Published in final edited form as:

J Proteome Res. 2007 November ; 6(11): 4230–4244. doi:10.1021/pr070313u.

On-Line LC-MS Approach Combining Collision-Induced Dissociation (CID), Electron-Transfer Dissociation (ETD), and CID of an Isolated Charge-Reduced Species for the Trace-Level Characterization of Proteins with Post-Translational Modifications

Shiaw-Lin Wu¹, Andreas F.R. Hühmer², Zhiqi Hao², and Barry L. Karger^{1,a}

¹ Barnett Institute, Northeastern University, Boston, MA 02115.

² Thermo Fisher Scientific, San Jose, CA 95134.

SUMMARY

We have expanded our recent on-line LC-MS platform for large peptide analysis to combine CID, electron transfer dissociation (ETD), and CID of an isolated charge-reduced (CRCID) species derived from ETD to determine sites of phosphorylation and glycosylation modifications, as well as the sequence of large peptide fragments (i.e. 2,000 to 10,000 Da) from complex proteins, such as β -casein, epidermal growth factor receptor (EGFR), and tissue plasminogen activator (t-PA) at the low fmol level. The incorporation of an additional CID activation step for a charge-reduced species, isolated from ETD fragment ions, improved ETD fragmentation when precursor ions with high m/z ($\sim >1000$) were automatically selected for fragmentation. Specifically, the identification of the exact phosphorylation sites was strengthened by the extensive coverage of the peptide sequence with a near continuous product ion series. The identification of N-linked glycosylation sites in EGFR and an O-linked glycosylation site in t-PA were also improved through the enhanced identification of the peptide backbone sequence of the glycosylated precursors. The new strategy is a good starting survey scan to characterize enzymatic peptide mixtures over a broad range of masses using LC-MS with data-dependent acquisition, as the three activation steps can provide complementary information to each other. In general, large peptides can be extensively characterized by the ETD and CRCID steps, including sites of modification from the generated near continuous product ion series, supplemented by the CID-MS² step. At the same time, small peptides (e.g. $\leq 2+$ ions), which lack extensive ETD or CRCID fragmentation, can be characterized by the CID-MS² step. A more targeted approach can then be followed in subsequent LC-MS runs to obtain additional information, if needed. Overall, the recently introduced ETD provides not only useful structural information but also enhances the confidence of all assignments. The sensitivity of this new approach on the chromatographic time scale is similar to the previous Extended Range Proteomic Analysis (ERPA) using CID-MS² and CID-MS³. The new LC-MS platform can be anticipated to be a useful approach for the comprehensive characterization of complex proteins.

INTRODUCTION

The two most common mass spectrometric approaches for the characterization of proteins are direct analysis of intact proteins (top-down)^{1, 2} and analysis of a separated mixture of peptides resulting from a tryptic digest (bottom-up).^{3, 4} High sequence coverage has been the focus of top-down proteomics using high resolution mass spectrometers, e.g. FTMS.^{1, 5, 6} While

a To whom correspondence should be addressed. b.karger@neu.edu..

impressive results have been obtained, the method is relatively insensitive (at hundreds of fmols or higher), is not readily applicable to proteins with heterogeneous modifications above 50 kDa, and has for the most part not been applied to glycosylation analysis. On the other hand, the bottom-up approach is typically highly sensitive for peptide detection (at a few fmols or lower) but often suffers from low sequence coverage and is limited in providing comprehensive characterization of post-translational modifications of proteins.

Recently, we introduced an intermediate LC-MS approach using a hybrid FTICR MS with a linear ion trap employing collision-induced dissociation (CID) with MS² and MS³ steps - Extended Range Proteomic Analysis (ERPA) - that combines the advantages of reduction in the size and complexity of the sample with improved chromatographic and mass ionization efficiency of modified peptides.⁷ ERPA generally employs proteolytic enzymes such as Lys-C (C-terminal K) to cut proteins less frequently than trypsin (C-terminal R and K). As a consequence, the average molecular weight distribution of peptide fragments is typically greater than that with tryptic digests, leading to larger fragments and simpler mixtures (average 2 to 3 times larger in mass and 2 to 3 times lower in number of peptide fragments). Using the ERPA approach with a 50 μ m ID polystyrene-divinyl benzene (PS-DVB) monolithic column, high sequence coverage (~95%) at the low fmol level for the tyrosine kinase membrane protein, epidermal growth factor receptor (EGFR), was obtained in addition to information associated with the specific sites and structure of phosphorylation and glycosylation modifications.^{7, 8}

Electron transfer dissociation (ETD) is a newly developed fragmentation method^{9, 10}, which is related to electron capture dissociation (ECD)^{11, 12} in that labile post-translational modifications are preserved while the backbone of the peptide is fragmented to yield c and z product ions. Often, peptides with charge states of 3+ or higher are required for effective fragmentation.^{13 - 15} Small peptides with predominantly 2+ charge states have been shown to exhibit poorer fragmentation efficiency with ETD or ECD.^{13 - 15} Large proteolytic peptides (e.g. Lys-C digestion) inherently carry additional charges to yield peptide charge states of 3+ or higher.^{16, 17} Thus, ETD should be well suited to the large peptides.¹⁷ In addition, because of the ability of ETD fragmentation to retain labile post-translational modifications, the deglycosylation step, often required to determine peptide backbone sequence in glycopeptide identification (for N- and O-linked glycopeptides), may be unnecessary.

The purpose of the present paper is to examine the use of both CID and ETD activation steps for characterization of complex proteins, particularly of large peptides (e.g. 2,000 to 10,000 Da) using on-line LC-MS. We explore the advantages of ETD relative to our previous platform which consisted of CID-MS² and CID-MS³ in conjunction with high resolution and accurate precursor mass measurement. To provide a basis of comparison, particularly on the chromatographic time scale, we have selected the previously studied proteins, β -casein and epidermal growth factor receptor at 50 to 75 fmol level^{7, 8}, using an LC column of 50 μ m i.d. polystyrene-divinyl benzene (PS-DVB) monolith. We have also studied tissue plasminogen activator to determine O-linked glycosylation sites. In some cases, if precursor ions with lower charge states are automatically selected for fragmentation in data-dependent acquisition, poor ETD fragmentation efficiency is developed, with significant product ions being charge-reduced (odd electron) species. Thus, we examine the capability of an additional CID activation step on a charge-reduced species isolated from the ETD fragment ions. Although in many cases, ETD fragmentation alone is significant^{17, 32}, this step, fragmentation of the isolated charge-reduced (CR) species by CID (CRCID), is shown to provide additional product ion series (c and z ions), particularly for large m/z peptide ions (m/z ~ >1000), to augment information from the ETD fragmentation process in the identification of the specific sites of modification. A related method is use of a supplemental activation step to enhance fragmentation of all ETD or ECD product ions.^{13, 14} The merits of on-line fragmenting a single isolated charge-reduced

species, as in the present work, to generate cleaner and easier to interpret spectra are discussed in the following.

EXPERIMENTAL PROCEDURES

Reagents

Achromobacter protease I (Lys-C) was obtained from Wako Co. (Richmond, VA). The proteins, β -casein from milk and human epidermal growth factor receptor (EGFR) from an A431 cancer cell line, as well as dithiothreitol (DTT), iodoacetamide (IAA), fluoranthene, guanidine hydrochloride and ammonium bicarbonate, were obtained from Sigma-Aldrich (St. Louis, MO). Recombinant human tissue plasminogen (t-PA) was obtained as a gift from Genentech, Inc. (So. San Francisco, CA). Formic acid, acetone and acetonitrile were purchased from Fisher Scientific (Fair Lawn, NJ), and the HPLC-grade water, used in all experiments, was from J.T. Baker (Bedford, MA).

Enzymatic digestion

For β -casein (1 mg/mL), the endoproteinase Lys-C was added in a 1:100 (w/w) ratio, and the solution was incubated for 4 hrs at 37 °C. EGFR was received as a lyophilized powder containing 500 units of the protein. Recombinant t-PA was received as a lyophilized powder containing 2 mg of the protein. The powder (~1 pmol of EGFR or t-PA) was reconstituted with 200 μ L of 6 M guanidine hydrochloride, reduced with 20 mM DTT for 30 min at 37 °C and alkylated in the dark with 50 mM of IAA for 1.5 hr at room temperature. After desalting over a Microcon spin column (10 kDa MWCO; Millipore, Bedford MA), the endoproteinase Lys-C (1:100 w/w) was added to digest the protein for 4 hr at 37 °C. Digestion was stopped by addition of 1% formic acid.

