Waters

A Systematic Study into the Reproducibility of HPLC and Mass Spectrometry for the Analysis of Complex Protein Tryptic Digests by LC-MS

Chris Hughes[1], Roy Martin[2], Therese McKenna[1], Iain Campuzano[1] and Jim Langridge[1]
[1]Waters MS Technologies, Manchester, UK, [2]Waters Corporation, Beverly, MA.

Overview

- Previously we have shown that relative quantification of peptides across samples can be accomplished by LC-MS, without the use of isotope labeling
- In this paper we have investigated the chromatographic and mass spectrometric reproducibility that can be obtained for the LC-MS analysis of complex peptide mixtures and studied the effect of these parameters on the matching of peptide ions between runs.
- Automated algorithms were used to compare and contrast the peptide ions that could be matched across multiple injections
- Finally, we have compared LC-MS data acquired at mass resolutions of >10000 FWHM and >17500 FWHM

Introduction

The coupling of liquid chromatography with mass spectrometry is now firmly established as a routine method for the identification of proteins that have been subjected to enzymatic digestion. In an on-line LC-MS experiment, the column eluent is coupled to the electrospray source via an emitter with tryptic peptides mass analyzed as they elute from the HPLC column. Should there be any co-eluting species in the HPLC eluent, these will be separated in the mass analyzer by their mass-to-charge (m/z) ratio.

It has become increasingly clear that relative quantification of protein expression changes is important in modern biology and medicine. Several current approaches have been developed that utilize stable isotope labeling of samples in combination with tryptic digestion, separation and subsequent analysis by mass spectrometry [1-2]. However, we have recently described an LC-MS strategy [3] where quantification is achieved via normalization of the LC-MS datasets and comparison of the peptide intensities (of the observed tryptic peptides) across samples is performed. In this type of experiment, it is desirable to perform replicate injections and this places a requirement upon good chromatography, especially in terms of retention time reproducibility. In addition, exact mass measurement of the eluting ions is required as well as the generation of reproducible and reliable peak intensity, or area, calculations for the eluting tryptic peptides. The ability to measure the mass to charge ratios of ions accurately, between injections and across samples, increases the confidence that the same ions have been matched from each sample injection.

In this study, we have investigated the reproducibility of an LC-MS system to determine the levels that are required, and that may be expected. In particular we have investigated chromatographic reproducibility (at the 100µM and 75µM scales), on the clustering of known peptide ions, from run-to-run using an automated algorithm. The effect of changing HPLC column, and also the complete analytical system (LC and MS) on the number of ions and the intensity of the ions that can be identified and matched has been investigated. In addition, we have studied the effect of mass spectrometer resolution on the mass measurement accuracy obtained and the number of peptide species that may be confidently assigned.

Methods

Nanoscale HPLC (LCMS reproducibility)

- Waters® nanoACQUITY UPLC™ System configured in direct mode with a nanoscale analytical column, as detailed below.
- Analytical column: Waters NanoEase™ Atlantis™ dC18, 100µM x 100mm.
- The HPLC pump was operated at a flow rate of 400nL/min.
- The analytical column was directly coupled to a NanoLC sprayer, attached to the Nanolockspray ESI source of the Q-Tof mass spectrometer.
- A= Aqueous 0.1% formic acid; B= acetonitrile + 0.1% formic acid
- Gradient from 3%B to 40%B in 30-minutes.
- Then to 85% B and subsequently reduced back to 3% to condition the column for 30 minutes prior to injecting the next sample.

Nanoscale HPLC (EMRT analyses / MS Stability)

- Waters CapLC™ XE system configured in trapping mode with a C18 precolumn and a nanoscale analytical column, as detailed below.
- Trapping column 320µm ID x 5mm, Waters Symmetry® C18
- Analytical column: Waters NanoEase Atlantis dC18 3mM 75µM x 100mm.
- The HPLC pumps were operated at a flow rate of 5µL/min whilst a flow splitter gave a resultant flow through the analytical columns of 300 nL/min.
- The analytical column was directly coupled to a NanoLC sprayer, attached to the NanoLockSpray™ ESI source of the Q-Tof™ mass spectrometer.
- Samples were injected via an autosampler and after a three minute loading / washing period, the gradient was started
- A= 95% water / 5% acetonitrile + 0.1% formic acid; B= 95% acetonitrile / 5% water + 0.1% formic acid
- Gradient from 2%B to 40%B in 90 minutes.
- Then to 90%B and subsequently reduced back to 2% to condition the column for 20 minutes prior to injecting the next sample.

