Journal of proteome research

Article

pubs.acs.org/jpr

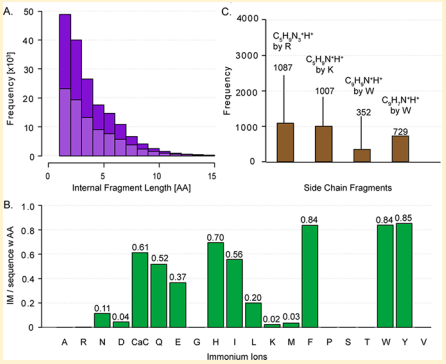
A Systematic Investigation into the Nature of Tryptic HCD Spectra

Annette Michalski, Nadin Neuhauser, Jürgen Cox, and Matthias Mann*

Department of Proteomics and Signal Transduction, Max-Planck Institute of Biochemistry, Martinsried, Germany

Supporting Information

ABSTRACT: Modern mass spectrometry-based proteomics can produce millions of peptide fragmentation spectra, which are automatically identified in databases using sequence-specific *b*- or *y*-ions. Proteomics projects have mainly been performed with low resolution collision-induced dissociation (CID) in ion traps and beam-type fragmentation on triple quadrupole and QTOF instruments. Recently, the latter has also become available with Orbitrap instrumentation as higher energy collisional dissociation (HCD), routinely providing full mass range fragmentation with high mass accuracy. To systematically study the nature of HCD spectra, we made use of a large scale data set of tryptic peptides identified with an FDR of 0.0001, from which we extract a subset of more than 16 000 that have little or no contribution from cofragmented precursors. We employed a newly developed computer-assisted "Expert System", which distills our experience and literature knowledge about fragmentation pathways. It aims to automatically annotate the peaks in high mass accuracy fragment spectra while strictly controlling the false discovery rate. Using this Expert System we determined that sequence specific regular ions covering the entire sequence were present for almost all peptides with up to 10 amino acids (median 100%). Peptides up to 20 amino acid length contained sufficient fragmentation to cover 80% of the sequence. Internal fragments are common in HCD spectra but not in high resolution CID spectra (10% vs 1%). The low mass region contains abundant immonium ions (6% of fragment ion intensity), the characteristic a_2, b_2 ion pair (72% of spectra), side chain fragments and reporter ions for peptide modifications such as tyrosine phosphorylation. *B*- and *y*-ions account for only 20% of fragment ions by number but 53% by ion intensity. Overall, 84% of the fragment ion intensity was unambiguously explainable. Thus high mass accuracy HCD and CID data are near comprehensively and automatically interpretable.



KEYWORDS: *tandem mass spectrometry, fragmentation mechanisms, shotgun proteomics, ion types, CID, HCD, Expert System, spectrum annotation*

INTRODUCTION

Rapid technological development of mass spectrometric instrumentation in conjunction with advanced bioinformatics analysis capabilities now allow relatively streamlined and in depth analysis of proteomic samples.[1−3] Modern proteomics projects routinely generate millions of fragmentation spectra, making entirely automated software tools a necessity. These include search engines that match MS/MS spectra to the most probable peptide sequence in a database, typically relying on sequence-specific backbone fragments, referred to as "regular ions" in this article, as well as associated neutral losses.[4] However, there are many other fragment ions in tandem mass spectra, and it has been argued that detailed interpretation of at least the more abundant peaks should be a requirement for confident peptide assignment.[5] Likewise, detailed understanding of the fragmentation process and discovery of potential new fragment types requires knowledge of the identity of the majority of fragmentation peaks.

While there are many different ways to fragment peptides, in proteomics collision-induced fragmentation has by far been the most frequently used technique (for a recent tutorial of peptide fragmentation and spectrum interpretation, see ref 6). While there are differences in how the ions are activated, the general ion types are the same and are summarized in Figure 1. The backbone fragments are designated as *a*, *b*, *c* for N-terminal and *x*, *y*, *z* for C-terminal types depending on the cleavage position on the peptide backbone.[7−10] A full series of either *b*- or *y*-type ions in principle allows reading out the entire amino acid sequence from a fragment ion spectrum. In collision-induced fragmentation techniques, cleavage of the peptide bond is preferred, but labile post-translational modifications such as phosphorylation or glycosylation also partially or (rarely) completely detach. While the chemistry involved in peptide fragmentation is still not completely understood, the mobile proton model is currently the most widely accepted framework to describe the dissociation process.[11,12] Moreover, different fragmentation pathways of protonated peptides have been extensively investigated and modeled with respect to both kinetic and thermodynamic aspects.[13]

In addition to the standard backbone ions, tandem mass spectra can contain many additional fragment ions.[14] Numerous studies of peptide dissociation behavior have been carried out to investigate the abundance and structure of ion types such as internal ions, immonium ions or neutral losses from these (Figure 1).[15,16] Some programs such as Protein

Received: July 30, 2012
Published: September 23, 2012

dx.doi.org/10.1021/pr3007045 | J. Proteome Res. 2012, 11, 5479−5491

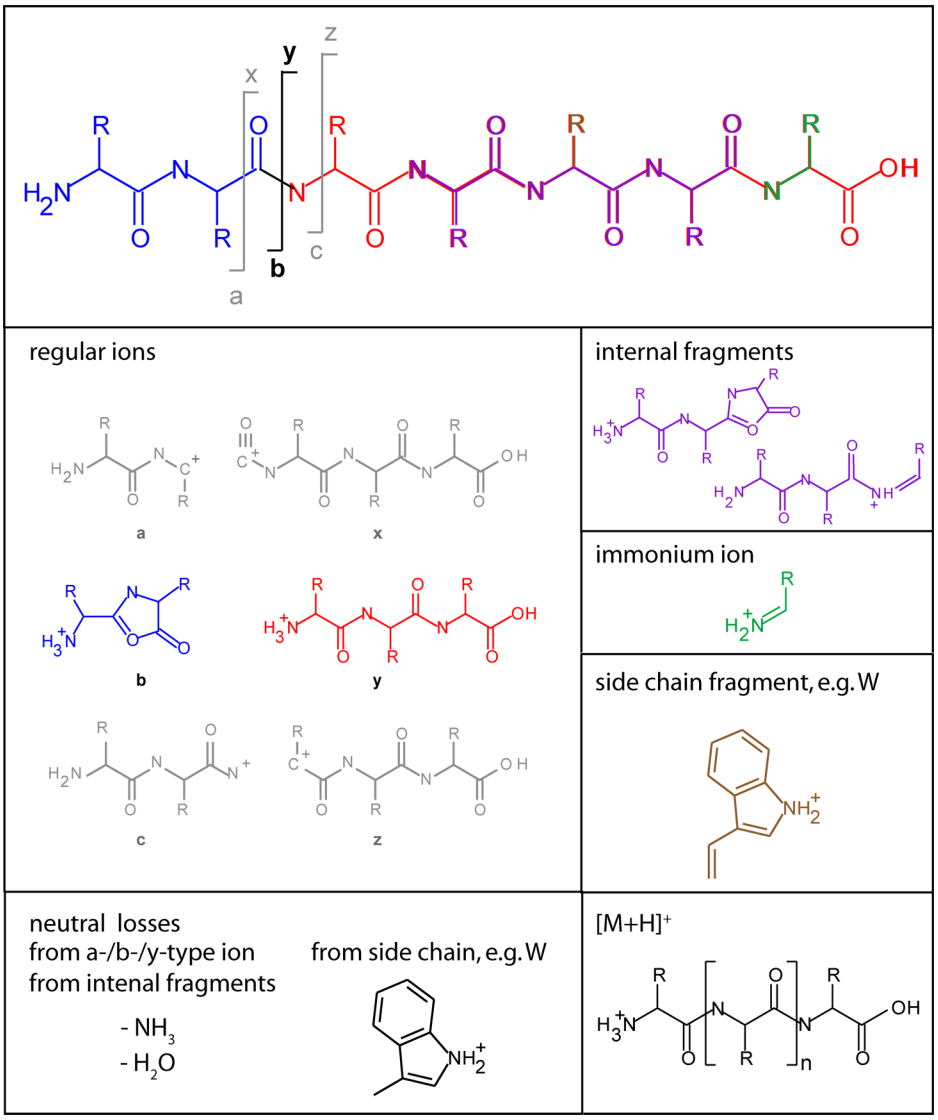


Figure 1. Cleavage sites of the peptide backbone giving rise to N-terminal *a*-, *b*- or *c*-type ions and the corresponding C-terminal *x*-, *y*- or *z*-type ions, respectively. The most prominent cleavage in CID and HCD fragmentation happens at the peptide bond. The boxes below represent the most frequent ion types of collision induced fragmentation processes; the color code provides their origin in the peptide sequence.

