# IsoCorrectoR: Isotope correction in stable isotope labeling experiments in metabolomics

Dissertation zur Erlangung des Doktorgrades der Naturwissenschaften (Dr. rer. nat.) der Fakultät für Chemie und Pharmazie

der Universität Regensburg

vorgelegt von
Paul Heinrich aus

München

im Jahr 2022

Promotionsgesuch eingereicht am: 22.02.2022
Die Arbeit wurde angeleitet von: PD Dr. Katja Dettmer-Wilde

# Eidesstattliche Erklärung

1. Ich erkläre hiermit an Eides statt, dass ich die vorliegende Arbeit ohne unzulässige Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe; die aus anderen Quellen direkt oder indirekt übernommenen Daten und Konzepte sind unter Angabe des Literaturzitats gekennzeichnet.
2. Bei der Auswahl und Auswertung folgenden Materials haben mir die nachstehend aufgeführten Personen in der jeweils beschriebenen Weise unentgeltlich geholfen:
   (a) Christian Kohler: Mitentwicklung des IsoCorrectoR R-packages (Entwicklung der Software, jedoch nicht Entwicklung des Algorithmus)
   (b) Dr. Maren Schmidt (vormals Feist): Zellkultur- und molekularbiologische Experimente im Rahmen des Kollaborationsprojektes zum Metabolismus von P493-6 B-Zellen
   (c) Lisa Ellmann: Durchführung der Laborarbeiten zur experimentellen Validierung von IsoCorrectoR mittels Isotopologen-Mixturen von bekannter Zusammensetzung
3. Weitere Personen waren an der inhaltlich-materiellen Herstellung der vorliegenden Arbeit nicht beteiligt. Insbesondere habe ich hierfür nicht die entgeltliche Hilfe eines Promotionsberaters oder anderer Personen in Anspruch genommen. Niemand hat von mir weder unmittelbar noch mittelbar geldwerte Leistungen für Arbeiten erhalten, die im Zusammenhang mit dem Inhalt der vorgelegten Dissertation stehen.
4. Die Arbeit wurde bisher weder im In- noch im Ausland in gleicher oder ähnlicher Form einer anderen Prüfungsbehörde vorgelegt.

Regensburg, den 22.02.2022

Paul Heinrich

**Gewidmet meinem Großvater Werner**

The best-laid plans of mice and men
Go oft awry
*Robert Burns, 1785*

# Danksagung

Ein besonders großes Dankeschön geht zu allererst an PD Dr. Katja Dettmer-Wilde für die schwerpunktmäßige Betreueung meiner Arbeit. Sie hat mich dazu ermuntert, meine Doktorarbeit - trotz anfänglicher Zweifel - mit einem Schwerpunkt auf die Entwicklung von IsoCorrectoR fortzuführen, und hat mich bei allen damit im Zusammenhang stehenden Problemen hervorragend unterstützt. Ein großer Dank geht auch an Prof. Dr. Peter J. Oefner, zunächst einmal dafür, dass ich die Doktorarbeit an seinem Lehrstuhl durchführen durfte. Zudem für die Begleitung der Arbeit mit ebenso kritischen wie hilfreichen Anmerkungen und Einwänden, und nicht zuletzt für den Einfall, Christian Kohler in die Entwicklung von IsoCorrectoR mit einzubeziehen. Wäre diese Entscheidung nicht getroffen worden, hätten wir jetzt möglicherweise kein fertiges R-package in Bioconductor. Das bringt mich direkt zu Christian: Ein ganz großes Dankeschön für die gewaltige Arbeit, die du in die Entwicklung unseres gemeinsamen R-packages gesteckt hast, ohne Dich wären wir nie so weit gekommen! Mein Dank geht an dieser Stelle auch an Paul Kürner für die Entwicklung eines GUI-Prototypen für IsoCorrectoR. Vielen, vielen Dank auch an Prof. Dr. Wolfram Gronwald für die intensive Auseinandersetzung mit meinen doch recht umfangreichen Ausführungen zum IsoCorrectoR-Algorithmus, und die Prüfung meiner theoretischen Überlegungen auf Plausibilität. Ein ganz besonderer Dank geht an dieser Stelle auch an Prof. Dr. Frank-Michael Matysik für die Bereitschaft, die beträchtliche Arbeit des Zweitgutachters zu übernehmen. Ebenso danken möchte ich Lisa Ellmann für Ihre Unterstützung bei den Experimenten mit den IsoCorrectoR-Validierungsmixen, sowie Prof. Dr. Rainer Spang für sein Feedback zum IsoCorrectoR-Manuskript und hilfreiche Gespräche. Vielen Dank auch an Dr. Raffaela Berger und Xueni Sun für das Nutzen von IsoCorrectoR in der Testphase, und entsprechendes Feedback. Ganz allgemein geht mein Dank an alle am Institut für Funktionelle Genomik und in der Statistischen Bioinformatik, für Unterstützung, Feedback, Gespräche und nicht zuletzt für unterhaltsame Mittags- und Kaffeepausen!

Außerdem möchte ich mich bei meinen Kooperationspartnern in Göttingen, Prof. Dr. Dieter Kube und Dr. Maren Schmidt (vormals Feist), für die gute Zusammenarbeit bedanken. Darüber hinaus geht mein Dank natürlich an das Bundesministerium für Bildung und Forschung, das meine Arbeiten im Rahmen des Netzwerks eMed MMML-Demonstrator (BMBF-FKZ 031A428A, Rainer Spang, Katja Dettmer-Wilde) gefördert hat, sowie an die Deutsche Forschungsgemeinschaft, von der Teile meiner Arbeit im Rahmen des Projektes KFO-262 (Peter Oefner, Katja Dettmer-Wilde, Rainer Spang) gefördert wurden.

Ein besonders großes Dankeschön zuletzt an meine Frau Ursula, die mir in den Jahren der Promotion immer zur Seite gestanden ist, auch wenn die Zeiten gelegentlich schwierig waren.

# Summary

In metabolomics, stable isotope labeling experiments with tracer isotopes like $^{13}$C or $^{15}$N can provide very valuable information. This includes, for example, the assessment of nutrient contributions to the synthesis of certain metabolites, contributions to metabolic pathways, or the estimation of metabolic fluxes via formal metabolic flux analysis. However, if mass spectrometry (MS) data from stable isotope labeling experiments is used without a prior correction for the abundance of naturally occurring stable heavy isotopes, and, possibly, the isotopic purity of the tracer substrate, this may lead to incorrect conclusions and flux estimates. This is because the mass shift introduced by the incorporation of the tracer isotope into the measured metabolites is also produced naturally, due to the natural abundance (NA) of stable heavy isotopes like $^{13}$C, $^{15}$N, $^{2}$H or $^{18}$O. Similarly, isotopic impurity of the tracer substrate can lead to a "loss" of label and mass shift, which also distorts the measured data. IsoCorrectoR, which is presented with this thesis, is a comprehensive R-based solution for the correction of stable isotope labeling data for NA and tracer purity. While there are already several tools available for that purpose, for example IsoCor (v2), ICT or PyNAC, IsoCorrectoR brings the advantage of combining many state-of-the-art correction features in a single implementation. This includes the low-resolution and resolution-dependent NA correction of MS and MS/MS (tandem MS) data, the NA correction of ultra-high resolution (UHR) data from experiments employing multiple tracer isotopes (*e.g.*, $^{13}$C and $^{15}$N) simultaneously, as well as tracer purity correction and the applicability to any tracer isotope. Additionally, two modes of correction that IsoCorrectoR offers are unique, or were unique at the date of initial publication. The first is a UHR correction approach for UHR data from multiple-tracer experiments that also accounts for tracer purity. Correction for tracer purity can be crucial, especially in the case of multiple tracer isotopes, where impurities of different tracers may add up. While IsoCorrectoR was the first tool to provide this correction feature, the very recently published correction tool AccuCor2 is also capable of correcting multiple-tracer data for tracer purity. It however performs resolution-dependent correction, and not UHR correction. The second novel mode that IsoCorrectoR offers is the resolution-dependent correction of MS/MS data. If MS/MS data is corrected with a MS correction algorithm, this can result in substantial deviations. As a consequence, a correction approach specifically suited for MS/MS data must be applied. If MS/MS data is additionally acquired at high resolution (*e.g.*, with orbitrap or FT-ICR devices), the correction procedure must adapt to this situation as well. In low-resolution correction, all natural abundance contributions that match the nominal mass shift resulting from tracer isotope incorporation are corrected for. In contrast, when applying resolution-dependent correction, natural abundance contributions are considered for correction depending on the instrument resolution at the given molecular ion or fragment *m/z*, because some of the NA contributions can already be resolved spectrometrically and, therefore, do not have to be corrected for. This is taken into account in resolution-dependent MS (which is also available in other tools, *e.g.*,

IsoCor v2) and MS/MS correction (which is a unique feature of IsoCorrectoR), and can have marked effects on the correction results. When coupled to online chromatographic separation, the resolution of orbitrap and FT-ICR devices diminishes in tandem MS, since the transient duration originally available for the full MS spectrum basically has to be split by the number of precursors to be measured (given the number of points across the chromatographic peak is to be kept constant). However, analyses in this thesis have shown that marked differences between low-resolution and resolution-dependent MS/MS correction can still be found down to a resolution of 20000 (at 200 $m/z$), underscoring that this mode of correction should nevertheless be relevant. Even more so, given the rising adoption of orbitrap devices in the metabolomics community and the potential for technological advancements that further increase resolution. Another interesting finding are pronounced differences detected in analyses comparing UHR correction to resolution-dependent correction. This indicates that the basic assumption of UHR correction - that UHR data only needs to be corrected for the NA of the tracer isotope, and no other isotopes - does not necessarily hold in practice. Therefore, UHR correction should be applied with care. To ensure correct functioning, IsoCorrectoR has been thoroughly validated against other available tools (where possible), manual calculations and validation mixtures of known isotopologue composition, all of which yielded very good agreement. More pronounced - but still relatively small - differences have been found in a few cases when comparing the low-resolution MS/MS correction modes of IsoCorrectoR and ICT. These can be attributed to differences in dealing with measurement uncertainty, while the approaches of both tools can be considered viable. Validation against isotopologue mixtures of known composition showed that absolute measurement errors seem to translate more or less directly from uncorrected to corrected values. This potentially magnifies the relative error in corrected data. As an application of isotope correction, a $^{13}$C stable isotope labeling experiment conducted in P493-6 B-cells is part of this thesis. These cells have an inducible *MYC* allele, where the quickly proliferating MYC-high state may serve as a model for Burkitt lymphomas, while the hardly proliferating MYC-low state resembles normal B-cells. Combined addition of the microenvironmental factors IL10 and CpG to MYC-low cells induces proliferation, however. As this cellular state may also serve as a model for certain B-cell lymphomas, and since malignancy is closely connected to metabolic alterations, analyses have been performed to assess the metabolism of MYC-low, MYC-high and IL10+CpG stimulated MYC-low cells in a collaborative study. $^{13}$C tracer analysis with the tracer substrate U-$^{13}$C-glutamine revealed that the usage of glutamine for the biosynthesis of glutamate-derived amino acids and TCA cycle intermediates increases substantially from MYC-low over IL10+CpG stimulated MYC-low to MYC-high cells. Also, the usage of glutamine for oxidative phosphorylation seems to increase in the same manner, as deduced from the fractional abundance of the m+4 indicator labeling pattern of TCA cycle intermediates and aspartate. Combining these results with other analyses (*e.g.*, nutrient deprivation and rescue, oxygen consumption rate and inhibitor/RNAi-knockdown experiments) performed by co-authors of the study, glutaminolysis appears to be of high importance for both MYC-high and IL10+CpG stimulated MYC-low cells. However, MYC-high cells appear to be dependent on glutamine mainly for oxidative phoshorylation via 2-ketoglutarate. IL10+CpG stimulated MYC-low cells, on the other hand, appear to require glutamine to a large degree for nucleotide synthesis via aspartate. In this context, also the aspartate aminotransferase GOT2 seems to play a crucial role.

# List of peer-reviewed publications

## Correcting for natural isotope abundance and tracer impurity in MS-, MS/MS- and high-resolution-multiple-tracer-data from stable isotope labeling experiments with IsoCorrectoR

Heinrich, P., Kohler, C., Ellmann, L., Kuerner, P., Spang, R., Oefner, P.J., and Dettmer, K. (2018). Scientific Reports 8, 17910.

**Abstract**: Experiments with stable isotope tracers such as $^{13}C$ and $^{15}N$ are increasingly used to gain insights into metabolism. However, mass spectrometric measurements of stable isotope labeling experiments should be corrected for the presence of naturally occurring stable isotopes and for impurities of the tracer substrate. Here, we analyzed the effect that such correction has on the data: omitting correction or performing invalid correction can result in largely distorted data, potentially leading to misinterpretation. IsoCorrectoR is the first R-based tool to offer said correction capabilities. It is easy-to-use and comprises all correction features that comparable tools can offer in a single solution: correction of MS and MS/MS data for natural stable isotope abundance and tracer impurity, applicability to any tracer isotope and correction of multiple-tracer data from high-resolution measurements. IsoCorrectoR's correction performance agreed well with manual calculations and other available tools including Python-based IsoCor and Perl-based ICT. IsoCorrectoR can be downloaded as an R-package from: http://bioconductor.org/packages/release/bioc/html/IsoCorrectoR.html.

**Contributions**: Development of the IsoCorrectoR algorithm and Matlab prototype, co-development of the IsoCorrectoR R-package with Christian Kohler, writing of manuscript, performance of analyses.

## Cooperative STAT/NF-$\varkappa$B signaling regulates lymphoma metabolic reprogramming and aberrant GOT2 expression

Feist, M., Schwarzfischer, P., Heinrich, P., Sun, X., Kemper, J., Bonin, F. von, Perez-Rubio, P., Taruttis, F., Rehberg, T., Dettmer, K., et al. (2018). Nat Commun 9, 1–14.

**Abstract**: Knowledge of stromal factors that have a role in the transcriptional regulation of metabolic pathways aside from c-Myc is fundamental to improvements in lymphoma therapy. Using a *MYC*-inducible human B-cell line, we observed the cooperative activation of STAT3 and NF-$\varkappa$B by IL10 and CpG stimulation. We show that IL10 + CpG-mediated cell proliferation of $MYC^{low}$ cells depends on glutaminolysis. By $^{13}C$- and $^{15}N$-tracing of glutamine

metabolism and metabolite rescue experiments, we demonstrate that GOT2 provides aspartate and nucleotides to cells with activated or aberrant Jak/STAT and NF-κB signaling. A model of GOT2 transcriptional regulation is proposed, in which the cooperative phosphorylation of STAT3 and direct joint binding of STAT3 and p65/NF-κB to the proximal GOT2 promoter are important. Furthermore, high aberrant GOT2 expression is prognostic in diffuse large B-cell lymphoma underscoring the current findings and importance of stromal factors in lymphoma biology.

**Contributions**: $^{13}$C tracer analysis in P493-6 B-cells with the tracer substrate U-$^{13}$C-glutamine: $^{13}$C tracer analysis experiment planning in collaboration with Maren Feist, sample preparation for LC-MS, LC-MS/MS measurement of samples, data analysis and interpretation.

## Quantification and $^{13}$C-Tracer analysis of total reduced glutathione by HPLC-QTOFMS/MS

Sun, X., Heinrich, P., Berger, R.S., Oefner, P.J., and Dettmer, K. (2019). Analytica Chimica Acta 1080, 127–137.

**Abstract**: Glutathione is an essential intra- and extracellular antioxidant. The level of glutathione in the body is highly related to different disease states and is a useful indicator of disease risk and oxidative stress status. We have developed a sensitive, selective, and comprehensive LC-MS/MS method for the absolute quantification and $^{13}$C-tracer analysis of total glutathione using dithiothreitol for the reduction of glutathione disulfide. The limit of detection (LOD) was 0.01 $\mu$M, while the lower limit of quantification (LLOQ) was 0.78 $\mu$M, with the linear (R = 0.9997) range extending up to 100 $\mu$M. The intra-run and inter-run coefficients of variation of 2.49% and 2.04%, respectively, attest to high repeatability. Mean (± SD) recoveries of three different concentrations (low, medium, high) of GSH spiked into aliquots of HCT116 cells prior to cell extraction were 108.9% (± 2.1), 100.8% (± 8.3), and 99.9% (± 7.1), respectively. Finally, using a 20 Da wide Q1 window in MRM mode, we were able to detect and relatively quantify all isotopic labeling states of GSH extracted from HCT116 cells fed with either $^{13}$C-labeled glucose or glutamine.

**Contributions**: Development of a reduction protocol to obtain total reduced glutathione that is compatible with HPLC-QTOFMS measurement, co-development of a method for the absolute quantification of total reduced glutathione with Xueni Sun, co-writing of manuscript.

## Optimized Protocol for the In Situ Derivatization of Glutathione with N-Ethylmaleimide in Cultured Cells and the Simultaneous Determination of Glutathione/Glutathione Disulfide Ratio by HPLC-UV-QTOF-MS

Sun, X., Berger, R.S., Heinrich, P., Marchiq, I., Pouyssegur, J., Renner, K., Oefner, P.J., and Dettmer, K. (2020). Metabolites 10, E292.

**Abstract**: Glutathione (GSH) and glutathione disulfide (GSSG) are commonly used to assess the oxidative status of a biological system. Various protocols are available for the analysis of

GSH and GSSG in biomedical specimens. In this study, we present an optimized protocol for the in situ derivatization of GSH with N-ethylmaleimide (NEM) to prevent GSH autooxidation, and thus to preserve the GSH/GSSG ratio during sample preparation. The protocol comprises the incubation of cells in NEM containing phosphate buffered saline (PBS), followed by metabolite extraction with 80% methanol. Further, to preserve the use of QTOF-MS, which may lack the linear dynamic range required for the simultaneous quantification of GSH and GSSG in non-targeted metabolomics, we combined liquid chromatographic separation with the online monitoring of UV absorbance of GS-NEM at 210 nm and the detection of GSSG and its corresponding stable isotope-labeled internal standard by QTOF-MS operated with a 10 Da Q1 window. The limit of detection (LOD) for GS-NEM was 7.81 $\mu$M and the linear range extended from 15.63 $\mu$M to 1000 $\mu$M with a squared correlation coefficient $R^2$ of 0.9997. The LOD for GSSG was 0.001 $\mu$M, and the lower limit of quantification (LLOQ) was 0.01 $\mu$M, with the linear ($R^2 = 0.9994$) range extending up to 10 $\mu$M. The method showed high repeatability with intra-run and inter-run coefficients of variation of 3.48% and 2.51% for GS-NEM, and 3.11% and 3.66% for GSSG, respectively. Mean recoveries of three different spike-in levels (low, medium, high) of GSSG and GS-NEM were above 92%. Finally, the method was applied to the determination of changes in the GSH/GSSG ratio either in response to oxidative stress in cells lacking one or both monocarboxylate transporters *MCT1* and *MCT4*, or in adaptation to the NADPH (nicotinamide adenine dinucleotide phosphate) consuming production of D-2-hydroxyglutarate in cells carrying mutations in the isocitrate dehydrogenase genes *IDH1* and *IDH2*.

**Contributions**: Consulting with regard to data interpretation and manuscript writing.

## Srebp-controlled glucose metabolism is essential for NK cell functional responses

Assmann, N., O'Brien, K.L., Donnelly, R.P., Dyck, L., Zaiatz-Bittencourt, V., Loftus, R.M., Heinrich, P., Oefner, P.J., Lynch, L., Gardiner, C.M., et al. (2017). Nature Immunology 18, 1197–1206.

**Abstract**: Activated natural killer (NK) cells engage in a robust metabolic response that is required for normal effector function. Using genetic, pharmacological and metabolic analyses, we demonstrated an essential role for Srebp transcription factors in cytokine-induced metabolic reprogramming of NK cells that was independent of their conventional role in the control of lipid synthesis. Srebp was required for elevated glycolysis and oxidative phosphorylation and promoted a distinct metabolic pathway configuration in which glucose was metabolized to cytosolic citrate via the citrate-malate shuttle. Preventing the activation of Srebp or direct inhibition of the citrate-malate shuttle inhibited production of interferon-$\gamma$ and NK cell cytotoxicity. Thus, Srebp controls glucose metabolism in NK cells, and this Srebp-dependent regulation is critical for NK cell effector function.

**Contributions**: Correction of $^{13}$C stable isotope labeling data with the Matlab prototype of IsoCorrectoR, consulting with regard to the mass spectrometric measurement of samples from stable isotope labeling experiments.

# Contents

# Chapter I

# Introduction

## 1 Metabolomics

Metabolomics is the comprehensive study of metabolites in biological specimen. Of all -omics approaches, metabolomics can be considered the one closest to the actual phenotype. Thus, it has great potential to contribute to the understanding of phenotypic changes in response to disease, genetic alterations, nutritional, environmental, toxicological or pharmacological influences. Fields of application range from cancer and metabolic diseases over bioengineering to ecology. One of the key difficulties encountered in metabolomics is the inherent heterogeneity of the subject matter, the metabolites. While proteins and especially nucleic acids have relatively low variability in their general chemical stucture, metabolites span a wide range of different compound classes with potentially very distinct chemical and physicochemical properties. Further, they also vary strongly in abundance. As a consequence, there is no single analytical approach that can be employed to cover the entirety of metabolites. Rather, the analytical method has to be tailored towards the metabolite or metabolite class of interest and covers only a subset of the metabolome. Chromatography- and sample preparation techniques that work well for small polar molecules like citric acid cycle intermediates or amino acids are *e.g.* not likely to be the a good choice in lipid analysis (Dettmer et al. (2007), Jang et al. (2018), Clish (2015)).

The analytical platforms predominantly used in metabolomics are nuclear magnetic resonance spectroscopy (NMR) and mass spectrometry (MS). The advantages of NMR are minimal sample preparation requirements, non-destructive and non-discriminating measurement, structure determination capabilities, robustness and the potential for high-throughput fingerprinting. However, sensitivity is relatively low, so that only medium- and highly abundant metabolites are detected. Further, spectral resolution is low in $^1H$ 1D-NMR, making it difficult to identify individual metabolites in complex mixtures. When applying 2D-NMR to improve resolution, measurement becomes relatively time consuming. MS, on the other hand, provides high sensitivity. Mass spectral resolution depends on the mass analyzer employed and can range from low (*e.g.*, quadrupole) to very high (fourier-transform ion cyclotron resonance, FT-ICR). However, regardless of mass analyzer, the capability to resolve individual metabolite species in complex samples can be strongly enhanced by coupling MS online to a separation technology like (high-performance-) liquid chromatography (LC), gas chromatography (GC) or capillary electrophoresis, due to additional separation in the time dimension. Further, it is possible

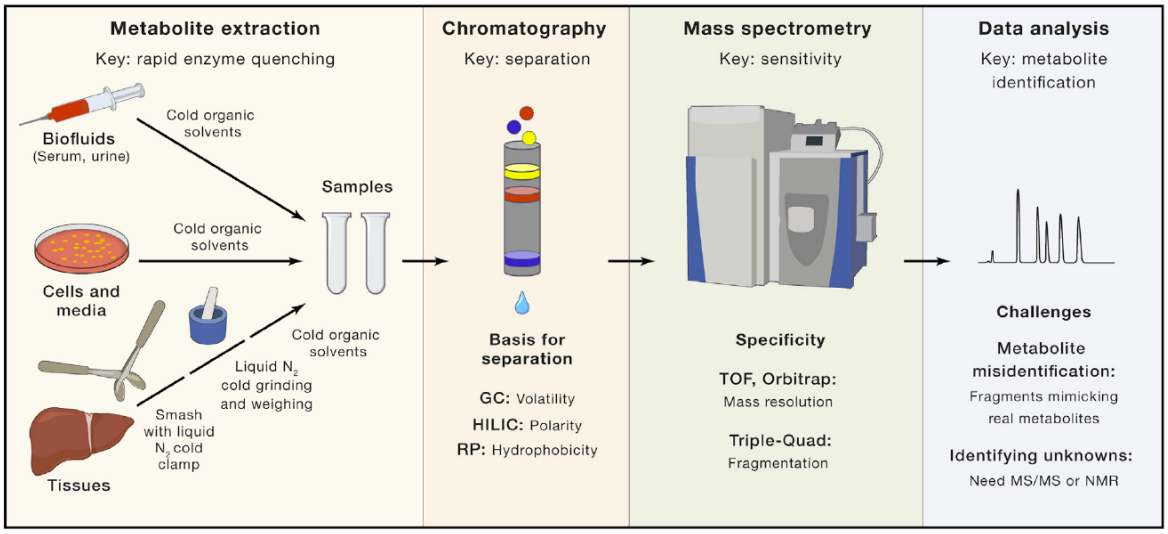



Figure I.1: ***Steps in metabolomic analysis.*** *The figure shows the steps involved in metabolomic analysis: Sample preparation/metabolite extraction, chromatography and mass spectrometry in the case of LC-MS measurement (NMR being an alternative), and data analysis. It must be noted, however, that metabolite identification need not be key in data analysis. Metabolic fingerprinting to detect features of interest and absolute metabolite quantification are also of high relevance.* **(Figure taken from Jang et al. (2018) with permission from Elsevier, legend rewritten)**

to very selectively and sensitively measure only defined species using tandem MS. Here, a precursor ion at a defined $m/z$ ratio is selected by a first mass analyzer. It is subsequently fragmented, and only fragment product ions that match another defined $m/z$ ratio can passage through a second mass analyzer and be detected (both steps may also be performed in a single mass analyzer, *e.g.* in ion trap instruments). Disadvantages of MS in metabolomics are the requirement of sample preparation steps like derivatization that can cause metabolite losses, sample destruction upon measurement and discrimination of metabolite classes. Ionization efficiency at the ion source can vary greatly depending on the properties of the metabolite, the ionization technique used and the sample matrix, which may lead to largely different signal intensities across different metabolites and matrices. This makes the use of (stable isotope labeled) internal standards mandatory when aiming for absolute quantification, especially in the case of electrospray ionization (ESI). Another source of discrimination is the separation technique employed. *E.g.*, a major prerequisite for gas chromatography is sufficient vapor pressure and thermal stability of the analytes. Also, the separation quality obtained with a given technique can vary greatly with the metabolite class (Dettmer et al. (2007), Markley et al. (2017), Bingol (2018)). Figure I.1 from Jang et al. (2018) illustrates the steps involved in metabolomic analysis.

Metabolomics can generally be subdivided into two distinct approaches: Metabolic profiling and metabolic fingerprinting. Metabolic profiling is a targeted approach - a specific set of metabolites is measured quantitatively *e.g.* because the metabolites belong to a certain metabolic pathway of interest or because they serve as biomarkers for clinical diagnosis. Although metabolic profiling of a small subset of metabolites on its own can not be seen as a metabolomic analysis, combining multiple profiling approaches to assess different biochemical

pathways can yield a true metabolomics approach. Metabolic fingerprinting, on the other hand, directly qualifies as an -omics methodology. It is untargeted and unbiased, comparing metabolite patterns among different samples or groups of samples, without any previous knowledge about which metabolites cause the observed patterns. In the MS case, metabolic fingerprints are just described by ion intensities detected at given $m/z$ values and, when a separation technique is employed, retention times. This results in a 3-dimensional dataset for each measured sample. If the aim is to find an initial grouping for the samples measured, metabolic fingerprinting data can be subjected to unsupervised analysis like principal component analysis (PCA) or clustering techniques. PCA finds directions (principal components, PCs) in the high-dimensional feature space for which the variability among samples is maximal. Thus, the PCs can indicate groups of samples that can be formed based on the fingerprinting data and which (combinations of) fingerprinting features are responsible for separating the samples into the observed groups. By identifying the metabolites behind those features, it is possible to understand what differentiates the groups in terms of metabolism and biology. If the goal is to find the metabolic differences between groups of samples that are known *a priori*, like disease vs. healthy, a supervised classification algorithm (*e.g.*, logistic regression) can be used. It can find fingerprinting features (and thus, metabolites) that are most explanatory for differentiating the *a priori* known groups, making this a viable approach for biomarker discovery (Dettmer et al. (2007)).

## 2 Stable isotope labeling experiments in metabolomics

There are many parameters that can be assessed to characterize cellular metabolism. This includes metabolite levels, nutrient uptake and metabolite secretion rates, intracellular metabolic rates (fluxes), nutrient contributions to biosynthesis and pathway activities. Classical metabolomics focuses on metabolite levels. It is well suited for detecting differences in metabolic states between conditions (*e.g.* disease vs. healthy or treatment vs. no treatment). However, metabolite levels only provide a static view of the metabolic processes under study and do not allow to make conclusions about the metabolic rates of conversion between metabolites (fluxes) that produce the measured concentrations. An increase of metabolite concentration may be indicative of an increased activity of metabolite producing enzymes, but it could also be a result of decreased activity of metabolite consuming enzymes. Consequently, high concentrations need not be accompanied by high fluxes. A suitable analogy for this aspect of metabolism is traffic. Traffic (that is, flux) increases with car density (concentration), however only until traffic slows. At this point, high car density is clearly not synonymous with high traffic. Similarly, high concentrations of a metabolite may be present, but producing reactions may operate at a low rate at the same time, *e.g.* due to negative feedback to the producing enzymes (to remain with the traffic analogy). Figure I.2 from Jang et al. (2018) illustrates this subject. Thus, while classical metabolomics can detect differences between states, it is not capable of explaining which changes in metabolic routes are responsible for the observed outcome (Buescher et al. (2015), Jang et al. (2018)). However, such information is crucial *e.g.* when trying to understand biological mechanisms and for developing targeted disease therapies. If, for example, it is known which metabolic routes are especially active and essential for proliferation in specific types of cancer cells, enzymes along these routes qualify as potential drug targets in personalized medicine approaches (Hiller & Metallo (2013)).

A set of approaches exists to assess the dynamics of metabolism. A very macroscopic picture

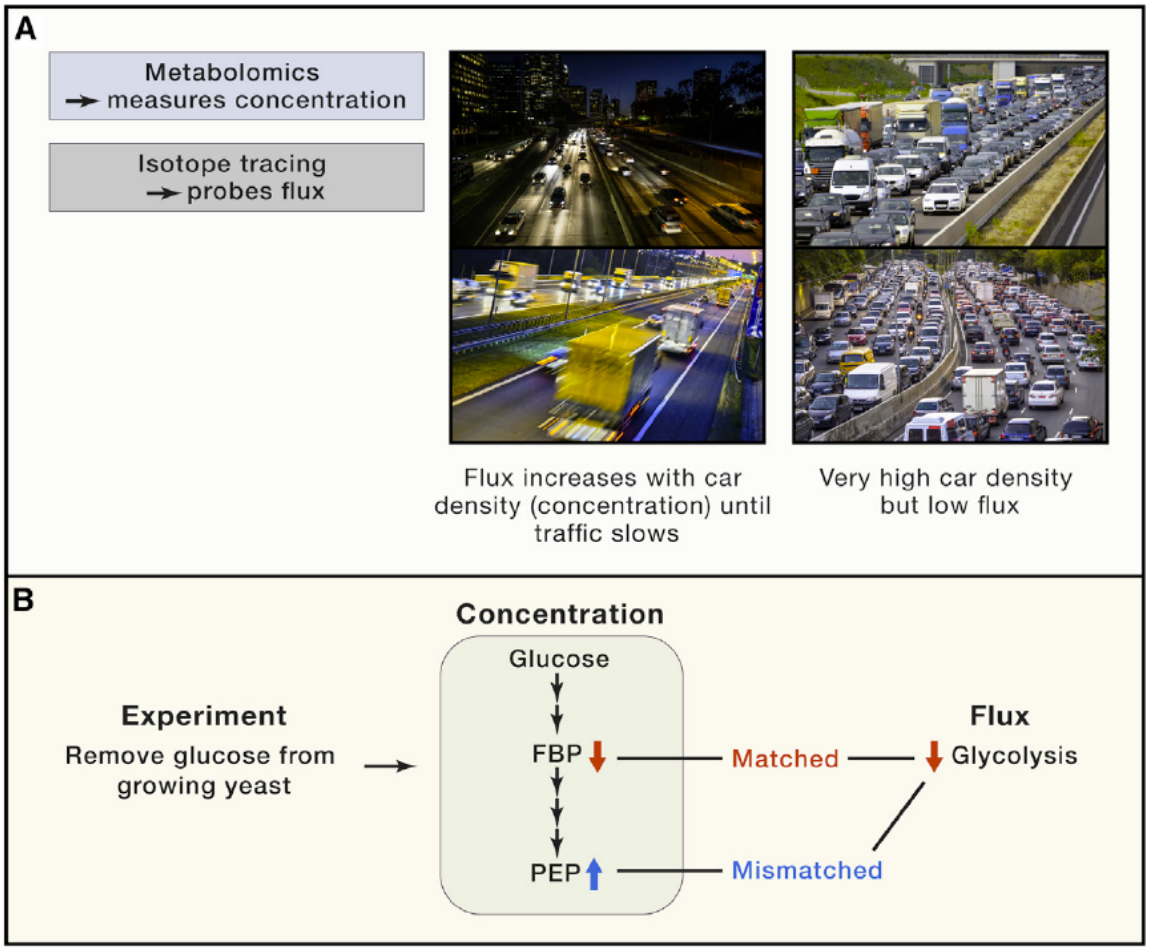
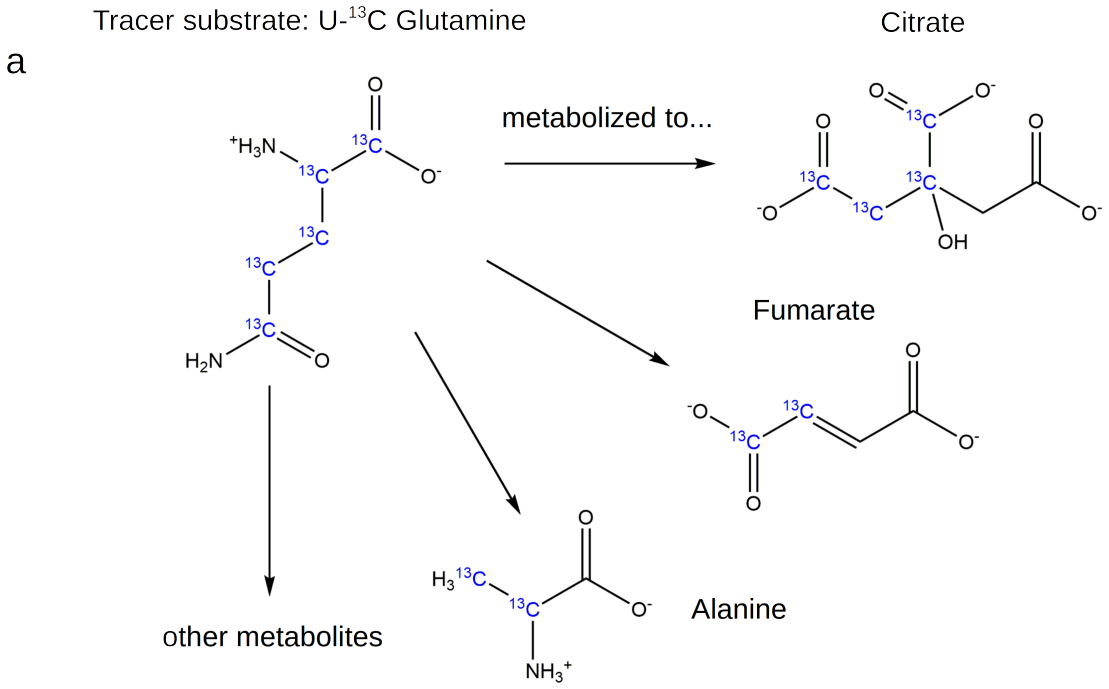



Figure I.2: ***Metabolite levels versus metabolic flux.*** *(A) Concentration and flux are distinct properties. An analogy for this is traffic, which slows at high car densities. (B) Biological example of divergence between concentration and flux. Glucose removal decreases flux throughout glycolysis, but some glycolytic intermediates increase. FBP, fructose-1,6-bisphosphate; PEP, phosphoenolpyruvate.* **(Figure and legend taken from Jang et al. (2018) with permission from Elsevier, legend modified)**

of dynamics can be provided by metabolite uptake and secretion rates in combination with growth rates. However, in many cases, a more detailed understanding of metabolic processes is desired. This can be achieved by employing stable isotope labeled tracers. In a stable isotope labeling experiment, a nutrient labeled with a stable heavy isotope (*e.g.*, $^{13}$C) is provided to the biological system under study. Most often these are cells in cell culture, but stable isotope labeling is also possible in whole organisms. When the labeled nutrient (tracer or tracer substrate) is metabolized, the stable isotopes are transferred to other metabolites, according to the metabolic routes that convert the tracer. These labeled metabolites can then be measured via MS (the incorporation of stable heavy isotopes results in a mass shift) or NMR. The labeling patterns obtained then allow to infer information on nutrient contribution to certain metabolites or pathways, or to assess fluxes (Buescher et al. (2015)). Figure I.3 illustrates how the tracer isotope $^{13}$C is transferred from the tracer substrate U-$^{13}$C-glutamine to other metabolites in a $^{13}$C stable isotope labeling experiment, and how MS data obtained from a stable isotope labeling experiment can be visualized.

If the goal is to quantitatively determine metabolic rates, formal metabolic flux analysis (MFA or $^{13}$C MFA, if $^{13}$C is the tracer isotope) needs to be employed. In MFA, *a priori* information on the biochemical network (*i.e.*, a network model), metabolite labeling patterns and cellular uptake and secretion rates are required as input information. The fluxes, which are the parameters of the biochemical network model, can be obtained by fitting the model to the measured data. While providing very detailed quantitative information on the usage of metabolic routes, this approach can be very time and data-intensive in practice. However, in many cases direct interpretation of labeling patterns without employing formal MFA can be sufficient to obtain the desired insights. This approach is termed (stable isotope) tracer analysis ($^{13}$C tracer analysis, if $^{13}$C is the tracer isotope). Using tracer analysis, nutrient contribution to the production of different metabolites, relative pathway activities and qualitative changes in pathway contributions (via alternative metabolic routes) can be assessed. Under certain circumstances, it is even possible to infer quantitative fluxes through tracer analysis, using labeling patterns and metabolite levels (Buescher et al. (2015), Jang et al. (2018)).

A requirement for most tracer analysis approaches is metabolic steady state (MSS), meaning that both intracellular metabolite levels and intracellular fluxes of a cell (population) are constant. However, it is usually sufficient that the system under study is at metabolic pseudo-steady state (MPSS), *i.e.*, changes in metabolite levels and metabolic rates are minimal on the experimental time-scale. In monolayer cell culture, the exponential growth phase is usually considered to reflect MPSS. A prerequisite is that nutrients do not become limiting, so sufficient supply must be given. MPSS can be verified by suitable experiments which show that there is no change in the levels of the metabolites of interest over time. For stable isotope labeling experiments, also isotopic steady state (ISS) is of special importance. When metabolites become labeled in a stable isotope labeling experiment, their enrichment with the tracer isotope will increase over time, up to a point when the levels of the different isotopologues that are produced become constant. This state is defined as isotopic steady state, which can be assessed in a similar way as MPSS. It is important to note that the time required to reach ISS depends on both the fluxes and the size of the metabolite pools, and that different pathways reach ISS at substantially different time scales. For example, for glycolysis intermediates, ISS is usually reached within seconds, while it may be hours for tricarboxylic acid (TCA) cycle intermediates. Figure I.4, taken from Buescher et al. (2015), illustrates metabolic and isotopic steady state. It further explains the mass distribution vector term and shows the effect that cellular compartments can have on labeling patterns. Patterns may differ strongly between



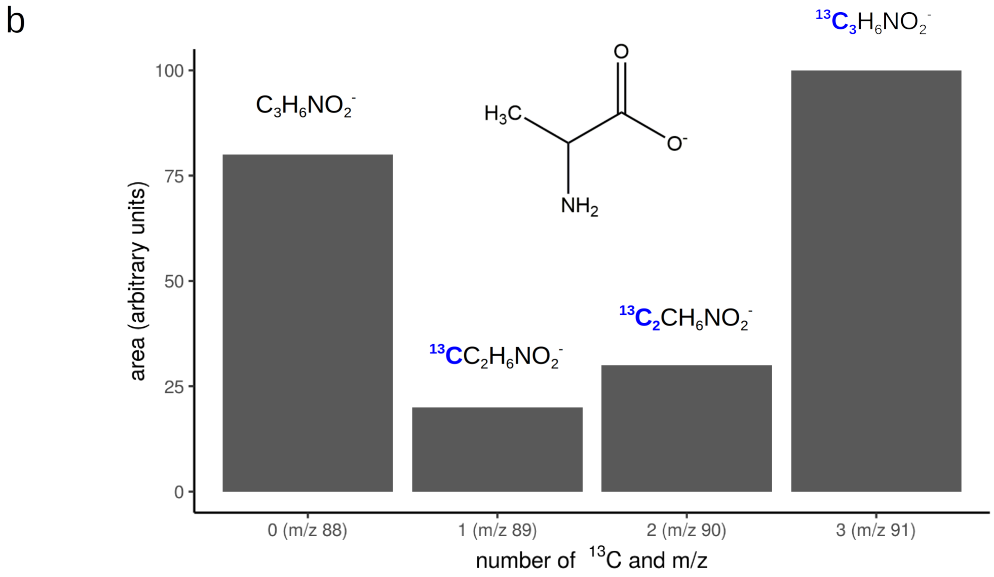




Figure I.3: ***Stable isotope labeling experiments in metabolomics.*** *Panel a illustrates how the tracer isotope $^{13}C$ is transferred from the tracer substrate U-$^{13}C$-glutamine to other metabolites in a $^{13}C$ stable isotope labeling experiment. Panel b shows (simulated) example results from MS measurements of isotopologues of the amino acid alanine ($C_3H_6NO_2^-$, negative mode) in a $^{13}C$ stable isotope labeling experiment. The $^{13}C$ tracer isotope introduces a mass shift into the metabolite, so that the different isotopologues can be measured at different m/z values.*

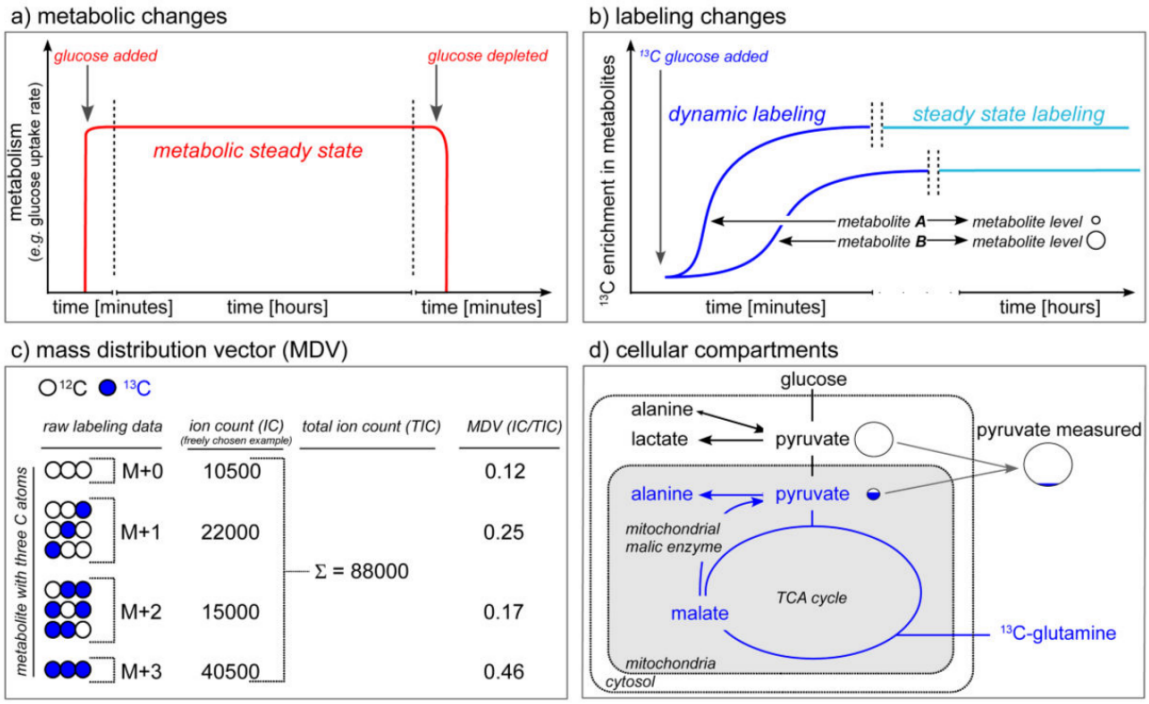



Figure I.4: ***Metabolic and isotopic steady state, mass distribution vectors and cellular compartments.*** *a) Time dependent metabolic changes: metabolism reaches a metabolic steady state when the parameters of interest (e.g. glucose uptake rate) are constant over time. b) Time dependent labeling changes: upon addition of an isotopically labeled carbon source, the isotopic enrichment will increase in the metabolites until the steady state enrichment is reached. c) Mass distribution vector (MDV) (also known as mass distribution (MID) vector): Labeling patterns are MDVs that consist of the fractional abundance of each isotopologue (mass isotopomer). M denotes mass of the unlabeled metabolite. d) Cellular compartmentalization: Most labeling pattern detection methods cannot resolve different cellular compartments, thus the whole cell average labeling pattern is measured.* **(Figure and legend taken from Buescher et al. (2015) with permission from Elsevier)**

compartments. This has to considered when interpreting labeling data, which most often is the result of a mix of compartment labeling patterns. The mass distribution vector (MDV, also termed mass isotopomer distribution, MID) is the quantitative expression of the labeling pattern of a metabolite. MDVs describe the fractional abundance of each isotopologue (also termed mass isotopomer) of a metabolite expected to occur through labeling. It is computed by normalizing the counts/area obtained for each isotopologue to the sum counts/area of all possible isotopologues (Buescher et al. (2015)). Isotopologues are molecules that differ only in their isotopic composition, but have the same elemental composition and bonding structure. Isotopomers have equal isotopic composition, but vary in the positional arrangement of isotopes. A single isotopologue may comprise different isotopomers. In stable isotope labeling metabolomics, the terms isotopologue and mass isotopomer are used interchangeably, so mass isotopomers are not to be confused with actual isotopomers.

If the tracer is uniformly labeled (*e.g.*, U-$^{13}C_6$-glucose) and if there is no unlabeled tracer available to the cells, the tracer isotope enrichment of metabolites observed at ISS reflects the

quantitative contributions of the tracer substrate to the synthesis of the analyzed metabolites. Mostly, metabolites will not be 100% labeled at ISS, because not only the tracer contributes to their synthesis (*e.g.*, both glucose and glutamine can contribute to the synthesis of certain amino acids like aspartic acid) and due to tracer impurity. Nutrient contribution can also be interpreted as the flux of the tracer to the metabolites, relative to all unlabeled sources. It does, however, allow no conclusions about absolute flux. Addressing the question of nutrient contribution is one of the practically most widely applicable approaches in tracer analysis. Information obtained from such analyses can *e.g.* be used to infer which metabolites in cancer cells are especially dependent on certain nutrients. Such knowledge can be very valuable for identifying new targets for therapy. Another widely applicable method is the qualitative assessment of pathway contribution. It can be used to infer whether a given metabolite is synthesized through a specific pathway or not. This is possible if the production of the metabolite through the pathway of interest yields a unique labeling pattern when employing a given tracer. Examples are M+3 malate from U-$^{13}C_6$-glucose, which suggests pyruvate carboxylase-driven anaplerosis of the TCA cycle, or M+5 citrate from U-$^{13}C_5$-glutamine, which suggests reductive carboxylation. An extensive listing of tracers with associated indicator labeling patterns of metabolites and interpretations can be found in Jang et al. (2018). Under more specific circumstances, also a quantitative assessment of relative pathway activity is possible via tracer analysis. Prerequisites are that a metabolite can be synthesized via two distinct pathways that start with the same substrate, and that the labeling patterns of this metabolite differ depending on the pathway of origin. Then, relative pathway activity can be inferred by calculating a split ratio from the representative labeling patterns. An example for this is the biosynthesis of pyruvate from glucose through a) glycolysis and b) the oxidative pentose phosphate pathway. Different labeling patterns in pyruvate are obtained through the two pathways when employing 1,2-$^{13}C_2$-glucose as the tracer substrate (Buescher et al. (2015), Jang et al. (2018)).

The quantitative tracer analysis approaches discussed above (nutrient contribution, relative pathway activity) require isotopic steady state. However, quantitative information can also be inferred at isotopic non-steady state under certain circumstances, through dynamic labeling. Assuming M(P)SS, if the direct substrate metabolite of the reaction of interest is 100% labeled (*e.g.*, because it is the tracer) and if the metabolite of interest is not the product of a condensation reaction, the absolute flux in moles or molarity per cell and unit time from substrate to metabolite, $F_{in}$, can be calculated according to $F_{in} = ln(2) \cdot \frac{[Y]}{t^{1/2}}$. Here, $t^{1/2}$ is the labeling half-time and $[Y]$ is the pool size in molarity or moles per cell. To be able to obtain $t^{1/2}$, the degree of label incorporation must be measured at different time points starting with application of the tracer to cell culture. As $[Y]$ is needed, a measurement of absolute metabolite levels is required as well. Difficulties arise if the processes for which dynamic labeling should be measured are very fast, like transformations in glycolysis. In dynamic labeling cases more complicated than the one described, fluxes can't be inferred from tracer analysis and non-stationary MFA must be employed (Buescher et al. (2015), Jang et al. (2018)).

Regardless of whether formal metabolic flux analysis or stable isotope tracer analysis is performed, data from stable isotope labeling experiments should always be corrected for natural isotope abundance and, possibly, tracer purity. In mass spectrometry, the incorporation of a tracer isotope like $^{13}C$ introduces a mass shift. This mass shift can, however, also occur without any labeling, as stable isotopes of higher mass like $^{13}C$, $^{15}N$, $^{2}H$ or $^{17}O$ are present naturally, at a given natural abundance. To avoid misinterpretations and incorrect flux estimates, data should be corrected for the presence of these naturally occurring heavy isotopes.

Impurities of the tracer substrate can have a similar effect due to a "loss" of label, and should be corrected for as well (Midani et al. (2017), Buescher et al. (2015)). A more detailed introduction to isotope correction in stable isotope labeling experiments will be given in chapter II.

# 3 The scope of this thesis

In this thesis, IsoCorrectoR is presented, an R-based tool for correcting stable isotope labeling data for natural isotope abundance and tracer purity. While there are already several other tools available for that purpose (*e.g.*, IsoCor - Millard et al. (2012, 2019), ICT - Jungreuthmayer et al. (2016) or PyNAC - Carreer et al. (2013)), these still lack features like resolution-dependent MS/MS correction or tracer purity correction when performing UHR correction on data from multiple-tracer experiments. IsoCorrectoR combines many different modes of natural isotope abundance and tracer purity correction (*e.g.*, MS/MS and resolution-dependent correction) in a single, comprehensive solution, while also extending functionality over what currently available tools can offer. Chapter II covers a detailed description of the algorithm behind IsoCorrectoR, as well an analysis of the application of IsoCorrectoR's various modes of correction to stable isotope labeling data. Further, IsoCorrectoR is validated in various scenarios. As an area of application, chapter III then deals with a $^{13}C$ stable isotope labeling experiment conducted in a P493-6 B-cell line. This cell line has an inducible *MYC*-allele and can serve as a model for B-cell lymphoma.

# Chapter II

# IsoCorrectoR

## 1 Introduction: Isotope correction in stable isotope labeling experiments

As illustrated in section 2 in the general introduction, stable isotope labeling experiments employed in metabolomics can provide a wealth of information. This includes pathway contributions, (quantitative) nutrient contributions and metabolic rates of conversion in metabolic networks (fluxes). However, if MS data from stable isotope labeling experiments is interpreted or used for metabolic flux analysis (MFA) without prior data correction, this may lead to false conclusions or incorrect flux estimates. The incorporation of a tracer isotope like $^{13}C$ from the tracer substrate introduces a mass shift into a given metabolite with respect to the unlabeled state. The different isotopologues (or mass isotopomers) originating from labeling can then be measured via MS, yielding an intensity value - *e.g.*, peak area - for each species. However, without correction, these values do not reflect the true abundance of the labeled species, the main reason for this being the natural abundance (NA) of stable heavy isotopes. This is illustrated in Figure II.1, which shows (simulated) exemplary results of MS measurements of alanine isotopologues ($C_3H_6NO_2^-$, negative mode) from a $^{13}C$ stable isotope labeling experiment. When measuring at an $m/z$ value of 90 with nominal resolution, the initial expectation may be that the area value obtained corresponds solely to the alanine species that has incorporated 2 $^{13}C$ from labeling through the tracer substrate (assuming alanine can be resolved from other metabolites in the sample). However, as shown in the figure, it is not only the $^{13}C_2{}^{12}C^1H_6{}^{14}N^{16}O_2{}^-$ species (all isotopes except the label are in their most abundant and lightest form) that corresponds to the value. Also other alanine isotopologues are measured at an $m/z$ value of 90. This occurs, because stable isotopes of higher mass also exist naturally, at a defined natural abundance (*e.g.*, 0.0107 for $^{13}C$ or 0.00364 for $^{15}N$). Thus, species that have obtained 1 $^{13}C$ from labeling and additionally contain an isotope of higher mass due to NA also contribute to the 90 $m/z$ measurement, as well species that have obtained no label at all and show a mass shift of +2 purely due to NA. To obtain the true abundance of the different labeled species, the measured data has to be corrected for those NA contributions.

Initial methods to correct stable isotope labeling data for NA relied on the measurement of an unlabeled standard of the respective molecule. In the most straightforward approach, a mass isotopomer distribution (MID) of a given metabolite is corrected by essentially subtracting the normalized MID of its unlabeled standard in successive steps. However, this approach is

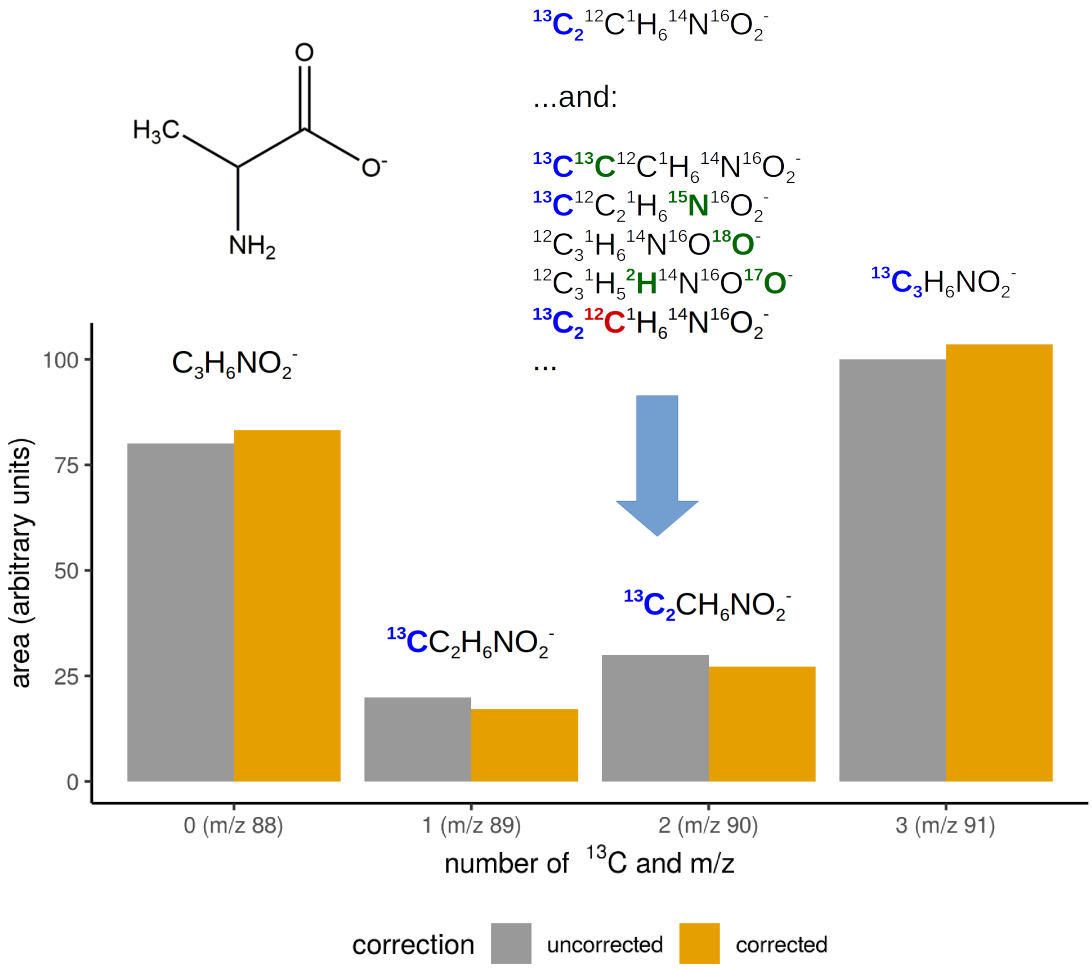



Figure II.1: ***Natural isotope abundance and tracer impurity in MS measurements from stable isotope labeling experiments.*** *The figure shows (simulated) example results from MS measurements of isotopologues of the amino acid alanine ($C_3H_6NO_2^-$, negative mode) in a $^{13}C$ stable isotope labeling experiment. The $^{13}C$ tracer isotope introduces a mass shift into the metabolite, so that the different isotopologues can be measured at different m/z values. However, not only the expected labeled species contributes to the measurement. This is illustrated for the measurement of the 2 $^{13}C$ species at an m/z value of 90. All the alanine isotopologues depicted above the measurement contribute to the value at nominal mass resolution. This is however only an exemplary subset of all possible contributions. A blue colored isotope corresponds to label incorporated from the tracer substrate. A green color identifies stable isotopes of higher mass that are present due to natural isotope abundance. A red color defines an "impure" tracer isotope of lower mass obtained from the tracer substrate due to tracer impurity. Values corrected for NA and tracer purity were obtained by employing IsoCorrectoR's low-resolution MS correction approach, assuming a tracer purity of 99%.*

incorrect, as it does not account for the fact that the MID of the unlabeled standard differs from the MID of species that have incorporated one or more $^{13}$C from labeling. In these cases, the probability of obtaining a $^{13}$C due to NA is lower, because less C atom positions are available - a certain number is already occupied due to labeling. This "skewness" effect associated with MIDs of labeled species can be accounted for by measuring standards of all labeled species and using them for correction, instead of only the unlabeled standard. However, acquiring spectra of standards from all possible labeled species originating from a stable isotope labeling experiment is difficult and likely not even feasible in many cases (Midani et al. (2017)). An alternative is the computation of a correction matrix based on reported natural isotope abundance values, as introduced in Fernandez et al. (1996). This approach was combined with a non-linear fitting procedure to estimate the true isotopic abundances in the sample, using the measurement of an unlabeled standard.

Another factor that leads to the deviation of measured values from the true abundance of labeled species is the impurity of the tracer substrate (Millard et al. (2012), Midani et al. (2017)). Due to tracer impurity, a $^{13}$C labeled tracer substrate, for instance, may contain $^{12}$C at positions that should actually be labeled. If a given metabolite receives *e.g.* 3 carbon atoms from a tracer substrate that should be fully $^{13}$C labeled, it would be expected to contribute to the measurement of the 3 $^{13}$C species at a mass shift of +3. However, if one of the carbons from the tracer substrate is "impure" (*i.e.*, $^{12}$C instead of $^{13}$C), the species will falsely contribute to the measurement of the 2 $^{13}$C species at a mass shift of +2, although it actually obtained 3 carbons from the tracer substrate. This is illustrated in Figure II.1: The red-colored $^{12}$C in one of the alanine isotopologues contributing to the $m+2$ value is due to tracer impurity in the 3 $^{13}$C species. Without impurity, the species would contribute to the $m+3$ value.

When performing correction of MS data for NA, different correction approaches are required depending on the resolution of the mass analyzer: low resolution, high resolution or ultra-high resolution. In the context of NA correction, low resolution means that the incorporation of isotopes which introduce the same nominal mass shift into the molecule cannot be resolved by the mass spectrometer. For example, it would not be possible to distinguish the incorporation of a $^{13}$C tracer isotope from the incorporation of a $^{15}$N isotope in a $^{13}$C stable isotope labeling experiment. Thus, data acquired with quadrupole, quadrupole ion trap and many TOF devices can usually be considered low-resolution. In low-resolution NA correction, all NA derived species that have the same nominal mass shift as a given expected labeled species have to be corrected for (see the isotopologues in Figure II.1). IsoCor (IsoCor v1, Millard et al. (2012)) is a Python-based tool that can be used to correct low-resolution MS data from stable isotope labeling experiments for both NA and tracer purity, using any tracer isotope ($^{13}$C, $^{15}$N...). IsoCor's correction algorithm is based on an approach suggested by van Winden et al. (2002). Here, a correction matrix is constructed from matrix multiplication of individual correction matrices for the different elements in the molecule, all of which are computed using reported values of natural isotope abundance. Next, correction is performed by solving the system of linear equations originating from $v_m = P \cdot v_c$, where $v_m$ is the vector of measured values (isotopologues), $v_c$ is the vector of corrected values and $P$ is the correction matrix. To deal with measurement noise, the equation is solved for $v_c$ using a least-squares method that minimizes the sum of squared errors, with the constraint that corrected values must be greater than 0.

While IsoCor is an excellent tool for dealing with low-resolution MS data from stable isotope labeling experiments, it is unable to handle MS/MS data. As discussed in Niedenführ et al. (2016) and Jungreuthmayer et al. (2016), MS/MS labeling data requires a different algorith-

mic approach to correction to account for the labeling of fragments. Both the Matlab-based MS-X-Corr (Niedenführ et al. (2016)) and the Perl-based ICT (Isotope Correction Toolbox, Jungreuthmayer et al. (2016)) can deal with MS/MS data from stable isotope labeling experiments. However, MS-X-Corr has no functionality to deal with tracer impurity and can't be used for correcting the NA of the tracer element (*e.g.*, $^{13}$C in a $^{13}$C tracing experiment) in the core molecule. Here, the core molecule shall be defined as the endogenous metabolite, excluding parts that stem from derivatization. Thus, ICT is the only tool currently available that provides full correction functionality on (low-resolution) MS/MS labeling data. Similarly to IsoCor, it is applicable to any tracer isotope and operates by computing a correction matrix and solving the resulting system of linear equations. It does, however, not use a least-squares approach to deal with measurement noise, but rather directly sets negative values to 0 in the process of solving the equations.

Given sufficient detection time, FT-ICR and orbitrap devices can attain ultra-high resolution (UHR) in the range of 400000 (400 $m/z$) or more. At such high resolution, it is considered possible to reliably resolve mass defect related differences between the incorporation of isotopes that yield the same nominal mass shift, like $^{13}$C and $^{15}$N or $^{2}$H (Carreer et al. (2013)). In terms of stable isotope labeling MS data acquired at ultra-high resolution, this means that it should be possible to resolve NA contributions that do not come from the tracer isotope itself already spectrometrically. When considering Figure II.1, only the isotopologue that carries an additional $^{13}$C due to NA and the species that has "lost" a $^{13}$C due to tracer impurity would contribute to the measurement, but none of the others. Thus, correcting such data using a low-resolution approach may yield substantial overcorrection. A reasonable simplification to perform valid correction at this level resolution is to account only for the NA of the tracer isotope, which is implemented in the Python-based tool PyNAC (Carreer et al. (2013)). An additional implication of being able to reliably resolve mass defect related differences is the possibility of performing labeling experiments where multiple tracer isotopes - *e.g.*, $^{13}$C and $^{15}$N - are used simultaneously. At said resolutions, it should be possible to resolve the peaks derived from labeling with $^{13}$C, $^{15}$N and combinations of the isotopes (Moseley (2010)). The term stable isotope resolved metabolomics (SIRM) is used for such analyses (Le et al. (2012)). PyNAC is capable of correcting data from such multiple-tracer experiments for NA, it does, however, lack the functionality to also correct for tracer purity.

Mostly, stable isotope labeling data acquired with the increasingly used orbitrap devices falls into the resolution range between low and ultra-high resolution (140000 is a common reference resolution at 200 $m/z$), and is considered high, but not ultra-high. At this level of resolution, some NA contributions that share the nominal mass shift of a given labeled species can be resolved, while others cannot. When again considering Figure II.1 as an example, *e.g.* the species containing $^{15}$N may be resolved, while the species containing $^{17}$O and $^{2}$H are not. Thus, a correction algorithm suited for such data must resolution-dependently correct for contributions that cannot be resolved spectrometrically, but exclude resolved species from correction. Such an approach was first implemented in the R-based tool AccuCor (Su et al. (2017)), designed for correcting stable isotope labeling data acquired with orbitrap devices. Similarly to IsoCor, it builds a correction matrix through matrix multiplication of individual element correction matrices. Depending on the instrument resolution provided, it however excludes isotopes that can generally be resolved from the tracer isotope from the computation of the matrices. ElemCor (Du et al. (2019)) builds upon the algorithm implemented in AccuCor and adds a small improvement. Additionally, it provides the option to construct a correction matrix from the measured spectrum of an unlabeled standard (while accounting for the skewness-effect

explained earlier). However, the algorithmic approach followed by AccuCor (and ElemCor) is slightly flawed, as shown in Millard et al. (2019). While individual isotopes may be generally resolved, this is not necessarily true for combinations of those isotopes with other isotopes. Here, mass defect related differences may add up or cancel each other out. A new version of IsoCor (IsoCor v2, Millard et al. (2019)) takes these peculiarities into account. Thus, it appears that currently only IsoCor v2 can be considered to perform the resolution-dependent correction of high-resolution labeling data correctly. As stated previously, MS/MS labeling data should be corrected using an MS/MS correction algorithm. While ICT and MS-X-Corr can perform correction on low-resolution MS/MS data, there is currently no tool that allows for the resolution-dependent correction of high-resolution MS/MS data.

With this thesis, IsoCorrectoR is presented, a comprehensive R-based tool for the correction of stable isotope labeling data for NA and tracer purity. IsoCorrectoR covers both the correction of low-resolution MS and MS/MS data. Additionally, it is capable of correcting ultra-high-resolution labeling data, including data from multiple-tracer experiments (SIRM), similarly to PyNAC. However, in contrast to PyNAC, IsoCorrectoR can correct such data also for impurities of the tracer substrate. As shown in the results section, this can be crucial, especially as impurities of multiple simultaneously used tracer elements can add up to have substantial impact. Furthermore, IsoCorrectoR is able to correct high-resolution MS labeling data in a resolution-dependent manner, just as IsoCor v2. Importantly, though, IsoCorrectoR's algorithm extends resolution-dependent correction also to MS/MS data. Many orbitrap devices are capable of tandem MS, and the positional labeling information provided by MS/MS measurements can be very valuable. This is especially true for metabolic flux analysis, where the additional information can help in estimating fluxes more precisely (Choi & Antoniewicz (2019)).

The following chapter deals with IsoCorrectoR and its application to stable isotope labeling data. First, the algorithmic approach behind IsoCorrectoR is presented in detail. Then, the impact of IsoCorrectoR's various modes of correction on stable isotope labeling data is assessed. This includes, for example, the effect of natural isotope abundance and tracer purity correction, an analysis of the error introduced when applying MS correction on MS/MS labeling data and a comparison of low-resolution correction and UHR correction to resolution-dependent correction. Finally, IsoCorrectoR is validated by comparison to other available tools (IsoCor, ICT, PyNAC), the results expected from validation mixtures of known composition and manual calculations.

# 2 Contributions

The algorithm and a Matlab-based prototype of IsoCorrectoR have been developed by Paul Heinrich. The Matlab prototype contains all correction functionality except resolution-dependent correction (low-resolution correction, low-resolution MS/MS correction, tracer purity correction, correction for different tracer elements, ultra-high-resolution correction of experiments with multiple tracers). Because Matlab is commercial software (and thus not free to use), the Matlab prototype has been rewritten in R by Christian Kohler from the Statistical Bioinformatics department. Paul Heinrich and Christian Kohler then jointly extended this R prototype to the R package IsoCorrectoR, which was subsequently published in Bioconductor, an open-source repository of bioinformatics software developed in R. Additionally, a graphical

user interface (GUI) for IsoCorrectoR has been developed by Paul Kürner from the Statistical Bioinformatics department under the guidance of Paul Heinrich and Christian Kohler. This GUI prototype was further developed by Paul Heinrich and subsequently published to Bioconductor as the R package IsoCorrectoRGUI. Paul Heinrich developed IsoCorrectoR's algorithm for resolution-dependent correction and resolution-dependent MS/MS correction, and implemented this algorithm in the IsoCorrectoR R package. Christian Kohler maintains the IsoCorrectoR R packages at Bioconductor.

All analyses regarding the impact of correction and the validation of IsoCorrectoR have been performed by Paul Heinrich. The lab experiments for the experimental validation of IsoCorrectoR have been performed by Lisa Ellmann and Katja Dettmer from the Institute of Functional Genomics. Parts of the results presented here were published in Heinrich et al. (2018). They are marked accordingly if taken directly from the published manuscript.

# 3 Results and Discussion

## 3.1 IsoCorrectoR's algorithm

When using IsoCorrectoR to perform correction of stable isotope labeling data for natural stable isotope abundance (and tracer purity), the basic equation expressing the correction procedure is always the same, irrespective of the specific correction task or parameters used. This means that regardless of whether correction at low resolution, high resolution or ultra-high resolution is performed, or whether the data to be corrected are MS or MS/MS data, the correction can always be expressed through Equation II.1. Equation II.2 shows the same expression in matrix form. In order to be able to explain the algorithm behind IsoCorrectoR in more detail, the term labeling state needs to be defined at first. A labeling state is a (hypothetical) molecule species containing a specific number of tracer isotopes (*e.g.*, $^{13}C$) originating from the tracer substrate (*e.g.*, U-$^{13}C$-glutamine). All atoms except the tracer isotopes are isotopically undefined, *e.g.* for a given N atom it is not defined whether it is a $^{14}N$ or a $^{15}N$ isotope. An example for this is the following alanine labeling state with 1 $^{13}C$ tracer isotope: $^{13}CC_2H_6NO_2^-$ (anion, negative mode). The mass shift of a labeling state is defined as the mass shift that the given labeling state has with respect to the completely unlabeled labeling state of the same molecule. In the provided example, the (nominal) mass shift would be 1.

$$v_m = P \cdot v_c \tag{II.1}$$

$$\begin{pmatrix} v_{m_1} \\ \vdots \\ v_{m_i} \\ \vdots \\ v_{m_k} \end{pmatrix} = \begin{pmatrix} p_{11} & \cdots & p_{1j} & \cdots & p_{1k} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ p_{i1} & \cdots & p_{ij} & \cdots & p_{ik} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ p_{k1} & \cdots & p_{kj} & \cdots & p_{kk} \end{pmatrix} \cdot \begin{pmatrix} v_{c_1} \\ \vdots \\ v_{c_j} \\ \vdots \\ v_{c_k} \end{pmatrix} \tag{II.2}$$

In Equation II.1, $v_m$ is the result of the mass spectrometry measurement of the isotopologues of a given metabolite in a stable isotope labeling experiment. In the MS case, it is a vector that

contains, at a given index $i$, the integrated area at the mass shift associated with a labeling state (with index) $i$. The vector $v_c$ is the result of the correction procedure. A given entry $j$ corresponds to the corrected area of a labeling state (with index) $j$. In the MS/MS case, an element of $v_m$ at the index $i$ holds the area integrated for a transition with the product ion and neutral loss mass shifts associated with a given labeling state $i$. At a given index $j$, $v_c$ then holds the corrected area of a labeling state $j$, which is defined by a specific number of label in product ion and neutral loss.

$P$ is the probability or correction matrix. A single entry $p_{ij}$ in $P$ defines the fraction of the mass distribution of the $j$-th labeling state that corresponds to the mass shift (or mass shifts, in the MS/MS case) of the $i$-th measured value. Consequently, a given column $p_{*,j}$ of $P$ is a mass distribution of the labeling state $j$. Thus, to create the probability matrix $P$, the mass distributions of all $k$ labeling states of a molecule that can originate from tracer isotope incorporation must be computed. In the MS case, $k = l + 1$, where $l$ is the maximum number of label the metabolite can carry. Mass distributions arise because molecules are composed of varying combinations of naturally occurring isotopes, the frequency of individual isotopes being defined by their natural abundance (NA). The same is of course true for labeled species, at atom positions not already occupied by the tracer isotope. These isotope combinations and the incorporated tracer isotopes determine the species mass. If the tracer substrate is impure, the mass can additionally be modulated through the incorporation of "impure" tracer isotopes. See Figure II.2 for an illustration of a probability matrix and the labeling state indices $i$ and $j$ using the previously described alanine example molecule.

The definition of $p_{ij}$ can also be worded differently, as follows: The fraction of all isotopologues that can arise from a labeling state $j$ (due to natural isotope abundance and tracer impurity) and contribute to the measurement (at the mass shift) of a labeling state $i$. By solving Equation II.1 for $v_c$, the contributions of all other labeling states to a labeling state $i$ are removed. Further, the contributions of the labeling state $i$ to other labeling states are re-added to $i$ itself. As in IsoCor (Millard et al. (2012)), the system of linear equations arising from Equation II.1 is solved in a least-squares approach, with the constraint that the corrected values cannot be $< 0$. Without such a constraint, corrected values that are below 0 may occur due to inaccurate measurements in $v_m$, or isotope abundance values used for correction that do not match those actually present in the measured sample. Exactly how a given element $p_{ij}$ of $P$ is calculated depends strongly on the type of correction performed. Different approaches are used for low-resolution correction, resolution-dependent correction and ultra-high-resolution (multiple-tracer) correction. Further, MS- and MS/MS correction differ in calculation. In the following, the algorithmic approaches employed for computing $P$ for the different types of correction will be presented. See tables II.8 and II.9 for a summary of descriptions of the variables used throughout this section.

### 3.1.1 Low-resolution correction

In this context, low resolution means that the mass analyzer employed generally cannot resolve the incorporation of the tracer isotope (*e.g.*, $^{13}$C) into the metabolite from the incorporation of other isotopes with the same nominal mass shift (*e.g.*, $^{15}$N, $^{17}$O, $^{2}$H). This is usually the case for quadrupole, quadrupole ion trap, and TOF (time-of-flight) devices. However, high-resolving TOF devices can be at the borderline of resolving some of the isotopes. A more detailed discussion of this issue can be found in the section dealing with resolution-dependent

**Probability matrix *P* of alanine ($C_3H_6NO_2^-$)**

$^{13}CC_2H_6NO_2^-$ is **Ala_1**, the labeling state (with index) ***j***

$^{13}C_2CH_6NO_2^-$ is **Ala_2**, the labeling state (with index) ***i***

| | **Ala_0** | **Ala_1 (*j*)** | **Ala_2** | **Ala_3** |
|---|---|---|---|---|
| **Ala_0** | 0.9593 | 0 | 0 | 0 |
| **Ala_1** | 0.0360 | 0.9697 | 0 | 0 |
| **Ala_2 (*i*)** | 0.0044 | 0.0259 | 0.9802 | 0 |
| **Ala_3** | 0.0001 | 0.0042 | 0.0156 | 0.9908 |

**Blue column**: Mass distribution of ***j***

Matrix entry with **orange background**: Fraction of mass distribution of ***j*** that contributes to the measurement of ***i***

Figure II.2: ***Example illustration of the probability matrix P.*** *The figure shows an example probability matrix P for alanine, $C_3H_6NO_2^-$. Values were computed with IsoCorrectoR (low-resolution correction) and rounded to the fourth digit. The blue column is the mass distribution of the labeling state (with index) j, $^{13}CC_2H_6NO_2^-$. The brown row shows the contributions of the mass distributions of all alanine labeling states to the labeling state (with index) i, $^{13}C_2CH_6NO_2^-$. Please note that only the columns are mass distributions, the rows are not. The matrix entry with orange background corresponds to $p_{ij}$, which is the contribution of the mass distribution of labeling state j to the measurement of labeling state i (at low/nominal mass resolution). In this specific case, the indices $i = 2$ and $j = 1$ also correspond to the number of label incorporated in the respective labeling states. There are however cases where this does not hold true, for example when considering MS/MS correction.*

correction (3.1.4).

If measurements from a stable isotope labeling experiment were acquired at low resolution, the contributions of all naturally occurring stable isotopes of the elements in the given metabolite have to be corrected for. This means that a given entry $p_{ij}$ of $P$ contains contributions from all the isotopic combinations across the different elements in the labeling state $j$ that can produce the nominal mass shift of labeling state $i$. To determine $p_{ij}$, the first step is to find all possible isotopic combinations for the individual elements in the labeling state $j$. Lets assume that labeling state $j$ is the alanine labeling state from the previous example ($^{13}CC_2H_6NO_2^-$). The possible isotope combinations of $j$ are shown in Table II.1. Please note that the isotope combinations of the tracer element, C, are limited to 2 carbons, although there are in total 3 carbons in the molecule. This is because 1 carbon position is already occupied by the tracer isotope $^{13}$C in this labeling state. Also, only isotope combinations with a maximum mass shift of 3 are included, as combinations with a higher mass shift can not contribute to any other labeling state. When not considering tracer purity, this is already true for mass shifts $> 2$, however due to tracer impurity, mass shifts up to 3 become relevant.

Next, the probability and the (nominal) mass shift of the isotope combinations has to be computed. The probability is calculated as suggested in van Winden et al. (2002). Let $f_{h,q}$ be the frequency of isotope $h$ in a given isotope combination $q$ and let $r$ be the total number of different isotopes of the given element. Further, let $a_h$ be the natural abundance of isotope $h$ and let $n_q$ be the total number of isotope positions that can be assigned to an isotope. The probability of obtaining a certain isotope combination $q = [f_{1,q}, \ldots, f_{r,q}]$ when assigning $n_q$ isotope positions with replacement and without considering ordering is a multinomial probability (binomial in the special case of $r = 2$). Thus, the isotope combination probability $\theta_p(q)$ of an isotope combination $q$ of a given element can be expressed as shown in Equation II.3. The mass shift $\theta_m(q)$ of an isotope combination is simply the sum of the mass shifts $\mu_h$ of the isotopes in the combination with respect to the most abundant isotope (*e.g.*, for C, $^{12}$C is the most abundant isotope, so $\mu_h$ for $^{13}$C is $13 - 12 = 1$).

$$\theta_p(q) = n_q! \cdot \prod_h \frac{{a_h}^{f_{h,q}}}{(f_{h,q})!}, \qquad \theta_m(q) = \sum_h f_{h,q} \cdot \mu_h \tag{II.3}$$

The probabilities and mass shifts in the isotope combination table (Table II.1) have been computed according to the formulae in Equation II.3. In a next step, the isotope combinations have to be combined to element combinations to yield a completely defined isotopologue of the labeling state. Table II.2 shows exemplary element combinations of the alanine labeling state $^{13}CC_2H_6NO_2^-$. Any combination of isotope combinations of the different elements yields an element combination $u$ (*i.e.*, an isotopologue). The isotope combination probability of the isotope combination of a given element $e$ in $u$ will be written as $\theta_p(u, e)$. The element combination probability $\omega_p(u)$ of a specific element combination $u$ consisting of those individual isotope combinations can then be expressed as shown in Equation II.4. The mass shift $\omega_m(u)$ of a given element combination $u$ is simply the sum of the mass shifts of its isotope combinations $\theta_m(u, e)$ plus the intrinsic nominal mass shift of the labeling state $j$, $m_j$.

$$\omega_p(u) = \prod_e \theta_p(u, e), \qquad \omega_m(u) = \sum_e \theta_m(u, e) + m_j \tag{II.4}$$

| **Isotope combination $q$, tracer C** | **Frequency $^{12}C$ ($f_{1,q}$)** | **Frequency $^{13}C$ ($f_{2,q}$)** | **Probability $\theta_p(q)$** | **Mass Shift (nominal) $\theta_m(q)$** |
|---|---|---|---|---|
| C1 | 2 | 0 | 9.787E-01 | 0 |
| C2 | 1 | 1 | 2.117E-02 | 1 |
| C3 | 0 | 2 | 1.145E-04 | 2 |
| **Isotope combination $q$, tracer purity C** | **Frequency non-tracer $^{12}C$ ($f_{1,q}$)** | **Frequency tracer $^{13}C$ ($f_{2,q}$)** | **Probability $\theta_p(q)$** | **Mass Shift (nominal) $\theta_m(q)$** |
| pur1 | 0 | 1 | 9.900E-01 | 0 |
| pur2 | 1 | 0 | 1.000E-02 | -1 |
| **Isotope combination $q$, N** | **Frequency $^{14}N$ ($f_{1,q}$)** | **Frequency $^{15}N$ ($f_{2,q}$)** | **Probability $\theta_p(q)$** | **Mass Shift (nominal) $\theta_m(q)$** |
| N1 | 1 | 0 | 9.963E-01 | 0 |
| N2 | 0 | 1 | 3.680E-03 | 1 |
| **Isotope combination $q$, H** | **Frequency $^{1}H$ ($f_{1,q}$)** | **Frequency $^{2}H$ ($f_{2,q}$)** | **Probability $\theta_p(q)$** | **Mass Shift (nominal) $\theta_m(q)$** |
| H1 | 6 | 0 | 9.993E-01 | 0 |
| H2 | 5 | 1 | 6.896E-04 | 1 |
| H3 | 4 | 2 | 1.983E-07 | 2 |
| H4 | 3 | 3 | 3.041E-11 | 3 |

| **Isotope combination $q$, O** | **Frequency $^{16}O$ ($f_{1,q}$)** | **Frequency $^{17}O$ ($f_{2,q}$)** | **Frequency $^{18}O$ ($f_{3,q}$)** | **Probability $\theta_p(q)$** | **Mass Shift (nominal) $\theta_m(q)$** |
|---|---|---|---|---|---|
| O1 | 2 | 0 | 0 | 9.951E-01 | 0 |
| O2 | 1 | 0 | 1 | 4.090E-03 | 2 |
| O3 | 1 | 1 | 0 | 7.582E-04 | 1 |
| O4 | 0 | 1 | 1 | 1.558E-06 | 3 |
| O5 | 0 | 2 | 0 | 1.444E-07 | 2 |

Table II.1: ***Isotope combinations of the different elements of the alanine labeling state $^{13}CC_2H_6NO_2^-$, including tracer purity combinations.*** *For each element, the table contains the possible isotope combinations together with their isotope count, isotope combination probability $\theta_p$ and isotope combination mass shift $\theta_m$. $\theta_p$ and $\theta_m$ are shown as computed by IsoCorrectoR, with digits after the decimal point of $\theta_p$ limited to 3 for readability. Only isotope combinations with a maximum mass shift of 3 are included, as combinations with a higher mass shift can not contribute to any other labeling state. In the tracer isotope combinations, the intrinsic mass shift of the labeling state is not included.*

| **Element Combination $u$** | **Isotopologue Molecular Formula** | **Probability $\omega_p(u)$** | **Mass Shift (nominal) $\omega_m(u)$** |
|---|---|---|---|
| Ele1 | $^{13}C^{12}C_2{}^{1}H_6{}^{14}N^{16}O_2^-$ | $\theta_p(C1)\cdot\theta_p(N1)\cdot$ $\theta_p(H1)\cdot\theta_p(O1)$ | $\theta_m(C1)+\theta_m(N1)+$ $\theta_m(H1)+\theta_m(O1)\ +m_j=1$ |
| Ele2 | $^{13}C^{13}C^{12}C^{1}H_6{}^{15}N^{16}O_2^-$ | $\theta_p(C2)\cdot\theta_p(N2)\cdot$ $\theta_p(H1)\cdot\theta_p(O1)$ | $\theta_m(C2)+\theta_m(N2)+$ $\theta_m(H1)+\theta_m(O1)\ +m_j=3$ |
| Ele3 | $^{13}C^{12}C_2{}^{1}H_6{}^{14}N^{16}O^{18}O^-$ | $\theta_p(C1)\cdot\theta_p(N1)\cdot$ $\theta_p(H1)\cdot\theta_p(O2)$ | $\theta_m(C1)+\theta_m(N1)+$ $\theta_m(H1)+\theta_m(O2)\ +m_j=3$ |
| Ele2_pur1 | $^{13}C^{13}C^{12}C^{1}H_6{}^{15}N^{16}O_2^-$ | $\theta_p(C2)\cdot\theta_p(N2)\cdot$ $\theta_p(H1)\cdot\theta_p(O1)\cdot$ $\theta_p(pur1)$ | $\theta_m(C2)+\theta_m(N2)+$ $\theta_m(H1)+\theta_m(O1)+$ $\theta_m(pur1)+m_j=3$ |
| Ele2_pur2 | $^{12}C^{13}C^{12}C^{1}H_6{}^{15}N^{16}O_2^-$ | $\theta_p(C2)\cdot\theta_p(N2)\cdot$ $\theta_p(H1)\cdot\theta_p(O1)\cdot$ $\theta_p(pur2)$ | $\theta_m(C2)+\theta_m(N2)+$ $\theta_m(H1)+\theta_m(O1)+$ $\theta_m(pur2)+m_j=2$ |

Table II.2: ***Exemplary element combinations of the alanine labeling state $^{13}CC_2H_6NO_2{}^-$, including element combinations accounting for tracer purity.*** *The table includes 3 exemplary element combinations not accounting for tracer purity: Ele1, Ele2 and Ele3. The tracer isotope of the labeling state (1 $^{13}C$) is marked in blue, all natural abundance derived isotopes are shown uncolored. The columns $\omega_p$ and $\omega_m$ show how element combination probability and mass shift are computed for the given element combination. Each element combination consists of specific isotope combinations (e.g., C1, N2, H1). Details for those isotope combinations are given in Table II.1. The intrinsic mass shift of the labeling state, $m_j$, is included in the computation of the element combination mass shifts. An additional 2 element combinations including tracer purity are given, Ele2_pur1 and Ele2_pur2. They are both derived from the element combination Ele2. The isotopic composition of Ele2_pur1 is identical to Ele2, as the tracer is not impure in the tracer purity combination pur1 (see Table II.1). However, tracer purity combination probability and mass shift are added to the calculation of $\omega_p$ and $\omega_m$, respectively. Ele2_pur2 contains the tracer purity combination pur2, where the $^{13}C$ label is impure and thus replaced by a $^{12}C$, marked in red.*

| | Ala_0 | Ala_1 | Ala_2 | Ala_3 |
|---|---|---|---|---|
| Ala_0 | 9.593E-01 | 0.000E+00 | 0.000E+00 | 0.000E+00 |
| Ala_1 | 3.606E-02 | 9.697E-01 | 0.000E+00 | 0.000E+00 |
| Ala_2 | 4.446E-03 | 2.597E-02 | 9.802E-01 | 0.000E+00 |
| Ala_3 | 1.499E-04 | 4.213E-03 | 1.565E-02 | 9.908E-01 |

Table II.3: ***Probability matrix $P$ of the alanine example $C_3H_6NO_2^-$, as computed by IsoCorrectoR.*** *Column names and row names of the matrix indicate the alanine labeling states, where Ala_x is the labeling state containing x $^{13}C$ tracer isotopes. Digits after the decimal point are limited to 3 for readability.*

| **Isotope** | **Abundance** |
|---|---|
| $^{12}C$ | 0.9893 |
| $^{13}C$ | 0.0107 |
| $^{1}H$ | 0.999885 |
| $^{2}H$ | 0.000115 |
| $^{16}O$ | 0.99757 |
| $^{17}O$ | 0.00038 |
| $^{18}O$ | 0.00205 |
| $^{14}N$ | 0.99632 |
| $^{15}N$ | 0.00368 |

Table II.4: ***Isotope abundance values.*** *Isotope abundance values (taken from Rosman & Taylor (1998)) used in the example calculations in this section.*

To yield all possible element combinations for a given labeling state $j$, the isotope combinations of all elements have to be combined combinatorically (*i.e.*, constructing the cartesian product from the isotope combination sets of the different elements). This yields the set of all element combinations of labeling state $j$, $U_j$. Now, $p_{ij}$ is the sum of the element combination probabilities $\omega_p(u)$ of labeling state $j$ where the element combination mass shift $\omega_m(u)$ is equal to the intrinsic nominal mass shift of labeling state $i$ (Equation II.5). The subset of element combinations satisfying this condition is $U_{ij}$. For example, as shown in Table II.2, the element combinations Ele2 and Ele3 of the labeling state $^{13}CC_2H_6NO_2^-$ both have a nominal mass shift of 3. Their element combination probabilities thus have to be added up when computing the probability that $^{13}CC_2H_6NO_2^-$ contributes to the labeling state $^{13}C_3H_6NO_2^-$. When all $p_{ij}$ have been computed, the result is the final probability matrix $P$. For the alanine example ($C_3H_6NO_2^-$), $P$ is shown in Table II.3, as computed by IsoCorrectoR when not accounting for tracer purity. As can be seen, the probability that a labeling state contributes to another labeling state with less label is always 0, making $P$ a lower triangular matrix.

$$U_{ij} = \{u \in U_j | \omega_m(u) = m_i\}, \qquad p_{ij} = \sum \{\omega_p(u) | u \in U_{ij}\} \tag{II.5}$$

### 3.1.2 Tracer purity correction

Due to tracer impurity, the tracer isotope positions of the labeling state may contain non-tracer isotopes instead of tracer isotopes. This is because impure sections of the tracer substrate can

be incorporated into the measured metabolite. As a consequence of tracer impurity, a labeling state $j$ can contribute to the measurement of a labeling state $i$ that has a lower mass shift than $j$. If correction for tracer purity is performed, the purity state is simply considered a "special" isotope combination. The possible isotopes of this isotope combination are the tracer isotope (*e.g.*, $^{13}C$) and the most abundant non-tracer isotope (*e.g.*, $^{12}C$) of the given tracer element. The isotope abundance values $a_h$ are the purity of the tracer (0.99, in this example) and the probability of having a non-tracer isotope at a tracer isotope position due to impurity (1 - purity). The mass shift of the tracer isotope in a tracer purity isotope combination is defined as 0 - for a given element combination, the mass shift associated with label incorporation is already provided by the intrinsic mass shift of the labeling state (see Table II.2), and needs not be added via a tracer purity isotope combination. The mass shift of the non-tracer isotope species is the negative of the tracer isotope mass shift relative to the most abundant isotope (*e.g.*, -1 in the case of a $^{13}C$ tracer). Thus, impurity leads to a decrease in mass shift. For $n_q$, the total number of tracer isotopes in the given labeling state is used (1 in the provided example of alanine). See Table II.1 for the tracer purity isotope combinations of $^{13}CC_2H_6NO_2^-$. When computing the set of element combinations of a given labeling state, the tracer purity isotope combinations are treated just as isotope combinations of real elements. Thus, they are combinatorically combined with the isotope combinations of all other elements to yield element combination probabilities and mass shifts that account for tracer purity. See Table II.2 for example element combinations including tracer purity and Table II.5 for a probability matrix $P$ of the alanine example that accounts for tracer purity. When considering the matrix, it becomes evident that it is completely filled with non-zero values and not a lower triangular matrix. The probability matrix not including tracer purity (see Table II.3), on the other hand, is a lower triangular matrix. This difference is observed because a given labeling state can contribute to labeling states with less label incorporated when tracer purity is accounted for.

