Evaluation of search engines for phosphopeptide identification and quantitation

Xiaoyue Jiang[1], David Horn[1], Ryan Bomgarden[2] ,Tara Schroeder[3], Rosa Viner[1], Andreas FR Huhmer[1]
[1]Thermo Fisher Scientific, San Jose, CA; [2]Thermo Fisher Scientific, Rockford, IL; [3]Thermo Fisher Scientific, Somerset, NJ

Poster Note 64793

ABSTRACT

Purpose: To compare different fragmentation methods including HCD, CID, EThcD for phosphopeptide identification. The performance of search engine platforms for phosphopeptide identification and quantitation were evaluated and optimized.

Methods: Enriched and multiplexed quantitative phosphopeptides were analyzed by mass spectrometry (MS) and searched using the SequestHT, Byonic and Mascot nodes in Proteome Discoverer software and the Andromeda search engine in MaxQuant software .

Results: Search engines performed differently for multiple fragmentation methods on the same sample. The quantitation accuracy was comparable across search engines. Novel search workflows were designed in Proteome Discoverer to maximize identification and to optimize quantitation accuracy.

INTRODUCTION

The identification of phosphoproteomes is of keen interest to biologists because of the significance of phosphorylation in numerous cellular processes. The quantitation of protein phosphorylation and the correct localization of phosphorylation modification sites are equally important to understand protein function. However, the identification, quantitation, and localization of biological post-translational modifications are challenging for both mass spectrometer acquisition and proteomics search engines. Fragmentation of phosphopeptides often results in spectra that are dominated by the neutral loss of the phosphate group. Even with new advances in complementary fragmentation techniques, the database search is still accompanied with incorrect assignment and missed identifications. As the number, types and combinatorial variations of PTMs expand it becomes even more challenging. In this study, we evaluated multiple search algorithms on different fragmentation methods for phosphopeptide identification, site localization and quantitation accuracy in large phosphoproteomics.

MATERIALS AND METHODS

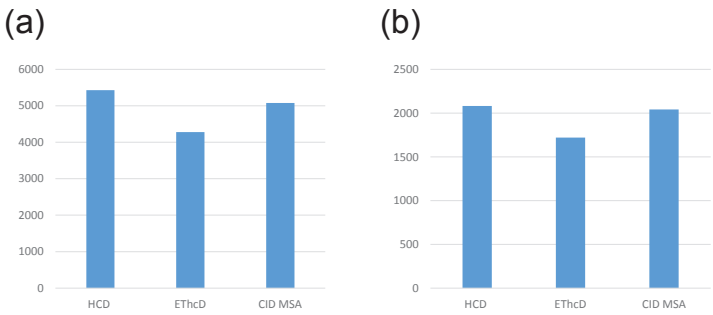
Sample Preparation

HeLa cells were lysed, digested, and further enriched for phosphopeptides using the Thermo Scientific™ Pierce™ Fe-NTA Phosphopeptide Enrichment Kit. The same digest was labeled by Thermo Scientific™ TMT10plex™ reagents, and mixed at ratios of 16:8:4:2:1:1:2:4:8:16. Yeast digest (Promega) was labeled with the last 5 channels mixed equimolar (0:0:0:0:0:1:1:1:1:1) and spiked into the aforementioned TMT-labeled HeLa digest sample. This multiplexed quantitative sample was also enriched for phosphopeptide analysis.

Liquid Chromatography and Mass Spectrometry

The samples were analyzed on a Thermo Scientific™ Orbitrap Fusion™ Lumos™ mass spectrometer coupled to the Thermo Scientific™ Easy-nLC™ 1000 chromatography system with a 50 cm Thermo Scientific™ EASY-Spray™ column. Different fragmentation methods (see results for details) were applied to compare the identifications and quantitation accuracy.