Prospector provide comprehensive lists of produced ion types for different fragmentation mechanisms and instrument types and even consider the latter for scoring of tandem mass spectra.[17] Furthermore, special types of ions have been characterized, for instance, b_1 ions of N-terminally acetylated ions,[18] c_1 ions in case glutamine is the second amino acid from the N-terminus,[19,20] specific side chain losses such as from oxidized methionine[21] and many more. Finally, novel fragmentation processes continue to be discussed controversially, such as the extent of scrambling of *b*-ions due to their formation of a cyclic peptide structures followed by random cleavage, which could interfere with determination of the correct amino acid sequence from the data.[22−25]

Furthermore, the observed types of fragment ions in a tandem mass spectrum depend on the instrument type. Triple quadrupole and quadrupole time-of-flight (TOF) fragmentation are beam-type dissociation processes,[26] where primary fragments retain kinetic energy and are therefore more likely to fragment again in the multiple collision conditions typical of these instruments. In 3D or 2D ion traps the excitation and activation step is only applied to the selected precursor mass. Any primary fragmentation product is off-resonance with the applied radio frequency and therefore usually remains intact. When collision-induced dissociation is performed in ion traps (often primarily associated with CID fragmentation), the low mass fragments are typically not retained, leading to a low mass cutoff in the tandem mass spectra.[27]

Higher energy collisional dissociation (HCD), first described in 2007, made beam type fragmentation available on the Orbitrap analyzer platforms.[28] Recently, HCD fragments have also been analyzed at low resolution in an ion trap[29,30] but are in general always detected in the Orbitrap analyzer at high resolution and mass accuracy. Since the introduction of the LTQ Orbitrap Velos mass spectrometer, which features improved sensitivity and HCD capability compared to its predecessors, routine acquisition of tandem mass spectra in the Orbitrap analyzer has become feasible.[31] This approach is termed "high−high" strategy because both the full scans (MS) and the fragment ion scans (MS/MS) have high resolution and high mass accuracy in comparison to previous strategies with

dx.doi.org/10.1021/pr3007045 | J. Proteome Res. 2012, 11, 5479−5491

acquisition of CID scans (MS/MS) in the ion trap ("high–low").[32] Note that high–high strategies have been the default in quadrupole-TOF instruments for many years; however, this did not necessarily imply high mass accuracy in the MS/MS mode, primarily due to issues with ion statistics. Because of the dedicated collision cell, HCD fragment ion spectra cover nearly the entire mass range and are therefore particularly suitable for observing the low mass region, which contains an a_2/b_2 pair, immonium ions, fragments resulting from the amino acid side chains as well the reporter ions[18] used for quantification in the TMT or iTRAQ methods.[33−35] Importantly, high mass accuracy of fragment ions helps to unambiguously annotate the fragment ion peaks. Especially in the low mass region, an accurate mass measurement may even uniquely determine the elemental composition of the fragment.

In contrast to ion trap CID data, high resolution HCD has been relatively little studied. Although HCD ion types are expected to recapitulate fragmentation rules known from older CID type instruments, those have not been tested on large-scale and high accuracy data. Here, we wished to take advantage of the excellent signal-to-noise, dynamic range and mass accuracy of HCD spectra on the Orbitrap analyzer to systematically investigate features of HCD spectra. This was facilitated by a rule-based "Expert System", which was developed in an iterative manner with this study and is described elsewhere.[36] This Expert System synthesizes well-established knowledge about peptide fragmentation pathways mechanisms. It is capable of annotating large-scale MS/MS data sets based on the rules chosen by the researcher. We apply the Expert System for a comprehensive statistical investigation into the nature of HCD tandem mass spectra of tryptic peptides.

EXPERIMENTAL PROCEDURES

Sample Preparation

Total cell extracts of *E. coli*, yeast and HeLa cells were separated by 1D-SDS PAGE (4–12% Novex mini-gel, Invitrogen) in three separate lanes. Colloidal Coomassie (Invitrogen) was used for staining of the proteins before each lane was cut into 8 or 10 slices. All gel slices were subjected to reduction of the proteins with 10 mM DTT in 50 mM ammonium bicarbonate and subsequently alkylated with 55 mM IAA in 50 mM ammonium bicarbonate. In-gel digestion with 12.5 ng/μL trypsin (Promega) in 50 mM ammonium bicarbonate was carried out at 37 °C for 12 h followed by extraction of the tryptic peptides with 3% TFA in 30% ACN.[37] Peptides were loaded on C_{18} StageTips[38] before eluting them with 80% ACN in 0.5% acetic acid prior to analysis.

HeLa cell lysate was digested according to the filter-aided sample preparation (FASP) method.[39] Briefly, the lysate was solubilized in SDS-containing buffer and loaded onto Microcon YM-30 devices (Millipore, Billerica, MA, USA) to remove SDS and exchange it by urea. The protein mixture was alkylated with 50 mM iodoacetamide before urea was replaced with 20 mM ammonium bicarbonate. The proteins were digested overnight at 37 °C with trypsin (Promega) (1 μg of trypsin/100 μg of protein). Peptides were collected from the filter after centrifugation. For enrichment of phosphorylated peptides, the mixture was acidified with trifluoroacetic acid to pH 2.7 and ACN was added to a final concentration of 30%. Incubation with TiO_2 beads[40] (MZ Analysentechnik, Germany) prepared in 30 mg/mL solution of dihydrobenzoic acid (Sigma) was carried out for 30 min, before the beads were washed with 30% ACN and 3% TFA (twice) followed by two washes with 75% ACN and 0.3% TFA. The phosphopeptides were eluted with buffer containing 15% ammonium hydroxide and 40% ACN. Finally, the eluted phosphopeptides were loaded on C_{18} StageTips before they were eluted with 60% ACN in 0.5% acetic acid prior to analysis.

LC–MS/MS Analysis

For the analysis of proteome samples, the peptide mixture was separated on a C18-reversed phase column (15 cm, 75 μm ID, packed in-house with ReproSil-Pur C_{18}-AQ 3 μm resin, Dr. Maisch GmbH). An Easy-nLC (Thermo Scientific, Odense) with IntelliFlow system was used for sample loading and operated at a constant flow rate of 250 nL/min during the 110 min linear gradient of 8–60% buffer B (80% ACN and 0.5% acetic acid). A nanoelectrospray ion source (Thermo Scientific, Odense) was used for online coupling to the LTQ Orbitrap Velos mass spectrometer.[31] Mass spectra were measured in positive ion mode applying a data-dependent "top 10" method for the acquisition of a survey scan followed by MS/MS spectra of the 10 most abundant precursors. High resolution data was acquired in the Orbitrap analyzer with a resolution of 30 000 (m/z 400) for MS and 7500 (m/z 400) for MS/MS scans. For peptide fragmentation higher energy collisional dissociation (HCD) was used applying a normalized collision energy of 40 eV. The minimal signal threshold required was set to 5000. The target value in the Orbitrap analysis was 1×10^6 for the MS scans and 5×10^4 for the MS/MS scans with 2 Th isolation window and the first mass was set to 80 Th for HCD spectra. Fragmented precursors were dynamically excluded from targeting for 90 s. High resolution CID data was acquired on an Orbitrap Elite (Thermo Scientific) the same parameters; however, the resolution for MS scans was 120 000 (m/z 400) and for MS/MS scans 15 000 (m/z 400); the normalized collision energy was set to 35 eV.

For the phosphoproteome data, the enriched peptide mixtures were separated on a C_{18}-reversed phase column (20 cm, 75 μm ID, packed in-house with ReproSil-Pur C_{18}-AQ 1.8 μm resin, Dr. Maisch GmbH) applying a 90 min linear gradient of 5–30% buffer B (80% ACN and 0.1% formic acid) and analyzed on the Orbitrap Elite instrument[41] that was online-coupled to an Easy-nLC 1000 (Thermo Scientific, Odense). The MS data was acquired with resolution of 120 000 (m/z 400) and target value of 1×10^6 and MS/MS (HCD fragmentation) with resolution of 15 000 (m/z 400) and target value of 5×10^4 in a data-dependent "top 15" method with a dynamic exclusion of 30 s. The signal threshold was set to 5000 for an isolation window of 2 Th and the first mass of HCD spectra to 80 Th. The collision energy was set to 35 eV.

Data Analysis

All spectra were processed with MaxQuant[42] version 1.2.5.2 using the Andromeda search engine[43] to search the MS/MS spectra with trypsin specificity against the IPI human database (version 3.68, 87 061 entries) combined with 262 common contaminants. We allow for up to 2 missed cleavages and N-terminal acetylation and methionine oxidation were selected as variable, carbamidomethylation of cysteine was selected as fixed modification. For MS spectra an initial mass accuracy of 7 ppm was allowed, and the MS/MS tolerance was set to 20 ppm for fragment detection in the Orbitrap analyzer for high resolution CID and HCD. A sliding mass window was applied to filter the MS/MS spectra for the 10 most abundant peaks in 100 Th. For identification, the peptide FDR was set to 0.0001. (The protein

