

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/5537200>

# Automatic Validation of Phosphopeptide Identifications by the MS2/MS3 Target-Decoy Search Strategy

ARTICLE *in* JOURNAL OF PROTEOME RESEARCH · MAY 2008

Impact Factor: 4.25 · DOI: 10.1021/pr700675j · Source: PubMed

---

CITATIONS

36

---

READS

38

7 AUTHORS, INCLUDING:



**Shun Feng**

Xinjiang University, Urumqi, Xinjiang, China

51 PUBLICATIONS 1,592 CITATIONS

SEE PROFILE



**Mingliang Ye**

Dalian Institute of Chemical Physics

180 PUBLICATIONS 5,503 CITATIONS

SEE PROFILE

## Automatic Validation of Phosphopeptide Identifications by the MS2/MS3 Target-Decoy Search Strategy

Xinning Jiang,<sup>†,‡</sup> Guanghui Han,<sup>†,‡</sup> Shun Feng,<sup>†</sup> Xiaogang Jiang,<sup>†</sup> Mingliang Ye,<sup>\*,†</sup>  
Xuebiao Yao,<sup>§</sup> and Hanfa Zou<sup>\*,†</sup>

*National Chromatographic R&A Center, Dalian Institute of Chemical Physics, The Chinese Academy of Sciences, Dalian 116023, China, Graduate School of Chinese Academy of Sciences, Beijing 100049, China, and Laboratory of Cell Dynamics, School of Life Sciences, University of Science and Technology of China, Hefei 230027, China*

Received October 17, 2007

Manual checking is commonly employed to validate the phosphopeptide identifications from database searching of tandem mass spectra. It is very time-consuming and labor intensive as the number of phosphopeptide identifications increases greatly. In this study, a simple automatic validation approach was developed for phosphopeptide identification by combining consecutive stage mass spectrometry data and the target-decoy database searching strategy. Only phosphopeptides identified from both MS2 and its corresponding MS3 were accepted for further filtering, which greatly improved the reliability in phosphopeptide identification. Before database searching, the spectra were validated for charge state and neutral loss peak intensity, and then the invalid MS2/MS3 spectra were removed, which greatly reduced the database searching time. It was found that the sensitivity was significantly improved in MS2/MS3 strategy as the number of identified phosphopeptides was 2.5 times that obtained by the conventional filter-based MS2 approach. Because of the use of the target-decoy database, the false-discovery rate (FDR) of the identified phosphopeptides could be easily determined, and it was demonstrated that the determined FDR can precisely reflect the actual FDR without any manual validation stage.

**Keywords:** phosphoproteome analysis • MS2/MS3 • target-decoy search • automatic validation

### Introduction

Protein phosphorylation is one of the most important reversible, covalent protein posttranslational modifications (PTMs), and it plays a key role in eukaryotic signal transduction, gene regulation, and metabolic control in cells.<sup>1</sup> Fundamental processes such as cell proliferation, adaptation, and differentiation are governed by reversible phosphorylation at specific serine, threonine, and tyrosine residues in proteins. As phosphorylation is usually present at low stoichiometry, a variety of enrichment strategies have been developed to improve the detection of phosphopeptides. Phosphopeptides are usually enriched by the following chromatographic techniques: immobilized metal ion affinity chromatography (IMAC),<sup>2–5</sup> strong cation exchange chromatography,<sup>5–7</sup> titanium dioxide (TiO<sub>2</sub>) microcolumns,<sup>7–9</sup> and zirconium dioxide (ZrO<sub>2</sub>) microcolumns.<sup>10</sup>

Owing to its reliability, speed, and sensitivity, tandem mass spectrometry in combination with affinity-based phospho-

peptide enrichment methods has become a powerful tool for global profiling and quantification of protein phosphorylation.<sup>4–9,11,12</sup> One of the challenges in phosphorylation analysis by mass spectrometry is the fact that phosphorylation is generally a labile modification. In mass spectrometry driven proteomics, protein identification is based on peptide fragmentation by collisionally activated tandem mass spectrometry (MS/MS or MS2)<sup>13,14</sup> where cleavages of covalent bonds mainly occur through the lowest energy pathways. Because the energy needed to dissociate the phosphorylation bond is much lower than that of a peptide amide bond, phosphopeptides often lose phosphoric acid (H<sub>3</sub>PO<sub>4</sub>, 98 Da) and produce a single dominant peak corresponding to the neutral loss. Thus, a large percentage of phosphopeptides undergo a significant neutral loss of phosphoric acid, and fragmentation of the peptide backbone yields few or no sequence ions.<sup>5</sup> This severely hampers efficient backbone fragmentation by MS/MS and reduces the ability of database searching algorithms to unambiguously identify phosphopeptides. To obtain more fragmentation information, the neutral loss ion in MS2 stage can be further fragmented to acquire the MS/MS/MS (MS3) spectrum.<sup>6,12,15</sup> However, as the third stage MS, only the highly abundant neutral loss ion can generate decent MS spectra and thus lacks enough sensitivity. Despite extensive effort made in the past several

\* Corresponding authors. Prof. Dr. Hanfa Zou. Tel.: +86-411-84379610. Fax: +86-411-84379620. E-mail: hanfazou@dicp.ac.cn. Prof. Dr. Mingliang Ye. Tel.: +86-411-84379620. Fax: +86-411-84379620. E-mail: mingliang@dicp.ac.cn.

<sup>†</sup> The Chinese Academy of Sciences.

<sup>‡</sup> Graduate School of Chinese Academy of Sciences.

<sup>§</sup> University of Science and Technology of China.

years, reliable analysis of protein phosphorylation remains a daunting challenge.

A widely used approach for the automated evaluation of FDR for peptide identifications is the target-decoy strategy which is based on the principle that incorrect matches have an equal probability of being derived from either the target or the decoy database. This strategy has been validated by many laboratories.<sup>16–19</sup> After database searching against a composite protein database including both target (forward) and decoy (reversed) sequences, FDR can be easily determined through the number of decoy identifications. Using the target-decoy search strategy for the acquired spectra, a data set of peptide identifications with low FDR (for example, 1%) could be easily established through postsearch filtering with easily accessible criteria. Recently, methods incorporating a very high accuracy mass spectrometer with MS2 target-decoy search strategy to get high confidence phosphopeptide identification and precise site location without manual validation have been reported.<sup>7,16</sup> However, very high mass accuracy is only available from a few specialized types of mass spectrometers. For the MS2 spectra obtained from a low accuracy mass spectrometer like an ion trap mass spectrometer, the MS2 target-decoy search strategy still lacks enough sensitivity because of the suppression of phosphopeptide match scores assigned by the current database searching algorithm and the similarities of these scores when a phosphopeptide has more than one potential phosphorylation site.<sup>15,16</sup> Large-scale experiments which focused on the phosphopeptide identifications by LC-MS/MS relied mainly on manual validation to control for error during the past few years.<sup>4,6,11,12,20–22</sup> However, this is very labor intensive and becomes more and more impractical as data sets have grown progressively, and the error rate can hardly be estimated. Recently, statistical approaches were developed to assess whether a search result was likely to be phosphorylated; however, they cannot verify the sequence and phosphorylation sites of phosphopeptide identification.<sup>23</sup>

It was reported that the confidence was greatly improved when a phosphopeptide was independently identified by both MS2 and its MS3 spectra.<sup>24,25</sup> In these studies, the MS2 and MS3 search results were typically filtered and validated separately before combining the results. Due to the poor quality of spectra for phosphopeptides, a very few number of phosphopeptides could be identified by both MS2 and its MS3 spectra. Therefore, phosphoproteome analysis was still mainly based on MS2 spectra, and MS3 spectra were only used as a complement to MS2 spectra in most of the research works until now.<sup>4,16</sup> In this work, a simple fully automated phosphopeptide validation method was developed by combining the results of the MS2 target-decoy search and the MS3 target-decoy search. The search results were combined before filtering which allowed phosphopeptides with poor MS2 and MS3 spectra to be identified and thus significantly improved the sensitivity. As the target-decoy database was used, the confidence of phosphopeptide identifications could be determined automatically. Three new defined scores were introduced as filtering criteria to generate phosphopeptide identifications with specific FDR in the MS2/MS3 target-decoy search approach. It was found that the determined FDR was demonstrated to precisely reflect to the actual FDR.

