mappings for ^{13}C metabolic flux analysis over the web. VANTED and YANAsquare are very helpful in visualization and analysis of networks containing experimental data and biological networks.

3 Least Square MIDA

Mass isotopomer distribution analysis (MIDA) measures mixtures of different molecules by quantification of the relative abundances of isotopomers in a molecular species typically by mass spectrometry. Mass isotopomer distribution analysis (MIDA) is an excellent approach to estimate rates of metabolite formation. Specifically, ^{13}C mass isotopomer distribution helps in metabolic network analysis for flux estimation by examining pathway activities and enzymes pass through by ancestor molecule (Bequette et al., 2006).

A crucial step in this experimental approach is the management of the complex experimental data including normalization of the mass isotopomer fractions for resolving the biosynthetic history of each metabolite under study. This section reports on an open-source software tool, the Least Square Mass Isotopomers Analyzer (LS-MIDA). From the original relative mass abundances of a given compound, it calculates the isotopomer enrichments in ^{13}C-labeled metabolites. The implemented design and graphical user interface is presented in Unified Modeling Language (UML). Efficiency is tested on different data sets by successful prediction of natural relative abundances, calculated relative abundances and of relative enrichments in ^{13}C labeled metabolites with visualization of the obtained results. Currently, only commercial software or user-specific approaches are available for the conversion of mass intensities (provided by the specific software implemented to the mass spectrometer) to the relative and molar isotopomer enrichments such as tandem mass spectrometric data computing for positional isotopomer distributions (Rantanen et al., 2002), measurements of mass distributions by mass spectrometry (Winden et al. 2002), isotopomer analysis using GC-MS (Christensen et al., 1999) (+Lee et al and our MS2X) and GC-MS analysis for isotopomer balancing (Dauner et al., 2000).

An improved method as an enhancement over existing approaches (Michael et al. 2008; Massila et al. 2008; Papageorgopoulos et al. 1999) is proposed to estimate isotopomer distributions of metabolites from experimental raw data produced by mass spectrometry. The provided features for primary data processing in LS-MIDA are not currently available in standard packages for metabolite modelling such as Metatool (Pfeiffer et al. 1999), Yanasquare (Schwarz et al. 2007), Gepasi (Mendes et al. 1993) or FiatFlux (Nicola et al. 2005) (here fluxes are predicted after the isotopologue data have been processed). Furthermore, LS-MIDA is the only tool providing a file based data management system for experimental metabolic mass

isotopomers based data. There is fast processing speed (only seconds) and calculation complexity scales O(n2) with the number of carbon atoms per isotopologue).

3.1 Methodology

The software treats experimental raw data from mass spectrometry. Specifically, mass intensities of metabolic products (typically ^{13}C labelled) are analyzed on the basis of their m/e values, and the number of C atoms in the given molecule, derivative or fragment thereof. Overall ^{13}C enrichment and the relative and molar contribution of isopomers is then calculated using Braumann's least squares algorithm (Brauman, 1966). It can also be used to analyze other isotopes than ^{13}C such as 15N Furthermore this method can deal with complex spectra based on fragmentation of molecules (e.g. OYX1; see notation below) containing in addition heteroatoms (atoms which may not be carbon or hydrogen itself but can occur in several different isotopes). These are challenges in fragmentation spectra (Brauman, 1966).

Considering that the GC-MS can detect only weight differences one is often not sure about the distribution of labeled (Y) or unlabeled (X) atoms. In the following, the algorithm calculates the partition of label regarding the results of no label (e.g. 000 for three carbons in alanine), one label (XXY for alanine, Y can be at any place in this chain of three carbons), two labels (XYY, the unlabeled atom X can again be at any place in alanine) and full labeling (all three carbons in the alanine are labeled).

The software has two parts: generation of an appropriate set of linear simultaneous equations and solution of these equations. The proposed matrix equation to calculate relative intensities of the fragments is

$$X = A^T (A^T * A)^{-1} P \qquad \text{(Eq. 4)}$$

Where in Eq. 4, X = Calculated Relative Abundance (Intensity values), A = Abundance Matrix of Natural Abundance Values and P = Actual Relative Intensities. The number of fragments contributes to natural abundance distributions. It is nevertheless worthwhile to look for contaminating fragment ions in the mass spectrum, as its impact is substantially corrected through the subtraction of natural abundance values. The proposed binomial expression (Lee et al., 1991) calculating natural abundances is

$$A = n! / [(i!) * (n - i)!] * P_o^{(n-i)} * P_1^{i} \qquad \text{(Eq. 5)}$$

Where in Eq.5, A = Calculated Relative Natural Abundance, n = number of fragments, i = count value, (loop from 0 till n-1) and p= proportion of labeled carbon. The reliability in the results of this technique depends upon a number of factors: (i) the analysis is based on the assumption that the fragmentation patterns for all heteroatom isotopes are identical (i.e., no isotope effect), (ii) the actual relative abundance of the heteroatom isotopes is known and (iii) the natural abundances are either known or measured.

The proposed equation to compute Absolute ^{13}C enrichment of both natural and relative abundances is

$$Abs\ ^{13}C = (\sum A_{0\ldots n} * n)/a \qquad \text{(Eq. 6)}$$

Absolute ^{13}C Enrichment is equal to the sum of all labeled isotopomers multiplied with their position number (0 to n), divided by number of carbon atom of the amino acid fragment. In Eq. 6, A = labeled isotopomer, n = 0 till the carbon atom of the amino acid fragment number, and a = carbon atom of the amino acid fragment number. Furthermore, to have better analytical measurement view, the percentages of outcomes of these three mathematical equations (natural abundance values, relative abundance values, absolute ^{13}C enrichment values and relative intensity values with respect to m/z value) are also computed. Using atomic mass value M_{-1} and M_{max} values are also computed.

3.2 UML Description

Today, what could be the better way of designing architectures of a newly proposed software application other than using UML? It is a modeling language, a well suited and the standard way of designing software application by creating different abstract models. UML is capable of facilitating software engineers stand alone and interconnected semiformal (Meta) design views for modeling software architectures (Medvidovic et al., 2002). Here software designs are created using UML principles to have better understanding of LS-MIDA in terms of its implementation, usage and working, Designed UML diagrams are describing over all feature based functionality, user accessibility, experimental data flow, internal system work flow, system sequence, involved component's integration and source code structure (Ahmed et al., 2012a). In this thesis we presents following LS-MIDA UML diagrams: Use Case, Data Flow, Internal Work Flow, System Sequence, Class and Component Diagrams. We are designing product line architecture (Ahmed, 2010a) to make software application flexible enough to easily adopt future updates and additional features.

3.2.1 Use Case Diagram

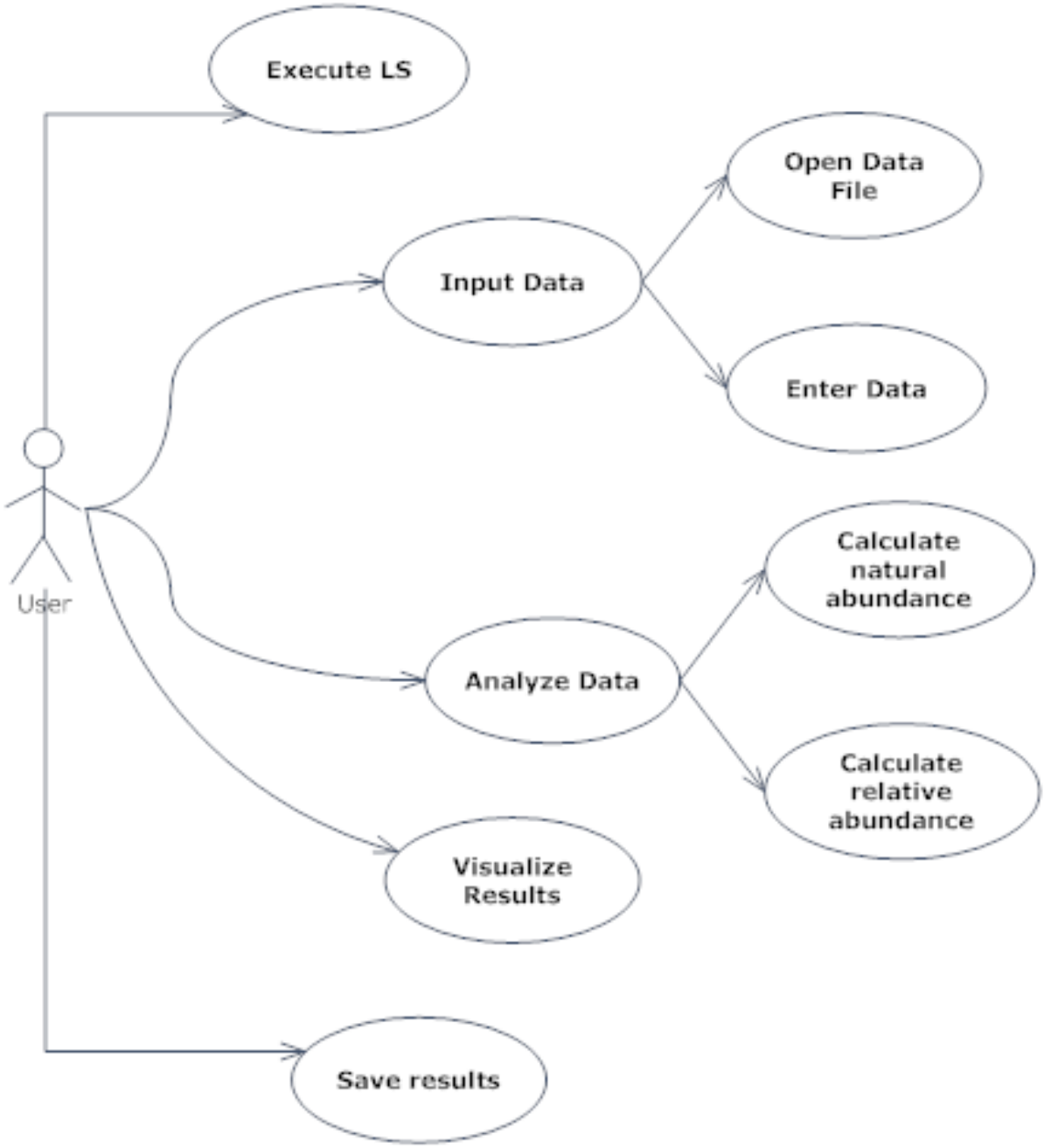


Figure 16: LS-MIDA; UML Use Case Diagram.

Figure Legend 16: Use case diagram of LS-MIDA is consisting of a User, five direct and four remote (indirect) activities.

Use case is the specific textual and visual method of presenting software application's functionalities comprising all ways of user system interactions (Jacobson et al., 1992). It consists of two main symbolic notations i.e. Actor and Activities. In most of the cases actor is either user or system itself as a remote actor. Activity is the event triggered by the system in response to the request by actor for some action.

We have designed a use case diagram (Figure 16) and explained in detail (Table 4). The designed use case diagram describes the user system communication for the isotopomers experimental data analysis, which consists of a user (actor), five direct activities (Execute LS, Input Data, Analyze data, Visualize Results, Save results) and four indirect activities (Open Data File, Enter data, Calculate natural abundance, Calculate relative).

Use Case	**Details**
Number	1
Name	LS-MIDA data analyzer
Application	LS-MIDA
Description	This use case consists of a User, five direct and four remote (indirect) activities. This describes the user (actor) system (LS-MIDA) communication for the isotopomers experimental data analysis.
Primary Actor	User (1 Actor)
Precondition	Software application successfully running.
Trigger / Events	1. Execute LS 2. Input Data 3. Open Data File 4. Enter data 5. Analyze data 6. Calculate natural abundance 7. Calculate relative 8. Visualize Results 9. Save results

Basic Flow	Basic flow consists of following steps: • Start software application • Enter input data by either loading from data file or by manually entering. • Analyze input data. • Visualize results • Observe predicted results (text and image). • Save obtained results for reuse.
Alternate Flows	Exception will be notified to the user.

Table 4: LS-MIDA; Use Case Description

This use case explains the basic flow of user system interaction starting with the execution of software application. At first user needs to enter the input (experimental isotopomer) data by either loading from data file or by manually entering which then can be analyzed, visualized, saved and reused.

3.2.2 Data Flow Diagram

The data flow diagram is the way of producing visual presentation of data flow inside a software application (Bruza and Weide, 1993). This data flow diagram presents internal data flow of LS-MIDA data analyzer (Figure 17). At first data is loaded from the Data File as input (Input/Output (I/O) Data), which is then analyzed by the system for the calculations of natural abundance values and relative abundance values prediction.

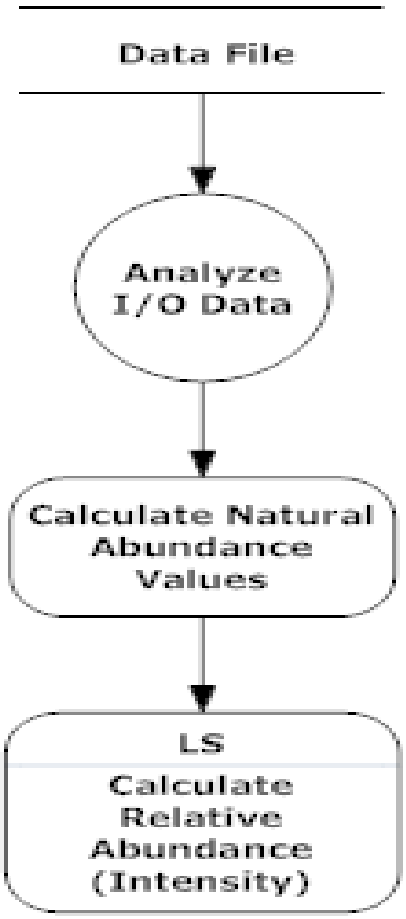


Figure 17: LS-MIDA; UML Data Flow Diagram.

Figure Legend 17: The data flow diagram of LS-MIDA is consisting of a File (Data File), one main Function (Analyze I/O Data) and two internal functions (Calculate Natural Abundance Values and Calculate Relative Abundance Values).

3.2.3 Internal Work Flow Diagram

Internal flow chart is also known as simply the Flow chart, a step by step visual representation of defined interlinked processes (operations) in a software application (Marilyn, 1978), categorized in different shaped boxes representing different kinds of operations connected by directional and unidirectional (associated) arrows.

The implemented mathematical procedure in LS-MIDA starts with the input (I/O) consisting of metabolite's experimental raw data observed during GC-MS i.e. metabolite information (e.g. name), m/e values, experimental relative intensity (R_i) values, mass fragment number and mass value (Figure 18).

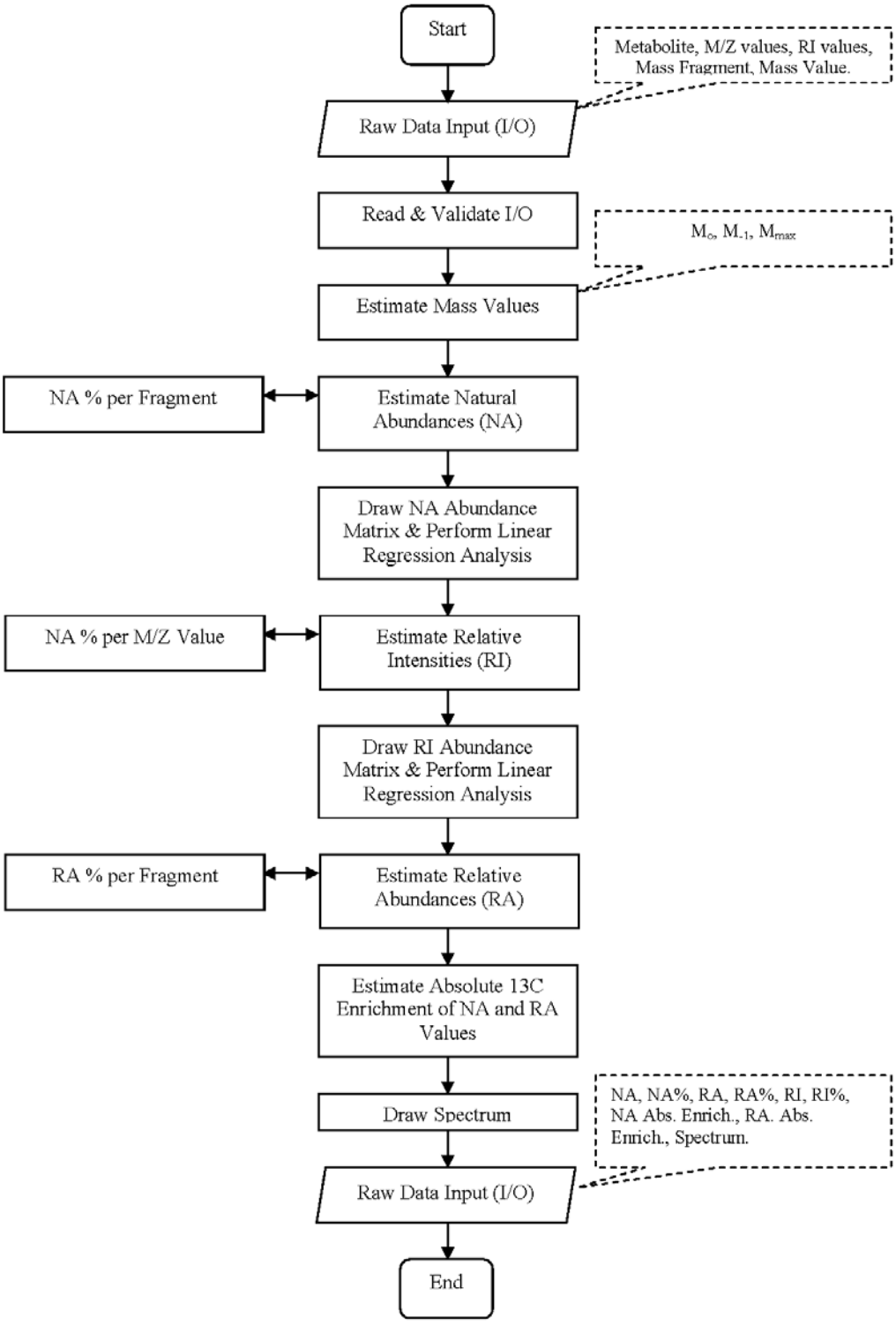


Figure 18: LS-MIDA; UML Flow Chart.

Figure Legend 18: Visual presentation of the unified mark-up language (UML) based flow chart. The implemented flow of operations performed during experimental data input, processing, analysis and visualization is given.

After I/O validation, at first the mass values M_o, M_{-1}, M_{max} are estimated to get the range to pick optimal intensities: These give expected mass M_o, lower (M_{-1}) and upper bound (M_{max}) where to expect the given fragment. Then using binomial expansion, relative natural abundances Na with percentages per each fragment are estimated. Next linear regression analysis is performed and abundance matrix is drawn with the application of Brauman's least square method using estimated Na values to estimate relative intensity values per each m/e value.

These estimated relative intensity values R_i are then used to perform, once again similar linear regression analysis by drawing the abundance matrix with the implementation of Brauman's least square method to estimate natural abundances Na and relative abundances Ra including their percentages with respect to each fragment. Using the calculated Na and Ra values, absolute ^{13}C enrichment is calculated for each amino acid. The output (N_a and R_a values) is presented in numeric format and in special notation format (based on the number of C atoms in the fragments). A graphic is drawn for the user based on m/e, experimental and relative intensity values

3.2.4 System Sequence Diagram (SSD)

The (SSD) represents a particular scenario (text or graphic) defined by use case, especially for transaction oriented systems (Latronico and Koopman, 2001). A SSD consists of actors (users), messages (methods) called by the actors, return values (optional, if any) and loop indicators. The main reason of using SSD is to explore the logic of multifaceted operations (procedures or functions).

The system sequence of LS-MIDA Data Analyzer consists of three sequential steps with individual tasks (Figure 19): The experimental data based information is given as an input file by the user. It consists of metabolite, mass to charge ratio (m/e) values, actual relative intensity values and number of fragments.

The software first validates the input data. They are next sent to Least Square for calculation of relative abundances. Least Square, at the third step, sends information based on number mass values and the number of fragments to the Binomial expression, which calculates natural abundances (n) and sends them to Least Square. With the use of natural abundances, an abundance matrix is then generated (An), applying as main algorithm the Least Square method. The Transpose of generated matrix is calculated (TAn), then the inverse of TAn times An (ITAn). In the last step relative abun-dances are calculated by

multiplying ITAn times TAn and actual intensity values. These calculated abundances (Ra) are sent to the graphical user interface of LS-MIDA including visualization of obtained results.

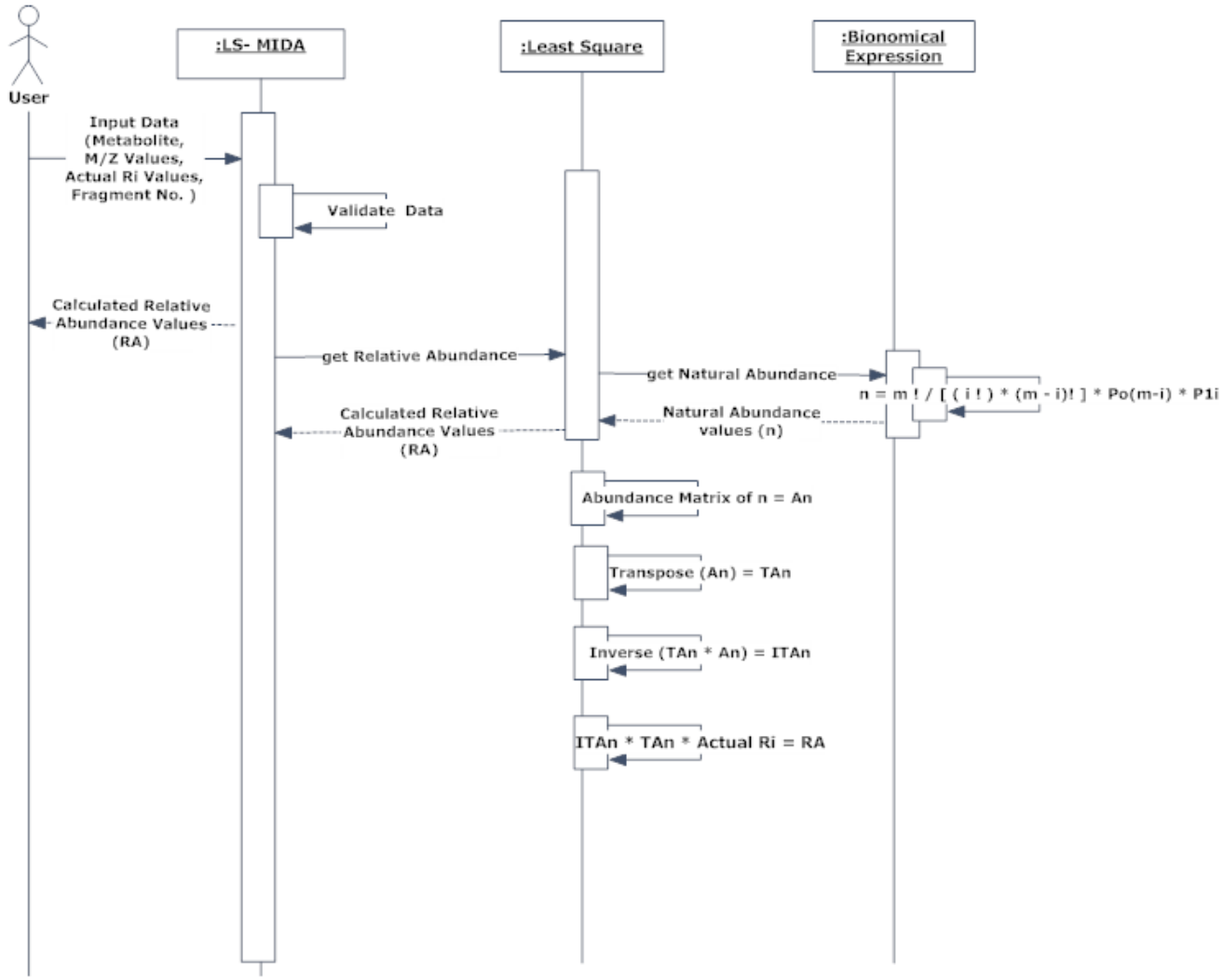


Figure 19: LS-MIDA; UML System Sequence Diagram.

Figure Legend. 19: The abstract system sequence diagram of LS-MIDA is consisting of three main steps (LS-MIDA, Calculate Natural Abundance Values and Calculate Relative Abundance Values) with several directing arrows in between.

3.2.5 Component Diagram

Component diagram is the visual presentation of assembled constituents representing structural relationship between service provider and consumer (Ambler, 2005). It allows the designer to confirm system's functionality to be implemented in the form of components using internal and third party

services (e.g. programming languages, libraries, executables, application programming interfaces, frameworks etc.) in terms of nature and behavior.

LS-MIDA is mainly consists of two main modules i.e. LS-MIDA Data Analyzer and LS-MIDA Data Manager (Figure 20). User can access these both modules to perform mass isotopomers distribution estimation and file based experimental data management. LS-MIDA, as whole application, is developed using an object oriented platform independent language C Sharp (C#) and Microsoft Dot Net technology.

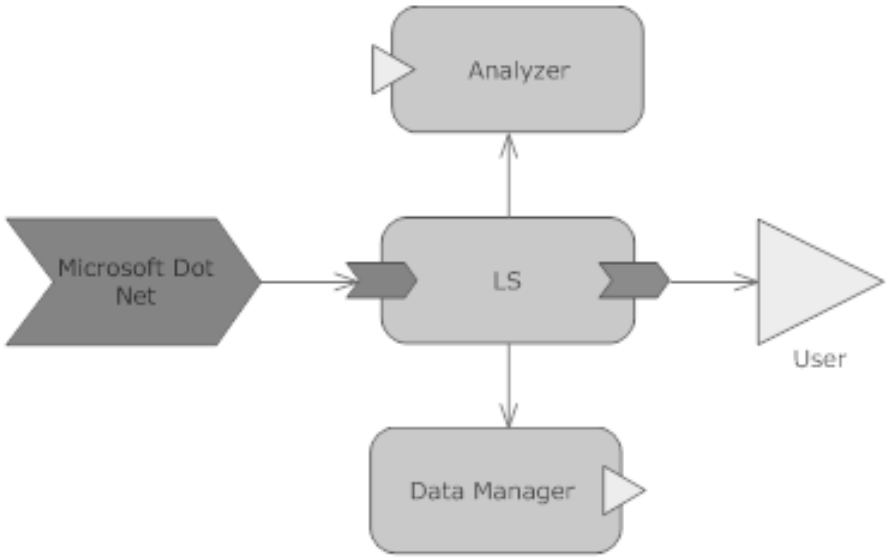


Figure 20: LS-MIDA; UML Component Diagram.

Figure Legend 20: The component diagram of LS-MIDA is consisting of one platform component (Microsoft Dot Net), one main component (LS), two subcomponents (Analyzer and Data Manager).

3.2.6 Class Diagram

Class diagram is the static representation of relationships between defined classes for the development of a software application (Berardi et al., 2005; Grant et al., 2006). The source code of developed LS-MIDA is divided into three namespaces i.e. SBEDA, System and ZEDGraph. SBEDA is the main namespace containing all related and newly developed source code classes, System is the by default namespace provided by C-Sharp language used during the software development, and this namespace is responsible for providing access to default language based controls and components.

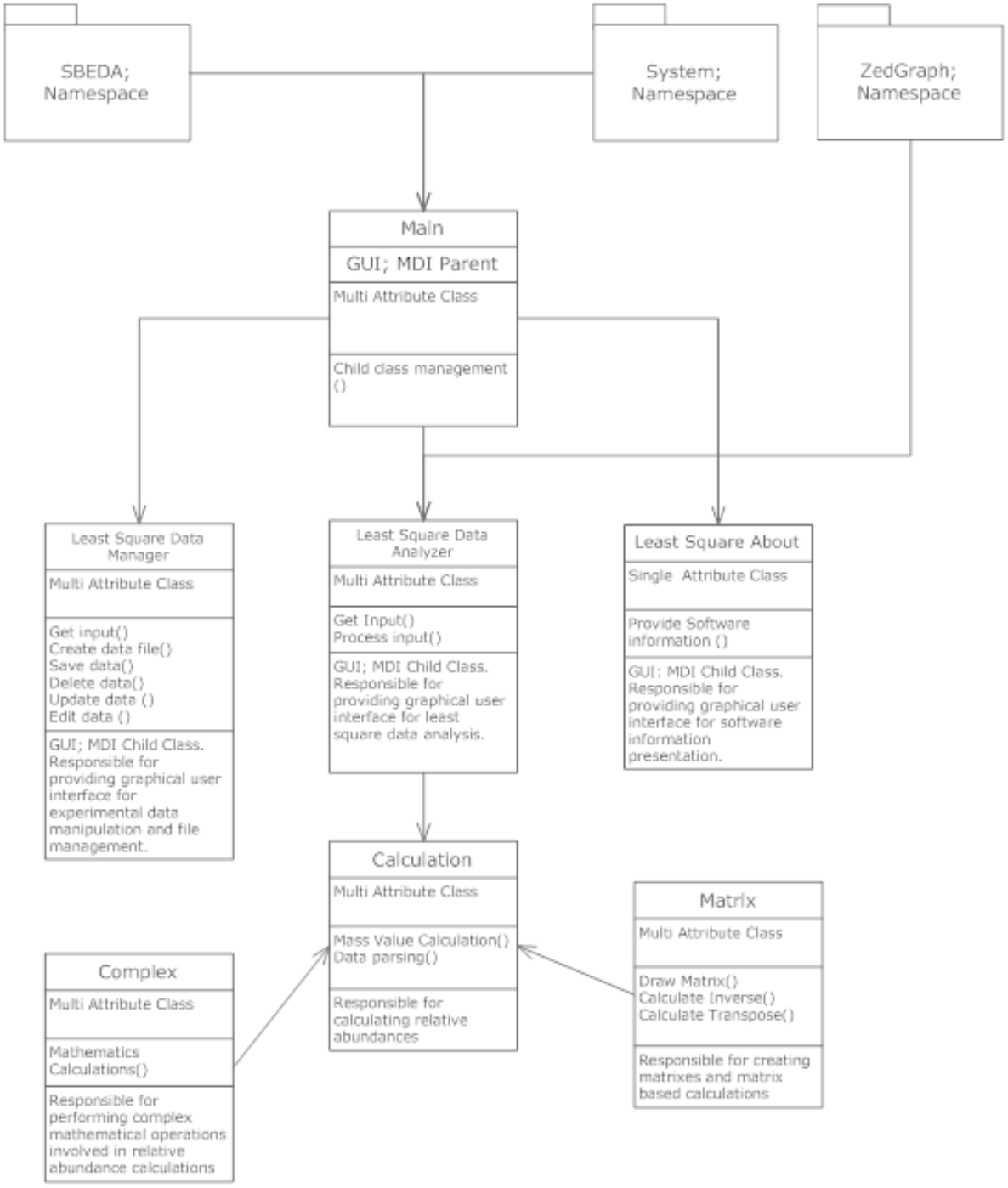


Figure 21: LS-MIDA; UML Class Diagram.

Figure Legend 21: The class diagram of LS-MIDA consisting of three name spaces (SBEDA, System and ZEDGraph), one main class (Main), six multi attribute classes (LeastSquareDataAnalyzer, LeastSquareDataManager, LeastSquareAbout, Calculation, Complex and Matrix) and one single attribute class (LeastSquareAbout).

Namespace REDGraph is a third party application programming interface used mainly for the development of graphical visualization of statistical, mathematical and experimental data in the form of two and three dimensional colored bar charts.

There are seven newly developed interlinked classes i.e. Main, LeastSquareDataAnalyzer, LeastSquareDataManager, LeastSquareAbout, Calculation, Complex and Matrix (Figure 21). As LS-MIDA is a multi document interface (MDI) application, Main MDI parent class which contains all other child classes. LeastSquareDataAnalyzer is the multi attribute class developed as the graphical user interface of the LS-MIDA analyzer which provides all visual options to the user to load, edit, analyze and visualize experimental data and observed results. LeastSquareDataManager is the multi attribute class developed as the graphical user interface of the LS-MIDA Data Manager which provides all visual options to the user for file based experimental data management and manipulation including entering, loading, editing, updating, deleting, merging, replacing and saving data in files.

Calculations is the multi attribute class developed for performing all mathematical operations including mass value estimations, relative abundances, data parsing and different data format conversions. Matrix is the multi attribute class developed for performing matrix operations including drawing simple matrix of NxM rows and columns, calculation inverse and transpose of matrix.

Complex is the multi attribute class developed for difficult mathematical operations including square root, absolute, tangent and operator overloading. LeastSquareAbout is the single attribute class, providing information LS-MIDA and development team and research group. Main sequence of classes, as shown in Figure 21, starts with Main container class, which provides other graphical user interface based classes LeastSquareDataAnalyzer, LeastSquareDataManager and LeastSquareDataAbout. LeastSquareDataAnalyzer perform user system communication, let user enter, edit and visualize experimental data, and analyze experimental data by directly using class Calculations which the uses classes i.e. Matrix and Complex. LeastSquareDataManager is an independent multi attribute class performing operations including user system communication for file based data management and manipulations.

3.3 LS-MIDA Implementation

The above mathematics was implemented using presented UML designs in the form of a software "LS-MIDA". The implementation of Brauman's least square method and the inclusion of binomial expression allow accurate isotopologue calculations according to experimental data and with correction for the different isotopes and isotopomer abundances. The available and tested version of LS-MIDA provides two main modules i.e. data analyzer and data manager. The data analyzer is capable of processing experimental data (metabolite information, m/e values and experimental relative intensity values. It then estimates mass values (M_o, M_{-1}, M_{max}), predicts relative natural abundance values, relative abundance values for each fragment and the percentage of relative intensity values per m/e values.

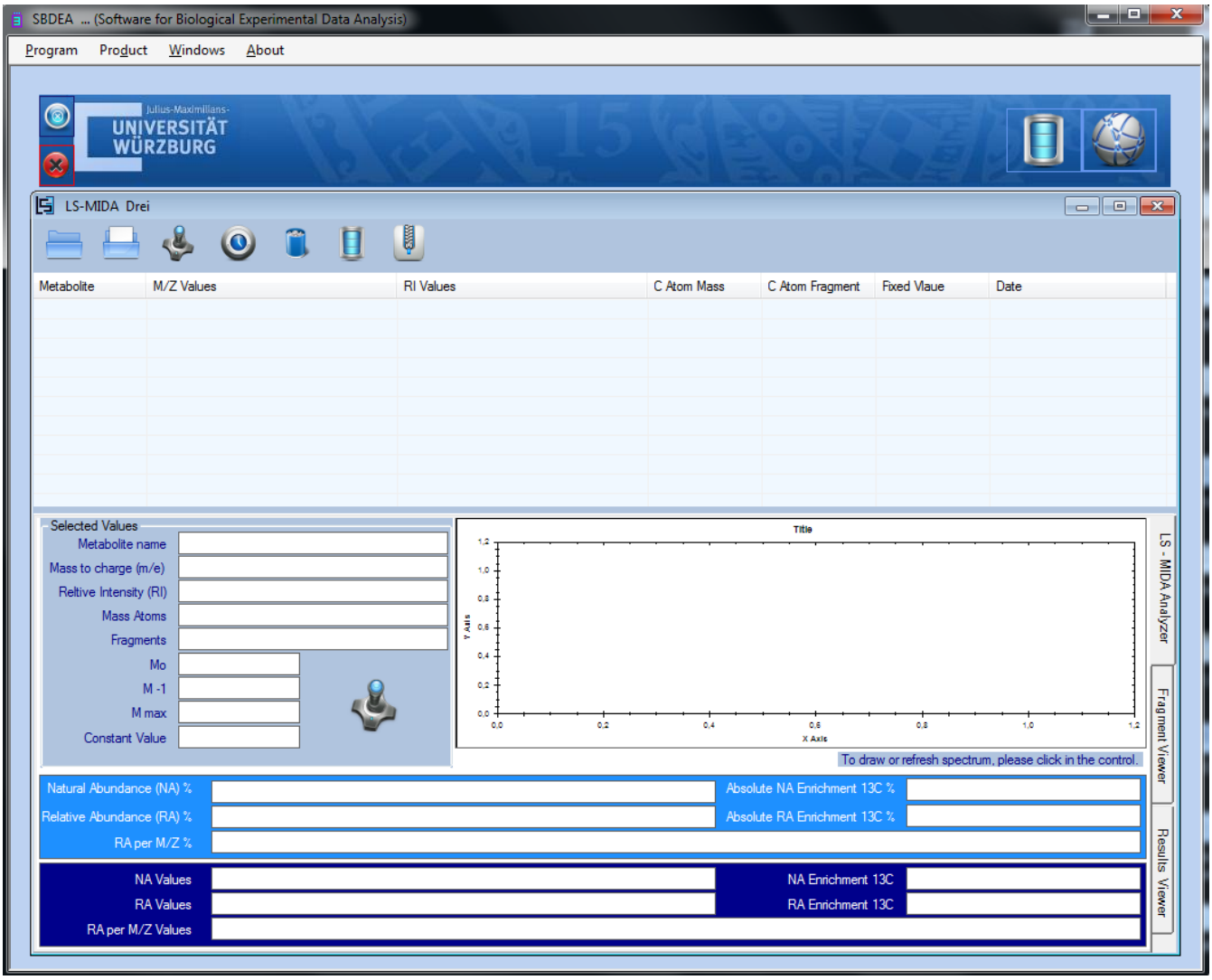


Figure 22 (a): LS-MIDA; Data Analyzer

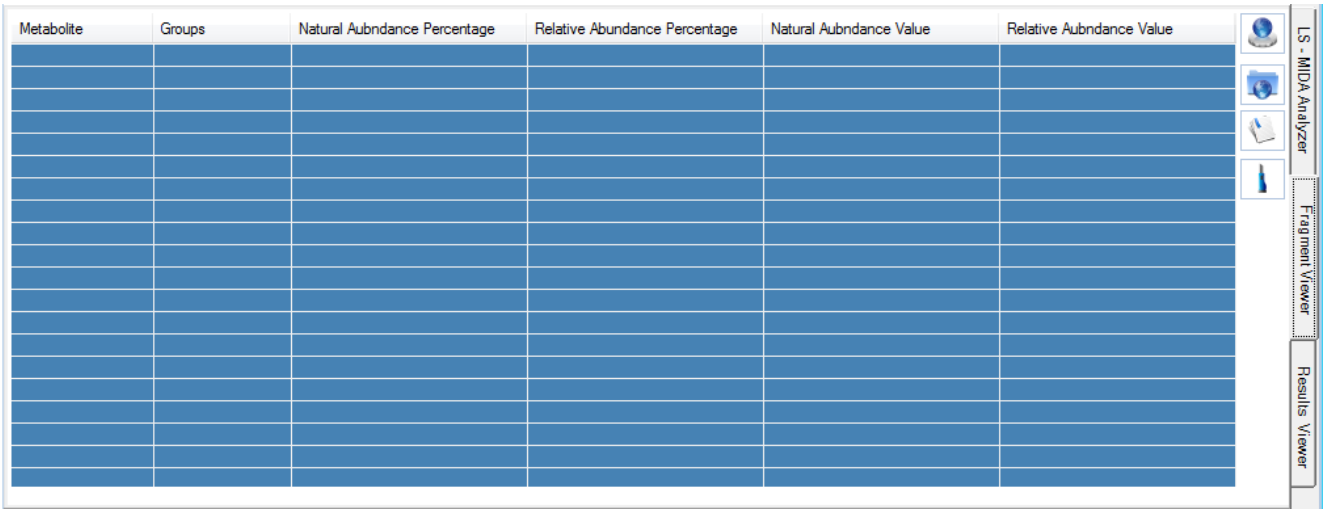


Figure 22 (b): Labeled Isotopomers Viewer

Figure 22(c): Processed Data Viewer

Figure 22: LS-MIDA; GUI of Data Analyzer.

Figure Legend 22: The LS-MIDA; GUI of Data Analyzer (a) presents the main graphical user interface responsible for handling user data input, analyzing and producing spectrum, along with the tab based graphical user interface module for isotopomer data (b) and complete application's output data viewing and manipulation.

It also draws the mass spectrum of the calculated relative intensity values and against constant m/e values. The graphical user interface of LS-MIDA Data Analyzer consists of 9 main controls: open data file, clear all text controls, measure selected data, measure all data, delete selected data, open data manager, exit, present input data and selected values. Moreover the graphical interface is divided into

three views: LS-MIDA Analyzer, Fragment Viewer and Result Viewer as shown in Figure 22 (a, b, c) and described in Table 5.

No.	Features	Descriptions
1	open data file	Opens directory browser to select input data file from attached repositories and loads data from data file into data viewer.
2	clear all text controls	Deletes all loaded data and clears all text controls.
3	measure selected data	Process selected data entry (only one at a time) from data view and perform MIDA.
4	measure all data	Processes all loaded data (all data entries) at once.
5	delete selected data	Deletes selected data entry from data view.
6	open data manager	Open LS-MIDA Data Manager's graphical user interface.
7	exit	Closes the LS-MIDA Data Analyzer.
8	present input data	Provides textual visualization of loaded data (option1). Allow user to select one or multiple values (option 3, 4 and 5) from loaded data.
9	Selected Values	Provides text boxes for experimental data manipulation by editing selected input data entry values from data viewer or by entering new experimental data for measurement analysis. It provides following text boxes. • Metabolite; name of the metabolite. • M/Z Values; mass to charge ratio values. • RI Values; relative intensity values.

		• C Atom Mass: atom number. • C Atom Fragment; number of fragments • M0; mass value • M-1; mass value minus 1 • Mmax; maximum mass values • Constant Value; set mass value
10	LS-MIDA Analyzer	Provides textual visualization of resultant information obtained after data processing, consisting of following elements: • Natural Abundance (NA) %; measured percentage of natural abundance with respect to the input number of fragments. • Absolute NA Enrichment ^{13}C %; measured percentage of absolute natural abundance ^{13}C. • Relative Abundance (RA) %; measured percentage of relative abundance with respect to the input number of fragments. • Absolute RA Enrichment ^{13}C %; measured percentage of absolute relative abundance ^{13}C. • RA per M/Z%, NA Values; measured percentage of relative intensity values per M/Z values • NA Values; measured actual natural abundance values. • NA Enrichment ^{13}C; measured actual NA Enrichment 13 C value. • RA Values; measured actual relative abundance values. • RA Enrichment ^{13}C; measured actual RA Enrichment 13 C

		value. • RA per M/Z Values; actual relative intensity values. Furthermore, LS-MDIA Analyzer viewer provides visual presentation of measured relative intensity values with respect to the mass to charge ratio values, in a spectrum.
11	Fragment Viewer	Provides the information about measured abundances (natural and relative) with respect to the number of fragments. It consists of four controls as well • Export File; allows user to export measure fragment based output in a new file. • Import File; allows user to import already estimated data. • Clear Text; allows user to clear the view by deleting all the data. • Delete Selected Data; allows user to delete particular (selected) data
12	Result Viewer	Provides the information of complete output including Metabolite name, input values (M/Z values, Fragment Number) and measured values (NA, RA and RA/MZ).It consists of four controls as well. • Export File; allows user to export measure fragment based output in a new file. • Import File; allows user to import already estimated data. • Clear Text; allows user to clear the view by deleting all the data. • Delete Selected Data; allows user to delete particular (selected) data

Table 5: LS-MIDA Data Analyzer; Control Descriptions.

The graphical user interface of LS-MIDA Data Manager consists of 14 main controls: open data file, clear all text controls, close isotopo data manager, add new values, update edited values, clear text fields, save data in file, select values to edit, delete values, create new data file, select source directory, save file, cancel creating file and data view as shown in Figure. 23 and described in Table 6.

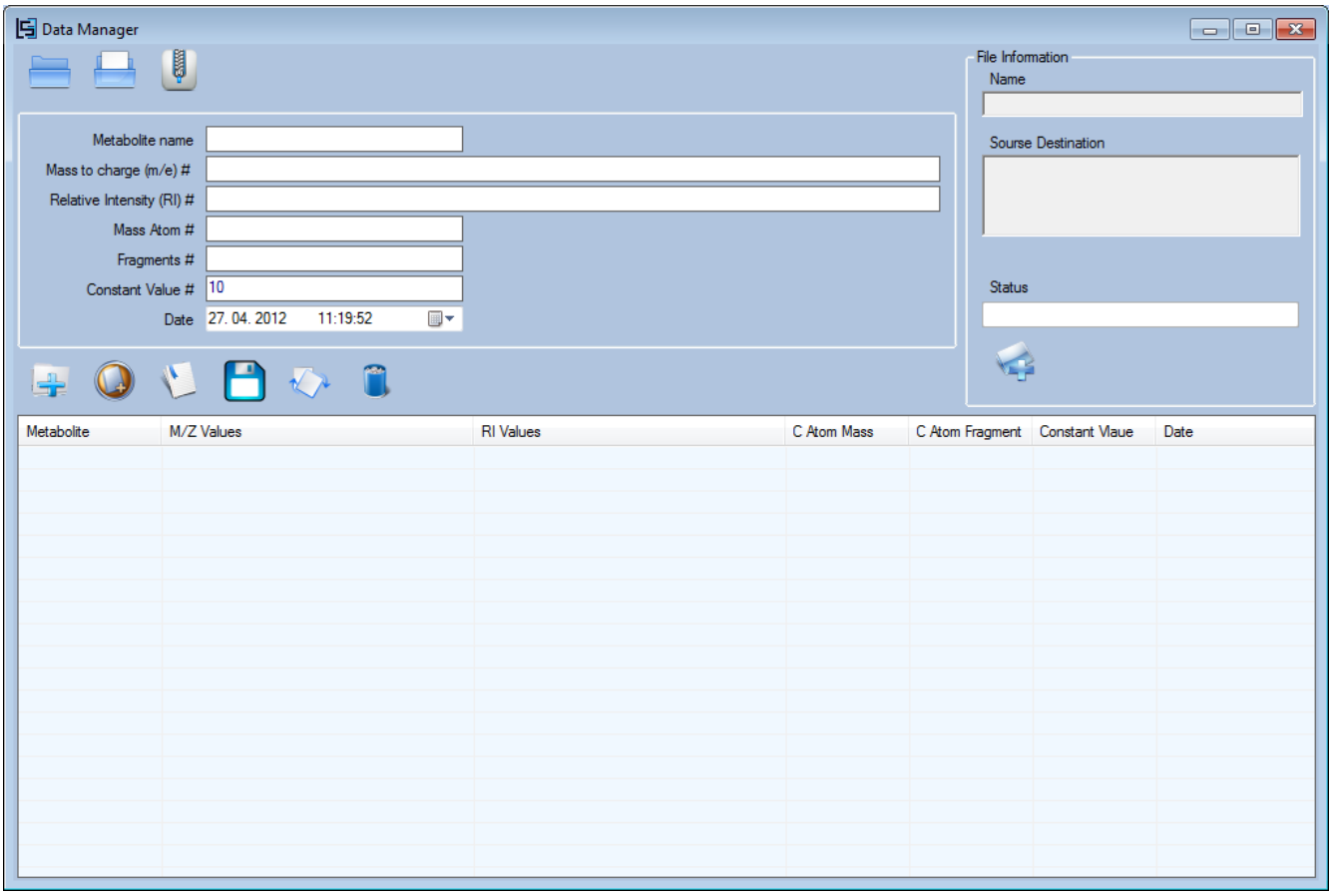


Figure 23: LS-MIDA; GUI of Data Manager.

Figure Legend 23: The LS-MIDA; GUI of Data Manager presents the main graphical user interface responsible for handling user data input and providing data management and manipulation options.

No.	Features	Descriptions
1	open data file	Opens directory browser to select input file from attached repositories and loads data from data file into data viewer.

2	clear all text controls	Deletes all loaded in data and clears all text controls.
3	close LS-MIDA data manager	Closes the LS-MIDA Data Manager.
4	add new values	Add newly entered values in text boxes to data view.
5	update edited values	Updates edited values in to data view
6	clear text fields	Deletes data a from text controls.
7	save data in file	Saves data into file.
8	select values to edit	Allows user to select one value from data view to edit existing values.
9	delete values	Deletes selected data entry from data view.
10	create new data file	Allows user to create new data (input) file.
11	select source directory	Opens directory browser to select the directory to store newly created data file. This option is only enabled and visible when user will click option 10.
12	save file	Allows user to save newly created file. It also allows user to save exiting file data e.g. if some data file is also open and user want to creates a new file, then system will ask if he want to merge existing data (in data view) to newly created file or not. This option is only enabled and visible when user will click option 10.
13	cancel creating file	Allows user to cancel new file creation process. This option is only enabled and visible when user will click option 10.
14	data view	Provides the textual view of all loaded, added or updated data.

Table 6: LS-MIDA Data Manager; Control Descriptions.

LS-MIDA is a multiple document interface (MDI) software application (public and freely distributed for academic use), developed following the principle of embedding children windows under a single parent

window by creating nested hierarchies. To meet aforementioned goals of LS-MIDA development, the graphical user interface of this application is divided into two main modules i.e. Data Analyzer and Data Manager. Data analyzer is responsible for providing options for experimental data load, analysis and visualization, whereas data manager is for biological experimental data manipulation and management.

The graphical user interface of data analyzer consists of eight main features i.e. language, control box, data view, selected values, analysis, calculated abundances, calculation log and visualization. These options help in loading experimental data, analyzing it and visualizing results. The graphical user interface of data manager consists of seven main options i.e. language, file control box, file information, create new data file, experimental data, selected data and data manipulation options. These options allows user to create new data files, manage created data files, merge new or already made data files into one or more new or already created files data files and manipulate entries data files. It is an independent file based data management system, doesn't require any external or third party database to install and use.

Some different mathematical algorithms: Binomial Expansion, Brauman's least square, Absolute ^{13}C Enrichment, Quantitative linear regression and Abundance matrix are used to measure mass isotopomers distribution from spectral data by analyzing each peak of given mass and each mass atom fragment. These algorithms are capable of analyzing (isotopic) labeled (Carbon) metabolite based experimental raw data (different amino acids) by measuring mass values (minimum and maximum), predicts natural abundance values for pure compounds, draws linear relationships between masses within the range of a compound and performs linear regression analysis to predict RI values with respect to the each m/z value.

Binomial Expansion uses actual (observed during GC-MS experimentation) RI values and number of mass fragments to measure natural abundance values. Brauman's least square method is used to compute Ri values per each m/z values. Computed natural abundance values and Ri values are used in quantitative linear regression analysis to predict relative abundance values and their percentages per each mass fragment. Absolute ^{13}C enrichment method is used to calculate absolute natural abundance enrichment and absolute relative abundance enrichments for concrete experimentation, we have used different amino acid's based experimental raw data consisting of m/z values, actual Ri values per each m/e value, number of C atoms in respective fragments, number of C atoms in compound and mass value.

3.4 LS-MIDA Modelling Experimentation

To apply the LS-MIDA, some experimental data has to be collected first. The whole experimentation process using LS-MIDA consists of three steps i.e. preparation of data set, input data file preparation and management, and data analysis. Observed data during actual experimentation (GC-MS) are collected during the preparation of data set. During input data file preparation and management at first data manager is used to structure data by organizing in an experimental data file which is later used by data analyzer for analysis. Throughout the experimental data analysis, each amino acid is individually analyzed.

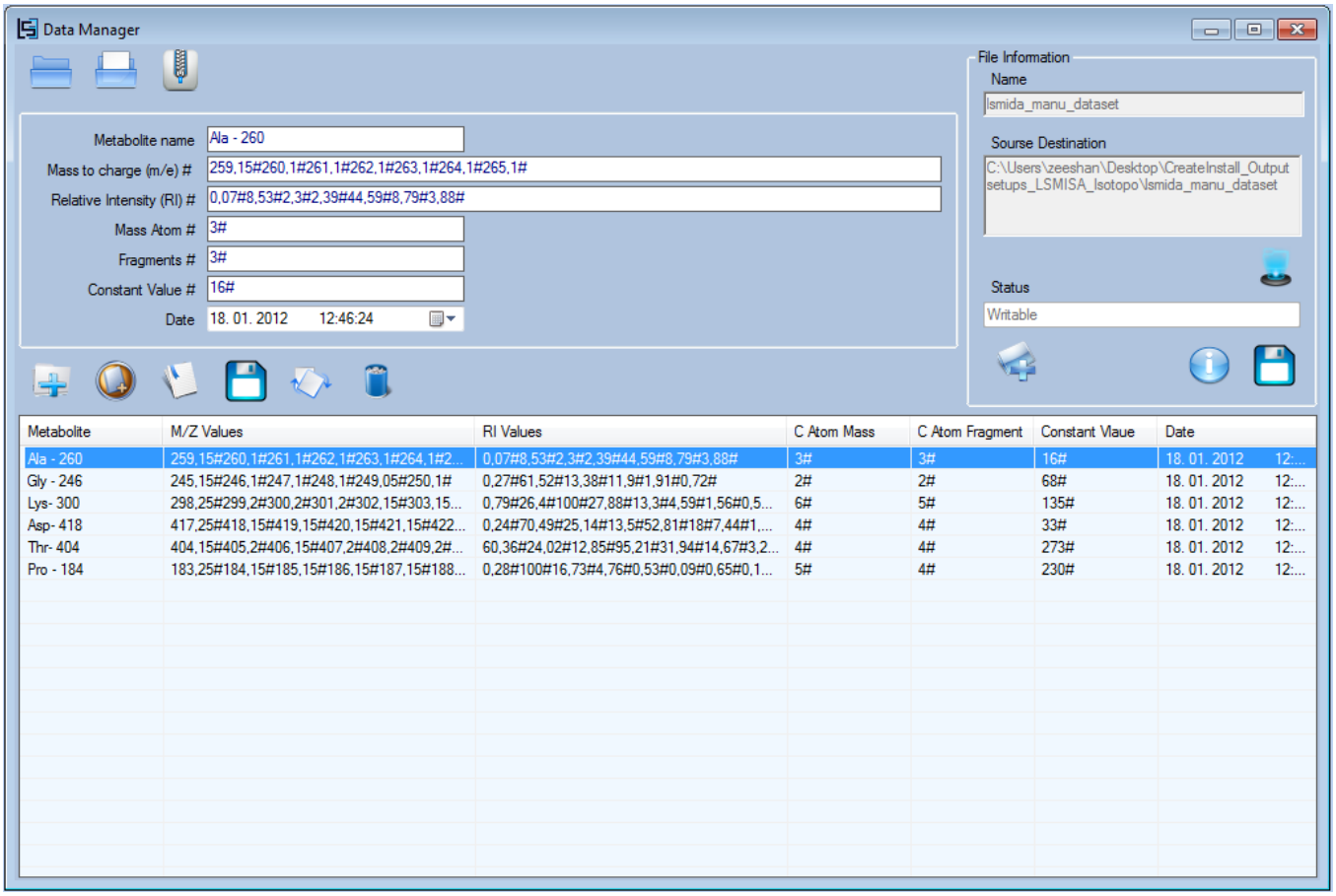


Figure 24: Data Manger; Inputted data management.

Figure Legend 24: Data Manger; Inputted data management presents the inputted user data consisting of six metabolites: Alanine (Ala – 260), Glycine (Gly – 246), Lysine (Lys- 300), Aspartic Acid (Asp- 418), Threonine- 404, Proline (Pro-184). A new data file "lsmida_manu_dataset.ls" is created for file based data management and further data analysis using LS-MIDA Data Analyzer.

Observed results after different experiments of metabolic isotopomers analysis with different metabolites using GC-MS are collected for mass isotopomers predictions using LS-MIDA. The collected data consists of six different amino acids i.e. Alanine (Ala – 260), Glycine (Gly – 246), Lysine (Lys- 300), Aspartic Acid (Asp- 418), Threonine (Thr- 404), Proline (Pro-184), described in Table 7; consists of following experimental elements i.e. m/z values, Ri values, atomic mass values and atomic fragment numbers. A data input file is created using data manager, as shown in Figure 24.

No.	Metabolite	m/e values	Ri values	C Atom Mass	Fragment
1	Ala- 260	259,15#260,1#261,1#262,1#263,1#264,1#265,1#	0,07#8,53#2,3#2,39#44,59#8,79#3,88#	3#	3#
2	Gly - 246	245,15#246,1#247,1#248,1#249,05#250,1#	0,27#61,52#13,38#11,9#1,91#0,72#	2#	2#
3	Lys- 300	298,25#299,2#300,2#301,2#302,15#303,15#304,15#305,15#306,15#	0,79#26,4#100#27,88#13,3#4,59#1,56#0,53#0,16#	6#	5#
4	Asp- 418	417,25#418,15#419,15#420,15#421,15#422,15#423,1#424,1#	0,24#70,49#25,14#13,5#52,81#18#7,44#1,66#	4#	4#
5	Thr- 404	404,15#405,2#406,15#407,2#408,2#409,2#410,2#	60,36#24,02#12,85#95,21#31,94#14,67#3,21#	4#	4#
6	Pro - 184	183,25#184,15#185,15#186,15#187,15#188,15#189,1#190,1#191,1#192,1#	0,28#100#16,73#4,76#0,53#0,09#0,65#0,14#0,07#0,01#	5#	4#

Table 7: LS-MIDA: Experimental Data Set.

