COMPARISON OF PRODUCT ION SPECIFICITY IN LC-MS DATA INDEPENDENT ACQUISITION (DIA) BETWEEN MS^E AND MS^E WITH M/Z SELECTIVE INTENSITY ENCODING

Waters™

Keith Richardson, Martin Green, Christopher Hughes
Waters Corporation, Altrincham Road, Wilmslow, UK

OVERVIEW

PURPOSE: Investigation of specificity of MS^E with m/z selective intensity encoding relative to a standard MS^E experiment for a complex mixture using a Q-ToF instrument.

METHODS: Precursors were encoded by amplitude modulation using dipole excitation in a modified quadrupole ion guide. Data were decoded using a non-negative least squares approach.

RESULTS: Significant Improvements in signal to noise were observed in extracted ion chromatograms, extracted DIA spectra and reconstructed MS/MS spectra.

INTRODUCTION

Data Independent Acquisition (DIA) tandem (LC-)MS workflows have become increasingly popular owing to the comprehensive and relatively unbiased nature of the data they produce.

In the simplest approach (MS^E), a wide band of precursors is subjected to alternate low and high collision energies. Fragments are linked to precursors through their elution profiles. This approach has many strengths, providing MS/MS information for all precursors with high duty cycle.

As complexity increases however, ambiguity begins to limit the confidence of the parent-fragment assignments. Algorithms designed to search traditional MS/MS data can give low quality results when searching MS^E data. Furthermore, the background present in the high energy data can limit the quantitative dynamic range attainable for complex samples.

We have previously described (Proc. ASMS 2020) a novel method using a modified quadrupole mass filter with frequency scanning dipolar excitation to add specificity to an MS^E like experiment without significantly compromising sensitivity or changing the basic instrument geometry.

Here we compare data from LC-MS separation of complex protein digests between MS^E and the new method and examine its ability to mine deeper into product ion data.

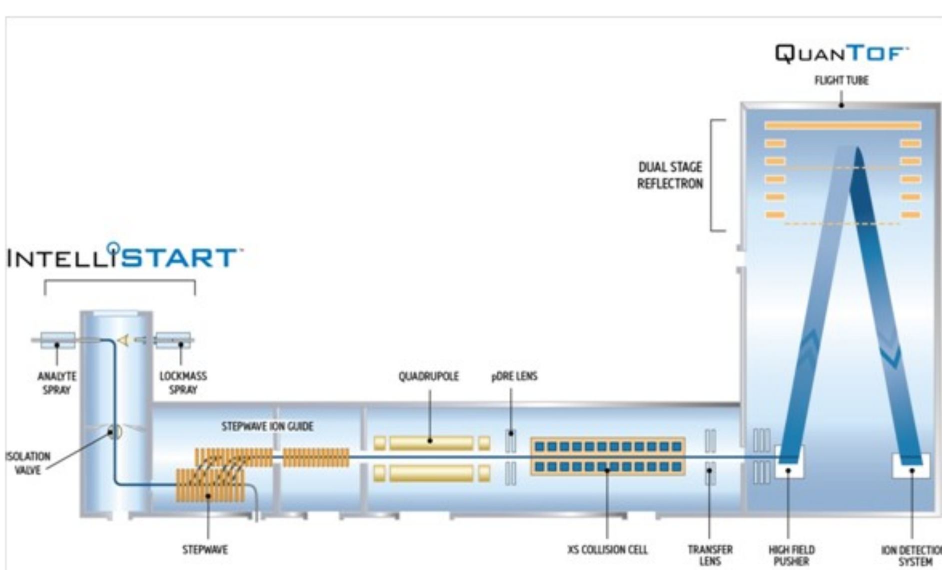

Figure 1. Schematic of the Waters™ Xevo™ G2-XS QToF

METHODS

Samples

E.Coli tryptic digest standard (Waters)

Liquid Chromatography

LC:	Waters NanoACQUITY™ UPLC™
Analytical Column:	Waters CSH™ 300 µm x 10 cm
Gradient:	5 to 40% acetonitrile + 0.1% FA
Gradient Length:	45minutes + re-equilibration
Flow rate:	7µL/min

Mass Spectrometry

The quadrupole mass filter on a Waters Xevo G2-XS QToF (Figure 1) was modified to allow the application of a dipole excitation waveform. The system was operated in an MS^E-like mode and three analyte functions were acquired: 1) low collision energy (CE) 2) 22V trap CE and 3) 32V transfer CE. A lock spray function was also acquired. Two discrete collision energies were used rather than the more usual ramp to avoid interference with the encoding pattern.

