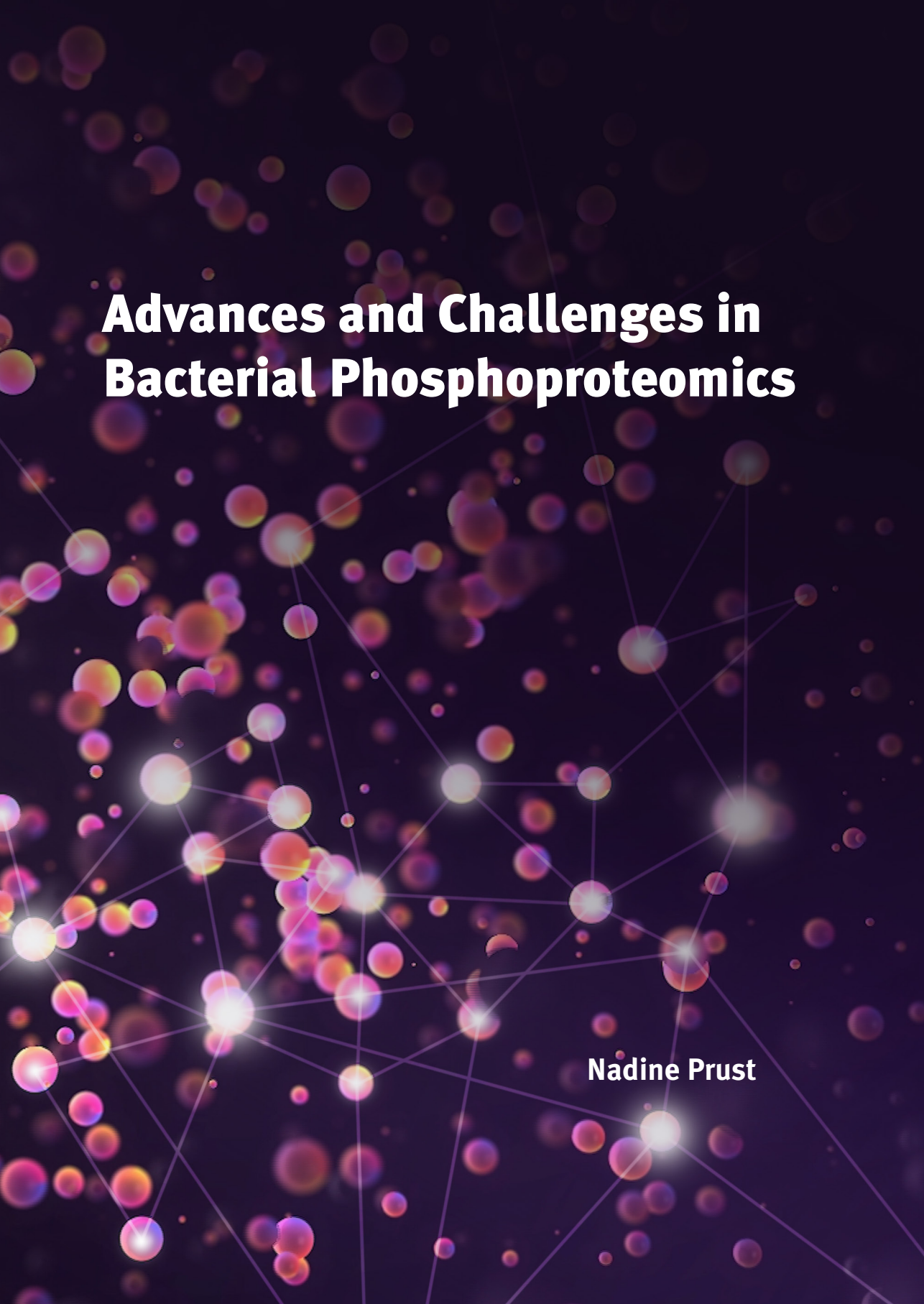
Advances and Challenges in Bacterial Phosphoproteomics



Nadine Prust

Advances and Challenges in Bacterial Phosphoproteomics

Nadine Prust

The research presented in this thesis was performed in the Biomolecular Mass Spectrometry & Proteomics group at the Bijvoet Center of Biomolecular Research, Utrecht University, The Netherlands.

Author:	Nadine Prust
ISBN/ENA:	978-94-6423-693-4
Cover design by:	Wendy Schoneveld | www.wenzidee.nl
Printed by:	ProefschriftMaken || www.proefschriftmaken.nl

Advances and Challenges in Bacterial Phosphoproteomics

Ontwikkelingen en uitdagingen in bacteriële fosfoproteomics

(met een samenvatting in het Nederlands)

Fortschritte und Herausforderungen in der bakteriellen Phosphoproteomik

(mit einer deutschen Zusammenfassung)

Proefschrift

ter verkrijging van de graad van doctor aan de
Universiteit Utrecht
op gezag van de
rector magnificus, prof.dr. H.R.B.M. Kummeling,
ingevolge het besluit van het college voor promoties
in het openbaar te verdedigen op

maandag 21 maart 2022 des middags te 12.15 uur

door

Nadine Prust

geboren op 5 november 1990
te Duisburg, Duitsland

Promotor:

Prof. dr. A.J.R. Heck

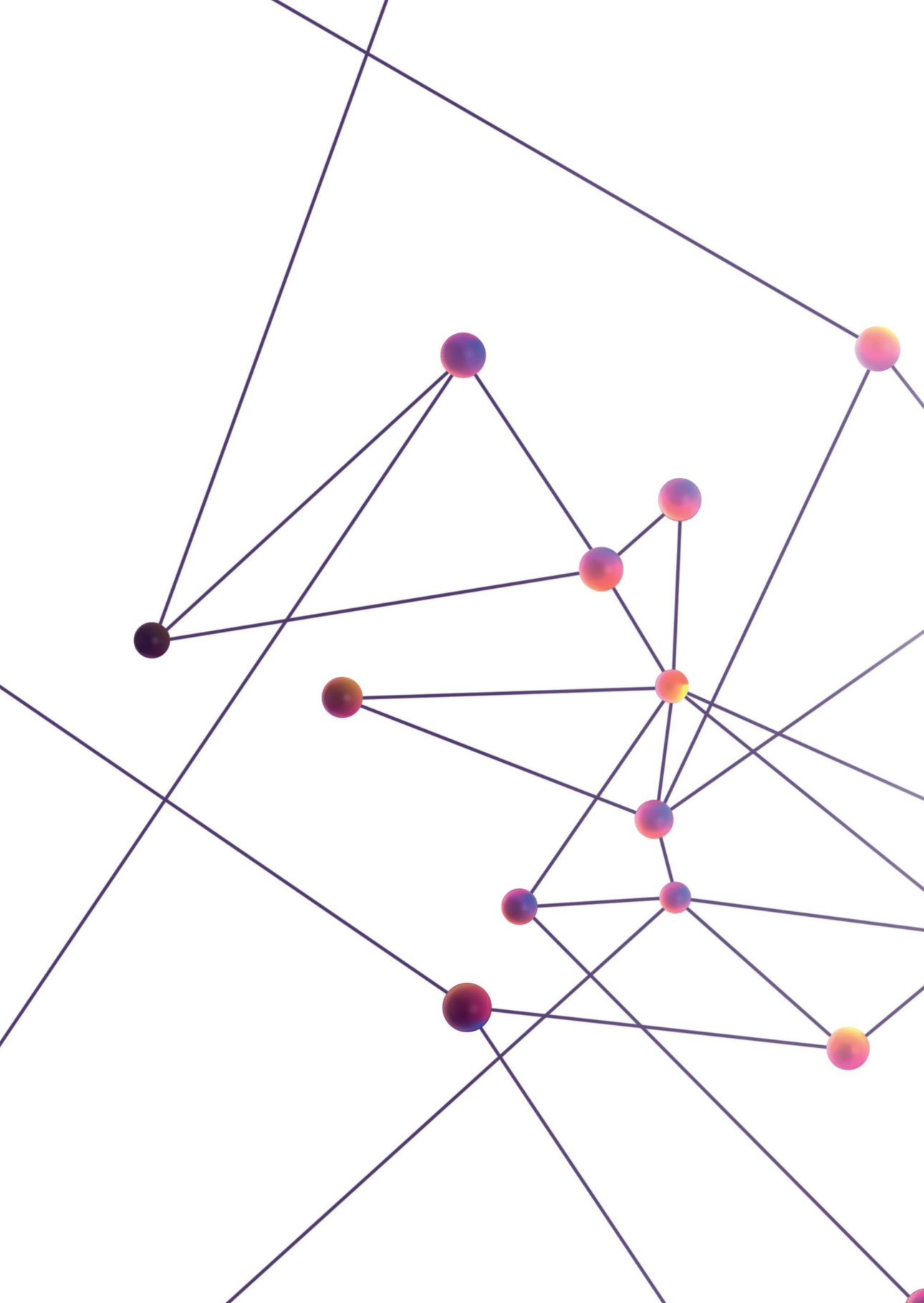
Copromotor:

Dr. S.M. Lemeer

'Imagine how we would be if we were less afraid'

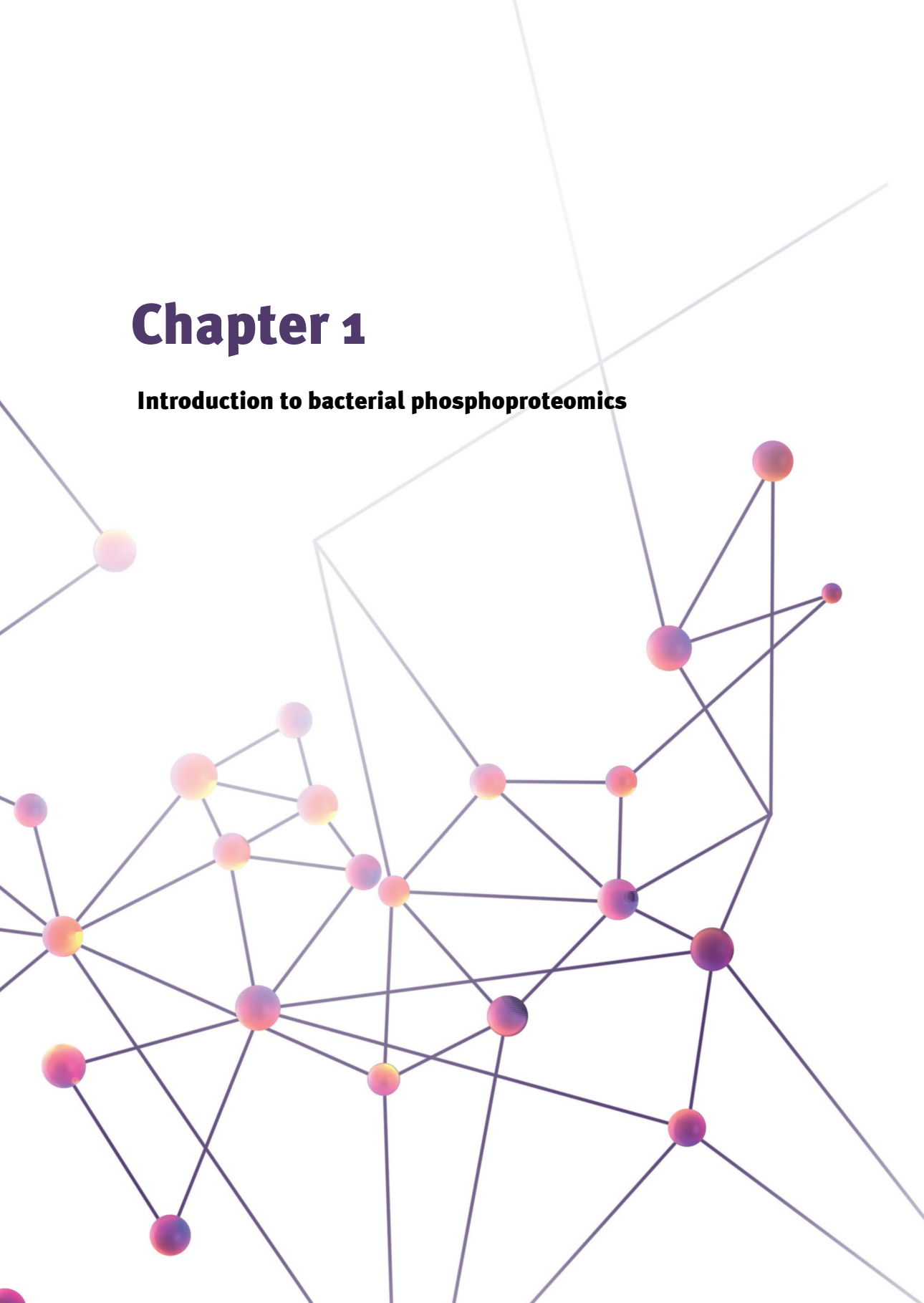
Charlie Mackesy, 'The Boy, the Mole, the fox and the Horse'

Table of content



Chapter 1

Introduction to bacterial phosphoproteomics



1 Introduction to protein phosphorylation

The central dogma of molecular biology postulated by Crick describes that genetic information will be transcribed from DNA into messenger RNA (mRNA) which is further translated into proteins[1]. This translation is facilitated by ribosomes that assemble amino acids via peptide bonds into big macromolecules that form the functional entity of all cells[1]. According to this dogma it was long thought that most diseases could be explained solely by gene mutations. However, today we know that gene mutations do not account for all protein alterations and that for example alternative splicing events, enzymatic processing and post-translational modifications (PTMs) expand the number of estimated proteoforms to above one million[2]. In contrast to the genome, the proteome, consisting of all expressed proteins at any given time and condition, is highly dynamic. Protein abundance, protein-protein interactions (PPIs), cellular localization and PTMs can change over time and thus provide crucial information about cellular processes and their regulation[3].

One of the most important and widespread PTMs that regulates those dynamic processes is reversible protein phosphorylation[4]. It is estimated that 75% of all human proteins are being phosphorylated at some point and it has been shown that protein phosphorylation is regulating a plethora of cellular processes[4–6]. This “success” of protein phosphorylation is based on multiple reasons: (I) phosphate salts are highly abundant on earth, (II) phosphate is chemically versatile and formed phosphomonoesters and phosphoanhydrides are stable under ambient temperatures in water and yet easily hydrolyzed (III) phosphorus can form five covalent bonds, (IV) phosphate is forming a large hydrated ionic shell and thus induces steric hindrance[5]. These unique properties allow the regulation of biological functions by alteration of the protein structure, PPIs, interaction with other small molecules, stabilization of proteins or cellular localization[7]. In order to facilitate this, nature makes use of a variety of enzymes, so-called protein kinases that enable the transfer of the γ-phosphate of ATP to specific amino acid residues in the acceptor protein[8]. Phosphatases can counteract this reaction by catalyzing the dephosphorylation of proteins and thus allow a perfectly balanced on/off regulation of specific cellular processes. Dysfunction of this tight regulation is implicated in a multitude of different diseases and highlighted elsewhere[9].

Mass spectrometry (MS) based proteomics not only offers a high-throughput approach to analyze complex information unrevealed in the proteome but can also be used to identify and localize PTMs such as phosphorylation. Thanks to technological advances over the last decades, mass spectrometry is now considered the gold standard in proteomics research.

2 Mass-spectrometry-based proteomics

MS-based proteomics includes but is not limited to the large-scale study of proteins, protein complexes, PPIs or PTMs. Depending on the desired information different complementary MS-based approaches can be utilized[10]. Bottom-up, also called shotgun proteomics, is the most commonly used type. In short, in shotgun proteomics proteins are enzymatically digested into peptides, in case of phosphoproteomics, phosphopeptides. The latter are enriched before being separated via liquid chromatography (LC) and subsequently analyzed by tandem mass spectrometry (MS/MS). The resulting fragment ion spectra are processed using computational software tools to identify the amino acid sequence of the respective peptides and eventually the protein of origin. In the following sections the different steps of this typical workflow depicted in Figure 1 will be summarized.

2.1 Sample preparation

A typical LC-MS/MS workflow uses cell tissue, cell/bacteria culture etc. as starting material. To extract proteins from this matrix, generally a mix of chemical lysis and mechanical lysis is used to break the cell membranes. The chosen method greatly depends on the origin of the sample. While Gram-negative bacteria for instance only contain a thin peptidoglycan layer that is surrounded by an outer liposaccharide membrane, Gram-positive bacteria lack this outer membrane but instead possess a thick peptidoglycan layer[11]. These structural differences demand harsher conditions for the lysis of Gram-positive bacteria. Also, the protein type of interest is a critical parameter in determining the optimal lysis buffer composition. For example, hydrophobic transmembrane proteins benefit from higher concentrations of chaotropes and the addition of detergents for proper solubilization. Another aspect that needs to be taken into consideration during lysis is the endogenous protease activity after proteins have been extracted. To prevent proteolysis, protease inhibitors are typically added to the lysis buffer. Equally, phosphatase inhibitors are added when interested in protein phosphorylation. Beyond the respective lysis buffer other parameters should be optimized including the method of lysis (e.g., sonication, bead-beating or other shearing forces) or the removal of potential contaminants such as lipids, DNA or RNA. These optimizations greatly influence the success of a proteomics experiment[12]. In the next step the extracted proteins are denatured, and cysteine residues are reduced and alkylated in order to break and prevent the reformation of disulfide bridges and thus the refolding of proteins. Proteins are further digested into peptides using site specific proteases, most commonly trypsin. Trypsin cleaves specifically on the carboxyl terminus of lysine and arginine resulting in short peptides suitable for effective fragmentation[13]. To remove salts or other buffer compounds that could interfere with the subsequent LC-MS/MS analysis the resulting peptide mixture is cleaned up, most commonly using C18 material.

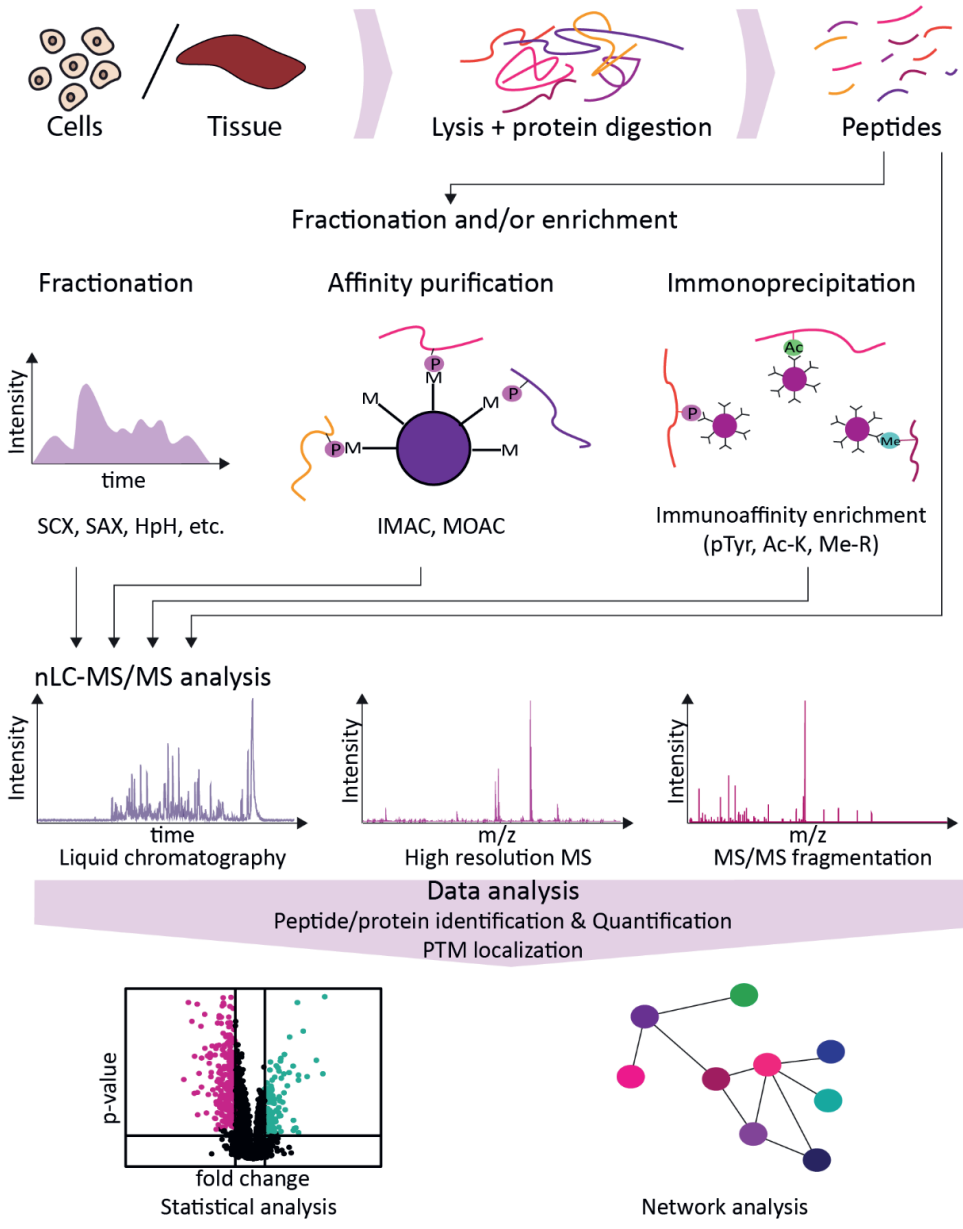


Figure 1: Schematic representation of shotgun MS-based proteomics. All experiments entail sample preparation, mass spectrometric analysis and data analysis. Depending on the exact experiment the sample preparation requires different steps, such as fractionation or enrichment steps. Figure is inspired by Altelaar et al.[10] and adapted from Clement Potel.

2.2 Enrichment and fractionation

The resulting peptide mixtures are generally high in complexity and span a wide dynamic range. Even though mass analyzers have become faster and more sensitive, this high complexity can greatly impair the MS/MS analysis. By performing LC-MS/MS analysis, samples are commonly separated by C18 based reversed-phase high-performance liquid chromatography (RP-HPLC). The hydrophobic C18 material in the stationary phase and the predominantly hydrophilic organic mobile phase separate the peptides based on their hydrophobicity[14]. This decreases the sample complexity presented to the mass spectrometer at any point of time during the gradient and can provide additional information about the peptide's hydrophobicity based on their retention time (RT). However, the detection of low abundant peptides such a phosphopeptides in a background of more abundant peptides is analytically challenging and the sub-stoichiometric nature of protein phosphorylation necessitates a highly efficient enrichment method prior to mass spectrometric analysis. Depending on the type of phosphorylation to be investigated, different approaches exist. Most enrichment techniques exploit the affinity of negatively charged phosphopeptides towards metal oxides such as TiO_2, which is mostly used for metal oxide affinity chromatography (MOAC)[15,16], or metal ions such as Ti^{4+} and Fe^{3+} that are used for immobilized metal affinity chromatography (IMAC)[17,18]. Alternatively, phosphotyrosine immunoprecipitation[19], strong cation exchange (SCX)[20,21] or strong anion exchange chromatography (SAX)[22,23] have been proven to be successful for phosphopeptide enrichment.

To decrease the complexity further and to achieve in-depth analysis of the (phospho)proteome, additional online or offline first dimension chromatographic fractionation can be applied prior to the LC-MS/MS analysis. While online approaches involve the direct transfer from the first to the second dimension, offline methods are based on fraction collection[14]. Every collected fraction contains a peptide pool with slightly different characteristics based on the selected fractionation method. These fractionation strategies comprise for example high-pH reversed-phase (HpH)[24,25] which separates peptides based on their hydrophobicity, ion exchange chromatography (IEX)[20] which is based on peptide charges or size-exclusion chromatography (SEC)[26,27] based on peptide size. SCX, SAX and HpH have been proven as good first dimension separation method for phosphopeptides[21,23,28]. Those fractions can be subsequently subjected to LC-MS/MS analysis and therefore drastically reduce the complexity of the measured sample. Although fractionation improves the (phospho)proteome coverage, it also increases MS-measurement time. To decrease the MS-measurement time without losing the advantage of fractionation, concatenation strategies where fractions from different parts of the gradient are pooled together should be applied[28]. Even though offline approaches are more prone to sample loss due to more sample handling, they offer more flexibility since each

dimension can be independently optimized and fractions can be manipulated[14]. Online approaches on the other hand prevent sample loss but are more restricted.[29] First dimension solvents need to be adapted to not interfere with the second dimension which needs to be relatively fast to not lose the first-dimension resolution[14].

2.3 MS/MS analysis

The basic principle of mass spectrometry is the detection of the mass-to-charge (*m/z*) ratio of analytes, in our case peptides, in the gas phase by manipulation of electric and magnetic fields. However, this information is not sufficient for reliable peptide identification. Therefore, to obtain appropriate sequence information, several fragmentation methods can be used. While there are several different MS instruments available to date, all consist of three main components: (I) an ion source (II) a mass analyzer and (III) a detector. In the following, those components will be described in more detail focusing on the types and methods used for the work described in this thesis.

2.3.1 Ion source and Ionization

Since the introduction of two soft ionization techniques, matrix assisted laser desorption/ionization (MALDI)[30] and electrospray ionization (ESI)[31,32], that induce little or no fragmentation and enabled to bring larger biomolecules into the gas-phase, MS became available for a wide range of applications. As ESI allows ionization of liquid analytes, it is easily coupled to LC systems. Thus, it is the most common ionization method for liquid samples[33]. The principle of ESI is based on applying a high voltage of several kV under atmospheric pressure between a column emitter containing the liquid sample and the MS inlet[34]. The electric potential results in the accumulation of charges at the capillary tip and the dispersion of the solution into charged droplets under formation of a Taylor cone[34]. Heating of the MS inlet to 150-350°C facilitates solvent evaporation to shrink the charged droplets while the charge density increases until the Rayleigh limit is reached[32]. At this point the Coulomb repulsion equals the surface tension and ultimately results in a so-called Coulomb explosion that generates free ions[32] (see Figure 2). The finale mechanism for the ionization of gas-phase analytes is not yet fully understood.

ESI can be performed either in positive or negative ion mode, which results in the formation of either positive or negative ions, respectively. However, the positive ion mode has been proven beneficial for the analysis of most peptides and proteins.

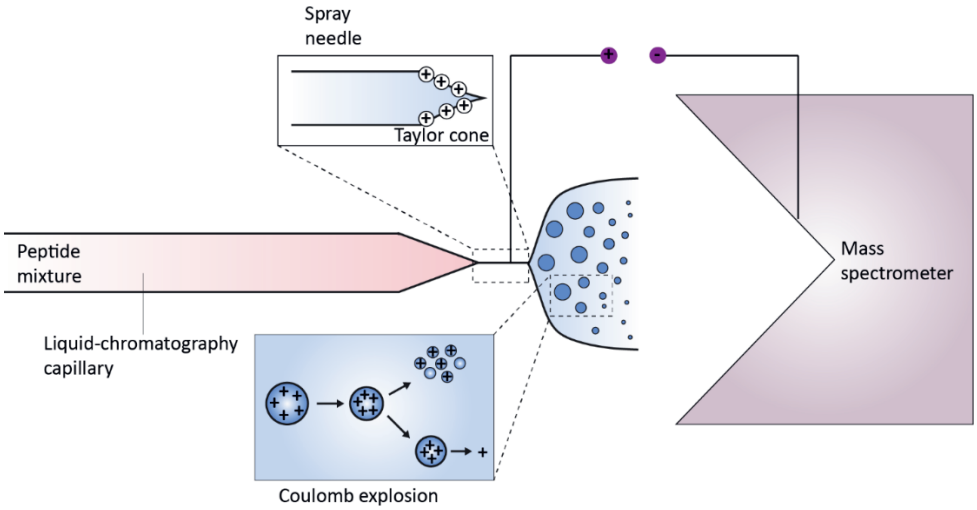


Figure2: Schematic representation of electrospray ionization. The peptide mixture reaches the emitter (spray needle) and the applied electric potential results in the accumulation of charges at the needle tip under formation of the Taylor cone and thus dispersion of the solution. Solvent evaporation leads ultimately to the Coulomb explosion releasing free ions that reach the MS. Figure adapted from H. Steen and M. Mann[35].

2.3.2 Mass analyzers and mass filter

After ionization, the ions are focused into the instrument where ion optics manipulate, separate, and guide the ions through the instrument to finally be measured by the mass analyzer[36,37]. The mass analyzer determines the *m/z* ratio and thus is the heart of every mass spectrometer. Today, several different mass analyzers with distinct characteristics in regard to mass range, mass accuracy, resolution, sensitivity, and speed can be used depending on the experimental setup[38]. For shotgun (phospho)proteomics especially, the speed is an important parameter since it determines the depth of analysis. Modern instruments, so-called hybrid instruments, combine multiple mass analyzers and mass filters. In this thesis Q-Exactive HF, Fusion, Fusion Lumos and Orbitrap Exploris™ 480 mass spectrometers (all of Thermo Fisher Scientific) were used. All those instruments make use of a Quadrupole Mass Filter (QMF) and an Orbitrap mass analyzer, the latter combines high resolution with great mass accuracy (< 5 ppm)[39].

Quadrupole Mass Filter: A quadrupole consists of four parallel, equally spaced electrodes of which two opposing electrodes are electrically connected (see Figure 3a). In general, a quadrupole can serve three different functions, I) as an ion guide system, II) as *m/z* filter and III) as a mass analyzer[40,41]. When applying only an alternating current (AC or main RF) to the electrodes, the ions are radially confined while at the same time the main RF is inducing an ion motion[42]. This ion motion can be described as a cork-screw motion (secular motion) and is proportional to the RF amplitude and mass of the ion[42]. Under these conditions a wide variety of *m/z* ions will be guided through the quadrupole. Applying an

additional direct current (DC) allows the filtering for specific *m/z* values. Again, the RF induces in a secular motion along the x-z axis, with smaller ions having a higher motion than lager ions[42]. If the amplitude of this motion gets too high, ions can collide with the electrodes and thus adjusting the RF can serve as a lower mass cut-off filter[42,43]. The DC works on the y-z axis pulling the ions towards the electrodes while at the same time the radial confinement of the RF forces those ions back in the defined trajectory[42,43]. However, this radial confinement might not be strong enough for larger ions that will eventually collide with the electrodes. Thus, adjusting the DC can serve as a high mass cut-off filter[42]. This allows ultimately the selection of defined *m/z* populations to be guided through the mass analyzer.

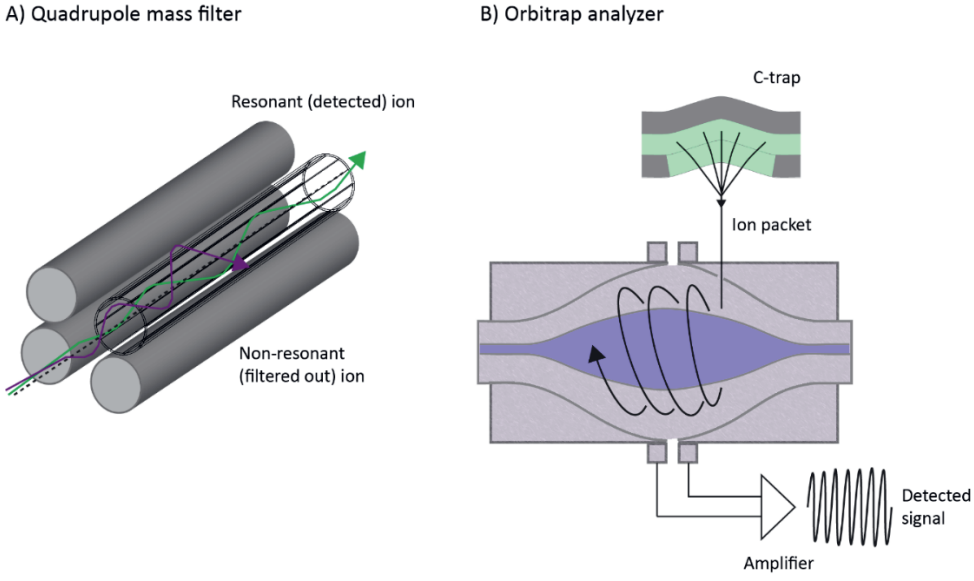


Figure 3: Mass filter and analyzer. A) Schematic representation of quadrupole mass filter. Resonant ions travel on stable trajectories through the four metal rod electrodes while non-resonant ions are filtered out. Reproduced from Savaryn, Toby and Kelleher[42] B) Cross section of an Orbitrap mass analyzer and C-trap. The Orbitrap consists of an outer barrel-shaped and central spindle-shaped electrode. Ion packets are submitted from the C-trap to the Orbitrap and circulate in a harmonic oscillation around the central electrode. After stabilization of the circular motion, the amplifier detects the induced current on the split outer electrodes of the trap. The detected transient is Fourier transformed to generate an *m/z* spectrum. Reproduced from Eliuk and Markarov[44].

Orbitrap: In 2000 Alexander Makarov introduced the Orbitrap mass analyzer[45] which was commercialized in 2005 when implemented in a hybrid instrument[39]. In the orbitrap analyzer the *m/z* ratio is determined by the frequency of harmonic ion oscillation[45]. The orbitrap consists of an outer barrel-shaped and central spindle-shaped electrode (see Figure 3b). Once the ions are accelerated and injected into the orbitrap, ions are trapped due to a balance between electrostatic forces and initial centrifugal forces acting on the ions, resulting in longitudinal harmonic oscillation along the spindle-shaped electrode[39]. The ions

oscillatory frequency along the axis is solely dependent on the *m/z* ratio and detected by the outer electrode resulting in an image current or transient that is fast Fourier transformed (FT) into a high-resolution mass spectrum[45].

2.3.3 Peptide fragmentation

Information about the *m/z* ratio of peptides, even with high mass accuracy, is unfortunately not sufficient for correct peptide assignment and thus proper protein identification. To gain more confidence about the peptide at hand, additional information such as peptide sequence and potential PTMs, and their site localization, is necessary. To obtain this information, tandem mass spectrometry (MS/MS) is performed. Hereto a full MS spectrum (MS1) for all eluting peptides from the LC over the gradient is acquired. This scan allows the selection of certain precursors that are to be submitted for fragmentation analysis. This selection is mostly based on the Top N precursor methods, meaning that the N most intense ions from the MS1 scan will be selected and isolated within the first mass analyzer, e.g. quadrupole mass filter. The isolated ion populations can then be submitted for fragmentation along the peptide backbone. To avoid continuous fragmentation of the same precursor, each precursor is excluded from fragmentation for a defined time after being selected, known as dynamic exclusion. The fragment ions are then transferred to the mass analyzer e.g. Orbitrap and a second MS spectrum, a fragmentation spectrum, will be acquired that allows the identification of the (partial) amino acid sequence as well as potential PTMs. Over the years different fragmentation methods also focusing on the more labile phosphopeptides have been developed. The most common ones are described below:

CID: Collision-Induced Dissociation (sometimes also defined as collisionally activated dissociation, CAD)[46] is induced by accelerating gas-phase protonated peptides by an electric potential in a vacuum, resulting in multiple collision with inert gas molecules (e.g. nitrogen, argon, or helium). This collision increases the internal energies of the peptide ion, distributed over the molecule, by partial conversion of kinetic energy and resulting eventually in the breakage of bonds and thus the dissociation of fragment ions[47]. The resulting fragmentation pattern depends on numerous parameters such as amino acid composition, peptide size, excitation methods, charge, and CID parameters[48,49]. In (phospho)proteomics CID mostly refers to low-energy CID using collision energies of 1-200 eV[50]. The low energy necessitates multiple collisions (tens to hundreds[50]) to transfer sufficient kinetic energy. This energy is randomized amongst all degrees of freedom of the respective peptide ion and vibrational energy induces the dissociation of the peptide ion[51].

Fragmentation under low-energy CID is most acceptably explained by the “mobile proton” model introduced by Gaskell and Wysocki[52]. Peptide ions produced by soft ionization techniques such as ESI are relatively low in energy. Thus, protons need to adapt the energetically most favorable position, this being the N-terminus and basic residue side

chains (e.g. arginine (Arg), lysine (Lys) or histidine (His)). An incremental increase of the internal energy during CID allows the mobilization of the proton along the peptide backbone to less favorable positions generating a population of different protonated forms[50]. Protonation of the backbone, especially the amide nitrogen induces charge-directed cleavage by weakening the amide bond to generate characteristic b- and y-ion[50,53] according to the nomenclature introduced by Roepstorff[54] (see Figure 4).

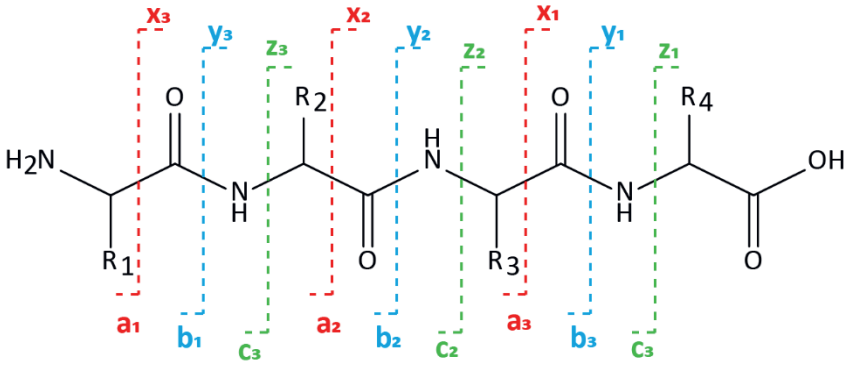


Figure 4: Peptide fragmentation nomenclature according to Roepstorff[54] indicating complementary a/x, b/y and c/z ions. CID/HCD result predominantly in b- and y-ions whereas ETD generates c- and z-ions.

Typically, when performing CID one can distinguish between resonance excitation and beam-type CID. Ion trap CID (IT-CID) is an example of resonance excitation. Peptide ions are trapped and exited using a precursor specific energy corresponding to its secular frequency, subsequently collision with the inert gas induces fragmentation[49,50]. However, the ions can only be exited to a few eV during collision. Thus IT-CID requires multiple collisions over a relatively long time during which extensive rearrangement of the precursor ion can occur before ion dissociation is induced[50]. In the context of phosphoproteomics this rearrangement can include the transfer of phosphate groups to other non-modified serine or threonine residues. However, it has been shown that it has a limited effect on phosphosite localization[55,56]. Another drawback in the analysis of phosphopeptides, when using resonance CID, is that it follows primarily low-energy pathways. The labile nature of protein phosphorylation offers such a low-energy path competing with backbone fragmentation and results in the generation of non-informative fragment ions corresponding to the phosphate neutral loss(es) (HPO_3:80 Da; H_3PO_4:96 Da or H_5PO_5:116 Da) of the precursor ion[53]. Those fragment ions are usually high in intensity and can dominate the CID spectrum, unfortunately providing limited sequence information[57]. In beam-type CID (including *higher energy collision induced dissociation, HCD*[58]) ions are not excited by resonance but by accelerating those ions through an inert gas within a multipole collision cell[50]. This decreases the activation time and thus also decreases the likelihood of gas-phase rearrangements to occur, such as phosphate group transfer[55]. The shorter

activation time and the possibility to optimize collision energies, especially for HCD, also reduces the dominance of phosphate neutral losses compared to IT-CID and common beam-type CID[56,59]. Also, the acceleration of all ions in beam-type CID compared to precursor excitation, allows further fragmentation of neutral loss fragments and therefore provides more sequence information than obtained by resonance CID[60]. Contrary to CID that has a low mass cut-off filter[42,43], the dedicated collision cell for HCD fragmentation allows the screening of almost the entire mass range and thus the identification of signature immonium ions[61]. Immonium ions are the result of further fragmentation of a- or y-ions or of the N-terminal residue in beam-type CID[62]. Since immonium ions contain only one specific amino acid residue, those ions are quite informative. In the context of phosphoproteomics, immonium ions, originating from phosphotyrosine (pTyr)[63] and phosphohistidine (pHis)[64], have a unique *m/z* value and thus serve as an important diagnostic tool. The identification of those ions can not only help to localize phosphosites, but they can also be used to scan specifically for pTyr and pHis and / or trigger their fragmentation and improve confident analysis[65,66].

ExD: As addressed above, for phosphopeptide analysis CID and to a slightly smaller extent HCD suffer from extensive phosphoric neutral loss. Electron-based dissociation methods such electron capture/transfer dissociation (ECD[67] or ETD[68], respectively) have been shown to induce fragmentation, keeping labile PTMs such as phosphorylation unaffected. In comparison to CID, fragmentation is induced without extensive redistribution of intramolecular vibrational energies, but by the uptake of an electron, either a low-energy (< 1 eV), ECD, or from a reagent radical anion (ETD) to generate a radical species that triggers backbone fragmentation[67–69]. The generated radical makes the fragmentation highly specific for N-C_α cleavages producing c'- and z-dot ions and to a lesser extent c-dot and z'-ions with only little side-chain fragmentations[67–69]. Therefore, ExD spectra should in theory result in richer MS spectra with better sequence coverage[70]. Mechanistically both ECD and ETD are quite similar and even though the exact mechanism is not fully understood it is extensively studied[67,71,72]. The two most accepted fundamental concepts are the 'Cornell mechanism'[67] and the 'Utah-Washington mechanism'. McLafferty and co-workers[67] proposed the Cornell mechanism, where the electron attaches to the protonated side. As described earlier these are mostly amine groups of the N-terminus or basic amino acids (Arg, Lys, His) resulting in hypervalent radical N-species. Ground state relaxation induces a hydrogen transfer to an amid oxygen and the formation of a carbon centered aminoketyl radical intermediate[69,71]. Ultimately, the N-C_α bond right-hand side to the radical will be cleaved resulting in the typical c'- and z-dot ions[71]. The Utah-Washington mechanism proposed by Simons and co-workers[72] as well as Turecek[73] and co-workers assumes an immediate formation of an aminoketyl radical intermediate by direct electron capture into the π*-orbital of an amide group[69]. Subsequently, a proton from a charged side is

transferred and induces the N-C_α bond cleavage[69]. Even though ExD fragmentations are mechanistically similar, they deviate from an instrumentational point of view. The low mass cut-off filter of the RF field of quadrupole ion traps does not allow the efficient trapping of electrons. Therefore, ECD is usually implemented in Fourier-transform ion cyclotron resonance (FTICR) mass spectrometers where the strong magnetic fields allow simultaneous trapping of analytes and electrons[74]. However, Fort and co-workers modified a Q Exactive Orbitrap mass spectrometer to extend its fragmentation to ECD and therefore extending the applicability of ECD fragmentation[75]. ETD on the other hand makes use of anion reagents (typically fluoranthene radical anion) as electron donors that can be trapped by RF fields and thus enables the usage of ion trap instruments[68]. Since ion trap instruments are more robust and widespread, ETD has become the ExD method of choice[69,74].

Besides to the obvious advantages of ExD over CID which are, in particular, the theoretically better sequence coverage and preservation of labile phosphosites, ExD also holds drawbacks. While the N-C_α bond cleavage is highly specific for ExD fragmentation, the proline N-C_α bond is an exemption due to it cyclic structure[76]. After the cleavage of the N-C_α bond the generated fragment remains covalently bound via the proline side-chain[68]. This phenomenon, also known as 'proline effect', hampers the fragmentation of proline containing peptides. Furthermore, the efficiency of ETD is highly dependent on the ratio of charge to the number of amino acid residues *i.e*. the charge density[77]. At lower charge densities, non-covalent interactions might result in an electron transfer without dissociation, also known as ETnoD, therefore ETD is performing best for $z \geq 3$[68,77]. This is in particular problematic, as trypsin is the enzyme of choice resulting in mainly doubly protonated peptides[13].

EThcD: To minimize ETnoD events, ETD can be supplied with more energy, supplemental activation, such as beam-type collision dissociation (EThcD, electron- transfer and higher-energy collision dissociation)[78] and thereby increasing the efficiency of ETD-based fragmentation for low charge density populations. Introduced in 2012 by Frese and co-workers[78], unreacted as well as ETD generated fragment ions are subsequently subjected to HCD producing a comprehensive fragmentation pattern of c'- and z-dot as well as b- and y-ions. Implementation of EThcD for phosphoproteomics showed increased sequence coverage when compared to HCD or ETD without supplemental activation, together with improved phosphosite localization[79,80]. In addition, supplemental activation has not been shown to induce extensive phosphoric neutral loss as seen for HCD[80,81]. However, the benefit of EThcD as well as ETD compared to HCD is still controversial with other groups showing a clear advantage of HCD in regard to number of identified phosphopeptides with only slightly worse phosphosite localization[82]. Therefore, for optimal identification and / or phosphosite localization the choice of fragmentation should be adjusted to the respective

analytical question and obviously depends on the available instrumentation and possibly also the operators.

2.3.4 Peptide and Protein identification

Tandem-MS allows to distinguish peptides with the same *m/z* ratio based on their fragmentation spectra that enable the identification of the exact amino acid sequence. To do so, one can follow two different approaches: *de novo* sequencing or *in silico* database supported interpretation. Knowing which fragment ions are produced using the different fragmentation techniques and the residue masses of all 20 amino acids at hand, *de novo* sequencing directly interprets the acquired fragmentation spectra based on the fragment ion peaks and their differences. As described before, HCD produces mainly b- and y- ion series and in an ideal scenario all ions would be identified. Thus, the amino acid sequence of the respective peptide should be determinable by connecting fragments with increasing size from the N-terminus (y-ions) or C-terminus (b-ions)[83]. PTMs such as phosphorylation result in a mass shift (e.g. 79.97 Da, HPO_3) of the targeted amino acids and need to be taken into account. Even though most of the 20 amino acids should be easily distinguishable based on their residue masses, there are four that are analytically challenging because of their isobaric nature, meaning that they have either identical masses (leucine, Leu and isoleucine, Ile) or that their mass difference is only marginal (glutamine, Glu and lysine, Lys). Hence, confident *de novo* sequencing requires high mass resolution and accuracy to distinguish e.g. isobaric amino acids such as Glu/Lys. The main advantage of *de novo* sequencing is that no *a priori* knowledge of the protein/peptide sequence is necessary, allowing for an unbiased interpretation of the MS spectra. This is especially important for the identification of novel peptides and PTMs. Even though manual *de novo* sequencing is still valuable for the validation of identified PTMs, for complex mixtures the sheer amount of high-throughput data disqualifies the manual approach and different algorithms were developed over the years that have all been nicely reviewed[84]. However, this is computationally intensive and most often the question in shotgun experiments is not only which peptide was identified, but rather which proteins were identified in the sample. This means further processing of *de novo* identified peptides using e.g. NCBI protein BLAST (https://blast.ncbi.nlm.nih.gov/Blast.cgi) to identify potential proteins of origin. Therefore, more commonly *in silico* database supported spectral interpretation is used that allows high-throughput analysis. Here, different software tools, MASCOT[85], SEQUEST[86] or Andromeda[87], amongst others, allow a comparison of the acquired spectra to theoretical spectra. Those theoretical spectra are constructed after *in silico* digestion of a reference proteome e.g. an organism specific proteome or a subset of a proteome mostly publicly available from UniProtKB or NCBI RefSeq. The reference proteins are *in silico* digested making use of the cleavage rules of the same proteolytic enzyme as experimentally used and theoretical fragmentation patterns are calculated according to the above-mentioned

fragmentation rules. Experiment specific parameters, such as PTMs, taxonomy and missed cleavages can be specified. Each tool has its own scoring algorithm to evaluate the likelihood of correct matches between the real and theoretical spectrum. In this thesis, Andromeda was the only search engine used. This probability-based scoring algorithm calculates an Andromeda score for each peptide-spectrum comparison, determining how likely an acquired fragmentation spectrum matches a theoretical spectrum by chance[87]. Thus, the spectrum with the lowest probability score, the lowest probability to match by chance, is considered as the most significant hit[88].

To limit false positive matches Andromeda makes use of a target-decoy approach. A decoy database contains reversed or randomized protein sequences from the target database and both decoy and target databases are merged to a composite database[89]. All spectra are matched against the composite database enabling an estimation of the number of false positives (random matches) and thus the *False Discovery Rate* (FDR), the proportion of false peptide spectrum matches (PSM) among positive PSMs[89,90]. Per default Andromeda accepts peptides up to an 1% FDR and has thereby a second statistical measure for confident identification[91]. After confident peptide identification, peptides need to be assigned to proteins of origin. This is one of the biggest challenges of shotgun proteomics, as the same peptide can be present in multiple proteins or protein isoforms and thus leading to ambiguities in determining the identities of sample proteins[92]. This issue is also known as the protein inference problem. In order to deal with this issue, most search engines, including Andromeda implemented in MaxQuant[91], group proteins which can be explained by shared peptides and are thus indistingusihable[91].

2.3.5 Phosphosite localization

In phosphoproteomics identification of phosphorylated peptides and proteins is, however, not sufficient to understand the molecular importance of the respective phosphorylation. Only further knowledge about the exact localization and occupancy provides significant information to understand its function. Unfortunately, gaining this sort of information proved to be even more complex than identifying phosphopeptides in the pool of unmodified peptides. As described earlier, phosphopeptides are prone to phosphoric neutral loss especially when using collision induced fragmentation[53]. Even though it does not necessarily need to impair phophopeptide identification, it can hinder phosphosite localization as resulting fragments and non-modified fragments can have the same mass and thus be indistinguishable[56]. Further, confident phosphosite localization requires comprehensive peptide sequence coverage to distinguish between multiple phosphosite candidates. For reliable phosphosite localization the different search engines exploit mainly two different strategies: (I) for each phosphopeptide isoform the probability of an incorrect match is calculated or (II) the score difference between different phosphopeptide isoforms

is calculated. PTM-score[93] implemented in MaxQuant/Andromeda uses the probabilistic approach. The different localization algorithms have been nicely reviewed elsewhere[88,94] but can unfortunately result in quite a discrepancy. This is due to different prerequisites such as which type of fragment ions are considered for the scoring, whether phosphate neutral losses are taken into account, or the amount of the most intense peaks considered per *m/z* window[88]. In addition, it is important whether the algorithm was designed for high or low mass accuracy[88]. Contrary to peptide identifications where the FDR serves as second statistical discrimination tool, such a *false localization rate* (FLR) is not employed by most phosphosite localization algorithms, instead arbitrary cut-offs are typically chosen, such as the 0.75 Andromeda localization probability corresponding to so-called class I phosphosites[93]. The main reason why most algorithms do not make use of a FLR is that peptide identification with incorrect phosphosite localizations are not random matches and too similar to the correct match and thus the decoy-based FDR cannot provide an error estimate[88]. Therefore, random manual inspection of phosphopeptide spectra, especially of less common phosphorylations identified without an FLR should still be considered. Alternatively, the use of synthetic analogous with known phosphosite localization can serve as validation tool. The acquired spectra of those synthetic analogous can either be manually compared[64] to the endogenous spectra or within a spectral library[95]. Spectral libraries contain high-quality identified spectra e.g. synthetic phosphorylated peptides that can be used similarly to a reference proteome database[96]. The advantage of spectral libraries compared to an *in silico* generated database is a smaller search space and thus improved search speed and sensitivity[97]. Additionally, the endogenous peptide spectra are compared to a real spectrum allowing to take peak intensity as well as minor peaks into account and therefore optimizing the identification[97].

2.4 Quantitative phosphoproteomics

Proteins and PTMs, especially protein phosphorylations, are highly dynamic and change not only over time, but also under different conditions and thus provide crucial information about cellular processes[3]. To understand those changes it is not sufficient to simply identify proteins and phosphorylations, but rather to quantify those. However, MS *per se* is not a quantitative method, since e.g. different peptides have unique characteristics determining their behavior in MS detection. Therefore, several specific quantitative strategies have been developed over the years. These can be categorized into label free and stable isotope labeling. Labeling can be performed at the protein level (e.g. SILAC[98]) or the peptide level (e.g. dimethyl labeling[99] or tandem mass tags, TMT[100]). Those stable isotopes increase the mass by fixed incremental amounts thus allowing multiplexing that enables the quantification over several different conditions within a single LC-MS/MS run[101]. This quantification can either take place at the MS1 or MS2 level. MS1 quantification uses the elution profiles of differently labeled peptides as comparison, whereas MS2 quantification

compares the intensity of specific reporter ions in the fragmentation spectrum[101]. An advantage that is dependent on the labeling method and thus the time point when the label is introduced, is the minimization of experimental variation and quantification errors[102] (see Figure 5). The sooner samples are combined in a quantitative proteomics experiment, the less experimental variation can be introduced. Since in this thesis only label-free quantification (LFQ) was used, I just refer here to excellent reviews for further comparisons of different labeling strategies[101–103]. In LFQ two approaches can be distinguished: spectral counting or spectral intensity (also known as area under the curve). Both allow a relative comparison between different samples without introducing any isotope label. Spectral counting is less common nowadays and is based on counting the number of fragment spectra that identify a specific protein assuming that more abundant proteins and thus more abundant peptides will be selected more often for fragmentation[103]. Assuming that the area of a peak reflects the number of ions being detected for the particular *m/z*, spectral intensity uses, as the above mentioned MS1 quantification strategies, peptide elution profiles in different LC-MS/MS runs for comparison[101,103]. The advantage of LFQ is that the number of conditions that can be compared is not predefined by the number of available labels and thus allows a rather simple way to compare different conditions. However, this simplicity comes also along with two basic prerequisites: (I) high mass resolution and mass accuracy to minimize interfering signals from co-eluting peptides and (II) robust chromatographic setup with narrow LC peak width and high retention time stability for proper alignment[103]. Moreover, one should not neglect the increased measurement time, since every sample needs to be measured individually. Therefore, not only sample preparations but also MS measurements especially need to be consistent between different samples. To account for differences due to sample preparation, normalization should be performed as implemented in MaxQuant's MaxLFQ[104]. The simplicity and flexibility of LFQ along with constant improvement of instrumentations and thus higher mass resolution, mass accuracy and high peptide identification made it one of the most popular quantification methods.

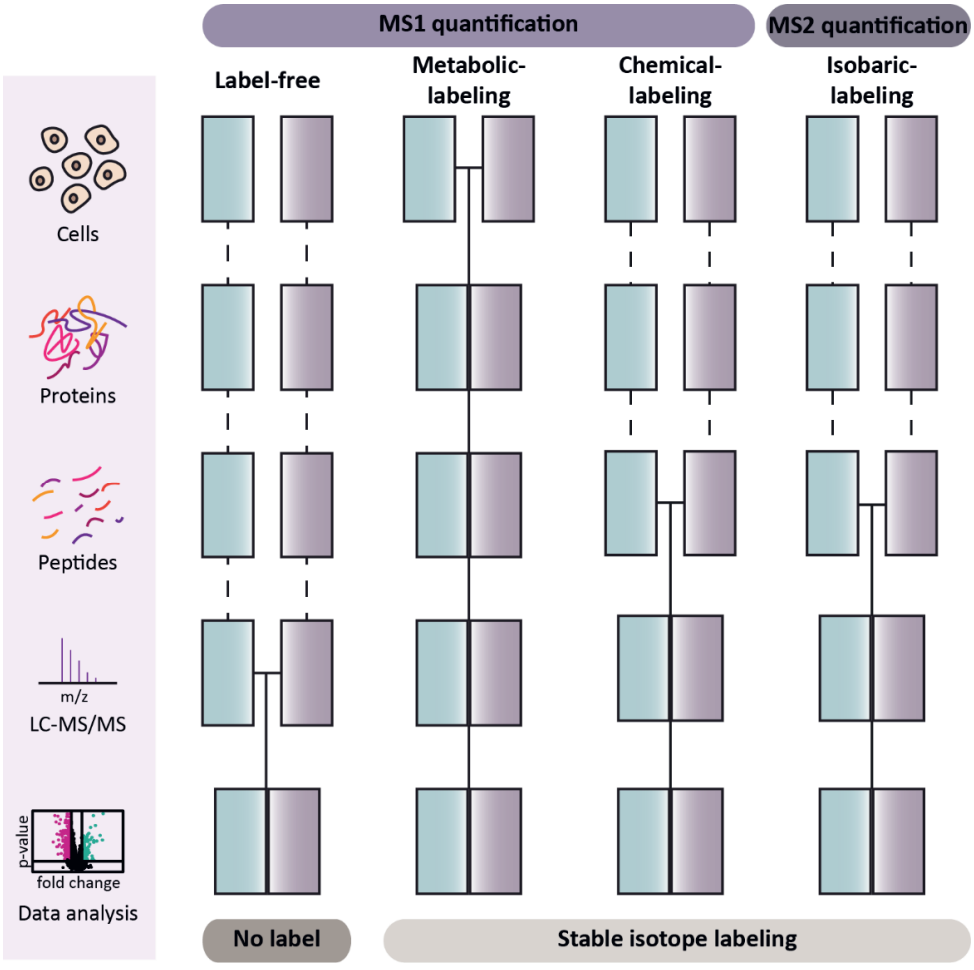


Figure 5: Different quantification strategies used in (phospho)proteomics. The strategies can be differentiated based on the level of the analysis by MS level, MS1 vs MS2 quantification and whether a form of stable isotope labeling is used. Blue and purple boxes represent different experimental conditions. Horizontal lines indicate at which point samples are combined. Dashed lines indicate steps where experimental variation and thus quantification errors can be introduced. Adapted from M. Bantscheff, S. Lemeer, M. Savitski et al.[103]

3 Bacterial phosphoproteomics

Similar to eukaryotes or maybe even to a larger extent, bacteria need to aid quick and efficient cellular signaling in order to adapt to e.g. new environments after host infiltration. To respond to the new surroundings, they need to be able to regulate gene transcription, protein expression, protein localization, or secretion of virulence factors etc[105–108]. These adaptations and cellular changes are largely facilitated by protein phosphorylation. Due to its intrinsic characteristics phosphorylation offers a quick, regulated and reversible mechanism to transfer information within a cascade of proteins from one cellular compartment to another. Protein phosphorylation can occur on nine different amino acids

(Ser, Thr, Tyr, His, Lys, Arg, Asp, Glu and Cys) under the formation of four different linkages, all with different thermodynamic characteristics[109] (see Figure 6). Ser, Thr and Tyr are forming a phosphomonoester bond that is thermodynamically relatively acid stable but alkaline instable[109,110]. Addition and thus formation of a phosphomonoester is achieved by different protein kinases that transfer the γ-phosphate from ATP to the respective amino acid. Corresponding phosphatases can remove the phosphorylation by hydrolyzing the phosphomonoester. His, Lys and Arg can form high energetic phosphoramidate bonds which are acid labile and alkaline stable[111]. Thermodynamically comparable to phosphoramidates, Asp and Glu form a mixed anhydride or acyl phosphate. These different chemical properties result in an altered phosphate transfer compared to phosphomonoesters and allow an even faster transfer[112]. Further, the high energy bond of phosphorylated arginine has been shown to be even sufficient to prompt the conversion from ADP to ATP[111]. Lastly, Cys can form phosphothiolates which have a relatively high free energy as well[109]. The higher the free energy, the more labile these phosphorylations, which makes it more difficult to detect phosphoramidates, acyl phosphates or phosphothiolates compared to phosphomonoesters[109]. The different thermodynamics and kinetics define the stabilities of these distinct nine phosphorylation sites and thus also their biological functions.

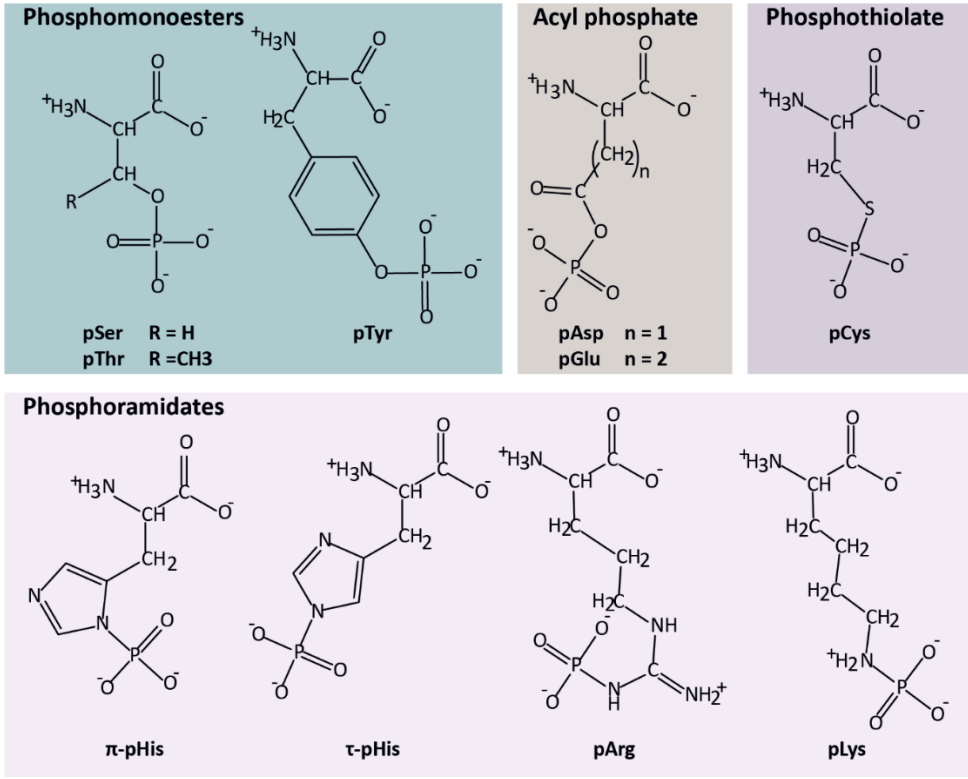


Figure 6: Chemical representation of the nine known phosphorylated amino acids under the formation of four different linkages, phosphomonoester (pSer, pThr and pTyr), acyl phosphate (pAsp and pGlu), phosphothiolate (pCys) and phosphoramidates (pHis, pArg and pLys).

3.1 Protein histidine phosphorylation

In prokaryotes signal transduction is largely regulated by protein histidine phosphorylation within the so-called two-component systems (TCS)[108,113]. TCS are identified in almost all prokaryotes and comprise a membrane embedded sensory histidine kinase and an intracellular response regulator (RR) (see Figure 7). Upon extracellular stimuli the histidine kinase is activated, dimerizes, and induces auto-phosphorylation of its intracellular kinase domain facilitated by ATP[114]. This phosphorylation is then rapidly transferred to a conserved Asp residue of a respective RR[114]. Phosphorylation of the RR induces a conformational change that mainly results in the regulation of its DNA binding capacity and thus an alteration of gene transcription[115]. TCS are the first and best characterized phosphorylation systems in prokaryotes and several excellent reviews can be found[107,108,116,117]. TCS have been implicated in the regulation of cell wall synthesis[115], expression of virulence factors[118], biofilm formation[119] or resistance to oxidative stress and vancomycin resistance[120] for instance. The number of TCS can drastically differ between organisms and might be correlated to the variety of environmental signals they are exposed to[108]. For example, though highly dependent on the strain, in *Escherichia coli* (*E. coli*) around 36 TCS[105,121] have been identified, whereas only 16 TCS[122] have been reported in *Staphylococcus aureus* (*S. aureus*).

3.2 Protein serine/threonine phosphorylation

While phosphorylation on Ser, Thr and Tyr has been very well studied in eukaryotes already for years, it was long assumed this type of phosphorylation is not present in prokaryotes. With technical advances and the identification of the first eukaryotic-like serine/threonine kinases (eSTKs) or also called Hanks-type kinases this perception changed and Ser/Thr phosphorylation is now known to play a crucial role in prokaryotes[123,124]. Hanks-type kinases show structural similarities to the eukaryotic protein kinase superfamily[123]. The kinase domain is usually built up by 12 subdomains, also called Hanks domains, that form a characteristic two-lobed catalytic core structure with the active site hidden in deep cleft between the lobes[125,126]. Membrane embedded Hanks-type kinases have an extracellular domain consisting of multiple repeats also known as PASTA (penicillin-binding protein and serine/threonine kinase associated) domains[127] (see Figure 7). These PASTA domains are characteristic for prokaryotes, especially for Gram-positive bacteria and can vary in their number of repeats between different kinases[127]. Ligand binding induces dimerization and autophosphorylation of the kinase domain, promoting activation of several signaling pathways involving cell growth, cell division, biofilm formation, PPIs etc. To ensure a tight

regulation of those mechanisms, prokaryotes, similar to eukaryotes, use respective phosphatases (eukaryotic-like Ser/Thr phosphatases, eSTP) that have been identified in almost all prokaryotic phyla to facilitate the hydrolysis of the respective phosphorylations[128].
Further, cross-talk between Hanks-type kinases and TCS have been reported in different organisms. For instance, Serine-Threonine kinase 1 (Stk1) in *S. aureus* phosphorylates the RR of the TCS GraRS. GraRS regulates modification of wall teichoic acid (WTA) by $_D$-Ala and phosphorylation by Stk1 enables a secondary activation mechanism independent of the signals sensed by GraS[115].

In addition to Hanks-type kinases, protein serine/threonine phosphorylation can also be induced by bifunctional kinase/phosphatases. Those enzymes can catalyze both the phosphorylation as well as dephosphorylation. One example is Hpr kinase/phosphorylase (HprK/P) which is involved in the phosphotransferases system (PTS) that enables the transfer of sugars across the cell membrane.

3.3 Protein tyrosine phosphorylation

The majority of protein tyrosine phosphorylation in bacteria is facilitated by bacterial protein tyrosine kinases (BY-kinases) consisting of an extracellular loop domain and a cytosolic catalytic domain[129]. Their catalytic domain is characterized by four conserved structural motifs, a Walker A and B motif that constitute the ATP binding site, an additional Walker B motif termed A' and a less conserved C motif containing multiple Tyr sites representing the autophosphorylation site[129,130]. BY-kinases can either be present as one single-membrane protein or as two individual polypeptides which become active upon interaction[129] (see Figure 7). BY-kinases have been shown to be involved in several signaling pathways e.g. CapAB is involved in the peptidoglycan biosynthesis of *S. aureus*[131]. As for Hanks-type kinases BY-kinases have respective phosphatases that hydrolyze tyrosine phosphorylation, mainly low-molecular-weight proteases (LMW PTPs)[132]. Interestingly, some pathogens secrete tyrosine phosphatases after host infiltration such as *Mycobacterium tuberculosis* to escape apoptotic activity of microphages[133].

3.4 Protein arginine phosphorylation

The first protein arginine kinases, McsB, in *Bacillus subtilis* was only identified in 2009 by Fuhrmann and co-workers[134] and a first structure was only elucidated in 2019[135]. McsB composes a N-terminal catalytic ATP: guanido phosphotransferase domain (PD) and a C-terminal dimerization domain (DD) which forms a linear dimer (see Figure 7) and thus shows phosphagen kinase (PhK) activity[135]. The respective phosphatase was identified in 2012 being a LMW-PTP, YwlE, previously identified as tyrosine phosphatase[136–138]. Similarly, tyrosine phosphatase PtpB has been identified as arginine phosphatase in *S. aureus*[95].

Recent studies on protein arginine phosphorylation have linked McsB intricately to several cellular stress responses and protein quality control such as the ClpP-ClpC degradation system[136,139,140]. McsB regulates the expression of ClpCP by phosphorylating their transcription repressor CtsR and by arginine phosphorylated induced activation of ClpC[134]. Further, pArg serves as degradation tag for proteins within the ClpCP system[141]. This demonstrates that arginine phosphorylation can serve at least two different functions, (I) a tight regulation to activate or inhibit certain proteins such as CtsR and (II) a general, more unspecific, function enabling the degradation of proteins especially within the context of extracellular stresses.

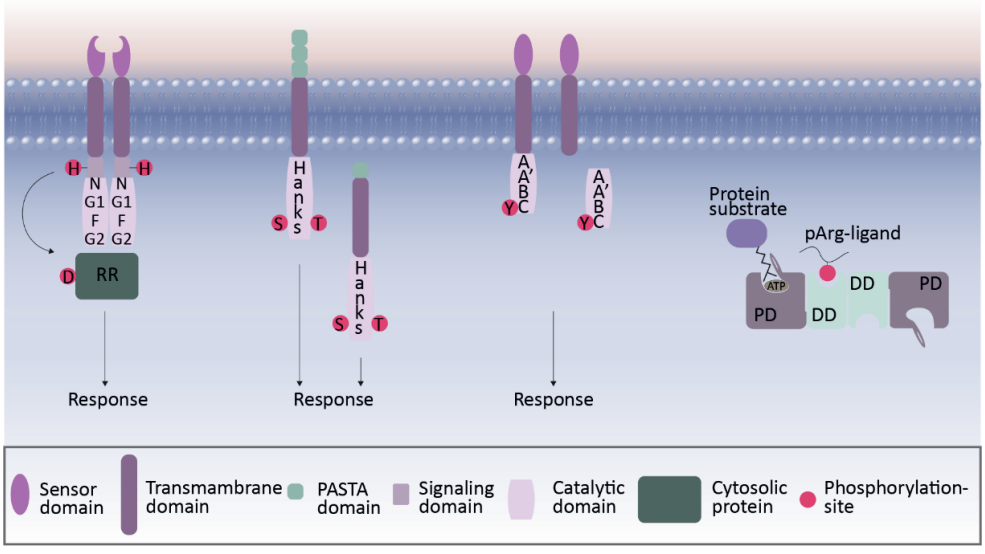


Figure 6: Overview of four types of bacterial phosphorylation systems. From left to right: I) Sensory histidine kinases within the two-component systems. The sensory histidine kinase is composed of an extracellular sensor domain, a transmembrane domain, a signaling domain containing a conserved histidine residue and the catalytic domain with the signature sequences N, G1, F and G2 that enable ATP binding. Upon extracellular stimuli, dimerization and autophosphorylation of the cytosolic catalytic domain is induced. The phosphorylation is rapidly transferred to the response regulator (RR). II) Hanks-type Ser/Thr kinase can be membrane embedded or cytosolic and contain the twelve characteristic Hanks domains. Especially membrane embedded kinases contain several PASTA domains that can function as sensory domain. III) BY-kinases are characterized by their four cytosolic Walker motifs A, A', B and C. BY-kinases can either be present as a single membrane protein or as two individual polypeptides that get activated upon interaction. IV) Arginine kinase McsB is a cytosolic protein that contains a terminal catalytic ATP:guanido phosphotransferase domain (PD) and a C-terminal dimerization (DD) domain which form a linear dimer. Inspired by G. Pagano and R. Arsenault[142] as well as I. Mijakovic, C. Grangeasse and K. Turgay[130].

3.5 Advances and challenges in bacterial phosphoproteomics

As outlined above, nowadays MS-based proteomics offers great advantages for the identification and localization of phosphorylation events. Advances in enrichment strategies as well as fragmentation, mass spectrometric detection, and phosphorylation site localization assessment made sensitive, accurate and robust large scale phosphoproteomics

analysis feasible[101,143,144]. However, bacterial phosphoproteomics was, and still is, much more challenging. Firstly, due to the even lower abundance of phosphorylation, bacterial phosphoproteomes are more difficult to explore compared to eukaryotic phosphoproteomes[145]. Secondly, since eukaryotes mainly contain phosphomonoesters, sample preparation workflows and enrichment strategies are optimized for this group of phosphorylations. Protein histidine and aspartate phosphorylation, long assumed to be the most prominent type of bacterial phosphorylation, pose an intrinsic problem of low stability especially under conditions used for standard phosphopeptide enrichments[109,145]. Thanks to optimizations especially in sample preparation[12] and enrichment strategies[12,139,146] knowledge about bacterial phosphorylation has greatly expanded over the last years. In particular, removal of interfering components such as phospholipids or nucleic acids, e.g. indicated by the presence of cAMP (330.06 *m/z*), prior to phosphopeptide enrichments have been shown to improve the phosphopeptide identification by a factor 10 compared to a standard protocol[12]. Also, reversing the perception that pHis is the most abundant type of phosphorylation, showing that pSer and pThr are much more abundant, changed the research focus. Bonne Kohler and co-workers[147] provide an extensive overview on how many phosphosites in different bacteria based on gel-free MS approaches have been identified. This revealed huge discrepancies between the different organisms reaching from 25 phosphosites for *Leptospira interrogans*[148] to 2,500 phosphosites in *E. coli*[149] demanding more organism specific workflows. Even though bacterial phosphoproteomics substantially progressed in recent years we are still far away from a comprehensive understanding of bacterial phosphorylation and phosphorylation mediated signaling. This conclusion actually formed the foundation for the research described here in my thesis.

Scope of this thesis

Despite recent instrumentational, methodological and bioinformatics advances, bacterial phosphoproteomics remains a challenging field where we are still trying to understand the role the different types of phosphorylation events are playing. In contrast to eukaryotes, bacteria utilize a more comprehensive signal transduction machinery that allows not only quick and efficient signaling but also adaptations to new surroundings after e.g. host infiltrations. Furthermore, bacteria are not only able to phosphorylate and recognize phosphorylated bacterial proteins, but also host proteins. This makes the bacterial phosphoproteome much more complex, as it can change completely after infiltration of different host organisms. To gain a comprehensive view of signaling cascades within bacteria we first need to be able to identify as many phosphorylation events as possible before we can analyze the phosphoproteome after host interaction and thus their pathogenicity.

In **chapter 2**, I discuss the optimization of a sample preparation for Gram-positive bacteria, especially *S. aureus.* Based on a sample preparation workflow for Gram-negative bacteria I adjusted the cell lyses to the much thicker peptidoglycan layer of Gram-positive bacteria. Removal of interfering components such as phospholipids and nucleic acids increased the phosphopeptide identification significantly compared to reported literature data. Using a transposon mutant of the only known Hanks-type Ser/Thr kinase (Stk1) and its respective phosphatase (Stp1) for the strain *S. aureus* USA300 I was able to identify new targets of Stk1 and Stp1. Further, based on the high number of identified Ser and Thr phosphorylations, even in the KO strain, we hypothesize that Stk1 is likely not the only Hanks-type Ser/Thr kinase.