## Experimental Procedure

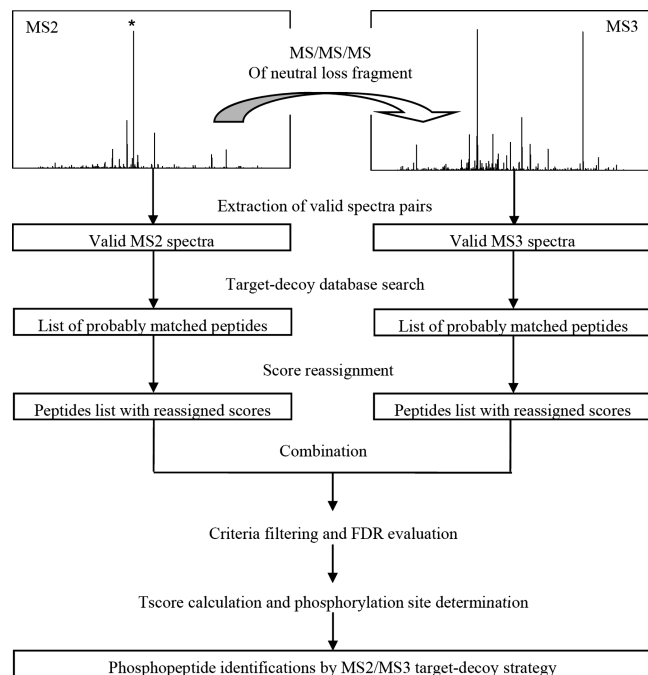
**Preparation of Protein Samples.**  $\alpha$ -Casein was dissolved in 0.5 mL of 50 mM Tris-HCl (pH 8.1) buffer at 10 mM concentra-

tion, then trypsin was added at a protein/enzyme ratio of 50:1 by weight. The solution was incubated at 37 °C for 16 h. Finally, the solution was acidified with 0.1% FA for further usage.

HeLa cells were placed in ice-cold homogenization buffer consisting of 8 M urea, 4% CHAPS w/v, 65 mM DTT, 1 mM EDTA, 0.5 mM EGTA, a mixture of protease inhibitor (1 mM PMSF), and phosphatase inhibitors (0.2 mM  $\text{Na}_3\text{VO}_4$ , 1 mM NaF), and 40 mM Tris-HCl at pH 7.4. The suspension was homogenized for approximately 1 min, sonicated for 180 s at 400 W, and centrifuged at 25 000g for 1 h. The supernatant contained the total HeLa cell soluble proteins. An appropriate volume of supernatant was precipitated to remove insoluble substance, lyophilized to dryness, and redissolved in denaturing solution (8 M urea, 100 mM ammonium bicarbonate, at pH 8.3) with the protein concentration determined to be 17.8 mg/mL. Then, 300 mg of this protein sample was mixed with 1 mL of 1 M DTT. The mixture was incubated at 37 °C for 2.5 h, and then 5 mL of 1 M IAA was added and incubated for an additional 30 min at room temperature in the dark. The protein mixtures were diluted by 10-fold with 100 mM ammonium bicarbonate buffer (pH 8.5) and incubated with trypsin (25:1 w/w) at 37 °C overnight. The trypsin digest was stored in the refrigerator under –30 °C for further usage.

**Enrichment of Phosphopeptides by  $\text{ZrO}_2$ .** The enrichment procedure using  $\text{ZrO}_2$  nanoparticles was similar to that reported before.<sup>10</sup> Digests (25  $\mu\text{g}$ ) of proteins extracted from HeLa cells were diluted with loading buffer containing 10% acetic acid and 50% ACN (v/v) (pH 3–4). It was then mixed with 25  $\mu\text{L}$  of  $\text{ZrO}_2$  nanoparticle suspension (30 mg/mL). The resulting solution was incubated for 30 min at room temperature, and it was then centrifuged at 30 000g for 5 min. The  $\text{ZrO}_2$  nanoparticles with trapped phosphopeptides were deposited on the bottom of the tube, and the supernatant was removed by an Eppendorf pipet. Subsequently, the isolated nanoparticles were rinsed with 50  $\mu\text{L}$  of solution of 10% acetic acid containing 50% ACN to remove nonspecifically adsorbed peptides. The  $\text{ZrO}_2$  nanoparticles were again isolated with centrifugation at 30 000g for 5 min. The supernatant was once again removed with a pipet. The trapped phosphopeptides on  $\text{ZrO}_2$  nanoparticles were eluted using 15  $\mu\text{L}$  of  $\text{NH}_4\text{OH}$  (pH 11.5) under sonication for 10 min. After centrifugation, the supernatant was lyophilized and dissolved in 0.1% formic acid for nano-LC MS/MS and MS3 analysis.

**HPLC and MS<sup>n</sup> Analysis.** A Finnigan surveyor MS pump (ThermoFinnigan, San Jose, CA) was used to deliver the mobile phase. The pump flow rate was split by a cross to achieve a column flow rate of about 200 nL/min. The fused-silica capillary (75  $\mu\text{m}$  i.d.) was manually pulled to a fine point of  $\sim 5 \mu\text{m}$  with a flame torch at one end and packed with C18 AQ particles (5  $\mu\text{m}$ , 120 Å) from Michrom BioResources (Auburn, CA, USA) to the length of 12 cm. The nano-RPLC column was directly coupled to a LTQ linear IT MS from ThermoFinnigan with a nanospray source. The LTQ instrument was operated at positive ion mode with a spray voltage of 1.8 kV. The scan range of each full MS scan was  $m/z$  400–1700. ACN gradients of 5–35% for 30 and 50 min at a flow rate of 200 nL/min were applied for the separation of phosphopeptides from  $\alpha$ -casein and HeLa cell extract, respectively. For the detection, the MS was set as a full scan MS followed by three data dependent MS2 events. A subsequent MS3 event was triggered upon detection when a neutral loss of –49 or –32.7 (loss of  $\text{H}_3\text{PO}_4$  for the +2 and +3 charged ions, respectively) was detected among the top 10 most intense ions in MS2. A dynamic



**Figure 1.** Flowchart illustrating the computing algorithm of this study.

exclusion window was applied which prevented the same  $m/z$  from being selected for 1 min after its acquisition. This entire LC-MS<sup>n</sup> system was controlled under Xcalibur software 2.0 (ThermoFinnigan, San Jose, CA).

**Database Search.** The MS2 and MS3 spectra were searched using SEQUEST<sup>26</sup> (v 0.27) against a composite database including both the original database and the reversed version of this original database with the following parameters: peptide mass tolerance, 2 Da; MS2 and MS3 fragment ion mass tolerance, 1 Da; enzyme set as trypsin and allowance up to two missed cleavages; for  $\alpha$ -casein, no static modification was selected, while for the HeLa cell lysate, static modification was cysteine carbamidomethylation (+57 Da); dynamic modifications were methionine oxidation (+16 Da), phosphorylation on serine, threonine, and tyrosine (+80 Da); for MS3 data, besides the above modifications, variable modifications of −18 Da ( $\beta$ -elimination of phosphoric acid) on serine and threonine residues were also selected. The original database used for HeLa cells was a human proteome sequence database (v3.17) from the European Bioinformatics Institute (<http://www.ebi.ac.uk/IPI/IPIhuman.html>).

**Phosphopeptide Identification Strategies.** The scheme of the MS2/MS3 target-decoy search approach was illustrated in Figure 1. New defined scores for MS2/MS3 pairs were first generated by combining the scores from MS2 and MS3 database searches, and then the filter was set with the new scores for MS2/MS3 spectra pairs to reach a specific FDR. There are four steps in detail: (A) Extraction of valid MS2/MS3 pairs and database searching. Because the charge state of the precursor ion cannot be determined with low mass accuracy MS, more than one DTA file with different precursor charge states (commonly 2+ and 3+, respectively) were exported for one tandem spectrum. By Combining MS2 spectra and corresponding neutral loss MS3, charge states of precursor ions can be determined from the  $m/z$  value of neutral loss: −49 indicated +2 charged precursor ions, while −32.7 was for +3 charged ions. Only a DTA spectrum with neutral loss peak of

at least 50% of the base peak in intensity was considered. After removal of MS2/MS3 pairs with incorrect charge states, MS2 with no MS3, and MS2/MS3 pairs with neutral loss intensity less than 50% of the base peak in MS2 spectrum, the remaining MS2 and MS3 DTA spectra with specific precursor charge states were searched against the same target-decoy database, respectively.

