Dimethyl Isotope Labeling Assisted De Novo Peptide Sequencing

Marco L. Hennrich,[a,b] Shabaz Mohammed,[a,b] A. F. Maarten Altelaar,[a,b] and Albert J. R. Heck[a,b,c]

[a] Biomolecular Mass Spectrometry and Proteomics Group, Bijvoet Center for Biomolecular Research and Utrecht Institute for Pharmaceutical Sciences, Utrecht University, Utrecht, The Netherlands
[b] The Netherlands Proteomics Center, Utrecht, The Netherlands
[c] Center for Biomedical Genetics, Utrecht, The Netherlands

Here, we explore a de novo sequencing strategy in which we combine Lys-N protein digestion with differential isotopic dimethyl labeling to facilitate the (de novo) identification of multiply charged peptides in ESI-MS, both under CID and ETD conditions. For a large fraction of the Lys-N generated peptides, all primary amines are present at the N-terminal lysine, enabling specific labeling of the N-terminus. Differential derivatization of only the peptide N-terminus in combination with the simultaneous fragmentation of the corresponding isotopologues allows the straightforward distinction of N-terminal fragments from C-terminal and internal fragments. Furthermore, also singly and multiply charged N-terminal fragments can easily be distinguished due to the mass differences of the isotope labeled fragment pairs. As a proof of concept, we applied this approach to proteins isolated from an avocado fruit, and were able to partially de novo sequence and correctly align, with green plant homologues, a previously uncharacterized avocado ascorbate peroxidase. (J Am Soc Mass Spectrom 2010, 21, 1957–1965) © 2010 Published by Elsevier Inc. on behalf of American Society for Mass Spectrometry

In mass spectrometry (MS) based proteomics, protein identification is typically accomplished by proteolytic digestion with trypsin followed by tandem MS (MS/MS) of the resultant proteolytic peptides [1]. The peptides are typically separated by nanoflow liquid chromatography and fragmented by collision induced dissociation (CID), followed by matching of the generated fragment spectra against in silico derived spectra from large protein sequence databases [2–4]. This conventional proteomics workflow reaches its boundaries when unexpected peptide modifications are present or when the required databases are either incomplete, error-prone, or not present at all. These boundaries can be overcome by de novo sequencing, where spectra are interpreted based on known fragmentation rules. However, the complexity of the obtained peptide fragmentation spectra poses a severe challenge, in particular for de novo sequencing strategies [5, 6], since the unknown origin (N-, C-terminal, internal) and charge state of the observed fragments generally complicates correct peak annotation. Therefore, many chemical strategies to modify peptides have been introduced to simplify MS/MS spectra. The goal is, often, to make the fragmentation spectra consist of a single but complete ion series allowing easier interpretation via de novo sequencing approaches [7–12].

Alternative to the above described chemical approaches, we recently introduced a biochemical approach to manipulate the basicity of the peptide termini, using the metalloendopeptidase Lys-N [13–15]. This enzyme has cleavage specificity N-terminal of lysine residues [16], creating peptides with a strong basicity on the N-terminus caused by the presence of the N-terminal lysine. Fragmentation of singly charged Lys-N peptides that contain a lysine as the only basic residue generates spectra dominated by N-terminal fragment ions. The simplification of the spectra is caused by the strong basicity of the N-terminus, attracting the single proton. This feature is quite apparent in MALDI-CID spectra, wherein the fragment spectra are dominated by b-ions, as recently reported by Boersema et al. [13] Interestingly, doubly charged Lys-N peptides after ESI are also amenable to N-terminal directed dissociation [14], after charge reduction and electron-transfer induced dissociation (ETD) [17–19]. We speculated that in single lysine containing Lys-N peptides, the remaining single proton is directed by the strong basicity of the N-terminus, in a similar way as known for MALDI-CID, generating dominant, easy to interpret c-ion ladders in ESI-ETD [14]. This phenomena can be further enhanced by chemically increasing the basicity of the N-terminus by imidazolinylation or guanidina-

Address reprint requests to Dr. Albert J. R. Heck or Dr. A. F. Maarten Altelaar, Utrecht University, Biomolecular Mass Spectrometry and Proteomics Group, Padualaan 8, H.R. Kruytgebouw, 3584 CH Utrecht, The Netherlands. E-mail: a.j.r.heck@uu.nl or m.altelaar@uu.nl

1044-0305/10/$32.00
doi:10.1016/j.jasms.2010.08.007

Published online August 21, 2010
Received June 9, 2010
Revised July 29, 2010
Accepted August 12, 2010

tion of the lysine side chain or dimethylation of the N-terminal primary amines [20]. Alternative strategies to facilitate de novo sequencing are based on differential labeling of the N- or C-terminus, providing mass signatures in MS/MS. Several of such methods have been described using specific labeling of the N- or C-terminus to create specific isotopic patterns in MS/MS spectra. An easy way of specific C-terminal labeling has been accomplished by proteolytic ^{18}O incorporation with the protease trypsin. Fragmentation of the ^{16}O and ^{18}O labeled peptides, either individually and/or simultaneously, enables the identification of fragments from the $^{16}O/^{18}O$ C-terminus and has been used for de novo sequencing [21–24]. Most of the N-terminal labeling strategies target primary amines, as several specific derivatization procedures are known [10, 11, 25]. These strategies are complicated by the fact that primary amines are not only present at the N-terminus but also on lysines. For specific N-terminal labeling two modification steps are needed, first the modification of the ε-amine of lysine at the protein level and then modification at the peptide level of the proteolytically formed N-termini. Examples are modification of lysine residues with succinic anhydride followed by digestion with Asp-C and N-terminal derivatization with 1-([H4/D4]-nicotinoyloxy)succinimide [11]. Other strategies use guanidination of lysine followed by N-terminal derivatization by reductive amination [26] or derivatization using sulfonate groups [10, 27]. The most recent developments in de novo sequencing strategies are often MS based, i.e., recording of MS/MS spectra with high-resolution [28, 29] or applying complementary fragmentation techniques like CAD and ECD/ETD [30, 31].

Here, we explore stable isotope dimethyl labeling in combination with Lys-N protein digestion to facilitate de novo sequencing. Since all primary amines after Lys-N cleavage are clustered at the N-terminus, specific terminal labeling can be achieved in a single step by reductive amination. Reductive amination with formaldehyde and sodiumcyanoborohydride is facile, highly selective, allows near complete derivatization and has proven popular for quantitative proteomics [25, 32–34]. This derivatization introduces two methyl groups to each primary amine and is therefore often also called dimethylation or dimethyl labeling. Labeling of Lys-N peptides leads to the introduction of four methyl groups to the N-terminus of each Lys-N peptide. As described before [20], an additional benefit of dimethyl labeling of Lys-N peptides is the increased basicity of the N-terminus. We make use of a specific differential isotopic dimethyl labeling of the N-terminus of Lys-N peptides, as this leads to a characteristic isotopic pattern on all N-terminal fragments of the MS/MS spectra. The isotopic pattern is observed for all Lys-N peptides, independent of charge state and in both CID and ETD fragmentation spectra. This strategy enables us to both identify N-terminal fragments and to distinguish between singly and doubly charged N-terminal fragments even with low-resolution MS/MS using a conventional ion trap for fragmentation.

