Implementation of Bioinformatics Methods for miRNA and Metabolic Modelling

Dissertation zur Erlangung des

naturwissenschaftlichen Doktorgrades

der Julius-Maximilians-Universität Würzburg

vorgelegt von

Saman Zeeshan

(Maiden Majeed)

aus Rawalpindi, PK

Würzburg, 2014

Eingereicht am: …**4 September 2014** …

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

Mitglieder der Promotionskommission:

        Vorsitzende: …
        1. Gutachter: Prof. Dr. Thomas Dandekar
        2. Gutachter: Prof. Dr. Harald Schulze

Tag des öffentlichen Promotionskolloquiums: …**13 October 2014** ...

Doktorurkunde ausgehändigt am: …**13 October 2014** ...
wissenschaftliche Befähigung nachgewiesen und dabei das Gesamturteil
‚**magna cum laude'**

verleiht unter dem Präsidenten Professor Dr. rer. nat. Alfred Forchel
durch den Dekan der Fakultät für Biologie Professor Dr. rer. nat. Markus Engstler

AFFIDAVIT / EIDESSTATTLICHE ERKLÄRUNG

English:

I hereby confirm that my thesis entitled "*Implementation of Bioinformatics Methods for miRNA and Metabolic Modelling*" 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.

Place, Date   Signature:

Wuerzburg, 04/09/ 2014**.**   Saman Zeeshan (maiden Majeed)

German:

Hiermit erkläre ich an Eides statt, die Dissertation "*Die Umsetzung der Bioinformatik-Methoden für miRNA-und der Metabolischen Modellierung*" 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.

Ort, Datum   Unterschrift:

Wuerzburg, 04/09/ 2014**.**   Saman Zeeshan (maiden Majeed)

SUMMARY

Dynamic interactions and their changes are at the forefront of current research in bioinformatics and systems biology. This thesis focusses on two particular dynamic aspects of cellular adaptation: miRNA and metabolites.

miRNAs have an established role in hematopoiesis and megakaryocytopoiesis, and platelet miRNAs have potential as tools for understanding basic mechanisms of platelet function. The thesis highlights the possible role of miRNAs in regulating protein translation in platelet lifespan with relevance to platelet apoptosis and identifying involved pathways and potential key regulatory molecules. Furthermore, corresponding miRNA/target mRNAs in murine platelets are identified. Moreover, key miRNAs involved in aortic aneurysm are predicted by similar techniques. The clinical relevance of miRNAs as biomarkers, targets, resulting later translational therapeutics, and tissue specific restrictors of genes expression in cardiovascular diseases is also discussed.

In a second part of thesis we highlight the importance of scientific software solution development in metabolic modelling and how it can be helpful in bioinformatics tool development along with software feature analysis such as performed on metabolic flux analysis applications. We proposed the "Butterfly" approach to implement efficiently scientific software programming. Using this approach, software applications were developed for quantitative Metabolic Flux Analysis and efficient Mass Isotopomer Distribution Analysis (MIDA) in metabolic modelling as well as for data management. "LS-MIDA" allows easy and efficient MIDA analysis and, with a more powerful algorithm and database, the software "Isotopo" allows efficient analysis of metabolic flows, for

instance in pathogenic bacteria (Salmonella, Listeria). All three approaches have been published

(see Appendices).

ZUSAMMENFASSUNG

Dynamische Wechselwirkungen und deren Veränderungen sind wichtige Themen der aktuellen Forschung in Bioinformatik und Systembiologie. Diese Promotionsarbeit konzentriert sich auf zwei besonders dynamische Aspekte der zellulären Anpassung: miRNA und Metabolite.

miRNAs spielen eine wichtige Rolle in der Hämatopoese und Megakaryozytopoese, und die Thrombozyten miRNAs helfen uns, grundlegende Mechanismen der Thrombozytenfunktion besser zu verstehen.

Die Arbeit analysiert die potentielle Rolle von miRNAs bei der Proteintranslation, der Thrombozytenlebensdauer sowie der Apoptose von Thrombozyten und ermöglichte die Identifizierung von beteiligten Signalwegen und möglicher regulatorischer Schlüsselmoleküle. Darüber hinaus wurden entsprechende miRNA / Ziel-mRNAs in murinen Thrombozyten systematisch gesammelt. Zudem wurden wichtige miRNAs, die am Aortenaneurysma beteiligt sein könnten, durch ähnliche Techniken vorhergesagt. Die klinische Relevanz von miRNAs als Biomarker, und resultierende potentielle Therapeutika, etwa über eine gewebsspezifische Beeinflussung der Genexpression bei Herz-Kreislauf Erkrankungen wird ebenfalls diskutiert.

In einem zweiten Teil der Dissertation wird die Bedeutung der Entwicklung wissenschaftlicher Softwarelösungen für die Stoffwechselmodellierung aufgezeigt, mit einer Software-Feature-Analyse wurden verschiedene Softwarelösungen in der Bioinformatik verglichen. Wir vorgeschlagen dann den "Butterfly"-Ansatz, um effiziente wissenschaftliche Software-Programmierung zu implementieren. Mit diesem Ansatz wurden für die quantitative Stoffflussanalyse mit Isotopomeren effiziente Software-Anwendungen und ihre Datenverwaltung

entwickelt: LS-MIDA ermöglicht eine einfache und effiziente Analyse, die Software "Isotopo" ermöglicht mit einem leistungsfähigeren Algorithmus und einer Datenbank, eine noch effizientere Analyse von Stoffwechselflüssen, zum Beispiel in pathogenen Bakterien (Salmonellen, Listerien). Alle drei Ansätze wurden bereits veröffentlicht (siehe Appendix).

ACKNOWLEDGEMENT

I dedicate this doctoral thesis to my beloved family.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

1 Introduction

1.1 Motivation and scope

The field of bioinformatics entails implementing principles and tools of computer science, statistics and mathematics into the research methods and processes of biology and related sciences. It is an interdisciplinary field in which physical sciences, life sciences, computer science, and engineering are merged to solve both fundamental and applied problems in biology and medicine. The outcomes of bioinformatics and computational biology provide novel and global perspectives into the organization and functioning of biological data. Thus, bioinformatics is emerging as a strategic discipline at the frontier between biology, biochemistry, biomedicine, bioengineering, computer science, mathematics and statistics, impacting fundamental science, medicine, and the society.

Although research in bioinformatics has soared in the last decade, much of the focus has been on high performance computing, such as optimizing algorithms and large-scale data storage techniques. Only recently have studies on end-user programming (Letondal, 2005) (Massar et al., 2005) and information activities in bioinformatics (Bartlett and Toms, 2005), (MacMullen and Denn, 2005) started to emerge. The approach in this thesis is based on our understanding of problems in bioinformatics software development, how domain knowledge among molecular biology and biochemistry and computer science experts is exchanged, and how the software development process in this domain can be improved.

Synthesizing metabolites (chemical compound produced by metabolic reaction) and resulting metabolic fluxes provide 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).

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). During this research we have investigated different bacterial systems and analysed bacterial data obtained from staphylococcus 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).

In the scope of this thesis, methods were designed and developed that allow improvement in development of biological softwares. It involved proposition of a new software model "*Butterfly*" and also refining measurements of metabolites via metabolic modelling analyzing flux pathways and predicting actual fluxes under a given situation. We performed feature comparison and analysis of flux pathways in applications C13, Classical and Dynamic FBA, MetaTool, FiatFlux, BioOPT, Rematch, VANTED and YANASquare, highlighting their advantages and limitations in metabolic

modelling. Their drawbacks included third party software usage, absence of a comprehensive database management system, lack of standard formatting for data intake and result generation, and unfriendly graphical user interface making data visualization complicated.

In this research the aim was to find better analytical approaches using mathematics and informatics for analyzing experimental results. There is also no efficient solution available for extensive metabolic flux data management and manipulation. We took the challenge to design a new data management system for managing such experimental data. The other part of this research is also focused on miRNA study in different cells and their analysis. During this research the role of miRNAs in platelet gene expression was investigated, also exploring platelet differentiation with respect to miRNA regulation of platelet apoptosis/ cell death and cell survival/ proliferation. In the long run manipulating the expression of miRNAs and downstream signaling cascades, will open up new possibilities to prevent and cure cardiovascular diseases.

The thesis deals with topics in biology and software design and engineering and is divided into three chapters. Starting from a general introduction of the theoretical foundations, the first chapter gives an account of the work performed towards scientific software design and engineering and the approaches used in their implementation. It also gives an overview of the published work highlighting some of the concepts followed. Chapter two discusses the miRNA study with different research groups and on different cells. The thesis ends with conclusions and suggestions for future work.

1.2 Theoretical Foundations

1.2.1 Platelet biology

Platelets have attracted a growing interest among basic scientists and clinicians, as they have been shown to play an important role in many physiological and pathophysiological conditions (de Groot et al., 2012). They are anucleate blood cells, but contain organelles and structures, such as mitochondria, microtubules, and granules. They have a characteristic discoid shape. A wide variety of transmembrane receptors cover the platelet membrane, including many integrins, leucine-rich repeat receptors, G protein-coupled receptors, proteins belonging to the immunoglobulin superfamily, tyrosine kinase receptors, and a variety of other types (Cimmino and Golino, 2013). Megakaryocytes are polyploid cells that develop primarily in the bone marrow, generate platelets by extruding cytoplasmic structures called proplatelets into the circulation. These filamentous processes are sheared off to become preplatelets, which undergo maturation and fission to produce individual platelets (de Groot et al., 2012). The production of platelets requires an intricate series of remodelling events that result in the release of thousands of platelets from a single megakaryocyte (Patel et al., 2005a). The whole process of assembly and release of platelets is as follows: first the megakaryocytes tailor their cytoplasm and membrane systems for platelet biogenesis, they become polyploid by endomitosis (DNA replication without cell division) and then undergo a maturation process in which the bulk of their cytoplasm is packaged into multiple long thin processes called proplatelets, and the nucleus is extruded.
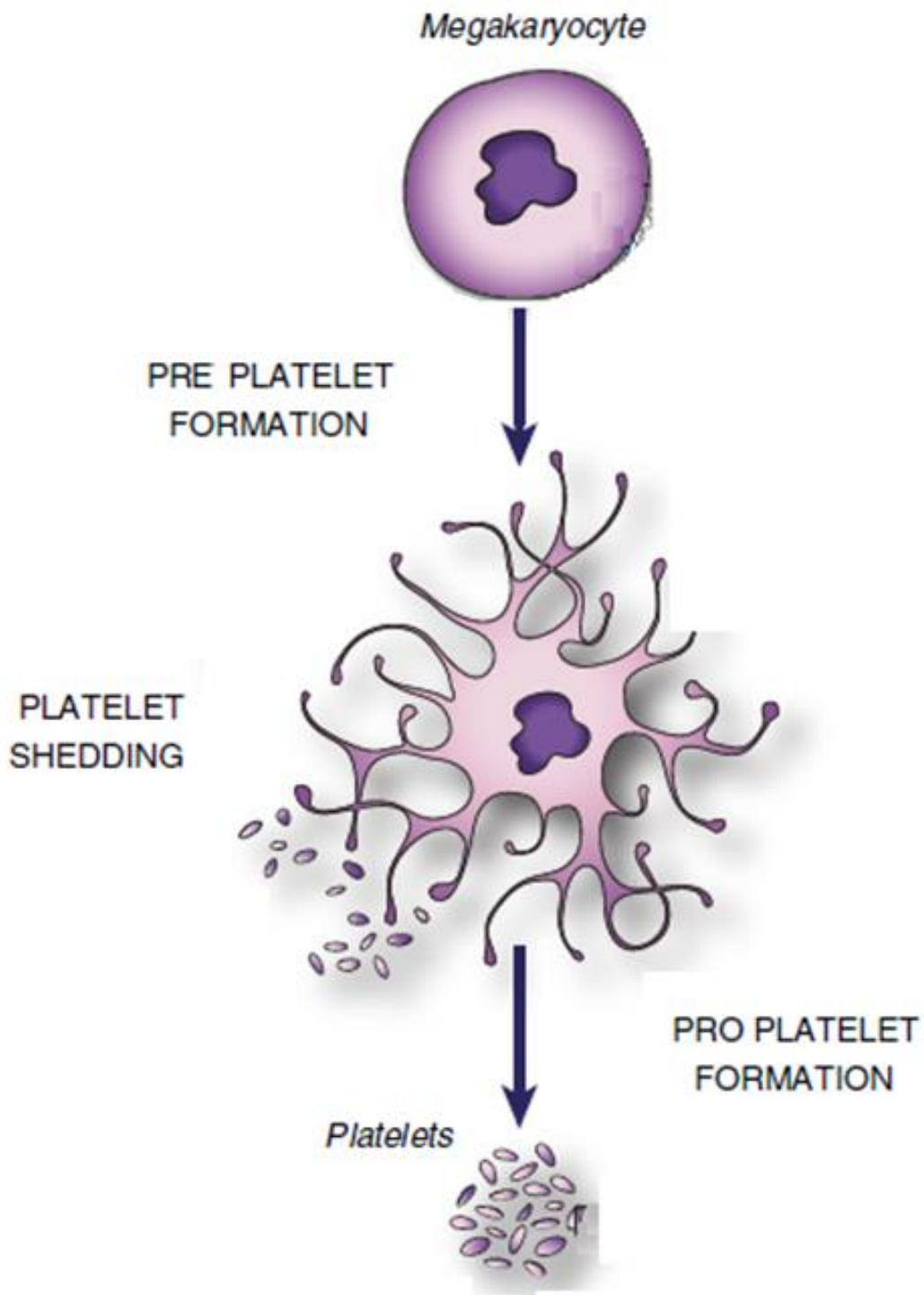
Figure 1: Platelet formation from Megakaryocytes

Proplatelets are a complex cytoskeletal protein network made of microtubules (Patel et al., 2005b). Platelets form selectively at the tips of proplatelets (Italiano et al. 1999). These filamentous processes are sheared off to become preplatelets, which undergo maturation and fission to generate discoid platelets (Figure 1).

1.2.2   miRNA

The focus of this research is on the biology of miRNAs and their role in platelets and the possible role of miRNAs in regulating protein translation in platelet lifespan with relevance to platelet apoptosis. The clinical relevance of miRNAs as biomarkers, targets, therapeutics, and tissue specific restrictors of gene expression in cardiovascular diseases is also studied.

Small regulatory RNAs include microRNAs (miRNAs), endogenous small interfering RNAs, piwi-interacting RNA, small nucleolar RNAs, and others (Mattick and Makunin, 2006). The best studied noncoding RNA are miRNAs. MiRNAs were initially discovered in C. elegans in 1993 by the laboratories of Victor Ambros and Gary Ruvkin (Wightman et al., 1993) (Lee et al., 1993). The discovery by Ambros laboratory on the role of the lin-4 and lin-14 genes in temporal control of development in the model organism Caenorhabditis elegans found that lin-4 gene does not encode a protein product, but instead gives rise to a 61-nt precursor gene that matured to a more abundant 22-nt transcript, the Ruvkun laboratory found that LIN-14 protein synthesis is regulated post transcriptionally and that LIN-14 levels are inversely proportional to those of lin-4 RNA. Sequence analysis revealed that the lin-4 RNA has sequence complementarity to the 3' untranslated region (UTR) of the lin-14 gene, leading to the hypothesis that lin-4 regulated LIN-14 through Watson–Crick base pairing, revealing the first miRNA and mRNA target interaction.

Approximately 30% of miRNA genes are located in intergenic regions, and approximately 70% are located within introns or exons of protein coding genes. They are approximately 22 nucleotides

in length and exist in premature form. RNA polymerase II is responsible for most miRNA transcription, although RNA polymerase III has been identified transcribing miRNAs located within repetitive elements (Brochert et al., 2006). A primary transcript (pri miRNA) is transcribed from DNA genes and processed in the nucleus. The primiRNA is capped and polyadenylated, and forms a hairpin structure (Cai et al., 2004) (Lee et al., 2004). The pri-miRNA is cleaved into an approximately 60- to 70-bp pre-miRNA by Drosha, an RNase type III endonuclease, which is complexed with DiGeorge syndrome critical region 8 (DGCR8) (Edelstein and Bray, 2011).

The pre-miRNA is transported out of the nucleus via exportin 5. In the cytoplasm, the 3' overhang of the pre-miRNA is recognized by the Dicer-TAR RNA-binding protein (TRBP) complex. Dicer is another RNase type III endonuclease, which generates the miRNA duplex for many miRNAs. The strands separate and the mature miRNA associates with a macromolecular complex called the RNA-induced silencing complex (RISC), which guides the miRNA to its mRNA target. The process of miRNA generation (Figure 2) is regulated at both transcriptional and posttranscriptional levels, occasionally composed of positive feed-forward or negative feedback circuitry, in which the miRNA targets a transcriptional activator or repressor of itself.

The defining RISC protein is Argonaute 2 (Ago2), which targets mRNAs for degradation by catalyzing mRNA cleavage and/or translational inhibition. Notably, Ago2-dependent, Dicer-independent miRNA processing pathways have been identified (Cheloufi et al., 2010) (Cifuentes et al., 2010). Work by Guo et al. 2010 indicates that miRNA knockdown of mammalian protein expression is primarily via mRNA degradation (Laffont et al., 2013; Edelstein et al., 2013).

The function of a miRNA is ultimately defined by the genes it targets and its effects on their expression. A given miRNA can be predicted to target several hundred genes, and 60% of mRNAs have predicted binding sites for 1 or multiple miRNAs in their UTR. Two major silencing

mechanisms have been identified for miRNAs: miRNAs can inhibit translation by inhibition of translation initiation or can target mRNAs for degradation. Under normal conditions, miRNAs act as moderate regulators that act as a regulator for gene expression, and under conditions of stress or disease, this function becomes more pronounced and important (Edelstein and Bray, 2011).

It is important to emphasize the nature of software predicted miRNA–mRNA pairs. In general, software tends to extra predict miRNA-mRNA pairs as the default settings are often put such that no potential is overlooked (low false negative rate and high false positive rate). Thus experiments are necessary for validation. Often, potentially important miRNAs are identified in an association study of a particular disease or physiologic trait.

Unlike mRNAs, the function of the miRNA is not immediately clear because miRNAs target multiple mRNAs. Most investigators utilize one or more target prediction algorithms to help understand miRNA function and to drive mechanistic experiments in cells. However, there is a lack of consensus on the optimal prediction method (Rigoutsos et al., 2011). Serving as a guidance, these predictions remain to be validated experimentally which, in some cases, may represent a relatively challenging task. Because these programs only predict putative targets, it is important to confirm these predictions using miRNA target validation techniques (van Rooij, 2011).

Combining the results of different target prediction programs to look for overlap in predicted targets between the different programs will result in the highest specificity but lowest sensitivity. On the other hand, combining the results of all programs will lead to the highest sensitivity but lowest specificity. Based on experimentally validated data sets, it has been recommended that intersecting Targetscan and PicTar predictions often results in both high sensitivity and specificity (Baek et al., 2008; Selbach et al., 2008; Sethupathy et al., 2006).
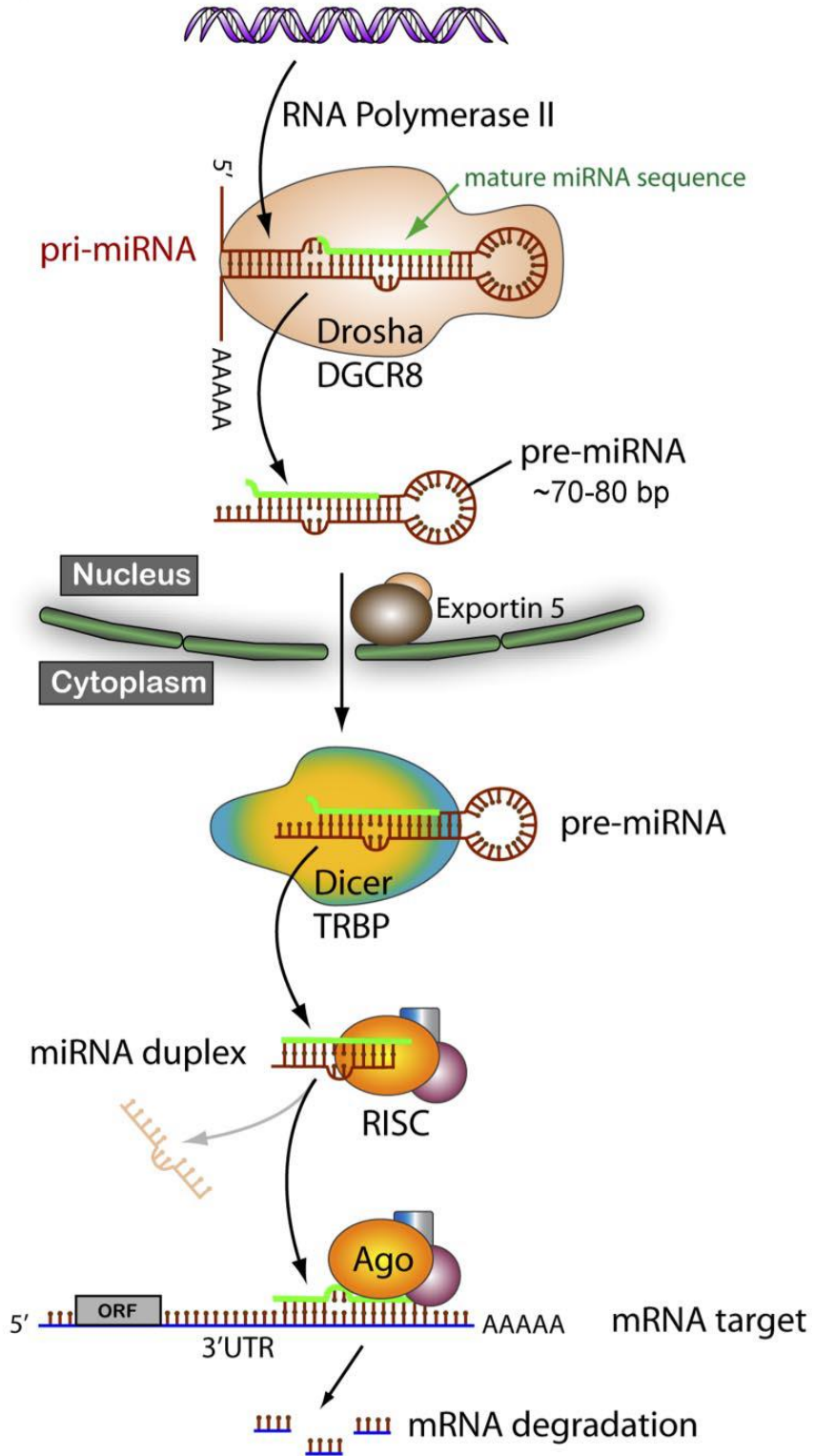
Figure 2: miRNA synthesis and function (Edelstein et al., 2013)

1.2.3 Platelet miRNAs

MicroRNAs (miRNAs) are 21-23 nucleotides RNAs that regulate more than 60% of mammalian protein coding genes. miRNAs play critical roles in hematopoiesis and megakaryocyte function and development.

Platelets are anucleated blood cells that play a crucial role in thrombosis and haemostasis. Despite their lack of nuclear DNA, platelets contain significant amounts of microRNA (miRNA) that may have vital functions in post-transcriptional gene regulation (Nagalla et al, 2011). They are also said to retain a small amount of megakaryocyte-derived messenger RNA (mRNA) that can regulate protein expression (Tanriverdi et al., 2006). Although the biology of platelet miRNAs is largely unexplored recently, platelets have been found to contain small (23 base pair) non-coding microRNAs (miRNAs) derived from hairpin-like precursors. In 2006 Tanriverdi et al., determined the presence of this system in platelets. It was found that platelets have miRNA essential proteins and their activation can regulate miRNA expression. These findings suggested a potentially new mechanism for platelet control of function via protein expression. In 2009 the regulation control of gene expression in platelet through mRNA splicing was studied by Landry et al, 2009. They proposed a gene regulatory pathway for platelets based on miRNA interactions.

MiRNAs can specifically silence their mRNA targets regulating mRNA translation. Platelet miRNAs are reported to bind to important target mRNAs involved in reactivity. They also regulate important functions such as shape change, granules secretion, and activation (Dimitrios et al., 2013).

In addition to possessing functional miRNA processing machinery, have been shown to target mRNAs that encode proteins that alter platelet function. (Edelstein and Bray, 2012).

Platelet profiling has been performed by many groups. Nagalla et al., 2011 identified 284 miRNAs in platelets. Osman and Falker (2011) identified 281 transcripts, including 228 mature miRNAs and 53 minor miRNAs. Although 284 miRNAs are described to be present in platelets, their role is mostly unknown. The most abundant miRNA in platelets is miR-223 followed by miR-126. The miR-96, miR-200b, miR- 495, miR-107 and miR-223 are critically involved in platelet reactivity, aggregation, secretion and adhesion (Gatsiou et al., 2012).

1.2.4 Metabolic modelling and spectral data analysis

There are five different and important levels of information in the cell, as shown in Figure 3 (Matsuoka and Kazuyuki, 2012). These levels are Genomics, Transcriptomics, Proteomics, Metabolomics and Fluxomics. As miRNAs, the dynamic regulation of metabolites has become a recent focus of omics investigations.
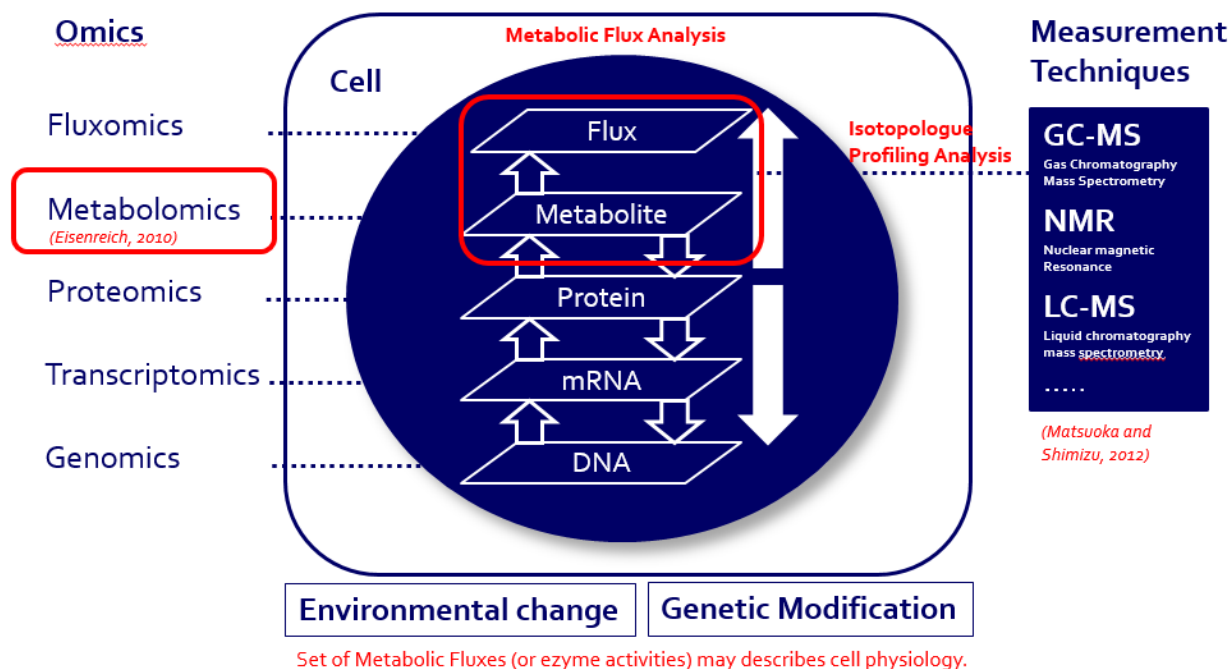
Figure 3: OMICS; different levels of information in the cell

A second focus of this doctoral study is hence to perform intensive research and development in the fields of Quantitative Metabolic Flux Analysis and Mass Isotopomers Distribution Analysis. As shown in Figure 4, the study plan started with related literature reviewer of Metabolic Flux and Isotopomer Distributions Analysis, experimental data is then analysed using different scientific software solutions implementing different mathematical and statistical methodologies. Based on the resulting goals and limitations to the existing solutions, some new software solutions and data management systems were planned, designed and implemented.
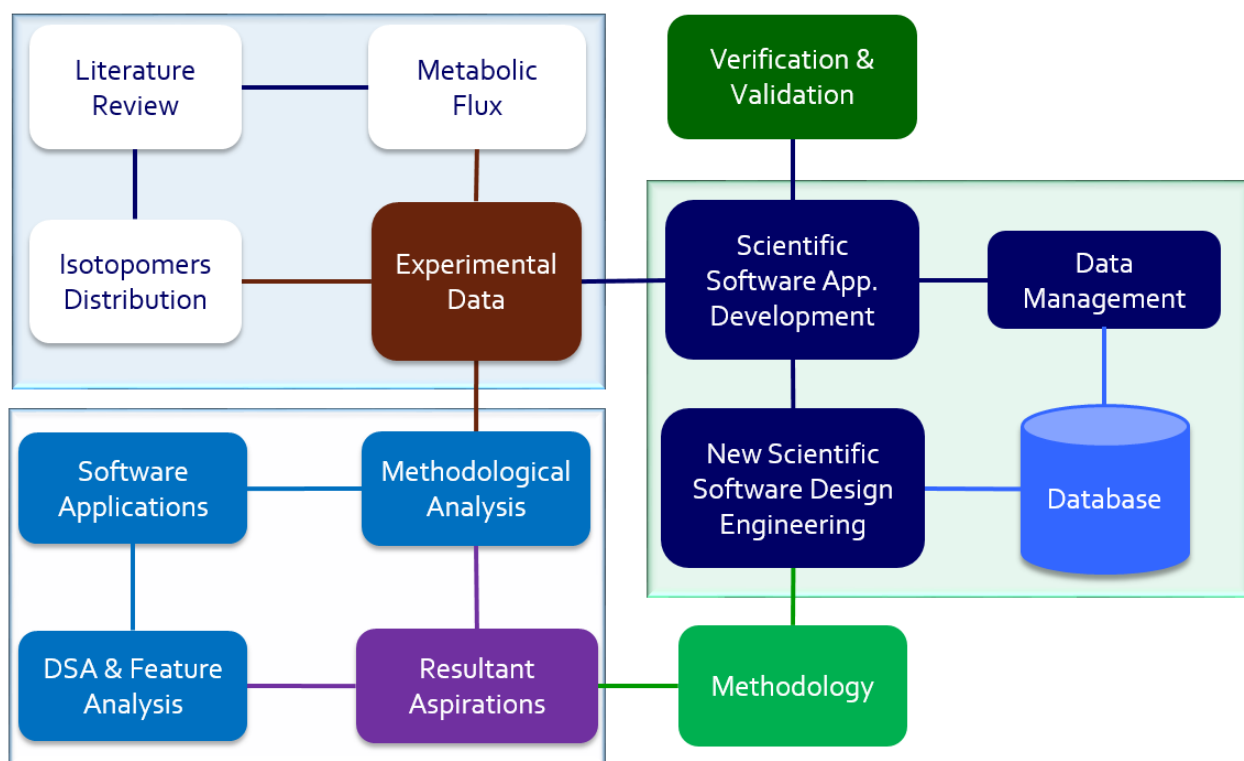
Figure 4: Study plan, metabolic modelling and spectral data analysis

1.2.4.1 Metabolic modelling

Metabolites and their pathway are central for pathophysiological adaptation of cells. Recent methodological advances allow direct measurements of metabolites. Metabolic modelling 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. Here, the focus of research is towards the Metabolic Flux Balance Analysis (FBA); rate of turnover of molecules regulated by the involved enzymes, through a metabolic pathway (Eisenreich et al. 2010; Eylert et al 2010). FBA is for analyzing the metabolism without any knowledge of enzyme kinetics. In other words it is the rate of movement of molecules across unit areas. Metabolic pathway is the series of chemical reactions occurring in a cell to modify principal chemical and maintain homeostasis within an organism. In case, if there is more than one metabolic pathway in a cell, it forms a network called Metabolic Network. We tried to find better analytical approaches towards Metabolic Pathway analysis, with the involvement of mathematics and informatics to analyse obtained results from performed experiments.

Metabolic Flux is basically divided into two types: Reactions and Metabolites. Reactions are the columns in stoichiometry matrix, further divided into two different types: Reversible based on forward and reverse steps and irreversible based on reverse rate lower than the forward. *Metabolic flux networks* are composed of nodes by edges inferring a degree of co expression consisting of methods for the calculation of correlation and defined threshold. Metabolic Flux Network analysis is very important to be considered in case we are interested in understanding the complexity of biological systems especially if systems are based on large datasets of biological components interaction with each other. To have in depth analysis of organism, in case the metabolic network is complex and consisting of multiple pathways, metabolic network can be reconstructed by

breaking down pathways with respect to the reactions and enzymes (correlation between genome and metabolism must be made).

Elementary Mode Analysis; elementary modes (Papin et al., 2004) are the flux distributions of the metabolic network in a steady state with the fluxes through irreversible reactions going in the appropriate direction (Schuster *et al*., 2002). Elementary flux modes are calculated to decompose the biochemical network into small parts to represent essential structural features of the biochemical network. Elementary modes analysis is to systematically enumerate stoichiometrically and thermodynamically feasible and independent (minimal) pathways through a network, consisting of stoichiometries and reversibility reactions (consisting of metabolites and reversible and irreversible enzymes) and classifies metabolites (internal and external), majorly used for metabolic network analysis. Elementary modes analysis is mainly used to test a set of enzymes for production of a desired product and detect non-redundant pathways, test a set of enzymes and pathways to determine maximal molar yields and pathway alternatives and reconstruction of metabolism from annotated genome sequences and analysis of effect of enzyme deficiency. Moreover it can also be used for the reduction of drug effects and drug target identification.

For detailed network analysis, high quality network data visualization is helpful. Mathematics, graph theory and computer science in system biology improve network visualization, remove noisy, spurious edges and randomly linked nodes with the use of several existing and different algorithms, techniques and statics. Targeting this need, many software applications have been designed and developed by several academic and commercial organizations for biological data based network and pathways e.g. MetaTool (Schuster and Schuster, 1993), C13 (Wiechert et al. 1997), 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) and COBRA Toolbox (Cvijovic et al. 2010).

1.2.4.2 Isotopologue labeling and metabolite analysis

Metabolite synthesis and balance is central for cellular survival. Irregularities in the transport or utilization of molecules can lead to disease e.g. metabolic disorders (such as diabetes and hypercholesterolemia). Moreover, changes in metabolism are a crucial step in the adoption of bacteria to environmental changes. For more than seventy years many attempts have been made to measure metabolic fluxes using isotopes, but still currently a comprehensive, general approach is still missing. A major reason for this in comparison to static measurements such as radio immuno assays is the difficulty to deconvolute the full isotopic composition in the metabolic network under physiological relevant conditions (Hellerstein and Nesse, 1992).

In the past (till 1992), generally, two isotopic methodologies were proposed and widely used for quantitative polymer analysis i.e. first by quantifiably incorporating labelled 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. Later on, after 1992, a new approach was proposed (Hellerstein and Nesse, 1992) using stable isotopes known as mass isotopomer distribution analysis (MIDA); a technique for biological polymer synthesis measurement (e.g. lipids, carbohydrates, and proteins) involving mass spectrometry for quantifying different relative abundances of molecular species of a polymer (Hellerstein et al., 1999). During MIDA the mass isotopomer distribution is analysed using a combinatorial probability model by comparing estimated abundance values to the predicted theoretical distribution (Hellerstein et al., 1999). With the application of probability analysis, MIDA is capable

of calculating actual biosynthesized precursor molecule subunits but the major constraint it encounters is the use of polymer molecules consisting of two or more identical reiterating precursor subunits. Three different MIDA challenges have to be addressed: (i) possible combinations of an identical subunit must be considered, (ii) the subpopulations of molecules must be distinguished and quantified and that (iii) the dilution of the monomeric and polymeric pools can affect the abundance distributions differently (Hellerstein and Nesse, 1992).

Newly synthesized polymers (labeled) are mixed into a compound with pre-existing polymer molecules to measure the fraction of molecules in the mixture. Theoretically speaking, according to the precursor product relationship (Hellerstein and Nesse, 1992), by determining the rate of the labeled polymer population in the mixture the amount of pre-existing unlabelled molecules can then be estimated. But in practice it is very difficult to measure the population of newly synthesized molecules, as there is no such physical technique for this purpose. Meeting this goal of MIDA, different calculation algorithms (Lee WN et al., 1991), (Lee WN et al., 1992) have already been proposed, published and compared. They share a common model, use overlapping solutions to computational problems and generate overall identical results. MIDA is a validated and internally consistent approach. Furthermore, to put more value to the approach's acceptability, MIDA has been quantitatively compared with some independent methods, applying several biological validations for biosynthesis polymer measurement (Hellerstein and Nesse, 1992).

