universität wien

DISSERTATION

Titel der Dissertation

„Enhanced Analysis of Peptides with Multiple, Phosphorylated Amino Acids and Nitrogen-bound Protein Phosphorylations by Titanium Dioxide Enrichment and Mass Spectrometry“

>1/1<

Andreas Schmidt

Diplomchemiker

angestrebter akademischer Grad

Doktor der Naturwissenschaften (Dr. rer.nat.)

Wien, 2010

Studienkennzahl lt. Studienblatt:	A 091 419
Dissertationsgebiet lt. Studienblatt:	Naturwissenschaften Chemie
Betreuerin / Betreuer:	Prof. Dr. Gustav Ammerer

Declaration

I, Andreas Schmidt (born 4th April 1981), declare to be the only author of the thesis "Enhanced Analysis of Peptides with Multiple Phosphorylated Amino Acids and Protein Phosphorylations of Histidine and Arginine by Titanium Dioxide Enrichment and Mass Spectrometry". All sources of scientific material, which I used during preparation of the thesis are cited in the text and listed in the corresponding references section. In case of adapted figures, I tried to identify the author, to get his permission to use the figure. All other figures were prepared by myself. A part of the thesis describes experimental data of a cooperation with Jakob Fuhrmann in the group of Tim Clausen, which where adapted from the publication in Science (Fuhrmann J. *et al.*, 2009). Further, I am permitted to present data from Christoph Stingl, which are also part of this section. These data are located at the beginning of part 2 and are especially signed to be contributed the mentioned colleagues.

Vienna, 07th June 2010, Andreas Schmidt

Acknowledgement

First of all I would like to thank my family for a lot of help and support to start my PhD in a foreign country. Further, I want to thank Nina for cheering me up in critical times and being the centre of my private life.

I would like to thank the members of the Christian-Doppler-Laboratory for Proteome Analysis, the Protein Chemistry Facility at the IMP of Vienna, the University Mass Spectrometry Facility of the Department of Life Sciences and Biochemistry, and Gustav Ammerer's group at the University of Vienna for helping a lot with sample preparation and analysis, many fruitful discussions on improvements and a especially the pleasant and encouraging working atmosphere. Especially, I would like to thank Edina Csazsar, Christoph Stingl, and Ines Steinmacher for introduction to MS techniques and instruments, Michi Mazanek for sharing his knowledge on TiO_2, a technique which helped a lot to solve the problems, I encountered. I would like to gratitude Goran Mitulovic for his patience in all discussions and explanations on all kind of HPLC ideas or problems. Especially, I would like to thank Mathias Madalinski for synthesis of any kind of peptide sequence and lots of good discussions.

I would like to acknowledge, Prof. Dr. Andrea Sinz, Prof. Dr. Kristina Djinovic-Carugo, Prof. Dr. Emanuelle Charpentier, Dr. Lothar Brecker, Dr. Tim Clausen, Dr. Thomas Marlovits, Dr. Christian Ihling, Dr. Masato Yonezawa, Dr. Jakob Fuhrmann, Dr. Jürgen Hartler, Dr. Mario Richter, Oliver Schraidt, Matthias Brunner, Mathias Müller, and Silvia Spiess for initiating and contributing to scientific cooperation.

I want to thank Johann Holzmann, Karin Großteßner-Hain, James Hutchins, and Ilse Dohnal for proof-reading and many suggestions to improve the thesis text.

Finally, I would like to thank my supervisors Prof. Dr. Gustav Ammerer and Karl Mechtler for constant scientific support and many helpful discussions, but also for allowing enough freedom to develop my own ideas and test my hypothesis with the provided scientific material.

I want to thank the Christian-Doppler Society and the IMP for financial support and allowance to visit the IMSC 2006 and the ASMS 2008 conferences to present my scientific work.

Abstract

One key principle of protein regulation is post-translational modification such as methylation or phosphorylation, which can alter activity, substrate interaction and subcellular localization of proteins. Protein phosphorylation is one of the key regulatory events driving cell growth and division, metabolism, and motility. Also many signal transduction processes require the activity of protein kinases and phosphatases. These enzymes catalyse the transfer of phosphate moieties from adenosine triphosphate as to specific amino acid side chains to form phosphoric acid esters, anhydrides, or phospho-amidates and thereby creating a negative charge at the amino acid side chain. Addition of the phospho-moiety requires physical interaction of kinase and target protein, which makes this modification an important marker for protein-protein interactions. Technical improvements of the sample preparation such as specific enrichment techniques as well as novel detection methods for mass spectrometry provide easier access to protein phosphorylation data and even allow quantitative analysis.

Physico-chemical properties of the different phospho-modifications on amino acids require different experimental strategies for detection by mass spectrometry. Therefore two strategies for phosphopeptide analysis are presented in individual sections of the thesis. The first part deals with the problem of peptides with multiple phosphorylations, which are quite often underrepresented in phospho-proteomic datasets. Furthermore, such peptides have often underlying combinatorial complexity, which is caused by the distribution of the phospho-moieties onto different putative acceptors. In the second section I report the first identification of a protein-arginine kinase as a fundamental signal transducer in bacteria. Further, I established a protocol which allows detection of this arginine modification by proteomic methods.

With regard to the first part, one of the problems for reliable detection of peptides with multiple phosphorylation sites by mass spectrometry is their low stoichiometry in comparison to unphosphorylated and singly phosphorylated peptides. Chemical properties which give rise to multiple charged isoforms, due to similar pKa values of the phospho-moieties, require specific separation conditions in chromatography and optimized mass spectrometric detection. Our protocol combines the previously reported titanium dioxide enrichment procedure with separation of phospho-peptides using a monolithic column followed by matrix-assisted laser desorption/ionization mass spectrometry. Using this method, we are able to localize the phosphorylated amino acid in peptide sequences with 3 or more phospho-moieties reliably.

The second part of the thesis is focused on phosphorylation of histidines and arginines in peptides and proteins. Detection of this modification has been challenging because of its low stability under acidic conditions. Nevertheless, recently developed methods in mass spectrometry allow unambiguous identification of such phosphorylated amino acids. We developed a protocol which preserves the modification during sample preparation, thereby allowing us to detect arginine phosphorylation on a bacterial translation factor CtsR after *in vitro* phosphorylation with McsB. Using synthethic peptides we demonstrate the specificity of the arginine kinase McsB that is expressed in some Gram-positive bacteria.

Further, we optimized our protocol to purify arginine and histidine phosphorylated peptides from *in vitro* phosphorylation samples. Using these purified model peptides we monitored the hydrolysis at different temperatures and pH values, allowing us to estimate the limits of acidic treatment. To improve mass spectrometric detection, we applied different fragmentation conditions in collision-induced and electron-induced dissociation methods. Under collision-induced dissociation processes, the fragmentation behaviour of phospho-arginine containing peptides differs significantly from that of serine or threonine phosphorylated peptides, although all modifications show extensive elimination of phosphoric acid.

Based on the stability experiments, an enrichment procedure was established which allows purification of peptides with nitrogen-bound phosphorylations. Furthermore, we studied the influence of analysis method and search parameters on identification of the correct phospho-amino acid. This protocol confirmed the arginine phosphorylation of the bacterial protein CtsR and revealed a high similarity between the phosphorylation patterns of this protein from B. subtilis and B. stearothermophilus after *in vitro* phosphorylation. The developed protocol will also help to study these modifications in cell lysates of other bacteria and eukaryotic cells.

Zusammenfassung

In Zellen gibt es unzählige, chemische Prozesse, wie z.B. Zellteilung oder Wachstum, welche ausschließlich von Proteinen gesteuert werden und strikter Regulierung bedürfen. Einen sehr wichtigen Beitrag zur Regulierung von Proteinaktivitäten liefern post-translationale Modifikationen, welche in Form von reversiblen, an Aminosäureseitenketten gekoppelte Molekülen, wie bei der Acetylierung oder Phosphorylierung, oder als irreversible Änderungen auftreten können, z.B. bei der Abspaltung eines Signalpeptids. Die Protein-Phosphorylierung stellt zum jetzigen Zeitpunkt die am besten untersuchte post-translationale Modifikation dar und spielt eine sehr wichtige Rolle für den Ablauf der Zellteilung sowie in fast allen Signaltransduktionswegen innerhalb der Zellen. Die Phosphorylierung wird von der Enzymfamilie der Proteinkinasen übertragen, welche einen Phospho-Ester an den Seitenketten der Aminosäuren Serin, Threonin, Tyrosin, ein Phospho-Anhydrid an Aspartat oder ein Phospho-Amidat an Stickstoff-Atomen von Histidin, Lysin oder Arginin synthetisieren. Dabei wird an vorher positiv geladenen oder ungeladenen Seitenketten eine negative Ladung angebracht, welche die Interaktion zu benachbarten Aminosäuren beeinflusst. Die Wichtigkeit dieser Modifikation für zelluläre Prozesse und die schwerwiegenden Effekte, die durch eine Missregulation entstehen, machen ein umfangreiches Verständnis dieser Modifikation erforderlich.

Durch Entwicklung neuer Methoden, die spezifisch auf die Anreicherung und Detektion post-translationaler Modifikationen ausgelegt sind, ist die Massenspektrometrie zu einer Standardmethode zum Nachweis von Proteinphosphorylierungen geworden. Aufgrund der Eigenschaften der verschiedenen phosphorylierten Aminosäuren müssen unterschiedliche experimentelle Bedingungen für die Bestimmung angewendet werden. Die Dissertation behandelt die Analyse von mehrfach phosphorylierten Peptiden und die Detektion von Stickstoff-gebundenen Phosphorylierungen in Peptiden in zwei Teilen. Im ersten Teil wird ein Protokoll vorgestellt mit dem Phosphorylierungsstellen in Peptiden besser identifiziert werden, wenn mehrere dieser Modifikationen in der Peptidsequenz vorhanden sind. Die Analyse von Stickstoff-gebundenen Phosphorylierungen ist im zweiten Teil detailliert beschrieben.

Peptide mit drei oder mehr Phosphorylierungen sind in den meisten Studien oft unterrepräsentiert, da sie einerseits nur zu einem geringen Prozentsatz vorhanden sind und andererseits aufgrund ihrer physikalisch-chemischen Eigenschaften oft nicht detektiert werden. Zur Verbesserung der Analyse solcher vielfach phosphorylierter Peptide haben wir

ein Protokoll entwickelt, welches Metalloxid-Affinitätschromatographie an Titandioxid mit der Sensitivität von Matrix-unterstützter Laser Desorption/Ionisations Massenspektrometrie verbindet. Dabei werden Phosphopeptide im ersten Schritt nach der enzymatischen Spaltung mittels Titanoxid angereichert. Aufgrund des ungenügenden Trennungsverhaltens vielfach phosphorylierter Peptide auf herkömmlichen "reversed phase" Chromatographiesäulen wurde die Elutionsmischung auf einer monolithischen Säule getrennt. Detektion und Identifikation erfolgt mittels Matrix-unterstützte Laser Desorption/Ionisations-Massenspektrometrie, wobei aufgrund des besonderen Fragmentationsverhaltens der einfach geladenen Peptidionen oft eine hohe Sequenzabdeckung und genaue Lokalisation der Phosphorylierungsstelle erreicht werden kann. Als Modellsystem haben wir die mehrfach phosphorylierten Peptide von Osteopontin und ß-Casein analysiert und das neue Protokoll mit den etablierten Methoden "multi-stage activation" kollisionsaktivierter Dissoziation und Elektronentransferdissoziation verglichen.

Bereits in den frühen 1970er Jahren wurden Phosphorylierungen an Seitenketten von Histidin, Lysin oder Arginin beschrieben und mittels Radioaktivitätsassays nachgewiesen, allerdings sind nur sehr wenige Phosphorylierungstellen bekannt. Im Allgemeinen werden N-Phosphorylierungen mit der Domäne der Prokaryonten verbunden und gelten als evolutionär früheste Form der Protein-Phosphorylierung. Über die Funktion von Stickstoff-gebundenen Phosphorylierungen (N-Phosphorylierungen) auf Proteinen ist bisher nur wenig bekannt, da diese Modifikation durch die Probenvorbereitung oft verloren geht. Werden Bakterien Stress-Situationen ausgesetzt, wie z.B. erhöhter Temperatur oder hohen Salzkonzentrationen, müssen sie sich schnell an die Bedingungen anpassen und eventuelle Schäden, wie Proteinaggregation oder Denaturierung, beheben.

Aufgrund der Stickstoff-Phosphor-Bindung, welche signifikant reaktiver ist als die Sauerstoff-Phosphor-Bindung in den Phosphoestern von Serin oder Threonin, kann diese Modifikation sehr leicht unter sauren Bedingungen abgespalten werden. Daher wurde die Probenvorbereitung für die Analyse dieser Modifikation verändert um sie einer massenspektrometrischen Detektion zugänglich zu machen. Bei der Analyse solcher Verbindungen wird die Phosphorylierung als Phosphorsäure abgespalten, was oft zu Fehlinterpretationen führt. Um die Analyse von N-Phosphorylierungen zu verbessern, haben wir ein Anreicherungsprotokoll für die Metalloxid-Affinitätschromatographie sowie Trennbedingungen für die Umkehrphasen-Chromatographie optimiert. Das verbesserte Protokoll wurde im Folgenden zur Detektion der Phosphorylierungsstellen in verschiedenen bakteriellen Proteinen angewandt. Im Fall der Arginin-Phosphorylierung ist nicht nur die

Probenvorbereitung von großer Wichtigkeit, sondern aufgrund des speziellen Fragmentationsverhaltens im Massenspektrometer wird auch die Interpretation der Daten erschwert. Um diese Modifikation zuverlässig von anderen Protein-Phosphorylierungen zu unterscheiden haben wir die Fragmentierung verschiedener Modellpeptide untersucht.

In dieser Dissertation werden zwei wichtige Probleme aktueller Phospho-Proteomstudien untersucht, die Analyse mehrfach phosphorylierter Peptide und die Eigenschaften von Arginin-gebundenen Proteinphosphorylierungen. Die Entdeckung der Arginin-Phosphorylierung lieferte aufschlussreiche Details zur Adaption von Bakterien an Stressbedingungen. Mit Hilfe einer optimierten Probenvorbereitung und Dateninterpretation kann eine Analyse Stickstoff-gebundener Proteinphosphorylierungen auch mit komplexen Proben wie Proteinkomplexen und Zell-Lysaten möglich werden.

Structure of the PhD thesis - Novel Protocols for the Identification of Protein Phosphorylation Sites on Multiply Phosphorylated Domains and Acid-Labile Phosphorylations of Histidine and Arginine in Proteins

Acknowledgements

Abstract

Zusammenfassung

Structure

Introduction Part 1

A Phosphorylation Processes in Cells

I Abundance and Function 001

II Protein Kinases 004

III Protein Phosphatases 005

B Analysis of Protein Phosphorylation by Mass Spectrometry

I Protein Identification by Mass Spectrometry 006

II Phosphopeptide Enrichment Techniques 011

III Identification by Mass Spectrometry 014

C Properties of Peptides with Multiple Phosphorylation 020

Results 024

Part 1 Enrichment of Multiply Phosphorylated Peptides Using Metal Affinity Chromatography on Titanium Dioxide

a Optimization of MALDI MS Detection and Identification 024

b Development of an Enrichment Procedure 029

c Reversed-Phase Separation of Multiple Phosphorylated Peptides 033

d Comparison to other methods for phosphopeptide analysis 036

Discussion – Part 1 043

Introduction Part 2

D Nitrogen-bound Protein Phosphorylation Processes in Cells 045

I Chemical Properties of Phospho-Amidates on Biomolecules 049

II Identification of N-bound Phosphorylations on Peptides and Proteins 051

III. Mass Spectrometry of Protein N-Phosphorylations 052

Results

Part 2. Identification and Characterization of Arginine-Phosphorylated Peptides and Proteins

a The CtsR Phospho-Switch 053

b Characterization of the Protein Arginine Kinase McsB 062

c Enrichment of Acid-Labile Phosphorylations on TiO_2 067

d Fragmentation of Peptides with Arginine-Phosphorylation 081

Discussion – Part 2 099

Outlook – Analysis of N-phosphorylated Peptides 102

Appendix A: Chemical Cross-Linking in Combination with Mass Spectrometry to Study Structures of Proteins and Protein Complexes 104

Appendix B: Detection of Histone Modification Activity by MALDI-TOF MS 118

References 119

Table of contents 129

figures 129

tables 130

abbreviations 131

Publications 133

In preparation 134

Curriculum Vitae 135

Introduction Part 1

A Phosphorylation Processes in Cells

I Abundance, Properties and Function

All cellular processes rely on protein activity, which makes these molecules indispensable for live. To adapt to environmental changes, cells require fast and reliable regulatory systems, which are often based on networks of interacting proteins. Since expression and degradation of proteins are long lasting processes, cells use enzyme-catalyzed, reversible, post-translational modifications such as phosphorylation, acetylation, methylation, and others to directly modify protein activity. Thereby, protein phosphorylation represents a key activator of protein activity in all kingdoms of live and has been shown to be the key regulator in cell division and signal transduction processes. On the other side, mis-regulation of protein phosphorylation can severely affect cellular processes and thereby lead to oncogenic transformation (Blume-Jensen P. and HunterT., 2001; Rush J. *et al.*, 2005). Recently, a plethora of protein phosphorylation sites has been reported in several large scale studies, which led to the estimate that up to 30% of all eukaryotic proteins encounter this modification during their lifetime (Hunter T., 2000). This makes phosphorylation the most widespread protein modification in eukaryotic cells. Phosphorylation is not only limited to proteins but also sugars, lipids, DNA and RNA are phosphorylated by kinases or phosphorylases (Voet D., 2002). A well studied example for sugar phosphorylation is that of glucose to glucose-6-phosphate, an initiation step of sugar metabolism. Furthermore, also the repair of DNA breaks requires specific phosphorylation to reconnect both strands.