In **chapter 3**, I show that the optimized sample preparation and enrichment strategy is not only suitable for the identification of phosphomonoesters, but also for the much more labile phosphoramidates, reporting the most comprehensive arginine phosphoproteome to date. Using synthetic pArg peptides I could validate the high abundance of pArg and the correct phosphosite localization. Further, by testing different fragmentation methods I show that HCD remains the gold standard in phosphoproteomics. Lastly, using the same Stk1 and Stp1 mutants as in **chapter 2** I show that the knockdown of Stp1 significantly increases the overall amount of arginine phosphorylation. However, by performing a dephosphorylation assay I could also show that Stp1 does not pose arginine phosphatase activity, but only indirectly influences the arginine phosphoproteome.

In **chapter 4**, I tried to show the universal applicability of the sample preparation workflow for Gram-negative as well as Gram-positive bacteria. Hereto, I used the Gram-negative bacteria *Pseudomonas aeruginosa* PA14 and PAO1, *Klebsiella pneumoniae* MGH as well as

Escherichia coli DH5a. Previously a strong correlation between the amount of nucleic acid contamination, indicated by the percentage of MS2 spectra containing the diagnostic ion of 330.06 *m/z*, and improved phosphopeptide identification has been shown. Here, I could show that this correlation is not as strong as previously thought and that this seems to be highly organism if not even strain specific. Further, I highlight that bacterial phosphoproteomics is still in need of more tailored protocols to reach a comprehensive picture of the bacterial phosphoproteoms.

Lastly, in **Chapter 5** I discuss the current state of bacterial phosphoproteomics. I highlight important advances as well as ongoing challenges in the field and provide my own perspective what this means for the field of bacterial phosphoproteomics and where we need to go to gain a comprehensive understanding of the bacterial phosphoproteom.

References

1. CRICK F. Central Dogma of Molecular Biology. *Nature*. 1970;227(5258):561-563. doi:10.1038/227561a0

2. Aebersold R, Agar JN, Amster IJ, Baker MS, Bertozzi CR, Boja ES, Costello CE, Cravatt BF, Fenselau C, Garcia BA, Ge Y, Gunawardena J, Hendrickson RC, Hergenrother PJ, Huber CG, et al. How many human proteoforms are there? *Nat Chem Biol*. 2018;14(3):206-214. doi:10.1038/nchembio.2576

3. Vogel C, Marcotte EM. Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nat Rev Genet*. 2012;13(4):227-232. doi:10.1038/nrg3185

4. Cohen P. The origins of protein phosphorylation. *Nat Cell Biol*. 2002;4(5):E127-E130. doi:10.1038/ncb0502-e127

5. Hunter T. Why nature chose phosphate to modify proteins. *Philos Trans R Soc B Biol Sci*. 2012;367(1602):2513-2516. doi:10.1098/rstb.2012.0013

6. Sharma K, D'Souza RCJ, Tyanova S, Schaab C, Wiśniewski J, Cox J, Mann M. Ultradeep Human Phosphoproteome Reveals a Distinct Regulatory Nature of Tyr and Ser/Thr-Based Signaling. *Cell Rep*. 2014;8(5):1583-1594. doi:10.1016/j.celrep.2014.07.036

7. Johnson LN, Lewis RJ. Structural basis for control by phosphorylation. *Chem Rev*. 2001;101(8):2209-2242. doi:10.1021/cr000225s

8. FISCHER EH, GRAVES DJ, CRITTENDEN ER, KREBS EG. Structure of the site phosphorylated in the phosphorylase b to a reaction. *J Biol Chem*. 1959;234(7):1698-1704. doi:10.1016/s0021-9258(18)69910-7

9. Cohen P. The role of protein phosphorylation in human health and disease: Delivered on June 30th 2001 at the FEBS meeting in Lisbon. *Eur J Biochem*. 2001;268(19):5001-5010. doi:10.1046/j.0014-2956.2001.02473.x

10. Altelaar AFMM, Munoz J, Heck AJRR. Next-generation proteomics: Towards an integrative view of proteome dynamics. *Nat Rev Genet*. 2012;14(1):35-48. doi:10.1038/nrg3356

11. Swoboda JG, Campbell J, Meredith TC, Walker S. Wall teichoic acid function, biosynthesis, and inhibition. *Chembiochem*. 2010;11(1):35-45. doi:10.1002/cbic.200900557

12. Potel CM, Lin M-H, Heck AJR, Lemeer S. Defeating Major Contaminants in Fe 3+ - Immobilized Metal Ion Affinity Chromatography (IMAC) Phosphopeptide Enrichment. *Mol Cell Proteomics*. 2018;17(5):1028-1034. doi:10.1074/mcp.TIR117.000518

13. Olsen J V., Ong SE, Mann M. Trypsin cleaves exclusively C-terminal to arginine and lysine residues. *Mol Cell Proteomics*. 2004;3(6):608-614. doi:10.1074/mcp.T400003-MCP200

14. Di Palma S, Hennrich ML, Heck AJR, Mohammed S. Recent advances in peptide separation by multidimensional liquid chromatography for proteome analysis. *J Proteomics*. 2012;75(13):3791-3813. doi:10.1016/j.jprot.2012.04.033

15. Pinkse MWH, Uitto PM, Hilhorst MJ, Ooms B, Heck AJR. 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(14):3935-3943. doi:10.1021/ac0498617

16. Larsen MR, Thingholm TE, Jensen ON, Roepstorff P, D. JJ. Highly Selective Enrichment of Phosphorylated Peptides from Peptide Mixtures Using Titanium Dioxide Microcolumns. *Mol Cell Proteomics*. 2005;4(7):873-886. doi:10.1074/mcp.T500007-MCP200

17. Villén J, Gygi SP. The SCX/IMAC enrichment approach for global phosphorylation analysis by mass spectrometry. *Nat Protoc*. 2008;3(10):1630-1638. doi:10.1038/nprot.2008.150

18. Posewitz MC, Tempst P. Immobilized gallium(III) affinity chromatography of phosphopeptides. *Anal Chem*. 1999;71(14):2883-2892. doi:10.1021/ac981409y

19. Pandey A, Andersen JS, Mann M. Use of mass spectrometry to study signaling pathways. *Sci STKE*. 2000;2000(37):1-13. doi:10.1126/stke.2000.37.pl1

20. Isobe T, Takayasu T, Takai N, Okuyama T. High-performance liquid chromatography of peptides on a macroreticular cation-exchange resin: Application to peptide mapping of Bence-Jones proteins. *Anal Biochem*. 1982;122(2):417-425. doi:10.1016/0003-2697(82)90304-9

21. SUI S, WANG J, LU Z, CAI Y, ZHANG Y, YU W, QIAN X. Phosphopeptide enrichment strategy based

on strong cation exchange chromatography. *Chinese J Chromatogr*. 2008;26(2):195-199. doi:10.1016/S1872-2059(08)60012-7

22. Dizdaroglu M. Weak Anion-Exchange Graphy of Peptides. *J Chromatogr*. 1985;334:49-69.

23. Han G, Ye M, Zhou H, Jiang X, Feng S, Jiang X, Tian R, Wan D, Zou H, Gu J. Large-scale phosphoproteome analysis of human liver tissue by enrichment and fractionation of phosphopeptides with strong anion exchange chromatography. *Proteomics*. 2008;8(7):1346-1361. doi:10.1002/pmic.200700884

24. Gilar M, Olivova P, Daly AE, Gebler JC. Orthogonality of Separation in Two-Dimensional Liquid Chromatography. *Anal Chem*. 2005;77(19):6426-6434. doi:10.1021/ac050923i

25. Gilar M, Olivova P, Daly AE, Gebler JC. Two-dimensional separation of peptides using RP-RP-HPLC system with different pH in first and second separation dimensions. *J Sep Sci*. 2005;28(14):1694-1703. doi:10.1002/jssc.200500116

26. Uversky VN. Use of Fast Protein Size-Exclusion Liquid Chromatography To Study the Unfolding of Proteins Which Denature through the Molten Globule. *Biochemistry*. 1993;32(48):13288-13298. doi:10.1021/bi00211a042

27. Hartmann WK, Saptharishi N, Yang XY, Mitra G, Soman G. Characterization and analysis of thermal denaturation of antibodies by size exclusion high-performance liquid chromatography with quadruple detection. *Anal Biochem*. 2004;325(2):227-239. doi:10.1016/j.ab.2003.10.031

28. Batth TS, Francavilla C, Olsen J V. Off-line high-pH reversed-phase fractionation for in-depth phosphoproteomics. *J Proteome Res*. 2014;13(12):6176-6186. doi:10.1021/pr500893m

29. Wagner K, Miliotis T, Marko-Varga G, Bischoff R, Unger KK. An automated on-line multidimensional HPLC system for protein and peptide mapping with integrated sample preparation. *Anal Chem*. 2002;74(4):809-820. doi:10.1021/ac010627f

30. Karas M, Hillenkamp F. Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. *Anal Chem*. 1988;60(20):2299-2301. doi:10.1021/ac00171a028

31. Meng CK, Mann M, Fenn JB. Of protons or proteins. *Zeitschrift für Phys D Atoms, Mol Clust*. 1988;10(2-3):361-368. doi:10.1007/BF01384871

32. Fenn J, Mann M, Meng C, Wong S, Whitehouse C. Electrospray ionization for mass spectrometry of large biomolecules. *Science (80-)*. 1989;246(4926):64-71. doi:10.1126/science.2675315

33. Wilm M. Principles of Electrospray Ionization. *Mol Cell Proteomics*. 2011;10(7):M111.009407. doi:10.1074/mcp.M111.009407

34. Konermann L, Ahadi E, Rodriguez AD, Vahidi S. Unraveling the mechanism of electrospray ionization. *Anal Chem*. 2013;85(1):2-9. doi:10.1021/ac302789c

35. Steen H, Mann M. The ABC's (and XYZ's) of peptide sequencing. *Nat Rev Mol Cell Biol*. 2004;5(9):699-711. doi:10.1038/nrm1468

36. Michalski A, Damoc E, Hauschild J-PPJ-PJ-P, Lange O, Wieghaus A, Makarov A, Nagaraj N, Cox J, Mann M, Horning S, Hauschild J-PPJ-PJ-P, Makarov A, Cox J, Lange O, Damoc E, et al. Mass Spectrometry-based Proteomics Using Q Exactive, a High-performance Benchtop Quadrupole Orbitrap Mass Spectrometer. *Mol Cell Proteomics*. 2011;10(9):M111.011015-M111.011015. doi:10.1074/mcp.M111.011015

37. Kelstrup CD, Bekker-Jensen DB, Arrey TN, Hogrebe A, Harder A, Olsen J V. Performance Evaluation of the Q Exactive HF-X for Shotgun Proteomics. *J Proteome Res*. 2018;17(1):727-738. doi:10.1021/acs.jproteome.7b00602

38. Glish GL, Burinsky DJ. Hybrid Mass Spectrometers for Tandem Mass Spectrometry. *J Am Soc Mass Spectrom*. 2008;19(2):161-172. doi:10.1016/j.jasms.2007.11.013

39. Hu Q, Noll RJ, Li H, Makarov A, Hardman M, Cooks RG. The Orbitrap: A new mass spectrometer. *J Mass Spectrom*. 2005;40(4):430-443. doi:10.1002/jms.856

40. Courant ED, Livingston MS, Snyder HS. The Strong-Focusing Synchroton—A New High Energy Accelerator. *Phys Rev*. 1952;88(5):1190-1196. doi:10.1103/PhysRev.88.1190

41. Paul W, Steinwedel H. Notizen: Ein neues Massenspektrometer ohne Magnetfeld. *Zeitschrift für Naturforsch A*. 1953;8(7):448-450. doi:10.1515/zna-1953-0710

42. Savaryn JP, Toby TK, Kelleher NL. A researcher's guide to mass spectrometry-based proteomics. *Proteomics*. 2016;16(18):2435-2443.

doi:10.1002/pmic.201600113

43. Campana JE. Elementary theory of the quadrupole mass filter. *Int J Mass Spectrom Ion Phys*. 1980;33(2):101-117. doi:10.1016/0020-7381(80)80042-8

44. Eliuk S, Makarov A. Evolution of Orbitrap Mass Spectrometry Instrumentation. *Annu Rev Anal Chem*. 2015;8(1):61-80. doi:10.1146/annurev-anchem-071114-040325

45. Makarov A. Electrostatic axially harmonic orbital trapping: A high-performance technique of mass analysis. *Anal Chem*. 2000;72(6):1156-1162. doi:10.1021/ac991131p

46. McLafferty FW, Bente PF, Kornfeld R, Tsai S-C, Howe I. Collisional activation spectra of organic ions. *J Mass Spectrom*. 1972;30(6):797-806. doi:10.1002/jms.1190300604

47. Shukla AK, Futrell JH. Tandem mass spectrometry: Dissociation of ions by collisional activation. *J Mass Spectrom*. 2000;35(9):1069-1090. doi:10.1002/1096-9888(200009)35:9<1069::AID-JMS54>3.0.CO;2-C

48. Paizs B, Suhai S, Palzs B, Suhal S. Fragmentation pathways of protonated peptides. *Mass Spectrom Rev*. 2005;24(4):508-548. doi:10.1002/mas.20024

49. Waldera-Lupa DM, Stefanski A, Meyer HE, Stühler K. The fate of b-ions in the two worlds of collision-induced dissociation. *Biochim Biophys Acta - Proteins Proteomics*. 2013;1834(12):2843-2848. doi:10.1016/j.bbapap.2013.08.007

50. Mitchell Wells J, McLuckey SA. Collision-induced dissociation (CID) of peptides and proteins. *Methods Enzymol*. 2005;402(1993):148-185. doi:10.1016/S0076-6879(05)02005-7

51. McLuckey SA. Principles of collisional activation in analytical mass spectrometry. *J Am Soc Mass Spectrom*. 1992;3(6):599-614. doi:10.1016/1044-0305(92)85001-Z

52. Dongré AR, Jones JL, Somogyi Á, Wysocki VH. Influence of peptide composition, gas-phase basicity, and chemical modification on fragmentation efficiency: Evidence for the mobile proton model. *J Am Chem Soc*. 1996;118(35):8365-8374. doi:10.1021/ja9542193

53. Boersema PJ, Mohammed S, Heck AJR. Phosphopeptide fragmentation and analysis by mass spectrometry. *J Mass Spectrom*. 2009;44(6):861-878. doi:10.1002/jms.1599

54. Roepstorff P, Fohlman J. Letter to the editors. *Biol Mass Spectrom*. 1984;11(11):601-601. doi:10.1002/bms.1200111109

55. Palumbo AM, Reid GE. Evaluation of gas-phase rearrangement and competing fragmentation reactions on protein phosphorylation site assignment using collision induced dissociation-MS/MS and MS3. *Anal Chem*. 2008;80(24):9735-9747. doi:10.1021/ac801768s

56. Kelstrup CD, Hekmat O, Francavilla C, Olsen J V. Pinpointing phosphorylation sites: Quantitative filtering and a novel site-specific x-ion fragment. *J Proteome Res*. 2011;10(7):2937-2948. doi:10.1021/pr200154t

57. Villén J, Beausoleil SA, Gygi SP. Evaluation of the utility of neutral-loss-dependent MS3 strategies in large-scale phosphorylation analysis. *Proteomics*. 2008;8(21):4444-4452. doi:10.1002/pmic.200800283

58. Olsen J V., Macek B, Lange O, Makarov A, Horning S, Mann M. Higher-energy C-trap dissociation for peptide modification analysis. *Nat Methods*. 2007;4(9):709-712. doi:10.1038/nmeth1060

59. Cui L, Yapici I, Borhan B, Reid GE. Quantification of competing H3PO4 Versus HPO 3 + H2O neutral losses from regioselective 18O-labeled phosphopeptides. *J Am Soc Mass Spectrom*. 2014;25(1):141-148. doi:10.1007/s13361-013-0744-4

60. Nagaraj N, D'Souza RCJ, Cox J, Olsen J V., Mann M. Feasibility of large-scale phosphoproteomics with higher energy collisional dissociation fragmentation. *J Proteome Res*. 2010;9(12):6786-6794. doi:10.1021/pr100637q

61. Michalski A, Neuhauser N, Cox J, Mann M. A systematic investigation into the nature of tryptic HCD spectra. *J Proteome Res*. 2012;11(11):5479-5491. doi:10.1021/pr3007045

62. Guo C, Guo XF, Zhao L, Chen DD, Wang J, Sun J. A Study on Immonium Ions and Immonium-Related Ions Depending on Different Collision Energies as Assessed by Q-TOF MS. *Russ J Bioorganic Chem*. 2018;44(4):408-415. doi:10.1134/S1068162018040088

63. HOFFMANN R, WACHS WO, BERGER RG, KALBITZER H-R, WAIDELICH D, BAYER E, WAGNER-REDEKER W, ZEPPEZAUER M. Chemical phosphorylation of the peptides GGXA (X = S, T, Y): an evaluation of different chemical approaches. *Int J Pept Protein Res*.

2009;45(1):26-34. doi:10.1111/j.1399-3011.1995.tb01564.x

64. Potel CM, Lin MH, Heck AJR, Lemeer S. Widespread bacterial protein histidine phosphorylation revealed by mass spectrometrybased proteomics. *Nat Methods*. 2018;15(3):187-190. doi:10.1038/nmeth.4580

65. Steen H, Küster B, Fernandez M, Pandey A, Mann M. Detection of Tyrosine Phosphorylated Peptides by Precursor Ion Scanning Quadrupole TOF Mass Spectrometry in Positive Ion Mode. *Anal Chem*. 2001;73(7):1440-1448. doi:10.1021/ac001318c

66. Potel CM, Lin MH, Prust N, Van Den Toorn HWP, Heck AJR, Lemeer S. Gaining Confidence in the Elusive Histidine Phosphoproteome. *Anal Chem*. 2019;91(9):5542-5547. doi:10.1021/acs.analchem.9b00734

67. Zubarev RA, Kelleher NL, McLafferty FW. Electron Capture Dissociation of Multiply Charged Protein Cations. A Nonergodic Process. *J Am Chem Soc*. 1998;120(13):3265-3266. doi:10.1021/ja973478k

68. Syka JEP, Coon JJ, Schroeder MJ, Shabanowitz J, Hunt DF. Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proc Natl Acad Sci*. 2004;101(26):9528-9533. doi:10.1073/pnas.0402700101

69. Zhurov KO, Fornelli L, Wodrich MD, Laskay ÜA, Tsybin YO. Principles of electron capture and transfer dissociation mass spectrometry applied to peptide and protein structure analysis. *Chem Soc Rev*. 2013;42(12):5014-5030. doi:10.1039/c3cs35477f

70. Wysocki VH, Resing KA, Zhang Q, Cheng G. Mass spectrometry of peptides and proteins. *Methods*. 2005;35(3 SPEC.ISS.):211-222. doi:10.1016/j.ymeth.2004.08.013

71. Zubarev RA, Haselmann KF, Budnik B, Kjeldsen F, Jensen F. Towards an understanding of the mechanism of electron-capture dissociation: A historical perspective and modern ideas. *Eur J Mass Spectrom*. 2002;8(5):337-349. doi:10.1255/ejms.517

72. Sawicka A, Skurski P, Hudgins RR, Simons J. Model calculations relevant to disulfide bond cleavage via electron capture influenced by positively charged groups. *J Phys Chem B*. 2003;107(48):13505-13511. doi:10.1021/jp035675d

73. Syrstad EA, Turecček F, Tureček F. Toward a general mechanism of electron capture dissociation. *J Am Soc Mass Spectrom*. 2005;16(2):208-224. doi:10.1016/j.jasms.2004.11.001

74. Qi Y, Volmer DA. Electron-based fragmentation methods in mass spectrometry: An overview. *Mass Spectrom Rev*. 2017;36(1):4-15. doi:10.1002/mas.21482

75. Fort KL, Cramer CN, Voinov VG, Vasil'Ev Y V., Lopez NI, Beckman JS, Heck AJR. Exploring ECD on a Benchtop Q Exactive Orbitrap Mass Spectrometer. *J Proteome Res*. 2018;17(2):926-933. doi:10.1021/acs.jproteome.7b00622

76. Zubarev RA. Electron-capture dissociation tandem mass spectrometry. *Curr Opin Biotechnol*. 2004;15(1):12-16. doi:10.1016/j.copbio.2003.12.002

77. Good DM, Wirtala M, McAlister GC, Coon JJ. Performance Characteristics of Electron Transfer Dissociation Mass Spectrometry. *Mol Cell Proteomics*. 2007;6(11):1942-1951. doi:10.1074/mcp.M700073-MCP200

78. Frese CK, Altelaar AFMM, van den Toorn H, Nolting D, Griep-Raming J, Heck AJRR, Mohammed S. Toward Full Peptide Sequence Coverage by Dual Fragmentation Combining Electron-Transfer and Higher-Energy Collision Dissociation Tandem Mass Spectrometry. *Anal Chem*. 2012;84(22):9668-9673. doi:10.1021/ac3025366

79. Frese CK, Zhou H, Taus T, Altelaar AFM, Mechtler K, Heck AJR, Mohammed S. Unambiguous phosphosite localization using electron-transfer/higher-energy collision dissociation (EThcD). *J Proteome Res*. 2013;12(3):1520-1525. doi:10.1021/pr301130k

80. Penkert M, Hauser A, Harmel R, Fiedler D, Hackenberger CPR, Krause E. Electron Transfer/Higher Energy Collisional Dissociation of Doubly Charged Peptide Ions: Identification of Labile Protein Phosphorylations. *J Am Soc Mass Spectrom*. Published online 2019. doi:10.1007/s13361-019-02240-4

81. Riley NM, Westphall MS, Hebert AS, Coon JJ. Implementation of Activated Ion Electron Transfer Dissociation on a Quadrupole-Orbitrap-Linear Ion Trap Hybrid Mass Spectrometer. *Anal Chem*. 2017;89(12):6358-6366. doi:10.1021/acs.analchem.7b00213

82. Marx H, Lemeer S, Schliep JE, Matheron L, Mohammed S, Cox J, Mann M, Heck AJR, Kuster

B. A large synthetic peptide and phosphopeptide reference library for mass spectrometry–based proteomics. *Nat Biotechnol*. 2013;31(6):557-564. doi:10.1038/nbt.2585

83. Medzihradszky KF, Chalkley RJ. Lessons in de novo peptide sequencing by tandem mass spectrometry. *Mass Spectrom Rev*. 2015;34(1):43-63. doi:10.1002/mas.21406

84. Muth T, Hartkopf F, Vaudel M, Renard BY. A Potential Golden Age to Come—Current Tools, Recent Use Cases, and Future Avenues for De Novo Sequencing in Proteomics. *Proteomics*. 2018;18(18):1-14. doi:10.1002/pmic.201700150

85. Perkins DN, Pappin DJC, Creasy DM, Cottrell JS. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis*. 1999;20(18):3551-3567. doi:10.1002/(SICI)1522-2683(19991201)20:18<3551::AID-ELPS3551>3.0.CO;2-2

86. Eng JK, McCormack AL, Yates JR. 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. doi:10.1016/1044-0305(94)80016-2

87. Cox J, Neuhauser N, Michalski A, Scheltema RA, Olsen J V., Mann M. Andromeda: A peptide search engine integrated into the MaxQuant environment. *J Proteome Res*. 2011;10(4):1794-1805. doi:10.1021/pr101065j

88. Chalkley RJ, Clauser KR. Modification site localization scoring: Strategies and performance. *Mol Cell Proteomics*. 2012;11(5):3-14. doi:10.1074/mcp.R111.015305

89. Elias JE, Gygi SP. Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nat Methods*. 2007;4(3):207-214. doi:10.1038/nmeth1019

90. Jeong K, Kim S, Bandeira N. False discovery rates in spectral identification. *BMC Bioinformatics*. 2012;13 Suppl 1(Suppl 16). doi:10.1186/1471-2105-13-S16-S2

91. 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(12):1367-1372. doi:10.1038/nbt.1511

92. Nesvizhskii AI, Aebersold R. Interpretation of shotgun proteomic data: The protein inference problem. *Mol Cell Proteomics*. 2005;4(10):1419-1440. doi:10.1074/mcp.R500012-MCP200

93. Olsen J V., Blagoev B, Gnad F, Macek B, Kumar C, Mortensen P, Mann M. Global, In Vivo, and Site-Specific Phosphorylation Dynamics in Signaling Networks. *Cell*. 2006;127(3):635-648. doi:10.1016/j.cell.2006.09.026

94. Potel CM, Lemeer S, Heck AJR. Phosphopeptide Fragmentation and Site Localization by Mass Spectrometry: An Update. *Anal Chem*. 2019;91(1):126-141. doi:10.1021/acs.analchem.8b04746

95. Junker S, Maaβ S, Otto A, Michalik S, Morgenroth F, Gerth U, Hecker M, Becher D. Spectral Library Based Analysis of Arginine Phosphorylations in *Staphylococcus aureus*. *Mol Cell Proteomics*. 2018;17(2):335-348. doi:10.1074/mcp.RA117.000378

96. Lam H. Building and searching tandem mass spectral libraries for peptide identification. *Mol Cell Proteomics*. 2011;10(12):R111.008565. doi:10.1074/mcp.R111.008565

97. Hu Y, Lam H. Expanding Tandem Mass Spectral Libraries of Phosphorylated Peptides: Advances and Applications. *J Proteome Res*. 2013;12(12):5971-5977. doi:10.1021/pr4007443

98. Ong S-E, Blagoev B, Kratchmarova I, Kristensen DB, Steen H, Pandey A, Mann M. Stable Isotope Labeling by Amino Acids in Cell Culture, SILAC, as a Simple and Accurate Approach to Expression Proteomics. *Mol Cell Proteomics*. 2002;1(5):376-386. doi:10.1074/mcp.M200025-MCP200

99. Hsu J-LL, Huang S-YY, Chow N-HH, Chen S-HH. Stable-Isotope Dimethyl Labeling for Quantitative Proteomics. *Anal Chem*. 2003;75(24):6843-6852. doi:10.1021/ac0348625

100. Thompson A, Schäfer J, Kuhn K, Kienle S, Schwarz J, Schmidt G, Neumann T, Hamon C. Tandem mass tags: A novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. *Anal Chem*. 2003;75(8):1895-1904. doi:10.1021/ac0262560

101. Riley NM, Coon JJ. Phosphoproteomics in the Age of Rapid and Deep Proteome Profiling. *Anal Chem*. 2016;88(1):74-94. doi:10.1021/acs.analchem.5b04123

102. Bantscheff M, Schirle M, Sweetman G, Rick J,

Kuster B. Quantitative mass spectrometry in proteomics: A critical review. *Anal Bioanal Chem*. 2007;389(4):1017-1031. doi:10.1007/s00216-007-1486-6

103. Bantscheff M, Lemeer S, Savitski MM, Kuster B. Quantitative mass spectrometry in proteomics: Critical review update from 2007 to the present. *Anal Bioanal Chem*. 2012;404(4):939-965. doi:10.1007/s00216-012-6203-4

104. Cox J, Hein MY, Luber CA, Paron I, Nagaraj N, Mann M. Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. *Mol Cell Proteomics*. 2014;13(9):2513-2526. doi:10.1074/mcp.M113.031591

105. Mizuno T. Compilation of All Genes Encoding Two-component Phosphotransfer Signal Transducers in the Genome of Escherichia coli. *DNA Res*. 1997;4(2):161-168. doi:10.1093/dnares/4.2.161

106. Laub MT, Goulian M. Specificity in Two-Component Signal Transduction Pathways. Published online 2007:121-145. doi:10.1146/annurev.genet.41.042007.170548

107. Tiwari S, Jamal SB, Hassan SS, Carvalho PVSD, Almeida S, Barh D, Ghosh P, Silva A, Castro TLP, Azevedo V. Two-Component Signal Transduction Systems of Pathogenic Bacteria As Targets for Antimicrobial Therapy: An Overview. *Front Microbiol*. 2017;8(October):1-7. doi:10.3389/fmicb.2017.01878

108. Jung K, Fried L, Behr S, Heermann R. Histidine kinases and response regulators in networks. *Curr Opin Microbiol*. 2012;15(2):118-124. doi:10.1016/j.mib.2011.11.009

109. Hauser A, Penkert M, Hackenberger CPR. Chemical Approaches to Investigate Labile Peptide and Protein Phosphorylation. *Acc Chem Res*. 2017;50(8):1883-1893. doi:10.1021/acs.accounts.7b00170

110. Shizuta Y, Beavo JA, Bechtel PJ, Hofmann F, Krebs EG. Reversibility of adenosine 3':5' monophosphate dependent protein kinase reactions. *J Biol Chem*. 1975;250(17):6891-6896. doi:10.1016/s0021-9258(19)41016-8

111. Besant P, Attwood P, Piggott M. Focus on Phosphoarginine and Phospholysine. *Curr Protein Pept Sci*. 2009;10(6):536-550. doi:10.2174/138920309789630598

112. Cieśla J, Frączyk T, Rode W. Phosphorylation of basic amino acid residues in proteins. *Acta Biochim Pol*. 2011;58(2). doi:201114789 [pii]

113. Villanueva M, García B, Valle J, Rapún B, Ruiz De Los Mozos I, Solano C, Martí M, Penadés JR, Toledo-Arana A, Lasa I. Sensory deprivation in Staphylococcus aureus. *Nat Commun*. 2018;9(1):1-12. doi:10.1038/s41467-018-02949-y

114. Dutta R, Qin L, Inouye M. Histidine kinases: Diversity of domain organization. *Mol Microbiol*. 1999;34(4):633-640. doi:10.1046/j.1365-2958.1999.01646.x

115. Fridman M, Williams GD, Muzamal U, Hunter H, Siu KWMWMMWM, Golemi-Kotra D. Two unique phosphorylation-driven signaling pathways crosstalk in staphylococcus aureus to modulate the cell-wall charge: Stk1/Stp1 meets GraSR. *Biochemistry*. 2013;52(45):7975-7986. doi:10.1021/bi401177n

116. Ninfa AJ. Use of two-component signal transduction systems in the construction of synthetic genetic networks. *Curr Opin Microbiol*. 2010;13(2):240-245. doi:10.1016/j.mib.2010.01.003

117. Junecko JM. Transcribing virulence in Staphylococcus aureus. *World J Clin Infect Dis*. 2012;2(4):63. doi:10.5495/wjcid.v2.i4.63

118. Rogasch K, Ruhmling V, Pane-Farre J, Hoper D, Weinberg C, Fuchs S, Schmudde M, Broker BM, Wolz C, Hecker M, Engelmann S. Influence of the Two-Component System SaeRS on Global Gene Expression in Two Different Staphylococcus aureus Strains. *J Bacteriol*. 2006;188(22):7742-7758. doi:10.1128/JB.00555-06

119. Burgui S, Gil C, Solano C, Lasa I, Valle J. A systematic evaluation of the two-component systems network reveals that ArlRS is a key regulator of catheter colonization by Staphylococcus aureus. *Front Microbiol*. 2018;9(MAR):1-11. doi:10.3389/fmicb.2018.00342

120. Falord M, Mäder U, Hiron A, Dbarbouillé M, Msadek T. Investigation of the Staphylococcus aureus GraSR regulon reveals novel links to virulence, stress response and cell wall signal transduction pathways. *PLoS One*. 2011;6(7). doi:10.1371/journal.pone.0021323

121. Oshima T, Aiba H, Masuda Y, Kanaya S, Sugiura M, Wanner BL, Mori H, Mizuno T. Transcriptome analysis of all two-component regulatory system mutants of Escherichia coli K-12. *Mol Microbiol*. 2002;46(1):281-291. doi:10.1046/j.1365-2958.2002.03170.x

122. Cheung AL. Global regulation of virulence determinants in Staphylococcus Aureus by the SarA protein family. *Front Biosci*. 2002;7(4):A882. doi:10.2741/A882

123. Munoz-Dorado J, Inouye S, Inouye M. Eukaryotic-like protein serine/threonine kinases inMyxococcus xanthus, a developmental bacterium exhibiting social behavior. *J Cell Biochem*. 1993;51(1):29-33. doi:10.1002/jcb.240510107

124. Pereira SFF, Goss L, Dworkin J. Eukaryote-Like Serine/Threonine Kinases and Phosphatases in Bacteria. *Microbiol Mol Biol Rev*. 2011;75(1):192-212. doi:10.1128/MMBR.00042-10

125. Hanks SK, Hunter T. The eukaryotic protein kinase superfamily: kinase (catalytic) domain structure and classification 1 . *FASEB J*. 1995;9(8):576-596. doi:10.1096/fasebj.9.8.7768349

126. Kornev AP, Taylor SS. Defining the conserved internal architecture of a protein kinase. *Biochim Biophys Acta - Proteins Proteomics*. 2010;1804(3):440-444. doi:10.1016/j.bbapap.2009.10.017

127. Zucchini L, Mercy C, Garcia PS, Cluzel C, Gueguen-Chaignon V, Galisson F, Freton C, Guiral S, Brochier-Armanet C, Gouet P, Grangeasse C. PASTA repeats of the protein kinase StkP interconnect cell constriction and separation of Streptococcus pneumoniae. *Nat Microbiol*. 2018;3(2):197-209. doi:10.1038/s41564-017-0069-3

128. Wright DP, Ulijasz AT. Regulation of transcription by eukaryotic-like serine-threonine kinases and phosphatases in gram-positive bacterial pathogens. *Virulence*. 2014;5(8):863-885. doi:10.4161/21505594.2014.983404

129. Mijakovic I, Petranovic D, Bottini N, Deutscher J, Jensen PR. Protein-tyrosine phosphorylation in Bacillus subtilis. *J Mol Microbiol Biotechnol*. 2006;9(3-4):189-197. doi:10.1159/000089647

130. Mijakovic I, Grangeasse C, Turgay K. Exploring the diversity of protein modifications: Special bacterial phosphorylation systems. *FEMS Microbiol Rev*. 2016;40(3):398-417. doi:10.1093/femsre/fuw003

131. Rausch M, Deisinger JP, Ulm H, Müller A, Li W, Hardt P, Wang X, Li X, Sylvester M, Engeser M, Vollmer W, Müller CE, Sahl HG, Lee JC, Schneider T. Coordination of capsule assembly and cell wall biosynthesis in Staphylococcus aureus. *Nat Commun*. 2019;10(1). doi:10.1038/s41467-019-09356-x

132. Grangeasse C, Doublet P, Vincent C, Vaganay E, Riberty M, Duclos B, Cozzone AJ. Functional characterization of the low-molecular-mass phosphotyrosine-protein phosphatase of Acinetobacter johnsonii. *J Mol Biol*. 1998;278(2):339-347. doi:10.1006/jmbi.1998.1650

133. Poirier V, Bach H, Av-Gay Y. Mycobacterium tuberculosis promotes anti-apoptotic activity of the macrophage by PtpA protein-dependent dephosphorylation of host GSK3α. *J Biol Chem*. 2014;289(42):29376-29385. doi:10.1074/jbc.M114.582502

134. Fuhrmann J, Schmidt A, Spiess S, Lehner A, Turgay K, Mechtler K, Charpentier E, Clausen T. McsB Is a Protein Arginine Kinase That Phosphorylates and Inhibits the Heat-Shock Regulator CtsR. *Science (80-)*. 2009;324(5932):1323-1327. doi:10.1126/science.1170088

135. Suskiewicz MJ, Hajdusits B, Beveridge R, Heuck A, Vu LD, Kurzbauer R, Hauer K, Thoeny V, Rumpel K, Mechtler K, Meinhart A, Clausen T. Structure of McsB, a protein kinase for regulated arginine phosphorylation. *Nat Chem Biol*. 2019;15(5):510-518. doi:10.1038/s41589-019-0265-y

136. Elsholz AKW, Turgay K, Michalik S, Hessling B, Gronau K, Oertel D, Mader U, Bernhardt J, Becher D, Hecker M, Gerth U. Global impact of protein arginine phosphorylation on the physiology of Bacillus subtilis. *Proc Natl Acad Sci*. 2012;109(19):7451-7456. doi:10.1073/pnas.1117483109

137. Kirstein J, Zühlke D, Gerth U, Turgay K, Hecker M. A tyrosine kinase and its activator control the activity of the CtsR heat shock repressor in B. subtilis. *EMBO J*. 2005;24(19):3435-3445. doi:10.1038/sj.emboj.7600780

138. Fuhrmann J, Mierzwa B, Trentini DB, Spiess S, Lehner A, Charpentier E, Clausen T. Structural Basis for Recognizing Phosphoarginine and Evolving Residue-Specific Protein Phosphatases in Gram-Positive Bacteria. *Cell Rep*. 2013;3(6):1832-1839. doi:10.1016/j.celrep.2013.05.023

139. Trentini DB, Fuhrmann J, Mechtler K, Clausen T. Chasing phosphoarginine proteins: Development of a selective enrichment method using a phosphatase trap. *Mol Cell Proteomics*.

2014;13(8):1953-1964. doi:10.1074/mcp.O113.035790

140. Junker S, Maaß S, Otto A, Hecker M, Becher D. Toward the Quantitative Characterization of Arginine Phosphorylations in Staphylococcus aureus. *J Proteome Res*. 2018;18(1):acs.jproteome.8b00579. doi:10.1021/acs.jproteome.8b00579

141. Trentini DB, Suskiewicz MJ, Heuck A, Kurzbauer R, Deszcz L, Mechtler K, Clausen T. Arginine phosphorylation marks proteins for degradation by a Clp protease. *Nature*. 2016;539(7627):48-53. doi:10.1038/nature20122

142. Pagano GJ, Arsenault RJ. Advances, challenges and tools in characterizing bacterial serine, threonine and tyrosine kinases and phosphorylation target sites. *Expert Rev Proteomics*. 2019;16(5):431-441. doi:10.1080/14789450.2019.1601015

143. Olsen J V., Mann M. Status of large-scale analysis of posttranslational modifications by mass spectrometry. *Mol Cell Proteomics*. 2013;12(12):3444-3452. doi:10.1074/mcp.O113.034181

144. Lemeer S, Heck AJ. The phosphoproteomics data explosion. *Curr Opin Chem Biol*. 2009;13(4):414-420. doi:10.1016/j.cbpa.2009.06.022

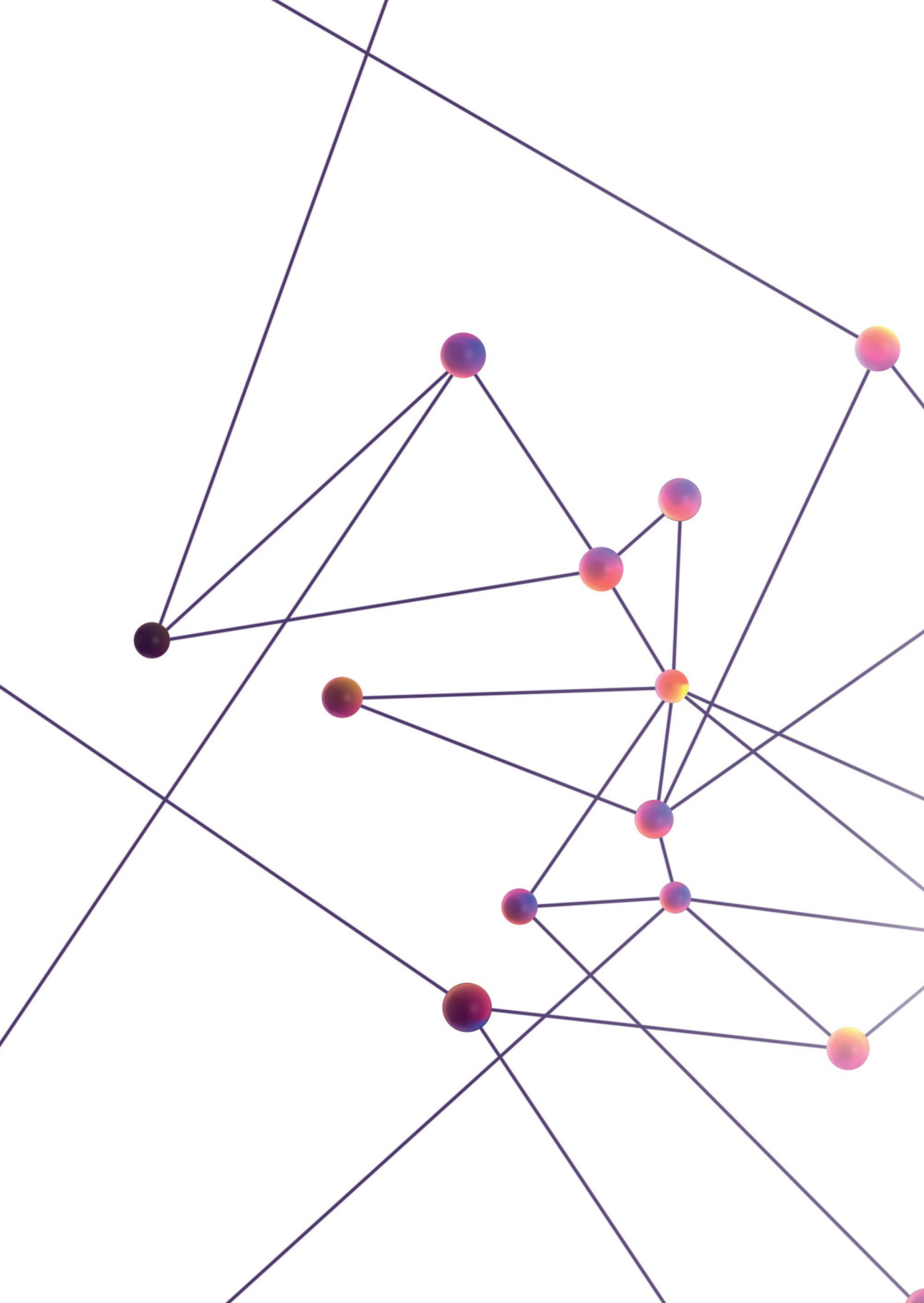
145. Semanjski M, Macek B. Shotgun proteomics of bacterial pathogens: Advances, challenges and clinical implications. *Expert Rev Proteomics*. 2016;13(2):139-156. doi:10.1586/14789450.2016.1132168

146. Ruprecht B, Koch H, Medard G, Mundt M, Kuster B, Lemeer S. Comprehensive and Reproducible Phosphopeptide Enrichment Using Iron Immobilized Metal Ion Affinity Chromatography (Fe-IMAC) Columns. *Mol Cell Proteomics*. 2015;14(1):205-215. doi:10.1074/mcp.M114.043109

147. Bonne Køhler J, Jers C, Senissar M, Shi L, Derouiche A, Mijakovic I. Importance of protein Ser/Thr/Tyr phosphorylation for bacterial pathogenesis. *FEBS Lett*. 2020;594(15):2339-2369. doi:10.1002/1873-3468.13797

148. Cao XJ, Dai J, Xu H, Nie S, Chang X, Hu BY, Sheng QH, Wang LS, Ning Z Bin, Li YX, Guo XK, Zhao GP, Zeng R. High-coverage proteome analysis reveals the first insight of protein modification systems in the pathogenic spirochete Leptospira interrogans. *Cell Res*. 2010;20(2):197-210. doi:10.1038/cr.2009.127

149. Lin M-H, Potel CM, Tehrani KHME, Heck AJR, Martin NI, Lemeer S. A New Tool to Reveal Bacterial Signaling Mechanisms in Antibiotic Treatment and Resistance. *Mol Cell Proteomics*. 2018;17(12):2496-2507. doi:10.1074/mcp.ra118.000880



Chapter 2

In-depth characterization of the *Staphylococcus aureus* phosphoproteome reveals new targets of Stk1

Nadine Prust[1,2], Saar van der Laarse[1,2], Henk van den Toorn[1,2], Nina M. van Sorge[3,4], Simone Lemeer[1,2#]

[1] Biomolecular Mass Spectrometry and Proteomics, Bijvoet Center for Biomolecular Research and Utrecht Institute for Pharmaceutical Sciences, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands
[2] Netherlands Proteomics Center, Padualaan 8, 3584 CH Utrecht, The Netherlands
[3] Medical Microbiology, University Medical Center Utrecht, Utrecht University, Heidelberglaan 100, 3584 CX, Utrecht, The Netherlands
[4] Department of Medical Microbiology and Infection Prevention and Netherlands Reference Laboratory for Bacterial Meningitis, Amsterdam University Medical Center, University of Amsterdam, Meibergdreef 9, 1105 AZ, Amsterdam, The Netherlands

Published in Molecular & Cellular Proteomics (2021), mcp.RA120.002232

Abbreviations

TCS	Two-component system
IMAC	Immobilized metal ion affinity chromatography
MOAC	Metal oxide affinity chromatography
PTM	Post-translational modification
eSTK	eukaryotic-type Serine/Threonine kinase
RR	Response regulator
MRSA	Methicillin resistant *Staphylococcus aureus*
LC	Liquid chromatography
HCD	Higher energy collision induced dissociation
FDR	False discovery rate
TCEP	tris(2-carboxyethyl)phosphine
CAA	2-chloroacetamide
SDC	Sodium deoxycholate

Abstract

Staphylococcus aureus is a major cause of infections worldwide and infection results in a variety of diseases. As of no surprise, protein phosphorylation is an important game player in signaling cascades and has been shown to be involved in *S. aureus* virulence. Albeit long neglected, eukaryotic-type serine/threonine kinases in *S. aureus* have been implicated in this complex signaling cascades. Due to the sub-stoichiometric nature of protein phosphorylation and a lack of suitable analysis tools, the knowledge of these cascades is however, to date, still limited.

Here, were apply an optimized protocol for efficient phosphopeptide enrichment via Fe^{3+}-IMAC followed by LC-MS/MS to get a better understanding of the impact of protein phosphorylation on the complex signaling networks involved in pathogenicity. By profiling a serine/threonine kinase and phosphatase mutant from a methicillin-resistant *S. aureus* mutant library, we generated the most comprehensive phosphoproteome dataset of *S. aureus* to date, aiding a better understanding of signaling in bacteria. With the identification of 3,800 class I p-sites we were able to increase the number of identifications by more than 21 times compared to recent literature. In addition, we were able to identify 74 downstream targets of the only reported eukaryotic-type Ser/Thr kinase of the *S. aureus* strain USA300, Stk1. This work allowed an extensive analysis of the bacterial phosphoproteome and indicates that Ser/Thr kinase signaling is far more abundant than previously anticipated in *S. aureus*.

Introduction

Protein phosphorylation is a major regulator of cellular processes in all kinds of organisms. Eukaryotic protein phosphorylation commonly occurs on serine (S), threonine (T) and tyrosine (Y) residues and it was long assumed that similar post-translational modifications (PTMs) only play a minor role in prokaryotes. Instead, it was proposed that protein histidine (H) phosphorylation, as for example used in the so-called two-component systems (TCSs), is the main regulatory PTM[1,2]. The discovery of eukaryotic-type serine/threonine kinases (eSTKs) and the identification of numerous pS/pT peptides changed this perception. The occurrence of STY phosphorylation has now been identified in both Gram-negative and Gram-positive bacteria including important human pathogens[3,2].

The sub-stoichiometric nature of protein phosphorylation not only necessitates a highly efficient sample preparation but also a highly efficient enrichment method before mass spectrometric analysis. To date, a lack of efficient methods to enrich for the sub-stoichiometric phosphopeptides from Gram-positive bacteria confined the information about signaling pathways in these bacteria. Most enrichment techniques exploit the affinity of negatively-charged phosphopeptides towards metal oxides such as TiO_2, which is mostly used for metal oxide affinity chromatography (MOAC)[4,5], or metal ions such as Ti^{4+} and Fe^{3+} that are used for immobilized metal affinity chromatography (IMAC)[6,7]. Recent developments in enrichment strategies as well as fragmentation, mass spectrometric detection and phosphorylation site localization assessment allow the identification of thousands of phosphorylation sites[8–10]. However, we recently showed that contamination by DNA/RNA and phospholipids in standard sample preparation protocols hamper the Fe^{3+}-IMAC enrichment[11]. Consequently, removing those contaminants drastically improved the identification for the Gram-negative bacterium *E. coli*[11]. This improvement has encouraged us to also further optimize the sample preparation for Gram-positive bacteria, which require a more stringent cell lysis due to the presence of the thick peptidoglycan layer.

This optimization of phosphopeptide identification from Gram-positive bacteria could subsequently help to understand signaling pathways of important human pathogens such as *Staphylococcus aureus* (*S. aureus*). *S. aureus* is a great threat to public health since it is a major cause of infections worldwide[12–14]. Due to a variety of virulence factors and its exceptional versatility, this pathogen is able to cause a broad spectrum of infections ranging from mild skin infections to life-threatening infections such as toxic shock syndrome and sepsis[13]. From a clinical point of view, the evolution of antibiotic resistant strains (e.g. methicillin-resistant *S. aureus* (MRSA)) is complicating both prevention and treatment of *S. aureus* infections, rendering *S. aureus* a high priority pathogen according to the World Health Organization[15]. In the search for new treatment strategies, major efforts have been devoted to elucidating the virulence and adaptation mechanisms of this pathogen[16–18].

Nevertheless, we are still far from a comprehensive understanding of the underlying mechanisms involved in the intracellular signaling pathways engaged in virulence.

S. aureus contains a large variety of virulence factors, which allows bacterial adaptation and survival in a temporal[19] and host tissue-specific manner[20]. Gene expression is regulated through sophisticated regulatory mechanisms, including protein phosphorylation via TCSs[16]. TCSs sense specific environmental changes such as pH or nutrient concentrations and translate these extracellular stimuli into intracellular responses by affecting gene transcription, amongst which are virulence factors[21]. *S. aureus* encompasses 16 to 18 known TCSs that are known to be involved in virulence gene regulations, cell wall metabolism, nutrient sensing and response to antimicrobial agents[22,23]. In addition to these TCSs, *S. aureus* USA300 also expresses two serine kinases (HprK and RsbW) and one eSTK [24–26], most importantly Serine/Threonine kinase 1 (Stk1, also known as PknB, protein kinase B). Stk1 and its corresponding phosphatase, Stp1, have been shown to fine-tune the response of certain TCS response regulators (RR) by adding/removing additional phosphate on serine/threonine residues[22]. Such regulation is clinically relevant since phosphorylation of the RRs VraR and GraR by Stk1 has been linked to vancomycin resistance[24]. Here, we optimized the sample preparation workflow for phosphoproteomic studies of Gram-positive bacteria. This optimized workflow was first tested on the non-pathogenic model organism *B. subtilis* and subsequently applied to *S. aureus*. Moreover, we performed a label-free quantitative (LFQ) phosphoproteomics study and identified 74 potential new targets of Stk1 and Stp1, which will help elucidating the role Stk1 and Stp1 in signal transduction and virulence. Furthermore, our data shows that more, unknown, Ser/Thr kinases, are involved in signaling in *S. aureus.*

Material and methods

Bacterial culture

B. subtilis 168 was grown overnight in 50 ml Luria broth (LB) at 37°C with agitation in n=3 biological replicates. Transposon mutants NE98 (disruption in unrelated surface protein encoding gene *sdrE*), NE217 (disruption in protein coding gene *pknB*) and NE1919 (disruption in gene SAUSA300_1112 encoding Stp1), all containing an erythromycin resistance marker, of the *S. aureus* USA300 JE2 strain were obtained from the Nebraska Transposon Mutant Library (NTML)[27] . Mutants were grown in n=4 biological replicates, overnight in 25 ml Todd Hewitt broth (THB) supplemented with 5 ug/ml erythromycin at 37°C with agitation. Bacteria were harvested by centrifugation (15 min, 3,200 rpm at 4°C) and the supernatant was subsequently removed.

Optimized cell lysis

Bacterial cell lysis was performed as described in Potel et *al*. (2018) with optimization for Gram-positive bacteria[11]. One volume bacteria pellet was resuspended in five volume of lysis buffer (100 mM Tris-HCl pH 8.5, 7 M Urea, 5 mM tris(2-carboxyethyl)phosphine (TCEP), 30 mM 2-Chloroacetamide (CAA), 10 U/ml DNase I, 1 mM magnesium chloride (Sigma-Aldrich, Steinheim, Germany), 1% (v/v) benzonase (Merck Millipore, Darmstadt, Germany), 1 mM sodium orthovanadate, phosphoSTOP phosphatases inhibitors (Roche) and complete mini EDTA free protease inhibitors). The lysis was performed by bead beating for 17.5 min (1.5 min on, 2 min off) at 2,850 rpm (Disruptor Genie, Scientific industries) in case of *B. subtilis* and 3,200 rpm (Mini-Beadbeater-24, Bio Spec Products Inc.) for *S. aureus*. Subsequently, the beads were pelleted by centrifugation (2 min at 3000 rpm) and 1% (v/v) Triton X-100 in case of *B. subtilis* and 1 % (v/v) Triton X-100 plus 1% (v/v) sodium deoxycholate (SDC, final concentration) in case of *S. aureus* were added to the bacteria lysate. Complete lysis was reached by sonication for 45 min (20 s ON, 40 s off) using a Bioruptor Plus. Cell debris was removed by ultracentrifugation (45,000 rpm for 1 hr at 4°C). Protein concentration of the supernatant was determined via a Bicinchonic Acid (BCA) assay. To decrease the SDC concentration to <0.4 %, the supernatant was diluted 2.5 times with dilution buffer (100 mM Tris-HCl pH 8.5, 7 M Urea, 5 mM TCEP, 30 mM CAA, 1 mM magnesium chloride (Sigma-Aldrich, Steinheim, Germany), 1 mM sodium orthovanadate, phosphoSTOP phosphatases inhibitors (Roche) and complete mini EDTA free protease inhibitors). 1 % (v/v) benzonase was added to the supernatant mixture and incubated for 2 h at room temperature. Subsequently, methanol/chloroform precipitation was performed as described earlier[11]. The precipitate was then resuspended in digestion buffer (100 mM Tris-HCL pH 8.5, 30 M CAA, 1% (v/v) SDC (Sigma-Aldrich) and 5 mM TCEP). Protein digestion was performed overnight at room temperature using a mix of trypsin and Lys-C in a ratio of 1:25 and 1:100 (w/w), respectively. Protein digests were acidified to pH 3.5 using 10% formic acid (Sigma-Aldrich) and precipitated SDC was removed by centrifugation (1,400 rpm, 5 min). The supernatant was loaded onto C18 Sep-Pak (3cc) resin columns (Waters) for desalting. The loaded samples were washed twice with 0.1 % (v/v) formic acid and bound peptides were eluted with 600 ul 40 % acetonitrile and 0.06 % formic acid. Eluted peptides were split into 2 mg fractions and samples for full proteome analysis were frozen in liquid nitrogen and freeze dried.

Phosphopeptide enrichment

Fe^{3+}-IMAC enrichments were performed as previously described[11]. In short, 2 mg lyophilized peptides were resuspended in loading buffer A (30% acetonitrile and 0.07% TFA) and, if necessary, the pH was adjusted to 2.3 using 10% TFA. The samples were loaded onto the Fe^{3+}-IMAC column (Propac IMAC-10 4 x 5 mm column, Thermo Fischer Scientific). Bound

phosphopeptides were eluted with elution buffer B (0.3 % NH_4OH). The respective gradient is described in supplementary Table S1. The UV-abs signal at a wavelength of 280 nm was recoded at the outlet of the column and eluting phosphopeptides were collected manually. Subsequently, phosphopeptides were frozen in liquid nitrogen and freeze dried.

LC-MS/MS

Nanoflow LC-MS/MS analysis was performed using an Agilent 1290 (Agilent technologies, Middelburg, The Netherlands) coupled to an Orbitrap Q-Exactive HF-X (Thermo Fisher Scientific, Bremen, Germany). Lyophilized phosphopeptides or full proteome samples were resuspended in 20 mM citric acid (Sigma-Aldrich), 1 % (v/v) formic acid or 2 % (v/v) formic acid, respectively. Resuspended phosphopeptide, corresponding to 1.6 mg or 200 ng full proteome samples were injected, trapped and washed on a trap-column (100 µm i.d. × 2 cm, packed with 3 µm C18 resin, Reprosil PUR AQ, Dr. Maisch, packed in-house) for 5 min at a flow rate of 5 µL/minute with 100 % buffer A (0.1 FA, in HPLC grade water). Peptides were subsequently transferred onto an analytical column (75 µm x 60 cm Poroshell 120 EC-C18, 2.7 µm, Agilent Technology, packed in-house) and separated at room temperature at a flow rate of 300 nL/min using a 85 min linear gradient from 8 % to 32 % buffer B (0.1 % FA, 80 % ACN) or a 115 min linear gradient from13 % to 44 % buffer B. Electrospray ionization was performed using 1.9 kV spray voltage and a capillary temperature of 320 °C. The mass spectrometer was operated in data-dependent acquisition mode: full scan MS spectra (m/z 375 – 1,600) were acquired in the Orbitrap at 60,000 resolution for a maximum injection time of 20 ms with an AGC target value of 3e6 charges. Up to 12 precursors for phosphoproteome samples and up to 15 precursors for full proteome samples were selected for subsequent fragmentation. High resolution HCD MS2 spectra were generated using a normalized collision energy of 27 %. The intensity threshold to trigger MS2 spectra was set to 2e5, and the dynamic exclusion to 12 or 16, respectively. MS2 scans were acquired in the Orbitrap mass analyzer at a resolution of 30,000 (isolation window of 1.4 Th) with an AGC target value of 1e5 charges and a maximum ion injection time of 50 ms. Precursor ions with unassigned charge state as well as charge state of 1+ or superior/equal to 6+ were excluded from fragmentation.

Data analysis

Raw files were processed using MaxQuant software (version 1.6.3.4) and the Andromeda search engine was used to search against either a *B. subtilis* 168 (Uniprot/TrEMBL, December 2017, 4,247 entries) or *S. aureus* USA300 database (Uniprot, June 2018, 5,954 entries) with the following parameters for phosphoproteome analysis: trypsin digestion with a maximum of 3 missed cleavages, carbamidomethylation of cysteines (57.02 Da) as a fixed modification, methionine oxidation (15.99 Da), N-acetylation of proteins N-termini (42.01 Da) and phosphorylation on serine, threonine, tyrosine and histidine residues

(79.96 Da) as variable modifications. Mass tolerance was set to 4.5 ppm at the MS1 level and 20 ppm at the MS2 level. The False Discovery Rate (FDR) was set to 1% for peptide-spectrum matches (PSMs) and protein identification using a target-decoy approach, a score cut-off of 40 was used in the case of modified peptides and the minimum peptide length was set to 7 residues. The match between run feature was enabled with a matching time window of 0.7 min and an alignment time window of 20 min. The MaxQuant generated tables "evidence.txt" and "phospho (HSTY)Sites.txt" were used to calculate the number of unique phosphopeptides and phosphosites identified, respectively, and known contaminants were filtered out. For full proteome analysis the following deviations were applied: trypsin digestion with a maximum of 2 missed cleavages, carbamidomethylation of cysteine's (57.02 Da) as a fixed modification, methionine oxidation (15.99 Da), N-acetylation of protein N-termini (42.01 Da) as variable modifications. Relative label-free quantification was performed using the MaxLFQ algorithm with the minimum ratio count set to 2.

Identification of DNA/RNA contamination in MS2 spectra

Raw files were converted into .mgf with Proteome Discoverer (Vers. 2.3.0.523), using de-isotoping with an isotope deviation tolerance of 25 mmu. Subsequently mgf-files were analyzed using an in-house made script searching MS2 spectra for a 330.06 m/z peak with a 0.02 Da tolerance.

Statistical data analysis

For the phosphoproteome and full proteome analysis the MaxQuant generated "phospho (HSTY)Sites.txt" and "proteinGroups.txt" file, respectively, were used for subsequent statistical data analysis in R studio (R version 3.6.0). Four biological replicates per mutant were analysed. The data was filtered for "Reversed" and "Potential contaminant". In case of the phosphoproteome, an Andromeda localization score greater than 0.75 was required. Intensities for the phosphoproteome data, or LFQ intensities in case of the full proteome analysis, were log2 transformed. For each phosphosite or protein, the median calculated per sample was subtracted to compensate for systematic measurement effects. Only proteins with at least three valid values in one condition and two valid values in at least one other condition were considered. Data was checked for normal distribution before one-way ANOVA on each phosphosite or protein was done, after which the p-values were adjusted with the Benjamini-Hochberg procedure. The post-hoc Tukey Honestly Significant Difference (HSD) method was used to identify changing p-sites between the individual groups. A Tukey HSD *p*-value cut-off of 0.05 and a fold change cut-off of the mean +/- one standard-deviation of the data was used to select for significantly changing phosphosites or proteins between two groups.

Experiment design and statistical rationale

Each sample was grown in n=4 biological replicates, enriched and injected separately into the LC-MS/MS system. Each raw file was separately processed using the MaxQuant software. This analysis was sufficient to saturate the number of phosphosites detected.

Phosphosite environment analysis

Phosphosites listed in "phospho (HSTY)Sites.txt" with a localization probability of at least 0.75 were used for further analysis in R. To assess whether phosphorylation events preferentially occurred on flexible regions of proteins, the relative occurrence of amino acids present in a window of 5 amino acids before and after the phosphosites was computed for *S. aureus*, *E. coli* and human phosphoproteomes. The resulting amino acid frequencies were grouped into five groups, namely flexible (A/G/P), acidic (D/E), basic (K/R), aromatic (F/W/Y) and others. To investigate the conservation of *S. aureus* phosphosites across different organisms, the identified phosphoproteins were mapped back to their corresponding gene names (Uniprot retrieve ID/mapping tool). The PSP database (downloaded may 8th, 2020) was also mapped back to gene names and for all gene names found in both the MS data and PSP database, all corresponding proteins in both sets were aligned (msa package available via Bioconductor, http://www.bioinf.jku.at/software/msa/[28]) using the ClustalW algorithm with default settings. For each gene name, identity scores were computed by calculating the fraction of fully conserved amino acids across the whole alignment (i.e. the presence of gaps in one of the aligned protein sequences reduces the identity score). Scripts can be made available upon request.

Results & Discussion

Enhanced phosphopeptide identification for Gram-positive bacteria B. subtilis and S. aureus

Reversible protein phosphorylation allows quick and effective signal transduction upon changing environmental conditions such as carbon accessibility or pH changes. This adaptation is especially important for bacterial colonization of different hosts but also after invasion and spread to different tissues. Recently, Potel *et al*. optimized sample preparation for the phosphopeptide identification from Gram-negative bacteria[11]. Here we present an extension of this protocol for application with Gram-positive bacteria, which contain a thick peptidoglycan layer and therefore require a more stringent lysis protocol. A three-step lysis protocol was developed, which combined chemical lysis using chaotropic agents and detergents with mechanical lysis by bead beating and subsequent sonication (see Material and Methods). After sample clean up and protein digestion, phosphopeptides were enriched via Fe^{3+}-IMAC and analyzed by LC-MS/MS (Figure 1a). Optimization was initially

tested on the Gram-positive model organism *B. subtilis* 168, before applying the optimized method to the lysis of the pathogenic bacteria *S. aureus* (transposon mutant NE98 of the strain *S. aureus USA300* JE3 containing a disruption in unrelated surface protein encoding gene *sdrE* (see Material and Methods)). The optimized lysis protocol for Gram-positive bacteria resulted in the identification of 283 phosphopeptides (pS, pT, pY and pH) for *B. subtilis* (Figure S1a, Supplementary Table 2). When filtering on the Andromeda localization probability, 176 class I phosphosites on 146 proteins were identified in our study. This is a slight increase compared to the study of Ravikumar *et al*[29] (Figure 1b). For *S. aureus* the improvement was more striking. Here, 3,800 phosphopeptides (pS, pT, pY and pH) were identified in at least two out of three biological replicates, making it the largest phosphosite dataset of *S. aureus* to date and representing a more than 21-fold increase compared to a recent study by Junker *et al.*, which identified only 173 phosphopeptides (pS, pT and pY) in *S. aureus* strain COL[30] (Figure S1b. Supplementary Table 2). Also the identification of class I phosphosites (2,852, Figure 1c), was improved by more than 17-fold compared to previous work[30]. To confirm that this improvement resulted from the reduction of DNA/RNA contamination, we looked for the diagnostic ion of 330.06 *m/z* in MS2 spectra. As previously demonstrated, the use of DNase and benzonase during sample preparation decreased the percentage of MS2 spectra containing the 330.06 *m/z* ion from around 35 % to 8 % for human samples and from 75 % to 13 % for *E. coli*, thereby enhancing phosphopeptide identification[11]. For the Gram-positive bacteria, the amount of contamination was in a similar range; 3.5 % of all MS2 spectra of *B. subtilis* and around 18 % of all MS2 spectra of *S. aureus* contained the 330.06 *m/z* ion (Figure 1d). In comparison, 50 % of all MS2 spectra of the Junker *et al*[30] study contained the 330.06 *m/z* ion (Figure 1d), revealing a clear co-enrichment of these contaminants. Therefore, the use of DNase and benzonase in our optimized sample preparation protocol improves phosphopeptide identification also for Gram-positive bacteria, by minimizing the binding of contaminants to the Fe^{3+}-IMAC column.

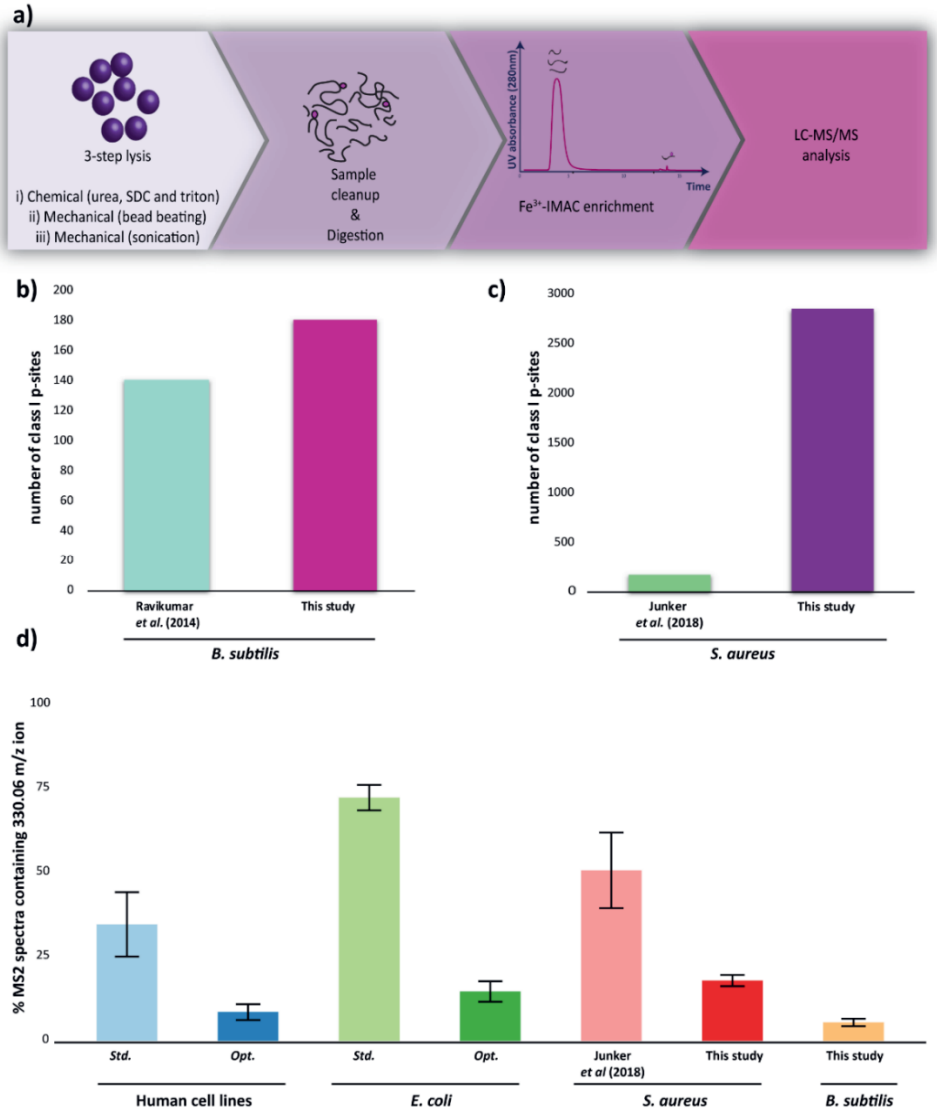


Figure 1: Experimental workflow and literature comparison. **a)** Cell lysis was optimized by introducing a three-step lysis using a mix of chemical and mechanical lysis. After sample cleanup and digestion phosphopeptides were enriched using a Fe^{3+}-IMAC column and analyzed via LC-MS/MS. **b)** The number of identified class I (Andromeda localization score >0.75) phosphosites (176) for *B. subtilis* was compared to a recent study by Ravikumar et al.[29] and showed a 1.3x increase. **c)** In total 2,852 class I phosphosites were identified for *S. aureus* resulting in a 17 fold increase compared to Junker et al.[30] **d)** Percentage of MS2 spectra containing the diagnostic ion 330.06 m/z for human cell lines, *E.coli*, *S. aureus* and *B. subtilis*.

Phosphosite distribution of S. aureus

The phosphosite distribution of eukaryotes and prokaryotes differs to a great extent. Whereas around 80 - 90 % of all eukaryotic phosphorylation sites are localized on serine, phosphorylation on serine only accounts for 40 – 60 % of all phosphorylation events in prokaryotes depending on the species[2,31,32]. In addition to a clear shift towards threonine phosphorylation, prokaryotes also show a preference towards protein histidine phosphorylation compared to eukaryotes. We recently confirmed these preferences by identifying 80 % of phosphorylation from human cell lines being localized on serine in contrast to only 57 % of *E. coli* peptides (Figure 2, a & b)[10]. A similar preference was observed for *S. aureus*, where only 55 % of all phosphorylation sites are localized on serine but almost 30 % on threonine (55 % pS, 29.6 % pT, 7.3 % pY and 7.9 % pH). Interestingly, *B. subtilis* showed a higher percentage of serine phosphorylation (73% pS, 12.7% pT, 7.5 pY and 6.7% pH (Figure 2, c & d)), in agreement with previous phosphoproteomic studies on *B. subtilis*[29,33]. Not only the phosphosite distribution but also the multiplicity of phosphorylation, differs between eukaryotes and prokaryotes[11]. Here, we could confirm that the multiplicity of phosphopeptides for Gram-positive bacteria is similar to the Gram-negative bacterium *E. coli* and favors singly phosphorylated peptides. Indeed, we found that more than 90 % of all identified prokaryotic phosphopeptides are singly phosphorylated (Figure 2e), whereas eukaryotes showed a broad range of multiple phosphorylated peptides.

Interestingly, we could show that *S. aureus* also favored multiple phosphorylated proteins almost to the same extent as singly phosphorylated proteins, whereas *E. coli* and *B. subtilis* showed a high preference for singly phosphorylated proteins (Figure 2f). Proteins with multiple phosphorylation sites are capable of generating different proteoforms, which are able to fulfil differential functions, which allows for a more complex and precise regulation of the pathogen[2]. Previous work, indicated that species with smaller genome sizes and lower number of bacterial transcription initiation factors (so called sigma factors) display more multiple phosphorylated proteins, suggesting regulation is mainly driven by PTMs in those bacteria[2]. *S. aureus* indeed has considerable smaller genome and lower number of sigma factors compared to both *E. coli* and *B. subtilis.*

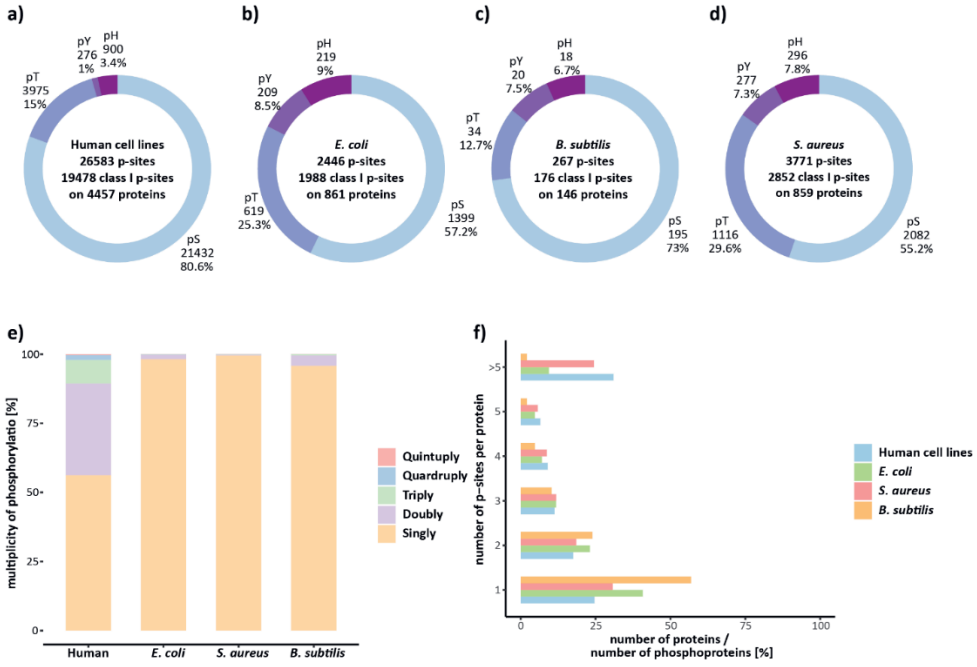


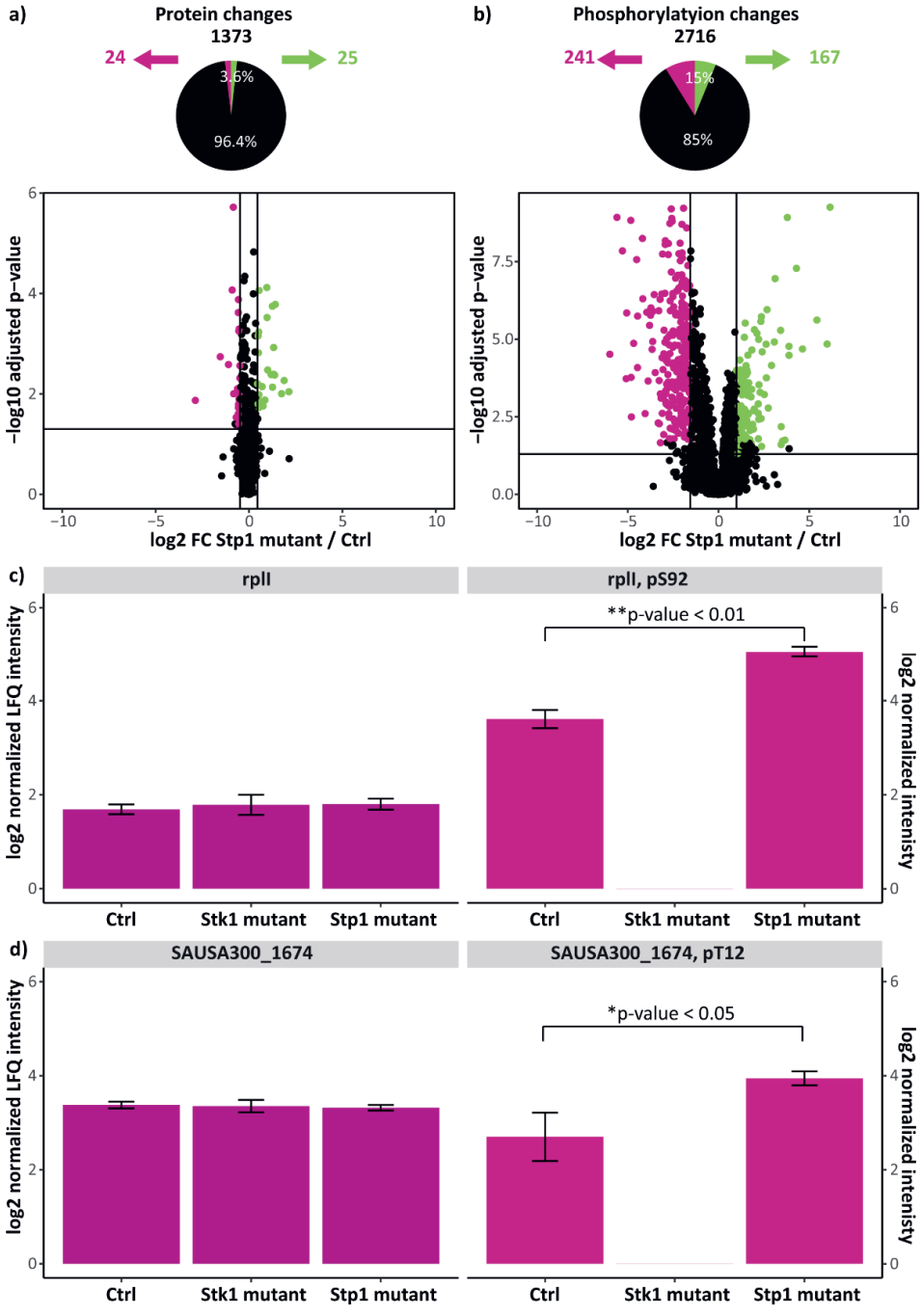
Figure 2: Comparison of phosphorylation characteristic for human cell lines[11], *E. coli*[11], *B. subtilis* and *S. aureus*. **a-d)** Number of identified phosphosites, class I phosphosites (Andromeda localization probability > 0.75) and phosphoproteins as well as the distribution of serine, threonine, tyrosine and histidine phosphosites identified. **e)** Percentage of singly, doubly, triply, quadruply and quintuply identified phosphorylated peptides for all four organisms. **f)** Percentage of phosphoproteins that have 1, 2, 3, 4, 5, or more phosphorylation sites for all four organisms.

