PhD-FSTC-2015-20
The Faculty of Sciences, Technology and Communication

# DISSERTATION

Defense held on 07/05/2015 in Esch-sur-Alzette

to obtain the degree of

## DOCTEUR DE L'UNIVERSITÉ DU LUXEMBOURG

## EN BIOLOGIE

by

Daniel WEINDL
Born on 23 January 1986 in Mühlacker, (Germany)

# NON-TARGETED MASS ISOTOPOLOME ANALYSIS TO GAIN INSIGHTS INTO CANCER CELL METABOLISM

## Dissertation defense committee

Dr. Karsten Hiller, Dissertation Supervisor
*Université du Luxembourg*

Dr. Hannelore Daniel
*Professor, Technische Universität München*

Dr. Rainer Schuhmacher
*Professor, Universität für Bodenkultur Wien*

Dr. Rudi Balling, Chairman
*Professor, Université du Luxembourg*

Dr. Reinhard Schneider, Vice Chairman
*Université du Luxembourg*

University of Luxembourg

Luxembourg Centre for Systems Biomedicine

Dissertation

# Non-targeted mass isotopolome analysis to gain insights into cancer cell metabolism

*Author:*
Daniel Weindl

*Supervisor:*
Dr. Karsten Hiller

### Previous publications

Parts of this dissertation have been previously published and are included in the Appendix where I also indicated my contributions to the individual publications.

### Publications in peer-reviewed journals

- **Weindl D**, Wegner A, Jäger C, Hiller K. Isotopologue ratio normalization for non-targeted metabolomics. *Journal of Chromatography A*, **2015**, *1389*, 112–9.
- Wegner A, Meiser J, **Weindl D**, Hiller K. How metabolites modulate metabolic flux. *Current Opinion in Biotechnology* **2015**, *34*, 16–22.
- Sapcariu SC, Kanashova T, **Weindl D**, Ghelfi J, Dittmar G, Hiller K. Simultaneous extraction of proteins and metabolites from cells in culture. *MethodsX* **2014**, *1*, 74–8.
- Wegner A, **Weindl D**, Jäger C, Sapcariu SC, Dong X, Stephanopoulos G, Hiller K. Fragment Formula Calculator (FFC): Determination of Chemical Formulas for Fragment Ions in Mass Spectrometric Data. *Analytical Chemistry* **2014**, *86*, 2221–8.
- Meiser J[1], **Weindl D**[1], Hiller K. Complexity of Dopamine Metabolism. *Cell Communication and Signaling* **2013**, *11*, 34.
- Wegner A, Sapcariu SC, **Weindl D**, Hiller K. Isotope cluster based compound matching in Gas Chromatography/Mass Spectrometry for non-targeted metabolomics. *Analytical Chemistry* **2013**, 85(8):4030–7.
- Hiller K, Wegner A, **Weindl D**, Cordes T, Metallo CM, Kelleher JK, Stephanopoulos G. NTFD—a stand-alone application for the non-targeted detection of stable isotope-labeled compounds in GC/MS data. *Bioinformatics* **2013**, *29*:1226–8.

### Book chapters

- **Weindl D**, Wegner A, Hiller K. Non-targeted tracer fate detection. *Methods in Enzymology*, 2015, *in press*.

### Manuscripts planned to be submitted for publications in peer-reviewed journals

- **Weindl D**, Hiller K. MIA: Non-targeted mass isotopolome analysis. To be submitted as application note to *Bioinformatics*.
- **Weindl D**, Cordes T, Wegner A, Hiller K. Non-targeted stable isotope labeling analysis. Manuscript in preparation.

[1]Equally contributing authors

## Presentations at scientific conferences

**Poster presentations:**

- **Weindl D**, Wegner A, Jäger C, Hiller K. Non-targeted quantification of metabolites using stable isotope dilution mass spectrometry and non-targeted tracer fate detection. *ISC 2014—30$^{th}$ International Symposium on Chromatography*, **2014**, Salzburg (Austria).

- **Weindl D**, Cordes T, Hiller K. Profiling metabolic networks using stable-isotope tracers and mass spectrometry. *2$^{nd}$ International Systems Biomedicine Symposium*, **2013**, Esch-sur-Alzette (Luxembourg).

- **Weindl D**, Cordes T, Hiller K. Profiling metabolic networks using stable-isotope tracers and mass spectrometry. *Life Sciences PhD Days*, **2013**, Luxembourg (Luxembourg).

- **Weindl D**, Hiller K. Comprehensive profiling of metabolic networks using stable-isotope tracers and mass spectrometry. *II International Workshop on Metabolomics & Proteomics*, 2012, Bilbao (Spain).

- **Weindl D**, Hiller K. Comprehensive profiling of metabolic networks using stable-isotope tracers and mass spectrometry. *Life Sciences PhD Days*, **2012**, Luxembourg (Luxembourg).

# Contents

# List of Figures

# List of Tables

# Acronyms

| | |
|---|---|
| $G_{a,\ldots,n}$ | Isotopologue of a compound containing $N$ possibly labeled atoms with atoms 1 ($a$), ..., $N$ ($n$) labeled (1) or unlabeled (0). Fully labeled isotopologue: $G_{1,\ldots,1}$. |
| A549 | human lung adenocarcinoma cells |
| aa | amino acid(s) |
| ACLY | ATP-dependent citrate lyase |
| ACSS | acyl-CoA synthetase |
| APCI | atmospheric pressure chemical ionization |
| ASPA | aspartoacylase |
| ATP | adenosine triphosphate |
| BCAA | branched-chain amino acid |
| BCAT | branched-chain amino acid aminotransferase |
| BCKDH | branched-chain $\alpha$-keto acid dehydrogenase |
| bp | base pair(s) |
| BPI | base peak intensity |
| C$a$:$b$ | fatty acid with $a$ carbons and $b$ double bonds |
| CI | chemical ionization |
| CoA | coenzyme A |
| csv | comma separated values |
| EI | electron ionization |
| FAD | flavin adenine dinucleotide |
| FAME | fatty acid methyl ester |
| FASN | fatty acid synthase |
| FBA | flux balance analysis |
| FFC | Fragment Formula Calculator |
| GC | gas chromatography |
| GC-MS | gas chromatography - mass spectrometry |

| | |
|---|---|
| GlcNAc | *N*-acetyl-D-glucosamine |
| GNE | UDP-*N*-acetylglucosamine 2-epimerase |
| GPD | glycerol-3-phosphate dehydrogenase |
| GUI | graphical user interface |
| HIF | hypoxia inducible factor |
| ICBM | isotope cluster-based compound matching |
| IDH | isocitrate dehydrogenase |
| IDMS | isotope dilution mass spectrometry |
| InChI | INternational CHemical Identifier |
| IUPAC | International Union of Pure and Applied Chemistry |
| LC-MS | liquid chromatography - mass spectrometry |
| LDH | lactate dehydrogenase |
| $m/z$ | mass-to-charge ratio |
| ManNAc | *N*-acetyl-D-mannosamine |
| MDH | malate dehydrogenase |
| MeOX | methyloxime |
| MFA | metabolic flux analysis |
| $mGluR_3$ | type 3 metabotropic glutamate receptor |
| MIA | Mass Isotopolome Analyzer |
| MID | mass isotopomer distribution |
| MS | mass spectrometry |
| MSTFA | *N*-methyl-*N*-(trimethylsilyl) trifluoroacetamide |
| MTBSTFA | *N*-*tert*-butyldimethylsilyl-*N*-methyltrifluoroacetamide |
| mTORC1 | mammalian/mechanistic target of rapamycin complex 1 |
| $I_n$ | Relative abundance of the raw mass isotopomer with a mass increment of $n$ as compared to the lightest isotopologue (the M+0 peak). |
| NAAGS | *N*-acetylaspartylglutamic acid synthetase |
| NAD | nicotinamide adenine dinucleotide |
| NADP | nicotinamide adenine dinucleotide phosphate |
| NAT8L | *N*-acetyltransferase 8-like |
| NeuNAc | *N*-acetylneuraminic acid |
| NMR | nuclear magnetic resonance |
| NNT | nicotinamide nucleotide transhydrogenase |

| | |
|---|---|
| NTFD | non-targeted tracer fate detection |
| OGDH | 2-oxoglutarate dehydrogenase |
| PCA | principal component analysis |
| PCC | propionyl-CoA carboxylase |
| PDH | pyruvate dehydrogenase |
| PDK | pyruvate dehydrogenase kinase |
| PKA | protein kinase A |
| PPP | pentose phosphate pathway |
| qPCR | quantitative real-time polymerase chain reaction |
| RI | retention index |
| ROS | reactive oxygen species |
| RT | retention time |
| SAM | *S*-adenosyl methionine |
| SD | standard deviation of the mean |
| SDH | succinate dehydrogenase |
| SHMT | serine hydroxymethyl transferase |
| SIM | selected ion monitoring |
| siRNA | small interfering RNA |
| SVG | scalable vector graphics |
| TBDMS | *tert*-butyltrimethylsilyl |
| TCA | tricarboxylic acid |
| THF | tetrahydrofolate |
| TIC | total ion current |
| TMS | trimethylsilyl |
| TOF | time of flight |
| UDP | uridine diphosphate |

# Summary

In this work, a novel computational tool, MIA, for the non-targeted analysis of complex stable isotope labeling gas chromatography electron impact mass spectrometry (GC-EI-MS) datasets is presented. MIA, the *Mass Isotopolome Analyzer*, extends the previously published non-targeted tracer fate detection (NTFD) tool to analyze multiple datasets in parallel. Furthermore, MIA allows not only for the non-targeted detection and quantification of stable isotope labeled compounds, but also for subsequent visual data analysis and for the non-targeted detection of metabolic flux changes based on mass isotopomer distribution (MID) variation analysis. Furthermore, it aids compound identification and pathway contextualization by generating MID similarity-based compound networks.

Applying MIA for the analysis of the metabolism of a human lung adenocarcinoma cell line (A549) allowed for the detection of hypoxia-induced metabolic flux changes and unanticipated metabolites. A549 cells were found to produce *N*-acetylaspartic acid, a metabolite known to be involved in neuronal-oligodendrocytic acetyl-shuttling. Moreover, several *N*-carboxy derivatives of amine metabolites were detected. Hypoxic cancer cells were shown to have an increased branched-chain amino acid (BCAA) catabolism and that the propionyl-coenzyme A (CoA) generated during catabolism of isoleucine gives rise to odd-chain fatty acids.

The observation of *N*-acetylaspartic acid production in A549 cells lead to the discovery of an alternative acetyl-shuttling mechanism across the mitochondrial membrane based on *N*-acetyltransferase 8-like (NAT8L) and *N*-acetylaspartic acid, circumventing ATP-dependent citrate lyase (ACLY) which was yet unknown in cancer cells.

A metabolomics data management platform, MetaboBase, was developed to keep track of tens of thousands of GC-MS measurements. The platform is integrated with data acquisition and preprocessing, and is easily operated via a user-friendly web interface. It allows for a central management of experimental metadata and mass spectra reference libraries which are exported to a format directly accessible with the MetaboliteDetector data analysis software.

In summary, this work illustrates how non-targeted analysis of stable isotope labeling can be a valuable tool for initial hypothesis generation in the study of metabolism. The detected unanticipated metabolites show once more that knowledge on species- and cell type-specific metabolism is still incomplete and that non-targeted approaches are necessary. The developed data analysis tools like MIA can be applied in further studies, also in contexts other than

cancer metabolism. The discovery of *N*-acetylaspartic acid-based acetyl transport within cancer cells may provide new targets for cancer treatment. However, further studies are needed.

# 1 Introduction

## 1.1 Metabolism and metabolic research

Every living organism is exchanging matter with its environment. Organic and inorganic substrates are taken up and converted into biomass or are used to obtain energy. This entirety of uptake, interconversion, and excretion of matter is called *metabolism*. Metabolism has long been studied to understand biological systems in health and disease.

### 1.1.1 Metabolism is strongly regulated

Most metabolic reactions are catalyzed by enzymes, only some occur spontaneously. Enzymes are mostly proteins, which are synthesized by translation of messenger ribonucleic acids (mRNAs) which are transcribed from genes (Figure 1). This gene expression is controlled by proteins and the epigenetic state, which can in turn be regulated by metabolites.[1] Epigenetic modifications, transcription, translation and enzymatic catalysis are strongly regulated and finally give rise to the entirety of metabolites, the *metabolome*. All these levels are strongly interconnected and their interactions are finely tuned but yet highly dynamic to adopt to changing physiological requirements. Together, they determine the final phenotype.

The metabolic network consists of metabolite pools which are interconnected by metabolic reactions or metabolic *fluxes*. These fluxes are even more dynamic and regulated than the metabolome itself. They give rise to the metabolome and are themselves regulated *by* the metabolome. In Wegner *et al.*[2] (section A.6) we review current literature and describe how metabolites regulate metabolic fluxes on all levels shown in Figure 1.

### 1.1.2 Dysregulated metabolism leads to disease

Metabolism is finely regulated by intricate mechanisms and metabolism itself regulates many cellular processes directly or indirectly. Therefore, it comes as no surprise that perturbations of metabolism can lead to disease. Many genetically caused inborn errors of metabolism are known[3] and some of them, like for example Tay-Sachs disease or Canavan's disease have devastating effects. In Tay-Sachs disease, it is a defect in the $\beta$-hexosaminidase gene[4] and in Canavan's disease, it is a defect in the aspartoacylase gene[5] that lead to severe neurological

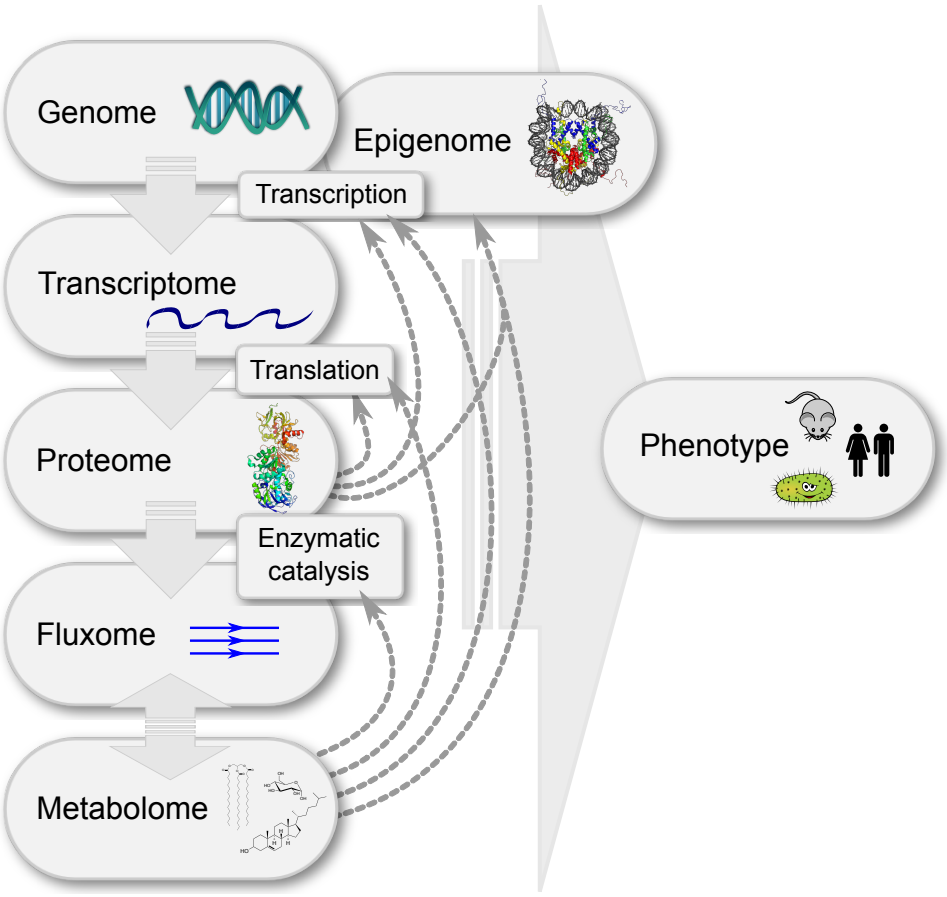



**Figure 1: All omes interact and determine the phenotype.** The fluxome is the most dynamic and probably the closest to the final phenotype.

impairment. Several other well known diseases like diabetes and phenylketonuria (PKU) are caused by pathological metabolism. In Meiser *et al.*[6] (section A.3) we review how dopamine metabolism is involved in the neurodegenerative Parkinson's disease.

### 1.1.3 The importance of metabolic research

Many diseases are caused by, and probably all diseases lead, either directly or indirectly, to altered metabolism. Therefore, metabolic research is highly important for the diagnosis and understanding of disease, and to finally develop therapeutic means to help the patients. Recently, metabolic research has provided good mechanistic insights into cancer metabolism,[7–9] provided diagnostic biomarkers,[10,11] and potential therapeutic targets.[12–15]

Cellular metabolism is most easily studied, but for the understanding of complex diseases, a systems understanding of whole-body metabolism is required. There is much textbook knowledge on metabolism in general,[16,17] but for both ends of the scale, for cellular and for whole-body metabolism, knowledge is still incomplete. There are large gaps in knowledge on subcellular metabolism, cell-type and tissue-specific metabolism and their consequences

for systemic metabolism. This lack of knowledge is to a large extent caused by a lack of appropriate tools to study those aspects. However, subcellular metabolism is attracting more attention,[18] especially with the application of $^{13}$C-metabolic flux analysis (MFA) (section 1.3.8) to mammalian cells.[19,20]

## 1.2 Metabolomics — the study of metabolite levels

Due to the technological progress in analytical instruments over the recent decades, the study of metabolism and the measurement of metabolite levels on a large scale became more and more feasible and is now popular as *metabolomics*. Metabolomics has become an important tool in systems biology[21] and pharmacological research.[22]

Metabolomics measurements can be performed on different analytical platforms. The most prominent are gas chromatography - mass spectrometry (GC-MS), liquid chromatography - mass spectrometry (LC-MS) and nuclear magnetic resonance (NMR), each with their own strengths and weaknesses.[23] Metabolite measurements presented in this dissertation have been performed on GC-MS instruments which I therefore describe in more detail.

### 1.2.1 GC-MS

GC-MS is a commonly used analytical platform in metabolomics. For GC-MS analysis a liquid sample is evaporated and transferred onto a capillary column. This column is coated with a stationary phase, often consisting of polysiloxanes. There, compounds are separated based on their volatility and differential retention through interactions with the stationary phase. Finally, the separated analytes enter the mass spectrometer where they are ionized and analyzed by their mass-to-charge ratio. The advantages of GC-MS are the high peak capacity and high reproducibility. A disadvantage is, that to be analyzed by gas chromatography, samples need to be transferred to the gas phase. As most metabolites are not volatile *per se*, they need to be chemically modified to increase their volatility.

#### Derivatization

In an additional sample preparation step, polar groups are masked to prevent hydrogen bonding which is responsible for the strong intramolecular interactions leading to low volatility. Depending on the nature of the analyte there are many types of derivatizations possible.[24] Very versatile and commonly used methods are derivatization of hydrogen bond donors with silylating agents like *N*-methyl-*N*-(trimethylsilyl) trifluoroacetamide (MSTFA) or *N*-*tert*-butyldimethylsilyl-*N*-methyltrifluoroacetamide (MTBSTFA) (Figure 2A). Reactions with MSTFA and MTBSTFA lead to trimethylsilyl (TMS)- and *tert*-butyltrimethylsilyl (TBDMS)-derivatives respectively.

Carbonyl groups can react with methoxyamine to form methoximes. Methoximation also lowers the ability to form hydrogen bonds, but more importantly, it prevents keto-enol-tautomerism by locking molecules in their keto-form. Methoximation also prevents hemi-acetal formation, and thus, keeps *e.g.* sugars in their open chain form. This reduces the number of different derivatives formed from a single precursor, but the methoxime itself can exist in two different configurations ($E/Z$) (Figure 2B).

Derivatization reactions are often not exhaustive, so that a single analyte can give rise to multiple derivatives. This is the case for primary amines for which often both the 1TMS and the 2TMS derivative is detected.

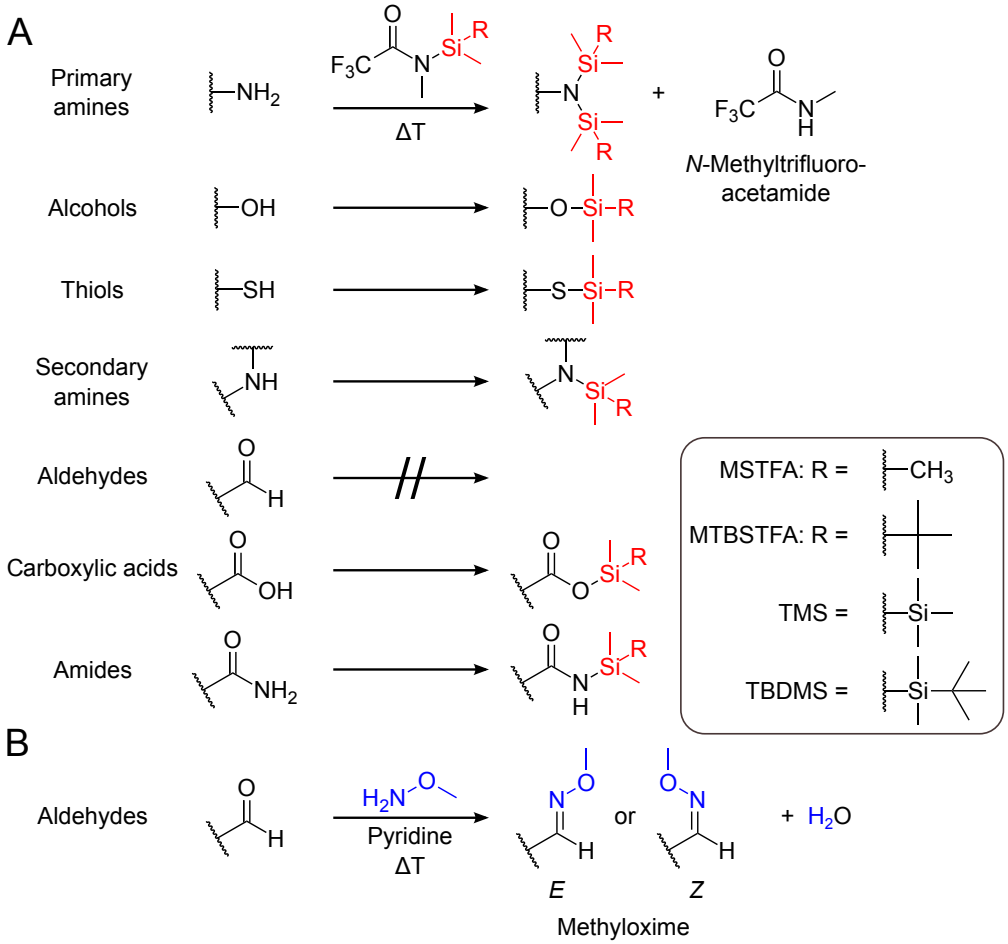



**Figure 2: Possible derivatization reactions for GC-MS sample preparation.** A) Alkylsilylation using M(TB)STFA works on a wide range of functional groups. B) Aldehydes and ketones can be methoximated, preventing keto-enol-tautomerism and acetal formation. Two stereoisomers of the methyloxime can occur.

### 1.2.2 Electron ionization mass spectrometry

For GC-MS multiple ionization techniques can be used: electron ionization (EI), chemical ionization (CI) or atmospheric pressure chemical ionization (APCI). For EI, which is very

commonly used, analytes entering the mass spectrometry (MS) are ionized by electron impact. Electrons emitted from a filament in the ion source of the mass spectrometer are accelerated to hit the analytes and remove an electron. This process leaves a radical cation (the molecular ion, $M^{\bullet+}$) which is usually unstable and undergoes subsequent fragmentation and rearrangement reactions.[25] For that reason EI is called a hard ionization technique. Fragmentation always leads to both, charged and neutral products. While the neutral products are lost, the ions are analyzed and detected according to their mass-to-charge ratio ($m/z$). This can lead to rather complex mass spectra. However, the fragmentation of the analyte is structure-specific and its mass spectrum is like a fingerprint. This allows for an easy identification of compounds by matching them against mass spectral libraries (subsection 1.2.3, section A.1). As metabolite identification is a bottleneck in current metabolomics analyses (subsection 1.2.7), this relatively simple compound identification is one big advantage of EI-MS.

As an example for a typical mass spectrum, the spectrum of the 2TMS derivative of glycine is shown in Figure 3. The molecular ion at $m/z$ 219 fragments quickly and was not detected. The heaviest ion is formed by the loss of a methyl group from one of the TMS groups, a common loss for TMS derivatives. In the lighter fragments, parts of the glycine backbone are lost. The fragment with $m/z$ 147 is formed from all TMS derivatives and contains only a single oxygen atom from the native metabolite. All these fragments show satellite peaks due to incorporation of naturally occurring stable isotopes.

In Wegner *et al.*[26] (section A.4) we present Fragment Formula Calculator (FFC), a tool that uses a combinatorial graph-based approach to determine the elemental composition and potential substructures in such mass spectrometric fragments (Figure 3). These information are valuable in combination with stable isotope labeling, since they can be used to obtain positional information on isotopic enrichment as I show in subsection 3.3.1.

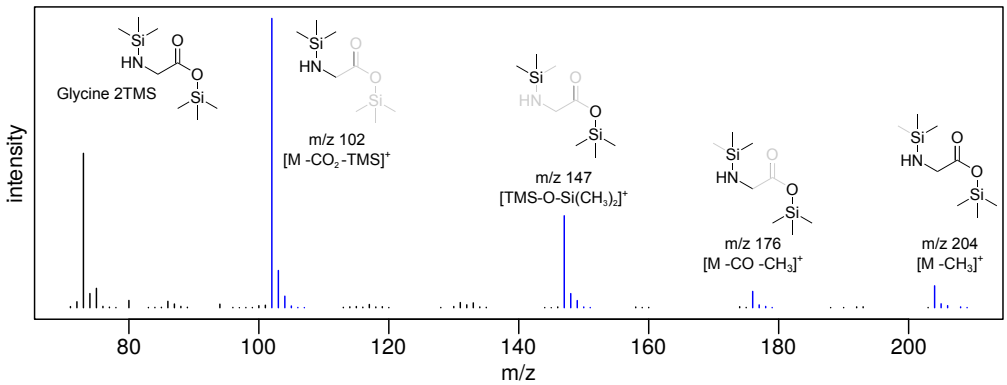



**Figure 3: The EI-MS spectrum of glycine 2TMS.** Substructures retained (black) and lost (grey) in the indicated fragment ions have been determined with FFC (section 3.3). Peak clusters occurring due to natural isotope abundance are colored blue.

### 1.2.3 Compound identification

#### Mass spectrum reference libraries

Compound identification in MS-based metabolomics is usually achieved by matching mass spectra against reference libraries. There are several commercial[27–29] and free[30] reference libraries available. Every mass spectrum of interest is compared to all mass spectra in the reference library and a score is determined based on the spectrum similarity and often also the retention index (RI).[31,32]

Mass spectra from complex metabolite samples are often noisy and require robust spectrum matching algorithms. In Wegner *et al.*[33] (section A.1) we present isotope cluster-based compound matching (ICBM), a novel algorithm that improves this mass spectrum matching by making use of peak patterns arising from natural isotope abundance. Furthermore, ICBM allows a more efficient and faster spectrum matching against large libraries which is otherwise computationally very expensive. It allows to use the commercial NIST library,[28] which currently contains mass spectra of more than 270,000 compounds (Nist14), in routine analyses.

#### Compounds not present in reference libraries

Although there are extensive mass spectrum reference libraries available, it is probable that many metabolites still remain to be discovered and are therefore not present in any of these libraries. Unlike NMR spectroscopy, mass spectrometry cannot be used for unambiguous *ab initio* structure elucidation, but it can provide good hints. Soft ionization techniques like APCI or electrospray ionization (ESI), ideally in combination with high-resolution mass spectrometry, can provide the mass of the molecular ion or even its elemental composition.[34] If the accurate mass alone is ambiguous, certain heuristics can be applied to determine the putative elemental composition.[35] Once the elemental composition is determined, it can be used for further database searches. Furthermore, stable isotope labeling can provide the minimum number of atoms of the isotopic elements in a given compound and can reveal the origin of the given atoms which can be used to rule out certain candidates (section A.7).

Structurally related compounds often yield similar mass spectra with compound class-specific ions or neutral losses.[25,36–38] For example TMS derivatives of carbohydrates can easily be recognized by their characteristic ions.[39,40] Therefore, mass spectrum reference databases can also provide valuable hints for compound identification, even if they do not contain the compound of interest itself.

A problem for compound identification, in particular in GC-MS metabolomics, is that most compounds are not native metabolites but chemical derivatives thereof (section 1.2.1). However, most compound databases only contain the native metabolite but *e.g.* not its

TMS derivative. Therefore, the elemental composition of the native compound needs to be determined. In the case of TMS derivatives the number of TMS groups can to some extent be deduced from the relative abundance of the M+1 and M+2 peaks of the $M^{\bullet+}$ or $[M\text{-}CH_3]^+$ ions which are mostly a result of the high natural isotope abundance of silicon (Table 1) and the high number of introduced carbons. A more reliable approach is the use of stable isotope labeled derivatization agents. For example deuterated MSTFA-$d_9$ is commercially available and the induced mass shift divided by 9 yields the number of TMS groups in the molecule.[41] The mass and composition of these derivatization groups can be subtracted to get to the native mass or elemental composition. Additionally, the number of derivatization sites already provides information on functional groups present in the molecule. Similar tools are available for different derivatization reagents like for example methoxyamine-$d_3$.

### 1.2.4 Quantitative metabolomics and comparability of measurements

In MS-based metabolomics analyses, analyte abundance is quantified by the intensity of the detector signal caused from ions of specific mass-to-charge ratios. Within a limited range, the *linear range*, the detector signal is linearly dependent on the amount of the analyte. Below or above the linear range, the signal is skewed due to a low signal-to-noise ratio, detector memory, or detector saturation (Figure 4A).[42]

The absolute detector signal is furthermore dependent on the analyte, the analytical instrument, its configuration, and the sample matrix (Figure 4B). These factors pose a problem to the comparability of the semi-quantitative raw signal intensities between different analytes and between signal intensities across different measurements.

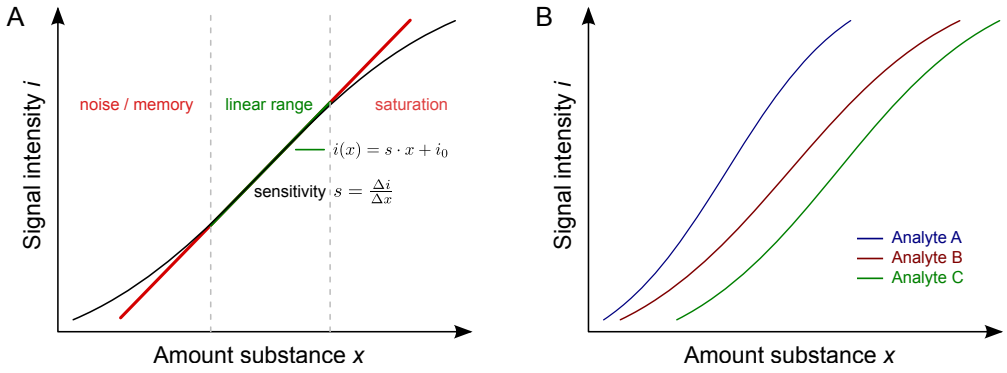



**Figure 4: The MS detector signal as a function of amount of substance.**
A) The detector signal is a function of the amount of substance introduced. Over a limited range there is a linear relationship between the signal intensity $i$ and the amount of substance $x$. Outside this range, the signal is skewed because of detector memory, noise, or saturation.
B) The sensitivity, the change in signal intensity due to changed amount of substance, depends on the analyte and its amount.

### 1.2.5 Absolute quantification

The best means to compare analyte levels is their absolute quantification, *e.g.* to determine their amount in moles. In targeted metabolomics approaches, when all analytes of interest are known beforehand, absolute quantification is possible through external or internal calibration. For external calibration, reference samples containing known quantities of the analyte are measured along with the sample of interest. The measured signal of the known amount of substance can be used to quantify the same analyte in the sample of interest (Figure 5A). This method may not well account for technical variation introduced during sample preparation. That problem is reduced by internal calibration, where a defined amount of one or multiple exogenous compounds, the internal standards, is introduced into the sample, at the earliest possible stage of sample processing. From a measurement of a defined amount of both internal standards and analytes, a response factor $r$ is determined that can be used in subsequent measurements to quantify analytes in the sample of interest (Figure 5B).

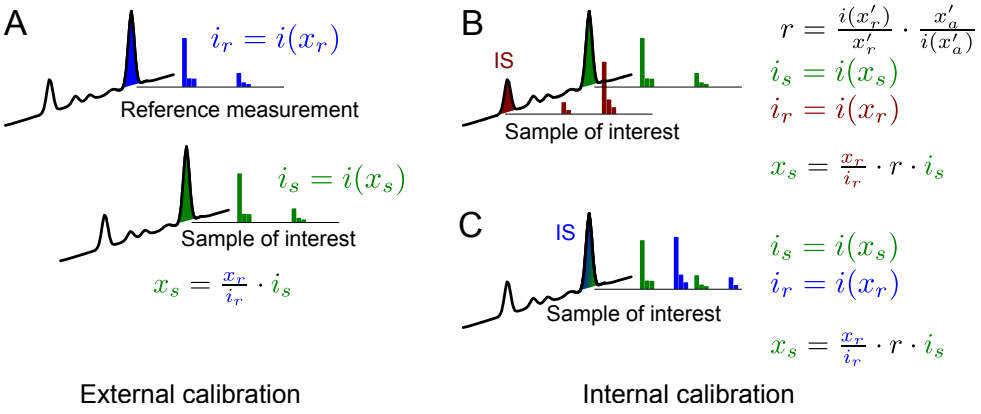



**Figure 5: Quantitative mass spectrometry.**
A) The signal intensity $i_r$ caused by a defined amount of analyte $x_r$ in a reference measurement can be used to quantify the same analyte in the sample of interest ($x_s$).
B,C) A defined amount $x_r$ of an internal standard (IS) can be spiked into the sample as quantification reference. After measurement of a defined amount of the analyte, a compound-specific response factor $r$ can be determined from measurement of a defined amount of the analyte $a'_r$ together with a defined amount of the internal standard $x'_r$, and used for quantification of the compound in the sample of interest ($x_s$). The internal standard can be an isotopologue of the analyte of interest (C) or a different compound with similar physicochemical properties (B).

A single internal standard can be used for all analytes, or group-specific internal standards can be used for analytes of certain compound classes. However, such approaches cannot account for matrix effects or analytical discrimination of individual analytes. Therefore, internal standards for quantification should be as similar as possible in the physicochemical properties to the analyte of interest. To this end, stable isotope labeled analogues of the analyte, which are subject to the same analytical biases, are usually applied as internal standards. The analyte is quantified from the intensity ratio of its labeled and unlabeled

form and a previously determined response factor. Such approaches are referred to as isotope dilution mass spectrometry (IDMS) (Figure 5C).

### 1.2.6 Normalization

In non-targeted metabolomics experiments, many analytes usually remain unidentified. For these compounds no absolute quantification is possible and, therefore, no statements can be made on the relative absolute abundance of different unidentified compounds. Only the relative levels of the same compound across different samples can be compared. This comparison of raw signal intensities of an analyte is usually only possible across a limited number of measurements, because the detector signal induced by a certain amount of substance in the sample can change over time due to *e.g.* different instrument tuning or deteriorating inlet or column performance. Therefore, additional measures need to be taken to allow for a comparison of analyte levels across different measurements from different batches or different instruments.

Several normalization approaches have been applied to account for such variation. The simplest means is total ion current (TIC) normalization, that is the normalization of any measured signal to the overall signal of the sample. However, this approach is sensitive to strong changes in levels of single, high-abundant analytes. Another approach, probabilistic quotient normalization,[43] addresses this problem by determining a most probable dilution factor of a given sample as compared to a representative reference. Such a reference is obtained from pooling a large number of different samples or measurements which is only possible for large studies. Normalization on stable isotope labeled internal standards would avoid these shortcoming, but no defined mixture of internal standards can be provided for unidentified analytes. Metabolite extracts from fully stable isotope labeled organisms have been identified as a source for a large number of internal standards.[44,45] For high resolution LC-MS, tools have been developed to automatically detect these internal standards and to use them as normalization reference.[46,47] However, these tools cannot operate on data from low resolution MS or from hard ionization techniques like EI-MS. The problem is that current approaches detect the analytes and corresponding internal standards from characteristic isotope patterns in the mass spectrum. Since EI mass spectra contain a large number of mass peaks from fragmentation and no accurate mass measurements are available, these fragment peaks cannot easily be distinguished from isotopic peak patterns.

### 1.2.7 Chances and limitations

Metabolite levels can be helpful to detect changes in metabolism and can serve as diagnostic biomarkers.[10,48,49] Yet, metabolite levels alone are often not easy to interpret. To get from

the phenotype to a more mechanistic understanding, it is important to know whether a certain metabolite accumulates due to its increased production or due to its reduced consumption. These cases, increased production or decreased consumption, cannot be distinguished based on metabolite levels alone.

The ultimate goal of metabolomics is to quantify every compound throughout the metabolome. Depending on the analytical platform and the sample type, current routine metabolomics measurements provide information on a few dozens to a few thousands of compounds.[46,50–52] Identification of these compounds is crucial for data interpretation. However, this identification step (subsection 1.2.3) is still a major bottleneck in metabolomics analyses.[53,54] In non-targeted metabolomics data, a large fraction of the detected compounds usually remains unidentified and can therefore only provide limited biological insights.

Another big issue in non-targeted metabolomics is the lack of robust normalization methods that allow for the comparison of measurements from different batches, instruments or laboratories. Since absolute quantification of unidentified compounds is not possible, other approaches are required (subsection 1.2.4). In section 3.4, I describe a method based on IDMS that allows for a more robust normalization of non-targeted metabolomics data.

## 1.3 Stable isotopes & stable isotope-assisted metabolic research

### 1.3.1 Stable isotopes

Isotopes are nuclides of an element with the same number of protons, but varying numbers of neutrons (*e.g.* $^{12}$C and $^{13}$C).[55] There exist stable and non-stable (*i.e.* radioactive) isotopes. In this work, I only consider stable isotopes. Stable isotopes occur naturally, but for most elements they are of very low abundance (Table 1). They were discovered in the beginning of the twentieth century and the term isotope was coined in 1913.[56]

Stable isotopes can be introduced into molecules as chemical labels. Such labeled molecules differ in their masses, but are very similar in their other physical and chemical properties (exceptions in subsection 1.3.3).

### 1.3.2 Stable isotope-assisted metabolic research

Stable isotope labeled compounds can be applied to any organism and they are metabolized almost like their natural occurring analogues. Once these tracers enter metabolism, the isotopes will be incorporated into downstream metabolites and can be detected using mass spectrometry[58] or NMR spectroscopy.[59]

**Table 1: Average natural stable isotope abundance of biologically most relevant elements.**[57] Some elements have none, one, or multiple naturally occurring stable heavy isotopes.

| Nuclide | Molar fraction |
|---|---|
| $^{1}$H | 0.9999 |
| $^{2}$H (D) | 0.0001 |
| $^{12}$C | 0.9893 |
| $^{13}$C | 0.0107 |
| $^{14}$N | 0.9964 |
| $^{15}$N | 0.0036 |
| $^{16}$O | 0.9976 |
| $^{17}$O | 0.0004 |
| $^{18}$O | 0.0021 |

| Nuclide | Molar fraction |
|---|---|
| $^{28}$Si | 0.9222 |
| $^{29}$Si | 0.0468 |
| $^{30}$Si | 0.0309 |
| $^{31}$P | 1 |
| $^{32}$S | 0.9499 |
| $^{33}$S | 0.0075 |
| $^{34}$S | 0.0425 |
| $^{36}$S | 0.0001 |

Since mid of the last century, stable isotopes have been used in a biological context to trace atoms through metabolism.[60] Except for deuterium, stable isotopes are non-toxic, even in large doses.[61,62] Therefore, they can also be applied *in vivo*, without the ethical concerns that arise from radioactive tracers.

Stable isotope labeling experiments can provide novel insights into cellular metabolism of less well studied organisms.[63] The labeling patterns of tracer-derived metabolites can help to determine active metabolic pathways under a given condition. In humans, there are diverse clinical applications of stable isotope labeling,[64] such as the measurement of gluconeogenesis or glucose and alanine turnover.[65,66]

Stable isotope labeled compounds have been used as internal standards for IDMS-based quantification of analytes. IDMS provides more robust results than other quantification approaches, since isotopic peak ratios in a mass spectrum are very stable (section 3.4). Besides individual labeled compounds, whole organisms have been isotopically labeled, which provides a cheap source for a large number of internal standards.[44,45] In addition to more precise quantification, such fully labeled organisms are also valuable to distinguish native metabolite from analytical background.[46,47]

In summary, the main applications of stable isotope labeled compounds in metabolomics research are as tracers for metabolic flux analysis, as internal standards for quantification, for structure elucidation and elucidation of reaction mechanisms.[67,68]

Commonly used stable isotopes in metabolic research are $^{13}$C, $^{15}$N and $^{2}$H, the main elements of biological systems. Rarely sulfur isotopes are used.[69] Phosphorous is monoisotopic, that

means there exists no heavy stable isotope. Although deuterated tracers can be easily produced and are less expensive, mostly $^{13}C$ tracers are used because of their lower kinetic isotope effect in biochemical reactions as well as in chromatographic retention.

### 1.3.3 Kinetic isotope effect

As previously mentioned, stable isotope labeled compounds are mostly behaving like their unlabeled counterparts, but there can be some differences. Different isotopologues can exhibit different reaction rates. This phenomenon is called *kinetic isotope effect*. Theoretically this effect can be very big[62] but in practice the difference in reaction rates is only around a few percent for most isotopes.[70] Nevertheless, these small differences can provide insights into ecological networks or geochemistry.[71–73] There have also been speculations, that consumption of heavy isotopes may reduce the susceptibility to oxidative damage and may therefore increase life span.[74]

The extent of the kinetic isotope effect depends on the isotopic species as well as on the significance of the respective atoms in the given reaction. For example, with $^{13}C$-labeling the isotope effect is usually negligible, but can become significant in reactions where the carbon backbone is modified. That becomes apparent in the ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO)-reaction in photosynthesis, where there is a strong discrimination of $^{13}CO_2$.[75]

In stable isotope-assisted metabolomics, the kinetic isotope effect is usually neglected and it is commonly assumed that the isotopic tracer is metabolized just like its unlabeled analogue. This includes $^{13}C$-MFA where the kinetic isotope effect may well cause slight errors in the estimated fluxes.[70]

Isotope labeling can also have an impact on chromatographic retention. This is especially true for deuterated compounds. Hydrogen atoms on the periphery of the molecules can interact more with the stationary phase than for example carbon atoms of the molecular backbone which are more shielded.[76] Depending on the extent of deuteration this can lead to splitting of chromatographic peaks or even baseline separation.[77] For that reason carbon or nitrogen labeling is usually preferred for GC- or LC-MS experiments, although deuterated tracers are easier to synthesize and thus cheaper.

In other contexts, the strong kinetic isotope effect of deuterium is of interest, since it can provide insights into reaction mechanisms.[78,79]

### 1.3.4 Incorporation of isotopes leads to isotopic isomerism

Through the incorporation of different isotopes of the elements into molecules, a new isomerism arises — isotopic isomerism. Common terms for isotopic isomers are *isotopomers*, *isotopologues*

and *mass isotopomers*. These terms are used inconsistently in current literature.[67,80–84] For that reason, I will provide clear definitions which I use throughout this work.

I will illustrate the different groups of isotopic isomers with the help of a relatively simple molecule, glycolonitrile ($C_2H_3NO$, Figure 6). This compound is of no further biological relevance, but offers a good example, since it has a manageable number of isotopic isomers.

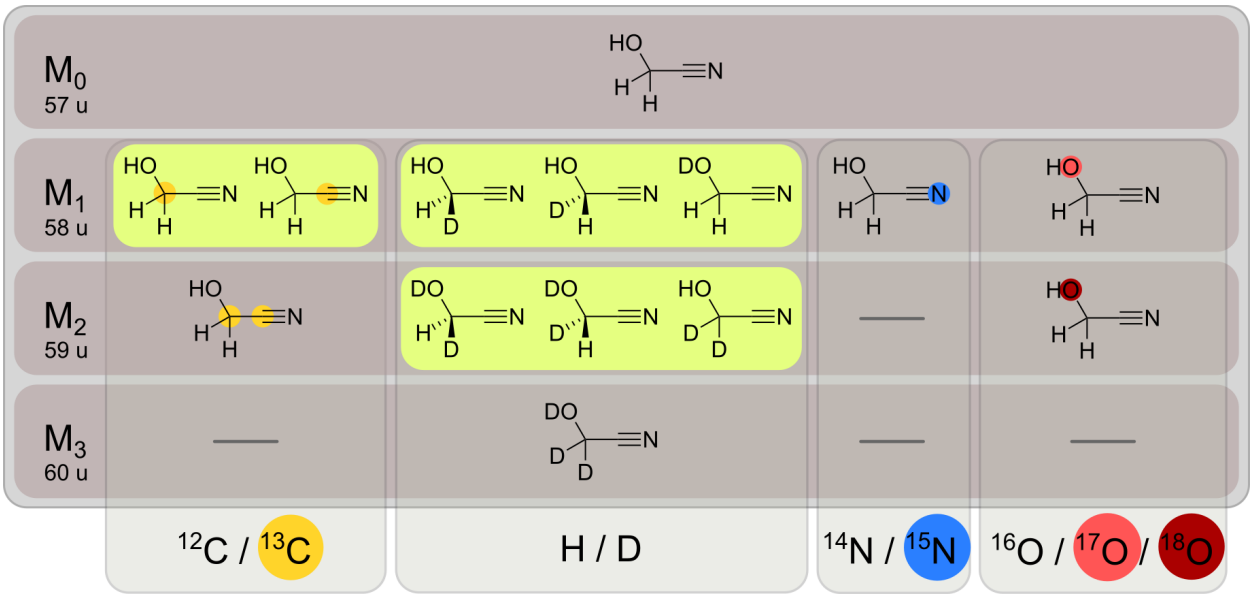



**Figure 6: Some isotopic isomers (isotopologues) of glycolonitrile.** Pure $^{13}C$, $^{2}H$, $^{15}N$, $^{17}O$ and $^{18}O$ isotopologues are shown, mixed isotopologues are omitted. Isotopomers have the same elemental and isotopic composition but can differ in the position of the isotopic substitution (green boxes). The $M+n$, $M_n$ or $Mn$ mass isotopomer comprises all isotopologues with a mass increment of $n$ as compared to the lightest isotopologue (dark gray boxes).

### Isotopologues

For the definition of *isotopologues*[1], I follow the recommendations of the International Union of Pure and Applied Chemistry (IUPAC)[86] that states that an isotopologue is a

> molecular entity that differs only in isotopic composition (number of isotopic substitutions), e.g. $CH_4$, $CH_3D$, $CH_2D_2$.

This recommendation is not explicit on whether the number of isotopic substitutions may or must differ, and on whether the isotopic substitutions may or may not be on different elements. For example, it is not clear whether $^{13}CH_4$ is isotopologic with $CD_4$. However, it is clarified later[85] that

> isotopolog ions [...] are ions that differ only in the isotopic composition of one or more of the constituent atoms

[1] *Isotopologue* is short for *isotopic homologue*[85]

and that

> the ions forming an *isotope cluster* [in a mass spectrum] corresponding to the *natural isotopic abundances* of the constituent atoms [are isotopologic].

Therefore, "isotopologue" comprises all kinds of isotopic substitutions (Figure 6). In other words, all compounds that consist of the same elements but possibly different isotopes while otherwise exhibiting the same connectivity and spatial arrangement are isotopologic (are isotopologues). Therefore, isotopologues can differ in their masses (Figure 6, all depicted molecules are isotopologic).

For practical reasons isotopologues are often grouped by their nominal masses (section 1.3.5). Such groups of isotopologues will be addressed[2] as $M_n$ where $n$ is the number of additional neutrons in the molecule, or the increase in nominal mass as compared to the lightest isotopologue ($M_0$). $n$ does not necessarily correspond to the number of isotopes. For example in the case of $^{18}O$ labeling, there is only a single $^{18}O$ isotope present in the $M_2$ isotopologues (Figure 6).

**Number of isotopologues** Of a molecule with $n$ atoms of an element with $m$ possible isotopic substitutions, there can exist $m^n$ isotopologues. In the case of $^{13}C$ labeling $n$ equals two ($^{12}C$, $^{13}C$). For the example in Figure 6 there are therefore $2^2 = 4$ carbon isotopologues. For hydrogen there are $2^3 = 8$ isotopologues, for nitrogen $2^1 = 2$ and for oxygen ($^{16}O$, $^{17}O$, $^{18}O$, $n = 3$) $3^1 = 3$. Of course, the different isotopic species can occur in parallel. For the combined occurrence of $^{13}C$, $^{2}H$, $^{15}N$, $^{17}O$ and $^{18}O$ there exist $2^2 \cdot 2^3 \cdot 2^1 \cdot 3^1 = 4 \cdot 8 \cdot 2 \cdot 3 = 192$ different isotopologues. The number is lower for compounds with symmetry elements.

### Isotopomers

The term *isotopomer* is defined by the IUPAC[86,87] as follows:

> [Isotopomers are] Isomers having the same number of each isotopic atom but differing in their positions. The term is a contraction of 'isotopic isomer'. Isotopomers can be either constitutional isomers (e.g. $CH_2DCH{=}O$ and $CH_3CD{=}O$) or isotopic stereoisomers [e.g. (*R*)- and (*S*)-$CH_3CHDOH$ or (*Z*)- and (*E*)-$CH_3CH{=}CHD$].

Isotopomers are positional isomers. Isotopomeric compounds have the same isotopic composition and, thus, the same mass (Figure 6, green boxes). There are no $M_0$ isotopomers as there exists only a single isotopologue.

[2] Often also addressed as M$n$ or M+$n$.

### Mass isotopomers

For the term *mass isotopomer* there exists unfortunately no IUPAC recommendation and in current literature it is used in different ways.[67,80–84] I use the term "mass isotopomer" to address a group of isotopologues that have a certain mass increment as compared to the lightest isotopologue. Therefore, "$M_1$ mass isotopomers" would be synonymous with "$M_1$ isotopologues" (Figure 6). This definition follows mostly the more pragmatic one provided by Hellerstein & Neese[81], who make the term dependent on the resolving power of the mass spectrometer. Compounds that are mass isotopomeric have the same nominal mass, but possibly different numbers and kinds of isotopes incorporated. In the case of $^{13}C$ labeling experiments, the term mass isotopomer is usually only used for the $^{13}C$ mass isotopomers arising from tracer-derived artificial enrichment, excluding isotopologues arising from incorporation of different isotopic elements.

In summary, all isotopomeric compounds are also mass isotopomeric and all mass isotopomeric compounds are also isotopologic (Figure 6). However, the inverse is not necessarily true. More formally expressed: isotopomers $\subset$ mass isotopomers $\subset$ isotopologues.