In proteins, phosphorylation is often attached in the form of phospho-esters to the OH-moieties in the side chains of the amino acids serine (Ser), threonine (Thr) or tyrosine (Tyr) (figure A_01; Hunter T., 2000). To a significantly lower extend histidine (His), arginine (Arg), aspartate (Asp), cysteine (Cys), and lysine (Lys) residues can also be phosphorylated. These five amino acid modifications are highly labile under acidic pH conditions that they are often lost durinsample preparation. Furthermore, most of these modifications are reported to be present in bacteria and plants, while their existence in eukaryotes is only assumed. Of the acid-labile phosphorylations, histidine and aspartate phosphorylation have been intensely studied, since they occur in conjunction in the so called two-component signalling systems (Mitrophanov A.Y. and Groisman E.A., 2008). Protein kinases for Arg, Lys, and Cys are

rarely reported (Besant P.G. *et al.*, 2009, Attwood P.V. *et al.*, 2007). The phosphorylations of Ser, Thr and Tyr are quite stable during the chemical treatment most of the common sample preparation methods but can easily be removed by phosphatases, both *in vivo* and *in vitro*. Since both hydrogen atoms in the applied phospho-moiety are acidic, this modification adopts a doubly negative charge state under physiological pH conditions. Such a strong negative charge has powerful influence on the binding of other proteins, which could either be impaired or promoted. For instance 14-3-3 proteins or proteins with polo box domains specifically bind to phosphorylated peptide sequences (Thomas D. *et al.*, 2005; Lowery D.M. *et al.*, 2004). Not only protein-protein interactions are altered upon phosphorylation events, but also structural changes within the phospho-protein itself have been observed which might be re-localization of an inhibitory domain to allow access to the active centre, e.g. activation of calmodulin-dependent kinases.

Protein phosphorylation has important functions in every cell compartment from the transcriptional machinery in the nucleus to secreted proteins in the extra-cellular matrix (Hunter T., 2000). Interestingly, while eukaryotes have a very complex network of phosphorylation pathways, recent data show that this modification is only rudimentary present in bacteria (Macek B. *et al.*, 2007, Miller M.L. *et al.*, 2009). In eukaryotes, signal transduction from the cell membrane to the nucleus is meditated by many protein phosphory-lation events which lead to activation of other kinases and can thereby form phosphorylation cascades. Prominent members of such cascades are the

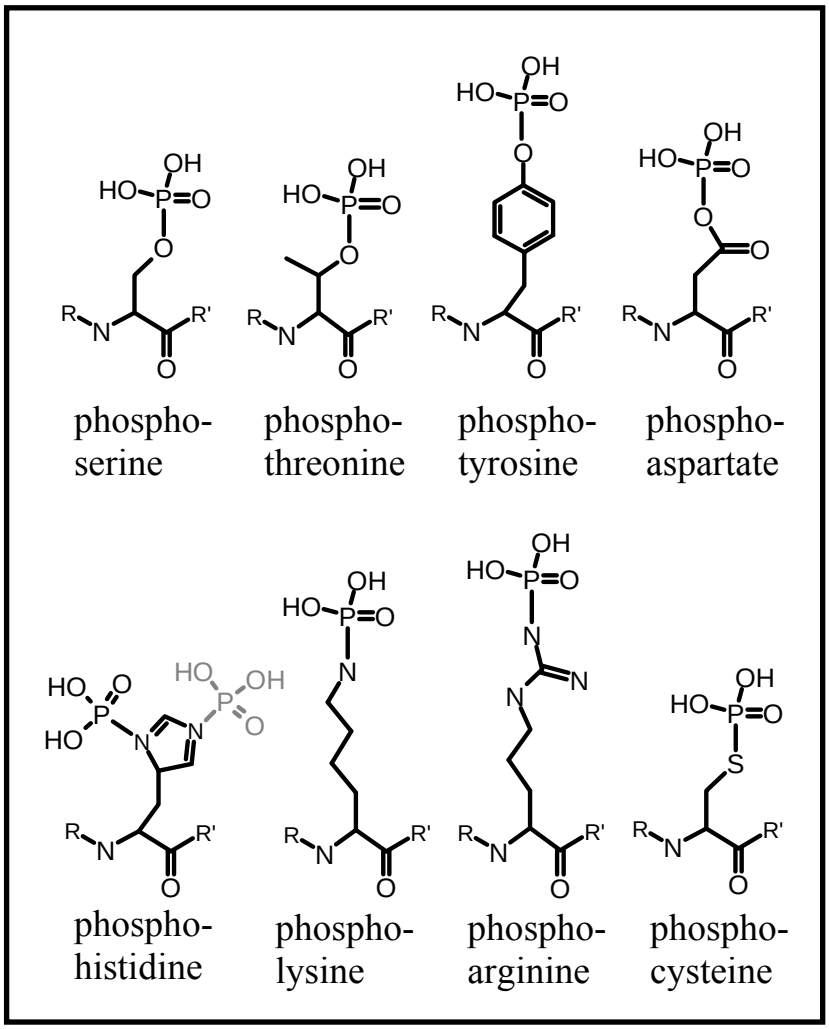


Fig. A-01 Phosphorylated amino acids in proteins

receptor tyrosine kinases (figure A_02) and members of MAP-kinase pathways (mitogen-activated protein kinase) which mediate signal transduction from the cell surface to the transcription machinery (Hunter T., 2000). While the former Tyr-kinases are important to sense growth factors and other extra-cellular signals, the latter MAP-kinases are messengers within the cell which activate transcription factors and thereby regulate gene expression. Phosphorylation cascades consists of several kinases, for instance MAP-kinase, MAP-kinase-kinase, and MAP-kinase-kinase-kinase, which are activated upon phosphorylation and subsequently phosphorylate the next downstream kinase (Teis D. *et al.*, 2002). Also cell division processes depend on the activity of different protein kinases and the appropiate phosphorylation state of key proteins (King R. *et al.*, 1994; Nigg E.A., 2001). As consequence of these essential roles in maintaining the viability of cells, misregulation and inactivation of protein kinases can lead to neoplastic transformation or the promotion of tumor growth, as demonstrated for receptor tyrosine kinases and several members of the MAP kinase family (Blume-Jensen P. and Hunter T., 2001).

Protein phosphorylation is not only able to mediate protein-protein-interactions, it is also a good indicator of kinase-substrate interactions as demonstrated by the signalling cascades. Studying protein phosphorylation can help to explain different states of protein activity at the molecular level, and also enlucidate the pathways behind such important processes like cell division and oncogenic transformation.

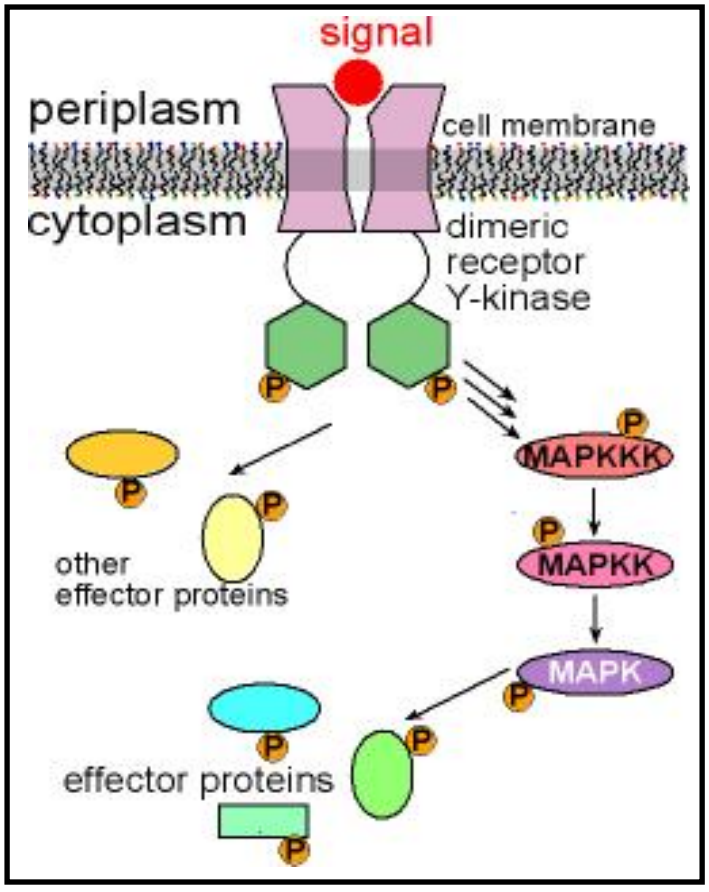


Figure A_02 **Schematic Phosphorylation Cascade in Eukaryotic Cells.**
Interaction with an extra-cellular signal leads to receptor dimerization and activates the Tyr-kinase activity. Both receptors auto-phosphorylate each other in their intra-cellular domains and thereby become active for docking of other proteins. These effector proteins are phosphorylated by the kinase domain and thereby initiate MAP kinase cascades, to transduct the signal to effector proteins, altering gene expression.

II Protein Kinases

Protein kinases are enzymes that catalyze the transfer of the γ-phosphoryl moiety from an adenosyl triphosphate (ATP) molecule onto a specific amino acid in a target protein (figure A_03). Three domains are ubiquitously found in protein kinases: the ATP binding domain, the active centre, and the target protein recognition domain, which can be extended by a regulatory domain. These domains have often evolutionarily conserved amino acid sequences. Since the chemical natures of Ser, Thr, Tyr, and His differ significantly from each other, various classes of kinases have evolved. Furthermore, kinases are normally not able to modify just any putative target amino acid in a protein sequence, for instance serines and threonines. Often only a specific peptide strands or even specific amino acids become phosphorylated. For some kinases these strands have developed specific recognition sequences with certain amino acids at fixed positions close to the target site for phosphorylation. The yeast MAP kinases, for example, require a proline residue C-terminal to the Ser or Thr, but also more complex motifs have been described recently, for instance for protein kinase A or polo-like kinase (Nakajima H. et al., 2003). The aim of this diversity is a tight control over kinase activity which allows the activation of specific cellular pathways. In human cells 518 different kinases are encoded by the DNA and assumed to be expressed (Linding R. *et al*., 2008). Many of them have been proposed based on their amino acid sequences, but not all of these proteins have been detected by biochemical methods. The biochemical data are still required to demonstrate catalytic activity. A different approach for kinase identification uses specific kinase inhibitors, which are often structurally related to ATP, to immobilize kinases onto a solid phase. The bound proteins are identified by mass spectrometry and can be tested by other biochemical techniques to reveal the kinase activity (Bantscheff M. *et al.*, 2007).

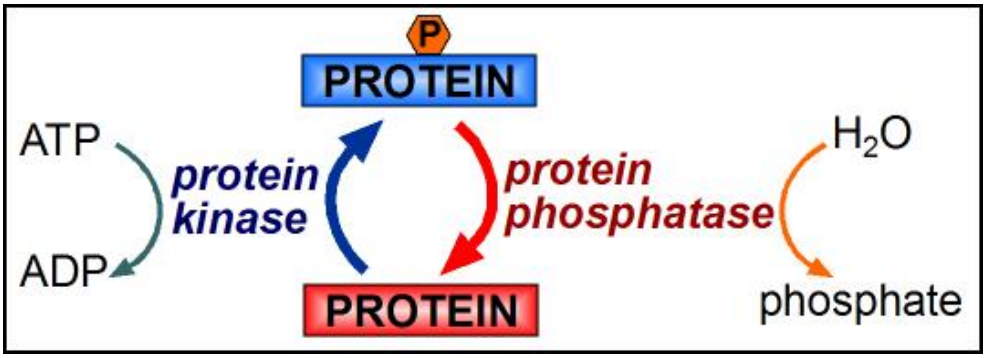


Figure A_03 **Protein Phosphorylation by Kinases and Dephosphorylation by Phosphatases Are Antagonistic Processes in Cells.** Kinases belong to the enzyme class of transferases, protein phosphatases are hydrolases.

III Protein Phosphatases

Similar to the protein kinase family, protein phosphatases have evolved, to generate high diversity for tight control of phosphorylation processes. Nevertheless, some phosphatases have a more general activity and are therefore able to dephosphorylate many different substrates. Since one of the substrates is water, these enzymes belong to the hydrolase family and they are classified by the type of amino acid that becomes dephosphorylated (figure A-03, Theodosiou A. and Ashworth A., 2002). Up to now 4 classes have been described, the Ser/Thr phosphatases, the Tyr phosphatases and the His phosphatases, which act on the individual phospho-amino acids, respectively. The fourth class are dual-specificity phosphatases, which are able to hydrolyze the phospho-ester bond on phosphorylated Ser, Thr, and Tyr (Keyse S.M., 1995). In cells, phosphatases are the antagonists of kinases, and their function is to shut the kinase-activated pathways off when the signal is no longer required. To allow a rapid kinase signalling as required in case of stress, the general phosphatase activity in cells has to be slightly higher than the kinase activity to maintain the supply of kinase substrates.

B Analysis of Protein Phosphorylation by Mass Spectrometry

Incorporation of radioactive phosphorous ^{32}P is still the most prominent technique for visualizing kinase activity (Wålinder O. *et al.*, 1969). Unfortunately, this method is not able to determine the phosphorylated amino acid and has to be combined with mutational studies where putatively phosphorylated amino acids are exchange to such which cannot be modified. Despite the sensitivity of detection systems for radioactive markers other methods have been established to study protein phosphorylation, since radioactive radiation is ionizing and can cause health damage, when working with it over long time. Specific antibodies for phosphotyrosine (phospho-Tyr) can be used in Western Blot experiments to detect proteins with this modification. Phosphoserine (phospho-Ser) and phosphothreonine (phospho-Thr) can be detected by phosphorylation specific stainings, since available antibodies cannot universally detect this modification. Other antibodies have to be generated to detect specific phosphorylated peptide strands in selected proteins.

Recently, technological advances and novel enrichment methods made mass spectrometry (MS) a highly important tool for phospho-protein research. This technique provides the ability to observe multiple phosphorylation events on many proteins within a single experiment and is therefore often faster than mutating specific amino acids. The introduction of label-free or isotope-based relative quantification methods even allows distinction of the phosphorylation state of a protein in different cell states or at different timepoints (Agrawal G.K. and Thelen J.J., 2006; Villen, J. et al., 2007).

I Protein Identification by Mass Spectrometry

Currently, protein identification by mass spectrometry (MS) is a routine method in biochemistry. Individual proteins of interest or complex mixtures of proteins and cell lysates are enzymatically cleaved into short peptide fragments see figure A_04A. Prior to the digestion step, proteins are denatured and disulfide bonds that stabilize the three-dimensional structure are reduced to sulfhydryls and modified by alkylation, for instance with iodoacetamide. Then a protease, which hydrolyses amide bonds at specific positions within the amino acid sequence, is added to the mixture. A commonly-used protease is trypsin, which cleaves C-terminally to Arg and Lys and thereby produces a distinct peptide pattern. The resulting peptide mixture is further fractionated by chromatography steps; usually “reversed-phase” liquid chromatography (RP-LC) is the last chromatography step prior to

mass spectrometric detection of the peptides. Nowadays, nano-scale high performance liquid chromatography (HPLC) systems are directly linked to the inlet of the mass spectrometer which allows on-line-detection and analysis of peptides even from quite complex mixtures. In the late 1980s two major ionization techniques have been invented, electro-spray ionization and matrix-assisted laser desorption/ionization. Both are so-called soft ionization techniques because relatively labile biomolecules can be transferred into the gas-phase without undergoing extensive fractionation, which was observed with other ionization techniques (Karas M. and Hillenkamp F., 1988; Fenn J.B. *et al.*, 1989). Upon ESI, dissolved analytes are sprayed from the tip of a capillary, connected to an electrode, towards the entrance of the mass spectrometer, which is the counter electrode. During this process the solvent evaporates and in case of peptides leaves often multiply-charged analyte ions. The MALDI process requires co-crystallization of the sample with an excess of matrix, which is a substance that can be irradiated by UV lasers. The energy uptake upon laser exposure leads to a local temperature increase and evaporation of matrix molecules, also transferring analytes into the gas phase. Application of an electric field accelerates the analyte ions into direction of the mass analyzer. Whereas ESI is easily coupled to an HPLC system, MALDI sample preparation requires a spotting device that fractionates the sample onto the sample plate.

Following ionization, the analyte ions are separated by the mass analyzer and finally detected. The principles of mass analyzers are based on the behaviour of ions in electric or magnetic fields or the flight time through a fieldless tube. These processes are combined into computer software, allowing automatic acquisition of multiple MS spectra during HPLC separation. For protein identification, acquired mass spectra are typically compared to a database that contains all expected protein sequences. According to the specificity of the proteolytic enzyme, a list of theoretically observable peptides is generated and matched to the experimental data. If a high number of theoretical peptides is found in the MS data, the protein is more reliably identified. Due to the fact that the protein sequence is reconstructed from its peptides this method is called *bottom-up* approach and a scheme is shown in figure A_04A.

