Top-down Protein Sequencing and MS³ on a Hybrid Linear Quadrupole Ion Trap-Orbitrap Mass Spectrometer*§

Boris Macek‡§, Leonie F. Waanders‡§, Jesper V. Olsen‡¶, and Matthias Mann‡¶

Top-down proteomics, the analysis of intact proteins (instead of first digesting them to peptides), has the potential to become a powerful tool for mass spectrometric protein characterization. Requirements for extremely high mass resolution, accuracy, and ability to efficiently fragment large ions have often limited top-down analyses to custom built FT-ICR mass analyzers. Here we explore the hybrid linear ion trap (LTQ)-Orbitrap, a novel, high performance, and compact mass spectrometric analyzer, for top-down proteomics. Protein standards from 10 to 25 kDa were electrospayed into the instrument using a nanoelectrospray chip. Resolving power of 60,000 was ample for isotope resolution of all protein charge states. We achieved absolute mass accuracies for intact proteins between 0.92 and 2.8 ppm using the “lock mass” mode of operation. Fifty femtomole of cytochrome c applied to the chip resulted in spectra with excellent signal-to-noise ratio and only low atomole sample consumption. Different protein charge states were dissociated in the LTQ, and the sensitivity of the orbitrap allowed routine, high resolution, and high mass accuracy fragment detection. This resulted in unambiguous charge state determination of fragment ions and identification of unmodified and modified proteins by database searching. Protein fragments were further isolated and fragmented in the LTQ followed by analysis of MS³ fragments in the orbitrap, localizing modifications to part of the sequence and helping to identify the protein with these small peptide-like fragments. Given the ready availability and ease of operation of the LTQ-Orbitrap, it may have significant impact on top-down proteomics. Molecular & Cellular Proteomics 5: 949–958, 2006.

Major goals in every mass spectrometry-based proteomic experiment are protein identification and characterization. Almost invariably, proteins are enzymatically degraded to peptides, which are much more amenable to mass spectrometric investigation (1). Further advantages of this “bottom-up proteomics” approach are that one protein generates many peptides, providing many opportunities to identify or quantify it. However, identified peptides rarely cover the whole sequence of a given protein often leading to difficulties in protein characterization, particularly in determination of posttranslational modifications (PTMs).1 In the alternative approach, termed “top-down proteomics,” intact proteins are ionized, physically fragmented, and analyzed in the mass spectrometer (for reviews, see Refs. 2–4). Because this approach starts from MS detection of the intact, fully modified protein, it has the potential for full protein characterization. Although analysis of intact proteins has been reported for almost all mass analyzers, to date only one, the FT-ICR analyzer, has sufficient resolving power and mass accuracy to efficiently analyze large protein ions. In addition, several methods especially useful for fragmentation of whole proteins have been developed for the FT-ICR analyzers, such as infrared multiphoton dissociation, sustained off-resonance irradiation, and, in particular, electron capture dissociation (ECD), which is non-photonic in nature and in some cases can cleave almost any peptide bond in proteins of up to 50 kDa (5). A similar fragmentation method, electron transfer dissociation, was recently introduced for ion traps and, in combination with charge state reduction, shows great promise for top-down proteomics on these mass analyzers (6).

CID is known to efficiently fragment proteins in ion traps, but these mass spectrometers lack sufficient resolution to resolve large protein fragment ions and their charge states. With the hybrid ion trap-FTICR mass spectrometer (LTQ-FT, Thermo Electron Corp., Bremen, Germany) it is possible to fragment large peptides or even protein ions in the ion trap and detect them with high resolution and accuracy by FT-ICR (7, 8). However, in our experience, and as shown here, the LTQ-FT is less suitable for detection of fragments produced in the LTQ due to lower sensitivity and time-of-flight effects. On the other hand, the ions can be fragmented in the ICR cell using methods like ECD and infrared multiphoton dissociation. Furthermore although the LTQ-FT is a commercial and

Published, MCP Papers in Press, February 13, 2006, DOI 10.1074/mcp.T500042-MCP200

© 2006 by The American Society for Biochemistry and Molecular Biology, Inc. This paper is available on line at http://www.mcponline.org


1 The abbreviations used are: PTM, posttranslational modification; LTQ, linear quadrupole ion trap; AGC, automatic gain control; PCM, polycyclodimethylsiloxane; ECD, electron capture dissociation; S/N, signal-to-noise ratio.
robust instrument, the necessity for a high magnetic field detector and relatively high maintenance costs tend to limit its use to specialized laboratories.