### Isotopic stereoisomers

The incorporation of stable isotopes into a molecule can also give rise to stereo isomerism. In the example in Figure 6, C-2 of glycolonitrile is prochiral. Replacing one of the two $\alpha$-hydrogens with deuterium will give rise to two isotopomeric enantiomers. Since enzymes are chiral molecules themselves, they distinguish between the different substituents of prochiral compound.[88] In such cases, only a single enantiomer is formed from a pro-chiral compound. This has consequences for the interpretation of stable isotope labeling patterns for metabolic flux analyses.[89]

## 1.3.5 Measures of isotopic enrichment

In stable isotope-assisted metabolomics the relative abundances of the individual isotopologues are of interest. However, complete isotopologue distributions, *i.e.* the relative abundance of each isotopologue, cannot, or not easily be determined for most compounds (subsection 3.3.1). Therefore, other measures are needed to assess isotopic enrichment.

### Mass isotopomer distributions

Complete isotopologue distributions cannot easily be determined for most compounds, but MIDs, that is the isotopologue distribution aggregated by the mass increment as compared to the $M_0$ isotopologue, can easily be determined from a mass spectrum (Figure 7A,B). MIDs *per se* do not contain any positional information on isotopic enrichment or the nature of

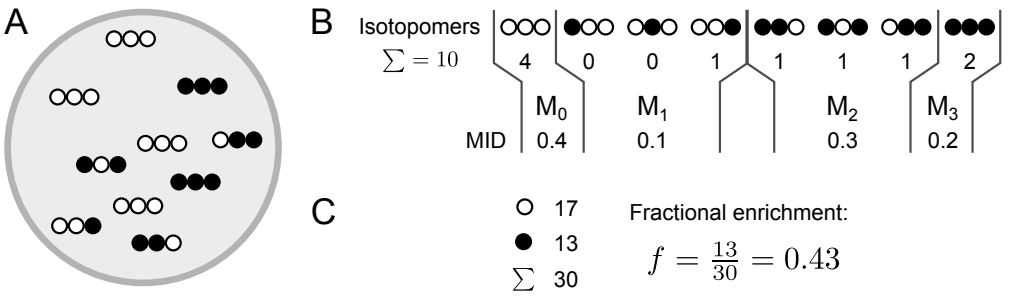



**Figure 7: Measures of isotopic enrichment.**
A) Stable isotope labeling experiments usually yield a mixture of different isotopologues. ○: light isotope; ●: heavy isotope.
B) Full isotopologue distributions usually cannot be obtained from MS measurements. Instead, isotopic enrichment is measured in the form of mass isotopomer distributions (MIDs). MIDs *per se* do not hold information of positional isotopic enrichment.
C) The fractional enrichment of ● describes the average abundance of the heavy isotope in the overall pool of a compound.

the isotopic species, but positional labeling information can be derived from MIDs of mass spectrometric fragment ions if the fragmentation pathway is known or at least the atoms of the original molecule that are retained within a certain fragment are known.[26,90] I will show how such positional labeling can provide additional biological insights (section A.9) and how in some cases even the whole isotopologue distribution can be determined from mass spectrometric data (subsection 3.3.1).

### Fractional enrichment

In certain contexts, the exact labeling pattern of a compound is less important than the overall extent of isotopic enrichment. This can be determined from the MID and illustrated as the average number of *e.g.* $^{13}$C isotopes per molecule:

$$\tilde{f} = \sum_{i=1}^{N} i \cdot M_i \tag{1.1}$$

Another measure, which is easier to compare across different compounds, is the average relative heavy isotope abundance, or the *fractional enrichment* (Figure 7):

$$f = \frac{1}{N} \sum_{i=1}^{N} i \cdot M_i \tag{1.2}$$

The fractional enrichment of a compound from a fully $^{13}$C-labeled tracer describes how much of its carbon is derived from that tracer.

### 1.3.6 Determination of isotopic enrichment

In a stable isotope labeling experiment, an isotopically enriched tracer is applied to an organism. After metabolization of this tracer, labeled compounds contain isotopic species from artificial enrichment of the tracer, as well as from natural isotope abundance (Table 1). However, for the biological analysis, we are only interested in the isotopic enrichment that is caused by incorporation of tracer-derived isotopes. Thus, the "raw" MIDs which are directly accessible from the mass spectrum are corrected for natural isotope abundance (subsection 1.3.6). This can lead to some confusion, as the corrected MID is usually just referred to as MID. The term "label distribution" would be more appropriate to discriminate natural isotope occurrence from artificial isotopic enrichment. However, for consistency with current literature, I use "mass isotopomer distribution" for the "corrected MID" and "raw MID" to refer to the uncorrected MID.

In the case of carbon labeling on metabolites the raw and the corrected MIDs are very similar, because the natural abundances of isotopes of the other elements is comparably low (Table 1). However, in the case of GC-MS analysis, where there is often a high number of additional carbons introduced from the derivatization agents (section 1.2.1), the contribution of natural isotopes is significant and it is therefore vital to correct for this natural contribution (subsection 1.3.6).

There are two main approaches to perform this correction for MS-derived MIDs: One uses the theoretical mass isotopomer abundance of a given mass spectrometric fragment and therefore requires the elemental composition of the observed fragment;[91,92] the other calculates the isotopic enrichment from the comparison of the mass spectra of the labeled and unlabeled compound and therefore does not require any additional information.

#### From elemental composition

In a $^{13}C$ labeling experiment, the incorporation of tracer-derived $^{13}C$ into a compound with $n$ enrichable carbons can give rise to $n+1$ mass isotopomers disregarding the natural isotope abundances of any other elements. All non-enriched positions within these isotopologues will also contain stable isotopes according to the natural isotope abundance of the respective elements.

Therefore, the raw MID, or the relative intensities in an isotopic cluster $I^l$ (Figure 3) in the mass spectrum of a labeled fragment are the average of the MIDs of the molecules with $0...n$ "artificial" $^{13}C$ isotopes, weighted by the corrected MID M:

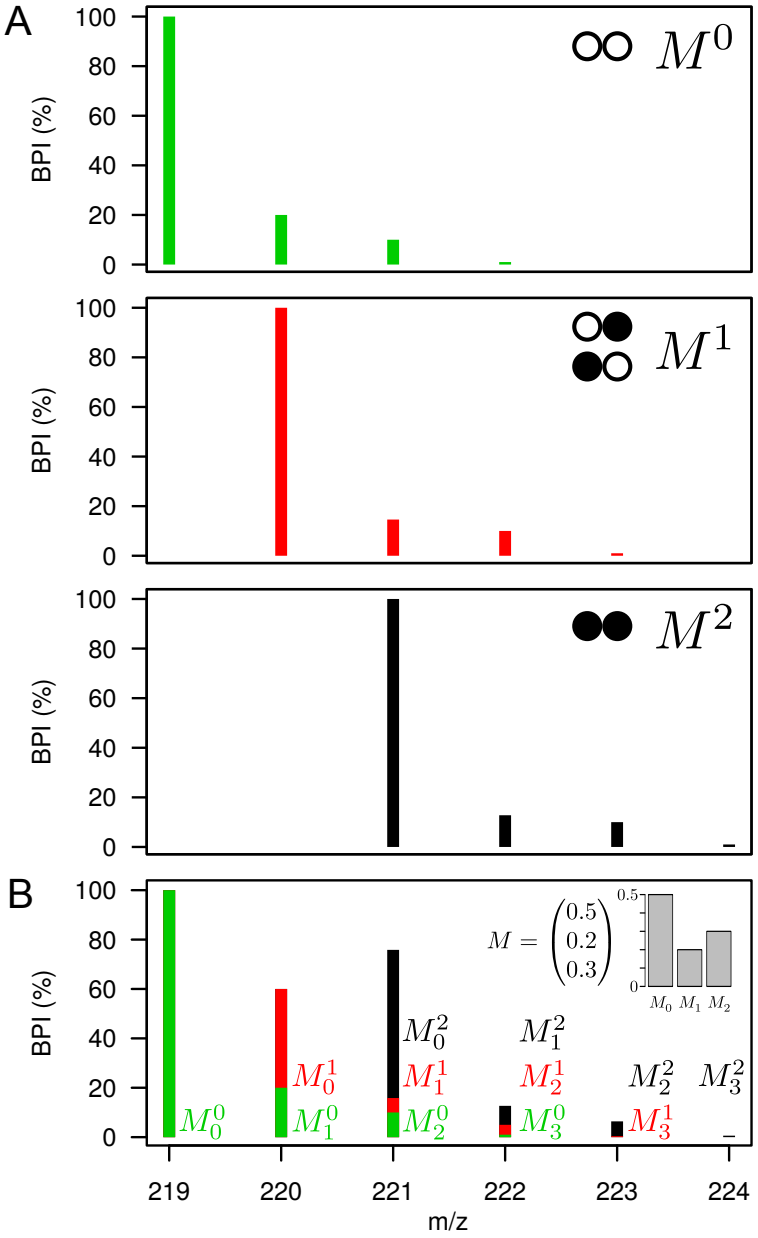



**Figure 8: Mass spectrum of a mixture of different isotopologues of an isotopically enriched compound using the example of $^{13}$C labeling of glycine 2TMS.** ○: $^{12}$C; ●: $^{13}$C.

A) Stable isotope labeling of a compound with two enrichable positions. The mass spectra of the $M_0$, $M_1$, and $M_2$ isotopologues show isotopic peaks due to the natural isotope abundance in the non-enriched positions. The natural isotope contribution, *i.e.* the relative intensity of the M+1 peak, decreases with the increasing artificial enrichment. Relative M+0 intensity increases by the same value.

B) After a stable isotope labeling experiment, the mass spectrum of the given compound is a mixture of the raw MIDs of the three artificial mass isotopomers.

$$I^l = \begin{pmatrix} \mathrm{M}_0^0 & 0 & \cdots & 0 \\ \mathrm{M}_1^0 & \mathrm{M}_0^1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{M}_n^0 & \mathrm{M}_{n-1}^1 & \cdots & \mathrm{M}_0^n \end{pmatrix} \cdot \mathrm{M} = \begin{pmatrix} \mathrm{M}_0^0 \cdot \mathrm{M}_0 \\ \mathrm{M}_1^0 \cdot \mathrm{M}_0 + \mathrm{M}_0^1 \cdot \mathrm{M}_1 \\ \vdots \\ \mathrm{M}_n^0 \cdot \mathrm{M}_0 + \mathrm{M}_{n-1}^1 \cdot \mathrm{M}_1 + \cdots + \mathrm{M}_0^n \cdot \mathrm{M}_n \end{pmatrix} \tag{1.3}$$

with $\mathrm{M}_b^a$ as the relative abundance of $b$th natural mass isotopomer of the compound containing $a$ tracer isotopes. When the relative abundances of all $\mathrm{M}_b^a$ are known, and $I^l$ is available from a mass spectrometric measurement, then corrected MIDs can be obtained as

$$\mathrm{M} = \mathbf{C}^{-1} \cdot I^l \tag{1.4}$$

This equation is usually solved by least squares regression.

To populate the correction matrix $\mathbf{C}$, we need the natural MIDs for the enriched isotopologues. When the elemental composition of a molecule is known, its theoretical MID occurring due to natural isotope abundance can be determined.[93] Therefore, the probabilities for the inclusion of each isotopic element need to determined according to the elemental composition. The probability $P(x, p)$ for the presence of $x_1, \ldots, x_K$ atoms of each of the $K$ isotopes of an element with the natural abundance of $p_1, \ldots, p_K$ is given by the multinomial or polynomial distribution[94,95]:

$$P(x, p) = \frac{(\sum_{i=1}^{K} x_i)!}{\prod_{i=1}^{K} (x_i!)} \cdot \prod_{i=1}^{K} p_i^{x_i} \tag{1.5}$$

For di-isotopic elements, Equation 1.5 simplifies to the binomial distribution.[92,94] The probability $P(x)$ to encounter $x$ isotopes that occur with the abundance or probability of $p$ among a total of $n$ atoms is then given by:

$$P(x) = \binom{n}{x} \cdot p^x \cdot (1-p)^{n-x} \tag{1.6}$$

The probabilities for all different combinations of isotopes leading to a given mass isotopomer have to be summed up for each element and then multiplied to obtain the theoretical relative mass isotopomer abundance. Such calculations can be performed for all $n + 1$ carbon mass isotopomers that can arise from stable isotope labeling. The resulting MID vectors can then be put together to obtain the correction matrix $\mathbf{C}$.

**An example** I will demonstrate this method with the example of glycine 2TMS (Figure 3). The ion at $m/z$ 204 is the $[\mathrm{M\text{-}CH_3}]^+$ fragment ($[\mathrm{C_7H_{18}NO_2Si_2}]^+$) and contains the complete carbon backbone of glycine. It needs to be corrected for the natural isotope abundance of C,

N, O and Si.

Assuming the masses 204–208 have been measured, the following equation system needs to be solved:

$$\begin{pmatrix} \mathrm{M}_0^0 & 0 & 0 \\ \mathrm{M}_1^0 & \mathrm{M}_0^1 & 0 \\ \mathrm{M}_2^0 & \mathrm{M}_1^1 & \mathrm{M}_0^2 \\ \mathrm{M}_3^0 & \mathrm{M}_2^1 & \mathrm{M}_1^2 \\ \mathrm{M}_4^0 & \mathrm{M}_3^1 & \mathrm{M}_2^2 \end{pmatrix} \cdot \begin{pmatrix} \mathrm{M}_0 \\ \mathrm{M}_1 \\ \mathrm{M}_2 \end{pmatrix} = \begin{pmatrix} I_{204} \\ I_{205} \\ I_{206} \\ I_{207} \\ I_{208} \end{pmatrix} \tag{1.7}$$

Although the maximal artificial enrichment will only lead to $\mathrm{M}_2$ ($I_{206}$), the further measurements ($I_{207}$ and $I_{208}$) make the system overdetermined and will make the fitting more robust.

The probability for $\mathrm{M}_0^0$ of glycine 2TMS $m/z$ 204 ($[\mathrm{C_7H_{18}NO_2Si_2}]^+$) is given with Equation 1.5 and 1.6 as:

$$P(\mathrm{M}_0^0) = P(\mathrm{C_7H_{18}NO_2Si_2}) = P(7, p_{^{12}\mathrm{C}}) \cdot P(18, p_{^{1}\mathrm{H}}) \cdot P(1, p_{^{14}\mathrm{N}}) \cdot P(2, p_{^{16}\mathrm{O}}) \cdot P(2, p_{^{28}\mathrm{Si}}) \tag{1.8}$$

The probability for $M_0^1$ is calculated the same way, but for the sum formula $\mathrm{C_6H_{18}NO_2Si_2}$, since one carbon is already considered as labeled and not subject to natural isotope abundance:

$$P(\mathrm{M}_0^1) = P(\mathrm{C_6H_{18}NO_2Si_2}) = P(6, p_{^{12}\mathrm{C}}) \cdot P(18, p_{^{1}\mathrm{H}}) \cdot P(1, p_{^{14}\mathrm{N}}) \cdot P(2, p_{^{16}\mathrm{O}}) \cdot P(2, p_{^{28}\mathrm{Si}}) \tag{1.9}$$

The probability of $\mathrm{M}_0^1$ is higher than for $\mathrm{M}_0^0$, since there is already one $^{13}$C present and chance to include another $^{13}$C in 6 instead of 7 positions is lower.

For $\mathrm{M}_1^0$ there are multiple combinations of isotopes that lead to the same mass. The probabilities of these combinations need to be summed up:

$$\begin{aligned} P(\mathrm{M}_1^0) = \quad & P(^{13}\mathrm{C}_1{}^{12}\mathrm{C}_6{}^{1}\mathrm{H}_{18}{}^{14}\mathrm{N}^{16}\mathrm{O}_2{}^{28}\mathrm{Si}_2) \\ + \quad & P(^{12}\mathrm{C}_7{}^{2}\mathrm{H}^{1}\mathrm{H}_{17}{}^{14}\mathrm{N}^{16}\mathrm{O}_2{}^{28}\mathrm{Si}_2) \\ + \quad & P(^{12}\mathrm{C}_7{}^{1}\mathrm{H}_{18}{}^{15}\mathrm{N}^{16}\mathrm{O}_2{}^{28}\mathrm{Si}_2) \\ + \quad & P(^{12}\mathrm{C}_7{}^{1}\mathrm{H}_{18}{}^{14}\mathrm{N}^{17}\mathrm{O}^{16}\mathrm{O}^{28}\mathrm{Si}_2) \\ + \quad & P(^{12}\mathrm{C}_7{}^{1}\mathrm{H}_{18}{}^{14}\mathrm{N}^{16}\mathrm{O}_2{}^{29}\mathrm{Si}^{28}\mathrm{Si}) \end{aligned}$$

For subsequent mass isotopomers, the calculation is more complicated, since also A+2

isotopes ($^{18}O$, $^{30}Si$) need to be considered. Even for the relatively small ion $[C_7H_{18}NO_2Si_2]^+$, there can exist 16 $M_2$ isotopologues:

$$
\begin{aligned}
P(M_2^0) = \quad & P(^{13}C_2{}^{12}C_5{}^{1}H_{18}{}^{14}N^{16}O_2{}^{28}Si_2) + \quad P(^{13}C^{12}C_6{}^{2}H^{1}H_{17}{}^{14}N^{16}O_2{}^{28}Si_2) \\
+ \quad & P(^{13}C^{12}C_6{}^{1}H_{18}{}^{15}N^{16}O_2{}^{28}Si_2) + \quad P(^{13}C^{12}C_6{}^{1}H_{18}{}^{14}N^{17}O^{16}O^{28}Si_2) \\
+ \quad & P(^{13}C^{12}C_6{}^{1}H_{18}{}^{14}N^{16}O_2{}^{29}Si^{28}Si) + \quad P(^{13}C^{12}C_6{}^{2}H_2{}^{1}H_{16}{}^{14}N^{16}O_2{}^{28}Si_2) \\
+ \quad & P(^{12}C_7{}^{2}H^{1}H_{17}{}^{15}N^{16}O_2{}^{28}Si_2) + \quad P(^{12}C_7{}^{2}H^{1}H_{17}{}^{14}N^{17}O^{16}O^{28}Si_2) \\
+ \quad & P(^{12}C_7{}^{2}H^{1}H_{17}{}^{14}N^{16}O_2{}^{29}Si^{28}Si) + \quad P(^{12}C_7{}^{1}H_{18}{}^{15}N^{17}O^{16}O^{28}Si_2) \\
+ \quad & P(^{12}C_7{}^{1}H_{18}{}^{15}N^{16}O_2{}^{29}Si^{28}Si) + \quad P(^{12}C_7{}^{1}H_{18}{}^{14}N^{17}O_2{}^{28}Si_2) \\
+ \quad & P(^{12}C_7{}^{1}H_{18}{}^{14}N^{17}O^{16}O^{29}Si^{28}Si) + \quad P(^{12}C_7{}^{1}H_{18}{}^{14}N^{16}O_2{}^{29}Si_2) \\
+ \quad & P(^{12}C_7{}^{1}H_{18}{}^{14}N^{16}O_2{}^{30}Si^{28}Si) + \quad P(^{12}C_7{}^{1}H_{18}{}^{14}N^{18}O^{16}O^{28}Si_2)
\end{aligned}
$$

For large molecules these calculations are very complex and become computationally expensive.[93]

### From reference spectrum

The previously described approach to quantify isotopic labeling has the drawback that it requires the elemental composition of the ion of interest. It can therefore not be used to determine isotopic enrichment of unidentified compounds, or of fragment ions of unknown composition.

However, at times it can be interesting to determine the isotopic enrichment of an unknown or unidentified compound. Therefore, we need other means to correct for natural isotope abundance. Jennings & Matthews[96] described an elaborate approach based on an unlabeled reference spectrum, that only requires the natural isotope abundance of the tracer element as additional information. This way to correct MIDs is the basis for non-targeted stable isotope labeling analysis approaches (subsection 1.3.7). I will describe a simplified version of this approach which is used in NTFD and MIA (section A.2 and A.8).

To correct a raw MID for natural isotope abundance, we are again confronted with the problem in Equation 1.4. However, if we do not have the elemental composition available, we need to populate the correction matrix $\mathbf{C}$ by other means. The first column of $\mathbf{C}$, $M^0$, is the natural MID of the molecule without any artificial enrichment. We have these values available from the mass spectrum $I^{ul}$ of the unlabeled compound and can get to:

$$\mathbf{C} = \begin{pmatrix} s_0^{ul} & 0 & \cdots & 0 \\ s_1^{ul} & \mathrm{M}_0^1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ s_n^{ul} & \mathrm{M}_{n-1}^1 & \cdots & \mathrm{M}_0^n \end{pmatrix} \tag{1.10}$$

The remaining $\mathrm{M}_b^a$ can be replaced in a similar way. The spectrum of labeled compounds $\mathrm{M}^1, \ldots, \mathrm{M}^n$ looks similar to the unlabeled spectrum, but it is shifted to higher mass (Figure 8). Furthermore, due to the fixed inclusion of stable isotopes, the $\mathrm{M}_0$ abundance will increase and the other mass isotopomer abundances will decrease.

In the case of $^{13}$C labeling of a compound with $n$ carbons that are subject to labeling, the relative abundance of $\mathrm{M}_0$ is the probability of the simultaneous occurrence of $n$ $^{12}$C isotopes and heavy isotopes in any other elements:

$$\mathrm{M}_0^0 = p_{^{12}\mathrm{C}}^n \cdot c \tag{1.11}$$

with $c$ as the probability of the occurrence of the lightest isotopes of all atoms that are not subject to artificial enrichment. The relative abundance of $\mathrm{M}_1^0$ is the summed probability of the occurrence of one $^{13}$C isotope in any of the $n$ enrichable positions or of a heavy isotope with a mass increment of one in any of the other positions:

$$\mathrm{M}_1^0 = n \cdot p_{^{13}\mathrm{C}} \cdot p_{^{12}\mathrm{C}}^{n-1} \cdot c + p_{^{12}\mathrm{C}}^n \cdot s \tag{1.12}$$

with $s$ as the probability for the occurrence of a heavy isotope with a mass increment of one in any of the atoms not subject to artificial enrichment.

When $a$ carbons are substituted by $^{13}$C, the probability of $\mathrm{M}_0^a$ increases as compared to $\mathrm{M}_0^0$ by the factor $p_{^{12}\mathrm{C}}$, since it only requires the simultaneous occurrence of $(n-a)$ $^{12}$C isotopes:

$$\mathrm{M}_0^a = p_{^{12}\mathrm{C}}^{n-a} \cdot c \overset{(1.11)}{=} p_{^{12}\mathrm{C}}^{-a} \cdot \mathrm{M}_0^0 \tag{1.13}$$

The $\mathrm{M}_1^a$ abundance increases accordingly, because the probability of the occurrence of one $^{13}$C isotope in only $(n-a)$ positions is lower than in $n$ positions:

$$\begin{aligned}
\mathrm{M}_1^a &= (n-a) \cdot p_{^{13}\mathrm{C}} \cdot p_{^{12}\mathrm{C}}^{n-a-1} \cdot c + p_{^{12}\mathrm{C}}^{n-a} \cdot s && (1.14)\\
&= p_{^{12}\mathrm{C}}^{-a} \left( (n-a) \cdot p_{^{13}\mathrm{C}} \cdot p_{^{12}\mathrm{C}}^{n-1} \cdot c + p_{^{12}\mathrm{C}}^n \cdot s \right) && (1.15)\\
&= p_{^{12}\mathrm{C}}^{-a} \left( n \cdot p_{^{13}\mathrm{C}} \cdot p_{^{12}\mathrm{C}}^{n-1} \cdot c - a \cdot p_{^{13}\mathrm{C}} \cdot p_{^{12}\mathrm{C}}^{n-1} \cdot c + p_{^{12}\mathrm{C}}^n \cdot s \right) && (1.16)\\
&\overset{(1.11),(1.12)}{=} p_{^{12}\mathrm{C}}^{-a} \left( \mathrm{M}_1^0 - a \cdot p_{^{13}\mathrm{C}} \cdot p_{^{12}\mathrm{C}}^{-1} \cdot \mathrm{M}_0^0 \right) && (1.17)\\
&\overset{(1.11),(1.13)}{=} \frac{\mathrm{M}_0^a}{\mathrm{M}_0^0} \cdot \left( \mathrm{M}_1^0 - a \cdot p_{^{13}\mathrm{C}} \cdot p_{^{12}\mathrm{C}}^{-1} \cdot \mathrm{M}_0^0 \right) && (1.18)
\end{aligned}$$

Since the natural abundance of $^{13}C$ is rather low, we assume that only the $M_1$ abundance is affected by the substitution of up to $n$ carbons with $^{13}C$ isotopes and neglect the insignificant changes in the abundances of heavier mass isotopomers. Rearranging (1.18) provides a correction term for the $M_1$ to $M_0$ ratio of the labeled mass isotopomers:[96]

$$\frac{M_1^a}{M_0^a} = \frac{M_1^0}{M_0^0} - a \cdot \frac{p_{^{13}C}}{p_{^{12}C}} \tag{1.19}$$

When a $^{13}C$ isotope is artificially included in the molecule, this position is no longer subject to natural isotope abundance. Therefore, the relative natural $M_0$ abundance of these mass isotopomers increases and the relative $M_1$ abundance decreases (Figure 8).

We define the correction term in (1.19) as:

$$c_a = a \cdot \frac{p_{^{13}C}}{p_{^{12}C}} \tag{1.20}$$

Since we assume, that the relative abundances of the heavier mass isotopomers remain constant, the $M_0$ abundance increases by the same value as the $M_1$ abundance decreases. Therefore, their sum will not change:

$$M_1^a + M_0^a = M_1^0 + M_0^0 \tag{1.21}$$

Combining Equation 1.19, 1.20 and 1.21 provides a correction term for $M_0$:

$$M_0^a = M_0^0 + \frac{M_0^0 \cdot c_a}{\frac{M_0^0}{M_1^0} + 1 - c_a} \tag{1.22}$$

Since the sum of $M_0$ and $M_1$ needs to be constant (1.21), $M_1$ needs to decrease by the same amount as $M_0$ increases:

$$M_1^a = M_1^0 - \frac{M_0^0 \cdot c_a}{\frac{M_0^0}{M_1^0} + 1 - c_a} \tag{1.23}$$

Now we can populate the correction matrix and determine the isotopic enrichment without requiring any further knowledge on the compound.

I explained this method for the case of $^{13}C$ labeling, but it can analogously be applied to any other isotopic element. Depending on the natural isotope abundance of the isotopic element, a $M_2$ correction needs to be performed additionally as described by Jennings & Matthews[96]. A further benefit of this reference spectrum based approach for MID correction is, that it corrects for any measurement biases, ionization effects, and deviations of natural isotope abundance from the global average.[96]

#### Positional information on isotopic enrichment

Position-specific isotopic enrichment can be easily determined by NMR spectroscopy, but is generally not available from MS measurements. However, MIDs from mass spectrometric fragments can provide such positional isotopic enrichment if fragmentation pathways, or more precisely, the atoms retained in a certain fragment are known.[97] It has been shown by Choi *et al.*[90] how all 16 carbon isotopologues of aspartic acid can be determined from tandem MS ($MS^2$) spectra of its TBDMS derivative. Antoniewicz *et al.*[98] showed how fragment MIDs from three glucose derivatives can be combined to obtain the relative abundance of all informative hydrogen isotopologues of glucose. More impressively, Di Donato *et al.*[99] combined several derivatization techniques to finally obtain the relative abundances of all 32 carbon isotopologues of glutamic acid.

For smaller molecules the full isotopologue distribution can be obtained more easily. Such knowledge on positional enrichment can provide more biological information and additional constraints for $^{13}$C-MFA (section 1.3.8).[90,97,100] In subsection 3.3.1, I show how the complete isotopologue distribution of some interesting metabolites can be determined without additional experimental effort from a single EI-MS measurement of commonly used derivatives.

### 1.3.7 Non-targeted detection of stable isotope labeling

Whereas radioactive labeling experiments have been analyzed in a non-targeted manner for a long time, stable isotope labeling analysis has usually been rather targeted. The reason is, that heavy stable isotopes do not exhibit any readily detectable qualitative differences to their light counterparts. Therefore, dedicated algorithms are necessary to detect stable isotopic enrichment in mass spectrometric data in a non-targeted manner.

Over the last few years, several tools have been developed to detect stable isotope labeling metabolome-wide in a non-targeted manner. One can discriminate between qualitative and quantitative approaches. Qualitative detection of stable isotope labeling is for example sufficient to find out which compounds are derived from a given tracer, or in the case of complete labeling of a sample, to distinguish between analytes of interest and analytical background.[46,47,50,101] The quantitative detection of isotopic enrichment in the form of MIDs is often more informative. There are algorithms and tools available for the non-targeted detection and quantification of stable isotope labeling in GC-[102,103] and LC-MS[104–106] data.

The field of non-targeted stable isotope labeling analysis is still young and not many applications have been published yet. Some of those studies have only focused on the qualitative detection of isotopic enrichment,[101] or have not gone beyond the mere detection of numerical differences in relative mass isotopomer abundances between different experimental conditions.[50] More biological insights have been obtained by Gaglio *et al.*[7], who applied non-targeted tracer fate detection (NTFD) to detect differential transaminase activities in

cancer cells. Isotopic enrichment in lipids has been analyzed by Li *et al.*[107] using LC-MS. In a non-targeted approach, by using principal component analysis (PCA) to analyze mass isotopomer abundances after different timepoints, they revealed differential dynamics of different lipid species. Furthermore, Nakayama *et al.*[108] presented non-targeted metabolic turnover analysis as a novel strategy to detect metabolic pathways affected by external perturbations. They generated MID time-series data and performed PCA on the relative mass isotopomer abundances to reveal the metabolic vicinity of labeled compounds and to group unidentified compounds into discrete pathways.

The first promising approaches for the non-targeted analysis of stable isotope labeling data have been reported, but there is still a lack of adequate data analysis software and workflows.[109] Without such tools, data analysis can be tedious and when comparing multiple experimental conditions, one is easily overwhelmed by the complexity of the data.

In section 3.5 I describe how, using a newly developed software tool (section 3.2), quantitative stable isotope labeling data can be analyzed in a non-targeted manner, how it can be used to detect changes in metabolic fluxes, and how it can aid the identification or biochemical classification of the detected compounds. In the context of the analysis of hypoxic cancer cells, I demonstrate how such a data-driven analysis is a valuable tool for hypothesis generation and can grant biological insights.

#### NTFD - Non-targeted tracer fate detection

NTFD was the first approach to quantitatively detect isotopic enrichment in a non-targeted manner. Since much of the work presented here is built on the NTFD algorithm,[102] I will give a short introduction here. Further details of the NTFD algorithm and practical considerations for experimentation and data analysis are provided in the book chapter in section A.10.

NTFD operates on GC-MS measurements of an isotopically enriched and a non-enriched sample (Figure 9). The mass spectra from these two measurements are matched and subtracted. If the mass spectrum from the isotopically enriched sample is non-enriched, it is identical to the one from the non-enriched sample, and thus the difference spectrum will have all zero intensities. However, in the case of isotopic enrichment, there will be characteristic peaks in the difference spectrum for each labeled fragment. Detection of these peaks is the most important step for the automated detection of isotopic enrichment. I elaborate more on this step in subsection 3.2.3 where I present some enhancements of the NTFD implementation.

The peaks in the difference spectrum provide the $m/z$ ranges of each isotopically enriched fragment (section A.10). These fragments are used in a second step to quantify this enrichment by the procedure described above (section 1.3.6). This way, NTFD detects all isotopically

enriched compounds, and determines the MIDs of all their mass spectrometric fragments.

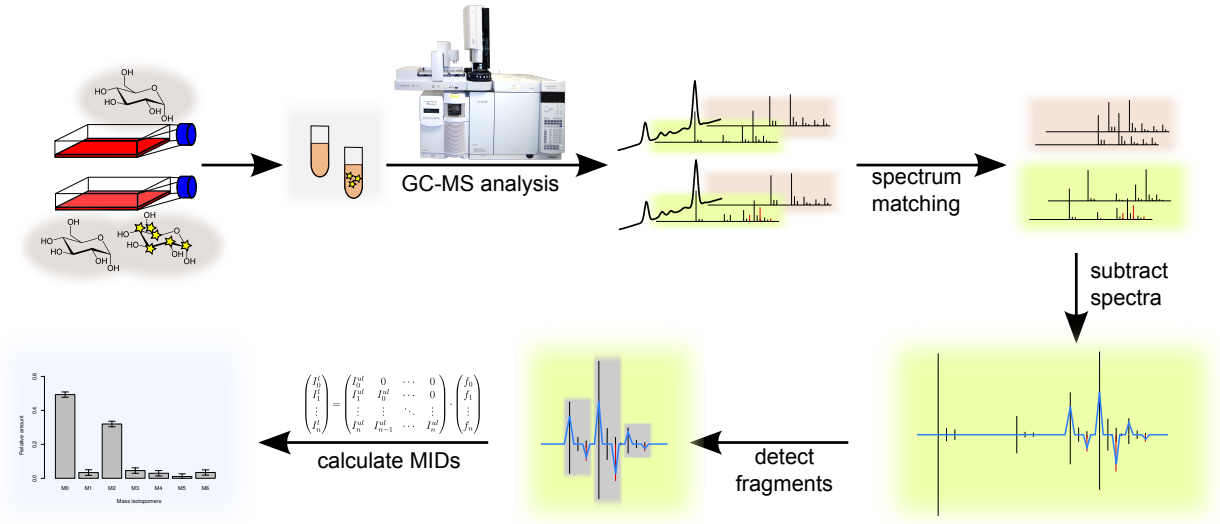



**Figure 9: The NTFD algorithm.** After GC-MS measurement of a isotopically enriched and non-enriched sample, mass spectra are matched and all enriched fragments are detected from the difference spectrum. For these fragments, the MIDs are calculated from the difference of labeled and unlabeled spectra.

### 1.3.8 Metabolic flux analysis

A problem of metabolomics studies which focus only on metabolite levels is that they cannot capture changes in metabolic fluxes. Metabolism is more dynamic and not only determined by metabolite levels alone. Metabolic fluxes are more sensitive to perturbations and can change without leaving any noticeable changes in metabolite levels. Metabolic networks are strongly interconnected. Especially within intermediary metabolism, there are often multiple pathways to get from one metabolite to another.[110] By just observing metabolite levels, one would not necessarily notice if certain pathways are blocked or induced. However, it can be very important to know the exact pathway that is taken to produce a certain metabolite. Firstly, changed metabolic fluxes can be a sign of pathological changes which are important for diagnostics; secondly, specific pathways can be important drug targets.[111]

This is one of the reasons why there is increasing interest in the *fluxome*, the fluxes (reaction rates) of all biochemical reactions. The fluxome is the result of a complex interplay of all other "*omes*" (Figure 1) and often seen as the endpoint of all cellular regulation and, thus, the closest connected to the given phenotype. Yet, none of these *omes* should be regarded in isolation. However, multi-omics studies are demanding, time-consuming, and cost-intensive and not routinely feasible. Since the fluxome probably provides the most direct readout of the current state of the organism, it is a good starting point to detect any changes. Furthermore, metabolic fluxes are much more sensitive than many other biological read-outs as for example

shown by Niklas *et al.*[112] who demonstrated that subtoxic effects of xenobiotics, effects which are usually overlooked in toxicity screenings, can be detected by $^{13}$C-MFA.

Metabolic fluxes can be very informative, but they cannot be measured as such. Cellular exchange fluxes with the environment (uptake or secretion) can be measured directly, but for intracellular fluxes one needs different means. Therefore, diverse MFA techniques have been developed.

One big group of flux analysis techniques is based on stoichiometric modeling.[113] The most common approach is flux balance analysis (FBA). FBA uses genome scale metabolic models[114] and measurements of exchange fluxes with the environment. It relies on a defined biological objective and uses linear optimization techniques to determine fluxes to optimally fulfill this objective. A problem is, that the objective function of an organism is not known. For bacteria the objective function is usually assumed to be the maximization of biomass production. Other objective functions can be production of adenosine triphosphate (ATP) or other metabolites. Due to the size of the metabolic networks and the associated high degree of freedom, FBA does usually not provide a single solution, but a huge solution space of different flux sets that fulfill the given objective equally well. Furthermore, pure stoichiometric modeling cannot distinguish between parallel pathways and cannot resolve metabolic cycles.[115]

To address these shortcomings, stable isotope-assisted $^{13}$C-MFA approaches have been developed. In addition to exchange fluxes, $^{13}$C-MFA uses stable isotope labeling as constraints for flux inference and can therefore provide more information on intracellular fluxes (section 1.3.8). It uses much smaller metabolic networks and leads to a smaller solution space, ideally to one single set of fluxes. $^{13}$C-MFA proved to be a valuable tool for metabolic engineering of bacteria,[116] but is also used in human[8,19] and plant systems.[117]

There are stationary and non-stationary MFA techniques available.[118] Stationary MFA analyzes organisms in metabolic steady state, that means that metabolite concentrations do not change over time ($\frac{dc}{dt} = 0$). For stationary $^{13}$C-MFA, additionally isotopic steady state is assumed, that means that isotopic labeling may not change over time ($\frac{dM}{dt} = 0$). Non-stationary MFA (INST-MFA) is used to analyze metabolic networks when they are in metabolic but not in isotopic steady state or neither in metabolic nor isotopic steady state. Such non-stationary techniques can be more informative, but are inherently more complex, since they additionally need to consider metabolite pool sizes and the temporal dimension.

#### From fluxes to MIDs

The mass isotopolome, *i.e.* the entirety of MIDs, of a given metabolic network in metabolic and isotopic steady state is fully determined by the associated metabolic fluxes and the isotopic tracer. When a tracer is taken up and metabolized by a cell the MIDs of downstream

metabolites will be identical to those of the tracer, until there is a loss or gain of labeled atoms, or a dilution by a converging flux carrying a different MID. I will illustrate that with $^{13}$C labeling in a simple metabolic model in metabolic and isotopic steady state (Figure 10). All reactions are assumed to be irreversible. When glutamine is used as a tracer, glutamic acid will always have the same MID as glutamine, no matter how the GLS flux changes. On the other hand, the MID of 2-oxoglutarate is strongly dependent on the flux ratio of GLS and isocitrate dehydrogenase (IDH). These fluxes lead to differently labeled 2-oxoglutarate. The MID of 2-oxoglutarate is the average of its MIDs when derived from either flux, weighted by the flux ratios (Figure 10). Thus, if all producing fluxes change proportionally, they will leave the MID of the product unchanged (Figure 10 A,C). Therefore, MIDs can only provide relative metabolic flux information.[119] Furthermore, this means that MIDs are fully determined by fluxes. If fluxes are not changing, neither will MIDs.

In summary, this leads to the following consequences for flux information contained in MIDs:

- MIDs alone can only provide relative flux information.
- Changes in MIDs must be caused by changes in metabolic fluxes.
- Not all changes in metabolic fluxes manifest in MID changes.

### From MIDs to fluxes — $^{13}$C-MFA

As described above, MIDs are fully defined by the underlying metabolic fluxes. Although this functional relationship is not invertible, MIDs often hold information on metabolic fluxes. This is exploited in $^{13}$C-MFA. Just as FBA, $^{13}$C-MFA requires a metabolic network model for which it determines an optimal set of fluxes. However, instead of fulfilling a biological objective, $^{13}$C-MFA determines a set of fluxes that can best explain the experimentally observed MIDs. This best fitting flux set is assumed to represent the intracellular fluxes. Since MIDs can only provide relative flux information (see above), $^{13}$C-MFA requires the measurement of at least one absolute flux, such as glucose consumption, to make the flux estimates absolute.

MIDs can be calculated if the metabolic network, its atom transitions, and its fluxes are known. Initially, the labeling of all isotopologues in the metabolic network was simulated, but as this is computationally very costly, diverse methods have been developed to optimize this simulation of isotope labeling systems. These methods include the cumomer,[120] fluxomer,[121] and elementary metabolite units[89] frameworks. With these approaches only a fraction of variables and computation time is necessary.

For $^{13}$C-MFA, usually an initial random flux vector $\nu$ is used to simulate the MIDs of the measured metabolites. These simulated MIDs are compared to the experimentally observed ones and fluxes are iteratively changed until they represent the measured MIDs. This flux

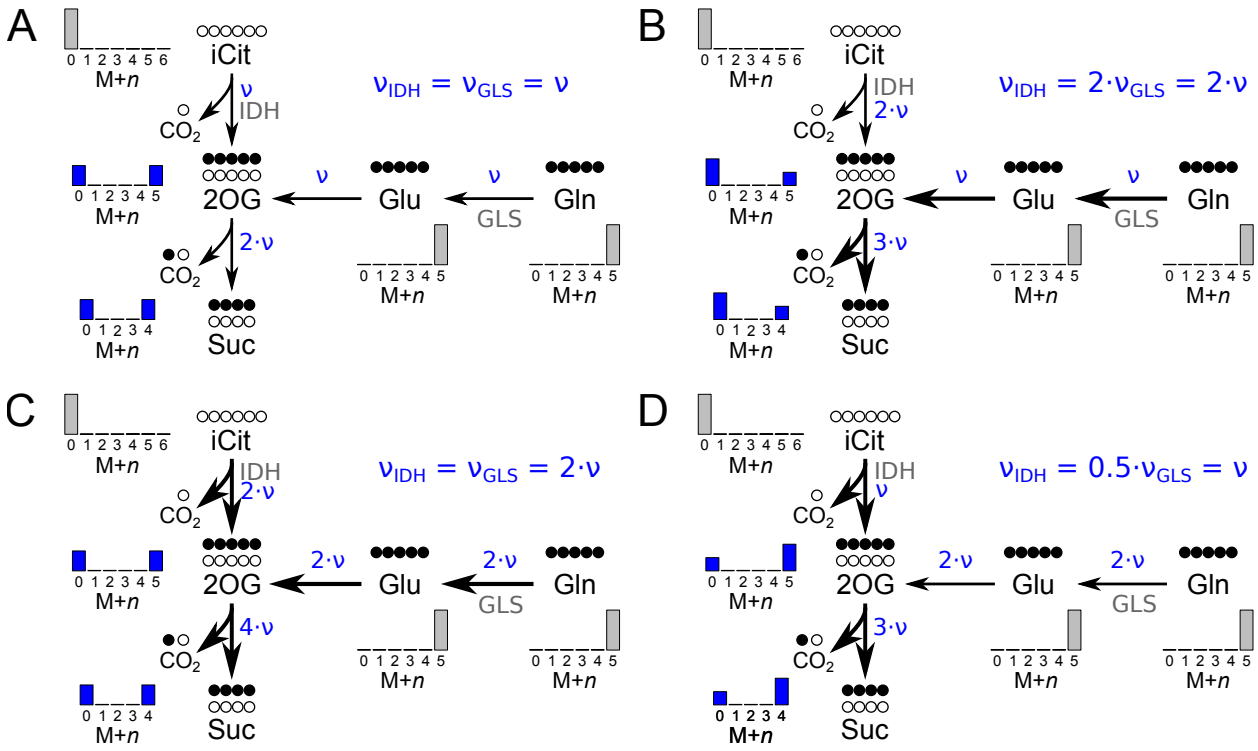



**Figure 10: MIDs are determined by metabolic fluxes.**
MIDs after stable isotope labeling with [U-$^{13}$C]glutamine in glutamine catabolism and a simplified section of the tricarboxylic acid (TCA) cycle. Only these isolated reactions are considered and treated as irreversible and in isotopic and metabolic steady state (metabolite labeling and MIDs are constant over time). Isotopic labeling is fully defined by the tracer and metabolic fluxes. MIDs alone can only provide relative flux information: If both IDH and GLS fluxes change proportionally, *i.e.* their ratio is constant, this change cannot be detected from the MIDs (A,C). However, if their flux ratio changes, the MIDs of downstream metabolites will change accordingly (A,B,D, blue MIDs). Grey MIDs do not depend on either IDH or GLS flux, and thus, are not informative.

fitting is performed by constrained non-linear optimization with an objective function similar to:

$$\min_{\nu} |X_{obs} - X_{sim}|$$

where $X_{obs}$ are the experimentally observed MIDs and $X_{sim}$ are the simulated ones. To account for measurement errors, the individual deviations in mass isotopomer abundances are usually weighted by the standard deviation of the measurements.[89,122,123] Several software implementations for $^{13}$C-MFA are available, like for example 13CFLUX2,[124] INCA,[59] and OPENFLUX.[122]

Once estimates of intracellular fluxes are available, the *in silico* simulation of MIDs can furthermore be used for experimental design of stable isotope labeling experiments to show which tracer, or which combination of tracers is ideally suited to resolve certain fluxes.[125]

### 1.3.9 Metabolic flux analysis techniques are highly targeted

The described MFA techniques are highly targeted and require a high amount of *a priori* knowledge. They require knowledge on active metabolic reactions and atom transitions ($^{13}$C-MFA) or an objective function (FBA). Although it is self-evident that no fluxes can be determined quantitatively without proper knowledge of the metabolic network, these prerequisites can be problematic: For many organisms or cell types the metabolic network is still not fully known and neither is the subcellular organization of metabolism in eukaryotic cells.[126]

Furthermore, current flux analysis techniques cannot account for unknown or unanticipated reactions or metabolites, and are therefore not suitable for discovery-based approaches, except in very well-studied organisms. However, non-targeted flux profiling techniques are needed for less well-studied systems to detect more subtle changes in metabolism that do not manifest in changed metabolite levels and, thus, cannot be analyzed using classical metabolomics approaches.

I address this issue in section A.9 and show how MIDs can also be used in a non-targeted flux profiling approach to detect metabolic flux changes without the need of a metabolic model.

## 1.4 Cancer metabolism

There were 14.1 million new cases of cancer diagnosed worldwide in the year 2012.[127] During the same time 8.2 million people died from cancer.[127] Although the incidence is decreasing, the absolute number of cancer patients is continuing to increase.[128] Besides the obvious individual hardships, the high cancer prevalence also poses a burden on healthcare systems. To alleviate these problems, a better understanding of pathological mechanisms is necessary to allow for both early diagnostics and effective therapies.

Cancer is usually characterized by abnormal cell division leading to tumor formation and cell migration (metastasis). These changes are accompanied by metabolic abnormalities which have been noticed already long ago.[129,130] Cancer cells were found to exhibit high rates of lactic acid production, even at high oxygen levels,[131] a high rate of glutaminolysis,[8,132,133] and increased *de novo* fatty acid biosynthesis.[13,134]

### 1.4.1 Cancer cells need biomass precursors and are often hypoxic

Tumor cells have a high demand of biomass precursors to maintain their high proliferation rate. Cell proliferation requires mostly amino acids for protein biosynthesis, nucleotides for DNA replication and transcription, and fatty acids for membrane lipids. Providing these

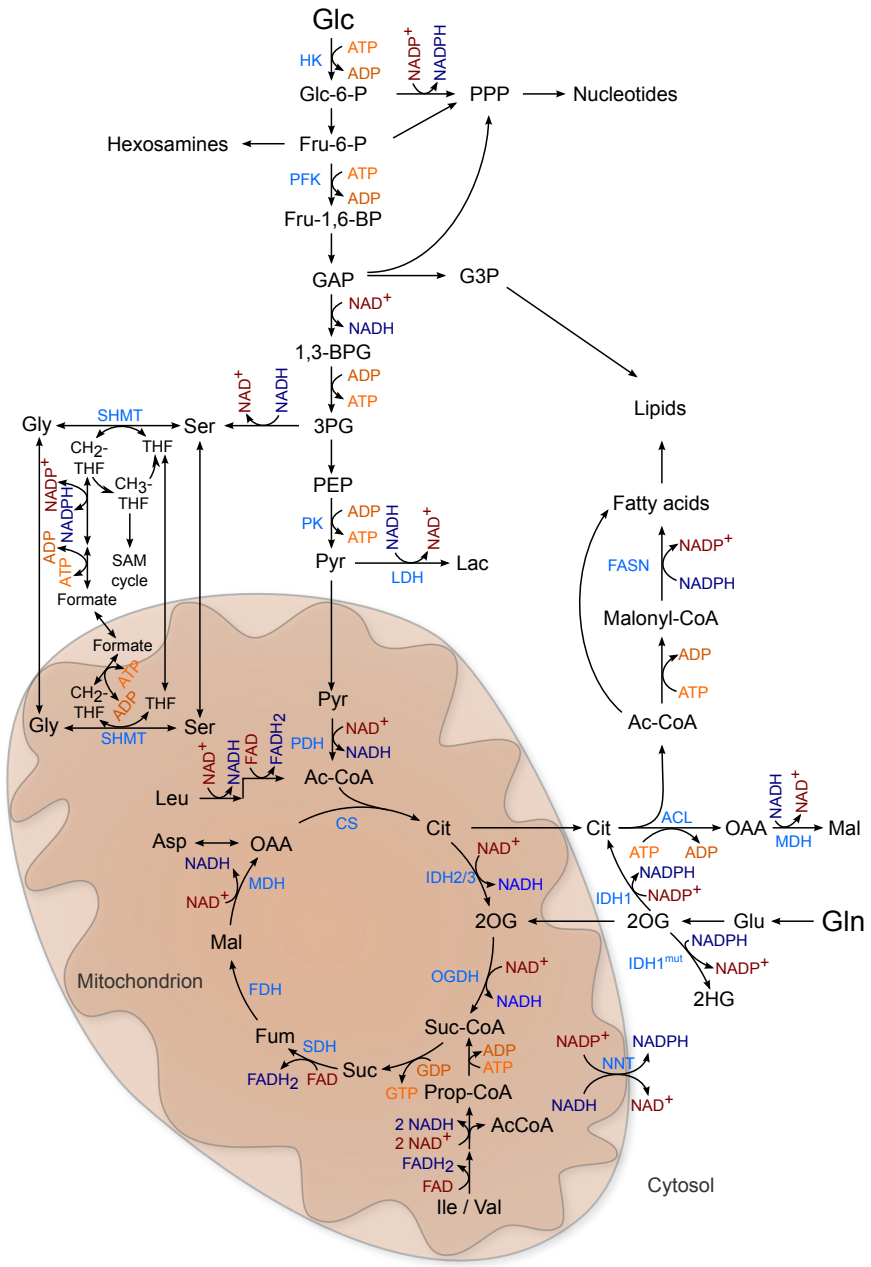



**Figure 11: Cancer cell metabolism.** Cancer cells require high amounts of anabolic precursors for proliferation. Glucose and glutamine are the major carbon sources. Glycolytic flux is induced to provide ATP and amino acid precursors. Lactate production and excretion is high to recover nicotinamide adenine dinucleotide $(NAD)^+$. Pentose phosphate pathway (PPP), nicotinamide nucleotide transhydrogenase (NNT) and one-carbon metabolism provide nicotinamide adenine dinucleotide phosphate (NADP)H. The TCA cycle is mostly decoupled from glycolysis through pyruvate dehydrogenase (PDH) inhibition, TCA cycle reactions are fueled by glutamine instead. Reductive glutamine metabolism provides acetyl-coenzyme A (CoA) for lipid biosynthesis, especially under hypoxic conditions. Branched-chain amino acid (BCAA) catabolism may replenish mitochondrial acetyl-CoA (details in Figure 12). Not all intermediates are shown.
Glc: glucose, Fru: fructose, P: phosphate, GAP: glyceraldehyde-3-phosphate, 1,3-BPG: 1,3-bisphosphoglycerate, 3PG: 3-phosphoglycerate, Pyr: pyruvate, G3P: glycerol-3-phosphate, Cit: citrate, 2OG: 2-oxoglutarate, Suc: succinate, Mal: malate, Fum: fumarate, OAA: oxaloacetic acid, HK: hexokinase, PFK: phosphofructo kinase, PK: pyruvate kinase.

precursors is energy demanding and requires large amounts of high energy substrates like ATP and NADPH. Metabolism of cancer cells needs to be adjusted to meet these demands. An additional metabolic burden of tumor cells is that they are usually exposed to a hypoxic microenvironment.[135] The fast cell proliferation excels vascularization and leads to low oxygen supply. Therefore, many tumor cells are exposed to oxygen levels of below 2%[136] which requires special metabolic adaptations.[137]

### 1.4.2 High glycolytic flux increases ATP production and provides catabolic precursors

The first metabolic feature noticed in cancer cells was their increased rate of lactic acid excretion.[131] Most cancer cells produce high amounts of lactic acid even under normoxic conditions when lactic acid fermentation usually is inhibited. Glycolytic ATP production is much less efficient than citric acid cycle reactions and oxidative phosphorylation.[138] However, when glucose is not limiting, or oxidative phosphorylation is already saturated, then increasing the glycolytic flux by means of lactic acid production is a way to increase the total ATP yield, but to the price of a lower efficiency.[139]

Especially under hypoxia, the increased expression of glucose transporters, lactate dehydrogenase (LDH), and monocarboxylate transporters (MCT) for lactate excretion increase the glycolytic flux by a push-pull-effect. Under hypoxia, oxidative phosphorylation is limited since there is a lack of $O_2$ which is needed as the final electron acceptor at complex IV of the electron transport chain. Therefore, the importance of glycolytic ATP generation is increasing. On the other hand, cellular respiration under hypoxia seems to be not only limited by $O_2$ availability but to be actively repressed[140]: Respiration is uncoupled from glycolysis and glucose carbon is directed away from the TCA cycle, preventing production of reactive oxygen species (ROS) at the respiratory chain.[141]

The high rate of lactic acid production lowers the intracellular pH and needs to be compensated. Therefore, expression of carbonic anhydrase is induced. The carbonic anhydrase enzyme is bound to the extracellular side of the plasma membrane where it catalyzes bicarbonate formation from carbon dioxide and water. The resulting bicarbonate is imported into the cell, is protonated to carbonic acid, and decomposes to water and $CO_2$, thereby effectively removing one proton per molecule of imported bicarbonate.[142]

### 1.4.3 Hypoxia inducible factor (HIF)-1 is a key regulator of cancer cell metabolism

The aforementioned metabolic adaptations are directly regulated by the HIF-1 transcription factor.[143] HIF-1-mediated expression of pyruvate dehydrogenase kinase (PDK)1 leads to

PDH inhibition[141] which, together with simultaneous LDHA expression,[14] leads to the high lactic acid production observed in cancer cells. HIF-1 also induces the monocarboxylate transporter to excrete lactate, and carbonic anhydrase to buffer the pH decrease caused by lactic acid production.[143]

HIF-1 was identified as a key mediator in the metabolic reprogramming of cancer cells, both under hypoxia and normoxia.[143–147] It is a heterodimeric transcription factor consisting of the HIF-1$\alpha$ and HIF-1$\beta$ subunits.[148] HIF-1 levels are highly sensitive to oxygen and inversely correlated with oxygen concentration.[149] Although HIF-1$\alpha$ is constitutively expressed, it is readily degraded in the presence of molecular oxygen. HIF-1$\alpha$ is targeted by $O_2$-dependent prolylhydroxylases, *i.e.* dioxygenases that use molecular oxygen to oxidize prolyl residues in proteins. The resulting hydroxyprolyl residues are then recognized by the von-Hippel-Lindau (VHL) protein which mark it for proteasomal degradation.[150,151] Defects in this HIF-1 regulation like loss of PHD enzymes[147] or loss of VHL[152] lead to constitutive stabilization of HIF-1 and are known to promote tumor growth.