As stated, tracer purity isotope combinations consider only two isotopic states: tracer and non-tracer. Clearly, a tracer element can have more than two isotopes (*e.g.*, when using $^{18}O$ as a tracer). In this case, the tracer purity correction may not be fully correct, because not only the most abundant isotope of the tracer element (*e.g.*, $^{16}O$) but also other, less abundant isotopes (*e.g.*, $^{17}O$) can be incorporated instead of the tracer isotope due to impurity. This is not accounted for in the presented purity model. However, manufacturer information on tracer purity usually is not as detailed as providing abundances of individual tracer element isotopes in the tracer substrate. Rather, only the abundance of the tracer isotope itself is given as a percentage, reflecting the probability that a given tracer element position in the tracer substrate will be labeled - the tracer purity value. Furthermore, the most commonly used tracer elements C and N only have 2 stable isotopes.

### 3.1.3 Low-resolution MS/MS correction

MS/MS data where label can occur both in product ion and neutral loss has to be corrected using an MS/MS correction algorithm. When continuing with the alanine example from section 3.1.1, a 2-$^{13}CC_2H_6NO_2^-$ precursor labeling state might fragment to yield the $^{13}CCH_6N^-$ and $CO_2$ product ion and neutral loss labeling states (while a 1-$^{13}CC_2H_6NO_2^-$ precursor would fragment into the $C_2H_6N^-$ and $^{13}CO_2$ labeling states under the same conditions). Figure II.3 illustrates the principle of tandem MS with a labeled molecule and its implications for the correction procedure. The actual labeled species is an alanine isotopologue with 3 $^{13}C$, which

| | Ala_0 | Ala_1 | Ala_2 | Ala_3 |
|---|---|---|---|---|
| Ala_0 | 9.593E-01 | 9.697E-03 | 9.802E-05 | 9.908E-07 |
| Ala_1 | 3.606E-02 | 9.603E-01 | 1.941E-02 | 2.943E-04 |
| Ala_2 | 4.446E-03 | 2.575E-02 | 9.610E-01 | 2.913E-02 |
| Ala_3 | 1.499E-04 | 4.172E-03 | 1.541E-02 | 9.615E-01 |

Table II.5: ***Probability matrix $P$ of the alanine example $C_3H_6NO_2^-$, as computed by IsoCorrectoR when accounting for tracer purity (0.99).*** *Column names and row names of the matrix indicate the alanine labeling states, where Ala_x is the labeling state containing x $^{13}C$ tracer isotopes. Digits after the decimal point are limited to 3 for readability.*

corresponds to a mass shift of 3 in the precursor, 2 in the product ion and 1 in the neutral loss (compared to the unlabeled species). As shown in the figure, NA contributions which share the same nominal precursor mass shift of 3 cannot be resolved by the precursor mass analyzer. In the MS case, all these species would contribute to the mass spectrometric peak of the actual labeled species and would have to be corrected for. However, in the MS/MS case, some of the species can be resolved at the product ion stage, because they have differing nominal product ion mass shifts. Thus, only NA contributions where both the precursor and the product ion mass shift matches that of the labeled species have to be corrected for in MS/MS NA correction. As a consequence, performing MS correction on MS/MS labeling data (using the precursor molecular formula) would lead to overcorrection.

In a first step of MS/MS correction, the mass distributions of all product ion and neutral loss labeling states of the metabolite are computed just as described for the MS case. This yields the individual probability matrices $P_{prod}$ and $P_{loss}$ for product ion and neutral loss, respectively. For correcting MS/MS data according to Equation II.1, $P_{prod}$ and $P_{loss}$ cannot be used directly, however. Rather, a specific probability matrix $P_{MS/MS}$ is required for this purpose. It must hold, at a given entry $p_{ij}$, the probability that a given MS/MS labeling state $j$ - defined by its label in both product ion and neutral loss - is found at the product ion and neutral loss mass shifts of an MS/MS labeling state $i$.

In the case of low-resolution MS/MS correction, an entry $p_{ij}$ of $P_{MS/MS}$ is derived by multiplying the probability that the product ion labeling state $j_{prod}$ of $j$ will produce the mass shift of the product ion labeling state $i_{prod}$ of $i$ (probability derived from $P_{prod}$) with the probability that the neutral loss labeling state $j_{loss}$ of $j$ will produce the mass shift of the neutral loss labeling state $i_{loss}$ of $i$ (probability derived from $P_{loss}$). This is expressed in Equation II.6. In MS/MS correction, the maximum index $k$ of $P$ (*i.e.*, $P_{MS/MS}$) is $k = (l_{prod} + 1) \cdot (l_{loss} + 1)$ and corresponds to the total number of MS/MS labeling states that can originate, given the fragmentation pattern of the metabolite. Here, $l_{prod}$ is the maximum number of label in the product ion and $l_{loss}$ is the maximum number of label in the neutral loss. Consequently, for the alanine example with fragments $C_2H_6N^-$ and $CO_2$ we would expect $k = (2+1) \cdot (1+1) = 6$ MS/MS labeling states. Table II.6 shows the probability matrix $P$ for the alanine MS/MS example, as computed by IsoCorrectoR when not accounting for tracer purity.

$$P_{MS/MS\,ij} = P_{prod\,i_{prod}\,j_{prod}} \cdot P_{loss\,i_{loss}\,j_{loss}} \tag{II.6}$$

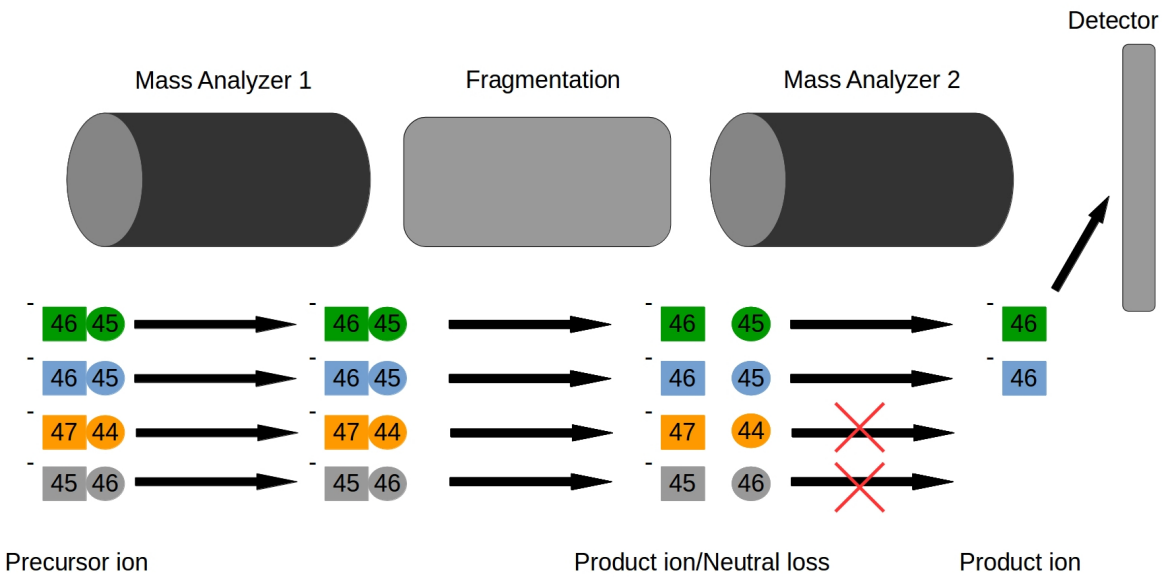



Figure II.3: ***Illustration of tandem MS of a labeled molecule with a low-resolution instrument.*** *The figure shows the schematic setup of a mass spectrometer capable of tandem MS (triple quadrupole setup). It contains a mass analyzer that can filter precursor ions (Mass Analyzer 1), a means for the fragmentation of precursor ions, a mass analyzer to select product ions (Mass Analyzer 2) and a detector. The green, blue, orange and gray symbols above the label "Precursor ion" are isotopologues of an alanine molecule: $C_3H_6NO_2^-$, nominal $m/z$ of unlabeled species = 88. All isotopologues share the same nominal mass shift of 3 in the precursor (nominal $m/z$ = 91), so that they cannot be resolved from each other by Mass Analyzer 1. When the isotopologues get fragmented, the rectangle is the product ion portion ($C_2CH_6N^-$), while the circle is the neutral loss portion ($CO_2$). The numbers inside the symbols are the nominal $m/z$ values (or mass, in the case of neutral loss) of the respective fragment. The green (first) precursor is the actual labeled species $i$ - the only isotope of higher mass it contains is the tracer isotope $^{13}C$. The other species $u_1, u_2, u_3$ (blue, orange, gray) may also contain other isotopes of higher mass. The arrows indicate passage through a mass analyzer to the next stage or fragmentation. A red cross across an arrow means that passage through the given mass analyzer is not possible for the given species, because it can be resolved from $i$/the product ion of $i$. While Mass Analyzer 1 could not resolve the 4 species due to equal nominal precursor $m/z$, Mass Analyzer 2 can resolve the product ions of the orange and gray species from the product ion of $i$, because they differ in nominal $m/z$. However, the product ion of the blue species has the same nominal $m/z$ as the product ion of $i$ and thus can not be resolved. As a consequence, the green and the blue species pass through to the detector and both contribute to the mass spectrometric peak expected for labeling state $i$. Thus, a MS/MS correction algorithm must correct for the contribution of $u_1$ to $i$, but not for the contribution of $u_2$ and $u_3$ to $i$, because they can be resolved at the product ion stage. This is in contrast to MS correction, where the contribution of all species would be corrected for.*

| | Ala_0.0 | Ala_1.0 | Ala_1.1 | Ala_2.1 | Ala_2.2 | Ala_3.2 |
|---|---|---|---|---|---|---|
| Ala_0.0 | 9.593E-01 | 0.000E+00 | 0.000E+00 | 0.000E+00 | 0.000E+00 | 0.000E+00 |
| Ala_1.0 | 1.111E-02 | 9.697E-01 | 0.000E+00 | 0.000E+00 | 0.000E+00 | 0.000E+00 |
| Ala_1.1 | 2.496E-02 | 0.000E+00 | 9.697E-01 | 0.000E+00 | 0.000E+00 | 0.000E+00 |
| Ala_2.1 | 2.889E-04 | 2.523E-02 | 1.123E-02 | 9.802E-01 | 0.000E+00 | 0.000E+00 |
| Ala_2.2 | 2.058E-04 | 0.000E+00 | 1.474E-02 | 0.000E+00 | 9.802E-01 | 0.000E+00 |
| Ala_3.2 | 2.383E-06 | 2.081E-04 | 1.706E-04 | 1.490E-02 | 1.135E-02 | 9.908E-01 |

Table II.6: ***Probability matrix $P$ of the alanine MS/MS example (product ion: $C_2H_6N^-$, neutral loss: $CO_2$), as computed by IsoCorrectoR when not accounting for tracer purity.*** *Column names and row names of the matrix indicate the alanine MS/MS labeling states, where Ala_x.y is the labeling state containing x $^{13}C$ in the precursor and y $^{13}C$ in the product ion. Digits after the decimal point are limited to 3 for readability.*

### 3.1.4 Resolution-dependent correction

Low-resolution correction is adequate for mass analyzers that cannot distinguish isotopes that introduce the same nominal mass shift into the molecule (*e.g.*, $^{13}C$ and $^{15}N$). However, high-resolving orbitrap and FT-ICR (fourier-transform ion cyclotron resonance) mass analyzers are capable of resolving some or even almost all isotopes that share the same nominal mass shift, based on the subtle mass defect related differences. For example, the mass difference between $^{13}C^{12}C_2{}^1H_6{}^{14}N^{16}O_2{}^-$ and $^{12}C_3{}^1H_6{}^{15}N^{16}O_2{}^-$ is the difference between the exact mass shift of $^{13}C$ and the exact mass shift of $^{15}N$: $1.003355 - 0.997035 = 0.00632$. Thus, at a nominal $m/z$ of 89, a minimal resolution of about 15000 would already be sufficient to spectrometrically resolve the two isotopologues. This is easily obtained by virtually all orbitrap and FT-ICR devices, and even by some TOF mass analyzers. The absolute mass difference between incorporating $^{13}C$ and $^{17}O$, on the other hand, is much smaller: 0.000861. Here, a minimal resolution of roughly 105000 at 89 $m/z$ would be required. It must be noted, though, that also mass accuracy and the $m/z$ window used for peak extraction play an important role in that respect. Further, with increasing metabolite $m/z$, orbitrap and FT-ICR resolution diminishes (Zubarev & Makarov (2013)). As a consequence, some isotopic contributions may be resolved for molecules in the lower $m/z$ range, while they cannot be resolved for species with high $m/z$ values.

Clearly, being able to resolve some of the isotopic contributions spectrometrically has an impact on how correction for natural isotope abundance should be performed. If low-resolution correction is performed on high-resolution data, overcorrection will likely be the result. This is because isotopic contributions that are actually not part of the integrated isotopologue peaks due to the high instrument resolution are nevertheless removed during correction. Thus, a special approach to natural isotope abundance correction is needed in the case of high-resolution data. Solutions to this problem have recently been provided, however only the one implemented in IsoCor v2 (Millard et al. (2019)) appears to be fully correct, and IsoCorrectoR implements the approach of IsoCor v2.

The other available tools for high-resolution correction, AccuCor (Su et al. (2017), R) and ElemCor (Du et al. (2019), Matlab), both employ a similar algorithm based on matrix multiplication of individual element correction matrices. In this approach, a correction matrix is computed for each element in the metabolite of interest. These correction matrices are similar to the correction matrix $P$ from Equation II.2, however they only consider the contributions of

a single element (*e.g.*, N or O). Thus, a correction matrix $P_O$ for O would contain, at a given entry $p_{ij}$, the probability that a labeling state $j$ matches the mass shift of a labeling state $i$ **only** through the natural abundance of O, not considering any other elements. To compute the final correction matrix $P$ (see Equation II.7), matrix multiplication among the individual element correction matrices $P_e$ is performed. If tracer purity correction is desired, the resulting matrix is multiplied with a purity correction matrix in a last step (the multiplication is not commutative).

$$P = (\prod_{e} P_e) \cdot P_{purity} \tag{II.7}$$

When multiplying the correction matrices of all elements, the result is the correction matrix for low-resolution correction. It takes the isotopic contributions from all elements into account. In fact, this matrix multiplication approach to create the low-resolution correction matrix is implemented in the first version of IsoCor (IsoCor v1, Millard et al. (2012)). The result of this approach is equivalent to the result of IsoCorrectoR's approach to low-resolution correction described in 3.1.1. The rationale behind using the matrix multiplication approach in resolution-dependent correction is the following: If all isotopes of a given element $e$ can be resolved from the tracer isotope at the given resolution of the mass analyzer, this element correction matrix $P_e$ can be excluded from the calculation of $P$ and $P$ does not correct for natural isotope abundance contributions of $e$. Further, in the approach of AccuCor (Su et al. (2017)), if some of the isotopes of a given element $e$ can be resolved and some can not, the isotopes that can be resolved are not included when computing $P_e$, but $P_e$ is still used for calculating $P$. This is not entirely correct, as certain combinations of the isotopes $a$ and $b$ of $e$ may remain unresolved even if $a$ or $b$ can be resolved individually. This error is corrected in the algorithm used in ElemCor (Du et al. (2019)).

However, even when considering this improvement of ElemCor over AccuCor, both approaches remain flawed, as detailed in Millard et al. (2019). While individual isotopes may be resolved from the tracer isotope at a given resolution, combinations of different (and probably individually resolved) isotopes in an element combination may produce a mass shift that can not be resolved from that of the given labeling state. The reason for this is that the mass defects of the individual isotopes in an isotopologue (*i.e.*, an element combination) may cancel each other out. This potentially yields a lower mass difference of the given element combination to the considered labeling state than that of the individual (and potentially resolved) isotopes to the tracer isotope.

Rather than using matrix multiplication according to Equation II.7, each element combination of a given labeling state $j$ must be computed individually to obtain valid correction results, as described for IsoCorrectoR in Section 3.1.1. This way, the exact (not nominal) mass shift of each element combination can be calculated. In a next step, a given entry $p_{ij}$ of $P$ is obtained similarly to Equation II.5 (low resolution). However, rather than summing up the probability of all element combinations of $j$ where the mass shift matches that of labeling state $i$, a slightly different approach is used. In the resolution-dependent case, all element combinations $u$ of $j$ where the exact mass shift $\omega_m(u)$ is within a certain interval $\delta_{min}$ around the exact mass shift $m_i$ of labeling state $i$, are selected. As in the low-resolution case, the probabilities of the element combinations in the resulting set $U_{ij}$ are then summed up to yield $p_{ij}$ (Equation II.8). As shown in Equation II.9, the size of the interval $\delta_{min}$ is defined by the resolution of the mass analyzer at the $m/z$ of $i$ (given as $FWHM$, full width at half maximum of the

mass spectrometric peak) and by the charge $z$ of $i$ (Millard et al. (2019)). An additional factor of 1.66 makes sure that only isotopologues that should be well separated in the mass spectrum are excluded from the correction procedure. As previously stated, resolution is $m/z$-dependent for orbitrap and FT-ICR mass analyzers. In the case of FT-ICR analyzers, resolution is negatively proportional to $m/z$, whereas for orbitrap analyzers, resolution is negatively proportional to the square root of $m/z$ (Zubarev & Makarov (2013)). If a reference resolution $R_{ref}$ at a reference $m/z$ $m_{ref}$ is provided, the orbitrap $FWHM$ at a given metabolite $m/z$ can be computed according to Equation II.10 (Su et al. (2017)). Equation II.10 also shows a similar approach for FT-ICRs, based on the previously stated $m/z$ dependency of resolution. $FWHM$ may also be set to a constant value, according to $FWHM_{const} = \frac{m_{ref}}{R_{ref}}$. This may be useful if resolution-dependent correction should be performed with data from a TOF device, or if resolution is limited in the data analysis stage by the $m/z$ range associated with an extracted ion chromatogram.

$$U_{ij} = \{u \in U_j | \, (|\omega_m(u) - m_i| < \delta_{min})\}, \; p_{ij} = \sum \{\omega_p(u) | u \in U_{ij}\} \tag{II.8}$$

$$\delta_{min} = 1.66 \cdot FWHM \cdot |z| \tag{II.9}$$

$$FWHM_{orbitrap} = \frac{(m/z)^{3/2}}{R_{ref} \cdot \sqrt{m_{ref}}}, \quad FWHM_{ft-icr} = \frac{(m/z)^2}{R_{ref} \cdot m_{ref}} \tag{II.10}$$

The algorithm for resolution-dependent correction just described is implemented in both IsoCorrectoR and IsoCor v2. For more information about why the matrix multiplication approach to correction implemented in AccuCor and ElemCor is flawed and for a practical example illustrating error magnitude, please refer to the publication introducing IsoCor v2 (Millard et al. (2019)).

### 3.1.5 Resolution-dependent MS/MS correction

Many orbitrap instruments are capable of acquiring tandem mass spectra, and gathering the positional labeling information that such spectra provide can be highly beneficial, especially in metabolic flux analysis. To be able to properly correct such data for NA and tracer purity, it is likely that an approach to correct MS/MS data from stable isotope labeling experiments in a resolution-dependent manner will be required. Compared to resolution-dependent correction of MS data, not only 1 but 2 mass analyzers - namely, precursor and product ion mass analyzer, (which can, technically, be the same device) - can filter NA contributions to a given labeling state $i$ in the MS/MS case. This is illustrated in Figure II.4. For correction, it may appear straightforward to directly combine the approach for resolution-dependent MS correction (described in 3.1.4) with the approach to low-resolution MS/MS correction (section 3.1.3). When doing this, a first step would be to compute the individual product ion and neutral loss probability matrices $P_{prod}$ and $P_{loss}$ in a resolution dependent manner, just as described in section 3.1.4. For computing $P_{prod}$, $\delta_{min_{prod}}$ would be obtained by gathering the product ion mass analyzer (usually orbitrap) $FWHM$ at the $m/z$ of the product ion. For $P_{loss}$ the precursor mass analyzer (usually quadrupole or comparable) $FWHM$ at the precursor $m/z$ would analogously be used to calculate $\delta_{min_{prec}}$. Then, $P_{prod}$ and $P_{loss}$ could be combined as

described in section 3.1.3 to obtain $P$ and ultimately correct the data. Bear in mind, however, that $P_{prod}$ holds the probabilities $p_{i_{prod}\, j_{prod}}$ that the product ion labeling state $j_{prod}$ of the MS/MS labeling state $j$ produces an exact mass shift within the interval $\delta_{min_{prod}}$ (product ion mass analyzer) around the exact mass shift of the product ion labeling state $i_{prod}$ of MS/MS labeling state $i$. Accordingly, $P_{loss}$ holds the probabilities $p_{i_{loss}\, j_{loss}}$ that the neutral loss labeling state $j_{loss}$ of the MS/MS labeling state $j$ produces an exact mass shift within the interval $\delta_{min_{prec}}$ (precursor mass analyzer) around the exact mass shift of the neutral loss labeling state $i_{loss}$ of MS/MS labeling state $i$. This is where a fundamental problem arises. The mass shift of a given neutral loss isotopologue species derived from the neutral loss labeling state $j_{loss}$ is not the mass shift that is recognized by the precursor mass analyzer. The mass shift actually present at the mass analyzer is the mass shift of the precursor ion $u$ (*i.e.*, a given isotopologue derived from MS/MS labeling state $j$ as a whole), which is the sum of the mass shifts of product ion and neutral loss. As a direct consequence, also mass defect related differences in product ion and neutral loss add up in the precursor mass shift.

This results in two situations where the discussed approach fails. The first case leads to overcorrection. The neutral loss mass shift $\omega_m(u_{loss})$ produced by a given isotopologue $u_{loss}$ of the neutral loss labeling state $j_{loss}$ may be unresolved from the mass shift $m_{i_{loss}}$ of neutral loss labeling state $i_{loss}$ by the precursor mass analyzer. However, technically, the mass analyzer does not distinguish between $\omega_m(u_{loss})$ and $m_{i_{loss}}$, but rather between $\omega_m(u)$ and $m_i$ (the mass shift of MS/MS labeling state $i$), where $\omega_m(u) = \omega_m(u_{prod}) + \omega_m(u_{loss})$. When $|\omega_m(u_{loss}) - m_{i_{loss}}| < \delta_{min_{prec}}$, this does not necessarily mean that $|\omega_m(u) - m_i| < \delta_{min_{prec}}$. The mass defect related differences $|\omega_m(u_{prod}) - m_{i_{prod}}|$ of the product ion and $|\omega_m(u_{loss}) - m_{i_{loss}}|$ of the neutral loss add up in $\omega_m(u)$ so that they may exceed the interval determined by $\delta_{min_{prec}}$. With the approach to resolution-dependent MS/MS correction discussed, this would mean that we may include certain isotopologues $u$ resulting from MS/MS labeling state $j$ in the correction of $i$ because $|\omega_m(u_{prod}) - m_{i_{prod}}| < \delta_{min_{prod}}$ and $|\omega_m(u_{loss}) - m_{i_{loss}}| < \delta_{min_{prec}}$, when actually $|\omega_m(u) - m_i| > \delta_{min_{prec}}$ - which means that the precursor mass analyzer can resolve MS/MS labeling state $i$ from $u$ of $j$. This results in overcorrection, as $u$ does not contribute to $i$ in the mass spectrum and should not be corrected for.

The following example illustrates this situation. Consider trimethylsilyl - (TMS-) derivatized alanine: Assume that the precursor $C_9H_{24}O_2NSi_2^+$ fragments into the product ion $C_5H_{14}NSi^+$ containing 2 C from alanine and the neutral loss $C_4H_{10}O_2Si$ containing 1 C from alanine. We will look at the contributions of the MS/MS labeling state $j$, $^{13}CC_4H_{14}NSi^+/C_4H_{10}O_2Si$, to the MS/MS labeling state $i$, $^{13}C_2C_3H_{14}NSi^+/^{13}CC_3H_{10}O_2Si$ (product ion/neutral loss). An isotopologue $u$ resulting from $j$ that may contribute to $i$ is $^{13}C^{12}C_4{}^1H_{14}{}^{15}N^{28}Si^+/^{12}C_4{}^1H_{10}{}^{16}O_2{}^{29}Si$, having an exact mass shift of $\omega_m(u) = 2.99995786$ compared to the completely unlabeled molecule, resulting from 1 $^{13}C$ and 1 $^{15}N$ in the product ion ($\omega_m(u_{prod}) = 2.00038973$) and 1 $^{29}Si$ in the neutral loss ($\omega_m(u_{loss}) = 0.99956813$). The mass shifts associated with MS/MS labeling state $i$ are $m_i = 3.010064505$, $m_{i_{prod}} = 2.00670967$, $m_{i_{loss}} = 1.003354835$. Assume now, that $\delta_{min_{prec}} = \delta_{min_{prod}} = 0.008$. The mass shift difference between $u_{prod}$ and $i_{prod}$, $|\omega_m(u_{prod}) - m_{i_{prod}}| = 0.00631994$ is smaller than $\delta_{min_{prod}}$ and thus can not be resolved. The same is true for the mass difference between $u_{loss}$ and $i_{loss}$, $|\omega_m(u_{loss}) - m_{i_{loss}}| = 0.003786705$. Thus, according to the algorithm presented, $i$ would be corrected for the contribution of $u$ from $j$ because neither product ion nor neutral loss can be resolved. However, when considering the mass difference the precursor mass analyzer is actually confronted with, $|\omega_m(u) - m_i| = 0.010106645$, we can see that this actually exceeds $\delta_{min_{prec}}$. Thus, $u$ from $j$ can actually be resolved from $i$ because the mass difference between precursors is large enough for the mass

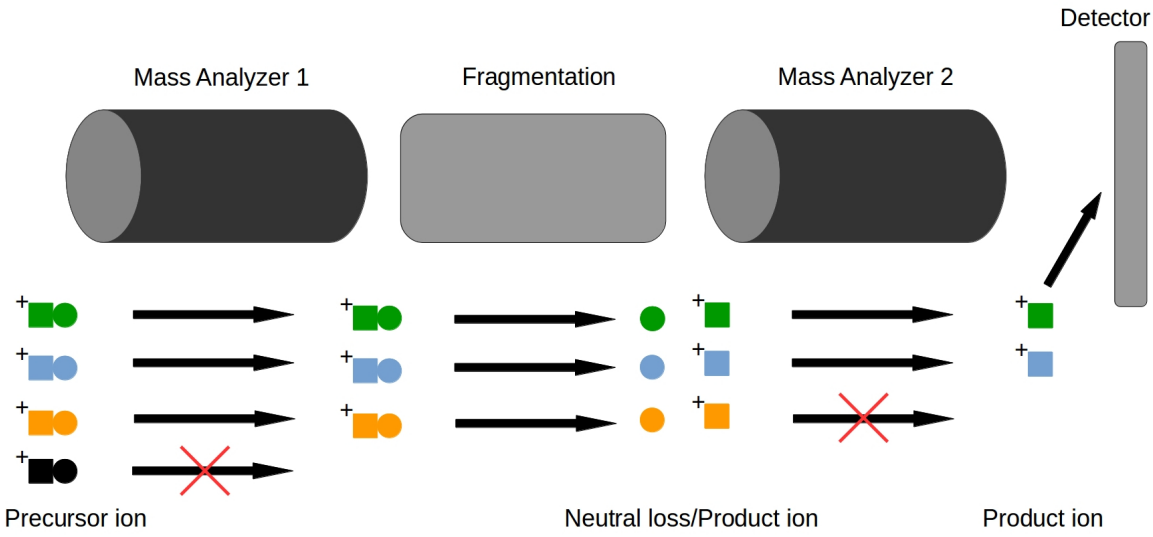



Figure II.4: ***Illustration of tandem MS of a labeled molecule with a high-resolution instrument.*** *The figure shows the schematic setup of a mass spectrometer capable of tandem MS. It contains a mass analyzer that can filter precursor ions (Mass Analyzer 1), a means for the fragmentation of precursor ions, a mass analyzer to select product ions (Mass Analyzer 2) and a detector. For the sake of better illustration, a triple quadrupole setup is shown, even though such devices can actually not attain high resolutions. The green, blue, orange and black symbols below are isotopologues of a hypothetic metabolite that all share the same nominal mass shift in product ion and neutral loss (and thus in the precursor). A low-resolution mass spectrometer would not be able to resolve them, the only mass differences are mass defect related. The rectangle is the product ion portion, the circle is the neutral loss portion. The green (first) precursor is the actual labeled species $i$ - the only isotopes of higher mass it contains are tracer isotopes (e.g., $^{13}C$). The other species $u_1, u_2, u_3$ (blue, orange, black) also contain other isotopes of higher mass. The arrows indicate passage through a mass analyzer to the next stage or fragmentation. A red cross across an arrow means that passage through the given mass analyzer is not possible for the given species, because it can be resolved from $i$/the product ion of $i$. The black species can already be resolved in the precursor isolation stage, because the resolution of mass analyzer 1 is sufficient. The blue and orange species pass through mass analyzer 1 together with the green species $i$ and all get fragmented into product ion and neutral loss. Mass analyzer 2 has sufficient resolution to distinguish the product ion of the orange species from the product ion of $i$. However, the mass difference to the product ion of the blue species is too low and can not be resolved. As a consequence, the green and the blue species pass through to the detector and both contribute to the mass spectrometric peak expected for labeling state $i$. Thus, a resolution-dependent MS/MS correction algorithm must correct for the contribution of $u_1$ to $i$, but not for the contribution of $u_2$ and $u_3$ to $i$, because they can be resolved.*

analyzer to distinguish - while the individual mass differences of product ion and neutral loss are not. This illustrates that product ion and neutral loss can not be considered separately, as in the low-resolution case, but that the mass shift of every precursor species $u$ that can result from an MS/MS labeling state $j$ has to be computed individually. Then it is necessary to assess whether it can be resolved from the precursor of $i$ by the precursor mass analyzer, or not. The previous example covered overcorrection, when the mass defect related differences of product ion and neutral loss add up to exceed $\delta_{min_{prec}}$. However, undercorrection is also possible if $u_{loss}$ of $j$ can be resolved from $i_{loss}$, but neither $u_{prod}$ can be resolved from $i_{prod}$ nor $u$ from $i$. Then the algorithm assumes that $u$ does not contribute to $i$ in the mass spectrum when it actually does. This happens if the mass defect related differences of product ion and neutral loss cancel each other out to fall below $\delta_{min_{prec}}$. An example for this can be found if we assume $\delta_{min_{prec}} = \delta_{min_{prod}} = 0.0035$ with the MS/MS labeling state $i$ from the previous example and a species $u$ $^{13}C^{12}C_4{}^2H^1H_{13}{}^{14}N^{28}Si^+/{}^{12}C_4{}^1H_{10}{}^{16}O_2{}^{29}Si$ (1 $^{13}C$ and 1 $^2H$ in the product ion, 1 $^{29}Si$ in the neutral loss) produced by $j$. Here, the mass difference actually present at the precursor mass analyzer is 0.0008647941 and thus clearly below $\delta_{min_{prec}}$, while the difference considering the neutral loss alone (0.003786705) is above $\delta_{min_{prec}}$, falsely indicating that the precursor mass analyzer can filter $u$ from $i$.

Given the current MS instrumentation available, the above examples are not too likely to occur - usually, a high-resolving orbitrap analyzer is used for assessing product ion $m/z$ and a low-resolving quadrupole analyzer for precursor isolation. High-resolving precursor mass analyzers are not commonly found. However, the discussed cases can only occur if the precursor isolation step can be performed at rather high resolution (low $\delta_{min_{prec}}$). Otherwise, if, as in the first example, the neutral loss mass difference is unresolved, the precursor mass difference will also be unresolved under all viable circumstances. This is because the addition of product ion and neutral loss mass defect related differences is usually by far not sufficient to exceed a fairly high value of $\delta_{min_{prec}}$ like 0.2 or greater. This is also the reason why the approach is valid for low-resolution MS/MS data.

Nevertheless, it is possible - to the best of our knowledge - to perform a resolution-dependent correction of MS/MS data that is valid under all circumstances. To this end, for a given MS/MS labeling state $j$, the product ion and neutral loss probabilities and exact mass shifts of all species $u_{prod}$ and $u_{loss}$ that can originate from $j$ have to be computed, as described previously. Then, all possible pairs of $u_{prod}$ and $u_{loss}$ have to be formed to obtain the corresponding precursors $u$ (the set of the $u$ formed is $U_j$). Then, the precursor mass shifts are calculated by adding product ion and neutral loss mass shift: $\omega_m(u) = \omega_m(u_{prod}) + \omega_m(u_{loss})$. Now, to compute the probability of contribution of $j$ to another MS/MS labeling state $i$, $p_{ij}$, it is first checked whether $|\omega_m(u_{prod}) - m_{i_{prod}}| < \delta_{min_{prod}}$. If this is the case, the product ions can not be resolved from each other. For the species $u_{prod}$ where this applies, it is additionally checked whether $|\omega_m(u) - m_i| < \delta_{min_{prec}}$ - *i.e.*, whether the precursor mass analyzer can resolve the precursor ions or not. Only species $u$ for which both checks are true, meaning that they can neither be resolved from the precursor $i$ nor from the product ion $i_{prod}$, are included in the set $U_{ij}$ (Equation II.11). This set is used to calculate the probability $p_{ij}$ by multiplying the probability of the product ion portion of each $u$ in $U_{ij}$, $\omega_p(u_{prod})$, with the probability of the corresponding neutral portion $\omega_p(u_{loss})$ and then summing over all products (Equation II.12). The algorithm just described is implemented in IsoCorrectoR. Figure II.5 shows a flowchart

of the IsoCorrectoR algorithm for the different modes of correction previously described.

$$U_{ij} = \{u \in U_j | \, (|\omega_m(u_{prod}) - m_{i_{prod}}| < \delta_{min_{prod}}) \; and \; (|\omega_m(u) - m_i| < \delta_{min_{prec}})\} \tag{II.11}$$

$$p_{ij} = \sum \{\omega_p(u_{prod}) \cdot \omega_p(u_{loss}) | u \in U_{ij}\} \tag{II.12}$$

### 3.1.6 Ultra-high-resolution correction and correction of data from multiple-tracer experiments

FT-ICR and orbitrap devices are capable of achieving ultra-high resolution (UHR) in the range of 400000 (400 $m/z$) or more, given sufficient acquisition time. At this level of resolution, it may be a viable simplification to assume that mass defect related differences between the incorporation of different isotopes with the same nominal mass shift into a molecule can generally be resolved. Following this assumption, it would not be necessary to correct ultra-high-resolution stable isotope labeling data for the NA of isotopes other than the tracer isotope itself, because these contributions would already be resolved spectrometrically (Carreer et al. (2013). This simplifies the correction procedure substantially, as only the NA of the tracer isotope and tracer purity have to be considered, but no other elements. Thus, to correct ultra-high-resolution data based on the simplification of (near) infinite resolution, all isotope combinations of the tracer element (*e.g.*, C or N) and all tracer purity combinations are computed for a given labeling state, just as in the low-resolution case (see Equation II.3). However, the isotope combinations of all other elements in the given molecule may be ignored. Also, in the case that the tracer element has more than 2 isotopes, isotope combinations containing non-tracer isotopes of higher mass (*e.g.*, $^{17}$O in the case of an $^{18}$O tracer) need not be computed. These, too, should be resolved by the mass analyzer under the given assumption. Further, instead of computing a mass shift for each combination, the total number of label in the combination is computed instead. Probabilities are calculated just as in the low-resolution approach. In a next step, all possible combinations of the tracer element isotope combinations and the tracer purity combinations are computed to obtain an UHR equivalent to the element combinations of the low-resolution case, together with their probabilities (Equation II.4). These UHR element combinations will be termed NA-purity combinations, as they do not actually contain different elements. The mass shift is again replaced by the total number of label in the combination (NA adds label, while impurity removes label). The set of all NA-purity combinations of a labeling state $j$ is $U_j$. Similarly to Equation II.5 for the low-resolution algorithm, NA-purity combinations of a given labeling state $j$ that have the same number of label as a labeling state $i$ are grouped into a set $U_{ij}$. The sum over this set then yields an entry $p_{ij}$ of the probability matrix.

As already stated in the chapter introduction, being able to reliably resolve mass defect related differences between the incorporation of isotopes that provide the same nominal mass shift yields the opportunity to perform stable isotope labeling experiments with multiple tracers. Then, *e.g.* $^{13}$C and $^{15}$N can be used simultaneously, as the peaks of the labeling states they produce can be resolved (Moseley (2010)). To perform NA and tracer purity correction on such multiple-tracer data, a corresponding multiple-tracer probability matrix $P_{mult}$ has to be computed. A single value $P_{mult\,ij}$ of that matrix is the probability that a given multi-tracer

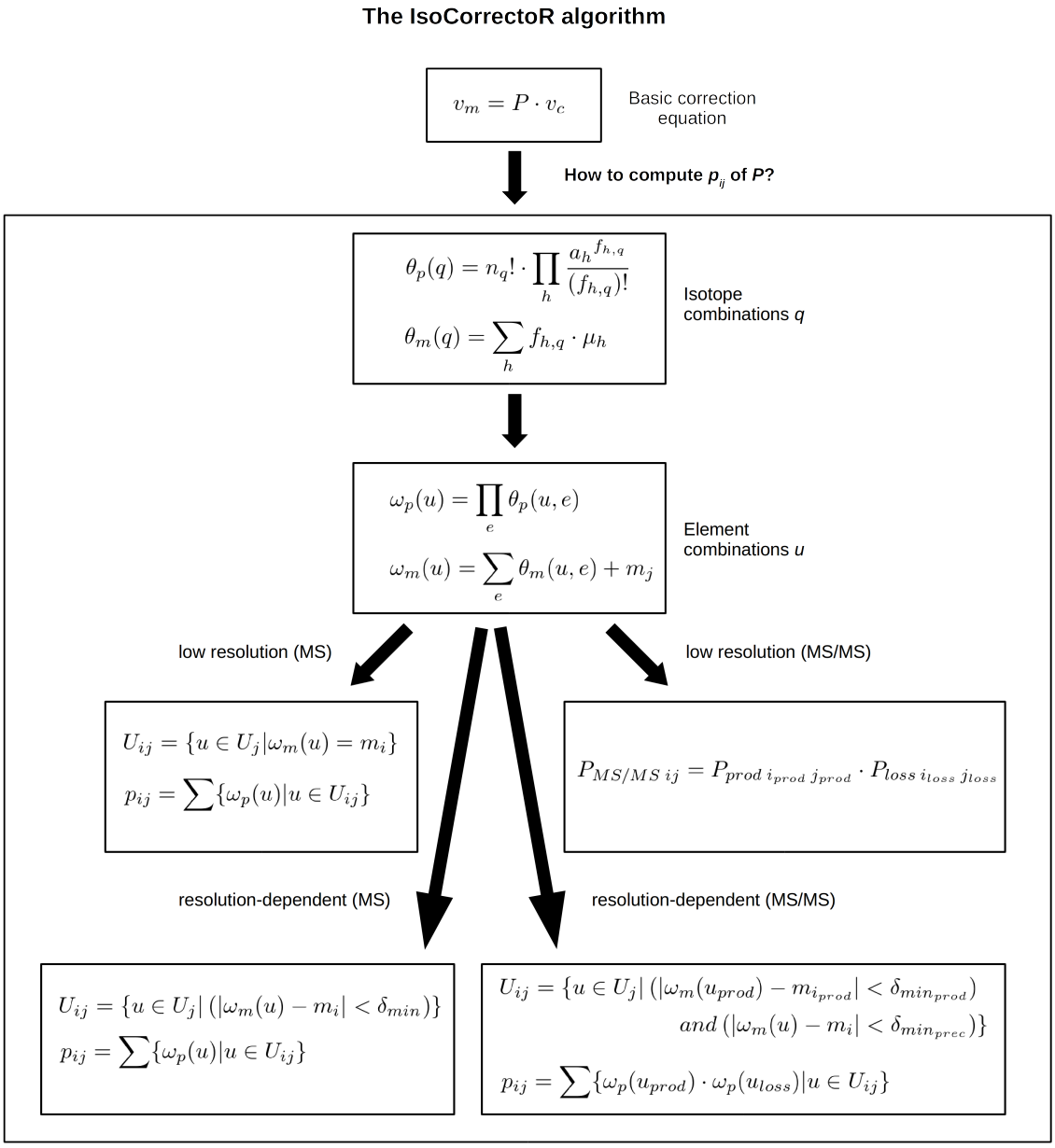



Figure II.5: ***Flowchart of the IsoCorrectoR algorithm.*** *A flowchart of the IsoCorrectoR algorithm that illustrates how the individual values $p_{ij}$ of the probability matrix $P$ are computed for different modes of correction: low-resolution MS and MS/MS correction and resolution-dependent MS and MS/MS correction. The algorithm generally starts by computing isotope combination probabilities $\theta_p(q)$ and mass shifts $\theta_m(q)$. In a next step, those values are used to calculate the probabilities $\omega_p(u)$ and mass shifts $\omega_m(u)$ of element combinations that can be formed from combinations of isotope combinations of the different elements in the molecule. Tracer purity correction is done by computing specific tracer purity isotope combinations. Thus, this mode of correction is covered in the "Isotope combinations" section of the flowchart as well. In the MS/MS case, isotope combinations and element combinations are computed individually for product ion and neutral loss of the given molecule. How IsoCorrectoR computes $p_{ij}$ from the element combinations $u$ depends on the individual mode of correction. The correction of ultra-high-resolution data is not included in this figure.*

labeling state $j$ produces the exact labeling pattern of another multi-tracer labeling state $i$ through NA (and possibly tracer purity). An example for such a labeling state $j$ in a combined $^{13}$C and $^{15}$N labeling experiment may be $^{13}CC_2H_6NO_2^-$, while the labeling state $i$ might be $^{13}C_2CH_6{}^{15}NO_2^-$ (note, however, that H and O are actually irrelevant in this model). The first step to obtain $P_{mult}$ is to compute a probability matrix $P_t$ for each individual tracer element with index $t$, as described for UHR correction with a single tracer in the previous paragraph. Then, $P_{mult}$ can be computed in analogy to the matrix $P_{MS/MS}$ in the low-resolution MS/MS case. To this end, a given multi-tracer labeling state is split into its single-tracer labeling states, just as an MS/MS labeling state is split into product ion and neutral loss labeling states. *E.g.*, $^{13}CC_2H_6NO_2^-$ has the single-tracer labeling states $^{13}CC_2$ for carbon and N for nitrogen, while $^{13}C_2CH_6{}^{15}NO_2^-$ can be split into $^{13}C_2C$ and $^{15}$N. In the case of 2 tracer elements, a single entry $p_{ij}$ of $P_{mult}$ is derived by multiplying the probability that the tracer-element-1 (*e.g.*, carbon) labeling state $j_1$ of $j$ produces the number of label in the tracer-element-1 labeling state $i_1$ of $i$ with the probability that the tracer-element-2 (*e.g.*, nitrogen) labeling state $j_2$ of $j$ produces the number of label in the tracer-element-2 labeling state $i_2$ of $i$. This is shown for the generic case of $t$ tracer elements in Equation II.13. The maximum index $k$ of $P_{mult}$ (*i.e.*, the number of expected multi-tracer labeling states) is given by $k = \prod_t l_t + 1$, where $l_t$ is the maximum number of label possible for the tracer element with index $t$ in the given molecule. Table II.7 shows $P_{mult}$ for the alanine $^{13}$C and $^{15}$N labeling example, as computed by IsoCorrectoR when assuming a purity of 0.99 for both tracer elements.

$$P_{mult\,ij} = \prod_t P_{t\,i_t\,j_t} \tag{II.13}$$

The Python-based tool PyNAC (Carreer et al. (2013)) is also capable of correcting UHR labeling data from experiments with multiple tracers. However, the algorithmic approach is a different one, and it is, in its current form, not compatible with tracer purity correction. PyNAC does not use a correction matrix approach for correcting the data. Instead, correction is performed in a step-wise fashion, assuming that no correction is necessary for the labeling state that contains no label at all. Clearly, this assumption directly excludes the possibility of tracer purity correction, as correction would be necessary for the unlabeled species in that case (it receives impurity contributions from labeled species). PyNACs incremental UHR correction algorithm begins with correcting the species containing 1 label using the value of the unlabeled species (which needs not be corrected). Then it uses the values of the species with 0 and 1 label (corrected) to correct the species with 2 label incorporated. This process is repeated until all labeling states are corrected for NA. In the case of multiple-tracer data, for the correction of a given multiple-tracer labeling state like $^{13}C_2CH_6{}^{15}NO_2^-$, the corrected values of all species with less or equal amounts of label from either tracer element are required (*e.g.*, $^{13}CC_2H_6{}^{15}NO_2^-$, $^{13}C_2CH_6NO_2^-$ or $^{13}CC_2H_6NO_2^-$). If negative corrected values arise, they are flattened to 0 (Moseley (2010)). This is in contrast to the least-squares based approach enforcing values $\geq 0$ in IsoCorrectoR.

### 3.1.7 Correction residuals

As stated previously, the system of linear equations defined in Equation II.1 is solved with the constraint that no value in $v_c$ (corrected values) may be smaller than 0. Thus, if an exact solution of the system of linear equations would lead to negative corrected values due to

| | C0.N0 | C0.N1 | C1.N0 | C1.N1 | C2.N0 | C2.N1 | C3.N0 | C3.N1 |
|---|---|---|---|---|---|---|---|---|
| C0.N0 | 9.647E-01 | 9.682E-03 | 9.751E-03 | 9.787E-05 | 9.857E-05 | 9.893E-07 | 9.963E-07 | 1.000E-08 |
| C0.N1 | 3.563E-03 | 9.586E-01 | 3.602E-05 | 9.689E-03 | 3.641E-07 | 9.794E-05 | 3.680E-09 | 9.900E-07 |
| C1.N0 | 3.130E-02 | 3.142E-04 | 9.656E-01 | 9.691E-03 | 1.952E-02 | 1.959E-04 | 2.959E-04 | 2.970E-06 |
| C1.N1 | 1.156E-04 | 3.110E-02 | 3.566E-03 | 9.594E-01 | 7.209E-05 | 1.939E-02 | 1.093E-06 | 2.940E-04 |
| C2.N0 | 3.385E-04 | 3.398E-06 | 2.088E-02 | 2.096E-04 | 9.663E-01 | 9.698E-03 | 2.929E-02 | 2.940E-04 |
| C2.N1 | 1.250E-06 | 3.364E-04 | 7.713E-05 | 2.075E-02 | 3.569E-03 | 9.601E-01 | 1.082E-04 | 2.911E-02 |
| C3.N0 | 1.221E-06 | 1.225E-08 | 1.129E-04 | 1.133E-06 | 1.045E-02 | 1.049E-04 | 9.667E-01 | 9.703E-03 |
| C3.N1 | 4.508E-09 | 1.213E-06 | 4.171E-07 | 1.122E-04 | 3.859E-05 | 1.038E-02 | 3.571E-03 | 9.606E-01 |

Table II.7: ***Probability matrix $P$ for the UHR correction of a multiple-tracer alanine example ($^{13}C$ and $^{15}N$ labeling), as computed by IsoCorrectoR when assuming a tracer purity of 0.99 for both tracer elements.*** *Column names and row names of the matrix indicate the alanine multi-tracer labeling states, where Cx.Ny is the labeling state containing x $^{13}C$ and y $^{15}N$. Digits after the decimal point are limited to 3 for readability.*

measurement/integration errors (incorrect $v_m$) or unsuitable isotope abundance values (incorrect $P$), an exact solution of Equation II.1 is not possible because of the constraint. As a consequence, correction residuals result. These residuals are given by $v_m - P \cdot v_c$, using the corrected values $v_c$ obtained via correction under the constraint. The magnitude of correction residuals can be an indicator of how well the correction worked on the data. It must be noted, however, that the absence of residuals does not mean that the correction results are reliable. Residuals only result if measured values are close to 0/mainly composed of NA contributions. If all measured values are relatively high, both $v_m$ and $P$ may be incorrect, but no residuals are seen. If substantial residuals can be observed, though, this indicates that there is an issue.

### 3.1.8 Handling of missing values

IsoCor and ICT are unable to handle missing values in the measured data $v_m$. If a value could not be measured, it has to be set to 0 explicitly, otherwise correction cannot proceed. This is reasonable, if the value is missing because it is actually close to 0 and probably below the lower limit of quantification (LLOQ). However, the true value will never actually be 0 due to NA and tracer purity contributions of other species. When using a least-squares approach to deal with measurement noise as in IsoCor and IsoCorrectoR, stating an explicit value of 0 in $v_m$ will thus distort the correction results in $v_c$ to some degree, as the algorithm attempts to adapt other values to fit the (false) value of 0 that has been provided. To deal with this issue, the IsoCorrectoR algorithm does not require that values for all expected labeling states are provided. If the value of a labeling state $i$ is missing, its corresponding row and column at index $i$ are simply removed from the probability matrix and thus from the system of linear

equations described by Equation II.2. This way, the (missing) value of labeling state $i$ is not corrected, nor its contribution to other labeling states. To alert the user, IsoCorrectoR will issue a warning for each missing value. If an explicit and false value of 0 was stated for $i$, however, $i$ would be part of the correction procedure and thus distort the correction results to some degree. Clearly, though, regardless of whether a missing value of a given labeling state $i$ is set to 0 or if it is excluded from correction completely, the data will not be corrected for contributions of $i$ to other labeling states. Thus, if the true value of $i$ is actually not close to 0, but *e.g.* missing because integration failed due to peak overlap in the chromatogram, the results of correction may still be recognizably distorted by NA, depending on the magnitude of NA contributions of $i$ to other species. This should always be considered when performing correction on data with missing values. Also, calculations of mean isotopic enrichment should be performed with care if values of expected labeling states are missing.

### 3.1.9 Omitting the tracer element in the core molecule when correcting

In a derivatized molecule, only the tracer element atoms in the part of the molecule that does not come from derivatization can be labeled through metabolism. This part of the molecule will be termed the core molecule. However, the atoms in the derivatizing groups influence the magnitude of NA contributions just as the atoms in the core molecule. In some cases, *e.g.*, if the derivatizing groups contain Si or add recognizable amounts of C atoms to the molecule, the influence of derivatization on NA may even exceed that of the core molecule. Some tools used for metabolic flux analysis have a built-in functionality to correct for the NA contributions of (only) the tracer element atoms in the core molecule (Midani et al. (2017)). In such cases, one may want to correct for the remaining NA with a tool like IsoCorrectoR, and it has to be possible to exclude the part of correction performed by the tool used for metabolic flux analysis. Otherwise, correction for NA contributions of the tracer element atoms in the core molecule will be performed twice, leading to wrong results. Thus, IsoCorrectoR, as well as IsoCor and ICT, can be set to omit the tracer element NA in the core molecule when correcting. Internally, this is done by simply removing the unlabeled core tracer element atoms from the labeling state molecular formulas used for computing the probability matrix. For example, the labeling state $^{13}CC_2H_6NO_2^-$ would be turned to $^{13}CH_6NO_2^-$, $C_3H_6NO_2^-$ would be converted to $H_6NO_2^-$. If the molecule is derivatized, tracer element atoms from derivatization remain.

### 3.1.10 Implementation

IsoCorrectoR's algorithm has been implemented in an R-package with the same name. The package is available from Bioconductor, an open-source respository for R-based bioinformatics software (*https://bioconductor.org/packages/release/bioc/html/IsoCorrectoR.html*). It is also available on request from the Institute of Functional Genomics at the University of Regensburg, Faculty of Medicine. As of the date of thesis submission, the version available on Bioconductor lacks the functionality of resolution-dependent MS and MS/MS correction, while a prototype R-package that contains all functionality described in this section is available from the Institute of Functional Genomics. IsoCorrectoR depends on the *pracma* R-package (Practical Numerical Math Functions). It uses *pracma*'s *lsqlincon* function for solving Equation II.1 with the constraint that all values in $v_c$ are $\geq 0$ in a least-squares approach. Apart from that, IsoCorrectoR only depends on *tidyverse* packages for data handling and

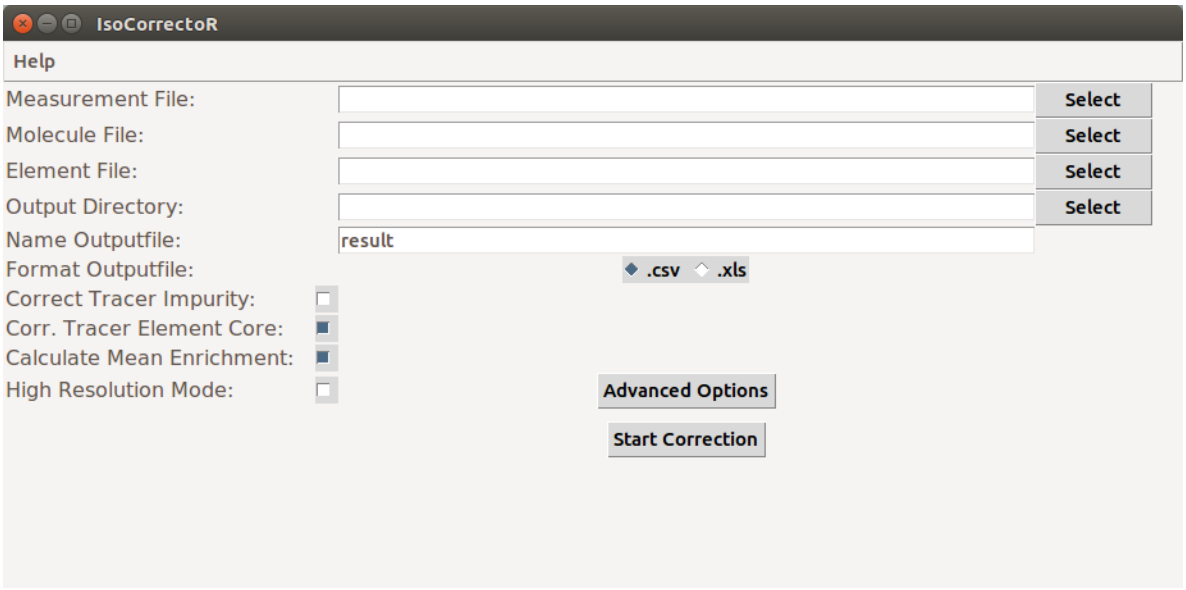



Figure II.6: ***IsoCorrectoRGUI: A graphical user interface to IsoCorrectoR.*** *The figure shows a screenshot of the graphical user interface to IsoCorrectoR that is provided by the IsoCorrectoRGUI-package.*

string operations, as well as packages for reading and writing (Microsoft Excel) data files (*readr*, *readxl*, *WriteXLS*). The package IsoCorrectoRGUI, which is also available on Bioconductor (*https://bioconductor.org/packages/release/bioc/html/IsoCorrectoRGUI.html*) and from the Institute of Functional Genomics provides an optional graphical user interface to IsoCorrectoR (see Figure II.6). It depends on the *tcltk*/*tcltk2* packages which are used for creating graphical user interfaces. The original prototype of IsoCorrectoR was implemented in Matlab. This prototype is capable of performing all types of correction described in this section, except resolution-dependent MS and MS/MS correction. It is available upon request from the Institute of Functional Genomics.

IsoCorrectoR can be used either via the R console or via the graphical user interface provided with IsoCorrectoRGUI. As input, it requires a measurement file containing measurements of labeling states (intensity values or fractions) labeled with an ID, a molecule file containing molecular formula information on the molecules (or their fragments, in MS/MS mode) in the measurement file, and an element file containing information on natural isotope abundance and mass shifts/absolute masses of the elements and tracer purity. Among the additional settings are the options to correct for tracer purity, to omit the tracer element atoms in the core molecule when correcting, and to supply resolution information in the case that resolution-dependent correction is desired. Resolution information consists of a reference resolution, a reference $m/z$ and a resolution formula. The resolution formula gives the dependency of $FWHM$ on $m/z$, as described in section 3.1.4 for orbitrap and FT-ICR devices. For example, a possible input for an orbitrap device would be a reference resolution of 140000 at a reference $m/z$ of 200, using the orbitrap resolution formula (see Equation II.10). By default - *i.e.*, no resolution information is provided - IsoCorrectoR performs low-resolution correction. IsoCorrectoR also offers the possibility to perform UHR correction on UHR (and possibly multiple tracer) data. Additionally, IsoCorrectoR allows the user to set a threshold value for internal calculations. If it is set, no probabilities below this threshold will be computed by IsoCor-

rectoR when constructing the probability matrix, which can save computational resources. By default, the calculation threshold is set to 1E-08 in the current implementation, which has virtually no effect on the results (see 3.3.6 in the validation section). This default was also used for almost all analyses in this thesis, see the corresponding part of the methods section (4.1.1) for more details. Other NA correction tools use similar thresholds, although they are usually fixed and not a modifiable parameter. IsoCorrectoR's UHR correction uses a calculation threshold that is different from the one that has just been described. In UHR correction, computations are not restricted to probabilities above a fixed threshold. Instead, the relevant metric is the difference in the amount of label incorporated between two labeling states. Using the default value of 8 for the UHR calculation threshold as an example, this means that the contributions of a multi-tracer labeling state with 1 $^{13}$C tracer isotope and 1 $^{15}$N tracer isotope to a labeling state with 6 $^{13}$C tracer isotopes and 2 $^{15}$N tracer isotopes will be computed, as the total labeling difference is 6 ($\leq 8$). However, the (negligibly small) contributions to a labeling state with 8 $^{13}$C tracer isotopes and 3 $^{15}$N tracer isotopes will not be computed, as the labeling difference of 9 exceeds the threshold of 8.

| Variable | Description |
|---|---|
| $v_m$ | Vector of measured values |
| $v_c$ | Vector of corrected values |
| $P$ | Probability or correction matrix |
| $p_{ij}$ | Single entry of $P$ |
| $i, j$ | Indices of $P$/labeling state indices |
| $q$ | Isotope combination (tuple of frequencies of isotopes of an element) |
| $\theta_p(q)$ | Isotope combination probability |
| $\theta_m(q)$ | Isotope combination mass shift |
| $n_q$ | Total number of isotopes in isotope combination $q$ |
| $h$ | Index of isotope in an isotope combination |
| $a_h$ | Natural abundance of isotope with index $h$ |
| $\mu_h$ | Mass shift of isotope with index $h$ |
| $f_{h,q}$ | Frequency of isotope with index $h$ in $q$ |
| $u$ | Element combination or isotopologue of a molecule/fragment |
| $\omega_p(u)$ | Element combination probability |
| $\omega_m(u)$ | Element combination mass shift |
| $e$ | Element |
| $\theta_p(u, e)$ | Probability of the isotope combination of element $e$ in $u$ |
| $\theta_m(u, e)$ | Mass shift of the isotope combination of element $e$ in $u$ |
| $m_{i/j}$ | Intrinsic mass shift of labeling state $i/j$ |
| $U_j$ | Set of all element combinations of a labeling state $j$ |
| $U_{ij}$ | Set of all element combinations of a labeling state $j$ that cannot be resolved from a labeling state $i$ |

Table II.8: ***Description of variables.*** *Description of variables used in this section.*

| Variable | Description |
|---|---|
| $P_{MS/MS}$ | Probability or correction matrix, MS/MS correction |
| $P_{MS/MS\,ij}$ | Single entry of $P_{MS/MS}$ |
| $P_{prod}$ | Probability or correction matrix, product ion only |
| $P_{loss}$ | Probability or correction matrix, neutral loss only |
| $i_{prod}/j_{prod}$ | Index associated with product ion of labeling state $i/j$ in $P_{prod}$ |
| $i_{loss}/j_{loss}$ | Index associated with neutral loss of labeling state $i/j$ in $P_{loss}$ |
| $\delta_{min}$ | Minimum mass difference required to resolve a given labeling state from another species |
| $z$ | Molecule charge |
| $R_{ref}$ | Reference resolution used for computing $FWHM$ |
| $m_{ref}$ | Reference $m/z$ used for computing $FWHM$ |
| $u_{prod}$ | Product ion element combination associated with an element combination $u$ |
| $u_{loss}$ | Neutral loss element combination associated with an element combination $u$ |
| $m_{i_{prod}}$ | Intrinsic mass shift of the product ion of a labeling state $i$ |
| $\delta_{min_{prod}}$ | Minimum mass difference required to resolve the product ion of a given labeling state from another product ion species |
| $\delta_{min_{prec}}$ | Minimum mass difference required to resolve the precursor ion of a given labeling state from another precursor ion species |
| $P_{mult}$ | Probability or correction matrix, UHR multiple-tracer correction |
| $P_{mult\,ij}$ | Single entry of $P_{mult}$ |
| $P_t$ | Probability or correction matrix of a single tracer element with index $t$ in UHR (multiple-tracer) correction |
| $i_t/j_t$ | Index in $P_t$ that is associated with tracer element $t$ labeling state of multi-tracer labeling state $i/j$ |

Table II.9: ***Description of variables.*** *Description of variables used in this section.*

## 3.2 Application of IsoCorrectoR

With the algorithm behind IsoCorrectoR explained in the previous section, this section deals with the application of IsoCorrectoR for the correction of data from stable isotope labeling experiments. The different modes of correction of IsoCorrectoR (*e.g.*, low-resolution correction of MS and MS/MS data, tracer purity correction, resolution-dependent correction) and their impact on the data will be assessed with practical examples.

### 3.2.1 Natural isotope abundance correction

The most basic feature of IsoCorrectoR is the correction of low to medium resolution MS data (*e.g.* acquired with quadrupole, linear ion trap or TOF instruments) from stable isotope labeling experiments for natural isotope abundance (NA). At this level of resolution, isotopic contributions that provide the same nominal mass shift (*e.g.*, $^{13}$C and $^{15}$N) usually cannot be resolved spectrometrically. As a consequence, all isotopic contributions have to be considered in correction, and IsoCorrectoR's correction approach for low-resolution MS data has to be applied.

To illustrate the impact of basic NA correction, Figure II.7 shows data from a stable isotope labeling experiment conducted in a P493-6 B-cell line. This cell line can serve as a model for B-cell lymphoma (see chapter III). Cells were fed with U-$^{13}$C-glutamine, and the incorporation of $^{13}$C into different amino acids and TCA cycle intermediates was assessed via LC-MS/MS of cell extracts, using a low-resolution triple-quadrupole device. Prior to measurement, amino acids were derivatized with propylchloroformate (PCF) to improve separation in reversed phase HPLC (the derivatization reaction is illustrated in Figure IV.1 a in the methods chapter for a generic amino acid). The amino acids glycine, ornithine, proline and serine were measured in MRM mode, however, the neutral loss comes purely from derivatization and cannot contain label. As a consequence, it is valid to perform MS correction on the data using the product ion molecular formula, and MS/MS correction is not required.

Figure II.7 shows the uncorrected mass isotopomer distribution (MID) in comparison to the MID resulting from NA correction for different amino acids. Each isotopologue fraction is the mean of biological triplicates. When considering glycine, ornithine and serine, only a very low degree of labeling can be observed in the uncorrected MID. After NA correction, the MID consists almost entirely of the m+0 isotopologue. What becomes especially evident is the strong decrease in the fractions of the m+1 and m+2 isotopologues, suggesting that the uncorrected fractions came almost exclusively from natural isotope abundance. This is supported by the MIDs of the same amino acids from the medium control samples, which can be seen in Figure II.8. In cell culture medium controls, no labeling should have occured and, in theory, MIDs should consist only of the m+0 isotopologue after NA correction. In fact, the medium control MIDs of gylcine and serine match the MIDs from the cell extracts very closely, suggesting that no or hardly any labeling of those amino acids has occured. However, comparing the m+1 isotopologue of glycine between cell extract and medium control in a two-tailed two-sample Welch's t-test (we don't assume equal group variances) yields a significant difference ($p = 0.0312$). This may indicate a very small degree of labeling or a small but systematic integration error.

Considering ornithine, on the other hand, small m+3 and m+5 fractions remained after correction and do not seem to be present in the medium control. This indicates a small degree

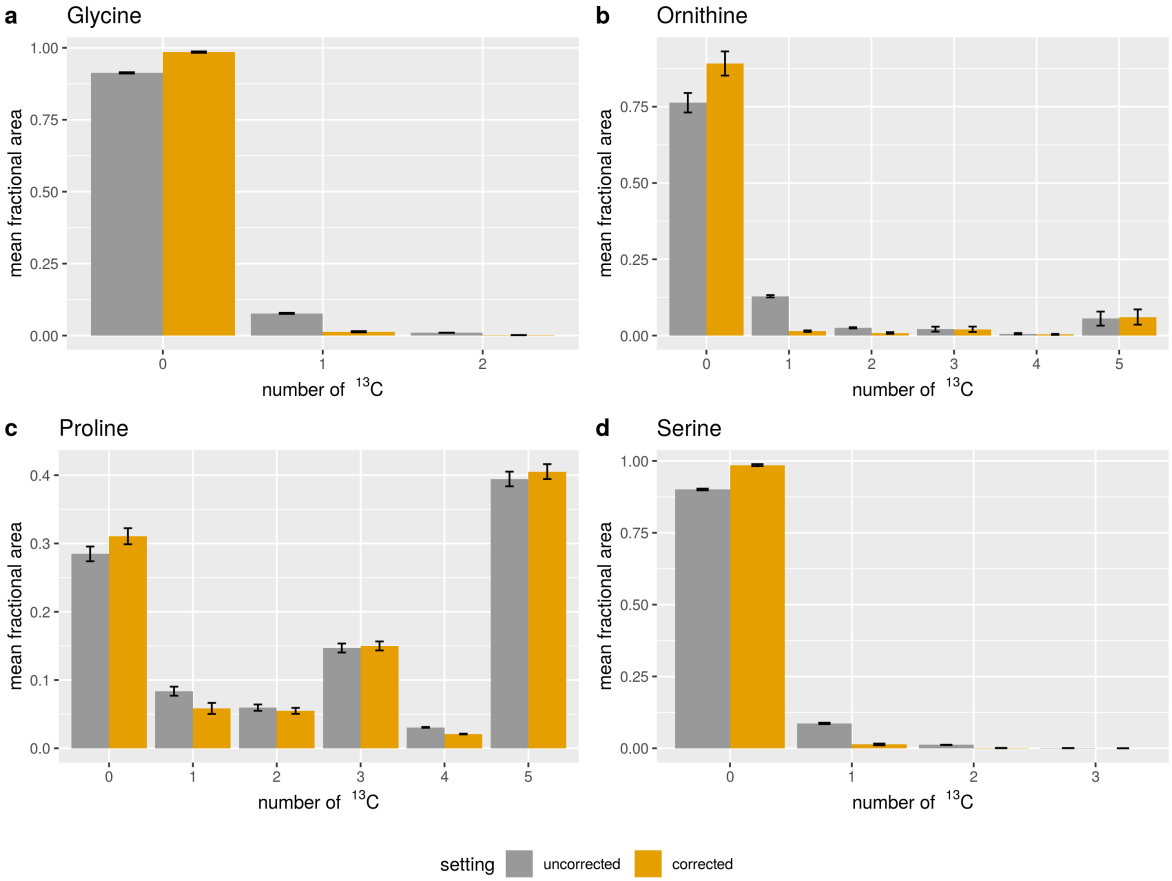



Figure II.7: ***Correction of PCF-derivatized amino acids for natural isotope abundance, cell extract.*** *The figure shows the results of natural isotope abundance correction on isotopologues of PCF-derivatized amino acids from cell extracts of a stable isotope labeling experiment. Panels* ***a - d*** *show the corrected and uncorrected mass isotopomer distribution (MID) for the different amino acids. The x-axis labels correspond to the number of $^{13}C$ label incorporated in the respective isotopologue. Samples were measured in (biological) triplicates, means of isotopologue fractions +/- SD are shown.*

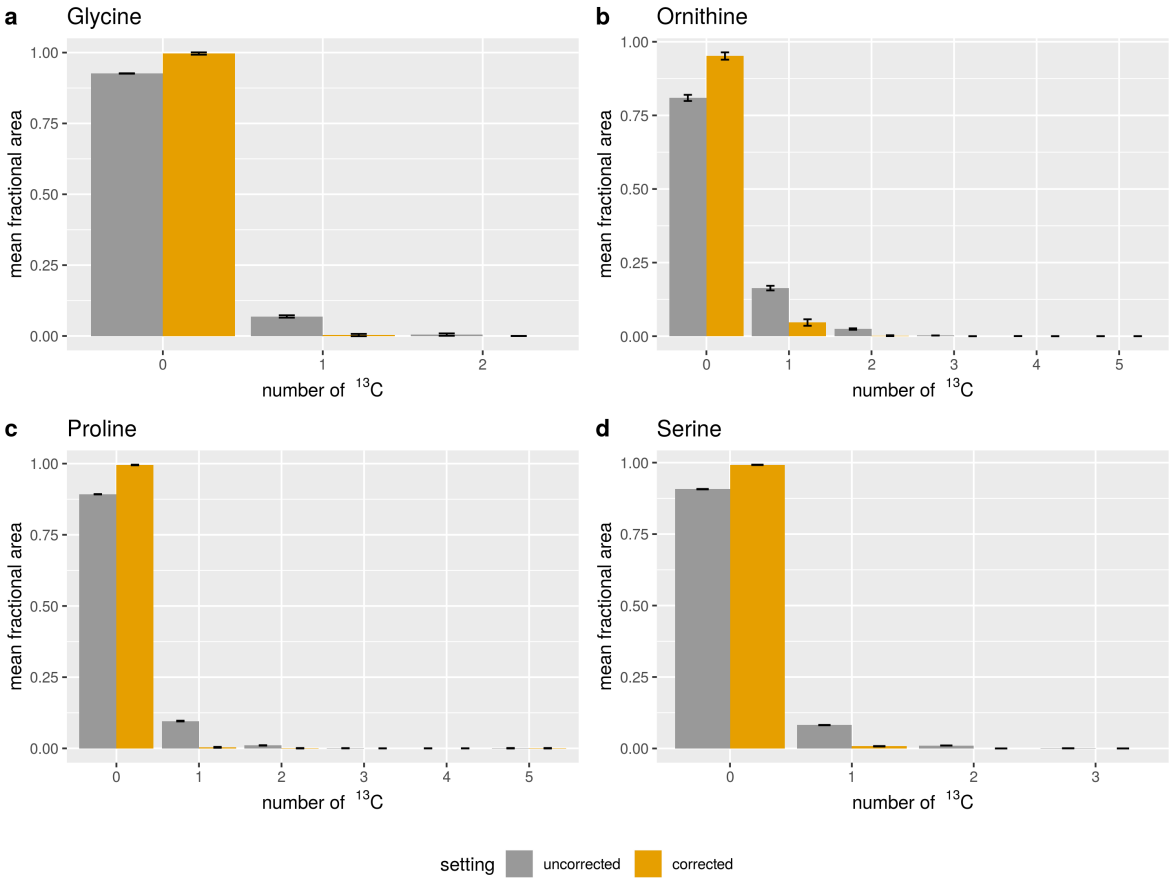



Figure II.8: ***Correction of PCF-derivatized amino acids for natural isotope abundance, medium control.*** *The figure shows the results of natural isotope abundance correction on isotopologues of PCF-derivatized amino acids from the medium control samples of a stable isotope labeling experiment. Metabolites in medium control samples are not expected to be labeled. Panels* ***a - d*** *show the corrected and uncorrected mass isotopomer distribution (MID) for the different amino acids. The x-axis labels correspond to the number of $^{13}C$ label incorporated in the respective isotopologue. Samples were measured in (biological) triplicates, means of isotopologue fractions +/- SD are shown.*

of labeling of ornithine through the glutamine tracer. However, using a two-tailed two-sample Welch's t-test on the corrected data, the difference between medium control and cell extract is not significant ($p = 0.0551$ and 0.0521 for m+3 and m+5, respectively). This is likely because of the small sample size ($n = 3$) and the high variance of ornithine m+3 and m+5 values in cell extracts. Prior to performing the t-test, the Shapiro-Wilk test was applied to the data of both groups to assess normality. The null hypothesis of normally distributed data could be accepted for both cell extract and medium control for the m+5 isotopologue. For the m+3 isotopologue, it had to be rejected for the medium control. However, this appears to be a result of correction, as the Shapiro-Wilk test is insignificant ($p = 0.845$) for uncorrected medium control data of the m+3 isotopologue. Upon correction, the m+3 values of some of the medium control samples become exactly 0. This indicates that the correction forced values that would otherwise have been negative to 0, at the expense of yielding correction residuals (in fact, there are residuals for both samples where the corrected value is exactly 0). It appears reasonable that this can in turn result in a deviation from the normal distribution. If the same values were sampled repeatedly in new, similar experiments, we would expect the mean of the corrected m+3 isotopologue in the medium control to always be very close to 0. As a consequence, the difference between means is essentially defined by the mean of the cell extract group. Further, the variance in the medium control group should be negligible compared to the variance in the cell extract group. Thus, the two-sample t-test should approximate a one-sample t-test where the mean of the (normally distributed) cell extract group is compared to a constant value of 0. In fact, performing such a one-sample t-test yields a p-value of 0.0549, which is very close to the p-value obtained with the two-sample test, 0.0551. When comparing to a constant of 0, we actually only test for the mean of the cell extract values being greater than zero, not less. Thus, a one-tailed one-sample t-test can be applied, which then yields the significant p-values 0.0274 and 0.026 for ornithine m+3 and m+5, respectively. Accounting for all this, applying t-tests to the data seems reasonable, even though the medium control group violates the normality assumption. Further, the significant differences found indicate actual ornithine labeling.

This example also illustrates a relevant notion of NA correction: In the classical linear algebra based correction of very small values, small random or systematic errors could actually lead to negative corrected values. However, IsoCorrectoR's algorithm does not allow negative values and forces them to 0 in a least-squares optimization approach. Due to this perturbation, the corrected data of very small values may appear to violate the normality assumption, even if it could be considered normally distributed prior to correction. However, this should not be a crucial issue, as the mean of groups with values that are close to or exactly 0 can be approximated by a constant value of 0 for statistical procedures.

Regarding proline, there is substantial labeling from glutamine, which is clearly visible before and after NA correction, but not in the medium control. However, also for proline, a substantial decrease especially in the m+1 and m+4 fractions can be found after correction, exemplifying the impact of the correction procedure on the data. When considering Figure II.8 (which shows the medium control) again, there are still small non-zero fractions of isotopologues other than m+0 after correction, especially ornithine m+1. If this is the case, the cause are likely measurement/integration inaccuracies. Another explanation are natural isotope abundance values used for correction that do not exactly match the actual isotope abundance in the specific samples.

The magnitude of NA correction mainly scales with two factors: 1) The number of atoms of

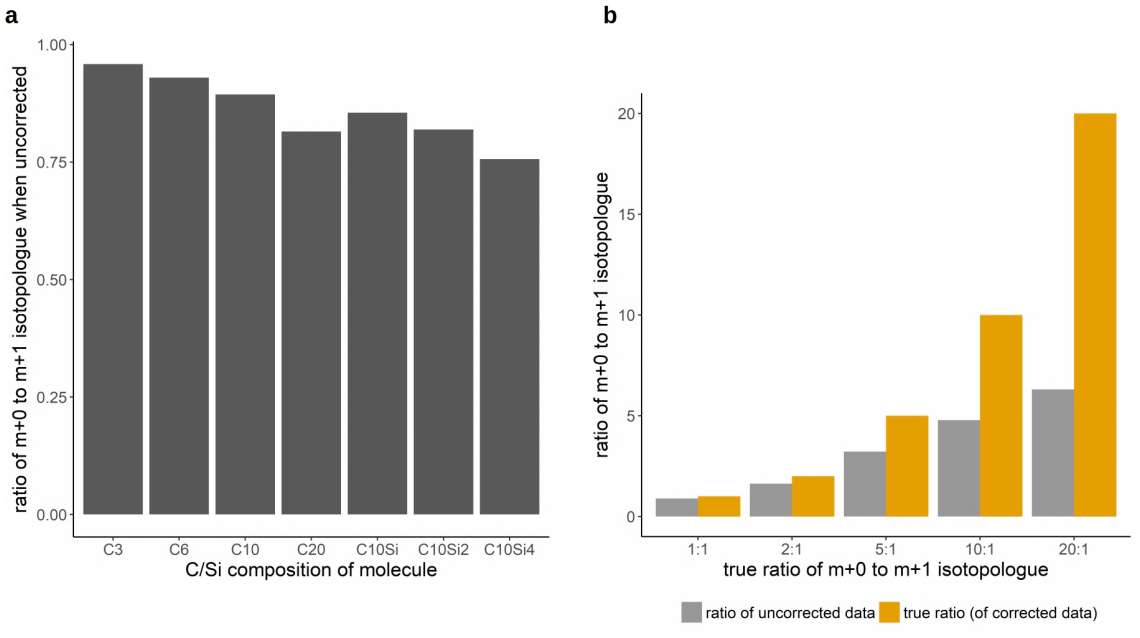



Figure II.9: ***Factors that determine the magnitude of NA correction.*** *Both panels show the ratio of the isotopologues m+0 to m+1 (0 $^{13}C$ and one $^{13}C$, respectively) in simulated uncorrected data. Panel a depicts this ratio for hypothetical molecules with differing C/Si composition. The true ratio is 1:1. In panel b the ratio of uncorrected data is shown for increasing true ratios of isotopologues m+0 to m+1 for a hypothetic molecule containing 10 C.* ***(Figure and legend taken from Heinrich et al. (2018), legend slightly modified)***

elements with abundant isotopes of higher mass in the molecule and 2), the relative abundance of the m+n with regard to the m+n+1 fraction. This is illustrated in Figure II.9. Panel a shows the uncorrected m+0/m+1 ratio for hypothetical molecules containing differing numbers of C and Si atoms. The true ratio after NA correction is 1 for all molecules. As can be seen, the more C atoms the molecule contains, the more the uncorrected ratio deviates from 1 (the NA of $^{13}C$ is 0.0107). The ratio decreases even more if Si atoms are additionally present in the molecule (the NA of $^{29}Si$ and $^{30}Si$ is 0.047 and 0.031, respectively). The second main factor affecting correction is illustrated in panel b. Here, the uncorrected m+0/m+1 ratio of a hypothetical molecule with 10 carbon atoms is shown for increasing true m+0/m+1 ratios. Clearly, the uncorrected ratio deviates more strongly from the true (corrected) ratio, the higher the m+0 isotopologue fraction is in relation to the m+1 isotopologue fraction. The deviations already become very clearly visible at a ratio of 2:1 and increase up to a factor of more than three for the 20:1 ratio. The more abundant the m+0 isotopologue, the more NA contributions it can make to the measurements of the m+1 isotopologue, and the lower the abundance of the m+1 isotopologue, the higher the NA based proportion of the m+1 measurement which is removed upon correction.

### 3.2.2 Tracer purity correction

Apart from NA, also the isotopic impurity of the tracer substrate can contribute to the distortion of MIDs. This distortion arises because molecules that incorporate a certain number of tracer element atoms from the tracer substrate through metabolism can also receive "impure" tracer element atoms if the tracer substrate is impure. This is illustrated in Figure II.10 b, which shows the metabolism of a U-$^{13}$C-glutamine tracer substrate to proline. The tracer substrate has a purity of 99%, so each of its five C atoms has a 1% chance of being a $^{12}$C instead of a $^{13}$C. If such a glutamine tracer substrate that contains *e.g.* one $^{12}$C due to tracer impurity is metabolized to proline, the result is a proline isotopologue that contains only four $^{13}$C, even though five C were transferred from the tracer substrate. While this species should be found in the measurement of the m+5 fraction from a biological viewpoint, it falsely contributes to the measurement of the m+4 fraction. This distorting effect is corrected for when applying tracer purity correction.

The a panel of Figure II.10 shows the proline MID from section 3.2.1 (Figure II.7), however, now MIDs corrected for both NA and tracer purity are included, assuming purities of 99% or 98%. Strikingly, the m+4 fraction completely disappears, already when correcting for a purity of 99%. This has direct biological implications with regard to data interpretation: When not correcting for tracer purity, one might have reasoned that it is possible to produce a proline species that carries four C atoms from glutamine by the cells used in the experiment. However, after correction, the fraction does not exist anymore, which suggests that the occurence of the m+4 fraction is merely an artifact introduced by tracer impurity. In fact, the true purity of the U-$^{13}$C-glutamine tracer substrate, as stated by the manufacturer, is 99%, which supports the assumption. This also means that assuming a purity of 98% should lead to overcorrection of the data, producing a slightly distorted MID. Thus, tracer purity correction should be performed with care, and only if reliable purity information is available. When it comes to the magnitude of tracer purity correction, it basically depends on similar factors as NA correction. It increases with the number of tracer element atoms that can be labeled and the impurity of the tracer substrate, as well as with the ratio of the m+n to the m+n-1 isotopologue fraction. The direction of tracer impurity contributions is opposite to NA contributions, from higher mass shift to lower mass shift.

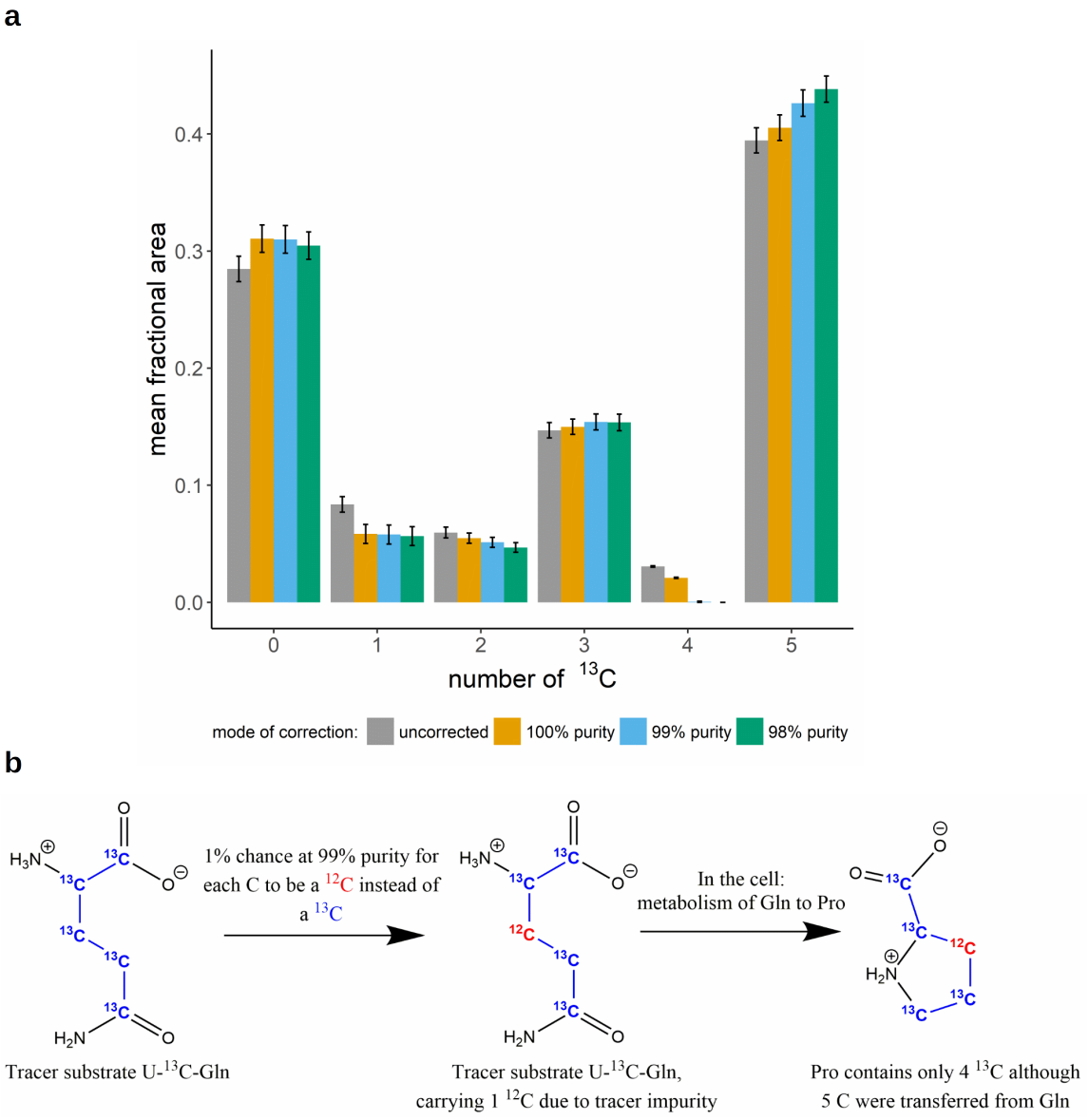



Figure II.10: ***Correction of PCF-derivatized proline for natural isotope abundance and tracer impurity.*** *Panel a shows the results of natural isotope abundance and tracer purity correction on PCF-proline isotopologues from a stable isotope tracer experiment. The x-axis labels 0-5 correspond to the proline isotopologues with 0 – 5 $^{13}C$ incorporated. Samples were measured in (biological) triplicates, means of isotopologue fractions +/- SD are shown. Correction was performed assuming different isotopic purities of the tracer substrate (100%, 99%, 98%). Panel b illustrates how isotopic impurities in the tracer substrate U-$^{13}C$-glutamine can lead to a distortion of the proline labeling patterns.* ***(Figure and legend taken from Heinrich et al. (2018), slightly modified)***

### 3.2.3 MS/MS correction

Figure II.11 shows examples of MS/MS measurements of amino acids from the $^{13}C$ stable isotope labeling experiment described in 3.2.1. For these amino acids, both product ion and neutral loss may contain label, as the neutral loss $C_4H_8O_2$ contains 1 carbon from the core molecule (the fragmentation is illustrated in Figure IV.1 a in the methods chapter for a generic amino acid). The figure contains both uncorrected MIDs and MIDs corrected for NA and tracer purity (99% purity). As can be seen, a measurement is not only defined by the total number of $^{13}C$ in the precursor ion, but by the number of $^{13}C$ in both precursor (or neutral loss) and product ion. This yields several transitions for each metabolite, and their number exceeds the number of measurements in MS labeling data. Apart from increased specificity and sensitivity, MS/MS measurements provide the advantage of positional labeling information, which can be very valuable, especially when modeling fluxes in $^{13}C$ MFA.

