

SUPPLEMENTAL DATA

Spectral library based analysis of arginine phosphorylations in *Staphylococcus aureus*

**Sabryna Junker, Sandra Maaß, Andreas Otto, Stephan Michalik, Friedrich Morgenroth, Ulf
Gerth, Michael Hecker, and Dörte Becher***

Institute for Microbiology, University of Greifswald, Germany

***Corresponding Author information:** Prof. Dr. Dörte Becher, Institute for Microbiology,
University of Greifswald, Friedrich-Ludwig-Jahn-Straße 15, 17487 Greifswald, Germany,
phone: 0049-3834-4205903, fax: 0049-3834-4205902, mail: dbecher@uni-greifswald.de

Abbreviations

BIP	2,2'-bipyridine
HCD	higher-energy collision-activated dissociation
LTQ	linear trap quadrupole
MSA	multistage activation
NCE	normalised collision energy
NL	neutral loss
spec	spectinomycin
SB	Super Broth
TMAD	3-(dimethylcarbamoylimino)-1,1-dimethylurea
USM	user specific modification
WBA	wideband activation

***In vitro* studies with recombinant PtpB**

For the preparation of recombinant purified PtpB, the gene *ptpB* was fused with a Strep-Tag using *Bam*HI/*Kpn*I and cloned into pRSETwow (in house derivative of overexpression plasmid pRSETA (Invitrogen, Thermo Fisher Scientific, Waltham, USA)). Overexpression was performed using *E. coli* BL21 DE3 pLysS. Cells were cultivated in 1 l SB up to OD₆₀₀ = 1.0 with subsequent induction by 0.1 mM IPTG. After two hours, cells were harvested and disrupted using a Frenchpress with buffer W (100 mM NaCl, 150 mM Tris/HCl pH 8.0, 1 mM DTT, 4 °C). Proteins were purified via Strep-Tag column (perfusor flow rate = 1 ml/h) and buffer W. Elution was performed with 3 mM Desthiobiotin and fractions were collected in 50% glycerine/1 mM DTT. Identification was performed using SDS-PAGE and mass spectrometry.

Detection of phosphatase activity and substrate specificity was performed measuring the release of inorganic phosphate as part of an molybdate:malachite green:phosphate complex-based optical assay modified from Carter *et al.* (1) To adapt it to the special needs of arginine phosphorylations, solution A (0.57% (v/v) HCl, 57 mM (w/v) Na₂MoO₄ x 2 H₂O) was prepared beforehand and stored for later use, solution B contained 0.15% (w/v) malachite green in distilled water which was diluted 1:3.57 immediately before use and solution C contained 1.4 M H₂SO₄. There were no additional incubation times between addition of solution A, B and C. After 45 minutes, samples were measured at 600 nm. To enable enzymatic activity as well as linearity of the optical assay, the enzymatic reaction was performed in 50 mM Tris/HCl pH 6.52 and with or without inhibitor (5 mM C₃H₇O₆PNa₂, Na₄P₂O₇, Na₃VO₄ and NaF). Three synthetic peptides with equal peptide sequences, but phosphorylation at either arginine or adjacent serine or threonine residues were used as substrate and supplemented in equal amounts to start the reaction (supplemental data table 11S). Enzymatic activity was measured after 120 minutes (37 °C) using the phosphate assay. Negative controls contained either no

enzyme, no substrate or the enzyme was replaced by a dummy protein (equal amounts of lysozyme from chicken egg). Phosphate concentration was measured using a calibration curve within a range of 9.5 nM KH_2PO_4 – 0.15 mM KH_2PO_4 ($R^2 = 0.9997$).

MS measurements

All proteomic samples were measured online by LC-ESI-mass spectrometry using an Easy-nLCII or Easy-nLC1000 (Thermo Fisher Scientific, Waltham, USA) with self-packed analytical columns (100 μm x 20 cm) containing C18 material (Phenomenex, Aschaffenburg, Germany) coupled to a LTQ Orbitrap Velos, Orbitrap Velos Pro or Orbitrap Elite (Thermo Fisher Scientific, Waltham, USA) applying an 85 min gradient, using 0.1% (v/v) acetic acid as buffer A and 99.9% (v/v) ACN with 0.1% (v/v) acetic acid as buffer B and a flow rate of 300 nl/min. For CID measurements, full scan at a resolution of 60,000 (120,000 for Orbitrap Elite), recorded in the Orbitrap analyzer, was followed by the analysis of the 20 most intense precursor ions (in the LTQ) in data-dependent MS/MS mode with lockmass option enabled (2). For phosphoproteome samples, wideband activation (WBA) at LTQ Orbitrap Velos was enabled and fragment spectra were recorded by enabling neutral loss (NL) option. Multistage activation (MSA) at -97.98, -48.99, -32.66 and -24.49 Th was applied in all MS/MS events to improve the fragmentation for putative phosphorylated ions (3). For HCD measurements, full scan at a resolution of 30,000 (60,000 for Orbitrap Elite) was carried out. The top 10 precursor ions were fragmented with 40% normalised collision energy (NCE) and MS/MS Scans were recorded at a resolution of 7,500 (15,000 for Orbitrap Elite).

MS measurements of synthetic peptides were performed with an Easy-nLC1000 (Thermo Fisher Scientific, Waltham, USA) coupled to an Orbitrap Elite (Thermo Fisher Scientific, Waltham, USA) applying the same setup and parameters as for proteome samples except that

for HCD fragmentation, NCE was varied from 24% to 40% and the gradient was applied for 30 minutes only. All spectra were recorded without dynamic exclusion option.

Construction of spectral libraries

For spectral library construction, raw files were converted to mzXML (MSConvert GUI, ProteoWizard (4)). Phosphoproteome samples were analysed using the search engines Comet (v. 2014.02 rev. 2), X! Tandem, both as part of the Trans Proteomics Pipeline (5) (TPP, v. 4.8.0) and Sorcerer/Sequest (v. 4.0.4), classical proteome samples were processed with Comet and X! Tandem and synthetic peptide measurements were processed with Comet subjected against the database from Uniprot (release May 2015), containing 2680 proteins of *S. aureus* COL (common laboratory contaminants added). For the respective parameter files, precursor mass tolerance was set to 10 ppm and fragment mass tolerance was set to 1.0005 Da (CID) or 20 ppm (HCD). Full tryptic specificity with a maximum of two missed cleavage sites was applied as well as variable modifications (oxidation on methionine, carbamidomethylation on cysteine and phosphorylation on serine, threonine, tyrosine or arginine). For statistical validation, decoy hits were used to calculate false positives. In addition, for HCD data, accurate mass binning using ppm was applied and only the Expect Score was used as discriminant.

All pep.xml files were compiled into SpectraST libraries (6). Spectral libraries for synthetic peptides consider hits with probability ≥ 0.9 and spectral libraries for non-phosphorylated peptides consider hits with probability ≥ 0.95 . Raw spectral libraries for phosphopeptides were built by implementing serine, threonine and tyrosine phosphorylations as well as the user specific modification (USM_R|+79.966331) to take into account arginine phosphorylations and contain all phosphorylated spectra with a probability $> 1 \times 10^{-7}$. All MaxQuant spectra passing the manual validation were kept in the libraries and the final

libraries were converted to consensus libraries, combined with equal number of spectra used for consensus and added with decoys. All consensus spectra derived from synthetic peptides were validated manually according to the criteria described in the methods section. To enhance proteome coverage and number of possible phosphosites, the gel free raw data used for the studies of Bäsell *et al.* (7) were also added to the library.