Optimized sample preparation allowed identification of in total 74 potential Stk1 targets

Even though it was long assumed that prokaryotes mainly exploit protein histidine phosphorylation comprised in TCSs, it is now obvious that Ser/Thr kinases play an important role in protein signaling as well[34]. Depending on the strain *S. aureus* comprises 16 to 18 TCSs but only one eSTKs[22,23,26]. Therefore, we performed a phosphoproteome study to elucidate the targets of Stk1 as well as the cognate phosphatase Stp1. Hereto, the transposon mutants NE217 and NE1919 obtained from the Nebraska Transposon Mutant Library (NTML)[27] were used. NE217 and NE1919 have an inactivating disruption of the gene *pknB* (encoding Stk1) or the gene coding for the protein phosphatase 2C domain-containing protein Stp1. These mutants will be further referred to as Stk1 mutant and Stp1 mutant, respectively. NE98, was used as a control sample, allowing the analysis of phosphoproteome changes related to the activity of these two enzymes. Gene disruption and absence of those three proteins in the respective mutants was confirmed by PCR, full proteome as well as phosphoproteome analysis (Figure S2, Figure S3).

To identify the changes upon *S. aureus* mutants, the three strains (n=4 biological replicates) were cultured, lysed and digested. Phosphopeptides were enriched using a Fe^{3+}-IMAC column and analyzed via LC-MS/MS. Since Stk1 and Stp1 are being reported as a writer-eraser pair[35], potential targets of the kinase or phosphatase should show opposite behavior. Therefore, phosphosites over-represented in the Stp1 mutant compared to the control and under-represented or even absent in the Stk1 mutant and showing no significant change on the full proteome level are most likely to be substrates of these enzymes (Figure S4). We identified and quantified 2,716 phosphosites in both the Stp1 mutant and the control (see Material and Methods, Figure 3, Supplementary Table 3). 15 % of these sites were significantly changing (Tukey HSD *p*-value cut-off of 0.05 and a fold change cut-off of of $\bar{x} \pm \sigma$ of the data). 241 of these phosphosites were significantly over-represented and 167 phosphosites under-represented in the Stp1 mutant compared to the control (Figure 3b). On the full proteome level 1,373 proteins were quantified, of which 3.7 % were significantly different between the two conditions (Figure 3a, Supplementary Table S4). Hence, most of the changes are occurring on the phosphoproteome level instead of the proteome level, which is in line with our expectations when Stk1 or Stp1 are absent. Phosphorylation on S92 of the 50S ribosomal protein L9 (rplI) showed a significant over-representation (Tukey HSD *p*-value < 0.01) in the Stp1 mutant compared to the control, whereas this phosphorylation could not be identified for the Stk1 mutant. On the full proteome level this protein did not change significantly according to Tukey HSD (Figure 3c). The same behavior was observed for the phosphosite pT12 of the Putative serine protease HtrA (SAUSA300_1674, Figure 3d).

Figure 3: Identified phosphorylation and protein changes in *S. aureus*. **a)** Volcano plot comparing proteins between Stp1 mutant and Ctrl, identifying 51 proteins with significantly changing (Tukey HSD p-value cut-off of 0.05 and a fold change cut-off of $\bar{x} \pm \sigma$ of the data). Twenty-four proteins were under-represented (pink dots) and 25 proteins were over-represented (green dots). **b)** Volcano plot comparing the phosphorylation sites, identifying 408 changes between Stp1 mutant and Ctrl (Tukey HSD p-value cut-off of 0.05 and a fold change cut-off of $\bar{x} \pm \sigma$ of the data). Two hundred forty-one phosphosites were under-represented (pink dots) and 167 phosphosites were over-represented (green dots). **c)** Overrepresentation of S92 phosphorylation in RplI in the Stp1 mutant compared to Ctrl (Tukey HSD p-value < 0.01) and absence of phosphorylation on the Stk1 mutant (right panel). No change in protein abundance was observed between the three mutants (left panel). **d)** Overrepresentation of Thr12 phosphorylation on SAUSA300_1674 in the Stp1 mutant compared to Ctrl (Tukey HSD p-value < 0.05) and absence of phosphorylation in the Stk1 mutant (right panel). On the proteome level no significant changes were detected in the three mutants (left panel). For all bar charts the center values represent the mean and error bars the standard deviation (s.d.) for n=4 biological independent replicates.


a)
Protein changes
1373
24
25
3.6%
96.4%
b)
Phosphorylatyion changes
2716
241
167
15%
85%
−log10 adjusted p−value
log2 FC Stp1 mutant / Ctrl
c)
rpll
rpll, pS92
**p-value < 0.01
log2 normalized LFQ intensity
log2 normalized intenisty
Ctrl
Stk1 mutant
Stp1 mutant
d)
SAUSA300_1674
SAUSA300_1674, pT12
*p-value < 0.05

In total 20 phosphosites were identified showing this characteristic (see Table 1 and Table 2, supplementary Table S2, Figure S5 and Figure S6). Moreover, 54 phosphosites were exclusively identified in the Stp1 mutant, whereas the protein level was not affected significantly on the full proteome level or was not detected (Supplementary Table S2). Thus, LFQ quantitative analysis of the three mutants (Stp1, Stk1 and Ctrl) allowed the identification 74 potential Stk1/Stp1-related phosphosites which were distributed over the whole dynamic range of identified phosphopeptides as well as phosphoproteins (Supplementary Figure S7).

Table 1. Phosphosites over-represented in Stp1 mutant/Ctrl but not identified in Stk1 without significant change on full proteome level.

Gene	Uniprot ID	Protein name	Position within protein
ftnA	Q2FFK2	Bacterial non-heme ferritin	T98
map	A0A0H2XG88	Methionine aminopeptidase (MAP) (MetAP)	T153
rplI	Q2FKP1	50S ribosomal protein L9	S92
rplQ	Q2FER6	50S ribosomal protein L17	T33
ruvB	Q2FG86	Holliday junction ATP-dependent DNA helicase RuvB	S9
SAUSA300_0236	A0A0H2XGK7	PTS system, IIBC components	S475
SAUSA300_0905	A0A0H2XGC0	Putative adenylate cyclase	T3
SAUSA300_1230	A0A0H2XIS0	Uncharacterized protein	T5
SAUSA300_1240	A0A0H2XFT2	UPF0154 protein SAUSA300_1240	T80
SAUSA300_1674	A0A0H2XGV4	Putative serine protease HtrA	T12
SAUSA300_1856	A0A0H2XHR8	Uncharacterized protein	T89
SAUSA300_1909	A0A0H2XEY2	Uncharacterized protein	T143
secA1	Q2FIN8	Protein translocase subunit SecA 1	S702
secA1	Q2FIN8	Protein translocase subunit SecA 1	S565

Table 2. Phosphosites over-represented in Stp1 mutant/Ctrl but not identified in Stk1 and no identification on full proteome level.

Gene	Uniprot ID	Protein name	position within protein
pknB	A0A0H2XGG5	Protein kinase	T161
pknB	A0A0H2XGG5	Protein kinase	T166
pknB	A0A0H2XGG5	Protein kinase	T288
pknB	A0A0H2XGG5	Protein kinase	T314
SAUSA300_1574	Q2FGA9	UPF0297 protein SAUSA300_1574	T7
rplS HMPREF0776_2241	A0A0E1VS63	50S ribosomal protein L19 (Fragment)	T11

Phosphoproteomic analysis reveals new targets of Stk1

We identified the sensor protein kinase HptS (part of the HptRS TCS) as a target of Stk1 (pT272, Supplementary Table S3). Previous studies showed that Stk1 phosphorylates RRs of at least two TCSs; GraRS and VraRS[24,25]. With our identification of HptS as a specific target, we expand the repertoire of TCSs that are regulated by Stk1 to HptRS. But in contrast to GraRS and VraRS, in which the RR is phosphorylated, here Stk1 phosphorylates the sensor-like histidine kinase HtpS. HptRS belongs to the hexose phosphate transport system (HPT) that is partly responsible for the uptake of extracellular sugars that can vary between different host organisms[36–39]. Therefore, being able to switch quickly and efficiently between available carbon sources is critical for the survival and colonization of bacteria.

Transcription of HptRS is regulated by Catabolite control protein A (CcpA). CcpA has been reported to be phosphorylated in an Stk1-dependent manner on T18 and T33 in *S. aureus* strain N315[40]. Both phosphosites are located within the DNA-binding site of CcpA and phosphorylation seems to interfere with DNA binding[40]. Here, we identified eight serine and threonine phosphosites, including pT18, but none of those sites were significantly changing between the three mutants. Therefore, we hypothesize that Stk1/Stp1 is probably not the only regulator of phosphorylation of CcpA.

CcpA activation can also occur via a sugar dependent manner which involves Histidine-containing protein (HPr)[40–42]. In the presence of the primary carbon source glucose, HPr is phosphorylated by HPr kinase/phosphatase (HPrK/P) on S46. Only pS46-HPr is able to bind to CcpA thereby enabling its binding to the catabolite response element (*cre)* on the chromosomal DNA, which can either lead to repression or activation of the respective gene transcription[38,39,42]. pS46-HPr was identified in all three mutants without any significant change. However, phosphorylation of HPrK/P on S294 was identified exclusively in the Stp1

mutant, without changes in protein expression levels. Therefore, we suggest that phosphorylation of HPrK/P by Stk1 does not negatively influence its ability to phosphorylate HPr on S46. Even though Stk1 seems to phosphorylate important proteins within the CcpA regulatory machinery, we could not correlate those phosphorylation events to any changes in the abundance of proteins regulated by CcpA. Still, our data indicates extensive regulation of the *ccpA* regulon on multiple levels, which highlights its importance.

Finally, Elongation factor tu (EF-Tu) was one of the proteins with phosphosites exclusively identified in the Stp1 mutant. EF-Tu was previously reported to be phosphorylated by Stk1 but without information on the phosphorylation site. Here, pT34 and pT257 were identified as two phosphosites on EF-Tu (Supplementary Table S3). EF-Tu is one of the most abundant bacterial proteins that consists of three functional domains, domain I (amino acids 1-199) comprising the GTP/GDP binding domains, domain II (amino acids 200- 299) and domain III (amino acids 300-394) that both regulate the activity of domain I [43–45]. The main and most studied function of this protein is mediating protein translational elongation. Sajid *et al.* showed that *Mycobacterium tuberculosis* EF-Tu is phosphorylated by an Stk1 homolog on 11 threonine residues[46]. One of those, pT259, is homologous to pT257 identified in EF-Tu for *S. aureus*. However, Sajid *et al.* could not determine a negative effect of the phosphorylation on protein synthesis in *M. tuberculosis*. In our study, no growth defects on blood agar plates or in the liquid cultures using full media were observed for both Stk1 and Stp1 mutants, which is in line with previous studies[47]. Given the essential role of EF-Tu in protein synthesis, it remains elusive how Stk1/Stp1 influence EF-Tu regulation.

Previous work has indicated that Stk1 and Stp1 mutants display defects in cell division as well as cell wall structure, affecting antibiotic susceptibility[48]. With the identification of new Stk1/Stp1 substrates, future research can be more focused on how these target proteins and their phosphorylation sites are involved in these cellular processes.

Extensive S/T phosphorylation indicates high S/T kinase activity in S. aureus strain USA300

The identification of more than 3,000 phosphosites localized on Ser and Thr (Figure 1d) indicates that Ser/Thr kinase signaling, and activity is much higher than previously anticipated. By deleting the only known eSTK, Stk1, we could still identify around 2,700 phosphosites localized on Ser/Thr (see Figure S8, Supplementary Table S2). In addition, the quantification of only 74 potential Stk1 targets out of 2,977 phosphosites localized on Ser or Thr raises the question what the source of these other phosphosites is. The *S. aureus* strain USA300 has two other Serine-protein kinases, HprK and RsbW. However, the number of currently reported targets for these kinases is low [49]. This suggest that *S. aureus* USA300 must have other, currently unknown, Ser/Thr kinases.

In an effort to support this hypothesis, we tried to identify specific motifs and analyzed the flanking amino acids (+/- five amino acids) around pSer and pThr and compared this to well-studied human phosphosite data (HeLa phosphoproteome[11]). As expected, we could see a higher frequency for proline and serine adjacent to the phosphosites in human cell lines. This pattern was not observed in *S. aureus* and *E. coli* where acidic amino acids (aspartic and glutamic acid) as well as basic amino acids (lysine and arginine) were slightly more abundant adjacent to the phosphosite (see supplementary Figure S9). Our analysis also directly showed that the identified phosphopeptides do not have an obvious motif, which was confirmed by the absence of enriched motifs using either iceLogo[50] or pLogo[51]. Therefore, most of the identified phosphorylation events in *S. aureus* do not seem to occur within highly specific protein motifs. In addition, we were unable to obtain potential phosphorylation motifs using MEME[52]. However, we cannot exclude that more specific motifs are buried in this large number of analyzed phosphosites.

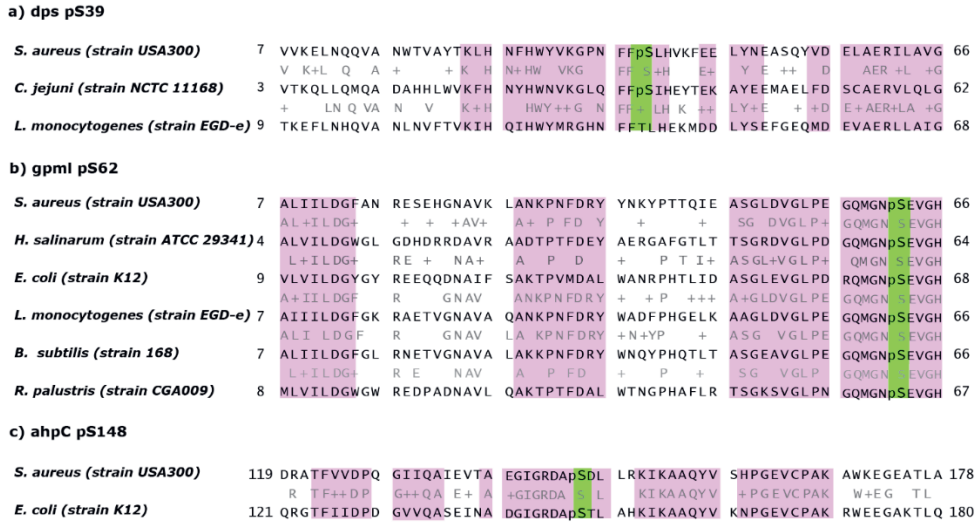


Figure 4: Sequence alignments obtained with the NCBI Blastp (https://blast.ncbi.nlm.nih.gov/Blast.cgi) search for **a)** Dps, **b)** GpmI and **c)** AhpC. Conserved phosphosites are highlighted in green and regions of high conservation (more than 5 subsequent amino acids) are highlighted in pink. pS39 in Dps is conserved in *C. jejuni* and even though no phosphorylation is reported so far *L. monocytogenes* contains S/T substitution at this position. pS62 in GpmI is conserved *in H. salinarum*, *E. coli*, *L. monocytogenes*, *B. subtilis* and *R. palustris*. pS148 in AhpC is conserved in *E. coli*. All three proteins show a high degree of conservation around the phosphosite.

We also compared the identified pSer and pThr sites identified in this study with the entries in dbPSP[32]. Here we could match 286 phosphoproteins based on their gene name. Proteins not having a gene name, but only a gene locus ID were not included in the analysis. From the 286 genes, we identified 15 phosphosites that were conserved between different species. The alignments of those proteins show conservation hotspots (more than five subsequent amino acids, pink) especially localized around the conserved phosphosite

(green) (Figure 4, Figure S10). One of those conserved phosphosites, pS62, on GpmI (Figure 4b) is known to be a phosphoserine intermediate (Uniprot, UniRule: UR000100531), however that was the only identified site reported as a metabolic phosphoserine intermediate. Even though we were only able to map 15 out of the more than 3,000 identified phosphosites to already reported phosphorylation events, the observed conservation hotspots emphasized a high degree of conservation especially around the phosphosites. Further we could also identify conservation hotspots, for which we identified phosphosites in *S. aureus*, but no phosphorylation was detected in other species (Figure 5).

rplS

Figure 5: Sequence alignment of RplS obtained with NCBI BlastP (https://blast.ncbi.nlm.nih.gov/Blast.cgi). Ser/Thr phosphosites are highlighted in green and regions of high conservation (more than 5 subsequent amino acids) are highlighted in pink. In total we, identified seven phosphosites of which three (pS13, pS21 and pS38) were conserved according to dbPSP[32] and NCBI BlastP.

This rather low overlap in identified phosphosites but high conservation of certain regions is a reflection of the poor coverage of bacterial phosphoproteomes. This clearly highlights the importance of this study, in which improved sample preparation reveals an extreme underestimation of bacterial phosphorylation mediated signaling. Despite the low overlap, our results still show that phosphorylation events are not just random even though specific motif were not identified. When looking for additional kinases by homology searches using NCBI protein Blast (https://blast.ncbi.nlm.nih.gov/Blast.cgi) or with motif searches using pfam[53], we did not obtain any results. Therefore, we recommend thorough characterization of the *S. aureus* proteome in order to identify novel kinases. Those kinases could play an important role in the virulence and versatility of this pathogen and hence identification of those kinases could help to understand and combat its pathogenicity.

Acknowledgments

We acknowledge Pieter van Breugel for technical assistance. S.L. acknowledges support from the Netherlands Organization for Scientific Research (NWO) through a VIDI grant (project 723.013.008).

Competing financial interest

The authors declare no competing financial interest

Data availability

All raw data that support the findings of this study have been deposited in ProteomeXchange with the accession number PXD020226.

References

1. Ninfa AJ. Use of two-component signal transduction systems in the construction of synthetic genetic networks. *Curr Opin Microbiol*. 2010;13(2):240-245. doi:10.1016/j.mib.2010.01.003

2. Ge R, Shan W. Bacterial Phosphoproteomic Analysis Reveals the Correlation Between Protein Phosphorylation and Bacterial Pathogenicity. *Genomics Proteomics Bioinformatics*. 2011;9(4-5):119-127. doi:10.1016/S1672-0229(11)60015-6

3. Manai M, Cozzone AJ. Analysis of the protein-kinase activity of Escherichia coli cells. *Biochem Biophys Res Commun*. 1979;91(3):819-826. doi:10.1016/0006-291X(79)91953-3

4. Pinkse MWH, Uitto PM, Hilhorst MJ, Ooms B, Heck AJR. 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(14):3935-3943. doi:10.1021/ac0498617

5. Larsen MR, Thingholm TE, Jensen ON, Roepstorff P, D. JJ. Highly Selective Enrichment of Phosphorylated Peptides from Peptide Mixtures Using Titanium Dioxide Microcolumns. *Mol Cell Proteomics*. 2005;4(7):873-886. doi:10.1074/mcp.T500007-MCP200

6. Villén J, Gygi SP. The SCX/IMAC enrichment approach for global phosphorylation analysis by mass spectrometry. *Nat Protoc*. 2008;3(10):1630-1638. doi:10.1038/nprot.2008.150

7. Posewitz MC, Tempst P. Immobilized gallium(III) affinity chromatography of phosphopeptides. *Anal Chem*. 1999;71(14):2883-2892. doi:10.1021/ac981409y

8. Olsen J V., Mann M. Status of large-scale analysis of posttranslational modifications by mass spectrometry. *Mol Cell Proteomics*. 2013;12(12):3444-3452. doi:10.1074/mcp.O113.034181

9. Riley NM, Coon JJ. Phosphoproteomics in the Age of Rapid and Deep Proteome Profiling. *Anal Chem*. 2016;88(1):74-94. doi:10.1021/acs.analchem.5b04123

10. Lemeer S, Heck AJ. The phosphoproteomics data explosion. *Curr Opin Chem Biol*. 2009;13(4):414-420. doi:10.1016/j.cbpa.2009.06.022

11. Potel CM, Lin M-H, Heck AJR, Lemeer S. Defeating major contaminants in Fe 3+-IMAC phosphopeptide enrichment. *Mol Cell Proteomics*. Published online 2018:mcp.TIR117.000518. doi:10.1074/mcp.TIR117.000518

12. Lee AS, Huttner B, Harbarth S. Prevention and Control of Methicillin-Resistant Staphylococcus aureus in Acute Care Settings. *Infect Dis Clin North Am*. 2016;30(4):931-952. doi:10.1016/j.idc.2016.07.006

13. Foster TJ. The Staphylococcus aureus " superbug " Find the latest version : The Staphylococcus aureus " superbug ." 2004;114(12):1693-1696. doi:10.1172/JCI200422123.7.

14. Wertheim HF, Melles DC, Vos MC, van Leeuwen W, van Belkum A, Verbrugh HA, Nouwen JL. The role of nasal carriage in Staphylococcus aureus infections. *Lancet Infect Dis*. 2005;5(12):751-762. doi:10.1016/S1473-3099(05)70295-4

15. Tacconelli E, Carrara E, Savoldi A, Harbarth S, Mendelson M, Monnet DL, Pulcini C, Kahlmeter G, Kluytmans J, Carmeli Y, Ouellette M, Outterson K, Patel J, Cavaleri M, Cox EM, et al. Discovery, research, and development of new antibiotics: the WHO priority list of antibiotic-resistant bacteria and tuberculosis. *Lancet Infect Dis*. 2018;18(3):318-327. doi:10.1016/S1473-3099(17)30753-3

16. Mainiero M, Goerke C, Geiger T, Gonser C, Herbert S, Wolz C. Differential Target Gene Activation by the Staphylococcus aureus Two-Component System saeRS. *J Bacteriol*. 2010;192(3):613-623. doi:10.1128/JB.01242-09

17. Ziebandt A-K, Weber H, Rudolph J, Schmid R, Höper D, Engelmann S, Hecker M. Extracellular proteins ofStaphylococcus aureus and the role of SarA andσB. *Proteomics*. 2001;1(4):480-493. doi:10.1002/1615-9861(200104)1:4<480::AID-PROT480>3.0.CO;2-O

18. Tuchscherr L, Medina E, Hussain M, Völker W, Heitmann V, Niemann S, Holzinger D, Roth J, Proctor RA, Becker K, Peters G, Löffler B. Staphylococcus aureus phenotype switching: An

effective bacterial strategy to escape host immune response and establish a chronic infection. *EMBO Mol Med*. 2011;3(3):129-141. doi:10.1002/emmm.201000115

19. Jenul C, Horswill AR. Regulation of Staphylococcus aureus Virulence. Microbiol Spectr. 2019 Apr 5;7(2):10.1128/microbiolspec.GPP3-0031-2018. doi: 10.1128/microbiolspec.GPP3-0031-2018. PMID: 30953424; PMCID: PMC6452892

20. Foster TJ, Höök M. Surface protein adhesins of Staphylococcus aureus. 1998;(98):484-488.

21. Gotoh Y, Eguchi Y, Watanabe T, Okamoto S, Doi A, Utsumi R. Two-component signal transduction as potential drug targets in pathogenic bacteria. *Curr Opin Microbiol*. 2010;13(2):232-239. doi:10.1016/j.mib.2010.01.008

22. Junecko JM. Transcribing virulence in Staphylococcus aureus. *World J Clin Infect Dis*. 2012;2(4):63. doi:10.5495/wjcid.v2.i4.63

23. Villanueva M, García B, Valle J, Rapún B, Ruiz De Los Mozos I, Solano C, Martí M, Penadés JR, Toledo-Arana A, Lasa I. Sensory deprivation in Staphylococcus aureus. *Nat Commun*. 2018;9(1):1-12. doi:10.1038/s41467-018-02949-y

24. Fridman M, Williams GD, Muzamal U, Hunter H, Siu KWM, Golemi-Kotra D. Two unique phosphorylation-driven signaling pathways crosstalk in staphylococcus aureus to modulate the cell-wall charge: Stk1/Stp1 meets GraSR. *Biochemistry*. 2013;52(45):7975-7986. doi:10.1021/bi401177n

25. Canova MJ, Baronian G, Brelle S, Cohen-Gonsaud M, Bischoff M, Molle V. A novel mode of regulation of the Staphylococcus aureus Vancomycin-resistance-associated response regulator VraR mediated by Stk1 protein phosphorylation. *Biochem Biophys Res Commun*. 2014;447(1):165-171. doi:10.1016/j.bbrc.2014.03.128

26. Cheung A, Duclos B. Stp1 and Stk1 : The Yin and Yang of Vancomycin Sensitivity and Virulence in Vancomycin-Intermediate Staphylococcus aureus Strains. 2012;205:1625-1627. doi:10.1093/infdis/jis255

27. Fey PD, Endres JL, Yajjala VK, Widhelm TJ, Boissy RJ, Bose JL, Bayles KW. A Genetic Resource for Rapid and Comprehensive Phenotype Screening of Nonessential Staphylococcus aureus Genes. *MBio*. 2013;4(1):e00537-12-e00537-12. doi:10.1128/mBio.00537-12

28. Bodenhofer U, Bonatesta E, Horejš-Kainrath C, Hochreiter S. Msa: An R package for multiple sequence alignment. *Bioinformatics*. 2015;31(24):3997-3999. doi:10.1093/bioinformatics/btv494

29. Ravikumar V, Shi L, Krug K, Derouiche A, Jers C, Cousin C, Kobir A, Mijakovic I, Macek B. Quantitative Phosphoproteome Analysis of *Bacillus subtilis* Reveals Novel Substrates of the Kinase PrkC and Phosphatase PrpC. *Mol Cell Proteomics*. 2014;13(8):1965-1978. doi:10.1074/mcp.M113.035949

30. Junker S, Maaβ S, Otto A, Michalik S, Morgenroth F, Gerth U, Hecker M, Becher D. *Spectral Library Based Analysis of Arginine Phosphorylations in* Staphylococcus Aureus. Vol 17.; 2018. doi:10.1074/mcp.RA117.000378

31. Sharma K, D'Souza RCJ, Tyanova S, Schaab C, Wiśniewski J, Cox J, Mann M. Ultradeep Human Phosphoproteome Reveals a Distinct Regulatory Nature of Tyr and Ser/Thr-Based Signaling. *Cell Rep*. 2014;8(5):1583-1594. doi:10.1016/j.celrep.2014.07.036

32. Pan Z, Wang B, Zhang Y, Wang Y, Ullah S, Jian R, Liu Z, Xue Y. DbPSP: A curated database for protein phosphorylation sites in prokaryotes. *Database*. 2015;2015(August 2018):1-8. doi:10.1093/database/bav031

33. Macek B, Mijakovic I, Olsen J V., Gnad F, Kumar C, Jensen PR, Mann M. The Serine/Threonine/Tyrosine Phosphoproteome of the Model Bacterium Bacillus subtilis. *Mol Cell Proteomics*. 2007;6(4):697-707. doi:10.1074/mcp.M600464-MCP200

34. Cozzone AJ. Role of protein phosphorylation on serine/threonine and tyrosine in the virulence of bacterial pathogens. *J Mol Microbiol Biotechnol*. 2006;9(3-4):198-213. doi:10.1159/000089648

35. Zheng W, Cai X, Li S, Li Z. Autophosphorylation mechanism of the Ser/Thr Kinase Stk1 from Staphylococcus aureus. *Front Microbiol*. 2018;9(APR):1-10. doi:10.3389/fmicb.2018.00758

36. Park JY, Kim JW, Moon BY, Lee J, Fortin YJ, Austin FW, Yang SJ, Seo KS. Characterization of a novel two-component regulatory system, HptRS, the regulator for the hexose phosphate transport system in Staphylococcus aureus. *Infect Immun*. 2015;83(4):1620-1628. doi:10.1128/IAI.03109-14

37. Yang Y, Hu M, Yu K, Zeng X, Liu X. Mass spectrometry-based proteomic approaches to study pathogenic bacteria-host interactions. *Protein Cell*. 2015;6(4):265-274. doi:10.1007/s13238-015-0136-6

38. Reed JM, Olson S, Brees DF, Griffin CE, Grove RA, Davis PJ, Kachman SD, Adamec J, Somerville GA. Coordinated regulation of transcription by CcpA and the Staphylococcus aureus two-component system HptRS. Becker K, ed. *PLoS One*. 2018;13(12):e0207161. doi:10.1371/journal.pone.0207161

39. Schumacher MA, Allen GS, Diel M, Seidel G, Hillen W, Brennan RG. Structural Basis for Allosteric Control of the Transcription Regulator CcpA by the Phosphoprotein HPr-Ser46-P. *Cell*. 2004;118(6):731-741. doi:10.1016/j.cell.2004.08.027

40. Leiba J, Hartmann T, Cluzel M-E, Cohen-Gonsaud M, Delolme F, Bischoff M, Molle V. A Novel Mode of Regulation of the Staphylococcus aureus Catabolite Control Protein A (CcpA) Mediated by Stk1 Protein Phosphorylation. *J Biol Chem*. 2012;287(52):43607-43619. doi:10.1074/jbc.M112.418913

41. Deutscher J, Küster E, Bergstedt U, Charrier V, Hillen W. Protein kinase-dependent HPr/CcpA interaction links glycolytic activity to carbon catabolite repression in Gram-positive bacteria. *Mol Microbiol*. 1995;15(6):1049-1053. doi:10.1111/j.1365-2958.1995.tb02280.x

42. Görke B, Stülke J. Carbon catabolite repression in bacteria: Many ways to make the most out of nutrients. *Nat Rev Microbiol*. 2008;6(8):613-624. doi:10.1038/nrmicro1932

43. Parmeggiani A, Swart GWM, Mortensen KK, Jensen M, Clark BF, Dente L, Cortese R. Properties of a genetically engineered G domain of elongation factor Tu. *Proc Natl Acad Sci U S A*. 1987;84(10):3141-3145. doi:10.1073/pnas.84.10.3141

44. Widjaja M, Harvey KL, Hagemann L, Berry IJ, Jarocki VM, Raymond BBA, Tacchi JL, Gründel A, Steele JR, Padula MP, Charles IG, Dumke R, Djordjevic SP. Elongation factor Tu is a multifunctional and processed moonlighting protein. *Sci Rep*. 2017;7(1):1-17. doi:10.1038/s41598-017-10644-z

45. Prezioso SM, Brown NE, Goldberg JB. Elfamycins: inhibitors of elongation factor-Tu. *Mol Microbiol*. 2017;106(1):22-34. doi:10.1111/mmi.13750

46. Sajid A, Arora G, Gupta M, Singhal A, Chakraborty K, Nandicoori VK, Singh Y. Interaction of Mycobacterium tuberculosis Elongation Factor Tu with GTP Is Regulated by Phosphorylation. *J Bacteriol*. 2011;193(19):5347-5358. doi:10.1128/JB.05469-11

47. Débarbouillé M, Dramsi S, Dussurget O, Nahori M-AMA, Vaganay E, Jouvion G, Cozzone A, Msadek T, Duclos B, Debarbouille M, Dramsi S, Dussurget O, Nahori M-AMA, Vaganay E, Jouvion G, et al. Characterization of a Serine/Threonine Kinase Involved in Virulence of Staphylococcus aureus. *J Bacteriol*. 2009;191(13):4070-4081. doi:10.1128/JB.01813-08

48. Beltramini AM, Mukhopadhyay CD, Pancholi V. Modulation of cell wall structure and antimicrobial susceptibility by a Staphylococcus aureus eukaryote-like serine/threonine kinase and phosphatase. *Infect Immun*. 2009;77(4):1406-1416. doi:10.1128/IAI.01499-08

49. Miyazaki E, Chen JM, Ko C, Bishai WR. The Staphylococcus aureus rsbW (orf159) gene encodes an anti-sigma factor of SigB. *J Bacteriol*. 1999;181(9):2846-2851. http://www.ncbi.nlm.nih.gov/pubmed/10217777

50. Colaert N, Helsens K, Martens L, Vandekerckhove J, Gevaert K. Improved visualization of protein consensus sequences by iceLogo. *Nat Methods*. 2009;6(11):786-787. doi:10.1038/nmeth1109-786

51. O'Shea JP, Chou MF, Quader SA, Ryan JK, Church GM, Schwartz D. PLogo: A probabilistic approach to visualizing sequence motifs. *Nat Methods*. 2013;10(12):1211-1212. doi:10.1038/nmeth.2646

52. Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, Noble WS. MEME Suite: Tools for motif discovery and searching. *Nucleic Acids Res*. 2009;37(SUPPL. 2):202-208. doi:10.1093/nar/gkp335

53. El-Gebali S, Mistry J, Bateman A, Eddy SR, Luciani A, Potter SC, Qureshi M, Richardson LJ, Salazar GA, Smart A, Sonnhammer ELL, Hirsh L, Paladin L, Piovesan D, Tosatto SCE, et al. The Pfam protein families database in 2019. *Nucleic Acids Res*. 2019;47(D1):D427-D432. doi:10.1093/nar/gky995

Supplementary information

Further supplementary information can be found online: doi: 10.1074/mcp.RA120.002232

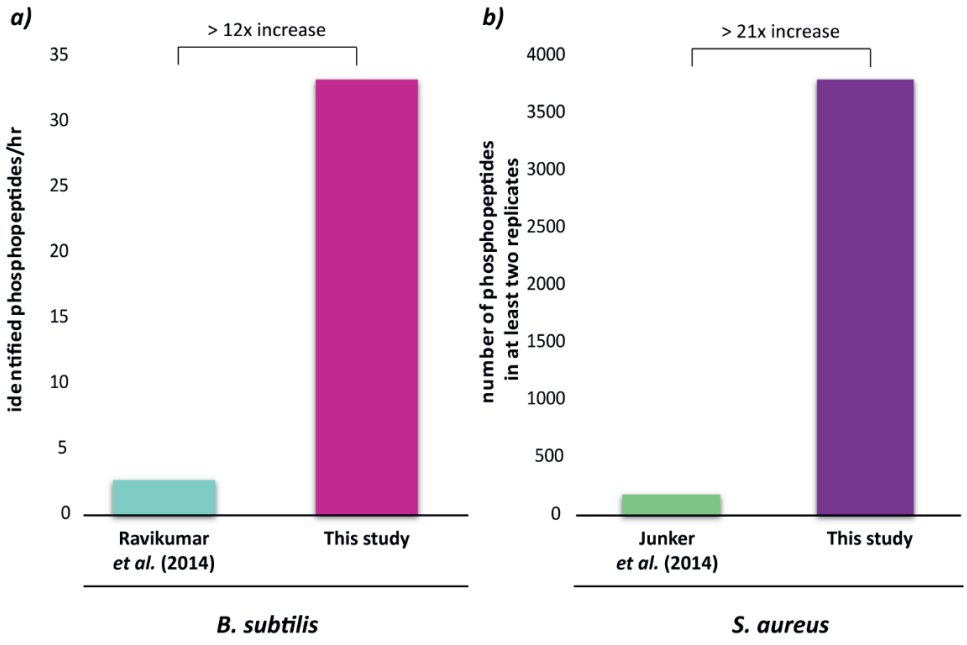
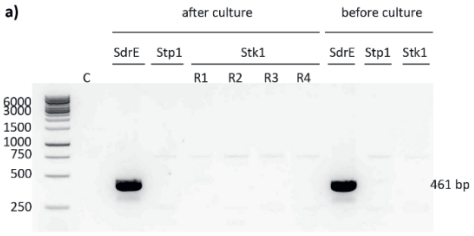
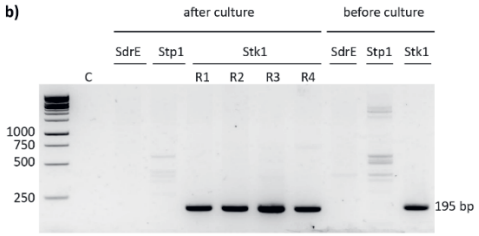
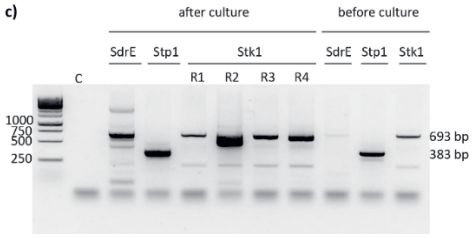


Figure S1: Literature comparison of the number of identified phosphopeptides. **a)** shows the number phosphopeptides per hour for *B. subtilis* identified by Ravikumar et al. compared to this study after applying the three-step lysis. Ravikumar et al. used 67 raw files resulting in 145.7 hrs run time in total[29]. In this study we used 3 raw files resulting in 8.75 hrs total run time. **b)** shows the number of phosphopeptides for *S. aureus* identified in at least two replicates by Junker et al. compared to this study (n=4). Junker *et al.* used a combination of SCX and TiO_2 whereas in this study no fractionation and Fe^{3+} enrichment was applied[30].







d)

Primer number	Base pair sequence	Location
2361	agtgttgcaccaaaacgtgt	SdrE_Fw1
2365	ggtggcatgagtaccgttta	PknB/Stk1_Fw1
2369	gggtacaacatgtgtttgtgca	Stp1/ PrpC_Fw1
2373	gctttttctaaatgtttttaagtaaatcaagtacc	Transposon (Buster)

Figure S2: PCR analysis to verify the transposons within the three different mutants (SdrE, Stk1 and Stp1). For all three mutants a sample before and after the liquid culture was taken to confirm the presence of the transposon insert and thus the disruption of the corresponding gene. For Stk1 all four replicates after the culture were tested. **a)** A forward primer within the sdrE gene (primer number 2361) and a reverse primer within the transposon (primer number 2373) was used. This should result in a fragment of 461 bp as seen for both SdrE samples. **b)** A forward primer within the stk1 gene (primer number 2365) and a reverse primer within the transposon was used (primer number 2373). This should result in a fragment of 195 bp as seen for all Stk1 samples. **c)** A forward primer within the stp1 gene (primer number 269) and a reverse primer within the transposon (primer number 2373) was used. This should result in a fragment of 383 bp as seen for both Stp1 samples. Due to the proximity of stk1 and stp1 the PCR product can also result in a 693 bp fragment if the tn is inserted in stk1, as seen in all Stk1 samples. **d)** List of used primers and base pair sequences.

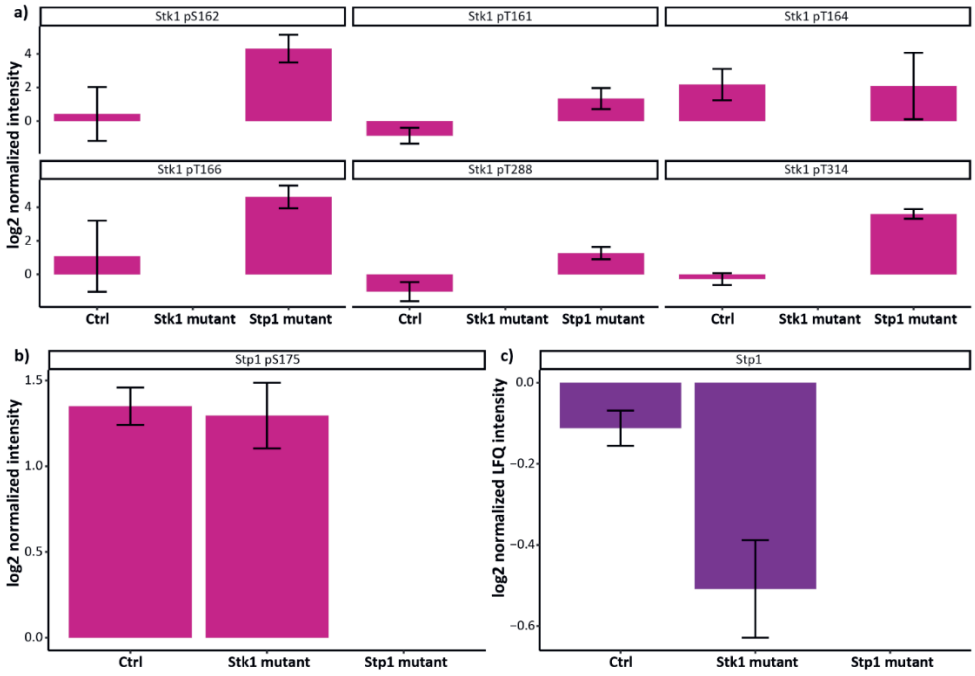
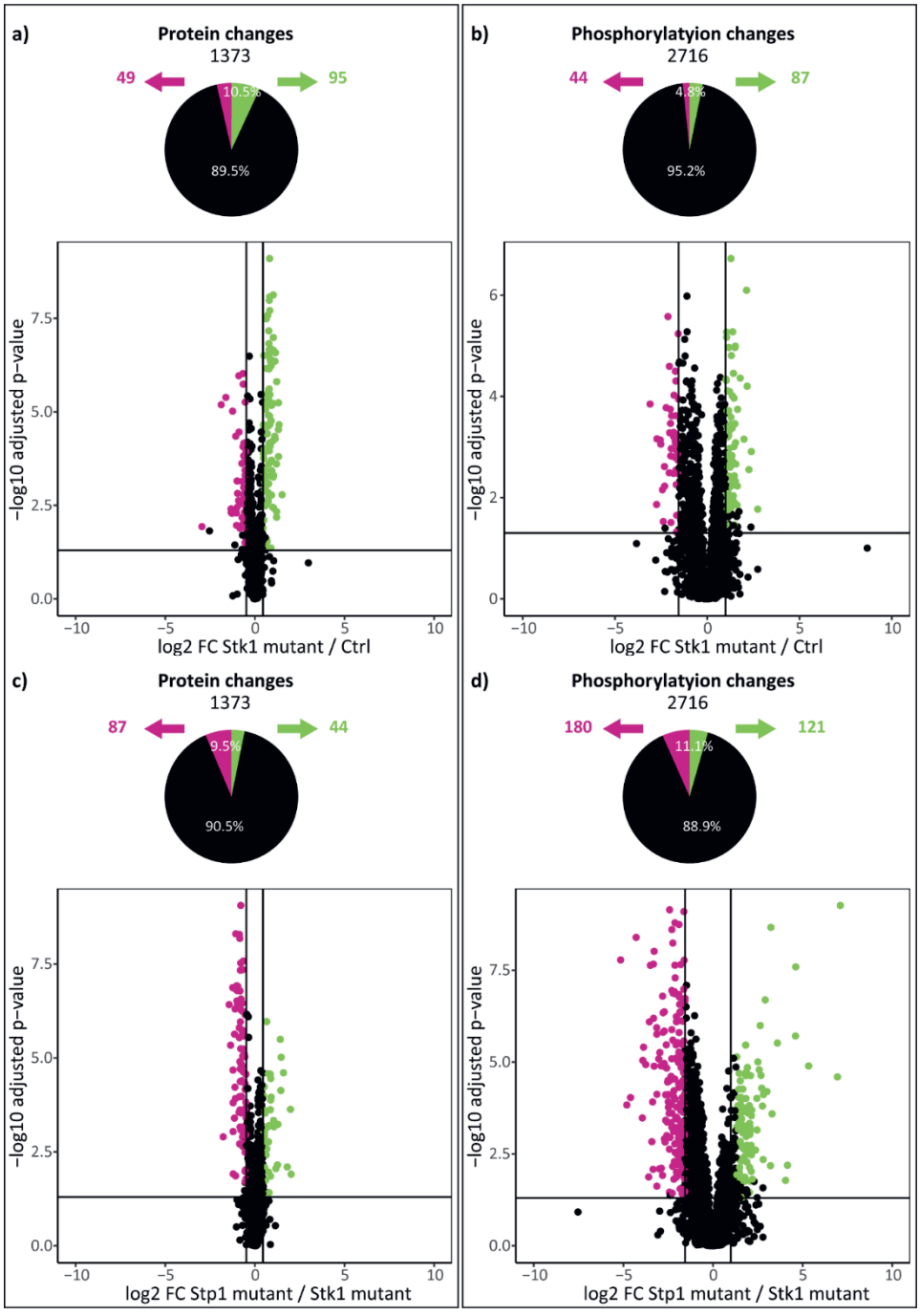


Figure S3: Intensity indicating the presence or absence of Stk1 and Stp1 in the respective mutants. **a)** Log2 normalized intensity of all identified phosphosites on Stk1. All phosphosites were exclusively identified in the control or the Stp1 mutant. **b)** log2 normalized intensity of the only identified phosphosite of Stp1, S175. This site was exclusively identified in the control and the Stk1 mutant. **c)** log2 normalized LFQ intensity of Stp1. Stp1 was exclusively identified in the control and the Stk1 mutant. For all bar charts the center values represent the mean and error bars the s.d. for n=4.

Figure S4: Identified protein and phosphorylation changes. **a)** The volcano plot for the comparison of the Stk1 mutant compared to the Ctrl showed that 144 quantified proteins (found in at least three replicates in one group and in two replicates in the other group) are significantly changing (Tukey HSD p-value cut-off of 0.05 and a fold change cut-off of $\bar{x} \pm \sigma$ of the data). 49 were under-represented (pink dots) and 95 were over-represented (green dots). **b)** The volcano plot shows that 131 localized phosphorylation sites (found in at least three replicates in one group and in two replicates in the other group) were significantly changing (Tukey HSD p-value cut-off of 0.05 and a fold change cut-off of $\bar{x} \pm \sigma$ of the data). 44 were under-represented (pink dots) and 87 were over-represented (green dots). **c)** The volcano plot for the comparison of the Stp1 mutant compared to the Stk1 mutant showed that 131 quantified proteins (found in at least three replicates in one group and in two replicates in the other group) are significantly changing (Tukey HSD p-value cut-off of 0.05 and a fold change cut-off of $\bar{x} \pm \sigma$ of the data). 87 were under-represented (pink dots) and 44 were over-represented (green dots). **d)** The volcano plot shows that 301 localized phosphorylation sites (found in at least three replicates in one group and in two replicates in the other group) were significantly changing (Tukey HSD p-value cut-off of 0.05 and a fold change cut-off of $\bar{x} \pm \sigma$ of the data). 180 were under-represented (pink dots) and 121 were over-represented (green dots).


a)
Protein changes
1373
49
95
10.5%
89.5%
−log10 adjusted p−value
log2 FC Stk1 mutant / Ctrl
b)
Phosphorylatyion changes
2716
44
87
4.8%
95.2%
−log10 adjusted p−value
log2 FC Stk1 mutant / Ctrl
c)
Protein changes
1373
87
44
9.5%
90.5%
−log10 adjusted p−value
log2 FC Stp1 mutant / Stk1 mutant
d)
Phosphorylatyion changes
2716
180
121
11.1%
88.9%
−log10 adjusted p−value
log2 FC Stp1 mutant / Stk1 mutant

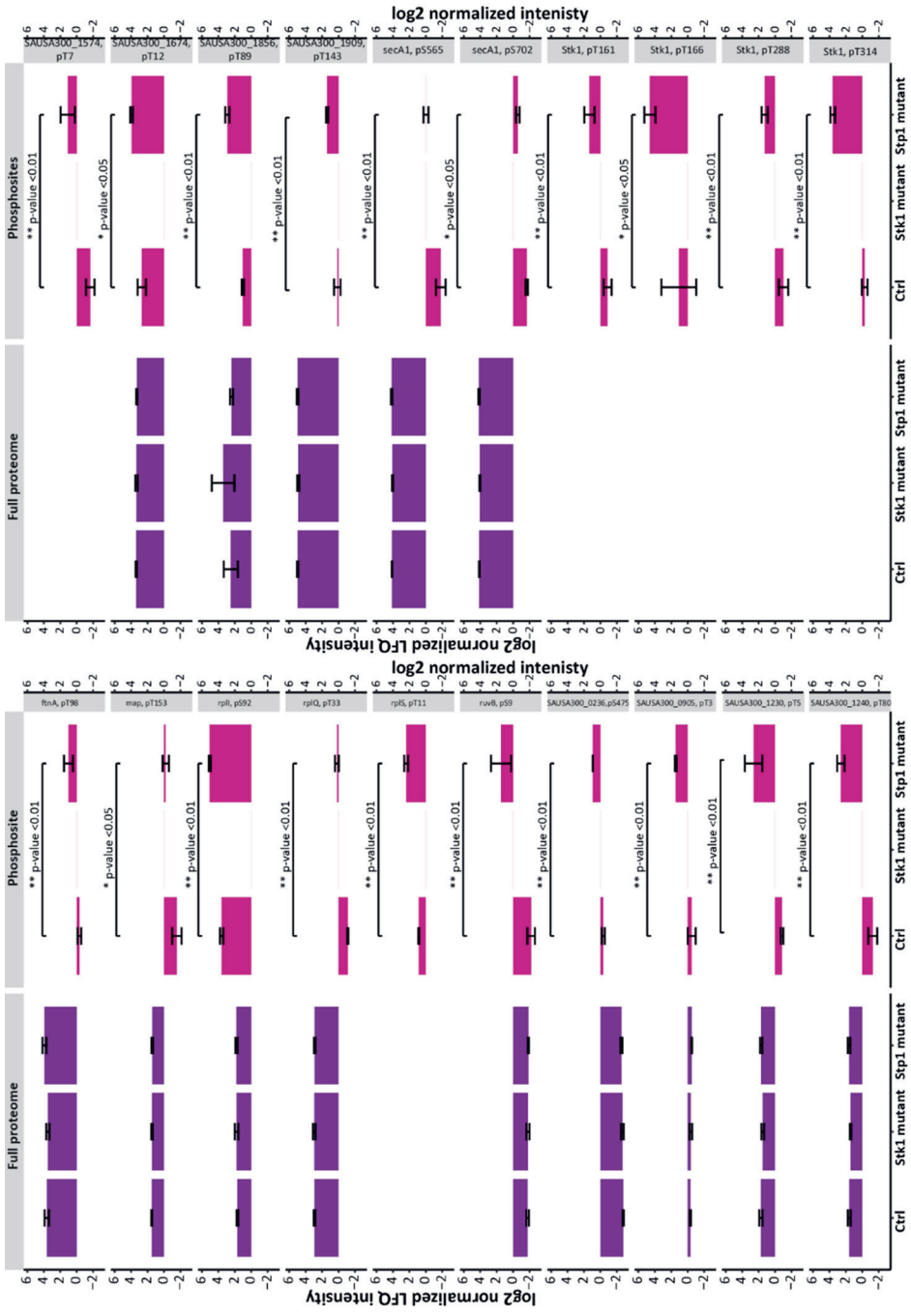


Figure S5: Intensity comparison of all 20 phosphosites significantly overrepresented in the Stp1 mutant compared to the Ctrl. Shown is the log2 normalized LFQ intensity for the protein (left, full proteome) and the log2 normalized intensity for the phosphosite (right, phosphosite). P-value significance is indicated with * (Tukey HSD p-value < 0.05) or ** (Tukey HSD p-value < 0.05 <0.01). The proteins rplS, SAUSA300_1574 and Stk1 could not be identified on the full proteome. For all bar charts the center values represent the mean and error bars the s.d. for n=4.

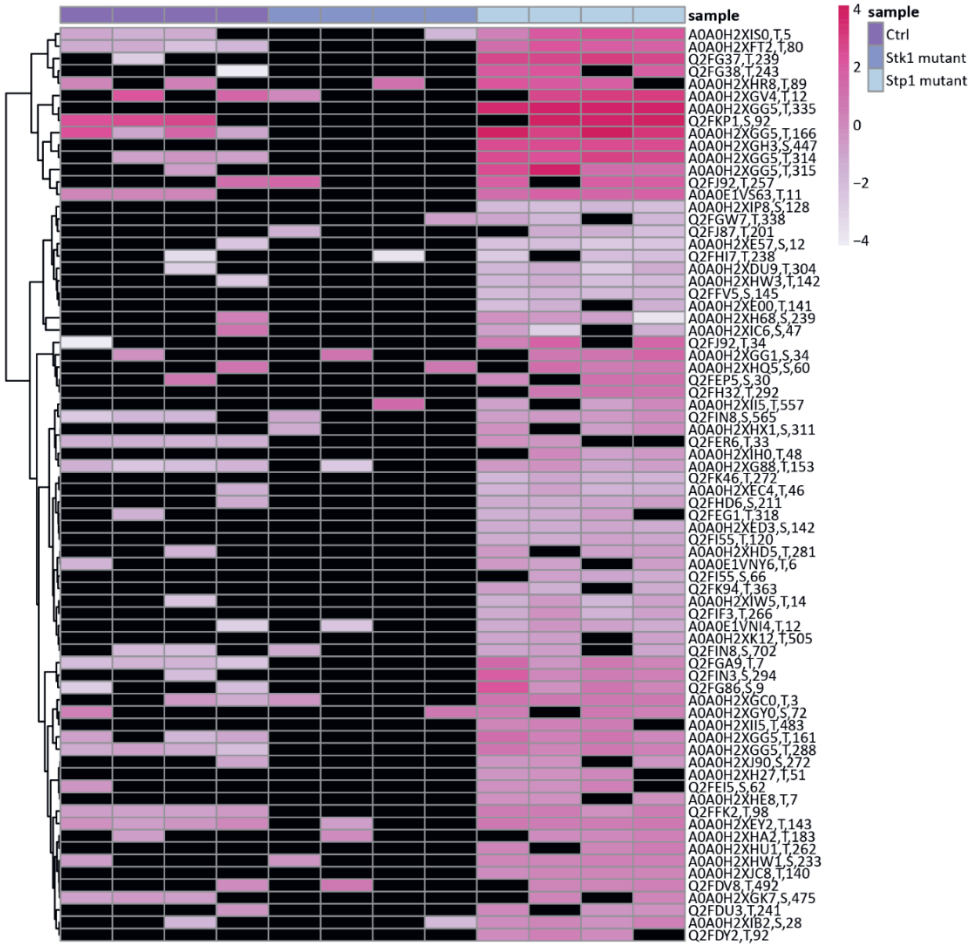
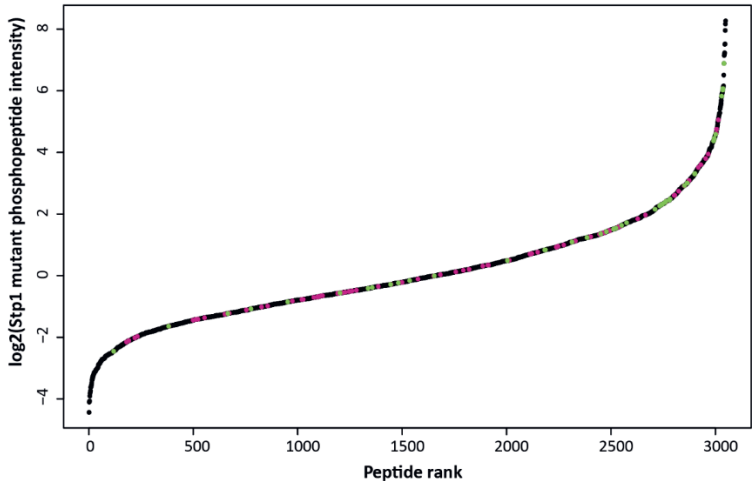


Figure S6: Heat map of all 74 potential Stk1/Stp1 target sites. Shown is the log2 normalized intensity of the phosphosites, NA values are shown in black. K-means clustering was used within the pheatmap package (vers. 1.0.12) to cluster phosphosites.

a)



b)

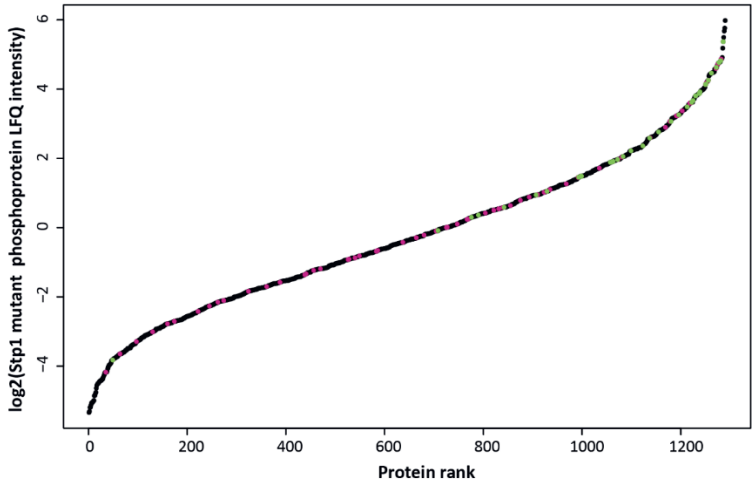


Figure S7: Dynamic range of identified phosphopeptides and phosphoproteins in the Stp1 mutant. **a)** The average log2 phosphopeptide intensity of all phosphopeptides identified in the Stp1 mutant (n=4) plotted against their peptide rank (black). Identified Stk1/Stp1 targets are highlighted in pink (74 phosphosites). Peptides that were also identified in the study by Junker *et al*. are highlighted in green (42 phosphosites)[1]. **b)** The average log2 LFQ intensity of all phosphoproteins identified in the Stp1 mutant (n=4) plotted against their protein rank (black). Identified Stk1/Stp1 targets are highlighted in pink (65 phosphoproteins). Proteins that were also identified in the Study by Junker et al. are highlighted in green (41 phosphoproteins)[1].

[1]Junker S, Maaβ S, Otto A, et al. Spectral Library Based Analysis of Arginine Phosphorylations in Staphylococcus Aureus. Vol 17.; 2018. doi:10.1074/mcp.RA117.000378

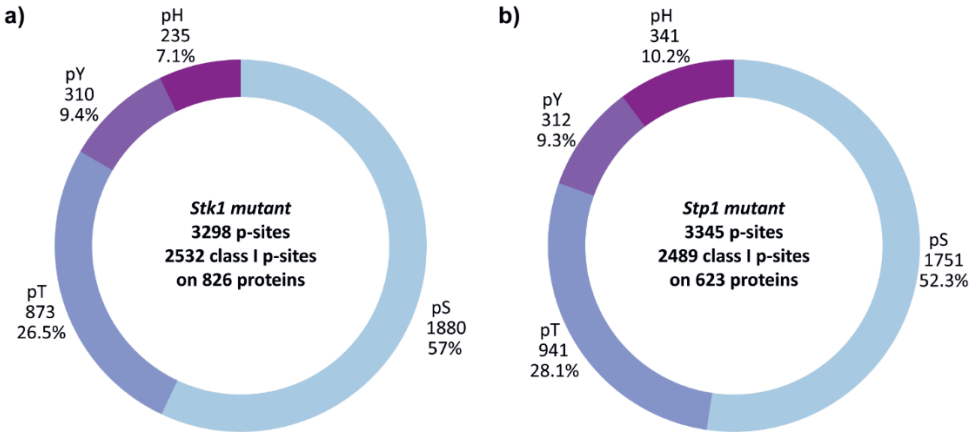


Figure S8: S. Number of identified phosphosites, class I phosphosites (Andromeda localization probability > 0.75) and phosphoproteins as well as the distribution of serine, threonine, tyrosine and histidine phosphosites identified for **a)** Stk1 mutant and **b)** Stp1 mutant

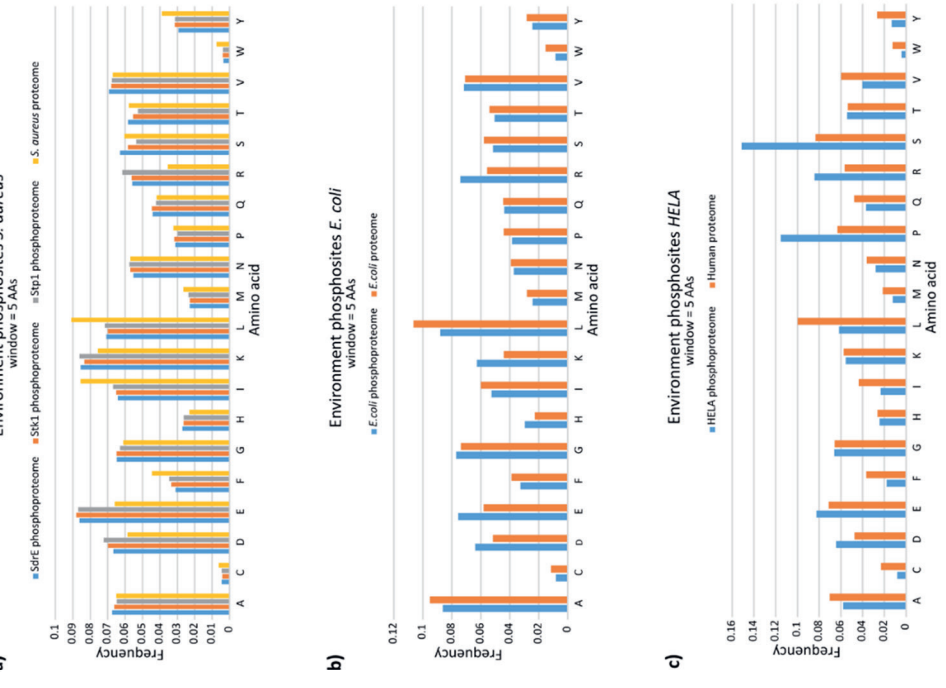

a)
Environment phosphosites S. aureus
window = 5 AAs
SdrE phosphoproteome
Stk1 phosphoproteome
Stp1 phosphoproteome
S. aureus proteome
Frequency
Amino acid
b)
Environment phosphosites E. coli
window = 5 AAs
E.coli phosphoproteome
E.coli proteome
Frequency
Amino acid
c)
Environment phosphosites HELA
window = 5 AAs
HELA phosphoproteome
Human proteome
Frequency
Amino acid
d)
Phosphosite environment S. aureus
window = 5 AAs
SdrE phosphoproteome
Stk1 phosphoproteome
Stp1 phosphoproteome
S. aureus proteome
Frequency
flexible (A/G/P)
acidic (D/E)
basic (R/K)
aromatic (W/F/Y)
other
e)
Phosphosite environment E. coli
window = 5 AAs
E. coli phosphoproteome
E. coli proteome
Frequency
flexible (A/G/P)
acidic (D/E)
basic (R/K)
aromatic (W/F/Y)
other
f)
Phosphosite environment HELA
window = 5 AAs
HELA phosphoproteome
Human proteome
Frequency
flexible (A/G/P)
acidic (D/E)
basic (R/K)
aromatic (W/F/Y)
other

Figure S9: Phosphosite environment for *S. aureus, E. coli* and HeLa. The frequency of all 20 common amino acids (AA) adjacent (+/- 5 AA) to a phosphosite was plotted and compared to the general frequency in **a)** *S. aureus* **b)** *E.coli* and **c)** HeLa protome. The AA were grouped in flexible, acidic, basic, aromatic and other and the summed frequency was plotted for **d)** *S. aureus* **e)** *E. coli* and **f)** HeLa

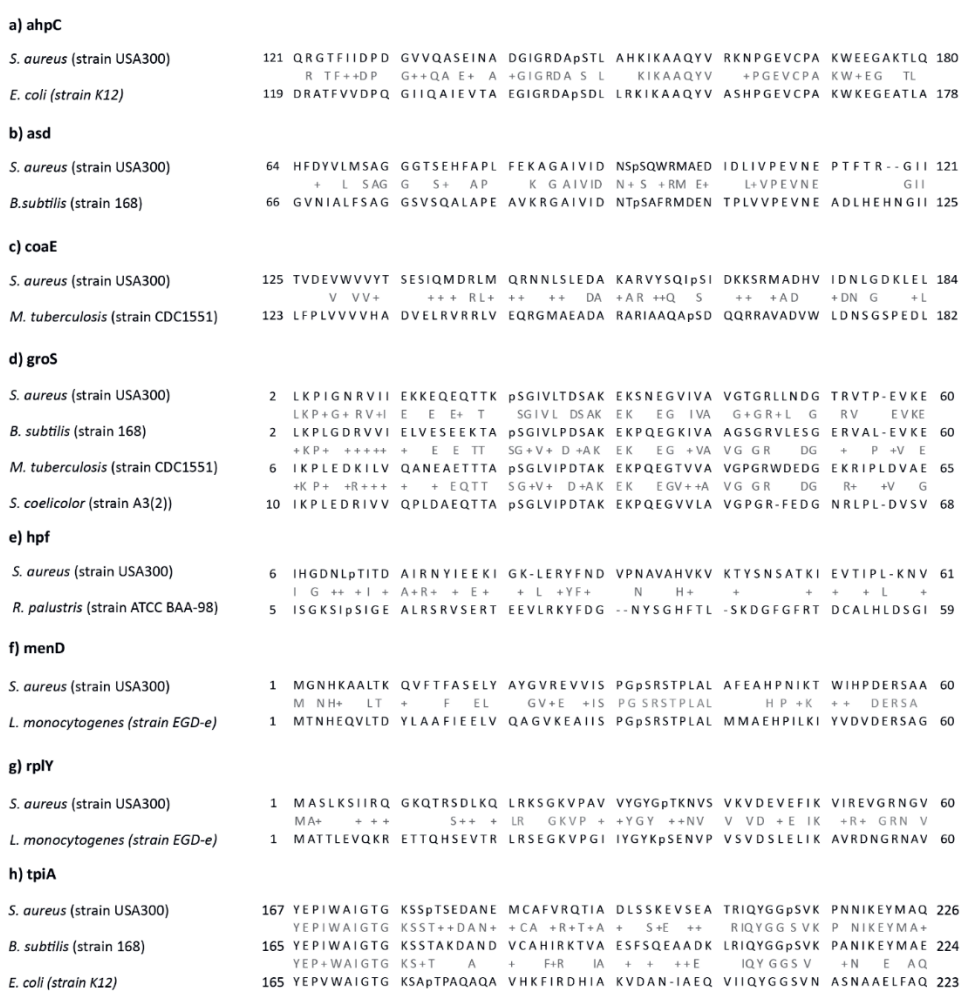
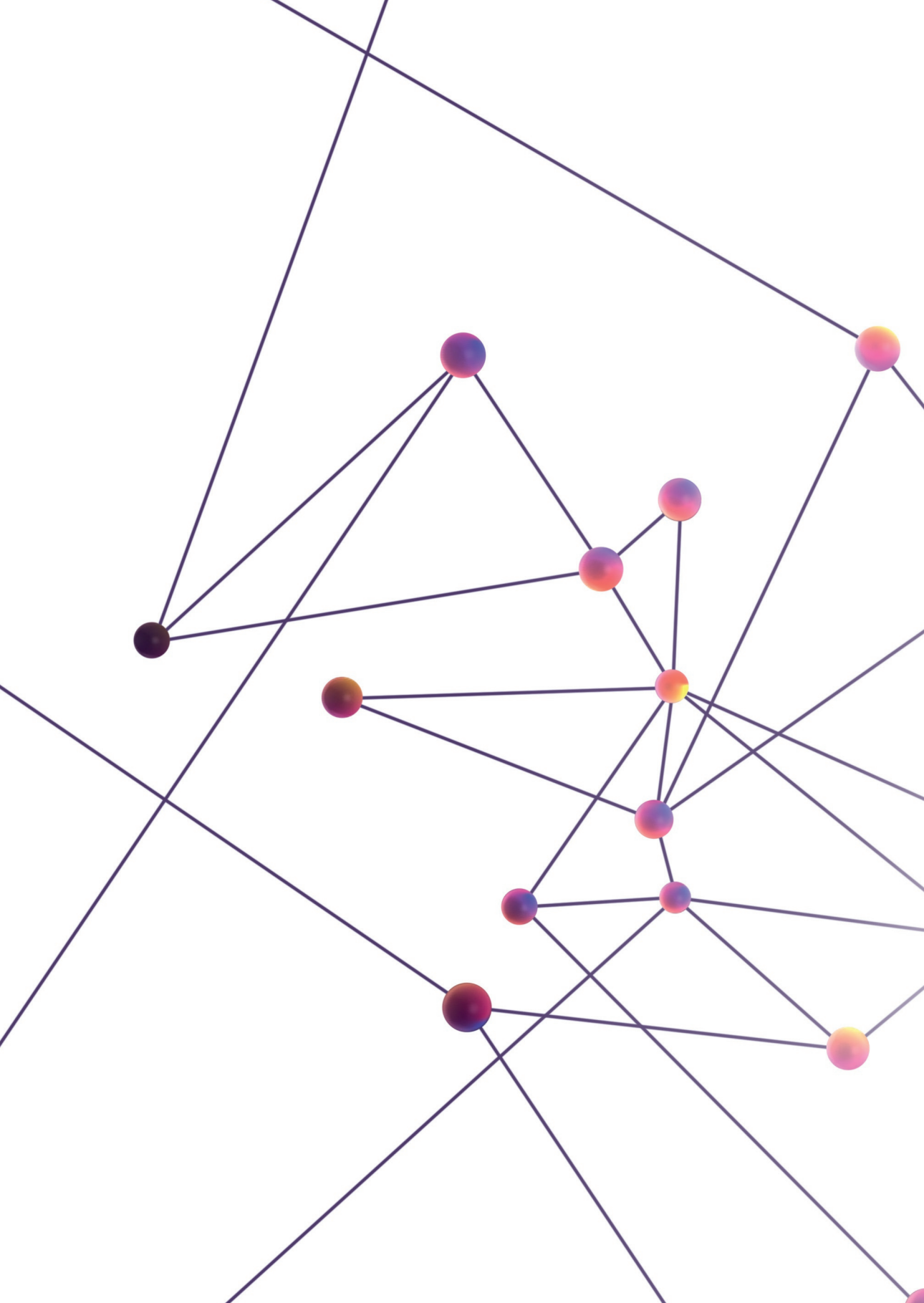


Figure S10: Sequence alignments obtained with NCBI blastp (https://blast.ncbi.nlm.nih.gov/Blast.cgi) for proteins with conserved phosphosites identified with dbPSP[3] **a)** ahpC **b)** asd, **c)** coaE, **d)** groS **e)** hpf, **f)** menD, **g)** rplY, **h)** tpiA. Phosphosites are indicated as pS or pT.