dx.doi.org/10.1021/pr3007045 | J. Proteome Res. 2012, 11, 5479–5491

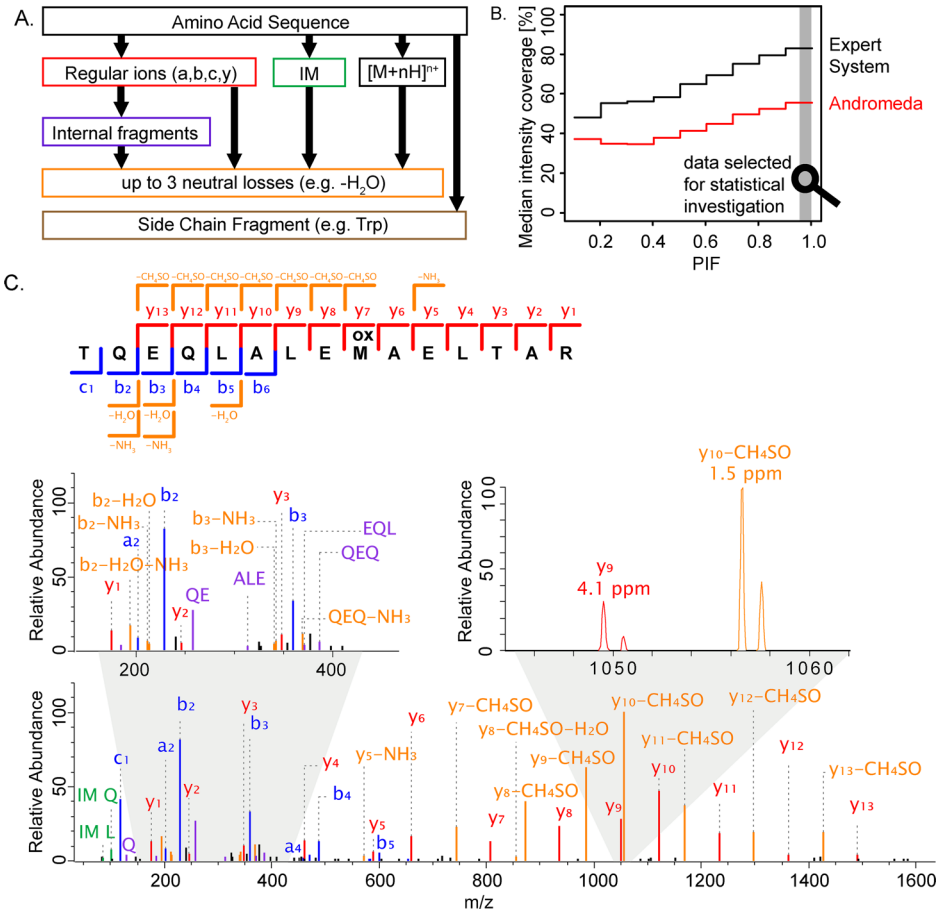


Figure 2. Peak annotation by the Expert System. (A) Ranking of the six major ion types: intact precursor mass $[M+nH]^{n+}$, regular ions, immonium ions (IM), internal fragments, neutral losses and side chain fragments that are considered for peak annotation by the Expert System. (B) Average intensity coverage of the total intensity of >100 000 MS/MS spectra by standard search engine annotation (Andromeda, red line) and by the Expert System (black line) vs the precursor intensity fraction "PIF" provides a measure for the purity of precursor isolation. The high quality data set (16 000 spectra) that was selected for statistical investigation is highlighted in gray. (C) Typical MS/MS spectrum with PIF 0.99 annotated by the Expert System reaching an intensity coverage of 87%. A zoom window displays the high mass accuracy of two fragment ion peaks.

FDR remained at the standard setting of 0.01, but protein identifications were not directly used in this paper.) The shortest peptide length was set to 6 amino acids, and the Max Quant feature to treat the isobaric amino acids leucine and isoleucine as indistinguishable for improved statics was disabled. This setting ensures that either amino acid matches the fragmentation spectrum as HCD in our setup cannot distinguish them; however, side chain losses can then be assigned correctly because the isoleucine/leucine ambiguity is absent after database search. MaxQuant and Andromeda data processing provides access to the peptide sequences that were identified from the MS/MS spectra. Detailed annotation of the MS/MS spectra was then carried out using the Expert System.[36] Results were further analyzed within the R scripting and statistical environment.[44] Raw mass spectrometric data are available at Tranche (www.proteomecommons.org) using the following hash code:

pI2oaLaSi7gPxUWNbesdXCgR17sWvMY6qVkHL+MtWA0Q5sqn/UxZVSjk3KpFTfrmDYpf3y/Iv6WfaAi6-HaILdZL0YocAAAAAAAAT7Q==

RESULTS AND DISCUSSION

Generation of a High Quality Data Set

To produce a diverse set of fragmentation spectra of tryptic peptides, we separated proteomes of *E. coli*, yeast and HeLa cells by one-dimensional gel electrophoresis, excised eight slices and in-gel digested them (Experimental Procedures). This generated a total of 24 complex peptide mixtures, which were analyzed using a "high−high" strategy on a linear ion trap−Orbitrap instrument (LTQ Orbitrap Velos) using HCD as the fragmentation method. For a smaller number of fractions, we also employed CID fragmentation followed by high resolution detection of fragments in the Orbitrap analyzer (Experimental Procedures).

We wished to work with an extremely high quality set of fragmentation spectra in order to enable us to unambiguously attribute the observed fragments to the precursors. Therefore, we set the false discovery rate (FDR) for peptide identification by MaxQuant using the Andromeda search engine[42,43] to 0.0001 rather than the customary 0.01. From our data set, we obtained more than 100 000 MS/MS spectra that were identified with this very stringent criterion. We and others have recently introduced the notion of the precursor intensity fraction (PIF),[45] chimeric or mixture MS/MS spectra,[46,47] which refers to the fact that precursor ions are frequently

dx.doi.org/10.1021/pr3007045 | J. Proteome Res. 2012, 11, 5479−5491

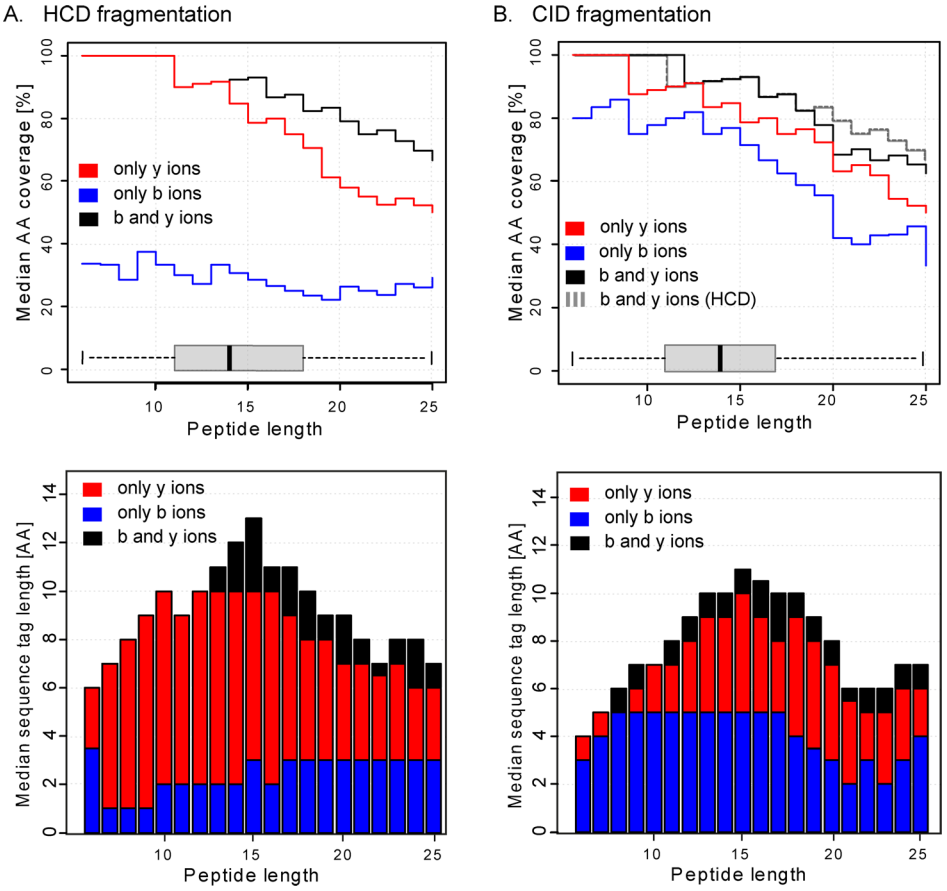


Figure 3. Sequence information content in HCD and CID spectra (A) Median coverage of amino acids by *y*-type ions (red), *b*-type ions (blue) and both together (black) in the upper panel. The boxplot displays the distribution of the peptide length within the data set (>16 000 spectra). The lower panel shows the median length of the longest sequence tag based on *y*-type ions (red), *b*-type ions (blue) and both together (black). (B) Same as (A) for a data set of 3290 high resolution CID spectra. The dashed gray line in the upper panel repeats the median amino acid coverage in HCD from panel (A) for comparison.

cofragmented unintentionally in the analysis of complex peptide mixtures. For our purposes we needed to minimize the occurrence of coeluting precursor ions in the isolation window so that they could not "contaminate" the MS/MS spectra with unassignable peaks. This was achieved by only retaining spectra with a PIF greater than 0.95. If there was more than one spectrum for a particular sequence, the one with the highest PIF was kept. Furthermore, we required the peptide length to be smaller than 26 amino acids and the charge state to be 2+, 3+ or 4+. These filters reduced the number of MS/MS spectra to about 16 000, which were nearly free of any contaminating peaks and which represented a broad sampling of typical tryptic peptides.

Computer-Assisted Annotation by the Expert System

We recently developed a computer Expert System,[36] which is now integrated into the Viewer component of the MaxQuant software environment. Briefly, the Expert System features a knowledgebase that was supplied with peptide fragmentation mechanisms described in the literature (see Introduction) and with knowledge gained from manual evaluation of small and large-scale HCD data sets. These facts are implemented in a rule-engine that assigns annotations to the peaks in the MS/MS spectra. In order to avoid incorrect assignments, the Expert System follows strict dependencies among its rules. We derived a rigorous FDR for peak annotation, which made it possible to derive a minimal yet relative comprehensive set of rules.[36]

Some MS/MS peaks can have an elemental composition that corresponds to more than one ion type, and we have developed a strict ranking of the possible annotations to address this particular issue (Figure 2A). On the basis of the identified peptide sequence, regular ions that result from cleavage of peptide bonds (*b*- and *y*-type ions), *a*-type ions that derive from the corresponding *b*-type ion by losing CO and *c*-type ions that occur in specific cases,[20] are assigned the highest priority for annotation. The chemical structures of regular ions and immonium ions are different, and as a consequence, there is no possible overlap between them. Therefore the order of assignment is of no consequence, and they are treated with the same priority. The second step covers annotations of neutral losses and internal fragment ions; these types derive from regular backbone ions. Importantly, neutral losses are specific to N- or C-termini of fragments or to a single or several amino acids. These are required to be contained in the peptide sequence to allow an annotation. Internal fragment ions originate from regular ions that have undergone a second cleavage of the peptide backbone. The side chains of the amino acids tryptophan (W), arginine (R) and lysine (K) are prone to produce specific fragment ions that can carry a proton because of the heteroatom in their chemical structure. Their mass is sufficiently large (>100 Da) that they are recorded in HCD fragment ion spectra as side chain fragment ions. They are assigned a low priority because they are independent of any

dx.doi.org/10.1021/pr3007045 | J. Proteome Res. 2012, 11, 5479−5491

other ion type. Finally, incomplete fragmentation results in protonated precursor ions remaining in the MS/MS spectra, which are annotated as $[M+nH]^{n+}$.

