Systematic and quantitative comparison of digest efficiency and specificity reveals the impact of trypsin quality on MS-based proteomics

Julia Maria Burkhart\textsuperscript{a}, Cornelia Schumbrutzki\textsuperscript{a}, Stefanie Wortelkamp\textsuperscript{a}, Albert Sickmann\textsuperscript{a, b}, René Peiman Zahedi\textsuperscript{a, *}

\textsuperscript{a}Leibniz-Institut für Analytische Wissenschaften – ISAS - e.V., Dortmund, Germany
\textsuperscript{b}Medizinisches Proteom-Center (MPC), Ruhr-Universität, Bochum, Germany

Abstract

Trypsin is the most frequently used proteolytic enzyme in mass spectrometry-based proteomics. Beside its good availability, it also offers some major advantages such as an optimal average peptide length of ~14 amino acids, and typically the presence of at least two defined positive charges at the N-terminus as well as the C-terminal Arg/Lys, rendering tryptic peptides well suited for CID-based LC-MS/MS. Here, we conducted a systematic study of different types of commercially available trypsin in order to qualitatively and quantitatively compare cleavage specificity, efficiency as well as reproducibility and the potential impact on quantitation and proteome coverage. We present a straightforward strategy applied to complex digests of human platelets, comprising (1) digest controls using a monolithic column HPLC-setup, (2) SCX enrichment of semitryptic/nonspecific peptides, (3) targeted MRM analysis of corresponding full cleavage/missed cleavage peptide pairs as well as (4) LC-MS analyses of complete digests with a three-step data interpretation. Thus, differences in digest performance can be readily assessed, rendering these procedures extremely beneficial to quality control not only the trypsin of choice, but also to effectively compare as well as optimize different digestion conditions and to evaluate the reproducibility of a dedicated digest protocol for all kinds of quantitative proteome studies.

© 2011 Elsevier B.V. All rights reserved.

Keywords:
Trypsin digestion quantitative proteomics quality control digest control SCX

1. Introduction

In the last decade, due to their enormous potential for biomedical and biochemical research, quantitative strategies have become increasingly important in proteomic research. For a plethora of biological questions various chemical [1–6] and metabolic labeling [7] as well as label free techniques [8–12] have been established. In particular improvements in technical instrumentation [13,14] and software algorithms [15,16] have led to increased reproducibility for reliable and robust quantitative results. However, despite stable and sensitive LC-MS systems and elaborate software, quantitative analyses may fail and yield unreliable results in case of inefficient and irreproducible sample preparation.

Abbreviations: 1-DE, one dimensional gel electrophoresis; IAA, iodoacetamide; SCX, strong cation exchange; BCA, bicinchoninic acid; Gu-HCl, guanidinium hydrochloride; FA, formic acid; MRM, multiple reaction monitoring.

* Corresponding author at: Leibniz-Institut für Analytische Wissenschaften – ISAS - e.V., Otto-Hahn-Str. 6b, D-44227 Dortmund, Germany. Tel.: +49 231 1392 4143; fax: +49 231 1392 4850.
E-mail address: rene.zahedi@isas.de (R.P. Zahedi).

1874-3919/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.jprot.2011.11.016
Generally, sample preparation has to be optimized separately for each specimen with regard to sampling, lysis, fractionation and finally proteolytic cleavage. To this end trypsin, which was the first proteolytic enzyme ever discovered – by Wilhelm Kühne in 1876 in pancreatic juice [17] – is still the most commonly-used cleavage agent in protein chemistry, and particularly in mass spectrometry-based proteomics. It cleaves C-terminal to Lys and Arg residues and, in the literature, its specificity has been reported to range from almost 100% [18] to significantly lower (~75% based signal intensities) [19], while showing also transpeptidase activity [20]. One of its main advantages is the generation of peptides with an average length of ~14 amino acids (based on an in-silico digestion of the human Uniprot database) usually containing at least two defined positive charges: at the N-terminus and the C-terminal Arg/Lys, respectively, rendering tryptic peptides perfectly suited for typical CID-based LC-MS/MS analyses. Beside minor side-activities [20,21], trypsin specificity and efficiency can be further influenced by enzyme contamination [22], inappropriate storage [23] as well as inappropriate digestion conditions (temperature, pH, time, presence of protease inhibitors or detergents, etc.) [24].