With growing knowledge of the proteome, it became obvious that identification of a protein based on the peptide masses from the enzymatic cleavage is not accurate enough, since it determines only the amino acid composition. This is often fulfilled by several peptides of different proteins. Higher assurance for the peptide identity is provided, if each amino acid can be localized within the sequence (Arnott D. *et al.*, 1993). To gain this information, peptides have to be sequenced within the mass spectrometer, which is achieved by

dissociation of an amide bond in the peptide backbone in tandem MS experiments (MS/MS). Special fragmentation methods have been introduced for this purpose, which either increase the internal energy of the peptide, for instance by collisions with gas molecules (collision-induced dissociation, CID) or infrared radiation (infrared multi-photon dissociation, IRMPD), or induce radical dissociations by uptake of low-energy electrons (Hunt D.F. *et al.*, 1986; Stensballe A. *et al.*, 2000). Methods which increase the internal energy of the precursor ion to induce fragmentations are "so-called" thermal activation methods and lead to dissociation of the weakest bond in the molecule (see figure A_04B), which is the peptide bond in case of unmodified peptides (Hunt D.F. *et al.*, 1986). This fragmentation type predominantly leads to formation of "so-called" b- and y-type ions and it became the standard method in proteomics (Roepstorff P. and Fohlman J., 1984). It is available in every mass spectrometer, easily implemented into the analysis program and delivers highly reliable peptide identifications. Obtained MS/MS data are also searched against a database but now each peptide is validated individually so that peptides with equal amino acid composition but different sequence can be distinguished. Identification of amino acid sequences reduced the number of false positive identifications dramatically.

To circumvent the laborious digestion step and to reveal the mass of the intact protein, a different methodology is available, *top-down* analysis (see figure A_04C, Reid G.E. *et al.*, 2001; Ge Y. *et al.*, 2002). Although intact proteins can be analyzed with almost all instrument types, this technique is routinely applied on Fourier-transform (FT) mass spectrometers, which possess highest mass accuracy and highest resolution. *Top down* MS approaches were also demonstrated for ion trapping and quadrupole-TOF (qToF) instruments. Mass differences of measured and calculated protein mass provide hints for putative amino acid modifications and cleavage of signal peptides. Fragmentation of the whole protein leads to localization of the modification site within the protein sequence (Ge Y. *et al.*, 2003). Due to the physical parameters for the ion transfer into the mass analyzer the protein size is limited to approximately 50 kDa for commercially available instruments; custom built mass spectrometers are able to analyze much larger proteins.

QTOF instruments have an almost unlimited mass range but lack the resolution and accuracy. The high mass range provides analysis of intact protein complexes (Rostom A.A. and Robinson C.V., 1999; Ruotolo B.V. and Robinson C.V., 2006). Upon CID activation, individual subunits of the complex can be released, elucidating specific protein-protein interactions and the structural association of the complex.

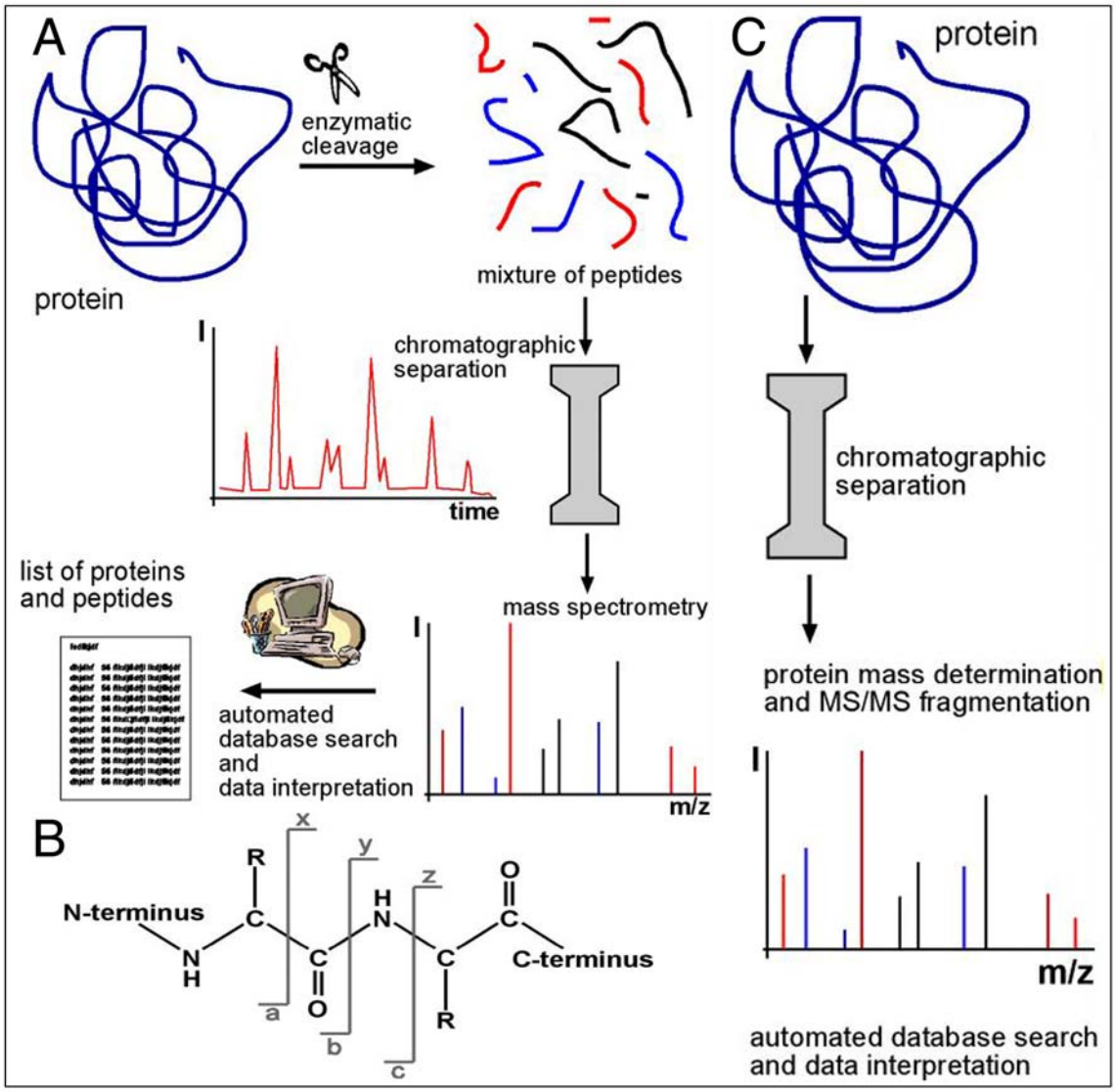


Figure A_04 **Protein Analysis by Mass Spectrometry**. A. Schematic representation of "bottom up"-proteomics approach. After enzymatic cleavage of the protein, peptides are separated by chromatographic methods and detected in the mass spectrometer. Proteins are identified by their respective peptides. B. Fragmentation of oligopeptides in tandem MS processes leads to breaking of the peptide backbone, resulting in characteristic fragment ions (daughter ions). Small characters represent the nomenclature of peptide fragment ions. C. Analysis of full-length proteins by *top down* mass spectrometry is applied to identify protein sequences and modifications, but requires purified protein samples.

II Phosphopeptide Enrichment Techniques

Protein phosphorylation in cells is not a stoichiometric process but more often this modification is applied to only a small fraction of the observed protein. Furthermore, sample preparation for mass spectrometry requires enzymatic cleavage of the proteins to small peptide fragments. Since only specific amino acids are phosphorylated in a protein molecule, a large fraction of the peptide mixture forms a biological background, which interferes with the analysis of phosphorylated peptides. This background of non-phosphorylated peptides influences the analysis by reducing the ionization efficiency for phosphopeptides and the chance to select phosphopeptide ions for MS/MS analysis (Janek K. *et al.*, 2001; Steen H. *et al.*, 2006).

To reduce this biological background and improve phosphopeptide detection, special enrichment protocols have been developed recently. The aim of each experimental setup is to physically isolate phosphopeptides from all other peptides. Chromatographic techniques that rely on a specific affinity of the phospho-moiety to the solid phase have been successfully applied to enrich phosphopeptides from highly complex samples and become standard techniques in phosphoproteomics (Thingholm T.E. *et al.*, 2009). Other techniques such as isoelectric focussing, thin-layer chromatography, phospho-amidate chemistry, or calcium hydroxide precipitation have not been widely used but also proved good enrichment results (Xu C.F. *et al.*, 2007; Kochin V. *et al.*, 2006; Bodenmiller B. *et al.*, 2007, Zhang X. *et al.*, 2007).

Phospho-tyrosine has also an outstanding position, since it is the only phosphorylation that can be isolated by antibody-affinity chromatography, routinely (Rush J. *et al.*, 2005). Therefore, anti-phosphotyrosine antibodies are immobilized on a chromatographic resin such as modified silica and this material can then be mixed with the sample. Under specific elution conditions the antibody releases the peptide, which is further analyzed by mass spectrometry.

Strong Cation Exchange (SCX) and Electrostatic-Repulsion Hydrophilic-Interaction (ERLIC) Chromatography

Due to the negative charge of the phospho-moiety even at low pH, interaction with the already negatively charged strong-cation exchange chromatography resin is impaired, and furthermore repulsion of the peptide could be observed as schematically demonstrated in figure A_05A (Trinidad J.C. *et al.*, 2006). Other peptides adopt a positive charge, since carboxylic acids and amine functions are protonated and are therefore retained on the column.

This method allows partial enrichment of phosphopeptides within the flow-through and the first fractions. Phosphopeptides with more than one positively charged amino acid in their sequence are also retained and might elute together with unphosphorylated peptides.

A sophisticated procedure which uses ion pairing is electrostatic-repulsion hydrophilic-interaction chromatography (ERLIC) which introduces the idea of competitive binding of small organic acids to a weak anion exchange column (WAX) to prevent unphosphorylated peptides from interaction with the stationary phase (see figure A_05B, Alpert A.J., 2008). Ammonium formiate and ammonium methylphosphonate have been depicted as masking reagents because they have a lower affinity towards the stationary phase than the phospho-moiety and guarantee stable pH conditions in the range of pH 2-3. Elution of the peptides can be achieved with a gradient of increasing phosphate concentration which outcompetes the phosphopeptides for the binding sites and allows even a separation due to the number of phospho-moieties on the peptide. Since establishing this method is quite laborious, the potency of this method is yet not fully explored.

Immobilized-Metal Ion Affinity Chromatography (IMAC)

Since phospho-moieties are negatively charged down to very low pH values (the first pKa of the phospho-moiety is about 1.2) they can form salt bridges to positively charged metal ions. A particularly high affinity was thereby observed for three-valent metal ions such as iron Fe^{3+}, gallium Ga^{3+} and aluminium Al^{3+} (Ficarro S.B. et al, 2002). This interaction requires only two free ligand binding sites on the metal ion, so that the other sites can be used to immobilize the metal ion on a chromatographic resin. Under acidic loading conditions, phosphopeptides occupy the positively charged binding sites much stronger than peptides without phosphorylated amino acids. In contrast, these peptides compete with the buffer molecules for residual free binding sites due to their low affinity and can be washed from the column with few column volumes of loading buffer. Afterwards, phosphopeptides are selectively eluted with a sodium phosphate solution at neutral or basic pH. To reconstitute the column material, the respective metal salt solution is applied under loading conditions. The carrier for the metal ion is an organic molecule with several coordination sites, which chemical nature could be carboxylic acids or nitrogen atoms. The structure of that ligand is similar to that of chelating reagents in solution as for instance imido-diacetic acid derivatives (Ficarro S.B. et al, 2002). The carboxylic acids of this chelate are still sensitive towards protonation at very low pH, thereby allowing an application of the IMAC method only at pH values above 3.5. The pH limitation causes a partial deprotonation of acidic amino acids,

aspartate and glutamate, which then become able to interact with the positively charged metal ions on the resin and enrich highly acidic peptides (Ficarro S.B. et al, 2002).

Later on, an effective method to avoid binding of unphosphorylated peptides was introduced by Ficarro *et al.*; modification of the carboxylic acid side chains in peptides by a methyl ester prior to the enrichment step (Ficarro S.B. et al, 2005). Using this protocol 13000 phosphorylation sites have been identified in the mouse brain (Villen J. and Gygi S.P., 2008). Since the elution solvent contains a high salt concentration and the polymeric stationary phase is sensitive to high pressure, this method is predominantly applied in batch mode on the bench and not in automated HPLC systems. Nevertheless, it is difficult to adopt this method to varying sample amounts. A schematic illustration of the phosphopeptide binding is demonstrated in figure A_05C.

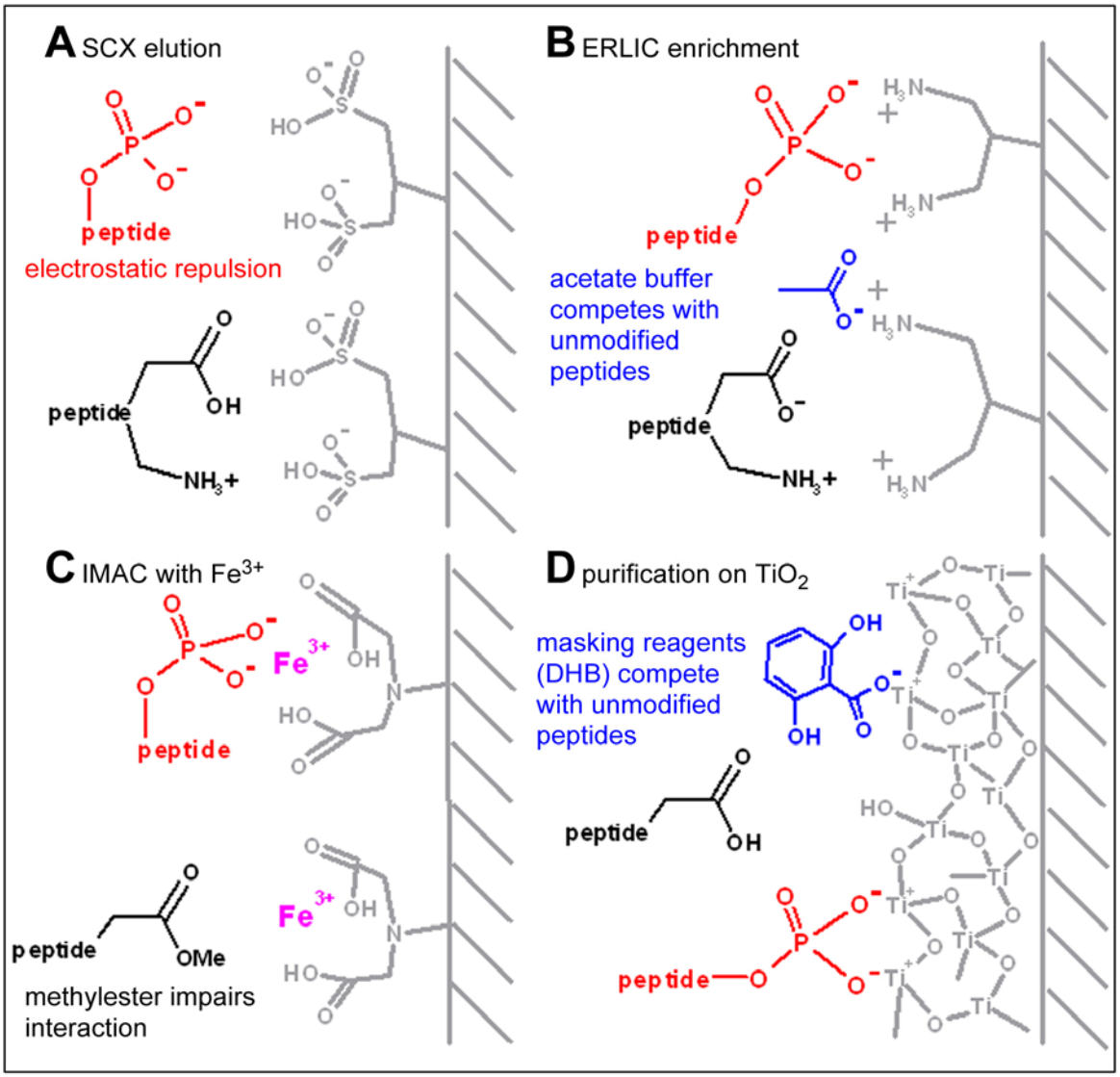


Figure A-05 **Enrichment Techniques for Phosphopeptides**. While SCX (A) chromatography elutes phosphopeptides earlier than unphosphorylated peptides, ERLIC (B),

IMAC (C), and MOAC-based (D) methods retain phosphorylated peptides on the column. Phosphopeptides stay attached to the resin under loading and washing conditions and are eluted with a specific elution buffer. Masking substances as buffer components competitively remove unphosphorylated peptides during loading and washing steps in ERLIC, IMAC, and MOAC-based protocols.

Metal oxide-affinity chromatography (MOAC) on titanium dioxide TiO_2, zirconium dioxide ZrO_2, and tin oxide SnO_2

The enrichment procedure with the highest impact in the recent years was metal oxide-affinity chromatography (MOAC) on spheres of metal oxides from Ti, Zr, or Sn. Theroretical arguments and affinity experiments of inorganic and organic ions (Cl^-, SO_4^{2-}, NO_3^-, CO_3^{2-}, acetate, formiate, PO_4^{3-}) towards these metal oxides revealed that only phosphate was strongly adsorbed on the surface and every other anion could be released by washing with a phosphate solution (Kweon H.K. and Hakansson K., 2006). Since phosphopeptides represent the monoesters of phosphoric acid, these molecules should also be able to attach to the surface of these metal oxides with high affinity, which provides a good method to separate these analytes from the majority of unphosphorylated peptides (Cantin G.T. *et al.*, 2007).

