A Framework for the Rapid Development of New Data-Dependent Acquisition Strategies

***Vinny Davies**, Joe Wandy, Justin J.J. van der Hooft, Stefan Weidt, Alice Miller, Rónán Daly and Simon Rogers {vinny.davies, simon.rogers}@glasgow.ac.uk*

ViMMS
VIRTUAL MS
https://github.com/sdrogers/vimms

1. Introduction

In untargeted metabolomics, tandem mass spectrometry (LC-MS/MS) is widely used to identify unknown ions in complex samples. One standard approach for acquiring data is Data Dependent Acquisition (DDA), but despite its limitations, little work has been done to develop improved strategies.

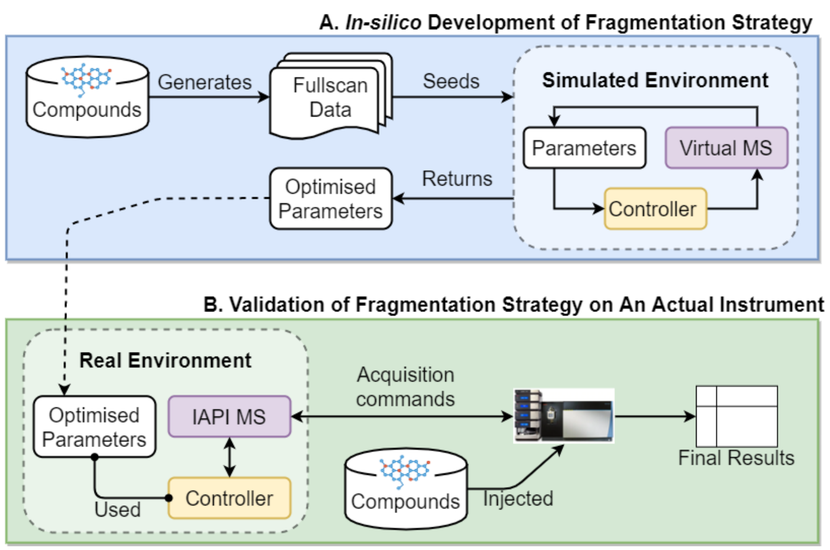
In this work we provide:

- A development framework for new acquisition strategies.
- Two new improved DDA methods as a proof of concept.
- A procedure to compute optimal DDA performance.

2. ViMMS Development Framework

The framework for the rapid development of new DDA methods is part of the Virtual Metabolomics Mass Spectrometer (ViMMS) simulator, and begins with a full scan .mzML file from a real sample. The process works as follows:

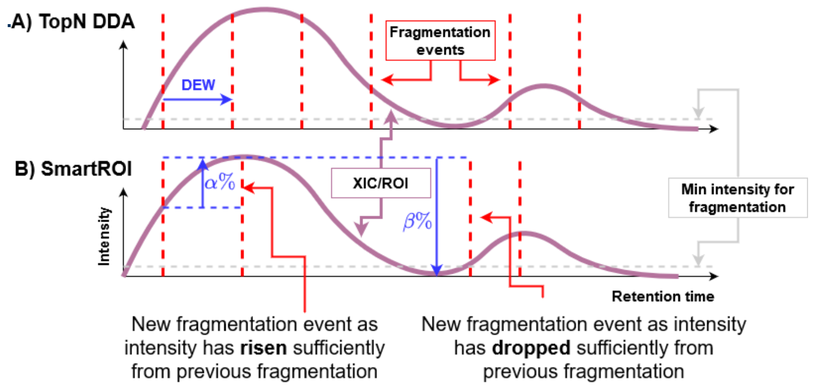
- Seed the simulator with the real data.
- Develop a new fragmentation controller within the ViMMS framework in the Python programming language.
- Evaluate the controller in the simulated environment, optimising the controller's parameters.
- Use the same controller and the optimised parameters on the actual Thermo Fusion instrument using the Instrument Application Programming Interface (IAPI).



ViMMS Development Framework.