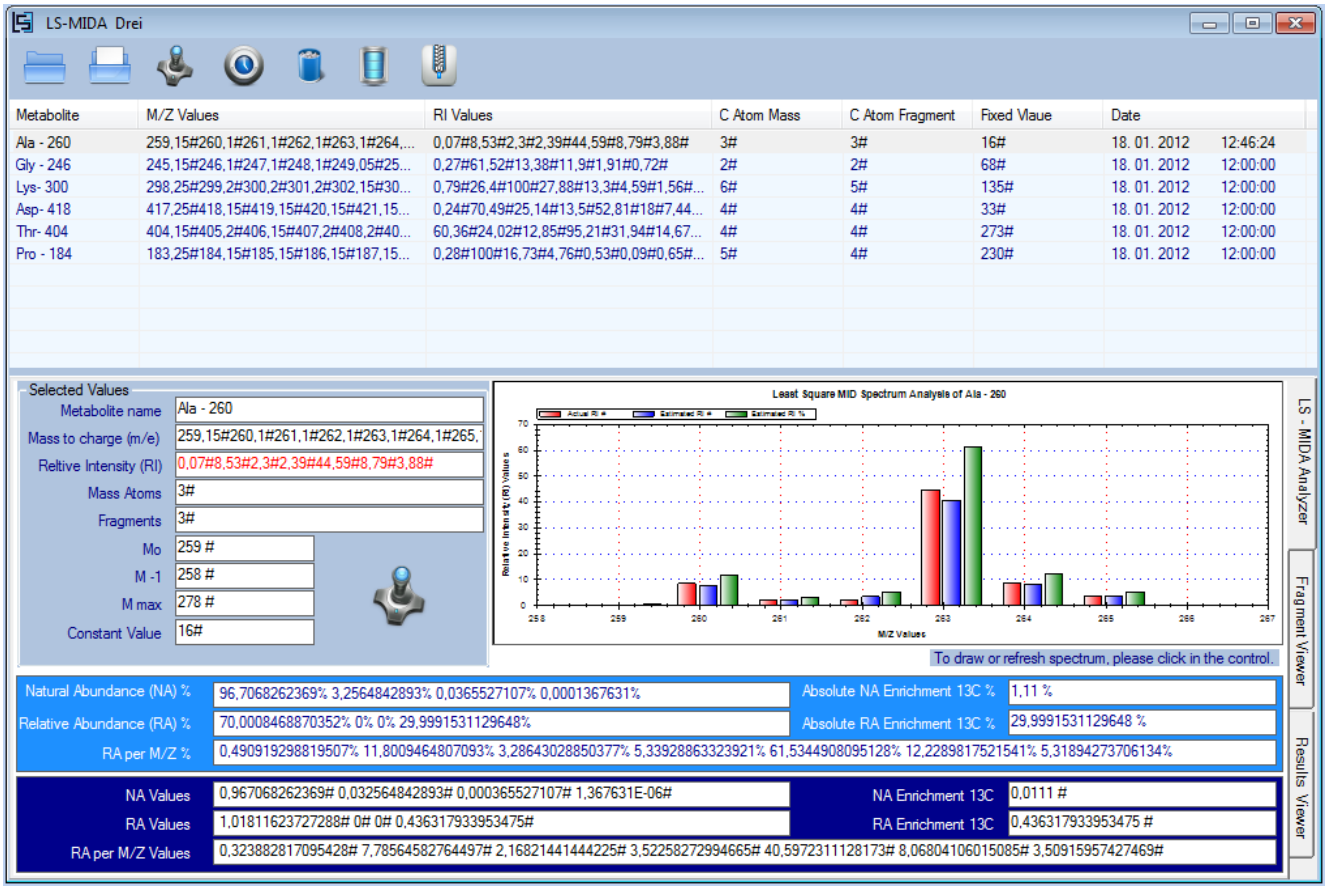


Figure 25 (a): LS-MIDA- Data Analyzer; Analyzing Alanine

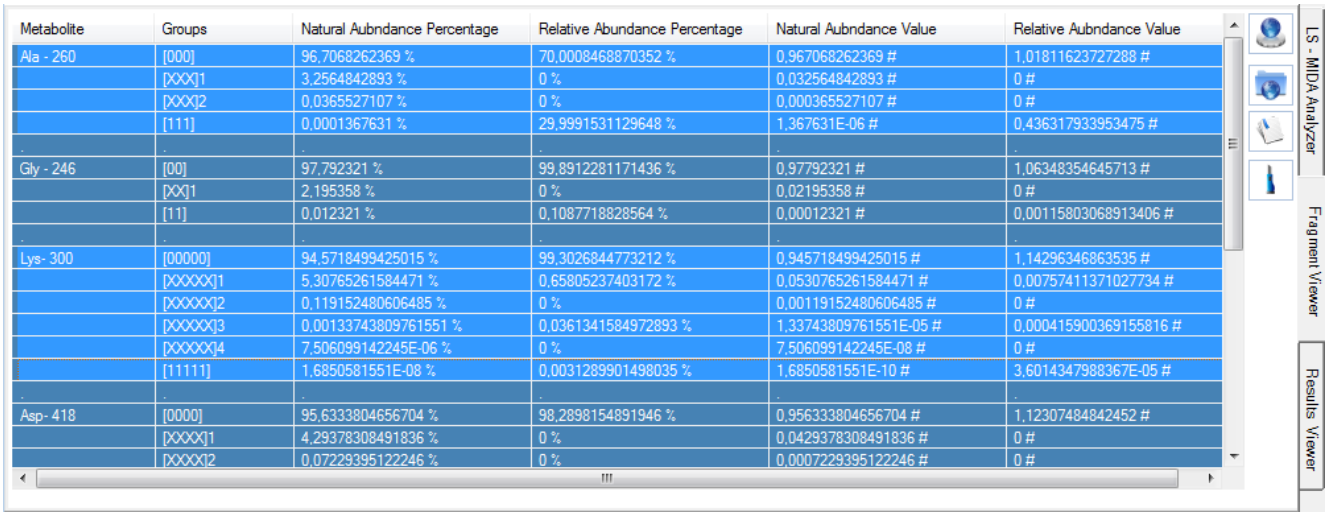


Figure 25 (b): Fragment Viewer

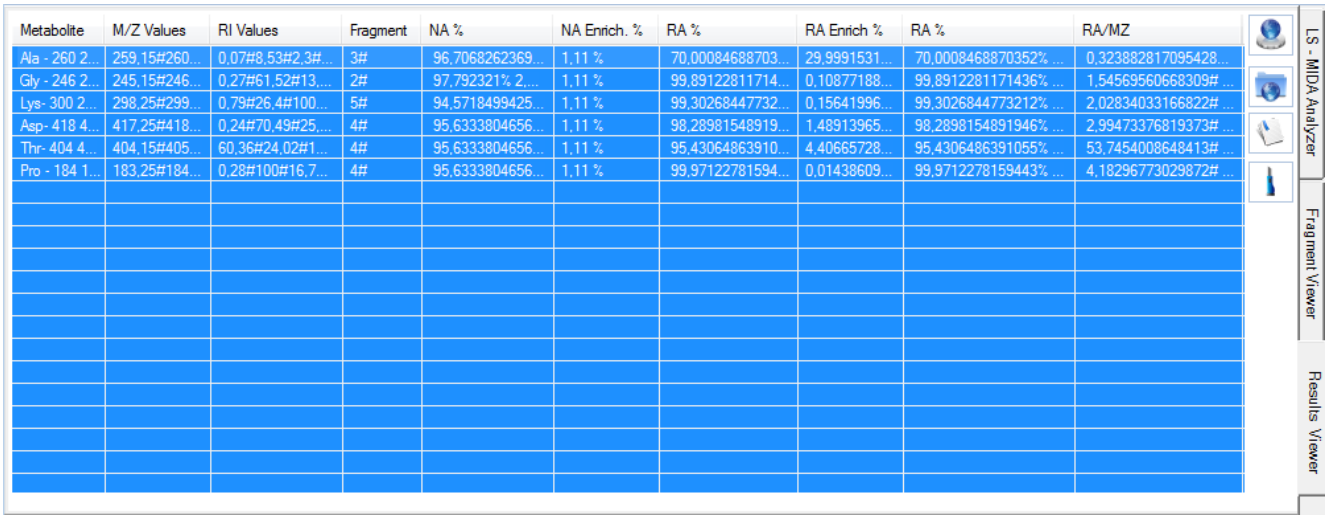


Figure 25 (c): Result Viewer

Figure 25: LS-MIDA- Data Analysis.

Figure Legend 25: LS-MIDA- Data Analyzer; Analyzing Alanine (a) presents the obtained results after metabolite Ala-260 analysis. Labeled fragments based output (relative and natural abundance) is presented in fragment viewer (b) and complete output information is presented in result viewer (c).

The sequence of input file creation started with the creation of a new file and entering manually each metabolite based experimental data. Before starting analysis using data analyzer, newly created input data file is validated by data analyzer and as shown in Figure 25 (a, b, c). During experimental data analysis, the prepared data file is input and analyzed using data analyzer item by item. Each input item of experimental data consists of five main information elements i.e. metabolite name, m/z values, RI values, atomic mass values and atomic fragment numbers. Evaluation process using data analyzer starts with the use of data entries of earlier mentioned complete experimental data set. The data entry of each metabolite (Ala, Gly, Lys, Asp, Thr, Pro) is selected and processed for analysis (one by one or at once all) as shown in Figure 26.

Metabolite / Abundances	Ala-260	Gly - 246	Lys- 300	Asp- 418	Thr- 404	Pro - 184

Natural Abundances	96,7068262369% 3,25648428 93% 0,03655271 07% 0,00013676 31%	97,792321% 2,195358% 0,012321%	94,5718499425015% 5,30765261584471% 0,11915248060 6485% 0,00133743809761551% 7,506099142245E-06% 1,6850581551E-08%	95,6333804656704% 4,29378308491836% 0,07229395122246% 0,00054098011836% 1,51807041E-06%	95,6333804656704% 4,29378308491836% 0,07229395122246% 0,00054098011836% 1,51807041E-06%	95,6333804656704% 4,29378308491836% 0,07229395122246% 0,00054098011836% 1,51807041E-06%
Relative Abundances	70,0008468870352% 0% 0% 29,9991531129648%	99,8912281171436% 0% 0,1087718828564%	99,3026844773212% 0,65805237403172% 0% 0,036134158 4972893% 0% 0,00312899014 98035%	98,2898154891946% 0% 0% 0,88417944 1492093% 0,82600506 9313319%	95,4306486391055% 0% 0% 0,65077632 1011312% 3,91857503988317%	99,9712278159443% 0% 0,028772184055688 8% 0% 0%
Percentage of Relative Abundances per M/Z Values	0,490919298819507% 11,8009464807093% 3,28643028850377% 5,3392886	1,80213609628263% 67,4093295593138% 14,8808366101696% 13,0223740883095%	1,29493119012556% 17,3056208343025% 54,8537432146748% 15,4613301062492%	1,73052015748485% 36,2024002594465% 13,0326729659361% 8,0283604597961%	24,5258042348218% 9,84608905503391% 6,84461974156777% 38,5776284857007%	3,71315835064298% 78,22634037855 9% 13,1554719031474% 3,71421169606794%

	3323921% 61,534490 8095128% 12,228981 7521541% 5,3189427 3706134%	2,100250050 63644% 0,785073595 288046%	7,322134001 30023% 2,526226758 4692% 0,858566818 538945% 0,291047522 20747% 0,086399554 1321148%	27,10220258 51006% 9,266910238 56856% 3,797946677 63311% 0,838986656 034203%	13,0119545 375927% 5,91270814 602815% 1,28119579 925491%	0,4150132 97114186 % 0,0926173 04892388 1% 0,5095836 9233405% 0,1111425 05509892 % 0,0546968 36831434 7% 0,0077640 34900685 52%

Table 8: Calculated Abundances using LS-MIDA.

The resultant information from data analyzer consists of three different items i.e. natural abundances, relative abundances and percentage of calculated relative abundances per m/z values, in each experiment, described in Table 8. Furthermore based on observed results a spectrum is drawn.

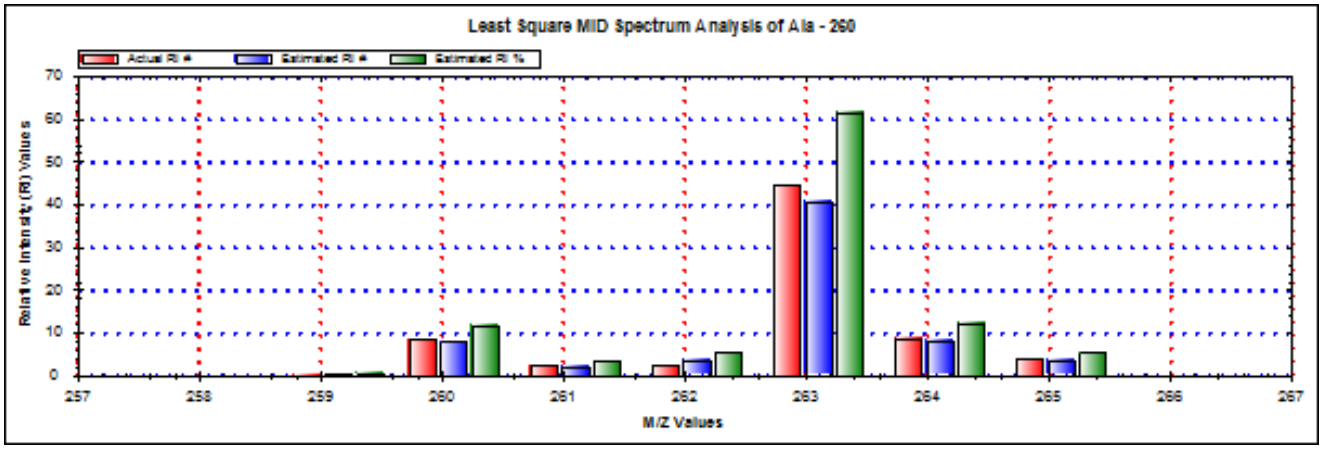


Figure 26 (a): LS-MIDA; Ala 260 Drawn Spectrum

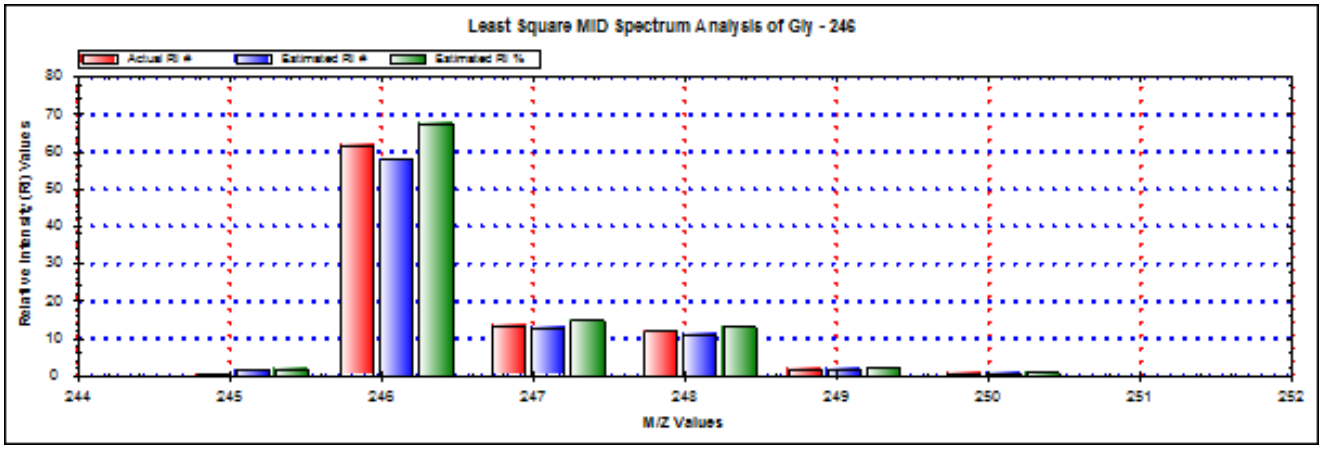


Figure 26 (b): LS-MIDA; Gly 246 Drawn Spectrum

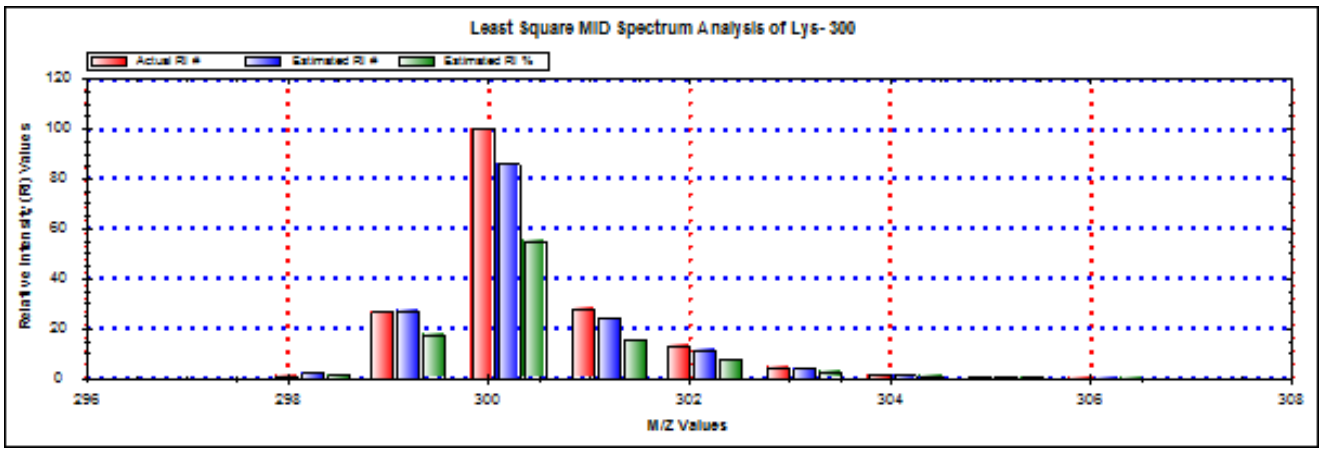


Figure 26 (c): LS-MIDA; Lys 300 Drawn Spectrum

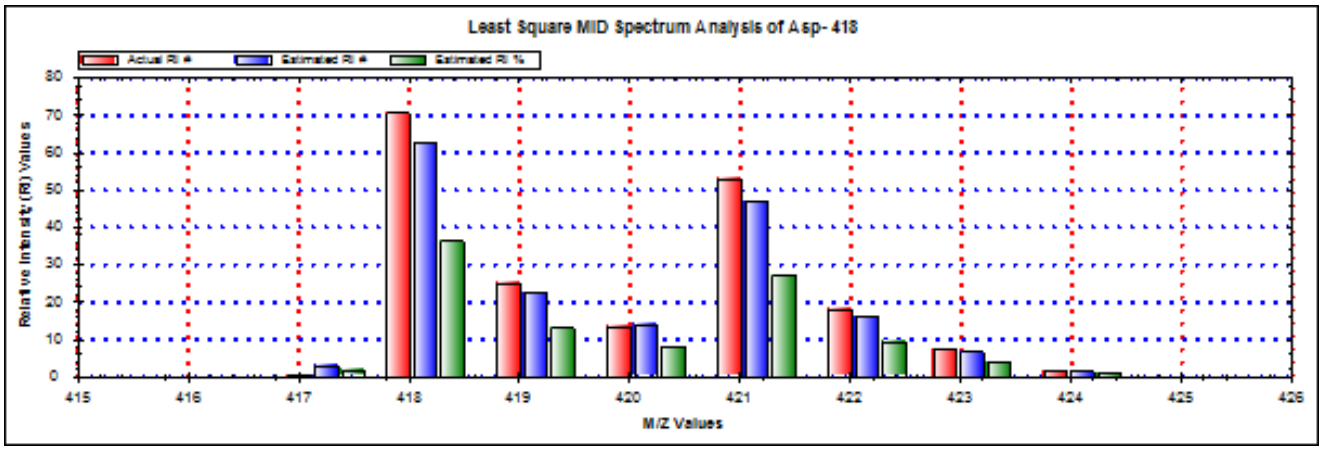


Figure 26 (d): LS-MIDA; Asp 418 Drawn Spectrum

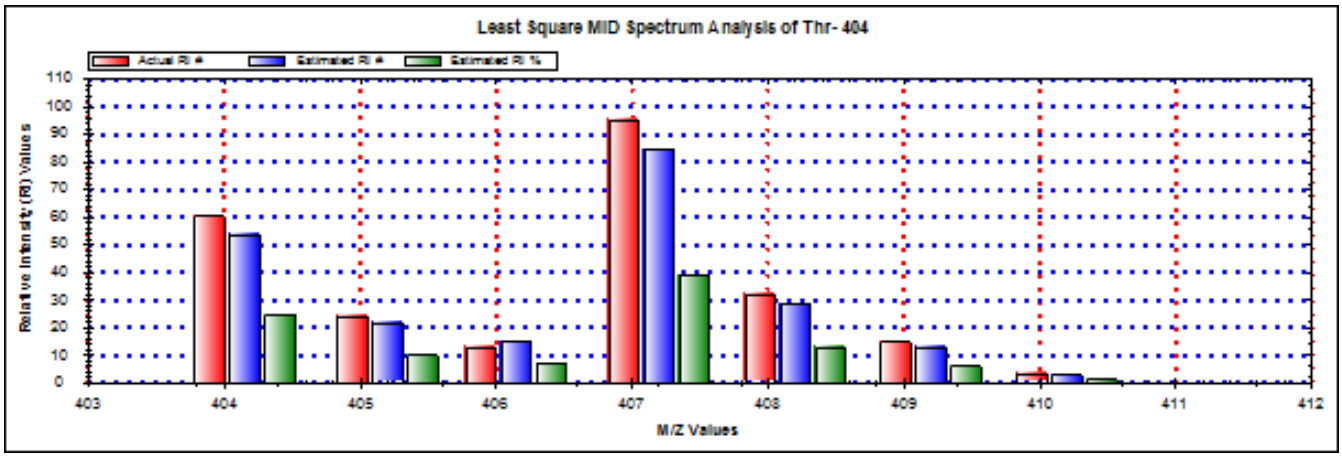


Figure 26 (e): LS-MIDA; The 404 Drawn Spectrum

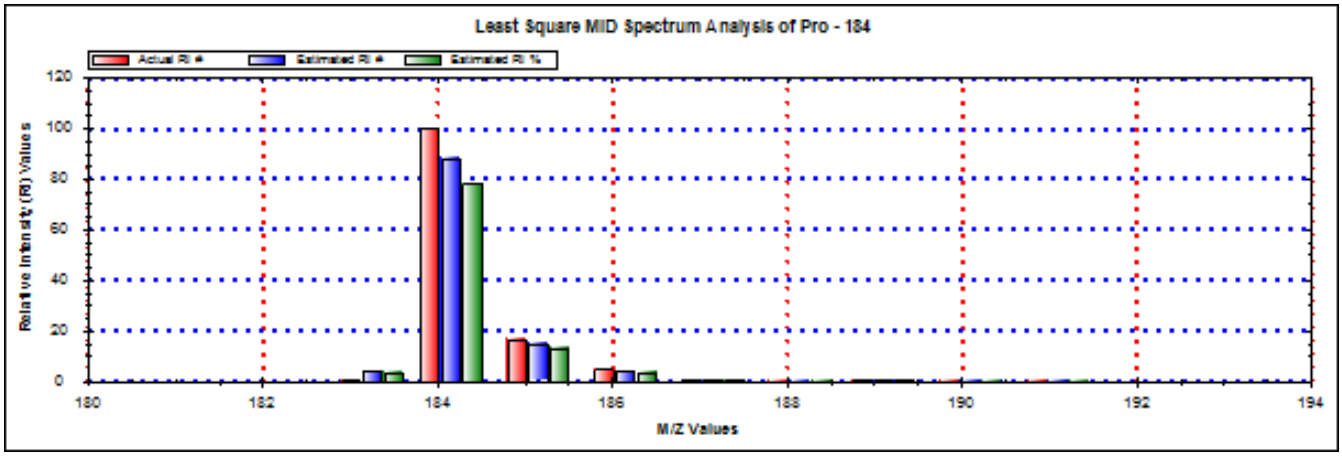


Figure 26 (f): LS-MIDA; Pro 184 Drawn Spectrum

Figure 26: LS-MIDA; Drawn Spectrums.

Figure Legend 26: This Figure presents the drawn spectrums, as the results of analyzed metabolites Ala–260 (a), Gly – 246 (b), Lys- 300 (c), Asp- 418 (d), Thr- 404 (e) and Pro-184 (f). The parameters of drawn spectrum are relative intensity values and mass to charge ratio values. Red colored bar presents actual relative intensity values observed during experimentation, blue colored bar presents estimated relative intensity values and green colored relative intensity values presents percentage of observed relative intensity values.

3.5 LS-MIDA Evaluation

The methodology includes statistical calculations and we give the sequence of steps for the implementation process. The accuracy of the mathematical equation can be determined by iteratively analyzing the estimated abundance resonances (Baverel et al., 2003). We tested the preparation of

different data sets; input data file preparation and management, experimentation and data analysis. Example data and results are presented from Salmonella labeling experiments for different metabolites such as analysis of various amino acids.

During input data file preparation and management the LS-MIDA Data Manager structures data into experimental data files which are later used by LS-MIDA Data Analyzer for analysis. Throughout the experimental data analysis, each observed resultant data set during experimentation is individually analyzed using LS-MIDA. Based on the obtained results, a clear difference is observed in actual and calculated abundances, as each calculated relative abundance value is less than the actual abundance value at each m/z value, in each metabolite. Furthermore, there are two software solutions available for isotopologue data processing i.e. Envelop (Michael et al., 2008) and Isotope Pattern Calculator (Massila et al., 2008), but none uses binomial expression for data extension.

The implementation of Brauman's least square method with the inclusion of binomial expression, however, allows our isotopologue data to be more accurate. LS-MIDA is the only tool providing a file based data management system for experimental metabolic mass isotopomers based data. There is fast processing speed (only seconds), calculation complexity scales O(n2) with the number of carbon atoms per isotopologue. Furthermore, LS-MIDA is the only tool providing a file based data management system for experimental metabolic mass isotopomers based data. There is fast processing speed (only seconds), calculation complexity scales O(n2) with the number of carbon atoms per isotopologue.

All data examples given are from Salmonella isotopologue measurements. Accumulating such measurements and taking further data on their metabolism into account allows insights into Salmonella and their metabolism during infection (Eisenreich et al., 2010). Thus glucose, glucose-6P and gluconate present possible carbon sources for the intracellular pathogen. Under such sources enzymes and fluxes for glycolysis and for the Entner–Doudoroff pathway are up-regulated in these bacteria. In contrast, most enzymes and the fluxes in the TCA cycle are down-regulated.

3.6 Installation

Take the setup executable SBEDA (Software for Biological Experimental Data Analysis) framework.

To run LS-MIDA Data Analyzer click on the “Blue Globe” icon in the main bar and to run Data Manager click on “Cylinder Database” icon.

For a simple example and guided tour, load the example data into Data Analyzer by clicking the “Open Data File” icon, and process these by clicking “Measure Selected Data” icon. To obtain a spectrum click on the white control.

The SBEDA framework (including LS-MIDA) is developed using the Microsoft C# (sharp) programming language and Microsoft Dot Net framework 2008. It is compatible (install and use) for all Microsoft Windows operating systems.

4 Isotopo

The major objective of metabolic isotope mixture studies is to quantify the fraction of synthesized molecules in the mixture, as labeled substrates are more often analyzed because of their more volatile derivatives. Isotope is an atom with the same number of protons of particular element but with different numbers of neutrons. Isotopic labeling is one referenced technique for the quantification of any compound, as it uses a known amount of stable isotopically substituted analog to a sample. During this technique the primary ion for each isotopic species is needed to be monitored and correct ion intensities needed to be measured due to ion overlapping with other isotopic species.

Here, after shedding of LS-MIDA (section 3), we developed a new software for MIDA, capable of easily processing experimental data showing and managing better and using an alternative algorithm applying pseudo inverse and partial matrix calculation as well as iterative refinement for best processing isotopologue data, here this is called Isotopo. Input includes: metabolite information (ion), m/z values, actual relative intensities (up to three entries for one m/z value), standard relative intensities and the number of carbon atom fragments. During GC-MS experiments, the m/z values are dimensionless quantities formed by dividing the mass number of the ion by its charge number, the actual relative intensities are the different intensity values for individual ions measured. “Natural abundance” denotes a theoretical value calculated by Isotopo which is the complete population of isotopomers in the molecules of a given compound (including label derived isotopomers but without artificially added isotopomers), “relative abundance” is a vector calculated that refers to the population of artificially labeled isotopomers (e.g. by ^{13}C) in the molecules of a compound. Relative intensities per m/z value are calculated next. Here the “fractional molar abundance” means the concentration of a molecular species as a fraction of the total number of molecules15. The software estimates mass values (M_0, M_{-1}, M_{max}): These are three values estimated from m/z values, M_0 is the first fraction less m/z, M_{-1} is the first m/z value minus 1 and M_{max} is estimated maximum m/z value. Furthermore Isotopo calculates the minimal value (see eq. 9), subtracting relative intensity values from the fractional molar abundances. Finally, the absolute enrichment of natural abundances, absolute enrichment of relative abundances are calculated. Using the implemented application up to three observed relative intensity value against one m/z values can be efficiently processed. Corresponding mean and standard deviations are calculated, the results are drawn as a spectrum of calculated relative intensity values. To standardize and maintain the experimental metabolite data (inputted and observed resultant data during experimentation), a file-based (independent of third

party tools) data manipulation and management system is also implemented. This new data management system allows the user to create new experimental data-based files to manager existing data files into existing files and to manipulate file data.

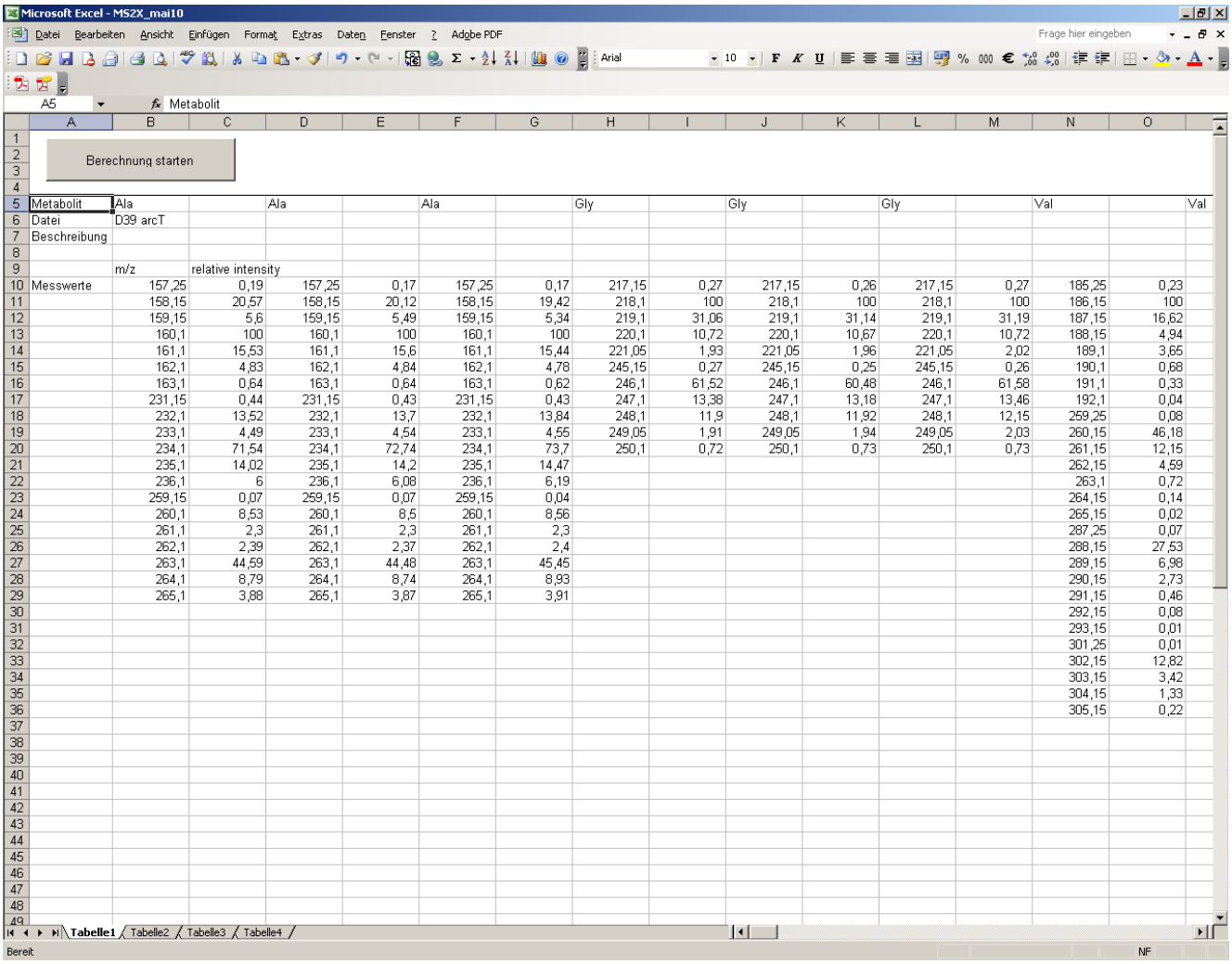
The motivation for our research and new software development is to obtain a comprehensive easy-to-use software application, including different calculation options, free for any user, with a data management system and all options as a key step to study complex metabolic fluxes such as the intertwined metabolism of host and microbial pathogen. Focusing on identifying the quantity of population of labeled isotopomers for resolving the exact rate of synthesized fractions present in the mixture and metabolic experimental data management, a new software application named "Isotopo" is proposed and designed. Isotopo is an application with proposed abilities of performing quantitative mass spectrometry to readily mixtures of materials labeled with stable isotopes. This can be very important for both biomedicine and biochemistry. Most recent version of Isotopo has the ability of processing experimental isotopomers data and estimating mass values and relative intensities. Using formal mathematical algorithms which generate an appropriate set of linear simultaneous equations, it predicts natural abundance values, relative isotopic abundance values and fractional molar abundance values for each fragment from labeled substance based experimental data elements. Using Isotopo it is also possible to process data sets with multiple data entries up to three actual intensity values against one mass to charge ratio values, estimate absolute enrichment, mean and standard deviation of both natural and relative isotopic abundances. Isotopo also provide the standardization of experimental data with a file based record keeping system for experimental data manipulation and management.

In this section of thesis, justifying the need of a new software application along with the presentation of an existing similar solution, we present the followed V-Model, formal UML designs (including use case, data flow, flow chart, system sequence, component and class diagrams), and designed graphical user interface of Isotopo. Going into the details,

we present implemented methodology, designed relational database, currently available most recent version of prototype application and validates it potential strength using some experimental isotopomers data.

4.1 Existing Similar Solutions

There is an existing solution provided by Prof Wolfgang Eisenreich at Lehrstuhl für Biochemie, Technische Universität München Germany, implementing the similar methodology but with completely different way of development and usage.



Microsoft Excel - MS2X_mai10

Datei Bearbeiten Ansicht Einfügen Format Extras Daten Fenster ? Adobe PDF

A5 Metabolit

Berechnung starten

Metabolit	Ala		Ala		Ala		Gly		Gly		Gly		Val		Val
Datei	D39 arcT														
Beschreibung															
	m/z	relative intensity													
Messwerte	157,25	0,19	157,25	0,17	157,25	0,17	217,15	0,27	217,15	0,26	217,15	0,27	185,25	0,23	
	158,15	20,57	158,15	20,12	158,15	19,42	218,1	100	218,1	100	218,1	100	186,15	100	
	159,15	5,6	159,15	5,49	159,15	5,34	219,1	31,06	219,1	31,14	219,1	31,19	187,15	16,62	
	160,1	100	160,1	100	160,1	100	220,1	10,72	220,1	10,67	220,1	10,72	188,15	4,94	
	161,1	15,53	161,1	15,6	161,1	15,44	221,05	1,93	221,05	1,96	221,05	2,02	189,1	3,65	
	162,1	4,83	162,1	4,84	162,1	4,78	245,15	0,27	245,15	0,25	245,15	0,26	190,1	0,68	
	163,1	0,64	163,1	0,64	163,1	0,62	246,1	61,52	246,1	60,48	246,1	61,58	191,1	0,33	
	231,15	0,44	231,15	0,43	231,15	0,43	247,1	13,38	247,1	13,18	247,1	13,46	192,1	0,04	
	232,1	13,52	232,1	13,7	232,1	13,84	248,1	11,9	248,1	11,92	248,1	12,15	259,25	0,08	
	233,1	4,49	233,1	4,54	233,1	4,55	249,05	1,91	249,05	1,94	249,05	2,03	260,15	46,18	
	234,1	71,54	234,1	72,74	234,1	73,7	250,1	0,72	250,1	0,73	250,1	0,73	261,15	12,15	
	235,1	14,02	235,1	14,2	235,1	14,47							262,15	4,59	
	236,1	6	236,1	6,08	236,1	6,19							263,1	0,72	
	259,15	0,07	259,15	0,07	259,15	0,04							264,15	0,14	
	260,1	8,53	260,1	8,5	260,1	8,56							265,15	0,02	
	261,1	2,3	261,1	2,3	261,1	2,3							287,25	0,07	
	262,1	2,39	262,1	2,37	262,1	2,4							288,15	27,53	
	263,1	44,59	263,1	44,48	263,1	45,45							289,15	6,98	
	264,1	8,79	264,1	8,74	264,1	8,93							290,15	2,73	
	265,1	3,88	265,1	3,87	265,1	3,91							291,15	0,46	
													292,15	0,08	
													293,15	0,01	
													301,25	0,01	
													302,15	12,82	
													303,15	3,42	
													304,15	1,33	
													305,15	0,22	

Tabelle1 Tabelle2 Tabelle3 Tabelle4

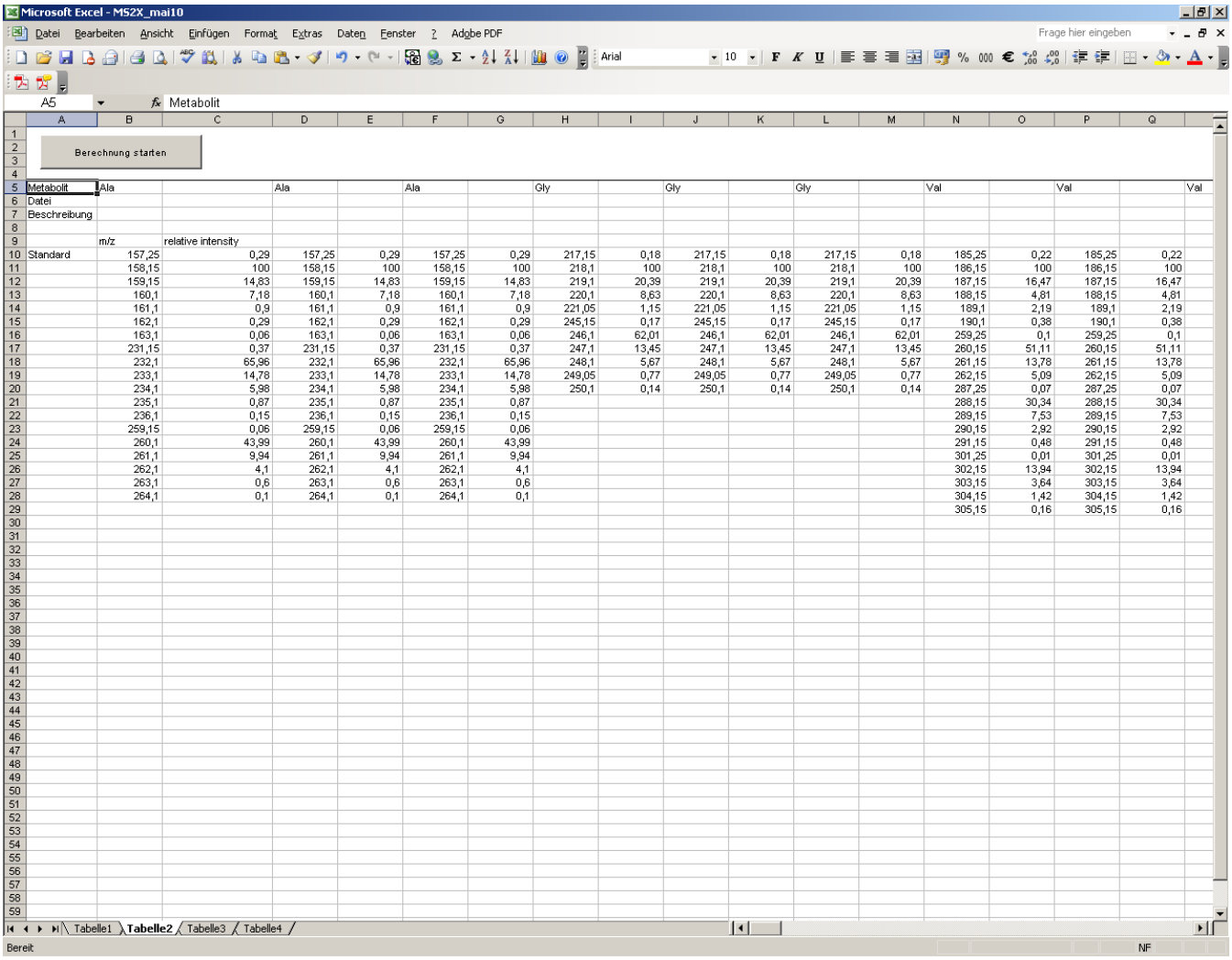
Bereit NF

Figure 27: Excel Sheet 1; Metabolite Experimental Data.

Figure Legend 27: Excel Sheet 1 presents experimental data consisting of several metabolites (e.g. Ala, Gly, Val etc.) with information about mass to charge ratio and relative intensity values observed during experimentation.

The previously implemented and available solution is basically a Microsoft excel macro. To process experimental metabolite data using already developed macro, user at first needs to store experimental data in Microsoft excel sheets in a particular user unfriendly three tables (pages or sheets) order i.e.

In the first excel page actual metabolite based information is stored including metabolite name, date, actual m/z values, and actual relative intensity values of each metabolite, as shown in Figure 27.



Microsoft Excel - MS2X_mai10

Datei Bearbeiten Ansicht Einfügen Format Extras Daten Fenster ? Adobe PDF

Frage hier eingeben

A5: Metabolit

Berechnung starten

Metabolit	Ala		Ala		Ala		Gly		Gly		Gly		Val		Val		Val
Datei																	
Beschreibung																	
	m/z	relative intensity															
Standard	157,25	0,29	157,25	0,29	157,25	0,29	217,15	0,18	217,15	0,18	217,15	0,18	185,25	0,22	185,25	0,22	
	158,15	100	158,15	100	158,15	100	218,1	100	218,1	100	218,1	100	186,15	100	186,15	100	
	159,15	14,83	159,15	14,83	159,15	14,83	219,1	20,39	219,1	20,39	219,1	20,39	187,15	16,47	187,15	16,47	
	160,1	7,18	160,1	7,18	160,1	7,18	220,1	8,63	220,1	8,63	220,1	8,63	188,15	4,81	188,15	4,81	
	161,1	0,9	161,1	0,9	161,1	0,9	221,05	1,15	221,05	1,15	221,05	1,15	189,1	2,19	189,1	2,19	
	162,1	0,29	162,1	0,29	162,1	0,29	245,15	0,17	245,15	0,17	245,15	0,17	190,1	0,38	190,1	0,38	
	163,1	0,06	163,1	0,06	163,1	0,06	246,1	62,01	246,1	62,01	246,1	62,01	259,25	0,1	259,25	0,1	
	231,15	0,37	231,15	0,37	231,15	0,37	247,1	13,45	247,1	13,45	247,1	13,45	260,15	51,11	260,15	51,11	
	232,1	65,96	232,1	65,96	232,1	65,96	248,1	5,67	248,1	5,67	248,1	5,67	261,15	13,78	261,15	13,78	
	233,1	14,78	233,1	14,78	233,1	14,78	249,05	0,77	249,05	0,77	249,05	0,77	262,15	5,09	262,15	5,09	
	234,1	5,98	234,1	5,98	234,1	5,98	250,1	0,14	250,1	0,14	250,1	0,14	287,25	0,07	287,25	0,07	
	235,1	0,87	235,1	0,87	235,1	0,87							288,15	30,34	288,15	30,34	
	236,1	0,15	236,1	0,15	236,1	0,15							289,15	7,53	289,15	7,53	
	259,15	0,06	259,15	0,06	259,15	0,06							290,15	2,92	290,15	2,92	
	260,1	43,99	260,1	43,99	260,1	43,99							291,15	0,48	291,15	0,48	
	261,1	9,94	261,1	9,94	261,1	9,94							301,25	0,01	301,25	0,01	
	262,1	4,1	262,1	4,1	262,1	4,1							302,15	13,94	302,15	13,94	
	263,1	0,6	263,1	0,6	263,1	0,6							303,15	3,64	303,15	3,64	
	264,1	0,1	264,1	0,1	264,1	0,1							304,15	1,42	304,15	1,42	
													305,15	0,16	305,15	0,16	

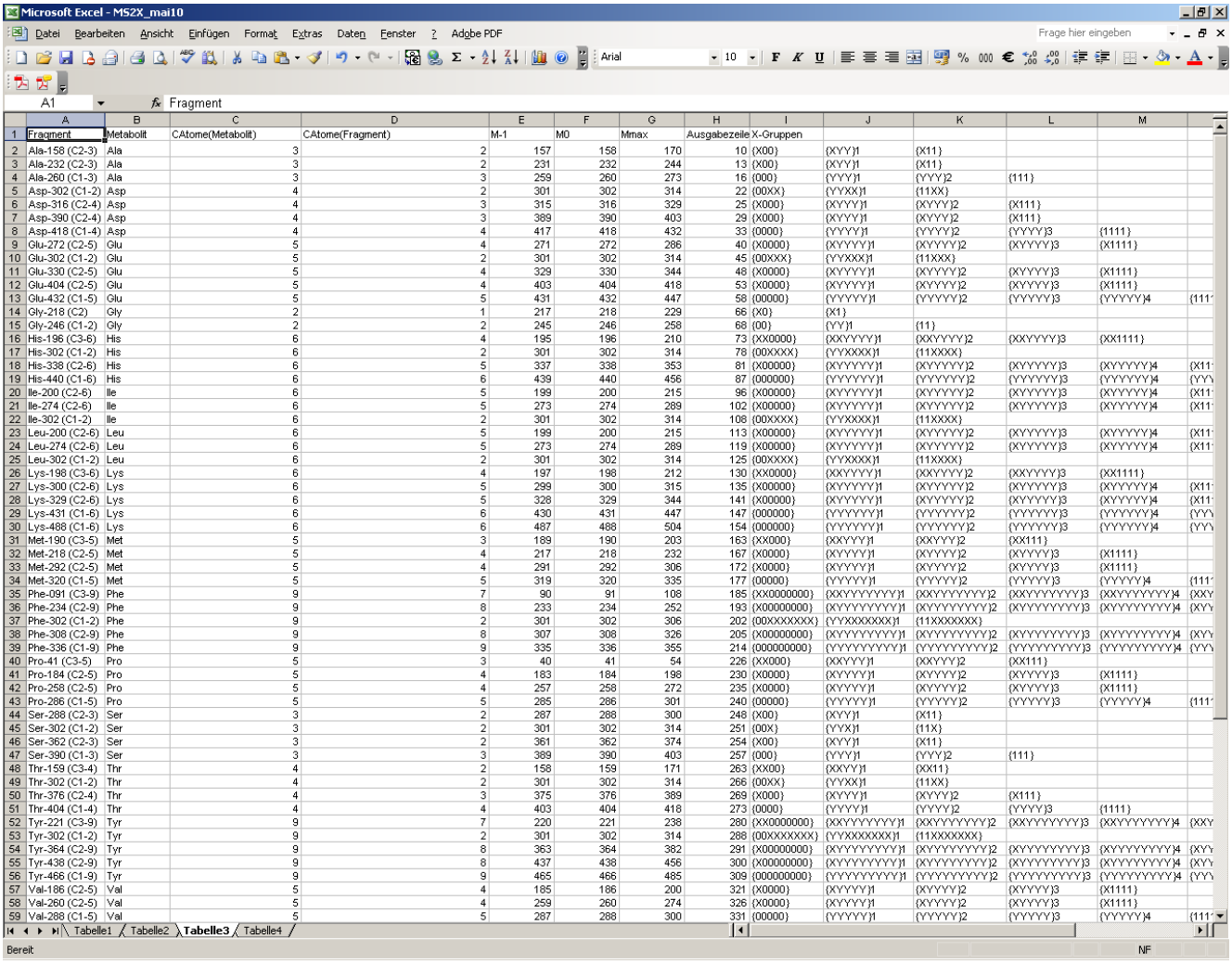
Tabelle1 Tabelle2 Tabelle3 Tabelle4

Bereit NF

Figure 28: Excel Sheet 2; Standard Metabolite Data.

Figure Legend 28: Excel Sheet 2 presents standard experimental data consisting of several metabolites (e.g. Ala, Gly, Val etc.) with information about mass to charge ratio and relative intensity values observed during experimentation.

In the second excel sheet standard metabolite based information is stored including metabolite name, date, standard m/z values, and standard relative intensity values of each metabolite, as shown in Figure 28.



Fragment	Metabolit	CAtome(Metabolit)	CAtome(Fragment)	M-1	M0	Mmax	Ausgabezeile	X-Gruppen				
Ala-158 (C2-3)	Ala	3	2	157	158	170	10	{X00}	{XYY}1	{X11}		
Ala-232 (C2-3)	Ala	3	2	231	232	244	13	{X00}	{XYY}1	{X11}		
Ala-260 (C1-3)	Ala	3	3	259	260	273	16	{000}	{YYY}1	{YYY}2	{111}	
Asp-302 (C1-2)	Asp	4	2	301	302	314	22	{00XX}	{YYXX}1	{11XX}		
Asp-316 (C2-4)	Asp	4	3	315	316	329	25	{X000}	{XYYY}1	{XYYY}2	{X111}	
Asp-390 (C2-4)	Asp	4	3	389	390	403	29	{X000}	{XYYY}1	{XYYY}2	{X111}	
Asp-418 (C1-4)	Asp	4	4	417	418	432	33	{0000}	{YYYY}1	{YYYY}2	{YYYY}3	{1111}
Glu-272 (C2-5)	Glu	5	4	271	272	286	40	{X0000}	{XYYYY}1	{XYYYY}2	{XYYYY}3	{X1111}
Glu-302 (C1-2)	Glu	5	2	301	302	314	45	{00XXX}	{YYXXX}1	{11XXX}		
Glu-330 (C2-5)	Glu	5	4	329	330	344	48	{X0000}	{XYYYY}1	{XYYYY}2	{XYYYY}3	{X1111}
Glu-404 (C2-5)	Glu	5	4	403	404	418	53	{X0000}	{XYYYY}1	{XYYYY}2	{XYYYY}3	{X1111}
Glu-432 (C1-5)	Glu	5	5	431	432	447	58	{00000}	{YYYYY}1	{YYYYY}2	{YYYYY}3	{YYYYY}4
Gly-218 (C2)	Gly	2	1	217	218	229	66	{X0}	{X1}			
Gly-246 (C1-2)	Gly	2	2	245	246	258	68	{00}	{YY}1	{11}		
His-196 (C3-6)	His	6	4	195	196	210	73	{XX0000}	{XXYYYY}1	{XXYYYY}2	{XXYYYY}3	{XX1111}
His-302 (C1-2)	His	6	2	301	302	314	78	{00XXXX}	{YYXXXX}1	{11XXXX}		
His-338 (C2-6)	His	6	5	337	338	353	81	{X00000}	{XYYYYY}1	{XYYYYY}2	{XYYYYY}3	{XYYYYY}4
His-440 (C1-6)	His	6	6	439	440	456	87	{000000}	{YYYYYY}1	{YYYYYY}2	{YYYYYY}3	{YYYYYY}4
Ile-200 (C2-6)	Ile	6	5	199	200	215	96	{X00000}	{XYYYYY}1	{XYYYYY}2	{XYYYYY}3	{XYYYYY}4
Ile-274 (C2-6)	Ile	6	5	273	274	289	102	{X00000}	{XYYYYY}1	{XYYYYY}2	{XYYYYY}3	{XYYYYY}4
Ile-302 (C1-2)	Ile	6	2	301	302	314	108	{00XXXX}	{YYXXXX}1	{11XXXX}		
Leu-200 (C2-6)	Leu	6	5	199	200	215	113	{X00000}	{XYYYYY}1	{XYYYYY}2	{XYYYYY}3	{XYYYYY}4
Leu-274 (C2-6)	Leu	6	5	273	274	289	119	{X00000}	{XYYYYY}1	{XYYYYY}2	{XYYYYY}3	{XYYYYY}4
Leu-302 (C1-2)	Leu	6	2	301	302	314	125	{00XXXX}	{YYXXXX}1	{11XXXX}		
Lys-198 (C3-6)	Lys	6	4	197	198	212	130	{XX0000}	{XXYYYY}1	{XXYYYY}2	{XXYYYY}3	{XX1111}
Lys-300 (C2-6)	Lys	6	5	299	300	315	135	{X00000}	{XYYYYY}1	{XYYYYY}2	{XYYYYY}3	{XYYYYY}4
Lys-329 (C2-6)	Lys	6	5	328	329	344	141	{X00000}	{XYYYYY}1	{XYYYYY}2	{XYYYYY}3	{XYYYYY}4
Lys-431 (C1-6)	Lys	6	6	430	431	447	147	{000000}	{YYYYYY}1	{YYYYYY}2	{YYYYYY}3	{YYYYYY}4
Lys-488 (C1-6)	Lys	6	6	487	488	504	154	{000000}	{YYYYYY}1	{YYYYYY}2	{YYYYYY}3	{YYYYYY}4
Met-190 (C3-5)	Met	5	3	189	190	203	163	{XX000}	{XXYYY}1	{XXYYY}2	{XX111}	
Met-218 (C2-5)	Met	5	4	217	218	232	167	{X0000}	{XYYYY}1	{XYYYY}2	{XYYYY}3	{X1111}
Met-292 (C2-5)	Met	5	4	291	292	306	172	{X0000}	{XYYYY}1	{XYYYY}2	{XYYYY}3	{X1111}
Met-320 (C1-5)	Met	5	5	319	320	335	177	{00000}	{YYYYY}1	{YYYYY}2	{YYYYY}3	{YYYYY}4
Phe-091 (C3-9)	Phe	9	7	90	91	108	185	{XX0000000}	{XXYYYYYYYY}1	{XXYYYYYYYY}2	{XXYYYYYYYY}3	{XXYYYYYYYY}4
Phe-234 (C2-9)	Phe	9	8	233	234	252	193	{X00000000}	{XYYYYYYYY}1	{XYYYYYYYY}2	{XYYYYYYYY}3	{XYYYYYYYY}4
Phe-302 (C1-2)	Phe	9	2	301	302	306	202	{00XXXXXXX}	{YYXXXXXXX}1	{11XXXXXXX}		
Phe-308 (C2-9)	Phe	9	8	307	308	326	205	{X00000000}	{XYYYYYYYY}1	{XYYYYYYYY}2	{XYYYYYYYY}3	{XYYYYYYYY}4
Phe-336 (C1-9)	Phe	9	9	335	336	355	214	{000000000}	{YYYYYYYYY}1	{YYYYYYYYY}2	{YYYYYYYYY}3	{YYYYYYYYY}4
Pro-41 (C3-5)	Pro	5	3	40	41	54	226	{XX000}	{XXYYY}1	{XXYYY}2	{XX111}	
Pro-184 (C2-5)	Pro	5	4	183	184	198	230	{X0000}	{XYYYY}1	{XYYYY}2	{XYYYY}3	{X1111}
Pro-258 (C2-5)	Pro	5	4	257	258	272	235	{X0000}	{XYYYY}1	{XYYYY}2	{XYYYY}3	{X1111}
Pro-286 (C1-5)	Pro	5	5	285	286	301	240	{00000}	{YYYYY}1	{YYYYY}2	{YYYYY}3	{YYYYY}4
Ser-288 (C2-3)	Ser	3	2	287	288	300	248	{X00}	{XYY}1	{X11}		
Ser-302 (C1-2)	Ser	3	2	301	302	314	251	{00X}	{YYX}1	{11X}		
Ser-362 (C2-3)	Ser	3	2	361	362	374	254	{X00}	{XYY}1	{X11}		
Ser-390 (C1-3)	Ser	3	3	389	390	403	257	{000}	{YYY}1	{YYY}2	{111}	
Thr-159 (C3-4)	Thr	4	2	158	159	171	263	{XX00}	{XXYY}1	{XX11}		
Thr-302 (C1-2)	Thr	4	2	301	302	314	266	{00XX}	{YYXX}1	{11XX}		
Thr-376 (C2-4)	Thr	4	3	375	376	389	269	{X000}	{XYYY}1	{XYYY}2	{X111}	
Thr-404 (C1-4)	Thr	4	4	403	404	418	273	{0000}	{YYYY}1	{YYYY}2	{YYYY}3	{1111}
Tyr-221 (C3-9)	Tyr	9	7	220	221	238	280	{XX0000000}	{XXYYYYYYYY}1	{XXYYYYYYYY}2	{XXYYYYYYYY}3	{XXYYYYYYYY}4
Tyr-302 (C1-2)	Tyr	9	2	301	302	314	288	{00XXXXXXX}	{YYXXXXXXX}1	{11XXXXXXX}		
Tyr-364 (C2-9)	Tyr	9	8	363	364	382	291	{X00000000}	{XYYYYYYYYY}1	{XYYYYYYYYY}2	{XYYYYYYYYY}3	{XYYYYYYYYY}4
Tyr-438 (C2-9)	Tyr	9	8	437	438	456	300	{X00000000}	{XYYYYYYYYY}1	{XYYYYYYYYY}2	{XYYYYYYYYY}3	{XYYYYYYYYY}4
Tyr-466 (C1-9)	Tyr	9	9	465	466	485	309	{000000000}	{YYYYYYYYY}1	{YYYYYYYYY}2	{YYYYYYYYY}3	{YYYYYYYYY}4
Val-186 (C2-5)	Val	5	4	185	186	200	321	{X0000}	{XYYYY}1	{XYYYY}2	{XYYYY}3	{X1111}
Val-260 (C2-5)	Val	5	4	259	260	274	326	{X0000}	{XYYYY}1	{XYYYY}2	{XYYYY}3	{X1111}
Val-288 (C1-5)	Val	5	5	287	288	300	331	{00000}	{YYYYY}1	{YYYYY}2	{YYYYY}3	{YYYYY}4

Figure 29: Excel Sheet 3; Experimental Data Information.

Figure Legend 29: Excel Sheet 3 presents experimental data consisting of several metabolites (e.g. Ala, Gly, Val, Leu, Pro, Thr etc.) with information about Fragments, Groups, Constant and Mass values.

Whereas in third excel sheet as shown in Figure 29, further metabolite based information is stored including metabolite short name, number of fragments, Carbon atom (metabolite) value, Carbon atom fragment value, Ion values (Mass -1, Mass 0 and Mass maximum values) and fixed output line value.

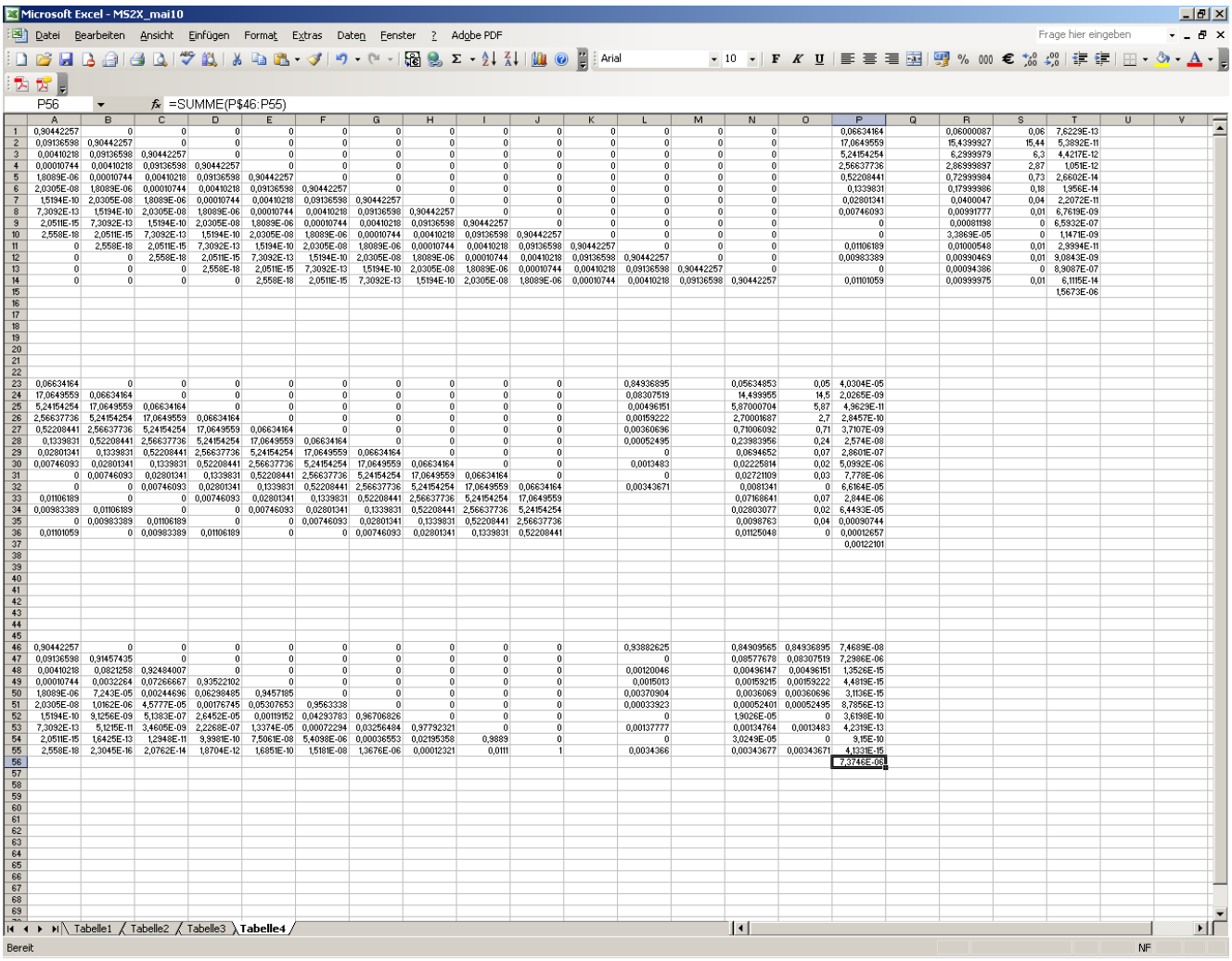


Figure 30: Excel Page 4; Outputted Information.

Figure Legend 30: Excel Sheet 4 presents measurement analysis process performing linear regression analysis and drawing matrix to measure relative abundance values.

Excel sheet number four contains the resultant data obtained as the outcome of data processing using developed macro and as shown in Figure 30 the output information is placed in sheet in following top down left to right order i.e.

- Abundance matrix generated of calculated relative natural abundance values.
- Calculated relative abundance values
- Calculated Minimum values

- New Abundance matrix generated of calculated relative abundance values
- Calculated new relative abundance values
- Calculated new Fractional Molar Abundance Values
- Calculated new Minimum values

Furthermore, another Microsoft excel file is automatically generated by the macro, containing detailed output consisting of number of percentage of calculated relative intensity values per mass to charge ration. This newly generated excel file is also consists of four more excel sheets. In the first excel sheet it just repeats the inputted experimental data, in the second, third and fourth excel sheet the resultant data in percentage is provided in an undefined order. During this all data storage, processing and resultant information analysis process, we have faced following problems i.e.

