

1 **Application of the half decimal place rule to increase the peptide identification rate**

2

3 Christian J Koehler, Ravi Chand Bollineni and Bernd Thiede\*

4

5 Department of Biosciences, University of Oslo, Oslo, Norway

6

7 **Keywords:** data-dependent acquisition, half decimal place rule, mass defect, mass  
8 spectrometry, peptide spectrum matches

9

10 \*To whom the correspondence should be addressed:

11 Bernd Thiede, Department of Biosciences, University of Oslo, P.O. Box 1066 Blindern, 0316  
12 Oslo, Norway, Tel.: +47-22840533; Fax +47-22840501; E-mail: bernd.thiede@ibv.uio.no

13

14 **Abbreviations:** DDA, data-dependent acquisition; HDPR, half decimal place rule; LC, liquid  
15 chromatography; MS, mass spectrometry; PSM, peptide spectrum match; UHPLC, ultra high  
16 performance liquid chromatography

17

18

19

1 **Abstract**

2 RATIONALE: Many MS2 spectra in bottom-up proteomics experiments remain unassigned.  
3 To improve proteome coverage, we applied the half decimal place rule (HDPR) to remove  
4 non-peptidic molecules. The HDPR considers the ratio of the digits after the decimal point to  
5 the full molecular mass and results in a relatively small permitted mass window for most  
6 peptides.

7

8 METHODS: First, the HDPR mass filter was calculated for the human and other proteomes.  
9 Subsequently, the HDPR was applied to three technical replicates of an in-solution tryptic  
10 digest of HeLa cells which were analysed by LC-MS using a quadrupole-orbitrap mass  
11 spectrometer (Q Exactive). In addition, the same sample was analysed three times with a  
12 fixed exclusion list. The exclusion list was based on only choosing doubly charged ions for  
13 fragmentation.

14

15 RESULTS: The peptide spectrum match (PSM) rate increased by 2-4% applying HDPR  
16 filters from 0.1-0.25 Da and 75-150 ppm, respectively. Excluding all MS2 events by applying  
17 an HDPR filter of doubly charged ions, we were able to improve PSMs by 0.9 % and the  
18 PSM rate by 2.5%.

19

20 CONCLUSIONS: An algorithm to filter precursors based on the HDPR was established to  
21 improve the targeting of the acquisition of MS2 spectra in data-dependent acquisition (DDA)  
22 experiments. According to our data, a total gain of PSMs of 1-5% might be achievable if the  
23 HDPR filter would already be applied during MS data acquisition.

## 1 Introduction

2 Data-dependent acquisition (DDA) is an operation mode in mass spectrometry predominantly  
3 used in discovery proteomics where algorithms decide which precursor is subjected to further  
4 fragmentation and MS2 acquisition. The target is typically not known and the aim is to  
5 acquire spectra from as many different peptides as possible to obtain a map of the sample  
6 that is as complete as possible. Typically, the DDA algorithm ranks peaks in an MS spectrum  
7 by intensity. Dynamic exclusion of already taken precursors is used to maximise the peptide  
8 identification rate. Furthermore, additional rules can be specified to include or exclude  
9 possible precursors based on the charge state, minimum intensity and mass range of the  
10 precursor. Nevertheless, undersampling of complex proteomic samples remains one of the  
11 major issues in DDA-based proteomics. Furthermore, a significant proportion of MS2 spectra  
12 in bottom-up proteomics experiments remain unassigned despite significant improvements of  
13 acquisition speed and resolution of mass spectrometers during the last years. These  
14 unexplained MS2 spectra might originate from e.g., unexpected posttranslational  
15 modifications, chemical modifications, sequencing errors, unknown alternative splice variants  
16 and polymorphisms <sup>[1]</sup>. Another reason might be that the analysed molecules are not  
17 peptides and thus worthless for proteome analyses. Actually, it was observed that up to 30%  
18 of detected molecules can be of nonpeptidic origin <sup>[2]</sup>.

19 The mass defect is the difference in mass between an atom and the sum of the masses of  
20 the protons and neutrons. Each isotope has a unique mass defect with carbon-12 (zero mass  
21 defect), hydrogen-1 (0.0078), nitrogen-14 (0.0031), oxygen-16 (-0.0051) and sulfur-32 (-  
22 0.0279) <sup>[3]</sup>. The mass distribution of the theoretically possibly tryptic peptides revealed gaps  
23 and low-populated areas <sup>[4] [5] [6] [7] [8]</sup>. Furthermore, the statistical occurrences of average  
24 amino acids with  $C_{4.9384}H_{7.7583}N_{1.3577}O_{1.4773}S_{0.0417}$  <sup>[9]</sup>, results in a slightly increased mass defect  
25 with increased molecular mass. It has been demonstrated that the mass defect can be used  
26 to improve identification of proteins by peptide mass fingerprinting using MALDI-TOF <sup>[10] [11]</sup>  
27 and MALDI-FTICR <sup>[12] [13]</sup> excluding non-peptide signals. A model of peptide mass cluster  
28 centres was developed, considering amino acid frequencies, average length of proteins in  
29 the database, and cleavage specificity of endoproteinases, and the distance to the nearest  
30 cluster was used to calibrate peptide peak lists and to identify non-peptide peaks <sup>[14]</sup>.  
31 Furthermore, it can be used to distinguish modified from unmodified peptides as shown for  
32 phosphopeptides <sup>[15]</sup>, glycopeptides <sup>[16]</sup> and oligonucleotide-peptide cross-links <sup>[17]</sup>. A non-  
33 linear Random Forest classification and a discrete mapping approach were developed to  
34 increase fragmentation efficiency on-the fly for specific subsets of analytes such as shown  
35 for phosphorylated and iodinated peptides <sup>[18]</sup>. Moreover, the HDPR was defined to take  
36 advantage of the mass defect <sup>[10]</sup>. It is based on the observation that the first digit of the