As detailed in the algorithm section (3.1.3), MS/MS data with potential label in both product ion and neutral loss has to be corrected using an MS/MS NA correction algorithm. Performing MS correction on MS/MS labeling data leads to overcorrection when using the precursor molecular formula for correction, and undercorrection when using the product ion molecular formula. In the special case where only one of the two fragments (either product ion or neutral loss) can contain label, MS correction using only the molecular formula of the respective fragment is valid. To be able to perform MS correction on MS/MS data, the MS/MS data first has to be converted to pseudo MS data by summing up the area values of transitions with the same precursor $m/z$. This is shown in Figure II.12 for simulated alanine example data. Positional labeling information is lost in the process.

The potential magnitude of overcorrection when correcting MS/MS data with an MS correction algorithm (using the precursor molecular formula) is illustrated in Figure II.13. Here, uncorrected MS/MS data have been simulated for carbon chains of varying length (see panel a). The carbon chains always consist of a 10 C product ion that can be labeled at 4 positions. The number of C that can be labeled in the neutral loss is always 2, however, the total chain length of the neutral loss varies up to 10 C. Carbons that cannot be labeled may be derived from a derivatizing group. Panel b shows the error introduced when performing MS correction on the simulated MS/MS data. In the simulated data, the true abundance (after MS/MS correction) of all isotopologues is equal. The figure shows the relative deviation from these expected values after MS correction. As can be seen, MS correction works fine for species containing 0 to 2 $^{13}C$ from labeling (up to 4 $^{13}C$ for NL-C2). The error then increases with increasing label incorporation. This happens, because up to an amount of 2 $^{13}C$, all possible NA contributions that are measured in MS mode are also measured in MS/MS mode. However, beginning with 3 $^{13}C$, there are NA contributions that are no longer measured in MS/MS because they exceed the maximum label expected for the neutral loss, which is 2 $^{13}C$. For example, the species from panel a containing 3 $^{13}C$ in the neutral loss portion of the molecule would be measured in MS, but not in MS/MS mode. Similarly, NA derived species with a mass shift exceeding that of 4 $^{13}C$ in the product ion are not measured in MS/MS (the reason why correction on NL-C2 works fine up to 4 $^{13}C$ is that it cannot contain more than 2 $^{13}C$ in the neutral loss, so only the product ion is relevant with regard to overcorrection). MS correction however corrects for NA contributions of species that are not measured in MS/MS, resulting in overcorrection. Overcorrection increases with the amount of label because the amount of NA contributions that are not measured in MS/MS increases. It also increases with the chain length (number of carbons that cannot be labeled), because the magnitude of NA correction

(and thus overcorrection) in general increases with the number of atoms. Panel c of Figure II.13 exemplifies how high the correction error can become under certain circumstances. It shows data from the same hypothetical molecules as in panel b, however, it is limited to the isotopologue labeled with 6 $^{13}$C. This is the isotopologue where the error was highest in the previous analysis. Further, the true amount of the 6 $^{13}$C species is only 1/10 of the amount of the other isotopologues that contribute to it through NA, to further increase the impact of correction. Under these circumstances, the relative error of MS correction on MS/MS data can be as high as 75%.

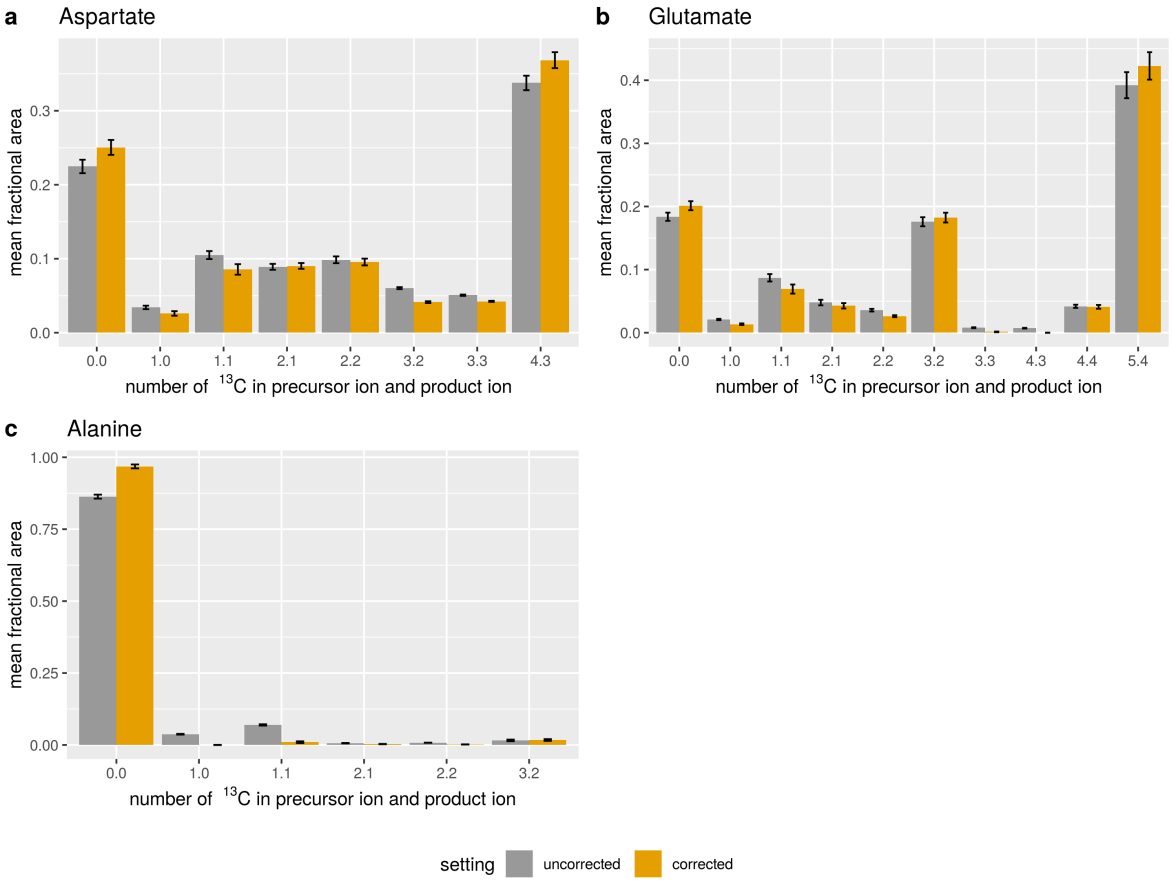



Figure II.11: ***MS/MS correction of PCF-derivatized amino acids for natural isotope abundance and tracer purity, cell extract.*** *The figure shows the results of MS/MS natural isotope abundance and tracer purity correction (purity 99%) on MS/MS data of PCF-derivatized amino acids from cell extracts of a stable isotope labeling experiment. Panels* ***a - c*** *show the corrected and uncorrected mass isotopomer distribution (MID) for the different amino acids. The x-axis labels n.m correspond to the isotopomers with n $^{13}C$ in the precursor ion and m $^{13}C$ in the product ion. Samples were measured in (biological) triplicates, means of isotopologue fractions +/- SD are shown.*

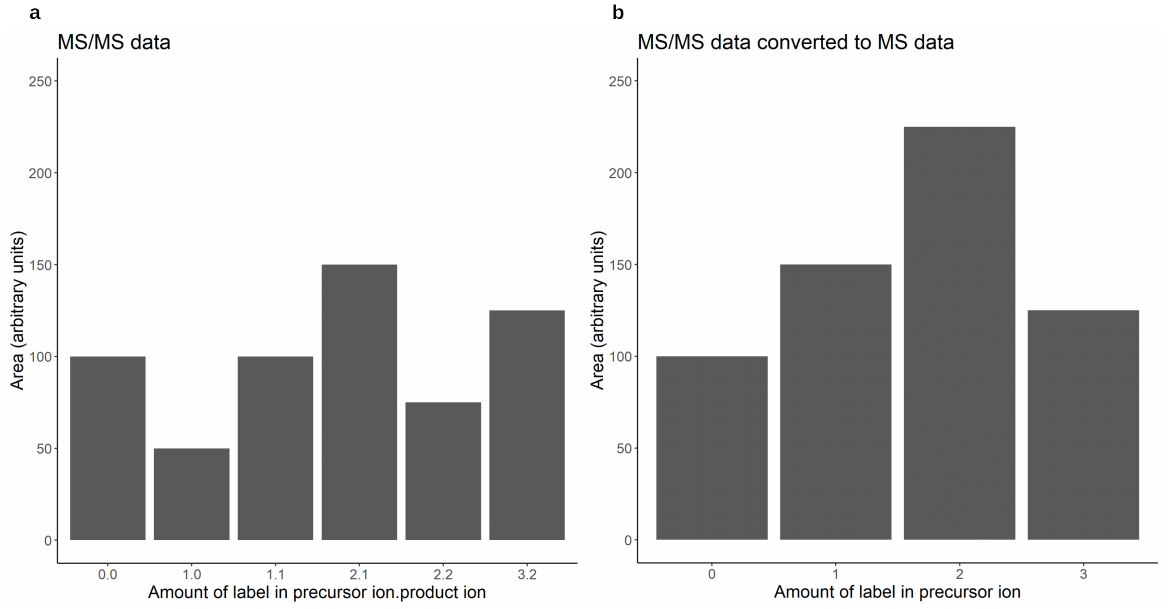



Figure II.12: ***Example of conversion of MS/MS data to MS data to be able to correct for natural isotope abundance with a MS correction algorithm.*** *The bar chart in panel a shows MS/MS data from alanine from a hypothetic $^{13}C$ stable isotope labeling experiment. The $C_3H_6NO_2^-$ precursor fragments into $C_2H_6N^-$ and $CO_2$. The x-axis labels correspond to the number of $^{13}C$ in the precursor ion and in the product ion of the given transition, separated by a dot. Panel b shows the data from panel a after it has been converted to MS data by adding the area values of transitions with an equal amount of label in the precursor ion (transitions 1.0 + 1.1. and transitions 2.1 + 2.2).* ***(Figure and legend taken from Heinrich et al. (2018) supplementary material, slightly modified)***

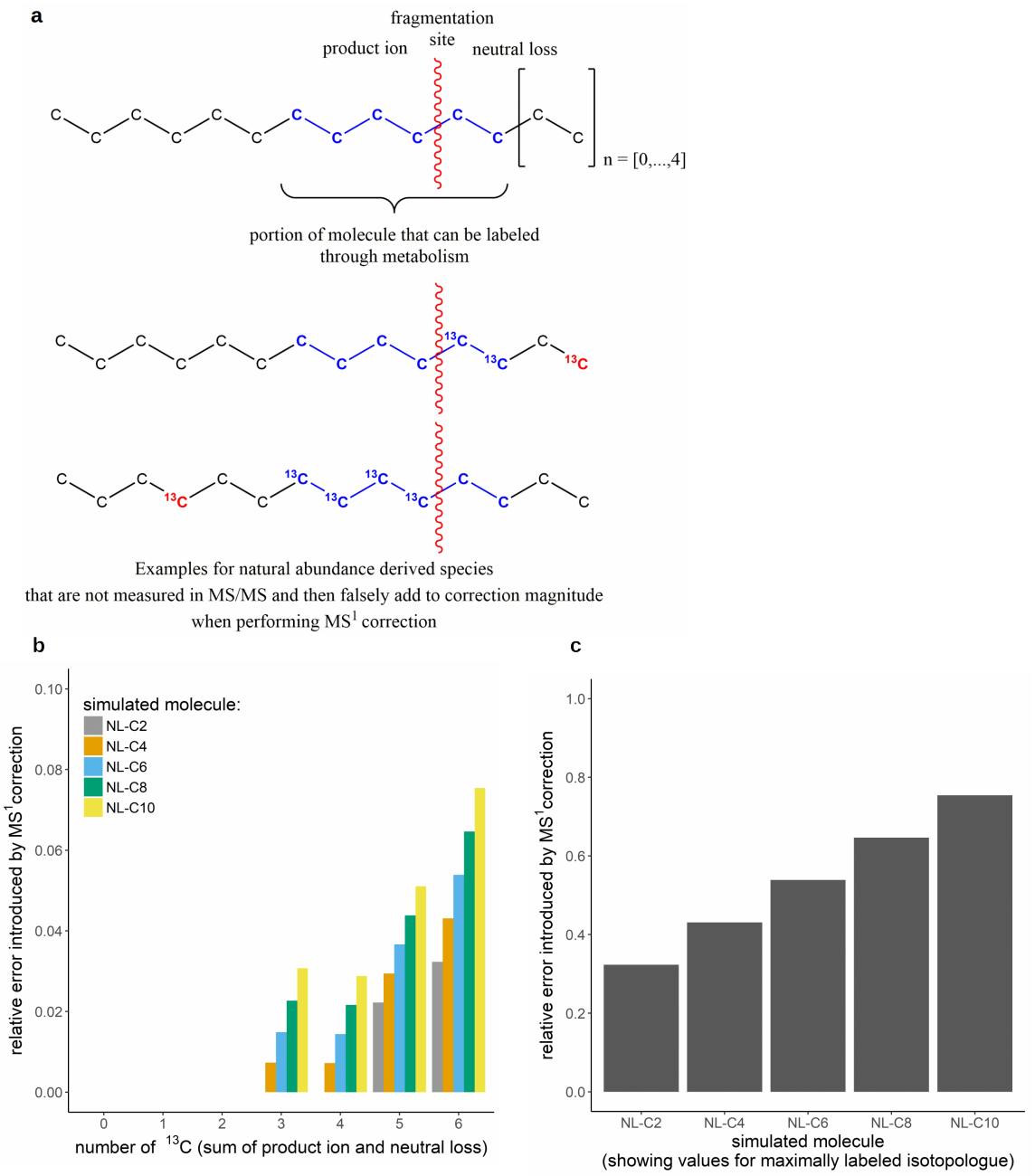



Figure II.13: ***The error introduced when performing MS correction on MS/MS data.*** *Panel a shows the structure of the hypothetic molecules used for systematic simulation of the effect of MS correction on MS/MS data. The section containing the first 12 C (6 of which can be labeled) is the same for each species, they only differ in the length of the neutral loss carbon chain. Panel a also depicts examples for species that are assumed to have been measured in a MS measurement and, thus, are corrected for performing MS correction. However, they are usually not measured in MS/MS approaches. In Panel b the relative error introduced by MS correction of MS/MS data is shown for the different isotopologues of the hypothetic molecules. The true (corrected) abundance of all isotopologues is equal. NL-Cn indicates a neutral loss chain length of n. Panel c illustrates the same error as panel b, except that only the values for the maximally labeled isotopologue (6 $^{13}C$) are shown for each hypothetic molecule. Furthermore, the true abundance of the 6 $^{13}C$ species is only 1/10 of that of the other isotopologues.* ***(Figure and legend taken from Heinrich et al. (2018) supplementary material, slightly modified)***

### 3.2.4 Resolution-dependent correction

As pointed out in the algorithm section dealing with resolution-dependent NA correction (3.1.4), stable isotope labeling MS data acquired with a high-resolution mass analyzer (*e.g.* , orbitrap, resolution range around 100000 - 200000) should not be corrected using a low-resolution correction algorithm. This is because some (or even many) of the NA contributions that would be corrected for in a low-resolution approach may already be resolvable spectrometrically, depending on mass analyzer resolution and molecule $m/z$. As a consequence, employing low-resolution correction on high-resolution data may lead to pronounced overcorrection.

Figure II.14 illustrates the potential error magnitude. It shows exemplary MIDs of (a) trimethylsilyl - (TMS-) derivatized proline ($C_{11}H_{26}NO_2Si_2^+$, $[M+H]^+$) and (b) propylchloroformate - (PCF-) derivatized proline ($C_{12}H_{22}NO_4^+$, $[M+H]^+$) from a simulated $^{13}C$ labeling experiment. The uncorrected MID was devised manually and is the same for both molecules. The MIDs have been corrected using IsoCorrectoR's low-resolution MS correction approach and the approach to resolution-dependent MS correction, assuming a tracer purity of 0.99. For resolution-dependent correction, a resolution of 140000 at 200 $m/z$ was assumed. For TMS-derivatized proline (panel a), there is a substantial difference between low-resolution and resolution-dependent correction for many of the mass fractions. In resolution-dependent correction, less NA contributions have to be removed, because they can already be resolved by the mass analyzer. This can lead to markedly higher mass fractions for some species compared to the low-resolution approach. A main reason for the pronounced differences observed for TMS-proline is the presence of Si in the molecule. At an orbitrap resolution of 140000 (200 $m/z$), the $\delta_{min}$ value is 0.0035 at the molecular ion (nominal) $m/z$ of 260 (see Equations II.10 and II.9). Thus, the incorporation of $^{29}Si$ can be resolved from the incorporation of $^{13}C$ (mass difference 0.0038), and the incorporation of $^{30}Si$ can be resolved from the incorporation of 2 $^{13}C$ (mass difference 0.0099). As a consequence, potential NA contributions that carry a Si isotope of higher mass usually do not have to be corrected for at this resolution. Since $^{29}Si$ and $^{30}Si$ exhibit a very high NA of 0.047 and 0.031, respectively, it is reasonable that this leads to pronounced differences.

Figure II.14 b shows the same comparison, however for PCF-proline instead of TMS-proline. Here, the differences between low-resolution and resolution-dependent correction are only very small. No Si atoms are present in the molecule, and the only isotope of higher mass that may reliably be resolved from $^{13}C$ at the given resolution and molecular ion $m/z$ (244, nominal) is $^{15}N$. For $^{2}H$, $^{17}O$ and $^{18}O$, the resolution is generally not sufficient. However, even if they could be resolved, their NA would be too low to produce really pronounced differences. The same is the case for $^{15}N$. Thus, the MIDs resulting from low-resolution and resolution-dependent correction are nearly the same, in contrast to the TMS-derivatized molecule. The same comparison is shown in Figure A1 in the appendix for a different uncorrected MID, yielding similar results. However, molecules need not contain Si for large differences to arise. Considering the example of acetyl-CoA ($C_{23}H_{39}N_7O_{17}P_3S^+$, $[M+H]^+$) from a simulated $^{15}N$ labeling experiment (the uncorrected MID was generated randomly, see 3.3 in the methods section) in Figure II.15, large NA contributions arise from the 23 carbons the molecule contains. In contrast to $^{13}C$ labeling experiments, where it is impossible to resolve $^{13}C$ NA contributions from the ($^{13}C$) labeled species, $^{13}C$ NA contributions can be resolved from the ($^{15}N$) labeled species in $^{15}N$ labeling experiments if the resolution is sufficiently high. As acetyl-CoA has a rather high molecular mass ($m/z$ of 810 for the molecular ion), a resolution in the range of 500000 (at 200 $m/z$) is needed to resolve the incorporation of $^{15}N$ from the incorporation

of $^{13}C$. At this resolution (panel b), there are stark differences between low-resolution and resolution-dependent correction (another isotope responsible for high NA contributions, $^{34}S$, still cannot be resolved from 2 $^{15}N$, though). If the resolution is insufficient, as shown in panel a (resolution of 140000 at 200 $m/z$), the results of the two modes of correction are nearly the same. Figure A2 in the appendix shows the same comparison for a different uncorrected MID.

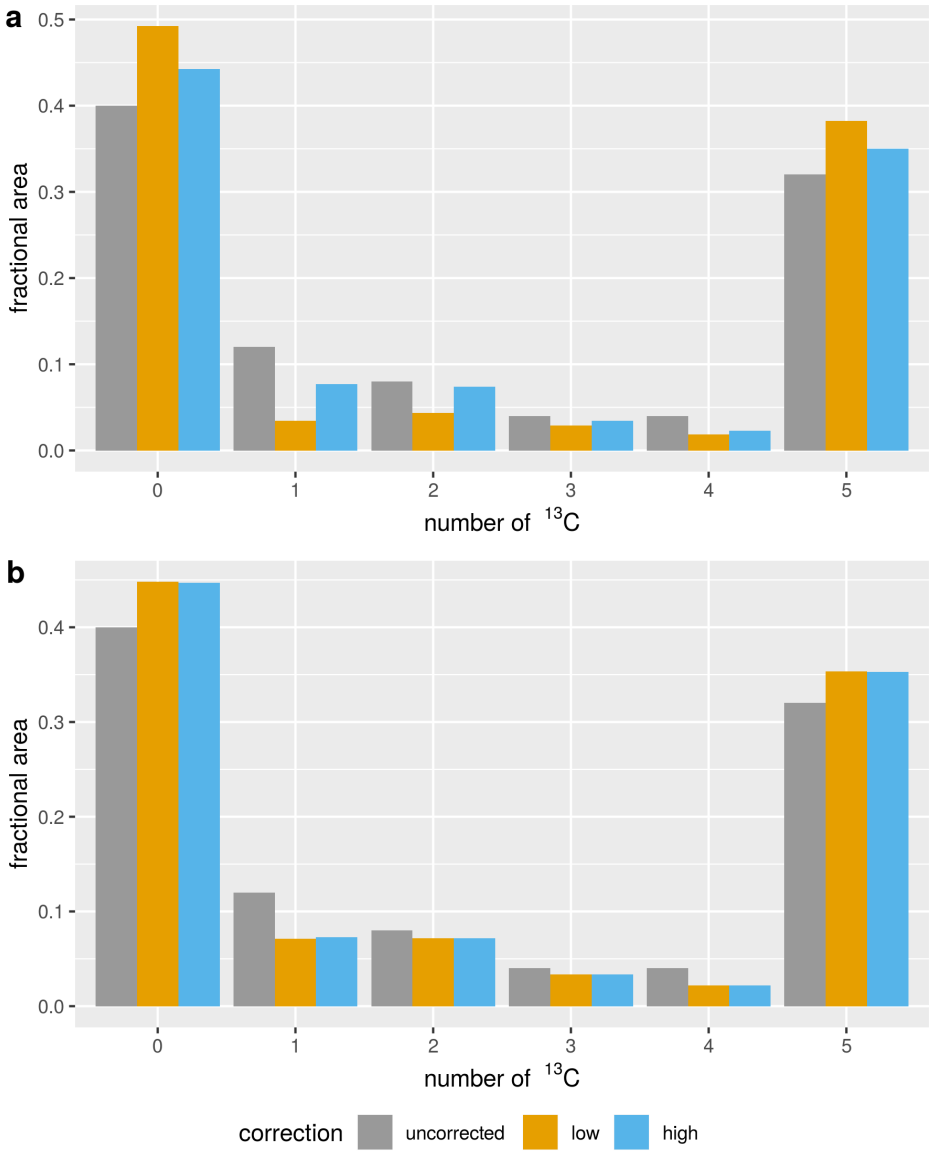



Figure II.14: ***Illustration of the impact of resolution-dependent MS correction on example data, $^{13}C$ labeled derivatized proline.*** *The figure shows the results obtained from correcting simulated MS data of TMS- (a) and PCF-derivatized (b) proline. The uncorrected MID is the same for both molecules. The x-axis labels correspond to the number of $^{13}C$ label incorporated in the respective isotopologue. Data are shown uncorrected, corrected with IsoCorrectoR's low-resolution MS correction algorithm (low) and corrected with IsoCorrectoR's resolution-dependent MS correction algorithm (high). Resolution-dependent correction was performed assuming an orbitrap mass analyzer with a resolution of 140000 at 200 m/z.*

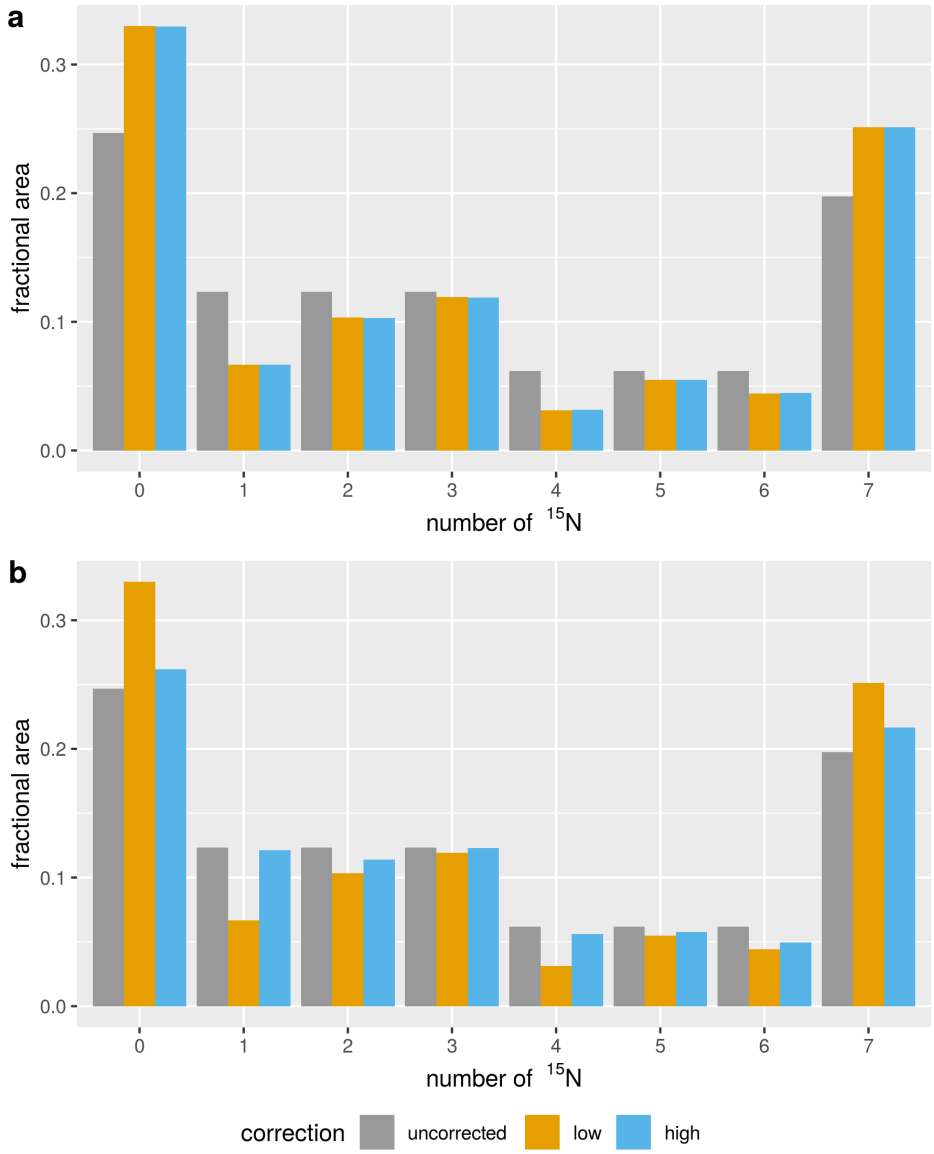



Figure II.15: ***Illustration of the impact of resolution-dependent MS correction on example data from $^{15}$N labeled underivatized acetyl-CoA.*** *The x-axis labels correspond to the number of $^{15}$N label incorporated in the respective isotopologue. Data are shown uncorrected, corrected with IsoCorrectoR's low-resolution MS correction algorithm (low) and corrected with IsoCorrectoR's resolution-dependent MS correction algorithm (high). Resolution-dependent correction was performed assuming an orbitrap mass analyzer with a resolution of 140000 at 200 m/z (a) and 500000 at 200 m/z (b). At the lower of the two resolutions, $^{13}$C and $^{15}$N can't be resolved for the 810 m/z acetyl-CoA molecular ion.*

### 3.2.5 Resolution-dependent MS/MS correction

Not only in MS, but also in MS/MS NA correction it may become necessary to correct high-resolution data in a resolution-dependent manner. In common orbitrap setups, a low-resolution quadrupole or linear ion trap mass analyzer for precursor isolation is coupled to a high-resolution orbitrap mass analyzer used for determining product ion *m/z*. Mass spectrometer setups where both precursor and product ion mass analyzer operate at high resolution are uncommon, but should basically also be possible, *e.g.* when performing fragmentation inside of a high-resolution mass analyzer after using it for precursor ion selection (Scigelova et al. (2011)). For both cases, resolution-dependent MS/MS correction may be needed. As with low-resolution MS/MS correction, it takes into account that not all NA contributions from MS correction have to be corrected for in the MS/MS case, because some of them are filtered through the tandem MS setup. Further, NA contributions that would be considered in low-resolution MS/MS correction may also be filtered at both the precursor and the product ion stage, if the mass analyzer resolution is sufficiently high to resolve them. Not accounting for this and correcting MS/MS data acquired with a high-resolution device with a low-resolution MS/MS correction approach may result in overcorrection. This is explained in detail in the algorithm section (3.1.5).

Figure II.16 gives an impression of the impact of resolution-dependent MS/MS correction. It shows exemplary MIDs of (a) trimethylsilyl - (TMS-) derivatized alanine and (b) propylchloroformate - (PCF-) derivatized alanine from a simulated $^{13}C$ labeling experiment. The uncorrected MID was devised manually and is the same for both molecules. TMS-derivatized alanine ($C_9H_{24}NO_2Si_2^+$, $[M+H]^+$) fragments at its C1-C2 bond to yield the product ion $C_5H_{14}NSi^+$ and the neutral loss $C_4H_{10}O_2Si$. PCF-derivatized alanine ($C_{10}H_{20}NO_4^+$, $[M+H]^+$) also fragments at its C1-C2 bond and yields $C_6H_{12}NO_2^+$ as the product ion and $C_4H_8O_2$ as the neutral loss (the fragmentation is illustrated in Figure IV.1 a and b in the methods chapter for a generic amino acid for PCF- and TMS-derivatization, respectively). The MIDs have been corrected using IsoCorrectoR's low-resolution MS/MS correction approach and the approach to resolution-dependent MS/MS correction, assuming a tracer purity of 0.99. Resolution-dependent correction was performed for the usual setup of having a low-resolution precursor mass analyzer (1000 at *m/z* 200, *FWHM* constant over the *m/z* range) and a high-resolution product ion mass analyzer (orbitrap, 140000 resolution at *m/z* 200), labeled "high - product ion", and for the rather theoretical case of both analyzers operating at high resolution, labeled "high - both analyzers". When considering TMS-derivatized alanine in Figure II.16a, there are substantial differences between the MIDs after low-resolution and after resolution-dependent correction. For the alanine isotopomer 1.1 (n.m means n $^{13}C$ in the precursor ion and m $^{13}C$ in the product ion), the isotopomer fraction after resolution-dependent correction (high - product ion) is about 155% of the fraction after low-resolution correction. For the 2.2 species, it is about 178%. The fraction after resolution-dependent correction is higher because less NA contributions have to be removed - it is assumed that the mass spectrometer is already able to resolve them during the measurement. Similarly to the resolution-dependent MS correction example in Figure II.14a (TMS-derivatized proline), a major factor that drives the substantial difference in this MS/MS example is the presence of Si in the molecule (due to the TMS-derivatization). When applying low-resolution MS/MS correction, the NA contributions of all isotopologues of labeling states 0.0 and 1.1 that carry a Si isotope of higher mass in their product ion portion are removed from 1.1 and 2.2. However, when performing resolution-dependent correction, it is assumed that the product ion mass analyzer employed (orbitrap,

140000 resolution at $m/z$ 200) can resolve the incorporation of Si isotopes of higher mass from the incorporation of $^{13}C$. With such a mass analyzer, the resolution should be around 184000 at the product ion $m/z$ of 116 (nominal) (formula II.10). This then results in a $\delta_{min}$ of about 0.001 for the product ion (formula II.9). At this value of $\delta_{min}$, the mass difference of about 0.0038 between the incorporation of $^{13}C$ and $^{29}Si$ should easily be resolved. The same is the case for the mass difference of about 0.0099 between the incorporation of 2 $^{13}C$ and 1 $^{30}Si$. Thus, potential NA contributions that carry a Si isotope of higher mass in their product ion usually do not have to be corrected for at this resolution (see Figure II.4).

Another observation of interest in Figure II.16a (TMS-derivatized alanine) is that the fraction of labeling state 1.0 (1 $^{13}C$ in the neutral loss) is substantially higher for the (rather theoretical) case of both precursor and product ion mass analyzer operating at high resolution - roughly 145% compared to only the product ion mass analyzer operating at high resolution. In this setup, NA contributions can already be filtered at the precursor stage (see Figure II.4). As a direct consequence, potential NA contributions that carry a Si of higher mass in their neutral loss portion are usually not removed during correction, because they can be resolved by the precursor mass analyzer. This yields a much higher isotopomer fraction than for the correction setups assuming a low-resolution precursor mass analyzer. In these cases, the fraction is corrected for all possible NA contributions to the neutral loss portion, because the mass analyzer can't filter them.

Similarly to the resolution-dependent MS correction example in Figure II.14b, which deals with PCF-derivatized proline, also the MS/MS correction example of PCF-derivatized alanine in Figure II.16b only shows minor differences between low-resolution and resolution-dependent correction, because only low abundant isotopic species can be resolved. This is different when considering the resolution-dependent MS/MS correction of acetyl-CoA MS/MS data from a simulated $^{15}N$ labeling experiment (the uncorrected MID was generated randomly, see 3.3 in the methods section) in Figure II.17. The precursor $C_{23}H_{39}N_7O_{17}P_3S^+$ ($[M+H]^+$) fragments into the product ion $C_{10}H_{16}N_5O_{10}P_2{}^+$ (3'-phosphonucleoside monophosphate) and the neutral loss $C_{13}H_{23}N_2O_7PS$. Again, as in the case of resolution-dependent MS correction of acetyl-CoA (Figure II.15), being able to resolve $^{13}C$ NA from the $^{15}N$ labeled species has a substantial impact on the data. In the MS/MS case, the orbitrap mass analyzer analyzes the 428 $m/z$ product ion, and not the 810 $m/z$ molecular ion, as in the MS case. As a consequence, a resolution of 200000 at 200 $m/z$ is already sufficient to resolve $^{13}C$ from $^{15}N$ and yield pronounced differences between low-resolution and resolution-dependent correction (see Figure II.17b). For the MS case, such substantial differences are not observed until a resolution of around 500000 at 200 $m/z$. Regarding $^{34}S$ in the precursor ion/neutral loss, even a resolution of 200000 (200 $m/z$) in the precursor mass analyzer is not sufficient to resolve this highly abundant isotope from 2 $^{15}N$. Figure II.17a shows the correction of the same MS/MS acetyl-CoA MID at a resolution of 140000 at 200 $m/z$, which is not sufficient to resolve $^{13}C$ from $^{15}N$ in the product ion. In this case, the observed differences between low-resolution and resolution-dependent MS/MS correction are much smaller. Figures A3 and A4 in the appendix show the correction of additional MIDs for the alanine and acetyl-CoA examples used in this section. The heatmaps in Figures A5 and A6 in the appendix show a comparison of low-resolution MS/MS correction with resolution-dependent MS/MS correction for (low-resolution) MS/MS data from the $^{13}C$ stable isotope labeling experiment in P493-6 B-cells which were already used as an example in 3.2.1 and 3.2.3. The heatmaps were generated by subtracting the MIDs from low-resolution correction from the MIDs resulting from resolution-dependent correction. Here, maximum differences between low-resolution and

resolution-dependent correction in the range of 0.005 to 0.009 arise.

There is a very important practical aspect regarding resolution-dependent MS/MS correction that has not been considered in the analyses presented so far. The resolution of orbitrap and FT-ICR mass analyzers depends linearly on the transient duration, that is, the time that ion oscillations are measured in the mass analyzer to subsequently generate a single mass spectrum via fourier transformation (Zubarev & Makarov (2013)). In full MS mode, transient lengths sufficient to provide high resolution are usually attainable also in combination with online chromatographic separation (Su et al. (2017)). However, in the case of tandem MS, a high resolution spectrum has to be acquired for the fragments of each precursor ion, splitting the transient duration available in full MS by the total number of precursors measured in the cycle. When measuring isotopologues from stable isotope labeling experiments, there are usually many precursors, easily dividing the time available for fragment spectra acquisition by a factor of 25 or more if the number of points across the chromatographic peak is to be kept constant. Thus, a resolution of 140000 (200 $m/z$) would diminish to about 5500. Options to increase resolution may be a reduction of points across the peak or reducing the total number of measured precursor ions according to where they are expected in the chromatographic separation, so that the mass spectrometer does not always cycle over all possible precursors. However, even if only one 5 carbon metabolite is expected at a given retention time in a $^{13}C$ labeling experiment, this would still mean that 6 precursors have to be measured to cover all of the MID. If, for example, decreasing the number of points across the peak by a factor of 2 and decreasing the number of precursors to 10, this would yield a resolution of about 28000. This raises an essential question regarding resolution-dependent MS/MS correction: Is this mode of correction needed given the drop of resolution in orbitrap and FT-ICR devices operating in MS/MS mode, or is low-resolution MS/MS correction actually sufficient in the expected resolution range of roughly about 5000 to 50000 (200 $m/z$)? To address this issue, results of resolution-dependent MS/MS correction assuming a low-resolution precursor mass analyzer (1000 at $m/z$ 200, $FWHM$ constant over the $m/z$ range) and a high resolution orbitrap product ion mass analyzer operating at different resolutions (140000, 80000, 50000, 20000, 10000 or 5000 at $m/z$ 200) were compared to results of low-resolution MS/MS correction. This was performed for the simulated labeling data already discussed in this section, from $^{13}C$ labeled TMS- and PCF-derivatized alanine and $^{15}N$ labeled acetyl-CoA. Underivatized alanine (Precursor: $C_3H_6NO_2^-$, product ion: $C_2H_6N^-$, neutral loss: $CO_2$) was also added to the comparison as an example for a low $m/z$ species.

The heatmap in Figure II.18 shows the $log_{10}$ transformed differences when subtracting MIDs resulting from low-resolution MS/MS correction from MIDs that are obtained via resolution-dependent correction when assuming a resolution of 140000 ($m/z$ 200) for the orbitrap analyzer. This is the same resolution that was also used for most of the previous analyses, so Figure II.18 mainly shows results previously depicted as bar plots as a difference heatmap. As previously discussed, the highest differences between low-resolution and resolution-dependent correction are found for TMS-derivatized alanine, mainly due to the ability or inability to resolve Si isotopes from $^{13}C$ (maximum difference: 0.024). The differences for the other molecules are mostly in the range of 1E-04 to 3E-03. Figure II.19 shows the same comparison if orbitrap resolution diminishes to 50000. Here, the differences for TMS-derivatized alanine are still high (maximum difference: 0.0235), and also the differences for PCF-derivatized and underivatized alanine remain largely on the same scale as with a resolution of 140000. For the high $m/z$ molecule acetyl-CoA, however, differences decrease to a negligible magnitude. This is likely because of the $m/z$ dependency of resolution in orbitrap devices, lowering the actual resolution

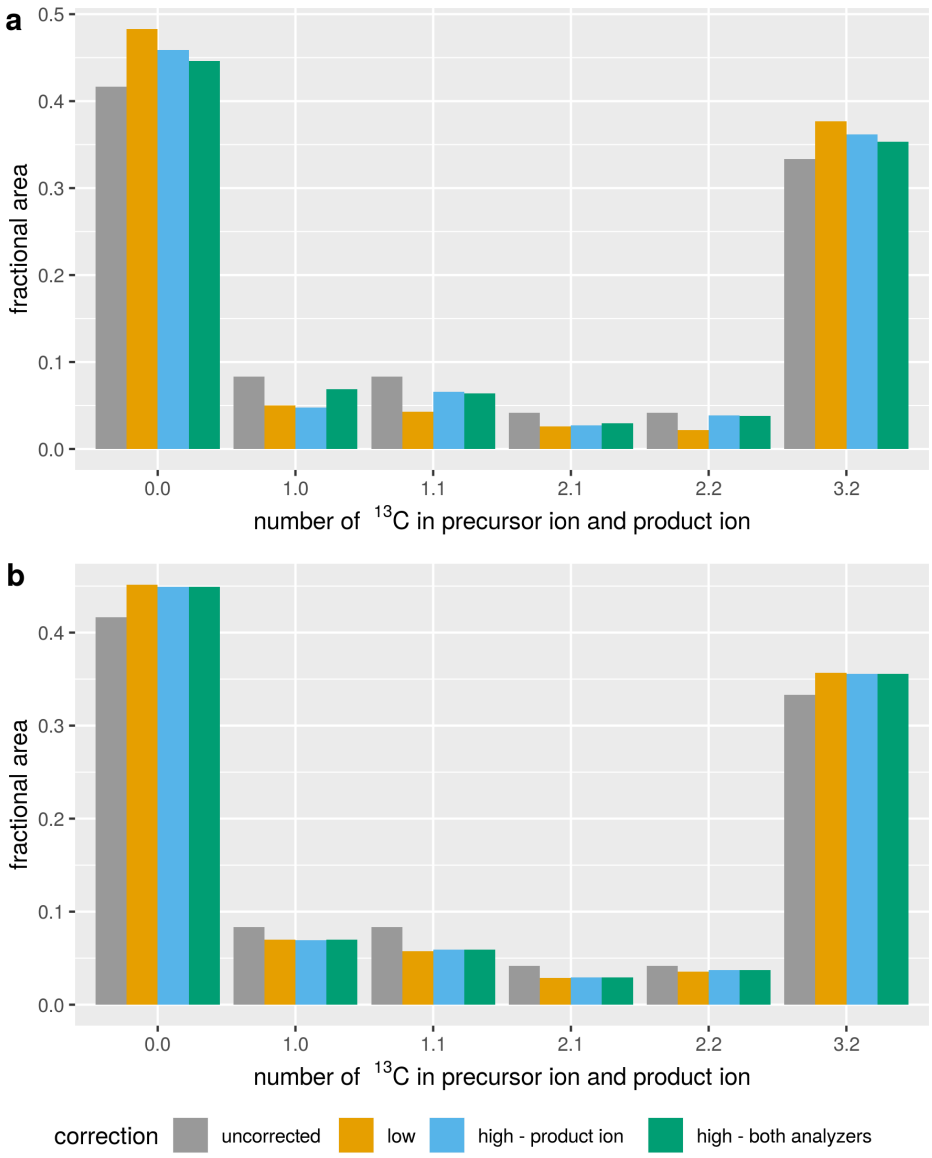



Figure II.16: ***Illustration of the impact of resolution-dependent MS/MS correction on example data.*** *The figure shows the results obtained from correcting simulated MS/MS data of TMS- (a) and PCF-derivatized (b) alanine. The uncorrected MID is the same for both molecules. The x-axis labels n.m correspond to the MS/MS transitions with n $^{13}C$ in the precursor ion and m $^{13}C$ in the product ion. Data are shown uncorrected, corrected with IsoCorrectoR's low-resolution MS/MS correction algorithm (low) and corrected with IsoCorrectoR's resolution-dependent MS/MS correction algorithm (high). Resolution-dependent correction was performed for the usual case of having a low-resolution precursor mass analyzer (1000 at m/z 200, FWHM constant over the m/z range) and a high resolution product ion mass analyzer (140000 at m/z 200, orbitrap m/z dependency of FWHM), labeled "high - product ion", and for the theoretical case of both analyzers operating at high resolution, labeled "high - both analyzers".*

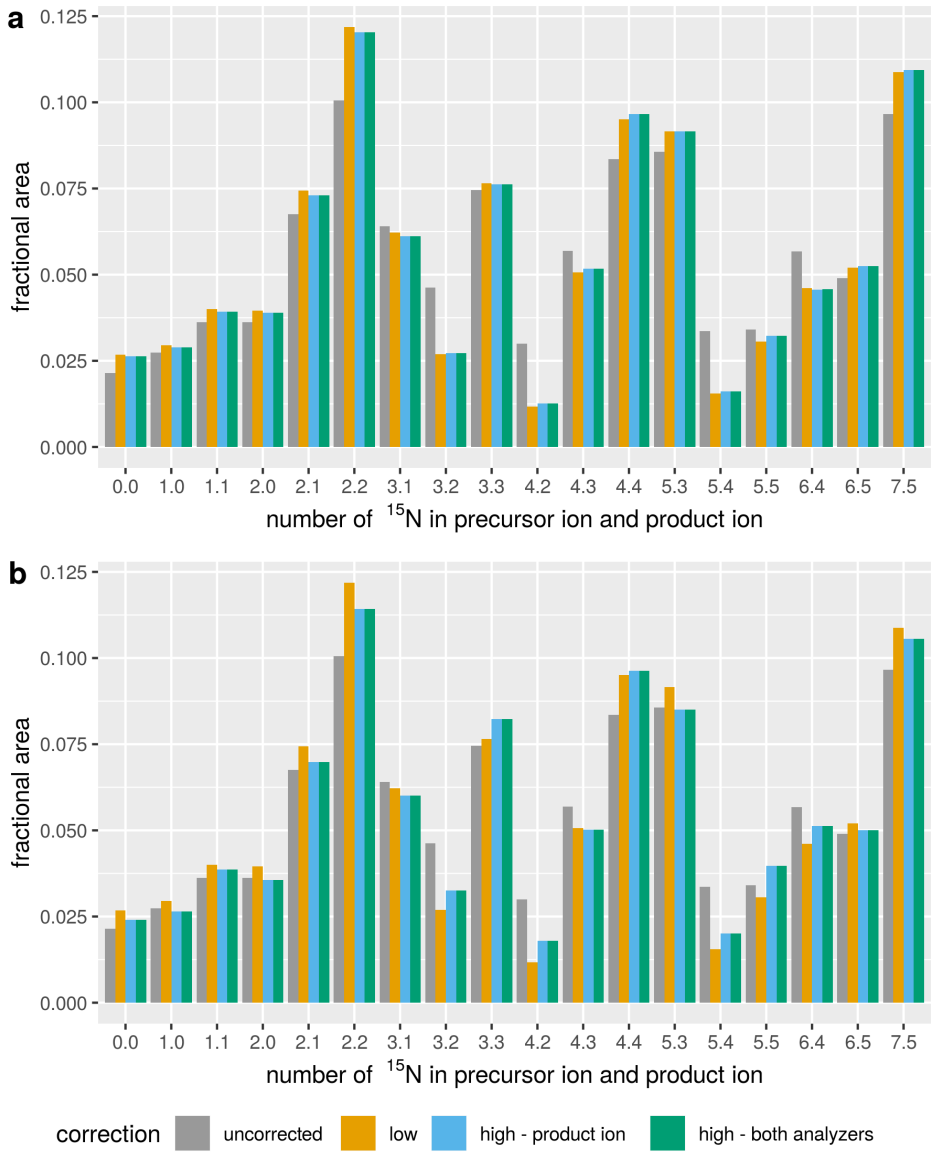



Figure II.17: ***Illustration of the impact of resolution-dependent MS/MS correction on example data from $^{15}N$ labeled underivatized acetyl-CoA.*** *Precursor: $C_{23}H_{39}N_7O_{17}P_3S^+$, product ion: $C_{10}H_{16}N_5O_{10}P_2^+$ (3'-phosphonucleoside monophosphate), neutral loss: $C_{13}H_{23}N_2O_7PS$. The x-axis labels n.m correspond to the MS/MS transitions with n $^{15}N$ in the precursor ion and m $^{15}N$ in the product ion. Data are shown uncorrected, corrected with IsoCorrectoR's low-resolution MS/MS correction algorithm (low) and corrected with IsoCorrectoR's resolution-dependent MS/MS correction algorithm (high). Resolution-dependent correction was performed for the usual case of having a low-resolution precursor mass analyzer (1000 at m/z 200, FWHM constant over the m/z range) and a high resolution product ion mass analyzer, labeled "high - product ion", and for the theoretical case of both analyzers operating at high resolution, labeled "high - both analyzers". Panels a and b correspond to different resolutions of the high-resolution orbitrap mass analyzer(s): a) 140000 at m/z 200, b) 200000 at m/z 200. At the lower of the two resolutions, $^{13}C$ and $^{15}N$ can't be resolved for the acetyl-CoA product ion (nominal m/z of 428). Even at a resolution of 200000, the two isotopes cannot be resolved for the precursor (nominal m/z of 810).*

at the acetyl-CoA product ion $m/z$ of 428 (nominal, unlabeled) to about 34000. However, the results for species with lower $m/z$ show that resolution-dependent correction can still be very relevant also at a resolution of 50000. A similar heatmap for a resolution of 80000 can be found in Figure A8 in the appendix. At an orbitrap resolution of 20000, shown in Figure II.20, the differences for acetyl-CoA and PCF-derivatized alanine drop to 0. TMS-derivatized alanine still shows a maximum difference of 0.0156, likely because $^{30}$Si can still be resolved, even if $^{29}$Si cannot. For underivatized alanine, the differences are similar to those at higher resolutions, most probably because the actual resolution for the alanine product ion with an $m/z$ of only 44 (nominal, unlabeled) in the orbitrap is around 40000. If resolution diminishes further to 10000 (see Figure A8 in the appendix), differences between low-resolution and resolution-dependent MS/MS correction remain only for underivatized alanine, which still shows differences comparable to higher resolutions (maximum difference: 1.4E-03). At a resolution of 5000, which is essentially low-resolution, all differences become 0, also for underivatized alanine.

Thus, for the example molecules considered, even at a resolution of 10000 for determining product ion $m/z$, there is still a difference between low-resolution and resolution-dependent MS/MS correction. This difference is relatively small, constrained to the low $m/z$ product ion of underivatized alanine, and may or may not be critical in practice at this magnitude, but it clearly cannot be considered negligible. When resolution increases to 20000, differences become more marked, and even reach the order of magnitude observed at a resolution of 140000 for TMS-derivatized alanine. This is most likely due to the presence of $^{30}$Si in the molecule, which is resolved from (2) $^{13}$C relatively easily. At a resolution of 50000, the differences observed for the small or medium sized metabolites are clearly in the range of those at a resolution of 140000. Resolutions of 140000 or higher are only realistic in the tandem MS measurement of MIDs when operating in direct infusion mode, and most likely not in combination with online chromatography. However, the resolutions that may be attainable in that setup can still be high enough to make resolution-dependent MS/MS correction recommendable. Higher differences between low-resolution and resolution-dependent correction may be constrained to specific cases, *e.g.* the presence of Si in the molecule or molecules with low $m/z$, but these cases do occur in practice. Another aspect to consider following this analysis is the relevance of resolution-dependent MS and MS/MS correction for TOF devices. These operate roughly in the resolution range of 10000 to 50000. For this resolution range, the potential relevance of resolution-dependent MS/MS correction has been shown in the orbitrap case. When attempting to transfer these insights to TOF devices, it has to be considered, however, that, other than for orbitrap or FT-ICR mass analyzers, TOF resolution does not increase with lower $m/z$, but is relatively constant throughout the $m/z$ range (Zubarev & Makarov (2013)). Thus, similar differences for low $m/z$ species (*e.g.*, underivatized alanine) likely would not be seen in the case of a TOF mass analyzer. Regarding the relevance of resolution-dependent MS/MS correction, a potential increase in the resolution capabilities of orbitrap devices in the future should be kept in mind. The reduction of orbitrap dimensions and the application of enhanced fourier transformation have already brought major improvements in this area, and further developments in engineering and/or computation may increase resolution even more (Zubarev & Makarov (2013), Eliuk & Makarov (2015)).

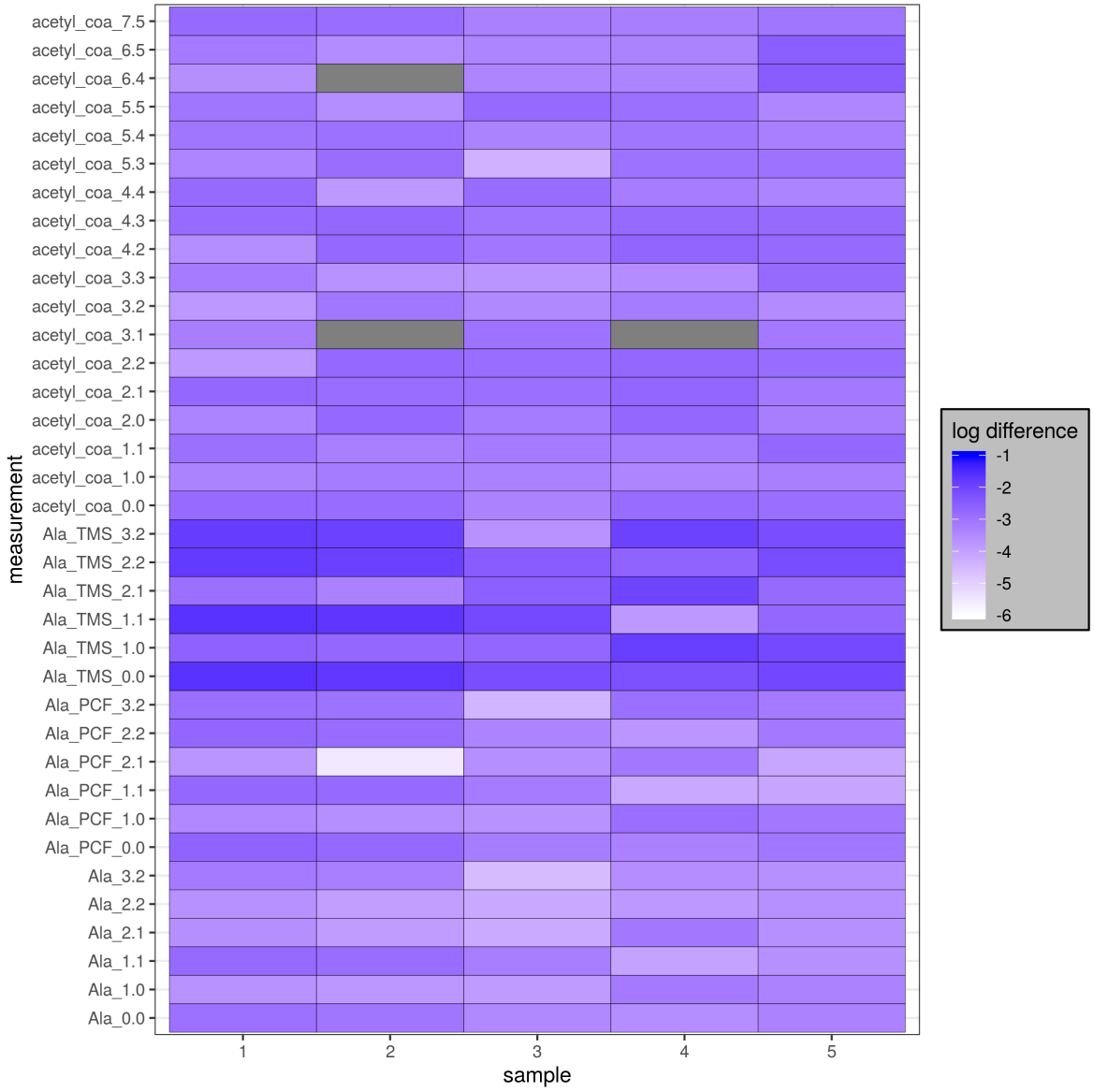



Figure II.18: ***Heatmap showing differences between low-resolution and resolution-dependent MS/MS correction (product ion mass analyzer resolution of 140000 at m/z 200).*** *Simulated $^{13}C$ labeling data of underivatized and PCF-/TMS-derivatized alanine (Ala, Ala_PCF, Ala_TMS), simulated $^{15}N$ labeling data of acetyl-CoA. A tracer purity of 0.99 was assumed. The heatmap was generated by subtracting MIDs corrected via IsoCorrectoR's low-resolution MS/MS correction approach from MIDs corrected via the resolution-dependent approach. For resolution-dependent MS/MS correction, the precursor mass analyzer was assumed to operate at a resolution of 1000 at m/z 200 (FWHM constant over the m/z range), while the product ion mass analyzer was assumed to be an orbitrap operating at a resolution of 140000 at m/z 200. The (absolute) differences were $log_{10}$ transformed, gray fields indicate an untransformed difference of 0 (equality of values). Columns represent samples, while rows represent measurements of MS/MS transitions. The numbers n.m after the metabolite names correspond to the MS/MS transitions with n label in the precursor ion and m label in the product ion.*

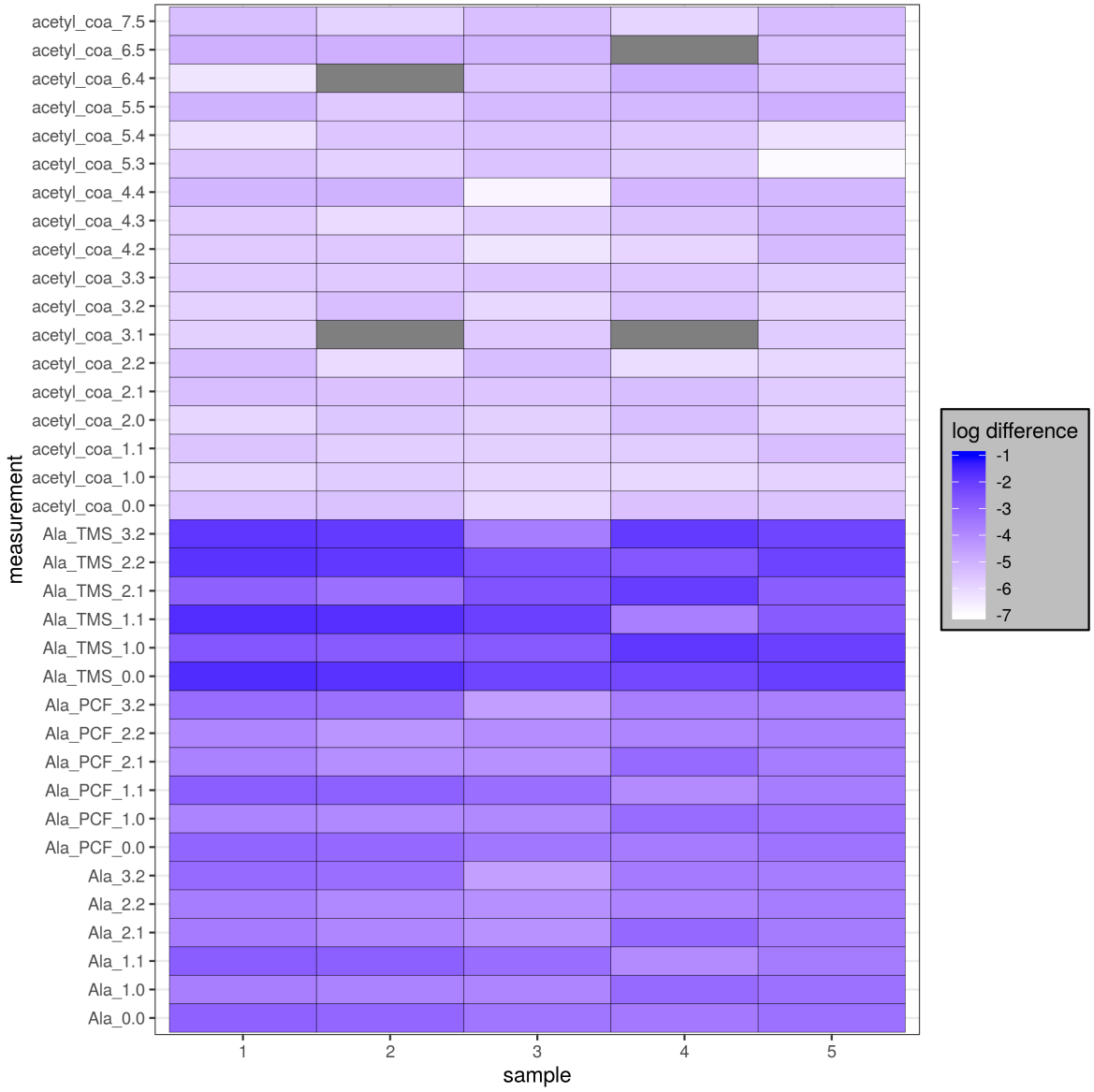



Figure II.19: ***Heatmap showing differences between low-resolution and resolution-dependent MS/MS correction (product ion mass analyzer resolution of 50000 at m/z 200).*** *Simulated $^{13}C$ labeling data of underivatized and PCF-/TMS-derivatized alanine (Ala, Ala_PCF, Ala_TMS), simulated $^{15}N$ labeling data of acetyl-CoA. A tracer purity of 0.99 was assumed. The heatmap was generated by subtracting MIDs corrected via IsoCorrectoR's low-resolution MS/MS correction approach from MIDs corrected via the resolution-dependent approach. For resolution-dependent MS/MS correction, the precursor mass analyzer was assumed to operate at a resolution of 1000 at m/z 200 (FWHM constant over the m/z range), while the product ion mass analyzer was assumed to be an orbitrap operating at a resolution of 50000 at m/z 200. The (absolute) differences were $log_{10}$ transformed, gray fields indicate an untransformed difference of 0 (equality of values). Columns represent samples, while rows represent measurements of MS/MS transitions. The numbers n.m after the metabolite names correspond to the MS/MS transitions with n label in the precursor ion and m label in the product ion.*

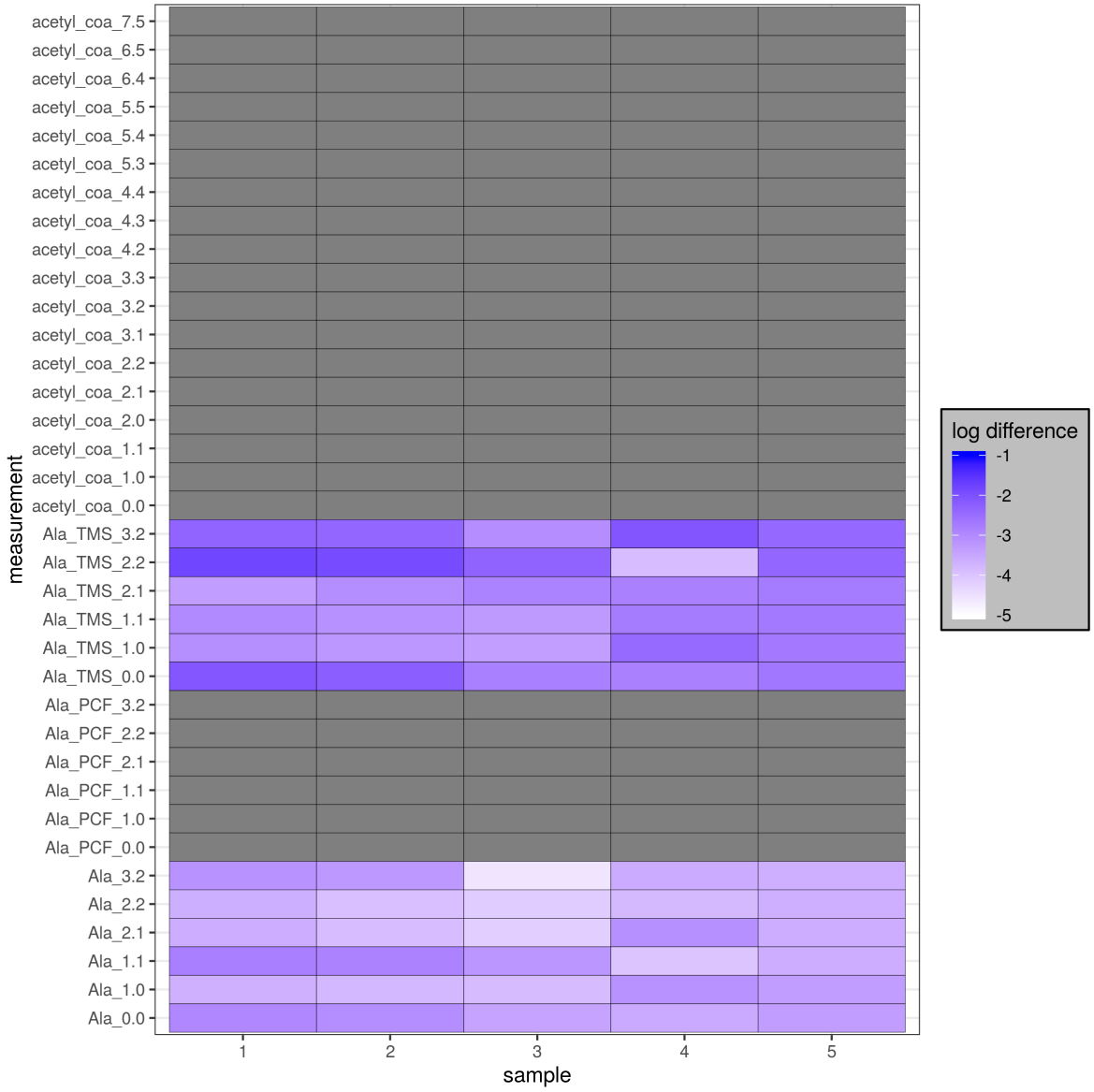



Figure II.20: ***Heatmap showing differences between low-resolution and resolution-dependent MS/MS correction (product ion mass analyzer resolution of 20000 at m/z 200).*** *Simulated $^{13}C$ labeling data of underivatized and PCF-/TMS-derivatized alanine (Ala, Ala_PCF, Ala_TMS), simulated $^{15}N$ labeling data of acetyl-CoA. A tracer purity of 0.99 was assumed. The heatmap was generated by subtracting MIDs corrected via IsoCorrectoR's low-resolution MS/MS correction approach from MIDs corrected via the resolution-dependent approach. For resolution-dependent MS/MS correction, the precursor mass analyzer was assumed to operate at a resolution of 1000 at m/z 200 (FWHM constant over the m/z range), while the product ion mass analyzer was assumed to be an orbitrap operating at a resolution of 20000 at m/z 200. The (absolute) differences were $log_{10}$ transformed, gray fields indicate an untransformed difference of 0 (equality of values). Columns represent samples, while rows represent measurements of MS/MS transitions. The numbers n.m after the metabolite names correspond to the MS/MS transitions with n label in the precursor ion and m label in the product ion.*

### 3.2.6 Correction of ultra-high-resolution MS data from multiple-tracer experiments

As already mentioned in previous sections, resolutions exceeding 400000 (at 400 *m/z*) may be considered ultra-high resolution. Such resolutions are attainable by FT-ICR and orbitrap devices, given sufficient transient duration. It has been stated in the literature (Moseley (2010), Carreer et al. (2013)) that ultra-high resolution should be sufficient to reliably resolve mass defect related differences between the incorporation of isotopes that produce the same nominal mass shift. Thus, under this assumption, correcting for NA contributions of isotopes other than the tracer isotope would not be necessary (Moseley (2010)). Further, this level of resolution provides the opportunity to perform stable isotope labeling experiments where multiple tracer isotopes (*e.g.*, $^{13}$C and $^{15}$N) are used simultaneously. IsoCorrectoR is capable of correcting UHR data from such multiple-tracer experiments, under the assumption that the incorporation of all isotopes except the tracer isotope(s) can be resolved spectrometrically. This is similar to PyNAC (Carreer et al. (2013)), a tool specifically designed for UHR correction. However, IsoCorrectoR can additionally correct the data for the purity of the tracer substrates, which is not possible when using PyNAC.

Figure II.21 shows the UHR correction of example data from PCF-derivatized asparagine from a simulated $^{13}$C and $^{15}$N stable isotope labeling experiment (uncorrected data were devised by hand). Data has been corrected without considering tracer purity (100% purity) and assuming 99% or 98% purity (for both tracer elements, it would also be possible to set a different value for each tracer). When considering the species C3.N0 (3 $^{13}$C, 0 $^{15}$N) and C3.N1 (3 $^{13}$C, 1 $^{15}$N) in panel a, the differences between performing and not performing tracer purity correction are especially evident. Without tracer purity correction, the corrected value of the C3.N0 species is about 145% of the corrected value at 98% purity. For the C3.N1 species, it is 190%. Thus, given the tracer elements had the commonly observed purity of 98%, not accounting for tracer purity would lead to substantially different results for some species. Assuming 99% purity, the results would still deviate remarkably. In the case of the given MID, the reason for the magnitude of deviations are mainly tracer purity contributions of the C3.N2 ($^{15}$N impurity) and C4.N1 ($^{13}$C impurity) species to C3.N1, as well as the impurity contribution of C4.N0 to C3.N0. As the abundance of the contributing labeling states is high in relation to the receiving labeling states, the relative magnitude of tracer purity contributions is also high. The example of C3.N1 further illustrates how impurity contributions of multiple tracer elements can add up in a multi-tracer experiment. Thus, being able to perform tracer purity correction should be especially crucial in such a scenario. Panel b shows the same analysis for a different MID. In this case, the species that mainly provide tracer purity contributions are less than or equally abundant as the species receiving contributions, resulting in less pronounced differences between performing and not performing impurity correction. Figure A9 in the appendix shows the UHR correction of PCF-asparagine for two additional MIDs.

To test whether it is a viable approximation to correct only for the NA of the tracer isotope(s) and not for contributions of any other isotopes in UHR correction, UHR correction was compared to the application of resolution-dependent correction with resolutions that are assumed to be ultra-high. To this end, example datasets provided with AccuCor (Su et al. (2017)) were used. One of the datasets was derived from a $^{13}$C labeling experiment, while the other comes from a $^{15}$N labeling experiment. Data were acquired with an orbitrap mass spectrometer at a resolution of 140000 (200 *m/z*) in negative mode. IsoCorrectoR's UHR correction algorithm, as well as IsoCorrectoR's algorithm for resolution-dependent correction

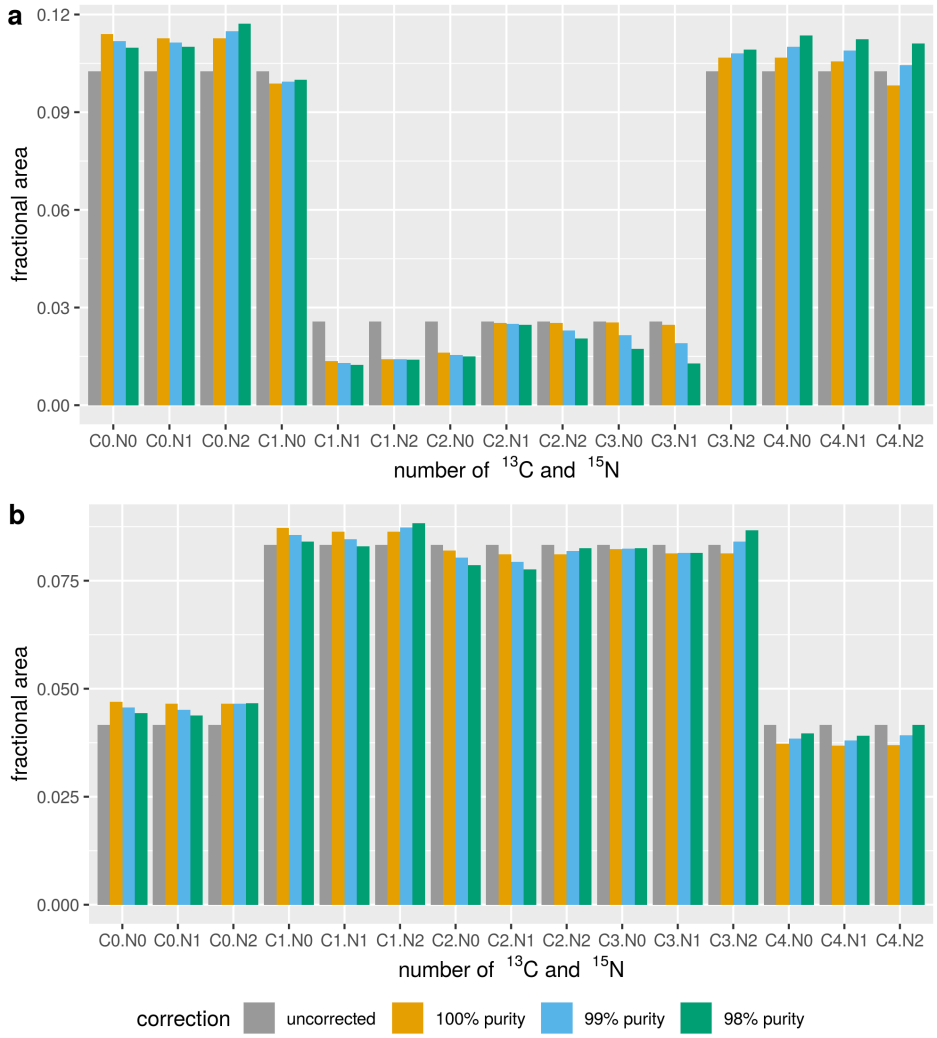



Figure II.21: ***Illustration of the impact of UHR (and tracer purity) correction on simulated example data from simultaneously $^{13}C$ and $^{15}N$ labeled PCF-derivatized asparagine.*** *The x-axis labels Cx.Ny correspond to x $^{13}C$ label and y $^{15}N$ label incorporated in the respective isotopologue. Data are shown uncorrected and corrected with IsoCorrectoR's UHR correction algorithm at different assumed purities of the tracer substrate. The panels a and b correspond to different uncorrected MIDs.* ***(Panel a taken from Heinrich et al. (2018), slightly altered)***

were applied to the data (assumed tracer purity 99%). For resolution-dependent correction, an FT-ICR mass analyzer with a resolution of either 400000 or 800000 at 400 $m/z$ was assumed (the $FWHM$-dependency on $m/z$ is given by Equation II.10). In order to obtain "true" corrected values for the given datasets, resolution-dependent correction would actually have to be performed assuming an orbitrap mass analyzer with a resolution of 140000 at 200 $m/z$. However, the "true" corrected values are not of interest in this analysis. Only the difference between the two correction approaches is relevant, so applying the "correct" resolution for correction is not required. Of course, overcorrection resulting in correction residuals (see 3.1.7 in the algorithm section) may be a problem when using the wrong resolution for correction, but since the resolutions used for correction are higher than the resolution at which the data were acquired, this is not an issue.

The heatmap in Figure II.22 shows the $log_{10}$-transformed differences between UHR correction and resolution-dependent correction (resolution 400000 at 400 $m/z$) for a set of metabolites from the $^{13}$C labeling dataset. The maximum difference is in the range of 5.34E-03 (0.00534). This is not extremely high, but might already be relevant to biological interpretation in some cases. Differences between corrected MIDs that get as high as the third decimal place are found only for heavier metabolites like NADH, sedoheptulose-7-phosphate or fructose-1,6-bisphosphate (nominal molecular masses of 664, 290 and 340, respectively). For the smaller metabolites, maximum differences are in the range of 2E-05 (4.5E-11 for 3-phosphoglycerate). At the higher $m/z$ values, $^{17}$O is usually unresolved from $^{13}$C: The mass shift difference between $^{17}$O and $^{13}$C is 0.0008623, while the minimal resolvable mass difference for sedoheptulose-7-phosphate (m+0) is 0.0008725). This should be responsible for most of the observed differences. Further, at $m/z$ values as high of those of NADH, $^{2}$H should mostly be unresolved, too, which explains why differences are most pronounced for that metabolite (even though the NA of $^{2}$H is relatively low, the presence of 29 H atoms in NADH can still result in notable differences). At a resolution of 800000 (400 $m/z$, FT-ICR), also $^{17}$O can usually be resolved from $^{13}$C for all molecules except NADH. The heatmap in Figure II.23 shows the associated differences between UHR and resolution-dependent correction. The maximum difference is in the range of 6.6E-04, which is roughly an order of magnitude lower than at a resolution of 400000. Differences are substantially less pronounced in general, while NADH still shows the highest deviations. In the cases where $^{17}$O can normally be resolved, the remaining deviations are likely caused by NA contributions where mass defect related differences between isotope mass shifts cancel each other out (differences for fructose-1,6-bisphosphate are presumably more pronounced than for sedoheptulose-7-phosphate because its $\delta_{min}$ is higher due to the higher $m/z$). These contributions are corrected for in the resolution-dependent, but not in the UHR approach.

The deviations associated with UHR correction compared to resolution-dependent correction on the $^{13}$C labeling dataset can not be considered minor. However, they should still be mostly acceptable in practice, even at a resolution of 400000 (400 $m/z$). However, a different picture emerges when considering the heatmap associated with the comparison of UHR to resolution-dependent correction on the $^{15}$N labeling dataset (Figure II.24, no log-transformation). When assuming an FT-ICR analyzer with a resolution of 400000 (400 $m/z$) for resolution-dependent correction, the maximum difference between corrected MIDs gets as high as 0.217 for acetyl-CoA. Clearly, deviations on that scale should have a tremendous impact on any effort regarding biological interpretation of the data. Finding such pronounced differences is, however, reasonable, as $^{13}$C NA cannot be resolved from the incorporation of the $^{15}$N tracer isotope at an $m/z$ as high as 808 for the acetyl-CoA molecular ion (mass shift difference of 0.00632 vs. minimal

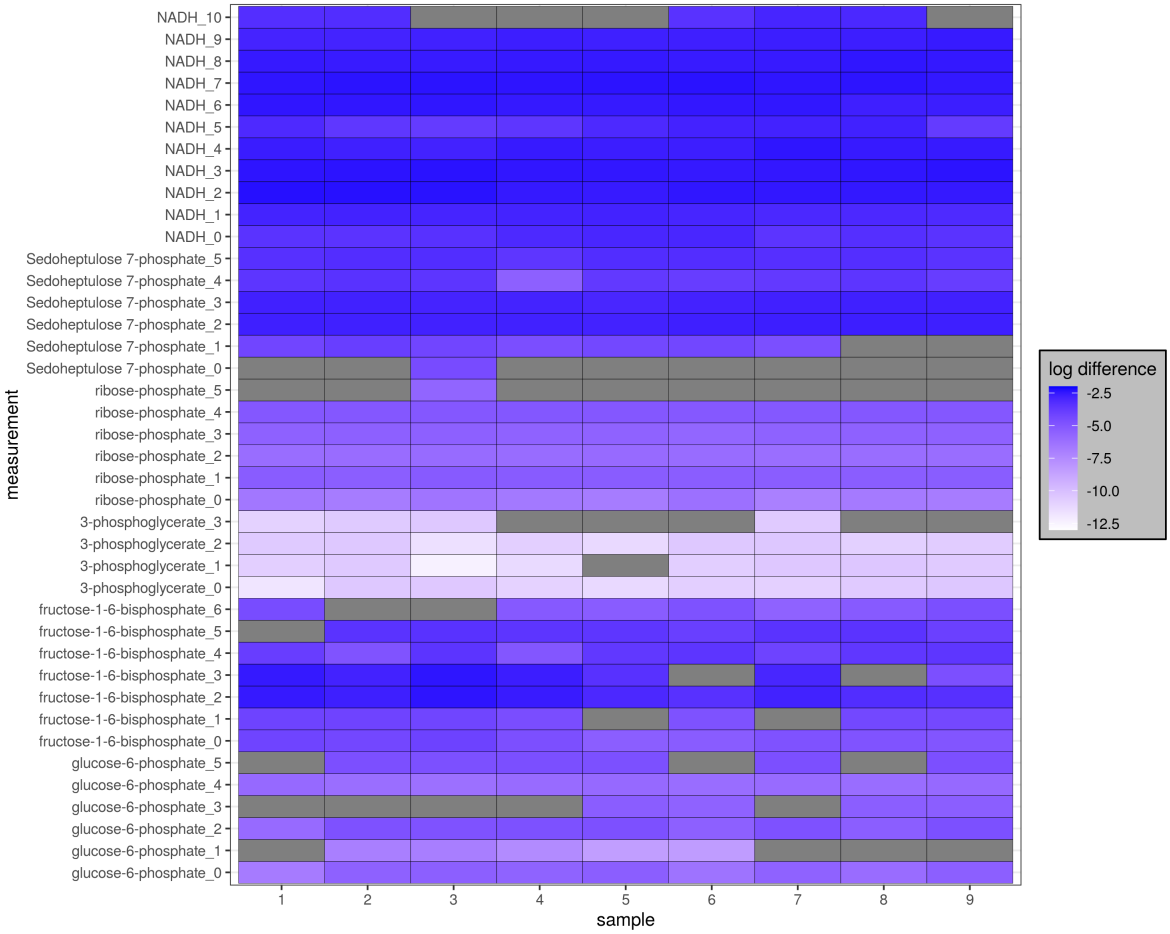



Figure II.22: ***Heatmap showing differences between UHR and resolution-dependent correction of data from a $^{13}C$ labeling experiment with IsoCorrectoR.*** *Data used are example data supplied with AccuCor (Su et al. (2017)). Correction was performed assuming a tracer purity of 0.99. An FT-ICR mass analyzer with a resolution of 400000 at 400 m/z was assumed for resolution-dependent correction. The heatmap was generated by subtracting MIDs obtained by UHR correction from MIDs obtained through resolution-dependent correction. The (absolute) differences were $log_{10}$ transformed, gray fields indicate an untransformed difference of 0 (equality of values). Columns represent samples, while rows represent measurements of mass isotopomers. The number after the metabolite name corresponds to the number of label incorporated.*

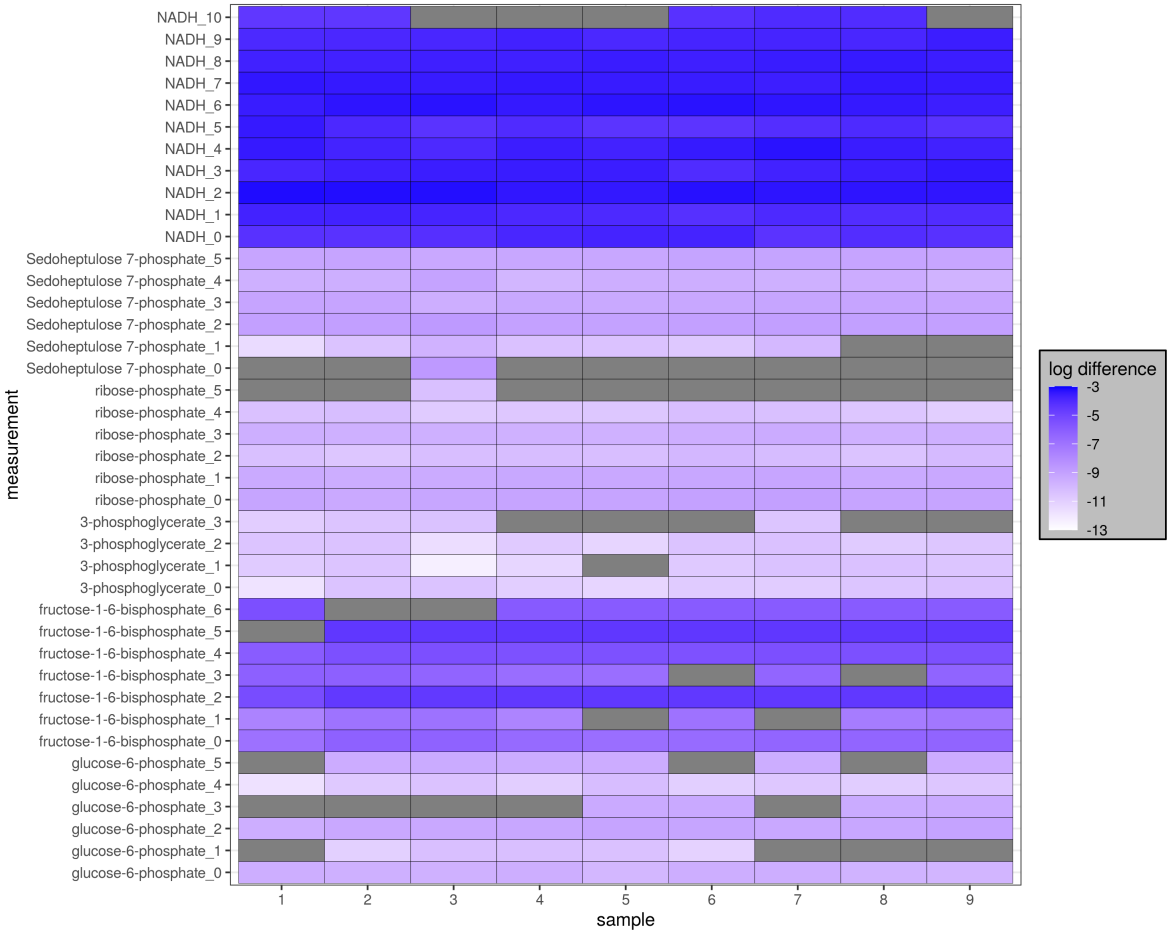



Figure II.23: ***Heatmap showing differences between UHR and resolution-dependent correction of data from a $^{13}C$ labeling experiment with IsoCorrectoR.*** *Data used are example data supplied with AccuCor (Su et al. (2017)). Correction was performed assuming a tracer purity of 0.99. An FT-ICR mass analyzer with a resolution of 800000 at 400 m/z was assumed for resolution-dependent correction. The heatmap was generated by subtracting MIDs obtained by UHR correction from MIDs obtained through resolution-dependent correction. The (absolute) differences were $log_{10}$ transformed, gray fields indicate an untransformed difference of 0 (equality of values). Columns represent samples, while rows represent measurements of mass isotopomers. The number after the metabolite name corresponds to the number of label incorporated.*

resolvable mass difference of 0.00677). Further, with 23 carbons present in the molecule, the influence of $^{13}$C NA should be substantial. For glutathione disulfide, which also shows marked deviations (maximum: 0.063), $^{13}$C can usually be resolved. However, resolution is not sufficient to resolve $^{15}$N from $^{33}$S and 2 $^{15}$N from $^{34}$S. As the NA of both isotopes (especially $^{34}$S) is high, the observed magnitude of differences seems reasonable. Assuming a resolution of 800000 (400 *m/z*, Figure II.25), the maximum difference between UHR and resolution-dependent correction diminishes to 0.0573 (glutathione disulfide). $^{13}$C can be resolved from $^{15}$N for acetyl-CoA at this resolution, however, $^{34}$S still cannot be resolved from $^{15}$N for both acetyl-CoA and glutathione disulfide. Thus, even at a resolution of 800000, deviations are still so high that they are unacceptable in practice. Figures A10 and A11 in the appendix show the heatmaps of the $^{15}$N labeling analysis in log-scale.

Judging from this analysis, UHR correction should be performed with extreme care. The assumption that it is possible to spectrometrically resolve all non-tracer isotopes is often not valid at high *m/z* values. Further, the mass defect related mass shift differences between the incorporation of some isotopes are so small that they are often not resolved (*e.g.*, $^{13}$C vs. $^{17}$O, or $^{15}$N vs. $^{33}$S or $^{34}$S). Thus, if stable isotope labeling data were acquired with a (ultra-) high-resolution device, it is strongly recommended to employ resolution-dependent correction instead of UHR correction. Until very recently, there was no solution that offered resolution-dependent correction for multiple-tracer data, and UHR correction was the only option in these cases. To make sure that the correction error is still acceptable when applying UHR correction, it is advised to either assess correction using unlabeled standards of molecules with higher *m/z* values, or to manually calculate whether especially abundant isotopes like $^{13}$C, $^{33}$S or $^{34}$S can be resolved at the given molecule *m/z*. This can be done for orbitrap and FT-ICR devices based on known exact isotope masses and the formulas presented in Equations II.9 and II.10. However, with AccuCor2, a first tool for resolution-dependent correction of multiple-tracer data has now been developed, and it is recommended to use this software instead of applying UHR correction. In the manuscript presenting AccuCor2, comparisons between resolution-dependent and UHR correction result in similar conclusions as those presented in this thesis (Wang et al. (2021)). Further, erroneous assumptions that have been made in the correction algorithm of the single-tracer resolution-dependent correction tool AccuCor (Su et al. (2017)), and that have been pointed out in Millard et al. (2019), as well as in section 3.1.4 in this thesis, have been revised in AccuCor2. Yet, even though AccuCor2 clearly seems to be an appropriate tool for the correction of multiple-tracer data, making sure that the different tracer isotopes employed (*e.g.*, $^{13}$C together with $^{15}$N) can generally be resolved from each other is still the primary prerequisite for successfully carrying out a multiple-tracer experiment. If this mandatory condition is not fulfilled, using even the most sophisticated correction algorithm will not yield reasonable results.