Very recently a new hybrid mass spectrometer, the LTQ-Orbitrap (Thermo Electron), was introduced (9). It consists of a linear quadrupole ion trap (LTQ) coupled to a novel mass analyzer, the orbitrap, invented by A. Makarov (10–12). In the orbitrap, ion packages circle between two concentric electrodes, and their axial motion is detected, as in the FT-ICR instrument, by recording their image currents followed by Fourier transformation of the time domain signal to obtain the mass spectrum. Importantly the orbitrap is very compact and requires no magnetic field or special maintenance. LTQ and orbitrap are coupled via the C-trap, an intermediate radio frequency-only storage device, which can also be used to store background ions of known composition. When analyte ions are added and analyzed together with this “lock mass,” sub-ppm mass accuracy for peptides is achievable (13).

No systematic analysis dealing with intact proteins and their fragments in the LTQ-Orbitrap has been reported so far. In this study, we explored the utility of the LTQ-Orbitrap mass spectrometer for top-down analysis of proteins ranging in mass from 10 to 25 kDa. Our results show that the instrument is capable of routinely achieving high sensitivity (attomole to femtomole range), high mass accuracy (low ppm), and isotope resolution of small proteins. In addition, selected multiply charged protein ions can be successfully fragmented in two stages, MS\(^2\) and MS\(^3\), in the linear ion trap, and their fragments can be transferred and measured in the orbitrap. We demonstrate ready identification of modified and unmodified proteins by MS\(^2\) and MS\(^3\) data.

**EXPERIMENTAL PROCEDURES**

All protein standards were obtained from Sigma. The following proteins were analyzed: bovine cytochrome c (catalog number C-3131), bovine α-crystallin A and B chains (catalog number C-41633), bovine β-lactoglobulin (catalog number L-3908), bovine β-casein (catalog number C-6905), and recombinant human ubiquitin (catalog number U-5382).

Proteins were dissolved in methanol/water with 0.5% formic acid immediately before analysis. Sample concentration ranged from 100 fmol/μl (acquisition of whole protein spectra) to 1 pmol/μl (acquisition of MS\(^n\) spectra). A sample volume of 0.5 μl was delivered to the mass spectrometer using a NanoMate 100 system (Advinon Biosciences, Ithaca, NY). A NanoMate low flow 2.5 μM ESI chip was used as static nanoelectrospray emitter, providing a stable flow of 20–30 nl/min.

**CID of selected protein charge states was performed in the LTQ, and fragments were subsequently transferred and measured in the ICR cell.** Isolation of protein charge states was performed in LTQ. Isolation width of 106 charges for full scan and 2 × 105 for MS\(^n\) scan. Protein mass spectra were acquired at a resolving power of 60,000, and MS\(^n\) spectra were acquired at 60,000, 30,000, or 15,000. Lock mass option was enabled in all measurements unless otherwise stated, and polydimethylcyclosiloxane (PCM) background ions (at m/z 445.120025 and 429.088735) were used for real time internal calibration as described previously (13). Unless otherwise stated, all orbitrap scans consisted of 10 microscans.

**Protein masses were determined either by deconvolution using the integrated Xcalibur Extract software (Thermo Electron) or by direct integration from the peak positions and charge states.** Expected protein fragment masses were calculated using PILGrinder software (developed in house by Peter Mortensen), Protein Prospector software (prospector.ucsf.edu) (14), or GPMAW (General Protein/Mass Analysis for Windows) (Lighthouse data) (15). MS\(^2\) spectra were searched with the web-based ProSight PTM (prosightptm.scs.uiuc.edu) (16) against either the human UniProt or a custom bovine database in “Absolute Mass” mode. A wide tolerance of up to 2000 Da was used for the protein mass to allow for differences between measured and theoretical masses due to protein modification. Fragment ion mass tolerance was in all cases set to 5 ppm, and at least five matched fragments were required for protein identification.

