

RESEARCH ARTICLE

Optimization of TripleTOF spectral simulation and library searching for confident localization of phosphorylation sites

Ayano Takai¹, Tomoya Tsubosaka¹, Yasuhiro Hirano¹, Naoki Hayakawa¹, Fumitaka Tani¹, Pekka Haapaniemi², Veronika Suni^{2,3}, Susumu Y. Imanishi^{1*}

1 Faculty of Pharmacy, Meijo University, Nagoya, Japan, **2** Turku Bioscience Centre, University of Turku and Åbo Akademi University, Turku, Finland, **3** Turku Centre for Computer Science, Turku, Finland

* susima@meijo-u.ac.jp



OPEN ACCESS

Citation: Takai A, Tsubosaka T, Hirano Y, Hayakawa N, Tani F, Haapaniemi P, et al. (2019) Optimization of TripleTOF spectral simulation and library searching for confident localization of phosphorylation sites. *PLoS ONE* 14(12): e0225885. <https://doi.org/10.1371/journal.pone.0225885>

Editor: Frederique Lisacek, Swiss Institute of Bioinformatics, SWITZERLAND

Received: September 21, 2019

Accepted: November 14, 2019

Published: December 2, 2019

Copyright: © 2019 Takai et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: Raw mass spectrometry data, protein sequence database, spectral libraries, and search results have been deposited to the jPOST Repository (<https://repository.jpostdb.org>) with the data set identifier JPST000548 (ProteomeXchange identifier: PXD012433).

Funding: This work was supported by the JSPS KAKENHI (16K08206 and 19K07017 to SYI; <https://www.jps.go.jp>) and the Research Institute

Abstract

Tandem mass spectrometry (MS/MS) has been used in analysis of proteins and their post-translational modifications. A recently developed data analysis method, which simulates MS/MS spectra of phosphopeptides and performs spectral library searching using SpectraST, facilitates confident localization of phosphorylation sites. However, its performance has been evaluated only on MS/MS spectra acquired using Orbitrap HCD mass spectrometers so far. In this study, we have investigated whether this approach would be applicable to another type of mass spectrometers, and optimized the simulation and search conditions to achieve sensitive and confident site localization. Synthetic phosphopeptides and enriched K562 cell phosphopeptides were analyzed using a TripleTOF 6600 mass spectrometer before and after enzymatic dephosphorylation. Dephosphorylated peptides identified by X! Tandem database searching were subjected to spectral simulation of all possible single phosphorylations using SimPhospho software. Phosphopeptides were identified and localized by SpectraST searching against a library of the simulated spectra. Although no synthetic phosphopeptide was localized at 1% false localization rate under the previous conditions, optimization of the spectral simulation and search conditions for the TripleTOF datasets achieved the localization and improved the sensitivity. Furthermore, the optimized conditions enabled sensitive localization of K562 phosphopeptides at 1% false discovery and localization rates. These results suggest that accurate phosphopeptide simulation of TripleTOF MS/MS spectra is possible and the simulated spectral libraries can be used in SpectraST searching for confident localization of phosphorylation sites.

Introduction

Tandem mass spectrometry (MS/MS) and sequence database searching have widely been used in the field of proteomics [1–3]. The recent advances in those techniques, and also in sample preparation, have enabled identification and quantification of thousands of phosphopeptides [4–8]. Furthermore, assisting tools such as Ascore and phosphoRS have been developed for confident localization of phosphorylation sites [9–13]. Since reversible protein phosphorylation at

of Meijo University (SYI; <https://www.meijo-u.ac.jp>). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

specific serine, threonine, tyrosine residues is a key determinant in many cellular functions [14–22], many phosphoproteomics studies have been conducted [4–8, 23–25].

In addition to database searching, spectral library searching has been available for peptide identification [26–29]. Since spectral library searching matches MS/MS spectra of peptides to the ones previously identified, it takes advantage of comparison of specific spectral features including peak intensities and neutral losses from fragments. Furthermore, open modification searches offer identifications of various post-translational modifications on library spectra [30–33]. Spectral library searching has been proven to be effective also in identification of phosphopeptides [34–36]. Hu et al. has reported computational simulation of known phosphorylation sites onto nonphosphorylated peptides, which was performed by a simple +80 Da shift of fragment ions generated by ion trap collision-induced dissociation (CID) [37, 38]. Searching by SpectraST software [28] against their spectral library supplemented with the simulated spectra was highly sensitive for phosphopeptide identifications compared to database searching [38]. However, accuracy of phosphorylation site localization by this approach was uncertain, as many false localizations appeared in our evaluation [39].

Recently, we have developed a phosphorylation site localization method using spectral library searching. It was demonstrated that, although a prominent neutral loss of phosphoric acid from phosphorylated serine/threonine residues readily occurs, fragmentation patterns of serine/threonine/tyrosine phosphorylated peptides and their dephosphorylated counterparts are similar in beam-type CID spectra acquired in a quadrupole time-of-flight (Q-TOF) instrument [40]. Based on this, a software tool SimPhospho was developed for beam-type CID spectra acquired in Orbitrap higher energy collisional dissociation (HCD) instruments, which enables spectral simulation of all possible single phosphorylations by predicting the diagnostic fragmentation events onto enzymatically dephosphorylated peptides (Fig 1) [39]. SpectraST searching against a spectral library of the simulated phosphopeptides achieved sensitive detection and confident localization of phosphorylation sites, even if the library was merged with publicly available large libraries. Furthermore, this simulated spectral library searching could be used to supplement database search results prior to label-free quantification [23]. Very recently, an improved version of SimPhospho has been developed with a user interface [41].

Here, we have investigated the applicability of simulated spectral library searching to beam-type CID spectra obtained using a TripleTOF 6600 Q-TOF mass spectrometer (refer to Fig 2). TripleTOF mass spectrometers are capable of high-speed MS/MS (maximum 100 spectra/second), which allows comprehensive and specific peptide quantification by data-independent acquisition, such as SWATH-MS [42]. However, also in quantitative phosphoproteomics studies, confident and sensitive identification of phosphopeptides is required prior to quantification. In this study, we have acquired TripleTOF CID spectra of synthetic phosphopeptides and K562 cell tryptic phosphopeptides before and after enzymatic dephosphorylation, and performed spectral simulation and searching for those data using SimPhospho and SpectraST software. However, this approach developed with Orbitrap HCD spectra was not readily applicable to TripleTOF CID spectra even though both fragmentation patterns appear to be similar. Therefore, we have evaluated different scoring to analyze TripleTOF CID spectra and optimized simulation and search conditions. The performance of the optimized searching was evaluated in comparison to MaxQuant searching [43].

Materials and methods

Sample preparation

Human phosphopeptide samples were prepared according to Kauko et al. [23] with some modifications. Briefly, K562 cell lysate (1 mg protein) obtained from Promega was mixed with

