Discrimination of isoleucine and leucine by dimethylation-assisted MS3
Sheila Maibom-Thomsen, Søren Heissel, Ejvind Mortz, Peter Hojrup, and Jakob Bunkenborg

Anal. Chem., Just Accepted Manuscript • DOI: 10.1021/acs.analchem.8b01375 • Publication Date (Web): 17 Jul 2018

Downloaded from http://pubs.acs.org on July 24, 2018

“Just Accepted” manuscripts have been peer-reviewed and accepted for publication. They are posted online prior to technical editing, formatting for publication and author proofing. The American Chemical Society provides “Just Accepted” as a service to the research community to expedite the dissemination of scientific material as soon as possible after acceptance. “Just Accepted” manuscripts appear in full in PDF format accompanied by an HTML abstract. “Just Accepted” manuscripts have been fully peer reviewed, but should not be considered the official version of record. They are citable by the Digital Object Identifier (DOI®). “Just Accepted” is an optional service offered to authors. Therefore, the “Just Accepted” Web site may not include all articles that will be published in the journal. After a manuscript is technically edited and formatted, it will be removed from the “Just Accepted” Web site and published as an ASAP article. Note that technical editing may introduce minor changes to the manuscript text and/or graphics which could affect content, and all legal disclaimers and ethical guidelines that apply to the journal pertain. ACS cannot be held responsible for errors or consequences arising from the use of information contained in these “Just Accepted” manuscripts.
Discrimination of isoleucine and leucine by dimethylation-assisted MS3

Sheila Maibom-Thomsen1,2, Søren Heissel1, Ejvind Mørtz2, Peter Højrup1, Jakob Bunkenborg2.

1: Department of Biochemistry and Molecular Biology, University of Southern Denmark, Campusvej 55, 5230 Odense M, Denmark
2: Alphalyse A/S, Unsbjergvej 4D, 5220 Odense SØ, Denmark
*Corresponding author: bunkenborg@alphalyse.com

Abstract

Protein sequencing by mass spectrometry has transformed the field of biopharmaceutical analysis but a missing part in the analytical toolkit is the ability to distinguish between the isomeric residues isoleucine and leucine since it is a requisite for efficient analysis of the primary structure of proteins. To address this need, we have developed a novel mass spectrometric method that combines reductive dimethylation and MS3 fragmentation with LCMS peptide mapping. The dimethylation of peptide N-termini leads to intense a1 ions upon collision induced fragmentation and further fragmentation of the isoleucine/leucine a1-ion leads to informative spectra with fragments that can discriminate between the two isomers. The methodology of a1-directed MS3 was applied to two antibodies in combination with the proteases trypsin, thermolysin, chymotrypsin and pepsin to generate peptides exposing N-terminal I/L residues.

Introduction
Determination of the complete amino acid sequence of proteins remains a challenge in analytical chemistry. Considerable effort is devoted to sequencing novel proteins with potential therapeutic benefit, such as new monoclonal antibodies, and for quality control monitoring of recombinant proteins. To establish the primary structure of proteins it is paramount to be able to distinguish between isoleucine and leucine residues. The classic method of Edman degradation utilizes the retention time difference of phenylthiohydantoin (PTH)-derivated Ile/Leu residues. However, the two isomeric amino acids are not distinguishable by mass due to their identical elementary composition. Several mass spectrometry-based methods have been suggested to differentiate between Ile and Leu using dissociation methods that induce side chain fragmentation. Fragmentation of the isomeric m/z 86 immonium ion from leucine and isoleucine yields a diagnostic ion at m/z 69 with higher intensity for isoleucine. For a given peptide sequence, this means that MSn can only be used to differentiate immonium ions from peptide fragments that contain a single Ile/Leu residue. Alternatively, ECD/ETD fragmentation of z-ions with leucine/isoleucine residues at the N-terminus can be used to distinguish between the isomeric residues.

Here, leucine can fragment to a w-ion by isopropyl radical loss, whereas isoleucine predominantly loses an ethyl group thus leading to fragments with distinguishable mass. Bagal et al. developed a comprehensive LCMS method for de novo sequencing of monoclonal antibodies using a peptide sequence dependent decision tree combination of MS3 or side-chain losses of z-ions employing ETD-HCD that can be performed in a matter of days. Another approach combines germline homology data with the LCMS analysis after cleavage with chymotrypsin and leucine aminopeptidase that to some extent prefer leucine over isoleucine.