- As this process is dependent on Microsoft Excel (sheets and library elements), it is not easy for a new user to get used to of it because of completely user unfriendly graphical user interface with hectic data entering system.
- It is not easily possible to store and manage large amount different metabolite based experimental data.
- Output presentation is completely not understandable, especially for a new user, as there is no such labeling provided which can define and help user in understanding different data provisions e.g. it's not possible for a new user to know about generated abundance matrix, and to differentiate between other generated results e.g. what are the calculated relative intensity values, what are the fractional molar value and how next abundance matrix is generated etc.
- As the data processing mechanism (mathematic) is not visible, it is not possible for the user to validate obtained results (only final output is presented).
- This source code of developed and in use macro is open but still some of the source code function are restricted e.g. EUROTOOL, ATPVBAEN, FUNCRES, furthermore as this macro uses some

built-in Microsoft excel library function e.g. Microsoft Excel Solver, it is also not possible to improve and validate implemented source code for mathematical enhancements.

- One of the biggest problems is regarding the deployment of this developed macro, as it is not easily possible it to move to other machine and configure the settings, especially a user who is unfamiliar with Microsoft Excel in details.

Although the methodology could be very good to follow but implemented software system is not easily adoptable and verifiable. Furthermore it is not possible to validate generated output by the system which could be a big drawback regarding the authenticity of results produced by this application using experimental metabolite data.

4.2 Methodology

For probabilistic estimates, some already well known, established, validated and published mathematical algorithms were considered. The implementation of such a software application makes it capable of analyzing metabolite based experimental data for the measurement of natural or theoretical abundance values, relative intensity or vector values, fractional molar abundance values, minimum values and percentage of calculated relative intensity values with respect to each mass to charge ratio. A binomial expansion is used for the measurement of natural abundance values. An abundance matrix is drawn and multiple regression analysis performed. We used the partial Brauman's least square algorithm for the measurement of relative intensity values with respect to each m/z values. Using estimated relative intensity values, there is a newly drawn abundance matrix and Pseudo inverse matrix calculated; we have estimated actual values and percentages of relative abundances, fractional molar abundances and minimum values with respect to the number of fragments. This whole procedure is repeated twice to obtain efficient values. A third (optional) repetition validates results and convergence. Using resultant natural, relative and fractional Molar abundances, absolute ^{13}C enrichment, mean and standard deviation is measured.

The reliability in the results produced by this technique depends on a number of factors. At first the analysis is based on the assumption that the fragmentation patterns for all heteroatom isotopes are

identical (i.e. no differential isotope effect), the relative abundance (actual) values of the isotopes is known and either the natural abundances are known or measured in some way.

4.2.1 Natural Abundance Value Calculation

In case if the natural (theoretical) abundance values are not available (or estimated), the first step towards mass isotopomers distribution analysis is to determine the natural abundance values because these estimated abundances will then be used for the construction of abundance matrix for multiple regression analysis to estimate contribution of isotopes from derivative compounds to mass spectrum (Wolfgang et al., 1984).

Natural abundance values can be estimated by the isotope contents of biosynthesized subunits from polymerized product (Zilversmit et al. 1943). We used for this the binomial expansion (Eq. 5: Section 3.1) for the measurement of natural abundances (Hellerstein et al. 1992).

4.2.2 Abundance Matrix

To predict the relative isotopomers distribution, at first linear regression analysis is performed (Eq. 7) using computed natural abundance values by binomial expression.

$$\begin{pmatrix} A_n 0 0 0 0 \\ A_{n+1} A_n 0 0 0 \\ A_{n+n} A_{n+2} A_{n+1} 0 0 \\ A_{n+n} A_{n+3} A_{n+2} A_{n+1} 0 \\ A_{n+n} A_{n+4} A_{n+3} A_{n+2} A_{n+1} \\ \cdots \end{pmatrix} \quad \text{(Eq. 7)}$$

4.2.3 Relative Isotopic Abundance Value Calculation

Brauman's algorithm (Brauman, 1966; Korzekwa et al., 1990) is a least squares technique to calculate relative isotopic abundances (also called as calculated relative intensity values) by simplifying the mass spectra of molecules containing elements with many isotopes by dealing with complex spectra based on fragmentation of molecules (containing heteroatom). The method has been divided into two parts i.e. first the generation of an appropriate set of linear simultaneous equations and second the solution of these

equations. The complete Brauman's least square algorithm is presented in Eq. 4 (Section 3.1). But we are using the partial Brauman's least square algorithm (equation) to calculate relative intensity values, presented in Eq. 8.

$$X = A^{-1} P \qquad \text{(Eq. 8)}$$

X is equal to the product of A inverse times P. Where 'X' is equal to the calculated relative intensity values, 'A' is the drawn (square) abundance matrix of estimated natural abundance values and 'P' is the set of actual relative intensities as observed during a GC-MS experiment.

To compute relative intensity values, linear regression analysis is performed using drawn abundance matrix (Eq. 7), with the partial implementation of Brauman's least square method (Eq. 8).

$$\begin{pmatrix} A_n 0000 \\ A_{n+1} A_n 000 \\ A_{n+n} A_{n+2} A_{n+1} 00 \\ A_{n+n} A_{n+3} A_{n+2} A_{n+1} 0 \\ A_{n+n} A_{n+4} A_{n+3} A_{n+2} A_{n+1} \\ \dots \end{pmatrix} \Rightarrow A^{-1} * P \Rightarrow \begin{pmatrix} Ri_1 \\ Ri_2 \\ Ri_3 \\ Ri_4 \\ \dots \\ Ri_n \end{pmatrix} \qquad \text{(Eq. 9)}$$

As presented in Eq. 9, the length of the drawn abundance matrix depends upon the total number of mass to charge ratio and actual relative intensity values. In Eq. 9, Ri_1, Ri_2, Ri_3, Ri_4 … Ri_n, are estimated relative intensity values, estimated with respect to the each m/z values. 'n' in this case is the number of total number of m/z values and length of Abundance Matrix, and 'P' is the actual observed relative intensity values.

4.2.4 Relative Abundance Value Calculation

To compute relative abundance values, again linear regression analysis is performed by again drawing a new abundance matrix (second threshold) but this time the input values are estimated relative intensity values (Ri) using Eq. 10 and the length of the abundance matrix depends on the user inputted number of

fragments. For example if the number of fragments runs from 0 to 3 and computed relative intensity values are five, then a non square matrix of four by five will be drawn.

In Eq. 9, Ra_1, Ra_2, Ra_3 … Ra_n, are estimated relative abundance values, estimated with respect to the number of fragments. "n" in this case is the number of total number of fragments values and length of newly drawn (second) Abundance Matrix, and "P" is the observed (standard) relative intensity values.

$$\begin{pmatrix} Ri_n 0000 \\ Ri_{n+1} Ri_n 000 \\ Ri_{n+n} Ri_{n+2} Ri_{n+1} 00 \\ Ri_{n+n} Ri_{n+3} Ri_{n+2} R_{n+1} 0 \\ \dots \end{pmatrix} \Rightarrow A^{-1} P \Rightarrow \begin{pmatrix} Ra_1 \\ Ra_2 \\ Ra_3 \\ \dots \\ Ri_n \end{pmatrix} \qquad \text{(Eq. 10)}$$

In most of the cases the number of fragments and relative abundance values will not be equal (in number). Then, instead of a square abundance matrix a non-square abundance matrix will be drawn. Mathematical it is not possible to implement Eq.5 as a non-square matrix because the inverse of a matrix can only be computed if it will be a square matrix (number of rows must be equal to the number of columns). To resolve this issue we have used the Pseudo Inverse (also called Generalized Inverse Matrix) of the drawn abundance matrix, as presented in Eq. 11.

$$X = \text{Pseudo Inverse } (A) * P \qquad \text{(Eq. 11)}$$

X is equal to the product of Pseudo Inverse of A times P. In Eq. 11, A is the second drawn abundance matrix consisting of computed relative intensity values and P are the actual observed relative intensity values.

4.2.5 Fractional Molar Abundance Value Calculation

Fractional molar abundance is the mass fraction of one element to the total mass of a compound. To calculate fractional molar abundance, the equation is

$$F = Ri * Ra \qquad \text{(Eq. 12)}$$

In Eq. 12, F is equal to the calculated relative abundance values times calculated abundance matrix of relative intensity values. Here, 'Ri' is the drawn abundance matrix consisting of Ri values, and 'Ra' is the measured relative abundance values.

4.2.6 Minimum Value Calculation

To calculate minimum values, the equation is

$$MinVal = F - Ri \qquad \text{(Eq.13)}$$

In Eq. 13, MinVal is equal to the calculated fractional molar abundance values minus relative intensity values (standard in first transition and actual in second transition). There 'F' denotes the calculated fractional molar abundance values and 'Ri' the respective relative intensity values.

4.2.7 Mathematical Validation

To mathematically validate the calculations, a new linear regression analysis is performed. A new abundance matrix is drawn, whose length is equal to the number of fragments but the difference is that each of its columns consists of new values e.g. if the number of fragments is 2 then three different natural abundance values will be estimated using binomial expansion three times, at first n will be set to 3, then to 2 and then to 1.

$$\begin{pmatrix} N1_1 & 0 & 0 & 0 \\ N1_2 & N2_1 & 0 & 0 \\ N1_2 & N2_2 & N3_{n+1} & Nn_{n+1} \\ \cdots & & & \\ N1_n & N2_n & N3_n & Nn_n \end{pmatrix} \qquad \text{(Eq.14)}$$

In Eq. 14, N_1, N_2, N_3 … N_n, are the estimated natural abundance values. Likewise before, new relative abundance values are estimated, and then the difference is calculated between the newly calculated

fractional molar abundance and calculated relative abundance values. A minimum difference validates the result and indicates convergence.

4.3 V-Model

The software development of any kind should be done following some process model. There are already some well established development models existing and followed e.g. Waterfall model, Spiral model, Iterative and incremental development, Agile development, Code and fix, and some Process improvement models. During our software design and development, we are following a well established software development model i.e. V-Model; an extended form of waterfall model proposed by Paul Rook.

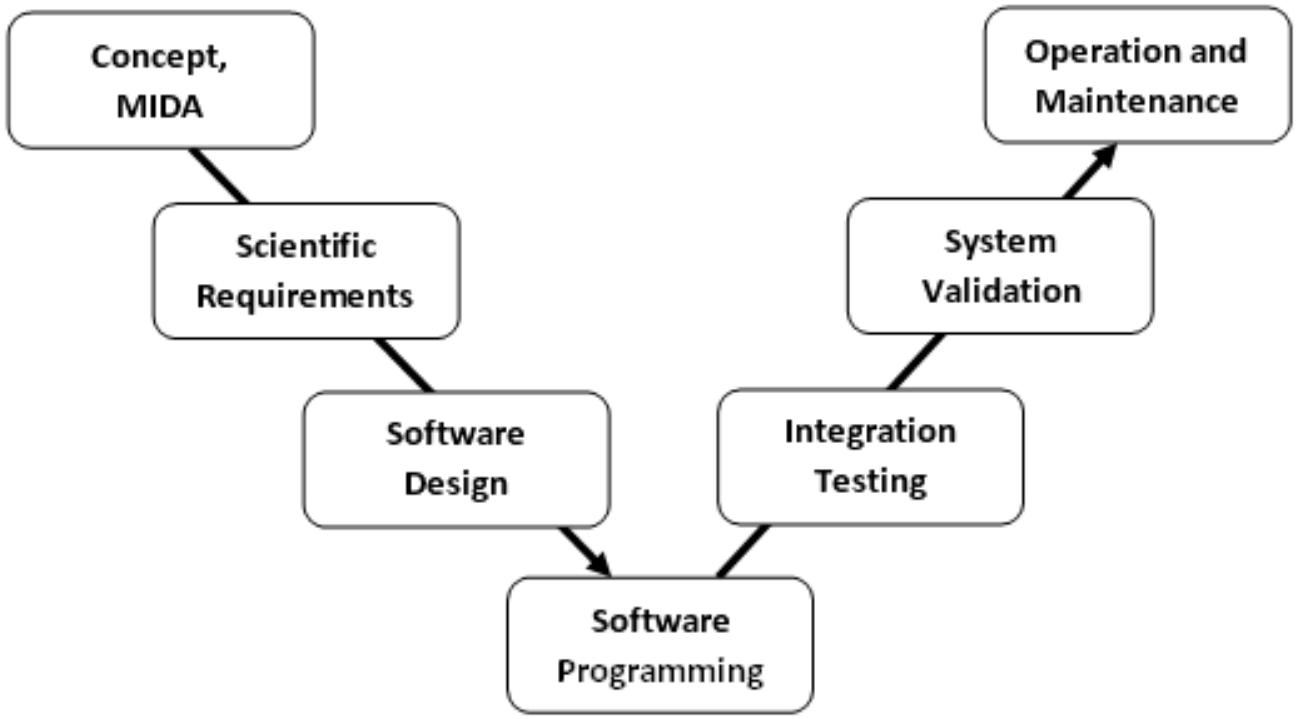


Figure 31: Isotopo; V-Model Software Development Process

Figure Legend 31: Software development model consisting of seven phases: Concept MIDA, Scientific Requirements, Software Design, Software Programming, Integration Testing, System Validation and Operation and Maintenance.

The V-Model expresses the relationships between each phase of the development life cycle forming typical V shape (Figure 31). The overall job of Isotopo V-Model software development process starts with the initialization of main concept (which in our case was MIDA), then scientific requirements for

operational scenarios have to be clearly described and to be strictly followed to model Isotopo. Later on following architected software designs, a real time system has to be developed using programming, which then has to be in house tested (integrated) and validated by scientists. The final step is to maintain Isotopo and if needed then repeat V-Model for software releases with more computational and feature updates.

4.4 UML Description

As computational and empirical software systems development becomes more complex, scientific academias as well as commercial organizations require high-quality products in short time. Unfortunately usually wrong presumptions leads to direct software development without adopting software development life cycle and formal design modelling which gives a temporary and limited (scripted) solution and in the long run it is quite difficult to enhance and improve it. Software design modelling helps in dealing with complexity as the meta-model architecture provides abstraction and modification techniques which allows the designer to concentrate on the basis of a problem by reducing gratuitous details.

Today, a better way of architecture modelling for a newly proposed software application is available in the form of Unified Modelling Language (UML). It is a modelling language, a well suited and the standard way of designing software application by creating different abstract models. UML is capable of facilitating software engineers stand alone and interconnected semiformal (Meta) design views for modelling software architectures.

Here software designs are created using UML principles to have better understanding of Isotopo in terms of its implementation, usage and working, Designed UML diagrams describe over all feature based functionality, user accessibility, experimental data flow, internal system work flow, system sequence, involved component's integration and source code structure. In this manuscript we present following Isotopo UML diagrams (Ahmed et al., 2012b): Use Case, Data Flow, System Sequence, Internal Work Flow, Component and Class Diagrams. This logical design presentation will give an overall physical view of the Isotopo focusing on its technical architecture, grouped functionalities, flow of information, operational perspective focusing on interface requirements and involved technologies during software design, development, deployment and testing.

4.4.1 Use Case Diagram

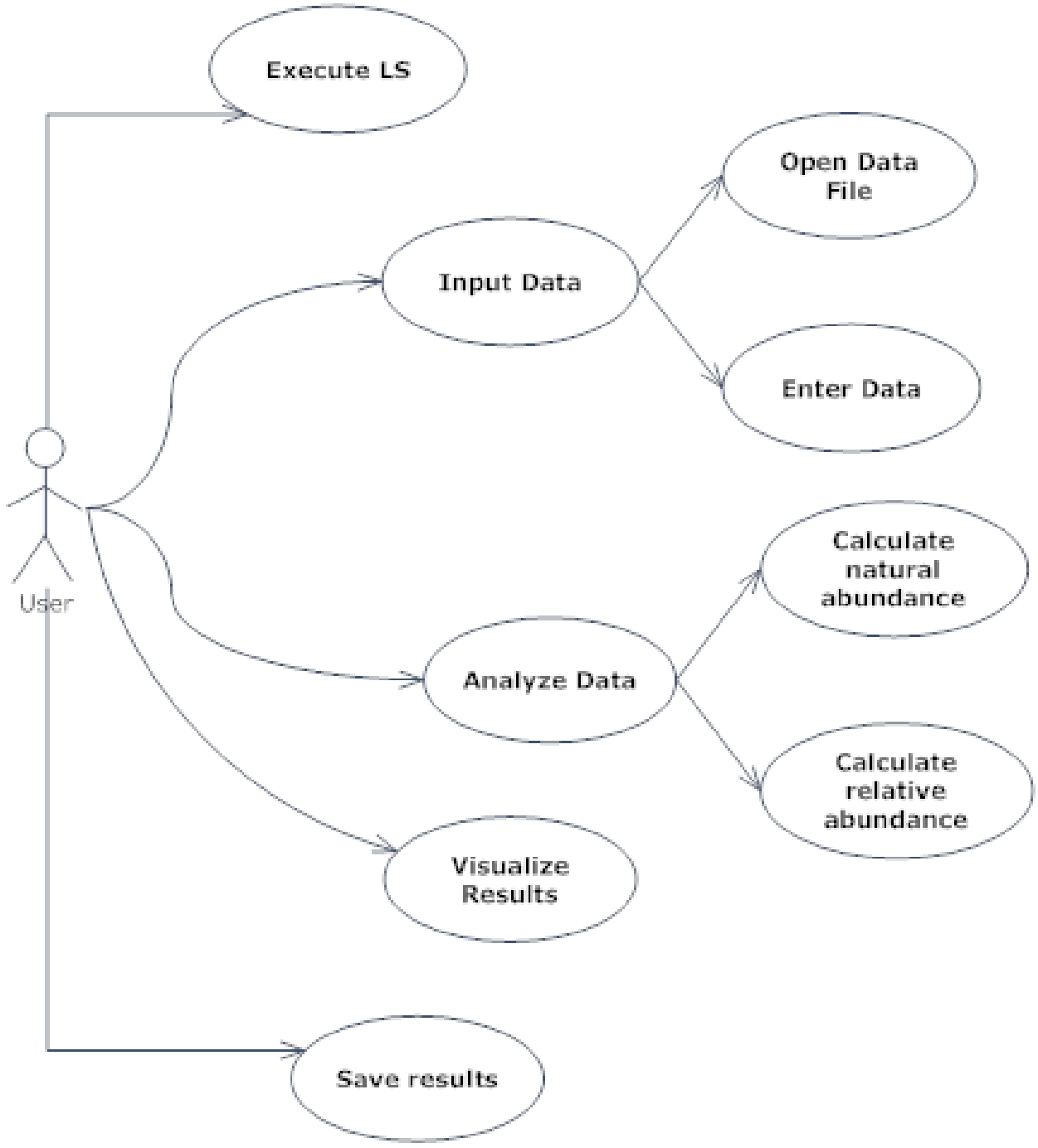


Figure 32: Isotopo; UML Use Case Diagram.

Figure Legend 32: Use case diagram of Isotopo is consisting of a User, five direct and four remote (indirect) activities.

We have designed a use case diagram (Figure 32) and explained in detail (Table 9). The designed use case diagram describes the user system communication for the isotopomers experimental data analysis, which consists of a user (actor), five direct activities (Execute LS, Input Data, Analyze data, Visualize Results, Save results) and four indirect activities (Open Data File, Enter data, Calculate natural abundance, Calculate relative).

Use case diagram explains over user system interaction. As shown in Figure 32, at first user needs to execute the software application Isotopo, then user can input experimental data in two ways to the software application for analysis, by entering manually and by loading experimental data file. Later after inputting data, user can analyze it to calculate relative abundances. After then, user can visualize obtained results by drawing a mass spectrum and also can save results in the form of an image file.

Use Case	**Details**
Number	1
Name	ISOTOPO data analyzer
Application	ISOTOPO
Description	This use case consists of a User, five direct and four remote (indirect) activities. This describes the user (actor) system (Isotopo) communication for the isotopomers experimental data analysis.
Primary Actor	User (1 Actor)
Precondition	Software application successfully running.
Trigger / Events	1. Execute Isotopo 2. Input Data 3. Open Data File 4. Enter data 5. Analyze data 6. Calculate natural abundance 7. Calculate relative 8. Visualize Results

	9. Save results
Basic Flow	Basic flow consists of following steps: • Start software application • Enter input data by either loading from data file or by manually entering. • Analyze input data. • Visualize results • Observe predicted results (text and image). • Save obtained results for reuse.
Alternate Flows	Exception will be notified to the user.

Table 9: Isotopo; Use Case Description.

4.4.2 Data Flow Diagram (DFD)

Data Flow diagram presents the basic data flow inside the Isotopo analyzer, as shown in the Figure 33, data is loaded from the Data File as input so called I/O Data, which is then analyzed by the system . Systematic analysis procedure is divided in to two levels. First level starts by calculating relative and natural abundances using actual intensities using user inputted experimental. Then fractional molar abundance values and minimum abundance values are calculated using already calculated relative and natural abundance values.

In second level, again, new relative natural abundance values are calculated using previously calculated relative abundance values in level 1. Then relative abundance values are calculated using standard intensity values, inputted by user. Using these two newly calculated relative and natural abundance values, likewise level 1, fractional molar abundance and minimum values are calculated. In third level relative difference between the observations of level 1 and 2 is calculated.

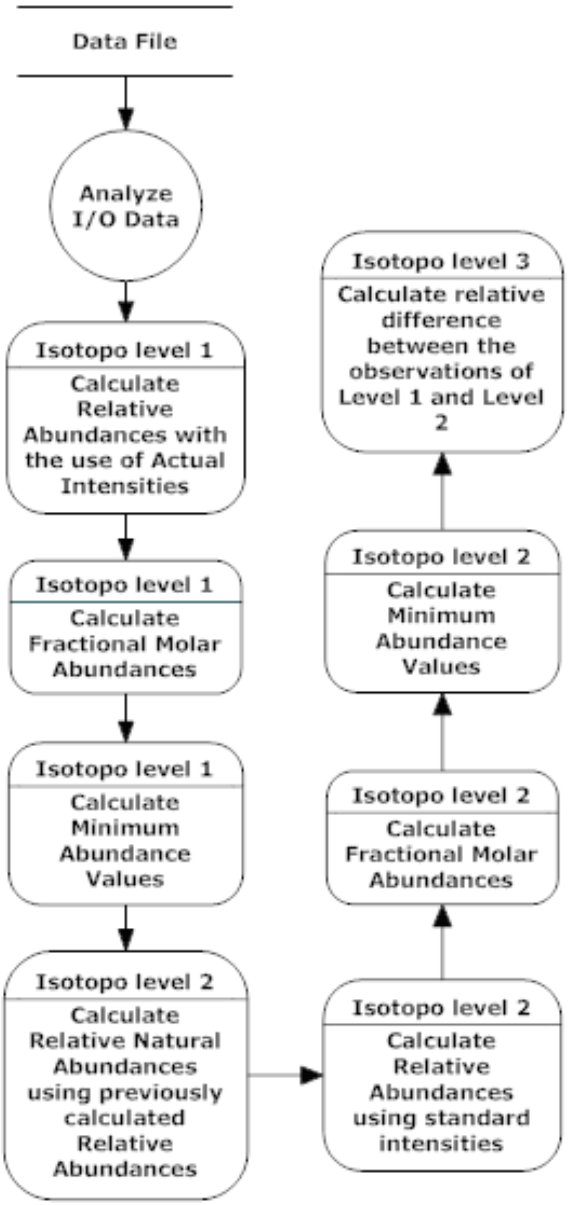


Figure 33: Isotopo; UML Data Flow Diagram.

Figure Legend 33: The data flow diagram of Isotopo is consisting of a File (Data File), one main Function (Analyze I/O Data) and eight internal functions: Calculate Natural Abundance Values and Calculate Relative Abundance Values, Calculate Fractional Molar Abundance, Calculate Minimum Abundance Values, Calculate Relative Natural Abundances using previously calculated Relative Abundances, Calculate relative difference between the observations of Level 1 and Level 2, Calculate new Minimum Abundance, Calculate new Fractional Molar Abundance and Calculate Relative Abundances using standard intensities.

4.4.3 Flow Chart

Isotopo is a MDI software application, developed following the principles of embedding child windows under a single parent window by creating nested hierarchies. To meet the aforementioned goals of Isotopo development, the graphical user interface of this application is divided into two main modules i.e. Isotopo Analyzer and Isotopo Data Manager.

Isotopo Analyzer is the module responsible for providing options for experimental data loading, analysis and visualization, whereas Isotopo Data Manager is the module responsible for providing options for experimental data manipulation and management. Isotopo is licensed software but freely available for academic use (on request).

The flow of events of the implemented Isotopo Analyzer is according to the previously discussed mathematical algorithm but technically it consists of a sequence of operations performed (Figure 34). At first the experimental data is put in: the information which metabolite, its actual m/z values, its actual relative intensity values, standard relative intensity values and the number of fragments obtained.

The system first reads and analyzes the input and performs data validation (data processing is only performed when input data is successfully validated). In the second step, after successful data validation, at first Mass (M_o, M_{-1} and M_{max}) values are calculated, then using natural abundance values along with the percentage at each fragment are calculated.

Using the estimated natural abundance values, a square abundance matrix is drawn to compute relative intensity values and their percentage with respect to each m/z value. A new abundance matrix is drawn using estimated relative intensity values to measure relative abundances using the Pseudo Inverse with actual relative intensity values. Then using all estimated abundances, the fractional molar abundances, minimum values (differences), mean, standard deviation and 13C absolute enrichments are calculated as the final outputs.

In one complete processing cycle, all actual relative intensity values against one m/z value are processed. Later average abundances are calculated and final calculations (13C Absolute Enrichments, Standard Deviation and Mean) are based on average relative abundances (please see above mentioned definitions).

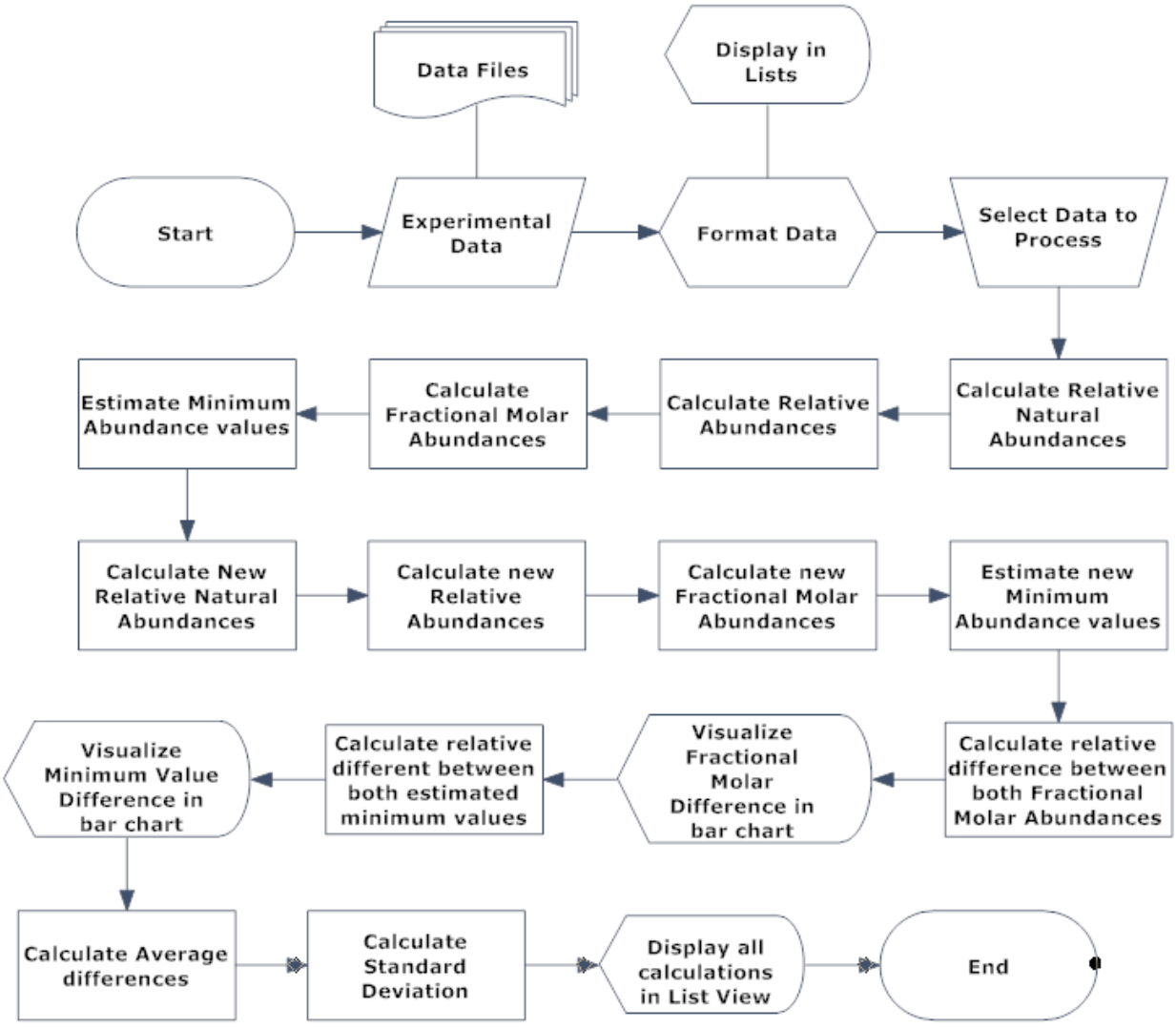


Figure 34: Isotopo; UML Flow chart of Data Analyzer.

Figure Legend 34: Visual presentation of the UML based flow chart. The implemented flow of operations performed during experimental data input, processing, analysis and visualization is given. The flow chart of Isotopo Data Analyzer is consisting of one starting point (Start), one Input point (Experimental Data; Data Files), one formatting point (Format Data; Display in lists), one data selection point (Select Data to Process), processing units (Calculate Natural Abundance Values, Relative Abundance Values, Fractional Molar Abundance, Minimum Abundance Values, Average Differences and Standard Deviation), one visualization mode (Draw Spectrum) and one ending point (End).

Figure 35: Isotopo; UML Flow chart of Data Manager.

Figure Legend 35: The flow chart of Isotopo Data Manager is consisting of one starting point (Start), one Input point (Experimental Data; Data Files), one formatting point (Format Data; Display in lists), one data file creation process, one visualization mode and one ending point (End).

Internal work flow of the Isotopo data manager (Figure 35) starts with experimental data manipulation, which leads to the experimental data extraction from existing data files and storing data by creating new data files. Furthermore it displays data loaded from data file into system and let user manipulate it by adding some new data, merging data from other files, deleting some data, and updating data.

4.4.4 System Sequence Diagram (SSD)

The system sequence of Isotopo Analyzer is consists of seven sequential steps with individual tasks. As shown in Figure 36, as first the experimental data based on the information of metabolite, actual m/z values, actual Ri values, standard m/z values, standard Ri values and number of fragments is inputted to the system i.e. Isotopo, via the user via graphical user interface. System at first analyzes inputted data by validating it. In the second step, when the data is successfully validated, and sent to Isotopo Analyzer, to calculate abundance values. Isotopo Analyzer, as the third step, sends information based on number of fragments to the relative natural abundances, which calculates relative natural abundances (n) and send them to Isotopo Analyzer.

Then, to calculate relative abundance values, Isotopo Analyzer sends calculated natural abundance values and actual Ri values to the step 4 i.e. relative abundance values calculator, which then first calculates abundance matrix of calculated natural abundance values (Mn) using step 5 of system sequence i.e. abundance martix. Later after performing mathematical operations, sends back the resultant relative abundance values (RMn) to the Isotopo Analyzer.

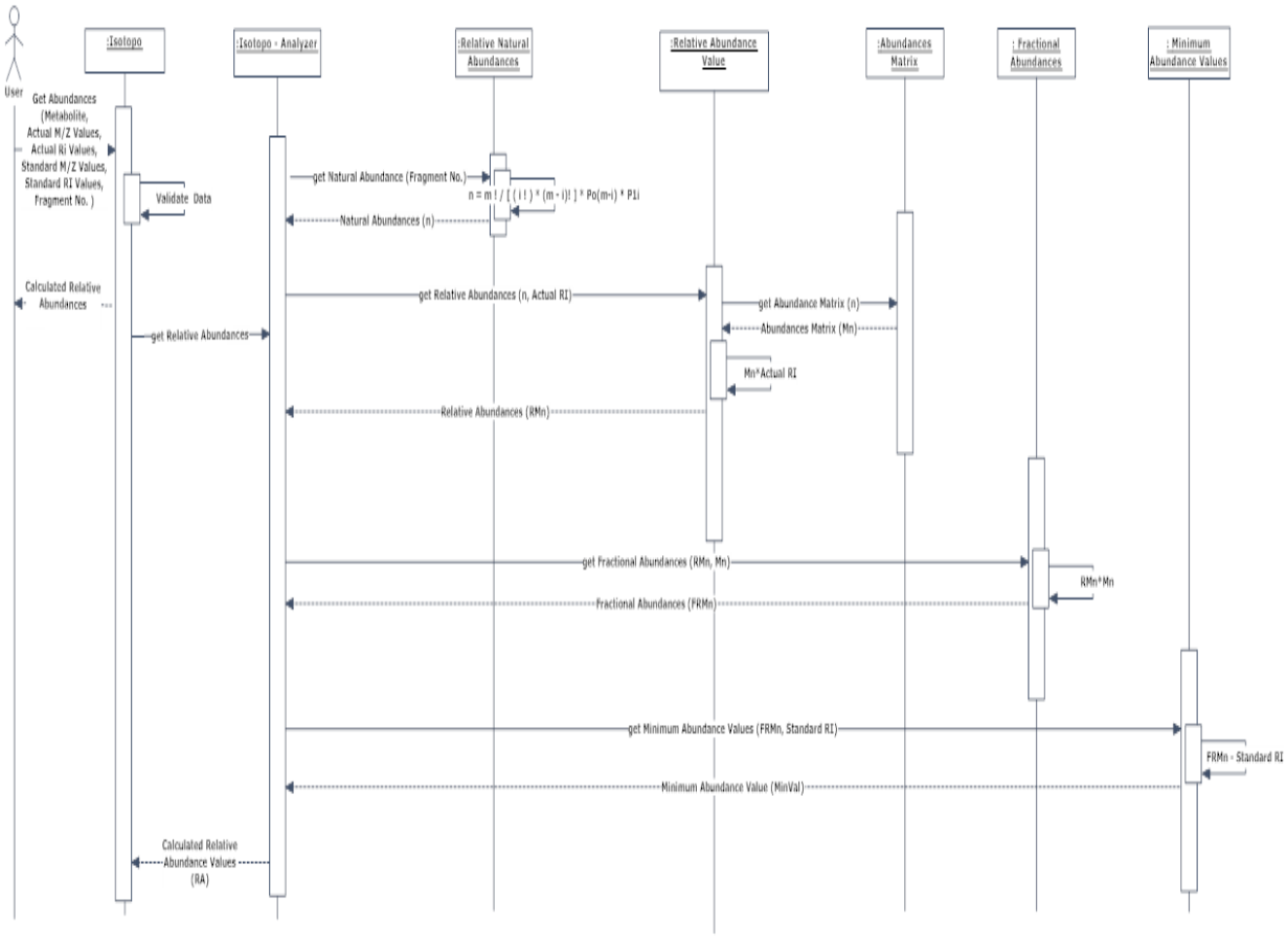


Figure 36: Isotopo; UML System Sequence Diagram.

Figure Legend. 36: The abstract system sequence diagram of Isotopo is consisting of seven steps (Isotopo, Analyzer, Relative Natural Abundances, Relative Abundance Values, Abundance Matrix, Fractional Abundances and Minimum Abundance values) with several directing arrows in between.

Then, to calculate Fractional Molar Abundance values, Isotopo Analyzer sends natural abundance matrix values and calculated relative abundance values to the step 6 i.e. Fractional Abundances, which later after the mathematical operation performance sends back resultant fractional molar abundance values (FRMn) to the Isotopo Analyzer. Later after, to calculate minimum abundance values, Isotopo Analyzer sends calculated fractional molar abundance values, standard relative intensity values to step 7 i.e., Minimum Abundance Values, which then later after the calculation of minimum values send back resultant values

(Min Val) to the Isotopo Analyzer. This was the first complete transaction of different abundance value calculations.

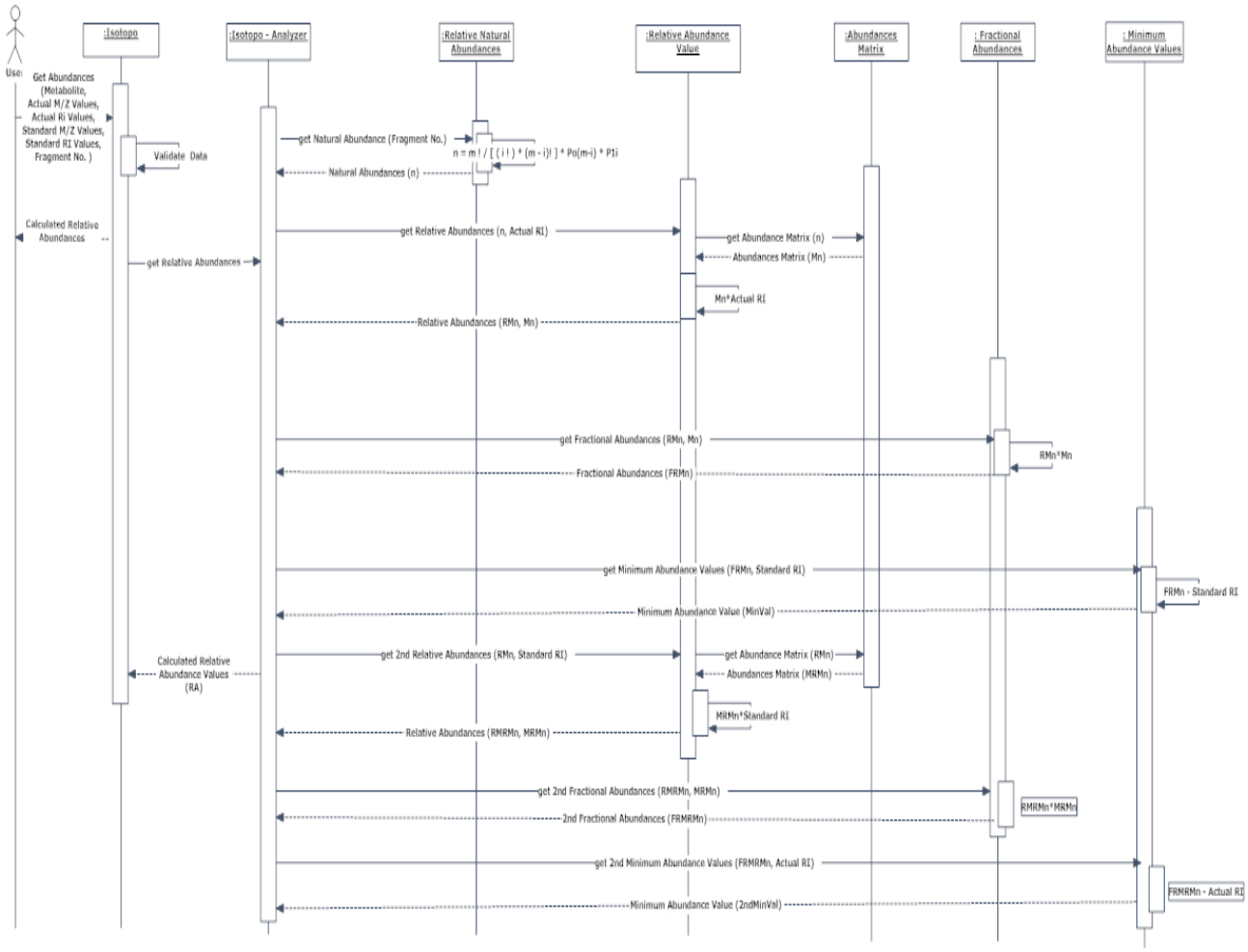


Figure 37: Isotopo; UML SSD with all Abundances.

Figure Legend. 37: The full system sequence diagram of Isotopo is consisting of repetitive iterative seven steps (Isotopo, Analyzer, Relative Natural Abundances, Relative Abundance Values, Abundance Matrix, Fractional Abundances and Minimum Abundance values) with several directing arrows in between.

Here, now as shown in Figure 37, Isotopo Analyzer will repeat the same mechanism once again with different values. Likewise before at first new relative abundance values (RMRMn) are calculated with the use of already calculated relative abundance values and Standard relative intensity values, by calculating new abundance matrix (MRMn). Then likewise before new Fractional Molar Abundance values

(FRMRMn) are calculated with the use of newly calculated relative abundance values and Abundance matrix values. Then again likewise before, with the use of newly calculated Fractional Molar Abundance values and actual relative intensity values, new minimum values (2ndMinVal) are calculated.

4.4.5 Class Diagram

The source code of developed Isotopo is divided into three namespaces i.e. SBEDA, System and ZEDGraph. SBEDA is the main namespace containing all related and newly developed source code classes, System is the by default namespace provided by C-Sharp language used during the software development, and this namespace is responsible for providing access to default language based controls and components. Namespace REDGraph is a third party application programming interface used mainly for the development of graphical visualization of statistical, mathematical and experimental data in the form of two and three dimensional colored bar charts.

There are seven newly developed interlinked classes i.e. Main, IsotopoDataAnalyzer, Isotopo DataManager, Isotopo About, Calculation, Complex and Matrix, as shown in Figure 38. As Isotopo is a MDI application, Main MDI parent class which contains all other child classes. IsotopoDataAnalyzer is the multi attribute class developed as the graphical user interface of the Isotopo analyzer which provides all visual options to the user to load, edit, analyze and visualize experimental data and observed results.

IsotopoDataManager is the multi attribute class developed as the graphical user interface of the Isotopo Data Manager which provides all visual options to the user for file based experimental data management and manipulation including entering, loading, editing, updating, deleting, merging, replacing and saving data in files. Calculations is the multi attribute class developed for performing all mathematical operations including mass value estimations, relative abundances, data parsing and different data format conversions. Matrix is the multi attribute class developed for performing matrix operations including drawing simple matrix of NxM rows and columns, calculation inverse and transpose of matrix. Complex is the multi attribute class developed for difficult mathematical operations including square root, absolute, tangent and operator overloading. IsotopoAbout is the single attribute class, providing information ISOTOPO and development team and research group.

Main sequence of classes starts with Main container class, which provides other graphical user interface based classes IsotopoDataAnalyzer, IsotopoDataManager and IsotopoDataAbout. IsotopoDataAnalyzer

perform user system communication, let user enter, edit and visualize experimental data, and analyze experimental data by directly using class Calculations which the uses classes i.e. Matrix and Complex. IsotopoDataManager is an independent multi attribute class performing operations including user system communication for file based data management and manipulations.

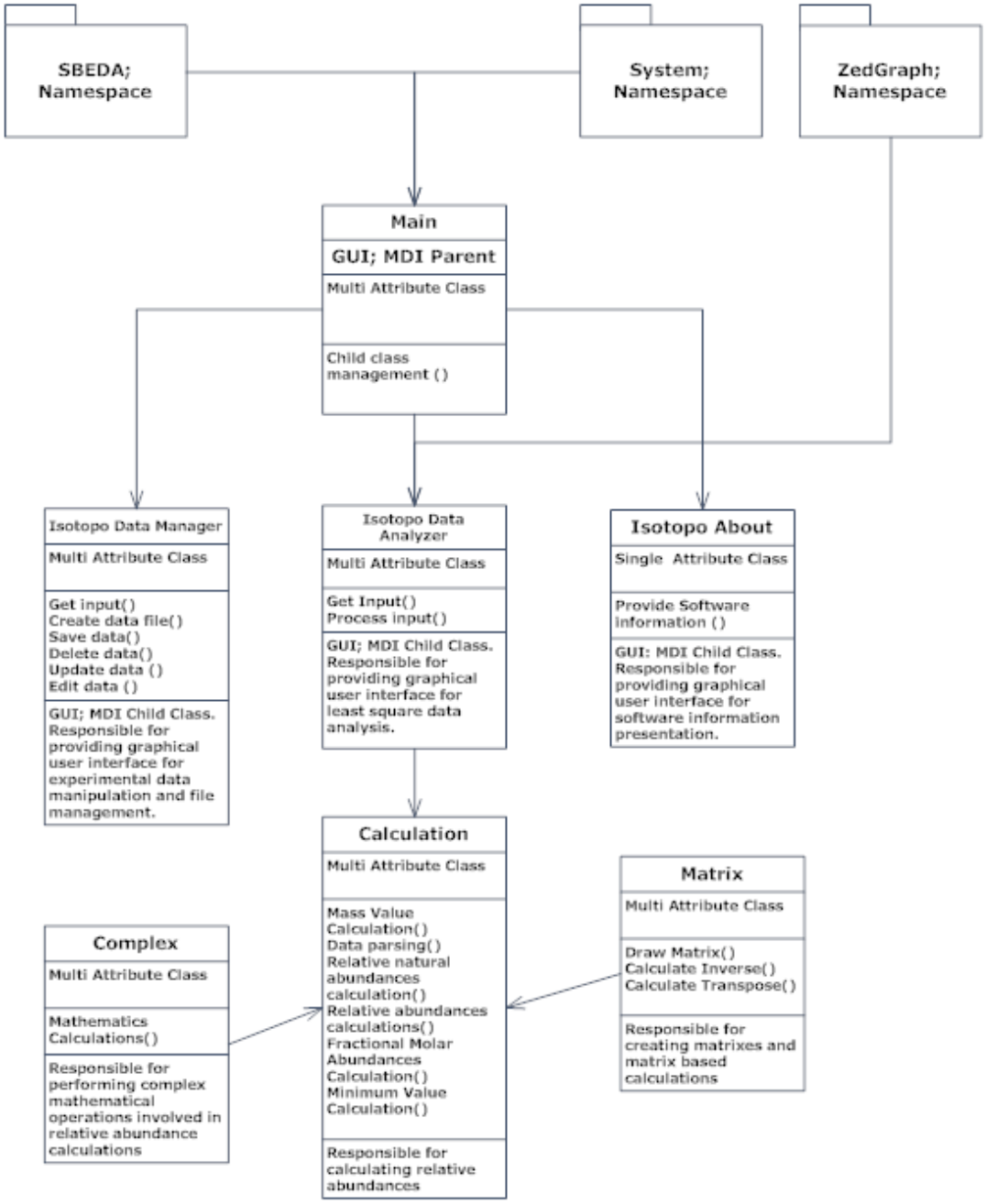


Figure 38: Isotopo; UML Class Diagram.

Figure Legend 38: The class diagram of LS-MIDA consisting of three name spaces (SBEDA, System and ZEDGraph), one main class (Main), six multi attribute classes (IsotopoDataAnalyzer, IsotopoDataManager, IsotopoAbout, Calculation, Complex and Matrix) and one single attribute class (IsotopoAbout).

4.4.6 Component Diagram

Isotopo is mainly consists of two modules i.e. Isotopo Data Analyzer and Isotopo Data Manager as shown in Figure 39. User can access these both modules to perform mass isotopomers distribution estimation and file based experimental data management. Isotopo, as whole application, is developed using an object oriented platform independent language C Sharp (C#) and Microsoft Dot Net technology.

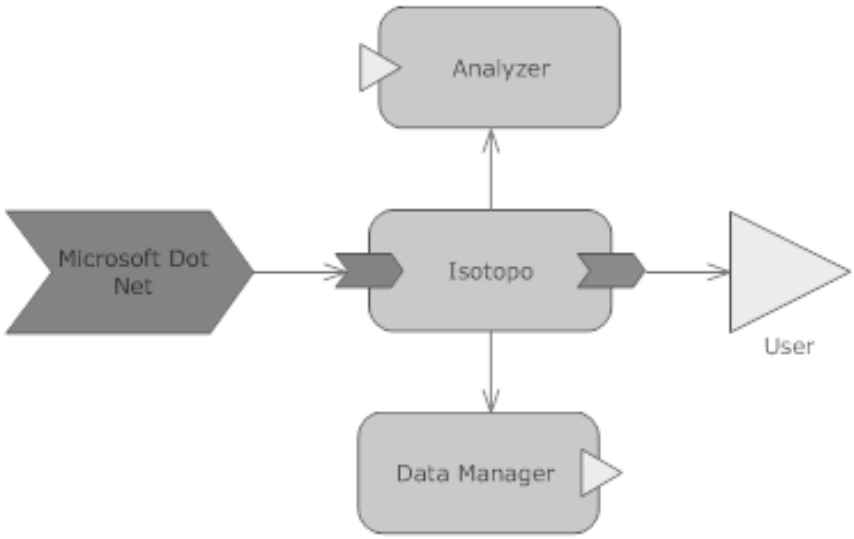


Figure 39: Isotopo; UML Component Diagram.

Figure Legend 39: The component diagram of Isotopo is consisting of one platform component (Microsoft Dot Net), one main component (LS), two subcomponents (Analyzer and Data Manager).

4.5 Isotopomers Database

The scope of research and development of designed modules is limited with respect to its platform dependency, insecurity and incapability of maintaining large data sets. Extending the scope of this research towards the proposition of an efficient schematic database structure, this section presents a platform independent relational database management system to gather, record and maintain direct self derived experimentally data by research scientists and from explicitly publicly available online repositories.

Spectral isotopomers experimental preprocessed (input) data consists of metabolite name, date, actual m/z values, standard relative intensity values, actual relative intensity values, number of fragments, carbon atom (metabolite) value, carbon atom fragment value, and fixed output line value of each metabolite.

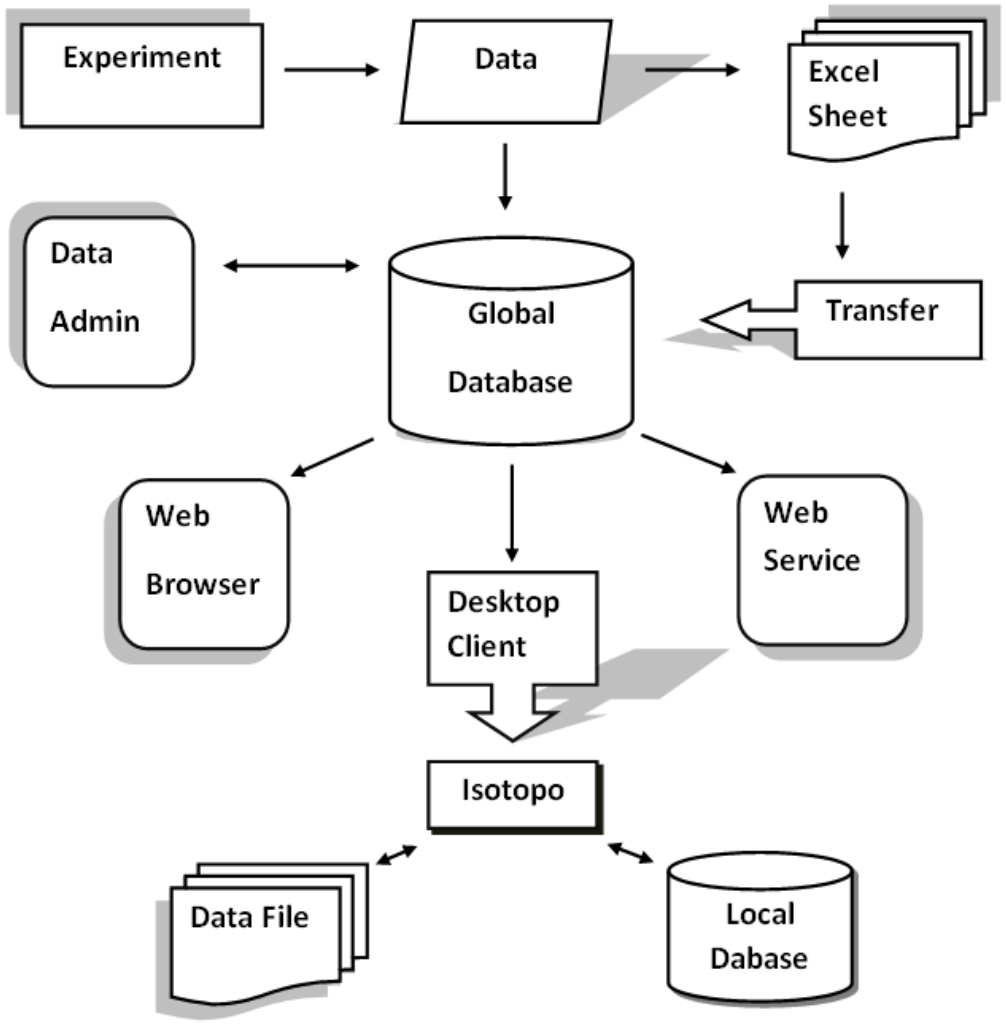


Figure 40: IsoDB; Schematic diagram of database structure.

Figure Legend 40: This Figure presents designed schematic structure for isotopomers experimental data extraction, storage, management, sharing and analysis.

However processed isotopomers experimental data (output) includes ion values (M_{-1}, M_0 and M_{max} values), relative intensity values per m/z value, natural abundance values, relative abundance values,

fractional molar abundance values per number of fragments, absolute enrichment values from both estimated natural and relative abundance values with standard deviation and mean of the output. It is not easily possible to store and manage large amount different metabolite based experimental data. If data (preprocessed and processed) is not properly maintained then it makes difficult to use it for statistical and probabilistic analysis. A new normalized database (Codd 1970), (Codd 1971), (Kent 1983), (Date 1986) and (Silberschatz et al. 2006) schematic structure (Figure 40) is designed for isotopomers experimental data extraction, storage, management, sharing and analysis.

Designed IsoDB schema is divided into three main parts. First part is about experimental data extraction and storage, second part is global data sharing and third part is data analysis. After experimentation data is extracted from scientific labs and stored directly in to the global database.

In other cases, data can also first stored in Microsoft excel sheet (or any other related medium), later then can be transferred into the global database. Archived data can be completely accessed by the Data Administrator. Furthermore data will be online and will be extracted using standard web browsers (e.g. Mozilla or Google Chrome etc.) or accessed using provided web service (Ahmed and Saman, 2011c). The stored data can be in the compatible format to be imported and exported directly by the Isotopo Data Analyzer and Data Manager modules, as desktop client applications. Later after, as earlier discussed, it can also be possible to analyze using Isotopo and store into a local database or in the form of data files, to be privately accessed by user.

The relational schema of global database is mainly divided into two parts: IsoDB Relational Experimental Data Schema and IsoDB Personal Relational Schema. IsoDB Relational Experimental Data Schema is designed to store preprocessed and processed experimental metabolite data.

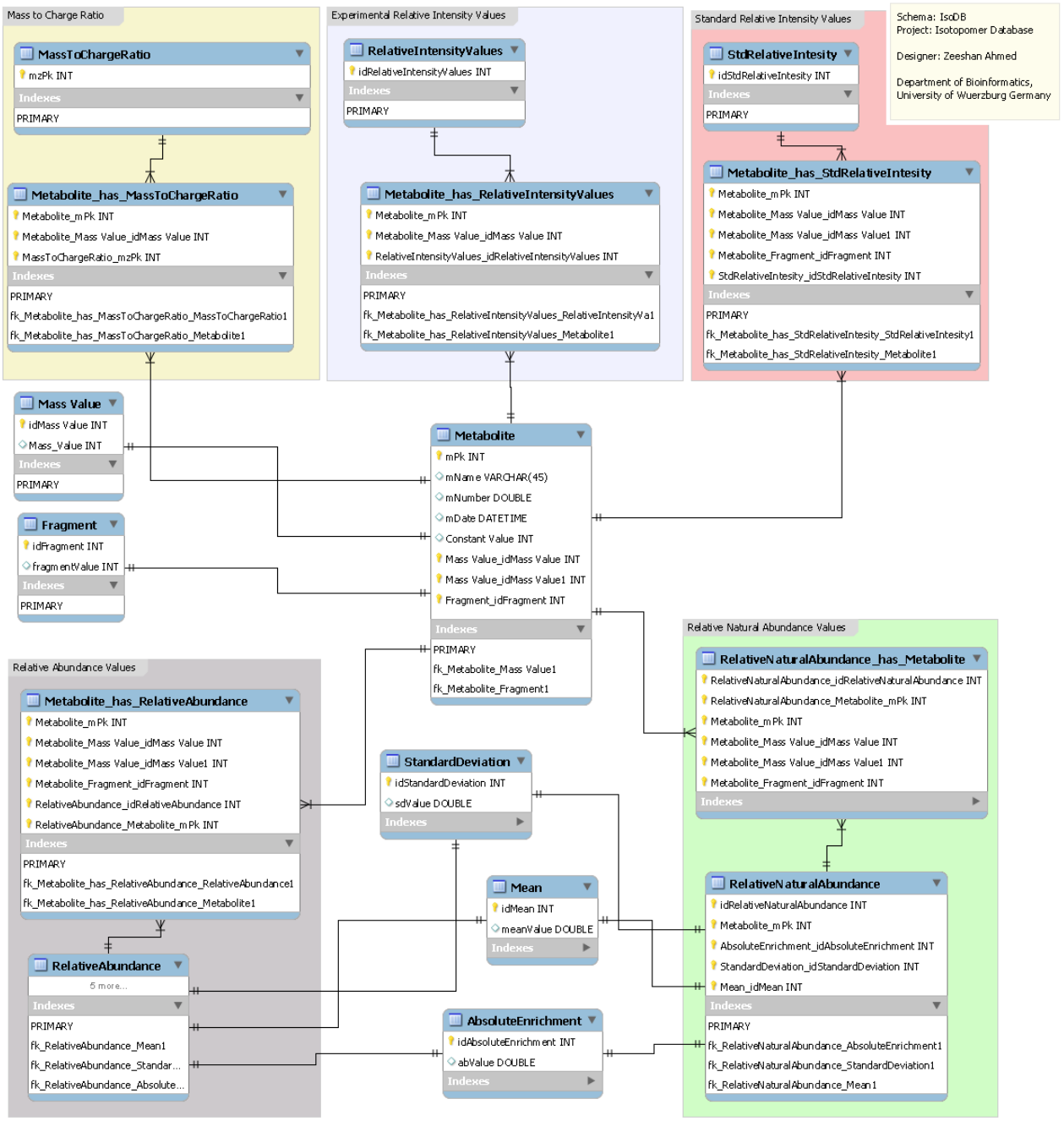


Figure 41: IsoDB; Relational Experimental Data Schema.

Figure Legend 41: This Figure contain following relations: Metabolite, MassToChargeRatio, Metabolite_has_MassToChargeRatio, RelativeIntensityValues,Metabolite_has_RelativeIntensityValues, StdRelativeIntesity,Metabolite_has_StdRelativeIntesity, Mass Value, FragmentMetabolite_ has_RelativeAbundance, RelativeAbundance, RelativeNaturalAbundance_has_Metabolite, RelativeNaturalAbundance, Standard Deviation, Mean, AbsoluteEnrichment.

No.	Relations	Type	Descriptions
1	Metabolite	One to Many	This relation is to maintain metabolite information consisting of name, number, date, time and constant value.
2	MassToChargeRatio	One to Many	This relation is to maintain information of mass to charge ratio values.
3	Metabolite_has_ MassToChargeRatio	Many to Many	This is an intermediate table between Metabolite and Mass to Charge Ratio, to maintain many to many relationships.
4	RelativeIntensityValues	One to Many	This relation is to maintain information about relative intensity values.
5	Metabolite_has_ RelativeIntensityValues	Many to Many	This is an intermediate table between Metabolite and Relative Intensity Values, to maintain many to many relationships.
6	StdRelativeIntesity	One to Many	This relation is to maintain information about standard relative intensity values.
7	Metabolite_has_ StdRelativeIntesity	Many to Many	This is an intermediate table between Metabolite and Standard Relative Intensity Values, to maintain many to many relationships.
8	Mass Value	One to Many	This relation is to maintain information about ion mass values.
9	Fragment	One to Many	This relation is to maintain information about number of fragments.

10	Metabolite_has_ RelativeAbundance	Many to Many	This is an intermediate table between Metabolite and Relative Abundance Values, to maintain many to many relationships.
	RelativeAbundance	One to Many	This relation is to maintain information about relative abundance values.
12	RelativeNatural Abundance_has_ Metabolite	Many to Many	This is an intermediate table between Metabolite and Relative Natural Abundance Values.
13	RelativeNaturalAbundance	One to Many	This relation is to maintain information about relative natural abundance values.
14	StandardDeviation	One to Many	This relation is to maintain information about estimated standard deviation values.
15	Mean	One to Many	This relation is to maintain information about estimated mean values.
16	AbsoluteEnrichment	One to Many	This relation is to maintain information about estimated absolute enrichment values.

Table 10: IsoDB: Experimental Data Schema Description.

However IsoDB Personal Relational Schema is to maintain user data for security purposed. Sixteen relations are created in IsoDB Relational Experimental Data Schema (Figure 41, Table 10) where as four relations are created in IsoDB Personal Relational Schema (Figure 42, Table 11). Both schemas are created using MySQL Workbench and database is created in Relational Database Management System (RDBMS) Oracle database.