1 mantissa of the mass of a peptide is near the half of the first digit of mass values between  
2 500 and 999 and near the half of the first two digits of mass values between 1000 and 1999,  
3 and near the half of the first digit of the peptide mass in the mass range 2000 to 3000 <sup>[10]</sup> <sup>[11]</sup>.  
4 This means that for peptides with molecular masses between 100 and 999 Da, the mantissa  
5 is typically between 0.05 (at 100 Da) and 0.5 (at 999 Da), and for peptides with molecular  
6 masses between 1000 and 1999 Da, the mantissa is typically between 0.5 and 0.99. The  
7 main application of the HDPR is so far on peptide mass fingerprints acquired by MALDI-MS  
8 to filter chemical contaminants and to quickly evaluate instrument calibration.  
9 In this report, we performed a new calculation of the HDPR. The HDPR was applied to a  
10 complex proteome derived from HeLa cells analysed by LC-MS to select precursors  
11 subjected to MS2 acquisition with the aim to increase the peptide spectrum identification rate.  
12

## 13 **Materials and Methods**

### 14 ***Cell culture***

15 HeLa cells were grown as a monolayer in RPMI-1640 supplemented with 10 % fetal bovine  
16 serum and maintained in a humid incubator at 37°C in a 5 % CO<sub>2</sub> environment to reach a  
17 density of 1x10<sup>5</sup> cells/ml. Cell pellets were frozen and stored in liquid nitrogen. Pellets of cells  
18 were thawed on ice and 800 µl SILAC Phosphoprotein lysis buffer B (Invitrogen, Oslo,  
19 Norway) was added. The cell slurry was homogenized with a pestle (20x) for mechanical  
20 breakage of the cells followed by sonication using an Ultrasonic processor (UP400s, Dr.  
21 Hielscher). Samples were centrifuged at 16,000 g for 20 minutes at 4°C in a Heraeus Biofuge  
22 pico (Kendro, Hanau, Germany) and the supernatant was aliquoted in 40 µl aliquots.  
23  
24

### 25 ***In-solution digestion***

26 To one HeLa cell aliquot, 400 µl of ice cold acetone/methanol (1:1) acidified with 0.1 %  
27 hydrochloric acid (v/v) was added, vortexed and precipitated at -20°C over night. Samples  
28 were centrifuged at 16,000 g for 20 minutes at 4°C (Heraeus Biofuge pico) and the  
29 supernatant was discarded. Proteins were re-dissolved in 50 µl 6M urea and 100 mM  
30 ammonium bicarbonate, pH 7.8. For reduction and alkylation of cysteines, 5 µl of 200 mM  
31 DTT was added and the samples were incubated at 37°C for 1 h followed by addition of 20 µl  
32 200 mM iodoacetamide for 1 h at room temperature in the dark. The alkylation reaction was  
33 quenched by adding 20 µl 200 mM DTT. The proteins were digested with trypsin in a final  
34 volume of 225 µl for 16 h at 37°C. The digestion was stopped by adding 100 µl 1 % formic  
35 acid and the generated peptides were purified using a Strata C18-E SPE column

1 (Phenomenex, Værløse, Denmark), and dried using a Speed Vac concentrator (Eppendorf,  
2 Hamburg, Germany).

3

#### 4 **Liquid chromatography-mass spectrometry (LC-MS)**

5 Three technical replicates tryptic digest of the HeLa proteome were analyzed using an  
6 Ultimate 3000 nano-UHPLC system (Dionex, Sunnyvale, CA, USA) connected to a Q  
7 Exactive mass spectrometer (ThermoElectron, Bremen, Germany) equipped with a nano  
8 electrospray ion source. For liquid chromatography separation, an Acclaim PepMap 100  
9 column (C18, 3  $\mu\text{m}$  beads, 100  $\text{\AA}$ , 75  $\mu\text{m}$  inner diameter) (Dionex, Sunnyvale CA, USA)  
10 capillary of 50 cm bed length was used. A flow rate of 300 nL/min was employed with a  
11 solvent gradient 3-35% B in 97 min, to 50% B in 13 min and then to 80% B in 2 min. Solvent  
12 A was 0.1% formic acid and solvent B was 0.1% formic acid/90% acetonitrile.

13 The mass spectrometer was operated in the data-dependent mode to automatically switch  
14 between MS and MS2 acquisition. Survey full scan MS spectra (from m/z 400 to 2000) were  
15 acquired with the resolution  $R = 70,000$  at m/z 200, after accumulation to a target of  $1e6$ . The  
16 maximum allowed ion accumulation times were 100 ms. The method used allowed  
17 sequential isolation of up to the ten most intense, doubly charged ions, depending on signal  
18 intensity (intensity threshold  $1.7e4$ ), for fragmentation using higher-energy collision induced  
19 dissociation (HCD) at a target value of 10,000 charges and a resolution  $R = 17,500$ . Only  
20 double charged precursors were considered for MS2 to assure comparability between  
21 different experiments. The peptide match filter was set to off. Target ions already selected for  
22 MS/MS were dynamically excluded for 30 sec. The isolation window was m/z = 2 without  
23 offset. The maximum allowed ion accumulation for the MS/MS spectrum was 60 ms. For  
24 accurate mass measurements, the lock mass option was enabled in MS mode and the  
25 polydimethylcyclosiloxane ions generated in the electrospray process from ambient air were  
26 used for internal recalibration during the analysis.