LC-MS

LC-MS experiments were performed on a prototype LTQXL with ETD (Thermo Fisher Scientific, San Jose, CA), consisting of a newly developed linear ion trap (LTQXL) with an additional chemical ionization source to generate fluoranthene anions within the CI source, as described previously.¹⁶ An Agilent 1100 capillary system (Agilent Technologies, Palo Alto, CA) was used to separate the samples with a monolithic column (polystyrene-divinylbenzene, PS-DVB, 50 μ m i.d. $\times$ 10 cm) prepared in-house.¹⁸ The column was coupled on-line to the LTQXL with ETD mass spectrometer. Mobile phase A was 0.1% formic acid in water, while mobile phase B was 0.1 % formic acid in acetonitrile. The gradient consisted of: (i) 20 minutes at 0 % B for sample loading, (ii) linear from 0 to 40% B over 40 min, then (iii) linear from 40 to 80% B over 10 min, and finally (iv) isocratic at 80% B for 10 min. The flow rate of the column (at the initial mobile phase condition) was measured as ~100 nL/min.

The mass spectrometer was operated in the data-dependent mode to switch automatically between MS (scan 1), CID-MS² (scan 2), ETD-MS² (scan 3), and CID-MS³ (scan 4) (see Figure 1). The CID-MS³ step (scan 4) is called the charge-reduced CID (CRCID) step to fragment the charge-reduced species. Briefly, after a survey full-scan MS spectrum from m/z 400 to 2000 in the linear ion trap (at a target value of 30,000 ions), subsequent CID-MS² (at a target value of 30,000 ions and 35% normalized collision energy) and ETD-MS² (at a target value of 30,000 ions) activation scan steps were performed on the same precursor ion over the same m/z scan range as that used for the full scan MS spectrum. The precursor ion was isolated using the data-dependent acquisition mode with a $\pm$ 2.5 m/z isolation width to select automatically and sequentially a specific ion (starting with the most intense ion) from the survey scan. Then, an additional CRCID step (at a target value of 30,000 ions and 10% normalized collision energy with the decrease of the activation Q value from 0.25 to 0.15, 2 microscans) was performed on an isolated precursor ion with a $\pm$ 5 m/z isolation width and

with the highest intensity from the ETD-MS² scan. Scans 2, 3, and 4 were repeated an additional 2 times in sequence to select for fragmentation the second and third highest intensity precursor ions from the first survey scan. The CI source parameters, such as ion optics, filament emission current, anion injection time (anion target value set at 2e5 ions), fluoranthene gas flow, and CI gas flow were optimized automatically. The ion/ion reaction duration time was maintained constant throughout the experiment at 100 ms. In most cases, the generation of several charge-reduced species with high intensity in the ETD spectrum allowed the determination of the charge state of the large peptide (precursor ion). The intensity of the charge-reduced species could be enhanced further, if needed (e.g. decreased the ion/ion reaction duration time to 30 ms). In order to label each assignment clearly, some of the background noise in the figures have been reduced. For further confirmation, an LTQ-FT MS (Thermo Fisher Scientific) with an Ultimate 3000 nanoLC pump (Dionex, Mountain View, CA) and a home-made monolithic column (PS-DVB, 50 μ m i.d. $\times$ 10 cm) was at times used to acquire full mass spectra in the FT-ICR (400 to 2000 m/z) at 100,000 resolution (at a target value of 2 million ions) to determine the accurate mass and charge states of the precursor ions generated under the similar conditions on the LTQXL-ETD MS instrument. If two or more similar m/z precursor ions appear at a similar retention time, we then add the CID-MS² spectrum pattern to track the correct m/z precursor ion between the LTQ-ETD and LTQ-FT runs.

Peptide assignment

Spectra generated on the LTQXL with ETD MS instrument were filtered using BioWorks software (3.3.1, Thermo Fisher Scientific) that has the Sequest algorithm incorporated to assign fragmentation spectra to the most probable peptide sequence. Briefly, the spectra generated in CID step were searched against spectra of theoretical fragmentations (b and y ions) of a human SwissProt annotated database downloaded in Jan 2006 which contains 14094 protein entries with a mass tolerance $\pm$ 1.4 Da (for both precursor and fragment ion tolerance) and with Lys-C specificity (2 missed cleavages). The resultant spectra were filtered using the scores of Xcorr (1+ precursor ion $\geq$ 1.5, 2+ $\geq$ 2.0, and 3+ and above $\geq$ 2.5). The spectra generated in the ETD or CRCID steps were searched against spectra of theoretical fragmentations (c and z ions) of the same human SwissProt database but filtered using the scores of Xcorr ($\geq$ 1). Final confirmation of the most probable peptide assignments was obtained by inspection of individual spectra with the preferred fragmentation patterns in the observed CID-MS², ETD-MS² and CRCID spectra, as detailed in the Results Section. Glycopeptides were manually assigned, as described previously.^{7, 8}

RESULTS

In the following, we examine the analysis of three complex proteins, β -casein, epidermal growth factor receptor (EGFR), and tissue plasminogen activator (t-PA) by LC-MS using a combination of CID and ETD activation. Several characteristic large peptides of each protein, with and without post-translational modifications, are used to illustrate the fragmentation of CID and ETD. As mentioned earlier, if precursor ions with higher m/z (lower charge states) are automatically selected for fragmentation in data-dependent acquisition, limited ETD fragmentation can result.^{14, 15} As described below, we implement on-line an additional activation by CID in the MS³ mode to fragment an isolated charge-reduced species from the ETD fragmentation, thus providing an additional means of fragmentation of peptides with large m/z ($\sim$ >1000) that may not exhibit significant fragmentation by ETD activation alone.

Data acquisition strategies for ERPA using the LTQXL with ETD

The LTQXL MS with ETD is a linear ion trap utilizing two different ion activation processes, CID and ETD, both of which can be operated in the dependent and/or independent mode. When the instrument is operated in the dependent mode, the fragment ions generated from a given

activation process can be further fragmented by either CID or ETD. On the other hand, when operated in the independent mode, the same precursor ion can be fragmented by both CID and ETD in separate scan events.

The operation scheme for data acquisition in this work combines both in the dependent and independent modes, as shown in Figure 1. After the first survey scan (scan 1), CID and ETD are operated in the independent mode, selecting the same precursor ion (starting from the highest intensity ion) for fragmentation, as shown in scans 2 (CID) and 3 (ETD), respectively. After the ETD activation step (scan 3), the CID is operated in the dependent mode to select the most intense fragment ion which is isolated from the ETD scan for further fragmentation (scan 4). This last activation step (scan 4) is termed the charge-reduced CID-MS³ or CRCID step, where dissociation of the charge-reduced species (generally the highest intensity ion) created during the ETD activation step takes place. Scans 2, 3, and 4 are then repeated as scans 5, 6, and 7 to fragment the second highest precursor ion from the initial MS survey scan. Similarly, scans 8, 9, and 10 (not shown in the figure) represent a third repeat to fragment the third highest precursor ion from the first MS scan. The full cycle (10 scans – 1 survey MS scan plus 3 repeats of 3 different types of ion activation steps), requiring approximately 3 seconds, is continuously repeated during the entire LC-MS run under data-dependent and dynamic exclusion conditions. Using the data acquisition strategy in Figure 1, peptides with complex post-translational modifications, such as multiply phosphorylated or glycosylated peptides, can be substantially characterized in a single LC-MS run. It is also important to note that the generation of several charge-reduced species in the ETD spectrum generally allows one to deduce the precursor ion charges of large peptides even with the low resolution and limited mass accuracy in the linear ion trap, as illustrated below. A separate LC-MS run can be performed, if desired, using a high resolution mass spectrometer (e.g. LTQ-FT) to confirm the charge states and molecular weights of the precursor ions. In the near future, ETD coupled to the high resolution and accurate mass spectrometer, Orbitrap³³⁻³⁵, will become available, the charge state of the precursor ion can be directly measured in the same run.

The data acquisition scheme described in Figure 1 was used to analyze β -casein, EGFR, and t-PA. Both β -casein and EGFR were previously characterized using CID activation alone.^{7, 8} In that work, the experimental analysis scheme consisted of a combination of one survey scan using the FTICR at 100,000 mass resolution with 4 paired CID-MS² and CID-MS³ scans using the linear ion trap. The full cycle time (9 scans - 1 survey FTMS scan plus 4 repeats of CID-MS² and CID-MS³ ion activation steps) was approximately 2.7 seconds, a time comparable to the scheme in Figure 1. These two proteins were selected to provide a basis of comparison between the information obtained in the CID and CID/ETD survey scan approaches.

