RAPID COMMUNICATIONS IN MASS SPECTROMETRY
Rapid Commun. Mass Spectrom. 2006; **20**: 1558–1562
Published online in Wiley InterScience (www.interscience.wiley.com) DOI: 10.1002/rcm.2476

Detection of artifacts and peptide modifications in liquid chromatography/mass spectrometry data using two-dimensional signal intensity map data visualization

Matthias Berg[1]*, Axel Parbel[1], Harald Pettersen[2], David Fenyö[2] and Lennart Björkesten[2]

[1]GE Healthcare, Oskar Schlemmer Str. 11, D-80807 München, Germany
[2]GE Healthcare, Björkgatan 30, Uppsala, Sweden

Received 27 December 2005; Revised 7 March 2006; Accepted 10 March 2006

We demonstrate how visualization of liquid chromatography/mass spectrometry data as a two-dimensional signal intensity map can be used to assess the overall quality of the data, for the identification of polymer contaminants and artifacts, as well as for the confirmation of post-translational modifications. Copyright © 2006 John Wiley & Sons, Ltd.

Liquid chromatography combined with mass spectrometry (LC/MS) has become an essential part of biological research in the post-genome era.[1] An LC/MS experiment commonly starts with the enrichment of a mixture of the proteins that are relevant to the biological hypothesis being tested. The proteins are digested and the resulting peptides are then typically separated by reversed-phase liquid chromatography (LC) to reduce the complexity of the analyte. Finally, the masses of the proteolytic peptides and their fragments are measured by mass spectrometry (MS). LC/MS is a powerful technique that can provide information on the identity and quantity of a fraction of the proteins in the sample. Protein identification is achieved by comparing the measured masses with calculated masses of peptides or their fragments derived from a protein sequence collection.[2] Scores are calculated for each comparison and the protein sequences in the collection are ranked according to the scores. Different search engines calculate the score in different ways. The search engine dependent scores can, however, be converted into a measure of the significance of the protein identification result, i.e. the probability that the identification is false.[3–5] LC/MS can also be used for quantitative sample-to-sample comparisons. One set of methods include labeling the samples with different isotopes and subsequently mixing them to minimize any variation in the sample handling and measurement.[6] Improvements in the stability of LC/MS systems have also allowed differential analysis of protein amounts without the use of isotopic labels.[7,8]

Many analytical tools for LC/MS data provide graphical views of the data including total ion chromatograms and single or average MS and MS/MS spectra. However, these graphs only allow a limited view of the rich and multifaceted data that is collected during an LC/MS run. Image visualization of mass spectrometric data has previously been used in combination with, e.g., two-dimensional (2D) gel[9] and immobilized pH gradient (IPG) strip[10] data. Image visualization of LC/MS data as a 2D signal intensity map has also been demonstrated.[11,12] In this paper we present examples of how such a visualization approach can be utilized to judge the overall success of an LC/MS experiment as well as to identify contaminants and artifacts as well as to confirm post-translational modifications.

EXPERIMENTAL

A mixture of three recombinant proteins (glucose oxidase, beta-lactoglobulin and triacylglycerollipase) was digested using trypsin and analyzed by one-dimensional LC/MS using an Ettan™ MDLC system (GE Healthcare) in high-throughput configuration directly connected to a Finnigan™ LTQ™ system (Thermo Electron). Samples were concentrated and desalted on RPC trap columns (Zorbax™ 300 SB C18, 0.3 mm × 5 mm; Agilent Technologies), and the peptides were separated on a nano-RPC column (Zorbax 300 SB C18, 0.075 mm × 100 mm; Agilent Technologies) using a linear acetonitrile (ACN) gradient (GE Healthcare, 1% ACN increase/min). All buffers used for nano-LC separation contained 0.1% formic acid (Fluka) as the ion pairing reagent. Full scan mass spectra were recorded in profile mode and tandem mass spectra in centroid mode.

DeCyder™ MS Differential Analysis software (DeCyder MS; GE Healthcare) was used to visualize and analyze the signal intensity maps from various LC/MS experiments. The LC/MS data was displayed as 2D intensity maps with m/z and retention time on the two axes and using a grey scale to represent the intensity of the ion current at each m/z and retention time point (darker at higher intensity).