While the first protocols achieved only slight depletion of unmodified peptide due to the fact that carboxylic acid can also bind to these materials, introduction of masking reagents such as acetic acid, oxalic acid, pthalic acid, dihydroxybenzoic acid, or lactic acid resulted in almost complete depletion of unmodified peptides with high recovery of phosphopeptides (Pinkse M.W. *et al.*, 2004, Larsen M.R. *et al.*, 2005). Currently, the binding mode of the phosphoryl group is heavily discussed; one theory states that two oxygen atoms of the phospho-moiety are able to act as a bi-dentate ligand which coordinates two binding sites on Ti-ions while carboxylic acids occupy only a single coordination site (Larsen M.R. *et al.*, 2005). Masking reagents bind competitively to the same sites, which can be occupied by unmodified peptides and the excess of these additives shifts the equilibrium towards release of unmodified peptides. The phosphate-metal ion complex has a higher stability so that enrichment of phosphopeptides is not disturbed. Another theory suggests the binding towards binding sites with different chemical nature, so that the masking reagents can successfully block the binding sites for COOH-moieties, while the phospho-binding sites are not affected (Larsen M.R. *et al.*, 2005). The binding mode is schematically drawn in figure A_05D. Interestingly, this procedure requires a high amount of organic solvent for sample loading to

impair undesired hydrophobic interactions of unphosphorylated peptides and to increase the solubility of peptides. Detergents and other hydrophobic sample additives stay also in the mobile phase during sample washing which allows further purification of phosphopeptides from these substances (Jensen S.S. and Larsen M.R., 2007). Low pH conditions also improve the binding, since the affinity of the carboxylate towards TiO_2 is dramatically decreased by protonation. Phosphopeptides are eluted by a pH shift from acidic to alkaline, which induces dissociation of the surface OH-groups and electrostatic repulsion of the phosphopeptides.

Recent protocols for TiO_2 based phosphopeptide enrichment demonstrated that even enzymatic digests of cell lysates can be pepared with this method. This enrichment effect could not be demonstrated with the earlier protocols, since highly acidic, unphosphorylated peptides were not depleted without effective masking reagents (Pinske M.W. *et al.*, 2008). Another remaining question was, if highly acidic phosphopeptides and peptides with multiple phosphorylations are partially lost on the surface of the metal oxide due to irreversible binding. This observation might result from the elution conditions, since most early protocols suggested 10% ammonia for phosphopeptide elution (Jensen S.S. and Larsen M.R., 2007). We and others could show that elution with ammonium phosphate also releases such peptides quantitatively, so that nowadays metal-oxide enrichment of phosphopeptides is a standard method next to the IMAC procedure. It was further shown that a combination of both methods even increased the phosphopeptide yield due to a better fractionation of phosphopeptides with different number of phospho-moieties (Thingholm T.E. *et al.*, 2008).

III Identification by Mass Spectrometry

Identification of phosphorylation sites of a protein by proteomic methods is still a challenging task, because phosphopeptides behave differently from unmodified ones. Not only the chromatographic behaviour is significantly altered but most importantly the ionization efficiency is significantly reduced, especially if peptides without phospho-moieties are present (Steen H. *et al.*, 2006; Nabetani T. *et al.*, 2006; Gao Y. and Wang Y., 2007). Due to the chemical nature of the phospho-ester of Ser and Thr and the phospho-amidate bond of His, these covalent bonds dissociate more effectively upon thermal activation than the peptide bonds between the individual amino acids. As shown in figure A_06A, CID-MSMS fragmentation of phosphopeptides leads to dissociation in the phosphorylated side chain. The majority of the phosphopeptide ions undergoes a neutral loss of phosphoric acid in positive ionization mode (Schlosser A. *et al.*, 2001; Boersema P.J. *et al.*, 2009). Several mechanisms

have been proposed for this reaction, which in case of Ser and Thr is based on parallel elimination of bi-phosphate and the hydrogen atom in β-position, thereby forming a double bond (Boersema P.J. *et al.*, 2009). For His this mechanism is not yet fully understood, but might involve a transfer of the phospho-moiety onto an OH group on another amino acid (Medzihradszky K.F. *et al.*, 1997; Zu X.L. *et al.*, 2007) Excessive loss of H_3PO_4 causes a low number and intensity of fragment ion signals, so that spectra yield only little sequence information. The aromatic system in side chain of tyrosine impairs the β-elimination process of the phenolic C-O bond and thereby increases the stability of such phospho-esters. Therefore tandem MS analysis of Tyr-phosphorylated peptides leads mainly to dissociation of amid bonds between the amino acids (Steen H. *et al.*, 2001). Several adaptations have been introduced recently to improve fragmention phosphorylated peptides, which led to broad application and discovery of numerous novel phosphorylation sites.

Phosphatase treatment

A quite simple method to detect phosphopeptides is to compare mass spectra of the enzymatically cleaved sample with a sample treated with an unspecific phosphatase (figure A_06B). Due to the enzymatic activity, phosphoryl groups are removed from the phosphopeptides and the mass of the peptide ion is decreased by the mass of the phosphoryl group (80 Da) (Raggiaschi R. *et al.*, 2006). Furthermore, the dephosphorylated peptide can easily be sequenced by tandem MS, thereby revealing their amino acid sequence. Since this process requires the comparison of two different samples a mass spectrometer with good mass accuracy and resolution is required to determine both peptide forms correctly. Unfortunately, the information of the location of the modified amino acid is lost during the hydrolysis.

Neutral Loss and Immonium Ion Screening

The occurring neutral losses of the phospho-moiety can also be used to find phosphopeptides in mixture with unmodified peptides (figure A_06C). Precursor ion scanning methods on triple quadrupole mass spectrometers are applied to detect the loss of H_3PO_4 upon CID activation in positive ionization mode or the formation of the HPO_3-ion in negative mode (Schlosser A. *et al.*, 2001). This method is also valuable for detection of peptides with Tyr-phosphorylation, but instead of neutral loss ions, the immonium ion of phosphotyrosine at 216.11 Da is detected (Steen H. *et al.*, 2001). Unfortunately, peptide sequence and the location of the phosphorylated amino acid remains unkown without further analyses.

Similarly, switching the ionization mode from positive to negative ionization was reported to result in significantly higher signal intensities for phosphorylated peptides (Janek K. *et al.*, 2006; Nabetani T. *et al.*, 2006). Comparison of MS data of both ionization conditions further helps to validate the phosphorylation state of a peptide.

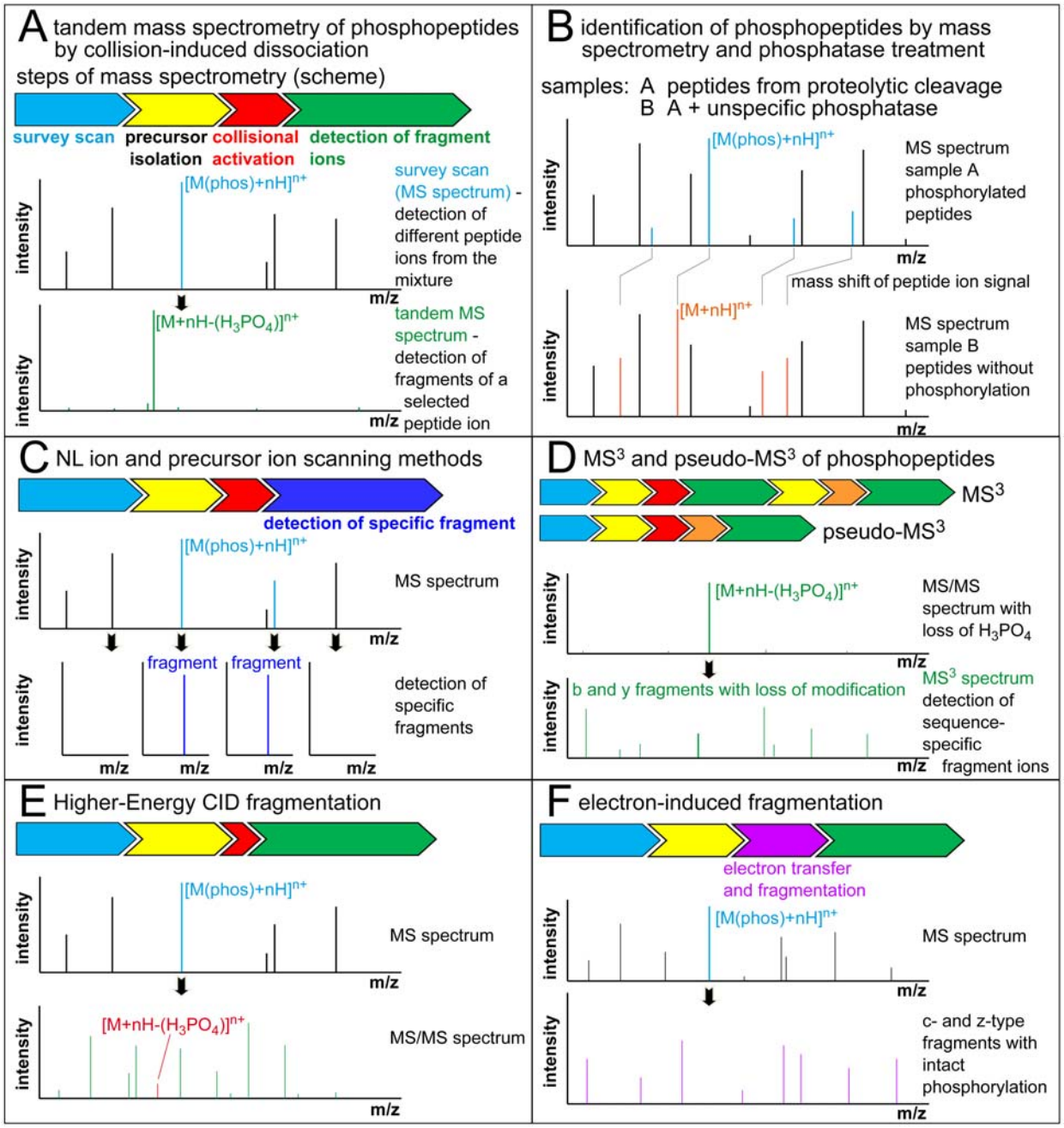


Figure A_06 **Mass Spectrometry of Phosphorylated Peptides**. A. Tandem MS analysis by thermal activation methods of most phosphopeptides results in extensive elimination of phosphoric acid, thereby impairing detection of backbone fragment ions and identification of the amino acid sequence. B. Phosphatase treatment can identify the number of

phosphorylations and the phosphorylated sequence, but the information of the phosphorylated amino acid is lost. C. Precursor scanning methods specifically identify phosphopeptides, but require a combination with standard tandem MS analysis to reveal the amino acid sequence. D. A second fragmentation step is introduced in MS^3 and pseudo-MS^3 methods in which the dephosphorylated peptide ion (shown in A) is selected for further fragmentation to reveal the amino acid sequence. E. Higher fragmentation energies induce backbone cleavages with higher frequency than low-energy methods. Therefore backbone fragments gain higher signal intensities, but also different neutral loss pathways can be activated. F Electron-induced fragmentations preserve the phospho-moiety on the fragment ions which makes this method mostly suited for phosphopeptide analysis.

MS^n and pseudo-MS^n

Despite their low mass resolution and accuracy, ion trapping mass spectrometers have the unique opportunity to perform multiple stages of fragmentation which allows very specific investigation of peptide sequences or reliable MS/MS spectra of peptides with labile modifications (figure A_06D) (Beausoleil S.A. *et al.*, 2004). While MS^n methods require the detection of fragmentation products after each fragmentation step to select the ion of choice, pseudo-MS^n activates putative product ions without isolation of the respective ion (Kalkum M. *et al.*, 2003; Chang E.J. *et al.*, 2004). In comparison to real MS^n, pseudo- MS^n is faster because it does not require refilling of the ion trap after the first fragmentation step, but the obtained mass spectra show fragment products of each fragmentation step which makes data interpretation more difficult. In real MS^n methods the obtained mass spectrum consists only of the fragment ions of the respective precursor ion. A special method for phosphopeptide research is the neutral loss-triggered MS^3 method which acquires MS^3 spectra of phosphopeptides only when they exhibit a strong loss of H_3PO_4 in the MS/MS step. This could also save time since not every peptide has to be fragmented in two adjacent steps (Chang E.J. *et al.*, 2004).

These methods are the standard techniques in phosphoproteomics, since the neutral loss species are further fragmented and their amino acid sequence is accessible. In contrast to phosphatase treatment and neutral loss scanning only a single MS experiment is required. β-elimination of H_3PO_4 in Ser and Thr residues nominally leads to a loss of water in comparison to the unmodified amino acid, which allows localization of the phospho-amino acid in an MS^n spectrum (Claverol S. *et al.*, 2003). Since water losses can also occur from other residues the specificity of these signals is currently discussed (Palumbo A.M. and Reid G.E., 2009).

Pseudo-MS^n spectra contain also fragment ions with intact modifications which remain from the first fragmentation step and often allow reliable identification of the modified amino acid.

Another benefit of the MS^n techniques is easy implementation into any HPLC-MS method without data analysis to find putative phosphopeptides from the pre-screen. This facilitates database search and identification of the phosphopeptides.

Higher energy collision-induced dissociation (HE-CID)

In comparison to the collisional activation in ion trapping instruments, mass spectrometers with multi-polar mass analyzers like quadrupoles deliver significantly higher collision energy in a relatively short activation time (figure A_06E, Chang E.J. *et al.*, 2004). The energy uptake is high enough to induce dissociation of bonds, which require more energy than a break at the amide bond, for instance in amino aid side chains, or to trigger dissociation of multiple bonds (Chang E.J. *et al.*, 2004). Mass spectra which are acquired under these conditions have a high diversity of fragment ions; in case of phosphopeptides also fragments with intact phosphorylation sites can be identified. Peptide ion signals with neutral loss of modifications occur with reduced intensity compared to ion trap MS/MS spectra as the result of a more average dissociation of covalent bonds due to the high energy (Olsen J.V. *et al.*, 2007).

Similarly, fragmentation of peptide ions in MALDI instruments also depends on the uptake of high amounts of energy during the ionization which leads to formation of multiple different fragment ion types in post-source decay (PSD) (Annan R.S. and Carr S.A., 1996; Schmidt A. *et al.*, 2008). Phosphopeptides undergo also tremendous neutral loss of phosphoric acid in MALDI-PSD-MS, but at low intensities sequence specific fragment ions can be found.

Another method for thermal activation uses infrared radiation, which increases the internal energy and leads to fragmentations similar to that observed in CID processes (Flora J.W. and Muddiman D.C., 2001). Since the implementation of infrared multi-photon dissociation (IRMPD) requires an external IR-laser as source and special geometry of the mass spectrometer, it is almost exclusively applied to FT-ICR mass spectrometers.

Electron-induced fragmentations

Electron-induced fragmentation procedures follow completely different fragmentation rules. In positive ionization mode a peptide acts as acceptor of low energy electrons, which are caught by positively charge side chains like lysine and arginine. Free, emitted electrons in

electron-capture dissociation (ECD) or chemically ionized polyaromatic anions such as fluoranthen in electron-transfer dissociation (ETD) act as electron donors (Stensballe A. *et al.*, 2000; Syka J.E. *et al.*, 2003; Gunawardena H.P. *et al.*, 2006). After translocation of the electron to the peptide bond, radical rearrangement results in dissociation of the N-Cα-bond and thereby forms c and z ions (see figure A_04B and figure A_06F) (Stensballe A. *et al.*, 2000). This type of fragmentation requires no increase of internal energy and does not rely on the dissociation energy of a specific bond, which is demonstrated in average signal intensities as well as integrity of almost all labile peptide modifications such as phosphorylation or glycosylation (Sleno L. and Vollmer D.A., 2004). In comparison to ECD, which is almost exclusively available in FT-ion cyclotron resonance (ICR) MS instruments, ETD is applied in ion trap mass spectrometers and has higher fragmentation efficiency, due to the cation-anion interaction (Syka J.E. *et al.*, 2003).

Unfortunately, these methods are not as robust as the MS^n technique and they have significantly lower fragmentation efficiency. Furthermore, singly charged ions are completely excluded from fragmentation analysis because they lose their charge upon the electron uptake. Fragmentation of doubly charged peptides often results in low sequence information. However, electron-induced fragmentation methods are highly efficient to study large peptides or small proteins in top down experiments.

Studies of electron transfer and detachment in negative ionization mode have been conducted, but so far no common mechanism has been reported (Gunawardena H.P. *et al.*, 2006; Kleinijenhuis A.J. *et al.*, 2008). With regard to phosphopeptide analysis, electron-detachment dissociation (EDD) delivered high quality fragment ion spectra that allowed identification of the modified amino acid in almost all cases. Only the negative ionization mode, which is seldomly used in common proteomic protocols, prevents a wider application of this technique.

