

Phosphorylation site localization in peptides by MALDI MS/MS and the Mascot Delta Score

Simone Lemeer · Elena Kunold · Susan Klaeger · Monika Raabe · Mark W. Towers ·
Emmanuelle Claudes · Tabiwang N. Arrey · Kerstin Strupat · Henning Urlaub ·
Bernhard Kuster

Received: 1 July 2011 / Revised: 20 September 2011 / Accepted: 3 October 2011 / Published online: 25 October 2011
© Springer-Verlag 2011

Abstract Owing to its broad biological significance, the large-scale analysis of protein phosphorylation is more and more getting into the focus of proteomic research. Thousands of phosphopeptides can nowadays be identified using state-of-the-art tandem mass spectrometers in conjunction with sequence database searching, but localizing the phosphate group to a particular amino acid in the peptide sequence is often still difficult. Using 180 individually synthesized phosphopeptides with precisely known phosphorylation sites (p-sites), we have assessed the merits of the Mascot Delta Score (MD score) for the assignment of phosphorylation sites from tandem mass spectra (MS/MS) generated on four different matrix-assisted laser desorption ionization (MALDI) mass spectrometers including tandem

time-of-flight (TOF/TOF), quadrupole time-of-flight, and ion trap mass analyzers. The results show that phosphorylation site identification is generally possible with false localization rates of about 10%. However, a comparison to previous work also revealed that phosphorylation site determination by MALDI MS/MS is less accurate than by ESI-MS/MS particularly if several and/or adjacent possible phosphorylation acceptor sites exist in a peptide sequence. We are making the tandem MS spectra and phosphopeptide collection available to the community so that scientists may adapt the MD scores reported here to their analytical environment and so that informatics developers may integrate the MD score into proteomic data analysis pipelines.

Published in the *10th Anniversary Issue*.

Simone Lemeer, Elena Kunold, and Susan Klaeger contributed equally to this work.

Electronic supplementary material The online version of this article (doi:10.1007/s00216-011-5469-2) contains supplementary material, which is available to authorized users.

S. Lemeer · E. Kunold · S. Klaeger · B. Kuster (✉)
Chair of Proteomics and Bioanalytics,
Technische Universität München,
Emil Erlenmeyer Forum 5,
85354 Freising, Germany
e-mail: kuster@wzw.tum.de

M. Raabe · H. Urlaub
Bioanalytical Mass Spectrometry Group,
Max Planck Institute for Biophysical Chemistry,
Am Fassberg 11,
37077 Göttingen, Germany

M. W. Towers · E. Claudes
Waters Corporation,
Atlas Park, Simmons Way,
Manchester M22 5PP, UK

T. N. Arrey · K. Strupat
Thermo Fisher Scientific,
Hanna-Kunath-Strasse 11,
28199 Bremen, Germany

H. Urlaub
Bioanalytics, Department of Clinical Chemistry,
University Medical Center Göttingen,
Robert Koch Strasse 42,
37075 Göttingen, Germany

B. Kuster
Center for Integrated Protein Science Munich (CIPSM),
Munich, Germany

Keywords Mass spectrometry · Phosphorylation · Proteomics · False localization rate

Abbreviations

CID	Collision-induced dissociation
ESI	Electrospray ionization
ETD	Electron transfer dissociation
ETDSA	Electron transfer dissociation with supplemental activation
FLR	False localization rate
HCD	Higher energy collision-induced dissociation
LC-MS/MS	Liquid chromatography–tandem mass spectrometry
MALDI	Matrix-assisted laser desorption ionization
MD score	Mascot Delta score
Mgf	Mascot generic format
MSA	Multistage activation
PKL	Peak list file
PSD	Post-source decay
p-sites	Phosphorylation sites
PTMs	Posttranslational modifications
QTOF	Quadrupole time of flight
TOF/TOF	Tandem time of flight

Introduction

Owing to their widespread occurrence and broad biological significance, posttranslational modifications (PTMs) of proteins are increasingly studied by proteomic methods [1]. The development of efficient biochemical enrichment procedures [2, 3] and advances in nanoscale chromatography and tandem mass spectrometry now allow the analysis of thousands of phosphopeptides from a given biological system [4–7]. Concomitant with the ability to analyze phosphopeptides en masse, our ability to verify phosphorylation site assignments by manual inspection of tandem mass spectra is diminishing rapidly [8]. Therefore, we and others have developed computational approaches that aim to alleviate this problem. Broadly speaking, these methods can be divided into two categories. Well-known representatives of the first category are the Ascore and PTMs score that interpret the tandem mass spectra based on empirical fragmentation rules and score the likelihood that a particular spectrum represents a certain phosphopeptide [5, 9–14]. In the second group, the phosphorylation site localization score uses the results of a sequence database search engine [15–19], and these tools score the likelihood that a particular phosphorylation site isomer of a peptide is best represented by the spectrum that generated the match.

Albeit powerful, all of these methods were trained on data generated by electrospray ionization (ESI) [20]

coupled to several forms of tandem mass spectrometry available on a number of mass analyzer types. Matrix-assisted laser desorption ionization (MALDI) [21] is the second important ionization technique for biomolecules including phosphopeptides [22–24] for which the aforementioned site localization methods are either not applicable or have not been systematically investigated. In contrast to ESI, MALDI predominantly generates singly charged peptides, and therefore, their fragmentation behaviors in the gas phase can be quite distinct. In addition to the typical y- and b-type sequence ions, MALDI tandem mass spectra often contain so-called internal fragment ions that arise from cleavage of two peptide bonds along the peptide backbone [22, 25]. A common feature of ESI and MALDI tandem MS spectra of phosphopeptides is the strong loss of the phosphate or phosphoric acid group in the gas phase, which often renders the tandem MS spectra of phosphopeptides difficult to interpret [26]. This complicates not only peptide identification but also phosphorylation site localization because the detection of just one or few particular fragment ions is often required for unambiguous site assignments. The situation is further complicated by the frequent presence of multiple potential sites of modification in the same peptide.

Based on a set of 180 individually synthesized phosphopeptides with precisely known phosphorylation sites and thousands of tandem mass spectra generated by all common types of ESI MS/MS methods, we have recently described and evaluated the Mascot Delta Score (MD score) for phosphorylation site localization [18, 27, 28]. The MD score calculates the difference between the top 2 Mascot ion scores of alternative phosphorylation sites in the same peptide sequence. We found that the MD score is very discriminating and can robustly distinguish phosphorylation site isomers for all of the different fragmentation types investigated. Using the same set of 180 synthetic peptides, we investigated in this work whether the MD score was also applicable to phosphorylation site determination of MALDI-generated ions and fragmented by low- and high-energy collision-induced dissociation (CID), low-energy resonance CID, and post-source decay (PSD) on TOF/TOF, QTOF, or Orbitrap instruments.

The results show that the MD score is also generally applicable for MALDI tandem MS analysis of phosphopeptides, which makes it a useful tool for proteomic studies employing LC-MALDI MS/MS workflows. The analysis also highlights that significant differences in the fragmentation behavior exist on the investigated MS platforms, which in turn allowed the computation of platform-specific false localization rates (FLR). The practical significance of this work is that our data may provide guidance for scientists with respect to determining FLRs in their own particular phosphoproteomics studies. In addition, we are

making the tandem mass spectra available to the community as database searchable text files so that developers of proteomic software can benchmark the performance of their tools against a standard data set. Finally, the physical set of phosphopeptides used in this study is available so that any laboratory can determine and implement MD score characteristics for their particular analytical setup.