Isotopologue measurements can rapidly translate into metabolic flux prediction applying the effective computation software platform with good application potential for microbiology and biotechnology. As shown in the Figure 5, the ^{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).
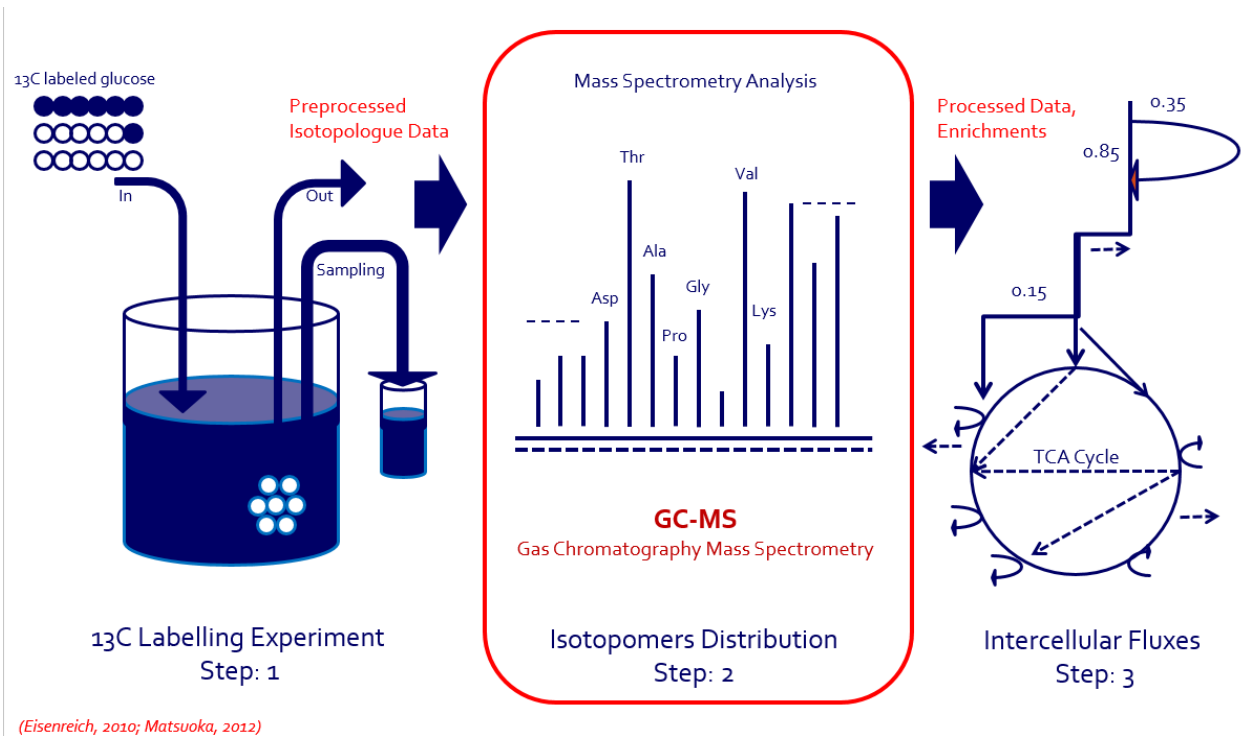
Figure 5: Spectral, preprocessed isotopologue data analysis of TCA cycle

The motivation for our research and new software applications 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. The major development goals are to implement a solution towards:

1. 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.

2. Abundance Measurements e.g. natural abundance, relative abundances, fractional molar abundances.

3. Enrichment Estimations e.g. global isotope enrichments (absolute etc.).

4. Visualization e.g. Spectrum bar, line, curve charts etc.

17

5. Spectral Data Management e.g. File based data handling, normalized relations, database management system.

This doctoral research presents two user-friendly platforms for Mass Isotopomer Distribution Analysis (MIDA), techniques that enables the determination of metabolic fluxes on the basis of labeling experiments using 13C-enriched precursors. Two new software applications named 'LS-MIDA' and 'Isotopo' are developed with facile data management and robustness to quantify the populations of isotopomers in mixtures of 13C-labelled amino acids. Both applications are 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.

LS-MIDA is using Brauman's least square method (Brauman, 1966) of linear regression based on the experimental data elements.

Isotopo is the advanced version of LS-MIDA, implementing a partial least square method with wider scope and better results. Both estimate mass values (M_0, M_1, $M_{maximum}$) and predict 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. Further, Isotopo 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.

1.2.5 Software development in scientific academia

Software development in scientific academia was central to all bioinformatics efforts done in this thesis and we developed a new approach also for this.

Scientists (e.g. biologist, neurologist, chemist, physicist etc.) are usually involved in performing different kinds of experiments in the laboratories. The results of the experiments drive the generation of different kinds of biological raw data. Nowadays with the help of advanceed computational machinery and network data sharing resources, it is possible to perform quick and efficient analysis of the biological raw data with the availability of different software and hardware oriented solutions.

The scientists are mainly interested in obtaining the end results but not the computational solution itself (e.g. whether it's desktop or web or ubiquitous etc. incorporating mathematics or statistics etc.). Their common interests lie among the processing, analyzing, visualizing, managing, sharing, experimenting and in some cases even generating new raw data (e.g. in case of behavioral or neural research on insects etc.). The lack of interest in computational solution development (e.g. during planning, requirements finalization, implementation and pre-experimental testing phases etc.) at scientist's end, increases the different levels of complexities at developer's end (especially if the scientific solution developer is without much biological knowledge).

In the beginning of scientific software solution implementation, most of the times things look ideally interesting and fascinating. With the passage of time the problems start appearing. This ambiguity causes an increase in the level of confusion at the developer's end. Moreover, it could be even worse, if the desired end product is a solution which contains the third party elements and requires the information to be included from unfamiliar or unspecialized fields to the scientists or developer or both (e.g. if the scientist is from the field of behavioral research, developers are from

informatics and required system is needed to be implemented using different kinds of knowledge from the fields of mathematics, statistics, engineering etc.). In such situations it's really hard for the developers to implement such system which can perform scientist's desired operations and produce required results.
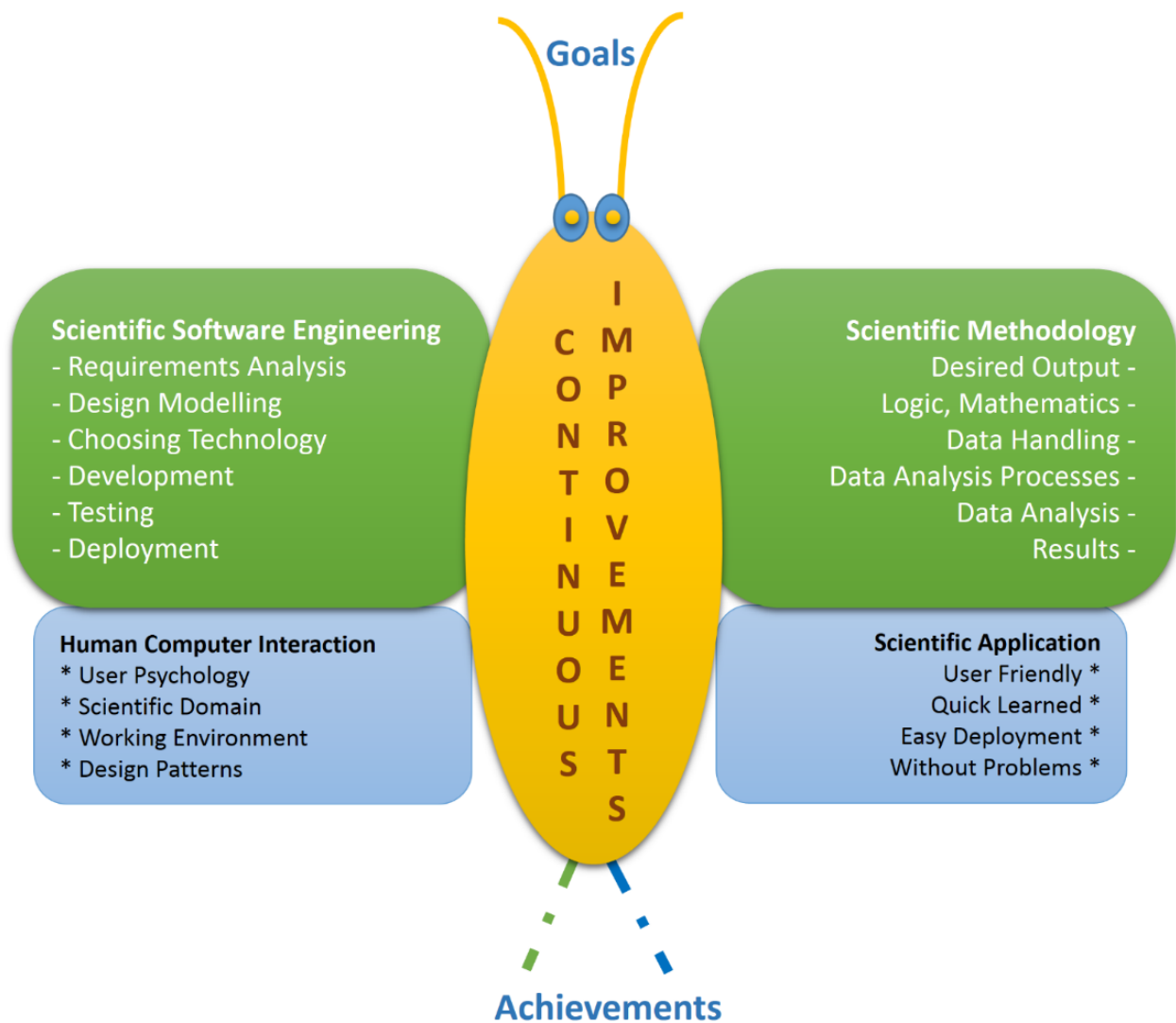
Figure 6: *Butterfly* **paradigm**

Vice versa in some cases problems could be of the reverse nature, where the implemented scientific solutions are producing anticipated results but the scientists are unable to use the software. In such

situations, there is a high possibility that the developer is not a professional or graduate of informatics but a bioinformatician is etc.

There could be many other kinds of the complexities and problems as well, which can only be solved or reduced using the concepts of the main building block of the computer science, its subfield i.e. Software Engineering (SE) (Boehm, 1976). To cope with such developmental issues, we have proposed one new paradigm '*Butterfly*' (shown in Figure 6). It is a science-oriented, software development model which can be helpful by generalizing the use of major developmental aspects correlating to the important scientific needs of the target system.

2   Material & Methods

2.1   Platelet web knowledgebase

The PlateletWeb knowledgebase (http://plateletweb.bioapps.biozentrum.uni-wuerzburg.de/PlateletWeb.php) provides direct user access to well curated known protein-protein interactions in the platelet (Boyanova et. al. 2014; Boyanova et. al. 2012). This resource was extensively exploited for details of all the apoptotic and anti-apoptotic proteins.

The first round of prediction covered extensive study of The PlateletWeb for all the proteins known to be involved in platelet apoptosis and anti-apoptosis. In the second step these proteins were then specifically analysed for their protein interactions. A list was compiled of all the unique protein interactions and their network generated using The PlateletWeb knowledgebase.

2.1.1   miRNA profiling

The miRNA study was carried out in context of the research performed by Landry et al, 2009. They performed miRNA profiling of platelets yielding 94 miRNAs linked to platelet functioning. Out of these miRNAs, most highly expressed miRNAs were separated out and verified by the miRBase: the microRNA database (http://www.mirbase.org/). They were analysed for their target proteins (based on probability >0.99).

The target proteins involved in platelet apoptosis and anti-apoptosis pathways were predicted from TargetScan (http://www.targetscan.org/) (Friedman et al., 2009) and their proteins interactions were established in the platelet network.

2.2 Methods miRNA

2.2.1 MicroRNA target determination

A fundamental aspect of miRNA function relates to mRNA targeting: One of the most interesting aspects of miRNA biology is that miRNAs are predicted to target multiple mRNAs, and most mRNAs have predicted targets for many miRNAs. Thus making miRNAs one of the most potent biological regulators. However, defining the gene targets through which a miRNA functions is probably also the most tedious aspect of miRNA research. Initial insight into miRNA targets can be obtained bioinformatically through a number of freely available softwares and applications that predict potential mRNA targets for individual miRNAs.

mRNA targets of platelet microRNAs were predicted by a variety of Web-based bioinformatic tools, such as miRBase Targets (http://www.mirbase.org) (Griffiths-Jones et al., 2008), DIANA-microT-CDS (http://diana.cslab.ece.ntua.gr/) (Reczko et al., 2011), PicTar (http://pictar.mdc-berlin.de/) (Krek et al., 2005), Miranda (www.microrna.org) (John et al., 2004), TargetScan (http://www.targetscan. org/) (Friedman et al., 2009) and miRecords (http://mirecords.biolead.org) (Xiao et al., 2009).

These target prediction programs use several characteristics to determine whether a miRNA can potentially target an mRNA. One common approach for target prediction has been 'seed driven', in which the reverse complement of nucleotides 2–8 in the miRNA is searched for in the mRNA target sequences, particularly in the 3′- UTR (Grimson et al., 2007). TargetScan is an example of an algorithm using this approach.

2.2.2 TargetScan

TargetScan predicts biological targets of miRNAs by searching for the presence of conserved 8mer and 7mer sites that match the seed region of each miRNA (Lewis et al., 2005). As an option, nonconserved sites are also predicted. Also identified are sites with mismatches in the seed region that are compensated by conserved 3' pairing (Friedman et al., 2009). In mammals, predictions are ranked based on the predicted efficacy of targeting as calculated using the context scores of the sites (Grimson et al., 2007). TargetScanHuman considers matches to annotated human UTRs and their orthologs, as defined by UCSC whole-genome alignments. Conserved targeting has also been detected within open reading frames (ORFs).

The 5' seed region of the miRNA (bases 2 to 8) must show sequence complementarity to the 3' UTR of a target gene, and the target site within the mRNA should be conserved among different species. Often, the thermal stability of the mRNA/miRNA duplex, together with the absence of complicated secondary structures surrounding the miRNA binding site, is taken into account to predict whether a miRNA is likely to target an mRNA.

TargetScan also adds a "context score," which considers features in the surrounding mRNA, including local A-U content and location (near either end of the 3' UTR is preferred) and improves predictions for nonconserved sequences. mRNAs that have a high context score or multiple predicted miRNA-binding sites are more likely to be true targets. Additionally, TargetScan includes a special class of seed matches with a hexamer match in positions 2 to 7, plus an adenosine at position 1.

2.2.3 PicTar

However, in addition to these common characteristics in target prediction, there are some distinct differences between the several approaches. Although all of the prediction algorithms use the seed

24

sequence as the main determinant of target site recognition, PicTar additionally allows for both perfect and imperfect seed complementarity. The perfect seed is defined as perfect Watson–Crick base pair complementarity of seven nucleotides, starting at either the first or second base of the 5' end of the miRNA. Imperfect complementarity allows for an insertion or mutation as long as the free energy of binding of the miRNA/ mRNA duplex does not increase or does not contain a G U base-pairing. Both TargetScan and PicTar improve their predictions by taking into account evolutionary conservation.

2.2.4 miRanda

Although seed pairing is weighed more strongly than pairing elsewhere, 'Non-seed' target prediction methods consider the free energy of binding between the miRNA and target, as well as conserved sequence features of mature miRNAs. These latter methods, such as miRanda and RNA22, do not require seed base pair matching. Algorithms use additional information, including conservation of target sites across species and the RNA profiles of miRNAs and mRNAs that are co-expressed in the same cell (Rigoutsos et al., 2011).

The miRanda algorithm aligns a miRNA to the target mRNA to identify highly complementary sequences, whereby allowing for seed G U wobbles and mismatches. High-scoring targets are then filtered on a secondary criterion of free energy (G), whereas only conserved predictions are considered. Because miRanda does not require exact seed pairing, it predicts sites such as the two let-7 sites in the Caenorhabditis elegans gene lin-41, which contain either a bulge or a G U wobble in the seed region (Thomas et al., 2010) (John et al., 2004). Computational approaches for the identification of these tools combine different parameters of the sequence requirements for miRNA-mRNA binding and calculate the free energy of the interaction as a predictive method to identify mRNA targets with a relatively high degree of confidence (Ple´ et al., 2012).

2.2.5 Cytoscape

Cytoscape (http://www.cytoscape.org/) is an open source bioinformatics software platform for visualizing molecular interaction networks and biological pathways and integrating these networks with annotations, gene expression profiles and other state data. Although Cytoscape was originally designed for biological research, now it is a general platform for complex network analysis and visualization. The Cytoscape software provides many features for data integration, visualization and analysis.

2.3 Metabolic modelling software

Metabolic modelling includes metabolic flux analysis as a key method and provides important insights into metabolism and adaptations of different organisms (Eisenreich et al., 2010), (Eylert et al., 2010). In the following description we compare, analyse and combine available softwares to obtain suggestions for further development and improvement in this area.

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 it is produced. The stable and balanced metabolite flux achieved by such a combination of enzymes is called as 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 can 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 reduced set of such elementary modes is sufficient to reproduce all other modes by linear combination of the modes, a so called convex basis (Schuster et al., 2000).

Modifications and extensions regarding metabolic flux analysis have continuously been proposed e.g. (Kaleta et al., 2009; Rahman and Schomburg, 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) and Classical and Dynamic FBA (Zamboni, 2007). Another problem is to map fluxes to actual observed experimental data. For this solutions such as Rematch (Pitkänen et al., 2008). 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. Many groups in genome-based bioinformatics development rapidly establishes enzyme networks for genome annotation and comparisons (including KEGGbrowser (Goto et al., 2000)). Hence, the problem is important but too complex for a fair comparison between all available tools. Furthermore, groups involved in major genome annotation efforts established quite advanced user interfaces and annotation systems such as MAGPIE (Gaasterland and Sensen, 1996), GenDB (Meyer et al., 2003), GENEDB (Logan-Klumpler et al., 2011), PEDANT (Frishman et al., 2003), large-scale interactome searching tools (e.g. STRING (Szklarczyk et al., 2011)) as well as important large-scale repositories (e.g. EBI, NCBI). There are various further tools for

annotation and sequence analysis such as SMART (Letunic et al., 2002), COG (Tatusov et al., 2000), Daileon (Perillo et al., 2009) and E2D (Lee et al., 2006).

Finally, no matter which of these preparatory steps is performed, the graphical representation of the found pathways is again important and some of the discussed software provide such visualization. Major current graphics related challenges are: new techniques for information visualization (including 2D and 3D vector data), immersive virtual presentation of complex pathways, 3D reconstruction of large amount of data (including text, images and video), visualization of univariate 3D and multivariate scalar data, abstraction and effective use of colors etc.

During recent years, a lot of well contributing computational research and development has been done in the field of bioinformatics e.g. methods for chromosomal translocations (Hoffmann-LA and Roche, 2012), heavy and complex data analysis using distributed processing (Hubbard, 2012), secretory protein of bacterial pathogens (Brett et al., 2012), genotypic determination using probe array data (Earl and Simon, 2012), Image analysis for measuring global nuclear patterns (Sammak et al., 2012), detecting splice variants (Williams et al., 2012), data relationship determination (Graham et al., 2012), compositions for diagnosing conditions (Fan and Marina, 2005), gene expression signature for assessing pathway activity (Loboda et al., 2010), sequential pattern data mining and visualization (Wong et al., 2012).

The focus of our research is towards the analysis of various software applications developed for pathway analysis e.g. BioMet Toolbox (Cvijovic et al., 2010), CellNetAnalyzer (formerly FluxAnalyzer) (Steffen and Axel, 2002), COPASI (Hoops et al., 2006), Jarnac (Sauro, 2000), Pathway Analyser (Oehm et al., 2008), MMT (Hurlebaus et al., 2002), SBTOOLBOX2 (Henning

and Mats, 2005), TinkerCell (Deepak, 2009), WebCell (Lee et al., 2006) and SCAMP (Herbert, 1993).

All of these softwares are used for certain aspects of pathway analysis including regulation, involved interaction networks as well as kinetics. For this thesis we used a powerful combination for metabolic modelling starting from raw data (e.g. isotopologue measurements) and ending up with a description of flux distributions. However, this is a dynamic field and its usage and preferences (including our own) for software depend also on the familiarity of the user with a particular software, operating system preferences, the scientific question and purpose in mind and of course the type of data to be analyzed. Hence, we had for each step of analysis two alternatives. Our selection covers C13 (Wiechert and de Graaf, 1997), Classical and Dynamic FBA (Mahadevan et al., 2002), MetaTool (Schuster and Schuster, 1993), BioOPT (Cvijovic et al., 2010), FiatFlux (Zamboni et al., 2005), (Sammak et al., 2012), Rematch (Pitkänen et al., 2008) and BioLayout Express3D (Theocharidis et al., 2009).

First, there is the construction of the metabolic network using e.g. KEGGconverter (Konstantinos, 2009), Metannogen (Gille et al., 2007), KEGGbrowser (Goto et al., 2000) followed by 13C-constrained flux analysis software such as C13 (Wiechert and de Graaf, 1997), Openflux (Quek et al., 2009) and ReMatch (Pitkänen et al., 2008). Next flux balance analysis by Metatool (Schuster and Schuster, 1993), BioOpt (Cvijovic et al., 2010) or Classical and dynamical FBA (Mahadevan et al., 2002) enumerates all metabolic pathways available to the system investigated. Finally, the fit to data using FiatFlux (Zamboni, 2007), points out which of the calculated pathways is actually active under a given condition. The tests and data provided by the different applications often support each other for subsequent analysis steps. This makes it difficult to give a fair comparison choosing a specific test data set and applying it to all (Ahmed et al. 2014a).

Furthermore, we combined software YANA (Schwarz et al., 2005), YANAsquare (Schwarz et al., 2007) and YANAvergence (Liang et al., 2011) for metabolic pathway analysis (Schauer et al., 2010). The YANA programs present also a pipeline to construct metabolic models rapidly and apply experimental data to calibrate the network, estimate the metabolic fluxes. Various optimization methods fit experimental data to calculated fluxes, including the genetic algorithm, non-linear optimization and L-BFG-S routines which have been integrated into YANAvergence. This software was helpful in many ways for metabolic flux analysis and visualization. Thus C13 can be used for metabolic flux analysis by stationary carbon isotope labelling 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 13C-metabolic flux analysis over the web.

The discussed software in Table 1 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.

The conclusion are:

1. The efficient Metabolic reconstruction can be achieved using e.g. KEGG-DB & Browser, Yana processes SBML

2. EFM tool, iMAT, ReMatch can be applied for better analysis of network structure and modelling.

30

3. FBA, OpenFlux, BioOpt are good tools for the Flux balance and constraint based modelling

4. Cobra toolbox, Yana, METATOOL can be used for detailed modelling of metabolism and its kinetics.

5. FiatFlux, C13 can be adopted for flux predictions by processing Isotopologue.

Table 1: Software Comparison

Software	Description
C13	Estimate fluxes satisfying stoichiometric 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 to1000 expression arrays and using MCL produces cluster of graphs (30000 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 the best set of gene deletions for a 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 modelling 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 from13C 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 Readymade metabolic networks for more than 160 organisms
Metatool	Provides fast and simple elementary flux mode calculations, parses reaction equations and translates them into a stoichiometric 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 stoichiometric 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.

3   Results: miRNA

3.1   Role of platelet miRNAs in apoptosis

The phenomenon of life and death has fascinated scientists since old times. Cell proliferation, differentiation and death are fundamental processes in multicellular organisms. Programmed cell death occurs in maturation and aging by different cellular programmes (Zhang et al., 2007).

In platelets apoptosis such as programmed cell death refers to cell death events mediated by activation of proteolytic pathways. Several studies have demonstrated that platelet senescence is associated with changes characteristic of apoptosis. The Bcl-2 family of proteins (i.e., Bax, Bak, Bcl-XL, and Bcl-2) expressed in platelets, is the key regulator of mitochondria-dependent apoptosis in nucleated cells and consists of both antiapoptotic (Bcl-XL, Bcl-2, Bcl-w, A1, Mcl-1) and proapoptotic (Bak, Bax, Bid, Bim, Bad, Bik, Bmf, Noxa, Puma) members. The cascades involving these pro-survival and antiapoptotic proteins were closely analysed and all key molecules and pathways were included in the analysis. These included caspases 3 and 9, also expressed in platelets but their exact role and effects are still controversial as well as their extant of participation in apoptosis or apoptosis-like pathways.

Our study also verified the correlation of miRNA expression with platelet activation or inhibition. Here we have tried to create that connection in platelets (Figure-10b) between 3 of our predicted proteins as putative miRNA targets and the one identified by Landry et al., 2009. The results show that platelet activation leads to platelet migration and shape change with inhibition of apoptosis and miRNA expression for gene regulation. Validating the protein cascades involved in platelet activation and studying the role of miRNA targeted proteins in inhibiting apoptosis like events in platelets. miRNA study revealed 6 miRNAs, hsa-let-7d, hsa-let-7f, hsa-let-7g, hsa-miR-98, hsa-

miR-130a, and hsa-miR-130b, with three platelet protein targets, IGF-1, IGF-1R, and TGFβ-R1 (Table 2).

The mechanism for platelet control of function via miRNA expression is still not well known. However increase in miRNA expression on platelet activation with thrombin or immune stimulation (Tanriverdi et al., 2008) and miRNA control of platelet gene expression (Landry et al., 2009) has been recently studied. This study also verifies the correlation of miRNA expression with platelet activation or inhibition. Validating the protein cascades involved in platelet activation and studying the role of miRNA targeted proteins in inhibiting apoptosis like events in platelets.

3.2   Factors involved in platelet apoptosis

Apoptosis is a complex process and the decision to undergo apoptosis involves many pathways and an intricate signaling machinery. On the other hand, survival or proliferation of a cell comprises of many interacting networks. Both cell proliferation and apoptosis are very closely linked by receptors and signaling cascades.

This study aims to explore and understand these processes in platelets. Platelets are already "born" by an apoptosis-like process from mature megakaryocytes (Patel et al., 2005a). For activation of platelets (e.g. in blood clotting), the cascade is switched on a second time but how and why not earlier? For this protein interactions, mRNA translation (Weyrich et. al. 2004) and again mRNA regulation by miRNAs are all interconnected.

We considered the expression of Bcl-2 family proteins (i.e., Bax, Bak, Bcl- XL, and Bcl-2) in platelets indicating that platelets contain the necessary apoptotic machinery to execute programmed cell death (Zhang et al., 2007). However the exact role of these pro-survival (Bcl-2, Bcl-xL, the latter suggested as platelet timer e.g. by Mason et al., 2007) and proapoptotic (Bax, Bak, the latter

35

a key counter player for Bcl-xL in platelets) is still not clear. In addition, platelets possess other key components of the pathway including caspase-9. However, it has been difficult to determine whether these factors have detrimental or beneficial effects on platelet survival or biogenesis (Qi and Hardwick, 2007). Recent experimental studies suggest important modulatory functions for about 200 microRNAs in platelets (Landry et al., 2009). These recent developments are incremental in that they only partially explore the apoptotic process in platelets. However to find the key players in bringing about this process has strong medical implications e.g. to regulate the amount of circulating platelets.

Integrins play a very important role in thrombosis, the processes of platelet adhesion and aggregation (Ni and Freedman, 2003). Resting platelets express integrins α2β1, α5β1, α6β1, αvβ3, and αIIbβ3. In addition to these, αLβ2 and αMβ2 expression on activated platelets has also been reported (Raab et al., 2010). GPIIb-IIIa is the most abundantly present and best studied member of the adhesion molecule family of integrins. It was later named as αIIbβ3. It is expressed exclusively on megakaryocytes and platelets, accounting for about 17% of the total platelet membrane protein (Ni and Freedman, 2003). Abnormalities of the platelet specific integrin, αIIbβ3, prevent platelet aggregation following vascular injury (Takagi et al., 2002).

However, integrins do not function alone. The signals from these adhesion receptors are integrated with those originating from growth factor receptors and G-protein receptors in order to organize the cytoskeleton, stimulate cell proliferation and rescue cells from matrix detachment and induced programmed cell death (Schwartz and Ginsberg, 2002).

Integrin activation induces a wide range of intracellular signaling events, including the activation of Ras; MAP kinase; focal adhesion kinase (FAK); Src, Rac, Rho, and Cdc42 GTPases;

phosphatidylinositol-3-kinase (PI3-kinase), Abl, and integrin-linked kinase (Schwartz and Shattil, 2000).

Platelets are rich in growth factors (PDGF, IGF-1, TGF-β, FGF, VEGF, EGF, HGF) that can stimulate myogenesis and decrease inflammation. They also contain TGF-β secreted on platelet activation and aggregation. It stimulates endothelial cell proliferation (Trikha et al., 2002).

ADP-induced platelet aggregation requires coactivation of the Gq-coupled P2Y1 and Gi-coupled P2Y12 receptors. Studies have convincingly shown that stimulation of P2Y1 leads to platelet shape change, while stimulating P2Y12 promotes platelet aggregation (Eckly et al., 2001), and both P2Y1 and P2Y12 play an important role in shear stress-induced platelet aggregation (Mazzucato et al., 2004) (Turner et al., 2001).

The P2Y12 receptor plays a central role in platelet activation, in the recruitment of other platelets to the site of injury subsequent to the adhesion of platelets to vWF and collagen. Despite the established role of the P2Y12 receptor in the hemostatic response, the full implications of P2Y12 receptor antagonism in the prevention of thrombosis remain incompletely understood (Dorsam and Kunapuli, 2004).

P2Y12 receptor mRNA has been identified to be a putative target of miR-223 miRNA. This claim was verified and the connection of P2Y12 with integrin-growth factor crosslink was established. All these factors were analysed, detailed data are given later.

3.3   Results for Platelet Apoptosis Study

The previously done miRNA study on platelets (Landry et al., 2009) was based only on experimentation data and no bioinformatics analysis was performed. Platelet activation regulates miRNA expression in platelets. The study by Landry et al., 2009 was not performed systematically

to disclose other possible miRNA targets. Our study comprises of an organized research revealing putative targets. miRNA analysis of 30 highly expressing platelet miRNAs (Appendix-1) from the set of 94 platelet miRNAs identified by Landry et al., 2009 was performed. This research revealed 3 platelet protein targets and many non-platelet protein targets (Table 2).

miRNA-targets screening steps are as follows:

- Only top miRNAs were considered for our study. miRNA profiling of platelets yielded 94 miRNAs (Landry et al, 2009) linked to platelet functioning. Out of these miRNAs, the most highly expressed top 30 miRNAs were picked out for our study (Appendix-1).

- The first round of platelet protein prediction revealed 105 anti-apoptotic proteins and 501 apoptotic proteins from The PlateletWeb knowledgebase. At every step each and every one of these proteins was analysed for their known non platelet interacting proteins. The non-platelet proteins were also listed with their platelet interaction functions as many of them were found to be involved in cellular processes of apoptosis, survival and their regulation. The protein functions were retrieved from Entrez Gene (http://www.ncbi.nlm.nih.gov/gene). This lead to a comprehensive set of platelet proteins and platelet interacting proteins involved in a complex and an extensive network for platelet apoptosis and antiapoptosis.

- These set of platelet proteins were mapped against the protein targets of the highly expressed set of miRNAs identified from Landry's list of platelet miRNAs. The result was three platelet proteins mapped with six platelet miRNAs. A networking pathway was constructed around these three platelet protein targets using their set of platelet and non-platelet interacting proteins. This was only possible combining both manual search and screening with software prediction and visualization. The three platelet protein targets are:

- Insulin-Like Growth Factor 1 Receptor (IGF-1R) with 39 interacting proteins in the network (Figure 7)

- Transforming Growth Factor, Beta Receptor 1 (TGFB-R1) with 78 interacting proteins in the network (Figure 8)

- Insulin-like growth factor 1 (IGF-1) with 6 interacting proteins in the network (Figure 9)

Table 2: The platelet protein targets for the platelet miRNAs

hsa-let-7d	IGF-1R and TGFβ-R1
hsa-let-7f	
hsa-let-7g	
hsa-miR-98	
hsa-miR-130a	IGF-1
hsa-miR-130b	

IGF-1R: Insulin-Like Growth Factor 1 Receptor

TGFB-R1: Transforming Growth Factor, Beta Receptor 1,

IGF-1: Insulin-like growth factor 1

hsa: Homo sapiens.
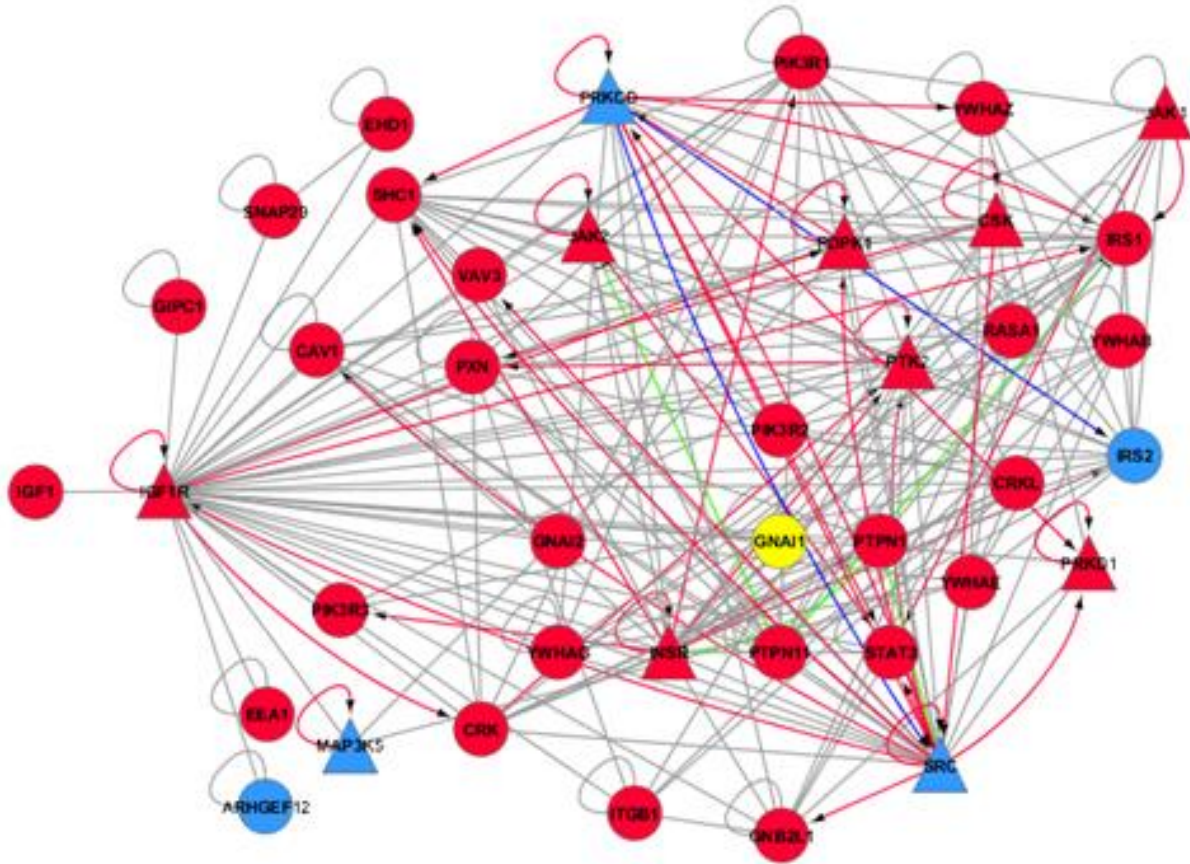
Figure 7: Protein-protein interaction network for IGF1R.

