Automated Analysis for Multiplet Identification from Ultra-High Resolution 2D-^{13}C, ^{1}H -HSQC NMR Spectra

Laura Ferrante[1], Kashif Rajpoot[2], Mark Jeeves[3], Christian Ludwig[4]

[1]School of Computer Science, University of Birmingham, Birmingham, UK - [2]School of Computer Science, University of Birmingham Dubai, Block 2, Dubai International Academic city, [3]Institute of Cancer and genomic Sciences, University of Birmingham, Birmingham, UK - [4]Institute of Metabolism and Systems Research, University of Birmingham, Birmingham, UK

Nuclear magnetic resonance (NMR) based stable isotope resolved metabolism analysis is a very powerful tool. To derive this information, signal line shapes from high-resolution 2D-^{1}H,^{13}C-HSQC NMR are quantitatively analysed. However, metabolites identification and multiplet quantification can be very challenging because of for example chemical shift changes due to pH changes. Here we present a data-driven, computational approach to automate the analysis of these signals, which is currently a bottleneck in the analysis workflow as currently data is usually analysed manually. This requires not only a large degree of expertise from the user, but also can lead to a biased signal analysis, which will have a huge impact downstream, where the NMR multiplet data is combined with mass spectrometry (MS) data to derive model-free isotopomer distributions for the metabolites analysed.

Metabolism is at the core of every cell's ability to survive and proliferate. Just as healthy metabolism requires the efficient functioning of the metabolic network, metabolic perturbations within the cell can result in disease by spreading through and altering the entire metabolic network. Feeding metabolic precursors, such as glucose or glutamine, enriched with a low-abundance, stable isotope, such as ^{13}C, enables the tracing of metabolic pathways. While a single metabolite can be made from different sources, the contribution of the different metabolic pathways leading to the production of this metabolite can be determined through the analysis of the ^{13}C distribution within the metabolite [1-2].

However, NMR based stable isotope metabolism analysis requires an in-depth knowledge of 2D-NMR spectroscopy, such as 2D-^{13}C,^{1}H-HSQC NMR spectra, of complex mixtures. Signal annotation and multiplet analysis are the most important and error-prone steps within the analysis, as there are effects such as changes of the exact resonance position as a result of matrix effects or due to very minor pH changes of the mixture. An additional challenge is that the shape of the multiplets can differ quite drastically depending on the tracer chosen as well as the activity of metabolic pathways.

In this work, the problem of metabolite identification is firstly cast into a signal unmixing problem, aiming at decomposing the spectrum into statistically fundamental components, without using a specific mathematical model or prior knowledge. A 2D-^{1}H,^{13}C-HSQC NMR spectrum of a complex metabolite mixture is a linear superposition of the individual metabolite spectra. Therefore, it is theoretically possible to disentangle major components in a data-driven manner. The problem can be formulated as an inverse modelling problem, in which multivariate observations (e.g: the 2D-^{1}H,^{13}C-HSQC NMR spectrum) relates to lower dimensional vectors of statistically independent variables (e.g: the contribution of each metabolite to the 2D NMR spectrum), using a linear model. Secondly, once a metabolite multiplet has been identified and localized within the spectrum, the multiplet analysis assumes that the experimental ^{13}C-NMR signal is obtained as a linear combination of multiplet components, which can be theoretically simulated using the PyGamma library [3]. A constrained least square method is used to estimate the parameters representing the contribution of the components to the final pattern. The software developed within this work builds upon the MetaboLab software package for NMR data processing [4]. The signal annotation and multiplet analysis is reliable, with an accuracy >80%. The quality of the automated data analysis is assessed via confidence levels.

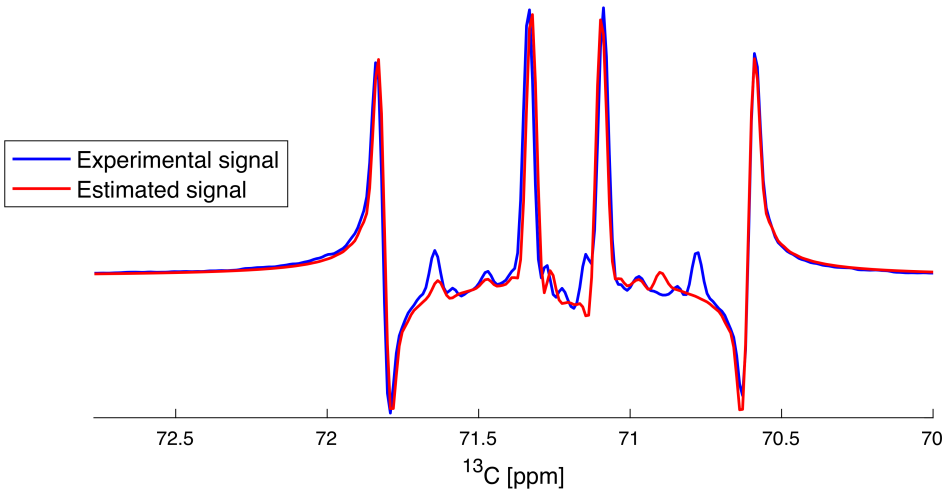

Figure 1: Result of automated multiplet analysis. The experimental NMR signal of Lactate (carbon 2) is displayed in blue, while the estimated signal using multiplet simulation is shown in red.

[1] Fan TW, Lane AN, Higashi RM, Farag MA, Gao H, Bousamra M, Miller DM, Altered regulation of metabolic pathways in human lung cancer discerned by 13C stable isotope-resolved metabolomics (SIRM). Mol. Cancer, 2009, 8, 4.

[2] Chong M, Jayaraman A, Marin S, Selivanov V, da Auturi Carla PR, Tennant DA, Cascante M, Günther UL, Ludwig C. Combined Analysis of NMR and MS Spectra (CANMS). Angewandte Chemie - International Edition, 2017, 56(15), 4140-4144.

[3] Wiechert W, Möllney M, Isermann N, Wurzel M, De Graaf AA, Biotechnol. Bioeng., 1999, 66, 69 – 85.

[4] Ludwig C, Günther UL. MetaboLab - advanced NMR data processing and analysis for metabolomics. BMC Bioinformatics, 2011. 12(1):366-371.