Materials and Methods

Bovine Serum Albumin (BSA), iodoacetamide, formaldehyde (37% solution in H_2O), triethylammonium bicarbonate (1 M solution), 3-[(3-Cholamidopropyl) dimethylammonio]-1-propanesulfonate (CHAPS) and thiourea were supplied by Sigma-Aldrich (Steinheim, Germany). Formaldehyde 99% ^{13}C (20% solution in water) was purchased from Isotec (Miamisburg, OH, USA). Acetonitrile was supplied by Biosolve B.V. (Valkenswaard, The Netherlands). Ammonia (25% solution in water), acetone, urea, and formic acid were obtained from Merck (Darmstadt, Germany). Ammonium bicarbonate, dithiothreitol (DTT), poly(vinylpolypyrrolidone) (PVPP), and sodium cyanoborohydrate were purchased from Fluka (Buchs, Switzerland). Precast Criterion XT gels, XT sample buffer 4×, and XT MOPS running buffer 20× were purchased from Bio-Rad Laboratories (Hercules, CA, USA), and GelCode Blue Stain Reagent was purchased from Thermo Scientific (Rockford, IL, USA). The water used in these experiments was obtained from a Milli-Q purification system (Millipore, Bedford, MA, USA). Avocado (*Persea americana* Mill.) fruit cv. Ryan was purchased at the local store. Metalloendopeptidase Lys-N (*Grifola frondosa*) was obtained from Seikagaku Corporation (Tokyo, Japan).

Protein Extraction and Digestion

BSA was reduced and carbamidomethylated using DTT and iodoacetamide followed by digestion with Lys-N at 37 °C overnight. A protein extract of an avocado fruit mesocarp was prepared as described previously [35]. Briefly, the mesocarp was ground to a fine powder in liquid nitrogen. The powder was suspended in lysis buffer and after lysis proteins were extracted by acetone precipitation following separation by SDS-PAGE. Two bands at ~26 kDa and 31 kDa were excised. After reduction and carbamidomethylation using DTT and iodoacetamide, the proteins were digested in-gel with Lys-N at 37 °C overnight [14].

Labeling of Peptides

The samples were dried down, reconstituted in 200 μL triethylammonium bicarbonate and split in two. Either conventional or ^{13}C formaldehyde (4 μL of a 4% solution) and 4 μL of 0.6 M cyanoborohydrate solution were added to 100 μL peptide solution and was left shaking for 1 h at room temperature. The reaction was terminated by addition of 16 μL 1% ammonia solution.

NanoLC-ESI-CID/ETD-MS/MS

Peptides were subjected to nanoLC-MS/MS analysis, performed on an Agilent 1200 HPLC system (Agilent Technologies, Waldbronn, Germany) connected to an Orbitrap XL Mass Spectrometer equipped with an ETD source, Thermo Fisher (Bremen, Germany). The instrument was equipped with a 20 mm × 100 μm i.d. Reprosil C18 trap column and a 400 mm × 50 μm i.d. Reprosil C18 RP analytical column (Dr. Maisch, Ammerbuch-Entringen, Germany). Trapping and washing of the sample was performed at 5 μL/min for 10 min with 100% Solvent A, and elution was achieved during a 52 min gradient from 13% to 32% Solvent B at 350 μL/min (with Solvent B being 0.1 M acetic acid in 80% ACN and Solvent A being 0.1 M acetic acid). The flow rate was passively split from 350 μL/min to 100 nL/min and the column effluent was directly introduced into the ESI source of the MS using an in-house pulled fused silica emitter, gold-coated by a Scancoat six Pirani 501 sputter coater (Edwards Laboratories, Crawley, UK), biased to 1.7 kV. The mass spectrometer was operated in positive ion mode, from 350 to 1500 *m/z* in MS mode and with an automatic gain control (AGC) value of 5.00e + 05 and a max injection time of 250 ms. Parent ions were fragmented by CID and ETD in data dependent mode with an AGC value of 1.00e + 04 and a max injection time of 100 ms. ETD fragmentation was performed with supplemental activation for doubly charged species, fluoranthene was used as reagent anion and ion/ion reaction time in the ion trap was charge state dependent with 50 ms reaction for 2+ ions, 33.3 ms for 3+ ions and 25 ms for 4+ ions.

Data Analysis

De novo sequencing of peptides from Avocado fruit was performed manually. The obtained sequences were searched against green plants and the whole NCBI database using BLAST. Consensus sequences were identified between the peptides of the 26 kDa band and ascorbate peroxidases, chitinases, proteasome subunit α of different plants, and autocleavage products of Lys-N. Subsequently, all sequenced peptides were aligned against ascorbate peroxidases from green plants using ClustalX 2.0 [36]. BSA peptides were used as a test bed. MS/MS spectra were manually de

