Large-scale analysis of peptide sequence variants: The case for high field asymmetric waveform ion mobility spectrometry

Andrew J. Creese, Jade Smart and Helen J. Cooper*

Supplemental Files:

Supplemental information: Compensation Voltage Range Optimisation.

Supplemental Figure 1 a: Compensation voltage spectrum (extracted ion chromatogram) (single replicate) of phosphopeptide [APLSFRGpSLPKpSYVK]$^{2+}$ ions. CV values for the peak maxima were determined by Gaussian fitting. **b**: Mean compensation voltages (n=10) versus *m/z* value for the phosphorylated (solid diamonds) and non-modified (hollow diamonds) peptides. Each box relates to one peptide and contains data points corresponding to the maxima of the different FAIMS peaks. Error bars represent standard deviation (n=10).

Supplemental Figure 2: The compensation voltage spectra (extracted ion chromatograms) (n=10) for the peptide [APLSFRGpSLPKpSYVK]$^{2+}$.

Supplemental figure 3: Strong cation exchange UV chromatogram of the combined phosphopeptide libraries.

Supplemental figure 4 a: % amino acid composition at the variable sites in the peptides identified across the compensation voltage range. **b:** % amino acid composition at variable site 1 (nearest N-terminus). **c:** % amino acid composition at variable site 2 (central). **d:** % amino acid composition at variable site 3 (nearest C-terminus).

Supplemental table 1: Number of isomers and the corresponding number of unique peptide masses.

Supplemental table 2: The 322 unique peptides identified in the RPLC MS/MS experiment.

Supplemental table 3: The 1212 unique peptides identified in the SCXLC RPLC MS/MS experiment when each raw file was analysed separately.

Supplemental table 4: The 693 unique peptides identified in the SCXLC RPLC MS/MS experiment when all 9 raw files were analysed together.

Supplemental table 5: The 2097 unique peptides identified in the FAIMS RPLC MS/MS experiment when each raw file was analysed separately.

Supplemental table 6: The 2495 unique peptides identified in the FAIMS RPLC MS/MS experiment when all 9 raw files were analysed together.

Supplemental table 7: Fragment ions identified from all of the possible isomers of the peptide GPSGYpSAAQLYK.

Compensation Voltage Range Optimisation.

Experimental

Protein digestion

Alcohol dehydrogenase (yeast), cytochrome C (equine) and serum albumin (bovine) were purchased from Sigma Aldrich (Leicester, UK). Proteins (27 mM of alcohol dehydrogenase, 85 mM of cytochrome C and 14 mM of serum albumin) were reduced with 50 mM dithiothreitol (Sigma) and alkylated with 20 mM iodoacetamide (Sigma Aldrich) and digested overnight at 37 ^{o}C with trypsin Gold (Promega, Madison, USA). The tryptic peptides were diluted to concentrations of 1.3 µM, 4.1 µM and 0.7 µM (alcohol dehydrogenase, cytochrome C and serum albumin respectively) in methanol:water (70:30) (JT Baker, Netherlands) with 2% formic acid (Sigma Aldrich). The following tryptic peptides were used for optimisation of FAIMS compensation voltage (CV) range: YSGVCHTDLHAWHGDWPLPTK (3+), ATDGGAHGVINVSVSEAAIEASTR (2+), LFTFHADICTLPDTEK (3+), SANMAGHWVAISGAAGGLGSLAVQYAK (2+), GITWEETLMEYLENPK (2+), LGEYGFQNALIVR (2+), KTGQAPGFSYTDANK (3+), VVGLSTLEIYEK (2+), TVMENFVAFVDK (2+), TGPNLGHGLFGER (3+), SISIVGSYVGNR (2+), RPCFSALTPDETYVPK (3+), LVNELTEFAK (2+), DAIPENLPPLTADFAEDK (2+) and ANELLINVK (2+). These peptides were the 15 highest scoring peptides from a SEQUEST search of protein digests analysed by direct infusion MS/MS.