Chemical derivatization that preserves the primary structure of peptides is often used to encode reactive sites with the aim of extracting additional information. Dimethyl labeling adds two methyl groups to the α-amino group of N-terminal residues and the ε-amino group of lysine residues. The mass added by the
methyl groups can be modified by using heavy stable isotopes, creating three compositionally different methyl groups: light (CH3), intermediate (CD2H) and heavy (13CD3), which adds 28, 32 and 36 Da, respectively. Stable isotope dimethyl labeling is commonly used in quantitative proteomics, as it is fast, easy, and inexpensive to use and offers almost 100% complete labeling11. The intense a1 ion observed in HCD spectra of dimethyl labeled peptides reflects the N-terminal amino acid, and this information has proven valuable in peptide identification12, disulfide-bridge determination13 and PTM validation14. We noticed the similarity between the immonium ion from Ile/Leu residues and the dimethylated a1-ion. In this study, we measured the Ile/Leu a1-ion fragmentation pattern by MS3 of stable isotope dimethyl labelled reference peptides. The reference peptides were derivatized with three different isotopic dimethyl labelling forms to elucidate the fragmentation behavior of the I/L a1-ions. The MS3 fragment ion pattern of the dimethylated N-terminal I/L-residues showed fragmentation ions that can be used to discriminate between the isomeric side-chains. Based on these results, we created an LCMS method where a1-ion directed MS3 events are triggered by the presence of isomeric isoleucine and leucine a1 ions in the MS2. We tested the method on two antibodies and the analysis of the antibodies was performed with two different isotopic dimethyl labelling forms of four different proteolytic digests to test the method.

**Experimental**

**Materials**

All chemicals were purchased from Sigma-Aldrich unless otherwise noted. The 6-peptide mixture was the Hi3 Phos B Standard from Waters (186006011). The monoclonal antibodies Silu™Lite SigmaMAb Universal human Antibody Standard was from Sigma Aldrich (Germany) and Trastuzumab was a gift from Chromacon (Switzerland)
Methods

Antibody digestion

25µg of each monoclonal antibody was dissolved in Ultra-High Quality Milli-Q water (UHQ) (Elga, 18.2MΩ), reduced in 10mM dithiothreitol (DTT) for 20min at 56°C and alkylated in 20mM iodoacetamide (IAA) for 30min at room temperature (RT) in the dark. For trypsin (Promega, USA) chymotrypsin (Worthington, USA) digestions, the antibody was diluted with UHQ and 1 M triethylammonium bicarbonate (TEAB) to a final protein concentration of 1µg/µL in 100mM TEAB. For thermolysin digestion the antibody was diluted with UHQ and 1M TEAB, 50mM CaCl₂ (Merck Millipore, Germany) to a final protein concentration of 1µg/µL in 100mM TEAB, 5mM CaCl₂. The sample used for pepsin digestion was acidified by diluting with 0.1% formic acid (FA) to a final protein concentration of 1µg/µL (pH approximately 3). 50µL immobilized pepsin (Thermo Fischer Scientific, USA) was added to the acidified sample. Digestion for the remaining samples were carried out with an enzyme to substrate ratio of 1:20 (w/w). The trypsin, chymotrypsin and pepsin digestion samples were incubated at 37°C overnight. The thermolysin digestion sample was incubated at 70°C for 2h. Following digestion, the trypsin, chymotrypsin and thermolysin digested samples were split in 2 samples, each containing 12.5µg peptides, in preparation for dimethyl labeling. The liquid from the pepsin digestion sample was transferred to a new tube and the remaining resin washed with 1M TEAB, which was then added to the pepsin digestion liquid. The pH of this sample was checked and corrected with 1M TEAB until pH=8, and the sample was split in 2 for dimethyl labeling.