Materials and methods

Phosphopeptide synthesis

The choice and synthesis of the set of 180 phosphopeptides used in the study has been described elsewhere [18]. Briefly, based on a list of naturally occurring phosphopeptides [29], 180 peptides including positional p-site isomers were synthesized individually by solid-phase synthesis at a scale of 2 μ mol on a parallel peptide synthesizer (Intavis, Cologne, Germany) following the standard Fmoc strategy (see Electronic supplementary material (ESM) Table S1 for a complete list of all sequences and phosphorylation sites). Fmoc-protected amino acids were obtained from Intavis. Crude peptides were quality controlled by MALDI-TOF-MS and LC-ESI-MS/MS (see ESM Fig. S3 for ESI-CID tandem mass spectra) and used for subsequent experiments without further purification. The synthesized peptides vary in length between 5 and 28 residues (800–3,300 Da), which is similar to what is typically found in large-scale phosphoproteomics studies [30]. Of the peptides, 33% contain one missed protease cleavage site (5% contain two such sites), which is also similar to other phosphorylation studies using trypsin as the protease. Furthermore, 129 peptides are phosphorylated at Ser or Thr residues, 48 on Tyr, and 3 represent mixed pSer/pThr/pTyr peptides. One hundred sixty-four peptides are singly and 16 are doubly phosphorylated, and about 50% of all peptides represent positional isomers. The higher incidence of pY-containing peptides in our set compared with that typically found in large-scale studies was driven by the need to investigate a sufficient number of these peptides in order to arrive at general conclusions. Multiply phosphorylated peptides are underrepresented in our study (~10% here vs. ~20% in other studies) [30]. Therefore, MD score thresholds should be carefully assessed in these cases.

Sample preparation for mass spectrometry

Phosphopeptides were dissolved in 0.1% trifluoroacetic acid (TFA) before use. For phosphopeptide sample preparation, microscale reversed-phase columns were prepared by applying 40 μ g OligoR3 material (Applied Biosystems, Darmstadt, Germany) on top of a micro ZipTip (Merck

Millipore, Darmstadt, Germany). Columns were equilibrated by 40 μ l 1% TFA and 1 pmol of each synthetic peptide applied. Columns were washed with 40 μ l 1% TFA and peptides were eluted directly onto MALDI targets using 4 μ l of the matrix solution (2 mg/ml α -cyano-4-hydroxycinnamic acid in 50% acetonitrile and 0.5% TFA).

Generation of tandem mass spectra

All 180 phosphopeptides were analyzed on all four different MALDI tandem mass spectrometers: (a) an ultrafleXtreme TOF/TOF (Bruker Daltonics, Bremen, Germany); (b) a 4800 MALDI-TOF/TOF Analyzer (Applied Biosystems); (c) a MALDI Synapt G2 QTOF (Waters, Manchester, UK); and (d) a LTQ Orbitrap XL (Thermo Fisher Scientific, Bremen GmbH, Bremen, Germany). Peptides were selected for tandem MS and tandem mass spectra were accumulated until a stable signal-to-noise level was obtained (either by accumulating spectra or, in case of the LTQ Orbitrap XL, by using the automatic gain control function). Measurements on the ultrafleXtreme TOF/TOF were performed at the TU München Laboratory and employed a form of PSD [31] without the use of a collision gas. The data for the 4800 TOF/TOF instrument were generated at the MPI for Biophysical Chemistry in Göttingen and employed high-energy CID with nitrogen as the collision gas. Experiments on the Synapt G2 QTOF instrument were performed at the Waters site in Manchester and utilized argon for low-energy CID. Lastly, LTQ Orbitrap XL measurements were performed at the Thermo Scientific site in Bremen and employed helium for resonance low-energy CID in conjunction with multistage activation which co-fragments the precursor-98/-196 ion (loss of one or two phosphoric acid moieties).

Data processing and database searching

UltrafleXtreme TOF/TOF tandem MS spectra were smoothed and baseline-subtracted and peaks with signal-to-noise ratios of <3 were removed. The spectra were then converted into mascot generic format files (mgf) using FlexAnalysis (v. 3.3). 4800 TOF/TOF spectra were converted into mgf files using the 4000 Series Explorer (v 3.5.3) using the same criteria. LTQ Orbitrap XL spectra were converted into mgf files using Mascot Distiller (v.2.3). Synapt G2 QTOF spectra were deisotoped, base line-subtracted, and converted into peak list files (pkl) using MassLynx (4.1). Precursor masses were recalibrated on the known masses of the analyzed peptides. All subsequent data analysis was centrally performed in the TU München Laboratory. All mgf and pkl files were searched against the human International Protein Index database (v. 3.68) combined with a decoy version thereof [32] using the

Mascot server (version 2.3). This database contains a total of 174,122 protein sequences (50% forward, 50% reverse). Global search parameters were as follows: carbamidomethyl cysteine as fixed modification; oxidized methionine and phosphorylation of serine, threonine, and tyrosine as variable modifications. Tryptophane oxidation (to kynurenin) was considered for tryptophane-containing peptides. Trypsin was specified as the proteolytic enzyme, and up to two missed cleavages were allowed. The mass tolerance for precursor/fragment ions was set in an instrument-specific manner: ultrafleXtreme TOF/TOF, 100 ppm/0.6 Da; 4800 TOF/TOF, 100 ppm/0.6 Da; Synapt G2 QTOF, 100 ppm/0.6 Da; LTQ Orbitrap XL, 5 ppm/0.6 Da. In addition, the Synapt G2 QTOF data were also searched with 50 ppm/0.1 Da.

Phosphorylation site localization

The MD score was computed from the Mascot search result files by determining the difference between the best and second (third, etc.) best Mascot ion scores for the correct and alternative phosphorylation site localizations on an otherwise identical peptide sequence. FLR calculation for all scores was performed by dividing the number of incorrect site assignments by the total number of assignments.

Data and reagent availability

All Mascot searchable files are available as part of the supplement. All synthetic peptides used in this study are available from Intavis AG (Cologne, Germany; <http://www.intavis.com>). Unprocessed mass spectrometric data are available from the authors upon request.