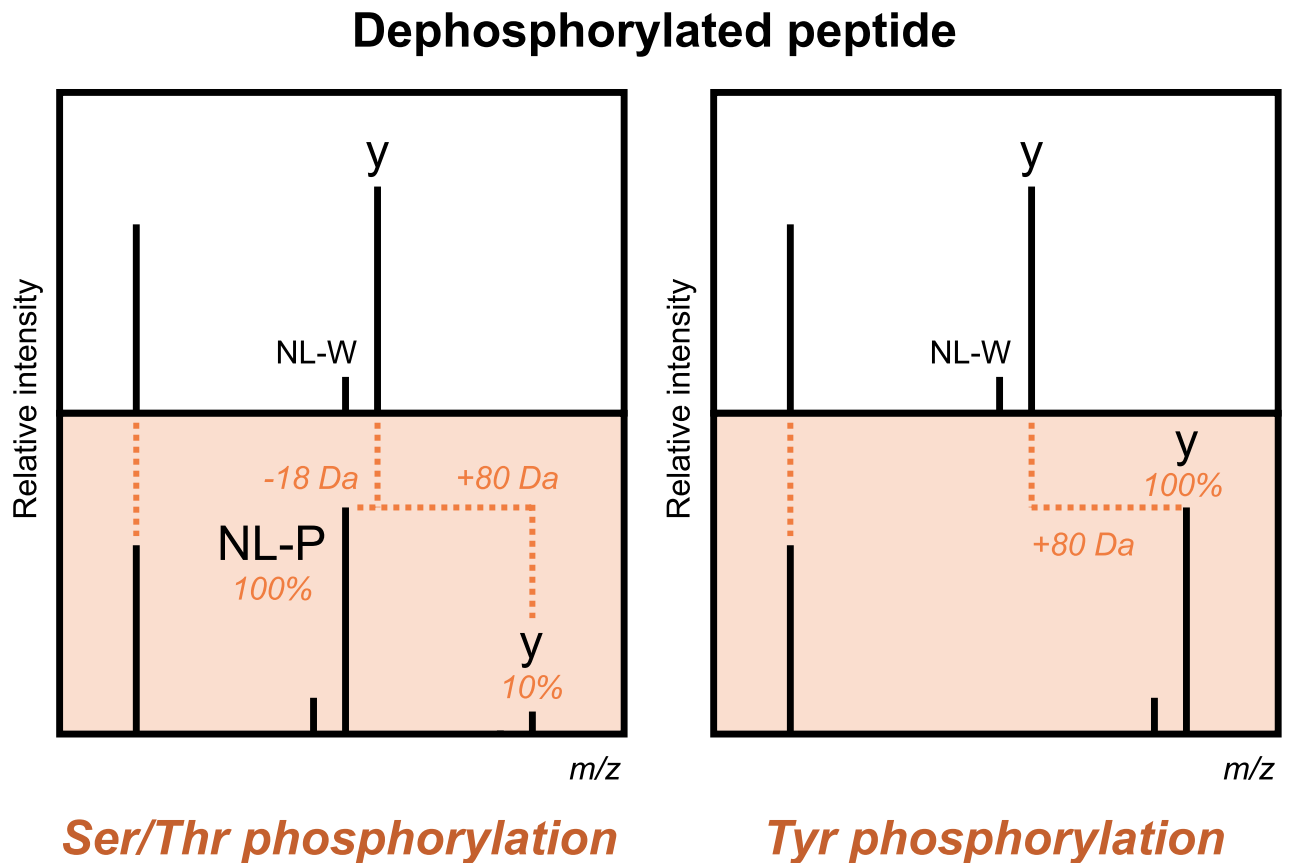


Fig 1. Simulation of single phosphorylation on beam-type CID spectra. From nonphosphorylated fragment ions observed in beam-type CID spectra of enzymatically dephosphorylated peptides, such as a, b, and y-series ions and their ammonia and water neutral loss ions (NL-A and -W), SimPhospho predicts serine/threonine phosphorylated ions (pS/pT intact ions) and their phosphoric acid neutral loss ions (NL-P ions) by a +80 Da shift and a -18 Da shift, respectively. Note, fragment ions with the combination of NL-A and NL-W were not taken into consideration. Tyrosine phosphorylated ions (pY intact ions) are predicted from the nonphosphorylated ions only by a +80 Da shift. In our previous study [39], intensities of pS/pT intact ions, NL-P ions, and pY intact ions relative to the nonphosphorylated ions were set to 10%, 100%, and 100%, respectively.

<https://doi.org/10.1371/journal.pone.0225885.g001>

a phosphoprotein standard bovine α -casein (10 μ g, Sigma-Aldrich), reduced with dithiothreitol, alkylated with iodoacetamide, and then digested with trypsin. After acidification with trifluoroacetic acid, aliquots of the digest (20 of 1000 μ L) were desalted with a microcolumn packed with Empore C18 disk (3M) [44], for checking the digestion. The remaining digests (980 μ L) were desalted with an Empore C18-SD 10mm/6mL cartridge (3M), followed by phosphopeptide enrichment with a microcolumn packed with Sachtopore-NP TiO₂ beads (20 μ m, 300 \AA ; ZirChrom) [40]. The enriched phosphopeptides were immediately desalted with the C18 microcolumn. Half of the phosphopeptide sample was subjected to enzymatic dephosphorylation with calf intestinal alkaline phosphatase (Roche), followed by desalting with the C18 microcolumn. Both the samples before and after the dephosphorylation were dissolved with 20 μ L of 0.1% formic acid.

Synthetic phosphopeptides (PEPotec) containing 62 singly phosphorylated peptides (24 sequences) were obtained from Thermo Fisher Scientific. These peptides were mixed into 3 groups, where the peptide concentration was 200 pmol each based on the manufacturer information and phosphopeptide isoforms (different phosphorylation sites on the same peptide sequences) were separated from each other (see S1 Table). The peptide mixtures were desalted

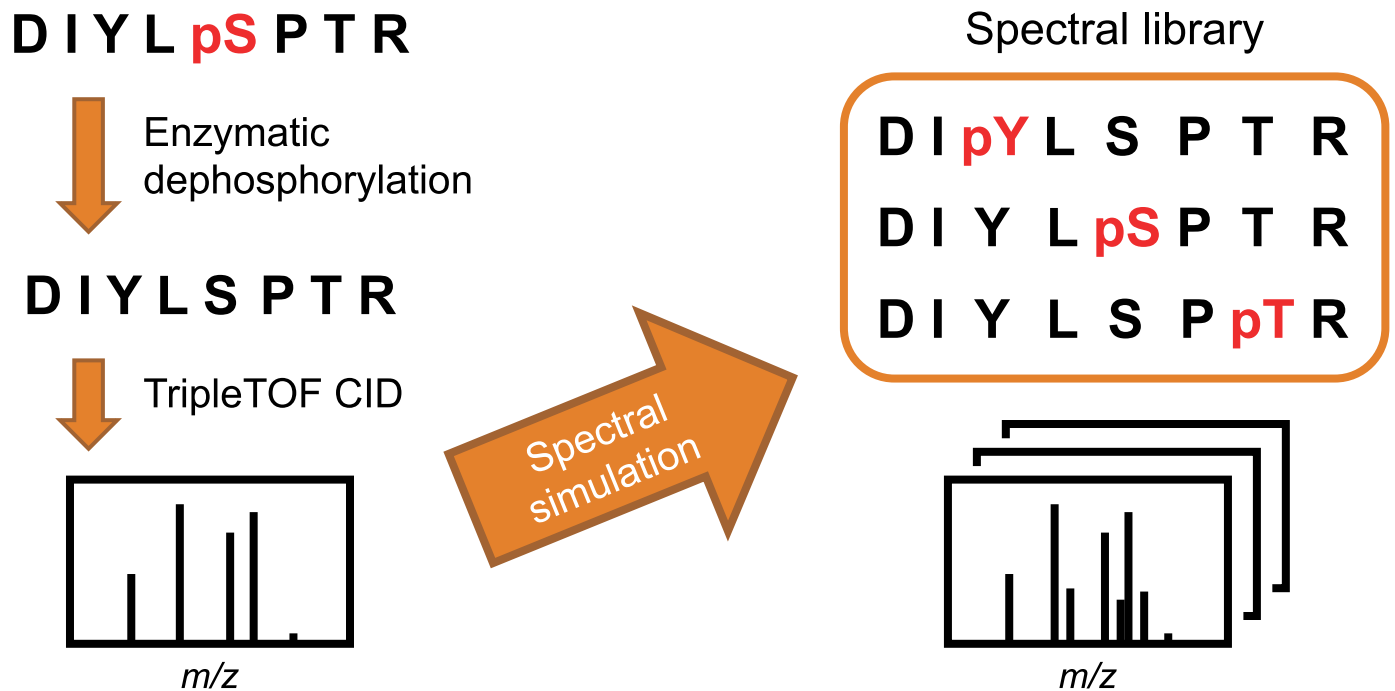


Fig 2. Experimental scheme of phosphopeptide simulation on TripleTOF CID spectra. SimPhospho enables spectral simulation of all possible single phosphorylations. A spectral library of simulated phosphopeptides can be used in SpectraST searching. pS: phosphoserine, pT: phosphothreonine, and pY: phosphotyrosine.

<https://doi.org/10.1371/journal.pone.0225885.g002>

before and after enzymatic dephosphorylation as described above. Both the mixtures were dissolved with 100 μ L of 0.1% formic acid.

Mass spectrometry

LC-MS/MS analysis for the K562 peptides (5 μ L injection) was performed in duplicates, using an Eksigent Ekspert 425 cHiPLC nanoflow LC instrument coupled to a TripleTOF 6600 Q-TOF mass spectrometer (Sciex). A 200 μ m x 0.5 mm trap column and a 75 μ m x 15 cm analytical column (ChromXP C18-CL, 3 μ m, 120 \AA ; Sciex) were used. The mobile phases were 0.1% formic acid in water (A) and 0.1% formic acid in acetonitrile (B). A flow rate was 300 nL/min. LC gradient elution condition was initially 5% B to 30% B (90 min) and 80% B (95–105 min). Data dependent acquisition was performed in positive ion mode. MS spectra were acquired from m/z 400 to m/z 1250 with an accumulation time of 0.1 s. The 10 most abundant ions of which charge states were from 2+ to 4+ were selected for subsequent fragmentation with rolling collision energy ($0.0625 \times m/z - X$ (V); X = 3 (2+), 5 (3+), or 6 (4+)). MS/MS spectra were acquired from m/z 100 to m/z 1600 with an accumulation time of 0.1 s in high sensitivity mode. Exclusion time was 20 s.