Identification of Phosphopeptides (β -casein)

β -casein at the level of ~50 to 75 fmol per injection in a 50 μ m ID PS-DVB monolithic column, similar to our previous study⁷, was used in the following. The advantages of employing a narrow-bore monolithic column for large peptide separation have been discussed previously.^{7, 8} Since a main feature of ETD is that it can preserve labile modifications, we will focus on the identification of the key phosphopeptides of bovine β -casein in the following.

From the base ion chromatogram of Lys-C digested β -casein at the indicated elution time of 35.00 min (Figure 2A), a precursor ion in the 2+ charge state and an m/z of 1031.91, was selected for analysis. The ion was isolated using the data-dependent acquisition mode and subjected to CID fragmentation in the linear ion trap (Figure 2B). As expected for the CID-MS² fragmentation of this monophosphorylated peptide, the phosphorylation site labeled as **pS** in the sequence **FQpSEEQQTEDELQDK** (2062 Da), revealed a small number of high-

intensity, neutral loss fragments, typical for doubly charged phosphopeptides fragmenting by CID.^{7, 19}

Using the identical isolation procedure as for CID, the same precursor ion (m/z of 1031.91) was next selected for ETD fragmentation; however, as seen in Figure 2C, only a few low intensity fragments were produced. A significant level of unfragmented precursor ion remained after ETD activation. Next, this unfragmented precursor ion was isolated, and an additional CID step was applied to yield a fragmentation pattern in Figure 2D similar to the CID-MS² spectrum shown in Figure 2B (neutral loss, y , and b ions). The results of Figure 2C are anticipated as peptides with 2+ precursor ions have been shown to produce much poorer ETD or ECD fragmentation efficiency than the same peptides with higher charge states (or lower m/z).¹³⁻¹⁵

The 3+ charge state (lower m/z , 688.76) of the same monophosphorylated peptide was next examined in the same LC-MS run following the identical procedure as in Figure 2. As shown in Figure 3B, CID fragmentation of the precursor ion of the monophosphorylated peptide in the 3+ charge state again yielded only several high-intensity preferred-cleavage fragments (i.e. neutral loss ions), comparable to the fragmentation pattern observed in Figure 2B. On the other hand, the ETD fragmentation of the same 3+ ion now produced a greatly increased c and z ion series (compare Figure 3C to Figure 2C). In the ETD fragmentation spectrum, the charge-reduced 2+ ion of m/z at 1031.74 (3+ ion with an odd electron) was detected within the mass window as the highest intensity ion (Figure 3C). This ion was isolated for further fragmentation by the additional CID step (CRCID), as shown in Figure 3D. Importantly, a near complete ion series with c and z ions within the mass detection window of the linear ion trap was observed (compare Figure 3D to Figure 3C). The charge-reduced species is likely an ETD fragmented peptide species that is held together by intramolecular non-covalent forces of van der Waals and/or hydrogen bonding.^{9, 14, 15} Thus, by addition of kinetic energy to the charge-reduced species through the CRCID step, the ions break apart with a substantial ETD fragmentation pattern.

It is interesting to note in passing that in other studies^{14, 15}, peptides with 2+ precursor ions have been provided with additional activation energy after the ETD activation step to fragment the charge-reduced species (i.e. the 1+ ions) resulting in a c and z ion series. However, in the present 2+ precursor ion example (Figure 2C), the 1+ charge-reduced species (m/z at 2062) was beyond the mass detection window of the ion trap mass spectrometer (limited to m/z at 2000 in this experiment) and was therefore not available for isolation and further fragmentation. Enlarging the acquisition mass window to 4000 m/z did allow observation and fragmentation of the 1+ charge-reduced species (m/z at 2062) but the detection sensitivity was significantly decreased ~5 to 10 fold (data not shown). With the 3+ charge state precursor ion (Figure 3), however, the charge-reduced species, which was the highest intensity ion, was found within the mass detection window (m/z at 1031.74 in Figure 3C), and was then automatically selected for CRCID fragmentation. .

In Figure 3D, the phosphorylated serine (pS) site was clearly identified by high abundant product ions using the CRCID activation step, i.e. the z_{13} and z_{14} ions. These bond cleavages pinpointed the modification (+80 Da) at the S site, not at adjacent amino acids. If one only considers the phosphorylation modification, an incomplete ion series produced by CID may be sufficient⁷, since there are only two possibilities, given the peptide sequence. For monophosphorylated peptides (at a similar amount per injection), in many cases, we found that CID-MS² and CID-MS³ steps in our previous study were quite comparable to ETD and CRCID to identify the site of phosphorylation. However, when multiple phosphorylations occur on closely spaced multiple amino acid residues of S, T and Y on the same peptide, a comprehensive

coverage in the ion series is necessary for determination of the exact site(s) of modification, and ETD can clearly be very useful.

To illustrate the advantage of ETD to pinpoint specific modification sites for multiple residues, we next examined the tetraphosphopeptide of β -casein, which was present in the same Lys-C digested, using the analysis scheme detailed in Figure 1. In the same LCMS run as for the monophosphopeptide, at the indicated elution time of 61.03 min (Figure 4A), a precursor ion of the known tetraphosphorylated peptide RELEELNVPGEIVEpSLpSpSpSEESITRINK (3477 Da) with the 3+ charge state ion at 1160.12 m/z was selected for CID fragmentation. As shown in Figure 4B, the MS² spectrum of this peptide produced only a small number of high-intensity neutral loss ions. Similar to the previous observation in Figure 2C, the ETD fragmentation of this tetraphosphopeptide precursor ion (Figure 4C) yielded a much smaller number of low-intensity fragments than with CID fragmentation. A high level of unfragmented precursor ion still remained after ETD activation. Isolation of the 2+ charge-reduced species (m/z of 1740.04), followed by further fragmentation by CRCID, resulted in the spectrum shown in Figure 4D. A near complete ion series (c and z ions) within the mass detection window was now observed. The four pS sites (3 adjacent and 1 with a single amino acid residue spacing) were unambiguously identified by the peptide bond cleavages (i.e. z13, z12, z11, z10 and z9 ions), demonstrating the power of ETD.

The remaining unfragmented 3+ charged precursor ion (m/z of 1159.94 in Figure 4C), in contrast to its 2+ charge-reduced ion (m/z of 1740.04 in Figure 4C), produced a typical CID fragmentation pattern (neutral loss, b and y ions) (data not shown). This major difference in fragmentation pattern, between the unfragmented and charge-reduced species, was identical to the pattern observed for the monophosphopeptide in Figures 2 and 3.

It is important to note that as the molecular weight of the peptide becomes larger, the required charge state of the peptide for effective ETD fragmentation increases, as previously found in ECD fragmentation.^{12, 13} When the ETD fragmentation of the precursor ion with 4+ charge (870.25 m/z) of the same tetraphosphopeptide was examined, a higher number of bond cleavages were observed when compared to ETD fragmentation of the 3+ charge precursor ion (1160.12 m/z) (data not shown). In general, peptides with greater m/z ($\sim >1000$) appear to yield more charge-reduced species and less fragment ions (c and z ions) in ETD fragmentation. The bond cleavages for the CRCID step were nevertheless similar for the two precursors (3+ and 4+ charge states) when a common charge-reduced species (2+ ion of m/z at 1740.04) was selected for additional fragmentation (data not shown). However, the highest intensity for this tetraphosphopeptide was found for the 3+ charged species (the 4+ or higher charge species was approximately only 20 % of the 3+ precursor ion). The ability to select a higher charged species (low m/z) for effective ETD fragmentation was thus limited by the lower intensity of the peptide at higher charge states. Nevertheless, the combination of ETD activation with an additional CID activation (CRCID) step, regardless of the selection of charge state (or m/z), generally appears to lead to substantial c and z ion series to compensate for the disadvantage of the intensity-based selection process.