Since differential studies, in particular, require an optimal and robust digestion of proteins – to improve not only inter- run reproducibility but also overall sequence coverage – proteolytic digestion is one of the limiting steps in peptide-centric quantitative proteomics. Various vendors offer trypsins that can strongly differ in price, activity and side activity. Especially for large amounts of protein or extensive sample cohorts, the use of less expensive trypsins might be preferred to reduce experimental costs. Yet the actual choice of trypsin - as well as inappropriate storage, shipment or digestion conditions - can have severe consequences on the outcome of a proteomics study.

Initiated by an incident in our laboratory, when a new batch of an inexpensive trypsin which for many years has been successfully utilized for large-scale digests, suddenly yielded an unexpected but reproducible decrease in digest efficiency as well as specificity, we conducted a systematic study comparing six different trypsins (see Table 1 and Fig. 1) from different vendors and price ranges. The main goal was to develop a strategy to assess differences in efficiency, specificity and reproducibility of tryptic cleavage and the resulting impact on proteome coverage as well as quantitation. Hence, our trypsin evaluation comprises (a) digest controls using a monolithic column setup [25], (b) global proteome analyses combined with digest-specific data interpretation, (c) specific analyses of fractions enriched in semitryptic peptides and (d) quantitative analyses of reproducibility and yield of missed cleavage occurrence, based on selected reaction monitoring of specific peptide pairs such as DFVSFDK/DFGSFDKFK (from human superoxide dismutase) and ITIADCQGQL/ITIIADCGQLE (from human peptidylprolyl isomerase A-like, isoform CRA_c ) in a complex digest of human platelets.

In accordance with Picotti et al. [19], our data indicate, that, in general, all trypsins generate a certain amount of semitreptic cleavages, non-specific cleavages and missed cleavages. Consequently, for the trypsins evaluated here, clear differences with respect to specificity, efficiency and reproducibility can be detected: interestingly, the prices do not fully correlate with performance.

Based on our findings, we recommend conducting some of these straightforward experiments (a) to evaluate, optimize and compare digestion conditions in general and (b) to regularly monitor the performance of existing or new trypsin stocks before using them for any kind of important/limited/expensive samples.

### 2. Material and methods

Different trypsins were acquired from Promega (Mannheim, Germany) and Sigma Aldrich (Steinheim, Germany); namely Promega sequencing grade (Trypsin 1), Sigma T-8658 (Trypsin 2), Sigma T-1426 (Trypsin 3), Sigma T-0303 (Trypsin 4), Sigma 93614 (Trypsin 6). Finally, we assayed a trypsin from Fluka which was stored at −30°C for more than 7 years and which is no longer commercially available (Trypsin 5). The numbers assigned here will be used throughout the rest of the manuscript.

Ammonium bicarbonate (NH₄HCO₃), iodoacetamide (IAA), guanidinium hydrochloride (Gu-HCl) and formaldehyde were purchased from Sigma-Aldrich (Steinheim, Germany). All equipment for one-dimensional electrophoresis (1-DE) was obtained from NuPAGE®, Invitrogen (Karlsruhe, Germany). High purity chemicals for silver staining, sodium di-hydrogen phosphate (NaH₂PO₄), di-sodium hydrogen phosphate (Na₂HPO₄), sodium chloride (NaCl), and calcium chloride (CaCl₂) were purchased from Merck (Darmstadt, Germany). Dithiothretiol (DTT) was acquired from Roche Diagnostics (Mannheim, Germany), bicinechonic acid (BCA) assay from Thermo Scientific and Spec C18AR tips from