27

#### 28 ***Data analysis***

29 Data were acquired using Xcalibur v2.5.5 and raw files were processed to generate peak list  
30 in Mascot generic format (\*.mgf) using ProteoWizard release version 3.0.7230. Database  
31 searches were performed using Mascot in-house version 2.4.0 to search the SwissProt  
32 database (Human, 21.01.2016, 20187 proteins) assuming the digestion enzyme trypsin, at  
33 maximum one missed cleavage site, fragment ion mass tolerance of 0.05 Da, parent ion  
34 tolerance of 10 ppm and oxidation of methionines, and acetylation of the protein N-terminus  
35 as variable modifications. Scaffold (version Scaffold\_4.4.3, Proteome Software Inc., Portland,  
36 OR) was used to validate MS2 based peptide and protein identifications. Peptide

1 identifications were accepted if they could be established at greater than 95.0% probability  
2 by the Scaffold Local FDR algorithm. Protein identifications were accepted if they could be  
3 established at greater than 99.9% probability. The mass spectrometry proteomics data have  
4 been deposited to the ProteomeXchange Consortium <sup>[19]</sup> via the PRIDE partner repository  
5 with the dataset identifier PXD004900 (Username: reviewer91076@ebi.ac.uk; password:  
6 0EdGEubA).

7

### 8 **Calculation of the HDPR ratio**

9 Based on the descriptive observation of the HDPR <sup>[9]</sup>, we inferred a continuous function (1.1)  
10 which describes the HDPR over the whole mass range. From the initial description, a  
11 denominator of 2000 was found to match best the described phenomena.

12

$$13 \quad \text{calculated mantissa} = \left( \frac{\text{peptide mass}}{2000} \right) - \text{floor} \left( \frac{\text{peptide mass}}{2000} \right) \quad (1.1)$$

14

15 However, the denominator does not need to be exactly 2000. Therefore a more general  
16 description of the equation was formulated (1.2).

17

$$18 \quad \text{calculated mantissa} = \left( \frac{\text{peptide mass}}{\text{denominator}} \right) - \text{floor} \left( \frac{\text{peptide mass}}{\text{denominator}} \right) \quad (1.2)$$

19

20 The difference between the calculated mantissa and the measured mantissa was determined  
21 to be used as a qualifier for peptides. Notably, the difference between 1999.9 and 2000.1 is  
22 0.2, however formula 1.1 calculates the mantissa only and has no information about the  
23 number before the digit. Thus, the difference in this example only using the mantissae would  
24 result in 0.8 (difference of 0.9 and 0.1) (Fig. 1A and 1B, top). To overcome this issue, a  
25 minimal distance of two mantissae (disregarding a carry) needs to be calculated, resulting in  
26 a positive number between 0 and 0.5. The absolute value of the minimal distance between  
27 two mantissae (x, y) can be calculated as:

28

$$29 \quad \text{distance}(x, y) = 0.5 - ||x - y| - 0.5| \quad (1.3)$$

30

31 This formula was used to calculate the minimal distance between two mantissae by  
32 disregarding the floor units (units before the decimal) (Fig. 1A and 1B, bottom).

33

## 34 **Results and Discussion**

35

### 1 **Calculation of the denominator for the HDPR mass filter of the human proteome**

2 An *in silico* tryptic digest without missed cleavages of the reviewed human Swiss-Prot  
3 proteome database was performed. From the resulting set of all human tryptic peptides, a  
4 subset was created using only peptides with a minimum number of five amino acids per  
5 peptide and all peptides smaller than 600 Da or larger than 4,000 Da were disregarded.  
6 Assuming a minimum charge of 2+, this subset contains 639,699 of useful detectable  
7 peptides in a mass spectrometry experiment utilizing a 300 - 2,000 m/z scan range. All  
8 further calculations were based on this set of human tryptic peptides. Notably, certain  
9 posttranslational modifications (PTMs), such as glycosylation and phosphorylation lead to a  
10 significant change of the HDPR ratio in comparison to corresponding non-modified peptides.  
11 For PTM-specific experiments, the HDPR filter might be changed accordingly to improve the  
12 selection of modified peptides for fragmentation.

13 Every individual peptide has its own ideal denominator where the distance between the  
14 calculated mantissa and the real mantissa becomes zero. For a proteomics experiment, the  
15 best denominator for the HDPR rule is where the average of all peptides share a minimal  
16 distance. Therefore, we calculated the denominator for which the difference between the  
17 calculated and actual mantissa is minimized for each peptide of the human tryptic peptidome.  
18 Because every peptide mass has several minima (Fig. 1A and 1B), the value nearest to the  
19 original, previous observed denominator of 2000 was used as the optimal denominator. The  
20 average of all these denominators of the whole human tryptic peptidome was 2007 without  
21 and 2011 with carbamidomethylation (Fig. 2A). Other organisms revealed similar values  
22 using the SwissProt database, e.g., 2010 for mouse (533,985 tryptic peptides), 2005 for  
23 *Arabidopsis thaliana* (374,786 tryptic peptides), 2008 for rat (227,089 tryptic peptides), 1991  
24 *Schizosaccharomyces pombe* (139,510 tryptic peptides), 2005 for bovine (139,188 tryptic  
25 peptides), 1990 for *Dictyostelium discoideum* (122,496 tryptic peptides), 2009 for *Drosophila*  
26 *melanogaster* (110,397 tryptic peptides), and 2019 for *Caenorhabditis elegans* (107,559  
27 tryptic peptides).