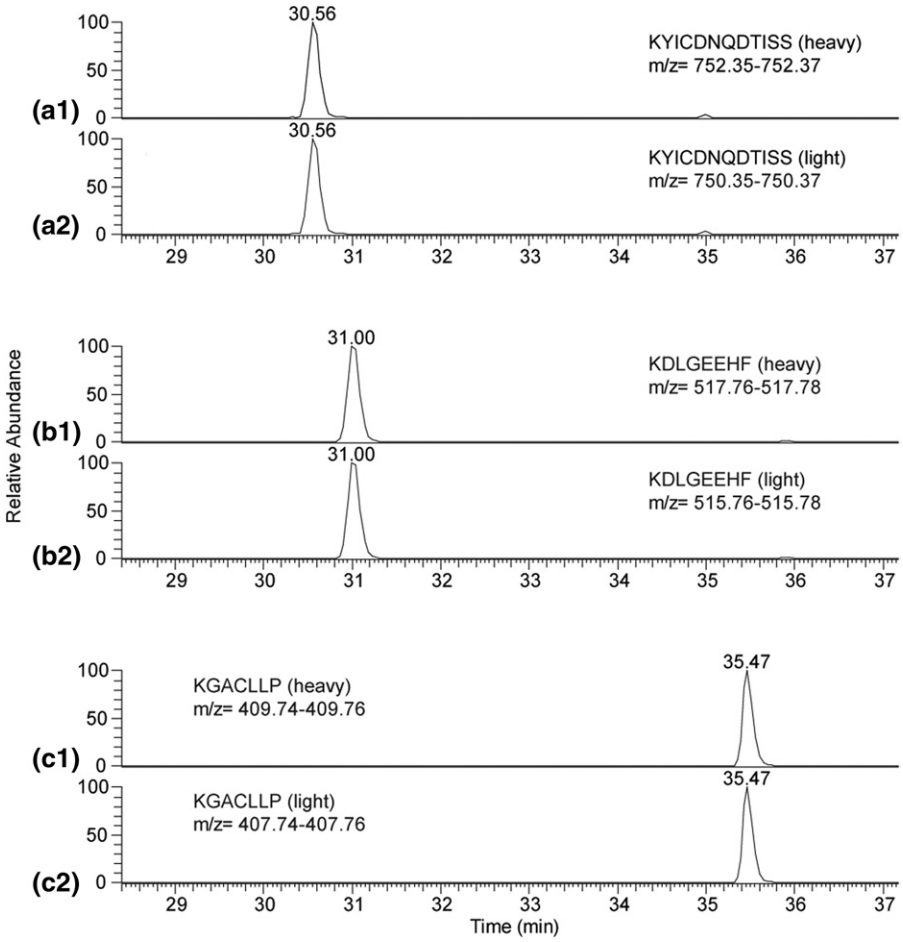


Figure 1. Extracted ion chromatograms of the monoisotopic precursors of three doubly charged peptides (**a**)–(**c**) from a Lys-N digestion of BSA. The peptides are either (1) heavy or (2) light dimethyl labeled at the N-terminus.

novo sequenced and additionally converted to DTA files by Bioworks 3.3.1 followed by searching against a homemade BSA database using the Mascot search engine software ver. 2.2.0. with carbamidomethylcysteine as a fixed modification and light and heavy dimethylation and oxidized methionine as variable modification. Lys-N was used as protease and one missed cleavage was allowed.

Results

Here, we present a method in which we aim at the simplification of peptide fragmentation spectra using the protease Lys-N and a dimethyl isotope labeling strategy to introduce a mass signature in MS/MS. Our approach is based on the simultaneous fragmentation of isotopologues of differential isotopic N-terminal labeled Lys-N peptides. In LC-MS/MS simultaneous fragmentation is only possible, when the corresponding differential labeled peptides are close in mass and exhibit identical chromatographic retention behavior. Therefore, carbon isotopologues were chosen for differential isotopic dimethyl labeling, as they are known to have nearly identical behavior in reversed-phase liquid chromatography [37]. The protein standard bovine serum albumin (BSA) was used as a test sample to set up the method. The protein was digested with Lys-N and subsequently dimethyl labeled by either ^{12}C formaldehyde or ^{13}C formaldehyde in presence of cyanoborohydride. The peptides labeled with ^{12}C formaldehyde will be referred to as "light" peptides, whereas the peptides labeled with ^{13}C formaldehyde are further called "heavy" peptides. Extracted ion chromatograms of the monoisotopic mass of three corresponding light and heavy peptides from BSA are shown in Figure 1. As expected, the retention times and peak shapes of the corresponding peptides are identical. This confirms the identical behavior of the corresponding light and heavy peptides in our reversed-phase liquid chromatography system. On an ETD enabled Orbitrap XL mass spectrometer, simultaneous fragmentation was achieved by applying a broad precursor ion isolation width of 6 *m/z* units. This isolation width has been chosen after testing a precursor ion isolation width of 3, 4, 6 and 8 *m/z* units.

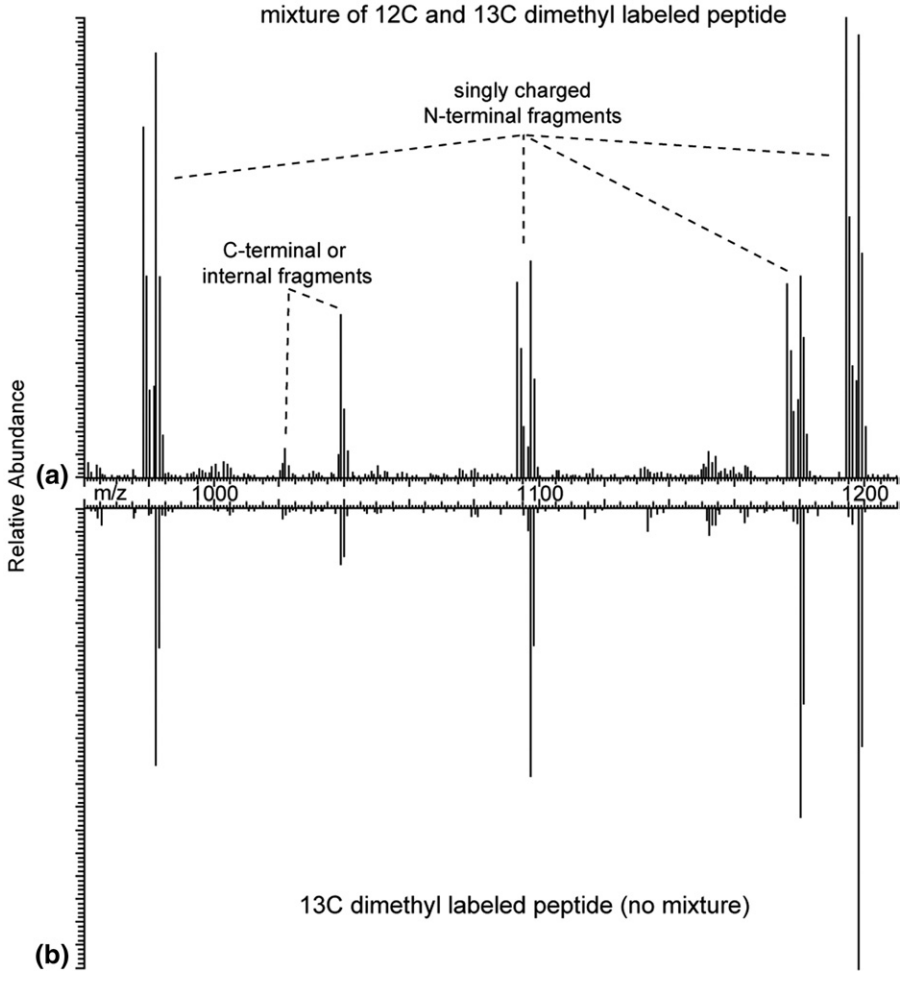


Figure 2. Comparison of a part of the CID fragmentation spectra of (**a**) a mixture of the light and heavy form and (**b**) only the heavy form of the doubly charged peptide KYICDNQDTISS from BSA.