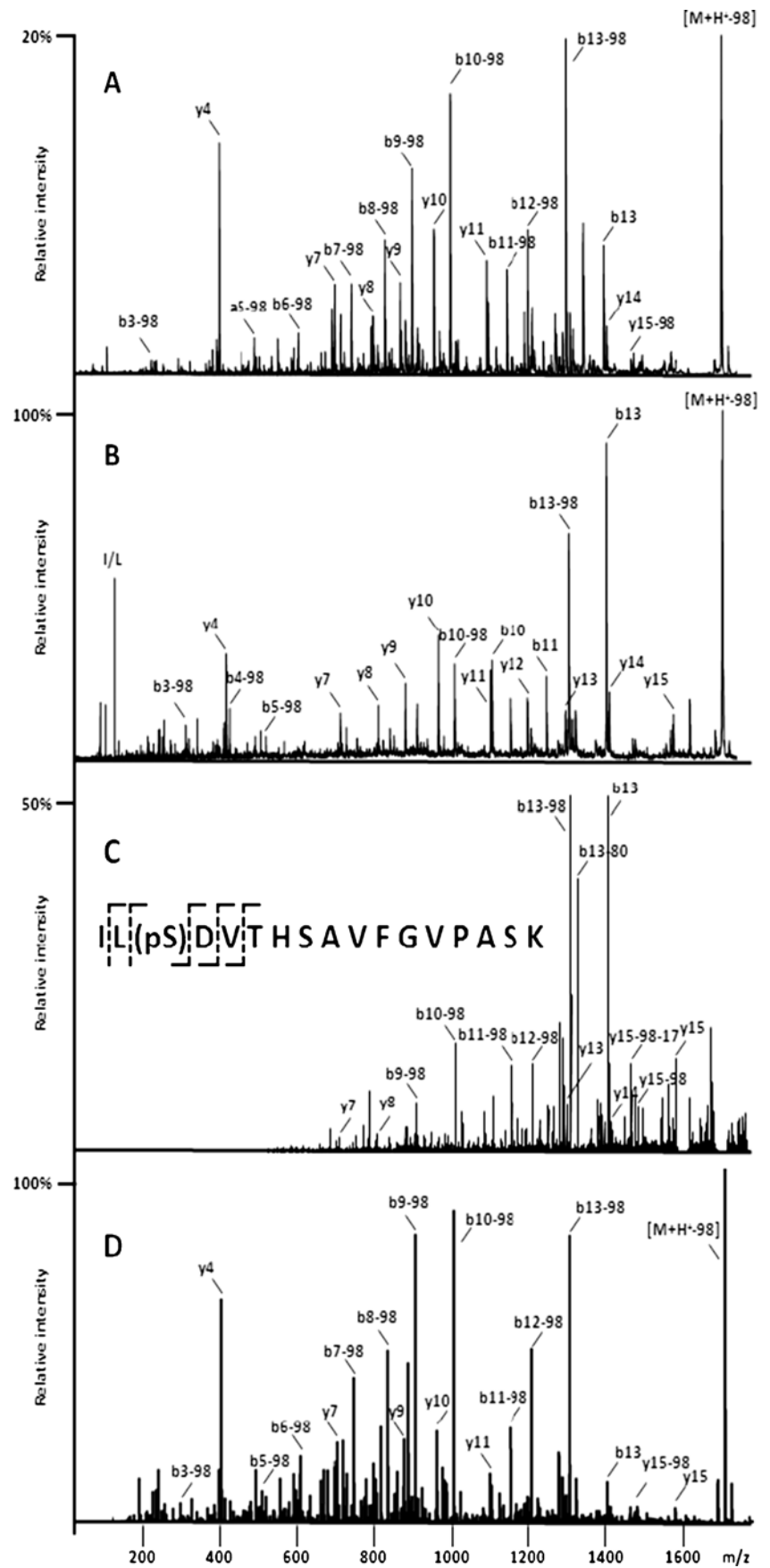
Results and discussion

MALDI tandem mass spectra of phosphopeptides

In this study, we used four distinct MALDI MS/MS platforms for the analysis of phosphopeptides with particular emphasis on phosphorylation site determination using the Mascot delta score. The types of fragmentation methods offered by these four platforms are quite distinct. The Bruker ultrafleXtreme TOF/TOF instrument operates in the so-called LIFT mode which essentially collects fragment ions that are generated by spontaneous post-source decay from high kinetic energy precursor ions [31]. The Applied Biosystems 4800 TOF/TOF uses high-energy CID (2 kV) in a nitrogen-filled collision cell [33]. The Waters Synapt G2 QTOF [34] operates with argon in a classical low-energy collision cell (30–200 V), and the LTQ Orbitrap XL

utilizes low-energy resonance CID in a helium-filled linear ion trap [35]. While both TOF/TOF instruments employ timed ion selection for precursor isolation (leading to a quite broad precursor ion selection of 0.1–1% of the precursor m/z), the latter two instruments are capable of selecting precursor ions with higher resolution. The three time-of-flight instruments can record fragment ions across the entire m/z range, while only fragment ions with $m/z > 28\%$ of the precursor ion m/z value can be stably trapped in the linear ion trap of the LTQ Orbitrap XL instrument (the so-called low mass cutoff). As an example for the overall appearance and information content of the above fragmentation types, Fig. 1 (and ESM Fig. S1) shows tandem mass spectra of the synthetic phosphopeptide IL(pS)DVTHSAVFGVPASK. All four spectra show a strong neutral loss of phosphoric acid from the precursor or fragment ions (except for the LTQ Orbitrap XL spectrum in which the loss of 98 Da is co-activated with the precursor in order to increase the fragment ion yield of other ions). All spectra also show sufficient numbers of b and y ions from which the peptide can be unambiguously identified by database searching and the phosphorylation site localized to the serine in position 3. The extent to which sequence informative fragment ions are present in the spectra is quite different between the platforms. Of the 16 possible y and b ions of this peptide (not considering whether or not these are neutral losses of 98), 13 y ions (10 b ions) are detected on the ultrafleXtreme TOF/TOF (leading to a Mascot ion score of 60), 12 y ions (12 b ions) on the 4800 TOF/TOF (Mascot ion score of 73), 11 y ions (13 b-ions) on the Synapt G2 QTOF (Mascot ion score of 69), and 7 y ions (7 b ions) on the LTQ Orbitrap XL (Mascot ion score 45). For the latter instrument, the detection of fewer sequence ions is explained by the fact that four of all possible b and y ions are outside the stable detectable fragment ion mass range (here $< 485 m/z$). This is generally not an issue for simple peptide identification, but can be a disadvantage for phosphorylation site determination, as discussed below. Not all sequence ions are useful for phosphorylation site localization. In this particular example, the ions y12 through to y-16 or the b3 to b5 ions (or the respective loss of -98 or the respective a ions) would each localize the phosphorylation site to the serine residue in position 3. Seven such supporting ions are detected with varying intensities in the spectrum of the ultrafleXtreme TOF/TOF, eight in the spectrum of the 4800 TOF/TOF, six in the spectrum of the Synapt G2 QTOF, and three in the linear ion trap spectrum of the LTQ Orbitrap XL. All platforms thus unambiguously identified the phosphorylation site in this example. However, the different numbers of detected p-site-relevant sequence ions directly translate into the confidence with which the site can be localized. As a result, the highest p-site localization confidence was obtained by the 4800 TOF/TOF (MD score of 62), followed by the

Fig. 1 MALDI tandem mass spectra of the phosphorylated peptide IL(pS)DVTHSAVFGV-PASK recorded on four different mass spectrometers. Bruker ultrafleXtreme TOF/TOF (A), Applied Biosystems 4800 TOF/TOF (B), Thermo LTQ Orbitrap XL (C), and Waters Synapt G2 QTOF (D). The main sequence ions as well as phosphorylation site-localizing ions are marked



ultrafleXtreme TOF/TOF (MD score of 41), the Synapt G2 QTOF (MD score of 21), and the LTQ Orbitrap XL (MD score of 10, see also below).