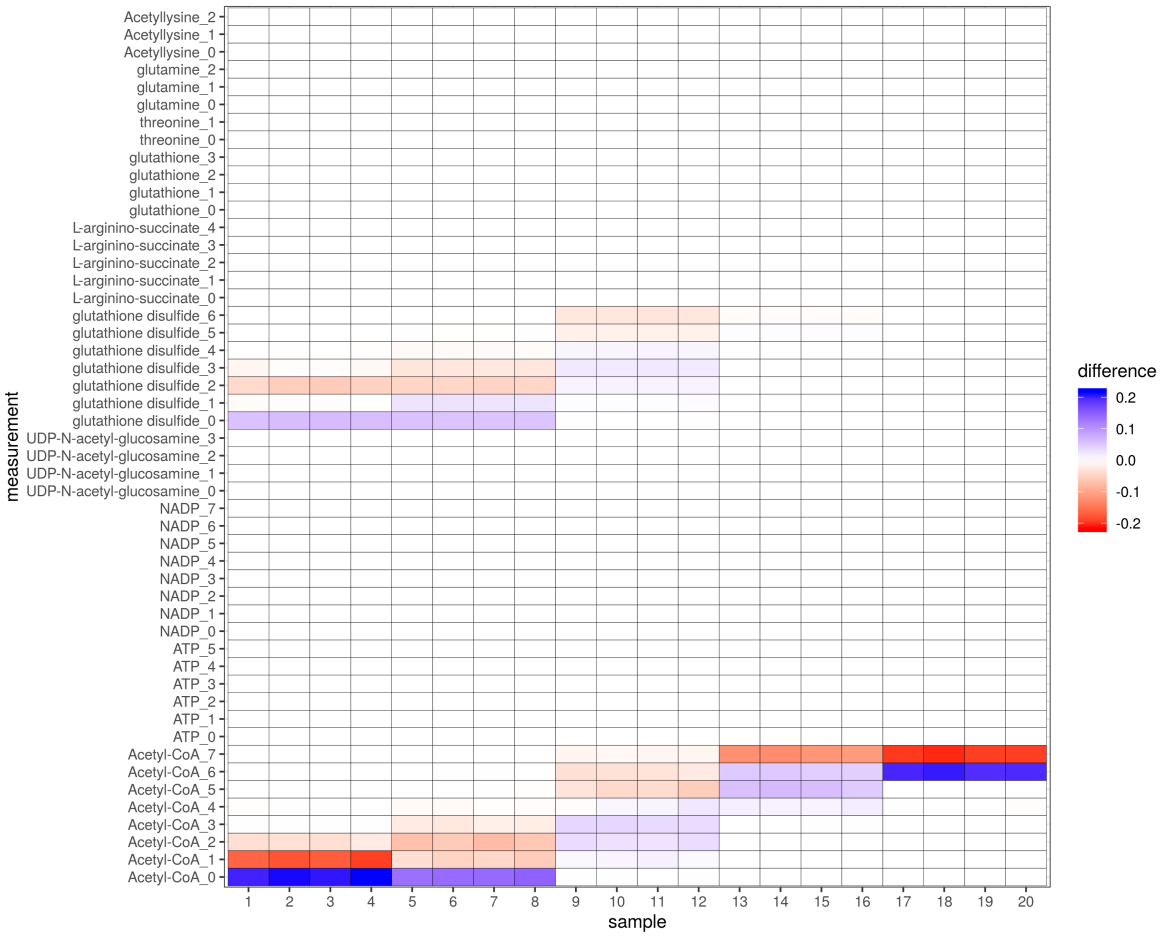



Figure II.24: ***Heatmap showing differences between UHR and resolution-dependent correction of data from a $^{15}N$ labeling experiment with IsoCorrectoR.*** *Data used are example data supplied with AccuCor (Su et al. (2017)). Correction was performed assuming a tracer purity of 0.99. An FT-ICR mass analyzer with a resolution of 400000 at 400 m/z was assumed for resolution-dependent correction. The heatmap was generated by subtracting MIDs obtained by UHR correction from MIDs obtained through resolution-dependent correction. Columns represent samples, while rows represent measurements of mass isotopomers. The number after the metabolite name corresponds to the number of label incorporated.*

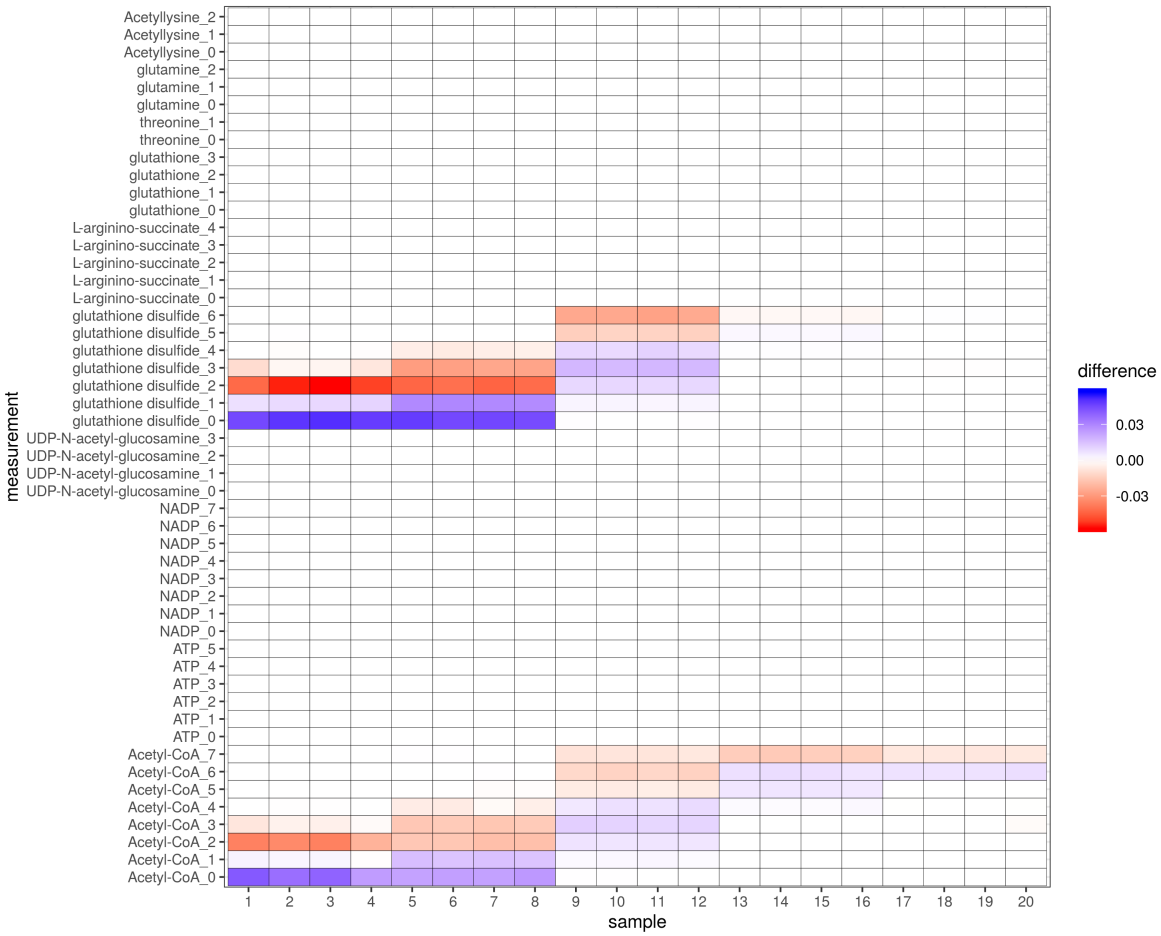



Figure II.25: ***Heatmap showing differences between UHR and resolution-dependent correction of data from a $^{15}$N labeling experiment with IsoCorrectoR.*** *Data used are example data supplied with AccuCor (Su et al. (2017)). Correction was performed assuming a tracer purity of 0.99. An FT-ICR mass analyzer with a resolution of 800000 at 400 m/z was assumed for resolution-dependent correction. The heatmap was generated by subtracting MIDs obtained by UHR correction from MIDs obtained through resolution-dependent correction. Columns represent samples, while rows represent measurements of mass isotopomers. The number after the metabolite name corresponds to the number of label incorporated.*

## 3.3 Validation of IsoCorrectoR

This section deals with the validation of the different correction modes of IsoCorrectoR. IsoCorrectoR was validated by comparison to other tools used for NA correction, by comparison to manual calculations and experimentally, through validation mixtures of known isotopologue composition.

### 3.3.1 Low-resolution MS correction

**Comparison to IsoCor** IsoCorrectoR's low-resolution MS correction approach was compared to that of IsoCor v2 using different datasets. The first dataset originates from a $^{13}C$ stable isotope labeling experiment conducted in a P493-6 B-cell line (see 3.1 in the methods chapter). These data were already used in section 3.2.1. Cells were fed with U-$^{13}C$-glutamine, and the incorporation of $^{13}C$ into different metabolites was assessed via LC-MS/MS with a low-resolution triple-quadrupole device. The amino acids glycine, ornithine, proline and serine were measured in tandem MS, however, the neutral loss cannot contain label. As a consequence, it is valid to perform MS correction on the data. The heatmap in Figure II.26 covers data from these amino acids in 26 samples, including cell extracts, supernatants and medium controls. It was generated by subtracting MIDs corrected by IsoCor v2 from MIDs corrected by IsoCorrectoR, assuming a tracer purity of 0.99. The maximum difference found is marginal (in the range of 8E-08) and most likely results from small numeric deviations, illustrating the excellent agreement between IsoCorrectoR and IsoCor v2. Figures A12 and A13 in the appendix show the same comparison when not correcting for tracer purity and when not correcting for tracer element NA in the core molecule (the part of the molecule not coming from derivatization), respectively. The maximum differences found are on the same scale as in the previous comparison.

The other two datasets are example data supplied with AccuCor (Su et al. (2017)). One dataset covers a $^{13}C$ stable isotope labeling experiment, the other measurements from a $^{15}N$ stable isotope labeling experiment. Both datasets were acquired with a high-resolution orbitrap device at a resolution of 140000 at 200 $m/z$. Thus, the data should not be corrected using a low-resolution correction approach, as this would result in overcorrection. However, if the goal is simply a comparsion of two tools, overcorrection of the data should not be a problem, but instead serve as a stress-test with regard to how the tools perform when corrected values $< 0$ would result in some cases. Both IsoCorrectoR and IsoCor v2 use an algorithm that does not allow negative values when solving the correction equation (Equation II.1). Figures II.27 and II.28 contain heatmaps showing the difference resulting from subtracting the IsoCor v2 MID from the IsoCorrectoR MID for the $^{13}C$ stable isotope labeling dataset. A tracer purity of 0.99 was assumed for correction, the two heatmaps show the results for different metabolites. A maximum difference in the range of 9E-08 again illustrates excellent agreement. To also cover $^{15}N$ labeling, the heatmaps in Figures II.29 and II.30 show the difference between IsoCor v2 and IsoCorrectoR low-resolution MS correction on the $^{15}N$ stable isotope labeling data. Here, maximum differences are in the range of 2.5E-07 and thus slightly more pronounced than in the other datasets. Yet, they are still extremely small in practice. Figures A14 and A15 in the appendix show the comparison of the two tools on data from simulated $^{18}O$ (glucose) and $^{2}H$ (palmitate) stable isotope labeling experiments. The data were simulated by randomly generating uncorrected MIDs (see 3.3 in the methods section). Again, the comparisons show excellent agreement between IsoCor v2 and IsoCorrectoR.

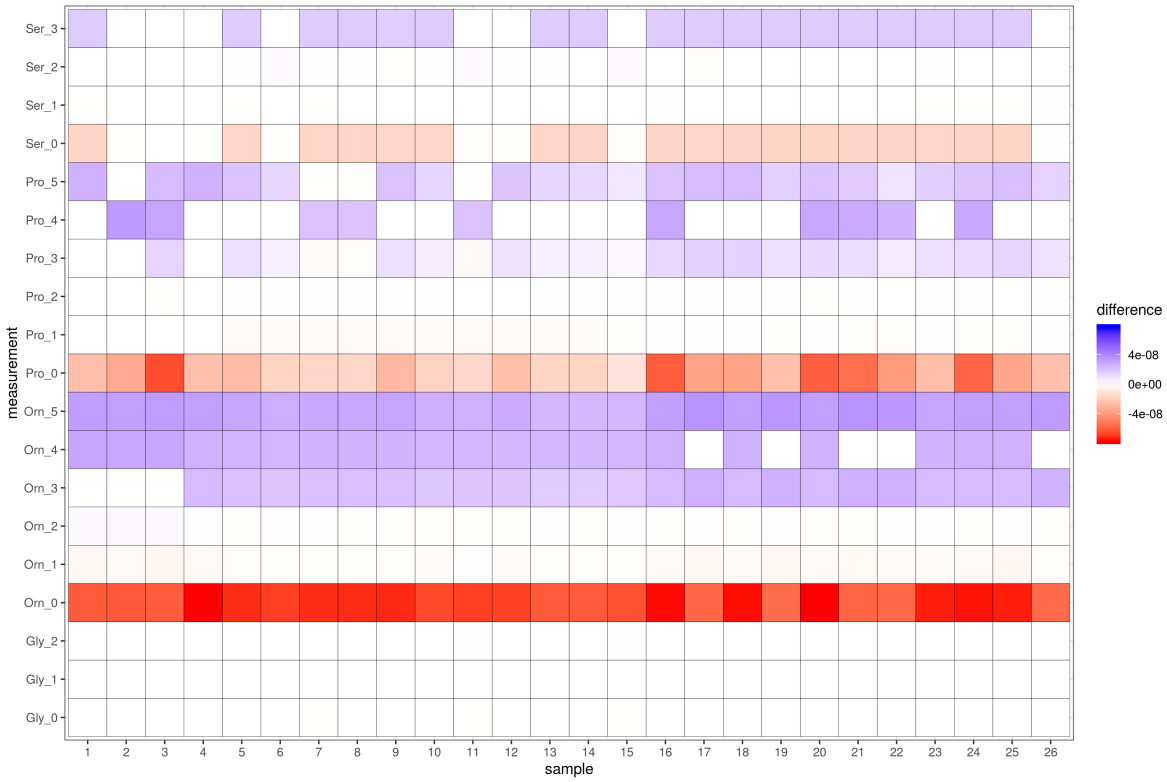



Figure II.26: ***Heatmap showing differences between IsoCorrectoR and IsoCor v2 for the low-resolution MS correction of data from a $^{13}C$ labeling experiment.*** *Amino acid data from a $^{13}C$ labeling experiment with the tracer substrate U-$^{13}C$-glutamine in a P493-6 B-cell line. A tracer purity of 0.99 was assumed. The heatmap was generated by subtracting MIDs corrected by IsoCor v2 from MIDs corrected by IsoCorrectoR. Columns represent samples, while rows represent measurements of mass isotopomers. The number after the metabolite name corresponds to the number of label incorporated. Abbreviated metabolite names: Gly - Glycine, Orn - Ornithine, Pro - Proline, Ser - Serine.*

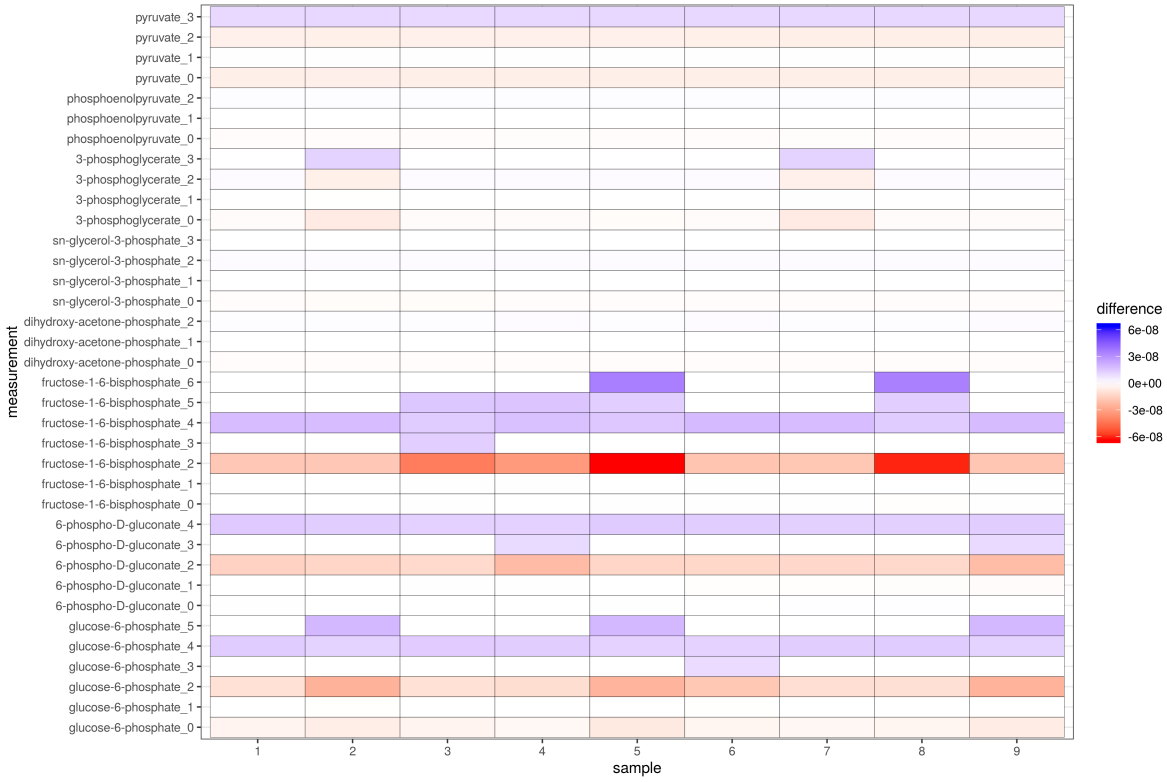



Figure II.27: ***Heatmap showing differences between IsoCorrectoR and IsoCor v2 for the low-resolution MS correction of data from a $^{13}C$ labeling experiment.*** *Data used are example data supplied with AccuCor (Su et al. (2017)). A tracer purity of 0.99 was assumed. The heatmap was generated by subtracting MIDs corrected by IsoCor v2 from MIDs corrected by IsoCorrectoR. Columns represent samples, while rows represent measurements of mass isotopomers. The number after the metabolite name corresponds to the number of label incorporated.*

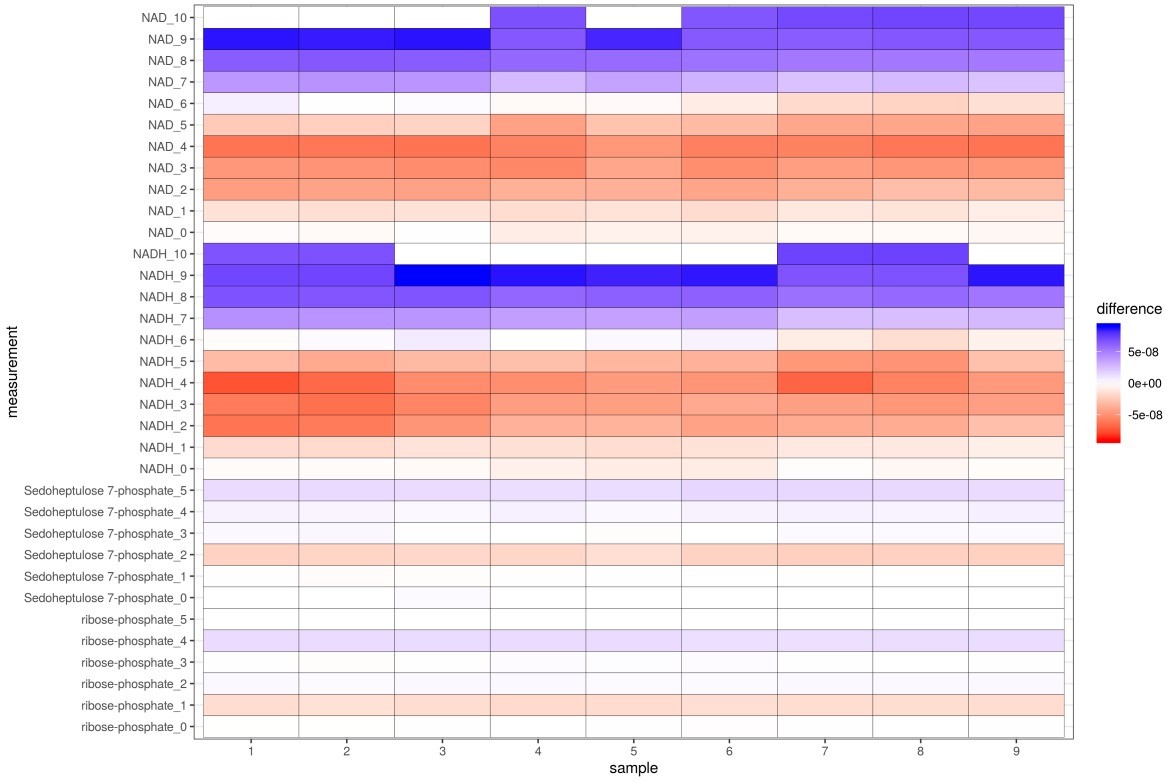



Figure II.28: ***Heatmap showing differences between IsoCorrectoR and IsoCor v2 for the low-resolution MS correction of data from a $^{13}C$ labeling experiment.*** *Data used are example data supplied with AccuCor (Su et al. (2017)). A tracer purity of 0.99 was assumed. The heatmap was generated by subtracting MIDs corrected by IsoCor v2 from MIDs corrected by IsoCorrectoR. Columns represent samples, while rows represent measurements of mass isotopomers. The number after the metabolite name corresponds to the number of label incorporated.*

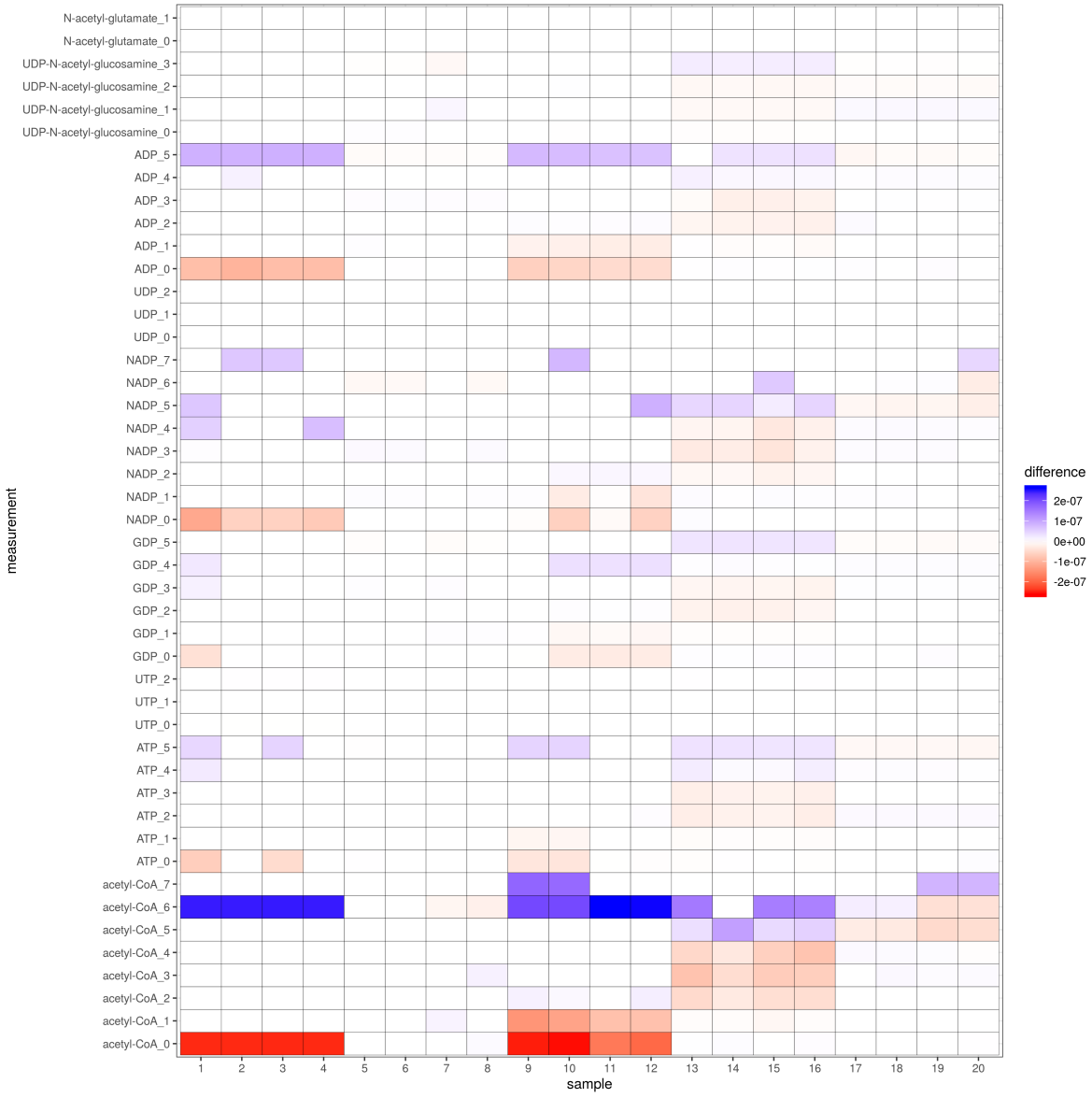



Figure II.29: ***Heatmap showing differences between IsoCorrectoR and IsoCor v2 for the low-resolution MS correction of data from a $^{15}$N labeling experiment.*** *Data used are example data supplied with AccuCor (Su et al. (2017)). A tracer purity of 0.99 was assumed. The heatmap was generated by subtracting MIDs corrected by IsoCor v2 from MIDs corrected by IsoCorrectoR. Columns represent samples, while rows represent measurements of mass isotopomers. The number after the metabolite name corresponds to the number of label incorporated.*

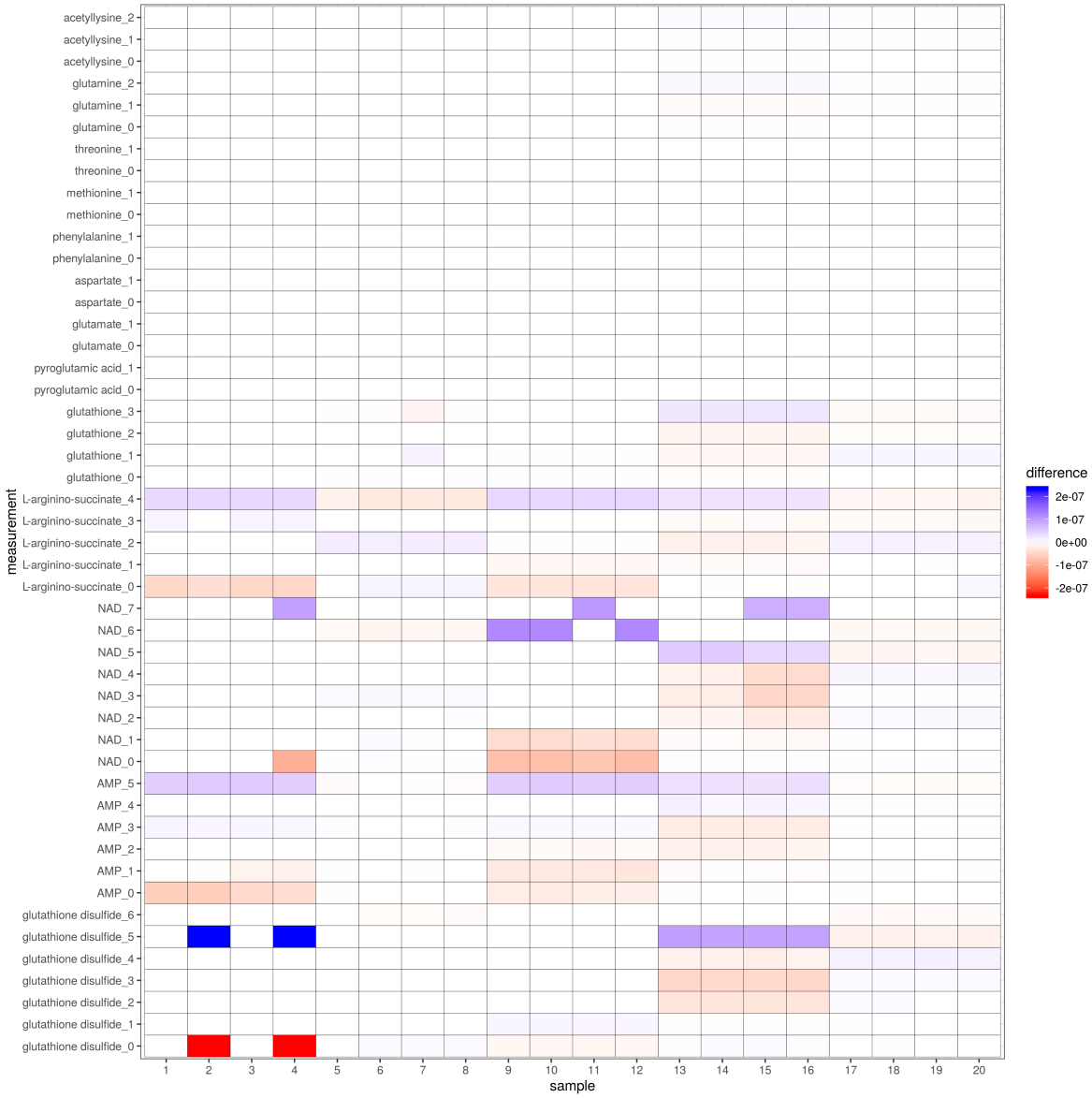



Figure II.30: ***Heatmap showing differences between IsoCorrectoR and IsoCor v2 for the low-resolution MS correction of data from a $^{15}N$ labeling experiment.*** *Data used are example data supplied with AccuCor (Su et al. (2017)). A tracer purity of 0.99 was assumed. The heatmap was generated by subtracting MIDs corrected by IsoCor v2 from MIDs corrected by IsoCorrectoR. Columns represent samples, while rows represent measurements of mass isotopomers. The number after the metabolite name corresponds to the number of label incorporated.*

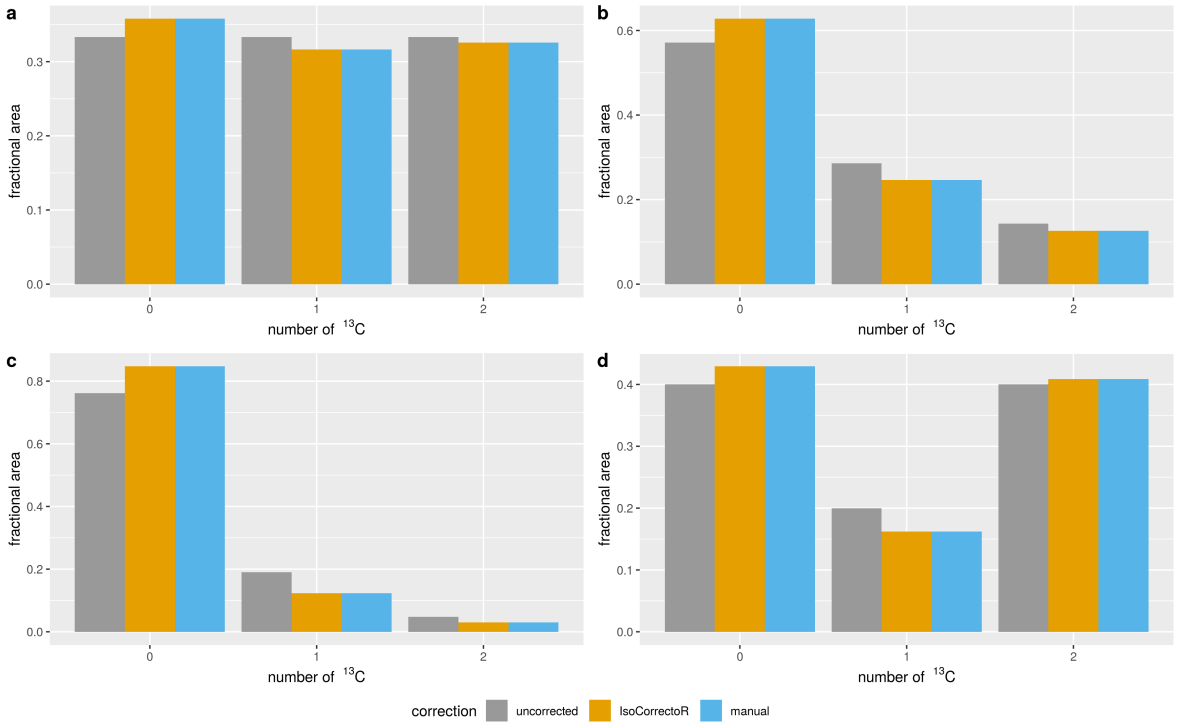



Figure II.31: ***Comparison of low-resolution MS correction to manual calculations.*** *Simulated $^{13}C$ stable isotope labeling data from PCF-derivatized glycine was corrected using the low-resolution MS correction approach of IsoCorrectoR and through manual calculations. A tracer purity of 99% was assumed. The x-axis labels 0 - 2 correspond to the glycine isotopologues with 0 - 2 $^{13}C$ incorporated. The panels show the results for different uncorrected MIDs.* ***(Figure and legend partly taken from Heinrich et al. (2018) supplementary material)***

**Comparison to manual calculations** To validate IsoCorrectoR's low-resolution MS correction approach, we also performed a manual correction of simulated $^{13}$C stable isotope labeling data from PCF-derivatized glycine. The probability matrix was generated using Microsoft Excel. It was then imported together with simulated, uncorrected values into Matlab to solve the system of linear equations for the corrected values. The process of manual probability matrix calculation is laborious and error-prone. However, for small core molecules like glycine it is still manageable and the derivatization, as long as it does not introduce new elements into the molecule, does not add substantially to the complexity of the calculation. Figure II.31 depicts the comparison for different simulated uncorrected MIDs of PCF-glycine (tracer purity assumed for correction was 99%). The results of IsoCorrectoR match the results from the manual calculations perfectly, with a maximum difference in the range of 5E-09. **(Paragraph taken from Heinrich et al. (2018) supplementary material and modified)**

**Experimental validation** To assess the validity of the correction procedure also experimentally, different $^{13}$C labeled alanine species were combined to varying mixtures of known composition. The samples were subjected to silylation to yield TMS-derivatized alanine and analyzed by GC-APCI-TOFMS. Each mixture was prepared and analyzed in triplicates. The diagrams shown in Figure II.32 contain the fractional measured uncorrected data for each

isotopologue mass window, the fractions of the labeled species expected after the correction (corresponding to their fractional concentrations in the mixture), the expected uncorrected data (derived by performing IsoCorrectoR's algorithm in reverse, thus creating uncorrected values from known true corrected values), and the correction results of IsoCorrectoR and IsoCor (v1) at 99% tracer purity. In line with previous examples, the correction results of the two tools match very well. With regard to the corrected values expected according to the composition of the mixture, Figure II.32a shows relatively pronounced deviations for the species with 2 and 3 $^{13}$C. After correction with IsoCorrectoR/IsoCor, the 2 $^{13}$C species shows about 90% of the expected value, while roughly 160% of the expected value is found for the 3 $^{13}$C species. When comparing the measured uncorrected data to the expected uncorrected data, it becomes evident that the deviations found have the same direction and overall magnitude as those found between the expected corrected values and the corrected values derived from IsoCorrectoR/IsoCor. This indicates that, given the results of IsoCorrectoR and IsoCor are correct, deviations in the uncorrected values translate more or less directly to the corrected values. The cause of such differences between expected uncorrected values and measured uncorrected values are most likely measurement or peak integration inaccuracies. Such a deviation may be of small or medium size in relation to the uncorrected value. Compared to a substantially smaller corrected value, however, it can be huge. This exemplifies the effect of natural abundance/tracer purity correction in conjunction with inaccurate measurements (which can be especially pronounced in the lower dynamic range). For isotopologues with a low fractional abundance, the relative magnitude of a bias existing in the uncorrected data is expected to increase substantially in the corrected data. On the other hand, the diagrams in Figure II.32 indicate that performing no correction will result in a much greater bias. Figure II.32b shows the results for a different mixture of alanine isotopologues, none of which has markedly low fractional abundance. As can be seen, the bias of the corrected data relative to the expected values is far less pronounced than in Figure II.32a. The diagrams for two additional mixtures can be found in Figure A16 in the appendix. The results are similar to those shown in Figure II.32. **(Paragraph taken from Heinrich et al. (2018) and slightly modified)**

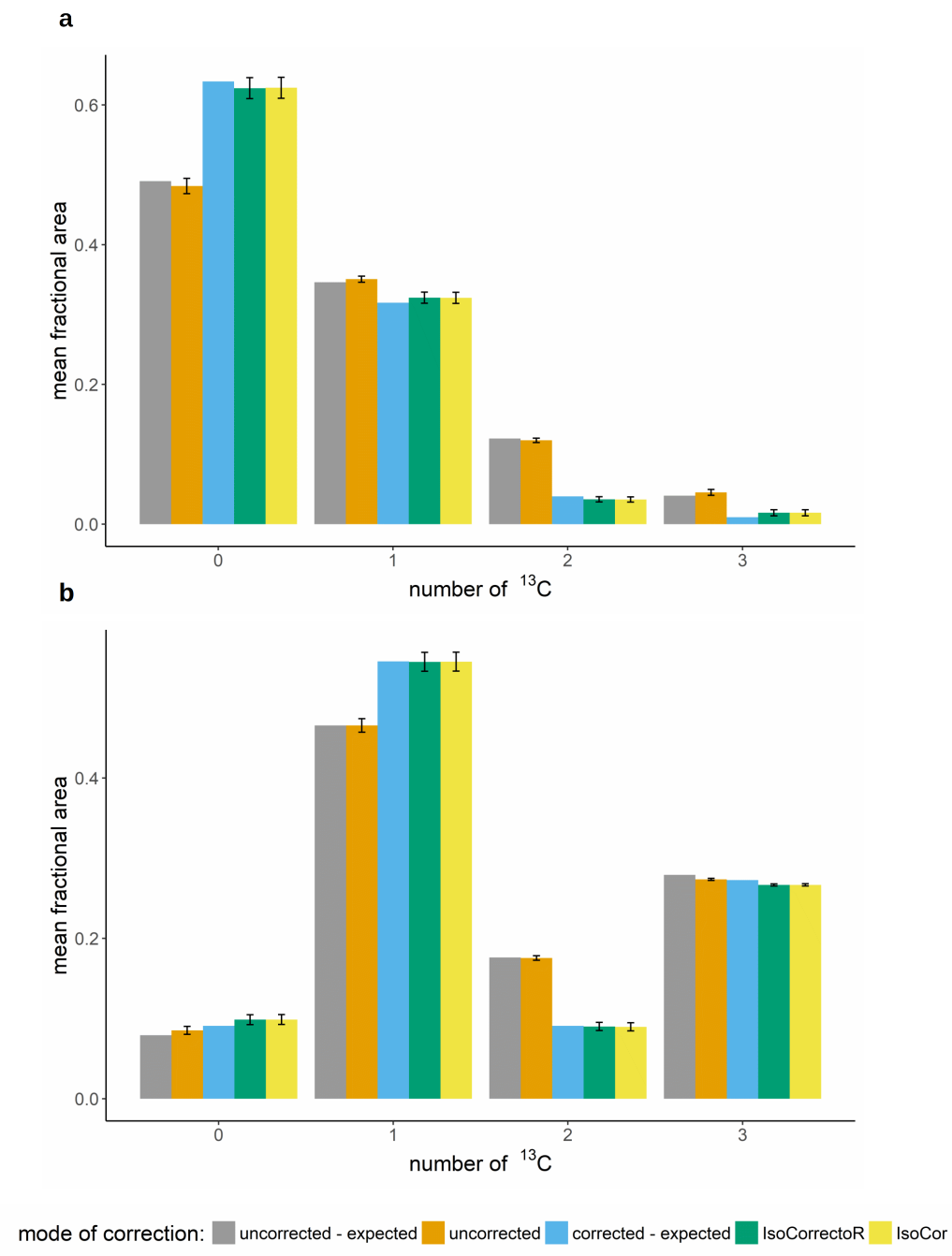



Figure II.32: ***Experimental validation of low-resolution MS correction with isotopologue mixtures of known composition.*** *Low-resolution MS correction of GC-MS data of mixtures of TMS-derivatized alanine isotopologues. The mixtures have known composition. The x-axis labels 0 - 3 correspond to the TMS-alanine isotopologues with 0 - 3 $^{13}C$ incorporated. The grouped bar charts depict the correction performance of IsoCorrectoR or IsoCor in comparison to the expected values. Additionally, the expected uncorrected values (reverse application of IsoCorrectoR functions) are compared to the uncorrected values actually measured. Samples were measured in technical triplicates, means of isotopologue fractions +/- SD are shown. Panels a and b show different mixtures.* ***(Figure and legend taken from Heinrich et al. (2018) and slightly modified)***

### 3.3.2 Low-resolution MS/MS correction

**Comparison to ICT** To compare IsoCorrectoR's low-resolution MS/MS correction approach to the Perl-based MS/MS correction tool ICT, organic acid and amino acid data from a $^{13}C$ stable isotope labeling experiment conducted in a P493-6 B-cell line were used (see section 3.1 in the methods chapter). Cells were fed with U-$^{13}C$-glutamine, and the incorporation of $^{13}C$ into different metabolites was assessed via LC-MS/MS with a low-resolution triple-quadrupole device. Data from that experiment were already employed in various previous sections. The heatmap in Figure II.33 shows the differences between IsoCorrectoR's and ICT's MS/MS correction for low-resolution MS/MS data of organic acids. The corrected MIDs obtained from ICT were subtracted from the corrected MIDs produced by IsoCorrectoR, and the (absolute) differences were $log_{10}$ transformed to better capture their scale. Gray fields indicate an untransformed difference of 0 between the two tools. Differences get as high as 8E-04, which is nearing the range of biological relevance. However, when considering the color pattern visible in the heatmap, it becomes evident that more pronounced differences are always found in samples that contain at least one value where the difference between tools is exactly 0. This usually occurs in samples where corrected values $< 0$ would result. As both ICT and IsoCorrectoR do not allow corrected values $< 0$, both tools force those values to 0, thus yielding an exact match. In the case of IsoCorrectoR, this yields correction residuals, as exact solving of $v_m = P \cdot v_c$ is not possible.

Figure II.34 shows these residuals for the given data, indicating a clear match between samples with correction residuals and samples where IsoCorrectoR and ICT differ more strongly. Most likely, differences in the algorithmic approach that enforces corrected values $\geq 0$ are the cause of the observed mismatch between the two tools. Both IsoCorrectoR and IsoCor employ a least-squares optimization procedure (minimizing the squared residuals) with the constraint of obtaining only positive values when solving $v_m = P \cdot v_c$. This is likely the reason why the results of IsoCorrectoR and IsoCor match so well also in the case of overcorrecting high-resolution data with a low-resolution approach (see section 3.3.1). ICT, on the other hand, uses a different method to ensure positive corrected values. Here, the system of linear equations in $v_m = P \cdot v_c$ is solved analytically, equation by equation, following gaussian elimination to bring the correction matrix into upper triangular shape. While the system of equations is solved row by row, each time a negative corrected value would occur it is directly set to 0, and this value of 0 is used for solving the remaining equations (Jungreuthmayer et al. (2016) and ICT documentation). Clearly, such differences in solving $v_m = P \cdot v_c$ between two tools should lead to some degree of deviation between the corrected MIDs. Differences between the correction results of ICT and IsoCor on MS (not MS/MS) data have been reported, too, in the ICT documentation for cases where correction would yield negative values. In samples where it is not necessary to deal with negative corrected values, the results of ICT and IsoCorrectoR match perfectly, with a maximum difference in the range of 2E-08. Figure II.35 shows a heatmap comparing ICT and IsoCorrectoR on amino acid data from the same labeling experiment. Again, higher deviations are solely found in samples where correction would result in negative values (see Figure A17 in the appendix for a heatmap of the residuals). Additionally, the low-resolution MS/MS correction of ICT and IsoCorrectoR has been compared on acetyl-CoA MS/MS data from a simulated $^{15}N$ labeling experiment (random generation of uncorrected MIDs). A heatmap of the comparison is provided in Figure A18 in the appendix. It shows results similar to the $^{13}C$ labeling data. Again, major differences between the two tools are only found in samples where corrected values $< 0$ would result.

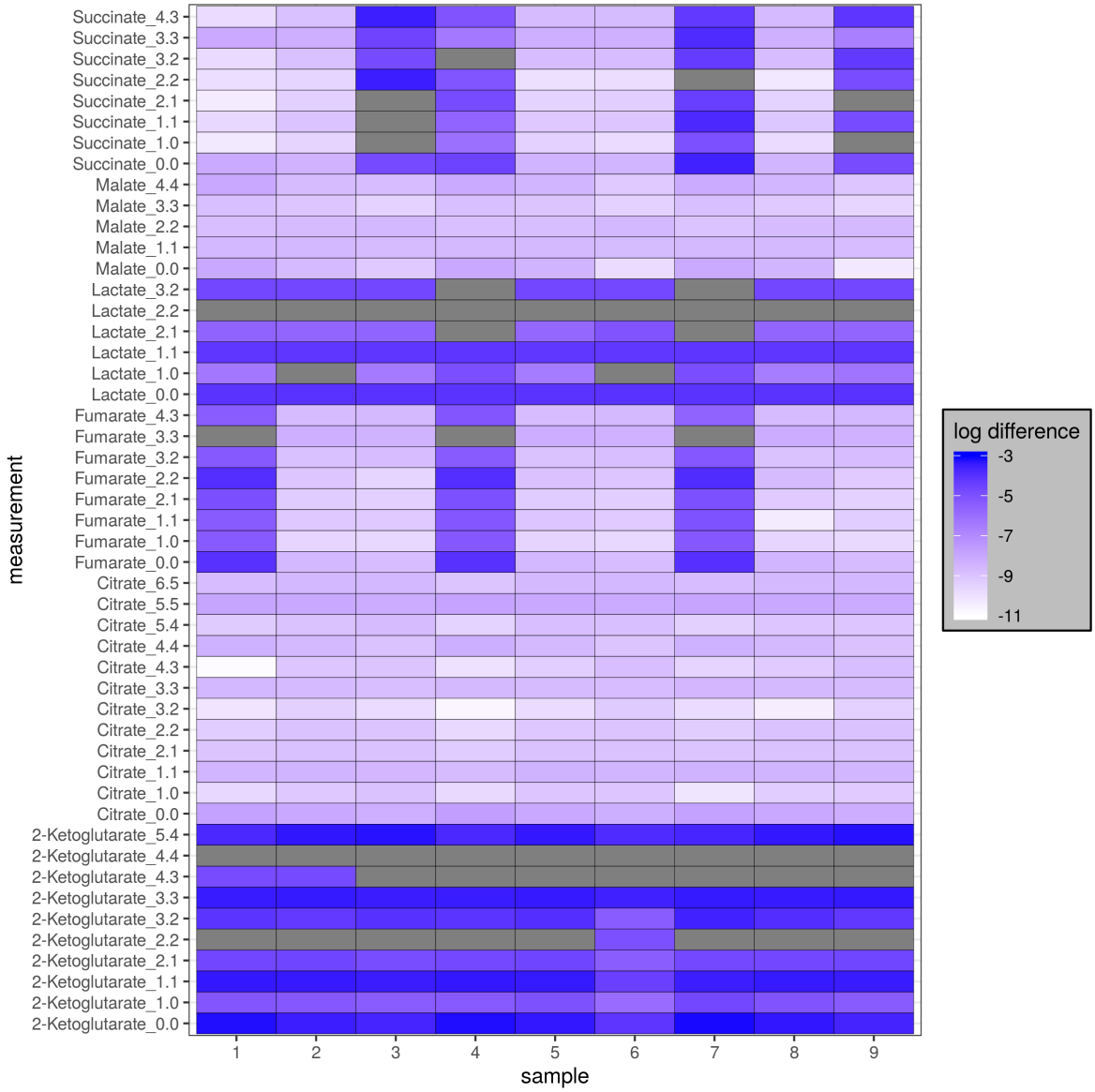



Figure II.33: ***Heatmap showing differences between IsoCorrectoR and ICT for the low-resolution MS/MS correction of data from a $^{13}C$ labeling experiment.*** *Organic acid data from a $^{13}C$ labeling experiment with the tracer substrate U-$^{13}C$-glutamine in a P493-6 B-cell line. A tracer purity of 0.99 was assumed. The heatmap was generated by subtracting MIDs corrected by ICT from MIDs corrected by IsoCorrectoR. The (absolute) differences were $log_{10}$ transformed, gray fields indicate an untransformed difference of 0 (equality of values). Columns represent samples, while rows represent measurements of MS/MS transitions. The numbers n.m after the metabolite names correspond to the MS/MS transitions with n label in the precursor ion and m label in the product ion.*

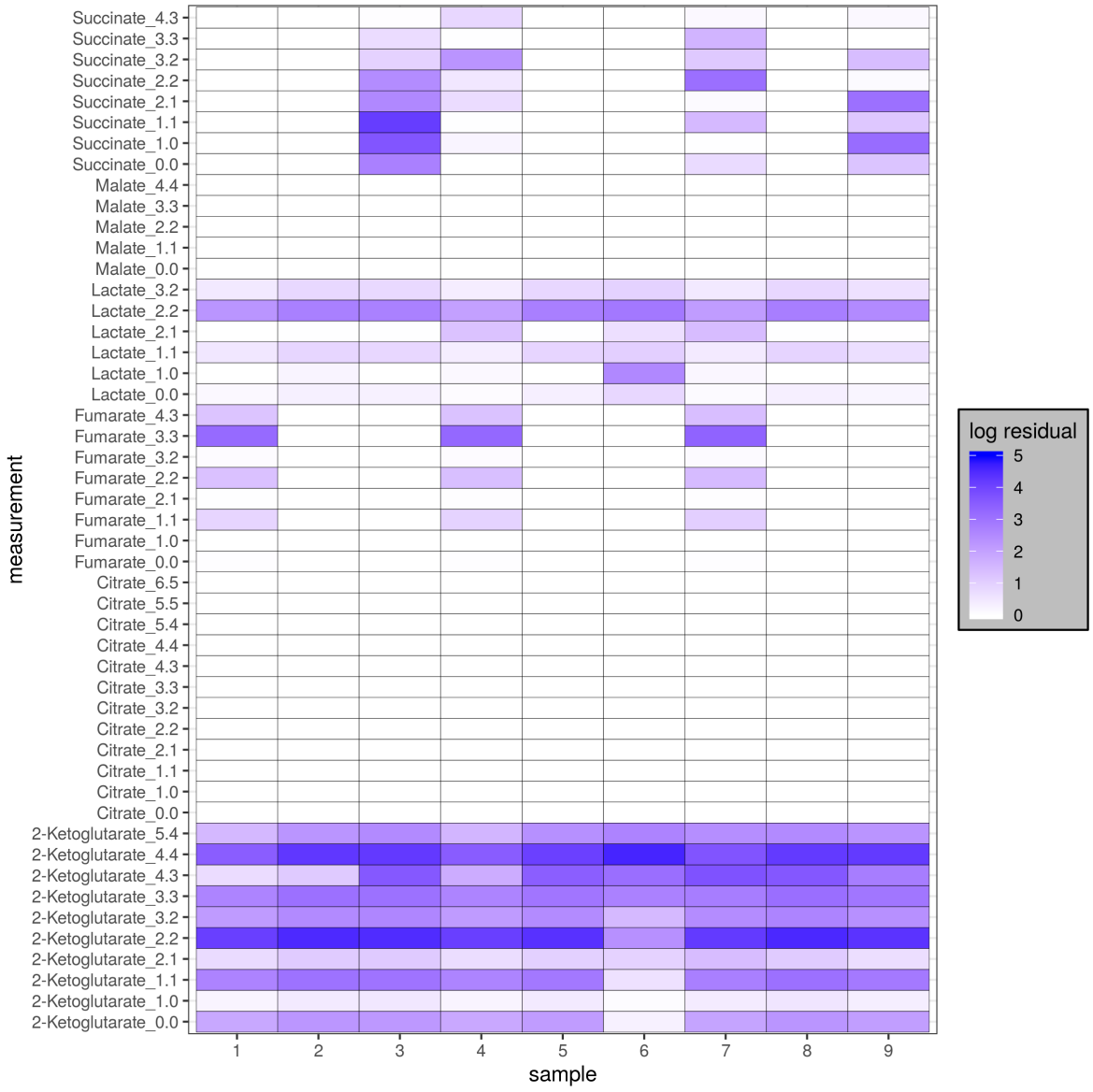



Figure II.34: ***Heatmap showing the correction residuals for the low-resolution MS/MS correction of data from a $^{13}C$ labeling experiment with IsoCorrectoR.*** *Organic acid data from a $^{13}C$ labeling experiment with the tracer substrate U-$^{13}C$-glutamine in a P493-6 B-cell line. A tracer purity of 0.99 was assumed. The (absolute) values were added a constant of 1 and $\log_{10}$ transformed. Columns represent samples, while rows represent measurements of MS/MS transitions. The numbers n.m after the metabolite names correspond to the MS/MS transitions with n label in the precursor ion and m label in the product ion.*

Which method of dealing with negative corrected values is actually superior is debatable. Using ICT's approach, a negative corrected value of a given species is set to 0 and all other corrected values remain unaffected. Thus, in such a model, the sole cause of a negative corrected value that is considered is a (measurement/integration) error in the corresponding uncorrected value of the same species. However, a negative corrected value in a given species $x$ may also be caused by measurement/integration errors in the uncorrected values of other species that contribute to $x$ through NA or tracer impurity. This possibility is covered in the least-squares optimization approach employed in IsoCor and IsoCorrectoR, where also other corrected values are adapted during residual minimization if they contribute to the negative corrected value. Thus, in cases where the only reason for a negative corrected value is an error in the uncorrected value of the same species, ICT's approach should be superior. If there are more contributions, the least-squares optimization approach may cover the situation more appropriately.

During the comparison of IsoCorrectoR and ICT, it became evident that correction with ICT seems to take substantially more time (correction was performed on the system described in 4.3 in the methods chapter). While the correction of the organic acid and amino acid data discussed in this section took only a few seconds when using IsoCorrectoR, it took several minutes when ICT was used for correction. While a run time in the range of minutes is still completely acceptable for an isotope correction task in practice, it appears that the run time of ICT increases drastically with more complex molecules. The correction of the acetyl-CoA MS/MS data used in this section took 3 hours at 100% CPU usage when correcting with ICT, while IsoCorrectoR finished the task, again, in several seconds. While the correction results of IsoCorrectoR and ICT match very well except for the different approach to dealing with negative corrected values, the very high run time when dealing with more complex molecules is a clear disadvantage of ICT.

The reasons for these huge run time differences may be found in the algorithmic approach or in the implementation. As described in section 3.1.3, IsoCorrectoR computes element combinations for all product ion and neutral loss labeling states and combines species with the same nominal mass shift to obtain mass distributions for the labeling states. These product ion and neutral loss mass distributions are then multiplied to yield the entries $p_{ij}$ of the probability/correction matrix $P$. ICT, on the other hand, computes element combinations for all precursor labeling states. For each precursor element combination, it then individually computes the probability that it will produce a certain product ion element combination by random drawing from the precursor composition. For a given labeling state (defined by the number of label in precursor and product ion), element combinations that share the same mass shift in precursor ion and product ion are then summed up to produce an entry $p_{ij}$ of $P$. The process of drawing multiple product ion element combinations from each precursor element combination is not needed in IsoCorrectoR. However, this approach does not appear to be substantially less efficient than that of IsoCorrectoR, as the process of drawing itself should not be very computationally expensive. Thus, it is more likely that the reason for the run time differences lies in the implementation. Probably, ICT makes use of data structures that store references to (numeric) values (equivalent to R lists) rather than the values themselves (equivalent to R vectors and matrices) during performance critical computation. As a result, data required for a given calculation (*e.g.*, combinatorical multiplication of elements in two vectors) would be scattered in computer memory rather than located in a single block or a few blocks of memory, requiring frequent reloading of CPU cache, which can tremendously slow computation. Especially during the computation of element combination probabilities and

mass shifts, where many values are combinatorically multiplied and added, scattering these values across memory by using data structures that store references instead of actual values can be very costly.

**Comparison to manual calculations** Similarly to low-resolution MS correction, also low-resolution MS/MS correction was validated via manual calculations. The manual calculations were performed just as described previously in section 3.3.1 and in the methods chapter. Simulated $^{13}C$ stable isotope labeling MS/MS data from PCF-derivatized alanine (uncorrected MIDs were devised manually) was corrected using the low-resolution MS/MS correction approach of IsoCorrectoR and through manual calculations (assuming a tracer purity of 99 %). The precursor $C_{10}H_{20}NO_4^+$ fragments at its C1-C2 bond and yields $C_6H_{12}NO_2^+$ as the product ion and $C_4H_8O_2$ as the neutral loss. The results of the comparison are shown for in Figure II.36 for different uncorrected MIDs. The match between manual calculations and IsoCorrectoR is excellent, showing a maximum difference in the range of 5E-09.

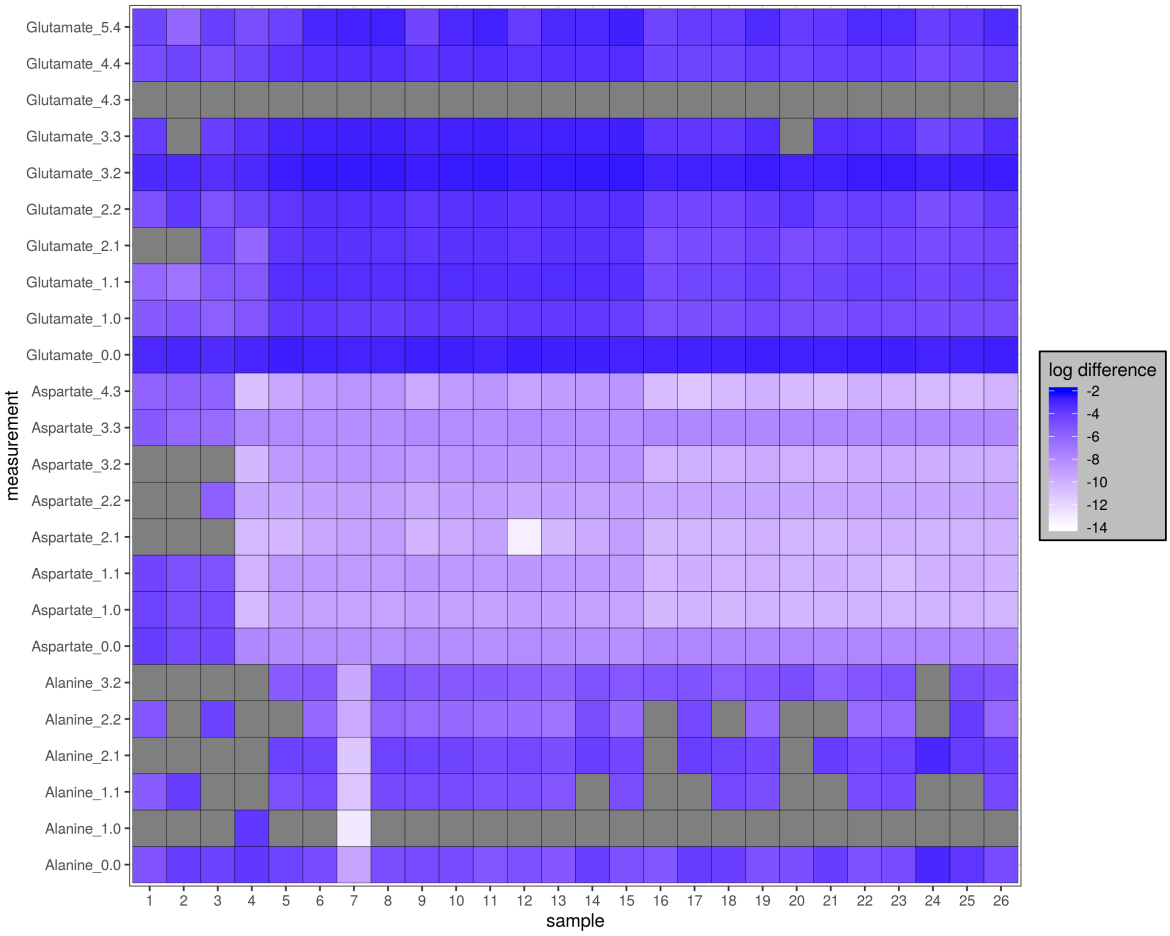



Figure II.35: ***Heatmap showing differences between IsoCorrectoR and ICT for the low-resolution MS/MS correction of data from a $^{13}C$ labeling experiment.*** *Amino acid data from a $^{13}C$ labeling experiment with the tracer substrate U-$^{13}C$-glutamine in a P493-6 B-cell line. A tracer purity of 0.99 was assumed. The heatmap was generated by subtracting MIDs corrected by ICT from MIDs corrected by IsoCorrectoR. The (absolute) differences were $log_{10}$ transformed, gray fields indicate an untransformed difference of 0 (equality of values). Columns represent samples, while rows represent measurements of MS/MS transitions. The numbers n.m after the metabolite names correspond to the MS/MS transitions with n label in the precursor ion and m label in the product ion.*

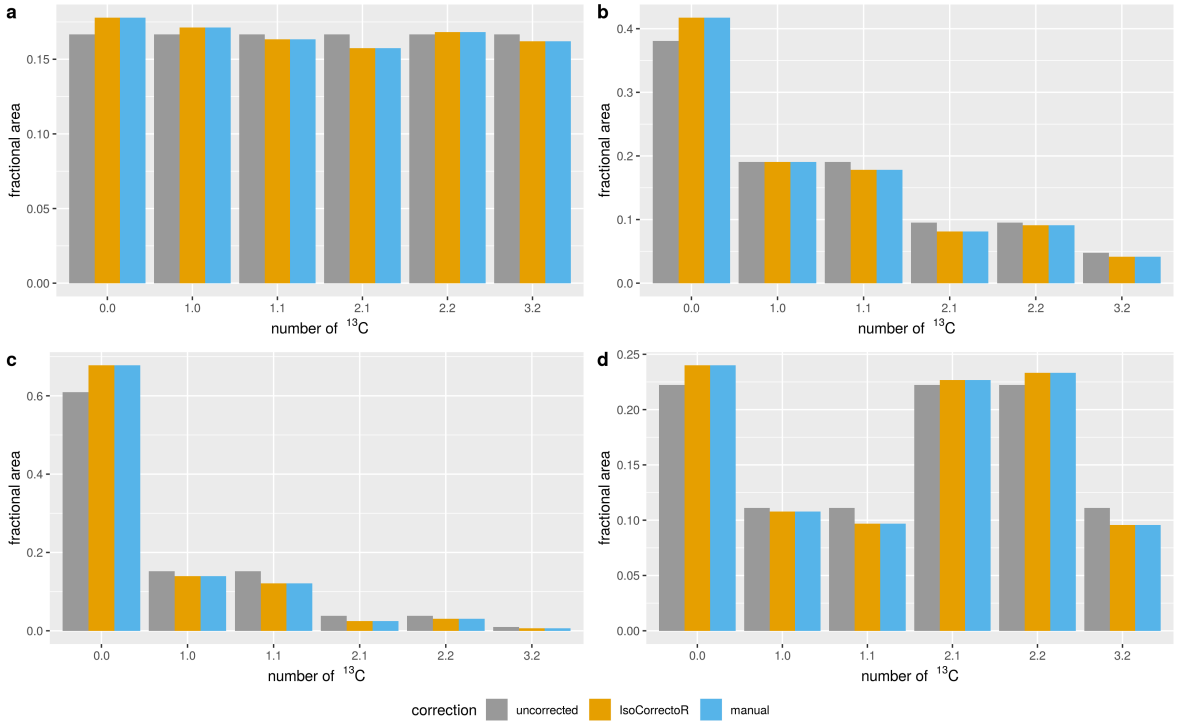



Figure II.36: ***Comparison of low-resolution MS/MS correction to manual calculations.*** *Simulated $^{13}C$ stable isotope labeling MS/MS data from PCF-derivatized alanine was corrected using the low-resolution MS/MS correction approach of IsoCorrectoR and through manual calculations. The precursor $C_{10}H_{20}NO_4^+$ fragments at its C1-C2 bond and yields $C_6H_{12}NO_2^+$ as the product ion and $C_4H_8O_2$ as the neutral loss. A tracer purity of 99% was assumed. The x-axis labels n.m correspond to the MS/MS transitions with n label in the precursor ion and m label in the product ion. The panels show the results for different uncorrected MIDs.*

### 3.3.3 Resolution-dependent correction

**Comparison to IsoCor** To validate its approach to resolution-dependent MS correction, IsoCorrectoR was compared to IsoCor v2. Currently, IsoCor v2 is the only tool that can be considered to perform resolution-dependent correction correctly, as AccuCor and ElemCor make false algorithmic assumptions (see 3.1.4 in the algorithm section). These do usually not lead to drastic differences, but are still incorrect (Millard et al. (2019)). Comparison was performed on a $^{13}C$ labeling dataset and a $^{15}N$ labeling dataset supplied as example data with AccuCor (Su et al. (2017)). Both datasets were acquired with an orbitrap device at a resolution of 140000 at 200 *m/z*. The heatmap in Figure II.37 shows the results for the $^{13}C$ labeling dataset. The heatmap was generated by subtracting MIDs corrected by IsoCor v2 from MIDs corrected by IsoCorrectoR. Maximum differences are in the range of 8E-08, suggesting an excellent match. Figure A19 in the appendix shows the same comparison for other molecules from the same dataset. A similar heatmap was generated for the $^{15}N$ labeling dataset and is shown in Figure II.38. For better illustrating the scale of differences, the values have been $log_{10}$ transformed. Figure A20 in the appendix shows the same analysis for other molecules from the dataset. As can be seen clearly, the differences are substantially more pronounced than for the $^{13}C$ labeling dataset, with a maximum in the range of 5E-05. Deviations in that range are only found for the comparably large metabolites acetyl-CoA and $NAD^+$.

While such differences should still be acceptable in practice, they nevertheless raise the question of their origin. Because of the still subtle magnitude of the differences, their cause is unlikely a major algorithmic difference. When performing low-resolution MS correction on the same dataset (see Figure II.29), the differences between IsoCor v2 and IsoCorrectoR are far less pronounced. This indicates that the deviations could be caused by a component specific to resolution-dependent MS correction. If there was a major issue with calculating *e.g.* $delta_{min}$ or the exact mass shifts of isotopologues, the differences should be more pronounced. Relatively small differences, such as the ones observed, may *e.g.* arise if the *m/z* value of the $m+0$ isotopologue is generally used for calculating the minimal resolvable mass difference $delta_{min}$ actually associated with the measurement of a $m+i$ isotopologue (see section 3.1.4). In fact, in the presentation of the IsoCor v2 algorithm in Millard et al. (2019), the $delta_{min}$ (or $m_{min}$) value does not seem to depend on $i$ but appears to be constant for each metabolite. Figures II.39 and A21 (appendix) show the same analysis as before, however, the IsoCorrectoR algorithm employed has been slightly altered to always use the *m/z* value of the $m+0$ isotopologue (instead of the *m/z* of the respective $m+i$ species) for computing $FWHM/delta_{min}$ of the labeled species. As a result, the maximum difference diminishes from 5E-05 to 6E-08, yielding an excellent match, as with the $^{13}C$ labeling dataset.

Hence, these results indicate that IsoCor v2 uses a constant $delta_{min}$ value for all labeling states of a given metabolite, not accounting for the *m/z* differences introduced by the tracer isotope. This has additionally been confirmed in personal communication with Pierre Millard, the author of IsoCor. While the approach of IsoCor v2 is formally incorrect, the resulting differences are minor. Further, it seems unlikely that it should be possible to determine the "true" $FWHM/delta_{min}$ value of the instrument used at a given *m/z* with very high accuracy by simply using Equation II.10. Most probably, the error in $FWHM$ resulting from using this formal calculation markedly exceeds the error produced by not accounting for the increasing *m/z* of $m+i$ species when computing $FWHM/delta_{min}$.

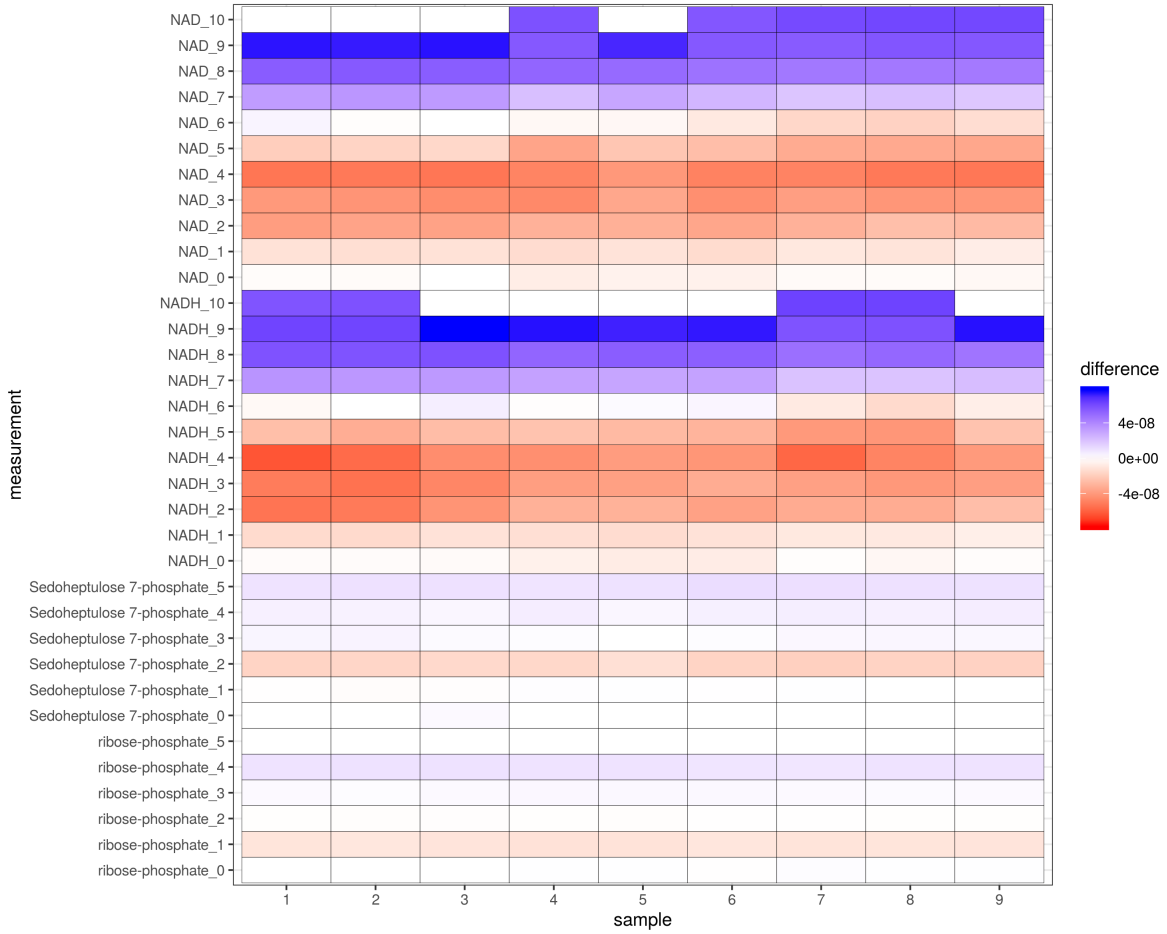



Figure II.37: ***Heatmap showing differences between IsoCorrectoR and IsoCor v2 for the resolution-dependent MS correction of data from a $^{13}C$ labeling experiment.*** *Data used are example data supplied with AccuCor (Su et al. (2017)). A tracer purity of 0.99 and an orbitrap mass analyzer with a resolution of 140000 at 200 m/z were assumed. The heatmap was generated by subtracting MIDs corrected by IsoCor v2 from MIDs corrected by IsoCorrectoR. Columns represent samples, while rows represent measurements of mass isotopomers. The number after the metabolite name corresponds to the number of label incorporated.*

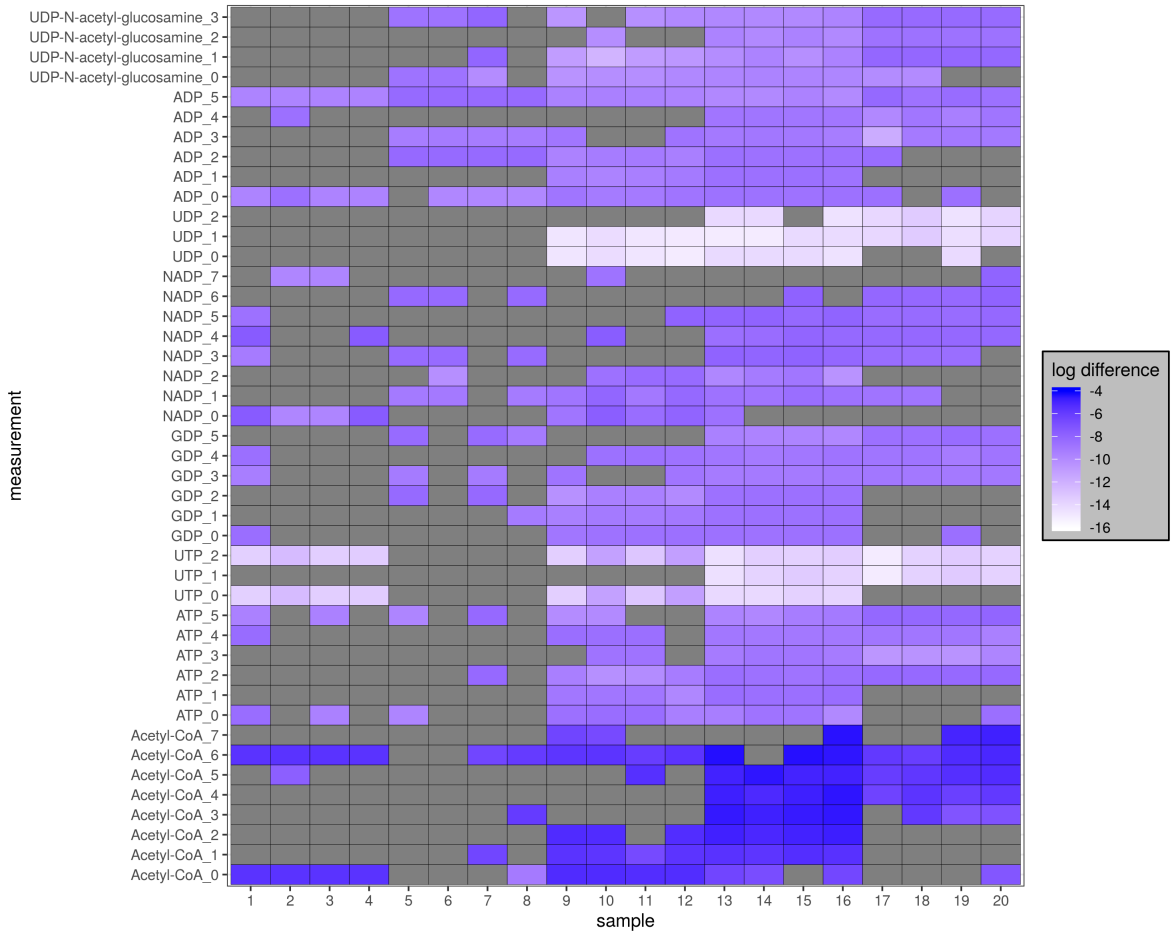



Figure II.38: ***Heatmap showing differences between IsoCorrectoR and IsoCor v2 for the resolution-dependent MS correction of data from a $^{15}N$ labeling experiment.*** *Data used are example data supplied with AccuCor (Su et al. (2017)). A tracer purity of 0.99 and an orbitrap mass analyzer with a resolution of 140000 at 200 m/z were assumed. The heatmap was generated by subtracting MIDs corrected by IsoCor v2 from MIDs corrected by IsoCorrectoR. The (absolute) differences were $log_{10}$ transformed, gray fields indicate an untransformed difference of 0 (equality of values). Columns represent samples, while rows represent measurements of mass isotopomers. The number after the metabolite name corresponds to the number of label incorporated.*

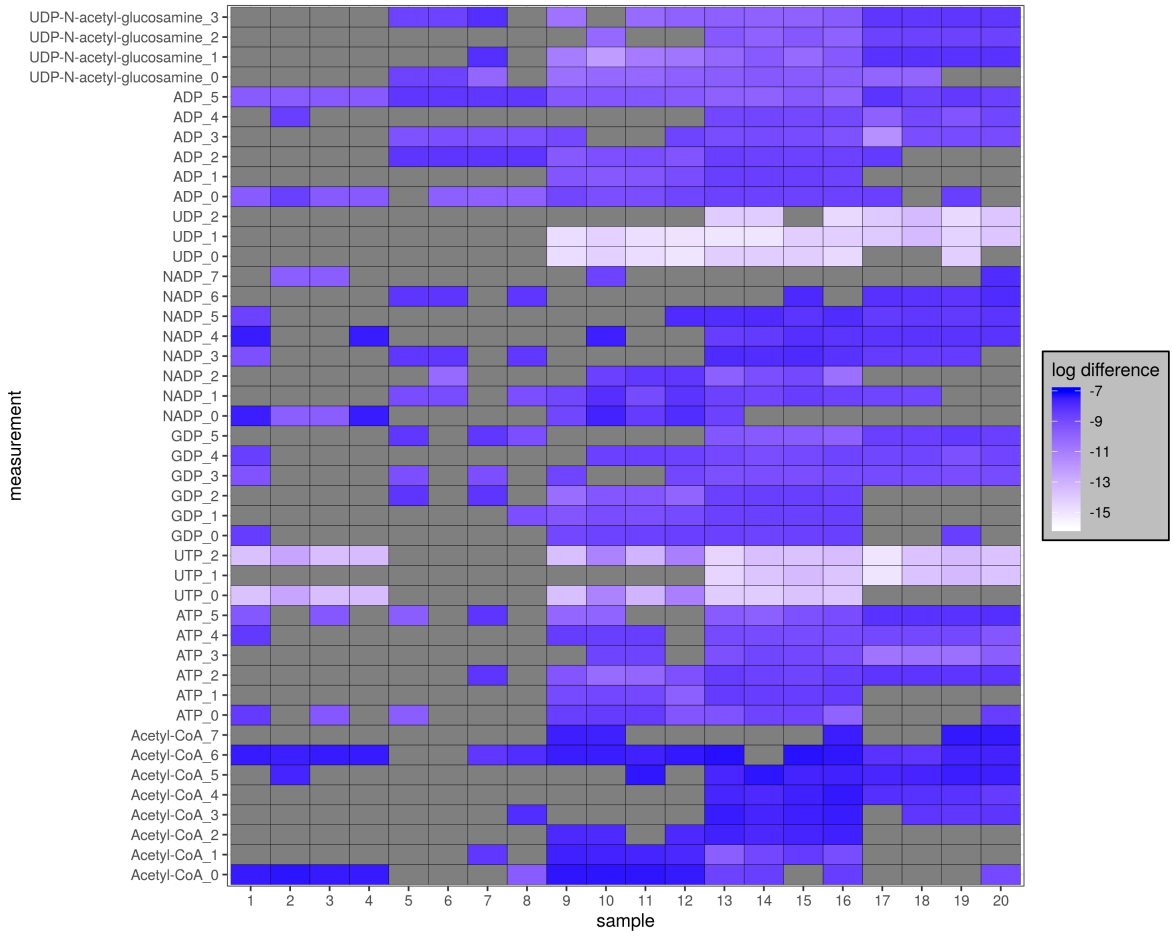



Figure II.39: ***Heatmap showing differences between IsoCorrectoR and IsoCor v2 for the resolution-dependent MS correction of data from a $^{15}N$ labeling experiment, altered IsoCorrectoR algorithm.*** *Data used are example data supplied with AccuCor (Su et al. (2017)). A tracer purity of 0.99 and an orbitrap mass analyzer with a resolution of 140000 at 200 m/z were assumed. IsoCorrectoR's algorithm was adjusted to always use the m/z of the m+0 mass isotopomer of a given metabolite (instead of the respective m+i mass isotopomer) for computing FWHM. The heatmap was generated by subtracting MIDs corrected by IsoCor v2 from MIDs corrected by IsoCorrectoR. The (absolute) differences were $log_{10}$ transformed, gray fields indicate an untransformed difference of 0 (equality of values). Columns represent samples, while rows represent measurements of mass isotopomers. The number after the metabolite name corresponds to the number of label incorporated.*

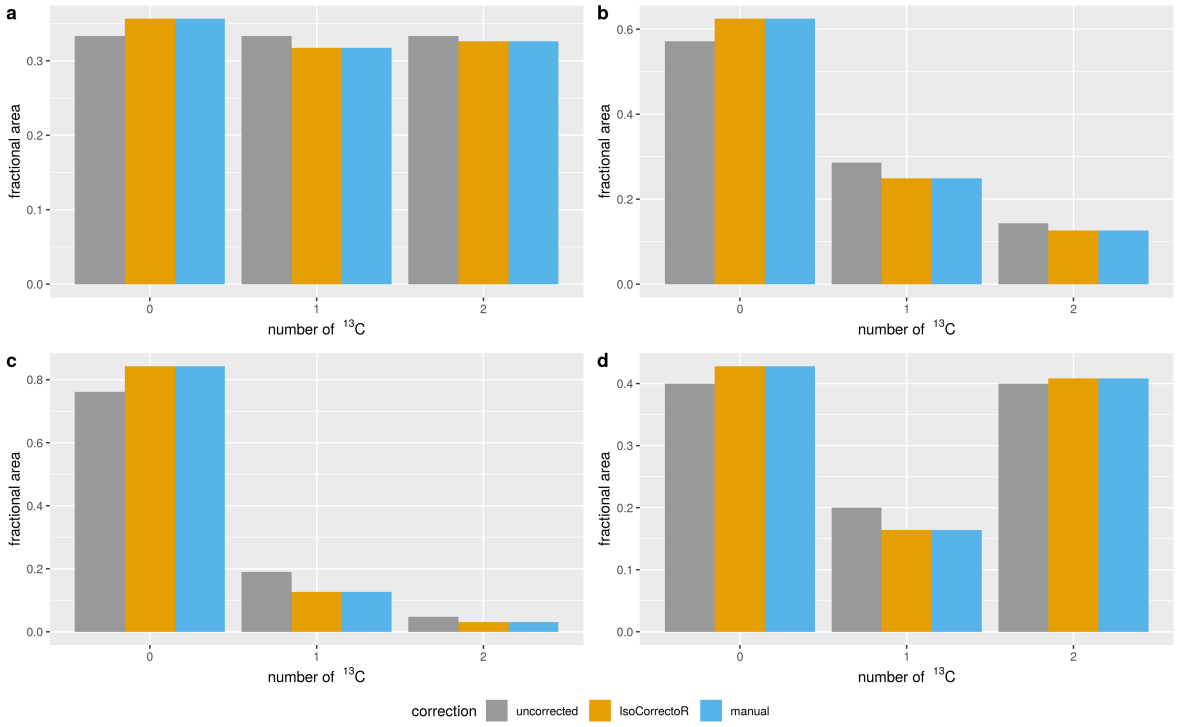



Figure II.40: ***Comparison of resolution-dependent MS correction to manual calculations.*** *Simulated $^{13}C$ stable isotope labeling data from PCF-derivatized glycine was corrected using the resolution-dependent MS correction approach of IsoCorrectoR and through manual calculations. An orbitrap mass analyzer with a resolution of 140000 at 200 m/z and a tracer purity of 99% were assumed for correction. The x-axis labels 0 - 2 correspond to the glycine isotopologues with 0 - 2 $^{13}C$ incorporated. The panels show the results for different uncorrected MIDs.*

**Comparison to manual calculations** Similarly to validating IsoCorrectoR's low-resolution MS correction approach, a manual correction of simulated $^{13}$C stable isotope labeling data from PCF-derivatized glycine (uncorrected MIDs were derived manually) was performed to validate resolution-dependent MS correction. The probability matrix was generated using LibreOfficeCalc. For resolution-dependent correction, the process is yet more complicated than for low-resolution correction. The probability matrix and uncorrected values were then supplied to the *lm* base function of R to solve the system of linear equations and obtain corrected values. Figure II.40 depicts the comparison for different simulated uncorrected MIDs of PCF-glycine. An orbitrap mass analyzer with a resolution of 140000 at 200 $m/z$ and a tracer purity of 99% were assumed for correction. The results of IsoCorrectoR match the results from the manual calculations perfectly (no difference). For this analysis, the calculation threshold parameter of IsoCorrectoR was set to 0 instead of using the 1E-08 default (see 4.1.1 in the methods chapter) to show that a difference of 0 can actually be obtained. However, when using the default threshold parameter, the difference would still be negligibly small (see 3.3.6).

### 3.3.4 Resolution-dependent MS/MS correction

For validating IsoCorrectoR's resolution-dependent MS/MS correction approach, a manual correction of simulated $^{13}C$ stable isotope labeling MS/MS data from PCF-derivatized alanine was performed (uncorrected MIDs were derived manually). The precursor $C_{10}H_{20}NO_4^+$ fragments at its C1-C2 bond and yields $C_6H_{12}NO_2^+$ as the product ion and $C_4H_8O_2$ as the neutral loss. The probability matrix was generated using LibreOfficeCalc. MS/MS correction adds an additional layer of complexity to the manual calculations compared to resolution-dependent MS correction. The probability matrix and uncorrected values were supplied to the *lm* base function of R to solve the system of linear equations and obtain corrected values. Figures II.41 and II.42 depict the comparison between IsoCorrectoR's results and manual correction for different simulated uncorrected MIDs. The mass spectrometer setup assumed for correction was a low-resolution precursor mass analyzer (resolution of 1000 at 200 *m/z*) combined with an orbitrap mass analyzer with a resolution of 140000 at 200 *m/z* for determining product ion *m/z*. Tracer purity was assumed to be 99%. As can be seen, there is an excellent match between the results of IsoCorrectoR and the manual calculations, with a maximum difference in the range of 1E-10. As in the comparison of resolution-dependent MS correction to manual calculations, the calculation threshold parameter of IsoCorrectoR was set to 0 instead of using the 1E-08 default. Since IsoCorrectoR is currently the only tool capable of correcting MS/MS data in a resolution-dependent manner, a comparison of resolution-dependent MS/MS correction with other tools is not possible. However, IsoCorrectoR's approach to resolution-dependent MS correction, which is the algorithmic basis for the MS/MS approach, has been thoroughly validated against IsoCor v2 (3.3.3), showing very good agreement.