28

### 29 **Application of the HDPR filter after data acquisition**

30 To assess the HDPR algorithm, three technical replicates of a complex proteome derived  
31 from an in-solution digest of HeLa cells were analysed using two hours LC runs connected to  
32 a quadrupole orbitrap mass spectrometer (Q Exactive). After conventional processing, the  
33 resulting peak list (mgf format) was further processed with an in-house python script filtering  
34 all MS2 spectra based on the HDPR rule, resulting in peak lists which contained only  
35 peptides in which the distance of the true and the theoretical mantissa were below the  
36 defined threshold. In this way, we simulated if a MS2 spectrum would have been acquired

1 (passed the HDPR) or rejected. The unfiltered and HDPR filtered mgf files were searched  
2 using Mascot. This approach consequently results in a reduction of searchable MS2 spectra  
3 and therefore the loss of peptide spectrum matches (PSMs). However, the relation of the  
4 filtered out not-matching and matching spectra will give a measure of the quality of the HDPR  
5 algorithm. We compared different filter thresholds, both fixed mass filters allowing a fixed  
6 distance between the mantissae (0.1 Da, 0.15 Da, 0.2 Da and 0.25 Da) as well as  
7 proportional mass filters with a distance based on the acquired peptide mass (75 ppm, 100  
8 ppm, 125 ppm, 150 ppm) (Table 1).

9 With more restrictive filters, the identified number of proteins and number of PSMs  
10 decreased (3,111 to 3,006 for Da and 3,109 to 3,040 for ppm, respectively), as less MS2  
11 spectra were available. However, the identification rate (PSMs/MS2 spectra (%)) increased  
12 from 51.3% to 52.7% for Da and 50.8% to 52.5% for ppm, respectively (Table 1B). Notably, it  
13 was always higher than in the original dataset with 48.7% (Table 1A).

14 Considering only the rejected data (Total - HDPR), a distinct lower identification rate  
15 (PSMs/MS2 spectra (%); 4.2% - 34.3% for Da and 5.6% - 28.7% for ppm) than for the  
16 unfiltered data (48.7%) (Table 1A) was obtained. This result shows that more non-matching  
17 MS2 spectra were removed in comparison to the unfiltered data set.

18 Based on the identification rate for each filter, a maximal gain of PSMs can be calculated  
19 (Table 1D). This gain is a theoretical calculation and based on the assumption that acquired  
20 precursors which did not pass the HDPR would be rejected and replaced with the next  
21 precursor in the DDA algorithm which passes the HDPR. To calculate the maximal gain of  
22 PSMs, precursors which fail the HDPR but result in a PSM need to be subtracted.

23 As an example, using an HDPR mass filter of 0.15 Da, 33,519 precursors remained which  
24 led to 17,535 PSMs (52.3% identification rate) (Table 1B). As a consequence, 3,795 MS2  
25 spectra were discarded, of which 641 MS2 spectra led to PSMs (Table 1C). Hence, the filter  
26 incurred the loss of 83.1% of non-usable precursors and of 16.9% of peptide precursors  
27 (1.7% of all MS2 spectra) (Table 1C). In case of an MS instrument operating with an HDPR  
28 filter, all discarded MS2 spectra would give time to acquire MS2 spectra with an increased  
29 chance of peptide identification. If the 3,795 filtered precursors with HDPR of 0.15 Da would  
30 have been used to acquire MS2 spectra with an identification rate of 52.3%, theoretically  
31 1,985 MS2 spectra would be identified. Subtracting the 641 peptide identifications which  
32 were discarded by the algorithm, a theoretical total gain of 1,345 additional peptides can be  
33 obtained with HDPR filter 0.15 Da, which corresponds to a gain of 7.4% of total PSMs (Table  
34 1D). However, the theoretical gain is a too high estimation as less intense precursors will be  
35 chosen for MS2 fragmentation (top10 + n peaks) and not all MS2 spectra which led initially to  
36 an identification event will be replaced by a PSM. The more stringent the HDPR filter is



1 applied, the more precursors are disregarded and the chance to supplement these free with  
2 a peptide matching HDPR precursor becomes increasingly difficult. A more relaxed filter on  
3 the other hand leads to a smaller gain. The optimal values have to be determined  
4 experimentally and depend highly on the sample complexity and duty cycle.

5 Apart from the experimental determination of the optimal HDPR filter stringency, a theoretical  
6 approach of a useful threshold was considered. Using all theoretical peptides from the  
7 human database, we calculated the threshold where the distance of theoretical and true  
8 mantissae includes 95% of the human tryptic peptides using the determined optimal HDPR  
9 ratio of 2011 (Fig. 2B). As a result, a distance of  $\pm 0.1497$  Da for the fixed mass filter and  
10 102 ppm for the proportional mass filter, respectively, were calculated. A total deviation of  $\pm$   
11 0.15 Da and  $\pm 100$  ppm, respectively, was found to be a suitable approximation (Fig. 3A  
12 and 3B).

13 The use of a proportional mass filter would be dependent on the mass of each peptide in  
14 contrast to the fixed mass filter. Using 100 ppm, on average a theoretical total gain of 1,241  
15 peptide matches (6.8% of total peptide identifications) was calculated, consistent for all three  
16 replicates (Table 1D). Using 100 ppm, the filtering effects were slightly more moderate than  
17 using the fixed mass filter of 0.15 Da. While the fixed mass filter is biased towards small  
18 masses being more stringent to high masses, the reverse is true for the proportional mass  
19 filter (Fig. 3C, 3D, 3E). Covering approximately 95% of the tryptic proteome, a fixed mass  
20 filter of 0.15 Da resulted in a higher gain in PSMs than a 100 ppm proportional mass filter. A  
21 probable explanation is the discrepancy between the theoretical calculation and the  
22 experimental higher abundance of low mass peptides of the peptides being available for MS  
23 and the just noted preference.