(B) Score rearrangement for peptides identified from spectra. The top 10 hit peptides from a database search for a spectrum were considered in this approach. Score reassignment for these top 10 hit peptides was achieved as: (i) Ranks of the adjacent peptides with the same amino acid sequence (with different phosphorylation site locations) were set to the same value as the first hit in the identifications, and the new ranks were defined as rank'; (ii)  $\Delta Cn'$ , which was defined as the  $\Delta Cn$  score to the first nonidentical sequence in the matches, was used to replace the  $\Delta Cn$  value as described by Gygi et al.<sup>5</sup>

(C) Matching and selection of the best candidate for a spectra pair. Peptide identifications from a pair of spectra (MS2 and its corresponding MS3) were combined. Only peptides which were identified from both of the spectra (MS2 and MS3) were retained, and  $Xcorr_s$  scores were calculated for these matched peptides as follows

$$Xcorr_s = Xcorr_{MS2} + Xcorr_{MS3}$$

where  $Xcorr_{MS2}$  was the  $Xcorr$  value of the identification of this peptide from MS/MS and  $Xcorr_{MS3}$  was that of MS3. The matched peptide in a spectra pair with the highest  $Xcorr_s$  score was defined as the top matched peptide for the spectra pair and selected for filter afterward (if any).

(D) Generating peptide identification with specific FDR. FDR was evaluated by the target-decoy strategy.<sup>16,17</sup> Matched peptides which were generated from C for pairs of spectra were filtered to specific FDR using criteria formed by the following three cutoff scores: (i) Rank'<sub>m</sub>, which was the minimum rank' for the peptide identification in MS2 and MS3, was set to 1 for filter (peptide identifying with rank' equals 1 in MS2 or MS3, or both was selected); (ii)  $\Delta Cn'_m$  was defined as the  $\Delta Cn'$  value for the peptide identifications in the spectra pair with a rank' of 1 or the maximum value of the two  $\Delta Cn'$  values if both the ranks' of peptide identifications in MS2 and MS3 were 1, and  $\Delta Cn'_m \geq 0.1$  was used to filter the identifications; (iii)  $Xcorr'_s$  was calculated in the same way from  $Xcorr_s$  to reduce the dependence of  $Xcorr_s$  on the length of the peptide by  $\ln(Xcorr_s)/\ln(\text{length})$ , as reported by Keller et al.<sup>27</sup> The most possible phosphorylation site localizations were then determined by Tscore as described below.

A conventional approach was also used for comparison. All MS2 and MS3 spectra without any prefiltering were searched against the same target-decoy database separately (two charge states for the precursor, i.e., 2+ and 3+, were considered). Then, the MS2 and MS3 search results were separately filtered by standard criteria using a strategy similar to that described:<sup>28</sup>  $Xcorr$  cutoffs for doubly and triply charged peptides were set constantly as 2.5 and 3.8, then the  $\Delta Cn$  cutoff was manually determined so that FDR of the final identified peptides was lower than a specific value (in this study <1%). To make the data comparative and objective, after the FDR evaluation by the target-decoy strategy, no further manual validation was employed. As the target-decoy database was also used, the above phosphopeptide identification approach was also termed as the MS2 or MS3 target-decoy search approach.



**Determination of the Phosphorylation Site.** For peptides with multiple possible phosphorylation sites, Mann and Blagoev have defined the PTM localization probability score (PTM score) to determine the most probable phosphorylation site localization based on the binomial probability  $P$  of randomly matching for phosphopeptides by comparing the matched ions in the tandem spectra to random chance.<sup>7</sup> The PTM score of a phosphopeptide for the spectrum was then calculated for easy reading by  $-10 \log(P)$ . As the majority of phosphopeptides was identified by either MS2 or MS3, the PTM score determined only from MS2 or MS3 was used to locate the possible phosphorylation site in their study. In this study, each phosphopeptide was identified by both MS2 and MS3. Therefore, two random probabilities for a candidate phosphopeptide were calculated and indicated as  $P_{ms2}$  and  $P_{ms3}$ . Even though both MS2 and MS3 spectra originated from the same phosphopeptide ion in MS1, they were from different precursor ions (intact and neutral loss ions) with different fragment characteristics. After performing database searching independently, the dependence of the phosphopeptide identifications achieved by MS2 and MS3 should be very low and so could be approximately evaluated as two independent events in statistics. Nesvizhskii et al. have determined this phenomenon and used this strategy for calculation of the combined peptide probability from MS2 and MS3.<sup>29</sup> Thus, the probability of random matching of a phosphopeptide,  $P_{total}$ , can be calculated as the product of  $P_{ms2}$  and  $P_{ms3}$ . The new localization probability score, Tscore, was defined as  $-10 \log(P_{total})$ . Therefore, for given phosphopeptides, its Tscore is the sum of its MS2 and MS3 PTM scores. Neutral loss induced peaks ( $MH-H_3PO_4$ ,  $MH-H_3PO_4-H_2O$ ,  $MH-H_3PO_4-NH_3$  and, etc.) were removed while calculating the Tscore. For the phosphopeptide with two or more phosphorylation sites, Tcores of all candidate sequences with different phosphorylation site combinations for this phosphopeptide were calculated. And then the Tscore of a given site was computed by summing the Tcores of all candidate sequences containing this site. Phosphorylation sites with top  $n$  (equal to the number of possible phosphorylation sites) Tcores were considered as the most likely phosphorylation site localizations.

**Software.** The Automatic Phosphopeptide Identification Validating algorithm for SEQUEST (APIVASE) was implemented in Java using the Java 2 standard edition develop kit 6.0. It is available from <http://bioanalysis.dicp.ac.cn/proteomics/software/APIVASE.html> freely for academic users.