[3]Pan Z, Wang B, Zhang Y, et al. dbPSP: A curated database for protein phosphorylation sites in prokaryotes. Database. 2015;2015(August 2018):1-8. doi:10.1093/database/bav03



Chapter 3

Widespread arginine phosphorylation in *Staphylococcus aureus*

Nadine Prust[1,2], Pieter C. van Breugel[1,2] and Simone Lemeer[1,2#]

[1] Biomolecular Mass Spectrometry and Proteomics, Bijvoet Center for Biomolecular Research and Utrecht Institute for Pharmaceutical Sciences, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands
[2] Netherlands Proteomics Center, Padualaan 8, 3584 CH Utrecht, The Netherlands

Submitted and under revision in Molecular & Cellular Proteomics

Abbreviations

AmBiC	Ammonium bicarbonate
Arg	Arginine
Asp	Aspartic acid
ETD	Electron transfer dissociation
FA	Formic acid
FDR	False discovery rate
Glu	Glutamic acid
HCD	Higher energy collision induced dissociation
His/H	Histidine
IMAC	Immobilized metal ion affinity chromatography
LC	Liquid chromatography
LMW-PTP	low-molecular-weight protein tyrosine phosphatase
Lys	Lysine
MnCl2	Manganese-II-chloride
PCR	Polymerase chain reaction
Ser/S	Serine
Thr/T	Threonine
Tyr/Y	Tyrosine

Abstract

Arginine phosphorylation was only recently discovered to play a significant and relevant role in the Gram-positive bacterium *B. subtilis*. In addition, arginine phosphorylation was also detected in *S. aureus*, suggesting a widespread role in bacteria. However, the large-scale analysis of protein phosphorylation and especially those which involve a phosphoramidate bond, comes along with several challenges. The sub-stoichiometric nature of protein phosphorylation requires proper enrichment strategies prior to LC-MS/MS analysis and the acid instability of phosphoramidates was long thought to impede those enrichments. Furthermore, good spectral quality is required which can be impeded by the presence of neutral losses of phosphoric acid upon higher energy collision induced dissociation. Here we show that pArg is stable enough for commonly used Fe^{3+}-IMAC enrichment followed by LC-MS/MS and that HCD is still the gold standard for the analysis of phosphopeptides. By profiling a serine/threonine kinase (Stk1) and phosphatase (Stp1) mutant from a methicillin-resistant *S. aureus* mutant library, we identified 1,062 pArg sites and thus the most comprehensive arginine phosphoproteome to date. Using synthetic arginine phosphorylated peptides we validated the presence and localization of arginine phosphorylation in *S. aureus*. Finally, we could show that the knockdown of Stp1 significantly increases the overall amount of arginine phosphorylation in *S. aureus*. However, our analysis also shows that Stp1 is not a direct protein-arginine phosphatase, but only indirectly influences the arginine phosphoproteome.

Introduction

Protein phosphorylation is one of the most important reversible post-translational modifications (PTMs), regulating a magnitude of cellular processes[1,2]. Hereto, a variety of protein kinases enable the transfer of the γ-phosphate of ATP to specific amino acid residues in the acceptor protein[3]. The negative charge of the phosphoryl group can induce conformational changes that enable protein-protein interactions, changes in protein activity or subcellular localizations[4]. Phosphatases can counteract this reaction by catalyzing the dephosphorylation of the proteins and thus allowing for a perfectly balanced on/off regulation of specific cellular processes. Whereas eukaryotes mainly utilize phosphorylation on hydroxyl groups to form phosphomonoester (pSer, pThr and pTyr)[5] prokaryotes largely also exploit phosphorylation on amide groups to form phosphoramidates (pHis and pArg)[6,7]. Phosphoramidate bonds, contrary to phosphomonoester bonds, are rather unstable under acidic conditions[8,9]. Since most enrichment and fractionation strategies, prior to mass spectrometry based analysis, are performed under strong acidic conditions those modifications remained long unidentified. This despite the notion that phosphoramidates, such as histidine phosphorylation, are known to fulfill important physiological roles, especially in bacteria and lower eukaryotes[6,10]. Recent studies on protein-histidine phosphorylation however showed that it is stable enough for commonly applied Fe^{3+}-IMAC enrichments[11,12].

Protein-histidine phosphorylation is known to be involved in two-component system (TCS) signal transduction systems and plays a major role in prokaryotic signal transduction[6]. TCS consist of a receptor histidine kinase sensing the outside environment and a corresponding response regulator, which mediates the intracellular response[6,13]. The autophosphorylated receptor histidine kinases, transfers the phosphate group to an aspartic residue in the response regulator (RR). Phosphorylation of the RR induces conformational changes that can trigger further downstream processes by e.g. binding to DNA[13]. To date more than 16 histidine kinases are known for the Gram-positive bacterium *S. aureus*[13] and at least one of those TCS systems has been shown to crosstalk with reversible phosphorylation via the kinase Stk1 and phosphatase Stp1[14].

The other phosphoramidate, protein arginine phosphorylation, only received attention of the scientific community over the last decade after the identification, of the first and yet only, protein-arginine kinase, McsB, in the Gram-positive bacterium *B. subtilis*[7]. McsB shows homology to phosphagen kinases and was previously reported to act as tyrosine kinase[15,16]. The phosphatase YwlE was identified as tyrosine phosphatase and classified as low-molecular-weight protein tyrosine phosphatase (LMW-PTP) based on its amino acid sequence[16–18]. However, Elsholz *et al*. could show that YwlE acts as protein-arginine phosphatase and thus counteracts the protein-arginine kinase activity of McsB[19].

Adaptations of common enrichments strategies to less acidic conditions[20,21] or using a phosphatase trap mutant[22] enabled first insights in the arginine phosphoproteome of Gram-positive bacteria, resulting in the identification of around 200 pArg sites for *B. subtilis*[8,19] and *S. aureus*[20,21]. Arginine phosphorylation was shown to play a functional role in the stress response, as well as in degradation pathways by marking proteins for degradation within the ClpC-ClpP proteasome, such as the transcriptional repressor CtsR[7,19,23,24]. Similar to histidine phosphorylation, reports of arginine phosphorylation in eukaryotes are scarce. A few studies reported the presences of pArg in rat liver[25] or mouse leukemia cells[26], however a protein-arginine kinase has not been identified yet.

In addition to suitable enrichment strategies, the right fragmentation and automated spectrum interpretation are crucial for the identification and localization of arginine phosphorylated peptides. Higher energy collision induced dissociation (HCD) fragmentation is known to generate extensive neutral loss fragments and therefore hampers the exact phosphosite localization by generating "unmodified" peptides. Electron transfer dissociation (ETD) in contrast, is preserving labile phosphosites and thus allows an unambiguous phosphosite localization[27,28]. ETD fragmentation has especially been shown to be beneficial for preservation of the instable phosphoramidate bond of pArg during fragmentation [29].

Using an optimized sample preparation and subsequent Fe^{3+}-IMAC enrichment we could show extensive protein-arginine phosphorylation in *S. aureus,* being as prevalent as threonine phosphorylation. Knockdown of the serine/threonine phosphatase Stp1 seemed to increase the amount of arginine phosphorylated proteins. Using synthetic peptides, we confidentially validated the presence of pArg phosphorylation and concomitantly analyzed different fragmentation methods for confident localization of arginine phosphorylation. Finally, using synthetic peptides and purified recombinant Stp1, we could convincingly show that Stp1 is not a direct arginine phosphatase but rather has a secondary effect on the arginine phosphoproteome.

Material & Methods

Bacterial culture

B. subtilis 168 was grown overnight in 50 ml Luria broth (LB) at 37°C with agitation in n=3 biological replicates. Transposon mutants NE98 (disruption in unrelated surface protein encoding gene *sdrE*), NE217 (disruption in protein coding gene *pknB*) and NE1919 (disruption in gene SAUSA300_1112 encoding Stp1), all containing an erythromycin resistance marker, of the *S. aureus* USA300 JE2 strain were obtained from the Nebraska Transposon Mutant Library (NTML)[30]. Mutants were grown in n=4 biological replicates, overnight in 25 ml Todd Hewitt broth (THB) supplemented with 5 ug/ml erythromycin at

37°C with agitation as described published[11,31]. Bacteria were harvested by centrifugation (15 min, 3,200 rpm at 4°C) and the supernatant was subsequently removed.

Optimized Cell lysis

Bacterial cell lysis was performed as described previously[31]. One volume bacteria pellet was resuspended in five volume of lysis buffer (100 mM Tris-HCl pH 8.5, 7 M Urea, 5 mM tris(2-carboxyethyl)phosphine (TCEP), 30 mM 2-Chloroacetamide (CAA), 10 U/ml DNase I, 1 mM magnesium chloride (Sigma-Aldrich, Steinheim, Germany), 1% (v/v) benzonase (Merck Millipore, Darmstadt, Germany), 1 mM sodium orthovanadate, phosphoSTOP phosphatases inhibitors (Roche) and complete mini EDTA free protease inhibitors). The lysis was performed by bead beating for 17.5 min (1.5 min on, 2 min off) at 2,850 rpm (Disruptor Genie, Scientific industries) in case of *B. subtilis* and 3,200 rpm (Mini-Beadbeater-24, Bio Spec Products Inc.) for *S. aureus*. Subsequently, the beads were pelleted by centrifugation (2 min at 3000 rpm) and 1% (v/v) Triton X-100 in case of *B. subtilis* and 1 % (v/v) Triton X-100 plus 1% (v/v) sodium deoxycholate (SDC, final concentration) in case of *S. aureus* were added to the bacteria lysate. Complete lysis was reached by sonication for 45 min (20 s ON, 40 s off) using a Bioruptor Plus. Cell debris was removed by ultracentrifugation (45,000 rpm for 1 hr at 4°C). Protein concentration of the supernatant was determined via a Bicinchonic Acid (BCA) assay. To decrease the SDC concentration to <0.4 %, the supernatant was diluted 2.5 times with dilution buffer (100 mM Tris-HCl pH 8.5, 7 M Urea, 5 mM TCEP, 30 mM CAA, 1 mM magnesium chloride (Sigma-Aldrich, Steinheim, Germany), 1 mM sodium orthovanadate, phosphoSTOP phosphatases inhibitors (Roche) and complete mini EDTA free protease inhibitors). 1 % (v/v) benzonase was added to the supernatant mixture and incubated for 2 h at room temperature. Subsequently, methanol/chloroform precipitation was performed as described earlier[11]. The precipitate was then resuspended in digestion buffer (100 mM Tris-HCL pH 8.5, 30 M CAA, 1% (v/v) SDC (Sigma-Aldrich) and 5 mM TCEP). Protein digestion was performed overnight at room temperature using a mix of trypsin and Lys-C in a ratio of 1:25 and 1:100 (w/w), respectively. Protein digests were acidified to pH 3.5 using 10% formic acid (Sigma-Aldrich) and precipitated SDC was removed by centrifugation (1,400 rpm, 5 min). The supernatant was loaded onto C18 Sep-Pak (3cc) resin columns (Waters) for desalting. The loaded samples were washed twice with 0.1 % (v/v) formic acid and bound peptides were eluted with 600 ul 40 % acetonitrile and 0.06 % formic acid. Eluted peptides were split into 2 mg fractions and samples for full proteome analysis were frozen in liquid nitrogen and freeze dried.

Phosphopeptide enrichment

Fe^{3+}-IMAC enrichments were performed as previously described[11]. In short, 2 mg lyophilized peptides were resuspended in loading buffer A (30% acetonitrile and 0.07% TFA) and, if necessary, the pH was adjusted to 2.3 using 10% TFA. The samples were loaded onto the

Fe^{3+}-IMAC column (Propac IMAC-10 4 x 5 mm column, Thermo Fischer Scientific). Bound phosphopeptides were eluted with elution buffer B (0.3 % NH_4OH). The respective gradient is described in supplementary Table S1. The UV-abs signal at a wavelength of 280 nm was recoded at the outlet of the column and eluting phosphopeptides were collected manually. Subsequently, phosphopeptides were frozen in liquid nitrogen and freeze dried.

LC-MS/MS

Nanoflow LC-MS/MS analysis was performed using an Agilent 1290 (Agilent technologies, Middelburg, The Netherlands) coupled to an Orbitrap Q-Exactive HF-X (Thermo Fisher Scientific, Bremen, Germany). Lyophilized phosphopeptides or full proteome samples were resuspended in 20 mM citric acid (Sigma-Aldrich), 1 % (v/v) formic acid or 2 % (v/v) formic acid, respectively. Resuspended phosphopeptide, corresponding to 1.6 mg or 200 ng full proteome samples were injected, trapped and washed on a trap-column (100 µm i.d. × 2 cm, packed with 3 µm C18 resin, Reprosil PUR AQ, Dr. Maisch, packed in-house) for 5 min at a flow rate of 5 µL/minute with 100 % buffer A (0.1 FA, in HPLC grade water). Peptides were subsequently transferred onto an analytical column (75 µm x 60 cm Poroshell 120 EC-C18, 2.7 µm, Agilent Technology, packed in-house) and separated at room temperature at a flow rate of 300 nL/min using a 85 min linear gradient from 8 % to 32 % buffer B (0.1 % FA, 80 % ACN) or a 115 min linear gradient from13 % to 44 % buffer B. Electrospray ionization was performed using 1.9 kV spray voltage and a capillary temperature of 320 °C. The mass spectrometer was operated in data-dependent acquisition mode: full scan MS spectra (*m/z* 375 – 1,600) were acquired in the Orbitrap at 60,000 resolution for a maximum injection time of 20 ms with an AGC target value of 3e6 charges. Up to 12 precursors for phosphoproteome samples and up to 15 precursors for full proteome samples were selected for subsequent fragmentation. High resolution HCD MS2 spectra were generated using a normalized collision energy of 27 %. The intensity threshold to trigger MS2 spectra was set to 2e5, and the dynamic exclusion to 12 or 16, respectively. MS2 scans were acquired in the Orbitrap mass analyzer at a resolution of 30,000 (isolation window of 1.4 Th) with an AGC target value of 1e5 charges and a maximum ion injection time of 50 ms. Precursor ions with unassigned charge state as well as charge state of 1+ or superior/equal to 6+ were excluded from fragmentation.

E. coli samples

E. coli samples published by Potel and co-workers[11] were reanalyzed for pArg. Detailed sample preparation can be found in the respective publication. In short, *E. coli* strain W3110 was used and grown in M9 minimal medium, consisting of M9 salts (6 g/L Na_2HPO_4, 3 g/L KH_2PO_4, 0.5 g/L NaCl, 1 g/L NH_4Cl) supplemented with additional 0.5% (w/v) glucose, 1 mM $MgSO_4$, 0.1 mM $CaCl_2$, with vigorous shaking at 37°C. Cells were collected by centrifugation at stationary phase (OD600 = 1.2) and washed three times with ice-cold PBS. Sample were lysed as described above under optimized sample preparation, however without the addition of SDC and additional bead beating.

Synthetic peptides

Synthetic arginine peptide analysis

Seventeen synthetic peptides phosphorylated on arginine were ordered from Pepscan (Lelystad, the Netherlands, see supplementary Table S2). Counterparts phosphorylated on Ser, Thr or Tyr were ordered from JPT Peptide Technologies GmbH, Berlin, Germany (supplementary Table S2). Peptides were reconstituted in 0.1 M AmBiC, 20% acetonitrile and diluted to 400 fmol/ul. A peptide mix containing all 17 pArg-peptides was prepared with adjusted concentrations for best peptide identification. All 45 pSTY peptides were mixed in a 1:1 ratio of 400 fmol per peptide.

Stability test of synthetic pArg-peptides

Synthetic peptides (n = 4) were constituted in 52.5 ul 20 mM citric acid (Sigma-Aldrich), 2% (v/v) FA and 250 ng *E. coli* digest was spiked in. The peptide mix was spun down for 3 min at 2,0817 x g, 4°C and loaded on the plate 0 min, 15 min, 30 min, 60 min or 120 min prior to LC-MSMS analysis. Nanoflow LC-MS/MS analysis was performed using an Agilent 1290 (Agilent technologies, Middelburg, The Netherlands) coupled to an Orbitrap Fusion Lumos (Thermo Fisher Scientific, Bremen, Germany). Resuspended phosphopeptides, corresponding to 16 % of the peptide mix were injected, trapped and washed on a trap-column (100 µm i.d. × 2 cm, packed with 3 µm C18 resin, Reprosil PUR AQ, Dr. Maisch, packed in-house) for 5 min at a flow rate of 5 µL/minute with 100 % buffer A (0.1 FA, in HPLC grade water). Peptides were subsequently transferred onto an analytical column (75 µm x 60 cm Poroshell 120 EC-C18, 2.7 µm, Agilent Technology, packed in-house) and separated at room temperature at a flow rate of 300 nL/min using a 40 min linear gradient from 10 % to 40 % buffer B (0.1 % FA, 80 % ACN). Electrospray ionization was performed using 1.9 kV spray voltage and a capillary temperature of 320 °C. The mass spectrometer was operated in data-dependent acquisition mode: full scan MS spectra (*m/z* 375 – 1,500) were acquired in the Orbitrap at 60,000 resolution for a maximum injection time of 50 ms with an AGC target value of 4e5 charges. High resolution HCD MS2 spectra were generated

using a normalized collision energy of 35 %. The intensity threshold to trigger MS2 spectra was set to 5e5, and the dynamic exclusion 2. MS2 scans were acquired in the Orbitrap mass analyzer at a resolution of 30,000 (isolation window of 1.6 Th) with an AGC target value of 1e5 charges and a maximum ion injection time of 100 ms. Precursor ions with unassigned charge state as well as charge state of 1+ or superior/equal to 6+ were excluded from fragmentation.

Retention time analysis was performed using an Ultimate 3000 (Thermo Fisher Scientific, Bremen, Germany) coupled to an Orbitrap Exploris™ 480 (Thermo Fisher Scientific, Bremen, Germany). Resuspended phosphopeptides, corresponding to 16% of the pArg-peptide mix or 8% pSTY peptide mix both containing 4 ng/ul *E. coli* digest were injected, trapped and washed on a trap-column (µ-Precolumn, 300 µm i.d. x 5mmC18 PepMap100, 5 µm, 100 Å (Thermo Scientific, P/N 160454)) for 5 min at a flow rate of 5 µL/minute with 92 % buffer A (0.1 FA, in HPLC grade water). Peptides were subsequently transferred onto an analytical column (75 µm x 50 cm Poroshell 120 EC-C18, 2.7 µm, Agilent Technology, packed in-house) and separated at 40°C at a flow rate of 0.3 µl/min using a 40 min linear gradient from 9 % to 36% buffer B (0.1 % FA, 80 % ACN). Electrospray ionization was performed using 1.9 kV spray voltage and a capillary temperature of 275 °C. The mass spectrometer was operated in data-dependent acquisition mode: full scan MS spectra (*m/z* 375 – 1,600) were acquired in the Orbitrap at 60,000 resolution for a maximum injection time set to auto-mode with a standard AGC target. High resolution HCD MS2 spectra were generated using a normalized collision energy of 28%. The intensity threshold to trigger MS2 spectra was set to 5e4, and the dynamic exclusion 2. MS2 scans were acquired in the Orbitrap mass analyzer at a resolution of 30,000 (isolation window of 1.4 Th) with a normalized AGC target of 200% and an automatic maximum injection time. Precursor ions with unassigned charge state as well as charge state of 1+ or superior/equal to 6+ were excluded from fragmentation.

ETD, EThcD and HCD fragmentation

Nanoflow LC-MS/MS analysis was performed using an Agilent 1290 (Agilent technologies, Middelburg, The Netherlands) coupled to an Orbitrap Fusion (Thermo Fisher Scientific, Bremen, Germany). Resuspended phosphopeptide, corresponding to 80% of the peptide mix were injected, trapped and washed on a trap-column (100 µm i.d. × 2 cm, packed with 3 µm C18 resin, Reprosil PUR AQ, Dr. Maisch, packed in-house) for 5 min at a flow rate of 5 µL/minute with 100 % buffer A (0.1 FA, in HPLC grade water). Peptides were subsequently transferred onto an analytical column (75 µm x 50 cm Poroshell 120 EC-C18, 2.7 µm, Agilent Technology, packed in-house) and separated at room temperature at a flow rate of 300 nL/min using a 40 min linear gradient from 10 % to 40 % buffer B (0.1 % FA, 80 % ACN). Electrospray ionization was performed using 2 kV spray voltage and a capillary temperature of 275 °C. The mass spectrometer was operated in data-dependent acquisition mode: full

scan MS spectra (*m/z* 375 – 1,500) were acquired in the Orbitrap at 60,000 resolution for a maximum injection time of 50 ms with an AGC target value of 4e5 charges. High resolution EThcD spectra were generated using a supplemental activation collision energy of 35% and high resolution HCD spectra were generated using a normalized collision energy of 35 %. For all three fragmentation methods, the intensity threshold to trigger MS2 spectra was set to 5e5, and the dynamic exclusion 2. MS2 scans were acquired in the Orbitrap mass analyzer at a resolution of 30,000 (isolation window of 1.6 Th) with an AGC target value of 1e5 charges and a maximum ion injection time of 100 ms. Precursor ions with unassigned charge state as well as charge state of 1+ or superior/equal to 6+ were excluded from fragmentation.

Plasmid construction

pET28(+) eGFP, pET28a(+) Stp1 and pET28a(+) PknB/Stk1.

To generate pET28a(+)-EGFP, the eGFP gene was excised out from pEGFP-N3 (Clontech/TaKaRa) by restriction digestion and cloned into pET28a(+) (Merck-Novagen, Cat# 69864) using BamHI and NotI (ThermoFisherScientific). To generate expression vectors pET28a(+) Stp1 and pET28a(+) PknB/Stk1, respective genes were amplified by PCR using chromosomal DNA form *S. aureus* as template and indicated primers (supplementary Table S3) ordered at Merck-Sigma (see table). PCR generated DNA fragments were cloned into pET28a(+) by restriction digestion using NdeI and BamHI. All constructs were sequence verified by Sanger sequencing.

Protein overexpression and purification

For the pre-culture 20 ml YT 2x medium supplemented with 50 ug/mL kanamycin and 25 ug/mL chloramphenicol were inoculated with BL21(DE3)pLysS + pET28 (+), BL21(DE3)pLysS + pET28 (+)eGFP or BL21(DE3)pLysS + pET28 (+)Stp1 and incubated over-night at 37°C, 190 rpm. Pre-cultures were pelleted at 4,816 x g for 15 min, the pellets were suspended in 20 ml fresh medium and OD_{600} was measured. The 20 ml pre-cultures were added to 800 ml YT 2x medium supplemented 50 ug/mL kanamycin and 25 ug/mL chloramphenicol and incubated at 37°C, 190 rpm until an OD_{600} of 0.6 was reached. Protein expression was induced by adding a final concentration of 1 mM IPTG. Proteins were expressed for approximately 16 hrs at room temperature and overexpression was confirmed by SDS-PAGE and subsequent staining with Imperial™ Protein Stain (Thermo Fischer Scientific).

For the protein purification, pellets of a 400 ml culture were suspended in 5 ml lysis buffer per g wet cell paste (50mM Sodium dihydrogen phosphate (NaH2PO4), 0.5M sodium chloride (NaCl), 1% (v/v) Trition x-100, 0.1% (v/v) benzonase (Merck Millipore, Darmstadt, Germany), 10u/ml DNase 1, 2mM magnesium chloride (Sigma-Aldrich, Steinheim,

Germany) and completeMini EDTA-free protease inhibitors (Roche)). Samples were incubated for 1 hr on ice before sonication was performed using the Dr. Hielscher, UP100H with Micro tip MS3 (5 cycles of 1 min on: cycle 0.5, amplitude 90% and 1.5 min off). Insoluble parts and cell debris was pelted for 30 min at 4694 x g at 4°C and the soluble fraction was taking for purification using Ni-NTA Agarose beads (Qiagen). 200 ul Ni-NTA beads were loaded on Bio-Rad Poly-Prep Columns 10-ml and first equilibrated with 600 ul distilled water and subsequently with 1x binding buffer (0.5 M NaCl, 20 mM Tris-HCl pH 8, 5 mM imidazole). Soluble fraction was supplemented with a final concentration of 5 mM imidazole, loaded on the column and incubated for 1 hr a 4°C over-head shaking. The beads were washed twice with 1 ml 1x binding buffer and 4x with 500 ul 1x wash buffer (0.5 M NaCl, 30 mM imidazole and 20 mM Tris-HCL pH 8). Bound proteins were eluted five times with 500 ul 1x elution buffer (0.5 M imidazole, 0.5 M NaCl and 20 mM Tris-HCl pH 8). To confirm the purification a SDS-PAGE with subsequent staining using Imperial™ Protein Stain (Thermo Fischer Scientific) was performed. To stabilize the purified protein 50% glycerol was added to the Stp1 eluates.

Dephosphorylation assay

24.8 pmol synthetic pArg or 22.4 pmol pSTY peptides (n=4) were reconstituted in 100 ul reaction buffer (50mM Tris-HCl pH 8, 100 mM NaCl, 2 mM Manganese(II) chloride (MnCl2), 1 mM dithiothreitol (DTT)). Peptide solutions were either incubated for 1 hr at 20 °C with 2 ug Stp1, 10 units Shrimp alkaline phosphatase (New England BioLabs Inc.) or without the addition of any phosphatase. Subsequently samples were lyophilized. For LC-MS/MS analysis synthetic pArg-peptides were reconstituted as described for the stability analysis.

Data analysis

Raw files were processed using MaxQuant software (version 1.6.3.4 and 1.6.17.0 for HCD data and version 1.5.3.30 for ETD, EThcD and HCD data) and the Andromeda search engine was used to search against either *S. aureus* USA300 (Uniprot, June 2018, 5,954 entries), *E. coli* (Uniprot, March 2016, 4,434 entries) or *B. subtilis* (Uniprot/TrEMBL, December 2017, 4,247 entries) or a data base containing all 17 synthetic pArg-peptides with the following parameters for phosphoproteome analysis: trypsin digestion with a maximum of 3 missed cleavages, carbamidomethylation of cysteines (57.02 Da) as a fixed modification, methionine oxidation (15.99 Da), N-acetylation of proteins N-termini (42.01 Da) and phosphorylation on serine, threonine, tyrosine, histidine and arginine residues (79.96 Da) in case of pArg searches as variable modifications. Mass tolerance was set to 4.5 ppm at the MS1 level and 20 ppm at the MS2 level. The False Discovery Rate (FDR) was set to 1% for peptide-spectrum matches (PSMs) and protein identification using a target-decoy approach, a score cut-off of 40 was used in the case of modified peptides and the minimum peptide length was set to 7 residues. The match between run feature was enabled with a

matching time window of 0.7 min and an alignment time window of 20 min for the endogenous phosphoproteome. The MaxQuant generated tables “evidence.txt” and “phospho (HSTY)Sites.txt” were used to calculate the number of unique phosphopeptides and phosphosites identified, respectively, and known contaminants were filtered out. For full proteome analysis the following deviations were applied: trypsin digestion with a maximum of 2 missed cleavages, carbamidomethylation of cysteine’s (57.02 Da) as a fixed modification, methionine oxidation (15.99 Da), N-acetylation of protein N-termini (42.01 Da) as variable modifications. Relative label-free quantification was performed using the MaxLFQ algorithm with the minimum ratio count set to 2.

Statistical data analysis of endogenous peptides

All used scripts are published (https://github.com/hecklab/Protein-arginine-phosphorylation) and can be downloaded. For more detailed information on the endogenous data we refer you to our previous study[31]. The MaxQuant generated “phospho (RHSTY)Sites.txt” and “proteinGroups.txt” file, respectively, were used for subsequent statistical data analysis in R studio (R version 3.6.0). Four biological replicates per mutant were analysed. The data was filtered for “Reversed” and “Potential contaminant”. In case of the phosphoproteome, an Andromeda localization score greater than 0.75 was required. Intensities for the phosphoproteome data, or LFQ intensities in case of the full proteome analysis, were log2 transformed. For each phosphosite or protein, the median calculated per sample was subtracted to compensate for systematic measurement effects. Only proteins with at least three valid values in one condition and two valid values in at least one other condition. Data was checked for normal distribution before one-way ANOVA on each phosphosite or protein was done, after which the p-values were adjusted with the Benjamini-Hochberg procedure. The post-hoc Tukey Honestly Significant Difference (HSD) method was used to identify changing p-sites between the individual groups. A Tukey HSD *p*-value cut-off of 0.05 and a fold change cut-off of the mean +/- one standard-deviation of the data was used to select for significantly changing phosphosites or proteins between two groups.

Experiment Design and Statistical Rationale

Each sample was grown in n=3 or n=4 biological replicates, enriched and injected separately into the LC-MS/MS system. Synthetic peptides were run in n=4 technical replicates. Each raw file was separately processed using the MaxQuant software. This analysis was sufficient to saturate the number of phosphosites detected.

GO-term analysis

All identified pArg proteins were used to identify over-represented GO-terms and protein classes using PANTHER[32]. For the Fischer exact test, the whole *S. aureus* genome (2,889

entries) was used as a background. Only genes having a gene name and not only a gene Locus ID were considered (74 genes). As comparison all pST (303 genes) proteins were also used for an enrichment. An FDR cut-off of 0.05% was used.

Analysis of synthetic pArg-peptides

To analyze the stability of synthetic arginine phosphorylation, the MaxQuant generated "evidence.txt" and "msms.txt" files were used within the R studio (R version 3.6.0). Four technical replicates per time point were analyzed. The data was filtered for "Reversed", "Potential contaminant", "MULTI-MSMS" spectra and the intensity was log2 transformed. Identifications in the "evidence.txt" and "msms.txt" files were linked via the "Best.MS.MS" identifier.

Stability test

The intensity of unique peptides were summed and the average per time point calculated. Unique peptides are defined based on the Modified sequence. The ratio for pArg to unmodified peptides was determined as the difference between the respective log2 intensity averages per peptides.

Retention time analysis

The average RT per unique peptide was calculated taking different charge states into account. The RT of pArg-peptides were plotted as function of the RT of unmodified peptides. Respectively, peptides identified only as arginine phosphorylated or unmodified, were not considered for the analysis. Further pArg and pSTY peptide mixes were spiked in with 1ul of 1x iRT peptides and the RT of pArg-peptides were plotted as function of the RT of synthetic pSTY peptides. iRT peptides were used to compare the RT between different LC runs.

Fragmentation analysis

The sequence coverage for the three different fragmentation methods was calculated based on all ions contributing to the sequence coverage. For HCD y- and b-ions as well as the respective neutral-loss y* ions were considered. For ETD c- and z-radical ions as well as the less common c-radical and z-ions were considered. For EThcD respectively all of the above mentioned ions were considered. In case one fragment ions was identified based on different types e.g. y2 and y*2, this was counted as one ion to calculate the percentage based on the theoretically observed number of ions. The average sequence coverage, Andromeda score as well as localization probability was calculated for each fragmentation method as well as for each unique peptide.

Neutral-loss and immonium-ion analysis

Raw files from the fragmentation analysis and endogenous samples were converted into .mgf with Proteome Discoverer (Vers. 2.3.0.523). Subsequently mgf-files were analyzed using an in-house made script searching PSMs matching the different phosphorylated amino acids using the MaxQuant generated "evidence.txt" and "msmsScans.txt" file for neutral-loss ions (HPO_3 = 79.966331 Da, H_3PO_4 = 97.976896 Da and H_5PO_5 = 115.98746 Da) and potential immonium-ions.

To analyze the occurrence of potential immonium-ion the extracted MS2 spectra matching the respective phosphorylated amino acids were searched with an in-house made script for potential pArg iommium-ions (237.0747 *m/z* or 209.0798 *m/z*) and pHis immonium-ion (190.0376 *m/z*).

Spectra comparison

To compare MS2 spectra from the synthetic pArg-peptides with the endogenous identified counterparts, the MaxQuant generated "evidence.txt" files were loaded in an in-house developed software FragmentLab (https://scheltemalab.com/software) (v2.4.1.0). The synthetic peptides were used to create a peptide library that was used to validate the endogenous identified pArg-peptides. FragmentLab calculates a quality score for the spectra comparison. Spectra mass lists were exported to plot the spectra. All comparisons for identified endogenous and synthetic peptides were extracted.

Results & Discussion

Stp1 transposon mutant shows extensive pArg phosphorylation

Protein-arginine phosphorylation in Gram-positive bacteria gained attention since its physiological relevance for the bacterial stress response was shown in *B. subtilis*[8,19]. In this system, arginine phosphorylation is regulated by the only known protein-arginine kinase McsB and its respective protein-arginine phosphatase YwlE[7,19]. By analyzing a *S. aureus* COL ΔptpB mutant, Junker and co-workers[21] showed that the previous annotated tyrosine phosphatase PtpB acts as protein-arginine phosphatase and is therefore counterbalancing the staphylococcal protein-arginine kinase McsB[21]. They exclusively identified 207 pArg sites in the ΔptpB mutant, which represented the largest staphylococcal arginine phosphoproteome[21]. Spurred by this observation we searched for arginine phosphorylation in three different *S. aureus* mutant strains[31], and found that arginine phosphorylation is even more abundant than previously shown *in S. aureus*.

Surprisingly, we found 891 pArg sites (470 class I) in the NE98 mutant (lacking the membrane protein SdrE), making arginine phosphorylation almost as abundant as threonine phosphorylation (45.4% pS, 24% pT, 5% pY, 5.4% pH and 20.2% pR). Even more strikingly we identified 1,062 pArg sites (651 class I) in the Stp1 mutant (NE1919). With 26% of the phosphorylation sites localized on arginine, this is the second most abundant phosphorylation type after protein serine phosphorylation in this mutant (40.2% pS, 21.2% pT, 6.1% pY, 6.5% pH and 26% pR) (Figure 1c). In the Stk1 mutant 681 sites (350 class I) were identified (48.1% pS, 22% pT, 6.7% pY, 5.2% pH and 18% pR). This makes the percentage of pArg sites in the Stp1 mutant 6-8% higher compared to the other mutants (Figure 1a-c). In addition, we also compared the intensity distribution of Ser, Thr, Try, His and Arg phosphosites showing that pArg-sites are not having a higher intensity compared to the other phosphosites, but are simply more widespread (supplementary Figure S1).

Encouraged by the number of identified pArg-sites we also looked for arginine phosphorylation in other Gram-positive (*B. subtilis*) as well as Gram-negative (*E. coli*) bacteria. Even though *E. coli* possesses an arginine kinase homologous to McsB (see supplementary Figure S2), here, we could identify less than 3% class I pArg-sites for *E. coli* (supplementary Figure S2). For *B. subtilis* we could not identify any class I pArg-sites. This is in line with previous work where arginine phosphorylation could only be detected in the arginine phosphatase mutant (ΔYwlE) in *B. subtilis*[8]. These results show that arginine phosphorylation seems to play a prominent role in *S. aureus* and that, unexpectedly, Stp1 appears to play a functional role in regulating arginine phosphorylation in *S. aureus*. This was further underlined by an increased number of exclusively arginine phosphorylated proteins in the Stp1 mutants (supplementary Figure S3) compared to the control and the

Stk1 mutant. Based on the data we obtained, we identified a small part of the proteome that is uniquely phosphorylated on an arginine residue (supplementary Figure S3).

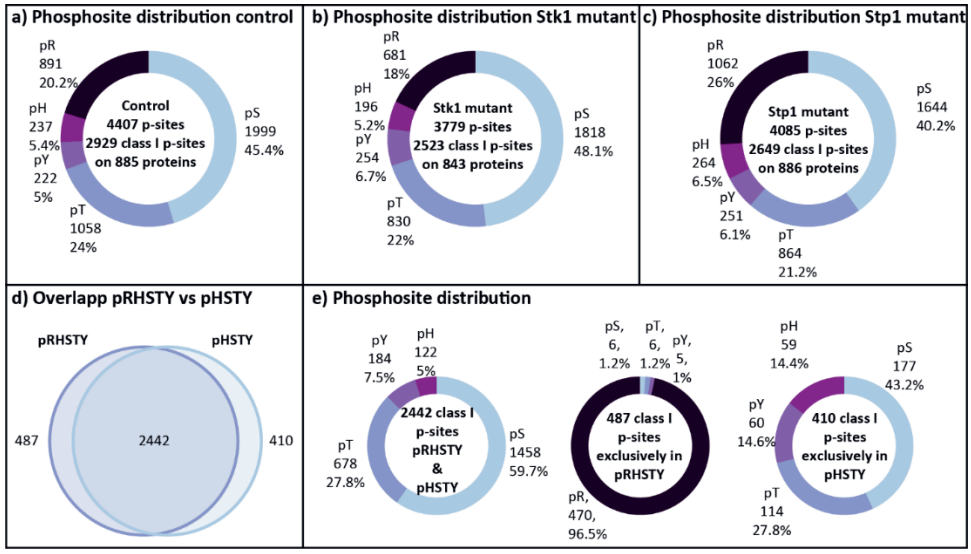


Figure 1: Arginine phosphorylation distribution in the control, Stk1 mutant as well Stp1 mutant of *S. aureus* and comparison of identified class I p-sites after MaxQuant search for pRHSTY and pHSTY for the control (n=4). **a)-c)** Number of identified phosphosites, class I phosphosites (Andromeda localization probability > 0.75) and phosphoproteins as well as the distribution of serine, threonine, tyrosine, histidine and arginine phosphorylation in the different mutant strains. **d)** Overlap of identified class I p-sites after either pRHSTY or pHSTY search for the control. **e)** Phosphorylation site distribution for the 2,442 p-sites identified in both searches as well as for the 487 and 410 class I p-sites identified exclusively after pRHSTY or pHSTY search respectively.

To avoid over interpretation of these results we checked for pHSTY phosphorylated counterpart peptides of the identified pArg-peptides. For 386 of the 1,294 identified pArg-peptides in at least one of the three mutants, we could identify a pHSTY counterpart. The Andromeda localization probability for those 386 pArg-peptides was lower than for their pHSTY counter parts (supplementary Figure S4a). However, for peptides exclusively identified as pArg the Andromeda localization probability was even slightly better than for pHSTY peptides. The average Andromeda score distribution was similar between all phosphopeptides (supplementary Figure S4b).

To further disprove that the identified pArg-sites are not the result of mis-localization we compared the class I sites identified for pRHSTY and pHSTY searches in the control (SdrE) mutant[31] and found that 83% of the identified sites overlapped. 487 class I phosphosites were exclusively identified in the data-set searched for pRHSTY of which 96.5% were localized on Arg stressing that the larger search space is almost exclusively giving rise to pArg sites (Figure 1d and e). Of these newly identified sites, 177 were reported to be

phosphorylated on a different amino acid in the pHSTY search, based on the identified peptides sequence (supplementary Figure S5a and c). We also identified a small number of class I sites that resulted in a different localization in the two searches, based on the same spectrum (supplementary Figure S5b). Even though this shows a certain degree of mis-localization either towards pHis or pSTY, we see that the vast majority of pArg sites seem to be in fact newly (and additionally) identified phosphosites.

Phosphorylation on arginine presumably increases the number of missed cleavages after tryptic digestion. Trypsin, the most commonly used protease for shotgun proteomics, forms a deep, narrow, negative charged binding pocket that enables ionic-interactions with long basic amino acids and thus cleaves specifically after lysine and arginine[33]. The addition of a negative charged phosphate group is therefore thought to impair the tryptic cleavage and studies on arginine phosphorylation increased the number of allowed missed cleavages for database searches[8,21,34]. Here, we show that indeed the majority of pArg-peptides (90%) showed at least one missed cleavage, whereas 70% of pHSTY peptides were identified without missed cleavage. For class I pArg–sites the percentage of non-missed cleaved decreased even further and 97% of the identified peptides showed at least one missed cleavage (supplementary Figure S6a). In addition, we also analyzed where the phosphorylation is located within the pArg-peptides. Comparing all pArg sites and class I pArg sites, showed that only a minority of class I pArg sites is located at the C-terminus and that the majority is one to five amino acids away from the C-terminus (supplementary Figure S6b). These findings indeed confirm that phosphorylation of arginine impairs tryptic digestions, probably due to electrostatic repulsion and or steric hindrance and in addition provides compelling evidence that identified (class I) pArg sites are not a product of mis-localization.

Arginine phosphorylation forms an acid labile phosphoramidate bond, similar to histidine phosphorylation that was long thought to impede the enrichment and analysis of this PTM. In addition, and again similar to what is known for histidine phosphorylation, neutral loss of the phosphate group under fragmentation conditions might complicate correct site-localization. Mis-localization of histidine phosphorylation is a major issue and can in some cases lead to extensive over-estimation of the amount of phosphorylation. To exclude mis-localization, we carefully evaluated the behavior of arginine phosphorylated peptides under enrichment and fragmentation conditions, in order to provide confidence in proper site localization.

Arginine phosphorylation is stable under acidic conditions

Several groups tried to modify existing phosphopeptide enrichment strategies by using slightly less acidic environments, to make them more suitable for enriching the acid labile arginine phosphorylation[8,21]. In order to demonstrate that pArg is actually stable under the

acidic conditions that have been used for pHis enrichments[12] and to validate the identification of new protein-arginine phosphorylation in *S. aureus,* we analyzed seventeen synthetic peptides representing randomly picked endogenous phosphopeptides identified in this study. First, we analyzed the chemical stability of arginine phosphorylation. Schmidt et al. demonstrated that the phosphoramidate bond is rapidly hydrolyzed under acidic conditions (pH < 3)[8]. However, it was recently shown that the phosphoramidate bond of pHis is relatively stable under acidic conditions (pH 2.3, room temperature)[12]. We analyzed the chemical stability of pArg under acidic conditions (pH 2) at 4°C over a time course of two hours (0 min, 15 min, 30 min, 60 min, 120 min). These conditions mimic the phosphopeptide enrichment conditions as well as loading conditions for the subsequent LC-MS/MS analysis. Under those conditions the synthetic pArg-peptides proved to be rather stable (Figure 2a). In order to see if the unmodified peptides became more predominant over time, the ratio of pArg-peptide to unmodified peptides over the time course of two hours was determined (Figure 2b). This demonstrated that all peptide ratios remain stable even after two hours, indicating the stability of synthetic pArg under acidic conditions. Next, we incubated the pArg-peptides under acidic conditions for 24 hrs at room temperature. Only five pArg-peptides could be identified after 24 hrs and for those a clear shift towards the unmodified peptide was observed (Figure 2c). These results convincingly show that the phosphoramidate bond of arginine is indeed acid labile, but sufficiently stable to allow for phosphopeptide enrichment as well as LC-MS/MS analysis, similar to what has been shown for histidine phosphorylation.

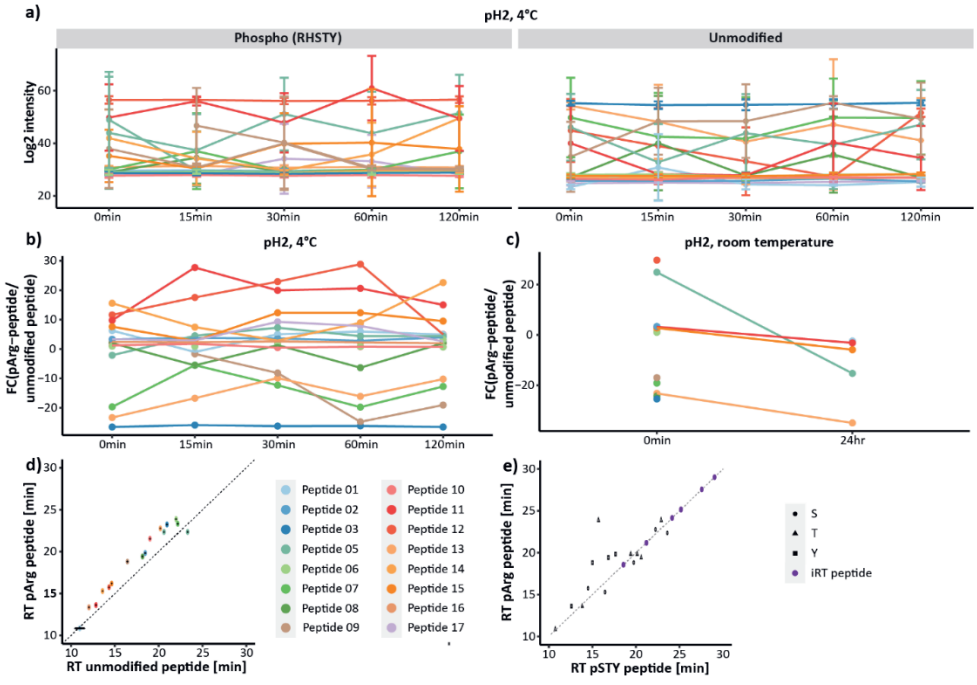


Figure 2: Stability of synthetic arginine phosphorylate peptides. **a)** Log2 intensity of the 16 identified synthetic pArg-peptides over a time course of 120 min under acidic conditions (pH 2) at 4°C and their respective unmodified counterpart. **b)** Fold-change of the respective pArg and unmodified peptides over the time course of 120 min at 4°C. The ratio of pArg to unmodified peptide was also determined after incubating the peptides under acidic conditions (pH 2) at room temperature for 24hrs **c)**. **d)** Retention of pArg-peptides in relation to retention time of unmodified peptides. The dashed line is showing the ideal behavior if both peptide types would show the same RT characteristics. **e)** Retention time of pArg peptides in relation to their pSer (circle), pThr (triangle) or pTyr (square) counterpart peptides. iRT peptides used for correlation of the RT are depicted in purple and the linear regression based on the RT of iRT peptides is shown as grey dashed line. For all plots dots represent the average and if applicable the standard deviation is displayed as error bars (n=4).

We also analyzed the retention time (RT) behavior of pArg-peptides in comparison to the unmodified peptides. pArg-peptides elute later than their respective unmodified counterpart peptides (Figure 2d). This is in line with previous studies showing that phosphorylation in general increases the retention time in reversed phase chromatography[35,36]. Previously, we showed that pHis peptides elute predominantly later than their pSTY counterpart peptide[37]. Unmodified histidine is contributing the least to retention on C18[35] and thus phosphorylation of histidine might increases the RT due to charge neutralization[36,37]. Unmodified arginine is only contributing slightly more to the RT on C18[35], therefore we investigated if pArg is similar to pHis eluting predominantly later than their pSTY counterpart. This comparison showed a slight trend towards later elution (Figure 2e), however this trend is not as strong as observed for pHis[37].

HCD out performs ETD and EThcD in the identification of synthetic pArg-peptides

Over the years different fragmentation strategies were developed, that especially focus on the identification of labile phosphopeptides[27,38]. ETD fragmentation, which predominantly occurs along the backbone (N-Cα) is known to preserve labile phosphorylations, however the fragmentation efficiency is highly dependent on the charge state, performing best for $z \geq 3$[27,28]. The problem of ETnoD events for lower charge states was solved by introducing EThcD that further improved the phosphorylation site localization compared to commonly used HCD fragmentation which shows a prominent neutral-loss[38,39].

Here, we compared HCD, ETD, and EThcD for the identification of synthetic pArg-peptides. A peptide mix combining all 17 synthetic peptides was analyzed using an Orbitrap Fusion with Orbitrap readout (n = 4). The acquired raw files were searched against a .fasta file containing all 17 synthetic peptides (see Material and Methods) and the resulting "evidence.txt" and "msms.txt" files were used to examine the sequence coverage and phosphosite localization. Taking all fragment ions contributing to the sequence identification into account, the average sequence coverage for the three different fragmentation methods was compared. Figure 3a shows that by far the highest sequence coverage was reached by HCD (70 %) against 48 % by ETD and only 30 % by EThcD. Additionally, in total 14 peptides were identified using HCD whereas both other techniques identified less than half of the synthetic pArg-peptides (ETD: 5, EThcD: 7, supplementary

Figure S6). To further assess the quality of the fragmentation and identification we compared the Andromeda scores, showing that ETD reaches the highest average Andromeda score with 204.9 (HCD: 154.5, EThcD: 94.7, Figure 3b). These results are not in full agreement with previous studies comparing different fragmentation methods. Previous reports showed that collisional activation of ET products increased the sequence coverage compared to ETD or HCD alone especially for short, doubly protonated tryptic peptides[38,40]. Also, EThcD seems to be superior to HCD with respect to phosphosite assignment. For synthetic pArg-peptides the highest identification rate was previously observed by using ETD compared to HCD[29]. ETD is known to perform significantly worse on peptides containing proline. The cleavage of the N-Cα bond keeps the resulting fragment attached to the remaining atoms of the ring and therefore, c- and z-ions containing the N- and C-termini of proline are not observed[27]. Indeed seven of our 17 synthetic peptides contain at least one proline (supplementary Table S2). However, also peptides not containing any proline were not identified. Even though these findings are in contrast with commonly accepted notions, we are not the first ones seeing HCD outperforming especially ETD. A similar trend was observed when using a library of > 100,000 unmodified and modified peptides[35]. Also in agreement with the work of Marx *et al.*, unambiguous localization (Andromeda localization probability of 1.0) of pArg was observed when using ETD and EThcD (Figure 3c). However as pointed out before, the number of identified pArg-peptides was low. In comparison, HCD performed slightly worse with approx. 80 % of all identified pArg-peptides being class I phosphosites (≥0.75) and 60% being unambiguously identified (Andromeda localization probability of 1.0). These findings underline the superiority of ETD and ETchD fragmentation in phosphosite localization as reported earlier[38,40], however the overall identification of our synthetic pArg-peptides was clearly hampered using ETD and EThcD compared to HCD. The superiority of phosphosite assignment has been shown to result from less phosphoric neutral-losses when using ETD and ETchD[38,40]. Here, we can show that indeed no neutral-loss was observed using ETD and that HCD proved to exhibited the highest amount of neutral-loss (HPO_3 = 79.966331 Da, H_3PO_4 = 97.976896 Da and H_5PO_5 = 115.98746 Da, Figure 3d). Using the endogenous identified phosphopeptides, we compared the amount of phosphoric neutral-loss for the different amino acids (STYHR) and could show that the labile phosphoramidates (pHis and pArg) exhibit significantly more neutral-loss, especially H_3PO_4-loss (Figure 3e). This is in line with previous studies showing a significant neutral-loss for pHis[37,40] and pArg[29]. Even though a triplet neutral-loss is more characteristic for pHis and pArg it does not represent unique evidence for the identification of phosphoramidates since only around 20% of PSMs matching pArg or pHis peptides showed the respective loss. Analyzing synthetic and endogenous pHis peptides we could previously show the existence of a pHis-immonium ion that helps with the identification of the labile protein histidine phosphorylation[37]. Accordingly, we also searched for respective pArg-immonium ions that

could improve the localization. Indeed, we could identify two ions (237.0747 *m/z* and 209.0798 *m/z*) present in around 30 % of MS2 spectra from synthetic pArg-peptides (supplementary Figure S8). Searching MS2 spectra of synthetic pHSTY peptides revealed that those ions are not present in the respective MS2 spectra. However, the potential pArg-immonium ions were only present in less than 3% of MS2 from endogenous pArg-peptides. In comparison 20% MS2 spectra from synthetic pHis peptides and 16.5% MS2 spectra from endogenous pHis peptides in *E.coli* showed the pHis-immonim ion of 190.0376 *m/z*.

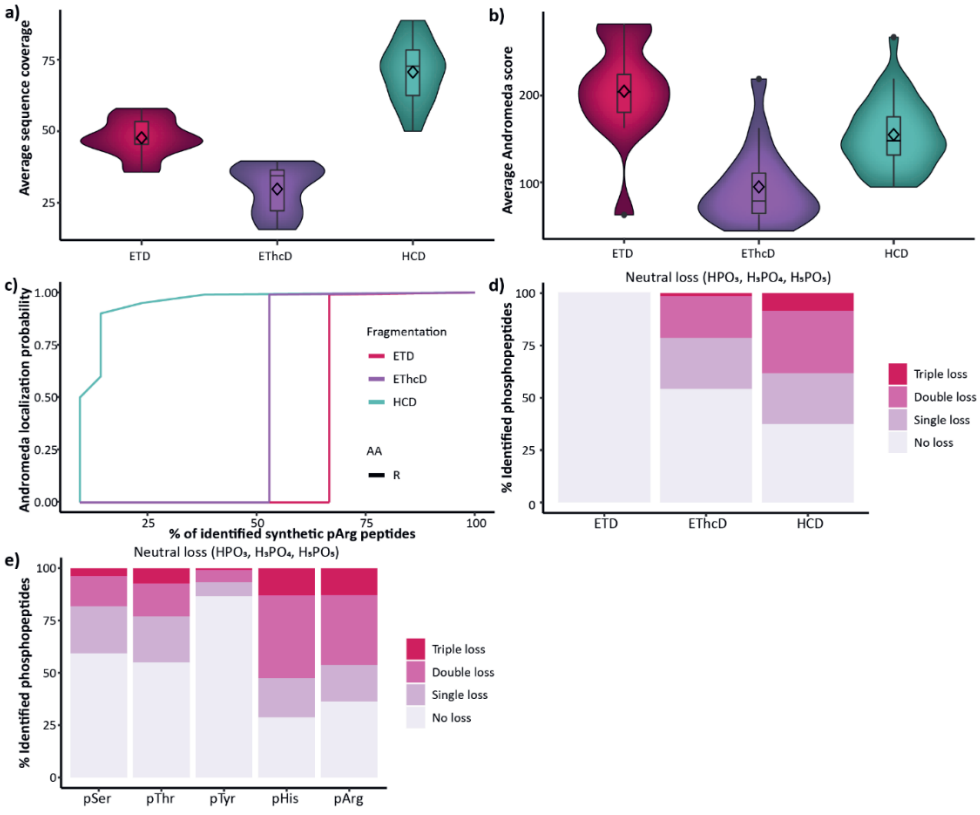


Figure 3: Fragmentation characteristics of synthetic pArg peptides. **a)** Violin plot showing the average sequence coverage identified with ETD (pink), EThcD (purple) and HCD (turquoise) fragmentation. **b)** Violin plot showing the average Andromeda score after ETD (pink), EThcD (purple) and HCD (turquoise) fragmentation. The mean values are highlighted as diamond. The median, first and third quantile are depicted in the boxplot. Outliers are highlighted as individual points. **c)** Andromeda localization probability for the three fragmentation methods as function of the percentage of identified synthetic pArg peptides. **d)** Phosphate neutral-loss triplets (i.e., 80, 96, and 116 Da, respectively) observed in spectra derived from ETD, EThcD and HCD fragmentation of synthetic pArg peptides. Percentages of peptide-spectrum matches (PSMs) corresponding to synthetic pArg phosphopeptides that exhibit neutral losses. **e)** Here, the percentages of peptide-spectrum matches (PSMs) corresponding to endogenous phosphopeptides identified in *S. aureus* samples that exhibit neutral losses are displayed for pSer, pThr, pTyr, pHis and pArg peptides.

Synthetic peptides confirm identification of endogenous pArg

To further validate endogenous pArg phosphorylation, we compared the acquired HCD spectra of our synthetic peptides with the spectra of our endogenous peptides (supplementary Figure S9). Figure 4a shows the spectrum comparison for the 50S ribosomal protein L17 (RplQ). The endogenous peptide was identified with an Andromeda score of 113.69 and a localization probability of 0.99999. The synthetic peptides had an Andromeda score of 191.53 and a localization probability of 1. For both peptides almost the complete y-ion series (y1-y8) including y3-ion allowing the unambiguous localization of the phosphorylation site as well as two b-ions (b2 and b3) were identified showing a high similarity between the two spectra and therefore confirming the identification of pR36 on RplQ. For the Acetyltransferase SAUSA300_2505 the overlap of identified fragments was not as high as for RplQ, still we could show a high degree of similarity between the spectra of the synthetic and endogenous peptide (Figure 4b). The localization of the phosphorylation site is high for both peptides (synthetic peptide: 1, endogenous peptide 0.99981). Whereas the Andromeda score of the synthetic peptides was 157.97, the endogenous peptide only reached 56.122. The endogenous spectrum shows more peaks especially in the lower *m/z* range which can simply be the result of co-eluting peptides. Nevertheless, the identified b- and y-ion match perfectly to the fragment ions identified for the synthetic peptide and thus providing strong evidence for the accurate assignment of the spectrum. A peptide of the MutT/nudix family protein (Uniprot ID: A0A0H2XHE8) was identified to be phosphorylated on Thr as well as Arg. This allowed us to compare both spectra to the synthetic pArg-peptide, showing a clear deviation in the b4 ion for the endogenous pThr peptide (Figure S10) compared to the synthetic pArg-peptide where all identified b-and y-ions for the endogenous pArg-peptide matched. We also compared the synthetic pThr peptide to the endogenous pThr peptide which showed greater similarity than compared to the synthetic pArg-peptide (supplementary Figure S9). This shows that the identification and phosphorylation site assignment of the endogenous peptides is highly accurate and clearly supports the high number of identified pArg-peptides in this study.

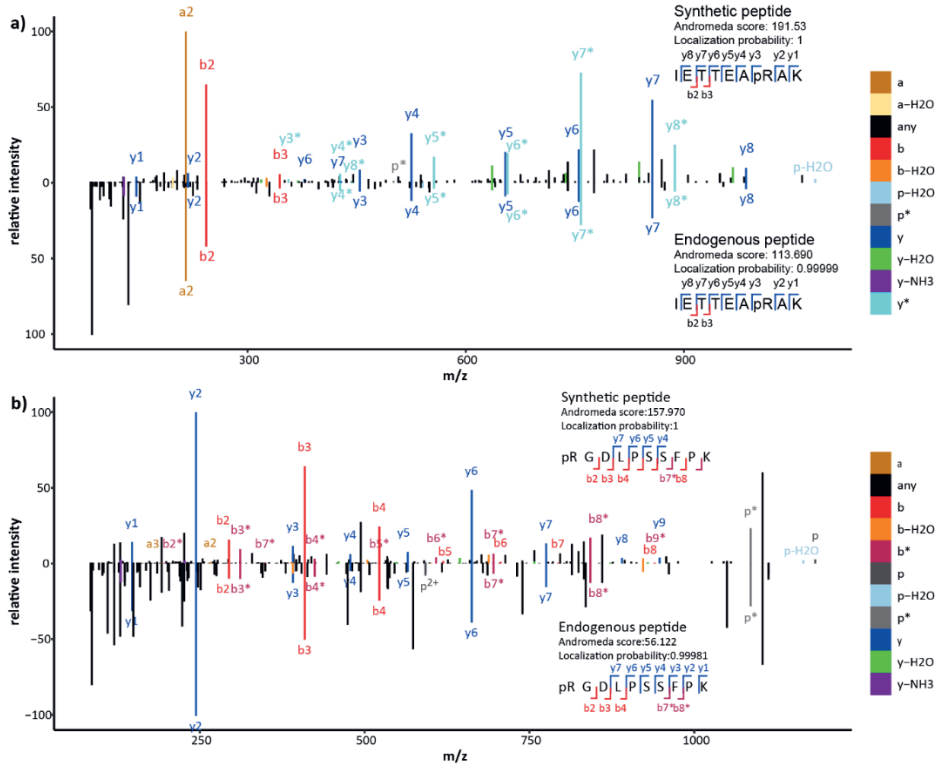


Figure 4: Comparison of the acquired fragmentation spectra of synthetic and the respective endogenous peptides for **a)** RplQ and **b)** SAUSA300_2505. b- and y-ions are highlighted in red and blue.

Stp1 is not a direct arginine phosphatase

Quantitative analysis of arginine phosphorylation in the three mutant strains revealed that forty-four class I pArg sites were exclusively identified in the Stp1 mutant and 72 class I pArg sites were significantly over-represented compared to the control (Tukey HSD *p*-value cut-off of 0.05 and a fold change cut-off of of $\bar{x} \pm \sigma$ of the data) (Figure 5a). Overall, all pArg-sites identified were skewed towards an over-representation in the Stp1 mutant compared to the control (Figure 5a). At the same time pArg-sites were not shown to be regulated between the Stk1 mutant and the control (Figure 5b). The majority of identified pArg-sites in the Stp1 mutant showed an over-representation compared to the control. Whereas, a true on/off regulation of target sites was previously reported for the ΔPtpB (arginine phosphatase) mutant[21]. Still the over-representation in the Stp1 mutant provides strong evidence for a regulatory effect of Stp1 on the arginine phosphoproteome in *S. aureus* USA300.

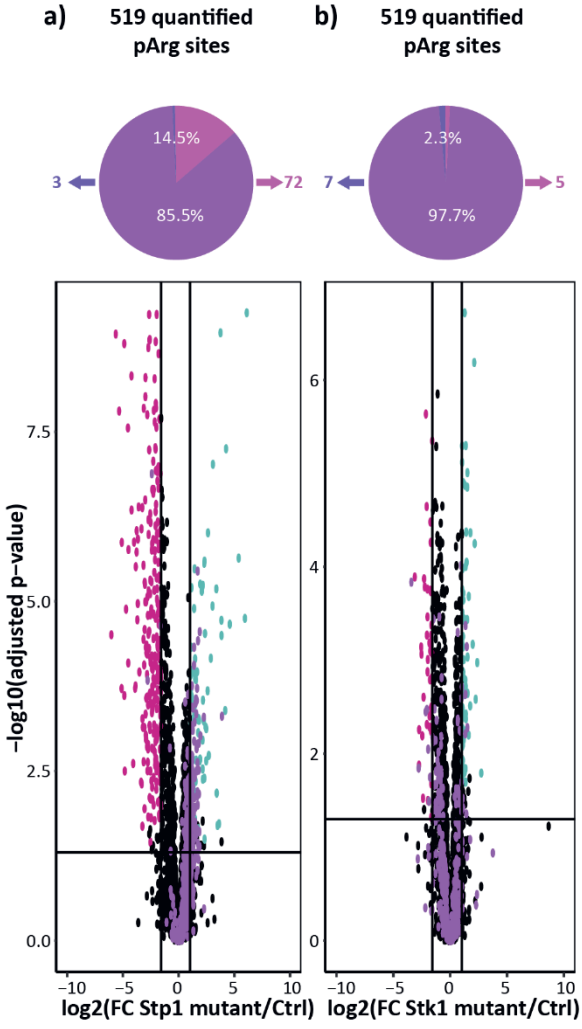


Figure 5: **a)** Pie chart showing the changes in pArg phosphorylation and volcano plot showing the quantitatively changing phosphorylation sites between the Stp1 mutant and Ctrl; identifying 411 changes (Tukey HSD p-value cut-off of 0.05 and a fold change cut-off of $\bar{x} \pm \sigma$ of the data). Quantified pArg sites are highlighted in purple and show a clear over-representation in the Stp1 mutant. Under-represented p-sites are shown in pink and over-represented sites in turquoise. Three pArg sites were under-represented and 72 pArg sites were over-represented. **b)** Pie chart showing the changes in pArg phosphorylation and volcano plot comparing the phosphorylation sites between Stk1 mutant and Ctrl, identifying 117 changes (Tukey HSD p-value cut-off of 0.05 and a fold change cut-off of $\bar{x} \pm \sigma$ of the data). Seven pArg sites were under-represented and 5 pArg sites over-represented. All identified pArg sites are equally distributed between Stk1 mutant and Ctrl (purple dots).

To get more insight into the functional relevance of arginine phosphorylated proteins we performed a GO-term as well protein class enrichment using PANTHER[32] and compared it to Ser/Thr phosphorylated proteins (supplementary Figure S11). GO-terms related to protein transcription/translation as well as metabolic pathways were enriched according to a Fischer exact test. Protein-arginine phosphorylation has been shown to be involved in the general stress response[8,23] as well as marking proteins for degradation within the ClpCP proteasome[34], therefore it is also of no surprise that those terms were enriched. Next we tried to identify sequence motifs surrounding pArg-sites. In line with previous studies on *B. subtilis*[8] and *S. aureus*[21] we could not identify a preferred sequence motif using pLogo[41] neither for sites solely identified in the Stp1 mutant nor for all identified pArg-sites (supplementary Figure S12). This supports the hypothesis that protein-arginine kinases do

not necessarily have a substrate specificity[8]. It is hypothesized that arginine kinases require additional regulatory mechanisms to achieve substrate specificity such as temporal activation, cellular localization or protein-protein-interactions[8]. The clear presence of a specific pArg proteome supports this hypothesis.

To investigate whether Stp1 has arginine phosphatase activity, we overexpressed and purified Stp1 (supplementary Figure S13) and performed a dephosphorylation assay using synthetic pArg, pSer, pThr or pTyr peptides as well as endogenous phosphopeptides. While Stp1 did not show any influence on arginine or tyrosine phosphorylation (Figure 6a and c), incubation of Stp1 with pST peptides led to a decrease of identified synthetic pST peptides by almost 44% (Figure 6b). As a positive control the shrimp alkaline phosphatase (rSAP) showed a decrease of identified phosphopeptide by up to 60% for all three types of phosphopeptides. Stp1 treatment also resulted in a decrease in phosphorylation after incubation with endogenous phosphopeptide, whereas the total number of identified peptides remained the same (Figure 6d). Even though in this case phosphorylation generally decreased after treatment with Stp1, a clear preference for pS and pT dephosphorylation was observed (Figure 6e). These results indicate that Stp1 does not have specific arginine phosphatase activity but rather has a secondary effect on the arginine phosphoproteome.

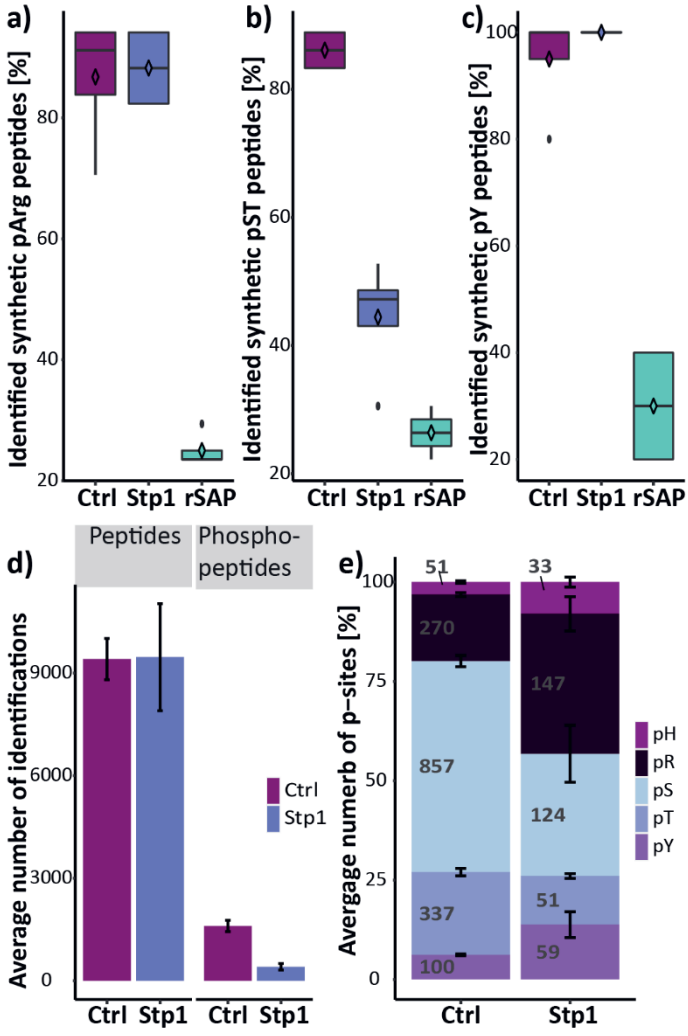


Figure 6: Dephosphorylation of synthetic and endogenous phosphopeptides. Percentage of identified synthetic **a)** pArg, **b)** pSer/Thr and **c)** pTyr after incubation without phosphatase, recombinant Stp1 or shrimp alkaline phosphatase rSAP (n=4). **d)** Average number of identified peptides and phosphopeptides after incubation with or without Stp1. **e)** Distribution of serine, threonine, tyrosine histidine and arginine phosphosites identified. The stack bars represent the average percentage of phosphosites (n=3) and the standard deviation is shown as error bars. Numbers depict the number of identified sites.

Conclusions

Here we report the largest arginine phosphoproteome for *S. aureus* USA300 to date. In the future this dataset will help to shed light on the biological role of arginine phosphorylation that thus far is limited to a general-stress response and marking proteins for degradation[8,23,34]. Using synthetic pArg-peptides we were able to confirm that arginine phosphorylation is stable enough to be enriched under conditions suitable for histidine phosphorylation[12]. Also the extensive endogenous arginine phosphorylation was confirmed by comparing MS2 spectra to the synthetic counterparts supporting the confident phosphorylation site assignment. Using synthetic peptides we also showed that HCD is still the gold standard in phosphoproteomics and outperforms ETD as well as EThcD for the

identification of 17 synthetic arginine phosphorylated peptides, while site assignment was comparable between the three fragmentations methods. Finally, we present strong evidence that the eukaryotic-like Ser/Thr phosphatase Stp1, even though not a direct pArg phosphatase, influences the *S. aureus* USA300 arginine phosphoproteome and thereby introduces an additional regulatory mechanism for arginine phosphorylation next to McsB and PtpB.

Acknowledgments

We acknowledge Henk van den Toorn and Franziska Völlmy for bioinformatics support, Nina van Sorge for providing the *S. aureus* samples and laboratory space for sample preparations and Albert Heck for scientific discussions. S.L. acknowledges support from the Netherlands Organization for Scientific Research (NWO) through a VIDI grant (project 723.013.008).

Competing financial interest

The authors declare no competing financial interest

Data availability

All raw data that support the findings of this study have been deposited in ProteomeXchange with the accession number PXD026981.

References

1. Cohen P. The origins of protein phosphorylation. *Nat Cell Biol*. 2002;4(5):E127-E130. doi:10.1038/ncb0502-e127

2. Hunter T. Why nature chose phosphate to modify proteins. *Philos Trans R Soc B Biol Sci*. 2012;367(1602):2513-2516. doi:10.1098/rstb.2012.0013

3. FISCHER EH, GRAVES DJ, CRITTENDEN ER, KREBS EG. Structure of the site phosphorylated in the phosphorylase b to a reaction. *J Biol Chem*. 1959;234(7):1698-1704. doi:10.1016/s0021-9258(18)69910-7

4. Johnson LN, Lewis RJ. Structural basis for control by phosphorylation. *Chem Rev*. 2001;101(8):2209-2242. doi:10.1021/cr000225s

5. Sharma K, D'Souza RCJ, Tyanova S, Schaab C, Wiśniewski J, Cox J, Mann M. Ultradeep Human Phosphoproteome Reveals a Distinct Regulatory Nature of Tyr and Ser/Thr-Based Signaling. *Cell Rep*. 2014;8(5):1583-1594. doi:10.1016/j.celrep.2014.07.036

6. Ninfa AJ. Use of two-component signal transduction systems in the construction of synthetic genetic networks. *Curr Opin Microbiol*. 2010;13(2):240-245. doi:10.1016/j.mib.2010.01.003

7. Fuhrmann J, Schmidt A, Spiess S, Lehner A, Turgay K, Mechtler K, Charpentier E, Clausen T. McsB Is a Protein Arginine Kinase That Phosphorylates and Inhibits the Heat-Shock Regulator CtsR. *Science (80-)*. 2009;324(5932):1323-1327. doi:10.1126/science.1170088

8. Schmidt A, Trentini DB, Spiess S, Fuhrmann J, Ammerer G, Mechtler K, Clausen T. Quantitative Phosphoproteomics Reveals the Role of Protein Arginine Phosphorylation in the Bacterial Stress Response. *Mol Cell Proteomics*. 2014;13(2):537-550. doi:10.1074/mcp.M113.032292

9. Sickmann A, Meyer HE. Phosphoamino acid analysis. *Proteomics*. 2001;1(2):200-206. doi:10.1002/1615-9861(200102)1:2<200::AID-PROT200>3.0.CO;2-V

10. Ge R, Shan W. Bacterial Phosphoproteomic Analysis Reveals the Correlation Between Protein Phosphorylation and Bacterial Pathogenicity. *Genomics, Proteomics Bioinforma*. 2011;9(4-5):119-127. doi:10.1016/S1672-0229(11)60015-6

11. Potel CM, Lin M-H, Heck AJR, Lemeer S. Defeating Major Contaminants in Fe 3+ - Immobilized Metal Ion Affinity Chromatography (IMAC) Phosphopeptide Enrichment. *Mol Cell Proteomics*. 2018;17(5):1028-1034. doi:10.1074/mcp.TIR117.000518

12. Potel CM, Lin MH, Heck AJR, Lemeer S. Widespread bacterial protein histidine phosphorylation revealed by mass spectrometrybased proteomics. *Nat Methods*. 2018;15(3):187-190. doi:10.1038/nmeth.4580

13. Park JY, Kim JW, Moon BY, Lee J, Fortin YJ, Austin FW, Yang SJ, Seo KS. Characterization of a novel two-component regulatory system, HptRS, the regulator for the hexose phosphate transport system in Staphylococcus aureus. *Infect Immun*. 2015;83(4):1620-1628. doi:10.1128/IAI.03109-14

14. Fridman M, Williams GD, Muzamal U, Hunter H, Siu KWMWMMWM, Golemi-Kotra D. Two unique phosphorylation-driven signaling pathways crosstalk in staphylococcus aureus to modulate the cell-wall charge: Stk1/Stp1 meets GraSR. *Biochemistry*. 2013;52(45):7975-7986. doi:10.1021/bi401177n

15. Kruger E. Clp-mediated proteolysis in Gram-positive bacteria is autoregulated by the stability of a repressor. *EMBO J*. 2001;20(4):852-863. doi:10.1093/emboj/20.4.852

16. Kirstein J, Zühlke D, Gerth U, Turgay K, Hecker M. A tyrosine kinase and its activator control the activity of the CtsR heat shock repressor in B. subtilis. *EMBO J*. 2005;24(19):3435-3445. doi:10.1038/sj.emboj.7600780

17. Zhang Z-YY. Mechanistic Studies on Protein Tyrosine Phosphatases. In: *Progress in Nucleic Acid Research and Molecular Biology*. Vol 73. ; 2003:171-220. doi:10.1016/S0079-6603(03)01006-7

18. Musumeci L, Bongiorni C, Tautz L, Edwards RA, Osterman A, Perego M, Mustelin T, Bottini N. Low-Molecular-Weight Protein Tyrosine Phosphatases of Bacillus subtilis. *J Bacteriol*.