3. SmartROI

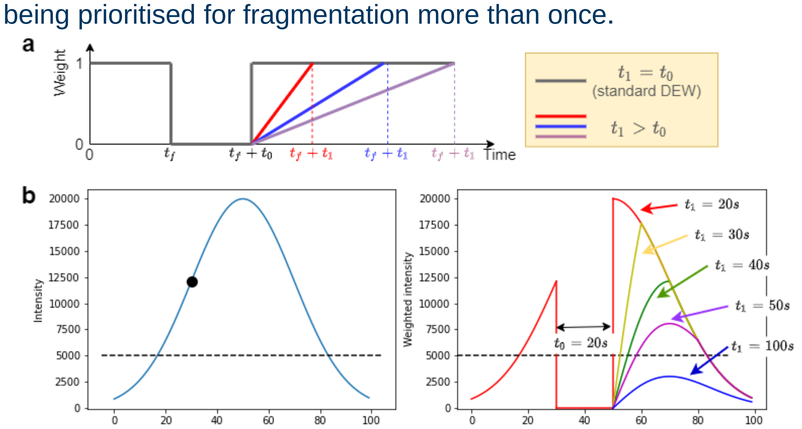
SmartROI is a new fragmentation strategy which is informed by regions of interest (ROIs) constructed in real time. The method is based on a traditional Top-N DDA duty cycle, with additional rules designed to prevent unnecessary repeated fragmentations of a chromatographic peak. SmartROI keeps track of the intensities within each ROI meaning that it still allows multiple peaks within an ROI to be fragmented when intensities profiles suggest it is needed



SmartROI compared with a Top-N strategy.

4. WeightedDEW

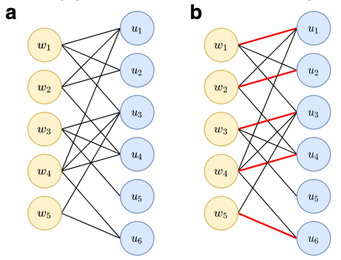
In standard DDA, ions can be thought of as being prioritised by their intensity multiplied by an indicator as to whether the ion is excluded (0 if excluded, 1 otherwise). WeightedDEW generalises this idea to non binary weights. When $t_0 = t_1$, the method is equivalent to a standard DDA approach, while higher values of t_1 reduce the chances of a chromatogram being prioritised for fragmentation more than once.



The weight function in the WeightedDEW method.

5. Optimal DDA Results

The theoretical optimal DDA performance can be computed by firstly creating a bipartite graph between MS2 scans and picked peak bounding boxes. Optimal performance can then be calculated using a maximal matching algorithm, giving us an upper bound on DDA performance.



Bipartite graph G and (b) G with a maximum matching shown in red. W={w1,w2,w3,w4,w5} corresponds to MS2 scans, U ={u1,u2,u3,u4,u5, u6} corresponds to peak bounding boxes (from peak picking).

6. Evaluation

Performance was evaluated by looking at the number of picked peaks (MZMine2) that were fragmented:

- The WeightedDEW method gives the best performance
- Both new DDA methods outperformed a standard Top-N
- The methods perform worse than the upper bound for DDA performance calculated using the matching algorithm.
- Further development of improved strategies is needed

Method	Beer (6267 peaks)	Serum (4481 peaks)
Top N	1509	1184
WeightedDEW	**2180**	**1534**
SmartROI	1982	1281
Optimal	2736	1767

Table showing the coverage (number of picked peaks fragmented) after a time shift correction (in brackets).

7. Conclusions

We have introduced a framework for the development of new DDA methods within ViMMS, but also allows any developed method to run on an actual instrument through Thermo's IAPI. To demonstrate the utility of the framework, we have developed two new DDA methods, SmartROI and WeightedDEW. These methods were developed and optimised within ViMMS and validated on two complex samples to show the improved performance.

***Funding**: EPSRC project EP/R018634/1 and ASDI eScience grant, ASDI.2017.030, from the Netherlands eScience Center—NLeSC.*