## Results

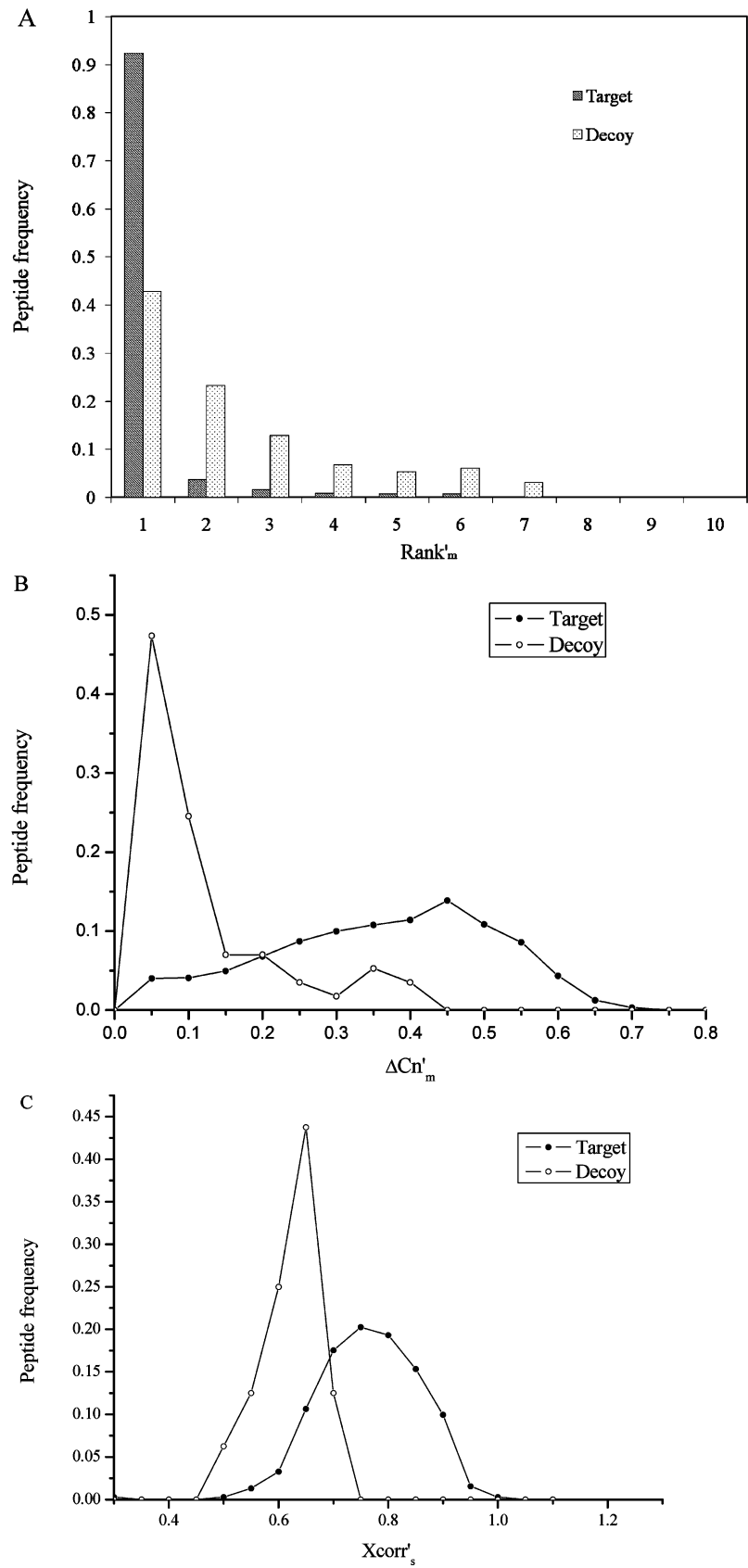
**MS2/MS3 Target-Decoy Search Strategy.** In this study, we developed an automatic approach for identification and validation of phosphopeptides under specific confidence levels. The flowchart was illustrated in Figure 1: (1) evaluation of charge state to remove invalid MS2/MS3 pairs; (2) performing MS2 and MS3 target-decoy database searches, separately; (3) reassignment of the peptide scores in SEQUEST output to generate a list of peptide identifications for a pair of MS2/MS3 spectra; (4) filtering the candidate phosphopeptides with criteria to achieve phosphopeptide identification with low FDR; (5) determining the most probable phosphorylation site by Tscore. This strategy combined the information obtained from neutral loss MS3 and its corresponding MS2 to improve the confidence of phosphopeptide identifications; thus, this strategy was also termed as the MS2/MS3 target-decoy search approach or the MS2/MS3 approach in short.

Phosphopeptides enriched by  $ZrO_2$  from the lysate of mitotic HeLa cells were analyzed by LC-MS/MS with three replicate runs, resulting in the collection of 23 768 MS2 and 11 889 MS3 spectra. After charge evaluation and removal of spectra pairs without significant neutral loss peaks, 4522 spectra pairs with unambiguous charge states were obtained. After database searching and combination, 1859 out of 4522 valid spectra pairs have matched peptide identifications. It was found that the top matched peptides (with highest  $Xcorr'_s$  values) for the spectra pairs were all phosphopeptides. The FDR was determined to be 14.3% (133 decoy hits out of 1859 target hits) without setting any other criteria. The low FDR without any further filtering demonstrated that the strategy combining the MS2 and MS3 target-decoy database searches can effectively remove false positive identifications and improve the confidence level in phosphopeptide identification. However, there were still about 15% false positive phosphopeptide identifications. To remove these false positive identifications, further filtering is required.

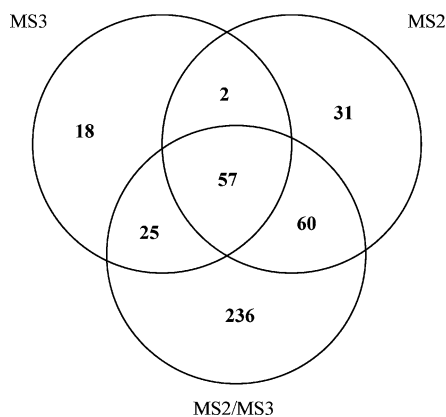
To further control the FDR of peptide identification, the scores used to filter the data set should be determined. The  $Xcorr$  value reflects the correlation between the acquired spectrum and the theoretical spectrum. The high quality spectrum of the peptide should be assigned with a high  $Xcorr$  value; otherwise, the low quality spectrum which may have resulted from inefficient fragment should be assigned with a low  $Xcorr$  value. As the MS2 and MS3 spectra are complementary, the sum of their  $Xcorr$  values, i.e.,  $Xcorr_s$ , should reflect the overall correlation between the pair of the spectra and the peptide candidate. Thus,  $Xcorr_s$  may be a good score to filter the data set. Then the less length dependent score,  $Xcorr'_s$ , was used for filtering.  $Rank'$  indicates the ranking of the peptide hit, and  $\Delta Cn'$  reflects the difference between the peptide hit and its next hit. As each peptide hit had two  $rank'$  and two  $\Delta Cn'$  corresponding to MS2 and MS3 spectra, to simplify the filtering procedure, the better scores as described in the method section, i.e.,  $Rank'_m$  and  $\Delta Cn'_m$ , were selected for filtering.

The distributions of the three scores, i.e.,  $Rank'_m$ ,  $\Delta Cn'_m$ , and  $Xcorr'_s$ , for the top peptide matches obtained from the combination of the results from MS2 and MS3 target-decoy searches were illustrated in Figure 2A, B, and C, respectively. As shown in Figure 2A, a majority (92.4%) of the true positive identifications were found with  $Rank'_m$  equals to 1, while nearly 60% of the false positive identifications can be removed by this filter ( $Rank'_m = 1$ ).  $\Delta Cn'_m$  distributions for target and decoy identifications were shown in Figure 2B. After  $Rank'_m$  filtering, 91.3% of the left target identifications were identified with  $\Delta Cn'_m$  values bigger than 0.1, while only 32.9% of the decoy identifications had  $\Delta Cn'_m$  values bigger than 0.1. Figure 2C shows the  $Xcorr'_s$  distribution of target and decoy identifications. Therefore, combinational using of these three scores should be able to filter the data set with desirable FDR. In this study, the criteria were set as follows:  $Rank'_m = 1$ ,  $\Delta Cn'_m \geq 0.1$ , and the  $Xcorr'_s$  cutoff value was determined by increasing its value until the observed FDR was just lower than the specified FDR.

Peptide identifications from the 4522 spectra pairs were filtered with the above filter ( $Xcorr'_s$  cutoff was determined as 0.60). Finally, 1383 spectra pairs were assigned as phosphopeptides with FDR of 1.01%. The phosphorylation sites were then determined by Tscore, and 378 unique phosphopeptides were identified. To demonstrate the performance of this method, the MS2 target-decoy search approach and MS3 target-decoy search approach were also used to generate



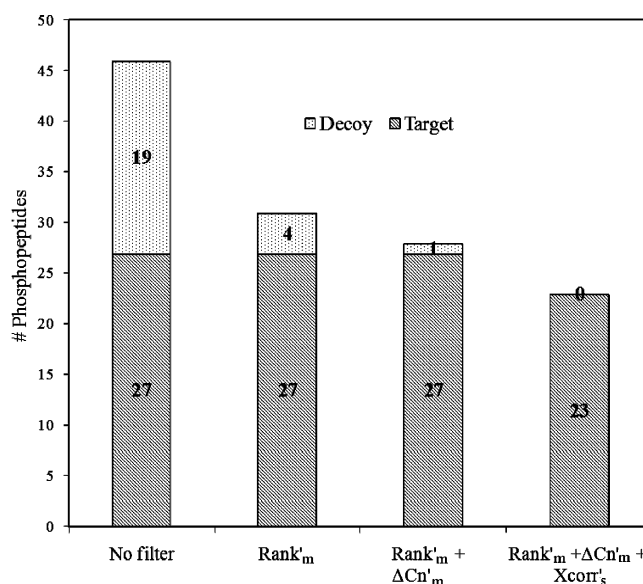
**Figure 2.** (A) Frequency distribution of the rank'\_m of target and decoy identifications obtained after combination. (B) Frequency distribution of the ΔCn'\_m of target and decoy identifications after rank'\_m filtering. (C) Frequency distribution of the Xcorr'\_s of target and decoy identifications after rank'\_m and ΔCn'\_m filtering.