For this tetraphosphopeptide, the CID-MS² and even CID-MS³ steps in our previous study mainly produced multiple neutral loss ions.⁷ The ambiguity for the assignment of phosphorylation sites by CID, however, was partially compensated by the fact that the four phosphorylation sites were 100% modified. The ETD/CRCID steps in the present study greatly enhanced the confidence of the assignment through the observation of continuous bond cleavages at these phosphorylation sites. To our knowledge, this is the first time clear and direct evidence has been provided to assign the tetraphosphorylation sites of β -casein by mass spectrometry. Note that there are 6 possible S and T sites close to each other on the peptide. A substantial ion series generated by ETD/CRCID steps should be even more important for partial

PTM modification, where the spectral complexity would be much greater relative to full phosphorylation, e.g. kinases *in vivo*.

To explore further the fragmentation strategy in Figure 1, we next examined the epidermal growth factor receptor (EGFR) kinase, containing heterogeneous and partially modified phosphorylation and glycosylation structures.

Identification of Phosphopeptides (EGFR)

Similar to the above study for β -casein, a Lys-C digest of EGFR (50 to 75 fmol per injection using a 50 μ m I.D. PS-DVB monolithic column), was first evaluated for the identification of phosphorylation sites using the strategy in Figure 1. At the elution time of 56.70 min in the base peak ion chromatogram of the Lys-C digest of EGFR (Figure 5A), a precursor ion with a 5+ charge state and m/z of 759.24 was selected for CID fragmentation (the charge state was determined from the charge-reduced species in the ETD spectrum discussed below). As shown in Figure 5B, the highly charged phosphopeptide **RTLRRLLQERELVEPLpTPSGEAPNQALLRILK** (3789 Da) produced a small number of high-intensity neutral loss ions. The ETD fragmentation of the 5+ phosphopeptide precursor ion produced an ion series at the N and C terminals along with the charge-reduced species of the peptide, see Figure 5C. The generation of several charge-reduced species with high intensities in the ETD spectrum (labeled as $[M+5H]++++\bullet$, $[M+5H]+++$, and $[M+5H]++$) allowed the determination of the charge state of the precursor ion to be 5+, in agreement with what was found previously on the LTQ-FT MS.⁷ After isolating the charge-reduced species, the 2+ ion at m/z 1897.23 (highest intensity ion) was further fragmented by CRCID, as shown in Figure 5D. A large number of peptide bond cleavages with a continuous ion series encompassing the middle region of the peptide were observed. Interestingly, although the ETD and CRCID activation steps produced c and z ions, the peptide bond cleavages did not significantly overlap for this large phosphopeptide, as shown in Figure 5C and 5D. This additional information provided by CRCID can be important in the characterization of complex peptides. It is also to be noted, the other highly abundant fragment ion in Figure 5C (m/z 1135.54), which was not a charge-reduced species of this precursor (based on molecular weight) could be additional precursor ion with 3+ charge (based on molecular weight) to be co-isolated and fragmented in ETD. Nevertheless, the CRCID step, which only isolated and fragmented the charge-reduced species of the 5+ precursor, appeared to minimize this overlap problem.

The exact pT site of this phosphopeptide was assigned and confirmed by the combination of CRCID and CID cleavages (z14 and z16 ions in Figure 5D, and y18 and y15 ions in Figure 5B). Since the pT site was adjacent to a proline residue, the ETD (or CRCID) activation step could not break this bond (z15 ion not present).⁹ On the other hand, CID activation produced a highly abundant product ion at this cleavage site (y15 ion in Figure 5B). Since pT or pS are often located in close proximity to proline residues in many kinase proteins²⁰, a preferred cleavage by CID at proline peptide bonds can be important in assigning phosphorylation sites, particularly when there are two prolines surrounding the pT site, as for this phosphopeptide. In this case, our previous ERPA strategy with CID-MS² and CID-MS³ steps provided site assignment (high abundant and characteristic preferred cleavages).⁷ Nevertheless, the combination of CID, ETD, and CRCID activation steps, which all indicated phosphorylation at the pT site, enhanced confidence in the phosphorylation assignment.

It is significant to note that the additional charge-reduced species in Figure 5C, $[M+5H]+++ + \bullet$ (948.26 m/z) and $[M+5H]++++\bullet$ (1264.61 m/z) ions, when isolated in a subsequent run, produced similar but slightly different bond cleavages (c and z fragment ions) in comparison to $[M+5H]+++$ (1897.23 m/z) charge-reduced species in CRCID step, specifically in the low m/z range (molecular weight cutoff in the ion trap mass spectrometer) (data not shown).

Nevertheless, the 5+ precursor ion provided three charge-reduced species within the mass detection window available for further fragmentation in the CRCID step, enhancing the confidence of the assignment through the repeated observation of overlapping key fragment ions.

We next examined the ability of the current approach to identify a highly complicated phosphopeptide of EGFR. The sequence of the phosphopeptide (Lys-C fragment) is EDSFLQRYSSDPTGALTEDSIDDTFLPVPEYINQSVPK (4353 Da). There are 10 potential phosphorylation sites in this peptide sequence (amino acid sites in bold), with three sites adjacent to each other (underlined). For this phosphopeptide, the 3+ and 4+ charged precursor ions had similar intensity, and the automated data-dependent MS/MS mode selected both charge states for fragmentation. The 3+ precursor ion (1452.68 m/z) was found to have greater CID fragmentation than the 4+ precursor ion. On the other hand, the 4+ precursor ion (1086.75 m/z, lower m/z) had greater ETD fragmentation than its 3+ precursor ion (higher m/z). Thus, the 3+ precursor ion (for CID fragmentation) and the 4+ precursor ion (for ETD fragmentation) are presented in Figure 6.

In the same LC-MS run as in Figure 5, the precursor ion with a 3+ charge state and an m/z of 1452.68 at the indicated elution time of 63.89 min (Figure 6A) was selected for CID fragmentation. As shown in Figure 6B, CID-MS² of this phosphopeptide (the 3+ precursor ion) produced a small number of high-intensity preferred cleavage fragments (i.e. neutral loss ions) rather than an ion series. The observation of b8 and b9 ions in Figure 6B indicated the phosphorylation location at the pS site (pS in the peptide sequence of Figure 6B). However, in contrast to the previous phosphopeptide examples, a neutral loss of 80 Da (i.e. HPO₃), instead of the typical 98 Da (i.e. H₃PO₄), was observed. Interestingly, this 80 Da neutral loss (breaking the bond between O and P in the phospho-group) often occurs in CID fragmentation when the phosphate is attached to Y sites.¹⁹ This neutral loss could thus suggest assignment of the phosphorylation site to the adjacent pY instead of pS, particularly, if the signals of b8 and b9 ions are relatively low (see Figure 6B). In our previous work, the CID-MS³ step did not yield useful site information (b8 or b9 ion intensity are too low for further fragmentation). For this assignment, we needed LC separation. When EGFR was stimulated with EGF, the pY monophosphorylated site eluted at a different retention time than the pS site, and this allowed us in previous work to indirectly assign the pS site.⁸

To avoid the ambiguity from the CID fragmentation pattern, we next examined the phosphopeptide by ETD fragmentation of the 4+ precursor ion and produced an ion series at the N and C terminals along with the charge-reduced species of the peptide, see Figure 6C. From the generation of the 3+ charge-reduced species ([M+4H]^{+++•}) in the ETD spectrum, we were able to deduce the precursor ion with 4+ charge. As seen in Figure 6C, key fragment ions necessary to locate the phosphorylation site were not observed in the ETD spectrum. The other highly abundant fragment ion in Figure 6C (m/z 1632.79), which could be another precursor ion with 3+ charge (based on molecular weight) appeared to be co-isolated and fragmented in ETD, as described in the previous example for Figure 5C. We next examined the CRCID of the charge-reduced 3+ ion species ([M+4H]^{+++•}). As shown in Figure 6D, the phosphorylation site was pinpointed at the indicated pS site by the observation of c7, c8, c10, z28, z29, z30, and z31 ions in the CRCID spectrum. The observation of key bond cleavages from both the C- and N-terminal ends (i.e. both c and z ions) eliminated many potential phosphorylation sites for this peptide and greatly enhanced the confidence of the specific assignment.

As previously noted⁸, the estimated stoichiometry of EGFR phosphorylation was quite low prior to EGF stimulation, i.e. ~0.5% for the pS1046 site. With the total level of 50 to 75 fmol

per injection, the amount of this phosphopeptide was estimated to be at the high attomole level, which should be close to the limit of detection using ETD with a 50 μm i.d. monolithic column.