The scanning notch pattern employed in both low and high energy in this experiment (see Figure 2A) was designed to simultaneously remove five precursor m/z ranges between m/z 300 and m/z 1100. The pattern was adjusted during the scan time of 0.5s to ensure that every m/z value was removed a total of five times. The width of the notches is such that the duty cycle of the experiment remains comparable to an MS^E experiment (>75%).

The SONAR acquisition mode was used to produce and store 200 sub-scans profiling the scanning notch pattern, for an acquisition rate of 400Hz during the scan time.

Model and Data Processing

The data were initially processed as an ordinary MS^E dataset (ignoring the encoding dimension) using Waters ProteinLynx Global Server (PLGS) v3.0.3. and the resulting time-aligned product ion spectra with associated precursor masses were exported in the Mascot™ [1] MGF file format.

A model of the notch pattern was constructed from dataset specially acquired under conditions designed to ensure a densely populated m/z range prior to the quadrupole. This data was then normalised, smoothed and rebinned to produce the desired model (**Figure 2A**). The bin width in the model is 2Th in the notch scan dimension, and 0.25Th in the ToF dimension. This model predicts the intensity modulation pattern for a fragment of a precursor of a given m/z (**Figure 2B**).

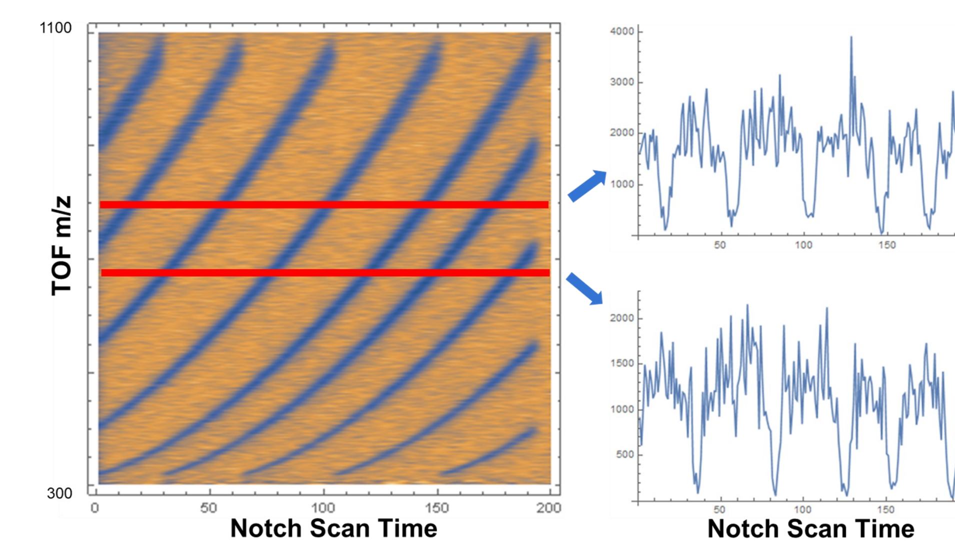

Figure 2A) Model of notch positions B) Raw low-energy data showing the characteristic notch pattern for two precursors having different m/z values.

The encoded data were first smoothed in the RT dimension, and each scan of the resulting data was subsequently decoded relative to the model (Figure 2A) using a non-negative least squares (NNLS) approach. The resulting data was stored in a custom compressed file format for efficient access. Faster direct (regularised) inversion methods were tried and found to yield significantly inferior results. The fully decoded data set has three analytical dimensions (RT, precursor m/z, fragment m/z) and resembles scanning quadrupole DIA data (SONAR). Routines were written to allow calibration of the m/z dimensions of the decoded data, efficient extraction of chromatograms and reconstruction of spectra for chosen precursor m/z ranges.

Owing to the binning employed, the decoded data has fewer bins in the ToF m/z dimension than the original dataset. Digitization could be improved, but with a corresponding increase the decoding time required. It was therefore decided for this preliminary study to use information from the relatively low resolution decoded data to refine the original processed high resolution MS^E data.

Software was written to read product ion spectra from the MGF files produced by PLGS. For each of these spectra, a corresponding low-resolution decoded product ion spectrum was obtained at the appropriate retention time and precursor m/z. Where the precursor had significant abundance at two charge states, two decoded spectra were formed and combined. Decoded data from both collision energies were also combined at this stage. The decoded data was thresholded at a level corresponding approximately to a single ion arrival.