The Expert System greatly improves on the number and intensity coverage of assigned peaks in the fragmentation spectra calculated by adding the signal for the 10 largest peaks per sliding 100 Th window. Standard annotation by the Andromeda search engine results in an intensity coverage of up to 58% for pure spectra (PIF > 0.95; highlighted in gray in Figure 2B). Including the additional ion types that are covered by the Expert System increased the intensity coverage to 84%. With the Expert System in hand, we next annotated all of the about 100 000 high scoring fragment spectra in the initial set. This showed that even for impure MS/MS spectra (PIF less than 0.5), the intensity coverage of assigned peaks in MS/MS spectra was still above 50%. A typical MS/MS spectrum with a high PIF precursor that was comprehensively annotated by the Expert System is displayed in Figure 2C. Virtually all major peaks are correctly annotated and fragment intensity coverage reaches 87%. The figure also illustrates the mass accuracy typically achieved in our experiment. Even though the lock mass feature during data acquisition was enabled,[48] data analyzed with Max Quant is routinely independently recalibrated.[49]

Sequence Related Information Content of HCD Spectra

The most important information imbedded in tandem mass spectra relates directly to the amino acid sequence of the peptide. Cleavage of all peptide bonds, resulting in *b*- and *y*-type ion series, would in principle allow read out of the peptide sequence from the MS/MS spectrum in two directions starting from the N- or the C-terminus, respectively. Moreover, combining the *b*- and *y*-ion series highlights complementary *b*- and *y*-type ions pairs that together match the mass of the unfragmented peptide. Complementary pairs provide strong constraints for correct peptide identification and can be used in scoring algorithms even of multiplexed spectra.[50]

In our large collection of HCD data, we found nearly universal evidence for such pairs. Typical spectra have the prominent a_2/b_2 pair (observed in 72% of the peptide sequences) followed by at least a few more *b*-ions. *Y*-ion series were very abundant in our spectra, especially in the middle mass range (450−800 Da). For peptides that were not too long (<20 amino acids), the low mass *b*-ion series almost always had a corresponding, complementary *y*-ion series of high intensity. These trends are well-known from triple quadrupole and quadrupole time-of-flight spectra.

We next evaluated all 16 000 HCD spectra in the collection (Figure 3A). Remarkably, for peptides up to 12 amino acids the *y*-ion series alone provided for at least half of the sequences complete sequence coverage (median 100%), indicating that complete sequencing of such peptides even in routinely acquired large-scale data sets is in principle possible. This includes the order of the two first amino acids, which is normally inaccessible because of the missing y_{n-1} and b_1 ions (see below). With increasing peptide length, the amino acid coverage slowly drops to a median of 50% at a peptide length of 25 amino acids, which was the upper limit in our collection (Figure 3A). The *b*-ion series, in contrast, remains at a constant level, providing about 30% amino acid coverage independent of the peptide length. Taking both ion series together yields median amino acid coverage of 80% percent even for a peptide length of 20 AA.

Besides the percentage of the sequence that is covered by backbone fragmentation, another important parameter is the number of amino acids that can be read out from the MS/MS spectrum as an uninterrupted part of the sequence, i.e., the maximum sequence tags length.[51] A sequence tag of six amino acids is generally unique in the human genome even without added peptide mass information.[6,52] In addition to peptide identification, such stretches are useful for partial de novo sequencing or homology searching. The lower panel in Figure 3A depicts the median sequence tag lengths based on the two different ion series of the identified sequence. Peptides up to 10 amino acids contain a complete *y*-ion based sequence tag, but above this length, the $y_n - 1$ ion is often of too low intensity to be recorded. Even small peptides contain short sequence tags of three amino acids, which are sufficient for peptide identification. When combined with the *y*-ion series, the *b*-ion series helps to increase the sequence tag length for peptides larger than 14 amino acids. The largest median sequence tag length is about 12 amino acids, and it starts to drop from a peptide length of 16 amino acids.

We next compared the sequence related information content of HCD with that of high resolution CID spectra both acquired in the Orbitrap analyzer. A prominent difference is the much larger contribution of the *b*-ion series in CID spectra (Figure 3B). This is due to the higher stability of *b*-ions in ion trap fragmentation processes. Although lower than the *y*-ion series, the *b*-ion series continued to provide a median of more than 50% sequence coverage up to a peptide length of 19 amino acids. Nevertheless, the combined contribution from *y*-ions and *b*-ions was slightly higher for HCD than for CID, which partly reflects the more extensive fragmentation in beam type instruments and the fact that ion series in CID spectra are limited by the low mass cutoff that is inherent to ion trap fragmentation. As a consequence, maximum sequence tag length was likewise higher in HCD spectra.

We have previously investigated maximum sequence tag lengths in low resolution CID spectra. In more than 85% of the identified spectra sequence tags of at least three amino acids and only in half of the spectra sequence tags of six or more amino acids were detected.[52] Despite the potential for overcounting due to the lower mass accuracy, these sequence tags were substantially shorter than tags from either high resolution HCD or high resolution CID.

Neutral Loss Fragments in HCD

During collision-induced dissociation processes, peptides can follow numerous fragmentation pathways and consequently give rise to various ion types beyond those produced by the typical peptide backbone cleavage. A large class of such ions are those involving neutral losses from different fragment species. These occur from nearly all ion types, however, the chemical structures of the diverse ion types as well as the amino acid side chains allow specific neutral losses (Figure 2A). In some cases, these can result either from the peptide terminus or from one of the side chains of the amino acids, and localization of the origin is not straightforward. However, such losses can still be unambiguously assigned to the fragment ion. We carried out a systematic study considering 45 possible chemical compositions that could formally occur as neutral losses from amino acid residues. We then used our large scale data set to determine the primary neutral losses for all of the fragments in the collection that contained the amino acid in question. The median absolute mass accuracy of all neutral losses is 2.7 ppm with 97.5% of the

dx.doi.org/10.1021/pr3007045 | J. Proteome Res. 2012, 11, 5479−5491

peaks within 5 ppm, therefore they can unambiguously be connected to their precursor fragments. Only first neutral losses which happened in at least 5% of the cases were considered and encoded in the Expert System.[36] Table 1 summarizes the observed frequencies of the primary neutral losses that occur in different combinations in more than 270 000 fragments. While *b*-type ions frequently lose a water molecule, the chemical structure of *y*-type ions allows both water and ammonia losses. These are by far the most frequent neutral losses. Furthermore, acidic amino acids as well as serine and threonine are likely to lose water. However, it was possible in about 48% of the cases to assign the neutral loss to either a specific amino acid or the C-terminus of the fragment, because there was only one possible origin for the water loss. At least 33% of the spectra from sequences that contain glutamic acid, serine or threonine exhibit water losses from those amino acids. This is the case in only 29% of spectra where the water loss can be confidently assigned to aspartic acid. The rate of ammonia losses is comparable to water losses and this also holds true for confidently assignable losses from glutamine (29%), asparagine (30%) and arginine (21%). Further frequently observed neutral losses that are specific to certain amino acids include CH_3NO from glutamine (20%) and from asparagines (29%) or C_2H_4O from threonine (26%). While other neutral losses may exist, our large data set suggests that they are unlikely to occur at substantial frequencies in HCD spectra.

Table 1. Neutral Losses Considered by the Expert System and Fraction of Spectra That Contain the Loss from the Corresponding Amino Acid[a]

NH_3	45% (N-term)	30% (N)	29% (Q)	21% (R)	
H_2O	48% (C-term)	37% (S)	44% (T)	21% (D)	33% (E)
CO	84% (internal)[b]				
CO_2	5% (D)				
CH_2N_2	8% (R)				
CH_3NO	29% (N)	20% (Q)			
CH_4O	5% (S)				
CH_4SO	89% (Mox)				
C_2H_4	5% (I)				
C_2H_5NO	9% (N)	6% (Q)			
C_2H_4O	26% (T)				
$C_2H_4O_2$	6% (D)	6% (E)			
C_3H_6	6% (L)				
$C_3H_9N_3$	6% (R)				
C_3H_6SO	6% (Mox)				
C_3H_8SO	12% (Mox)				
C_4H_8	5% (L)				
C_8H_7N	6% (W)				
C_9H_9N	12% (W)				

[a]Only examples allowing unambiguous localization of the origin of the neutral loss were considered. [b]This ion is formally equivalent to an *a*-type internal fragment.