C Properties of Peptides with Multiple Phosphorylation

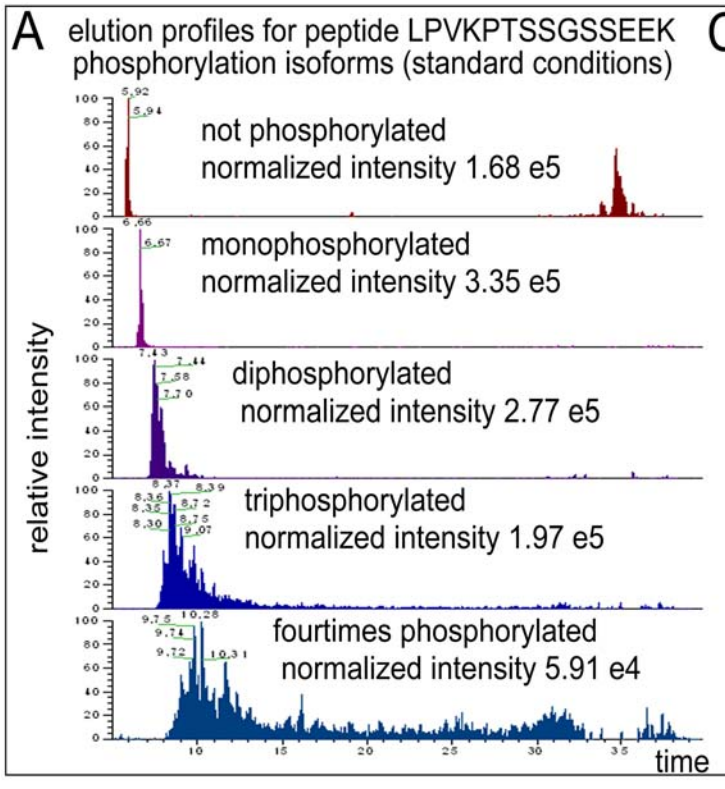
RP chromatography of multi-phosphorylated peptides

Chromatographic separation of peptides with multiple phosphate moieties under standard nanoRP-HPLC-MS conditions is often very difficult because such peptides tend to elute over a broad time window, even if relatively short gradients are applied (see figure A_07A, Choi H. *et al*., 2008). This "smearing" is caused by undesired hydrophilic interaction with the silica material, especially enclosed metal atoms, as well as the abundance of different "charge isoforms" of the same peptide. These isoforms are a result from dissociation of the phosphate moieties, since the phospho-ester moiety is a stronger acid than formic acid. The first pKa of the phosphorylation is at 1.2 and the second at 4.6. Separation conditions for RP-HPLC prior to MS of peptides often include 0.1% formic acid and a low percentage of acetonitrile which has a pH of approx. 2.6 and lies between the first and the second pKa value of the phosphoester. Increase of the organic modifier during the gradient elution will raise the pH by approximately 0.5 to one unit. Since pH conditions of the separation are near to the first equivalence point each phospho-moiety is mainly present in the singly de-protonated state, but could also adopt full protonation at the beginning of the gradient. At high percentage of organic solvent, singly or doubly negative charged forms occur, since the pH is slightly raised. High amounts of organic modifier in the solvent reduce the protonation state of all charge groups. Solvation by the liquid phase is limited due to the lower amount of water, thereby impairing further deprotonation and establishing of a doubly negative charge. The increased pH on the other side promotes deprotonation, especially since the phospho-moiety represents the stronger acid. A further effect of the organic solvent is increased stability of electrostatic interactions of the phospho-moiety with the silica surface which results in a prolonged retention of phosphopeptides. Therefore, pH conditions during chromatography allow the phosphate moieties to adopt a fully protonated, singly deprotonated, as well as doubly deprotonated state.

As an example, figure A_07B shows uncharged to fourtimes-negatively charged peptide isoforms for a fourtimes phosphorylated peptide, when only fully protonated and singly deprotonated forms of phosphate moieties are considered. All putative charge states of such a peptide are shown in a simplified version in figure A_07C where 0 represents fully protonated state; - and -- denote singly and doubly negative charged phospho-groups, respectively. As the table shows, 69 different charge isoforms are possible under standard chromatographic conditions of which most have slightly different retention effects.

Uncharged and singly negative charged molecules have higher hydrophobicity than highly charged molecules, whereas these molecules interact intensely with hydrophilic or charged surfaces, as for instance the fused silica material and especially metal ions, which can be found in low concentration in almost all stationary phases. Even different charge distributions will cause a difference in retention times.

Since peak broadening reduces the peak height, subsequent ESI-MS analysis of these peptide species is often not achieved. As demonstrated in figure A_07A, while unmodified and singly phosphorylated peptides elute in a sharp peak, elution profiles for higher phosphorylated forms become very broad. Guessing from the individual peak heights only, the monophosphorylated peptide seems to represent the most abundant species, whereas the peak area clearly shows that the fourtimes phosphorylated form has the highest abundance. During MS acquisition in data-dependent manner, the instrument is not able to identify such broad elution profiles thereby resulting either in missed identification due to low intensity or a multiple selection for MS/MS of the same peptide and increased undersampling of other peptides. The fourtimes phosphorylated peptide shown in figure A_07A was selected for MS/MS analysis at least 19 times over the whole elution profile. Similar retention behaviour was observed also for other multiply phosphorylated peptides. Since peptides with multiple phosphorylations have a very strong negative charge, it was believed, that they bind irreversibly to specific surfaces, such as TiO_2 (Larsen M.R. *et al.*, 2005). Such effects would limit the applicability of specific enrichment protocols tremendously.



C

	position of phosphate moiety					position of phosphate moiety			
	A	B	C	D		A	B	C	D
1	0	0	0	0	36	0	--	-	0
2	-	0	0	0	37	0	--	0	-
3	0	-	0	0	38	-	0	--	0
4	0	0	-	0	39	0	-	--	0
5	0	0	0	-	40	0	0	--	-
6	-	-	0	0	41	-	0	0	--
7	-	0	-	0	42	0	-	0	--
8	-	0	0	-	43	0	0	-	--
9	0	-	-	0	44	--	-	-	0
10	0	-	0	-	45	--	-	0	-
11	0	0	-	-	46	--	0	-	-
12	-	-	-	0	47	-	--	-	0
13	-	-	0	-	48	-	--	0	-
14	-	0	-	-	49	0	--	-	-
15	0	-	-	-	50	-	-	--	0
16	-	-	-	-	51	-	0	--	-
17	--	0	0	0	52	0	-	--	-
18	0	--	0	0	53	-	-	0	--
19	0	0	--	0	54	-	0	-	--
20	0	0	0	--	55	0	-	-	--
21	--	--	0	0	56	--	-	-	-
22	--	0	--	0	57	-	--	-	-
23	--	0	0	--	58	-	-	--	-
24	0	--	--	0	59	-	-	-	--
25	0	--	0	--	60	--	--	-	-
26	0	0	--	--	61	--	-	--	-
27	--	--	--	0	62	--	-	-	--
28	--	--	0	--	63	-	--	--	-
29	--	0	--	--	64	-	--	-	--
30	0	--	--	--	65	-	-	--	--
31	--	--	--	--	66	--	--	--	-
32	--	-	0	0	67	--	--	-	--
33	--	0	-	0	68	--	-	--	--
34	--	0	0	-	69	-	--	--	--
35	-	--	0	0					

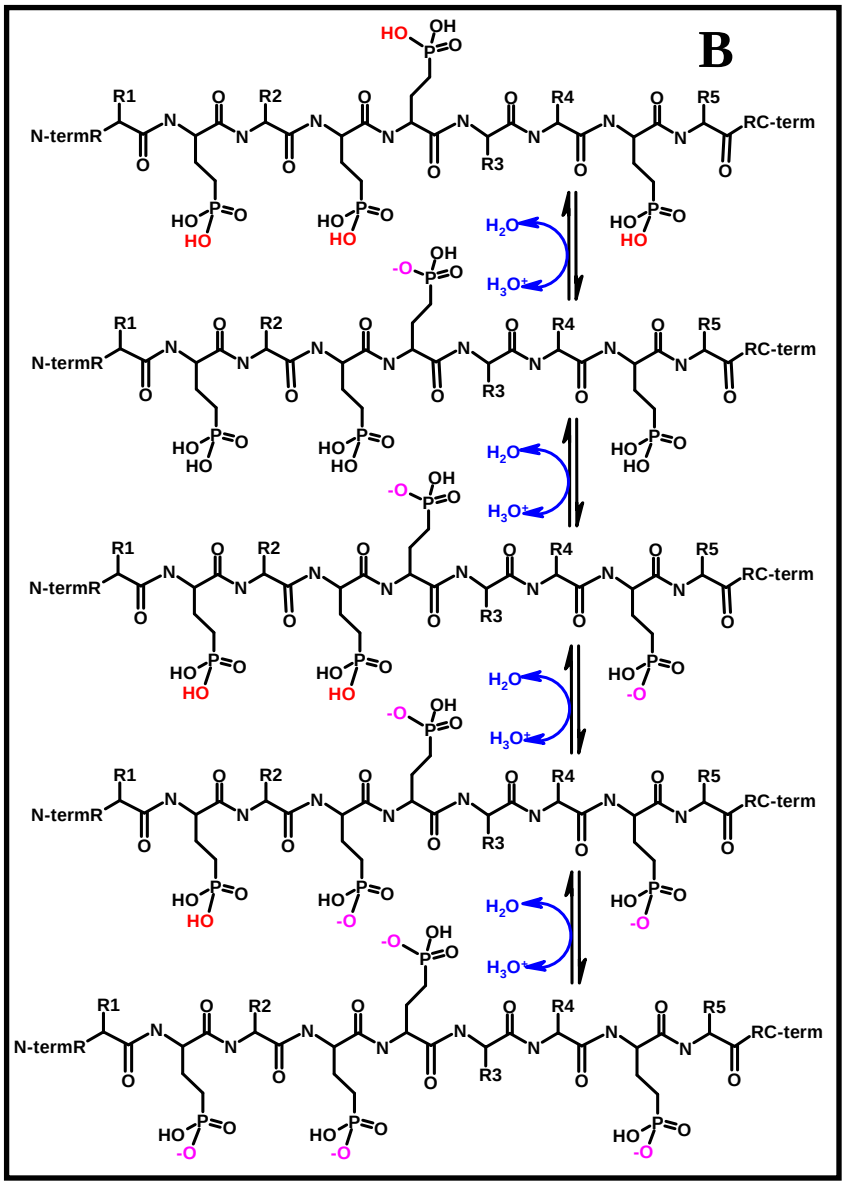


Figure A_07. **Dissociation of Phospho-Moieties in HPLC-Separation.** A. Growing number of phospho-sites on the peptide LPVKPTSSGSSEEK results in broader elution profiles. While unphosphorylated and monophosphorylated peptides elute as sharp peaks, the fully phosphorylated form is distributed over the whole gradient and was selected 19 times for MS/MS analysis. B. Partial dissociation of a fourtimes phosphorylated peptide from uncharged to to fourfold negative charge with only a single charge on each phosphosite. C. A fourtimes phosphorylated peptide has 69 different charge isoforms under standard conditions. (symbols 0 = uncharged, - = singly negative charged, -- = double negative charge.

Mass spectrometry of peptides with multiple phosphorylations

Mass spectrometry of peptides with multiple phosphorylations has two major bottlenecks: First, often the peptides undergo parallel, multiple neutral losses of phosphoric acid, which cannot be triggered individually and suppress formation of sequence specific fragments. Second, due to the low quality of MS/MS spectra, pinpointing of all individual phosphosites is often not achieved.

With regard to the first problem, all techniques applied for detection of mono-phosphopeptides described in 1.B.III can also be applied for the analysis of peptides with

multiple phosphorylations but they require adaptations. Upon phosphatase treatment the number of phospho-moieties, which are removed, is higher, thereby increasing the mass difference between unphosphorylated and phosphorylated peptide. MS^n methods have to apply more fragmentation steps to remove all phosphorylations efficiently, which requires longer analysis time. Therefore, the intensity of MS^n fragment ion spectra is often very low, since other fragmentation processes and neutral loss of water molecules reduce the number of ions that are available for each fragmentation process. This behaviour results in fragment ion spectra with only little sequence information. In pseudo-MS^n spectra often signals of multiple eliminations of phosphoric acid or combination of phosphoric acid and water are observed with high intensity, which impair fragment ion detection in this spectral region. Multiple activation processes of this technique induce formation of fragment ions with varying phosphorylation state, which requires a complex data interpretation.

Most important, higher phosphorylation states reduce the occupation of higher charge states even for large peptides, therefore impairing the application of electron-induced fragmentation methods for exact pinpointing of each phosphorylated amino acid.

An origin of the second challenge, impaired identification of the phosphorylated amino acids, is that excessive phosphorylation affects enzymatic protein cleavage, therefore leading to larger peptide fragments. Next to the traditional handicaps of HPLC-MS analysis of large peptides, such peptides have often more putatively phosphorylated amino acid residues (STY) than phosphorylation sites, which results in a combinatorial complexity. Assignment of a fragment ion spectrum to a distinct isoform of a multiphosphorylated peptide is therefore often impossible for MS^n and pseudo-MS^n fragment ion mass spectra, because not all possible fragment ions are detected. Another problem is the low mass accuracy of ion tap mass spectra which often causes multiple assignments of a single signal to different fragment ions. Most reliable are spectra with intact phosphorylated fragments, which can be achieved by MALDI-MS/MS and electron-induced dissociation methods, since extensive elimination of the phosphate moiety is impaired under both fragmentation conditions. Moreover, both methods predominantly form singly charged fragment ions, thereby reducing the complexity of the obtained tandem MS spectrum.

Results

Part 1 Enrichment of Multiply Phosphorylated Peptides Using Metal Affinity Chromatography on Titanium Dioxide

a Optimization of MALDI MS Detection and Identification

Identification of phosphopeptides by MALDI-MS has been described in some publications, but there is still no protocol that allows reliable detection of such peptides. The initial step in establishing the protocol was to test various, recently described MALDI matrices on their potency to ionize phosphopeptides from a phosphopeptide mixture and from a sample which also included unmodified peptides. Further, crystallization conditions should be optimized with focus on the connection to HPLC separation. In the second step, tandem MS conditions were screened for the individual matrix substances, since identification should be achieved by high-energy fragmentation in the MALDI mass spectrometer.

For optimization of the MALDI-MS protocol I tested two well-known MALDI matrices: α-cyano-4-hydroxycinnamic (CHCA) acid and 2,5-dihydroxybenzoic acid (gentisic acid, DHB), which have been described to give good results also for phosphopeptides. Furthermore, two other substances were described in single studies, which focussed on phosphopeptide identification. These matrices were indolylacrylic acid and 2,5-hydroxyacetophenone. While sample preparation has been widely studied and optimized for CHCA, significantly less is known for the optimal conditions for the other three substances, apart from the fact that they are applied in significantly higher concentrations. For both 2,5-hydroxyacetophenone and DHB, 20 mg/ml have been described recently as optimal concentration, indolylacrylic acid should be used at 5 to 25 mg/ml. In comparison, a CHCA concentration of 2 to 5 mg/ml is suggested for the newest type of MALDI mass spectrometers. Since our studies were carried out on the recently introduced MALDI TOF/TOF mass spectrometer 4800 MALDI TOF/TOF Analyzer (Applied Biosystems, Darmstadt, Germany, now LifeTech) optimization of the crystallization conditions was necessary. The crystal formation and incorporation of the analyte is very important for the MALDI process, while CHCA forms very small crystals, the other matrices aggregate to significantly larger assemblies.

Two phosphopeptide mixtures, which consisted of peptides 1-14 (mix A) from table TA_1 and 15-23 (mix B) were used in 50 nM concentration to test which matrix was able to ionize the phosphopeptides also at low concentrations. To simulate a sample with biolocical background, a tryptic digest of cytochrome C was added to mixture A in an equimolar ratio.

1µl of the sample was crystallized with 1µl matrix. The matrices DHB, indolylacrylic acid, and 2,5-hydroxyacetophenone were prepared in concentrations of 2, 5, 10, 15, 20, and 50 mg/ml, CHCA was used in 2 and 5 mg/ml concentrations. All matrix substances were dissolved in 70% ACN with 0.2% TFA. For MS analysis 4000 shots were acquired per spot. For each matrix, signal intensities of the individual peptide ions were monitored from 10 different spots. Good ionization was only achieved by CHCA and DHB, whereas 2,5-hydroxyacetophenone (mass spectrum not shown) and indolylacrylic acid gave only very poor signal intensities at all matrix concentrations (figure A_08A). The optimization was focussed on DHB and CHCA, since these matrices performed significantly better. Initial experiments with indolylacrylic acid revealed good ionization at higher concentrations of phosphopeptides, but working with this matrix had the severe side effect, that after few laser shots the intensity dropped dramatically due to formation of a polymer and loss of the crystalline properties. Polymerization also impaired automated MS/MS analysis, since fragment ion spectra are acquired from the same sample spots that were used for MS analysis before.

DHB had very good results in the range between 5 and 20 mg/ml and CHCA gave high intensities for both preparations. Low concentrations of DHB (2-5 mg/ml) poorly crystallized on the MALDI sample plate and led to large spot areas with few, sparse crystals. The MALDI spectra of DHB (10 mg/ml) clearly showed higher ionization efficiency for all phosphopeptides in comparison to CHCA (5 mg/ml) (figure A_08B). Moreover only DHB was able to ionize the multiply phosphorylated peptides TwoP_1 and ThreeP_1. Unfortunately, the second triphosphorylated peptide ThreeP_2 was never observed at this low concentrations. Mixtures of both matrices were reported to even improve phosphopeptide detection, but this was not observed for the described phosphopeptide/cytochrome C sample (Laugensen S. and Roepsdorff P., 2003). Based on these data, we selected DHB at 10 mg/ml as optimal concentration for MALDI-MS analysis of multiphosphorylated peptides.

Kjellström and Jensen demonstrated that addition of phosphoric acid to the matrix solution up to 1% v/v increases phosphopeptide ion signals by masking the metal ions on the surface of the stainless steel MALDI target and thereby impairing adsorption of phosphopeptides (Kjellström S. and Jensen O.N., 2004). The composition of the matrix solution was therefore again optimized and contained finally 80% ACN (v/v), 1% TFA (v/v), and 0.2% H_3PO_4 (v/v), since higher concentrations of H_3PO_4 showed clusters of phosphate in the lower mass range.