**MS\(^3\) spectra were searched against the National Center for Biotechnology Information non-redundant (NCBI nr) protein database (April 15, 2005; 244,0425 sequences) using the Mascot search engine (Matrix Science, London, UK) (17).** Search criteria were as follows: no enzyme specificity; precursor mass tolerance, 5 ppm; and fragment mass tolerance, 0.01 Da. Because Mascot cannot handle MS\(^3\) data, we manually added the mass of H\(_2\)O (18.0106 Da) to all precursor ions that gave good quality MS\(^2\) spectra but did not result in protein identification. This formally turns b-type ions or internal fragments into peptide precursors. To match MS\(^3\) fragments by Mascot when the precursor was a b-ion, y-ions were allowed to match only with H\(_2\)O loss.

**LTQ-FTICR Mass Spectrometry—Measurements were performed on an LTQ-FT mass spectrometer (Thermo Electron) in the positive ion mode.** Bovine cytochrome c, bovine β-lactoglobulin, and bovine β-casein were prepared for measurement and delivered to the mass spectrometer using the NanoMate 100 system in a manner as identical as possible to the LTQ-Orbitrap measurements. The instrument was fully calibrated prior to all measurements according to the manufacturer’s instructions.

**CID of selected protein charge states was performed in the LTQ, and fragments were subsequently transferred and measured in the ICR cell.** An isolation width of m/z = 10–15 was used for selected protein charge states, which were subsequently activated for 30 ms using 30% normalized collision energy and an activation q of 0.25. The instrument was controlled using TunePlus 1.1 (beta 4), and the acquired spectra were evaluated using Xcalibur 1.4 software.

**The AGC target values were set to 2 × 10\(^6\) for full scan and 2 × 10\(^5\) for the MS\(^n\) scan.** A resolving power of 100,000 was used in acquisition of protein spectra, whereas 50,000 was used for MS\(^n\) spectra. All FT scans consisted of 10 microscans.

**RESULTS AND DISCUSSION**

The utility of the LTQ-Orbitrap mass spectrometer for top-down analysis of proteins was assessed at three levels: (a) high accuracy MS measurement of the whole protein in the Orbitrap, (b) MS\(^2\) (CID) fragmentation of a single charge state
of the protein of interest in the LTQ with subsequent detection of resulting fragments in the orbitrap, and (c) MS^3 of selected CID fragment(s) in the LTQ and their detection in the orbitrap.

First we wanted to establish optimal conditions for protein measurement on the LTQ-Orbitrap. Nanoelectrospray (18, 19) is commonly used in top-down proteomics because it allows detailed investigation of complex samples for extended periods of time and because it is very sensitive. Here we used a newly developed very low flow nanoelectrospray chip (NanoMate, 2.5-μm nozzle inner diameter, Advion Bioscience), which supports stable flow rates in the true nanoelectrospray range of 20–30 nl/min. This allowed us to routinely use a volume of 0.5 μl and still acquire data for more than 15 min, more than sufficient time for all measurement sequences.

We investigated optimal parameters for acquisition of whole protein spectra in the orbitrap mass analyzer. The best sensitivity, S/N, and accuracy were obtained when each scan consisted of 10 microscans. In this regime, transients of 10 consecutive microscans are added to form a final transient on which Fourier transformation is performed. Data acquisition time was between 10 and 30 s at the resolution chosen (see below), so it was still very short compared with total available spray time. At the concentrations used, fill times for the 10^6 target value chosen was between 0.2 and 4 s, comparable to the transient time of 750 ms.

In non-mass-resolved mode, target values of up to 10^7 are possible in the LTQ part of the instrument, values much higher than the limit of about 10^6 of the C-trap. However, in mass-resolved mode, for example when storing a charge state for subsequent dissociation, only about 10^5 ions can be accumulated. In this mode, the ability to sequentially fill the C-trap would be useful as pointed out previously (13).