### Table 1 – Overview of the evaluated trypsins.

<table>
<thead>
<tr>
<th>no.</th>
<th>manufacturer</th>
<th>ordering no.</th>
<th>side activity according to manufacturer</th>
<th>remarks</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Promega</td>
<td>V5111</td>
<td>none</td>
<td>sequencing grade modified Trypsin, TPCK treated</td>
</tr>
<tr>
<td>2</td>
<td>Sigma</td>
<td>T-8658</td>
<td>none</td>
<td>suitable for protein sequencing TPCK treated</td>
</tr>
<tr>
<td>3</td>
<td>Sigma</td>
<td>T-1426</td>
<td>chymotryptic ≤0.1 units/mg protein</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Sigma</td>
<td>T-0303</td>
<td>chymotryptic ≤1.0 units/mg protein</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Fluka</td>
<td>No longer commercially available</td>
<td>used as a cell-detachment agent (trypsinization of tissue fragments)</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Sigma</td>
<td>93614</td>
<td>chymotryptic ≤0.2%</td>
<td></td>
</tr>
</tbody>
</table>
Agilent (Darmstadt, Germany). All solvents for HPLC were obtained from Biosolve (Valkenswaard, The Netherlands).

2.1. Platelet isolation and purification

Human platelets from apheresis concentrates were isolated and purified according to the protocol in Moebius et al. [26]. Platelets were resuspended and lysed in 800 μl 2 M Gu-HCl, 50 mM Na₂HPO₄, pH 7.8. Protein concentration was determined using the BCA assay. Disulfide bonds were reduced with 10 mM DTT for 30 min at 56 °C. Afterwards free sulfhydryl groups were carbamidomethylated using 30 mM IAA for 30 min at room temperature in the dark.

2.2. Digest and digest control

Aliquots of 500 μg patelet lysate were treated identically using different trypsins (fixed protease:protein ratio of 1:30). All digests were performed independently in triplicate. Therefore, Gu-HCl was diluted to a concentration of 0.2 M using 50 mM Na₂HPO₄, pH 7.8. Protein concentration was determined using the BCA assay. Disulfide bonds were reduced with 10 mM DTT for 30 min at 56 °C. Afterwards free sulfhydryl groups were carbamidomethylated using 30 mM IAA for 30 min at room temperature in the dark.

2.3. SCX chromatography

Prior to strong cation exchange chromatography (SCX) samples were desalted using Spec C18AR tips according to the manufacturer’s instructions and dried under vacuum. SCX was performed using a 150 mm x 1 mm PolySULFOETHYL column (200 Å pore size, 5 μm particle size, PolyLC, USA) in combination with an inert Ultimate 3000 HPLC system (Dionex, Germering, Germany). Therefore, 100 μg of desalted sample was resuspended in 5 mM NaH₂PO₄, pH 2.7 (SCX buffer A), each. Peptides were eluted at a flow rate of 50 μL/min with increasing salt content (SCX buffer B: 5 mM Na₂HPO₄, 500 mM NaCl, 15% ACN, pH 2.7). One-minute fractions were collected and subsequently combined to yield three fractions covering charge states +1 to +2 (SCX 1, SCX 2 and SCX 3), as depicted in Fig. 2. For all eighteen digest protocols and replicates, 0.6%, 0.75% and 0.2% (v/v), respectively of fractions 1, 2 and 3 were analyzed by nano-LC-MS/MS on an LTQ-Orbitrap XL.

2.4. nano-LC-MS/MS

nano-LC-MS/MS was performed on LTQ-Orbitrap XL or LTQ-Orbitrap Velos mass spectrometers (Thermo Fisher Scientific, Bremen, Germany) coupled to Ultimate 3000 Rapid Separation Liquid Chromatography (RSLC) systems (Dionex, Germering, Germany). Briefly, peptides were preconcentrated on a reversed-phase (RP) trapping column (Acclaim PepMap,
75 μm × 2 cm C18, 100 Å, Dionex) in 0.1% TFA followed by RP separation (Acclaim PepMap RSLC 75 μm × 15 cm, 2 μm, 100 Å, Dionex) using a binary gradient (solvent A: 0.1% FA, solvent B: 0.1% FA, 84% ACN) from 5% to 50% B at a flow rate of 300 nL/min in 50 min, 90 min or 180 min.

MS survey scans were acquired within the Orbitrap from m/z 300 to 2000 at a resolution of 60,000 using the polysiloxane ion at m/z 445.120030 as lock mass [27]. The ten (Orbitrap XL: five) most intense signals were subjected to collision-induced dissociation (CID) in the ion trap taking into account a dynamic exclusion of 12 s. CID spectra were acquired with a normalized CE of 35%, a default charge state of 2 and an activation time of 30 ms. AGC target values were set to 10^6 for Orbitrap MS and 10^4 for ion trap MS^n scans.