### 1.4.4 Amino acids, nucleotides and NADPH are generated from glycolytic intermediates

The increased glucose uptake of cancer cells is not only used for ATP generation, but also for other biosynthetic precursors. Nucleotide biosynthesis requires ribose-5-phosphate which is produced via the PPP. The reductive part of the PPP and sometimes also the oxidative part is induced in cancer cells.[153,154]

Purine biosynthesis requires glycine and formyl-tetrahydrofolate (THF). These are intermediates of one-carbon metabolism and are produced from serine which is synthesized from the glycolytic intermediate 3-phosphoglyceric acid (Figure 11).[155] Besides the carbon contribution, one-carbon metabolism was also shown to provide a significant share of NADPH[156] which is required for fatty acid biosynthesis (see below). This explains the observation that high nutritional supplementation with folic acid, an important cofactor in one-carbon metabolism, can promote tumor progression.[157] Overall, one-carbon metabolism seems to play an important role in many cancers.[158]

### 1.4.5 Acetyl-CoA is produced from glutamine and acetate

Many biosynthetic precursors like fatty acids and several amino acids are produced from TCA cycle intermediates, but in cancer cells, glucose is mainly metabolized to lactate (see above). To compensate for the reduced pyruvate flux into the TCA cycle due to PDH inhibition, cancer cells show a high rate of glutamine consumption and glutamine provides a significant amount of carbon for their increased *de novo* fatty acid biosynthesis.[7,8,13,133,159] Furthermore,

acetate has recently been identified as a potential source of acetyl-CoA in cancer cells.[160–162]

Glutamine can be metabolized oxidatively (glutaminolysis)[132] or reductively (Figure 11). Reductive glutamine metabolism is caused by changing substrate concentrations that shift the IDH reaction equilibrium. Such a change can be caused by reduced 2-oxoglutarate dehydrogenase (OGDH) activity. HIF-1 induces the degradation of OGDH2[163] which leads to an accumulation of its substrate 2-oxoglutarate, shifting the equilibrium of the IDH reaction towards carboxylation of 2-oxoglutarate to isocitrate.[164] This isocitrate is then converted to citrate and cleaved by ATP-dependent citrate lyase (ACLY) to produce cytosolic acetyl-CoA for fatty acid biosynthesis.[165] The second product of the ACLY reaction is oxaloacetate. It is partially aminated to aspartic acid, which is needed for protein biosynthesis as well as for pyrimidine nucleotides, or reduced to malate, which can be transported back to the mitochondrium and is the substrate of malic enzyme to yield pyruvate.

The carboxylation of 2-oxoglutarate by IDH1 and IDH2 requires NADPH. It has recently been shown that a significant share of this NADPH is derived from one-carbon metabolism.[156] Also NNT, an enzyme catalyzing the hydride transfer from NADH to $NADP^+$, was found to be highly important in cancer cell metabolism, especially for reductive carboxylation.[159] Other contributors to NADPH production are malic enzyme and the oxidative branch of the PPP.[16]

### 1.4.6 Cancer cells exhibit increased *de novo* fatty acid biosynthesis

The high proliferation rate of cancer cells requires high amounts of fatty acids for the synthesis of membrane lipids. To satisfy this demand, cancer cells show both increased *de novo* fatty acid biosynthesis as well as fatty acid import.[137,166,167]

*De novo* fatty acid biosynthesis is catalyzed by the fatty acid synthase (FASN) complex. The FASN substrates are acetyl-CoA, malonyl-CoA, and NADPH. An initial acetyl-CoA molecule is elongated by $C_2$ units derived from malonyl-CoA. During each elongation step an NADPH-dependent reduction has to take place. The FASN complex stops elongation at the stage of palmitic acid (C16:0).[168] Longer chain fatty acids are produced by dedicated elongases which extend the chain with additional $C_2$ units.[169] In cancer cells, the FASN complex is induced, provides the major part of fatty acids, and has been identified as a potential therapeutic target.[12,13]

FASN only produces saturated fatty acids. However, membrane integrity relies on a proper ratio of saturated and unsaturated lipids.[170] Desaturation is performed by dedicated enzymes such as stearoyl-CoA desaturases (SCD). SCD1 was recently shown to play a key role in cancer-initiation.[171]

To maintain the high *de novo* fatty acid biosynthesis, cancer cell metabolism needs to provide sufficient acetyl-CoA, ATP and NADPH. In non-cancer cells most acetyl-CoA is

derived from the PDH reaction followed by citrate synthesis and citrate cleavage by cytosolic ACLY (Figure 11). However, in cancer cells and especially in hypoxic cancer cells, reductive glutamine metabolism is the major source of acetyl-CoA (see above). Secondary acetyl-CoA sources can be the catabolism of BCAAs, lysine, aromatic amino acids, ketone bodies or the $\beta$-oxidation of fatty acids, but so far there are no reports on their significance for fatty acid biosynthesis in cancer cells.

### 1.4.7 Branched chain amino acid catabolism in cancer

There is increasing evidence that BCAA metabolism plays an important role in cancer.[172,173] Recently, elevated plasma levels of the BCAAs leucine, isoleucine, and valine have been associated with a significantly increased risk to be diagnosed with pancreatic adenocarcinoma.[10] Furthermore, Tönjes *et al.*[174] found that branched-chain amino acid aminotransferase (BCAT)1 plays an important role in proliferation and metastasis of glioblastoma. BCAT1 is catalyzing the deamination of BCAAs and therefore a key enzyme of their catabolism (Figure 12).

Recently, leucyl-tRNA synthetase has been identified as the long sought-after metabolite sensor of the mammalian/mechanistic target of rapamycin complex 1 (mTORC1) signaling pathway.[175] Intracellular levels of leucine function as a proxy for the overall availability of anabolic substrates and for cellular energy status. MTORC1 activates diverse cellular anabolic processes and deactivates catabolic processes depending on the energy status.[176] This role makes mTORC1 central also to the proliferation of cancer cells and has therefore been identified as a therapeutic target.[177]

In mammals, the proteinogenic BCAAs are fully derived from diet and are essential. As for other amino acids, there is no cellular storage pool for BCAAs. Surplus BCAAs are therefore degraded. The degradation of BCAAs starts with their deamination (Figure 12).[16] The resulting $\alpha$-keto acids are oxidatively decarboxylated by branched-chain $\alpha$-keto acid dehydrogenase (BCKDH). During a subsequent reaction sequence borrowed from the $\beta$-oxidation of fatty acids, acetyl-CoA and acetoacetate or propionyl-CoA are generated. The toxic[178] propionyl-CoA, which is also generated from the degradation of odd-chain fatty acids,[17] is detoxified via the methylmalonate pathway and enters the TCA cycle as succinyl-CoA (Figure 12). All enzymes of BCAA metabolism are located in the mitochondria, except for BCAT, for which there exists both, a cytosolic (BCAT1) and a mitochondrial isoform (BCAT2).

The mechanistic significance of BCAA catabolism for cancer cells is not well understood. BCAT and BCKDH, the enzymes catalyzing the key steps, are highly regulated.[179,180] In respect to the high demand of anabolic substrates of cancer cells, the induction of BCAA degradation seems wasteful.[174] However, the degradation of BCAAs provides mitochondrial

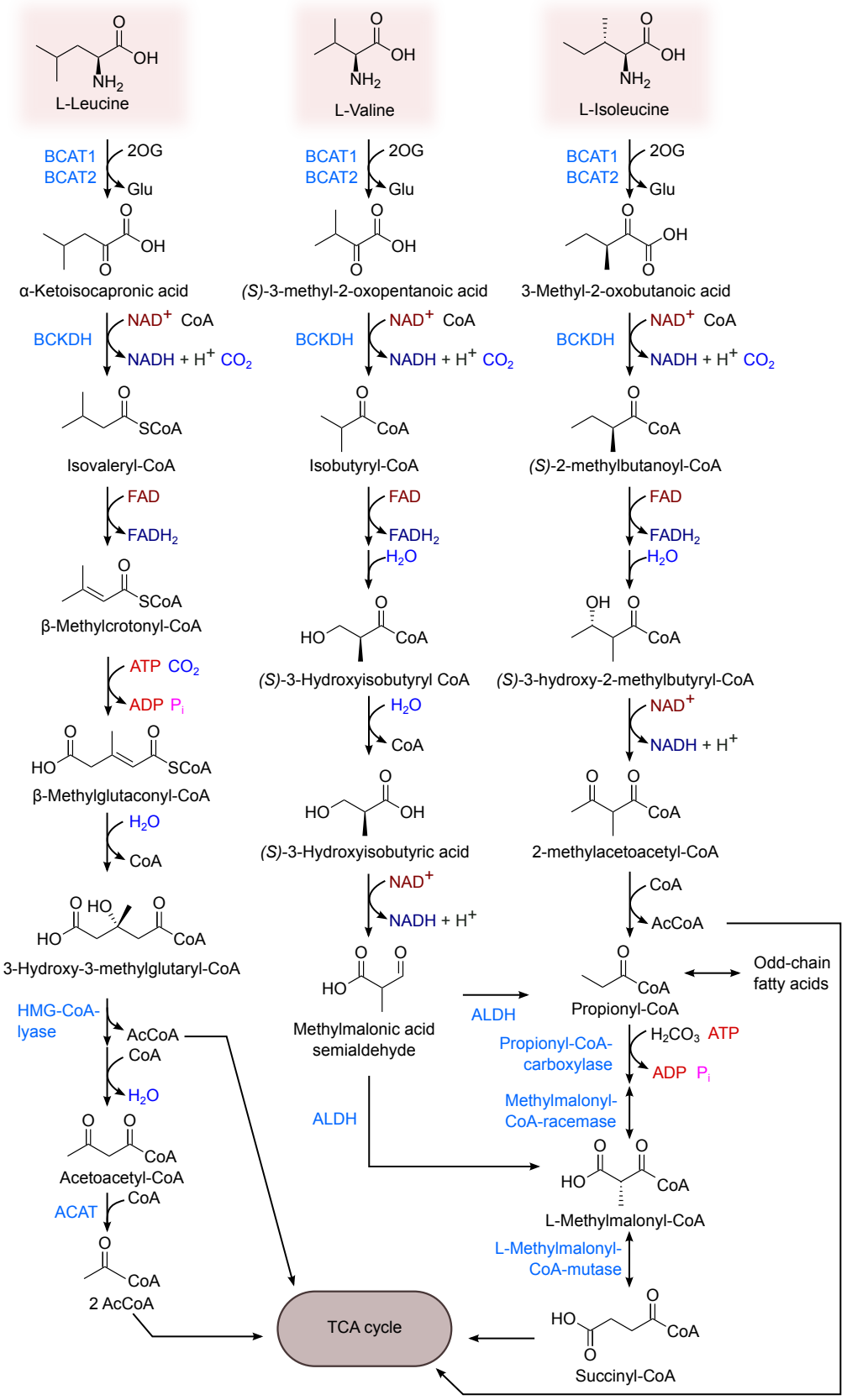



**Figure 12: BCAA catabolism provides acetyl-CoA and succinyl-CoA for the TCA cycle, as well as NADH and flavin adenine dinucleotide (FAD)$H_2$.**[16] After deamination and oxidative decarboxylation the branched-chain acyl-CoAs undergo $\beta$-oxidation. From valine and isoleucine propionyl-CoA is produced which can be detoxified via the methylmalonyl pathway and enter the TCA cycle as succinyl-CoA.

acetyl-CoA, NADH and $FADH_2$ which may be more important to tumor cells under certain circumstances, as for example when PDH is inhibited and therefore mitochondrial acetyl-CoA levels are low.

### 1.4.8 Summary: cancer metabolism

Much research has been done over the recent decades that has deepened our understanding of cancer cell transformation. Diverse risk factors have been identified, prospective biomarkers have been found,[10,11] and stable isotope labeling and metabolic flux analysis have provided interesting insights into cancer metabolism.[181] Mechanistic insights led to the discovery of various drug targets[13,15,143,182] and therapeutic agents have been developed.[14,15] This has improved the life of many patients and led to increasing survival rates.[128] However, the high cancer prevalence together with the high cost for treatment remains a significant burden for healthcare systems. There is still a long way to go and more research is required to allow for an earlier diagnosis and a more efficient treatment of cancer patients.

## 1.5 Objectives of this work

Despite major progresses towards the understanding of cancer, it still remains a wide-spread disease with an increasing number of cases every year.[128] Altered energy metabolism is an emerging hallmark of cancer cells (section 1.4).[183] Such alterations are necessary to meet the high demand of biosynthetic precursors required for rapid cell division. Additionally, the hypoxic microenvironments that many tumor cells are exposed to, require further metabolic adaptations.[135] Although most tumors form hypoxic microenvironments and many tumor cells are exposed to oxygen levels of below 2%,[136] pronounced tumor hypoxia is considered a bad prognostic sign for cancer patients, as hypoxia was shown to increase the motility of lung cancer cells *in vitro*[184] and to contribute to metastasis *in vivo*.[185] Therefore, the in-depth analysis of hypoxia-induced metabolic changes is highly relevant for the understanding of cancer cell metabolism and to advance cancer therapy.

Over the recent decades, targeted and non-targeted metabolomics techniques have been used to analyze metabolite levels and have provided valuable insights into cancer metabolism. Although metabolite levels are an adequate readout to detect for example biomarkers, metabolic flux changes may go unnoticed when analyzing only metabolite levels. Yet, metabolic fluxes are highly informative, because they are more sensitive to perturbations than metabolite levels and can provide a more mechanistic understanding of metabolism. Many recent advances in the understanding of cancer metabolism have been obtained from stable isotope labeling experiments.[181] However, current MFA approaches all have in common that they are very targeted and cannot account for unanticipated reactions. Therefore, such approaches miss potentially important but unexpected metabolic reactions.

Stable isotope labeling experiments can provide metabolic flux information, and there are algorithms for the non-targeted metabolome-wide detection of stable isotope labeling available (subsection 1.3.7). However, there have been only few non-targeted stable isotope labeling analyses focusing on metabolic fluxes, mostly due to the high complexity of the data and the lack of adequate software tools and workflows. Therefore, *the overall objective of this work was to develop tools and workflows for non-targeted stable isotope labeling analysis, and to apply them to analyze the metabolism of human cancer cells.*

To fulfill this objective, the first step was to develop a tool to detect and comparatively analyze isotopic enrichment across multiple datasets. This tool should implement novel workflows to 1) detect metabolic flux changes by differential MID analysis, and 2) to estimate and visualize metabolic proximity between different compounds by analyzing the similarity of their MIDs.

This software was then to be applied to gain biological insights into cancer cell metabolism, specifically to analyze metabolome-wide stable isotope labeling data in a non-targeted manner to reveal metabolic flux changes in response to hypoxia and to detect unexpected metabolic

reactions.

# 2 Materials & methods

Materials and methods not provided here, are described in the manuscripts in the Appendix. Unless specified differently, all chemicals have been obtained from Sigma-Aldrich.

## 2.1 Used organisms and culture conditions

Yeast strains and culture conditions used in section 3.4 are described in section A.7. Cell lines and culture conditions used in section 3.5 are described in section A.9 and the dissertation of Dr. Cordes[186].

## 2.2 Metabolite measurements

### 2.2.1 Metabolite extraction

We published a detailed description of the metabolite extraction from cell culture used for section 3.5 in Sapcariu *et al.*[187] (section A.5). Metabolite extraction from yeast is described in section A.7.

### 2.2.2 Fatty acid methyl ester preparation

For the analysis of total cell lipids as fatty acid methyl esters (FAMEs), the organic phase forming during metabolite extraction (section A.5) was left to dry at room temperature or dried in a vacuum centrifuge. For transesterification 500 µl 0.5 M NaOH in methanol were added and incubated for 1 h at room temperature. The solution was acidified using 2 drops of 6 M hydrochloric acid. FAMEs were extracted by adding 300 µl of hexane and thorough mixing. For GC-MS analysis, 250 µl of the (upper) hexane phase were transferred to a glass vial with micro insert and the vial was sealed air tight. If not analyzed directly, extracts were stored at −80 °C.

### 2.2.3 GC-EI-MS

#### Full scan measurements of polar metabolites

Derivatization conditions and v parameters for the analysis of methoxyamine and MSTFA derivatized polar metabolite samples are described in section A.4 and A.5.

#### SIM measurement for MID determination

For targeted MID analysis, polar metabolite extracts, derivatized with methoxyamine and TBDMS, were measured in a dedicated 26 min GC-MS selected ion monitoring (SIM) method.

Metabolite extracts were derivatized using a multi-purpose sampler (GERSTEL). Dried samples were dissolved in 15 µl pyridine containing 20 mg ml$^{-1}$ methoxyamine hydrochloride, and incubated under shaking for 60 min at 55 °C. After adding 15 µl MTBSTFA, samples were incubated for additional 60 min at 55 °C under continuous shaking.

GC-MS analysis was performed using an Agilent 7890A GC coupled to an Agilent 5975C inert XL mass selective detector (Agilent Technologies). A sample volume of 1 µl was injected into a split/splitless inlet, operating in splitless mode at 270 °C. The GC was equipped with a 30 m DB-35MS capillary column and a 5 m DuraGuard capillary in front of the analytical column (Agilent J&W GC Column). Helium was used as carrier gas with a constant flow rate of 1 ml min$^{-1}$. The GC oven temperature was held at 100 °C for 2 min, increased to 300 °C at 10 °C min$^{-1}$ and held for 4 min. The transfer line temperature was set constantly to 280 °C. The MS was operating under electron ionization at 70 eV. The MS source was held at 230 °C and the quadrupole at 150 °C. SIM parameters are listed in Table 2.

## 2.3 Molecular biology

### 2.3.1 RNA extraction and reverse transcription

Total RNA extraction from interphases formed during metabolite extraction (section A.5) was performed using an RNeasy Mini Kit (Qiagen) according to the manufacturer's instructions. RNA concentration was determined using a NanoDrop (Thermo Scientific). RNA samples were stored at −80 °C.

For reverse transcription 0.1 µg of total RNA were combined with 1 µl of oligo-dT (50 µM) and 1 µl of dNTP mix (10 µM) in a 1.5 ml tube and filled with RNase free water to 13 µl. RNA was denatured at 65 °C for 5 min, cooled down on ice for 1 min and centrifuged. A master mix was prepared of 4 µl 5×First-Strand Buffer (Invitrogen), 1 µl 0.1 M dithiothreitol, 1 µl SuperScript III RT (200 U µl$^{-1}$; Invitrogen) and 1 µl RNase OUT (40 U µl$^{-1}$; Invitrogen) per sample and 7 µl were added to the denatured RNA sample. After reverse transcription

**Table 2:** Settings for SIM-method for TBDMS derivatives of intermediates of central carbon metabolism. The monitored fragments contain the full carbon backbone of the respective metabolite. The provided sum formula was used for correction of MIDs for natural isotope abundance.

| **Compound** | **Ions** | **Exact ion** | **dwell time (ms)** | **Sum formula** | **RT (min)** |
|---|---|---|---|---|---|
| Pyruvic acid 1MeOX 1TBDMS | 174–180 | 0 | 15 | $C_6H_{12}O_3NSi$ | 6.62 |
| Lactic acid 2TBDMS | 261–267 | 0.1 | 15 | $C_{11}H_{25}O_3Si_2$ | 8.35 |
| Alanine 2TBDMS | 260–266 | 0.1 | 15 | $C_{11}H_{26}NO_2Si_2$ | 8.89 |
| Glycine 2TBDMS | 246–252 | 0.1 | 15 | $C_{10}H_{24}NO_2Si_2$ | 9.40 |
| $\gamma$-Aminobutyric acid 2TBDMS | 274–281 | 0.2 | 15 | $C_{12}H_{28}NO_2Si_2$ | 11.74 |
| Succinic acid 2TBDMS | 289–296 | 0.1 | 15 | $C_{12}H_{25}O_4Si_2$ | 12.04 |
| Fumaric acid 2TBDMS | 287–294 | 0.1 | 15 | $C_{12}H_{23}O_4Si_2$ | 12.18 |
| Serine 2TBDMS | 390–396 | 0.2 | 15 | $C_{17}H_{40}NO_3Si_3$ | 13.54 |
| Methionine 2TBDMS | 320–328 | 0.2 | 10 | $C_{13}H_{30}NO_2SSi_2$ | 14.17 |
| 2-Oxoglutaric 1MeOX 2TBDMS | 346–354 | 0.2 | 10 | $C_{14}H_{28}NO_5Si_2$ | 14.86 |
| Malic acid 3TBDMS | 419–426 | 0.2 | 15 | $C_{18}H_{39}O_5Si_3$ | 15.06 |
| Aspartic acid 3TBDMS | 418–425 | 0.2 | 15 | $C_{18}H_{40}NO_4Si_3$ | 15.48 |
| Glutamatic acid 3TBDMS | 432–440 | 0.3 | 10 | $C_{19}H_{42}NO_4Si_3$ | 16.56 |
| Glutamine 3TBDMS | 431–439 | 0.3 | 10 | $C_{19}H_{43}N_2O_3Si_3$ | 18.22 |
| Citric acid 4TBDMS | 591–600 | 0.3 | 10 | $C_{26}H_{55}O_7Si_4$ | 19.17 |

during 60 min incubation at 50 °C, the reaction was stopped by 15 min incubation at 70 °C. Samples were diluted with 180 µl of RNase-free water and stored at −20 °C.

### 2.3.2 qPCR

For quantitative real-time polymerase chain reaction (qPCR) 2 µl of diluted cDNA (see above) were mixed with 10 µl 2× SYBR-green, 7 µl water, 0.5 µl of each 10 pmol µl$^{-1}$ forward and reverse primer. qPCR was carried out using a LightCycler 480 II (Roche) operated with the LightCycler 480 software (Version 1.5.x.x) in 96- or 384-well plates using the following program:

| Step | Temperature | Duration | Heating | Repetitions |
|---|---|---|---|---|
| Initial denaturation | 95 °C | 5 min | 4.4 °C s$^{-1}$ | 1 |
| Amplification | | | | 44 |
| Denaturation | 95 °C | 30 s | 4.4 °C s$^{-1}$ | |
| Annealing | 60 °C | 30 s | 2.2 °C s$^{-1}$ | |
| Elongation | 72 °C | 30 s | 4.4 °C s$^{-1}$ | |
| Melting curve | | | | 5 |
| | 95 °C | 15 s | 4.4 °C s$^{-1}$ | |
| | 40 °C | 1.5 min | 2.2 °C s$^{-1}$ | |
| | 95 °C | | | |

HPLC-purified primers have been purchased from Eurogentec. Primer sequences used for gene expression analysis are listed in Table 3. Gene expression levels were determined relative to the ribosomal protein L27 gene (RPL27) as internal standard:

$$\text{Expression of } \textit{target} \text{ relative to L27} = 2^{Ct_{L27} - Ct_{target}}$$

**Table 3:** QPCR primer sequences for human genes (sense strand "fwd" and anti-sense strand "rev").

| Gene symbol | | Sequence |
|---|---|---|
| ACLY | fwd | 5′-AATTTCAGAGCAGACGGGCA-3′ |
| | rev | 5′-GACCCCAACGAGACCAAGTT-3′ |
| ACSS1 | fwd | 5′-GAAGTGAGGATCACCTACAGGG-3′ |
| | rev | 5′-ACCTTGCACTTGGCATCATTG-3′ |
| ACSS2 | fwd | 5′-GCCACATGTTGACTCCCCTT-3′ |
| | rev | 5′-GGCTGCTTGAACACCAGATA-3′ |
| ASPA | fwd | 5′-ACCCCCGGGATGAAAATGGA-3′ |
| | rev | 5′-AGTGGGATCGTCTTCCCATC-3′ |
| BCAT1 | fwd | 5′-GCTCTGGTACAGCCTGTGTT-3′ |
| | rev | 5′-GCACAATTGTCCAGTCGCTC-3′ |
| BCAT2 | fwd | 5′-CTCAACATGGACCGGATGCT-3′ |
| | rev | 5′-AGCAACTCCAGCTTGTCGAA-3′ |
| BCKDHA | fwd | 5′-CTTCAACTTCGCTGCCACAC-3′ |
| | rev | 5′-TCCACGCGGATTGACATGAT-3′ |

**Table 3:** QPCR primer sequences for human genes. *(continued)*

| Gene symbol | | Sequence |
|---|---|---|
| BCKDHB | fwd | 5′-GATCAAAACAGGGCGACTGC-3′ |
| | rev | 5′-CGAAGGGCATCATAACACTTCC-3′ |
| CS | fwd | 5′-TTCCGACCCTTACCTGTCCT-3′ |
| | rev | 5′-ATAGCCTGGAACAACCCGTC-3′ |
| GPD1 | fwd | 5′-CAACCATTGGCTGCAAGGAC-3′ |
| | rev | 5′-CCCACGGCCACTACATTCTT-3′ |
| GPD1L | fwd | 5′-AGCGCTGGGAATCACCCTC-3′ |
| | rev | 5′-GCTCCCACAGCTACGATGTTC-3′ |
| GPD2 | fwd | 5′-GATGACGATCGCTGGCACTA-3′ |
| | rev | 5′-AGAGGACGGATTCCACTCCA-3′ |
| RPL27 | fwd | 5′-CTGGTGGCTGGAATTGAC-3′ |
| | rev | 5′-ACAGAGTACCTTGTGGGC-3′ |
| NAT8L[188] | fwd | 5′-TGTGCATCCGCGAGTTCCGT-3′ |
| | rev | 5′-CGGAAGGCCGTGTTAGGGAT-3′ |
| NAT8L exon 1 | fwd | 5′-GTGCATCCGCGAGTTCCG-3′ |
| | rev | 5′-CTCCATGATGCCGTCGTAGAA-3′ |
| NAT8L exon 2 | fwd | 5′-CTACTACAGCCGCAAGGTGAT-3′ |
| | rev | 5′-TGTAGTACTGCTCGATGTCCG-3′ |
| PCCA | fwd | 5′-CGGATGAGGCTGTCTGTGTT-3′ |
| | rev | 5′-AACGACATCTTCTGCTGCCA-3′ |
| PCCB | fwd | 5′-CATCTGTGAAAGGGGCTCGT-3′ |
| | rev | 5′-CTGTGCTGTGCCAGGTAGAA-3′ |
| PPARA | fwd | 5′-GCAAACTAAGACCTGGGGAGG-3′ |
| | rev | 5′-CCTCCTCCTCACATTTGACCC-3′ |

### 2.3.3 Cell transfection

For knockdown experiments, small interfering RNAs (siRNAs) were reverse transfected into A549 cells using Lipofectamine RNAiMAX (Invitrogen/Life Technologies) according to the

Invitrogen protocol.[1]

5 nmol siRNA were resuspended in 250 µl water and stored at −80 °C. For each well (12-well plate), 20 pmol (1 µl) siRNA were diluted in 200 µl Opti-MEM I reduced-serum medium (Invitrogen/Life Technologies), supplemented with 2.5 µl Lipofectamine RNAiMAX, gently mixed, and incubated for 20 min at room temperature. The prepared solution was spread in a well 5 min before 25,000–100,000 cells in 800 µl of DMEM 5796 growth medium containing 10% FBS were added. The plate was gently mixed and incubated (37 °C, 5% $CO_2$) for up to 72 h until further use. For 6-well plates, the volumes were doubled.

ON-TARGETplus non-targeting, ACLY, and *N*-acetyltransferase 8-like (NAT8L) siRNA were obtained from Dharamacon/GEHealthcare. The target sequences of the used siRNAs are listed in Table 4.

**Table 4:** siRNA target sequences for gene silencing as part of ON-TARGETplus smart pools (Dharamacon/GEHealthcare).

| **Gene symbol** | **Target sequence** |
|---|---|
| *non-targeting* | 5′-UGGUUUACAUGUCGACUAA-3′ |
| | 5′-UGGUUUACAUGUUGUGUGA-3′ |
| | 5′-UGGUUUACAUGUUUUCUGA-3′ |
| | 5′-UGGUUUACAUGUUUUCCUA-3′ |
| ACLY | 5′-GCACGAAGUCACAAUCUUU-3′ |
| | 5′-CGAGUGAAGUCGAUAAACA-3′ |
| | 5′-GAGAGCAAUUCGAGAUUAC-3′ |
| | 5′-CCACUCCUCUGCUCGAUUA-3′ |
| NAT8L | 5′-CUUUAAUUCUUGGGACAAA-3′ |
| | 5′-CGGACAUCGAGCAGUACUA-3′ |
| | 5′-GUUUGUACCCUAAGACACA-3′ |
| | 5′-AGUUCGCCGUGGUGCACAA-3′ |

## 2.4 Software

Agilent/HP ChemStation was used for data acquisition from GC-EI-MS measurements and for the export to the open netCDF format. Further data analysis was performed in

[1] http://www.lifetechnologies.com/content/dam/LifeTech/migration/en/filelibrary/pdf/protocols.par.55829.file.dat/a549-rnaimax.pdf

MetaboliteDetector[189]. Correction of MIDs for natural isotope abundance was also performed within MetaboliteDetector for targeted analyses or with NTFD[102,103] and MIA (section A.8). Third party libraries used in MIA are described in the application note (section A.8); those for FFC in Wegner *et al.*[26]. FFC was used to determine elemental composition and retained atoms in mass spectrometric fragments. Further statistical analysis was performed in the R statistical environment[190] and in OpenOffice Calc. Figures were created using Inkscape, R, and MIA.

# 3 Results & discussion

Parts of my work and contributions to joint projects are described in previous publications[2,6,26,33,103,187,191] which are reproduced in the Appendix A, along with manuscripts that are in preparation for submission. Here, I will summarize the motivation and findings of the respective projects and provide additional results and discussion not included in those publications.

In sections 3.1, 3.2, and 3.3, I will present the more technical or theoretical aspects of my work and the tools that have been developed. In section 3.5, I will show how I successfully applied these tools for the analysis of cancer cell metabolism. In section 3.6, I will present MetaboBase, a metabolomics data management platform I designed.

## 3.1 NTFD — a tool for the non-targeted detection of isotopic enrichment

Until recently, the analysis of isotopic enrichment of metabolites after stable isotope labeling experiments has been highly targeted. However, as described in the introduction, NTFD[102] was developed to address this problem. NTFD is an algorithm for the non-targeted detection of isotopic enrichment (subsection 1.3.7). To facilitate its application, we developed and published a user-friendly software implementation[1] of this algorithm for the analysis of GC-MS data in Hiller *et al.*[103] (section A.2). After selecting recorded GC-MS data and setting the parameters for the detection of isotopic enrichment, NTFD provides a list of all enriched compounds and MIDs of all isotopically enriched mass spectrometric fragments. If replicate measurements are available, NTFD can also provide quality measures of MID calculation like confidence intervals or coefficients of determination ($R^2$). All workflows in NTFD are embedded into an intuitive graphical user interface (GUI) (Figure 13).

We described both the NTFD algorithm as well as the software implementation in more detail in a book chapter where we also presented diverse applications of non-targeted stable isotope labeling analyses (section A.10). Among other applications, NTFD is a valuable tool for the non-targeted profiling of metabolic capabilities of different organisms. While genome or transcriptome analysis can provide the theoretical metabolic capabilities, it cannot predict

[1] Available at: http://ntfd.mit.edu/

whether a certain pathway is active or carries a significant flux under the given condition. Even if an enzyme is produced, it can still be inactive due to post-translational modifications. However, all compounds that are isotopically enriched after incubating an organism in the presence of a certain isotopic tracer, are metabolically connected to that tracer. Therefore, it can be concluded that one or multiple pathways connecting these compounds are active under the given conditions. The labeling patterns of these compounds may further help to distinguish between different metabolic pathways leading to the same product. Such a tracer fate detection is also interesting for pharmacokinetic analyses, since it can reveal the metabolic fate of an isotopically enriched drug.[192]

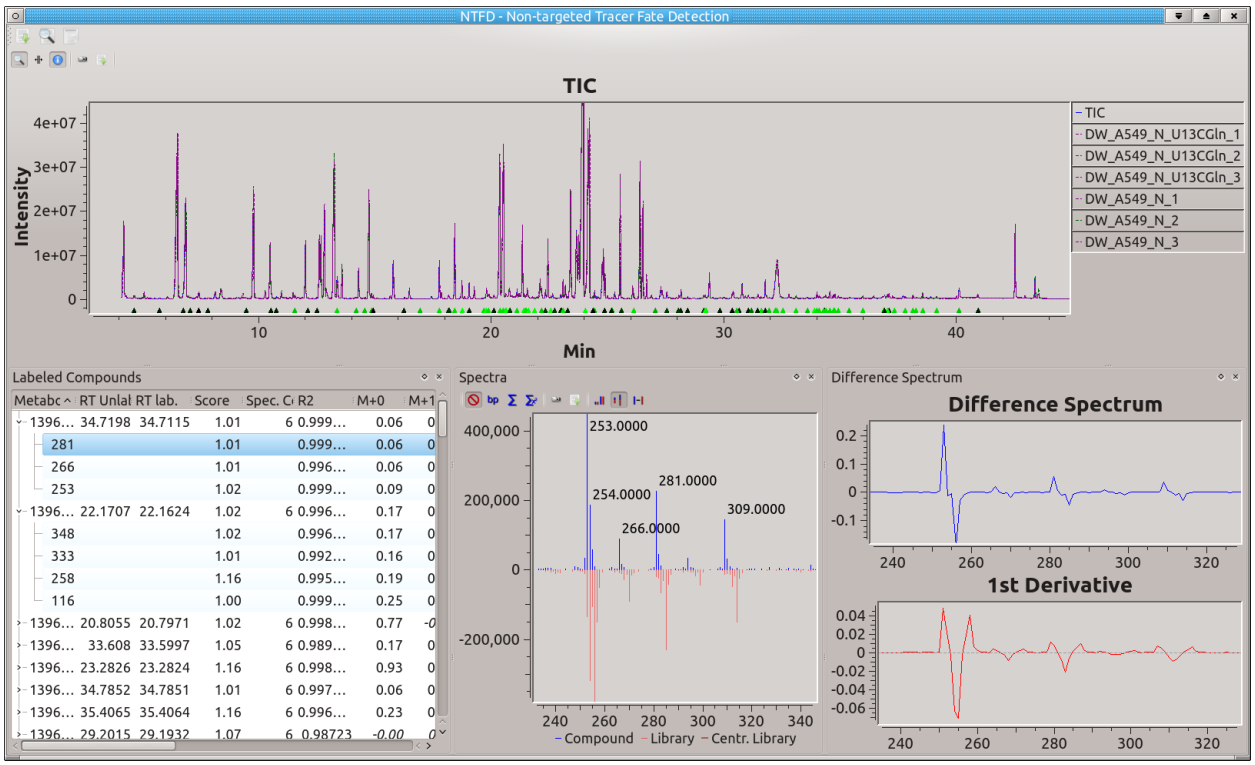



**Figure 13: The graphical user interface of NTFD for the non-targeted detection of isotopic enrichment.** Labeled and unlabeled chromatograms with isotopically enriched compounds are shown (top). MIDs for all isotopically enriched mass spectrometric fragments are presented (bottom left). The labeled and unlabeled mass spectra, their difference spectrum, its first derivative are shown to verify the results (bottom center, bottom right).

## 3.2 MIA — a tool for the non-targeted analysis of isotopic enrichment

The analysis of stable isotope labeling data can provide much information on active metabolic pathways within a cell under any given condition. Furthermore MIDs can be compared across different experimental conditions to detect changes in metabolic fluxes (subsection 3.2.1). If isotopic enrichment is detected in a non-targeted manner, this approach provides the means for qualitative non-targeted metabolic flux analysis. However, so far there have not been many studies applying truly non-targeted isotope labeling analyses to obtain biological insights. Most analyses stopped after the qualitative detection of isotopic enrichment [101] or detected changes in MIDs but did not trace them back to the underlying metabolic fluxes. [50] To my knowledge, none of the current studies that employ non-targeted detection of isotopic labeling did consider the MIDs of unidentified compounds for data analysis. This is mostly due to the lack of adequate software and workflows for such non-targeted analyses. [109]

Although NTFD is a powerful tool for the non-targeted detection of isotopic enrichment, it lacks data analysis and visualization capabilities. Moreover, there is a general lack of adequate tools for the analysis of non-targeted stable isotope labeling data. For all isotopically enriched compounds, NTFD provides the RI or retention time (RT), the $m/z$ range, and MIDs for all labeled fragments. Already for a simply designed experiment with only two groups a differential analysis can be tedious: All mass spectra of the isotopically enriched compounds need to be compared between the groups to correctly match the MIDs. Once they are matched, the relative mass isotopomer abundances need to be compared for each fragment. For experiments involving multiple tracers or treatments this becomes relatively complex. To address the urgent need for adequate tools for the non-targeted analysis of stable isotope labeling data, I developed MIA (Figure 15).

The objective of MIA is to facilitate the analysis of non-targeted stable isotope labeling data from a single or multiple experiments and at the same time minimizing user intervention. The goal was to have an easy-to-use GUI that visualizes stable isotope labeling across multiple datasets and implements novel data analysis workflows (subsection 3.2.1, 3.2.2, Figure 14).

I implemented MIA in C++ to make use of the MetaboliteDetector- and NTFD-framework and to allow for a fast detection of isotopic enrichment. Most computationally expensive steps are parallelized to efficiently analyze bigger datasets. All results generated with MIA can be exported as either raw MID data or as vector graphics. MIA is well integrated with MetaboliteDetector which can easily be used for downstream analyses.

In the course of developing MIA, I detected some weaknesses in the detection of isotopic enrichment in NTFD. To overcome these problems, I modified its label detection algorithm to increase both sensitivity and specificity (subsection 3.2.3).

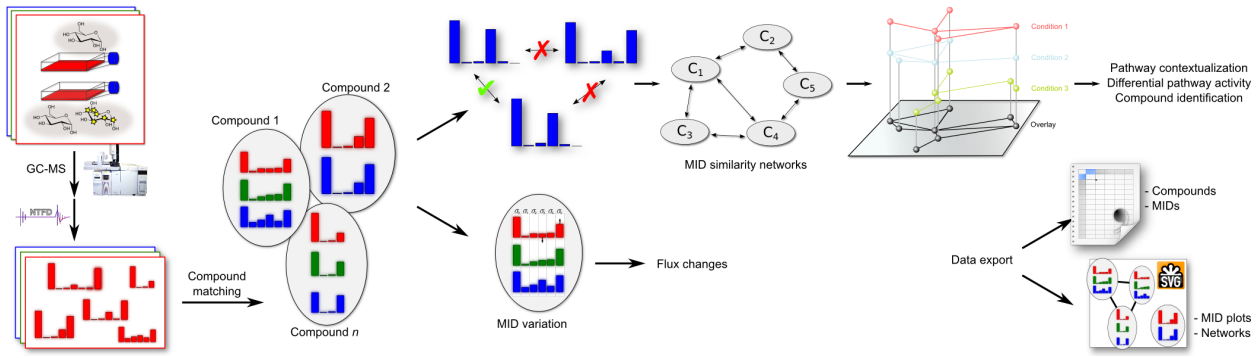



**Figure 14: MIA workflow.** After performing a stable isotope labeling experiment isotopic enrichment is detected and quantified in a non-targeted manner. Labeled compounds are matched across all datasets and MIDs are visualized. Data can be filtered and analyzed by changes in MIDs which indicate metabolic flux changes or MIDs similarity can be visualized to reveal metabolic similarity of the isotopically enriched compounds. Finally, MIDs can be exported either as spreadsheet or vector graphics for further use.

In section 3.5, I demonstrate how MIA is a valuable tool for non-targeted mass isotopolome analysis, helps to detect changes in metabolic fluxes and aids compound identification.

### 3.2.1 MID variation analysis to detect changes in metabolic fluxes

After applying a stable isotope labeled tracer and under the assumption of isotopic steady-state, the MIDs of all compounds in a metabolic network are a function of the fluxes through this network (section 1.3.8). Thus, changes in MIDs across different experimental conditions can only be a consequence of altered metabolic fluxes. These changes in metabolic fluxes can therefore be detected by analyzing changes in the mass isotopolome. Such a mass isotopomer abundance variation analysis can be performed with MIA.

To detect changes in MIDs, the MIDs of identical compounds are matched across different experimental conditions. Each mass isotopomer is then analyzed to detect significant differences in its relative abundance. As a measure of variation, for each isotopically enriched compound, I calculated the maximal standard deviation of relative mass isotopomer abundance across the different experimental conditions. I assumed that large flux changes will lead to large changes in mass isotopomer abundance. Thus, to find the most significant flux changes, metabolites are ranked by their variation score. Like any MID analysis, this approach is subject to the limitations described in section 1.3.8. Apart from that, this systematic analysis of mass isotopomer abundance variation detects flux changes without the requirement of any biochemical *a priori* knowledge on the system of interest. It will consider any unanticipated reactions or metabolites which cannot be accounted for in current flux analysis techniques, and it is only biased by analytical restrictions and the choice of the isotopic tracer.

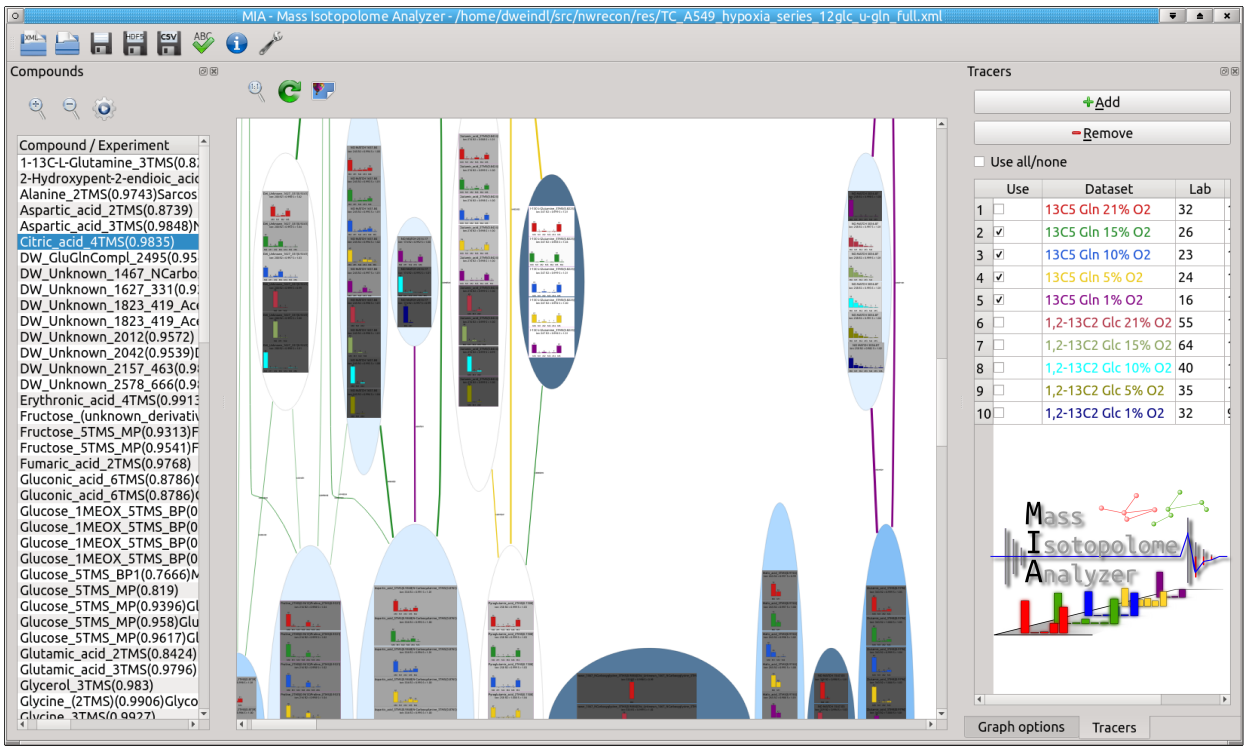



**Figure 15: The graphical user interface of MIA, a tool for the non-targeted analysis of isotopic enrichment.** Different datasets can be analyzed together (right panel). MIDs of all labeled compounds (left panel) are presented as barplots and MID similarity is visualized as a network (center).

### 3.2.2 MID similarity analysis to detect metabolically connected compounds

For the interpretation of the detected changes in mass isotopomer abundances (subsection 3.2.1) the respective compounds need to be identified. This is usually done by matching their mass spectra against reference libraries. However, the available libraries are far from comprehensive, in fact, compound identification is a major bottleneck in current metabolomics research (subsection 1.2.3). Without identification of at least their compound classes or associated pathways, these features can provide only very limited insights. Hence, compound identification is, however cumbersome, still highly important.

For compounds that are not present in reference libraries, other means are needed for identification. Here I will describe an approach based on MID-similarity. As described in section 1.3.8, the MID of any metabolite is determined by the MIDs of its precursors and the flux ratios of any producing reactions. Within linear pathways, the MIDs of all compounds are identical under isotopic steady-state. MIDs are only changing if isotopically enriched fragments are added or removed from the molecule, or when the metabolite pool is diluted by a converging flux.

The reverse conclusion, that compounds with identical or highly similar MIDs are more

likely to be part of the same pathway, is employed to aid compound identification. By analyzing the MID similarity of different compounds, they can be grouped by metabolic pathways. Strictly speaking, this high MID similarity is only granted in linear pathways. However, in converging pathways, if one flux is much larger than the other, or there is only a dilution with the unlabeled isotopologue, then the labeling pattern of the dominating precursor is mostly conserved in the product MID and the reaction sequence can be seen as quasi-linear pathway. In this case, the MID similarity is still significant. Empirically, this is the case for many metabolic reactions. Even in highly branched and interconnected pathways, such as the TCA cycle, the MIDs of its intermediates can be very similar (Figure 16).

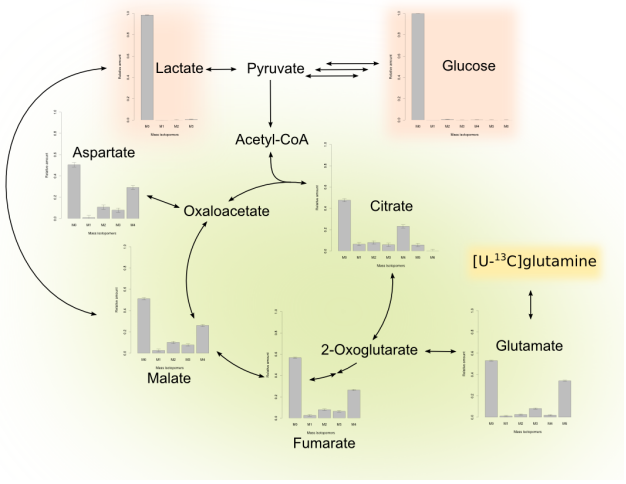



**Figure 16: Metabolic vicinity leads to similar MIDs.** Closely related compounds show high MID similarity (green). Labeling of human A549 lung cancer cells with [U-$^{13}$C]glutamine tracer (yellow box). Unlabeled compounds are not produced from the applied tracer (red boxes).

To use MID similarity for pathway contextualization of unidentified compounds, MIA pairwisely compares the MIDs of all isotopically enriched compounds (Figure 17A). This pairwise comparison results in a distance or similarity matrix. After applying an empirically determined distance cut-off, MIA creates a network of compounds with higher MID similarity. The resulting graph is likely to show metabolically connected compounds. However, the MID similarity can, dependent on the applied tracer and metabolic pathways, be ambiguous. The specificity can be increased by using distinct tracers and multiple experimental conditions (Figure 17A). Edges in the graph occurring in multiple conditions are more likely to be biologically meaningful.

