Introduction

Metabolomics studies have gained major importance in food applications. Quality control of food samples can be established on a large scale via state-of-the-art metabolic profiling based on statistical data evaluation. However, the metabolomics workflow puts high demands on an analytical system in terms of its robustness, dynamic range and identification capabilities. The Bruker compact™ QTOF System combines the robustness of Bruker orthogonal-Time-of-Flight instruments with new 10-bit digitizer technology that allows for compound detection and identification across an unrivalled dynamic range.

After water and black tea, coffee represents the third most important beverage with a worldwide consumption of 4.5 million tons per year [1]. Analytically, it constitutes a very complex mixture of small molecules which differ in composition and quantities based on the different coffee cultivars, cultivation regions and processing procedures.

As a proof of concept study for the compact™, we analyzed 13 different types of coffee capsule extracts assigned by their manufacturer to different intensity categories. The major analytical task was to correlate high resolution LC-MS data to the manufacturer’s description via a non-targeted metabolomics approach.

Authors

Dr. Verena Tellström, Dr. Alexander Harder, Klaus Meyer, Dr. Aiko Barsch
Bruker Daltonik GmbH, Bremen, Germany

<table>
<thead>
<tr>
<th>Keywords</th>
<th>Instrumentation and Software</th>
</tr>
</thead>
<tbody>
<tr>
<td>compact</td>
<td>compact</td>
</tr>
<tr>
<td>Metabolomics</td>
<td>DataAnalysis</td>
</tr>
<tr>
<td>Structure Elucidation</td>
<td>SmartFormula 3D</td>
</tr>
<tr>
<td>Coffee</td>
<td>MetFrag</td>
</tr>
<tr>
<td>PCA</td>
<td>FragmentationExplorer</td>
</tr>
<tr>
<td>ProfileAnalysis</td>
<td></td>
</tr>
</tbody>
</table>
**Results and Discussion**

Evaluating the **compact** QTOF performance for the analysis of large metabolomics sample sets

In metabolomics studies, large numbers of samples are often analyzed with each consisting of a complex mixture of small molecules across a wide range of abundances. In order to monitor the system stability, more and more researches analyze so-called quality control (QC) samples before, within and after the real analytical samples. Pools of all analytical samples were collected for a common quality control sample since they represent a mixture similar in complexity and chemical composition. The Base Peak Chromatograms of 15 QC sample injections spanning a sequence of ~150 coffee extract injections is shown in Figure 1 A. Peak shape and intensity remain unchanged revealing the stability of the **compact** instrument which is an important prerequisite for analyzing large metabolomics data sets.

Applying sophisticated peak picking algorithms is a first crucial step in data pre-processing and it forms the basis for statistical analysis of metabolomics datasets. The FindMolecularFeatures (FMF) peak finder combines ions belonging to one compound, such as common adducts (e.g. +Na, +K, +NH4), isotopes and charge states. Molecular features were extracted using this strategy within high resolution full scan MS data acquired for the QC and analytical samples. Based on the extracted features, a pareto scaled Principal Component Analysis (PCA) was then calculated. The PCA scores plot shown in Figure 1 B reveals a close clustering of the QC samples (red circles) indicating that the data acquisition using the **compact** instrument which is an important prerequisite for analyzing large metabolomics data sets.

The extreme dynamic range of analyte concentrations in complex biological samples represents one of the most crucial challenges in metabolomics applications. To access the dynamic range of the **compact** QTOF instrument, which is equipped with a new 10-bit digitizer technology, we performed a proof of concept study. A mixture of two small molecules, alanine and caffeine, was analyzed by direct infusion at an acquisition rate of 1Hz. As shown in Figure 2 the low concentrated alanine was detected with an intensity of 46 cts whereas the caffeine peak had an intensity of 7,155,285 cts. The intensity ratio between these two values, \( \frac{7,155,285}{46} = 1.6 \times 10^5 \), demonstrates the unique capability of the **compact** to detect target compounds on an LC timescale across a dynamic range > 5 orders of magnitude. Furthermore, the mass accuracy for both compounds was lower than 1 ppm thus enabling a reliable compound identification.