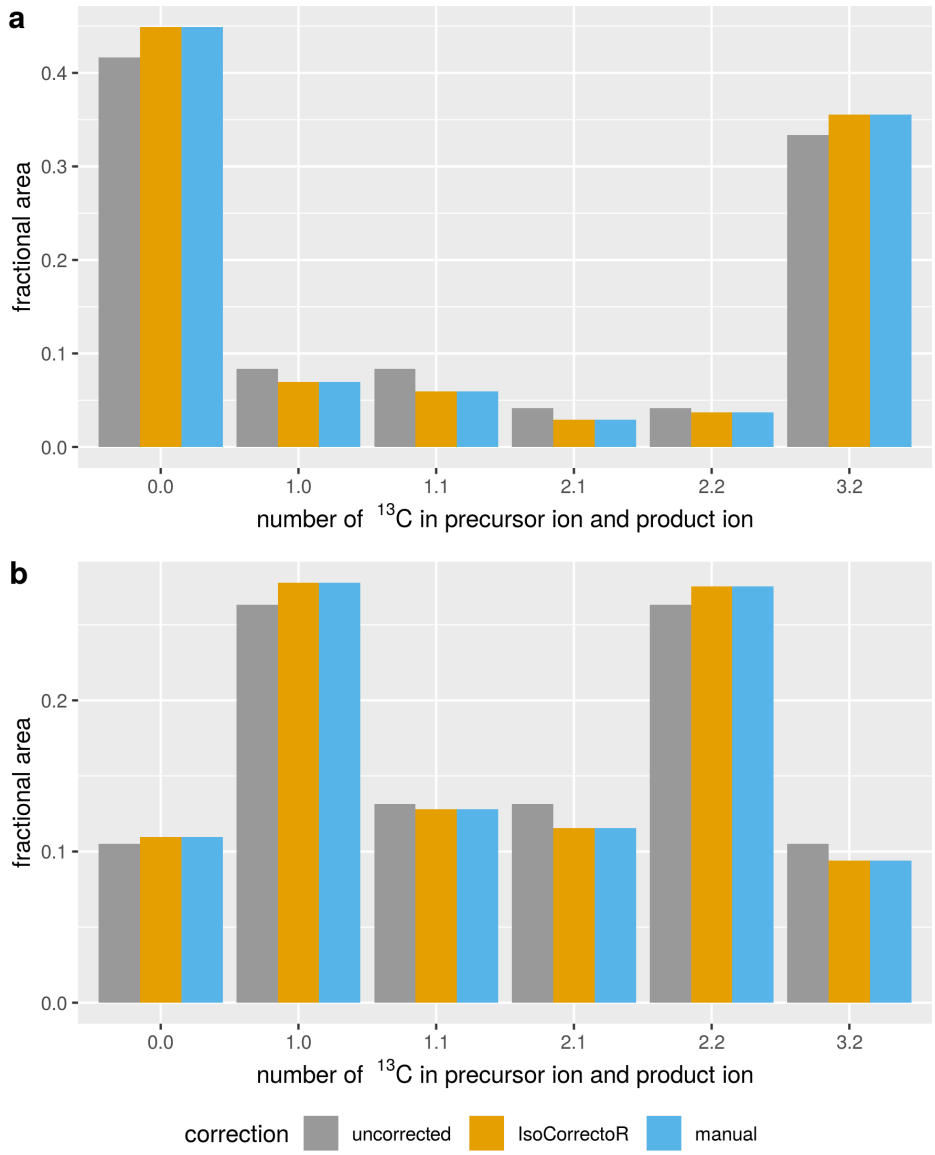



Figure II.41: ***Comparison of resolution-dependent MS/MS correction to manual calculations on example data.*** *The figure shows the results obtained from correcting simulated MS/MS data of PCF-derivatized alanine. The x-axis labels n.m correspond to the MS/MS transitions with n $^{13}C$ in the precursor ion and m $^{13}C$ in the product ion. Data are shown uncorrected, corrected using IsoCorrectoR's resolution-dependent MS/MS correction mode and corrected via manual calculations. Resolution-dependent correction was performed for the usual case of having a low-resolution precursor mass analyzer (1000 at m/z 200, FWHM constant over the m/z range) and a high resolution product ion mass analyzer (140000 at m/z 200, orbitrap m/z dependency of FWHM). Panels a and b correspond to different uncorrected MIDs.*

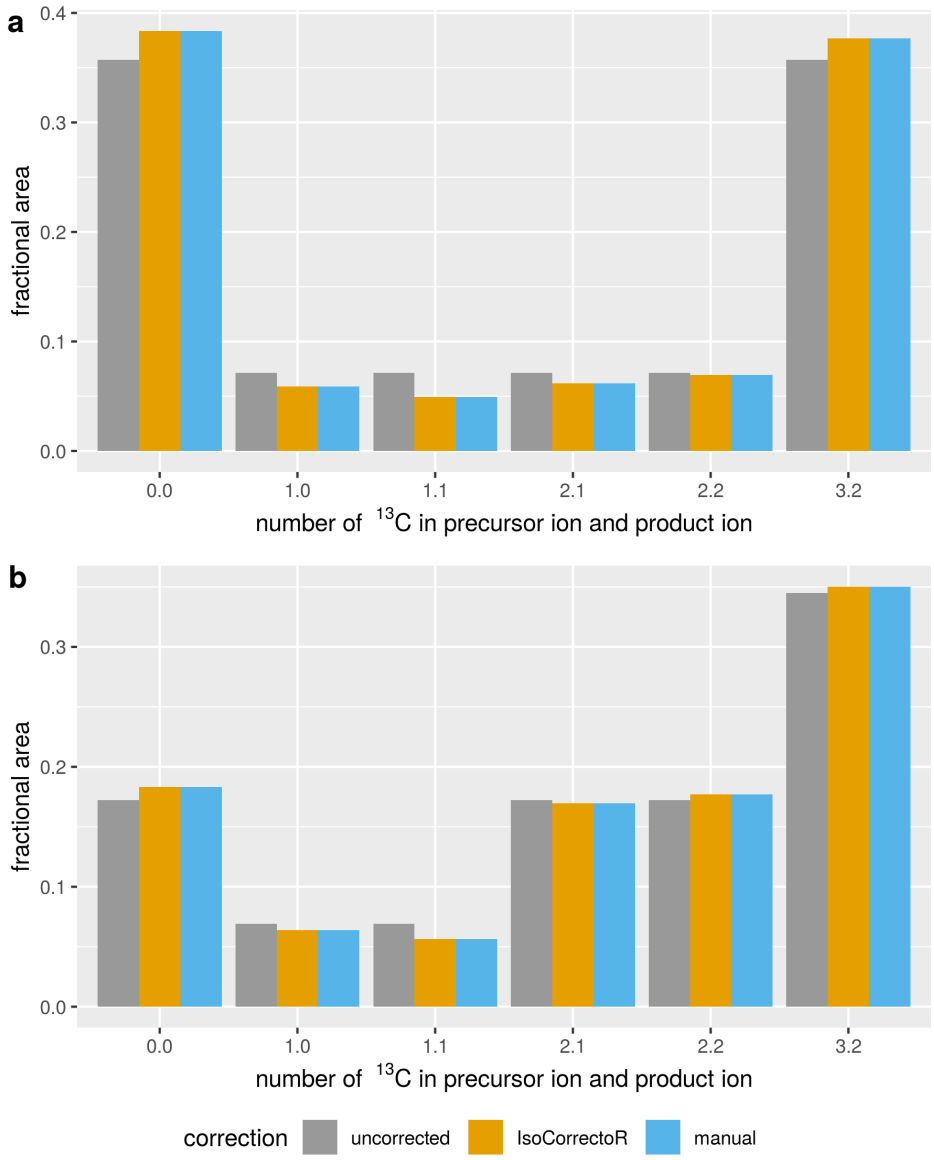



Figure II.42: ***Comparison of resolution-dependent MS/MS correction to manual calculations on example data.*** *The figure shows the results obtained from correcting simulated MS/MS data of PCF-derivatized alanine. The x-axis labels n.m correspond to the MS/MS transitions with n $^{13}C$ in the precursor ion and m $^{13}C$ in the product ion. Data are shown uncorrected, corrected using IsoCorrectoR's resolution-dependent MS/MS correction mode and corrected via manual calculations. Resolution-dependent correction was performed for the usual case of having a low-resolution precursor mass analyzer (1000 at m/z 200, FWHM constant over the m/z range) and a high resolution product ion mass analyzer (140000 at m/z 200, orbitrap m/z dependency of FWHM). Panels a and b correspond to different uncorrected MIDs.*

### 3.3.5 Ultra-high-resolution correction of multiple-tracer data

**Comparison to PyNAC** The UHR correction approach of IsoCorrectoR was validated against PyNAC, however without considering tracer purity, as this option is not available in PyNAC. Simulated UHR data of simultaneously $^{13}C$ and $^{15}N$ labeled PCF-derivatized glycine, asparagine and arginine was used for the comparison (uncorrected MIDs were derived manually). The heatmap in Figure II.43 shows the results when subtracting MIDs obtained by UHR correction with PyNAC from the MIDs obtained through UHR correction with IsoCorrectoR. A maximum difference in the range of 2E-11 indicates excellent agreement between the two tools.

**Comparison to manual calculations** UHR correction was also validated via manual calculations. The manual calculations were performed as described previously in section 3.3.1 and in the methods chapter, by manually calculating a probability matrix in a Microsoft Excel spreadsheet and subsequently passing the matrix and uncorrected values to the *linsolve* function of Matlab. Simulated UHR data of simultaneously $^{13}C$ and $^{15}N$ labeled PCF-derivatized glycine was used for the comparison (uncorrected MIDs were derived manually). A tracer purity of 99% was assumed for both tracer elements for correction. Figure II.44 shows the results for two different uncorrected MIDs and suggests an excellent match. The maximum observed difference between IsoCorrectoR and manual correction was in the range of 5.5E-11. Figures A22 and A23 in the appendix show the comparison for additional uncorrected MIDs.

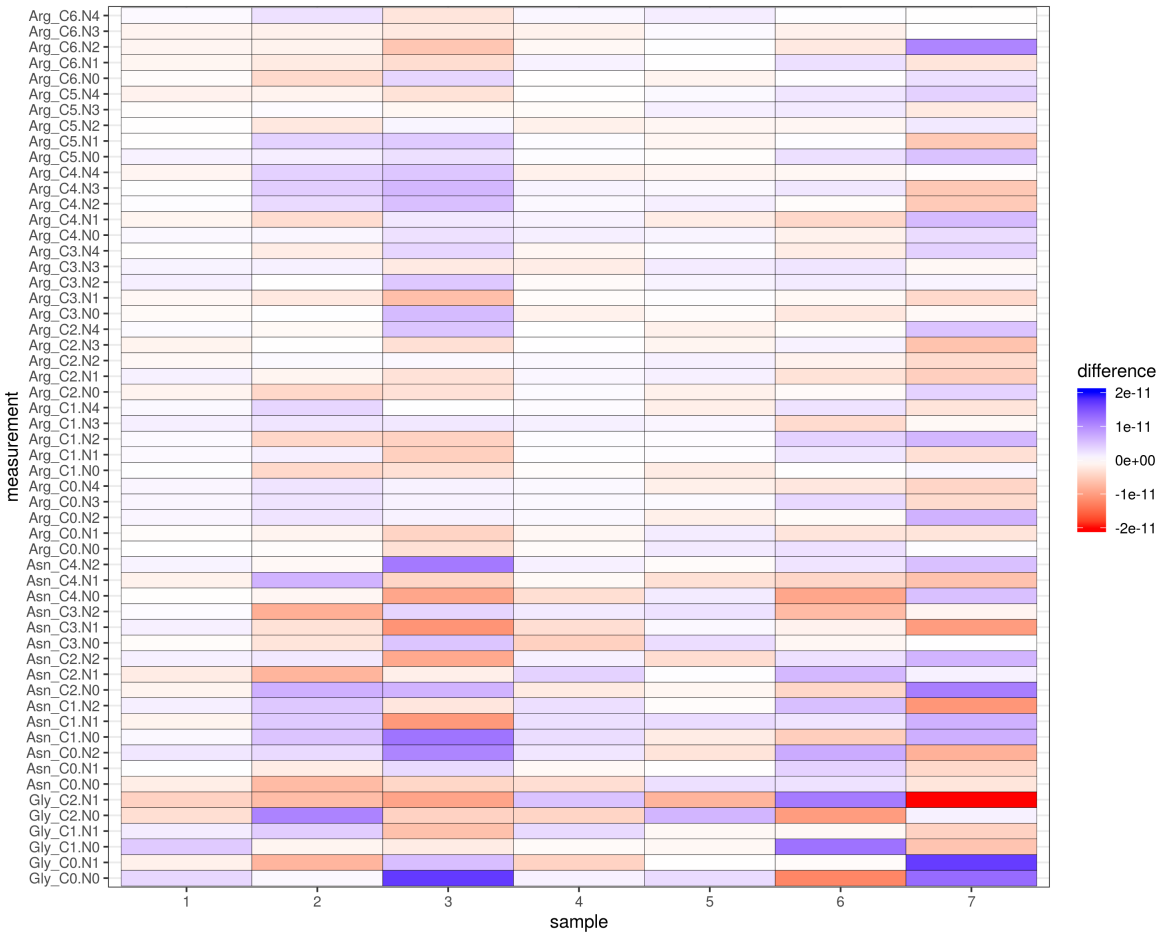



Figure II.43: ***Heatmap showing differences between IsoCorrectoR and PyNAC for the UHR correction of data from a simultaneous $^{13}C$ and $^{15}N$ labeling experiment.*** *Uncorrected UHR data of simultaneously $^{13}C$ and $^{15}N$ labeled PCF-derivatized glycine, asparagine and arginine was simulated. Correction was performed without accounting for tracer purity. The heatmap was generated by subtracting MIDs obtained by UHR correction with PyNAC from the MIDs obtained through UHR correction with IsoCorrectoR. Columns represent samples, while rows represent measurements of isotopologues. The number after the metabolite name corresponds to the number of label incorporated.*

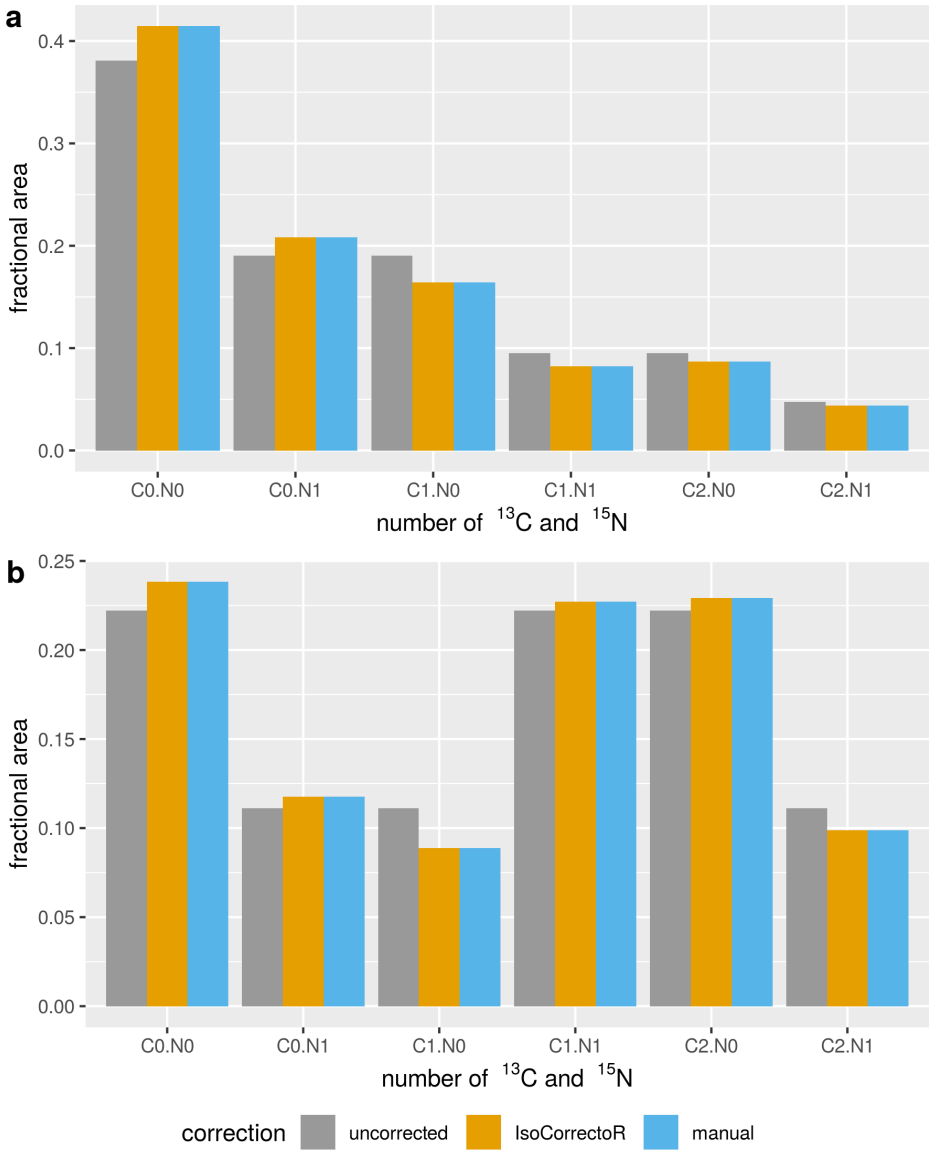



Figure II.44: ***Comparison of UHR correction to manual calculations on simulated UHR example data of simultaneously $^{13}C$ and $^{15}N$ labeled PCF-derivatized glycine.*** *A tracer purity of 99% was assumed for correction. The x-axis labels Cx.Ny correspond to x $^{13}C$ label and y $^{15}N$ label incorporated in the respective isotopologue. Data are shown uncorrected, corrected with IsoCorrectoR's UHR correction algorithm and corrected through manual calculations. The panels a and b correspond to different uncorrected MIDs.* ***(Data similar to panel b are also shown in supplementary figure 4 of Heinrich et al. (2018), albeit with an assumed tracer purity of 98%)***

### 3.3.6 Impact of the calculation threshold parameter

The calculation threshold parameter of IsoCorrectoR can be used to limit probability computations to values above a given threshold (for all modes of correction except UHR correction) to save computational resources. The default value is 1E-08. UHR correction uses a different type of calculation threshold, with a default value of 8. For more information on these calculation thresholds, see 3.1.10 in the algorithm section. To assess the difference between the default value of 1E-08 and using no threshold at all (a value of 0), $^{13}$C stable isotope isotope labeling data supplied with AccuCor (Su et al. (2017) have been corrected with IsoCorrectoR's low-resolution MS correction mode using either a calculation threshold value of 1E-08 or 0. In low-resolution correction, the differences should be most pronounced, as all NA contributions that match the nominal mass shift of a given labeling state are removed during correction (when not using a calculation threshold). When applying a threshold, contributions with a probability below the threshold value will not be computed and thus not be removed. See the heatmap in Figure II.45 for a visualization of the results. The maximum difference of 9E-08 can be considered negligible. It is found for NADH, which can produce a very large set of NA derived species due to its size, some of them with very low probability. Figure A24 in the appendix shows the results for a different set of molecules from the dataset, yielding similar results. Even though they have been acquired at high resolution, the AccuCor data have been selected for this analysis because of the variation in molecular species (and, specifically, the presence of NADH data). A similar analysis has been performed for assessing the impact of the UHR calculation threshold, correcting *e.g.* simultaneosly $^{13}$C and $^{15}$N labeled arginine (maximum label: 6 $^{13}$C and 4 $^{15}$N) with the UHR correction mode of IsoCorrectoR. However, when comparing the default threshold of 8 to a threshold of 100, no differences were found. It is likely that the probabilities associated with a labeling difference of $> 8$ are so low that they are in the range of numeric imprecision, resulting in an effective probability value of 0 in subsequent calculations.

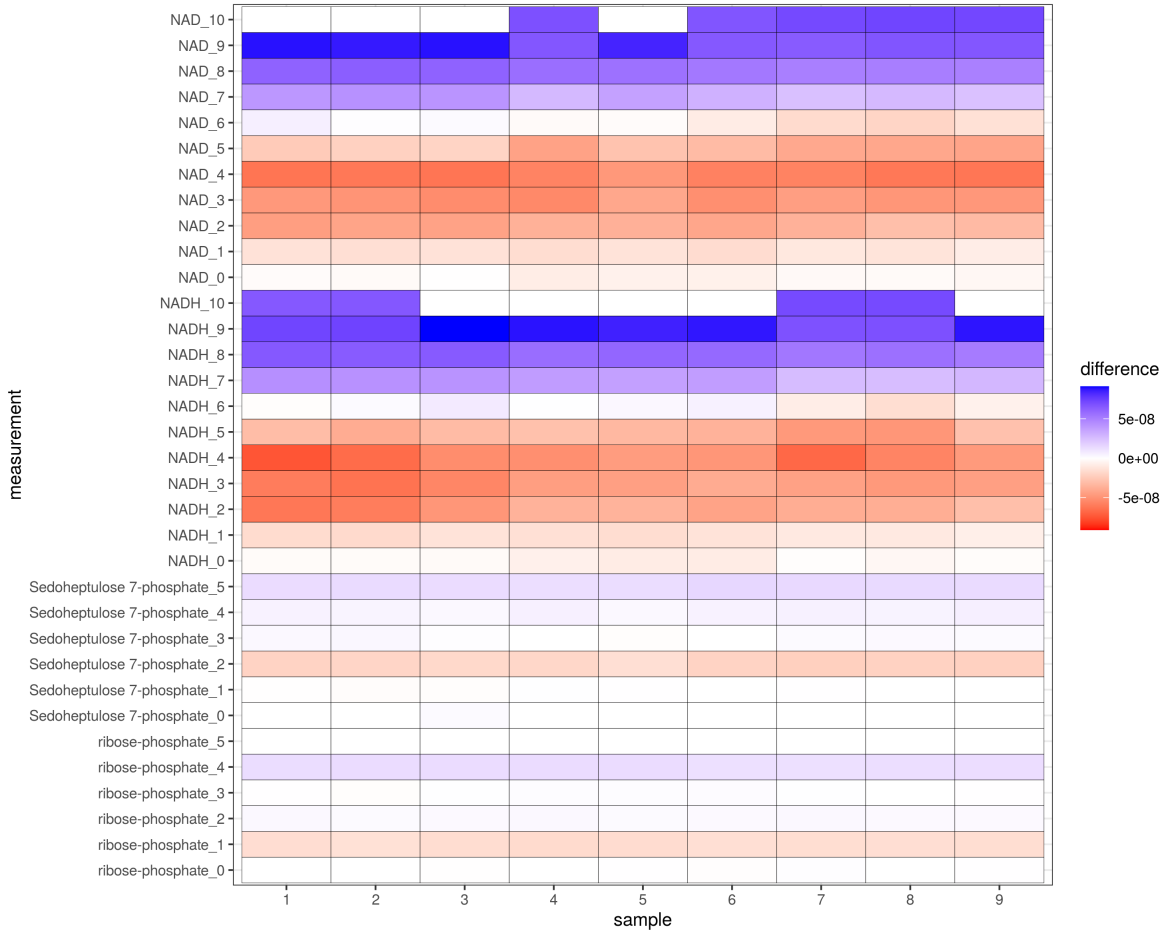



Figure II.45: ***Heatmap showing differences between performing IsoCorrectoR's low-resolution MS correction with the default calculation threshold and without calculation threshold on data from a $^{13}C$ labeling experiment.*** *Data used are example data supplied with AccuCor (Su et al. (2017)). The default calculation threshold is 1E-08, a tracer purity of 0.99 was assumed. The heatmap was generated by subtracting MIDs corrected without threshold from MIDs corrected with the default threshold. Columns represent samples, while rows represent measurements of mass isotopomers. The number after the metabolite name corresponds to the number of label incorporated.*

# Chapter III

# $^{13}$C tracer analysis in a B-cell lymphoma model cell line

## 1 Introduction: Cancer metabolism and B-cell lymphoma

Altered metabolism is considered one of the hallmarks of cancer, as tumor cells reprogram metabolism to satisfy the core metabolic needs of cancer cells: energy production, biomass generation and redox control (Hanahan & Weinberg (2011)). A classical example for such metabolic reprogramming often observed in malignant cells is the Warburg effect, which is also termed aerobic glycolysis. Here, cells take up large amounts of glucose and produce large amounts of lactate, even in the presence of oxygen (DeBerardinis & Chandel (2016)). The main goal of this process seems to be the provision of building blocks for growth and proliferation, as glucose does not only serve energy metabolism, but also *e.g.* provides acetyl-CoA for fatty acid synthesis, precursors of non-essential amino acids and ribose (Vander Heiden et al. (2009)). Apart from glucose, also glutamine is often critical for cancer cell proliferation, as it is used *e.g.* for TCA cycle anaplerosis, energy production, the synthesis of several non-essential amino acids and nucleotide biosynthesis (Jin et al. (2016), Yuneva et al. (2007)). Especially products of the TCA cycle and nucleotide biosynthesis can be limiting for proliferation (Heiden & DeBerardinis (2017)). Figure III.1 from Heiden & DeBerardinis (2017) shows a map of metabolic pathways important in cancer.

Often, alterations of metabolism in cancer cells can be attributed to oncogenes or tumor suppressor genes which serve as upstream regulators of metabolism. The transcription factor c-myc, for instance, regulates the expression of metabolite transporters as well as enzymes required for glucose and glutamine metabolism. Another example for a central coordinator of metabolism is mTOR complex 1. It is activated through growth factor signaling networks and increases protein, nucleic acid and lipid synthesis through various mechanisms. However, also the cell lineage, the tumor microenvironment and nutrient availability play a major role in determining the metabolic routes predominantly used in malignant cells. Eventually, characterizing tumor metabolism may lead to novel therapeutic approaches that are based on specific metabolic features of the given type of cancer (Heiden & DeBerardinis (2017)).

Lymphomas are solid tumors of the immune system. The classification distinguishes between Hodgkin and non-Hodgkin lymphomas, the latter accounting for about 90% of all cases and

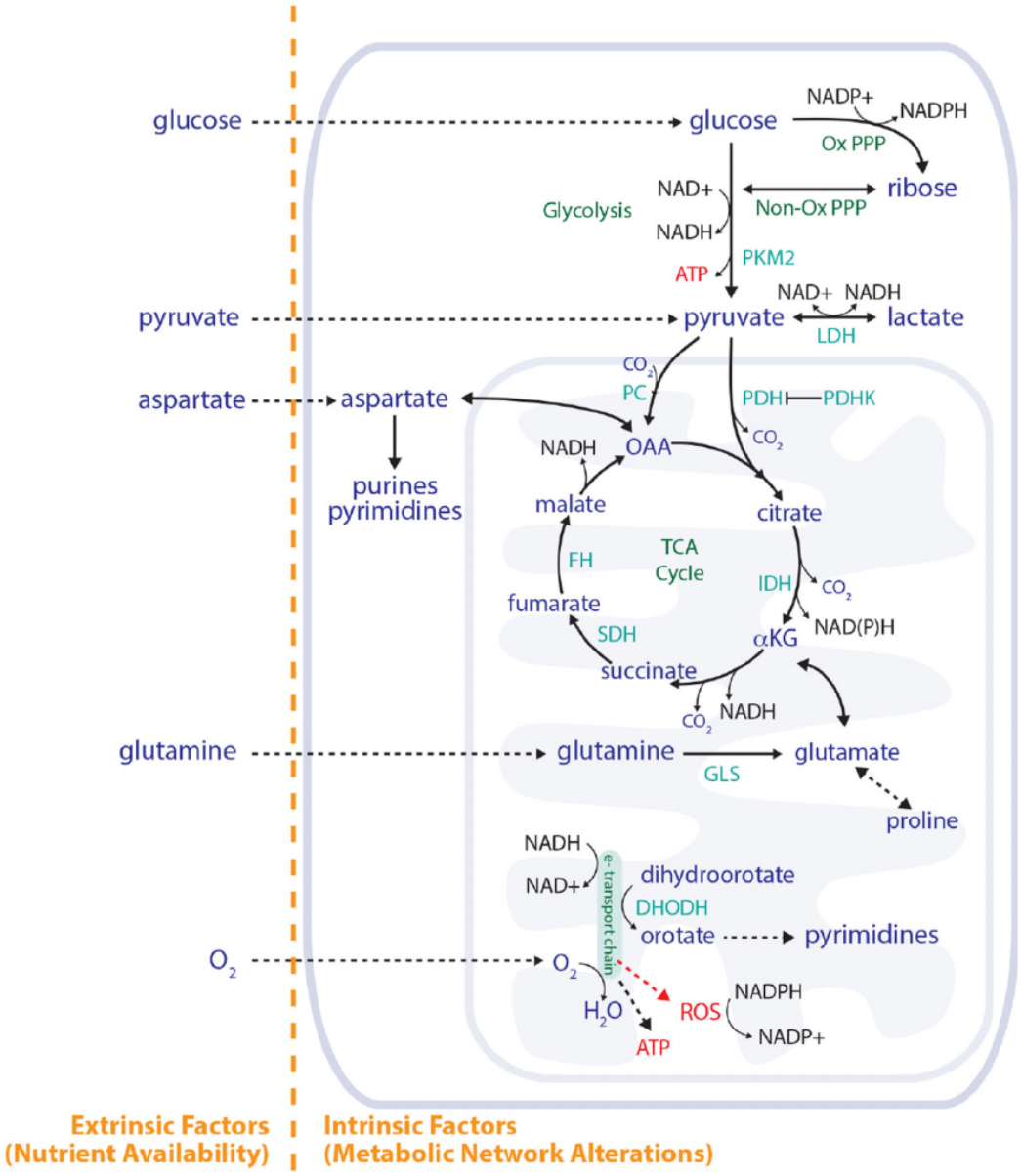



Figure III.1: ***Map of metabolic pathways.*** *Both nutrient availability and metabolic network configuration affect how cells use metabolism to produce ATP, generate macromolecules, and regulate redox state. Key reactions in central carbon metabolism are shown, including how the TCA cycle and the electron transport chain are involved in purine and pyrimidine synthesis. Abbreviations: PKM2, pyruvate kinase M2; LDH, lactate dehydrogenase; PDH, pyruvate dehydrogenase; PDHK, pyruvate dehydrogenase kinase; PC, pyruvate carboxylase; IDH, isocitrate dehydrogenase; SDH, succinate dehydrogenase; FH, fumarate hydratase; GLS, glutaminase; DHODH, dihydroorotate dehydrogenase; Ox PPP, oxidative pentose phosphate pathway; Non-Ox PPP, non-oxidative pentose phosphate pathway; aKG, a-ketoglutarate; OAA, oxaloacetate; ROS, reactive oxygen species.* ***(Figure and legend taken from Heiden & DeBerardinis (2017) with permission from Elsevier, legend modified)***

being the fifth most frequently diagnosed cancer in the UK. Today, the proportion of non-Hodgkin lymphoma patients that is expected to survive for longer than 10 years is 50.8% (data from England and Wales, Shankland et al. (2012)). Non-Hodgkin lymphomas are a very heterogeneous group of cancers, they usually develop in the lymph nodes but can occur in almost any tissue. The group comprises for example follicular lymphoma which are relatively indolent and the more aggressive Burkitt lymphomas and diffuse large B-cell lymphomas (DLBCL). DLBCLs are the most abundant type of non-Hodgkin lymphoma, making up about one third of the cases diagnosed (Shankland et al. (2012)).

Burkitt lymphomas (BL) are highly aggressive and derived from germinal center B cells. BL development is dependent on and characterized by inappropriately high levels of the transcription factor c-myc, which regulates proliferation, differentiation and apoptosis. It is encoded by the *MYC* gene. The most common genetic lesion that results in *MYC* overexpression is the translocation of the long arm of chromosome 8 (containing *MYC*) and the Ig heavy chain gene on chromosome 14 (Kalisz et al. (2019)). Lymphoma classified as DLBCL vary strongly with respect to clinical presentation, morphology and molecular pathogenesis. In the last few years, the understanding of the molecular basis of different DLBCL subtypes has increased substantially, enabling novel diagnostic and therapeutic approaches. At least three different DLBCL subtypes are distinguished according to their gene expression profiles, namely germinal center B-cell like (GCB), activated B-cell like (ABC) and primary mediastinal B-cell lymphoma (PMBL). They differ significantly in their response to standard treatment: While a combined treatment approach employing the anti-CD20 antibody rituximab and chemotherapy is effective for GCBs, survival for ABC cases is only 50%. Thus, the molecular differences between the specific subtypes need to be considered to be able to target the malignancies more effectively. GCBs likely originate from germinal center B-cells. They very often show a constitutive activation of the anti-apoptotic BCL2 protein and a down-regulation of PTEN. PTEN otherwise counteracts the activation of the PI3K/mTOR signaling pathway which mediates proliferation, cell growth and survival. Gene expression profiles of ABC DLBCLs suggest that this subtype is derived from B cells that are in the process of differentiation to plasma cells. Here, constitutive activation of the proliferation- and survival promoting NF-$\varkappa$B signaling pathway is an important pathogenic feature, sometimes due to chronic active B-cell receptor (BCR) signaling (Frick et al. (2011)).

The B-cell line P493-6 has been established as a Burkitt lymphoma model cell line. It is based on the EBV- (Epstein-Barr-virus) positive EREB2-5 B-cell line, as EBV is often an important cofactor in BL pathogenesis. To create P493-6, EREB2-5 has been stably transfected with a DNA construct that allows conditional *MYC* expression under the control of a tetracycline-regulated promoter. In the absence of tetracyline, *MYC* is constitutively expressed (MYC-high state), while promoter activation is abolished if tetracycline is present (MYC-low state). In the MYC-high state, P493-6 cells proliferate and mimic BL cells, while the MYC-low state resembles resting, non-proliferating B-cells in G0/G1-arrest (Pajic et al. (2000)). However, it has been shown that microenvironmental stimuli can induce proliferation in MYC-low cells, comparable to the MYC-high state. Specifically, a combination of the Interleukin-10 receptor agonist Interleukin 10 (IL10) and the Toll-like receptor 9 agonist CpG could synergistically promote proliferation via joint STAT3/NF-$\varkappa$B activation. Individual application of IL10 or CpG alone did not have this effect (Feist et al. (2017)). An important question arising from these findings is the following: How is the metabolic network in P493-6 B-cells reprogrammed to enable proliferation in response to those stimuli? This may provide valuable insights with regard to B-cell lymphomas with aberrant STAT3 and/or NF-$\varkappa$B signaling.

It has been shown that *MYC*-overexpressing cells can be particularly sensitive to glutamine withdrawal (Yuneva et al. (2007)). To assess the usage of the central nutrient glutamine in P493-6 cells in the MYC-high, MYC-low and IL10+CpG-stimulated state, a $^{13}C$ tracer analysis experiment was conducted with U-$^{13}C$-glutamine as the tracer substrate. In the following, the results from this tracing experiment will be presented. Furthermore, the findings will be discussed in the context of other experiments from Feist et al. (2018), including nutrient deprivation and rescue experiments or the inhibition of metabolic pathways.

## 2 Contributions

This work was part of a collaborative study on the metabolism of P493-6 B-cells, "Cooperative STAT/NF-κB signaling regulates lymphoma metabolic reprogramming and aberrant GOT2 expression" (Feist et al. (2018)). $^{13}C$ tracer analysis in P493-6 B-cells with the tracer substrate U-$^{13}C$-glutamine was performed by Paul Heinrich. This included planning the $^{13}C$ tracer analysis experiments in cooperation with Maren Feist (Clinic of Haematology and Medical Oncology, University Medical Centre Göttingen), preparation of cell pellets and cell supernatants for LC-MS/MS analysis, LC-MS/MS measurement of samples as well as the analysis and interpretation of the $^{13}C$ tracer analysis results. In order to assess isotopic steady state for several relevant metabolites, Paul Heinrich also planned the associated experiments together with Maren Feist, prepared the LC-MS samples and performed LC-MS/MS measurements and data analysis.

The cell culture work with P493-6 B-cells, including nutrient withdrawal and metabolite rescue cell proliferation experiments and oxygen consumption rate experiments, were performed by Maren Feist. Maren Feist also performed gene expression analyses, RNAi knockdown experiments and other molecular biology related work. Absolute quantification of metabolites was performed by Philipp Schwarzfischer (Institute of Functional Genomics). $^{15}N$ tracer analysis with α-$^{15}N$-glutamine, as well as tracer analysis with 1-$^{13}C$-glutamine was performed by Xueni Sun (Institute of Functional Genomics).

## 3 Results and Discussion

P493-6 B-cells in the MYC-low, MYC-high and IL10+CpG stimulated states were cultured in the presence of the tracer substrate U-$^{13}C$-glutamine. The aim was to assess the usage of glutamine for the biosynthesis of amino acids and TCA cycle intermediates/lactate, and to analyze how it differs between the different cellular states. Preliminary nutrient withdrawal experiments by Maren Feist at the University Medical Centre Göttingen had shown that proliferation of MYC-high and IL10+CpG stimulated MYC-low cells (abbreviated with IL10+CpG) drastically decreased upon glutamine withdrawal, whereas the effects of glucose withdrawal were not as pronounced. Further, the intracellular amounts of amino acids and TCA cycle intermediates (analyses by Philipp Schwarzfischer, Institute of Functional Genomics), as well as the expression of glutaminolysis and glycolysis genes were markedly increased by IL10+CpG costimulation compared to MYC-low cells. This is summarized in Figure A25 in the appendix, taken from Feist et al. (2018). These findings were accompanied by increased cellular protein amounts and a cell size increase of IL10+CpG cells comparable to the size of MYC-high cells (Feist et al. (2018), gene expression, protein amount and cell size analyses by Maren Feist

at the University Medical Centre Göttingen). To determine the degree of $^{13}C$ labeling and labeling patterns in amino acids and TCA cycle intermediates/lactate in cell extracts and supernatants of P493-6 cells, an LC-MS/MS approach employing a low-resolution QTRAP mass spectrometer was used (see section 2.1 in the methods chapter). For each metabolite of interest, transitions were measured to cover all distinct isotopologues that may originate from the labeling experiment. Correction for NA and tracer purity was performed using IsoCorrectoR's low-resolution MS/MS correction algorithm and assuming a tracer purity of 99%.

## 3.1 Assessment of isotopic steady state

As stated previously in section 2 in the general introduction, isotopic steady state (ISS) is usually necessary to reliably assess the quantitative contribution of the tracer substrate to the synthesis of a given metabolite. Otherwise, differences in the degree of labeling between different cellular conditions may *e.g.* simply arise due to different rates of proliferation. In the case of P493-6 cells, highly proliferating MYC-high cells would be expected to incorporate substantially more tracer per unit time than MYC-low cells due to their increased need for biosynthesis. Thus, at a given point in time prior to ISS, the degree of labeling in MYC-high cells may be markedly higher than in MYC-low cells. However, this does not necessarily mean that the relative contribution of the tracer to the biosynthesis of a given metabolite is actually higher, but can also be a result of the slower labeling rate of metabolite pools in MYC-low cells. If a metabolite pool is in ISS, however, the degree of labeling reflects the nutrient contribution of the tracer substrate and thus its importance for the synthesis of that metabolite. In order to be able to determine the presence of ISS for the various metabolites analyzed, a time course experiment was conducted. P493-6 cells were cultured with U-$^{13}C$-glutamine in the MYC-low, MYC-high or IL10/CpG state for 30 h, and cells were harvested at the time points 18 h, 22 h, 24 h, 26 h and 30 h. Transitions of expected isotopologues in the cell extracts (biological triplicates) were then measured via LC-MS/MS. For assessing isotopic steady-state, transitions with the same number of label in the precursor were summed up, as there is no noteworthy gain in interpreting transitions individually in that setting (MS/MS was chosen merely for sensitivity and specificity, not for gathering positional labeling information).

Figures A26 to A38 in the appendix show the results. When considering the amino acids, the mass isotopomer distributions (MIDs) seem to be fairly stable over the time course. There may be a minor trend towards lower label incorporation for MYC-low aspartate and glutamate, MYC-high ornithine and for proline in general. However, these are only very small effects and it should be viable to assume approximate ISS. A small trend towards lower labeling may be indicative of a slow transition from glutamine-based biosynthesis to other sources in the course of the cell culture experiment, so metabolic steady state is probably not entirely granted. Again, given the small scale, this should not be an issue for the following analyses. This is also supported by statistical tests performed on the time course data. For each labeling state of each molecule and cellular state (MYC-low, MYC-high, IL10/CpG), an ANOVA was performed across the different time points. After correction for multiple testing via the Benjamini-Hochberg procedure, only the tests for MYC-high m+1 glutamate and MYC-high m+1 aspartate showed significant differences between time points (FDR of 0.05, see Tables A1 to A3 in the appendix for detailed testing results). A one-way ANOVA across the time points of each labeling state was chosen instead of a two-way ANOVA across both labeling states and time points, because testing for differences between labeling states was not of interest. Thus, excess statistical testing that would reduce statistical power was avoided. A note on statistical

testing in this specific context: The absence of significant differences between time points of course does not proof that there are no differences between time points. Absence of evidence is not evidence of absence. What can be stated is that the observed time course data are not unlikely under the assumption that values are equal across time points.

Regarding the TCA cycle intermediates 2-ketoglutarate, succinate, fumarate and malate, measurements partly showed very high variability among the biological replicates, making it difficult to reliably assure ISS. In the case of malate, the high variability seems to originate from large differences between only the m+0 values, which were probably caused by an impurity in the samples coming from one of the solutions used for cell extraction. This is supported by the observation that very high m+0 values are found in the samples processed on the first day of sample preparation, while the samples processed on the second day show normal m+0 values. Thus, for assessing malate ISS, the m+0 values were removed from the MIDs. MIDs constructed from the remaining mass isotopomer fractions appear very stable over the time course, indicating ISS is reached for malate (Figure A37 in the appendix). Fumarate, 2-ketoglutarate and succinate were measured twice, as the original MS/MS method for measuring TCA cycle intermediates was incorrect. It lacked several transitions for 2-ketoglutarate and fumarate, and was therefore updated to cover all relevant mass isotopomers (see section 2.1 in the methods chapter). However, in the repeated analysis, fumarate data showed very high variability, so that the original, more stable measurements were used for assessing ISS, lacking the transitions 1.1 and 2.2 (x.y means x label in the precursor and y label in the product ion). The remaining fumarate transitions (Figure A34 in the appendix) are stable over the time course and should be sufficient to reliably assure ISS. Succinate was remeasured because of high variability in the original measurements. However, the second measurement led to no improvement. ISS may be assumed for succinate MYC-low and IL10+CpG, but it is questionable in the case of MYC-high. For 2-ketoglutarate, 4 out of 10 transitions were missing in the original MS/MS method (1.0, 1.1, 3.2, 3.3). In the measurements with the new method (Figure A35 in the appendix), variability was still comparably high as in the original measurements, especially for MYC-low. It should however be viable to assume ISS for IL10+CpG and MYC-high. This is also supported by statistical tests performed analogously to those on the amino acid data. No significant differences between time points could be observed at a FDR of 0.05. The test for MYC-low m+2 2-ketoglutarate could not be performed because of too many zeros in the data, however (see Tables A4 to A6 in the appendix). A reason for the relatively high variability among measurements of TCA cycle intermediates may be that they are often not as abundant as amino acids. Since only $2x10^6$ cells were harvested per sample for the time course experiment (opposed to $4x10^6$ cells for the single 24 h time point used for the remaining labeling analyses), it is likely that measurements close to or below LLOQ (lower limit of quantification) substantially added to the variation for some species. Indeed, the area values observed for succinate, 2-ketoglutarate and fumarate were roughly between 1 and 2 orders of magnitude lower than those of citrate and malate and most amino acids. The MIDs of citrate and lactate were stable over time, so ISS may be assumed for these metabolites.

## 3.2 Analysis of label incorporation

### 3.2.1 Mean $^{13}C$ enrichment

To analyze the contribution of glutamine to the biosynthesis of a given metabolite, P493-6 cells were cultured with U-$^{13}C$-glutamine in the MYC-low, MYC-high or IL10/CpG cellu-

lar states for 24 h, followed by LC-MS/MS measurement of the metabolite labeling patterns. Figures III.2 and III.3 show the mean $^{13}$C enrichment of amino acids and TCA cycle intermediates/lactate in cell extracts, respectively. The mean (isotopic) enrichment of a metabolite corresponds to the total fraction of labeled atoms of the tracer element it contains and thus reflects the degree of label incorporation. It is computed from the MID as follows: $ME = \sum_{i=0}^{n} \frac{f_i \cdot i}{n}$. Here, $ME$ is the mean enrichment, $f_i$ is the fraction of the isotopologue with $i$ labeled atoms in the MID, $i$ is the number of labeled atoms, and $n$ is the maximum number of labeled atoms the metabolite can contain.

From Figures III.2 and III.3 it is evident that glutamate, aspartate, proline and the TCA cycle intermediates are substantially $^{13}$C enriched regardless of cellular state. Mean enrichment for MYC-low is usually in the range of 25%, while it is around 45% and 60% for IL10+CpG and MYC-high, respectively. Hardly any enrichment can be found for alanine, glycine, serine and lactate. For ornithine, there appears to be minor enrichment. These results are not unexpected, as the tracer substrate glutamine can be directly converted into glutamate via glutaminase once it has been imported into the cells. Subsequently, glutamate can be transformed to either proline or ornithine via the intermediate glutamate-5-semialdehyde, retaining all of its carbon backbone and thus propagating $^{13}$C label directly to these amino acids. Alternatively, glutamate may be converted to the TCA cycle intermediate 2-ketoglutarate through the action of transaminases or glutamate dehydrogenase. Oxidative decarboxylation of 2-ketoglutarate via the 2-ketoglutarate dehydrogenase complex and further TCA transformations then yield succinate, fumarate, malate and oxaloacetate, explaining the high degree of $^{13}$C enrichment in TCA cycle intermediates. Oxaloacetate can be transaminated to aspartate via the action of aspartate aminotransferase, obtaining the amino group from glutamate. It is thus reasonable that the enrichment in aspartate resembles that of the four carbon TCA cycle intermediates fumarate and malate, which are direct predecessors of oxaloacetate. The labeling of oxaloacetate itself could not be determined due to difficulties regarding the LC-MS/MS measurement. In the oxidative TCA, citrate is formed via the condensation of (usually) glucose or fatty acid derived acetyl-CoA and oxaloacetate. As the added acetyl-CoA will usually be unlabeled, the observed lower degree of labeling in citrate compared to fumarate, malate and 2-ketoglutarate is not unexpected (p = 7.06E-05 for Welch t-test between MYC-high citrate and fumarate mean enrichment). Lactate results from the reduction of pyruvate through lactate dehydrogenase. Alanine is also derived from pyruvate, via alanine aminotransferase, while serine is synthesized from 3-phosphoglycerate. Glycine is produced from serine. As 3-phosphoglycerate and pyruvate are both products of glycolysis, lactate, alanine, serine and glycine should be primarily derived from (unlabeled) glucose, explaining the very low degree of labeling in these metabolites. Although ornithine can be synthesized from glutamate, and the time-course of ornithine labeling patterns (Figure A30 in the appendix) indicated ISS, the labeling in ornithine is low compared to the other glutamate-derived amino acids. Probably, ornithine is mainly synthesized from unlabeled arginine from the cell culture medium, via the urea cycle enzyme arginase.

It should be noted that a glutamate mean enrichment of about 20% was found in the cell culture medium control samples, where no biological/metabolic activity should have occured (see Figure A39 in the appendix, showing the mean enrichment of amino acids in cell supernatant and medium control samples). This is likely the result of spontaneous hydrolysis of the glutamine tracer substrate to glutamate. Indeed, when investigating the exact labeling patterns of glutamate in the medium control (see Figure A43 in the appendix), only the m+0 and m+5 species are present (after NA and tracer purity correction), suggesting direct spon-

taneous hydrolysis instead of biological activity, where a more complex labeling pattern would be expected. Regardless of origin, though, the labeled glutamate in the medium control may affect the investigated labeling patterns. Glutamate imported by the cells may be used for the synthesis of other amino acids and TCA cycle intermediates directly, instead of or in addition to glutamine. However, even if the synthesis of a given metabolite was purely based on directly imported glutamate and not glutamine (which is unlikely, as glucose and glutamine are the most commonly consumed nutrients), a maximum enrichment of around 20% could not be exceeded. As the labeling in most species with relatively high enrichment clearly exceeds 20%, glutamine must have a substantial role in their biosynthesis. Thus, while the labeled fraction of a given metabolite is derived from either labeled glutamine or glutamate, the unlabeled fraction may *e.g.* be derived from glucose or unlabeled amino acids from the cell culture medium.

When focusing on differences in mean enrichment between the different cellular states of P493-6 cells, it becomes evident that mean enrichment generally increases from MYC-low over IL10+CpG stimulated MYC-low to MYC-high, for metabolites that show marked enrichment (glutamate, aspartate, proline, citrate, 2-ketoglutarate, fumarate, malate). The presence of statistically significant differences between mean enrichment group means (MYC-low, IL10+CpG and MYC-high) was tested for with a one-way ANOVA (analysis of variance) approach. In the case of normally distributed ANOVA model residuals (insignificant Shapiro-Wilk test), a Welch-type ANOVA (variances could not be assumed to be equal) was used. The non-parametric Kruskal-Wallis ANOVA was used in the case that the normality assumption had to be rejected. Subsequently, the p-values obtained from the ANOVA tests were corrected for multiple testing using the FDR (false discovery rate) approach of Benjamini and Hochberg. Results were considered to be statistically significant if their associated q-value (FDR) was $\leq 0.05$. This way, the differences for all metabolites except glycine, ornithine, serine and lactate were found to be statistically significant. Subsequently, for metabolites that showed a significant ANOVA result, post-hoc tests for pairwise comparisons were used to test for significant differences between pairs of groups. To this end, Welch's t-test was used regardless of whether the normality assumption had to be rejected in the Shapiro-Wilk test for one or both of the compared groups. Using the non-parametric Mann-Whitney-U test would not have been sensible given the low sample size (see 4.2 in the Methods section). After Benjamini-Hochberg FDR correction for multiple testing, all pairwise comparisons except succinate IL10+CpG vs. MYC-high were significant ($q \leq 0.05$). The detailed results of the performed significance tests are shown in Tables III.1 and III.2 for ANOVA and post-hoc tests, respectively. The normality assumption had to be rejected for the citrate IL10+CpG group and the 2-ketoglutarate MYC-low group. However, rejecting normality based on a sample size of 3 is not necessarily reasonable. In the case of citrate and 2-ketoglutarate, normality is likely rejected because 2 of the 3 data points are very close to each other, something that is not unlikely to happen when measuring, also if measurements are usually normally distributed.

Thus, for glutamate, aspartate, proline, citrate, 2-ketoglutarate, fumarate and malate, the mean enrichment in IL10+CpG stimulated MYC-low cells is significantly higher than in MYC-low cells, while the mean enrichment in MYC-high cells is again significantly higher than in IL10+CpG stimulated cells. This means that, if ISS can be assumed, the usage of glutamine for the biosynthesis of said metabolites increases from MYC-low over IL10+CpG to the MYC-high state, and is likely substantial, mostly in the range of 45% or 60% enrichment for IL10+CpG and MYC-high, respectively. It has to be considered, however, that labeled glutamate from the cell culture medium may also contribute, and that ISS could not be assured for MYC-low

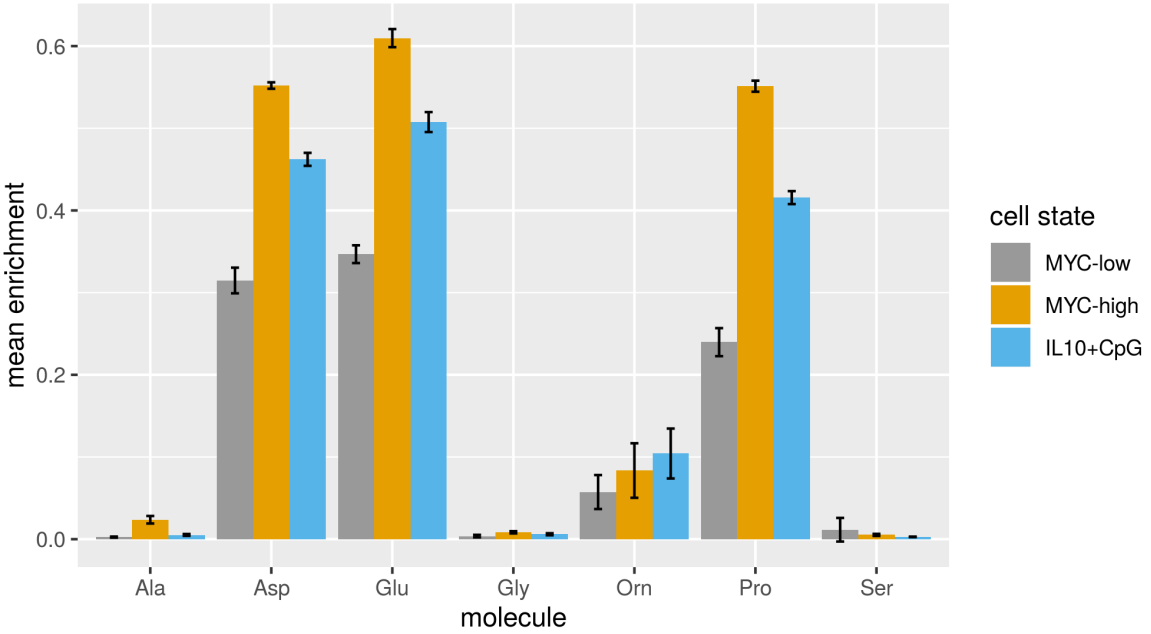



Figure III.2: ***$^{13}$C tracer analysis with the tracer substrate U-$^{13}$C-glutamine in P493-6 cells, mean $^{13}$C enrichment of amino acids.*** *The figure shows the mean $^{13}$C enrichment of PCF-derivatized amino acids from cell extracts for the different cell states MYC-low, MYC-high and IL10+CpG stimulation. Cells were harvested after 24 h of incubation with the tracer substrate. Data were corrected for NA and tracer purity using IsoCorrectoR, assuming a tracer purity of 99%. Samples were measured in (biological) triplicates, means of mean enrichment among replicates +/- SD are shown.* ***(These data are also shown in Figure 2a of Feist et al. (2018), but the figure was recreated for this thesis.)***

2-ketoglutarate. For succinate, IL10+CpG enrichment and MYC-high enrichment could be shown to be significantly higher than the enrichment in MYC-low. However, as mentioned above, succinate IL10+CpG vs. MYC-high was insignificant. Further, ISS is questionable for MYC-high succinate. The discussed results regarding the differences between MYC-low and MYC-high are supported by a another $^{13}$C tracer analysis study in P493-6 cells with glutamine as the tracer substrate. In that study, the labeling of TCA cycle intermediates was greatly enhanced in the MYC-high compared to the MYC-low state, too (Le et al. (2012)). In an additional study on $^{13}$C metabolic flux analysis in P493-6 cells, employing $^{13}$C labeled glucose as the tracer substrate, most TCA fluxes were found to be increased near 4-fold in MYC-high compared to MYC-low cells (Murphy et al. (2013)). These strongly increased TCA fluxes, combined with the finding that glutamine contributes substantially to the synthesis of TCA intermediates under MYC-high conditions, indicate that large amounts of glutamine derived carbon enter and leave the TCA, presumably serving both energy production and anabolic processes.

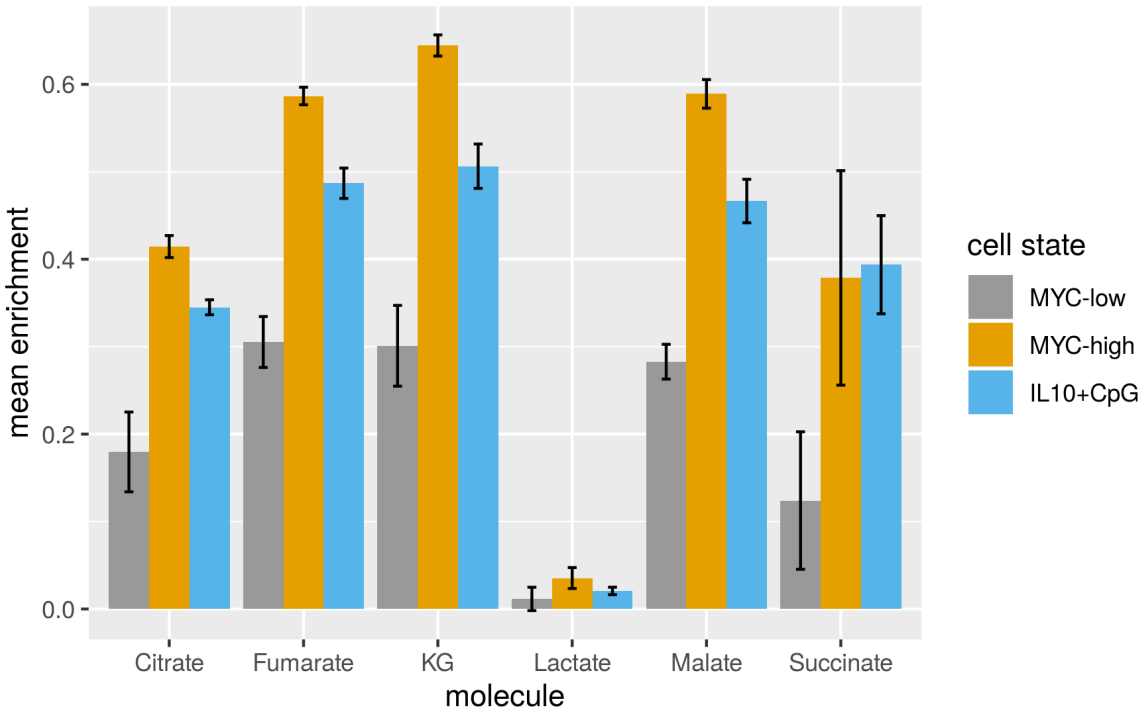



Figure III.3: ***$^{13}C$ tracer analysis with the tracer substrate U-$^{13}C$-glutamine in P493-6 cells, mean $^{13}C$ enrichment of TCA cycle intermediates and lactate.*** *The figure shows the mean $^{13}C$ enrichment of TCA cycle intermediates (KG: 2-ketoglutarate) and lactate from cell extracts for the different cell states MYC-low, MYC-high and IL10+CpG stimulation. Cells were harvested after 24 h of incubation with the tracer substrate. Data were corrected for NA and tracer purity using IsoCorrectoR, assuming a tracer purity of 99%. Samples were measured in (biological) triplicates, means of mean enrichment among replicates +/- SD are shown.* ***(These data are also shown in Figure 2a of Feist et al. (2018), but the figure was recreated for this thesis.)***

| Molecule | p-value | q-value | significant | Statistical test |
|---|---|---|---|---|
| Ala | 7.493E-03 | 1.392E-02 | yes | anova welch |
| Asp | 1.373E-04 | 6.251E-04 | yes | anova welch |
| Glu | 2.732E-02 | 4.416E-02 | yes | kruskal |
| Gly | 5.137E-02 | 6.679E-02 | no | anova welch |
| Orn | 2.319E-01 | 2.319E-01 | no | anova welch |
| Pro | 3.370E-05 | 4.382E-04 | yes | anova welch |
| Ser | 9.915E-02 | 1.172E-01 | no | kruskal |
| KG | 1.478E-03 | 3.844E-03 | yes | anova welch |
| Citrate | 3.104E-03 | 6.726E-03 | yes | anova welch |
| Fumarate | 6.120E-04 | 1.989E-03 | yes | anova welch |
| Lactate | 2.306E-01 | 2.319E-01 | no | anova welch |
| Malate | 1.443E-04 | 6.251E-04 | yes | anova welch |
| Succinate | 3.057E-02 | 4.416E-02 | yes | anova welch |

Table III.1: ***ANOVA tests to compare mean $^{13}C$ enrichment between the MYC-low, IL10+CpG and MYC-high cellular states in P493-6 cells.*** *The p-value is the p-value that results directly from the ANOVA test, the q-value corresponds to the FDR when correcting for multiple testing according to the Benjamini-Hochberg approach. The test was considered to be significant when $q \leq 0.05$. Depending on whether or not ANOVA model residuals were normally distributed, either the Welch-type ANOVA (anova welch) or the non-parametric Kruskal-Wallis ANOVA (kruskal) were used. KG: 2-ketoglutarate.*

| Molecule | Group a | Group b | p-value | q-value | significant | Normality |
|---|---|---|---|---|---|---|
| Ala | MYC-low | IL10+CpG | 3.398E-02 | 3.669E-02 | yes | yes |
| Ala | MYC-low | MYC-high | 1.383E-02 | 1.698E-02 | yes | yes |
| Ala | IL10+CpG | MYC-high | 1.549E-02 | 1.818E-02 | yes | yes |
| Asp | MYC-low | IL10+CpG | 7.452E-04 | 1.975E-03 | yes | yes |
| Asp | MYC-low | MYC-high | 8.045E-04 | 1.975E-03 | yes | yes |
| Asp | IL10+CpG | MYC-high | 4.560E-04 | 1.759E-03 | yes | yes |
| Glu | MYC-low | IL10+CpG | 7.486E-05 | 5.053E-04 | yes | yes |
| Glu | MYC-low | MYC-high | 7.817E-06 | 2.111E-04 | yes | yes |
| Glu | IL10+CpG | MYC-high | 4.386E-04 | 1.759E-03 | yes | yes |
| Pro | MYC-low | IL10+CpG | 7.189E-04 | 1.975E-03 | yes | yes |
| Pro | MYC-low | MYC-high | 2.305E-04 | 1.245E-03 | yes | yes |
| Pro | IL10+CpG | MYC-high | 2.758E-05 | 3.723E-04 | yes | yes |
| KG | MYC-low | IL10+CpG | 5.943E-03 | 8.445E-03 | yes | no |
| KG | MYC-low | MYC-high | 3.860E-03 | 6.065E-03 | yes | no |
| KG | IL10+CpG | MYC-high | 4.043E-03 | 6.065E-03 | yes | yes |
| Citrate | MYC-low | IL10+CpG | 2.137E-02 | 2.404E-02 | yes | no |
| Citrate | MYC-low | MYC-high | 8.506E-03 | 1.148E-02 | yes | yes |
| Citrate | IL10+CpG | MYC-high | 2.288E-03 | 4.412E-03 | yes | no |
| Fumarate | MYC-low | IL10+CpG | 1.869E-03 | 3.883E-03 | yes | yes |
| Fumarate | MYC-low | MYC-high | 1.523E-03 | 3.426E-03 | yes | yes |
| Fumarate | IL10+CpG | MYC-high | 2.556E-03 | 4.601E-03 | yes | yes |
| Malate | MYC-low | IL10+CpG | 7.162E-04 | 1.975E-03 | yes | yes |
| Malate | MYC-low | MYC-high | 4.359E-05 | 3.923E-04 | yes | yes |
| Malate | IL10+CpG | MYC-high | 3.514E-03 | 5.929E-03 | yes | yes |
| Succinate | MYC-low | IL10+CpG | 1.081E-02 | 1.390E-02 | yes | yes |
| Succinate | MYC-low | MYC-high | 4.790E-02 | 4.974E-02 | yes | yes |
| Succinate | IL10+CpG | MYC-high | 8.595E-01 | 8.595E-01 | no | yes |

Table III.2: ***Post-hoc pair-wise Welch t-tests to compare mean $^{13}C$ enrichment between the MYC-low, IL10+CpG and MYC-high cellular states in P493-6 cells after a significant ANOVA test.*** *Group a and Group b correspond to the groups compared in the significance test. The p-value is the p-value that results directly from the test, the q-value corresponds to the FDR when correcting for multiple testing according to the Benjamini-Hochberg approach. The test was considered to be significant when $q \leq 0.05$. "Normality" indicates whether the normality null hypothesis could be accepted for the compared groups when applying the Shapiro-Wilk test.*

### 3.2.2 $^{13}$C labeling patterns

While they are clearly very important when fitting models for metabolic flux analysis, the detailed labeling patterns of metabolites can provide relevant information in addition to the analysis of mean enrichment also in stable isotope tracer analysis approaches (Jang et al. (2018)). For the comparison of P493-6 cellular states, these patterns are shown in Figures III.4 to III.7 for amino acids and TCA cycle intermediates/lactate in cell extracts (Figures A40 to A43 in the appendix show the amino acid labeling patterns in cell supernatants and the medium control). Metabolites were measured in MS/MS mode, however, for the analysis of labeling patterns, transitions with the same number of label in the precursor ion were summed up after MS/MS NA and tracer purity correction. Using a low-resolution mass spectrometer, MS/MS measurements were not performed to obtain more detailed information on labeling patterns, but rather to increase specificity and sensitivity. As a consequence, transitions were chosen based on intensity, and not based on potential positional labeling information they may provide. An interesting observation when considering citrate, malate, fumarate and aspartate labeling is that the fraction of the m+4 isotopologue increases from MYC-low over IL10+CpG to MYC-high. These differences were significant in FDR-controlled ANOVA and post-hoc pairwise Welch t-tests with a significance level of 0.05. As in the case of some previously discussed mean enrichment comparisons, the normality hypothesis was rejected in the Shapiro-Wilk test also for MYC-low aspartate m+4, because the values of 2 out of 3 measurements were very close to each other. The detailed results of the performed significance tests are shown in Tables III.3 and III.4 for ANOVA and post-hoc tests, respectively.

Species of citrate, malate, fumarate or aspartate that contain 4 $^{13}$C result when glutamine-tracer derived fully labeled 2-ketoglutarate is subjected to oxidative decarboxylation via the TCA 2-ketoglutarate dehydrogenase complex. Thus, a higher fraction of m+4 indicates that citrate, malate, fumarate and aspartate are synthesized to a larger fraction through oxidation of glutamine derived 2-ketoglutarate, rather than through oxidation of 2-ketoglutarate that is (at least partially) glucose or fatty acid derived. This also indicates that oxidative phosphorylation may be more strongly glutamine dependent in IL10+CpG and MYC-high cells, because a larger proportion of respiratory chain NADH and $FADH_2$ is obtained from glutamine derived 2-ketoglutarate. Species with less than 4 (down to 0) $^{13}$C can result from continuous cycling of labeled species in the TCA cycle, losing both acetyl-CoA derived unla-

| Measurement | p-value | q-value | significant | Statistical test |
|---|---|---|---|---|
| Citrate m+4 | 5.826E-03 | 9.709E-03 | yes | anova welch |
| Malate m+4 | 2.118E-04 | 1.059E-03 | yes | anova welch |
| Fumarate m+4 | 1.153E-03 | 2.883E-03 | yes | anova welch |
| Asp m+4 | 2.732E-02 | 2.732E-02 | yes | kruskal |
| Citrate m+5 | 2.063E-02 | 2.578E-02 | yes | anova welch |

Table III.3: ***ANOVA tests to compare $^{13}$C labeling patterns between the MYC-low, IL10+CpG and MYC-high cellular states in P493-6 cells.*** *The p-value is the p-value that results directly from the ANOVA test, the q-value corresponds to the FDR when correcting for multiple testing according to the Benjamini-Hochberg approach. The test was considered to be significant when $q \leq 0.05$. Depending on whether or not ANOVA model residuals were normally distributed, either the Welch-type ANOVA (anova welch) or the non-parametric Kruskal-Wallis ANOVA (kruskal) were used.*

| Measurement | Group a | Group b | p-value | q-value | significant | Normality |
|---|---|---|---|---|---|---|
| Citrate m+4 | MYC-low | IL10+CpG | 1.078E-02 | 1.348E-02 | yes | yes |
| Citrate m+4 | MYC-low | MYC-high | 1.302E-03 | 2.789E-03 | yes | yes |
| Citrate m+4 | IL10+CpG | MYC-high | 1.750E-02 | 2.019E-02 | yes | yes |
| Malate m+4 | MYC-low | IL10+CpG | 1.079E-03 | 2.698E-03 | yes | yes |
| Malate m+4 | MYC-low | MYC-high | 7.205E-05 | 5.403E-04 | yes | yes |
| Malate m+4 | IL10+CpG | MYC-high | 1.631E-03 | 3.058E-03 | yes | yes |
| Fumarate m+4 | MYC-low | IL10+CpG | 3.906E-03 | 6.381E-03 | yes | yes |
| Fumarate m+4 | MYC-low | MYC-high | 3.273E-04 | 1.636E-03 | yes | yes |
| Fumarate m+4 | IL10+CpG | MYC-high | 4.254E-03 | 6.381E-03 | yes | yes |
| Asp m+4 | MYC-low | IL10+CpG | 4.372E-04 | 1.640E-03 | yes | no |
| Asp m+4 | MYC-low | MYC-high | 4.496E-05 | 5.403E-04 | yes | no |
| Asp m+4 | IL10+CpG | MYC-high | 6.447E-04 | 1.934E-03 | yes | yes |
| Citrate m+5 | MYC-low | IL10+CpG | 1.060E-02 | 1.348E-02 | yes | yes |
| Citrate m+5 | MYC-low | MYC-high | 3.475E-02 | 3.723E-02 | yes | yes |
| Citrate m+5 | IL10+CpG | MYC-high | 5.158E-01 | 5.158E-01 | no | yes |

Table III.4: ***Post-hoc pair-wise Welch t-tests to compare $^{13}C$ labeling patterns between the MYC-low, IL10+CpG and MYC-high cellular states in P493-6 cells after a significant ANOVA test.*** *Group a and Group b correspond to the groups compared in the significance test. The p-value is the p-value that results directly from the test, the q-value corresponds to the FDR when correcting for multiple testing according to the Benjamini-Hochberg approach. The test was considered to be significant when $q \leq 0.05$. "Normality" indicates whether the normality null hypothesis could be accepted for the compared groups when applying the Shapiro-Wilk test.*

beled carbon and glutamine derived labeled carbon through oxidative decarboxylation. If a 4 $^{13}$C species passes through the TCA once, it will lose at least 1 $^{13}$C in the isocitrate dehydrogenase reaction, so the m+4 species can only result from direct oxidation of glutamine-tracer derived 2-ketoglutarate. Other sources that provide unlabeled carbon to the TCA cycle are glucose based TCA anaplerosis via pyruvate carboxylase derived oxaloacetate, or oxaloacetate obtained from unlabeled amino acids contained in the cell culture medium.

Another labeling pattern of interest is the presence of m+5 citrate. This species cannot be derived from the oxidative TCA in the quantity observed, as this would require substantial labeling of acetyl-CoA, whereas lactate is hardly labeled. Thus, it is presumably produced via reductive carboxylation of glutamine-tracer derived 2-ketoglutarate. This is further supported by a tracing experiment with 1-$^{13}$C-glutamine in P493-6 cells performed by Xueni Sun in the Feist et al. (2018) study. If 1-$^{13}$C-2-ketoglutarate resulting from 1-$^{13}$C-glutamine is processed in the TCA cycle, the label is lost during the 2-ketoglutarate dehydrogenase reaction. Thus, citrate that is produced via oxidative processing of 2-ketoglutarate in the TCA cycle cannot contain label. The label only remains if citrate is directly synthesized from 2-ketoglutarate via reductive carboxylation. Supplementary figure 5b from Feist et al. (2018) shows that recognizable amounts of m+1 citrate can be found in all cellular states (MYC-low, MYC-high, IL10+CpG) in the 1-$^{13}$C-glutamine tracing experiment, which is a strong indicator of reductive carboxylation. Considering results from a previous study (Yoo et al. (2008)), reductive carboxylation likely occurs via the action of a NADPH/NADP$^+$-dependent isocitrate dehydrogenase (isoforms IDH1 and IDH2, Cairns & Mak (2013)), and most probably via the mitochondrial isoform (IDH2). The large fraction of m+4 species produced by 2-ketoglutarate oxidation in the P493-6 data indicates that the TCA cycle operates strongly in the oxidative direction, which must be driven by a low mitochondrial NADH to NAD$^+$ ratio. As a consequence, the NADH/NAD$^+$-dependent isocitrate dehydrogenase IDH3 will perform pronounced (iso-)citrate oxidation instead of 2-ketoglutarate reduction. This claim is also supported by the substantial fraction of m+3 2-ketoglutarate observed in all cellular states. Thus, reductive carboxylation must be carried out by other protagonists, namely the NADPH/NADP$^+$-dependent isoforms. The fraction of the m+5 citrate isotopologue and thus the proportion of reductive carboxylation is significantly higher in the IL10+CpG and MYC-high states than in the MYC-low state (FDR-controlled ANOVA and post-hoc Welch t-tests). This way, glutamine carbons may be directly funneled into fatty acid synthesis as acetyl-CoA via ATP-citrate lyase, without a prior oxidation step in the TCA cycle.

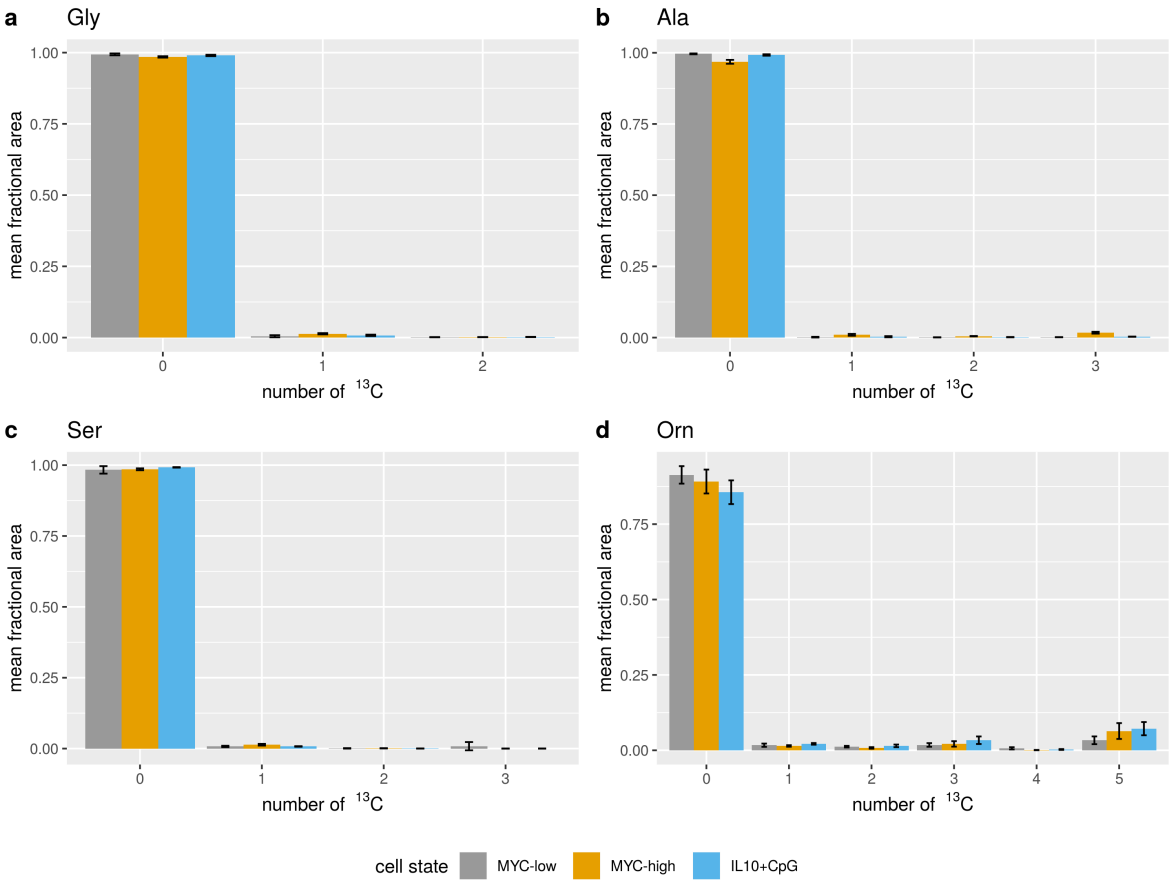



Figure III.4: ***$^{13}C$ tracer analysis with the tracer substrate U-$^{13}C$-glutamine in P493-6 cells, labeling of amino acids.*** *The figure shows the MIDs of PCF-derivatized amino acids from cell extracts for the different cell states MYC-low, MYC-high and IL10+CpG stimulation. Cells were harvested after 24 h of incubation with the tracer substrate. Data were corrected for NA and tracer purity using IsoCorrectoR, assuming a tracer purity of 99%. The x-axis labels correspond to the number of $^{13}C$ label incorporated in the respective isotopologue. Samples were measured in (biological) triplicates, means of isotopologue fractions +/- SD are shown.*

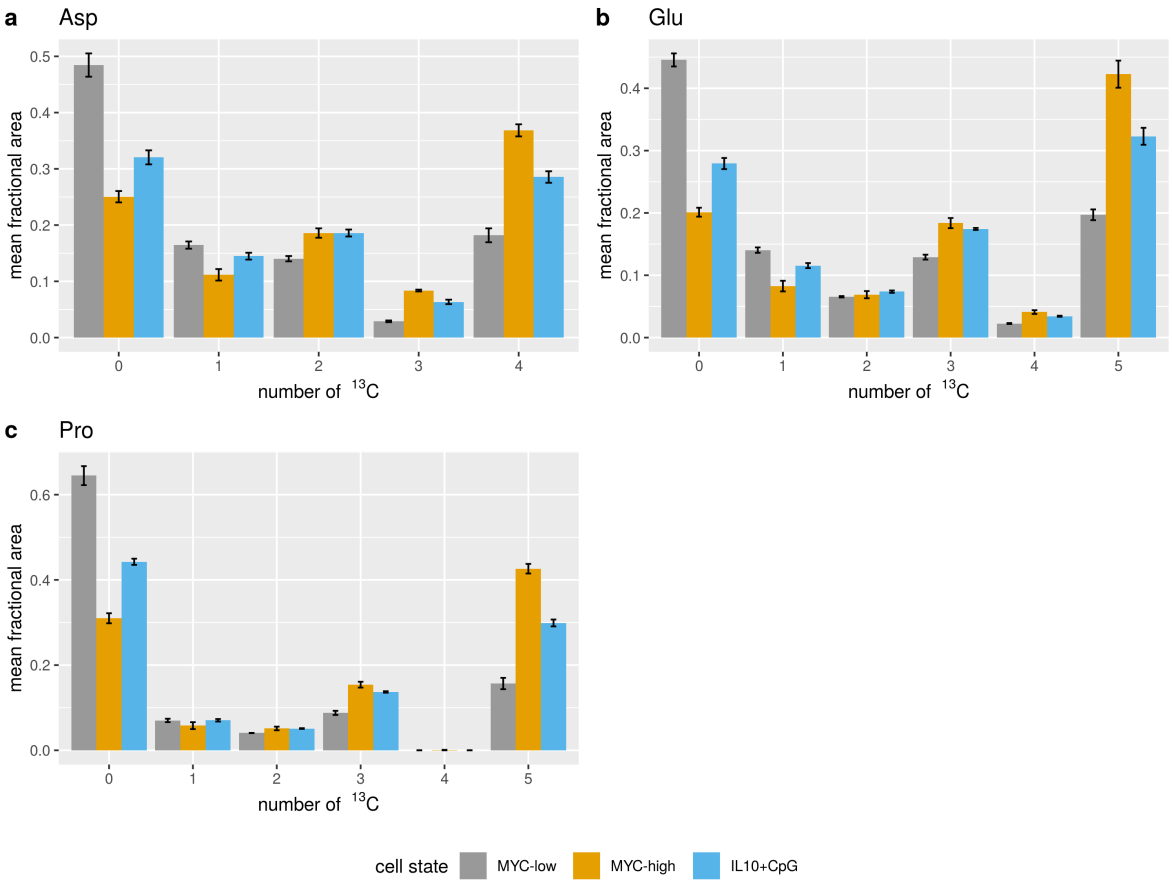



Figure III.5: ***$^{13}$C tracer analysis with the tracer substrate U-$^{13}$C-glutamine in P493-6 cells, labeling of amino acids.*** *The figure shows the MIDs of PCF-derivatized amino acids from cell extracts for the different cell states MYC-low, MYC-high and IL10+CpG stimulation. Cells were harvested after 24 h of incubation with the tracer substrate. Data were corrected for NA and tracer purity using IsoCorrectoR, assuming a tracer purity of 99%. The x-axis labels correspond to the number of $^{13}$C label incorporated in the respective isotopologue. Samples were measured in (biological) triplicates, means of isotopologue fractions +/- SD are shown.*

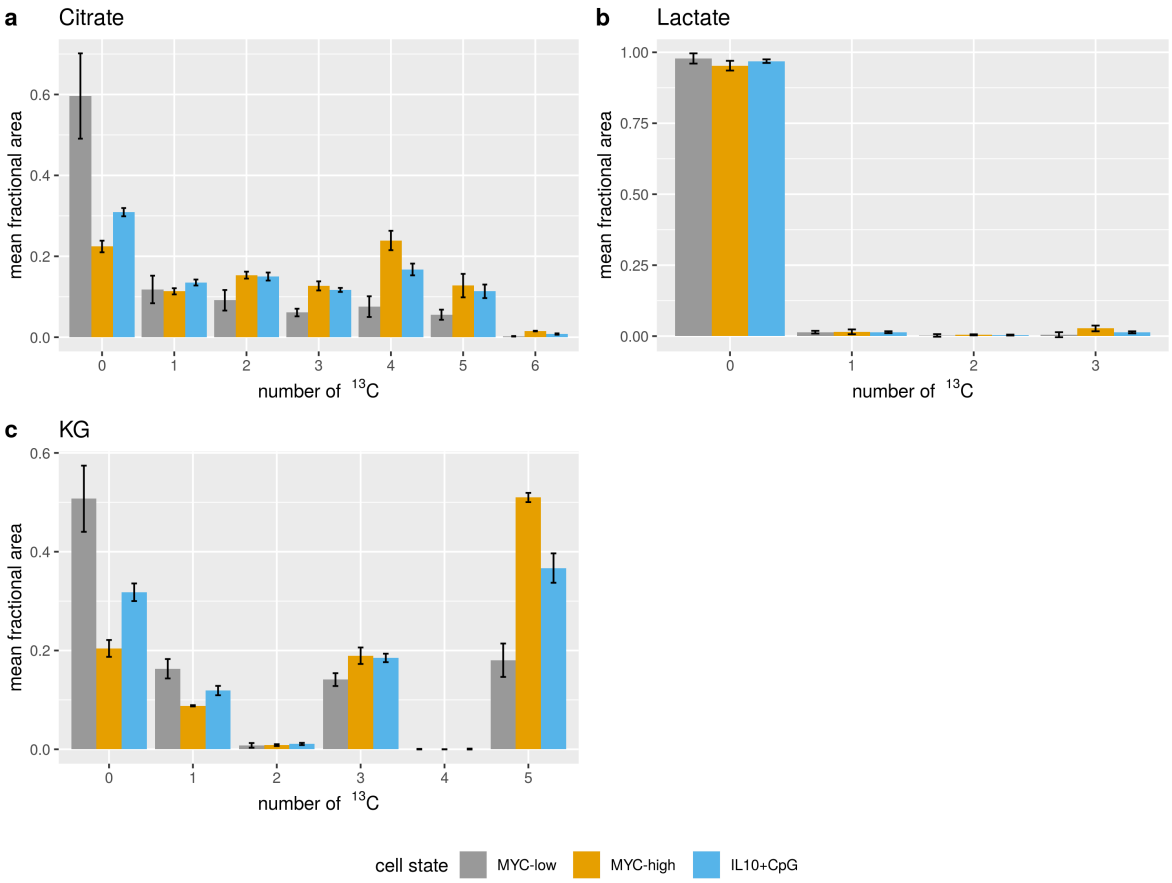



Figure III.6: ***$^{13}C$ tracer analysis with the tracer substrate U-$^{13}C$-glutamine in P493-6 cells, labeling of 2-ketoglutarate (KG), citrate and lactate.*** *The figure shows the MIDs of TCA cycle intermediates/lactate from cell extracts for the different cell states MYC-low, MYC-high and IL10+CpG stimulation. Cells were harvested after 24 h of incubation with the tracer substrate. Data were corrected for NA and tracer purity using IsoCorrectoR, assuming a tracer purity of 99%. The x-axis labels correspond to the number of $^{13}C$ label incorporated in the respective isotopologue. Samples were measured in (biological) triplicates, means of isotopologue fractions +/- SD are shown.* ***(Data on citrate are also shown in Figure 2c of Feist et al. (2018), but the figure was recreated for this thesis.)***

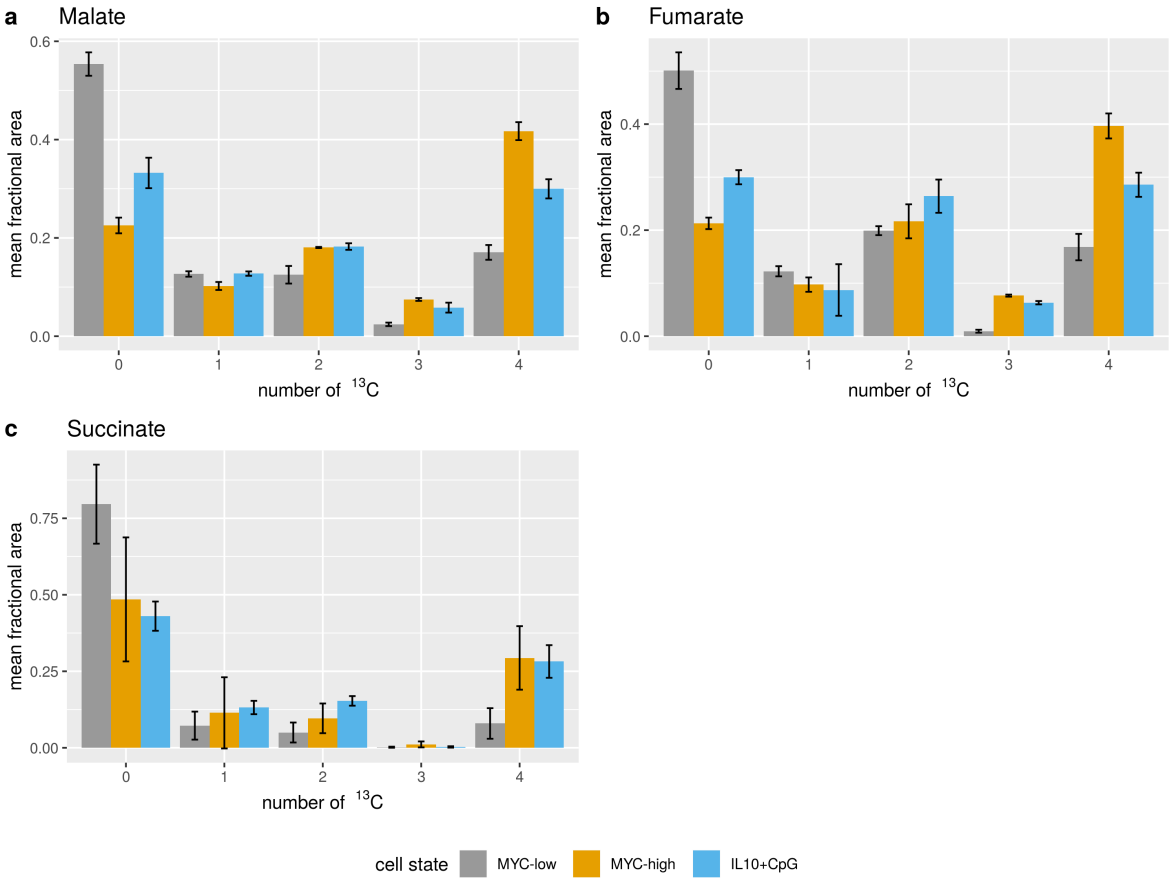



Figure III.7: ***$^{13}$C tracer analysis with the tracer substrate U-$^{13}$C-glutamine in P493-6 cells, labeling of malate, fumarate and succinate.*** *The figure shows the MIDs of TCA cycle intermediates from cell extracts for the different cell states MYC-low, MYC-high and IL10+CpG stimulation. Cells were harvested after 24 h of incubation with the tracer substrate. Data were corrected for NA and tracer purity using IsoCorrectoR, assuming a tracer purity of 99%. The x-axis labels correspond to the number of $^{13}$C label incorporated in the respective isotopologue. Samples were measured in (biological) triplicates, means of isotopologue fractions +/- SD are shown.*

## 3.3 $^{13}C$ tracer analysis results in a broader biological context

Changes in metabolism and nutrient usage are not isolated, but occur in conjunction with other physiological changes and usually serve a certain biological purpose. To put the results from $^{13}C$ tracer analysis in a broader biological context, the data will be integrated with the results from other analyses performed by co-authors within the scope of the Feist et al. (2018) study. Both glucose and glutamine consumption were found to be significantly increased in IL10+CpG and MYC-high cells compared to MYC-low cells. Further, removal of glutamine from the cell culture medium drastically decreased proliferation in MYC-high and IL10+CpG cells, while the effect of glucose removal was not as pronounced. In addition, intracellular amounts of TCA cycle intermediates and amino acids were substantially increased in IL10+CpG stimulated cells compared to MYC-low cells, together with an increased expression of glycolysis and glutaminolysis genes (see Figure A25 in the appendix). When combining all this information with the $^{13}C$ tracer analysis result that glutamine carbon contribution to the biosynthesis of TCA cycle intermediates and glutamate derived amino acids increases substantially from MYC-low over IL10+CpG to MYC-high, it seems likely that marked amounts of glutamine carbon are diverted into these metabolite pools in Il10+CpG and MYC-high cells.