Isolation widths smaller than 6 *m/z* units led to unequal intensities of the corresponding fragment pairs, while a greater isolation width was avoided to minimize potential unwanted co-fragmentation of co-eluting peptides. Applying an isolation window of 6 *m/z*, we observed that it is beneficial when the heavy peptide is chosen as precursor. When the light peptide was chosen, the heavy fragments were often observed with decreased intensities. Therefore, the amount of heavy peptides in the mixture was increased by 10% relative to the light peptides to enhance the possibility that the data dependent precursor selection software chooses the heavy peptide as precursor. This resulted in the clear observation of fragment ion pairs with differences of four (singly charged) and two *m/z* units (doubly charged).

In Figure 2 a zoomed in part of the CID spectrum resulting from the simultaneous fragmentation of a mixture of the light and heavy form of the doubly charged peptide KYICDNQDTISS from BSA (Figure 2a) is compared with the same part of the spectrum of only the heavy form of the same peptide (Figure 2b). C-terminal and internal fragments of the corresponding light and heavy peptide have identical masses and are observed as identical peaks. N-terminal fragments contain the isotope label and are therefore observed as fragment pairs, due to the light and heavy form of the fragment, as visualized in Figure 2. N-terminal fragments can be easily distinguished from others, facilitating annotation and spectral interpretation.

Next, we explored the potential of our method for the de novo sequencing of proteins from the avocado fruit for which currently no genome is available. The proteins from avocado fruit mesocarp were extracted and separated by SDS-PAGE. Extraction was carried out by acetone precipitation after grinding in liquid nitrogen to a powder and reconstitution in lysis buffer as described before [35]. It was observed, that addition of polyvinyl polypyrrolidone (PVPP) to the lysis buffer was crucial to remove polyphenols, and get sharp bands in SDS-PAGE. The extracted proteins were separated by SDS-PAGE and two gel bands, one at ~26 kDa and one at ~31 kDa were excised. The proteins were further processed using the workflow depicted in Figure 3. Briefly, the proteins were reduced, alkylated and digested in gel using Lys-N, followed by differential labeling. The light and heavy peptides are mixed in a 10:11 ratio respectively and the mixture is analyzed by LC-MS/MS using alternating CID and ETD for every precursor selected. Applying an isolation width of 6 *m/z* led to spectra resulting from the simultaneous fragmentation of light and heavy peptide. A typical CID spectrum obtained from one of the peptides originating from the avocado sample is depicted in Figure 4. Although this spectrum is quite complicated it could be manually interpreted using the unique signatures of the N-terminal fragments. Also, ion pairs with a mass difference of 4 and 2 *m/z* could be easily assigned as

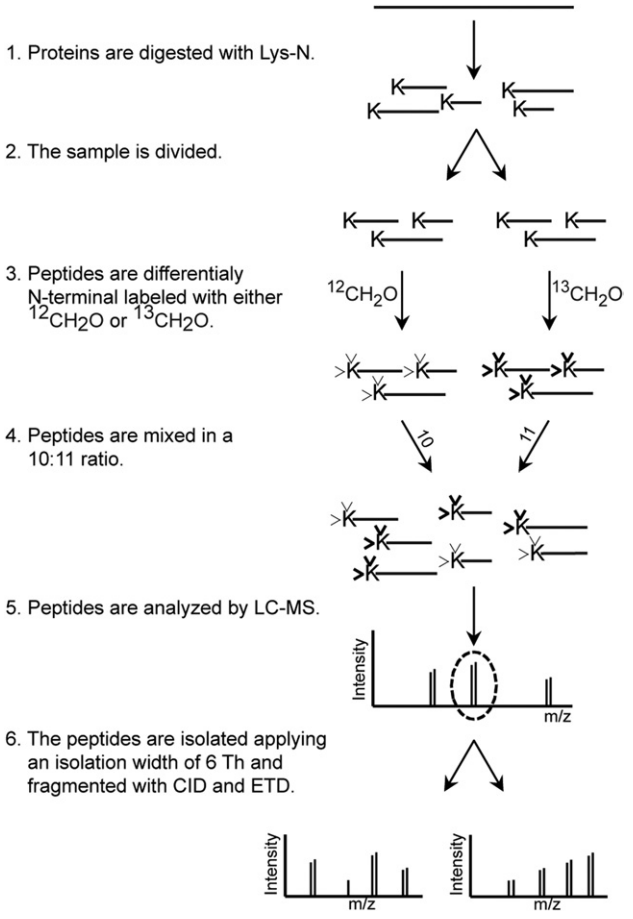
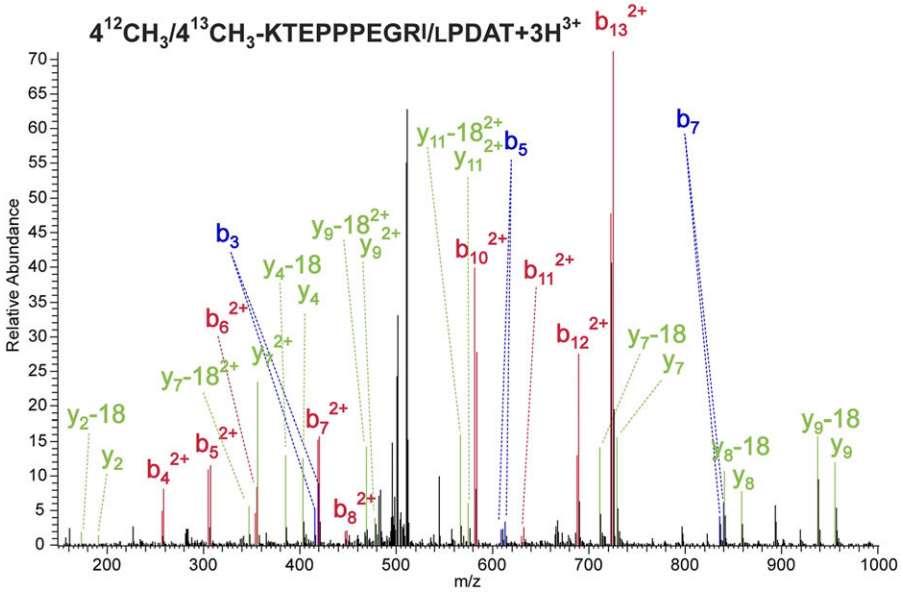