Red color: platelet proteins with phosphorylation sites detected in human cells

Yellow color: platelet proteins with no detected phosphorylation sites

Blue color: platelet proteins with phosphorylation sites detected in platelets

Circle: Proteins

Triangle/Square (new version): Kinases

Grey arrow: Protein-protein interaction

Red arrow: Phosphorylation reaction

Blue arrow: Kinase phosphorylation reaction

Green arrow: Dephosphorylation reaction

Loops: Autophosphorylation

Figure 8: Protein-protein interaction network for TGFBR1

Red color: platelet proteins with phosphorylation sites detected in human cells

Yellow color: platelet proteins with no detected phosphorylation sites

Blue color: platelet proteins with phosphorylation sites detected in platelets

Circle: Proteins

Triangle/Square (new version): Kinases

Grey arrow: Protein-protein interaction

Red arrow: Phosphorylation reaction

Blue arrow: Kinase phosphorylation reaction

Green arrow: Dephosphorylation reaction

Loops: Autophosphorylation
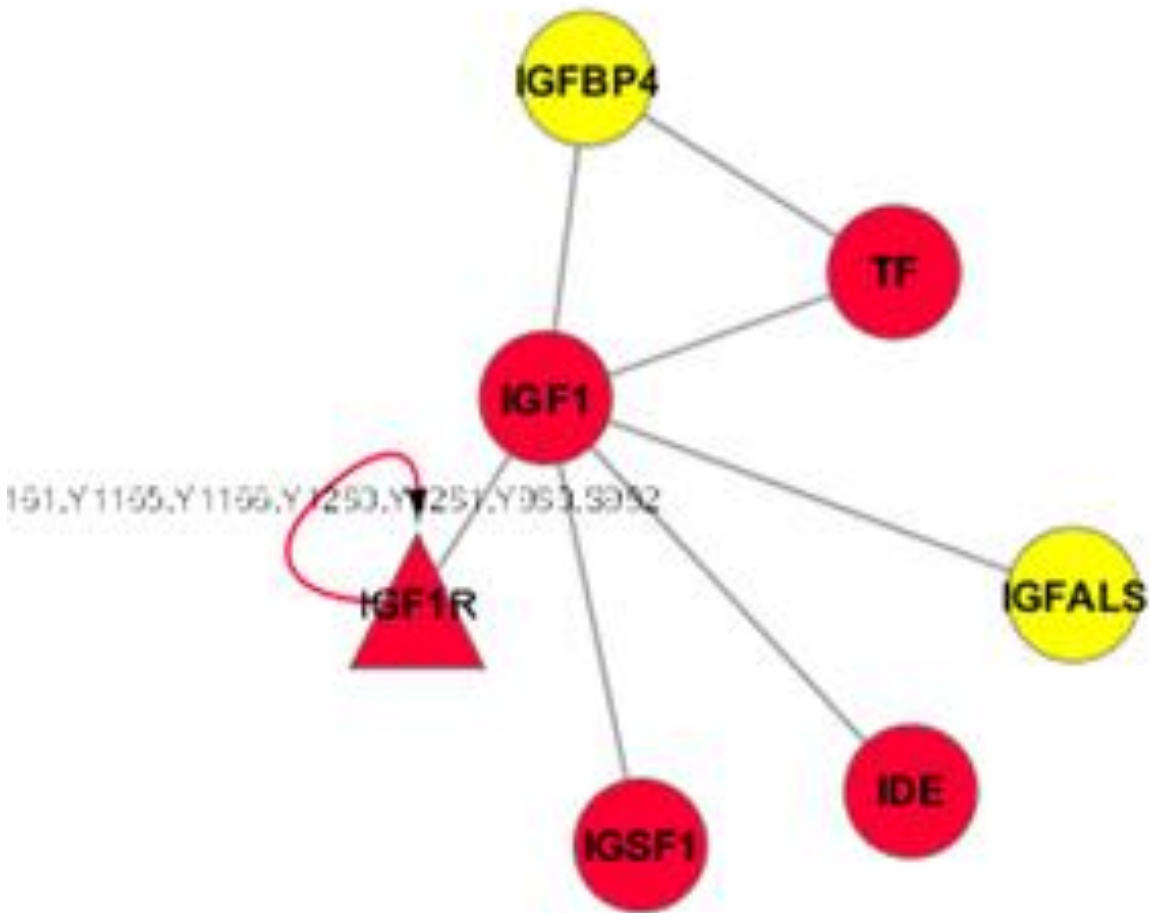
Figure 9: Protein-protein interaction network for IGF1 with phosphorylation sites on IGF1R

Red color: platelet proteins with phosphorylation sites detected in human cells

Yellow color: platelet proteins with no detected phosphorylation sites

Blue color: platelet proteins with phosphorylation sites detected in platelets

Circle: Proteins

Triangle/Square (new version): Kinases

Grey arrow: Protein-protein interaction

Red arrow: Phosphorylation reaction

Blue arrow: Kinase phosphorylation reaction

Green arrow: Dephosphorylation reaction

Loops: Autophosphorylation

Figure 10a: Crosslinking pathway between platelet IGF-1R, IGF-1 and TGFB-R1 (The PlateletWeb version)

Red color: platelet proteins with phosphorylation sites detected in human cells

Yellow color: platelet proteins with no detected phosphorylation sites

Blue color: platelet proteins with phosphorylation sites detected in platelets

Circle: Proteins

Triangle/Square (new version): Kinases

Grey arrow: Protein-protein interaction

Red arrow: Phosphorylation reaction

Blue arrow: Kinase phosphorylation reaction

43

Green arrow: Dephosphorylation reaction

Loops: Autophosphorylation

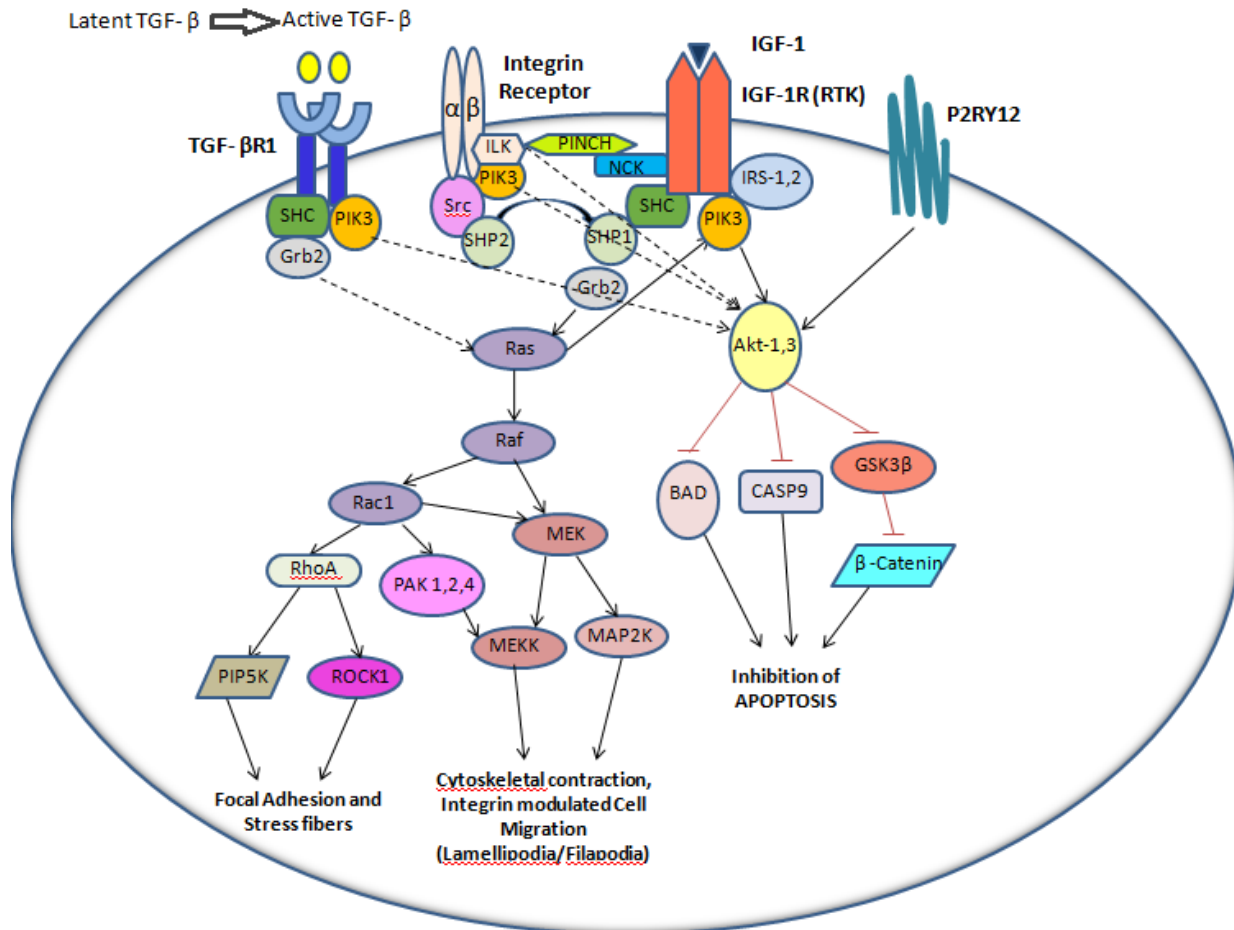**Figure 10b: Predicted protein connections**

Figure 10: Crosslinking pathway (10a) and Predicted protein connections (10b)

IGF-1, IGF-1R and TGFB-R1 have been known to have an important role in platelet apoptosis and anti-apoptosis pathways. These miRNA target proteins have already been identified from the PlateletWeb knowledgebase and are present in the above generated networks. Protein interactions

for these targets were extracted (Appendix-2) from the above mentioned extensive networks using visualization tool Cytoscape.

The crosslinking between the three miRNA targets IGF-1R, IGF-1 and TGFB-R1 leads to a network which was first generated by The PlateletWeb, Figure 10a, but was later on redesigned explicitly for understanding in Figure 10b. Proteins involved in the cross linking pathway have been tabulated in Appendix-3. Although not all proteins from Appendix-3 have been included in the network to keep the network simple, their relation to the network has been enlisted along too.

3.4   Results for murine platelet study

A part of our miRNA study was examining microRNA expression in murine megakaryocytes. miRNAs are also expressed in megakaryocytes and platelets (Opalinska et al., 2010). Many of the changes observed in murine megakaryocytes are consistent with miRNA profiling studies of human megakaryocytes and platelets. Previous studies have defined how miRNAs are likely to regulate mature platelet functions both positively and negatively. We investigated in this area to provide new insights into the basic biology of cell death in platelets and the pathologic conditions associated with it.

The study was extended to the the platelet protein targets for the platelet miRNAs, looking for protein targets for these miRNAs in mice. Common protein targets were found for the miRNAs both in human and mice (Appendix-4). Let-7d, let-7g, let-7f, miR-130a and miR-13b were found to share targets between human and murine miRNA (Figure 11-15) however miR.-98 did not have any common target.

The screened miRNAs (Table 2) analysed for target protein both in humans and mice (based on probability >0.99). TargetScan software (http://www.targetscan.org/) (Friedman et al., 2009) was

used to find these predicted hits. Theses miRNA hits were matched against each other to come up with common proteins targets. The following results were generated using Cytoscape.

Figure 11: Protein targets for let-7d

Green squares: protein targets present in both humans and mice

Light blue squares: murine protein targets

Aqua blue squares: human protein targets

Figure 12: Protein targets for let-7f

Green squares: protein targets present in both humans and mice

Royal blue squares: murine protein targets

Aqua blue squares: human protein targets

Figure 13: Protein targets for let-7g

Green squares: protein targets present in both humans and mice

Aqua blue squares: murine protein targets

Purple squares: human protein targets

48

Figure 14: Protein targets for miR-130a

Green squares: protein targets present in both humans and mice

Pink squares: murine protein targets

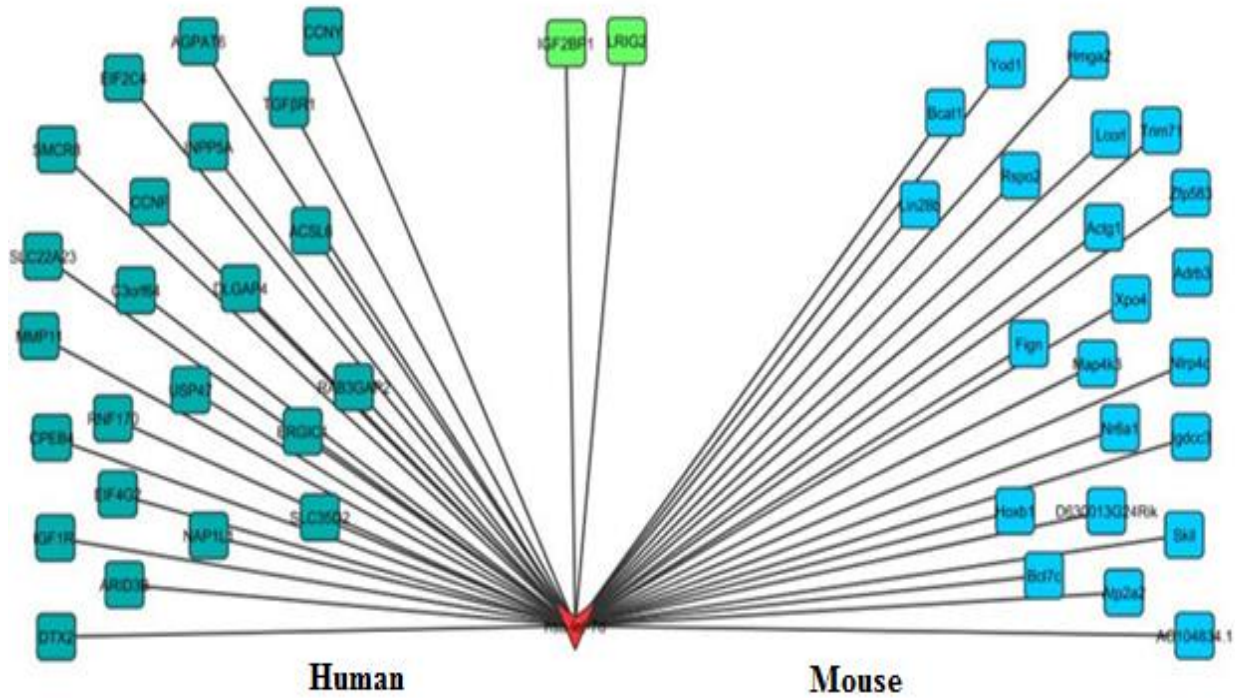Aqua blue squares: human protein targets
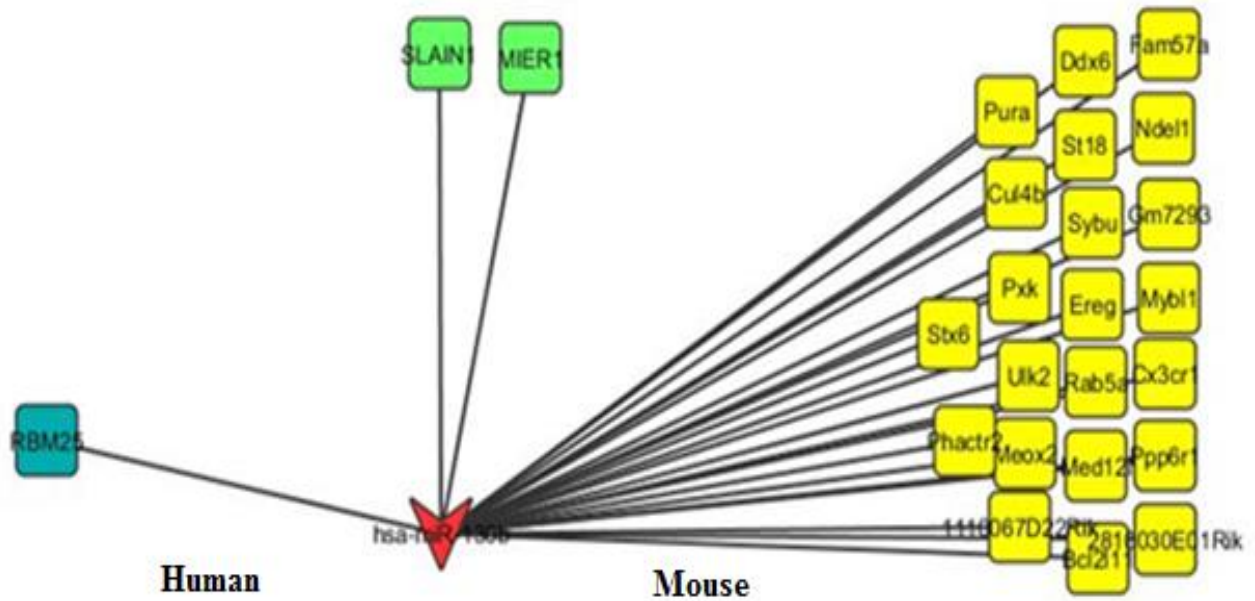
Figure 15: Protein targets for miR-130b

Green squares: protein targets present in both humans and mice

Yellow squares: murine protein targets
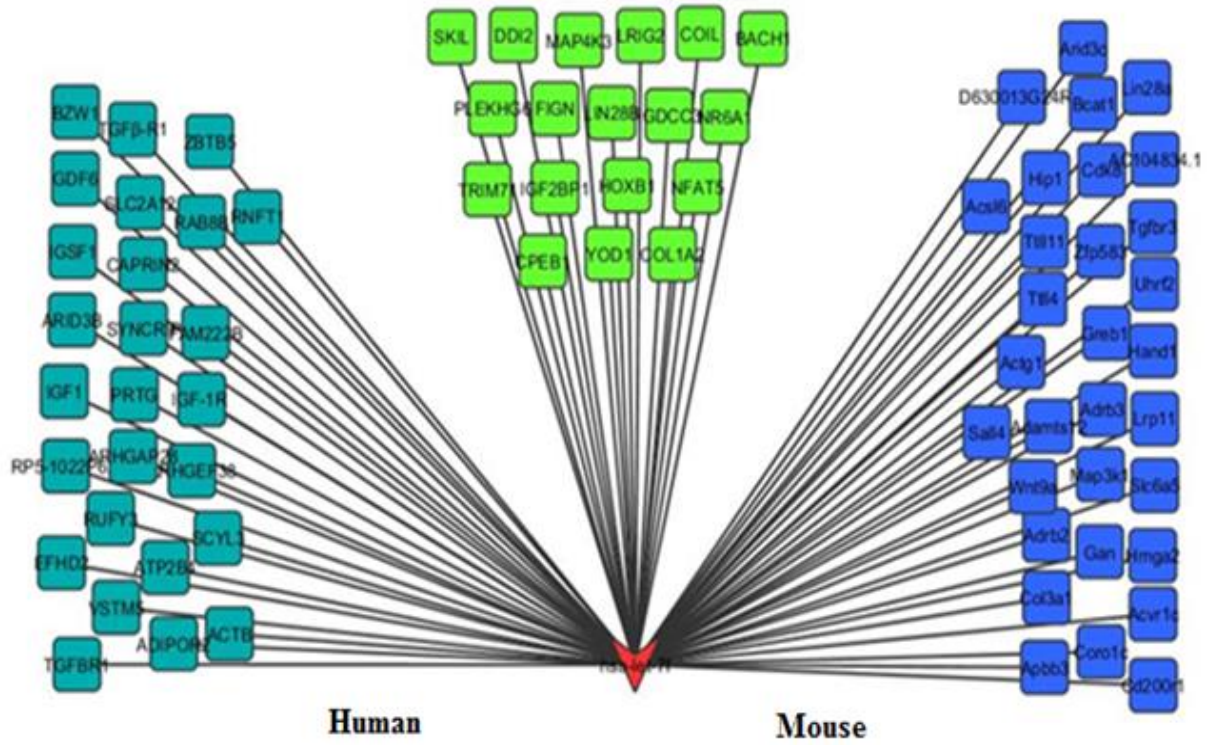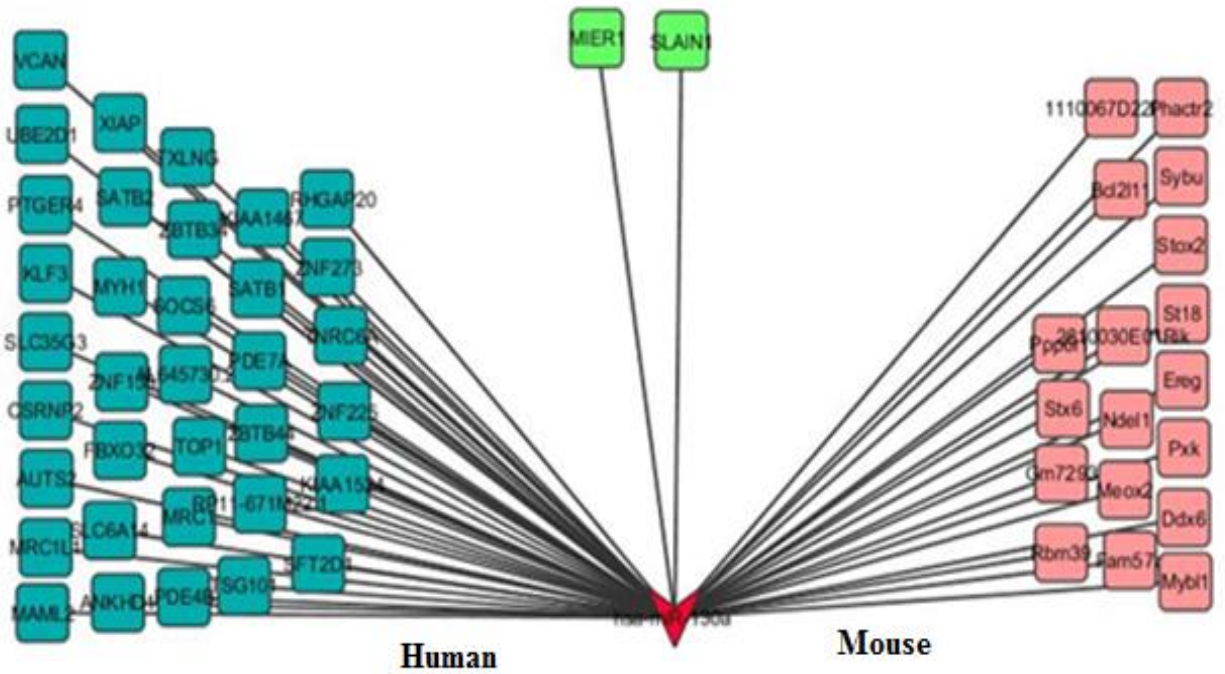
Aqua blue squares: human protein targets

Murine counterparts may be helpful for experimental validation, as genetics is possible in mice and as Dr. Schulze has expression vector technology available for siRNA knockout both in human and murine platelets. We have here no valid hit but we found a major apotosis pathway in platelets conserved between mouse and man and directly regulated by a number of miRNAs, again conserved. The first human platelet miRNA profiling study was performed in 2008 by Bruchova et al as part of a study testing for differentially expressed miRNAs in patients with polycythemia vera (Bruchova et al., 2008). These investigators found that miR-26b was significantly higher in polycythemia vera platelets than in platelets from healthy donors. Landry et al also described how miR-223 could target

the mRNA of the adenosine 5' diphosphate receptor, P2Y12, and that P2Y12 mRNA was found in Ago2 immunoprecipitates in both megakaryocytes and platelets.

3.5   Results for miRNA study in abdominal aortic aneurysm

Considering the complexity of mRNA regulation by miRNAs, whereby a single miRNAs may regulate hundreds of different mRNAs with varying efficiency, and each mRNA can be regulated by several microRNAs, the repertoire of potential platelet and extra-platelet mRNA targets of platelet microRNAs may be rather large (Ple´ et al., 2012).

Bioinformatic target prediction is often the first step towards defining the function of a specific miRNA. Among the different in silico algorithms are publicly available that predict miRNA binding sites, we utilized DIANA LAB microT, Miranda, TargetScan, and miRBase.

An enhanced understanding of the pathogenetic pathways has led to significant research and development of new molecules, which can inhibit these pathways and delay the expansion of Abdominal Aortic Aneurysm (AAA). We look into likely interactions of miRNAs that may have a beneficial role in preventing the progression of AAA. Dr Albert Busch had a set of mRNAs highly expressed in aortic aneurysm and our study is in the phase of experimentation. The bioinformatical results suggest down-regulation of several protein targets in abdominal aneurysm by these miRNAs. As a first approach to identifying potential miRNA target binding sites for these highly expressed protein mRNA, the full-length transcripts were used to search the TargetScanHuman database (version 6.2, http://www.targetscan.org/). The results identified target binding sites for the following miRNAs respectively (Table 3).

Table 3: Proteins with more than one targeting miRNA.

Protein Targets	Targeting miRNA	Seed Match	KEGG Pathway
TRPS1	mir-194, mir-19b	mir-19b (7mer-m8)	
NAA50	mir-194 (8mer), mir-362	mir-194 (8mer)	
EDN1	mir-194, mir-19b		HIF-1 signaling pathway, TNF signaling pathway, Melanogenesis
WDFY1	mir-194, mir-19b		
SBNO1	mir-362, mir-19b		
BCL11B	mir-550a, mir-19b		
TRAK2	mir-550a, mir-19b (8mer)	mir-19b (8mer)	Metabolic pathways, GABAergic synapse

Table 4: miRNA's with multiple protein targets

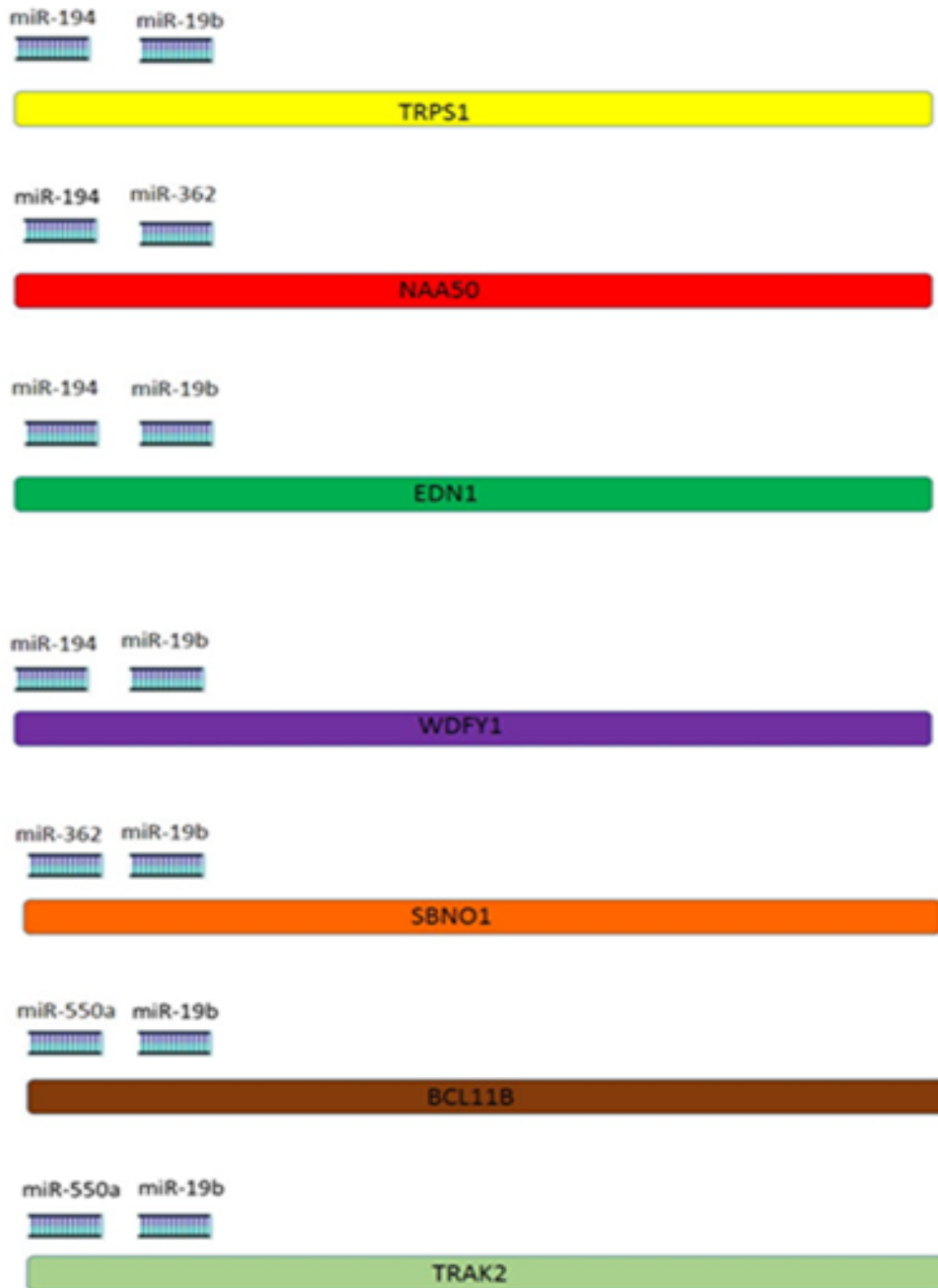| Targeting miRNA	Protein Targets
hsa-miR-194-5p	TRPS1, NAA50, EDN1, WDFY1
hsa-miR-362-3p	NAA50, SBNO1
hsa-miR-550a-5p	BCL11B, TRAK2
hsa-miR-19b-1-5p	TRPS1, EDN1, WDFY1,SBNO1, BCL11B, TRAK2
hsa-miR-769-5p	-

Figure 16: Proteins with their targeting miRNAs

To confirm the initial TargetScan results for the microRNAs, a series of other bioinformatics databases were also consulted like miRBase. No success was there to identify a potential platelet target binding domain for hsa-miR-769-5p however, a putative target could be identified for all other miRNAs (Table 4).

These bioinformatics approaches identify a refined and highly suggestive set of candidates for miRNA experimental analysis. These results prompted us to consider this smaller set of platelet miRNAs. This prediction analysis suggests that miRNA expression may be involved in down-signalling in aortic vesicular walls. These results increase the evidence supporting a possible involvement of mir-19b and mir-194 in AAA.

It is an impressive result that a set of the experimentally identified upregulated proteins are predicted to be regulated by a comprehensive set of just two miRNAs. These miRNAs can result in down-regulation of proteins destroying the aortic walls during heart disease and can reduce the damage caused by the aneurysm. Whether these miRNAs are involved as causative agents in the process of aneurysm or, on the contrary are counter reaction or regulation beneficial to the high levels of the proteins in aneurysm (Table 3) still remains to be determined. Nevertheless, monitoring the miRNA levels may be diagnostic or also prognostic of aortic aneurysm and even be an early marker. These levels are easily monitored in blood stream and detailed follow up of all these questions and predictions will be done by Dr. Busch in his laboratory.

4   Results: Isotopologue Data Analysis

Comparative analysis of software for metabolic modelling and metabolite data analysis (Table 4) show limitations in terms of technical solutions, which are needed to be addressed.

Firstly the need to have a quantitative analytical approach to completely analyse a cell of a living organism to provide information about its metabolism (Meyer et al., 2003) with (re)construction and simulation of the complex information based cellular function (Nahnsen et al., 2011) and also the capacity to improve the tendency to accommodate high diversity of biomolecules (Perillo et al., 2009).

These solutions may have improved the crosslinking and data compatibility between different metabolite databases like KEGG LIGAND (Eisenreich et al., 2010) but due to the heterogeneous naming conventions of metabolites in different databases they are time consuming and complex to construct biochemical reactions based network model (Schauer et al., 2010). Carbon atom mapping from substrates to products is difficult during metabolic flux analysis, as it requires a model (Schauer et al., 2010). Construction of new models for 13C metabolic flux analysis is tidious with the lack of good quality metabolic model repositories. (Eylert et al., 2010). Manually adding carbon mappings to stoichiometric model is a complex, lengthy and error prone task because even small central carbon metabolism models contain hundreds to thousands of carbon mappings (Schauer et al., 2010). Online repositories of SMBL models are needed for rapid metabolic network model construction, carbon mapping and information sharing with accurate prediction of metabolite concentrations and calculation of flux distribution (Schauer et al., 2010). Integration, analysis and quality visualization of disparate data types consisting of raw data expressions and big graphs

comprising many thousands of nodes and edges, to improve biological systems understanding mechanism is needed (Gaasterland and Sensen, 1996).

Third party software tool dependency is one of the biggest draw back now a days, as most of the applications are depending on other software applications to run e.g. METATOOL 5.0 and Fiat Flux needs Mat lab to execute, which in return costs extra effort of spending to acquiring and learning a new tool, which may not be freely accessible.

Many modelling based approaches have been developed to quantify metabolic flux but still these are unable to resolve high complexity datasets with maximum pre purification (Perillo et al., 2009). With the above mentioned glitches to the community of flux analyzers, some areas that need to be explored with respect to the flux analysis and visualizations are:

1. Comparison between the outcomes of existing approaches (algorithms) to conclude with the most optimal methodology.

2. Comparison between the outcomes of existing visualization tools to have the most presentable visualization for better analysis of results

4.1 Motivation for new software in isotopologue data analysis

The objective of this research was to perform Mass Isotopomers Distribution Analysis (MIDA) (Hellerstein MK and Neese RA, 1999) and to study metabolic isotope to quantify the fraction of metabolites of interest in the mixture typically by tracing isotopes. Isotopologues (isotopic isomers) are the species of a compound that differ only in their isotopic composition (Brenninkmeijer et al., 2003).

MIDA measures the mixtures of polymers (e.g. lipids, carbohydrates and proteins) by quantifying relative abundances of molecular species with Mass Spectrometry (converts individual molecules

into ions to direct them in magnetic fields using Mass Spectrometer) (Van-Bramer SE, 1998).

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 fragment with its peak.

All mass spectrometers consist of three main sections: ionizer, ion analyzer and detector (Hites RA, 1997). Electron impact ionization is performed by Ionizer with a gas chromatograph using a high-energy electron beam to collect molecular ions and fragments. Ion Analyser accelerates obtained molecular ions and fragments by maneuvering the charged particles using mass spectrometer, eliminating uncharged molecular ions and fragments. The job of the Detector is to generate an electronic signal at every ion hit. During this process mass analyser classifies the ions with respect to the mass to charge ratio values and detector extracts the abundance values of each m/z ratio value.

Estimating mass isotopomers distribution from spectral data is an extension of the quantitative mass spectrometric method to a multi component mixture analysis (Lee WN et al., 1991). Mass spectrometric analysis describes relative abundances quantitatively, based on combinatorial probabilities. It is essential to identify the quantity of the labelled isotopomer population to resolve the exact amount present in the mixture. The exact identification of the number of labelled 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., 1999). These generate identical results. A formal mathematical algorithm generates an appropriate set of linear simultaneous equations and finds their solutions, for the calculation of both natural isotope abundances and relative isotope abundances.

Some software applications do exist for mass isotopomer distribution analysis and metabolic modelling. As it is a broad field contributing towards the development of understanding between complex interactions, control and regulation of metabolic networks. Metabolic flux analysis is one of the key methods of metabolic modelling (Schlatter et al., 2012). We (our group) have also developed some computational software applications (e.g. Isotopo, Yana (Schwarz et al., 2005), Yanasquare (Schwarz et al., 2005) and Yanavergence (Schwarz et al., 2007) etc.) for complex pathway analysis and isotopic distribution prediction (Eisenreich et al., 2006). Most importantly there is an existing solution provided by Priv. Doz. Wolfgang Eisenreich's group at the Institute of Biochemistry Technical University Munich, Germany, implementing the similar methodology but with a completely different way of development and usage; in the form of a Microsoft excel macro. To process experimental metabolite data using a developed macro, a user at first needs to store experimental data in Microsoft excel sheets in a particular user unfriendly way and has to set many paths and configurations. To meet the aforementioned goals of the research, a new software application is needed to be designed and developed.

One of the objectives of this research was 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. Focusing on identification of the quantity of population of labelled isotopomers for resolving the exact rate of synthesized fractions present in the mixture and metabolic experimental data management, two new software applications named "LS-MIDA" and "Isotopo" are developed. These software applications have a facile data management and robustness to quantify the populations of isotopomers in mixtures of 13C-labelled amino acids. Isotopo is the upgraded version of LS-MIDA.

4.2 LS-MIDA

Least Square Mass Isotopomers Distribution Analyzer (LS-MIDA) is a software application with the ability of performing quantitative mass spectrometry readily to mixtures of materials labeled with stable isotopes that can be very important for both bio-medicine and biochemistry. LS-MIDA is developed to estimate mass isotopomer distributions from spectral data, by analyzing each peak of given mass. The currently available version (ver. 1.0.0) is capable of processing experimental data consisting of metabolite information, mass to charge ratio (m/z) values, actual observed relative intensities and number of fragments, and as a result it calculates mass values (Mo, M-1 and Mmax), natural abundances for pure compounds, linear relationships between masses within range of a compound by drawing Abundance Matrixes, and Relative Abundances. In addition it visualizes the obtained results in an intuitive way and provides a special interface for experimental data manipulation and management (Ahmed et al. 2013a).

4.2.1 LS-MIDA, method

Brauman´s algorithm (Brauman, 1966) uses the least squares technique to calculate relative isotopic abundances by simplifying the mass spectra of molecules containing elements with many isotopes. Furthermore this method can deal with complex spectra based on fragmentation of molecules containing heteroatoms (atoms which may not be carbon or hydrogen itself but can occur in several different isotopes). The method has two parts: generation of an appropriate set of linear simultaneous equations and solutions of these equations. The proposed equation to calculate relative abundances is

$$X = A^T (A^{T} * A)^{-1} P$$

(Eq 1): Where 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 however advisable to look for contaminating fragment ions in the mass spectrum, as its impact is considerably corrected through the subtraction of natural abundance values. The proposed binomial expression (Lee et al., 1991) calculating natural abundances is

$$A = n! \, / \, [(i!) * (n - i)!] * P_0^{(n-i)} * P_1^i$$

(Eq 2): Where A = Calculated Natural Relative Abundance, n = number of fragments, i = count value, (loop from 0 till n-1) and p= proportion of labeled carbon.

The reliability in the outcome by 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.

LS-MIDA was implemented using 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 LS-MIDA was extensively tested and results are presented from Salmonella labeling experiments for different metabolites such as analysis of various amino acids in Table 5. The statistical information is 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 (Ahmed et al. 2013a).