LC-MS/MS analysis for the synthetic peptides was performed in triplicates as described above, but LC gradient elution condition was initially 5% B to 40% B (10 min) and 80% B (12–20 min). The accumulation time for MS/MS and the exclusion time were modified into 0.2 s and 0 s, respectively.

X!Tandem database search

TripleTOF data (wiff and wiffscan files) of the dephosphorylated peptides (K562 and synthetic peptides) were converted into mzXML files with Trans-Proteomic Pipeline (TPP) [45, 46]

version 4.8.0. These files were subjected to database searching with X!Tandem [47] included in TPP, against a concatenated forward-reverse SwissProt database (2010_03 release, *Homo sapiens*, total 40530 sequences) used in the previous study [39]. The following search condition was used: carbamidomethylation (C) as a fixed modification, and oxidation (M), phosphorylation (S, T, and Y), acetylation (protein N-terminus), and pyroglutamine (E, Q, and carbamidomethyl C) as variable modifications. Trypsin was specified as an enzyme and two missed cleavage sites were allowed. Precursor and fragment ion mass tolerances were set to 50 ppm and 0.05 Da, respectively. Probability values and false discovery rates (FDRs) were estimated using the target-decoy strategy by PeptideProphet [48] included in TPP. The searches were also tested against a SwissProt database (2018_02 release, *Homo sapiens*) supplemented with common contaminants (forward-reverse, total 40762 sequences).

Spectral library creation

The X!Tandem search results (pepXML files) and the mzXML files of dephosphorylated peptides were subjected to spectral simulation with SimPhospho version 1 [41], which generates new pepXML and mzXML files containing only the phosphopeptides simulated from nonphosphorylated peptides. Consensus spectral libraries were built from the generated files and supplemented with decoy entries using SpectraST [28, 49, 50] version 5.0 included in TPP. For building these libraries, the nonphosphorylated peptides with a minimum probability of 0.95 were taken as the simulation templates. Different simulation conditions were used on SimPhospho and the obtained SpectraST search results were compared (see details in the Result section).

SpectraST spectral library search

TripleTOF data of the K562 and synthetic phosphopeptides were converted into mzXML files with TPP and searched against the corresponding simulated spectral libraries with SpectraST. The following search condition was used: version 4 scoring (based on dot products of square-root intensities), precursor ion mass tolerance 3 m/z , and fragment ion bins 20 per m/z ; or version 5 scoring (rank-based similarity scoring), precursor ion mass tolerance 3 m/z , and fragment ion mass tolerance 0.05 m/z . An FDR of 1% for the K562 phosphopeptides was estimated by PeptideProphet as a cutoff; however, no FDR cutoff was applied for the synthetic phosphopeptides because of the known peptide sequences and data simplicity. To estimate false localization rates (FLRs), deltadot score and F-value were used. Deltadot score reflects how much the first spectral match differs from the second best match in dot product scores and contributes to the total discriminant score F-value [51]; however, deltadot score in the default condition of SpectraST does not consider homologous peptides until the fourth match (HOM4 deltadot score). Therefore, the searches were repeated to calculate HOM1 deltadot score by comparing the first and second matches even if they represent the same peptide sequence with only difference in phosphorylation sites. Narrower precursor ion mass tolerance in SpectraST searching may affect the scoring for FLR calculation (S1 Fig); therefore, in this study we decided to use 3 m/z tolerance as described above, and 0.05 m/z tolerance was subsequently applied to search results prior to the localization. Search results obtained under different simulation conditions were compared (see details in the Result section).

MaxQuant database search and site localization

The TripleTOF data of K562 and synthetic phosphopeptides were searched against the SwissProt 2010_03 database with MaxQuant [43] version 1.5.5.1. The following search condition was used: carbamidomethylation (C) as a fixed modification, and oxidation (M),

phosphorylation (S, T, and Y) and acetylation (protein N-terminus) as variable modifications. Trypsin was specified as an enzyme and two missed cleavage sites were allowed. Precursor and fragment ion mass tolerances were set to 0.05 Da. A 1% FDR cutoff was applied for the K562 phosphopeptides, but not for the synthetic phosphopeptides. PTM score [10] was used for FLR estimation.

Results and discussion

Simulated spectral library searching on synthetic phosphopeptide datasets acquired using TripleTOF instrument

To investigate whether the simulated spectral library approach would be applicable to TripleTOF CID data, we used 62 singly phosphorylated synthetic peptides (62 phosphorylation sites on 24 sequences, refer to [S1 Table](#)). The tryptic phosphopeptides, which tended to be differentially localized/identified from the human HeLa cell dataset reported in our previous study [39], were selected for synthetic phosphopeptides to clearly observe localization performance of simulated spectral library searching. For creating a spectral library of simulated phosphopeptides, the synthetic phosphopeptides were enzymatically dephosphorylated and analyzed by LC-MS/MS using a TripleTOF 6600 instrument. The analysis was performed in triplicates without using exclusion time, resulting in acquisition of many spectral replicates for each peptide. X!Tandem database searching was performed against the sequence database (SwissProt 2010_03 human) used for obtaining the HeLa dataset in the previous study [39]. From the dephosphorylated sample data, 3737 peptide spectra were identified with PeptideProphet minimum probability of 0.95, of which nonphosphorylated peptides (3728 spectra, 0.63% FDR) were subjected to spectral simulation of single phosphorylation using SimPhospho [39, 41]. Since the latest version of SimPhospho offers various options on its user interface [41], different simulation conditions can be used. Serine/threonine phosphorylated fragment ions (pS/pT intact ions), their phosphoric acid neutral loss ions (NL-P ions), and tyrosine phosphorylated fragment ions (pY intact ions) were predicted from nonphosphorylated fragment ions (a, b, and y-series) and their ammonia and water neutral loss ions (NL-A and -W) (refer to [Fig 1](#)). As the default condition, intensities of pS/pT intact ions, NL-P ions, and pY intact ions relative to nonphosphorylated ions (a, b, y, and NL-AW) were set to 10%, 100%, and 100%, respectively, as used in the original study [39]. The simulated spectra were used to build a spectral library (181 phosphopeptide consensus spectra, 181 decoys) using SpectraST. LC-MS/MS analysis of the synthetic phosphopeptides was performed in triplicates as well, and the acquired TripleTOF CID spectra were subjected to SpectraST searching (version 5 scoring) against the library. Because of the known sequences and data simplicity, no FDR cutoff was applied to the search results. Representative TripleTOF spectral matches shown in [Fig 3](#) demonstrate well-predicted fragment ions in simulated spectra.

TripleTOF spectral matches of the synthetic phosphopeptides were sorted by the total discriminant score F-value, and then FLRs across F-value were calculated, as described previously for Orbitrap HCD spectral matches [39]. However, only 1 correct spectral match was observed at 1% FLR ([Fig 4A](#)), which was due to many false localization matches showing high F-value ([Fig 4B](#)). Therefore, the searches were repeated to calculate another discriminant score HOM1 delatadot, of which concept is similar to Mascot delta score [11]. Sorting of the TripleTOF spectral matches by HOM1 delatadot score showed the greater number of correct spectral matches (i.e. true localizations) than the F-value sorting at 0–32% FLRs ([Fig 4A](#)). HOM1 delatadot score of 0.116 as a cutoff satisfied 1% FLR for the search result ([Fig 4C](#)), with 318 correct spectral matches ([Fig 4A](#)).