References

1. Carter, S. G., and Karl, D. W. (1982) Inorganic phosphate assay with malachite green: An improvement and evaluation *. *J. Biochem. Biophys. Methods* 7, 7–13
2. Olsen, J. V., de Godoy, L. M. F., Li, G., Maček, B., Mortensen, P., Pesch, R., Makarov, A., Lange, O., Horning, S., and Mann, M. (2005) Parts per Million Mass Accuracy on an Orbitrap Mass Spectrometer via Lock Mass Injection into a C-trap. *Mol. Cell. Proteomics* 4, 2010–21
3. Schroeder, M. J., Shabanowitz, J., Schwartz, J. C., Hunt, D. F., and Coon, J. J. (2004) A Neutral Loss Activation Method for Improved Phosphopeptide Sequence Analysis by Quadrupole Ion Trap Mass Spectrometry. *Anal. Chem.* 76, 3590–8
4. Chambers, M. C., Maclean, B., Burke, R., Amodei, D., Ruderman, D. L., Neumann, S., Gatto, L., Fischer, B., Pratt, B., Egertson, J., Hoff, K., Kessner, D., Tasman, N., Shulman, N., Frewen, B., Baker, T. A., Brusniak, M.-Y., Paulse, C., Creasy, D., Flashner, L., Kani, K., Moulding, C., Seymour, S. L., Nuwaysir, L. M., Lefebvre, B., Kuhlmann, F., Roark, J., Paape, R., Suckau, D., Hemenway, T., Huhmer, A., Langridge, J., Connolly, B., Chadick, T., Holly, K., Eckels, J., Deutsch, E. W., Moritz, R. L., Katz, J. E., Agus, D. B., MacCoss, M., Tabb, D. L., and Mallick, P. (2012) A cross-platform Toolkit for Mass Spectrometry and Proteomics. *Nat. Biotechnol.* 30, 918–20
5. Deutsch, E. W., Mendoza, L., Shteynberg, D., Farrah, T., Lam, H., Tasman, N., Sun, Z., Nilsson, E., Pratt, B., Prazen, B., Eng, J. K., Daniel, B. M., Nesvizhskii, A., and Aebersold, R. (2010) A Guided Tour of the Trans-Proteomic Pipeline. *Proteomics* 10, 1150–9
6. Lam, H., Deutsch, E. W., Eddes, J. S., Eng, J. K., King, N., Stein, S. E., and Aebersold, R. (2007) Development and validation of a spectral library searching method for peptide identification from MS/MS. *Proteomics* 7, 655–67
7. Bäsell, K., Otto, A., Junker, S., Zühlke, D., Rappen, G.-M., Schmidt, S., Hentschker, C., Maček, B., Ohlsen, K., Hecker, M., and Becher, D. (2014) The phosphoproteome and its physiological dynamics in *Staphylococcus aureus*. *Int. J. Med. Microbiol.* 304, 121–32
8. Vizcaíno, J. A., Csordas, A., Del-Toro, N., Dianes, J. A., Griss, J., Lavidas, I., Mayer, G., Perez-Riverol, Y., Reisinger, F., Ternent, T., Xu, Q.-W., Wang, R., and Hermjakob, H. (2016) 2016 update of the PRIDE database and its related tools. *Nucleic Acids Res.* 44, 447–56

Supplemental data figures legends

Supplemental data figure 1S: Number of identified phosphopeptides identified in one or \geq two biological replicates of the mutant Δ *ptpB* (A) and the wild type (B)

Phosphopeptides are defined as unique if they differ in peptide sequence or phosphosite. Different charge states as well as oxidation on methionine/ carbamidomethylation are not considered for this figure. This way, the same phosphosite can be identified in the form of several phosphopeptides. The number of phosphopeptides identified in one or \geq two biological replicates is displayed on the left pie whereas the right pie highlights the distribution of all phosphopeptides identified in \geq two biological replicates according to the respective amino acid.

Supplemental data figure 2S: Functional groups of identified proteins phosphorylated on arginine residues

Each protein phosphorylated at arginine residues was matched to one or more functional groups, according to TIGRfam or KEGG based annotation. The green line shows the absolute number of identified arginine phosphorylated proteins for the main functional groups whereas the orange line represents the percentage coverage of identified arginine phosphorylated proteins with respect to the theoretical number of proteins belonging to this group.

Supplemental data figure 3S: Comparison of probabilities and dot values of different spectral libraries

Compared are only peptide sequences which can be found with exactly the same sequence, phosphosite and charge state in both experimental and synthetic datasets. The comparison

has only been done for CID measurements. Samples 1-4 represent four biological replicates of the mutant $\Delta ptpB$. **A Comparison of mean probabilities of all samples searched against the different libraries** The mean probabilities of all comparable phosphopeptides are compared when the same data were subjected to the search with four different spectral libraries: The experimental spectral library contains only spectra originating from phosphoenriched samples and the synthetic spectral library contains only spectra originating from measurements of synthetic arginine phosphorylated peptides. The combined library with disparate number of spectra used for consensus combines both libraries using different numbers of spectra originating from experimental or synthetic data if available whereas the combined library with equal numbers of spectra used for consensus uses equal numbers of spectra originating from different sources. That means, if only one phosphospectrum of the respective sequence can be found within the experimental data set of phosphoenriched samples, there is also only one synthetic spectrum taken for the construction of the consensus spectrum. **B Comparison of probabilities between experimental and combined library with equal number of spectra used for consensus** Comparison of the probability values the phosphopeptide gets after the construction of consensus spectra in experimental library (blue) or combined library with equal number of spectra used for consensus (orange). **C Comparison of mean dot values of all samples searched against the different libraries** The mean dot values of all comparable phosphopeptides are compared when the same data were subjected to searches with four different spectral libraries.

Supplemental data figure 4S: Comparison of arginine phosphoproteomes

The Venn diagram displays the overlap between all arginine phosphoproteins identified in *S. aureus* (this study) and *B. subtilis* (Elsholz *et al.* and Schmidt *et al.*). Homologues for 39 proteins are phosphorylated on arginine residues in both organisms.

Supplemental data figure 5S: Comparison of proteins phosphorylated at arginine residues in *S. aureus* and *B. subtilis*

The pie chart displays the percentage value of proteins that are identified with arginine phosphosites either only in *S. aureus* (coloured orange), in *S. aureus* and *B. subtilis*, but at different phosphosites (coloured yellow) or in *S. aureus* and *B. subtilis* with at least one identical or adjacent phosphosite (coloured blue).