Synthetic peptides

Thirteen individual phosphopeptides were synthesised (Alta Biosciences, Birmingham, UK): APLsFRGSLPKSYVK (3+), APLSFRGsLPKSYVK (3+), APLSFRGSLPKsYVK (3+), APLSFLGSLPKsYVK

(3+), APLsFLGSLPKSYVK (3+), LLGssFSsGPVADGIIR (3+), LLGssFsSGPVADGIIR (3+), LLGssFSSGPVADGIIR (3+), LLGSSFssGPVADGIIR (3+), INNIDYyKKTTNG (2+), INNIDyyKKTTNG (2+), APLSFRGsLPKsYVK (2+, 3+) and HYGItSPISLAAPK (2+, 3+) (lower case s,t,y indicate site of phosphorylation). Peptides LFtGHPESLER (2+) and LFTGHPEsLER (2+) were a gift from Professor Gavin Reid. Each peptide was resuspended in methanol:water (70:30) with 2% formic acid to a final concentration of 2 µM. The peptides were analysed in three groups of 8. Group 1: APLsFRGSLPKSYVK, APLSFLGSLPKsYVK, LLGssFSsGPVADGIIR, LLGssFSSGPVADGIIR, INNIDYyKKTTNG, APLSFRGsLPKsYVK, HYGItSPISLAAPK and LFtGHPESLER. Group 2: APLSFRGsLPKSYVK, APLsFLGSLPKSYVK, LLGssFsSGPVADGIIR, LLGSSFssGPVADGIIR, INNIDyyKKTTNG, APLSFRGsLPKsYVK and HYGItSPISLAAPK. Group 3: APLSFRGSLPKsYVK, APLSFLGSLPKsYVK, LLGssFSsGPVADGIIR, LLGssFSSGPVADGIIR, INNIDYyKKTTNG, APLSFRGsLPKsYVK, HYGItSPISLAAPK and LFtGHPESLER. These peptides were used to optimise the CV range.

Direct infusion FAIMS analysis

The peptides were infused into the mass spectrometer (Thermo Fisher Velos Orbitrap ETD) via the HESI ion max source. Peptides were infused at a flow rate of 3 µL/min with a spray voltage of 4 kV. The sheath gas was set to 5 (arbitrary units) and the auxiliary gas was set to 2. The capillary temperature was 275 ^{o}C. All spectra were recorded in the Orbitrap at a resolution of 100,000 (at *m/z* 400). Each mass spectrum consisted of 1 microscan. Automatic gain control (AGC) target was 1,000,000 with a maximum fill time of 1,000 ms.. The FAIMS parameters were as follows: Gas flow rate of 3.5 L/ min and a composition of 50:50 He/N. The dispersion voltage (DV) was set to -5000 V and inner and outer electrode

temperatures were 70^{o}C and 90^{o}C respectively. The dwell time was set to 50 ms. The compensation voltage (CV) was scanned from -60 V to -10 V in 0.3 V steps.

PeakFit analysis

Extracted ion chromatograms were imported directly into Peakfit 4.12 (Systat Software) from Xcalibur V2.1.0.1139 (Thermo Fisher Scientific). Peak fitting was performed using the residual method within the software. The data were smoothed with an FFT function (20% smoothing). Multiple Gaussian peaks were fitted to the raw data to give the greatest correlation (r^2 >0.9) with the fewest number of peaks.

Results

Optimising CV range

We have shown previously[1] that the optimum FAIMS method for the analysis of complex peptide mixture is an 'external CV stepping' method in which multiple LC MS/MS analyses are performed at fixed CV values. To determine the optimum CV range for the analysis of phosphopeptides, replicate (n=10) direct infusion FAIMS analyses of 15 synthetic phosphopeptides and 15 non-phosphorylated tryptic peptides were performed. Gas-phase peptide ions are often present as multiple conformers resulting in multiple, and often overlapping, FAIMS peaks[2, 3]. To determine the maximum CV of the FAIMS peaks for the individual (phospho)peptides, the data was manually analysed by approximation to a Gaussian distribution by use of Peakfit software (Systat Software). (Note that to correct for the possibility that CV peaks from cylindrical FAIMS devices can deviate from a Gaussian shape[4], a tolerance of r^2 >0.9 was applied). **Supplemental figure 1a** shows an extracted ion