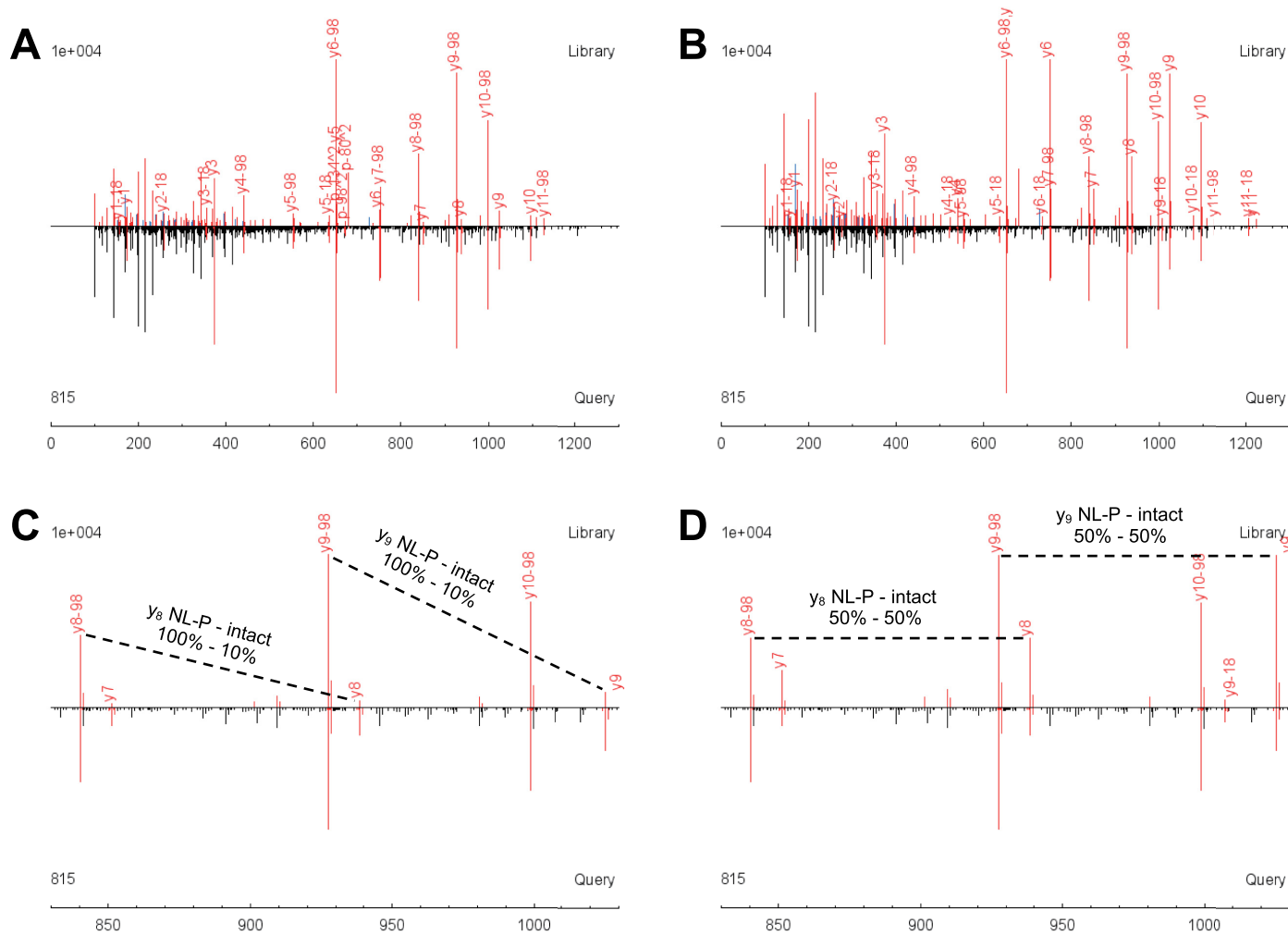
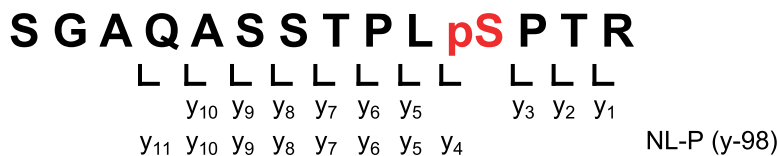


Fig 3. Phosphopeptide simulation on TripleTOF CID spectra. Phosphopeptides were subjected to enzymatic dephosphorylation, TripleTOF CID analysis, and then spectral simulation of single phosphorylations (refer to Fig 2). (A) A representative SpectraST spectral match of a synthetic phosphopeptide (lower spectrum; SGAQASSTPLpSPTR, pS: phosphoserine) with a simulated phosphopeptide (upper spectrum) is shown. Serine-phosphorylated fragment ions (pS intact ions) and their phosphoric acid neutral loss ions (NL-P ions) were predicted from nonphosphorylated ions with 10% and 100% intensities, respectively (refer to Fig 1). (B) The same match is shown, but with the peptide simulated using 50%-50% intensities. (C and D) Fragment ions at *m/z* 830–1030 in the spectral matches A and B are shown, respectively. In the synthetic phosphopeptide spectra, *y*-series ions matched to those of simulated phosphopeptides are shown in red.

<https://doi.org/10.1371/journal.pone.0225885.g003>

To optimize the simulated spectral library search conditions for TripleTOF CID data, the synthetic phosphopeptide spectra were searched using two versions of SpectraST scoring (4 and 5) against 16 spectral libraries created under different simulation conditions. The simulation conditions used were as follows: intensities of pS/pT intact ions, NL-P ions, and pY intact ions were 10%-100%-100%, 50%-100%-100%, 50%-50%-100%, and 50%-50%-50%, and NL conditions were NL-PAW, NL-PA, NL-PW, and NL-P. Total 32 searches were performed, and

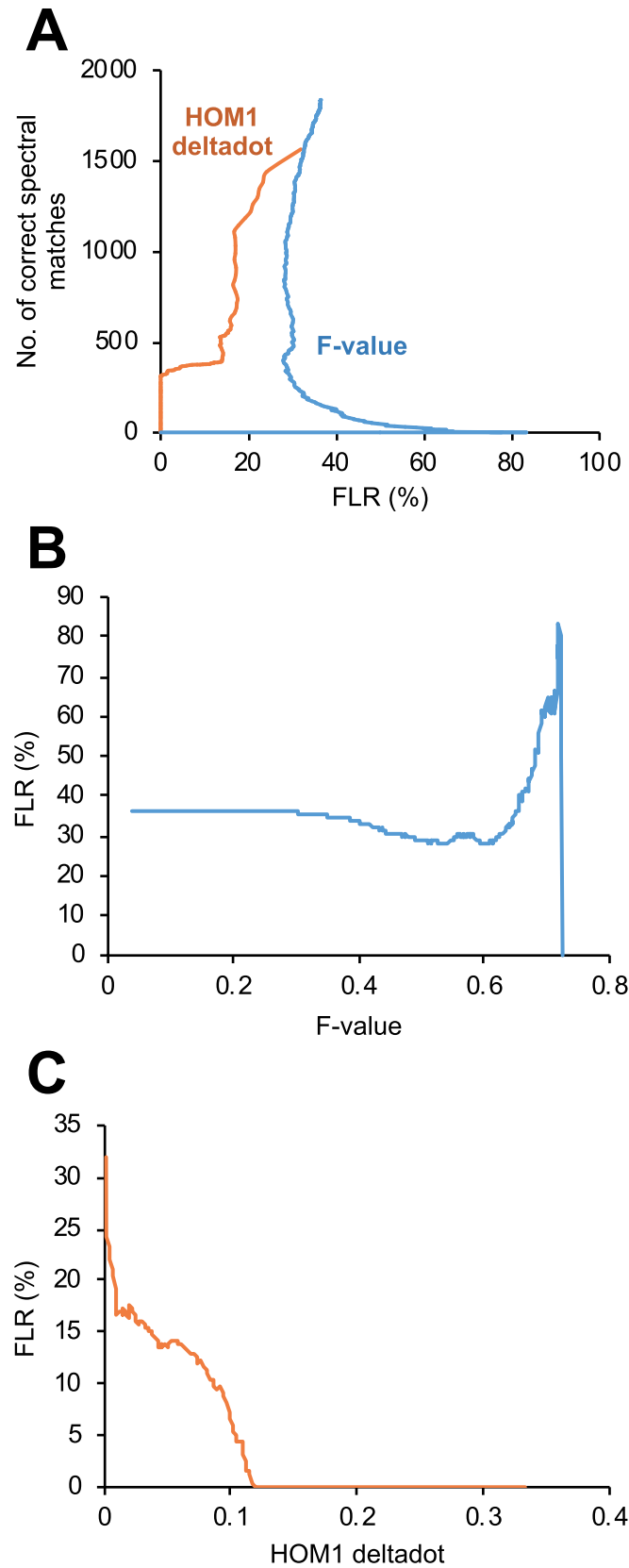


Fig 4. Phosphorylation site localization on synthetic phosphopeptides by simulated spectral library searching. (A) The synthetic phosphopeptides analyzed by TripleTOF CID were searched by SpectraST against the simulated spectral library and the results were sorted by one of the SpectraST scores, F-value or HOM1 deltatdot. (B) FLRs as a function of F-value are shown. (C) FLRs as a function of HOM1 deltatdot score are shown.