**Measurement of Intact Proteins**

**Sensitivity**—At the intact protein level, well defined “envelopes” arising from detection of multiple charge states were routinely obtained in the orbitrap for all investigated proteins at concentrations of 100–500 fmol/μl and total protein amounts of less than 250 fmol (Table I). The lowest amount

<table>
<thead>
<tr>
<th>Molecular mass</th>
<th>Mass accuracy</th>
<th>Amount measured</th>
</tr>
</thead>
<tbody>
<tr>
<td>Theoretical</td>
<td>Measured</td>
<td>−LM ppm</td>
</tr>
<tr>
<td>Cytochrome c</td>
<td>12,224.209</td>
<td>12,224.175</td>
</tr>
<tr>
<td>α-Crystallin (A)</td>
<td>19,820.867</td>
<td>19,820.838</td>
</tr>
<tr>
<td>α-Crystallin (B)</td>
<td>20,067.393</td>
<td>20,067.404</td>
</tr>
<tr>
<td>β-Lactoglobulin</td>
<td>18,266.394</td>
<td>18,266.396</td>
</tr>
<tr>
<td>β-Casein</td>
<td>23,969.226</td>
<td>23,969.198</td>
</tr>
<tr>
<td>Ubiquitin</td>
<td>10,026.329</td>
<td>10,026.484</td>
</tr>
</tbody>
</table>

**Table I**

**Masses and accuracies of intact proteins measured in the LTQ-Orbitrap (with and without lock mass (LM)) and amounts of proteins used for analysis**

Two charge states, about 10 isotopes, and two different scans were used for mass determination. Standard deviation was calculated from these 40 values.
analyzed in this study was 50 fmol of cytochrome c (Fig. 1). Note that this was the total protein amount used for measurement and that high S/N protein mass spectra were obtained even after one MS scan (10 microscans) on a population of about $2 \times 10^7$ ions (low attomole range) and with an acquisition time of 6 s. The same amount of cytochrome c was detected on the LTQ-FT instrument under the same measurement conditions albeit with lower intensity and S/N (Supplemental Fig. 1).

High sensitivity measurements at the whole protein level have been reported before on FT-ICR instrumentation. For example, Valaskovic et al. (20) have reported subattomole detection of 8–30-kDa proteins using capillary electrophoresis separation coupled with FT-ICR mass detection, and Belov et al. (21) have reported low zeptomole detection of proteins in the same mass range. However, these high sensitivities were not achieved routinely and required either up-front analyte separation and concentration or extensive instrument optimization. The static nanoelectrospray measurements in this study show that the orbitrap sensitivity for proteins compares favorably with other instrumentation routinely used in top-down proteomics.

Resolution—All orbitrap protein spectra presented here were recorded at 60,000 resolution at $m/z$ 400, which is the specified resolving power of the orbitrap mass analyzer. This resolution requires “1-s” measurements (750-ms transients), whereas the maximum possible measurement time of 1.8 s leads to a resolution of 100,000. Although this value is less than the maximum possible with high magnetic field FT-ICR analyzers, it was sufficient for isotope resolution of all investigated proteins (Figs. 1 and 2) and provided a good duty cycle. Furthermore resolution in the orbitrap is inversely proportional to the square root of the $m/z$ value rather than to the $m/z$ value directly as it is in the FT-ICR analyzers (10), resulting in slower decrease of observed resolution across the mass range. In practice this means that at $m/z$ values above 1111.1 resolution power of 60,000 ($m/z$ 400, 750-ms transient) of the orbitrap analyzer exceeds resolution power of 100,000 of the 7-tesla LTQ-FTICR instrument. This unique feature is useful in analysis of intact protein mass spectra where protein charge states are often observed above $m/z$ 1000. It has to be noted, however, that longer transient acquisition in the LTQ-FT instrument leads to significantly higher resolution (specified to up to 500,000), whereas in the orbitrap mass analyzer the transient acquisition time is limited to 1.8 s, and consequently its resolving power is limited to about 100,000. Isotope resolution directly enables the detection of modifications such as disulfide bridges ($\Delta m = 2$ Da) or deamidation ($\Delta m = 1$ Da) of small proteins, whereas the ability to resolve isotopic clusters of different protein charge states across the mass range is important for proper charge state assignment in complex spectra caused by overlapping protein charge distributions.
Mass Accuracy—For high mass accuracy, orbitrap spectra in this study were acquired in the lock mass mode of operation using background PCM ions from the ambient air as internal calibrants as described recently (13). In the lock mass mode, each orbitrap scan is preceded by an injection of a defined number of calibrant ions into the C-trap. Analyte ions are then added into the C-trap and injected together with the calibrant ion into the orbitrap. The calibrant ion is recognized on the fly, and masses of all ions are corrected in real time. For the intact protein measurements, the PCM ion at $m/z$ 445.120025 was used as the reference, whereas for the MS^n measurements the corresponding neutral loss ($-\text{H}_2\text{O}$) PCM ion at $m/z$ 429.088735 was used instead due to its higher abundance in the MS^n spectra. This improved mass accuracy up to 5-fold (see Table I).