Figure 1. Numbers of (a) phosphopeptide groups and (b) protein groups identified by Sequest HT search engine for different fragmentation techniques



Data Analysis

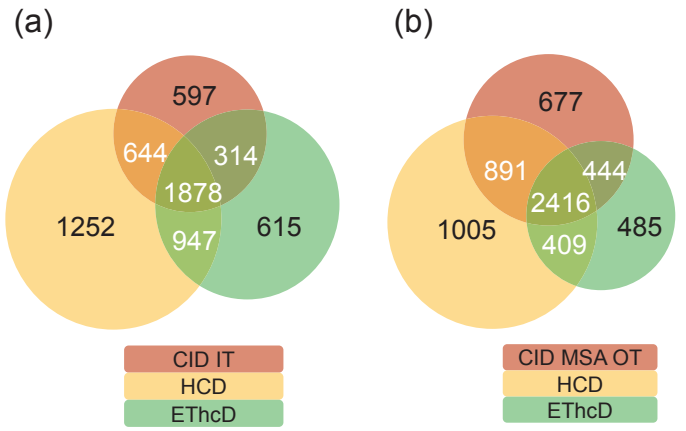
The data were analyzed using Thermo Scientific™ Proteome Discoverer™ 2.1 software and MaxQuant™ software 1.5.3.51. The search algorithms used in the study were Sequest HT™, Mascot™ v2.4, Byonic™ v2.8 as part of the Proteome Discoverer platform, and MaxQuant with the Andromeda search engine. Carbamidomethylation(C) and Thermo Scientific™ TMT 6-plex™ (N-term and K) were set as fixed modifications, while oxidation (M), deamidation (NQ) and phospho (STY) were set as dynamic modifications. A mass tolerance of 10ppm was set for MS1 and 0.02 Da for MS2 if using the orbitrap or 0.6Da for MS^2 for the ion trap analyzer. For multiplexed quantitative experiment, the isotopic correction factors for the TMT reagents were applied. A FDR of 1% at the peptide level and phosphorylation were used to filter the results (In addition, 1% protein FDR was used for Byonic and MaxQuant search engines by default). ptmRS was utilized to calculate the site localization probabilities of all the PTMs. A probability of 90% or higher was considered as a confident phosphorylation site. All search engines used the same SwissProt Human fasta database. The results from Proteome Discoverer were imported into Thermo Scientific™ ProteinCenter ™ software for comparison.

RESULTS

Identification

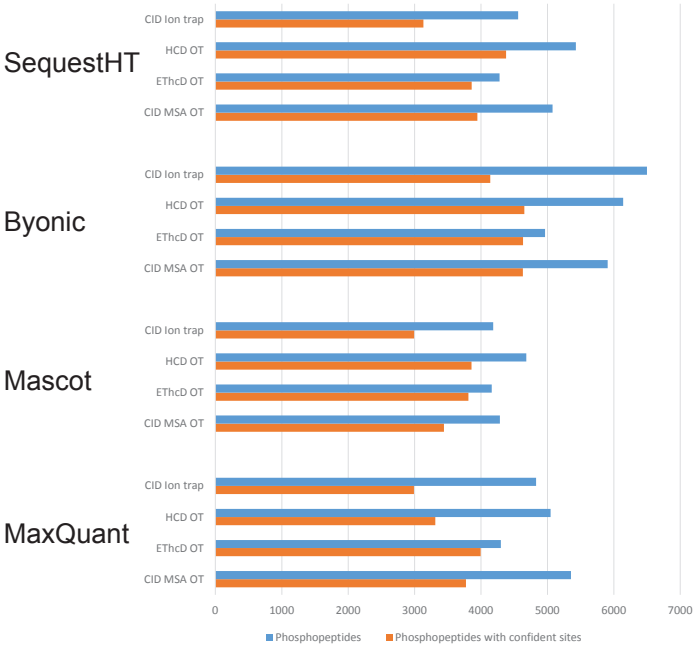
Different fragmentation methods including collision induced dissociation (CID), high energy collisional dissociation (HCD), and electron transfer dissociation (ETD) for phosphopeptide identification have been compared in the past[1,2]. It was found that CID resulted in significant neutral loss for phosphopeptides and HCD was the best on the latest generation of orbitrap instrument. Recently, some new fragmentation techniques have been developed, including EThcD and multistage activation (MSA) CID. The EThcD method uses HCD as supplemental activation for ETD. It allows more fragments to be generated and matched to the precursor ion as compared to ETD only[3]. Another technique, MSA ,involves simultaneous activation of the precursor ion and the resultant neutral loss product ion during a single CID-MS/MS event[4]. The comparison of these two new fragmentation methods with HCD fragmentation are shown in Figure 1. EThcD still suffered 21% loss of identifications compared to HCD, mainly due to the slow scan speed. MSA CID with the Orbitrap analyzer achieved 94% of the identification by HCD. MSA significantly increased the overlap of CID fragmentation with HCD and EThcD methods, as illustrated in Figure 2.