Supplemental data figure 6S: Voronoi treemaps of combined spectral library

Supplemental data figure 6S provides the Voronoi treemaps of the combined spectral library as high quality figure. Figure 6S A displays the main functional categories. Figure 6S B shows the functional subcategories. Figure 6S C contains the respective SACOL numbers. Each cell displays a single protein. Light grey coloured proteins are part of the spectral library and if one or more phosphopeptides are part of the library, the respective protein is depicted in light blue. Proteins containing at least one arginine phosphorylation are depicted in dark blue. The figures are provided in high quality as separate .pdf files.

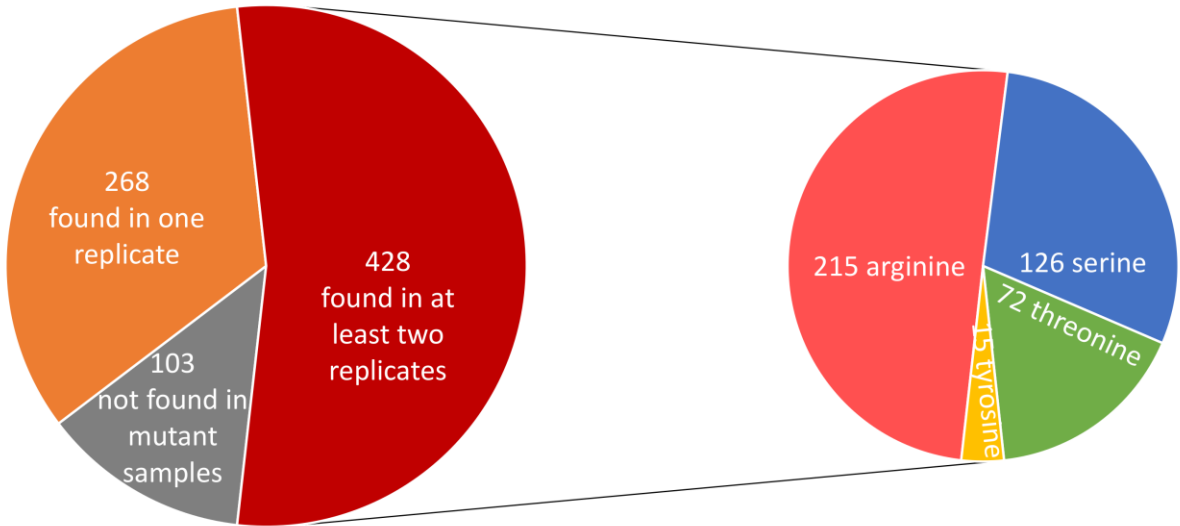
Supplemental data figure 7S: Annotated spectra of phosphorylated peptides

Supplemental data figure 7S A-E displays five examples for annotated spectra of identified phosphorylated peptides. The figures are additionally provided in high quality as separate .pdf files. Furthermore, for each phosphopeptide identification listed in supplemental data table 8S or 9S, one annotated spectrum is provided via the PRIDE partner repository (8) with the dataset identifier PXD007167. The documents are called “annotated_spectra_CID.pdf” and “annotated_spectra_HCD.pdf”.

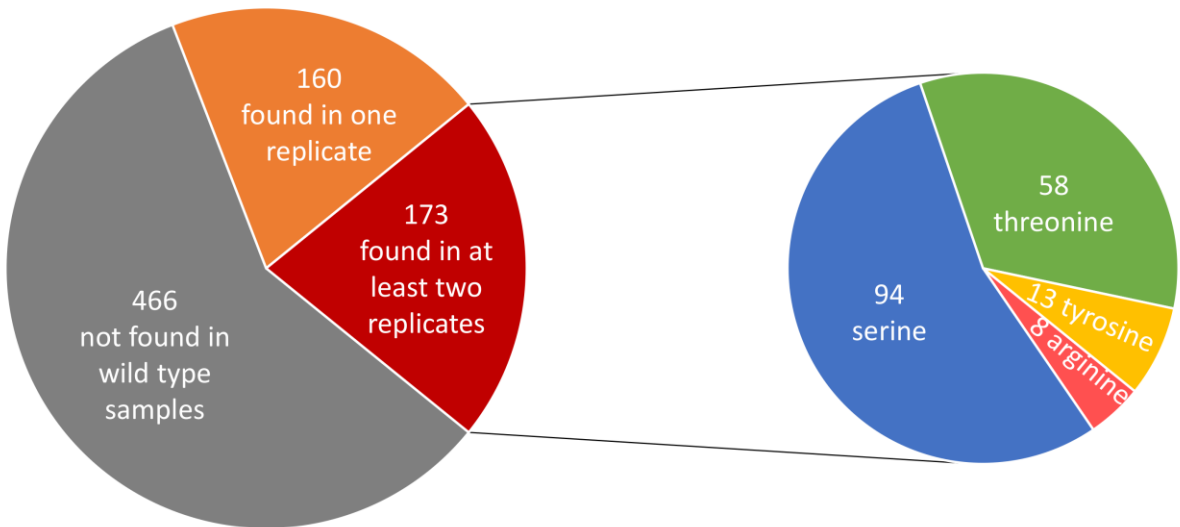
Supplemental data Figures

Supplemental data figure 1S:

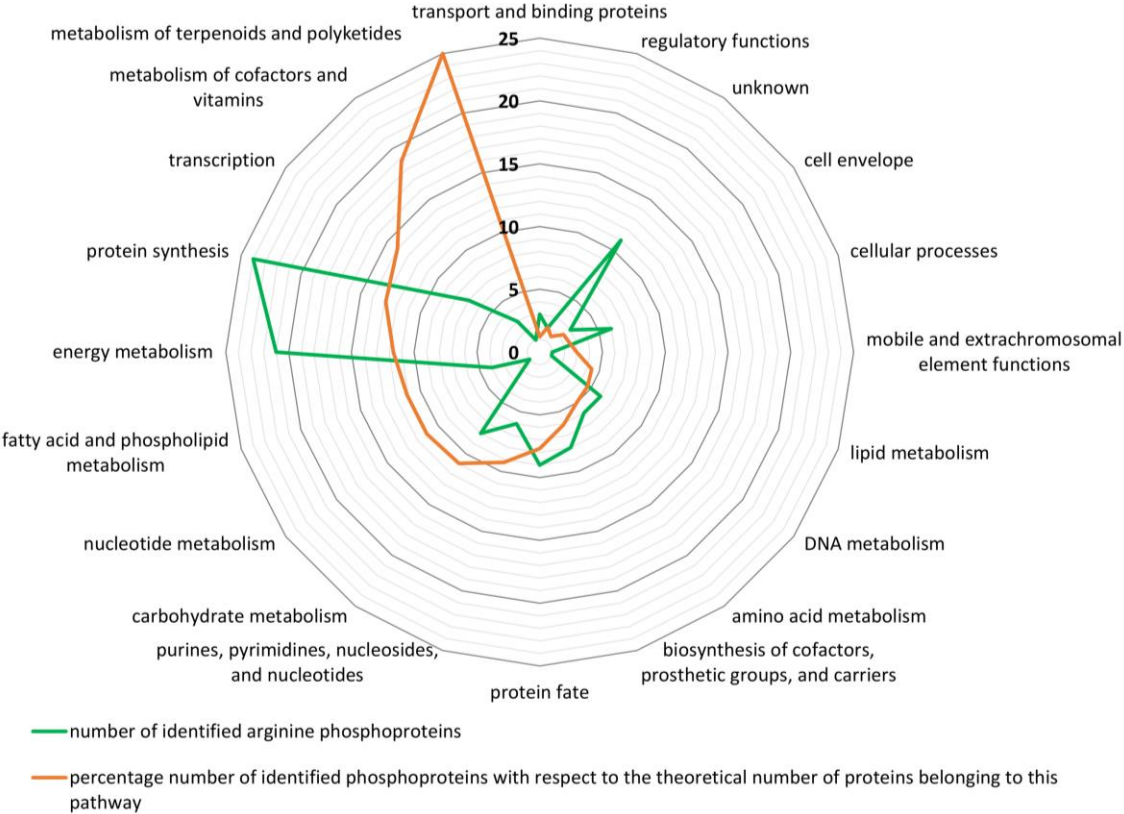
A



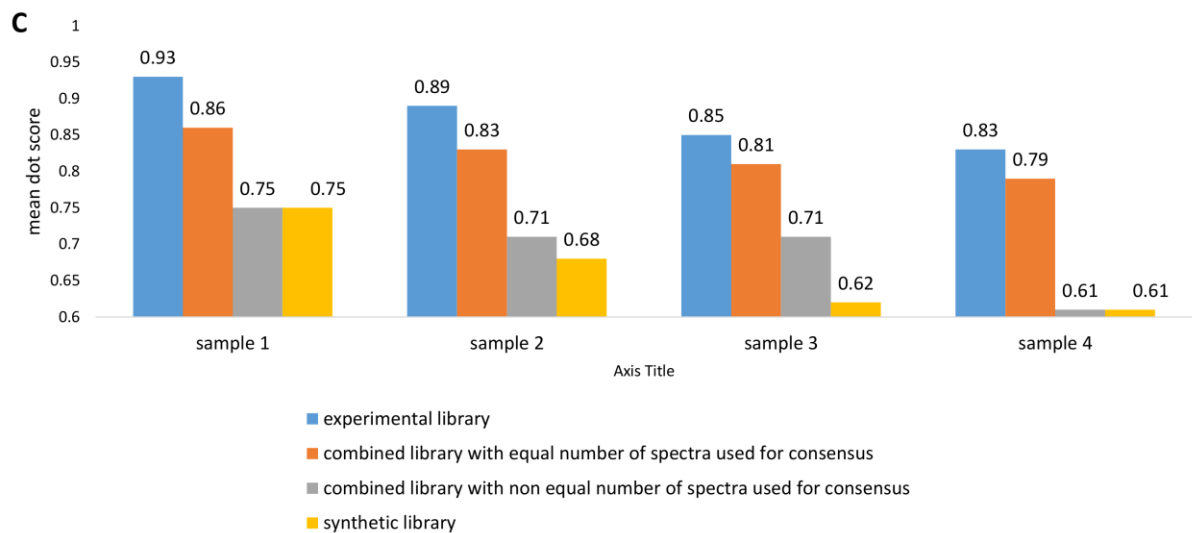
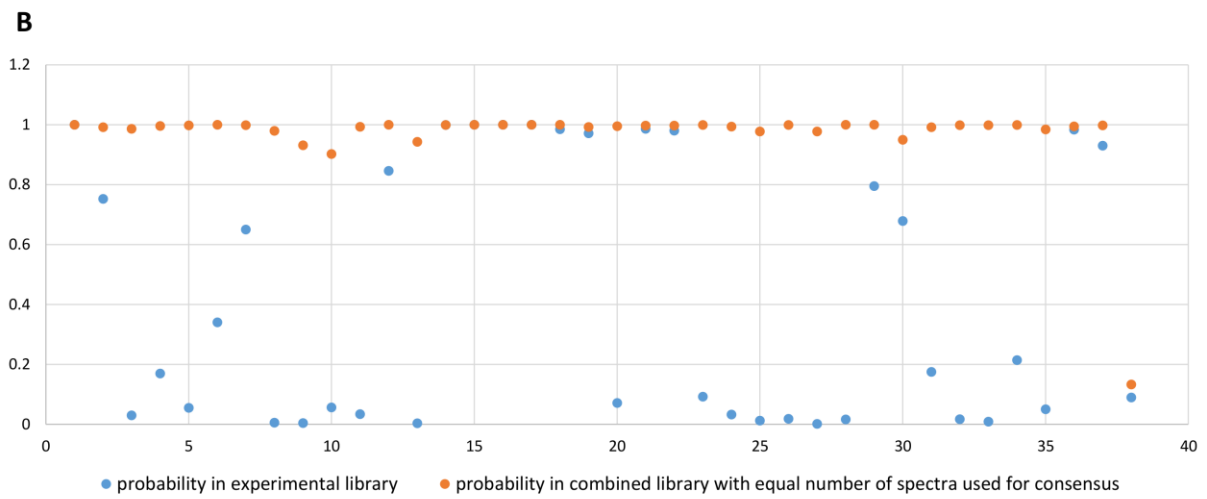
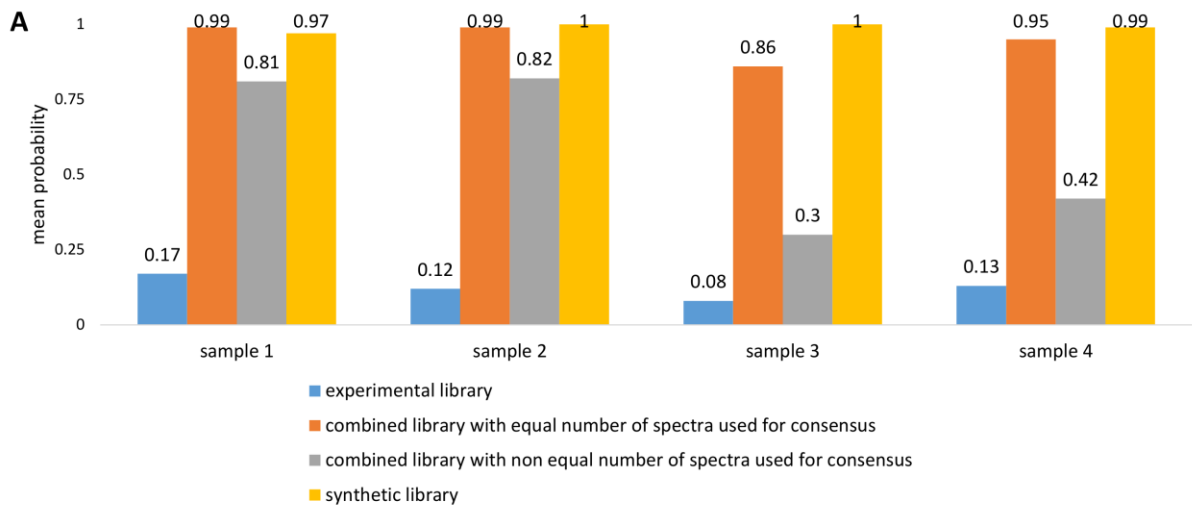
B