However, what is the purpose of the observed increased glutaminolysis? To address this question, metabolite rescue experiments have been performed with cells cultured in glutamine-free medium. For IL10+CpG stimulated cells, proliferation could be restored almost to the level of cells cultured with glutamine using either oxaloacetate, aspartate, adenine or thymine. Interestingly, for MYC-high cells, especially the addition of a membrane permeable form of 2-ketoglutarate restored proliferation, while the metabolites rescuing IL10+CpG cells were relatively ineffective. This indicates that glutamine is used for different main purposes in IL10+CpG and MYC-high cells. The possibility to rescue with aspartate or nucleobases in IL10+CpG cells suggests that a main use of glutamine in this cellular state is the synthesis of nucleotides and nucleic acids to supply proliferation. Aspartate, which has been found to be highly enriched with glutamine $^{13}C$, is a main carbon donor in pyrimidine synthesis, and a nitrogen donor in both pyrimidine and purine synthesis. The observation that oxaloacetate is also effective at rescuing IL10+CpG cells suggests that the carbon supplied via glutamine is probably more important than the nitrogen in that setting. However, since purine bases like adenine only contain carbons from glycine, tetrahydrofolate (providing a one-carbon unit originally taken from serine) and $HCO_3^-$, and none directly taken from oxaloacetate or aspartate, it is reasonable to ask why the addition of adenine can rescue proliferation if mainly glutamine carbon is relevant. Probably, low levels of aspartate due to glutamine withdrawal are also limiting for purine base synthesis, because at least one of the nitrogens in all purine bases is derived from a molecule of aspartate (two in the case of adenine). Given enough nitrogen is available in the system in general, the addition of oxaloacetate may then increase aspartate levels to supply purine synthesis with nitrogen. Further, the addition of oxaloacetate or aspartate may also increase glutamine levels. Glutamine is especially important in purine base synthesis as a nitrogen donor since it directly provides between two and three (side chain amide derived) nitrogens depending on the base. Thus, providing oxaloacetate and aspartate may rescue proliferation by rising aspartate and glutamine levels so that they are sufficient to supply both pyrimidine and purine base synthesis, while the addition of adenine may help proliferation by allowing low existing pools to be used mainly for pyrimidine synthesis. To assess also the usage of glutamine derived nitrogen, a $^{15}N$ tracer analysis with α-$^{15}N$-glutamine was performed by Xueni Sun. Especially glutamate, aspartate, proline and alanine were found

to be $^{15}$N enriched (up to about 50% mean enrichment for glutamate and aspartate in MYC-high), while the enrichment again increased from MYC-low over IL10+CpG to MYC-high, similar to the $^{13}$C labeling experiments. Some $^{15}$N enrichment was also found in nucleobases, however to a lower than expected degree, probably because ISS had not been reached for these metabolites. It must also be considered that nucleobase nitrogen is derived to a large degree from the side chain amide of glutamine, and only to a lesser extent from its α-amino group (via aspartate).

As MYC-high cells can mainly be rescued via the addition of 2-ketoglutarate, it is possible that this cellular state is more strongly dependent on oxidative catabolism of glutamine derived 2-ketoglutarate to supply oxidative phosphorylation. This is supported by the observation that the fractions of the 4 $^{13}$C species of citrate, malate, fumarate and aspartate - derived from the oxidation of glutamine derived 2-ketoglutarate - are significantly increased in MYC-high cells compared to IL10+CpG cells and especially MYC-low cells. Additionally, upon glutamine withdrawal, the OCR (oxygen consumption rate) was decreased significantly down to the level of MYC-low cells in MYC-high cells, but remained unchanged in IL10+CpG stimulated cells. Thus, albeit cells in the IL10+CpG state also use glutamine to a large degree for oxidative phosphorylation if it is supplied, they do not seem to be as dependent on it for respiration as MYC-high cells.

To assess the relevance of key enzymes for the metabolism of the different P493-6 cellular states, inhibitor and RNAi-knockdown studies were performed by Maren Feist. Both the addition of the glutaminase inhibitor CB-839 as well as the general aminotransferase inhibitor aminooxyacetate (AOA) led to a strong decrease in proliferation in IL10+CpG stimulated and MYC-high cells, and intracellular aspartate levels dropped substantially. Also in this setting, the addition of aspartate, adenine or thymine could restore proliferation in IL10+CpG stimulated cells, underscoring the importance of aminotransferases in funneling glutamine (via aspartate) into nucleotide synthesis. To assess the role of the aspartate aminotransferases GOT1 (cytosolic) and GOT2 (mitochondrial), RNAi knowckdown experiments specifically targeting the mRNA of either of these enzymes were performed. Both enzymes catalyze the conversion of glutamate and oxaloacetate to 2-ketoglutarate and aspartate, and vice versa. The GOT2 knockdown led to a marked decrease in proliferation in both IL10+CpG and MYC-high cells, whereas the GOT1 knockdown only significantly affected MYC-high cells. Proliferation of IL10+CpG cells could be fully restored by addition of adenine or thymine, and partially by addition of aspartate. In a previous study (Feist et al. (2017)) it had been shown that combined IL10 and CpG signaling induces proliferation in MYC-low cells via the synergistic action of the transcription factors STAT3 and NFκB. In the current study (Feist et al. (2018)), it has additionally been found that combined STAT3 and NFκB signaling increases the expression of *GOT2*, and that STAT3 and NFκB jointly bind to the *GOT2* promoter.

In conclusion, these results suggest that glutaminolysis is very important for both the MYC-high cellular state of P493-6 cells, mimicking Burkitt lymphomas, as well as for IL10+CpG stimulated MYC-low cells, which can serve as a model for B-cell lymphomas with aberrant STAT3 and NFκB signaling. The analyses indicate that MYC-high and IL10+CpG stimulated cells use glutamine in a different manner, though. TCA cycle intermediates and glutamate derived amino acids are strongly enriched with glutamine derived carbon in both cellular states. However, while MYC-high cells appear to be very dependent on glutamine derived 2-ketoglutarate for oxidative phosphorylation, IL10+CpG cells appear to employ glutamine especially for the synthesis of nucleotides via aspartate. Further, the aspartate aminotrans-

ferase GOT2 appears to have an important role in diverting glutamine to nucleotide synthesis, and its expression is upregulated by the joint action of STAT3 and NFκB. These findings may aid in developing therapies targeting the metabolism of specific subtypes of B-cell lymphomas. However, care must be taken when attempting to translate results from cell culture experiments to medicine, as the environment of cancer model cell lines in cell culture can vary drastically from the actual tumor environment in a tissue. This includes *e.g.* the availability of oxygen, the supply with various nutrients and, importantly, the interaction with other cells in the tissue (Heiden & DeBerardinis (2017)). However, in the current study, high aberrant *GOT2* expression has actually been shown to be significantly associated with shorter overall survival in DLBCL patients that have received an R-CHOP (Rituximab, Cyclophosphamide, Hydroxydaunomycin, Oncovin, Prednisone) treatment. This is a first indication that it may be possible to translate the results to actual medical applications.

# Chapter IV

# Materials and methods

## 1 Materials

| **Material** | **Manufacturer** | **Usage** |
|---|---|---|
| Pure water, PURELAB Plus system | ELGA LabWater (Celle, Germany) | |
| 10 kDa Amicon ultra centrifugal filters | Merck (Darmstadt, Germany) | Sample preparation |
| Methanol for HPLC Ultra LC-MS grade, HiPerSolv Chromanorm | VWR International (Darmstadt, Germany) | Metabolite extraction, LC-MS, PCF derivatization |
| 1-Propanol for LC-MS, LiChrosolv | Merck (Darmstadt, Germany) | PCF derivatization |
| 3-Picoline (3-Methylpyridine) $\geq$ 99.5%, LiChrosolv | Merck (Darmstadt, Germany) | PCF derivatization |
| Isooctane, AnalaR Normapur | VWR International (Darmstadt, Germany) | PCF derivatization |
| Propylchloroformate (PCF) 98% | VWR International (Darmstadt, Germany) | PCF derivatization |
| Ethylacetate for LC-MS, HiPerSolv Chromanorm | VWR International (Darmstadt, Germany) | PCF derivatization |
| Chloroform, $\geq$ 99.9% | Sigma-Aldrich (Taufkirchen, Germany) | PCF derivatization |

Table IV.1: ***Laboratory materials and usage.***

| **Material** | **Manufacturer** | **Usage** |
|---|---|---|
| Heptafluorobutyric acid (HFBA) suitable for Ion Chromatography, $\geq$ 99.5% | Sigma-Aldrich (Taufkirchen, Germany) | LC-MS |
| Ammonium formate (AF) $\geq$ 99.995% trace metal basis | Sigma-Aldrich (Taufkirchen, Germany) | LC-MS |
| Acetonitrile Hypergrade for LC-MS | Supelco (Belafonte, PA, USA) | LC-MS |
| Formic acid 98-100% for LC-MS, LiChropur | Merck (Darmstadt, Germany) | LC-MS |
| N-Methyl-N-(trimethylsilyl)-trifluoracetamide (MSTFA) | Macherey-Nagel (Dueren, Germany) | methoximation - trimethylsilylation |
| Methoxyamine hydrochloride | Sigma-Aldrich (Taufkirchen, Germany) | methoximation - trimethylsilylation |
| Pyridine, 99,8% | Sigma-Aldrich (Taufkirchen, Germany) | methoximation - trimethylsilylation |
| Undecanoic Acid, analytical standard | Sigma-Aldrich (Taufkirchen, Germany) | methoximation - trimethylsilylation |
| Isooctane, $\geq$ 99.9% | VWR International (Darmstadt, Germany) | methoximation - trimethylsilylation |
| L-alanine, Fluka | Sigma-Aldrich (Taufkirchen, Germany) | Alanine validation mixes |
| L-alanine-1-$^{13}$C, 99% $^{13}$C purity | Cortecnet (Voisins-Le-Bretonneux, France) | Alanine validation mixes |
| L-alanine-3-$^{13}$C, 99% $^{13}$C purity | Cortecnet (Voisins-Le-Bretonneux, France) | Alanine validation mixes |
| L-alanine-1,2-$^{13}C_2$, 99.2% $^{13}$C purity | Cortecnet (Voisins-Le-Bretonneux, France) | Alanine validation mixes |
| L-alanine-$^{13}C_3$, 97.6% $^{13}$C purity | Cortecnet (Voisins-Le-Bretonneux, France) | Alanine validation mixes |

Table IV.2: ***Laboratory materials and usage.***

# 2 Experimental methods

## 2.1 Stable isotope tracing in a P493-6 B-cell lymphoma model cell line

A stable isotope labeling experiment was conducted in a P493-6 B-cell lymphoma model cell line. The cell line has an inducible *MYC* allele, which allows it to mimick both normal B-cells and MYC-driven lymphomas. Cells were fed with U-$^{13}$C-glutamine, and the incorporation of $^{13}$C into different amino acids and TCA cycle intermediates was assessed via LC-MS/MS of cell extracts and supernatants.

### 2.1.1 Cell culture

Cell culture work was performed by Maren Feist at the Clinic of Haematology and Medical Oncology at the University Medical Centre Göttingen, as described in Feist et al. (2018). Briefly, P493-6 cells were cultured in RPMI 1640 medium with 10% FCS and 1:100 penicillin/streptomycin at 37 °C with 5% $CO_2$. Normally, P493-6 cells show *MYC* overexpression. This state is termed MYC-high. To obtain the MYC-low state of P493-6 cells (low *MYC* expression), *MYC* expression was repressed via treatment with 1 ng/ml doxycycline for 24 h. For stable isotope labeling analysis, cells were then washed in warm phosphate-buffered saline (PBS), counted, centrifuged and seeded in fresh Gln-free cell culture medium at a density of 1x10$^6$ cells/ml, with or without doxycycline. U-$^{13}$C-glutamine was added to a concentration of 2 mM as the tracer substrate (tracer purity: 99%), and IL10, CpG, or IL10 + CpG were added for cell stimulation in the corresponding samples. After culturing for an additional 24 h, cells were counted and harvested. 4x10$^6$ cells were centrifuged (300 x g, 5 min, 4 °C), and the supernatant was collected. The pellet was washed two times with cold PBS and then resuspended in 1 ml of ice-cold 80% MeOH to stop enzymatic activity. To assess isotopic steady state, cells were cultured up to 30 h, and 2x10$^6$ cells were harvested at 18 h, 22 h, 24 h, 26 h and 30 h just as described before.

### 2.1.2 Sample preparation for metabolomic analysis

Supernatants were filtered by centrifugation (30 min, 4000 x g, 4 °C) in 10 kDa Amicon ultra centrifugal filters (Merck, Darmstadt, Germany). Polar metabolites in the cell pellets were extracted using 80% MeOH. First, pellets were vortexed in 80% MeOH and then centrifuged (10,000 x g, 6 min, 4 °C). The resulting supernatant was transferred to a glass tube, and the cell pellets were washed two additional times with 80% MeOH, combining all supernatants. The combined extracts were then dried in a vacuum evaporator for 2 h, followed by resuspension of the residue in 100 $\mu$l of water.

For the analysis of amino acids, a derivatization with propylchloroformate (PCF) was carried out to improve retention and thus separability in reversed-phase HPLC. Derivatization was performed as described in Goot et al. (2012). Briefly, 10 $\mu$l of sample (cell extract or supernatant) are diluted with water to a final volume of 200 $\mu$l. Then, 80 $\mu$l derivatizing reagent 1 (77% n-propanol, 23% 3-picoline (vol/vol)) is added, followed by 50 $\mu$l derivatizing reagent 2 (17.4% propylchloroformate, 11% isooctane, 71.6% chloroform (vol/vol)). The mixture is vortexed, and the derivatives are extracted by the addition of 250 $\mu$l ethylacetate. Following vigorous mixing, 200 $\mu$l upper organic phase are taken. The extract is then dried under a

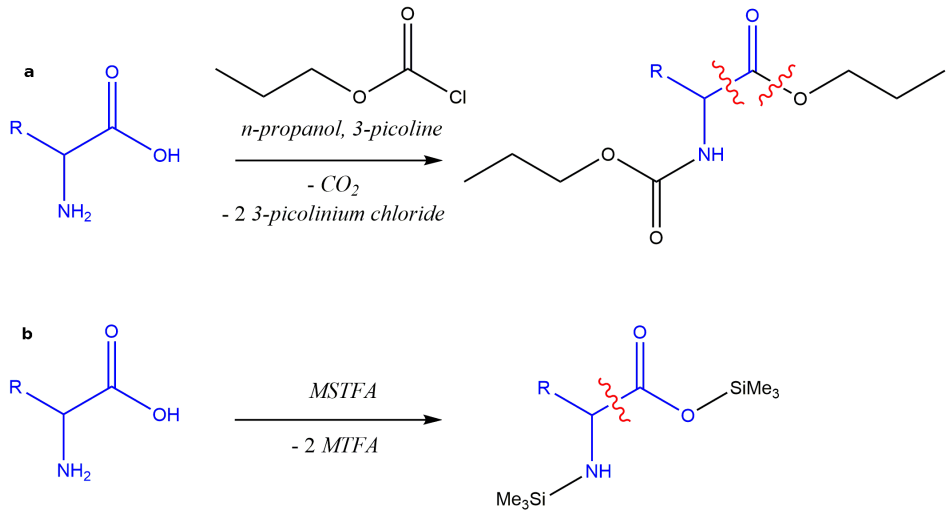



Figure IV.1: ***PCF- and TMS-derivatization of amino acids.*** *PCF- (a) and TMS (b) derivatization are shown for a generic amino acid. Carboxyl- and amino groups in the amino acid residue are derivatized in the same way. For TMS-derivatization, N-methyl-N-(trimethylsilyl)trifluoracetamid (MSTFA) was used, which is converted to N-methyltrifluoracetamid (MTFA) during the reaction. Fragmentation sites in tandem MS are shown as dashed red lines. The fragmentation of PCF-derivatized amino acids can yield, among others, propylformate or propanol as the neutral loss. A study on the mechanism of alkyl chloroformate-mediated esterification in aqueous media has recently shown that the reaction at the carboxyl group proceeds via the formation of N-acylpyridinium and alkyl carbonate anion intermediates when employing pyridine. These intermediates result from the reaction between pyridine and the carboxylic anhydride initially formed from the carboxylic acid and alkyl chloroformate. Said intermediates are then further converted through decarboxylation of the alkyl carbonate anion species and a conversion of N-acylpyridinium to an ester via nucleophilic attack of an alcohol species (either from the decarboxylated alkyl carbonate species or from solution) or to the original carboxylic acid via hydrolysis (Opekar et al. (2021)). In the case of the particular PCF-derivatization reaction performed, a N-acylpicolinium intermediate is likley formed analogously to the N-acylpyridinium intermediate observed in the study.*

nitrogen stream and the residue is redissolved in 100 $\mu$l water/methanol (38%/62%, vol/vol). The derivatization reaction is illustrated in Figure IV.1 a for a generic amino acid.

### 2.1.3 LC-MS/MS

For the HPLC-MS/MS analysis of PCF-derivatized amino acids from $^{13}$C labeling, an Agilent 1200 HPLC (Agilent, Waldbronn, Germany) coupled to a low-resolution API 4000 QTRAP mass spectrometer (AB SCIEX, Darmstadt, Germany) operating in positive ionization mode was used. The mass spectrometer is a triple-quadrupole device that also has linear ion trap functionality, but it was used only in triple-quadrupole mode. An EZ:faast AAA-MS (250 x 3 mm i.d., 4 $\mu$m, Phenomenex) reversed-phase column was used for HPLC and kept at 30 °C. Mobile phase A was 10 mM ammonium formate and 0.1 % (vol/vol) heptafluorobutyric acid in water, mobile phase B was 10 mM ammonium formate and 0.1 % (vol/vol) heptafluorobutyric acid in methanol. The column flow rate was 350 $\mu$l/min, with an injection volume of 5 $\mu$l for

cell pellet extracts and 10 $\mu$l for cell supernatants. The HPLC gradient was set up as follows: Starting with 62 % B, B was increased to 79 % within 12 min, with a final increase to 98 % B within 0.01 min. This condition was then kept for 3 min. In the following, the column was reconditioned with 62 % B for 8 min. The mass spectrometer was operated in MRM mode, and the transitions for each amino acid were set up so that the mass shifts in precursor and product ion match the mass shifts of the isotopologues expected from $^{13}$C labeling. As illustrated in Figure IV.1 a, the acquired transitions of PCF-derivatized amino acids cover either fragmentation at the C1-C2 bond of the amino acid (neutral loss: propylformate, $C_4H_8O_2$) or fragmentation at the C1-O bond linking the amino acid to the propanol-residue introduced via derivatization (neutral loss: propanol, $C_3H_8O$). In amino acids that carry a carboxylic acid group in their residue, fragmentation may also occur there. For the amino acids alanine, aspartate and glutamate, the neutral loss is propylformate and contains one carbon of the amino acid. For glycine, ornithine, proline and serine, the neutral loss propanol comes purely from derivatization and cannot contain label. Thus, it is valid to correct the $^{13}$C labeling data from these amino acids using an MS instead of an MS/MS correction algorithm, as pointed out in section 3.2.3. Peak integration in extracted ion chromatograms was performed using the Analyst Software (version 1.6.2, AB Sciex). The described method for the analysis of amino acids from a $^{13}$C labeling experiment is based on the method for absolute amino acid quantification in unlabeled samples published in Goot et al. (2012).

The analysis of (underivatized) organic acids/TCA cycle intermediates from $^{13}$C labeling was performed with the same LC-MS/MS instrument as the analysis of amino acids (triple-quadrupole mode). However, the mass spectrometer was operated in negative ionization mode. For chromatography, a Discovery HS F5-3 HPLC column (15 cm x 2.1 mm, 3 $\mu$m, Supelco) equipped with a Security Guard column (C18, Phenomenex) was employed, with mobile phase A being 0.1 % formic acid in water (v/v) and mobile phase B being 100 % acetonitrile. The column was kept at 30 °C. The HPLC gradient started with 0 % B for 5 min at a flow rate of 300 $\mu$l/min. This was followed by a linear increase of B to 100 % within 2 min. 100 % B was kept for 4 min, followed by a switch to 0 % B within 0.1 min. The flow rate was then increased to 300 $\mu$l/min and 0 % B was held for 18 min for equilibration. The injection volume was 10 $\mu$l. As with amino acid analysis, MRM transitions were set up to cover all expected $^{13}$C labeling states of a given metabolite. Table IV.3 shows precursor and product ion $m/z$ as well as product ion and neutral loss composition of the measured organic acids (m+0 isotopologues). In the original MRM method for analyzing $^{13}$C organic acid labeling, transitions for some labeling states were missing or measured in duplicate (labeled fumarate and succinate have overlapping nominal precursor and product ion $m/z$ for some transitions). Thus, missing transitions were added and duplicates were removed (duplicated measurements result in less dwell time per transition). The remaining transition of a duplicate pair was renamed such that it is clear that the transition corresponds to labeling states of different metabolites. The affected transitions are listed in Table IV.4.

Prior to analysis, data were corrected for NA and tracer purity with IsoCorrectoR, assuming a purity of 99% for the tracer substrate U-$^{13}$C-glutamine. As data were acquired in MS/MS mode with a low-resolution mass spectrometer, IsoCorrectoR's low-resolution MS/MS algorithm was used to correct the data. To this end, product ion and and neutral loss chemical formulas of all metabolites were supplied to IsoCorrectoR. For species where the neutral loss cannot contain label, using the MS (instead of MS/MS) correction algorithm would have been appropriate, too.

| **Molecule** | **Precursor m/z** | **Product ion m/z** | **Product ion composition** | **Neutral loss composition** |
|---|---|---|---|---|
| 2-Ketoglutarate | 145 | 101 | $C_4H_5O_3^-$ | $CO_2$ |
| Citrate | 191 | 111 | $C_5H_3O_3^-$ | $CH_4O_4$ |
| Fumarate | 115 | 71 | $C_3H_3O_2^-$ | $CO_2$ |
| Lactate | 89 | 43 | $C_2H_3O^-$ | $CH_2O_2$ |
| Malate | 133 | 115 | $C_4H_3O_4^-$ | $H_2O$ |
| Succinate | 117 | 73 | $C_3H_5O_2^-$ | $CO_2$ |

Table IV.3: ***Transitions of organic acids/TCA cycle intermediates.*** *Transitions of organic acids/TCA cycle intermediates in the LC-MS/MS method employed for analyzing $^{13}C$ organic acid labeling, only m+0 isotopologue shown.*

| **Transition ID** | **Precursor m/z** | **Product ion m/z** | **Issue** |
|---|---|---|---|
| 2-Ketoglutarate 1.0 | 146 | 101 | missing |
| 2-Ketoglutarate 1.1 | 146 | 102 | missing |
| 2-Ketoglutarate 3.2 | 148 | 103 | missing |
| 2-Ketoglutarate 3.3 | 148 | 104 | missing |
| Fumarate 1.1 | 116 | 72 | missing |
| Fumarate 2.2 | 117 | 73 | missing |
| Fumarate 3.2 | 118 | 73 | duplicate of Succinate 1.0 |
| Fumarate 3.3 | 118 | 74 | duplicate of Succinate 1.1 |
| Fumarate 4.3 | 119 | 74 | duplicate of Succinate 2.1 |

Table IV.4: ***Missing and duplicate transitions in the original MRM method for analyzing $^{13}C$ labeled organic acids/TCA cycle intermediates.*** *Transitions of organic acids/TCA cycle intermediates in the LC-MS/MS method employed for analyzing $^{13}C$ organic acid labeling that were missing or measured in duplicate. The numbering x.y in the transition IDs corresponds to x label in the precursor ion and y label in the product ion.*

| **Analyte [$\mu$M]** | **Mixture 1** | **Mixture 2** | **Mixture 3** | **Mixture 4** |
|---|---|---|---|---|
| L-Alanine | 50.000 | 20.000 | 10.000 | 100.000 |
| L-Alanine-1-$^{13}$C | 12.500 | 20.000 | 20.000 | 1.000 |
| L-Alanine-3-$^{13}$C | 12.500 | 20.000 | 40.000 | 1.000 |
| L-Alanine-2,3-$^{13}C_2$ | 3.125 | 20.000 | 10.000 | 1.000 |
| L-Alanine-$^{13}C_3$ | 0.781 | 20.000 | 30.000 | 1.000 |

Table IV.5: ***Mixtures of alanine isotopologues.*** *Mixtures of alanine isotopologues (and isotopomers) used for the experimental validation of isotope correction. Concentrations in $\mu M$ are given.* ***(Table taken from Heinrich et al. (2018))***

### 2.2 Measurement of alanine isotopologue mixtures of known composition via GC-APCI-TOFMS

Single stock solutions (500 mM) of L-alanine (Fluka /Sigma-Aldrich, Taufkirchen, Germany), L-alanine-1-$^{13}$C (99% $^{13}$C purity, Cortecnet, Voisins-Le-Bretonneux, France), L-alanine-3-$^{13}$C (99% $^{13}$C purity, Cortecnet), L-alanine-1,2-$^{13}C_2$ (99.2% $^{13}$C purity, Cortecnet), and L-alanine-$^{13}C_3$ (97.6% $^{13}$C purity, Cortecnet) were prepared in pure water (PURELAB Plus system, ELGA LabWater, Celle, Germany). Single stocks were further diluted and four different mixtures containing the different isotopomers in varying concentrations were prepared (see Table IV.5). A 10 $\mu$l aliquot of each mixture was evaporated to complete dryness using a vacuum evaporator (CombiDancer, Hettich AG, Bäch, Switzerland), subjected to silylation and analyzed by GC-APCI-TOFMS employing the derivatization protocol and instrumental setup previously described in Wachsmuth et al. (2015). Specifically, a 450-GC (Bruker Daltonics GmbH, Bremen, Germany) coupled to a microTOF orthogonal acceleration TOF mass spectrometer (Bruker Daltonics) via an atmospheric pressure chemical ionization source (APCI II) was used. One $\mu$l of the derivatized sample was injected using splitless mode. The temperature program started at 50 °C (1 min), was ramped at 5 °C/min to 120 °C and then at 8 °C/min to 30 °C (5 min). Each sample was prepared and analyzed in triplicates. Area integrals of the $[M+H]^+$ ions of the different isotopologues were determined using Bruker Quant Analysis 2.2 (Bruker Daltonik GmbH, Bremen, Germany). Note that L-alanine-1-$^{13}$C and L-alanine-3-$^{13}$C cannot be distinguished and both contribute to the m+1 signal. GC-APCI-TOFMS mass spectra of the silylated alanine standards used for preparing the mixtures are shown in Figures A44 to A46 in the appendix. **(Paragraph taken from Heinrich et al. (2018), slightly modified)**

## 3 Datasets

### 3.1 Data from stable isotope tracing in a P493-6 B-cell line

For analyzing the effect of isotope correction and for validating IsoCorrectoR, several MS datasets from stable isotope labeling experiments were used. Two of the datasets come from the $^{13}$C stable isotope labeling experiment conducted in a P493-6 B-cell line that is described

in section 2.1. The data were acquired at the Institute of Functional Genomics, University of Regensburg, with a low-resolution QTRAP device in MS/MS mode. One dataset covers amino acids (PCF-derivatized, measured in positive mode), while the other covers organic acids/TCA cycle intermediates (underivatized, measured in negative mode). While all metabolites were measured in tandem MS, the neutral loss does not contain label in some cases (glycine, proline, ornithine, serine). Because of that, it is possible to correct these data using an MS instead of an MS/MS correction algorithm.

### 3.2 AccuCor datasets

Two additional datasets were provided as example data together with the AccuCor tool for NA correction (Su et al. (2017)). Data have been acquired via LC-MS, with HILIC LC coupled to a Q Exactive PLUS hybrid quadrupole-orbitrap mass spectrometer (Thermo Scientific) in negative ionization mode. The MS resolution was 140000 at 200 $m/z$ (high resolution). Both a $^{13}C$ and a $^{15}N$ labeling dataset have been provided, all metabolites were measured in underivatized form.

### 3.3 Simulated stable isotope labeling data

In cases where experimentally acquired data were not sufficient for the analyses (*e.g.*, ultra-high-resolution data from multiple-tracer experiments or specific examples to illustrate the impact of correction), stable isotope labeling data were simulated. This was done either manually, by devising uncorrected MIDs suitable for the given analysis, or by randomly generating uncorrected MIDs using R. Random simulation was chosen in cases where the exact setup of the MIDs was not important for the analysis. When randomly simulating MIDs, random numbers within a certain interval between a minimum and a maximum number were chosen for the different expected measurements of a given MID. The minimum was usually chosen so that negative corrected values should not or only rarely occur, except negative corrected values were explicitly intended.

## 4 Computational methods

### 4.1 Correction of stable isotope labeling data

#### 4.1.1 Correction using tools

Correction of stable isotope labeling data for NA and tracer purity was performed using IsoCorrectoR, IsoCor, ICT or PyNAC, or manually.

When correcting with IsoCorrectoR, either IsoCorrectoR version 0.1.12 together with R version 3.3.3 was used (analyses that have been taken from Heinrich et al. (2018)), or IsoCorrectoR version res_dep_v0.9 with R version 3.6.2. This is the prototype capable of resolution-dependent MS and MS/MS correction. When using IsoCorrectoR, a threshold can be set to limit internal calculations to probabilities above this threshold when computing the probability matrix. The default value for this calculation threshold is 1E-08. This default value was used for all

analyses in this thesis, except for comparisons of resolution-dependent MS and MS/MS correction to manual calculations, for creating the example probability matrices in the algorithm section and for the analysis on the impact of the calculation threshold parameter. In these cases, the threshold was set to 0 (*i.e.*, no threshold). For UHR correction with IsoCorrectoR, a different type of threshold is used, as described in the algorithm section. Here, the default value of 8 was used for all analyses, except for the analysis explicitly assessing the effect of the calculation threshold parameter.

IsoCor v1 (Millard et al. (2012), version 1.0, Python version 2.6.5) was used for analyses from Heinrich et al. (2018), while IsoCor v2 (Millard et al. (2019), version 2.1.3, Python version 3.7.3) was used for other analyses. Correction with ICT (Jungreuthmayer et al. (2016)) was performed using ICT version 0.04 with Strawberry Perl version 5.24.1.1 (Windows) or Perl version 5.22.1 (Linux). PyNAC (Carreer et al. (2013)) was used together with Python version 2.6.5. Depending on the analysis, slightly different values for natural isotope abundance have been used. When comparing against IsoCor v2, for analyses of resolution-dependent correction and for comparing UHR correction to resolution-dependent correction, the isotope abundance values (and isotope masses) also used by IsoCor v2 were employed. When comparing against PyNAC, isotope abundance values used by PyNAC were used. For all other analyses, IsoCorrectoR's default NA values were employed. Table A7 in the appendix shows the corresponding NA and mass values of the isotopes. In low-resolution correction, only nominal mass shifts and no exact absolute isotope masses are used. Additional settings associated with a given correction procedure are specified in the corresponding section in the results.

### 4.1.2 Manual correction

Manual correction of data was performed by first calculating the probability matrix $P$ either in Microsoft Excel (Office 365) or LibreOfficeCalc (version 5.1.6.2). While manual computation of $P$ in a spreadsheet is laborious and error-prone, it is still manageable for small core molecules like glycine. If it does not add additional elements to the molecule, derivatization does not substantially increase the complexity of calculations. To derive corrected values $v_c$ according to $v_m = P \cdot v_c$, $P$ and uncorrected values $v_m$ were supplied to either the *linsolve* function of Matlab (version R2016a, analyses taken from Heinrich et al. (2018)) or the *lm* function of R (version 3.6.2) to solve the resulting system of linear equations.

### 4.1.3 Computing uncorrected values from known corrected values

To compute uncorrected values $v_m$ from known corrected values $v_c$, IsoCorrectoR functions were used to calculate the corresponding probability matrix $P$. Then, $v_m$ could be derived through matrix multiplication via $v_m = P \cdot v_c$.

## 4.2 Statistics

All statistics were performed using base functions of R, version 3.6.2.

### 4.2.1 Testing for normally distributed data

To test whether data can be assumed to be normally distributed or whether the normality assumption has to be rejected, the Shapiro-Wilk test was used. In the case of $p > 0.05$, normality was assumed, in the case of $p \leq 0.05$, the null hypothesis of normal distribution was rejected and the data were assumed to follow a different distribution. It should be noted, however, that the Shapiro-Wilk test has low power to reject the null hypothesis if sample sizes are small, which usually is the case with the labeling data analyzed in this thesis.

### 4.2.2 Testing for differences between two groups

If not stated otherwise, to test for statistically significant differences between two (independent) groups of samples, Welch's t-test for unequal variances was used. In the case of data that cannot be assumed to be normally distributed, a non-parametric test like the Mann-Whitney-U test should usually be applied. However, all significance tests performed in this thesis were performed on groups with a sample size of 3. With a sample size of 3, the minimal p-value that can be obtained with the Mann-Whitney-U test is 0.1, the reason being the mathematical definition of the test. As a consequence, at a significance level of 0.05, such a test could never reject the null hypothesis. Thus, also in the case of data where normal distribution was in doubt, Welch's t-test was performed, albeit with a notification that the normality assumption may be violated.

If equal group variances cannot be safely assumed, instead of generally using Welch's t-test, it would also be possible to first test for equality of variances (e.g. with Levene's test), and then use either Student's t-test (which requires equal variances) or Welch's t-test depending on the outcome. However, the power of Levene's test to reject the null hypothesis of equal variances is low for small sample sizes, resulting in a substantial proportion of Type II errors, where equal variance would be assumed when variances are in fact not equal. If Student's t-test is performed in these cases, its Type I error rate will increase, producing more false positive results than expected. When generally using Welch's t-test if equal variances are in doubt, however, this increase of Type I error can be avoided. And if variances are in fact equal, the power of Welch's t-test is almost the same as that of Student's t-test (Delacre et al. (2017)).

### 4.2.3 Correction for multiple testing

Correction for multiple testing was performed via the FDR approach of Benjamini and Hochberg (Benjamini & Hochberg (1995)). If not stated otherwise, tests with a q-value (FDR) $\leq 0.05$ were considered to be statistically significant.

### 4.2.4 Testing for differences between more than two groups and post-hoc pairwise tests

The presence of statistically significant differences between more than 2 group means was tested for with a one-way ANOVA (analysis of variance) approach. In the case of normally distributed ANOVA model residuals, a Welch-type ANOVA (variances were assumed to be unequal) was used, if not stated otherwise. The non-parametric Kruskal-Wallis ANOVA was

used in the case that the normality assumption had to be rejected (significant Shapiro-Wilk test for the ANOVA model residuals). If multiple ANOVAs were performed in a given analysis, the obtained p-values were corrected for multiple testing as described in the previous paragraph. Results were considered to be statistically significant if their associated q-value (FDR) was $\leq 0.05$. Subsequently, for significant ANOVA results, post-hoc Welch's t-tests for pairwise comparisons were applied to test for significant differences between pairs of groups. The p-values obtained from the post-hoc pairwise tests were corrected for multiple testing as described above. If not stated otherwise, differences with a q-value (FDR) $\leq 0.05$ were considered to be statistically significant.

## 4.3 System

Analyses were run on a machine with a Intel Core i7-4510 CPU (4x 2.00 GHz) and 16 GB RAM. As operating system, either Windows 10 or Ubuntu 16.04 LTS was used.

## 4.4 Additional software

R (version 3.3.3 or 3.6.2) together with ggplot2 (3.2.0) and cowplot (1.0.0) were used for analyzing and plotting data. Microsoft Excel (Office 365) or LibreOfficeCalc (version 5.1.6.2) were used for generating csv and xls input files for IsoCorrectoR and for setting up files for data analysis in R. ChemDraw (version 16.0.1.4 or 18.1) was used for drawing chemical structures.

# Chapter V

# Conclusion and outlook

In metabolomics, stable isotope labeling experiments with tracer isotopes like $^{13}C$ or $^{15}N$ can provide a wealth of information on nutrient contributions to metabolite synthesis, contributions to certain pathways or the dynamics of metabolism. To avoid misinterpretations of data or incorrect flux estimates, it is however imperative to correct MS data from stable isotope labeling experiments for natural isotope abundance (NA), and, possibly, for impurities of the tracer substrate. This is necessary, since the mass shift that is produced by the incorporation of tracer isotopes into the metabolites of interest also arises naturally, because stable heavy isotopes like $^{13}C$, $^{15}N$ or $^{2}H$ are naturally present in all molecules, the quantity defined by their NA. Additionally, impurities of the tracer substrate can lead to a "loss" of label and mass shift, so that species that would actually carry a higher amount of label contribute to the measurements of species with less amounts of label.

IsoCorrectoR, which is presented with this thesis, is an R-based tool for the correction of stable isotope labeling data for NA and tracer purity. While there are already several other tools available for this purpose, like Python-based IsoCor (v2) and PyNAC or Perl-based ICT, IsoCorrectoR provides a comprehensive set of correction features in a single solution, was designed for ease-of-use, and, importantly, adds new modes of correction. IsoCorrectoR features low-resolution and resolution-dependent MS and MS/MS correction, ultra-high resolution (UHR) correction of UHR data (also from stable isotope resolved metabolomics (SIRM) experiments employing multiple tracers like $^{13}C$ and $^{15}N$ simultaneously), tracer purity correction and compatibility with any tracer isotope (*e.g.*, $^{13}C$, $^{15}N$ or $^{2}H$).

A mode of correction that was novel when IsoCorrectoR was first released with Heinrich et al. (2018) is the UHR correction of data from stable isotope labeling (SIRM) experiments not only for NA, but also for tracer purity. The only other tool designed for the UHR correction of data from SIRM experiments, where multiple tracer isotopes are used simultaneously, is PyNAC, and it lacks the option to correct for tracer purity. Very recently, AccuCor2 (Wang et al. (2021)) has been released, which can perform resolution-dependent correction on data from SIRM experiments and is also capable of tracer purity correction. As shown in this thesis (and before in Heinrich et al. (2018)), performing tracer purity correction can be crucial, especially in the case of multiple tracers.

The other novel mode of correction that IsoCorrectoR can offer is resolution-dependent MS/MS correction. It has been stated in the literature and shown with quantitative simulations in this thesis that MS/MS stable isotope labeling data (where labeling can occur in both fragments)

should not be corrected with a correction procedure intended for MS data. In these cases, it is necessary to apply a correction algorithm specifically tailored for MS/MS data. While the tools ICT and MS-X-Corr offer MS/MS correction for low-resolution MS/MS data (MS-X-Corr being limited in its overall functions, though), only IsoCorrectoR can correct also high-resolution MS/MS data in a resolution-dependent manner. The resolution-dependent correction of high-resolution data takes into account that, for devices in the resolution range of orbitraps (and partly also FT-ICRs), some NA contributions can be resolved, while others can not, depending on the instrument resolution available for the measurement of the given molecular ion or fragment. Clearly, this has to be accounted for in the correction process. As shown in this thesis, if a casual low-resolution algorithm is applied to such data, this can lead to marked overcorrection due to the removal of NA contributions that were already resolved spectrometrically. To the best of our knowledge, the only tool (aside of IsoCorrectoR) with a correct algorithmic approach to the resolution-dependent correction of high-resolution data is IsoCor v2 - there are also other tools available for this purpose (AccuCor, ElemCor), but their approach is slightly flawed. IsoCor v2 is only capable of correcting MS data in a resolution-dependent manner, however, while it lacks the functionality of resolution-dependent MS/MS correction. Thus, IsoCorrectoR is the only tool that provides such a feature. With increased adoption of orbitrap mass spectrometers in metabolomics and the benefits that positional labeling information from tandem MS measurements can provide especially for flux modeling, being able to correct high resolution MS/MS data from stable isotope labeling experiments in a resolution-dependent manner should be a relevant addition to the existing set of correction approaches.

However, an important aspect that must be considered in this regard is the drop in resolution in orbitrap and FT-ICR devices that accompanies tandem MS measurements. When not operating the devices in direct infusion mode, the transient duration originally available for full scan MS has to be split by the number of precursors that should be measured. Analyses comparing low-resolution MS/MS correction and resolution-dependent MS/MS correction at lower resolutions in the range of 5000 - 80000 (at 200 $m/z$) however showed that there can still be marked differences down to a resolution of 20000 (and possibly lower, in specific cases) for orbitrap devices. Thus, the use of resolution-dependent correction can also be advisable in this medium resolution range, especially for low $m/z$ species. Also, future technological advancements may lead to a further increase of resolution capabilities in orbitrap devices. Possibly, the relevance of resolution-dependent correction for orbitraps operating in a medium resolution range also has implications for the correction of TOF data from stable isotope labeling experiments, as TOF devices operate roughly in the resolution range of 10000 - 50000. It has to be considered, however, that TOF resolution does not increase with lower $m/z$, which likely makes resolution-dependent correction less relevant in this case. Not only the comparison of low-resolution correction to resolution-dependent correction, but also the comparison of resolution-dependent correction to UHR correction yielded important insights. As analyses in this thesis have shown, the assumption of UHR correction - that stable isotope labeling data acquired at ultra-high resolution (400000 at 400 $m/z$, or higher), only has to be corrected for NA contributions of the tracer isotope itself, and no other isotopes - does not hold in any case. Especially for high $m/z$ metabolites, and in $^{15}$N labeling experiments, the error introduced can be substantial, up to a point that meaningful biological interpretation of data should no longer be possible. Thus, extreme caution is advised when correcting UHR data, and, if possible, resolution-dependent correction should be employed instead of UHR correction.

In addition to analyses of its various modes of correction on example data, IsoCorrectoR has

also been thoroughly validated against other available tools, manual calculations, and validation mixtures of known composition, all of which yielded very good agreement. More pronounced (but still relatively small) differences where found in the comparison of low-resolution MS/MS correction between IsoCorrectoR and ICT. These differences could be attributed to the different strategies of the two programs to deal with measurement uncertainty that would result in negative corrected values. Effectively, however, both approaches can be considered viable. The analysis employing validation mixtures of known isotopologue composition showed that relative errors in measurement can be magnified through the correction process, as absolute errors in the uncorrected values translate more or less directly to the corrected values.

Albeit IsoCorrectoR already offers a comprehensive array of correction modes, there is of course still room for improvement in the domain of isotope correction for stable isotope labeling experiments. State-of-the-art resolution-dependent correction works with a hard $\delta_{min}$ cut-off to determine whether isotopic contributions have to be corrected for or not. Given the shape of mass spectrometric peaks, it is questionable whether this approach is optimal. Often, peaks may overlap only partially and only to a very low degree, while 100% of the NA contribution is removed by the correction algorithm if the $m/z$ difference is below $\delta_{min}$. This can result in overcorrection in the case of peaks where the absolute difference in $m/z$ is roughly in the range of $FWHM$. If the difference is much higher, the peaks should clearly be resolved, if it is much lower (close to 0), the peaks will approximately overlap to 100%. A way to overcome this issue may be to perform a partial correction for a given NA contribution, depending on an assumed peak shape (*e.g.*, gaussian) and the $m/z$ difference between peaks in relation to $FWHM$ (*i.e*, a measure of peak spread). However, doing this for all relevant NA contributions may be computationally expensive, and computationally efficient solutions for determining partial NA contributions would be needed. Another useful improvement in the area of isotope correction clearly is the possibility to correct multiple-tracer data in a resolution-dependent manner. As analyses in this thesis have shown, UHR correction can lead to results that are markedly wrong, and until very recently, only UHR correction was available for data from SIRM experiments. However, as addressed in section 3.2.6, AccuCor2 has now been developed, and it is capable of resolution-dependent correction of multiple tracer data (Wang et al. (2021)). Clearly, of course, a mandatory prerequisite also for applying resolution-dependent correction to multiple-tracer data is that at least the tracer isotopes (*e.g.*, $^{13}$C and $^{15}$N) can be reliably resolved. And even with AccuCor2 being able to correct multiple-tracer data in a resolution-dependent manner, a comprehensive solution to isotope correction that combines all commonly required correction approaches in a unified algorithmic theory and implementation is still lacking. As the authors of AccuCor2 state themselves, it is *e.g.* not advised to employ AccuCor2 for the correction of single-tracer data, and AccuCor2 is further limited to the correction of only specific combinations of tracer isotopes, lacking for instance the combinations $^{13}$C and $^{18}$O or $^{2}$H and $^{15}$N. Also, MS/MS correction is not covered. Thus, when considering the current landscape of isotope correction tools, it resembles a patchwork of tools with various algorithmic approaches and purposes, and IsoCorrectoR may still be the most comprehensive solution to date.

As an application of isotope correction, a $^{13}$C stable isotope labeling experiment in P493-6 B-cells was presented in this thesis. These cells have an inducible *MYC* allele. If *MYC* is induced (MYC-high state), the cells mimick Burkitt lymphomas, if *MYC* is repressed (MYC-low state), the cells resemble normal B-cells. In the MYC-high state, cells proliferate quickly, while there is hardly any proliferation in the MYC-low state. However, as shown in a previous study (Feist et al. (2017)), if the microenvironmental factors IL10 and CpG are added to MYC-low

P493-6 cells, they proliferate almost with the same rate as MYC-high cells. As the combined addition of IL10 and CpG activates joint STAT3/NFκB signaling, the IL10+CpG stimulated state of MYC-low P493-6 cells may serve as a model for B-cell lymphomas with aberrations in that part of cellular signaling. As malign transformations are usually accompanied and supported by metabolic alterations, a $^{13}$C tracer analysis in P493-6 cells with U-$^{13}$C-glutamine as the tracer substrate was performed to check for differences in the usage of the major nutrient glutamine between the different cellular states MYC-low, MYC-high and IL10+CpG stimulated MYC-low. Key results were that glutamine usage for the synthesis of glutamate-derived amino acids and TCA cycle intermediates increases substantially and significantly from MYC-low over the IL10+CpG state to MYC-high. Additionally, the usage of glutamine for oxidative phosphorylation increases in the same order, as inferred from the fractional abundance of m+4 TCA cycle intermediates and m+4 aspartate. Within the scope of the Feist et al. (2018) study, the $^{13}$C tracer analysis was performed in conjunction with other analyses, including nutrient deprivation and rescue, oxygen consumption rate and inhibitor experiments. Taken together, the results indicate that glutaminolysis plays a very important role in both MYC-low IL10+CpG stimulated cells and MYC-high cells, but they use glutamine in a different manner. MYC-high cells seem to be very dependent on the usage of glutamine for oxidative phosphorylation, via 2-ketoglutarate. In IL10+CpG stimulated cells, on the other hand, glutamine appears to have an especially dominant role in nulceotide synthesis via aspartate, and the aspartate aminotransferase GOT2 seems to be crucial in that regard. Because of the general difficulties when attempting to transfer results from cell culture to medical insights, it remains to be seen whether the IL10+CpG stimulated cellular state of P493-6 MYC-low cells can really serve as a model for certain types of B-cell lymphomas. However, in the Feist et al. (2018) study, it was possible to significantly associate aberrant *GOT2* expression with shorter overall survival in DLBCL patients that received an R-CHOP treatment, which may be a first indication of the medical relevance of the established cell culture model.

# Chapter VI

# Bibliography


Benjamini, Y. & Hochberg, Y. (1995), 'Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing', *Journal of the Royal Statistical Society: Series B (Methodological)* **57**(1), 289–300. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.2517-6161.1995.tb02031.x.
**URL:** *https://onlinelibrary.wiley.com/doi/abs/10.1111/j.2517-6161.1995.tb02031.x*

Bingol, K. (2018), 'Recent Advances in Targeted and Untargeted Metabolomics by NMR and MS/NMR Methods', *High-Throughput* **7**(2).
**URL:** *https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6023270/*

Buescher, J. M., Antoniewicz, M. R., Boros, L. G., Burgess, S. C., Brunengraber, H., Clish, C. B., DeBerardinis, R. J., Feron, O., Frezza, C., Ghesquiere, B., Gottlieb, E., Hiller, K., Jones, R. G., Kamphorst, J. J., Kibbey, R. G., Kimmelman, A. C., Locasale, J. W., Lunt, S. Y., Maddocks, O. D. K., Malloy, C., Metallo, C. M., Meuillet, E. J., Munger, J., Nöh, K., Rabinowitz, J. D., Ralser, M., Sauer, U., Stephanopoulos, G., St-Pierre, J., Tennant, D. A., Wittmann, C., Vander Heiden, M. G., Vazquez, A., Vousden, K., Young, J. D., Zamboni, N. & Fendt, S.-M. (2015), 'A roadmap for interpreting 13C metabolite labeling patterns from cells', *Current opinion in biotechnology* **34**, 189–201.
**URL:** *https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4552607/*

Cairns, R. A. & Mak, T. W. (2013), 'Oncogenic Isocitrate Dehydrogenase Mutations: Mechanisms, Models, and Clinical Opportunities', *Cancer Discovery* **3**(7), 730–741. Publisher: American Association for Cancer Research Section: Review.
**URL:** *https://cancerdiscovery.aacrjournals.org/content/3/7/730*

Carreer, W. J., Flight, R. M. & Moseley, H. N. B. (2013), 'A Computational Framework for High-Throughput Isotopic Natural Abundance Correction of Omics-Level Ultra-High Resolution FT-MS Datasets', *Metabolites* **3**(4), 853–866.
**URL:** *https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3882318/*

Choi, J. & Antoniewicz, M. R. (2019), 'Tandem Mass Spectrometry for 13C Metabolic Flux Analysis: Methods and Algorithms Based on EMU Framework', *Frontiers in Microbiology* **10**.
**URL:** *https://www.frontiersin.org/articles/10.3389/fmicb.2019.00031/full*

Clish, C. B. (2015), 'Metabolomics: an emerging but powerful tool for precision medicine', *Cold Spring Harbor Molecular Case Studies* **1**(1).
**URL:** *https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4850886/*

DeBerardinis, R. J. & Chandel, N. S. (2016), 'Fundamentals of cancer metabolism', *Science Advances* **2**(5).
**URL:** *https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4928883/*

Delacre, M., Lakens, D. & Leys, C. (2017), 'Why Psychologists Should by Default Use Welch's $t$-test Instead of Student's $t$-test', *International Review of Social Psychology* **30**(1), 92–101. Number: 1 Publisher: Ubiquity Press.
**URL:** *http://www.rips-irsp.com/article/10.5334/irsp.82/*

Dettmer, K., Aronov, P. A. & Hammock, B. D. (2007), 'Mass spectrometry-based metabolomics', *Mass Spectrometry Reviews* **26**(1), 51–78.
**URL:** *https://onlinelibrary.wiley.com/doi/abs/10.1002/mas.20108*

Du, D., Tan, L., Wang, Y., Peng, B., Weinstein, J. N., Wondisford, F. E., Su, X. & Lorenzi, P. L. (2019), 'ElemCor: accurate data analysis and enrichment calculation for high-resolution LC-MS stable isotope labeling experiments', *BMC Bioinformatics* **20**(1), 89.
**URL:** *https://doi.org/10.1186/s12859-019-2669-9*

Eliuk, S. & Makarov, A. (2015), 'Evolution of Orbitrap Mass Spectrometry Instrumentation', *Annual Review of Analytical Chemistry (Palo Alto, Calif.)* **8**, 61–80.

Feist, M., Kemper, J., Taruttis, F., Rehberg, T., Engelmann, J. C., Gronwald, W., Hummel, M., Spang, R. & Kube, D. (2017), 'Synergy of interleukin 10 and toll-like receptor 9 signalling in B cell proliferation: Implications for lymphoma pathogenesis', *International Journal of Cancer* **140**(5), 1147–1158. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/ijc.30444.
**URL:** *https://onlinelibrary.wiley.com/doi/abs/10.1002/ijc.30444*

Feist, M., Schwarzfischer, P., Heinrich, P., Sun, X., Kemper, J., Bonin, F. v., Perez-Rubio, P., Taruttis, F., Rehberg, T., Dettmer, K., Gronwald, W., Reinders, J., Engelmann, J. C., Dudek, J., Klapper, W., Trümper, L., Spang, R., Oefner, P. J. & Kube, D. (2018), 'Cooperative STAT/NF-$\varkappa$B signaling regulates lymphoma metabolic reprogramming and aberrant GOT2 expression', *Nature Communications* **9**(1), 1–14.
**URL:** *https://www.nature.com/articles/s41467-018-03803-x*

Fernandez, C. A., Rosiers, C. D., Previs, S. F., David, F. & Brunengraber, H. (1996), 'Correction of 13C Mass Isotopomer Distributions for Natural Stable Isotope Abundance', *Journal of Mass Spectrometry* **31**(3), 255–262.

Frick, M., Dörken, B. & Lenz, G. (2011), 'The molecular biology of diffuse large B-cell lymphoma', *Therapeutic Advances in Hematology* **2**(6), 369–379. Publisher: SAGE Publications.
**URL:** *https://doi.org/10.1177/2040620711419001*

Goot, A. T. v. d., Zhu, W., Vázquez-Manrique, R. P., Seinstra, R. I., Dettmer, K., Michels, H., Farina, F., Krijnen, J., Melki, R., Buijsman, R. C., Silva, M. R., Thijssen, K. L., Kema, I. P., Neri, C., Oefner, P. J. & Nollen, E. A. A. (2012), 'Delaying aging and the aging-associated decline in protein homeostasis by inhibition of tryptophan degradation', *Proceedings of the*

*National Academy of Sciences* **109**(37), 14912–14917.
**URL:** *https://www.pnas.org/content/109/37/14912*

Hanahan, D. & Weinberg, R. A. (2011), 'Hallmarks of cancer: the next generation', *Cell* **144**(5), 646–674.

Heiden, M. G. V. & DeBerardinis, R. J. (2017), 'Understanding the intersections between metabolism and cancer biology', *Cell* **168**(4), 657–669.
**URL:** *https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5329766/*

Heinrich, P., Kohler, C., Ellmann, L., Kuerner, P., Spang, R., Oefner, P. J. & Dettmer, K. (2018), 'Correcting for natural isotope abundance and tracer impurity in MS-, MS/MS- and high-resolution-multiple-tracer-data from stable isotope labeling experiments with IsoCorrectoR', *Scientific Reports* **8**(1), 17910.
**URL:** *https://www.nature.com/articles/s41598-018-36293-4*

Hiller, K. & Metallo, C. M. (2013), 'Profiling metabolic networks to study cancer metabolism', *Current Opinion in Biotechnology* **24**(1), 60–68.
**URL:** *http://www.sciencedirect.com/science/article/pii/S0958166912001772*

Jang, C., Chen, L. & Rabinowitz, J. D. (2018), 'Metabolomics and Isotope Tracing', *Cell* **173**(4), 822–837.
**URL:** *http://www.sciencedirect.com/science/article/pii/S0092867418303878*

Jin, L., Alesi, G. & Kang, S. (2016), 'Glutaminolysis as a target for cancer therapy', *Oncogene* **35**(28), 3619–3625.
**URL:** *https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5225500/*

Jungreuthmayer, C., Neubauer, S., Mairinger, T., Zanghellini, J. & Hann, S. (2016), 'ICT: isotope correction toolbox', *Bioinformatics* **32**(1), 154–156.
**URL:** *https://academic.oup.com/bioinformatics/article/32/1/154/1742487*

Kalisz, K., Alessandrino, F., Beck, R., Smith, D., Kikano, E., Ramaiya, N. H. & Tirumani, S. H. (2019), 'An update on Burkitt lymphoma: a review of pathogenesis and multimodality imaging assessment of disease presentation, treatment response, and recurrence', *Insights into Imaging* **10**.
**URL:** *https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6529494/*

Le, A., Lane, A. N., Hamaker, M., Bose, S., Gouw, A., Barbi, J., Tsukamoto, T., Rojas, C. J., Slusher, B. S., Zhang, H., Zimmerman, L. J., Liebler, D. C., Slebos, R. J., Lorkiewicz, P. K., Higashi, R. M., Fan, T. W. M. & Dang, C. V. (2012), 'Glucose-independent glutamine metabolism via TCA cycling for proliferation and survival in B-cells', *Cell Metabolism* **15**(1), 110–121.
**URL:** *https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3345194/*

Markley, J. L., Brüschweiler, R., Edison, A. S., Eghbalnia, H. R., Powers, R., Raftery, D. & Wishart, D. S. (2017), 'The future of NMR-based metabolomics', *Current Opinion in Biotechnology* **43**, 34–40.
**URL:** *http://www.sciencedirect.com/science/article/pii/S0958166916301768*

Midani, F. S., Wynn, M. L. & Schnell, S. (2017), 'The importance of accurately correcting for the natural abundance of stable isotopes', *Analytical biochemistry* **520**, 27–43.
**URL:** *https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5343595/*

Millard, P., Delépine, B., Guionnet, M., Heuillet, M., Bellvert, F. & Létisse, F. (2019), 'IsoCor: isotope correction for high-resolution MS labeling experiments', *Bioinformatics* .

Millard, P., Letisse, F., Sokol, S. & Portais, J.-C. (2012), 'IsoCor: correcting MS data in isotope labeling experiments', *Bioinformatics (Oxford, England)* **28**(9), 1294–1296.

Moseley, H. N. (2010), 'Correcting for the effects of natural abundance in stable isotope resolved metabolomics experiments involving ultra-high resolution mass spectrometry', *BMC Bioinformatics* **11**, 139.
**URL:** *https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2848236/*

Murphy, T. A., Dang, C. V. & Young, J. D. (2013), 'Isotopically nonstationary 13C flux analysis of Myc-induced metabolic reprogramming in B-cells', *Metabolic engineering* **15**, 206–217.
**URL:** *https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3535509/*

Niedenführ, S., Pierick, A. t., Dam, P. T. N. v., Suarez-Mendez, C. A., Nöh, K. & Wahl, S. A. (2016), 'Natural isotope correction of MS/MS measurements for metabolomics and 13C fluxomics', *Biotechnology and Bioengineering* **113**(5), 1137–1147.
**URL:** *https://onlinelibrary.wiley.com/doi/abs/10.1002/bit.25859*

Opekar, S., Kvíčala, J., Moos, M., Pejchal, V. & Šimek, P. (2021), 'Mechanism of Alkyl Chloroformate-Mediated Esterification of Carboxylic Acids in Aqueous Media', *The Journal of Organic Chemistry* **86**(23), 16293–16299. Publisher: American Chemical Society.
**URL:** *https://doi.org/10.1021/acs.joc.1c01546*

Pajic, A., Spitkovsky, D., Christoph, B., Kempkes, B., Schuhmacher, M., Staege, M. S., Brielmeier, M., Ellwart, J., Kohlhuber, F., Bornkamm, G. W., Polack, A. & Eick, D. (2000), 'Cell cycle activation by c-myc in a Burkitt lymphoma model cell line', *International Journal of Cancer* **87**(6), 787–793.

Rosman, K. J. R. & Taylor, P. D. P. (1998), 'Isotopic compositions of the elements 1997 (Technical Report)', *Pure and Applied Chemistry* **70**(1), 217–235.
**URL:** *https://www.degruyter.com/document/doi/10.1351/pac199870010217/html*

Scigelova, M., Hornshaw, M., Giannakopulos, A. & Makarov, A. (2011), 'Fourier Transform Mass Spectrometry', *Molecular & Cellular Proteomics : MCP* **10**(7).
**URL:** *https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3134075/*

Shankland, K. R., Armitage, J. O. & Hancock, B. W. (2012), 'Non-Hodgkin lymphoma', *The Lancet* **380**(9844), 848–857.
**URL:** *http://www.sciencedirect.com/science/article/pii/S0140673612606059*

Su, X., Lu, W. & Rabinowitz, J. D. (2017), 'Metabolite Spectral Accuracy on Orbitraps', *Analytical chemistry* **89**(11), 5940–5948.
**URL:** *https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5748891/*

van Winden, W. A., Wittmann, C., Heinzle, E. & Heijnen, J. J. (2002), 'Correcting mass isotopomer distributions for naturally occurring isotopes', *Biotechnology and Bioengineering* **80**(4), 477–479.

Vander Heiden, M. G., Cantley, L. C. & Thompson, C. B. (2009), 'Understanding the Warburg Effect: The Metabolic Requirements of Cell Proliferation', *Science (New York, N.Y.)*

**324**(5930), 1029–1033.
**URL:** *https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2849637/*

Wachsmuth, C. J., Hahn, T. A., Oefner, P. J. & Dettmer, K. (2015), 'Enhanced metabolite profiling using a redesigned atmospheric pressure chemical ionization source for gas chromatography coupled to high-resolution time-of-flight mass spectrometry', *Analytical and Bioanalytical Chemistry* **407**(22), 6669–6680.
**URL:** *https://doi.org/10.1007/s00216-015-8824-x*

Wang, Y., Parsons, L. R. & Su, X. (2021), 'AccuCor2: isotope natural abundance correction for dual-isotope tracer experiments', *Laboratory Investigation* **101**(10), 1403–1410.
**URL:** *https://www.nature.com/articles/s41374-021-00631-4*

Yoo, H., Antoniewicz, M. R., Stephanopoulos, G. & Kelleher, J. K. (2008), 'Quantifying reductive carboxylation flux of glutamine to lipid in a brown adipocyte cell line', *The Journal of Biological Chemistry* **283**(30), 20621–20627.

Yuneva, M., Zamboni, N., Oefner, P., Sachidanandam, R. & Lazebnik, Y. (2007), 'Deficiency in glutamine but not glucose induces MYC-dependent apoptosis in human cells', *The Journal of Cell Biology* **178**(1), 93–105.
**URL:** *https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2064426/*

Zubarev, R. A. & Makarov, A. (2013), 'Orbitrap Mass Spectrometry', *Analytical Chemistry* **85**(11), 5288–5296.
**URL:** *http://pubs.acs.org/doi/10.1021/ac4001223*

# Abbreviations

| | |
|---|---|
| ANOVA | Analysis of variance |
| AOA | Aminooxyacetate |
| BL | Burkitt lymphoma |
| DLBCL | Diffuse large B-cell lymphoma |
| ESI | Electrospray ionization |
| FDR | False discovery rate |
| FT-ICR | Fourier-transform ion cyclotron resonance |
| FWHM | Full width at half maximum |
| GC | Gas chromatography |
| HILIC | Hydrophilic interaction chromatography |
| HPLC | High performance liquid chromatography |
| IDH | Isocitrate dehydrogenase |
| ISS | Isotopic steady state |
| LC | Liquid chromatography |
| LLOQ | Lower limit of quantification |
| MDV | Mass distribution vector |
| MFA | Metabolic flux analysis |
| MID | Mass isotopomer distribution |
| MPSS | Metabolic pseudo-steady state |
| MS | Mass spectrometry |
| MSS | Metabolic steady state |
| MSTFA | N-methyl-N-(trimethylsilyl)trifluoracetamid |
| MS/MS | Tandem mass spectrometry |
| MTFA | N-methyltrifluoracetamid |
| NA | Natural isotope abundance |
| NMR | Nuclear magnetic resonance spectroscopy |
| OCR | Oxygen consumption rate |
| PCA | Principal component analysis |
| PCF | Propylchloroformate |
| RP | Reversed phase |
| SIRM | Stable isotope resolved metabolomics |
| TCA cycle | Tricarboxylic acid cycle |
| TMS | Trimethylsilyl - |
| TOF | Time-of-flight |
| UHR | Ultra-high resolution |

# Appendix

## A1 IsoCorrectoR

### A1.1 Application

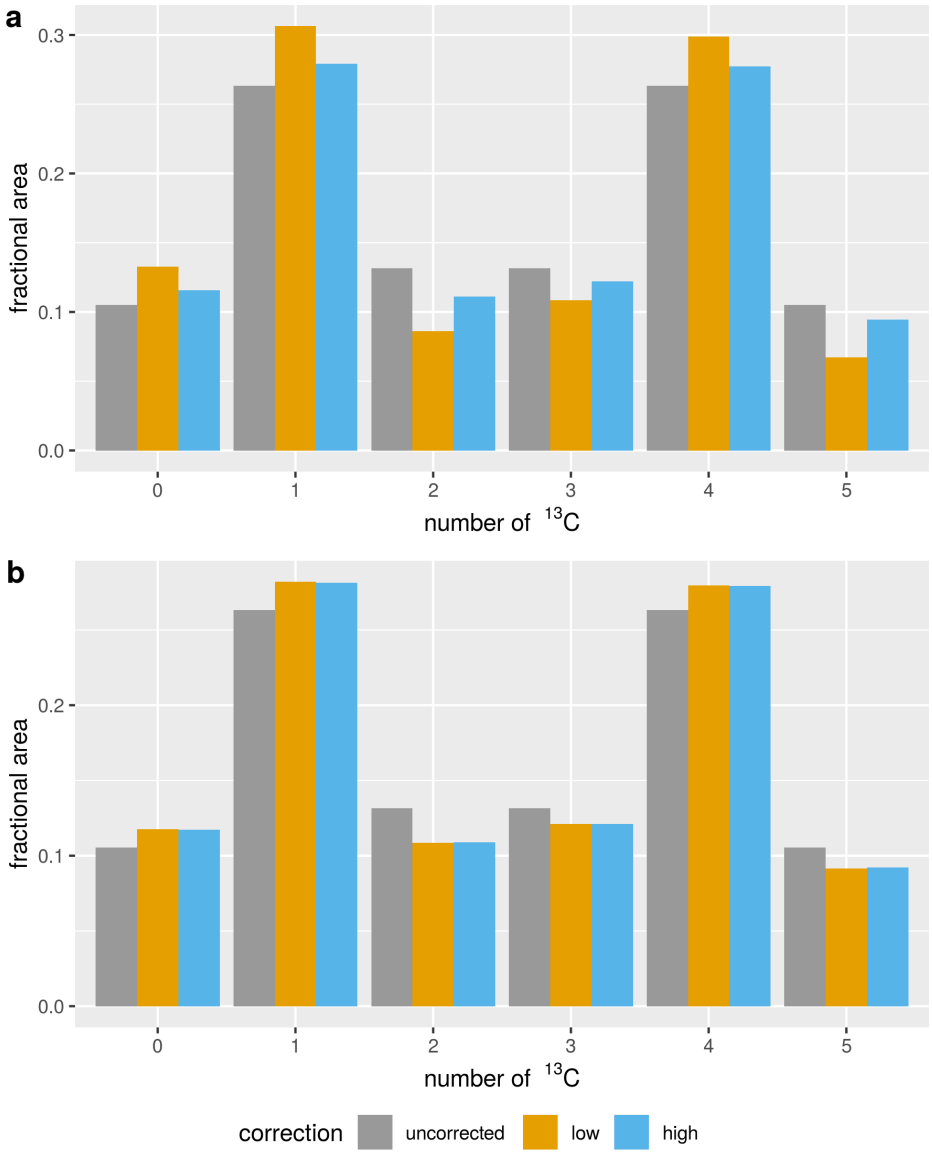



Figure A1: ***Illustration of the impact of resolution-dependent MS correction on example data, $^{13}C$ labeled derivatized proline.*** *The figure shows the results obtained from correcting simulated MS data of TMS- (a) and PCF-derivatized (b) proline. The uncorrected MID is the same for both molecules. The x-axis labels correspond to the number of $^{13}C$ label incorporated in the respective isotopologue. Data are shown uncorrected, corrected with IsoCorrectoR's low-resolution MS correction algorithm (low) and corrected with IsoCorrectoR's resolution-dependent MS correction algorithm (high). Resolution-dependent correction was performed assuming an orbitrap mass analyzer with a resolution of 140000 at 200 m/z.*

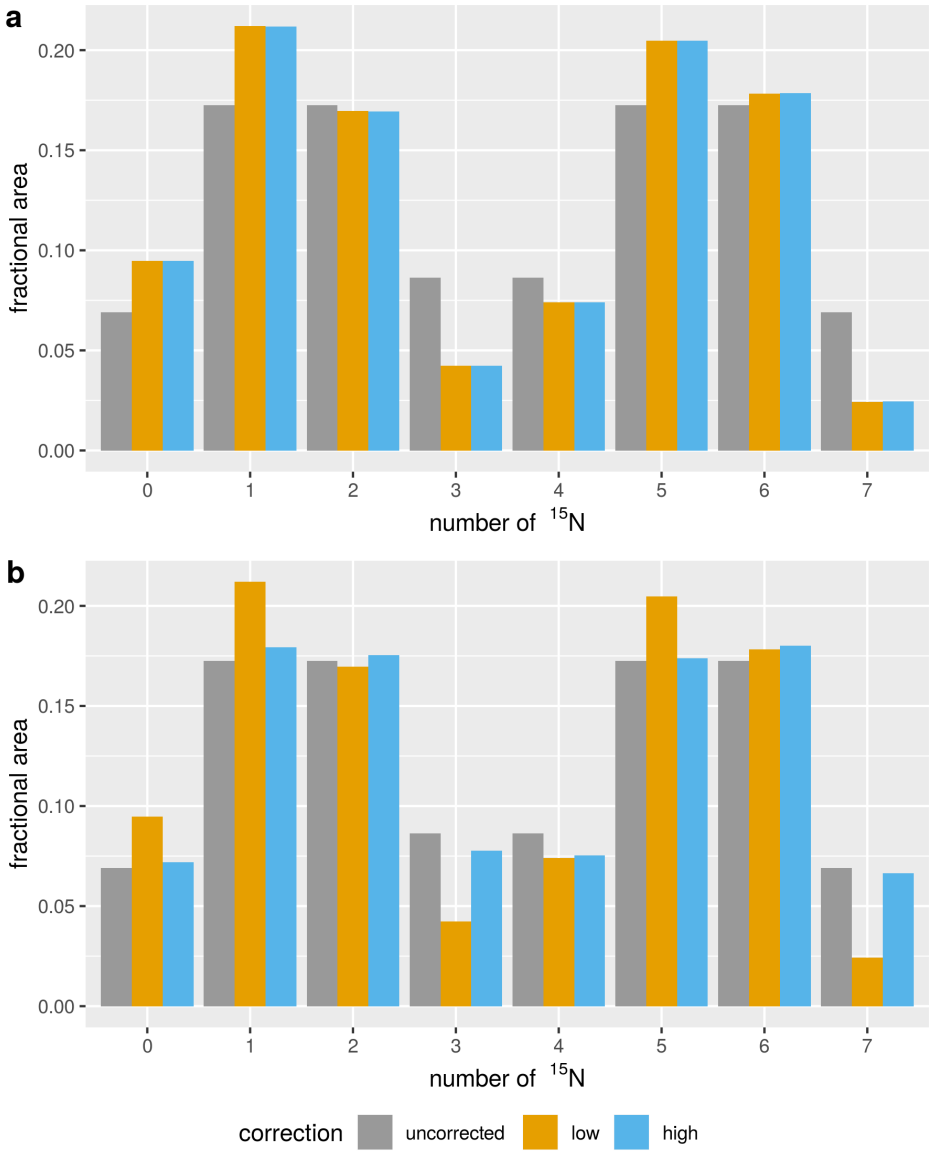



Figure A2: ***Illustration of the impact of resolution-dependent MS correction on example data from $^{15}$N labeled underivatized acetyl-CoA.*** *The x-axis labels correspond to the number of $^{15}$N label incorporated in the respective isotopologue. Data are shown uncorrected, corrected with IsoCorrectoR's low-resolution MS correction algorithm (low) and corrected with IsoCorrectoR's resolution-dependent MS correction algorithm (high). Resolution-dependent correction was performed assuming an orbitrap mass analyzer with a resolution of 140000 at 200 m/z (a) and 500000 at 200 m/z (b). At the lower of the two resolutions, $^{13}$C and $^{15}$N can't be resolved for the 810 m/z acetyl-CoA molecular ion.*

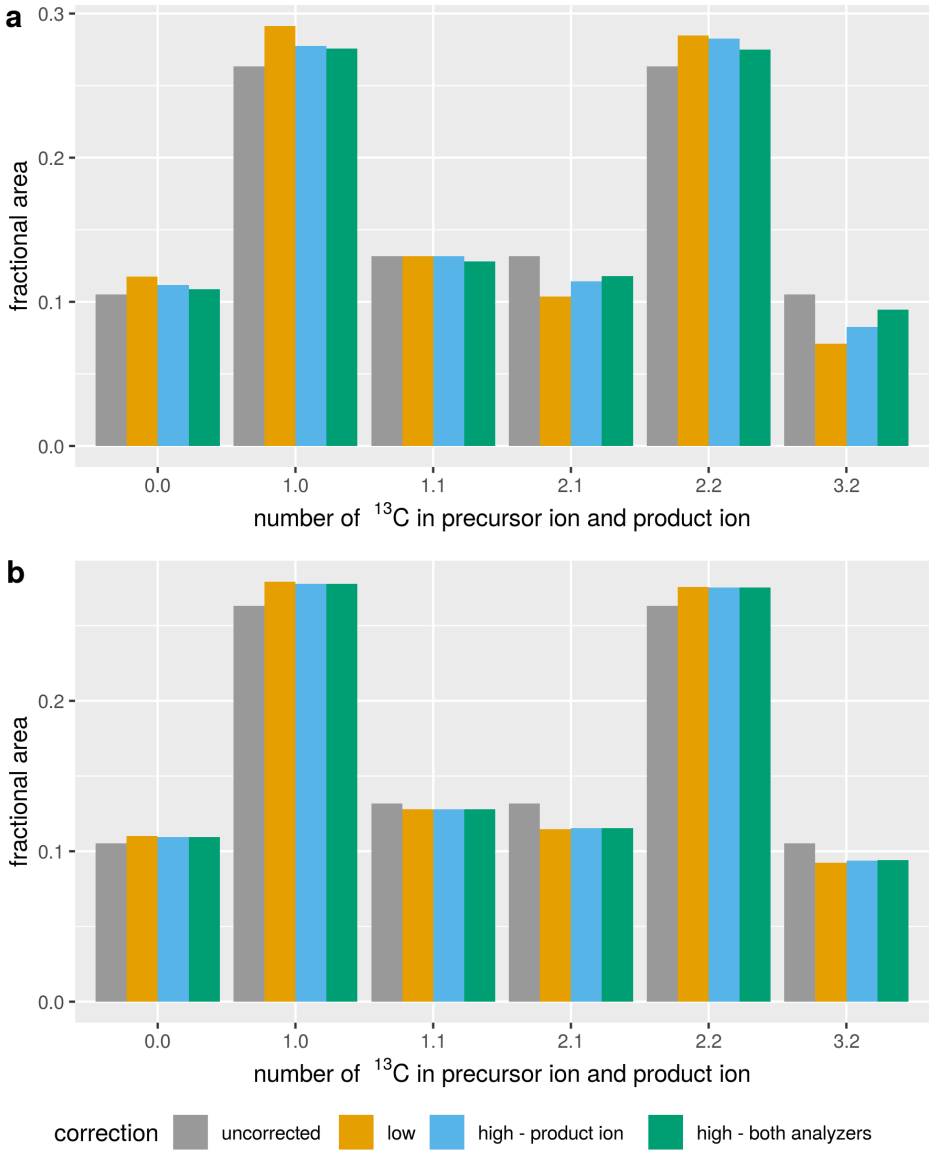



Figure A3: ***Illustration of the impact of resolution-dependent MS/MS correction on example data.*** *The figure shows the results obtained from correcting simulated MS/MS data of TMS- (a) and PCF-derivatized (b) alanine. The uncorrected MID is the same for both molecules. The x-axis labels n.m correspond to the MS/MS transitions with n $^{13}C$ in the precursor ion and m $^{13}C$ in the product ion. Data are shown uncorrected, corrected with IsoCorrectoR's low-resolution MS/MS correction algorithm (low) and corrected with IsoCorrectoR's resolution-dependent MS/MS correction algorithm (high). Resolution-dependent correction was performed for the usual case of having a low-resolution precursor mass analyzer (1000 at m/z 200, FWHM constant over the m/z range) and a high resolution product ion mass analyzer (140000 at m/z 200, orbitrap m/z dependency of FWHM), labeled "high - product ion", and for the theoretical case of both analyzers operating at high resolution, labeled "high - both analyzers".*

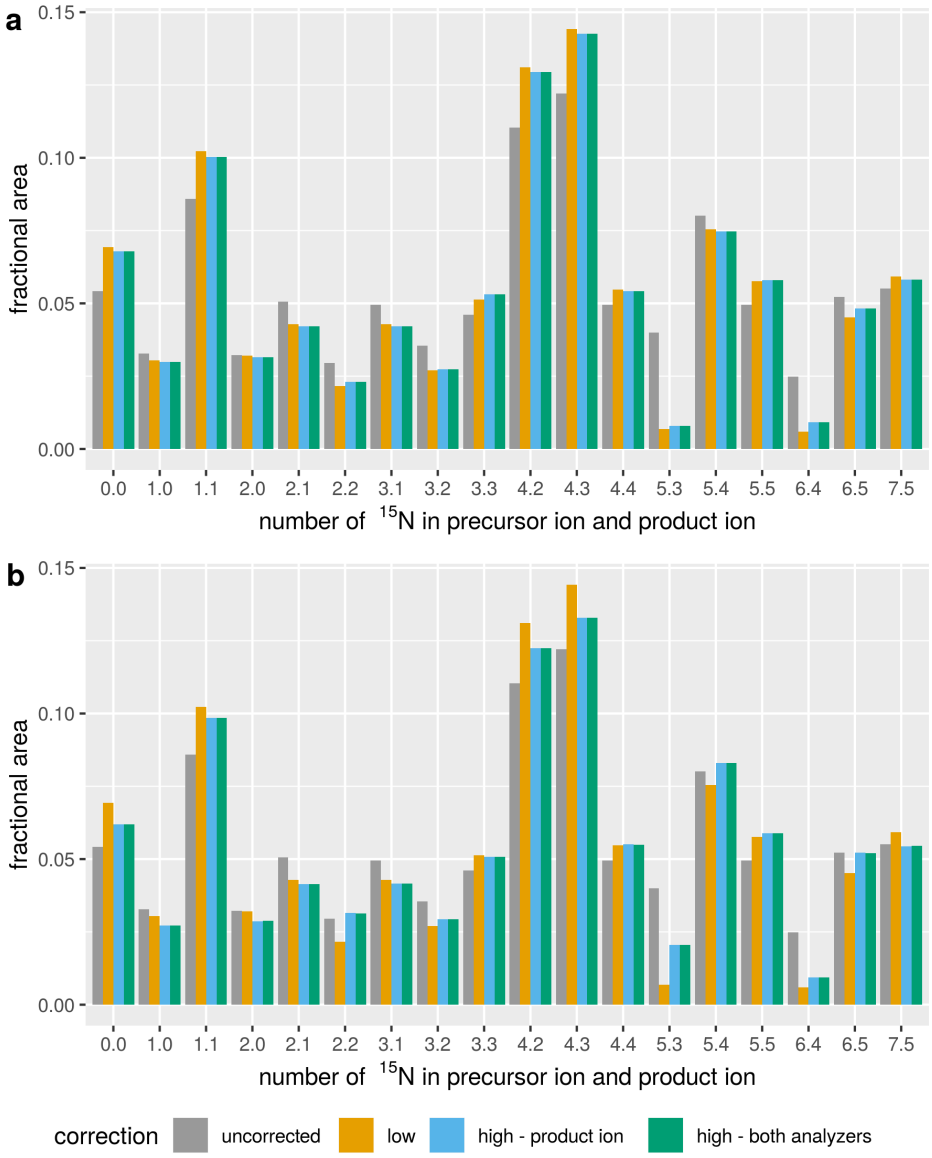



Figure A4: ***Illustration of the impact of resolution-dependent MS/MS correction on example data from $^{15}N$ labeled underivatized acetyl-CoA.*** *Precursor: $C_{23}H_{39}N_7O_{17}P_3S^+$, product ion: $C_{10}H_{16}N_5O_{10}P_2^+$ (3'-phosphonucleoside monophosphate), neutral loss: $C_{13}H_{23}N_2O_7PS$. The x-axis labels n.m correspond to the MS/MS transitions with n $^{15}N$ in the precursor ion and m $^{15}N$ in the product ion. Data are shown uncorrected, corrected with IsoCorrectoR's low-resolution MS/MS correction algorithm (low) and corrected with IsoCorrectoR's resolution-dependent MS/MS correction algorithm (high). Resolution-dependent correction was performed for the usual case of having a low-resolution precursor mass analyzer (1000 at m/z 200, FWHM constant over the m/z range) and a high resolution product ion mass analyzer, labeled "high - product ion", and for the theoretical case of both analyzers operating at high resolution, labeled "high - both analyzers". Panels a and b correspond to different resolutions of the high-resolution orbitrap mass analyzer(s): a) 140000 at m/z 200, b) 200000 at m/z 200. At the lower of the two resolutions, $^{13}C$ and $^{15}N$ can't be resolved for the acetyl-CoA product ion (nominal m/z of 428). Even at a resolution of 200000, the two isotopes cannot be resolved for the precursor (nominal m/z of 810).*

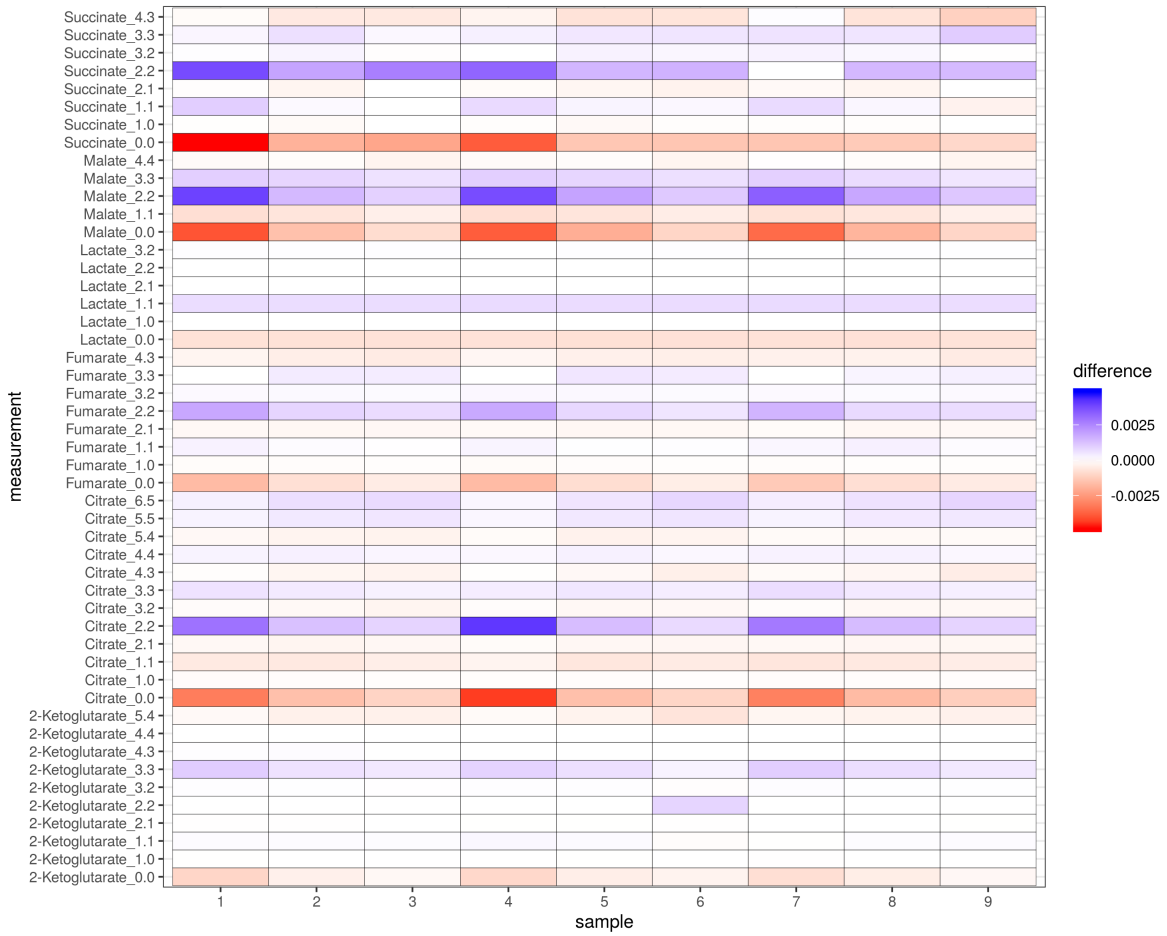



Figure A5: ***Heatmap showing differences between IsoCorrectoR's low-resolution and resolution-dependent MS/MS correction on data from a $^{13}C$ labeling experiment.*** *Organic acid data from a $^{13}C$ labeling experiment with the tracer substrate U-$^{13}C$-glutamine in a P493-6 B-cell line. Resolution-dependent correction was performed for the usual case of having a low-resolution precursor mass analyzer (1000 at m/z 200, FWHM constant over the m/z range) and a high resolution product ion mass analyzer (140000 at m/z 200, orbitrap m/z dependency of FWHM). A tracer purity of 0.99 was assumed. The heatmap was generated by subtracting MIDs from low-resolution correction from the MIDs resulting from resolution-dependent correction. Columns represent samples, while rows represent measurements of MS/MS transitions. The numbers n.m after the metabolite names correspond to the MS/MS transitions with n label in the precursor ion and m label in the product ion.*

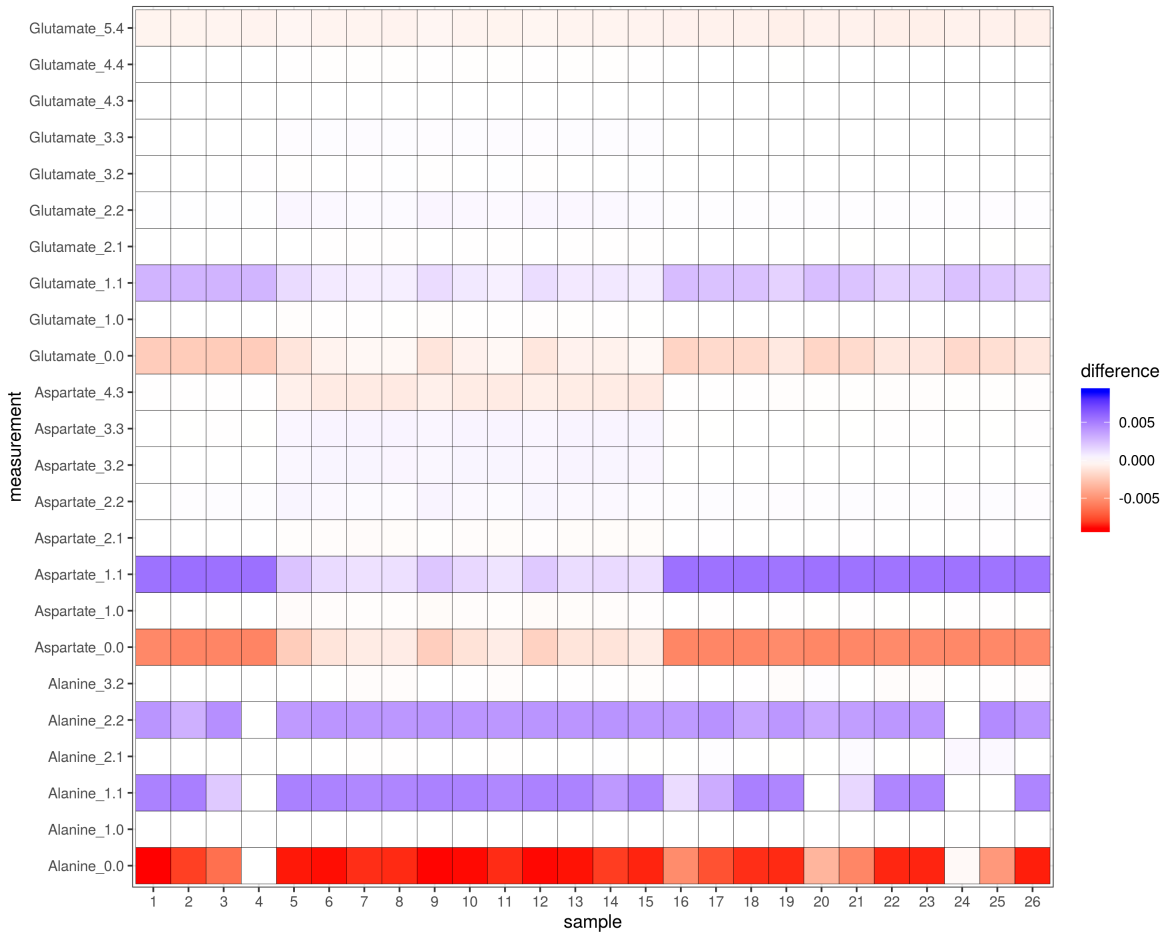



Figure A6: ***Heatmap showing differences between IsoCorrectoR's low-resolution and resolution-dependent MS/MS correction on data from a $^{13}C$ labeling experiment.*** *Amino acid data from a $^{13}C$ labeling experiment with the tracer substrate U-$^{13}C$-glutamine in a P493-6 B-cell line. Resolution-dependent correction was performed for the usual case of having a low-resolution precursor mass analyzer (1000 at m/z 200, FWHM constant over the m/z range) and a high resolution product ion mass analyzer (140000 at m/z 200, orbitrap m/z dependency of FWHM). A tracer purity of 0.99 was assumed. The heatmap was generated by subtracting MIDs from low-resolution correction from the MIDs resulting from resolution-dependent correction. Columns represent samples, while rows represent measurements of MS/MS transitions. The numbers n.m after the metabolite names correspond to the MS/MS transitions with n label in the precursor ion and m label in the product ion.*