Mass spectrometry

- Three different mass spectrometers were used during the course of this study; Q-Tof Ultima™ Global, Q-Tof Ultima API and Q-Tof Premier™ (Waters, Manchester, UK).
- All mass spectrometers were operated with a NanoLockSpray source to provide the exact mass of all eluting peptide species.
- The reference probe of the NanoLockSpray source was set up to continually infuse a solution containing Glu-Fibrinopeptide b $[M+H]^{2+}$=785.8426 amu.
- Data was acquired using an alternating low collision energy (8eV) and elevated collision energy function. During the course of the elevated energy scan the collision energy was stepped from 23 to 33eV.
- An integration time of 1.5 seconds was used for each scan. The reference, lock mass, channel was sampled every 30 seconds.
- Data was acquired with the oa-Tof operating in either the V-mode of operation, at a resolution of >10000 FWHM, or in the W-optics mode of operation with a mass resolution of >17500 FWHM.

Samples

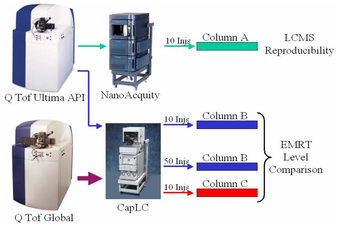
Samples used in the experiments were standard tryptic digests of the proteins Yeast Enolase and Alcohol Dehydrogenase (ADH). Yeast Enolase was also analysed after doping into a complex tryptic digest from an *E. Coli* cytosolic cell fraction (Waters, Milford, MA).

Bioinformatics

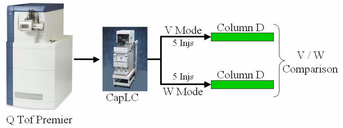
- The LC-MS data contains the exact masses of all detectable peptide molecular ions.
- In addition the data from the parallel high-energy function provides information about the amino acid sequence for the peptide.
- Interrogation of the acquired raw data, via an automated algorithm [3], extracts all of the relevant peptide information, including the de-isotoped monoisotopic exact mass, HPLC retention time and peak intensity/ area for each detectable peptide. This is known as an Exact Mass Retention Time signature (EMRT)
- Subsequent comparison of these EMRT signatures across injections (replicates) and conditions (different samples) can be performed via bioinformatics, and this provides a list of the EMRT "clusters". The algorithm used to match the EMRT's can take account of variability in retention time reproducibility, and mass measurement accuracy from experiment to experiment.
- The data from replicate runs can be interrogated to provide information about the total number of ions and, from this, the number of clustered ions replicating across injections can be calculated.

System reproducibility and Stability tests

- LCMS reproducibilithas been examined by performing replicate injections of Yeast Enolase on the nanoACQUITY UPLC system, **Figure** y **1**:



- Three experimental conditions have been examined to determine the reproducibility of the LC-MS systems at the EMRT level **Figure 1**. The sample used in each of these experiments was a tryptic digest mixture containing 100fmol of both Yeast Enolase and ADH.
- The stability of the mass measurement accuracy was investigated by measurement of peptide ions eluting from the column during injections every 2 hours, over a 24 hour period.
- To determine the effect of resolution on the peptide signatures, several injections of a Yeast Enolase tryptic digest contained in a complex *E. Coli* cell lysate were carried out on the Q-Tof Premier at >10000 FWHM (V mode) and >17500 FWHM (W mode), **Figure 2**:



Results

LC-MS Reproducibility

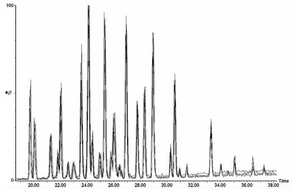


Figure 3. Overlaid Base Peak Intensity (BPI) chromatograms obtained from the replicate LC-MS analyses (n=4) of a tryptic digest of 250fmol Yeast Enolase. Data was obtained on a nanoACQUITY UPLC system.

LC-MS data threshold

In processing the LC-MS data to identify the peptide related ions, one of the most critical parameters is the threshold applied to the data. This allows chemical noise to be removed, whilst providing detection of the peptide related signal. Two approaches to this signal thresholding have been implemented. The first is to set a manual fixed threshold above which an ion must rise to be considered as significant. The second approach is to use an automatically determined threshold that is set, dependant upon the data quality on a spectrum-to-spectrum basis.

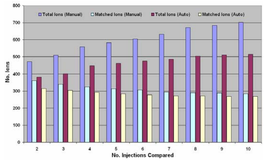
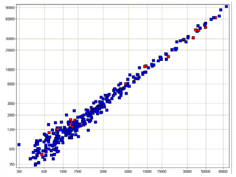


Figure 4 shows a comparison between the two approaches, where in the first situation the threshold is set manually at a fixed value and, in the second, automatic determination of the threshold on a spectrum-to-spectrum basis is employed. The data presented highlights that even after 2 injections good reproducibility of matched ions between runs can be obtained. After 3 injections the number of peptides identified and matched is well characterized. It can also be seen that the use of auto thresholding is efficient at reducing the number of ions that don't match between injections.