Masses and standard deviations for each protein standard were calculated from up to 10 members of each isotopic distribution in at least two different charge states and two different scans. For larger proteins, where the monoisotopic peak could not readily be determined, the isotope state was calculated using the top-fitting method described by Zubarev et al. (22) with the cutoff level of $h = 50\%$. Alternatively the protein mass was determined from the protein spectra deconvoluted by Xcalibur software. As expected, all protein standards were measured with very high mass accuracy, between 0.92 ppm (S.D., 0.79 ppm) and 2.80 ppm (S.D., 1.02 ppm) (Table I). The overall (combined) absolute mass accuracy was 2.25 ppm with average S.D. of 1.46 ppm. The mass accuracy of the cytochrome $c$ measured in the LTQ-FTICR analyzer was 5.1 ppm (S.D., 2.7 ppm), which was lower than that measured in the orbitrap, probably due to relatively low signal intensity.

Although FT-ICR MS is capable of achieving extremely high mass accuracies, there are only a few reports where accuracy was better than 10 ppm at the protein level. The main reason for this is the space-charging effect, which depends on the ion population in the ICR cell. Although several strategies have been proposed to overcome this effect (23, 24), the best results were obtained when the ion population was well controlled or when internal calibration was performed. For example, Lee et al. (25) have used a dual ESI source to infuse $\text{H}_2\text{O}$-endorphin as internal calibrant and enabled “mass locking” to its doubly charged ion. In a high throughput LC-FT-ICR analysis of intact proteins of the yeast large ribosomal subunit they have identified about a hundred proteins with average absolute accuracy of about 2.5 ppm and S.D. of 2 ppm in cases when protein monoisotopic mass could be measured (25).

Although a detailed study of space-charging effects in the orbitrap mass analyzer has not yet been published, we have not observed systematic mass shifts as a function of ion number up to the limit imposed by the C-trap of about $10^6$ charges. Thus in our experiments mass accuracies for all six standard proteins were never worse than 7.5 ppm even without lock mass option. In addition, protein masses were measured with extremely high precision, almost always better than 2 ppm (Table I). Together these data demonstrate that the LTQ-Orbitrap can routinely achieve extremely high mass accuracy at the protein level without any software or hardware modification.

Analysis of PTMs at the Protein Level—One of the strengths of the top-down approach in MS-based proteomics is potentially comprehensive characterization of PTMs. Three of the proteins analyzed in this study, namely $\beta$-casein and $\alpha$-crys-
Spectrom (14) or GPMAW software (15). As demonstrated for y-ions, internal fragments were calculated using Protein Pro-}

measurement, typically better than 2 ppm, and straightforward interpretation of protein fragment spectra without ion-}

rameter. In the LTQ-Orbitrap fragmentation must be performed outside of the orbitrap mass analyzer; therefore two fragment-

tation types are currently possible. Proteins can be fragment-

tated by nozzle-skipper (in-source) dissociation or CID in the LTQ with subsequent detection of fragment ions either in the ion trap or in the orbitrap. Our initial studies using nozzle-

skimmer fragmentation resulted in very little fragmentation without useful sequence information (data not shown); there-

Therefore all fragmentation was performed by CID in the LTQ.

Measurement of Protein MS 2 Fragments in the Orbitrap

Accurate protein mass alone would be of little use in the analysis of unknown proteins or protein mixtures. As dis-

cussed before, various fragmentation methods have therefore been applied to intact proteins (or their isolated charge states) to obtain at least partial information on protein primary structure. In the LTQ-Orbitrap fragmentation must be performed outside of the orbitrap mass analyzer; therefore two fragmentation types are currently possible. Proteins can be fragmented by nozzle-skipper (in-source) dissociation or CID in the LTQ with subsequent detection of fragment ions either in the ion trap or in the orbitrap. Our initial studies using nozzle-skipper fragmentation resulted in very little fragmentation without useful sequence information (data not shown); therefore all fragmentation was performed by CID in the LTQ.