*Correspondence to: M. Berg, Oskar Schlemmer Str. 11, D-80807 München, Germany.
E-mail: matthias.berg@ge.com

RESULTS AND DISCUSSION

The visualization of an LC/MS run as a signal intensity map gives a good overview of the sample complexity and was used to assess the overall quality of the LC separation with regards to variations in the background level before any further analysis was done. Rapid visual comparison of replicate runs can be used to quickly identify problems with individual LC runs (Figs. 1(a) and 1(b)) and to identify possible polymer contaminations like PEG (Fig. 1(c))–a contamination that results in a characteristic pattern in the intensity map which would not be easily found using conventional visualization.

The feature to overlay the MS/MS events on the signal intensity map in the software was used to optimize the data-dependent acquisition of MS/MS spectra in subsequent runs in order to maximize the number of different peptides that were selected for MS/MS analysis and to assess which peak in the isotopic distribution was selected for fragmentation (Fig. 1(d)).

The effect on peptides in the LC/MS analysis due to post-translational modifications[13] including oxidation, methylation, and deamidation was studied by visually inspecting the signal intensity map. The expected shift in the m/z value depends simply on the mass of the modification and the charge of the peptide, e.g. a peptide with one oxidized methionine shifts +16 or +8 if it is singly or doubly charged, respectively (for deamidation the corresponding shifts in m/z are +1 and +0.5 and for methylation +14 and +7). Deamidated and methylated peptides elute later than the corresponding unmodified peptide (Figs. 2(a) and 2(e)) while peptides with oxidized methionines elute earlier than the unoxidized form of the peptide (Fig. 2(c)). The shift in retention time[14] due to a modification can be predicted but with much lower accuracy than for the m/z shift. However, when both modified and unmodified forms of a peptide are present, the corresponding pattern in the signal intensity map can be used to easily confirm the database search results or to locate potential modifications. This information can then be utilized when searching protein sequence collections with the corresponding MS/MS spectra to increase the sensitivity and specificity of the database search result.

A partially deamidated peptide will result in a characteristic twin feature in the intensity map (Fig. 2(a)). The MS/MS spectra corresponding to the potential deamidation of SNGIEASLLTDPK (Fig. 2(b)) show that all the b-ions of the deamidated peptide have a 1 Da shift for all observed b-ions (including the b2 ion) indicating a deamidation of asparagine in position 2 from the N-terminal of the peptide. Figure 2(c) shows the shift of several minutes that is associated with oxidation of methionine. The corresponding and annotated MS/MS spectra for the unmodified peptide and the oxidized version of the peptide IIVTQTMK are shown in Fig. 2(d), confirming the oxidation. Looking at the