A rather striking difference between the platforms concerns the presence (and intensity) of internal fragment ions arising from the cleavage of two peptide bonds (ESM Fig. S2). While many such ions are detected particularly in the lower half of the ultrafleXtreme TOF/TOF ($n=29$) and Synapt G2 QTOF ($n=22$) spectra, their occurrence in the high-energy CID spectrum of the 4800 TOF/TOF instrument is much lower ($n=11$), and these ions are virtually absent in the linear ion trap spectrum of the LTQ Orbitrap XL ($n=3$). Because the presence of these ions cannot easily be correlated with the energy regime or timescale of the MS/MS experiment, we can currently not offer any plausible explanation for this observation. Internal ions can be useful for peptide identification or for phosphorylation site determination in case they cover the amino acid sequence in which the modification occurs. In this particular example, this is not the case because the modified amino acid is located close to the peptide N-terminus for which no internal ion is detected. However, for peptides containing Pro residues, internal fragment ions may often help in localizing a phosphorylation site. Mass accuracy is another parameter that influences peptide identification and modification site localization [18, 36]. Across all analyzed peptides, the LTQ Orbitrap XL instrument obtained the best precursor ion mass measurements (~ 1 ppm) followed by the 4800 TOF/TOF and the ultrafleXtreme TOF/TOF (~ 40 ppm). Given that the Synapt G2 QTOF data were recalibrated on the known peptide masses, these data cannot be directly compared with the other platforms, but the initial precursor mass accuracy was generally better than 40 ppm. The observed errors in fragment ion mass measurements were generally between 50 and 100 ppm for the Synapt G2 QTOF, 100 and 200 ppm for the LTQ Orbitrap XL, and between 300 and 700 ppm for the two TOF/TOF instruments. In the interest of comparability of the phosphorylation site localization data (see section below), we searched all tandem mass spectra produced on the TOF instruments with 100 ppm on the precursor and 0.6 Da on the fragment ions and all spectra produced on the LTQ Orbitrap XL instrument with 5 ppm on the precursor and 0.6 Da on the fragment ions. In addition, the Synapt G2 QTOF spectra were also searched with 50-ppm precursor and 0.1-Da fragment ion tolerance.

Phosphorylation site determination using the Mascot Delta Score

The peptide identification scores of Mascot or other search engines are not in themselves necessarily a good or sufficient indicator for the correct localization of a

phosphorylation site within a peptide sequence (Fig. 2A). However, we and others have shown for a number of ESI-MS/MS platforms that the Mascot Delta Score, which simply reflects the difference of Mascot ion scores between the highest and the second highest ion scores for candidate phosphorylation sites on an identical peptide sequence in a database search, is a powerful criterion for site localization [18]. In this study, we have extended the evaluation of the MD score to four MALDI MS/MS platforms using the same set of 180 individually synthesized and characterized (but naturally occurring) phosphopeptides [18, 29]. These 180 phosphopeptides range in mass from 800 to 3,300 Da and include 129 pS/pT, 48 pY, and 3 mixed pS/pT/pY peptides. One hundred sixty-four peptides are singly and 16 are doubly phosphorylated, and about 50% of all peptides represent positional isomers. We measured all the peptides on all of four MALDI MS/MS platforms, searched the generated tandem mass spectra against a human protein sequence database using Mascot to identify the (known) peptide sequence, computed the MD score from the Mascot result files for the alternative phosphorylation sites, and derived the false localization rates for each platform. These data are summarized in Table 1 (see ESM Tables S1–S5 for MD scores of all peptides on all platforms). Interestingly, the three TOF platforms gave very similar results. For $\sim 80\%$ of the identified peptides, the phosphorylation site was also correctly assigned. For about 10% of the peptides, the site assignment was incorrect; for the remaining 10%, Mascot failed to assign the phosphorylation site to any amino acid (MD score of 0). The false localization rates of phosphopeptide assignment for the non-trivial cases (i.e., peptides containing more than one possible phosphorylation site) were 10% (ultrafleXtreme TOF/TOF), 11% (4800 TOF/TOF), and 12% (Synapt G2 QTOF), respectively. The data also confirm earlier results that localizing p-sites is more difficult if two possible acceptor sites are directly adjacent compared with cases in which the two acceptor sites are further apart [18]. This is because fewer sequence ions are available for correct site assignment for adjacent sites compared with cases in which the sites are further apart. Although this does not drastically change the observed FLRs, the majority of cases (16 out of 18 for ultrafleXtreme, 6 out of 11 for 4800 TOF/TOF, and 8 out of 9 for Synapt G2 QTOF) in which Mascot failed to assign any phosphorylation site (MD score of 0) occur in peptides with adjacent acceptor sites.

In contrast, the data obtained from the ion trap part of the LTQ Orbitrap XL (operated in multistage activation mode in which precursors are fragmented in the ion trap by low-energy resonance CID) show a lower peptide identification and site localization rate compared with the above. From the 142 analyzed phosphopeptides, 93 were identified and the correct p-site was assigned in about 50% of the cases.

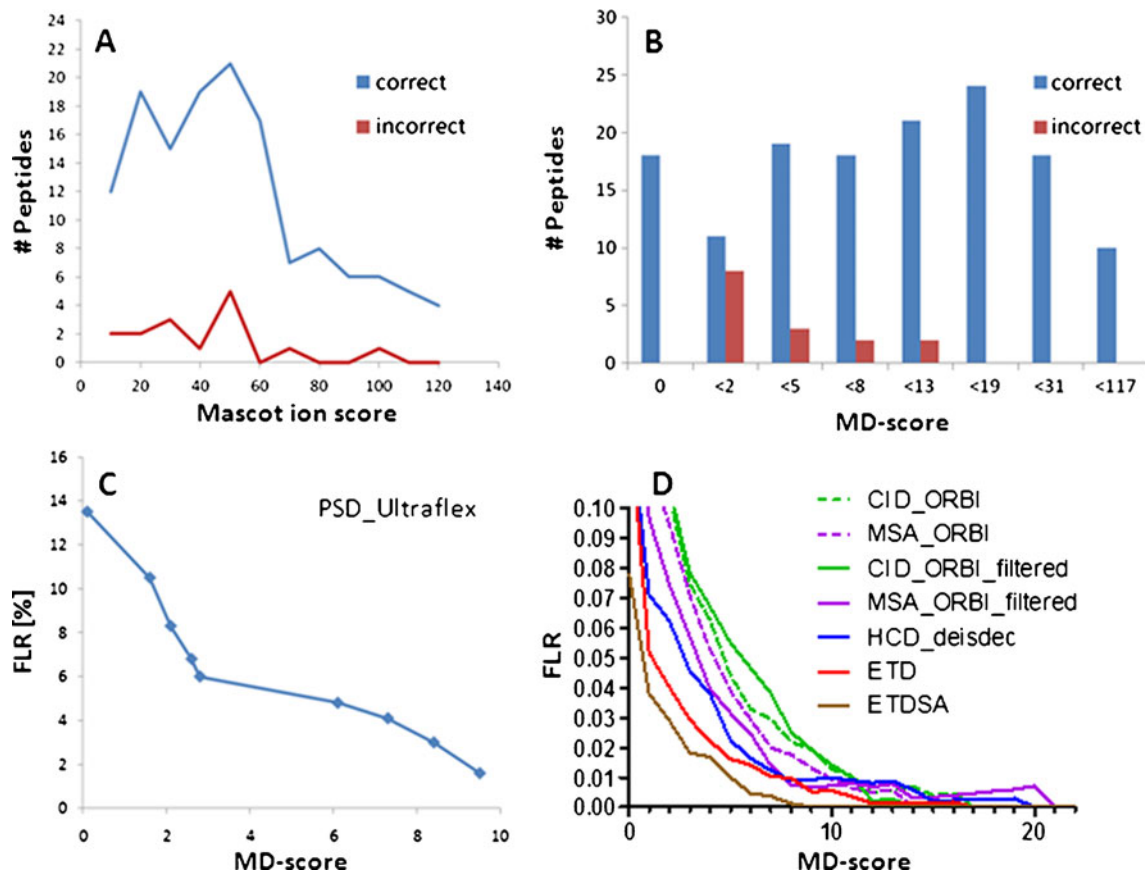


Fig. 2 The Mascot Delta Score can identify sites of phosphorylation in MALDI tandem mass spectra. **A** Mascot ion score distribution for correct and incorrect phosphorylation site assignments derived from MALDI MS/MS data collected by post-source decay on an Ultraflex instrument. **B** Mascot Delta score distribution for correct and incorrect phosphorylation site assignments of the same data. **C** False localization rate as a function of the MD score of the same data. **D** For comparison: false localization rate as a function of the MD score for ESI MS/MS data collected for different fragmentation techniques collected on an LTQ Orbitrap XL ETD (reproduced from