As already reported in experiments on the LTQ-FTICR instru-

ment, intact proteins readily fragment in the LTQ under the same conditions normally used for peptide sequencing (7, 8). Various charge states of all analyzed proteins produced multiply charged fragment ions that were isotopically resolved and showed excellent S/N (Figs. 3 and 4). As in the case of intact protein measurements, spectra were acquired as the sum of 10 microscans. Efficient transfer from the LTQ to the C-trap and orbitrap and absence of time-of-flight effects in-

sured that fragments could be acquired in a single mass range. MS 2 fragments were matched by ProSight (Single Pro-

tein mode) (16). Because ProSight considers only b- and y-ions, internal fragments were calculated using Protein Pro-

spector (14) or GPMAW software (15). As demonstrated for \( \beta \)-lactoglobulin and \( \beta \)-casein, fragments were measured with average absolute accuracy better than 2 ppm in the orbitrap (Supplemental Table 1).

CID of intact proteins has been performed in ion trap instru-

ments before, producing multiply charged fragment ions (for a review, see Ref. 2). Unfortunately this has had little application so far due to the inability of ion traps to resolve charge states of fragment ions and necessity to perform charge state reduction by introduction of anions into the mass analyzer. The combination of a linear ion trap and the high resolution orbitrap analyzer now enables straightforward interpretation of protein fragment spectra without ion-ion reactions.

Fragmentation Patterns—Fig. 3A shows an MS 2 spectrum of \( \beta \)-lactoglobulin upon isolation and CID fragmentation of its \([M + 15H]^{15+}\) charge state in the LTQ and subsequent de-

tection of fragment ions in the orbitrap. The fragmentation patterns of proteins in our study were in agreement with the ones reported previously for ion trap CID of intact proteins (7, 8, 27). Cleavages C-terminal to charged residues, in particular Asp, Glu, and Lys, as well as N-terminal to Pro were the major fragmentation channels, although other cleavages were observed as well (Figs. 3B and 4B). In MS 2 spectra of cyto-

chrome c, ubiquitin, and \( \beta \)-lactoglobulin, complementary b/y fragments were observed. However, \( \beta \)-casein fragmented to \( \sim 70 \) residues from the N terminus and 25 residues from the C terminus, leaving the central portion of the molecule uncovered (Fig. 4). Importantly all detected b-ions formed by CID of its \([M + 22H]^{22+}\) charge state that encompassed the five predicted phosphorylation sites indeed had all five phosphate groups attached. Together with low mass, unmodified b-ions, this locates all phosphorylation sites to between residues 9 and 57 of the sequence. Because the loss of phosphoric acid from the peptide backbone, commonly observed in the “bot-

tom-up” approach, was not observed after CID of intact pro-

teins, modified fragments could be fragmented further to give insight into the exact locations of protein modifications.

It is well known that CID fragmentation of whole proteins depends largely on the protein structure (such as the positions of disulfide bonds) and size and that larger proteins (>20 kDa) tend to fragment mostly in terminal regions. It is both an advantage and a disadvantage of CID that it concentrates fragmentation products into a few preferred channels. This increases the sensitivity for detecting these fragments but will often preclude complete characterization with single residue resolution. ECD can potentially cleave almost all peptide bonds in a protein (28); however, it suffers from relatively low efficiency, which results in lower sensitivity and increased acquisition times. Therefore, these methods are complement-

ary rather than competing. Electron transfer dissociation, a recently developed fragmentation technique for ion traps, would be very suitable for the LTQ-Orbitrap because it would combine the advantages of nonergodic fragmentation (high sequence coverage, preservation of labile bonds, etc.) with the high resolution, sensitivity, and accuracy of the orbitrap mass analyzer demonstrated here.