**Figure 3.** Overlap of phosphopeptides identified from HeLa cells by three approaches: the MS2, MS3, and MS2/MS3 approach.

peptide identification with the same confidence level. The phosphorylation sites for the phosphopeptides identified by the MS2 approach or MS3 approach were determined by the PTM score as reported by Mann et al.<sup>7</sup> The overlap of the identified phosphopeptides by the three approaches was shown in Figure 3. For the MS2 target-decoy search approach, all of the 23 768 MS2 spectra were searched against the same target-decoy database by SEQUEST, and the generated peptide identifications were filtered in 1% FDR using the following criteria:  $Xcorr \geq 2.5$  and  $3.8$  for charge states of  $2+$  and  $3+$ , respectively, and  $\Delta Cn \geq 0.27$ . Finally, 150 unique phosphopeptides were identified, which was only 40% of the number of identified phosphopeptides by the MS2/MS3 approach, and 79% of phosphopeptides (119 from 150) identified from the MS2 approach could also be identified from the MS2/MS3 approach. Similar to MS2 spectra, the MS3 target-decoy search strategy was also applied to processes all the 11 889 MS3 spectra and resulted in the identification of 102 unique phosphopeptides. Among the 102 unique phosphopeptides, 59 were already identified by the MS2 approach. The total number of identified phosphopeptides increased to 193 when the identified phosphopeptides from MS2 and MS3 approaches were pooled together. However, this number is still only 51% of the number of phosphopeptides identified by the MS2/MS3 approach. Combining all the phosphopeptides identified by the three approaches only resulted in the increase of 51 phosphopeptides (13% of the total number of phosphopeptides identified by the MS2/MS3 approach). This indicated that significant loss of potential phosphopeptide identifications did not occur even though 81% of the MS2 spectra were discarded because they did not have valid MS3 spectra or the neutral loss peak in MS3 spectra was not significant. All the identified phosphopeptides with their distinct phosphorylation sites by these three strategies were shown in Supplementary Tables 1, 2, and 3, respectively.

In this work, we determined 451 phosphorylation sites by the MS2/MS3 approach, including 3 pY, 63 pT, and 385 pS sites (all the pY sites were detected from peptides containing another nontyrosine phosphorylation site where the neutral loss occurred, e.g., IGEGpTpYGVVYK). Thus, the distribution of pY, pT, and pS sites in HeLa cells determined in this study was 0.7%, 13.9%, and 85.4%, very close to the distribution which has been reported before<sup>7,30,31</sup> (in a recent study by Mann and Blagoev, the distribution was determined to be 1.8%, 11.8%, and 86.4% for pY, pT, and pS, respectively<sup>7</sup>). The consistently distributed serine and threonine phosphorylation sites deter-



**Figure 4.** Target and decoy phosphopeptide identifications for  $\alpha$ -casein using the MS2/MS3 target-decoy search strategy. Different combinations of three cutoff scores were used to filter phosphopeptide identifications: (1) without score cutoff; (2) using  $rank'_m = 1$ ; (3)  $rank'_m = 1$ ,  $\Delta Cn'_m \geq 0.1$ ; (4)  $rank'_m = 1$ ,  $\Delta Cn'_m \geq 0.1$  and  $Xcorr'_s \geq 0.64$ . High confidence identification was generated when three cutoff scores were used.

mined in this study clearly demonstrated the confidence of the phosphopeptide identification obtained by the MS2/MS3 target-decoy approach. It is known that phosphotyrosine (pY) containing peptides are relatively stable and often do not lose phosphoric acid to form predominant neutral loss peaks.<sup>32</sup> Therefore, peptides which were phosphorylated only at the tyrosine site would not be identified in this strategy. However, because the phosphorylation in normally growing cells only contains about 1% phosphotyrosine,<sup>7,30</sup> the loss of tyrosine phosphorylation should not significantly reduce the sensitivity in this approach.

**Validation of the MS2/MS3 Target-Decoy Strategy.** Tryptic digestion of bovine  $\alpha$ -casein, one model phosphoprotein that has been extensively characterized, was used as the sample. An amount of 1 pmol ( $\sim 25$  ng) of  $\alpha$ -casein digest was directly analyzed by the data-dependent neutral loss LC-MS/MS strategy. The acquired MS2 spectra were then searched against a composite database of the forward database containing the sequences of  $\alpha$ -s1-casein,  $\alpha$ -s2-casein, trypsin, all yeast proteins, and the reversed sequences of the above proteins. Conventional approaches ( $Xcorr \geq 2.5$  and  $3.8$  for charge states of  $2+$  and  $3+$ , respectively,  $\Delta Cn$  must be bigger than 0.29 for MS2 and 0.19 for MS3) were used to filter this data set so that no peptide from yeast or decoy database was identified. It was found that only three monophosphopeptides (VPQLEIVNPpSAEER, YKVPQLEIVNPpSAEER, and KYKVPQLEIVNPpSAEER) from  $\alpha$ -s1-casein were identified by the MS2 approach and the MS3 approach. These three identified phosphopeptides resulted in the same phosphorylation site identification.

Then, the same data were processed using the MS2/MS3 target-decoy database search strategy, and 46 peptide identifications including 27 target (peptide identifications from  $\alpha$ -s1-casein or  $\alpha$ -s1-casein) and 19 decoy (identifications of yeast or reversed sequences) identifications were obtained after the combination of spectra pairs. Filtered with  $rank'_m$  equal to 1, all of the 27 target identifications were retained, and 15 decoy

**Table 1.** Phosphopeptides Identified from α-Casein by a Conventional Strategy and the MS2/MS3 Target-Decoy Strategy

no.	MH+ (Da)	no. of phosphorylation sites	peptide sequence	MS2/MS3 target/decoy strategy	standard criteria
P1	1661	1	VPQLEIVNPpSAEER	✓	✓
P2	1952	1	YKVPQLEIVNPpSAEER	✓	✓
P3	2080	1	KYKVPQLEIVNPpSAEER	✓	✓
P4	1467	1	TVDMEpSTEVFTK	✓	
P5	1595	1	KTVDMEpSTEVFTK	✓	
P6	1848	1	DIGSEpSTEDQAMEDIK	✓	
P7	1928	2	DIGpSEpSTEDQAMEDIK	✓	
P8	2937	3	EKVNELpSKDIGpSEpSTEDQAMEDIK	✓	

identifications were removed. After  $\Delta Cn'_m (\geq 0.1)$  and  $Xcorr'_s (\geq 0.64)$  filtering, 23 of the 27 target identifications were generated with no decoy identifications (Figure 4). All of the 23 target identifications were manually validated as true positive identifications (<http://bioanalysis.dicp.ac.cn/proteomics/Publications/MS2MS3/alpha-casein.htm>). Eight nonredundant phosphopeptides were generated including two monophosphopeptides (TVDMEpSTEVFTK and KTVDMEpSTEVFTK) from α-s2-casein and six phosphopeptides from α-s1-casein (DIGpSEpSTEDQAMEDIK, DIGSEpSTEDQAMEDIK, KYKVPQLEIVNPpSAEER, VPQLEIVNPpSAEER, YKVPQLEIVNPpSAEER, EKVNELpSKDIGpSEpSTEDQAMEDIK). Compared with the MS2 approach, five more phosphopeptides were identified using the MS2/MS3 strategy. All the phosphopeptides identified by this strategy were shown in Table 1. The analysis of the standard phosphoprotein demonstrated the high classification performance and high sensitivity of the MS2/MS3 approach.