<https://doi.org/10.1371/journal.pone.0225885.g004>

spectral matches of the synthetic phosphopeptides were sorted by HOM1 deltatdot score (see [S2 Table](#) for 1% FLR cutoffs). In the number of correct spectral matches at 1% FLR, the searching with version 4 scoring was better than the newest scoring version 5 for all tested simulation conditions ([Fig 5](#)). In particular, both 50%-50%-100% and 50%-50%-50% with NL-P showed the highest numbers at 1% FLR, i.e. 893 correct spectral matches (HOM1 deltatdot score of 0.017) and 701 correct spectral matches (HOM1 deltatdot score of 0.046) with the scoring versions 4 and 5, respectively. This result is consistent with the previous evaluation as 50%-50%-50% was the best condition for the Orbitrap HCD dataset of 20 synthetic phosphopeptides [[41](#)]. However, the intensity condition was not enough, and the thorough comparison of the combinations of conditions found that the NL conditions were also critical as 50%-50%-50% with NL-P resulted in 1.8–2.1 times more correct spectral matches than those with NL-PAW used in the previous study [[41](#)]. These results suggest that considering NL-A and NL-W ions to be phosphorylated in the simulation may introduce possible simulation errors, and/or somehow interfere with subsequent spectral searching due to unexpected fragment ion matching. The synthetic phosphopeptide results obtained under one of the optimized conditions (50%-50%-100%, NL-P, version 4) are listed in [S3 Table](#).

Simulated spectral library searching on K562 phosphopeptide datasets acquired using TripleTOF instrument

Simulated spectral library searching for TripleTOF CID data was evaluated on singly phosphorylated peptides obtained from human cell proteins. Phosphopeptides enriched from a

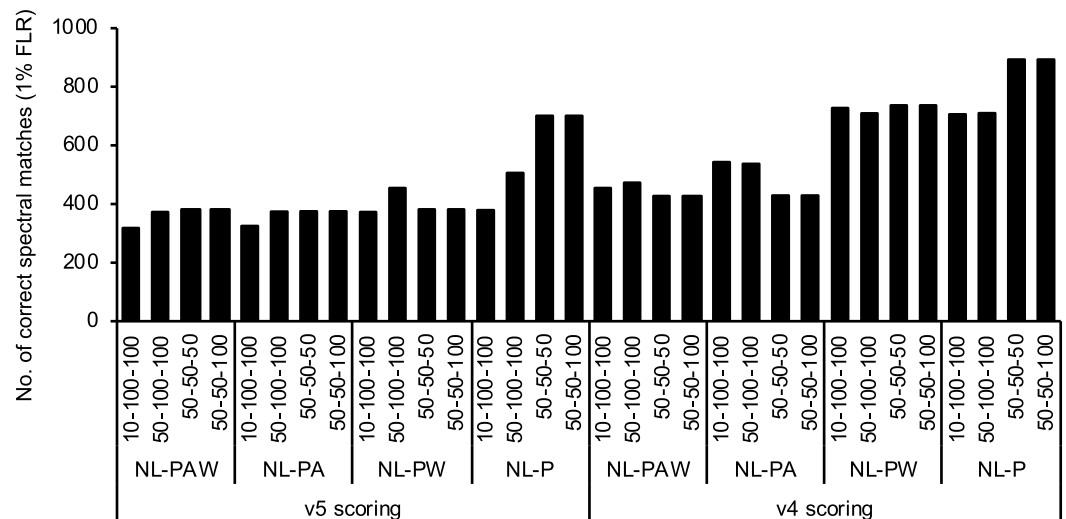


Fig 5. Optimization of simulation conditions for localizing phosphorylation sites on synthetic phosphopeptides. Different simulation conditions were used for creating TripleTOF spectral libraries from the dephosphorylated synthetic peptides. Four different intensities were used for pS/pT intact ions, NL-P ions, and pY intact ions: 10%-100%-100%, 50%-100%-100%, 50%-50%-50%, and 50%-50%-100% (refer to [Fig 3](#) and [[39](#)]). Those ions were predicted from nonphosphorylated fragment ions: a, b, y, and their ammonia (A) and water (W) NL ions. Four different NL conditions were used: NL-PAW, NL-PA, NL-PW, and NL-P. Against the 16 libraries, the synthetic phosphopeptides were searched by SpectraST with scoring versions 4 and 5.

<https://doi.org/10.1371/journal.pone.0225885.g005>

K562 tryptic digest and their enzymatically dephosphorylated peptides were analyzed by TripleTOF LC-MS/MS in duplicates using dynamic exclusion. X!Tandem searching of the dephosphorylated sample data identified 13190 human peptide spectra with PeptideProphet minimum probability of 0.95, of which nonphosphorylated peptides (13103 spectra, 0.77% FDR) were subjected to SimPhospho spectral simulation under the default and optimized conditions. Although the K562 proteins were supplemented with a phosphoprotein standard bovine α -casein (S2 Fig), it was not taken into consideration since the number of human peptide spectra obtained was sufficient for the purpose of this study. Against libraries created with the simulated spectra (16752 phosphopeptide consensus spectra for 4182 sequences, 16752 decoys, in each library), SpectraST searching of the phosphopeptide sample data was performed with the scoring versions 4 and 5. At 1% FDR estimated by PeptideProphet, the default condition (10%-100%-100%, NL-PAW) with the version 4 scoring showed the highest number of spectral matches, i.e. 10609 spectral matches (Fig 6). However, after applying the respective 1% FLR cutoffs (HOM1 delatadot scores) obtained from the synthetic phosphopeptide data (S2 Table), one of the optimized conditions (50%-50%-100%, NL-P) with the version 4 scoring showed the highest number (5703 spectral matches), which was 2.8 times higher than that obtained under the default condition with the newer scoring version (2066 spectral matches). As seen on the synthetic phosphopeptide data (Fig 5), the version 4 scoring was better than the version 5 under the three simulation conditions tested on the K562 data (Fig 6). But the newer scoring may have worked less efficiently only with the libraries tested in this study, and further evaluations would be required for their performance comparison, e.g. with larger libraries. A recent sequence database (SwissProt 2018_02 release) was also tested for creating a simulated spectral library but it showed similar results (S2 Fig). The K562 results obtained under the optimized condition (50%-50%-100%, NL-P, version 4) are listed in S4 Table.

To comparatively evaluate the performance of simulated spectral library searching optimized on the TripleTOF CID data, the same data were analyzed using a widely used software

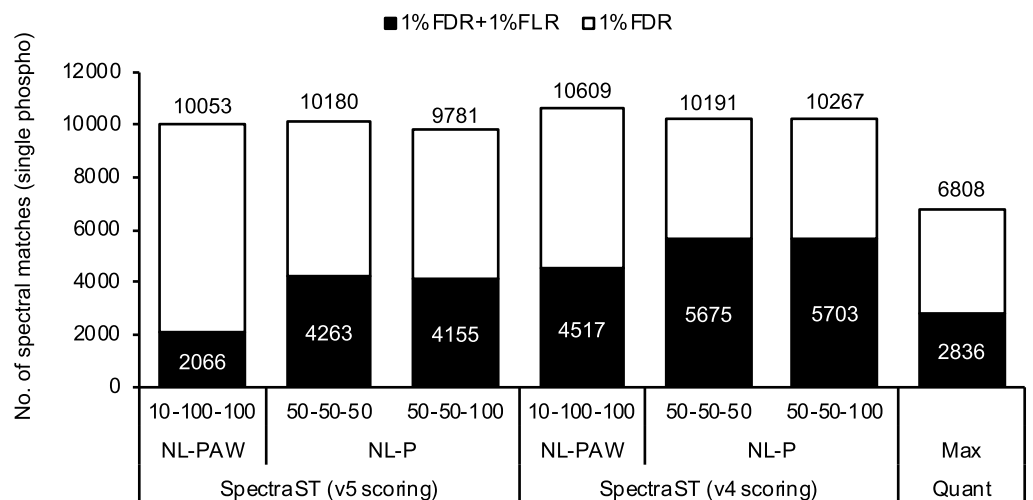


Fig 6. Phosphorylation site localization on K562 phosphopeptides by simulated spectral library searching and MaxQuant searching. From the K562 phosphopeptides analyzed by TripleTOF CID, singly phosphorylated peptides were identified at 1% FDR by SpectraST with the simulated spectral library and by MaxQuant, followed by applying 1% FLR cutoff. Three simulated spectral libraries generated under the default condition (10%-100%-100% intensities, NL-PAW) and the optimized conditions (50%-50%-50% and 50%-50%-100% intensities, NL-P; refer to Fig 5) were used for SpectraST searching with scoring versions 4 and 5.