System reproducibility

Figure 5 is an intensity plot (ln) of the matched peptide (EMRT) components between two injections of the Enolase and ADH tryptic digest. The data presented shows good correlation between the intensity of the two injections.



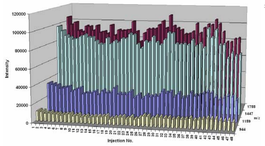
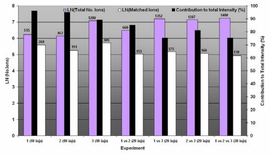


Figure 6 The intensity variation of four selected peptide ions over 48 injections, equivalent to approximately 104 hours. It can be seen that even prior to normalization, the intensities are extremely reproducible over this extended time period.

Figure 7 Display of the number of EMRT's identified across replicate injections. When 10 injections are combined for each experiment, there appears to be a large number of peptides that do not appear in every run.

However, closer examination of the peptides by calculating the percentage contribution of intensities of the matched ions to the total intensity of all ions, suggests that the matched ions are the most abundant species and the unmatched ions are of low intensity. This also appears to be the case when comparisons are made across the three experimental conditions.



Mass analyzer stability

An important factor in the matching of ions across multiple runs and over extended time periods is the stability of the mass analyzer, both in terms of mass measurement accuracy and resolution.

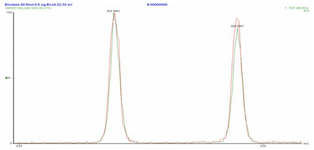


Figure 8 The stability of the mass spectrometer resolution. Displayed is a doubly charged ion from a tryptic peptide of an *E.coli* protein, shown 28 hours apart. In addition the resolution changes over this time period are minimal.

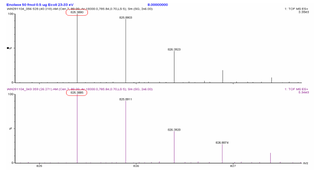
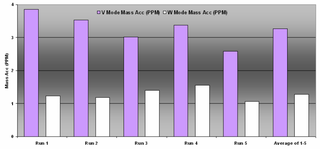


Figure 9 The lock mass corrected mass for a doubly charged ion and the mass difference observed over the 28-hour period, (–0.5mDa).

Comparison of V to W-OPTICS mode on the Q-Tof Premier.

Five injections of Yeast Enolase in an *E. Coli* tryptic digest were performed, in both the V and W-mode, to compare the protein identifications obtained from the mass spectrometer.

Figure 10 displays the mass accuracies obtained by averaging the mass errors from the two highest scoring peptides in the top ten proteins identified in each of the ten injections. Clearly the W-mode mass accuracies are superior to the V mode measurements.



	V Mode		W Mode	
Injection	No. Peptides	Score	No. Peptides	Score
1	17	313	15	426
2	15	300	15	417
3	18	325	14	505
4	15	218	17	487
5	16	335	16	490
	Total = 81		Total = 77	

Figure 11 A comparison of the databank searching results obtained in the V and W-mode. Shown is the results for Yeast Enolase, whilst present in the complex *E. coli* mixture. Displaying the total number of peptides identified in W-mode compared to V-mode. The total number of peptides is lower in W-mode, however, the improved mass measurement accuracy in W-mode leads to considerably higher protein scores and hence more confident identification. By comparing the Enolase identifications from injection 3 in both cases, the peptides not apparent in W mode but identified in the V mode are of low intensity and provide very little fragment data.

Conclusions

- The chromatographic reproducibility exhibited by several experimental conditions and the Protein Expression algorithm allows data to be compared across the different platforms.
- The peptides matched across the experiments are the most abundant ions that elute from the columns.
- Excellent mass analyzer stability is important in allowing the EMRTs to be generated and matched.
- 'W-mode' optics, where the mass spectrometer resolution is >17500 FWHM, leads to better mass measurement than 'V-mode' optics at a resolution of >10000 FWHM. This consequently leads to more confident assignments of protein identifications.

References

[1] Gygi SP, Rist B, Gerber SA, Turecek F, Gelb MH, Aebersold R.; Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat. Biotechnol.* 1999 Oct.17(10): 994-9
[2] T.P. Conrads, K. Alving, T.D. Veenstra, M.E. Belov, G.A. Anderson, D.J. Anderson, M.S. Lipton, L. Pasa-Tolic, H.R. Udseth, W.B. Chrisler, B.D. Thrall and R.D. Smith, *Anal. Chem.* **73** (2001), p. 2132
[3] T. McKenna et al; Proceedings of the 52nd ASMS conference, Nashville, TN

©2004 Waters Corporation 720001114EN