(A) Main bar

(B) LS-MIDA data analyzer main

(C) LS-MIDA data analyzer, fragment viewer

(D) LS-MIDA data analyzer main results viewer

Figure 17: LS-MIDA; analyzer with inputted data analysis

Figure 18: LS-MIDA; data manager with inputted data management

The graphical user interface of LS-MIDA is based on the principle of embedding child windows under a single parent window by creating nesting hierarchies. LS-MIDA is a multiple document interface (MDI) software application so it is divided into two main modules i.e. LS-MIDA data analyzer and LS-MIDA data manager.

LS-MIDA Data Analyzer module is responsible for providing options for experimental data loading, analysis and visualization, whereas LS-MIDA Data Manager is responsible for providing options for experimental data manipulation and management. LS-MIDA Main is the parent window of the application embedding all the module interfaces and responsible for providing options for child windows manipulation and closing over all applications. Finally, LS-MIDA About is a simple interface providing basic information about the current version of in use LS-MIDA version (Ahmed et al. 2013a).

4.2.2   LS-MIDA, Experimentation and Results

Experimental data has to be collected to run the LS-MIDA. The process consists of three major steps i.e. preparation of a data set, input data file preparation and management, and evaluation and data analysis. Observed data during actual experimentation are collected during the preparation of a data set which is later used by LS-MIDA Data Analyzer, Figure 42, for analysis. Input data file preparation and management requires LS-MIDA Data Manager, Figure 43, to be used to structure data by organizing an experimental data file. Observed outcome data during different experiments of metabolic isotopomers analysis with different metabolites is collected for mass isotopomers predictions using LS-MIDA. The collected data consists of seven different results observed using seven different metabolites during Evaluations i.e. Alanine (Ala), Glycine (Gly), Valine (Val), Leucine (Leu), Isoleucine (Ile), Proline (Pro), Serine (Ser), as shown in Table 5.

Table 5: Experimental dataset

No.	Metabolite	m/z Values	RI Values	Atomic Mass Values	Atomic Fragment Numbers
1	Ala	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#	3#	2#
2	Gly	217,15#218,1#219,1#220,1#221,05#	0,27# 100# 31,06# 10,72# 1,93#	2#	1#
3	Val	185,25# 186,15# 187,15# 188,15# 189,1# 190,1# 191,1# 192,1#	0,23#100#16,62#4,94#3,65#0,68#0,33#0,04#	5#	4#
4	Leu	301,25#302,15#303,15#304,15#305,15#306,15	0,05#27,44#7,25#2,83#0,47#0,08#	6#	2#

5	Ile	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#	6#	5#
6	Pro	285,15#286,15#287,15#288,1#289,1#290,1#292,1#293,1#	2,06#16,11#4,07#1,49#0,21#0,06#0#0,02#	5#	5#
7	Ser	287,25#288,15#289,15#290,15#291,15#292,15#	0,36#80,22#20,67#9,86#1,79#0,41#	3#	2#

The metabolite based data set example consists of following experimental elements i.e. (m/z) values, Relative intensity (RI), Atomic Mass Values and Atomic Fragment Numbers.

The metabolite (seven in total in the example) based data is taken as input and analysed using LS-MIDA Data Analyzer item by item during experimental data analysis. Each item given as input consists of five main information elements i.e. metabolite name, m/z values, relative intensity values, atomic mass values and atomic fragment numbers. The resultant information from LS-MIDA Data analyzer, Table 6, consists of three different items i.e. natural abundances, relative abundances and percentage of calculated relative abundances per m/z values, in each experiment. Furthermore based on observed results i.e. calculated relative abundance values, actual relative

abundance values and percentage of relative abundance per mass to charge ratio values (abundance versus m/z values), a spectrum is drawn graphically for better analysis. Drawn spectra are shown in Figure. 44 (a, b, c, d, e, f and g): blue colored bars represent calculated relative abundance values, yellow bars are representing actual relative abundance values and green bars are the percentage of calculated relative abundances per m/z values.
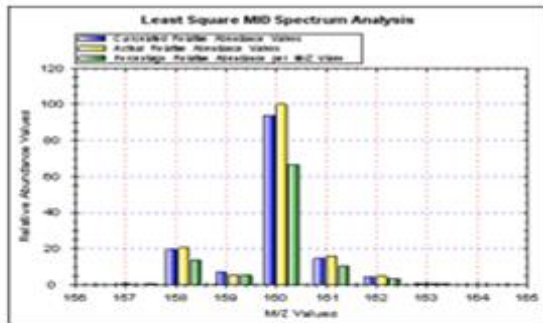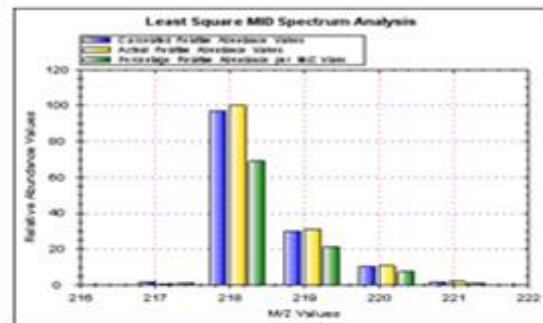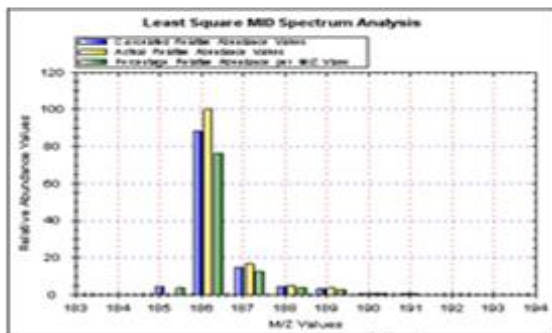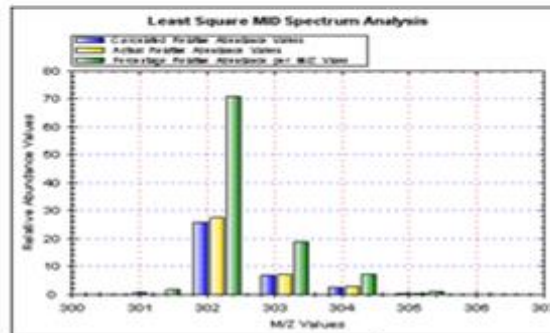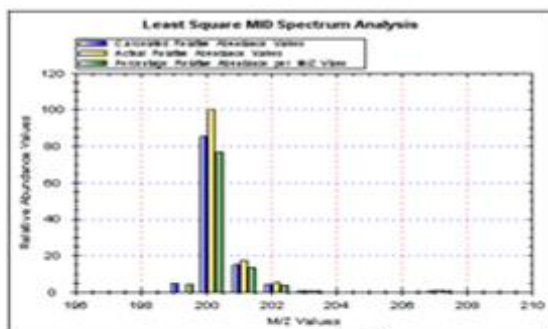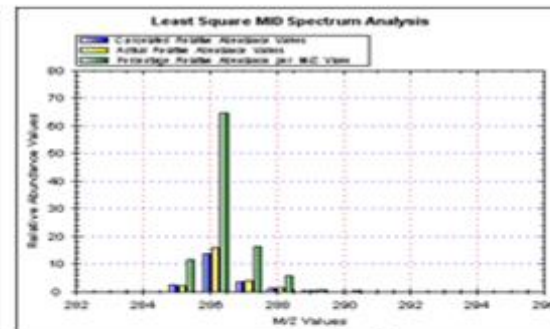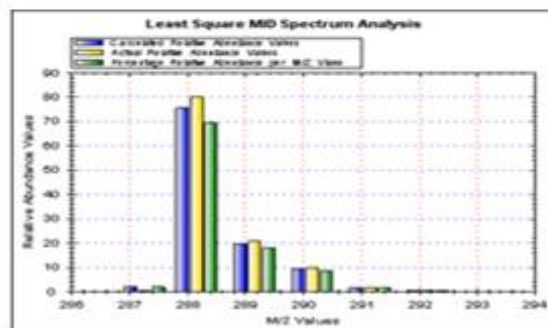
Figure 19: LS-MIDA data analyzer; metabolite spectrum analysis

Table 6: Calculated abundances of input metabolites using LS-MIDA data analyzer

Metabolite/ Abundances	Ala	Gly	Val	Leu	Ile	Pro	Ser
Natural Abundances	0,97792321# 0,02195358# 0,00012321#	10,9889# 0,0111#	0,9563338046567 04# 0,0429378308491 836# 0,0007229395122 246# 5,4098011836E-06# 1,51807041E-08#	0,97792321# 0,02195358# 0,00012321#	0,945718499425015# 0,053076526158447 1# 0,001191524806064 85# 1,33743809761551E-05# 7,506099142245E-08# 1,6850581551E-10#	0,945718499425015# 0,053076526158447 1# 0,001191524806064 85# 1,33743809761551E-05# 7,506099142245E-08# 1,6850581551E-10#	0,97792321# 0,02195358# 0,00012321#

| Relative Abundances | 0,61021797758509 8#
19,36685117801#
7,33856270174111#
93,84872311106719#
14,62546352693118#
4,53055431 47529 | 1,34660319393963#
97,0439809103189#
30,1535052677855#
10,3879218215235#
1,86644174637217# | 4,13916401598954#
88,1197854124893#
14,7328992933026#
4,46449966534509#
3,21935190634397#
0,607739581444572#
0,290201453624253#
0,034985 | 0,62371676475032 8#
25,8150117518592#
6,8398235150824#
2,6565525513644#
0,44123347737325#
0,0748176819265 6118# | 4,96883125695095#
85,4216661577551#
15,0825925419121#
4,41483685712431#
0,60022954672092#
0,147157161815149#
0,0615490581271288#
0,0661999107177709#
0,72741810237568 8# | 2,51152933631727#
13,8211983003602#
3,51350441636556#
1,27032695758267#
0,18047383590742 5#
0,050771413767800 4#
0,000949415363602 28#
0,016916700055248 5# | 2,02333005658576#
75,4585579852112#
19,5382396970383#
9,25890856942684#
1,68265356288747#
0,38344061987337 3# |

	8# 0,598541455412094#		5345553543#				
Percentage of Relative Abundances per m/z Values	0,433027731420865% 13,7432588655743% 5,20764919316322% 66,5976768271987% 10,37	0,956404822511829% 68,924039209234% 21,4160771217053% 7,37786644996354% 1,325	3,58032452101624% 76,2224998285309% 12,743771544618% 3,86173574281097% 2,78469868010487% 0,5256870511612	1,71110284995867% 70,820831949266% 18,7643529424077% 7,28797893307261% 1,21047870331454% 0,205254	4,4567314002956% 76,6179011006398% 13,5281438035092% 3,95983301320681% 0,538368427085172% 0,131990786147647% 0,0552056622228945%	11,7549755853737% 64,688811805628% 16,4446252075402% 5,94564521142425% 0,844690724577112% 0,237630801538049% 0,00444364883903296%	1,86748591968645% 69,6464692443778% 18,0333344087459% 8,54575422768446% 1,55304955124576% 0,353906648259696%

86376 74867 1% 3,215 00796 28273 6% 0,424 74174 49485 05%	61239 65853 1%	13% 0,251020 5868701 42% 0,030262 0448876 922%	6219805 42%	0,0593771 86613034 4% 0,6524486 20279876 %	0,0791770 15079634 1%	

Data set consists of Salmonella isotopolog measurements. Collecting these measurements along with further in depth data on their metabolism permits insights into Salmonella and their metabolism during infection (Eisenreich et al., 2010). So glucose, glucose-6P and gluconate provide possible carbon sources for the intracellular pathogen. In disease condition 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 (Ahmed et al. 2013a).

4.3 Isotopo

We developed a new software applications for MIDA, capable of easily processing experimental data. Input data includes: metabolite information (ion), mass to charge ratio (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 gas chromatography – mass spectrometry (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 13C) 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 molecules. The software estimates mass values (M0, M-1, Mmax): These are three values estimated from m/z values, M0 is the first fraction less m/z, M-1 is the first m/z value minus 1 and Mmax 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 (input 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.

4.3.1   Isotopo, Method

For probabilistic estimates, some already 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.

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. Natural abundance values can be estimated by the isotope contents of biosynthesized subunits from polymerized product. We used for this the binomial expansion for the measurement of natural abundances (eq. 2).

Here 'A' represents the calculated natural relative abundance values, 'n' is equal to the number of fragment(s), 'i' is index running from 0 till n-1 and 'P' is the distributed proportion of mass isotopomers.

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

$$
\begin{pmatrix}
A_n & 0 & 0 & 0 & 0 \\
A_{n+1} & A_n & 0 & 0 & 0 \\
A_{n+n} & A_{n+2} & A_{n+1} & 0 & 0 \\
A_{n+n} & A_{n+3} & A_{n+2} & A_{n+1} & 0 \\
A_{n+n} & A_{n+4} & A_{n+3} & A_{n+2} & A_{n+1} \\
\dots & \dots & \dots & \dots & \dots
\end{pmatrix}
$$

(eq. 3): Where in eq.2, An, An+1, An+2 … An, are estimated natural abundance values plotted in the form of an Abundance Matrix.

The population of (first) abundance matrix depends upon the number of m/z values e.g. if the number of m/z values is five, then a square matrix of five rows and five columns will be drawn. Brauman's algorithm (Brauman, 1996 ; 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 heteroatoms). The method is 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. 1. But we are using the partial Brauman's least square algorithm (equation) to calculate relative intensity values, presented in eq. 4.

$$X = A^{-1} P$$

(eq. 4): 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 calculate relative intensity values, linear regression analysis is performed using drawn abundance matrix (eq. 2), with the partial implementation of Brauman's least square method (eq. 4).

$$
\begin{pmatrix}
A_n & 0 & 0 & 0 & 0 \\
A_{n+1} & A_n & 0 & 0 & 0 \\
A_{n+n} & A_{n+2} & A_{n+1} & 0 & 0 \\
A_{n+n} & A_{n+3} & A_{n+2} & A_{n+1} & 0 \\
A_{n+n} & A_{n+4} & A_{n+3} & A_{n+2} & A_{n+1} \\
\ldots & \ldots & \ldots & \ldots & \ldots
\end{pmatrix}
=> A\text{-}1^*P =>
\begin{pmatrix}
Ri_1 \\
Ri_2 \\
Ri_3 \\
Ri_4 \\
\ldots \\
Ri_n
\end{pmatrix}
$$

(eq. 5): To compute relative abundance values, again linear regression analysis is performed by drawing a new abundance matrix (second threshold) but this time the input values are estimated relative intensity values (Ri) using eq. 5 and the length of the abundance matrix depends on the user input 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.

$$\begin{pmatrix} Ri_n & 0 & 0 & 0 & 0 \\ Ri_{n+1} & Ri_n & 0 & 0 & 0 \\ Ri_{n+n} & Ri_{n+2} & Ri_{n+1} & 0 & 0 \\ Ri_{n+n} & Ri_{n+3} & Ri_{n+2} & R_{n+1} & 0 \\ \dots & \dots & \dots & \dots & \dots \end{pmatrix} => A\text{-}1\ P => \begin{pmatrix} Ra_1 \\ Ra_2 \\ Ra_3 \\ \dots \\ Ri_n \end{pmatrix}$$

(eq. 6): In eq.6, Ra1, Ra2, Ra3 … Ran, 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.

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. Mathematically it is not possible to implement eq.4 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. 7.

$$X = \text{Pseudo Inverse } (A) * P$$

(eq. 7): X is equal to the product of Pseudo Inverse of A times P. In eq.7, A is the second drawn abundance matrix consisting of computed relative intensity values and P are the actual observed relative intensity values.

The above mathematics was implemented in the software "Isotopo". The available and tested version of Isotopo provides two main modules i.e. Isotopo Analyzer and Isotopo Data Manager, Figure 20 and 21.
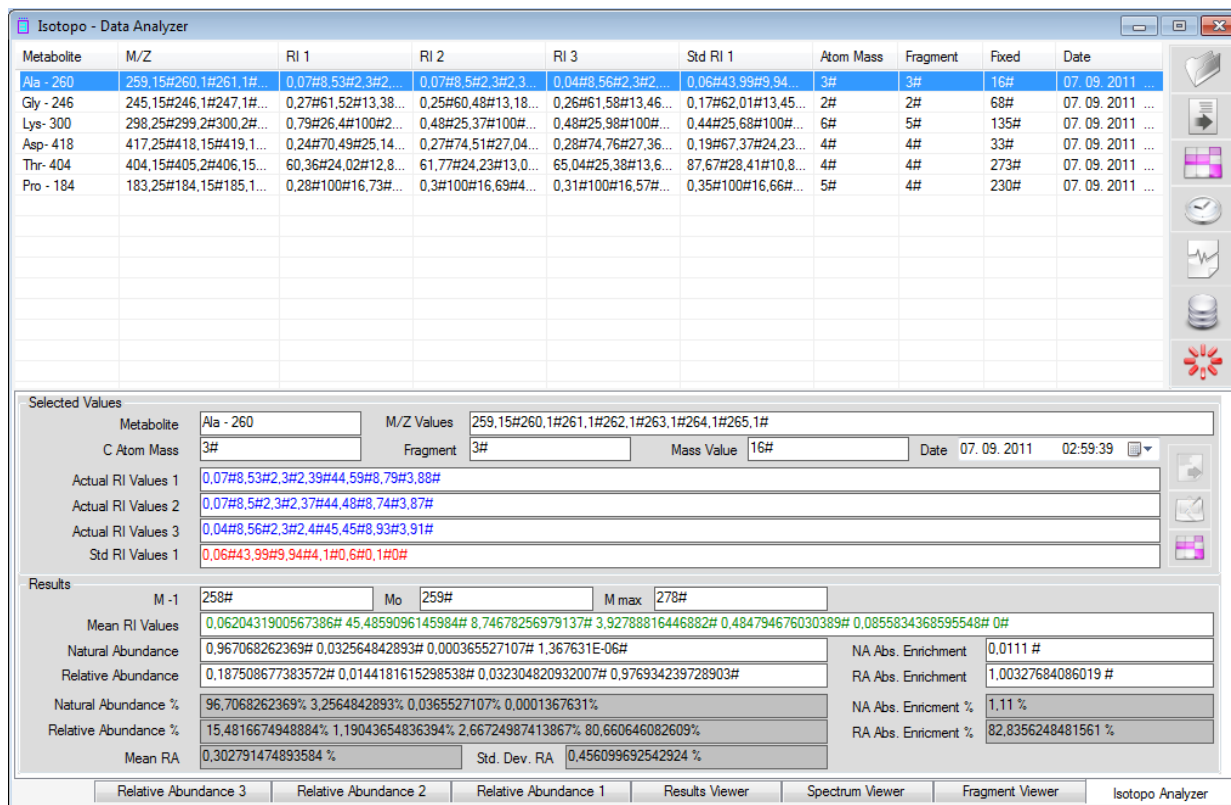
Figure 20: Isotopo; analyzer with inputted data analysis

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 (M0, M-1, Mmax), predicts natural abundance values, relative abundance values and calculates 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

77

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 (Ahmed et al. 2012b).

Figure 21: Isotopo; data manager with inputted data management

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 (Ahmed et al. 2014b).

4.3.2 Isotopo, Experimentation and Results

To apply Isotopo, experimental data have to be collected first from GC-MS experimentation. The experimental process using Isotopo consists of three major steps: preparation of data set, input data file preparation and management, and evaluation and data analysis. The observed data during actual experimentation are collected during the preparation of the data set. During input data file preparation and management at first Isotopo Data Manager is used to structure data by organizing an experimental data file which is later used by the Isotopo Analyzer for analysis. Throughout the experimental data analysis, each data point is individually analysed and evaluated. Data for different experiments are collected by Isotopo Analyzer.

It includes statistical calculations. We have successfully tested the preparation of different data sets; input data file preparation and management, experimentation and data analysis using Isotopo. Example data and results are presented from Salmonella labeling experiments for different metabolites. For detailed step by step calculation example we have used experimental raw data of alanine consisting of metabolite information, mass to charge ratio (m/z) values, three actual relative intensity values, standard relative intensity values, atomic mass value and atomic number of fragments, presented in Table 7.

The example Ala 260 dataset consists of seven m/z values (259.15, 260.1, 261.1, 262.1, 263.1, 264.1, 265.1) as well as standard relative intensity and three fragments (0-3), seven standard relative intensity values and seven relative intensity values respectively (please see above mentioned definitions).

Table 7: Isotopo Data experimental data set Ala (260).

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#2,3 7#44,48# 8,74#3,8 7#	0,04#8, 56#2,3 #2,4#4 5,45#8, 93#3,9 1#	0,06#43, 99#9,94# 4,1#0,6# 0,1#0# #	3#	3#

For the example (Ala 260) the calculated results are shown in Figure 22a-g: M0, M-1, Mmax, relative intensity values, natural abundance values, percentage of natural abundance at each fragment, three different relative abundance values with respect to each RI value, three fractional molar abundance values with respect to each RA value, average relative abundance values, percentage of average relative abundance at each fragment, absolute 13C enrichment of natural abundance, absolute 13C enrichment, standard deviation and mean of average relative abundances (Ahmed et al. 2012b).

(a). Data analyzer

Metabolite	Groups	Natural Aubndance Percentage	Relative Abundance Percentage	Natural Aubndance Value	Relative Aubndance Value
Ala - 260	[0000]	96,7068262369 %	15,4816674948884 %	0,967068262369 #	0,187508677383572 #
	[XXXX]1	3,2564842893 %	1,19043654836394 %	0,032564842893 #	0,0144181615298538 #
	[XXXX]2	0,0365527107 %	2,66724987413867 %	0,000365527107 #	0,032304820932007 #
	[1111]	0,0001367631 %	80,660646082609 %	1,367631E-06 #	0,976934239728903 #
.	.	.	.	.	.

Relative Abundance 3 | Relative Abundance 2 | Relative Abundance 1 | Results Viewer | Spectrum Viewer | Fragment Viewer | Isotopo Analyzer

(b). Fragment viewer
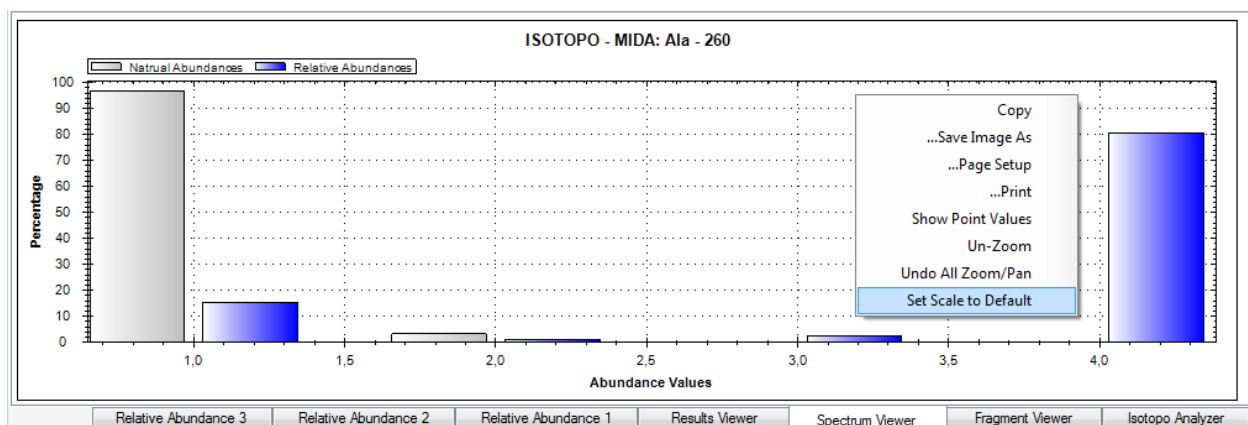
(c). Spectrum viewer

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 %	RA /
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...	82,8
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...	10,2
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...	3,03
Asp- 418...	417,25...	4#	4#	07. 09. ...	0,24#7...	0,27#7...	0,28#7...	0,19#6...	416#	417#	454#	0,1986...	95,633...	1,11 %	56,249...	31,3
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...	46,3
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...	1,20

Relative Abundance 3 | Relative Abundance 2 | Relative Abundance 1 | Results Viewer | Spectrum Viewer | Fragment Viewer | Isotopo Analyzer

(d). Results viewer

(e). Relative abundance 1

(f). Relative abundance 2

(g). Relative abundance 3

Figure 22: Isotopo; analyzer with inputted data analysis Ala 260

Table 8: Isotopo data analyzer's output after processing Ala (260)

Metabolite	Mo	M-1	M max	NA	RA	NA Abs. Enrichment	RA Abs. Enrichment	RI Values
Alanine (Ala – 260)	259	258	278	96.7068262369% 3.2564842893% 0.0365527107% 0.0001367631%	15.48166749488884% 1.19043654836394% 2.66724987413867% 80.660646082609%	1,11 %	82,8356248481561 %	0.0620431900567386, 45.4859096145984, 8.74678256979137, 3.92788816446882, 0.484794676030389, 0.0855834368595548, 0

The resultant mass values are: M-1 at 258, M0 at 259 and Mmax at 278. The natural abundance of the first fragment is 96.7068262369%, at second 3.2564842893%, at third 0.0365527107% and at fourth and last 0.0001367631%. Likewise average relative abundance at first fragment is 15.4816674948884%, at second 1.19043654836394%, at third 2.66724987413867% and at fourth is 80.660646082609%. The highest observed natural abundance is at first fragment and highest

average relative abundance is at fourth fragment. Conclusively the resultant 13C absolute enrichment of natural abundance is 1.11% and average relative abundance is 82.8356248481561 %, and standard deviation of average relative abundance is 0.456099692542924 % and mean is 0.302791474893584 % (please see above mentioned definitions).

We tested the processing 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, Isotopo Data Manager Structures the data into an experimental data file which later is used by Isotopo Analyzer for analysis. Throughout the experimental data analysis, each observed resultant data set is individually analysed. Results by Isotopo Analyzer are presented in Table 8 and Table 9. These results were validated in an experimental setting regarding metabolism of Salmonella under different growth conditions.

Table 9: Isotopo Analyzer Output (RA1, RA2 and RA3) after processing Ala 260

Ala 256, RI	1	2	3
Relative Abundance	15,55001272333797% 1,19508703038 98% 2,68720500478513% 80,5676952414454%	16,09359350277 63% 1,19181064136482% 1,79122318111892% 80,92337267474%	16,09359350277 63% 1,191810641364 82% 1,791223181118 92% 80,92337267474 %
Fractional Molar Abundance	0,187508582982646# 0,0144108612382297# 0,0324034463247449# 0,971519100186158#	0,186849716055398# 0,014547280874517# 0,0319686204759015# 0,969065867576428#	0,188167733112 672# 0,014296342476 8147# 0,032542395995

			3742# 0,990217751424 123#

5 Discussion

Computer science has revolutionized almost all other fields of life. Common man including engineers, doctors, artists, technicians and scientist etc., somehow, every one's life is now partially depending on the usage of informatics. In the past (1980s), the informatics (IT) issues were related to the development of the large sized but small scaled applications. Later on (1990s), with the passage of time, systems started becoming complex but smaller in size, especially with the evolvement of the concept i.e. Component Based Systems (CBS) (Mahmood and Lai, 2013) and the innovations of advanced programming tools and technologies (Szyperski et al., 2002) e.g. Enterprise Java Beans, Microsoft COM and CORBA etc. So far the focus of the last decade (2000s), was to develop smart, intelligent and robotic applications.

Particularly in life science, with the front runner field bioinformatics, the world has been changed by small, efficient, fast, logical, embedded and intelligent software, databases and management systems. Even last year's (2013) Nobel Prize winners (Arieh Warshel, Martin Karplus, Michael Levitt) in the field of chemistry relied on powerful computational programs to understand and predict biochemical processes and molecular dynamics, giving testimony to the novelty and innovation of bioinformatics.

In all the previous mentioned major phases of time, where developmental programming languages, tools and technologies have been improved a lot, some things remained the same and need to be discussed and matured, and one of them is the software engineering of scientific solutions, academic applications.

One has to understand the differences between the academic prototype solution and the commercial software system development. Here, the focus is on the comprehensive and stable academic

scientific solutions' development in different fields of life e.g. biology, physics, chemistry and medicine etc. which includes the principles of informatics, engineering, statistics and mathematics e.g. bioinformatics, neuroinformatics, health informatics, biomedical engineering, computational biology, systems biology, biotechnology etc.

In the scientific world, the major concern is finding solutions to problems and not the full fledge system development. Being asked to do so in little time and resources but with high expectations. Ignoring these exceptions, most of the software developers are not the professional programmers or with good informatics background but only the researchers with some interest and little informatics backgrounds. In comparison to the professional or commercial software development world, it seems a lot different, as there the achievable goals are set with finalized requirements analysis, the system is fully or at least partially well designed, implemented using most recent or reliable technologies, assured quality application is deployed and users are trained to use it.

These could be the reasons that most solutions developed in academia are the open source scripts, small databases, web pages etc., implemented without formal requirement analysis and design, conceived on the basis of ideas. Moreover, the target of most of the scientific people is not to have a comprehensive prototype software application, well usable and extendable but to have it published. This goal ultimately kills all the software developmental goals, need of required expertise and end product, when the software project work hasn't even started.

This section is about to briefly highlight the importance of the proper software application development in scientific academia, with the involvement of professional software designers and developers. Conceptual hypothesis of this research states that, doing so, not only the comprehensive software solution development will be achieved but it will also expedite the processes of application usage, maintenance and future extensions. Moreover in return, it will not only save

time and efforts, but will also help in efficient experimental data analysis, storage, manipulation, maintenance, sharing and reuse.

5.1 Software engineering and development life cycles

To establish and expedite the processes of software engineering (SE), many Software Development Life Cycles (SDLC) and process improvement models have been introduced (Rook, 1986).

SDLC is a goal-oriented approach towards software development. Almost all of the proposed SDLC models provide distinct processes for software implementations, and depending upon the nature of the end product, the right model has to be chosen and applied. Based on the process artefacts and the logical steps for developing a software project (e.g. time, quality, size, development effort etc.), it is not easy to compare different SDLCs. But doing so, where one will conclude with the variabilities, there, will also be commonalties (Ahmed, 2010).

The five modular approaches follow the procedures of some life cycle management approaches, which can help them in performing their individual functionalities as well as regulating tasks in cyclic chain processes.

5.1.1 Requirement engineering

This denotes the process of gathering specifications of the end system (project or product) to be developed (Lee and Xue, 1999). It involves various methods to collect the system's requirement e.g. by interviewing individuals and teams, brain storming, documentations etc. One important task during requirement engineering is the requirement analysis, which helps in examining the provided user end specifications and classifying the functional (which can be implemented) and the non-functional (which cannot be implemented) requirements. In most cases, the overall requirements of the whole system are decided at the beginning of the project, before moving towards design and

implementation. A critical step is strong interaction with the user, patiently noting his observations and requests, but bearing in mind the distinction between functional and non-functional requirements before wasting unnecessary programming effort.

5.1.2 Design modelling

Design modelling starts from the finalized functional requirements and achieves the development of a software application. It consists of the sets of the substantial decisions, selection of structural elements, collaborations among components and necessary constraints to be followed (Sommerville, 2005).

Reimbursements are fast implementations based on a big picture of the noted requirements. Incremented reliabilities differentiate the functional from the non-functional requirements from the start, reduce the level of complexities in pre-processed source code write-up and help in testing. Different languages have been proposed to be of advantage in formally designing a software system i.e. Architecture Description Languages (ADLs) e.g. UML, ACME, C2, Rapide, Wright etc.

A variety of academic as well as commercial tools are available for design modelling that implement these languages (Egyed and Kruchten, 1999) e.g. Rational Rose, Astah, UModel, UML etc., which help in designing software application' functionalities: This includes user access by textual and visual representation of relationships among the users and the use cases (Pfleeger and Bohner, 1990), optimized work flows regarding the flow of states and activities in computational and organizational internal processes (Jacobson et al., 1992), system sequences with particular scenarios between different objects at specific times (Bruza and van-der-Weide, 1993) and preprocessed source code structures regarding static representations of the relationships between defined classes in namespaces or packages (Marilyn, 1978), and data flow diagrams. Finally, there is compilation, execution and integration with other components involved.

5.1.3 Programming

The art of writing sets of machine communicating instructions formulizing scientific, computing problems to solution by providing executables is programming. Originating from low to high, many different programming languages (e.g. Binary, Assembly, LISP, Cobol, Fortran, Pascal, C, C++, Java, C# etc.) have been introduced advancing system, embedded, statistical, computational, distributed, intranet, enterprise, database, web, semantic web, mobile and bioinformatical system development. Each programming language syntax is comprised of a particular grammar specifics: expression, atom, number, symbol and lists, which are used to define the semantic of the programs. Furthermore there are different (proposed, patent) software development methods which facilitates context understanding for developing a software, model based solution development, application development for object model, multilingual build integration for compiled applications, implementing instrumentation code, monitoring in-development solutions.

5.1.4 Testing

This ensures the quality of the software application with the implementation of the processes to verify, validate and measure the performance of the system up to the claimed and expected functionalities, and to avoid unexpected system behaviors in real time functioning environment. There are different types of testing procedures which help to ensure the quality of the software system at different levels of developments e.g. black box testing, white box testing, unit testing, incremental integration testing, integration testing, functional testing, system testing, end-to-end testing, sanity testing, regression testing, acceptance testing, load testing, stress testing, performance testing, usability testing, install/uninstall testing, recovery testing, security testing, compatibility testing, comparison testing, alpha testing, beta testing etc. Nowadays different tools

are also available to perform software quality assurance e.g. Beye, Cubic Test, Duma, Fastest, Framework for Integrated Testing, Java Path Finder Open Java, Quick Check etc.

5.1.5 Deployment

The last expected step of the software development life cycle, releasing and installing qualitatively ensured software applications at the user end to meet the particular developmental goals based on the functional requirements initially defined and committed to.

These five modular SE approaches remains the same, when it comes to the software engineering of the scientific software solution development in academia but in contrast to the commercial software application development, the major and the only change is the inconsistency in all phases of the SDLC e.g. requirements are unclear throughout the developmental process, programming structure can become complex with the possibilities of error proneness and deployment procedures can also be complex and time consuming. Importantly testing of integrated and individual modules becomes time consuming, as new test cases have to be rewritten all the times or the application exists with a high expectation of ripple effects (i.e. unidentified logical or syntax errors in the system which arise while fixing the identified logical or syntax errors). The quality of a software application decreases with an increase in the ripples a change in software creates. Moreover measured optimum software maintenance can only be achieved with the accessibility of the concrete information about the ripples effect in the system. Depending upon the nature of the system, many approaches have been proposed to improve the software quality measurement processes, towards the commercial software development but one can also use these in the scientific software solution's quality assurance and for improvements as well.

Different SDLCs have been proposed so far and well used in commercial software development e.g. Waterfall Model (Petersen et al., 2009), V-Model (Rook, 1986), Spiral Model (Boehm, 1988),

Iterative and Incremental Model (Larman and Basili, 2003), Rapid Prototype Model (Hull et al., 1995), Agile Development Model (Ambler, 2002), Extreme Programming Model (Ambler, 2002), Evolutionary Model (Pei et al., 2011), Code and Fix Model (Pei et al., 2011).

(1)   Waterfall model, a sequential design process with following phases: Requirement Gathering and analysis (possible requirements of the system), System Design (based on requirement specifications), Implementation (system design development), Integration and Testing (all units developed are qualitatively assured), Deployment of system (after functional and non-functional testing), and Maintenance.

(2)   V-Model is the an extended form of the waterfall model proposed by Paul Rook, initiates with the system requirements, then defines the requirements, states preliminary requirements, give detailed design, asks for code and debugging, then unit, component, integration and acceptance testing and in the last system integration.

(3)   Spiral Model, a firm Product Line Architecture based model, which positions on four main pillars: determine objectives, identify and resolve risk, development and test, and plan the next iteration.

(4)   Iterative and Incremental Model syndicates iterative design or iterative method for the comprehensions system development. It consists of the following phases: initial planning, planning, requirements, analysis, implementation, testing, evaluations and deployment.

(5)   Rapid Prototype Model is mainly proposed to take advantage in quick software development. It's based on following five phases: requirements capture, quick design, build prototype, customer's evaluations and engineer final product.

(6)   Agile Development Model is mainly based on the iterative (cyclic) and incremental development, towards rapid and flexible software system development. It is closer to the followed developmental approaches, which consists of less documentation and analysis but direct user feedback and implementation. Its phases are: planning, requirements analysis, designing, building and testing.

(7)   Extreme Programming Model, it is a similar SDLC model to the Agile SDLC, as it also promotes the fast development. It is a cyclic model, consisting of the following six phases: select user stories, break down stories, plan release, develop/integrate/test, and release software and evaluate system.

(8)   Evolutionary Model, it is also called as the Evolutionary Prototyping, consists of four phases: initial concept, design and implement initial prototype, refine prototype until acceptable, and complete and release prototype.

(9)   Code and Fix Model, it is also a kind of model which supports quick development, consisting of the following four phases: implement the first version, modify until client is satisfied, post-delivery maintenance and retirement.