24 Notably, the ideal tolerance is dependent on the peptide mass, with a higher spread for  
25 higher masses <sup>[10]</sup> (Fig. 3B). Based on the HDPR ratio 2011, the tolerable deviation can be  
26 narrowed to  $\pm 0.1323$  Da to cover 95% of the human tryptic peptidome if only masses less  
27 than 3,000 Da are considered. For peptides less than 2,000 Da, even a deviation of 0.1130  
28 Da would be enough whereas a deviation of  $\pm 0.2736$  Da is required to include 95% of the  
29 peptides between 3,000 Da and 4,000 Da.

30

### 31 ***Application of the HDPR filter during data acquisition***

32 We were not able to change the acquisition software of our mass spectrometer to implement  
33 an HDPR filter. However, we applied an exclusion list to demonstrate the feasibility of this  
34 filter. Due to the linear dependency of the mantissa from the mass defined by the HDPR rule  
35 and a fixed tolerance of e.g.,  $\pm 0.15$  Da (Figure 3A), it is possible to create an exclusion list  
36 which contains the masses between consecutive HDPR masses. These masses have a fixed

1 tolerance which fills the gap between accepted mass areas (Figure 4). Thus, exclusion lists  
2 can be defined with an absolute tolerance. As the tolerance depends on the charge state, it  
3 is only possible to create an exclusion list for one charge state at a time and for the linear  
4 mass filter only. For this reason, all experiments had been carried out allowing only doubly  
5 charged precursors to be considered for MS2 fragmentation.

6 First, we implemented an exclusion list based on the HDPR tolerance +/- 0.15 Da, acquired  
7 three technical replicates and compared them against three replicates acquired without  
8 exclusion list and were able to improve the PSM rate, but the absolute number of PSMs was  
9 not improved (Table 1E). Therefore, we used a more relaxed exclusion list based on the  
10 HDPR tolerance of +/- 0.20 Da and repeated the experiment. Without exclusion, on average  
11 37,315 MS2 spectra were acquired resulting in 18,034, 18,146 and 18,346 PSMs per  
12 replicate (on average  $18,175 \pm 129$ ) (Table 1A). Using the exclusion list, on average 35,782  
13 MS2 spectra were acquired leading to 18,329, 18,373, and 18,413 PSMs per replicate (on  
14 average  $18,342 \pm 74$ ). The identification rate increased by 2.6% (51.3% vs. 48.7%) because  
15 1,533 less MS2 spectra were acquired but 167 more PSMs were identified (Table 1E). The  
16 experimental gain appears to be moderate compared to the theoretical gain with only 0.9%  
17 increase in total PSMs. A similar gain in protein identification has been achieved with peptide  
18 mass fingerprinting using MALDI-TOF instrumentation <sup>[14]</sup>. Still, a comparison is difficult  
19 because removing non-peptide peaks from peptide mass fingerprints is performed post-  
20 acquisition and applied to single proteins. Anyhow, we were able to increase the total PSMs  
21 simply by the application of the HDPR implemented as a fixed exclusion list. Nevertheless,  
22 we could only compare acquisitions based on one charge state which limits the number of  
23 available precursors. In the experiment using the exclusion list, the frequency of recording of  
24 ten consecutive MS2 spectra was decreased compared to normal acquisitions. Furthermore,  
25 we could only compare acquisitions with the filter turned on (exclusion list) or off (normal  
26 acquisition). If it had been possible to apply the filter as long as there are precursors to  
27 choose from and otherwise continue with less favourable precursors, the PSMs would have  
28 been increased more without negative consequences. Ultimately, the benefit of the HDPR  
29 filter in a real-time decision tree is difficult to predict without an actual implementation.

30 If the HDPR filter was be implemented into the acquisition software, it must observe m/z  
31 values, determine the charge, and calculate the actual mass, all is typically included into  
32 current mass spectrometry software. Apart from that, a lookup table of allowed and  
33 disallowed mass regions could be calculated based on the parameters "denominator" and  
34 the tolerance in Da or ppm.

35

1 **Conclusions**

2 We developed a simple algorithm to filter precursors based on the idea of HDPR in order to  
3 improve the targeting of the acquisition of MS2 spectra in DDA experiments and to increase  
4 the number of peptide identifications amongst the chosen precursors. We were not able to  
5 implement the filter into the acquisition parameters of our mass spectrometer, but have  
6 shown the feasibility by an exclusion list applied to the same samples. According to our data,  
7 a total gain of PSMs of 1-5% might be achievable by including HDPR-based filtering in MS  
8 acquisition for standard peptide identification. This gain would just require changes in the MS  
9 acquisition software. However, the HDPR-based filter should be optional because it might not  
10 be suitable for the analysis of certain modified peptides which lead to relative large variations  
11 of the HDPR denominator such as phosphorylated, glycosylated, and lipidated peptides.  
12 Finally, the HDPR ratio must be determined empirically, because PSMs of precursors which  
13 do not fulfil the HDPR filter must be replaced with low intensity MS2 spectra, which cannot be  
14 theoretically predicted.

15

16 **Acknowledgements:** Financial support from the Norwegian Cancer Society (Project  
17 4514636) is gratefully acknowledged.