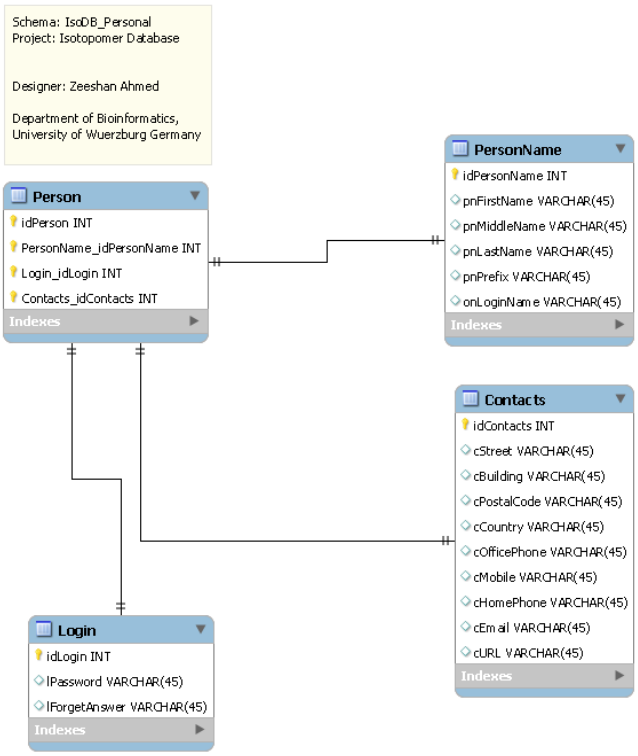


Figure 42: IsoDB_Personal; Relational Schema Personal.

Figure Legend 42: This Figure contain following relations Person, PersonName, Login, Contacts.

No.	Relations	Type	Descriptions
1	Person	One to Many	This relation is to maintain information about user.
2	PersonName	One to Many	This relation is to maintain information about user name.
3	Login	One to Many	This relation is to maintain information about user contacts.
4	Contacts	One to Many	This relation is to maintain information about user login.

Table 11: IsoDB: Personal Schema Description.

4.6 Isotopo Implementation

Isotopo is a MDI software application, developed following the principle of embedding child windows under a single parent window by creating nesting hierarchies. To meet aforementioned goals of Isotopo development, the graphical user interface of this application is divided into two main modules i.e. Data analyzer and Data manager.

Isotopo Data Analyzer is the module responsible for providing options for experimental data loading, analysis and visualization, whereas Data Manager is the module responsible for providing options for experimental data manipulation and management.

Isotopo Main is the parent window of the application embedding all the module interfaces and responsible for providing options for child widows manipulation and closing over all application. Finally, Isotopo About is a child interface providing basic information about the current version of in use Isotopo version.

The graphical user interface of Isotopo data analyzer consists of 10 main controls: open data file, clear all text controls, measure selected data, process all data, remove selected data, open data manager, close Isotopo, selected values and results. Moreover the graphical interface is divided into seven views: Isotopo Analyzer, Fragment Viewer, Spectrum Viewer, Result Viewer, Relative Abundance 1, Relative Abundance 2 and Relative Abundance 3 as shown in Figure 43 (a, b, c, d, e, f, g) and described in Table 12.

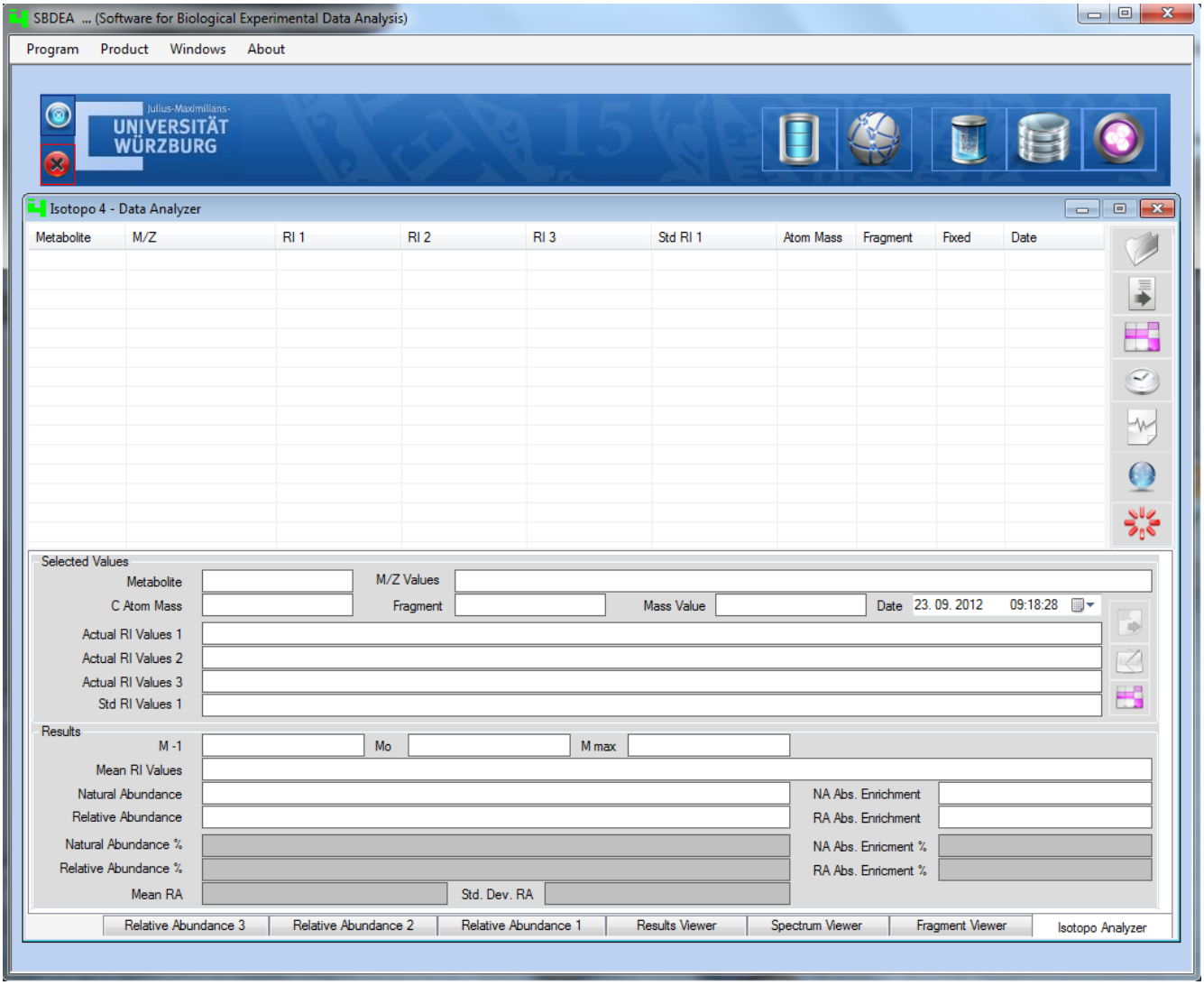


Figure 43 (a): Isotopo; Data Analyzer

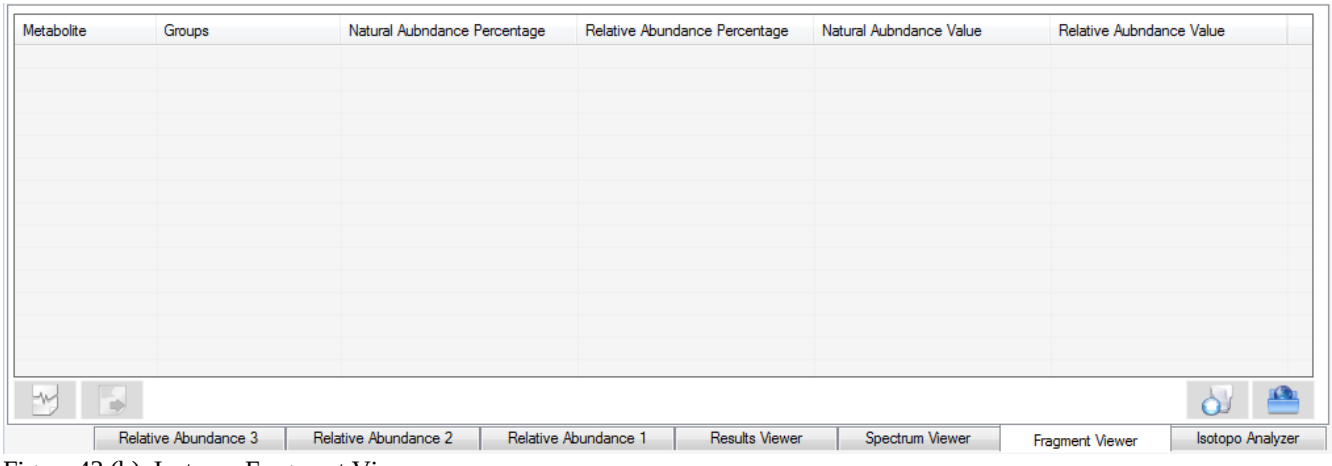


Figure 43 (b): Isotopo; Fragment Viewer

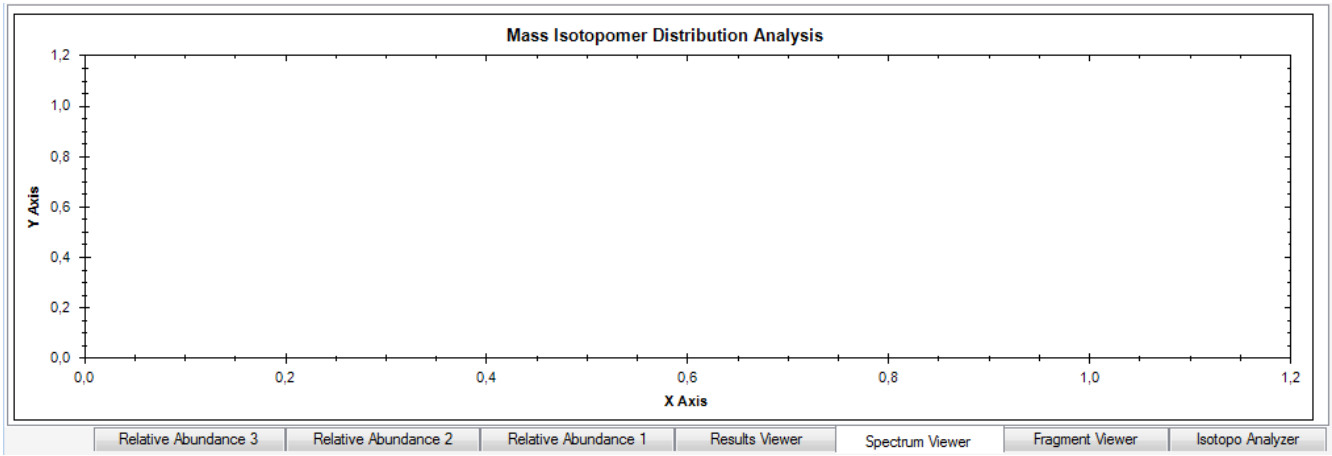


Figure 43 (c): Isotopo; Spectrum Viewer

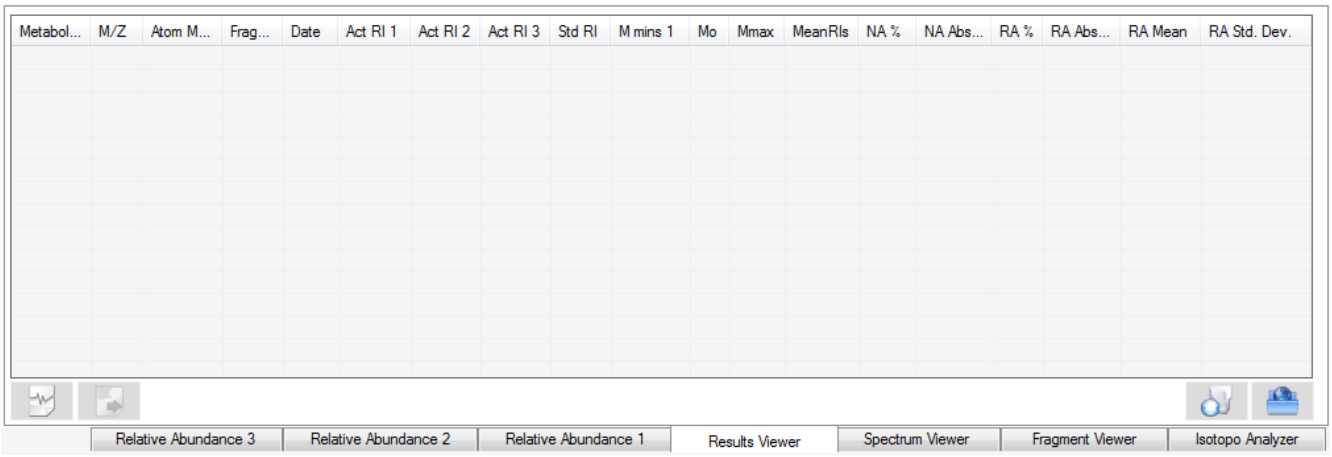


Figure 43 (d): Isotopo; Result Viewer

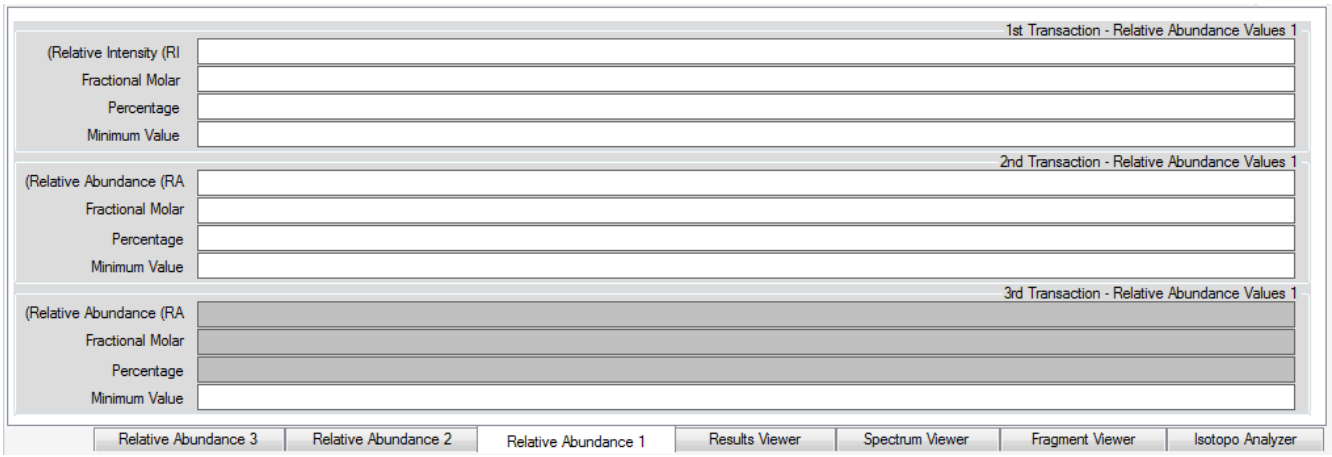


Figure 43 (e): Isotopo; Relative Abundance 1

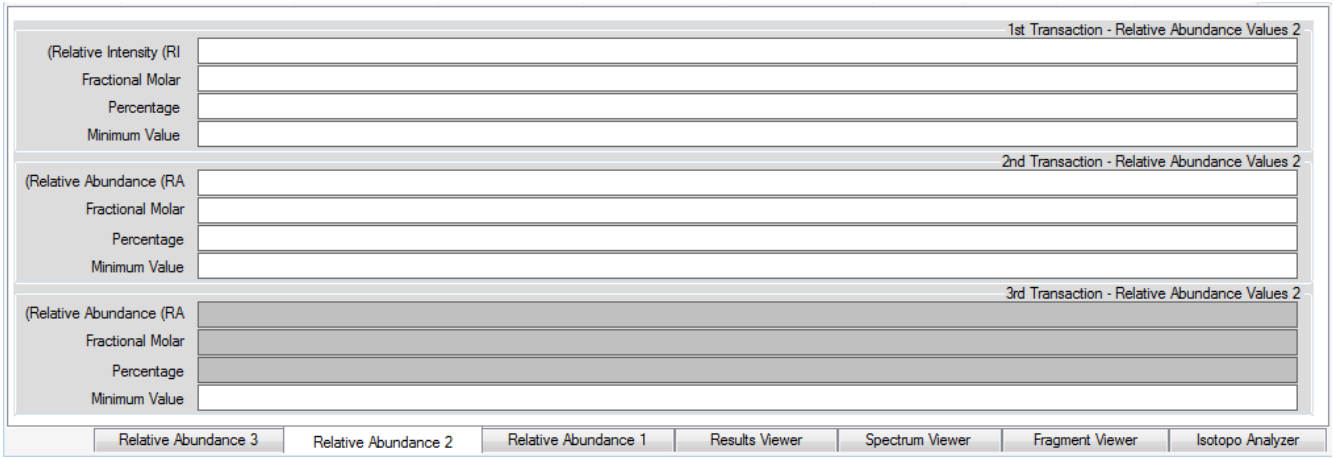


Figure 43 (r): Isotopo; Relative Abundance 2

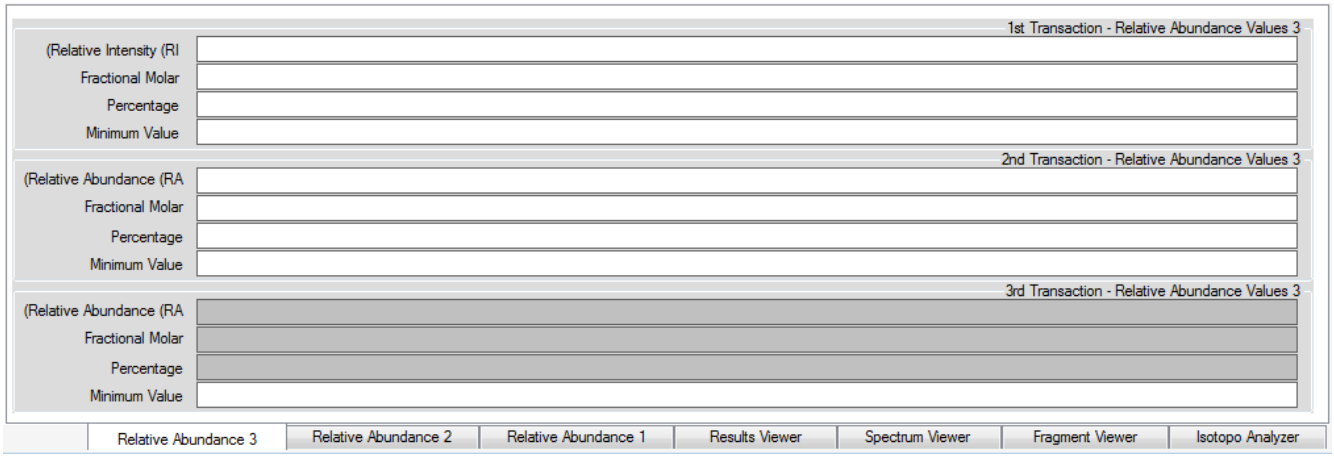


Figure 43 (g): Isotopo; Relative Abundance 3

Figure 43: Isotopo; GUI Data Analyzer.

Figure Legend 43: The Isotopo GUI of Data Analyzer (a) presents the main graphical user interface responsible for handling user data input, analyzing and producing spectrum, along with (b): Isotopo; Fragment Viewer, (c): Isotopo; Spectrum Viewer, (d): Isotopo; Result Viewer, (e): Isotopo; Relative Abundance 1, (r): Isotopo; Relative Abundance 2, (g): Isotopo; Relative Abundance 3.

No.	Features	Descriptions
1	open data file	Opens directory browser to select input data file from attached repositories and loads data from data file into data viewer.
2	clear all text controls	Deletes all loaded data and clears all text controls.

3	measure selected data	Process selected data entry (only one at a time) from data view and perform MIDA.
4	process all data	Processes all loaded data (all data entries) at once.
5	remove selected data	Deletes selected data entry from data view.
6	connect to database	Connects to the database server and retrieves experimental isotopomers data.
7	close Isotopo	Closes the Isotopo Data Analyzer.
8	selected values	Provides text boxes for experimental data manipulation by editing selected input data entry values from data viewer or by entering new experimental data for measurement analysis. It provides following text boxes • Metabolite; name of the metabolite. • M/Z Values; mass to charge ratio values. • C Atom Mass: atom number. • C Atom Fragment; number of fragments • Date • R_{I1} Values; relative intensity values. • R_{I2} Values; relative intensity values. • R_{I3} Values; relative intensity values. • Mass Value; set mass value.
9	results	Provides text boxes presenting measured results: • M_0; mass value

		• M_{-1}; mass value minus 1 • Mmax; maximum mass values • Mean Ri Values • Natural Abundance Values • Relative Abundance Values • Percentages of Natural Abundance Values • Percentages of Relative Abundance Values • Absolute Enrichment value of Natural Abundance • Absolute Enrichment value of Relative Abundance • Percentage of Absolute Enrichment value of Natural Abundance • Percentage of Absolute Enrichment value of Relative Abundance • Mean Relative Abundance (calculated from three estimated Ra) • Standard Deviation of Relative Abundance (calculated from three estimated RAs)
10	Isotopo Analyzer	It is the default view (containing options 1,2,3,4,5,6,7,8,9)
11	Fragment Viewer	Provides the information about measured abundances (natural and relative) with respect to the number of fragments. It consists of four controls as well • Export File; allows user to export measure fragment based output in a new file.

		• Import File; allows user to import already estimated data. • Clear Text; allows user to clear the view by deleting all the data. • Delete Selected Data; allows user to delete particular (selected) data.
12	Spectrum Viewer	Draws a spectrum (visualization) of estimated natural and relative abundance values with respect to the number of fragments.
13	Result Viewer	Provides the information of complete output including Metabolite name, input values (e.g. m/z values, Fragment Number etc.) and measured values (e.g. Na, Ra and Ra per m/z etc.). It consists of four controls as well • Export File; allows user to export measure fragment based output in a new file. • Import File; allows user to import already estimated data. • Clear Text; allows user to clear the view by deleting all the data. • Delete Selected Data; allows user to delete particular (selected) data
14	Relative Abundance 1	Provides measured Relative Intensity, Fractional Molar Abundance, Minimum Values using inputted actual Ri 1 values.
15	Relative Abundance 2	Provides measured Relative Intensity, Fractional Molar Abundance, Minimum Values using inputted actual Ri_2 values.
16	Relative Abundance 3	Provides measured Relative Intensity, Fractional Molar Abundance, Minimum Values using inputted actual Ri_2 values.

Table 12: Isotopo Analyzer; Control Descriptions.

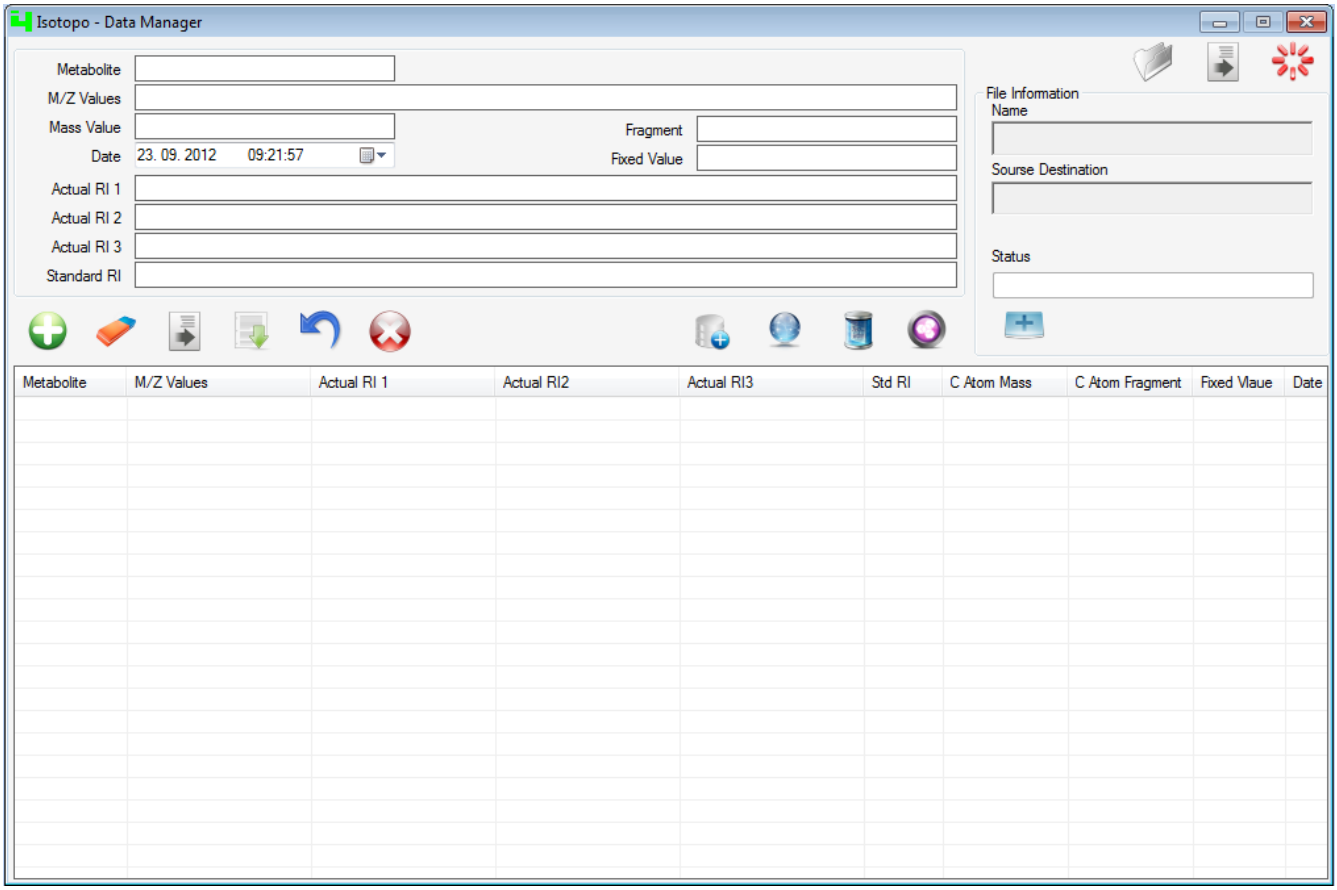


Figure 44: Isotopo; GUI of Isotopo Data Manager.

Figure Legend 44: The Isotopo; GUI of Data Manager presents the main graphical user interface responsible for handling user data input and providing data management and manipulation options.

No.	Features	Descriptions
1	open data file	Opens directory browser to select input file from attached repositories and loads data from data file into data viewer.
2	clear all text controls	Deletes all loaded in data and clears all text controls.
3	close isotopo data manager	Closes the Isotopo Data Manager.
4	add new values	Add newly entered values in text boxes to data view.

5	update edited values	Updates edited values in to data view
6	clear text fields	Deletes data a from text controls.
7	save data in file	Saves data into file.
8	select values to edit	Allows user to select one value from data view to edit existing values.
9	delete values	Deletes selected data entry from data view.
10	create new data file	Allows user to create new data (input) file.
11	select source directory	Opens directory browser to select the directory to store newly created data file. This option is only enabled and visible when user will click option 10.
12	save file	Allows user to save newly created file. It also allows user to save exiting file data e.g. if some data file is also open and user want to creates a new file, then system will ask if he want to merge existing data (in data view) to newly created file or not. This option is only enabled and visible when user will click option 10.
13	cancel creating file	Allows user to cancel new file creation process. This option is only enabled and visible when user will click option 10.
14	data view	Provides the textual view of all loaded, added or updated data.
15	Open Isotopo Data Analyzer	Opens the graphical user interface of Isotopo Data Analyzer.
16	Open Isotopo Data Viewer	Opens the graphical user interface of Isotopo Data Viewer.

17	Get data from database	Connects to the database server and retrieves experimental isotopomers data.
18	Insert into database	Connects to the database server and insert newly entered/edited experimental isotopomers data in to the database.

Table 13: Isotopo Data Manager; Control Descriptions.

The graphical user interface of Isotopo Data Manager consists of 16 main controls: open data file, clear all text controls, close isotopo data manager, add new values, update edited values, clear text fields, save data in file, select values to edit, delete values, create new data file, select source directory, save file, cancel creating file, data view, Open Isotopo Data Analyzer and Open Isotopo Data Viewer as shown in Figure 44 and described in Table 13.

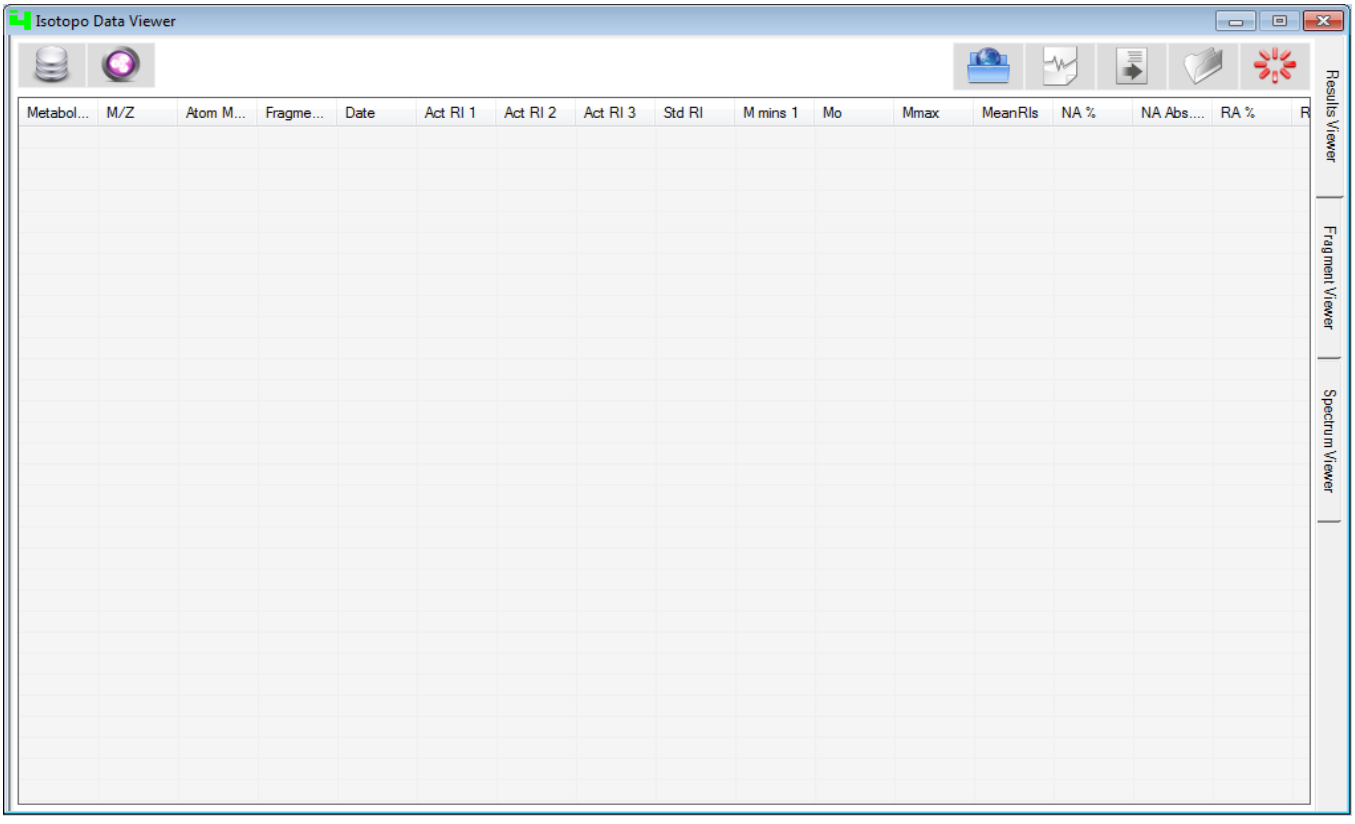


Figure 45 (a): Isotopo; GUI of Isotopo Data Viewer (results)

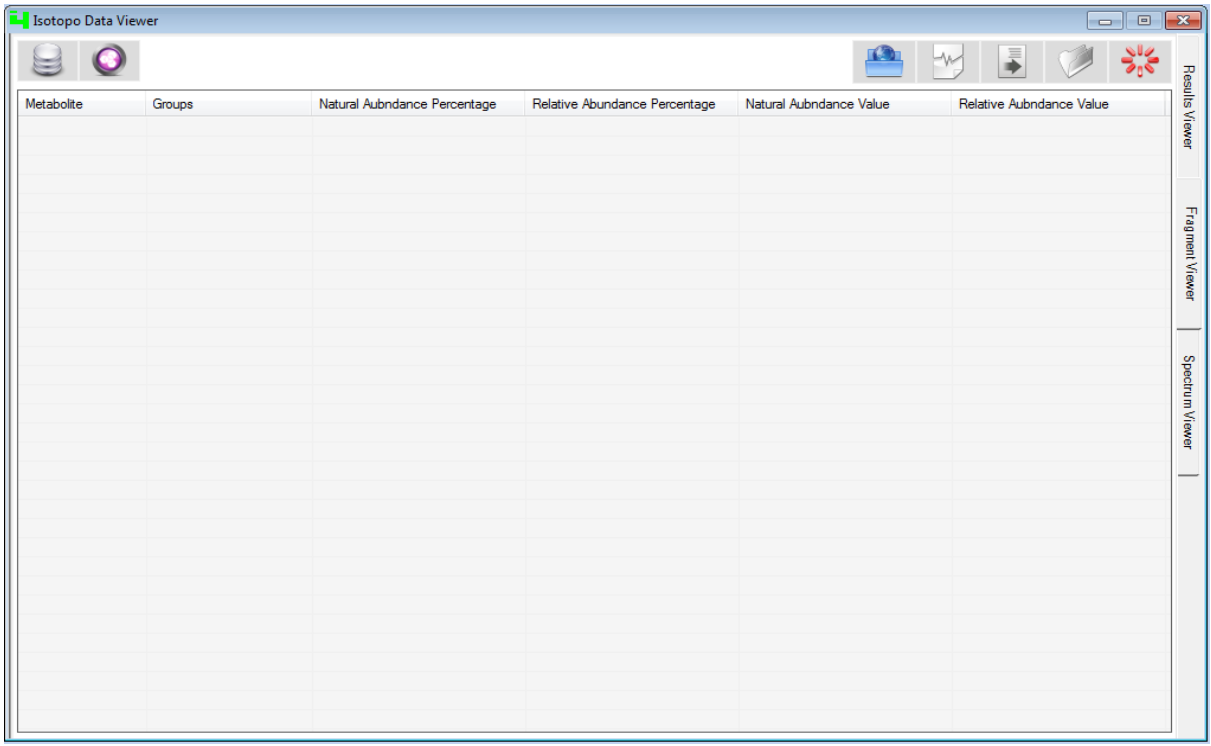


Figure 45 (b): Isotopo; GUI of Isotopo Data Viewer (fragments)

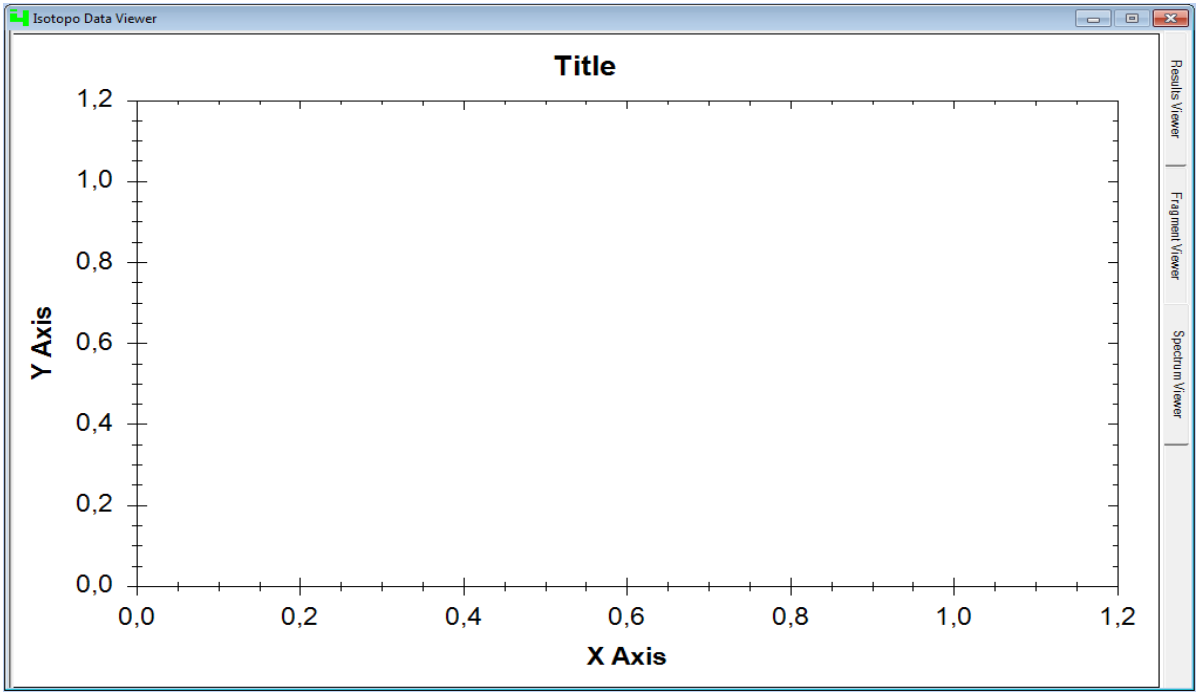


Figure 45 (c): Isotopo; GUI of Isotopo Data Viewer (spectrum)

Figure 45: Isotopo; GUI of Data Viewer.

Figure Legend 45: The Isotopo; GUI of Data Viewer presents the main graphical user interfaces for 41 (a): Isotopo; GUI of Isotopo Data Viewer (results) and (b): Isotopo; GUI of Isotopo Data Viewer (fragments).

The graphical user interface of Isotopo Data Viewer consists of seven main controls: Export Data, Remove Selected Values, Clear Values, Import Data, Close Data Viewer, Open Isotopo Data Analyzer and Open Isotopo Data Manager (Table 14). Moreover the graphical interface is divided into three (tabs) views: Result Viewer (Figure 45a), Fragment Viewer (Figure 45b) and Spectrum Viewer (Figure 45c).

No.	Features	Descriptions
1	Export Data	Allows user to export data (into file).
2	Remove Selected Values	Removes selected data from list view.
3	Clear Values	Deletes all contained data from all controls.
4	Import Data	Allows user to import data in to data viewer.
5	Close Data Viewer	Exits isotopo data viewer
6	Open Isotopo Data Analyzer	Opens the graphical user interface of Isotopo Data Analyzer.
7	Open Isotopo Data Manager	Opens the graphical user interface of Isotopo Data Manager.
8	Result Viewer	Default view, allows user to view, import and export resultant data (obtained from Isotopo Data Analyzer).
9	Fragment Viewer	Allows user to view, import and export resultant fragment based data obtained from Isotopo Data Analyzer (natural and relative abundance values).
10	Spectrum Viewer	Draws a spectrum (visualization) of estimated natural and relative abundance values with respect to the number of fragments.

Table 14: Isotopo Data Viewer; Control Descriptions.

4.7 Isotopo Modelling Experimentation

The above mathematics was implemented in the software "Isotopo". The available and tested version of Isotopo provides two main modules i.e. Isotopo Analyzer (Figure) and Isotopo Data Manager (Figure) with one additional module i.e. Isotopo Output Viewer (Figure). Isotopo Analyzer is capable of processing experimental data (metabolite information, mass to charge ratio (m/z) values, actual relative intensity values and standard relative intensity values and number of carbon atom fragments).

It then estimates (definitions see above) mass calculations (M_o, M_{-1}, M_{max}), predicts natural abundance values, relative abundance values (Lee et al., 1991) and cculates fractional molar abundance values, the percentage of relative abundance per m/z value and the minimal value. It also draws the spectrum of the calculated relative abundance values.

Isotopo Analyzer repeats this whole procedure twice to get new relative abundance values (up to the threshold level of two) because according to the first rule of MIDA, during combinatorial polymer analysis at least two repeats of a probabilistically identical subunit must be present (Hellerstein et al., 1999). The Isotopo Data Manager is a supporting utility, developed as a user friendly file based experimental data management system. It allows the user to create new experimental data files that later can be used for the analysis using Isotopo Analyzer. It allows user to perform data manipulation by reading, adding, editing, updating, deleting and merging data (from other source files of the same extension) into a file.

To apply the Isotopo, some experimental data have to be collected first. The process consists of three major steps i.e. preparation of data set, input data file preparation and management, and evaluation and data analysis. Observed data during actual experimentation are collected during the preparation of data set. During input data file preparation and management at first Data Manager is used to structure data by organizing an experimental data file (Figure 46) which is later used by Data Analyzer for analysis (Figure 47).

Throughout the experimental data analysis, each observed resultant data during Evaluation is individually analyzed using Data Analyzer and results are discussed.

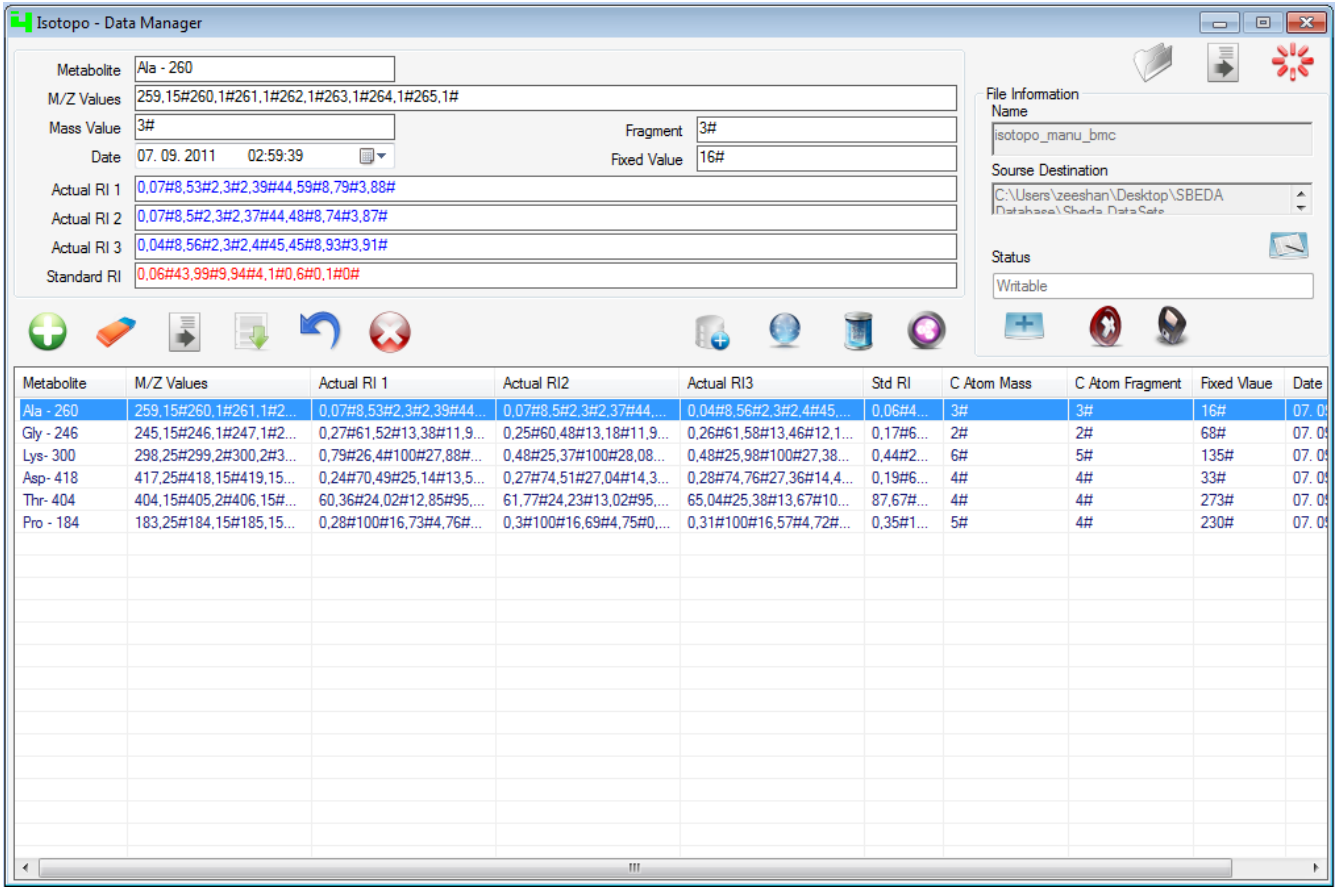


Figure 46: Isotopo; Data Manager- Input Creation.

Figure Legend 46: Data Manger; Inputted data management presents the inputted user data consisting of six metabolites: Alanine (Ala – 260), Glycine (Gly – 246), Lysine (Lys- 300), Aspartic Acid (Asp- 418), Threonine- 404, Proline (Pro-184). A new data file "Isotopo_manu_dataset.ls" is created for file based data management and further data analysis using Isotopo Data Analyzer.

Observed resultant data during different experiments of metabolic isotopomers analysis with different metabolites are collected for mass isotopomers predictions using Isotopo. The collected data consists of five different results observed using seven different metabolites during Evaluations i.e. Alanine (Ala – 260), Glycine (Gly – 246), Lysine (Lys- 300), Aspartic Acid (Asp- 418), Threonine (Thr- 404) and Proline (Pro – 184), presented in Table 15. The metabolite based data set example presented in table consists of following experimental elements i.e. m/z values, relative intensity values (Ri), Atomic Mass Values and Atomic Fragment Numbers.

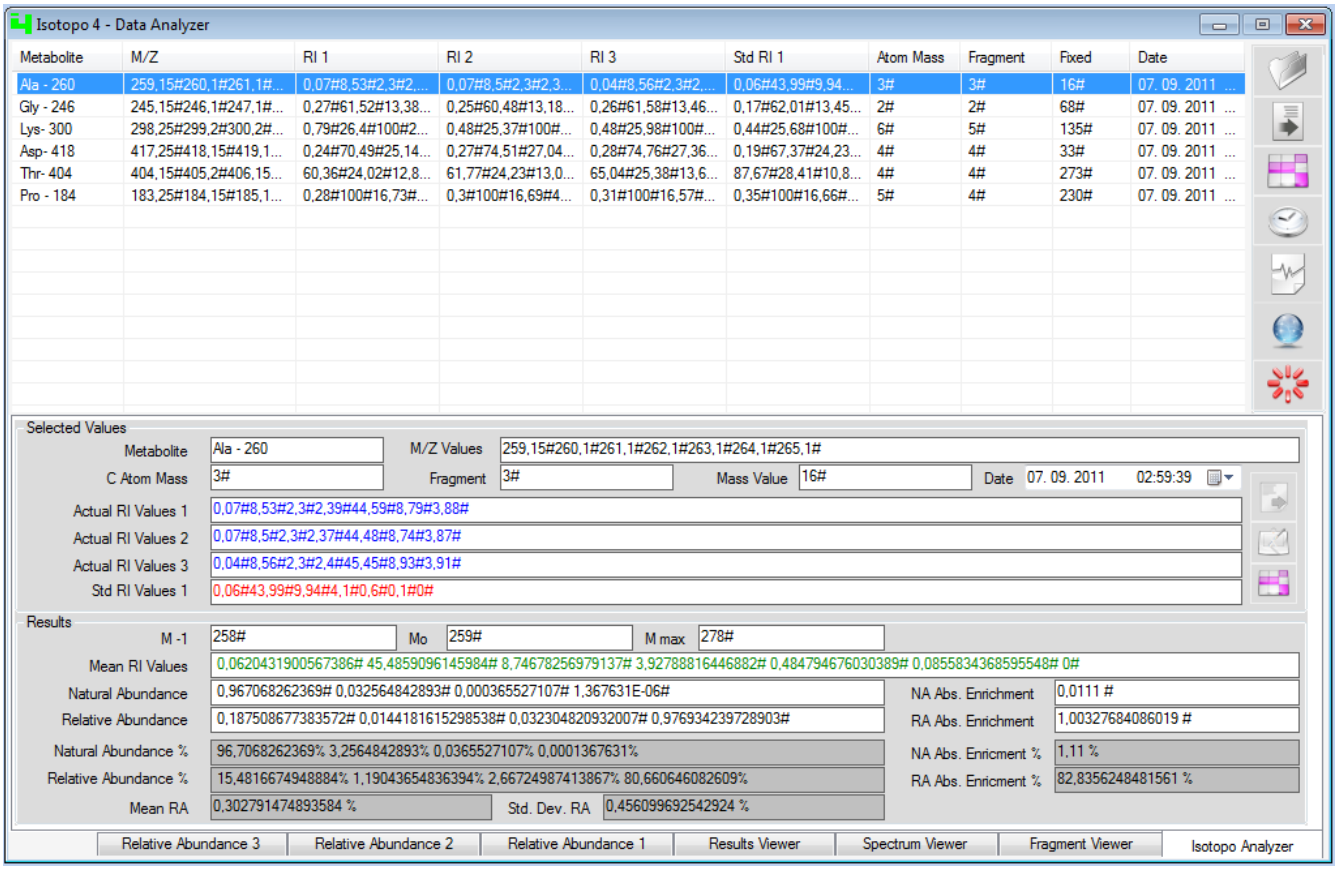


Figure 47: Isotopo; Analyzer- Inputted Data Analysis.

Figure Legend 47: This figure shows the Isotopo Analyzer loaded with the experimental data using the already created data file using the Isotopo Data Manager (earlier discussed in figure 46). This figure also shows a successful data processing during experimentation using alanine. Resultant data are provided and visible in respective fields.

No.	Metabolite	M/Z	RI1	RI 2	RI 3	Std. RI	Mass Values	Fragments
1	Ala - 260	259,15#260,1#261,1#262,1#263,1#264,1#265,1#	0,07#8,53#2,3#2,39#44,59#8,79#3,88#	0,07#8,5#2,3#2,37#44,48#8,74#3,87#	0,04#8,56#2,3#2,4#45,45#8,93#3,91#	0,06#43,99#9,94#4,1#0,6#0,1#0#	3#	3#

2	Gly - 246	245,15#246,1#247,1#248,1#249,05#250,1#	0,27#61,52#13,38#11,9#1,91#0,72#	0,25#60,48#13,18#11,92#1,94#0,73#	0,26#61,58#13,46#12,15#2,03#0,73#	0,17#62,01#13,45#5,67#0,77#0,14#	2#	2#
3	Lys-300	298,25#299,2#300,2#301,2#302,15#303,15#304,15#305,15#306,15#	0,79#26,4#100#27,88#13,3#4,59#1,56#0,53#0,16#	0,48#25,37#100#28,08#13,61#5,55#1,61#1,23#0,1#	0,48#25,98#100#27,38#13,3#5,52#1,6#1,21#3#	0,44#25,68#100#26,94#12,23#3,45#0,89#0,23#0,11#	6#	5#
4	Asp-418	417,25#418,15#419,15#420,15#421,15#422,15#423,1#424,1#	0,24#70,49#25,14#13,5#52,81#18#7,44#1,66#	0,27#74,51#27,04#14,32#53,75#17,92#7,62#1,72#	0,28#74,76#27,36#14,47#56,25#19,02#8,01#1,81#	0,19#67,37#24,23#11,36#2,71#0,68#0,11#0,01#	4#	4#
5	Thr-404	404,15#405,2#406,15#407,2#408,2#409,2#410,2#	60,36#24,02#12,85#95,21#31,94#14,67#3,21#	61,77#24,23#13,02#95,41#32,19#14,66#3,38#	65,04#25,38#13,67#100#33,72#15,5#3,45#	87,67#28,41#10,88#2,67#0,41#0,13#0,13#	4#	4#
6	Pro - 184	183,25#184,15#185,15#186,15#187,15#188,15#189,1	0,28#100#16,73#4,76#0,53#0,09#0,65#0,14#0,07#0,01#	0,3#100#16,69#4,75#0,53#0,09#0,64#0,15#0,07#0	0,31#100#16,57#4,72#0,53#0,09#0,64#0,15#0,07#	0,35#100#16,66#4,71#0,51#0#0,59#0,13#0,07#	5#	4#

		#190,1#19 1,1#192,1#		,01#	0#	0#		

Table 15: Isotopo processing experimental data set

Note: Symbol "# ", is the number separator. The implemented Isotopo is designed for Germany and English systems. Symbol "," (comma) is used instead of "." (dot), so to differentiate (especially in German Systems) values we have standardized "#" (hash) symbol. Table 11 provides information obtained Relative Abundances (1, 2 and 3) using three Relative Intensity Values (Ri_1, Ri_2 and Ri_3).

A data input file for Isotopo Data Analyzer is created using Isotopo Data Manager. During experimental data analysis, the metabolites are taken as input and analyzed using Data Analyzer item by item. The resultant information from Data analyzer consists of relative intensity per m/z values, natural abundances, relative abundances per fragments, and absolute C13 enrichment, presented in Table 16 and Figure 48 (a, b, c, d, e, f, g).

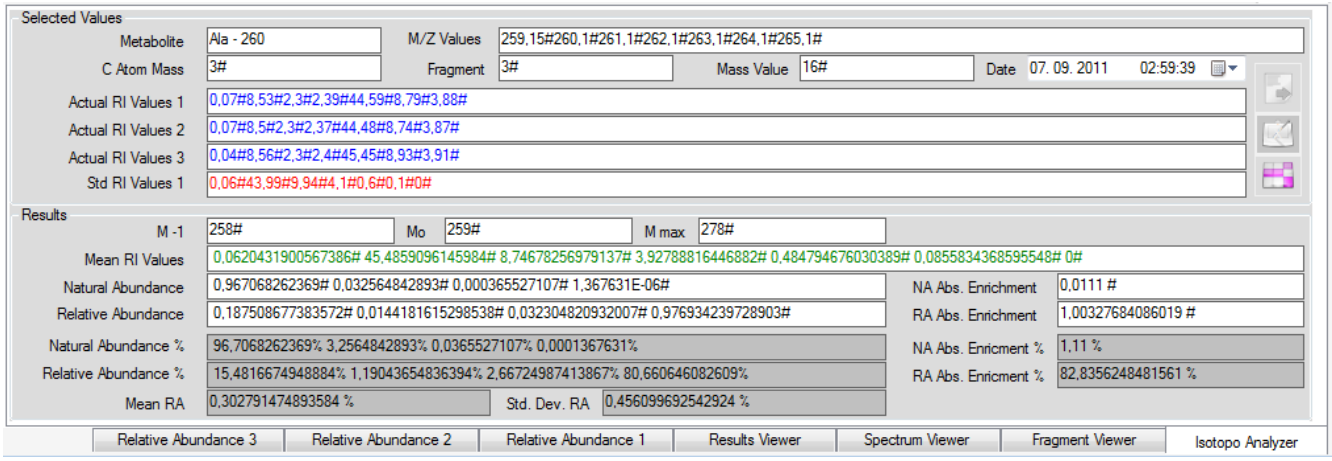


Figure 44 (a) – Isotopo Analyzer

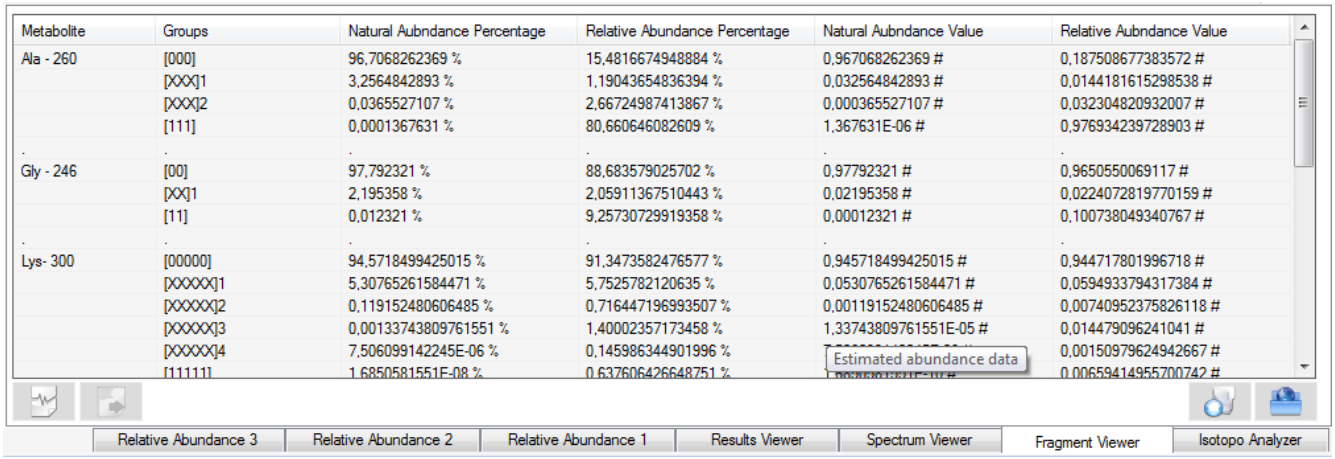


Figure 44 (b) – Fragment Viewer

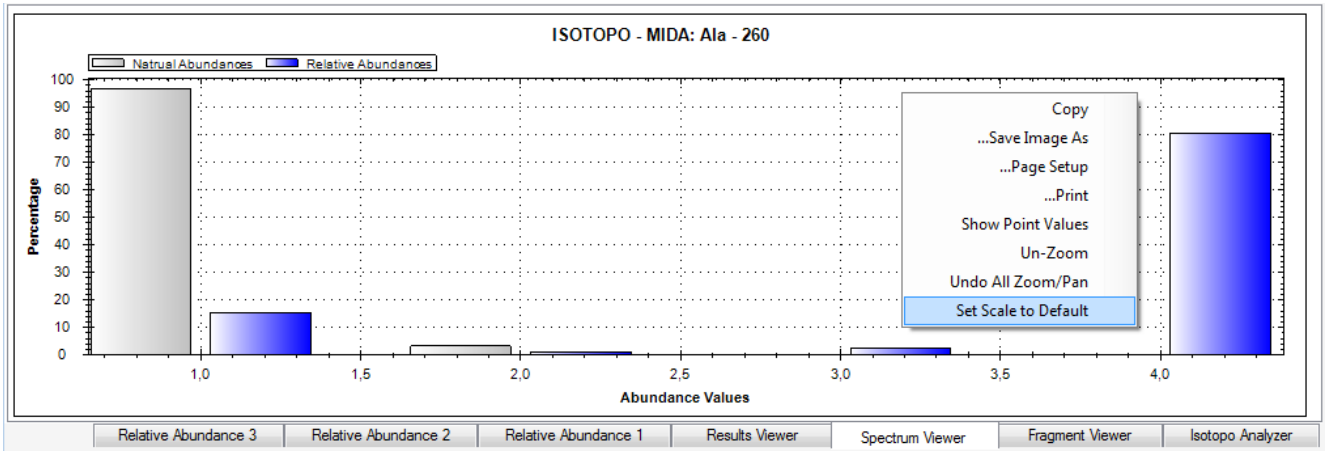


Figure 44 (c) – Spectrum Viewer

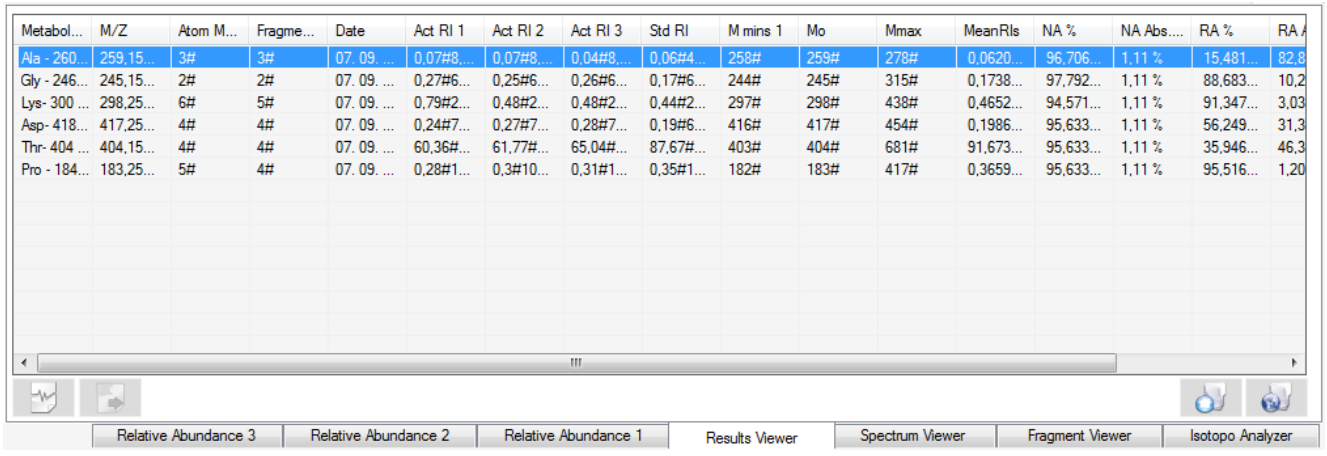


Figure 44 (d) –Result Viewer

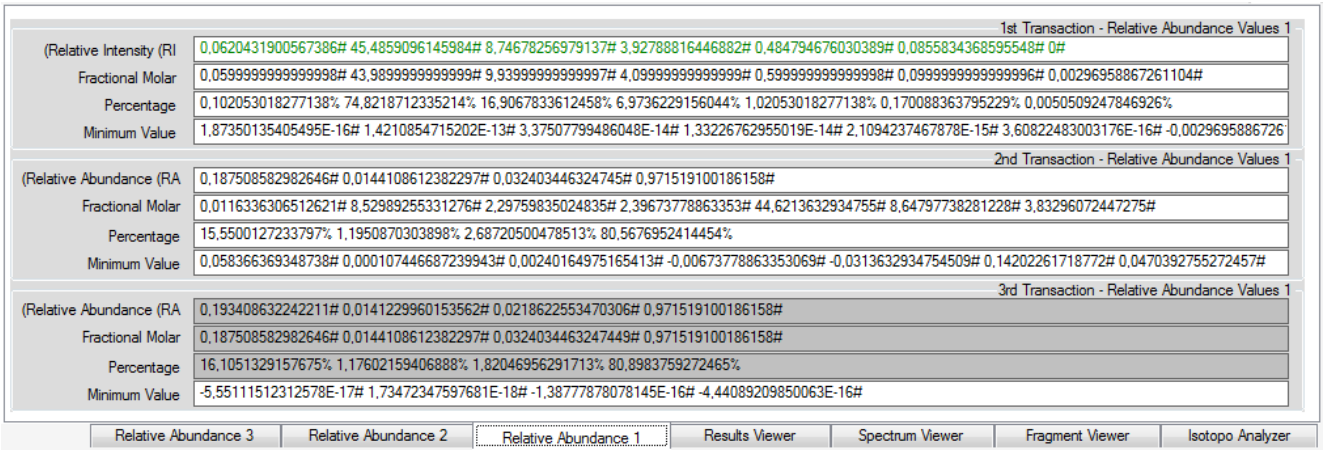


Figure 44 (e) – Relative Abundance 1

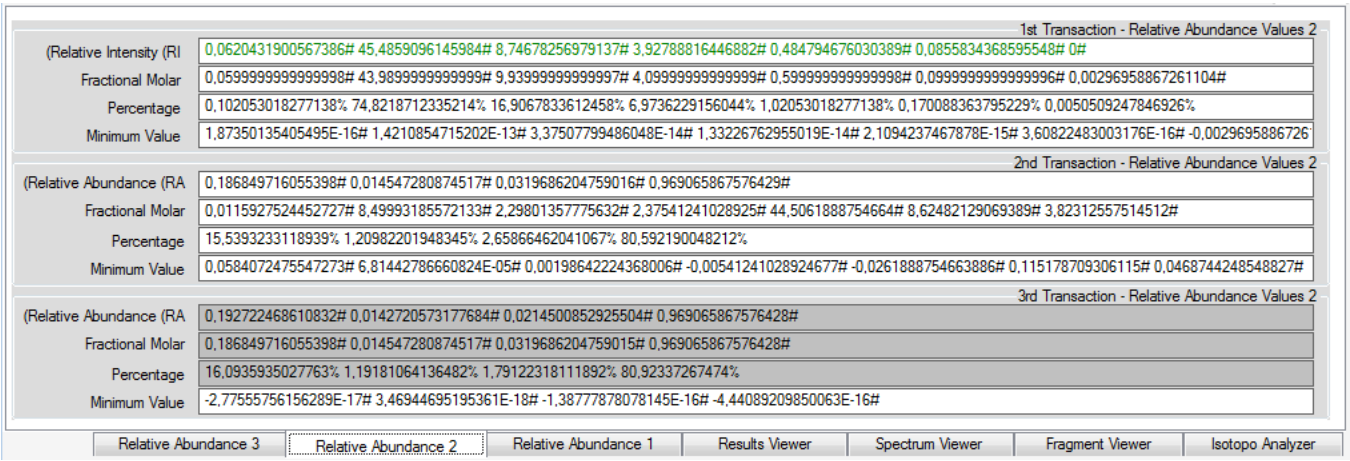


Figure 44 (f) – Relative Abundance 2

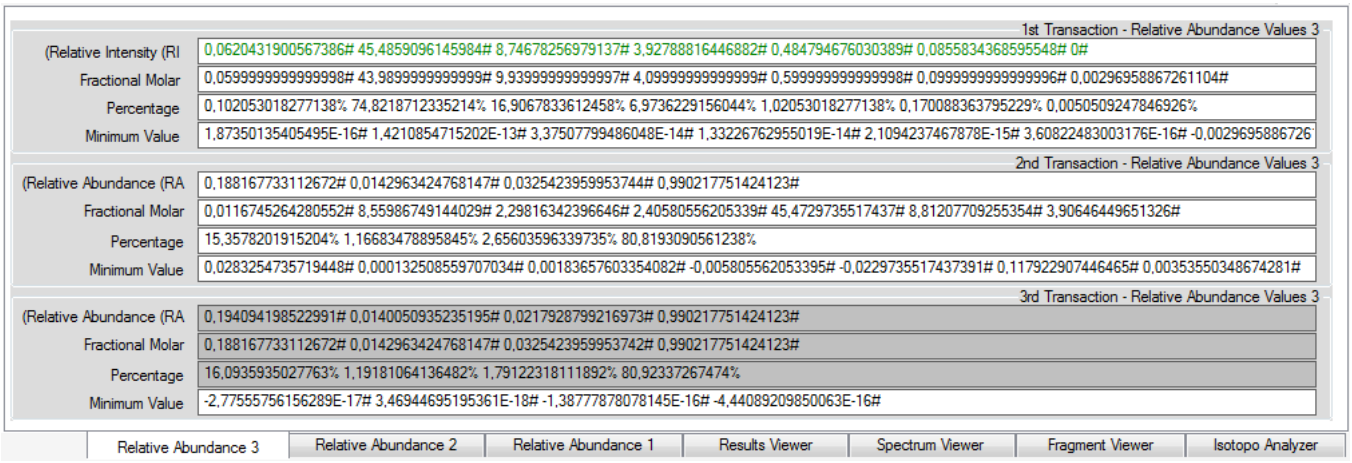


Figure 44 (g) – Relative Abundance 3

Figure 48: Isotopo; Analyzer Data Analysis – Ala 260.