5.1.6   Evaluating SDLCs

Based on the process' artifacts and logical steps for developing a software project (e.g. time, quality, size, development effort etc.), it is not easily possible to compare different SDLCs, but doing so reveals differences. Depending upon the distinguished novelty, workflow and the level complexity, a comparative analysis on these models is performed and resulted with the information

that all of them are mainly proposed for the commercial software engineering purposes. The major focus of all is the comprehensive or quick software development.

The comparative study was based on the following, author defined twenty four metrics:

(1)   Software Engineering Approach: it's about the inclusion of the processes to verify whether the targeted and developed application is a software (Boehm, 1976).

(2)   Initial, Developmental Plan: it's about verifying if there is a documented preliminary plan to be followed and used during the software development.

(3)   Software Requirements Engineering: it's about looking for the preliminary application of the requirement engineering principles during the software developmental processes.

(4)   In Depth Requirements Analysis: it's about determining whether intense requirement analysis processes were asked to be implemented, or if only some preliminary requirements analysis was required. It includes the criteria to find errors in the software application, or in the implementation of analysis procedures for state-machine modelling and semantic analysis (Jaffe et al., 1991) .

(5)   Requirement Validation: it's about confirming whether the decided requirements needed be validated and classified into functional and non-functional requirements.

(6)   Functionals, Risk Analysis: it's about verifying whether the model enforces the functional, risk analysis to identify the possible risks (e.g. labour, technology, finance) during the implementation of software.

(7)   Software Design: it's about verifying whether the SDLC is in support of designing preliminary software models (conceptual, work flow etc.) before jumping into the developmental stage.

(8)   Software Architecture Design: it's about implementing software architecture designs e.g. Component-Based Architecture, Client/Server, Domain Driven Design. Layered Architecture, Tier based, Procedural, Object-Oriented (OO), and Service-Oriented Architecture (SOA).

(9)   In Depth Software Design Modelling: it's about verifying whether the chosen SDLC enforces in depth software design modelling with the inclusion of the mandatory software developmental designs, e.g. Use Case, Activity, Class, Collaboration, System Sequence, Components, Data Flow, Deployment.

(10)   Reusable Designing: it's about verifying whether the SDLC is promoting the use of product line architecture designing. Moreover it's also about implementing systematic software reuse processes (Application system reuse, Component reuse, Object and Function reuse) to achieve better software more quickly and at a lower cost, with increased dependability, reduced process risk, accelerated development, and lack of tool support.

(11)   Developmental Plan: it's about checking whether there was a documented, comprehensive plan to be followed and used during software development. It mainly includes: Phase Plan (work breakdown structure of Inception Phase, Elaboration Phase, Construction Phase and Transition Phase), Iteration Plans, Project Monitoring and Control (Schedule, Quality Control), and Risk Management Plan.

(12)   Tools and Technology Selection and Analysis: it's about checking whether the chosen SDLC is promoting the selection of proper tools and technologies for different kind of projects (based on the scientific requirements and available resources).

(13)   Graphical User Interface Design: it's about to find if the concerned SDLC is enforcing the HCI design patterns, principles and guide line in designing a user friendly graphical interface of the application (Ahmed et al., 2008).

(14)   Preprocessed Source Code Writing: it's about verifying whether there are some preprocessed source writing patterns to be followed (Ahmed, 2010).

(15)   Integrated Programming: extending reusable designing, it's about checking whether the SDLC is enforcing the modular integration or not (supporting syntax driven editing different languages within one session, multi-user access, reusable components etc.).

(16)   Software Testing: it's about verifying whether the SDLC is promoting the use of initial software testing (black box) processes to gather information about software bugs and the quality of the application, with an objective and independent view.

(17)   In Depth Software Testing: it's about verifying whether the SDLC is promoting the use of extreme software testing to find out the detailed information about defects, failures, ripples in the software application. It includes black, white and gray box testing with the use of software testing norms (including test case against written use cases).

(18)   Customer's Evaluation: it's about evaluating the customer's feedback and improve the software accordingly, before obtaining the final product.

(19)   Deployment Procedures: it's about implementing standard deployment procedures at the user end (software configuration and installation) e.g. to provide image based help, promote tested production, give upgraded policy sequence to available revision.

(20)   Maintenance: it's about verifying whether the SDLC is promoting the software maintenance.

(21)   Software Re-Engineering: it's also called as the reverse engineering and it's about verifying the implementation of the processes for examining a system at a higher level of abstraction (Chikofsky and Cross, 1990).

(22)   Cyclic or Repetition: it's about checking whether SDLC is a recurring, which means it promotes never ending software development (or at least some more iterations, till the final version is achieved).

(23)   Easy to learn and Use: it verifies whether the SDLC promotes the use of HCI guidelines in making the software application easy to learn, adopt and use.

(24)   User Training: it's the last step to check whether the SDLC promotes the user training of programmed application (e.g. with the provision of some user manuals, presentations).
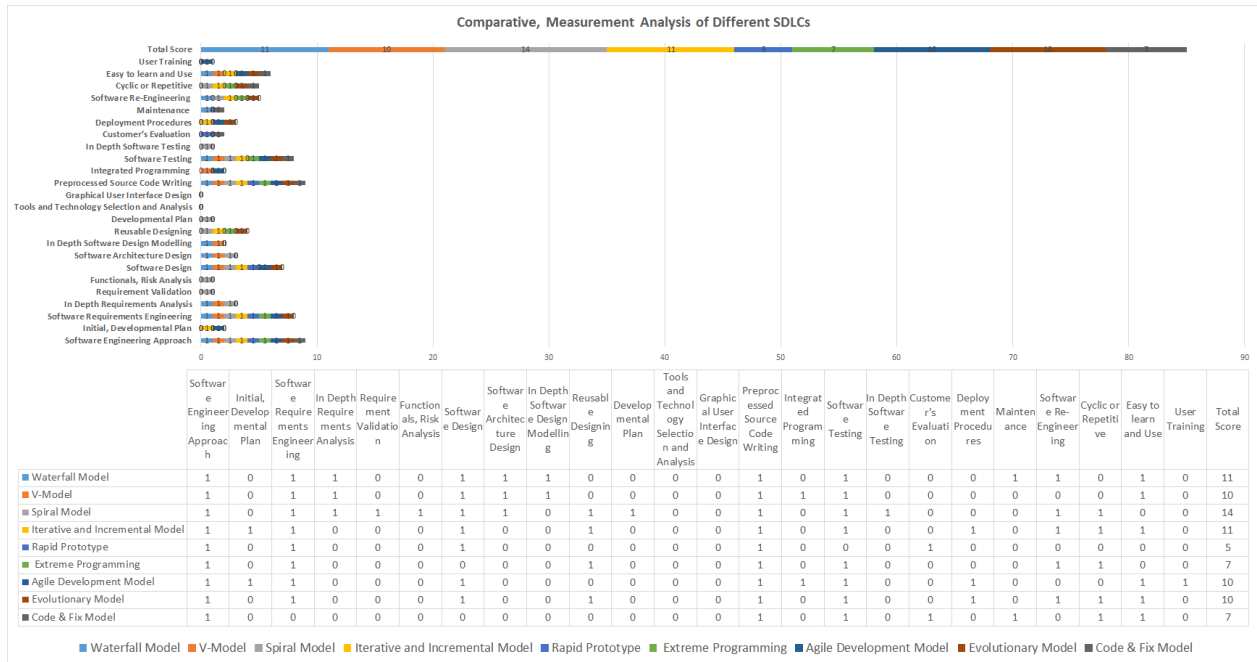
Figure 23: Comparative, measurement analysis of different SDLCs

	Software Engineering Approach	Initial, Developmental Plan	Software Requirements Engineering	In Depth Requirements Analysis	Requirement Validation	Functionals, Risk Analysis	Software Design	Software Architecture Design	In Depth Software Design Modelling	Reusable Designing	Developmental Plan	Tools and Technology Selection and Analysis	Graphical User Interface Design	Preprocessed Source Code Writing	Integrated Programming	Software Testing	In Depth Software Testing	Customer's Evaluation	Deployment Procedures	Maintenance	Software Re-Engineering	Cyclic or Repetitive	Easy to learn and Use	User Training	Total Score
Waterfall Model	1	0	1	1	0	0	1	1	1	0	0	0	0	1	0	1	0	0	0	1	1	0	1	0	11
V-Model	1	0	1	1	0	0	1	1	1	0	0	0	0	1	1	1	0	0	0	0	0	0	1	0	10
Spiral Model	1	0	1	1	1	1	1	1	0	1	1	0	0	1	0	1	1	0	0	0	1	1	0	0	14
Iterative and Incremental Model	1	1	1	0	0	0	1	0	0	1	0	0	0	1	0	1	0	0	1	0	1	1	1	0	11
Rapid Prototype	1	0	1	0	0	0	1	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	5
Extreme Programming	1	0	1	0	0	0	0	0	0	1	0	0	0	1	0	1	0	0	0	1	1	0	0	0	7
Agile Development Model	1	1	1	0	0	0	1	0	0	0	0	0	0	1	1	1	0	0	1	0	0	0	1	1	10
Evolutionary Model	1	0	1	0	0	0	1	0	0	1	0	0	0	1	0	1	0	0	1	0	1	1	1	0	10
Code & Fix Model	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	1	0	1	0	1	1	0	7

Figure 23 shows the results in tabular and graphical chart formats. All (twenty four) measures were applied (kind of majority voting model to support the asset) to the chosen (nine) SDLCs, and based on the application of each measure, 0 or 1 score was attributed to the SDLC. Total scores show that the most appropriate SDLC towards scientific software solution development could be the Spiral Model (score 14) and the least suitable is the Rapid Prototype Model (score 5). However Waterfall Model (score 11), Iterative and Incremental Mode (score 10), Agile Development Model (score 10) and Evolutionary Mode (score 10) could also be well considered.

Comparative analysis concludes with the information that none of them is especially proposed towards the scientific software solution development in academia. Moreover two important missing elements in all of the SDLCs are selection of the appropriate developmental technologies and user

friendly graphical interface development, based on the scientists informatics background, scientific domain and available technical resources.

The applied (voting) method may not be very accurate, as it implements the same criterion with the similar level of importance to all SDLCs. We know that each SDLC has been proposed for a special and given criteria. The main reason for putting all in one category is to find out the most reliable SDLCs, to be suggested and followed in interdisciplinary scientific software solution development, especially in unpredictable and variable conditions.

Considering the fact that this thesis belongs to the field of bioinformatics, and the conducted research and development work belongs to the fields of life sciences but it is strongly influenced from the field of informatics, mathematics and statistics. Due to the heavy involvement of informatics in the science, the software design and its engineering becomes essential for bioinformatic software impact. Giving high importance to the modelling of scientific software solutions, in this thesis we have tried to target the key developmental points: intuitive, graphical user interface design, stable methodical implementation and comprehensive output presentation. Furthermore, we tried to address the following key points: (1) Differences and different challenges required to change from traditional to scientific software engineering, (2) Scientific software solution development needs feedback and control loops following basic engineering principles for implementation, (3) The Butterfly software design and approach helps in developing and implementing a comprehensive scientific software solution.

Most exciting part of this research was to work on and propose a new method for the betterment of scientific software solution developed in bioinformatics. Then, validate the approach by first comparing with (old and new) other approaches, and then implement some new bioinformatics software solutions to practically prove the strength. Moreover, additional work was about to bring

new proposition and implemented solutions to the public forums (by making good publications). The proposed approach and almost all implemented bioinformatic tools (using it) are now well published or in the process of publication with some well reputed scientific and computer science journals including Database by Oxford University Press, Bioinformatics by Oxford University Press, BMC Bioinformatics by BioMed Central, Frontiers in Neuroinformatics, Recent Patents in Computer Science and Faculty 1000 Research.

5.2   Butterfly model

When it comes to the scientific software solution development, ignoring all SDLC phases, most adopted way of development is the direct programming and debugging, without any formal requirements analysis, proper design modelling, suitable choice of developmental technologies, testing & quality assurance and expected deployment procedures (Ahmed et al., 2014). It might work in case of small scripts but won't be helpful in comprehensive prototype development. Furthermore, achieved solutions in quick and dirty ways cannot be preserved longer, reused and enhanced by other community members.

The need of the time is not a completely new way or unique paradigm to be adopted in scientific academia but a simple, accumulative approach which can combine the different benefiting aspects of the software development models together with the missing elements, with the abridged way of acceptance.

Butterfly is a science oriented, software development model which can be helpful by generalizing the use of major developmental aspects correlating to the important scientific needs of the target system. The *Butterfly* workflow design accentuates experience from previous software developments including a number of larger efforts. Most of these are team efforts which simply

have come close to the *Butterfly* paradigm, but by chance and pressure, not by explicitly following a scientific approach. With rapid development of new software applications, the need to formalize the software solution development principles increases to ensure that all scientific applications follow the standard scientific paradigms.
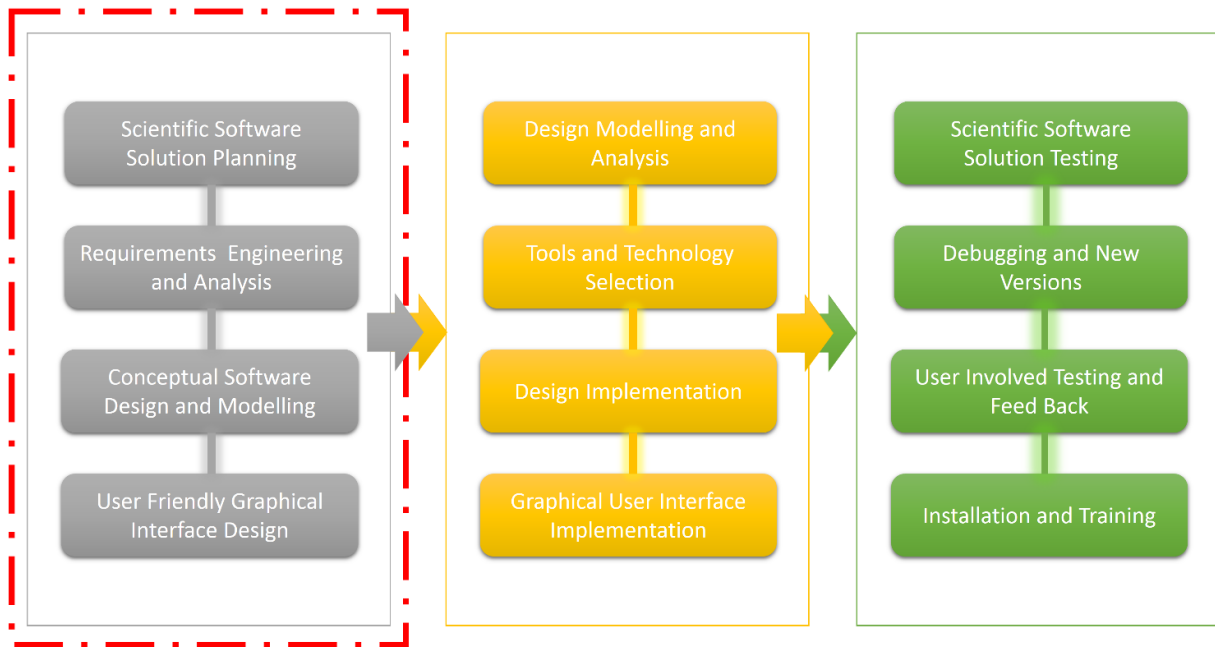
Figure 24: The *Butterfly* three layer model

It consists of four phases: software engineering, scientific methodology, human computer interaction and scientific application. Correlating the software engineering with the scientific methodology phase, the requirement analysis is performed to get maximum possible functional information about the system, producing desired output. Design modelling required to draw the main logic of the system. Developmental technologies must be chosen according to the system and the nature of the data. Development should occur according to the defined data analysis processes.

Systems should be tested using real time data or in real time environment. Deployed (final) system should be capable of producing the required results.

To implement the *Butterfly* model, it is divided in three layered architecture (Figure 24), going from abstract planning (gray) to designers and developers (yellow) to implementation and user (green). Scientific software solution planning (gray layer) is the first step towards a new scientific application development. It requires good knowledge of the field (e.g. biochemistry, neurobiology, genetics, metabolomics, proteomics etc.) as well as project related information (e.g. what could be the end product, input to the system, expected output from the system, methodology, ideas, user opinions etc. It is the most important layer (red marked; Figure 24) because if the requirements of the targeted system are unclear then the end product will not result in the solutions, what the end users were looking for.

The next important phase is to perform requirements engineering and analysis. The third phase is the conceptual software design and modelling. Before moving ahead, first go for some abstract designs based on functional requirements and discuss these in your team. The last phase is the design of a user friendly graphical interface following HCI design patterns, principles and guidelines. Furthermore different methods (patents, published and well used) have been proposed towards enhanced user interface (e.g. software, hardware, ubiquitous, mobile etc.) implementation for better human machine interactions.

The yellow layer involves the designers and developers. It consists of four steps: design and modelling and analysis, tools and technology selection, design implementation and graphical user interface implementation. It main requires to model the finalized functional requirements and implement them using most suitable tools and technologies. Furthermore using the concepts of HCI, design graphical user interface, establishing the efficient communication protocols between

humans and machines. Additionally, if the target scientific software solution has the focus towards database implementation, then it is required to properly model the database schema (entity relationship model) by reducing the levels of data redundancy and dependency, using data normalization forms (William, 1983).

The last, green layer includes the scientific software solution's testing, debugging and creation of new versions. Furthermore it asks for the users' involved testing and feedback, and finally installation and training.
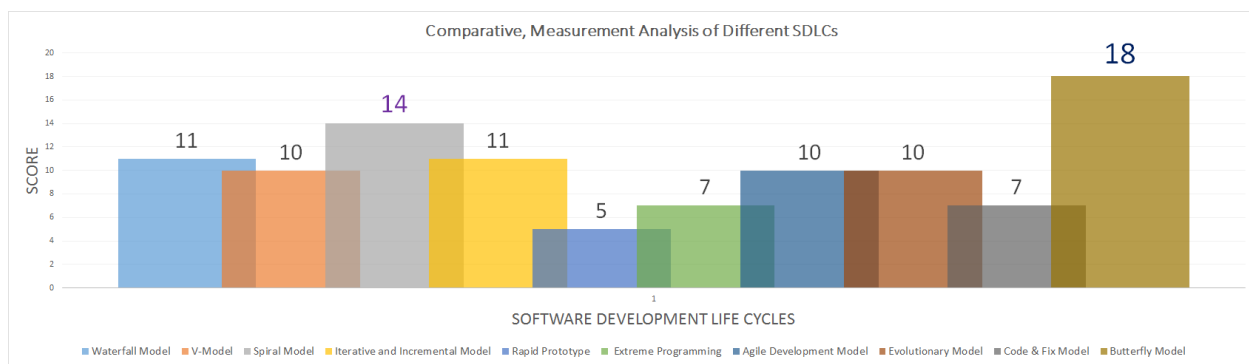
Figure 25: Comparative, measurement analysis of different SDLCs with the *Butterfly* model

Ignoring all SDLC phases and neglecting all expected problems (before, during and after development), the easiest and the most frequently adopted way of scientific software solution development, is direct programming without any formal requirement analysis, proper design and testing. This may work in the case of small and urgent software solutions, but won't help in comprehensive, reliable, flexible, maintainable, manageable and long-lived scientific software solution development. For these reasons the full implementation of software engineering principles and SDLC is essential. In this paper, many examples of the published and unpublished previous

academic scientific software solutions can be presented without targeting any particular application. The focus is on new approaches to the development of scientific software solutions.

Of course it is not the user´s duty to provide all information required for software development, but it is the responsibility of the developer to provide this systematically. Here the *Butterfly* paradigm is very helpful. Similarly, we do not intend to criticize academic scientific software developers who use other current and popular approaches to software development, including very pragmatic approaches. The *Butterfly* paradigm highlights all the necessary steps in scientific software solution development with a high dividend after additional investment in time and effort involved in obtaining sustainable, user-friendly software solutions, which maintain the potential for further development, interoperability and high quality.

As this is a general and open comparison, depending on the nature of the scientific software application, one can further analyse and pick what suits best. Furthermore, we considered only the typical effort and strengths for each of these software development paradigms. A meticulous developer can of course take special care and spend more time on any of the features not typically covered by the software paradigm he follows, and turn the "no" for this feature into a "yes" simply by this additional effort during SDLC (for instance, regarding agile programming – for that matter, extreme programming can also be considered as a type of agile development). The goal for our "*Butterfly*" paradigm is a SDLC paradigm that fulfils all of the features regarding life cycle management of the resulting software.

Earlier discussed 24 measures are also applied to the *Butterfly* model, and it is compared with earlier mentioned nine SDLCs as well. The Figure 25 shows the comparative analysis in both tabular and graphical chart formats. The score of *Butterfly* model is 18, which is more than any

other SDLC, and considering this fact, one can adopt and apply the *Butterfly* model for the scientific software solution development.

5.2.1 Examples using Butterfly model

Adopting the concepts of the Butterfly model, some new scientific software applications have already been proposed, designed, implemented, tested and are in use.

These applications are:

- LS-MIDA (Ahmed et al, 2013a), (Ahmed et al, 2012a); earlier discussed in chapter 4 (4.2), software application towards GC-MS data analysis, manipulation and management.

- Isotopo (Ahmed et al., 2012b) (Ahmed, 2014a) (Ahmed et al., 2014b); earlier discussed in chapter 4 (4.3), software application towards GC-MS data analysis, manipulation and management.

- DroLIGHT (Ahmed et al., 2013b) (Ahmed and Helfrich-Förster, 2013c) (Ahmed and Helfrich-Förster, 2013d); scientific, computational solution towards neurobiology and photobiology, capable of controlling an LED-based lighting system to automatically simulate various light-dark cycles with high quality and fidelity (Figure 26).

- Lipid-Pro; bioinformatics tool towards LC-MS data analysis, visualization management and Lipid Annotation (Figure 27).

- Ant-App-Database and Dataplus (Ahmed, 2014c); combination of smart phone, tablet and desktop application towards the neural and behavioral research on insects. Especially designed and implemented for the efficient data management during the experimentation on the desert Ants and approximate solar estimations (Figure 28).
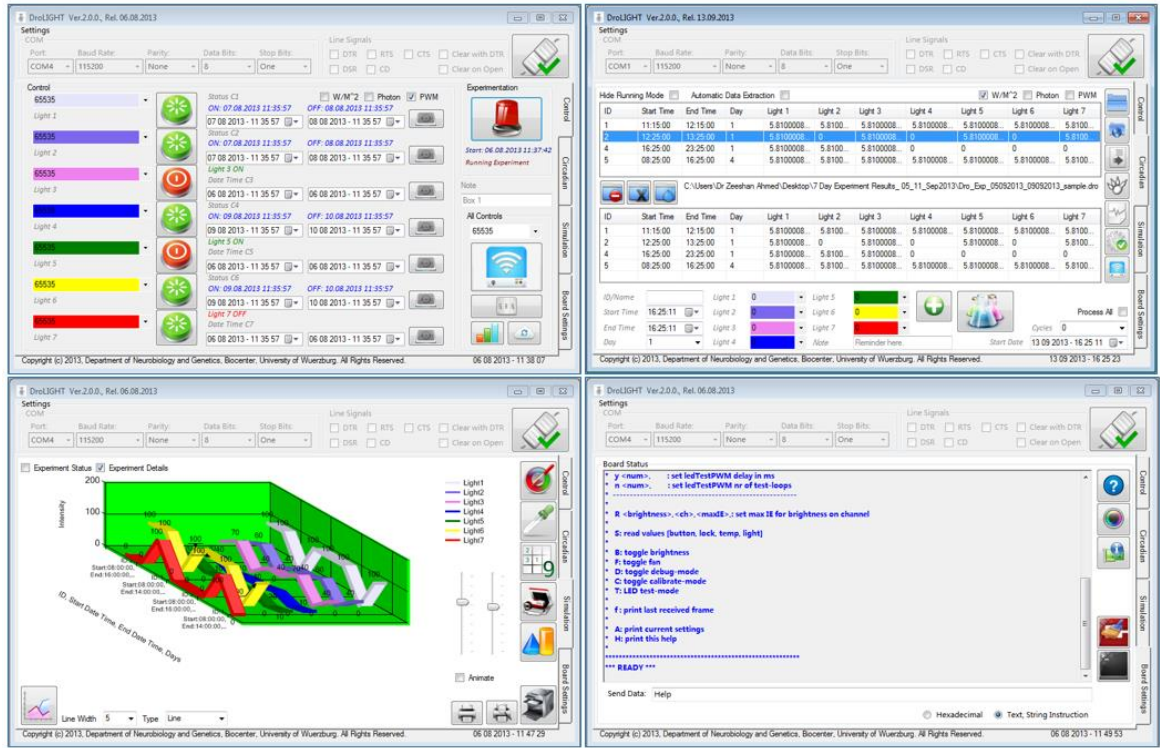
Figure 26: DroLIGHT; control, circadian, visualization and board control
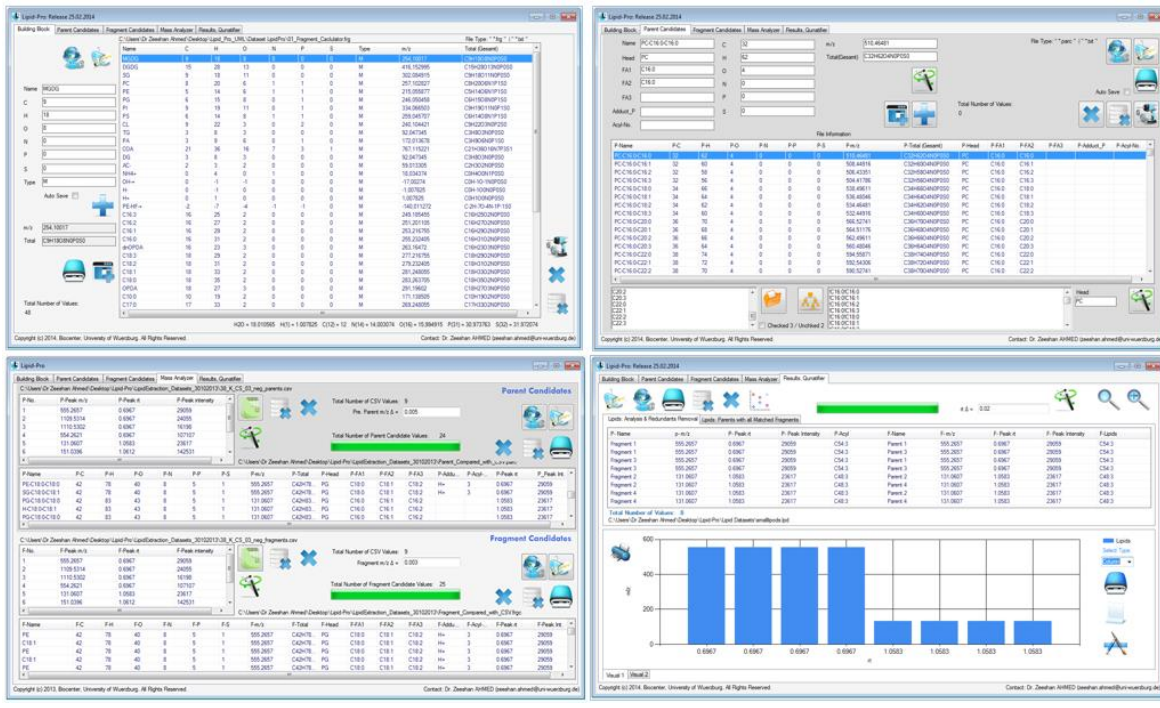
Figure 27: Lipid-Pro; building block, fragments, mass analyzer and lipids
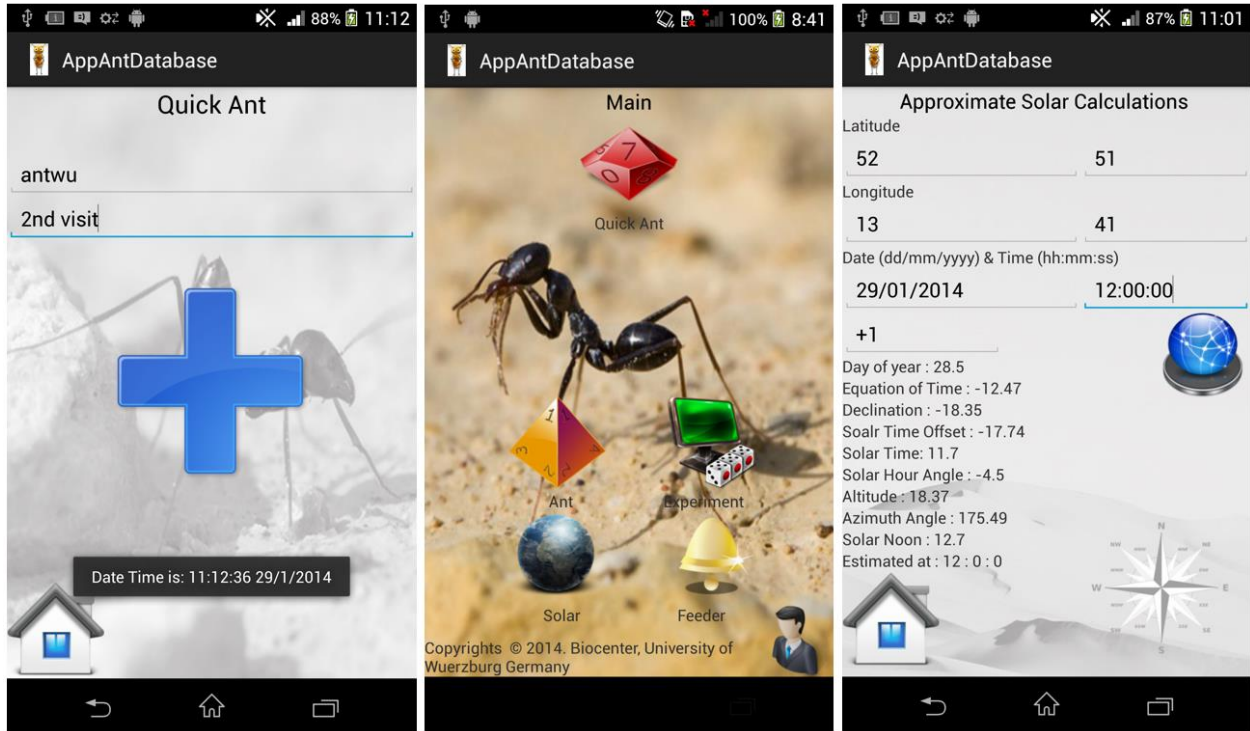
Figure 28: AntAppDB: main and solar calculation

5.3 Cross Linking Network in Platelet Apoptosis

Efficient platelet activation requires a coordinated signaling function of integrins, growth factors (IGF-1 and IGF-1R) and P2Y12 by platelet agonist ADP. Significant advances have been made in the last few years towards understanding the mechanism of IGF1R signaling, and revealing a significant degree of cross-talk with integrins that provides a good example of the intimate relationship between integrins and GFR-mediated signaling (Eliceiri, 2001).

The antiapoptotic function of IGF1-R is mediated via the activation of phosphatidylinositol 3-kinase (PI3K), Akt/ PKB (Protein Kinase B), and the phosphorylation and inactivation of Bad. Akt phosphorylates Bad which is unable to neutralize the antiapoptotic effect of either Bcl-xL or Bcl-

2 (Hirai and Wang, 2001). It also causes cell proliferation by activating the Mapk/Erk pathway via Ras-Raf activation (Aleman, 2005).

Our study reveals that platelet aggregation is brought by integrin mediated growth factor signaling but also involves activation of αIIbβ3 integrin by the secondary mediator, ADP thus presenting P2Y12 signaling as a possible association to this crosslinking pathway (Figure 10b). P2Y12 can stimulate αIIbβ3 receptor activation independently at high concentration of ADP through a PI3K dependent signaling pathway (Kauffenstein et al., 2001). It has been shown that activation of the P2Y12 receptor leads to Akt activation through IGF-1 receptor cross-talk (Kim et al., 2007).

The experimental study by Landry et al., 2009 leaves some vacant connections and raises many questions on the potentially new mechanism for platelet control of function via miRNA expression. In particular Landry et al. only observed an unexpected high number of miRNAs in the platelet but did not do in their study a systematic analysis for all possible targets. Instead, they considered one potentially important target, the P2Y12 receptor and showed that there is a connection to miRNA regulation. However, our effort tries to fill out this shortcoming of the above study by a general bioinformatics screen on all mRNAs that can be targeted by the miRNAs of the platelet. As each of the potential targets has to be screened individually and manually, we considered only 30 high expressed miRNAs and only the best targets with probability >0.99 were selected for further study. Predicting potential targets with TargetScan, it was possible to derive a complete list of best hit targetted by highly expressed miRNAs in the platelet.

This is an important extension of the previous analysis and helps to get a clear picture of platelet functioning involving miRNA regulation. Interestingly, the top regulated mRNAs encode exactly key players involved in platelet activation. We think the following scenario happens:

Activated platelets change shape and its proteins undergo mobilization with transport of many proteins like PI3K to the cell membrane. Down signaling through PI3K results in activation of Akt which delays platelet apoptosis by inhibiting Bad until platelets are fully activated. Simultaneously the activation of the Raf-Ras pathway causes platelet migration. These results increase the evidence supporting the possible role of hsa-let-7d, hsa-let-7f, hsa-let-7g, hsa-miR-98, hsa-miR-130a, and hsa-miR-130b, in targeting and degradation of mRNA for 3 platelet proteins IGF-1, IGF-1R, and TGFβ-R1 and hence mediating delayed apoptosis on the transcription level. This can only work if these protein targets have some turnover in the platelets.

5.4   Integrin – Growth Factors (GF) Cross Talk

The IGF system has a broad and potent ability to prevent apoptosis both during normal development and during stress or disease, in a wide array of cell types as diverse as osteoblasts, melanoma cells, cardiac myoblasts, neuronal cells, and epithelium. IGF-1R (IGF-1 Receptor), activated by its ligand IGF-1, plays a pivotal role in tissue homeostasis, regulating cell proliferation, differentiation and migration during development and in the adults. Activation of IGF-1R is a particularly important survival-promoting signal (Aleman, 2005).

The main signaling pathway for IGF1R-mediated protection from apoptosis rests on the activation of PI3K and Akt/PKB. The IGF-1R also activates alternative pathways for protection from apoptosis and in some cases, involved in cell proliferation and differentiation. One of these pathways leads to the activation of MAPKs, while a third pathway results in the calcium mobilization inside the cell. All these pathways, however, result in maintenance of cell survival by antagonizing the processes and proteins involved in apoptosis. The multiplicity of signaling

pathways used by the IGF-1R may explain why this receptor has such a powerful and widespread antiapoptotic activity (Shelton et al., 2004).

The best-defined pathways by which IGF-1R signaling can prevent apoptosis is mediated by PI3K. Activated platelets change shape and its proteins undergo mobilization with transport of many proteins like PI3K to the cell membrane. PI3K activates Akt (Jackson et al., 2004). Activated Akt phosphorylates and inactivates several proteins that are involved in apoptosis. A primary target is the Bad. In its non-phosphorylated state, Bad locates at the mitochondrial membrane where it interacts with Bcl2 (B-Cell CLL/Lymphoma-2) and prevents it from performing its anti-apoptotic functions. Akt can also prevent the initiation of the caspase cascade through phosphorylation and inactivation of caspase 9. In addition to the inhibition of pro-apoptotic transcription factors, the activity of Akt also increases the levels of anti-apoptotic proteins including Bcl2 and Bcl-X and several extracellular matrix adhesion molecules (Vincent and Feldman, 2002).

Several studies have demonstrated that the Bcl-2 family proteins are also expressed in platelets. The Bcl-2 family of proteins are the key regulator of apoptosis in nucleated cells and consists of both antiapoptotic (Bcl-XL, Bcl-2, Bcl-w, A1, Mcl-1) (Li et. al. 2000) (Bertino et. al. 2003) (Clarke et. al. 2003) and proapoptotic (Bak, Bax, Bid, Bim, Bad, Bik, Bmf, Noxa, Puma) members (Gelinas and White 2005). Bcl-2 may play a minor role in platelet survival as the inhibitor of anti apoptotic factors initiates the apoptotic cascade in isolated platelets. In comparison, Bcl-XL may play a larger role in platelet survival (Zhang et. al. 2007) (Vanags et. al. 1997). Studies examining apoptotic markers in stored platelets indicates that Bcl-XL levels rapidly decrease as platelets age and lose viability in vitro, further supporting a role for Bcl-XL (Bertino et. al. 2003) (Vogler et. al. 2011).

The role of caspases in platelet production and function, and function of apoptotic caspases in megakaryocytes was addressed by White et. al. 2012. Active caspases are found in megakaryocytes

lacking Bcl-xL, or megakaryocytes treated with a potent antagonist of Bcl-2, Bcl-X(L) but their contribution to the cell-death process was not determined. It is also possible that megakaryocytes possess a classic apoptotic caspase cascade, with caspase-9 at its apex. This was established by apoptotic caspases, including caspase-9, facilitating platelet shedding by megakaryocytes. They are known to be activated during platelet apoptosis, and have also been implicated in platelet hemostatic responses. Both megakaryocytes and platelets possess a functional apoptotic caspase cascade downstream of Bcl-2 family-mediated damage and caspase-9 is the initiator caspase. Platelets with active caspase 9 are fully capable of physiologic hemostatic responses and functional regulation of adhesive integrins in response to agonists (White et al. 2012).