18

19

1 **Figure legends**

2

3 **Figure 1: Graphical representation of the determination of HDPR ratios.**

4 Graphical representations of the iterative calculations of two peptides with different masses  
5 ( $m/z$  815.4137 and  $m/z$  2442.2763) using formula 1.2 are displayed in A and B, respectively.  
6 The denominator is displayed on the x-axis, and the mantissa (top) or distance (bottom) is  
7 shown on the y-axis. The real decimal place of the peptide masses is graphed as a  
8 horizontal dashed line. The higher the mass the more often the calculated decimal place  
9 traverses the real decimal place (top, A vs. B). The distance between the calculated and real  
10 decimal place becomes zero whenever the function traverses the true decimal place (A and  
11 B, bottom). To calculate the HDPR ratio, the denominator creating a minimal distance to the  
12 real decimal place and being closest to 2000 was calculated. For the two presented  
13 examples, these values are indicated by arrows and resulted in HDPR ratios of 1971 (A) and  
14 1913 (B).

15

16 **Figure 2: Histogram of the frequency of HDPR ratios and distances of the calculated**  
17 **mantissae.**

18 The HDPR ratio was determined for all human tryptic peptides and the distribution of the  
19 HDPR ratios is shown in A. The resulting average HDPR ratio was 2011. 95% of all tryptic  
20 peptides with a minimum of five amino acids, and 600 to 4000 Da are within a tolerance of  
21 0.1497 Da (light grey) using this HDPR ratio (B).

22

23 **Figure 3: The mantissae and frequency of the human tryptic peptides in dependence**  
24 **of molecular mass.**

25 The mantissae vs. masses are shown. The greyscale of the dots indicates the frequency.  
26 The HDPR filter of 0.15 Da (A) and 100 ppm (B), respectively, are shown by black lines. The  
27 frequency of all peptides (C), peptides deviating more than 0.15 Da (D), and peptides  
28 deviating more than 100 ppm (E) are depicted using HDPR ratio 2011. A bias towards small  
29 masses can be observed for the proportional mass filter (ppm) and towards high masses for  
30 the fixed filter (Da).

31

32 **Figure 4: Implementation of the linear HDPR filter by using an exclusion list.**

33 HDPR masses can be calculated based on equation 1.2. In a linear HDPR filter, HDPR  
34 masses with their tolerances create accepted mass areas which match the HDPR (bright  
35 area). Consequently, the areas in between are non-accepted mass areas with a central mass  
36 defined in the exclusion list to implement the HDPR filter (dark gray). Here, uncharged

1 masses are displayed. For a filter, the uncharged masses and the tolerances have to be  
2 converted in  $m/z$  values for every charge. Therefore, the tolerances vary with charge and  
3 only one charge state can be defined in an exclusion list to create a HDPR filter.

4

5

6

1 **Table 1: The effect of the HDPR filter**

2

3 Average numbers of three technical replicates are shown. The results obtained without  
 4 HDPR filter (A), and with different HDPR filters (B) are presented. In C, the difference  
 5 between A and B is displayed. Furthermore, the maximal theoretical gain in PSMs is shown  
 6 in D. The data obtained after using an exclusion list for all doubly charged ions is shown in E.  
 7 Protein ID, number of identified proteins; ID, number of identifications; Max. gain = MS2  
 8 spectra (Total - HDPR) x PSMs/MS2 spectra (HDPR) - PSMs (Total - HDPR).  
 9

	HDPR filter	Protein ID	MS2 spectra (%)	PSMs (%)	PSM/MS2 spectra [%]
<b>A. Total</b>	no	3180	37315	18175	48.7
<b>B. HDPR</b>	0.10 Da	3006	29189	15391	52.7
	0.15 Da	3088	33519	17535	52.3
	0.20 Da	3108	34768	17999	51.8
	0.25 Da	3111	35230	18089	51.3
	75 ppm	3040	31339	16458	52.5
	100 ppm	3088	33763	17568	52.0
	125 ppm	3099	34972	17946	51.3
	150 ppm	3109	35579	18078	50.8
<b>C. Total - HDPR (%)</b>	0.10 Da	-174	8126 (21.8)	2784 (15.3)	34.3
	0.15 Da	-92	3796 (10.2)	640 (3.5)	16.9
	0.20 Da	-72	2547 (6.8)	176 (1.0)	6.9
	0.25 Da	-69	2085 (5.6)	86 (0.5)	4.2
	75 ppm	-140	5976 (16.0)	1717 (9.4)	28.7
	100 ppm	-92	3552 (9.5)	607 (3.3)	17.1
	125 ppm	-81	2343 (6.3)	229 (1.3)	9.8
	150 ppm	-71	1736 (4.7)	97 (0.5)	5.6
<b>D. Max.gain</b>	0.10 Da			1501 (8.3)	
	0.15 Da			1345 (7.4)	
	0.20 Da			1142 (6.3)	
	0.25 Da			984 (5.4)	
	75 ppm			1421 (7.8)	
	100 ppm			1241 (6.8)	
	125 ppm			973 (5.4)	
	150 ppm			785 (4.3)	
<b>E. Total with exclusion</b>	0.15 Da	3172	35224	18022	51.2
	0.20 Da	3176	35782	18342	51.3