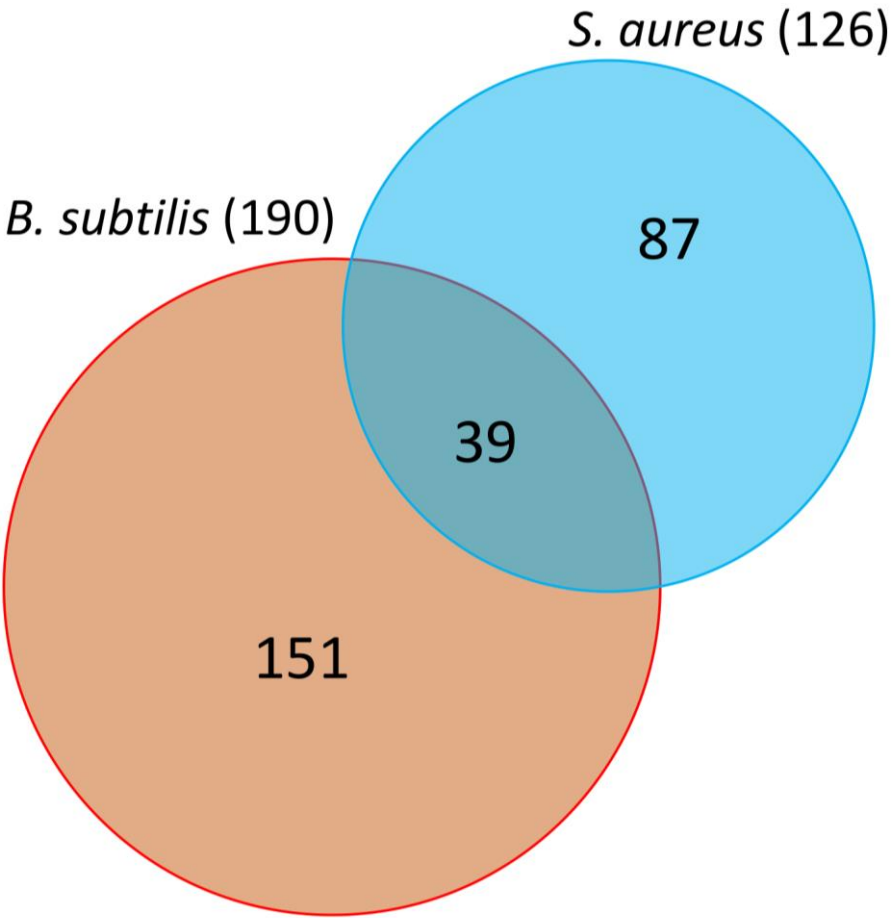
Supplemental data figure 2S:



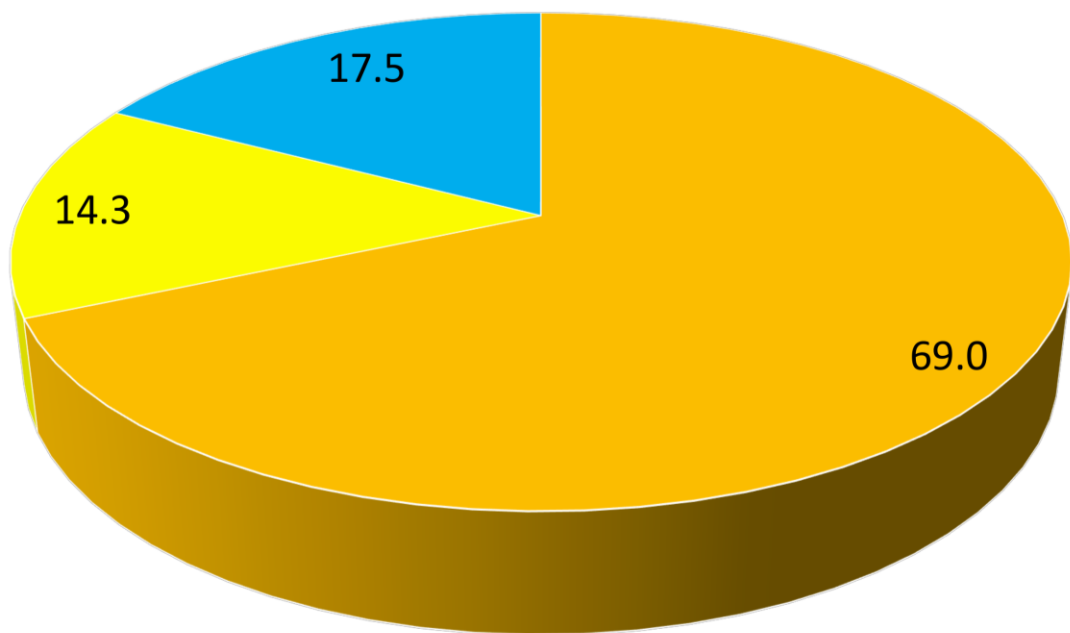
Supplemental data figure 3S:



Supplemental data figure 4S:



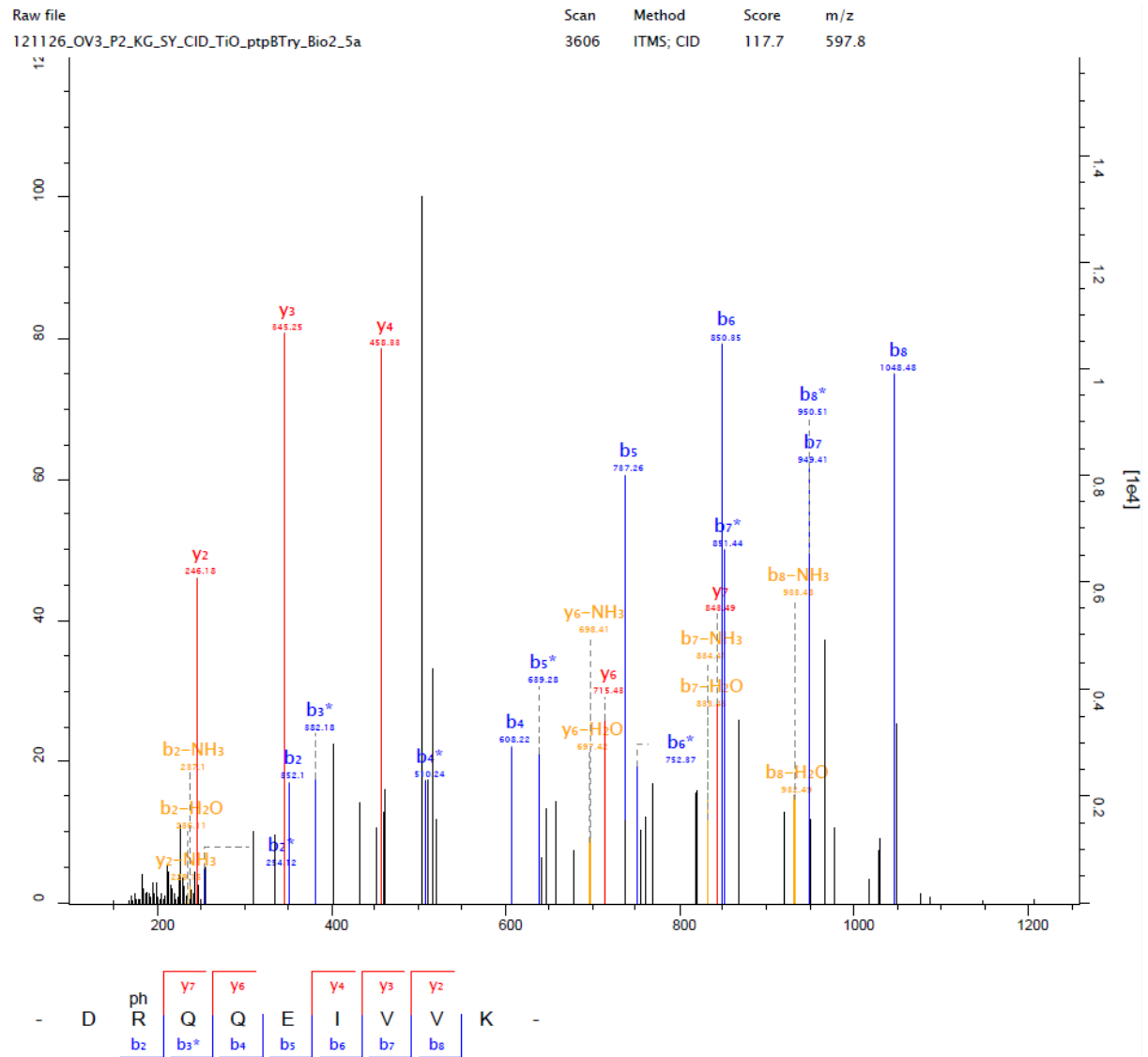
Supplemental data figure 5S:



- in *S. aureus* only
 - in *S. aureus* and *B. subtilis*, but at totally different phosphosites
 - in *S. aureus* and *B. subtilis* with at least one identical or adjacent phosphosite
-

Supplemental data figure 7S:

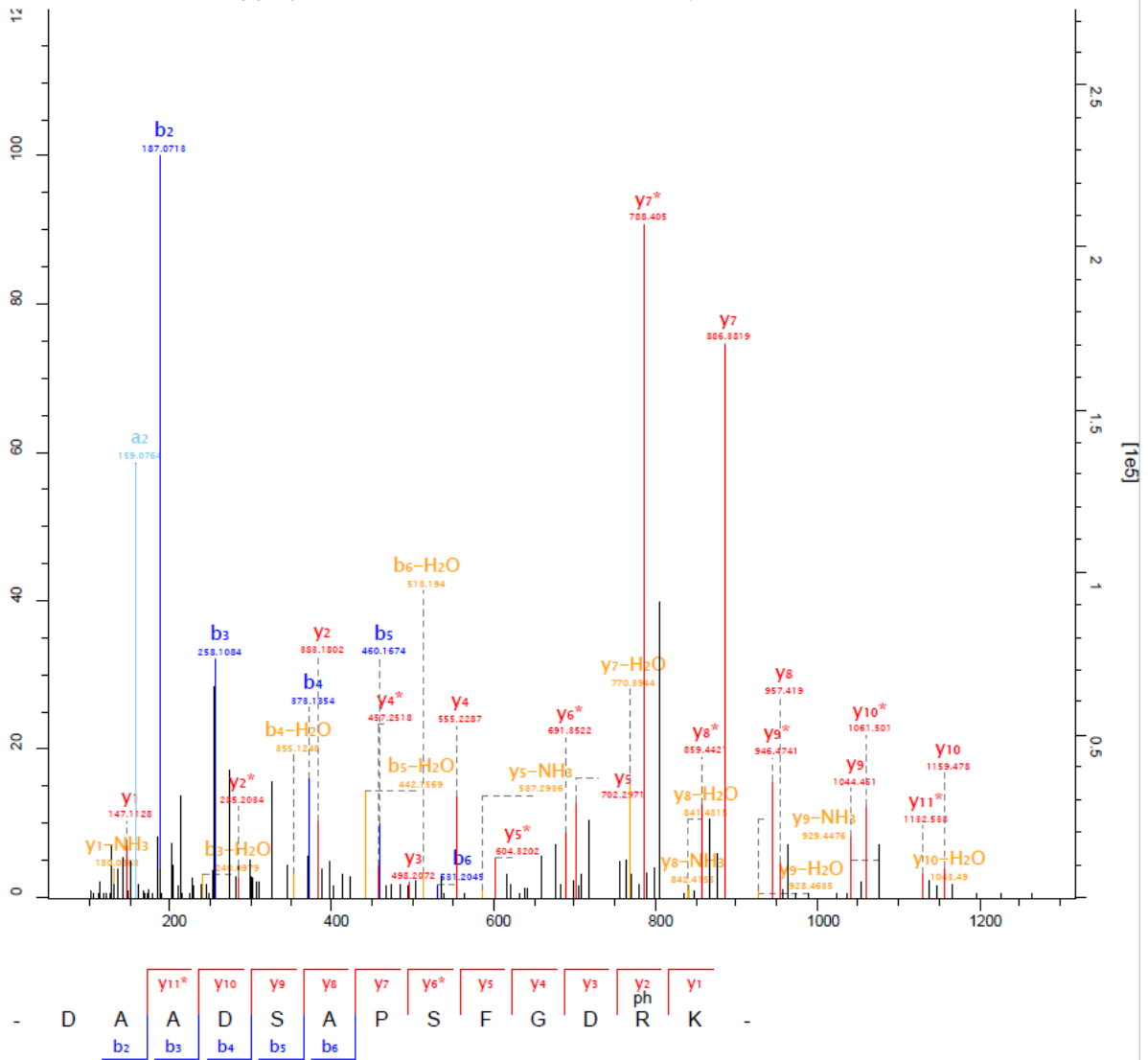
A



B

Raw file
121126_OV3_P2_KG_SY_HCD_TiO_ptpBTry_Bio2_5a