To further validate this strategy, the data set from HeLa cells was modified by adding 4 *m/z* units to each peak in every MS2 and MS3 spectrum to form the entirely falsified spectra. After a database search, the identified peptides should be random matches as have been demonstrated by Gygi et al.<sup>16</sup> However, when the same criteria ( $Xcorr$  bigger than 2.5 and 2.8 for 2+ and 3+,  $\Delta Cn \geq 0.27$ ) were used to filter search results of falsified MS2 spectra, 22 peptides passed the cutoff values even though the criteria were very strict. And it was very interesting that all of the 22 randomly matched peptides were identified as phosphopeptides even though there were no actual phosphopeptides presented. Therefore, manual validation for phosphopeptide identifications should always be needed for the MS2 approach. After combining the search results of falsified spectra pairs of MS2 and MS3, the resulting data set was filtered by the criteria used in the MS2/MS3 target-decoy search ( $rank'_m = 1$ ,  $\Delta Cn'_m \geq 0.1$ , and  $Xcorr'_s \geq 0.60$ ). It was found that there was no peptide passing the criteria. These results indicated that random matches can be easily removed (no random match passed filtering criteria in this approach) when the MS2/MS3 target-decoy approach was used, and the phosphopeptide identifications with high confidence could be generated.

To demonstrate whether the FDR determined in this work using the target-decoy strategy can reflect the true FDR, manual validation for the phosphopeptides identified from HeLa cell lysate was performed to evaluate the actual FDR. Manual validation was performed using the following criteria:<sup>20,22</sup> (i) the MS2 or MS3 spectrum in a spectra pair must be of good quality, with fragment ions clearly above the baseline noise level; (ii) the corresponding phosphoric acid neutral loss peak to phosphoserine and phosphothreonine must be a dominant signal; (iii) there must be some sequential members of b- or y-series ions in MS2 or MS3 spectra; (iv) for peptides with more

than two phosphorylation sites of phosphoserine or phosphothreonine, dominant neutral loss signals must also be found in MS3. 500 of the identified phosphopeptides were randomly selected and validated manually, and 496 phosphopeptides were validated as true positive. The actual FDR was 0.8%, which was very close to the predicted FDR by the target-decoy strategy. The above results clearly demonstrated that the phosphopeptide identifications were very confident, and the FDR determined by the MS2/MS3 target-decoy strategy can precisely reflect to the actual FDR. All of the spectra validated in this study could be found at <http://bioanalysis.dicp.ac.cn/proteomics/Publications/MS2MS3/hela.htm>.

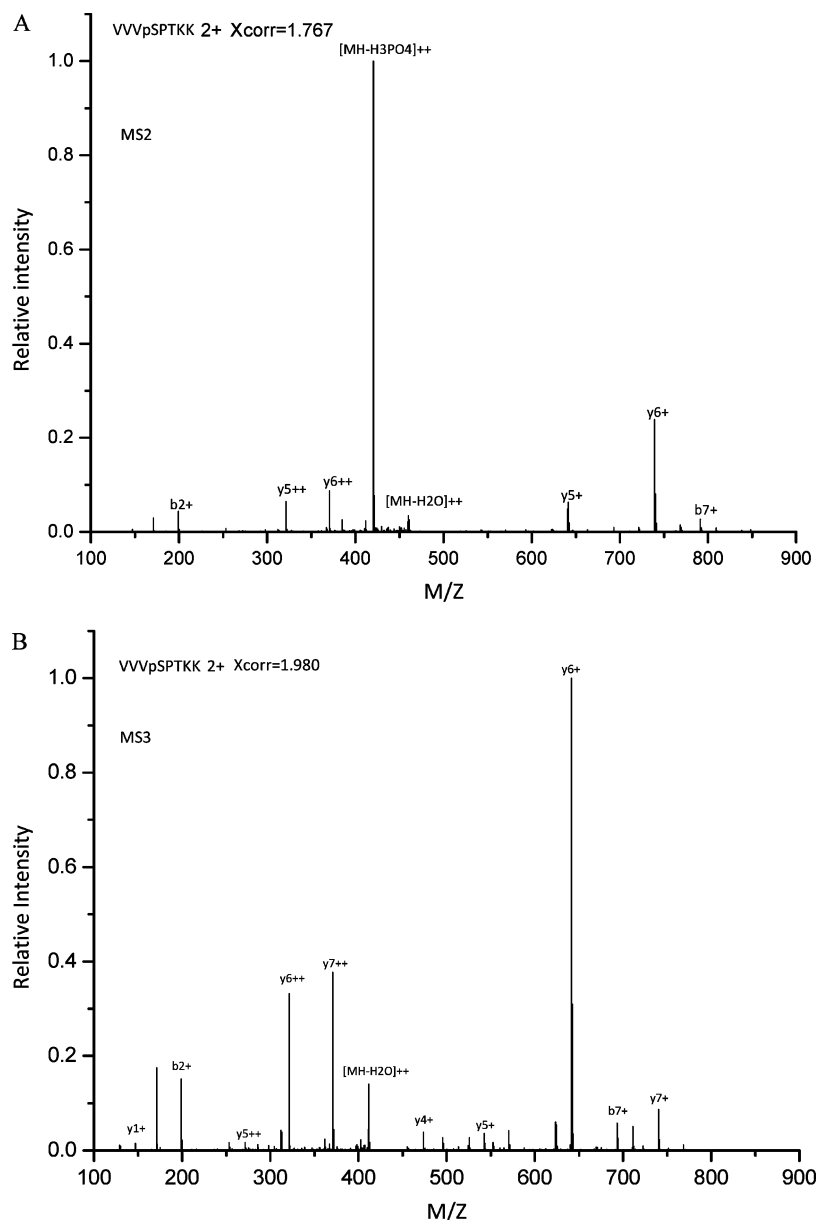
To further investigate the reliability of the identified phosphopeptides and the precision of the phosphorylation sites determined by the MS2/MS3 target-decoy strategy, PhosphoSite (<http://www.phosphosite.org/>) was used to investigate how many phosphorylation sites were also detected by others. Among the 451 phosphorylated sites identified, 77.4% (349 sites) were also reported by others previously, and 22.6% (102 sites) were identified as novel phosphorylation sites (see Supplementary Table 1). Obviously, a majority of the phosphorylation sites were also determined by others, indicating that the phosphorylation sites determined by the MS2/MS3 strategy are of high confidence.

Discussion

Commonly, manual checking was employed to validate the phosphopeptide identifications. It is very time-consuming and labor intensive. It is not fit for large scale phosphoproteome analysis where thousands of phosphopeptide identifications need to be validated, and the phosphopeptide identifications obtained by manual validation were not objective as it largely depends on the experience of the researcher. Thus, data sets of phosphopeptide identifications generated by different laboratories using manual checking are often not comparable. The FDR for phosphopeptide identifications obtained by manual checking is often difficult to assess. To overcome the aforementioned limitations, a new approach termed as the MS2/MS3 target-decoy search approach was developed in this study to automatically validate the phosphopeptide identifications. In this approach, the results of MS2 and MS3 target-decoy searches were combined, and then the suitable filter was used to generate phosphopeptide identifications with desirable FDR. This approach is fully automated, and no human intervention is required as long as the filter criteria are set. It was also demonstrated that the phosphopeptides identified by this approach are very confident, and the determined FDR value reflects the actual FDR.

Compared with unmodified peptides, the database search for phosphopeptide identification is very slow because variable





**Figure 5.** MS2 and MS3 spectra for a phosphopeptide identified with (A) MS2 and (B) MS3 by the MS2/MS3 target-decoy searching strategy. Identified phosphopeptide, VVVpSPTKK; charge state, 2+; Xcorr values, 1.767 (MS2) and 1.980 (MS3). Evidently, this phosphopeptide would not be identified in a conventional approach (Xcorr cutoff in the method was commonly set to 2.5 for 2+ peptides).

modifications were set to S, T, Y residues. In conventional approaches, all MS2 and MS3 spectra were submitted to a database search, and because the charge state of the precursor is not determined, all spectra were searched twice with different charge states (2+ and 3+). Therefore, this approach is very time-consuming. However, for the MS2/MS3 target-decoy search approach, as the charge state of the precursor is already determined, only spectra pairs with a specific charge state and with a significant neutral loss peak were submitted to the database search. This new approach should be very time saving. In this study, there were 35 675 MS2 and MS3 spectra, among which only 4522 valid MS2/MS3 spectra pairs were reserved after the removal of invalid DTA files. This means that 71 350 spectra (two times the total number of acquired spectra because of the two possible charge states) were subjected to database searching in the conventional approach, while only 9044 spectra were searched in the MS2/MS3 strategy. As only

about 12.7% of data were submitted to database searching, a significant reduction of the database searching time was achieved. And though a small portion of spectra were submitted to the database search, the number of identified unique phosphopeptides by the new approach was increased by 2.5 times compared with that all MS2 spectra were submitted to database searching using the MS2 target-decoy search approach. This indicated the new approach significantly improved the sensitivity for phosphopeptide identifications with the target-decoy search.