Each fragment m/z value from the MGF file was then compared with the corresponding decoded product ion spectrum. Any fragments lying in bins having zero intensity in the thresholded decoded data were removed from the list, and a filtered MGF file was generated.

Databank Searching

The original and filtered MGF files were searched against the Ecoli K12 databank using the online MASCOT MS/MS ion search[1]. A precursor tolerance of 8ppm and a fragment tolerance of 15ppm were used.

RESULTS

Figure 3 shows the original and decoded data at the elution time of a peptide from malate dehydrogenase. The diagonal line in Figure 3B corresponds to unfragmented precursors while the vertical bands in Figure 3D are fragments sharing common precursor m/z encoding. These plots are similar to plots obtained directly from raw SONAR data.

The number of identifications obtained from the Mascot MS/MS search of the unfiltered MS^E data was lower than would normally be expected from an MS^E based search. This is partly due to the fact that collision energy could not be ramped in this experiment and only one function (with the lower 22V CE) was used to produce the initial MGF file. However the decrease is mainly a consequence of the complexity of the time-aligned MS/MS spectra: an MS^E search in PLGS returned more protein identifications. Our aim is to investigate whether the scanning notch encoding method can increase specificity of DIA experiments, so we focus here on relative rather than absolute performance.

The search of the filtered MGF file produced 33% more protein identifications than the original unfiltered version. For the top scoring protein (glycerol kinase), one additional peptide was identified, and the peptide scores for more than 75% of peptides in common between the two results were increased. Overall, the number of peptide sequence matches above the homology threshold was increased by over 40%. Mass accuracy for annotated precursors and fragments was identical because the filter simply removes existing peaks. However the simplification produced by the filtering process resulted in the annotation of some additional fragments by Mascot.

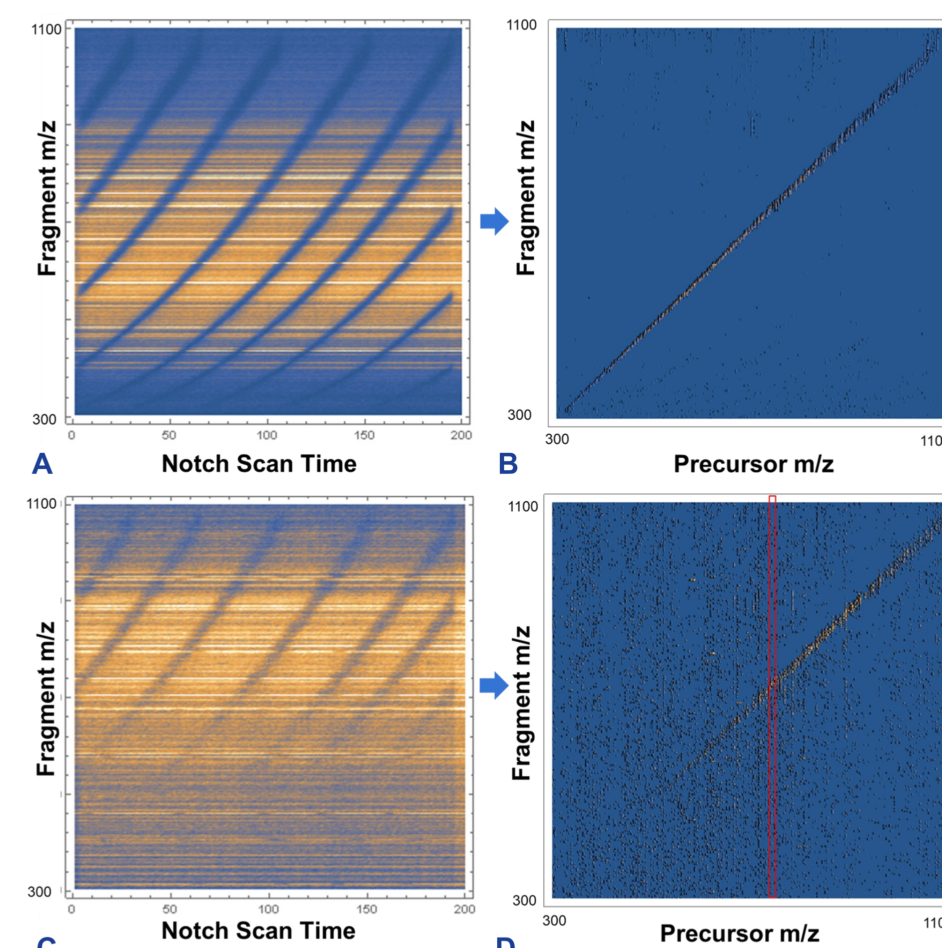

Figure 3A) Raw low energy data B) Deconvolved low-energy data C) Raw high-energy data D) Deconvolved high-energy data. The boxed region contains the MS/MS spectrum from the 2+ precursor of the pe tide SDLFNVNAGIVK.