Figure 3. Schematic overview of the workflow. The proteins are first digested with Lys-N. The resulting peptides are divided in two, dimethyl labeled with either ^{12}C formaldehyde or ^{13}C formaldehyde and combined in a 10:11 ration. The mixture is analyzed by LC-MS and the peptide isotopologues are fragmented simultaneously using both CID and ETD for each precursor.

singly and doubly charged N-terminal fragments, respectively. The complex spectrum in Figure 4 is dominated by doubly charged b-ions and near complete sequence coverage was obtained solely on the basis of N-terminal fragment ions. After identification of b-ions the masses of the corresponding singly and doubly charged y-ions as well as associated fragments resulting from water or ammonia loss could be calculated and assigned. All other single peaks were, subsequently, interpreted as C-terminal or internal fragments. Since, for every precursor selected for CID fragmentation an alternating ETD spectrum was recorded we used this complementary fragmentation data to increase the spectral annotation. Despite the presence of four proline residues in the sequence which led to a relatively poor ETD spectrum, [17, 19] sufficient sequence information was observed to complete the sequence of the peptide. The two isobaric amino acids isoleucine and leucine could not be distinguished, whereby the sequence of the peptide could be identified as KTEPPPEGR(I/L)PDAT.

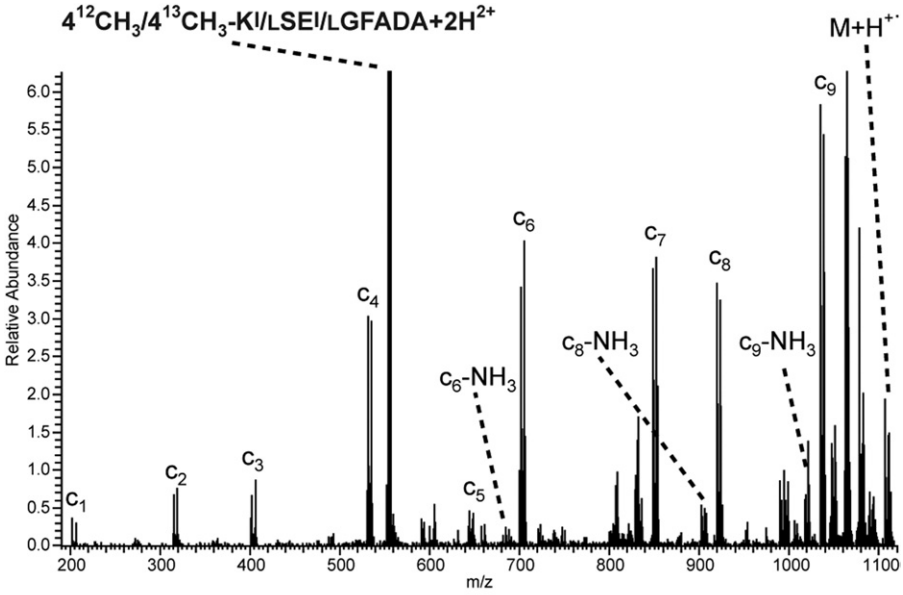


b_1	b_2	b_3	b_4	b_5	b_6	b_7	b_8	b_9	b_{10}	b_{11}	b_{12}	b_{13}	
b_1^{2+}	b_2^{2+}	b_3^{2+}	b_4^{2+}	b_5^{2+}	b_6^{2+}	b_7^{2+}	b_8^{2+}	b_9^{2+}	b_{10}^{2+}	b_{11}^{2+}	b_{12}^{2+}	b_{13}^{2+}	
K	T	E	P	P	P	E	G	R	I/L	P	D	A	T
	y_{13}	y_{12}	y_{11}	y_{10}	y_9	y_8	y_7	y_6	y_5	y_4	y_3	y_2	y_1
	y_{13}^{2+}	y_{12}^{2+}	y_{11}^{2+}	y_{10}^{2+}	y_9^{2+}	y_8^{2+}	y_7^{2+}	y_6^{2+}	y_5^{2+}	y_4^{2+}	y_3^{2+}	y_2^{2+}	y_1^{2+}

Figure 4. CID spectrum of simultaneous fragmentation of a mixture of light and heavy triply charged peptide KTEPPPEGR(I/L)PDAT from ascorbate peroxidase of avocado fruit with a table of the b- and y-ion fragments of the peptide. The colored b- and y-ions in the table have been identified in the spectrum.

A typical ETD spectrum of the light and heavy form of a doubly charged avocado peptide generated by Lys-N is shown in Figure 5. Also, in this case, full sequence coverage could be achieved now primarily on the basis of the complete and dominant c-ion series, leading to the sequence K(I/L)SE(I/L)GFADA. Furthermore, the use of a dimethyl label at the N-terminus lead to a more complete series of c-ions [20]. Also c_6, c_8, and c_9-ions with a loss of ammonia (formally b-ions) could be identified. ETD spectra of triply charged peptides were found to be more complex, as doubly charged fragments and z-ions are more abundant. Figure 6 is an example of such an ETD spectrum of a triply charged peptide from the digest of the 31 kDa band. Also, in-line with the CID data, this spectrum could be fully annotated on basis of the fragment pairs and the peptide was sequenced to completion, leading to the sequence KVADR(I/L)GFY. Our approach directly revealed that c_1, c_2, and c_4 are present as singly charged fragments, whereas c_5-c_8 are present as singly and doubly charged fragments, due to the presence of an internal arginine in the latter. The z-ions could be annotated on basis of the identified c-ions. The gap in the sequence because of the missing c_3-ion was compensated by the corresponding z-ion being present. The result from the ETD spectrum was confirmed by the corresponding CID spectrum (data not shown) and the peptide sequence matches to a 32 kDa avocado endochitinase [38]. The presented examples clearly demonstrate that the introduced mass signatures facilitate annotation and consequently sequence determination. The mass difference of 4 Da between corresponding light and heavy peptides has the advantage that even with low-resolution MS/MS, it is still possible to distinguish between singly and doubly charged fragments. Especially for annotation of fragments from triply and higher charged peptides this is a major advantage. All the obtained information was used to sequence the peptides and the sequence information of corresponding CID and ETD spectra were combined to complement one another to achieve the highest sequence coverage.