<https://doi.org/10.1371/journal.pone.0225885.g006>

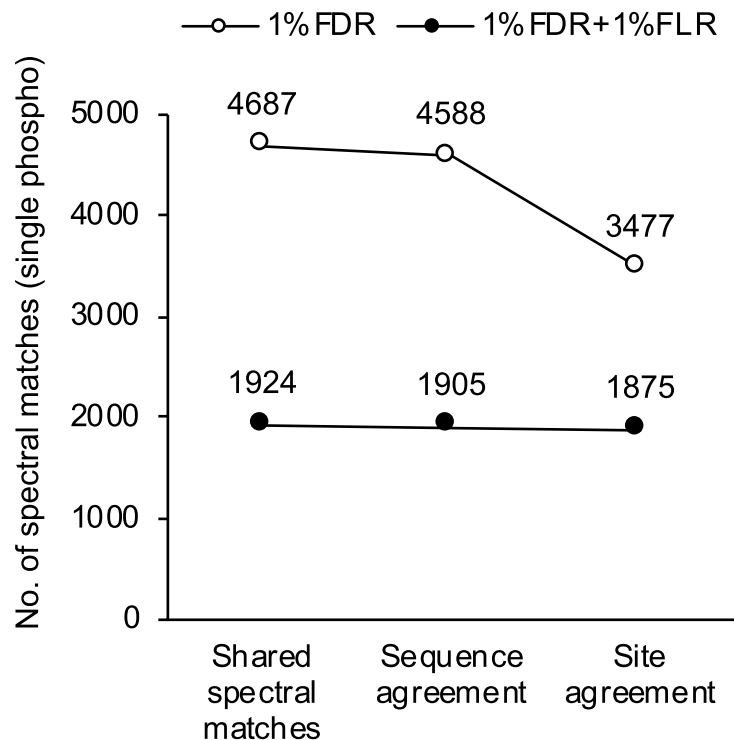


Fig 7. Sequence and site agreements of simulated spectral library searching and MaxQuant searching on K562 phosphopeptides. From the singly phosphorylated K562 peptides analyzed by TripleTOF CID, spectral matches shared by SpectraST searching against the simulated spectral library under the optimized condition (50%-50%-100% intensities, NL-P, versions 4 scoring) and MaxQuant searching were extracted at 1% FDR and 1% FLR (see Fig 6). Out of those, sequences and phosphorylation sites agreed by the two searches were counted.

<https://doi.org/10.1371/journal.pone.0225885.g007>

MaxQuant [43], where Andromeda engine for database searching [52] and PTM score for phosphorylation site localization [10] were implemented. MaxQuant searching was performed against the same sequence database (SwissProt 2010_03 human). An FLR of 1% on the 62 synthetic phosphopeptides identified with no FDR cutoff required PTM score of 0.977 (S3 Fig). The PTM score obtained on the TripleTOF dataset was somewhat less stringent compared to those previously obtained on the Orbitrap HCD datasets of 20 synthetic human phosphopeptides (0.992) [39] and >100000 synthetic phosphopeptides (0.995) [53]. At 1% FDR and 1% FLR (PTM score ≥ 0.977) on the K562 singly phosphorylated peptides, MaxQuant searching showed 2836 spectral matches, which were 2.0 times fewer than those obtained by simulated spectral library searching under the optimized condition (Fig 6). The superior identification sensitivity of simulated spectral library searching would largely be due to the limited search space, as spectral libraries contain only peptide spectra confidently identified by database searching. However, the confident localization on identified peptide sequences was achieved by optimization of spectral simulation and search conditions (refer to Figs 4–6).

The K562 results obtained by simulated spectral library searching under the optimized condition and MaxQuant searching shared 4687 spectral matches at 1% FDR (Fig 7). Among those, the two searches agreed on 4588 spectral matches for their peptide sequences (97.9%). Assuming that all the agreeing search results are correct identifications, the search disagreement (2.1%) is consistent with the expectation that each search may contain 1% false sequence identifications. The two searches agreed only on 3477 spectral matches for their phosphorylation sites

on the agreed sequences (75.8%). However, at 1% FDR and 1% FLR, the site agreement was clearly improved to 98.5% (i.e. 1.5% disagreement) as the two searches agreed on 1905 and 1875 spectral matches for their peptide sequences and phosphorylation sites, respectively. These observations support the validity of the FLR cutoff scores obtained on the synthetic phosphopeptides, and also the quality of the results obtained by simulated spectral library searching.

Conclusions

In this study, simulated spectral library searching previously proposed for Orbitrap HCD data was evaluated on TripleTOF CID data. SimPhospho enabled the simulation of single phosphorylations on TripleTOF CID spectra, and facilitated the use of various simulation conditions. After optimization for the TripleTOF CID data, simulated spectral library searching with SpectraST achieved highly confident and sensitive localization of phosphorylation sites. However, development of simulation algorithm for multiply phosphorylated peptides, evaluation of compatibility with alternative spectral library software, and further optimization for both TripleTOF CID and Orbitrap HCD remain to be considered. In principle, simulation of multiple phosphorylations might be performed by multiplying the mass shifts used for single phosphorylation, but careful evaluation and optimization would be needed with some model data. Although creation of simulated spectral libraries requires additional steps, such as enzymatic dephosphorylation, CID measurement, and finally spectral simulation in each experiment, the resulting libraries will cover all the possible phosphorylation sites on stored high-confidence peptide sequences and can be used repeatedly in a large-scale experiment, as demonstrated in our previous study [23]. Because of limited sizes of spectral libraries, simulated spectral library searching can be used to supplement database search results as mentioned above [23]. Alternatively, phosphopeptides may be simulated based on publicly available large spectral libraries such as PeptideAtlas [54]. Once de novo simulation of beam-type CID spectra for all the theoretical tryptic peptides would be achieved with well-predicted ion intensities, it may increase the impact of the phosphorylation simulation on sensitive and confident localization of phosphorylation sites.

Supporting information

S1 Table. List of synthetic phosphopeptides.
(XLSX)

S2 Table. HOM1 deltadot scores required for 1% FLR on synthetic phosphopeptides under different simulation conditions.
(XLSX)

S3 Table. Synthetic phosphopeptides identified and localized by simulated spectral library searching under the optimized condition (50%-50%-100%, NL-P, version 4).
(XLSX)

S4 Table. K562 phosphopeptides identified at 1% FDR and localized by simulated spectral library searching under the optimized condition (50%-50%-100%, NL-P, version 4).
(XLSX)

S1 Fig. Simulated spectral library searching of synthetic phosphopeptides under different mass tolerance conditions. Different precursor ion mass tolerance conditions (0.05 m/z , 0.1 m/z , 1 m/z , 3 m/z , and 5 m/z) were tested for simulated spectral library searching. For this evaluation, the previously reported Orbitrap HCD datasets were used (ref. 39: Suni *et al.*, 2015). A simulated spectral library was created from HeLa dephosphorylated peptides under the default

condition (10%-100%-100% intensities, NL-PAW), and used for SpectraST searching (version 4 scoring) of 20 synthetic phosphopeptides. The search results for the different mass tolerances were further filtered by the tolerance of 0.05 m/z . Those results were sorted by F-value for FLR calculation. The searching with 3 m/z tolerance showed more spectral matches at 1% FLR than those with the narrower tolerances, but the post-search filter by 0.05 m/z did not reduce the matches significantly. These results suggest that the mass tolerance condition in SpectraST searching may affect scoring used for FLR calculation. Therefore, we decided to use 3 m/z tolerance in combination with 0.05 m/z post-search filter for the localization in this study. (PDF)

S2 Fig. Searching of K562 and casein phosphopeptides against a simulated spectral library created with a recent sequence database. A SwissProt database (2018_02 release, *Homo sapiens*) supplemented with common contaminants (forward-reverse, total 40762 sequences) was tested for creating a simulated spectral library (50%-50%-100% intensities, NL-P). SpectraST searching (version 4 scoring) against the library showed 3266 spectral matches for casein phosphopeptides at 1% FDR, in addition to 10604 spectral matches for human phosphopeptides. Refer to Fig 6 for the search without casein phosphopeptides. (PDF)

S3 Fig. MaxQuant searching of synthetic phosphopeptides for calculating FLRs across PTM score. The TripleTOF 6600 dataset (9 wiff files) of the 62 singly phosphorylated synthetic peptides (24 human peptide sequences) was searched with MaxQuant. The searches resulted in 1058 spectral matches for the synthetic peptides, and 1% FLR required PTM score of 0.977. (PDF)

Acknowledgments

We thank Tatsuko Sakai for instrument support at the Analytical Center of Meijo University.