Although the association between platelets activation and apoptosis is not well defined (Gyulkhandanyan et. al. 2013) but caspase-9 pathway is required for the efficient death of MKs and platelets rather than for production. Furthermore, some platelet agonists can stimulate both caspase-9 and -3 activation in human platelets (Lin et. al.2009).

The binding of IGF-1 or IGF-1R activates the receptor's intrinsic tyrosine kinase activities, which results in the phosphorylation of the IRSs (Insulin Receptor Substrates) (Tseng et al., 2002). On phosphorylation, IRSs interact with signaling molecules, including Akt, Ras/Raf, and Rac. Activation of the PI3K and Akt pathway causes release of cytochrome c, thereby inhibiting caspase-3 activation and apoptosis. The Ras/Raf pathway is critical for proliferative responses, whereas activation of Rac is important for cell migration (Delafontaine et al., 2004).

Upon IGF-1R autophosphorylation the protein Shc (SH2 Containing Protein) is recruited to the receptor and becomes phosphorylated on tyrosine residues. Activated Shc then binds the adaptor Grb2 (Growth Factor Receptor Bound Protein-2), recruiting the Sos in an IRS-independent manner. This complex then activates Ras and initiates sequential phosphorylation cascades involving

serine/threonine kinase Raf, MEK1/2 (MAP Kinase Kinases), and ERK1/2 (Extracellular Signal Regulated Kinases). This pathway of IGF-1R signaling has been most closely associated with cell differentiation and migration, but in some cases also can regulate the machinery of apoptosis. An endpoint of the MAPK pathway is modification of transcription factor activity, such as activation of Elk transcription factors. Similar to the Akt pathway, the downstream target of ERKs that prevents apoptosis is Bad.

The Ras-->Raf-->ERK1/2 pathway is also activated by the tyrosine phosphorylation of IRSs, with the resulting formation of the IRS-Grb2-Sos complex that activates Ras, which in turn binds to and activates Raf, subsequently phosphorylating and activating MEKs, followed by ERKs. Phosphorylated ERKs, in turn, transmit signals to the nucleus, with resulting mitogenic response: progression of the cell cycle and cell proliferation (Mora et al., 2005; Zheng et al., 2002).

IGF binding proteins (IGFBPs) have also an important role because of their capacity to bind to IGF-I with high affinity. In addition to IGFBPs, extracellular matrix (ECM) proteins have been shown to play a role in modulating cellular responses to IGF-1. One of such ECM protein are integrins (Clemmons and Maile, 2006).

Shc protein binds to either IGF-IR or distinct integrin receptors, suggesting a collaboration between integrins and growth factors in a platelet signalling pathway. Shc then associates with Grb2 and with phospho-ERK, leading to activation of the Ras/MAP kinase signalling pathway. Shc has its effect on downstream signal transduction by interacting with the distinct cell surface integrins, but not with all.

Ligand occupancy of the integrin influences the recruitment of the phosphatase SHP-2 to the IGF-1R receptor and enhances IGF-1R mediated intracellular signaling (Maile and Clemmons, 2002). SHP-2 is a tyrosine phosphatase whose recruitment to signaling molecules is stimulated by growth

112

factors including IGF-I. When integrin ligand occupancy is inhibited, there is no recruitment of SHP-2 to the integrin receptor and its transfer to downstream signaling molecules is blocked. Subsequently, SHP-2 is transferred to another transmembrane protein, SHPS-1. The transfer of SHP-2 to SHPS-1 is also required for recruitment of Shc to SHPS-1 (Delafontaine et al., 2004). Integrin αIIbβ3 which is the highly expressed platelet integrin becomes tyrosine phosphorylated during platelet aggregation, causing Shc and myosin to interact with the beta-integrin cytoplasmic domain showing that the signaling adapter protein Shc, has an important role in platelet stimulation (Phillips et al., 2001).

Therefore, it is likely that Shc may be a common point in these two signal-transduction pathways. The synergistic interaction between integrins and growth factor receptors phosphorylates MAP kinases, which control a variety of cell functions, and that this collaboration is only possible if the integrins are already aggregated and occupied by a ligand. Thus the integrins and growth factor receptors synergize to enhance Ras-MAPK activation and also promote cell migration. Some integrins generate formation of filopodia and some lamellipodia in the presence of αIIbβ3 integrin (Watson et al., 2005).

The precise mechanism by which integrins transduce signals has not been fully elucidated. There is abundant evidence that PI 3-kinases play a major role in integrin αIIbβ3 outside in signalling relevant to several key functional responses, including primary platelet adhesion, spreading and aggregation. However, the mechanism by which integrins and specific growth factors, such as IGF-I, regulate platelet activation at present is not fully understood (Jackson et al., 2004).

PI3K is also associated with integrin-associated focal adhesion complexes and provides protective signal acting through Akt/ PKB which blocks entry into apoptosis. So PI3K/Akt pathway stimulates cell proliferation and suppresses apoptosis (Martelli et al., 2007). The kinase domain of Integrin

linking kinase (ILK) also binds to the cytoplasmic tails of beta-integrins and the kinase substrate Akt and Pxn (Paxillin).

The role of TGF-β in this cross linking pathway is not that much clear. TGF-β is activated from a latent state. Latent TGF-β activation is catalyzed by specific integrin on epithelial cell membranes. The activated TGF-βRI then phosphorylates its downstream targets, the members of the SMAD (Sma and Mad Related Family) family of signal transducers. TGF-β also induces other non-SMAD signaling pathways, which include activation of several MKKs (MAP kinase Kinase) and MEKs (MAPK/ERK Kinase) pathways (JNK/SPAK, p38, and ERK1/2) through upstream mediators RhoA, Ras, TAK1 (TGF-β Activated Kinase) (Derynck et al., 2001) (Wakefield and Roberts, 2002).

Because of its critical role in cell fate determination, TGF-β signaling is subject to many levels of positive and negative regulation, targeting both the receptors and the intracellular mediators. TGF-β also induces a variety of complex cellular responses, depending on the cell type, most notably growth arrest in late G1 involving PI3K and PP2A (Protein Phosphatase-2A), changes in differentiation programs, and apoptosis (Derynck et al., 2001). Other growth factors also regulate TGF-β mediated signaling through Grb2 - Sos activation of Ras (Moustakas et al., 2002).

Interestingly, many of the signaling pathways and effectors, which are activated by integrin interaction, are also activated after growth factor stimulation. This suggests that integrin and growth factor mediated cellular responses may synergize and may function to coordinate biochemical responses in multiple cell types (Eliceiri, 2001). The initial signals promoting inside-out signaling can be triggered by a diverse array of adhesive substrates and soluble stimuli that activate platelets through either tyrosine kinase linked adhesion receptors like growth factor

receptors or G-protein-coupled agonist receptors respectively like P2Y receptors (Jackson et al., 2004).

ADP-induced platelet aggregation requires coactivation of both the Gq-coupled P2Y1 receptor and the Gi-coupled P2Y12 receptor. The latter is responsible for a significant proportion of activation of Akt, and ADP-induced Akt phosphorylation requires Gi signalling. In contrast the P2Y1 receptor is essential for intracellular calcium mobilization and platelet shape change, whereas the P2Y12 receptor leads to inhibition of adenylyl cyclase. Though selective P2Y1 receptor or P2Y12 receptor activation alone has been shown to be unable to cause platelet aggregation it has also been found that IGF-1 can potentiate platelet aggregation induced by a number of agonists, including ADP. Platelets release IGF-1 from alpha granules upon activation. Although IGF-1 alone fails to induce platelet aggregation, IGF-1 potentiates platelet aggregation by synergizing with Gi but not Gq signaling through the PI3-K pathway. This cross talk implicates that IGF-1 plays an important role in platelets because IGF-1 is released from the platelet immediately after injury directly into the tissue, where additive and synergistic actions with other growth factors in alpha granules, including PDGF, EGF, and TGF beta can promote tissue repair (Kim et al., 2007).

The anucleate status of platelets has stereotyped it as a cell without ability to synthesize. It is now clear that platelets can translate biologically relevant proteins that are regulated via gene expression programs at the translational level. This process does not require a nucleus; instead, it uses mRNAs and other translational factors that are passed on from megakaryocytes as they generate platelets during thrombopoiesis (Weyrich et. al. 2004). The role of miRNAs in megakaryocytopoiesis was first reported by Garzon et al 2006. They also highlighted the down regulation of several miRNAs like miR-10a, miR- 126, miR-106, miR-10b, miR-17, and miR-20 during megakaryocytopoiesis.

The first evidence of the existence of miRNAs in human platelets was provided by Burchova et. al. 2008 with their study on patients with polycythemia vera.

This investigation revealed a strong synergistic cooperation between integrin and growth factor receptor mediated signaling, which maintains optimized signal transduction in platelet.

5.5 microRNA expression in murine megakaryocytes

The process of thrombopoiesis culminates in the release of platelets. The regulatory architecture including the modulatory role of microRNAs, which underlies megakaryocytic maturation and platelet formation, is incompletely understood. (Emmrich et al., 2012). Megakaryocytes invest platelets with requisite translational machinery that includes ribosomes, initiation and termination factors, microRNAs (miRNAs), and template messenger RNAs (mRNAs) (Gatsiou et al., 2012).

We are investigating this process in mice and men because interspecies comparison could provide useful information on the evolution of miRNA functions. Moreover, changes in megakaryocyte miRNA expression that are conserved between species may reflect biological significance. Anucleate platelets are only present in mammals among all animals. Murine systems represent an excellent model to study megakaryopoiesis because progenitors are responsive to genetic manipulation both in vivo and in vitro (Opalinska et al., 2010). Another reason for utilizing this approach was while comprehensive profiling of miRNA in human megakaryocyte precursors has not been reported, Petriv et al. 2010, have published an extensive characterization of multiple murine hematopoietic precursors. miRNA regulation has been validated in murine megakaryocytes for many cellular functions and pathways but they have not been investigated for apoptosis. With the recent arrival of Dr. Schulze and establishment of his laboratory, we are in the phase of

preparation of oligonucleotides and with the good cell culture and efficient transfection system we hope to proceed with the experimentation.

Our group has investigated associations between miRNAs and platelet reactivity to better understand megakaryocyte/platelet gene expression. Our findings will extend the repertoire of miRNA targets that potentially regulate platelet production and provide a useful dataset for future studies on this topic.

5.6   Role of MicroRNAs in AAA

miRNAs repress gene translation upon binding to mRNAs and thereby influence the biological processes. Changes in miRNA expression profiles are noted during progression of heart disease and vascular disorders (Nurden, 2011) (Denis et al., 2005). Liu et al, 2010 profiled miRNA expression in a rat model of abdominal aortic aneurysm, using a miRNA microarray, and identified 15 dysregulated miRNAs with putative targets in multiple signaling pathways. Expression of miRNAs in abdominal aortic aneurysm has been established for many miRNAs with miR-133b, miR-133a, miR-204, miR-331-3p, and miR-30c-2, the five miRNAs with confirmed downregulated expression between AAA and positive regulation of apoptosis with smooth muscle cell apoptosis is a characteristic histological feature of aneurysmal aortic wall in humans (Pahl et. al. 2012). miR-21 and miR-29b have also been identified as potential therapeutic targets in an animal model of aortic aneurysms (Maegdefessel et. al. 2012a) (Maegdefessel et. al. 2012b) (Adam et. al. 2013).

Moreover, a single miRNA is capable of targeting multiple mRNAs, and a single mRNA may contain multiple miRNA binding sites (Bartel, 2009; Krek et al., 2005) Together, this suggests that

a common set of miRNAs can fine-tune the protein abundance of a cassette of specific genes that together influence specific cellular functions (Jones et al., 2011).

It was not until 2006 that the first cardiac miRNA-profiling study appeared, linking dysregulation of many different miRNAs to cardiac remodelling in both mice and humans (van Rooij et al., 2006). A detailed understanding of the biology of miRNAs and optimization of gene delivery vectors will determine the future of therapeutic miRNAs and their application in treating cardiovascular diseases such as AAA. (Maegdefessel et al., 2012b). The identification of molecular pathways that lead to the onset and progression of cardiovascular disorders in vivo is a prerequisite for uncovering novel targets for therapeutic intervention. No miRNA therapy developed so far to treat cardiovascular ailments has entered human clinical trials; nevertheless, the use of miRNA-based therapies in preclinical animal models provides an important proof of concept for the clinical use of such therapies in the future.

Mediators of aortic damage include angiotensin II, leukotriene- (LT)4, prostaglandin- (PG)E2, interleukins, tumor necrosis factor, tissue plasminogen activator, c-Jun N-terminal Kinase, NF-κB, Rho kinases, osteoprotegerin and chymases. They result is degradation of aortic wall proteins, extracellular matrix and apoptosis of vascular smooth muscle cells (Maegdefessel et al., 2012a).

As it is becoming increasingly apparent, miRNAs are important determinants of disease within the cardiovascular system. Clear roles for altered miRNA expression have been implicated in the regulation of smooth muscle cell phenotypes, angiogenesis, atherosclerosis, restenosis, and other vascular injury responses (Small et al., 2010; Qin and Zhang, 2011). Moreover, recent work has identified that changes in miRNA expression may contribute to the pathogenesis of aortic aneurysm or dissection (Jones et al., 2011).

Our miRNA study was also extended to examining likely interactions of miRNA with mediators of aortic damage with an enhanced understanding of the pathogenetic pathways and how to inhibit these pathways and delay the expansion of (AAA). We have come up with multiple protein targets for miRNAs (Figure 16) in focus and their murine study is in progress.

5.7 miRNA as biomarkers and therapeutic agents

MicroRNAs are an emerging class of highly conserved, non-coding small RNAs that regulate gene expression on the post-transcriptional level by inhibiting the translation of protein from mRNA or by promoting the degradation of mRNA. Unlike conventional drug molecules, miRNAs may target many mRNAs at the same time and therefore may influence the translation of multiple genes that contribute to common cellular and biological functions (Guo et al., 2010) (van Rooij and Olson, 2012).

miRNAs have been shown to play important roles in developmental biology, cellular stress, circadian rhythm, and immunology, as well as numerous disease states, including Alzheimer disease, cancer, and heart failure. Abundant evidence demonstrates a critical role for miRNAs in normal human hematopoiesis (Edelstein and Bray, 2011). Further, the possibility of using miRNAs as therapeutic tools has emerged. Using synthetic oligo-nucleotides that antagonize miRNAs binding to their mRNAs-targets or synthetic miRNAs mimics that enhance endogenous their function potentially will ultimately lead to the manipulation of platelet miRNAs expression and function with significant effects on specific protein levels and overall platelet reactivity (Dimitrios et al., 2013).

Platelet miRNAs have the potential to serve as biomarkers for platelet-mediated diseases and may also have future therapeutic application. Biomarkers can be diagnostic, prognostic, or predictive in

nature. In general, miRNAs are very biochemically stable and compared to mRNAs have superior performance characteristics as biomarkers for disease activity (Lu et al., 2008) (Kai and Pasquinelli, 2010) (Scholer et al., 2010).

There are potential uses of platelet miRNA as biomarkers and therapeutic agents. Due to the ability of platelets to release miRNA-containing microparticles at sites of activation, including angiogenic regions, tumors, and atherosclerotic plaques, there is the possibility of engineering platelets to deliver miRNA-based therapies to these disease sites (Edelstein and Bray, 2012).

5.8   Challenges in research

miRNA studies investigating the role of platelet in various pathophysiological conditions are increasingly published in recent years. Some of the initial studies therefore widely used computational prediction analysis and lacked direct mechanistic evidences in platelets. Despite the unique challenges, the platelet research area represents a great scope for novel findings to reveal yet unknown molecular mechanisms that may regulate platelet genesis and functions. Further advances in the platelet field may identify new therapeutic targets in platelets. The major shortcoming for our research is that the work is still in progress regarding the experimental work; platelet apoptosis is now being investigated in murine megakaryocytes by Dr. Schulze's group for our most promising miRNAs let-7d, let-7g, let-7f, miR-130a and miR-13b. Similarly Dr. Busch is studying abdominal aortic aneurysm in mouse models. Here the bioinformatics analysis gave him an important clue which miRNAs are probably implicated (miR-194-5p, miR-362-3p, miR-550a-5p, miR-19b-1-5p, miR-769-5p).

Metabolic modelling involves efforts in future to perform more intensive bioinformatics application analysis, contributing towards metabolic flux and mass isotopomers distribution

analysis, with the involvement of further well published and highly in use tools (e.g. COBRA toolbox (Becker et al., 2007), (Schellenberger, 2011), efmtool (Terzer and Stelling, 2008), Yana (Hoffmann-LA and Roche, 2012), Yanasquare (Schwarz et al., 2007), Yanavergence (Liang et al., 2011), and own software currently under extension for isotopologue analysis called Isotopo and LS-MIDA.

However, there are also general limitations apparent from our comparison such as the modelling of large-scale ("genome-sized") networks. These may be tackled using clever sampling procedures (Barrett et al., 2009), (Price et al., 2004) but this delivers not an ideal, detailed picture on metabolic adaptation potential suitable to model a more complex scenario than a lethal knockout (e.g. double knockout, partial enzyme inactivation etc.). Similarly, we are far from a detailed modelling of regulatory interactions. This includes further complexities such as protein stability and RNA turnover (Wiechert and de Graaf, 1997). The largest challenge is the kinetics of biochemical reactions. Only small enzyme networks have a chance to be modelled in such detail as so many parameters are unknown and the calculations become very complex. However, for each of these challenges there are continuous efforts to master these better (e.g. (Dugar and Stephanopoulos, 2011)) and rapid progress has been made in the last ten years.

6 Bibliography

A

- Ahmed Z, Zeeshan S, Huber C, Hensel M, Schomburg D, Münch R, Eisenreich W, Dandekar T. Software LS-MIDA for efficient Mass Isotopomer Distribution Analysis. BMC Bioinformatics, 14:2018, 2013a.

- Ahmed Z, Majeed S, Dandekar T. Unified Modeling and HCI Mockup Designing towards MIDA. Int. Jr. Emerg. Sci., 2:361-382, 2012a.

- Ahmed Z, Helfrich-Förster C, Dandekar T. Integrating Formal UML Designs and HCI Patterns with Spiral SDLC in DroLIGHT Implementation. Rec. Pat. Comp. Sci., 6:58-98, 2013b.

- Ahmed Z, Helfrich-Förster C. DroLIGHT: Real Time Embedded System towards Endogenous Clock Synchronization of Drosophila. Front. Neuroinform. 7, 2013c.

- Ahmed Z, Helfrich-Förster C. DroLIGHT-2: Real Time Embedded and Data Management System for Synchronizing Circadian Clock to the Light-Dark Cycles. Rec. Pat. Comp. Sci. 6:pp 191-205, 2013d.

- Ahmed Z, Majeed S, Dandekar T. Formal UML Modelling of Isotopo, Bioinformatical Software for Mass Isotopomers Distribution Analysis. Software Engin. 2:147-159, 2012b.

- Ahmed Z. Bioinformatics tool towards GC-MS data analysis, with extraction, classification, standardization, visualization and management features. Metabolomics.4:177, 2014a.

- Ahmed Z, Saman Z, Huber C, Hensel M, Schomburg D, Muench R, Eylert E, Eisenreich W., Dandekar T. "Isotopo" a Database Application for Facile Analysis and Management of Mass Isotopmer Data, Database, Oxford Univ. Press, 2014b.

- Ahmed Z. Ant-App-Database towards Neural, Behavioral Research on Deserts Ants and Approximate Solar Estimations. Front. Neuroinform. 8, 2014c.

- Ahmed Z. Towards Performance Measurement and Metrics based Analysis of PLA Applications. Int. J. Software Engin. App., 1:66-80, 2010.

- Ahmed Z, Ganti SK, Kyhlbäck H. Design Artifact's, Design Principles, Problems, Goals and Importance. In Fourth International Conference of Statistical Sciences Pakistan, pp.57-68, 2008.

- Ambler S. Agile Modeling: Effective Practices for eXtreme Programming and the Unified Process. Wiley Computer Publishing, 2002.

- Adam M, Raaz U, Spin JM, Tsao PS. MicroRNAs in Abdominal Aortic Aneurysm. Curr Vasc Pharmacol. 2013.

B

- Bartlett JC and Toms EG. Developing a protocol for bioinformatics analysis: An integrated information behavior and task analysis approach. JASIST, 56: 469-482, 2005.

- Bartel DP. MicroRNAs: target recognition and regulatory functions. Cell. 136:215–233, 2009.

- Borchert GM, Lanier W, Davidson BL. RNA polymerase III transcribes human microRNAs. Nat Struct Mol Biol. 13(12):1097-1101, 2006.

- Bruchova H, Merkerova M, Prchal JT. Aberrant expression of microRNA in polycythemia vera. Haematologica.93 (7):1009-1016, 2008.

- Baek D, Villen J, Shin C, Camargo FD, Gygi SP, Bartel DP. The impact of microRNAs on protein output. Nature. 455:64 –71, 2008.

- Boehm B. Software Engineering, IEEE Trans. On Computers, 12:1226-1242, 1976.

- Bruza PD and van-der-Weide TP. The Semantics of Data Flow Diagrams. In International Conference on Management of Data, 1993.

- Boehm BW. A spiral model of software development and enhancement. Computer. 21:61-72, 1988.

- Brett F et al. Bacterial virulence factors and uses thereof. Patent EP2462948A1, 2012.

- Becker SA et al. Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox. Nature Protocols. 2:727–738, 2007.

- Barrett CL, Herrgard MJ, Palsson BO. Decomposing complex reaction networksusing random sampling, principal component analysis and basis rotation. BMC Syst. Biol. 3:30, 2009.

- Brauman JI. Least Squares Analysis and Simplification of Multi-Isotope Mass Spectra. Anal. Chem. 38, 607–610, 1966.

- Boyanova D, Nilla S, Klau GW, Dandekar T, Müller T, Dittrich M. Functional Module Search in Protein Networks based on Semantic Similarity Improves the Analysis of Proteomics Data. Mol. Cell Proteomics. 13(7):1877-89, 2014.

- Boyanova D, Nilla S, Birschmann I, Dandekar T, Dittrich M. PlateletWeb: A systems biologic analysis of signaling networks in human platelets. Blood. 119(3):e22-34, 2012.

- Bertino AM, Qi XQ, Li J, Xia Y, Kuter DJ. Apoptotic markers are increased in platelets stored at 37 degrees C. Transfusion, 43 (7): 857–866, 2003.

C

- Cimmino G, Golino P. Platelet Biology and Receptor Pathways Journal of Cardiovascular Translational Research. 6(3):299-309, 2013.

- Cai X, Hagedorn CH, Cullen BR. Human micro- RNAs are processed from capped, polyadenylated transcripts that can also function as mRNAs. RNA.10(12):1957-1966, 2004.

- Cheloufi S, Dos Santos CO, Chong MM, Hannon GJ. A dicer-independent miRNA biogenesis pathway that requires Ago catalysis. Nature.465(7298):584-589, 2010.

- Cifuentes D, Xue H, Taylor DW, et al. A novel miRNA processing pathway independent of Dicer requires Argonaute2 catalytic activity. Science. 328(5986):1694-1698, 2010.

- Clemmons, DR. and Maile, LA. Interaction between Insulin-Like Growth Factor-I Receptor and V3 Integrin Linked Signaling Pathways: Cellular Responses to Changes in Multiple Signaling Inputs. Molecular Endocrinology, 19(1): 1–11, 2006.

- Cheriet H, Bounour N. Software evolution: Models and challenges. In International Conference on Machine and Web Intelligence (ICMWI), pp. 479-481, 2011.

- Chikofsky EH, Cross JH. Reverse Engineering and Design Recovery: A Taxonomy. IEEE Soft., 7:13–17, 1990.

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

- Clarke MCH, Savill J, Jones DB, Noble BS, Brown SB. Compartmentalized megakaryocyte death generates functional platelets committed to caspase-independent death. J. Cell Biol. 160 (4): 577–587, 2003.

D

- de Groot PG, Urbanus RT, Roest M. Platelet interaction with the vessel wall. Handb Exp Pharmacol. 210:87-110, 2012.

- Dangwal S, Thum T. microRNA Therapeutics in Cardiovascular Disease Models. Annu. Rev. Pharmacol. Toxicol. 54:185–203, 2014.

- Denis MM, Tolley ND, Bunting M, Schwertz H, Jiang H, Lindemann S, Yost CC, Rubner FJ, Albertine KH, Swoboda KJ, Fratto CM, Tolley E, Kraiss LW, McIntyre TM, Zimmerman GA, Weyrich AS. Escaping the nuclear confines: signal-dependent pre-mRNA splicing in anucleate platelets. Cell, 122: 379–391, 2005.

- Dorsam, RT. Kunapuli, SP. Central role of the P2Y12 receptor in platelet activation. J. Clin. Invest., 113: 340–345, 2004.

- Derynck R, Akhurs, RJ, Balmain A. TGF-b signaling in tumor suppression and cancer progression. Nat Genet., 29: 117-129, 2001.

- Delafontaine, P., Song, YH., and Li, Y. Expression, Regulation, and Function of IGF-1, IGF-1R, and IGF-1 Binding Proteins in Blood Vessels. Arterioscler Thromb Vasc Biol., 24: 435-444, 2004.

- Dumas M, ter-Hofstede AHM. UML Activity Diagrams as a Workflow Specification Language. In Fourth International Conference on The Unified Modeling Language, Modeling Languages, Concepts, and Tools, UK, pp.76-90, 2001.

- Deepak C, Bergmann FT, Sauro HM. TinkerCell: modular CAD tool for synthetic biology. J. Biol. Eng. 3:19, 2009.

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

E

- Eckly A, Gendrault JL, Hechler B, Cazenave JP, Gachet C. Differential Involvement of the P2Y1 and P2YT Receptors in the Morphological Changes of Platelet Aggregation. Thromb Haemost., 85: 694-701, 2001.

- Edelstein LC, F. Paul, M.D. Bray. Small RNAs as Potential Platelet Therapeutics. Antiplatelet Agents. Handbook of Experimental Pharmacology. 2012.

- Edelstein LC, McKenzie SE, Shaw C, Holinstat MA, Kunapuli SP, Bray PF. MicroRNAs in platelet production and activation. J Thromb Haemost. 11 (Suppl. 1): 340–50, 2013.

- Edelstein LC, Bray PF. MicroRNAs in platelet production and activation. Blood. 117: 5289-5296, 2011.

- Emmrich S, Henke K, Hegermann J, Ochs M, Reinhardt D, Klusmann JH. miRNAs can increase the efficiency of ex vivo platelet generation. Annals of Hematology. 91(11): 1673-1684, 2012.

- Eliceiri BP. Integrin and Growth Factor Receptor Crosstalk. Circ. Res., 89: 1104-1110, 2001.

- Egyed A, Kruchten PB. Rose/Architect: A Tool to Visualize Architecture. In Thirty Second Annual Hawaii Conference on Systems Sciences, USA, pp.8066, 1999

- Eisenreich W, T. Dandekar, J. Heesemann and W. Goebel. Carbon metabolism of intracellular bacterial pathogens and possible links to virulence. Nat. Rev. Microbiol, 8:401-412, 2010.

- Eylert E, Herrmann V, Jules M, Gillmaier N, Lautner M, Buchrieser C, Eisenreich W, Heuner K. Isotopologue profiling of Legionella pneumophila: role of serine and glucose as carbon substrates. J. Biol. Chem., 285:22232-22243, 2010.

- Earl HA, Simon C. System, method, and computer software product for genotype determination using probe array data. Patent US8200440, 2012.

- Edwards JS, Ibarra RU, Palsson NO. In silico predictions of Escherichia coli metabolic capabilities are consistent with experimental data. Nat Biotechnol. 19(2):125-30, 2001.

- Eisenreich W, Slaghuis J, Laupitz R, Bussemer J, Stritzker J, Schwarz C, Schwarz R, Dandekar T, Goebel W, Bacher A. 13C isotopologue perturbation studies of Listeria monocytogenes carbon metabolism and its modulation by the virulence regulator PrfA. Proc. Natl. Acad. Sci. USA. 103() no.7, pp.2040-2045, 2006.

F

- Friedman RC, Farh KK, Burge CB, Bartel DP. Most mammalian mRNAs are conserved targets of microRNAs. Genome Res 19: 92–105, 2009.

- Frishman et al. The PEDANT genome database. Nucleic Acids Res. 31:207–211, 2003.

- Fan JB, Marina B. Methods and Compositions for Diagnosing Conditions Associated with Specific DNA Methylation Patterns. Patent US2005/0026183A1, 2005.

G

- Gatsiou A, Boeckel JN, Randriamboavonjy V, Stellos K. MicroRNAs in Platelet Biogenesis and Function: Implications in Vascular Homeostasis and Inflammation. Current Vascular Pharmacology, 2012 VOL: 10, ISSUE: 5, Pages 524-531 (8)

- Guo H, Ingolia NT, Weissman JS, Bartel DP. 2010. Mammalian microRNAs predominantly act to decrease target mRNA levels. Nature 466:835–40

- Guo H, Ingolia NT, Weissman JS, Bartel DP. Mammalian microRNAs predominantly act to decrease target mRNA levels. Nature, 466: 835–40, 2010.

- Grimson A, Farh KK, Johnston WK, Garrett-Engele P, Lim LP, Bartel DP. MicroRNA targeting specificity in mammals: determinants beyond seed pairing. Mol Cell, 27: 91–105, 2007.

- Griffiths-Jones S, Saini HK, van Dongen S, Enright AJ. miRBase: tools for microRNA genomics. Nucleic acids research 36: D154–158, 2008.

- Grimson A1, Farh KK, Johnston WK, Garrett-Engele P, Lim LP, Bartel DP. MicroRNA Targeting Specificity in Mammals: Determinants beyond Seed Pairing. Molecular Cell, 27: 91-105, 2007.

- Gaudermann, P., Vogl, I., Zientz, E., Silva, F.J., Moya, A., Gross, R. Dandekar, T. Function predictions for conserved hypothetical proteins and proteins with putative function in Blochmannia floridanus. BMC Microbiology 6: 1, 2006.

- Goto S et al. KEGG/EXPRESSION: A Database for Browsing and Analysing Microarray Expression Data. Genome Informatics. 11:222-223, 2000.

- Graham B, Lee L, Christophe L. Time to Event Data Analysis Method and System. Patent US20120066163A1, 2012.

- Gille C, Hoffmann S, Holzhütter HG. METANNOGEN: compiling features of biochemical reactions needed for the reconstruction of metabolic networks. BMC Systems Biology. 1:5, 2007.

- Garzon R, Pichiorri F, Palumbo T, et al. MicroRNA fingerprints during human megakaryocytopoiesis. Proc Natl Acad Sci USA. 103: 5078-83, 2006.

- Gelinas C, White E. BH3-only proteins in control: specificity regulates MCL-1 and BAK-mediated apoptosis. Genes Dev. 19 (11): 1263–1268, 2005.

- Gyulkhandanyan AV, Mutlu A, Freedman J, Leytin V. Selective triggering of platelet apoptosis, platelet activation or both. Br J Haematol. 161: 245–254, 2013.

H

- Hellerstein MK, Neese RA. Mass isotopomer distribution analysis at eight years: theoretical, analytic, and experimental considerations. Am. J. Physiol.276:1146-70, 1999.

- Hellerstein MK, Neese RA. Mass isotopomer distribution analysis: a technique for measuring biosynthesis and turnover of polymers. Am. J. Physiol. 263:988–1001, 1992.

- Hirai I, Wang HG. Survival-factor-induced phosphorylation of Bad results in its dissociation from Bcl-x(L) but not Bcl-2. Biochem J., 359(2): 345-52, 2001.

- Hull C, Feygin M, Baron Y, Sanders R, Sachs E, Lightman A, Wohlers T. Rapid prototyping: current technology and future potential. Rapid Prototyping Journal. 1:11-19, 1995.

- Hoffmann-LA A, Roche AG. A Method of Analyzing Chromosomal Translocations and A System Therefore. Patent WO2012123387, 2012.

- Hubbard AE. Capability based distributed processing. Patent 8249940, 2012.

- Hoops S et al. COPASI—a COmplex PAthway Simulator. Bioinformatics. 22:3067-3074, 2006.

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

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

- Herbert MS. SCAMP: a general-purpose simulator and metabolic control analysis program. Comput. Appl. Biosci. 9:441-450, 1993.

- Hellerstein MK, Neese RA. Mass isotopomer distribution analysis at eight years: theoretical, analytic, and experimental considerations", Am. J. Physiol. 276(6):146-1170, 1999.

- Hites RA. Gas chromatography mass spectrometry. Handbook of instrumental techniques for analytical chemistry, ed. F.Settle. Upper Saddle River, N. J.: Prentice Hall. 609–626, 1997.

I

- Italiano JE Jr, Lecine P, Shivdasani RA, Hartwig JH. Blood platelets are assembled principally at the ends of proplatelet processes produced by differentiated megakaryocytes. J Cell Biol. 147(6):1299-312, 1999.

J

- Jones JA, Stroud RE, O'Quinn EC, Black LE, Barth JL, Elefteriades JA, Bavaria JE, Gorman JH 3rd, Gorman RC, Spinale FG, Ikonomidis JS. Selective microRNA suppression

in human thoracic aneurysms: relationship of miR-29a to aortic size and proteolytic induction. Circ Cardiovasc Genet. 6:605-13, 2011.

- John B, Enright AJ, Aravin A, Tuschl T, Sander C, Marks DS.. Human MicroRNA targets. PLoS Biol 2: e363, 2005.

- Jackson SP1, Yap CL, Anderson KE. Phosphoinositide 3-kinases and the regulation of platelet function. Biochemical Society Transactions, 32(2): 387– 392, 2004.

- Jaffe M.S., Leveson N.G., Heimdahl M.P.E., Melhart B.E. Software requirements analysis for real-time process-control systems. IEEE Transactions on Software Engineering, 17:241-258, 1991.

K

- Kai ZS, Pasquinelli AE. MicroRNA assassins: factors that regulate the disappearance of miRNAs. Nat Struct Mol Biol. 17: 5–10, 2010.

- Krek A, Grun D, Poy MN, Wolf R, Rosenberg L, Epstein EJ, Mac- Menamin P, da Piedade I, Gunsalus KC, Stoffel M, Rajewsky N. Combinatorial microRNA target predictions. Nat Genet.37:495–500, 2005.

- Krek A1, Grün D, Poy MN, Wolf R, Rosenberg L, Epstein EJ, MacMenamin P, da Piedade I, Gunsalus KC, Stoffel M, Rajewsky N. Combinatorial microRNA target predictions. Nature genetics 37: 495–500, 2005.

- Kauffenstein G, Bergmeier W, Eckly A, Ohlmann P, Léon C, Cazenave JP, Nieswandt B, Gachet C. The P2Y(12) receptor induces platelet aggregation through weak activation of the alpha(IIb)beta(3) integrin—a phosphoinositide 3-kinase-dependent mechanism. FEBS letters, 505(2): 281-90, 2001.

- Kim S, Garcia A, Jackson SP, Kunapuli SP. Insulin-like growth factor-1 regulates platelet activation through PI3-K isoform. Blood, 110: 4206-4213, 2007.

- Kaleta C, de Figueiredo LF, Schuster S. 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, 2009.

- Kanehisa M. The KEGG database. Novartis Found Symp. 247: 91-101, 2002.

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

- Korzekwa K, Howald WN, Trager WN. The Use of Brauman's Least Squares Approach for the Quantification of Deuterated Chlorophenols. Biomed. Environ. Mass Spectrom. 19:211-217, 1999.

L

- Laffont B1, Corduan A, Plé H, Duchez AC, Cloutier N, Boilard E, Provost P. Activated platelets can deliver mRNA regulatory Ago2•microRNA complexes to endothelial cells via microparticles. Blood. 122: 253-261, 2013.

- Landry P, Plante I, Ouellet DL, Perron MP, Rousseau G, Provost P. Existence of a microRNA pathway in anucleate platelets. Nat Struct Mol Biol. 16(9):961-966, 2009.

- Landry P, Plante I, Ouellet DL, Perron MP, Rousseau G, Provost P. Existence of a microRNA pathway in anucleate platelets. Nat Struct Mol Biol., 16(9): 961-966, 2009.

- Leonard C. Edelstein, Paul F. Bray M.D. Small RNAs as Potential Platelet Therapeutics. Antiplatelet Agents. Handbook of Experimental Pharmacology Volume 210, 2012, pp 435-445

- Letondal C. Participatory programming: Developing programmable bioinformatics tools for end-users. End-User Development, Springer, 2005.

- Liu G, Huang Y, Lu X, Lu M, Huang X, Li W, Jiang M. Identification and characteristics of microRNAs with altered expression patterns in a rat model of abdominal aortic aneurysms. Tohoku J Exp Med.222:187–193, 2010.

- Lee RC, Feinbaum RL, Ambros V. The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. Cell. 75: 843–54, 1993.