2005;187(14):4945-4956. doi:10.1128/JB.187.14.4945-4956.2005

19. Elsholz AKW, Turgay K, Michalik S, Hessling B, Gronau K, Oertel D, Mader U, Bernhardt J, Becher D, Hecker M, Gerth U. Global impact of protein arginine phosphorylation on the physiology of Bacillus subtilis. *Proc Natl Acad Sci*. 2012;109(19):7451-7456. doi:10.1073/pnas.1117483109

20. Bäsell K, Otto A, Junker S, Zühlke D, Rappen G-M, Schmidt S, Hentschker C, Macek B, Ohlsen K, Hecker M, Becher D. The phosphoproteome and its physiological dynamics in Staphylococcus aureus. *Int J Med Microbiol*. 2014;304(2):121-132. doi:10.1016/j.ijmm.2013.11.020

21. Junker S, Maaβ S, Otto A, Michalik S, Morgenroth F, Gerth U, Hecker M, Becher D. Spectral Library Based Analysis of Arginine Phosphorylations in Staphylococcus aureus. *Mol Cell Proteomics*. 2018;17(2):335-348. doi:10.1074/mcp.RA117.000378

22. Trentini DB, Fuhrmann J, Mechtler K, Clausen T. Chasing phosphoarginine proteins: Development of a selective enrichment method using a phosphatase trap. *Mol Cell Proteomics*. 2014;13(8):1953-1964. doi:10.1074/mcp.O113.035790

23. Junker S, Maaß S, Otto A, Hecker M, Becher D. Toward the Quantitative Characterization of Arginine Phosphorylations in Staphylococcus aureus. *J Proteome Res*. 2018;18(1):acs.jproteome.8b00579. doi:10.1021/acs.jproteome.8b00579

24. Wozniak DJ, Tiwari KB, Soufan R, Jayaswal RK. The mcsB gene of the clpC operon is required for stress tolerance and virulence in Staphylococcus aureus. *Microbiol (United Kingdom)*. 2012;158(10):2568-2576. doi:10.1099/mic.0.060749-0

25. LEVY-FAVATIER F, DELPECH M, KRUH J. Characterization of an arginine-specific protein kinase tightly bound to rat liver DNA. *Eur J Biochem*. 1987;166(3):617-621. doi:10.1111/j.1432-1033.1987.tb13558.x

26. Wakim BT, Aswad GD. Ca2+-calmodulin-dependent phosphorylation of arginine in histone 3 by a nuclear kinase from mouse leukemia cells. *J Biol Chem*. 1994;269(4):2722-2727. doi:10.1016/s0021-9258(17)42003-5

27. Syka JEP, Coon JJ, Schroeder MJ, Shabanowitz J, Hunt DF. Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proc Natl Acad Sci*. 2004;101(26):9528-9533. doi:10.1073/pnas.0402700101

28. Good DM, Wirtala M, McAlister GC, Coon JJ. Performance Characteristics of Electron Transfer Dissociation Mass Spectrometry. *Mol Cell Proteomics*. 2007;6(11):1942-1951. doi:10.1074/mcp.M700073-MCP200

29. Schmidt A, Ammerer G, Mechtler K. Studying the fragmentation behavior of peptides with arginine phosphorylation and its influence on phospho-site localization. *Proteomics*. 2013;13(6):945-954. doi:10.1002/pmic.201200240

30. Fey PD, Endres JL, Yajjala VK, Widhelm TJ, Boissy RJ, Bose JL, Bayles KW. A Genetic Resource for Rapid and Comprehensive Phenotype Screening of Nonessential Staphylococcus aureus Genes. *MBio*. 2013;4(1):e00537-12-e00537-12. doi:10.1128/mBio.00537-12

31. Prust N, van der Laarse S, van den Toorn HWP, van Sorge NM, Lemeer S. In-Depth Characterization of the Staphylococcus aureus Phosphoproteome Reveals New Targets of Stk1. *Mol Cell Proteomics*. 2021;20:100034. doi:10.1074/mcp.RA120.002232

32. Mi H, Muruganujan A, Huang X, Ebert D, Mills C, Guo X, Thomas PD. Protocol Update for large-scale genome and gene function analysis with the PANTHER classification system (v.14.0). *Nat Protoc*. 2019;14(3):703-721. doi:10.1038/s41596-019-0128-8

33. Olsen J V., Ong SE, Mann M. Trypsin cleaves exclusively C-terminal to arginine and lysine residues. *Mol Cell Proteomics*. 2004;3(6):608-614. doi:10.1074/mcp.T400003-MCP200

34. Trentini DB, Suskiewicz MJ, Heuck A, Kurzbauer R, Deszcz L, Mechtler K, Clausen T. Arginine phosphorylation marks proteins for degradation by a Clp protease. *Nature*. 2016;539(7627):48-53. doi:10.1038/nature20122

35. Marx H, Lemeer S, Schliep JE, Matheron L, Mohammed S, Cox J, Mann M, Heck AJR, Kuster B. A large synthetic peptide and phosphopeptide reference library for mass spectrometry–based proteomics. *Nat*

Biotechnol. 2013;31(6):557-564. doi:10.1038/nbt.2585

36. Steen H, Jebanathirajah JA, Rush J, Morrice N, Kirschner MW. Phosphorylation Analysis by Mass Spectrometry: Myths, Facts, and the Consequences for Qualitative and Quantitative Measurements. *Mol Cell Proteomics*. 2006;5(1):172-182. doi:10.1074/mcp.M500135-MCP200

37. Potel CM, Lin MH, Prust N, Van Den Toorn HWP, Heck AJR, Lemeer S. Gaining Confidence in the Elusive Histidine Phosphoproteome. *Anal Chem*. 2019;91(9):5542-5547. doi:10.1021/acs.analchem.9b00734

38. Frese CK, Altelaar AFMM, van den Toorn H, Nolting D, Griep-Raming J, Heck AJRR, Mohammed S. Toward Full Peptide Sequence Coverage by Dual Fragmentation Combining Electron-Transfer and Higher-Energy Collision Dissociation Tandem Mass Spectrometry. *Anal Chem*. 2012;84(22):9668-9673. doi:10.1021/ac3025366

39. Olsen J V., Macek B, Lange O, Makarov A, Horning S, Mann M. Higher-energy C-trap dissociation for peptide modification analysis. *Nat Methods*. 2007;4(9):709-712. doi:10.1038/nmeth1060

40. Penkert M, Hauser A, Harmel R, Fiedler D, Hackenberger CPR, Krause E. Electron Transfer/Higher Energy Collisional Dissociation of Doubly Charged Peptide Ions: Identification of Labile Protein Phosphorylations. *J Am Soc Mass Spectrom*. Published online 2019. doi:10.1007/s13361-019-02240-4

41. O'Shea JP, Chou MF, Quader SA, Ryan JK, Church GM, Schwartz D. PLogo: A probabilistic approach to visualizing sequence motifs. *Nat Methods*. 2013;10(12):1211-1212. doi:10.1038/nmeth.2

Supplementary information

Table S1: Fe3+-IMAC column enrichment gradient. Buffer A corresponds to the loading buffer (30 %ACN, 0.07 % TFA) and B to the elution buffer (0.3 % ammonia).

Time (min)	Flow rate (ml/min)	%A	%B
0 – 7.00	0.1	100	0
7.01 – 12.00	1	100	0
12.01 – 13.50	1	40	60
13.51 – 16.00	0.5	40	60
16.01 – 25.00	1	100	0

Table S2: Synthetic pArg and pSTY peptides

Peptide	Sequence	Amount per aliquot peptidemix	Fastafile identification
Peptide 01	H-IETTEAR(PO3)AK-OH	800 fmol	Peptide_01
Peptide 02	H-NAIAEAR(PO3)K-OH	1600 fmol	Peptide_02
Peptide 03	H-TIDR(PO3)TVIYNGK-OH	1600 fmol	Peptide_03
Peptide 04	H-EIADGTALVNLVSR(PO3)R-OH	2400 fmol	Peptide_04
Peptide 05	H-NFEGR(PO3)IHPLVK-OH	2400 fmol	Peptide_05
Peptide 06	H-R(PO3)GDLPSSFPK-OH	1600 fmol	Peptide_06
Peptide 07	H-IQDNALTR(PO3)QFR-OH	1600 fmol	Peptide_07
Peptide 08	H-KTEDYISGR(PO3)FG-OH	800 fmol	Peptide_08
Peptide 09	H-VLTYDLR(PO3)GHGK-OH	2400 fmol	Peptide_09
Peptide 10	H-QR(PO3)IYSIK-OH	800 fmol	Peptide_10
Peptide 11	H-PLIR(PO3)AIQK-OH	1600 fmol	Peptide_11
Peptide 12	H-HTYQNEPLPNR(PO3)K-OH	800 fmol	Peptide_12
Peptide 13	H-AVADR(PO3)LENSEAK-OH	1600 fmol	Peptide_13
Peptide 14	H-NMFIR(PO3)PESK-OH	1600 fmol	Peptide_14
Peptide 15	H-R(PO3)IIPSDK-OH	1600 fmol	Peptide_15
Peptide 16	H-AADIIR(PO3)GR-OH	800 fmol	Peptide_16
Peptide 17	H-FR(PO3)GHDGPIK-OH	800 fmol	Peptide_17
Synthetic pSTY peptide			
Peptide_001	TTLTAAIAtVLAK	800 fmol	Peptide_18
Peptide_002	TTVtGVEMFRK	400 fmol	Peptide_19
Peptide_003	tTVTGVEMFRK	400 fmol	Peptide_19
Peptide_004	DSLSQEQEVtRR	400 fmol	Peptide_20
Peptide_005	LFGsVSTK	400 fmol	Peptide_21
Peptide_006	SDFNHTDHSTtNHSQTPR	800 fmol	Peptide_22
Peptide_007	INQMMtMMNK	400 fmol	Peptide_23
Peptide_008	LNIMGINTAEEFsER	400 fmol	Peptide_24
Peptide_009	EEFEEIsVRGR	400 fmol	Peptide_25
Peptide_010	LDMSEYSDtTAVSK	400 fmol	Peptide_26
Peptide_011	KLsNENFVSK	400 fmol	Peptide_27
Peptide_012	IHsINIK	400 fmol	Peptide_28
Peptide_013	KADsILAASGVSDDAAK	400 fmol	Peptide_29
Peptide_014	SNDtIVETFK	800 fmol	Peptide_30
Peptide_015	GYSVSVFNRsSEK	400 fmol	Peptide_31
Peptide_016	GDVYQSLsQLPESVR	800 fmol	Peptide_32
Peptide_017	IHsINIK	400 fmol	Doubly synthezied
Peptide_018	DFAsLLHEDLK	800 fmol	Peptide_33
Peptide_019	NVTsGANPVGLR	400 fmol	Peptide_34

Peptide	Sequence	Amount per aliquot peptidemix	Fastafile identification
Peptide_020	QSEDAsYRQQYAK	400 fmol	Peptide_35
Peptide_021	LDMSEYsDTTAVSK	400 fmol	Peptide_26
Peptide_022	KADSILAAsGVSDDAAK	400 fmol	Peptide_29
Peptide_023	KADSILAASGVsDDAAK	400 fmol	Peptide_29
Peptide_024	GDVYQsLSQLPESVR	400 fmol	Peptide_32
Peptide_025	TIDRtVIYNGK	400 fmol	Peptide_03
Peptide_026	TIDRTVIyNGK	800 fmol	Peptide_03
Peptide_027	tIDRTVIYNGK	400 fmol	Peptide_03
Peptide_028	EIARtALVNLVSDGR	400 fmol	Peptide_36
Peptide_029	EIARTALVNLVsDGR	400 fmol	Peptide_36
Peptide_030	PGLRVyAK	400 fmol	Peptide_37
Peptide_031	RGDLPsSFPK	400 fmol	Peptide_06
Peptide_032	IQDNALtRQFR	400 fmol	Peptide_07
Peptide_033	VLtYDLRGHGK	400 fmol	Peptide_09
Peptide_034	VLTyDLRGHGK	400 fmol	Peptide_09
Peptide_035	QRIySIK	400 fmol	Peptide_10
Peptide_036	QRIYsIK	800 fmol	Peptide_10
Peptide_037	HtYQNEPLPNRK	400 fmol	Peptide_12
Peptide_038	HTyQNEPLPNRK	400 fmol	Peptide_12
Peptide_039	AVADRLENsEAK	400 fmol	Peptide_13
Peptide_040	NMFIRPEsK	400 fmol	Peptide_14
Peptide_041	RIIPsDK	800 fmol	Peptide_15
Peptide_042	IEtTEARAK	400 fmol	Peptide_01

Table S3: Chromosomal DNA templates from S. aureus and used primers

#	sequence	name
2410	acagcgCATATGataggtaaaataataaatgaacgatataaaattgtag	Fw_NdeI-PknB/Stk1 (SAUSA300_RS06025)
2411	acagcgGGATCCttatacatcatcatagctgacttctttttc	Rv_BamHI-PknB/Stk1 (SAUSA300_RS06025)
2412	acagcgCATATGctagaggcacaatttttttactgata	Fw_NdeI-Stp1 (SAUSA300_RS06020)
2413	acagcgGGATCCtcatactttatcaccttcaatagccg	Rv_BamHI-Stp1 (SAUSA300_RS06020)

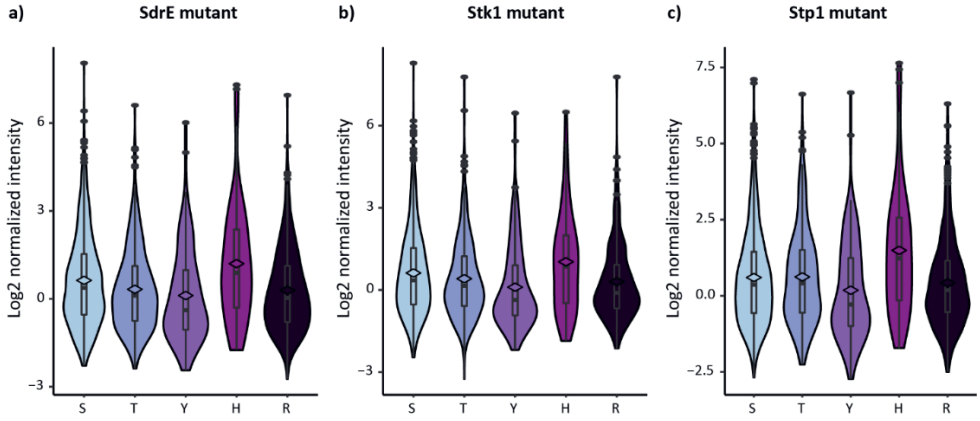
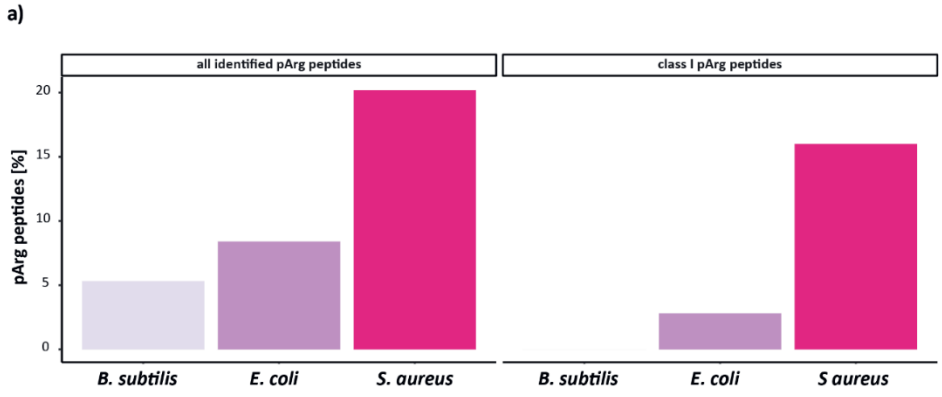


Figure S1: Violin plot showing the intensity distribution for Ser, Thr, Tyr, His and Arg phosphosites for **a)** the control SdrE mutant, **b)** the Stk1 mutant and **c)** the Stp1 mutant. Intensities were log2 transformed and median normalized for at least three valid values for n=4 biological replicates. The mean values are highlighted as diamond. The median, first and third quantile are depicted in the boxplot. Outliers are highlighted as individual points.



b) McsB alignment *S.aureus* vs *B. Subtilis*

S. aureus	1	MTNHIHDNISQWMK-SNEET PIVMSSRIRLARNLENHVHP LMYATENDGFRVINEVQD--57	
B. subtilis	3	LKHFIQDALSSWMKQKGPES DIVLSSRIRLARNFEHIRFP TRYSNE-EASSIIQQFEDQF61	
S. aureus	58	------ALPNFELMRLDQMD QQSKMKMVAKHLISPELIKQ PAAAVLVNDDESLSVMINEE111	
B. subtilis	52	SEQEIPGIGKFVLIRMNDAQ PLEKRVLVEKHLISPNLIES PFGGCLLSENEEVSVMLNEE121	
S. aureus	112	DHIRIQAMGTDTTLQALYNQ ASSIDDELDRSLDISYDEQL GYLTTCPTNIGTGMRASVML171	
B. subtilis	122	DHIRIQCLFPGFQLLEAMKA ANQVDDWIEEKVDYAFNEQR GYLTSCPTNVGTGLRASVMM181	
S. aureus	172	HLPGLSIMKRMTRIAQTINR FGYTIRGIYGEGSQVYGHTY QVSNQLTLGKSELEIIETLT231	
B. subtilis	182	HLPALVLTRQINRIIPAINQ LGLVVRGIYGEGSEAVGNIF QISNQITLGKSEQDIVEDLN241	
S. aureus	232	EVVNQIIHEEKQIRQKLDTY NQLETQDRVFRSLGILQNCR MITMEEASYRLSEVKLGIDL291	
B. subtilis	242	SVAAQLIEQERSAREAIYQT SKIELEDRVYRSYGVLSNCR MIESKETAKCLSDVRLGIDL301	
S. aureus	292	NYIE-LQNFKFNELMVAIQS PFLDEEDDKSVKEKRADIL R 331	
B. subtilis	303	GIIKGLSSNILNELMILTQ- PGFLQQYSGGALRPNERDIR R 341	

c) McsB alignment *S.aureus* vs *E. coli*

S. aureus	1	MTHNIHDNISQWMKSNEETP IVMSSRIRLARNLENHVHPL MYATENDGFRVINEVQDALP 60	
E. coli	1	MTHNLDQHMSDWMKKVNAQP VIMSSRIRLARNLENYVHPL MFPSEAEGHRVINEVQDVLT 60	
S. aureus	61	NFELMRLDQMDQQSKMKMVA KHLISPELIKQPAAAVLVND DESLSVMINEEDHIRIQAMG 120	
E. coli	61	DLKVLRLDELDQLSKYKLVA KHLISPELTKQPASAVLLNE DESVSVMVNEEDHIRIQAMG 120	
S. aureus	121	TDTTLQALYNQASSIDDELD RSLDISYDEQLGYLTTCPTNI GTGMRASVMLHLPGLSIMK 180	
E.coli	121	NDLSLNELYQKASAIDDILD RTLQISYDEQLGYLTTCPTNV GTGMRASVMLHLPGLSIMK 180	
S. aureus	181	RMTRIAQTINRFGYTIRGIY GEGSQVYGHTYQVSNQLTLG KSELEIIETLTEVVNQIIHE 240	
E. coli	181	RMNRIAQTINRFGFTIRGIY GEGSQVYGHIYQVSNQLTLG KSEQEIIEALTELVDQIIHE 240	
S. aureus	241	EKQIRQKLDTYNQ 253	
E. coli	241	ELALRERLNEHNH 253	

Figure S2: Number of total pArg and class I (Andromeda localization probability > 0.75) pArg phosphosites identified for *B. subtilis*, *E.coli* and *S. aureus* (**a**). Sequence alignment obtained with the NCBI Blastp (https://blast.ncbi.nlm.nih.gov/Blast.cgi) searching for *S. aureus* MscB homolouges in, *B. subtilis* (**b**) and *E. coli* (**c**).

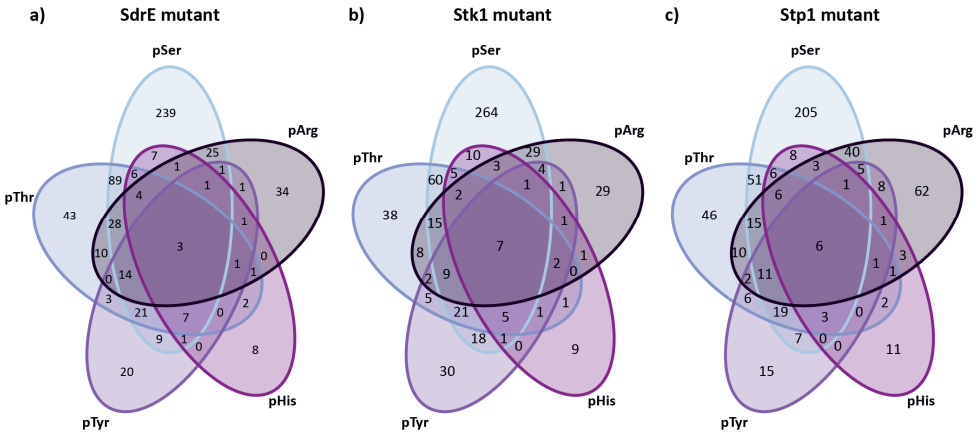


Figure S3: A Venn-diagram depicting the overlap of Ser, Thr, Tyr, His and Arg phosphorylated proteins in *S. aureus* is showing a unique pArg proteome. Phosphosites were filtered for at least three valid values (n=4 biological replicates).

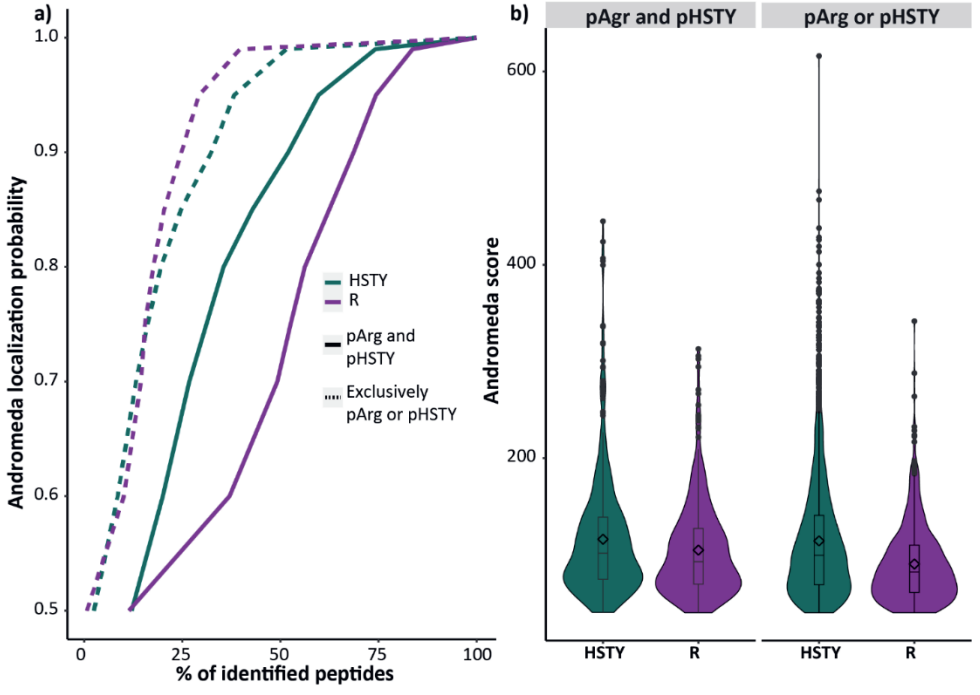


Figure S4: Comparison of **a)** Andromeda localization probability and **b)** Andromeda score for endogenous peptides that were either exclusively identified being phosphorylated on arginine and pArg peptides that were also identified to be phosphorylated on H, S, T, or Y. In total 386 pArg peptides out of 1,294 identified pArg peptides (three valid values in at least one group) were also identified to be phosphorylated on HSTY (n=4).

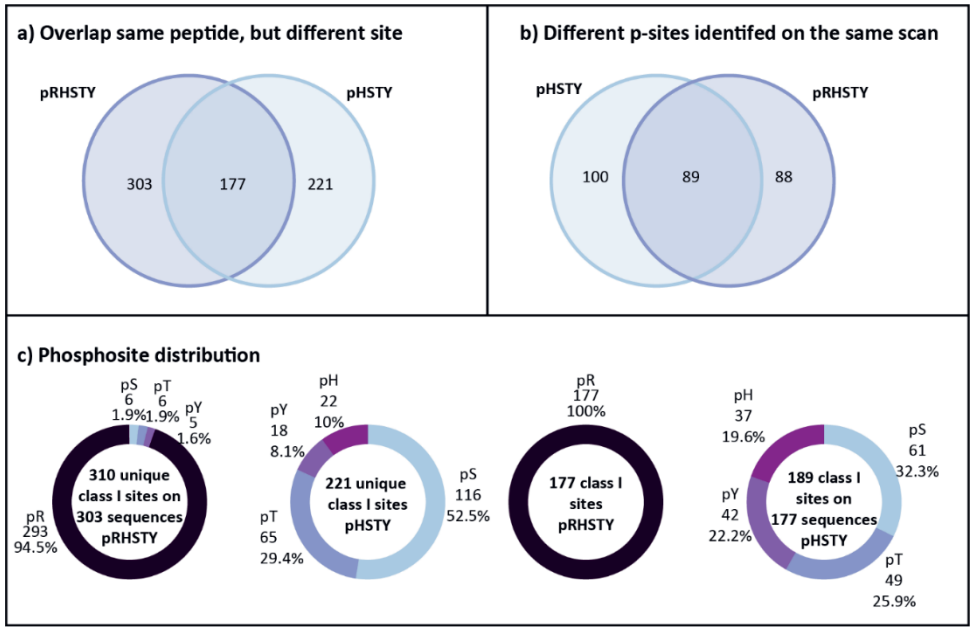


Figure S5: Comparison of identified class I p-sites after MaxQuant search for pRHSTY and pHSTY for the control mutant (n=4). **a)** Overlap of exclusively identified p-sites after pRHSTY or pHSTY search based on the peptide sequence, multiple sequences were not considered for the overlap. **b)** Overlap of peptides identified being phosphorylated on different sites, but based on the same MS2 spectrum. **c)** p-site distribution for the 303 and 221 exclusively identified phosphorylated peptides giving rise to 310 and 221 class I p-sites after pRHSTY and pHSTY search respectively. 177 peptide sequences were identified being phosphorylated in both searches. 12 peptides were identified being phosphorylated on multiple possible pHSTY sites, resulting in total 189 class I site.

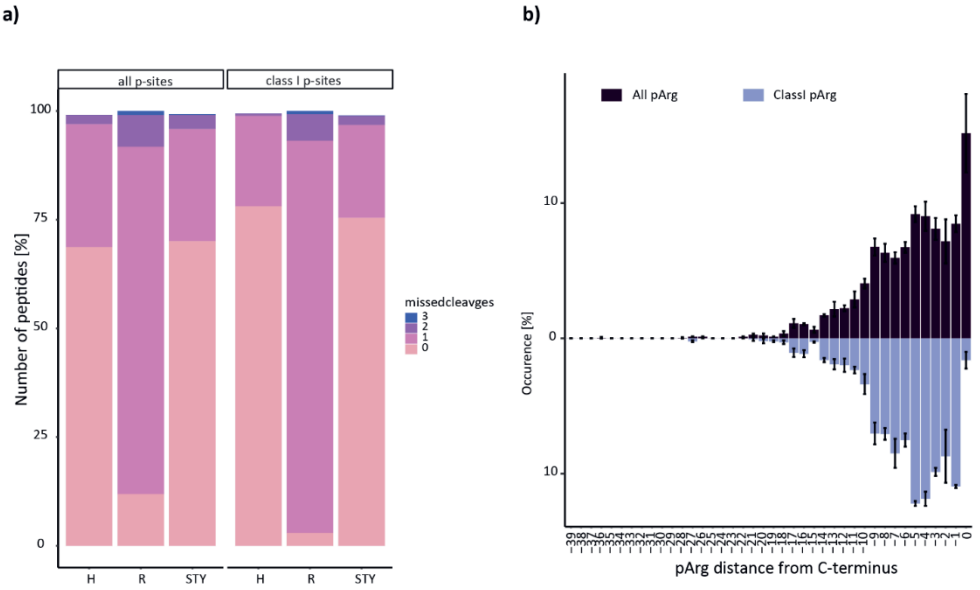


Figure S6: Missed cleavages and pArg localization. **a)** Percentage of endogenous phosphorylated histidine (pHis), arginine phosphorylated (pArg) and serine, threonine or tyrosine phosphorylated (pSTY) peptides identified with zero, one, two or three missed cleavages. **b)** Distance of pArg from C-terminus (0) for all pArg sites (purple) and class I pArg (violet).

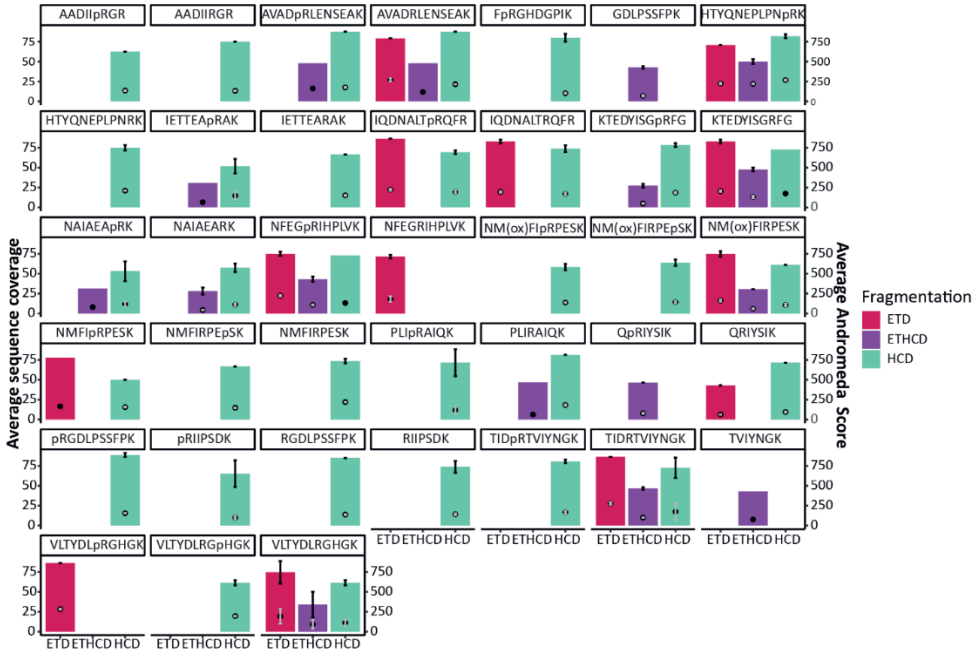


Figure S7: Fragmentation characteristics of synthetic pArg peptides. The average sequence coverage for ETD (pink), EThcD (purple) and HCD (turquoise) for all identified synthetic peptides is shown. Phosphorylation sites are indicated by "p" in front of the p-site. Oxidation of methionines is indicated as (ox). For all bar charts the center values represent the mean and error bars the s.d. for n=4. The average Andromeda score is shown as dot plot inside the bar charts. For all dot plots the center values represent the mean and error bars the s.d. for n=4.

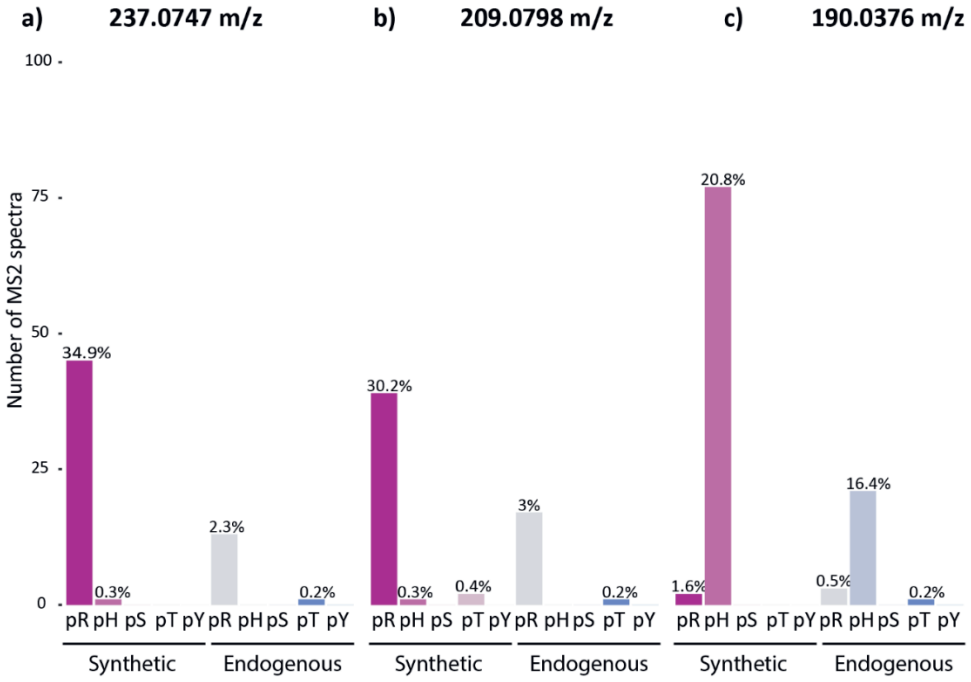
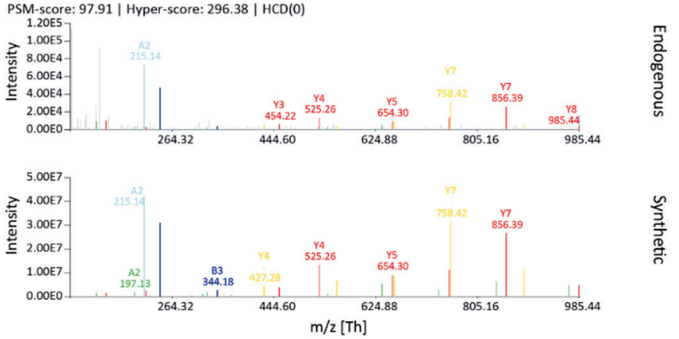


Figure S8: Number of MS2 spectra matching pRHSTY presenting the ions **a)** 237.0747 m/z, **b)** 209.0798 and **c)** the pHis immonium ion of 190.0376 m/z for synthetic peptides and endogenous peptides. The percentage of total MS2 spectra matching the respective phosphorylated amino acids is stated above the bar.

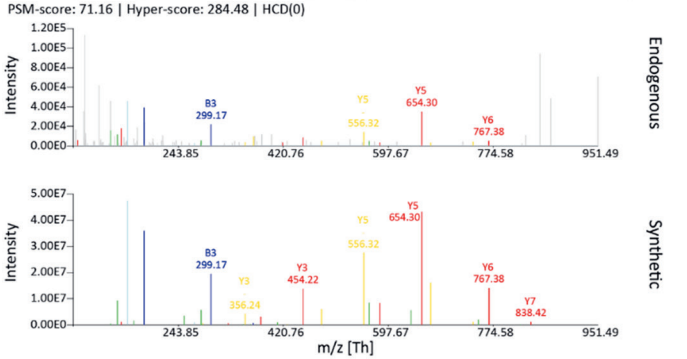
Peptide 01 - Q2FER6

20190215_OR15_Ag1_Prust001_SA_2mg_Stp1_1_IMAC_exp7 SN=7258 RT=19.73 MZ=549.76318 Charge=2+
PSM-score: 97.91 | Hyper-score: 296.38 | HCD(0)



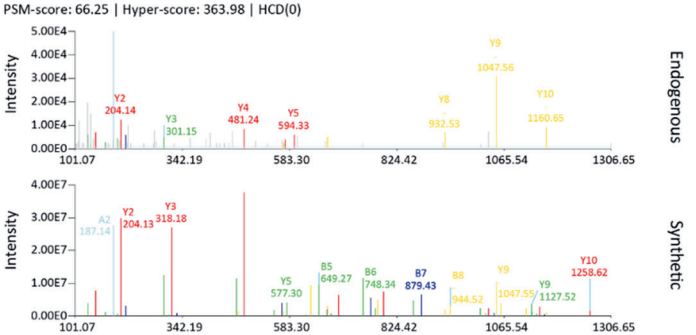
Peptide 02 - Q2FHI2

20190215_OR15_Ag1_Prust001_SA_2mg_Stp1_1_IMAC_exp7 SN=8428 RT=21.06 MZ=476.73422 Charge=2+
PSM-score: 71.16 | Hyper-score: 284.48 | HCD(0)



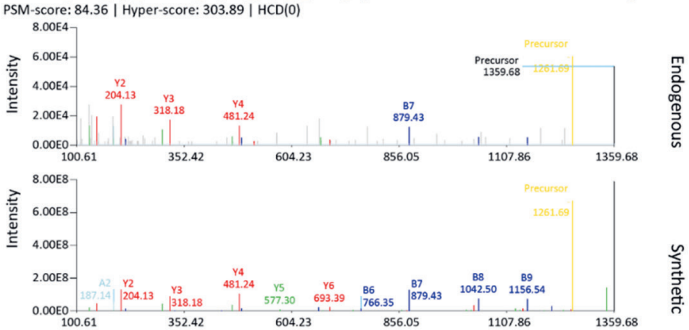
Peptide 03 - A0A0H2XHE8

20190215_OR15_Ag1_Prust001_SA_2mg_Stp1_4_IMAC_exp7 SN=25405 RT=42.12 MZ=453.89379 Charge=3+
PSM-score: 66.25 | Hyper-score: 363.98 | HCD(0)

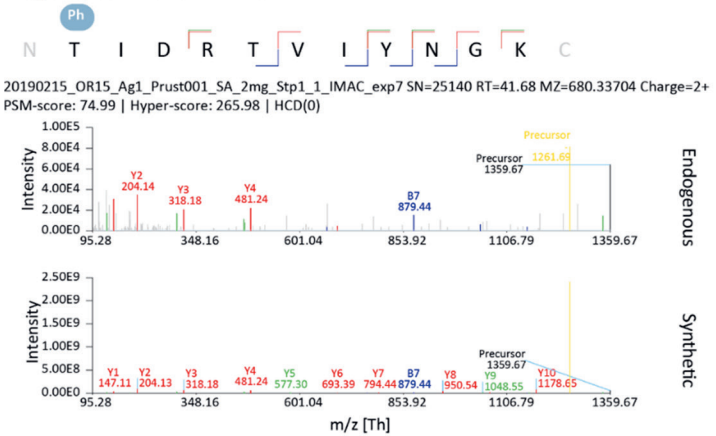


Peptide 03 - A0A0H2XHE8

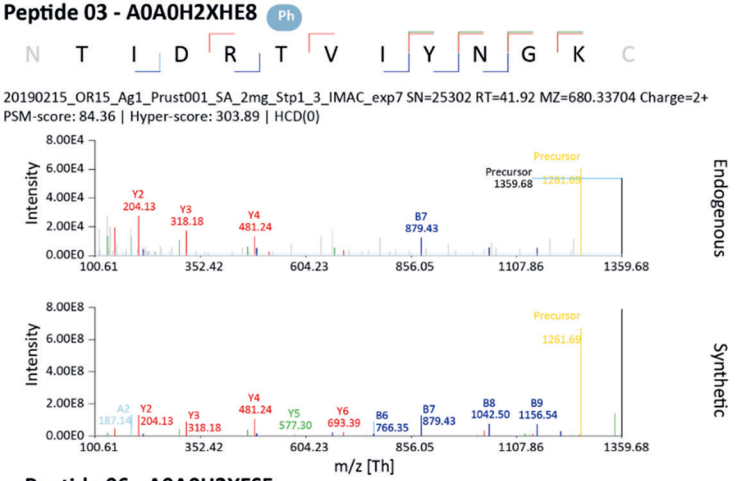
20190215_OR15_Ag1_Prust001_SA_2mg_Stp1_3_IMAC_exp7 SN=25302 RT=41.92 MZ=680.33704 Charge=2+
PSM-score: 84.36 | Hyper-score: 303.89 | HCD(0)



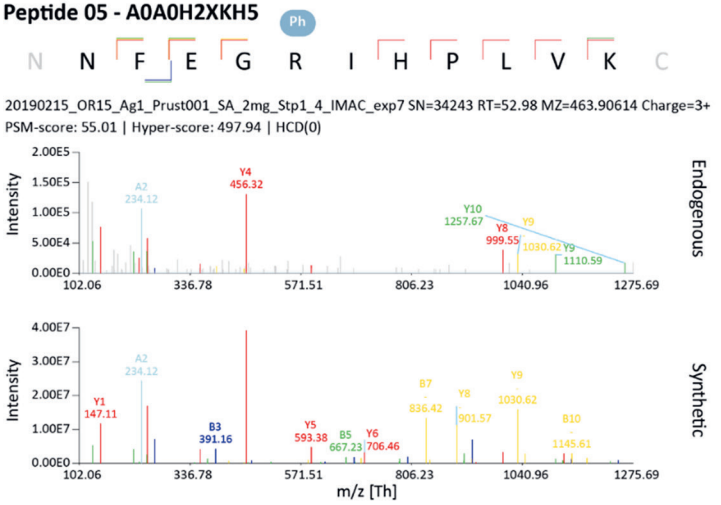
Peptide 03 - A0A0H2XHE8



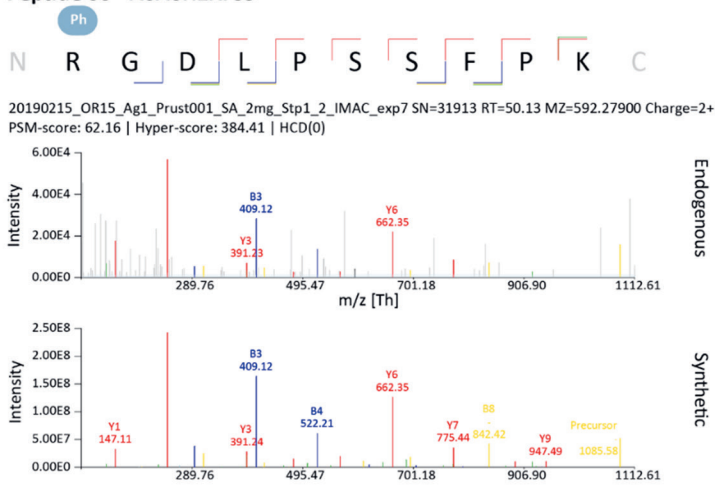
Peptide 03 - A0A0H2XHE8

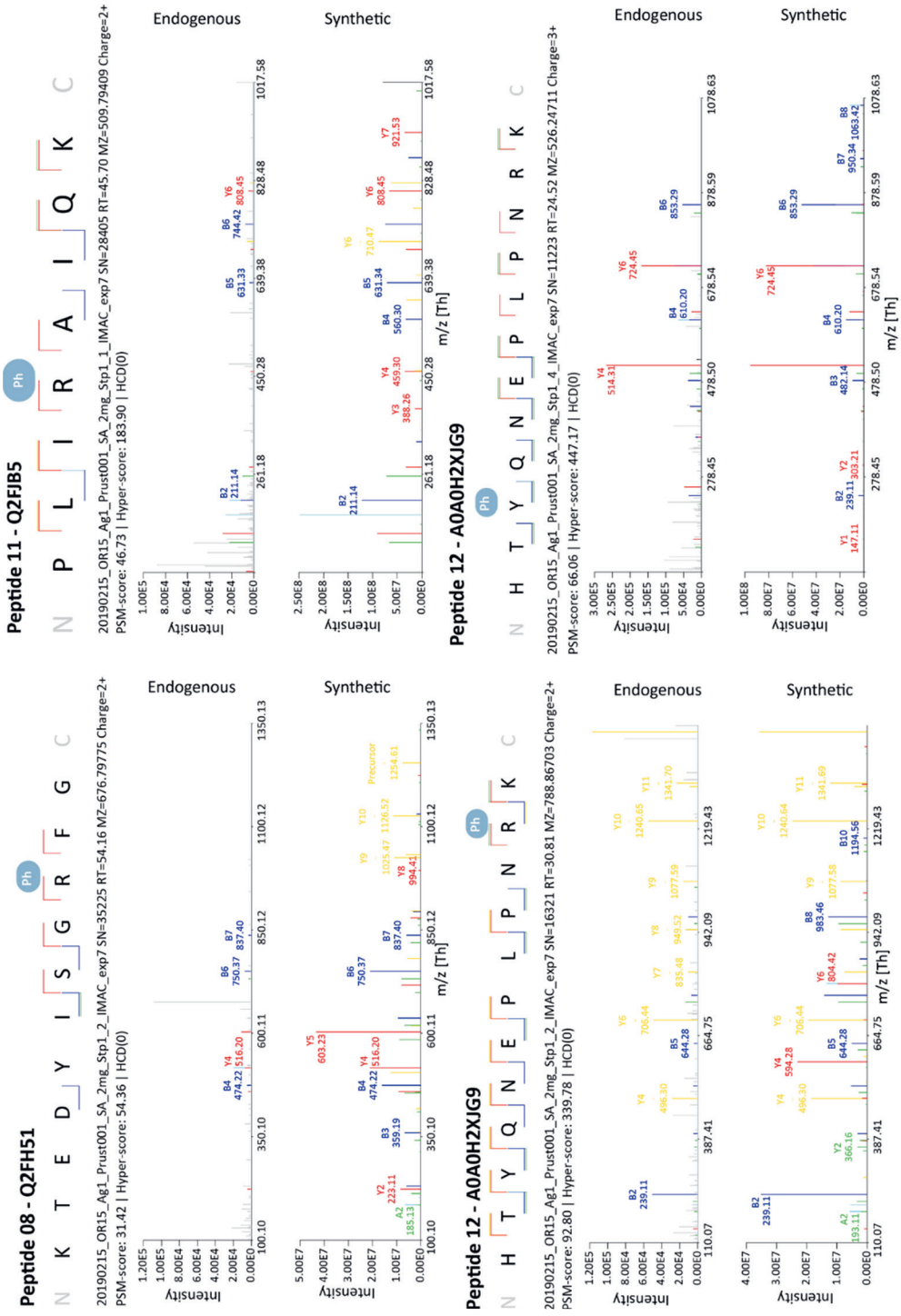


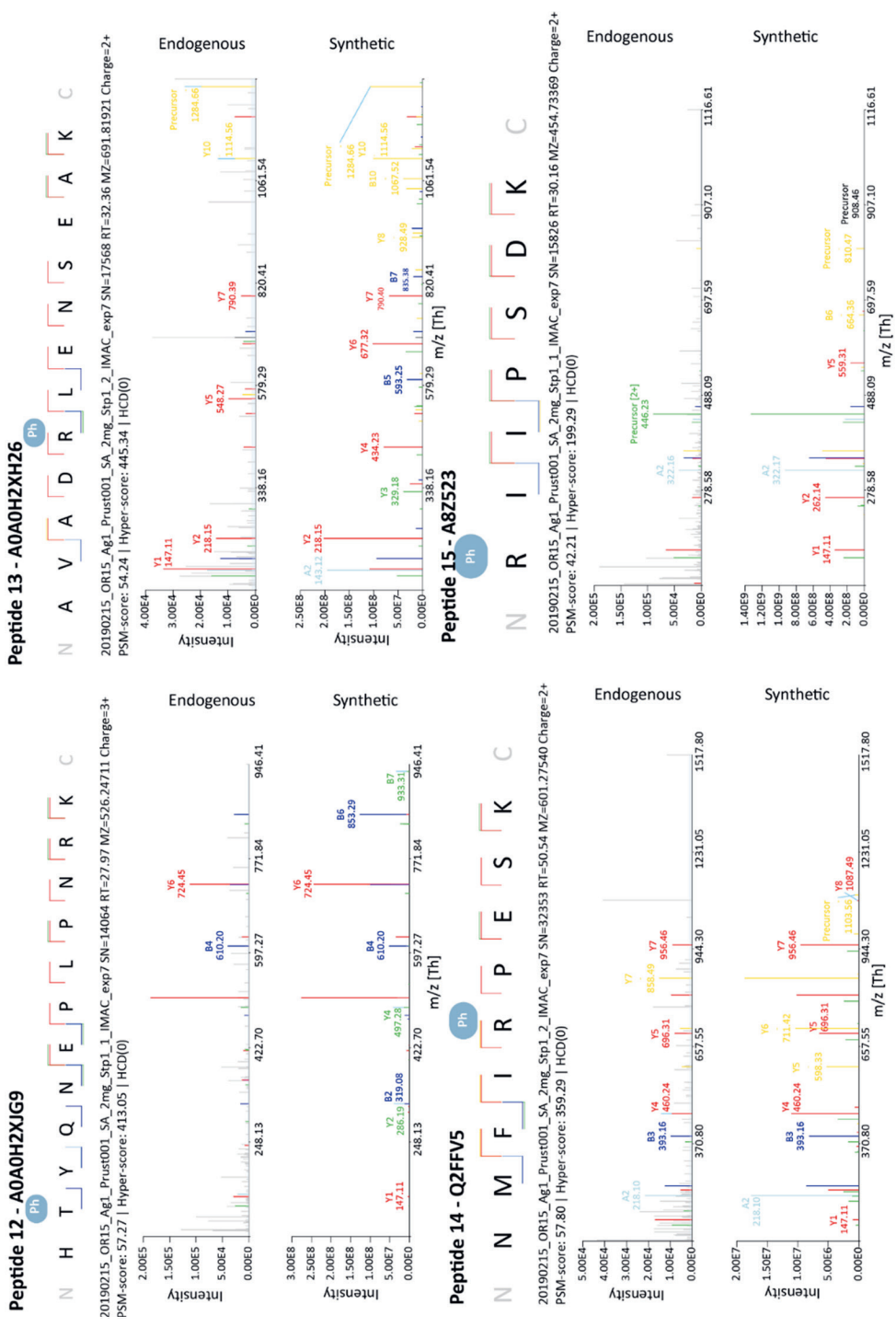
Peptide 05 - A0A0H2XKH5



Peptide 06 - A0A0H2XFS5

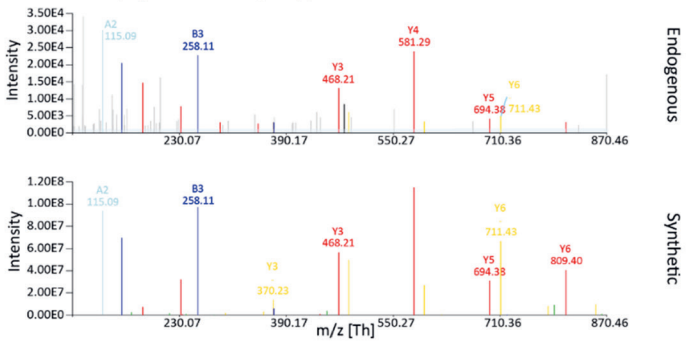



Peptide 08 - Q2FH51
N K T E D Y I S G R F G C
Ph
20190215_OR15_Ag1_Prust001_SA_2mg_Stp1_2_IMAC_exp7 SN=35225 RT=54.16 MZ=676.79775 Charge=2+
PSM-score: 31.42 | Hyper-score: 54.36 | HCD(0)
Endogenous
Synthetic
Intensity
m/z [Th]
Peptide 12 - A0A0H2XJG9
N H T Y Q N E P L P N R K C
Ph
20190215_OR15_Ag1_Prust001_SA_2mg_Stp1_2_IMAC_exp7 SN=16321 RT=30.81 MZ=788.86703 Charge=2+
PSM-score: 92.80 | Hyper-score: 339.78 | HCD(0)
Endogenous
Synthetic
Intensity
m/z [Th]
Peptide 11 - Q2FJB5
N P L I R A I Q K C
Ph
20190215_OR15_Ag1_Prust001_SA_2mg_Stp1_1_IMAC_exp7 SN=28405 RT=45.70 MZ=509.79409 Charge=2+
PSM-score: 46.73 | Hyper-score: 183.90 | HCD(0)
Endogenous
Synthetic
Intensity
m/z [Th]
Peptide 12 - A0A0H2XJG9
N H T Y Q N E P L P N R K C
Ph
20190215_OR15_Ag1_Prust001_SA_2mg_Stp1_4_IMAC_exp7 SN=11223 RT=24.52 MZ=526.24711 Charge=3+
PSM-score: 66.06 | Hyper-score: 447.17 | HCD(0)
Endogenous
Synthetic
Intensity
m/z [Th]


Peptide 12 - A0A0H2XJG9
Peptide 13 - A0A0H2XH26
Peptide 14 - Q2FFV5
Peptide 15 - A8Z523
Endogenous
Synthetic
Intensity
m/z [Th]

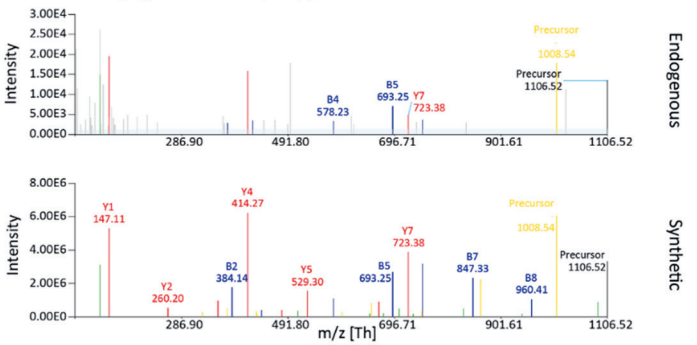
Peptide 16 - Q2FDP9

20190215_OR15_Ag1_Prust001_SA_2mg_Stp1_4_IMAC_exp7 SN=17561 RT=32.38 MZ=476.24222 Charge=2+
PSM-score: 98.94 | Hyper-score: 521.78 | HCD(0)



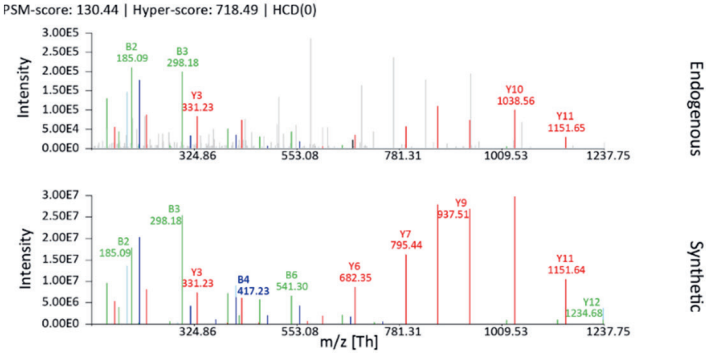
Peptide 17 - Q2FDP9

20190215_OR15_Ag1_Prust001_SA_2mg_Stp1_4_IMAC_exp7 SN=12542 RT=26.15 MZ=553.76077 Charge=2+
PSM-score: 52.55 | Hyper-score: 303.97 | HCD(0)



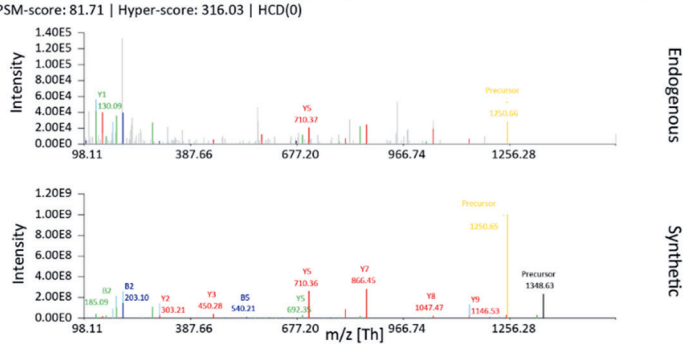
Peptide 18 - Q2FJ92

20190215_OR15_Ag1_Prust001_SA_2mg_Stp1_4_IMAC_exp7 SN=56328 RT=80.93 MZ=677.37309 Charge=2+
PSM-score: 130.44 | Hyper-score: 718.49 | HCD(0)



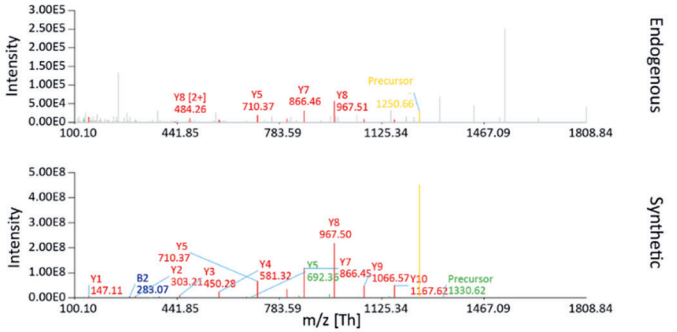
Peptide 19 - Q2FJ92

20190215_OR15_Ag1_Prust001_SA_2mg_Stp1_4_IMAC_exp7 SN=37381 RT=57.02 MZ=674.82017 Charge=2+
PSM-score: 81.71 | Hyper-score: 316.03 | HCD(0)



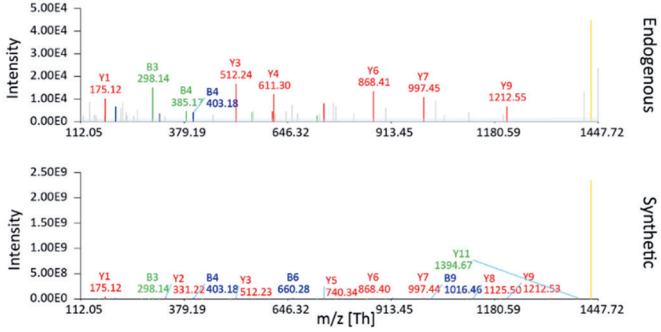
Peptide 19 - Q2FJ92

20190215_OR15_Ag1_Prust001_SA_2mg_Stp1_4_IMAC_exp7 SN=36903 RT=56.34 MZ=674.82017 Charge=2+
PSM-score: 40.55 | Hyper-score: 99.99 | HCD(0)

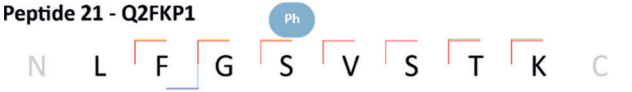


Peptide 20 - Q2FFK2

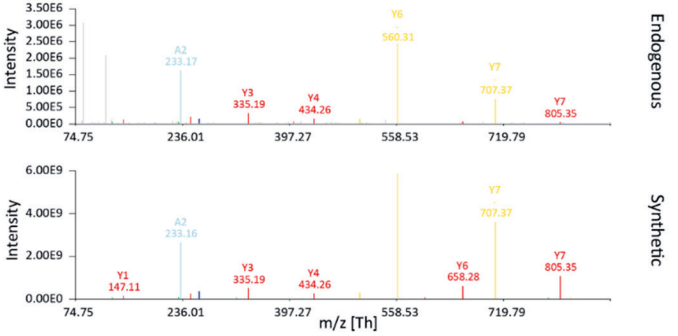
20190215_OR15_Ag1_Prust001_SA_2mg_Stp1_4_IMAC_exp7 SN=18694 RT=33.85 MZ=764.34358 Charge=2+
PSM-score: 89.65 | Hyper-score: 305.89 | HCD(0)



Peptide 21 - Q2FKP1

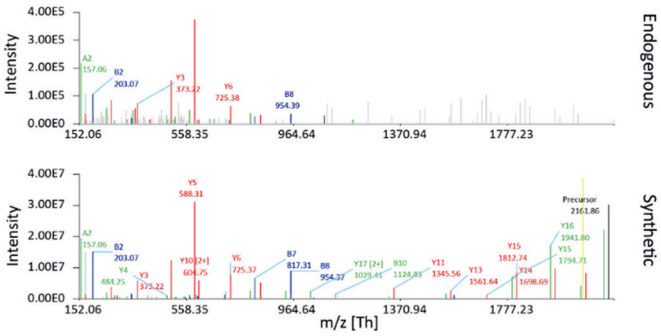


20190215_OR15_Ag1_Prust001_SA_2mg_Stp1_3_IMAC_exp7 SN=31869 RT=50.01 MZ=459.72025 Charge=2+
PSM-score: 81.87 | Hyper-score: 237.55 | HCD(0)



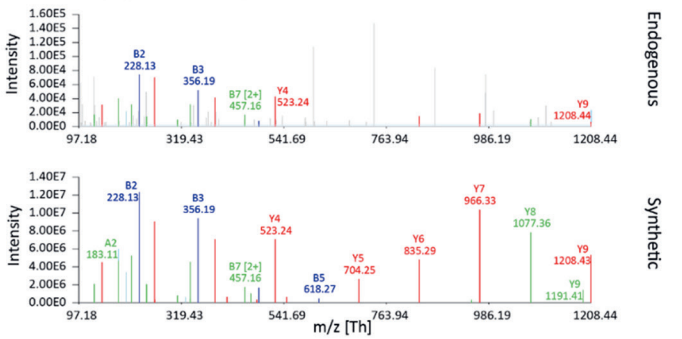
Peptide 22 - A0A0H2XGV4

20190215_OR15_Ag1_Prust001_SA_2mg_Stp1_2_IMAC_exp7 SN=8597 RT=21.14 MZ=541.22254 Charge=4+
PSM-score: 85.25 | Hyper-score: 569.82 | HCD(0)



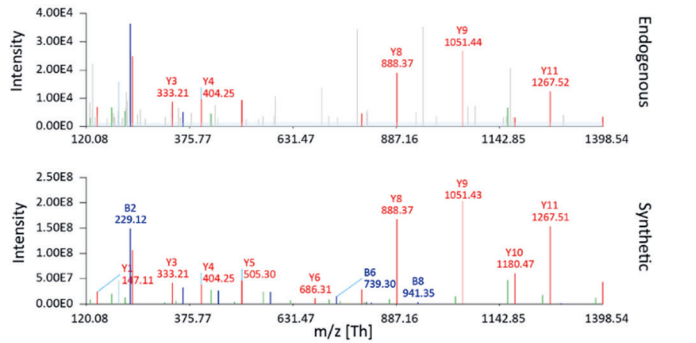
Peptide 23 - A0A0H2XFT2

20190215_OR15_Ag1_Prust001_SA_2mg_Stp1_4_IMAC_exp7 SN=40884 RT=61.22 MZ=661.26226 Charge=2+
PSM-score: 78.90 | Hyper-score: 413.48 | HCD(0)



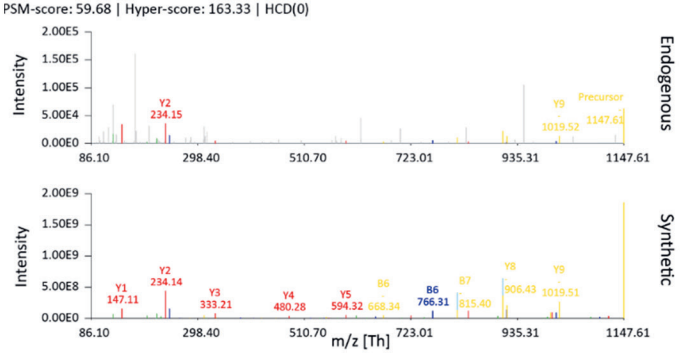
Peptide 26 - Q2FDV8

20190215_OR15_Ag1_Prust001_SA_2mg_Stp1_2_IMAC_exp7 SN=31635 RT=49.74 MZ=813.83387 Charge=2+
PSM-score: 69.33 | Hyper-score: 623.68 | HCD(0)



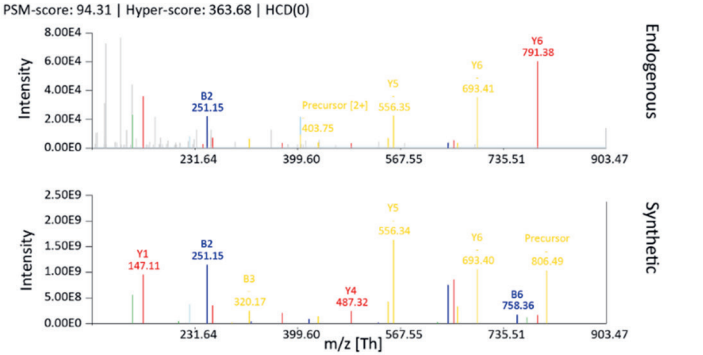
Peptide 27 - Q2FG72

20190215_OR15_Ag1_Prust001_SA_2mg_Stp1_3_IMAC_exp7 SN=22947 RT=39.04 MZ=623.29739 Charge=2+
PSM-score: 59.68 | Hyper-score: 163.33 | HCD(0)



Peptide 28 - A0A0H2XED3

20190215_OR15_Ag1_Prust001_SA_2mg_Stp1_3_IMAC_exp7 SN=24006 RT=40.29 MZ=452.73624 Charge=2+
PSM-score: 94.31 | Hyper-score: 363.68 | HCD(0)

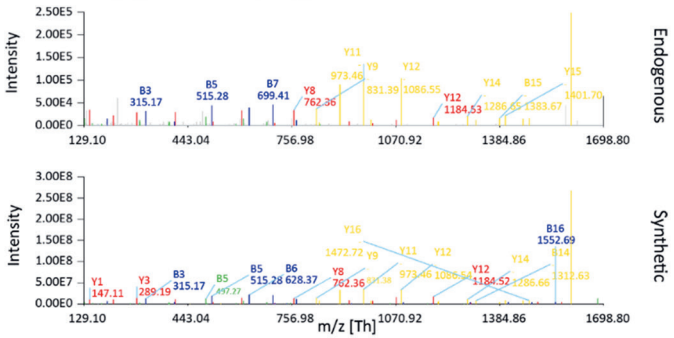


Peptide 29 - A0A0H2XGH3

N K A D S(Ph) I L A A S G V S D D A A K C

20190215_OR15_Ag1_Prust001_SA_2mg_Stp1_1_IMAC_exp7 SN=31085 RT=49.12 MZ=849.90093 Charge=2+

PSM-score: 273.46 | Hyper-score: 547.81 | HCD(0)

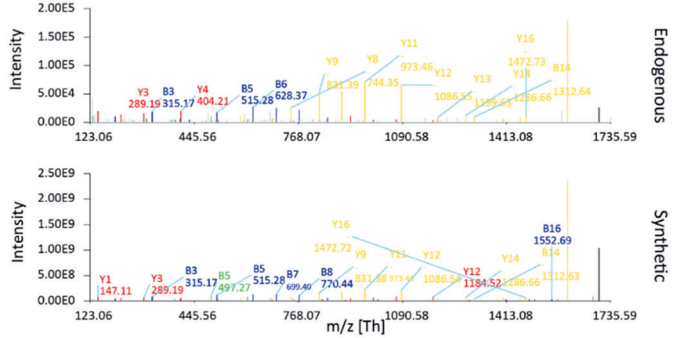


Peptide 29 - A0A0H2XGH3

N K A D S(Ph) I L A A S G V S D D A A K C

20190215_OR15_Ag1_Prust001_SA_2mg_Stp1_3_IMAC_exp7 SN=32353 RT=50.71 MZ=849.90093 Charge=2+

PSM-score: 229.46 | Hyper-score: 486.81 | HCD(0)

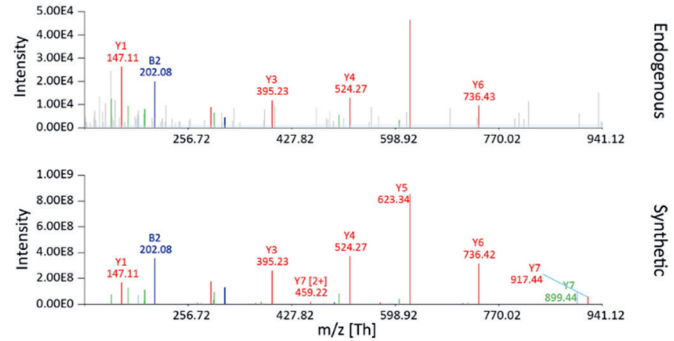


Peptide 30 - A0A0E1VNI4

N S N D(Ph) T I V E T F K C

20190215_OR15_Ag1_Prust001_SA_2mg_Stp1_1_IMAC_exp7 SN=38486 RT=58.07 MZ=617.27358 Charge=2+

PSM-score: 45.43 | Hyper-score: 414.67 | HCD(0)

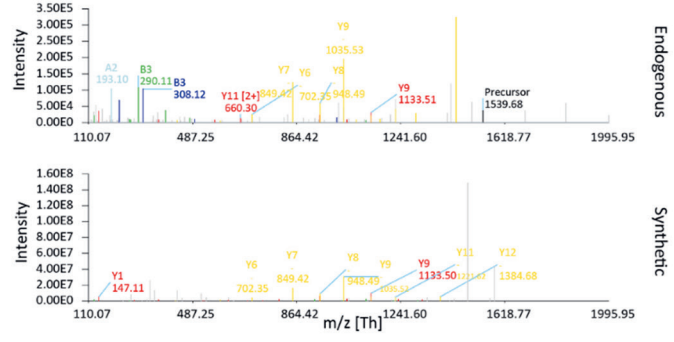


Peptide 31 - A0A0H2XGG1

N G Y S(Ph) V S V F N R S S E K C

20190215_OR15_Ag1_Prust001_SA_2mg_Stp1_1_IMAC_exp7 SN=34986 RT=53.70 MZ=770.34560 Charge=2+

PSM-score: 128.74 | Hyper-score: 440.19 | HCD(0)

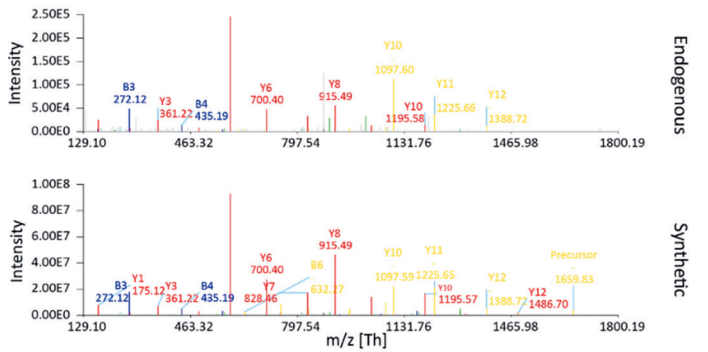


Peptide 32 - A0A0H2XGY0

N G D V Y Q S L S Q L P E S V R C (Ph)

20190215_OR15_Ag1_Prust001_SA_2mg_Stp1_1_IMAC_exp7 SN=48734 RT=70.78 MZ=879.40892 Charge=2+

PSM-score: 107.33 | Hyper-score: 425.68 | HCD(0)

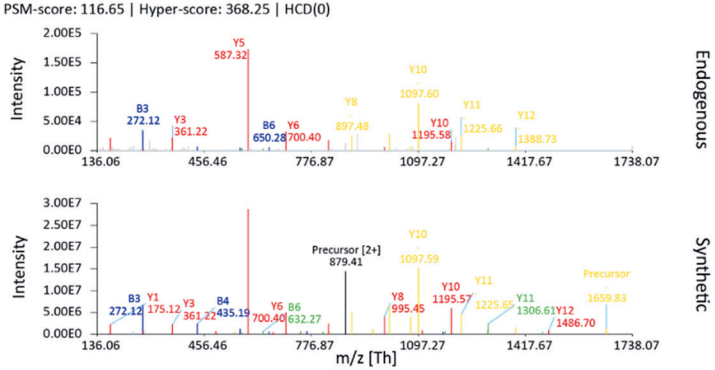


Peptide 32 - A0A0H2XGY0

N G D V Y Q S L S Q L P E S V R C (Ph)

20190215_OR15_Ag1_Prust001_SA_2mg_Stp1_3_IMAC_exp7 SN=48973 RT=70.95 MZ=879.40892 Charge=2+

PSM-score: 116.65 | Hyper-score: 368.25 | HCD(0)

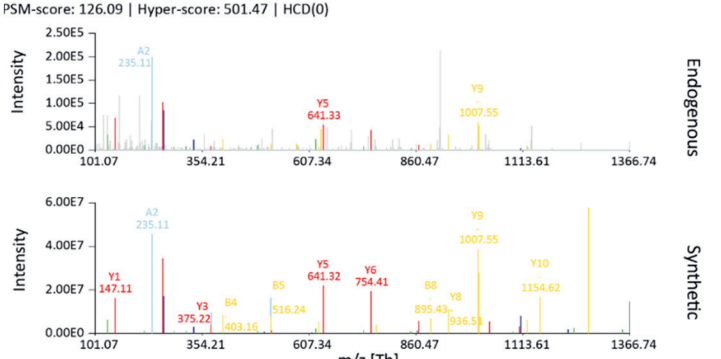


Peptide 33 - Q2FEP5

N D F A S L L H E D L K C (Ph)

20190215_OR15_Ag1_Prust001_SA_2mg_Stp1_4_IMAC_exp7 SN=56340 RT=80.88 MZ=684.31578 Charge=2+

PSM-score: 126.09 | Hyper-score: 501.47 | HCD(0)

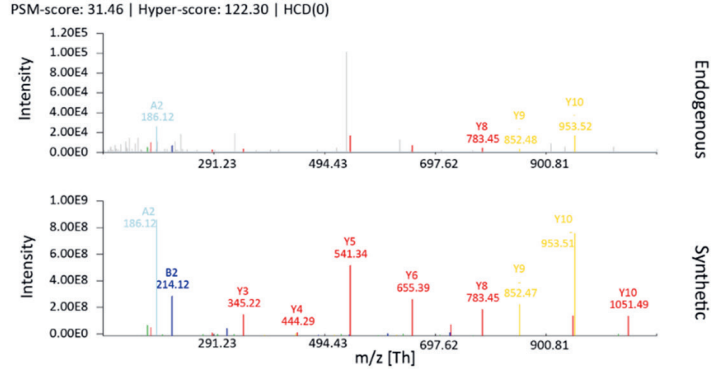


Peptide 34 - Q2FF95

N N V T S G A N P V G L R C (Ph)

20190215_OR15_Ag1_Prust001_SA_2mg_Stp1_2_IMAC_exp7 SN=31129 RT=49.03 MZ=632.80591 Charge=2+

PSM-score: 31.46 | Hyper-score: 122.30 | HCD(0)



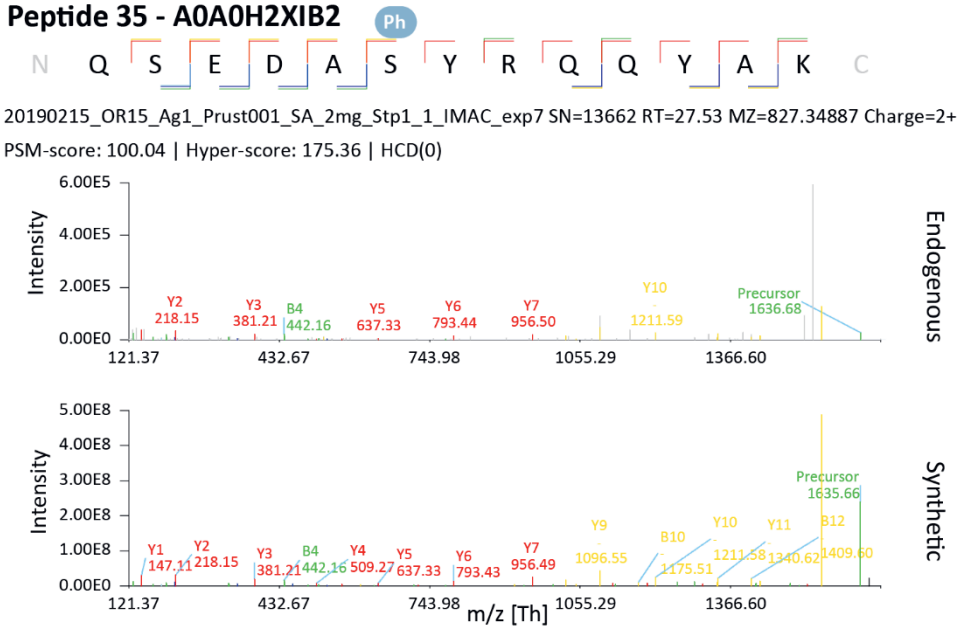


Figure S9: Spectra comparison for all identified synthetic and endogenous pArg as well pSTY peptides using FragmentLab (https://scheltemalab.com/software) (v2.4.1.0). The synthetic peptides were used to create a peptide library that was used to validate the endogenous identified peptides.