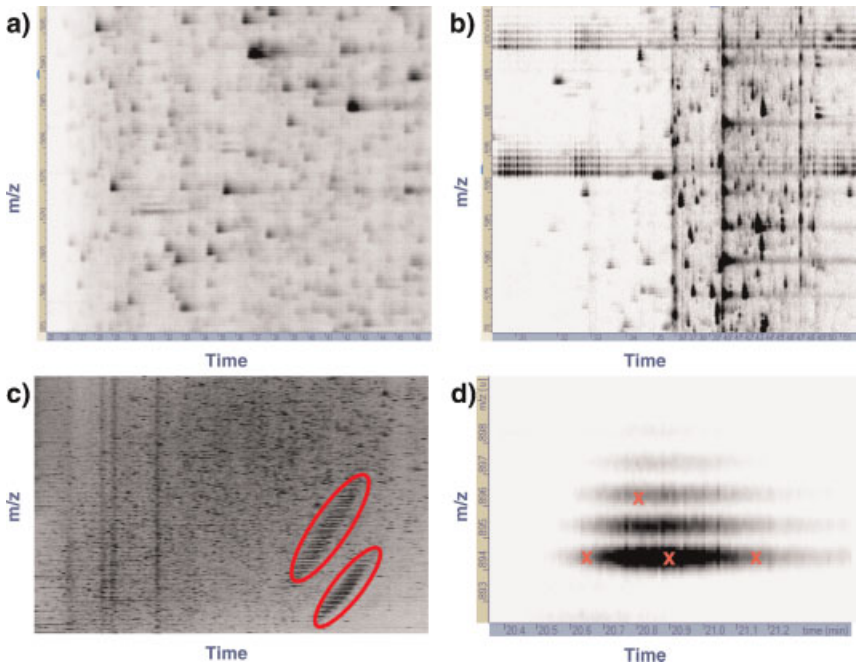

Figure 1. Visualization of LC/MS data as a signal intensity map can be used to assess the overall quality of the LC separation. Examples showing (a) well-optimized LC separation and (b) suboptimal LC separation. (c) Visualization of LC/MS data as a signal intensity map for the detection of polymer contaminants. The marked areas show contamination of the sample with PEG. (d) Detailed view of LC/MS data acquired by an LTQ instrument showing the isotopic envelope of a singly charged peptide. The crosses indicate the MS/MS events. In addition to the peptide ions containing only ^{12}C isotopes, also peptides containing two ^{13}C isotopes were fragmented. This may increase the probability of incorrect identification when searching a collection of protein sequences.