Figure 2. Venn Diagrams to compare phosphopeptide identifications from HCD, EThcD with (a) CID with ion trap as an analyzer and (b) CID MSA with an Orbitrap as an analyzer. The results were searched with Sequest HT engine.



While it was clear that HCD was the best fragmentation method for phosphopeptide identification using SequestHT, any potential bias generated by the search engine itself should be considered before making conclusion about the performance of the fragmentation technique. Different search algorithms and scoring theme can result in different identifications and therefore conclusions about the performance of different fragmentation techniques. To evaluate this, we performed the comparison of fragmentation methods for phosphopeptide identification on multiple search engines, including Sequest HT, Mascot and Byonic in Proteome Discoverer software and Andromeda in MaxQuant (Figure 3).

Figure 3. The comparison of multiple search engines and their effect on identification and site confidence of phosphopeptides analyzed by different fragmentation methods. OT: Orbitrap



HCD fragmentation was proven by multiple search engines to generate the highest or second highest number of phosphopeptide identifications compared to other fragmentation techniques. Traditional CID fragmentation generated fewer identifications by Sequest HT, Mascot and MaxQuant, but was favored by Byonic, possibly due to the score boost when a neutral loss peak was observed. EThcD was the fragmentation methods which gave the lowest phosphopeptide identifications as a result of a slower duty cycle during the data acquisition. MSA improved the identifications compared to CID using Sequest HT, Mascot and MaxQuant, but produced less identifications compared to traditional CID by Byonic, due to lack of diagnostic neutral loss ions. MaxQuant generated more identifications from CID MSA method over HCD fragmentation. As for phosphosites mapping, as expected EThcD spectra produced high confident site localizations for over 90% of all the identified phosphopeptides by all four search engines. Consistently, ion trap CID contributed the least site determination. Overall, Byonic provided the highest number of confidently identified phosphosites and phosphopeptides followed by Sequest HT and MaxQuant engines.

Quantitation

While identification of phophopeptides is important, quantitative phosphoproteomics allows elucidating changes in protein phosphorylation between different samples. Isobaric labeling, such as the TMT approach, has been widely used to compare the protein quantitation of several conditions in the same experiment. The higher number of phosphopeptide identifications facilitates the discovery of new biological targets only if the quantitation accuracy is not compromised.

Here we used the two proteome model to study the quantitation accuracy. The workflow is described in Figure 4. The Synchronous Precursor Selection (SPS) MS^3 method eliminated the yeast interference and resulted in accurate ratios for the human proteome. A theoretical log2 ratio of 3 was expected for 131/129N for true human proteins and peptides. On the other hand, if a yeast peptide was mistakenly identified as a human peptide, the log2 ratio would shift towards 0. After applying the correction factor, we did see a Gaussian distribution around 3 for human peptide ratios for all four search engine results, confirming the identifications were confident (Figure 5).

Figure 4. Experimental workflow to utilize quantitation to confirm identifications in phosphoproteomics.