Protein Identification Using MS 2 Data—High accuracy measurement, typically better than 2 ppm, and straightfor-
ward assignment of the charge states of fragment ions together with the accurately determined protein mass were used for protein identification. Monoisotopic masses of the MS² fragments were submitted to ProSight PTM (Absolute Mass mode), currently the only publicly available search engine for top-down proteomics. MS² spectra of recombinant human ubiquitin were searched against the human UniProt database, whereas all other proteins were searched against both a custom made bovine database and the human UniProt database. The protein mass tolerance was set to 2000 Da to allow for potential modifications, and a minimum of five matched fragments with accuracy of 5 ppm or better were required for a hit. These stringent criteria led to unambiguous identification of all analyzed proteins. In the analysis of FLAG-tagged ubiquitin, the protein was identified with 10 fragment ions (average absolute mass accuracy, 2.45 ppm; S.D., 0.92 ppm) and probability score of 7.4 x 10^19. As expected, its mass was 1597.42 Da higher than the theoretical mass, corresponding to the difference of the FLAG tag (DYKDDD-DKKLMV) plus linker sequence and Met at its N terminus. In the case of cytochrome c, the measured mass differed from the theoretical mass by 616.172 Da, which corresponds to the mass of the heme group. This not only shows the potential of the top-down approach to conclusively identify proteins but to point to modifications not considered in the database as well.

FIG. 4. Collision-induced dissociation of β-casein. A, single scan fragmentation spectrum of the [M + 22H]²²⁺ charge state of β-casein analyzed in the orbitrap shows fragmentation at both termini. Note that all five phosphate groups remained attached to MS² fragments; neutral loss of phosphate upon CID was not observed. The insets show typical isotope patterns of small and large fragments as well as the obtained mass accuracy; the scan consisted of one microscan and took 1.12 s to acquire (at resolution = 60,000). B, sequence of β-casein with phosphorylation sites indicated by circles and fragmentation pattern observed upon CID.
When searched against the human database MS² spectra of these proteins did not lead to any hits. This demonstrates that high accuracy MS² spectra of whole proteins acquired in the orbitrap can routinely be used for high stringency database search in top-down proteomics, leading to high confidence identification and greatly constraining the nature of a possible modification.

Comparison to LTQ-FT Performance in MS²—Detection of LTQ-derived CID fragments of bovine \(\beta\)-lactoglobulin and \(\beta\)-casein was directly compared between the orbitrap and the FT-ICR mass analyzers. As expected, under the same measurement conditions, fragment ions produced in the LTQ and transferred to the ICR cell resulted in spectra of lower S/N (Supplemental Fig. 2). This resulted in a lower number of identified CID fragments as compared with the orbitrap and hence demonstrated higher sensitivity of the orbitrap mass analyzer in the MS² mode. The low and high mass regions of the FT-ICR mass spectra were particularly weak due to the time-of-flight effect the ions experience in the about 1-m-long flight tube between the LTQ and the ICR cell. However, longer accumulation times and higher AGC MS² target values significantly improved the appearance of the FT-ICR MS² spectra (not shown).

Top-down Proteomics Using MS³ on the Orbitrap

We have shown recently that an additional level of peptide fragmentation (MS³) is feasible on the LTQ-FTICR at high sensitivity and chromatographic time scale (29). Together with a new scoring algorithm, these data significantly improved confidence of peptide identifications. With this background and the excellent quality of the MS² spectra, we decided to explore the feasibility of the LTQ-Orbitrap for top-down MS³ experiments.

Fig. 5A shows the MS² spectrum of ubiquitin. The predominant ion is \(y_{58}^7\) caused by proline-directed cleavage. This ion was isolated and dissociated in the LTQ, and resulting fragments were analyzed in the orbitrap (Fig. 5B). Even a single microscan acquisition (0.6 s) led to a good S/N spectrum, which was further improved by summing 21 spectra (Fig. 5B). The spectrum contains many informative \(b\) and \(y\) ions (\(b\) refers to the MS³ fragment ion generated from the end of the sequence truncated by the MS² cleavage, see Ref.
29). To generate MS\textsuperscript{3} spectra similar in quality to Fig. 5B typically required selecting one of the dominant ions of the MS\textsuperscript{2} spectrum and a total acquisition time less than 1 min. Thus, several different MS\textsuperscript{3} spectra can be acquired of a single protein loaded at a few hundred femtomoles into the NanoMate. We also explored higher stages of MS\textsuperscript{n}. Fig. 5C shows an MS\textsuperscript{4} spectrum analyzed in the LTQ and demonstrates that interpretable spectra can still be obtained after four rounds of isolation and fragmentation.