Alternatively, fragmentation of the three most intense signals was conducted in the higher energy collision dissociation (HCD) cell of the LTQ-Orbitrap Velos. HCD spectra were acquired with a normalized CE of 35%, a default charge state of 2 and an activation time of 0.1 ms with a resolution of 7,500. Orbitrap AGC target values were set to 10^6 for MS and 2×10^5 for MS^n.

Fig. 2 – SCX fractionation of trypsin digests. Already from the UV traces, a substantial increase in the fraction of singly charged peptides in SCX 1 is clearly detectable for all trypsins compared to Trypsin 1. Based on triplicate analyses for each trypsin and each fraction the median proportion as well as corresponding standard deviations (SD) of fully tryptic peptide identifications related to the total number of peptide identifications (tryptic/semitryptic/non-specific) are given. Using Trypsin 1, fractions SCX 1 and SCX 2, which are potentially enriched in semitryptic and non-specifically cleaved peptides, contain 84.2 ± 1.4% and 82.8 ± 1.5% fully-tryptic peptides, whereas using Trypsin 6 even fraction SCX 3, which should be enriched in fully-tryptic peptides, contains a share of only 53.2 ± 22.8%. These results strongly indicate a reduced specificity and efficiency of Trypsin 6. SG = sequencing grade.

2.5. Data interpretation

Data interpretation was accomplished with the help of Proteome Discoverer 1.2 (Thermo Scientific). Therefore, all data were searched with Mascot 2.3.2 (Matrix Science) against the human Uniprot database (November 04, 2010; 20,322 sequences) using the following settings: (1) trypsin with a maximum of two missed cleavages, (2) carbamidomethylation of cysteine as fixed and (3) phosphorylation of Ser/Thr/Tyr as well as oxidation of Met as variable modifications. To account for charge-reducing peptide modifications, SCX fractions corresponding to charge state +1 were searched with these additional dynamic modifications: (4) acetylation of protein N-termini and (5) N-terminal Glu→pyroGlu as well as (6) N-terminal Gln→pyroGlu conversion.

In general, to maximize the number of identifications and to account for non-specific cleavage, database searches were conducted in three successive steps: (i) all MS/MS spectra were searched using trypsin as protease. Subsequently, all spectra not identified thus, were (ii) researched with semitryptic cleavage. (iii) Finally, to account for the potential occurrence of totally non-specific cleavage events, all spectra not identified in the semitryptic search were then re-searched using “none” as protease setting. After database searching, the following filter criteria were applied to all results: (7) high confidence corresponding to an FDR <1%, (8) ≤4 ppm mass deviation, (9) peptide length >6 and <22 amino acids, (10) at least two unique peptides per protein.

2.6. MRM

Multiple reaction monitoring was conducted using a TSQ Vantage mass spectrometer coupled to an inert Ultimate 3000 RSLC nano system as described above, in conjunction with the Pinpoint software (both from Thermo Scientific). Briefly, Orbitrap Velos HCD data from entire platelet lysates were used to select sequences which could be detected with and without missed cleavage sites. For these peptides, the 4–5 most intense ions were chosen to build a specific and scheduled MRM method, using 1.5 s as fixed cycle time and a Q1 resolution of 0.2. Finally, all digest replicates were analyzed successively with this method and data interpretation was conducted using Pinpoint.
2.7. Trypsin purity analysis by 1-DE

1 μg of each trypsin were loaded on a precast 1-DE gel (NUPAGE, Invitrogen). After separation, bands were visualized by silver staining according to Blum et al. [28] and in-gel digestion of excised gel bands was accomplished as described previously [29]. Peptide extracts were analyzed on an LTQ Obitrap XL (49 min gradient, five most intense ions for CID) and data interpretation was done as described above, whereas here the entire Uniprot database (November 04, 2010; 522,019 sequences) was used.