Figure Legend 48: Isotopo Data Analyzer; Analyzing Alanine (a) presents the obtained results after metabolite Ala-260 analysis.

Metabolite	M_o	M_{-1}	M_{max}	Na	Ra	Na Abs. Enrich.	Ra Abs. Enrich	Ri Values
Ala – 260	259#	258#	278#	96,7068262369% 3,2564842893% 0,0365527107% 0,0001367631%	15,481667494888 4% 1,1904365483639 4% 2,6672498741386 7% 80,660646082609 %	1,11 %	82,835624848156 1 %	0,0620431900567386# 45,4859096145984# 8,74678256979137# 3,92788816446882# 0,484794676030389# 0,0855834368595548# 0#
Gly - 246	244	245	315	97,792321 % 2,195358 % 0,012321 %	88,683579025702 % 2,059113675104 43% 9,257307299193 58%	1,11 %	10,286864136745 8 %	0,17383778016680 9# 63,4059842371328# 12,3302015032638# 5,51320965535131# 0,662062317395308# 0,127603147274152#
Lys-300	297	298	438	94,5718499425015% 5,3076526	91,347358247657 7%	1,11 %	3,0315041668212 %	0,4652547246009 39# 27,12784608848

				1584471% 0,1191524 80606485 % 0,0013374 38097615 51% 7,5060991 42245E- 06% 1,6850581 551E-08%	5,752578 2120635 % 0,716447 1969935 07% 1,400023 5717345 8% 0,145986 3449019 96% 0,637606 4266487 51%			97# 104,2166288011 06# 22,60314635300 65# 11,53172319931 34# 2,970871474835 36# 0,759492384229 862# 0,196668429158 787# 0,104276236597 961#
Asp- 418	416	417	454	95,633380 4656704% 4,2937830 8491836% 0,0722939 5122246% 0,0005409 8011836% 1,5180704 1E-06%	56,24922 7851349 3% 2,650887 2701982 9% 1,250301 8539110 4% 38,98875 2440666 3% 0,860830 5838750 73%	1,11 %	31,39026 7658879 9 %	0,198675398772 716# 70,43719356288 38# 22,17367614085 53# 10,82988752355 77# 2,330333644699 66# 0,598107216629 968# 0,086345331182 2733# 0,006114342460

								89636#
Thr-404	403	404	681	95,6333804656704% 4,2937830849183 6% 0,07229395122246% 0,00054098011836% 1,51807041E-06%	35,946801126187 9% 4,1003783828499 1% 2,4423420754364 1% 53,814091552794 6% 3,6963868627311 8%	1,11 %	46,303221160757 8 %	91,67301163370 53# 25,59123144436 16# 10,15847598756 84# 2,315948908666 77# 0,316912752867 383# 0,119898322596 564# 0,130299727680 376#
Pro - 184	182	183	417	95,6333804656704% 4,2937830849183 6% 0,07229395122246% 0,00054098011836% 1,51807041E-06%	95,516794410392 3% 4,2916572270011 7% 0,1033890925752 67% 0,0165325811460 895% 0,0716266888851	1,11 %	1,208634977782 59 %	0,365980997739 213# 104,549567403 106# 12,72631346722 6# 4,274631610802 34# 0,331150730666 402# 0# 0,617480626620 422# 0,108223721212 432#

					002%			0,067870444390 6936# 0#

Table 16: Isotopo Analyzer results.

Figure 48 (a) shows the Isotopo Analyzer loaded with the Alanine 260 experimental data using already created data file using Isotopo Data Manager and processed outputs. Figure 48(b) shows the fragment Viewer, presenting natural and relative Abundance at each fragment, which can be maintained in a different file as well. Figure 48(c) shows the Spectrum Viewer, drawing a bar chart. The grey colored bar represents the percentage of natural abundance and blue colored bar represents percentage of relative abundance at each fragment. Figure 48(d) shows the Result Viewer, it presents the final output of each transaction, and this information can be maintained in the form of an output file as well. Figure 48(e), (f), (g) presents output information (relative abundance 1, 2 and 3) obtained after successful experimental data processing with relative intensity 1, 2 and 3, respectively.

The data entry of each metabolite (Ala, Gly, Lys, Asp, Thr, Pro) is selected and processed for analysis (one by one or at once all). Figure 49a,b,c,d,e,f presents measured natural and relative abundance values at each fragment, given in table 16.

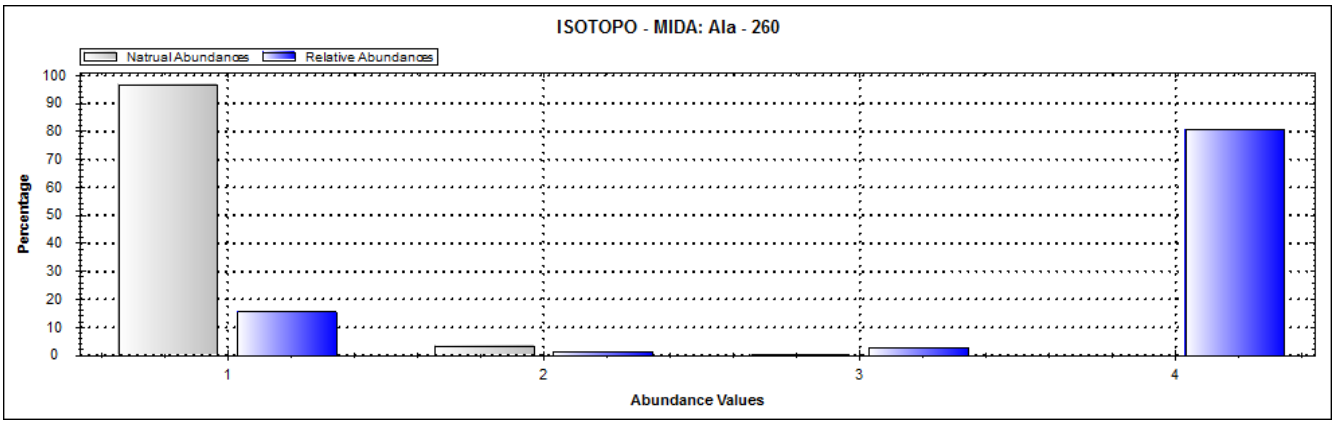


Figure 49 (a): Ala 260 Spectrum Analysis

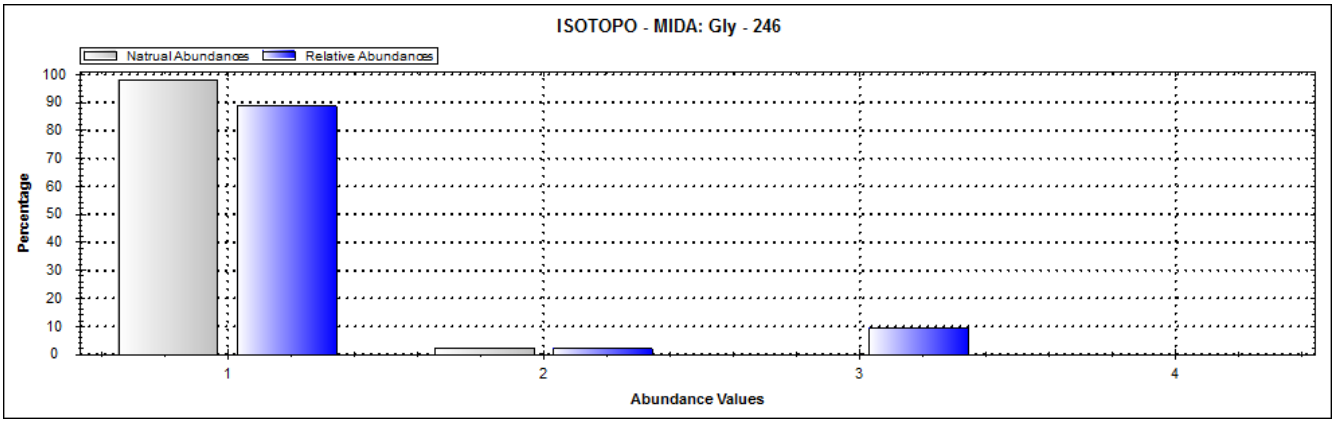


Figure 49 (b): Gly 246 Spectrum Analysis

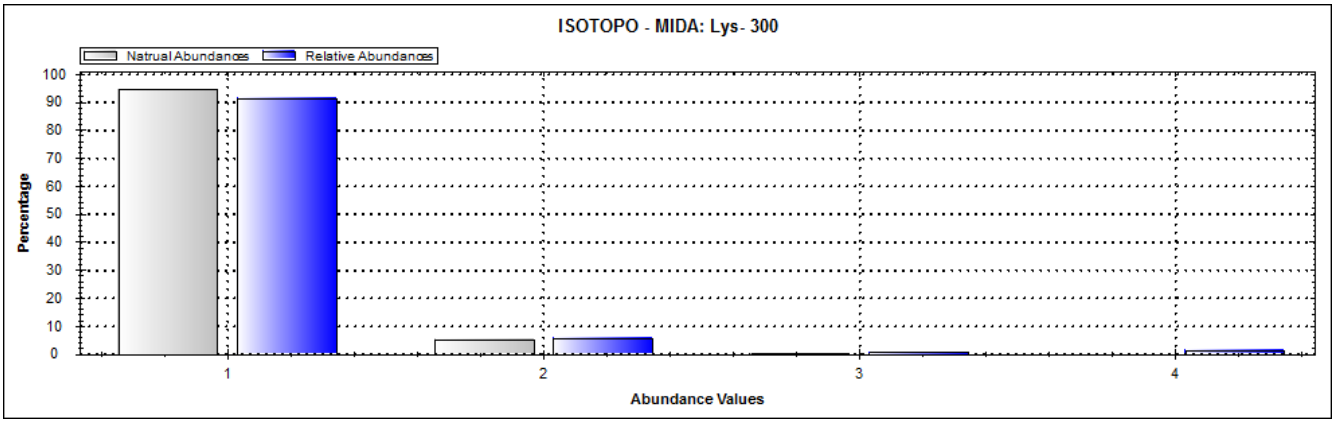


Figure 49 (c): Lys 300 Spectrum Analysis

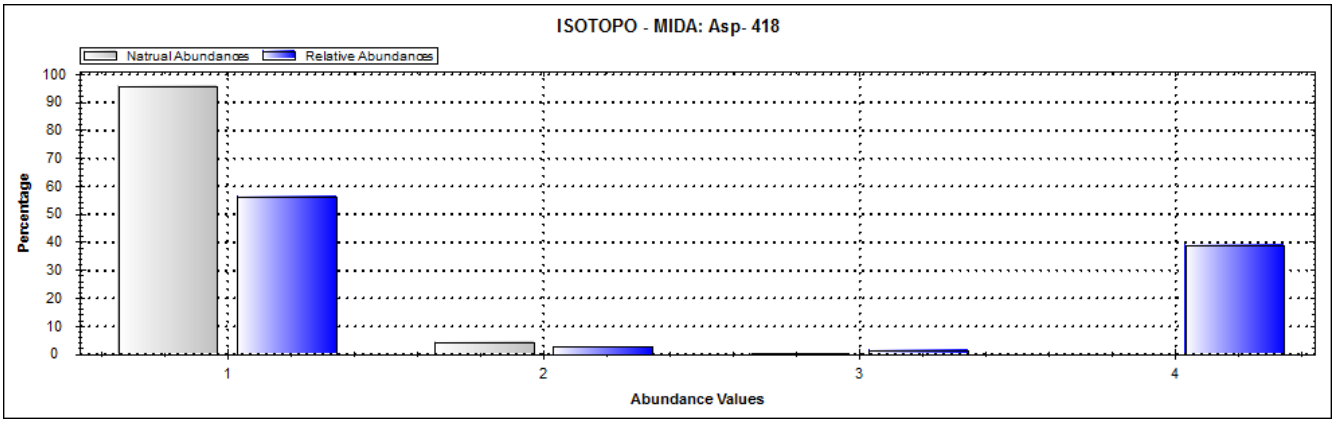


Figure 49 (d): Asp 418 Spectrum Analysis

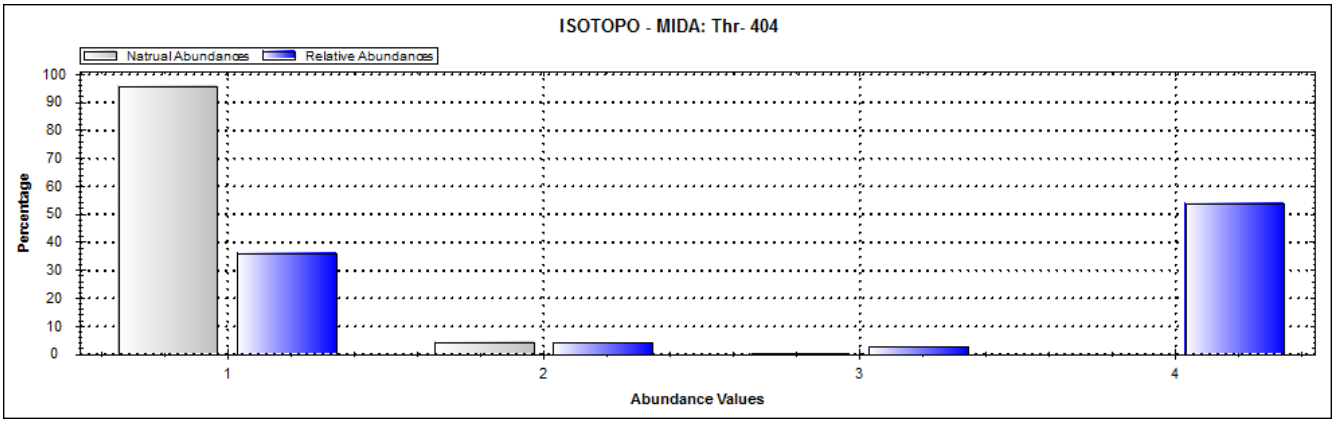


Figure 49 (e): Thr 404 Spectrum Analysis

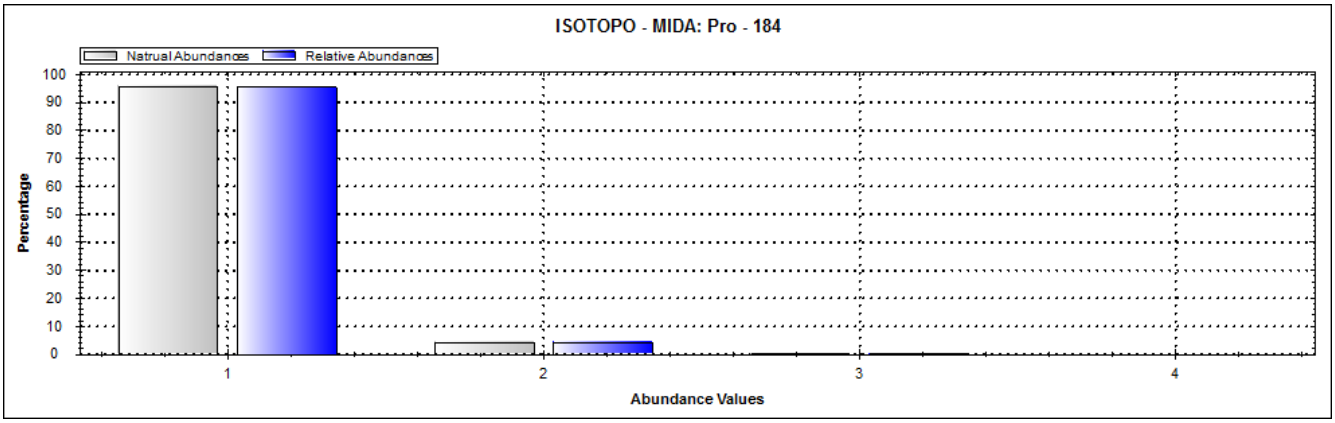
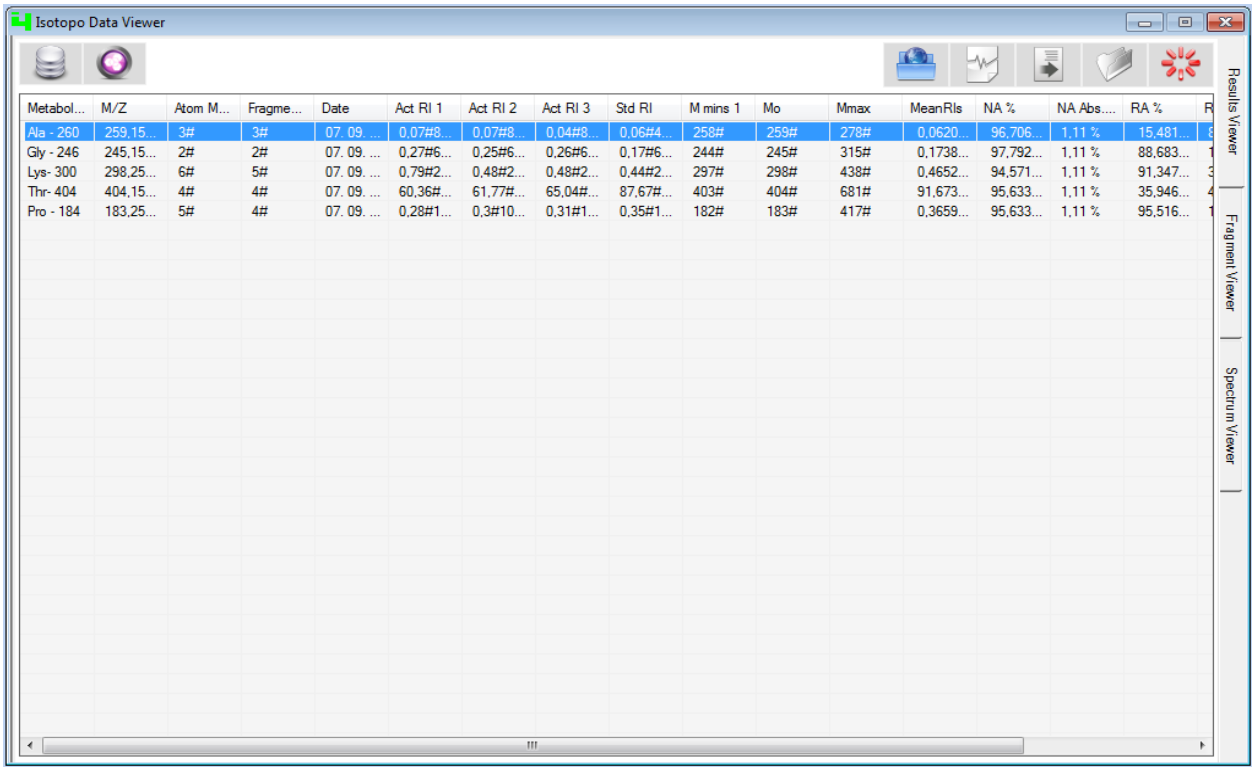


Figure 49 (f): Pro 184 Spectrum Analysis

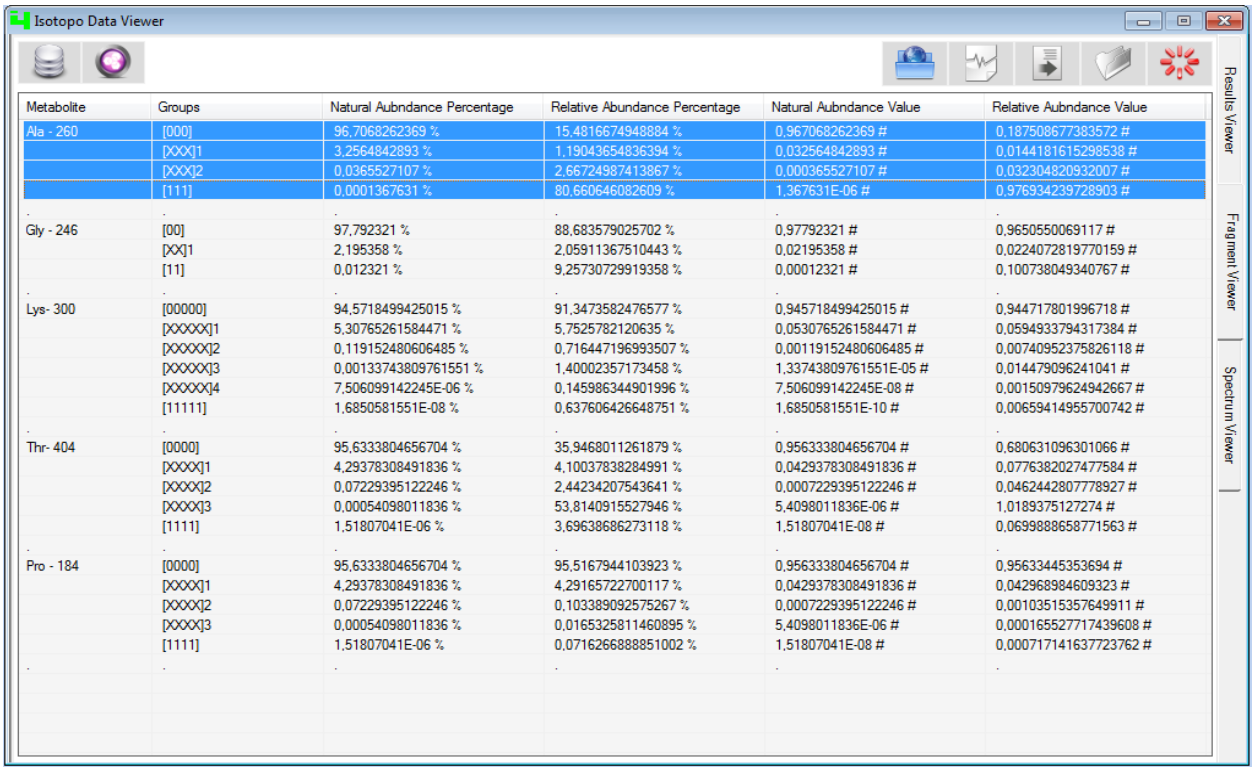
Figure 49: Isotopo; Drawn Spectrums.

Figure Legend 49: This Figure presents the drawn spectrums, as the results of analyzed metabolites Ala–260 (a), Gly – 246 (b), Lys- 300 (c), Asp- 418 (d), Thr- 404 (e) and Pro-184 (f). Gray colored bar presents estimated percentage of natural abundance values and blue colored bar presents estimated relative abundance values.



Metabol...	M/Z	Atom M...	Fragme...	Date	Act RI 1	Act RI 2	Act RI 3	Std RI	M mins 1	Mo	Mmax	MeanRIs	NA %	NA Abs....	RA %	R
Ala - 260	259,15...	3#	3#	07. 09. ...	0,07#8...	0,07#8...	0,04#8...	0,06#4...	258#	259#	278#	0,0620...	96,706...	1,11 %	15,481...	8
Gly - 246	245,15...	2#	2#	07. 09. ...	0,27#6...	0,25#6...	0,26#6...	0,17#6...	244#	245#	315#	0,1738...	97,792...	1,11 %	88,683...	1
Lys- 300	298,25...	6#	5#	07. 09. ...	0,79#2...	0,48#2...	0,48#2...	0,44#2...	297#	298#	438#	0,4652...	94,571...	1,11 %	91,347...	3
Thr- 404	404,15...	4#	4#	07. 09. ...	60,36#...	61,77#...	65,04#...	87,67#...	403#	404#	681#	91,673...	95,633...	1,11 %	35,946...	4
Pro - 184	183,25...	5#	4#	07. 09. ...	0,28#1...	0,3#10...	0,31#1...	0,35#1...	182#	183#	417#	0,3659...	95,633...	1,11 %	95,516...	1

Figure 50 (a) – Complete output viewer



Metabolite	Groups	Natural Aubndance Percentage	Relative Abundance Percentage	Natural Aubndance Value	Relative Aubndance Value
Ala - 260	[000]	96,7068262369 %	15,4816674948884 %	0,967068262369 #	0,187508677383572 #
	[XXX]1	3,2564842893 %	1,19043654836394 %	0,032564842893 #	0,0144181615298538 #
	[XXX]2	0,0365527107 %	2,66724987413867 %	0,000365527107 #	0,032304820932007 #
	[111]	0,0001367631 %	80,660646082609 %	1,367631E-06 #	0,976934239728903 #
.	.	.	.	.	.
Gly - 246	[00]	97,792321 %	88,683579025702 %	0,97792321 #	0,9650550069117 #
	[XX]1	2,195358 %	2,05911367510443 %	0,02195358 #	0,0224072819770159 #
	[11]	0,012321 %	9,25730729919358 %	0,00012321 #	0,100738049340767 #
.	.	.	.	.	.
Lys- 300	[00000]	94,5718499425015 %	91,3473582476577 %	0,945718499425015 #	0,944717801996718 #
	[XXXXX]1	5,30765261584471 %	5,7525782120635 %	0,0530765261584471 #	0,0594933794317384 #
	[XXXXX]2	0,119152480606485 %	0,716447196993507 %	0,00119152480606485 #	0,00740952375826118 #
	[XXXXX]3	0,00133743809761551 %	1,40002357173458 %	1,33743809761551E-05 #	0,0144790962410 41 #
	[XXXXX]4	7,506099142245E-06 %	0,145986344901996 %	7,506099142245E-08 #	0,00150979624942667 #
	[11111]	1,6850581551E-08 %	0,637606426648751 %	1,6850581551E-10 #	0,00659414955700742 #
.	.	.	.	.	.
Thr- 404	[0000]	95,6333804656704 %	35,9468011261879 %	0,956333804656704 #	0,680631096301066 #
	[XXXX]1	4,29378308491836 %	4,10037838284991 %	0,0429378308491836 #	0,0776382027477584 #
	[XXXX]2	0,07229395122246 %	2,44234207543641 %	0,0007229395122246 #	0,0462442807778927 #
	[XXXX]3	0,00054098011836 %	53,8140915527946 %	5,4098011836E-06 #	1,0189375127274 #
	[1111]	1,51807041E-06 %	3,69638686273118 %	1,51807041E-08 #	0,0699888658771563 #
.	.	.	.	.	.
Pro - 184	[0000]	95,6333804656704 %	95,5167944103923 %	0,956333804656704 #	0,956334453536 94 #
	[XXXX]1	4,29378308491836 %	4,29165722700117 %	0,0429378308491836 #	0,0429689846093 23 #
	[XXXX]2	0,07229395122246 %	0,103389092575267 %	0,0007229395122246 #	0,0010351535764991 1 #
	[XXXX]3	0,00054098011836 %	0,0165325811460895 %	5,4098011836E-06 #	0,0001655277174396 08 #
	[1111]	1,51807041E-06 %	0,0716266888851002 %	1,51807041E-08 #	0,0007171416377237 62 #
.	.	.	.	.	.

Figure 50 (b) – Fragment viewer

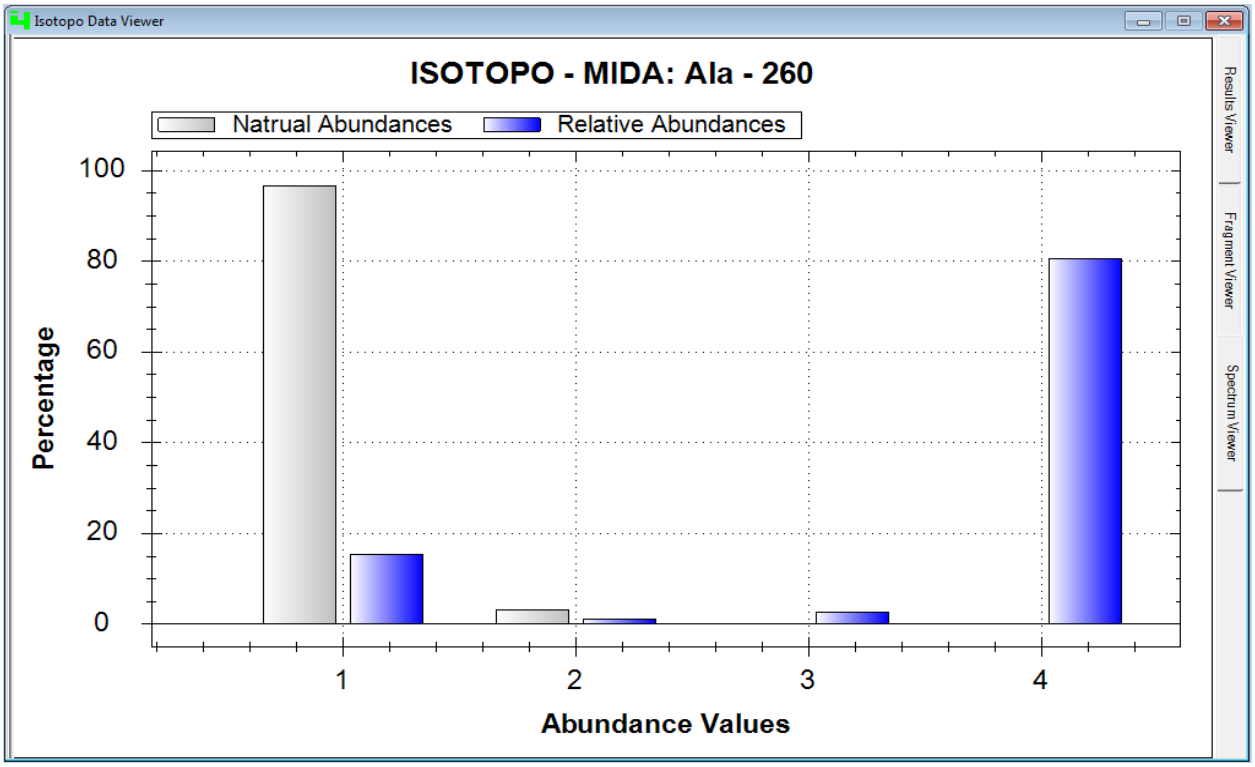


Figure 50 (c) – Spectrum

Figure 50: Isotopo; Data Viewer.

Figure Legend 50: Isotopo Data Viewer presents obtained (a): complete output and (b) estimated abundances with resptec to the number of fragment groups.

As earlier mentioned, the graphical user interface of Isotopo Data Viewer allows user to review and visualize obtained and saved file based output of the system. Figure 50a shows the obtained over all output of the Isotopo Data Analyzer and Figure 50b shows the measures natural abundance values and relative abundance values for each fragment. Whereas Figure 50c helps in user to visual, analyzer and print obtained results in the form of a spectrum bar chart with some textual information.

4.8 Installation

Take the setup executable SBEDA framework and install at Microsoft Windows operating systems. The SBEDA framework (including Isotopo) is developed using the Microsoft C# (sharp) programming language and Microsoft Dot Net framework 2008, that's why it is only compatible to Microsoft operating systems.

To run Isotopo Data Analyzer click on the “Pink Circled” icon in the main bar, to run Data Manager click on “Silver Cylinder Database” icon and to run Data Viewer click on “Blue Cylinder” icon. These all modules can also be run using main menu “Product →Isotopo→ Data Analyzer / Data Manager / Data Viewer”.

For a simple example and guided tour, load the example data into Isotopo Data Analyser by clicking the “Open Data File” icon, and process these by clicking “Process selected data” or “Process all data” icon.

5 Clinical Data Management

Using more than five years of personal (author) research and development in the field of Product Data Management, and considering the need for a new health care management system towards the problem of steadily increasing cardiovascular diseases in an aging population despite generally high health standards, proposed a new project i.e. Patient life cycle management system for cardiovascular disease. For platelets there is already a knowledge based solution available (produced by Prof. Dandekar's group of Functional genomics and systems Biology, Department of Bioinformatics, Biocenter at the University of Wuerzburg Germany) towards comprehensive protein-protein data interaction and management i.e. PlateletWeb (Boyanova et al., 2012) (http://plateletweb.bioapps.biozentrum.uni-wuerzburg.de/plateletweb.php). Here, further enhancing the scope of our research, after the proposition and implementation of a data management module for isotopologue data (section 4), we look forward towards a new clinical data management system proposition, capable of managing key cardiovascular diseases entities (stroke, thrombosis, failing heart, diabetic angiopathies) as well as individual case histories.

The contributions of efficient PDM systems (section 5.3) in industry for better product life cycle and data management is described (Ahmed et al., 2012c). Presented are importance, architectural designs and implementation details of some existing commercial and academic PDM Systems i.e. Windchill, CIM DATABASE, CPDM, WPDM and Promise PLM. To implement a new PDM System, this research presents major development guidelines, main components, mandatory functional requirements and some currently identified problems in PDM Systems. We support the medical community with the proposition and implementation of an innovative initiative, addressing major challenges of providing optimal management of a certain diagnosis, working rapidly under emergency conditions, protecting personal data, coupling individual patient data with general repositories, allowing therapy monitoring, analyzing individual variations with the and incorporation of golden standard therapy guidelines. The concepts of PDM transit the field of Bioinformatics for the implementation of an advanced clinical data analysis and management system. We propose layered conceptual and implementation architectures along with tools and technologies for the development of a prototype system.

5.1 Concept and objective

Cardiovascular diseases are steadily increasing in an aging European population despite generally high health standards. Together with increasing efforts in individualized medicine and direct burden of care for aged, multi-morbid patients, new, highly effective solutions for prevention and treatment of cardiovascular diseases in aging populations are needed. To achieve this we start from a scientific backbone of integrated genomics, proteomics and clinical laboratory data on key molecular cascades in thrombosis and hemostasis together with detailed data on individual genetic and postgenomic variation. Clinical data on key cardiovascular diseases entities (stroke, thrombosis, failing heart, diabetic angiopathies) as well as individual case histories are systematically collected and connected on the next level of investigation providing a systems biological view on disease states, changes and individual clinical variation. However, our project concept is to go beyond this and establish, integrating these two levels for better treatment of individual patients an intelligent individual health data management (PHS, Personal Health System) providing knowledge, prevention and optimal advice and care for individual patients. This is achieved using PDM techniques as framework on all levels considered, integrating both molecular data on pathophysiological cascades involved in thrombosis and haemostasis with laboratory and clinical data of key cardiovascular disease entities with the individual patient data and case histories for high quality healthcare in an aging population.

The resulting framework will include PGS, safety healthcare record information reuse (with disease management) and an integrated solution for the information infrastructure of larger repositories. World faces the challenge of a rapidly aging population, profiting from its good health system including state-provided comprehensive social care in its member states, a post-industrial low birth rate and good environmental conditions. Concomitantly, this can also lead to a high increase in cardiovascular disease where age is a strong and broad risk factor. Populations with an increased risk will be studied here. New possibilities to prevent and better treat cardiovascular disease are clearly sorely needed. Our proposed and applied research will also contribute to highly intelligent environments in support of the ageing population.

Starting from high end data on molecular cascades involved in cardiovascular diseases (PlateletWeb knowledgebase, human cardiogenomics, intelligent proteomics etc.) we can establish integrated software solutions to model and analyze these in the light of individual patient data and histories with predictions

regarding therapy monitoring and outcome. A key innovation will be to use techniques of PDM (product data management) for data integration including process life cycle management of patient data. This offers healthcare policy impact analysis and new personal health system development with the provision of a mechanism putting values (in the predictions) for treatment of diseases.

Exemplary medical problems will focus on key cardiovascular diseases with a focus on stroke and heart failure. We want to derive an integrated, scientific as well as doctor-patient-oriented data management system. It takes into account integrated modelling of key components (patient, cell types such as the platelet, doctor interaction, molecular processes, clinical and laboratory parameters). The data management system will have to cope with diverse and large-scale data with the aim to aid introduction of personalized medicine for the studied pathophysiological cascades. The proposed research can be divided into three main work packages (WPs): Scientific Data Integration (WP1), Clinical Data Integration (WP2) and Individualized patient history management (WP3). We will establish an integrated molecular database warehouse on critical signaling cascades in thrombosis and hemostasis with two main tasks: Implementation of an integrated data management and Integration, generation and analysis of experimental biological and in silico data by the other partners. WP2 can be divided into four tasks: PDM system development and deployment, friendly graphical user interface, semantic based search engine and patient data management system. WP3 Personal Health System includes integration of a PDM-minded user interface and natural language processing. We will have a PDM-based framework for data analysis. Clinical assays include different assays for hematological parameters, from standard assays to sophisticated phosphorylation assays on VASP phosphorylation. Such assays have furthermore been applied to specific patients and conditions, for instance diabetic patients. Deliverables expected include: Database system for experimental biological data management and manipulation, prototype PDM solution integrated with database management system and new PDM system managing individualized patient as well as in silico/ experimental generated molecular data. Finally, the PDM system development will achieve a flexible graphical multi user interface and semantic based search engine performing as a patient data management system with European scaling-up potential.

Clinics and university hospitals consist of different departments located in different places e.g. cities, countries etc. Every clinical place has its own setup, regulations, staff and patient data. There is no doubt that the treatment (practitioners, surgeons etc) departments are the important sections and expected to

play a vital role in the progress of a clinic by treating patients but at the same time the department taking care of patient data consisting of its personal and medical records (reports) cannot be ignored. In the past, there were no such systems available to store, track and manage patient data consisting of its personal and medical records (reports). This doesn't mean that there was no system for data management; there were some systems to store the information about clinics (or hospitals), personnel involved in different operations, financial details and patient data but there was no such comprehensive system to manage and share data globally. Each section of the health care management system has independent data and management system to maintain it. The job of existing data management systems is limited to storage, management and sharing of data for a limited number of people. Main data (associated to each section) are divided into two categories: private and public. In most of the cases private data is personal (patient, clinic and business) and not shared outside the boundaries unless it is really needed. In contrast public data is available with open access to people (e.g. notifications, warnings, guidelines and advertisements etc.).

The existing health care community (Europe or even Worldwide) advocates but does not have an implemented health care data management solution which could offer a life cycle management system, e.g. Porter's theory on organizational structure and information systems for health care. This proposal wants to achieve an integrated personal health system capable of providing the treatment life cycle data management of the patient (caused with any kind of disease), patient life cycle management maintaining individual patient data history, help in analyzing the individual and accumulative changes in a person's life over time caused by aging, hospital life cycle management and global information (only public) sharing among medical and related communities. Our research proposes an innovative approach bridging the world of research and decision making by providing relevant information about the medical, social, economic, legal and ethical issues related to the use of health technology. To support health care communities, primary care and hospitals, the clinical environments are heavily influenced by health informatics equipment e.g. digital health records, digital images, automated medication handling, work lists etc. To achieve the set objectives we present a systematic and global network based clinical data management system with conceptual architecture and the involvement of PDM system development concepts and methodological framework regarding complexity, efficiency, management and effectiveness.

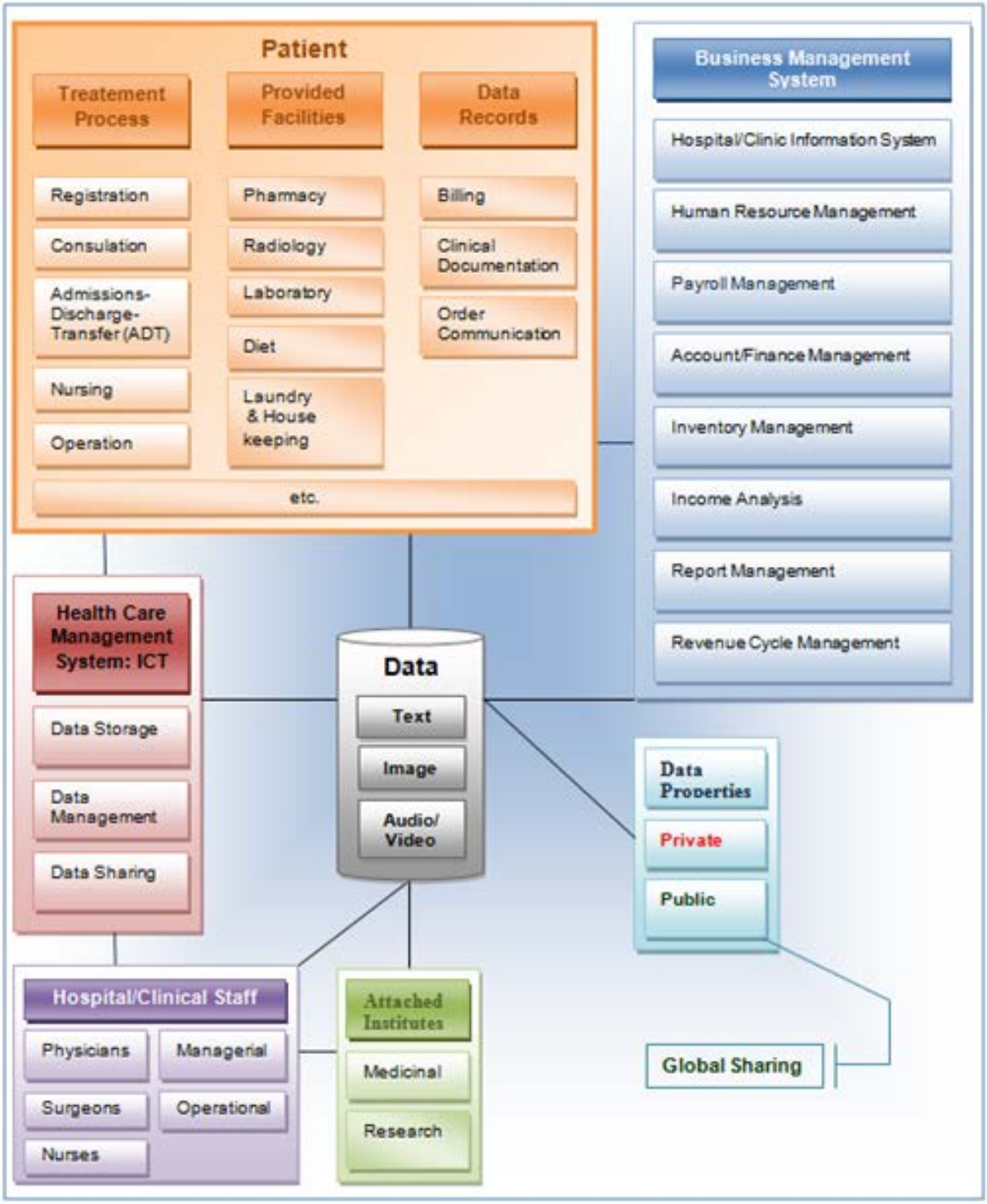


Figure 51: HCM System's basic Infrastructure

Figure Legend 51: This Figure consists of seven components: patient management, business management, hospital/clinical staff management, attached institutes, data, health care management system and data properties.

A further aim of this research is to address the challenges regarding the need for the continuous assessment of health technologies throughout their life cycles, the integration of social, organizational, ethical and legal aspects, and assessment of relative effectiveness to evaluate their implementation into European health service provisions (Figure 51) (Ahmed et al., 2012c).

5.2 Progress beyond the state-of-the-art

Societies (EU) are all facing the effects of an ageing populations with huge costs associated with age-related cardiovascular diseases. Thrombosis and hemostasis involve molecular cascades and pathophysiological conditions (Ikram et al., 2010) which lead to major cardiovascular pathologies such as stroke and failing heart. The widespread occurrence of cardiovascular diseases among the ageing population, is a new and strong challenge for European Medicine (Pamolli et al., 2011).

Though it is clear that pharmacological modulation of key molecular cascades is critical for progress and rational treatment, integrated data solutions for treatment approaches become only now feasible (Szent-Gyorgyi et al., 2011). We recently achieved the Platelet Knowledgebase as an important integrated solution (Boyanova et al., 2012). It offers the user multilevel access to molecular data integrating information on platelet action and pathophysiological cascades involved such as the key activatory GPIb/vWF pathway as well as inhibitory cyclic nucleotide pathways (Wangorsch et al., 2011). We include phosphorylation information and molecular sites (Boyanova et al., 2012). However, for the new project a broader view, more and diverse data as well as individual patient data information have to be integrated with clinical settings including participation of the clinic. In particular, early and correct treatment in pathological conditions such as acute ischemic stroke (Grube et al., 2011) as well as myocardial infarction (Petritsch et al., 2011) has to be closely monitored and managed. The proposal will focus on risk populations with risk factors for stroke as well as cardiovascular disease in general (Ovbiagele et al., 2011). Furthermore, genetic or pathophysiological conditions affecting hemostatic equilibrium or patient responsitivity to treatment pose a major risk and therefore demands early detection and specific regimen as well as anticoagulants and the corresponding increased risk of hemorrhage due to genetic polymorphism (Bonello et al., 2010; Collet et al., 2009; Fitzgerald et al. 2011).

Diagnostic tools facilitating the identification of potential therapy failure or drug resistance and appropriate data management are thus urgently needed (Lanktree et al., 2010). Highly sensitive (Mindukshev et al. 2012) and high-throughput ready techniques (Geiger et al., 2005; Geiger et al., 2010) have been developed to meet these requirements. PDM is a powerful approach to guarantee optimal product data management (Ahmed, Detlef., 2007) (Ahmed, 2010d) (Ahmed, 2011a). This has not yet been applied to this research question. Certainly, the challenge to provide a distributed, individualized data management solution both to scientists and medical doctors is enormous both regarding the amount of data to be processed (Pathak et al., 2012), as well as the spread of access nodes with different requirements involved (Peng et al., 2010): However, the consortium is unique to tackle this challenge as we provide novel and powerful solutions touching on latest developments in PDM such as intelligent agent-based technology (Gnosisplace; Ahmed, 2009), natural language processing (Ahmed et al., 2012e; G&L) and latest web-based problem-guided user interfaces (Ahmed, 2011b; Gnosisplace; PIAP). Furthermore, we combine this with latest data management solutions for genomics (Biobyte Solutions), proteomics (UMIST); diagnostic monitoring of cascades in thrombosis and hemostasis as well as clinical patient management involving large-scale data transfer and user communities (University clinics Würzburg, Rostock, Craiova, Rieti).

5.3 Product Data Management (PDM) System

In early 1970s there was no such system to automate the process of data management, then in 1980s Computer Integrated Manufacturing was introduced but seemed not to be successful in product data management. With emergence of Computer Aided Design (CAD) technologies PDM Systems are introduced and used to manage engineering data, activities and processes with enhanced control of engineering related data, process oriented changes, development activities, technical operations and artifact patterns. PDM is also renowned as Engineering Data Management and Engineering Document Management Systems because it provides better management and control over engineering data, activities, and changes related to design and manufacture of product. Product Lifecycle Management (PLM) is another acronym for PDM to manage the entire development life cycle of the product by integrating people, data, processes and business systems.

PDM functions like a spine throughout the product life cycle (development) by managing the manufacturing (e.g. documents, CAD etc.) and business related information. Moreover PDM also

supports product teams and techniques by providing Concurrent Engineering in improving engineering workflow. PDM systems also address the security, reuse and control related matters as well. PDM performs five main functions to integrate and manage all applications, information, and processes during the associated product life cycle i.e., Data vault and document management, Workflow and process management, Product structure management, Parts management and Program management (Sung and Sam, 2007).

The major objectives of PDM are to reduce the cost of engineering, reduce effort in product development life cycle, reduce time in change handling and new product development. This improves the quality and services of the product, delivers and support products at a given time, improves team coordination, increases customization of products and maintains product configuration based information. Informatics challenges involve management of large volumes of data generated by computer based systems, reduce engineering environment based problems, provide better access to information, provide better reuse of design information, provide common data warehousing, secure engineering data's originality, prevent error creation and propagation. All this combines for a strong effect on market shares. Moreover PDM is also supposed to handle business process work flows, change management, revision control, product configurations, product structure management, project tracking and resource planning.

Every PDM System consists of some basic components which most of the time need to be developed and improved. Data management; allowing user to use, enter, edit, store, incorporate, protect and recuperate data by administering in warehouse. Networking; connecting computer systems in a networked infrastructure (e.g. local or wireless area network etc). Interface; supporting user queries using menu(s) and form driven inputs and report. Information and Workflow Structure Definition; managing resources, events and responsibilities. Information Structure Management; producing exact structure based on maintained information in the system. Workflow Structure; defining flow of engineering activities. Workflow Control; controlling engineering process. System Administration; setting up administration and maintain the configuration of the system. Many commercial and academic PDM Systems exists. In this research paper we are only presenting some of the promising and well accepted PDM Systems in detail i.e. Windchill (Andreas et al., 2008), CIM Database (Claudia et al., 2007), Component-based Product Data Management system (CPDM) (Sung and Sam, 2007), WPDM (Huang et al., 2004) and Promise Product Life Cycle Management and Information Tracking Using Smart Embedded Systems

(Andreas et al., 2008). However some academic research projects are also contributing towards the advancement of PDM system implementation and usage e.g. I-SOAS (Ahmed, 2009).

5.4 Methodology

An integrated database management solution, integrating scientific and clinical management, demands for molecular cascades in thrombosis and haemostasis.

The concepts of product data management will help in innovatively implementing a new data management system both for individualized patient treatment as well as individualized cascades in thrombosis and haemostasis. The focus will be on key cascades (e.g. in platelets GPIb/vWF and cyclic nucleotide pathways. Next individualized patient data will be considered. Background variation will be analyzed from large-scale SNP data, concrete patient information and laboratory data on variation will come from the university clinic data already in place. Only then, when a frame-work for management of such data has been established, the clinic data on new patients will be incorporated. Furthermore, distributed web solutions will be carefully tested from the start. This will first include only very few nodes (consortium) but after extensive validation solutions will migrate to a few and then many wards in the university clinic partners. For the latter data integrity, in particular of the clinic data is a major issue.

The proposition allows fusion of two types of data: scientifically focused general knowledge on specific molecular cascades, data and databases, as well as more clinic-oriented, patient-specific modifications. Building on both and integrating individual patient interactions and therapy, our key goal is a good and powerful, PDM-driven personal health management system. Dynamical modelling is possible and included in preparatory work for the project. However, this will only be attempted in selected examples where large and accurate data sets are available (Boyanova et al., 2011; Wangorsch et al., 2011). All involved partners can build on own success-stories on integrated data solutions. For the clinical partners this involves important patient data and selected disease cohorts they can provide and previously managed diagnostically and therapeutical.

Furthermore, upcoming challenges and obstacles in the project regarding individual software solutions can be tackled as partners evolved independent data management solutions in previous work, for instance different types of agent-based simulations for PDM, various web user interfaces, database solutions and

patient ward management. Conducted research try to tackle also the computational challenges involved such as large scale data, multiple users and cloud computing.

5.5 Tasks

Main objectives are the implementation of a successful data management system development with multirole flexible graphical user interface development, natural language search engine and an individual patient health management system is in place with scaling-up potential. As this research is about to propose a new innovative approach bridging the world of research and decision making by providing relevant information about the medical, social, economic, legal and ethical issues related to the use of health technology. To support health care communities, primary care and hospitals, the clinical environments are heavily influenced with health informatics equipment e.g. digital health records, digital images, automated medication handling, work lists etc.

Databases, networking, graphical user interfaces, intelligent decision support systems and specialized programming languages are just a few of the technologies currently used in medical informatics. Mobility and ubiquity in healthcare systems, standardization of technologies and procedures, certification, privacy are some of the issues that medical informatics professionals are need to address in order to further promote ICT in healthcare.

This research implements the new European health system management strategy: PDM allows to realize the envisioned Patient Life Cycle Management; integrating patient information (throughout project time, but including individual case histories and open for continuous support in future) from multiple sources and location with active patient participation via ubiquitous access.

A data management approach is need to be implemented, providing a secure and global data (e.g. digital health records, digital images etc.) management, modelling, simulation and visualization with intelligent, multi role based, flexible and self learning graphical user interface. This proposed data management system will be a multi role (user could of any kind, belonging to any profession but able to use computer or mobile) desktop, web and ubiquitous system which could directly be accessed by using a desktop networked computer, by using web interface (via internet) and by using ubiquitous (e.g. mobile etc.).

This research is also to design and support the medicine community with the proposition and implementation of an innovative initiative, addressing major challenges of providing optimal management of a certain diagnosis, working rapidly under emergency conditions, protecting personal data and coupling individual patient data with general repositories. We propose a systematic and global network based data management system architecture with the involvement of PDM system development concepts and methodological framework regarding complexity, efficiency, management and effectiveness.

It consists of different services and functionalities at a time for many types of roles/users like medicine related staff e.g. physicians, surgeons, nurses etc., technical staff for the organizational record, operational and quality management, laboratory staff, and scientists from associated research institutes (in university hospitals). Core concept of a professional, friendly and flexible graphical user interface so then new and different users can easily learn, use and adopt. New PDM system where the user is provided with the flexibility of redesigning GUI by changing the default orientation of the provided controls (by adding or deleting) in default GUI of the PDM system according to his needs. Several innovations in preparatory work will help to establish a fast responding, optimal interactive system. Building and complementing the work of the clinical colleagues in WP2, here SCL will be the key clinical partner, judging from its broad spectre of individual patients with cardiovascular disease whether for our four focus disease entities in fact the system performs well and allows individual patient as well as doctor queries in an optimal way and with a strong and intuitive user-interface. Building on preparatory work, we will build a search engine following the concept of artificial intelligence, semantic web and natural language process to structure data in an organized way. Extract the Meta data out of data using semantic web and language processing technologies for a new PDM system. Evaluation and users provided by SCL as well as the other clinical partners.

A patient data management system is needed to be deployed, which should be capable of managing all clinic, staff and patients data. If this kind of system will be available then it will be a global step by involving the communities of different medicine related people together at one plate form, sharing patients data and managing patient’s recovery processes. It will not only help the established medical research institutes to get updated and managed but will also give the opportunity to mainstream organizations as well.

5.6 Potential of PDM for Clinical Patient Data Management

The challenge for high quality healthcare and ageing is addressed to provide a new comprehensive solution towards Personal Health Systems Implementation for PGS, safety healthcare record information reuse (with disease management) with integrated solution for the information infrastructure of larger repositories. The subject Health is aligned with the fundamental objectives of medicine related research to improve the health of living beings and increase the competitiveness of health related services, socio economic dimensions of health care and global clinical data sharing. Our focus is on one specific aspect of Bioinformatical analysis: the molecular processes involved in cardiovascular diseases. We offer an integrated software solution to model and analyze this together with an application oriented data framework. The aim of this research is to enable living beings to live longer, independently, in good health by increasing the average number of healthy life years to improve the sustainability and efficiency of our social and healthcare systems. Furthermore the scope is up to the proposition of creating innovative products and services (global health care management system) for international markets.