Dimethyl labeling
Hi3 PhosB peptides were dissolved in 25µL 100mM TEAB and 35pmol of each was reacted with the dimethyl labels. Light: 2µL 4% formaldehyde and 2µL 0.6M NaBH₃CN was added. Intermediate: 2µL 4% formaldehyde-₃² (Isotec, USA) and 2µL 0.6M NaBH₃CN was added. Heavy: 2µL 4% ¹³C-formaldehyde-₃² and 2µL 0.6M NaBD₃CN (Isotec, USA) was added. The samples were mixed and incubated at RT for 1h on a shaker. The labeling was quenched with 8µL 1% ammonia solution (EMD Millipore, Germany) and acidified for micropurification. The digested antibody peptides were diluted with 100mM TEAB to a final volume of 50µL. Light: 4µL 4% formaldehyde and 4µL 0.6M NaBH₃CN was added. Intermediate: 4µL 4% formaldehyde-₃² and 2µL 0.6M NaBH₃CN was added. The samples were incubated on a shaker at RT for 1h. The labeling was quenched with 16µL 1% ammonia solution and acidified with FA for micropurification.

**Micropurification**

Reversed phase columns for micropurification were made in-house using Rainin GLS-L10 pipette tips (Mettler Toledo, Germany) containing Empore C18 lining (3M Company, USA) below 1.5-2cm Poros R2 resin (20µm, Applied Biosystems, USA). The column was equilibrated with 100% acetonitrile (ACN), followed by 0.1% trifluoroacetic acid (TFA), the sample was loaded onto the column and washed with 0.1% TFA before the peptides were eluted using 60% ACN, 0.1% TFA. All purified peptides were lyophilized.

**Mass Spectrometry**

*Direct infusion MSⁿ analysis of peptides*

The lyophilized peptide mixtures were dissolved in 50% ACN, 0.1% FA to a final concentration of 100fmol/µL and transferred to a syringe and directly injected into a TSQ electrospray ionization (ESI)
source (Thermo Scientific, Germany) coupled to an Orbitrap Fusion Lumos Tribrid mass spectrometer (Thermo Scientific, Germany), using a Chemyx Fusion 101 syringe pump (Thermo Scientific, Germany) at a flowrate of 3µL/min. MS1 was detected in the Orbitrap (resolution: 120.000, scan range: 400-1000m/z, RF lens: 30%, AGC target: 500.000, minimum injection time: 100ms, 1 microscan). Relevant peptide ions were manually selected for HCD MS2/MS3 analysis performed in the Orbitrap. For the unlabeled peptide mixture Ile/Leu immonium ions (m/z 86.09) were selected and HCD MS3 analysis was performed in the Orbitrap. For the dimethyl-labeled peptide mixtures, the a1 ions from the relevant peptides were selected for MS3 at m/z 114.13, 118.15 and 122.17 for the light, intermediate and heavy labeled samples, respectively, and MS3 events were performed on these in the scan range 50-150 m/z.

**nanoLC-MS analysis of digested antibodies**

For antibody peptide analysis, the digested and dimethyl labeled peptides were dissolved in 0.1% FA (solvent A) and approximately 1µg peptides were analyzed using an EASY-nanoLC 1000 system (Thermo Scientific, Germany) coupled to the Orbitrap Fusion Lumos Tribrid mass spectrometer. The peptides were loaded onto a 2.5cm in-house packed Reprosil-Pur 120 C18-AQ (5µm; Dr. Maisch GmbH, Germany) precolumn with an internal diameter of 100µm and eluted directly onto a 19cm in-house packed Reprosil-Pur C18-AQ column (3 µm; Dr. Maisch GmbH, Germany) with an internal diameter of 75µm. A 104min HPLC gradient with a flowrate of 250 nL/min was used with a solvent B profile (95% ACN, 0.1% FA) in the following increments: 1-3% for 3min, 3-25% for 80min, 25-45% for 10min, 45-100% for 3min and 100% for the final 8min. The peptides from each proteolytic digestion and each labelling form were run separately.
The Orbitrap Fusion Lumos mass spectrometer was operated in data dependent acquisition mode with an MS1 Orbitrap scan (resolution: 120.000, scan range: 300-2000 m/z, AGC target: 400.000, maximum injection time: 100ms, RF lens: 30%, 1 microscan). For each 3s cycle, ions above a 50.000 intensity threshold and with a charge state of >+1 were selected for MS2. Monoisotopic peak determination was set to peptide and dynamic exclusion switched on (exclude after 1 time, exclusion duration: 15s, mass tolerance: ±10ppm, exclude isotopes). The selected MS2 precursors were isolated in the quadrupole at m/z 1.0, fragmented using HCD at a collision energy of 30 and detected in the Orbitrap (resolution: 30.000, first mass: 100 m/z, AGC target: 200.000, maximum injection time: 54s). For each MS2 event one m/z targeted trigger is assigned; 114.1283 m/z for the light and 118.1526 m/z for the intermediate dimethyl labeled peptides. The presence of the targeted fragment triggers an MS3 event of the ion with scan range 50-125 m/z. The MS3 precursor is fragmented by HCD and analyzed by the Orbitrap (MS1 and MS2 isolation window: 0.8 m/z, Collision energy: 30%, AGC target: 5.0e4, Resolution: 15.000, maximum injection time: 200ms, first mass: 50 m/z, 1 microscan).