3. Results

To quantitatively compare the different trypsins, we monitored efficiency as well as specificity of tryptic digests generated from human platelets using a multi-pronged approach, summarized in Fig. 1. (1) All digests were analyzed using a monolithic column setup allowing for parallel separation and sensitive detection of peptides as well as proteins within a single run, as depicted in Fig. 3. Thus, different proteolytic digests can be readily and thoroughly compared, which is also an important prerequisite for any kind of quantitative proteomics workflow. (2) 1 μg aliquots of each sample were directly analyzed by LC-MS/MS in conjunction with a three-step database search as described in the experimental section. (3) In addition, all digests were separated by SCX at pH 2.7, and three pooled fractions covering charge states +1 to +2 were analyzed by nano-LC-MS/MS. Indeed, fraction +1 is of particular interest, since here a large share of the semi-tryptic peptides that usually contain only one positive charge (at the N-terminus) will be enriched in addition to other charge-reduced peptide species (phosphorylated peptides, sialylated glycopeptides, N-terminally acetylated peptides, etc.) [30,31]. Since peptides which are charge reduced due to post-translational modifications are usually present substoichiometrically, UV signal intensity in the +1 charge state area of the SCX gradient directly correlates with decreased digest specificity, as depicted in Fig. 2. (4) Moreover, corresponding peptide pairs (same sequence with 0 and 1 missed cleavages) were relatively quantified in all digest samples by MRM to assess the level of reproducibility as well as cleavage efficiency for all trypsins. (5) Finally, all evaluated trypsins were analyzed by 1-DE and subsequent LC-MS analyses of excised gel bands.

As illustrated in Fig. 3, a monolithic column setup can be utilized for evaluation of digest efficiency and reproducibility. From these separations it can be readily concluded that missed cleavage peptide patterns vary substantially between the different trypsins. Whereas Trypsin 6 generated a large share of non- or incompletely digested species, indicated by intense peaks at late retention times, Trypsin 1 provided the highest reproducibility.

To avoid accompanying changes in peptide patterns, for direct nano-LC-MS/MS analyses, platelet digests were analyzed without prior desalting. As illustrated in Fig. 4, to evaluate cleavage specificity, we searched LC-MS/MS data three times using different protease settings: (1) all MS/MS spectra against “Trypsin”, (2) MS/MS not-assigned with Trypsin, against “semitryptic”, (3) MS/MS spectra neither assigned with Trypsin nor semitryptic-settings, against “none” for non-specific digestion.

(1) Trypsin: Except for Trypsin 6 (with 16,224 spectra) and Trypsin 1 (with 18,592 spectra), the number of MS/MS spectra searched was ~20,000 in most cases. The number of peptide identifications after filtering varied significantly with Trypsin 1 yielding the highest amount of 7,630 identified peptides (41% of searches), followed by Trypsin 2 with 6,276 peptides (32%) whereas Trypsin...
6 yielded only 1,399 identified peptides (8%). On the protein level, when taking into account only identifications with at least two different peptides, Trypsin 1 yielded 709 proteins, followed by Trypsin 2 with 612, whereas Trypsin 1 only identified 188 proteins.

2) Semitryptic: As depicted in Fig. 4c, afore unassigned spectra show an inverse pattern after semitryptic search: Whereas Trypsin 1 now yields the lowest number of searches (5,003) in conjunction with significantly lower numbers of identified proteins and peptides around 1,200 and 200, respectively for Trypsin 4 and Trypsin 5.

3) Nonspecific: As demonstrated in Fig. 4d, again the number of searches increases from Trypsin 1 with 3,745 to Trypsin 6 with 6,789, with the number of identified peptides (15 for Trypsin 1 to 241 for Trypsin 6) as well as proteins showing the same trends (2 for Trypsin 1 and 42 for Trypsin 6).

These results strongly indicate that even when using the same conditions and the same samples for digestion, clear differences can be reproducibly detected between various trypsins with the proportion of semitryptic and non-specific peptides ranging from 4% (Trypsin 1) to 43% of the identified peptides (Trypsin 6).
Moreover, for tryptic searches, average protein and peptide scores correlate with the number of peptide identifications, as can be seen from Fig. 4b. Thus, digests providing a higher number of peptide IDs concurrently provide higher scores for those as well, clearly illustrating considerable differences in protease performance which already affect qualitative proteomic studies. In accordance with these results, clear differences in the charge state-dependant intensity distribution can be observed as illustrated in Supplemental Fig. 1.