Author Contributions

Conceptualization: Susumu Y. Imanishi.

Data curation: Ayano Takai.

Formal analysis: Ayano Takai, Tomoya Tsubosaka, Yasuhiro Hirano, Naoki Hayakawa.

Funding acquisition: Susumu Y. Imanishi.

Investigation: Ayano Takai, Tomoya Tsubosaka, Yasuhiro Hirano, Fumitaka Tani, Pekka Haapaniemi.

Methodology: Ayano Takai, Tomoya Tsubosaka, Yasuhiro Hirano, Susumu Y. Imanishi.

Project administration: Susumu Y. Imanishi.

Software: Veronika Suni.

Supervision: Susumu Y. Imanishi.

Validation: Ayano Takai, Yasuhiro Hirano, Naoki Hayakawa.

Visualization: Ayano Takai, Tomoya Tsubosaka, Susumu Y. Imanishi.

Writing – original draft: Ayano Takai, Tomoya Tsubosaka, Susumu Y. Imanishi.

Writing – review & editing: Susumu Y. Imanishi.

References

1. Aebersold R, Mann M. Mass spectrometry-based proteomics. *Nature*. 2003; 422:198–207. <https://doi.org/10.1038/nature01511> PMID: 12634793
2. Steen H, Mann M. The ABC's (and XYZ's) of peptide sequencing. *Nat Rev Mol Cell Biol*. 2004; 5:699–711. <https://doi.org/10.1038/nrm1468> PMID: 15340378
3. Zhang Y, Fonslow BR, Shan B, Baek M-C, Yates JR III. Protein Analysis by Shotgun/Bottom-up Proteomics. *Chem Rev*. 2013; 113:2343–2394. <https://doi.org/10.1021/cr3003533> PMID: 23438204
4. Engholm-Keller K, Larsen MR. Technologies and challenges in large-scale phosphoproteomics. *Proteomics*. 2013; 13:910–931. <https://doi.org/10.1002/pmic.201200484> PMID: 23404676
5. Roux PP, Thibault P. The coming of age of phosphoproteomics—from large data sets to inference of protein functions. *Mol Cell Proteomics*. 2013; 12:3453–3464. <https://doi.org/10.1074/mcp.R113.032862> PMID: 24037665
6. von Stechow L, Francavilla C, Olsen JV. Recent findings and technological advances in phosphoproteomics for cells and tissues. *Expert Rev Proteomics*. 2015; 12:469–487. <https://doi.org/10.1586/14789450.2015.1078730> PMID: 26400465
7. Riley NM, Coon JJ. Phosphoproteomics in the Age of Rapid and Deep Proteome Profiling. *Anal Chem*. 2016; 88:74–94. <https://doi.org/10.1021/acs.analchem.5b04123> PMID: 26539879
8. Potel CM, Lemeer S, Heck AJR. Phosphopeptide fragmentation and site localization by mass spectrometry; an update. *Anal Chem*. 2018; 91:126–141. <https://doi.org/10.1021/acs.analchem.8b04746> PMID: 30457327
9. Beausoleil SA, Villén J, Gerber SA, Rush J, Gygi SP. A probability-based approach for high-throughput protein phosphorylation analysis and site localization. *Nat Biotechnol*. 2006; 24:1285–1292. <https://doi.org/10.1038/nbt1240> PMID: 16964243
10. Olsen JV, Blagoev B, Gnäd F, Macek B, Kumar C, Mortensen P, et al. Global, in vivo, and site-specific phosphorylation dynamics in signaling networks. *Cell*. 2006; 127:635–648. <https://doi.org/10.1016/j.cell.2006.09.026> PMID: 17081983
11. Savitski MM, Lemeer S, Boesche M, Lang M, Mathieson T, Bantscheff M, et al. Confident phosphorylation site localization using the Mascot Delta Score. *Mol Cell Proteomics*. 2011; 10:M110 003830. <https://doi.org/10.1074/mcp.M110.003830> PMID: 21057138
12. Taus T, Köcher T, Pichler P, Paschke C, Schmidt A, Henrich C, et al. Universal and confident phosphorylation site localization using phosphoRS. *J Proteome Res*. 2011; 10:5354–5362. <https://doi.org/10.1021/pr200611n> PMID: 22073976
13. Fermin D, Walmsley SJ, Gingras AC, Choi H, Nesvizhskii AI. LuciPHOR: algorithm for phosphorylation site localization with false localization rate estimation using modified target-decoy approach. *Mol Cell Proteomics*. 2013; 12:3409–3419. <https://doi.org/10.1074/mcp.M113.028928> PMID: 23918812
14. Hunter T. Signaling—2000 and beyond. *Cell*. 2000; 100:113–127. [https://doi.org/10.1016/s0092-8674\(00\)81688-8](https://doi.org/10.1016/s0092-8674(00)81688-8) PMID: 10647936
15. de Thonel A, Ferraris SE, Pallari H-M, Imanishi SY, Kochin V, Hosokawa T, et al. Protein Kinase C zeta regulates Cdk5/p25 signaling during myogenesis. *Molecular Biology of the Cell*. 2010; 21:1423–1434. <https://doi.org/10.1091/mbc.E09-10-0847> PMID: 20200223
16. Tyagarajan SK, Ghosh H, Yevenes GE, Imanishi SY, Zeilhofer HU, Gerrits B, et al. Extracellular signal-regulated kinase and glycogen synthase kinase 3 beta regulate gephyrin postsynaptic aggregation and GABAergic synaptic function in a calpain-dependent mechanism. *Journal of Biological Chemistry*. 2013; 288:9634–9647. <https://doi.org/10.1074/jbc.M112.442616> PMID: 23408424
17. Kochin V, Shimi T, Torvaldson E, Adam SA, Goldman A, Pack C-G, et al. Interphase phosphorylation of lamin A. *Journal of Cell Science*. 2014; 127:2683–2696. <https://doi.org/10.1242/jcs.141820> PMID: 24741066
18. Sjoqvist M, Antfolk D, Ferraris S, Rraklli V, Haga C, Antila C, et al. PKC zeta regulates Notch receptor routing and activity in a Notch signaling-dependent manner. *Cell Research*. 2014; 24:433–450. <https://doi.org/10.1038/cr.2014.34> PMID: 24662486
19. Hyder CL, Kempainen K, Isoniemi KO, Imanishi SY, Goto H, Inagaki M, et al. Sphingolipids inhibit vimentin-dependent cell migration. *Journal of Cell Science*. 2015; 128:2057–2069. <https://doi.org/10.1242/jcs.160341> PMID: 25908861
20. Lindqvist J, Imanishi SY, Torvaldson E, Malinen M, Remes M, Orn F, et al. Cyclin-dependent kinase 5 acts as a critical determinant of AKT-dependent proliferation and regulates differential gene expression by the androgen receptor in prostate cancer cells. *Molecular Biology of the Cell*. 2015; 26:1971–1984. <https://doi.org/10.1091/mbc.E14-12-1634> PMID: 25851605