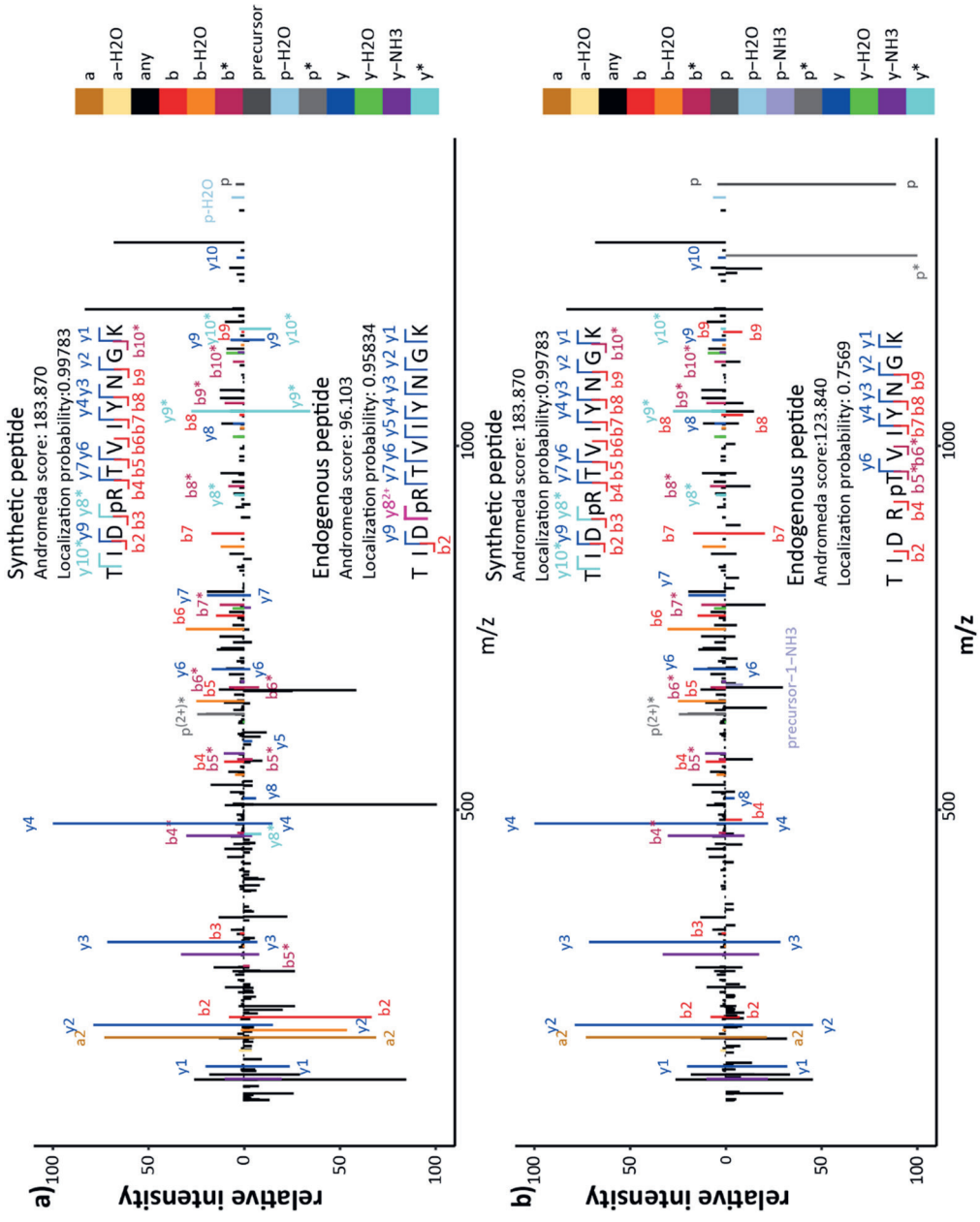


Figure S10: Comparison of the acquired fragmentation spectra of synthetic and the respective endogenous peptide for SAUSA300_1449 phosphorylated on **a)** Arg and **b)** Thr. B- and y-ions are highlighted in red and blue.

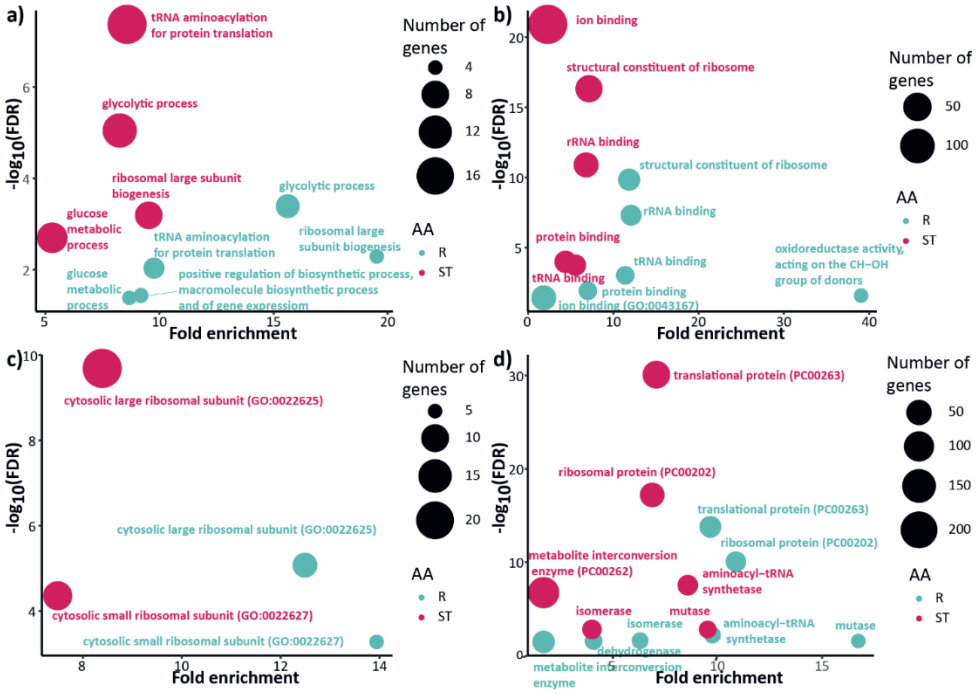


Figure S11: Statistical overrepresentation of GO-terms and protein classes for all identified pArg as well as pSer/Thr proteins after Fischer's exact test using Panther (vers. 16.0). All identified pArg proteins or pSer/Thr proteins with official gene name (74 pArg genes and 303 pSer/Thr genes) were compared to the whole genome. The $-\log_{10}$(FDR) is plotted as function of the Fold enrichment for the main GO-terms for **a)** biological process, **b)** molecular function, **c)** cellular compartment and **d)** Panther protein classes. The size of points is corresponding to the number of genes contributing to the respective GO term or protein classes. The enrichment for pArg and pSer/Thr is highlighted in turquois and pink respectively.

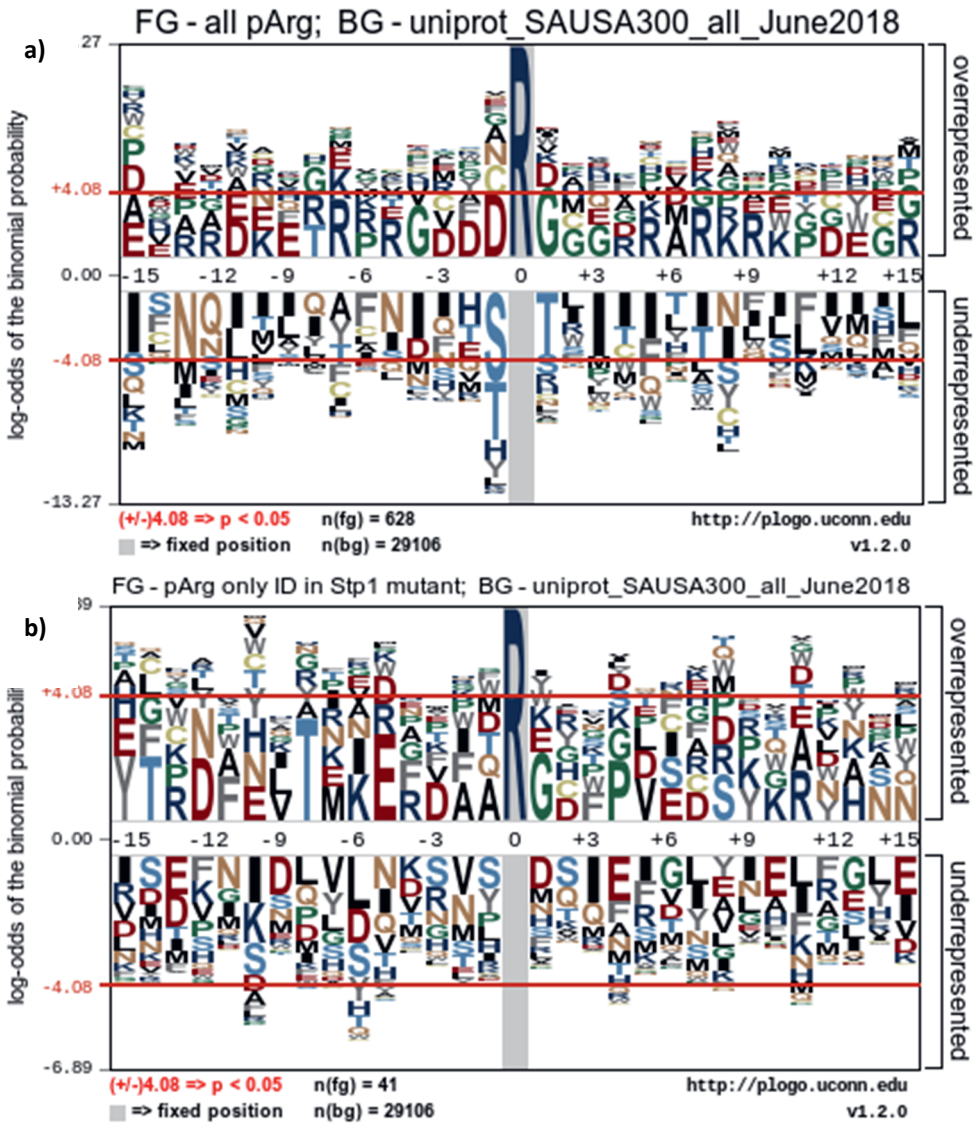


Figure S12: Sequence motifs generated using pLogo[1] for **a)** all identified arginine phosphorylated peptides and **b)** arginine phosphorylated peptides only identified in the Stp1 mutant. The whole *S. aureus* USA300 proteome was used a background.

[1] O'Shea JP, Chou MF, Quader SA, Ryan JK, Church GM, Schwartz D. PLogo: A probabilistic approach to visualizing sequence motifs. *Nat Methods*. 2013;10(12):1211-1212. doi:10.1038/nmeth.2646

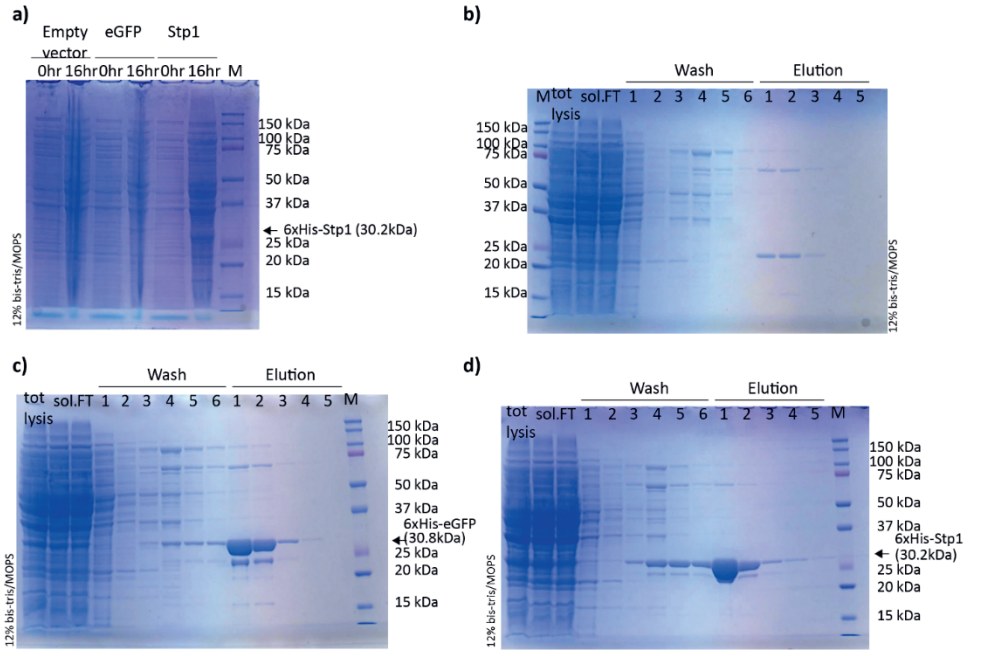
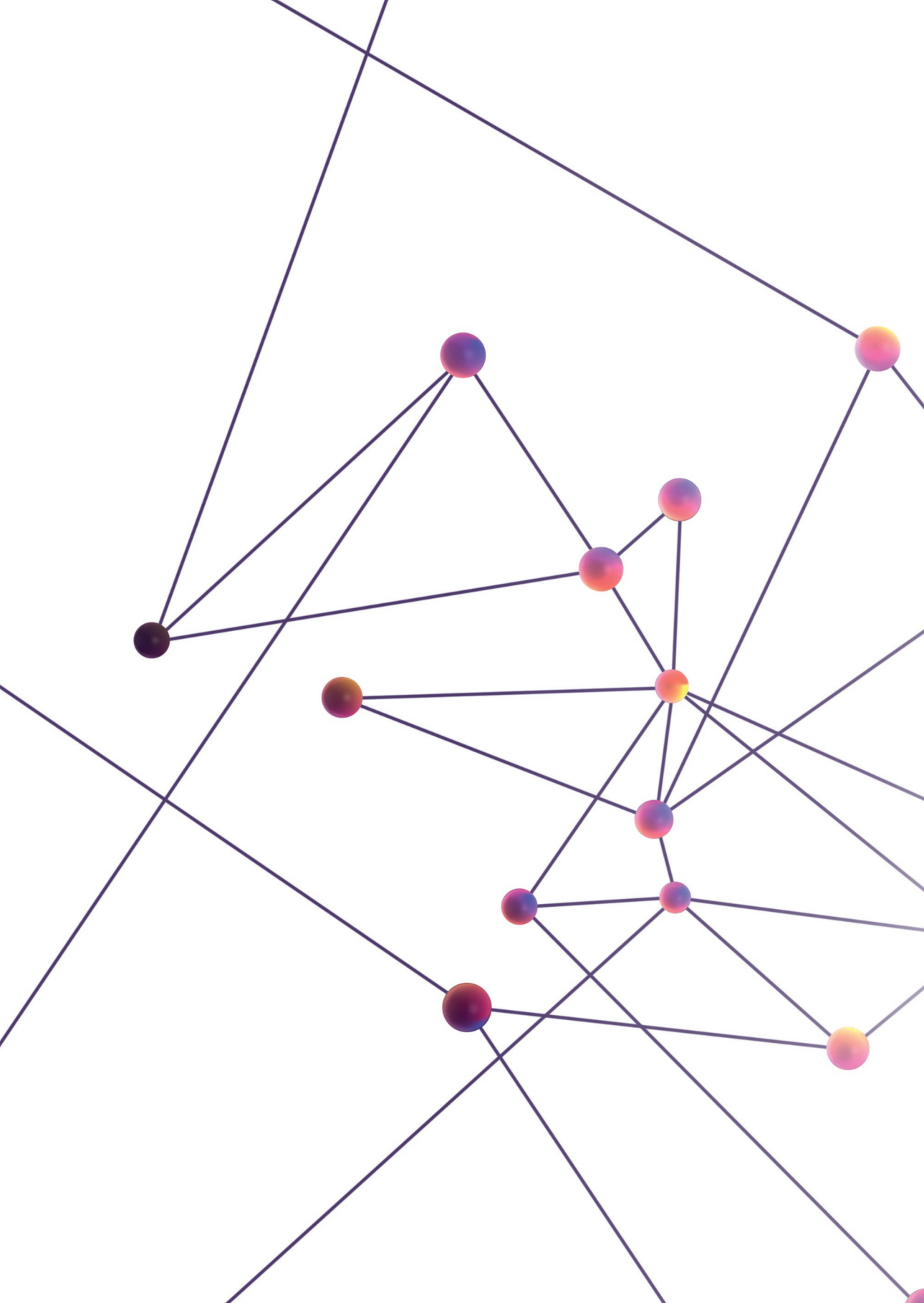


Figure S13: Imperial™ Protein Stain (Thermo Fischer Scientific) stained SDS-PAGE. **a)** 0hr and 16hrs after protein induction with 1mM IPTG for the empty vector, eGFP and Stp1. SDS-PAGE after purification with Ni-NTA beads of **b)** the empty vector, **c)** eGFP and **d)** Stp1. 10ul of the whole lysate, the soluble faction, flow through, wash 1-5 and eluate 1-5 were loaded together with 10ul 2x XT sample buffer (BIO-RAD) containing 25mM final concentration DTT.



Chapter 4

Challenges in bacterial phosphoproteomics – One size does not fit all

Nadine Prust[1,2] and Simone Lemeer[1,2]

[1] Biomolecular Mass Spectrometry and Proteomics, Bijvoet Center for Biomolecular Research and Utrecht Institute for Pharmaceutical Sciences, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands
[2] Netherlands Proteomics Center, Padualaan 8, 3584 CH Utrecht, The Netherlands

Abbreviations

ACN	Acetonitrile
AmBiC	Ammonium bicarbonate
Arg	Arginine
Asp	Aspartic acid
CAA	2-chloroacetamide
cAMP	cyclic adenosine monophosphate
ECM	Extracellular matrix
FA	Formic acid
FDR	False discovery rate
HAMMOC	hydroxy acid–modified metal oxide chromatography
HCD	Higher energy collision induced dissociation
His/H	Histidine
IMAC	Immobilized metal ion affinity chromatography
LC	Liquid chromatography
OD	Optical density
p-sites	Phosphorylation sites
SDC	Sodium deoxycholate
Ser/S	Serine
TCEP	tris(2-carboxyethyl)phosphine
TCS	Two-component system
Thr/T	Threonine
Tyr/Y	Tyrosine

Abstract

The ability of prokaryotes to adapt to changing environments and the rapid development of antibiotic resistance has spurred the analysis of bacterial proteomes and phosphoproteomes. In contrast to eukaryotes, the analysis of bacterial phosphoproteomes is far from straightforward. The general low abundance and increased amount of challenging molecular contaminants when compared to especially eukaryotic cell lines has so far hampered comprehensive analysis of bacterial phosphoproteomes. Recently, improved sample preparation protocols for both Gram-negative as well as Gram-positive bacteria have been introduced. These protocols improved general lysis conditions, reduced the amount of contaminating biomolecules, mostly DNA and RNA, and at the same time allowed the analysis of acid-labile protein phosphorylations, such as the phosphoramidates pHis and pArg.

Spurred by this success, in the here described work, we attempted to apply the described protocols to other bacterial species, more specifically the Gram-negative bacteria *P. aeruginosa* and *K. pneumoniae*. Unfortunately, the previously developed protocols proved not be universally applicable and the number of phosphosites identified in the current study are relatively low. Our analysis indicates a persisting source of DNA/RNA contamination, despite DNase/benzonase treatment, which supposedly hampers the fragmentation and subsequent identification. However, this is probably not the only complicating factor, as only a limited correlation between the success of p-site identification and the contamination marker (330.06 *m/z*) could be found. Surprisingly the success of phosphopeptide enrichment was shown not only to be bacterial species, but also bacterial strain specific. Our results indicate that in bacterial phosphoproteomics one size does unfortunately fit not all, and that tailored protocols are needed for different bacterial species, strains and maybe even growth conditions. Hence, these persisting complications in sample preparation protocols for bacterial species continue to hamper comprehensive as well as high-throughput analysis, urging for continuous research and novel technical approaches, to advance this important research.

Introduction

Even though every prokaryote has a preference for a specific natural habitat in regard to for example pH or temperature, most prokaryotes are able to adapt quickly and efficiently to different surroundings[1,2]. This is also important after e.g. host infiltrations, where bacteria are not only exposed to a different pH range and temperature, but also to the host's immune response[3,4]. To infiltrate a host and actually survive e.g. antibiotic treatment, bacteria are dependent on quick alterations in gene transcription and protein expression as well as metabolic changes[4,5]. One regulatory mechanism that facilitates those processes is reversible protein phosphorylation. It was long assumed that protein-histidine (His) and aspartate (Asp) are the main types of protein phosphorylation within prokaryotes[6]. Both are involved in the so-called two-component systems (TCS) that transduce extracellular stimuli into intracellular responses such as changes in gene transcription of virulence factors[6–8]. However, today we know that protein phosphorylation on serine (Ser), threonine (Thr) and tyrosine (Tyr) is playing an equally important role and the abundance of pTyr has been linked to higher virulence[9–13]. Over the last decade great effort has been taken to elucidate the important role of protein phosphorylation in prokaryotes. Nevertheless, certain prerequisites hampered and still hamper the comprehensive analysis of bacterial phosphoproteomes: i) instrumentational limitations ii) the even lower abundance of protein phosphorylation compared to eukaryotes, and iii) sample preparation and enrichment protocols tailored to eukaryotes. Improvements in phosphopeptide fragmentation, mass spectrometric detection and phosphorylation site localization assessment currently allow the identification of thousands of phosphorylation sites, especially for eukaryotes[14–17]. However, the nature and even lower abundance of bacterial phosphorylations require a more tailored approach. Most enrichment protocols were optimized for eukaryotes and thus adapted to enrich for phosphomonoesters (pSer, pThr and pTyr)[18–20]. While those are relatively stable under acidic conditions, phosphoramidates (pHis and pArg) are acid labile[21]. Therefore, the identification of phosphoramidates was long impeded by the acidic conditions during sample preparation, enrichment and LC-MS/MS analysis. Adaptations to less acidic conditions have been proven successful for the identification of both pHis[22] as well as pArg[23,24]. Another issue, that is most likely a cumulative effect of the overall lower abundance of phosphorylation and the cellular differences between prokaryotes and eukaryotes, is a relatively higher contamination of phospholipids and nucleic acid containing biomolecules. Those components contain, similar to phosphopeptides, phosphoryl groups and thus are competing for binding to the e.g. Fe^{3+}-IMAC resin during the enrichment. The lower abundance of phosphopeptides in prokaryotes allows thus a simultaneous enrichment of those components, which has been linked to inefficient phosphopeptide identification[25]. Optimizations in sample preparation to remove those contaminations have been shown to increase the phosphopeptide identification rate by decreasing the

contamination, when applied to *E. coli* K12[25]. We could also show that further optimization of this workflow allowed a drastic increase in phosphopeptide identifications for *S. aureus* USA300[26]. Further optimization was necessary due to the innate differences between Gram-negative and Gram-positive bacteria. Gram-negative bacteria contain a thin peptidoglycan layer that is surrounded by an outer liposaccharide membrane[27]. Gram-positive bacteria on the other hand lack this outer membrane but instead possess a thick peptidoglycan layer[27]. Consequently, Gram-positive bacteria demand harsher cell lysis conditions compared to Gram-negative bacteria. Encouraged by the success for Gram-negative as well as Gram-positive bacteria, we wanted to test the universal application of those two protocols. Therefor, we performed phosphoproteomics analysis on the Gram-negative bacteria *Pseudomonas aeruginosa* PA14 and PAO1, *Klebsiella pneumoniae* MGH as well as *Escherichia coli* DH5a. Our results indicate that for bacterial phosphoproteomics one size unfortunately does not fit all, and species and strain specific contaminations hamper comprehensive phosphoproteome analysis, thereby limiting high-throughput analysis of bacterial phosphoproteomes.

Material and methods

Bacterial culture

Pseudomonas aeruginosa PA14, PA01 *and Klebsiella pneumoniae* MGH were streaked out on blood agar plates and incubated overnight at 37°C. One colony was used to inoculate a 3 ml pre-culture using LB medium and incubated overnight at 37°C and 180 rpm. The pre-culture was diluted in 20 ml LB-medium to reach an optical density, OD_{600} = 0.05. The culture was incubated at 37°C and 180 rpm till an OD_{600}= 2 in the log-phase was reached. Bacteria were harvested by centrifugation at 4,000 x g for 5min and the pellet was washed with ice cold PBS three times to remove remaining LB medium.

Cell lysis

Bacterial cell lysis was performed as described in Potel et al. (2018)[25] or as previously published for Gram-positive bacteria[26] with slight adaptations. In short, one volume bacteria pellet was resuspended in five volume of lysis buffer i) Gram-negative protocol: 100 mM Tris-HCl pH 8.5, 7 M Urea, 5 mM tris(2-carboxyethyl)phosphine (TCEP), 30 mM 2-Chloroacetamide (CAA), 1% (v/v) Triton X-100, 10 U/ml DNase I, 1 mM magnesium chloride (Sigma-Aldrich, Steinheim, Germany), 1% (v/v) benzonase (Merck Millipore, Darmstadt, Germany), 1 mM sodium orthovanadate, phosphoSTOP phosphatases inhibitors (Roche) and complete mini EDTA free protease inhibitors. ii) Gram-positive protocol: 100 mM Tris-HCl pH 8.5, 7 M Urea, 5 mM tris(2-carboxyethyl)phosphine (TCEP), 30 mM 2-Chloroacetamide (CAA), 1% (v/v) Triton X-100, 1% (v/v) sodium deoxycholate (SDC), 10 U/ml DNase I, 1 mM magnesium chloride (Sigma-Aldrich, Steinheim, Germany), 1% (v/v)

benzonase (Merck Millipore, Darmstadt, Germany), 1 mM sodium orthovanadate, phosphoSTOP phosphatases inhibitors (Roche) and complete mini EDTA free protease inhibitors. Complete lysis was reached by sonication for 45 min (20 s ON, 40 s off) using a Bioruptor Plus. Cell debris was removed by ultracentrifugation (140.000 x g for 1 hr at 4 °C). Protein concentration of the supernatant was determined via a Bicinchoninic Acid (BCA) assay. To decrease the SDC concentration to <0.4 %, the supernatant was diluted 2.5 times with dilution buffer (100 mM Tris-HCl pH 8.5, 7 M Urea, 5 mM TCEP, 30 mM CAA, 1 % Triton x-100, 1 mM magnesium chloride (Sigma-Aldrich, Steinheim, Germany), 1 mM sodium orthovanadate, phosphoSTOP phosphatases inhibitors (Roche) and complete mini EDTA free protease inhibitors). Either 1 % (v/v) benzonase, 2% (v/v) benzonase or 1 % (v/v) benzonase plus 5 % DMSO were added to the supernatant mixture and incubated for 2 h at room temperature. Subsequently, methanol/chloroform precipitation was performed as described earlier[25]. The precipitate was then resuspended in digestion buffer (100 mM Tris-HCL pH 8.5, 30 M CAA, 1% (v/v) SDC (Sigma-Aldrich) and 5 mM TCEP). Protein digestion was performed overnight at room temperature using a mix of trypsin and Lys-C in a ratio of 1:25 and 1:100 (w/w), respectively. Protein digests were acidified to pH 3.5 using 10% formic acid (Sigma-Aldrich) and precipitated SDC was removed by centrifugation (1,400 rpm, 5 min). The supernatant was loaded onto C18 Sep-Pak (3cc) resin columns (Waters) for desalting. The loaded samples were washed twice with 0.1 % (v/v) formic acid and bound peptides were eluted with 600 ul 30 % acetonitrile and 0.06 % formic acid. Eluted peptides were split into 2 mg fractions and samples for full proteome analysis were frozen in liquid nitrogen and freeze dried.

Extracellular matrix (ECM) extraction

To remove the ECM and thus the formed biofilm prior to cell lysis one volume cell pellet was suspended in 1.5 M NaCl and the bacteria were spun down for 10 min at 5,000 rpm and 4°C. Subsequently the bacteria were lysed according to the Gram-positive protocol. 10 ul of each the ECM fraction as well as the lysates were supplemented with 2 x XT-sample buffer containing 25 mM DTT and run on a 12% bis-Tris gel.

Phosphopeptide enrichment

Fe^{3+}-IMAC enrichments were performed as previously described[25]. In short, 2 mg lyophilized peptides were resuspended in loading buffer A (30% acetonitrile and 0.07% TFA) and, if necessary, the pH was adjusted to 2.3 using 10% TFA. The samples were loaded onto the Fe^{3+}-IMAC column (Propac IMAC-10 4 x 5 mm column, Thermo Fischer Scientific). Bound phosphopeptides were eluted with elution buffer B (0.3 % NH_4OH). The respective gradient is described in supplementary Table S1. The UV-abs signal at a wavelength of 280 nm was recoded at the outlet of the column and eluting phosphopeptides were collected manually. Subsequently, phosphopeptides were frozen in liquid nitrogen and freeze dried.

LC-MS/MS

Microflow LC-MS/MS was performed using an Ultimate 3000 (Thermo Fisher Scientific, Bremen, Germany) coupled to an Orbitrap Exploris™ 480 or Orbitrap Fusion Lumos (both Thermo Fisher Scientific, Bremen, Germany). Lyophilized phosphopeptides were resuspended in 20 mM citric acid (Sigma-Aldrich), 1 % (v/v) formic acid. Resuspended phosphopeptide, corresponding to 1.6 mg were injected, trapped and washed on a trap-column (µ-Precolumn, 300 µm i.d. x 5mmC18 PepMap100, 5 µm, 100 Å (Thermo Scientific, P/N 160454)) for 5 min at a flow rate of 5 µL/minute with 92 % buffer A (0.1 FA, in HPLC grade water). Peptides were subsequently transferred onto an analytical column (75 µm x 50 cm Poroshell 120 EC-C18, 2.7 µm, Agilent Technology, packed in-house) and separated at 40°C at a flow rate of 0.3 µl/min using a 85 min linear gradient from 9 % to 36% buffer B (0.1 % FA, 80 % ACN). Electrospray ionization was performed using 1.9 kV spray voltage and a capillary temperature of 275 °C. The mass spectrometer was operated in data-dependent acquisition mode: full scan MS spectra (m/z 375 – 1,600, 375-1,500 respectively) were acquired in the Orbitrap at 60,000 resolution for a maximum injection time set to auto-mode with a standard AGC target or 4e5 with a normalized AGC target of 100%, respectively. High resolution HCD MS2 spectra were generated using a normalized collision energy of 28% or 35% respectively. The intensity threshold to trigger MS2 spectra was set to 5e4, and the dynamic exclusion 14s and 12s respectively. MS2 scans were acquired in the Orbitrap mass analyzer at a resolution of 30,000 (isolation window of 1.4 Th or 1.6 Th) with a standard AGC target and an automatic maximum injection time or an AGC target of 5e4 and a maximum injection time of 54ms. Precursor ions with unassigned charge state as well as charge state of 1+ or superior/equal to 6+ were excluded from fragmentation.

Data analysis

Raw files were processed using MaxQuant software (version 1.6.17.0) and the Andromeda search engine was used to search against either a *P. aeruginosa* PA14 (Uniprot/TrEMBL, July 2020, 5,888 entries), *Klebsiella pneumonia* MGH database (Uniprot/TrEMBL, February 2021, 5,126 entries) or *E. coli* (Uniprot, March 2016, 4,434 entries) with the following parameters for phosphoproteome analysis: trypsin digestion with a maximum of 3 missed cleavages, carbamidomethylation of cysteines (57.02 Da) as a fixed modification, methionine oxidation (15.99 Da), N-acetylation of proteins N-termini (42.01 Da) and phosphorylation on serine, threonine, tyrosine and histidine residues (79.96 Da) or serine, threonine, tyrosine histidine and arginine residues as variable modifications. Mass tolerance was set to 4.5 ppm at the MS1 level and 20 ppm at the MS2 level. The False Discovery Rate (FDR) was set to 1% for peptide-spectrum matches (PSMs) and protein identification using a target-decoy approach, a score cut-off of 40 was used in the case of modified peptides and the minimum peptide length was set to 7 residues. The MaxQuant generated tables “evidence.txt” and

"phospho (HSTY)Sites.txt" were used to calculate the number of unique phosphopeptides and phosphosites identified, respectively, and known contaminants were filtered out.

Identification of DNA/RNA contamination in MS2 spectra

Raw files were converted into .mgf files with Proteome Discoverer (vers. 2.3.0.523), using de-isotoping with an isotope deviation tolerance of 25 mmu. Subsequently .mgf-files were analyzed using an in-house made script searching MS2 spectra for a 330.06 m/z peak with a 0.02 Da tolerance and no S/N cut-off.

MaxQuant output processing

Further data processing was done using R studio with R version 4.1.0. All scripts are available upon request. In addition to the basic r functions the following packages were used: tidyverse (v1.3.1), reshape2 and ggpubr. MaxQuant output files, "evidence.txt", "Phospho (HSTY)Sites.txt" as well as "summary.txt" were used. The "evidence.txt" was filtered for "reversed" and "potential contaminants" as well as for unique peptides based on their modified sequence. Number of peptides and phosphopeptides as well as the enrichment efficiency (phosphopeptides/peptides) was determined per replicate and the average as well as standard deviation was calculated. "Phospho (HSTY)Sites.txt" was filtered for "reversed", "potential contaminants" as well as valid "Number of Phospho (HSTY)". Class I sites were identified by filtering for a localization probability >= 0.75 and the average as well as standard deviation was calculated. To compare the number of MS and MS2 spectra as well as the identification rate, the "summary.txt" was used and the average and standard deviation of recorder "MS" and "MSMS" as well as "MSMS.submitted" and "MSMS.identified" was calculated. The identification rate was determined by the ratio of "MSMS.identified"/"MSMS".

To determine the depth of the PAO1 phosphoproteomes the "Phospho (HSTY)Sites.txt" of the combined search (n=4) as well as combined PA14 search (n=3) was filtered for "reversed" and "potential contaminants". The intensity was log2 transformed and averaged. The data set was either filtered for at least three valid values or further processed without valid value filtering. p-sites were sorted according their log2(intensity) and the intensity was plotted against their rank. Similar, the p-sites were sorted according to the Andromeda Score and the score was plotted against their rank. Overlapping p-sites between PAO1 and PA14 were identified using an identifier consisting of the Uniprot ID, the amino acids as well as the position with the protein. The p-site distribution was calculated based on this and without valid value filter. Additionally, "Phospho (RHSTY)Sites.txt" was used and filtered as above to determine the pArg phosphoproteome distribution. Lastly, the "Phospho (HSTY)Sites.txt" and "Phospho (RHSTY)Sites.txt" for the combined *K. pnaeuminae*

MGH searches (n=4) were filtered as described above and the p-site distribution determined.

Pearson correlation

To determine a possible correlation between the amount of contamination and better identification Pearson correlation was performed using the PerformanceAnalytics r package (https://CRAN.R-project.org/package=PerformanceAnalytics). For the correlation when using the old and the Gram-negative protocol for *E. coli* K12 as previously published[25] three replicates for each condition were used. To establish the correlation using different conditions and different organisms in total 18 replicates were used.

Spectra comparison

To identify precursor identified and assigned in PAO1 but not assigned in PA14 mgf files were generated using and in-house developed software FragmentLab (https://scheltemalab.com/software) (v2.6.1.9) and the MSnbase[28] r package was used to load the complete mgf files as well as the mgf files containing assigned spectra. Precursor *m/z* were extracted, and precursor masses assigned in PAO1, but not in PA14 were identified. MSnbase[3] was further used to calculate expected b- and y-ion and to plot the annotated spectra.

Results & Discussion

Over the last decades, several groups adjusted the typical eukaryotic sample preparation and enrichment strategies to the lower abundance of bacterial phosphorylation as well as phosphoramidates, that are only scarcely reported in eukaryotes[11,22–26]. This broadened our understanding of the bacterial phosphoproteome; however, we are still far from a comprehensive understanding of those complex signaling events. Recently, an optimized sample preparation workflow for *E. coli*[25] was published which we extended to Gram-positive bacteria, mainly *S. aureus*[26]. In both cases, optimization of the workflow resulted in an enormous increase in identified phosphosites. The success was partly associated with a decrease in contamination of nucleic acid containing biomolecules. It was hypothesized that the phosphoryl groups of DNA and RNA are competing with phosphopeptides and thus impede their enrichment. Further, high amounts of those molecules are believed to result in insufficient ionization due to their negative charges and hamper spectral identification supposedly caused by co-fragmenting DNA and/or RNA molecules[25]. Figure 1 shows a characteristic spectrum for samples containing large amounts of contaminations and thus suffer from identification. A mix of DNase I and benzonase to cleave DNA and RNA with subsequent chloroform/methanol precipitation has been shown to efficiently remove the contamination, thereby increasing the identification rate[25]. Therefore, removal of those nucleic acid containing biomolecules should not only improve the phosphopeptide

enrichment, but also ionization and identification. We could show that lower amounts of contamination, indicated by the presence of cAMP (330.06 *m/z*), is indeed correlating with better identification for *E. coli* K12 (see supplementary Figure S1).

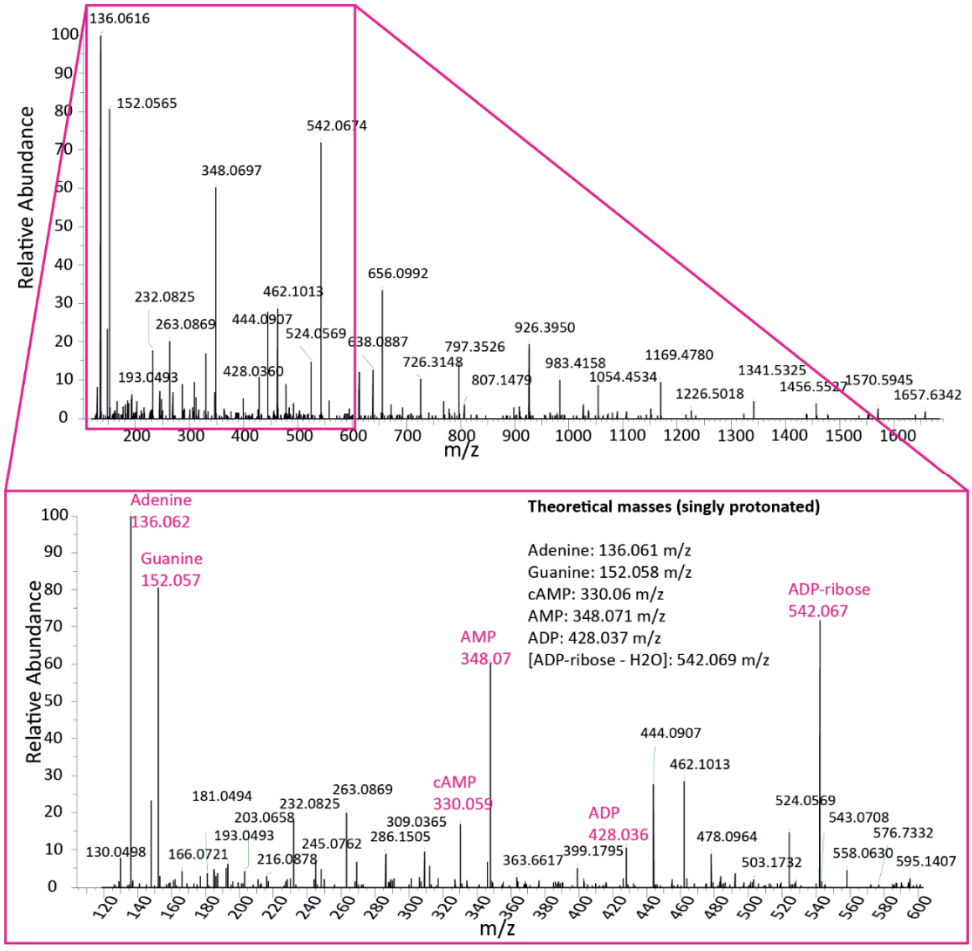


Figure 1: Marker fragment ions of nucleic acid containing biomolecules present in HCD spectra. The phosphoryl group of nucleic acids can compete with phosphopeptides for binding to the Fe^{3+}-IMAC resin. In addition, co-fragmentation of DNA and/or RNA molecules are thought to hamper spectral identification, due to an increased complexity of the MSMS spectrum by additional high intensity DNA/RNA fragment masses.

Applying improved sample preparation on P. aeruginosa and K. pneumoniae

Encouraged by this success we wanted to test the universal applicability of this protocol by performing a phosphoproteomic analysis of the Gram-negative bacteria *P. aeruginosa* as well as *K. pneumoniae*.

The main difference between the sample preparation for Gram-negative and Gram-positive bacteria is the additional use of SDC as second detergent next to Triton X-100 and the combination of two different mechanical lysis steps (bead beating and sonication) to create harsher lysis conditions for the thicker peptidoglycan layer of Gram-positive bacteria. Here, we compared the identification of phosphopeptides and phosphorylation sites (p-sites) for *P. aeruginosa* PA14 after lysis with the optimized protocol for Gram-negative bacteria (referred to as Gram-negative protocol) to the additional use of SDC, however without bead beating (referred to as Gram-positive protocol). Using the Gram-negative protocol around 72 ± 9 phosphopeptides and 58 ± 14 p-sites were identified of which on average 47 ± 12 class I p-site (Figure 2a and b). This is slightly lower than what is described by Ouidir and co-workers who identified 84 p-sites for *P. aeruginosa* PA14[29]. The Gram-positive protocol led to the identification of 173 ± 35 phosphopeptides and 142 ± 29 p-sites of which 115 ± 29 class I p-sites. This indicated a general lower identification after using the Gram-negative protocol. Both lysis conditions resulted in a similar enrichment efficiency (phosphopeptides/peptides) of 2.5% and 2.7% respectively (Figure 2a). However, even though the identification increased when using the Gram-positive protocol, the contamination, which is indicated by MS2 spectra containing the diagnostic ion of 330.06 *m/z*, was high for both lysis conditions (64% and 67% respectively, Figure 2b). Hence, surprisingly, the approach which led to a significant improvement in phosphopeptide enrichment in *E. coli* K12 and *S. aureus*, was not sufficient to reduce contamination in *P. aeruginosa* PA14 after lysis, which we believe is a major factor in identifying relatively much less phosphopeptides in *P. aeruginosa* PA14.

Finding the source of contamination

Compared to *E. coli* K12 and *S. aureus* which have a 50.8%[30] and 32.8%[31] GC content respectively, *P. aeruginosa* has a higher GC content with around 66%[32] (Table 1). GC-rich DNA has been shown to have a higher melting point[33,34] and is thus more difficult to brake. To test whether the higher GC content caused the high contamination, we analyzed more favorable conditions to cleave DNA/RNA being i) a higher percentage of benzonase, 2% instead of 1% (referred to as 2% benzonase) and ii) the addition of 5% DMSO to decrease the stability of GC-rich DNA (referred to as DMSO).

Table 1: GC content of different organisms

Organism	%GC content	Ref
E. coli K12	50.8	B. Ely[30]
S. aureus USA300	32.8	J McClure, K Zhang[31]
P. aeruginosa PA14	66.4	D. Lee[32]
P. aeruginosa PAO1	66.6	D. Lee[32]
K. pneumonia MGH	57.7	KM Wu[35]

The Gram-positive protocol resulted in the identification of around 102 ± 5 phosphopeptides and 68 ± 8 p-sites (n=3). Surprisingly this is less than half of the number of peptides and sites identified in the previous batch. At the same time, the contamination increased to 82% of the spectra, which is around 15% higher than in the previous batch and might explain the lower identification. Increased concentration of benzonase and the addition of DMSO increased the number of identified phosphopeptides (144 ± 11 and 152 ± 2 respectively) and p-sites (101 ± 20 and 125 ± 1, respectively) for both conditions compared to the Gram-positive protocol (Figure 2c and d), while the contamination was present in 61% of the spectra (Figure 2c). However, the standard deviation of MS2 spectra containing the diagnostic ion was relatively high with 14% and 11%, respectively. Furthermore, the amount of identified p-sites and peptides after using 2% benzonase or 5% DMSO was comparable to that after using the Gram-positive protocol for the first sample batch (Figure 2b), indicating batch to batch variation. However, we also need to stress that the overall number of identified phosphopeptides is relatively low and small deviations might weigh stronger than when the number of identified phosphopeptides is high.

P. aeruginosa produces a biofilm that consists, especially during the early stages of its formation, of extracellular DNA (eDNA) that hold the different cells together[36,37]. We thus hypothesized that this extracellular DNA might be the source of the high contamination and thus the cause for the lower number of p-sites identified. To remove this biofilm, including extracellular DNA that could interfere with the enrichment completely, we extracted the extracellular matrix (ECM) prior to lysis using the Gram-positive protocol and compared it to the lysis without ECM extraction. After ECM extraction 114 ± 9 phosphopeptides and 87 ± 2 p-sites were identified while having 97% contamination. Using the Gram-positive protocol 234 ± 29 phosphopeptides and 189 ± 27 p-sites were identified while having a slightly lower contamination of around 87% (Figure 2e and f). One explanation for the overall lower identification after ECM extraction could be a partial lysis of the cells. SDS-PAGE showed no obvious differences between the lysates with and without ECM extraction (supplementary Figure S2). However, the ECM fraction contains more protein bands than expected from the literature protocol[38], indicating indeed the possibility of partial cell lysis during ECM extraction. One should keep in mind though that the ECM extraction protocol

was optimized for *S. aureus* MR23 and that biofilms can already vary greatly between different strains just as between different species[38].

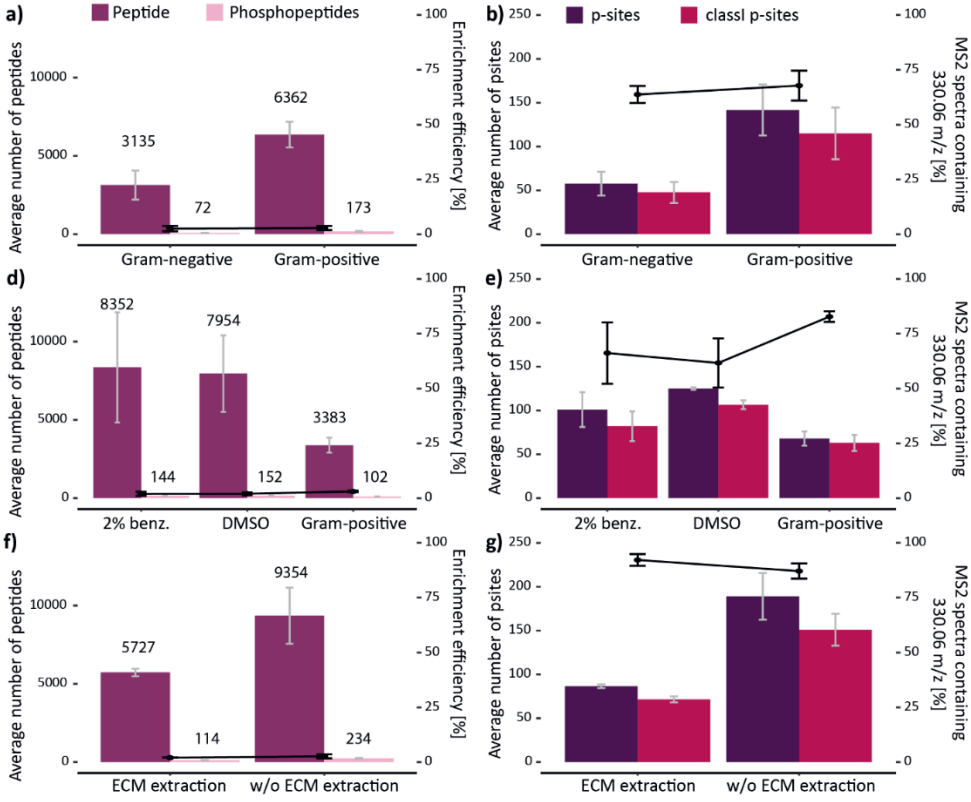


Figure 2: Peptide identification for *P. aeruginosa*. The average number of identified peptides and phosphopeptides is shown as bar graph and the efficiency of the enrichment determined by the ratio of phosphopeptides to peptides [%] is shown as point and line graph (**a, c, d**). The average number of identified p-sites is shown as bar graph and the number of MS2 spectra containing the diagnostic ion of 330.06 *m/z* indicative for the DNA/RNA contamination is depicted as point and line graph (**b, d, f**). **a)** and **b)** show the comparison for the Gram-negative to the Gram-positive protocol (n=3). **c)** and **d)** compare the identification after using a higher amount of benzonase (2% benz.) to the use of 5% DMSO (both n=2) as well as the Gram-positive protocol (n=3). **e)** and **f)** show the identification using the Gram-positive protocol with or without previous ECM extraction (n=3). All bar graphs depict the mean value, and the standard deviation is given as error bars.

Since we observed quite a deviation in the number of identified peptides, phosphopeptides and consequently p-sites, not only between the various conditions, but also between different sample batches while the enrichment efficiency remained almost the same, we sought to elucidate on which level these issues occur. A general lower identification, indicated by the number of identified peptides and phosphopeptide as for example seen for the comparison of the Gram-negative and Gram-positive protocol (Figure 2a), could result from a non-optimal Fe^{3+}-IMAC column, resulting in a lower number of peptides subjected

to LC-MS/MS analysis and thus resulting in an overall lower number of MS spectra. However, this variation between column performances is minimal for samples that are enriched at the same time, as is the case here. Alternatively, poorer ionization of the peptides presumably due to the negatively charged DNA/RNA contamination could also result in an overall lower number of MS spectra and thus lower identification. Additionally, lower identification could originate from insufficient fragmentation and thus inadequate identification of spectra by the chosen search engine. To check this, we first compared the TIC of the Gram-negative and Gram-positive protocol. Both samples showed great similarity with respect to intensity, retention time as well as most abundant base peaks (see supplementary Figure S3). Furthermore, the numbers of MS and MS2 spectra did not show any obvious differences for the various conditions as well as batches (see supplementary Figure S4). At the same time, we observed substantial differences on the level of identification associated to the level of contamination (see supplementary Figure S4).

Possible strain specific challenges in identification

To see if the higher contamination and thus most likely worse identification is strain specific, we also analyzed *P. aeruginosa* PAO1. Even though PAO1 and PA14 share more than 90%[32] of their genome, the strains differ, for example, in their virulence with PA14 showing a higher virulence in most hosts compared to PAO1[32]. Intriguingly, the number of peptides (8,488 $\pm$ 479) and especially phosphopeptides (389 $\pm$ 24) identified for PAO1 was higher than for strain PA14 (see above and Figure 3a). Consequently, we identified almost six-times more p-sites for PAO1 compared to PA14 and compared to literature (333 $\pm$ 12, 58 $\pm$ 14 and 55[39], respectively). The increased identification came along with a drastic decrease in contamination to 30% (Figure 3b), again emphasizing the role of the contamination in the success of p-site identification in bacteria. Since both strains have a similar %GC-content (see Table 1) this also disproves an effect of the %GC for the benzonase treatment, and rather points to a strain specific problem. To test if the contamination might be indeed strain or organism specific, we also analyzed the Gram-negative bacterium *K. pneumoniae* MGH using the Gram-negative as well as the Gram-positive protocol. First, only minor differences between the Gram-negative and Gram-positive protocol were observed with a slight preference for the Gram-negative protocol leading to the identification of 8,656 $\pm$ 1164 peptides and 590 $\pm$ 90 phosphopeptides (Figure 3c). In total 439 $\pm$ 68 p-sites (Figure 3d) were identified which is similar to the total number identified by Lin and co-workers[11] who identified 559 p-sites using HAMMOC with titanium dioxide. However, around 80% of all MS2 spectra contained the diagnostic ion of 330.06 *m/z*, suggesting a hampered identification due to higher contamination. To validate that the Gram-negative protocol is indeed still applicable for *E. coli* and that we are not dealing with a systematic error we applied both the Gram-negative as well as the Gram-positive protocol for the lysis of *E. coli* DH5a. Here, we identified 5,512 $\pm$ 160 peptides and 787 $\pm$ 32 phosphopeptides

and even though those numbers are slightly lower than published by Potel and co-workers[25] with 7,095 ± 628 peptides and 1,035 ± 145 phosphopeptides, the enrichment efficiency with around 14% was similar (Figure 3e). At the same time, while *E. coli* K12 only showed 15% contamination using the Gram-negative protocol we identified 33% for *E. coli* DH5a, which could explain the reduction in number of identifications. Interestingly, even though the Gram-positive protocol increased the contamination to 67%, the number of identified p-sites (659 ± 24 and 667 ± 46 for the Gram-negative and Gram-positive protocol, respectively) and class I sites (554 ± 17 and 550 ± 36, respectively) was similar, independent of the used protocol. Higher contamination, when using the Gram-positive protocol could result from the use of SDC. At higher concentrations, SDC is known to inhibit the benzonase activity (manufacture information, Merck). However, by lowering the SDC concentration to 0.4%, benzonase activity is only reduced to 70% (manufacture information, Merck). This remaining benzonase activity has previously been shown to be sufficient to remove the majority of the contamination[26].

Further, even though the contamination is almost in the range as observed for the old protocol used for Gram-negative bacteria (72%), there is a great difference in the number of identified p-sites (119 ± 59). The fact that the enrichment efficiency for *E. coli* K12 using the Gram-negative protocol and *E. coli* DH5a is comparable supports again the assumption that the identification is not necessarily hampered at the level of the enrichment, but at the level of identification. This was again supported when comparing the number of recorded MS and MS2 spectra (see supplementary Figure S5). For *P. aeruginosa* and *K. pneumoniae* the number of recorded MS and MS2 spectra were comparable, while the number of identified MS2 spectra was differing especially for the strains PAO1 and PA14 with better identification for PAO1. Additionally, the TIC comparison for PAO1 and PA14 did not show any significant differences with respect to intensity and most abundant base peaks (see supplementary Figure S3). For *E. coli* K12 using the Gram-negative protocol more MS spectra were recorded compared to the old protocol as well as *E. coli* DH5a. Interestingly, more MS2 spectra were recorder for *E. coli* DH5a compared to *E. coli* K12 while the number of identified MS2 spectra was comparable. Consequently *E. coli* K12 using the Gram-negative protocol has the best ID rate (supplementary Figure S5). To see if there is indeed a correlation between the amount of contamination and recorded as well as identified spectra, we performed Pearson correlation. Here, we could identify only a moderate positive correlation ($r=0.57$, $p \leq 0.05$) between the amount of contamination and the number of MS2 spectra submitted for analysis as well as a moderate negative correlation ($r=-0.56$, $p \leq 0.05$) between the contamination and ID rate (MS2 identified/MS2 submitted) (see supplementary Figure S6). Similarly, only a moderate correlation was observed between the contamination and identified p-sites ($r=-0.52$, $p \leq 0.05$, supplementary Figure S7). This is contrary to the correlation observed for *E. coli* K12 (supplementary Figure S1) as

previously published[25] where a strong negative correlation between the amount of contamination and the identification could be observed. This suggests i) that the amount of contamination and identification is not correlating as strongly as previously thought, ii) that the success of the optimized sample preparation is highly species if not even strain specific and iii) that other, environmental factors, such as growth medium, temperature or growth phase are also having an essential influence on the phosphoproteome and thus its analysis.

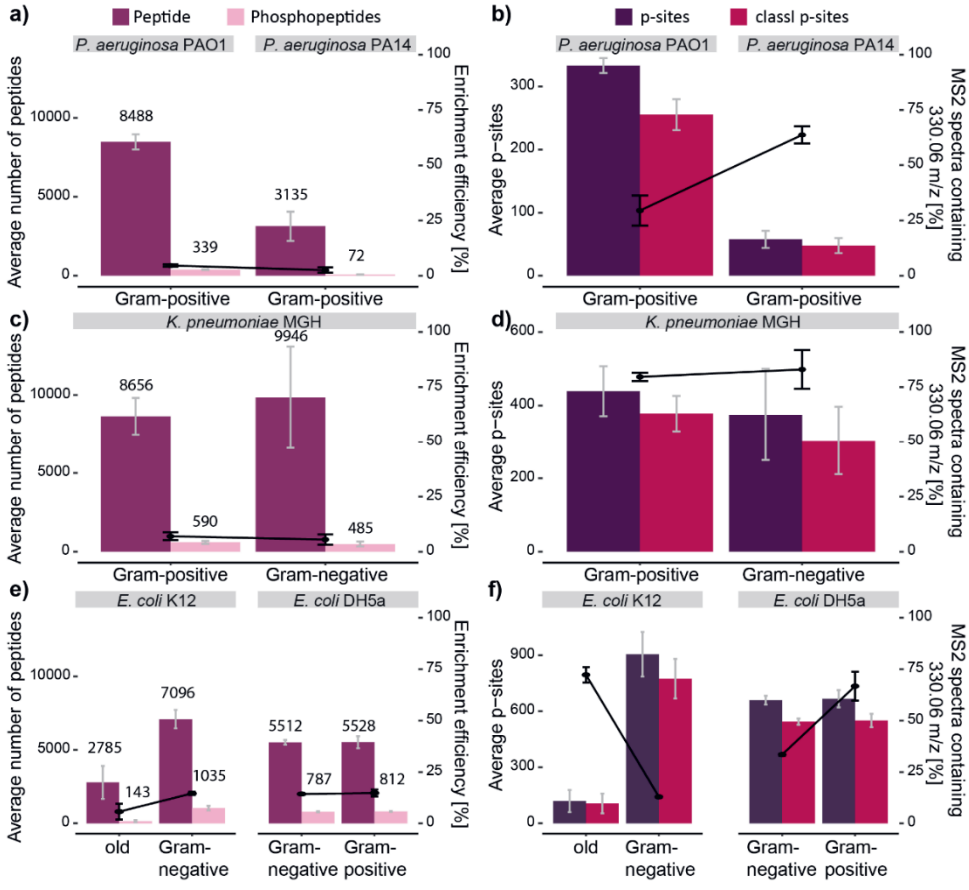
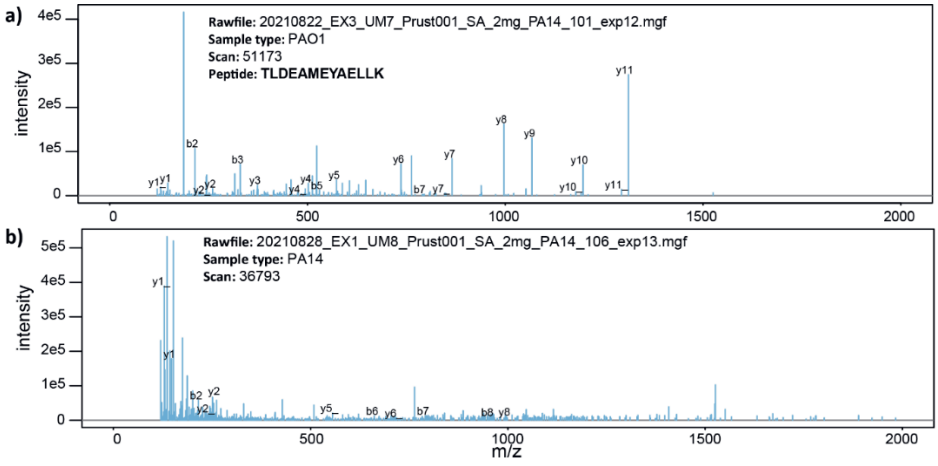


Figure 3: Peptide identification for *P. aeruginosa* PAO1, PA14, *K. pneumoniae* MGH as well as *E. coli* K12 and DH5a. The average number of identified peptides and phosphopeptides is shown as bar graph and the efficiency of the enrichment determined by the ratio of phosphopeptides to peptides [%] is shown as point and line graph (**a, c, d**). The average number of identified p-sites is shown as bar graphs and number of MS2 spectra containing the diagnostic ion of 330.06 *m/z* indicative for the DNA/RNA contamination is depicted as point and line graph (**b, d, f**). **a)** and **b)** show the comparison *P. aeruginosa* PAO1 (n=4) and PA14 (n=3). **c)** and **d)** compare the identification after using the Gram-negative or Gram-positive protocol for *K. pneumonia* MGH (n=4). **e)** and **f)** show the identification for *E. coli* K12 using an old protocol compared to the Gram-negative protocol (as published by Potel and co-workers[25]) and the comparison of the Gram-negative and Gram-positive protocol for *E. coli* DH5a.

Since not necessarily the enrichment itself seems to be the issue, but rather the identification, we were wondering how many of the assigned spectra contained the contamination. Here we observed for both strains that spectra that were assigned by MaxQuant contained a lower percentage of contaminant spectra compared to the percentage in all recorded spectra (see supplementary Figure S8), again hinting towards sub-optimal identification. We were thus speculating whether the identification can be improved by removing all peaks related to DNA/RNA from the .mgf files and then searching those using Proteome Discoverer. However, unfortunately no improvement compared to the original .mgf file was observed (data not shown). Next, we tried to identify precursor masses in PAO1 that were identified by MaxQuant and present in PA14, but not identified. Here, we found 46 precursor masses that were indeed identified by MaxQuant for PAO1 (replicate 1) and not for PA14 (replicate 1). For some of those spectra we could identify the corresponding fragmentation peaks as shown in Figure 4. However, the overall number of precursor masses not identified in PA14 is far too small that this would make up for the great difference in peptide identification. Furthermore, we saw a generally higher noise level containing the contamination peaks in PA14 compared to PAO1. Even though we saw a good correlation between the contamination and identification for *E. coli* K12, we could not confidently extend this to all prokaryotes analyzed in this study. Consequently, we cannot conclude that contamination as displayed by the diagnostic ion 330.06 *m/z* is the only reason nor a bona fide indicator for the limited p-site identification in *P. aeruginosa* as well as *K. pneumoniae*, although a high contamination suggests inefficient and thus non-comprehensive p-site identification.



c)

mz	intensity	ion	type	pos	z	seq	error
129.1025	389223	y1_	y_	1	1	K	-0.00024
147.1131	178417	y1	y	1	1	K	-0.00028
215.1395	58347.9	b2	b	2	1	TL	-0.00046
242.1505	19716.2	y2_	y_	2	1	LK	0.035796
260.1978	23749	y2	y	2	1	LK	-0.00095
555.2899	20879.6	y5_	y_	5	1	AELLK	0.060196
661.3043	12267.2	b6	b	6	1	TLDEAM	-0.01811
718.3494	6967.31	y6_	y_	6	1	YAELLK	0.063956
790.3829	12623.8	b7	b	7	1	TLDEAME	-0.05413
953.3871	8933.58	b8	b	8	1	TLDEAMEY	0.004986
996.432	10515.4	y8	y	8	1	MEYAELLK	0.075081

Figure 4: Spectra comparison for TLDEAMEYAELLK identified by MaxQaunt in PAO1 **a)** and the respective spectra of the same precursor *m/z*, but not identified by MaxQuant in PA14, **b)**. The MSnbase[3] r package was used to calculate the expected b- and y-ion and to plot the annotated spectra. The identified ions for the PA14 spectrum are displayed in c).

Species and strain specific phosphoproteomes

Despite these remaining challenges in bacterial sample preparation, still differences were observed for PAO1 and PA14. Previous studies suggested a link between a higher phosphorylation rate, especially pTyr, and a higher virulence[12,40,41]. Thus, one would expect more phosphorylation or at least more pTyr in the more virulent strain PA14 compared to PAO1. When comparing the p-site distribution this is indeed the case. While PA01 has only 7.8% pTyr (Figure 5c), PA14 contains 11.8% pTyr (see supplementary Figure S9a). Interestingly, the percentage of pTyr for PA14 was similar when using the Gram-positive (supplementary Figure S9a and b) and Gram-negative protocol (supplementary Figure S9c and d), while the percentage of identified pSer changed between those two protocols. However, the lower numbers of p-sites could skew the distribution to more abundant p-sites. This was indeed confirmed when comparing the phosphoproteome depth of both species. Sixty-two p-sites were identified in both strains that are mainly high abundant sites (pink dots), showing a greater phosphoproteome depth for PAO1 than PA14. (Figure 5a).

Similarly, the Andromeda score plotted against their p-site rank showed that those p-sites have overall a relatively high score. This however does not hint towards a poor data quality in PAO1, as 75 % of p-sites identified in PAO1 (filtered for at least three valid values) are unambiguously identified (170 class I sites, Figure 5c). This indicates that the p-site identification in PA14 is indeed limited to high abundant p-sites. This was again confirmed when comparing the identified p-sites for PA14 after using the two different protocols. Here we could see that 80% of the p-sites identified using the Gram-negative protocol (filtered for three valid values) were also identified using the Gram-positive protocol (filtered for three valid values, see supplementary Figure S9f). Based on this we hypothesize that the PA14 phosphoproteome is larger than what is discovered today and that its identification is hampered, not only, but at least partially by DNA/RNA contamination.

While PAO1 was previously described to have 14% pTyr[39], we show here that only 7.8% of all p-sites are localized on Tyr (52.2 pSer, 34.5% pThr, 7.8% pTyr and 5.4% pHis, Figure 5c). These differences can be explained by the overall higher number of p-sites that should provide a better representation of the p-site distribution. Even though we applied strict valid value filtering (each site had to be identified in at least three out of four replicates), the proportion of pTyr did not change when considering all sites (see supplementary Figure S10). Since we previously observed extensive arginine phosphorylation in *S. aureus* we were interested how much pArg is present in PAO1. Here we identified in total 13 phosphorylation sites at least present in three out of four replicates located to pArg making up for 6.2% of all identified p-sites (48.8% pSer, 33.2% pThr, 7.6% pTyr, 4.3% pHis, 6.2% pArg, Figure 5d). This shows that pArg is indeed present in *P. aeruginosa*, but to a far lesser extent than in *S. aureus* (up to 26%).

Even though we identified less p-sites for *K. penumoniae* we were also interested in the p-site distribution and especially the presence of pArg. Here, only 6.7% were localized on pArg underlining again that the larger proportion of arginine phosphorylation in *S. aureus* is indeed strain specific. Additionally, we found around 13% pTyr phosphorylation independent of the used lysis protocol (Figure 5f and supplementary Figure S11). Of all identified pTyr sites either after using the Gram-negative or Gram-positive protocol (filtered for at least three valid values) around 60% were identified in both sample sets (70% without valid value filtering, see supplementary Figure S12). Together with an overall overlap of 60% identified p-sites after using both protocols (see supplementary Figure S112), this shows a high similarity between the two protocols for the phosphoproteome analysis of *K. pneumoniae*.

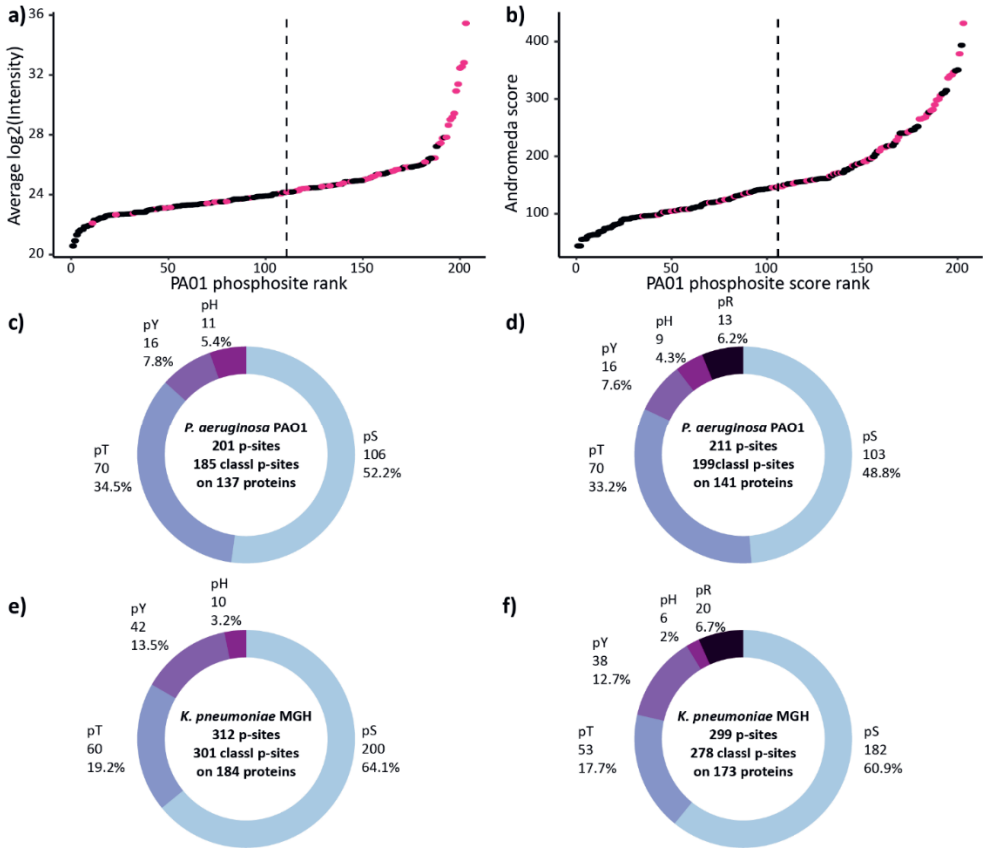


Figure 5: Phosphoproteome overview of *P. aeruginosa* and *K. pneumonia* MGH. Dynamic range of the phosphoproteome from *P. aeruginosa* PA01 depicted as the average log2(intensity) against the intensity rank (**a**) or the Andromeda score against the score rank (**b**). 62 of 79 p-sites in PA14 using the Gram-positive protocol were also identified in PAO1 using the Gram-negative protocol and these are depicted in pink. On the right-hand site of the dashed line 75% of the PA14 p-sites are located. P-site distribution for *P. aeruginosa* PAO1 (**c, d**) and *K. pneumonia* MGH (**e, f**) either searched for pHSTY or pRHSTY, using the Gram-negative protocol and filtering for at least 3 valid values (n=4).

One size does not fit all

By profiling different species and different strains, using the published protocol for Gram-negative bacteria[25] as well as for Gram-positive bacteria[26], we could show that those optimizations are highly species if not even strain specific. Furthermore, the correlation between a decrease in contamination and improved p-site identification for the different organisms was not as striking as observed for *E. coli* K12. Since we mainly observed differences on the level of identification and not on the level of the enrichment, we believe that the contamination is mostly hampering spectral identification. Most striking was the difference between the strains PAO1 and PA14, where the contamination was 2.3 times less abundant in PAO1 than PA14. Even though PAO1 and PA14 share more than 90% of their

genome, those strains do not only differ in their virulence, but also in the biofilm formation. In PAO1 surface attachment and extracellular matrix secretion occurs relatively quickly under control of c-di-GMP[42,43]. PA14 biofilm formation on the other hand takes longer, where an increase in cAMP precedes a c-di-GMP increase followed by biofilm formation[42,44]. Those innate difference could be one explanation for the drastic differences observed for the contamination in PAO1 and PA14. While the cAMP contamination in PAO1 is resulting from DNA/RNA contamination, the elevated intracellular cAMP levels in PA14 could give rise to the here observed cAMP contamination in the noise level and thus hamper the p-site identification. Interestingly, *K. pneumoniae*, which also showed high level of contaminations, makes direct use of c-di-GMP for its biofilm formation[45]. The question remains though, why the sample preparation does not result in improved p-site identification, while it was successfully applied for *E. coli* K12[25] as well as *S. aureus* USA300[26]. Based on the results obtained here, we can only speculate that the DNA/RNA contamination is only partly influencing the p-site identification and that other organism specific factors are also still hampering the identification. Those intrinsic factors could be extremely strain specific, as here hypothesized for the intercellular cAMP levels of PA14. One could also speculate that the here identified phosphoproteome of *P. auerginosa* PA14 and *K. pneumoniae* MGH is the almost complete phosphoproteome. However, Lin and co-workers already identified a larger phosphoproteome for *K. pneumoniae* by using HAMMOC with TiO_2[11]. Lastly, one should keep in mind that the phosphoproteome is a highly dynamic system and thus dependent on various factors such as e.g. growth medium, growth phase or temperature as well enrichment strategies. In short, we showed here that one optimized sample preparation and enrichment protocol cannot simply be applied to different species. Intrinsic differences between different species or even different strains call for highly tailored approaches to obtain a comprehensive view on their phosphoproteome.

Acknowledgments

We acknowledge Julius Ned Witte for providing all bacterial samples and support from Epic-XS (Project number: EPIC-XS-0000009). S.L. acknowledges support from the Netherlands Organization for Scientific Research (NWO) through a VIDI grant (project 723.013.008).