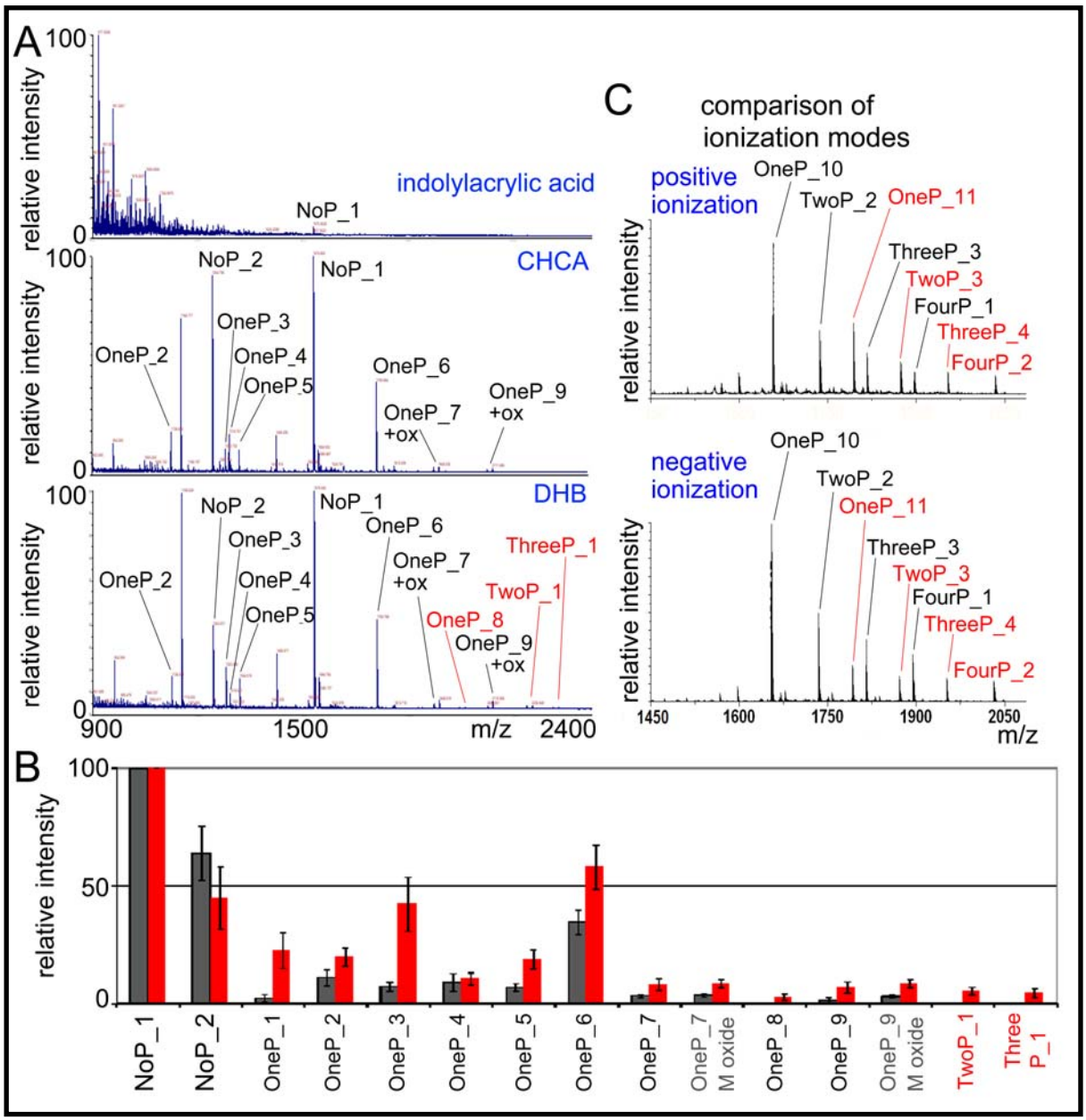


Figure A_08 **MALDI TOF-MS Analysis of Phosphopeptides.** A. Only DHB and CHCA are able to ionize low amounts of phosphopeptides efficiently. However, peptides with multiple phosphosites are suppressed by CHCA (red). B. Individual phosphopeptides have higher signal intensity in DHB crystallized samples (red bars) in comparison to CHCA (gray bars) when using optimized ionization conditions. C. The relative signal intensity of peptides with multiple phosphorylations does not depend on the ionization mode. Singly modified peptides have the highest signal intensity in positive and negative ionization, but amino acids with high proton affinity seem to suppress negative ionization. Signals with multiple negative charges were not observed, even for the fourtimes phosphorylated forms.

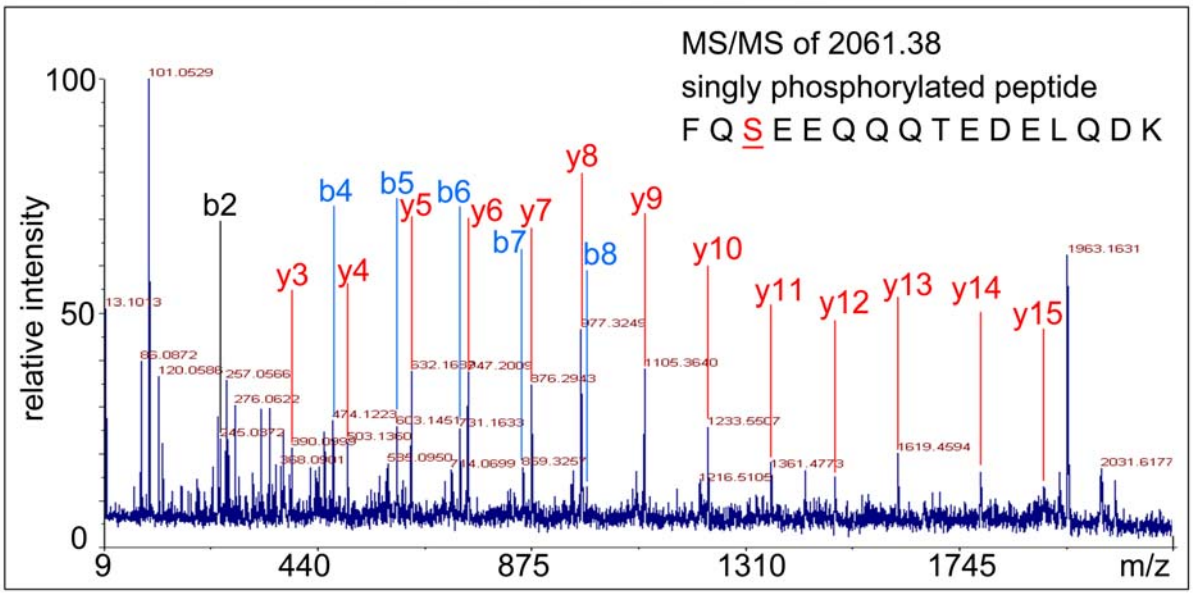
Table TA_1 **Synthetic Phosphopeptides for Optimization of MALDI-TOF/TOF-Analysis of Protein Phosphorylations**. Mix A consist of peptides 1 to 14, mix B contains peptides 15-23.

#	Name	Sequence	[M-H]+
1	**NoP_1**	EGVNDNEEGFFSAR	1570,6769
2	**NoP_2**	APPDNLPSPGGSR	1264,6281
3	**OneP_1**	SV**S**DYEGK	964,3660
4	**OneP_2**	N**S**VEQGRRL	1138,5365
5	**OneP_3**	LIEDNE**Y**TAR	1303,5566
6	**OneP_4**	R**S**DGGHTVLHR	1314,6063
7	**OneP_5**	APPDNLPSPGG**S**R	1344,5944
8	**OneP_6**	SVENLPEAGI**T**HEQR	1759,8011
9	**OneP_7**	ENIMR**S**ENSESQLTSK	1932,8369
10	**OneP_8**	THILLFLPK**S**VSDYEGK	2027,0249
11	**OneP_9**	TASDTDSS**Y**AIPTAGMSPSR	2094,8686
12	**TwoP_1**	QLGEPEK**S**QDSSPVL**S**ELK	2230,9880
13	**ThreeP_1**	QLGEPEK**S**QD**S**SPVL**S**ELK	2310,9544
14	**ThreeP_2**	FG**SS**NTDSAGALG**T**LR	1793,6544
15	**OneP_10**	WWGSGPSGSGG**S**GGGKR	1656.6915
16	**OneP_11**	HWWGSGPSGSGG**S**GGGKR	1793.7504
17	**TwoP_2**	WWGSGPSG**S**GG**S**GGGKR	1736,6578
18	**TwoP_3**	HWWGSGPSG**S**GG**S**GGGKR	1873.7167
19	**ThreeP_3**	WWGSGP**S**G**S**GG**S**GGGKR	1816.6241
20	**ThreeP_4**	HWWGSGP**S**G**S**GG**S**GGGKR	1953.6830
21	**FourP_1**	WWG**S**GP**S**G**S**GG**S**GGGKR	1896,5905
22	**FourP_2**	HWWG**S**GP**S**G**S**GG**S**GGGKR	2033.6494
23	**FourP_3**	RELLNVPGE**S**IVE**S**L**SS**EESITR	2864,1805

As reported by Janek *et al.*, ionization of phosphopeptides is improved in negative ionization mode in comparison to unmodified peptides (Janek K., *et al.*, 2001). Negative ionization was tested with both previously mentioned phosphopeptide mixtures and a sample containing solely peptides 16-22 in 100 nMol concentration. For both standard peptide mixtures, negative ionization mass spectra did not differ dramatically from that acquired in positive ionization mode. All spectra were acquired from the same sample spots. And increased signal intensity for phosphorylated peptides was not obvious.

Interestingly, signal intensities for peptides 16-22 significantly depended on the number of phospho-moieties. In both ionization modes monophosphorylated peptides had the most intense signals, while multiphosphorylated forms were strongly suppressed, although they share the same sequence background. By comparison of mass spectra from positive and negative mode, it was obvious that multiply phosphorylated peptides do not gain signal intensity in negative ionization mode, as demonstrated in Figure A_08C. The s/n ratios for the two least intense signals, FourP_1 and FourP_2 in positive ionization mode and ThreeP_3 and FourP_2 in negative ionization mode, were in average approx. 20-30 for both ionization modes. This result was unexpected, since higher phosphorylation should improve negative ionization. Interestingly, also peptide ions with higher negative charge states were completely absent. In negative ionization the peptide with the longer sequence is significantly suppressed even for the monophosphorylated form. Since both sequences differ only in a single histidine residue, we assume that the high proton affinity of this residue might suppress the formation of negatively charged peptide ions. Similar to the positive ionization mode, PSD fragmentation in negative ionization led to extensive elimination of phosphoric acid and only few peptide fragment ions. The masses of these fragments corresponded to singly negatively charged b- and y-type fragments. Comparison of positive and negative ionization MS/MS spectra showed always higher number of sequence related ions for the positive ionization mode. Since negative ionization showed significant disadvantages in acquisition and interpretation of tandem MS spectra, we optimized the MALDI protocol for positive ionization.

The second important aspect of the project was fragmentation of the phosphopeptide for identification of the amino acid sequence and localization of the phosphate moiety. For this purpose, tandem MS methods were optimized with focus on fragment ion intensity and s/n ratio for CHCA and DHB matrix preparations. Both matrices performed similarly well, whereas signal intensities were slightly higher for DHB. Neutral loss of H_3PO_4 was observed in almost all phosphopeptide spectra, except for tyrosine phosphorylated peptides. Interestingly, the number of phosphorylated Ser and Thr residues could be estimated by the neutral loss ions, since multiple losses of phosphoric acid were also observed. Figure A_09 shows an example spectrum for the Ser-phosphorylated peptide FQ**S**EEQQQTEDELQDK from β-casein (aa 48-63) after tryptic digestion, which was crystallized with 10 mg/ml DHB. Interestingly, the b-ion series completely lost the phosphorylation due to β-elimination, while the y-ions exhibited the intact phospho-amino acid on fragments y 14 and y15.



A_09 **MS/MS Spectrum of the Singly Phosphorylated Peptide (aa 33–48) from Tryptically Digested β-casein.** While the b-ion series was observed in dephosphorylated form (-98Da for b4-b8 blue), y ions (red) 14 and 15 had intact phospho-moieties.

Since spotting should be automated and connected to a RP-HPLC system, the matrix composition was also optimized with regard to these special conditions. Therefore the matrix should be mixed with the column eluent in a micro-T flow path before deposition on the sample carrier and it should crystallize rapidly in small spots after deposition. We optimized the matrix flow rate for different time steps, since the spotting device was operated to fractionate the sample every 4-20 seconds. For this purpose, we evaluated crystallization by eye and tested the MS ionization efficiency for different conditions. Again the matrix solution with 10 mg/ml DHB dissolved in 80% ACN (v/v), 1%TFA (v/v), and 0.1% H_3PO_4 (v/v) gave excellent results and was further used to prepare the LC-MALDI MS samples.

b Development of an Enrichment Procedure

As MALDI MS spectra demonstrated, ionization of multiply phosphorylated peptides is dramatically reduced in the presence of unmodified peptides and peptides with one phospho-moiety. Based on the protocols by Mazanek and Larsen, we optimized an offline TiO_2 enrichment procedure that allows analysis of multiphosphorylated peptides. The workflow is schematically drawn in figure A_10 (Mazanek M. et al., 2007; Larsen M.R. *et al.*, 2005). TiO_2 enrichment was performed on 5µm TiO_2 microspheres from Glygen (Columbia, USA) which were packed into top tips.

Our data demonstrate that sample loading onto TiO_2 with a loading buffer which consists of 200 mg/mL DHB, 35 mM OcSNa (octylsulfonate sodium salt), 23% acetic acid (AcOH), 40% ACN and 0.1 % heptafluorobutyric acid (HFBA) showed the best removal of unmodified peptides and is compatible with MALDI ionization. Peptide samples were first diluted with 50% ACN, containing 10% AcOH, to ensure a high amount of organic solvent, to promote phosphopeptide binding to TiO_2 and in a second step with 300 mg/mL DHB, 50 mM OcSNa in 40% AcOH, 40% ACN, 0.2% HBFA in a 1:2 v/v (sample/buffer) ratio. This mixture was rinsed through the TiO_2 tip slowly over a time span of 15-30 min to allow efficient adsorption of phosphopeptides. Unmodified peptides were removed by washing with 200 mg/mL DHB, 35 mM OcSNa, 23% AcOH, 40% ACN and 0.1 % HFBA and additionally with 100 mL of 70% ACN containing 0.2% HFBA to remove components of the loading buffer, which might interfere with the adjacent LC-MS analysis.

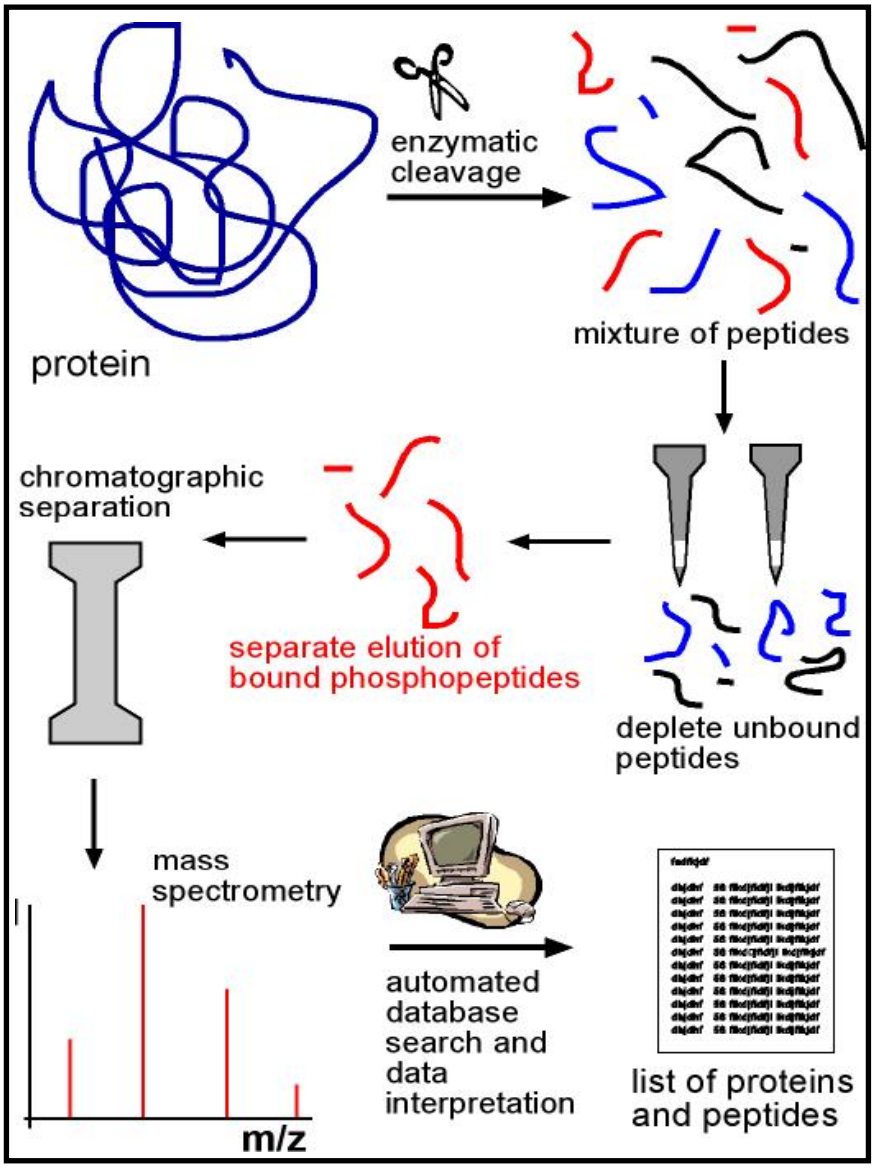


Figure A_10 **Strategy for Phosphopeptide Identification by Off-line TiO_2-MOAC Enrichment and Subsequent HPLC-MS/MS Analysis.**

Effective elution of peptides with multiple phosphorylations could not be achieved with ammonia alone, therefore we used 50 mM $(NH_4)_2HPO_4$ (pH 11 with ammonia), which introduces phosphate as competitive binder to TiO_2. Elution time and volume of elution buffer were optimized in several experiments, resulting in a minimal volume of 25µl of buffer to release the phosphopeptides in a minimal time span of 10 min.