After de novo sequencing a homology-search of the identified peptides from two bands, 26 kDa and 31 kDa, of avocado fruit was performed using NCBI-BLAST. The sequences were searched against the entire NCBI database and, in a second search, against the NCBI database limited to green plants. Out of the sequenced peptides from the 26 kDa band, five peptides were annotated that had high sequence homology or were even identical with stretches from ascorbate peroxidases of different plants. One of these five is the peptide KTEPPPEGR(I/L)PDAT presented in Figure 4. The

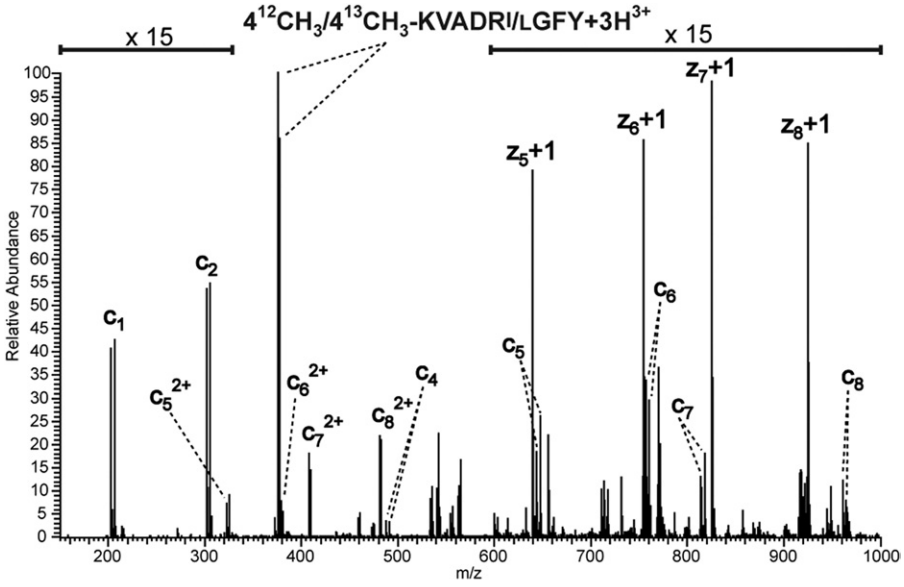


c_1	**c_2**	**c_3**	**c_4**	**c_5**	**c_6**	**c_7**	**c_8**	**c_9**	
c_1-NH_3	c_2-NH_3	c_3-NH_3	c_4-NH_3	c_5-NH_3	**c_6-NH_3**	c_7-NH_3	**c_8-NH_3**	**c_9-NH_3**	
K	**I/L**	**S**	**E**	**I/L**	**G**	**F**	**A**	**D**	**A**
	z_9+1	z_8+1	z_7+1	z_6+1	z_5+1	z_4+1	z_3+1	z_2+1	z_1+1

Figure 5. ETD spectrum of simultaneous fragmentation of a mixture of light and heavy doubly charged peptide K(I/L)SE(I/L)GFADA from ascorbate peroxidase from avocado fruit with a table of the c-, c-NH3, and z+1-ion fragments of the peptide. The fragment ions depicted in bold on a grey background have been identified in the spectrum.

N-terminal amino acids of this peptide are in a low conserved region. Hence, this peptide cannot be identified in a standard database search, as the sequence is not present in any known ascorbate peroxidase in green plants. With our method it was possible to de novo sequence this peptide and identify it as belonging to an ascorbate peroxidase, by homology-searching. Clustal X [36] alignment of all de novo sequenced peptides from the 26 kDa band against ascorbate peroxidases from green plants, resulted in the alignment of two additional semi-sequenced peptides. The multiple sequence alignment of the total of these seven de novo sequenced peptides and cytosolic ascorbate peroxidases from papaya, cotton, maize, soybean, tobacco, and arabidopsis is presented in Figure 7. The seven peptides that had homology with stretches of ascorbate peroxidases match to cytosolic as well as to thylakoid bound and stromal ascorbate peroxidases. Most of the known cytosolic ascorbate peroxidases have a length of ~250 amino acids and their average weight is between 26 and 28 kDa. The protein in the gel band has an expected weight of ~26 kDa, which corresponds with cytosolic ascorbate peroxidases. Thylakoid bound and stromal ascorbate peroxidases are longer and none of the peptides align with the regions which are specific for thylakoid bound or stromal ascorbate peroxidases. Therefore, we conclude that the peptides from the 26 kDa band belong to an as yet undescribed cytosolic ascorbate peroxidase from avocado. The alignment gave further insights into the unidentified stretches of the sequenced peptides. The combined mass of the two unidentified amino acids (xx) in the first peptide with the sequence KSxxTVSEEY equals the mass of tyrosine plus proline, which is a conserved motif in cytosolic ascorbate peroxidases of green plants. Therefore, we suggest the sequence of this peptide to be KSYPTVSEEY. As mentioned above, the two isobaric amino acids leucine and isoleucine could not be distinguished with our method. However, in peptide 2 (K(I/L)RGxxAE) leucine at position 2 and isoleucine at position 6 are conserved and in the peptide 4, 5, and 7 leucine is dominant among the species. The unidentified amino acids at position 5 and 6 of peptide 2 have the mass of phenylalanine plus leucine/isoleucine. Based on the homology, the suggested sequence of peptide 2 (K(I/L)RGxxAE) is KLRGFIAE, which fits the mass of the precursor. In peptide 6 (KYxxDEDAFF) the sum of the two unidentified amino acids equals the sum of alanine plus valine. This is an uncommon combination upon cytosolic ascorbate peroxidases, but as hydrophobic amino acids are common at this position in other species, such a variation is not exceptional.

In total, we were able to achieve a sequence coverage of over 23% by de novo sequencing and identified the protein as a cytosolic ascorbate peroxidase from avocado fruit. Besides the seven peptides from ascorbate



c_1	c_2	c_3	c_4	c_5	c_6	c_7	c_8	
c_1^{2+}	c_2^{2+}	c_3^{2+}	c_4^{2+}	c_5^{2+}	c_6^{2+}	c_7^{2+}	c_8^{2+}	
K	V	A	D	R	I/L	G	F	Y
	z_8+1	z_7+1	z_6+1	z_5+1	z_4+1	z_3+1	z_2+1	z_1+1

Figure 6. ETD spectrum of simultaneous fragmentation of a mixture of light and heavy triply charged peptide KVADR(I/L)GFY from avocado fruit with a table of the c-, doubly charged c- and z+1-ion fragments of the peptide. The low and high mass range is 15-fold magnified as indicated. The fragment ions depicted in bold on a grey background have been identified in the spectrum.

peroxidase, we de novo sequenced peptides from proteins in the 26 kDa band that have high sequence homology either to chitinases or to a proteasome subunit alpha. The abundant peptide KVADR(I/L)GFY from the 31 kDa band (Figure 6) had sequence homology with a peptide of a known and already described 32 kDa endochitinase from avocado [38].