Identification of N-linked Glycopeptides (EGFR)

In our previous study of EGFR⁷, a deglycosylation step was necessary for identification of the peptide and the site of glycosylation. The reanalysis of the deglycosylated sample assumed that the deglycosylated peptide eluted at approximately the same retention time as its glycosylated counterpart.^{7, 8, 20} After obtaining the backbone sequence of the peptide from the deglycosylated species, the glycan modification on the peptide was then estimated by subtraction of the molecular weight of the peptide backbone sequence from that of the precursor ion (glycopeptide). The deglycosylation step, however, can be quite complicated to analyze for a complex mixture containing a number of co-migrating glycopeptides. Moreover, it will generally not be successful for O-linked glycopeptides because of the lack of available glycosidases for enzymatic deglycosylation or suitable chemical deglycosylation (e.g. β -elimination) without interferences from phosphorylation modifications.²² Thus, to avoid the deglycosylation step, the fragmentation strategy in Figure 1 is employed in the following.

Since EGFR contains both N-linked glycosylation as well as phosphorylation modifications, the glycopeptides of the Lys-C digest of EGFR can be identified in the same LC-MS run as for the phosphopeptides. As shown in the base ion chromatogram of Figure 7A, a precursor ion with a 5+ charge state with m/z of 1142.73 at the indicated elution time of 62.95 min, was selected for CID fragmentation. In the CID MS² (Figure 7B), the glycosylation site labeled as N* in the sequence N*CTISGDLHILPVAFRGDSFTHTPPLDPQELDILK (5706 Da), produced, as expected, almost no peptide backbone cleavage, but rather glycosidic bond cleavages. Without significant peptide backbone cleavage, the peptide sequence could not be assigned. As noted, the CID-MS³ step in our previous study produced further glycosidic cleavages and/or limited peptide backbone sequence with partially cleaved glycan still attached to the product ions. Both fragmentation processes could only be used for structure confirmation but not peptide sequence determination.^{7, 8}

The determination of the backbone peptide sequence using the ETD activation step was next explored. As shown in Figure 7C, ETD fragmentation of the 5+ glycopeptide precursor ion produced an ion series at the C terminal along with the charge-reduced species, 4+ ([M+5H]⁺⁺•) and 3+ ([M+5H]⁺⁺⁺••) of the peptide. As described earlier, the charge state (5+) of the precursor ion could be determined from these charge-reduced species. The 3+ charge-reduced ion of m/z at 1903.38 was isolated for further fragmentation by CRCID, as shown in Figure 7D. A large number of peptide bond cleavages with a continuous ion series at the N-terminal portion of the peptide were found. Due to the high molecular weight of the glycan located on the N-terminus, only z ions were observed in this mass window. In comparison to the CID-MS² step, little glycosidic cleavages were observed with either the ETD or CRCID activation.^{26, 27} However, in contrast to the phosphorylation assignment, the molecular weight of the glycosylation is often unknown and not readily predictable. Thus, the determination of the backbone peptide sequence was still not possible by peptide identification software (e.g. Sequest) using a reasonable assumption of molecular weights for modifications in the database search. Manual inspection (*de novo* sequencing) of the fragment ions in Figures 7C and 7D led to three recognizable partial sequences, “ELDI”, “FR”, and “TSISG”, which were generated among other possible candidates. Note: replacing “I” (isoleucine) with “L” (leucine) or “D” (aspartic acid) with “N” (asparagine) must also be considered because neither the isobaric isomers (I, L) nor the molecular weights (D, N) which differ by 1 dalton could be differentiated by the linear ion trap. Nevertheless, using all possible candidates from the two largest sequence tags, along with the predicated Lys-C cleavage (i.e. C-terminal K) within the limited precursor molecular weight range, it was possible to assign the correct peptide sequence from the entire

Swissprot human database. Additionally, the identification was further confirmed by the presence of the known consensus sequence (N - X - S/T) for the N-linked glycopeptide.

After determining the peptide backbone sequence, we then returned to interpret the glycan structure from the glycosidic cleavages generated in Figure 7B.^{7, 24} A peptide with high-mannose (Man-8) glycostructure at the exact location (N337) was determined without the need of the deglycosylation step nor the need to know the attached glycan molecular weight. In this glycopeptide assignment, the ETD and CRCID fragmentation steps were mainly used for the peptide backbone sequence and glycosylation site assignment, and the CID fragmentation for assignment of the glycan structure. The ETD, CRCID, and CID activation steps produced complementary information necessary for the glycosylation structure determination, similar to the work described by others using the combination of ECD and CID.^{25, 26}

Beyond high mannose structures, the identification of other glycan classes generally requires higher stages of CID fragmentation.^{7, 8} The analysis of complex-type glycostructures depends mainly on the fragmentation from CID, since ETD under our conditions generally produces little glycan fragmentation. Moreover, the determination of the peptide backbone sequence by ETD (or ECD) is the same for N-linked^{23, 27-29} or O-linked glycopeptides.^{16, 25, 26} Thus, we may anticipate that the process of determination of all types of glycopeptides will be similar to that in Figure 7, with the incorporation of CID-MS³ or higher steps for further glycan fragmentation in additional runs. Since glycopeptide assignments involve many steps (e.g. *de novo* sequencing, searching by sequence tags, and glycan fragmentation matching), the future development of effective software to streamline these steps will be very important to facilitate the identification process. Fragmentation of charge-reduced species to produce a significant coverage of the ion series (or sequence tags) will often be necessary, because the ability to assign the correct peptide backbone sequence will depend heavily on sufficient information in the spectra generated in the ETD and CRCID steps. With this in mind, a highly charged Lys-C fragment will likely have more useful precursor ions or produce more charge-reduced species within the mass detection window than a small tryptic peptide modified with high molecular weight glycans.

Identification of O-linked Glycopeptides (t-PA)

It can be anticipated that the elimination of the deglycosylation step will be even more beneficial for determination of O-linked glycopeptides since there is currently no simple way to distinguish deglycosylation of O-linked glycans from phosphorylation modifications.²² To illustrate the power of the strategy in Figure 1, we chose as example a glycoprotein with a known O-linked glycosylation site, tissue plasminogen activator (t-PA).³⁰

At the elution time of 58.40 min in the base peak ion chromatogram of the Lys-C digest of t-PA (Figure 8A), a precursor ion with a 4+ charge state (determined from the charge-reduced species in the ETD spectrum) and *m/z* of 1004.18 was selected for CID fragmentation. As shown in Figure 8B, the O-linked glycopeptide SCSEPRCFNGGT*CQQALYFSDVCQCPEGFAGK (4009 Da) produced only a small number of high-intensity neutral loss ions. The loss of fucose in every parent and product ion associated with the fucosylation site produced a spectrum that was difficult to interpret without pre-knowledge of the fucose component. As shown in Figure 8C, the ETD fragmentation of the 4+ glycopeptide precursor ion produced an ion series at the C terminal of the peptide along with the charge-reduced species, see Figure 8C. Again, the generation of 3+ charge-reduced species ($[M+4H]^{+++•}$) in the ETD spectrum allowed one to deduce the precursor ion with 4+ charge. After isolating the charge-reduced species, the 3+ ion at *m/z* 1339.17 was further fragmented by CRCID, as shown in Figure 8D. A large number of additional peptide backbone cleavages with a near continuous ion series encompassing the middle region of the peptide was observed. Manual inspection of the fragment ions in Figure 8C and 8D, led to three partial

sequence tags, “FGE”, “QQALYF”, and “FNG” to assign the correct peptide sequence from a Swissport human database, as described earlier..

It should be noted that in the CID-MS² spectrum (Figure 8B), the neutral loss of the O-linked peptide at threonine differs from the neutral loss of phosphothreonine or phosphoserine in that the latter are often accompanied by an additional water loss to form a dehydroalanine-like threonine or serine product ion. This dehydroalanine-like product ion can then be further fragmented in the CID-MS³ step to pinpoint the location of phosphorylation site.^{7, 8} Without the additional water loss in this O-linked peptide, it can be difficult to locate the fucosylation site even using the CID-MS³ step, as observed by others as well.³¹ It should also be noted that in the CID-MS² spectrum, almost every product ion with the fucosylation site had this atypical neutral loss of fucose (Figure 8B); in contrast, very little fucose cleavage was observed with either ETD or CRCID activation. Thus, the initial peptide backbone sequence and the site of modification can rely heavily on the fragmentation of ETD and CRCID steps for the O-linked glycopeptide assignment, similar to the fragmentation pattern observed for the N-linked glycopeptides described above. After determining this atypical neutral loss, the CID-MS² spectrum (in Figure 8B) can then be used to assure the final assignment of the peptide backbone sequence along with the accurate precursor mass measurement.