[18]. *CID_Orbi* fragmentation by collision-induced dissociation, fragment ion detection in the ion trap, and precursor ion detection in the orbitrap; *MSA_Orbi* fragmentation using multistage activation; *CID_Orbi_filtered* fragmentation by filtering tandem mass spectra according to [41]; *MSA_Orbi_filtered* fragmentation using multistage activation by filtering tandem mass spectra; *HCD_deisdec* fragmentation by higher energy CID, fragment, and precursor ion detection in the orbitrap and deisotoping/charge deconvolution of tandem mass spectra according to [42]

The phosphorylation site assignment was incorrect for 33% of the non-trivial cases and no p-site could be assigned for 16% of the peptides. The reason for both observations (lower identification rate and higher p-site FLR) is the aforementioned low mass cutoff of ion trap mass spectrometers owing to the physical inability of these devices to trap fragment ions across the entire m/z range. For the singly charged precursor ions generated by MALDI, the low mass cutoff of the instrument ($\sim 28\%$ of the precursor m/z) leads to the progressive loss of many valuable sequence ions as the peptides get larger (i.e., higher m/z). For example, for a singly charged phosphopeptide of m/z 2,000, no ions below ~ 600 Da will be detected which corresponds to the lack of sequence information for the five to six most C- and/or N-terminal amino acids. Should the p-site be localized in this region, the tandem mass spectrum

cannot reveal it. Our data also suggest that the above loss of sequence information is not generally compensated for by the detection of complementary y and b ion series representing complementary parts of the peptide sequence. Further evidence for our interpretation of the lower performance of the MALDI LTQ Orbitrap XL comes from previously published data on the same set of phosphopeptides using resonance CID on the same mass spectrometer (LTQ Orbitrap XL) fitted with an ESI source [18]. In ESI, most tryptic peptides and phosphopeptides are doubly or triply charged. This greatly facilitates the detection of lower m/z fragment ions in the ion trap. Using the above example of a phosphopeptide of 2,000 Da, the detectable fragment ion range is extended down to ~ 300 Da (doubly charged precursor) and ~ 200 Da, respectively (triply charged precursor). Hence, many more sequence-related fragment

Table 1 Summary of phosphorylation site determination data for all employed MALDI mass spectrometers

Instrument	No. of MS/MS spectra	No. of identified peptides	No. of correct p-site assignments	No. of false p-site assignments	No. of unassigned p-sites	Global FLR (%)
Bruker UltrafleXtreme TOF/TOF: search parameters 100 ppm/0.6 Da						
Total	172	154	121	15	18	10
Non-trivial	162	146	113	15	18	10
Adjacent p-sites	78	66	43	7	16	11
Isomers	82	77	54	9	14	12
Applied Biosystems 4800 TOF/TOF: search parameters 100 ppm/0.6 Da						
Total	169	151	124	16	11	11
Non-trivial	158	141	114	16	11	11
Adjacent p-sites	76	67	51	10	6	15
Isomers	79	73	54	12	7	16
Thermo LTQ Orbitrap: search parameters 5 ppm/0.6 Da						
Total	142	93	54	29	10	31
Non-trivial	134	89	50	29	10	33
Adjacent p-sites	61	42	22	12	8	29
Isomers	82	51	25	20	6	39
Waters Synapt G2 QTOF: search parameters 100 ppm/0.6 Da						
Total	176	168	140	19	9	11
Non-trivial	165	158	130	19	9	12
Adjacent p-sites	80	77	63	6	8	8
Isomers	81	80	64	11	5	14
Waters Synapt G2 QTOF: search parameters 50 ppm/0.1 Da						
Total	176	165	136	18	11	11
Non-trivial	165	156	127	18	11	12
Adjacent p-sites	80	76	60	6	10	13
Isomers	81	79	63	11	5	14

Non-trivial indicates that there are at least two amino acid in the peptide sequence to which a single phosphate group may be localized. *Adjacent p-sites* refers to two (or more) possible directly adjacent amino acids on which the phosphorylation may occur. The term *isomers* refers to synthetic phosphopeptides with identical amino acid sequences but different positions where the phosphate group is actually localized

ions become detectable, which facilitates p-site localization. As a result, the ESI MS/MS data show an FLR of ~10% MS/MS [18], which is significantly better than that observed for the MALDI MS/MS data in this study (see also Fig. 2D). Several ways can be envisaged by which this situation may be improved. First, the LTQ Orbitrap XL allows the detection of fragment ions in the Orbitrap detector rather than the ion trap detector. Second, the fragments can be generated in a multipole collision cell (called HCD for the higher energy CID) [37]. Both these measures would overcome the low mass cutoff issue, but may also lead to lower sensitivity. A third possibility would be to employ the recently published iHCD method [38] in which the precursor ions are isolated in the ion trap, then fragmented in the ion injection path of the instrument with subsequent readout of the fragment ions in the ion trap. This alternative is particularly attractive as it does not appear to lose sensitivity, but this scan type has not yet been

implemented on the commercial MALDI LTQ Orbitrap XL system used in this study.

The information in Table 1 reports global FLR values for the four investigated platforms irrespective of the absolute values of the MD scores that underlie the FLR calculation. We have recently shown that it is possible to compute and predict FLRs for each MD score value provided that a statistically sufficient (i.e., high) number of tandem mass spectra are available. Owing to its discontinuous nature, comparatively fewer tandem mass spectra are typically generated on MALDI MS/MS instruments. Still, as shown in Fig. 2 (and ESM Fig. S2), it is possible to identify an MD score range that indicates a rough FLR estimate for individual peptides. For the ultrafleXtreme TOF/TOF instrument, spectra with MD scores larger than 13 do not contain any false assignments (52 peptides). The MD score range of 8–13 contains 21 correct and 2 incorrect assignment (FLR = 9%), MD scores between 5 and

8 indicate an FLR of 10% (18 correct, 2 incorrect), MD scores of 2–5 correspond to an FLR of 14% (19 correct, 3 incorrect), and for MD scores lower than 2 (but not 0) the FLR rises to close to 50% (11 correct, 8 incorrect). Cumulatively, MD scores in our data set of >8 indicate an FLR of 3% (73 correct, 2 incorrect assignments) and MD scores of >5 correspond to an FLR of 4%. The corresponding data for the other platforms are shown in ESM Fig. S2. Briefly, the MD score characteristic for the 4800 TOF/TOF is broadly similar to that of the ultrafleXtreme TOF/TOF, and an MD score of >10 is required to reach 3% global FLR. Owing to the low number of identified and correctly localized peptides, we were unable to define a meaningful MD score threshold to reach a certain FLR for the LTQ Orbitrap XL data. However, the data suggest that an MD score of at least 20 is required to reach an FLR of <10%. Although the Synapt G2 QTOF instrument was very successful in the identification of phosphopeptides and correct site localization (highest absolute number of IDs and correct localizations both at search tolerances of 0.6 and 0.1 Da; Table 1), there are a couple of outliers that are incorrectly localized at high MD scores. This leads to a requirement of MD scores of >20 to reach a global FLR of 3%. When disregarding those outliers, the MD score threshold approximates that obtained for the two TOF/TOF platforms.

