The development of an LC-FAIMS-MS metabolomics workflow: a new tool for untargeted metabolite profiling to diagnose disease

Lauren Brown; Kayleigh Arthur; Alasdair Edge; Aditya Malkar; Paul Nasca

Owlstone Medical Ltd., 162 Cambridge Science Park, Cambridge, CB4 0GH, UK

For further information, email: **ultrafaims@owlstone.co.uk**

1. Introduction

In non-targeted 'omics applications, liquid or gas chromatography (LC or GC) is typically combined with mass spectrometry (MS) for analysis of complex biological samples.

LC-MS molecular features can be missed due to:

- Trace component suppression by chemical noise.
- Chromatographically unresolved isomeric species.

Orthogonality between FAIMS, LC and MS provides additional unique compound identifiers with detection of features based on:

- Retention time (RT)
- FAIMS dispersion and compensation fields (DF and CF)
- Mass-to-charge (m/z) (Figure 1)

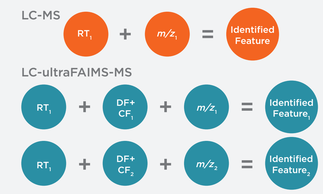

Figure 1. LC-MS and LC-ultraFAIMS-MS feature determination

A recent study[1] reported a threefold increase in features detected in non-targeted profiling of human urine with the addition of FAIMS to LC-MS analysis.

Here we describe an optimized workflow to produce three-dimensional metabolomics data sets and its application to urinary analysis of a colorectal cancer cohort.

2. Methods

Analysis was performed on a 6230 TOF-MS and a 1290 series LC (Agilent, Santa Clara, US) combined with an ultraFAIMS device (Owlstone Medical Ltd, Cambridge, UK).

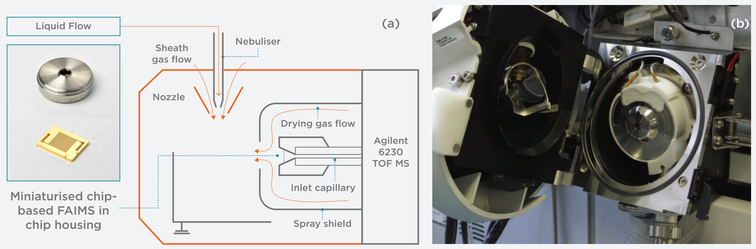

Figure 2. (a) ultraFAIMS chip and source schematic (b) installed on Agilent 6230 ToF-MS

The key dimensions of the ultraFAIMS device are the 100 µm electrode gap and 700 µm path length. The small scale is key to the ability to integrate into the LC-MS workflow.

- The short ion residence time means an entire CF scan per second can be achieved, making the scanning approach compatible with chromatographic timescales.
- The CF values are synchronized with MS acquisition so a spectrum is acquired for each CF (Figure 3).
- Synchronization was provided via a contact closure interface from the binary pump on the LC.
- LC was performed using a reversed phase Poroshell 120 EC-C18, 2.1 x 100 mm, particle size 2.7 µm (Agilent Technologies).

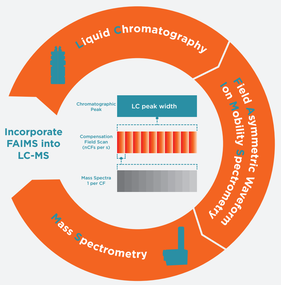

Figure 3. Integrated LC-ultraFAIMS-MS analysis

Optimization of chromatography to achieve optimal retention of small polar molecules and coverage of features across the chromatographic range was carried out prior to FAIMS optimization experiments.

References

1. Arthur, KA, Turner, MA, Reynolds, JC, Creaser, CS.; Anal. Chem., 2017, 89, 3452–3459.

3. Data analysis

Proteowizard's "msconvert" was used to convert the Agilent MassHunter .d data into mzML.

An in-house developed Python based tool was then used to split the LC-FAIMS-MS data and convert it to individual .mzML files associated with each FAIMS CF setting.

The saved .mzML files were subjected to feature extraction using XCMS package in R.

The workflow for feature extraction from LC-FAIMS-MS files is detailed in Figure 4.