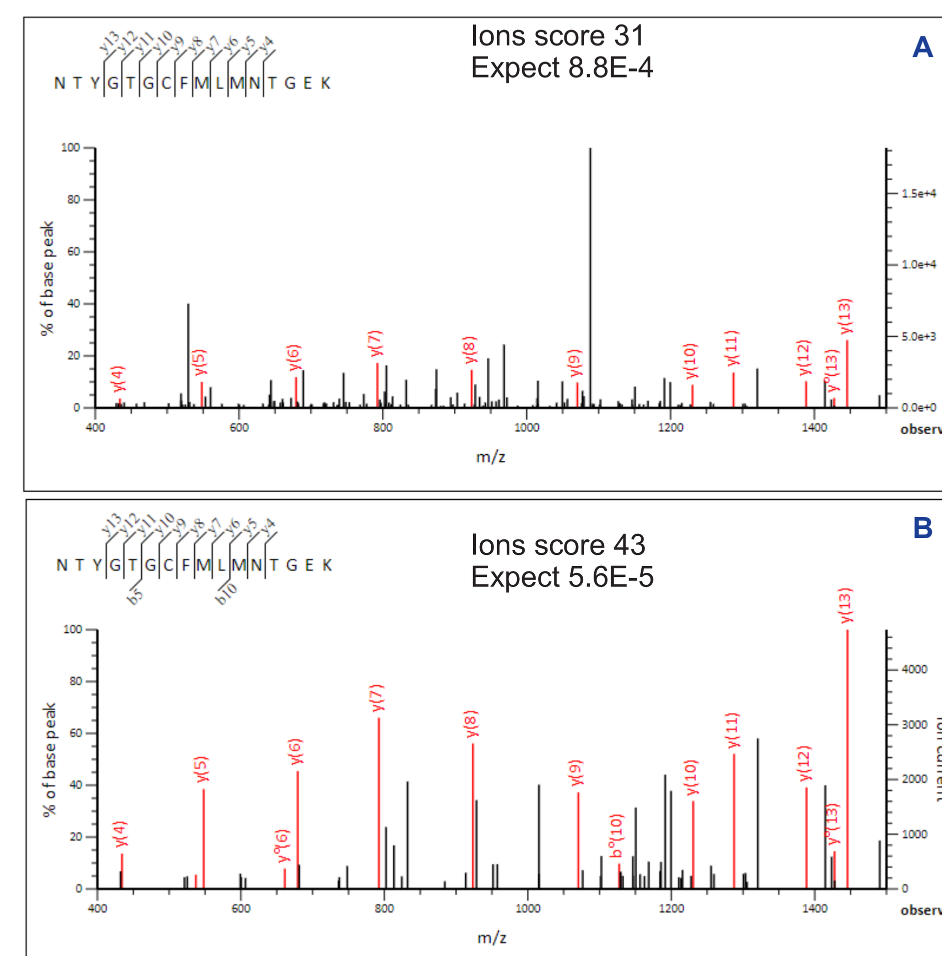

Figure 4. Annotated Mascot match for the peptide NTYGTGCFMLMNTGEK from glycerol kinase A) The product ion spectrum produced directly from RT-aligned MS^E data B) following application of the filter.

Figure 4 compares the annotated sequence match for a peptide from glycerol kinase between the original and filtered datasets. The signal to noise of the filtered data is markedly improved, and a number of additional fragments have been annotated by Mascot.