21. Virtakoivu R, Mai A, Mattila E, De Franceschi N, Imanishi SY, Corthals G, et al. Vimentin-ERK signaling uncouples Slug gene regulatory function. *Cancer Research*. 2015; 75:2349–2362. <https://doi.org/10.1158/0008-5472.CAN-14-2842> PMID: 25855378
22. Santio NM, Landor SK-J, Vahtera L, Ylä-Pelto J, Paloniemi E, Imanishi SY, et al. Phosphorylation of Notch1 by Pim kinases promotes oncogenic signaling in breast and prostate cancer cells. *Oncotarget*. 2016; 7:43220–43238. <https://doi.org/10.18632/oncotarget.9215> PMID: 27281612
23. Kauko O, Laajala TD, Jumppanen M, Hintsanen P, Suni V, Haapaniemi P, et al. Label-free quantitative phosphoproteomics with novel pairwise abundance normalization reveals synergistic RAS and CIP2A signaling. *Sci Rep*. 2015; 5:13099. <https://doi.org/10.1038/srep13099> PMID: 26278961
24. Nguyen EV, Imanishi SY, Haapaniemi P, Yadav A, Saloheimo M, Corthals GL, et al. Quantitative site-specific phosphoproteomics of *Trichoderma reesei* signaling pathways upon induction of hydrolytic enzyme production. *Journal of Proteome Research*. 2016; 15:457–467. <https://doi.org/10.1021/acs.jproteome.5b00796> PMID: 26689635
25. Soderholm S, Kainov DE, Ohman T, Denisova OV, Schepens B, Kuleskiy E, et al. Phosphoproteomics to characterize host response during influenza A virus infection of human macrophages. *Molecular & Cellular Proteomics*. 2016; 15:3203–3219.
26. Frewen BE, Merrihew GE, Wu CC, Noble WS, MacCoss MJ. Analysis of peptide MS/MS spectra from large-scale proteomics experiments using spectrum libraries. *Anal Chem*. 2006; 78:5678–5684. <https://doi.org/10.1021/ac060279n> PMID: 16906711
27. Craig R, Cortens JC, Fenyo D, Beavis RC. Using annotated peptide mass spectrum libraries for protein identification. *J Proteome Res*. 2006; 5:1843–1849. <https://doi.org/10.1021/pr0602085> PMID: 16889405
28. Lam H, Deutsch EW, Eddes JS, Eng JK, King N, Stein SE, et al. Development and validation of a spectral library searching method for peptide identification from MS/MS. *Proteomics*. 2007; 7:655–667. <https://doi.org/10.1002/pmic.200600625> PMID: 17295354
29. Griss J. Spectral library searching in proteomics. *Proteomics*. 2016; 16:729–740. <https://doi.org/10.1002/pmic.201500296> PMID: 26616598
30. Ye D, Fu Y, Sun R, Wang H, Yuan Z, Chi H, et al. Open MS/MS spectral library search to identify unanticipated post-translational modifications and increase spectral identification rate. *Bioinformatics*. 2010; 26:i399–406. <https://doi.org/10.1093/bioinformatics/btq185> PMID: 20529934
31. AhmÉ E, Nikitin F, Lisacek F, Müller M. QuickMod: A tool for open modification spectrum library searches. *J Proteome Res*. 2011; 10:2913–2921. <https://doi.org/10.1021/pr200152g> PMID: 21500769
32. Ma CWM, Lam H. Hunting for unexpected post-translational modifications by spectral library searching with tier-wise scoring. *J Proteome Res*. 2014; 13:2262–2271. <https://doi.org/10.1021/pr401006g> PMID: 24661115
33. Horlacher O, Lisacek F, Müller M. Mining Large Scale Tandem Mass Spectrometry Data for Protein Modifications Using Spectral Libraries. *J Proteome Res*. 2016; 15:721–731. <https://doi.org/10.1021/acs.jproteome.5b00877> PMID: 26653734
34. Bodenmiller B, Malmstrom J, Gerrits B, Campbell D, Lam H, Schmidt A, et al. PhosphoPep—a phosphoproteome resource for systems biology research in *Drosophila* Kc167 cells. *Mol Syst Biol*. 2007; 3:139. <https://doi.org/10.1038/msb4100182> PMID: 17940529
35. Bodenmiller B, Campbell D, Gerrits B, Lam H, Jovanovic M, Picotti P, et al. PhosphoPep—a database of protein phosphorylation sites in model organisms. *Nat Biotechnol*. 2008; 26:1339–1340. <https://doi.org/10.1038/nbt1208-1339> PMID: 19060867
36. Hummel J, Niemann M, Wienkoop S, Schulze W, Steinhauser D, Selbig J, et al. ProMEX: a mass spectral reference database for proteins and protein phosphorylation sites. *BMC Bioinformatics*. 2007; 8:216. <https://doi.org/10.1186/1471-2105-8-216> PMID: 17587460
37. Hu Y, Li Y, Lam H. A semi-empirical approach for predicting unobserved peptide MS/MS spectra from spectral libraries. *Proteomics*. 2011; 11:4702–4711. <https://doi.org/10.1002/pmic.201100316> PMID: 22038894
38. Hu Y, Lam H. Expanding tandem mass spectral libraries of phosphorylated peptides: advances and applications. *J Proteome Res*. 2013; 12:5971–5977. <https://doi.org/10.1021/pr4007443> PMID: 24125593
39. Suni V, Imanishi SY, Maiolica A, Aebersold R, Corthals GL. Confident site localization using a simulated phosphopeptide spectral library. *J Proteome Res*. 2015; 14:2348–2359. <https://doi.org/10.1021/acs.jproteome.5b00050> PMID: 25774671
40. Imanishi SY, Kochin V, Ferraris SE, de Thonel A, Pallari HM, Corthals GL, et al. Reference-facilitated phosphoproteomics: fast and reliable phosphopeptide validation by microLC-ESI-Q-TOF MS/MS. *Mol Cell Proteomics*. 2007; 6:1380–1391. PMID: 17510049

41. Suni V, Suomi T, Tsubosaka T, Imanishi SY, Elo LL, Corthals GL. SimPhospho: a software tool enabling confident phosphosite assignment. *Bioinformatics*. 2018; 34:2690–2692. <https://doi.org/10.1093/bioinformatics/bty151> PMID: 29596608
42. Gillet LC, Navarro P, Tate S, Röst H, Selevsek N, Reiter L, et al. Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis. *Mol Cell Proteomics*. 2012; 11:O111 016717. <https://doi.org/10.1074/mcp.O111.016717> PMID: 22261725
43. Cox J, Mann M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol*. 2008; 26:1367–1372. <https://doi.org/10.1038/nbt.1511> PMID: 19029910
44. Rappsilber J, Ishihama Y, Mann M. Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics. *Anal Chem*. 2003; 75:663–670. <https://doi.org/10.1021/ac026117i> PMID: 12585499
45. Keller A, Eng J, Zhang N, Li XJ, Aebersold R. A uniform proteomics MS/MS analysis platform utilizing open XML file formats. *Mol Syst Biol*. 2005; 1:2005.0017. <https://doi.org/10.1038/msb4100024> PMID: 16729052
46. Deutsch EW, Mendoza L, Shteynberg D, Farrah T, Lam H, Tasman N, et al. A guided tour of the Trans-Proteomic Pipeline. *Proteomics*. 2010; 10:1150–1159. <https://doi.org/10.1002/pmic.200900375> PMID: 20101611
47. Craig R, Beavis RC. TANDEM: matching proteins with tandem mass spectra. *Bioinformatics*. 2004; 20:1466–1467. <https://doi.org/10.1093/bioinformatics/bth092> PMID: 14976030
48. Keller A, Nesvizhskii AI, Kolker E, Aebersold R. Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. *Anal Chem*. 2002; 74:5383–5392. <https://doi.org/10.1021/ac025747h> PMID: 12403597
49. Lam H, Deutsch EW, Edes JS, Eng JK, Stein SE, Aebersold R. Building consensus spectral libraries for peptide identification in proteomics. *Nat Methods*. 2008; 5:873–875. <https://doi.org/10.1038/nmeth.1254> PMID: 18806791
50. Lam H, Deutsch EW, Aebersold R. Artificial decoy spectral libraries for false discovery rate estimation in spectral library searching in proteomics. *J Proteome Res*. 2010; 9:605–610. <https://doi.org/10.1021/pr900947u> PMID: 19916561
51. Baumgardner LA, Shanmugam AK, Lam H, Eng JK, Martin DB. Fast parallel tandem mass spectral library searching using GPU hardware acceleration. *J Proteome Res*. 2011; 10:2882–2888. <https://doi.org/10.1021/pr200074h> PMID: 21545112
52. Cox J, Neuhauser N, Michalski A, Scheltema RA, Olsen JV, Mann M. Andromeda: a peptide search engine integrated into the MaxQuant environment. *J Proteome Res*. 2011; 10:1794–1805. <https://doi.org/10.1021/pr101065j> PMID: 21254760
53. Marx H, Lemeer S, Schliep JE, Matheron L, Mohammed S, Cox J, et al. A large synthetic peptide and phosphopeptide reference library for mass spectrometry-based proteomics. *Nat Biotechnol*. 2013; 31:557–564. <https://doi.org/10.1038/nbt.2585> PMID: 23685481
54. Desiere F, Deutsch EW, King NL, Nesvizhskii AI, Mallick P, Eng J, et al. The PeptideAtlas project. *Nucleic Acids Res*. 2006; 34:D655–658. <https://doi.org/10.1093/nar/gkj040> PMID: 16381952