10

11

12

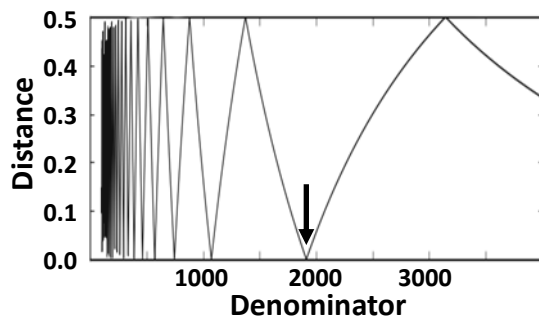
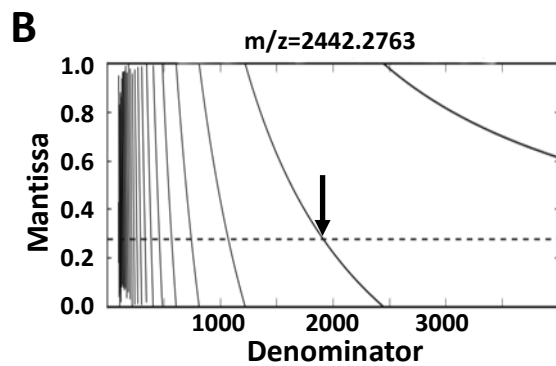
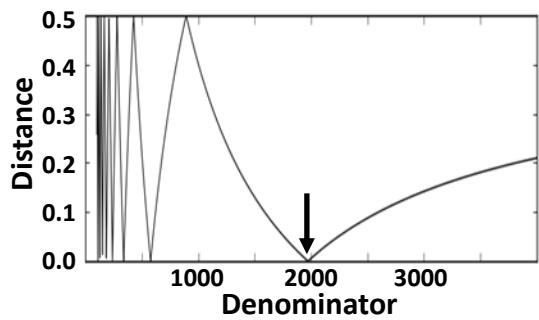
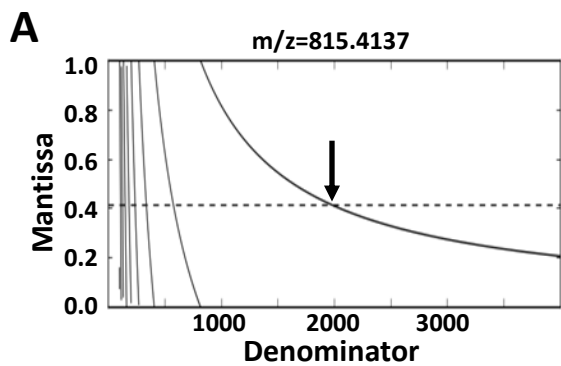
13

## 1 References

- 2 [1] J. M. Chick, D. Kolippakkam, D. P. Nusinow, B. Zhai, R. Rad, E. L. Huttlin, S. P. Gygi,  
3 *Nat Biotechnol* **2015**, 33, 743.
- 4 [2] A. Michalski, J. Cox, M. Mann, *J Proteome Res* **2011**, 10, 1785.
- 5 [3] L. Sleno, *J Mass Spectrom* **2012**, 47, 226.
- 6 [4] M. Mann, in *Annual Conference on Mass Spectrometry and Allied Topic, American*  
7 *Society of Mass Spectrometry: Atlanta, GA, 1995*.
- 8 [5] R. A. Zubarev, P. Håkansson, B. Sundqvist, *Anal Chem* **1996**, 68, 4060.
- 9 [6] S. Gay, P. A. Binz, D. F. Hochstrasser, R. D. Appel, *Electrophoresis* **1999**, 20, 3527.
- 10 [7] A. V. Nefedov, I. Mitra, A. R. Brasier, R. G. Sadygov, *J Proteome Res* **2011**, 10, 4150.
- 11 [8] I. Mitra, A. V. Nefedov, A. R. Brasier, R. G. Sadygov, *Anal Chem* **2012**, 84, 3026.
- 12 [9] M. W. Senko, S. C. Beu, F. W. McLafferty, *J Am Soc Mass Spectrom* **1995**, 6, 229.
- 13 [10] F. Schmidt, M. Schmid, P. R. Jungblut, J. Mattow, A. Facius, K. P. Pleissner, *J Am*  
14 *Soc Mass Spectrom* **2003**, 14, 943.
- 15 [11] J. A. Karty, M. M. Ireland, Y. V. Brun, J. P. Reilly, *J Chromatogr B Analyt Technol*  
16 *Biomed Life Sci* **2002**, 782, 363.
- 17 [12] E. D. Dodds, H. J. An, P. J. Hagerman, C. B. Lebrilla, *J Proteome Res* **2005**, 5, 1195.
- 18 [13] E. D. Dodds, B. H. Clowers, P. J. Hagerman, C. B. Lebrilla, *Anal Biochem* **2008**, 372,  
19 156.
- 20 [14] W. E. Wolski, M. Farrow, A. K. Emde, H. Lehrach, M. Lalowski, K. Reinert, *Proteome*  
21 *Sci* **2006**, 4, 18.
- 22 [15] C. Bruce, M. A. Shifman, P. Miller, E. E. Gulcicek, *Anal Chem* **2006**, 78, 4374.
- 23 [16] W. D. Lehmann, A. Bohne, C. W. von der Lieth, *J. Mass Spectrom* **2000**, 35, 1335.
- 24 [17] S. Pourshahian, P. A. Limbach, *J. Mass Spectrom* **2008**, 43, 1081.
- 25 [18] M. Kirchner, W. Timm, P. Fong, P. Wangemann, H. Steen, *Bioinformatics* **2010**, 26,  
26 791.
- 27 [19] J. A. Vizcaino, E. W. Deutsch, R. Wang, A. Csordas, F. Reisinger, D. Rios, J. A.  
28 Dianes, Z. Sun, T. Farrah, N. Bandeira, P. A. Binz, I. Xenarios, M. Eisenacher, G.  
29 Mayer, L. Gatto, A. Campos, R. J. Chalkley, H. J. Kraus, J. P. Albar, S. Martinez-  
30 Bartolome, R. Apweiler, G. S. Omenn, L. Martens, A. R. Jones, H. Hermjakob, *Nat*  
31 *Biotechnol* **2014**, 32, 223.