- Lee Y, Kim M, Han J, et al. MicroRNA genes are transcribed by RNA polymerase II. EMBO J.23(20):4051-4060, 2004.

- Lu J, Guo S, Ebert BL, Zhang H, Peng X, Bosco J, Pretz J, Schlanger R, Wang JY, Mak RH, Dombkowski DM, Preffer FI, Scadden DT, Golub TR. MicroRNA-mediated control of cell fate in megakaryocyte-erythrocyte progenitors. Dev Cell. 14: 843–53, 2008.

- Lewis BP, Burge CB, Bartel DP. Conserved Seed Pairing, Often Flanked by Adenosines, Indicates that Thousands of Human Genes are MicroRNA Targets. Cell, 120: 15-20, 2005.

- Lee J, and Xue NL. Analyzing user requirements by use cases: a goal-driven approach. IEEE Softw., 16: 92–101, 1999.

- Larman C, Basili VR. Iterative and Incremental Development: A Brief History. Computer, 36: 47-56, 2003.

- Logan-Klumpler FJ et al. GeneDB—an annotation database for pathogens. Nucl. Acids Res. 40: 98-108, 2011.

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

- Lee GH et al. E2D: A Novel Tool for Annotating Protein Domains in Expressed Sequence Tags. CIBCBIEEE, pp. 1-6, 2006.

- Loboda A et al. Gene Expression Signature for Assessing RAS Pathway Activity. Patent US20100280987A1, 2010.

- Lee DY et al. WebCell: a web-based environment for kinetic modeling and dynamic simulation of cellular networks. Bioinformatics. 22:1150-1151, 2006.

- Liang C et al. Staphylococcus aureus physiological growth limitations: insights from flux calculations built on proteomics and external metabolite data. Proteomics. 11:1915-1935, 2011.

- Lee WN, Byerley LO, Bergner EA, Edmond J. Mass isotopomer analysis: theoretical and practical considerations. John Wiley & Sons, Inc., 20(8):451-458, 1991.

- Lee WN, Bergner EA. Guo ZK. Mass isotopomer pattern and precursor-product relationship. Biol. Mass. Spectrom. 21, 114–122, 1992.

- Li J, Xia Y, Bertino AM, Coburn JP, Kuter DJ. The mechanism of apoptosis in human platelets during storage. Transfusion, 40 (11): 1320–1329, 2000.

- Lin KH, Chang HC, Lu WJ, Jayakumar T, Chou HC, Fong TH et al. Comparison of the relative activities of inducing platelet apoptosis stimulated by various platelet-activating agents. Platelets, 20: 575–581, 2009.

M

- Mattick JS, Makunin IV. Non-coding RNA. Hum Mol Genet. 15 Spec 1:R17-R29, 2006.

- Massar JP, Travers M, Elhai J, Shrager J. BioLingua: A programmable knowledge environment for biologists. Bioinformatics, 21(2):199-207, 2005.

- MacMullen WJ, Denn SO. Information problems in molecular biology and bioinformatics. JASIST, 56: 447-456, 2005.

- Maegdefessel L, Azuma J, Toh R, Deng A, Merk DR, Raiesdana A, Leeper NJ, Raaz U, Schoelmerich AM, McConnell MV, et al.: MicroRNA-21 blocks abdominal aortic aneurysm development and nicotine-augmented expansion. miR-21 was shown to be upregulated in human AAA tissue Sci Transl Med. 4:122, 2012a.

- Maegdefessel L, Azuma J, Toh R, Merk DR, Deng A, Chin JT, Raaz U, Schoelmerich AM, Raiesdana A, Leeper NJ, McConnell MV, Dalman RL, Spin JM, Tsao PS. Inhibition of microRNA-29b reduces murine abdominal aortic aneurysm development. J Clin Invest. 122(2):497-506, 2012b.

- Mason KD, Carpinelli MR, Fletcher JI, Collinge JE, Hilton AA, Ellis S, Kelly PN, Ekert PG, Metcalf D, Roberts AW, Huang DC, Kile BT. Programmed Anuclear Cell Death Delimits Platelet Life Span. Cell, 128: 1173–1186, 2007.

- Mazzucato M, Cozzi MR, Pradella P, Ruggeri ZM, De Marco L. Distinct Roles of ADP Receptors in Von Willebrand Factor-Mediated Platelet Signaling and Activation Under High Flow. Blood, 104: 3221-3227, 2004.

- Moustakas A, Pardali K, Gaal A, Heldin CH. Mechanisms of TGFbeta signaling in regulation of cell growth and differentiation. Immunol Lett., 82(1- 2): 85-91, 2002.

- Martelli AM, Cocco L, Capitani S, Miscia S, Papa S, Manzoli FA. Nuclear phosphatidylinositol 3,4,5-trisphosphate, phosphatidylinositol 3-kinase, Akt, and PTen: emerging key regulators of anti-apoptotic signaling and carcinogenesis. Eur J Histochem., 1: 125-31, 2007.

- Marilyn B. A guide for programmers. Prentice-Hall. 1978.

- Maile, LA., and Clemmons, DR. Regulation of insulin-like growth factor Ireceptor dephosphorylation by SHPS-1 and the tyrosine phosphatase SHP-2. J Biol Chem., 277: 8955–8960, 2002.

- Mora, A., Sakamoto, K., McManus, EJ., and Alessi, DR. Role of the PDK1-PKB-GSK3 pathway in regulating glycogen synthase and glucose uptake in the heart. FEBS Lett., 579(17): 3632-8, 2005.

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

- Meyer F et al. GenDB--an open source genome annotation system for prokaryote genomes. Nucleic Acids Res. 31:2187-2195, 2003.

- Matsuoka Y, Shimizu K. 13C-Metabolic Flux Analysis and Metabolic Regulation. Chemical Biology, 2012.

N

- Nagalla S, Shaw C, Kong X, Kondkar AA, Edelstein LC, Ma L, Chen J, McKnight GS, López JA, Yang L, Jin Y, Bray MS, Leal SM, Dong JF, Bray PF. Platelet microRNA-mRNA coexpression profiles correlate with platelet reactivity. Blood. 117: 5189-5197, 2011.

- Ni H, Freedman J. Platelets in hemostasis and thrombosis: role of integrins and their ligands. Transfusion and Apheresis Science, 28: 257–264, 2003.

- Nurden AT. Platelets, inflammation and tissue regeneration. Thromb Haemost.;105 (Suppl 1): S13–33, 2011.

- Nahnsen S, Bertsch A, Rahnenführer J, Nordheim A, Kohlbacher O. Probabilistic Consensus Scoring Improves Tandem Mass Spectrometry Peptide Identification. J. Proteome Res. 10:3332-3343, 2011.

O

- Opalinska JB, Bersenev A, Zhang Z, Schmaier AA, Choi J, Yao Y, D'Souza J, Tong W, Weiss MJ. MicroRNA expression in maturing murine megakaryocytes. Blood. 116: e128-e138, 2010.

- Oehm S et al. Comparative Pathway Analyzer--a web server for comparative analysis, clustering and visualization of metabolic networks in multiple organisms. Nucleic Acids Res. 36:433-437, 2008.

P

- Patel SR, Hartwig JH, Italiano JE. The biogenesis of platelets from megakaryocyte proplatelets. J. Clin. Invest. 115:3348–3354.

- Patel SR., Richardson JL, Schulze H, Kahle E, Galjart N, Drabek K, Shivdasani RA, Hartwig JH, Italiano JE. Differential roles of microtubule assembly and sliding in proplatelet formation by megakaryocytes. Blood. 106:4076–4085.

- Petriv OI, Kuchenbauer F, Delaney AD, Lecault V, White A, Kent D, Marmolejo L, Heuser M, Berg T, Copley M, Ruschmann J, Sekulovic S, Benz C, Kuroda E, Ho V, Antignano F, Halim T, Giambra V, Krystal G, Takei CJ, et al. Comprehensive microRNA expression profiling of the hematopoietic hierarchy. Proc Natl Acad Sci U S A, 107: 15443–8, 2010.

- Plé H, Landry P, Benham A, Coarfa C, Gunaratne PH, Provost P. The Repertoire and Features of Human Platelet microRNAs. PLoS ONE 7(12): e50746, 2012.

- Phillips, DR., Prasad, KS., Manganello, J., Bao, M., and Nannizzi-Alaimo, L. Integrin tyrosine phosphorylation in platelet signaling. Curr Opin Cell Biol., 13: 546–554, 2001.

- Pfleeger SL and Bohner SA. A Framework for Software Maintenance Metrics. IEEE Transactions on Software Engineering, pp. 320-327, 1990.

- Petersen K, Wohlin C, Baca D. The Waterfall Model in Large-Scale Development. Product-Focused Software Process Improvement, Lecture Notes in Business Information Processing, 32:386-400, 2009.

- Pei Y, Wei Y, Furia CA, Nordio M, Meyer B. Code-Based Automated Program Fixing. In 26th IEEE/ACM International Conference on Automated Software Engineering (ASE), pp. 392 – 395, 2011.

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

- Pitkänen,E. et al. 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, 2008.

- Previs, S. F., Fernandez C. A., Yang D., Soloviev M. V., France D., Brunengraber H. Limitations of the Mass Isotopomer Distribution Analysis of Glucose to Study Gluconeogenesis. J. Biol. Chem. 277: 1998.

- Papin JA, Stelling J, Price NJ, Klamt S, Schuster S, Palsson BO. Comparison of network-based pathway analysis methods. Trends Biotechnol. 22: 400–405, 2004.

- Perillo JR, Guerra EM, Fernandes CT. Daileon: A Tool for Enabling Domain Annotations. Proceedings of the Workshop on AOP and Meta-Data for Software Evolution, pp.1-4, 2009.

- Price ND, Schellenberger J, Palsson BO. Uniform sampling of steady-state flux spaces: means to design experiments and to interpret enzymopathies. Biophys. J., 87:2172-86, 2004.

- Pahl MC, Derr K, Gäbel G, Hinterseher I, Elmore JR, Schworer CM, Peeler TC, Franklin DP, Gray JL, Carey DJ, Tromp G, Kuivaniemi H. MicroRNA expression signature in human abdominal aortic aneurysms. BMC Med Genomics, 5:25, 2012.

Q

- Qin S, Zhang C. MicroRNAs in vascular disease. J Cardiovasc Pharmacol.57:8 –12, 2011.

- Qi B, Hardwick JM. A Bcl-xL Timer Sets Platelet Life Span. Cell, 128: 1035-1036, 2007.

- Quek LE et al. OpenFLUX: efficient modelling software for 13C-based metabolic flux analysis. Microbial Cell Factories, 8:25, 2009.

R

- Raab M1, Daxecker H, Edwards RJ, Treumann A, Murphy D, Moran N. Protein interactions with the platelet integrin IIb regulatory motif. Proteomics. 15:2790-800, 2010.

- Richardson JL, Shivdasani RA, Boers C, Hartwig JH, Italiano JE. Mechanisms of organelle transport and capture along proplatelets during platelet production. Blood. 106:4066–4075.

- Rigoutsos I, Tsirigos A. MicroRNA target prediction. In: Slack FJ, ed. MicroRNAs in Development and Cancer. Hackensack, NJ: Imperial College Press,237–64, 2011.

- Reczko M, Maragkakis M, Alexiou P, Papadopoulos GL, Hatzigeorgiou AG. Accurate microRNA Target Prediction Using Detailed Binding Site Accessibility and Machine Learning on Proteomics Data. Frontiers in genetics 2: 103, 2011.

- Rook P. Controlling software projects. Softw. Eng. J. 1:7-16, 1986.

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

S

- Scholer N, Langer C, Dohner H, Buske C, Kuchenbauer F. Serum microRNAs as a novel class of biomarkers: a comprehensive review of the literature. Exp Hematol, 38: 1126–30, 2010.

- Selbach M, Schwanhausser B, Thierfelder N, Fang Z, Khanin R, Rajewsky N. Widespread changes in protein synthesis induced by microRNAs. Nature. 455:58–63, 2008.

- Sethupathy P, Megraw M, Hatzigeorgiou AG. A guide through present computational approaches for the identification of mammalian microRNA targets. Nat Methods. 3:881–886, 2006.

- Small EM, Frost RJ, Olson EN. MicroRNAs add a new dimension to cardiovascular disease. Circulation. 121:1022–1032, 2010.

- Stakos DA, Gatsiou A, Stamatelopoulos K, Tselepis AD, Stellos K. Platelet microRNAs: From platelet biology to possible disease biomarkers and therapeutic targets. Platelets, 24 (8), 579-589, 2013.

- Shelton, JG., Steelman, LS., White, ER., and McCubrey, JA. Synergy between PI3K/Akt and Raf/MEK/ERK pathways in IGF-1R mediated cell cycle progression and prevention of apoptosis in hematopoietic cells. Cell Cycle, 3(3): 372-9, 2004.

- Schwartz, MA., and Ginsberg, MH. Networks and crosstalk: integrin signaling spreads. Nat Cell Biol., 4(4): E65-8, 2002.

- Schwartz, MA., and Shattil, SJ. Signaling networks linking integrins and rho family GTPases. Trends Biochem Sci., 25: 388–391, 2000.

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

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

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

- Schuster,R., Schuster,S. 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, 1993.

- Schuster S, Fell DA, Dandekar T. A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks. Nat. Biotechnol. 18:326-332, 2000.

- Sajjad Mahmood, Richard Lai. RE-UML: A Component-Based System Requirements Analysis Language, Comput. J. 56(7): 901-922, 2013.

- Szyperski C, Gruntz D and Murer S. Component Software—Beyond Object-Oriented Programming. Addison-Wesley, 2002.

- Sommerville I. Integrated requirements engineering: a Tutorial. IEEE Softw., 22:16–23, 2005.

- Sammak PJ et al. Image-based methods for measuring global nuclear patterns as epigenetic markers of cell differentiation. Patent US8189900, 2012.

- Steffen K, Axel VK. An application programming interface for CellNetAnalyzer. Biosystems. 105:162-168, 2002.

- Sauro SH. Jarnac: a system for interactive metabolic analysis. In Proceedings of the 9th International Meeting on BioThermoKinetics, pp. 221–228, 2000.

- Schellenberger J et al. Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0. Nature Protocols. 6:1290–1307, 2011.

- Schlatter R, Philippi N, Wangorsch G, Pick R, Sawodny O, Borner C, Timmer J, Ederer M, Dandekar T. Integration of Boolean models exemplified on hepatocyte signal transduction. Briefings in Bioinformatics. 13(3):365-76, 2012.

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

T

- Tanriverdi K, Iafrati MD, Rex S, Blair PS, Freedman JE. Platelet MicroRNA is Altered by Thrombin-Induced Aggregation. Circulation, 114: 27-28, 2006.

- Tanriverdi K, Morin M, Beaulieu LM, Freedman JE. Platelet Activation Regulates Levels of MicroRNA. Circulation. 118: S-407, 2008.

- Takagi J, Petre BM, Walz T, Springer TA.Global conformational rearrangements in integrin extracellular domains in outside-in and inside-out signaling. Cell, 110: 599-611, 2002.

- Thomas M, Lieberman J, Lal A. Desperately seeking microRNA targets. Nat Struct Mol Biol. 17:1169 –1174, 2010.

- Trikha M, Zhou Z, Timar J, Raso E, Kennel M, Emmell E, Nakada MT. Multiple Roles for Platelet GPIIb/IIIa and v3 Integrins in Tumor Growth, Angiogenesis, and Metastasis. Cancer Research, 62: 2824–2833, 2002.

- Turner NA1, Moake JL, McIntire LV. Blockade of Adenosine Diphosphate Receptors P2Y(12) and P2Y(1) Is Required to Inhibit Platelet Aggregation in Whole Blood Under Flow. Blood, 98: 3340-3345, 2001.

- Torres-Aleman, I. Role of insulin-like growth factors in neuronal plasticity and neuroprotection. Adv Exp Med Biol., 567: 243-58, 2005.

- Theocharidis A, van Dongen S, Enright AJ, Freeman TC. Network visualization and analysis of gene expression data using BioLayout Express3D. Nature Protocols. 4:1535-1550, 2009.

- Tatusov RL et al. The COG database: a tool for genome-scale analysis of protein functions and evolution. Nucleic Acids Res.28:33–36, 2000.

- Theocharidis A et al. Network visualization and analysis of gene expression data using BioLayout Express3D. Nature Protocols. 4:1535-1550, 2009.

- Terzer M, Stelling J. Large-scale computation of elementary flux modes with bit pattern trees. Bioinformatics.24:2229-35, 2008.

V

- van Rooij E, Olson EN. MicroRNA therapeutics for cardiovascular disease: opportunities and obstacles. Nat. Rev. Drug Discov. 11:860–72, 2012.

- van Rooij E. The art of microRNA research. Circ Res. 108(2):219-34, 2011.

- van Rooij E, Sutherland LB, Liu N, Williams AH, McAnally J, Gerard RD, Richardson JA, Olson EN. A signature pattern of stress-responsive microRNAs that can evoke cardiac hypertrophy and heart failure. Proc Natl Acad Sci USA. 103:18255–18260, 2006.

- Vincent, AM., and Feldman, EL. Control of cell survival by IGF signaling pathways. Growth Horm IGF Res., 12(4): 193-7, 2002.

- Van-Bramer SE. Introduction to Mass Spectrometry. Widener University, Department of Chemistry, September 2, 1998.

- Vogler M, Hamali HA, Sun XM, Bampton ET, Dinsdale D, Snowden RT, Dyer MJ, Goodall AH, Cohen GM. BCL2/BCL-X(L) inhibition induces apoptosis, disrupts cellular calcium homeostasis, and prevents platelet activation. Blood. 30;117(26):7145-54, 2011.

- Vanags DM, Orrenius S, Aguilar-Santelises M. Alterations in Bcl-2/Bax protein levels in platelets form part of an ionomycin-induced process that resembles apoptosis. Br J Haematol. 99 (4): 824–831, 1997.

W

- Wightman B, Ha I, Ruvkun G. Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in C. elegans. Cell. 75: 855–62, 1993.

- Wakefield LM, Roberts AB. TGF-b signaling: positive and negative effects on tumorigenesis. Curr Opin Genet Dev., 12: 22-29, 2002.

- Watson SP1, Auger JM, McCarty OJ, Pearce AC. GPVI and integrin aIIbb3 signaling in platelets. Journal of Thrombosis and Haemostasis, 3: 1752–1762, 2005.

- Weyrich AS, Lindemann S, Tolley ND, Kraiss LW, Dixon DA, Mahoney TM, Prescott SP, McIntyre TM, Zimmerman GA. Change in protein phenotype without a nucleus: translational control in platelets. Semin Thromb Hemost. 30(4): 491-8, 2004.

- William K. A simple guide to five normal forms in relational database theory. Commun. ACM. 26:120-125, 1983.

- Williams AJ et al. Methods and computer software for detecting splice variants. Patent US8170808, 2012.

- Wong PC et al. Sequential pattern data mining and visualization. Patent US8073859, 2012.

- Wiechert W, de Graaf AA. Bidirectional reaction steps in metabolic networks: I. Modeling and simulation of carbon isotope labeling experiments. Biotechnol Bioeng. 55:101-117, 1997.

- Wiechert W, de Graaf AA. Bidirectional reaction steps in metabolic networks: I. Modeling and simulation of carbon isotope labeling experiments. Biotechnol Bioeng. 55:101-117, 1997.

- White MJ, Schoenwaelder SM, Josefsson EC, Jarman KE, Henley KJ, James C, Debrincat MA, Jackson SP, Huang DC, Kile BT. Caspase-9 mediates the apoptotic death of megakaryocytes and platelets, but is dispensable for their generation and function. Blood, 119(18):4283-90, 2012.

X

- Xiao F, Zuo Z, Cai G, Kang S, Gao X, Li T. miRecords: an integrated resource for microRNA-target interactions. Nucleic acids research 37: D105–110, 2009.

Z

- Zhang H, Nimmer PM, Tahir SK, Chen J, Fryer RM, Hahn KR, Iciek LA, Morgan SJ, Nasarre MC, Nelson R, Preusser LC, Reinhart GA, Smith ML, Rosenberg SH, Elmore SW, Tse C. Bcl-2 family proteins are essential for platelet survival. Cell Death and Differentiation, 14: 943–951, 2007.

- Zheng, WH., Kar, S., and Quirion, R. Insulin-like growth factor-1-induced phosphorylation of transcription factor FKHRL1 is mediated by phosphatidylinositol 3-kinase/Akt kinase and role of this pathway in insulin-like growth factor-1-induced survival of cultured hippocampal neurons. Mol Pharmacol., 62(2): 225-33, 2002.

- Zamboni N, Fischer E, Sauer U. FiatFlux – a software for metabolic flux analysis from 13C-glucose experiments. BMC Bioinformatics.6:1-8, 2005.

- Zilversmit DB, Entenman C, Fishler M. On The Calculation of "Turnover Time" and "Turnover Rate" from Experiments involving the use of Labeling Agents. J Gen Physiol. 26:325–331, 1943.

7 Nomenclatures

Abundance Matrix (An)

Abundance Matrix (MRMn)

Allele specific oligonucleotide (ASO)

Alanine (Ala)

Aspartic Acid (Asp)

Computer Aided Design (CAD)

Constraint Based Modelling (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)

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)

Patient Guidance Services (PGS)

Phenotype phase plane (PhPP)

Phosphotransferase System (PTS)

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 Modelling Language (UML)

8   Curriculum Vitae

Mrs. Saman ZEESHAN
(maiden MAJEED)

Date of Birth: 11 April 1983

Marital Status: Married to Dr Zeeshan Ahmed
*   and have one son (Jibrael Zeeshan)*

Currently Resident: Wuerzburg, Germany
Email: saman.majeed@uni-wuerzburg.de
Objective: To pursue a challenging and successful career in Bioinformatics.

Education

- *2011-2014*: PhD Student (Promotion thesis) - Bioinformatics, Department of Bioinformatics, Biocenter, University of Wuerzburg, Germany.

- *2009-2011*: PhD Pre-Requisite (Promotionszulassungs thesis) - Bioinformatics, Department of Bioinformatics, Biocenter, University of Wuerzburg, Germany.

- *2009-2010*: Erasmus Mundus – Masters training course (EM-ABG) - Bioinformatics, Swedish University of Agricultural Sciences Uppsala University Sweden, Department of Animal Breeding and Genetics Wageningen University Netherland.

- *2003-2008*: Bachelor of Science - Bioinformatics (BS-BI), Department of Bioinformatics, COMSATS University, Islamabad Pakistan.

Positions and Employment

- *2011-2014*: Doctoral Researcher, Prof. Dr. Thomas Dandekar's Group, Department of Bioinformatics, University of Wuerzburg, Wuerzburg, Germany

- *2012*: Doctoral Researcher, Institute of Molecular and Translational Therapeutic Strategies (IMTTS), Hannover Medical School, Hannover, Germany

- *2009-2011*: Pre-Doctoral Researcher, Prof. Dr. Thomas Dandekar's Group, Department of Bioinformatics, University of Wuerzburg, Wuerzburg, Germany

- *2008-2009*: Bioinformatician, NUST Centre of Virology and Immunology, National University of Science and Technology, Pakistan

Honors

- Erasmus Mundus Scholarship for European Union Masters Training at Wageningen University Netherland and Swedish University of Agricultural Sciences Uppsala University Sweden, 2009.

- Gold Medal handed by Prime Minister Pakistan for securing Highest CGPA in Faculty in Bachelors by COMSATS Institute of Information Technology, 2008.

- Silver Medal handed by Prime Minister Pakistan for securing Highest CGPA amongst all Faculties in Bachelors by COMSATS Institute of Information Technology, 2008.

- Scholarship for outstanding performance throughout the 4 years Bachelors program, COMSATS Institute of Information Technology, 2004-2008.

- 2nd Position in Software Competition, COMSATS Institute of Information Technology, 2007.

Languages

- Deutsch als Fremdsprache, CEF Grundstufe A1.2, gut (2), and A1.1, sehr gut (1-)

- 2010: IELTS, Over all Band: 7.5

- 2009: French Language Certificate (Results: 93.5 %)

Place, Date   Signature:

Wuerzburg, 04/09/ 2014**.**   Saman Zeeshan (maiden Majeed)

9   List of Publications

Scientific, Journal Manuscripts

1. Ahmed Z, Fleischmann P, **Saman Z**, Rösseler W, Dandekar T (2014) "Ant-App-DB: A Smart Ecological Solution for Monitoring the Arthropods' Activities, Experimental Data Management and Solar Calculations without GPS in Behavioural Field Studies". [Submitted]

2. Ahmed Z, Michel M, **Saman Z**, Dandekar T, Mueller MJ, Fekete A (2014) "Lipid-Pro: A computational lipid identification solution for untargeted lipidomics on data-independent acquisition tandem mass spectrometry platforms", Bioinformatics. Oxford Journals, Oxford University Press. [Accept with Revision]

3. Ahmed Z, **Saman Z**, Huber C, Hensel M, Schomburg D, Münch R, Eylert E, Eisenreich W, Dandekar T (2014). "Isotopo Database and Tool for Facile Analysis and Management of Mass Isotopomer Data", Database: The Journal of Biological Databases and Curation, Oxford Journals, Oxford University Press.

4. Dandekar T, Fielsmann A, **Saman Z**, and Ahmed Z (2014). "Software Applications toward Quantitative Metabolic Flux Analysis and Visualization". Briefings in Bioinformatics, Oxford Journals, 15:1, 91-107, Oxford University Press.

5. Ahmed Z, **Saman Z***, Dandekar T. (2014) "Developing sustainable software solutions for bioinformatics by the "Butterfly" paradigm", F1000 Research, Faculty 1000 Publishers, 3:71.

6. Ahmed Z, **Saman Z***, Huber C, Hensel M, Schomburg D, Muench D, Eisenreich W, and Dandekar T (2013). "Software LS-MIDA for Efficient Mass Isotopomer Distribution Analysis in Metabolic Modeling". BMC Bioinformatics, 14:218. [Joint First Author]

7. Ahmed Z, **Saman Z***, and Dandekar T. (2012). "Computational Feature Performance and Domain Specific Architecture Evaluation of Software Applications towards Metabolic Flux Analysis". Recent Patents on Computer Science, 5:3, 65-176.

8. **Saman Z*** (2011). "Towards the Contribution of NEMDBs in Global Genetic Heterogeneity", Int. Jr. of Emerging Sciences, 1: 3.

Computer Science Oriented, Journal Manuscripts

9. Ahmed Z, and **Saman Z***. (2014) "Measurement, Analysis with Visualization for better Reliability", Book: Artificial Intelligence and Hybrid Systems, ISBN: 978-14775547-3-9. [Book Chapter]

10. Ahmed Z, **Saman Z***. (2014) "Applying WEKA towards Machine Learning with Genetic Algorithm and Back-Propagation Neural Networks", Journal of Data Mining in Genomics & Proteomics.

11. Ahmed Z, **Saman Z***. (2014) "Cultivating Software Solutions Development in the Scientific Academia", F1000 Research, Recent Patents on Computer Science.

12. Ahmed Z, Dandekar T, and **Saman Z*** (2012). "ADAM: Transiting PDM into Clinical Patient Data Management". Int. Jr. of Emerging Sciences, 2:2.

13. Ahmed Z, **Saman Z***, and Dandekar T (2012). "Formal UML Modelling of Isotopo, Bioinformatical Software for Mass Isotopomers Distribution Analysis", Soft. Eng., 2:4.

14. Ahmed Z, **Saman Z***, and Dandekar T (2012). "Unified Modeling and HCI Mockup Designing towards MIDA". Int. Jr. Emerg. Sci. 2:3.

15. Ahmed Z, Dandekar T, and **Saman Z*** (2012). "Semantic web; Ontology Specific Languages for Web Application Development". Int. Jr. Web App. 4:1.

16. Ahmed Z, and **Saman Z*** (2011). "Middleware Technologies; Chain Web Grid Services". International Journal of Web Applications, 3:4,.

Book

- **Saman Z***. (2010) Interactive Web Graphical User Interface (GUI) For Phylogenetic Trees: Web Graphical User Interface (GUI) for displaying Phylogenetic Trees using Perl, specifically BioPerl programming language. ISBN: 978-3838385174. [Book].

Conference Paper

- **Saman Z***, Mehtab Aziz, M Ansar (2007). "GDP-A Database of Genetic Diseases in Pakistan",   In 1st Symposium on Microbiology and Molecular Genetics on Genomics, Proteomics, Metabolomics: Recent Trends in Biotechnology.

Abstracts and Poster Papers

- Ahmed Z, **Saman Z**, Eisenreich W., Dandekar T. (2012): "Isotopo towards Quantitative Mass Isotopomers Distribution Analysis using Spectral Data", The 9th Symposium on the Practical Applications of Mass Spectrometry in the Biotechnology Industry, F1000Posters 4:1422, USA, 11-14 September 2012. [Abstract, Poster]

- Ahmed Z, **Saman Z**, 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), USA , 2nd February 2012. [Abstract, Poster]

- Ahmed Z, **Saman Z**, 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, F1000Posters 4:951, Luxembourg, 26 September 2011. [Abstract, Poster].

10 Appendix

A. Appendices

A.1 Most Related Scientific Publications

A.1.1 Manuscript 1

Dandekar T., Astrid F., **Saman Majeed**, *and Ahmed Z. (2014).* **"Software Applications toward Quantitative Metabolic Flux Analysis and Visualization".** *Briefings in Bioinformatics, Oxford Journals, Oxford University Press, Vol. 15, No.1.*

Software applications toward quantitative metabolic flux analysis and modeling

Thomas Dandekar, Astrid Fieselmann, Saman Majeed and Zeeshan Ahmed

Abstract

Metabolites and their pathways are central for adaptation and survival. Metabolic modeling elucidates *in silico* all the possible flux pathways (flux balance analysis, FBA) and predicts the actual fluxes under a given situation, further refinement of these models is possible by including experimental isotopologue data. In this review, we initially introduce the key theoretical concepts and different analysis steps in the modeling process before comparing flux calculation and metabolite analysis programs such as C13, BioOpt, COBRA toolbox, Metatool, efmtool, FiatFlux, ReMatch, VANTED, iMAT and YANA. Their respective strengths and limitations are discussed and compared to alternative software. While data analysis of metabolites, calculation of metabolic fluxes, pathways and their condition-specific changes are all possible, we highlight the considerations that need to be taken into account before deciding on a specific software. Current challenges in the field include the computation of large-scale networks (in elementary mode analysis), regulatory interactions and detailed kinetics, and these are discussed in the light of powerful new approaches.

Keywords: metabolism; flux modes; metabolites; metabolic network

THEORETICAL BACKGROUND: FLUXES, NETWORKS AND METABOLITES

Metabolic flux analysis is currently a key method in metabolic modeling, its application has provided important insights into metabolism and adaptation of different organisms [1, 2]. Metabolic modeling is a broad field and involves a number of specific tasks. Here, we describe metabolite flux modeling tools and also the incorporation of isotopologue data. While textbooks describe the central metabolic pathways such as glycolysis in form of linear and distinct pathways, it is clear that such linear pathways are embedded in comprehensive metabolic networks. Therefore, the design of specific algorithms is required to calculate all possible pathways in the context of a complex network.

Before discussing and comparing individual software solutions, we briefly introduce some of the theoretical concepts underlying such modeling work: (i) First the network of participating metabolites and enzymes has to be established, this is referred to as the metabolic reconstruction. (ii) Then the network can be analyzed in terms of structure, resources and adaptability. (iii) Next, a more dynamic step is modeling of metabolic fluxes according to the enzyme repertoire as well as the modeling of the metabolic adaptation to different external constraints. (iv) The incorporation of experimental data such as isotopologue data for validation and refinement of the calculated models and fluxes is important. (v) Finally, detailed dynamic analyses of enzymes, such as metabolic control coefficients or detailed modeling of metabolic subnetworks including

Corresponding author. Thomas Dandekar, Department of Bioinformatics, Biocenter, University of Wüerzburg, Am Hubland, 97074 Wuerzburg, Germany. Tel.: +49-931-318-4551; Fax: +49-931-318-4552; E-mail: dandekar@biozentrum.uni-wuerzburg.de
Thomas Dandekar is the chair of the Department of Bioinformatics at the University of Würzburg.
Astrid Fieselmann is a doctoral student in the Department of Bioinformatics at the University of Würzburg.
Saman Majeed is a doctoral student in the Department of Bioinformatics at the University of Würzburg.
Zeeshan Ahmed is a scientific software engineer and doctoral student in the Department of Bioinformatics at the University of Würzburg.

A.1.2 Manuscript 2

*Ahmed Z., **Saman Zeeshan**, Huber C., Hensel M., Schomburg D., Münch R., Eylert E., Eisenreich. W, Dandekar T. (2014). **"'Isotopo' Database and Tool for Facile Analysis and Management of Mass Isotopomer Data",** Database: The Journal of Biological Databases and Curation, Oxford Journals, Oxford University Press, Vol. 2104.*

Database tool

'Isotopo' a database application for facile analysis and management of mass isotopomer data

Zeeshan Ahmed[1,2], Saman Zeeshan[1,3], Claudia Huber[4], Michael Hensel[5], Dietmar Schomburg[6], Richard Münch[7], Eva Eylert[4], Wolfgang Eisenreich[4] and Thomas Dandekar[1,8,*]

[1]Department of Bioinformatics, Biocenter, University of Würzburg, Am Hubland, 97074 Wuerzburg, Germany, [2]Department of Neurobiology and Genetics, Biocenter, University of Wuerzburg, Am Hubland, 97074 Wuerzburg, Germany, [3]Institute of Molecular and Translational Therapeutic Strategies, OE 8886, Hannover Medical School, Carl-Neuberg-Str. 1, D-30625 Hanover, Germany, [4]Lehrstuhl für Biochemie, Center of Isotopologue Profiling, Lichtenbergstraße 4, Technische Universität München, D-85747 Garching, Germany, [5]Division of Microbiology, Barbarastraße 11, Gebäude 36, University of Osnabrück, 49076 Osnabrück, Germany, [6]Department of Bioinformatics and Biochemistry, Langer Kamp 19B, Technical University Braunschweig, D-38106 Braunschweig, Germany, [7]Institute for Microbiology, Biozentrum, 2. Obergeschoss Spielmannstraße 7, Technical University Braunschweig, 38106 Braunschweig, Germany and [8]Computational biology and structures program, European Molecular Biology Laboratory, Meyerhofstr. 1, 69117 Heidelberg, Germany

*Corresponding author: Tel: ++49-931-318-4551; Fax: ++49-931-318-4552; Email: dandekar@biozentrum.uni-wuerzburg.de

Abstract

The composition of stable-isotope labelled isotopologues/isotopomers in metabolic products can be measured by mass spectrometry and supports the analysis of pathways and fluxes. As a prerequisite, the original mass spectra have to be processed, managed and stored to rapidly calculate, analyse and compare isotopomer enrichments to study, for instance, bacterial metabolism in infection. For such applications, we provide here the database application 'Isotopo'. This software package includes (i) a database to store and process isotopomer data, (ii) a parser to upload and translate different data formats for such data and (iii) an improved application to process and convert signal intensities from mass spectra of ^{13}C-labelled metabolites such as tertbutyldimethylsilyl-derivatives of amino acids. Relative mass intensities and isotopomer distributions are calculated applying a partial least square method with iterative refinement for high precision data. The data output includes formats such as graphs for overall enrichments in amino acids.

Appendix Page: iv

A.1.3   Manuscript 3

Ahmed Z., **_Saman Zeeshan_***, Huber C., Hensel M., Schomburg D., Münch R., Eisenreich. W,*

Dandekar T. (2013). ***"Software LS-MIDA for efficient Mass Isotopomer Distribution Analysis".***

BMC Bioinformatics, BioMed Central Publishers, Vol. 14, No. 218.

BMC Bioinformatics

SOFTWARE   Open Access

Software LS-MIDA for efficient mass isotopomer distribution analysis in metabolic modelling

Zeeshan Ahmed[1,7†], Saman Zeeshan[1,8†], Claudia Huber[5], Michael Hensel[2], Dietmar Schomburg[3], Richard Münch[4], Wolfgang Eisenreich[5] and Thomas Dandekar[1,6*]

Abstract

Background: The knowledge of metabolic pathways and fluxes is important to understand the adaptation of organisms to their biotic and abiotic environment. The specific distribution of stable isotope labelled precursors into metabolic products can be taken as fingerprints of the metabolic events and dynamics through the metabolic networks. An open-source software is required that easily and rapidly calculates from mass spectra of labelled metabolites, derivatives and their fragments global isotope excess and isotopomer distribution.

Results: The open-source software "Least Square Mass Isotopomer Analyzer" (LS-MIDA) is presented that processes experimental mass spectrometry (MS) data on the basis of metabolite information such as the number of atoms in the compound, mass to charge ratio (m/e or m/z) values of the compounds and fragments under study, and the experimental relative MS intensities reflecting the enrichments of isotopomers in ^{13}C- or ^{15}N-labelled compounds, in comparison to the natural abundances in the unlabelled molecules. The software uses Brauman's least square method of linear regression. As a result, global isotope enrichments of the metabolite or fragment under study and the molar abundances of each isotopomer are obtained and displayed.