DOI: 10.1002/rcm

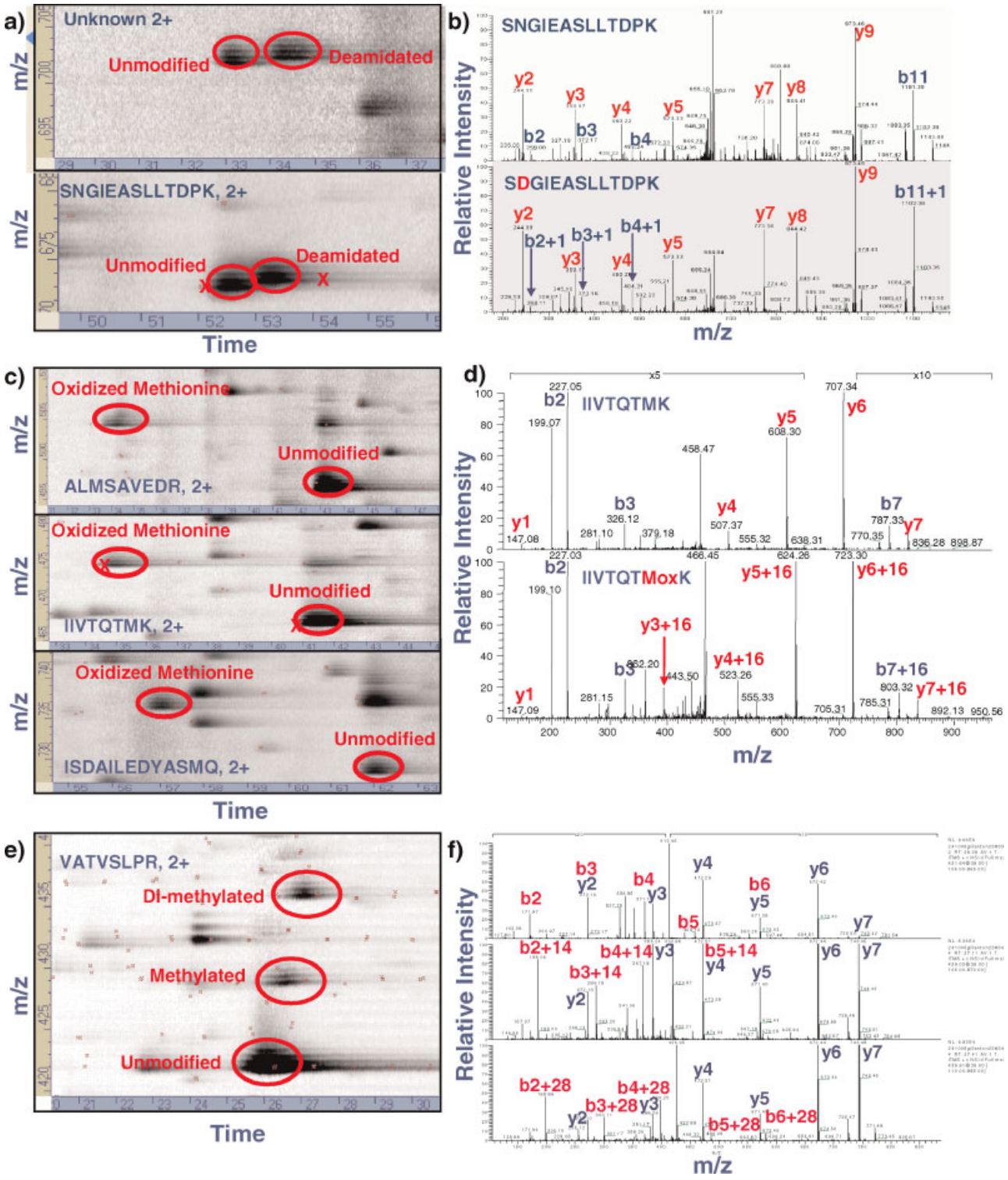

Figure 2. Visualization of LC/MS data as a signal intensity map for the detection of peptide modifications. (a) Details of two signal intensity maps illustrating the LC/MS behavior of two different peptides present in both unmodified and deamidated forms. (b) MS/MS spectra of the unmodified and deamidated form of the peptide with the sequence SNGIEASLLTDPK. The b-fragment ions of the deamidated peptide show a 1 Da shift for all observed b-ions (including b2) indicating a deamidation of asparagine in position 2 from the N-terminal of the peptide. (c) Details of a signal intensity map illustrating the reversed-phase chromatographic behavior of three peptides with and without oxidized methionine. (d) The MS/MS spectra for the oxidized and unmodified peptide with the sequence IIVTQTMK. (e) Details of a signal intensity map showing the LC/MS behavior of unmodified and the mono- and dimethylated forms of the peptide with sequence VATVSLPR. (f) The MS/MS spectra for the unmodified and the mono- and dimethylated forms of the peptide with sequence VATVSLPR.

DOI: 10.1002/rcm

Table 1. Deconvoluted masses and retention times for the unmodified and modified peptides in Fig. 2 and the corresponding mass and time shifts, Δ Mass and Δ Time

Unmodified peptide		Modified peptide					
Mass (Da)	Time (min)	Mass (Da)	Time (min)	Δ Mass (Da)	Δ Time (min)	Assignment	MS/MS
1343.8	52.70	1344.8	53.19	1.0	+0.49	Deamidation	Fig. 2(b)
1401.9	33.26	1402.6	34.48	0.7	+1.26	Deamidation	
990.5	43.33	1006.6	34.26	16.1	−9.07	Oxidation	
932.5	41.58	948.5	35.08	16.0	−6.05	Oxidation	Fig. 2(d)
1454.9	62.14	1470.8	57.12	15.9	−5.02	Oxidation	
841.5	26.18	855.5	26.56	14.0	+0.38	Methylation	Fig. 2(f)
841.5	26.18	869.5	27.1	28.0	+0.92	Dimethylation	Fig. 2(f)

mono- and dimethylation of the autocatalytic fragment from trypsin shows that due to the addition of methyl groups the peptides are eluting at later time points in the gradient (Fig. 2(e)). The corresponding MS/MS spectra are shown in Fig. 2(f).

A summary of the experimental observations is shown in Table 1. The three examples from the Met oxidation show clearly that the time shift between unmodified and modified peptides (between 5 and 9 min) is as expected very much sequence-dependent. It will therefore be difficult to identify these modifications automatically without getting many false positive hits, especially in more complex samples.

The electrospray ionization process following the chromatographic separation can introduce artifacts, including covalent modifications, non-covalent adduct formation and dimer formation as well as in-source fragmentation. These artifacts can be easily identified by visual inspection of the intensity map, because they co-elute with the unmodified precursor peptide. Figure 3(a) shows an example of a peptide that was identified to be a peptide from glucose oxidase