Internal Fragments

Internal fragments in the MS/MS spectra are characteristic of beam-type fragmentation because these result from ions undergoing a second cleavage resulting in a C-terminal carboxyl-group and an N-terminal oxazolone structure.[13,53] In our large-scale data set, the length of internal fragments varied between two and more than 10 amino acids, depending on peptide length. The majority of internal fragments, however, are shorter than five amino acids. Proline is most often the first amino acid of an internal fragment since N-terminal cleavage is very pronounced at this amino acid; this is called the proline effect.[54] However, we found that on the basis of peak presence, rather than peak intensity, proline initiated internal sequences were more than four times as common as those of the median of other amino acids (Supporting Information, Figure S1A). For cleavage at the C-terminal amino acid of an internal fragment there is a slight preference for aspartic acid, glutamic acid, glutamine, tryptophan and histidine (Supporting Information, Figure S1B). Proline is the least common amino acid at the C-terminus of internal ions.

Low Mass Region

HCD fragmentation takes place in a dedicated collision cell and is not subject to the low mass cutoff of ion trap CID spectra, therefore in principle allowing observation of the entire mass range. In practice, HCD spectra are normally acquired from m/z 100, but for a more extensive investigation of the low mass region we acquired data in our study from m/z 80, which was the lowest practical m/z without reducing the scan speed of the instrument. Therefore our data set does not contain immonium ions with an m/z lower than 80 Th.

Figure 4B displays the frequency of immonium ions in the MS/MS spectra. The most prominent immonium ions originate from phenylalanine (F), tryptophan (W) and tyrosine (Y) and can be observed in at least 84% of all peptide sequences containing the respective amino acid. This is due to their chemical structure containing both a heteroatom and an aromatic system that are prone to stabilize a positive charge and for the same reason, the immonium ion of histidine (H) is often present (70%). Carbamidomethylated cysteine (caC), glutamine (Q) and glutamic acid (E) immonium ions (61, 52, and 37%, respectively) can also be found relatively abundantly in the spectra. Aspartic acid (D) and asparagine (N) produce a significantly lower rate of immonium ions. Interestingly, immonium ions of isoleucine (I) and leucine (L) are detected in the MS/MS spectra with different frequencies. Immonium ions of glycine (G), alanine (A), serine (S), proline (P), valine (V) and threonine (T) are not observed in our, data as their m/z is below 80 Th. Arginine (R) and lysine (K) represent special cases due to their position at the N-termini of tryptic peptides. A very frequently observed ion is the immonium ion of lysine with an ammonia loss (IM K − NH_3). In fact, this ammonia loss often occurs even without immonium ion, and this was therefore implemented as an exception to the strict requirement for a detected precursor fragment in the Expert System. Immonium ions can be used to support the peptide sequence assignment. In special cases, such as phosphotyrosine (pY), immonium ions can be used as reporter ions to verify the existence and the nature a phosphorylation site (see below).[55,56]

Another fragment ion type in the low mass region are fragment ions that result from cleavage of amino acid side chains in which the molecular structure can stabilize a proton. This is the case for some of the amino acids that contain a nitrogen atom, such as arginine, lysine and tryptophan. The chemical compositions of the side chain fragments and their frequency of occurrence are displayed in Figure 5C. Note that these side chain fragments are different from the *v*-, *w*- and *d*-type ions from high energy CID dissociation carried out on TOF/TOF instruments.[57,58] In addition to the general ion types, certain amino acid side chains follow different

dx.doi.org/10.1021/pr3007045 | J. Proteome Res. 2012, 11, 5479−5491

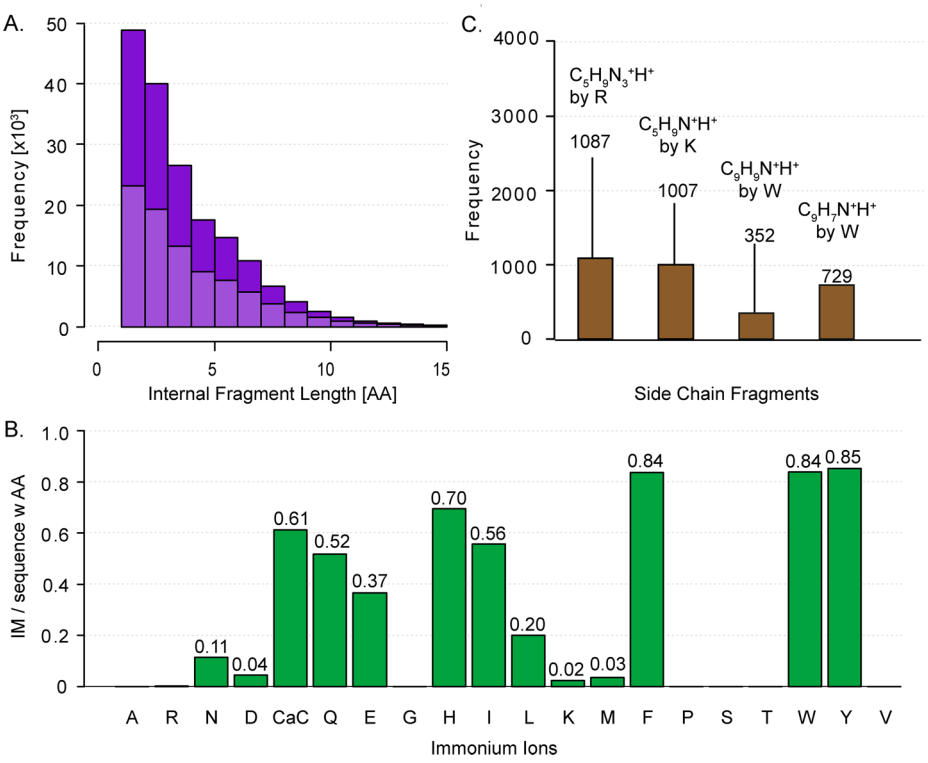


Figure 4. Statistics on the low mass region fragment ions from 16 000 MS/MS spectra. (A) Histogram of the length of all internal fragment ions in purple; the fraction of internal fragment ions starting with proline is highlighted by light color. (B) Percentage of immonium ion (IM) occurrence if the amino acid corresponding amino acid was at least once contained in the peptide sequence. Immonium ions of Alanine, Glycine, Proline, Serine and Threonine were not considered, because their m/z value is lower than 80 Th. (C) Bar plot displaying the five most abundant side chain fragment ions that are automatically assigned by the Expert System with their total number of occurrences within the data set and their chemical structures.

fragmentation pathways resulting in unusual ion types. Lehmann and co-workers observed c_1 ions resulting from the N-terminal amino acid of the peptide, if the second amino acid is glutamine (Q).[19,20,59] Along these lines, we investigated asparagine and carbamidomethylated cysteine and found the same behavior for these two candidates. Furthermore, b_1 ions are usually not observed because of their chemical instability. However, we did observe b_1 ions from acetylation of methionine, serine or alanine at the protein N-terminus. This is thought to be due to stabilization of this fragment by the acetyl group.[18,60] Besides the qualitative information contained in the variety of ion types of natural peptides, the low mass region in HCD fragmentation also gives access to the reporter ions for the TMT[61] and iTRAQ[33] quantification methods.[34,35] The reporter ions of TMT and iTRAQ are at m/z 126.1277, 127.1248, 128.1344, 129.1315, 130.1411, 131.1382 and m/z 114.1112, 115.1146, 116.1116, 117.1150, respectively. Investigation of our large-scale and high accuracy data set revealed no interfering ions of the same m/z. Therefore problems in quantification by these methods are confined to cofragmentation of other labeled peptides rather than other ion types that have the same mass as these reporter ions.

Global Composition of Tryptic HCD Spectra

The different ion types covered by the Expert System, such as regular ions, neutral losses, internal fragments, immonium ions, side chain fragments and the intact peptide mass $[M+nH]^{n+}$ by their nature occur in MS/MS spectra with different frequencies (Figure 5A). However, for high confidence of peptide identification it is predominantly the highly abundant MS/MS peaks that are of interest. Figure 5B displays the contribution of each of the ion types to the overall intensity coverage: Regular ions (a, b, c and y) account for 54% of total MS/MS spectra intensity and peaks that result from neutral losses for a further 15%. Immonium ions can originate from several amino acids, and these signals are added as singly charged peaks at defined masses in the low mass region. Together, their mean contribution to the total intensity coverage is 6%. Unlike immonium ions, internal fragments are spread over the low to middle mass range of the MS/MS spectrum because they can be generated by any two cleavages of the peptide backbone, and hence they are not as obvious in tandem mass spectra. As described above, in HCD internal fragment ions are frequently observed. However, their abundance is lower than that of immonium ions or y-ions, and together they contribute 10% to the total fragment intensity. The protonated unfragmented peptide precursor only has an average intensity coverage less than 1% in our data set. Side chain fragments account for only 0.1% of total peaks and an intensity coverage of less than 0.1% and are therefore not displayed in the pie chart. The fraction of unannotated peaks accounts for 44% on the basis of the 10 largest peaks per hundred Th but only for 15% with regard to total intensity coverage. This provides evidence that remaining peaks are mainly of low abundance. Note that those, beyond potentially being noise peaks, could also result from combinations of multiple neutral losses without precursor fragments or similar, which were not allowed by the Expert System to maintain a strict false positive rate. Furthermore, cofragmentation of other precursors still occurs in our data set to some degree. Together, our data suggests that nearly all fragment peaks in HCD are