Though phosphopeptide enrichment methods have experienced significant progress in recent years, the specificity of these methods is not high enough to only isolate phosphopeptides. Thus, there are still a lot of nonphosphopeptides presented in the enriched phosphopeptide sample.<sup>33</sup> After LC-MS/MS analysis, many MS spectra of nonphosphopeptides will be acquired. If these spectra were submitted to a database

search along with the phosphopeptide spectra, more false positive identifications will be generated. To reduce false positive phosphopeptide assignments, a dominant neutral loss peak was often used to confirm phosphopeptide identifications.<sup>20–22</sup> However, the dominant neutral loss peak only indicates the corresponding peptide phosphorylated, it does not provide any information on the sequence of the phosphopeptide. But when the neutral loss peak is further fragmented to generate the MS3 spectrum, the resulting MS3 spectrum contains a lot of sequence information. Therefore, use of MS3 spectra for the validation of phosphopeptides should be more effective than use of neutral loss peaks. Instead of searching all MS2 spectra, a good approach to reduce false positive identifications is to remove the MS2 spectra of nonphosphopeptides prior to the database search. In the MS2/MS3 target-decoy approach, only valid MS2/MS3 spectra pairs were submitted to the database search. The majority of the MS2 spectra was discarded, among which some were noises, some were nonphosphopeptides, and some were of low quality. After removing these MS2 spectra, the random assignments should be significantly reduced which further increased the confidence for phosphopeptide identifications by the MS2/MS3 target-decoy search approach.

MS2 spectra for phosphopeptides often lack enough fragment peaks due to neural loss, and MS3 spectra for neural loss peaks are often of low ion intensity because of the limited ion counts for the third stage MS. Phosphopeptides derived from these low quality spectra were therefore often assigned with low scores and cannot be identified by a conventional approach. Therefore, MS3 was often used as a complement for MS2 in most of the previous works. However, in the MS2/MS3 target-decoy search approach, because the filters were set at the last step for phosphopeptide identifications and most of the random matches or nonphosphopeptides can be removed at the combination step as they were less likely to be identified by both MS2 and MS3, a looser filter can be used to generated high confidence identification; therefore, higher sensitivity was achieved. It was found that the MS2/MS3 approach allowed more poor spectra to be identified as phosphopeptides if both MS2 and MS3 spectra were available for the phosphopeptide. For example, the MS2 and MS3 spectra of a phosphopeptide with a charge state of 2+ are shown in Figure 5. Xcorr values for this phosphopeptide identification in MS2 and MS3 were 1.77 and 1.98, respectively. As the scores were too low, it cannot be identified by a conventional approach based only on MS2 searching or MS3 searching. However, it was able to be identified by the MS2/MS3 target-decoy search approach.

SEQUEST is one of the most widely used database search algorithms in shotgun proteomics. Compared with other algorithms such as Mascot,<sup>34</sup> more scores are defined by SEQUEST after a database search, such as Xcorr,  $\Delta$ Cn, etc. And because all of these scores provide discrimination power, setting proper criteria becomes more complicated, especially for peptide identifications with variable modifications. Here, the MS2/MS3 target-decoy strategy was successfully developed to process SEQUEST search results; however, a similar strategy could also be developed to process the search results obtained by other algorithms such as Mascot.

During the preparation of this manuscript, Nesvizhskii et al.<sup>29</sup> presented a statistical model for adjusting peptide identification probabilities based on the combined information obtained by coupling peptide assignments of consecutive MS2 and MS3 spectra. Their model was also applied to improve the confi-

dence of phosphopeptide identification. Because we cannot access the algorithm implemented in their strategy, direct comparison was not able to be made. However, there are several advantages for the MS2/MS3 target-decoy strategy developed in this study: First, it considered the phosphopeptides that were not identified as the top ranked match which improved the sensitivity for phosphopeptide identification. It was found that there were 7% phosphopeptides not identified with the top 1 rank in search results of either MS2 or MS3 spectra. Second, the charge state for a precursor ion was determined before database searching. The DTA spectra with the wrong charge states, MS2 spectra with no corresponding MS3 spectra, and MS2/MS3 spectra pairs with a neutral loss peak less than 50% of the base peak in intensity for MS2 are removed, thus the time for database searching was greatly reduced (only 12% searching time in this study). Third, the FDR of the final phosphopeptide identifications was easily determined by the target-decoy strategy, which should be more reliable compared with empirical based statistical models. Fourth, Tscore was defined to determine the most possible phosphorylation site localizations. As it combined the fragment information from both MS2 and MS3, phosphorylation site localization should be determined more precisely.

## Conclusion

Here we have introduced an automatic approach, termed the MS2/MS3 target-decoy search approach, for validation of phosphopeptide identifications. This approach combined the results of MS2 and MS3 target-decoy searches and enabled identification of phosphopeptides with specific FDR. The combination of MS2 and neutral loss MS3 greatly reduced random matches and improved the confidence level in phosphopeptide identifications. The approach includes steps of charge state evaluation, invalid data removal, search results combination, and phosphorylation site determination. It is fully automated and provides great convenience for phosphopeptide identification, especially with low accuracy mass spectrometry. It is very suitable for processing data sets obtained for large scale phosphoproteome analysis.

**Abbreviations:** FDR, false discovery rate; Xcorr, SEQUEST cross-correlation score;  $\Delta$ Cn, SEQUEST delta correlation number.

**Acknowledgment.** Financial support from the National Natural Sciences Foundation of China (No. 20675081, 20735004), the China State Key Basic Research Program Grant (2005CB522701, 2007CB914102), the China High Technology Research Program Grant (2006AA02A309), the Knowledge Innovation program of CAS (KJCX2.YW.HO9), and the Knowledge Innovation program of DICP to H.Z. and National Natural Sciences Foundation of China (No. 20605022, 90713017) to M.Y. is gratefully acknowledged. We also thank Dr. Daniel Figeys at the University of Ottawa for his valuable comments and suggestions.

**Supporting Information Available:** Tables listing the identified phosphopeptides by the MS2/MS3 strategy, MS2 or MS3. This material is available free of charge via the Internet at <http://pubs.acs.org>.

## References

- (1) Hunter, T. Signaling - 2000 and beyond. *Cell* **2000**, *100* (1), 113–127.
- (2) Brill, L. M.; Salomon, A. R.; Ficarro, S. B.; Mukherji, M.; Stettler-Gill, M.; Peters, E. C. Robust phosphoproteomic profiling of