Since previous publications mentioned irreversible binding of peptides with multiple phosphosites to the TiO_2 surface, we tested the recovery of these peptides by quantitative mass spectrometry. For this purpose we labelled the peptides OneP_5, OneP_8, OneP_9, TwoP_1, TwoP_3, ThreeP_1, ThreeP_4, and FourP_1 with iTRAQ (isobaric tag for relative and absolute quantification) reagents 114 and 115. Peptides with the 115 label were enriched on TiO_2 and afterwards the same amount of iTRAQ-114 labelled peptides was added to the eluted phosphopeptides. MS/MS experiments were performed for all peptides and the intensity of the respective reporter ions was extracted from the raw data. Tandem MS analysis of the identified peptides demonstrates almost quantitative recovery for most of the phosphopeptides (Fig A_11A). Peptides with very hydrophobic nature, for instance peptide OneP_8 (THILLFLPK**S**VSDYEGK, peptide C), have a higher solubility in the loading solvent and might be lost partially in the loading and washing steps.

The efficiency of this method is demonstrated in figure A_11B which shows the mass spectra for two phosphoprotein digests before (spectra AI and BI) and after applying the enrichment protocol (spectra AII and BII). 2 pMol tryptically digested β-casein (A) and osteopontin (B) were treated according to the protocol and a fraction of 300 fMol was crystallized per spot on a MALDI target. Osteopontin is a secreted protein which is involved in bone cell differentiation and it has many phosphorylation sites spread over the whole protein sequence (Sørensen E.S. et al., 1995). The majority of the phosphorylation sites are applied by the casein kinase 2 and have the respective recognition sequence with an acidic amino acid two residues C-terminal from the observed phosphorylation site.

Although the same amount of the untreated protein digest was spotted, mass spectra before and after enrichment were strikingly different for both proteins. Apart from the monophosphorylated peptide FC**S**EEQQQTEDELQDK (amino acids 33–48, *m*/*z* 2061.83) which was just above detection limit, no phosphopeptides were identified in the untreated protein digests (AI and BI). After TiO_2 enrichment (AII) this peptide became base peak of the β-casein sample and the fourtimes phosphorylated, N-terminal peptide RELEELNVP-GEIVE**S**L**SSS**EESITR (aa 1-25, *m*/*z* 3122.26) was detected with more than 30% relative intensity. This sample also demonstrated that trypsin rarely cleaves the N-terminal arginine

residue of β-casein as demonstrated by the low intensity of the peptide 2-25 (*m*/z 2966.16), which is just above detection limit.

A similar enrichment effect was observed for osteopontin where only two peptides with multiple sodium adducts were detected in the untreated sample. Extraction of phosphopeptides from this sample leads to the detection of at least 10 multiphosphorylated peptides (Fig A-11B_BII), which are formed upon tryptic cleavage. The most abundant signal stems from the triply phosphorylated C-terminal peptide IRI**S**HELD**S**A**S**SEVN at 1896.72 m/z. Most of the observed peptide signals belong to phosphorylated peptides of osteopontin, as subsequent MS/MS analysis and comparison with theoretical cleavage products reveals.

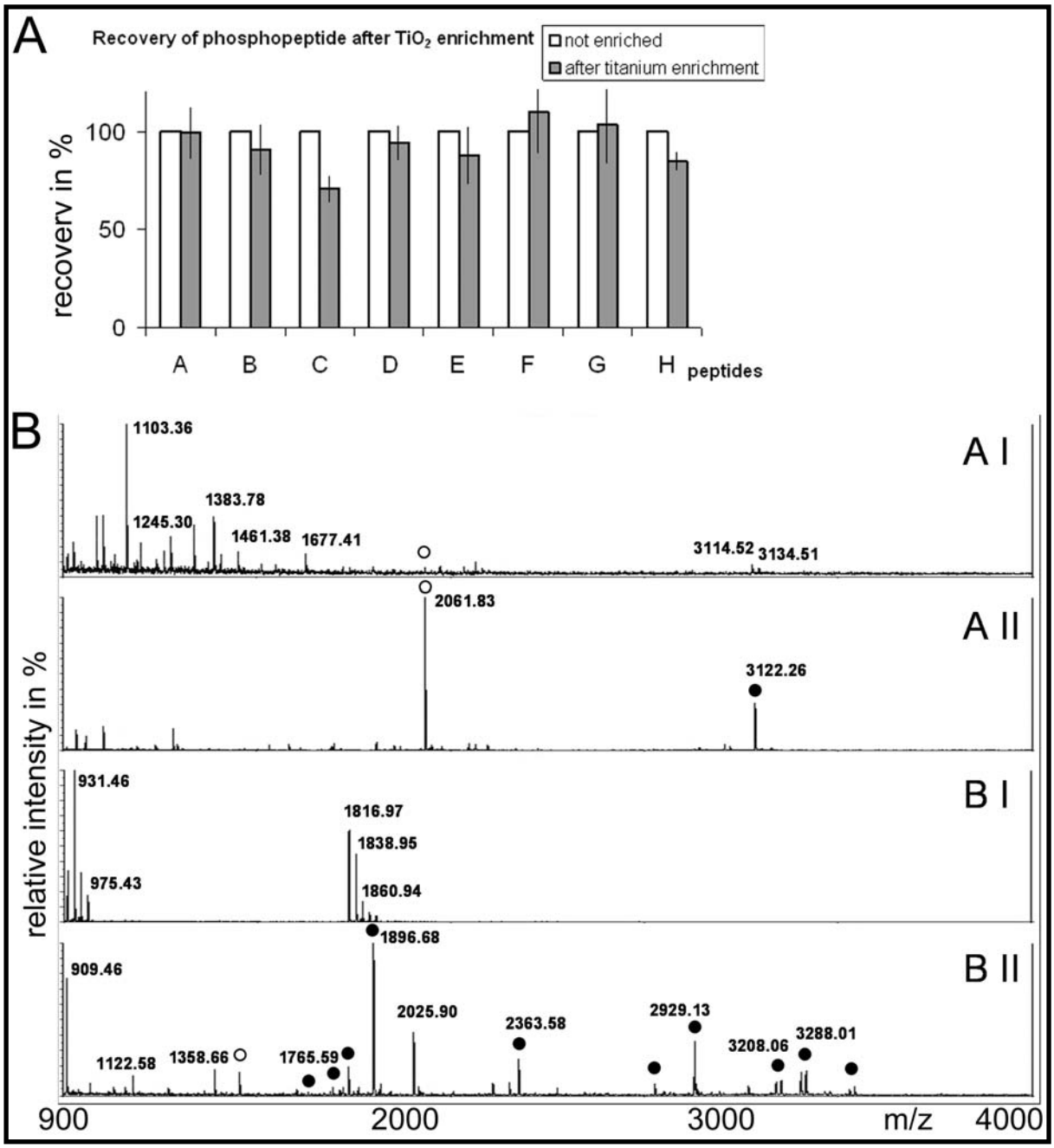


Figure A_11 **Results of Optimized TiO_2-Enrichment Procedure for MALDI-MS/MS Analysis**. A. TiO2 enrichment provides high recovery rates for phosphopeptides with different number of phosphorylations (OneP_5 (A), OneP_9 (B), OneP_8 (C), TwoP_1 (D), TwoP_3 (E), ThreeP_1 (F) ThreeP_4 (G) FourP_1 (H)). Hydrophobic peptides are partially soluble under washing conditions, resulting in lower recovery as demonstrated by peptide C. B. Phosphopeptides with various number of phosphorylation sites are efficiently enriched from tryptic digests of β-casein (A) and osteopontin (B) by the TiO_2 procedure, thereby allowing reliable detection of various phosphorylation products. ○ monophosphorylated, ● multiphosphorylated; AI, BI digests before enrichment; AII, BII after enrichment

c Reversed-Phase Separation of Multiple Phosphorylated Peptides

As demonstrated in figure A_7A, one major problem of phosphopeptide analysis results from the retention behaviour in standard RP chromatography. To determine the influence of unspecific binding to hydrophilic surfaces during chromatographic fractionation, we first added phosphoric acid instead of TFA or formic acid to the HPLC solvents. This addition should impair interaction with metal ions in the silica material. Since our protocol is designed for MALDI-MS detection, low amounts of phosphoric acid do not interfere with the ionization process. Other reports mention even enhanced ionization of phosphopeptides by addition of phosphoric acid and suppression of matrix clusters. In contrast, phosphate is able to form large cluster ions during ESI ionization, thereby resulting in huge signals in the lower m/z range. In comparison to formic acid and TFA, phosphoric acid has a lower UV absorption at the wavelength of 214 nm. Therefore, detection of phosphopeptides during LC separation is even possible at low concentrations.

To test the influence of the acidic modifier, mix_A and a fraction of TiO_2 purified phosphopeptides from osteopontin were HPLC separated on a 25 cm C18 column (Dionex, PepMapC18, 25cmx75µmx3µm, 100A) using solvents with TFA and phosphoric acid, respectively. Individual fractions were placed onto the MALDI sample plate together with DHB matrix every 15 seconds. Signal intensity and peak shape of phosphopeptides were determined from MALDI-MS spectra. Upon phosphoric acid addition, elution profiles of phosphopeptides improved slightly in comparison to TFA-containing buffers, but the signal intensity was quite low as demonstrated in figure A_12A. Interestingly, peptides with higher phosphorylation states have a stronger hydrophobic nature under these separation conditions than peptides with single or double phosphorylations, since multiphosphorylated peptides are

detected in later fractions. Introduction of phosphoric acid improved the chromatographic separation, but could not solve the charge-related distribution over a broad elution window. Peak broadening becomes more severe with increased gradient times, which are often used to achieve better chromatographic separation.

In comparison to nano-scale columns with silica stationary phase, the earlier mentioned monolithic columns require significantly less analysis time to separate peptide mixtures (Schley C. *et al.*, 2006; Yoo C. *et al.*, 2007). Unfortunately these columns are not able to fractionate samples with high complexity such as digests of whole cells, due to the reduced resolving power. Furthermore, they have only limited capacity, which interferes with trace analysis of complex samples. To optimize separation on monolithic columns, a mixture of synthetic phospho-peptides (see table TA_1, peptides 1-14) each 500 fMol was separated using different gradient conditions. Thereby, the sample was directly injected onto the separation column. A monolithic trapping column was not used, because the hydrophobic nature of the styrole-divinylbenzole monolith is significantly weaker than that of C18-resins, so that hydrophilic peptides might be lost (Schley C. *et al.*, 2006). For comparison with the C18 column, equal amounts of mix_A and osteopontin digest were also applied to monolithic separation. We observed sharp elution profiles for all phosphopeptides as demonstrated in figure A12_B. Moreover, monolithic separation resulted in dramatically increased peptide intensities for almost all observed phosphopeptides. Especially phosphopeptides of the osteopontin digest (lower traces) exhibited strong signals and amino acid sequence and phosphorylation sites were identified with high reliability from the obtained tandem MS spectra. Interestingly, if a peptide sequence exists in different phosphorylation states, first eluting the form with the highest number of phospho-moieties, while monophosphorylated and unphosphorylated forms are stronger retained.

Although monolithic separation leads to improved detection of highly phosphorylated peptides, a further problem, separation of regio-isomeres, where phospho-moieties are located on different amino acids, was seldomly observed. Due to the timely-limited chromatography and the rather large fractions applied by the spotting device, regio-isomeric, multiply phosphorylated peptides, which differ only in the localization of a single phosphorylation are not separated. Therefore, such peptides cannot be distinguished reliably and might result in mixed fragment ion spectra that contain fragments of both species.

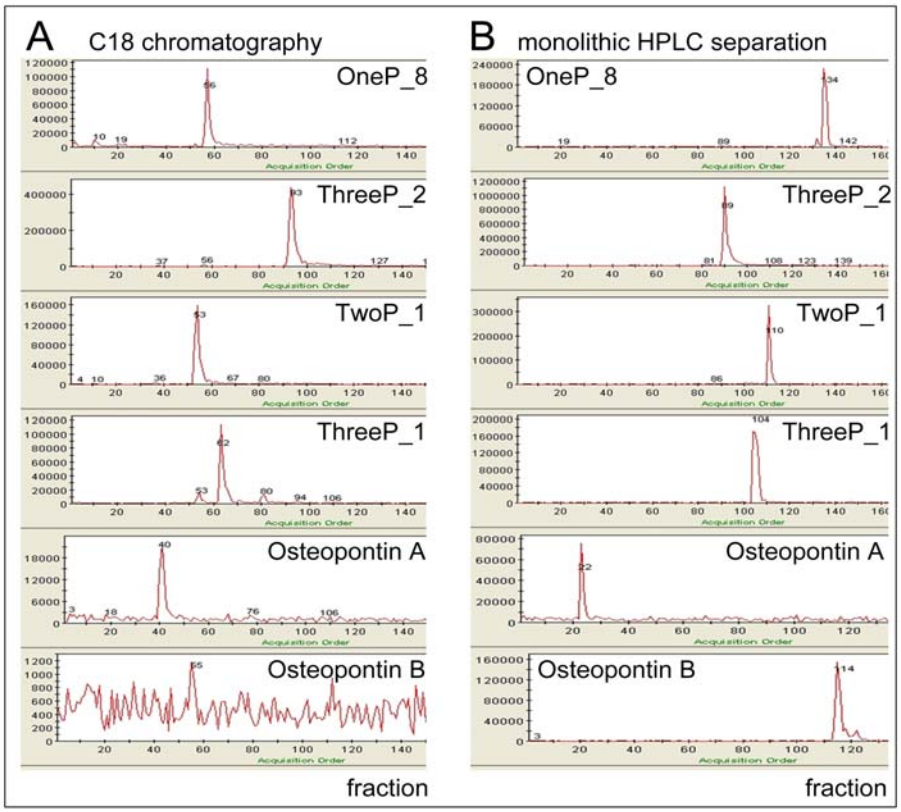


Figure A_12 **Optimized Chromatographic Conditions for Phosphopeptide Separation on RP Columns.** A. Replacing the acidic solvent modifier by phosphoric acid improves phosphopeptide elution, but peptides with multiple phosphorylations are still strongly retained. B. Observed phosphopeptide signals are significantly higher after monolithic chromatography and highly phosphorylated peptide forms elute first. Some peptides which could not be detected by C18-RPC were identified by separation on monolithic columns.

d Comparison to other methods for phosphopeptide analysis

Figure A_09 shows that MALDI TOF/TOF tandem MS is able to produce sequence specific fragment ions, which allow both identification of the peptide sequence and localization of the phosphorylated amino acid. The low intensity of these fragments is mainly caused by extensive loss of H_3PO_4 from the phosphorylated peptide, which often represented the dominant ion in the corresponding fragment mass spectra of peptides with Ser or Thr phosphorylation. This fragmentation behaviour was also observed for peptides with multiple phosphorylations, but significantly more neutral loss species occurred in the upper mass region, because of multiple eliminations of H_3PO_4. Furthermore, under the applied fragmentation conditions not only β-elimination mechanisms ($-H_3PO_4$, -98 Da) played a role, but also the loss of HPO_3 (-80 Da) could be observed with relatively high intensity. Signals, stemming from multiple eliminations of phospho-moieties, often occupied the complete high mass region of fragment ion spectra of highly phosphorylated peptides and could be used to estimate the minimal number of phosphorylations. Here, also combinations of H_3PO_4 and HPO_3 loss were observed. In case of Tyr-phosphorylation, neutral loss signals have significantly lower intensity compared to other fragment ions and did not interfere with peptide sequence identification.

Another important observation was the formation of fragment ions which had lost the phospho-moiety. These fragment ions can further help to confirm the amino acid sequence but their intensity varied strongly with the amino acid composition of the peptide.

Optimization of database searching parameters and comparison to standard HPLC-ESI-pseudoMS^n analysis

In contrast to ESI-MS/MS analysis, fragmentation of singly charged peptides in MALDI mass spectrometry locates peptide ions, which result from neutral loss of H_3PO_4, to the upper mass range, while most backbone fragments are found at lower masses. To improve data interpretation, we tested, if automatic interpretation of these spectra is improved by reduction of signal intensities of neutral loss peaks by a factor of 100 or complete deletion of these signals from the peak list. Neutral loss signals yield no sequence information, so database search algorithms are not able to identify them correctly. Since these signals cannot be explained by standard peptide fragmentation rules, they reduce the overall peptide score by not matching to the proposed amino acid sequence. Extracted MS/MS data from synthetic phosphopeptides were searched with altered NL-ion intensities and the corresponding

MASCOT scores were compared. As table TA_2 shows, reduction of intensities or even deletion of the neutral loss ions significantly enhanced sequence annotation by MASCOT (www.matrixscience.org). Application of this data interpretation step into the analysis of phosphopeptides from β-casein and osteopontin, resulted in significantly increased MASCOT scores for all observed peptides.

amount	sequence	Mascot Score		
		A	B	C
1 pmol	KFL**S**LASNPELLNLPS**S**VIK	54	56	61
0,1 pmol	KFL**S**LASNPELLNLPS**S**VIK	50	53	58
1 pmol	WWGSGPSG**S**GG**S**GGGKR	83	89	98
0,1 pmol	WWGSGPSG**S**GG**S**GGGKR	46	50	58
1 pmol	FG**SS**NTDSAGALG**T**LR	79	84	97
0,1 pmol	FG**SS**NTDSAGALG**T**LR	69	75	82
1 pmol	WWG**S**GP**S**G**S**GG**S**GGGKR	71	71	75
0,1 pmol	WWG**S**GP**S**G**S**GG**S**GGGKR	64	65	71
1 pmol	HWWG**S**GP**S**G**S**GG**S**GGGKR	78	87	115
0.1 pmol	HWWG**S**GP**S**G**S**GG**S**GGGKR	75	80	92
1 pmol	RELLNVPGE**S**IVE**S**L**SS**EESITR	32	36	43
0,1 pmol	RELLNVPGE**S**IVE**S**L**SS**EESITR	29	33	38

Table TA_2

Deletion of Neutral Loss Peaks Increases the MASCOT Score.