Recently Zabrouskov et al. (8), working on an LTQ-FTICR instrument, used MS\textsuperscript{3} of proteins for increased confidence of protein identification. Fragment and protein masses alone may not be sufficient to identify a protein in cases where overlapping protein envelopes lead to simultaneous fragmentation of two or more different proteins (such as in complex mixtures) or where CID of proteins results in a low number of detected fragments, leading to multiple hits and/or low-probability scores in database search. In these cases, unambiguous protein identification may require an additional level of structural information. Additionally as mentioned above, MS\textsuperscript{3} could help to pinpoint or restrict sites of modification in proteins.

Finally we decided to investigate whether small, peptide-like MS\textsuperscript{2} fragments of proteins could be fragmented and used for protein identification in a manner similar to protein identification in the bottom-up approach. The advantage of such a strategy would be that small fragments are less likely to be modified and that there are relatively few possible fragmentation channels. We chose a small fragment of \(\beta\)-casein that was easily determined to be doubly charged based on its isotope spacing (see Fig. 4A). This fragment was accumulated in the LTQ, fragmented, and analyzed in the orbitrap. Fig. 6 shows a relatively simple MS\textsuperscript{3} spectrum similar to the MS\textsuperscript{2} spectrum of a small peptide (see Supplemental Table 2 for the peak list). Several changes were made to the normal Mascot modus to be able to search these data. Because the MS\textsuperscript{3} precursor could be a \(b\) or \(y\) ion of the protein, both possibilities were checked. If a \(y\) ion did not lead to any matches, then the mass of water was added to the precursor mass to formally convert it from a \(b\) ion to a peptide precursor. Mascot was directed to match only \(b\) or \(y\) ions (these are the \(y\) MS\textsuperscript{3} fragments generated from \(b\) ion precursors (29)). Despite the short peptide length, a “non-enzyme” specificity search by Mascot in the non-redundant database (NCBI) yielded only two significant matches. The top match was located in the N-terminal region of the \(\beta\)-casein precursor. A check of the database confirmed that Mascot had identified the eight N-terminal residues of the mature protein. The second significant hit occurred on a different protein and had a related peptide sequence accounting for matched fragments despite the high mass accuracy but was located close to the center of that protein sequence and was therefore discarded. Because the protein was modified with five phosphogroups, this example shows an interesting additional way of identifying modified proteins.

**Conclusion**

Here we investigated whether a compact LTQ-Orbitrap mass spectrometer was suitable for top-down proteomics of small proteins. We interfaced the LTQ-Orbitrap to a low flow NanoMate static nanoelectrospray source providing more than 15-min measurement time for half a microliter of solution. Subpicomole amounts of intact proteins proved sufficient to obtain high quality MS, MS\textsuperscript{2}, and MS\textsuperscript{3} spectra with mass accuracies in the few ppm range. Intriguingly database
Top-down Protein Analysis on an Orbitrap

searches with the MS² data allowed straightforward identification not only of unmodified but also of modified proteins. Because the molecular weight of proteins seldom corresponds to that calculated from the database sequence, this should be very valuable for top-down proteomics. In particular we show that small peptide-like fragments generated by protein MS² can be fragmented similarly to peptides in the bottom-up approach, potentially combining advantages of both strategies. In the future, it will be interesting to use this instrument with on-line protein separation as well as to develop new algorithms tailored to the strength of the data that it generates. This technology may then allow routine, high accuracy, and high confidence analysis of the "peptidome" and small protein component of the proteome.

Acknowledgments—We thank colleagues in the Department for Proteomics and Signal Transduction, the Beijing Genome Institute, and Thermo Electron Corp. for fruitful discussions. We also thank Reinaldo Almeida from Advin Biosciences for technical support and Yong-Bin Kim, of the Kelleher group, for prompt creation of the Prosight PTM custom bovine database.

* Work on this project at Max Planck Institute of Biochemistry was supported in part by "Interaction Proteome," a 6th Framework grant from the European Union research directorate. Work at CEBI was supported by a generous grant from the Danish National Research Foundation. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

§ Both authors contributed equally to this work.

To whom correspondence should be addressed. Tel.: 49-89-8578-2557; Fax: 49-89-8578-2219; E-mail: mmann@biochem.mpg.de.

REFERENCES