Reactions that lead to losses or additions of isotopically enriched fragments change the MIDs of the respective molecules (Figure 10). To account for such losses or additions of isotopically enriched fragments which reduces MID similarity and could conceal metabolic proximity, MIA performs a Needleman-Wunsch alignment[193] on the MID vectors prior to the similarity calculation (Figure 17B). This step is for example necessary to reveal the high

similarity and metabolic proximity of isocitric acid and 2-oxoglutaric acid, or 2-oxoglutaric acid and succinic acid as shown in Figure 10. As a similarity measure, I chose the Canberra distances of all pairwisely aligned MIDs. The Canberra distance of two MID vectors $A$ and $B$ was calculated as

$$d_{A,B} = \sum_{i=1}^{n} \frac{|A_i - B_i|}{|A_i| + |B_i|} \tag{3.1}$$

and normalized by the sum of the dimensions of the MID vectors:

$$d_{A,B}^{norm} = \frac{d_{A,B}}{\dim A + \dim B} \tag{3.2}$$

By using the Canberra distance, small absolute differences are weighted more than when using *e.g.* the Euclidean distance. The normalization to the MID length accounts for the fact, that for longer MID vectors, the summed absolute differences are bigger than for shorter MID vectors.

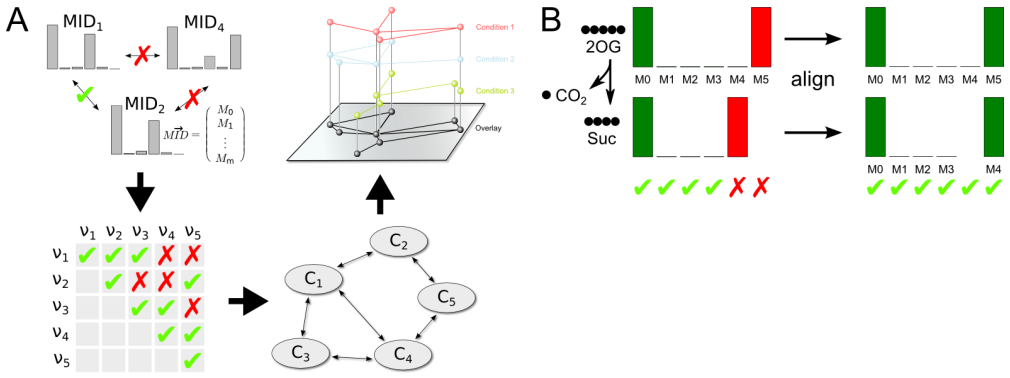



**Figure 17: MID similarity analysis for pathway contextualization and detection of metabolically related compounds.**
A) MIA determines the pairwise similarities of all MIDs. A similarity threshold is applied and compounds with highly similar MIDs are visualized as network. Networks derived from different experimental conditions can be superimposed for more information.
B) Before distance calculation, MIA aligns the MID vectors to account for gains or losses of labeled fragments, which could otherwise conceal the metabolic proximity of these compounds.

The resulting networks of labeled compounds indicate their proximity within the metabolic network. This is a substantial feature of MIA. It can be used to associate unidentified compounds to identified ones, and to map them to specific pathways. This itself is a valuable information and can furthermore be a strong hint for subsequent compound identification. For both identified and unidentified compounds such a nearest neighbor analysis based on MID similarity can reveal new biosynthetic pathways or help to distinguish between different known ones. In section 3.5, I will show how this MID similarity analysis was applied to identify unanticipated compounds in cancer cells.

### 3.2.3 Towards higher sensitivity and specificity in the detection of isotopic enrichment

In the course of developing MIA, I encountered the problem that not all isotopically enriched compounds were detected, or that certain fragments, especially those at high masses, were missing. The high-mass fragments are of major importance, since they usually retain the whole carbon backbone of a metabolite. Those missing high-mass fragments posed a problem especially for the MID similarity analysis (subsection 3.2.2). If MIDs of ions that arise from the loss of labeled fragments are analyzed, the similarity to other compounds often provides less information on the metabolic similarity.

To obtain a sufficient increase in sensitivity for the detection of labeled fragments specificity had to be sacrificed. Here, sensitivity means that labeled compounds or individual fragments thereof are detected as such, and specificity means that individual compounds or fragments thereof are only reported as labeled when they truly are. To obtain more informative and more comprehensive data for further analyses, I thoroughly analyzed the occurrences of false positives or false negatives in NTFD. The most common problems was the incorrect determination of fragments boundaries and the rejection of obviously labeled fragments. I modified the label detection algorithm of NTFD to improve and to provide better control over both sensitivity and specificity in such cases (Figure 18).

For the following discussion of NTFD's label detection algorithm, I refer to the detailed explanation provided in section A.10.

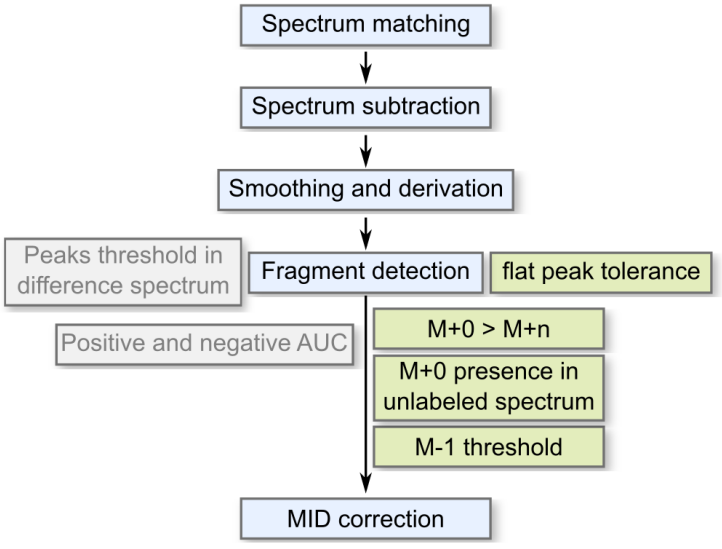



**Figure 18: Schema of the NTFD algorithm.** In order to increase NTFD's sensitivity certain filters were removed (grey boxes). In turn, new filters had to be implemented to recover specificity (green boxes). Details are described in the text.

### Increasing sensitivity

The second major step of the NTFD algorithm (Figure 9, section A.10), after matching mass spectra of the tracer and non-tracer samples, is the determination of individual isotopically enriched fragments. Therefore, the labeled and unlabeled spectra are both normalized to a TIC of one and subtracted. Labeled fragments will produce characteristic peaks in this difference spectrum, whereas unlabeled fragments will not be visible. A labeled fragment leads to a positive peak followed by a negative peak with the same area under the curve (Figure 19). The beginning and end of this peak pattern provides the beginning and the end of the labeled fragment respectively. The peak picking is more easily performed on the first derivative of the difference spectrum where the peak shape changes to two maxima separated by a minimum (Figure 19). I found that, at this stage, many labeled fragment were not detected.

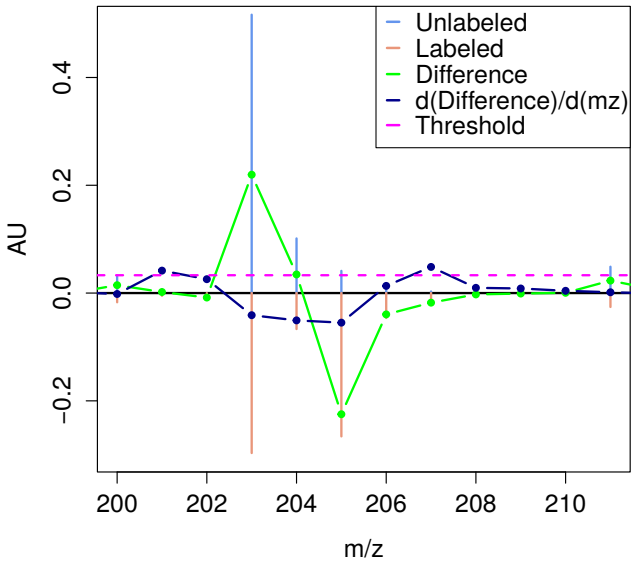



**Figure 19: NTFD detects labeled fragments from characteristic patterns in the difference spectrum.** The negative and positive area under the difference spectrum are identical. The actual peak-picking is performed on the first derivative of the smoothed difference spectrum (d(difference)/d(mz)). There, two maxima mark the beginning and end of the labeled fragment, shifted by two mass units (begin: $m/z$ 201→203, end: $m/z$ 207→209).

**No thresholding** To detect the peaks in the first derivative of the difference spectrum, the initial NTFD implementation used a fixed threshold value: A combination of a maximum above the threshold, minimum below the threshold and again a maximum above the threshold was considered as a labeled fragment (Figure 19, magenta line). This threshold had to be chosen by the user and was then used for all mass spectra throughout the dataset. On the one hand, this threshold had to be set high enough to reduce the number of false-positives, especially in the low-mass region ($m/z < 100$). In that region, there is a lot of overlap of

isotope clusters from different fragments, so that small changes in the relative abundance of certain fragments can cause a lot of noise in the difference spectrum (Figure 20). On the other hand, the threshold had to be set low enough to also include fragments at higher masses that usually have a much lower signal intensity, and thus, a lower peak amplitude in the difference spectrum. These fragments at higher $m/z$ values are most informative, because they usually retain a bigger part of the analyte and provide MIDs for the full backbone.[26] Besides the problem of very high and very low peaks within a difference spectrum, there was another issue caused by varying richness or sparsity of mass spectra. For the calculation of the difference spectrum, the initial spectra are normalized to a certain summed intensity. This normalization leads to different relative signal intensities, depending on the summed intensity of other peaks within the same spectrum. Therefore, in a spectrum with a high number of peaks, the average normalized peak height would be much smaller than in a spectrum with only a few peaks. This reduced peak height would propagate to the difference spectrum and require a lower threshold. The heterogeneity in fragment or peak number per spectrum makes it nearly impossible to find an adequate peak threshold. It soon became apparent that with these issues, sensitivity and specificity could not easily be well balanced with the initial fixed thresholding approach. Therefore, I removed this threshold, *i.e.* set this threshold to zero, the lowest meaningful value, to achieve maximal sensitivity. This change concomitantly lowered specificity and led to many incorrect detections of labeled fragments, especially in more complex regions of the mass spectra (Figure 20).

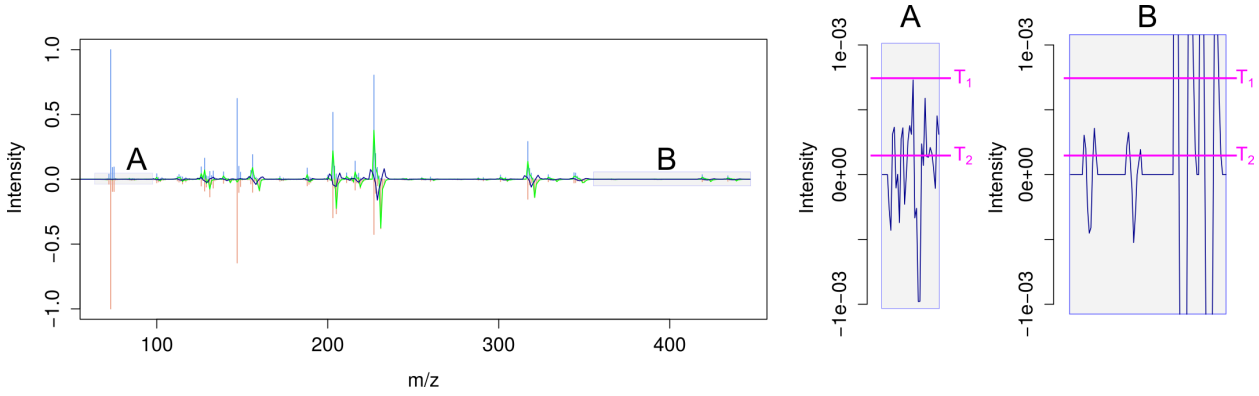



**Figure 20: For the detection of labeled fragments by NTFD, the derivative of the difference spectrum is scanned for characteristic peak patterns above a certain intensity threshold.** Selection of a proper threshold is difficult, because peak heights vary of multiple orders of magnitude. The threshold $T_2$ allows for the detection of the labeled fragments in (B), but will at the same time consider the peaks in the noisy region (A) as labeled fragments. A higher threshold $T_1$ will ignore the noisy region (A), but also miss some labeled fragments in (B).

**Area under the difference spectrum filter** Since the labeled and unlabeled mass spectra are normalized to the same summed intensity, theoretically, the positive and negative area

under the difference spectrum should be equal.[102] Therefore, the ratio of the positive and negative area under the curve was used as a filter criterion. All fragments for which the two areas differed by more than factor two were discarded. However, in practice this filter removed too many valid fragments. The cases where positive and negative areas do not match are probably artifacts from the TIC normalization of the spectra. Slight differences in relative fragment intensities can be caused by spectral skewing.[85] Because of these differences, the sum of the relative intensities of a fragment in the labeled or unlabeled spectrum may differ and would be wrongfully rejected.

**Flat peaks** In some cases, the fragments were correctly detected, but their borders were incorrectly determined. This occurred when the difference in the M+1 peaks was slightly higher than the difference in the M+0 peaks and thus the peaks in the difference spectrum were very flat but peaked at the M+1 position. In these cases, the determined beginning of the labeled fragment was one mass too high. Such misdetected fragments would lead to meaningless MIDs. To avoid these cases, a tolerance for such flat-top peaks was added. The user can then choose how high the M+1 to M+0 ratio in the derivative of the difference spectrum may be to still consider the M+0 as the true start of the fragment. A value of 1.1 turned out to lead to ideal results.

### Increasing specificity by additional filtering

The aforementioned changes in the detection of labeled fragments improved sensitivity a lot, however at the price of specificity. For example, the boundaries of overlapping fragments were detected incorrectly, or noise in the derivative of the difference spectrum was detected as labeled fragments (Figure 21). To improve specificity, I implemented a complementary filtering strategy later on (Figure 18). These filters include checks for maximal isotopic enrichment, isotope cluster integrity, and fragment overlap.

**Maximal isotopic enrichment** A newly implemented filter in the NTFD algorithm makes use of the known relative amount of tracer that was applied. Since the kinetic isotope effect in biochemical reaction is usually rather low, except maybe for deuterium (subsection 1.3.3), all metabolites downstream of the tracer can reach maximally the enrichment of the tracer. This information can be used to further eliminate false positives. For NTFD analysis the stable isotope labeled tracer is often applied in a 1:1 mixture with its unlabeled analogue. Excluding strong isotopologue discrimination, all metabolites downstream of that tracer can exhibit maximally the same 50% of isotopic enrichment. Hence, all fragments with a higher calculated isotopic enrichment must be false positives and can be discarded.

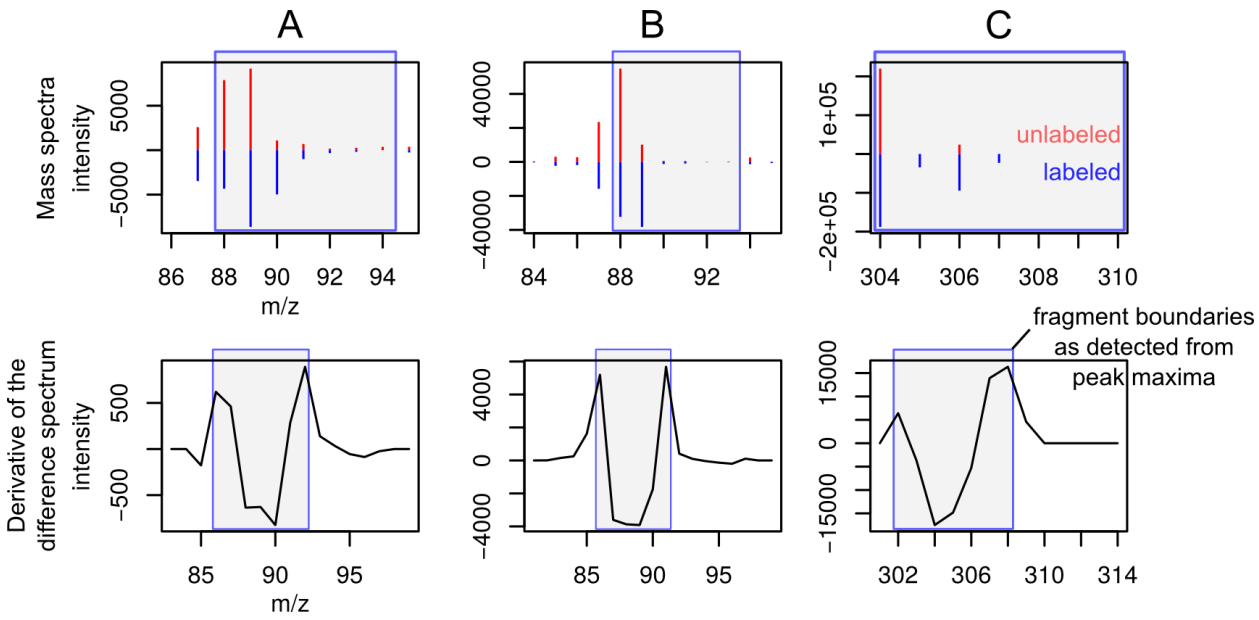



**Figure 21: Filtering of the putative boundaries of labeled fragment is necessary for good specificity.**
A,B) For overlapping fragments, correct MIDs cannot be determined. Indicators for such overlap are M+1 intensity in the putative $m/z$ range that is larger than M+0 (A) or a significant M-1 intensity (B).
C) Putative labeled fragments with no M+1 peak present in the unlabeled spectrum are of very low intensity or incorrect and should be rejected.

**M+0 presence in unlabeled spectrum** Peaks in the difference spectrum may occur at masses that were not detected in the original spectrum. Therefore, the beginning of a putative labeled fragment can be detected at such a position. However, if there is no signal in the unlabeled reference spectrum, this is impossible to be a real fragment. Consequently, every putative M+0 ion is required to be present within the unlabeled spectrum.

**Isotope cluster integrity** In small molecules, the M+0 peak is the highest peak within every isotope cluster, if they do not contain any elements with high natural isotope abundance like bromine or chlorine. Therefore, the M+0 peak is required to have the highest abundance within the isotope cluster in the non-enriched spectrum (Figure 21B). Furthermore, all mass spectrometric fragments should have a detectable M+1 peak (Figure 21C), therefore a minimal value for the M+1 intensity relative M+0 can be specified to filter out incorrect or very low intensity fragments.

**Overlapping fragments** Sometimes isotope clusters of different fragments are overlapping. Although these fragments may be interesting for a qualitative analysis, no meaningful MID can be calculated from them. Therefore, they should be excluded from a quantitative analysis. It is not always easy to detect such fragment overlap. However, when the M-1 peak, *i.e.* the

peak one mass lower than the putative beginning of a fragment is high in comparison to the M+0 peak, then it is likely, that the M+0 peak is "contaminated" by the isotopic peak of M-1 (Figure 21A). Therefore, I implemented a user-definable threshold for the maximal relative M-1 abundance, for which a value of 0.25 provides good results.

## Summary

**Table 5: Improving NTFD's sensitivity and specificity.** Modification of the NTFD's label detection algorithm significantly improved sensitivity while only slightly increasing the false discovery rate as compared to the original implementation. Fragments that were detected as labeled by either NTFD version were assumed to be the sum of TP and FN and were validated manually. TN is not available, because isotopic labeling cannot reliably be completely be excluded.
TP: true positives; FP: false positives; TN: true negatives; FN: false negatives.

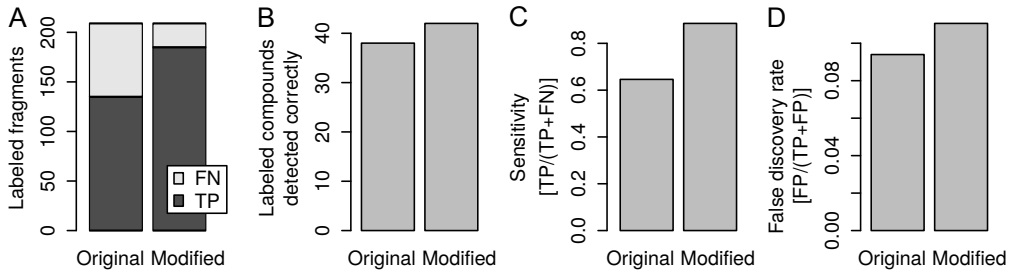
Test set: A549 cells labeled with [U-$^{13}$C]glucose, three replicates were used.
Settings before: *Max.* $1 - M_0$: 0.05; *Minimum number of fragments*: 1; *Maximum fragment deviation*: 0.2; $R^2$: 0.95; *Sensitivity*: 1.
Settings after: as "before", but without *sensitivity* and with: *Min.* $M_0$: 0.45; *Max. M-1*: 0.25; *Flat peak tolerance*: 1.1; *Min.* $M_1$: 0.01; *Ensure* $M_0$ *presence in unlabeled spectrum*: *true*; *Ensure* $M_1 < M_0$: *true*.

**(a)** Original NTFD algorithm

| | | **Manual inspection** | |
|---|---|---|---|
| | | labeled | unlabeled |
| **NTFD** | labeled | 135 (TP) | 14 (FP) |
| | unlabeled | 74 (FN) | *NA* (TN) |

**(b)** Modified NTFD algorithm

| | | **Manual inspection** | |
|---|---|---|---|
| | | labeled | unlabeled |
| **NTFD** | labeled | 185 (TP) | 23 (FP) |
| | unlabeled | 24 (FN) | *NA* (TN) |




**(c)** A) The number of correctly detected fragments increased significantly (+37%). B) The increase in correctly detected labeled compounds increased by 10%. Many of newly detected fragments belong to already detected compounds. C,D) Sensitivity increased much more than the false discovery rate.

In summary, I significantly improved the quality control of NTFD to increase its sensitivity

without major losses in specificity (Figure 18). These modifications allow to detect 37% more labeled fragments in a test data set (human A549 cells incubated with [U-$^{13}$C]glucose) without impairing specificity. For each additional false positive fragment there were 5 additional true positives detected. This significantly improved the data basis for subsequent metabolome-wide MID analyses (section 3.5). To further increase sensitivity in the comparative MID analysis with MIA, all labeled fragments detected in one dataset, were analyzed in all other datasets in an additional step. As a beneficial side-effect, my optimizations led to a significant speed-up of the label detection step (over 10×faster). Computation time was not limiting in NTFD, but became noticeable during the combined analysis of several datasets in MIA.

## 3.3 FFC — fragment formula calculator

For targeted stable isotope labeling analyses, MIDs are usually corrected for natural isotope abundance by comparison of the measured spectrum with the theoretical MID of the average natural isotope abundance (section 1.3.6). Therefore, the elemental composition of the ions of interest needs to be known. For the heaviest fragments in a mass spectrum this is usually straightforward as they occur from typical losses of groups from the derivatization reagents (subsection 1.2.2). However, when larger neutral losses occur, the fragmentation pathways are often harder to determine. Yet, such fragment ions are of interest for stable isotope labeling analyses. The MIDs of those smaller ions which contain only specific fragments of the native metabolite can provide positional information on isotopic enrichment if the retained atoms are known.[97]

To determine elemental composition and potential substructures of mass spectrometric fragment ions, we developed FFC[26] (section A.4). FFC requires the structure of the molecule of interest and uses a graph-based combinatorial approach to determine the simplest potential substructures of a given fragment ion (Figure 22). These substructures are an advantage over the mere determination of the elemental composition, since they provide information on specific atoms retained in a given fragment. In combination with stable isotope labeling, this information can be used to determine positional isotopic enrichment which is otherwise not determinable by MS.[97] For the analysis of isotopic enrichment of certain fragments, it is furthermore important to exclude any fragment overlap, since overlapping fragments would compromise the MID. Overlapping fragments may not be easily detected. However, since FFC uses a combinatorial approach, it will provide any combination of substructures that give rise to the same $m/z$, so that ambiguous fragments can be excluded from further analyses.

### 3.3.1 Complete carbon isotopologue distribution from mass spectrometric fragments

Generally, positional information on isotopic enrichment is not directly available from MS measurements (section 1.3.6). However, once the elemental compositions and the atoms retained in a mass spectrometric fragment ion are determined with FFC, the fragment MIDs can be used to analyze position-specific isotopic enrichment. In the following, I will demonstrate this approach to determine the complete isotopologue distribution from a single EI-MS spectrum of serine 3TMS and glycine 2TMS.

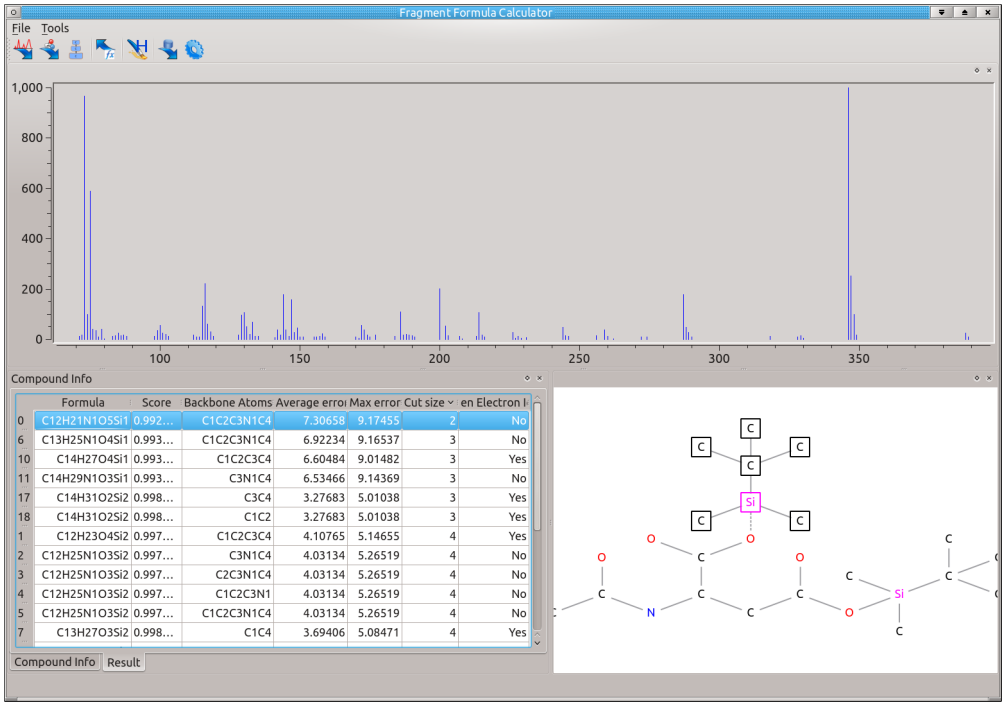



**Figure 22: The FFC GUI.** After providing the mass spectrum and the structure of the complete analyte, FFC determines all possible substructures for a given fragment. It provides several indicators for the likelihood of the individual substructures.[26]

### Serine 3TMS

The amino acid serine is an interesting metabolite since it is the main carbon donor for one-carbon metabolism. C-3 of serine is transferred to THF by serine hydroxymethyl transferase (SHMT) to yield glycine and $N^5$,$N^{10}$-methylene-THF and *vice versa* (Figure 23A).[155] If the enrichment of serine C-3 is known, it can be used as a proxy for the labeling of the methylene-THF pool under steady state conditions (see below).

I will first show how isotopologue abundances of serine determine its MID:

$$\begin{pmatrix} \mathrm{M}_0^{1-2-3} \\ \mathrm{M}_1^{1-2-3} \\ \mathrm{M}_2^{1-2-3} \\ \mathrm{M}_3^{1-2-3} \end{pmatrix} = \begin{pmatrix} G_{000} \\ G_{001} + G_{010} + G_{100} \\ G_{011} + G_{101} + G_{110} \\ G_{111} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} G_{000} \\ G_{001} \\ G_{010} \\ G_{100} \\ G_{011} \\ G_{101} \\ G_{110} \\ G_{111} \end{pmatrix} \tag{3.3}$$

$\mathrm{M}_{1-2-3}$ represents the MID of the molecule containing C-1, C-2 and C-3 (Figure 23B).

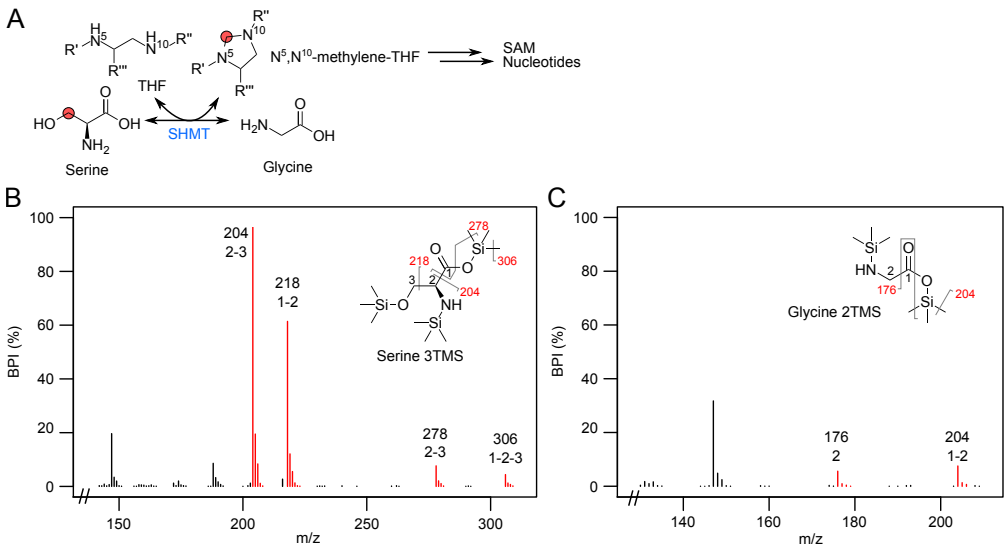



**Figure 23: Serine fuels $C_1$ metabolism.** A) C-3 of serine is transfered to THF which then acts as carbon donor for *S*-adenosyl methionine (SAM) and nucleotide biosynthesis. B,C) Mass spectrometric fragments of serine 3TMS and glycine 2TMS can provide positional information on isotopic enrichment.

$G_{abc}$ is the relative abundance of the isotopologue in which each carbon ($a$: C-1, $b$: C-2, $c$: C-3) is isotopically enriched (1) or not (0). It is obvious that the MID holds less information than the isotopologue distribution and thus cannot be used to determine the isotopologue distribution. However, we can include more information from fragment ions. If the MIDs of the intact serine are known and the MIDs of a fragment containing only C-1 and C-2, the labeling of C-3 can already be deduced. If additionally, the fragment with only C-2 and C-3 is included, the full isotopologue distribution can be resolved. For this purpose, I used FFC to determine the elemental composition and potential substructures of the major fragments of serine 3TMS (Figure 23B). The ion $m/z$ 306 contains the whole serine carbon backbone, $m/z$ 218 contains C-1 and C-2, $m/z$ 204 and $m/z$ 278 both contain C-2 and C-3.

These additional information can be added to the isotopologue mapping matrix $\mathbf{M_{map}}$ and the MID vector in Equation 3.3 to obtain:

$$\underbrace{\begin{pmatrix} \mathrm{M}_0^{1-2-3} \\ \mathrm{M}_1^{1-2-3} \\ \mathrm{M}_2^{1-2-3} \\ \mathrm{M}_3^{1-2-3} \\ \hline \mathrm{M}_0^{1-2} \\ \mathrm{M}_1^{1-2} \\ \mathrm{M}_2^{1-2} \\ \hline \mathrm{M}_0^{2-3} \\ \mathrm{M}_1^{2-3} \\ \mathrm{M}_2^{2-3} \end{pmatrix}}_{\mathbf{MIDs}} = \underbrace{\left(\begin{array}{cccccccc} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ \hline 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ \hline 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 \end{array}\right)}_{\mathbf{M_{map}}} \cdot \underbrace{\begin{pmatrix} G_{000} \\ G_{001} \\ G_{010} \\ G_{100} \\ G_{011} \\ G_{101} \\ G_{110} \\ G_{111} \end{pmatrix}}_{\mathbf{Isotopologues}} \tag{3.4}$$

$$\mathbf{MIDs} = \mathbf{M_{map}} \cdot \mathbf{Isotopologues} \tag{3.5}$$

This system is overdetermined and can be solved for the isotopologues using a least-squares approach:

$$\mathbf{Isotopologues} = \mathbf{M_{map}}^{-1} \cdot \mathbf{MIDs}$$

MIDs for additional fragments (*e.g.* $m/z$ 204 and $m/z$ 278) can be added to make the solution more robust. The overdetermined system furthermore allows for the determination of confidence intervals and a coefficient of determination to estimate the quality of the results.

### Glycine 2TMS

The $C_2$ amino acid glycine is the product of SHMT activity (Figure 23A). The mass spectrometric fragment $m/z$ 176 of glycine 2TMS is formed by the loss of carbon monoxide from the carboxylic acid group and a TMS-methyl group (Figure 23C). Since this fragment only contains C-2 of glycine, the positional enrichment at C-2 can be directly obtained from the MID of fragment $m/z$ 176. Furthermore, the relative abundances of all 4 carbon isotopologues of glycine can be obtained in an analogous manner as shown above for serine:

$$\underbrace{\begin{pmatrix} \mathrm{M}_0^{1-2} \\ \mathrm{M}_1^{1-2} \\ \mathrm{M}_2^{1-2} \\ \hline \mathrm{M}_0^{2} \\ \mathrm{M}_1^{2} \end{pmatrix}}_{} = \underbrace{\left(\begin{array}{cccc} 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ \hline 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{array}\right)}_{} \cdot \underbrace{\begin{pmatrix} G_{00} \\ G_{01} \\ G_{10} \\ G_{11} \end{pmatrix}}_{} \tag{3.6}$$

$$\mathbf{MIDs} = \mathbf{M_{map}} \cdot \mathbf{Isotopologues} \tag{3.7}$$

Again, this system is overdetermined and can be solved by least-squares regression.

### Biological interpretation

I used the aforementioned formulas to determine the complete carbon-isotopologue distributions of serine and glycine in A549 cells after [U-$^{13}$C]glucose labeling (Figure 24A). The unlabeled isotopologues are the most abundant and represent the fraction of serine and glycine that is taken up from the medium; the fully labeled isotopologues are synthesized *de novo*. Only [3-$^{13}$C]serine and [1,2-$^{13}C_2$]serine are produced in significant amounts out of the six possible [$^{13}C_1$]serine and [$^{13}C_2$]serine isotopologues. These isotopologues are formed by the serine-glycine-interconversion through SHMT and clearly show that C-3 of serine is transferred to or derived from the one-carbon pool (Figure 24B).

Glycine is either fully labeled or unlabeled, indicating that there is no significant glycine biosynthesis via the glycine cleavage system (Gly + THF + $NAD^+ \leftrightharpoons$ $N^5$,$N^{10}$-methylene-THF + $CO_2$ + $NH_3$ + NADH + $H^+$ ).[155] This reaction would otherwise produce [$^{13}C_1$]glycine, from $N^5$,$N^{10}$-[$^{13}$C]methylene-THF.

Under hypoxia, there was a statistically significant increase in the relative abundance of [U-$^{13}$C]serine and a concomitant decrease in relative [3-$^{13}$C]serine and [1,2-$^{13}C_2$]serine. This change indicates a lower formation of serine via SHMT (→[3-$^{13}$C]serine and [1,2-$^{13}C_2$]serine) than via *de novo* biosynthesis from 3-phosphoglycerate (→[U-$^{13}$C]serine). It was shown recently that hypoxic cancer cells generate NADPH from serine catabolism via SHMT and the glycine cleavage system.[194] Such an increased net flux from serine to glycine, is in line with the observed labeling pattern.

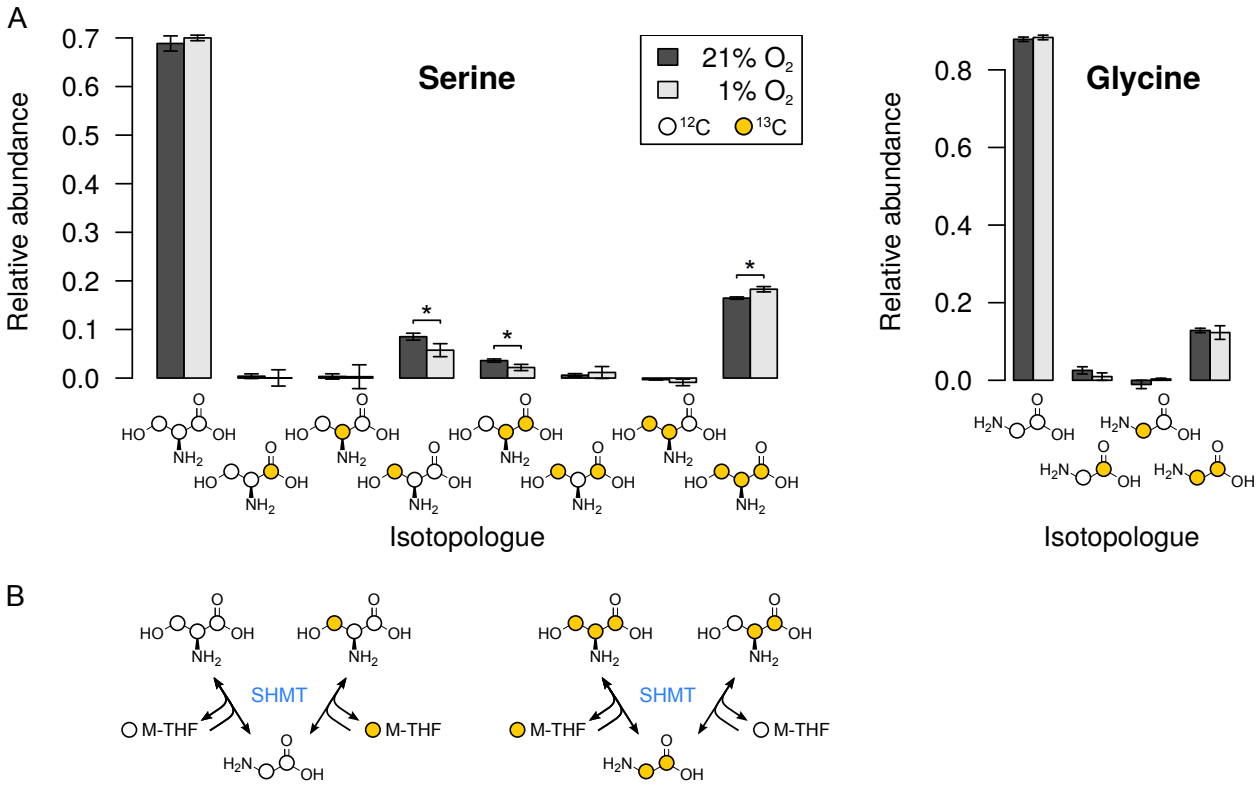



**Figure 24: Isotopologues of serine and glycine after [U-$^{13}$C]glucose labeling in A549 cells at different oxygen levels.**
A) The full isotopologue distributions of serine and glycine can be determined from MIDs of mass spectrometric fragments. Some isotopologues are not formed, or are very low abundant. The unlabeled fraction is taken up from the growth medium, the labeled fraction is synthesized by the cells. *: Welch's $t$-test $p < 0.05$, $n = 3$.
B) The different isotopologues of glycine and serine are interconverted by SHMT. Only C-3 of serine is transfered, the C-1—C-2 bond is not cleaved, therefore these atoms are either both labeled or both unlabeled.

## 3.4 Automatied metabolome-wide isotope dilution normalization for non-targeted GC-EI-MS metabolomics

A common problem in non-targeted metabolomics is the comparability of results across different measurements. Due to analytical variation and instrument drift, the comparability is limited. The problem is even bigger for large studies involving the measurement of a high number of samples in different batches across different instruments. Several more or less complex normalization techniques have been developed,[43,195] each with its specific strengths and weaknesses, but no single method is really suitable for all applications.

For non-targeted metabolomics it is infeasible to use a defined mixture of internal standards, since it is far too expensive and most analytes remain unidentified and thus cannot be purchased. To overcome this limitation, metabolite extracts of fully stable isotope labeled organisms have been used as complex internal standard mixtures.[44,45] These extracts, ideally derived from the organism of interest, contain a large number of metabolites. Although there is a larger number of internal standards in such mixes, data analysis in previous studies has mostly been rather targeted and only known compounds have been analyzed.[44,45] For a non-targeted analysis all isotopically enriched compounds need to be detected and appropriate quantification ions have to be chosen. Such a manual comparison of all mass spectra from the labeled and unlabeled metabolite extract is very tedious. For liquid chromatography high resolution MS (LC-HR-MS) there are tools available for the automated detection of stable isotope labeled compounds and the calculation of isotopologue ratios,[46,47] but not for GC-EI-MS. The hard ionization and the usually low-resolution analyzers in EI-MS make it more challenging to detect isotopic labeling and adequate quantification ions. Therefore, I developed an algorithm to automate the detection of potential internal standards and the selection of quantification ions in GC-EI-MS data (Figure 25).

I used NTFD for the detection of all isotopically enriched compounds in the reference mixture. These compounds can possibly be used as internal standards for quantification or normalization. For each compound, the main peaks of the labeled and unlabeled forms are detected and checked if they suffice certain quality requirements. All suitable compounds along with all valid pairs of quantification ions are added to a mass spectrum reference library which can be used with MetaboliteDetector to obtain the respective signals in the samples of interest. I implemented the algorithm as a new feature in the NTFD software.

This non-targeted IDMS-based normalization approach outperformed the commonly used TIC-normalization and the normalization to a single internal standard. The full methodology is described in Weindl *et al.*[191] (section A.7). There, I also showed that the application of a yeast-derived stable isotope labeled reference mixture is not limited to yeast samples, but can as well be applied to mammalian cells, for which complete labeling is difficult to achieve

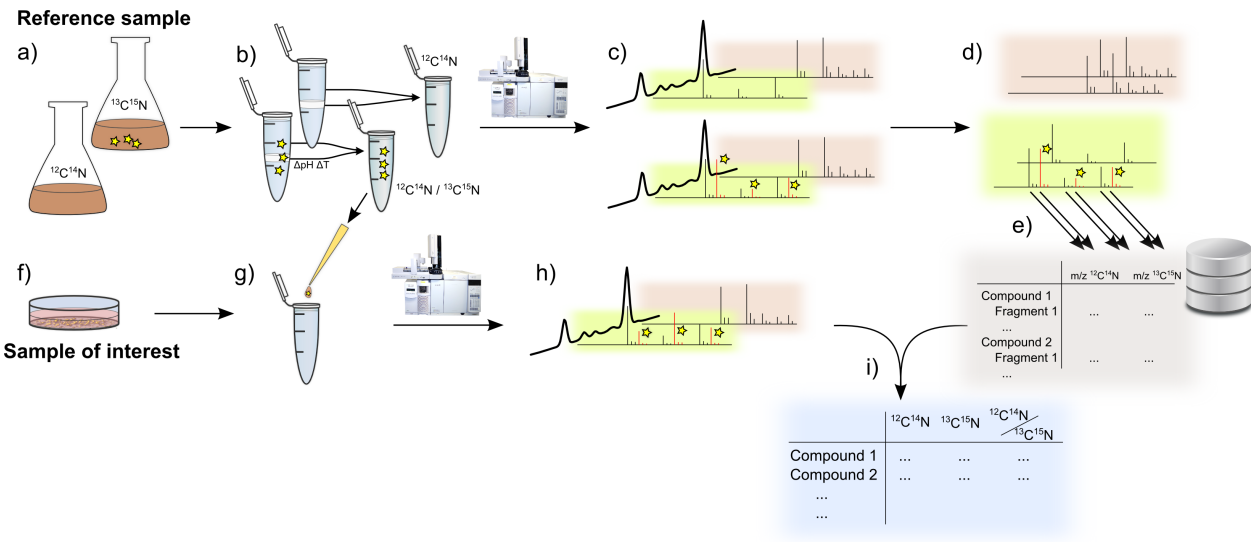



**Figure 25: Experimental setup.** a) A reference organism is grown simultaneously in defined medium and in a medium where all carbon and nitrogen sources are substituted by their fully stable isotope labeled analogues. b) Metabolites are extracted using water:methanol:chloroform. The protein- and nucleic acid-containing interphase is hydrolyzed, pooled with the polar metabolites and used as reference extract. c–d) NTFD is used to detect all stable isotope labeled compounds and fragments, as well as the $m/z$ ratios of their light and heavy isotopologues. Therefore, the unlabeled extract and a mixture of labeled and unlabeled extracts are measured with GC-EI-MS. e) The spectra of all these labeled compounds and selected quantification ions are collected in a reference library to be used to identify and quantify compounds within a sample of interest. f–g) An aliquot of the labeled reference mixture is added to a sample of interest. h–i) The previously determined ions are used for quantification of the detected compounds. The ratios of the intensities of light and heavy form for each metabolite provide a robust measure for the comparison of metabolite amounts across experiments. Reproduced from Weindl *et al.*[191].

and very expensive.

## 3.5 Non-targeted analysis of stable isotope labeling in hypoxic cancer cells

Although most tumors form a hypoxic microenvironment[135] and many tumor cells are exposed to oxygen levels of below 2%,[136] pronounced tumor hypoxia is considered a bad prognostic sign for cancer patients. Hypoxia was shown to increase the motility of lung cancer cells[184] through the induction of growth factors and possibly through induction of metalloproteinases which contribute to tumor metastasis.[185] Moreover, many cancer types show a pseudo-hypoxic phenotype, *i.e.* they show hypoxic metabolic features also under normoxic conditions.[18] Therefore, the in-depth analysis of hypoxia-induced metabolic changes is highly relevant for the understanding of cancer cell metabolism and to advance cancer therapy, because despite major progresses in its understanding, cancer still remains a wide-spread disease with an increasing number of cases every year.[128]

Metabolism of cancer cells is different from metabolism of non-transformed post-mitotic cells (section 1.4). For example, cancer cells show a much higher glutamine catabolism and changes in flux directionality of certain reactions.[8,196] Such changes in metabolic fluxes cannot be detected by only analyzing intracellular steady-state metabolite pool sizes. However, as described in subsection 1.3.8, stable isotope labeling experiments can provide information on changing metabolic fluxes.

Cancer cells do not only exhibit changed fluxes, but can also produce specific "oncometabolites". Cancer-associated mutations in IDH were shown to give rise to 2-hydroxyglutaric acid, a metabolite that is not produced under normal conditions.[197,198] To account for such unexpected metabolites, non-targeted approaches are required. However, while non-targeted analysis of metabolite levels is likely to miss metabolic flux changes in redundant pathways, targeted MFA is likely to miss unanticipated metabolic reactions, and unfortunately there are no non-targeted MFA techniques available. Therefore, we intended to develop a novel approach to obtain metabolic flux information in a non-targeted manner. Such flux information can be obtained from MIDs (section 1.3.8), and MIDs can be determined in a non-targeted manner (subsection 1.3.7). So far, no truly non-targeted mass isotopolome analyses have been performed. This can be mostly attributed to the lack of adequate data analysis tools.[109] To this end, I developed MIA, which allows for efficient non-targeted data analysis of complex stable isotope labeling datasets (section 3.2, section A.8).

To further the understanding of hypoxia-induced metabolic changes in cancer cells, my colleague, Dr. Thekla Cordes, cultivated human lung adenocarcinoma cells (A549)[199] at different $O_2$ concentrations ranging from atmospheric levels (21% $O_2$) down to severe hypoxia (1% $O_2$) to analyze their metabolic response.[186] She cultivated the cells in the presence of the stable isotope labeled tracers [1,2-$^{13}$C]glucose and [U-$^{13}$C]glutamine. Analysis of MIDs of

TCA cycle intermediates revealed a shift in fractional carbon contribution from glucose at high oxygen levels to glutamine at low oxygen levels.[186] Furthermore, reductive glutamine metabolism was found to be highly induced by hypoxia, which was in agreement with other studies.[8] These results were obtained from targeted analyses of MIDs of a predefined set of metabolites.

To extend this study, I applied MIA on that dataset for a non-targeted stable isotope labeling analysis. Although operating in non-targeted mode, I was able to confirm already known hypoxia-induced metabolic changes like reduced PDH flux, increased glutaminolysis and glutamine-fueled fatty acid biosynthesis (section A.9). In addition, I found unexpected metabolites like the neuronal metabolite *N*-acetylaspartic acid and *N*-carboxyamino acids (section 3.5.3).

Following these results, in subsequent experiments I was able to show that hypoxic A549 cells rely on increased BCAA catabolism, probably mediated by BCKDHA, possibly to provide mitochondrial acetyl-CoA and that BCAA-derived propionyl-CoA gives rise to odd-chain fatty acids, which were to my knowledge not yet reported to be produced by human cells under quasi-physiological conditions. The observed increase in BCAA catabolism in hypoxic A549 cells may serve as a drug target to treat lung cancer. It has been shown in glioma cell lines that suppression of BCAT1 reduces proliferation and invasiveness.[174]

I furthermore followed up on the surprising occurrence of *N*-acetylaspartic acid in A549 cells, thereby finding evidence for an alternative and yet unknown acetyl-shuttling mechanism across the mitochondrial membrane based on *N*-acetylaspartic acid, circumventing ACLY. I found *N*-acetylaspartic acid to be present also in other cancer lines and in primary tumor tissue (section 3.5.4). Using siRNA-mediated gene silencing, I found evidence that A549 cells produce *N*-acetylaspartic acid via NAT8L (section 3.5.4), the same enzyme that is responsible for neuronal *N*-acetylaspartic acid formation.

### 3.5.1 Isotopic steady state

The assumption underlying the mass isotopomer abundance variation analysis (subsection 3.2.1), that changes in MIDs must the caused by changes in metabolic fluxes, is only valid for a metabolic system in metabolic and isotopic steady-state. Until the system reaches isotopic steady state, MIDs will change over time even without any changes in metabolic fluxes.[118] Therefore, it is important to determine the time window of isotopic pseudo-steady-state between the initial isotopic enrichment phase and the metabolic non-stationarity caused by substrate limitation or accumulating excretion products.

To this end, A549 cells were grown for different periods in medium with either [U-$^{13}$C]glucose or [U-$^{13}$C]glutamine as tracer (Figure 26, experiment performed by Jenny Ghelfi). MIDs in intermediary metabolites turned out to be relatively stable from 18 hours until 72 hours after

application of either tracer. For subsequent stable isotope labeling experiments, cells were therefore incubated with the tracer for 24 hours or 48 hours.