Conclusion

Here, we have evaluated the use of Lys-N, differential isotopic dimethyl labeling and simultaneous peptide fragmentation of the corresponding light and heavy forms for de novo peptide sequencing. The resulting mass signature observed in the fragmentation spectra facilitates easy assignment of N-terminal fragments in both CID and ETD and allows distinguishing singly and doubly charged N-terminal fragments even with low-resolution mass spectrometers. As a proof-of-concept, we were able to identify a previously unreported cytosolic ascorbate peroxidase of avocado fruit and resolve more than 23% of the sequence of this protein. The presented method

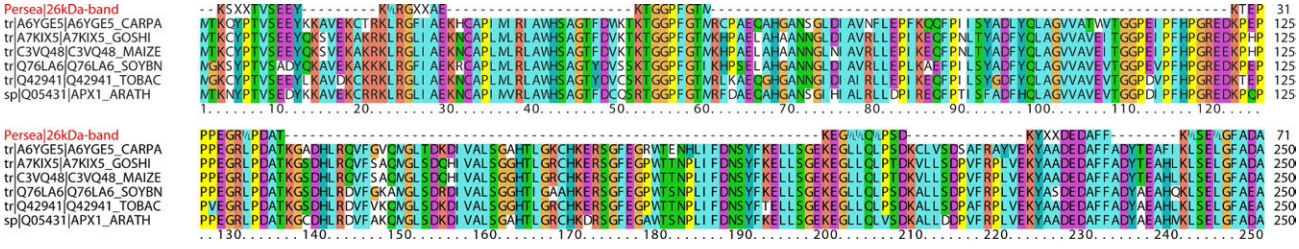


Figure 7. Multiple sequence alignment of de novo sequenced peptides from the 26 kDa band of avocado fruit and ascorbate peroxidase (a.p.) of papaya (CARPA), cytosolic a.p. 1 of cotton (GOSHI), a.p. of maize (MAIZE), cytosolic a.p. 2 of soybean (SOYBN), a.p. of tobacco (TOBAC), and cytosolic L-a.p. 1 of arabidopsis (ARATH).

is uncomplicated, cheap, and facilitates de novo sequencing.

Acknowledgments

The authors acknowledge the Heck-group members and in particular Pieter B. T. Neerincx for bioinformatic support. The Netherlands Proteomics Centre, embedded in the Netherlands Genomics Initiative, is kindly acknowledged for financial support.