chromatogram (single replicate) of the synthetic phosphopeptide [APLSFRGpSLPKpSYVK]$^{2+}$ ion (dots) along with the smoothed fit (solid line) and peaks as fitted by Peakfit (dashed lines). (For replicate analyses, see **supplemental figure 2**. Analysis of each extracted ion chromatogram for this peptide (n=10) identified two maxima with average CV values of -25.11 V and -22 V. **Supplemental figure 1b** shows the average CV values for the individual peptides (phosphorylated, solid diamonds and unmodified, hollow diamonds). The optimum CV range for these phosphopeptides is -45 to -25 V. This CV range matches well with that used by Bridon *et al*[5] (-24 to -42V) in their FAIMS analysis of phosphopeptides from *Drosophila melanogaster*. A CV range of -45 to -25 V was used for all subsequent FAIMS experiments.

1. Creese, A. J., Shimwell, N. J., Larkins, K. P. B., Heath, J. K., and Cooper, H. J. *J Am Soc Mass Spectr.* **2012**, *DOI: 10.1007/s13361-012-0544-2*,

2. Shvartsburg, A. A., Tang, K. Q., and Smith, R. D. *Anal Chem.* **2010**, *82*, 32-35.
3. Purves, R. W., Barnett, D. A., Ells, B., and Guevremont, R. *J Am Soc Mass Spectr.* **2001**, *12*, 894-901.
4. Guevremont, R., and Purves, R. *J Am Soc Mass Spectr.* **2005**, *16*, 349-362.
5. Bridon, G., Bonneil, E., Muratore-Schroeder, T., Caron-Lizotte, O., and Thibault, P. *J Proteome Res.* **2012**, *11*, 927-940.

Supplemental Tables:

Number of isomers	Unique peptide masses
1	28
3	170
4	12
6	116
9	122
12	48
15	12
18	48

Supplemental table 1: The number of unique masses per number of isomers.

Ions	GPSG**Y**VpS**A**AQL**Y**K	GPSG**Y**VpS**Y**AQL**A**K	GPSG**A**VpS**Y**AQL**Y**K	GPSG**S**VpS**F**AQL**Y**K	GPSG**S**VpS**Y**AQL**F**K	GPSG**F**VpS**S**AQL**Y**K	GPSG**F**VpS**Y**AQL**S**K	GPSG**Y**VpS**F**AQL**S**K	GPSG**Y**VpS**S**AQL**F**K
c_{12}	1291.84	1291.84	1291.84	1291.84	1291.84	1291.84	1291.84	1291.84	1291.84
c_{11}	1128.67	***1219.81***	1128.67	1128.67	*1143.65*	1128.67	*1203.72*	*1203.72*	*1143.65*
$c_{10}^{\bullet}$	1014.73	**1107.79**	1014.73	1014.73	*1030.63*	1014.73	1091.77	1091.77	*1030.63*
c_9	887.67	-	887.67	887.67	-	887.67	-	-	-
c_8	-	***907.55***	-	-	-	-	-	-	-
$z_{11}^{\bullet}$	1250.86	1250.86	1250.86	1250.86	1250.86	1250.86	1250.86	1250.86	1250.86
z_{10}	1164.71	1164.71	1164.71	1164.71	1164.71	1164.71	1164.71	1164.71	1164.71
z_9	1107.79	1107.79	1107.79	1107.79	1107.79	1107.79	1107.79	1107.79	1107.79
$z_8^{\bullet}$	943.71	943.71	**1035.81**	-	-	*960.71*	*960.71*	943.71	943.71
z_7	845.69	845.69	-	-	-	861.59	861.59	845.69	845.69
z_6	678.54	678.54	***769.67***	754.66	754.66	694.52	694.52	678.54	678.54
z_5	607.39	-	607.39	607.39	-	607.39	-	-	**591.53**
z_4	536.4	**444.62**	536.4	536.4	-	536.4	-	-	-

Supplemental Table 7: Fragment ions identified from all of the possible isomers of the peptide GPSGYpSAAQLYK. Values in bold represent unique ions identified. Values in italics represent hydrogen transfer compared to the original peptide.