When comparing the MD score thresholds obtained by MALDI MS/MS in this study (Fig. 2 and ESM Fig. S2) to those generated for a number of ESI MS/MS techniques from our previous work [18], we find that significantly higher MD scores are usually required for MALDI MS/MS data to reach a particular false localization rate (say 3% or 5%; Fig. 2D). For the TOF/TOF platforms, this likely is the result of the higher noise level in tandem mass spectra that arises from the co-fragmentation of molecular species in the wide precursor ion selection window (see also ESM Fig. S3 for the corresponding ESI MS/MS spectra for comparing spectral quality). A further possible influence may arise from the fact that MALDI-generated ions have (at least initially) more internal energy than those generated by ESI, which may lead to lower intensities of site-localizing fragment ions. Another, but possibly rather minor, contributing factor may be the much higher number of MS/MS attempts per peptide in LC-ESI MS/MS, which statistically increases the likelihood of generating an informative spectrum. We can also not entirely exclude the possibility that gas phase rearrangements of the phosphate group may lead to incorrect p-site assignments [39], but recent reports suggest that this is not of much practical consequence [28, 40].

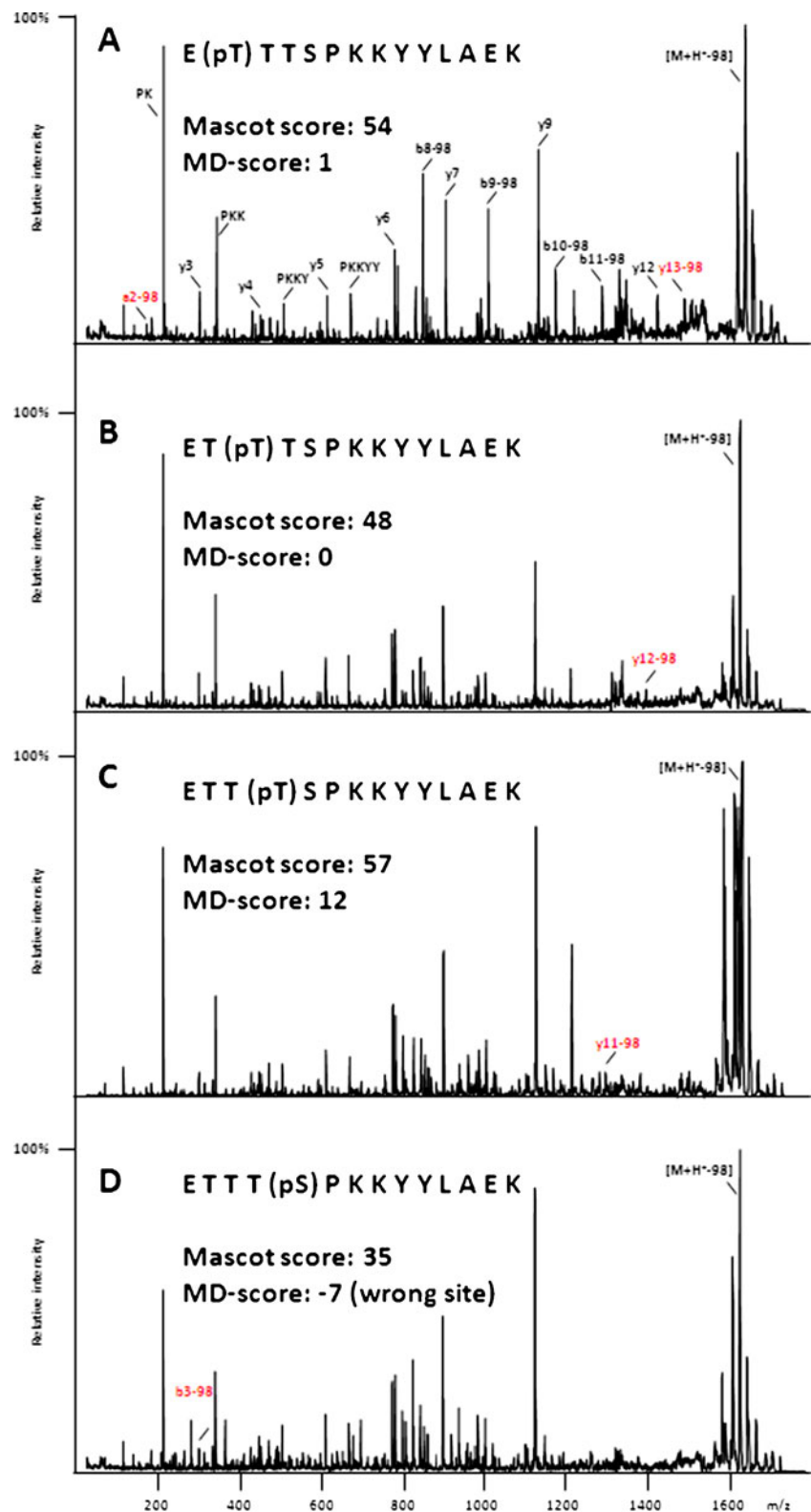
In any case, phosphorylation site localization by MALDI MS/MS is clearly possible albeit more difficult than for ESI MS/MS. For large-scale phosphorylation studies, global MALDI FLR values in the 3–10% range may be appropri-

ate to describe the quality of such data. We do, however, advise readers to use the more conservative FLR values (i.e., binned by MD score range; ESM Fig. S2) for studies in which individual phosphorylation site determinations have to be performed. Our set of phosphopeptides also contains 16 doubly phosphorylated peptides. This number is clearly too low to derive any general conclusions, but for completeness, we note here that for the ultrafleXtreme TOF/TOF data, nine out of ten detected doubly phosphorylated peptides were also correctly assigned by the MD score (5/6 for the 4800 TOF/TOF, 2/3 for the LTQ Orbitrap XL, and 9/12 for the Synapt G2 QTOF).