A feature was defined as a unique identifier for each component of a m/z, a retention time (t_R) and a compensation field (CF).

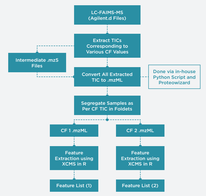

Figure 4. LC-ultraFAIMS-MS feature extraction workflow

4. LC-ultraFAIMS-MS optimization

Optimization of the acquisition of nested FAIMS data focussed on:

- Number and range of FAIMS CF settings to give optimal FAIMS separation
- Number of data points within the timescale of the chromatographic peaks
- Optimal sensitivity via chromatographic peak heights and number of TOF scans s^{-1}.

Full data acquisition parameters are shown in Table 1.

Table 1: LC-ultraFAIMS-MS optimization experiments

Test	MS scan rate (s^{-1})	per s	m/z start	m/z end	Index	Start CF (Td)	End CF (Td)	N CF steps	N CF actual	CF step size (Td)	N Repeats	Start DF (Td)	End DF (Td)	N DF Steps
1a	12	1	80	1500	0	-0.9	4.1	10	12	0.5	503	250	250	0
1b	12	1	80	1500	0	-0.9	4.1	10	12	0.5	503	240	240	0
1c	12	1	80	1500	0	-0.9	4.1	10	12	0.5	503	230	230	0
2	24	2	80	1500	0	-0.9	4.1	10	12	0.5	1007	Test 1	Test 1	0
3a	10	1	80	1500	0	-0.9	3.1	8	10	0.5	503	Test 1	Test 1	0
3b	20	2	80	1500	0	-0.9	3.1	8	10	0.5	1007	Test 1	Test 1	0
3c	18	1	80	1500	0	-0.9	3.1	16	18	0.25	503	Test 1	Test 1	0
3d	6	1	80	1500	0	-0.9	3.1	4	6	1	503	Test 1	Test 1	0
3e	12	2	80	1500	0	-0.9	3.1	4	6	1	1007	Test 1	Test 1	0
3f	18	3	80	1500	0	-0.9	3.1	4	6	1	1511	Test 1	Test 1	0

Optimal DF was determined based on trade off between selectivity and sensitivity.

- The CF vs features plot (Figure 5) shows good coverage across the analytical space in the range -1 to +3 Td.
- More features were detected in the higher CF region at the higher DFs as distribution increased with increasing DF.
- 240 Td was selected for further experiments based on widest distribution of detected features across the CF range.
- As the CF scan is synchronized with the MS acquisition, the number of data points across a chromatographic peak is dependent on the number of MS scans s^{-1} and the CF range.
- Increased MS scan rate means more data points can be acquired over a given CF range.
- Alternatively, a smaller CF step size could be applied, increasing the data points across the
- CF peak whilst maintaining the number of CF data points across the LC peak.
- Faster ToF scan rates do, however, reduce peak intensity (Figure 6).

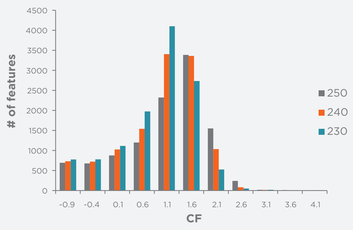

Figure 5. Comparison of features detected at DFs of 230, 240 and 250 Td

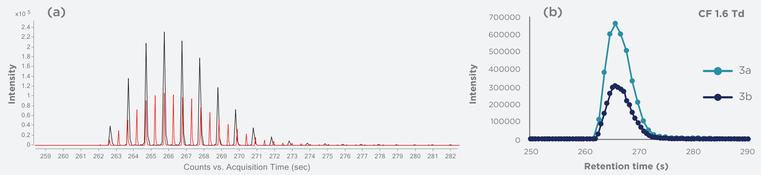

Figure 6. (a) Raw EIC for m/z 273 across all CFs (black) 10 spectra s^{-1} and (red) 20 spectra s^{-1}, showing twice as many CF scans across the LC peak and (b) deconvoluted LC-MS peak at CF of 1.6 Td showing twice as many data points at 20 spectra s^{-1}

- The effect of the different MS scan rate and CF ranges on feature detection was investigated.
- Acquiring data at 2 CF scans s^{-1} increased the number of features detected, in all cases, despite the decrease in peak intensity associated with higher acquisition speeds (Figure 7b, c).
- Increasing to 3 CF scans s^{-1} did not further increase the number of features detected.