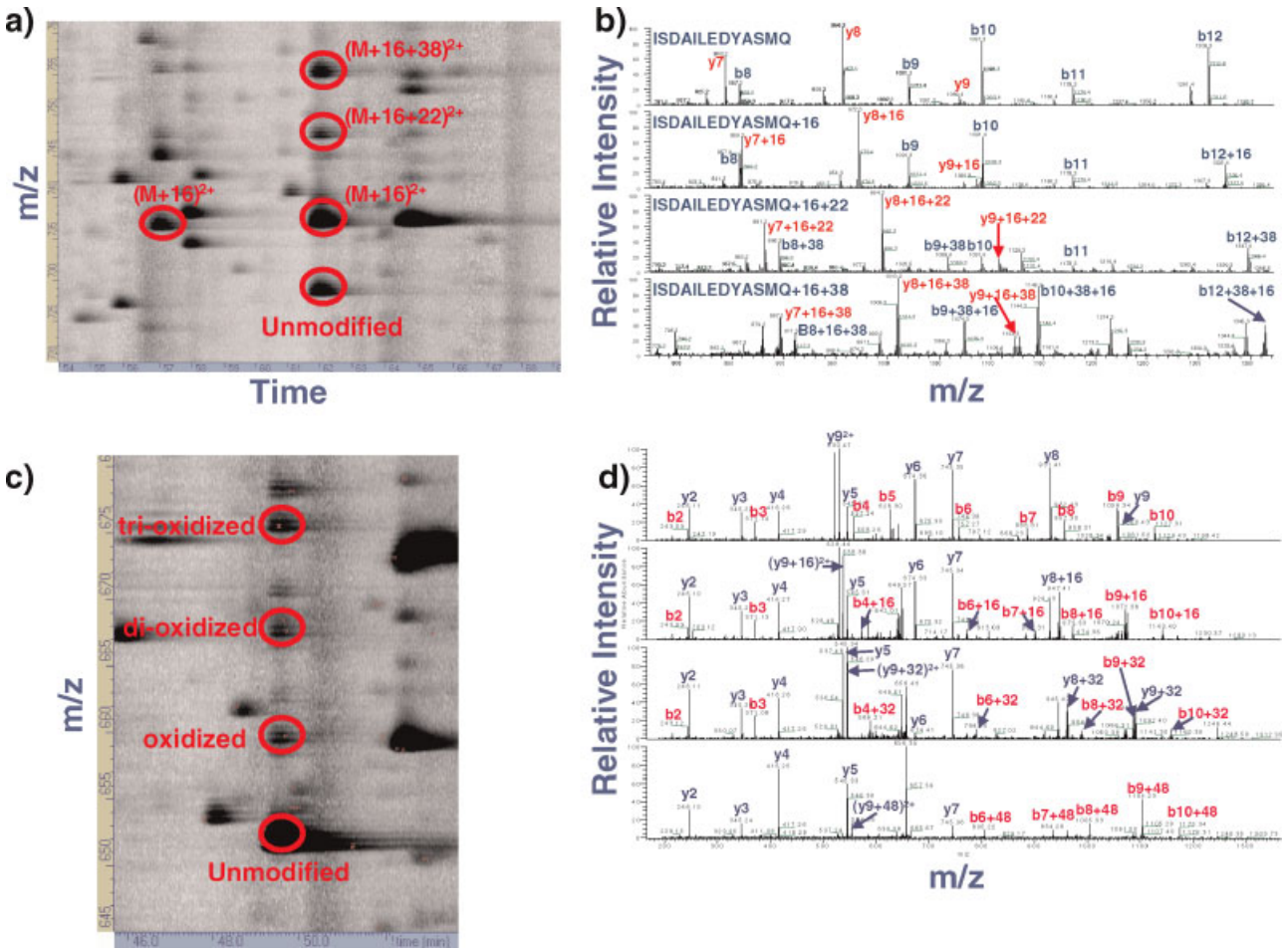

Figure 3. (a) Visualization of adducts formed during ionization of the peptide with sequence ISAILEDYASMQ. (b) Confirmation of adducts to the peptide with sequence ISAILEDYASMQ by MS/MS. (c) Visualization of tryptophan oxidation during ionization of the peptide with sequence LEQWAEEAVAR. (d) Confirmation of mono-, di-, and trioxidized tryptophan in the peptide with sequence LEQWAEEAVAR by MS/MS.

DOI: 10.1002/rcm

Table 2. Deconvoluted masses and retention times for the artifacts shown in Fig. 3 and the corresponding mass and time shifts, Δ Mass and Δ Time. In total 82 artifacts were automatically identified

Unmodified peptide		Modified peptide					
Mass (Da)	Time (min)	Mass (Da)	Time (min)	Δ Mass (Da)	Δ Time (min)	Assignment	MS/MS
1454.9	62.14	1471.5	62.13	16.6	−0.01	Monooxidation	Fig. 3(b)
1454.9	62.14	1492.9	62.12	38.0	−0.02	Potassium/sodium + monooxidation	Fig. 3(b)
1454.9	62.14	1508.6[a]	62.1[a]	53.7[a]	0.0[a]	Potassium + monooxidation/sodium + dioxidation	Fig. 3(b)
1300.5	49.67	1316.8	49.77	16.3	0.10	Monooxidation	Fig. 3(d)
1300.5	49.67	1332.8	49.69	32.3	0.02	Dioxidation	Fig. 3(d)
1300.5	49.67	1348.7	49.74	48.2	0.07	Trioxidation	Fig. 3(d)