Conclusions: The new software provides an open-source platform that easily and rapidly converts experimental MS patterns of labelled metabolites into isotopomer enrichments that are the basis for subsequent observation-driven analysis of pathways and fluxes, as well as for model-driven metabolic flux calculations.

Background

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 has been established to predict metabolic fluxes [1-4]. Implementing different mathematical parallel and sequential algorithms, several desktop and web based batch and interactive software applications [5] have been also developed towards quantitative metabolic flux analysis and modeling [6].

In contrast, only few methods allow direct determination of metabolic fluxes, one of which is based on *in vivo* experiments using stable isotope labelled 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 signals (typically of silylated derivatives, metabolites and fragments thereof in GC/MS experiments) have to be converted into relative and molar isotopomer abundances.

Isotopologues are species of a compound that differ only in their isotopic composition [7]. 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

* Correspondence: dandekar@biozentrum.uni-wuerzburg.de
†Equal contributors
¹Department of Bioinformatics, Biocenter, University of Würzburg, Würzburg, Germany
⁶EMBL, Structural and Computational Biology Unit, Heidelberg, Germany
Full list of author information is available at the end of the article

A.1.4   Manuscript 4

*Ahmed Z, Michel M., <u>**Saman Z.,**</u> Dandekar T., Mueller M. J., Fekete A. (2014)**. "Lipid-Pro: A computational lipid identification solution for untargeted lipidomics on data-independent acquisition tandem mass spectrometry platforms",** <u>Bioinformatics. Oxford Journals, Oxford University Press</u>. [Accept with Revision]

Lipid-Pro: A computational lipid identification solution for untargeted lipidomics on data-independent acquisition tandem mass spectrometry platforms

Zeeshan Ahmed[1, 2,*], Michel Mayr[3], Saman Zeeshan[2], Thomas Dandekar[2], Martin J. Mueller[3], Agnes Fekete[3]

[1]Department of Neurobiology and Genetics, Biocenter, University of Wuerzburg Germany

[2]Department of Bioinformatics, Biocenter, University of Wuerzburg Germany

[3] Department of Pharmaceutical Biology, Biocenter, University of Wuerzburg Germany

ABSTRACT

Summary:

A major challenge for mass spectrometric-based lipidomics, aiming at describing all lipid species in a biological sample, lies in the computational and bioinformatic processing of the large amount of data that arises after data acquisition. *Lipid-Pro* is a software tool that supports the identification of lipids by interpreting large datasets generated by liquid chromatography - tandem mass spectrometry (LC-MS/MS) using the advanced data-independent acquisition mode MS^E. In the MS^E mode, the instrument fragments all molecular ions generated from a sample and records time-resolved molecular ion data as well as fragment ion data for every detectable molecular ion. *Lipid-Pro* matches the retention time-aligned mass-to-charge ratio data of molecular- and fragment ions with a lipid database and generates a report on all identified lipid species. For generation of the lipid database, *Lipid-Pro* provides a module for construction of lipid species and their fragments using a flexible building block approach. Hence, *Lipid-Pro* is an easy to use analysis tool to interpret complex MS^E lipidomics data and also offers a module to generate a user-specific lipid database.

*To whom correspondence should be addressed.

1

A.1.5 Manuscript 5

Ahmed Z., **<u>Saman Majeed</u>**, *Dandekar T. (2012).* ***"Computational Feature Performance and DSA Evaluation of Applications towards MFA".*** <u>*Recent Patents on Computer Science*</u>, *Vol. 5, No. 3.*

Computational Feature Performance and Domain Specific Architecture Evaluation of Software Applications Towards Metabolic Flux Analysis

Zeeshan Ahmed*, Saman Majeed and Thomas Dandekar

Department of Bioinformatics, Biocenter, University of Wuerzburg, Germany

Abstract: 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 modelling. 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 along with the discussion of few of the patents related to Computational features.

INTRODUCTION

This paper analyses some most recent and previously done well contributing developments from the field of Bioinformatics, especially towards metabolic flux analysis from a computer science point of view. Moreover some domain specific architectures and computational feature performance evaluations are discussed in this paper, in brief.

Metabolic modelling includes metabolic flux analysis as a key method and provides important insights into metabolism and adaptations of different organisms [1, 2]. In the following we compare, analyze and combine available software to obtain suggestions for further development and improvement in this area.

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 [3].

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 can not be reduced or split further without loosing the ability to balance all used internal metabolites, which would then be an elementary mode [4]. Furthermore, a reduced set of such

---
*Address correspondence to this author at the Department of Bioinformatics, Biocenter, University of Wuerzburg, Germany;
E-mail: zeeshan.ahmed@uni-wuerzburg.de

elementary modes is sufficient to reproduce all other modes by linear combination of the modes, a so called convex basis [5].

Modifications and extensions regarding metabolic flux analysis have continuously been proposed (e.g. [6, 7]. 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 [8], Classical and Dynamic FBA [9] etc. Another problem is to map fluxes to actual observed experimental data. For this solutions such as ReMatch [10] 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 [11]. 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 [12]. Furthermore, there are various annotation tools to identify enzymes from the genome sequence. Many groups working in genome-based bioinformatics development to rapidly establish enzyme networks for genome annotation and comparisons (including KEGGbrowser [13]). Hence, the problem is important but too complex for a fair comparison between all available tools. Furthermore, groups involved in major genome annotation efforts established

A.1.6 Manuscript 6

*Ahmed Z, **<u>Saman Zeeshan</u>**, Dandekar T. (2014). **"Developing sustainable software solutions for bioinformatics by the "Butterfly" paradigm",** <u>F1000 Research, Faculty 1000 Publishers</u>, Vol. 3, No. 71.*

METHOD ARTICLE

REVISED

Developing sustainable software solutions for bioinformatics by the "*Butterfly*" paradigm [v2; ref status: indexed, http://f1000r.es/40q]

Zeeshan Ahmed[1,2], Saman Zeeshan[2], Thomas Dandekar[3]

[1]Department of Neurobiology and Genetics, Biocenter, University of Wuerzburg, Wuerzburg, 97074, Germany
[2]Department of Bioinformatics, Biocenter, University of Wuerzburg, Wuerzburg, 97074, Germany
[3]EMBL, Structural and Computational Biology Unit, Heidelberg, 69117, Germany

Abstract

Software design and sustainable software engineering are essential for the long-term development of bioinformatics software. Typical challenges in an academic environment are short-term contracts, island solutions, pragmatic approaches and loose documentation. Upcoming new challenges are big data, complex data sets, software compatibility and rapid changes in data representation. Our approach to cope with these challenges consists of iterative intertwined cycles of development ("*Butterfly*" paradigm) for key steps in scientific software engineering. User feedback is valued as well as software planning in a sustainable and interoperable way. Tool usage should be easy and intuitive. A middleware supports a user-friendly Graphical User Interface (GUI) as well as a database/tool development independently. We validated the approach of our own software development and compared the different design paradigms in various software solutions.

Open Peer Review

Referee Status: ☑☑

	Invited Referees	
	1	2
REVISED version 2 published 01 Aug 2014	☑	☑
version 1 published 13 Mar 2014	? report	? report

1   **Paul Vauterin**, University of Oxford UK

2   **Wolfgang Mueller**, Heidelberg Institute of Theoretical Studies gGmbH Germany

Discuss this article

Comments (0)

Appendix Page: xii

A.1.7 Manuscript 7

Ahmed Z, Fleischmann P., <u>***Saman Z***</u>*, Roessler W., Dandekar T. (2014)* ***"Ant-App-DB: A Smart Ecological Solution for Monitoring the Arthropods' Activities, Experimental Data Management and Solar Calculations without GPS in Behavioural Field Studies"****.* [Submitted]

1   # Ant-App-DB: A Smart Ecological Solution for Monitoring the
2   # Arthropods' Activities, Experimental Data Management and Solar
3   # Calculations without GPS in Behavioural Field Studies

4

5   **Zeeshan Ahmed[1, 3,]*, Pauline Fleischmann[2]2, Saman Zeeshan[3]3, Wolfgang Rössler[2]4, and Thomas
6   Dandekar[3]5**

7   [1]Department of Neurobiology and Genetics, Biocenter, University of Wuerzburg, Wuerzburg, Germany
8   [2]Behavioral Physiology and Sociobiology, Biocenter, University of Wuerzburg, Wuerzburg, Germany
9   [3]Department of Bioinformatics, Biocenter, University of Wuerzburg, Wuerzburg, Germany

10 *** Correspondence:** Zeeshan Ahmed, Department of Neurobiology and Genetics, Biocenter, University of Wuerzburg,
11 Wuerzburg, 97074, Germany
12 Zeeshan.ahmed@uni-wuerzburg.de

14
15 ## Abstract

16 Field studies on arthropod ecology and behaviour require simple and robust monitoring tools,
17 preferably with direct access to an integrated database. We have developed and here present a database
18 tool allowing smart-phone based monitoring of arthropods. This smart phone application provides an
19 easy solution to collect, manage and process the data in the field which has been a very difficult task
20 for field biologists using traditional methods. To monitor our example species, the desert ant
21 Cataglyphis fortis, we considered behaviour, nest search runs, feeding habits and path segmentations
22 including detailed information on solar position and Azimuth calculation, ant orientation and daytime.
23 For this we established a user friendly, embedded and biological database system integrating the Ant-
24 App-DB with a smart phone and tablet application combining experimental data manipulation with
25 data management and providing solar position and timing estimations without any GPS and GIS
26 system. Moreover, the new desktop application Dataplus allows efficient data extraction and
27 conversion from smart phone application to personal computers, for further ecological data analysis
28 and sharing. All features, software code and database as well as Dataplus application are made
29 available completely free of charge and sufficiently generic to be easily adapted to other field
30 monitoring studies on arthropods or other migratory organisms. The software applications Ant-App-
31 DB and Dataplus described here are developed using the Android SDK, Java, XML, C# and SQLite
32 Database.

A.2 Appendix 1

30 highly expressing platelet miRNAs from the set of 94 platelet miRNAs identified by Landry et al., 2009

Highly expressed platelet miRNAs
hsa-let-7d
hsa-let-7f
hsa-let-7g
hsa-miR-18a
hsa-miR-18b
hsa-miR-20a
hsa-miR-27b
hsa-miR-30b
hsa-miR-30d
hsa-miR-30e
hsa-miR-98
hsa-miR-106b

hsa-miR-130a
hsa-miR-130b
hsa-miR-136
hsa-miR-146a
hsa-miR-146b-5p
hsa-miR-151-3p
hsa-miR-151-5p
hsa-miR-155
hsa-miR-330-3p
hsa-miR-335
hsa-miR-374b
hsa-miR-376b
hsa-miR-432
hsa-miR-487b
hsa-miR-551b

miRPlus_17845
miRPlus_17900
miRPlus_17945

A.3   Appendix 2

Protein-Protein interactions for the miRNA target platelet proteins and their association with the platelet network derived from The PlateletWeb

Protein Targets	Interacting Platelet Proteins
IGF1R insulin-like growth factor 1 receptor	ARHGEF12-Rho guanine nucleotide exchange factor (GEF) 12 CAV1-caveolin 1, caveolae protein, 22kDa CRK-v-crk sarcoma virus CT10 oncogene homolog (avian) CRKL-v-crk sarcoma virus CT10 oncogene homolog (avian)-like CSK-c-src tyrosine kinase EEA1-early endosome antigen 1 EHD1-EH-domain containing 1 GIPC1-GIPC PDZ domain containing family, member 1 GNAI1-guanine nucleotide binding protein (G protein), alpha inhibiting activity polypeptide 1 GNAI2-guanine nucleotide binding protein (G protein), alpha inhibiting activity polypeptide 2 GNB2L1-guanine nucleotide binding protein (G protein), beta polypeptide 2-like 1 IGF1-insulin-like growth factor 1 (somatomedin C)

	INSR-insulin receptor
	IRS1-insulin receptor substrate 1
	IRS2-insulin receptor substrate 2
	ITGB1-integrin, beta 1 (fibronectin receptor, beta polypeptide, antigen CD29 includes MDF2, MSK12)
	JAK1-Janus kinase 1
	JAK2-Janus kinase 2
	MAP3K5-mitogen-activated protein kinase kinase kinase 5
	PDPK1-3-phosphoinositide dependent protein kinase-1
	PIK3R1-phosphoinositide-3-kinase, regulatory subunit 1 (alpha)
	PIK3R2-phosphoinositide-3-kinase, regulatory subunit 2 (beta)
	PIK3R3-phosphoinositide-3-kinase, regulatory subunit 3 (gamma)
	PRKCD-protein kinase C, delta
	PRKD1-protein kinase D1
	PTK2-PTK2 protein tyrosine kinase 2
	PTPN1-protein tyrosine phosphatase, non-receptor type 1
	PTPN11-protein tyrosine phosphatase, non-receptor type 11
	PXN-paxillin

	RASA1-RAS p21 protein activator (GTPase activating protein) 1
	SHC1-SHC (Src homology 2 domain containing) transforming protein 1
	SNAP29-synaptosomal-associated protein, 29kDa
	SRC-v-src sarcoma (Schmidt-Ruppin A-2) viral oncogene homolog (avian)
	STAT3-signal transducer and activator of transcription 3 (acute-phase response factor)
	VAV3-vav 3 guanine nucleotide exchange factor
	YWHAB-tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein, beta polypeptide
	YWHAE-tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein, epsilon polypeptide
	YWHAG-tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein, gamma polypeptide
	YWHAZ-tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein, zeta polypeptide
TGFBR1 transforming growth factor, beta receptor 1	ACVR1-activin A receptor, type I AP2B1-adaptor-related protein complex 2, beta 1 subunit

	ARHGAP15-Rho GTPase activating protein 15
	ARL8B-ADP-ribosylation factor-like 8B
	BTBD2-BTB (POZ) domain containing 2
	CAV1-caveolin 1, caveolae protein, 22kDa
	CD44-CD44 molecule (Indian blood group)
	CDK14-cyclin-dependent kinase 14
	CDK17-cyclin-dependent kinase 17
	CDK4-cyclin-dependent kinase 4
	CDK6-cyclin-dependent kinase 6
	CHUK-conserved helix-loop-helix ubiquitous kinase
	CLU-clusterin
	CSNK2A2-casein kinase 2, alpha prime polypeptide
	CTNNB1-catenin (cadherin-associated protein), beta 1, 88kDa
	DAB2-disabled homolog 2, mitogen-responsive phosphoprotein (Drosophila)
	EEF1A1-eukaryotic translation elongation factor 1 alpha 1
	ENG-endoglin
	FBXL12-F-box and leucine-rich repeat protein 12
	FKBP1A-FK506 binding protein 1A, 12kDa
	FNTA-farnesyltransferase, CAAX box, alpha

	GNA13-guanine nucleotide binding protein (G protein), alpha 13
	GNB2-guanine nucleotide binding protein (G protein), beta polypeptide 2
	GNB3-guanine nucleotide binding protein (G protein), beta polypeptide 3
	HSP90AA1-heat shock protein 90kDa alpha (cytosolic), class A member 1
	IKBKB-inhibitor of kappa light polypeptide gene enhancer in B-cells, kinase beta
	ITK-IL2-inducible T-cell kinase
	LIMS1-LIM and senescent cell antigen-like domains 1
	MAP3K7-mitogen-activated protein kinase kinase kinase 7
	MAPK14-mitogen-activated protein kinase 14
	MYO3A-myosin IIIA
	NEK6-NIMA (never in mitosis gene a)-related kinase 6
	NUP37-nucleoporin 37kDa
	OXSR1-oxidative-stress responsive 1
	PAK1-p21 protein (Cdc42/Rac)-activated kinase 1
	PIK3R1-phosphoinositide-3-kinase, regulatory subunit 1 (alpha)
	PIK3R2-phosphoinositide-3-kinase, regulatory subunit 2 (beta)

	PLEK-pleckstrin
	PPP1CA-protein phosphatase 1, catalytic subunit, alpha isozyme
	PPP2R2A-protein phosphatase 2, regulatory subunit B, alpha
	PPP2R2B-protein phosphatase 2, regulatory subunit B, beta
	PPP3CC-protein phosphatase 3, catalytic subunit, gamma isozyme
	PPP6C-protein phosphatase 6, catalytic subunit
	PREB-prolactin regulatory element binding
	PRPF4-PRP4 pre-mRNA processing factor 4 homolog (yeast)
	RAB13-RAB13, member RAS oncogene family
	RAB25-RAB25, member RAS oncogene family
	RAB33B-RAB33B, member RAS oncogene family
	RAB34-RAB34, member RAS oncogene family
	RAB38-RAB38, member RAS oncogene family
	RAB3B-RAB3B, member RAS oncogene family
	RAB6B-RAB6B, member RAS oncogene family
	RAN-RAN, member RAS oncogene family

	RAP2A-RAP2A, member of RAS oncogene family
	RGS19-regulator of G-protein signaling 19
	RHOA-ras homolog gene family, member A
	RHOG-ras homolog gene family, member G (rho G)
	RNF130-ring finger protein 130
	RPS27A-ribosomal protein S27a
	SKAP2-src kinase associated phosphoprotein 2
	SMAD2-SMAD family member 2
	SMAD3-SMAD family member 3
	SNX6-sorting nexin 6
	STRAP-serine/threonine kinase receptor associated protein
	STX8-syntaxin 8
	TGFB1-transforming growth factor, beta 1
	TGFB2-transforming growth factor, beta 2
	TGFBR2-transforming growth factor, beta receptor II (70/80kDa)
	TRAP1-TNF receptor-associated protein 1
	TTC1-tetratricopeptide repeat domain 1
	TTC27-tetratricopeptide repeat domain 27
	UBA52-ubiquitin A-52 residue ribosomal protein fusion product 1
	UBB-ubiquitin B

	UBE2E3-ubiquitin-conjugating enzyme E2E 3 (UBC4/5 homolog, yeast) UBXN1-UBX domain protein 1 WDR13-WD repeat domain 13 WDR61-WD repeat domain 61 WWOX-WW domain containing oxidoreductase
IGF1 insulin-like growth factor 1	IDE-insulin-degrading enzyme IGF1R-insulin-like growth factor 1 receptor IGFALS-insulin-like growth factor binding protein, acid labile subunit IGFBP4-insulin-like growth factor binding protein 4 IGSF1-immunoglobulin superfamily, member 1 TF-transferrin

A.4   Appendix 3

Proteins involved in cross linking pathway

Network Proteins Linking Proteins
IGF1 RAF1
IGF1R BAD
TGFBR1 PINCH (LIMS1)
IRS1
IRS2
SHC2
AKT1
GSK3A
GSK3B
CASP9
GRB2
RAC1
RHOA
PAK1
PAK2
PAK4

ROCK1	
NCK1	
ILK	
P2RY12	
CTNNB1	
PIK3CA	
RASA1	
PTPN11	

A.5 Appendix-4

Platelet protein targets for the selected miRNAs in mice

Highly expressed platelet miRNAs	Platelet Targets (Previous study)	Platelet Targets (New study)	Mouse Targets mmu-let mmu-miR	Common targets
hsa-let-7d	IGF-1R TGFβ-R1	DTX2 NAP1L1 C3orf64 ACSL6 AGPAT6 RNF170 SLC35D2 SLC22A23 DLGAP4 CPEB4 LRIG2 SMCR8 EIF2C4 USP47 CCNY RAB3GAP2	ENSMUSG0000 0095440 (AC104834.1) ENSMUSG0000 0063972 (NR6A1) ENSMUSG0000 0056758 (HMGA2) ENSMUSG0000 0029467 (ATP2A2) ENSMUSG0000 0046404 (YOD1)	LRIG2 IGF2BP1 (Figure-11)

		ERGIC1	ENSMUSG0000	
		CCNF	0063804	
		INPP5A	(LIN28B)	
		MMP11	ENSMUSG0000	
		EIF4G2	0013415	
			(IGF2BP1)	
			ENSMUSG0000	
		(IGF2BP1)	0062825	
		(ARID3B)	(ACTG1)	
			ENSMUSG0000	
			0032913	
			(LRIG2)	
			ENSMUSG0000	
			0032394	
			(IGDCC3)	
			ENSMUSG0000	
			0075324 (FIGN)	
			ENSMUSG0000	
			0021952	
			(XPO4)	

			ENSMUSG0000 0015882 (LCORL) ENSMUSG0000 0024242 (MAP4K3) ENSMUSG0000 0030268 (BCAT1) ENSMUSG0000 0079259 (TRIM71) ENSMUSG0000 0040969 (D630013G24RI K) ENSMUSG0000 0030443 (ZFP583) ENSMUSG0000 0034690 (NLRP4C)	

			ENSMUSG0000 0030814 (BCL7C) ENSMUSG0000 0027660 (SKIL) ENSMUSG0000 0018973 (HOXB1) ENSMUSG0000 0031489 (ADRB3) ENSMUSG0000 0051920 (RSPO2)	
hsa-let-7f	IGF-1R TGFβ-R1	TGFBR1 BZW1 IGF1 EFHD2 ADIPOR2 IGSF1 SCYL3 RUFY3	ENSMUSG0000 0031489 (ADRB3) ENSMUSG0000 0004530 (CORO1C)	BZW1 SLC2A12 DDI2 MAP4K3 COIL BACH1 LIN28B IGDCC3

		ATP2B4	ENSMUSG0000	PLEKHG6
		RAB8B	0095440	SKIL
		ZBTB5	(AC104834.1)	LRIG2
			ENSMUSG0000	YOD1
		DDI2	0052557 (GAN)	FIGN
		MAP4K3	ENSMUSG0000	IGF2BP1
		COIL	0026043	HOXB1
		SLC2A12	(COL3A1)	CPEB1
		RP5-1022P6.2	ENSMUSG0000	NFAT5
			0046404	NR6A1
		ENSG0000015627	(YOD1)	TRIM71
		3 (BACH1)	ENSMUSG0000	COL1A2
		ENSG0000018777	0078515 (DDI2)	(Figure-12)
		2 (LIN28B)	ENSMUSG0000	
		ENSG0000017449	0063972	
		8 (IGDCC3)	(NR6A1)	
		ENSG0000000832	ENSMUSG0000	
		3 (PLEKHG6)	0056758	
		ENSG0000013531	(HMGA2)	
		6 (SYNCRIP)	ENSMUSG0000	
		ENSG0000013660	0032913	
		3 (SKIL)	(LRIG2)	

		ENSG0000007562 4 (ACTB)	ENSMUSG0000 0040969 (D630013G24RI K)
		ENSG0000015646 6 (GDF6)	ENSMUSG0000 0066224 (ARID3C)
		ENSG0000019879 9 (LRIG2)	ENSMUSG0000 0063804 (LIN28B)
		ENSG0000018066 7 (YOD1)	
		ENSG0000018226 3 (FIGN)	
		ENSG0000015921 7 (IGF2BP1)	ENSMUSG0000 0029635 (CDK8)
		ENSG0000012105 8 (COIL)	ENSMUSG0000 0030443 (ZFP583)
		ENSG0000011088 8 (CAPRIN2)	
		ENSG0000016645 0 (PRTG)	ENSMUSG0000 0025586 (CPEB1)
		ENSG0000023669 9 (ARHGEF38)	ENSMUSG0000 0045730 (ADRB2)
		ENSG0000012009 4 (HOXB1)	

		ENSG0000021457 5 (CPEB1)	ENSMUSG0000 0013415 (IGF2BP1)
		ENSG0000010290 8 (NFAT5)	ENSMUSG0000 0021754
		ENSG0000001156 6 (MAP4K3)	(MAP3K1)
		ENSG0000017936 1 (ARID3B)	ENSMUSG0000 0032394
		ENSG0000014820 0 (NR6A1)	(IGDCC3) ENSMUSG0000 0050966
		ENSG0000008875 6 (ARHGAP28)	(LIN28A)
		ENSG0000021437 6 (VSTM5)	ENSMUSG0000 0024242
		ENSG0000020655 7 (TRIM71)	(MAP4K3) ENSMUSG0000 0030268
		ENSG0000018905 0 (RNFT1)	(BCAT1)
		ENSG0000016469 2 (COL1A2)	ENSMUSG0000 0079259
		ENSG0000017306 5 (FAM222B)	(TRIM71)

			ENSMUSG0000 0033983 (COIL)	
			ENSMUSG0000 0027660 (SKIL)	
			ENSMUSG0000 0018973 (HOXB1)	
			ENSMUSG0000 0027547 (SALL4)	
			ENSMUSG0000 0026885 (TTLL11)	
			ENSMUSG0000 0029661 (COL1A2)	
			ENSMUSG0000 0062825 (ACTG1)	
			ENSMUSG0000 0075324 (FIGN)	

			ENSMUSG0000 0000126 (WNT9A)	
			ENSMUSG0000 0037335 (HAND1)	
			ENSMUSG0000 0024817 (UHRF2)	
			ENSMUSG0000 0001379 (APBB3)	
			ENSMUSG0000 0025612 (BACH1)	
			ENSMUSG0000 0039959 (HIP1)	
			ENSMUSG0000 0047497 (ADAMTS12)	

			ENSMUSG0000 0020333 (ACSL6) ENSMUSG0000 0039728 (SLC6A5) ENSMUSG0000 0042942 (GREB1L) ENSMUSG0000 0038167 (PLEKHG6) ENSMUSG0000 0029287 (TGFBR3) ENSMUSG0000 0003847 (NFAT5) ENSMUSG0000 0019796 (LRP11)	

			ENSMUSG0000	
			0033257	
			(TTLL4)	
			ENSMUSG0000	
			0022667	
			(CD200R1)	
			ENSMUSG0000	
			0026834	
			(ACVR1C)	
hsa-let-7g	IGF-1R	IGF2BP1	ENSMUSG0000	COL1A2
	TGFβ-R1	LIN28B	0031489	MAP4K3
		HMGA2	(ADRB3)	LIN28B
		LRIG3	ENSMUSG0000	IGDCC3
		FIGNL2	0004530	PLEKHG6
		TRIM71	(CORO1C)	SKIL
		IGDCC3	ENSMUSG0000	LRIG2
		YOD1	0095440	YOD1
		ARHGAP28	(AC104834.1)	FIGN
		COL1A2	ENSMUSG0000	COIL
		MAP4K3	0052557 (GAN)	HOXB1
		LPGAT1		CPEB1
		STARD13		MAP4K3

		C14orf28	ENSMUSG0000 0046404 (YOD1)	NR6A1 TRIM71 (Figure-13)
		ENSG0000018777 2 (LIN28B)	ENSMUSG0000 0078515 (DDI2)	
		ENSG0000017449 8 (IGDCC3)	ENSMUSG0000 0063972	
		ENSG0000000832 3 (PLEKHG6)	(NR6A1)	
		ENSG0000013531 6 (SYNCRIP)	ENSMUSG0000 0056758 (HMGA2)	
		ENSG0000013660 3 (SKIL)	ENSMUSG0000 0032913	
		ENSG0000007562 4 (ACTB)	(LRIG2)	
		ENSG0000015320 7 (AHCTF1)	ENSMUSG0000 0040969 (D630013G24RI	
		ENSG0000015646 6 (GDF6)	K)	
		ENSG0000019879 9 (LRIG2)	ENSMUSG0000 0066224 (ARID3C)	

		ENSG0000018066 7 (YOD1)	ENSMUSG0000 0063804 (LIN28B)	
		ENSG0000018226 3 (FIGN)	ENSMUSG0000 0029635 (CDK8)	
		ENSG0000015921 7 (IGF2BP1)	ENSMUSG0000 0030443 (ZFP583)	
		ENSG0000012105 8 (COIL)	ENSMUSG0000 0025586 (CPEB1)	
		ENSG0000011088 8 (CAPRIN2)	ENSMUSG0000 0045730 (ADRB2)	
		ENSG0000016645 0 (PRTG)	ENSMUSG0000 0029287 (TGFBR3)	
		ENSG0000013095 8 (SLC35D2)	ENSMUSG0000 0032394 (IGDCC3)	
		ENSG0000023669 9 (ARHGEF38)		
		ENSG0000012009 4 (HOXB1)		
		ENSG0000021457 5 (CPEB1)		
		ENSG0000001156 6 (MAP4K3)		

		ENSG0000025852 9 (RP11-108O10.8)	ENSMUSG0000 0050966 (LIN28A)	
		ENSG0000017936 1 (ARID3B)	ENSMUSG0000 0024242 (MAP4K3)	
		ENSG0000014820 0 (NR6A1)	ENSMUSG0000 0030268 (BCAT1)	
		ENSG0000008875 6 (ARHGAP28)	ENSMUSG0000 0079259 (TRIM71)	
		ENSG0000021437 6 (VSTM5)	ENSMUSG0000 0033983 (COIL)	
		ENSG0000020655 7 (TRIM71)	ENSMUSG0000 0027660 (SKIL)	
		ENSG0000018905 0 (RNFT1)	ENSMUSG0000 0018973 (HOXB1)	
		ENSG0000016469 2 (COL1A2)	ENSMUSG0000 0027547 (SALL4)	
		ENSG0000017306 5 (FAM222B)		

			ENSMUSG0000 0026885 (TTLL11) ENSMUSG0000 0029661 (COL1A2) ENSMUSG0000 0062825 (ACTG1) ENSMUSG0000 0075324 (FIGN) ENSMUSG0000 0000126 (WNT9A) ENSMUSG0000 0024817 (UHRF2) ENSMUSG0000 0041488 (STX3) ENSMUSG0000 0001379 (APBB3)	

			ENSMUSG0000 0025612 (BACH1) ENSMUSG0000 0039959 (HIP1) ENSMUSG0000 0047497 (ADAMTS12) ENSMUSG0000 0026043 (COL3A1) ENSMUSG0000 0055994 (NOD2) ENSMUSG0000 0020333 (ACSL6) ENSMUSG0000 0039728 (SLC6A5)	

			ENSMUSG0000 0042942 (GREB1L) ENSMUSG0000 0038167 (PLEKHG6) ENSMUSG0000 0066892 (FBXL12) ENSMUSG0000 0021754 (MAP3K1) ENSMUSG0000 0003847 (NFAT5) ENSMUSG0000 0019796 (LRP11) ENSMUSG0000 0033257 (TTLL4)	

			ENSMUSG0000 0022667 (CD200R1) ENSMUSG0000 0037335 (HAND1) ENSMUSG0000 0026834 (ACVR1C) ENSMUSG0000 0022514 (IL1RAP)	
hsa-miR-98	IGF-1R TGFβ-R1	ENSG0000015627 3 (BACH1) ENSG0000018777 2 (LIN28B) ENSG0000017449 8 (IGDCC3) ENSG0000000832 3 (PLEKHG6) ENSG0000013531 6 (SYNCRIP)	ENSMUSG0000 0034751 (MAST4) ENSMUSG0000 0037234 (HOOK3) ENSMUSG0000 0071226 (CECR2)	No common targets

		ENSG0000014047 4 (ULK3)	ENSMUSG0000 0052144 (PPP4R2)	
		ENSG0000013660 3 (SKIL)	ENSMUSG0000 0019841 (REV3L)	
		ENSG0000007562 4 (ACTB)	ENSMUSG0000 0022306 (ZFPM2)	
		ENSG0000015646 6 (GDF6)	ENSMUSG0000 0040865 (INO80D)	
		ENSG0000019879 9 (LRIG2)	ENSMUSG0000 0036438 (CALM2)	
		ENSG0000018066 7 (YOD1)	ENSMUSG0000 0051674 (DCUN1D4)	
		ENSG0000018226 3 (FIGN)	ENSMUSG0000 0034518 (HMGXB4)	
		ENSG0000015921 7 (IGF2BP1)		
		ENSG0000004734 6 (FAM214A)		
		ENSG0000011088 8 (CAPRIN2)		
		ENSG0000016645 0 (PRTG)		

		ENSG0000013095 8 (SLC35D2)	ENSMUSG0000 0028248	
		ENSG0000012009 4 (HOXB1)	(SFRS18) ENSMUSG0000	
		ENSG0000021457 5 (CPEB1)	0028282 (CASP8AP2)	
		ENSG0000010290 8 (NFAT5)	ENSMUSG0000 0032030	
		ENSG0000001156 6 (MAP4K3)	(CUL5)	
		ENSG0000017936 1 (ARID3B)		
		ENSG0000014820 0 (NR6A1)		
		ENSG0000008875 6 (ARHGAP28)		
		ENSG0000021437 6 (VSTM5)		
		ENSG0000020655 7 (TRIM71)		
		ENSG0000018905 0 (RNFT1)		

		ENSG0000016469 2 (COL1A2) ENSG0000017306 5 (FAM222B)		
hsa-miR-130a	IGF-1	ENSG0000018438 4 (MAML2) ENSG0000018374 8 (MRC1L1) ENSG0000017152 2 (PTGER4) ENSG0000019632 3 (ZBTB44) ENSG0000019881 8 (SFT2D1) ENSG0000017712 5 (ZBTB34) ENSG0000007240 1 (UBE2D1) ENSG0000022783 9 (AL645730.2) ENSG0000010978 7 (KLF3)	ENSMUSG0000 0033740 (ST18) ENSMUSG0000 0028522 (MIER1) ENSMUSG0000 0025912 (MYBL1) ENSMUSG0000 0033885 (PXK) ENSMUSG0000 0026470 (STX6) ENSMUSG0000 0055717 (SLAIN1) ENSMUSG0000 0027381 (BCL2L11)	SLAIN1 MIER1 (Figure-14)

		ENSG0000020526 8 (PDE7A)	ENSMUSG0000 0032097 (DDX6)	
		ENSG0000017067 7 (SOCS6)	ENSMUSG0000 0054074 (2810030E01RI K)	
		ENSG0000018458 8 (PDE4B)		
		ENSG0000003842 7 (VCAN)	ENSMUSG0000 0042363 (1110067D22RI K)	
		ENSG0000013150 3 (ANKHD1)		
		ENSG0000008791 6 (SLC6A14)	ENSMUSG0000 0038143 (STOX2)	
		ENSG0000016350 7 (KIAA1524)		
		ENSG0000008444 4 (KIAA1467)	ENSMUSG0000 0018736 (NDEL1)	
		ENSG0000015680 4 (FBXO32)	ENSMUSG0000 0022340 (SYBU)	
		ENSG0000016472 9 (SLC35G3)		
		ENSG0000025629 4 (ZNF225)		

		ENSG0000019890 0 (TOP1) ENSG0000011904 2 (SATB2) ENSG0000018256 8 (SATB1) ENSG0000013772 7 (ARHGAP20) ENSG0000018420 6 (RP11- 671M22.1) ENSG0000009090 5 (TNRC6A) ENSG0000015832 1 (AUTS2) ENSG0000012058 6 (MRC1) ENSG0000011092 5 (CSRNP2) ENSG0000010196 6 (XIAP)	ENSMUSG0000 0078308 (GM7293) ENSMUSG0000 0036144 (MEOX2) ENSMUSG0000 0052296 (PPP6R1) ENSMUSG0000 0069808 (FAM57A) ENSMUSG0000 0062866 (PHACTR2) ENSMUSG0000 0029377 (EREG) ENSMUSG0000 0027620 (RBM39)	

		ENSG0000007431 9 (TSG101) ENSG0000010906 1 (MYH1) ENSG0000019700 8 (ZNF138) ENSG0000019803 9 (ZNF273) ENSG0000008671 2 (TXLNG) ENSG0000013973 7 (SLAIN1) ENSG0000019816 0 (MIER1)		
hsa-miR-130b	IGF-1	RBM25 ENSG0000013973 7 (SLAIN1) ENSG0000019816 0 (MIER1)	ENSMUSG0000 0031095 (CUL4B) ENSMUSG0000 0052336 (CX3CR1) ENSMUSG0000 0033740 (ST18)	SLAIN1 MIER1 (Figure-15)

			ENSMUSG0000	
			0028522	
			(MIER1)	
			ENSMUSG0000	
			0043991	
			(PURA)	
			ENSMUSG0000	
			0026470 (STX6)	
			ENSMUSG0000	
			0004798	
			(ULK2)	
			ENSMUSG0000	
			0027381	
			(BCL2L11)	
			ENSMUSG0000	
			0025912	
			(MYBL1)	
			ENSMUSG0000	
			0055717	
			(SLAIN1)	

			ENSMUSG0000	
			0032097	
			(DDX6)	
			ENSMUSG0000	
			0042363	
			(1110067D22RI	
			K)	
			ENSMUSG0000	
			0056476	
			(MED12L)	
			ENSMUSG0000	
			0054074	
			(2810030E01RI	
			K)	
			ENSMUSG0000	
			0018736	
			(NDEL1)	
			ENSMUSG0000	
			0052296	
			(PPP6R1)	

			ENSMUSG0000 0069808 (FAM57A) ENSMUSG0000 0078308 (GM7293) ENSMUSG0000 0017831 (RAB5A) ENSMUSG0000 0062866 (PHACTR2) ENSMUSG0000 0022340 (SYBU) ENSMUSG0000 0036144 (MEOX2) ENSMUSG0000 0029377 (EREG)	

			ENSMUSG0000 0033885 (PXK)	