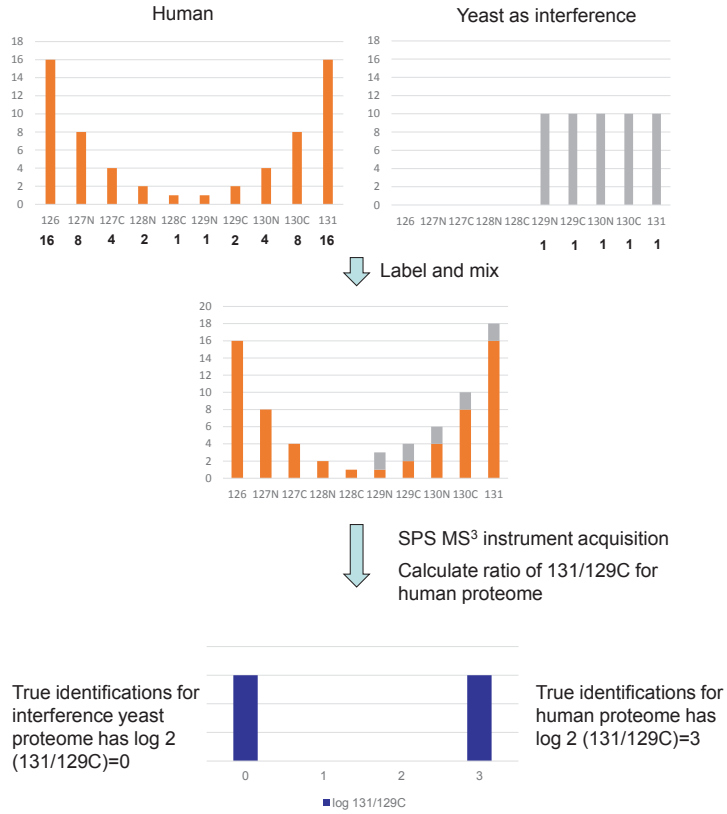
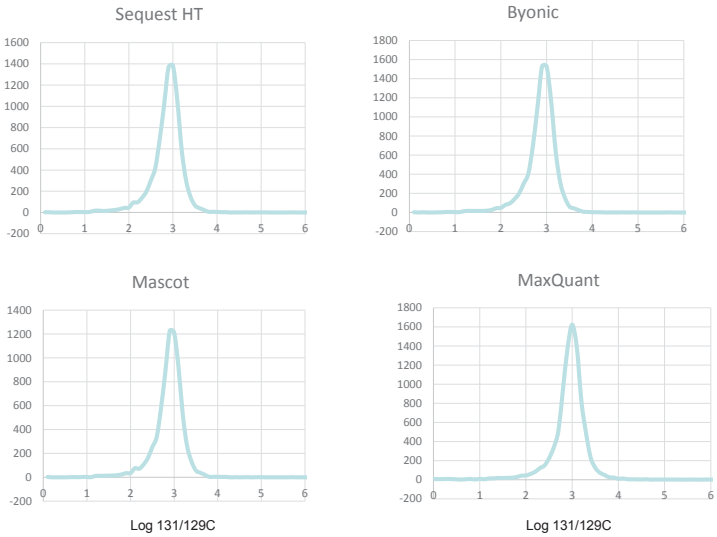


Figure 5. . Distributions of TMT ratios corresponding to human phosphopeptide with interference (131/129C). Expected Ratio is 8:1.