[a] Data for this modification was manually estimated from the signal intensity map rather than automatically calculated by the DeCyder MS software.

having a sequence of ISAILEDYASMQ. At an earlier retention time the oxidized form of the peptide is observed. At the same retention time as the unmodified peptide, several peaks are observed and a closer examination of the corresponding MS/MS spectra (Fig. 3(b)) reveals that these are all artifacts created during the ionization process by modification and adduct formation involving the original peptide. Another example of an artifact is shown in Figs. 3(c) and 3(d), where the unmodified and mono-, di-, and trioxidized forms of the peptide with sequence LEQWAEEAVAR appear at the same retention time, indicating that the oxidation of the tryptophan takes place after the separation, i.e. during the ionization.

These artifacts could be easily detected automatically using a simple peak pair algorithm (Table 2) due to the fact that they co-elute with the precursor ion and due to the fixed mass shift. In total, 82 artifacts were automatically detected in this way in the LC/MS data set by using a mass shift tolerance of ±0.7 Da and a time shift tolerance of ±0.11 min. The typical width of a peptide peak was 0.5 min.

The MS/MS spectra of these artifacts are similar to MS/MS spectra of regular peptides and can therefore lead to incorrect sequence assignments when searching sequence collections, if the information that they are adducts is disregarded.

CONCLUSIONS

Visualization of LC/MS data as a signal intensity map is a powerful tool to assess the success of an LC/MS experiment in order to determine the need for modifications to experimental conditions, including the optimization of the chromatographic conditions and the data-dependent acquisition of the fragment mass spectra. It is also amenable to quality control and for easily finding and confirming modified peptides. In addition, borderline identifications from searches of sequence collections can be checked by examining the intensity map to ensure that they are not the result of artifacts or incorrectly assigned monoisotopic masses.

REFERENCES

1. Aebersold R, Mann M. *Nature* 2003; **422**: 198.
2. Johnson RS, Davis MT, Taylor JA, Patterson SD. *Methods* 2005; **35**: 223.
3. Eriksson J, Chait BT, Fenyo D. *Anal. Chem*. 2000; **72**: 999.
4. Eriksson J, Fenyo D. *Proteomics* 2002; **2**: 262.
5. Fenyo D, Beavis RC. *Anal. Chem*. 2003; **75**: 768.
6. Julka S, Regnier F. *J. Proteome Res*. 2004; **3**: 350.
7. Schulz-Knappe P, Zucht HD, Heine G, Jurgens M, Hess R, Schrader M. *Comb. Chem. High Throughput Screen*. 2001; **4**: 207.
8. Wang W, Zhou H, Lin H, Roy S, Shaler TA, Hill LR, Norton S, Kumar P, Anderle M, Becker CH. *Anal. Chem*. 2003; **75**: 4818.
9. Binz PA, Muller M, Walther D, Bienvenut WV, Gras R, Hoogland C, Bouchet G, Gasteiger E, Fabbretti R, Gay S, Palagi P, Wilkins MR, Rouge V, Tonella L, Paesano S, Rossellat G, Karmime A, Bairoch A, Sanchez JC, Appel RD, Hochstrasser DF. *Anal. Chem*. 1999; **71**: 4981.
10. Ogorzalek Loo RR, Cavalcoli JD, VanBogelen RA, Mitchell C, Loo JA, Moldover B, Andrews PC. *Anal. Chem*. 2001; **73**: 4063.
11. Li XJ, Pedrioli PG, Eng J, Martin D, Yi EC, Lee H, Aebersold R. *Anal. Chem*. 2004; **76**: 3856.
12. Palagi PM, Walther D, Quadroni M, Catherinet S, Burgess J, Zimmermann-Ivol CG, Sanchez JC, Binz PA, Hochstrasser DF, Appel RD. *Proteomics* 2005; **5**: 2381.
13. Krishna RG, Wold F. Posttranslational modifications. In *Proteins: Analysis and Design*, vol. 1. Angeletti RH (ed). Academic Press: San Diego, 1998; 121–206.
14. Krokhin OV, Craig R, Spicer V, Ens W, Standing KG, Beavis RC, Wilkins JA. *Mol. Cell. Proteomics* 2004; **3**: 908.

DOI: 10.1002/rcm