**Experimental**

Capsules of 13 different types of coffee were extracted using 35 ml of water on a standard coffee capsule machine (Krups XN 301T Nespresso Pixie). Two replicates of each type were prepared. A QC sample was generated by pooling equal volumes of all coffee samples profiled in this study. Extracts were diluted 1:50 in water prior to analyzing 5µl of 3 replicates for each extract by UHPLC-MS. Chromatographic separation was carried out using an RS-LC system (Dionex) with a 50 x 2.1 mm BEH C18, 1.7 um column (Waters), at a flow rate of 0.45 mL/min, with Solvent A: Water + 0.1% HCOOH and Solvent B: methanol + 0.1% HCOOH. The following LC gradient program was used: linear increase from 2% B to 98% B (over 5 min), constant at 98%B (for 1 min).

MS detection was performed using a **compact** Qq-TOF mass spectrometer (Bruker Daltonik). The instrument was operated in ESI positive mode acquiring MS full scan and auto MS/MS data at a 3 Hz acquisition speed.

ProfileAnalysis 2.1 software (Bruker Daltonik) was used for statistical data analysis based on features extracted by the FindMolecularFeatures (FMF) algorithm which can combine all ions belonging to the same compound (isotopes, charge states, adducts and common neutral losses). Further data evaluation was performed using DataAnalysis 4.1 software (Bruker Daltonik). Molecular formula determination was carried out by combined evaluation of mass accuracy, isotopic patterns, adduct and fragment information using SmartFormula3D software. The direct link from SmartFormula3D to the open source MetFrag tool (http://msbi.ipb-halle.de/MetFrag/, [2]) allowed for the identification of target compounds by means of database search, in silico fragmentation and matching of expected and observed fragments. Fragmentation patterns were further evaluated using the ChemDraw™ based FragmentationExplorer within DataAnalysis. Compound identities were confirmed by comparing retention time, MS and MS/MS spectra to those observed for commercially available standards.
Figure 1: A: Overlay of Base Peak Chromatograms of interspersed quality control (QC) samples run during the acquisition of a coffee metabolomics sample batch demonstrates the stability of the Compact QTOF for analysing large sets of complex samples. B: A PCA scores plot of coffee metabolomics (green) and QC samples (red) reveals tight clustering of QC samples as an additional proof for reproducibility during data acquisition.

Figure 2: A mixture of alanine and caffeine measured at a 1 Hz acquisition rate in a direct infusion experiment demonstrates >5 orders of magnitude dynamic range with mass accuracies better than 1ppm.
Coffee samples cluster according to the assigned coffee intensity

For a detailed evaluation of the different coffee types analysed in this study, the QC samples were removed from the calculated PCA model. In total, 13 different coffee extracts, comprising espresso and lungo varieties from different blends and geographical regions were analysed. The two “biological” and three technical replicates for each sample type (highlighted by using the same colour and symbol) formed clusters in the PCA scores plot as seen in Figure 3 A. In order to avoid the dominance of the caffeine content on the clustering results, the corresponding peak was removed from the presented model. The manufacturer assigned different intensities to each coffee type (numbers from 3 (weak) to 10 (strong)). The PCA scores plot clearly separates coffee extracts described as strong (9 and 10) which cluster in the negative part of PC1, from those samples that are assigned as weak (3), and form clusters on the positive part of PC1. The medium intensity coffee varieties (4-8) cluster in the centre of the PCA scores plot.

These samples do not show a perfect alignment on PC1 according to the coffee intensity, although a trend can clearly be seen. Quite likely, the separation of samples is not only due to characteristics that contribute to intensity, but also to coffee plant varieties used for blending (data not shown) and geographical origin. Furthermore, lungo style samples are recommended by the manufacturer to be extracted in larger amounts of water compared to espressos indicating a difference in blending or roasting.

Analytes mainly contributing to the separation of samples in the PCA scores plot can be accessed from the corresponding loadings plot (Fig. 3 B). Two compounds marked X and Y are identified by the loadings plot to have a high content in strong and weak coffee samples, respectively. To directly access the relative amount for both metabolites in all samples, bucket statistics plots are shown in Fig. 3 C. These plots confirm the high abundance of compound X in strong coffees and of compound Y in weak samples.