Identification of Large Peptides Without PTM (EGFR)

Although a major strength of using ETD and CRCID is in the identification of the sites of PTM modifications, the precise assignment of unmodified peptides is still important for comprehensive protein characterization. For example, a possible truncation or mutation of a protein can be elucidated through determination of high sequence coverage.^{7, 8} In addition, the identification of large peptides will increase the confidence of the protein assignment. Thus, we next explored the ability of ETD and CRCID to identify high molecular Lys-C peptides without PTM modifications.

Large unmodified peptides of EGFR (Lys-C fragments) can be identified in the same LCMS run as above. As shown in the example of Figure 9A, a large peptide at the indicated elution time of 52.29 min (a precursor ion with a 6+ charge state and m/z of 1133.04) was selected for CID fragmentation. CID MS² of this peptide with the sequence RPAGSVQNPVYHNQPLNPAPSRDPHYQDPHSTAVGNPEYLNVTQPTCVNSTFDSPA HWAQK (6789 Da) produced only a small number of high-intensity preferred-cleavage fragments (Figure 9B). The ETD fragmentation of the 6+ peptide produced an ion series on the N terminal side of the peptide along with the charged reduced species, 4+ ([M+6H]⁺⁺⁺ +••) and 5+ ([M+6H]⁺⁺⁺⁺⁺ +•), as shown in Figure 9C. Again, the generation of 4+ and 5+ charge-reduced species in the ETD spectrum led to the precursor ion charge state of 6+. After isolating the charge-reduced species, the 4+ ion at m/z 1699.00 (the highest intensity ion) was selected for further fragmentation by CRCID, as shown in Figure 9D. A greater number of peptide bond cleavages with a near continuous ion series close to the C terminal side of the peptide was observed. A few characteristic CID fragment ions (i.e. b₂₃, y₆, and y₁₉ ions) were also found in this CRCID spectrum. The observation of a range of intensities of c and z ions in the ETD fragmentation would appear to be related to the location of the positively charged amino acids (i.e. R, K, and H) in the C or N-terminal side of peptide sequence, similar to our previous observation of b and y ions in CID fragmentation.⁷ As shown in Figure 9, a greater number of higher intensity c fragment ions in ETD/CRCID (as well as b ions in CID) were observed for this peptide sequence with the multiple arginine residues on the N-terminal side.

For this large peptide, a precursor ion with a 7+ charge state and m/z of 971.30 (~60% intensity of the 6+ precursor ion) was also automatically selected for fragmentation, as illustrated in Figure 10. Similar to the 6+ precursor ion in Figure 9B, CID MS² of this peptide (7+) produced only a small number of high-intensity preferred-cleavage fragments (Figure 10B). On the other

hand, the ETD fragmentation of the 7+ precursor with lower m/z (971.30) produced more extensive bond cleavages than its 6+ precursor ion with higher m/z (1133.04 m/z), compare Figure 10C and 10D. As previously, the high charge precursor ion (6+ in this case) did not yield efficient ETD fragmentation; large peptides with lower m/z (e.g. $< \sim 1000$ m/z) appear to have greater ETD fragmentation than the identical peptides with higher m/z . In addition, it appears that the ETD spectrum of the 7+ precursor ion (971.30 m/z) has even more extensive cleavages than CRCID of the charge-reduced species in this case (compare Figure 10C and Figure 9D). For comparison purposes, the ETD spectra of the 7+ precursor ion (Figure 10C) and its 6+ precursor ion (Figure 10D) are displayed. The bond cleavages for the CRCID step were similar when a common charge-reduced species ($[M+6H]^{++++\bullet\bullet}$ and $[M+7H]^{++++\bullet\bullet\bullet}$ ions of m/z at 1699) was selected for additional fragmentation (data not shown).

CONCLUSIONS

The on-line LC-MS fragmentation platform of Figure 1 has been shown to be a good starting point to characterize comprehensively peptide digest mixtures. Clearly, any enzymatic digestion of complex proteins will inevitably produce peptide fragments with a range of masses and m/z values in a mass spectrometer using electrospray ionization. Large peptides, with or without modification, can be extensively characterized by ETD (low m/z peptides) followed by CRCID (high m/z peptides), leading to comprehensive peptide bond coverage with extensive product ion series. In the same LC-MS run, the small peptides ($\leq 2+$ ions), which generally lack significant ETD or CRCID fragmentation can still be characterized by the CID-MS² step (see Figure 2). For modified peptides, the complementarity of CID, ETD, and CRCID can maximize the information content for the structure elucidation in the first survey run. As required, a more targeted approach (e.g. ETD/CRCID or CID-MSⁿ of specific ions) can then be followed to obtain additional information.

This paper has shown that the on-line isolation of the ETD-resultant charge-reduced species with additional activation by CID in the MS³ mode (CRCID) provides an additional means of efficient peptide fragmentation of large peptides (i.e. > 1000 m/z) that do not exhibit significant fragmentation by ETD activation alone. In some cases, the ability to select a higher charged species (low m/z) for effective ETD fragmentation was not feasible because the intensities of the peptide at higher charge states were low, particularly at the trace level as used in this study. In addition, even if ETD fragmentation is significant^{17, 32}, CRCID may still provide complementary backbone information. Others have suggested the use of a supplemental activation step to enhance ETD or ECD fragmentation^{13, 14, 23}, in which all product ions from ETD or ECD fragmentation are subjected to activation, not just an isolated charge-reduced species. While a supplemental activation step can be useful for doubly protonated peptide precursors, large peptides with PTM modifications typically generate cleaner and easier to interpret spectra from the activation of an *isolated* charge-reduced species in a separate scan. In particular, peptides with glycosylation modification rely heavily on the determination of peptide backbone sequence tags from *de novo* interpretation. Any glycosidic cleavages (i.e. from activating unfragmented species) may well interfere with the *de novo* interpretation of these peptide backbone cleavages.

For charge state determination of large peptides, the generation of charge-reduced species with high intensities in the ETD spectrum proved to be useful to deduce the precursor ion charges (up to 7+ charges in the current experiments). In the near future, when ETD is coupled to the high resolution Orbitrap mass spectrometer³³⁻³⁵, the charge state can be directly determined. Moreover, an appropriate chromatographic column with an open pore structure (i.e. the monolithic column used in this study), allowing high efficiency separation of large peptides, particularly glycosylated peptides, is also important.¹⁸

In the assignment of phosphorylation sites for monophosphorylated peptides or phosphorylation sites surrounded by prolines, our previous ERPA approach using CID-MS² and CID-MS³ steps seemed to be comparable to the current approach using ETD and CRCID steps. The great strength of the current strategy is in the identification of phosphorylation sites from closely spaced multiple amino acid residues of S, T and Y in a peptide sequence.

In the assignment of glycostructures and glycosylation sites of glycopeptides, the deglycosylation step is often required using CID-MS² and CID-MS³ steps. The current approach can determine the peptide backbone sequence in the ETD and CRCID steps without deglycosylation, which complements greatly the glycan structure information obtained by the CID steps. It should be noted that the glycosidic cleavages can be observed in the ETD or CRCID step, with an increase in ion-ion reaction time in ETD or the activation energy in the CRCID step, as well as inversely proportional to the peptide length (e.g. using tryptic instead of Lys-C fragments) (data not shown). Although these glycosidic cleavages may help in confirming the glycostructures, as reported recently²³, we attempted to minimize the glycosidic cleavages since the structure information on the carbohydrate can be obtained in the CID step. We can anticipate the current approach, as the initial survey scan, will maximize information on the glycostructure without the knowledge of the attached glycan molecular weight and without the deglycosylation step.

Overall, the current strategy which combines CID, ETD and CRCID activation steps on the chromatographic time scale for the comprehensive characterization of large peptides with phosphorylation and glycosylation modifications was achieved with similar sensitivity (~50 fmol per injection), as compared to the previous ERPA approach using CID-MS² and CID-MS³. Several EGFR phosphopeptides with a low estimated stoichiometry (~0.5%, attomole level) were identified in this study. The detection limit should be even lower when we apply this approach with even narrower porous-layer-open-tubular columns.³⁶ We can anticipate that in the future this new LC-MS approach will be a highly useful survey platform for comprehensive characterization of complex proteins with high sensitivity.