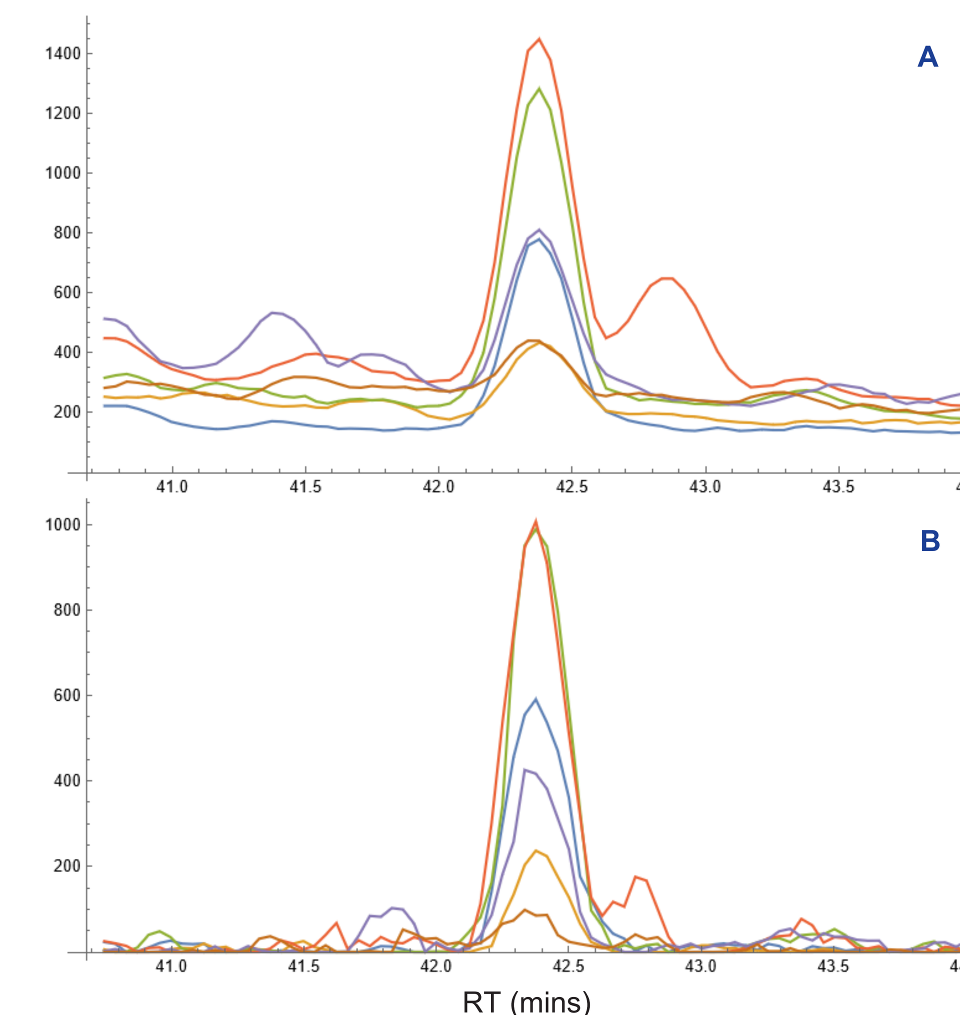

Figure 5 Extracted ion chromatograms generated from deconvolved data for six fragments of the peptide EMLI-ADGIDPNELLNSLAAVK from the DNA-binding protein H-NS. A) all precursor mass bins B) from central precursor mass bin +/- 3 bins.