Figure 3: A PCA scores plot reveals separation of samples on PC1 according to coffee intensity assigned by the coffee manufacturer (numbers from 3 (weak) – 10 (strong)). B: The corresponding PCA loadings plot points to compounds mainly contributing to this differentiation. The two analytes with highest contribution for strong and weak coffee are marked X and Y, respectively. C: Bucket statistics plots for two selected loadings plot visualize the relative abundance of the compounds “X” and “Y” across all samples. These plots indicate clear abundance trends in strong vs. weak samples and vice versa.
Identification of trigonelline and nicotinic acid

Compound identities were elucidated using the MS/MS capabilities of the Compact QTOF instrument. SmartFormula3D readily provides the correct molecular formula by combining accurate mass and isotopic pattern information for MS and MS/MS spectra (Fig. 4 A). Furthermore fragment formulae are easily accessed and can be used for gaining structural information. For compound X a unique elemental composition of C₆H₆NO₂ ([M+H]+) was generated. Via a direct link to MetFrag (Fig. 4 C), an open source tool for in-silico fragmentation, the molecular formula can be searched in multiple public databases so that possible structures can be determined [2]. All hits can then be fragmented in-silico and the resulting fragment masses can be matched to the measured fragment ions. The automated assignment of fragment formulae by SmartFormula3D enables to match the in-silico predictions with a “quasi” 0 ppm mass accuracy reducing false positive hits.

With the optimum score of 1.0, MetFrag suggested nicotinic acid as the most likely structure corresponding to compound X. In order to further evaluate the fragmentation pattern of this analyte, the structure of nicotinic acid was matched to the MS/MS fragment ions using the ChemDraw™ based FragmentationExplorer available within the DataAnalysis™ software. This tool allows for the annotation and reporting of MS/MS fragmentation patterns as seen in Fig. 4 D. Ultimately, the compound identity of nicotinic acid was confirmed by comparing retention time, accurate mass, isotopic pattern information and MS/MS fragmentation against the pure reference standard.

Using the workflow described above, the molecular formula for compound Y was identified as C₇H₈NO₂ ([M+H]+). In-silico fragmentation and comparison to a pure reference confirmed the analyte to be trigonelline.

Figure 4: Identification of Compound X: A+B: Based on accurate mass and isotopic pattern information in MS and MS/MS spectra for unknown compound X SmartFormula3D returned a single molecular formula for the precursor and corresponding formulae for the fragment ions. C: A direct link to MetFrag [2] enables the ability to send the molecular formula for the precursor and theoretical masses for fragment ions to the open source web application. MetFrag returned nicotinic acid as most likely structure by matching in-silico fragments with measured fragment ions. D: Detailed evaluation using the FragmentationExplorer supports the in-silico fragments generated by MetFrag. E: Verification of the identity of Compound X as nicotinic acid was achieved by comparing retention time, accurate mass and MS/MS spectra with the pure reference standard.
Many compounds that contribute to the typical flavour of coffee as a beverage are only formed during the high temperature treatment of the roasting process. Trigonelline is one of the major analytes in unprocessed coffee and is transformed during the roasting process mainly to pyridine and nicotinic acid [3, 5]. The obtained results perfectly match this prior knowledge. As seen in figure 5, trigonelline (compound Y) has a higher content in coffees classified with strength 3 suggesting a weaker roasting. In contrast, the trigonelline degradation product nicotinic acid (compound X) is more pronounced in the samples described as 9 and 10 in strength indicating a stronger roasting.

Figure 5: Schema of trigonelline degradation during coffee roasting (adapted from [3]). Trigonelline (corresponds to compound Y), a coffee metabolite, is known to be transformed to nicotinic acid (corresponds to compound X) proportional to the degree of roasting [4].

Conclusion

The compact QTOF provides unrivalled dynamic range (> 5 orders of magnitude) in combination with mass accuracy, sensitivity, MS/MS performance and robustness enabling this instrument to be the tool of choice for analyzing batches of highly complex metabolomics samples. Together with sophisticated software for statistical evaluation of metabolomics datasets, these features enabled quick pinpointing of relevant compounds contributing to coffee intensity. Molecular formula generation for MS as well as MS/MS spectra using SmartFormula3D together with the in-silico fragmentation tools MetFrag and FragmentationExplorer generated single structure candidates for two selected target compounds characteristic for weak and strong coffee samples.

This reliable proposal of compound identities helped to save analysis time and money spent for purchasing multiple references in order to confirm the identity of the target compounds.

References


For research use only. Not for use in diagnostic procedures.