Assigning phosphorylation sites in positional isomers

Protein phosphorylation is primarily found on serine, threonine, and tyrosine residues of which the former two amino acids are very abundant. Hence, many peptides contain more than one possible phosphorylation acceptor site, and in fact, alternative phosphorylation may frequently occur *in vivo*. It is therefore important to investigate to what extent positional isomers can be distinguished by tandem mass spectrometry. About 50% of our synthesized peptides represent positional isomers, and Fig. 3 shows the tandem mass spectra (ultrafleXtreme TOF/TOF) of the four example peptides E(pT)TTSPKKYYLAEK, ET(pT)TSPKKYYLAEK, ETT(pT)SPKKYYLAEK, and ETTT(pS)PKKYYLAEK. The b and y ions that would be required for unambiguous site localization for these peptides are also shown. Although all spectra exhibit a good signal-to-noise ratio and robustly identify the underlying peptide sequence, most of the strong fragment ions do not contribute to site localization. In fact, the site-determining ions are all of low signal intensity; therefore, the success of phosphorylation site localization is not very good. The peptide in which T2 is phosphorylated (Fig. 3A) has a MD score of 1 because only two weak ions (a2-98 and y13-98) are available for site localization. The spectrum of the isomer with T3 phosphorylation (Fig. 3B) has an MD score of 0 as no fragment ion is detectable above the noise that would allow unambiguous site localization. An MD score of 12 is obtained for the peptide with T4 phosphorylation (Fig. 3C), but site assignment is primarily based on the detection of the weak y11-98 ion. Finally, the MD score of 7 for the S5 phosphorylated peptide (Fig. 3D) indeed suggests phosphorylation at a different site probably because the b4 or y10 ions required for localization of S5 phosphorylation cannot be detected in the spectrum. Instead, there is a putative b3-98 ion that would suggest phosphorylation at T3. In this somewhat extreme example (six possible acceptor sites, four of which are directly adjacent), none of the four positional isomers could be assigned with good confidence. The same peptide

Fig. 3 Example of largely failed phosphorylation site determination in positional isomers with multiple adjacent acceptor sites. MALDI tandem mass spectra of the peptides: E (pT)TTSPKKYYLAEK (A), ET (pT)TSPKKYYLAEK (B), ETT (pT)SPKKYYLAEK (C), and ETTT(pS)PKKYYLAEK (D) were recorded on the Bruker ultrafleXtreme TOF/TOF instrument. The majority of all fragment ions is identical between the spectra. Therefore, these are only annotated in (A). Further site-localizing fragment ions are marked in *red*



isomers were analyzed by ESI MS/MS in our earlier study [18], and only electron transfer dissociation (ETD) was able to assign all isomers correctly. For many of the positional isomers in our peptide set, site localization is much more

straightforward, and the global FLR values are not radically different compared with the complete data (Table 1). Still, of the 82 positional isomers detected in the ultrafleXtreme TOF/TOF data, only 30 have MD scores of >8 (indicating

an FLR of <10%). Thirty-seven of 79 isomeric peptides detected by the 4800 TOF/TOF fulfill the 10% FLR criterion, and 36 of 81 reach that level of confidence in the Synapt G2 QTOF data set. In summary, the data on positional isomers indicate that caution should be exercised when interpreting MD scores as a means of phosphorylation site localization for peptides with many possible acceptor sites.

Concluding remarks

In this study, we have evaluated the ability of the MD score to identify phosphorylation sites in MALDI tandem MS spectra collected for 180 peptides with precisely known phosphorylation sites on four different instrument platforms. We find that the MD score genuinely reflects the information contained in the tandem mass spectra and that global false localization rates of <10% are attainable for current TOF/TOF and QTOF platforms. However, the performance of MALDI tandem MS for phosphorylation site localization is significantly lower both in quantity and quality than that of ESI tandem MS techniques, which indicates that critical caution is in order when interpreting MALDI tandem MS data for the purpose of phosphorylation site localization. Still, given that LC-MALDI workflows are frequently used in proteomics and that Mascot is one of the most frequently used database search engines, we feel that this work is of practical utility. We are making the mass spectrometry data available to the community so that other scientists or developers of informatics tools may be able to perform similar types of analysis as we did and adapt the reported scores to their analytical environment.

Acknowledgments The authors wish to thank Andrea Hubauer for expert technical assistance.

References

- Mallick P, Kuster B (2010) Proteomics: a pragmatic perspective. *Nat Biotechnol* 28(7):695–709. doi:10.1038/nbt.1658
- Pinkse MW, Uitto PM, Hilhorst MJ, Ooms B, Heck AJ (2004) Selective isolation at the femtomole level of phosphopeptides from proteolytic digests using 2D-NanoLC-ESI-MS/MS and titanium oxide precolumns. *Anal Chem* 76(14):3935–3943. doi:10.1021/ac0498617
- Thingholm TE, Jensen ON, Robinson PJ, Larsen MR (2008) SIMAC (sequential elution from IMAC), a phosphoproteomics strategy for the rapid separation of monophosphorylated from multiply phosphorylated peptides. *Mol Cell Proteomics* 7(4):661–671. doi:10.1074/mcp.M700362-MCP200
- Beausoleil SA, Jedrychowski M, Schwartz D, Elias JE, Villen J, Li J, Cohn MA, Cantley LC, Gygi SP (2004) Large-scale characterization of HeLa cell nuclear phosphoproteins. *Proc Natl Acad Sci U S A* 101(33):12130–12135. doi:10.1073/pnas.0404720101
- Olsen JV, Blagoev B, Gnäd F, Macek B, Kumar C, Mortensen P, Mann M (2006) Global, in vivo, and site-specific phosphorylation dynamics in signaling networks. *Cell* 127(3):635–648. doi:10.1016/j.cell.2006.09.026
- Swaney DL, Wenger CD, Thomson JA, Coon JJ (2009) Human embryonic stem cell phosphoproteome revealed by electron transfer dissociation tandem mass spectrometry. *Proc Natl Acad Sci U S A* 106(4):995–1000. doi:10.1073/pnas.0811964106
- Zhang Y, Wolf-Yadlin A, Ross PL, Pappin DJ, Rush J, Lauffenburger DA, White FM (2005) Time-resolved mass spectrometry of tyrosine phosphorylation sites in the epidermal growth factor receptor signaling network reveals dynamic modules. *Mol Cell Proteomics* 4(9):1240–1250. doi:10.1074/mcp.M500089-MCP200
- Nichols AM, White FM (2009) Manual validation of peptide sequence and sites of tyrosine phosphorylation from MS/MS spectra. *Methods Mol Biol* 492:143–160. doi:10.1007/978-1-59745-493-3_8
- Bailey CM, Sweet SM, Cunningham DL, Zeller M, Heath JK, Cooper HJ (2009) SLoMo: automated site localization of modifications from ETD/ECD mass spectra. *J Proteome Res* 8(4):1965–1971. doi:10.1021/pr800917p
- Beausoleil SA, Villen J, Gerber SA, Rush J, Gygi SP (2006) A probability-based approach for high-throughput protein phosphorylation analysis and site localization. *Nat Biotechnol* 24(10):1285–1292. doi:10.1038/nbt1240
- Payne SH, Yau M, Smolka MB, Tanner S, Zhou H, Bafna V (2008) Phosphorylation-specific MS/MS scoring for rapid and accurate phosphoproteome analysis. *J Proteome Res* 7(8):3373–3381. doi:10.1021/pr800129m
- Ruttenberg BE, Pisitkun T, Knepper MA, Hoffert JD (2008) PhosphoScore: an open-source phosphorylation site assignment tool for MSn data. *J Proteome Res* 7(7):3054–3059. doi:10.1021/pr800169k
- Schlosser A, Vanselow JT, Kramer A (2007) Comprehensive phosphorylation site analysis of individual phosphoproteins applying scoring schemes for MS/MS data. *Anal Chem* 79(19):7439–7449. doi:10.1021/ac0707784
- Wan Y, Cripps D, Thomas S, Campbell P, Ambulos N, Chen T, Yang A (2008) PhosphoScan: a probability-based method for phosphorylation site prediction using MS2/MS3 pair information. *J Proteome Res* 7(7):2803–2811. doi:10.1021/pr700773p
- Baker PR, Trinidad JC, Chalkley RJ (2011) Modification site localization scoring integrated into a search engine. *Mol Cell Proteomics* 10:M111.008078. doi:10.1074/mcp.M111.008078
- Lu B, Ruse C, Xu T, Park SK, Yates J 3rd (2007) Automatic validation of phosphopeptide identifications from tandem mass spectra. *Anal Chem* 79(4):1301–1310. doi:10.1021/ac061334v
- Perkins DN, Pappin DJ, Creasy DM, Cottrell JS (1999) Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* 20(18):3551–3567. doi:10.1002/(SICI)1522-2683(19991201)20:18<3551
- Savitski MM, Lemeer S, Boesche M, Lang M, Mathieson T, Bantscheff M, Kuster B (2011) Confident phosphorylation site localization using the Mascot Delta Score. *Mol Cell Proteomics* 10(2):M110.003830. doi:10.1074/mcp.M110.003830
- Yates JR 3rd, Eng JK, McCormack AL, Schieltz D (1995) Method to correlate tandem mass spectra of modified peptides to amino acid sequences in the protein database. *Anal Chem* 67(8):1426–1436
- Fenn JB, Mann M, Meng CK, Wong SF, Whitehouse CM (1989) Electrospray ionization for mass spectrometry of large biomolecules. *Science* 246(4926):64–71
- Karas M, Hillenkamp F (1988) Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. *Anal Chem* 60(20):2299–2301