Competing financial interest

The authors declare no competing financial interest.

References

1. Scheuerl T, Hopkins M, Nowell RW, Rivett DW, Barraclough TG, Bell T. Bacterial adaptation is constrained in complex communities. *Nat Commun*. 2020;11(1). doi:10.1038/s41467-020-14570-z

2. Rainey PB, Travisano M. Adaptive radiation in a heterogeneous environment. *Nature*. 1998;394(6688):69-72. doi:10.1038/27900

3. Zdziarski J, Brzuszkiewicz E, Wullt B, Liesegang H, Biran D, Voigt B, Grönberg-Hernandez J, Ragnarsdottir B, Hecker M, Ron EZ, Daniel R, Gottschalk G, Hacker J, Svanborg C, Dobrindt U. Host imprints on bacterial genomes-rapid, divergent evolution in individual patients. *PLoS Pathog*. 2010;6(8):95-96. doi:10.1371/journal.ppat.1001078

4. Tuchscherr L, Medina E, Hussain M, Völker W, Heitmann V, Niemann S, Holzinger D, Roth J, Proctor RA, Becker K, Peters G, Löffler B. Staphylococcus aureus phenotype switching: an effective bacterial strategy to escape host immune response and establish a chronic infection. *EMBO Mol Med*. 2011;3(3):129-141. doi:10.1002/emmm.201000115

5. Chuard C, Vaudaux PE, Proctor RA, Lew DP. Decreased susceptibility to antibiotic killing of a stable small colony variant of Staphylococcus aureus in fluid phase and on fibronectin-coated surfaces. *J Antimicrob Chemother*. 1997;39(5):603-608. doi:10.1093/jac/39.5.603

6. Stock JB, Stock AM, Mottonen JM. Signal transduction in bacteria. *Nature*. 1990;344(6265):395-400. doi:10.1038/344395a0

7. Villanueva M, García B, Valle J, Rapún B, Ruiz De Los Mozos I, Solano C, Martí M, Penadés JR, Toledo-Arana A, Lasa I. Sensory deprivation in Staphylococcus aureus. *Nat Commun*. 2018;9(1):1-12. doi:10.1038/s41467-018-02949-y

8. Jung K, Fried L, Behr S, Heermann R. Histidine kinases and response regulators in networks. *Curr Opin Microbiol*. 2012;15(2):118-124. doi:10.1016/j.mib.2011.11.009

9. Macek B, Mijakovic I, Olsen J V., Gnad F, Kumar C, Jensen PR, Mann M. The Serine/Threonine/Tyrosine Phosphoproteome of the Model Bacterium Bacillus subtilis. *Mol Cell Proteomics*. 2007;6(4):697-707. doi:10.1074/mcp.M600464-MCP200

10. Macek B, Gnad F, Soufi B, Kumar C, Olsen J V., Mijakovic I, Mann M. Phosphoproteome Analysis of E. coli Reveals Evolutionary Conservation of Bacterial Ser/Thr/Tyr Phosphorylation. *Mol Cell Proteomics*. 2008;7(2):299-307. doi:10.1074/mcp.M700311-MCP200

11. Lin M-H, Sugiyama N, Ishihama Y. Systematic profiling of the bacterial phosphoproteome reveals bacterium-specific features of phosphorylation. *Sci Signal*. 2015;8(394). doi:10.1126/scisignal.aaa3117

12. Grangeasse C, Cozzone AJ, Deutscher J, Mijakovic I. Tyrosine phosphorylation: an emerging regulatory device of bacterial physiology. *Trends Biochem Sci*. 2007;32(2):86-94. doi:10.1016/j.tibs.2006.12.004

13. Whitmore SE, Lamont RJ. Tyrosine phosphorylation and bacterial virulence. *Int J Oral Sci*. 2012;4(1):1-6. doi:10.1038/ijos.2012.6

14. Olsen J V., Mann M. Status of large-scale analysis of posttranslational modifications by mass spectrometry. *Mol Cell Proteomics*. 2013;12(12):3444-3452. doi:10.1074/mcp.O113.034181

15. Riley NM, Coon JJ. Phosphoproteomics in the Age of Rapid and Deep Proteome Profiling. *Anal Chem*. 2016;88(1):74-94. doi:10.1021/acs.analchem.5b04123

16. Lemeer S, Heck AJ. The phosphoproteomics data explosion. *Curr Opin Chem Biol*. 2009;13(4):414-420. doi:10.1016/j.cbpa.2009.06.022

17. Doll S, Burlingame AL. Mass spectrometry-based detection and assignment of protein posttranslational modifications. *ACS Chem Biol*. 2015;10(1):63-71. doi:10.1021/cb500904b

18. Loroch S, Zahedi RP, Sickmann A. Highly sensitive phosphoproteomics by tailoring solid-phase extraction to electrostatic repulsion-hydrophilic interaction chromatography. *Anal Chem*. 2015;87(3):1596-1604. doi:10.1021/ac502708m

19. Arike L, Peil L. *Shotgun Proteomics*. Vol 1156. (Martins-de-Souza D, ed.). Springer New York; 2014. doi:10.1007/978-1-4939-0685-7

20. Ruprecht B, Koch H, Medard G, Mundt M, Kuster B, Lemeer S. Comprehensive and Reproducible Phosphopeptide Enrichment Using Iron Immobilized Metal Ion Affinity Chromatography (Fe-IMAC) Columns. *Mol Cell Proteomics*. 2015;14(1):205-215. doi:10.1074/mcp.M114.043109

21. Besant P, Attwood P, Piggott M. Focus on Phosphoarginine and Phospholysine. *Curr Protein Pept Sci*. 2009;10(6):536-550. doi:10.2174/138920309789630598

22. Potel CM, Lin MH, Heck AJR, Lemeer S. Widespread bacterial protein histidine phosphorylation revealed by mass spectrometrybased proteomics. *Nat Methods*. 2018;15(3):187-190. doi:10.1038/nmeth.4580

23. Bäsell K, Otto A, Junker S, Zühlke D, Rappen G-M, Schmidt S, Hentschker C, Macek B, Ohlsen K, Hecker M, Becher D. The phosphoproteome and its physiological dynamics in Staphylococcus aureus. *Int J Med Microbiol*. 2014;304(2):121-132. doi:10.1016/j.ijmm.2013.11.020

24. Junker S, Maaβ S, Otto A, Michalik S, Morgenroth F, Gerth U, Hecker M, Becher D. Spectral Library Based Analysis of Arginine Phosphorylations in Staphylococcus aureus. *Mol Cell Proteomics*. 2018;17(2):335-348. doi:10.1074/mcp.RA117.000378

25. Potel CM, Lin M-H, Heck AJR, Lemeer S. Defeating Major Contaminants in Fe 3+ - Immobilized Metal Ion Affinity Chromatography (IMAC) Phosphopeptide Enrichment. *Mol Cell Proteomics*. 2018;17(5):1028-1034. doi:10.1074/mcp.TIR117.000518

26. Prust N, van der Laarse S, van den Toorn HWP, van Sorge NM, Lemeer S. In-Depth Characterization of the Staphylococcus aureus Phosphoproteome Reveals New Targets of Stk1. *Mol Cell Proteomics*. 2021;20:100034. doi:10.1074/mcp.RA120.002232

27. Swoboda JG, Campbell J, Meredith TC, Walker S. Wall teichoic acid function, biosynthesis, and inhibition. *Chembiochem*. 2010;11(1):35-45. doi:10.1002/cbic.200900557

28. Gatto L, Lilley KS. Msnbase-an R/Bioconductor package for isobaric tagged mass spectrometry data visualization, processing and quantitation. *Bioinformatics*. 2012;28(2):288-289. doi:10.1093/bioinformatics/btr645

29. Ouidir T, Jarnier F, Cosette P, Jouenne T, Hardouin J. Potential of liquid-isoelectric-focusing protein fractionation to improve phosphoprotein characterization of Pseudomonas aeruginosa PA14. *Anal Bioanal Chem*. 2014;406(25):6297-6309. doi:10.1007/s00216-014-8045-8

30. Ely B. Genomic GC content drifts downward in most bacterial genomes. *PLoS One*. 2021;16(5 May 2021):1-9. doi:10.1371/journal.pone.0244163

31. McClure J-A, Zhang K. Complete Genome Sequences of Two USA300-Related Community-Associated Methicillin-Resistant Staphylococcus aureus Clinical Isolates. Stewart FJ, ed. *Microbiol Resour Announc*. 2019;8(18):12-13. doi:10.1128/MRA.00404-19

32. Lee DG, Urbach JM, Wu G, Liberati NT, Feinbaum RL, Miyata S, Diggins LT, He J, Saucier M, Déziel E, Friedman L, Li L, Grills G, Montgomery K, Kucherlapati R, et al. Genomic analysis reveals that Pseudomonas aeruginosa virulence is combinatorial. *Genome Biol*. 2006;7(10). doi:10.1186/gb-2006-7-10-r90

33. Rouzina I, Bloomfield VA. Heat Capacity Effects on the Melting of DNA. 1. General Aspects. *Biophys J*. 1999;77(6):3242-3251. doi:10.1016/S0006-3495(99)77155-9

34. Frank-Kamenetskii MD. Simplification of the empirical relationship between melting temperature of DNA, its GC content and concentration of sodium ions in solution. *Biopolymers*. 1971;10(12):2623-2624. doi:10.1002/bip.360101223

35. Wu KM, Li NH, Yan JJ, Tsao N, Liao TL, Tsai HC, Fung CP, Chen HJ, Liu YM, Wang JT, Fang CT, Chang SC, Shu HY, Liu TT, Chen YT, et al. Genome sequencing and comparative analysis of Klebsiella pneumoniae NTUH-K2044, a strain causing liver abscess and meningitis. *J Bacteriol*. 2009;191(14):4492-4501. doi:10.1128/JB.00315-09

36. Sharma G, Rao S, Bansal A, Dang S, Gupta S, Gabrani R. Pseudomonas aeruginosa biofilm:

Potential therapeutic targets. *Biologicals*. 2014;42(1):1-7. doi:10.1016/j.biologicals.2013.11.001

37. Whitchurch CB, Tolker-Nielsen T, Ragas PC, Mattick JS. Extracellular DNA required for bacterial biofilm formation. *Science (80-)*. 2002;295(5559):1487. doi:10.1126/science.295.5559.1487

38. Chiba A, Sugimoto S, Sato F, Hori S, Mizunoe Y. A refined technique for extraction of extracellular matrices from bacterial biofilms and its applicability. *Microb Biotechnol*. 2015;8(3):392-403. doi:10.1111/1751-7915.12155

39. Ravichandran A, Sugiyama N, Tomita M, Swarup S, Ishihama Y. Ser/Thr/Tyr phosphoproteome analysis of pathogenic and non-pathogenic Pseudomonas species. *Proteomics*. 2009;9(10):2764-2775. doi:10.1002/pmic.200800655

40. Grangeasse C, Nessler S, Mijakovic I. Bacterial tyrosine kinases: Evolution, biological function and structural insights. *Philos Trans R Soc B Biol Sci*. 2012;367(1602):2640-2655. doi:10.1098/rstb.2011.0424

41. Whitmore SE, Lamont RJ. Tyrosine phosphorylation and bacterial virulence. *Int J Oral Sci*. 2012;4(1):1-6. doi:10.1038/ijos.2012.6

42. Kassety S, Katharios-Lanwermeyer S, O'Toole GA, Nadell CD. Differential surface competition and biofilm invasion strategies of Pseudomonas aeruginosa PA14 and PAO1. *J Bacteriol*. 2021;1248(September). doi:10.1128/JB.00265-21

43. Güvener ZT, Harwood CS. Subcellular location characteristics of the Pseudomonas aeruginosa GGDEF protein, WspR, indicate that it produces cyclic-di-GMP in response to growth on surfaces. *Mol Microbiol*. 2007;23(1):071119190133004-??? doi:10.1111/j.1365-2958.2007.06008.x

44. Luo Y, Zhao K, Baker AE, Kuchma SL, Coggan KA, Wolfgang MC, Wong GCL, O'Toole GA. A hierarchical cascade of second messengers regulates Pseudomonas aeruginosa Surface Behaviors. *MBio*. 2015;6(1):1-11. doi:10.1128/mBio.02456-14

45. Wilksch JJ, Yang J, Clements A, Gabbe JL, Short KR, Cao H, Cavaliere R, James CE, Whitchurch CB, Schembri MA, Chuah MLC, Liang Z-X, Wijburg OL, Jenney AW, Lithgow T, et al. MrkH, a Novel c-di-GMP-Dependent Transcriptional Activator, Controls Klebsiella pneumoniae Biofilm Formation by Regulating Type 3 Fimbriae Expression. Cheung A, ed. *PLoS Pathog*. 2011;7(8):e1002204. doi:10.1371/journal.ppat.1002204

Supplementary information

Table S1*:* Fe3+-IMAC column enrichment gradient. Buffer A corresponds to the loading buffer (30 %ACN, 0.07 % TFA) and B to the elution buffer (0.3 % ammonia).

Time (min)	Flow rate (ml/min)	%A	%B
0 – 7.00	0.1	100	0
7.01 – 12.00	1	100	0
12.01 – 13.50	1	40	60
13.51 – 16.00	0.5	40	60
16.01 – 25.00	1	100	0

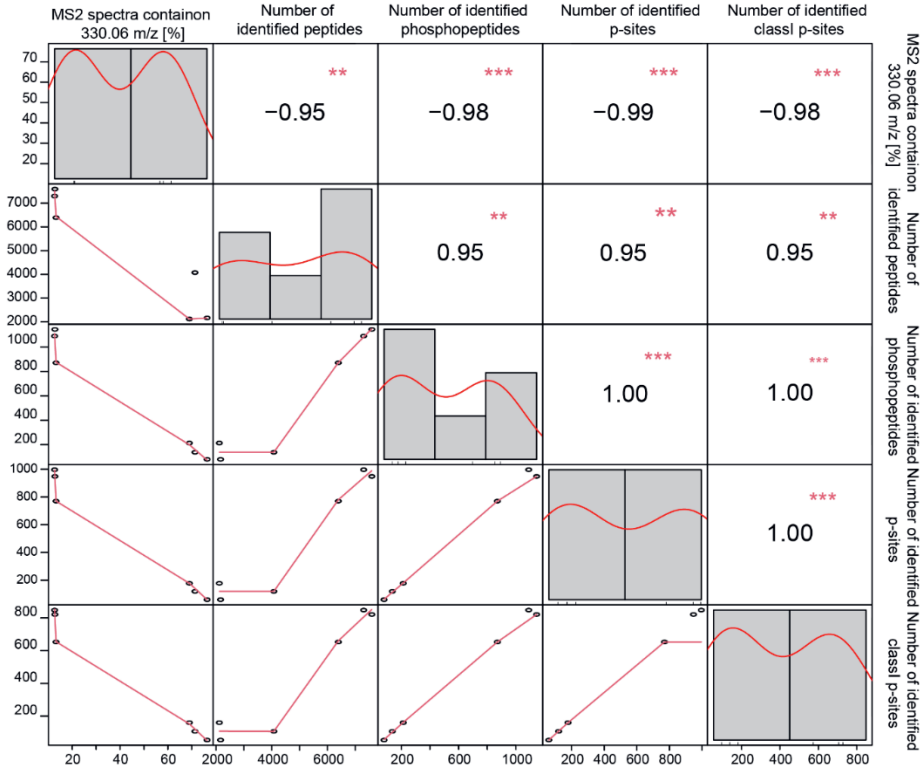


Figure S1: Correlation between MS2 spectra containing the diagnostic ion of 330.06 *m/z* and the identification in *E.coli* K12 using a standard (n=3) as well as the optimized Gram-negative protocol (n=3) as published by Potel and co-workers[1]. Pearson correlation using the PerformanceAnalytics r package[2] was used to determine if the contamination indicated by MS2 spectra containing the diagnostic ion of 330.06 *m/z* is an indication for better identification. The diagonal shows the distribution of all samples. Below the diagonal scatterplots including a fitted line (red) are displayed. Above the diagonal the correlation value, r, is depicted including the significance level. Each significance level is associated to a symbol: p-values (0, 0.001, 0.01, 0.05, 0.1, 1) <=> symbols (“***”, “**”, “*”, “.”, " “).

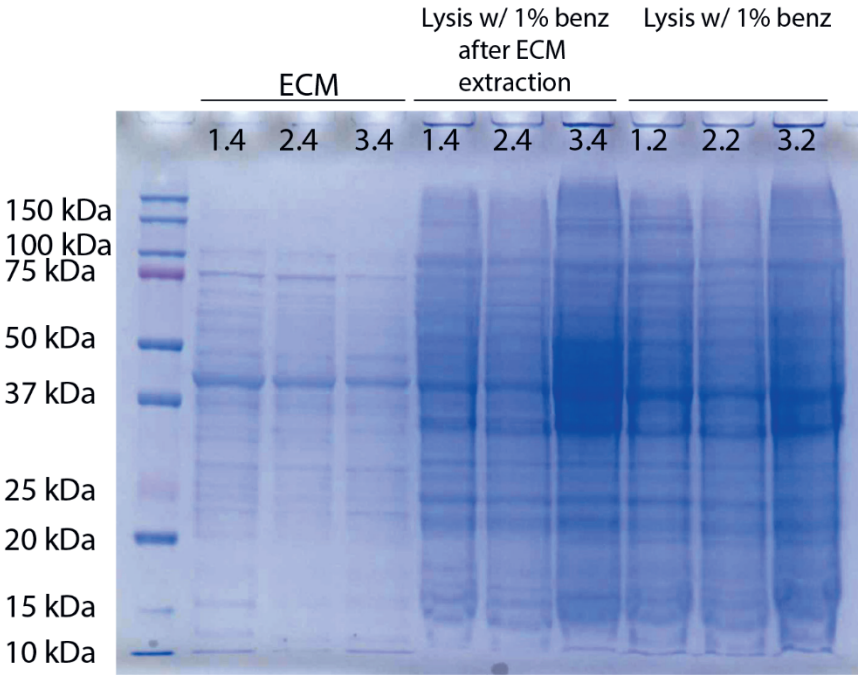


Figure S2: SDS-PAGE for the ECM fraction as well as lysates with or without ECM extraction. For all samples 10ul sample was supplemented with 10ul 2x XT-samples buffer containing 25 mM DTT and run on a 12% bis-Tris gel. 2.5ul Precision Plus Protein Dual Color Standard (BioRad) was used as Marker.

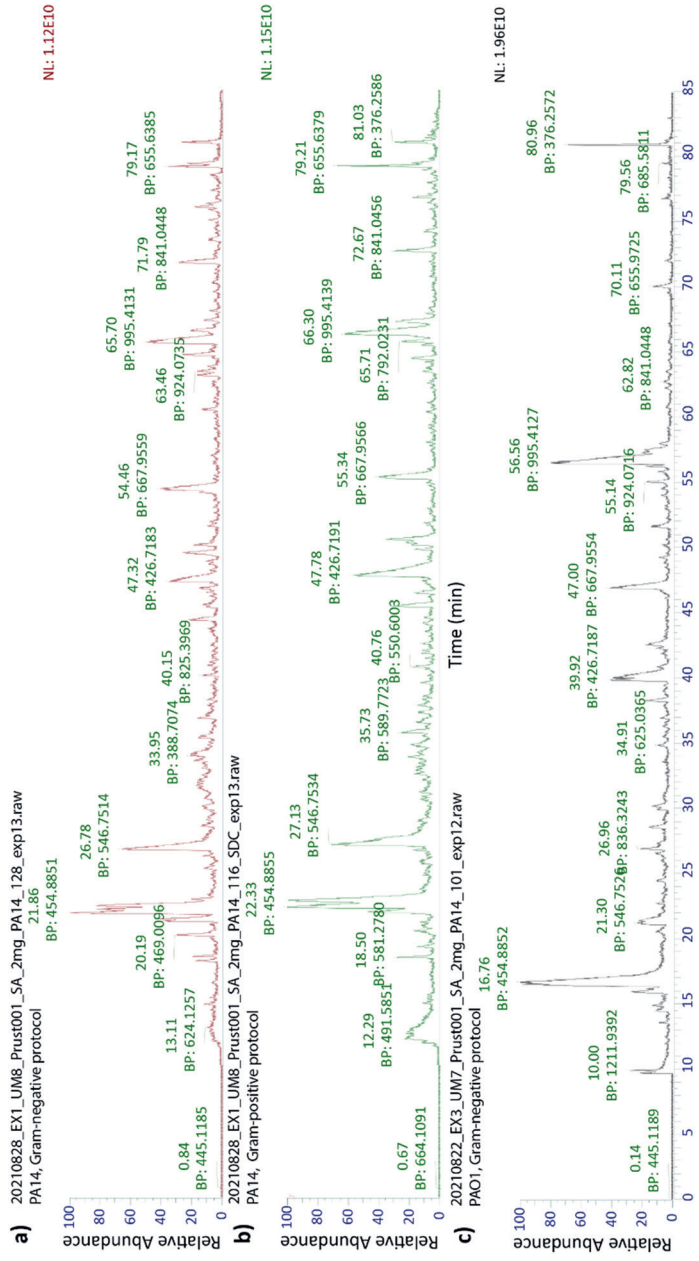


Figure S3: Total ion chromatogram (TIC) for *P. aeruginosa* PA14 after lysis using **a)** the Gram-negative and **b)** Gram-positive protocol as well as for *P. aeruginosa* PAO1 using the Gram-negative protocol, **c)**. All samples were measured using an Ultimate 3000 (Thermo Fisher Scientific, Bremen, Germany) coupled to an Orbitrap Exploris™ 480.

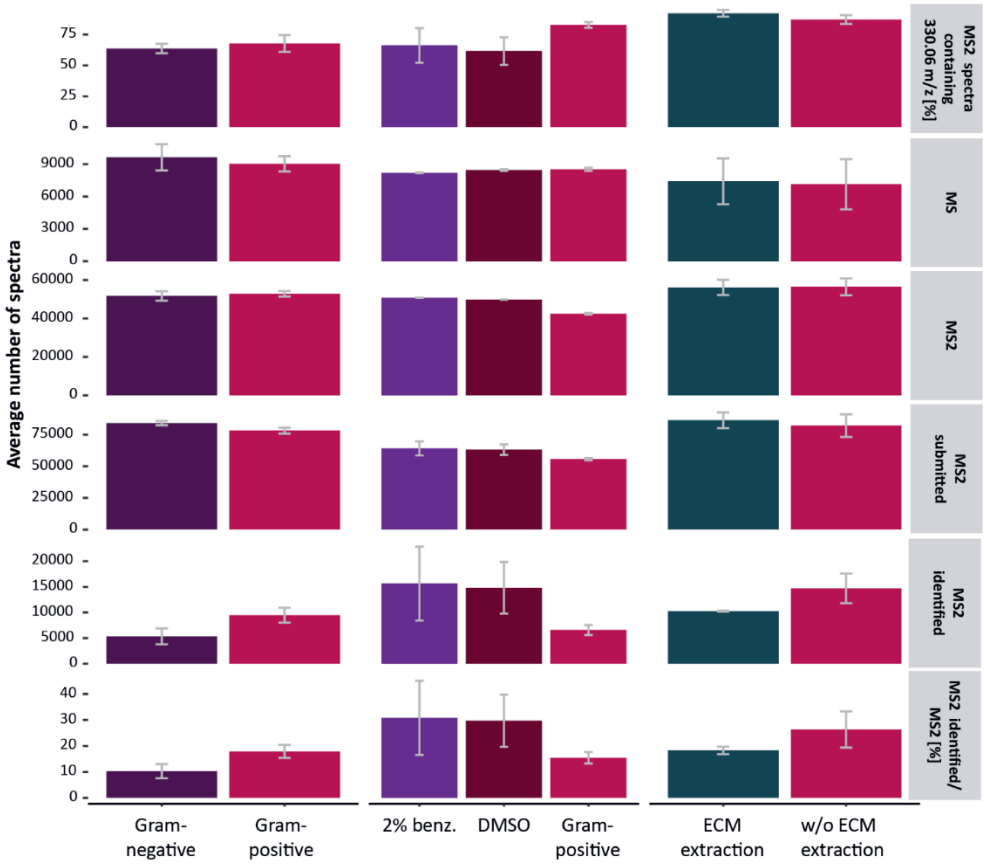


Figure S4: Overview of MS2 spectra containing the diagnostic ion 330.06 *m/z,* recorded MS and MSMS scans as well as number of MSMS scans submitted for analysis (including second peptide search) as well as identified MSMS scans. Numbers are grouped as in Figure 1 showing the comparison for the Gram-negative vs the Gram-positive protocol (n=3), 2% benzonase, 5% DMSO and Gram-positive (all n=2) as well as the Gram-positive protocol with and without ECM extraction (n=3). All bar graphs show the mean value and the standard deviation is depicted as error bar.

Figure S5: Overview of MS2 spectra containing the diagnostic ion 330.06 *m/z*, recorded MS and MSMS scans as well as number of MSMS scans submitted for analysis (including second peptide search) as well as identified MSMS scans. Numbers are grouped as in Figure 2 showing the comparison for *P. aeruginosa* PAO1 (n=4) and PA14 (n=3), *K. pneumonia* MGH (n=4) *E. coli* K12 and DH5a (n=3). All bar graphs show the mean value and the standard deviation is depicted as error bar.

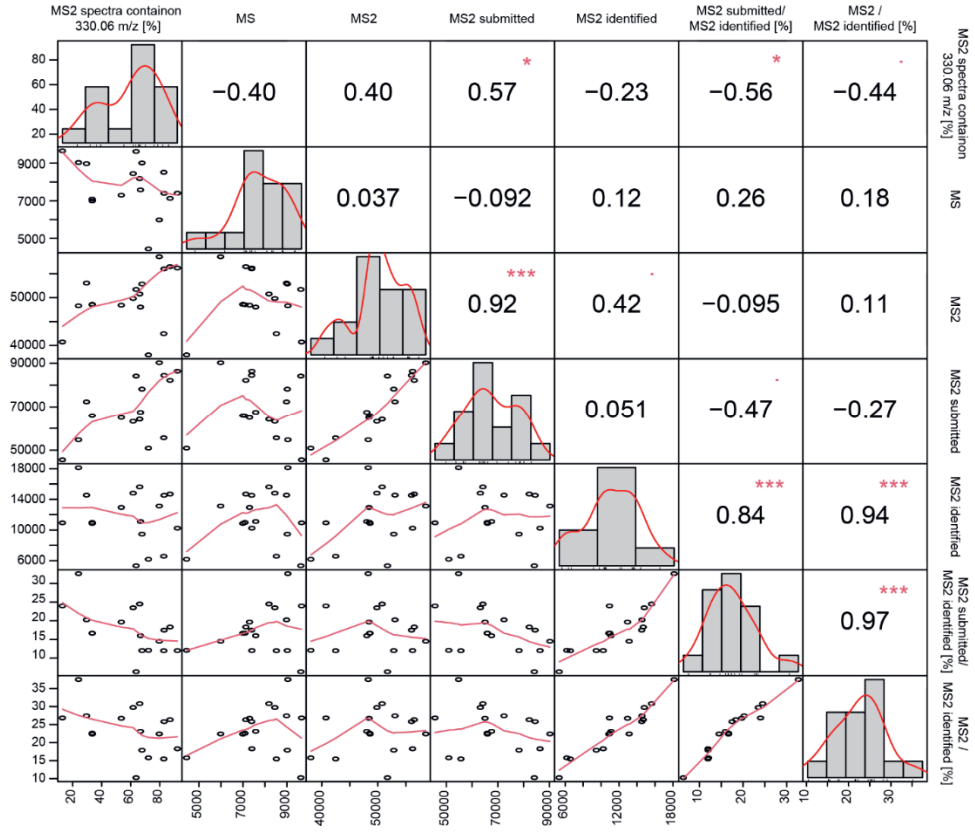


Figure S6: Correlation between MS2 spectra containing the diagnostic ion of 330.06 *m/z* and the number of recorded and identified spectra (n=18). Pearson correlation using the PerformanceAnalytics r package[2] was used to determine if the contamination indicated by MS2 spectra containing the diagnostic ion of 330.06 *m/z* is an indication for better identification. The diagonal shows the distribution of all samples. Below the diagonal scatterplots including a fitted line (red) are displayed. Above the diagonal the correlation value, r, is depicted including the significance level. Each significance level is associated to a symbol: p-values (0, 0.001, 0.01, 0.05, 0.1, 1) <=> symbols (“***”, “**”, “*”, “.”, " “).

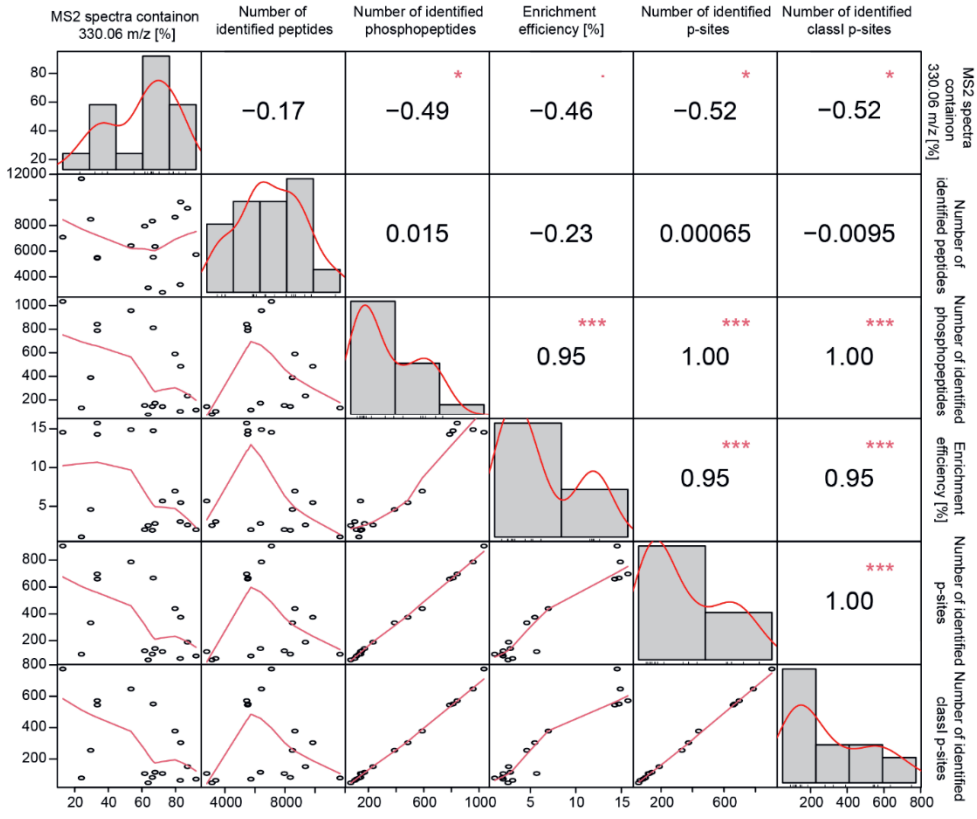


Figure S7: Correlation between MS2 spectra containing the diagnostic ion of 330.06 *m/z* and identification (n=18). Pearson correlation using the PerformanceAnalytics r package[2] was used to determine if the contamination indicated by MS2 spectra containing the diagnostic ion of 330.06 *m/z* is an indication for better identification. The diagonal shows the distribution of all samples. Below the diagonal scatterplots including a fitted line (red) are displayed. Above the diagonal the correlation value, r, is depicted including the significance level. Each significance level is associated to a symbol: p-values (0, 0.001, 0.01, 0.05, 0.1, 1) <=> symbols ("***", "**", "*", ".", " ").

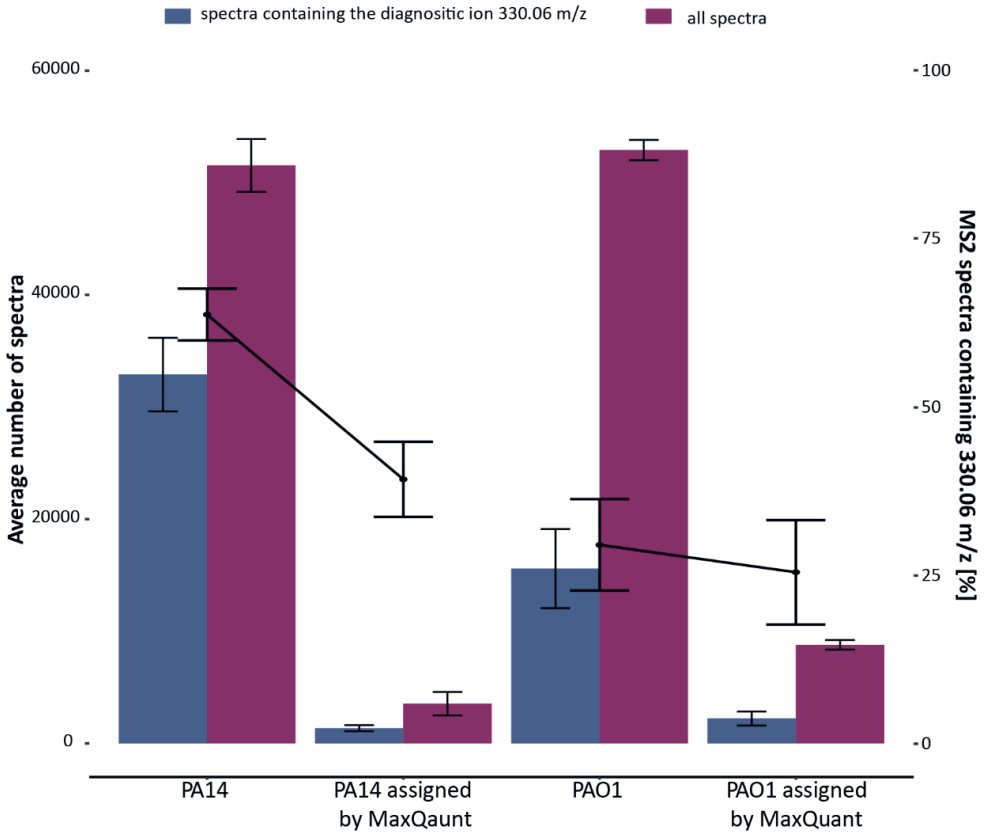


Figure S8: Average Number of recorded MS2 spectra and MS2 spectra containing the diagnostic ion 330.06 *m/z* for PA14 and PAO1 as well as the average number of MaxQuant assigned spectra and assigned spectra containing the diagnostic ion. Mgf files for assigned spectra were exported from FragmentLab. All bar graphs show the mean value and the standard deviation is depicted as error bar (PA14, n=3 and PAO1, n=4).

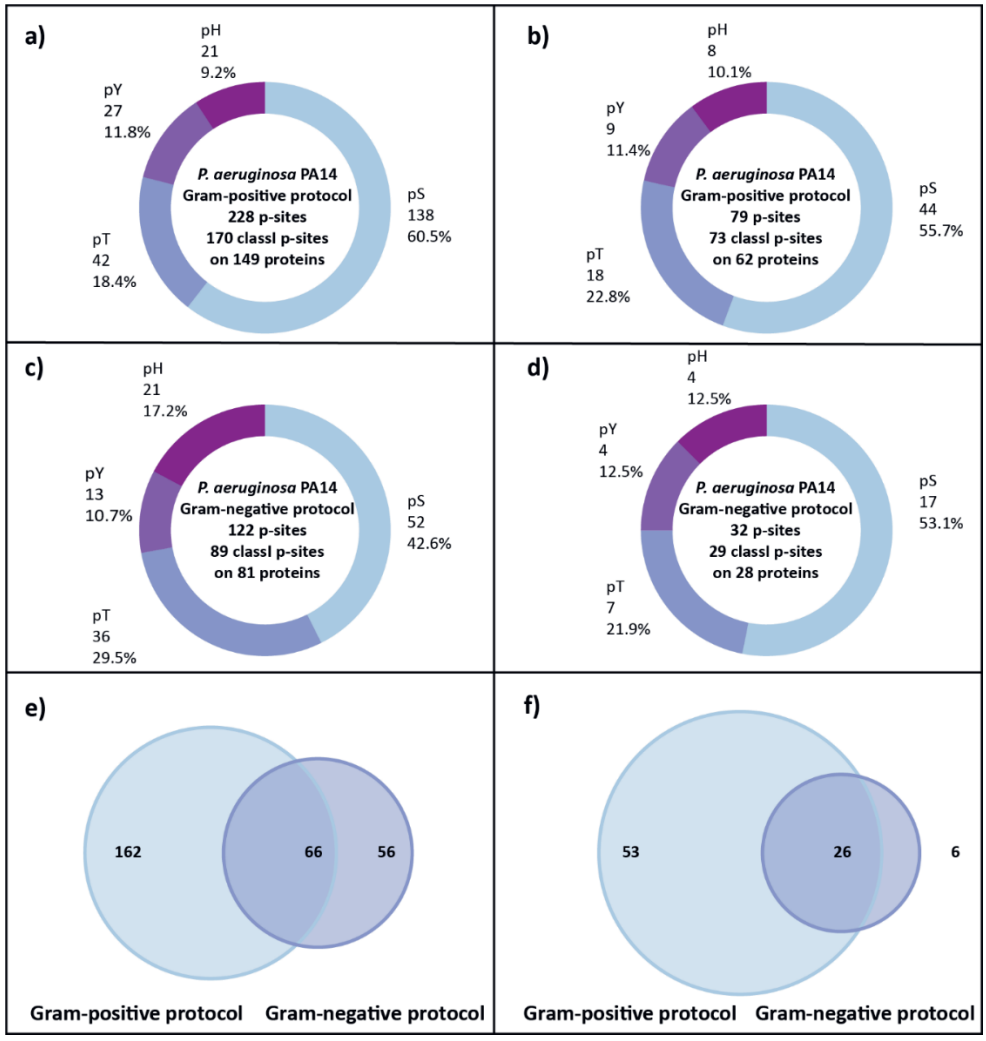


Figure S9: P-site distribution for *P. aeruginosa* PA14 using **a)** the Gram-positive protocol without valid value filtering **b)** the Gram-positive protocol with filtering for at least three valid values, **c)** the Gram-negative protocol without valid value filtering, **d)** the Gram-negative protocol with filtering for at least three valid values. The overlap of identified p-sites identified after either using the Gram-positive or Gram-negative protocol **e)** without valid value filtering or **f)** after filtering for at least three valid values (n=3).

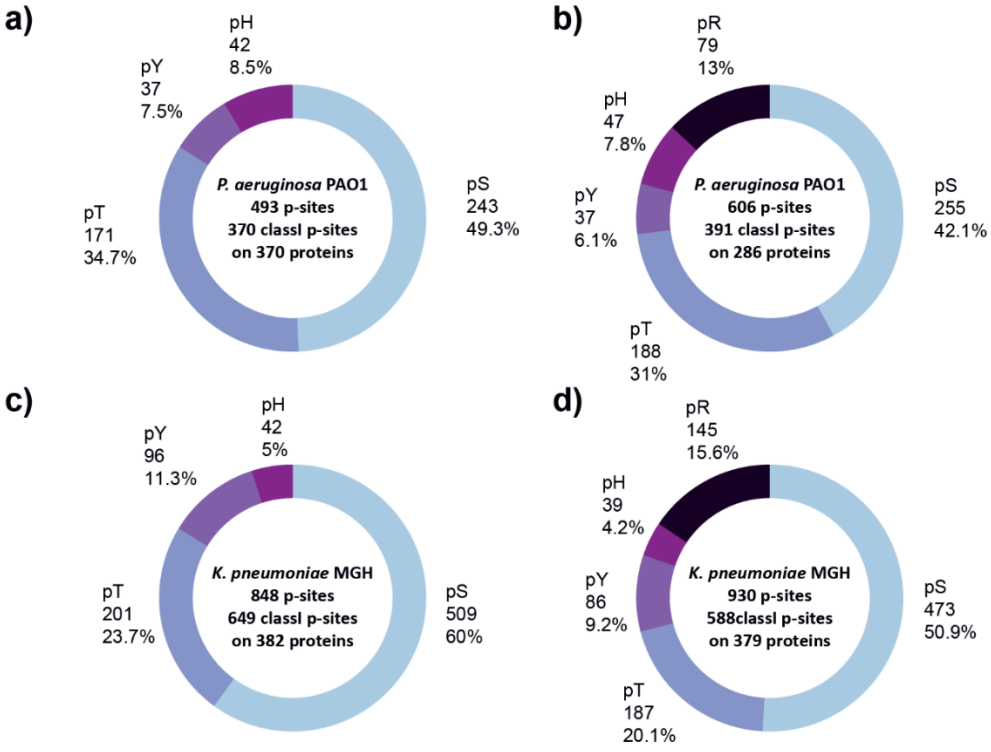


Figure S10: P-site distribution for *P. aeruginosa* PAO1 without valid value filtering searched either for pHSTY **a)** or pRHSTY **b)** and *K. pneumonia* MGH without valid value filtering either for pHSTY **c)** or pRHSTY **d)** without valid value filter (n=4). For both organisms the Gram-negative protocol was used.

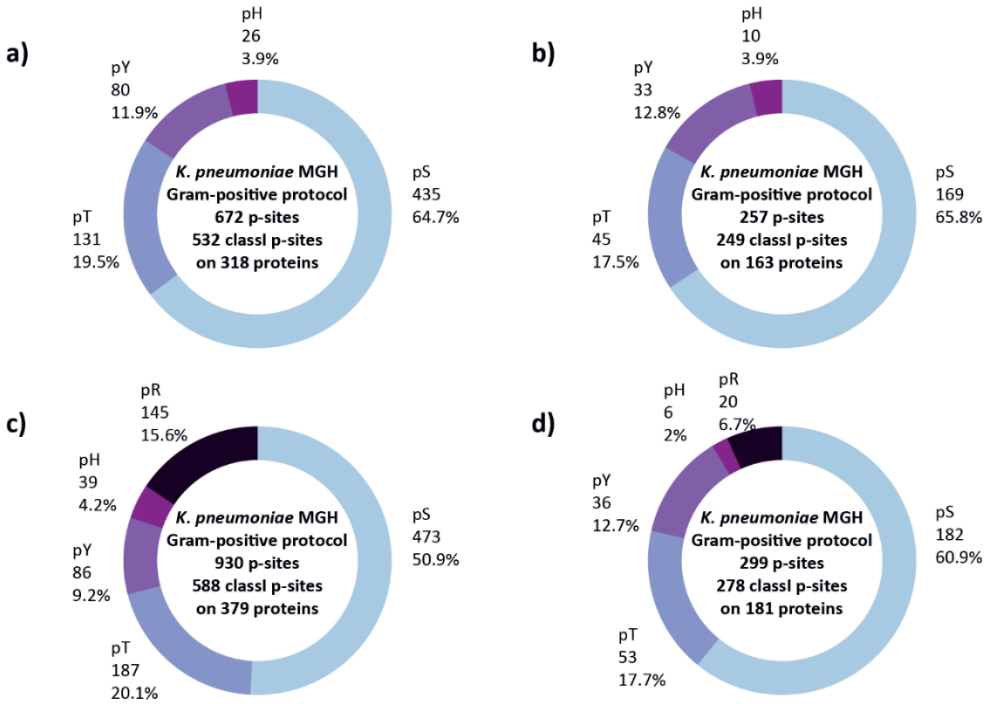


Figure S11: P-site distribution for *K. pneumonia* MGH using the Gram-positive protocol searched for either pHSTY without valid value filtering **(a)** or after filtering for at least three valid values **(b)** or searched for pRHSTY without valid value filter **(c)** or filtered for at least three valid values **(d)** (n=4).

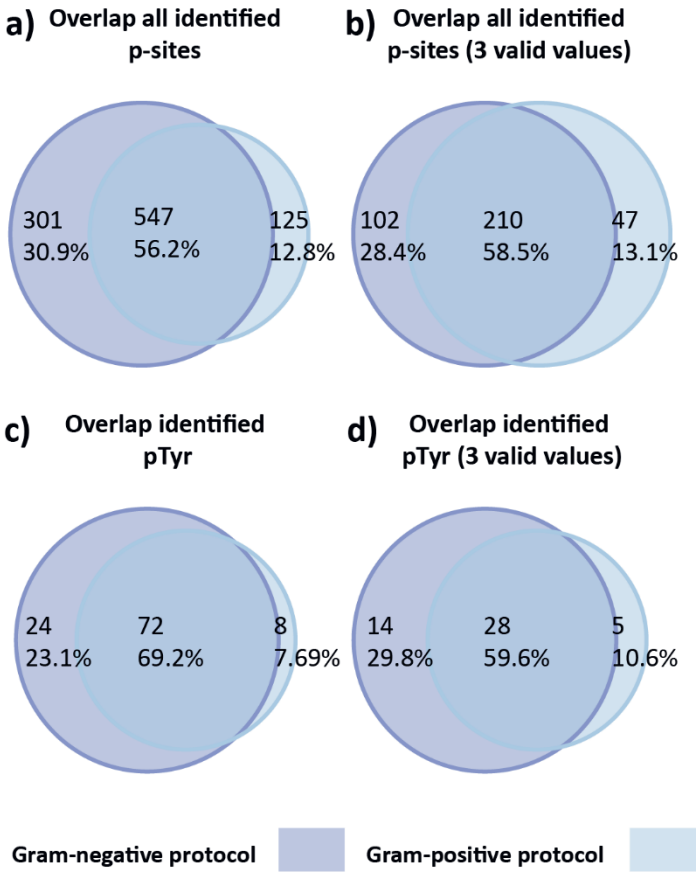
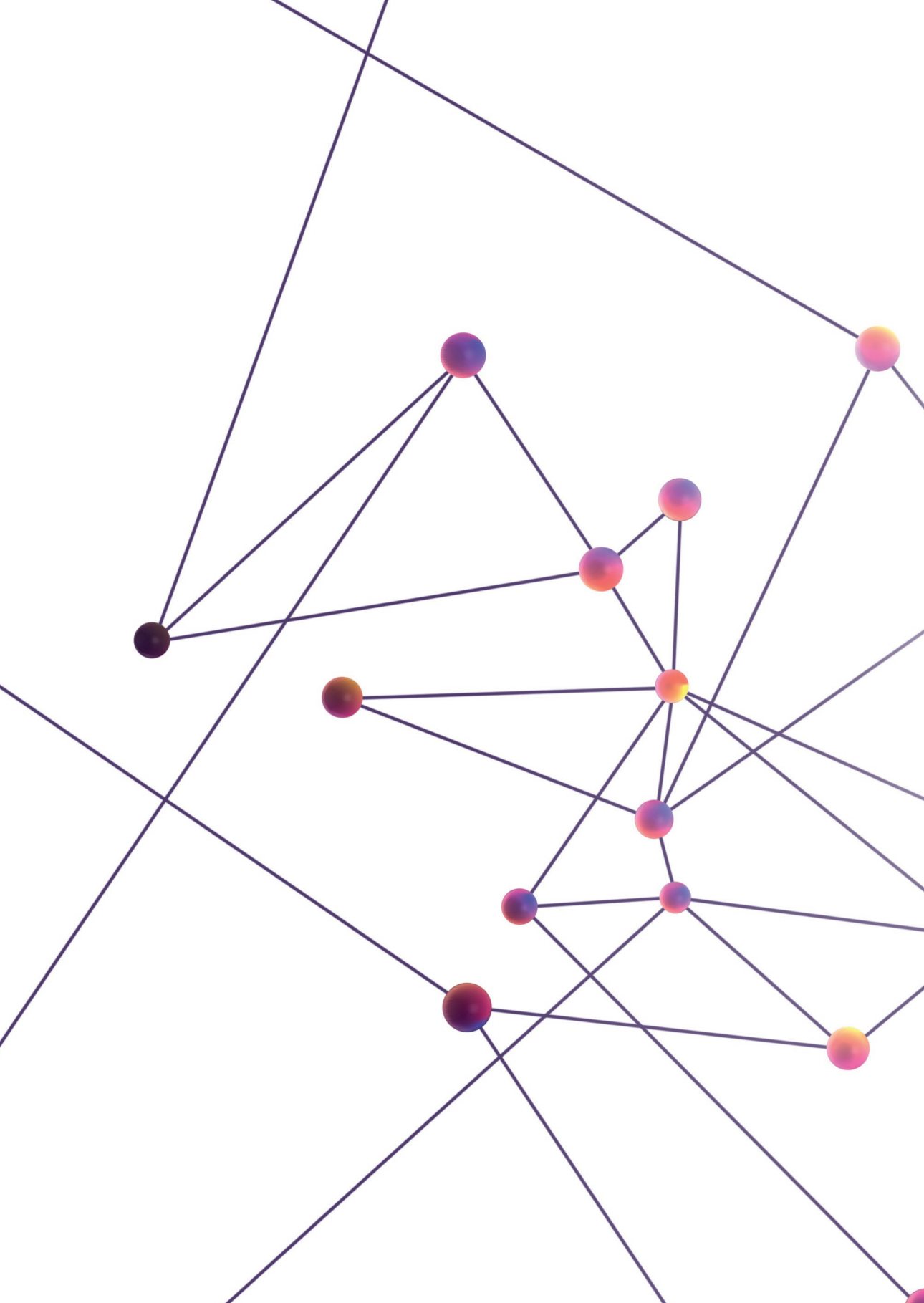
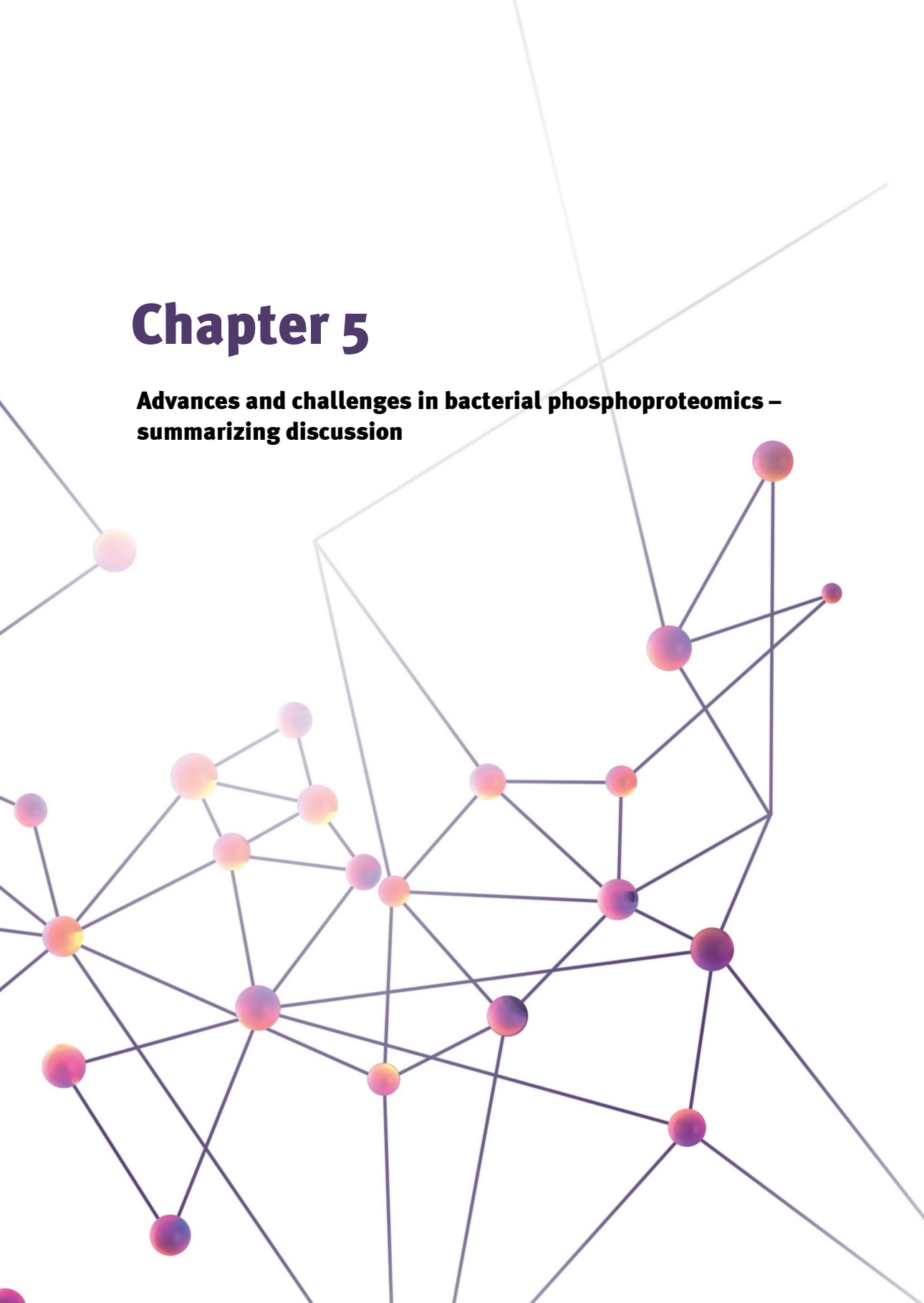


Figure S12: Overlap of identified p-sites for *K. pneumoniae* after either using the Gram-negative or Gram-positive protocol. Overlap of all identified p-sites searched for pHSTY without valid value filtering **(a)** and after filtering for at least three valid values **(b)**. Overlap of all identified pTyr sites without valid value filtering **(c)** and after filtering for at least three valid values **(d)**. The total number of sites is given as number and the sites overlapping or uniquely identified is given as percentage of the total number of identified sites.



Chapter 5

Advances and challenges in bacterial phosphoproteomics – summarizing discussion



Introduction

Bacteria can experience a multitude of natural or induced external stresses. Amongst them are changes in pH, temperature and antibiotic treatments after host infection[1–4]. To cope with these different environments, bacteria need a quick and effective way to communicate the extracellular information to intracellular responses. Part of this complex response system is controlled by reversible protein phosphorylations which are involved in the regulation of e.g. gene transcription, protein-protein interactions, protein localization or secretion of virulence factors[5–9]. Unfortunately, the sub-stoichiometric nature of protein phosphorylation complicates the analysis and consequently hampers the understanding of bacterial signaling and their virulence. To reach a comprehensive understanding of bacterial phosphoproteomics we need to overcome certain technical, as well as, intrinsic challenges.

Technical advances and current challenges

Today, thousands of phosphorylation sites can be identified in a single run and they allow a more comprehensive analysis of, especially, eukaryotic systems[10]. This is facilitated by instrumental advances, increased sensitivity, accuracy as well as increased speed, together with developments in fragmentation, enrichment strategies, as well as improved site localization[10,11]. Therefore, mass spectrometry is the gold standard in phosphoproteomics.

Sample preparation. One of the first gel-free studies using TiO_2 on *B. subtilis*[8] as well as *E. coli*[12] by Macek and co-workers only identified around 80 pSer/Thr/Tyr sites. Since then, not only due to optimized enrichment strategies, but also thanks to improved sample preparation, our knowledge of the bacterial phosphoproteome expanded greatly. It has been shown, that a good sample preparation prior to a phosphopeptide enrichment is of utmost importance for a successful enrichment[13–15]. Larger biomolecules such as phospholipids or nucleic acids that also contain phosphoryl groups can compete with phosphopeptides for the binding to metal ions such as Fe^{3+} [16]. Therefore, removal of those contaminants increased the identification of phosphorylation sites by a factor of three to around 1,300 class I sites[16]. In chapter 2, I showed that this optimization in combination with a more tailored sample preparation for Gram-positive bacteria indeed improved the identification of phosphorylation sites for *S. aureus* by a factor of 21[17] compared to recent literature[18]. Gram-positive and Gram-negative bacteria are inherently different in their cell wall compositions. While Gram-negative bacteria contain a thin peptidoglycan layer that is surrounded by an outer liposaccharide membrane, Gram-positive bacteria lack this outer membrane but instead possess a thick peptidoglycan layer[19]. Those differences need to be taken into account when optimizing their sample preparation. In this case, harsher lysis conditions are required for Gram-positive compared to Gram-negative bacteria. I also showed in chapter 4, that those optimizations are highly organism if not even strain specific.

Analyzing the Gram-negative bacteria *P. aeruginosa* as well as *K. pneumoniae* showed that the optimization does not have the same beneficial effect as seen before. Moreover, we could not observe a strong correlation between contamination, indicated by the diagnostic ion of 330.06 *m/z*, and better identification as seen for *E. coli*. This suggests that different organisms and even different strains demand more tailored protocols to meet their individual needs. *P. aeruginosa* PAO1 and PA14 share around 90% of their genome, at the same time they differ in their virulence which underlines their individuality[20]. While the same optimization on *P. aeruginosa* PAO1 seems to be beneficial for the identification, the improvement for PA14 is less strong. We hypothesize that their intrinsic differences in biofilm formation could partially contribute to the observed deviations. While the biofilm formation in PAO1 comes along with an increase in c-di-GMP, the increase in c-di-GMP is preceded by an increase in cAMP in PA14[21,22]. Those differences could explain the drastic differences in contamination measured by the presence of cAMP in MS2 spectra. This example emphasizes the need of tailored sample preparation workflows to get the most comprehensive view on different organism.

***Enrichment strategies*.** The sub-stoichiometric nature of protein phosphorylation, requires a highly effective enrichment prior to LC-MS/MS. Consequently several groups set out to optimize those methods over the last decade and one can choose from a variety of different methods nowadays[23–32]. However, metal-based affinity enrichment methods, such as immobilized metal affinity chromatography (IMAC)[23,24] and metal oxid affinity chromatography (MOAC)[25,26] gained most popularity. Both methods exploit the affinity of negatively charged phosphopeptides towards either metal ions such as Ti^{4+} and Fe^{3+} or metal oxides such as TiO_2, respectively. Even though those methods have been shown to enrich pSer, pThr and pTyr, immuno-affinity precipitation enrichment strategies, either performed individually or in combination with other methods, have been more successfully employed for pTyr analysis[33,34]. It should be noted, that all these methods have been optimized in the context of eukaryotic phosphorylation events, almost exclusively focusing of phosphomonoesters[25,26,35]. This bears several challenges for the analysis of bacterial phosphoproteomes: I) In contrast to eukaryotic systems, prokaryotes exploit phosphoramidates such as pHis and pArg or acyl phosphates such as pAsp, which are acid labile and alkaline stable[36]. Common enrichment strategies such as Fe^{3+}-IMAC operate under acid conditions and thus hamper the identification of phosphoramidates. II) Bacterial phosphorylation is even thought to be less abundant than eukaryotic phosphorylation and therefore necessitates even better enrichment strategies and larger input material[37,38]. Bäsell and co-workers[39] as well as Junker and co-workers[18] took those drawbacks into consideration and adapted their TiO_2 enrichment protocol to less acidic conditions. This allowed the identification of labile pArg for the Gram-positive bacterium *S. aureus*[18,39]. Similarly, Potel and co-workers[16] showed that shortening the gradient and less acid

conditions during the Fe^{3+}-IMAC enrichment can help with the identification of labile pHis. In chapter 3 I showed that those adjustments not only allow the identification of pHis, but also pArg. Alternatively, Trentini and co-workers[40] used a substrate-trapping mutant that can bind to phosphorylated proteins but lacks hydrolyzing activity. Those advances show the necessity of tailored protocols, especially for the analysis of bacterial phosphoproteomes. Another consideration in respect to phosphopeptide enrichments is their reproducibility and throughput. Rubrecht and co-workers showed that the Fe^{3+}-IMAC column format performs better in regard to reproducibility and analytical capacity compared to batch or tip fomats[35]. However, column-based enrichments are still labor intensive and thus not suited for high-throughput analyses. Birk and co- workers just recently published an automated phosphopeptide enrichment protocol for Gram-positive bacteria[41]. This allows for a robust, reproducible and high-throughput analysis[41]. However, due to the required buffer conditions and long loading and elution times (in total 40 min[41]), suitability for the enrichment of acid labile pHis and pArg still needs to be tested. If this method indeed would be suitable of phosphoramidates, it would allow both comprehensive as well as high-throughput screening of bacterial phosphoproteomes.

Acquisition modes. While advances over the last years allow deep proteome analysis reaching almost complete proteome coverage[42,43] and deep eukaryotic phosphoproteomic studies[44,45], the high dynamic range of bacterial phosphoproteomes remains a major challenge. This is especially the case for DDA ("data-dependent acquisition") shotgun proteomic studies. In DDA a precursor ion for fragmentation is selected based on the Top N most intense precursor in a full MS spectrum (MS1). This skews the identification to higher abundant peptides, which is especially problematic for bacterial phosphopeptides. For eukaryotic Fe^{3+}-IMAC enrichments we reach an efficiency (phosphopeptides/peptides) of more than 90%[16,35] and even for serum samples up to 60% were reached[46], whereas the efficiency for bacterial samples is around than 30%[47] or even less as shown in chapter 4. This is inherent to the even lower abundance of bacterial phosphorylation events compared to eukaryotic phosphorylation, leading to under-sampling, insufficient reproducibility and missing values. Alternative methods, such as parallel reaction monitoring (PRM)[48], selected reaction monitoring (SRM)[49] or data-independent acquisition (DIA)[50] have been developed to circumvent general random and under-sampling. PRM or SRM require *a prior* knowledge about targeted peptides and are thus not used for in-depth phosphoproteome studies, but rather the quantification of e.g. virulence factors. In DIA-MS-based acquisition the mass spectrometer cycles through predefined *m/z* segments and acquirers fragment ion spectra for all precursors in those segments, independent of their intensity[50]. These complex fragment ion maps require more elaborated data processing that is constantly improved over the recent years[51–53]. The application for eukaryotic phosphoproteomic studies has been demonstrated over the last years showing comparable if not even deeper

phosphoproteome coverage, however assignment of phosphorylation sites remains challenging[51,54]. Since bacterial peptides are mainly singly phosphorylated, their phosphorylation site assignment could be less challenging compared to eukaryotic samples. Further optimization of phosphorylation site assignment and thus reliable phosphopeptide identification using DIA-MS-based approaches could in the future improve the depth of current bacterial phosphoproteomes by circumventing under-sampling and increasing the dynamic range. The issue of correct phosphorylation site assignment is also a challenge in DDA measurements that rely on matching the identified spectra to *in-silico* generated spectra. The in silico spectra lack intensity information and do not consider a-typical fragments that can especially occur in phosphopeptide spectra. Junker and co-worker showed successfully that the use of a spectral library improved the confident identification of phosphopeptide in *S. aureus* COL and allowed the characterization of PtpB as an arginine phosphatase[55].

Technological advances as well as advances in sample preparation and enrichment strategies brought us already a long way from identifying only around 80 phosphorylation sites[12] to now more than 2,000 phosphorylation sites[56]. Yet, the field of phosphoproteomics especially bacterial phosphoproteomics is still under constant development and we need to continue to work on an improved workflow to increase the dynamic range to gain a comprehensive understanding of bacterial phosphorylation signaling.

Understanding the physiological role of bacterial phosphorylation

Over the last years our knowledge of prokaryotic phosphoproteomes has expanded and we are starting to gain a better understanding of their physiological role. Studies have linked phosphorylation events with rather basic biological functions such as metabolism and gene expression as well as bacterial pathogenicity[9,57,58]. While the physiological role of pHis is already understood for years; facilitating extracellular stimuli to intracellular responses within the two-component systems[5], we are just starting to understand the role of protein-Arg phosphorylation. Arginine phosphorylation has been linked to a general stress response and supposedly marks proteins for degradation[59–62]. To fully understand its role though, we still need to get a deeper understanding. This includes not only identifying which proteins are phosphorylated on arginine, but also quantifying pArg under different conditions to understand its physiological role. Furthermore, we also need to elucidate which organisms make use of pArg. Interestingly, we saw that *S. aureus* seems to utilize more arginine phosphorylation than *E. coli*, *B. subtilis*, *P. aeruginosa* or, *K. pneumoniae* (see Chapter 4). Therefore, elucidating the physiological role of pArg in *S. aureus* is of great importance to understand the observed differences. Long assumed to play only a minor role in prokaryotes, phosphorylations on Ser, Thr and Tyr have been shown to promote important cellular functions. pTyr has been shown to play a role in bacterial virulence and higher

abundances of pTyr is linked to virulence[9,63,64]. Ser and Thr phosphorylation has been linked to e.g. developmental processes[58,65], central metabolism[66] or cell wall synthesis[67]. Even though we already identified some physiological functions of bacterial phosphorylation, the function of the majority of the today identified phosphorylation sites remains elusive. Furthermore, for a large part the responsible kinases and phosphatases have not yet been identified. In chapter 2 we analyzed a knockout mutant of the only known Hanks-type Ser/Thr kinase of *S. aureus* USA300. Based on the large number of identified pSer and pThr sites in this mutant we hypothesized that this organism might possess yet unidentified kinases. It thus remains a future challenge to not only identify missing kinases and phosphatases but also understand the function of the "dark phosphoproteome". One should not neglect though that a great part of the identified phosphorylation sites might be non-functional and occur due to random kinase activities[68,69].

Functional analysis through homology. It has been shown that functional phosphorylation sites are more conserved than non-functional phosphorylation sites[68,69]. Inferring functionality through homology and conservation could thus help in identifying functional phosphorylation sites. Macek and co-workers showed that half of the identified *E. coli* phosphoproteins were conserved in other species[12]. Similarly it has been shown that pTyr is highly conserved, especially within closely related species[70]. Therefore, comparing identified phosphorylation sites to other studies and organisms can provide information of possible functions of a certain phosphorylation site. However, this requires comprehensive and well maintained databases that collect all identified and published phosphorylation sites throughout different species together with their functionality, if annotated. Databases such as PhosphoSitePlus[71] or PHOSIDA[72,73] collect this information for eukaryotes. Even though PHOSDIA also contains entries for prokaryotes, those are limited to *Lactococcus lactis*, *Bacillus subtilis*, *Escherichia coli* and *Halobacterium salinarium.* Furthermore, PHOSIDA has been last updated in 2012 which explains the low number of annotated phosphorylation sites for *E. coli* (81, October 2021). An alternative databases, specific for prokaryotes is the Database of Phosphorylation Sites in Prokaryotes (dbPSP), which is maintained by the Cuckoo lab[74]. The database was last updated in June 2019 and contains 19,296 experimentally identified phosphorylation sites (*E. coli* K12: 1,619 sites) and thus seems to be better maintained than PHOSIDA. This provides a good basis to find conserved phosphorylation sites to infer information from other studies and other organisms.

Host pathogen interaction. Unfortunately, identifying and understanding functional phosphorylation sites within different pathogens is not enough though. After host infiltration, bacteria are not only able to phosphorylate/dephosphorylate their own proteins, but secreted kinases/phosphatases allow the phosphorylation of host proteins and *vice versa*[75–78]. This complex interplay between host and pathogen can thus greatly influence each other's phosphoproteomes. Understanding this complex interaction is of

utmost importance to elucidate bacterial pathogenesis as well as the course of infection and should help identify new targets for antimicrobial therapies. In line with Bonne Kohler and co-worker[79] I believe only exhaustive analysis of host pathogen interactions will help identifying new drug targets which is especially important since antimicrobial resistance is on the verge[80–82].

Conclusion

Bacterial phosphoproteomics came a long way over the last decade and constant improvements in instrumentation and enrichment strategies together with sample preparations paved the way to a more comprehensive understanding of bacterial signaling. However, today we are still far from a complete understanding of phosphorylation dependent bacterial signaling. Bacteria are highly dynamic and can adapt to different environments quickly and efficiently. At the same time, they are also highly adapted to their occupied niches. This requires more tailored approaches in respect to sample preparation and possibly also enrichment strategies as shown in chapter 4. Overcoming those limitations should enable exhaustive, ideally, multi-'omics' analysis of host-pathogen interactions and thus identifying new points of attack to tackle the emerging antimicrobial resistances.