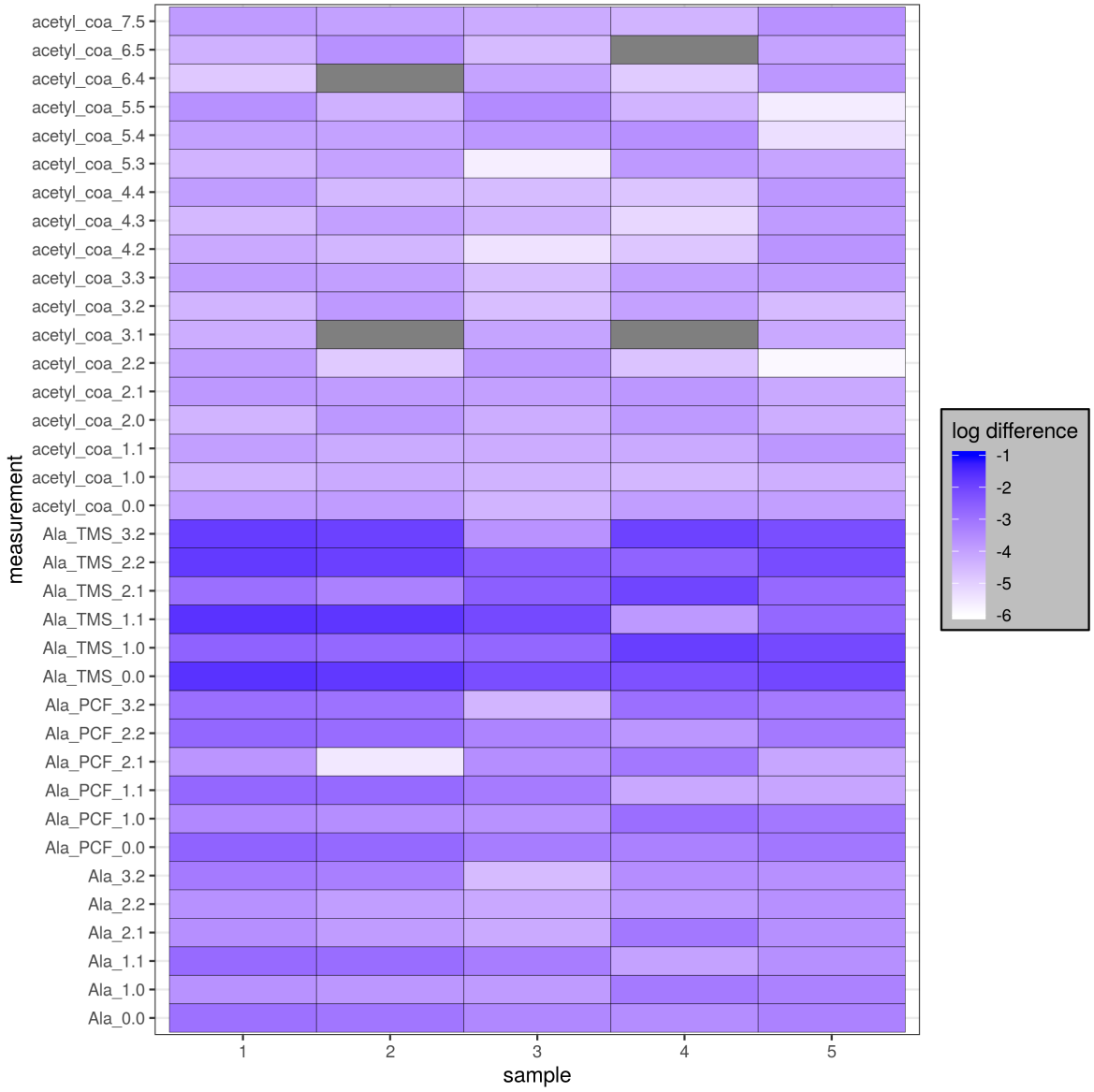



Figure A7: ***Heatmap showing differences between low-resolution and resolution-dependent MS/MS correction (product ion mass analyzer resolution of 80000 at m/z 200).*** *Simulated $^{13}C$ labeling data of underivatized and PCF-/TMS-derivatized alanine (Ala, Ala_PCF, Ala_TMS), simulated $^{15}N$ labeling data of acetyl-CoA. A tracer purity of 0.99 was assumed. The heatmap was generated by subtracting MIDs corrected via IsoCorrectoR's low-resolution MS/MS correction approach from MIDs corrected via the resolution-dependent approach. For resolution-dependent MS/MS correction, the precursor mass analyzer was assumed to operate at a resolution of 1000 at m/z 200 (FWHM constant over the m/z range), while the product ion mass analyzer was assumed to be an orbitrap operating at a resolution of 80000 at m/z 200. The (absolute) differences were $log_{10}$ transformed, gray fields indicate an untransformed difference of 0 (equality of values). Columns represent samples, while rows represent measurements of MS/MS transitions. The numbers n.m after the metabolite names correspond to the MS/MS transitions with n label in the precursor ion and m label in the product ion.*

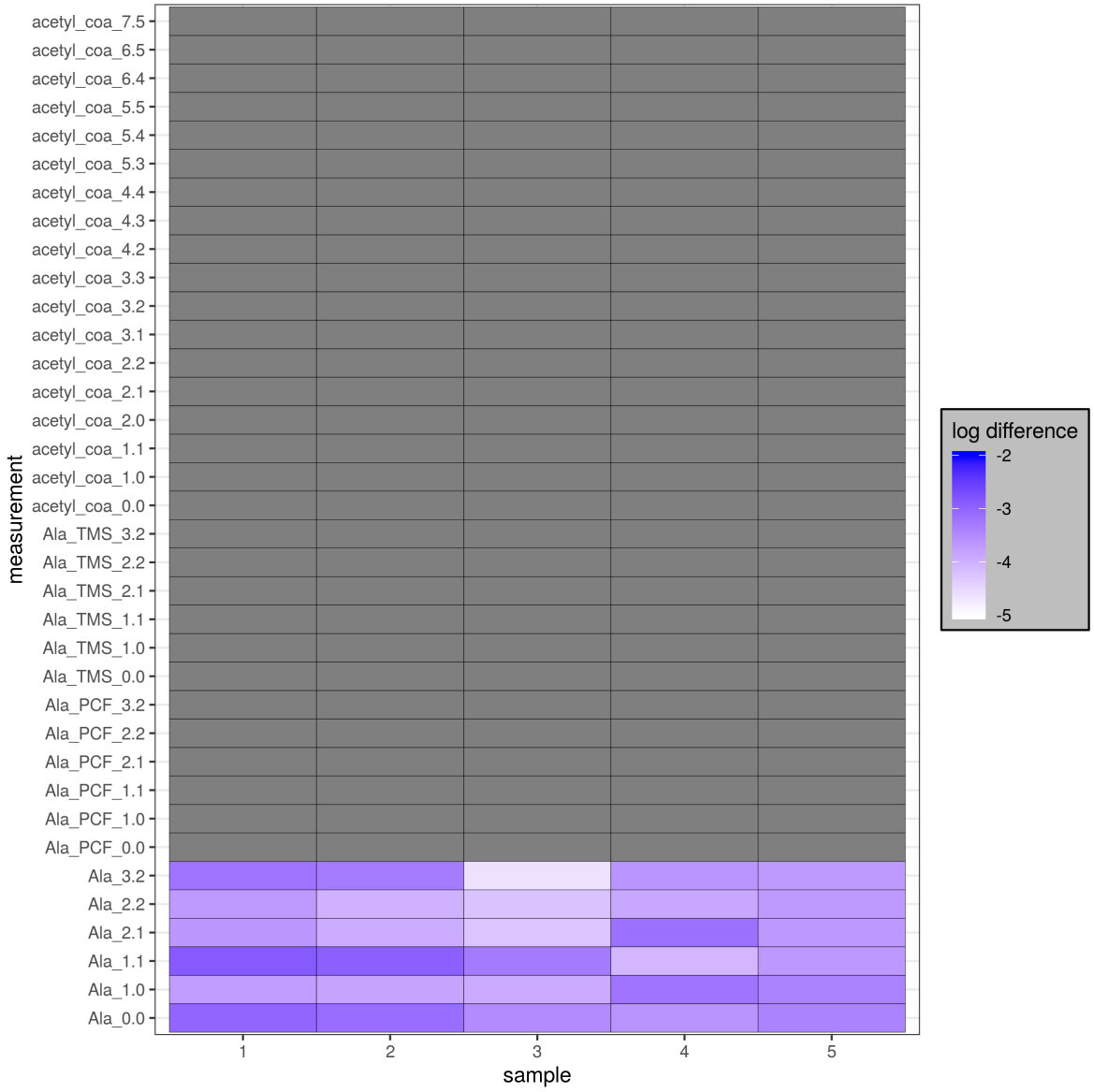



Figure A8: ***Heatmap showing differences between low-resolution and resolution-dependent MS/MS correction (product ion mass analyzer resolution of 10000 at m/z 200).*** *Simulated $^{13}C$ labeling data of underivatized and PCF-/TMS-derivatized alanine (Ala, Ala_PCF, Ala_TMS), simulated $^{15}N$ labeling data of acetyl-CoA. A tracer purity of 0.99 was assumed. The heatmap was generated by subtracting MIDs corrected via IsoCorrectoR's low-resolution MS/MS correction approach from MIDs corrected via the resolution-dependent approach. For resolution-dependent MS/MS correction, the precursor mass analyzer was assumed to operate at a resolution of 1000 at m/z 200 (FWHM constant over the m/z range), while the product ion mass analyzer was assumed to be an orbitrap operating at a resolution of 10000 at m/z 200. The (absolute) differences were $log_{10}$ transformed, gray fields indicate an untransformed difference of 0 (equality of values). Columns represent samples, while rows represent measurements of MS/MS transitions. The numbers n.m after the metabolite names correspond to the MS/MS transitions with n label in the precursor ion and m label in the product ion.*

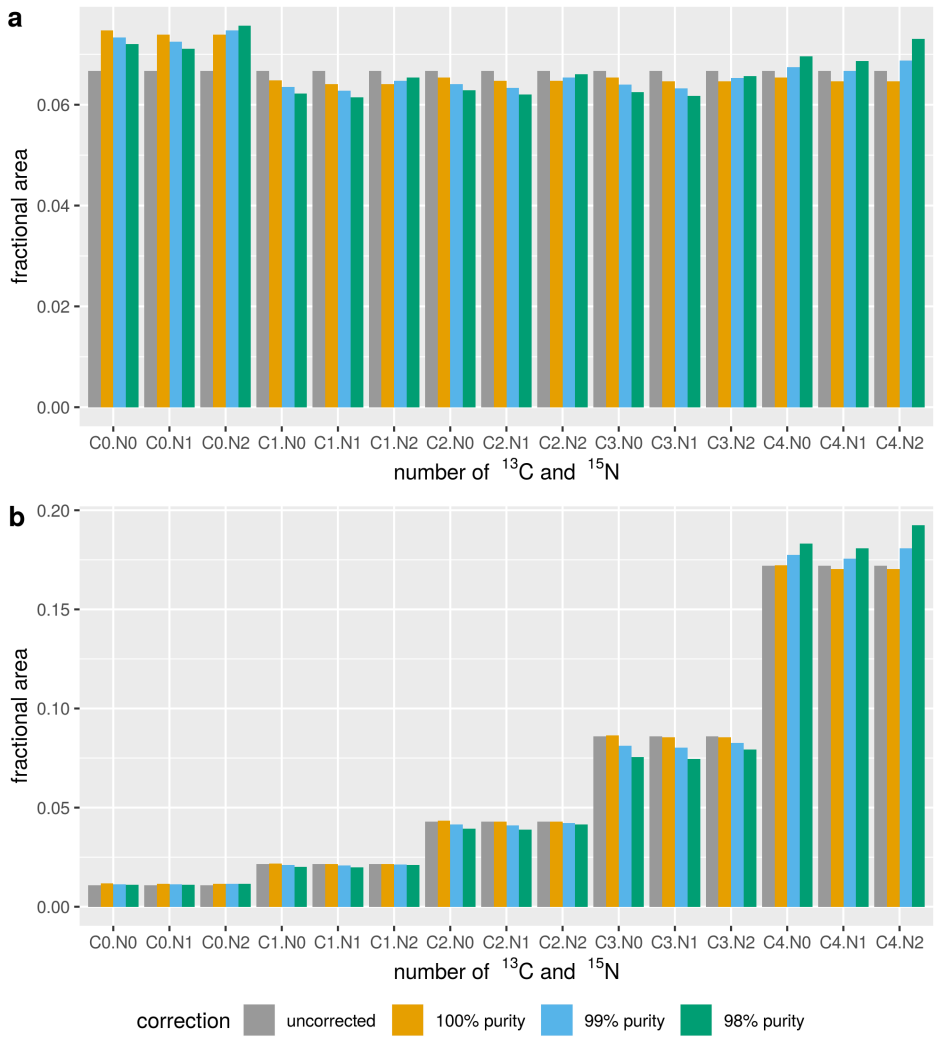



Figure A9: ***Illustration of the impact of UHR (and tracer purity) correction on simulated example data from simultaneously $^{13}C$ and $^{15}N$ labeled PCF-derivatized asparagine.*** *The x-axis labels Cx.Ny correspond to x $^{13}C$ label and y $^{15}N$ label incorporated in the respective isotopologue. Data are shown uncorrected and corrected with IsoCorrectoR's UHR correction algorithm at different assumed purities of the tracer substrate. The panels a and b correspond to different uncorrected MIDs.*

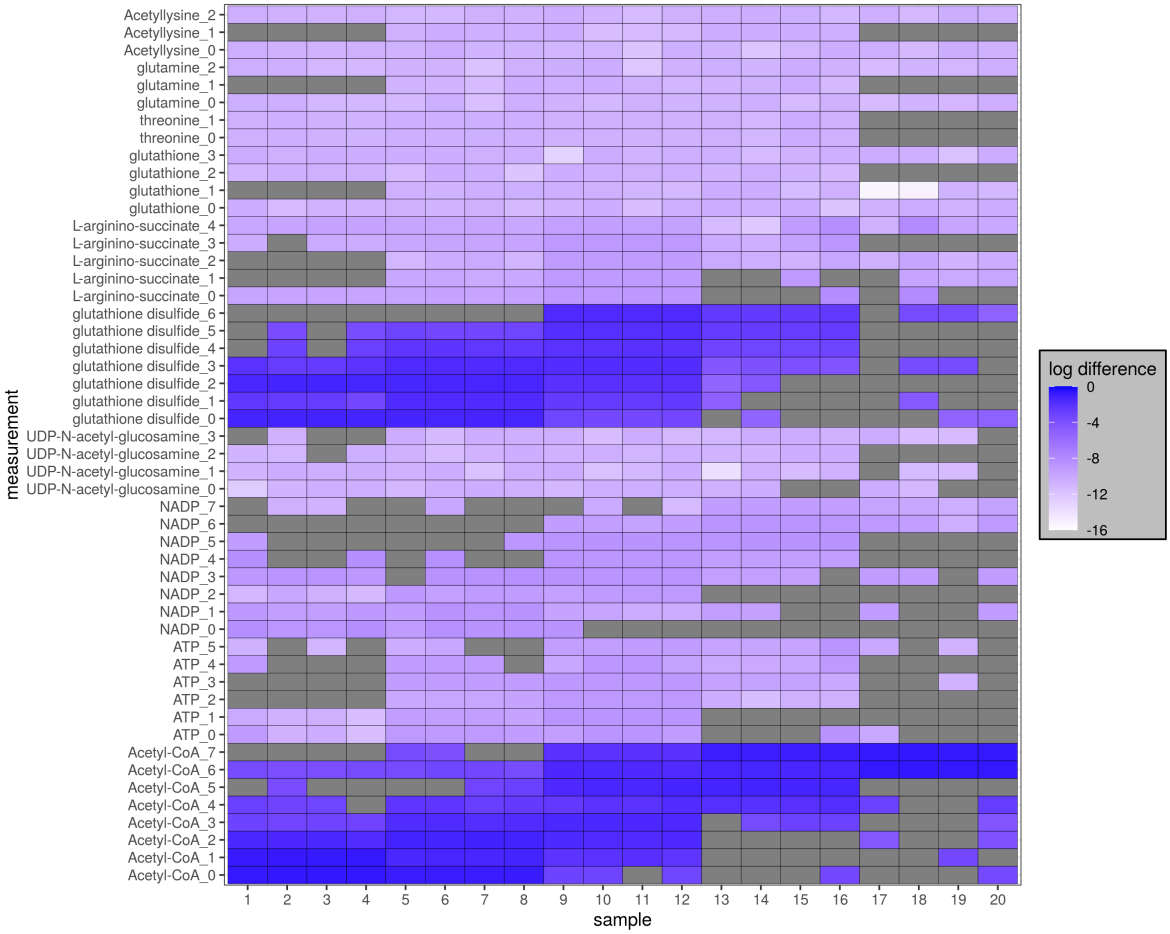



Figure A10: ***Heatmap showing differences between UHR and resolution-dependent correction of data from a $^{15}N$ labeling experiment with IsoCorrectoR.*** *Data used are example data supplied with AccuCor (Su et al. (2017)). Correction was performed assuming a tracer purity of 0.99. An FT-ICR mass analyzer with a resolution of 400000 at 400 m/z was assumed for resolution-dependent correction. The heatmap was generated by subtracting MIDs obtained by UHR correction from MIDs obtained through resolution-dependent correction. The (absolute) differences were $log_{10}$ transformed, gray fields indicate an untransformed difference of 0 (equality of values). Columns represent samples, while rows represent measurements of mass isotopomers. The number after the metabolite name corresponds to the number of label incorporated.*

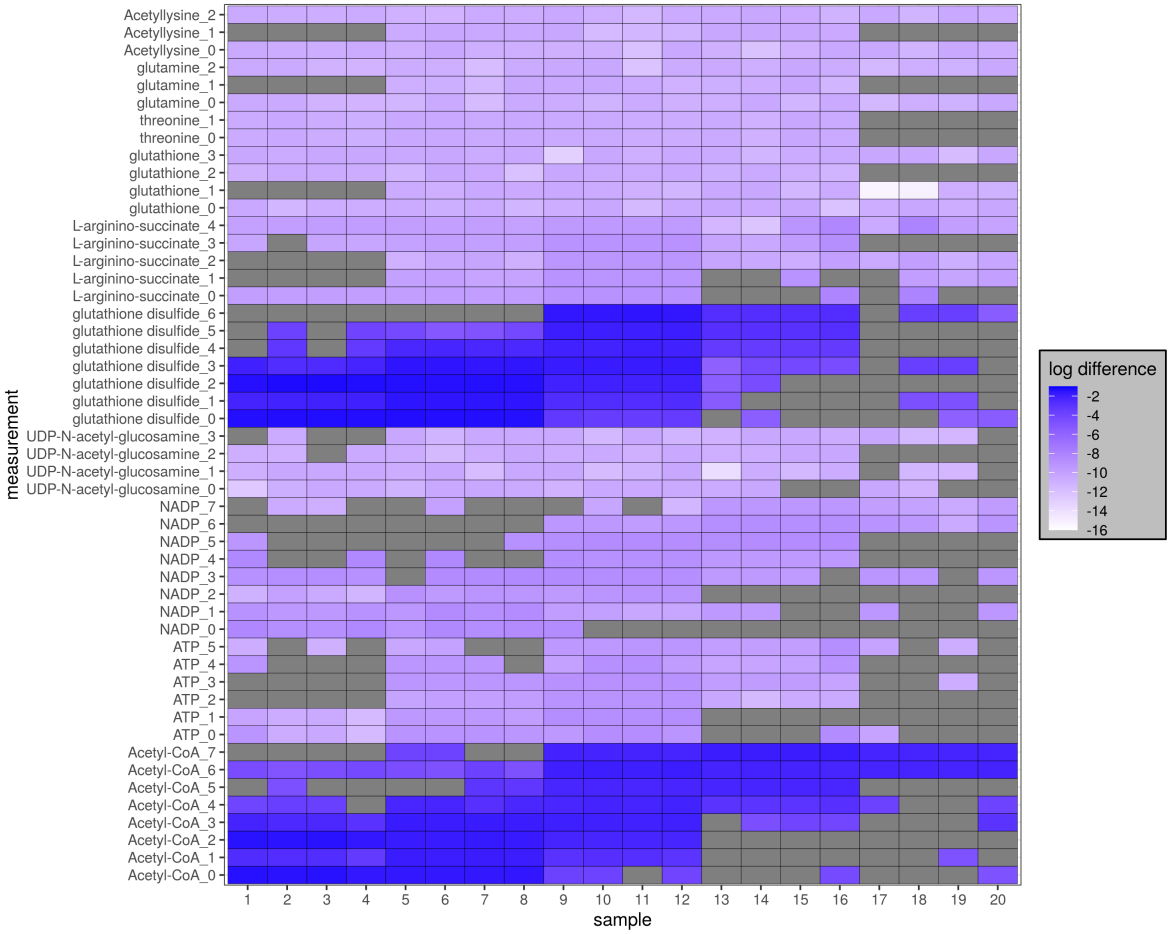



Figure A11: ***Heatmap showing differences between UHR and resolution-dependent correction of data from a $^{15}N$ labeling experiment with IsoCorrectoR.*** *Data used are example data supplied with AccuCor (Su et al. (2017)). Correction was performed assuming a tracer purity of 0.99. An FT-ICR mass analyzer with a resolution of 800000 at 400 m/z was assumed for resolution-dependent correction. The heatmap was generated by subtracting MIDs obtained by UHR correction from MIDs obtained through resolution-dependent correction. The (absolute) differences were $log_{10}$ transformed, gray fields indicate an untransformed difference of 0 (equality of values). Columns represent samples, while rows represent measurements of mass isotopomers. The number after the metabolite name corresponds to the number of label incorporated.*

## A1.2 Validation

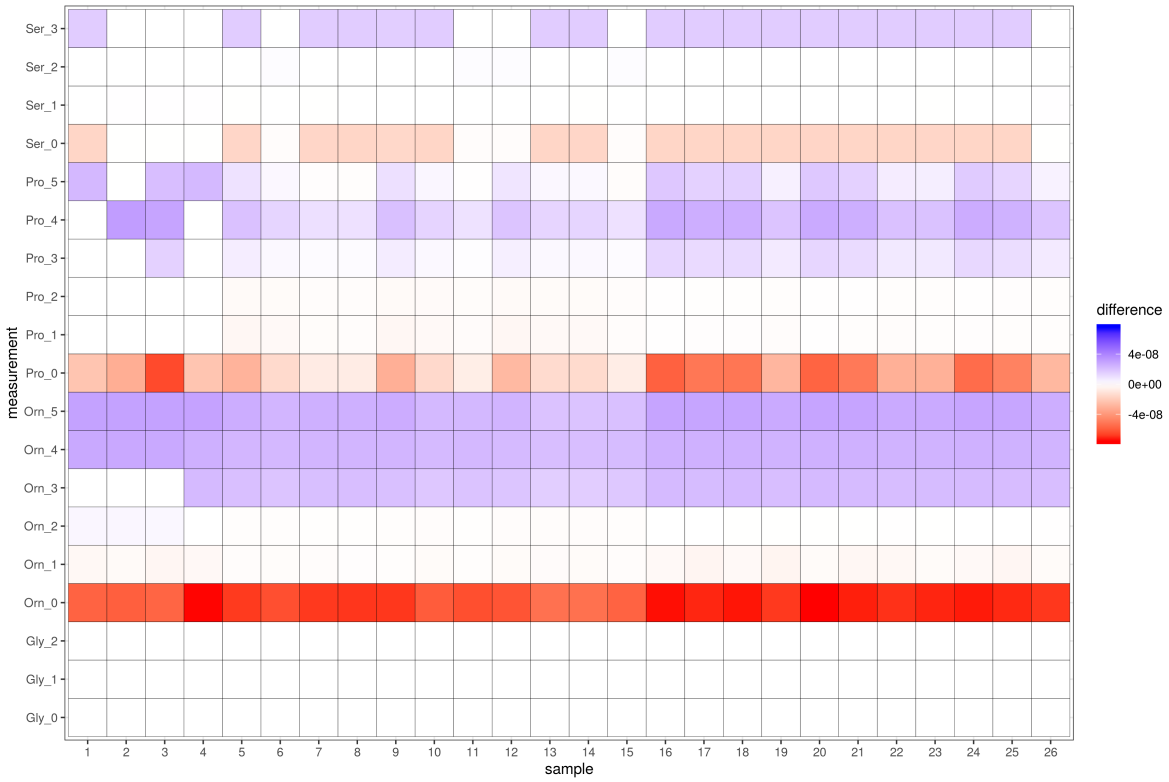



Figure A12: ***Heatmap showing differences between IsoCorrectoR and IsoCor v2 for the low-resolution MS correction of data from a $^{13}C$ labeling experiment.*** *Amino acid data from a $^{13}C$ labeling experiment with the tracer substrate U-$^{13}C$-glutamine in a P493-6 B-cell line. Data not corrected for tracer purity. The heatmap was generated by subtracting MIDs corrected by IsoCor v2 from MIDs corrected by IsoCorrectoR. Columns represent samples, while rows represent measurements of mass isotopomers. The number after the metabolite name corresponds to the number of label incorporated. Abbreviated metabolite names: Gly - Glycine, Orn - Ornithine, Pro - Proline, Ser - Serine.*

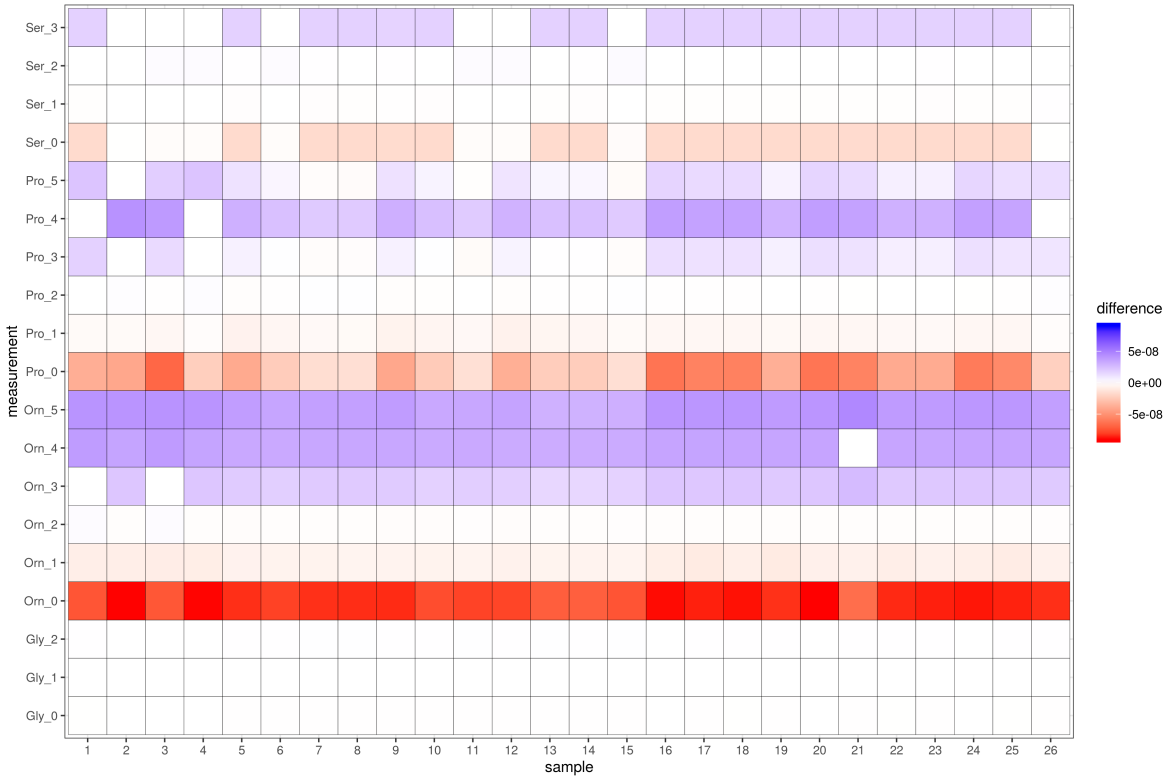



Figure A13: ***Heatmap showing differences between IsoCorrectoR and IsoCor v2 for the low-resolution MS correction of data from a $^{13}C$ labeling experiment.*** *Amino acid data from a $^{13}C$ labeling experiment with the tracer substrate U-$^{13}C$-glutamine in a P493-6 B-cell line. Data not corrected for tracer element natural abundance in the core molecule (part of the molecule not coming from derivatization). Tracer purity was assumed to be 0.99. The heatmap was generated by subtracting MIDs corrected by IsoCor v2 from MIDs corrected by IsoCorrectoR. Columns represent samples, while rows represent measurements of mass isotopomers. The number after the metabolite name corresponds to the number of label incorporated. Abbreviated metabolite names: Gly - Glycine, Orn - Ornithine, Pro - Proline, Ser - Serine.*

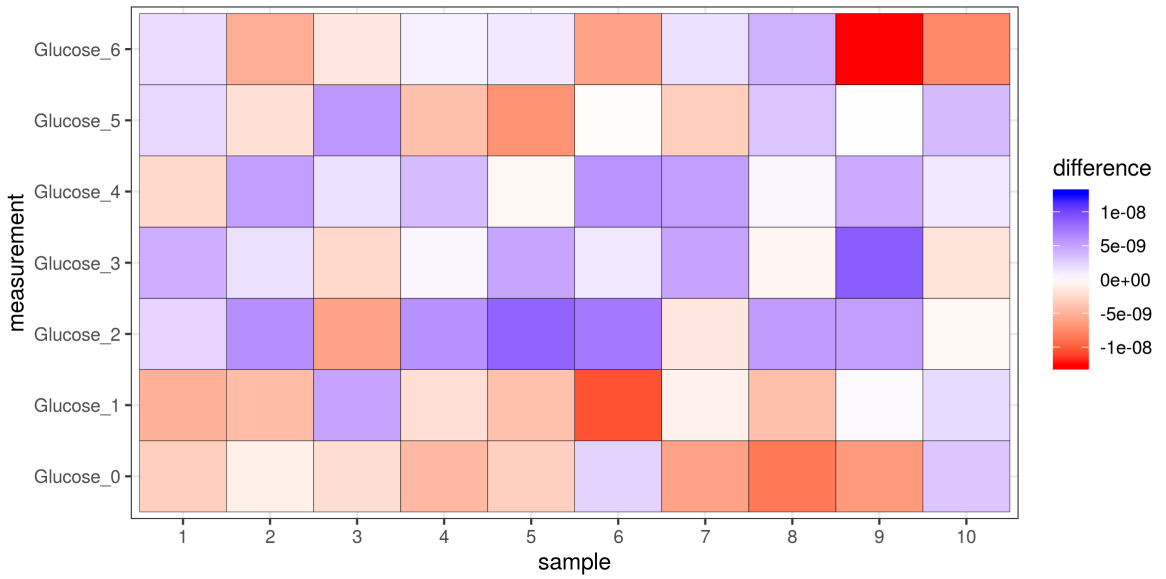



Figure A14: ***Heatmap showing differences between IsoCorrectoR and IsoCor v2 for the low-resolution MS correction of data from a simulated $^{18}O$ labeling experiment.*** *Correction of data from glucose ($C_6H_{11}O_6^-$), tracer purity was assumed to be 0.99. The heatmap was generated by subtracting MIDs corrected by IsoCor v2 from MIDs corrected by IsoCorrectoR. Columns represent samples, while rows represent measurements of mass isotopomers. The number after the metabolite name corresponds to the number of label incorporated.*

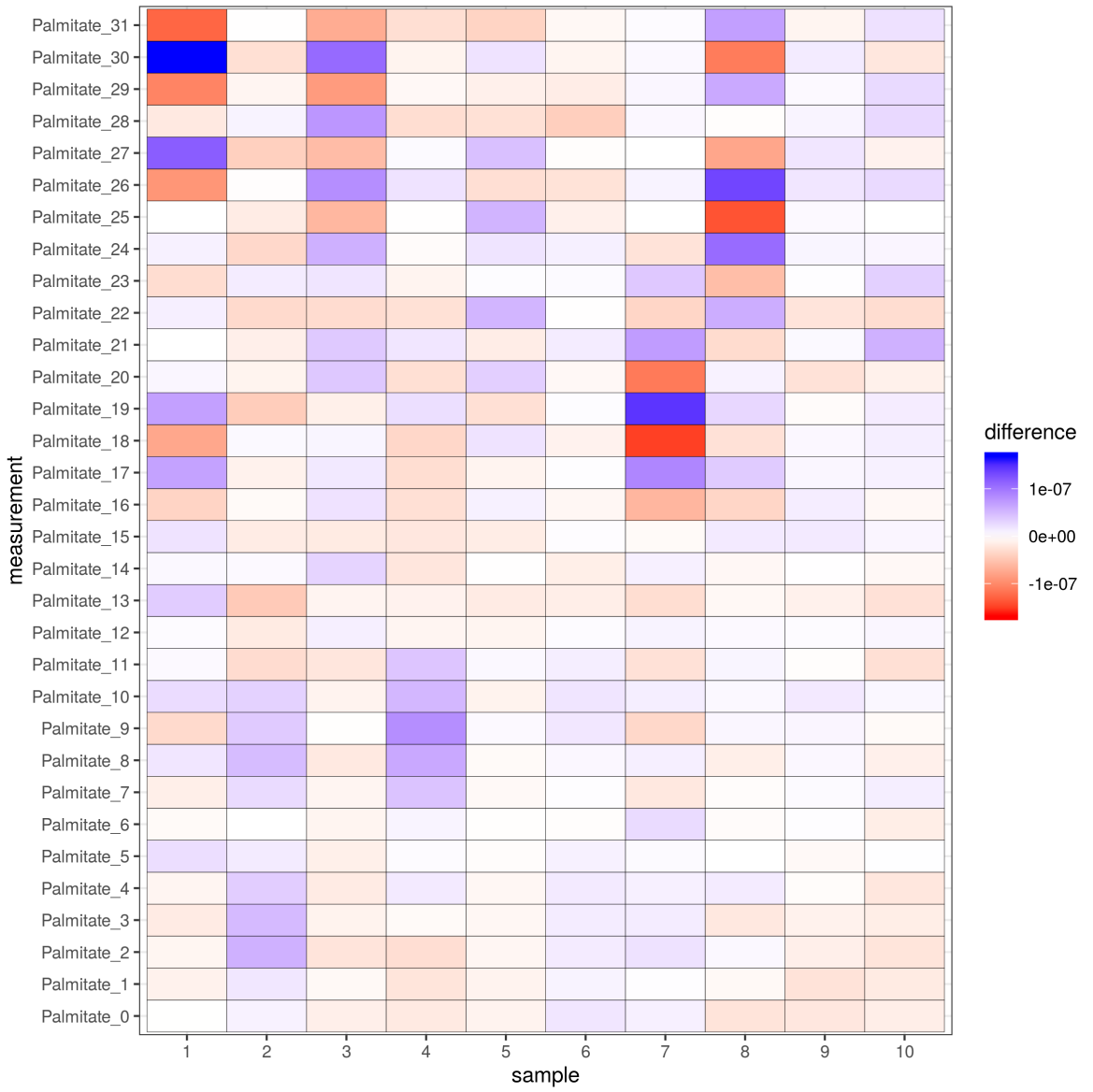



Figure A15: ***Heatmap showing differences between IsoCorrectoR and IsoCor v2 for the low-resolution MS correction of data from a simulated $^{2}H$ labeling experiment.*** *Correction of data from palmitate ($C_{16}H_{31}O_{2}^{-}$), tracer purity was assumed to be 0.99. The heatmap was generated by subtracting MIDs corrected by IsoCor v2 from MIDs corrected by IsoCorrectoR. Columns represent samples, while rows represent measurements of mass isotopomers. The number after the metabolite name corresponds to the number of label incorporated.*

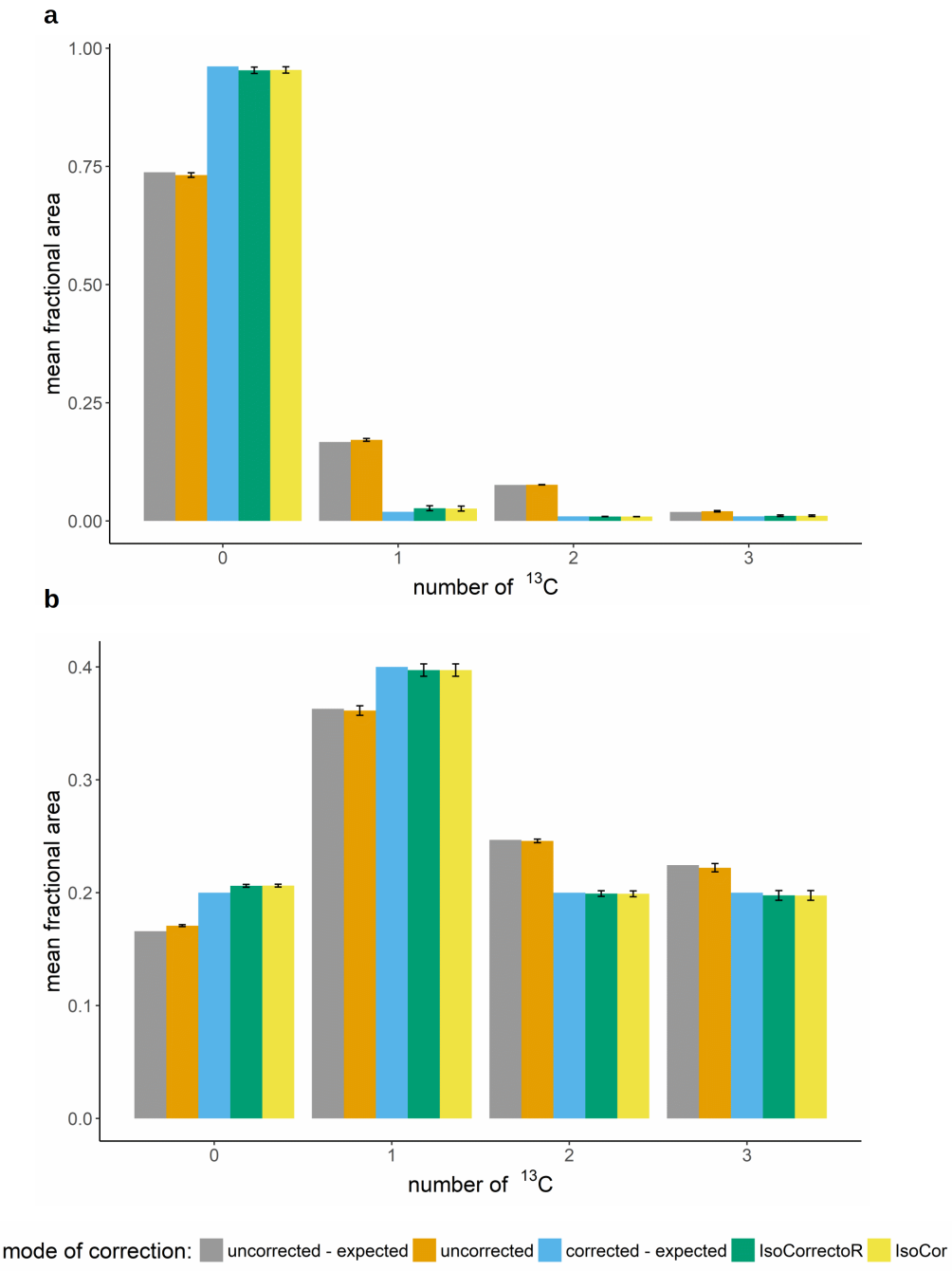



Figure A16: ***Experimental validation of low-resolution MS correction with isotopologue mixtures of known composition.*** *Low-resolution MS correction of GC-MS data of mixtures of TMS-derivatized alanine isotopologues. The mixtures have known composition. The x-axis labels 0 - 3 correspond to the TMS-alanine isotopologues with 0 - 3 $^{13}C$ incorporated. The grouped bar charts depict the correction performance of IsoCorrectoR or IsoCor in comparison to the expected values. Additionally, the expected uncorrected values (reverse application of IsoCorrectoR functions) are compared to the uncorrected values actually measured. Samples were measured in technical triplicates, means of isotopologue fractions +/- SD are shown. Panels a and b show different mixtures.* ***(Figure and legend taken from Heinrich et al. (2018) supplementary material and slightly modified)***

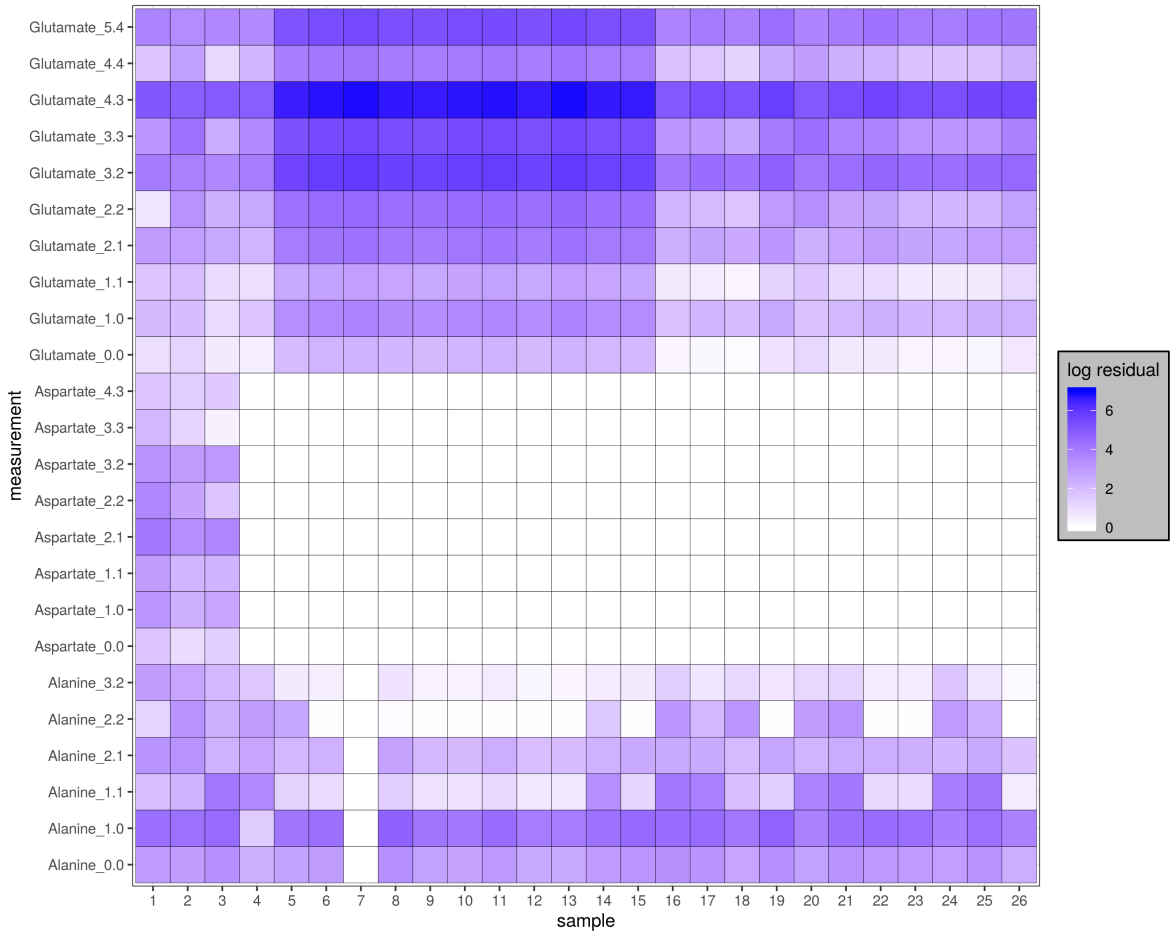



Figure A17: ***Heatmap showing the correction residuals for the low-resolution MS/MS correction of data from a $^{13}C$ labeling experiment with IsoCorrectoR.*** *Amino acid data from a $^{13}C$ labeling experiment with the tracer substrate U-$^{13}C$-glutamine in a P493-6 B-cell line. A tracer purity of 0.99 was assumed. The (absolute) values were added a constant of 1 and $\log_{10}$ transformed. Columns represent samples, while rows represent measurements of MS/MS transitions. The numbers n.m after the metabolite names correspond to the MS/MS transitions with n label in the precursor ion and m label in the product ion.*

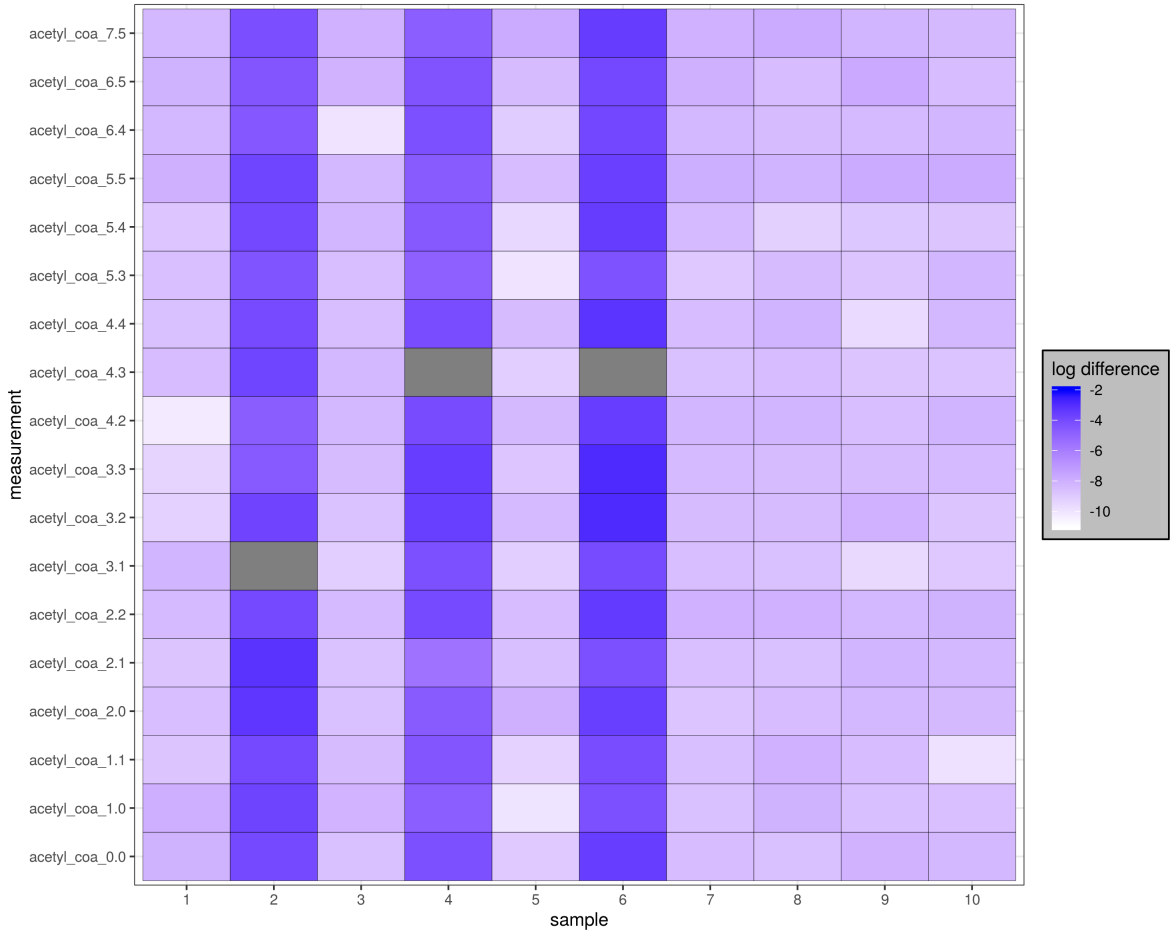



Figure A18: ***Heatmap showing differences between IsoCorrectoR and ICT for the low-resolution MS/MS correction of data from a $^{15}N$ labeling experiment.*** *Simulated acetyl-CoA data, the precursor $C_{23}H_{39}N_7O_{17}P_3S^+$ fragments into the product ion $C_{10}H_{16}N_5O_{10}P_2{}^+$ (3'-phosphonucleoside monophosphate) and the neutral loss $C_{13}H_{23}N_2O_7PS$. A tracer purity of 0.99 was assumed. The heatmap was generated by subtracting MIDs corrected by ICT from MIDs corrected by IsoCorrectoR. The (absolute) differences were $log_{10}$ transformed, gray fields indicate an untransformed difference of 0 (equality of values). Columns represent samples, while rows represent measurements of MS/MS transitions. The numbers n.m after the metabolite names correspond to the MS/MS transitions with n label in the precursor ion and m label in the product ion.*

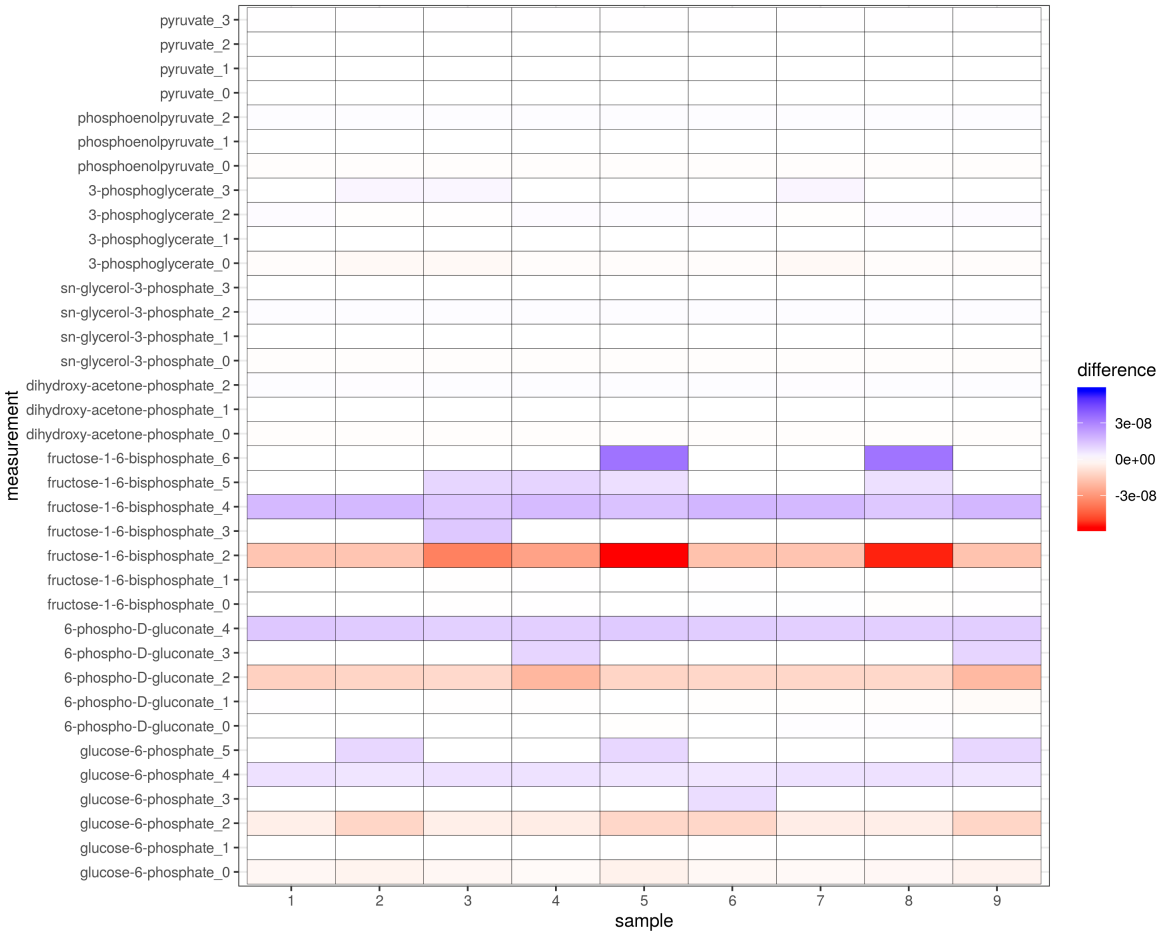



Figure A19: ***Heatmap showing differences between IsoCorrectoR and IsoCor v2 for the resolution-dependent MS correction of data from a $^{13}C$ labeling experiment.*** *Data used are example data supplied with AccuCor (Su et al. (2017)). A tracer purity of 0.99 and an orbitrap mass analyzer with a resolution of 140000 at 200 m/z were assumed. The heatmap was generated by subtracting MIDs corrected by IsoCor v2 from MIDs corrected by IsoCorrectoR. Columns represent samples, while rows represent measurements of mass isotopomers. The number after the metabolite name corresponds to the number of label incorporated.*

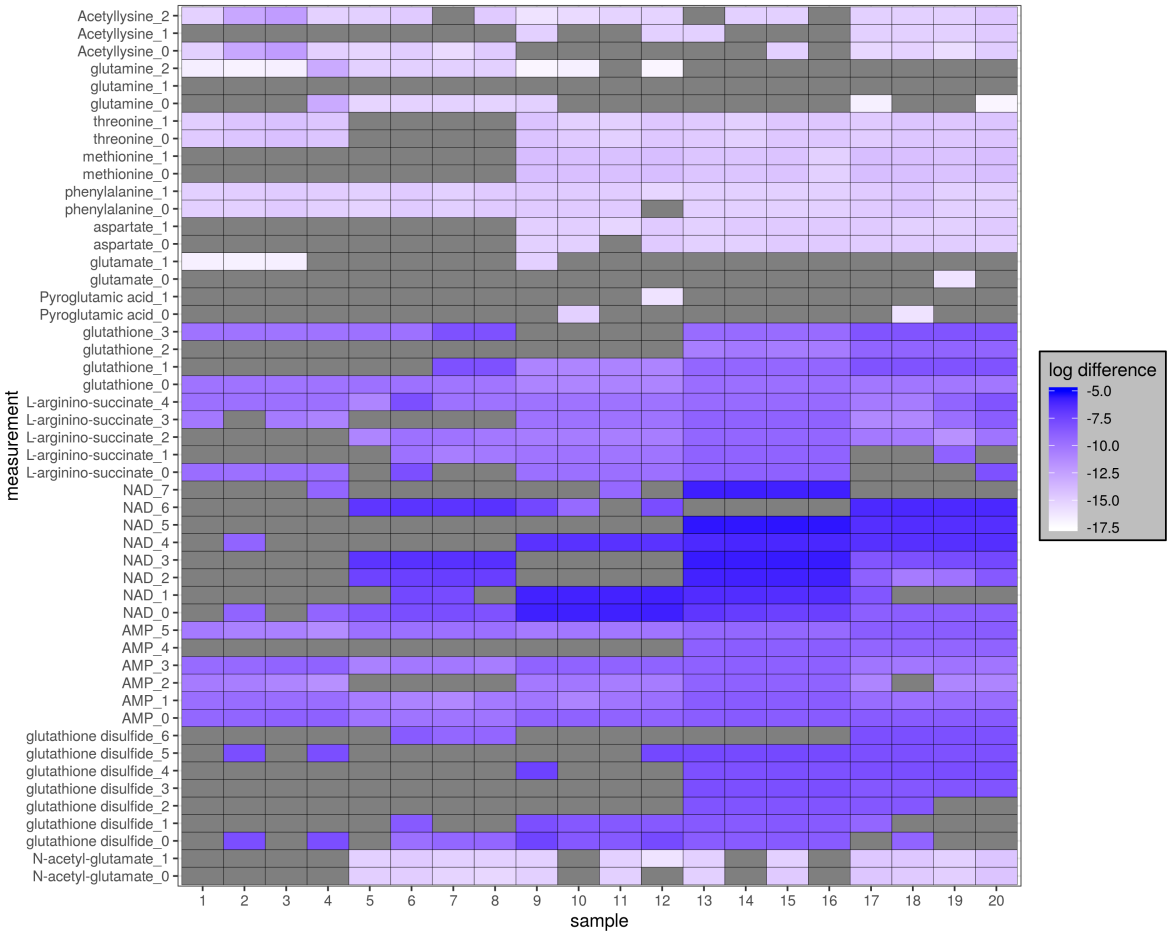



Figure A20: ***Heatmap showing differences between IsoCorrectoR and IsoCor v2 for the resolution-dependent MS correction of data from a $^{15}N$ labeling experiment.*** *Data used are example data supplied with AccuCor (Su et al. (2017)). A tracer purity of 0.99 and an orbitrap mass analyzer with a resolution of 140000 at 200 m/z were assumed. The heatmap was generated by subtracting MIDs corrected by IsoCor v2 from MIDs corrected by IsoCorrectoR. The (absolute) differences were $log_{10}$ transformed, gray fields indicate an untransformed difference of 0 (equality of values). Columns represent samples, while rows represent measurements of mass isotopomers. The number after the metabolite name corresponds to the number of label incorporated.*

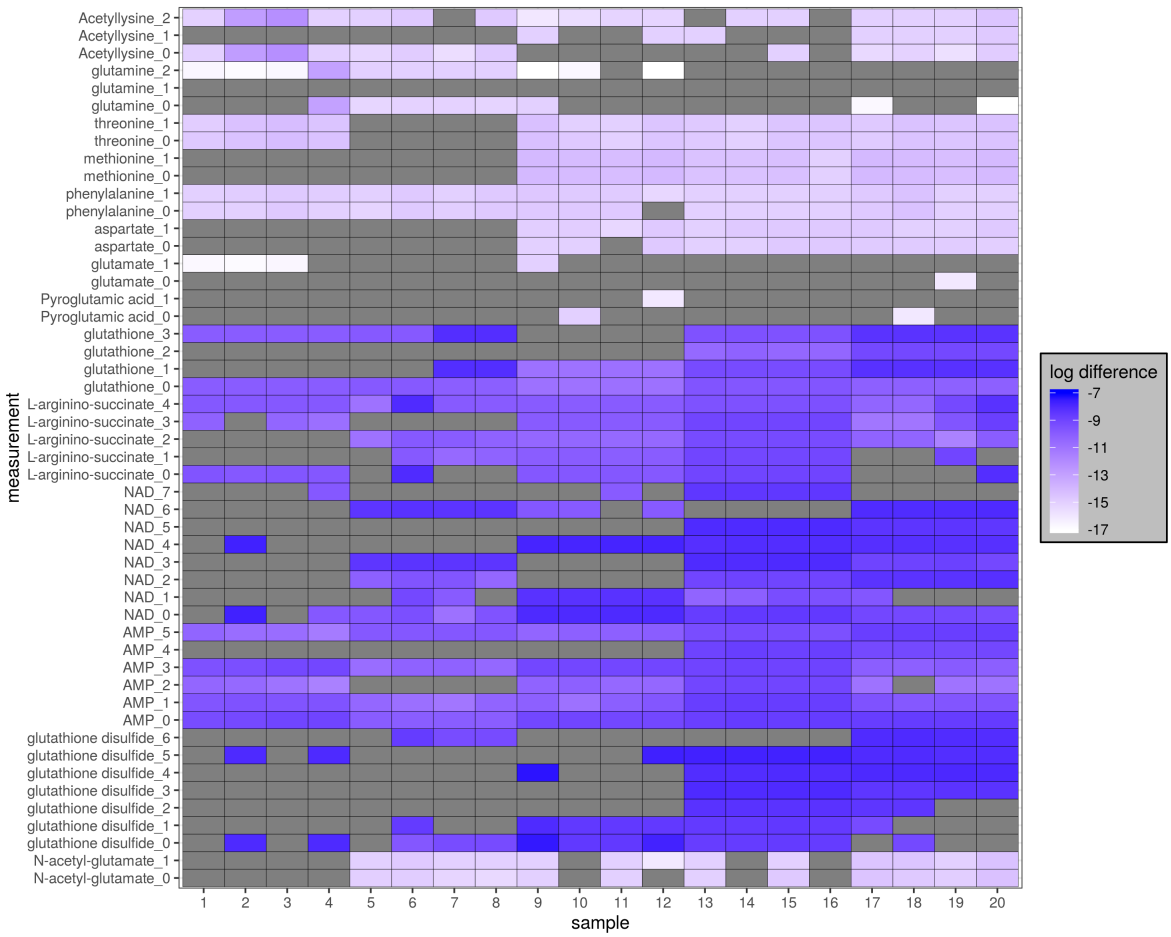



Figure A21: ***Heatmap showing differences between IsoCorrectoR and IsoCor v2 for the resolution-dependent MS correction of data from a $^{15}N$ labeling experiment, altered IsoCorrectoR algorithm.*** *Data used are example data supplied with AccuCor (Su et al. (2017)). A tracer purity of 0.99 and an orbitrap mass analyzer with a resolution of 140000 at 200 m/z were assumed. IsoCorrectoR's algorithm was adjusted to always use the m/z of the m+0 mass isotopomer of a given metabolite (instead of the respective m+i mass isotopomer) for computing FWHM. The heatmap was generated by subtracting MIDs corrected by IsoCor v2 from MIDs corrected by IsoCorrectoR. The (absolute) differences were $log_{10}$ transformed, gray fields indicate an untransformed difference of 0 (equality of values). Columns represent samples, while rows represent measurements of mass isotopomers. The number after the metabolite name corresponds to the number of label incorporated.*

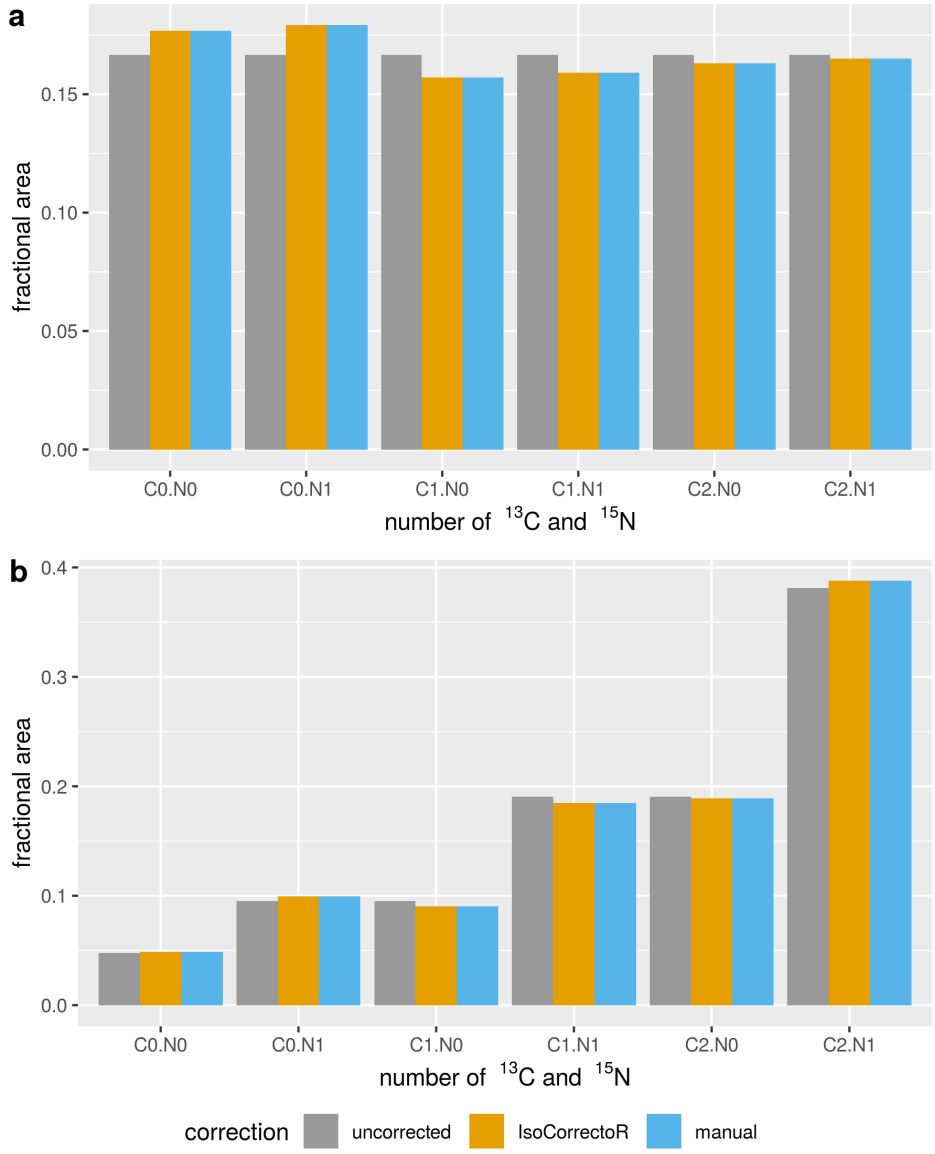



Figure A22: ***Comparison of UHR correction to manual calculations on simulated UHR example data of simultaneously $^{13}C$ and $^{15}N$ labeled PCF-derivatized glycine.*** *A tracer purity of 99% was assumed for correction. The x-axis labels Cx.Ny correspond to x $^{13}C$ label and y $^{15}N$ label incorporated in the respective isotopologue. Data are shown uncorrected, corrected with IsoCorrectoR's UHR correction algorithm and corrected through manual calculations. The panels a and b correspond to different uncorrected MIDs.*

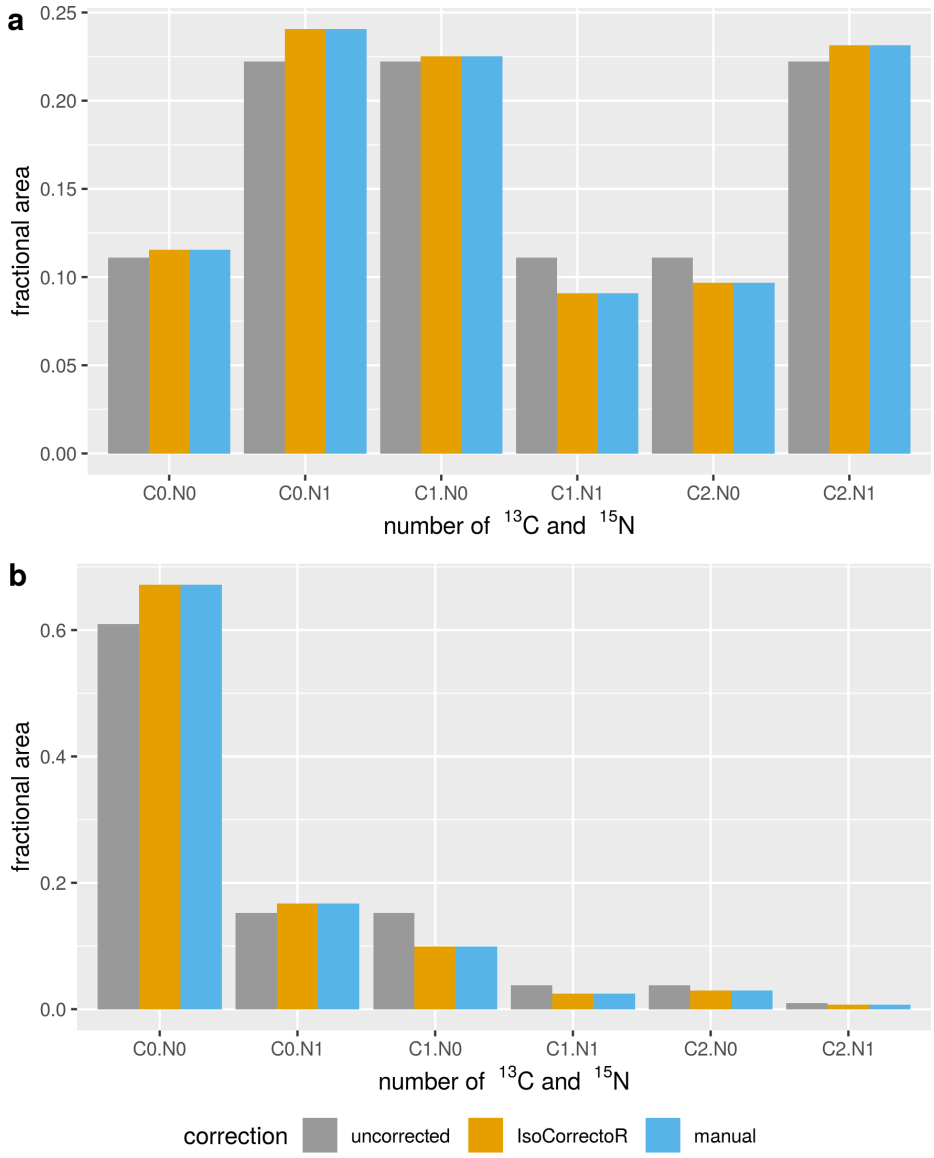



Figure A23: ***Comparison of UHR correction to manual calculations on simulated UHR example data of simultaneously $^{13}C$ and $^{15}N$ labeled PCF-derivatized glycine.*** *A tracer purity of 99% was assumed for correction. The x-axis labels Cx.Ny correspond to x $^{13}C$ label and y $^{15}N$ label incorporated in the respective isotopologue. Data are shown uncorrected, corrected with IsoCorrectoR's UHR correction algorithm and corrected through manual calculations. The panels a and b correspond to different uncorrected MIDs.*

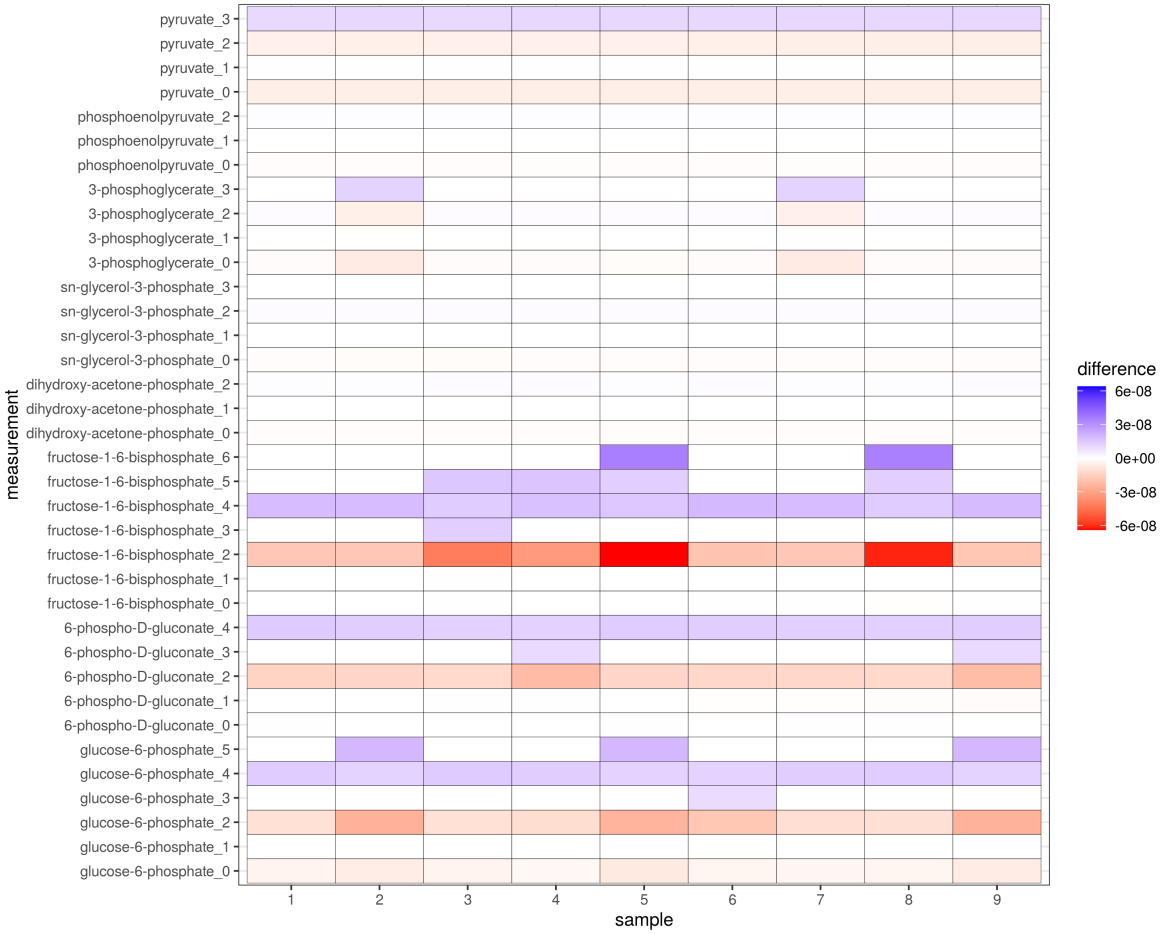



Figure A24: ***Heatmap showing differences between performing IsoCorrectoR's low-resolution MS correction with the default calculation threshold and without calculation threshold on data from a $^{13}C$ labeling experiment.*** *Data used are example data supplied with AccuCor (Su et al. (2017)). The default calculation threshold is 1E-08, a tracer purity of 0.99 was assumed. The heatmap was generated by subtracting MIDs corrected without threshold from MIDs corrected with the default threshold. Columns represent samples, while rows represent measurements of mass isotopomers. The number after the metabolite name corresponds to the number of label incorporated.*

## A2 $^{13}$C Tracer analysis

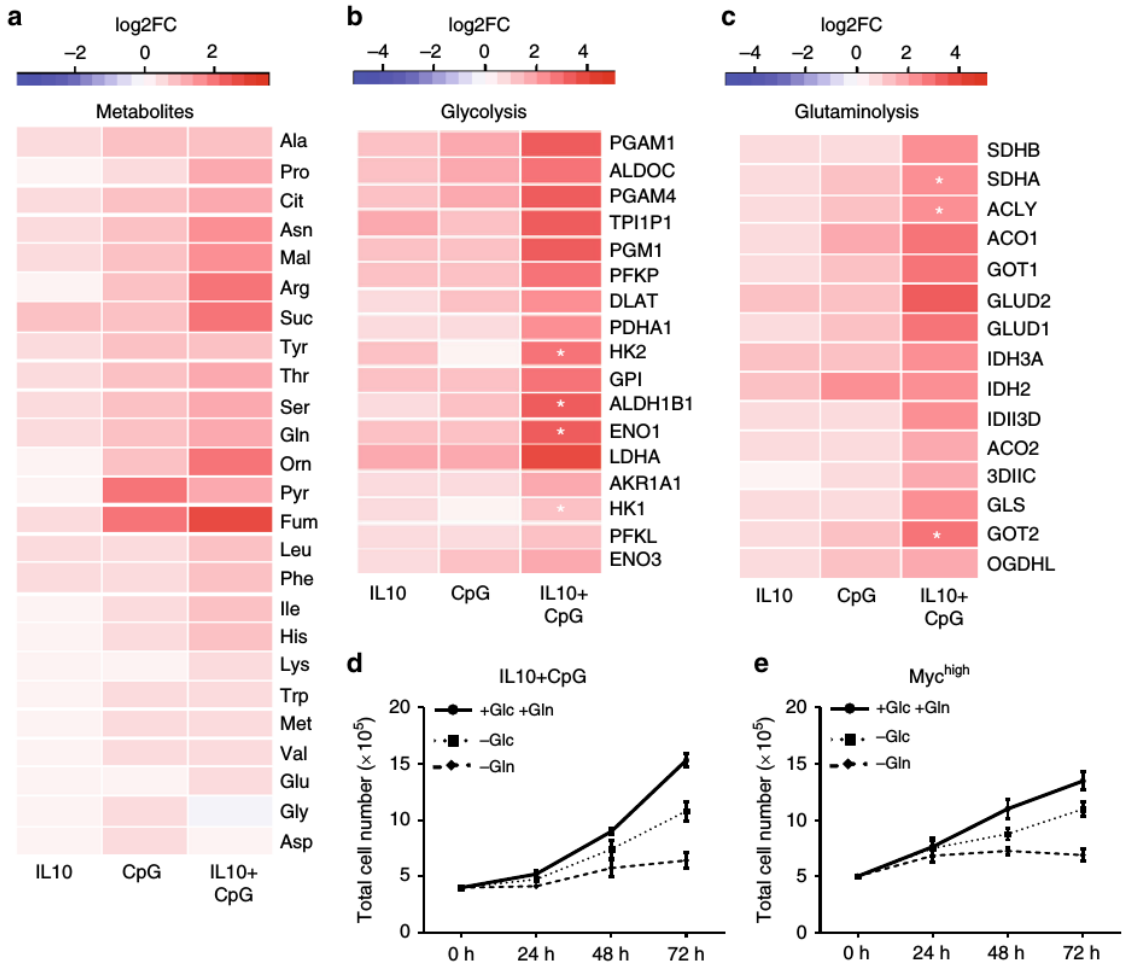



Figure A25: ***Activation of cell metabolism and global gene expression by combined stimulation of P493-6 cells with IL10 and CpG. a*** *Heatmap of changes in intracellular metabolite concentrations depicted as Log2FC after IL10 and/or CpG stimulation of P493-6 MYC-low cells in relation to unstimulated cells. Heat map of gene expression changes associated with* ***b*** *glycolysis (KEGG-term) and* ***c*** *glutaminolysis (KEGG-term), respectively, after IL10 and/or CpG stimulation of P493-6 MYC-low cells presented as Log2FC. Mean of three independent experiments is shown in all heat maps. Effects on gene expression were analyzed by linear regression and p-values for the IL10:CpG interaction terms were calculated (adjusted by Benjamini–Hochberg). Positive synergistic interactions are marked with a star (*$p < 0.05$). Relative cell counts of* ***d*** *IL10 + CpG-stimulated P493-6 MYC-low cells and* ***e*** *unstimulated P493-6 MYC-high cells grown in media with and without either glucose (–Glc) or glutamine (–Gln). Mean +/- SD of three independent experiments are given.* ***(Figure and legend taken from Feist et al. (2018))***

## A2.1 Assessment of isotopic steady state

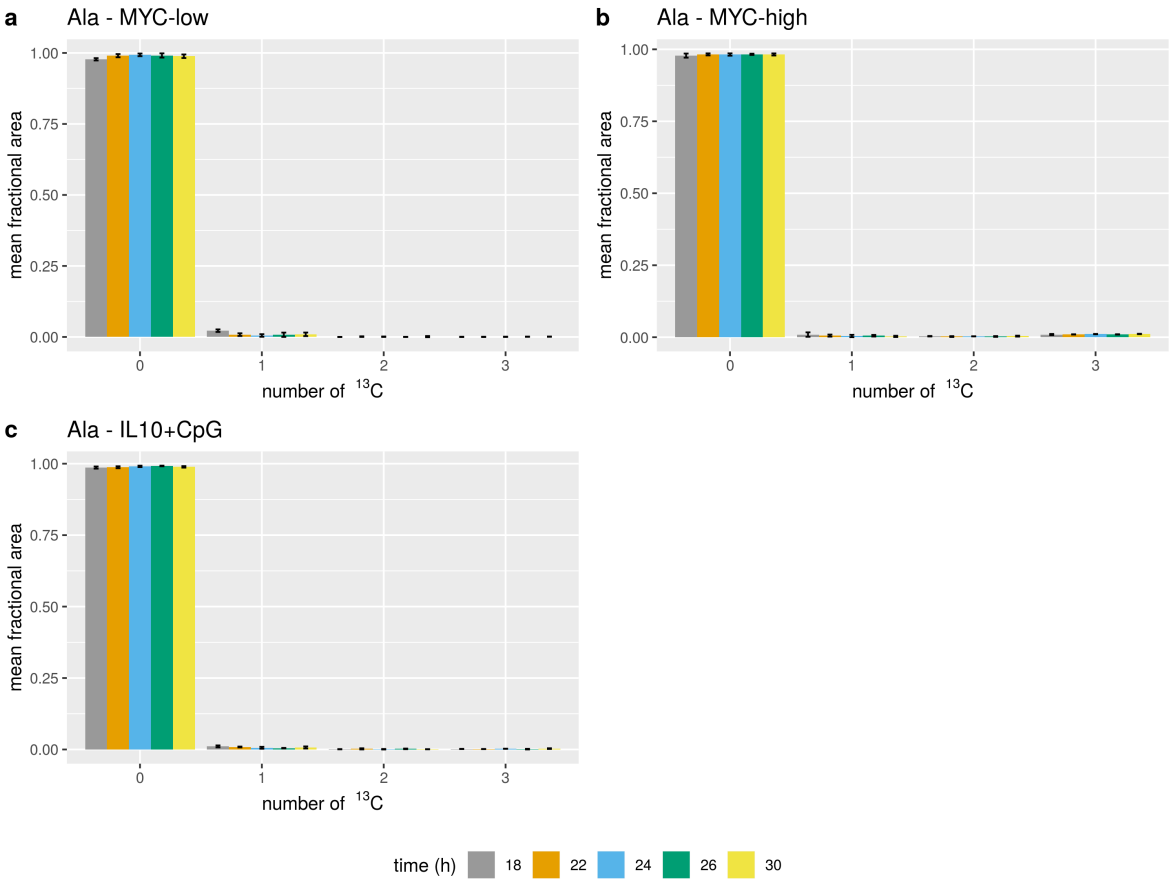



Figure A26: ***Assessment of isotopic steady state, $^{13}C$ tracer analysis with the tracer substrate U-$^{13}C$-glutamine in P493-6 cells.*** *The figure shows the MIDs of PCF-derivatized alanine from cell extracts for the different cell states MYC-low, MYC-high and IL10+CpG stimulation and for the different harvesting time points 18 h, 22 h, 24 h, 26 h and 30 h. Data were corrected for NA and tracer purity using IsoCorrectoR, assuming a tracer purity of 99%. The x-axis labels correspond to the number of $^{13}C$ label incorporated in the respective isotopologue. Samples were measured in (biological) triplicates, means of mass isotopomer fractions +/- SD are shown.*

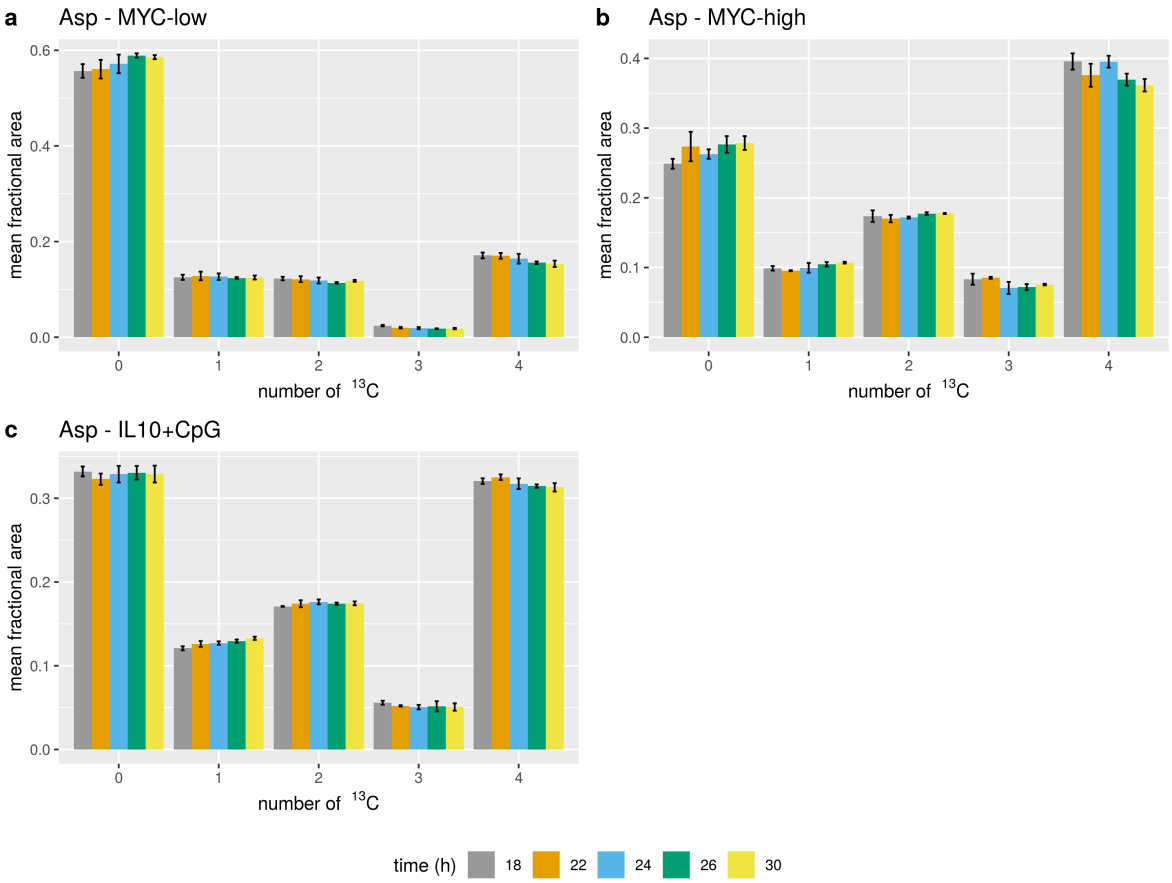



Figure A27: ***Assessment of isotopic steady state, $^{13}C$ tracer analysis with the tracer substrate U-$^{13}C$-glutamine in P493-6 cells.*** *The figure shows the MIDs of PCF-derivatized aspartate from cell extracts for the different cell states MYC-low, MYC-high and IL10+CpG stimulation and for the different harvesting time points 18 h, 22 h, 24 h, 26 h and 30 h. Data were corrected for NA and tracer purity using IsoCorrectoR, assuming a tracer purity of 99%. The x-axis labels correspond to the number of $^{13}C$ label incorporated in the respective isotopologue. Samples were measured in (biological) triplicates, means of mass isotopomer fractions +/- SD are shown.*

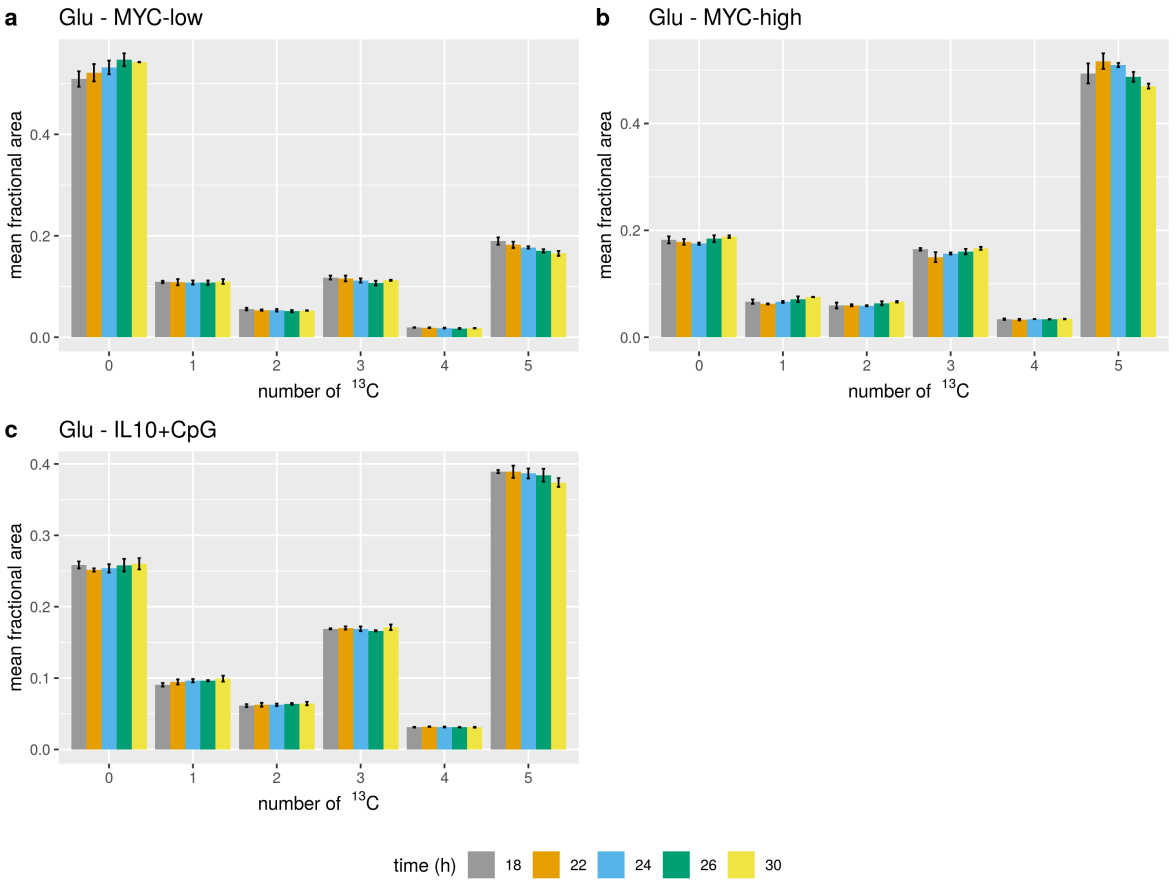



Figure A28: ***Assessment of isotopic steady state, $^{13}$C tracer analysis with the tracer substrate U-$^{13}$C-glutamine in P493-6 cells.*** *The figure shows the MIDs of PCF-derivatized glutamate from cell extracts for the different cell states MYC-low, MYC-high and IL10+CpG stimulation and for the different harvesting time points 18 h, 22 h, 24 h, 26 h and 30 h. Data were corrected for NA and tracer purity using IsoCorrectoR, assuming a tracer purity of 99%. The x-axis labels correspond to the number of $^{13}$C label incorporated in the respective isotopologue. Samples were measured in (biological) triplicates, means of mass isotopomer fractions +/- SD are shown.*