Maximizing phosphopeptide identifications and quantitation

Even though HCD was favored by all search engines for phosphopeptide identification and site localization, it suffered from distorted TMT quantitation in the complex sample. The Orbitrap Fusion and Lumos MS overcome this issue by implementing SPS MS3 method, but required CID IT for identification, which was proven to be the least favorable method for phosphopeptide identification. We developed two novel instrument methods to improve the identification for multiplexed quantitative phosphoproteomics[5]. In detail, one method was to couple CID MSA OT to SPS MS3. The search workflow in PD is illustrated in Figure 6. The other method was named neutral loss (NL) trigger method, which used a NL triggered MS3 event with moderate collision energy (HCD CE38) to obtain both the sequence information and the quantitation from MS3 spectrum. The benefits of two new methods over traditional SPS MS3 have been demonstrated [5]. The search workflows are shown in Figure 6 and 7. The NL method required identifications at both the MS2 and MS3 levels, which is only achievable in the Proteome Discoverer platform.

Figure 6. Proteome Discoverer search workflow for CID MSA OT method

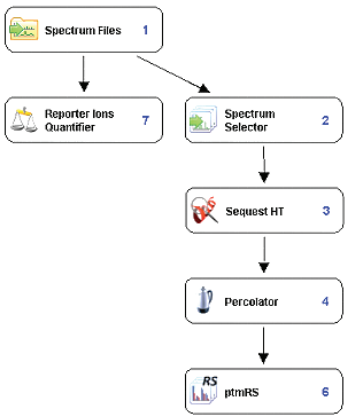
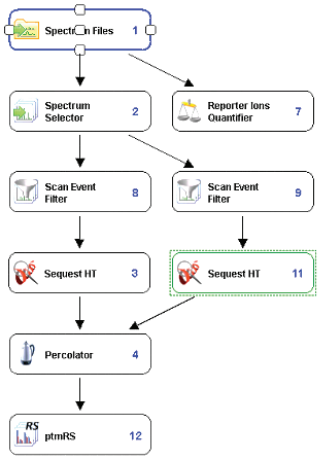


Figure 7. Proteome Discoverer search workflow for NL trigger method



Scan Event Filter 8: MS order is MS2
Scan Event Filter 8: MS order is MS3
SequestHT 3: Fragment mass tolerance is 0.6Da
SequestHT 11: Fragment mass tolerance is 0.02Da

CONCLUSIONS

- Different search engines favor different fragmentation methods for phosphopeptide identifications
- HCD fragmentation was favored by all search engines for phosphopeptide identification
- Byonic engine generated highest number of phosphopeptide identifications and confident sites
- The performance of multiplexed quantitation accuracy of phosphopeptides was comparable on all four search engines

REFERENCES

1. Marx, H. et al. 2013, *Nat Biotechnol.* **31(6):** 557-564
2. Nagaraj, N. et al. 2010, *J. Proteome Res.,* **9(12):** 6786-6794
3. Frese, C. et al. 2013, *J. Proteome Res.*, **12(3):** 1520-1525
4. Schroeder, M. et al 2004. *Anal. Chem.,* **76:** 3590–3598
5. Jiang, X. et al. 2016, *ASMS poster:* Sensitive and accurate quantitation of phosphopeptides using TMT isobaric labeling technique

Africa +43 1 333 50 34 0
Australia +61 3 9757 4300
Austria +43 810 282 206
Belgium +32 53 73 42 41
Brazil +55 11 2730 3006
Canada +1 800 530 8447
China 800 810 5118 (free call domestic) 400 650 5118
Denmark +45 70 23 62 60
Europe-Other +43 1 333 50 34 0
Finland +358 10 3292 200
France +33 1 60 92 48 00
Germany +49 6103 408 1014
India +91 22 6742 9494
Italy +39 02 950 591
Japan +81 6 6885 1213
Korea +82 2 3420 8600
Latin America +1 561 688 8700
Middle East +43 1 333 50 34 0
Netherlands +31 76 579 55 55
New Zealand +64 9 980 6700
Norway +46 8 556 468 00
Russia/CIS +43 1 333 50 34 0
Singapore +65 6289 1190
Sweden +46 8 556 468 00
Switzerland +41 61 716 77 00
Taiwan +886 2 8751 6655
UK/Ireland +44 1442 233555
USA +1 800 532 4752

PN10516-EN 0616S