Data analysis.

MS raw files were processed with Proteome Discoverer (version 2.0). The MS2 peak lists were searched using Mascot (v. 2.4.1) against a FASTA database containing the SiluMab and Trastuzumab antibody protein sequences and the protease sequences (10 entries). Carboxamidomethylation of cysteines and light/medium dimethylation of N-termini and lysine residues were set as fixed modifications and oxidation of methionine and deamidation of asparagine as variable modifications. Enzyme specificity was set as None and peptide identification was performed with a precursor mass tolerance of 10 p.p.m. and fragment mass tolerance of 0.05 Th. The MS3 spectra were extracted with a python script using Multiplierz and the intensities from the MS3 peak lists were used to assign fragment ion intensities: $I_{\text{lle}} \text{(light)} = I_{\text{mz 69}} + I_{\text{mz 85}} +$
I_{mz\ 99} ; I_{Ile\ (medium)} = I_{mz\ 69} + I_{mz\ 89} + I_{mz\ 103} ; I_{Leu\ (light)} = I_{mz\ 72} ; I_{Leu\ (medium)} = I_{mz\ 76}. The I-ratio is calculated for peptides with I_{Ile} + I_{Leu} > 0 as I_{Ile} / ( I_{Ile} + I_{Leu} ). The I/L score was calculated for peptides with a Mascot score above the Identity threshold and in most cases each assignment was supported by multiple observations of the I/L- residue from different peptide sequences, modification and charge states. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE\textsuperscript{15} partner repository with the dataset identifier PXD009295 and 10.6019/PXD009295

Results and discussion

To examine the fragmentation pattern, we performed three different states of stable isotopic dimethyl labeling on a standard peptide mixture (Hi3 PhosB, Waters). The derivatized peptides were analyzed by direct infusion on an Orbitrap Fusion Lumos, which has a lower mass range limit at m/z 50. MS3 spectra of the a1 ions from the dimethyl labeled peptides IGEEYISDLQDLRK and LITAIGDVNHPDVVGDR are shown in Figure 1 a and b, respectively. Both Leu and Ile MS3 spectra share a common fragment at m/z 58.065, 62.091 and 66.110 for the light, intermediate and heavy labeled peptides, respectively. The fragmented isoleucine a1 ion has two intense unique fragments: a peak at m/z 69.070 that is constant in all labeling conditions and a peak from the loss of 29 Da from an ethyl group. In addition, a less intense fragment originates from the loss of 15 Da from a methyl group. The leucine a1 MS3 HCD spectrum shows an intense diagnostic peak corresponding to the loss of a 42 Da propylene group rather than the loss an isopropyl group. The proposed bond cleavages are shown in Fig. 1c and the corresponding fragment masses are tabulated in Fig. 1d.
We chose two monoclonal antibodies (mAbs) with known sequence to test the methodology. SILuLite and Trastuzumab were digested, using the proteases trypsin, chymotrypsin, thermolysin, and pepsin, respectively, to generate peptides with N-terminally exposed I/L. The digested peptides were dimethyl-labeled with formaldehyde (light) and deuterated formaldehyde (intermediate). Each proteolytic digest and labelling state was analyzed separately by nanoLC-MS and to limit the number of samples we did not include the heavy label for this proof-of-concept analysis. Only one labelling state is necessary for the I/L discrimination experiment but the experimental setup with two separate isotopic modifications of the amine groups of the N-terminus and lysine residues has the advantages that the modifications can be considered as fixed modifications due to the efficient chemistry of the dimethyl labelling and that the parent and fragment ion assignments can be queried for consistency. We included a branch in the acquisition scheme (Figure 2c), so that if an HCD tandem MS spectrum contained the diagnostic a1 ion from I/L, this ion would be selected for HCD MS3 fragmentation. An example is shown in Figure 2a,b where the MS2 data of the dimethylated peptide can be confidently assigned and the intense a1-ion triggers an MS3 event. Intensities from the MS3 peak lists were used to create a simple metric: $I_{\text{Ile}}^{\text{light}} = I_{m/z \, 69} + I_{m/z \, 85} + I_{m/z \, 99}$ and $I_{\text{Leu}}^{\text{light}} = I_{m/z \, 72}$. The masses are shifted according to the stable isotope labeling scheme and the I-ratio is calculated as $I_{\text{Ile}} / (I_{\text{Ile}} + I_{\text{Leu}})$. Spectra with I-ratio greater than 0.8 are assigned as isoleucine and with a ratio less than 0.2 as leucine. The MS3 spectrum in Fig 2b has an I-ratio of 0 and is assigned as leucine in agreement with the peptide identification. Spectra with an I-ratio between 0.2-0.8 with mixed intensities from I and L a1-ions are not used for classification to avoid cases where there is a contribution from an a1-ion from co-eluting peptides within the MS2 precursor selection m/z window. Running two different states of dimethyl labelling changes the MS2 selection window and reduces the likelihood of having consistently confounding co-eluting peptides in both states. Using these criteria, we did not observe any inconsistent assignment of I/L between the two labelling forms. The peptide-to-
spectrum matches with an N-terminal I or L residue identified by Mascot database searching are documented in Supplementary Table 1 along with the a1-directed MS3 spectrum associated to it. Using this approach and the limited set of proteases, we could assign and confirm the identity of between 71-94% of the isoleucine/leucine residues in each subunit (Figure 2d).