With a view to achieve the objectives of this research, a key innovation will be to use techniques of PDM (product data management) towards better process life cycle management of healthcare organizations for trusted governance and policy impact analysis and new personal health systems development with the provision of a mechanism putting values (in the predictions) for treatment of diseases. We are trying to derive an integrated, scientific as well as doctor-patient-oriented data management system. It takes into account integrated modelling of key components (patient, cell types such as the platelet, doctor interaction, molecular processes, clinical and laboratory parameters). The data management system will have to cope with diverse and large-scale data with the aim to aid introduction of personalized medicine for the studied pathophysiological cascades.

If the clinic is small e.g. consisting of one or two special practitioners and taking care of a limited number of patients (on daily basis) and trying to cure not very serious diseases then its fine but if the clinical place is of large scale with many practitioners and surgeons, taking care of different kinds of patients including serious common as well as rare diseases, then it must be taking care of a huge amount of data which needs to be properly managed. To successfully cure a patient a good and on time medical treatment is needed, likewise, to successfully run a clinic (hospital) a patient data management system is also need to be deployed, which should be capable of managing all clinic, staff and patient data. In many disease

categories there is no such patient data management system available with the implementation of product data management principles, which could help in improving the mechanism of global information sharing and patient's recovery process improvements. If this kind of system will be available then it will be a global step by involving the communities of different medicine related people together at one platform, sharing patient's data and managing patient's recovery processes. It will not only help the established medical research institutes to get updated and managed but will also give the opportunity to mainstream organizations as well.

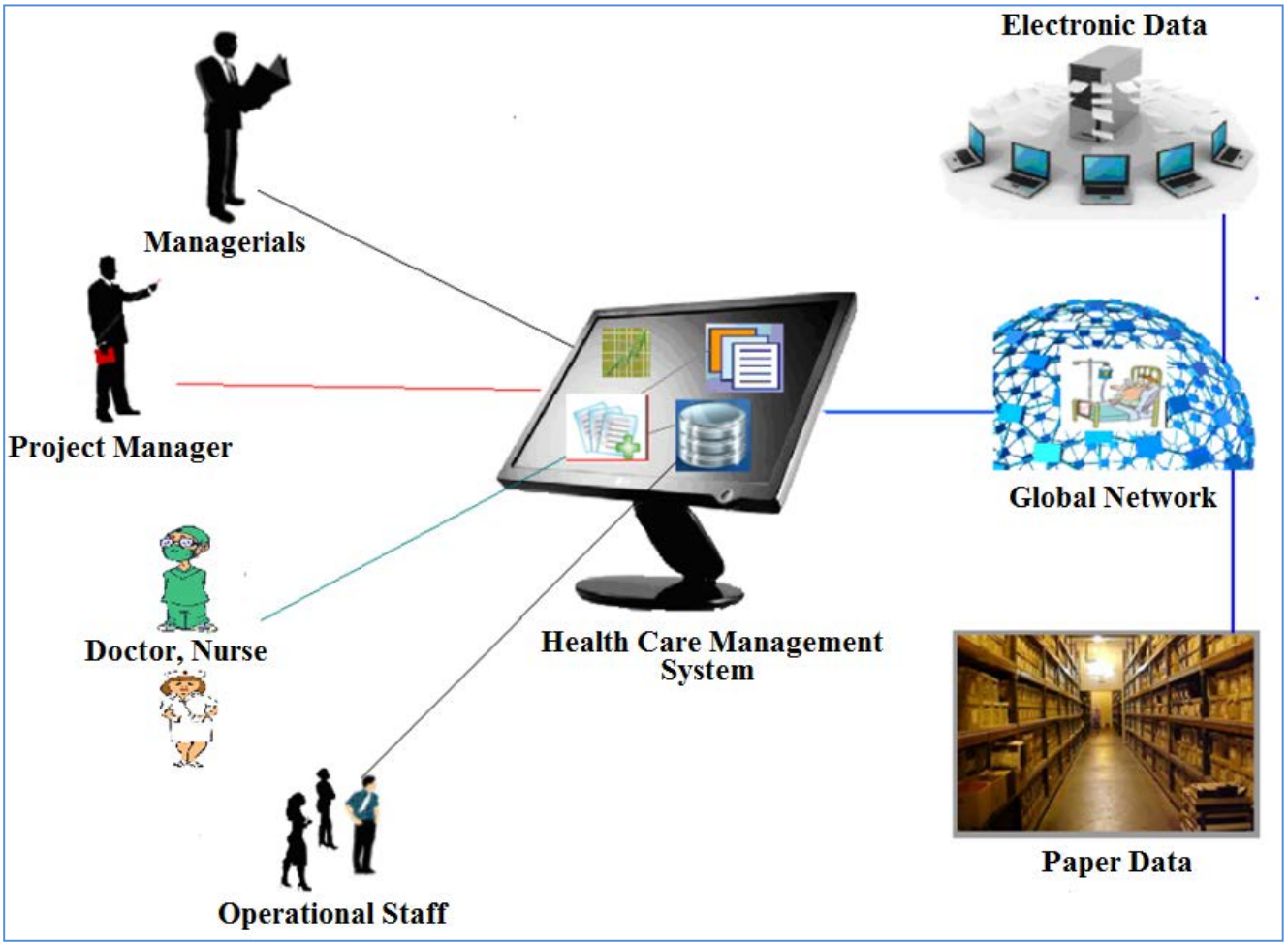


Figure 52: Multirole Health Care Management System

Figure Legend 52: This Figure is divided in four categories: man (managerial, project manager, doctor, nurse, operational staff), machine (software and hardware system), network (online and intranet) and archive (electronic database and physical data).

One of the main goals of this research is to cope with the above issues and provide a new patient data management system for clinical and university hospital data management. A successful PDM system development and deployment is a quite difficult task because it is time consuming, expensive and there is high chance that likewise other industries, most of the staff will not give importance to PDM system and without the involvement of organizational personals it is quite difficult to deploy it. As the new and proposed PDM systems will definitely consist of different services and functionalities at a time for many types of roles/users like medicine related staff e.g. physicians, surgeons, nurses etc., technical staff for the organizational record, operational and quality management, laboratory staff, and scientists from associated research institutes (in university hospitals), as shown in Figure 52 (Ahmed et al., 2012c). We believe if the product will be with many benefits but the graphical interface is not user friendly then it could be a disaster. So to avoid such future complications we will involve the core concepts of a professional, friendly and flexible graphical user interface so then new and different users can easily learn, use and adopt it.

The new PDM system will be designed for the different role-based client users (multiple users playing different roles with different rights in the same organizations). So the probability of predicting that every user does not need all the options of the system all the time is very high. Moreover we can also say that the massive availability of all the controls to all the users all the time will also reduce the speed and efficiency in the work of the users because if a user will only be provided with some limited options with respect to his nature of job, rights and responsibilities then it will be much faster and more convenient for him to use and perform in the system. Moreover if the user is also provided with the flexibility of redesigning GUI by changing the default orientation of the provided controls (by adding or deleting) in default GUI of the PDM system according to his ease and the need then it will also be a useful contribution at the user's end.

This new PDM system will definitely contain and manage heavy amount of data, which itself will be a big achievement but on the other hand the problem can start when user will need to find out some information out of this heavy data. If the data extraction process will be like old fashioned then it might increase the user headache by forcing him in filling different criteria and forms etc. To avoid such issue and provide efficient data extraction process, we will design a search engine following the concept of

artificial intelligence, semantic web (2001a ; 2001b) and natural language process to structure data in an organized way. Although designing an intelligent application capable of reading and analyzing a user's structured and unstructured natural language based requests and then extracting desired concrete and optimized results from database is still a challenging task for the designers because it is very difficult to completely extract the Meta data out of data but using some examining semantic web and language processing technologies we will try to implement a better search for this new PDM system.

The new PDM system will consider as the Product the optimal management of a certain diagnosis; furthermore it will be capable of working rapidly even under emergency conditions. It will protect personal data but at the same time it will couple individual patient data with general repositories. A well devised PDM system will even allow therapy monitoring. This will start from diagnostic monitoring. Next, by analyzing individual variations and incorporation of golden standard therapy guidelines (including statistics on actual treatment success) this will then translate into PDM managed treatment protocols.

5.7 Implementation

The equipment available for the project is in place and with the respective partners. This includes in particular high-end proteomics platform, top of the line haematological and laboratory equipment, PDM capacities and High power computing and biological databases. Proposed life cycle management system provides the treatment life cycle management of the patient, caused with any kind of disease. It gives a vision for the patient life cycle management by maintaining individual patient history, which will also help in analyzing the individual and accumulative changes in a person's life over time caused by aging. Furthermore, it gives the solution for hospital life cycle management and global information (only public) sharing among medicinal and related communities.

The proposed layered architecture is shown in Figure 53 (Ahmed et al., 2012c) is consists of three main layers. In the beginning the scientific data will be gathered and maintained in developed database using data management interface. In the second layer clinical data integration will be performed along with the implementation of PDM concepts for the development of a prototype PDM system for Patient data management with the availability of a flexible graphical user interface and natural language based search implementation. In the very last layer of the architecture, integrated PDM system will be available in the

form of a Health Care data management system taking care of data analysis, management and governance issues. The expected impacts for this PDM development include demonstration of a novel, highly integrated data management, development of PDM-oriented, individual patient data processing in clinic as well as in research as a basis for individualized medicine. We will bring new PDM solutions to a new and large market. This will be made possible by new technological advances in computing and man-machine interfaces.

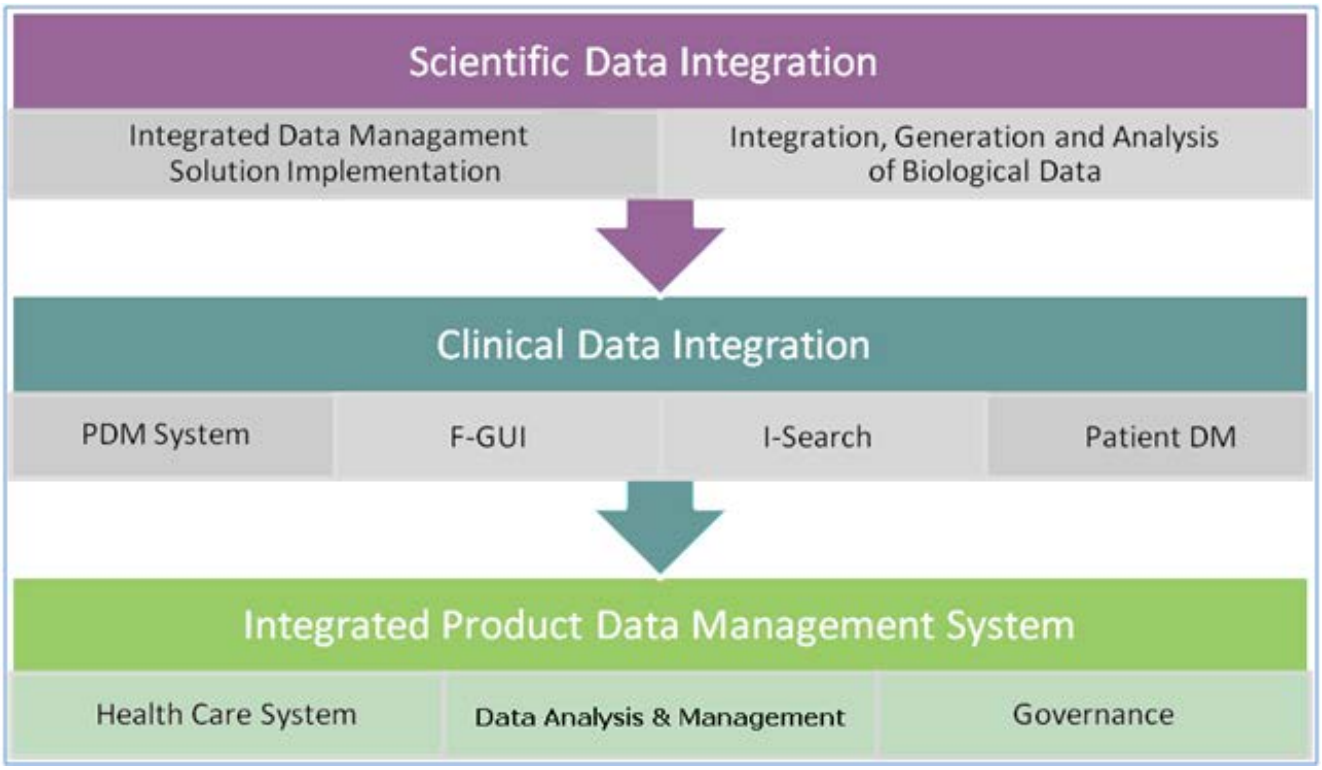


Figure 53: Layer Architecture; Clinical Data Management

Figure Legend 53. Figure presents three layer architecture divided into three main layers: Scientific Data Integration (integrated data management solution implementation), Clinical Data Integration (PDM System, F-GUI, I-Search, Patient DM) and Integrated Product Data Management System (Health Care System, Data Analysis & Management, Governance).

The implemented version may contribute in promoting innovation by strengthening the links between academia and industry, focusing on diagnosing infectious diseases, management of diabetes and health technology assessments. Furthermore it may also increase the development of new technologies for diagnosis, investigating clinical trials for treatment optimization, sustaining the competitiveness of the health with the involvement of different public health policies with intellectual property rights

management and knowledge transfer. It is obvious that this research work will contribute to various social innovations in the heath sector, independently of whether they are related to medical technologies, ageing, health care, specific diseases.

5.8 Technologies

5.8.1 Graphical User Interface (GUI) Development

Targeting the challenge of the proposition of designing a standardized desktop and web based graphical user interface, a review research is conducted in a chosen the field i.e. Human Computer Interaction (HCI), to have complete understanding of graphical user interface design and development. HCI renown as Human Machine Interface (HMI); the study of designing, evaluating and implementing interactive computing systems for human use (Ahmed, 2008). Designing High quality HCI design is difficult to implement because of many reasons i.e. market pressure of less time development, rapid functionality addition during development, excessive several iterations, competitive general purpose software and human behavior analysis.

Designing human computer interaction interface is an important and a complex task, but it could be simplified by decomposing task into subcomponents and maintaining relationships among those subcomponents. Task decomposition is a structured approach, applicable in both Software Engineering and HCI fields depending on specific processes and design artifacts. Using design artifacts applications could be made for analysis and design by making the hand draw sketches to provide high level of logical design based on user requirements, usage scenarios and essential use cases. To design hand draw sketches there are some strategies to be followed i.e. planning, sequential work flow, and levels of details.

To meet the requirement of implementing a flexible web based graphical user interface design, the field of Human Computer Interaction and Rich Internet Application will be considered researched. The term "Rich Internet Application" (RIA) was introduced in a white paper of March 2002 by Macromedia. RIA is web application with features and functionalities of traditional desktop applications as well as web applications. Traditional web applications centre all activities around client server architecture with a thin client where as RIA typically transfer the processing necessary for the user interface to the web client but keeps the bulk of the data (i.e., maintaining the state of the program) back on the application server. RIA shares one characteristic with other web development technologies, an intermediate layer of code often

called a Client Engine, between the user and the server. This client engine is usually downloaded as part of the instantiation of the application, and may be supplemented by further code downloads as use of the application progresses. The client engine acts as an extension of the browser, and usually takes over responsibility for rendering the application's user interface and for server communication. Using Client Engine RIA becomes richer, much responsive, balanced, asynchronous and efficient. There are several RIA technologies available i.e. FLEX (Adobe), AJAX, OpenLaszlo and Silverlight (Microsoft) (Zeeshan, Popov, 2010b), could be used in development. Preferably we are looking forward to choose Flex; because it is a highly productive, free open source framework for building and maintaining expressive web applications that deploy consistently on all major browsers, desktops, and operating systems. While Flex applications can be built using only the free Flex SDK, developers can use Adobe Flex Builder™ 3 software to dramatically accelerate development. Adobe Flex is a collection of technologies released by Adobe Systems for the development and deployment of cross platform rich Internet applications based on the proprietary Adobe Flash platform.

5.8.2 Intelligent Search

Semantic Web technologies will also be considered presenting information over the web in a format so that human being as well as machines can understand the semantic of context (Ahmed et al., 2012d). Semantic Web is a mechanism for presenting information over the web in such a form so that humans as well as machines can understand the semantic of the context. It is a linked mesh of information which could be processed. The aim of Semantic Web is to produce technologies and domain specific languages capable of reasoning on semi structured information. Semantic Web is an intelligent conception and advancement in World Wide Web to collect, manipulate and annotate information independently by providing effective access to the information. It provides categorization and uniform access to the resources and advances the transformation of World Wide Web into semantically modeled knowledge representation systems with a common framework which allows data to be shared and reused. It also gives the concept of semantic based web services for dynamic composition of service based applications. Semantic Web research depends on a number of key methodologies: Knowledge Representation Languages and Reasoning Algorithms.

The ultimate goal of semantic web is to structure the meaningful contents of unstructured published data over web to take advantage in improving the data extraction processes and to involve knowledge

management in creating an advanced knowledge modeled management systems. Currently Semantic Web is standing on a very important building block of Ontology, aims of structuring data into processable semantic models, as the collection of interrelated semantic oriented concepts. Ontology is the explicit representation and description of already available finite sets of terms and concepts used to make the abstract model of a particular domain. Ontology has become a favorite subject for different research communities e.g. computer science, philosophy, bioinformatics, knowledge engineering & management, natural language processing, information retrieval, cooperative information systems and information incorporation etc., because of its interdisciplinary nature. With variability in its usage, ontology has different definitions with respect to the different fields e.g. in computer science it is defined as the combination of concepts and relationships for domain modelling, in philosophy it is known as the study of "what there is", or a mathematical formulation of properties and relationships of certain entities, and so on. Ontologies can be constructed manually and automatically by using some ontology supporting languages i.e., XML, RDF and OWL (Web Ontology Language) offering ways of more explicitly structuring and richly annotating Web pages (Ahmed, Tacheva, 2010c).

5.8.3 Database Implementation

The database management system and knowledge base will be implemented using affordable Relational Database Management System, preferably Oracle, for efficient data storage, access and management.

6 Discussion and Conclusions

6.1 Software Comparison towards MFA

Metabolic flux analysis is complex and requires dedicated software. It is rapidly developing offering a multitude of different solutions to the user. New experimental data on metabolites and enzymes induce high interest in metabolic modelling including metabolic flux calculations. Data analysis of metabolites, calculation of metabolic fluxes, pathways and their condition-specific strengths is now possible by an advantageous combination of specific software. How can available software for metabolic modelling be improved from a computational point of view? A number of available and well established software solutions are first discussed individually. This includes information on software origin, capabilities, development and used methodology. Performance information is obtained for the compared software using provided example data sets.

A feature based comparison shows limitations and advantages of the compared software for specific tasks in metabolic modeling. Often found limitations include third party software dependence, no comprehensive database management and no standard format for data input and output. Graphical visualization can be improved for complex data visualization and at the web based graphical interface. Other areas for development are platform independency, product line architecture, data standardization, open source movement and new methodologies.

Due to limitations of space, we restrict detailed comparisons to a choice which cover a broad range of the different analysis steps in metabolic modeling. We opted for user-friendly programs with broad applicability and compare these to a background of alternatives. These have specific advantages and limitations as outlined. Typical general challenges remain. However, as the field is moving fast, the capabilities of metabolic modeling software are steadily improving.

The comparison shows clearly space for further software application development including steps towards an optimal user friendly graphical user interface, platform independence, database management system and third party independence especially in the case of desktop applications. The found limitations are not limited to the software compared and are of course also actively tackled in some of the most recent developments (see other references). Other improvements should aim at generality and standard data input formats, improved visualization of not only the input data set but also analyzed results.

With the implementation of these suggestions, metabolic software applications will become more professional, cheap, reliable and attractive for the user. Nevertheless, keeping these inherent limitations in mind, we are confident that the tools compared can be recommended for metabolic modeling covering broad aspects of the involved tasks and analysis steps.

6.2 LS-MIDA and Isotopo towards MIDA

MIDA is a technique towards the measurement of amalgamation of polymers by involving the process of quantification of relative abundances of molecular species with mass spectrometry. The objective of this research is to study metabolic isotope to quantify the fraction of metabolites of interest in the mixture typically by tracing isotopes. Estimating mass isotopomers distribution from spectral data is an extension of the quantitative mass spectrometric method to a multi component mixture analysis. We implemented the software application LS-MIDA for Brauman's least square method including binomial expansion for the prediction of natural and relative abundances using data from metabolic labeling experiments.

The effectiveness of the Least Square Mass Isotopomers Analyzer was extensively tested and results are presented from Salmonella labeling experiments for different metabolites such as analysis of various amino acids. The statistical information is calculated and provided in three different forms i.e. natural abundances, relative abundances and percentage of relative abundances per m/z values and a spectrum is drawn for better analysis.

Later than extending the scope of ongoing research and development, focusing on identification of the quantity of population of labeled isotopomers for resolving the exact rate of synthesized fractions present in the mixture and metabolic experimental data management, a new software application named "Isotopo" is developed. Isotopo is an application with the ability of performing quantitative mass spectrometry to readily mixtures of materials labeled with stable isotopes can be very important for both biomedicine and biochemistry.

Most recent version of Isotopo is with the ability of processing experimental isotopomers data and as the result of successful experimental data processing, Isotopo estimates mass values and relative intensities. It predicts natural abundance values, relative isotopic abundance values and fractional molar abundance values of each fragment from labeled substance based experimental data elements. Using Isotopo it is also possible to process data sets with multiple data entries up to three actual intensity values against one

mass to charge ratio values, estimate absolute enrichment, mean and standard deviation of both natural and relative isotopic abundances. Isotopo also provides the standardization of experimental data with a file based record keeping system for experimental data manipulation and management. The strength and effectiveness of Isotopo is validated using datasets based on observed values during GC-MS experimentation, consisting of the information obtained using different metabolites. Resultant outcomes are observed in both textual and visual (spectrometric visualization) formats, produced by Isotopo.

During evaluation of both LS-MIDA and Isotopo, all data examples given are from Salmonella isotopologue measurements. Accumulating such measurements and taking further data on their metabolism into account allows insights into Salmonella and their metabolism during infection (Figure 54; reference (Eisenreich et al., 2010)). The diagram (Figure 54) shows flux adaptation in Salmonella and here new from which part of the metabolic flux maps the examples given are derived. For instance, the data on alanine measurements can be processed following the short tutorial.

Glucose, glucose-6P and gluconate present possible carbon sources for intracellular pathogens. For such carbon sources the enzymes and fluxes for glycolysis and for the Entner–Doudoroff pathway are up-regulated in these bacteria. In contrast, most enzymes and the fluxes in the TCA cycle are down-regulated.

The Salmonella model (Figure 54) shows that here preferred carbohydrates are glucose and glycerol-3-phosphate. Furthermore, entry points for key amino acids into the metabolism during infection are given (blue boxes). The strength of the fluxes corresponds to the situation when these key carbon sources are there. Investigations in the presence of different amino acids with isotoplogue labeling and their metabolism lead to the data examples given here (see above for detailed processing in section 3 and 4).

The amino acids given are efficiently metabolized and enter the pathways roughly according to their position on the pathway map.

Currently we investigate the effect of different enzyme mutations on the metabolism, for instance regarding PEP carboxylase (Dandekar et al., 2012). If this enzyme is impaired, no anaplerotic reaction is possible. Similarly, mutations in the lower part of glycolysis are interesting, for instance around the metabolism of serin, alanine and glycerine aldehyde 3 phosphate in Salmonellae (Dandekar et al., 2012).

Some of these pathway mutations have already been shown to be important for survival in macrophages in Listeria (Schauer et al., 2010).

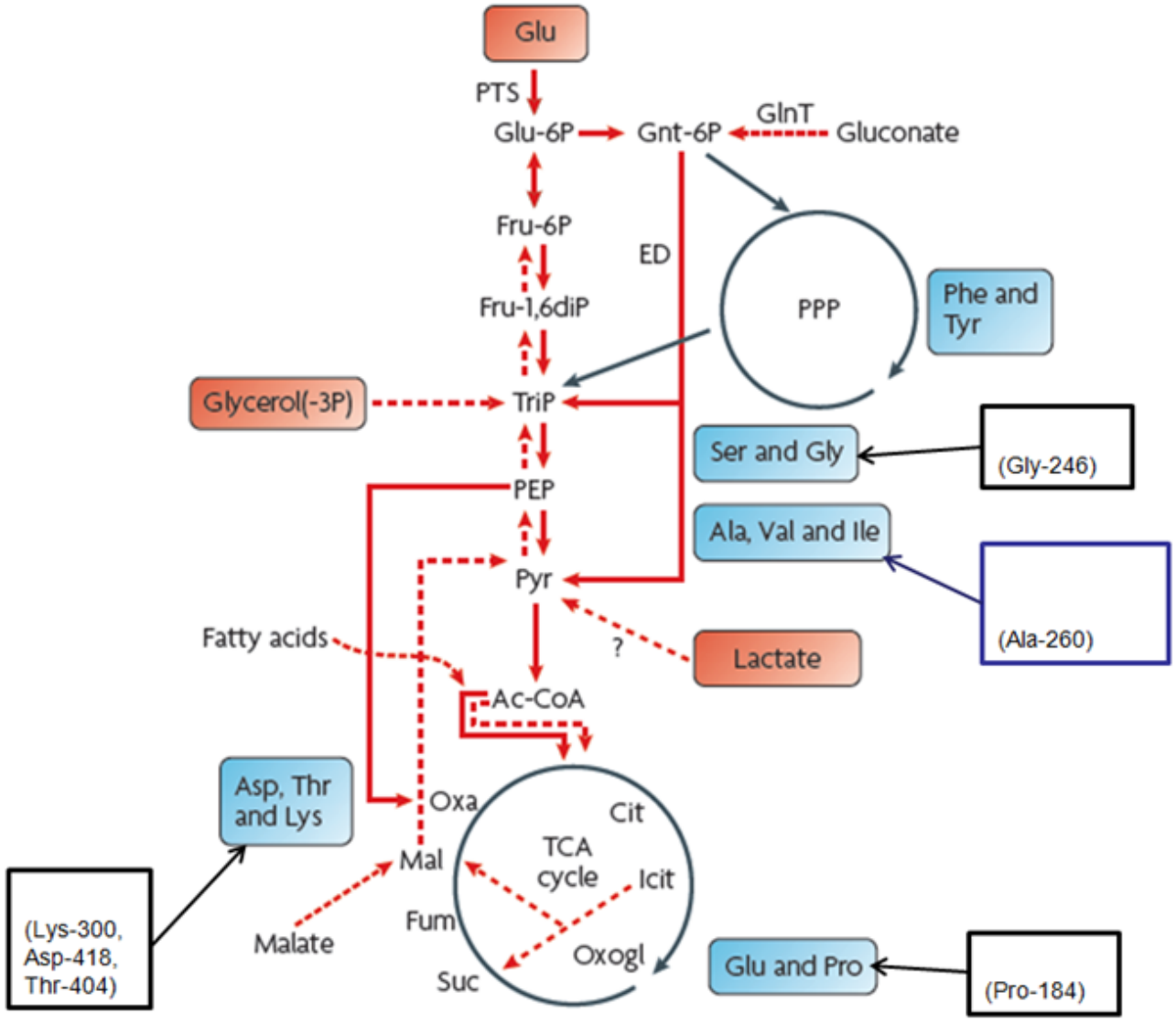


Figure 54: Salmonella Model.

Figure Legend 54. Starting from a basic model of Salmonella metabolism (Eisenreich et al., 2010) we show here current investigations and the position of the data examples given here on the simplified metabolic map of Salmonella.

There are many differences between LS-MIDA and Isotopo, with respect to the implemented mathematical methodology, design, implementation, usage and most of in the obtained richness and type

of results (see section 3 and 4 for details) processing experimental data examples from Salmonella isotopologue measurements.

Both software applications were tested with same data sets but with slightly different input because unlike Isotopo, LS-MIDA is not capable of processing actual and standard relative intensity values at the same time for mass to charge ratio values. ^{13}C labeled amino acid samples (analyzed as tert-butyldimethylsilyl (TBDMS) derivatives) were obtained from hydrolysates from cultures of Salmonella enterica grown in medium containing [U-$^{13}C6$] glucose (Dandekar et al., 2012). As an example, Table 1 and 8 (Column No. 1, Ala-260), and Figure 25 (LS-MIDA) and Figure 46 (Isotopo) show input parameters and experimental MS raw data of three fragments for TBDMS-alanine (Figure 55) such as metabolite information, m/e values of the relevant fragment, experimental intensity values, atomic mass values and the number of atoms in the fragment.

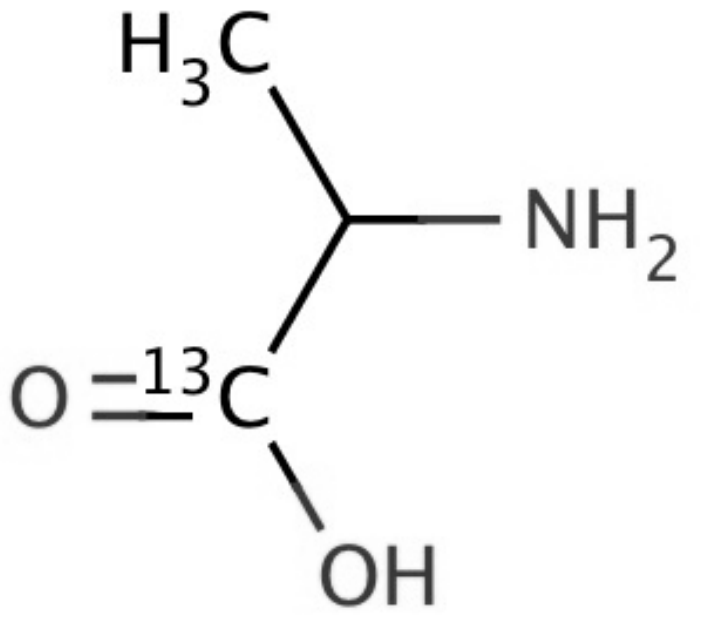


Figure 55: Alanin formula.

Figure Legend 55. Alanin (formular: CH3CH(NH2)COOH) as a non polar and non-essential amino acid used for natural and relative abundance prediction.

As the result of the measurements four different values are observed, provided in Table 8 (calculated Abundances using LS-MIDA) and Table 12 (Calculated Abundances using Isotopo). The estimated

absolute enrichment relative abundance by LS-MIDA is 29,9991531129648 % and by Isotopo is 82,8356248481561 % with the difference of 52,8364717351913 %. However the estimated absolute enrichment natural abundance using both software application is same (1,1%). During final output validation, estimated output using Isotopo proved to be more accurate than LS-MIDA.

Comparing LS-MIDA and Isotopo with some existing related software solutions, the provided features for primary data processing in both applications are not currently available in standard packages for metabolite modeling such as Metatool (Pfeiffer et al., 1999), Yanasquare (Schwarz et al., 2007), Gepasi (Mendez et al., 1993) or FiatFlux (Nicola et al., 2005; here fluxes are predicted after the isotopologue data have been processed).

There are alternative software solutions available for isotopologue data processing i.e. Envelop (Sykes et al., 2008) and Isotope Pattern Calculator (Kamalrudin et al., 2008), but none uses the binomial expression for data extension. Some currently available approaches i.e. PyMS (Callaghan et al., 2012), MSFACTs (Duran et al., 2003), AMDIS (Stein, 1999), TagFinder (Luedemann et al., 2008) and mMass (Strohalm et al., 2010) for raw data analysis observed during GC-MS experimentation but all completely differs in methodology to our solution and predictions using Isotopo seemed to be the more optimal depending upon the current observed, validated and presented results. Furthermore compared to the all mentioned software applications, Isotopo is the only software with an independent file based management system for experimental metabolic mass isotopomers raw data and with the most user friendly graphical interface for data analysis with fast processing speed (only seconds), can easily be deployed, learned and used.

The implementation of least square method with the inclusion of binomial expression, however, allows our isotopologue data to be more precisely calculated and Isotopo is in this definition a clear improvement over LS-MIDA. In future we are looking forward to enhance the abilities of Isotopo with the implementation of additional software features and will provide a new isotopomers database for online web and desktop based experimental isotopomers data management.

On a wider perspective these algorithms after applied in our shades on metabolism during bacterial infection. Research on microbial metabolism has great implications to human health. In particular research on bacterial metabolism holds great prospects in establishing the connection between pathogenesis, disease occurrence, and disease management. A complete picture of bacterial metabolism

helps in investigating and localizing the factors leading to the development of disease. Pathogens must be able to acquire nutrients and replicate successfully to invade their host. However the events that determine the dynamics of bacterial proliferation, distribution and incidence of disease are poorly understood and the data obtained from their study not well managed.

Metabolic modelling using LS-MIDA and Isotopo, leads to advances in the research on bacterial metabolism, life cycle in infection and highlight the multifactorial nature of infections they cause. Listeria monocytogenes is a bacterium that is capable of making the transition from a soil saprophyte into a virulent pathogen following its ingestion by susceptible humans. Salmonella causes extraintestinal infections, particularly blood stream infections utilizing the blood substrate. The pathogenicity of Staphylococci is related to acquisition of novel genetic elements that make it resistant and altered bacterial phenotypes that provide diversity in its virulence. Deciphering the intracellular metabolism of these pathogens can provide a link to specific virulence factors in the series of the events leading to the onset of disease. We believe that profound understanding of these phenomena with the help of well managed data warehousing is a prerequisite for the development of truly novel targeted preventive regimes.

6.3 Health Care Data Management

After molecular cascade and isotopologue data analysis, using previous PDM knowledge, this research lead to a new and exciting approach towards health care management system; patient life cycle management system in cardiovascular disease. The conducted research demonstrates a novel PDM application: product data management applied to molecular cascades for platelets and cardiovascular disease. This allows a boost for scientifically study, manipulate, test and describe such molecular cascades while getting in addition the typical benefits from PDM such as quality control, reliability analysis, quality management as well as integration of different data sources, cascades, experiments and experimental data. Furthermore, the impact will be even higher for clinical treatment. PDM treats here individualized management of cascades for optimal patient treatment and the aim for individualized medicine. It is thus expected that our interdisciplinary approach will boost not only PDM technologies but explore a new, large market for PDM solutions, both in bioinformatics and in health care.

Expected impacts for PDM development include demonstration of a novel, highly integrated data management, development of PDM-oriented, individual patient data processing in clinic as well as in research as a basis for individualized medicine. We will bring new PDM solutions to a new and large market. This will be made possible by new technological advances in computing and man-machine interfaces. We will bring new solutions to process large-scale data; the intended system development has high chances for success including a very careful and constantly validated plan for its scale-up; We chose our example, key cascades in thrombosis and hemostasis, with the idea to focus on a key problem in European health. Cardiovascular diseases are the major causes of death world-wide (about 70% of all death causes) but with an ageing population and a steady increase of this population segment in Europe this is becoming one of the largest areas where more investment into health care should go in Europe. This applies in particular regarding prevention and rapid, individualized medical treatment regarding involved pathophysiological cascades.

Global clinical impact: The proportion of aging people will grow rapidly within the near future. There will be fewer people to pay for the health and social care of the quickly aging population, furthermore, advancing in medical science will raise a new bar for quality in terms of the outcomes achieved in the treatment of illness. An aging population and rising public expectations will produce an increase in costs and will impede timely access to care, thus jeopardizing sustainability. EU National Health systems are an example of collapsing system: the aging EU workforce affects not only patient demographics but also the availability of clinicians. The upcoming wave of retiring health care professionals will occur at just the time when Italy will need more of them. To overcome these challenges, our health care systems will have to use its resources more effectively. Access and equity remain necessary characteristics of our health care system, but they are insufficient goals in terms of improving quality and achieving financial sustainability. The present point of view considers the treatments and the health service performances as the main objective and the final product of an efficient National Health System. They are the price of common wishful thinking about our health care service performances. Believing that treatments are always effective, that prevention always works are profoundly seductive to patients, who wish to benefit; to physicians, who want to be helpful; to politicians, who want the health service to be cost effective; and to the whole of the medical industrial complex, whose technologies and pharmaceuticals drive enormous profits. An example of this overoptimistic point of view is stroke medicine where the benefits of treatment are exaggerated and wishful thinking often trumps the published evidence. The concentration

on the hyper-acute care along with wishful thinking is having and has had a high price in stroke medicine without a true clinical benefit. The futility of this approach is due to a fragmented intervention instead of an integrated approach.

Specific impact: The early spin-off of this PDM approach in clinical practice will give value for the patient involves the full cycle of care, not just the outcome of a single intervention. In this process the "good" or product is health, not treatment. Better health is inherently less expensive than poor health and higher quality is often less costly than a low quality care. It was assumed that the quality of health care was good enough, and seen the problem as cost. However, the experience of the last decades has shown that more the focus is on driving down costs, the more costs tend to go up. This is because efforts to control costs often degenerate to cost shifting and eliminating discretionary but this monitoring services based approach has introduced only major inefficiencies in the system. In the long run, when the interventions on health status of a population are evaluated, the value of these interventions is obviously eroded.

The best way to reduce cost is to drive improvements in quality, but a true quality as measured by results. There is ample evidence that better quality care enables improved efficiency. So, more than in perhaps any other sector, better quality inherently reduces costs. Prevention and disease management cost less than acute treatment and rehabilitation. Treatment earlier in the causal chain is less costly. Getting the diagnosis right is more efficient than failed or unnecessary treatment. Fewer mistakes and complications cost less. Less invasive treatments enable less expensive recovery. Getting the right form of treatment to the right patients reduces the costs of failed or ineffective therapies. Faster recovery is less expensive than convalescence. Less disability means less long-term care. In synthesis, the PDM goal is to increase value and not just contain costs. Value is measured by the overall health outcome achieved relative to the total costs of care over the full cycle of the patient's illnesses. Value-based health care delivery seeks to minimize the overall cost of care, not focus just on minimizing the cost of individual services or interventions. Value-based care delivery spends more on appropriate services in order to save through early intervention, reducing mistakes, minimizing complications, and forestalling disease recurrences. This multidisciplinary approach underlines that care for a medical condition must integrate not only the activities directly related to disease, but also to related conditions, and defined for patient groups with similar needs. Care for a medical condition usually involves multiple specialties and numerous

interventions. Value for the patient is created not by any one intervention or specialty, but by the combined efforts of all of them. Patient outcomes will depend on a sequence of interventions often involving different sites and types of care — outpatient care, inpatient care, office visits, tests, rehabilitation, counseling, medications, procedures, and so on. The benefits of any one intervention for ultimate outcomes will depend on the effectiveness of other interventions throughout the care cycle.

7 Outlook and Conclusion

Computer Science approaches (software, database, management systems) are power tools to boost the research in metabolic modelling in infections as well as in health care management. Starting from a comparative analysis this thesis showed own steps and examples towards improvement in metabolic modelling software and health data management.

8 Bibliography

1. Ahmed. Z, Detlef. G (2007). Contributions of PDM Systems in Organizational Technical Data Management. In the proceedings of IEEE International Conference on Computer, Control & Communication, 12-13 November.

2. Ahmed. Z, Sudhir. G, Hans. K. (2008). Design Artifact's, Design Principles, Problems, Goals and Importance. In Proceedings 4th International Statistical Conference, 15, 57-68.

3. Ahmed. Z (2009). Proposing semantic oriented agent and knowledge base product data management. Information Management & Computer Security 17, 360–371, September 2009.

4. Ahmed. Z (2010a). Towards Performance Measurement and Metrics based Analysis of PLA Applications, International Journal of Software Engineering & Applications, 1(3): 66-80.

5. Ahmed. Z, Popov. V (2010b). Integration of Flexible RIA Based GUI in proceedings of I-SOAS. Innovative Production Machines and Systems - 6th I*PROMS Virtual Conference.

6. Ahmed. Z, Tacheva. I (2010c). Integration of Agile Ontology Mapping towards NLP Search in I-SOAS. Innovative Production Machines and Systems - 6th I*PROMS Virtual Conference.

7. Ahmed. Z. (2010d) Contributions to advance Product Data Management Systems (PDMs): Towards Flexible Graphical User Interface and Semantic Oriented Search for Web based PDMs. LAP Lambert Academic.

8. Ahmed. Z (2011a). I-HMI and LT-SOS in PDM Systems; Implementing AI Concepts for The Advancement of Multiple Role Based User Oriented Product Data Management Systems. VDM Verlag Dr. Müller.

9. Ahmed. Z (2011b). Designing Flexible GUI to Increase the Acceptance Rate of Product Data Management Systems in Industry. International Journal of Computer Science & Emerging Technologies 2, 100-109.

10. Ahmed. Z,, Saman M. (2011c). Middleware Technologies; Chain Web Grid Services. International Journal of Web Applications, 3, 197-205.

11. Ahmed. Z, Dandekar. T, Saman. M (2012a). Unified Modeling and HCI Mockup Designing towards MIDA. International Journal of Emerging Sciences, 2, 361-382.

12. Ahmed. Z. Saman. M., Dandekar. T (2012b). Formal UML Modelling of Isotopo, Bioinformatical Software for Mass Isotopomers Distribution Analysis, Software Engineering, Vol. 2 No. 5, Pages: 147-159, November 2012.

13. Ahmed. Z. Dandekar. T, Saman. M. (2012c). "ADAM: Transiting PDM into Clinical Patient Data Management". International Journal of Emerging Sciences, Vol. 2, No. 2, Pages: 280-299.

14. Ahmed. Z, Dandekar T., Saman M. (2012d). "Semantic web; Ontology Specific Languages for Web Application Development", International Journal of Web Applications, Vol. 4, No. 1, Pages: 33-41.

15. Ahmed. Z, Dandekar. T, Saman. M (2012e). Role of Ontology in NLP Grammar Construction for Semantic based Search Implementation in Product Data Management Systems, International Journals of Management, IT & Engineering, 2, 1-40.

16. Andreas. Er, Andreas. K, Michael. M (2008). DR10.5: Handbook for usage of the PDKM prototype version 2, PROMISE Consortium 2004-2008.

17. Ambler. SW (2005). The Elements of UML 2.0 Style, Cambridge University Press.

18. Bailey,J.E. (2001) Complex biology with no parameters. Nature Biotechnol,. 19, 503–504.

19. Baverel G., Conjard A., Chauvin M.F., Vercoutere B., Vittorelli A., Dubourg L., Gauthier C., Michoudet C., Durozard D., Martin G. (2003) Carbon 13 NMR spectroscopy: a powerful tool for studying renal metabolism. Biochimie. 85, 863–871.

20. Barrett,C.L., et al. (2009) Decomposing complex reaction networksusing random sampling, principal component analysis and basis rotation. BMC Syst Biol., 3, 30.

21. Berardi. D, Calvanese. D, Giacomo. GE (2005). Reasoning on UML class diagrams, Artif. Intell. 168(1): 70-118.

22. Becker,S. et al. (2007) Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox. Nat Prot., 2, 727–738.

23. Boyanova D, Nilla S, Birschmann I, Dandekar T, Dittrich M. (2012) PlateletWeb: A systems biological analysis of signaling networks in human platelets. Blood 119, 22-34.

24. Bequette .B. J, Sunny. N. E, El-Kadi. S. W. (2006) Owens. S. L. Application of stable isotopes and mass isotopomer distribution analysis to the study of intermediary metabolism of nutrients. J. Anim. Sci.. 84,E50-E59.

25. Bonello L, Tantry US, Marcucci R, Blindt R, Angiolillo DJ, Becker R, Bhatt DL, Cattaneo M, Collet JP, Cuisset T, Gachet C, Montalescot G, Jennings LK, Kereiakes D, Sibbing D, Trenk D, Van Werkum JW, Paganelli F, Price MJ, Waksman R, Gurbel PA. (2010) Consensus and future directions on the definition of high on-treatment platelet reactivity to adenosine diphosphate. J Am Coll Cardiol. 56, 919-33.

26. Bonarius. HP, Timmerarends. B, de Gooijer. CD, Tramper. J (1998). Metabolite-balancing techniques vs. 13C tracer experiments to determine metabolic fluxes in hybridoma cells. Biotechnol Bioeng. 5,258-262.

27. Bonarius, H. P. J., De Gooijer, C. D., Tramper, J., Schmid, G. (1995). Determination of the respiration quotient in mammalian cell culture in bicarbonate-buffered media. Biotechnol. Bioeng, 45, 524–535.

28. Brauman, J.I. (1966) Least Squares Analysis and Simplification of Multi-Isotope Mass Spectra. Anal. Chem. 38, 607–610.

29. Bruza. PD, van der Weide. T (1993). The Semantics of Data Flow Diagrams, in Proc. the International Conference on Management of Data.

30. Brenninkmeijer C.A., Janssen C., Kaiser J., Röckmann T., Rhee T.S., Assonov S.S. (2003) Isotope effects in the chemistry of atmospheric trace compounds. Chem. Rev. 103, 5125– 5161.

31. Byerley, L. O., S. Bassilian, E. A. Bergner, W. N. P. Lee. (1993) Use of D2O to quantitate cholesterol and fatty acid synthesis in tumor-bearing rats. FASEB J. 7, A289.

32. Centler,F. et al. (2010) A parallel algorithm to compute chemical organizations in biological networks. Bioinformatics, 26, 1788-9..

33. Chinkes, D. L., A. Aarsland, J. Rosenblatt, R. R. Wolfe. (1996) Comparison of mass isotopomer dilution methods used to calculate VLDL production in vivo. Am. J. Physiol. 271 (Endocrinol. Metab. 34), E373–E383.

34. Christensen B., Nielsen J. (1999) Isotopomer analysis using GC-MS. Metabolic Engineering 1, E8–16.

35. Christoph,G. et al. (2007) METANNOGEN: compiling features of biochemical reactions needed for the reconstruction of metabolic networks. BMC Systems Biology., 1, 5.

36. Claudia. SD, Jens. K, Achim. M (2007). Ways out of the mechatronics jungle. CONTACT Software CADplus Business+Engineering, 2:54-57Collet JP, Montalescot

G. (2009) Platelet function testing and implications for clinical practice. J Cardiovasc Pharmacol Ther. 14, 157-69.

37. Codd EF (1971). Further normalization of the data base relational model. IBM Research Report, San Jose, California, RJ909.

38. Codd. EF (1970). A relational model of data for large shared data banks. Communications of the ACM. 13(6): 377–387.

39. Cvijovic,M. et al. (2010) BioMet Toolbox: genome-wide analysis of metabolism. Nucleic Acids Research,. 38, 144-149.

40. Dandekar. T, Saman. M, Fieselmann. A, Ahmed. Z (2012a) Software Applications toward Quantitative Metabolic Flux Analysis and Visualization. Briefings in Bioinformatics, 13(6), 1-17.

41. Dandekar. T, Astrid F, Jasmin P, Hensel M (2012b). Salmonella enterica: a surprisingly well-adapted intracellular lifestyle. Front Microbiol. 3, 164.

42. Dauner M., Bailey J., Sauer U. (2001) Metabolic flux analysis with a comprehensive isotopomer model in Bacillus subtilis. Biotechnology and Bioengineering 76, 144–156.

43. Dauner M., Sauer U. (2000) GC-MS analysis of amino acids rapidly provides rich information for isotopomer balancing, Biotechnology Progress 16, 642–649.

44. Date. CJ (1986). An introduction to database system. Fourth edition, Addison Wesley.

45. Deepak,C. et al. (2009) TinkerCell: modular CAD tool for synthetic biology. Journal of Biological Engineering., 3, 19.

46. Dong-Yup,L. et al. (2006) WebCell: a web-based environment for kinetic modeling and dynamic simulation of cellular networks. Bioinformatics., 22, 1150-1151.

47. Dmitrij, F. et al. (2003) The PEDANT genome database. Nucleic Acids Res., 31, 207–211.

48. Duran. AL, Yang. J, Wang. L, Sumner. LW (2003). Metabolomics spectral formatting, alignment and conversion tools (MSFACTs). Bioinformatics. 19: 2283–2293.

49. Dugar, D., Stephanopoulos,G. (2011) Relative potential of biosynthetic pathways for biofuels and bio-based products. Nat Biotechnol., 29, 1074-1078.

50. Edwards,J.S. et al. (2001) In silico predictions of Escherichia coli metabolic capabilities are consistent with experimental data. Nat Biotechnol., 19,125-130.

51. Edwards,J.S. et al. (2002) Metabolic modelling of microbes: the flux-balance approach - (Minireview). Environmental Microbiology,.4, 133–140.

52. Eisenreich,W. et al. (2010) Carbon metabolism of intracellular bacterial pathogens and possible links to virulence. Nature reviews. Microbiology., 8, 401-412.

53. Eisenreich,W. et al. (2006) 13C isotopologue perturbation studies of Listeria monocytogenes carbon metabolism and its modulation by the virulence regulator PrfA. Proc Natl Acad Sci USA., 103, 2040–2050.

54. Emanuele,P. et al. (2008) FluXOR : Detecting and Monitoring Fast-Flux Service Networks. Detection of Intrusions and Malware and Vulnerability Assessment., 5137, 186-206.

55. Eylert,E. et al. (2010). Isotopologue profiling of Legionella pneumophila: role of serine and glucose as carbon substrates. The Journal of biological chemistry. 285, 22232-22243.

56. Fell, D. A., Small, J. A. (1986). Fat synthesis in adipose tissue. An examination of stoichiometric constraints. Biochem. J., 238, 781–786.

57. Flora,J.L. et al. (2012) GeneDB—an annotation database for pathogens. Nucl. Acids Res., 40,98-108.

58. Fischer,E. et al. (2004) High-throughput metabolic flux analysis based on gas chromatography-mass spectrometry derived 13C constraints. Analytical Biochemistry., 325, 308–316.

59. Fitzgerald R, Pirmohamed M. (2011) Aspirin resistance: effect of clinical, biochemical and genetic factors. Pharmacol Ther. 130, 213-25.

60. Gaasterland,T., Sensen, C.W (1996) Fully automated genome analysis that reflects user needs and preferences. A detailed introduction to the MAGPIE system architecture. Biochimie., 78, 302–310.

61. Geiger J, Brandmann T, Hubertus K, Tjahjadi B, Schinzel R, Walter U. (2010) A protein phosphorylation-based assay for screening and monitoring of drugs modulating cyclic nucleotide pathways. Anal Biochem. 407, 261-269.

62. Geiger J, Teichmann L, Grossmann R, Aktas B, Steigerwald U, Walter U, Schinzel R. (2005) Monitoring of clopidogrel action: comparison of methods. Clin Chem 51, 957-965.

63. Grant. ES, Chennamaneni. R, Reza. H (2006). Towards analyzing UML class diagram models to object-relational database systems transformations. in Proc. 24th IASTED international conference on Database and applications.

64. Grube MM, Dohle C, Djouchadar D, Rech P, Bienek K, Dietz-Fricke U, Jöbges M, Kohler M, Missala I, Schönherr B, Werner C, Zeytountchian H, Wissel J, Heuschmann PU. (2012) Evidence-Based Quality Indicators Stroke. 43, 142-6.

65. Hellerstein M.K., Neese R.A. (1999) Mass isotopomer distribution analysis at eight years: theoretical, analytic, and experimental considerations. Am J Physiol. Jun; 276 (6 Pt 1), E1146-70.

66. Herbert,M.S. (1993) SCAMP: a general-purpose simulator and metabolic control analysis program. Comput Appl Biosci., 9, 441-450.

67. Henning,S. Mats,J. (2005) Systems Biology Toolbox for MATLAB: a computational platform for research in systems biology Bioinformatics., 22, 514-515.

68. Hellerstein, M. K., R. Neese. (1992) Mass isotopomer distribution analysis: a technique for measuring biosynthesis and turnover of polymers. Am. J. Physiol. 263 (Endocrinol. Metab. 26), E988–E1001.

69. Hurlebaus,J. et al. (2002) MMT--a pathway modeling tool for data from rapid sampling experiments. Silico Biol., 2, 467-84.

70. Huang. MY, Lin. YJ, Hu-Xu (2004). A framework for web-based product data management using J2EE. The International Journal of Advanced Manufacturing Technology, 24(11-12):847-852.

71. Ikram MK, Sim X, Jensen RA, Cotch MF, Hewitt AW et al. (2010) Four novel Loci (19q13, 6q24, 12q24, and 5q14) influence the microcirculation in vivo. PLoS Genet. 6, 1001184.

72. Jacobson. I, Christerson. M., Jonsson. P, Övergaard. G (1992). Object-Oriented Software Engineering: A Use Case Driven Approach, Reading, MA: Addison-Wesley.

73. Jennings M.E., Matthews D.E. (2005) Determination of complex isotopomer patterns in isotopically labeled compounds by mass spectrometry. Analytical chemistry 77, 6435-6444.

74. Junker,B.H. et al. (2006). VANTED: a system for advanced data analysis and visualization in the context of biological networks. BMC Bioinformatics., 7, 1-13.

75. Kamp,A.V. et al. (2006) Metatool 5.0: fast and flexible elementary modes analysis. Systems biology., 22, 1930–1931.

76. Kaleta,C. et al. (2009) Can the whole be less than the sum of its parts? Pathway analysis in genome-scale metabolic networks using elementary flux patterns. Genome Res., 19,1872-1883.

77. Kanehisa,M. et al. (2008) KEGG for linking genomes to life and the environment. Nucleic Acids Research., 36: 480–484.

78. Kanehisa,M. (2002) The KEGG database. Novartis Found Symp., 247, 91-101

79. Kauffman,K.J. et al. (2003) Advances in flux balance analysis. Current Opinion in Biotechnology,. 14:491–496.

80. Kamalrudin. M, Cheng. SH (2008). Aziz A.A.; Sulong M.S. Reinforcing the concept of calculating isotope pattern using theoretical isotope generator (TIG). WSEAS Transactions on Information Science and Applications. 5: 949-958.

81. Kent. W (1983). A Simple Guide to Five Normal Forms in Relational Database Theory, Communications of the ACM. 26 (2):120-125.

82. Korzekwa, K., Howald, W. N., Trager W. N. (1990) The Use of Brauman's Least Squares Approach for the Quantification of Deuterated Chlorophenols. Biomed. Environ. Mass Spectrom. 19, 211-217.

83. Konstantinos, M. (2009) KEGGconverter: a tool for the in-silico modelling of metabolic networks of the KEGG Pathways database. BMC Bioinformatics., 10, 324.

84. Lake-Ee,Q. et al. (2009) OpenFLUX: efficient modelling software for 13C-based metabolic flux analysis. Microbial Cell Factories, 8, 25.

85. Lanktree MB, Hassell RG, Lahiry P, Hegele R.A. (2010) Phenomics: expanding the role of clinical evaluation in genomic studies.J Investig Med. 58, 700-6.

86. Latronico. E, Koopman. P (2001). Representing Embedded System Sequence Diagrams as a Formal Language, in Proc. 4th International Conference on The Unified Modeling Language, Modeling Languages, Concepts, and Tools.

87. Lee,G. et al. (2007) E2D: A Novel Tool for Annotating Protein Domains in Expressed Sequence Tags. CIBCBIEEE., 1-6.

88. Lee W.N., Byerley L.O., Bergner E.A., Edmond J. (1991) Mass isotopomer analysis: theoretical and practical considerations. Biol Mass Spectrom, 20, 451-458.

89. Lee, W. N. P., E. A. Bergner, Z. K. Guo. (1992) Mass isotopomer pattern and precursor-product relationship. Biol. Mass. Spectrom. 21: 114–122.

90. Letunic,I. et al. (2002) Recent improvements to the SMART domain-based sequence annotation resource. Nucleic Acids Res., 30, 242-244.

91. Liang C. et al. (2011) Staphylococcus aureus physiological growth limitations: Insights from flux calculations built on proteomics and external metabolite data. Proteomics., 10,1915-35.

92. Liang,C. et al. (2009) GENOVA: A rapid genome visualization and functional genomics software. Online Journal of Bioinformatics., 10, 201-217, 2009.

93. Liang,C., Dandekar T. (2006) inGeno – an integrated genome and ortholog viewer for improved genome to genome comparisons. BMC Bioinformatics., 7, 461.

94. Liang,C. et al. (2009) JANE: efficient mapping of prokaryotic ESTs and variable length sequence reads on related template genomes. BMC Bioinformatics., 10, 391.

95. Liang,C. et al. (2011) Staphylococcus aureus physiological growth limitations: insights from flux calculations built on proteomics and external metabolite data. Proteomics., 11, 1915-35.

96. Luedemann. A, Strassburg. K, Erban. A, Kopka. J (2008). TagFinder for the quantitative analysis of gas chromatography–mass spectrometry (GC-MS)-based metabolite profiling experiments. Bioinformatics. 24: 732–737.

97. Michael. F, Donald. M, Stephen. F (2006). Architecture and Sizing Guide for HP Enterprise Servers. PTC Windchill, PDMLink R8.0, Hewlett-Packet Company.

98. Mindukshev IV, Gambaryan S, Rukoyatkina N, Schütz C, Krivchenko AI, Walter U, Geiger J. (2012) A novel aggregometry technique with improved sensitivity based on particle scatter analysis. Clin Chem Lab Med 50, 7.

99. Mahadevan,R. et al. (2002) Dynamic Flux Balance Analysis of Diauxic Growth in Escherichia coli, Biophysical Journal,. 83, 1331–1340.

100. Marilyn. B (1978).A guide for programmers. Prentice-Hall.

101. Massila K., Soong H C., Azlianor A A., Muhammad S. S. (2008) Reinforcing the concept of calculating isotope pattern using theoretical isotope generator (TIG). WSEAS Transactions on Information Science and Applications, 5. 949.

102. Maddika. S et al. (2007). Cell survival, cell death and cell cycle pathways are interconnected: Implications for cancer therapy. Drug Resistance Updates, 10, 13–29.

103. Medvidovic. N, Rosenblum. DS, Redmiles. DF, Robbins. JE (2002) Modeling software architectures in the Unified Modeling Language, ACM Trans. Softw. Eng. Methodol. 11(1):2-57.

104. Meyer,F. et al. (2003) GenDB--an open source genome annotation system for prokaryote genomes., 31,2187-95.

105. Mendes, P. (1993) GEPASI: A software package for modelling the dynamics, steady states and control of biochemical and other systems. CABIOS, 9, 563-571.

106. Michael T. Sykes, James R. Williamson. (2008) Envelope: interactive software for modeling and fitting complex isotope distributions. BMC Bioinformatics. 9, 446.

107. Nahnsen, S. et al. (2011) Probabilistic Consensus Scoring Improves Tandem Mass Spectrometry Peptide Identification. J. Proteome Res., 10, 3332-3343.

108. Nicola Z., Eliane. F., Uwe. S. (2005) FiatFlux – a software for metabolic flux analysis from 13C-glucose experiments. BMC Bioinformatics, 6, 209, 2005.

109. Oehm,S. et al. (2008) Comparative Pathway Analyzer--a web server for comparative analysis, clustering and visualization of metabolic networks in multiple organisms. Nucleic Acids Res., 36, 433-4337.

110. O'Callaghan. S, DeSouza. DP, Isaac. A, Wang. Q, et al. (2012). PyMS: a Python toolkit for processing of gas chromatography--mass spectrometry (GC-MS) data. Application and comparative study of selected tools. BMC Bioinformatics. 13, 115.

111. Ovbiagele B, Diener HC, Yusuf S, Martin RH, Cotton D, Vinisko R, Donnan G.A., Bath P.M. (2011) PROFESS Investigators. Level of systolic blood pressure within the normal range and risk of recurrent stroke. JAMA. 306,2137-44.

112. Ohl. ME, Miller. SI. (2001). SALMONELLA: AModel for Bacterial Pathogenesis. Annu. Rev. Med., 52,259–274.

113. Papin,J.A. et al. (2004) Comparison of network-based pathway analysis methods. Trends Biotechnol., 22, 400–405.

114. Papageorgopoulos C., Caldwell K., Shackleton C., Schweingrubber H., Hellerstein M.K. (1999) Measuring Protein Synthesis by Mass Isotopomer Distribution Analysis (MIDA). Analytical Biochemistry 267, 1–16.

115. Pammolli F, Riccaboni M, Magazzini L. (2011) The sustainability of European health care systems: beyond income and aging. Eur J Health Econ. 52.

116. Pathak J, Wang J, Kashyap S, Basford M, Li R, Masys DR, Chute CG. (2011) Mapping clinical phenotype data elements to standardized metadata repositories and controlled terminologies: the eMERGE Network experience. J Am Med Inform Assoc. 18, 376-86.

117. Peng J, Babaguchi N, Luo H, Gao Y, Fan J. (2010) Constructing distributed Hippocratic video databases for privacy-preserving online patient training and counseling. IEEE Trans Inf Technol Biomed. 14, 1014-1026.

118. Perillo,J.R. et al. (2009) Daileon: A Tool for Enabling Domain Annotations. Proceedings of the Workshop on AOP and Meta-Data for Software Evolution., 1-4.