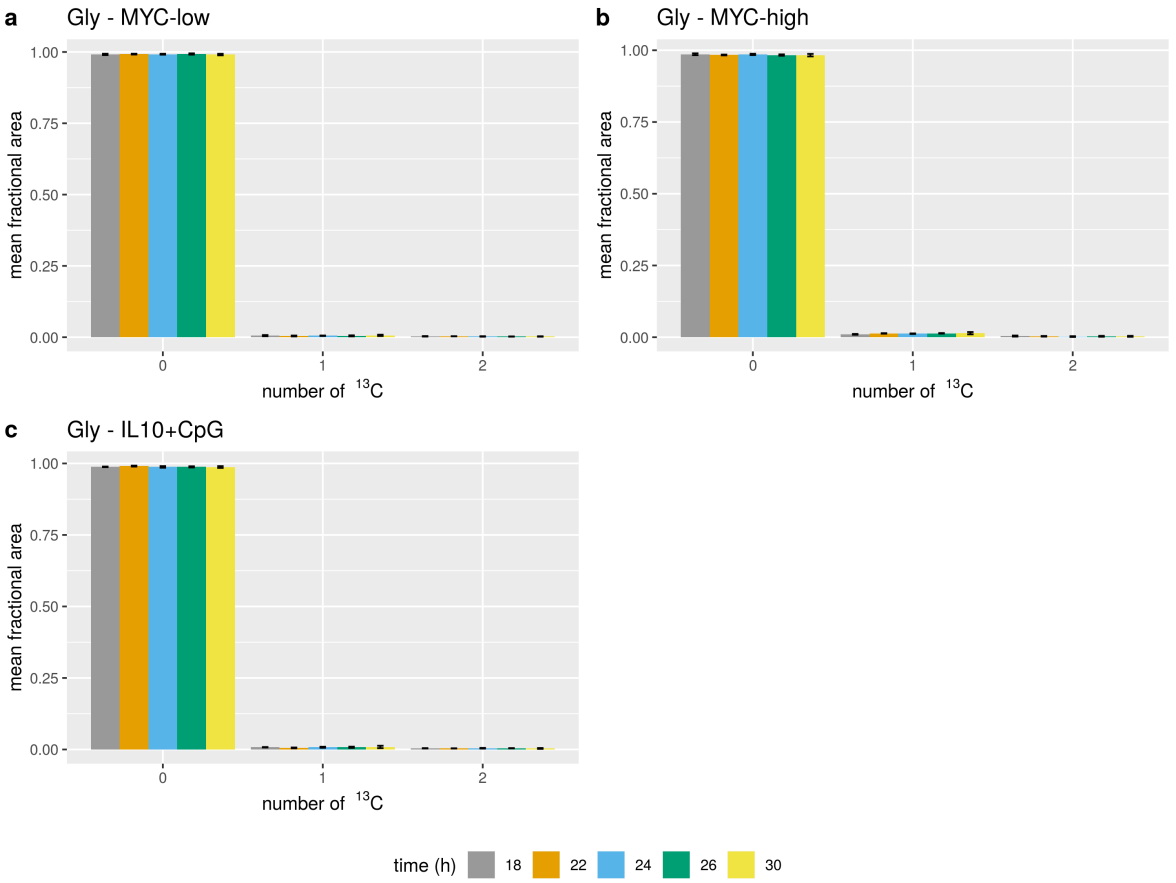



Figure A29: ***Assessment of isotopic steady state, $^{13}C$ tracer analysis with the tracer substrate U-$^{13}C$-glutamine in P493-6 cells.*** *The figure shows the MIDs of PCF-derivatized glycine from cell extracts for the different cell states MYC-low, MYC-high and IL10+CpG stimulation and for the different harvesting time points 18 h, 22 h, 24 h, 26 h and 30 h. Data were corrected for NA and tracer purity using IsoCorrectoR, assuming a tracer purity of 99%. The x-axis labels correspond to the number of $^{13}C$ label incorporated in the respective isotopologue. Samples were measured in (biological) triplicates, means of mass isotopomer fractions +/- SD are shown.*

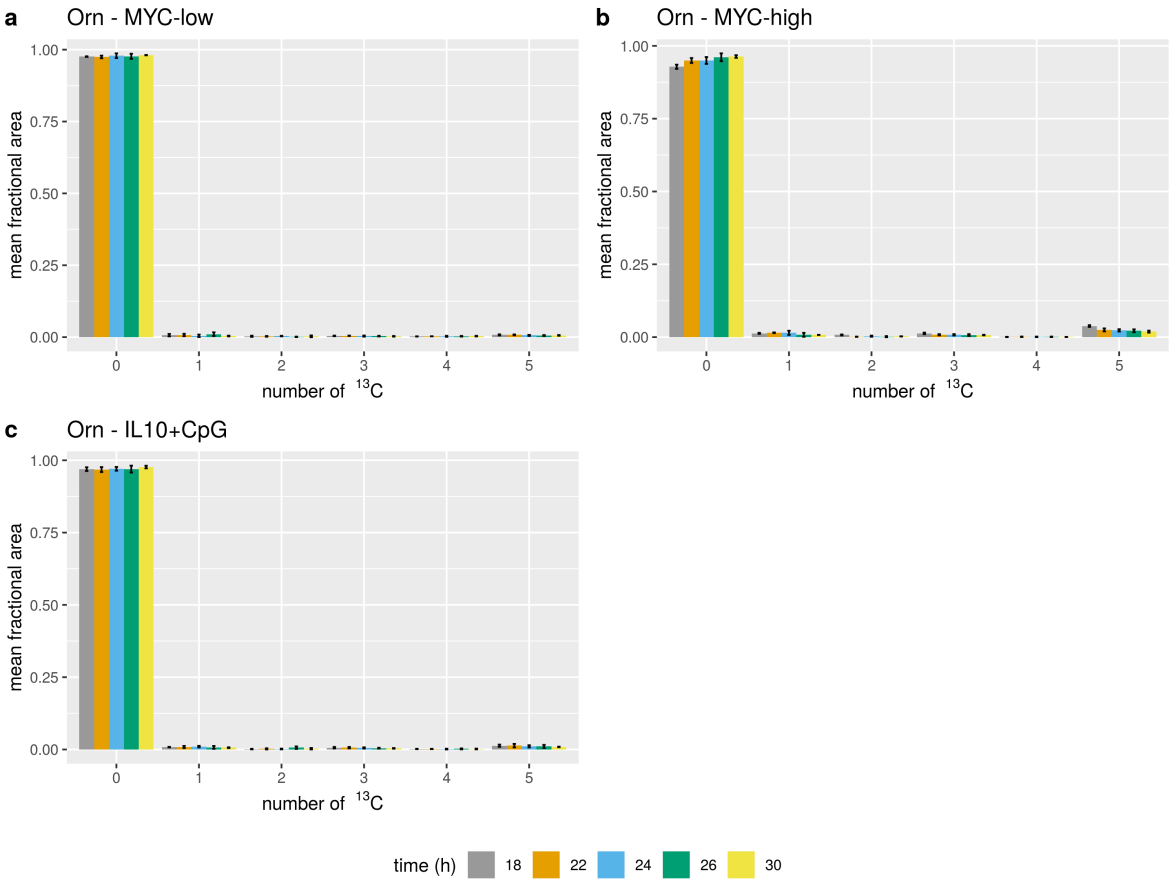



Figure A30: ***Assessment of isotopic steady state, $^{13}$C tracer analysis with the tracer substrate U-$^{13}$C-glutamine in P493-6 cells.*** *The figure shows the MIDs of PCF-derivatized ornithine from cell extracts for the different cell states MYC-low, MYC-high and IL10+CpG stimulation and for the different harvesting time points 18 h, 22 h, 24 h, 26 h and 30 h. Data were corrected for NA and tracer purity using IsoCorrectoR, assuming a tracer purity of 99%. The x-axis labels correspond to the number of $^{13}$C label incorporated in the respective isotopologue. Samples were measured in (biological) triplicates, means of mass isotopomer fractions +/- SD are shown.*

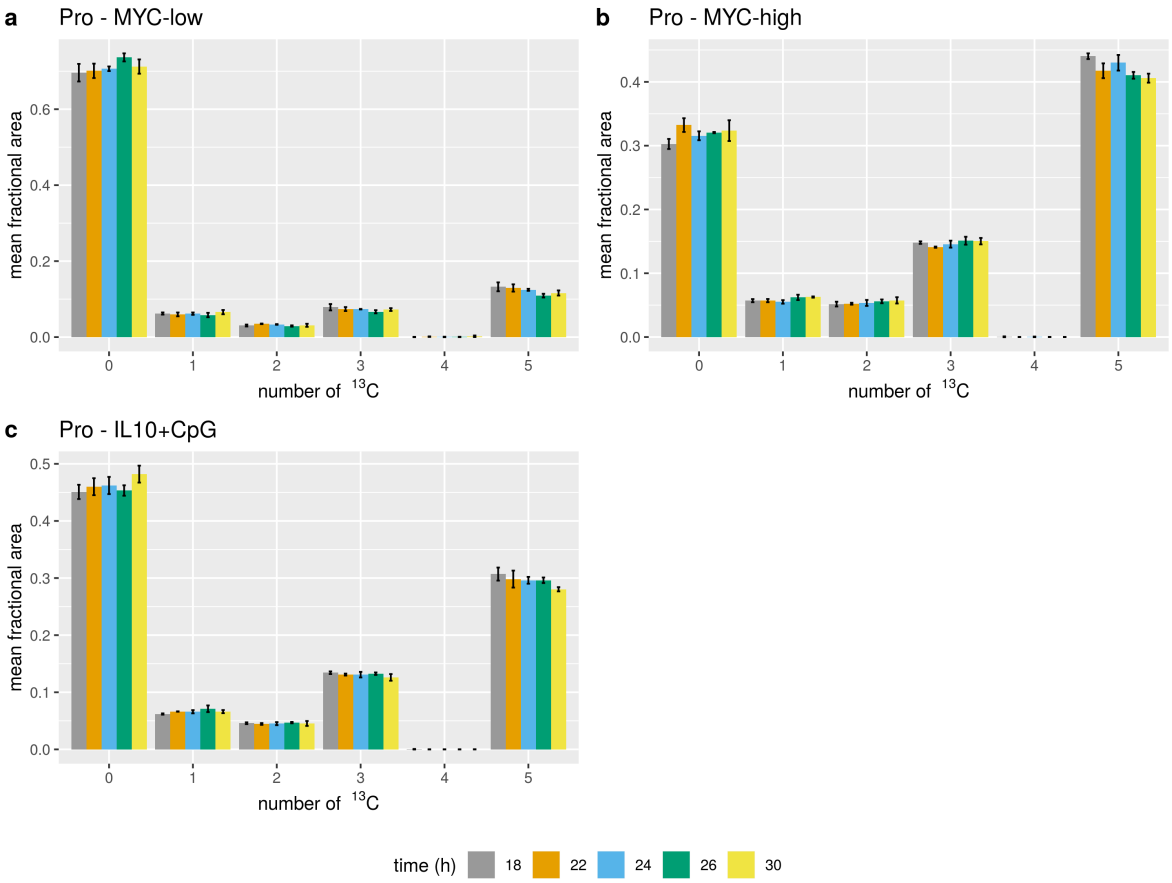



Figure A31: ***Assessment of isotopic steady state, $^{13}C$ tracer analysis with the tracer substrate U-$^{13}C$-glutamine in P493-6 cells.*** *The figure shows the MIDs of PCF-derivatized proline from cell extracts for the different cell states MYC-low, MYC-high and IL10+CpG stimulation and for the different harvesting time points 18 h, 22 h, 24 h, 26 h and 30 h. Data were corrected for NA and tracer purity using IsoCorrectoR, assuming a tracer purity of 99%. The x-axis labels correspond to the number of $^{13}C$ label incorporated in the respective isotopologue. Samples were measured in (biological) triplicates, means of mass isotopomer fractions +/- SD are shown.*

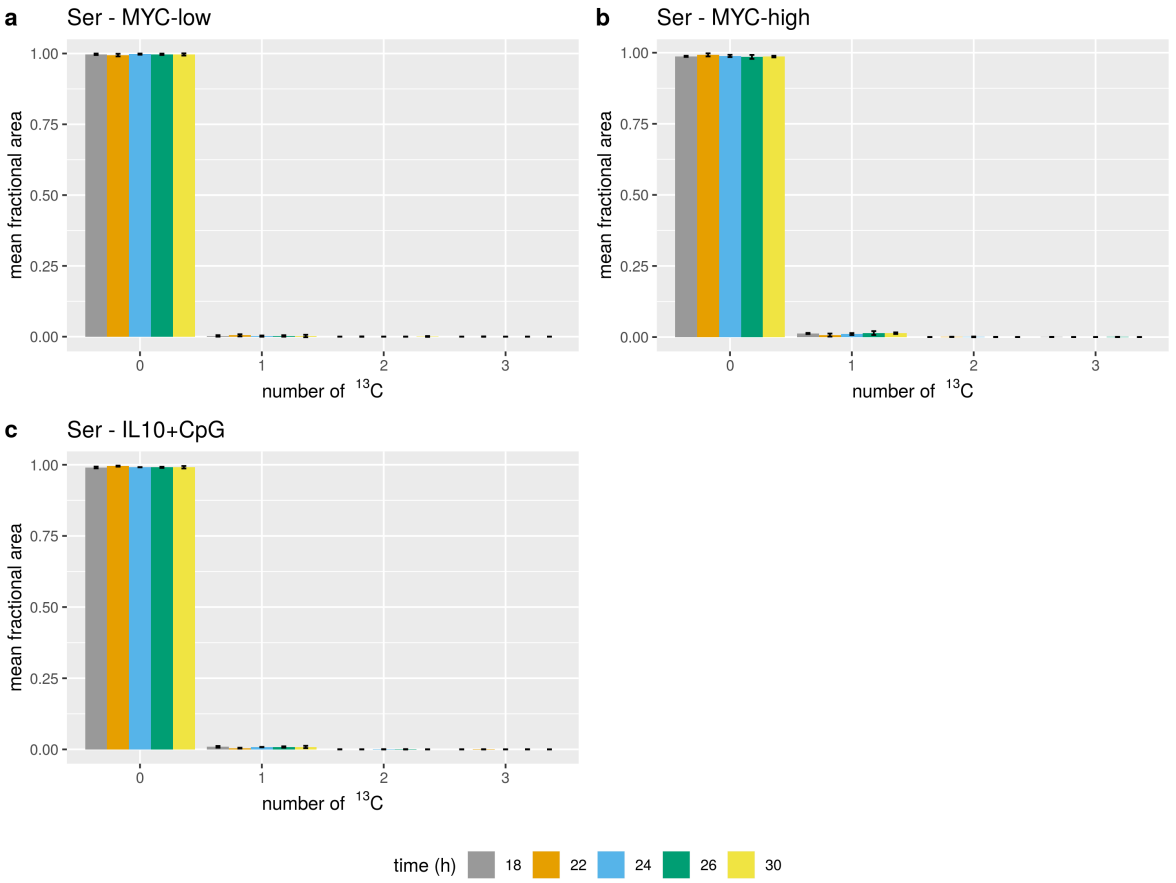



Figure A32: ***Assessment of isotopic steady state, $^{13}$C tracer analysis with the tracer substrate U-$^{13}$C-glutamine in P493-6 cells.*** *The figure shows the MIDs of PCF-derivatized serine from cell extracts for the different cell states MYC-low, MYC-high and IL10+CpG stimulation and for the different harvesting time points 18 h, 22 h, 24 h, 26 h and 30 h. Data were corrected for NA and tracer purity using IsoCorrectoR, assuming a tracer purity of 99%. The x-axis labels correspond to the number of $^{13}$C label incorporated in the respective isotopologue. Samples were measured in (biological) triplicates, means of mass isotopomer fractions +/- SD are shown.*

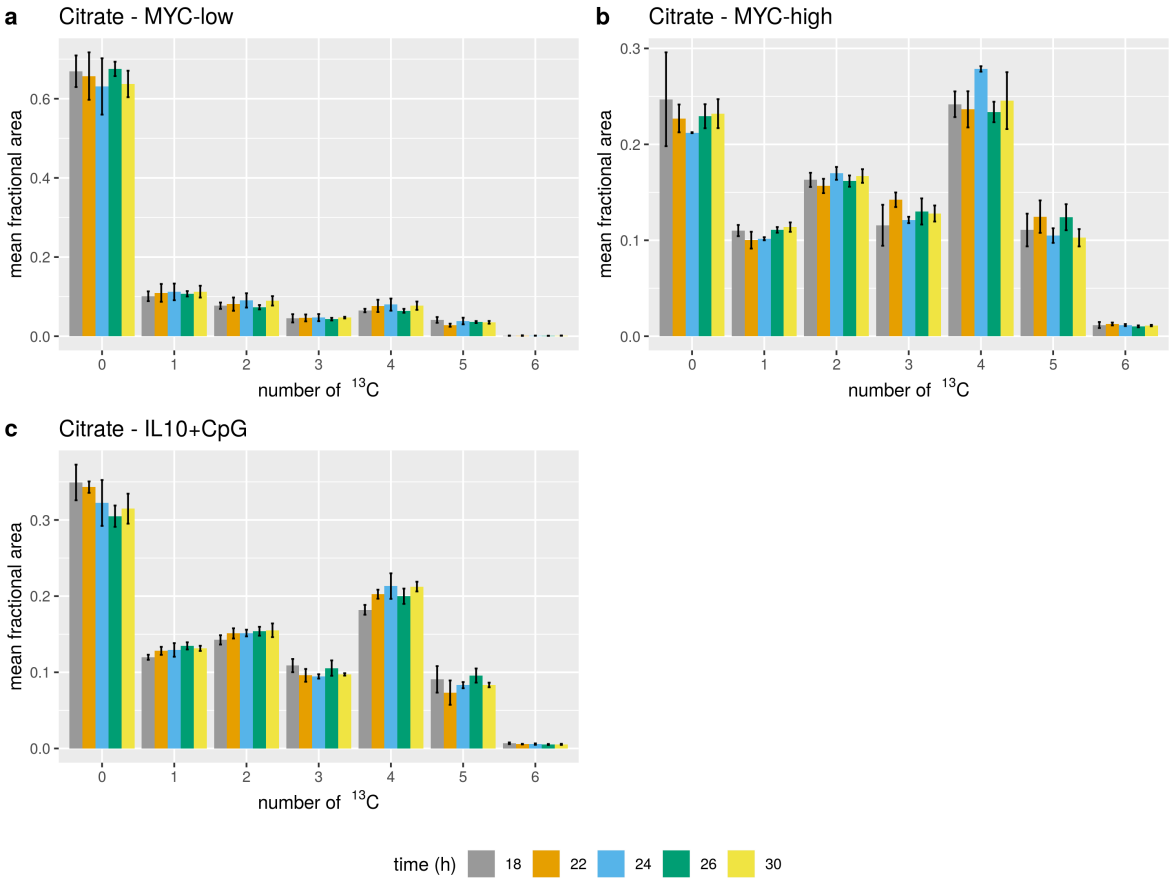



Figure A33: ***Assessment of isotopic steady state, $^{13}C$ tracer analysis with the tracer substrate U-$^{13}C$-glutamine in P493-6 cells.*** *The figure shows the MIDs of citrate from cell extracts for the different cell states MYC-low, MYC-high and IL10+CpG stimulation and for the different harvesting time points 18 h, 22 h, 24 h, 26 h and 30 h. Data were corrected for NA and tracer purity using IsoCorrectoR, assuming a tracer purity of 99%. The x-axis labels correspond to the number of $^{13}C$ label incorporated in the respective isotopologue. Samples were measured in (biological) triplicates, means of mass isotopomer fractions +/- SD are shown.*

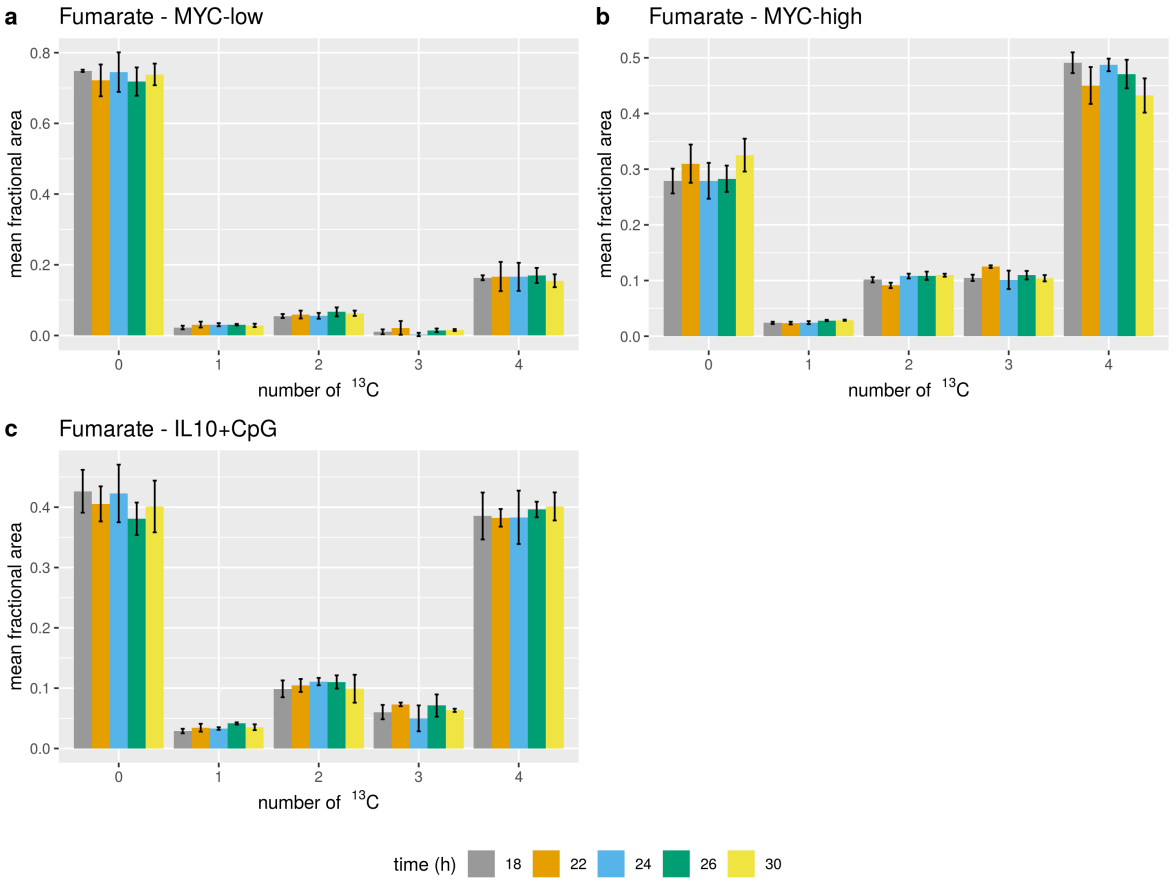



Figure A34: ***Assessment of isotopic steady state, $^{13}C$ tracer analysis with the tracer substrate U-$^{13}C$-glutamine in P493-6 cells.*** *The figure shows the MIDs of fumarate from cell extracts for the different cell states MYC-low, MYC-high and IL10+CpG stimulation and for the different harvesting time points 18 h, 22 h, 24 h, 26 h and 30 h. Data were corrected for NA and tracer purity using IsoCorrectoR, assuming a tracer purity of 99%. The x-axis labels correspond to the number of $^{13}C$ label incorporated in the respective isotopologue. Samples were measured in (biological) triplicates, means of mass isotopomer fractions +/- SD are shown.*

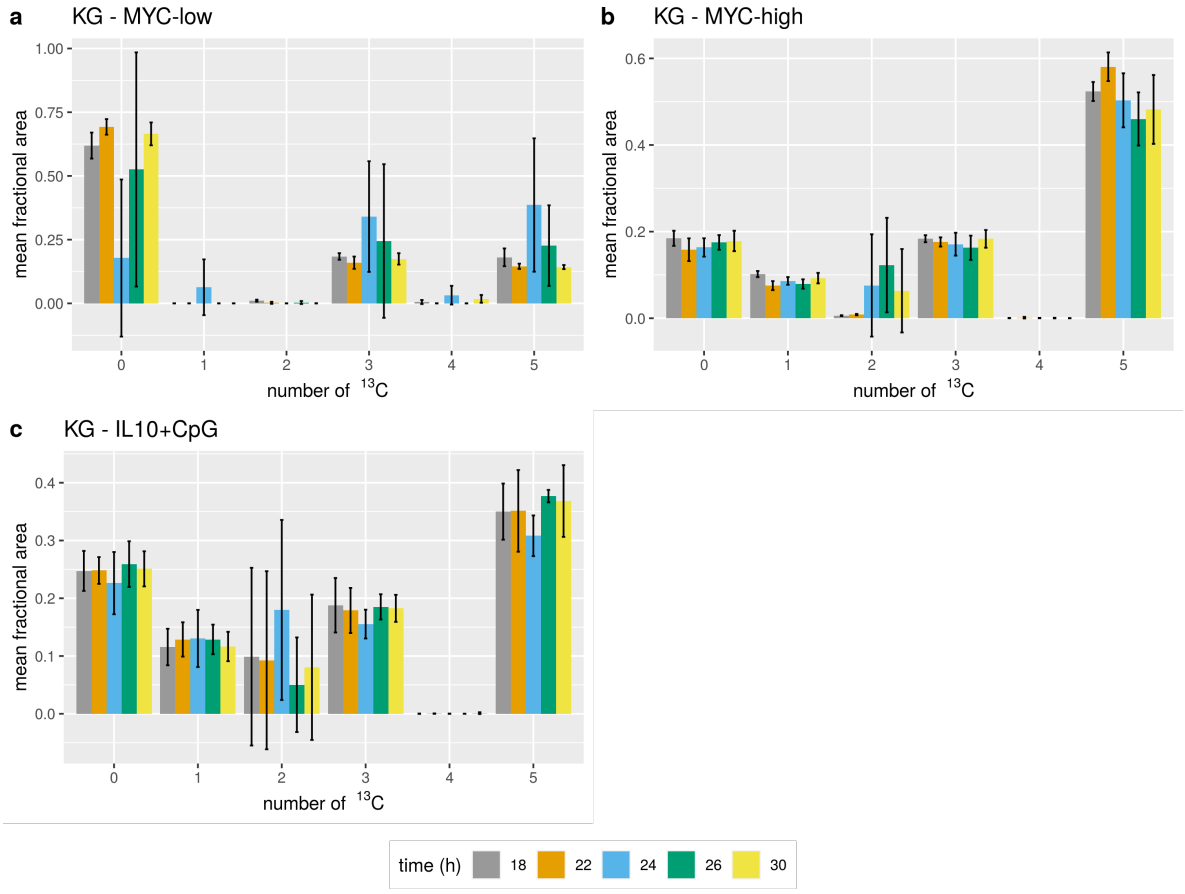



Figure A35: ***Assessment of isotopic steady state, $^{13}C$ tracer analysis with the tracer substrate U-$^{13}C$-glutamine in P493-6 cells.*** *The figure shows the MIDs of 2-ketoglutarate (KG) from cell extracts for the different cell states MYC-low, MYC-high and IL10+CpG stimulation and for the different harvesting time points 18 h, 22 h, 24 h, 26 h and 30 h. Data were corrected for NA and tracer purity using IsoCorrectoR, assuming a tracer purity of 99%. The x-axis labels correspond to the number of $^{13}C$ label incorporated in the respective isotopologue. Samples were measured in (biological) triplicates, means of mass isotopomer fractions +/- SD are shown.*

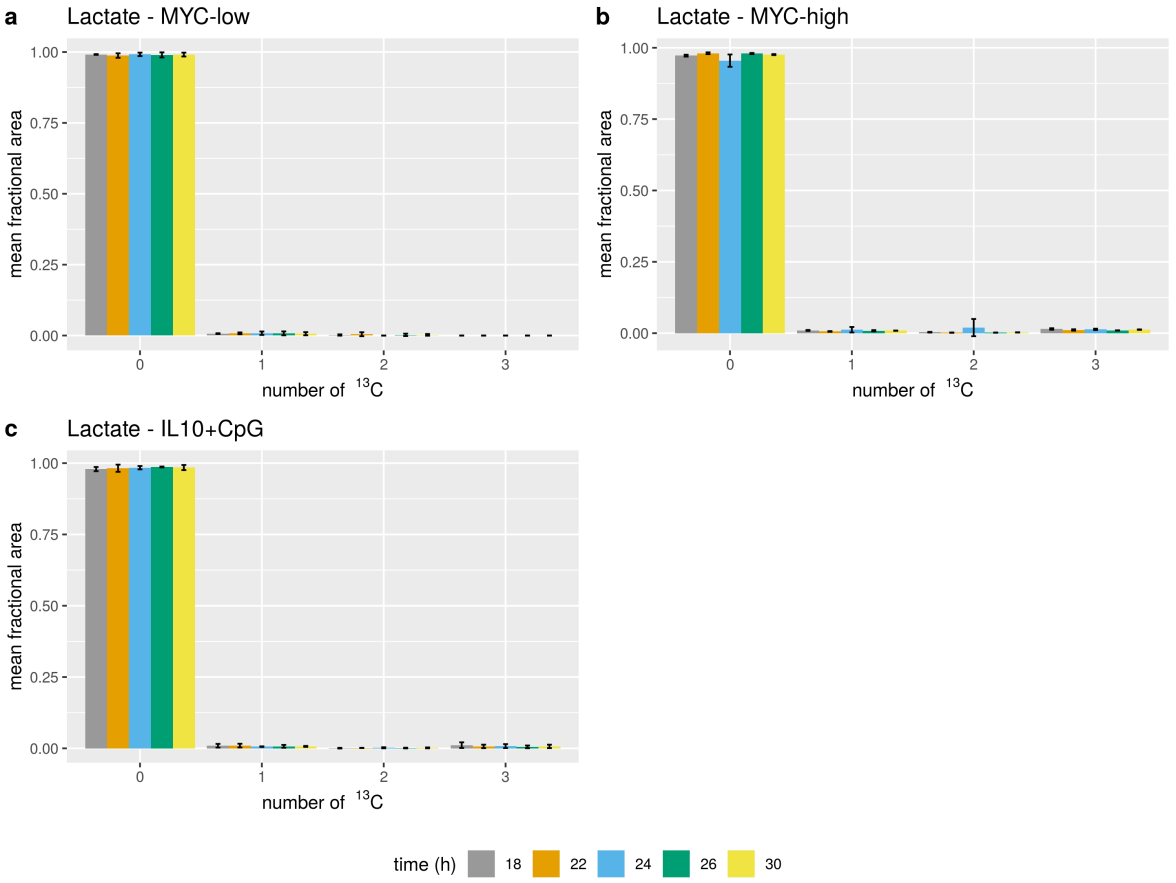



Figure A36: ***Assessment of isotopic steady state, $^{13}$C tracer analysis with the tracer substrate U-$^{13}$C-glutamine in P493-6 cells.*** *The figure shows the MIDs of lactate from cell extracts for the different cell states MYC-low, MYC-high and IL10+CpG stimulation and for the different harvesting time points 18 h, 22 h, 24 h, 26 h and 30 h. Data were corrected for NA and tracer purity using IsoCorrectoR, assuming a tracer purity of 99%. The x-axis labels correspond to the number of $^{13}$C label incorporated in the respective isotopologue. Samples were measured in (biological) triplicates, means of mass isotopomer fractions +/- SD are shown.*

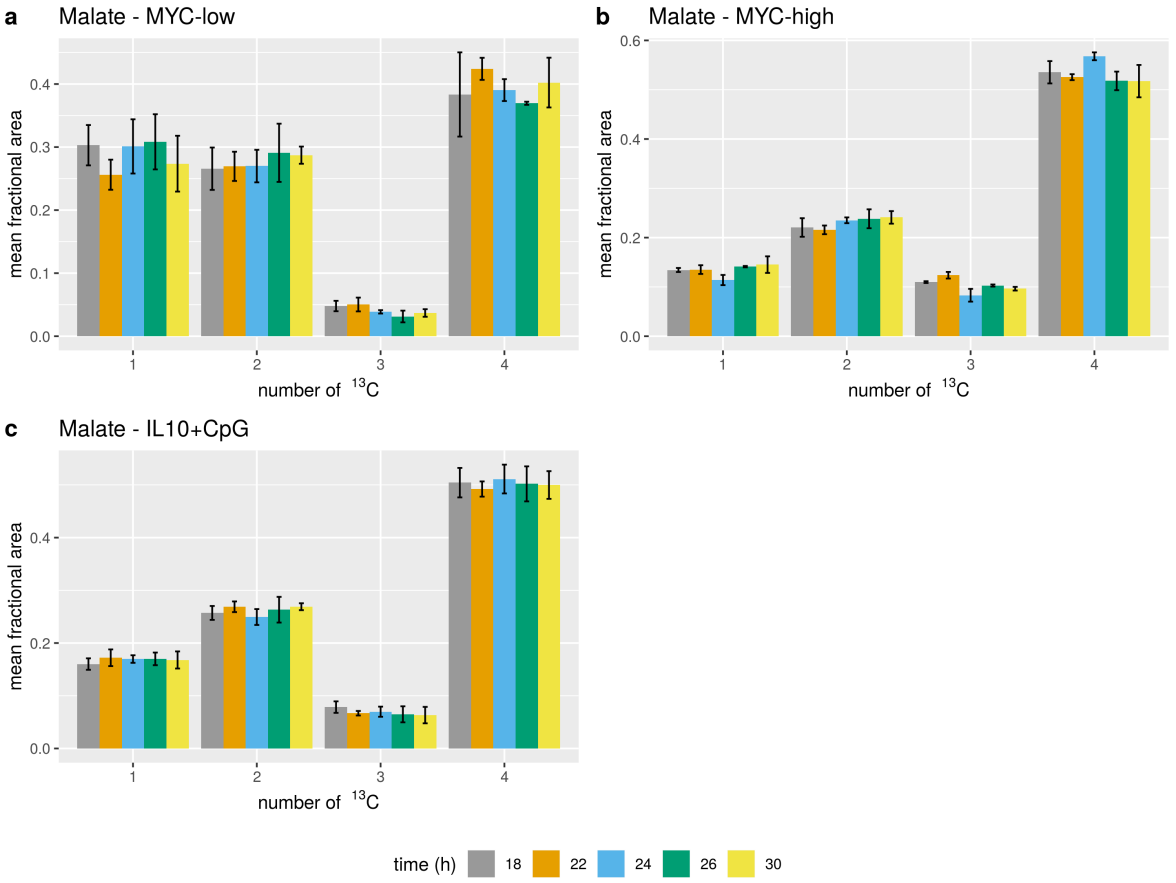



Figure A37: ***Assessment of isotopic steady state, $^{13}C$ tracer analysis with the tracer substrate U-$^{13}C$-glutamine in P493-6 cells.*** *The figure shows the MIDs of malate from cell extracts for the different cell states MYC-low, MYC-high and IL10+CpG stimulation and for the different harvesting time points 18 h, 22 h, 24 h, 26 h and 30 h. Data were corrected for NA and tracer purity using IsoCorrectoR, assuming a tracer purity of 99%. The x-axis labels correspond to the number of $^{13}C$ label incorporated in the respective isotopologue. Samples were measured in (biological) triplicates, means of mass isotopomer fractions +/- SD are shown.*

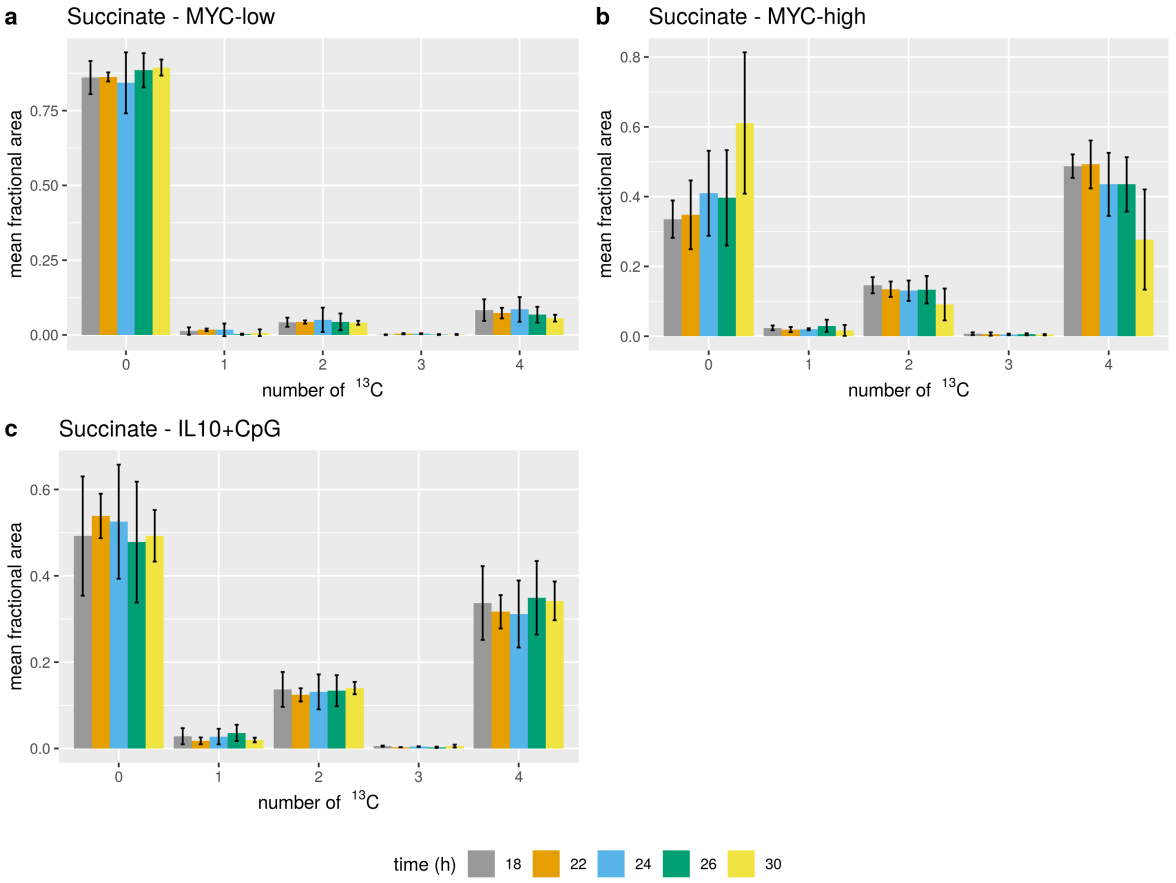



Figure A38: ***Assessment of isotopic steady state, $^{13}$C tracer analysis with the tracer substrate U-$^{13}$C-glutamine in P493-6 cells.*** *The figure shows the MIDs of succinate from cell extracts for the different cell states MYC-low, MYC-high and IL10+CpG stimulation and for the different harvesting time points 18 h, 22 h, 24 h, 26 h and 30 h. Data were corrected for NA and tracer purity using IsoCorrectoR, assuming a tracer purity of 99%. The x-axis labels correspond to the number of $^{13}$C label incorporated in the respective isotopologue. Samples were measured in (biological) triplicates, means of mass isotopomer fractions +/- SD are shown.*

| Cell state | Measurement | p-value | q-value | significant | Statistical test |
|---|---|---|---|---|---|
| MYC-low | Asp m+0 | 1.052E-01 | 2.135E-01 | no | anova welch |
| MYC-low | Asp m+1 | 9.293E-01 | 9.716E-01 | no | anova welch |
| MYC-low | Asp m+2 | 9.151E-02 | 1.973E-01 | no | anova welch |
| MYC-low | Asp m+3 | 2.325E-02 | 1.146E-01 | no | anova welch |
| MYC-low | Asp m+4 | 5.740E-02 | 1.584E-01 | no | anova welch |
| MYC-low | Glu m+0 | 7.095E-02 | 1.748E-01 | no | kruskal |
| MYC-low | Glu m+1 | 9.761E-01 | 9.935E-01 | no | anova welch |
| MYC-low | Glu m+2 | 4.912E-01 | 6.509E-01 | no | anova welch |
| MYC-low | Glu m+3 | 8.452E-02 | 1.881E-01 | no | kruskal |
| MYC-low | Glu m+4 | 1.310E-01 | 2.511E-01 | no | anova welch |
| MYC-low | Glu m+5 | 3.546E-02 | 1.225E-01 | no | anova welch |
| MYC-low | Orn m+0 | 1.970E-02 | 1.133E-01 | no | anova welch |
| MYC-low | Orn m+1 | 5.152E-01 | 6.509E-01 | no | anova welch |
| MYC-low | Orn m+2 | 1.640E-01 | 2.902E-01 | no | anova welch |
| MYC-low | Orn m+3 | 3.982E-01 | 5.846E-01 | no | anova welch |
| MYC-low | Orn m+4 | 6.264E-01 | 7.086E-01 | no | anova welch |
| MYC-low | Orn m+5 | 4.915E-01 | 6.509E-01 | no | anova welch |
| MYC-low | Pro m+0 | 1.021E-01 | 2.135E-01 | no | anova welch |
| MYC-low | Pro m+1 | 7.001E-01 | 7.668E-01 | no | anova welch |
| MYC-low | Pro m+2 | 4.077E-02 | 1.251E-01 | no | anova welch |
| MYC-low | Pro m+3 | 3.397E-01 | 5.382E-01 | no | anova welch |
| MYC-low | Pro m+4 | 4.798E-01 | 6.509E-01 | no | anova welch |
| MYC-low | Pro m+5 | 5.619E-02 | 1.584E-01 | no | anova welch |

Table A1: ***ANOVA tests to assess significant differences between time points for amino acid $^{13}C$ labeling states from a time course experiment in P493-6 cells, MYC-low cellular state.*** *"Measurement" corresponds to the labeling state for which the difference between time points was tested. The p-value is the p-value that results directly from the ANOVA test, the q-value corresponds to the FDR when correcting for multiple testing according to the Benjamini-Hochberg approach. The test was considered to be significant when* $q \leq 0.05$. *Depending on whether or not ANOVA model residuals were normally distributed, either the Welch-type ANOVA (anova welch) or the non-parametric Kruskal-Wallis ANOVA (kruskal) were used.*

| Cell state | Measurement | p-value | q-value | significant | Statistical test |
|---|---|---|---|---|---|
| MYC-high | Asp m+0 | 7.423E-02 | 1.766E-01 | no | anova welch |
| MYC-high | Asp m+1 | 1.300E-03 | 4.486E-02 | yes | anova welch |
| MYC-high | Asp m+2 | 2.926E-02 | 1.225E-01 | no | anova welch |
| MYC-high | Asp m+3 | 5.244E-03 | 7.237E-02 | no | anova welch |
| MYC-high | Asp m+4 | 4.170E-02 | 1.251E-01 | no | anova welch |
| MYC-high | Glu m+0 | 1.684E-02 | 1.127E-01 | no | anova welch |
| MYC-high | Glu m+1 | 1.726E-04 | 1.191E-02 | yes | anova welch |
| MYC-high | Glu m+2 | 1.446E-02 | 1.127E-01 | no | anova welch |
| MYC-high | Glu m+3 | 3.471E-02 | 1.225E-01 | no | anova welch |
| MYC-high | Glu m+4 | 7.382E-01 | 7.959E-01 | no | anova welch |
| MYC-high | Glu m+5 | 2.379E-03 | 5.473E-02 | no | anova welch |
| MYC-high | Orn m+0 | 1.703E-02 | 1.127E-01 | no | anova welch |
| MYC-high | Orn m+1 | 4.068E-03 | 7.018E-02 | no | anova welch |
| MYC-high | Orn m+2 | 2.302E-02 | 1.146E-01 | no | anova welch |
| MYC-high | Orn m+3 | 2.309E-01 | 3.887E-01 | no | anova welch |
| MYC-high | Orn m+4 | 8.566E-01 | 9.093E-01 | no | anova welch |
| MYC-high | Orn m+5 | 1.002E-02 | 9.876E-02 | no | anova welch |
| MYC-high | Pro m+0 | 1.279E-01 | 2.511E-01 | no | anova welch |
| MYC-high | Pro m+1 | 3.670E-02 | 1.225E-01 | no | anova welch |
| MYC-high | Pro m+2 | 4.290E-01 | 6.044E-01 | no | anova welch |
| MYC-high | Pro m+3 | 2.935E-02 | 1.225E-01 | no | anova welch |
| MYC-high | Pro m+4 | 9.935E-01 | 9.935E-01 | no | kruskal |
| MYC-high | Pro m+5 | 6.648E-03 | 7.645E-02 | no | anova welch |

Table A2: ***ANOVA tests to assess significant differences between time points for amino acid $^{13}C$ labeling states from a time course experiment in P493-6 cells, MYC-high cellular state.*** *"Measurement" corresponds to the labeling state for which the difference between time points was tested. The p-value is the p-value that results directly from the ANOVA test, the q-value corresponds to the FDR when correcting for multiple testing according to the Benjamini-Hochberg approach. The test was considered to be significant when* $q \leq 0.05$. *Depending on whether or not ANOVA model residuals were normally distributed, either the Welch-type ANOVA (anova welch) or the non-parametric Kruskal-Wallis ANOVA (kruskal) were used.*

| Cell state | Measurement | p-value | q-value | significant | Statistical test |
|---|---|---|---|---|---|
| IL10+CpG | Asp m+0 | 6.776E-01 | 7.541E-01 | no | anova welch |
| IL10+CpG | Asp m+1 | 1.796E-02 | 1.127E-01 | no | anova welch |
| IL10+CpG | Asp m+2 | 6.292E-02 | 1.635E-01 | no | anova welch |
| IL10+CpG | Asp m+3 | 3.531E-01 | 5.414E-01 | no | anova welch |
| IL10+CpG | Asp m+4 | 6.399E-02 | 1.635E-01 | no | anova welch |
| IL10+CpG | Glu m+0 | 3.432E-01 | 5.382E-01 | no | anova welch |
| IL10+CpG | Glu m+1 | 1.488E-01 | 2.702E-01 | no | anova welch |
| IL10+CpG | Glu m+2 | 5.658E-01 | 6.850E-01 | no | anova welch |
| IL10+CpG | Glu m+3 | 8.131E-02 | 1.870E-01 | no | anova welch |
| IL10+CpG | Glu m+4 | 2.124E-01 | 3.663E-01 | no | anova welch |
| IL10+CpG | Glu m+5 | 1.467E-01 | 2.702E-01 | no | anova welch |
| IL10+CpG | Orn m+0 | 5.777E-01 | 6.873E-01 | no | anova welch |
| IL10+CpG | Orn m+1 | 4.292E-01 | 6.044E-01 | no | anova welch |
| IL10+CpG | Orn m+2 | 5.972E-01 | 6.985E-01 | no | anova welch |
| IL10+CpG | Orn m+3 | 6.261E-01 | 7.086E-01 | no | anova welch |
| IL10+CpG | Orn m+4 | 9.882E-01 | 9.935E-01 | no | anova welch |
| IL10+CpG | Orn m+5 | 5.065E-01 | 6.509E-01 | no | anova welch |
| IL10+CpG | Pro m+0 | 2.787E-01 | 4.578E-01 | no | anova welch |
| IL10+CpG | Pro m+1 | 3.282E-02 | 1.225E-01 | no | anova welch |
| IL10+CpG | Pro m+2 | 5.580E-01 | 6.850E-01 | no | anova welch |
| IL10+CpG | Pro m+3 | 3.852E-01 | 5.778E-01 | no | anova welch |
| IL10+CpG | Pro m+4 | 5.188E-01 | 6.509E-01 | no | kruskal |
| IL10+CpG | Pro m+5 | 3.728E-02 | 1.225E-01 | no | anova welch |

Table A3: ***ANOVA tests to assess significant differences between time points for amino acid $^{13}C$ labeling states from a time course experiment in P493-6 cells, IL10+CpG stimulated MYC-low cellular state.*** *"Measurement" corresponds to the labeling state for which the difference between time points was tested. The p-value is the p-value that results directly from the ANOVA test, the q-value corresponds to the FDR when correcting for multiple testing according to the Benjamini-Hochberg approach. The test was considered to be significant when* $q \leq 0.05$. *Depending on whether or not ANOVA model residuals were normally distributed, either the Welch-type ANOVA (anova welch) or the non-parametric Kruskal-Wallis ANOVA (kruskal) were used.*

| Cell state | Measurement | p-value | q-value | significant | Statistical test |
|---|---|---|---|---|---|
| MYC-low | KG m+0 | 2.113E-01 | 5.956E-01 | no | anova welch |
| MYC-low | KG m+1 | 4.060E-01 | 7.840E-01 | no | kruskal |
| MYC-low | KG m+2 | NA | NA | NA | anova welch |
| MYC-low | KG m+3 | 6.929E-01 | 9.084E-01 | no | kruskal |
| MYC-low | KG m+4 | 1.031E-01 | 4.712E-01 | no | kruskal |
| MYC-low | KG m+5 | 6.566E-01 | 9.084E-01 | no | kruskal |
| MYC-low | Citrate m+0 | 6.441E-01 | 9.084E-01 | no | anova welch |
| MYC-low | Citrate m+1 | 9.054E-01 | 9.531E-01 | no | anova welch |
| MYC-low | Citrate m+2 | 4.410E-01 | 7.840E-01 | no | anova welch |
| MYC-low | Citrate m+3 | 7.267E-01 | 9.084E-01 | no | anova welch |
| MYC-low | Citrate m+4 | 3.725E-01 | 7.840E-01 | no | anova welch |
| MYC-low | Citrate m+5 | 1.970E-01 | 5.956E-01 | no | anova welch |
| MYC-low | Citrate m+6 | 3.791E-01 | 7.840E-01 | no | anova welch |
| MYC-low | Fumarate m+0 | 7.282E-01 | 9.084E-01 | no | anova welch |
| MYC-low | Fumarate m+1 | 4.312E-01 | 7.840E-01 | no | anova welch |
| MYC-low | Fumarate m+2 | 6.455E-01 | 9.084E-01 | no | anova welch |
| MYC-low | Fumarate m+3 | 1.403E-01 | 5.612E-01 | no | anova welch |
| MYC-low | Fumarate m+4 | 9.451E-01 | 9.802E-01 | no | anova welch |
| MYC-low | Malate m+1 | 4.264E-01 | 7.840E-01 | no | anova welch |
| MYC-low | Malate m+2 | 7.631E-01 | 9.194E-01 | no | anova welch |
| MYC-low | Malate m+3 | 3.172E-01 | 7.624E-01 | no | anova welch |
| MYC-low | Malate m+4 | 5.694E-02 | 4.181E-01 | no | anova welch |
| MYC-low | Succinate m+0 | 6.659E-01 | 9.084E-01 | no | anova welch |
| MYC-low | Succinate m+1 | 5.749E-02 | 4.181E-01 | no | anova welch |
| MYC-low | Succinate m+2 | 9.907E-01 | 9.907E-01 | no | anova welch |
| MYC-low | Succinate m+3 | 2.112E-01 | 5.956E-01 | no | anova welch |
| MYC-low | Succinate m+4 | 6.325E-01 | 9.084E-01 | no | anova welch |

Table A4: ***ANOVA tests to assess significant differences between time points for organic acid $^{13}C$ labeling states from a time course experiment in P493-6 cells, MYC-low cellular state.*** *"Measurement" corresponds to the labeling state for which the difference between time points was tested. The p-value is the p-value that results directly from the ANOVA test, the q-value corresponds to the FDR when correcting for multiple testing according to the Benjamini-Hochberg approach. The test was considered to be significant when* $q \leq 0.05$. *Depending on whether or not ANOVA model residuals were normally distributed, either the Welch-type ANOVA (anova welch) or the non-parametric Kruskal-Wallis ANOVA (kruskal) were used. ANOVA on KG m+2 could not be performed because of too many zeros in the data. KG: 2-ketoglutarate.*

| Cell state | Measurement | p-value | q-value | significant | Statistical test |
|---|---|---|---|---|---|
| MYC-high | KG m+0 | 7.128E-01 | 9.084E-01 | no | anova welch |
| MYC-high | KG m+1 | 1.121E-01 | 4.721E-01 | no | anova welch |
| MYC-high | KG m+2 | 2.748E-01 | 6.871E-01 | no | anova welch |
| MYC-high | KG m+3 | 7.700E-01 | 9.194E-01 | no | anova welch |
| MYC-high | KG m+4 | 7.320E-01 | 9.084E-01 | no | anova welch |
| MYC-high | KG m+5 | 2.051E-01 | 5.956E-01 | no | anova welch |
| MYC-high | Citrate m+0 | 1.912E-01 | 5.956E-01 | no | anova welch |
| MYC-high | Citrate m+1 | 3.648E-02 | 3.648E-01 | no | anova welch |
| MYC-high | Citrate m+2 | 2.642E-01 | 6.846E-01 | no | kruskal |
| MYC-high | Citrate m+3 | 1.060E-01 | 4.712E-01 | no | anova welch |
| MYC-high | Citrate m+4 | 1.060E-02 | 2.922E-01 | no | anova welch |
| MYC-high | Citrate m+5 | 3.240E-01 | 7.624E-01 | no | anova welch |
| MYC-high | Citrate m+6 | 4.228E-01 | 7.840E-01 | no | anova welch |
| MYC-high | Fumarate m+0 | 4.202E-01 | 7.840E-01 | no | anova welch |
| MYC-high | Fumarate m+1 | 6.508E-02 | 4.338E-01 | no | anova welch |
| MYC-high | Fumarate m+2 | 3.548E-02 | 3.648E-01 | no | anova welch |
| MYC-high | Fumarate m+3 | 1.257E-02 | 2.922E-01 | no | anova welch |
| MYC-high | Fumarate m+4 | 2.159E-01 | 5.956E-01 | no | anova welch |
| MYC-high | Malate m+1 | 7.401E-02 | 4.413E-01 | no | anova welch |
| MYC-high | Malate m+2 | 1.874E-01 | 5.956E-01 | no | anova welch |
| MYC-high | Malate m+3 | 9.972E-03 | 2.922E-01 | no | anova welch |
| MYC-high | Malate m+4 | 1.461E-02 | 2.922E-01 | no | anova welch |
| MYC-high | Succinate m+0 | 4.696E-01 | 8.154E-01 | no | anova welch |
| MYC-high | Succinate m+1 | 8.299E-01 | 9.352E-01 | no | anova welch |
| MYC-high | Succinate m+2 | 6.660E-01 | 9.084E-01 | no | anova welch |
| MYC-high | Succinate m+3 | 8.484E-01 | 9.426E-01 | no | anova welch |
| MYC-high | Succinate m+4 | 3.811E-01 | 7.840E-01 | no | anova welch |

Table A5: ***ANOVA tests to assess significant differences between time points for organic acid $^{13}C$ labeling states from a time course experiment in P493-6 cells, MYC-high cellular state.*** *"Measurement" corresponds to the labeling state for which the difference between time points was tested. The p-value is the p-value that results directly from the ANOVA test, the q-value corresponds to the FDR when correcting for multiple testing according to the Benjamini-Hochberg approach. The test was considered to be significant when* $q \leq 0.05$. *Depending on whether or not ANOVA model residuals were normally distributed, either the Welch-type ANOVA (anova welch) or the non-parametric Kruskal-Wallis ANOVA (kruskal) were used. KG: 2-ketoglutarate.*

| Cell state | Measurement | p-value | q-value | significant | Statistical test |
|---|---|---|---|---|---|
| IL10+CpG | KG m+0 | 9.627E-01 | 9.802E-01 | no | anova welch |
| IL10+CpG | KG m+1 | 9.680E-01 | 9.802E-01 | no | anova welch |
| IL10+CpG | KG m+2 | 6.295E-01 | 9.084E-01 | no | kruskal |
| IL10+CpG | KG m+3 | 6.566E-01 | 9.084E-01 | no | kruskal |
| IL10+CpG | KG m+4 | 1.964E-01 | 5.956E-01 | no | kruskal |
| IL10+CpG | KG m+5 | 2.653E-01 | 6.846E-01 | no | anova welch |
| IL10+CpG | Citrate m+0 | 9.383E-02 | 4.712E-01 | no | anova welch |
| IL10+CpG | Citrate m+1 | 5.729E-02 | 4.181E-01 | no | anova welch |
| IL10+CpG | Citrate m+2 | 3.935E-01 | 7.840E-01 | no | anova welch |
| IL10+CpG | Citrate m+3 | 1.967E-01 | 5.956E-01 | no | kruskal |
| IL10+CpG | Citrate m+4 | 3.192E-02 | 3.648E-01 | no | anova welch |
| IL10+CpG | Citrate m+5 | 4.197E-01 | 7.840E-01 | no | anova welch |
| IL10+CpG | Citrate m+6 | 5.417E-01 | 9.028E-01 | no | anova welch |
| IL10+CpG | Fumarate m+0 | 6.284E-01 | 9.084E-01 | no | anova welch |
| IL10+CpG | Fumarate m+1 | 2.631E-02 | 3.648E-01 | no | anova welch |
| IL10+CpG | Fumarate m+2 | 7.381E-01 | 9.084E-01 | no | anova welch |
| IL10+CpG | Fumarate m+3 | 1.015E-01 | 4.712E-01 | no | anova welch |
| IL10+CpG | Fumarate m+4 | 8.027E-01 | 9.308E-01 | no | anova welch |
| IL10+CpG | Malate m+1 | 8.300E-01 | 9.352E-01 | no | anova welch |
| IL10+CpG | Malate m+2 | 4.791E-01 | 8.154E-01 | no | anova welch |
| IL10+CpG | Malate m+3 | 6.792E-01 | 9.084E-01 | no | anova welch |
| IL10+CpG | Malate m+4 | 9.009E-01 | 9.531E-01 | no | anova welch |
| IL10+CpG | Succinate m+0 | 9.035E-01 | 9.531E-01 | no | anova welch |
| IL10+CpG | Succinate m+1 | 6.737E-01 | 9.084E-01 | no | anova welch |
| IL10+CpG | Succinate m+2 | 8.613E-01 | 9.439E-01 | no | anova welch |
| IL10+CpG | Succinate m+3 | 7.723E-02 | 4.413E-01 | no | anova welch |
| IL10+CpG | Succinate m+4 | 8.028E-01 | 9.308E-01 | no | kruskal |

Table A6: ***ANOVA tests to assess significant differences between time points for organic acid $^{13}C$ labeling states from a time course experiment in P493-6 cells, IL10+CpG stimulated MYC-low cellular state.*** *"Measurement" corresponds to the labeling state for which the difference between time points was tested. The p-value is the p-value that results directly from the ANOVA test, the q-value corresponds to the FDR when correcting for multiple testing according to the Benjamini-Hochberg approach. The test was considered to be significant when* $q \leq 0.05$. *Depending on whether or not ANOVA model residuals were normally distributed, either the Welch-type ANOVA (anova welch) or the non-parametric Kruskal-Wallis ANOVA (kruskal) were used. KG: 2-ketoglutarate.*

## A2.2 Analysis of label incorporation

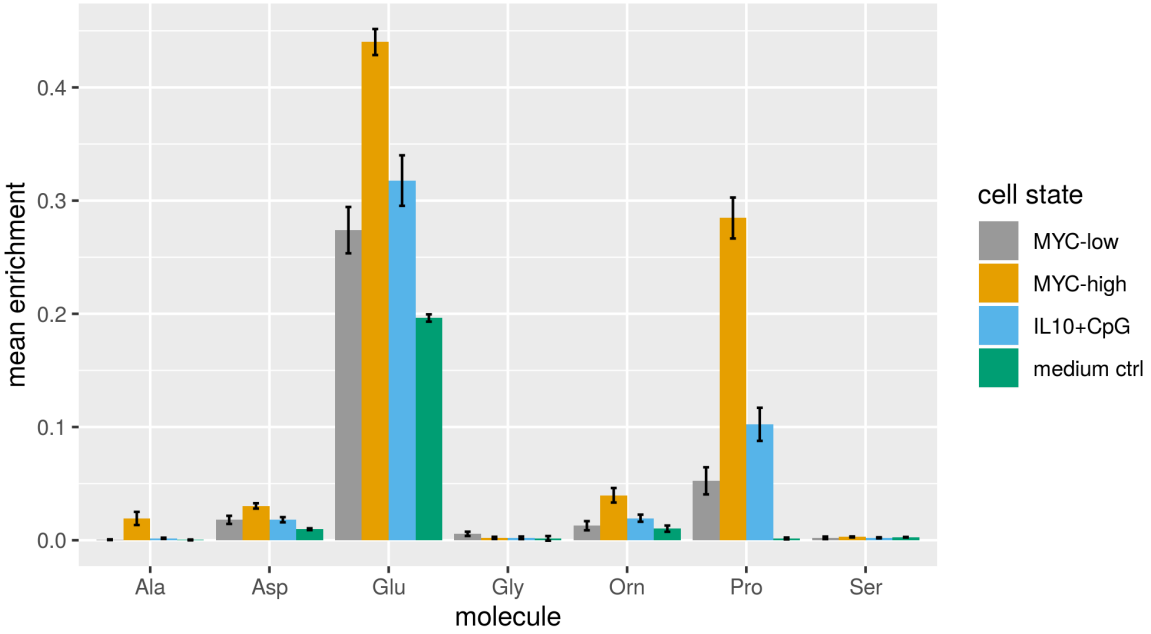



Figure A39: ***$^{13}C$ tracer analysis with the tracer substrate U-$^{13}C$-glutamine in P493-6 cells, mean $^{13}C$ enrichment of amino acids in cell supernatants.*** *The figure shows the mean $^{13}C$ enrichment of PCF-derivatized amino acids from cell supernatants and the medium control for the different cell states MYC-low, MYC-high and IL10+CpG stimulation. Cells were harvested after 24 h of incubation with the tracer substrate. Data were corrected for NA and tracer purity using IsoCorrectoR, assuming a tracer purity of 99%. Samples were measured in (biological) triplicates, means of mean enrichment among replicates +/- SD are shown.*

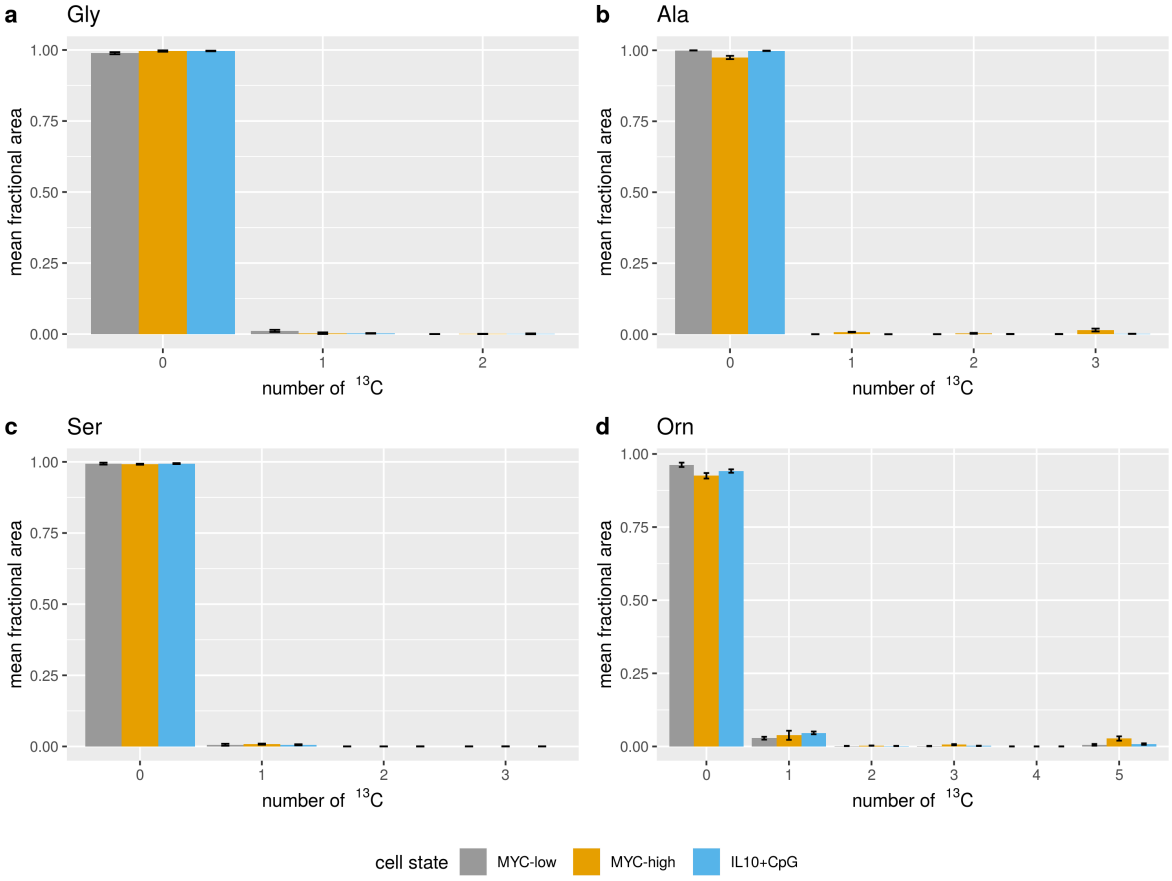



Figure A40: ***$^{13}$C tracer analysis with the tracer substrate U-$^{13}$C-glutamine in P493-6 cells, labeling of amino acids in cell supernatants.*** *The figure shows the MIDs of PCF-derivatized amino acids from cell supernatants for the different cell states MYC-low, MYC-high and IL10+CpG stimulation. Cells were harvested after 24 h of incubation with the tracer substrate. Data were corrected for NA and tracer purity using IsoCorrectoR, assuming a tracer purity of 99%. The x-axis labels correspond to the number of $^{13}$C label incorporated in the respective isotopologue. Samples were measured in (biological) triplicates, means of isotopologue fractions +/- SD are shown.*

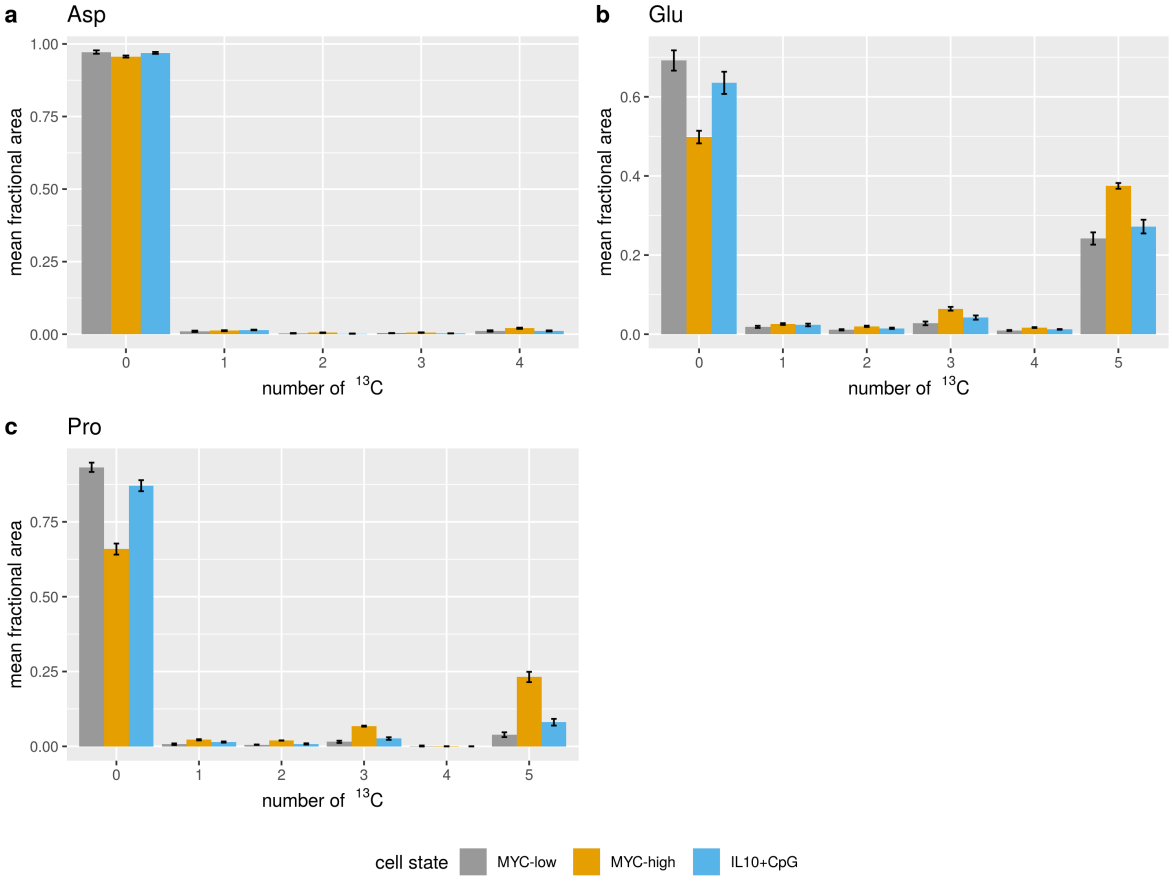



Figure A41: ***$^{13}C$ tracer analysis with the tracer substrate U-$^{13}C$-glutamine in P493-6 cells, labeling of amino acids in cell supernatants.*** *The figure shows the MIDs of PCF-derivatized amino acids from cell supernatants for the different cell states MYC-low, MYC-high and IL10+CpG stimulation. Cells were harvested after 24 h of incubation with the tracer substrate. Data were corrected for NA and tracer purity using IsoCorrectoR, assuming a tracer purity of 99%. The x-axis labels correspond to the number of $^{13}C$ label incorporated in the respective isotopologue. Samples were measured in (biological) triplicates, means of isotopologue fractions +/- SD are shown.*

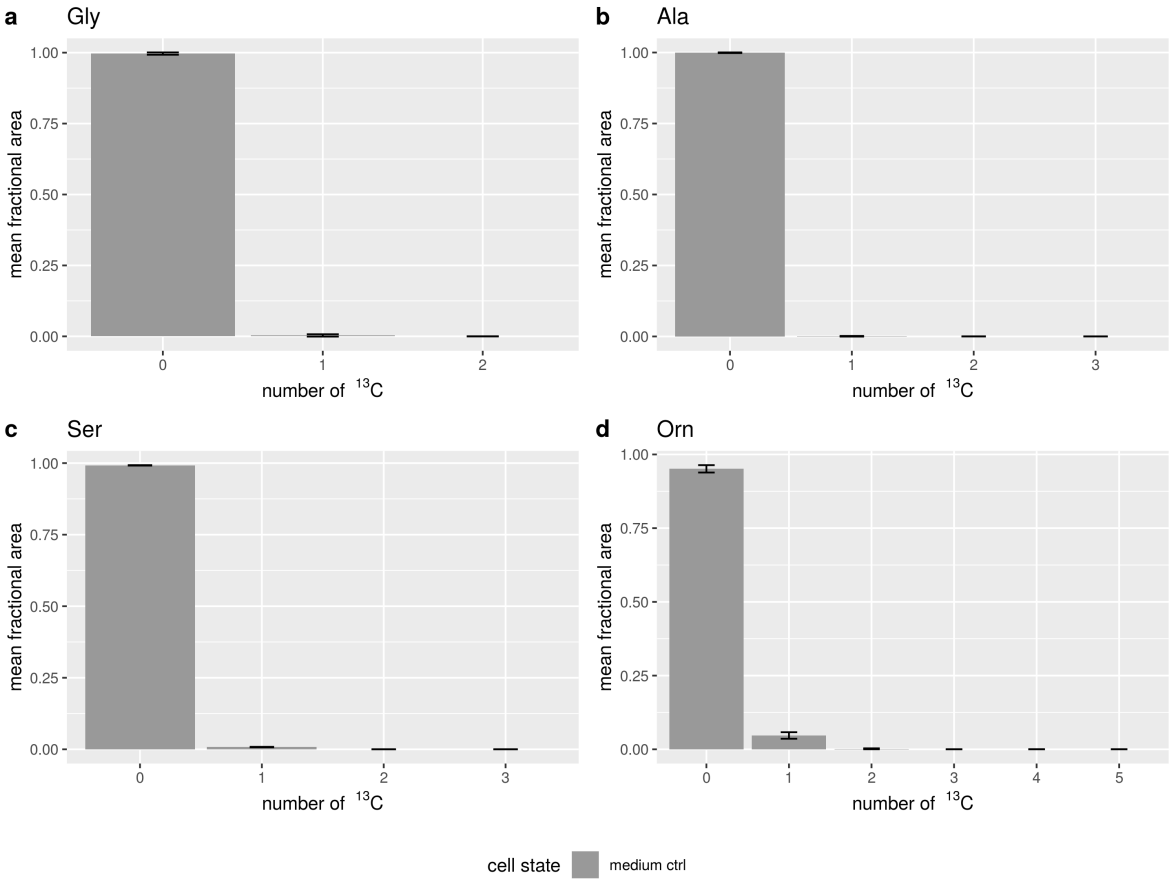



Figure A42: ***$^{13}$C tracer analysis with the tracer substrate U-$^{13}$C-glutamine in P493-6 cells, labeling of amino acids in the medium control.*** *The figure shows the MIDs of PCF-derivatized amino acids from the medium control (incubated for 24 h). Data were corrected for NA and tracer purity using IsoCorrectoR, assuming a tracer purity of 99%. The x-axis labels correspond to the number of $^{13}$C label incorporated in the respective isotopologue. Samples were measured in triplicates, means of isotopologue fractions +/- SD are shown.*

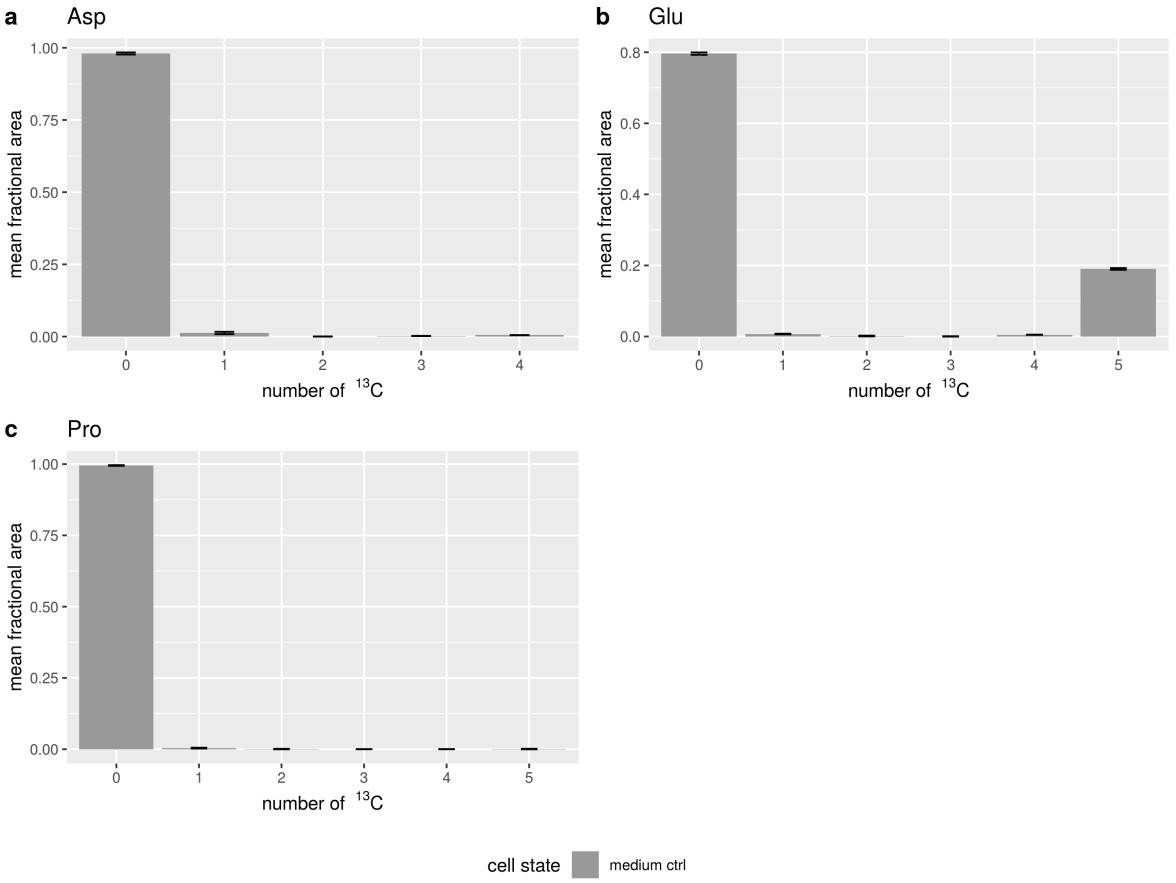
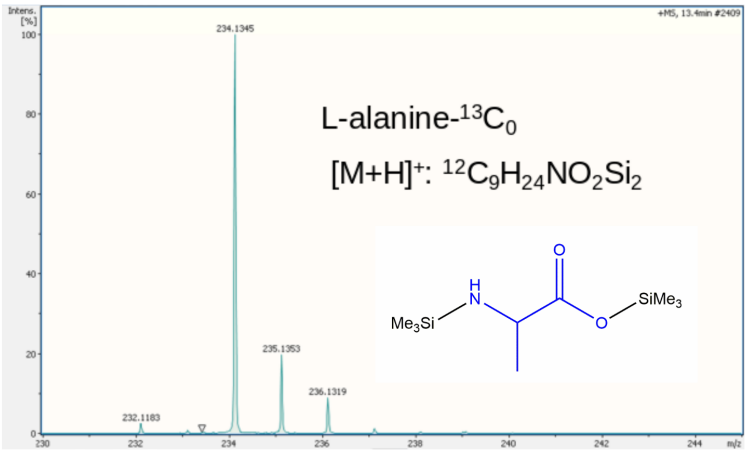


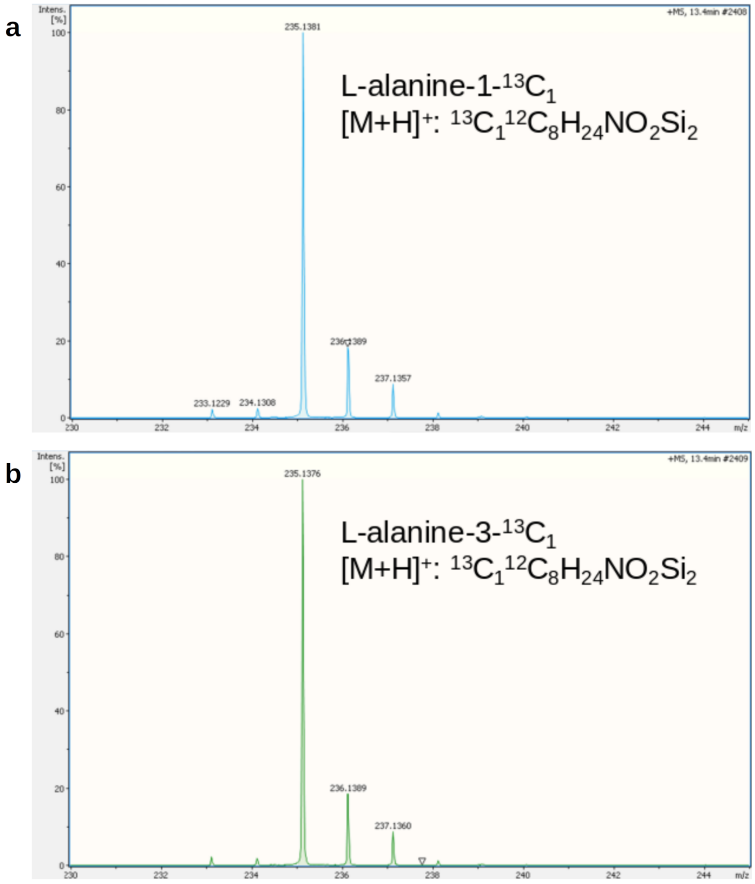

Figure A43: ***$^{13}C$ tracer analysis with the tracer substrate U-$^{13}C$-glutamine in P493-6 cells, labeling of amino acids in the medium control.*** *The figure shows the MIDs of PCF-derivatized amino acids from the medium control (incubated for 24 h). Data were corrected for NA and tracer purity using IsoCorrectoR, assuming a tracer purity of 99%. The x-axis labels correspond to the number of $^{13}C$ label incorporated in the respective isotopologue. Samples were measured in triplicates, means of isotopologue fractions +/- SD are shown.*

## A3 Methods



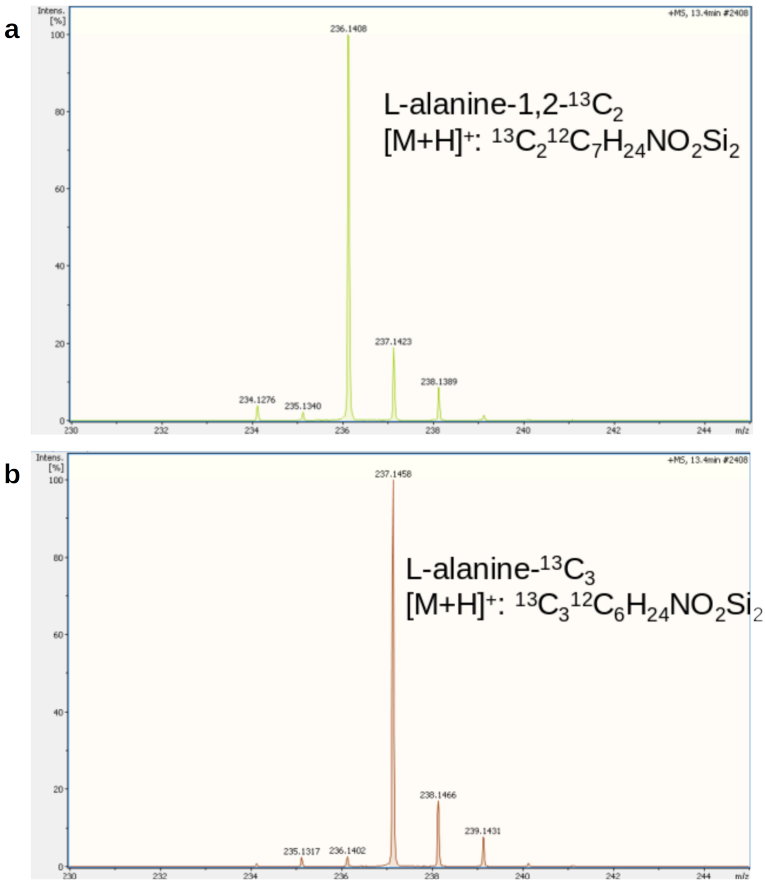

| m/z $[M+H]^+$ | Abundance % |
|---|---|
| 234.1340 | 100 |
| 235.1354 | 20.612 |
| 236.1325 | 8.929 |
| 237.1338 | 1.217 |
| 238.1317 | 0.219 |
| 239.1336 | 0.019 |

Figure A44: ***Mass spectra of the alanine standards used for generating mixtures of known composition, silylated L-alanine-$^{13}C_0$.*** *Standard solutions of the alanine isotopologues/isotopomers used for generating mixtures of known composition were measured via GC-APCI-TOFMS as described in the methods chapter. The table to the right of the mass spectrum depicts the expected masses and expected relative abundance of species from the respective isotopic distribution.* ***(Figure and legend taken from Heinrich et al. (2018) supplementary material and slightly modified)***




| m/z [M+H]+ | Abundance % |
|---|---|
| 235.1374 | 100 |
| 236.1387 | 19.53 |
| 237.1357 | 8.718 |
| 238.137 | 1.123 |
| 239.1348 | 0.207 |
| 240.1368 | 0.017 |

Figure A45: ***Mass spectra of the alanine standards used for generating mixtures of known composition, silylated L-alanine-1-$^{13}C_1$ and L-alanine-3-$^{13}C_1$.*** *Standard solutions of the alanine isotopologues/isotopomers used for generating mixtures of known composition were measured via GC-APCI-TOFMS as described in the methods chapter. The table to the right of the mass spectrum depicts the expected masses and expected relative abundance of species from the respective isotopic distribution.* ***(Figure and legend taken from Heinrich et al. (2018) supplementary material and slightly modified)***




| m/z [M+H]$^+$ | Abundance % |
|---|---|
| 236.1407 | 100 |
| 237.1419 | 18.449 |
| 238.1389 | 8.519 |
| 239.1402 | 1.03 |
| 240.1378 | 0.196 |
| 241.1400 | 0.015 |

| m/z [M+H]$^+$ | Abundance % |
|---|---|
| 237.1441 | 100 |
| 238.1451 | 17.367 |
| 239.1421 | 8.331 |
| 240.1434 | 0.94 |
| 241.1409 | 0.186 |
| 242.1433 | 0.013 |

Figure A46: ***Mass spectra of the alanine standards used for generating mixtures of known composition, silylated L-alanine-2,3-$^{13}C_2$ and L-alanine-$^{13}C_3$.*** *Standard solutions of the alanine isotopologues/isotopomers used for generating mixtures of known composition were measured via GC-APCI-TOFMS as described in the methods chapter. The table to the right of the mass spectrum depicts the expected masses and expected relative abundance of species from the respective isotopic distribution.* ***(Figure and legend taken from Heinrich et al. (2018) supplementary material and slightly modified)***

| **Isotope** | **Mass IsoCor v2** | **Abundance default** | **Abundance IsoCor v2** | **Abundance PyNAC** |
|---|---|---|---|---|
| $^{12}C$ | 12 | 0.9893 | 0.9893 | 0.98891 |
| $^{13}C$ | 13.003354835 | 0.0107 | 0.0107 | 0.01109 |
| $^{1}H$ | 1.0078250322 | 0.999885 | 0.999885 | 0.99985 |
| $^{2}H$ | 2.0141017781 | 0.000115 | 0.000115 | 0.00015 |
| $^{14}N$ | 14.003074004 | 0.99632 | 0.99636 | 0.9963 |
| $^{15}N$ | 15.000108899 | 0.00368 | 0.00364 | 0.0037 |
| $^{16}O$ | 15.99491462 | 0.99757 | 0.99757 | - |
| $^{17}O$ | 16.999131757 | 0.00038 | 0.00038 | - |
| $^{18}O$ | 17.999159613 | 0.00205 | 0.00205 | - |
| $^{31}P$ | 30.973761998 | 1 | 1 | - |
| $^{32}S$ | 31.972071174 | 0.9493 | 0.9499 | - |
| $^{33}S$ | 32.971458910 | 0.0076 | 0.0075 | - |
| $^{34}S$ | 33.9678670 | 0.0429 | 0.0425 | - |
| $^{36}S$ | 35.967081 | 0.0002 | 0.0001 | - |
| $^{28}Si$ | 27.976926535 | 0.922297 | 0.92223 | - |
| $^{29}Si$ | 28.976494665 | 0.046832 | 0.04685 | - |
| $^{30}Si$ | 29.9737701 | 0.030872 | 0.03092 | - |

Table A7: ***Natural isotope abundance values and absolute isotope masses used with IsoCorrectoR.*** *Isotope abundance values and absolute isotope masses as used by IsoCorrectoR. Default abundance: Values used by IsoCorrectoR by default (taken from Rosman & Taylor (1998)). IsoCor v2 and PyNAC abundance: Values used by the respective tools. As it performs only UHR correction, only the isotope abundance of potential tracer elements is relevant for PyNAC. It is limited to the tracer isotopes $^{13}C$, $^{15}N$ and $^{2}H$.*