dx.doi.org/10.1021/pr3007045 | J. Proteome Res. 2012, 11, 5479−5491

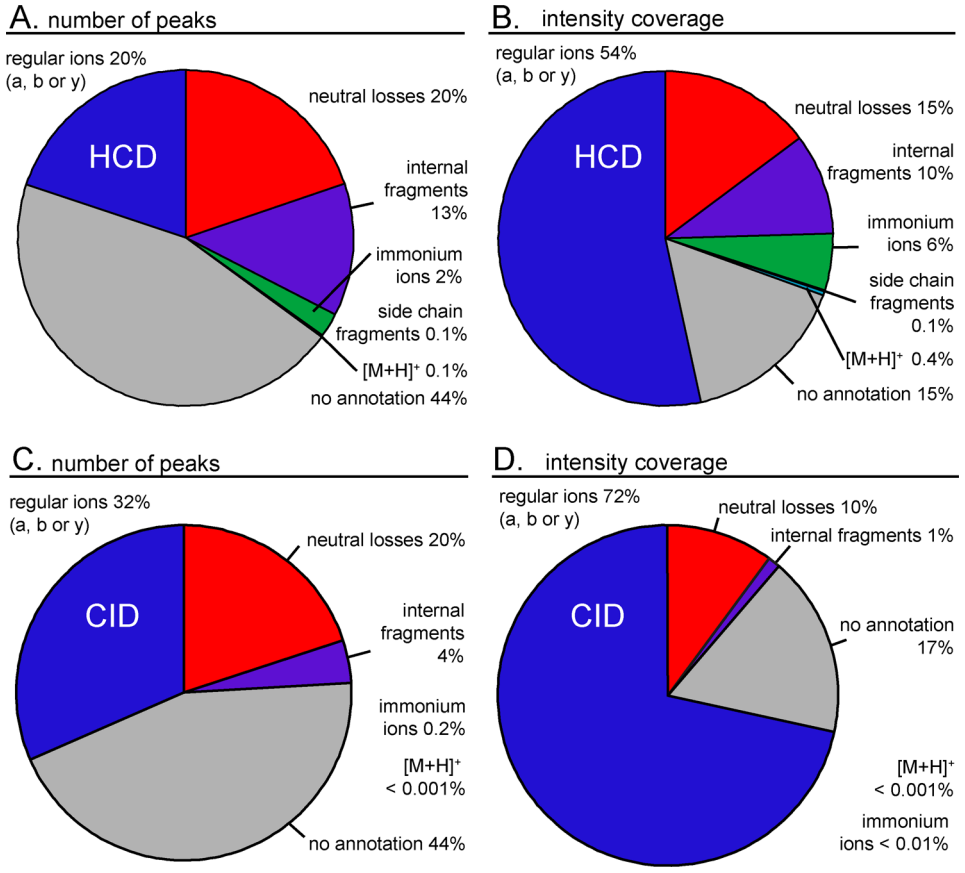


Figure 5. Intensity distribution of different ion types. (A) Average proportions of the six major ion types in HCD spectra by peak count based on a sliding mass window filtering for the 10 most abundant peaks per 100 Da; >16 000 tandem mass spectra. (B) Same as (A) but referring to the intensity coverage of the MS/MS spectrum. (C and D) Same as (A) and (B) for >3200 high resolution CID tandem mass spectra for comparison to the HCD ion type distribution.

explainable on the basis of current understanding of fragmentation pathways.

We next repeated the same analysis as above for high resolution CID spectra, which resulted in quite similar findings for the number of peaks. As expected, the number of immonium ions and internal fragments was drastically reduced since ion trap fragmentation is not capable of retaining the low mass region of the tandem mass spectra and their formation requires double cleavage. Together with the higher preponderance of high mass *b*-ions, this has the effect of increasing the fraction of regular ions to 32% as compared to the 20% of HCD fragmentation. On the basis of intensity coverage, this effect is less pronounced (72% for CID compared to 54% for HCD). Interestingly, using the Expert System the fraction of unannotated peaks by intensity is very similar between CID (17%) and HCD (15%).

Characteristics of Phosphorylated Peptides

Protein phosphorylation is among the most important and best studied post-translational modifications and is almost always located at serine, threonine or tyrosine in mammalian cells. Because of its chemical nature, the phosphogroup easily detaches from serine and threonine during collision induced fragmentation processes resulting in very characteristic and abundant neutral loss peaks such as HPO_3 and H_3PO_4. Furthermore, as already mentioned above, phosphotyrosine leads to a unique and characteristic immonium ion with m/z 216.0426.

We investigated large scale phosphorylation data with the Expert System, incorporating rules for the above-mentioned phosphospecific fragment ions. We found that the occurrence of both neutral losses from phosphorylated serine is about four times as high (65% for HPO_3 and 49% for H_3PO_4) as from threonine (18 and 12%, respectively). Table 2 summarizes the frequencies of these neutral losses. Their absolute number reveals an average of three H_3PO_4 losses and two HPO_3 losses per spectrum.

Table 2. Fraction of 1157 Spectra of Modified Sequences (Phospho STY) Containing Neutral Losses, Reporter Ions from Phosphorylated Serine (S) and Threonine (T) or the Characteristic *X*-Ion at Least Once[a]

	$-HPO_3$	$-H_3PO_4$	pS	pT	x_n (S,T)
S (1094)	65% (713)	49% (540)	29		279
T (585)	18% (103)	12% (68)		3	

[a]The first column lists the total number of sequences that contain the amino acid S or T at least once.

Finally, we investigated the frequency of x_n ions pinpointing the localization of a serine or threonine phosphor site in the peptide sequence very recently described by Kelstrup et al.[62] Our data set consisting of 1157 spectra of phosphorylated peptide sequences contains this characteristic x_n ion in 279 of the fragmentation patterns (24%).

dx.doi.org/10.1021/pr3007045 | *J. Proteome Res.* 2012, 11, 5479−5491

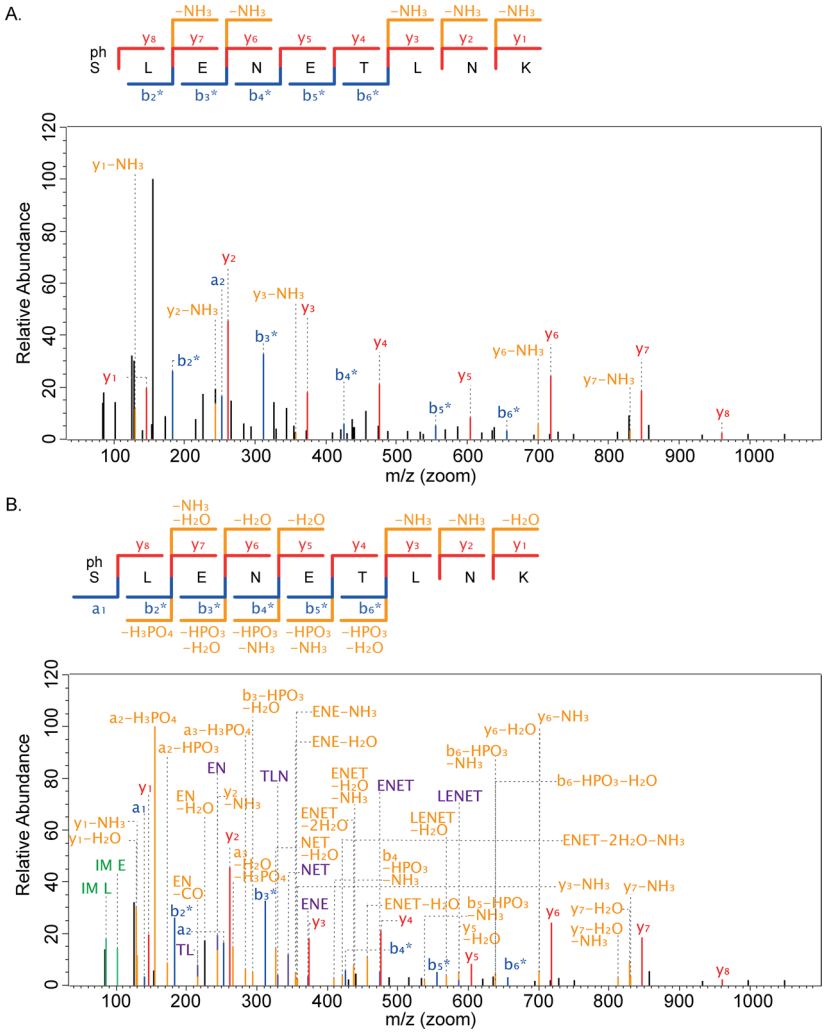


Figure 6. Annotated spectrum of phosphorylated peptide fragmented with HCD. (A) The phosphorylated peptide phSLENETLNK was identified and annotated by the Andromeda search engine assigning regular ions and single neutral losses. (B) The Expert System was modified for phosphorylated peptides to enable comprehensive annotation: Several additional neutral losses, internal fragments and immonium ions increase the intensity coverage to 82%.

CONCLUSION AND OUTLOOK

In 2007, beam-type fragmentation was introduced on Orbitrap instrumentation. This HCD mode of fragmentation has become especially popular since some limitations of ion source brightness and ion extraction from the collision cell were removed.[31] In our group, for instance, both proteome and PTM-based investigations are routinely done with HCD rather than low or high resolution CID. This was one reason why it was important to investigate the ion types that are produced by HCD. However, even though the general dissociation mechanisms operative in CID have been studied for decades,[63,64] large data sets with very high quality thresholds have previously not been studied. This was made possible here by very stringent filtering of peptide fragment spectra on the basis of identification score as well as near absence of cofragmenting peaks. Most importantly, we developed and made use of an Expert System, which annotated peptide peaks with high comprehensiveness but low false positive rates.