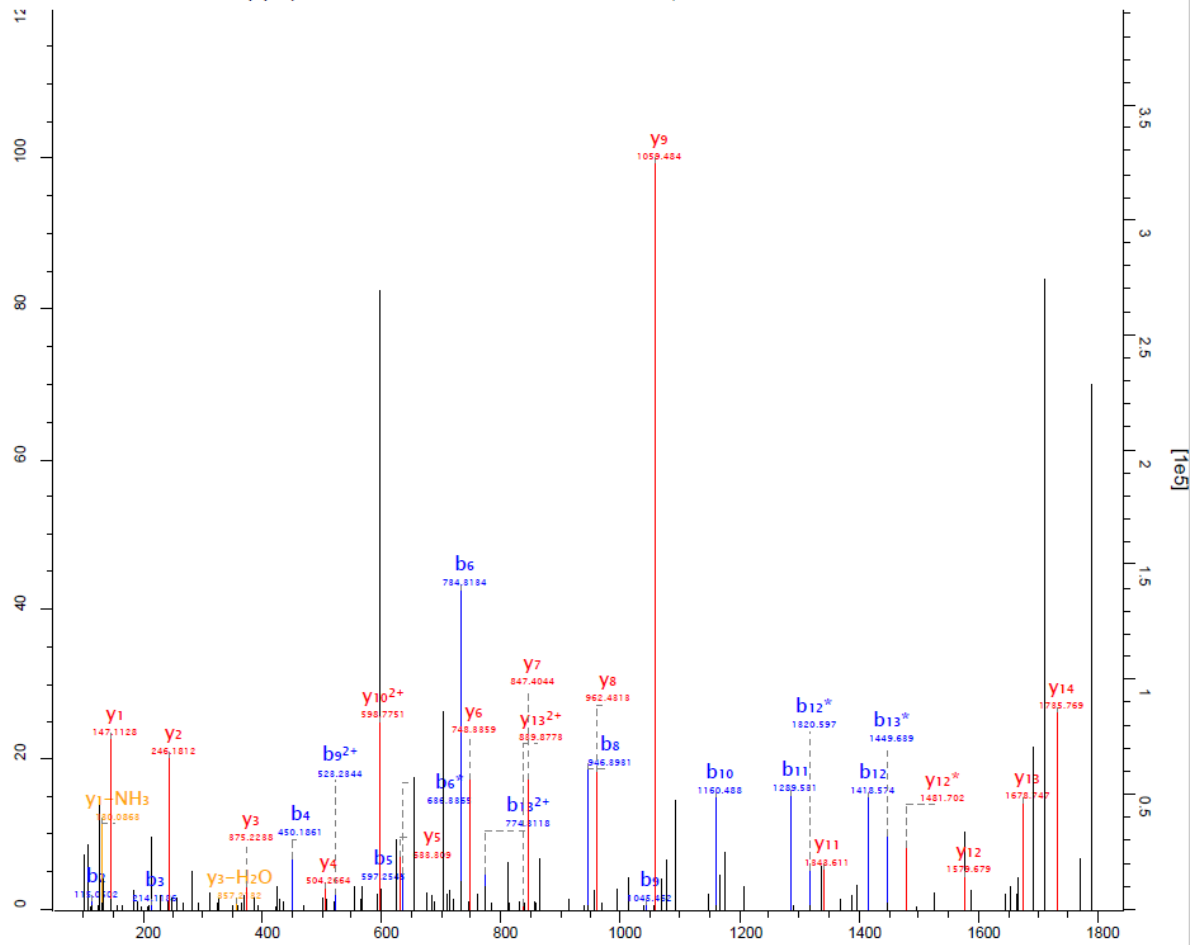
Scan Method Score m/z
3843 FTMS; HCD 140.81 708.79



C

Raw file
121126_OV3_P2_KG_SY_HCD_TiO_ptp8LysCa_9a

Scan 4960 Method FTMS; HCD Score 128.05 m/z 598.27



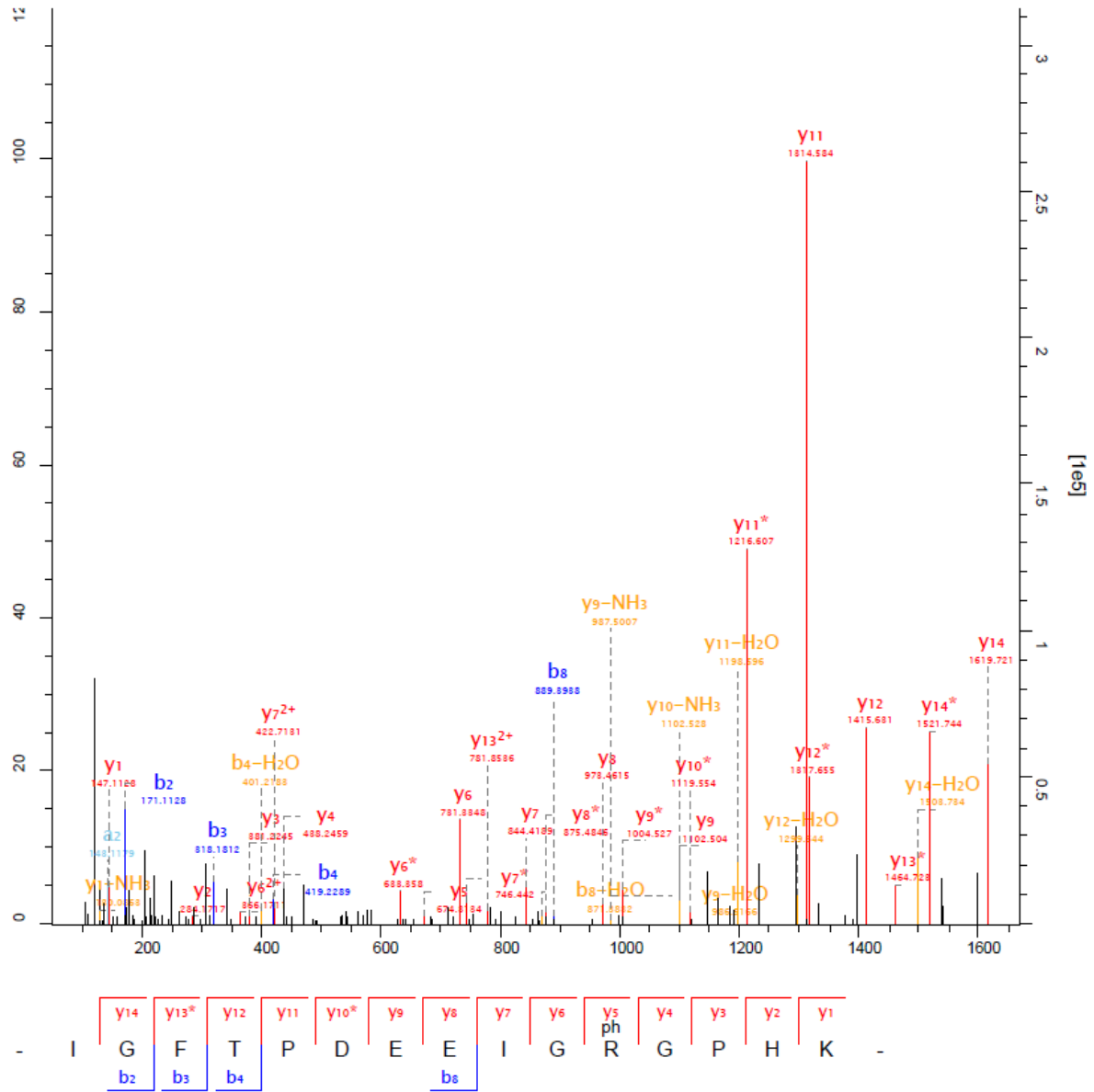
- G y14 y13 y12
ph y11 y10²⁺ y9 y8 y7 y6 y5 y4 y3 y2 y1 -