- tyrosine phosphorylation sites from human T cells using immobilized metal affinity chromatography and tandem mass spectrometry. *Anal. Chem.* **2004**, 76 (10), 2763–2772.
- (3) Ficarro, S. B.; McClelland, M. L.; Stukenberg, P. T.; Burke, D. J.; Ross, M. M.; Shabanowitz, J.; Hunt, D. F.; White, F. M. Phosphoproteome analysis by mass spectrometry and its application to *Saccharomyces cerevisiae*. *Nat. Biotechnol.* **2002**, 20 (3), 301–305.
  - (4) Gruhler, A.; Olsen, J. V.; Mohammed, S.; Mortensen, P.; Faergeman, N. J.; Mann, M.; Jensen, O. N. Quantitative phosphoproteomics applied to the yeast pheromone signaling pathway. *Mol. Cell. Proteomics* **2005**, 4 (3), 310–327.
  - (5) Villen, J.; Beausoleil, S. A.; Gerber, S. A.; Gygi, S. P. Large-scale phosphorylation analysis of mouse liver. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, 104 (5), 1488–1493.
  - (6) Beausoleil, S. A.; Jedrychowski, M.; Schwartz, D.; Elias, J. E.; Villen, J.; Li, J. X.; Cohn, M. A.; Cantley, L. C.; Gygi, S. P. Large-scale characterization of HeLa cell nuclear phosphoproteins. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, 101 (33), 12130–12135.
  - (7) Olsen, J. V.; Blagoev, B.; Gnäd, F.; Macek, B.; Kumar, C.; Mortensen, P.; Mann, M. Global, in vivo, and site-specific phosphorylation dynamics in signaling networks. *Cell* **2006**, 127, 635–648.
  - (8) Pinkse, M. W. H.; Uitto, P. M.; Hilhorst, M. J.; Ooms, B.; Heck, A. J. R. Selective isolation at the femtomole level of phosphopeptides from proteolytic digests using 2D-nanoLC-ESI-MS/MS and titanium oxide precolumns. *Anal. Chem.* **2004**, 76, 3935–3943.
  - (9) Larsen, M. R.; Thingholm, T. E.; Jensen, O. N.; Roepstorff, P.; Jorgensen, T. J. D. Highly selective enrichment of phosphorylated peptides from peptide mixtures using titanium dioxide microcolumns. *Mol. Cell. Proteomics* **2005**, 4, 873–886.
  - (10) Zhou, H. J.; Tian, R. J.; Ye, M. L.; Xu, S. Y.; Feng, S.; Pan, C. S.; Jiang, X. G.; Li, X.; Zou, H. F. Highly specific enrichment of phosphopeptides by zirconium dioxide nanoparticles for phosphoproteome analysis. *Electrophoresis* **2007**, 28 (23), 2201–2215.
  - (11) Trinidad, J. C.; Specht, C. G.; Thalhammer, A.; Schoepfer, R.; Burlingame, A. L. Comprehensive identification of phosphorylation sites in postsynaptic density preparations. *Mol. Cell. Proteomics* **2006**, 5 (5), 914–922.
  - (12) Lee, J.; Xu, Y. D.; Chen, Y.; Sprung, R.; Kim, S. C.; Xie, S. H.; Zhao, Y. M. Mitochondrial phosphoproteome revealed by an improved IMAC method and MS/MS/MS. *Mol. Cell. Proteomics* **2007**, 6 (4), 669–676.
  - (13) Aebersold, R.; Mann, M. Mass spectrometry-based proteomics. *Nature* **2003**, 422 (6928), 198–207.
  - (14) Steen, H.; Mann, M. The ABC's (and XYZ's) of peptide sequencing. *Nat. Rev. Mol. Cell Biol.* **2004**, 5, 699–711.
  - (15) DeGnove, J. P.; Qin, J. Fragmentation of phosphopeptides in an ion trap mass spectrometer. *J. Am. Soc. Mass Spectrom.* **1998**, 9 (11), 1175–1188.
  - (16) Beausoleil, S. A.; Villen, J.; Gerber, S. A.; Rush, J.; Gygi, S. P. A probability-based approach for high-throughput protein phosphorylation analysis and site localization. *Nat. Biotechnol.* **2006**, 24 (10), 1285–1292.
  - (17) Elias, J. E.; Gygi, S. P. Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nat. Methods* **2007**, 4 (3), 207–214.
  - (18) Higdon, R.; Kolker, E. A predictive model for identifying protein by a single peptide match. *Bioinformatics* **2007**, 23 (3), 277–280.
  - (19) Jiang, X. N.; Jiang, X. G.; Han, G. H.; Ye, M. L.; Zou, H. F. Optimization of filtering criterion for SEQUEST database searching to improve proteome coverage in shotgun proteomics. *BMC Bioinformatics* **2007**, 8, 323.
  - (20) Yang, F.; Stenoien, D. L.; Strittmatter, E. F.; Wang, J. H.; Ding, L. H.; Lipton, M. S.; Monroe, M. E.; Nicora, C. D.; Gristenko, M. A.; Tang, K. Q.; Fang, R. H.; Adkins, J. N.; Camp, D. G.; Chen, D. J.; Smith, R. D. Phosphoproteome profiling of human skin fibroblast cells in response to low- and high-dose irradiation. *J. Proteome Res.* **2006**, 5 (5), 1252–1260.
  - (21) Pan, C. S.; Ye, M. L.; Liu, Y. G.; Feng, S.; Jiang, X. G.; Han, G. H.; Zhu, J. J.; Zou, H. F. Enrichment of phosphopeptides by Fe<sup>3+</sup>-immobilized mesoporous nanoparticles of MCM-41 for MALDI and nano-LC-MS/MS analysis. *J. Proteome Res.* **2006**, 5 (11), 3114–3124.
  - (22) Feng, S.; Pan, C. S.; Jiang, X. G.; Xu, S. Y.; Zhou, H. J.; Ye, M. L.; Zou, H. F. Fe<sup>3+</sup> immobilized metal affinity chromatography with silica monolithic capillary column for phosphoproteome analysis. *Proteomics* **2007**, 7 (3), 351–360.
  - (23) Lu, B. W.; Ruse, C.; Xu, T.; Park, S. K.; Yates, J. Automatic validation of phosphopeptide identifications from tandem mass spectra. *Anal. Chem.* **2007**, 79 (4), 1301–1310.
  - (24) Olsen, J. V.; Mann, M. Improved peptide identification in proteomics by two consecutive stages of mass spectrometric fragmentation. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, 101 (37), 13417–13422.
  - (25) Feng, S.; Ye, M. L.; Zhou, H. J.; Jiang, X. G.; Jiang, X. N.; Zou, H. F.; Gong, B. L. Immobilized zirconium ion affinity chromatography for specific enrichment of phosphopeptides in phosphoproteome analysis. *Mol. Cell. Proteomics* **2007**, 6, 1656–1665.
  - (26) Eng, J. K.; McCormack, A. L.; Yates, J. R., III. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J. Am. Soc. Mass Spectrom.* **1994**, 5 (11), 976–989.
  - (27) Keller, A.; Nesvizhskii, A. I.; Kolker, E.; Aebersold, R. Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. *Anal. Chem.* **2002**, 74 (20), 5383–5392.
  - (28) Xie, H. W.; Griffin, T. J. Trade-off between high sensitivity and increased potential for false positive peptide sequence matches using a two-dimensional linear ion trap for tandem mass spectrometry-based proteomics. *J. Proteome Res.* **2006**, 5 (4), 1003–1009.
  - (29) Ulintz, P. J.; Bodenmiller, B.; Andrews, P. C.; Aebersold, R.; Nesvizhskii, A. I. Investigating MS2-MS3 matching statistics: A model for coupling consecutive stage mass spectrometry data for increased peptide identification confidence. *Mol. Cell. Proteomics*, mcp. M700128-MCP200.
  - (30) Hunter, T.; Sefton, B. M. Transforming Gene-Product of Rous-Sarcoma Virus Phosphorylates Tyrosine. *Proc. Natl. Acad. Sci. U.S.A.* **1980**, 77 (3), 1311–1315.
  - (31) Ishihama, Y.; Wei, F. Y.; Aoshima, K.; Sato, T.; Kuromitsu, J.; Oda, Y. Enhancement of the efficiency of phosphoproteomic identification by removing phosphates after phosphopeptide enrichment. *J. Proteome Res.* **2007**, 6 (3), 1139–1144.
  - (32) Bodenmiller, B.; Mueller, L. N.; Mueller, M.; Domon, B.; Aebersold, R. Reproducible isolation of distinct, overlapping segments of the phosphoproteome. *Nat. Methods* **2007**, 4 (3), 231–237.
  - (33) Dai, J.; Jin, W. H.; Sheng, Q. H.; Shieh, C. H.; Wu, J. R.; Zeng, R. Protein phosphorylation and expression profiling by Yin-yang multidimensional liquid chromatography (Yin-yang MDLC) mass spectrometry. *J. Proteome Res.* **2007**, 6 (1), 250–62.
  - (34) Perkins, D. N.; Pappin, D. J. C.; Creasy, D. M.; Cottrell, J. S. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **1999**, 20 (18), 3551–3567.

PR700675J