A significant advantage of the dimethylation prior to ionization and fragmentation is that it ensures that short peptides do not cyclize and scramble\textsuperscript{16} which can confound multistage CID-based analysis. Likewise, by fixing the readout to the N-terminal amino acid, the method does not suffer from the radical site migration problem, that has been reported for w-ion based isoleucine/leucine discrimination\textsuperscript{9}. The radical site migration appears when applying both ETD and collisional activation and can be reduced by minimizing the applied normalized collision energy\textsuperscript{9}. A limitation of the proposed method is that only Ile/Leu residues exposed at peptide N-termini can be distinguished. Finding means of exposing all leucine and isoleucine residues in a protein as N-terminal residues on peptides well suited for LCMS analysis will require optimization of digestion conditions and enzymes. Possibly this limitation can be circumvented because labeling the N-terminal residue by dimethylation changes the immonium ion of the first residue and the remaining residues could be analyzed in combination with a m/z 86 immonium ion based methods for peptides that contain both an N-terminal and an internal Ile/Leu residue.

The presented method to distinguish Leu and Ile residues in peptide LC MS analysis can be extended by increasing the panel of proteases and digestion conditions. To reduce analysis time it is possible to pool different proteases digestions prior to analysis with the caveat that increasing complexity will increase the likelihood of mixed I/L spectra from co-eluting co-fragmenting peptides. While not necessary for the I/L-
discrimination the use of stable isotope labelling allows the peptide identifications and ion assignments to be experimentally tested by the observation of mass changes consistent with the proposed peptide sequence. The sensitivity of the method benefits from the altered fragmentation pattern caused by the dimethylation which increases the signal of the a₁ ion and the simplicity of the resulting data lends itself to automation. Since the discrimination of Ile/Leu is independent of peptide identification and does not need sequence dependent fragmentation, the method can easily be applied to de-novo sequencing of proteins.

Figure 1.
**Figure 2**

**Discriminating Ile/Leu during LCMS analysis.** (a) HCD MS/MS spectrum of the dimethyl-labeled peptide identified as LFPPSSEELQANK. (b) The MS3 spectrum triggered by the presence of the intense
a1 ion. The presence of the m/z 72 ion - and the absence of isoleucine-specific fragment ions – identifies the N-terminal residue as leucine. (c) Acquisition scheme for the LCMS analysis. (d) Summary of the combined identifications from all proteases and labeling states.

Supporting information available:

Table of MS3 I/L identifications

Data are available via ProteomeXchange with identifier PXD009295.

Acknowledgements

The authors thank the Innovation Fund Denmark (5016-00002B) for the financial support, and the Villum Center for Bioanalytical Sciences. We thank ChromaCon AG for providing the Antibody Trastuzumab.

Competing financial interests

The authors declare no competing financial interests.

References


Table of Contents Figure