As can be concluded from the non-tryptic cleavage sites detected in these analyses, all trypsins have chymotryptic side activity, however to highly varying degree. This can be attributed to the autoproteolytic generation of pseudotrypsin having Chymotrypsin-like activity [21], which on the other hand can be compensated by methylation as well as by addition of L-1-tosylamido-2-phenylethyl chloromethyl ketone (TPCK) as Chymotrypsin inhibitor [32].

To assess the extent of direct Chymotrypsin and other protein contamination, we separated 1 μg of Trypsins 1-6 on a 1-DE gel followed by silver staining, excised all visualized bands and analyzed them by LC-MS subsequent to in-gel digestion. However, we could only identify minor Chymotrypsin contamination in Trypsin 6, which also revealed the highest degree of chymotryptic side activity.

To further examine the extent of semitryptic cleavage, we separated all digest replicates using SCX at pH 2.7 and for each analyzed three distinct fractions by LC-MS/MS, labeled as SCX 1, SCX 2 and SCX 3 in Fig. 2. Strong differences in the UV traces around the area where singly charged peptides are expected to elute (SCX 1 and SCX 2) already indicate a significant increase in semitryptic and non-specific cleavage events for certain Trypsins. In all three fractions, for Trypsin 1 the share of fully-tryptic peptides accounts for 84% (SCX 1) to 96% (SCX 2), demonstrating a high specificity even in the SCX 1 and SCX 2 fractions enriched with singly charged peptides, whereas the values for all other trypsins are considerably lower. Yet, for Trypsin 2 values of 66% and 69%, respectively can be achieved in fractions SCX 1 and SCX 2.

In concordance with the previous results, again Trypsin 6 has the lowest performance with only 30% and 37% of fully-tryptic peptides, respectively in fractions SCX 1 and SCX 2, and only 53% even in fraction SCX 3, which usually would be assumed to be enriched in fully-tryptic peptides.

From these results it is straightforward to conclude that Trypsin 1 yields the highest number of tryptic peptides and identified proteins while concurrently yielding the lowest number of semitryptic and non-specific peptides. However, we wondered whether this might be at the expense of more extensive missed cleavages. Therefore, to assess (a) the general impact of the different trypsins on quantitative proteomic studies and (b) the extent of missed cleavages for the different trypsins, we conducted a targeted MRM screening. Hence, based on LTQ Obitrap Velos HCD data from the analyzed digests, we selected pairs of peptides with and without missed cleavages that share same sequence parts such as DFGSFDK/DFGSFDFKFK and ITIADCGQLE/KITIADCGQLE. As these represent sequences with a generally higher tendency of missed cleavage, we used them as a direct measure of cleavage efficiency as well as reproducibility (full list given in supplemental table 1). However, due to different MS response of the two corresponding partners of a pair, relative inter-pair quantitation is not feasible without the use of stable-isotope-labeled internal standard peptides [33]. We therefore conducted relative quantitation of the single peptides between all digest replicates to assess whether peptides with missed cleavages are overrepresented in certain trypsin samples. Fig. 5 summarizes the data obtained from these analyses, illustrating that Trypsin 1 yielded not only the highest signal intensities in general, but also the lowest standard deviations (24%) and the lowest share of signal intensity contributed to missed cleavage peptides (18% of the total intensity). For all trypsins evaluated in this study, 0.1-0.2% of the identified peptides derived from K-P and R-P bond cleavages. However we refrained from SRM-based quantitation, since the corresponding peptide

Fig. 5 – Summary of MRM results. Here, specific sequences which were detected in the platelet digests with and without missed cleavage, such as DFGSFDK and DFGSFDFKFK were utilized to set up a specific MRM method. The overall sizes of the circles reflect the median areas for all monitored transitions within the respective digest replicates, comprising the sum of peptide areas for all peptides with missed cleavage and peptides without missed cleavage. Thus, Trypsin 1 generally yields the highest MRM signals, the lowest standard deviation of 24%, whereas 82% of the total area is attributed to peptides without missed cleavages, corresponding to the overall best specificity, reproducibility as well as cleavage efficiency.
sequences were identified in many different variants including missed cleavages and oxidized methionines rendering relative quantitation unfocused.