References

1. Aebersold, R.; Mann, M. Mass Spectrometry-Based Proteomics. *Nature* **2003,** *422,* 198–207.
2. Domon, B.; Aebersold, R. Review—Mass Spectrometry and Protein Analysis. *Science* **2006,** *312,* 212–217.
3. Cravatt, B. F.; Simon, G. M.; Yates, J. R. The Biological Impact of Mass-Spectrometry-Based Proteomics. *Nature* **2007,** *450,* 991–1000.
4. Tang, J.; Gao, M. X.; Deng, C. H.; Zhang, X. M. Recent Development of Multi-Dimensional Chromatography Strategies in Proteome Research. *J Chromatogr B Anal. Technol. Biomed. Life Sci.* **2008,** *866,* 123–132.
5. Roepstorff, P.; Fohlman, J. Proposal for a Common Nomenclature for Sequence Ions in Mass Spectra of Peptides. *Biomed. Mass Spectrom.* **1984,** *11,* 601.
6. Tang, X. J.; Thibault, P.; Boyd, R. K. Fragmentation Reactions of Multiply-Protonated Peptides and Implications for Sequencing by Tandem Mass-Spectrometry with Low-Energy Collision-Induced Dissociation. *Anal. Chem.* **1993,** *65,* 2824–2834.
7. Spengler, B.; Luetzenkirchen, F.; Metzger, S.; Chaurand, P.; Kaufmann, R.; Jeffery, W.; Bartlet-Jones, M.; Pappin, D. J. C. In *Peptide Sequencing of Charged Derivatives by Postsource Decay MALDI Mass Spectrometry*; Elsevier Science Bv: 1997, p. 127–140.
8. Keough, T.; Youngquist, R. S.; Lacey, M. P. A Method for High-Sensitivity Peptide Sequencing Using Postsource Decay Matrix-Assisted Laser Desorption Ionization Mass Spectrometry. *Proc. Natl. Acad. Sci. U.S.A.* **1999,** *96,* 7131–7136.
9. Brancia, F. L.; Oliver, S. G.; Gaskell, S. J. Improved Matrix-Assisted Laser Desorption/Ionization Mass Spectrometric Analysis of Tryptic Hydrolysates of Proteins Following Guanidination of Lysine-Containing Peptides. *Rapid Commun. Mass Spectrom.* **2000,** *14,* 2070–2073.
10. Keough, T.; Lacey, M. P.; Youngquist, R. S. Derivatization Procedures to Facilitate De Novo Sequencing of Lysine-Terminated Tryptic Peptides Using Postsource Decay Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry. *Rapid Commun. Mass Spectrom.* **2000,** *14,* 2348–2356.
11. Munchbach, M.; Quadroni, M.; Miotto, G.; James, P. Quantitation and Facilitated De Novo Sequencing of Proteins by Isotopic N-Terminal Labeling of Peptides with a Fragmentation Directing Moiety. *Anal. Chem.* **2000,** *72,* 4047–4057.
12. Standing, K. G. Peptide and Protein De Novo Sequencing by Mass sSpectrometry. *Current Opin. Struct. Biol.* **2003,** *13,* 595–601.
13. Boersema, P. J.; Taouatas, N.; Altelaar, A. F. M.; Gouw, J. W.; Ross, P. L.; Pappin, D. J.; Heck, A. J. R.; Mohammed, S. Straightforward and De Novo Peptide Sequencing by MALDI-MS/MS Using a Lys-N Metalloendopeptidase. *Mol. Cell. Proteom.* **2009,** *8,* 650–660.
14. Taouatas, N.; Drugan, M. M.; Heck, A. J. R.; Mohammed, S. Straightforward Ladder Sequencing of Peptides Using a Lys-N Metalloendopeptidase. *Nat. Methods* **2008,** *5,* 405–407.
15. Taouatas, N.; Altelaar, A. F. M.; Drugan, M. M.; Helbig, A. O.; Mohammed, S.; Heck, A. J. R.; Strong Cation. Exchange-Based Fractionation of Lys-N-generated Peptides Facilitates the Targeted Analysis of Post-Translational Modifications. *Mol Cell. Proteom.* **2009,** *8,* 190–200.
16. Nonaka, T.; Hashimoto, Y.; Takio, K. Kinetic Characterization of Lysine-Specific Metalloendopeptidases from *Grifola frondosa* and *Pleurotus ostreatus* Fruiting Bodies. *J. Biochem.* **1998,** *124,* 157–162.
17. Zubarev, R. A.; Kelleher, N. L.; McLafferty, F. W. Electron Capture Dissociation of Multiply Charged Protein Cations. A Nonergodic Process. *J. Am. Chem. Soc.* **1998,** *120,* 3265–3266.
18. Zubarev, R. A. Electron-Capture Dissociation Tandem Mass Spectrometry. *Cur. Opin. Biotechnol.* **2004,** *15,* 12–16.
19. Syka, J. E. P.; Coon, J. J.; Schroeder, M. J.; Shabanowitz, J.; Hunt, D. F. Peptide and Protein Sequence Analysis by Electron Transfer Dissociation Mass Spectrometry. *Proc. Natl. Acad. Sci. U.S.A.* **2004,** *101,* 9528–9533.
20. Hennrich, M. L.; Boersema, P. J.; van den Toorn, H.; Mischerikow, N.; Heck, A. J. R.; Mohammed, S. Effect of Chemical Modifications on Peptide Fragmentation Behavior Upon Electron Transfer Induced Dissociation. *Anal. Chem.* **2009,** *81,* 7814–7822.
21. Takao, T.; Gonzalez, J.; Yoshidome, K.; Sato, K.; Asada, T.; Kammei, Y.; Shimonishi, Y. Automatic Precursor-Ion Switching in a 4-Sector Tandem Mass-Spectrometer and Its Application to Acquisition of the MS/MS Product Ions Derived from a Partially O-18-Labeled Peptide for Their Facile Assignments. *Anal. Chem.* **1993,** *65,* 2394–2399.
22. Qin, J.; Herring, C. J.; Zhang, X. L. De Novo Peptide Sequencing in an Ion Trap Mass Spectrometer with O-18 Labeling. *Rapid Commun. Mass Spectrom.* **1998,** *12,* 209–216.
23. Shevchenko, A.; Chernushevich, I.; Ens, W.; Standing, K. G.; Thomson, B.; Wilm, M.; Mann, M. Rapid 'De Novo' Peptide Sequencing by a Combination of Nanoelectrospray, Isotopic Labeling, and a Quadrupole/Time-of-Flight Mass Spectrometer. *Rapid Commun. Mass Spectrom.* **1997,** *11,* 1015–1024.
24. Uttenweiler-Joseph, S.; Neubauer, G.; Christoforidis, A.; Zerial, M.; Wilm, M. Automated De Novo Sequencing of Proteins Using the Differential Scanning Technique. *Proteomics* **2001,** *1,* 668–682.
25. Hsu, J. L.; Huang, S. Y.; Chow, N. H.; Chen, S. H. Stable-Isotope Dimethyl Labeling for Quantitative Proteomics. *Anal. Chem.* **2003,** *75,* 6843–6852.
26. Ji, C. J.; Guo, N.; Li, L. Differential Dimethyl Labeling of N-Termini of Peptides After Guanidination for Proteome Analysis. *J. Proteome Res.* **2005,** *4,* 2099–2108.
27. Madsen, J. A.; Brodbelt, J. S. Simplifying Fragmentation Patterns of Multiply Charged Peptides by N-Terminal Derivatization and Electron Transfer Collision Activated Dissociation. *Anal. Chem.* **2009,** *81,* 3645–3653.
28. Frank, A. M.; Savitski, M. M.; Nielsen, M. L.; Zubarev, R. A.; Pevzner, P. A. De Novo Peptide Sequencing and Identification with Precision Mass Spectrometry. *J. Proteome Res.* **2007,** *6,* 114–123.
29. Chi, H.; Sun, R. X.; Yang, B.; Song, C. Q.; Wang, L. H.; Liu, C.; Fu, Y.; Yuan, Z. F.; Wang, H. P.; He, S. M.; Dong, M. Q. pNovo: De novo Peptide Sequencing and Identification Using HCD Spectra. *J. Proteome Res.* **2010,** *9,* 2713–2724.
30. Savitski, M. M.; Nielsen, M. L.; Kjeldsen, F.; Zubarev, R. A. Proteomics-Grade De Novo Sequencing Approach. *J. Proteome Res.* **2005,** *4,* 2348–2354.
31. Bertsch, A.; Leinenbach, A.; Pervukhin, A.; Lubeck, M.; Hartmer, R.; Baessmann, C.; Elnakady, Y. A.; Muller, R.; Bocker, S.; Huber, C. G.; Kohlbacher, O. De Novo Peptide Sequencing by Tandem MS Using Complementary CID and Electron Transfer Dissociation. *Electrophoresis* **2009,** *30,* 3736–3747.
32. Boersema, P. J.; Aye, T. T.; van Veen, T. A. B.; Heck, A. J. R.; Mohammed, S. Triplex Protein Quantification Based on Stable Isotope Labeling by Peptide Dimethylation Applied to Cell and Tissue Lysates. *Proteomics* **2008,** *8,* 4624–4632.
33. Raijmakers, R.; Berkers, C. R.; de Jong, A.; Ovaa, H.; Heck, A. J. R.; Mohammed, S. Automated On-Line Sequential Isotope Labeling for Protein Quantitation Applied to Proteasome Tissue-Specific Diversity. *Mol. Cell. Proteom.* **2008,** *7,* 1755–1762.
34. Boersema, P. J.; Raijmakers, R.; Lemeer, S.; Mohammed, S.; Heck, A. J. R. Multiplex Peptide Stable Isotope Dimethyl Labeling for Quantitative Proteomics. *Nat. Protocols* **2009,** *4,* 484–494.
35. Barraclough, D.; Obenland, D.; Laing, W.; Carroll, T. A General Method for Two-Dimensional Protein Electrophoresis of Fruit Samples. *Postharvest Biol. Technol.* **2004,** *32,* 175–181.
36. Larkin, M. A.; Blackshields, G.; Brown, N. P.; Chenna, R.; McGettigan, P. A.; McWilliam, H.; Valentin, F.; Wallace, I. M.; Wilm, A.; Lopez, R.; Thompson, J. D.; Gibson, T. J.; Higgins, D. G. Clustal W and Clustal X version 2.0. *Bioinformatics* **2007,** *23,* 2947–2948.
37. Zhang, R. J.; Regnier, F. E. Minimizing Resolution of Isotopically Coded Peptides in Comparative Proteomics. *J. Proteome Res.* **2002,** *1,* 139–147.
38. Sowka, S.; Hsieh, L. S.; Krebitz, M.; Akasawa, A.; Martin, B. M.; Starrett, D.; Peterbauer, C. K.; Scheiner, O.; Breiteneder, H. Identification and Cloning of Prs-a-1, a 32-kDa Endochitinase and Major Allergen of Avocado, and Its Expression in the Yeast *Pichia pastoris*. *J. Biol. Chem.* **1998,** *273,* 28091–28097.