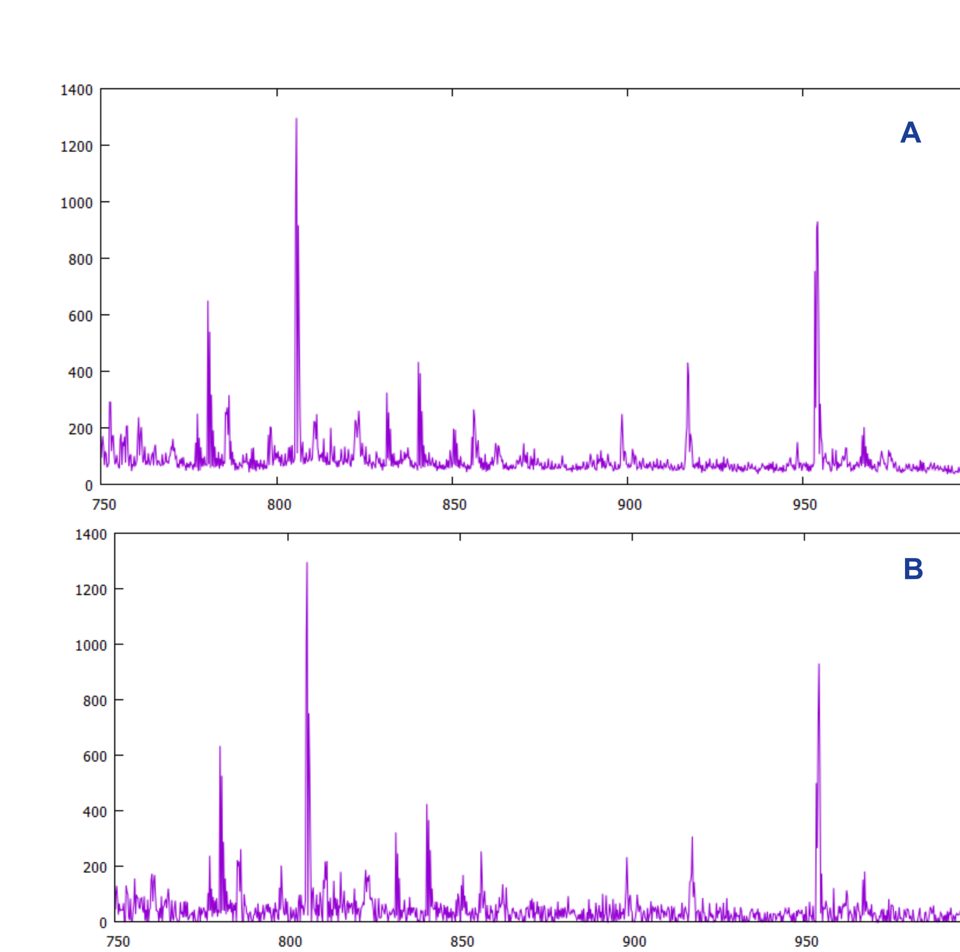

Figure 6. A) A portion a of a single decoded low-energy spectrum summed over all precursor m/z bins B) The same region of data following application of a filter that admits only data that has a notch encoding that matches the measured ToF m/z value.

Figure 5 shows extracted ion chromatograms from filtered and unfiltered data for six fragments of a peptide from the DNA-binding protein H-NS. In all cases, the baseline is significantly reduced, and the signal to noise is generally improved. Interestingly, a similar improvement can be obtained for extracted precursor ion chromatograms. To investigate this effect further we applied a scanning notch based filter to a complete decoded low energy spectrum. The effect of this filter is to retain only data from the low energy experiment for which the quadrupole encoding matches the measured ToF m/z. This corresponds to a selection along the prominent diagonal line of unfragmented precursors in the example in **Figure 3B**. **Figure 6A** and **6B** show the original unfiltered and filtered data. The filtered data has a noticeably reduced baseline compared with the original data.

DISCUSSION

It is clear from the results presented here that quadrupole-based precursor encoding can be used to improve the association between precursors and fragments in DIA analyses of complex mixtures. Simpler reconstructed product ion spectra can be obtained, improving the quality and confidence of databank search results, and the improved signal to noise of extracted ion chromatograms should benefit quantitative analysis. Some degree of cleanup of low energy data also seems to be possible. This must be a result of the rejection of species formed after the quadrupole, perhaps in part due to continued desolvation or breakup of solvent clusters.

It should be emphasised that this improvement in specificity has been obtained with a very modest compromise in sensitivity: only about 25% of the ions that would be present in the corresponding MS^E experiment have been removed. Furthermore, by ignoring the encoding dimension, the data can be processed as ordinary MS^E data by existing software.

One of the questions raised at the beginning of this work was whether the complexity of the high energy data would severely limit the dynamic range of fragments that could be successfully decoded. It is clear from **Figures 4** and **5** that many small fragments are successfully recovered. Although some annotated fragments were removed from the Mascot results, the improvement in signal to noise for those that remain more than compensates for this.

In order to realise the full potential of this technique it will be necessary to fully combine information obtained with different collision energies. This will allow a more systematic determination of the accessible dynamic range. It will also be interesting to explore higher resolution deconvolution using hardware acceleration (e.g. GPU based decoding). The high data rates produced by this technique mean that it would be desirable to perform decoding in real time, since the decoded data is considerably sparser than the original dataset (compare for example **Figures 3C** and **3D**).

CONCLUSION

- **m/z selective intensity encoding using a pattern of scanning notches significantly improves specificity of MS^E data in the analysis of a complex sample**
- **Improvements are observed in databank search results, both through increased protein identifications and peptide scores, extracted ion chromatograms, and a reduced baseline in low energy data.**
- **These improvements are obtained with a minimal reduction of sensitivity relative to conventional MS^E data.**

References

1. https://www.matrixscience.com/

TO DOWNLOAD A COPY OF THIS POSTER, VISIT WWW.WATERS.COM/POSTERS

Waters, Xevo, nanoACQUITY, UPLC, and CSH are trademarks of Waters Technologies Corporation. Mascot is a trademark of Matrix Science Ltd.
©2022 Waters Corporation