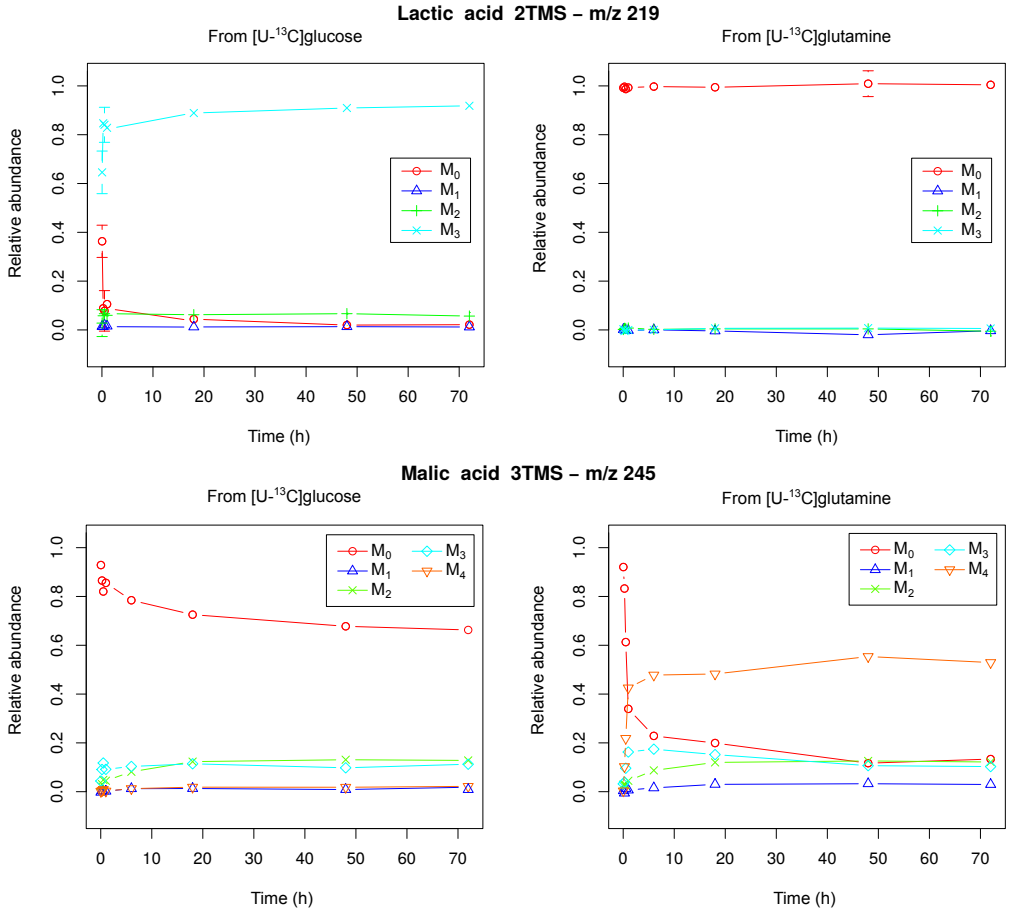



**Figure 26: Isotopic steady state in central carbon metabolism of A549 cells.** For glycolysis, steady state is reached after about one hour. After about 18 hours, MIDs of TCA cycle metabolites stabilize, with [U-$^{13}$C]glutamine slightly faster than with [U-$^{13}$C]glucose. Lactic acid is not enriched from [U-$^{13}$C]glutamine. Points are means of three replicates.

## 3.5.2 Fatty acid metabolism

The most significant flux changes observed in hypoxic A549 cells were associated with reductive glutamine metabolism and the transition from glucose to glutamine as the main carbon contributor for TCA cycle metabolites (section A.9). It has been reported before, that this reductive glutamine metabolism supplies acetyl-CoA for lipid biosynthesis.[8] Because lipid biosynthesis is an important metabolic feature of cancer cells,[137,166,167] I was interested in the carbon contribution of different substrates. Therefore, I analyzed lipid fatty acids as FAMEs on GC-MS after stable isotope labeling (Figure 27). The major carbon sources of palmitate (C16:0), were lipids from the growth medium, glucose and glutamine, providing about 90% of the lipid carbon in total. About 10% of the carbon, however, were of unknown

origin. The fractions changed significantly under hypoxia as previously reported:[8] The glucose-derived fraction was reduced by about 50% and glutamine contribution doubled. There was also a slight in increase in growth medium-derived fatty acids and in carbon of other origin.

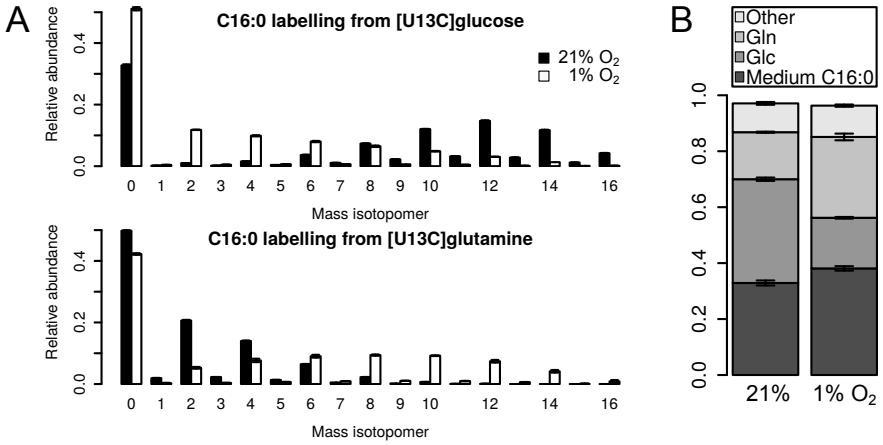



**Figure 27: Carbon origin of fatty acids under normoxia and hypoxia.**
A) Isotopic enrichment in C16:0 after [U-$^{13}$C]glucose or [U-$^{13}$C]glutamine labeling showing a strong shift from glucose- to glutamine-derived carbon.
B) Fractional fatty acid carbon origin. "Glc" and "Gln" represent fractional enrichment of C16:0 from [U-$^{13}$C]glucose or [U-$^{13}$C]glutamine, respectively. Fully unlabeled ($M_0$) after combined [U-$^{13}$C]glucose and [U-$^{13}$C]glutamine labeling was assumed to be fully derived from the growth medium; the remaining $^{12}$C-fraction not in $M_0$ is classified as "other". All fractional contributions are significantly changed with hypoxia (Welch's $t$-test, $p < 0.05$, $n = 3$).

### A549 cells produce odd chain fatty acids

Non-targeted stable isotope labeling analysis of FAMEs after [U-$^{13}$C]glucose and [U-$^{13}$C]glutamine labeling revealed isotopic enrichment in the odd-chain fatty acids pentadecanoic acid (C15:0) and heptadecanoic acid (C17:0). Furthermore, their levels were increased under hypoxia. Although rats were shown to produce these compounds upon injection of propionic acid[200], their *de novo* production in mammals was surprising because it has so far not been shown to occur in these organisms under physiological conditions.

A large fraction of these odd-chain fatty acids was unlabeled (about 85% $M_0$) and most probably derived from the growth medium, but stable isotope labeling confirmed the endogenous origin of a fraction of about 15% (Figure 28B,D). If all of the fatty acids were synthesized *de novo* from acetyl-CoA, the MID would be expected to be unimodal. However, even without considering $M_0$, the MID of C15:0 and C17:0 are still bimodal, indicating different biochemical origins of the fatty acids. For even-chain fatty acids, it is known that *de novo* biosynthesis by the fatty acid synthase complex stops at palmitic acid[168] and that longer chain fatty acids are synthesized by elongases, separate enzymes which add more

acetyl units.[169] For odd-chain fatty acids, there seems to be a similar mechanism, indicated by the relatively high isolated $M_2$ abundance, which is most probably due to elongation of shorter exogenous fatty acids by one acetyl unit. Over 10% of the total C17:0 is derived from exogenous C15:0 by such an elongation ($M_2$ in Figure 28B,D), about 5% C17:0 are synthesized *de novo*. Furthermore, mass isotopomers $>M_{12}$ ($>M_{14}$) of C15:0 (C17:0) were essentially absent, indicating that a C3 moiety is derived from a different source than acetyl-CoA.

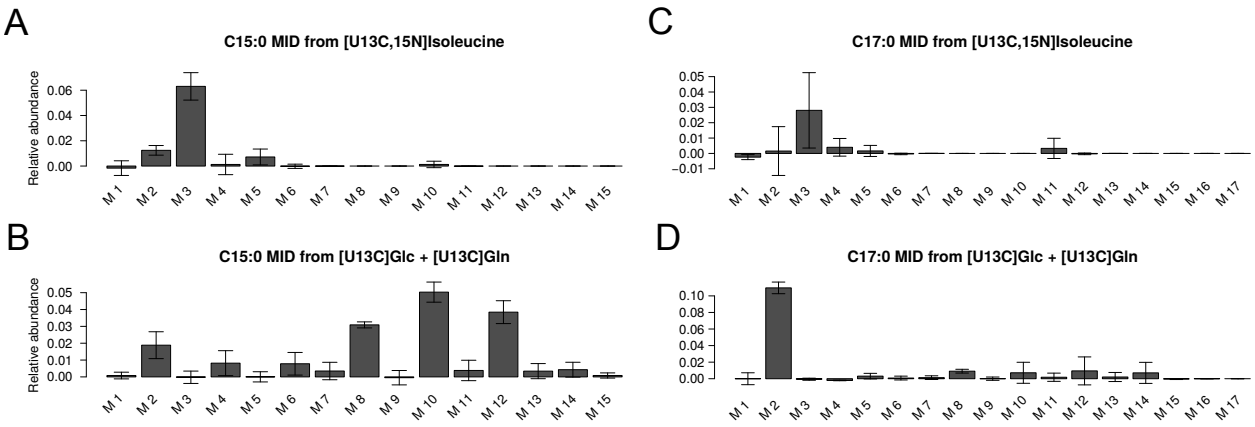



**Figure 28: A549 cells produce odd-chain fatty acids.** Isotopic enrichment of C15:0 (A,B) and C17:0 (C,D) after [U-$^{13}$C,$^{15}$N]isoleucine (A,C) and simultaneous [U-$^{13}$C]glucose/[U-$^{13}$C]glutamine labeling (B,D). $M_0$ fraction omitted.
A,C) Propionyl-CoA for odd-chain fatty acid biosynthesis is derived from isoleucine catabolism ($M_3$ from [U-$^{13}$C,$^{15}$N]isoleucine).
B,D) The main fraction of C15:0 and C17:0 is of exogenous origin (about 85% $M_0$).
D) Over 10% of the total C17:0 is elongated exogenous C15:0 ($M_2$ from glucose and glutamine labeling), about 5% are synthesized *de novo*.
Bars show mean±SD.

### Odd-chain fatty acids are derived from BCAA catabolism

Others have shown that odd-chain fatty acids can be produced in cell lysates upon supplementation with exogenous propionyl-CoA.[201] BCAA catabolism is known to yield propionyl-CoA (subsection 1.4.7). Other potential sources include methionine or threonine via 2-ketobutyric acid. To assess the contribution of BCAA carbon to odd-chain fatty acid and fatty acids in general, I applied [U-$^{13}$C]valine, [U-$^{13}$C]leucine, and [U-$^{13}$C,$^{15}$N]isoleucine to A549 cells and analyzed the isotopic enrichment in polar metabolites and lipids.

Stable isotope labeling with [U-$^{13}$C,$^{15}$N]isoleucine under hypoxia resulted in mostly $M_3$ enrichment of odd-chain fatty acids, confirming their biosynthesis starting from isoleucine-derived propionyl-CoA (Figure 28A,B). Complementarily, glucose and glutamine labeling resulted in maximum $M_{12}$ and $M_{14}$ labeling of C15:0 and C17:0, respectively.

From valine, there was no labeling detectable, suggesting that valine-derived methylmalonic acid is mostly directly converted to methylmalonyl-CoA without intermediary propionyl-CoA

(subsection 1.4.7).

Odd-chain fatty acids are known to be produced by bacteria in the rumen of ruminants and are used in humans as biomarkers of dairy fat intake.[202] However, their biological role is still unclear. Odd chain fatty acids may be produced as an additional measure to detoxify propionic acid besides the methylmalonyl pathway, or their production might only occur as by-product due to unspecific usage of propionyl-CoA instead of acetyl-CoA during the initiation of fatty acid biosynthesis. Besides any potential biological function, these endogenous odd-chain fatty acids need to be accounted for when they are used as nutritional biomarker or as internal standards for quantification.

### Hypoxia induces BCAA catabolism

To assess the carbon contribution from BCAA catabolism to the TCA cycle and lipid biosynthesis, I performed a BCAA labeling experiment at normoxia and hypoxia. Because there was no BCAA-free growth medium available, the tracers were added to the growth medium in addition to the unlabeled BCAAs in the same concentrations. The isotopic enrichment from BCAA tracers in polar metabolites was rather low, but significantly increased under hypoxia (Figure 29A). BCAA-carbon enters the TCA cycle as acetyl-CoA and succinyl-CoA as confirmed by the labeling of citrate, succinate, and palmitate. Although the relative carbon contribution to the these metabolites was relatively low, it doubled under hypoxia.

Transamination of BCAAs occurs to a high extent already under hypoxia: The $^{15}N$ enrichment in *e.g.* alanine was around 6% from [U-$^{13}C$,$^{15}N$]isoleucine alone (Figure 29A). Assuming valine, leucine, and isoleucine are degraded in similar amounts this would lead to $3 \cdot 6\% = 18\%$ enrichment, which is 36% when considering the tracer ratio. Therefore, about one third of the nitrogen in the "transamination pool", which is reversibly transferred by aminotransferases between amino acids and ketoacids, must be derived from BCAAs.

The relative hypoxia-induced increase in alanine labeling from [U-$^{13}C$,$^{15}N$]isoleucine nitrogen was rather low and, therefore, the reversible BCAT reactions seems not to be the rate-limiting step of BCAA breakdown. Gene expression analysis of BCAA catabolic genes showed a significant induction in BCKDHA expression under hypoxia, the protein of which catalyzes the irreversible rate-limiting oxidative decarboxylation in BCAA breakdown (Figure 29B). In line with the labeling data, BCAT expression did not change significantly, neither of the cytosolic, nor of the mitochondrial form, which both seem to be expressed in a similar range.

Propionyl-CoA carboxylase (PCC) is required for the carboxylation of propionyl-CoA. Its gene expression is not highly affected by hypoxia; while the $\alpha$-subunit is slightly up-regulated, the $\beta$-subunit is slightly down-regulated (Figure 29). Assuming substrate saturation of PCC, the increased propionyl-CoA production from degradation of isoleucine and valine, together with the non-induction of PCC, may be a reason for the increased odd-chain fatty acid

production under hypoxia.

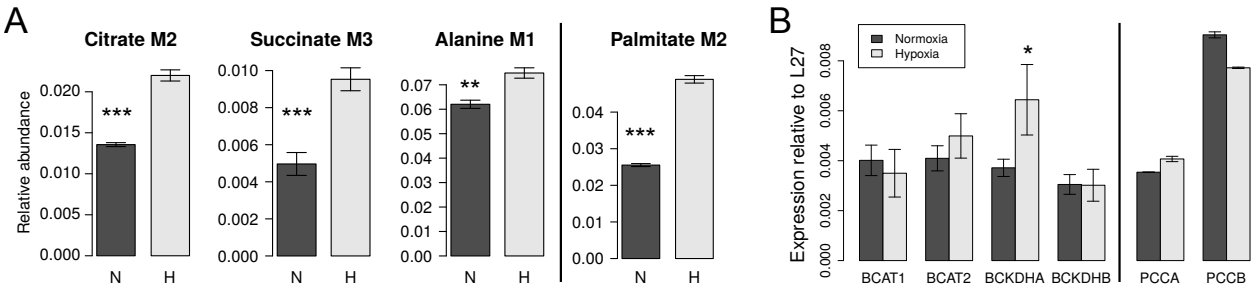



**Figure 29: Hypoxia induced BCAA catabolism can be seen in isotopic labeling and gene expression levels.**
A) Relative mass isotopomer abundances after [U-$^{13}$C]valine, [U-$^{13}$C]leucine, and [U-$^{13}$C,$^{15}$N]isoleucine labeling in A549 cells under normoxia ("N") and hypoxia ("H"). $M_2$ citrate is synthesized from labeled acetyl-CoA derived from mitochondrial leucine and isoleucine catabolism (Figure 12) which is further used for *de novo* fatty acid synthesis as visible in $M_2$ palmitate (palmitate data was obtained from a separate experiment with only [U-$^{13}$C,$^{15}$N]isoleucine labeling). $M_3$ succinate is derived from breakdown of valine and isoleucine via methylmalonyl-CoA. $M_1$ alanine represents the [$^{15}$N]alanine indirectly derived from deamination of [U-$^{13}$C,$^{15}$N]isoleucine. *: Welch's $t$-test $p < 0.05$ ($n = 3$), **: $p < 0.01$, ***: $p < 0.001$.
B) Effect of hypoxia on expression levels of BCAA catabolic genes. Neither cytosolic (BCAT1) nor mitochondrial (BCAT2) BCAT is significantly affected. Expression of the $\alpha$-subunit (BCKDHA) of BCKDH is significantly induced, but not its $\beta$-subunit (BCKDHB). Data are means ± SD of two independent experiments. *: Welch's $t$-test $p < 0.05$ ($n = 6$). Expression of $\alpha$- (PCCA) and $\beta$-subunits (PCCB) of PCC expression is slightly affected by hypoxia (separate experiment, $n = 2$).

## BCAA catabolism supports proliferation

The degradation of BCAA in proliferating cells seems counterintuitive, as high leucine levels are required to induce mTORC1 signaling promoting cell proliferation.[175] To assess the biological significance of BCAA catabolism, cell growth was observed in the present of an activator or inhibitor of BCAA catabolism (Experiment performed by Xiangyi Dong). Gabapentin inhibits BCAA degradation by competitive inhibition of BCAT1 which catalyzes the first step of their degradation, the transamination. Clofibric acid on the other hand promotes BCAA catabolism by activation of BCKDH[203] which catalyzes the second step of BCAA catabolism (Figure 12).

Gabapentin is showing a dose-dependent inhibitory effect on cell proliferation (Figure 30). After 72 h in the presence of 100 µM and 1 mM gabapentin, the number of cells was significantly decreased. Similar effects have been reported for glioblastoma cells.[174] Activation of BCAA catabolism by clofibric acid may convey a slight growth advantage, but differences in cell numbers were not statistically significant at concentrations of 100 µM and 1 mM clofibric acid. Higher concentrations seemed to be cytotoxic (Figure 30).

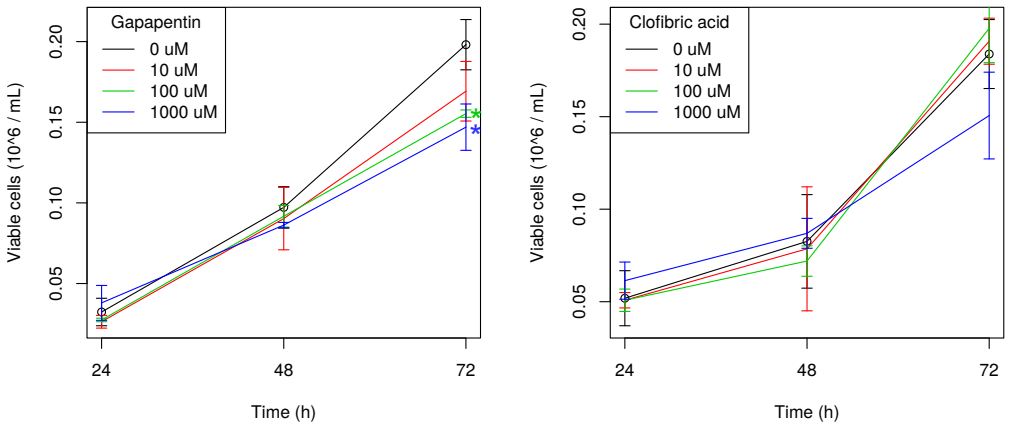



**Figure 30: Effect of an activator (clofibric acid) and an inhibitor (gabapentin) of BCAA catabolism on cell proliferation.**

In summary, stable isotope labeling with BCAAs revealed an increase in BCAA contribution to even-chain fatty acids under hypoxia. The relative increase is statistically significant, but overall contribution is still low. BCAA catabolism provides mitochondrial acetyl-CoA and NADH, but the biological significance is unclear. The observed decrease in cell proliferation due to BCAT inhibition, together with studies by others,[174] suggest an important role of BCAAs metabolism. Further studies are needed to assess the biological function of the increased BCAA uptake and catabolism under hypoxia and of cancer cells in general. BCAA catabolism may provide a therapeutic targeted for cancer treatment as previously suggested.[174]

### 3.5.3 Unidentified isotopically enriched compounds

Non-targeted stable isotope labeling analysis revealed a number of compounds which were not present in our reference library, but isotopically enriched and therefore likely to be derivatization products of cellular metabolites. Furthermore, several of these compounds showed changes in their MIDs in response to hypoxia, indicating that some flux in their biosynthesis was changing. However, without identification of the respective compounds, they only provide limited biological insights. To aid compound identification, I applied a strategy combining stable isotope labeling in cell culture, stable isotope labeled derivatization agents, GC-EI-MS and GC-APCI-TOF-MS techniques (Figure 31, subsection 1.2.3).

#### *N*-Carboxy amino acids

**RI 1920 — *N*-carboxyglutamic acid** One compound showing a strong hypoxia response in labeling was found to have a very high MID similarity to glutamic acid after either glucose or glutamine labeling (Figure 32A). Subsequent analysis of the mass spectrum revealed neutral losses very similar to glutamic acid (Figure 32B). I determined the sum formula of

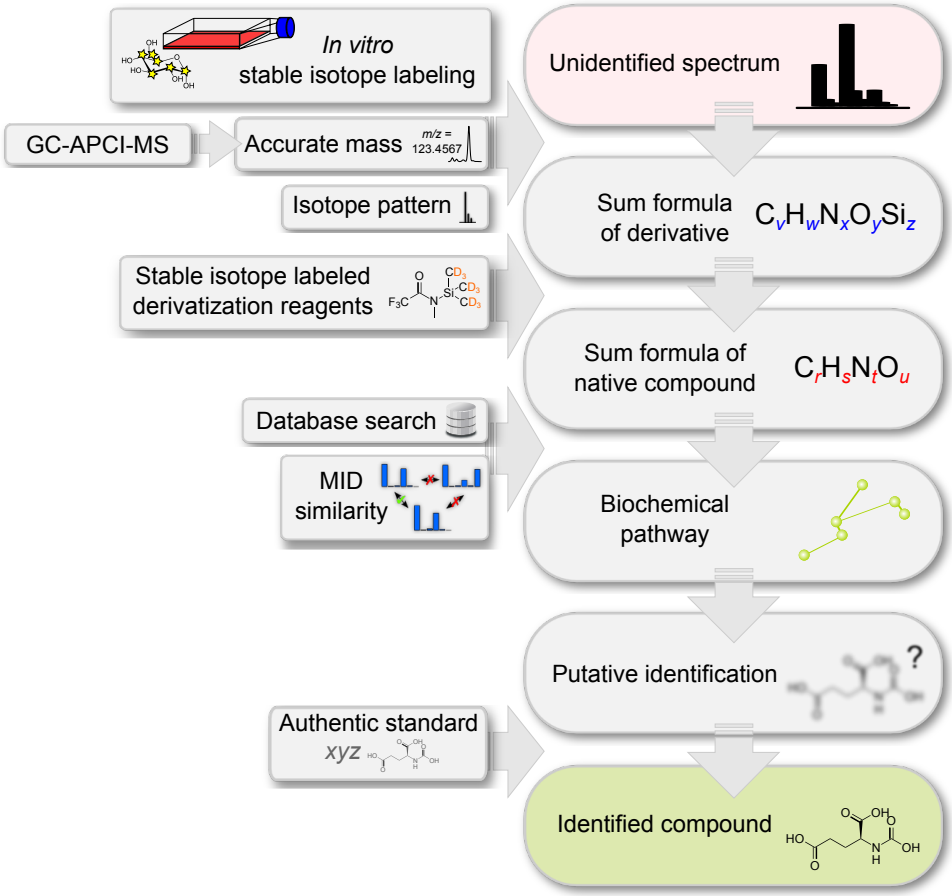



**Figure 31: Strategy for compound identification.** Stable isotope labeling and accurate mass measurements are used to narrow down potential sum formulas of the measured compound. Stable isotope labeled derivatization agents help to to determine the sum formula of the underivatized compound. Once this is available database searches may provide candidate structures which have to be confirmed with measurements of authentic standards.

the unidentified compound based on the minimum number of carbon atoms derived from isotope labeling, the odd-numbered mass of the molecular ion ($m/z$ 479), and number of TMS groups as derived from the natural mass isotopomer distribution. This sum formula, $C_{18}H_{41}NO_6Si_4$, was later confirmed by accurate mass measurement on a GC-APCI-TOF-MS instrument. It differed from the 3TMS derivative of glutamic acid ($C_{14}H_{33}NO_4Si_3$) by $CO_2$ and an additional TMS group. The high MID similarity to glutamic acid made a modification of the carbon backbone unlikely, so that it had to be a carboxy-derivative of glutamic acid. $\gamma$-carboxyglutamic acid was excluded since it is too unstable for GC analysis. The compound was later identified as *N*-carboxyglutamic acid. It is not clear whether this compound is a native metabolite or a derivatization artifact. *N*-carboxy derivatives of several primary and secondary amines are known to be formed during trimethylsilylation of amines in the presence of carbon dioxide,[204] but *N*-carboxyglutamic acid was not reported yet. On the other hand, *N*-carboxylation of amines also occurs physiologically,[205] is associated with

neurotoxicity of amino acids[206,207] and seems to play a significant role as activating step in the elimination of many xenobiotics.[208]

**Other *N*-carboxy amines** After having discovered *N*-carboxyglutamic acid, I was curious whether there were also other *N*-carboxyamino acids present in our samples. After the preparation of reference compounds from the incubation of several amino acids with sodium bicarbonate, I was able to identify more *N*-carboxy amino acids. In total, I found 27 *N*-carboxy derivatives of 13 amines (Figure 33 & Table 7). Some of them have been observed previously,[204] some have not been reported so far. Some of these compounds like *N*-carboxyglycine, *N*-carboxyserine and *N*-carboxythreonine were also present in A549 extracts (Table 7).

**Table 7: *N*-carboxyamines (Figure 33) and TMS-derivatives obtained from incubation of the amines and amino acids with $NaHCO_3$, derivatization with MSTFA and GC-MS measurement.**

| Native amino acid | *N*-Carboxyderivative | TMS derivatives |
|---|---|---|
| Alanine | *N*-Carboxyalanine | 2TMS; 3TMS[b] |
| Asparagine | *N*-Carboxyasparagine | 4TMS; 5TMS |
| Aspartic acid | *N*-Carboxyaspartic acid | 4TMS |
| Glutamic acid | *N*-Carboxyglutamic acid | 4TMS[b] |
| Glycine | *N*-Carboxyglycine | 3TMS |
| Lysine | *N*-Carboxylysine | 3TMS; 4TMS[b]; 5TMS |
| Methionine | *N*-Carboxymethionine | 3TMS |
| Ornithine | *N*-Carboxyornithine | 3TMS; 3TMS [-H2O]; 4TMS; 5TMS |
| | $N_1$,$N_2$-Biscarboxyornithine | 5TMS |
| Proline | *N*-Carboxyproline | 2TMS[b] |
| Serine | *N*-Carboxyserine | 4TMS[b]; 4TMS [-H2O] |
| Spermidine | *N*-Carboxyspermidine | 4TMS[a]; 5TMS[a] |
| Threonine | *N*-Carboxythreonine | 3TMS[b]; 4TMS[b] |
| Valine | *N*-Carboxyvaline | 3TMS[b] |

[a] Multiple derivatives.

[b] Detected in A549 extracts.

If these *N*-carboxy derivatives were merely derivatization artifacts, I would expect to see

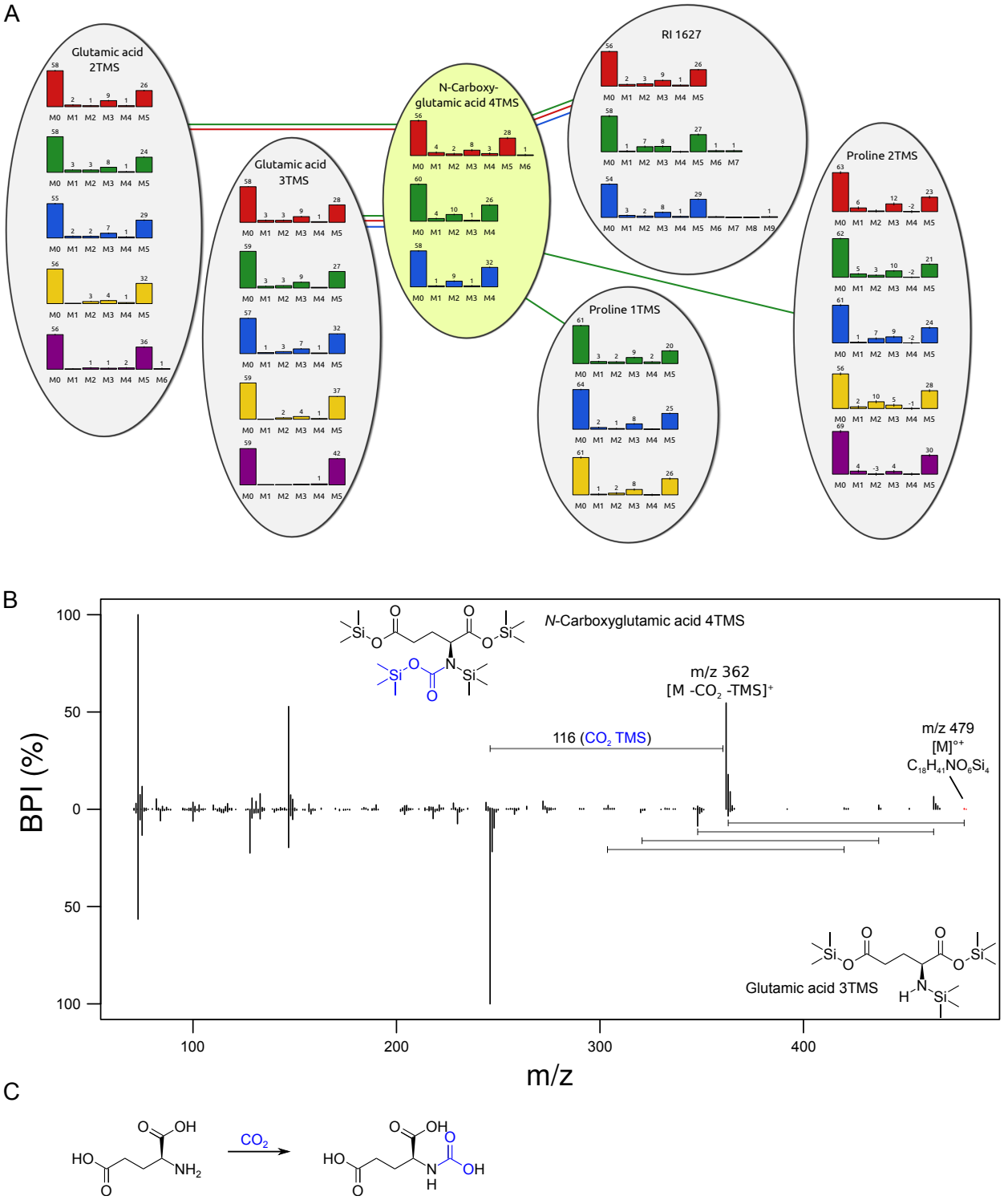



**Figure 32: Identification of *N*-carboxyglutamic acid.**
A) Analysis of MID similarities helped to identify *N*-carboxyglutamic acid 4TMS. Its MIDs from [U-$^{13}$C]glutamine labeling were found to be highly similar to those of glutamic acid, proline, and an unidentified compound. Color code indicates oxygen concentrations: 21%, 15%, 10%, 5%, 1%.
B) Mass spectra of *N*-carboxyglutamic acid 4TMS and glutamic acid 3TMS. The two compounds show highly similar fragmentation patterns, the fragments are shifted by 116 units.
C) *N*-carboxyglutamic acid is formed through the reaction of glutamic acid with carbon dioxide.

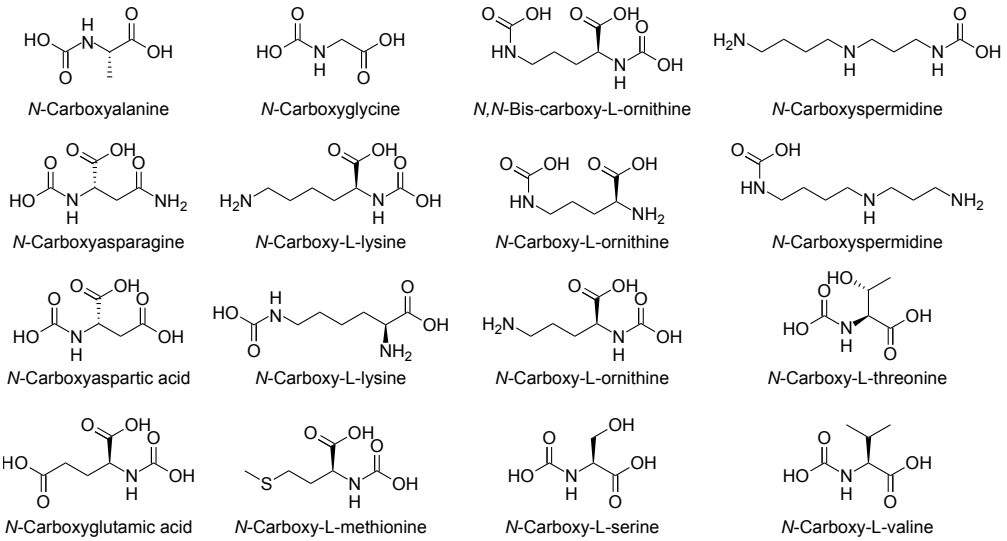



**Figure 33: *N*-carboxyamines formed during incubation of the amines and amino acids with $NaHCO_3$, derivatization with MSTFA and GC-MS measurement.**

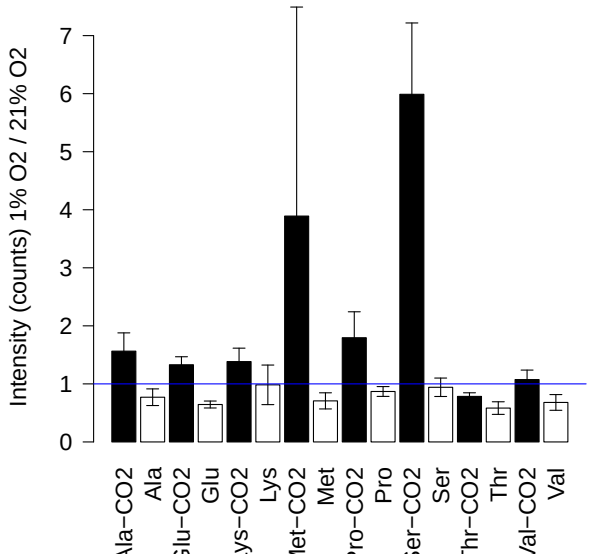



**Figure 34: Levels of *N*-carboxyamino acids ("-CO2") and native amino acids at 1% $O_2$ relative to 21% $O_2$.** *N*-carboxy compounds mostly do not show the same trend as their native counterparts. Error bars indicate mean±SD of the ratio.

a good correlation of the levels of the native amino acids and their *N*-carboxy derivatives, but they were often behaving contrarily, suggesting an, at least partial, metabolic origin (Figure 34). Although some experiments indicated an increase in *N*-carboxy amino acid levels under hypoxia, this trend was not fully reproducible. However, an increase in spontaneous *N*-carboxylation under hypoxic conditions could be explained as a result of pH regulating mechanisms which are especially important under hypoxia when protons are accumulating from lactic acid formation and other processes:[142] Decreasing intracellular pH is known to be balanced by bicarbonate influx through sodium-dependent bicarbonate transporters which gain even more importance under hypoxia.[209] The increased carbonate influx might increase formation of *N*-carboxy amino acids. Unfortunately, these compounds are not very stable and thus hard to study in more detail.[210]

### Putative 2-acetamidoglucal

The MIDs of the compound at RI 1823 were found to be strongly affected by hypoxia, both after [1,2-$^{13}$C]glucose and [U-$^{13}$C]glutamine labeling (Figure 35A,B). Fractional enrichment from glucose was decreasing with hypoxia and fractional enrichment from glutamine was strongly increasing, similar to other TCA cycle intermediates, hinting to a connection of the unknown compound with the TCA cycle. The heaviest mass isotopomer detected after [U-$^{13}$C]glutamine labeling was $M_2$.

The simplest $C_2$ fragment derived from the TCA cycle is an acetyl group from acetyl-CoA. The assumption of an acetyl group in the molecule is consistent with the observed decrease in fractional enrichment from [1,2-$^{13}$C]glucose and increase in fractional contribution from [U-$^{13}$C]glutamine (Figure 35A,B), which is caused by the well-known increase in glutamine contribution to citrate biosynthesis from which the acetyl-CoA is derived (section 1.4).

The $M_2$ abundance of RI 1823 was similar to that of glucose and less affected by hypoxia than the other mass isotopomers, indicating a substructure metabolically close to glucose (Figure 35B). This was further confirmed by the mass spectrum of RI 1823 which showed ions characteristic for TMS derivatives of hexoses ($m/z$ 117, 129, 131, 133, 204, 217, 218).

The heaviest fragments observed for RI 1823 were $m/z$ 404 and $m/z$ 419, with $m/z$ 419 probably being $M^{\bullet+}$ and $m/z$ 404 being $[M\text{-}CH_3]^+$ (Figure 35C). According to the nitrogen rule,[25] this compound most probably contains an odd number of nitrogens. The presence of one nitrogen atom was confirmed in a separate stable isotope labeling experiment with [amide-$^{15}$N]glutamine. Taken together, these findings pointed to a nitrogen-containing compound consisting of a sugar- and acetyl-moiety. Such compounds are found in the hexosamine pathway[16,17] which is known to play an important role in cancer metabolism.[211–213]

Assuming a hexosamine-like structure, the elemental composition of $M^{\bullet+}$ was determined to be $C_8H_{10}NO_5Si_3C_9H_{27}$ and later confirmed by GC-APCI-TOF-MS. Subtracting the 3 TMS

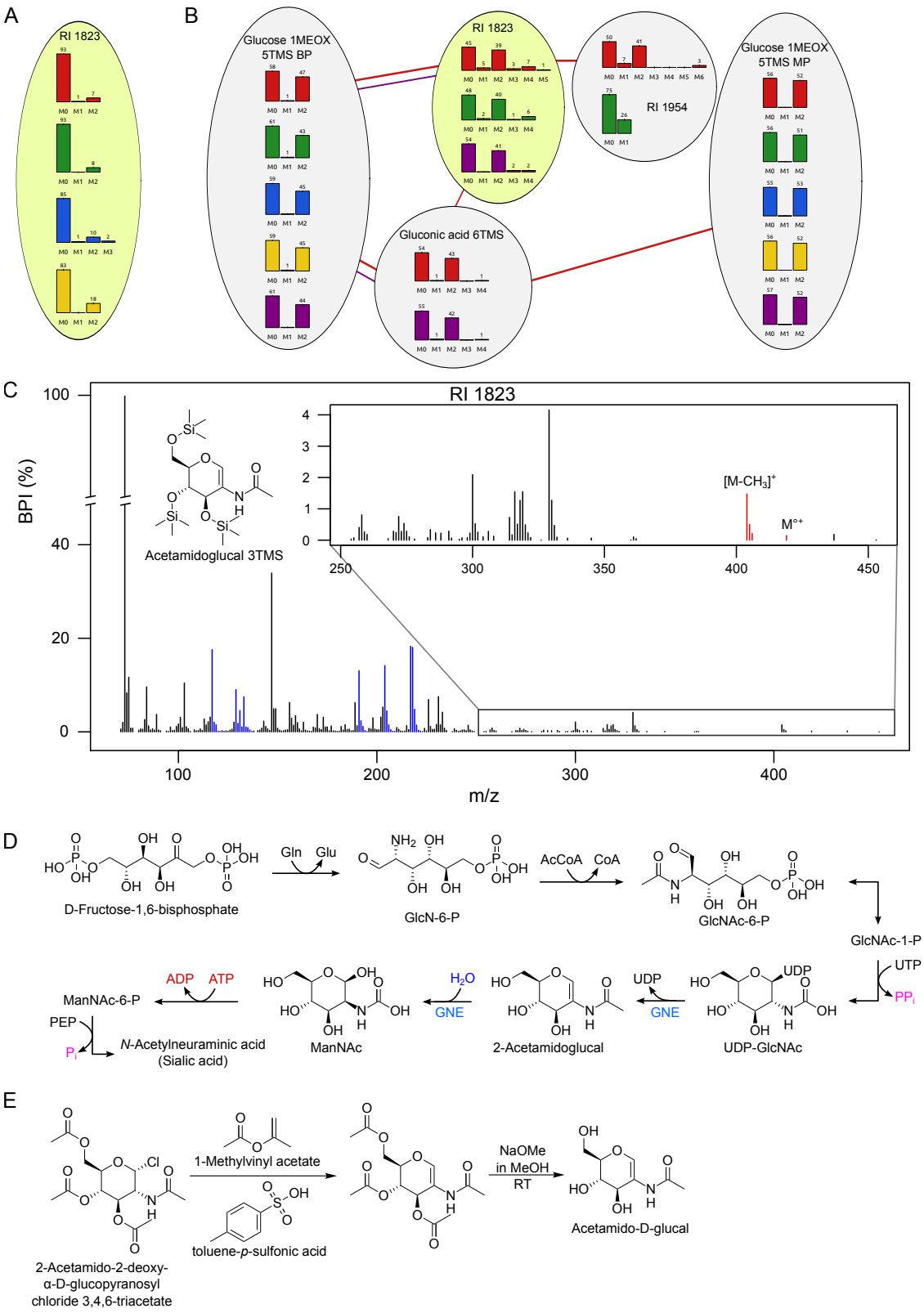



**Figure 35: The compound RI 1823 was putatively identified as 2-acetamidoglucal.**
A) Fractional contribution of [U-$^{13}$C]glutamine to RI 1823 is decreasing under hypoxia. Color code indicates oxygen concentrations: 21%, 15%, 10%, 5%, 1%.
B) MID similarity with glucose ([1,2-$^{13}$C]glucose tracer) hints to glucose-related substructure in RI 1823.
C) The mass spectrum of RI 1823 shows fragments that are characteristic to TMS derivatives of hexoses.
D) Acetamidoglucal is a metabolite of the hexosamine pathway. It is an intermediate in the epimerization of UDP-GlcNAc and is known to accumulate in patients with gene defects in UDP-*N*-acetylglucosamine 2-epimerase. [214]
E) Attempted synthesis of 2-acetamidoglucal.

groups leads to $C_8H_{13}NO_5$ for the underivatized molecule. A ChemSpider database search revealed 2-acetamidoglucal as a potential candidate (Figure 35C).

Unfortunately, there was no authentic standard of acetamidoglucal commercially available, to confirm the identity of RI 1823. The mass spectrum provided by Kamerling *et al.*[214] matched only partially. I attempted to prepare the compound from the precursor 2-acetamido-2-deoxy-$\alpha$-D-glucopyranosyl chloride 3,4,6-triacetate in a similar way as described by Pravdić *et al.*[215,216] (Figure 35E). However, after GC-MS analysis of the reaction mixture there was only a low-abundant peak with a high spectrum similarity to RI 1823. As the main product, I obtained *N*-acetyl-D-glucosamine (GlcNAc). Apparently, a nucleophilic substitution of chlorine occurred as the first reaction step instead of the intended elimination of HCl. Therefore, the identification of RI 1823 as 2-acetamidoglucal can only be seen as putative, but very likely.

2-Acetamidoglucal is an intermediate in the epimerization of uridine diphosphate (UDP)-GlcNAc to *N*-acetyl-D-mannosamine (ManNAc) by UDP-*N*-acetylglucosamine 2-epimerase (GNE)[217] (Figure 35D) and elevated levels of this compound have been detected in the urine of a patient with sialuria, a gene defect in GNE.[214] Besides these findings, there is surprisingly little literature—there are currently only 6 reports mentioning 2-acetamidoglucal indexed in PubMed.[2]

However, the hexosamine pathway is well known to play an important role in cancer. Serine and threonine residues in proteins can be post-translationally and reversibly modified with GlcNAc. It has been hypothesized that the level of protein *O*-GlcNAcetylation is a sensor of cellular energy status.[218] GlcNAc biosynthesis depends on both glucose and glutamine, the two major carbon substrates of mammalian cells, and acetyl-CoA, a metabolic hub of many anabolic and catabolic processes. Therefore, high availability of GlcNAc indicates a high energy status. Protein *O*-GlcNAcetylation has been shown to be necessary for increased glutamine uptake and catabolism in cancer cells,[211] probably via indirect stabilization of HIF-1.[212] Moreover, *N*-acetylneuraminic acid (NeuNAc) and its derivatives (sialic acids) are synthesized via 2-acetamidoglucal (Figure 35D). Sialic acids occur as terminal groups in glycans in the extracellular matrix[219] which are associated with metastasis.[220] Urinary NeuNAc has recently been identified as prognostic and diagnostic biomarker for lung cancer.[11]

### RI 1651 — *N*-acetylaspartic acid

The unidentified compound with RI 1651 was one of the compounds with strongly changing MIDs in response to hypoxia (section A.9). The analysis of MID similarity after [U-$^{13}$C]glutamine labeling indicated the vicinity of this compound to aspartic acid and citric acid (Figure 36. Especially one fragment ($m/z$ 245) of this compound showed a very high similarity

[2] http://www.ncbi.nlm.nih.gov/pubmed/?term=2-acetamidoglucal [Accessed: January 2015]

to aspartic acid. There are at least two more carbons present in the molecule, as apparent from the presence of the M+6 mass isotopomer of the heaviest fragment ($m/z$ 304). MID analysis after [1,2-$^{13}$C]glucose labeling showed high MID similarity to glycerol-3-phosphate indicating metabolic proximity to the lower part of glycolysis. Following these hints, the compound was eventually identified as the 3TMS derivative of *N*-acetylaspartic acid. A potential role of this compound in mitochondrial-cytosolic acetyl-shuttling is discussed in more detail in subsection 3.5.4.

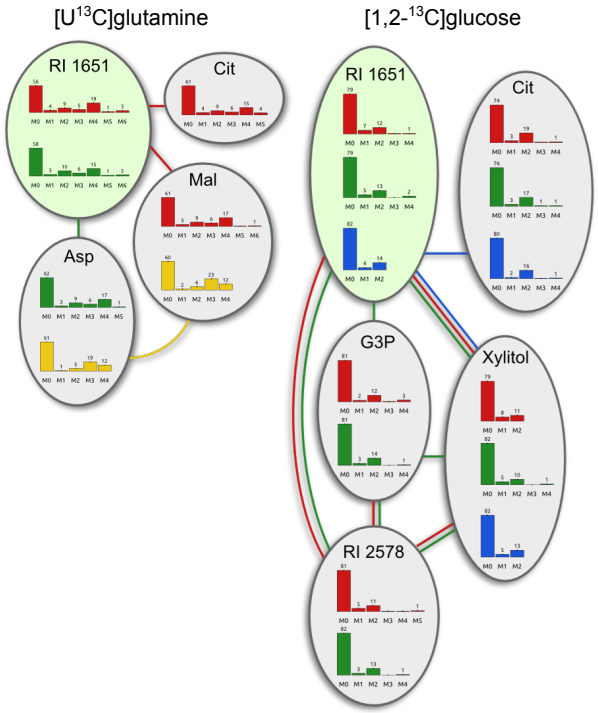



**Figure 36: MID similarity analysis for RI 1651 facilitated its identification as *N*-acetylaspartic acid.** MIDs of fragment $m/z$ 304 after [1,2-$^{13}$C]glucose and [U-$^{13}$C]glutamine labeling are shown, along with compounds with the most similar MIDs. Color code indicates oxygen concentrations: 21%, 15%, 10%, 5%.

### 3.5.4 NAT8L and *N*-acetylaspartic acid-based acetyl shuttling in lung cancer cells

*N*-acetylaspartic acid was one of the metabolites popping up in the mass isotopomer abundance variation analysis (section A.9). *N*-acetylaspartic acid is well-known to occur in very high abundance in neuronal tissue[221] where it is synthesized from aspartic acid and acetyl-CoA by the long unknown enzyme NAT8L, which is also known as Shati.[222] A function of *N*-acetylaspartic acid in non-neuronal mammalian cells has not been described so far. However, NAT8L expression was recently shown to play an important role in lipid turnover in brown

adipocytes by an unknown mechanism.[188] It was therefore surprising at first to find *N*-acetylaspartic acid biosynthesis in cancer cells. However, the measurement of an authentic standard confirmed its identity and stable isotope labeling clearly excluded an exogenous origin.

In the mammalian brain, *N*-acetylaspartic acid is known as the precursor of the neuropeptide *N*-acetylaspartylglutamic acid (Figure 37).[223] Additionally, *N*-acetylaspartic acid was shown to be an essential carrier of acetyl units for their transport from neurons to oligodendrocytes, where they are cleaved off by aspartoacylase (ASPA) to be used for myelin biosynthesis (Figure 37).[224,225] Perturbation of this transport is known to lead to pathological conditions known as Canavan's disease.[226]

Potential functions of neuronal *N*-acetylaspartic acid have been proposed,[227] but its role in cancer cells has not been elucidated yet. Its acetyl-transport function in the brain, together with the known increase in lipid biosynthesis of cancer cells, make it interesting to speculate about a similar role of *N*-acetylaspartic acid in tumor cells. As D'Adamo *et al.*[228] many years ago, I hypothesized a role of *N*-acetylaspartic acid in the transport of acetyl-CoA from the mitochondrial matrix to the cytoplasm.

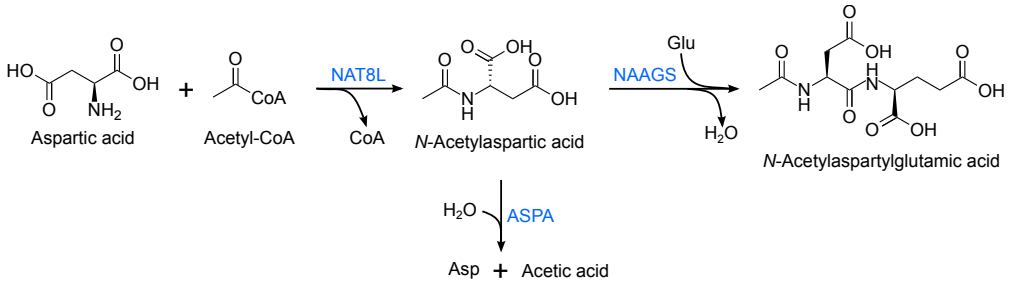



**Figure 37: Neuronal metabolism of *N*-acetylaspartic acid.** *N*-acetylaspartic acid is produced from acetyl-CoA and aspartic acid by NAT8L. It can be hydrolyzed to aspartic acid and acetic acid by ASPA or elongated to *N*-acetylaspartylglutamic acid by *N*-acetylaspartylglutamic acid synthetase (NAAGS).

### *N*-acetylaspartic acid is produced by NAT8L in A549 cells

To find out whether *N*-acetylaspartic acid is also produced by NAT8L in A549 cells, an siRNA-mediated knockdown of NAT8L was performed. Due to the lack of adequate primers for NAT8L, silencing efficiency could not be assessed. With the primers used by Pessentheiner *et al.*[188] we could amplify NAT8L in human neuroepithelial stem cells (cDNA kindly provided by Sarah Nickels), but did not yield any product in A549 cells. However, siNAT8L-transfection had a significant effect on the levels of several metabolites (Figure 38). *N*-acetylaspartic acid levels dropped significantly to 7.5% of the levels measured in the siCtrl-transfected cells, thus providing strong evidence that NAT8L is responsible for *N*-acetylaspartic acid production in A549 cells. Besides *N*-acetylaspartic acid, glycerol-3-phosphate levels were significantly

decreased to 13% of the levels in siCtrl-transfected cells. Another compound, glsRI 2718.62 was later identified as *N*-acetylaspartylglutamic acid 3TMS.