4. Discussion and conclusion

Reproducible, efficient and specific digestion of protein samples is among the key factors for successful quantitation in proteomics. Especially in complex samples such as cell lysates, slight changes in cleavage performance can dramatically alter peptide patterns. Hence, non-specific or inefficient cleavage can substantially increase the number of peptide features while concurrently reducing their average abundances. Thus, not only will identification and consequently quantitation rates suffer, but this will also lead, for instance, to complications in the alignment of peptide maps in label-free experiments.

In our group, for many years we successfully utilized an inexpensive trypsin for the digestion of large sample amounts or for proof-of-principle experiments. However, as could be seen from some of our established digest control workflows also described here, at once a new stock of this trypsin reproducibly generated lower amounts of fully tryptic peptides while concurrently increasing the proportion of semitryptic and non-specific peptides. This dramatic reduction in performance stimulated us to conduct a systematic comparison of different trypsins, qualitatively and quantitatively evaluating the occurrence of semitryptic, non-specific as well as missed cleavage events in a complex digest of human platelets.

In summary, our results are in accordance with Picotti et al. [19] who systematically analyzed a tryptic digest of a single protein, β-lactoglobulin and reported that fully-tryptic peptides accounted for ~75% of the total peptide intensity while many low intensity peptide signals could be attributed to non-tryptic cleavage events; all trypsins evaluated here generated a substantial proportion of not fully-tryptic peptides. In 2004 Olsen et al. reported a specificity of ~100% for trypsin analyzing the mouse liver proteome [18] – however sample complexity, MS duty cycle and dynamic range in conjunction with data-dependent acquisition might have lead to undersampling an consequently slight overestimation of enzyme specificity based on the generally higher intensities of fully-tryptic events, as also reported by Picotti et al.

In the present study, LC-MS/MS analyses of complete digests as well as SCX fractions enriched in semitryptic peptides revealed considerable differences in protease performance which do not fully correlate with price: The most expensive trypsin evaluated in this study was only 2nd best in performance. In all experiments, Sequencing Grade Promega Trypsin showed the best performance, however due to its very low price Sigma T-1426 (~1/1,000th of the cost of Promega per mg of enzyme) might be an alternative for conducting preliminary proof-of-principle experiments or for teaching purposes, whereas Sigma 93614 revealed a consistently poor performance and should not be used for any type of proteomics experiment.

Using stable isotope-labeled internal standard peptides, Proc et al. [24] clearly demonstrated that for specific peptides the final yield strongly depends on digest conditions and that for MRM-based assays of selected peptides this can have an enormous impact on detection limits and quantitation. Using our approach, even without stable isotope-labeled peptides, digest performance can be readily monitored without much effort.

Independent of the results presented here, we in general strongly recommend the following procedures as routine workflows for digestion and trypsin control, as they will substantially help to improve the quality as well as the coverage of quantitative studies.

1. Monolithic column-based separations as routine digest quality control as it is a robust, sensitive, fast and reproducible method, which requires only small amounts of sample (~0.1 μg) and furthermore is compatible with a large variety of substances that nevertheless might interfere with typical LC-MS/MS analyses.

2. Targeted LC-MS/MS analyses of SCX-enriched semitryptic fractions in conjunction with the three-step database search strategy, to evaluate cleavage specificity.

3. Targeted MRM analyses of selected peptide pairs to assess the degree and also reproducibility of missed-cleavage occurrence.

5. Conflict of interest statement

The authors have declared no conflict of interest.

Supporting information

All mgf files of the nano-LC-MS/MS analyses are publicly available via the PRIDE repository (http://www.ebi.ac.uk/pride/), accession number 18826.

Supplementary materials related to this article can be found online at doi:10.1016/j.jprot.2011.11.016.

Acknowledgment

The financial support by the Ministerium für Innovation, Wissenschaft und Forschung des Landes Nordrhein-Westfalen and by the Bundesministerium für Bildung und Forschung (MedSys Project SARA, 31PS800) is gratefully acknowledged.

References