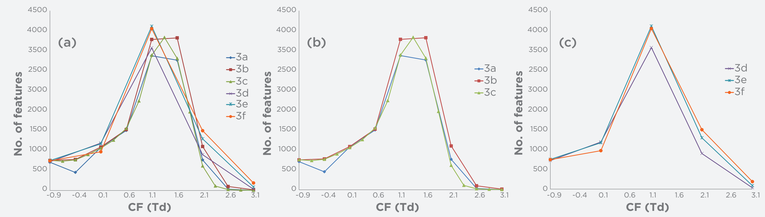

Figure 7. Comparison of the number of features detected in experiments 3 a-f as described in Table 1

5. Feature extraction

Features lists for each CF were aligned into one list based on retention time and m/z. Conditional formatting was applied to the features to deconvolute features in the FAIMS dimension. Features present in multiple CFs were identified as either adjacent or separated into multiple features in the CF dimension, such as in the case of isobars. An example of both is shown in Figure 8.

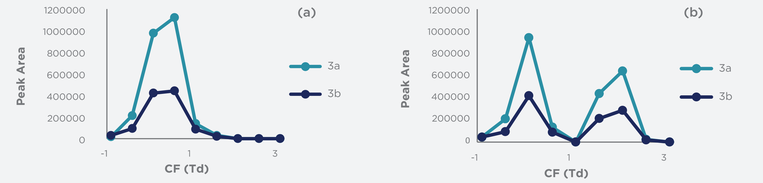

Figure 8. Example CF scans of detected features (a) single feature detected in adjacent CF ands (b) multiple features showing separation of isobars

At a S:N of 3, more features were detected using a 2 CF scans s^{-1} scan rate (Figure 9a).
To determine if this result was accurate, or a result of increased noise using the faster scan rate, the analysis was repeated at a S:N of 10 (Figure 9b).

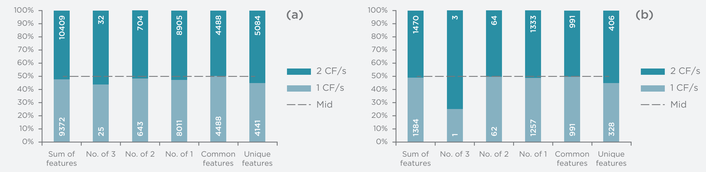

Figure 9. Comparison of features detected at 1 CF scan s^{-1} and 2 CF scans s^{-1} at (a) S:N of 3 and (b) S:N of 10

- Upon visual verification of the determined features, the lower sensitivity observed at the 2 CF scan s^{-1} rate resulted in increased noise in the baseline, accounting for the increased number of unique features detected at the S:N 3.
- 1 CF scan s^{-1} rate was therefore chosen for subsequent analysis, as a compromise between sensitivity and the number of data points across the peak for untargeted semi-quantitative analysis.

6. Conclusions

- The optimized FAIMS scan settings were used to generate a three-dimensional nested data set (Figure 10).
- The method is now being applied to the urinary analysis of a 674 patient colorectal cancer cohort in collaboration with University Hospital of Coventry and Warwickshire.
- The LC-ultraFAIMS-MS features list will be used to build a predictive model, or classifier, from which a probability of disease can be predicted. The additional features available in LC-ultraFAIMS-MS, compared to LC-MS alone, increase the likelihood of successfully building such a classifier.

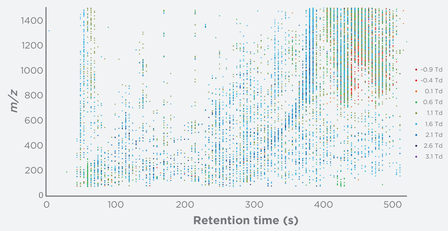

Figure 10. m/z vs t_R of a urine sample showing features detected at a S:N of 3; each CF is color differentiated