References

1. Scheuerl T, Hopkins M, Nowell RW, Rivett DW, Barraclough TG, Bell T. Bacterial adaptation is constrained in complex communities. *Nat Commun*. 2020;11(1). doi:10.1038/s41467-020-14570-z

2. Rainey PB, Travisano M. Adaptive radiation in a heterogeneous environment. *Nature*. 1998;394(6688):69-72. doi:10.1038/27900

3. Zdziarski J, Brzuszkiewicz E, Wullt B, Liesegang H, Biran D, Voigt B, Grönberg-Hernandez J, Ragnarsdottir B, Hecker M, Ron EZ, Daniel R, Gottschalk G, Hacker J, Svanborg C, Dobrindt U. Host imprints on bacterial genomes-rapid, divergent evolution in individual patients. *PLoS Pathog*. 2010;6(8):95-96. doi:10.1371/journal.ppat.1001078

4. Tuchscherr L, Medina E, Hussain M, Völker W, Heitmann V, Niemann S, Holzinger D, Roth J, Proctor RA, Becker K, Peters G, Löffler B. Staphylococcus aureus phenotype switching: an effective bacterial strategy to escape host immune response and establish a chronic infection. *EMBO Mol Med*. 2011;3(3):129-141. doi:10.1002/emmm.201000115

5. Stock JB, Stock AM, Mottonen JM. Signal transduction in bacteria. *Nature*. 1990;344(6265):395-400. doi:10.1038/344395a0

6. Villanueva M, García B, Valle J, Rapún B, Ruiz De Los Mozos I, Solano C, Martí M, Penadés JR, Toledo-Arana A, Lasa I. Sensory deprivation in Staphylococcus aureus. *Nat Commun*. 2018;9(1):1-12. doi:10.1038/s41467-018-02949-y

7. Jung K, Fried L, Behr S, Heermann R. Histidine kinases and response regulators in networks. *Curr Opin Microbiol*. 2012;15(2):118-124. doi:10.1016/j.mib.2011.11.009

8. Macek B, Mijakovic I, Olsen J V., Gnad F, Kumar C, Jensen PR, Mann M. The Serine/Threonine/Tyrosine Phosphoproteome of the Model Bacterium Bacillus subtilis. *Mol Cell Proteomics*. 2007;6(4):697-707. doi:10.1074/mcp.M600464-MCP200

9. Grangeasse C, Cozzone AJ, Deutscher J, Mijakovic I. Tyrosine phosphorylation: an emerging regulatory device of bacterial physiology. *Trends Biochem Sci*. 2007;32(2):86-94. doi:10.1016/j.tibs.2006.12.004

10. Doll S, Burlingame AL. Mass spectrometry-based detection and assignment of protein posttranslational modifications. *ACS Chem Biol*. 2015;10(1):63-71. doi:10.1021/cb500904b

11. Olsen J V., Mann M. Status of large-scale analysis of posttranslational modifications by mass spectrometry. *Mol Cell Proteomics*. 2013;12(12):3444-3452. doi:10.1074/mcp.O113.034181

12. Macek B, Gnad F, Soufi B, Kumar C, Olsen J V., Mijakovic I, Mann M. Phosphoproteome Analysis of E. coli Reveals Evolutionary Conservation of Bacterial Ser/Thr/Tyr Phosphorylation. *Mol Cell Proteomics*. 2008;7(2):299-307. doi:10.1074/mcp.M700311-MCP200

13. Thingholm TE, Jensen ON, Larsen MR. Analytical strategies for phosphoproteomics. *Proteomics*. 2009;9(6):1451-1468. doi:10.1002/pmic.200800454

14. Solari FA, Dell'Aica M, Sickmann A, Zahedi RP. Why phosphoproteomics is still a challenge. *Mol BioSyst*. 2015;11(6):1487-1493. doi:10.1039/c5mb00024f

15. Zhang X, Ye J, Jensen ON, Roepstorff P. Highly efficient phosphopeptide enrichment by calcium phosphate precipitation combined with subsequent IMAC enrichment. *Mol Cell Proteomics*. 2007;6(11):2032-2042. doi:10.1074/mcp.M700278-MCP200

16. Potel CM, Lin M-H, Heck AJR, Lemeer S. Defeating major contaminants in Fe 3+-IMAC phosphopeptide enrichment. *Mol Cell Proteomics*. Published online 2018:mcp.TIR117.000518. doi:10.1074/mcp.TIR117.000518

17. Prust N, van der Laarse S, van den Toorn HWP, van Sorge NM, Lemeer S. In-Depth Characterization of the Staphylococcus aureus Phosphoproteome Reveals New Targets of Stk1. *Mol Cell Proteomics*. 2021;20:100034. doi:10.1074/mcp.RA120.002232

18. Junker S, Maaβ S, Otto A, Michalik S, Morgenroth F, Gerth U, Hecker M, Becher D.

Spectral Library Based Analysis of Arginine Phosphorylations in *Staphylococcus aureus*. *Mol Cell Proteomics*. 2018;17(2):335-348. doi:10.1074/mcp.RA117.000378

19. Swoboda JG, Campbell J, Meredith TC, Walker S. Wall teichoic acid function, biosynthesis, and inhibition. *Chembiochem*. 2010;11(1):35-45. doi:10.1002/cbic.200900557

20. Lee DG, Urbach JM, Wu G, Liberati NT, Feinbaum RL, Miyata S, Diggins LT, He J, Saucier M, Déziel E, Friedman L, Li L, Grills G, Montgomery K, Kucherlapati R, et al. Genomic analysis reveals that Pseudomonas aeruginosa virulence is combinatorial. *Genome Biol*. 2006;7(10). doi:10.1186/gb-2006-7-10-r90

21. Kassety S, Katharios-Lanwermeyer S, O'Toole GA, Nadell CD. Differential surface competition and biofilm invasion strategies of Pseudomonas aeruginosa PA14 and PAO1. *J Bacteriol*. 2021;1248(September). doi:10.1128/JB.00265-21

22. Lee CK, Vachier J, de Anda J, Zhao K, Baker AE, Bennett RR, Armbruster CR, Lewis KA, Tarnopol RL, Lomba CJ, Hogan DA, Parsek MR, O'Toole GA, Golestanian R, Wong GCL. Social Cooperativity of Bacteria during Reversible Surface Attachment in Young Biofilms: a Quantitative Comparison of Pseudomonas aeruginosa PA14 and PAO1. Ausubel FM, ed. *MBio*. 2020;11(1):1-21. doi:10.1128/mBio.02644-19

23. Posewitz MC, Tempst P. Immobilized gallium(III) affinity chromatography of phosphopeptides. *Anal Chem*. 1999;71(14):2883-2892. doi:10.1021/ac981409y

24. Villén J, Gygi SP. The SCX/IMAC enrichment approach for global phosphorylation analysis by mass spectrometry. *Nat Protoc*. 2008;3(10):1630-1638. doi:10.1038/nprot.2008.150

25. Larsen MR, Thingholm TE, Jensen ON, Roepstorff P, D. JJ. Highly Selective Enrichment of Phosphorylated Peptides from Peptide Mixtures Using Titanium Dioxide Microcolumns. *Mol Cell Proteomics*. 2005;4(7):873-886. doi:10.1074/mcp.T500007-MCP200

26. Pinkse MWH, Uitto PM, Hilhorst MJ, Ooms B, Heck AJR. 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(14):3935-3943. doi:10.1021/ac0498617

27. Ficarro SB, McCleland ML, Stukenberg PT, Burke DJ, Ross MM, Shabanowitz J, Hunt DF, White FM. Phosphoproteome analysis by mass spectrometry and its application to Saccharomyces cerevisiae. *Nat Biotechnol*. 2002;20(3):301-305. doi:10.1038/nbt0302-301

28. Loroch S, Zahedi RP, Sickmann A. Highly sensitive phosphoproteomics by tailoring solid-phase extraction to electrostatic repulsion-hydrophilic interaction chromatography. *Anal Chem*. 2015;87(3):1596-1604. doi:10.1021/ac502708m

29. Smith PA, Koehler MFT, Girgis HS, Yan D, Chen Y, Chen Y, Crawford JJ, Durk MR, Higuchi RI, Kang J, Murray J, Paraselli P, Park S, Phung W, Quinn JG, et al. Optimized arylomycins are a new class of Gram-negative antibiotics. *Nature*. 2018;561(7722):189-194. doi:10.1038/s41586-018-0483-6

30. Batth TS, Francavilla C, Olsen J V. Off-line high-pH reversed-phase fractionation for in-depth phosphoproteomics. *J Proteome Res*. 2014;13(12):6176-6186. doi:10.1021/pr500893m

31. Han G, Ye M, Zhou H, Jiang X, Feng S, Jiang X, Tian R, Wan D, Zou H, Gu J. Large-scale phosphoproteome analysis of human liver tissue by enrichment and fractionation of phosphopeptides with strong anion exchange chromatography. *Proteomics*. 2008;8(7):1346-1361. doi:10.1002/pmic.200700884

32. SUI S, WANG J, LU Z, CAI Y, ZHANG Y, YU W, QIAN X. Phosphopeptide enrichment strategy based on strong cation exchange chromatography. *Chinese J Chromatogr*. 2008;26(2):195-199. doi:10.1016/S1872-2059(08)60012-7

33. Di Palma S, Zoumaro-Djayoon A, Peng M, Post H, Preisinger C, Munoz J, Heck AJR. Finding the same needles in the haystack? A comparison of phosphotyrosine peptides enriched by immuno-affinity precipitation and metal-based affinity chromatography. *J Proteomics*. 2013;91:331-337. doi:10.1016/j.jprot.2013.07.024

34. Lind SB, Molin M, Savitski MM, Emilsson L, Åström J, Hedberg L, Adams C, Nielsen ML, Engström Å, Elflneh L, Andersson E, Zubarev RA,

Pettersson U. Immunoaffinity enrichments followed by mass spectrometric detection for studying global protein tyrosine phosphorylation. *J Proteome Res*. 2008;7(7):2897-2910. doi:10.1021/pr8000546

35. Ruprecht B, Koch H, Medard G, Mundt M, Kuster B, Lemeer S. Comprehensive and Reproducible Phosphopeptide Enrichment Using Iron Immobilized Metal Ion Affinity Chromatography (Fe-IMAC) Columns. *Mol Cell Proteomics*. 2015;14(1):205-215. doi:10.1074/mcp.M114.043109

36. Besant P, Attwood P, Piggott M. Focus on Phosphoarginine and Phospholysine. *Curr Protein Pept Sci*. 2009;10(6):536-550. doi:10.2174/138920309789630598

37. Gnad F, Forner F, Zielinska DF, Birney E, Gunawardena J, Mann M. Evolutionary Constraints of Phosphorylation in Eukaryotes, Prokaryotes, and Mitochondria. *Mol Cell Proteomics*. 2010;9(12):2642-2653. doi:10.1074/mcp.M110.001594

38. Krüger R, Kübler D, Pallissé R, Burkovski A, Lehmann WD. Protein and Proteome Phosphorylation Stoichiometry Analysis by Element Mass Spectrometry. *Anal Chem*. 2006;78(6):1987-1994. doi:10.1021/ac051896z

39. Bäsell K, Otto A, Junker S, Zühlke D, Rappen G-M, Schmidt S, Hentschker C, Macek B, Ohlsen K, Hecker M, Becher D. The phosphoproteome and its physiological dynamics in Staphylococcus aureus. *Int J Med Microbiol*. 2014;304(2):121-132. doi:10.1016/j.ijmm.2013.11.020

40. Trentini DB, Fuhrmann J, Mechtler K, Clausen T. Chasing phosphoarginine proteins: Development of a selective enrichment method using a phosphatase trap. *Mol Cell Proteomics*. 2014;13(8):1953-1964. doi:10.1074/mcp.O113.035790

41. Birk MS, Charpentier E, Frese CK. Automated Phosphopeptide Enrichment for Gram-Positive Bacteria. *J Proteome Res*. 2021;20(10):4886-4892. doi:10.1021/acs.jproteome.1c00364

42. Hebert AS, Richards AL, Bailey DJ, Ulbrich A, Coughlin EE, Westphall MS, Coon JJ. The One Hour Yeast Proteome. *Mol Cell Proteomics*. 2014;13(1):339-347. doi:10.1074/mcp.M113.034769

43. Nagaraj N, Kulak NA, Cox J, Neuhauser N, Mayr K, Hoerning O, Vorm O, Mann M. System-wide perturbation analysis with nearly complete coverage of the yeast proteome by single-shot ultra HPLC runs on a bench top orbitrap. *Mol Cell Proteomics*. 2012;11(3):1-11. doi:10.1074/mcp.M111.013722

44. Sharma K, D'Souza RCJ, Tyanova S, Schaab C, Wiśniewski JR, Cox J, Mann M. Ultradeep Human Phosphoproteome Reveals a Distinct Regulatory Nature of Tyr and Ser/Thr-Based Signaling. *Cell Rep*. 2014;8(5):1583-1594. doi:10.1016/j.celrep.2014.07.036

45. Mertins P, Tang LC, Krug K, Clark DJ. Reproducible work fl ow for multiplexed deep-scale proteome and phosphoproteome analysis of tumor tissues by liquid chromatography – mass spectrometry. 2018;13(July). doi:10.1038/s41596-018-0006-9

46. Zhou W, Ross MM, Tessitore A, Ornstein D, VanMeter A, Liotta LA, Petricoin EF. An initial characterization of the serum phosphoproteome. *J Proteome Res*. 2009;8(12):5523-5531. doi:10.1021/pr900603n

47. Potel CM, Lin M-H, Heck AJR, Lemeer S. Defeating Major Contaminants in Fe 3+ - Immobilized Metal Ion Affinity Chromatography (IMAC) Phosphopeptide Enrichment. *Mol Cell Proteomics*. 2018;17(5):1028-1034. doi:10.1074/mcp.TIR117.000518

48. Peterson AC, Russell JD, Bailey DJ, Westphall MS, Coon JJ. Parallel reaction monitoring for high resolution and high mass accuracy quantitative, targeted proteomics. *Mol Cell Proteomics*. 2012;11(11):1475-1488. doi:10.1074/mcp.O112.020131

49. Picotti P, Rinner O, Stallmach R, Dautel F, Farrah T, Domon B, Wenschuh H, Aebersold R. High-throughput generation of selected reaction-monitoring assays for proteins and proteomes. *Nat Methods*. 2010;7(1):43-46. doi:10.1038/nmeth.1408

50. Gillet LC, Navarro P, Tate S, Röst H, Selevsek N, Reiter L, Bonner R, Aebersold R. 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(6):1-17. doi:10.1074/mcp.O111.016717

51. Bekker-Jensen DB, Bernhardt OM, Hogrebe A, Martinez-Val A, Verbeke L, Gandhi T, Kelstrup CD, Reiter L, Olsen J V. Rapid and site-specific deep phosphoproteome profiling by data-independent acquisition without the need for spectral libraries. *Nat Commun*. 2020;11(1):1-12. doi:10.1038/s41467-020-14609-1

52. Demichev V, Messner CB, Vernardis SI, Lilley KS, Ralser M. DIA-NN: neural networks and interference correction enable deep proteome coverage in high throughput. *Nat Methods*. 2020;17(1):41-44. doi:10.1038/s41592-019-0638-x

53. Searle BC, Lawrence RT, MacCoss MJ, Villén J. Thesaurus: quantifying phosphopeptide positional isomers. *Nat Methods*. 2019;16(8):703-706. doi:10.1038/s41592-019-0498-4

54. Shui W, Lou R, Liu W, Li R, Li S. DeepPhospho: Accelerate DIA phosphoproteome profiling by Deep Learning. Published online 2021:1-29.

55. Junker S, Maaβ S, Otto A, Michalik S, Morgenroth F, Gerth U, Hecker M, Becher D. Spectral Library Based Analysis of Arginine Phosphorylations in Staphylococcus aureus. *Mol Cell Proteomics*. 2018;17(2):335-348. doi:10.1074/mcp.RA117.000378

56. Lin M-H, Potel CM, Tehrani KHME, Heck AJR, Martin NI, Lemeer S. A New Tool to Reveal Bacterial Signaling Mechanisms in Antibiotic Treatment and Resistance. *Mol Cell Proteomics*. 2018;17(12):2496-2507. doi:10.1074/mcp.ra118.000880

57. Kalantari A, Derouiche A, Shi L, Mijakovic I. Serine/threonine/tyrosine phosphorylation regulates DNA binding of bacterial transcriptional regulators. *Microbiology*. 2015;161(9):1720-1729. doi:10.1099/mic.0.000148

58. Fridman M, Williams GD, Muzamal U, Hunter H, Siu KWMWMMWM, Golemi-Kotra D. Two unique phosphorylation-driven signaling pathways crosstalk in staphylococcus aureus to modulate the cell-wall charge: Stk1/Stp1 meets GraSR. *Biochemistry*. 2013;52(45):7975-7986. doi:10.1021/bi401177n

59. Fuhrmann J, Schmidt A, Spiess S, Lehner A, Turgay K, Mechtler K, Charpentier E, Clausen T. McsB Is a Protein Arginine Kinase That Phosphorylates and Inhibits the Heat-Shock Regulator CtsR. *Science (80-)*. 2009;324(5932):1323-1327. doi:10.1126/science.1170088

60. Elsholz AKW, Turgay K, Michalik S, Hessling B, Gronau K, Oertel D, Mader U, Bernhardt J, Becher D, Hecker M, Gerth U. Global impact of protein arginine phosphorylation on the physiology of Bacillus subtilis. *Proc Natl Acad Sci*. 2012;109(19):7451-7456. doi:10.1073/pnas.1117483109

61. Junker S, Maaß S, Otto A, Hecker M, Becher D. Toward the Quantitative Characterization of Arginine Phosphorylations in Staphylococcus aureus. *J Proteome Res*. 2018;18(1):acs.jproteome.8b00579. doi:10.1021/acs.jproteome.8b00579

62. Wozniak DJ, Tiwari KB, Soufan R, Jayaswal RK. The mcsB gene of the clpC operon is required for stress tolerance and virulence in Staphylococcus aureus. *Microbiol (United Kingdom)*. 2012;158(10):2568-2576. doi:10.1099/mic.0.060749-0

63. Grangeasse C, Nessler S, Mijakovic I. Bacterial tyrosine kinases: Evolution, biological function and structural insights. *Philos Trans R Soc B Biol Sci*. 2012;367(1602):2640-2655. doi:10.1098/rstb.2011.0424

64. Whitmore SE, Lamont RJ. Tyrosine phosphorylation and bacterial virulence. *Int J Oral Sci*. 2012;4(1):1-6. doi:10.1038/ijos.2012.6

65. Pereira SFF, Gonzalez RL, Dworkin J. Protein synthesis during cellular quiescence is inhibited by phosphorylation of a translational elongation factor. *Proc Natl Acad Sci U S A*. 2015;112(25):E3274-E3281. doi:10.1073/pnas.1505297112

66. Cowley S, Ko M, Pick N, Chow R, Downing KJ, Gordhan BG, Betts JC, Mizrahi V, Smith DA, Stokes RW, Av-Gay Y. The Mycobacterium tuberculosis protein serine/threonine kinase PknG is linked to cellular glutamate/glutamine levels and is important for growth in vivo. *Mol Microbiol*. 2004;52(6):1691-1702. doi:10.1111/j.1365-2958.2004.04085.x

67. Beltramini AM, Mukhopadhyay CD, Pancholi V. Modulation of cell wall structure and antimicrobial susceptibility by a Staphylococcus aureus eukaryote-like serine/threonine kinase

and phosphatase. *Infect Immun*. 2009;77(4):1406-1416. doi:10.1128/IAI.01499-08

68. Landry CR, Levy ED, Michnick SW. Weak functional constraints on phosphoproteomes. *Trends Genet*. 2009;25(5):193-197. doi:10.1016/j.tig.2009.03.003

69. Levy ED, Michnick SW, Landry CR. Protein abundance is key to distinguish promiscuous from functional phosphorylation based on evolutionary information. *Philos Trans R Soc B Biol Sci*. 2012;367(1602):2594-2606. doi:10.1098/rstb.2012.0078

70. Hansen AM, Chaerkady R, Sharma J, Díaz-Mejía JJ, Tyagi N, Renuse S, Jacob HKC, Pinto SM, Sahasrabuddhe NA, Kim MS, Delanghe B, Srinivasan N, Emili A, Kaper JB, Pandey A. The Escherichia coli Phosphotyrosine Proteome Relates to Core Pathways and Virulence. *PLoS Pathog*. 2013;9(6). doi:10.1371/journal.ppat.1003403

71. Hornbeck P V., Kornhauser JM, Tkachev S, Zhang B, Skrzypek E, Murray B, Latham V, Sullivan M. PhosphoSitePlus: A comprehensive resource for investigating the structure and function of experimentally determined post-translational modifications in man and mouse. *Nucleic Acids Res*. 2012;40(D1):261-270. doi:10.1093/nar/gkr1122

72. Gnad F, Ren S, Cox J, Olsen J V., Macek B, Oroshi M, Mann M. PHOSIDA (phosphorylation site database): Management, structural and evolutionary investigation, and prediction of phosphosites. *Genome Biol*. 2007;8(11). doi:10.1186/gb-2007-8-11-r250

73. Gnad F, Gunawardena J, Mann M. PHOSIDA 2011: The posttranslational modification database. *Nucleic Acids Res*. 2011;39(SUPPL. 1):253-260. doi:10.1093/nar/gkq1159

74. Pan Z, Wang B, Zhang Y, Wang Y, Ullah S, Jian R, Liu Z, Xue Y. DbPSP: A curated database for protein phosphorylation sites in prokaryotes. *Database*. 2015;2015(August 2018):1-8. doi:10.1093/database/bav031

75. DeVinney R, Nisan I, Ruschkowski S, Rosenshine I, Finlay BB. Tir tyrosine phosphorylation and pedestal formation are delayed in enteropathogenic Escherichia coli sepZ::TnphoA mutant 30-5-1(3). *Infect Immun*. 2001;69(1):559-563. doi:10.1128/IAI.69.1.559-563.2001

76. Wattiau P, Bernier B, Deslée P, Michiels T, Cornelis GR. Individual chaperones required for Yop secretion by Yersinia. *Proc Natl Acad Sci U S A*. 1994;91(22):10493-10497. doi:10.1073/pnas.91.22.10493

77. Selbach M, Moese S, Hauck CR, Meyer TF, Backert S. Src is the kinase of the Helicobacter pylori CagA protein in vitro and in vivo. *J Biol Chem*. 2002;277(9):6775-6778. doi:10.1074/jbc.C100754200

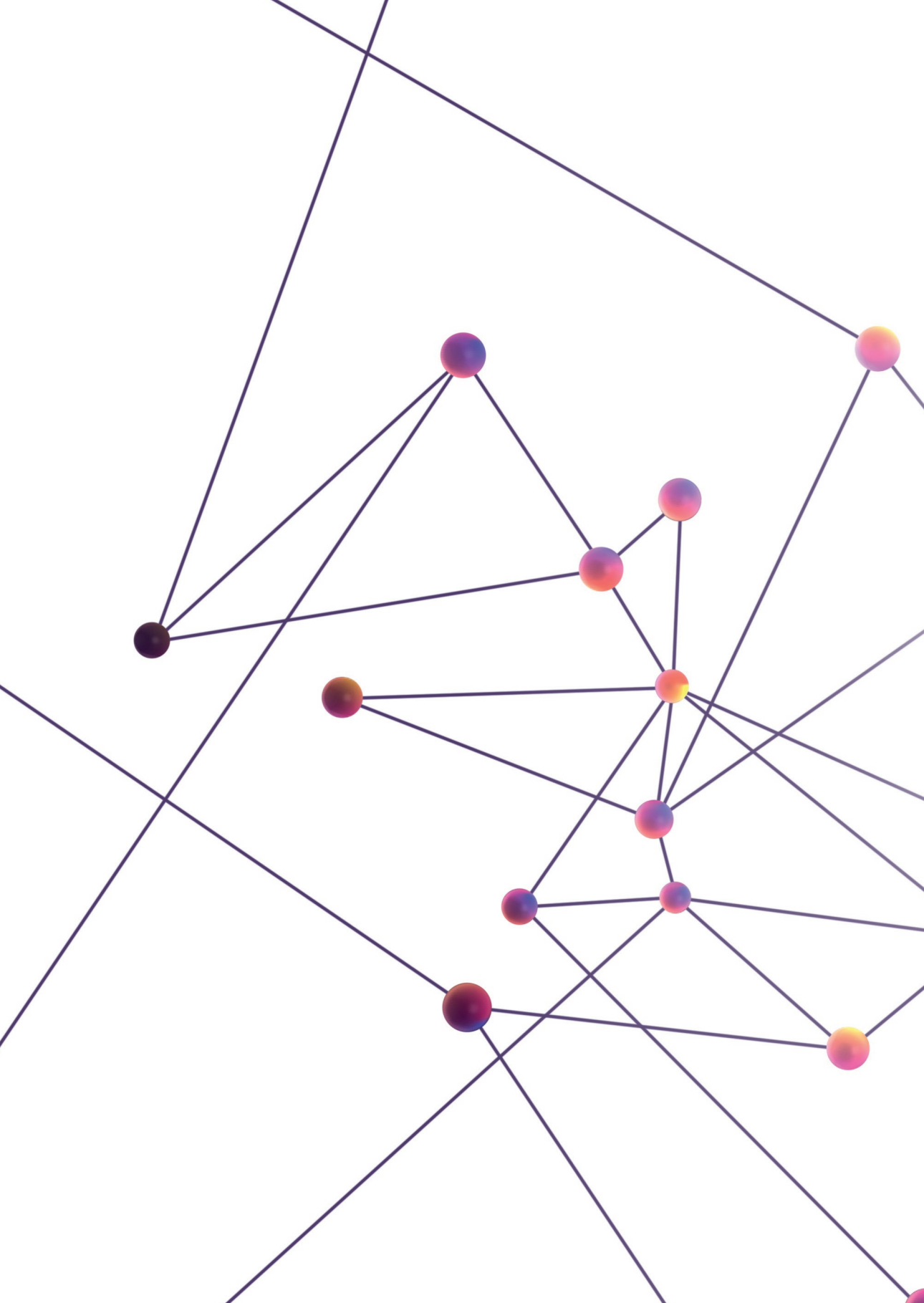
78. Stein M, Bagnoli F, Halenbeck R, Rappuoli R, Fantl WJ, Covacci A. c-Src/Lyn kinases activate Helicobacter pylori CagA through tyrosine phosphorylation of the EPIYA motifs. *Mol Microbiol*. 2002;43(4):971-980. doi:10.1046/j.1365-2958.2002.02781.x

79. Bonne Køhler J, Jers C, Senissar M, Shi L, Derouiche A, Mijakovic I. Importance of protein Ser/Thr/Tyr phosphorylation for bacterial pathogenesis. *FEBS Lett*. 2020;594(15):2339-2369. doi:10.1002/1873-3468.13797

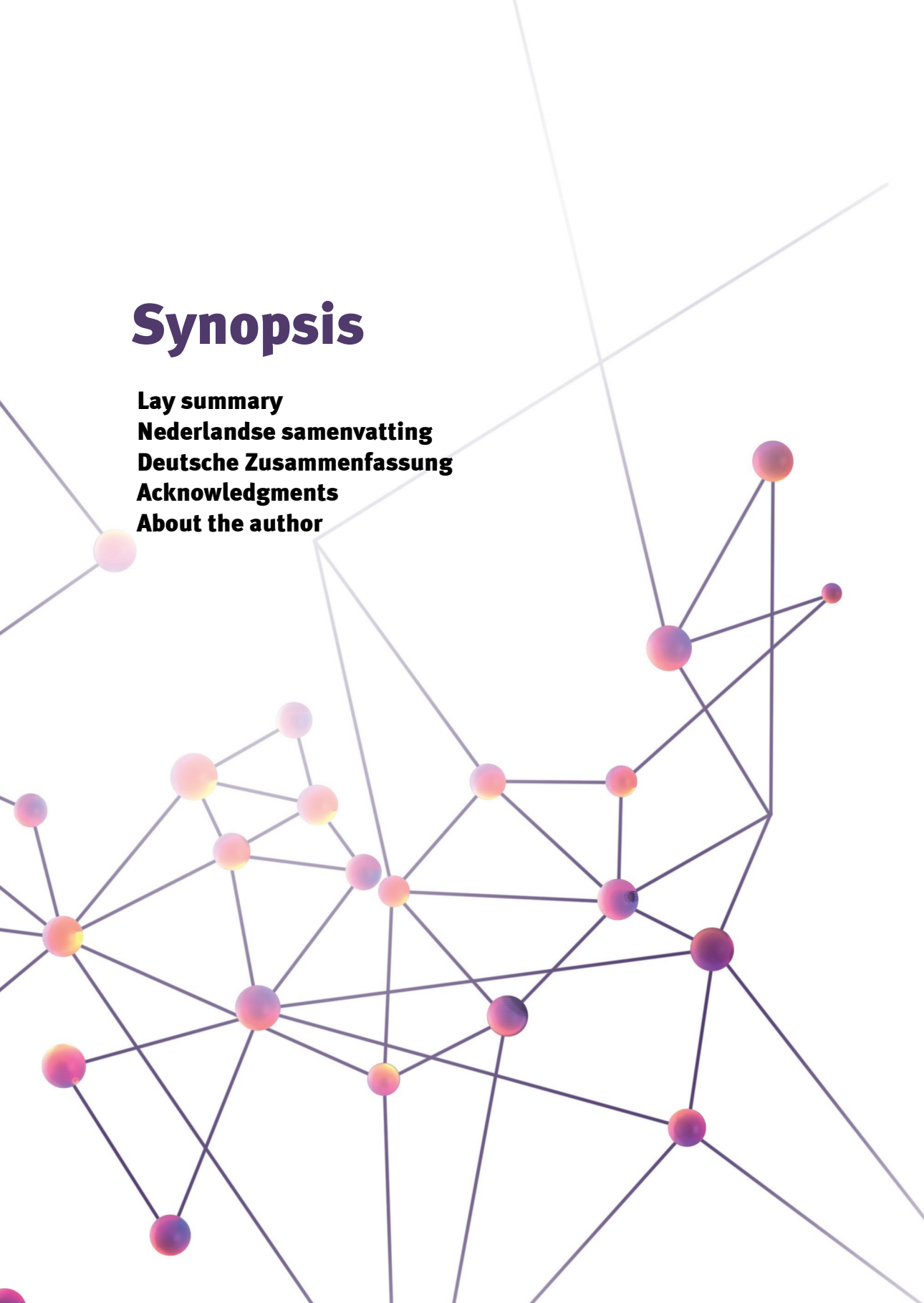
80. Mendelson M, Matsoso MP. The world health organization global action plan for antimicrobial resistance. *South African Med J*. 2015;105(5):325. doi:10.7196/SAMJ.9644

81. Clancy CJ, Buehrle DJ, Nguyen MH. PRO: The COVID-19 pandemic will result in increased antimicrobial resistance rates. *JAC-Antimicrobial Resist*. 2020;2(3):1-3. doi:10.1093/jacamr/dlaa049

82. Nelson RE, Hatfield KM, Wolford H, Samore MH, Scott RD, Reddy SC, Olubajo B, Paul P, Jernigan JA, Baggs J. National estimates of healthcare costs associated with multidrug-resistant bacterial infections among hospitalized patients in the United States. *Clin Infect Dis*. 2021;72(Suppl 1):S17-S26. doi:10.1093/cid/ciaa1581



Synopsis



Lay summary

Proteins are the workhorses of every single cell, no matter if eukaryotic or prokaryotic. They are involved in basically every biological process and thus determine the fate of every cell. The information to produce these proteins is stored in the DNA and reading out this information leads to the expression of the respective proteins. All proteins expressed at any given time or condition is called the proteome and their study proteomics. The proteome is carefully controlled, allowing the cells to adjust their expressed proteins to the cell's need and is thus highly dynamic. Identifying which proteins are expressed under which conditions can help to understand important cellular functions. For example, host cells change their protein expression to fight pathogenic bacteria. Thus, comparing host cells with infected host cells can shed light on how the cells are fighting the pathogen. Alternatively, bacteria can change their protein expression after treatment with antibiotics. Comparing treated and untreated bacteria can again help to understand cellular processes of how bacteria survive antibiotic treatment or even develop resistance. However, the abundance of a protein does not tell much about its function or activity. Proteins can be modified by small chemical groups, so-called **post-translational modifications (PTMs)**. The addition or removal of such PTMs can change features of the proteins, such as activate/deactivate a protein, localize it to a different position within the cell or change its stability. This process can occur repeatedly throughout a cell's lifespan and thus allows steady regulation of certain processes. One of those PTMs is protein **phosphorylation.** A specific enzyme, a protein **kinase**, adds a phosphate to specific amino acid residues (mainly serine, threonine, tyrosine, or histidine) of a protein. A respective **phosphatase** can reverse this reaction to remove the phosphorylation again and turn the protein to its original state. If those two enzymes are in perfect harmony, the cell can perform its normal cellular processes. However, as soon as this kinase and phosphatase system are disrupted, the entire cellular machinery can get out of balance. Dysregulation of protein phosphorylation is known to be involved in several severe diseases and understanding this tightly regulated process is of utmost importance. The study of these dynamic and important protein modifications is called **phosphoproteomics.**

The (phospho)proteome is a rich source of information. However, to access that information, several steps and considerations need to be taken. Firstly, the cells of interest need to be broken up to release the proteins. Here, one must consider the cell type or organism at hand to choose the best methods. Next, specific enzymes, proteases, cut the proteins into smaller fragments, so-called peptides. These peptides can then be measured on a **mass spectrometer (MS)** which determines its mass in relation to its charge. The peptide can also be further broken down into smaller fragments that allow then the

identification of the peptide sequence. Computational software helps to match the identified peptide back to the protein of origin.

In this thesis I optimized important steps for the sample preparation of different bacteria to gain more information about their phosphoproteome. I optimized the sample preparation for Gram-positive bacteria, identified new kinase targets in *Staphylococcus aureus*, highlighted the presence of protein arginine phosphorylation and lastly showed the need of more tailored protocols for different bacteria.

In **Chapter 1** I provided an overview about the fundamental principle of MS-based proteomics. In the context of shotgun (phospho)proteomics I focused on which considerations need to be taken into account when breaking open different cell types, how to digest proteins into peptides and how to reduce the complexity of the sample and specifically enrich for phosphorylated peptides. Next, I described how peptides can be ionized and transferred into the MS, different mass analyzers and how the peptide ions can be further fragmented using different fragmentation techniques. Furthermore, I highlighted a common database supported search algorithm to identify peptides and proteins as well as the exact position of phosphorylations. Lastly, I focused on bacteria. Bacteria are single celled organisms that are naturally present in the human body as for example in the gut system where they facilitate important processes. However, pathogenic bacteria can also cause infections. Here, I highlighted several types of phosphorylations present in bacteria and their importance as well as a brief overview about advances and challenges in their identification.

In **Chapter 2** I optimized the sample preparation for Gram-positive bacteria. Gram-positive and Gram-negative bacteria are inherently different in their cell membrane structure which requires different approaches to break open those cells. While Gram-negative bacteria are easier to break open, Gram-positive bacteria demand harsher conditions. This optimization together with removing certain compounds that are hampering the phosphopeptide identification improved the phosphopeptide identification for the Gram-positive bacterium *Staphylococcus aureus* USA300 by 21-times compared to recent literature. Lastly, using a knockout for the only known Hanks-type Ser/Thr kinase, Stk1, as well as phosphatase, Stp1, I performed a quantitative analysis. A knockout of Stk1 will result in an under-representation of phosphorylation sites introduced by Stk1 compared to a control. Respectively, a knockout of Stp1 will result in an over-representation of phosphorylation sites compared to a control. This allowed the identification of new targets of Stk1 and Stp1. At the same time, the high number of Ser and Thr phosphorylation after knockout of the only known kinase led us to the assumption that Stk1 is not the only Hanks-type kinase in *Staphylococcus aureus* USA300.

In Chapter 3 I showed that recent optimizations in phosphopeptide enrichment strategies together with the optimizations discussed in **Chapter 2** allow the identification of acid labile protein arginine phosphorylations. I reported the most comprehensive arginine phosphoproteome for *Staphylococcus aureus* USA300 to date. Since computer-based algorithms still struggle with highly confident localization of phosphorylation sites, especially for more rare types of phosphorylation, I used synthetic peptides phosphorylated on arginine to compare the acquired spectra to the acquired spectra for the *Staphylococcus aureus* USA300 samples. By doing so, I could validate the high abundance of arginine phosphorylation and correct localization by our search engine. I also tested different fragmentation methods. Those methods make use of different thermodynamically paths to fragment the peptides. Therefore, they result in different fragments that are supposedly either beneficial or disadvantageous for phosphoproteomics. Lastly, using the same Stk1 and Stp1 mutants as in **Chapter 2** I showed that the knockdown of Stp1 significantly increases the overall amount of arginine phosphorylation. To test whether Stp1 has arginine phosphatase activity, I used the synthetic as well as endogenous peptides and tried to remove the phosphorylations by using recombinant Stp1. By doing so, I could show that Stp1 does not pose arginine phosphatase activity and only indirectly influences the arginine phosphoproteome.

In **Chapter 4** I tried to show the universal application of the optimized sample preparation workflow, described in **Chapter 2**, as well as a recently published protocol. I used the Gram-negative bacteria *Pseudomonas aeruginosa* PA14 and PAO1, *Klebsiella pneumoniae* MGH as well as *Escherichia coli* DH5a. Previously, a correlation between a contamination of DNA and RNA fragments and phosphopeptide identification has been shown. Removing those contaminations increased the phosphopeptide identification. Testing different lysis conditions and taking organism specific features into account I could show that this correlation is not as strong as previously assumed and that it is highly organism if not even strain specific. Lastly, I highlighted that the used sample preparation needs to be tailored more to the organism at hand to gain a comprehensive understanding of the bacterial phosphoproteome.

Lastly, in **Chapter 5** I highlighted important advances as well as ongoing challenges in the field and provided my own perspective what this means for the field of bacterial proteomics and what we need to overcome to gain a comprehensive understanding of the bacterial phosphoproteome.

Nederlandse samenvatting

Of we het nu over eukaryoten of prokaryoten hebben – Eiwitten zijn de werkpaarden van elke cel. Ze zijn betrokken bij vrijwel elk biologisch proces en bepalen daarmee het lot van elke cel. De informatie voor het maken van deze eiwitten is te vinden in specifieke stukken DNA, de genen. Het uitlezen van een gen leidt tot de expressie van een specifiek eiwit. Alle tot expressie gebrachte eiwitten op elk gegeven moment en onder elk gegeven conditie wordt **proteoom** genoemd en de analyse ervan **proteomics**. Het proteoom is sterk gereguleerd, zodat het voor cellen mogelijk is hun eiwit expressie aan de gegeven conditie aan te passen; hierdoor is het proteoom extreem dynamisch. Om belangrijke cellulaire functies te begrijpen is het absoluut noodzakelijk om te weten, welke eiwitten onder welke condities tot expressie worden gebracht. Zo kunnen gastheercellen bijvoorbeeld hun proteoom veranderen om ziekteverwekkende bacteriën te bestrijden. Een vergelijking van besmette en niet besmette gastheercellen kan dus helpen te begrijpen, welke processen voor de bestrijding van bacteriën worden geactiveerd. Verder kunnen bacteriën tijdens een behandeling met antibiotica hun proteoom aanpassen. Een vergelijking van bacteriën met en zonder antibiotica behandeling dus helpen de onderliggende processen van antibiotica resistenties te belichten. Weten welke eiwitten in welke hoeveelheid voorkomen is vaak niet genoeg om hun activiteit of functie te begrijpen. Eiwitten (zowel eukaryoot als prokaryoot) kunnen modificaties in vorm van kleine chemische verbindingen bezitten, de zogenaamde **post-translationele modificaties (PTMs)**. Eigenschappen van eiwitten kunnen afhankelijk van de aan- of afwezigheid van PTMs veranderd worden. Dit omvat de activering of deactivering van eiwitten, de lokalisering in de cel of hun stabiliteit. Over de hele levenscyclus van een cel kunnen PTMs continu worden toegevoegd of verwijderd. Daardoor is een voortdurende regulering en aanpassing van cellulaire processen mogelijk. Een van deze PTMs is fosforylering. Specifieke enzymen, **kinasen**, kunnen fosfaat aan specifiek aminozuren (hoofdzakelijk serine, threonine, tyrosine en histidine) toevoegen. Passende **fosfatasen** kunnen deze reactie weer omkeren en het fosfaat verwijderen. Het eiwit wordt dan in zijn oorspronkelijk toestand teruggebracht. Zolang de activiteit van deze twee enzymen in evenwicht is, zal de cel haar normale processen uitoefenen. Maar zodra dit systeem uit balans raakt, wordt de cellulaire werking bemoeilijkt. Een ontregeling van fosforylering wordt in eukaryote cellen al in verband gebracht met de ontwikkeling van ernstige ziekten. Daarom, is het bestuderen van eiwit fosforylering heel belangrijk. De analyse van deze dynamische en belangrijke modificatie worden **fosfoproteomics** genoemd.

Het (fosfo)proteoom is dus een belangrijke bron van informatie over de toestand van de cel. Om toegang te krijgen tot deze informatie, moeten verschillende handelingen en procedures uitgevoerd worden. Ten eerste moeten de cellen worden opgebroken zodat de

eiwitten hieruit gezuiverd kunnen worden. Hierbij moet men op het specifieke cel-type en organisme letten, zodat de optimale methode voor de lysis kan worden gekozen. In de tweede stappen worden de eiwitten met behulp van specifieke enzymen, proteasen, in kleine fragmenten (peptiden) geknipt. Deze peptiden kunnen vervolgens in een massaspectrometer (MS) worden gemeten. Daarbij wordt de massa over lading van de peptiden gemeten. Door botsingen kunnen peptiden kleinere stukken worden gefragmenteerd waarvan de massa ook kan worden bepaald. Dit maakt het mogelijk de identiteit van het peptide te bepalen. Computer gebaseerde algoritmen maken het dan mogelijk de geïdentificeerde peptiden aan hun oorspronkelijken eiwitten toe te wijzen.

In dit proefschrift heb ik de belangrijkste werkstappen voor de analyse van het fosfoproteoom van verschillende bacteriën geoptimaliseerd. Deze optimalisatie was van cruciaal belang om toegang te krijgen tot de informatie in hun fosfoproteoom. Ik heb de monstervoorbereiding voor Gram-positieve bacteriën geoptimaliseerd en nieuwe kinase substraten in *Staphylococcus aureus* geïdentificeerd Verder heb ik laten zien dat fosforylering van het aminozuur arginine in *Staphylococcus aureus* vaker voorkomt en dat de monster voorbereiding beter aangepast moet worden op de specieke bacterie die men wil analyseren.

In het **1e hoofdstuk** heb ik de onderliggende principes van op massa spectrometrie gebaseerde eiwit analyse samengevat. In de context van shotgun (fosfo)proteomics heb ik eerst de belangrijksten stappen tijdens de monstervoorbereiding uitgelegd. Welke overwegingen moeten voor het openbreken van verschillenden bacteriën worden gemaakt? Hoe kunnen peptiden het beste geknipt worden? Hoe kan de complexiteit van de monsters worden gereduceerd? En hoe kunnen fosforpeptiden verrijkt worden? Daarnaast wordt uitgelegd hoe peptiden geïoniseerd worden zodat deze vervolgens in de massaspectrometer kunnen worden geleid. Hierbij worden de functies van de massa analysator alsook de verschillende fragmentatie methoden beschreven. Verder wordt een van de meest gebruikte databank gebaseerde zoek algoritmen beschreven. Deze maakt het mogelijk peptiden en eiwitten te identificeren en fosforyleringen te lokaliseren. Tenslotte richt ik me op bacteriën. Bacteriën zijn eencellige organismen die van nature in het menselijke lichaam voorkomen. Bacteriën vervullen bijvoorbeeld belangrijke functies in de darm. Evenwel kunnen ziekteverwekkende bacteriën de oorzaak van infecties zijn. Hier geef ik een kort overzicht van verschillende typen van fosforyleringen in bacteriën en welke functies deze uitoefenen alsmede een korte samenvatting over de huidige kennis die we hebben met betrekking tot bacteriële fosfoprotoemics

In het **2^{e} hoofdstuk** heb ik de monstervoorbereiding van Gram-positieve bacteriën geoptimaliseerd. Het celmembraan van Gram-positieve en Gram-negatieve bacteriën is fundamenteel verschillend. Daardoor zijn verschillende methoden nodig om deze bacteriën

open te breken. Terwijl Gram-positieve bacteriën makkelijker open te breken zijn, is er meer kracht nodig om Gram-negatieve bacteriën open te breken. Door deze optimalisering samen met het verwijderen van bepaalde componenten, die de verrijking van fosforpeptiden beïnvloedt, kon de identificering van fosfopeptiden verbeterd worden. Hier heb ik de identificering voor de Gram-positieve bacterie Staphylococcus aureus USA300, 21 maal verbetert in vergelijking met de actuele literatuur. Verder heb ik een kwantitatieve analyse van de KO-mutanten van de enige bekende Hanks-Ser/Thr Kinase, Stk1, en de bijbehorende fosfatase Stp1 uitgevoerd. Als Stk1 afwezig is, zullen fosforyleringen door Stk1 geïnduceerd, ondervertegenwoordigd zijn in vergelijking tot een controle monster. In overeenstemming hiermee zullen fosforyleringen verwijderd door Stp1 oververtegenwoordigd zijn, als Stp1 afwezig is. Dit maakte het mogelijk nieuwe substraten van Stk1 en Stp1 te identificeren. Tegelijkertijd laat de hoeveelheid aan fosforyleringen ondanks dat Stk1 afwezig is zien, dat Stk1 waarschijnlijk niet de enige Hanks-Ser/Thr Kinase in *Staphylococcus aureus* USA300 is.

Het **3e hoofdstuk** laat zien dat de nieuwste optimalisering van methoden voor het verrijken van fosfopeptiden samen met de in hoofdstuk twee besproken optimaliseringen voor de monstervoorbereiding de identificering van arginine fosforylering mogelijk maakt. Ik presenteer hier het op dit moment omvangrijkste arginine fosfoproteoom voor *Staphylococcus aureus* USA300. Een probleem van computer gebaseerde algoritmen is nog steeds de correcte lokalisering van fosforyleringen, vooral voor schaarse fosforyleringen zoals arginine fosforylering. Om deze computer gebaseerde algoritmen te valideren, heb ik spectra van synthetische arginine fosfopeptiden met de spectra van arginine fosforpeptiden van *Staphylococcus aureus* USA300 vergeleken. Hierdoor kon ik de hoeveelheid van arginine fosfopeptiden alsmede de correcte lokalisering door de computer gebaseerde algoritmen bevestigen. Verder heb ik verschillende fragmentie methoden getest. Deze methoden gebruiken verschillende thermodynamische eigenschapen en resulteren daardoor in verschillende fragmenten. Deze kunnen voor -en nadelen voor het identificeren van fosfopeptiden hebben. Ten slotte heb ik dezelfde Stk1 en Stp1 mutanten als in hoofdstuk twee gebruikt en kon laten zien dat de afwezigheid van Stp1 de hoeveelheid van arginine fosforylering verhoogt. Om te testen of Stp1 arginine fosfatase activiteit bezit, heb ik zowel synthetische als endogene peptiden gebruikt en geprobeerd de fosforyleringen middels Stp1 te verwijderen. Hierbij heb ik laten zien, dat Stp1 geen arginine fosfatase activiteit vertoont en dus alleen een indirect effect heeft op het arginine phosphoproteome.

In het **4e hoofdstuk** heb ik geprobeerd te laten zien, dat de in hoofdstuk twee beschreven methode voor de monstervoorbereiding en een tevoren gepubliceerd protocol universeel bruikbaar zijn. Hiervoor heb ik de Gram-negatieve bacteriën *Pseudomonas aeruginosa* PA14 en PAO1, *Klebsiella pneumoniae* MGH en *Escherichia coli* DH5a gebruikt. Hier wordt eerst een korrelatie tussen een contaminatie van DNA/RNA en het identificeren van

fosfopeptiden beschreven. Het verwijderen van deze contaminatie tijdens de monstervoorbereiding verbeterd de identificatie van fosforpeptiden. Hier heb ik verschillende cellysis condities getest en daarbij organisme specifieke eigenschapen in acht genomen. Ik heb kunnen aantonen dat deze correlatie niet zo sterk lijkt te zijn als eerder werd aangenomen. Bovendien lijken de vervuiling en de correlatie erg afhankelijk te zijn van organismen, zo niet zelfs van bacteriestammen. Uiteindelijk laat ik zien dat de monstervoorbereiding nog meer op de onderhevige bacteriestam moet worden aangepast, zodat een alomvattend beeld van het fosfoproteoom ontstaat.

Tenslotte plaats ik de belangrijkste vorderingen evenals voortdurende problemen van bacteriële fosfoproteomics in het **5e hoofdstuk** op de voorgrond. Daarnaast presenteer ik mijn eigen perspectief op wat dit betekent voor het huidig onderzoek en welke hindernissen we in de toekomst zullen moeten overwinnen om een uitgebreid begrip te krijgen van het bacteriële fosfoproteoom.

Deutsche Zusammenfassung

Unabhängig davon, ob wir über Eukaryoten oder Prokaryoten sprechen - Proteine sind die Arbeitstiere jeder Zelle. Sie sind praktisch an jedem biologischen Prozess beteiligt und bestimmen somit das Schicksal jeder Zeller. Die Informationen zur Produktion der Proteine sind auf der DNA gespeichert und das Auslesen der entsprechenden DNS-Fragmente induziert die Expression der entsprechenden Proteine. Alle exprimierten Proteine, zu jedem gegebenen Zeitpunkt und unter jeder Bedingung werden als **Proteom** bezeichnet und deren Analyse als **Proteomik**. Das Proteom ist streng reguliert, sodass es den Zellen möglich ist, die exprimierten Proteine den gegebenen Bedingungen anzupassen. Dies macht das Proteom hoch dynamisch. Um wichtige zelluläre Funktionen zu verstehen, ist es unabdingbar zu wissen, welche Proteine unter welchen Bedingungen exprimiert werden. Beispielsweise können Wirtszellen ihr Proteom anpassen, um pathogene Bakterien zu bekämpfen. Ein Vergleich von infizierten und nicht infizierten Wirtszellen kann somit Auskunft darüber geben, welche Prozesse zur Bekämpfung der Bakterien aktiviert werden. Des Weiteren können Bakterien in Folge einer Antibiotikatherapie ihr Proteom ändern. Ein Vergleich von Bakterien mit und ohne Antibiotikatherapie kann somit helfen, die zugrunde liegenden Prozesse von Antibiotikaresistenzen zu beleuchten. Das Wissen, welche Proteine in welchen Mengen wann vorliegen, ist meist jedoch nicht ausreichend, um dessen Aktivität oder Funktion zu verstehen. Proteine können Modifizierungen in Form von kleinen chemischen Verbindungen, sogenannten **post-translationale Modifizierungen (PTMs)** aufweisen. Die An- oder Abwesenheit dieser PTMs kann verschiedene Eigenschaften der Proteine verändern. Hierzu zählen unter anderem die Aktivierung oder Inaktivierung von Proteinen, die Lokalisierung innerhalb der Zelle oder die Stabilität. Über die gesamte Lebenszeit der Zelle können PTMs stetig hinzugefügt oder entfernt werden, sodass eine kontinuierliche Regulierung der zellulären Prozesse möglich ist. Eine dieser PTMs ist Proteinphosphorylierung. Spezielle Enzyme, **Proteinkinasen**, können Phosphate zu spezifischen Aminosäuren (hauptsächlich Serin, Threonin, Tyrosin und Histidin) des Proteins hinzufügen. Entsprechende **Phosphatasen** können diese Reaktion umkehren und die Phosphorylierung wieder entfernen. Das Protein wird somit wieder in seinen ursprünglichen Zustand zurückversetzt. Solange diese zwei Enzyme im Einklang sind, kann die Zelle ihre normalen Prozesse ausführen. Sobald dieses System jedoch aus der Balance gerät, wird die gesamte zelluläre Maschinerie unterbrochen. Eine Fehlregulierung von Proteinphosphorylierungen konnte bereits mit mehreren schwerwiegenden Erkrankungen in Verbindung gebracht werden. Daher ist eine strenge Regulierung von Proteinphosphorylierungen besonders wichtig. Die Analyse dieser dynamischen und wichtigen Modifizierung wird als **Phosphoproteomik** bezeichnet.

Das (Phospho)proteom ist eine wertvolle Informationsquelle. Um auf diese Information zuzugreifen müssen zunächst verschiedene Arbeitsschritte und Überlegungen getätigt werden. Als erstes müssen die Zellen aufgebrochen werden, um die Proteine zu extrahieren. Hierbei müssen der zu behandelnde Zelltyp sowie der Organismus beachtet werden, sodass die optimale Methode zur Zelllyse gewählt werden kann. Im nächsten Schritt werden die Proteine mithilfe von spezifischen Enzymen, sogenannten Proteasen, in kleinere Fragmente (Peptide), geschnitten. Diese Peptide können anschließend in einem **Massenspektrometer (MS)** vermessen werden. Hierbei wird die Masse der Peptide in Abhängigkeit ihrer Ladung gemessen. Die Peptide können zudem weiter in kleinere Fragmente fragmentiert werden, deren Masse ebenfalls ermittelt werden kann. Dies ermöglicht es, die Peptidsequenz zu identifizieren. Entsprechende Computer-basierte Algorithmen ermöglichen es anschließend, die identifizierten Peptide ihrem ursprünglichen Protein zuzuweisen.

In der hier vorliegenden Dissertation habe ich wichtige Arbeitsschritte zur Probenaufarbeitung von verschiedenen Bakterien für die anschließende Phosphoproteomik-Analyse optimiert. Dies ist wichtig, um mehr Informationen über das entsprechende Phosphoproteom zu erhalten. Ich habe die Probenaufarbeitung für Gram-positive Bakterien optimiert und neue Proteinkinase-*Targets* in *Staphylococcus aureus* identifiziert. Des Weiteren habe ich gezeigt, dass Proteinphosphorylierungen von Argininen in *Staphylococcus aureus* vermehrt vorkommen und wir die Probenaufarbeitung genauer auf die zu untersuchenden Bakterien zuschneiden müssen.

Im **1. Kapitel** habe ich die zugrunde liegenden Prinzipien der MS-basierten Proteomikanalyse zusammengefasst. Im Zusammenhang der *Shotgun* (Phospho)proteomik erläutere ich zunächst die wichtigsten Schritte während der Probenaufarbeitung. Welche Überlegungen müssen zum Aufbrechen verschiedener Bakterien getroffen werden? Wie können Proteine in Peptide verdaut werden? Wie kann die Komplexität der Proben reduziert werden? Und wie können Phosphopeptide am besten angereichert werden? Als nächstes beschreibe ich, wie Peptide ionisiert werden können, um diese anschließend ins MS zu leiten. Es werden die Funktion des Massenanalysators und verschiedene Fragmentierungsmethoden beschrieben. Außerdem wird einer der gängigsten Datenbank-basierten Suchalgorithmen, um Peptide und Proteine zu identifizieren sowie Phosphorylierungen zu lokalisieren, beschrieben. Zum Schluss richte ich den Fokus auf Bakterien. Bakterien sind einzellige Organismen, die natürlicherweise im menschlichen Körper vorhanden sind. Bakterien erfüllen zum Beispiel wichtige Funktionen im Darm. Pathogene Bakterien können aber auch die Ursache von verschiedenen Infektionen sein. Hier gebe ich eine kurze Übersicht der verschiedenen Arten von Phosphorylierungen in Bakterien und deren Aufgabe sowie eine kurze Zusammenfassung der aktuellen Fortschritte und Probleme im Bereich der bakteriellen Phosphoproteomik.

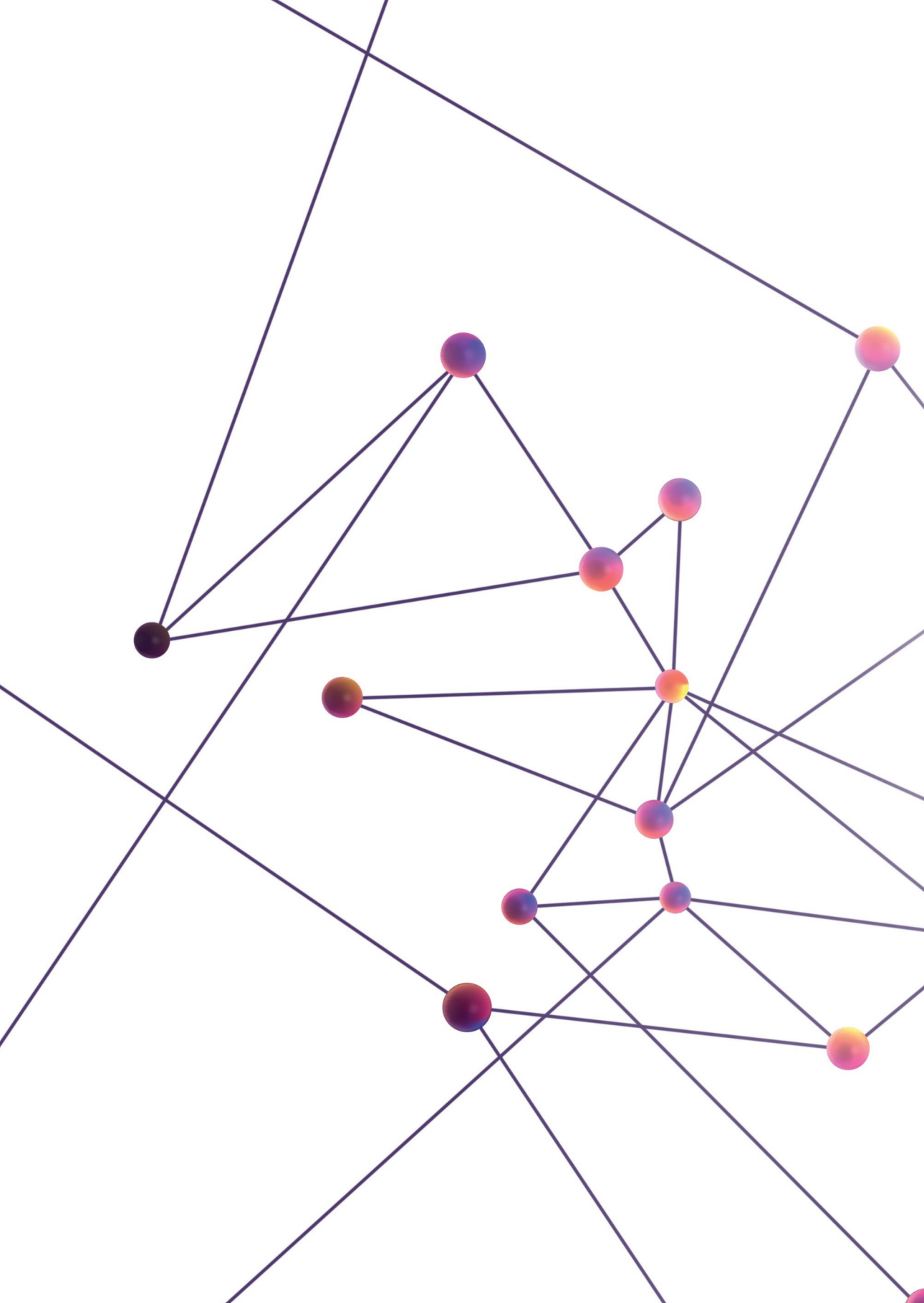
Im **2. Kapitel** habe ich die Probenaufarbeitung für Gram-positive Bakterien optimiert. Gram-positive und Gram-negative Bakterien unterscheiden sich grundlegend im Aufbau ihrer Zellmembran. Daher benötigen Gram-negative und Gram-positive Bakterien unterschiedliche Methoden, um ihre Zellen aufzubrechen. Gram-negative Bakterien sind einfacher aufzubrechen als Gram-positive Bakterien. Letztere benötigen daher härtere Bedingungen für die Zelllyse. Diese Optimierung zusammen mit dem Entfernen bestimmter Komponenten, die die Anreicherung von Phosphopeptiden behindern, hat die Identifizierung der Phosphopeptide verbessert. Hier habe ich die Anzahl an Identifizierungen von Phosphopeptiden für das Gram-positive Bakterium *Staphylococcus aureus* USA300 im Vergleich zur aktuellen Literatur um das 21-fache verbessert. Des Weiteren habe ich eine quantitative Analyse mit den KO-Mutanten der einzig bekannten Hanks-Ser/Thr Kinase, Stk1, sowie der entsprechen Phosphatase, Stp1, durchgeführt. Die Abwesenheit von Stk1 resultiert, im Vergleich zur Kontrolle, in eine Unterrepräsentation von Phosphorylierungen, die durch Stk1 hinzugefügt werden. Entsprechend führt die Abwesenheit von Stp1 zu einer Überrepräsentation im Vergleich zur Kontrolle. Dies ermöglichte es, neue *Targets* von Stk1 und Stp1 zu identifizieren. Gleichzeitig legt die hohe Anzahl an Phosphorylierungen, trotz Abwesenheit von Stk1, nahe, dass Stk1 nicht die einzige Hanks-Ser/Thr Kinase in *Staphylococcus aureus* USA300 ist.

Kapitel 3 zeigt, dass die neuesten Optimierungen von Anreicherungsmethoden für Phosphopeptiden zusammen mit den in **Kapitel 2** besprochenen Probenaufarbeitungsoptimierungen die Identifizierung von Säure-labilen Arginin-Phosphorylierungen ermöglichen. Ich beschreibe hier das zurzeit umfassendste Arginin-Phosphoproteome für *Staphylococcus aureus* USA300. Ein Problem Computer-basierter Algorithmen ist noch stets die korrekte Lokalisierung von Phosphorylierungen, besonders für seltener vorkommende Phosphorylierungen sowie der Phosphorylierung von Arginin. Um die Computer-basierte Lokalisierung zu validieren, habe ich die Spektren von synthetischen Arginin-Phosphopeptiden mit den Spektren von Arginin-Phosphopeptiden von *Staphylococcus aureus* USA300 verglichen. Hierdurch konnte ich die hohe Abundanz von Arginin-Phosphorylierungen sowie die korrekte Lokalisierung durch den Computer-basierten Algorithmus bestätigen. Des Weiteren habe ich verschiedene Fragmentierungsmethoden getestet. Diese Methoden nutzen verschiedene thermodynamische Charakteristika und resultieren daher in unterschiedlichen Fragmenten, die ihre Vor- und Nachteile für die Identifizierung von Phosphopeptiden haben. Letztlich habe ich dieselben Stk1 und Stp1 Mutanten wie im **2. Kapitel** verwendet und konnte zeigen, dass die Abwesenheit von Stp1 den Anteil an Arginin-Phosphorylierungen erhöht. Um zu testen, ob Stp1 Arginin-Phosphataseaktivität aufweist, habe ich sowohl die synthetischen als auch die endogenen Peptide verwendet und versucht, die entsprechenden Phosphorylierungen mittels rekombinanter Stp1 zu entfernen. Hierbei konnte ich zeigen,

dass Stp1 keine Arginin-Phosphataseaktivität besitzt und somit nur einen sekundären Effekt auf das Arginin-Phosphoproteom ausübt.

Im **4. Kapitel** habe ich versucht zu zeigen, dass die im **2. Kapitel** beschriebene Probenaufarbeitungsmethode sowie ein zuvor publiziertes Protokoll universell einsetzbar sind. Hierzu habe ich die Gram-negativen Bakterien *Pseudomonas aeruginosa* PA14 und PAO1, *Klebsiella pneumoniae* MGH sowie *Escherichia coli* DH5a verwendet. Zuvor wurde eine Korrelation zwischen einer Kontamination von DNS- und/oder RNS-Fragmenten und der Identifizierung von Phosphopeptiden beschrieben. Das Entfernen dieser Kontamination während der Probenaufarbeitung führte zu einer Verbesserung der Identifizierung von Phosphopeptiden. Hier habe ich verschiedene Lysebedingungen getestet und dabei Organismen-spezifische Eigenschaften berücksichtigt. Dabei konnte ich zeigen, dass diese Korrelation nicht so stark zu sein scheint, wie zuvor angenommen. Zudem scheint die Kontamination sowie die Korrelation sehr vom Organsimus, wenn nicht sogar Bakterienstamm, abhängig zu sein. Letztendlich zeige ich, dass die Probenaufarbeitung noch mehr auf den vorliegenden Bakterienstamm zugeschnitten werden muss, damit ein umfassender Blick auf das Phosphoproteom geworfen werden kann.

Zum Schluss hebe ich die wichtigsten Fortschritte sowie die andauernde Problematik im Bereich der bakteriellen Phosphoproteomik im **5. Kapitel** hervor. Zudem lege ich meine eigene Perspektive dar, was dies für die aktuelle Forschung bedeutet und welche Hürden wir in Zukunft überwinden müssen, um ein umfassendes Verständnis der bakteriellen Phosphoproteome zu erhalten.



Acknowledgments

Writing this chapter of my thesis means that I actually made it till the end and yes sometimes I was not that sure about this. The last years have been an incredible journey during which I not only learned how to be a better scientist, but also a lot about myself especially when facing the ups and downs only too familiar to every PhD student. I probably would not have reached the end of this journey without the endless support of several people and I am not sure if I can put in words how grateful I am for each and every one of you, but I will give it a try.

First of all, I have to thank **Simone** on so many levels. Your constant optimism drove me crazy from time to time but at the same time this optimism guided me through the last almost five years. You taught me to accept if things are not going as planned and still making the best out of it. I am grateful for every confidence boost, when I again doubted myself. No matter what bothered me, I could count on your optimism to lift my spirit. I am extremely thankful that you still took your time after you changed your position to finish this journey we started together.

Albert, even though we did not work together a lot, I knew that I could count on your honest opinion when I had questions about any of my projects. During one of our first meetings, you told me that I need to be more confident and believe in myself and the data I am presenting. Those things stick with me during the last years and even though I am still struggling sometimes, you made me believe more in my capabilities.

Katrin, since my bachelor studies your enthusiasm and power guided me through my scientific journey. Without you I would not even have applied for a PhD position and especially not in well-known lab as the Heck lab. Even today, I know that I can always reach out for advice. Thanks for your support over the last years.

Donna, we started our journey together and without you it definitely would have been less fun. Over the last years we shared laughter, sweat and tears together and you were my first go to person no matter what happened. You pulled me back up so many times, so thanks for all your support and motivation. You are truly an amazing person and your cheerful character brightens up the room. Do not let anyone steal your thunder. You got this girl! Thanks for finishing what we started together.

Marie, the moment you joined the lab I knew that we would be great friends. You are crazy and loud as I am and I could always rely on you, no matter what. Thanks for killing it during our rocycle sessions where we both got pushed to our limits to reach higher not only during spinning. I know you will find your way and obviously, I hope this will not be too far away. Thanks for being by my side during my defense.

Obviously, I cannot mention everyone personally, but I have to thank the whole Heck Lab former and present members. All of you made working, borrels, lunch, labutjes and

Sinterkerst fun and exciting. Even though I will highlight a few people, be assured that I am grateful for every one of you.

First, I must thank the Lemeer lab. **Celine**, I really enjoyed working with you or sharing a beer during our Friday borrels. Obviously, I also will never forget our legendary trip to the Zugspitze. **Miao** and **Clement**, thanks for all your wisdom. You taught me everything about your phospho protocol and all the tips and tricks to deal with the IMAC system. Clement, your unlimited MS knowledge was also highly appreciated. Your typical French love for food, wine, and beer, brought us to many restaurants and bars. **Niels**, we started around the same time and I really enjoyed working with you. We shared endless desperate moments in front of our princess or Horst and were sometimes close to just throw them out of the window. But nevertheless, we both made it! I really enjoyed our trip to Berlin and Potsdam. I know that your new job suits you well and I wish you all the best. **Sander**, you showed me around the mass specs and HPLC when I joined and I could always come along for a question. I am happy to see that you found your place.

Everyone should have a safe space where you can share frustrations and find support. I am happy that I had such a place in my office. So thanks to my past (**Linsey**) and present office mates **Juan M., Franziska** and **Laura.** You were indeed the perfect office mates. We always knew when we could distract each other for a ten minute chatter break to share our frustrations, confusions, or happy moments. Thanks for your support, fun moments, intense discussions (feminist rants) and borrels. Franziska, I do not know how many questions I asked you regarding R, but it was definitely too many to leave it unmentioned, that you helped me enormously during my PhD.

Esther, you always told me that I will be ok and what shall I say, I am ok. Thanks for your constant support. I enjoyed going out which quite often ended at the Bodytalk or all Eurovision parties. Yes, I know, 'Germany: zero points'. Dear **Kyle**, thank you so much for all your great stories and wisdom. There was never a dull borrel or lunch and for sure not a dull night out with you. **Domenico**, thanks for all the fun moments and all your wisdom you shared with me. **Julia**, you are a remarkable scientist and person! Thanks for all the fun moments and serious discussions. You will do great!

Thanks to all my fellow rainbow women in science**, Kelly D, Kelly S, Saar, Charlotte**, Franziska, Donna and Marie. One of us always knew how to cheer up our week, was available for coffee breaks, or a borrel at the end of the week. **Kelly S**, thanks for introducing me to OR13 together with Celine and then taking care of our drama queen OR15 together with **Max** and **Sem**. Thanks to the borrel people, especially **Gadi**, **Danique** and **Wouter**, for always knowing how to start the Friday evening right.

Henk, thanks for your help fixing my scripts and advising how to write better codes. **Wei**, thanks for all your mental support and wisdom. **Maarten**, thanks for adopting me for the last part of my journey, I really appreciated it.

Also, thanks to my fellow Germans, **Barbara** and **Johannes**, always nice to have someone around from home. Even though Johannes, till this day, I do not know why we constantly spoke English instead of German.

A huge thanks also to all the support staff, that was always available and kept the lab up and running. **Arjan**, thanks for all your help fixing a mass spec, but even more thanks for your mental support and cheering up every lunch, borrel or karaoke night. **Mirjam,** thank you for all the work you are doing for keeping the lab up and running, constant coffee supply and always being there when I needed to vent. **Harm**, **Ceri**, **Pieter**, **Dominique** and **Soenita**, thanks for all your support in cell culture and the MS lab. **Corine**, thanks for all your help with administrative work or just having a chat. **Geert**, thanks for keeping our computer and server up and running.

I also would like to thank my fellow PhD council members, especially, **Suze, Stephanie, Mui, Saskia, Kim and Anke**. It was a great pleasure working with you and I learned a lot. I hope the current council keeps up the good work.

Part of my work would not have been possible without successful collaborations. **Nina**, I would like to thank you for the great collaboration and providing your lab for my bacteria work. **Julius**, thanks for your help and providing the *P. aeruginosa* samples.

Luckily, there is also a life outside the lab and I was happy enough to have friends and family that were always around to unwind and lift my spirit. To address those people, I will continue in German.

Pia und **Janina,** ich bin froh, dass wir uns quasi seit dem ersten Tag unseres Studiums gefunden haben. Seitdem haben wir so einiges zusammen erlebet. Ich bin dankbar für eure Unterstützung und jegliche Motivation über die letzten Jahre. Pia, irgendwann laufen wir den Rest des West Highland Ways! **Prabal**, **Andy** und **Britta**, eure Unterstützung und Ermutigungen haben einen großen Anteil an dem, wo ich heute bin. Andy, du hast Simone auf der SummerSchool angesprochen und quasi die Grundlage für meine Bewerbung gelegt. Britta, unser gemeinsamer Aufenthalt in Brixen, war definitiv legendär. Auch wenn wir alle mittlerweile in Deutschland und den Niederlanden verstreut sind, bin ich dankbar, dass wir immer einfach da weiter machen können, wo wir aufgehört haben.

An meine Rudeldiere zuhause, **Markus, Nina, Maria und Jenny.** Danke für jedes aufbauende Wort, jede Diskussion über‘s Leben, jede Party und jedes Konzert. Ich glaube 2022 haben wir so einiges nachzuholen. **Sabrina, Dominik, Katrin, Kristin, Torsten und Kathrin**, obwohl

Utrecht nicht so weit weg ist, sehen wir uns deutlich weniger. Ich bin dankbar, dass wir es dennoch schaffen, einfach da weiterzumachen, wo wir das letzte Mal aufgehört haben.

Omimi, ich bin dir so dankbar für alles, was du für mich getan hast und noch stets tust und ich freu mich riesig, dass ich diesen Moment mit dir teilen kann.

Oppa, wie hast du dem Tag entgegengefiebert, an dem dein „holländisches Meisje" ihre Doktorarbeit verteidigt. Es bricht mir das Herz, dass du diesen Tag nicht mehr erleben darfst. Ich weiß aber, dass du da oben vor Stolz platzt und dass du immer an meiner Seit bist. Ich bin dir so unendlich dankbar für alles, was du für mich getan hast.

Jenny, du bist mein größter Fan und ich kann mich blind auf deine Unterstützung verlassen. Ich bin so froh und stolz, dich als meine große Schwester zu haben. Egal was ist, du bist sofort da und wenn es sein musst, verpasst du mir einen Tritt in den Hintern oder sagst mir, wie dumm es ist, dass ich mal wieder an mir selbst zweifle. Ich sag es viel zu selten, aber ich bin dir so dankbar!

Liebe **Mama,** lieber **Papa**, ohne euch wäre dies hier definitiv nicht möglich gewesen. Ihr habt mich immer in allem unterstützt und vor allem immer mehr an mich geglaubt, als ich selbst. Ich sag es viel zu selten und kann nicht in Worte fassen, wie dankbar ich euch bin. Mama, du bist zusammen mit **Jörg** selbst wörtlich bis ans andere Ender der Welt für mich geflogen. Ich weiß, dass ihr nicht die einzigen wart, die dies für mich getan hätten. Deswegen bin ich unendlich dankbar, zu wissen, dass ich mich immer auf meine Familie verlassen kann. Daher ein großer Dank an die ganze Sippe**: Oma, Opa, Norbert, Sabine, Marina, Astrid, Christina, Yannika, Florian, Hannah, Maik, Anke, Thorsten, Jörg, Steffi, Anja, Patrick, Jan und Lisa**.

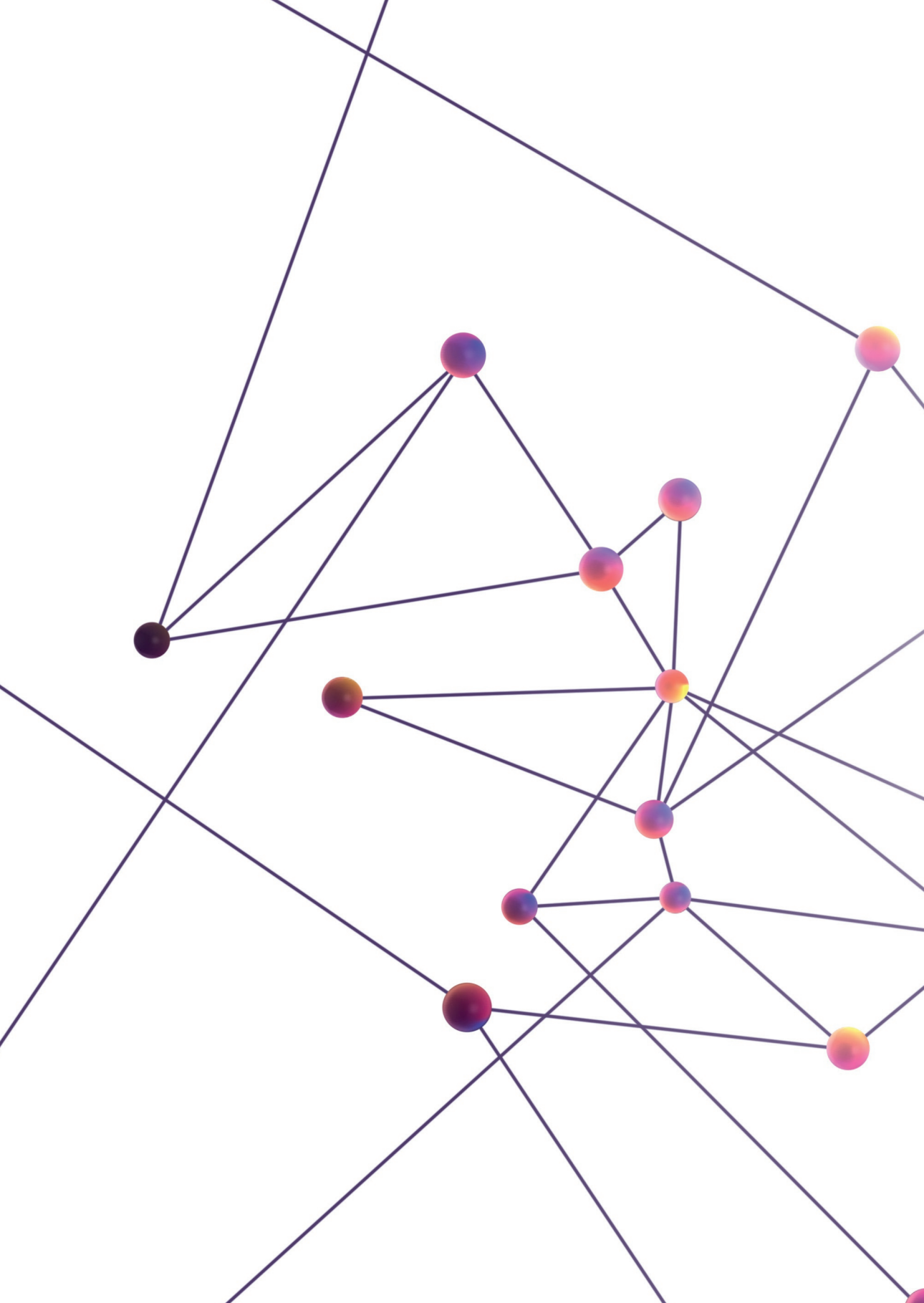
En tot slot lieve **Timo**, tijdens de afgelopen twee jaar heb jij mij enorm gesteund. We hebben elkaar ontmoet toen ik nog worstelde met de nasleep van Australië en ik druk bezig was met het afsluiten van mijn PhD. Jij moest mij vaak opbouwen als er weer iets niet lukte of alles een beetje te veel werd, maar jij wist mij altijd weer aan het lachen te maken. Ik ben zo blij dat jij er bent. Ik hou van jou.

Cheers,

Nadine

"'Sometimes I think you believe in me more than I do.' said the boy.

***'You'll catch up' said the horse."* – Charlie Mackesy**



About the author

List of publications

Prust N, van der Laarse SAM, van den Toorn H, van Sorge NM, Lemeer S. In depth characterization of the Staphylococcus aureus phosphoproteome reveals new targets of Stk1.
Mol Cell Proteomics. 2020;20:100034. doi:10.1074/mcp.RA120.002232

Potel CM, Lin MH, **Prust N**, van den Toorn HWP, Heck AJR, Lemeer S. (2019) Gaining confidence in the Elusive Histidine Phosphoproteome.
Anal Chem 7;91(9):5542-5547

Mulder C, **Prust N**, van Doorn S, Reinecke M, Kuster B, van Bergen En Henegouwen, Lemeer S. (2018) Adaptive Resistance to EGFR-Targeted Therapy by Calcium Signaling in NSCLC Cells.
Mol Cancer Res 16(11):1773-1784

Curriculum vitae

On November 5th 1990 I was born in Duisburg, Germany. After graduating from St. Hildegardis Gymnasium in Duisburg in 2010, I studied Biochemistry at the Ruhr-Universität Bochum, Germany. An Erasmus exchange in Glasgow, Scotland, allowed me to gain first experiences abroad during my Bachelor studies.

Already during my Bachelor studies I was fascinated by proteins and the possibilities of mass spectrometry. For both my Bachelor and Master thesis I had the privilege of working at the Medizinisches Proteom Center, Bochum, under supervision of Prof. Dr. Katrin Marcus. From the beginning I was intrigued by protein phosphorylation and had the chance to learn different techniques for their analysis in different contexts.

In January 2017 I started my PhD track in Prof. Dr. A.J.R. Heck's Biomolecular Mass Spectrometry & Proteomics group at Utrecht University. Under the supervision of Dr. Simone Lemeer I had the opportunity to extant my knowledge of phosphopoteomics and dived into the interesting field of bacterial phosphoproteomics, of which work is summarized in this thesis.