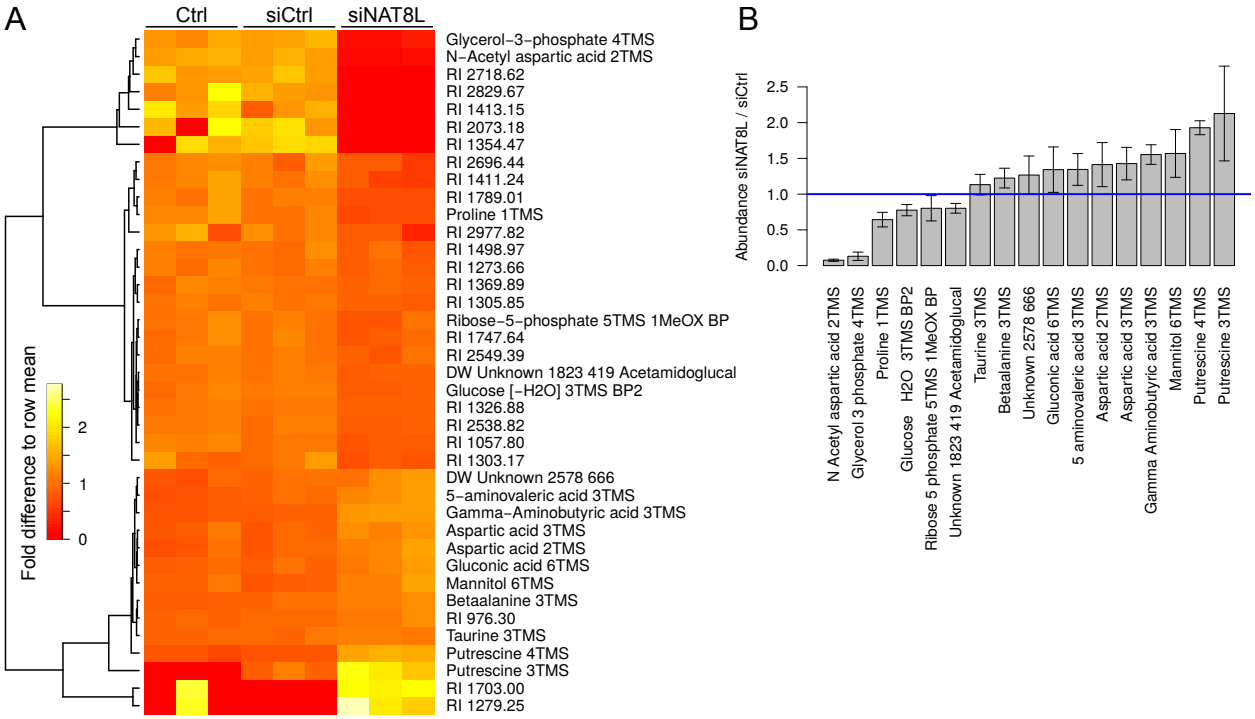



**Figure 38: Significantly changed metabolite levels** 48 h **after transfection with NAT8L-targeting siRNA.**
A) Heatmap of raw signal intensities normalized to row mean. Metabolites with Welch's *t*-test $p < 0.05$ are shown. Unidentified compounds are labeled with their RI. RI 2718.62 has been identified as *N*-acetylaspartylglutamic acid [-$H_2O$] 2TMS.
B) Identified compounds from (A). Signal intensity after knockdown normalized to siCtrl.

## Occurrence of *N*-acetylaspartic acid and NAT8L across other non-neuronal cell types

After having confirmed the identity and endogenous origin of *N*-acetylaspartic acid, as well as its production by NAT8L, I was interested in the occurrence of this metabolite and enzyme in other cell types and tissues. Therefore I analyzed other GC-MS measurements and found *N*-acetylaspartic acid to be present also in human hepatocellular carcinoma cells (HepG2; data generated by Nadia Battello) and human prostate epithelial cells (RWPE-1 and RWPE-2; data generated by Dr. Thekla Cordes). Interestingly, *N*-acetylaspartic acid levels in RWPE-2 cells were higher than in RWPE-1 cells (Figure 39A). The two cell lines are identical, except that RWPE-2 was further transformed with *v-Ki-ras*, rendering the cells tumorigenic.[229] Furthermore, I found *N*-acetylaspartic acid to be present in primary lung tissue (data generated by Jenny Ghelfi). For a previous study, lung biopsies from cancer patients were taken from healthy and tumor tissue of the same patient. Although not statistically significant (paired *t*-test, $n = 19$, $p = 0.11$), *N*-acetylaspartic acid levels were increased in most tumor

samples, pointing towards a role in cancer metabolism (Figure 39B). For a subset of these lung tissue samples, NAT8L mRNA levels were determined and found to be higher in tumor tissue with a median difference of 8%, but again not statistically significant (paired $t$-test, $n = 11$, $p = 0.19$; data not shown). Moreover, using the GENEVESTIGATOR[230,231] gene expression search engine, I found high NAT8L expression in a large number of cancer cell lines (data not shown).

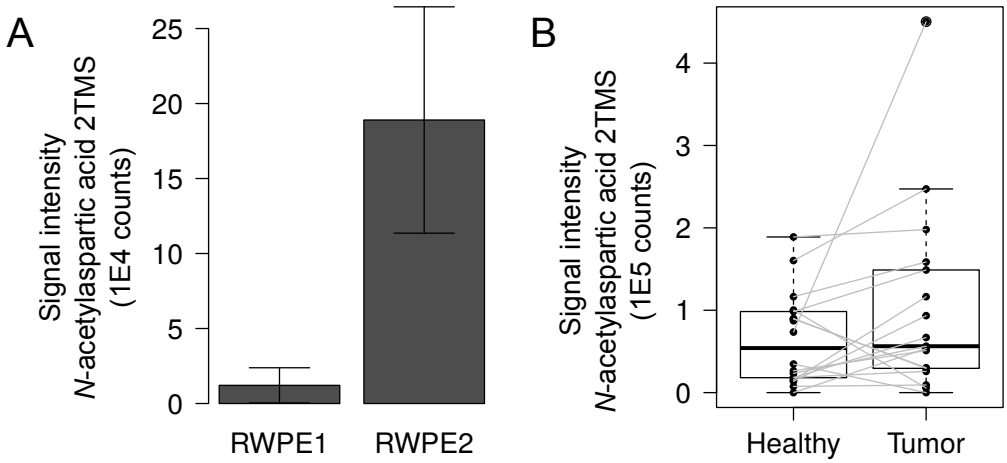



**Figure 39: Occurrence of *N*-acetylaspartic acid in cancer cells.**
A) Tumorigenic RWPE-2 cells show higher levels of $N$-acetylaspartic acid than non-tumorigenic RWPE-1 cells (Welch's $t$-test, $n = 3$, $p = 0.053$).
B) $N$-acetylaspartic acid levels in lung tissue of cancer patients. Abundance is higher in most tumor samples. Points are means of three technical replicates. One outlier not plotted ($24 \times 10^5$ signal intensity in tumor sample). Paired $t$-test, $n = 19$, $p = 0.11$.

Thorough literature search showed that $N$-acetylaspartic acid has also been detected in other metabolomics studies on cancer cells which, however, did not follow up on this finding.[231–234] Especially Sreekumar *et al.*[232], reporting sarcosine as an urinary biomarker for prostate cancer progression, also detected $N$-acetylaspartic acid. In their hierarchical cluster analysis, it clustered closely to sarcosine and its abundance is progressing with cancer stage but is not commented on. Reitman *et al.*[235] found $N$-acetylaspartic acid, as well as $N$-acetylaspartylglutamic acid which is produced[223] from $N$-acetylaspartic acid, to be present in glioma, but depleted in IDH1 mutated cells.

### Hints towards alternative splicing of NAT8L in A549 cells

When trying to verify the NAT8L silencing efficiency by qPCR, I initially failed to amplify NAT8L cDNA. Multiple primer pairs were tested. One pair of primers that failed to amplify NAT8L in A549 was working well for human neuroepithelial stem cells (cDNA kindly provided by Sarah Nickels) and was used by Pessentheiner *et al.*[188] in adipocytes.

*nat8l* codes for three exons (Figure 40). The amplicon of aforementioned primer pair spanned the exon 1 – exon 2 junction. To account for potential alternative mRNA splicing,

I designed primers for amplicons within either exon 1 or exon 2. Using the exon 2-specific primers, I was able to amplify NAT8L cDNA in A549 cell extracts, but not with the exon 1-specific primers, suggesting that exon 1 is not expressed in A549 cells. There are no alternative splicing variants of human NAT8L reported and the biological relevance is unclear.

Absence of exon 1 would leave a single translation initiation site with the same reading frame at 505 bp that would give rise to a truncated 134 aa long, 15.5 kDa heavy protein, still containing the *N*-acetyltransferase domain. This protein would lack the predicted transmembrane domain at aa 121 to aa 141. The N-terminal region may furthermore contain subcellular localization signals and alternative splicing could explain the different protein localization reported by different groups.[222,236–238] However, no localization signals were reported within the N-terminal region.

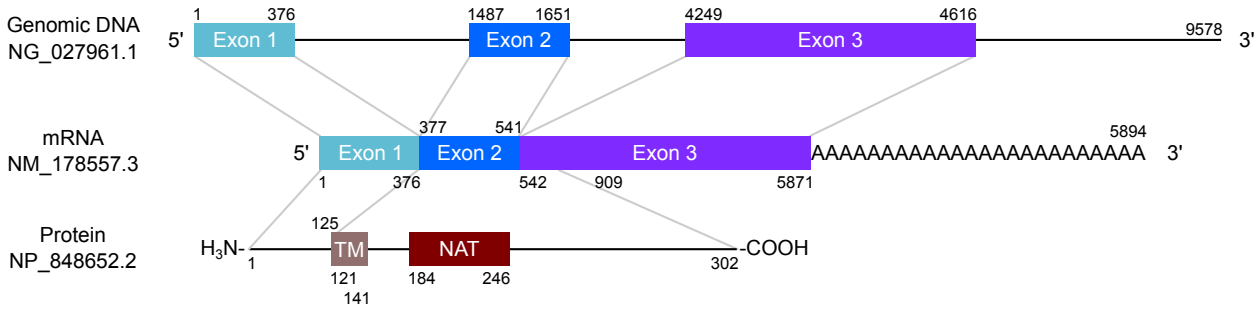



**Figure 40: NAT8L mRNA splicing and protein domains.** *nat8l* codes for three exons. The protein has a predicted transmembrane domain. TM: predicted transmembrane domain; NAT: *N*-acetyltransferase domain. (Source: NCBI protein/gene/nuccore databases)

## NAT8L and glycerol-3-phosphate

NAT8L knockdown significantly changed the levels of several metabolites (Figure 38). The metabolite, the levels of which were most affected, was glycerol-3-phosphate. A correlation between glycerol-3-phosphate and *N*-acetylaspartic acid concentrations has also been measured — but not discussed — by others.[231,232]

Glycerol-3-phosphate is a lipid building block as well as a key metabolite in electron shuttling to the mitochondrium via the glycerol-3-phosphate shuttle (Figure 41A). In humans, there exist multiple enzymes that are involved in glycerol-3-phosphate metabolism. Several *O*-acyltransferases are able to transfer fatty acids from and onto glycerol-3-phosphate, and glycerol-3-phosphate dehydrogenases catalyze the reversible $NAD^+$/NADH dependent oxidation to the glycolytic intermediate dihydroxyacetone phosphate. Glycerol-3-phosphate dehydrogenase (GPD)2 together with GPD1 or GPD1L constitute the glycerol-3-phosphate shuttle that transfers electrons from cytosolic NADH to mitochondrial FAD.[239] The resulting $FADH_2$ can transfer the electrons to the respiratory chain.

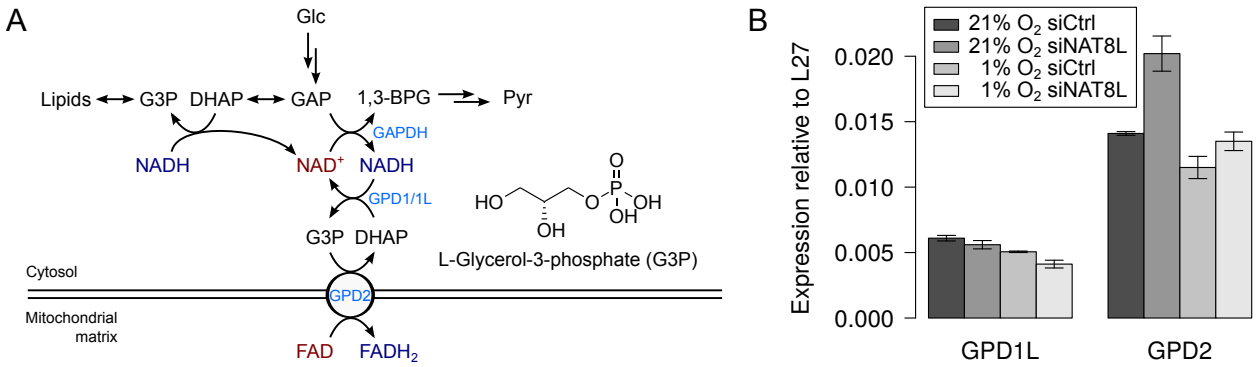



**Figure 41: Glycerol-3-phosphate (G3P) metabolism.**
A) G3P is a precursor for glycerolipids and part of the G3P shuttle. Cytosolic dihydroxyacetone phosphate (DHAP) can be reduced to G3P by NAD-dependent cytosolic GPD1 or GPD1L and reoxidized by mitochondrial membrane-bound GPD2 with concomitant reduction of FAD, effectively transferring two electrons to the mitochondrium.
B) GPD gene expression analysis in A549 cells. GPD1L expression is slightly reduced in siNAT8L transfected cells. GPD2 is induced in siNAT8L transfected cells under normoxia. Bars show mean±SD of two replicates (Welsh's $t$-test, $n = 2$, $p > 0.05$).

GPD1L and GPD2, but not GPD1 were found to be expressed in A549. GPD1L expression was only little affected by NAT8L silencing, under both normoxia and hypoxia (Figure 41B). GPD2 expression, however, increased by over 40% at normoxia. GPD2 is the mitochondrial membrane-bound, usually glycerol-3-phosphate consuming enzyme. Its induction could explain the observed reduction of glycerol-3-phosphate levels after NAT8L silencing (Figure 38).

### NAT8L, glycerol, and reductive glutamine metabolism

To obtain hints on the functional role of *N*-acetylaspartic acid and NAT8L in A549 cells and to find the link to glycerol-3-phosphate metabolism, a stable stable isotope labeling experiment was performed using [U-$^{13}$C]glucose and [U-$^{13}$C]glutamine, both under normoxia and hypoxia. NAT8L silencing was confirmed by qPCR and found to be more efficient under hypoxia (Figure 44).

Under normoxia, the most significantly changed MIDs were those of glycerol and glycerol-3-phosphate (Figure 42A). In siNAT8L-transfected cells, the relative abundance of fully labeled glycerol-3-phosphate increased significantly by about 5 percentage points, whereas the overall levels dropped dramatically by almost 90% (Figure 38). Fully labeled glycerol-3-phosphate is derived from the glycolytic intermediate dihydroxyacetone phosphate via GPD and the unlabeled isotopologue is derived from breakdown of unlabeled lipids from the growth medium (Figure 41). Therefore, the labeling experiment shows a relative increase in its *de novo* synthesis and a relative decrease in the contribution of lipid turnover.

The labeling of glycerol is puzzling. Although not statistically significant, the labeling of

glycerol changed inversely to glycerol-3-phosphate. This difference points to the irreversibility of glycerol-3-phosphate dephosphorylation, potentially due to the lack of glycerol kinase. However, isotopic enrichment in glycerol was higher than in glycerol-3-phosphate, which is surprising because it is derived from the latter and is therefore expected to be rather less enriched. Furthermore, the $M_2$ abundance in glycerol but not in glycerol-3-phosphate cannot be explained from its glycolytic origin, raising the question of other biosynthetic pathways leading to glycerol (Figure 42A).

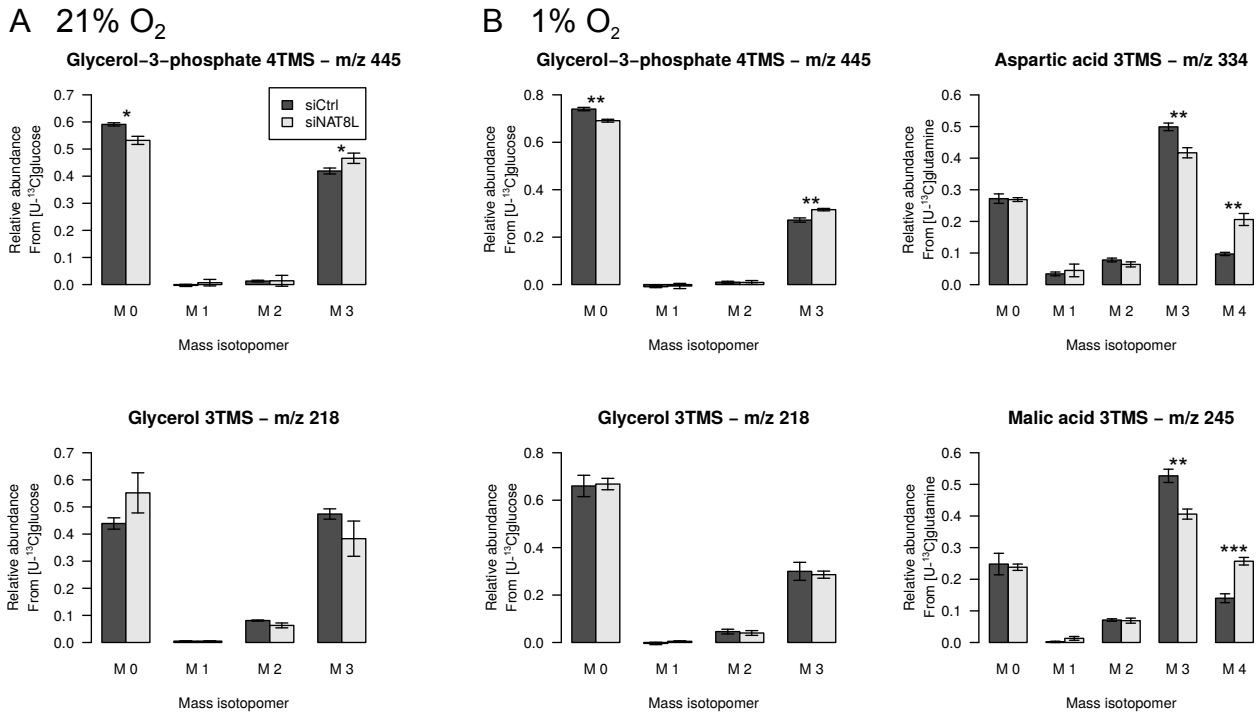



**Figure 42: Changes in MIDs due to NAT8L knockdown after labeling with [U-$^{13}$C]glucose and [U-$^{13}$C]glutamine under (A) normoxia and (B) hypoxia.**
Welch's $t$-test, $n = 3$, *: $p < 0.05$, **: $p < 0.01$, ***: $p < 0.001$.

Under hypoxia, glycerol-3-phosphate MIDs showed a similar change as under normoxia (Figure 42B). However, its overall isotopic enrichment was about 25% lower, indicating a higher contribution from lipid turnover than under normoxia. An increased turnover of exogenous lipids under hypoxia has also been observed by others.[167] For glycerol, no significant difference in isotopic enrichment, but again a $M_2$ fraction of unknown origin was observed.

Whereas under normoxia NAT8L knockdown barely affected MIDs of TCA cycle-associated metabolites, neither from [U-$^{13}$C]glucose nor [U-$^{13}$C]glutamine labeling, the effect under hypoxia was quite drastic (Figure 42). There was a massive shift from $M_3$ to $M_4$ abundance of malic acid and aspartic acid after [U-$^{13}$C]glutamine labeling in siNAT8L cells. [$^{13}C_3$]malic acid is produced from reductive glutamine metabolism via carboxylation of [$^{13}C_5$]2-oxoglutarate followed by ACLY-action; [$^{13}C_4$]malic acid is produced from glutamine oxidation in the TCA

cycle (section A.9). Therefore, the changes in MIDs indicate that NAT8L knockdown inhibits reductive glutamine metabolism or induces its oxidative metabolism. Reductive glutamine metabolism provides a means to sustain lipid biosynthesis fully on glutamine, independent of glycolysis and PDH activity.[8]

### Model of *N*-acetylaspartic acid function in cancer cells

Several potential functions of *N*-acetylaspartic acid, including a role as osmoregulator,[240] have been proposed,[227] but its role in cancer cells has not been elucidated yet. In earlier studies in rat brain, D'Adamo *et al.*[228] and Patel & Clark[241] suggested a role of *N*-acetylaspartic acid in mitochondrial-cytosolic carbon transport, refining the observation of D'Adamo & Yatsu[242], that *N*-acetylaspartic acid carbon is incorporated into brain lipids. Meanwhile, it has been shown by others that in the nervous system *N*-acetylaspartic acid is an essential carrier of acetyl units for their transport from neurons to oligodendrocytes where they are used for myelin biosynthesis.[224,225] Since then, the potential role of *N*-acetylaspartic acid in intracellular acetyl-transport was not further pursued.

The change in the ratio of oxidative and reductive glutamine metabolism upon NAT8L knockdown may be due to implications of NAT8L with mitochondrial-cytosolic-acetyl-transport (Figure 42B). Since cancer cells require high amounts of cytosolic acetyl-CoA for *de novo* lipid biosynthesis, such an acetyl-shuttling mechanism could be beneficial. Traditionally, cytosolic acetyl-CoA is thought to be derived from citrate by ACLY (section 1.4).

In the following, I postulate an *N*-acetylaspartic acid-based acetyl-shuttle and will discuss the aforementioned data in that light, starting with metabolic fluxes under hypoxia, where NAT8L knockdown led to larger MID changes.

**Reductive glutamine metabolism under hypoxia** With decreasing oxygen concentrations, the cytosolic acetyl-CoA pool is increasingly supplied from glutamine catabolism.[8] At the stage of 2-oxoglutaric acid, glutamine carbon can be oxidized by OGDH or reduced by IDH. Oxidative glutamine metabolism by OGDH, succinate dehydrogenase (SDH), malate dehydrogenase (MDH), malic enzyme, and PDH can produce mitochondrial acetyl-CoA (Figure 43A). This acetyl-CoA can be exported to the cytosol as citrate. The oxaloacetate resulting from ACLY action is transported back as malate. Reductive glutamine metabolism generates cytosolic citrate which can be cleaved by ACLY to provide acetyl-CoA. The resulting oxaloacetic acid can be transported to the mitochondrium, to form another molecule of acetyl-CoA via malic enzyme and PDH, which can be exported again as citrate to provide a second cytosolic acetyl-CoA molecule. However, in this scenario, citrate synthesis or transport may pose a bottleneck, because there are two competing reactions, via citrate synthase and IDH, producing citrate which could lead to product inhibition,[164] limiting cytosolic acetyl-CoA

production.

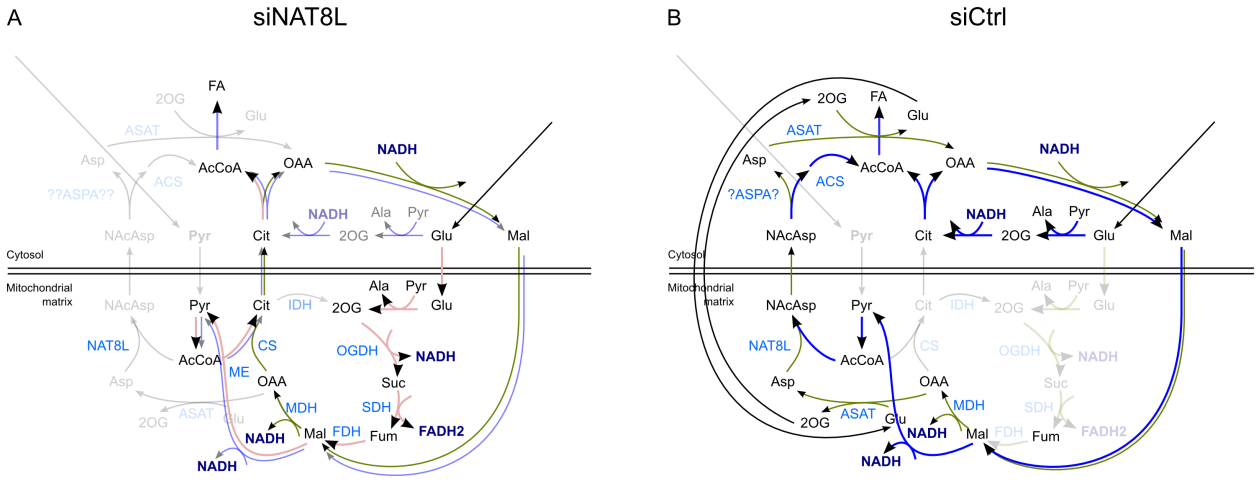



**Figure 43: Model of cytosolic-mitochondrial acetyl-CoA transport under hypoxia where acetyl-CoA-carbon is mainly derived from glutamine.**
A) Classical model for providing cytosolic acetyl-CoA via oxidative (red) or reductive (blue) glutamine metabolism, not considering a role of NAT8L. Both pathways rely on slow citrate formation via citrate synthase, and reductive IDH flux and citrate synthase reaction possibly suffer from product inhibition by citrate.
B) Proposed alternative model based on the *N*-acetylaspartic acid shuttle for acetyl transport. *N*-acetylaspartic acid is formed by NAT8L in the mitochondrium, transported to the cytosol where it is cleaved and the resulting acetate is used to provide acetyl-CoA. Here, no competing citrate-forming reactions are necessary.

**An *N*-acetylaspartic acid shuttle avoids product inhibition in reductive carboxylation of 2-oxoglutarate** *N*-acetylaspartic acid is likely produced in the mitochondrium,[188] although there are contradicting results.[222,227,236–238,243] Assuming its mitochondrial biosynthesis and a cytosolic hydrolyzing enzyme, which may or may not be identical with neuronal ASPA, an acetyl-shuttling from the mitochondrium would be possible (Figure 43B). Mitochondrial synthesis of *N*-acetylaspartic acid, export to the cytosol and deacetylation could provide cytosolic acetate which can be activated to acetyl-CoA by acyl-CoA synthetase (ACSS). Acetate was recently shown by multiple groups to be a lipogenic substrate in hypoxic cancer cells.[160–162] Although the net carbon transport in this scenario does not differ from the ACLY-based one (Figure 43B), the *N*-acetylaspartic acid model (Figure 43B) does not suffer from product inhibition by citrate and circumvents the citrate synthase reaction, which was shown to proceed rather slow in comparison to the transamination of oxaloacetate to aspartate.[244]

The proposed model can explain the observed relative decrease in reductive carboxylation of 2-oxoglutarate upon NAT8L silencing: Without the possibility of *N*-acetylaspartic acid-based acetyl-transport, reductive glutamine metabolism is inhibited by citrate production from citrate synthase and its export to the cytosol. However, as a consequence of NAT8L silencing,

ACLY should become more important for cytosolic acetyl-CoA. A respective induction of ACLY gene expression, was not observed (Figure 44) unlike in brown adipocytes as reported by Pessentheiner *et al.*[188]. The observed induction of both cytosolic and mitochondrial ACSS and the repression of citrate synthase in response to NAT8L silencing may also point towards a lack of acetyl-CoA (Figure 44): ACSS induction can replenish acetyl-CoA and reduced citrate synthesis via citrate synthase helps to conserve the available acetyl-CoA.

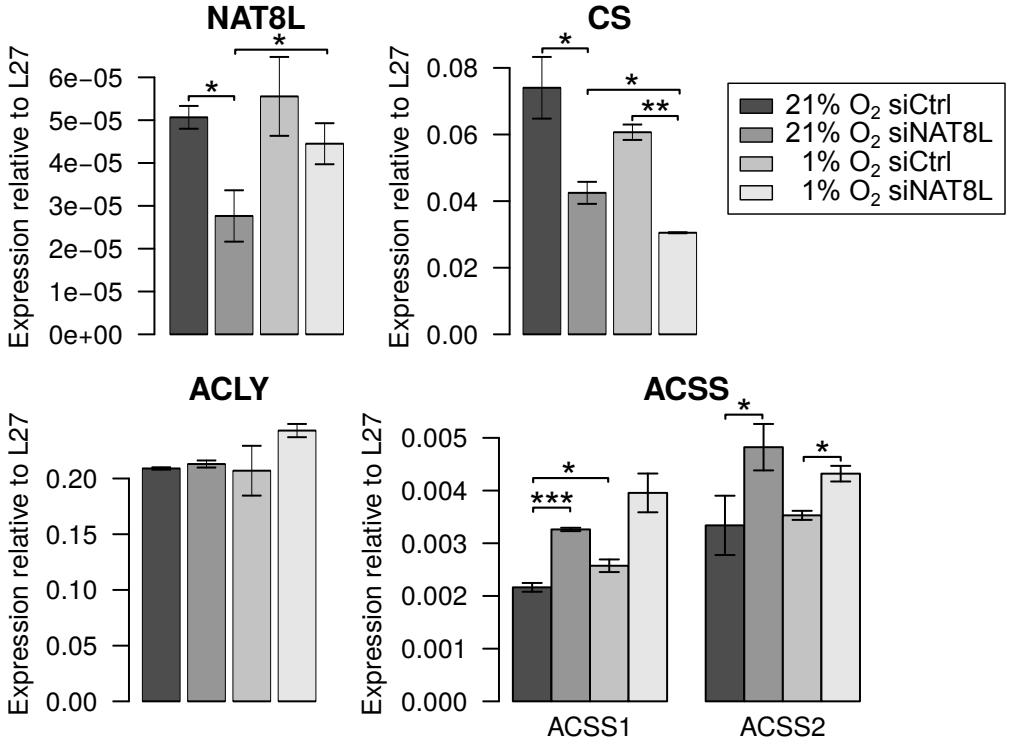



**Figure 44: Gene expression levels in response to NAT8L silencing.** NAT8L silencing is stronger under normoxia. It induces gene expression of ACSS1 and ACSS2 and represses citrate synthase expression. ACLY gene expression is not affected. *: Welch's $t$-test $p < 0.05$ ($n = 3$), **: $p < 0.01$, ***: $p < 0.001$. $n = 3$, except for ACLY expression and "1% $O_2$ siNAT8L" levels, where $n = 2$.

**Electron flow is affected under normoxia** Under normoxia, changes in MIDs were smaller than under hypoxia and there was no significant induction in ACLY expression. However, there was an induction in GPD2 gene expression. An *N*-acetylaspartic acid shuttle could explain these observations. Under normoxia, most cytosolic acetyl-CoA is derived from glucose via glycolysis, PDH, citrate synthase, and ACLY (Figure 43A). PDH flux, mitochondrial acetyl-CoA availability and citrate synthesis are high at normoxia; citrate and ACLY can satisfy the demand in cytosolic acetyl-CoA, detectable flux ratios and MIDs of TCA cycle intermediates do not change. However, the citrate detracted from the TCA cycle is not available for oxidation and concomitant reduction of $NAD^+$ and FAD to fuel oxidative phosphorylation. Therefore, GPD2 expression is induced to balance mitochondrial $FADH_2$

levels via the glycerol-3-phosphate shuttle and to regenerate cytosolic $NAD^+$ to sustain a high glycolytic rate (Figure 41).

The *N*-acetylaspartic acid shuttle mostly decouples acetyl-CoA export from NADH and $FADH_2$ production in the TCA cycle (Figure 43B). Provided there is enough mitochondrial acetyl-CoA, the *N*-acetylaspartic acid shuttle can deliver cytosolic acetyl-CoA and mitochondrial NADH, independently of the TCA cycle. The TCA cycle can produce more reducing equivalents for oxidative phosphorylation as no citrate is detracted.

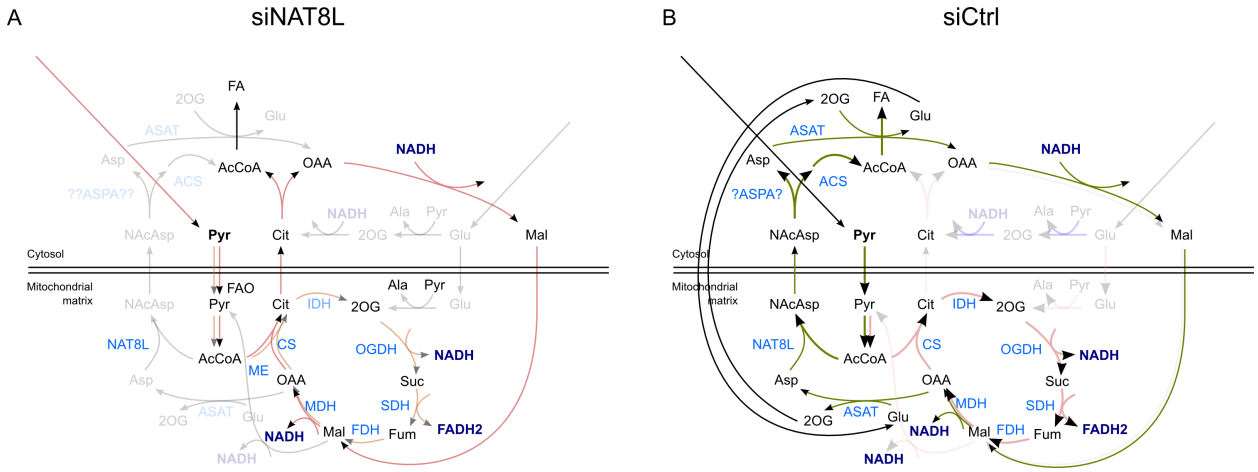



**Figure 45: Model of the role of *N*-acetylaspartic acid and NAT8L in normoxic cancer cells where cytosolic acetyl-CoA is mainly derived from glucose.**
A) Classical model for providing cytosolic acetyl-CoA via glycolysis, PDH, citrate synthase, and ACLY. Citrate synthesis is slow and a potential bottleneck for NADH and $FADH_2$ production when it is exported to the cytosol.
B) The alternative model based on the *N*-acetylaspartic acid shuttle can export acetyl-CoA from the mitochondrium without interfering with the TCA cycle, assuming that there is enough mitochondrial acetyl-CoA available. The slow [244] citrate synthase reaction is circumvented.

### NAT8L silencing lowers cell proliferation, even more than ACLY silencing

To further assess if the proposed model is adequate, I performed a NAT8L, ACLY, and combined NAT8L+ACLY silencing and monitored cell proliferation. I expected that the individual silencing of NAT8L and ACLY could be balanced by the respective other enzyme, but that their combined silencing, in which case the cells fully depend on acetate or fatty acid from the growth medium, would significantly reduce their proliferation rate.

Under normoxic conditions, the individual or combined gene silencing led to a growth inhibition of around 30% (Figure 46). The effect of the combined silencing was not significantly different from the individual ones.

Under hypoxia, the growth inhibitory effect of NAT8L silencing was 20% after 72 h and therefore lower than under normoxia (Figure 46). More surprisingly, ACLY silencing seemed to not affect cell growth under hypoxia.

The lower effect of NAT8L silencing under hypoxia could be explained by the increased reductive glutamine metabolism leading to citrate that can provide cytosolic acetyl-CoA by ACLY. However, then the combined silencing should result in increased growth inhibition which I did not observe.

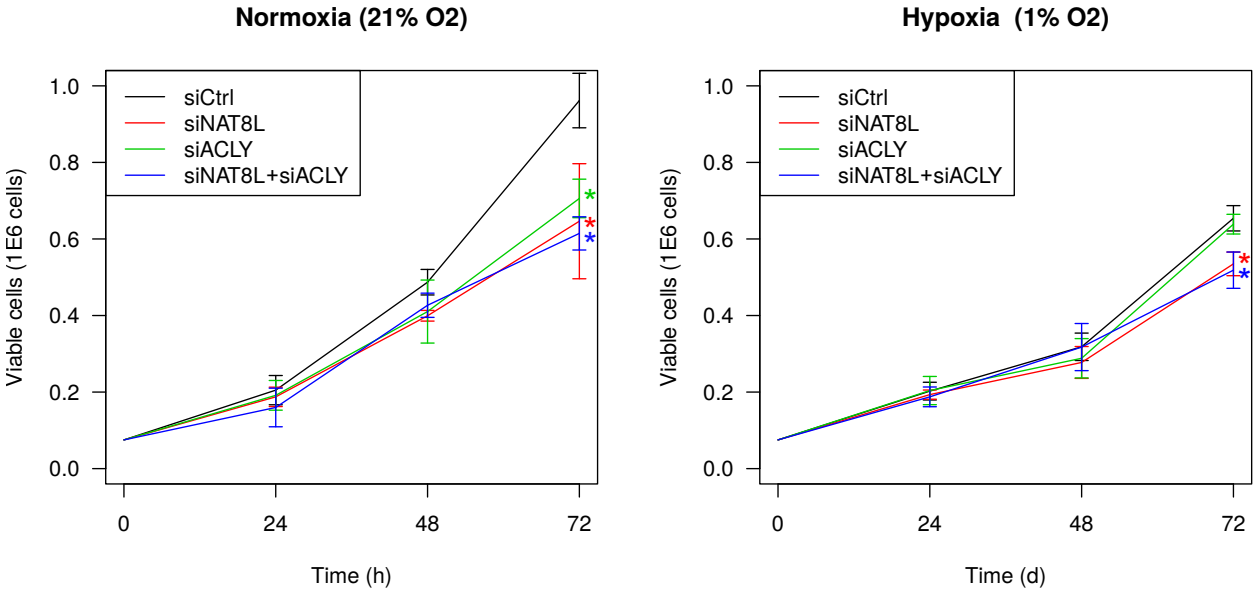



**Figure 46: A549 cell growth during 72 h of NAT8L, ACLY, and combined NAT8L+ACLY silencing under normoxic and hypoxic conditions.**
Mean±standard deviation of the mean (SD) of cell numbers of three replicates are shown. Welch's $t$-test, $n = 3$, *: $p < 0.05$.

**Validation of the model** To validate the postulated function of NAT8L and *N*-acetylaspartic acid, additional experiments are required. A stable isotope labeling experiment using *N*-([$^{13}C_2$]acetyl)-aspartic acid as tracer will show the fate of the acetyl-moiety of *N*-acetylaspartic acid in cancer cells, which is expected to mostly fuel fatty acid production and cytosolic or nuclear acetylation reactions (*e.g.* histones).

The effect of NAT8L overexpression in cancer cells should be analyzed — I expect increased NAT8L activity to confer a proliferative advantage over normal cells. Furthermore, if the proposed model is correct, NAT8L over-expression should rescue the impaired proliferation in response to ACLY-silencing. Downstream of NAT8L action, a knockdown of ACSS2 should be performed in an acetate free growth medium, where I expect it to have a similar effect on metabolic fluxes as the NAT8L knockdown.

For the suggested acetyl-shuttling mechanism, an *N*-acetylaspartic acid-hydrolyzing enzyme is required.Benuck & D'Adamo[245] already reported that *N*-acetylaspartic acid can be hydrolyzed in several peripheral tissues. The obvious enzyme candidate is ASPA, which is responsible for oligodendrocytic *N*-acetylaspartic acid hydrolysis. However, I was not able

to detect its expression in A549 cells by qPCR. Furthermore, the subcellular localization of NAT8L in A549 cells needs to be confirmed.

### Alternative role of *N*-acetylaspartic acid as precursor of the glutamate receptor agonist *N*-acetylaspartylglutamic acid

NAT8L may also play a role in providing the precursor for *N*-acetylaspartylglutamic acid biosynthesis (Figure 37). I found this metabolite also to be produced in A549 cells. *N*-acetylaspartylglutamic acid was described as an agonist of the extracellular type 3 metabotropic glutamate receptor ($mGluR_3$).[246,247] $mGluR_3$ receptors are $G_i$-coupled which, when activated, inhibits adenylate cyclase and subsequently cAMP-dependent protein kinase A (PKA). The $mGluR_3$ receptors are well-known in neurons, but were also found to be expressed in cancer cells,[248,249] where they modulate cell proliferation.[250] The reduction of intracellular cAMP and cGMP levels by *N*-acetylaspartylglutamic acid[247] may lower their inhibitory effect on cell proliferation and invasion.[251,252] Not only do *N*-acetylaspartylglutamic acid modulate PKA signaling, but also does PKA signaling regulates *N*-acetylaspartic acid and *N*-acetylaspartylglutamic acid levels, forming a negative feedback loop.[253] independently of PKA signaling, Guo *et al.*[254] found that $mGluR_3$ regulates proliferation of neural progenitor cells via JNK/ERK signaling, protein kinases that are well-known in cancer signaling.[255] Further evidence is needed to assess if the lack of $mGluR_3$ activation by *N*-acetylaspartylglutamic acid alone could explain the observed metabolic flux alterations observed in response to NAT8L silencing.

### Summary

The presented experiments provide evidence for an *N*-acetylaspartic acid-based mitochondrial-cytosolic acetyl-shuttle in cancer cells. Such a mechanism would explain the recently reported importance of ACSS for the activation of acetate.[160–162] However, not exogenous but *N*-acetylaspartic acid-derived, endogenous acetate is the main substrate of ACSS. NAT8L expression may play a role in the invasiveness of cancer cells as suggested by its elevated levels in tumorigenic RWPE-2 cells as compared to non-tumorigenic RWPE-1 cells, and by the correlation of *N*-acetylaspartic acid levels with prostate cancer progression as observed by others.[232] Furthermore, NAT8L or the not yet identified hydrolase may provide a potential drug target for peripheral cancer. Most non-cancer cells, besides adipocytes and oligodendrocytes, have a lower acetyl-CoA demand and might therefore not suffer from inhibition of the *N*-acetylaspartic acid-shuttle.

Although I presented evidence for an integral role of NAT8L in cellular metabolism, still many questions remain to be answered: Is ASPA the cytosolic *N*-acetylaspartic acid deacetylase? How is NAT8L regulated? Is the *N*-acetylaspartic acid-shuttle present in all

cell types, or did only cancer cells specifically hijack the neuronal transport system? How is *N*-acetylaspartic acid transported to the cytosol? By the mitochondrial dicarboxylate carrier SLC25A10,[256] or by the oligodendrocytic NaDC3[257]? Independently of these open question, the observed pronounced effects of NAT8L silencing on metabolite levels and metabolic fluxes despite the moderate silencing efficiency, point towards an integral function within cellular metabolism.

### 3.5.5 Acetylated compounds as proxies for acetyl-CoA labeling

Acetyl-CoA is at the interface of glycolysis, the TCA cycle, fatty acid biosynthesis, terpene biosynthesis, $\beta$-oxidation and several other pathways, therefore it represents an important metabolic hub. Isotopic enrichment of acetyl-CoA after stable isotope labeling experiments can provide important information on fractional contribution of the given tracer to the aforementioned pathways. Unfortunately, acetyl-CoA cannot be analyzed with GC-MS due to its low volatility. Other sophisticated methods like isotopomer spectral analysis (ISA)[258] and mass isotopomer distribution analysis[81] have been used to deduce acetyl-CoA labeling from the labeling of fatty acids.

In section A.9, I demonstrated how the labeling of mass spectrometric fragments of *N*-acetylaspartic acid can be used to infer acetyl-CoA labeling (Figure 47). This approach has the advantage, that it does not require the separate work-up and measurement of fatty acids. Besides *N*-acetylaspartic acid, any other purely endogenous acetylated compound holds information on acetyl-CoA labeling. To determine the labeling of the acetyl-moiety either suitable mass spectrometric fragments need to be found, or it must be ensured that the non-acetyl-moiety cannot be labeled from the applied tracer. A candidate for the latter case is the putative 2-acetamidoglucal (see above). For example in the case of glutamine labeling in non-gluconeogenetic cells under isotopic steady state the 2-acetamidoglucal MID represents the acetyl-CoA MID, since the glucal-carbon cannot be labeled from the glutamine tracer.

Moreover, the two compounds *N*-acetylaspartic acid and 2-acetamidoglucal may be suitable to determine compartment-specific acetyl-CoA MIDs. If the above model (Figure 43) is correct, then *N*-acetylaspartic acid is synthesized in the mitochondrium and, thus, provides labeling information on the mitochondrial acetyl-CoA pool, whereas hexosamine acetylation occurs in the cytosol[259] and 2-acetamidoglucal or other acetyl-hexosamines provide information on the cytosolic acetyl-CoA pool. The mitochondrial acetyl-CoA labeling provides information on acetyl-CoA derived from acetate via ACSS1, $\beta$-oxidation of fatty acids, and most importantly PDH. The cytosolic acetyl-CoA is a mixture of exported mitochondrial acetyl-CoA, and acetyl-CoA produced from reductive carboxylation of 2-oxoglutarate and ACLY action. The observed differences in acetyl-labeling of 2-acetamidoglucal and *N*-acetylaspartic acid support

this hypothesis: The enrichment from [U-$^{13}$C]glutamine is higher in the assumed cytosolic pool, which is composed of acetyl-CoA exported from the mitochondrium, but additionally enriched by labeled acetyl-CoA derived from reductive glutamine metabolism (Figure 47).

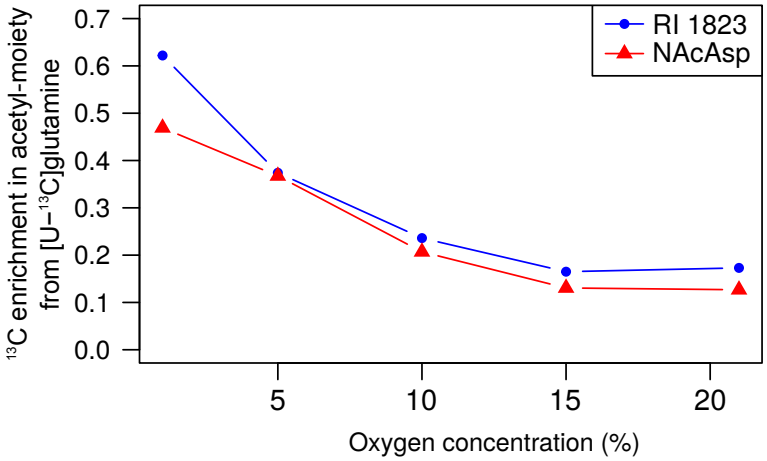



**Figure 47: Proxies for acetyl-CoA labeling.** From the mass spectra of the acetylated compounds RI 1823 and *N*-acetylaspartic acid, the labeling in their acetyl-moiety can be determined (section A.9). Graph shows the $^{13}$C enrichment after [U-$^{13}$C]glutamine labeling. The different enrichments point to two distinct subcellular acetyl-CoA pools.

### 3.5.6 Summary

This analysis of hypoxic cancer cells shows how non-targeted data acquisition and data analysis approaches are valuable tools to generate initial hypotheses leading to new biological insights. After performing stable isotope labeling experiments, mass spectrometric analysis, and the non-targeted detection of isotopically enriched compounds, qualitative analysis of isotopic enrichment provided information on active fluxes (*e.g.* production of *N*-acetylaspartic acid) and the general fate of the metabolic tracer (Figure 48B). MIDs from different experimental conditions have been systematically analyzed to detect changes in metabolic fluxes (subsection 3.2.1; *e.g.* decreased PDH flux and decreased IDH flux). Because MID similarity may indicate metabolic proximity (subsection 3.2.2), MIDs of compounds of interest have been compared to all other MIDs for pathway contextualization and to facilitate identification of unidentified compounds (*N*-carboxyglutamic acid, 2-acetamidoglucal; subsection 3.5.3).

Overall, this non-targeted approach (Figure 48B) provides information on: 1) active pathways, 2) changed fluxes and 3) compound identities. This information holds biological insights itself and will furthermore generate hypotheses for subsequent analyses (Figure 48A).

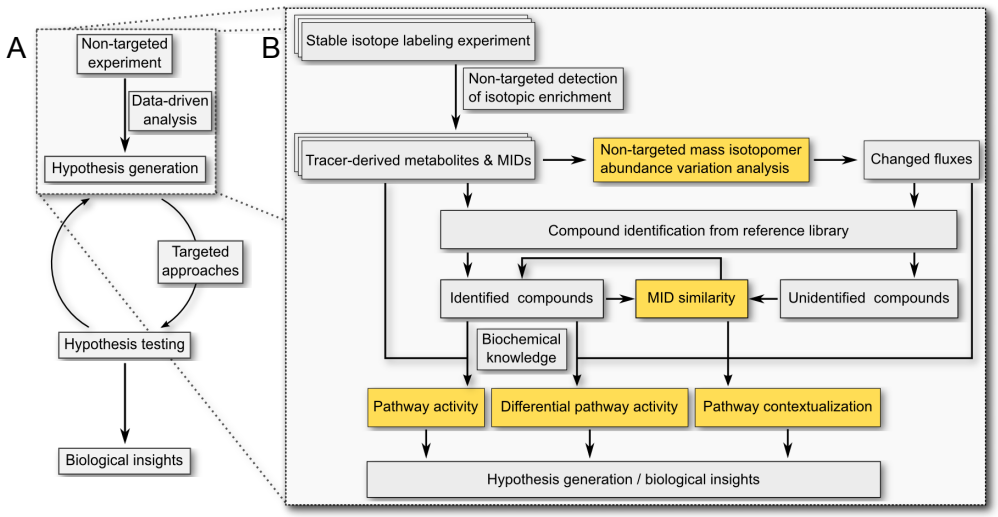



**Figure 48: Workflow for non-targeted mass isotopolome analysis.**
A) Non-targeted approaches are valuable scouting experiments for hypothesis generation. These hypotheses are tested by more targeted techniques leading to refined hypotheses and biological insights.
B) Data-driven analysis of non-targeted stable isotope labeling experiments. After stable isotope labeling experiments, active pathways and changed fluxes can be detected in a non-targeted manner from MIDs and changes therein. Compound identity and additional biochemical knowledge is only needed for further interpretation. Analysis of MID similarity between compounds can aid their identification or help to determine their biosynthetic pathway. Yellow boxes highlight novel analysis techniques presented in this work.

## 3.6 MetaboBase—Metabolomics data management

To assess the importance of specific compounds in a metabolite sample, especially for newly discovered compounds, it is valuable to know about their occurrence under specific conditions and in different sample types or organisms. Some tens of thousands of metabolite samples have been analyzed by GC-MS in the metabolomics group at the LCSB, providing an ideal data basis to answer such questions. However, to easily retrieve measurements for a given organism or condition, this data needed to be organized and annotated.

To make the accumulating GC-MS data of our group and the LCSB metabolomics platform more accessible, I set up, together with Jake Lin, a user-friendly web-accessible data management platform named MetaboBase.

This platform is used to store and retrieve information on more than 40,000 measured samples, making historic data more available. Having experimental meta-data in a well-accessible format will allow us to perform large-scale meta-analyses of different experiments.

MetaboBase is well integrated with our analytical instruments via the Agilent/HP GC/MSD ChemStation and our MetaboliteDetector software.[189]

For this project I set up the data interconversion, data import into the database, designed the database model and the workflows for the front-end, whereas Jake Lin was responsible for the actual implementation of most of the web-interface and back-end for the database connection.

MetaboBase provides the following features:

- Data warehousing
- Reference library management
- Reference library export
- Spectrum view
- Dataset annotation
- Mass spectrum annotation
- Atom transitions
- Data retrieval

### 3.6.1 Data warehousing

The data flow for MetaboBase starts directly at the acquisition computers of our GC-MS instruments (Fig. 49). After a measurement is finished, the data is directly converted to the MetaboliteDetector data format and analytical meta-information is added. The measurement data are transfered to a central server on regular intervals, multiple times a day. From there it is accessible for all group members and automatically backed up daily to an off-site server, ensuring high data security. Once the data files are on the server the

| | |
|---|---|
| GC-MS measurements | 52,203 |
| Experiments | 540 |
| Library compounds | 2940 |
| Annotated derivatives | 691 |

**Table 8: Some MetaboBase statistics [February 2015].**

associated meta-data is imported into a relational database from where it is available within the web interface (subsection 3.6.2).