Our investigation of HCD yielded a broad and quantitative overview of the ion types produced. It turns out that HCD spectra are somewhat more complex than CID spectra but that the peaks are assignable to the same degree. The low mass region is particularly straightforward to interpret given the very high resolution of the Orbitrap analyzer in this region, coupled to the high mass accuracy, which generally allows determination of the chemical composition of these fragments. The information content of HCD spectra is mostly related to very extensive series of *y*-ions, supplemented by relatively short series of low mass *b*-ions. This is in contrast to ion trap CID spectra, in which the high mass *b*-ions are also very prominent. Nevertheless, the coverage of peptide sequence overall and in particular with continuous ion series is somewhat higher in HCD than it is in CID. Remarkably, for tryptic peptides up to 15 amino acids, the fragment contents is almost complete, meaning that there is sufficient information in principle for de novo sequencing or at least very long sequence tags.

Our quantification of the overall contribution of different ion types to the entire MS/MS spectrum revealed that only a relatively small proportion remains unassigned by the rules that we have implemented into the Expert System. This proportion would further shrink if noise and remaining cofragmentation was further reduced and if the rules of the Expert System were relaxed. This means that the ion types produced in HCD and by extension by CID are already very well understood. New

dx.doi.org/10.1021/pr3007045 | *J. Proteome Res.* 2012, 11, 5479−5491

fragmentation pathways of standard peptides could of course be discovered in the future, but it is unlikely that such ions would contribute very much to the overall ion intensity. For modified peptides, our Expert System and quantification of fragmentation frequencies could help to discover potential new fragment types. In this connection, we have already demonstrated straightforward extension of our approach to phosphorylated peptides. In conclusion, we have here reported the most extensive investigation into HCD of peptides and hope that the results will be useful for both small and large scale investigation of the proteome.

■ ASSOCIATED CONTENT

Supporting Information

Figure S1: Barplot displaying the number of internal fragments of an amino acid as (A) first (N-terminal) (B) last (C-terminal) amino acid of an internal fragment divided by the number of occurences of the amino acid in the dataset. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: mmann@biochem.mpg.de. Fax: +49 89 8578 2219.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank our colleagues at the Max Planck Institute of Biochemistry for help and fruitful discussions, in particular Kirti Sharma for the preparation of the phosphoproteome samples. The research leading to these results has received funding from the European Commission's Seventh Framework Programme (Grant Agreement HEALTH-F4-2008-201648/PROSPECTS).

■ ABBREVIATIONS

CID, collision-induced dissociation; ETD, electron transfer dissociation; FDR, false discovery rate; FT, Fourier transform; HCD, higher energy collisional dissociation; HPLC, high performance liquid chromatography; ICR, ion cyclotron resonance; IM, immonium ion; IPI, international protein index; LTQ, linear trap quadrupole; MS/MS, tandem mass spectrometry; PIF, precursor intensity fraction; Q_TOF, quadrupole time-of-flight instrument; TOF, time of flight

■ REFERENCES

(1) Ahrens, C. H.; Brunner, E.; Qeli, E.; Basler, K.; Aebersold, R. Generating and navigating proteome maps using mass spectrometry. *Nat. Rev. Mol. Cell Biol.* **2010**, *11* (11), 789−801.

(2) Mallick, P.; Kuster, B. Proteomics: a pragmatic perspective. *Nat. Biotechnol.* **2010**, *28* (7), 695−709.

(3) Cox, J.; Mann, M. Quantitative, high-resolution proteomics for data-driven systems biology. *Annu. Rev. Biochem.* **2011**, *80*, 273−99.

(4) Nesvizhskii, A. I.; Vitek, O.; Aebersold, R. Analysis and validation of proteomic data generated by tandem mass spectrometry. *Nat. Methods* **2007**, *4* (10), 787−97.

(5) White, F. M. The potential cost of high-throughput proteomics. *Sci. Signaling* **2011**, *4* (160), pe8.

(6) Ma, B.; Johnson, R. De novo sequencing and homology searching. *Mol. Cell. Proteomics* **2012**, *11* (2), O111 014902.

(7) Biemann, K. Contributions of mass spectrometry to peptide and protein structure. *Biomed. Environ. Mass Spectrom.* **1988**, *16* (1−12), 99−111.

(8) Oomens, J.; Young, S.; Molesworth, S.; van Stipdonk, M. Spectroscopic evidence for an oxazolone structure of the b(2) fragment ion from protonated tri-alanine. *J. Am. Soc. Mass Spectrom.* **2009**, *20* (2), 334−9.

(9) Bythell, B. J.; Somogyi, A.; Paizs, B. What is the structure of b(2) ions generated from doubly protonated tryptic peptides? *J. Am. Soc. Mass Spectrom.* **2009**, *20* (4), 618−24.

(10) Perkins, B. R.; Chamot-Rooke, J.; Yoon, S. H.; Gucinski, A. C.; Somogyi, A.; Wysocki, V. H. Evidence of diketopiperazine and oxazolone structures for HA b(2)(+) Ion. *J. Am. Chem. Soc.* **2009**, *131* (48), 17528−9.

(11) Wysocki, V. H.; Tsaprailis, G.; Smith, L. L.; Breci, L. A. Mobile and localized protons: a framework for understanding peptide dissociation. *J. Mass Spectrom.* **2000**, *35* (12), 1399−406.

(12) Boyd, R.; Somogyi, A. The mobile proton hypothesis in fragmentation of protonated peptides: a perspective. *J. Am. Soc. Mass Spectrom.* **2010**, *21* (8), 1275−8.

(13) Paizs, B.; Suhai, S. Fragmentation pathways of protonated peptides. *Mass Spectrom. Rev.* **2005**, *24* (4), 508−48.

(14) Medzihradszky, K. F. Peptide sequence analysis. *Methods Enzymol.* **2005**, *402*, 209−44.

(15) Papayannopoulos, I. A. The interpretation of collision-induced dissociation tandem mass-spectra of peptides. *Mass Spectrom. Rev.* **1995**, *14* (1), 49−73.

(16) Falick, A. M.; Hines, W. M.; Medzihradszky, K. F.; Baldwin, M. A.; Gibson, B. W. Low-mass ions produced from peptides by high-energy collision-induced dissociation in tandem mass-spectrometry. *J. Am. Soc. Mass Spectrom.* **1993**, *4* (11), 882−93.

(17) Chalkley, R. J.; Baker, P. R.; Huang, L.; Hansen, K. C.; Allen, N. P.; Rexach, M.; Burlingame, A. L. Comprehensive analysis of a multidimensional liquid chromatography mass spectrometry dataset acquired on a quadrupole selecting, quadrupole collision cell, time-of-flight mass spectrometer: II. New developments in Protein Prospector allow for reliable and comprehensive automatic analysis of large datasets. *Mol. Cell. Proteomics* **2005**, *4* (8), 1194−204.

(18) Hung, C. W.; Schlosser, A.; Wei, J. H.; Lehmann, W. D. Collision-induced reporter fragmentations for identification of covalently modified peptides. *Anal. Bioanal. Chem.* **2007**, *389* (4), 1003−16.

(19) Winter, D.; Lehmann, W. D. Sequencing of the thirteen structurally isomeric quartets of N-terminal dipeptide motifs in peptides by collision-induced dissociation. *Proteomics* **2009**, *9* (8), 2076−84.

(20) Lee, Y. J.; Lee, Y. M. Formation of c1 fragment ions in collision-induced dissociation of glutamine-containing peptide ions: a tip for de novo sequencing. *Rapid Commun. Mass Spectrom.* **2004**, *18* (18), 2069−76.

(21) Reid, G. E.; Roberts, K. D.; Kapp, E. A.; Simpson, R. J. Statistical and mechanistic approaches to understanding the gas-phase fragmentation behavior of methionine sulfoxide containing peptides. *J. Proteome Res.* **2004**, *3* (4), 751−9.

(22) Harrison, A. G.; Young, A. B.; Bleiholder, C.; Suhai, S.; Paizs, B. Scrambling of sequence information in collision-induced dissociation of peptides. *J. Am. Chem. Soc.* **2006**, *128* (32), 10364−5.

(23) Bleiholder, C.; Osburn, S.; Williams, T. D.; Suhai, S.; Van Stipdonk, M.; Harrison, A. G.; Paizs, B. Sequence-scrambling fragmentation pathways of protonated peptides. *J. Am. Chem. Soc.* **2008**, *130* (52), 17774−89.

(24) Goloborodko, A. A.; Gorshkov, M. V.; Good, D. M.; Zubarev, R. A. Sequence scrambling in shotgun proteomics is negligible. *J. Am. Soc. Mass Spectrom.* **2011**, *22* (7), 1121−4.

(25) Yu, L.; Tan, Y.; Tsai, Y.; Goodlett, D. R.; Polfer, N. C. On the relevance of peptide sequence permutations in shotgun proteomics studies. *J. Proteome Res.* **2011**, *10* (5), 2409−16.

(26) Xia, Y.; Liang, X. R.; McLuckey, S. A. Ion trap versus low-energy beam-type collision-induced dissociation of protonated ubiquitin ions. *Anal. Chem.* **2006**, *78* (4), 1218−27.