32

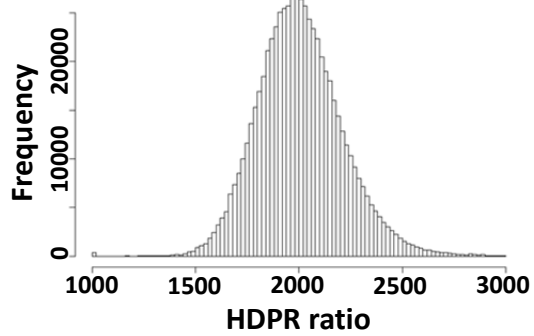
Figure 1



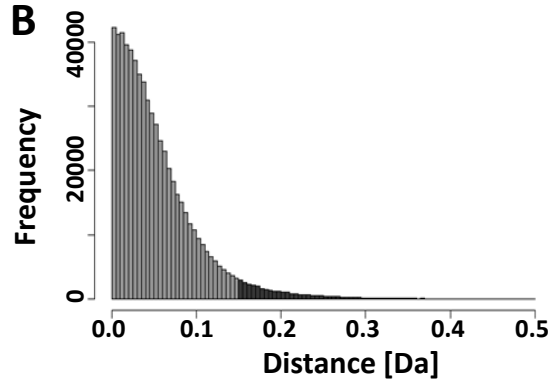


**Figure 2**

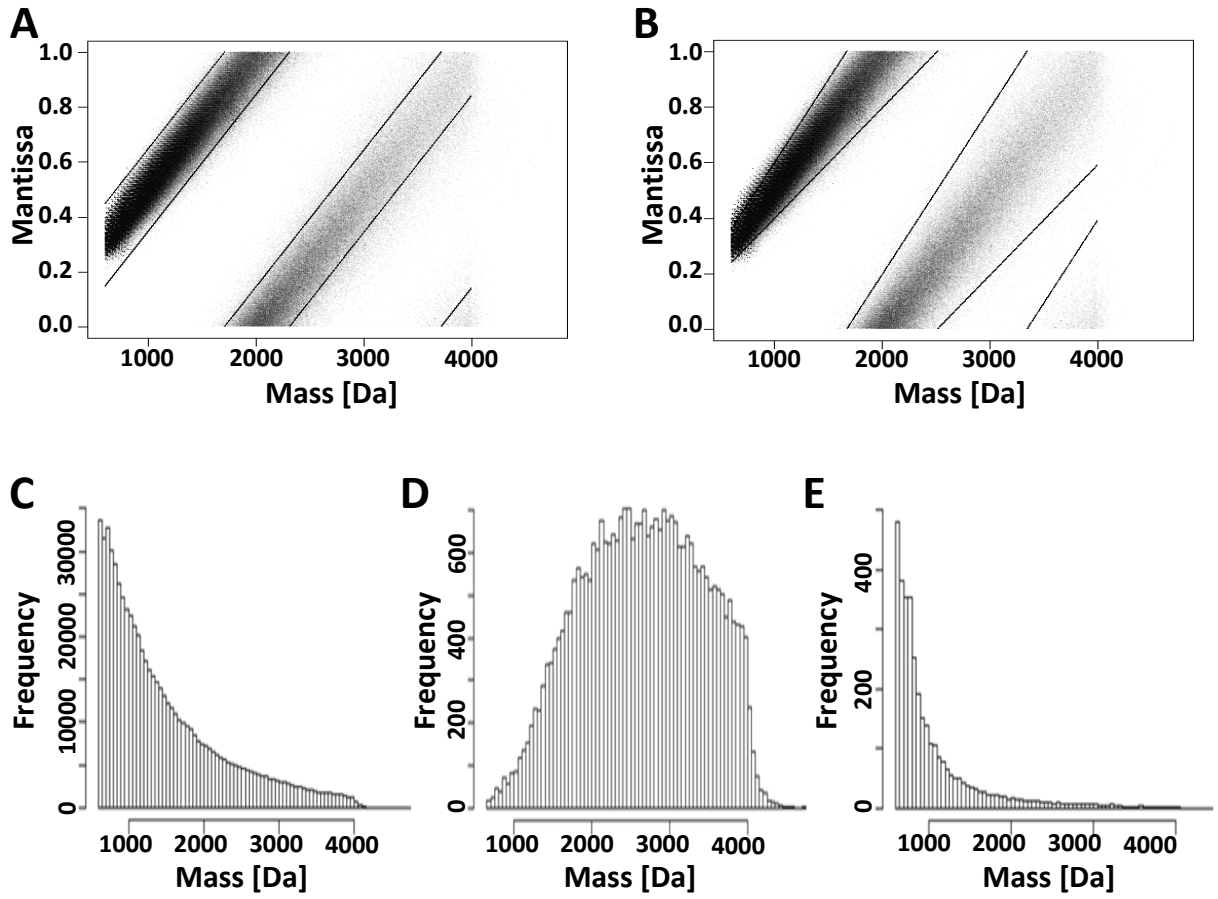
**A**



**B**



**Figure 3**



**Figure 4**