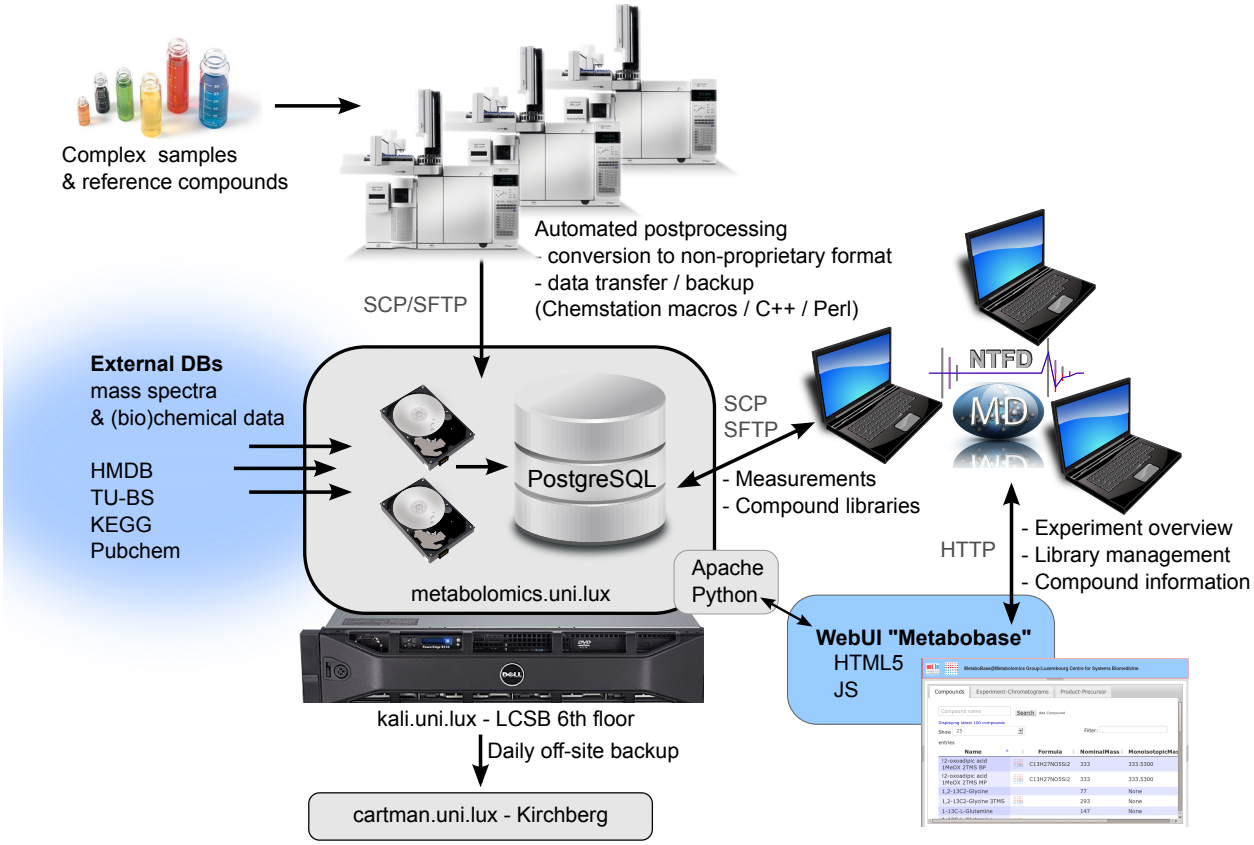



**Figure 49: MetaboBase overview.** MetaboBase is a metabolomics data management platform ensuring high data accessibility. It is well integrated with our data analysis tools and external databases.

### 3.6.2 Web front-end

The user interface of MetaboBase is a web front-end that allows for compound library management, spectrum annotation, data set annotation and data retrieval which I will explain below (Figure 50). The focus was put on user-friendliness, simple workflows and intuitive handling.

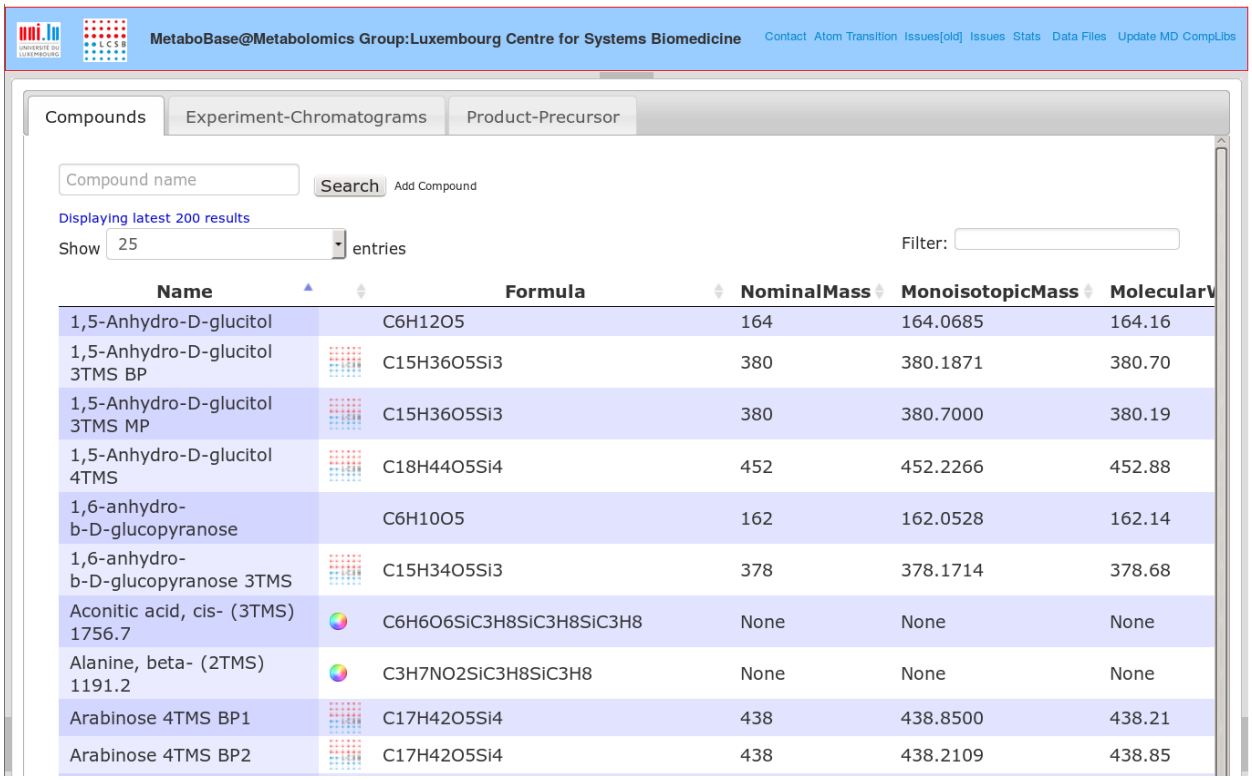



**Figure 50: MetaboBase compound library.**

## Compound library

One of the biggest issues in metabolomics still is the identification of measured compounds, which is usually done by comparing measured spectra with reference libraries. MetaboBase now allows the Metabolomics group to manage and extend these libraries centrally and more effectively.

**Compound information** Information on chemical compounds in the database can be viewed and modified via the user-friendly web interface (Figure 51). For each compound there is physical and chemical information available such as:

- name and synonyms
- molecular structure
- molecular weight, monoisotopic mass, nominal mass
- elemental composition
- INternational CHemical Identifiers (InChIs) [260]
- cross-references to other databases (KEGG, [261] HMDB, [262] PubChem, [263] ChemSpider[3])

The initial compound information was derived from HMDB [262] and subsequently supplemented with additional compounds, especially TMS and TBDMS derivatives of metabolites

[3] http://www.chemspider.com/

for GC-MS measurements. For all compounds with InChI identifiers, vector graphics of the molecular structure were generated in an automated manner.


**Figure 51:** For every compound, chemical information and cross-links to other databases are stored. For chemical derivatives for *e.g.* GC-MS measurement, the derivatization reagents and precursor metabolites are stored as well.

**Reference spectra** Mass spectra of measured authentic reference spectra can be added to a reference spectrum library. The RI and RT are assigned to all reference spectra and the spectra are linked to the original raw data files (Figure 52). All compounds that are derivatives of metabolites (section 1.2.1) are marked as such and are assigned to the native metabolites. This step is important to later map any measured derivative back to its precursor to locate it in metabolite databases and use it for bioinformatics analyses (see below).

**Reference spectra annotation** These reference spectra can be further annotated. For the determination of isotopic enrichment after stable isotope labeling experiments, the elemental composition of the fragment of interest is required (section 1.3.6 and 3.3). If such information is available, as *e.g.* obtained from FFC (section A.4), it can be directly added to the compound within the reference library (Figure 52).

For interpretation of MIDs of mass spectrometric fragments, it is furthermore necessary to know which atoms of the native metabolite are retained in the respective fragment. This information is also stored in the database. As there are no official universal identifiers for

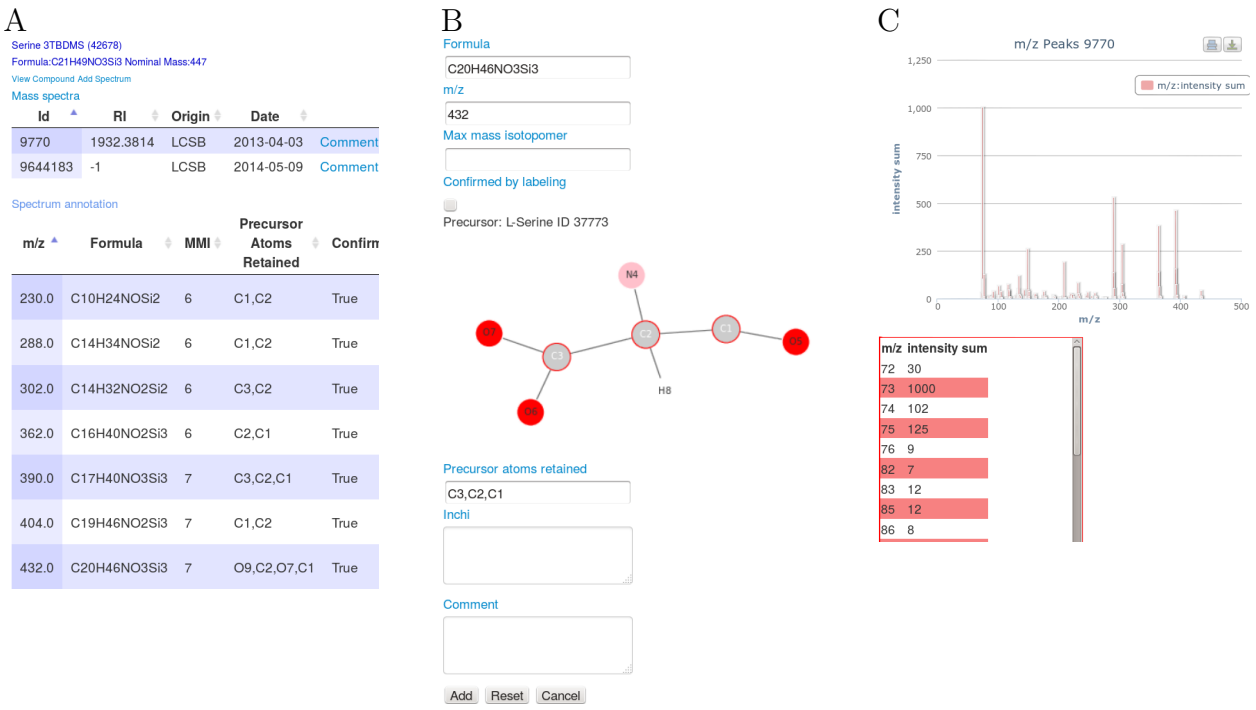



**Figure 52:** MetaboBase mass spectrum library to manage reference mass spectra.
A) The added mass spectra can be annotated with the elemental composition and atoms of the precursor retained in each fragment, which is important for MID correction (section 1.3.6 and 3.3).
B) Unambiguous selection of atoms retained within a fragment, by selecting them in the molecular graph.
C) Stored mass spectra can be viewed and downloaded as comma separated values (csv) file. Compound libraries are automatically exported for use with MetaboliteDetector.

the individual atoms within the molecule and the biochemical and IUPAC numbering often differs, it is not straightforward to unambiguously assign the atoms retained in a given fragment. Therefore, I exploited the numbering algorithm of the InChI.[260] This algorithm reproducibly assigns a unique number to every atom in the molecule. However, unfortunately, this numbering scheme is very unintuitive.[264] To counter this problem, I chose to allow for a simple and fail-safe graphical selection of retained atoms (Figure 52), which is only internally translated to the InChI numbers.

**Compound library export** All the aforementioned features are only of use when they are integrated with our data analysis toolbox. Therefore, all reference spectra in MetaboBase are automatically exported to compound libraries which are usable with MetaboliteDetector, NTFD and MIA. Separate compound libraries are generated for different derivatization methods (TMS- and TBDMS-derivatives and methyl esters). Furthermore, each library compound entry also contains the aforementioned chemical information, precursor metabolite and database identifiers. I added a new feature to MetaboliteDetector that allows to

export this data as part of the result table of a batch quantification. The included identifiers are valuable for further bioinformatics analyses like pathway mapping or enrichment analysis.[265–268]

### Dataset annotation

The more than 40,000 GC-MS measurements (Table 8) that our research group has accumulated over the last few years make it hard to keep an overview over all the experiments and datasets. It is therefore crucial to have a proper data management in place. MetaboBase allows to store and retrieve experiment meta-data to these samples. This includes a description of the experimental setup, the different groups and conditions, the sample type and origin, and most importantly a reference to the respective pages in laboratory notebooks (Figure 53). For stable isotope labeling experiments additional information on the applied tracer is included.

Having experimental meta-data in a well-accessible format makes it easy to retrieve historic data, which may be of interest for new studies. Furthermore, it can be the basis for large-scale meta-analyses of different experiments that eventually will lead to biomarker discoveries or the like.

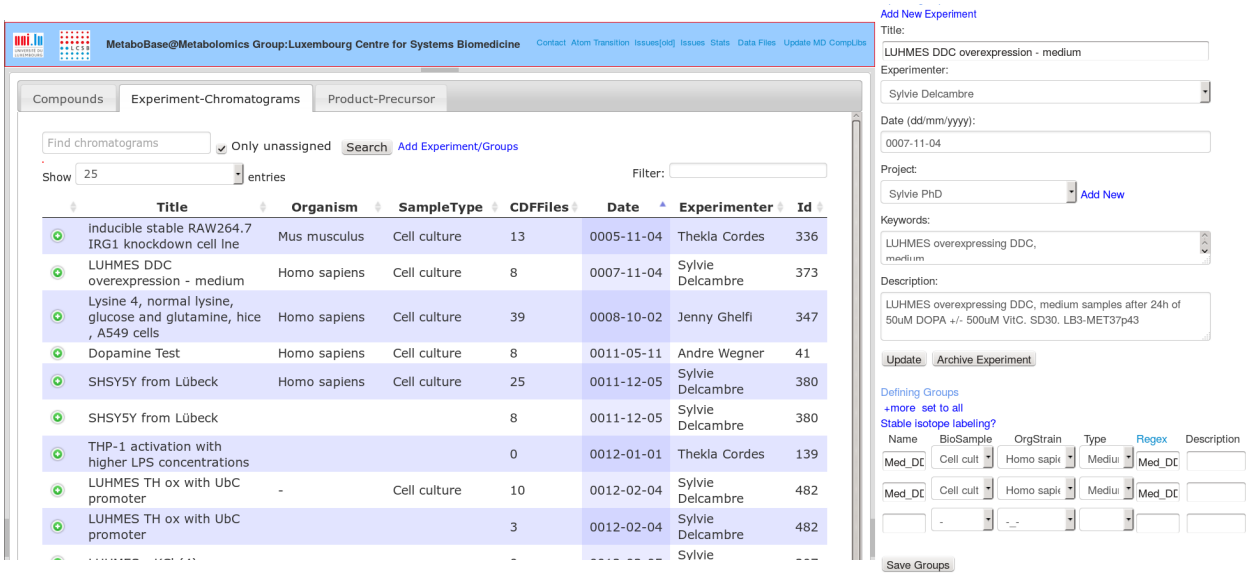



**Figure 53: MetaboBase allows to add experiment meta-data and use it to search for specific data sets.** The used organisms and strains, as well as treatment groups within an experiment are annotated in a machine-readable manner, allowing for future large scale analyses.

### 3.6.3 Implementation

**Relational database model** Except for the actual raw data files, all MetaboBase data is stored in a relational PostgreSQL[4] database. The database model is illustrated in Figure 54.

[4]http://www.postgresql.org/

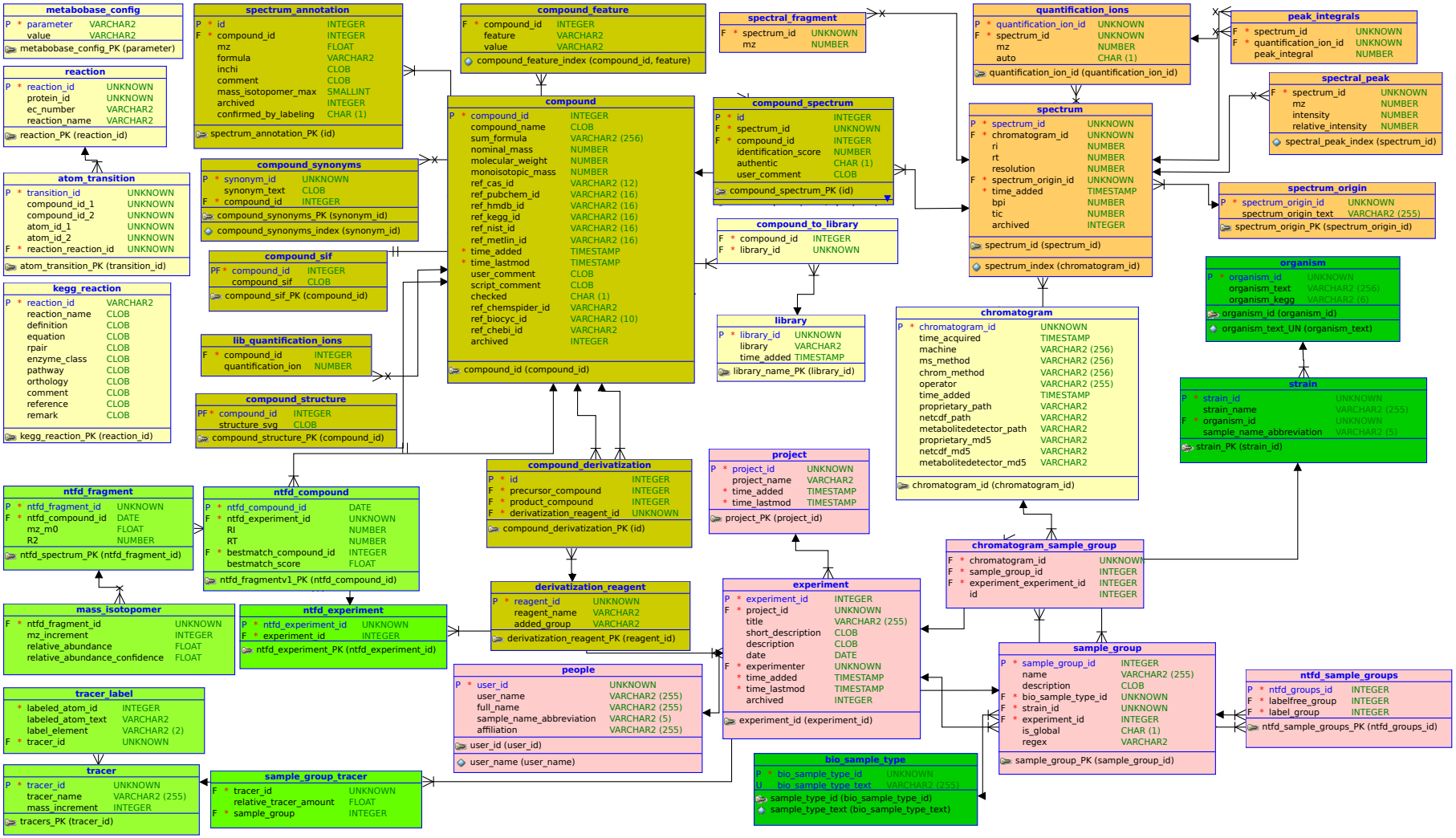



**Figure 54:** MetaboBase database model. olive: tables mainly for compound library; pink: tables mainly for experiment metadata; light green: tables for stable isotope labeling experiments; dark green: tables for biological materials.

**Web interface** The web front-end of MetaboBase is running on an Apache web server. It is written in HTML5 and connected to the PostgreSQL database via a python back-end and asynchronous JavaScript and JSON (JavaScript Object Notation) (AJAJ).

**Data transfer** After a measurement on our GC-MS instruments is finished, the data is converted from the proprietary ".D" format to the open netCDF[269,270][5] format using ChemStation macros (section B.1). By a gcmspreprocessing tool written in C++ by Dr. Karsten Hiller, meta-information on the data acquisition parameters is added to the data files before they are further converted to the MetaboliteDetector data format.

**Data import** The data import into MetaboBase is performed by a Perl script. This script is run every few hours as a Cron job and adds all new measurements to the database.

**Molecular structures** Molecular structures displayed in the compound library are generated from InChI identifiers in an automated manner using Open Babel.[271] The structures are stored in the database as scalable vector graphics (SVG) and served to the front-end by a PHP script.

**Compound library export** Compound libraries are exported by a dedicated C++ program which is based on MetaboliteDetector source code. It is run daily or on user request to generate libraries for different chemical derivatization methods.

**Molecular graphs** Molecular structures for display to select retained atoms in mass spectrometric fragments (see above) are generated by a dedicated C++ program. This program uses the InChI library[260] to extract atom numbers and connectivity from the stored InChI identifiers and generates a molecular graph in the simple interaction file (SIF) format[6] which is subsequently stored in the database.

### 3.6.4 Summary

MetaboBase allows for easy-to-use and yet sophisticated metabolomics data management. Its strength is the profound tailor-made integration with the proprietary data acquisition software as well as our in-house data analysis applications. This project is still ongoing to further improve the increasingly important data management infrastructure. Possible future applications can be meta-analyses across several data sets, to for example analyze the distribution of specific metabolites across different species or to detect metabolic signatures

[5] http://www.unidata.ucar.edu/packages/netcdf/index.html
[6] http://wiki.cytoscape.org/Cytoscape_User_Manual/Network_Formats

of different cell types or perturbations. Furthermore, data processing capabilities could be included for simple web-based data analysis as shown by others.[272]

# 4 Summary & Outlook

In this work I presented novel tools and workflows for metabolomics data analysis (section 3.4), especially for data derived from stable isotope labeling experiments (section 3.1, 3.2). I developed MIA, added new features to MetaboliteDetector and NTFD, and contributed to the development of ICBM and FFC (section 3.3, section A.1). I applied these tools and showed how they can be well combined to gain insights into biological phenomena (section 3.5). Furthermore, I developed MetaboBase, an integrated metabolomics data management platform (section 3.6).

To my knowledge, MIA is the first tool, that allows not only for the non-targeted detection and quantification of stable isotope labeled compounds, but also for subsequent data analysis. MIA allows for the non-targeted detection of metabolic flux changes based on MID variation analysis and aids compound identification by generating MID similarity-based compound networks.

Applying MIA for the analysis of cancer cell metabolism enabled me to detect hypoxia-induced metabolic flux changes and a non-linear change from glucose to glutamine as main substrate for fatty acid biosynthesis. I illustrated how non-targeted stable isotope labeling can be a valuable tool for initial hypothesis generation in the study of metabolism. I was able to detect unknown and unanticipated metabolites, and MIA provided vital information for their identification. I detected several *N*-carboxy derivatives of amine metabolites. Following these results I showed that under hypoxia, cancer cells exhibit increased BCAA uptake and catabolism. I found that the propionyl-CoA generated during catabolism of isoleucine and valine give rise to odd-chain fatty acids. I showed that A549 cells produce *N*-acetylaspartic acid and provided evidence that it is not only involved in neuronal-oligodendrocytic acetyl-shuttling, but also in intracellular mitochondrial-cytosolic transport in cancer cells. This pathway was not known to exist in cancer cells. It sheds new light on the recent findings on acetate utilization in cancer cells and may provide further drug targets.

MetaboBase, the metabolomics data management platform I designed, helped the metabolomics research group members to manage and retrieve their large amount of GC-MS data. This system keeps historic data together with experimental meta-data in a defined machine-readable format. These data are now more available for further analyses.

This work illustrates how non-targeted analysis of stable isotope labeling can be a valuable tool for initial hypothesis generation in the study of metabolism. The detected unanticipated

metabolites showed again that knowledge on species and cell type-specific metabolism is still incomplete and non-targeted approaches are necessary. There is for example no information on the role of odd-chain fatty acids in human metabolism available. It remains to be investigated whether they arise only due to unspecific enzymes, or whether they are specifically produced, for example to sequester toxic propionyl-CoA.

The developed data analysis tools like MIA can be applied in further studies, also in contexts other than cancer metabolism. The field of non-targeted stable isotope labeling analysis is very young, but seems to have great potential. Complementary approaches based on LC-MS could provide further interesting insights into areas of metabolism which are not accessible by GC-MS. Most tools in the field of non-targeted stable isotope labeling analysis have only been developed over the recent years and major progress is to be expected in the coming years.

# 5 Acknowledgements

To finish, I thank especially my supervisor Dr. Karsten Hiller for the supervision and the opportunity to work on an exciting project in the emerging field of non-targeted stable isotope labeling.

I thank Prof. Hannelore Daniel and Prof. Rainer Schuhmacher for taking over the external reviews of this thesis and joining my defense committee. I thank Prof. Rudi Balling and Dr. Reinhard Schneider for their time to be part of my *comité d'encadrement de thèse* (CET), taking over the internal reviews, and joining my defense committee.

I thank Jake Lin for the great collaboration on the MetaboBase project, I enjoyed putting up this platform together. Thanks to Dr. Thekla Cordes for providing the data on hypoxic cancer cells and other interesting datasets to work on and to test and validate MIA. Many thanks to Xiangyi Dong for the wet-lab support during the final weeks of preparing this thesis. A big thanks to Dr. Andre Wegner for valuable feedback and discussions on scientific writing. Thanks to Dr. Christian Jäger for all support and discussions on analytical and organic chemistry. Thanks to Dr. Johannes Meiser for great collaboration on the dopamine review. For proof-reading and helpful comments on my thesis, I thank Dr. Andre Wegner, Sylvie Delcambre, Dr. Christian Jäger, Dr. Johannes Meiser and Dr. Karsten Hiller.

Furthermore, I thank Prof. Dietmar Schomburg (TU Braunschweig) for the opportunity to perform GC-APCI-TOF-MS measurements in his laboratory and Dr. Christian Jäger for the realization of the actual measurements. For the financing of my PhD project as part of ATTRACT A10/03, I thank the Fonds National de la Recherche (FNR) Luxembourg.

Last but not least, I thank the whole metabolomics research group and other fellow LCSB colleagues for a very pleasant working environment and stimulating atmosphere over the last years.

# Appendices

# A Article manuscripts

## A.1

Original Article

---

Isotope cluster based compound matching in gas chromatography/mass spectrometry for non-targeted metabolomics

---

Wegner A, Sapcariu SC, Weindl D & Hiller K

2013

*Analytical Chemistry*, 85 (8), pp 4030–7

doi: 10.1021/ac303774z

### Contributions

- Preparation of Figure 1
- Critical revision of the manuscript

This manuscript has been omitted in the online version of this dissertation.

## A.2

Application Note

# NTFD — a stand-alone application for the non-targeted detection of stable isotope-labeled compounds in GC/MS data

Hiller K, Wegner A, Weindl D, Cordes T, Metallo CM, Kelleher JK & Stephanopoulos, G

2013

*Bioinformatics*, 29, pp 1226–8
doi: 10.1093/bioinformatics/btt119

### Contributions

- Preparation of documentation and tutorial
- Setting up website `http://ntfd.mit.edu/`
- Critical revision of the manuscript

This manuscript has been omitted in the online version of this dissertation.

# A.3

## Review Article

# Complexity of dopamine metabolism

Meiser J,[1] Weindl D[1] & Hiller K

2013

*Cell Communication and Signaling*, 11:34

doi: 10.1186/1478-811X-11-34

### Contributions

- Literature research & writing part of the manuscript
- Critical revision of the entire manuscript
- Figures as indicated in the manuscript

[1]Equally contributing authors

This manuscript has been omitted in the online version of this dissertation.

## A.4

Original Article

# Fragment Formula Calculator (FFC): Determination of chemical formulas for fragment ions in mass spectrometric data

Wegner A, Weindl D, Jäger C, Sapcariu SC, Dong X, Stephanopoulos G & Hiller K

2014

*Analytical Chemistry*, 86(4), 2221–8

doi: 10.1021/ac403879d

### Contributions

- Contributed to design of the algorithm
- Manuscript paragraph on fragmentation reactions & Figure 1
- Critical revision of the manuscript
- Setting up FFC website `http://ffc.lu/`

This manuscript has been omitted in the online version of this dissertation.

# A.5

Original Article

# Simultaneous extraction of proteins and metabolites from cells in culture

Sapcariu SC, Kanashova T, Weindl D, Ghelfi J, Dittmar G & Hiller K

2014

*MethodsX*, 1, pp 74–80

doi: 10.1016/j.mex.2014.07.002

**Contributions**

- Created figure 1
- Critical revision of the manuscript

This manuscript has been omitted in the online version of this dissertation.

## A.6 How metabolites modulate metabolic flux

Review Article

# How metabolites modulate metabolic flux

Wegner A, Meiser J, Weindl D, & Hiller K.

2015

*Current Opinion in Biotechnology*, 34, pp 16–22

doi: 10.1016/j.copbio.2014.11.008

### Contributions

- Contributed to literature research and design of the manuscript
- Critical revision of the manuscript

This manuscript has been omitted in the online version of this dissertation.

# A.7

## Original Article

# Isotopologue ratio normalization for non-targeted metabolomics

Weindl D, Wegner A, Jäger C, & Hiller K

*Journal of Chromatography A*, 1389, pp 112-9
doi: 10.1016/j.chroma.2015.02.025

### Contributions

- Conducting experiments
- Data analysis
- Computational implementation
- Writing and revising the initial manuscript and creating figures

This manuscript has been omitted in the online version of this dissertation.

## A.8

Application Note

# MIA: Non-targeted mass isotopolome analysis

Manuscript in Preparation

### Contributions

- Software design and implementation
- Documentation of the program
- Generation of sample data
- Writing and revising the initial manuscript and creating all figures

This manuscript has been omitted in the online version of this dissertation.

# A.9

## Article

---

---

Manuscript in Preparation

#### Contributions

- Data analysis
- Conducting some experiments
- Writing and revising the initial manuscript and creating the figures

This manuscript has been omitted in the online version of this dissertation.

## A.10

Book Chapter

# Non-targeted tracer fate detection

Weindl D, Wegner A & Hiller K

2015

Submitted to *Methods in Enzymology*

### Contributions

- Joint writing of the manuscript
- Preparing figures

This manuscript has been omitted in the online version of this dissertation.

# B Miscellaneous

## B.1 Chemstation macros

Post-processing macro for HP/Agilent ChemStation for the automated data export in the AIA/netCDF format and further conversion to the MetaboliteDetector format using an in-house tool.

```
! Convert .D to netCDF, correct date, export to MetaboliteDetector format
!
! 2013 Daniel Weindl <daniel.weindl@uni.lu>
!

Name dw_postprocess
  ! Convert .D to netCDF, correct date, export to MD format
  Parameter DPath$, DName$

  Local CDFFile$
  Local RawDataFullPath$
  Local MetaDataFullPath$
  Local CDFDir$ ! Where to place netCDF files
  Local MDDir$  ! Where to place MD files
  Local Sample$ ! Sample Name
  Local Cmd$    !
  Local DblQt$  ! Escapesequence workaround
  Local SubDir$ ! Subdirectory for netCDF/MD files e.g. 2012/01_Jan/...
  Local CDFFullPath$
  Local LogFile$

  Print "Handling " + DPath$ + DName$
  SubDir$ = DPath$[4:Len(DPath$)] ! strip "D:\"

  DblQt$ = Chr$(34)
  LogFile$ = "C:\gcmspreprocessing\log.txt"
  CDFDir$ = "D:\netCDF" + "\" + SubDir$
  MDDir$ = "D:\MetaboliteDetectorData" + "\" + SubDir$
```

```
  Sample$ = DName$[1:Len(DName$)-2]

! target directories must exist; must use mkdir wrapper for 0 exit code, macro
    otherwise aborted
  DOSRUN "C:\gcmspreprocessing\md.bat " + DblQt$ + CDFDir$ + DblQt$
  DOSRUN "C:\gcmspreprocessing\md.bat " + DblQt$ + MDDir$ + DblQt$

  RawDataFullPath$ = DPath$ + "\" + DName$ + "\data.ms"
  MetaDataFullPath$ = DPath$ + "\" + DName$ + "\acqmeth.txt"

  Local stat
  stat = FILESTAT(RawDataFullPath$) ! check if file exists; on error also
      calling macro will die

  If stat[7] = 0 Then ! not exists
        Print "Data file " + RawDataFullPath$ + " does not exist."
        Return
  EndIF

! convert to CDF
  CDFFullPath$ = CDFDir$ + "\" + Sample$ + ".cdf"

  Print "CDF..."
  HgXlate CDFFullPath$, RawDataFullPath$
  Print "CDF... done"

! invoke CDF date corrector and converter
  Cmd$ = "C:\gcmspreprocessing\gcmspreprocessing.bat -c " + DblQt$ +
      MetaDataFullPath$ + DblQt$ + " -o " + DblQt$ + MDDir$ + DblQt$ + " -l " +
       DblQt$ + LogFile$ + DblQt$ + " -a " + DblQt$ + MetaDataFullPath$ +
      DblQt$ + " " + DblQt$ + CDFFullPath$ + DblQt$

  Print Cmd$
  DOSRUN Cmd$

Return
```

# Bibliography


[1] Katada S, Imhof A, & Sassone-Corsi P. Connecting threads: epigenetics and metabolism. *Cell*, **2012**;148(1-2):24–28. doi: 10.1016/j.cell.2012.01.001.

[2] Wegner A, Meiser J, Weindl D, & Hiller K. How metabolites modulate metabolic flux. *Current Opinion in Biotechnology*, **2015**;34(0):16 – 22. doi: http://dx.doi.org/10.1016/j.copbio.2014.11.008.

[3] Sahoo S, Franzson L, Jonsson JJ, & Thiele I. A compendium of inborn errors of metabolism mapped onto the human metabolic network. *Mol Biosyst*, **2012**;8(10):2545–2558. doi: 10.1039/c2mb25075f.

[4] Mahuran DJ. Beta-hexosaminidase: biosynthesis and processing of the normal enzyme, and identification of mutations causing jewish tay-sachs disease. *Clin Biochem*, **1995**; 28(2):101–106.

[5] Matalon RM & Michals-Matalon K. Spongy degeneration of the brain, canavan disease: biochemical and molecular findings. *Front Biosci*, **2000**;5:D307–D311.

[6] Meiser J, Weindl D, & Hiller K. Complexity of dopamine metabolism. *Cell Commun Signal*, **2013**;11(1):34. doi: 10.1186/1478-811X-11-34.

[7] Gaglio D, Metallo CM, Gameiro PA, Hiller K, Danna LS, Balestrieri C, *et al.* Oncogenic k-ras decouples glucose and glutamine metabolism to support cancer cell growth. *Mol Syst Biol*, **2011**;7:523. doi: 10.1038/msb.2011.56.

[8] Metallo CM, Gameiro PA, Bell EL, Mattaini KR, Yang J, Hiller K, *et al.* Reductive glutamine metabolism by idh1 mediates lipogenesis under hypoxia. *Nature*, **2012**; 481(7381):380–384. doi: 10.1038/nature10602.

[9] Fan TWM, Lorkiewicz PK, Sellers K, Moseley HNB, Higashi RM, & Lane AN. Stable isotope-resolved metabolomics and applications for drug development. *Pharmacol Ther*, **2012**;133(3):366–391. doi: 10.1016/j.pharmthera.2011.12.007.

[10] Mayers JR, Wu C, Clish CB, Kraft P, Torrence ME, Fiske BP, *et al.* Elevation of circulating branched-chain amino acids is an early event in human pancreatic adenocarcinoma development. *Nat Med*, **2014**;20(10):1193–1198. doi: 10.1038/nm.3686.

[11] Mathé EA, Patterson AD, Haznadar M, Manna SK, Krausz KW, Bowman ED, *et al.* Noninvasive urinary metabolomic profiling identifies diagnostic and prognostic markers in lung cancer. *Cancer Res*, **2014**;74(12):3259–3270. doi: 10.1158/0008-5472.CAN-14-0109.

[12] Furuta E, Pai SK, Zhan R, Bandyopadhyay S, Watabe M, Mo YY, *et al.* Fatty acid synthase gene is up-regulated by hypoxia via activation of akt and sterol regulatory element binding protein-1. *Cancer Res*, **2008**;68(4):1003–1011. doi: 10.1158/0008-5472.CAN-07-2489.

[13] Menendez JA & Lupu R. Fatty acid synthase and the lipogenic phenotype in cancer pathogenesis. *Nat Rev Cancer*, **2007**;7(10):763–777. doi: 10.1038/nrc2222.

[14] Le A, Cooper CR, Gouw AM, Dinavahi R, Maitra A, Deck LM, *et al.* Inhibition of lactate dehydrogenase a induces oxidative stress and inhibits tumor progression. *Proc Natl Acad Sci U S A*, **2010**;107(5):2037–2042. doi: 10.1073/pnas.0914433107.

[15] Wilson WR & Hay MP. Targeting hypoxia in cancer therapy. *Nat Rev Cancer*, **2011**; 11(6):393–410. doi: 10.1038/nrc3064.

[16] Löffler G, Petrides PE, & Heinrich PC (eds.). *Biochemie und Pathobiochemie*. Springer Medizin Verlag Heidelberg, 8 edition, **2007**.

[17] Berg JM, Tymoczko JL, & Stryer L (eds.). *Biochemie*. Spektrum Akademischer Verlag, 6 edition, **2007**.

[18] Zecchini V, Madhu B, Russell R, Pértega-Gomes N, Warren A, Gaude E, *et al.* Nuclear arrb1 induces pseudohypoxia and cellular metabolism reprogramming in prostate cancer. *EMBO J*, **2014**;33(12):1365–1382. doi: 10.15252/embj.201386874.

[19] Vacanti NM, Divakaruni AS, Green CR, Parker SJ, Henry RR, Ciaraldi TP, *et al.* Regulation of substrate utilization by the mitochondrial pyruvate carrier. *Mol Cell*, **2014**;56(3):425–435. doi: 10.1016/j.molcel.2014.09.024.

[20] Nicolae A, Wahrheit J, Bahnemann J, Zeng AP, & Heinzle E. Non-stationary 13c metabolic flux analysis of chinese hamster ovary cells in batch culture using extracellular labeling highlights metabolic reversibility and compartmentation. *BMC Syst Biol*, **2014**; 8(1):50. doi: 10.1186/1752-0509-8-50.

[21] Weckwerth W. Metabolomics: an integral technique in systems biology. *Bioanalysis*, **2010**;2(4):829–836. doi: 10.4155/bio.09.192.

[22] Wei R. Metabolomics and its practical value in pharmaceutical industry. *Curr Drug Metab*, **2011**;12(4):345–358.

[23] Artati A, Prehn C, Möller G, & Adamski J. Assay tools for metabolomics. In K Suhre (ed.), *Genetics Meets Metabolomics*, 13–38. Springer New York, **2012**;doi: 10.1007/978-1-4614-1689-0_3.

[24] Drozd J. *Chemical derivatization in gas chromatography*, volume 19. Elsevier Scientific Publishing Company, **1981**.

[25] McLafferty FW & Tureček F. *Interpretation of mass spectra*. University Science Books, Sausalito, California, 4 edition, **1993**.

[26] Wegner A, Weindl D, Jäger C, Sapcariu SC, Dong X, Stephanopoulos G, *et al.* Fragment formula calculator (ffc): Determination of chemical formulas for fragment ions in mass spectrometric data. *Anal Chem*, **2014**;86(4):2221–2228. doi: 10.1021/ac403879d.

[27] Wiley. Wiley registry 10th edition. Software, **2013**.

[28] of Standards NI & Technology. Nist standard reference database 1a v14. Software, **2014**. URL http://www.nist.gov/srd/nist1a.cfm.

[29] Kind T, Wohlgemuth G, Lee DY, Lu Y, Palazoglu M, Shahbaz S, *et al.* Fiehnlib: mass spectral and retention index libraries for metabolomics based on quadrupole and time-of-flight gas chromatography/mass spectrometry. *Anal Chem*, **2009**;81(24):10038–10048. doi: 10.1021/ac9019522.

[30] Hummel J, Strehmel N, Bölling C, Schmidt S, Walther D, & Kopka J. *Mass Spectral Search and Analysis Using the Golm Metabolome Database*, chapter 18, 321–343. Wiley-VCH Verlag GmbH & Co. KGaA, **2013**;doi: 10.1002/9783527669882.ch18. URL http://dx.doi.org/10.1002/9783527669882.ch18.

[31] Stein S. An integrated method for spectrum extraction and compound identification from gas chromatography/mass spectrometry data. *Journal of the American Society for Mass Spectrometry*, **1999**;10(8):770 – 781. doi: 10.1016/S1044-0305(99)00047-1.

[32] Kováts E. Gas-chromatographische charakterisierung organischer verbindungen. teil 1: Retentionsindices aliphatischer halogenide, alkohole, aldehyde und ketone. *Helvetica Chimica Acta*, **1958**;41(7):1915–1932. doi: 10.1002/hlca.19580410703.

[33] Wegner A, Sapcariu SC, Weindl D, & Hiller K. Isotope cluster-based compound matching in gas chromatography/mass spectrometry for non-targeted metabolomics. *Anal Chem*, **2013**;85(8):4030–4037. doi: 10.1021/ac303774z.

[34] Jäger C. *Metabolomanalyse von Pseudomonas putida und Entwicklung eines Verfahrens zur Strukturanalyse unbekannter Verbindungen in bakteriellen Zellextrakten.* Ph.D. thesis, Technische Universität Carolo-Wilhelmina zu Braunschweig, **2012**.

[35] Kind T & Fiehn O. Seven golden rules for heuristic filtering of molecular formulas obtained by accurate mass spectrometry. *BMC Bioinformatics*, **2007**;8:105. doi: 10.1186/1471-2105-8-105.

[36] Williams DH. Mass spectrometry. In *Specialist Periodical Reports*, volume 1. The Chemical Society, Burlington House, London, **1971**;.

[37] Harvey DJ & Horning MG. Characterization of the trimethylsilyl derivatives of sugar phosphates and related compounds by gas chromatography and gas chromatography-mass spectrometry. *J Chromatogr*, **1973**;76(1):51–62.

[38] Miranda-Santos I, Gramacho S, Pineiro M, Martinez-Gomez K, Fritz M, Hollemeyer K, *et al.* Mass isotopomer analysis of nucleosides isolated from rna and dna using gc-ms. *Anal Chem*, **2014**;doi: 10.1021/ac503305w.

[39] Curtius HC, Müller M, & Völlmin J. Studies on the ring structures of ketoses by means of gas chromatography and mass spectroscopy. *Journal of Chromatography A*, **1968**; 37(0):216 – 224. doi: 10.1016/S0021-9673(01)99098-3.

[40] Karady S & Pines S. Mass spectrometry of the trimethylsilyl ethers of 2-ketohexoses. *Tetrahedron*, **1970**;26(19):4527 – 4536. doi: 10.1016/S0040-4020(01)93099-6.

[41] Urbach D. Mstfa und mstfa-d9 – unverzichtbare werkzeuge für die massenspektrometrische strukturaufklärung. *Toxichem Krimtech*, **2012**;79(3):137.

[42] Hübschmann HJ. *Handbook of GC/MS: Fundamentals and Applications*. WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim, 2 edition, **2009**.

[43] Dieterle F, Ross A, Schlotterbeck G, & Senn H. Probabilistic quotient normalization as robust method to account for dilution of complex biological mixtures. application in 1h nmr metabonomics. *Anal Chem*, **2006**;78(13):4281–4290. doi: 10.1021/ac051632c.

[44] Mashego MR, Wu L, Van Dam JC, Ras C, Vinke JL, Van Winden WA, *et al.* Miracle: mass isotopomer ratio analysis of u-13C-labeled extracts. a new method for accurate quantification of changes in concentrations of intracellular metabolites. *Biotechnol Bioeng*, **2004**;85(6):620–628. doi: 10.1002/bit.10907.

[45] Wu L, Mashego MR, van Dam JC, Proell AM, Vinke JL, Ras C, *et al.* Quantitative analysis of the microbial metabolome by isotope dilution mass spectrometry using uniformly 13C-labeled cell extracts as internal standards. *Analytical Biochemistry*, **2005**;336(2):164–171. doi: 10.1016/j.ab.2004.09.001.

[46] Bueschl C, Kluger B, Lemmens M, Adam G, Wiesenberger G, Maschietto V, *et al.* A novel stable isotope labelling assisted workflow for improved untargeted lc–hrms based metabolomics research. *Metabolomics*, **2013**;1–16. doi: 10.1007/s11306-013-0611-0.

[47] de Jong FA & Beecher C. Addressing the current bottlenecks of metabolomics: Isotopic ratio outlier analysis™, an isotopic-labeling technique for accurate biochemical profiling. *Bioanalysis*, **2012**;4(18):2303–2314. doi: 10.4155/bio.12.202.

[48] McHugh DMS, Cameron CA, Abdenur JE, Abdulrahman M, Adair O, Al Nuaimi SA, *et al.* Clinical validation of cutoff target ranges in newborn screening of metabolic disorders by tandem mass spectrometry: a worldwide collaborative project. *Genet Med*, **2011**;13(3):230–254. doi: 10.1097/GIM.0b013e31820d5e67.

[49] Schulze A, Lindner M, Kohlmüller D, Olgemöller K, Mayatepek E, & Hoffmann GF. Expanded newborn screening for inborn errors of metabolism by electrospray ionization-tandem mass spectrometry: results, outcome, and implications. *Pediatrics*, **2003**;111(6 Pt 1):1399–1406.

[50] Zhou R, Tseng CL, Huan T, & Li L. Isoms: automated processing of lc-ms data generated by a chemical isotope labeling metabolomics platform. *Anal Chem*, **2014**; 86(10):4675–4679. doi: 10.1021/ac5009089.

[51] Bouatra S, Aziat F, Mandal R, Guo AC, Wilson MR, Knox C, *et al.* The human urine metabolome. *PLoS One*, **2013**;8(9):e73076. doi: 10.1371/journal.pone.0073076.

[52] Psychogios N, Hau DD, Peng J, Guo AC, Mandal R, Bouatra S, *et al.* The human serum metabolome. *PLoS One*, **2011**;6(2):e16957. doi: 10.1371/journal.pone.0016957.

[53] Sévin DC, Kuehne A, Zamboni N, & Sauer U. Biological insights through non-targeted metabolomics. *Current Opinion in Biotechnology*, **2015**;34(0):1 – 8. doi: 10.1016/j.copbio.2014.10.001.

[54] Wishart DS. Advances in metabolite identification. *Bioanalysis*, **2011**;3(15):1769–1782. doi: 10.4155/bio.11.155.

[55] IUPAC. *Gold Book—Compendium of Chemical Terminology*. International Union of Pure and Applied Chemistry, **2012**. Version 2.3.2 2012-08-19.

[56] Budzikiewicz H & Grigsby RD. Mass spectrometry and isotopes: a century of research and discussion. *Mass Spectrom Rev*, **2006**;25(1):146–157. doi: 10.1002/mas.20061.

[57] Berglund M & Wieser ME. Isotopic compositions of the elements 2009. *Pure Appl Chern*, **2011**;83(2):397–410.

[58] Truong QX, Yoon JM, & Shanks JV. Isotopomer measurement techniques in metabolic flux analysis i: nuclear magnetic resonance. *Methods Mol Biol*, **2014**;1083:65–83. doi: 10.1007/978-1-62703-661-0_6.

[59] Young JD. Inca: a computational platform for isotopically non-stationary metabolic flux analysis. *Bioinformatics*, **2014**;30(9):1333–1335. doi: 10.1093/bioinformatics/btu015.

[60] Schoenheimer R & Rittenberg D. Deuterium as an indicator in the study of intermediary metabolism. i. *Journal of Biological Chemistry*, **1935**;111(1):163–168.

[61] Gregg CT, Hutson JY, Prine JR, Ott DG, & Furchner JE. Substantial replacement of mammalian body carbon with carbon-13. *Life Sci*, **1973**;13(7):775–782.

[62] Krumbiegel P. Large deuterium isotope effects and their use: a historical review. *Isotopes Environ Health Stud*, **2011**;47(1):1–17. doi: 10.1080/10256016.2011.556725.

[63] Yang S, Matsen JB, Konopka M, Green-Saxena A, Clubb J, Sadilek M, *et al.* Global molecular analyses of methane metabolism in methanotrophic alphaproteobacterium, methylosinus trichosporium ob3b. part ii. metabolomics and 13c-labeling study. *Front Microbiol*, **2013**;4:70. doi: 10.3389/fmicb.2013.00070.

[64] Bodamer OA & Halliday D. Uses of stable isotopes in clinical diagnosis and research in the paediatric population. *Arch Dis Child*, **2001**;84(5):444–448.

[65] Martineau A, Lecavalier L, Falardeau P, & Chiasson JL. Simultaneous determination of glucose turnover, alanine turnover, and gluconeogenesis in human using a double stable-isotope-labeled tracer infusion and gas chromatography-mass spectrometry analysis. *Anal Biochem*, **1985**;151(2):495–503.

[66] Landau BR. Quantifying the contribution of gluconeogenesis to glucose production in fasted human subjects using stable isotopes. *Proc Nutr Soc*, **1999**;58(4):963–972.

[67] Brunengraber H, Kelleher JK, & Rosiers CD. Applications of mass isotopomer analysis to nutrition research. *Annu Rev Nutr*, **1997**;17:559–596. doi: 10.1146/annurev.nutr.17.1.559.

[68] Hiller K, Metallo C, & Stephanopoulos G. Elucidation of cellular metabolism via metabolomics and stable-isotope assisted metabolomics. *Curr Pharm Biotechnol*, **2011**; 12(7):1075–1086.

[69] Ogra Y, Kitaguchi T, Suzuki N, & Suzuki KT. In vitro translation with [34s]-labeled methionine, selenomethionine, and telluromethionine. *Anal Bioanal Chem*, **2008**; 390(1):45–51. doi: 10.1007/s00216-007-1546-y.

[70] Wasylenko TM & Stephanopoulos G. Kinetic isotope effects significantly influence intracellular metabolite (13) c labeling patterns and flux determination. *Biotechnol J*, **2013**;doi: 10.1002/biot.201200276.

[71] Nadalig T, Greule M, Bringel F, Vuilleumier S, & Keppler F. Hydrogen and carbon isotope fractionation during degradation of chloromethane by methylotrophic bacteria. *Microbiologyopen*, **2013**;doi: 10.1002/mbo3.124.