(27) Schwartz, J. C.; Senko, M. W.; Syka, J. E. P. A two-dimensional quadrupole ion trap mass spectrometer. *J. Am. Soc. Mass Spectrom.* **2002**, *13* (6), 659−69.
(28) Olsen, J. V.; Macek, B.; Lange, O.; Makarov, A.; Horning, S.; Mann, M. Higher-energy C-trap dissociation for peptide modification analysis. *Nat. Methods* **2007**, *4* (9), 709−12.
(29) McAlister, G. C.; Phanstiel, D. H.; Brumbaugh, J.; Westphall, M. S.; Coon, J. J. Higher-energy collision-activated dissociation without a dedicated collision cell. *Mol. Cell. Proteomics* **2011**, *10* (5), O111 009456.
(30) Horner, J. A.; Remes, P.; Biringer, R.; Huhmer, A.; Specht, A. Achieving increased coverage in global proteomics survey experiments using higher-energy collisional dissociation (HCD) on a linear ion trap mass spectrometer. *Application Note 538*; Thermo Fisher Scientific: San Jose, CA, 2011.
(31) Olsen, J. V.; Schwartz, J. C.; Griep-Raming, J.; Nielsen, M. L.; Damoc, E.; Denisov, E.; Lange, O.; Remes, P.; Taylor, D.; Splendore, M.; Wouters, E. R.; Senko, M.; Makarov, A.; Mann, M.; Horning, S. A dual pressure linear ion trap Orbitrap instrument with very high sequencing speed. *Mol. Cell. Proteomics* **2009**, *8* (12), 2759−69.
(32) Mann, M.; Kelleher, N. L. Precision proteomics: the case for high resolution and high mass accuracy. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105* (47), 18132−8.
(33) Ross, P. L.; Huang, Y. L. N.; Marchese, J. N.; Williamson, B.; Parker, K.; Hattan, S.; Khainovski, N.; Pillai, S.; Dey, S.; Daniels, S.; Purkayastha, S.; Juhasz, P.; Martin, S.; Bartlet-Jones, M.; He, F.; Jacobson, A.; Pappin, D. J. Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents. *Mol. Cell. Proteomics* **2004**, *3* (12), 1154−69.
(34) Bantscheff, M.; Boesche, M.; Eberhard, D.; Matthieson, T.; Sweetman, G.; Kuster, B. Robust and sensitive iTRAQ quantification on an LTQ Orbitrap mass spectrometer. *Mol. Cell. Proteomics* **2008**, *7* (9), 1702−13.
(35) Pichler, P.; Kocher, T.; Holzmann, J.; Mohring, T.; Ammerer, G.; Mechtler, K. Improved precision of iTRAQ and TMT quantification by an axial extraction field in an Orbitrap HCD cell. *Anal. Chem.* **2011**, *83* (4), 1469−74.
(36) Neuhauser, N.; Michalski, A.; Cox, J.; Mann, M. Expert System for computer assisted annotation of MS/MS spectra. *Mol. Cell. Proteomics* **2012**.
(37) Shevchenko, A.; Tomas, H.; Havlis, J.; Olsen, J. V.; Mann, M. In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nat. Protoc.* **2006**, *1* (6), 2856−60.
(38) Rappsilber, J.; Ishihama, Y.; Mann, M. Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics. *Anal. Chem.* **2003**, *75* (3), 663−70.
(39) Wisniewski, J. R.; Zougman, A.; Mann, M. Combination of FASP and StageTip-based fractionation allows in-depth analysis of the hippocampal membrane proteome. *J. Proteome Res.* **2009**, *8* (12), 5674−8.
(40) Pinkse, M. W.; Uitto, P. M.; Hilhorst, M. J.; Ooms, B.; Heck, A. J. Selective isolation at the femtomole level of phosphopeptides from proteolytic digests using 2D-NanoLC-ESI-MS/MS and titanium oxide precolumns. *Anal. Chem.* **2004**, *76* (14), 3935−43.
(41) Michalski, A.; Damoc, E.; Lange, O.; Denisov, E.; Nolting, D.; Muller, M.; Viner, R.; Schwartz, J.; Remes, P.; Belford, M.; Dunyach, J. J.; Cox, J.; Horning, S.; Mann, M.; Makarov, A. Ultra high resolution linear ion trap Orbitrap mass spectrometer (Orbitrap Elite) facilitates top down LC MS/MS and versatile peptide fragmentation modes. *Mol. Cell. Proteomics* **2012**, *11* (3), O111 013698.
(42) Cox, J.; Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **2008**, *26* (12), 1367−72.
(43) Cox, J.; Neuhauser, N.; Michalski, A.; Scheltema, R. A.; Olsen, J. V.; Mann, M. Andromeda: a peptide search engine integrated into the MaxQuant environment. *J. Proteome Res.* **2011**, *10* (4), 1794−805.
(44) Ihaka, R.; Gentleman, R. R: A language for data analysis and graphics. *J. Comput. Graphical Stat.* **1996**, *5* (3), 16.
(45) Michalski, A.; Cox, J.; Mann, M. More than 100,000 detectable peptide species elute in single shotgun proteomics runs but the majority is inaccessible to data-dependent LC−MS/MS. *J. Proteome Res.* **2011**, *10* (4), 1785−93.
(46) Houel, S.; Abernathy, R.; Renganathan, K.; Meyer-Arendt, K.; Ahn, N. G.; Old, W. M. Quantifying the impact of chimera MS/MS spectra on peptide identification in large-scale proteomics studies. *J. Proteome Res.* **2010**, *9* (8), 4152−60.
(47) Wang, J.; Bourne, P. E.; Bandeira, N. Peptide identification by database search of mixture tandem mass spectra. *Mol. Cell. Proteomics* **2011**, *10*, 12.
(48) Olsen, J. V.; de Godoy, L. M.; Li, G.; Macek, B.; Mortensen, P.; Pesch, R.; Makarov, A.; Lange, O.; Horning, S.; Mann, M. Parts per million mass accuracy on an Orbitrap mass spectrometer via lock mass injection into a C-trap. *Mol. Cell. Proteomics* **2005**, *4* (12), 2010−21.
(49) Cox, J.; Michalski, A.; Mann, M. Software lock mass by two-dimensional minimization of peptide mass errors. *J. Am. Soc. Mass Spectrom.* **2011**, *22* (8), 1373−80.
(50) Ledvina, A. R.; Savitski, M. M.; Zubarev, A. R.; Good, D. M.; Coon, J. J.; Zubarev, R. A. Increased throughput of proteomics analysis by multiplexing high-resolution tandem mass spectra. *Anal. Chem.* **2011**, *83* (20), 7651−6.
(51) Mann, M.; Wilm, M. Error-tolerant identification of peptides in sequence databases by peptide sequence tags. *Anal. Chem.* **1994**, *66* (24), 4390−9.
(52) Cox, J.; Hubner, N. C.; Mann, M. How much peptide sequence information is contained in ion trap tandem mass spectra? *J. Am. Soc. Mass Spectrom.* **2008**, *19* (12), 1813−20.
(53) Ballard, K. D.; Gaskell, S. J. Sequential mass-spectrometry applied to the study of the formation of internal fragment ions of protonated peptides. *Int. J. Mass Spectrom.* **1991**, *111*, 173−189.
(54) Harrison, A. G.; Young, A. B. Fragmentation reactions of deprotonated peptides containing proline. The proline effect. *J. Mass Spectrom.* **2005**, *40* (9), 1173−86.
(55) Steen, H.; Kuster, B.; Fernandez, M.; Pandey, A.; Mann, M. Detection of tyrosine phosphorylated peptides by precursor ion scanning quadrupole TOF mass spectrometry in positive ion mode. *Anal. Chem.* **2001**, *73* (7), 1440−8.
(56) Boersema, P. J.; Mohammed, S.; Heck, A. J. Phosphopeptide fragmentation and analysis by mass spectrometry. *J. Mass Spectrom.* **2009**, *44* (6), 861−78.
(57) Johnson, R. S.; Martin, S. A.; Biemann, K. Collision-induced fragmentation of (M+H)+ ions of peptides—side-chain specific sequence ions. *Int. J. Mass Spectrom.* **1988**, *86*, 137−54.
(58) Medzihradszky, K. F.; Campbell, J. M.; Baldwin, M. A.; Falick, A. M.; Juhasz, P.; Vestal, M. L.; Burlingame, A. L. The characteristics of peptide collision-induced dissociation using a high-performance MALDI-TOF/TOF tandem mass spectrometer. *Anal. Chem.* **2000**, *72* (3), 552−8.
(59) Winter, D.; Seidler, J.; Hahn, B.; Lehmann, W. D. Structural and mechanistic information on c(1) ion formation in collision-induced fragmentation of peptides. *J. Am. Soc. Mass Spectrom.* **2010**, *21* (10), 1814−20.
(60) Yalcin, T.; Khouw, C.; Csizmadia, I. G.; Peterson, M. R.; Harrison, A. G. Why are B ions stable species in peptide spectra? *J. Am. Soc. Mass Spectrom.* **1995**, *6* (12), 1165−74.
(61) Thompson, A.; Schafer, J.; Kuhn, K.; Kienle, S.; Schwarz, J.; Schmidt, G.; Neumann, T.; Johnstone, R.; Mohammed, A. K.; Hamon, C. Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. *Anal. Chem.* **2003**, *75* (8), 1895−904.
(62) Kelstrup, C. D.; Hekmat, O.; Francavilla, C.; Olsen, J. V. Pinpointing phosphorylation sites: Quantitative filtering and a novel site-specific x-ion fragment. *J. Proteome Res.* **2011**, *10* (7), 2937−48.
(63) Khatun, J.; Ramkissoon, K.; Giddings, M. C. Fragmentation characteristics of collision-induced dissociation in MALDI TOF/TOF mass spectrometry. *Anal. Chem.* **2007**, *79* (8), 3032−40.
(64) Tabb, D. L.; Smith, L. L.; Breci, L. A.; Wysocki, V. H.; Lin, D.; Yates, J. R., 3rd Statistical characterization of ion trap tandem mass

spectra from doubly charged tryptic peptides. *Anal. Chem.* **2003**, *75* (5), 1155−63.

dx.doi.org/10.1021/pr3007045 | *J. Proteome Res.* 2012, 11, 5479−5491