22. Bennett KL, Stensballe A, Podtelejnikov AV, Moniatte M, Jensen ON (2002) Phosphopeptide detection and sequencing by matrix-assisted laser desorption/ionization quadrupole time-of-flight tandem mass spectrometry. *J Mass Spectrom* 37(2):179–190. doi:10.1002/jms.271
23. Condina MR, Gustafsson JO, Klingler-Hoffmann M, Bagley CJ, McColl SR, Hoffmann P (2010) EZYprep LC-coupled MALDI-TOF/TOF MS: an improved matrix spray application for phosphopeptide characterisation. *Proteomics* 10(13):2516–2530. doi:10.1002/pmic.200900800
24. Schmidt A, Csaszar E, Ammerer G, Mechtler K (2008) Enhanced detection and identification of multiply phosphorylated peptides using TiO₂ enrichment in combination with MALDI TOF/TOF MS. *Proteomics* 8(21):4577–4592. doi:10.1002/pmic.200800279
25. Griffin PR, MacCoss MJ, Eng JK, Blevins RA, Aaronson JS, Yates JR 3rd (1995) Direct database searching with MALDI-PSD spectra of peptides. *Rapid Commun Mass Spectrom* 9(15):1546–1551. doi:10.1002/rcm.1290091515
26. Lehmann WD, Kruger R, Salek M, Hung CW, Wolschin F, Weckwerth W (2007) Neutral loss-based phosphopeptide recognition: a collection of caveats. *J Proteome Res* 6(7):2866–2873. doi:10.1021/pr060573w
27. Boersema PJ, Mohammed S, Heck AJ (2009) Phosphopeptide fragmentation and analysis by mass spectrometry. *J Mass Spectrom* 44(6):861–878. doi:10.1002/jms.1599
28. Mischerikow N, Altelaar AF, Navarro JD, Mohammed S, Heck A (2010) Comparative assessment of site assignments in CID and ETD spectra of phosphopeptides discloses limited relocation of phosphate groups. *Mol Cell Proteomics* 9:2140–2148. doi:10.1074/mcp.M900619-MCP200
29. Bantscheff M, Eberhard D, Abraham Y, Bastuck S, Boesche M, Hobson S, Mathieson T, Perrin J, Raida M, Rau C, Reader V, Sweetman G, Bauer A, Bouwmeester T, Hopf C, Kruse U, Neubauer G, Ramsden N, Rick J, Kuster B, Drewes G (2007) Quantitative chemical proteomics reveals mechanisms of action of clinical ABL kinase inhibitors. *Nat Biotechnol* 25(9):1035–1044. doi:10.1038/nbt1328
30. Gnad F, Ren S, Cox J, Olsen JV, Macek B, Orosi M, Mann M (2007) PHOSIDA (phosphorylation site database): management, structural and evolutionary investigation, and prediction of phospho-sites. *Genome Biol* 8(11):R250. doi:10.1186/gb-2007-8-11-r250
31. Suckau D, Resemann A, Schuerenberg M, Hufnagel P, Franzen J, Holle A (2003) A novel MALDI LIFT-TOF/TOF mass spectrometer for proteomics. *Anal Bioanal Chem* 376(7):952–965. doi:10.1007/s00216-003-2057-0
32. Elias JE, Gygi SP (2007) Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nat Methods* 4(3):207–214. doi:10.1038/nmeth1019
33. Medzihradszky KF, Campbell JM, Baldwin MA, Falick AM, Juhasz P, Vestal ML, Burlingame AL (2000) The characteristics of peptide collision-induced dissociation using a high-performance MALDI-TOF/TOF tandem mass spectrometer. *Anal Chem* 72(3):552–558
34. Harvey DJ, Bateman RH, Bordoli RS, Tyldesley R (2000) Ionisation and fragmentation of complex glycans with a quadrupole time-of-flight mass spectrometer fitted with a matrix-assisted laser desorption/ionisation ion source. *Rapid Commun Mass Spectrom* 14(22):2135–2142. doi:10.1002/1097-0231(20001130)14:22<2135::AID-RCM143>3.0.CO;2-#
35. Verhaert PD, Pinkse MW, Strupat K, Conaway MC (2010) Imaging of similar mass neuropeptides in neuronal tissue by enhanced resolution MALDI MS with an ion trap–Orbitrap hybrid instrument. *Methods Mol Biol* 656:433–449. doi:10.1007/978-1-60761-746-4_25
36. Spengler B (2004) De novo sequencing, peptide composition analysis, and composition-based sequencing: a new strategy employing accurate mass determination by Fourier transform ion cyclotron resonance mass spectrometry. *J Am Soc Mass Spectrom* 15(5):703–714. doi:10.1016/j.jasms.2004.01.007
37. Nagaraj N, D’Souza RC, Cox J, Olsen JV, Mann M (2010) Feasibility of large-scale phosphoproteomics with higher energy collisional dissociation fragmentation. *J Proteome Res* 9(12):6786–6794. doi:10.1021/pr100637q
38. McAlister GC, Phanstiel DH, Brumbaugh J, Westphall MS, Coon JJ (2011) Higher-energy collision-activated dissociation without a dedicated collision cell. *Mol Cell Proteomics* 10(5):O111 009456. doi:10.1074/mcp.O111.009456
39. Edelson-Averbukh M, Shevchenko A, Pipkorn R, Lehmann WD (2009) Gas-phase intramolecular phosphate shift in phosphotyrosine-containing peptide monoanions. *Anal Chem* 81(11):4369–4381. doi:10.1021/ac900244e
40. Aguiar M, Haas W, Beausoleil SA, Rush J, Gygi SP (2010) Gas-phase rearrangements do not affect site localization reliability in phosphoproteomics data sets. *J Proteome Res* 9:3103–3107. doi:10.1021/pr1000225
41. Cox J, Mann M (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol* 26(12):1367–1372. doi:10.1038/nbt.1511
42. Savitski MM, Mathieson T, Becher I, Bantscheff M (2010) H-Score, a mass accuracy driven rescoring approach for improved peptide identification in modification rich samples. *J Proteome Res* 9:5511–5516. doi:10.1021/pr1006813