[72] del Rio CM, Wolf N, Carleton SA, & Gannes LZ. Isotopic ecology ten years after a call for more laboratory experiments. *Biol Rev Camb Philos Soc*, **2009**;84(1):91–111. doi: 10.1111/j.1469-185X.2008.00064.x.

[73] Thompson DR, Bury SJ, Hobson KA, Wassenaar LI, & Shannon JP. Stable isotopes in ecological studies. *Oecologia*, **2005**;144(4):517–519. doi: 10.1007/s00442-005-0171-8.

[74] Shchepinov MS. Do "heavy" eaters live longer? *Bioessays*, **2007**;29(12):1247–1256. doi: 10.1002/bies.20681.

[75] Farquhar GD, Ehleringer JR, & Hubick KT. Carbon isotope discrimination and photosynthesis. *Annual Review of Plant Physiology and Plant Molecular Biology*, **1989**; 40(1):503–537. doi: 10.1146/annurev.pp.40.060189.002443.

[76] Zhang R, Sioma CS, Thompson RA, Xiong L, & Regnier FE. Controlling deuterium isotope effects in comparative proteomics. *Anal Chem*, **2002**;74(15):3662–3669.

[77] Turowski M, Yamakawa N, Meller J, Kimata K, Ikegami T, Hosoya K, *et al.* Deuterium isotope effects on hydrophobic interactions: the importance of dispersion interactions in the hydrophobic phase. *J Am Chem Soc*, **2003**;125(45):13836–13849. doi: 10.1021/ja036006g.

[78] Wu G, Lü JM, van der Donk WA, Kulmacz RJ, & Tsai Al. Cyclooxygenase reaction mechanism of prostaglandin h synthase from deuterium kinetic isotope effects. *J Inorg Biochem*, **2011**;105(3):382–390. doi: 10.1016/j.jinorgbio.2010.11.015.

[79] Cleland WW. The use of isotope effects to determine enzyme mechanisms. *Arch Biochem Biophys*, **2005**;433(1):2–12. doi: 10.1016/j.abb.2004.08.027.

[80] Wittmann C. Fluxome analysis using gc-ms. *Microb Cell Fact*, **2007**;6:6. doi: 10.1186/1475-2859-6-6.

[81] Hellerstein MK & Neese RA. Mass isotopomer distribution analysis at eight years: theoretical, analytic, and experimental considerations. *Am J Physiol*, **1999**;276(6 Pt 1):E1146–E1170.

[82] Wolfe RR & Chinkes DL. *Isotope Tracers in Metabolic Research: Principles and Practice of Kinetic Analysis*. John Wiley & Sons, 2nd edition edition, **2005**.

[83] Papageorgopoulos C, Caldwell K, Shackleton C, Schweingrubber H, & Hellerstein MK. Measuring protein synthesis by mass isotopomer distribution analysis (mida). *Anal Biochem*, **1999**;267(1):1–16. doi: 10.1006/abio.1998.2958.

[84] Bian F, Kasumov T, Jobbins KA, Minkler PE, Anderson VE, Kerner J, *et al.* Competition between acetate and oleate for the formation of malonyl-coa and mitochondrial acetyl-coa in the perfused rat heart. *J Mol Cell Cardiol*, **2006**;41(5):868–875. doi: 10.1016/j.yjmcc.2006.08.011.

[85] Murray KK, Boyd RK, Eberlin MN, Langley GJ, Li L, & Naito Y. Definitions of terms relating to mass spectrometry (iupac recommendations 2013). *Pure and Applied Chemistry*, **2013**;86(7):1515–1609. doi: 10.1351/PAC-REC-06-04-06.

[86] Muller P. Glossary of terms used in physical organic chemistry (iupac recommendations 1994). *Pure Appl Chern*, **1994**;66(5):1077–1184.

[87] Moss GP. Basic terminology of stereochemistry (iupac recommendations 1996). *Pure Appl Chern*, **1996**;68(12):2193–2222.

[88] Sun T, Hayakawa K, Bateman KS, & Fraser ME. Identification of the citrate-binding site of human atp-citrate lyase using x-ray crystallography. *J Biol Chem*, **2010**; 285(35):27418–27428. doi: 10.1074/jbc.M109.078667.

[89] Antoniewicz MR, Kelleher JK, & Stephanopoulos333 G. Elementary metabolite units (emu): a novel framework for modeling isotopic distributions. *Metab Eng*, **2007**; 9(1):68–86. doi: 10.1016/j.ymben.2006.09.001.

[90] Choi J, Grossbach MT, & Antoniewicz MR. Measuring complete isotopomer distribution of aspartate using gas chromatography/tandem mass spectrometry. *Anal Chem*, **2012**; 84(10):4628–4632. doi: 10.1021/ac300611n.

[91] Fernandez CA, Des Rosiers C, Previs SF, David F, & Brunengraber H. Correction of 13c mass isotopomer distributions for natural stable isotope abundance. *J Mass Spectrom*, **1996**;31(3):255–262.

[92] Pickup JF & McPherson K. Theoretical considerations in stable isotope dilution mass spectrometry for organic analysis. *Analytical Chemistry*, **1976**;48(13):1885–1890. doi: 10.1021/ac50007a019.

[93] Valkenborg D, Mertens I, Lemière F, Witters E, & Burzykowski T. The isotopic distribution conundrum. *Mass Spectrom Rev*, **2012**;31(1):96–109. doi: 10.1002/mas.20339.

[94] Gross J. Isotopic composition and accurate mass. In *Mass Spectrometry*, 67–116. Springer Berlin Heidelberg, **2011**;doi: 10.1007/978-3-642-10711-5_3.

[95] stats version 302 RDP. The multinomial distribution. Electronic, **2014**.

[96] Jennings ME & Matthews DE. Determination of complex isotopomer patterns in isotopically labeled compounds by mass spectrometry. *Anal Chem*, **2005**;77(19):6435–6444. doi: 10.1021/ac0509354.

[97] Christensen B & Nielsen J. Isotopomer analysis using gc-ms. *Metab Eng*, **1999**; 1(4):282–290. doi: 10.1006/mben.1999.0117.

[98] Antoniewicz MR, Kelleher JK, & Stephanopoulos G. Measuring deuterium enrichment of glucose hydrogen atoms by gas chromatography/mass spectrometry. *Anal Chem*, **2011**;83(8):3211–3216. doi: 10.1021/ac200012p.

[99] Di Donato L, Des Rosiers C, Montgomery JA, David F, Garneau M, & Brunengraber H. Rates of gluconeogenesis and citric acid cycle in perfused livers, assessed from the mass spectrometric assay of the 13c labeling pattern of glutamate. *J Biol Chem*, **1993**; 268(6):4170–4180.

[100] Tepper N & Shlomi T. An integrated computational approach for metabolic flux analysis coupled with inference of tandem-ms collisional fragments. *Bioinformatics*, **2013**;29(23):3045–3052. doi: 10.1093/bioinformatics/btt516.

[101] Yang S, Hoggard JC, Lidstrom ME, & Synovec RE. Comprehensive discovery of 13c labeled metabolites in the bacterium methylobacterium extorquens am1 using gas chromatography-mass spectrometry. *J Chromatogr A*, **2013**;1317:175–185. doi: 10.1016/j.chroma.2013.08.059.

[102] Hiller K, Metallo CM, Kelleher JK, & Stephanopoulos G. Nontargeted elucidation of metabolic pathways using stable-isotope tracers and mass spectrometry. *Anal Chem*, **2010**;82(15):6621–6628. doi: 10.1021/ac1011574.

[103] Hiller K, Wegner A, Weindl D, Cordes T, Metallo CM, Kelleher JK, *et al.* Ntfd–a standalone application for the non-targeted detection of stable isotope-labeled compounds in gc/ms data. *Bioinformatics*, **2013**;29(9):1226–1228. doi: 10.1093/bioinformatics/btt119.

[104] Huang X, Chen YJ, Cho K, Nikolskiy I, Crawford PA, & Patti GJ. X13cms: global tracking of isotopic labels in untargeted metabolomics. *Anal Chem*, **2014**;86(3):1632–1639. doi: 10.1021/ac403384n.

[105] Creek DJ, Chokkathukalam A, Jankevics A, Burgess KEV, Breitling R, & Barrett MP. Stable isotope-assisted metabolomics for network-wide metabolic pathway elucidation. *Anal Chem*, **2012**;84(20):8442–8447. doi: 10.1021/ac3018795.

[106] Chokkathukalam A, Jankevics A, Creek DJ, Achcar F, Barrett MP, & Breitling R. mzmatch-iso: an r tool for the annotation and relative quantification of isotope-labelled mass spectrometry data. *Bioinformatics*, **2013**;29(2):281–283. doi: 10.1093/bioinformatics/bts674.

[107] Li J, Hoene M, Zhao X, Chen S, Wei H, Häring HU, *et al.* Stable isotope-assisted lipidomics combined with nontargeted isotopomer filtering, a tool to unravel the complex dynamics of lipid metabolism. *Anal Chem*, **2013**;85(9):4651–4657. doi: 10.1021/ac400293y.

[108] Nakayama Y, Tamada Y, Tsugawa H, Bamba T, & Fukusaki E. Novel strategy for non-targeted isotope-assisted metabolomics by means of metabolic turnover and multivariate analysis. *Metabolites*, **2014**;4(3):722–739. doi: 10.3390/metabo4030722.

[109] Chokkathukalam A, Kim DH, Barrett MP, Breitling R, & Creek DJ. Stable isotope-labeling studies in metabolomics: new insights into structure and dynamics of metabolic networks. *Bioanalysis*, **2014**;6(4):511–524. doi: 10.4155/bio.13.348.

[110] Sauer U. Metabolic networks in motion: 13c-based flux analysis. *Mol Syst Biol*, **2006**; 2:62. doi: 10.1038/msb4100109.

[111] Jang M, Kim SS, & Lee J. Cancer cell metabolism: implications for therapeutic targets. *Exp Mol Med*, **2013**;45:e45. doi: 10.1038/emm.2013.85.

[112] Niklas J, Noor F, & Heinzle E. Effects of drugs in subtoxic concentrations on the metabolic fluxes in human hepatoma cell line hep g2. *Toxicol Appl Pharmacol*, **2009**; 240(3):327–336. doi: 10.1016/j.taap.2009.07.005.

[113] Lewis NE, Nagarajan H, & Palsson BO. Constraining the metabolic genotype-phenotype relationship using a phylogeny of in silico methods. *Nat Rev Microbiol*, **2012**;10(4):291–305. doi: 10.1038/nrmicro2737.

[114] Thiele I, Swainston N, Fleming RMT, Hoppe A, Sahoo S, Aurich MK, *et al.* A community-driven global reconstruction of human metabolism. *Nat Biotechnol*, **2013**; 31(5):419–425. doi: 10.1038/nbt.2488.

[115] Wiechert W. 13c metabolic flux analysis. *Metab Eng*, **2001**;3(3):195–206. doi: 10.1006/mben.2001.0187.

[116] Rodrigues AL, Becker J, de Souza Lima AO, Porto LM, & Wittmann C. Systems metabolic engineering of escherichia coli for gram scale production of the antitumor drug deoxyviolacein from glycerol. *Biotechnol Bioeng*, **2014**;111(11):2280–2289. doi: 10.1002/bit.25297.

[117] Kruger NJ, Masakapalli SK, & Ratcliffe RG. Strategies for investigating the plant metabolic network with steady-state metabolic flux analysis: lessons from an arabidopsis cell culture and other systems. *J Exp Bot*, **2012**;63(6):2309–2323. doi: 10.1093/jxb/err382.

[118] Niedenführ S, Wiechert W, & Nöh K. How to measure metabolic fluxes: a taxonomic guide for 13c fluxomics. *Current Opinion in Biotechnology*, **2015**;34(0):82 – 90. doi: 10.1016/j.copbio.2014.12.003.

[119] Wiechert W & de Graaf AA. Bidirectional reaction steps in metabolic networks: I. modeling and simulation of carbon isotope labeling experiments. *Biotechnol Bioeng*, **1997**;55(1):101–117.

[120] Wiechert W, Möllney M, Isermann N, Wurzel M, & de Graaf AA. Bidirectional reaction steps in metabolic networks: Iii. explicit solution and analysis of isotopomer labeling systems. *Biotechnol Bioeng*, **1999**;66(2):69–85.

[121] Srour O, Young JD, & Eldar YC. Fluxomers: a new approach for 13c metabolic flux analysis. *BMC Syst Biol*, **2011**;5:129. doi: 10.1186/1752-0509-5-129.

[122] Quek LE, Wittmann C, Nielsen LK, & Krömer JO. Openflux: efficient modelling software for 13c-based metabolic flux analysis. *Microb Cell Fact*, **2009**;8:25. doi: 10.1186/1475-2859-8-25.

[123] Wiechert W, Möllney M, Petersen S, & de Graaf AA. A universal framework for 13c metabolic flux analysis. *Metab Eng*, **2001**;3(3):265–283. doi: 10.1006/mben.2001.0188.

[124] Weitzel M, Nöh K, Dalman T, Niedenführ S, Stute B, & Wiechert W. 13cflux2–high-performance software suite for (13)c-metabolic flux analysis. *Bioinformatics*, **2013**; 29(1):143–145. doi: 10.1093/bioinformatics/bts646.

[125] Crown SB & Antoniewicz MR. Selection of tracers for 13c-metabolic flux analysis using elementary metabolite units (emu) basis vector methodology. *Metab Eng*, **2012**; 14(2):150–161. doi: 10.1016/j.ymben.2011.12.005.

[126] Zamboni N. 13c metabolic flux analysis in complex systems. *Curr Opin Biotechnol*, **2011**;22(1):103–108. doi: 10.1016/j.copbio.2010.08.009.

[127] Ferlay J, Soerjomataram I, Ervik M, Dikshit R, Eser S, Mathers C, *et al.* Globocan 2012 v1.0, cancer incidence and mortality worldwide: Iarc cancerbase no. 11, lyon, france: International agency for research on cancer. Internet, **2013**. URL `http://globocan.iarc.fr`, accessed 17.12.2014.

[128] Edwards BK, Noone AM, Mariotto AB, Simard EP, Boscoe FP, Henley SJ, *et al.* Annual report to the nation on the status of cancer, 1975-2010, featuring prevalence of comorbidity and impact on survival among persons with lung, colorectal, breast, or prostate cancer. *Cancer*, **2014**;120(9):1290–1314. doi: 10.1002/cncr.28509.

[129] Warburg O, Posener K, & Negelein E. Über den stoffwechsel der karzinomzelle. *Biochem Zeitschr*, **1924**;152:309–44.

[130] Koppenol WH, Bounds PL, & Dang CV. Otto warburg's contributions to current concepts of cancer metabolism. *Nat Rev Cancer*, **2011**;11(5):325–337. doi: 10.1038/nrc3038.

[131] Warburg O. On the origin of cancer cells. *Science*, **1956**;123(3191):309–314.

[132] McKeehan WL. Glycolysis, glutaminolysis and cell proliferation. *Cell Biol Int Rep*, **1982**;6(7):635–650.

[133] Wise DR, Ward PS, Shay JES, Cross JR, Gruber JJ, Sachdeva UM, *et al.* Hypoxia promotes isocitrate dehydrogenase-dependent carboxylation of $\alpha$-ketoglutarate to citrate to support cell growth and viability. *Proc Natl Acad Sci U S A*, **2011**;108(49):19611–19616. doi: 10.1073/pnas.1117773108.

[134] Baenke F, Peck B, Miess H, & Schulze A. Hooked on fat: the role of lipid synthesis in cancer metabolism and tumour development. *Dis Model Mech*, **2013**;6(6):1353–1363. doi: 10.1242/dmm.011338.

[135] Balkwill FR, Capasso M, & Hagemann T. The tumor microenvironment at a glance. *J Cell Sci*, **2012**;125(Pt 23):5591–5596. doi: 10.1242/jcs.116392.

[136] McKeown SR. Defining normoxia, physoxia and hypoxia in tumours-implications for treatment response. *Br J Radiol*, **2014**;87(1035):20130676. doi: 10.1259/bjr.20130676.

[137] Ackerman D & Simon MC. Hypoxia, lipids, and cancer: surviving the harsh tumor microenvironment. *Trends Cell Biol*, **2014**;24(8):472–478. doi: 10.1016/j.tcb.2014.06.001.

[138] Vander Heiden MG, Cantley LC, & Thompson CB. Understanding the warburg effect: the metabolic requirements of cell proliferation. *Science*, **2009**;324(5930):1029–1033. doi: 10.1126/science.1160809.

[139] Koppenol WH & Bounds PL. The warburg effect and metabolic efficiency: Re-crunching the numbers. Science Magazine E-Letters, **14 September 2009**. URL `http://www.sciencemag.org/content/324/5930/1029/reply#sci_el_12397`.

[140] Semenza GL. Oxygen-dependent regulation of mitochondrial respiration by hypoxia-inducible factor 1. *Biochem J*, **2007**;405(1):1–9. doi: 10.1042/BJ20070389.

[141] Kim Jw, Tchernyshyov I, Semenza GL, & Dang CV. Hif-1-mediated expression of pyruvate dehydrogenase kinase: a metabolic switch required for cellular adaptation to hypoxia. *Cell Metab*, **2006**;3(3):177–185. doi: 10.1016/j.cmet.2006.02.002.

[142] Sedlakova O, Svastova E, Takacova M, Kopacek J, Pastorek J, & Pastorekova S. Carbonic anhydrase ix, a hypoxia-induced catalytic component of the ph regulating machinery in tumors. *Front Physiol*, **2014**;4:400. doi: 10.3389/fphys.2013.00400.

[143] Schulze A & Harris AL. How cancer metabolism is tuned for proliferation and vulnerable to disruption. *Nature*, **2012**;491(7424):364–373. doi: 10.1038/nature11706.

[144] Semenza GL. Hif-1: upstream and downstream of cancer metabolism. *Curr Opin Genet Dev*, **2010**;20(1):51–56. doi: 10.1016/j.gde.2009.10.009.

[145] Semenza GL. Hif-1 mediates metabolic responses to intratumoral hypoxia and oncogenic mutations. *J Clin Invest*, **2013**;123(9):3664–3671. doi: 10.1172/JCI67230.

[146] Amoêdo ND, Valencia JP, Rodrigues MF, Galina A, & Rumjanek FD. How does the metabolism of tumour cells differ from that of normal cells. *Biosci Rep*, **2013**;33(6). doi: 10.1042/BSR20130066.

[147] Henze AT, Garvalov BK, Seidel S, Cuesta AM, Ritter M, Filatova A, *et al.* Loss of phd3 allows tumours to overcome hypoxic growth inhibition and sustain proliferation through egfr. *Nat Commun*, **2014**;5:5582. doi: 10.1038/ncomms6582.

[148] Wang GL, Jiang BH, Rue EA, & Semenza GL. Hypoxia-inducible factor 1 is a basic-helix-loop-helix-pas heterodimer regulated by cellular o2 tension. *Proc Natl Acad Sci U S A*, **1995**;92(12):5510–5514.

[149] Jiang BH, Semenza GL, Bauer C, & Marti HH. Hypoxia-inducible factor 1 levels vary exponentially over a physiologically relevant range of o2 tension. *Am J Physiol*, **1996**; 271(4 Pt 1):C1172–C1180.

[150] Epstein AC, Gleadle JM, McNeill LA, Hewitson KS, O'Rourke J, Mole DR, *et al.* C. elegans egl-9 and mammalian homologs define a family of dioxygenases that regulate hif by prolyl hydroxylation. *Cell*, **2001**;107(1):43–54.

[151] Semenza GL. Hif-1, o(2), and the 3 phds: how animal cells signal hypoxia to the nucleus. *Cell*, **2001**;107(1):1–3.

[152] Kim WY & Kaelin WG. Role of vhl gene mutation in human cancer. *J Clin Oncol*, **2004**;22(24):4991–5004. doi: 10.1200/JCO.2004.05.061.

[153] Tong X, Zhao F, & Thompson CB. The molecular determinants of de novo nucleotide biosynthesis in cancer cells. *Curr Opin Genet Dev*, **2009**;19(1):32–37. doi: 10.1016/j.gde.2009.01.002.

[154] Patra KC & Hay N. The pentose phosphate pathway and cancer. *Trends Biochem Sci*, **2014**;39(8):347–354. doi: 10.1016/j.tibs.2014.06.005.

[155] Tibbetts AS & Appling DR. Compartmentalization of mammalian folate-mediated one-carbon metabolism. *Annu Rev Nutr*, **2010**;30:57–81. doi: 10.1146/annurev.nutr.012809.104810.

[156] Tedeschi PM, Markert EK, Gounder M, Lin H, Dvorzhinski D, Dolfi SC, *et al.* Contribution of serine, folate and glycine metabolism to the atp, nadph and purine requirements of cancer cells. *Cell Death Dis*, **2013**;4:e877. doi: 10.1038/cddis.2013.393.

[157] Deghan Manshadi S, Ishiguro L, Sohn KJ, Medline A, Renlund R, Croxford R, *et al.* Folic acid supplementation promotes mammary tumor progression in a rat model. *PLoS One*, **2014**;9(1):e84635. doi: 10.1371/journal.pone.0084635.

[158] Mehrmohamadi M, Liu X, Shestov AA, & Locasale JW. Characterization of the usage of the serine metabolic network in human cancer. *Cell Rep*, **2014**;9(4):1507–1519. doi: 10.1016/j.celrep.2014.10.026.

[159] Gameiro PA, Laviolette LA, Kelleher JK, Iliopoulos O, & Stephanopoulos G. Cofactor balance by nicotinamide nucleotide transhydrogenase (nnt) coordinates reductive carboxylation and glucose catabolism in the tricarboxylic acid (tca) cycle. *J Biol Chem*, **2013**;288(18):12967–12977. doi: 10.1074/jbc.M112.396796.

[160] Mashimo T, Pichumani K, Vemireddy V, Hatanpaa KJ, Singh DK, Sirasanagandla S, *et al.* Acetate is a bioenergetic substrate for human glioblastoma and brain metastases. *Cell*, **2014**;159(7):1603–1614. doi: 10.1016/j.cell.2014.11.025.

[161] Kamphorst J, Chung M, Fan J, & Rabinowitz J. Quantitative analysis of acetyl-coa production in hypoxic cancer cells reveals substantial contribution from acetate. *Cancer & Metabolism*, **2014**;2(1):23. doi: 10.1186/2049-3002-2-23.

[162] Schug ZT, Peck B, Jones DT, Zhang Q, Grosskurth S, Alam IS, *et al.* Acetyl-coa synthetase 2 promotes acetate utilization and maintains cancer cell growth under metabolic stress. *Cancer Cell*, **2015**;27(1):57–71. doi: 10.1016/j.ccell.2014.12.002.

[163] Sun RC & Denko NC. Hypoxic regulation of glutamine metabolism through hif1 and siah2 supports lipid synthesis that is necessary for tumor growth. *Cell Metab*, **2014**; 19(2):285–292. doi: 10.1016/j.cmet.2013.11.022.

[164] Fendt SM, Bell EL, Keibler MA, Olenchock BA, Mayers JR, Wasylenko TM, *et al.* Reductive glutamine metabolism is a function of the $\alpha$-ketoglutarate to citrate ratio in cells. *Nat Commun*, **2013**;4:2236. doi: 10.1038/ncomms3236.

[165] Zaidi N, Swinnen JV, & Smans K. Atp-citrate lyase: a key player in cancer metabolism. *Cancer Res*, **2012**;72(15):3709–3714. doi: 10.1158/0008-5472.CAN-11-4112.

[166] Zaidi N, Lupien L, Kuemmerle NB, Kinlaw WB, Swinnen JV, & Smans K. Lipogenesis and lipolysis: the pathways exploited by the cancer cells to acquire fatty acids. *Prog Lipid Res*, **2013**;52(4):585–589. doi: 10.1016/j.plipres.2013.08.005.

[167] Kamphorst JJ, Cross JR, Fan J, de Stanchina E, Mathew R, White EP, *et al.* Hypoxic and ras-transformed cells support growth by scavenging unsaturated fatty acids from lysophospholipids. *Proc Natl Acad Sci U S A*, **2013**;110(22):8882–8887. doi: 10.1073/pnas.1307237110.

[168] Smith S. The animal fatty acid synthase: one gene, one polypeptide, seven enzymes. *FASEB J*, **1994**;8(15):1248–1259.

[169] Guillou H, Zadravec D, Martin PGP, & Jacobsson A. The key roles of elongases and desaturases in mammalian fatty acid metabolism: Insights from transgenic mice. *Prog Lipid Res*, **2010**;49(2):186–199. doi: 10.1016/j.plipres.2009.12.002.

[170] Peng G, Li L, Liu Y, Pu J, Zhang S, Yu J, *et al.* Oleate blocks palmitate-induced abnormal lipid distribution, endoplasmic reticulum expansion and stress, and insulin resistance in skeletal muscle. *Endocrinology*, **2011**;152(6):2206–2218. doi: 10.1210/en.2010-1369.

[171] Noto A, Raffa S, De Vitis C, Roscilli G, Malpicci D, Coluccia P, *et al.* Stearoyl-coa desaturase-1 is a key factor for lung cancer-initiating cells. *Cell Death Dis*, **2013**;4:e947. doi: 10.1038/cddis.2013.444.

[172] Baracos VE & Mackenzie ML. Investigations of branched-chain amino acids and their metabolites in animal models of cancer. *J Nutr*, **2006**;136(1 Suppl):237S–242S.

[173] Argilés JM & López-Soriano FJ. The oxidation of leucine in tumour-bearing rats. *Biochem J*, **1990**;268(1):241–244.

[174] Tönjes M, Barbus S, Park YJ, Wang W, Schlotter M, Lindroth AM, *et al.* Bcat1 promotes cell proliferation through amino acid catabolism in gliomas carrying wild-type idh1. *Nat Med*, **2013**;19(7):901–908. doi: 10.1038/nm.3217.

[175] Han JM, Jeong SJ, Park MC, Kim G, Kwon NH, Kim HK, *et al.* Leucyl-trna synthetase is an intracellular leucine sensor for the mtorc1-signaling pathway. *Cell*, **2012**;149(2):410–424. doi: 10.1016/j.cell.2012.02.044.

[176] Dibble CC & Manning BD. Signal integration by mtorc1 coordinates nutrient input with biosynthetic output. *Nat Cell Biol*, **2013**;15(6):555–564. doi: 10.1038/ncb2763.

[177] Galluzzi L, Kepp O, Vander Heiden MG, & Kroemer G. Metabolic targets for cancer therapy. *Nat Rev Drug Discov*, **2013**;12(11):829–846. doi: 10.1038/nrd4145.

[178] Matsuishi T, Stumpf DA, Seliem M, Eguren LA, & Chrislip K. Propionate mitochondrial toxicity in liver and skeletal muscle: acyl coa levels. *Biochem Med Metab Biol*, **1991**; 45(2):244–253.

[179] Shimomura Y, Honda T, Shiraki M, Murakami T, Sato J, Kobayashi H, *et al.* Branched-chain amino acid catabolism in exercise and liver disease. *J Nutr*, **2006**;136(1 Suppl):250S–253S.

[180] Brosnan JT & Brosnan ME. Branched-chain amino acids: enzyme and substrate regulation. *J Nutr*, **2006**;136(1 Suppl):207S–211S.

[181] Duckwall CS, Murphy TA, & Young JD. Mapping cancer cell metabolism with(13)c flux analysis: Recent progress and future challenges. *J Carcinog*, **2013**;12:13. doi: 10.4103/1477-3163.115422.

[182] Cunningham JT, Moreno MV, Lodi A, Ronen SM, & Ruggero D. Protein and nucleotide biosynthesis are coupled by a single rate-limiting enzyme, prps2, to drive cancer. *Cell*, **2014**;157(5):1088–1103. doi: 10.1016/j.cell.2014.03.052.

[183] Hanahan D & Weinberg R. Hallmarks of cancer: The next generation. *Cell*, **2011**; 144(5):646 – 674. doi: 10.1016/j.cell.2011.02.013.

[184] Wang T, Niki T, Goto A, Ota S, Morikawa T, Nakamura Y, *et al.* Hypoxia increases the motility of lung adenocarcinoma cell line a549 via activation of the epidermal growth factor receptor pathway. *Cancer Sci*, **2007**;98(4):506–511. doi: 10.1111/j.1349-7006.2007.00428.x.

[185] Huang CH, Yang WH, Chang SY, Tai SK, Tzeng CH, Kao JY, *et al.* Regulation of membrane-type 4 matrix metalloproteinase by slug contributes to hypoxia-mediated metastasis. *Neoplasia*, **2009**;11(12):1371–1382.

[186] Cordes T. *Stable Isotope-assisted Metabolomics To Profile Mammalian Metabolism Of Inflammation And Cancer*. Ph.D. thesis, University of Luxembourg, **2014**.

[187] Sapcariu SC, Kanashova T, Weindl D, Ghelfi J, Dittmar G, & Hiller K. Simultaneous extraction of proteins and metabolites from cells in culture. *MethodsX*, **2014**;1(0):74 – 80. doi: http://dx.doi.org/10.1016/j.mex.2014.07.002.

[188] Pessentheiner AR, Pelzmann HJ, Walenta E, Schweiger M, Groschner LN, Graier WF, *et al.* Nat8l (n-acetyltransferase 8-like) accelerates lipid turnover and increases energy expenditure in brown adipocytes. *J Biol Chem*, **2013**;288(50):36040–36051. doi: 10.1074/jbc.M113.491324.

[189] Hiller K, Hangebrauk J, Jäger C, Spura J, Schreiber K, & Schomburg D. Metabolitedetector: comprehensive analysis tool for targeted and nontargeted gc/ms based metabolome analysis. *Anal Chem*, **2009**;81(9):3429–3439. doi: 10.1021/ac802689c.

[190] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, **2013**. URL http://www.R-project.org/.

[191] Weindl D, Wegner A, Jäger C, & Hiller K. Isotopologue ratio normalization for non-targeted metabolomics. *J Chromatogr A*, **2015**;1389:112–119. doi: 10.1016/j.chroma.2015.02.025.

[192] Iglesias J, Sleno L, & Volmer DA. Isotopic labeling of metabolites in drug discovery applications. *Curr Drug Metab*, **2012**;13(9):1213–1225.

[193] Needleman SB & Wunsch CD. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J Mol Biol*, **1970**;48(3):443–453.

[194] Ye J, Fan J, Venneti S, Wan YW, Pawel BR, Zhang J, *et al.* Serine catabolism regulates mitochondrial redox control during hypoxia. *Cancer Discov*, **2014**;4(12):1406–1417. doi: 10.1158/2159-8290.CD-14-0250.

[195] De Livera AM, Dias DA, De Souza D, Rupasinghe T, Pyke J, Tull D, *et al.* Normalizing and integrating metabolomics data. *Anal Chem*, **2012**;84(24):10768–10776. doi: 10.1021/ac302748b.

[196] Wise DR, DeBerardinis RJ, Mancuso A, Sayed N, Zhang XY, Pfeiffer HK, *et al.* Myc regulates a transcriptional program that stimulates mitochondrial glutaminolysis and leads to glutamine addiction. *Proc Natl Acad Sci U S A*, **2008**;105(48):18782–18787. doi: 10.1073/pnas.0810199105.

[197] Dang L, White DW, Gross S, Bennett BD, Bittinger MA, Driggers EM, *et al.* Cancer-associated idh1 mutations produce 2-hydroxyglutarate. *Nature*, **2009**;462(7274):739–744. doi: 10.1038/nature08617.

[198] Ward PS, Patel J, Wise DR, Abdel-Wahab O, Bennett BD, Coller HA, *et al.* The common feature of leukemia-associated idh1 and idh2 mutations is a neomorphic enzyme activity converting alpha-ketoglutarate to 2-hydroxyglutarate. *Cancer Cell*, **2010**;17(3):225–234. doi: 10.1016/j.ccr.2010.01.020.

[199] Giard DJ, Aaronson SA, Todaro GJ, Arnstein P, Kersey JH, Dosik H, *et al.* In vitro cultivation of human tumors: establishment of cell lines derived from a series of solid tumors. *J Natl Cancer Inst*, **1973**;51(5):1417–1423.

[200] Favarger P & Gerlach J. [synthesis of odd-numbered fatty acids with 1- or 2-c14-labelled propionate in the intact mouse]. *Bull Soc Chim Biol (Paris)*, **1960**;42:327–335.

[201] Horning MG, Martin DB, Karmen A, & Vagelos PR. Fatty acid synthesis in adipose tissue. ii. enzymatic synthesis of branched chain and odd-numbered fatty acids. *J Biol Chem*, **1961**;236:669–672.

[202] Wolk A, Vessby B, Ljung H, & Barrefors P. Evaluation of a biological marker of dairy fat intake. *Am J Clin Nutr*, **1998**;68(2):291–295.

[203] Kobayashi R, Murakami T, Obayashi M, Nakai N, Jaskiewicz J, Fujiwara Y, *et al.* Clofibric acid stimulates branched-chain amino acid catabolism by three mechanisms. *Arch Biochem Biophys*, **2002**;407(2):231–240.

[204] Greenaway W & Whatley F. Formation of n-carboxy compounds during trimethylsilanization of amino acids in the presence of carbon dioxide. *Journal of Chromatography A*, **1987**;409(0):383 – 389. doi: http://dx.doi.org/10.1016/S0021-9673(01)86817-5.

[205] Morrow JS, Keim P, & Gurd FRN. Co2 adducts of certain amino acids, peptides, and sperm whale myoglobin studied by carbon 13 and proton nuclear magnetic resonance. *Journal of Biological Chemistry*, **1974**;249(23):7484–7494.

[206] Nunn PB, Davis AJ, & O'Brien P. Carbamate formation and the neurotoxicity of l-alpha amino acids. *Science*, **1991**;251(5001):1619–1620.

[207] Olney JW, Zorumski C, Price MT, & Labruyere J. L-cysteine, a bicarbonate-sensitive endogenous excitotoxin. *Science*, **1990**;248(4955):596–599.

[208] Schaefer W. Reaction of primary and secondary amines to form carbamic acid glucuronides. *Current Drug Metabolism*, **2006**;7(8):873–881.

[209] Hulikova A, Harris AL, Vaughan-Jones RD, & Swietach P. Regulation of intracellular ph in cancer cell lines under normoxia and hypoxia. *J Cell Physiol*, **2013**;228(4):743–752. doi: 10.1002/jcp.24221.

[210] Davis A, Obrien P, & Nunn P. Studies of the stability of some amino acid carbamates in neutral aqueous solution. *Bioorganic Chemistry*, **1993**;21(3):309 – 318. doi: 10.1006/bioo.1993.1026.

[211] Wellen KE, Lu C, Mancuso A, Lemons JMS, Ryczko M, Dennis JW, *et al.* The hexosamine biosynthetic pathway couples growth factor-induced glutamine uptake to glucose metabolism. *Genes Dev*, **2010**;24(24):2784–2799. doi: 10.1101/gad.1985910.

[212] Ferrer CM, Lynch TP, Sodi VL, Falcone JN, Schwab LP, Peacock DL, *et al.* O-glcnacylation regulates cancer metabolism and survival stress signaling via regulation of the hif-1 pathway. *Mol Cell*, **2014**;54(5):820–831. doi: 10.1016/j.molcel.2014.04.026.

[213] Jóźwiak P, Forma E, Bryś M, & Krześlak A. O-glcnacylation and metabolic reprogramming in cancer. *Front Endocrinol (Lausanne)*, **2014**;5:145. doi: 10.3389/fendo.2014.00145.

[214] Kamerling JP, Strecker G, Farriaux JP, Dorland L, Haverkamp J, & Vliegenthart JF. 2-acetamidoglucal, a new metabolite isolated from the urine of a patient with sialuria. *Biochimica et Biophysica Acta (BBA) - General Subjects*, **1979**;583(3):403 – 408. doi: 10.1016/0304-4165(79)90465-3.

[215] Pravdić N, Franjić-Mihalić I, & Danilov B. An improved synthesis of the 2-acetamido-d-glucal derivative 3,4,6-tri-o-acetyl-2-(n-acetylacetamido)-1,5-anhydro-2-deoxy-d-arabino-hex-1-enitol. *Carbohydrate Research*, **1975**;45(1):302 – 306. doi: 10.1016/S0008-6215(00)85889-9.

[216] Pravdić N, Inch TD, & Fletcher H Jr. The formation of acetylated oxazolines through the action of zinc chloride and acetic anhydride on 2-acylamino-2-deoxyaldoses. *J Org Chem*, **1967**;32(6):1815–1818.

[217] Chou WK, Hinderlich S, Reutter W, & Tanner ME. Sialic acid biosynthesis: stereochemistry and mechanism of the reaction catalyzed by the mammalian udp-n-acetylglucosamine 2-epimerase. *J Am Chem Soc*, **2003**;125(9):2455–2461. doi: 10.1021/ja021309g.

[218] Hanover JA, Krause MW, & Love DC. The hexosamine signaling pathway: O-glcnac cycling in feast or famine. *Biochim Biophys Acta*, **2010**;1800(2):80–95. doi: 10.1016/j.bbagen.2009.07.017.

[219] Varki A, Cummings RD, Esko JD, Freeze HH, Stanley P, Bertozzi CR, *et al.* (eds.). *Essentials of Glycobiology*. 2009, Cold Spring Harbor (NY), 2 edition, **2009**. URL `http://www.ncbi.nlm.nih.gov/books/NBK1931/`.

[220] Häuselmann I & Borsig L. Altered tumor-cell glycosylation promotes metastasis. *Front Oncol*, **2014**;4:28. doi: 10.3389/fonc.2014.00028.

[221] Birken DL & Oldendorf WH. N-acetyl-l-aspartic acid: a literature review of a compound prominent in 1h-nmr spectroscopic studies of brain. *Neurosci Biobehav Rev*, **1989**; 13(1):23–31.

[222] Wiame E, Tyteca D, Pierrot N, Collard F, Amyere M, Noel G, *et al.* Molecular identification of aspartate n-acetyltransferase and its mutation in hypoacetylaspartia. *Biochem J*, **2010**;425(1):127–136. doi: 10.1042/BJ20091024.

[223] Becker I, Lodder J, Gieselmann V, & Eckhardt M. Molecular characterization of n-acetylaspartylglutamate synthetase. *J Biol Chem*, **2010**;285(38):29156–29164. doi: 10.1074/jbc.M110.111765.

[224] Chakraborty G, Mekala P, Yahya D, Wu G, & Ledeen RW. Intraneuronal n-acetylaspartate supplies acetyl groups for myelin lipid synthesis: evidence for myelin-associated aspartoacylase. *J Neurochem*, **2001**;78(4):736–745.

[225] Madhavarao CN, Arun P, Moffett JR, Szucs S, Surendran S, Matalon R, *et al.* Defective n-acetylaspartate catabolism reduces brain acetate levels and myelin lipid synthesis in canavan's disease. *Proc Natl Acad Sci U S A*, **2005**;102(14):5221–5226. doi: 10.1073/pnas.0409184102.

[226] Matalon R, Michals K, Sebesta D, Deanching M, Gashkoff P, Casanova J, *et al.* Aspartoacylase deficiency and n-acetylaspartic aciduria in patients with canavan disease. *American Journal of Medical Genetics*, **1988**;29(2):463–471. doi: 10.1002/ajmg.1320290234.

[227] Moffett JR, Ross B, Arun P, Madhavarao CN, & Namboodiri AMA. N-acetylaspartate in the cns: from neurodiagnostics to neurobiology. *Prog Neurobiol*, **2007**;81(2):89–131. doi: 10.1016/j.pneurobio.2006.12.003.

[228] D'Adamo J AF, Gidez L, & Yatsu F. Acetyl transport mechanisms. involvement of n-acetyl aspartic acid in de novo fatty acid biosynthesis in the developing rat brain. *Experimental Brain Research*, **1968**;5(4):267–273. doi: 10.1007/BF00235902.

[229] Bello D, Webber MM, Kleinman HK, Wartinger DD, & Rhim JS. Androgen responsive adult human prostatic epithelial cell lines immortalized by human papillomavirus 18. *Carcinogenesis*, **1997**;18(6):1215–1223.

[230] Hruz T, Laule O, Szabo G, Wessendorp F, Bleuler S, Oertle L, *et al.* Genevestigator v3: a reference expression database for the meta-analysis of transcriptomes. *Adv Bioinformatics*, **2008**;2008:420747. doi: 10.1155/2008/420747.

[231] Huang Y, Bell LN, Okamura J, Kim MS, Mohney RP, Guerrero-Preston R, *et al.* Phospho-$\delta$np63$\alpha$/srebf1 protein interactions: Bridging cell metabolism and cisplatin chemoresistance. *Cell Cycle*, **2012**;11(20):3810–3827. doi: 10.4161/cc.22022.

[232] Sreekumar A, Poisson LM, Rajendiran TM, Khan AP, Cao Q, Yu J, *et al.* Metabolomic profiles delineate potential role for sarcosine in prostate cancer progression. *Nature*, **2009**;457(7231):910–914. doi: 10.1038/nature07762.

[233] Fong MY, McDunn J, & Kakar SS. Identification of metabolites in the normal ovary and their transformation in primary and metastatic ovarian cancer. *PLoS One*, **2011**; 6(5):e19963. doi: 10.1371/journal.pone.0019963.

[234] Ben Sellem D, Elbayed K, Neuville A, Moussallieh FM, Lang-Averous G, Piotto M, *et al.* Metabolomic characterization of ovarian epithelial carcinomas by hrmas-nmr spectroscopy. *J Oncol*, **2011**;2011:174019. doi: 10.1155/2011/174019.

[235] Reitman ZJ, Jin G, Karoly ED, Spasojevic I, Yang J, Kinzler KW, *et al.* Profiling the effects of isocitrate dehydrogenase 1 and 2 mutations on the cellular metabolome. *Proc Natl Acad Sci U S A*, **2011**;108(8):3270–3275. doi: 10.1073/pnas.1019393108.

[236] Arun P, Moffett JR, & Namboodiri AMA. Evidence for mitochondrial and cytoplasmic n-acetylaspartate synthesis in sh-sy5y neuroblastoma cells. *Neurochem Int*, **2009**; 55(4):219–225. doi: 10.1016/j.neuint.2009.03.003.

[237] Ariyannur PS, Madhavarao CN, & Namboodiri AMA. N-acetylaspartate synthesis in the brain: mitochondria vs. microsomes. *Brain Res*, **2008**;1227:34–41. doi: 10.1016/j.brainres.2008.06.040.

[238] Lu ZH, Chakraborty G, Ledeen RW, Yahya D, & Wu G. N-acetylaspartate synthase is bimodally expressed in microsomes and mitochondria of brain. *Brain Res Mol Brain Res*, **2004**;122(1):71–78. doi: 10.1016/j.molbrainres.2003.12.002.

[239] Mráček T, Drahota Z, & Houštěk J. The function and the role of the mitochondrial glycerol-3-phosphate dehydrogenase in mammalian tissues. *Biochim Biophys Acta*, **2013**;1827(3):401–410. doi: 10.1016/j.bbabio.2012.11.014.

[240] McIntosh JC & Cooper JR. Studies on the function of n-acetyl aspartic acid in brain. *Journal of Neurochemistry*, **1965**;12(9-10):825–835. doi: 10.1111/j.1471-4159.1965.tb10267.x.

[241] Patel TB & Clark JB. Lipogenesis in the brain of suckling rats. studies on the mechansim of mitochondrial-cytosolic carbon transfer. *Biochem J*, **1980**;188(1):163–168.

[242] D'Adamo A Jr & Yatsu FM. Acetate metabolism in the nervous system. n-acetyl-l-aspartic acid and the biosynthesis of brain lipids. *J Neurochem*, **1966**;13(10):961–965.

[243] Tahay G, Wiame E, Tyteca D, Courtoy PJ, & Van Schaftingen E. Determinants of the enzymatic activity and the subcellular localization of aspartate n-acetyltransferase. *Biochem J*, **2012**;441(1):105–112. doi: 10.1042/BJ20111179.

[244] Yudkoff M, Nelson D, Daikhin Y, & Erecińska M. Tricarboxylic acid cycle in rat brain synaptosomes. fluxes and interactions with aspartate aminotransferase and malate/aspartate shuttle. *J Biol Chem*, **1994**;269(44):27414–27420.

[245] Benuck M & D'Adamo A Jr. Acetyl transport mechanisms. metabolism of n-acetyl-l-aspartic acid in the non-nervous tissues of the rat. *Biochim Biophys Acta*, **1968**; 152(3):611–618.

[246] Wroblewska B, Wroblewski JT, Pshenichkin S, Surin A, Sullivan SE, & Neale JH. N-acetylaspartylglutamate selectively activates mglur3 receptors in transfected cells. *J Neurochem*, **1997**;69(1):174–181.

[247] Neale JH. N-acetylaspartylglutamate is an agonist at $mglur_3$ in vivo and in vitro. *J Neurochem*, **2011**;119(5):891–895. doi: 10.1111/j.1471-4159.2011.07380.x.

[248] Pissimissis N, Papageorgiou E, Lembessis P, Armakolas A, & Koutsilieris M. The glutamatergic system expression in human pc-3 and lncap prostate cancer cells. *Anticancer Res*, **2009**;29(1):371–377.

[249] Stepulak A, Luksch H, Gebhardt C, Uckermann O, Marzahn J, Sifringer M, *et al.* Expression of glutamate receptor subunits in human cancers. *Histochem Cell Biol*, **2009**;132(4):435–445. doi: 10.1007/s00418-009-0613-1.

[250] Luksch H, Uckermann O, Stepulak A, Hendruschk S, Marzahn J, Bastian S, *et al.* Silencing of selected glutamate receptor subunits modulates cancer growth. *Anticancer Res*, **2011**;31(10):3181–3192.

[251] Shaw TJ, Keszthelyi EJ, Tonary AM, Cada M, & Vanderhyden BC. Cyclic amp in ovarian cancer cells both inhibits proliferation and increases c-kit expression. *Exp Cell Res*, **2002**;273(1):95–106. doi: 10.1006/excr.2001.5426.

[252] Ou Y, Zheng X, Gao Y, Shu M, Leng T, Li Y, *et al.* Activation of cyclic amp/pka pathway inhibits bladder cancer cell invasion by targeting map4-dependent microtubule dynamics. *Urol Oncol*, **2014**;32(1):47.e21–47.e28. doi: 10.1016/j.urolonc.2013.06.017.

[253] Arun P, Madhavarao CN, Moffett JR, & Namboodiri MAA. Regulation of n-acetylaspartate and n-acetylaspartylglutamate biosynthesis by protein kinase activators. *J Neurochem*, **2006**;98(6):2034–2042. doi: 10.1111/j.1471-4159.2006.04068.x.

[254] Guo J, Zhou X, Chen Y, Bai M, Yang X, Zhao K, *et al.* mglur3 promotes proliferation of human embryonic cortical neural progenitor cells by activating erk1/2 and jnk2 signaling pathway in vitro. *Cell Mol Biol (Noisy-le-grand)*, **2014**;60(2):42–49.

[255] Roberts PJ & Der CJ. Targeting the raf-mek-erk mitogen-activated protein kinase cascade for the treatment of cancer. *Oncogene*, **2007**;26(22):3291–3310. doi: 10.1038/sj.onc.1210422.

[256] Mizuarai S, Miki S, Araki H, Takahashi K, & Kotani H. Identification of dicarboxylate carrier slc25a10 as malate transporter in de novo fatty acid synthesis. *J Biol Chem*, **2005**;280(37):32434–32441. doi: 10.1074/jbc.M503152200.

[257] Huang W, Wang H, Kekuda R, Fei YJ, Friedrich A, Wang J, *et al.* Transport of n-acetylaspartate by the na(+)-dependent high-affinity dicarboxylate transporter nadc3 and its relevance to the expression of the transporter in the brain. *J Pharmacol Exp Ther*, **2000**;295(1):392–403.

[258] Kharroubi AT, Masterson TM, Aldaghlas TA, Kennedy KA, & Kelleher JK. Isotopomer spectral analysis of triglyceride fatty acid synthesis in 3t3-l1 cells. *Am J Physiol*, **1992**; 263(4 Pt 1):E667–E675.

[259] Boehmelt G, Fialka I, Brothers G, McGinley MD, Patterson SD, Mo R, *et al.* Cloning and characterization of the murine glucosamine-6-phosphate acetyltransferase emeg32. differential expression and intracellular membrane association. *J Biol Chem*, **2000**; 275(17):12821–12832.

[260] Stein SE, Heller SR, & Tchekhovskoi D. An open standard for chemical structure representation: The iupac chemical identifier. In *Proceedings of the 2003 International Chemical Information Conference (Nimes)*. Infonortics, **2003**; 131–143.

[261] Kanehisa M, Goto S, Sato Y, Kawashima M, Furumichi M, & Tanabe M. Data, information, knowledge and principle: back to metabolism in kegg. *Nucleic Acids Res*, **2014**;42(Database issue):D199–D205. doi: 10.1093/nar/gkt1076.

[262] Wishart DS, Jewison T, Guo AC, Wilson M, Knox C, Liu Y, *et al.* Hmdb 3.0–the human metabolome database in 2013. *Nucleic Acids Res*, **2013**;41(Database issue):D801–D807. doi: 10.1093/nar/gks1065.

[263] Bolton EE, Wang Y, Thiessen PA, & Bryant SH. Chapter 12 - pubchem: Integrated platform of small molecules and biological activities. In RA Wheeler & DC Spellmeyer (eds.), *Annual Reports in Computational Chemistry*, volume 4 of *Annual Reports in Computational Chemistry*, 217 – 241. Elsevier, **2008**;doi: 10.1016/S1574-1400(08)00012-1. URL http://www.sciencedirect.com/science/article/pii/S1574140008000121.

[264] Apodaca R. Inchi canonicalization algorithm. Website, **2006**. URL http://depth-first.com/articles/2006/08/12/inchi-canonicalization-algorithm/.

[265] Kamburov A, Cavill R, Ebbels TMD, Herwig R, & Keun HC. Integrated pathway-level analysis of transcriptomics and metabolomics data with impala. *Bioinformatics*, **2011**; 27(20):2917–2918. doi: 10.1093/bioinformatics/btr499.

[266] Xia J & Wishart DS. Msea: a web-based tool to identify biologically meaningful patterns in quantitative metabolomic data. *Nucleic Acids Res*, **2010**;38(Web Server issue):W71–W77. doi: 10.1093/nar/gkq329.

[267] Chagoyen M & Pazos F. Mbrole: enrichment analysis of metabolomic data. *Bioinformatics*, **2011**;27(5):730–731. doi: 10.1093/bioinformatics/btr001.

[268] Kankainen M, Gopalacharyulu P, Holm L, & Oresic M. Mpea–metabolite pathway enrichment analysis. *Bioinformatics*, **2011**;27(13):1878–1879. doi: 10.1093/bioinformatics/btr278.

[269] Rew RK & Davis GP. Netcdf: An interface for scientific data access. *IEEE Computer Graphics and Applications*, **1990**;10(4):76–82.

[270] Brown M S A, Folk GG, & Rew R. Software for portable scientific data management. *Computers in Physics*, **1993**;7(3):304–308.

[271] O'Boyle NM, Banck M, James CA, Morley C, Vandermeersch T, & Hutchison GR. Open babel: An open chemical toolbox. *J Cheminform*, **2011**;3:33. doi: 10.1186/1758-2946-3-33.

[272] Carroll AJ, Badger MR, & Harvey Millar A. The metabolomeexpress project: enabling web-based processing, analysis and transparent dissemination of gc/ms metabolomics datasets. *BMC Bioinformatics*, **2010**;11:376. doi: 10.1186/1471-2105-11-376.