119. Petritsch B, Wendel F, Leyh RG, Frantz S (2011) The broken heart. Circulation 123, 2020-2021.

120. Peroni, O., V. Large, M. Beylot. (1995) Measuring gluconeogenesis with [2-13C] glycerol and mass isotopomer distribution analysis of glucose. Am. J. Physiol. 269 (Endocrinol. Metab. 32), E516–E523.

121. Pfeiffer, T., Sanchez Valdenebro, I., Nuno, J. C., Montero, F., Schuster, S. (1999) METATOOL: for studying metabolic networks. Bioinformatics, 15, 251-257, 1999.

122. Ponting,C.P. et al. (1999) SMART: identification and annotation of domains from signalling and extracellular protein sequences.Nucleic Acids Res., 27,229-232.

123. Pitkänen,E. et al. (2008) ReMatch: a web-based tool to construct, store and share stoichiometric metabolic models with carbon maps for metabolic flux analysis, Journal of Integrative Bioinformatics., 5,1-13.Previs, S. F., Fernandez C. A., Yang D., Soloviev M. V., France D., Brunengraber H. (1998) Limitations of the Mass Isotopomer Distribution Analysis of Glucose to Study Gluconeogenesis. J. Biol. Chem. 277: 1998.

124. Price,N.D. et al. (2004) Uniform sampling of steady-state flux spaces: means to design experiments and to interpret enzymopathies. Biophys J., 87, 2172-86.

125. Rahman,S. and Schomburg, D. (2006) Observing local and global properties of metabolic pathways: 'load points' and 'choke points' in the metabolic networks. Bioinformatics, 22, 1767-1774.

126. Rantanen A., Rousu. J, Ketola R.A., Kokkonen J.T., Tarkiainen V. (2002) Computing positional isotopomer distributions from tandem mass spectrometric data. Metabolic Engineering 4, 285–294.

127. Rohn,H., Klukas, C., Schreiber, F. (2011) Creating Views on Integrated Multidomain Data; Bioinformatics, 27, 1839-1845.

128. Rousu J.A., Rantanen R.A., Ketola C., Juha T., Kokkonen C. (2005) Isotopomer distribution computation from tandem mass spectrometric data with overlapping fragment spectra. Spectroscopy 19, 53–67.

129. Roman,L.T. et al. (2000) The COG database: a tool for genome-scale analysis of protein functions and evolution. Nucleic Acids Res. 28, 33–36.

130. Sauro,H.M. (2000) Jarnac: a system for interactive metabolic analysis. Proceedings of the 9th International Meeting on BioThermoKinetics Stellenbosch University Press., 221–228

131. Savinell, J. M., Palsson, B. O. (1992). Network analysis of intermediary metabolism using linear optimization. J. Theor. Biol., 154, 421–473.

132. Schuster,S. et al. (2002) Reaction routes in biochemical reaction systems: algebraic properties, validated calculation procedure and example from nucleotide metabolism. J. Math. Biol., 45, 153–181.

133. Schwarz,R. et al. (2007) Integrated network reconstruction, visualization and analysis using YANAsquare. BMC Bioinformatics,. 8. 1-10.

134. Schuster,S. et al. (2000) A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks. Nat. Biotechnol., 18, 326-332.

135. Schauer K, Geginat G, Liang C, Goebel W, Dandekar T, Fuchs TM. Deciphering the intracellular metabolism of Listeria monocytogenes by mutant screening and modelling. BMC Genomics. 2010 Oct 18;11:573.

136. Schilling,C.H. et al. (1999) Metabolic pathway analysis: basic concepts and scientific applications in the post-genomic era.. Biotechnology progress., 15, 296-303.

137. Schwarz R. et al. (2009) Detecting species-site dependencies in large multiple sequence alignments. Nucleic Acids Res., 37, 5959-68.

138. Schwarz,R. et al. (2005) YANA - a software tool for analyzing flux modes, gene-expression and enzyme activities. BMC Bioinformatics., 6, 135-146.

139. Schäuble,S. et al. (2011) Hands-on metabolism analysis of complex biochemical networks using elementary flux modes. Methods Enzymol., 500, 437-456.

140. Schuster,R., Schuster,S. (1993) Refined algorithm and computer program for calculating all non-negative fluxes admissible in steady states of biochemical reaction systems with or without some flux rates fixed. Comput Appl Biosci., 9, 79-85.

141. Schwarz, R., Liang, C., Kaleta, C., Kuhnel, M., Hoffmann, E., Kuznetsov, S., Hecker, M., Griffith, G., Schuster, S., Dandekar, T. (2007) Integrated network recon-struction, visualization and analysis using YANAsquare. BMC Bioinformatics, 8, 313 (10 pp.).

142. Snoep, J.L. et al. (2002) Java Web Simulation (JWS): a web based database of kinetic models. Molecular Biology Reports., 29,259–263

143. Silberschatz. K, Sudarshan. S (2006). Database system Concepts. McGraw Hill international edition, Fifth edition..

144. Shulman, R. G., Brown, T. R., Ugurbil, S., Ogawa, S., Cohen, S. M., Den Hollander, J. A. (1979). Cellular applications of 31P and 13C NMR. Science, 205, 160–166.

145. Szklarczyk,D. et al. (2011) The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. Nucleic Acids Res., 39, 561-568.

146. Steffen,K., Axel,V.K. (2002) An application programming interface for CellNetAnalyzer. Biosystems., 105, 162-8.

147. Stefan,H. et al. (2006) COPASI—a COmplex PAthway SImulator. Bioinformatics., 22 ,3067-3074.

148. Steele. R. (1959) Influences of glucose loading and of injected insulin on hepatic glucose output. Ann. NY Acad. Sci. 82: 420-430.

149. Stein. SE (1999). An integrated method for spectrum extraction and compound identification from gas chromatography/mass spectrometry data. J. Am. Soc. Mass. Spectrom. 10: 770–781.

150. Strohalm. M, Kavan. D, Novak. P, Volny. M, Havlicek. V (2010). mMass 3: a cross-platform software environment for precise analysis of mass spectrometric data. Anal. Chem. , 82: 4648–4651

151. Sung. CS, Sam. JP (2007). A component-based product data management system. The International Journal of Advanced Manufacturing Technology. 33 (5): 614-626.

152. Sykes. MT, Williamson. JR (2008). Envelope: interactive software for modeling and fitting complex isotope distributions. BMC Bioinformatics, 9, 446.

153. Szent-Gyorgyi LE, Coblyn J, Turchin A, Loscalzo J, Kachalia A. (2011) Building a departmental quality program: a patient-based and provider-led approach. Acad Med. 286, 314-20.

154. Tim Berners-Lee (2001a). Hendler, Publishing on the Semantic Web. Nature 2001, 1023-1025.

155. Tim Berners-Lee, James Hendler, Ora Lassila, (2001b). The Semantic Web. Scientific American 2001, 29-37.

156. Urbanczik,R. et al. (2005) An improved algorithm for stoichiometric network analysis: theory and applications. Bioinformatics., 21, 1203-1210.

157. Van Gulik, W. M., J. J. Heijnen. (1995). A metabolic network stoichiometry analysis of microbial growth and product formation. Biotechnol. Bioeng., 48, 681–698.

158. Vallino, J. J., Stephanopoulos, G. (1990). Flux determination in cellular bioreaction networks: Applications to lysine fermentations. Frontiers in bioprocessing, 205–219.

159. Wiechert,W. et al. (1997) Bidirectional reaction steps in metabolic networks: I. Modeling and simulation of carbon isotope labeling experiments. Biotechnol Bioeng,. 55, 101-17.

160. Wiback,S.J. et al. (2004) Using metabolic flux data to further constrain the metabolic solution space and predict internal flux patterns: the Escherichia coli spectrum., Biotechnology and bioengineering., 86, 317-31.

161. Wolfe, R. R. (1984) Tracers in Metabolic Research. Radio-Isotope and Stable Isotope/Mass Spectrometric Methods. New York: Liss, 1984, p. 119–144.

162. Wangorsch G, Butt E, Mark R, Hubertus K, Geiger J, Dandekar T, Dittrich M. (2011) Time-resolved in silico modeling of fine-tuned cAMP signaling in platelets: feedback

loops, titrated phosphorylations and pharmacological modulation. BMC Syst Biol. 28, 178.

163. Winden W. V, Wittman C., Heinzle E., Heijnen J. (2002) Correcting mass isotopomer distributions for naturally occurring isotopes, Biotechnology and Bioengineering 80, 477–479.

164. Wiechert,W. (2002) An introduction to 13C metabolic flux analysis. Genetic Engineering., 24, 215–238.

165. Zamboni, N. et al. (2005) FiatFlux – a software for metabolic flux analysis from 13C-glucose experiments. BMC Bioinformatics., 6,1-8.

166. Zamboni,N. (2007) FiatFlux 1.6X – Getting started, Institute of Molecular Systems Biology ETH Zuerich, December 2007

167. Zilversmit, D. B. (1960) The design and analysis of isotope experiments. Am. J. Med. 29: 832–848, 1960.

168. Zilversmit. D. B., C. Entenman, M. Fishler. (1943) On The Calculation of "Turnover Time" and "Turnover Rate" from Experiments Involving the use of Labeling Agents. J Gen Physiol. 20; 26(3): 325–331.

169. Zur. H, Ruppin. E, Shlomi. T (2010). iMAT: an integrative metabolic analysis tool. Bioinformatics. 26: 3140-3142.

9 Nomenclatures

Abundance Matrix (An)

Abundance Matrix (MRMn)

Allele specific oligonucleotide (ASO)

Alanine (Ala)

Aspartic Acid (Asp)

Computer Aided Design (CAD)

Constraint Based Modeling (CBM)

Classical Flux Balance Analysis (CFBA)

Classical and Dynamic Flux Balance Analysis (CDFBA)

Citric Acid Cycle (TCA)

Common Data Form (CDF)

Deoxyribonucleic Acid (DNA)

Data Flow Diagram (DFD)

Dynamic Flux Balance (DFBA)

Dynamic Optimization Approach (DOA)

Elementary Flux Modes (EFMs)

European Union (EU)

Escherichia coli (E.coli)

Flux Balance Analysis (FBA)

FiatFlux (FF)

Fractional Molar Abundance (FRMn)

Gas Chromatography- Mass Spectrometry (GC-MS)

Glycine (Gly)

Graph Modelling Language (GML)

Graphical User Interface (GUI)

Health Care Management (HCM)

Health Care Data Management (HCDM)

Human Computer Interaction (HCI)

Human Machine Interface (HMI)

Information and Communications Technology (ICT)

Input/Output (I/O)

Inverse Transpose Abundance Matrix (ITAn)

Kyoto Encyclopedia of Genes and Genomes (KEGG)

Kyoto Encyclopedia of Genes and Genomes Browser (KEGGbrowser)

Least Square (LS)

Least Square Mass Isotopomers Analyzer (LS-MIDA)

Lysine (Lys)

MDVα (mass isotopomer distribution vector)

Mass to charge ratio (m/z)

Mass to charge ratio (m/e)

Mass Isotopomer Distribution Analysis (MIDA)

Mass Spectrometry (MS)

Metabolic Flux Analysis (MFA)

Minimum Values (MinVal)

Multiple Document Interface (MDI)

Natural Abundance (Mn)

Network Common Data Form (netDFC)

New Relative Abundance (RMRMn)

New Abundance matrix (MRMn)

New Fractional Molar Abundance values (FRMRMn)

New Minimum Values (2ndMinVal

Nuclear Magnetic Resonance (NMR)

Web Ontology Language (OWL)

Patient Guidance Services (PGS)

Patient Life Cycle Management (PLCM),

Product Lifecycle Management (PLM)

Personal Health System (PHS)

Phenotype phase plane (PhPP)

Phosphotransferase System (PTS)

Product Data Management (PDM)

Proline (Pro

Polymerase chain reaction (PCR)

Relative Abundances (Ra)

Relational Database Management System (RDBMS)

Resource Description Framework (RDF)

Relative Abundance (RMn)

Relative Natural Abundances (Na)

Relative Intensity (Ri)

Rich Internet Application (RIA)

Software for Biological Experimental Data Analysis (SBEDA)

Static Optimization Approach (SOA)

Systems Biology Markup Language (SBML)

System Sequence Diagram (SSD)

Threonine (Thr)

Time rate of change (d/dt)

Transpose Abundance Matrix (TAn)

Tert-butyldimethylsilyl (TBDMS)

Unified Modeling Language (UML)

Work Package (WP)

eXtensible Mark-up Language (XML).

10 Appendix

.

CURRICULUM VITAE: ZEESHAN AHMED

Date of Birth: 15 January 1983, Gujrat Pakistan. (29 Years old)
Marital Status: Married (to Saman Zeeshan: Bioinformatician), have one son (Jibrael Zeeshan)
Mobile: +49 (0) 151 57897184, Email: zeeshan@zeeshanahmed.info

I Mr. Zeeshan Ahmed, 29 years old, with more than 14 years of university education, have more than 4 years of industrial work experience and more than 6 years of academic interdisciplinary research and software development experience. I have written some publications only as the First/Main author including Books, Research Monographs, Book Chapters, Journals Manuscripts, Scientific Transactions, Conference Papers, Workshop Papers, Poster Papers, Scientific Articles, Thesis, Project Reports, Project Manuals and Seminar Papers & Presentations. I am the Editorial Board Member, Referee of some International journal publishers.

Furthermore, I have achieved some university distinctions, awarded some national and international prizes (including research, programming and talent).

Education

University Education

Date: 2010-2012 **Reg. No.:** 1728090 Doctoral Research	***Doctoral Research***- *at Julius-Maximilian's University of Würzburg; Faculty of Biology.* *Department of Bioinformatics, Biocenter.* • PhD Thesis: ***Bioinformatics Software for Metabolic and Health Care Data Management*** • Major Subjects: *Metabolic flux analysis, Mass Isotopomer Data Distribution Analysis, Health Informatics, Data Management (DM), Web technologies, Database, Mathematics and Software Engineering.* • Research Supervisors: *Prof. Dr. Thomas Dandekar (Uni. WÜ), PD. Dr. Wolfgang Eisenreich (TU Munich)*
Date: 2008-2010 **Reg. No.:** 0725085 Doctoral Research	***Doctoral Research***- *at Vienna University of Technology, Austria; Faculty of Mechanical and Industrial Engineering.* *Department of MIVP, Institute of Engineering Design and Logistics Engineering.* • Major Subjects: *Product Data Management, Human Machine Interaction, Semantic Web, Natural Language Processing, Information Technology and Software Engineering.* • *Holds distinction for completing PhD Course Work with highest possible grade "A" in all subjects- with 100% Marks.*
Date: 2005-2007 **Reg. No.:** 830115-P117 **Degree Duration**: 2 Years 2nd Master Degree	***Master of Science - Computer Science*** *at University of Blekinge Sweden; Faculty of Computer Science.* *Department of Artificial Intelligence, Blekinge Institute of Technology.* • Master Thesis: ***Integration of variants handling in M-system NT***. (Graded "A") • Major Subjects: *Applied Artificial Intelligence, Knowledge Engineering, Connectivity Software Technologies, Human Computer Interaction, Advance Topic in Computer and Information Science.* • *Holds distinction for completing course work earlier to the expected Record time of only one Semester, with 85% Marks (First Division), degree awarded in September 2006.*
Date: 2003-2005 **Reg. No.:** L1F03MSCS1023 **Degree Duration**: 2 Years 1st Master Degree	***Master of Science - Computer Science*** *at University of Central Punjab, Pakistan; Faculty of Information Technology.* *Department of Computer Science, Punjab Institute of Computer Science.* • Master thesis: ***Object Oriented Framework Design Patterns***. (Graded "A") • Major Subjects: *Advanced Software Engineering, Advanced Computer Architecture, Software Project Management, Data Warehousing, Software Quality Assurance and Testing.* • *Holds distinction for completing course work including final thesis earlier to the expected time with 82.3% Marks (First Division), awarded in April 2005.*
Date: 2000-2004 **Reg. No.:** L1F00BSCS0093 **Degree Duration**: 4 Years Bachelor Degree	***Bachelor of Science – Computer Science*** *at University of Central Punjab, Pakistan; Faculty of Information Technology.* *Department of Computer Science, Punjab Institute of Computer Science.* • Bachelor Thesis & Project: ***Smart House***. (Graded "A") • Major Subjects: *Programming Languages, Software Engineering and Artificial Intelligence.* • *Holds distinction for completing 4 years course work including final project earlier to the expected Record time of 2.8 Years (8/12 Semesters) with 75% Marks (First Division), degree awarded in July 2004.*

Higher Secondary and Secondary Degree Education

Date: 1998-2000 **Reg. No.:** 58781 **Degree Duration**: 2 Years	***Intermediate with Computer Science*** *Equivalent to A-Levels from Government Zamindar Degree College, The Board of Intermediate & Higher Secondary Education (BISE) Gujranwala, Pakistan* • *Holds distinction for completing Intermediate with First Division, degree awarded in April 2000.*
Date: 1996-1998 **Reg. No.:** 56594 **Degree Duration**: 2 Years	***Matriculation with Science Subjects*** *Equivalent to O-Levels from Grammar High School Gujrat, The Board of Intermediate & Secondary Education (BISE) Gujranwala, Pakistan* • *Holds distinction for completing Matriculation with Final Grade "A", degree awarded in July 1998.*

Research Training, Certification, Diploma Education

Date: 2006 **Training Duration**: 6 Months	***Master Research Thesis & Project*** *from Fraunhofer Institute for Experimental Software Engineering (IESE), Department Measurement, Prediction and Empiricism, Kaiserslautern Germany* • Started March 2006, completed and Certificate awarded in September 2006. *(Final Grade A)* • *Holds distinction for completing Research Thesis and Project with highest possible Grade "A".*
Date: 2001 **Cert. Duration**: 2 Months	***Certification in Visual C++*** *from Society of Advancement of Computer Science (SACS), Punjab Institute of Computer Science (PICS), University of Central Punjab, Lahore Pakistan* • Started 15th October 2001 and completed in 28th November 2001, awarded in December 2001.
Date: 1999-2000 **Diploma Duration**: 1 Year	***Diploma in Computer Application*** *from* Imperial College of Computer Science (ICCS), Gujrat Pakistan • Started August 1999, Completed with 100 % May 2000, awarded August 2000. • *Holds distinction for completing Diploma with highest possible Grade "A".*

Professional Work Experience

Scientific Software Engineer (Academia) **Date**: 07/2010 - On Job **Field**: Bioinformatics	***University of Wuerzburg Germany****; Prof. Thomas Dandekar's Functional Genomics and System Biology Group, Department of Bioinformatics, Biocenter.* • Providing services at DFG Projects: Z1- Bioinformatic tools, OMICs databases, WIKI and their integration. • Proposing and developing prototype software applications: *LS-MIDA, Isotopo, Isotopomer DB and ADAM*. • Analyzing mathematical algorithms and work on their optimization and implementation. • Doing technical documentation e.g. project reports, publications and user manuals etc. • Performing Scientific (Experimental) Data Analysis. • Gathering relevant information by software and document analysis.
Doctoral Researcher (Academia) **Date**: 12/2009 – 06/2010 **Field**: Bioinformatics	***University of Wuerzburg Germany****; Prof. Thomas Dandekar's Functional Genomics and System Biology Group, Department of Bioinformatics, Biocenter.* • Provided services as Researcher at DFG Project: Metabolism Bacterial Infections - SPP 1316. • Analyzed mathematical algorithms and worked on their optimization and implementation. • Gathered relevant information by software and document analysis. • Proposed and developed prototype software application: *DOC*. • Did technical documentation e.g. project reports, publications and user manuals etc.
University Assistant (Academia) **Date**: 10/2006-12/2009 **Field**: Machine Informatics	***Vienna University of Technology Austria****, Mechanical Engineering Informatics and Virtual Product Development, Institute for Engineering Design and Logistics Engineering.* • Provided services in defining and refining research objectives. • Analyzed and used appropriate methodologies to design and manage research projects. • Gathered relevant information including document analysis and case studies. • Prepared and reported findings. • Performed prototype rapid software development: I-SOAS, ZAJ, ZAC, VDI and PCM System. • Participated in lectures and labs of subject "Java Programming" and supervised Master, Bachelor and Diploma thesis.
Software Engineer (Industry & Academia) **Date**: 06/2006-10/2006 **Field**: 5D Visualization	***Fraunhofer Institute for Experimental Software Engineering, Germany*** • Analyzed and used appropriate methodologies. • Prepared and reported findings. • Provided services in research and development of a Five Dimensional Experimental and Statistical Data Visualization: 5D Visualizer.
Software Engineer (Industry) **Date**: 06/2004-07/2005 **Field**: Satellite and Broadcasting	***Wiz-Links*** *Pvt. Ltd Lahore Pakistan* • Analyzed and used appropriate methodologies. • Performed software technical documentation (Requirement Analysis, UML Design Modeling, and User Manuals). • Provided services in software design and development of different modules of the broadcasting solution and media management system for TV Channels: "Wiz-Air".
Software Engineer (Industry) **Date**: 01/2004-06/2004 **Field**: Digital Camera & DB	***NSE Technologies*** *Pvt. Ltd Lahore Pakistan* • Provided services in the design and development of an Enterprise System: "Oras Inventory Control System". • Worked as Quality Assurer for the testing of a special purpose product for image management using digital cameras: "I-Photo Grapher & Portal".
Lecturer(Academia) **Date**: 09/2009-12/2003	***College of management and information technology*** *Gujrat, Pakistan* • Delivered lectures of subjects: Artificial Intelligence, Database Management Systems, Advance computer concepts and Advance programming concepts, to graduates and under graduate students.

Internship; Software Developer (Industry) **Date**: 05/2003-09/2003 **Field**: Enterprise Applications	***Habib Rafiqe Technologies*** *Pvt. Ltd Lahore Pakistan* • Provided services in the development of architect software designs. • Provided services in the software development of an Enterprise Data Management System: "HR-DSL Inventory Control and Account System".
Lecturer(Academia) **Date**: 05/2002-08/08 *Part time job.*	***Imperial College of Computer Science*** *(ICCS) Gujrat, Pakistan* • Delivered lectures of subjects: C++ Programming language and Data Base Management Systems (Oracle 8), to under graduate students.
Teacher Assistant *(Academia)* **Date**: 05/2001-04/2002 *Part time job.*	***Punjab Institute of Computer Science*** *(PICS), University of Central Punjab, Lahore, Pakistan* • Delivered lectures of subjects: C++ Programming language and Data Base Management Systems (Oracle 8) to under graduates, as Assistant to Lecturer Sarfraz Bokhari.

Funded Research Projects

Project: ***SFB / Transregio 34, Z1: Bioinformatic tools, OMICs databases, a Staphylococcus aureus WIKI and their integration.***

- Supervisor: Prof. Dr. Thomas Dandekar
- Responsibilities: Programming, Research, Software Analysis
- Description**:** We will combine software and analysis algorithms to integrate data. Improved access to all data will exploit WIKI-like techniques for annotation, WEB services and enhanced large-content visualization algorithms.
- Organisation: Department of Bioinformatics, Biocenter, University of Wuerzburg Germany.

Project: ***Metabolism Bacterial Infections - SPP 1316, Deutsche Forschungsgemeinschaft (DFG).***

- Supervisor: Prof. Dr. Thomas Dandekar
- Responsibilities: Programming, Research, Software Analysis
- Description**:** This project is about to identify metabolic pathways that are important for the bacteria during infection and to determine the metabolic fluxes. The metabolic reactions of the host organisms and the genetic mechanisms of metabolic adaptation will be unraveled.
- Organisation: Department of Bioinformatics, Biocenter, University of Wuerzburg Germany

Academic & Industrial Interdisciplinary Software Projects

Project #18: ***Advance Product Data Analysis & Management -ADAM (2012)*** *(proposed by Zeeshan)*

- Organization: Department of Bioinformatics, University of Wuerzburg Germany.
- Responsibilities: Software Engineer (Research, Software Development and Technical Documentation)
- Description**:** This is published project, about to support the medical community with the proposition and implementation of an innovative initiative, addressing major challenges of providing optimal management of a certain diagnosis, working rapidly under emergency conditions, protecting personal data, coupling individual patient data with general repositories, allowing therapy monitoring, analyzing individual variations with the and incorporation of golden standard therapy guidelines.
- Technologies: Microsoft Dot Net, C#, MySQl, RIA, Flex, Ontology etc.

Project #17: ***Isotopomer Database (2012)*** *(proposed & developed by Zeeshan)*

- Organization: Department of Bioinformatics, University of Wuerzburg Germany.
- Responsibilities: Software Engineer (Software Analysis, Research, Development, Technical Documentation)
- Description**:** This software research project is about to design and implement a new database management system for biological experimental isotopomers data.
- Technologies: MySQL, Flex, C# etc.

Project #16: ***Isotopo (2010-2012)*** *(proposed & developed by Zeeshan)*

- Organization: Department of Bioinformatics, University of Wuerzburg Germany.
- Responsibilities: Software Engineer (Software Analysis, Research, Development, Technical Documentation)
- Description**:** This is published project, about to develop a software application to analyze isotopic (labeled) experimental data by implementing Mass Isotopomer Analysis to predict isotopic distribution. Furthermore it also provides the File based Data Management System for experimental data manipulation.
- Technologies: C# etc.

Project #15: ***LS-MIDA (2010-2012)*** *(proposed & developed by Zeeshan)*

- Organization: Department of Bioinformatics, University of Wuerzburg Germany.
- Responsibilities: Software Engineer (Software Analysis, Research , Development, Technical Documentation)
- Description**:** This is published project, about to develop a software application using Least Square Mass Isotopomer Analysis Method (J. H. Benyon 1960) to analyze distributed amount of isotopes in a compound by measuring intensities and generating MxN Abudance Matrix of each isotopic species at each Mass. Furthermore it also provides the File based Data Management System for experimental data manipulation.
- Technologies: C# etc.

Project #14: ***Optimal Data Classifier (2010-2011)*** *(proposed & developed by Zeeshan)*

- Organizations: Department of Bioinformatics, University of Wuerzburg Germany, Blekinge Institute of Technology Sweden.
- Responsibilities: Software Engineer (Software Analysis, Research , Development, Technical Documentation)
- Description**:** This is published project, about to implement an intelligent software application capable performing efficient data classification, optimization and learning behavior implementation with the use of mathematical algorithms e.g. Genetic Algorithm, Back Propagated Neural Network etc.

- Technologies: Java, WEKA etc.

Project #13: ***Intelligent - Semantic Oriented Agent Based Search (2007-2010)*** *(proposed & developed by Zeeshan)*

- Organization: TU Vienna Austria
- Responsibilities: Project Manager (Team lead, Design, Implementation, Quality testing, Deployment and technical documentation)
- Description**:** This is published project, about to implement an enhanced Product Data Management Application capable of intelligently handling user requests by providing intelligent and flexible graphical user interface and implement a natural language based search for finding records
- Technologies: Java, ANTLR, Flex, XML, RDF, OWL, MySQL etc.

Project #12: ***Java-Preprocessed source code analyzer (2007)*** *(proposed & developed by Zeeshan)*

- Responsibilities: Project Manager (Design, Implementation, Quality testing, Deployment and technical documentation)
- Description: This is published project, about to implement an intelligent software application capable of intelligently analyzing product line architecture based Java projects (preprocessed source code) by lexing, parsing and visualizing results in 2D and 3D diagrams (Graphs, Charts, Maps etc.)
- Technologies: Java, ANTLR, MySQL etc.

Project #11: ***Plotter Control and Management System (2007)*** *(proposed & developed by Zeeshan)*

- Organization: TU Vienna Austria
- Responsibilities: Project Manager (Design, Implementation, Quality testing and Deployment)
- Description: This published project is about to implement an intelligent software application capable of intelligently analyzing printer hardware (plotter) to extract printing based information and generating bills.
- Technologies: Java, MySQL etc.

Project #10: ***C-Preprocessed source code analyzer (2007)*** *(proposed & developed by Zeeshan)*

- Responsibilities: Project Manager (Design, Implementation, Quality testing, Deployment and technical documentation)
- Description: This is published project, about to implement an intelligent software application capable of intelligently analyzing product line architecture based C++ projects (preprocessed source code) by lexing, parsing and visualizing results in 2D and 3D diagrams (Graphs, Charts, Maps etc.)
- Technologies: Java, ANTLR, MySQL etc.

Project #9: ***VDI Bolt Calculation (2007)*** *(proposed & developed by Zeeshan)*

- Organizations: MIVP, TU Vienna Austria.
- Responsibilities: Project Manager (Design, Implementation, Quality testing, Deployment and technical documentation)
- Description: This project is about to implement an intelligent software application capable of calculating Lateral load in both ways i.e. Material and Dimension, using Force, Number of Screws, M values, Fitting (with/out), Pretention (with/out).
- Technologies: Java etc.

Project #8: ***5D Visualizer(2006)***

- Organization: Fraunhofer Institute Experimentells Software Engineering Kaiserslautern, Germany
- Responsibilities: Software Engineer (Implementation)
- Description: This project is about to implement an intelligent software application capable of taking five dimensional dataset as input and produce 3D and 5D Visualizations by drawing graphs using rendering, texturing and transparency effects.
- Technologies: Java etc.

Project #7*:* ***M-System-NT(2006)***

- Organization: Fraunhofer Institute Experimentells Software Engineering Kaiserslautern, Germany
- Responsibilities: Software Engineer (Design, Implementation, Quality testing, Deployment and technical documentation)
- Description: This is published project, about to implement an intelligent software application capable of intelligently analyzing C projects (preprocessed source code) by lexing, parsing and visualizing results in 2D and 3D diagrams (Graphs, Charts, Maps etc.)
- Technologies: Java, ANTLR, MySQL etc.

Project #6: ***Ini. Parser. M-System (2006)***

- Organization: Fraunhofer Institute Experimentells Software Engineering Kaiserslautern, Germany
- Responsibilities: Software Research Engineer
- Description: This project is about to implement a desktop software application capable of intelligently analyzing Ini - Files by lexing, parsing and visualizing results in 2D diagrams (Graphs, Charts, Maps etc.).
- Technologies: Java, MySQL etc.

Project #5: ***Wiz-Air (2004-2005)***

- Organization: Wiz-Links Pakistan & Dubai
- Responsibilities: Software Engineer (Implementation, Quality testing, Deployment and technical documentation)
- Description: This project is about to implement an intelligent software application capable of completely running a TV Channel's satellite, news, database, transmission, graphics, hardware and mixing departments . Complete broadcasting solutions for TV Channels
- Technologies: VC++, VB. Net, VC#, Oracle, Matrox, UML, Rational Rose etc.

Project #4: ***I-Photo Grapher & Portal(2004)***

- Organization: NSE Technologies, Pakistan & Singapore
- Responsibilities: Software Quality Engineer (Quality testing)
- Description: This project is about to implement an intelligent software application capable of providing a complete image management system for Camera (Proposed for Kodak Cameras)
- Technologies: VC++ etc.

Project #3: ***Oras Inventory Control System, NSE Technologies (2004)***

- Organization: NSE Technologies, Pakistan & Singapore
- Responsibilities: Software Engineer (Implementation, Quality testing, Deployment and technical documentation)
- Description: This project is about to implement a database/inventory control system capable of completely managing data of ORAS Company Pvt. Ltd.
- Technologies: VC- Sharp, Oracle etc.

Project #2: ***Habib Rafique DSL Accounts system (2003)***

- Organization: Habib Rafiqe Technologies Pvt. Ltd Lahore Pakistan
- Responsibilities: Software Engineer (Implementation, Quality testing, Deployment and technical documentation)
- Description: This project is about to implement a database/inventory control system capable of completely managing data of Habib Rafique DSL Department.
- Technologies: VC- Sharp, Oracle etc.

Project #1: ***Smart House(2001-2003)*** *(proposed & developed by Zeeshan)*

- Company: University of Central Punjab Pakistan
- Responsibilities: Research Software Engineer (Implementation, Quality testing, Deployment and technical documentation)
- Description: This is published project, a software and hardware application, solely based on personal efforts. It's about to implement a home automation system by automating house's electric appliances and providing security system.
- Technologies: C, C++, VB 6, Nokia Tool Kit, Hardware etc.

Distinguished Academic Software Projects

1. ***BTH Online Student Forum*** *(2006), programmed in Java, Blekinge Institute of Technology, Sweden. (Published)*
2. ***TAC Online Multi Agent System*** *(2005), programmed in Java, Blekinge Institute of Technology, Sweden. (Published)*
3. ***BTH HCI Chat*** *(2005), programmed in Java, Blekinge Institute of Technology, Sweden. (Published)*
4. ***AI 3D Cybog*** *(2005); War Game developed using CAICL, Blekinge Institute of Technology, Sweden. (Published)*
5. ***Artificial Life*** *(2006) programmed in C#, Blekinge Institute of Technology, Sweden.*
6. ***NHL ICE Hockey Team*** *(2005), programmed in Java, Blekinge Institute of Technology, Sweden.*
7. ***Knowledge Base*** *(2005), developed using C#, Ontology and Java, Blekinge Institute of Technology, Sweden.*
8. ***Kalaha AI Game*** *(2005), developed in Visual Basic Dot Net, Blekinge Institute of Technology, Sweden.*
9. ***System Controller*** *(2004), University of Central Punjab, Pakistan.*
10. ***Online Book Shop*** *(2002), developed in Java, University of Central Punjab, Pakistan.*
11. ***2D Data Structure Tutor*** *(2001), developed in C++, University of Central Punjab, Pakistan.*
12. ***Air War*** *(2001), 2D Game developed in C++, University of Central Punjab, Pakistan. (Published)*

Selected Publications List *Only Appeared as First Author / Last Author (Main/Supervisor Research)**

Written and Edited Books *(2010 - 2011)*

1. **Zeeshan A.** (2011). "***I-HMI and LT-SOS in PDM Systems****; Implementing AI Concepts for The Advancement of Multiple Role Based User Oriented Product Data Management Systems*", ISBN-13: 978-3-639-31070-2, Language: English, Publishers: VDM Verlag Dr. Müller, Pages: 240.
2. **Zeeshan A.** (2011). "***Integration of variants handling in M-system NT****; Empirically evaluated approach towards the identification of correlation between traditional and product line measures*", ISBN-13: 978-3-639-32553-9, language: English, Publishers: VDM Verlag Dr. Müller, Pages: 112.
3. **Zeeshan A.**, Mujtaba Ali (2011). "***Smart House***; Towards Artificially Intelligent Home Automation System", ISBN-13: 978-3-639-32410-5, Language: English, Publishers: VDM Verlag Dr. Muller, Pages: 181.
4. **Zeeshan A.** (2010). *"**Variant Handling using ZAC**: Variability Analysis of Product Line Architectures developed in C/C++",* ISBN-13: 978-3-8383-4103-3, language: English, Publishers: LAP LAMBERT Academic Publishing, Pages: 108.
5. **Zeeshan A.** (2010). "***Contributions to advance Product Data Management Systems (PDMs)****: Towards Flexible Graphical User Interface and Semantic Oriented Search for Web based PDMs",* ISBN-13: 978-3838324951, Language: English, Publisher: LAP Lambert Academic, Pages: 152.
6. **Zeeshan A.** (2010). "***Ele-Comp-Hus****: Digitally Mobile and Computerized House*", ISBN-10: 3838352092, ISBN-13: 978-3838352091, Language: English, Publisher: LAP Lambert Academic Publishing, Pages: 88.
7. Ina T. and **Zeeshan A**. (2010). "***Integration of NLP Search with I-SOAS****: Construction of Natural Language based Grammar and Mapping with Ontologies*", ISBN-13: 978-3843358514, Language: English, Publisher: LAP LAMBERT Academic Publishing, Pages: 68 pages.

Book Chapters *(2007 - 2012)*

1. **Zeeshan A.**, Saman M. "***Measurement, Analysis with Visualization for better Reliability***", Artificial Intelligence and Hybrid Systems, ISBN: 978-14775547-3-9, Vol.1, ID. 2, Edited by Claudio Rocha, iConcept Press Ltd. First Pub. Online 2012, *(in press)*
2. **Zeeshan A.**, Ina T. (2012)"***Integration of Flexible RIA Based GUI in I-SOAS***", Chapter: Production Organisation and Management, Book: Innovative Production Machines and Systems. Volume 6, ISBN: 978-1471050268, lulu Publishers, 2012.
3. **Zeeshan A.**, Vasil P. (2012) "***Integration of Agile Ontology Mapping towards NLP Search in I-SOAS***", Chapter: Intelligent Optimisation, Book: Innovative Production Machines and Systems. Volume 6, ISBN: 978-1471050268, lulu Publishers, 2012.
4. **Zeeshan A.**, Sudhir G. (2009) "***Removal of communication gap***", Chapter 3: HMI & HCI, In Innovative Production Machines and Systems, Volume 5, Pages 93-98, Vol. 5, ISBN 978-184995-006-0, Whittles Publishing, 2009.
5. **Zeeshan A.**, Gerhard D.(2008) "***Design Implementation of Semantic Oriented Agent and Knowledge based approach for Intelligent Human Machine Data Manipulation***", Innovative Production Machines and Systems, Volume 4, ISBN 978-1904445-81-4, Whittles Publishing,, 2008
6. **Zeeshan A.** (2007): "***Measurement Analysis and Fault Proneness Indication in Product Line Applications (PLA)***", Chapter 7, In Frontiers in

Artificial Intelligence and Applications, Volume 161, Pages 391-400, ISBN 978-1-58603-794-9 (print), ISBN 978-1-60750-281-4 (online), IOS Press, 2007.

Journal Manuscrits *(2009 - 2012)*

1. **Zeeshan A.**, et al. "***Supervised Classifier towards Effective Preprocessed Spectral Data Extraction and Management***". (*In Preparation)*
2. **Zeeshan A.**, et al. "***A platform for facile Quantitative Mass Isotopomers Distribution Analysis***". (*In Review)*
3. **Zeeshan A.**, et al. "***Software LS-MIDA for efficient Mass Isotopomer Distribution Analysis***". (*in revision*)
4. Dandekar T., Astrid F., Saman M., and **Zeeshan A**. (2013). "***Software Applications toward Quantitative Metabolic Flux Analysis and Visualization***". Briefings in Bioinformatics, Oxford University Press. First online published November 9, 2012. *(In press)*.
5. **Zeeshan A.**, Saman M., Dandekar T. (2012). "***Computational Feature Performance and DSA Evaluation of Applications towards MFA***". Recent Patents on Computer Science, Bentham Science Publishers, Vol. 5, No. 3, November.
6. **Zeeshan A.**, Saman M., Dandekar T. (2012). "***Formal UML Modeling of Isotopo, Bioinformatical Software for Mass Isotopomers Distribution Analysis***". Software Engineering, Vol. 2 No. 4, Pages: 147-159, November.
7. **Zeeshan A.** (2012). "***Distributed Real Time PCM System, UML Design and Development with Embedded Programming***", Software Engineering, Vol. 2 No. 4, Pages: 129-137, November.
8. **Zeeshan A.**, Saman M., Dandekar T. (2012). "***Unified Modeling and HCI Mockup Designing towards MIDA***". International Journal of Emerging Sciences, Vol. 2, No. 3, Pages 361-382, September.
9. **Zeeshan A.** (2012). "***Review Research: Learning using Web Application I-SOAS***", International Journal of Web Applications, Vol. 4, No. 3, Pages 151-167, September.
10. **Zeeshan A.**, Dandekar T., Saman M. (2012). "***ADAM: Transiting PDM into Clinical Patient Data Management***". International Journal of Emerging Sciences, Vol. 2, No. 2, Pages 280-299, June.
11. **Zeeshan A.**, Dandekar T., Saman M. (2012). "***Semantic web; Ontology Specific Languages for Web Application Development***". International Journal of Web Applications, Vol. 4, No. 1, Pages 33-41, March.
12. **Zeeshan A.**, Dandekar T., Saman M. (2012). "***Role of Ontology in NLP Grammar Construction for Semantic based Search Implementation in Product Data Management Systems***", International Journals of Management, IT & Engineering, Vol. 2, No. 2, Pages 1-40, Paper ID 706, February.
13. **Zeeshan A.**, Saman M. (2011). "***Middleware Technologies; Chain Web Grid Services***". International Journal of Web Applications, Vol. 3, No. 4, Pages 197-205, December.
14. **Zeeshan, A.**, Saman M. (2011). "***NLP Syntax Structure using ANTLR in I-SOAS***", International Journal of Information Technology and Engineering, Vol. 2. No. 2, Pages 51-57, December.
15. **Zeeshan, A.**, Ali M., Saman M., (2011). "***Implementing Computerized and Digitally Mobile Home Automation System towards Electric Appliance Control and Security System"***, International Journal of Emerging Sciences, Vol. 1, No. 3, Pages 487-503, Paper ID 10, September.
16. **Zeeshan, A.**, Saman M. (2011). "***Machine Learning and Data Optimization using BPNN and GA in DOC***", International Journal of Emerging Sciences, Vol. 1, No. 2, Pages 108-119, Paper ID 6, June.
17. **Zeeshan, A.** (2011). "***Importance of I-SOAS in PDM Community: Artificial intelligence based approach for Product Data Management"***, International Journal of Web Applications, Vol. 3, No.1, Page: 12-16, March.
18. **Zeeshan, A.** (2011). "***Designing Flexible GUI to Increase the Acceptance Rate of Product Data Management Systems in Industry***", International Journal of Computer Science & Emerging Technologies, Vol. 2, No.1., Page 100-109, February.
19. **Zeeshan, A.** (2011). "***Designing Knowledge Base towards PDMS***", Int. Journal of Information Technology and Engineering, Vol. 2, No. 1, Pages 9-12, July.
20. **Zeeshan, A.** (2010). "***Proposing LT based Search in PDM Systems for Better Information Retrieval***", International Journal of Computer Science & Emerging Technologies, Vol.1, No. 4, Pages 86-100, December.
21. **Zeeshan, A.**, Saman M. (2010): "***Towards Increase in Quality by Preprocessed Source Code and Measurement Analysis of Software Applications***", International Science & Tech. Transactions on Info. Tech. - Theory and Applications, Vol.1, No. 1, Pages 8-13, December.
22. **Zeeshan, A.**, Saman M., T. Dandekar (2010). "***Towards Design and Implementation of a Language Technology based Information Processor for PDM Systems***", International Science & Tech. Transactions on Info. Tech. - Theory and Applications, Vol. 1, No. 1, pp. 1-7, December.
23. **Zeeshan, A.**, Ina T. (2010), "***Integration of Natural Language Processing towards Semantic Oriented Search***", International Journal of Computer Science and Software Technology, Vol. 3, No.2, pp. 69-76, December.
24. **Zeeshan, A.** (2010): "***Towards Performance Measurement and Metrics based Analysis of PLA Applications***", International Journal of Software Engineering & Applications, Vol.1, No. 3, pp 66-80, July.
25. **Zeeshan, A.** (2009): "***Proposing Semantic Oriented Agent and Knowledge base Product Data Management***", Information Management and Computer Security Journal, Vol. 17, No. 5, pp: 360-371, *International Award for Excellence: the Donn B. Parker Award* to issue.*

Conference Full Papers, Published in Proceedings *(2007 - 2010)*

1. **Zeeshan A.**, Ina T. (2010): "***Integration of Agile Ontology Mapping towards NLP Search in I-SOAS***", In proceedings of 6th Virtual Conference, Innovative Production Machines and Systems, Section: Intelligent Optimization, European Commission, Cardiff Uni., Cardiff UK, 15-26 November. ***Potential Paper Award****.
2. **Zeeshan A.**, Popov V. (2010): "***Integration of Flexible RIA Based GUI in I-SOAS***", In proceedings of 6th Virtual Conference, Innovative Production Machines and Systems, Section: Production Organization & Management, European Commission, FP6 I*PROMS Network of Excellence, Cardiff University, 15-26 November.
3. **Zeeshan A.** (2009): "***Intelligent Human Machine Interface Design for Advanced Product Life Cycle Management Systems***", In the proceedings of International Conference on Frontiers of Information Technology, Special Interest Group on Artificial Intelligence (ACM-SIGART), No. 49, 16-18

December

4. **Zeeshan A.** (2009): "***Intelligent - Semantic Oriented Agent Based Search (I-SOAS)***", In the proceedings of ACM Doctoral Symposium at International Conference on Frontiers of Information Technology, Special Interest Group on Artificial Intelligence (ACM-SIGART), No. 55, ISBN:978-1-60558-642-7, 16-18 December.
5. **Zeeshan A.**, Sudhir G. (2009): "***Removal of Communication Gap***", In the proceedings of 5th Virtual Conference on Network of Excellence on Innovative Production Machines and Systems, ISBN 978-184995-006-0, Cardiff University, Whittles Publishing, Scotland UK, 6-17 July
6. **Zeeshan A.** (2009): "***Semantic Oriented Information Processing & Modeling***", In online proceedings of 5th Virtual Conference, Innovative Production Machines and Systems, European Commission, FP6 I*PROMS Network of Excellence, Cardiff University, Cardiff, 6-17 July.
7. **Zeeshan A.** (2009): "***PDM Problem and Solutions***", In online proceedings of 5th Virtual Conference, Innovative Production Machines and Systems, European Commission, FP6 I*PROMS Network of Excellence, Cardiff University, Cardiff, 6-17 July.
8. **Zeeshan A.** (2009): "***I-SOAS Data Repository for Advanced Product Data Management***", In the proceedings of IADIS European Conference Data Mining 2009 (part of MCCSIS 2009), P177-8, ISBN: 978-972-8924-88-1, Portugal, 18-20 June
9. **Zeeshan A.** (2009): "***SW, NLP and IUI using Intelligent Semantic Oriented Agent based Search (I-SOAS)***", In Doctoral Symposium at Semantics Week, Amsterdam Holland, 25-26 June
10. **Zeeshan A.** (2009): "***Home Automation***", In the proceedings of 9th National Research Conference, SZABIST | Shaheed Zulfikar Ali Bhutto Institute of Science and technology Research, 25 June
11. **Zeeshan A.** (2009): "***AI 3D Cybug Gaming***", In the proceedings of 9th National Research Conference, SZABIST | Shaheed Zulfikar Ali Bhutto Institute of Science and technology Research, 25 June
12. **Zeeshan A.** (2009): "***Aero Fighter -2D Gaming***", In the proceedings of 9th National Research Conference, SZABIST | Shaheed Zulfikar Ali Bhutto Institute of Science and technology Research, 25 June
13. **Zeeshan A.** Gerhard D. (2009), "***Design Implementation of I-SOAS IPM for Advanced Product Data Management***", In the proceedings of The Second IEEE International Conference On Computer, Control & Communication, ISBN: 978-1-4244-3313-1, May 17-18 February. ***Best Paper Award****.
14. **Zeeshan A.** (2009): "***PDM based I-SOAS Data Warehouse Design***", In the proceedings of FIFTH International Conference on Statistical Sciences: Mathematics, Paper ID 125, ISBN 978-969-8858-04-9, Vol. 17, 23-25 January
15. **Zeeshan A.**, Gerhard D. (2008): "***Design Implementation of Semantic Oriented Agent and Knowledge based approach for Intelligent Human Machine Data Manipulation***", In the proceedings of 4th Virtual International Conference on Innovative Production Machines and Systems (IPROMS), ISBN 978-1904445-81-4, USA ISBN 978-14398-0117-8, Cardiff University, Whittles Publishing, Scotland UK, 1-14 July
16. **Zeeshan A.**, Gerhard D. (2008): "***Information Engineering and Knowledge Management based approach towards Intelligent E-Learning Systems Development***", In proceedings of 4th Virtual Conference, Innovative Production Machines and Systems, European Commission, FP6 I*PROMS Network of Excellence, 1-14 July.
17. **Zeeshan A.**, Gerhard D. (2008): "***Semantic Oriented Agent based Approach towards Engineering Data Management, Web Information Retrieval and User System Communication Problems***", In the proceedings of 3rd International Conference for Internet Technology and Secured Transactions, , ICITST 08, IEEE - CST, Dublin Institute of Technology, Dublin Ireland, June 23-28
18. **Zeeshan A.**, Sudhir G., Hans K. (2008): "***Design Artifact's, Design Principles, Problems, Goals and Importance***", In Proceedings 4th International Statistical Conference, Paper ID 42, Vol. 15, 57-68, ISBN 978-969-8858-04-9, May 9-11
19. **Zeeshan A.** (2008): "***Statistical Semantic Modeling of Preprocessed Source Code to Identify, Measure and Visualize the Complexities in Software Product Line Applications***", In Proceedings 4th International Statistical Conference, Paper ID 19, Vol. 15, 1-12, ISBN 978-969-8858-04-9, 9-11 May
20. **Zeeshan A.** (2008): "***Statistical Trading Using Target Oriented Trading Agent***", In Proceedings 4th International Statistical Conference, Paper ID 20, Vol. 16, 355-358, ISBN 978-969-8858-04-9, 9-11 May
21. **Zeeshan A.**, Gerhard D. (2007): "***Contributions of PDM Systems in Organizational Technical Data Management***", In the proceedings of The First IEEE International Conference On Computer, Control & Communication, 12-13 November
22. **Zeeshan A.**, Gerhard D. (2007): "***An Agent based Approach towards Metadata Extraction, Modeling and Information Retrieval over the Web***", In the proceedings of First International Workshop on Cultural Heritage on the Semantic Web in conjunction with the 6th International Semantic Web Conference and the 2nd Asian Semantic Web Conference 2007, (ISWC + ASWC 2007), P 117, 12-15 November
23. **Zeeshan A.**, Gerhard D. (2007): "***Role of Ontology in Semantic Web Development***", In the proceedings of First International Workshop on Cultural Heritage on the Semantic Web in conjunction with the 6th International Semantic Web Conference and the 2nd Asian Semantic Web Conference 2007, (ISWC + ASWC 2007), P 119, 12-15 November
24. **Zeeshan A.** (2007): "***Measurement Analysis and Fault Proneness Indication in Product Line Applications (PLA)***", In the 6th International Conference on New Software Methodologies, Tools, and Techniques (SOMET-07), 7-9 November
25. **Zeeshan A.**, Gerhard D. (2007): "***Designing a Dynamic Components and Agent based Approach for Semantic Information Retrieval***", In the proceedings of 6th CIIT Workshop on Research in Computing , P 8, 27 October
26. **Zeeshan A.**, Gerhard D. (2007): "***How does Ontology Contribute in Semantic Web Development?***", In the proceedings of 6th CIIT Workshop on Research in Computing, P 3, 27 October
27. **Zeeshan A.**, Gerhard D. (2007): "***Web to Semantic Web & Role of Ontology***", In the proceedings of National Conference on Information and Communication Technologies, P 100, 9 June
28. **Zeeshan A.**, Gerhard D. (2007): "***Personal Assistant towards Semantic Information Retrieval***", In the proceedings of Fifth International Workshop on Ontologies and Semantic Web for E-Learning, 13th International Conference on Artificial Intelligence in Education (AIED 2007), P115, 9-13 July

Magazine Articles *(2011 - 2012)*

1. **Zeeshan A.** (2012). "***Natural Language Processing; a brief Introduction to ANTLR***", In NAYS e-Magazine by National Academy of Young Scientists, Section: Information Technology, Issue October 2012. *(Invited Author)*
2. **Zeeshan A.** (2012): "***Human Computer Interaction; Mockup***", In NAYS e-Newsletter, Issue: 14, by National Academy of Young Scientists, Section:

Information Technology, Page 10, Jan-Feb 2012. *(Invited Author)*

3. **Zeeshan A.** (2011): "***Web to Semantic Web; Online Data Search***", In NAYS e-newsletter, Issue: 13, by National Academy of Young Scientists, Section: Information Technology, Pages 25-27, Nov-Dec 2011. *(Invited Author)*
4. **Zeeshan A.** (2011): "***Chain SPL; Individual Efforts with Mutual Benefits by Reducing Gaps between Academia & Industry***", In NAYS e-newsletter, Issue: 12, by National Academy of Young Scientists, Pages 32-34, Sep-Oct 2011. *(Invited Author)*
5. **Zeeshan A.** (2011): "***Hypothesis; Computer Science, A Revolutionary Field***", In NAYS e-newsletter, Issue: 11, by National Academy of Young Scientists, Section: Information Technology, Pages 25-27, July-Aug 2011. *(Invited Author)*
6. **Zeeshan A.** (2011): "***Introduction to the field of Product Data Management (PDM)***", In NAYS e-newsletter, Issue: 10, by National Academy of Young Scientists, Section: Information Technology, Pages 18-20, May-June 2011. *(Invited Author)*

Conference Poster Papers and Presentations *(2009 - 2012)*

1. **Zeeshan A.**, Saman M, Eisenreich W., Dandekar T. (2012): "***Isotopo towards Quantitative Mass Isotopomers Distribution Analysis using Spectral Data***", Accepted at the 9th Symposium on the Practical Applications of Mass Spectrometry in the Biotechnology Industry, California USA, 11-14 September 2012.
2. **Zeeshan. A.**, (2012): "***Intelligent individual health data management in thrombosis and hemostasis***", The ICT Proposers Day, by European Commission, Polish Ministry of Science and Higher Education Warsaw Poland, 26th September 2012
3. **Zeeshan A.**, Saman M., Eisenreich W., Dandekar T. (2012): "***Isotopo: Software towards Quantitative Mass Isotopomers Distribution Analysis, Visualization and Data Management***", the Fifth Annual New England Database Summit (NEDB Summit), by EMC, Microsoft, and Paradigm4, Computer Science and Artificial Intelligence Laboratory (CSAIL), the Massachusetts Institute of Technology (MIT), 2 February USA.
4. **Zeeshan A.** (2012): "***Efficient Data Management, Intelligent Information Retrieval and Flexible User Interface in PDM Systems***", the Fifth Annual New England Database Summit (NEDB Summit), by EMC, Microsoft, and Paradigm4, Computer Science and Artificial Intelligence Laboratory (CSAIL), the Massachusetts Institute of Technology (MIT), 2 February USA.
5. **Zeeshan A.**, Saman M., Dandekar T. et al. (2011): "***Intelligent Information Management for efficient Computational Biology***", Information and Networking Day: Intelligent Information Management, by Information and Communication Technology Research (ICT), European Commission, at Jean Monnet Conference Centre, 26 September, Luxembourg.
6. **Zeeshan A.** (2011): "***Intelligent Information Retrieval and Flexible GUI in PDM Systems***", Information and Networking Day: Intelligent Information Management, by Information and Communication Technology Research (ICT), EU Commission, at Jean Monnet Conference Centre, 26 September, Luxembourg.
7. **Zeeshan A.**, Dandekar T.et al. (2011): "***Data integration and refined metabolic modeling***", Review Meeting of Priority Program! Host Adapted Metabolism of Bacterial Pathogens, (SPP 1316), by DFG, German Research Foundation, 06 April 2011, Bonn - Bad Godesberg Germany
8. **Zeeshan A.** (2011): "***Integration of DB in I-SOAS towards PDMs***", at NEDB Database Summit 2011 at Computer Science and Artificial Intelligence Laboratory (CSAIL), The Massachusetts Institute of Technology (MIT), January 2011, USA.
9. **Zeeshan A.** (2009): "***NLP TU I-SOAS***", In Startup Weekend Vienna, Microsoft Innovation Center, 05 June, Vienna Austria.
10. **Zeeshan A.** (2009): "***Intelligent Semantic Oriented Agent Based Search (I-SOAS)***", In Innovation Goes Business, by European Commission and INiTS, Austria Center, 7 May, Vienna Austria.
11. **Zeeshan A.** (2009): "***Zeeshan Ahmed Java Preprocessed Source Code Analyzer (ZAJ)",*** In Innovation Goes Business, by European Commission and INiTS, Austria Center, 7 May, Vienna Austria.