ACKNOWLEDGEMENTS

The authors thank NIH (GM 15847) for support of this research, and Genentech for the gift of recombinant tissue plasminogen activator. The authors also acknowledge Drs. Ian Jardine and Iain Mylchreest of Thermo Fisher Scientific for access to the LTQXL with ETD instrument for evaluation purposes. Contribution Number 889 from the Barnett Institute.

REFERENCES

1. McLafferty FW, Fridriksson EK, Horn DM, Lewis MA, Zubarev RA. Techview: biochemistry. Biomolecule mass spectrometry. Science 1999;284:1289. [PubMed: 10383309]
2. Meng F, Forbes AJ, Miller LM, Kelleher NL. Detection and localization of protein modifications by high resolution tandem mass spectrometry. Mass Spectrom Rev 2005;24:126. [PubMed: 15389861]
3. Link AJ, Eng J, Schieltz DM, Carmack E, Mize GJ, Morris DR, Garvik BM, Yates JR 3rd. Direct analysis of protein complexes using mass spectrometry. Nat Biotechnol 1999;17:676. [PubMed: 10404161]
4. Delahunty C, Yates JR 3rd. Protein identification using 2D-LC-MS/MS. Methods 2005;2005;35(3): 248. [PubMed: 15722221]
5. Sze SK, Ge Y, Oh H, McLafferty FW. Top-down mass spectrometry of a 29-kDa protein for characterization of any posttranslational modification to within one residue. Proc Natl Acad Sci 2002;99:1774. [PubMed: 11842225]
6. Wu SL, Jardine I, Hancock WS, Karger BL. A new and sensitive on-line liquid chromatography/mass spectrometric approach for top-down protein analysis: the comprehensive analysis of human growth

hormone in an E. coli lysate using a hybrid linear ion trap/Fourier transform ion cyclotron resonance mass spectrometer. *Rapid Commun Mass Spectrom* 2004;18(19):2201. [PubMed: 15384137]

7. Wu SL, Kim J, Hancock WS, Karger BL. Extended Range Proteomic Analysis (ERPA): a new and sensitive LC-MS platform for high sequence coverage of complex proteins with extensive post-translational modifications-comprehensive analysis of beta-casein and epidermal growth factor receptor (EGFR). *J Proteome Res* 2005;4(4):1155. [PubMed: 16083266]
8. Wu SL, Kim J, Bandle RW, Liotta L, Petricoin E, Karger BL. Dynamic profiling of the post-translational modifications and interaction partners of epidermal growth factor receptor signaling after stimulation by EGF using extended range proteomic analysis (ERPA). *Mol Cell Proteomics* 2006;5(9):1610. [PubMed: 16799092]
9. Syka JE, Coon JJ, Schroeder MJ, Shabanowitz J, Hunt DF. Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proc Natl Acad Sci* 2004;101(26):9528. [PubMed: 15210983]
10. Coon JJ, Ueberheide B, Syka JE, Dryhurst DD, Ausio J, Shabanowitz J, Hunt DF. Protein identification using sequential ion/ion reactions and tandem mass spectrometry. *Proc Natl Acad Sci* 2005;102(27):9463. [PubMed: 15983376]
11. McLafferty FW, Horn DM, Breuker K, Ge Y, Lewis MA, Cerda B, Zubarev RA, Carpenter BK. Electron capture dissociation of gaseous multiply charged ions by Fourier-transform ion cyclotron resonance. *J Am Soc Mass Spectrom* 2001;12(3):245. [PubMed: 11281599]Review
12. Zubarev RA. Reactions of polypeptide ions with electrons in the gas phase. *Mass Spectrom Rev* 2003;22(1):57. [PubMed: 12768604]Review
13. Horn DM, Ge Y, McLafferty FW. Activated ion electron capture dissociation for mass spectral sequencing of larger (42 kDa) proteins. *Anal Chem* 2001;72(20):4778. [PubMed: 11055690]
14. Swaney DL, McAlister GC, Schwartz JC, Syka JEP, Coon JJ. Supplemental activation method for high-efficiency electron-transfer dissociation of doubly protonated peptide precursors. *Anal Chem* 2007;79(2):477. [PubMed: 17222010]
15. Pitteri SJ, Chrisman PA, Hogan JM, McLuckey SA. Electron transfer ion/ion reactions in a three-dimensional quadrupole ion trap: reactions of doubly and triply protonated peptides with SO₂*-. *Anal Chem* 2005;77(6):1831. [PubMed: 15762593]
16. Schroeder MJ, Webb DJ, Shabanowitz J, Horwitz AF, Hunt DF. Methods for the detection of paxillin post-translational modifications and interacting proteins by mass spectrometry. *J Proteome Res* 2005;4(5):1832. [PubMed: 16212439]
17. Chi A, Huttenhower C, Geer LY, Coon JJ, Syka JE, Bai DL, Shabanowitz J, Burke DJ, Troyanskaya OG, Hunt DF. Analysis of phosphorylation sites on proteins from *Saccharomyces cerevisiae* by electron transfer dissociation (ETD) mass spectrometry. *Proc Natl Acad Sci* 2007;104(7):2193. [PubMed: 17287358]
18. Zhang J, Wu SL, Kim J, Karger BL. Ultratrace liquid chromatography/mass spectrometry analysis of large peptides with post-translational modifications using narrow-bore poly(styrene-divinylbenzene) monolithic columns and extended range proteomic analysis. *J Chromatogr A* 2007;1154(1-2):295. [PubMed: 17442327]
19. Tholey A, Reed J, Lehmann WD. Electrospray tandem mass spectrometric studies of phosphopeptides and phosphopeptide analogues. *J Mass Spectrom* Feb;1999 34(2):117. [PubMed: 12440389]
20. Schwartz D, Gygi SP. An iterative statistical approach to the identification of protein phosphorylation motifs from large-scale data sets. *Nat. Biotechnol* 2005;23(11):1391. [PubMed: 16273072]
21. Wang Y, Wu SL, Hancock WS. Approaches to the study of N-linked glycoproteins in human plasma using lectin affinity chromatography and nano-HPLC coupled to electrospray linear ion trap--Fourier transform mass spectrometry. *Glycobiology* 2006;16(6):514. [PubMed: 16497783]
22. Oda Y, Nagasu T, Chait BT. Enrichment analysis of phosphorylated proteins as a tool for probing the phosphoproteome. *Nat Biotechnol* 2001;19(4):379. [PubMed: 11283599]
23. Catalina MI, Koeleman CA, Deelder AM, Wuhler M. Electron transfer dissociation of N-glycopeptides: loss of the entire N-glycosylated asparagine side chain. *Rapid Commun Mass Spectrom* 2007;21(6):1053. [PubMed: 17311219]
24. Cooper CA, Gasteiger E, Packer NH. GlycoMod--a software tool for determining glycosylation compositions from mass spectrometric data. *Proteomics* 2001;1:340. [PubMed: 11680880]