b2 b3 b4 b5 b6 b6 b9 b10 b11 b12 b13* V K -

D

Raw file
121126_OV3_P2_KG_SY_HCD_TiO_ptp8Try_Bio2_9a

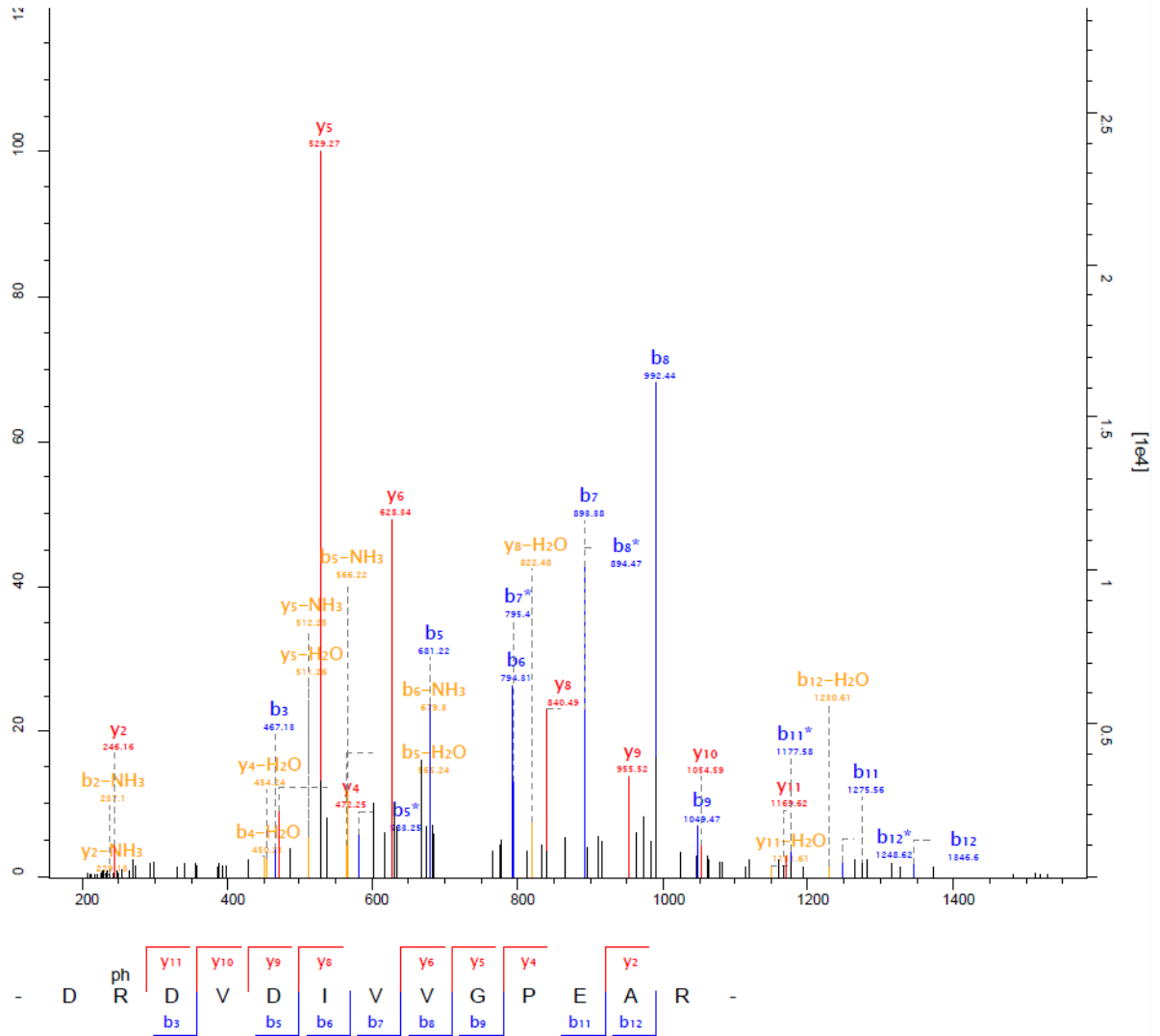
Scan Method Score m/z
5669 FTMS; HCD 207.53 578.61



E

Raw file
121126_OV3_P2_KG_SY_CID_TiO_ptpBTrY_Bio2_4a

Scan 7361 Method ITMS; CID Score 105.99 m/z 760.86



Supplemental data tables

Supplemental data table 1S: Primer sequences used for $\Delta ptpB$ construction

The following primer sequences were used to construct the deletion mutant $\Delta ptpB$

Supplemental data table 2S: Chemicals used for stress conditions

For construction of spectral libraries, different cultivation conditions were applied. Stress conditions were applied using the chemicals and concentrations listed below.

Supplemental data table 3S: Growth conditions for preparation of spectral libraries

For construction of spectral libraries, bacteria were grown according to the growth conditions listed below. The harvest at mid-exponential growth phase corresponds to OD_{540} nm 0.5 - 0.6 in LB and OD_{500} nm 0.5 - 0.6 in CDM. T180 = 180 min after entry into stationary growth phase

Supplemental data table 4S: Chemically synthesised arginine phosphorylated peptides

The table lists all arginine phosphorylated peptides that have been synthesised chemically. Sequences are derived from naturally occurring putative phosphorylated peptides which were successfully identified from biological samples before.

Supplemental data table 5S: Criteria for selection of synthetic peptides subjected to TiO_2 enrichment

The following criteria were considered to select five exemplary synthetic arginine phosphorylated peptides to subject them to TiO_2 enrichment.

Supplemental data table 6S: Fraction of TiO_2 enrichment subjected to LC-MS/MS measurement

The following fractions were separated and analysed by LC-MS/MS.

Supplemental data table 7S: Source of raw files

Origin of raw files for the construction of raw spectral libraries.

Supplemental data table 8S: Phosphorylated peptides identified within at least two replicates of the wild type

Supplemental data table 8S lists all phosphorylated peptides identified within at least two out of four biological replicates of the wild type. It is provided as separate spreadsheet file.

Supplemental data table 9S: Phosphorylated peptides identified within at least two replicates of the mutant $\Delta ptpB$

Supplemental data table 9S lists all phosphorylated peptides identified within at least two out of four biological replicates of the mutant $\Delta ptpB$. It is provided as separate spreadsheet file.

Supplemental data table 10S: Number of identified phosphosites

The table compares the number of identified phosphosites between the wild type and the mutant $\Delta ptpB$, differentiated according to the respective phosphorylated amino acid.

Supplemental data table 11S: Chemically synthesised peptides for PtpB specificity assay

The table lists the phosphorylated peptides used for *in vitro* investigation of substrate specificity of PtpB. Sequences of arginine phosphorylated peptides were derived from naturally occurring putative phosphorylated peptides which were successfully identified from biological samples before.

Supplemental data table 12S: Substrate specificity of PtpB

The table lists all samples prepared to investigate substrate specificity of PtpB. The columns “no enzyme” and “lysozyme instead of PtpB” and the row “no substrate” were used as negative controls. The column “PtpB” displays phosphatase activity when using substrates phosphorylated at arginine or serine/threonine residues. Further, the column “PtpB + inhibitor” was prepared to test whether phosphatase activity is truly performed by enzymatic activity instead of protein side effects. Due to limited available substrate amount, conditions

with “not performed” weren’t tested. Negative results are shown as “-“ whereas positive enzymatic reaction is displayed as “+”.