Technical Project Reports *(2002 - 2012)*

1. **Zeeshan, A.** (2012): "***ISOTOPO***", Department of Bioinformatics, Biocenter, University of Wuerzburg, Germany.
2. **Zeeshan, A.** (2011): "***Least Square Mass Isotopomer Distribution Analysis***", Department of Bioinformatics, University of Wuerzburg, Germany.
3. **Zeeshan, A.** (2011): "***Metabolic Flux Analysis***", Department of Bioinformatics, Biocenter, University of Wuerzburg, Germany.
4. **Zeeshan, A.** (2010): "***Intelligent Semantic Oriented Agent based Search***", MIVP, IKL, Vienna University of Technology, Vienna Austria.
5. **Zeeshan, A.** (2005): "***Wiz-Air***", Technical Project Report, at of Wiz-Links Pvt. Ltd, 2005
6. **Zeeshan, A.** (2004): "***Oras Inventory Control System***", at NSE Technologies Pvt. Ltd. 2004.
7. **Zeeshan, A.** (2003): "***HR DSL Account System***", at Habib Rafique Technologies, Pvt. Ltd. 2003
8. **Zeeshan, A.** (2002): "***Smart House***" at Faculty of Information Technology University of Central Punjab, 2002

Graduate Students, Supervised Thesis *(2009 - 2010)*

1. Vasil Popov (2010): "***Advanced User Interface Technology for I-SOA***S", ISBN-13: 978-3-8433-6892-6, Book language: English, Publishing house: LAP LAMBERT Academic Publishing, Number of pages: 60, Published at: 2010-12-29, [Book Publication]
2. Vasil Popov (2009): "***Advanced User Interface Technology for Intelligent Semantic Oriented Agent Based Search (I-SOAS)***", Published and Archived in Institute for Engineering Design and Logistic Engineering Vienna University of Technology Austria and Technical University Sofia, Sofia Bulgaria. ["A" Graded, Master Thesis]
3. Ina Tacheva (2009): "***Natural Language Processing Approaches for Intelligent Semantic Oriented Agent Based Search (I-SOAS)***", Published and Archived in Institute for Engineering Design and Logistic Engineering, Vienna University of Technology Austria and Technical University Sofia, Sofia Bulgaria. ["A" Graded, Bachelor Project]
4. M.T. Hashmi (2009):"***High IT Failure Rate: A Management Prospect***", Published and Archived at Institute for Engineering Design and Logistic Engineering, Vienna University of Technology Austria and Department of Business Administration, School of Management, Blekinge Institute of Technology, Ronneby Sweden. ["A" Graded, Master Thesis]
5. Omer Faraz (2009): "***Ontological, Warehoused and Knowledge Management Based Product Data Management***", Institute for Eng. Design and

Logistic Eng., Vienna University of Technology, and Department of Computer and System, Stockholm University, 2009. [Master Thesis Publication]

Written & Defended Thesis *(2002 - 2012)*

1. **Zeeshan, A.** (2006): "***Integration of variants handling in M-system NT***", Department of Computer Science, Blekinge Institute of Technology, University of Blekinge Sweden In Co-operation with Fraunhofer Institute of Experimentells Software Engineering Germany, Ronneby Sweden. ["A" Graded, Master Thesis]
2. **Zeeshan, A.** (2004): "***Object Oriented Framework Design Patterns, a Role Modeling Approach***", Department of Computer Science, Punjab Institute of Computer Science (PICS), Faculty of Informatics, University of Central Punjab ["A" Graded, Master Thesis]
3. **Zeeshan, A.** (2002): "***Smart House***", Punjab Institute of Computer Science (PICS), Department of Computer Science, Punjab Institute of Computer Science (PICS), Faculty of Informatics, University of Central Punjab ["A" Graded, Bachelor Thesis & Project]

Active Research Participations

International Journals; Editorship

1. *Editor: Special Journal Issue, Scientific & Academic Publishing (SAP), 2012.*
2. *Editorial Board Member; Progress in Machines and Systems, since year 2012.*
3. *Editorial Board Member; Electronic Devices, since year 2012.*
4. *Editorial Board Member; International Journal of Advances in Engineering & Technology, since year 2012.*
5. *Editorial Board Member; Software Engineering Journal, by Scientific & Academic Publishing Corporation, since years 2011.*
6. *Editorial Board Member; International Journal of Management, IT & Engineering, since years 2011.*
7. *Editorial Board Member; International Journal of Marketing and Technology, since years 2011.*
8. *Editorial Board Member; International Journal of Research in Social Sciences, since years 2011.*
9. *Editorial Board Member; International Journal of Information Technology and Engineering, sine years 2011.*
10. *Editorial Board Member; Computer Science Journal (CSJ), Innovative Research Foundation, since years 2010-12.*

International Journals; Referee

1. *Journal Referee; Science, Technology, and Development, Pakistan Council for Science and Technology (PCST, MST), since 2012*
2. *Journal Referee; International Journal of Advances in Engineering & Technology, since 2012*
3. *Journal Referee; Social Network Analysis and Mining, Jr. no. 13278, Springer Publishers, since 2011*
4. *Journal Referee; Behavior & Information Technology Journal since 2011.*
5. *Journal Referee; Symbiosis Center for Information Technology SCIT) Journal, since 2011*
6. *Journal Referee; LNCS Transactions on Computational Collective Intelligence, Springer Publishers, since 2011*
7. *Journal Referee; International Journal of Open Source Software and Processes, since 2011*
8. *Journal Referee; International Journal of Software Engineering & Applications (IJSEA), since 2010*
9. *Journal Referee; Journal of Engineering and Computer Innovations, Academics Journal, since 2010*
10. *Journal Referee; Multimedia Tools and Applications, An International Journal, Springer US, since 2010*
11. *Journal Referee; IST Transactions of Biomedical Sciences and Engineering, IST Press. since 2010*
12. *Journal Referee; International Journal of Computer Science and Software Technology (IJCSST), International Science Press, since 2010*
13. *Journal Referee; IST Transactions of Computer Vision Systems, Theory and Applications, IST Press, since 2010*
14. *Journal Referee; IST Transactions of Information Technology-, Theory and Applications, IST Press, since 2010*
15. *Journal Referee; The International Journal of Web Applications (IJWA), since 2010*

International Conferences; Session Chair

1. *Session Chairperson ISOSS 2008; Technical Session-15: Economic & Econometric Theory-I: Saturday, May 10, 2008 at 1200-1300 hours (Room-3), In Fourth international Conference on Statistical Sciences, University of Gujrat, May 9-11, Gujrat Pakistan 2008*

International Conferences; Program Committee Member (PCM)

1. *Technical PCM; International Symposium on Control, Automation, Industrial Informatics and Smart Grid, August 24-25, 2013.*
2. *PCM; 4th Int. Workshop on Software Quality (SQ 2013), 13th Int. Conf. Computational Science and Its Applications, June 24-27, 2013.*
3. *PCM; the International Conference on E-Technologies and Business on the Web (EBW), May 7-9, 2013.*
4. *PCM; Information Communication Technology-EurAsia Conference, Gajah Mada University, Indonesia, March 25-29, 2013.*
5. *PCM; the Second International Conference on Digital Enterprise and Information Systems (DEIS), March 4-6, 2013.*
6. *PCM; the Second International Conference on e-Technologies and Networks for Development (ICeND), March 4-6, 2013.*
7. *PCM; the Second International Conference on Cyber Security, Cyber Warfare and Digital Forensic, March 4-6, 2013.*
8. *PCM; the First International Conference on Green Computing, Technology and Innovation (ICGCTI), March 4-6, 2013.*
9. *Technical PCMs; 2nd International Conference on Information Technology, System and Management (ICITSM), March 4-5, 2013.*
10. *PCM; the Third International Conference on Digital Information Processing and Communications (ICDIPC), Jan 30 - Feb 1, 2013.*
11. *Technical PCM; 2nd International Conference on Advance Information System, E-Education & Development, January 15-16, 2013.*
12. *PCM; the IEEE International Conference on E-Learning and E-Technologies in Education, SDIWC, at TU Lodz Poland, September 2012.*
13. *Technical PCM; the Fifth International Conference on the Applications of Digital Information and Web Technologies, August 2012*
14. *Technical PCM; the 2012 IEEE Symposium on Humanities, Science and Engineering Research, Malaysia, June 2012*
15. *Technical PCM; SQ 2012: Third International Workshop on Software Quality, Brazil, June 2012*
16. *Technical PCM; Reviewer. IEEE Business Engineering and Industrial Applications Colloquium (IEEE BEIAC), April 2012*
17. *Technical PCM; the fourth International Conference on Network Digital Technologies, Canadian Uni. of Dubai, April 2012*
18. *Technical PCM; the first international conference on Innovative Computing Technology, Springer, December 2011*
19. *Technical PCM; IEEE -the 3rd International Congress On Engineering Education (ICEED 2011), Malaysia, December 2011*

20. *Technical PCM; 3rd IEEE International Conference on Privacy, Security, Risk and Trust, International Workshop on C4I Systems and Information Security, MIT Boston USA, 9-11 October 2011*
21. *Technical PCM; IEEE - International Symposium on Business Engineering and Industrial Applications (ISBEIA 2011), September 2011*
22. *PCM; 4th IEEE International Conference on the Applications of Digital Information and Web Technologies (ICADIWT 2011), August 2011*
23. *PCM; 4th International Conference on Network Security & Applications (CNSA-2011), LNCS- Springer Publishers, July 2011*
24. *PCM; 3rd International conference on Networks & Communications (NeCoM-2011), LNCS- Springer Publishers, July 2011*
25. *PCM; 3rd International Conference on Web & Semantic Technology (WeST-2011), LNCS- Springer Publishers, July 2011*
26. *PCM; 3rd International Conference on Wireless & Mobile Networks (WiMoN-2011), LNCS- Springer Publishers, July 2011*
27. *PCM; 3rd International Workshop on Software Engineering Processes and Applications (SEPA), ICCSA 2011, Spain, June 2011*
28. *PCM; 1st International Conference on Innovative Computing Technology (INCT 2011), June 2011*
29. *PCM; 2nd International Workshop on Software Quality, ICCSA 2011, University of Cantabria, Santander, Spain, June 2011*
30. *Technical PCM; IEEE International Symposium on Humanities, Science and Engineering (SHUSER), 2011*
31. *International Program Committee; 3rd International Conference on Network Digital Technologies, University of Macau, China, July 2011*
32. *PCM; CPR Conference and Doctoral Student Consortium, ACM SIGMIS, Ireland 2009*
33. *PCM; IEEE International level conference on Information & Communication Technologies, 2008*

International Conferences Reviewer (CR)

1. *CR - IEEE Business Engineering and Industrial Applications Colloquium (BEIAC), Malaysia, 8-9 April 2013.*
2. *CR - IEEE Symposium on Computers & Informatics, SCOPUS, Langkawi, Malaysia, April 7-9 2013.*
3. *CR - IEEE Colloquium on Humanities, Science and Engineering (CHUSER), Malaysia, 3-4 December 2012.*
4. *CR - IEEE International Conference on Power and Energy (PECON 2012), Malaysia, 2-5 December 2012.*
5. *CR - IEEE Symposium on Industrial Electronics & Applications (ISIEA 2012), Indonesia, 23–26 September 2012.*
6. *CR - IEEE Symposium on Business, Engineering and Industrial Applications (ISBEIA), Indonesia, 23-26 Sep 2012.*
7. *CR - IEEE Second International Conference on Digital Information and Communication Technology & its App., Thailand, 16-18 May 2012.*
8. *CR- the 6th Virtual Conference of Network of Excellence for Innovative Production Machines and Systems (I*PROMS), UK, 15-26 Nov. 2010.*
9. *CR - IEEE 2nd International Congress on Engineering Education (ICEED), Kaula-Lumpur 2010*
10. *CR - IEEE 1st International Conference on Integrated Intelligent Computing, IEEE Computational Intelligence Society, India 2010*
11. *CR - The Second International Conference on Network Digital Technologies (NDT 2010), Check Republic, 2010*
12. *CR - IEEE Global Communications Conference (GLOBECOM), Hilton Hawaii Village, Honolulu, Hawaii, USA, 30 Nov - 4 Dec 2009*
13. *CR - IEEE Second International Conference on the Applications of Digital Information and Web Technologies, 4-6 August, London UK 2009.*
14. *CR - the 5th Virtual Conference of Network of Excellence on Innovative Production Machines and Systems, UK, 6 - 17 July, 2009.*
15. *CR - IEEE International level conference on Information & Communication Technologies, Pakistan, 27 August 2008*

International Conferences; Session Reporter

1. *Conference Reporter; 5th International Conference on Statistical Sciences: Mathematics, Statistics and Applications, 2009.*
2. *Conference Session Chairperson and Reporter; 4th International Conference on Statistical Sciences: Mathematics, Statistics and Applications, 2008.*

Newsletters

1. *Guest Editor e-Newsletter; National Academy of Young Scientists (NAYS), NAYS Publishers, Issue 14, Edition Jan.Feb2012*
2. *Editor News, Media and Cultural Club; Punjab Institute of Computer Science, University of Central Punjab, Lahore Pakistan, 2002.*

Technical Professional & Research Skills

Research:	*Literature Reviewing, Software & Method Analysis, Prototype Development, Publication, Editing and Proposal Writing.*
Programming Languages:	***C#, Java**, C++, VB.Net, VB 6, VC++, Delphi, F#*
Tools / Technologies:	*DBGen, ANTLR, Java 3D, JPCT, JUN, JMonkey, WEKA, OSGI, COM, XML, RDF, OWL, HTML, SOAP, RSDL, RSS, MOS, Direct X & Show, Web Services, Grids, Matrox , Lixto, NetLogo, R, Perl*
Dev. Frameworks:	*MS. Visual Studio Dot Net 2005, MS Visual Studio 6, Eclipse 3.2, Delphi, Net Beans, Matlab*
Data Base (OLTP & OLAP):	*MS. SQL Server 2000, MySql 5.1, Oracle 9i & 10 G, MS. Access*
PDM Tools:	*CIM, PTC Windchill*
Bioinformatics Tools:	*Fiat Flux, Netto, Rematch, Yana, C13, FBA, MetaTool, BioOPT, FiatFlux, Rematch and BioLayout Express*
Graphical Tools:	*Adobe Photoshop, Corel Draw, Paint Brush and Pencil Drawing*
Technical Documentation:	*Ms. Office 2007- 2000, Ms Visio, UML, Ms. Project, ER-Win*
Reports Generation Tools:	*Crystal Reports 10, VB Data Reports, Quick Reports Delphi*
Operating Systems:	*UNIX Open Free BSD, Linux Debian 4, Linux Ubuntu, MS. Win 9X & 2000 & XP & Vista*
Statistics Methods:	*Brauman Least Square, Bayesian statistics, Decision Trees, Cross Validation, Regression Algorithms, Heuristics, A*, Hill Climbing, Simulated Annealing, Matrixes and Genetic Algorithm etc.*
Editor& Reviewer	*Editing of Research Papers, Journal/Magazine Issues and posters.*

Distinctions

Miscellaneous Awards *(Researcher, Manuscript, Grant and Programming)*

1. ***Best International Young Researcher*** *awarded at Second IEEE International Conference on Computer, Control & Communication, by Institute of Electrical and Electronics Engineers (IEEE), National University of Science and Technology (NUST), Higher Education Commission (HEC) Pakistan, February 2009. (Selected among the participation from more than 26 Countries.)*

Manuscripts & Related Awards

2. ***Manuscript in Top Ranked International Journal of Informatics & Bioinformatics****: Manuscript "Software Applications toward Quantitative Metabolic Flux Analysis and Visualization", et al. and Zeeshan (Last Author) at Briefings in Bioinformatics, Oxford University Press.*
3. ***Among Most Read Journal Paper****: Formal UML Modelling of Isotopo, Bioinformatical Software for Mass Isotopomers Distribution Analysis (DOI: 10.5923/j.se.20120204.08) by Zeeshan A. et al., at Software Engineering, 2012.*
4. ***Among Most Read Journal Paper****: Distributed Real Time PCM System, UML Design and Development with Embedded Programming (DOI: 10.5923/j.se.20120204.06.), by Zeeshan, at Software Engineering, 2012.*
5. ***Potential Paper*** *awarded to Paper: "Integration of Agile Ontology Mapping towards NLP Search in I-SOAS" by Zeeshan et al., Paper ID 13, at 6th I*PROMS Virtual International Conference on Innovative Production Machines and Systems, Section: Production Organization & Management, European Commission, FP6 I*PROMS Network of Excellence, Cardiff University, Whittles Publishing, Scotland UK, November, 2010.*
6. ***Best Paper Conference*** *awarded to Paper: "Design Implementation of I-SOAS IPM for Advanced Product Data Management" by Zeeshan et al., at Second IEEE International Conference on Computer, Control & Communication, IEEE- IC4, Karachi Pakistan, February 2009.*
7. ***The Donn B. Parker International Award for Excellence*** *to Emerald Information Management and Computer Security Journal, Vol.17, No.5, Paper: "Proposing Semantic Oriented Agent and Knowledge base Product Data Management" by Zeeshan, one of the article of awarded issue, Pages 360-417, September 2009.*
8. ***Best Author & Presenter*** *awarded at Second IEEE International Conference on Computer, Control & Communication, Session Computer Science, by Institute of Electrical and Electronics Engineers (IEEE), National University of Science and Technology (NUST), Higher Education Commission (HEC), Mobilink Pakistan 18th February 2009.*
9. ***Compliment Certificate; EPMA-PM*** *International Master Thesis Competition, awarded by The European Powdered Metallurgy Association (EPMA), to Thesis: "Integration of variants handling in M-system NT" by Zeeshan, Belgium & England, August 2008.*
10. ***Best Editor****: News Media and Cultural Club, Punjab Institute of Computer Science, Faculty of Information Technology, University of Central Punjab Pakistan, 2001.*

Best International Student Grant Awards

11. ***International Research Student Grant*** *awarded by Description Logics Steering Committee, The 25th International Workshop on Description Logics (DL 2012), The Sapienza University of Rome, Supported by Principles of Knowledge Representation and Reasoning, Incorporated (KR, Inc.) and Artificial Intelligence Journal Division (AIJD), Rome Italy, June 2012.*
12. ***International Research Student Grant*** *awarded by NMR Steering Committee, The 14th International Workshop on Non-Monotonic Reasoning (NMR 2012), Sapienza University of Rome, Supported by The Association for Logic Programming, Principles of Knowledge Representation and Reasoning Incorporated (KR, Inc). and Artificial Intelligence Journal Division (AIJD), Rome Italy, June 2012.*
13. ***International Talented Research Student Grant*** *(partial) awarded by STI International at the 6th Annual European Semantic Web Conference (ESWC2009). Heraklion Greece, 31 May - 4 June 2009****.***
14. ***Promising International Research Student Biggest Grant*** *awarded at Second IEEE International Conference on Computer, Control & Communication, supported by Institute of Electrical and Electronics Engineers (IEEE), National University of Science and Technology (NUST), Higher Education Commission (HEC) Pakistan, Karachi Pakistan, 18th February 2009.*
15. ***Gold Supporters General Student Grant WSDM & WAW*** *for International PhD Researcher at 2nd ACM International Conference on Web Search and Data Mining (WSDM), and The 6th Workshop on Algorithms and Models for the Web Graph (WAW), sponsored by Microsoft, Nokia, Google and Yahoo, Barcelona Spain, 9-12 February 2009.*

Programming & Related Awards

16. ***On Spot Programming Competition Winner****: organized by Society for Advancement of Computer Science (SACS) at Punjab Institute of Computer Science, Faculty of Information Technology, University of Central Punjab Pakistan. 2003.*
17. ***Certificate of Achievement****; Nation Wide Software Project Competition and Exhibition, awarded by Punjab Institute of Information Technology, University of Central Punjab, Lahore Pakistan, December 2003.*
18. ***Advance Programming Competition Winner****: organized by Society for Advancement of Computer Science (SACS) at Punjab Institute of Computer Science, Faculty of Information Technology, University of Central Punjab Pakistan.* ***2002****.*

University Distinctions

1. ***Doctor of Technology, Course Work*** *at Vienna University of Technology Austria with highest Possible Grade "A" (ECTS CGPA 4.0/4.0 Excellent, and Austrian/ German Grade 1.0 - Sehr Gut, 100% results),Vienna Austria,* ***2008****.*
2. ***Master of Science*** *- Computer Science with Major Subjects of Intelligent Software System's all Course Work in only 1st Semester (Winter 2005-06) with 85% results, final research thesis & project (Graded A), at Blekinge Institute of Technology, University of Blekinge, Karlskrona Sweden,* ***2006****.*
3. ***Master of Science*** *- Computer Science with Major Subjects of Software Engineering's all Course Work with 82.2% results, including final research thesis (Graded A), at University of Central Punjab, Lahore Pakistan,* ***2005****.*

4. ***Bachelor of Science*** *- Computer Science, 4 Years Graduation Program in 2.8 Years (started 08/2000 and completed 04/2003) with 75% including final research thesis & project (Graded A), at University of Central Punjab, Lahore Pakistan,* ***2004****.*

Distinguished Compliments & Nominations

Honors

1. ***Honored Compliment*** *in news about Research Publications and Awards, in the National Magazine: News & Views, Section: News Flash, Edition: February/March-2012, Page 23, Published by Higher Education Commission (HEC) Pakistan, April 2012.*
2. ***Honored Compliment for Outstanding Performance, Category: Excellent People February 2009****, in Journal Frei.Haus, Issue Nr. 10, Page 12, (Menschen Ausgezeichnet; Zeitschrift fuer Mitarbeiter TU Wien), by Vienna University of Technology Austria, April 2009*
3. ***Honored Compliment in News*** *about Research Awards (Best Paper, Best Research Student of Year, Best Presenter and Travel Grant), in Press Release Online & Print, by Vienna University of Technology Austria, 18 March 2009.*

Nominations

4. ***Received Compliments on Research Efforts*** *with the complimentary SCIT Journal Volume XI by SCIT Journal by Symbiosis Center for Information Technology, Symbiosis International University, 2012.*
5. ***Nominated for Young Scientist Presburger Award 2013****, European Association for Theoretical Computer Science, sponsored by Bertinoro International Center for informatics, 2012.*
6. ***Nominated for Biocenter Science Award;*** *Excellent Young Scientist of Year at Biocenter, University of Wuerzburg Germany, Highly appreciated by Award Committee, June 2012.*
7. ***Nominated for Young Scientist Presburger Award*** *to young scientist for outstanding contributions in theoretical computer science, highly appreciated by Presburger Award Committee, European Association for Theoretical Computer Science, Achen Germany 2010.*
8. ***Nominated Research Project****: I-SOAS(EN 19.002) proposed by Zeeshan, in the category of Best Research and Development Project World Wide, at Metal Working Production (MWP) Advanced Manufacturing Awards, Birmingham UK, 2010*
9. ***Nominated in INITS Thesis Award*** *by INITS, (to Thesis: "Integration of variants handling in M-system NT" by Zeeshan), International award for best research thesis worldwide, Vienna Austria, 2007*
10. ***Compliments on Research Efforts*** *in International Young Scientist Award Competition, by The 12th Rhythm Production and Perception Workshop (RPPW 2009), at International Symposium on Distorted time and motor Control, Lille France, 2009.*

Selected Talent Awards

1. ***Distinct Positions holder in Declamation Competition****, University of Central Punjab, Lahore Pakistan, Years: 2001,2002,2003, 2004*
2. ***Distinct Positions holder in Annual Debate Competition, Secondary School Education****, Grammar High School, Gujrat Pakistan, Years: 1997, 1998.*
3. ***Distinct Positions holder in Annual City/District Debate Competitions****, Gujrat Pakistan, Years: 1994, 1995 and 1996.*
4. ***Distinct positions holder in many school exams*** *and won some talent awards as well, Gujrat Pakistan1986-96.*

References

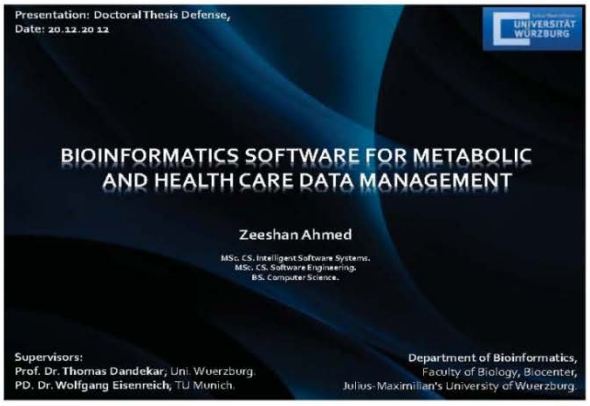
1. Prof. Dr. Thomas Dandekar (Doctoral Thesis Supervisor)
Chair Department of Bioinformatics, Biocenter, University of Wuerzburg.
Email: dandekar@biozentrum.uni-wuerzburg.de

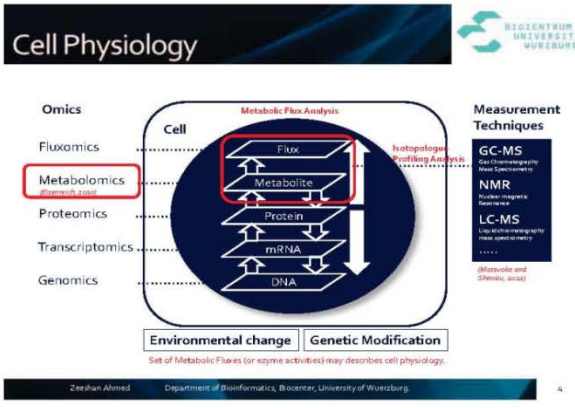
2. PD. Dr. Wolfgang Eisenreich (Doctoral Thesis Supervisor)
Department of Biochemistry, Technical University of Munich.
Email: wolfgang.eisenreich@mytum.de

Place, Date
Würzburg, 02 October 2012

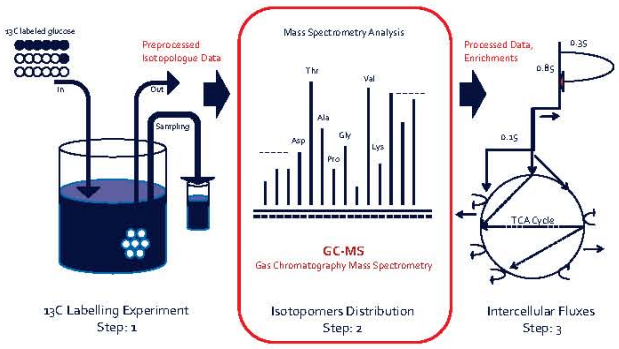
Signature:
(Zeeshan Ahmed)

Additional Material: Doctoral Thesis Presentation





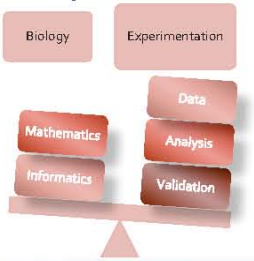
^{13}C Isotopologue Profiling Analysis



(Eisenreich, 2010; Matsuoka, 2012)

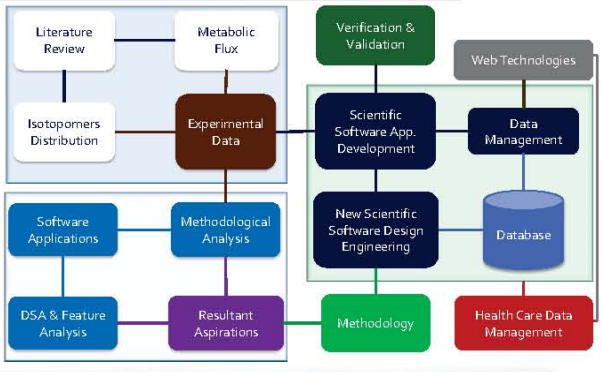
Motivation and Scope

Computer Science approaches (*software, database, management systems*) are powerful tools to advance Bioinformatic research. Here they are needed to be applied to metabolic modeling in infections as well as health care management.



- Metabolic flux prediction and pathway analysis
- Isotopomer distribution estimation from spectral data
- Quantification of relative abundances of molecular species
- Experimental isotopomer data extraction and management
- Product Data Management towards Health Care Management Systems

Research Plan



Software towards MFA

A key method providing important insights into metabolism and adaptations of different organisms in metabolic modeling. *(W. Eisenreich, T. Dandekar et al., 2010)*

Different Software Applications:

- **MetaTool** *(Schuster and Schuster, 1993)*,
- **C13** *(Wiechert et al. 1999)*,
- **Classical and Dynamic FBA** *(Mahadevan. et al. 2002)*,
- **VANTED** *(Junker, B.H., et al. 2006)*,
- **FiatFlux** *(Zamboni. et al. 2005; Zamboni. 2007)*,
- **YANASquare** *(Schwarz R, et al. 2007)*,
- **efmtool***(Terzer and Stelling, 2008)* ,
- **ReMatch** *(Esa Pitkänen. et al. 2008)*,
- **BioLayout Express 3d** *(Theocharidis A et al., 2009)*.
- **BioOPT** *(Cvijovic. et al. 2010)*,
- **iMAT** *(Zur et al., 2010)*,
- **COBRA Toolbox** *(Cvijovic et al. 2010)*

Key theoretical concepts e.g. metabolic modeling, flux analysis, networks and pathways etc.

Challenges e.g. calculation of metabolic fluxes, elementary mode analysis of large-scale networks, regulatory interactions and detailed kinetics etc.

Methodologies e.g. Bonarius et al.1996; Gulik & Heijnen 1995; Savinell & Palsson, 1992; Urbanczik et al., 2005; etc.

Feature & Architecture Analysis e.g. Data analysis steps, Graphical user interface, Running mode, algorithms, data standards, data management etc.

Evaluation e.g. Comparisons, aspirations, strengths, limitations, current and future developments etc.

Dandekar T, Fieselmann A, Saman M and Zeeshan A. *Software Applications toward Quantitative Metabolic Flux Analysis and Visualization*. Briefings in Bioinformatics, Oxford Uni. Press, First Online Published Nov. 9, 2011.
Zeeshan A, et al., *Computational Feature Performance and Domain Specific Architecture Evaluation of Software Applications towards Metabolic Flux Analysis*. Recent Patents on Computer Science, Bentham Science, 5(3),2012.

Software Specifications & Advantages

Software	Description
C13	Estimate fluxes satisfying stochiometric constraints, resolve limited enrichments by isotope balances around carbon atoms and computes deviation between fluxes and between fractional labeling
BioLayout Express 3D	Handles up to 500 to 5000 expression arrays and using MCL produces cluster of graphs (50-000 nodes), construct network graphs and Render pathways in 2D and 3D modes and calculates correlation matrix and identify genes of interest and export list of selected genes.
BioOpt	Calculates all internal mass balance fluxes using Metatool 4.3, identifies best set of gene deletions for given objective function value. Implements exhaustive combinatorial search for combinations of gene deletions.
COBRA Toolbox	Performs C13 analysis, metabolic engineering, simulation analysis of phenotypes and constraint-based modeling and reconstruction including network gap filling. Moreover creates genome-scale models and does omics guided analysis.
efmtool	Provides fast elementary mode calculation, introduces bit pattern trees and rank updating method well suited for parallel computation.
FiatFlux	Computes metabolic flux ratios exclusively from MS data. In the RATIO module, estimates net carbon fluxes within a comprehensive model of metabolite balances from measured extracellular fluxes and estimates error using Flux ratio from 13C labeling.
iMAT	Integrates transcriptomic and proteomic data with genome-scale metabolic network models and optimizes fit of gene expression or proteomic data to identify stronger and weaker fluxes. Moreover provides Ready made metabolic networks for more then 160 organisms
Metatool	Provides fast and simple elementary flux mode calculations, parses reaction equations and translates them into a stoichometric matrix, and capable of computing structural invariants like conservation relations, enzyme subsets and fits a power law to the connectivity distribution of metabolites.
ReMatch	Capable of metabolic network model construction, store and sharing. Generates stochiometric matrix and metabolic network visualizations. Integrates carbon mappings for 13C metabolic flux analysis and allows combining user developed models from several comprehensive metabolic data resources into a common repository for metabolic network models.
VANTED	Provides the visualization of flux distributions . Allows user to integrate complex structured data sets and connect several values to one single network element by presenting them as, e.g. line- or bar-charts. Supported input and output network formats are, e.g.GML, SBML and Pajek NET.
YANA programs	Provides the elementary mode specific visualization of biological networks by, e.g. distinguishing internal and external species with different node styles and colors. Performs internal elementary calculation using Metatool.

Dandekar T, Fieselmann A, Saman M and Zeeshan A. *Software Applications toward Quantitative Metabolic Flux Analysis and Visualization*. Briefings in Bioinformatics, 1st Pub., Nov. 9, 2011.

Computational Feature Comparison

Features / Software	C13	Metatool 4.3	Metatool 5.1	BioOpt	FiatFlux	ReMatch	Biolayout Express 3D
Desktop application	Yes (MatLab)	Yes	Yes (MatLab)	Yes (C++)	Yes	No	Yes
Web application	No	No	No	No	No	Yes	No
Database application	No	No	No	No	No	Yes	No
Third part tool depended	Yes (MatLab)	No	Yes (MatLab)	Yes	Yes (MatLab)	No	No
Reusable application	No	Yes	No	Yes	No	No	No
User friendly GUI	No	No	No	No	No	Yes	Yes
Platform independent	Yes	Yes	Yes	Yes	Yes	Yes	No
Standard input format	No	No	Yes (SBML)	No	Yes (CDF)	Yes (SBL)	Yes (Owl & SIF)
Standard output format	No	No	No	No	Yes(CSV & Excel)	Yes (SBML)	No
Visual output	No	No	No	No	Yes	Yes	Yes
Successful evaluation	No	Yes	Yes	Yes	Yes	Yes	Yes
Used methodologies	Isotopic Labelling	Schuster Algorithm	Null Space Mass Balance Equation	Isotopic labelling	Carbon Mapping	Markov Clustering	
Well explained for use	No	Yes	Yes	Yes	No	Yes	Yes
Running mode	Interactive	Interactive	Interactive	Batch	Interactive	Interactive	Interactive
Algorithm	Parallel	Sequential	Sequential	Sequential	Parallel	Sequential	Parallel
Publishing	Free	Free	Free	Free, Licensed	Free, Licensed	Free	Free, Licensed
Open Source	Yes	Yes	Yes	No	Yes	No	Yes

Zeeshan A, et al., *Computational Feature Performance and Domain Specific Architecture Evaluation of Software Applications towards Metabolic Flux Analysis*. RPCS, 5(3),2012.

Challenges and Solutions

The discussed software are beneficial in many ways for metabolic flux analysis and visualization. But the comparison shows clearly room for further software application development including steps towards an optimal user friendly graphical user interface, framework construction, database management system and third party independence especially in the case of desktop applications.

Solutions available for preprocessed isotopologue data processing, for raw data analysis observed during GC-MS:

- **AMDIS** *(Stein, 1993)*,
- **MSFACTs** *(Duran et al., 2003)*,
- **Envelop** *(Sykes and Williamson, 2008)*,
- **Isotope Pattern Calculator** *(Kamalrudin et al., 2008)*,
- **TagFinder***(Luedemann et al., 2008)*
- **InMass** *(Strohalm et al., 2010)* ,
- **PyMS** *(Callaghan et al., 2012)*

Metabolic reconstruction e.g. KEGG-DB & Browser, Yana processes SBML

Analysis of network structure and modeling e.g. EFM tool, iMAT, ReMatch

Flux balance and constraint based modeling e.g. FBA, OpenFlux, BioOpt,

Detailed modeling of metabolism and its kinetics e.g. Cobra toolbox, Yana, Metatool

Verifying flux predictions: Processed Isotopologue data e.g. FiatFlux, C13

Dandekar T, Fieselmann A, Saman M and Zeeshan A. *Software Applications toward Quantitative Metabolic Flux Analysis and Visualization*. Briefings in Bioinformatics, Oxford Uni. Press, First Online Published Nov. 9, 2011.
Zeeshan A, et al., *Computational Feature Performance and Domain Specific Architecture Evaluation of Software Applications towards Metabolic Flux Analysis*. Recent Patents on Computer Science, Bentham Science, 5(3),2012.

Isotopomer Distribution Analysis

- **Isotopologue** *(Isotopic Isomers) are species of a compound that differ only in their isotopic composition (Brennikmeijer et al., 2003).*
- **13C mass isotopomer distribution** *helps in metabolic network analysis for flux estimation by examining pathway activities and enzymes pass through by ancestor molecule (Biquette et al., 2006).*
- **Isotopologue measurements** can rapidly translate into metabolic flux prediction applying the effective computation software platform with good application potential for microbiology and biotechnology.

Spectral Data (GC-MS) Processing *e.g.* mass to charge ratio (m/e or m/z) values, fragments, experimental relative intensities (or abundance) of the labelled compounds.

Abundance Measurements *e.g.* natural abundance, relative abundances, fractional molar abundances.

Enrichment Estimations *e.g.* global isotope enrichments (absolute etc.).

Visualization *e.g.* Spectrum bar, line, curve charts etc.

Spectral Data Management *e.g.* File based data handling, normalized relations, database management system.

Natural Abundance Estimation

Definition:- *"Natural abundance" denotes a theoretical value, which is the complete population of isotopomers in the molecules of a given compound (including label derived isotopomers but without artificially added isotopomers).*

Binomial expression (*Lee et al., 1991*) used for calculating natural abundances.

$$A = n! / [(i!) * (n - i)!] * P_0^{(n-i)} * P_1^{i} \qquad \textit{(eq. 1)}$$

A = Calculated Relative Natural Abundance,
i = count value, (loop from 0 till n-1),
p= proportion of labeled carbon,
n = number of fragments.

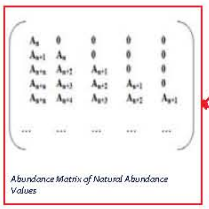
Zeeshan A. et al., "*Software LS MIDA for efficient Mass Isotopomer Distribution Analysis*". BMC Bioinformatics. (in review)
Zeeshan A. et al., "*A platform for facile Quantitative Mass Isotopomers Distribution Analysis*". Bioinformatics. (in submission)

Relative Abundance Estimation (s1)

Relative contribution of isotopomers is then calculated using Braumann's least squares algorithm (*Brauman, 1966*), (*Korzekwa et al., 1990*).

$$X = (A^T * A)^{-1} A^T P \qquad \textit{(eq. 2)}$$

X = Calculated Relative Abundance (Intensity values).
A = Abundance Matrix of Natural Abundance Values
P = Experimental Relative Intensities.



Abundance Matrix of Natural Abundance Values

$$\begin{matrix} X_0Y_0 & -Z_0 \\ X_0Y_1 + X_1Y_0 & -Z_1 \\ X_0Y_2 + X_1Y_1 + X_2Y_0 & -Z_2 \\ \cdots & \\ X_nY_{m-1} + X_{n-1}Y_m & -Z_{(n+m-1)} \\ X_nY_m & -Z_{(n+m)} \end{matrix}$$

(Brauman, 1966), (Korzekwa et al., 1990).

Relative Abundance Estimation (s2)

Definition: *"relative abundance" is a vector calculated that refers to the population of artificially labeled isotopomers (e.g. by 13C) in the molecules of a compound.*

Partial Braumann's least squares algorithm (*Brauman, 1966*), with pseudo inverse.

$$X = A^{-1} * P \qquad (eq.\ 3)$$

X = Calculated Relative Abundance (Intensity values),
A = Abundance Matrix of Natural Abundance Values,
P = Experimental Relative Intensities.

Inverse = Pseudo Inverse

Fractional Molar & Minimum Val.

Definition: *"fractional molar abundance" means the concentration of a molecular species as a fraction of the total number of molecules (Lee et al., 1990).*

To calculate fractional molar abundance, the equation is

$$F = Ri * Ra \qquad (eq.\ 4)$$

F = fractional molar abundance
Ri = relative intensity values,
Ra = measured relative abundance values.

A minimum difference validates the result and indicates convergence.

$$MinVal = F - R \qquad (eq.\ 5)$$

F = fractional molar abundance values,
R = relative intensity value or relative abundance values
Note: Standard in first transition and actual in second transition.

Absolute Enrichment

The equation to compute Absolute 13C enrichment of both natural and relative abundances is:

$$Abs\ ^{13}C = (\sum A_{o \ldots n} * n)/\alpha \qquad (eq.\ 6)$$

A labeled isotopomer,
n o, till the carbon atom of the amino acid fragment number,
α carbon atom of the amino acid fragment number.

Mathematical Validation

To mathematically validate the calculations, a new linear regression analysis is performed. A new abundance matrix is drawn:

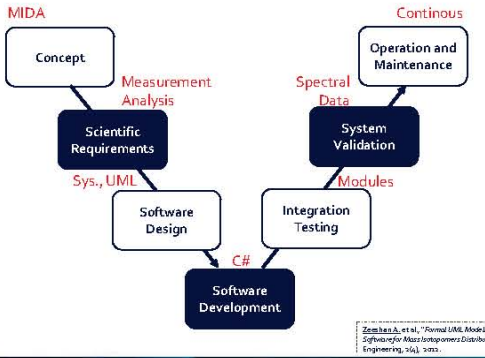
$$\begin{pmatrix} N_{11} & 0 & 0 & 0 \\ N_{12} & N_{21} & 0 & 0 \\ N_{12} & N_{22} & N_{3n+1} & N_{n+1} \\ \ldots & \ldots & \ldots & \ldots \\ N_{1n} & N_{2n} & N_{3n} & N_{nn} \end{pmatrix}$$

N1, N2, N3… Nn, are the estimated natural abundance values.

The length of Abundance Matrix is equal to the number of fragments.

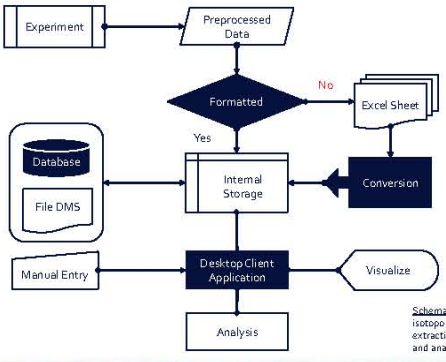
DESIGN & IMPLEMENTATION

Software Development, V-Model



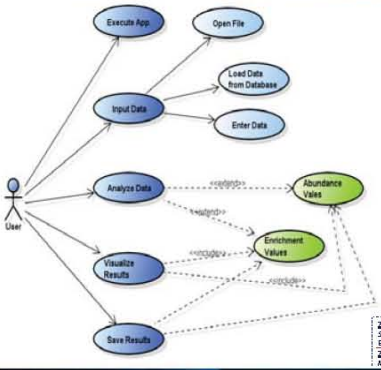
Zeeshan A. et al, "*Formal UML Modeling of Isotopo, Bioinformatical Software for Mass Isotopomers Distribution Analysis*". Software Engineering, 2(4), 2012.

Conceptual System Design



Schematic structure for isotopomers experimental data extraction, storage, management, and analysis.

UML Use Case



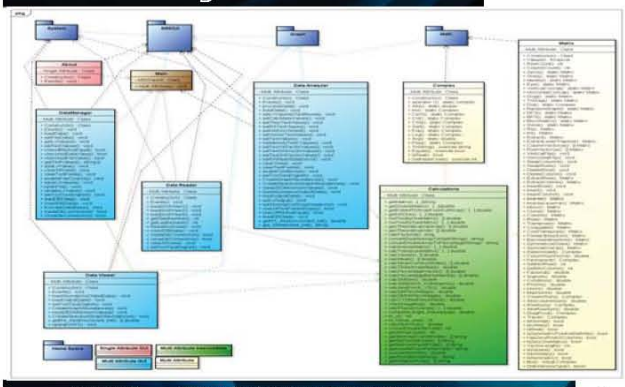
Use Case Table

Use Case	Details
Number	1.
Name	Data analyze
Application	MIDA
Description	This use case consist of 2 Users, five direct and four remote (indirect) activities. This describes, how user & the system communication for the isotopomers experimental data analysis.
Primary Actor	User(s Actor)
Precondition	Software application successfully running
Trigger / Events	1. Execute App. 2. Input Data 3. Open File 4. Load from Database 5. Enter data 6. Analyze data 7. Abundance Values 8. Enrichment Values 9. Visualize Results 10. Save Results.
Basic Flow	Basic flow consists of following steps: 1. Start software application 2. Enter input data by either loading from data file or by manually entering. 3. Analyze input data. 4. Visualize results. 5. Observe predicted results (text and image). 6. Save obtained results for reuse.
Alternate flows	Exception will be notified to the user.

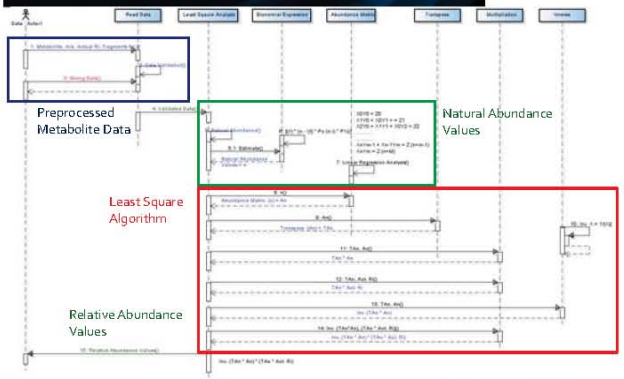
Zeeshan A. et al, "*Formal UML Modeling of Isotopo, Bioinformatical Software for Mass Isotopomers Distribution Analysis*". Software Engineering, 2(4), 2012.

Zeeshan A. et al, "*Unified Modeling and HCI Mockup Designing towards MIDA*". I. J. Eng. Sci, 1(6), 2012.

UML Class Diagram



UML Sequence Diagram (s1)

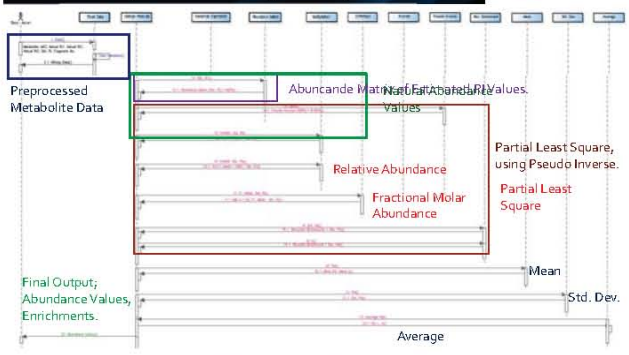


LS-MIDA (Prototype)

GUI of Data Analyzer (a) presents the main graphical user interface responsible for handling user data input, analyzing and producing spectrum, along with the tab based graphical user interface module for isotopomer data. (b) and complete application's output data viewing and manipulation.

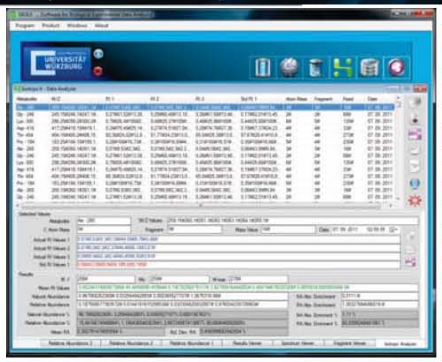
Zeeshan A. et al, "*Software LS-MIDA for efficient Mass Isotopomer Distribution Analysis*". BMC Bioinformatics. (in revision)

UML Sequence Diagram (s2)



Zeeshan A. et al, "*A platform for facile Quantitative Mass Isotopomers Distribution Analysis*". Bioinformatics. (in submission)

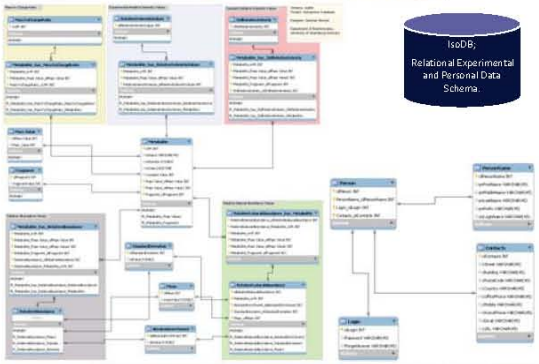
Isotopo Data Analyzer



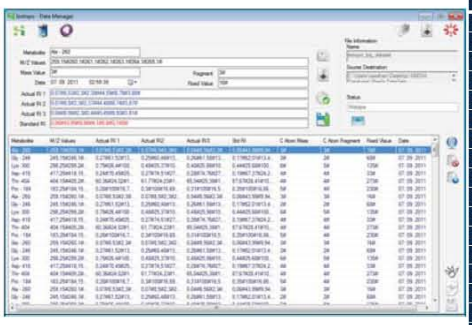
Features / Software	Isotopo
Desktop application	Yes
Web application	Yes (Client)
Database application	Yes
Third part tool depended	No
Reusable application	Yes
User friendly GUI	Yes
Platform independent	No
Standard input format	Yes
Standard output format	Yes
Visual output	Yes
Successful evaluation	Yes
Used methodologies	MVS.
Well explained for use	Yes
Running mode	Batch
Algorithm	Sequential
Publishing	Free, Licensed
Open Source	No

Zeeshan A. et al., *"A platform for facile Quantitative Mass Isotopomers Distribution Analysis"*. Bioinformatics. (in submission)

Normalized Entity Relationships



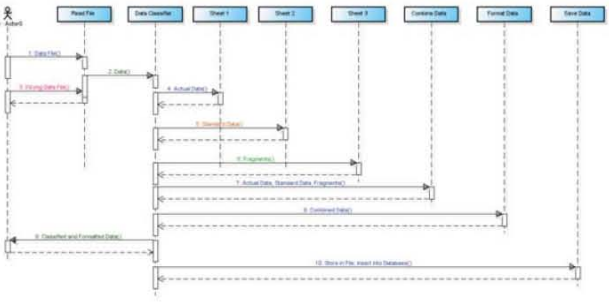
Isotopo Data Manager



Features / Software	Isotopo
Desktop application	Yes
Web application	Yes (Client)
Database application	Yes
Third part tool depended	No
Reusable application	Yes
User friendly GUI	Yes
Platform independent	No
Standard input format	Yes
Standard output format	Yes
Visual output	Yes
Successful evaluation	Yes
Used methodologies	NA
Well explained for use	Yes
Running mode	Batch
Algorithm	NA
Publishing	Free, Licensed
Open Source	No

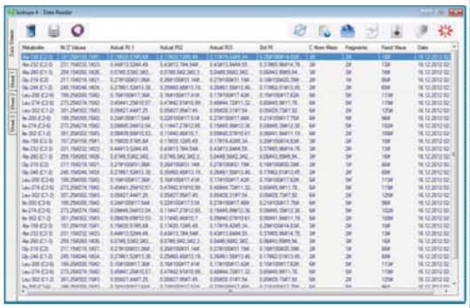
Zeeshan A. et al., *"A platform for facile Quantitative Mass Isotopomers Distribution Analysis"*. Bioinformatics. (in submission)

SSD Supervised Data Classifier



Zeeshan A. et al., *"Supervised Classifier towards Effective Preprocessed Spectral Data Extraction and Management"*. Database. (in submission)

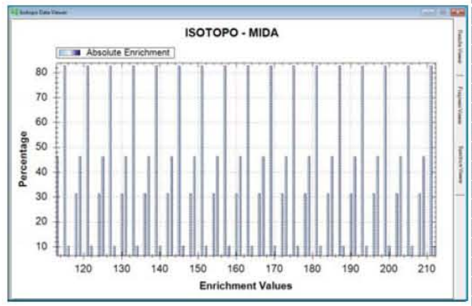
Isotopo Data Reader



Features / Software	Isotopo
Desktop application	Yes
Web application	Yes (Client)
Database application	Yes
Third part tool depended	No
Reusable application	Yes
User friendly GUI	Yes
Platform independent	No
Standard input format	Yes
Standard output format	Yes
Visual output	Yes
Successful evaluation	Yes
Used methodologies	NA
Well explained for use	Yes
Running mode	Batch
Algorithm	Sequential
Publishing	Free, Licensed
Open Source	No

Zeeshan A. et al., "*Supervised Classifier towards Effective Preprocessed Spectral Data Extraction and Management*". Database. (in submission)

Isotopo Data Viewer



Features / Software	Isotopo
Desktop application	Yes
Web application	Yes (Client)
Database application	Yes
Third part tool depended	No
Reusable application	Yes
User friendly GUI	Yes
Platform independent	No
Standard input format	Yes
Standard output format	Yes
Visual output	Yes
Successful evaluation	Yes
Used methodologies	NA
Well explained for use	Yes
Running mode	Batch
Algorithm	NA
Publishing	Free, Licensed
Open Source	No

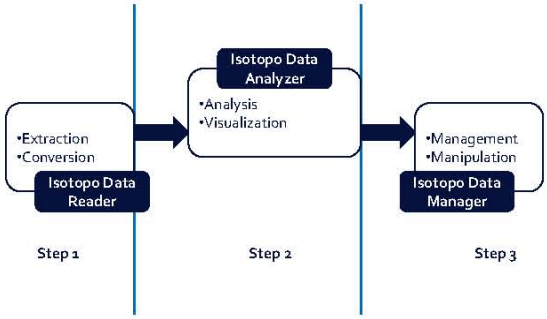
Zeeshan A. et al., "*A platform for facile Quantitative Mass Isotopomers Distribution Analysis*". Bioinformatics. (in submission)

EXPERIMENTATION & VALIDATION

Experimental Data; Salmonella

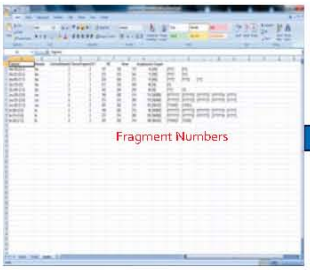
Metabolite	m/z	Ri1	Ri2	Ri3	Std. Ri	Atom Mass	Frag ment
Alanine (Ala-158)	157,25#158,15#159,15#160,1#161,1#162,1#163,1#	0,19#20,57#5,6#100#15,53#4,83#0,64#	0,17#20,12#5,49#100#15,6#4,84#0,64#	0,17#19,42#5,34#100#15,44#4,78#0,62#	0,29#100#14,83#7,18#0,9#0,29#0,06#	3#	2#
Alanine (Ala-232)	231,15#232,1#233,1#234,1#235,1#236,1#	0,44#13,52#4,49#71,54#14,02#6#	0,43#13,7#4,54#72,74#14,2#6,08#	0,43#13,84#4,55#73,7#14,47#6,19#	0,37#65,96#14,78#5,98#0,87#0,15#	3#	2#
Alanine (Ala-260)	259,15#260,1#261,1#262,1#263,1#264,1#265,1#	0,07#8,53#2,3#2,39#44,59#8,79#3,88#	0,07#8,5#2,3#2,37#44,48#8,74#3,87#	0,04#8,56#2,3#2,4#45,45#8,93#3,91#	0,06#43,99#9,94#4,1#0,6#0,1#0#	3#	3#
Glycine (Gly-218)	217,15#218,1#219,1#220,1#221,05#	0,27#100#31,06#10,72#1,93#	0,26#100#31,14#10,67#1,96#	0,27#100#31,19#10,72#2,02#	0,18#100#20,39#8,63#1,15#	2#	1#
Glycine (Gly-246)	245,15#246,1#247,1#248,1#249,05#250,1#	0,27#61,52#13,38#11,9#1,91#0,72#	0,25#60,48#13,18#11,92#1,94#0,73#	0,26#61,58#13,46#12,15#2,03#0,73#	0,17#62,01#13,45#5,67#0,77#0,14#	2#	2#
Leucine (Leu-200)	199,25#200,15#201,15#202,15#203,15#204,1#205,15#206,15#207,15#	0,15#100#17,36#5,59#0,69#0,13#0,02#0,01#0,01#	0,16#100#17,41#5,65#0,71#0,13#0,02#0,01#0,02#	0,17#100#17,42#5,67#0,71#0,13#0,02#0,01#0,01#	0,15#100#17,62#5,41#0,66#0,11#0,02#0,01#0,03#	6#	5#
Leucine (Leu-274)	273,25#274,15#275,15#276,15#277,15#278,15#279,15#280,15#281,15#	0,45#41,25#10,57#4,06#0,65#0,11#0,01#0#0#	0,47#42,91#10,99#4,22#0,68#0,11#0,01#0#0,01#	0,48#44,72#11,32#4,3#0,69#0,11#0,01#0#0,01#	0,65#45,9#11,78#4,54#0,75#0,13#0,01#0,01#0,01#	6#	5#
Leucine (Leu-302)	301,25#302,15#303,15#304,15#305,15#306,15#	0,05#27,44#7,25#2,83#0,47#0,08#	0,05#27,95#7,45#2,88#0,49#0,08#	0,05#28,31#7,54#2,92#0,49#0,09#	0,05#29,73#7,92#3,06#0,53#0,09#	6#	2#
Isoleucine (Ile-200)	199,25#200,15#201,15#202,15#203,15#204,1#205,15#206,15#207,15#	0,24#100#17,54#5,18#0,7#0,17#0,07#0,03#0,86#	0,22#100#17,51#5,25#0,68#0,18#0,04#0,02#0,69#	0,27#100#17,46#5,22#0,69#0,18#0,04#0,05#0,68#	0,21#100#17,75#5#0,64#0,07#0,44#0,11#0,03#	6#	5#
Isoleucine (Ile-274)	273,25#274,15#275,15#276,15#277,15#278,15#279,15#280,15#281,15#	0,09#45,04#12,04#4,53#0,77#0,16#0,07#0,01#0,01#	0,11#47,27#12,65#4,78#0,79#0,16#0,05#0,03#0,03#	0,15#45,99#12,36#4,66#0,82#0,17#0,05#0,02#0,03#	0,08#45,39#12,38#4,62#0,8#0,12#0#0,01#0,01#	6#	5#
Isoleucine (Ile-302)	301,25#302,15#303,15#304,15#305,15#306,15#	0,08#39,69#10,53#4,09#0,68#0,12#	0,11#40,46#10,7#4,15#0,72#0,11#	0,09#40,07#10,61#4,13#0,7#0,13#	0,06#41,94#11,19#4,34#0,82#0,13#	6#	2#

Validation Procedure

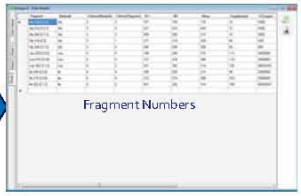


Data Extraction and Conversion

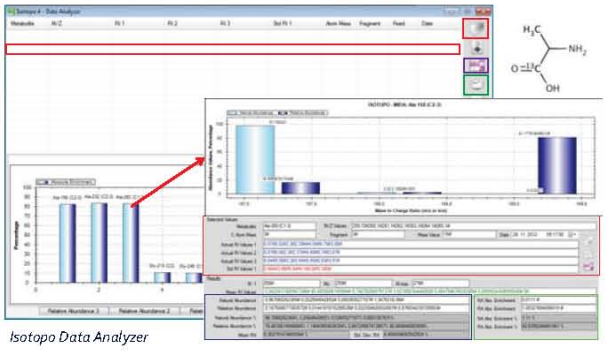
Preprocessed Metabolite Data



Isotopo Data Reader

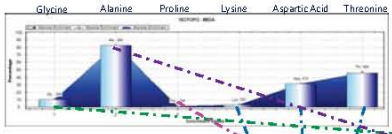


Data Analysis and Visualization



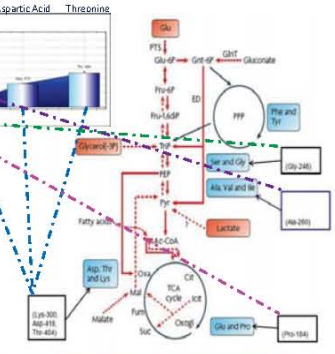
Isotopo Data Analyzer

Metabolic Modeling, Salmonella

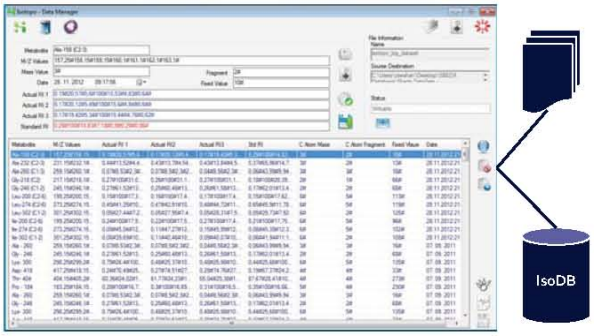


Starting from a basic model of Salmonella (enterica subsp. Enterica serovar Typhimurium) metabolism (Eisenreich et al., 2010) we show here current investigations and the position of the data examples given on the simplified metabolic map of Salmonella.

Slide 5:9C Isotopologue Profiling Analysis

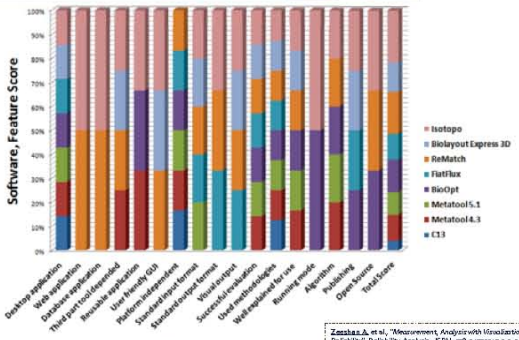


Data Management & Manipulation



Isotopo Data Manager

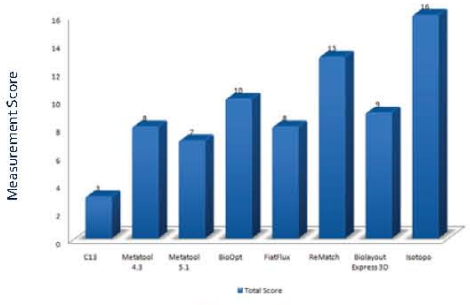
Performance Measurement



Zeeshan A. et al, *"Measurement, Analysis with Visualization for better Reliability"*. Reliability Analysis , ISBN: 978 1477554173 1, 2012.

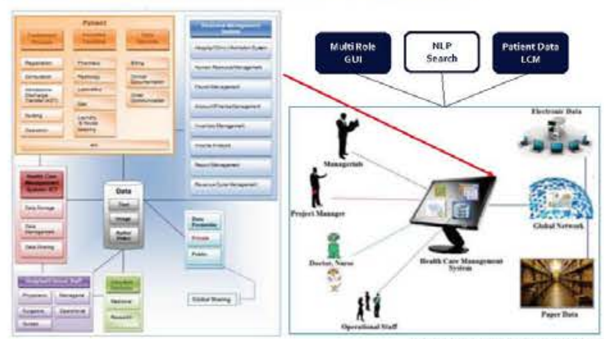
Zeeshan A., *"Towards Performance Measurement And Metrics Based Analysis of PLA Applications"*. Jr. Soft. Eng. App., Vol1, No 3, 2010.

Pareto Analysis



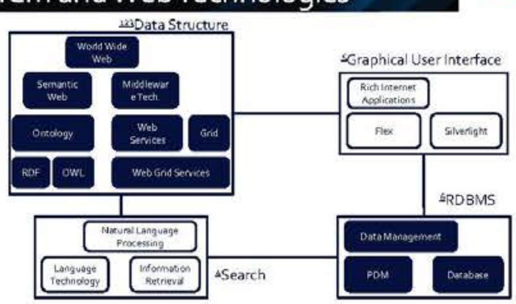
HEALTH CARE MANAGEMENT (HCM)

HCM System Basic Infrastructure



Zeeshan A., et al.,"ADAM: Transiting PDM into Clinical Patient Data Management". I. J. Eng. Sci., 1(1), 2012

HCM and Web Technologies



[1] Zeeshan A., et al. "Semantic web, Ontology Specific Languages for Web Application Development". Jr. Web App., 4(1), 2011

[2] Zeeshan A., et al. "Middleware Technologies, Chain Web Grid Services". Jr. Web App., V(4), 2011

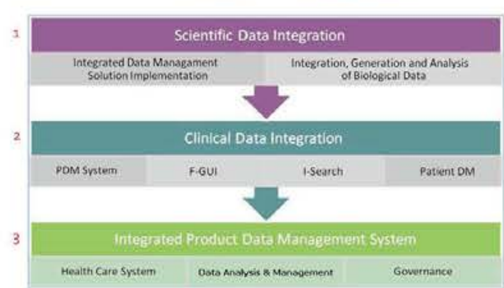
[3] Zeeshan A., "Web to Semantic Web and Role of Ontology in its Development", Semantic Web, 2010

[4] Zeeshan A., "Proposing LT based Search in PDM Systems for Better Information Retrieval", Comp. Sci. E. Tech., 1(4), 2010

[4] Zeeshan A., "Designing Flexible GUI to Increase the Acceptance Rate of Product Data Management Systems in Industry", C. Sci. E. Tech., 2(1), 2011

[4] Zeeshan A., "Proposing Semantic Oriented Agent and Knowledge base Product Data Management", IMCS Jr. I, 1(6), 2009

Layer Architecture



Zeeshan A., "I-Hit and LT-SOS in PDM Systems", ISBN-13: 978-3-639-31070-2, 2011

Zeeshan A., et al., "Contributions of PDM Systems in Organizational Technical Data Management", IEEE IC4, 2007

FUTURE RECOMMENDATIONS

Future Work

BIOZENTRUM UNIVERSITÄT WÜRZBURG

Isotopomer Distribution Analysis

- NMR Data Analysis,
- Other format parsers (SBML),
- Web data structuring, sharing and search.

Health Care Management

- System and UML designs,
- Implementations,
- Technical and research documentations.

CLOSING PRESENTATION

Acknowledgments

BIOZENTRUM UNIVERSITÄT WÜRZBURG

Prof. Dr. Thomas Dandekar
his group, department and especially Saman Majeed.

UNIVERSITÄT WÜRZBURG

PD. Dr. Wolfgang Eisenreich
his group, department and especially Eva Eylert.

TUM

Family, Friends and Colleagues

Deutsche Forschungsgemeinschaft DFG

Pathogen

Host

transregio 34

Bundesministerium für Bildung und Forschung