25. Nielsen ML, Savitski MM, Zubarev RA. Improving protein identification using complementary fragmentation techniques in fourier transform mass spectrometry. *Mol Cell Proteomics* 2005;4(6): 835. [PubMed: 15772112]
26. Renfrow MB, Cooper HJ, Tomana M, Kulhavy R, Hiki Y, Toma K, Emmett MR, Mestecky J, Marshall AG, Novak J. Determination of aberrant O-glycosylation in the IgA1 hinge region by electron capture dissociation fourier transform cyclotron resonance mass spectrometry. *J Biol Chem* 2005;280(19): 19136. [PubMed: 15728186]
27. Hogan JM, Pitteri SJ, Chrisman PA, McLuckey SA. Complementary structural information from a tryptic N-linked glycopeptide via electron transfer ion/ion reactions and collision-induced dissociation. *J. Proteome Res* 2005;4:628. [PubMed: 15822944]
28. Mirgorodskaya E, Roepstorff P, Zubarev RA. Localization of O-glycosylation sites in peptides by electron capture dissociation in a Fourier transform mass spectrometer. *Anal Chem* 1999;71(20): 4431. [PubMed: 10546526]
29. Hakansson K, Cooper HJ, Emmett MR, Costello CE, Marshall AG, Nilsson CL. Electron capture dissociation and infrared multiphoton dissociation MS/MS of an N-glycosylated tryptic peptide to yield complementary sequence information. *Anal Chem* 2001;73(18):4530. [PubMed: 11575803]
30. Harris RJ, Leonard CK, Guzzetta AW, Spellman MW. Tissue plasminogen activator has an O-linked fucose attached to threonine-61 in the epidermal growth factor domain. *Biochemistry* 1991;5(309): 2311. [PubMed: 1900431]
31. Khidekel N, Ficarro SB, Peters EC, Hsieh-Wilson LC. Exploring the OGlcNAc proteome: direct identification of O-GlcNAc-modified proteins from the brain. *Proc Natl Acad Sci* 2004;101(36): 13132. [PubMed: 15340146]
32. Molina H, Horn DM, Tang N, Mathivanan S, Pandey A. Global proteomic profiling of phosphopeptides using electron transfer dissociation tandem mass spectrometry. *Proc Natl Acad Sci* 2007;104(7):2199. [PubMed: 17287340]
33. Hunt, DF. Comparative Analysis of Post-Translationally Modified Proteins and Peptides by Mass Spectrometry: New Technology (Electron Transfer Dissociation) and Applications in the Study of Cell Migration, the Histone Code and Cancer Vaccine Development.. Plenary Lecture L2 in 17th International Mass Spectrometry Conference Prague; Czech Republic. Aug 27–Sep 1; 2006.
34. McAlister GC, Phanstiel D, Good DM, Berggren WT, Coon JJ. Implementation of electron-transfer dissociation on a hybrid linear ion trap-orbitrap mass spectrometer. *Anal Chem* 2007;79(10):3525–34. [PubMed: 17441688]
35. Thermo User Meeting at 55th ASMS Conference on Mass Spectrometry and Allied Topics; Indianapolis, Indiana. June 3–7, 2007;
36. Yue G, Luo Q, Zhang J, Wu SL, Karger BL. Ultratrace LC/MS Proteomic Analysis Using 10- μ m-i.d. Porous Layer Open Tubular Poly(styrene-divinylbenzene) Capillary Columns. *Anal Chem* 2007;79(3):938. [PubMed: 17263319]

Figure 1. Data acquisition scheme with CID and ETD used in this work

Using an LTQXL MS with ETD, the first survey MS (scan 1) is followed by 3 consecutive ion activation steps: the CID-MS² (scan 2), the ETD-MS² (scan 3), and the charge-reduced CID-MS³ (CRCID) (scan 4). Scans 2, 3, and 4 are repeated as scans 5, 6, and 7 to fragment the second highest precursor ion generated from the first MS scan. Similarly, the third iteration cycle corresponding to scans 8, 9, 10 (not shown in the figure) is used to fragment the third most abundant precursor ion generated from the first MS scan. The total cycle (10 scans) takes approximately 3 seconds and is continuously repeated for the entire LC-MS run under data-dependent conditions with dynamic exclusion. Separately, an LTQ-FT MS is used to acquire full mass spectra in the FTICR (400 to 2000 m/z) at 100,000 resolution to determine the charge states of the same precursor ions generated with the LTQXL MS with ETD instrument.

Figure 2. ERPA (CID / ETD) analysis of a monophosphorylated peptide (2+ charge state) from the Lys-C digest of β -casein

A: Base peak ion chromatogram. **B:** CID-MS² spectrum of the m/z 1031.91 (2+) ion eluted at 35.00 min. **C:** ETD-MS² spectrum of the m/z 1031.91 (2+) ion eluted at 35.01 min. **D:** CID-MS³ scan of the m/z 1031.56 ion isolated from the ETD spectrum, as indicated by the dotted circle. The peptide sequences with the observed fragment ions are shown in the insert: phosphoserine is indicated as pS. The neutral loss of phosphate is also shown.

Figure 3. ERPA (CID / ETD) analysis of a monophosphorylated peptide (3+ charge state) from the Lys-C digest of β-casein

A: Base peak ion chromatogram. **B:** CID-MS² spectrum of the m/z 1688.76 (3+) ion eluted at 35.04 min. **C:** ETD-MS² spectrum of the m/z 1688.76 (3+) ion eluted at 35.05 min. **D:** CRCID-MS³ scan of the m/z 1031.74 ion isolated from the ETD spectrum, as indicated by the dotted circle. The peptide sequences with the observed fragment ions are shown in the insert: phosphoserine is indicated as pS. The neutral loss of phosphate is also shown.

Figure 4. ERPA (CID / ETD) analysis of the tetraphosphorylated peptide from the Lys-C digest of β -casein

A: Base peak ion chromatogram. **B:** CID-MS² spectrum of the m/z 1160.12 (3+) ion eluted at 61.03 min. **C:** ETD-MS² spectrum of the m/z 1160.12 (3+) ion eluted at 61.04 min. **D:** CRCID-MS³ scan of the m/z 1740.04 ion (from the ETD spectrum as indicated by the dotted circle). The peptide sequences with the observed fragment ions are shown in the insert: phosphoserines are indicated as pS. The neutral losses of phosphate are also shown.

Figure 5. ERPA (CID / ETD) analysis of a threonine phosphorylated peptide (pT 669) from the Lys-C digest of EGFR
A: Base peak ion chromatogram. **B:** CID-MS² spectrum of the m/z 759.23 ion (5+). **C:** ETD-MS² spectrum of the m/z 759.23 ion. **D:** CRCID-MS³ scan of the m/z 1897.23 ion (from the ETD spectrum as indicated by the dotted circle). The peptide sequences with the observed fragment ions are shown in the insert: phosphothreonine is indicated as pT. The neutral loss of phosphate is also shown.

Figure 6. ERPA (CID / ETD) analysis of a serine phosphorylated peptide (pS 1046) from the Lys-C digest of EGFR

A: Base peak ion chromatogram. **B:** CID-MS² spectrum of the m/z 1452.68 (3+) ion. **C:** ETD-MS² spectrum of the m/z 1089.78 (4+) ion. **D:** CRCID-MS³ scan of the m/z 1452.75 ion. (from the ETD spectrum as indicated by the dotted circle). The peptide sequences with the observed fragment ions are shown in the insert: phosphoserine is indicated as pS. The neutral loss of phosphate is also shown.

Figure 7. ERPA (CID / ETD) analysis of an N-linked glycosylated peptide fragment modified with a high-mannose-type glycan from the Lys-C digest of EGFR

A: Base peak ion chromatogram. **B:** CID-MS² spectrum of the m/z 1142.73 ion. **C:** ETD-MS² spectrum of the m/z 1142.73 ion. **D:** CRCID-MS³ scan of the m/z 1903.38 ion (from the ETD spectrum as indicated by the dotted circle). The peptide sequence shown in the insert of panel **C** is identified through ETD fragmentation pattern. The glycosylation site is labeled N*. In the glycan structures, (●) represents mannose and (■) represents N-acetyl glucosamine. The sequential losses of terminal mannoses from the Man8 structure resulted in Man7, Man 6, etc., as indicated in panel **B**. The sequence tags of ELDI and CTSISGD are indicated in bold in the inserts of panels **C** and **D**, respectively.

Figure 8. ERPA (CID / ETD) analysis of an O-linked glycosylated peptide fragment modified with fucose from the Lys-C digest of t-PA

A: Base peak ion chromatogram. **B:** CID-MS² spectrum of the m/z 1004.18 ion. **C:** ETD-MS² spectrum of the m/z 1004.18 ion. **D:** CRCID-MS³ scan of the m/z 1339.17 ion (from the ETD spectrum as indicated by the dotted circle). The peptide sequences with the observed fragment ions are shown in the inserts. The fucosylation site is labeled T - Fu. The loss of fucose (-Fu) from the fragmentation in CID is indicated in panel B. The sequence tags of FGE and QQALYFS are indicated in bold in the inserts of panels C and D, respectively.

Figure 9. ERPA (CID / ETD) analysis of a large unmodified 6.7 kDa peptide from the Lys-C digest of EGFR

A: Base peak ion chromatogram. **B:** CID-MS² spectrum of the m/z 1133.04 ion. **C:** ETD-MS² spectrum of the m/z 1133.04 ion. **D:** CRCID-MS³ scan of the m/z 1699.00 ion (from the ETD spectrum as indicated by the dotted circle). The peptide sequences with the observed fragment ions are shown in the inserts.

Figure 10. ERPA (CID / ETD) analysis of a large unmodified 6.7 kDa peptide from the Lys-C digest of EGFR

A: Base peak ion chromatogram. **B:** CID-MS² spectrum of the m/z 971.30 ion. **C:** ETD-MS² spectrum of the m/z 971.30 ion. **D:** ETD-MS² spectrum of the m/z 1133.04 ion. The peptide sequences with the observed fragment ions are shown in the inserts.