High signals from neutral loss of H_3PO_4 in the upper mass area impair phosphopeptide identification by database search algorithms. Reduction of this peak area by factor 100 increases the score slightly (Column B), complete deletion gives the best identification scores (Column C).

To test the efficiency of the new protocol, we performed a comparison to the standard phosphopeptide analysis method in our laboratory, which is carried out by nanoC18RP-HPLC/nanoESI pseudo-MS^n. Therefore, either 1 or 3 pMol tryptic digest of osteopontin were enriched for phosphopeptides on TiO_2 and 1/3 of the sample (300 fMol and 1 pMol, respectively) was injected on each separation system, nanoC18RP-HPLC or monolithic HPLC with optimized buffer conditions. MALDI MS/MS identified the same number of phosphorylation sites (see figure A_13), but required only one third of the sample amount injected onto the separation system. All phosphopeptides identified by MALDI were from 300 fMol osteopontin digest, while ESI MS/MS experiments at this concentration did not identify peptides with multiple phosphorylations. Nevertheless, some singly phosphorylated peptides were observed, which corresponded to peptide sequences that were expected in higher

phosphorylation states. Therefore, 1 pMol of osteopontin was analyzed by "in house" standard protocol.

MALDI data delivered one additional fourtimes phosphorylated peptide (LPVKPT**SS**G**SS**EEK) of which only the doubly modified form was identified by the ESI method. One doubly phosphorylated peptide was additionally found in the ESI dataset which could not be detected by MALDI MS/MS. Both methods identified 15 out of 28 proposed phosphorylation sites of osteopontin, but many putative phospho-amino acids could not be identified due to the peptide size produced by tryptic cleavage. The amino acid sequence between S62 and R136 does not have a Lys or Arg residue, so that this region with 8 proposed phosphorylation sites could not be cut down to smaller peptides. For ESI data, peptides with MASCOT scores above 20 were inspected and accepted, MALDI results were manually validated for scores above 30.

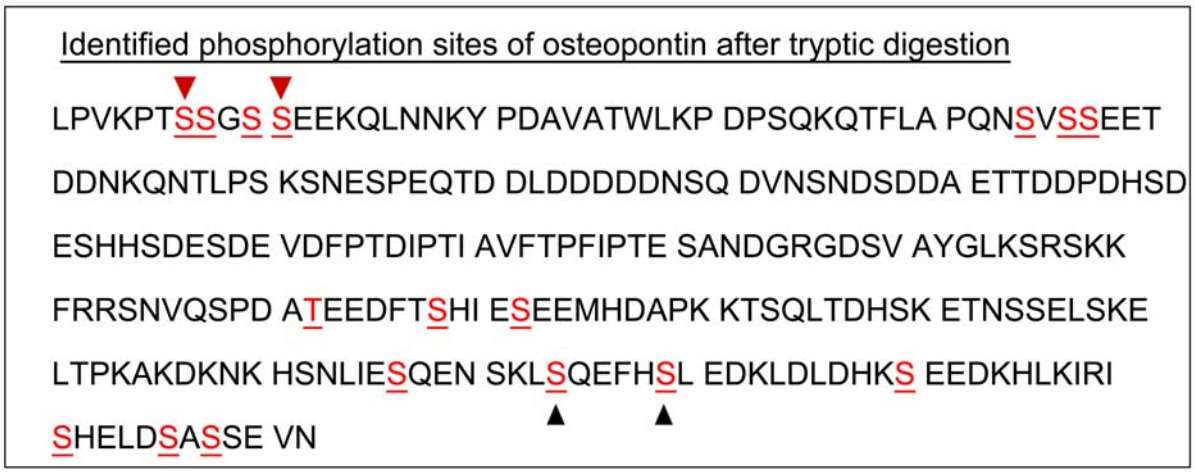


Figure A_13 **Identified Phosphorylation Sites of Osteopontin After Tryptic Cleavage.** Phosphorylated amino acids in red were identified by both types of analysis, MALDI-TOF/TOF and ESI-CID pseudoMS^n analysis. Phosphorylation sites with black triangles were exclusively found by the ESI method, while MALDI was able to identify the fourtimes phosphorylated N-terminus which was found only doubly modified in ESI.

Further, we compared the localization of phospho-moieties within the highest scoring peptide sequence with the proposed phospho-amino acids on the swissprot database entry (Sørensen E.S. et al., 1995). Distinct localization of a phospho-site is indicated by the difference of the MASCOT scores for individual isoforms of a peptide, where larger differences stand for higher localization confidence. It was obvious, that MALDI data not only resulted in higher peptide scores, but at the same time allowed more accurate identification of the phospho-sites, if multiple Ser, Thr, and Tyr residues are present (Schmidt

A. *et al.*, 2008). As demonstrated in table TA_3, most isoforms of the phosphopeptides assigned in the MALDI dataset have the expected phospho-site distribution, while ESI data often point to other distributions of phospho-sites within the peptide sequence. The correct phosphopeptide sequence (high-lighted in gray) received highest scores from MALDI data and was often unequivocally distinguished from all other putative phospho-isoforms. One explanation for the poor performance of the ESI tandem MS protocol is the bigger mass difference used for data interpretation of MS/MS spectra, thereby causing false positive annotation of fragment ions from noisy spectra. Interestingly, many phosphopeptides which were additionally identified from the ESI dataset missed one or two phosphorylations compared to the database entry.

Phosphosite distribution		MASCOT Score	
	Sequence A	**ESI**	**MALDI**
A-1	LPVKPT**S**S**G**S**S**EEK	5.9	50.5
A-2	LPVKP**T**S**S**G**S**S**E**EK	17.8	45.0
A-3	LPVKP**T**S**S**G**S**S**E**EK	11.4	32.3
A-4	LPVKP**T**S**S**G**S**S**E**EK	8.8	29.2
A-5	LPVKP**T**S**S**G**S**S**E**EK	8.8	27.5
	Sequence B	**ESI**	**MALDI**
B-1	IRI**S**HELD**S**A**S**SEVN	16.4	54.2
B-2	IRI**S**HELD**S**AS**S**EVN	10.2	44.1
B-3	IRI**S**HELDSA**S**SEVN	9.3	35.1
B-4	IRISHELD**S**A**S**SEVN	3.2	22.1
	Sequence C	**ESI**	**MALDI**
C-1	SNVQSPDA**T**EEDFT**S**HIE**S**EEMHDAPK	34.4	70.8
C-2	SNVQ**S**PDATEEDFT**S**HIE**S**EEMHDAPK	15.7	50
C-3	**S**NVQSPDATEEDFT**S**HIE**S**EEMHDAPK	6.5	50
C-4	SNVQSPDATEEDF**TS**HIE**S**EEMHDAPK	20.8	46.4
C-5	SNVQSPDA**T**EEDF**T**SHIE**S**EEMHDAPK	34.4	45
C-6	SNVQ**S**PDATEEDF**T**SHIE**S**EEMHDAPK	10.3	39.6
C-7	**S**NVQSPDATEEDF**T**SHIE**S**EEMHDAPK	4.4	39.6
C-8	SNVQSPDA**T**EEDF**TS**HIESEEMHDAPK	19.2	22.5
C-9	SNVQ**S**PDA**T**EEDFTSHIE**S**EEMHDAPK	2.7	22.2
C-10	**S**NVQSPDA**T**EEDFTSHIE**S**EEMHDAPK	n.a.	22.2
C11	SNVQ**S**PDATEEDF**TS**HIESEEMHDAPK	6.1	n.a.

Table TA_3 **MALDI Data Allow Localization of Phosphorylation Sites with Higher Confidence.** Phosphorylated amino acids of osteopontin are localized to the expected phosphosites of the database entry only from the MALDI data. ESI-MS/MS data point often to different distributions of the phosphosites within the peptide sequence.

Comparison to Electron-Transfer-Dissociation

Recently, novel tandem MS fragmentation techniques boosted phosphopeptide research and one of the most promising techniques is electron-transfer dissociation (ETD). Fragment ions (c- and z-type ions) formed by this process result from a cleavage of the N-Cα bond of the amide backbone and are therefore complementary to the b-and y-ion series of CID fragmentation (Syka J.E. *et al.*, 2003). Moreover, the fragmentation process is not due to an increase of the internal energy of the molecule, but involves radical rearrangements which leave fragment ions with intact phosphoester modifications. Upon electron uptake, the charge of the peptide is reduced by one which limits the application of this technique to peptide ions with charge states of 3+ or higher. Nevertheless, the implementation into ion trapping instruments makes this technique a sensitive and powerful tool for phosphopeptide research.

To compare both fragmentation techniques we prepared the peptide HWWG**S**GP**S**G**S**GG**S**GGGKR (FourP_1) with four phosphoserines in the sequence and crystallized 100 fMol with DHB matrix for the MALDI MS/MS experiment. 8000–12000 subspectra were summed up for the MS/MS spectrum and peaks were exported using the PEAKStoMASCOT tool from the 4000 Explorer Software (Applied Biosystems). The upper mass region starting from 1720 m/z, where exclusively neutral loss ions were observed, was removed from the mgf-file prior to database search. For ESI-ETD fragmentation a 2.5 µM solution was directly injected into the LTQ ETD mass spectrometer (Thermo-Fisher, Bremen). ETD fragmentation was performed on the doubly and triply charged precursor with the optimized activation time of 90 ms.

As demonstrated by figure A_14 both methods deliver high sequence coverage and provide identification of all phospho-sites by the respective fragment ions. The main signal of the MALDI spectrum stems from the elimination of one phosphoric acid molecule and neutral losses are also observed in the upper mass region. The localization of the positive charge in MALDI MS/MS leads mainly to intense y-ions, which can be observed from y1-y16 with a maximal mass error of 0.2 Da. The only b-ion which could be identified was b2.

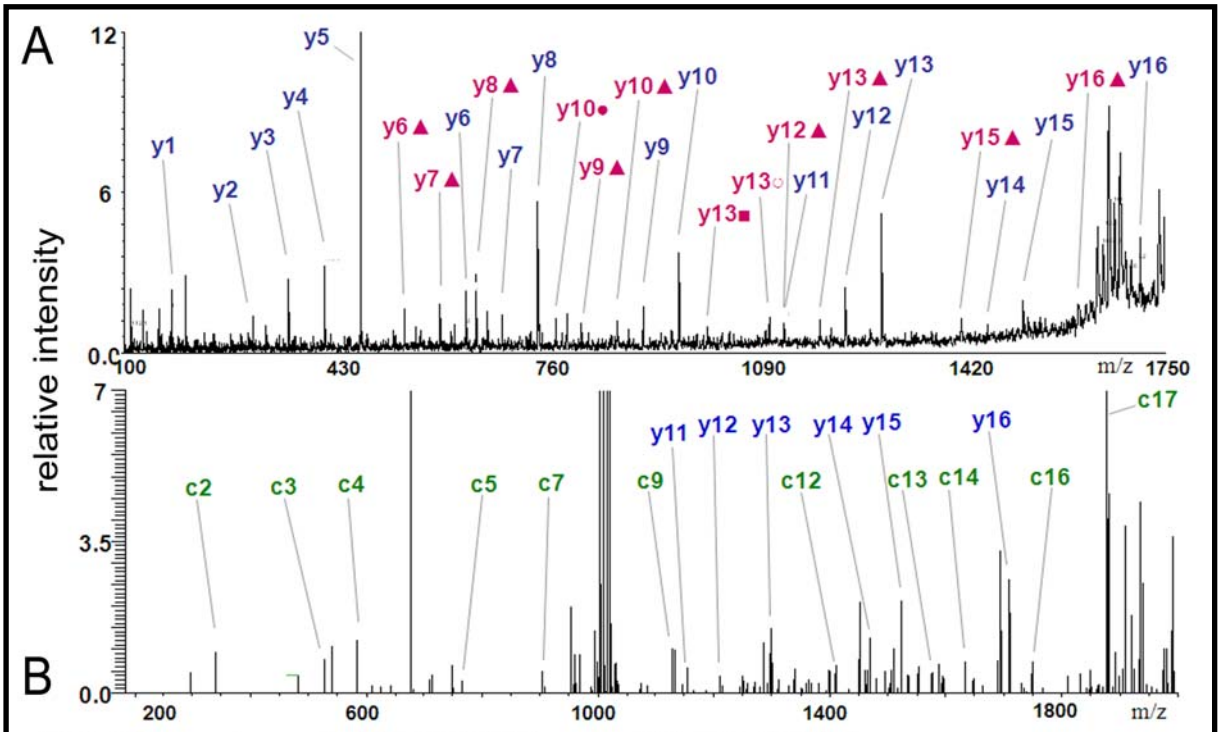
ETD MS/MS data were obtained by fragmentation of the triply charged precursor ion ([M+3H]3+), since the doubly charged precursor delivered almost no fragment ions. Interestingly, upon ETD activation sequence related fragment ions were found at lower intensities (max. 7%) compared to MALDI (12 %). The base peak of this analysis was the doubly charged radical peptide ion, which is formed upon electron uptake and a large fraction of the precursor ion was also present. This fragmentation did not result in a dominant ion

series like MALDI-MS/MS, but the identified c- and z-type fragment ions formed a complementary series so that all amino acid residues were covered by fragment ions.

Nevertheless, identification of all fragments required a mass window of 0.8 Da which is significantly larger than that of the MALDI spectrum (0.2 Da). Overall, both methods were able to map the complete peptide sequence and identify all phosphorylation sites. MALDI MS/MS has the higher mass accuracy, lower noise level and parallel fragment ion series which are formed upon neutral loss of a modification, this makes the MALDI analysis at least equal to ETD fragmentation, for the analysis of multiply phosphorylated peptides. Higher scores from database searching tools were observed for MALDI data which is due to the long y-ion series which receives an extra-score since the sequence is confirmed much better than by dispersed c- and z-fragments.

Next page:

Figure A_14 **Comparison of MALDI-TOF/TOF MS/MS and ESI-ETD MS/MS Analysis.** The fourtimes phosphorylated peptide HWWGSGPSGSGGSGGGKR was fragmented by both methods. A. Fragment ion spectra of both fragmentation principles result in high sequence coverage and all phosphorylation sites are covered by the identified fragment ions. The neutral loss region above 1750 m/z was truncated in the MALDI spectrum, because there were only peptide ion signals which resulted from neutral loss of phospho-moieties. MALDI fragmentation results in a second fragment ion series with a neutral loss of H_3PO_4. Both spectra are displayed with expanded view (MALDI 12% relative intensity, ETD 7%) on the identified fragment ions, since fragment ions were identified at low intensity. Fragment ions of the MALDI-TOF/TOF MS/MS spectrum fall into a 0.2 Da window, whereas 0.8 Da have to be applied to identify fragment signals of the ETD spectrum (symbols indicate loss of phosphoric acid: ▲ 1x, ●○ 2x, ■ 3x) B. MASCOT identification results for both fragmentation experiments and table of annotated fragment ions (red). Interestingly, many y-ions (blue) are observed in the ETD spectrum which were not expected from the proposed fragmentation mechanism.



MASCOT search results MALDI MS/MS
Ions Score: 92 Expect: 1.6e-10

#	Immon.	b	Seq.	y	y-80	y-98	#
1	110.1	138.1	H				18
2	159.1	324.1	W	1896.6	1816.6	1798.6	17
3	159.1	510.2	W	1710.5	1630.5	1612.5	16
4	30.0	567.2	G	1524.4	1444.4	1426.4	15
5	140.0	734.2	S	1467.4	1387.4	1369.4	14
6	30.0	791.3	G	1300.4	1220.4	1202.4	13
7	70.1	888.3	P	1243.4	1163.3	1145.3	12
8	140.0	1055.3	S	1146.3	1066.3	1048.3	11
9	30.0	1112.3	G	979.3	899.3	881.3	10
10	140.0	1279.3	S	922.3	842.3	824.3	9
11	30.0	1336.4	G	755.3	675.3	657.3	8
12	30.0	1393.4	G	698.3	618.3	600.3	7
13	140.0	1560.4	S	641.3	561.3	543.3	6
14	30.0	1617.4	G	474.3			5
15	30.0	1674.4	G	417.3			4
16	30.0	1731.4	G	360.2			3
17	101.1	1859.5	K	303.2			2
18	129.1		R	175.1			1

ESI-ETD MS/MS
Ions Score: 69 Expect: 1.5e-06

#	c	Seq.	y	z+1	z+2	z+2++	#
1	155.1	H					18
2	341.2	W	1896.6	1880.6	1881.6	941.3	17
3	527.3	W	1710.5	1694.5	1695.5	848.3	16
4	584.3	G	1524.4	1508.4	1509.4	755.2	15
5	751.3	S	1467.4	1451.4	1452.4	726.7	14
6	808.3	G	1300.4	1284.4	1285.4	643.2	13
7	905.4	P	1243.4	1227.4	1228.4	614.7	12
8	1072.3	S	1146.3	1130.3	1131.3	566.2	11
9	1129.4	G	979.3	963.3	964.3	482.7	10
10	1296.4	S	922.3	906.3	907.3	454.2	9
11	1353.4	G	755.3	739.3	740.3	370.7	8
12	1410.4	G	698.3	682.3	683.3	342.2	7
13	1577.4	S	641.3	625.3	626.3	313.6	6
14	1634.4	G	474.3	458.3	459.3	230.1	5
15	1691.5	G	417.3	401.2	402.3	201.6	4
16	1748.5	G	360.2	344.2	345.2	173.1	3
17	1876.6	K	303.2	287.2	288.2	144.6	2
18		R	175.1	159.1	160.1	80.6	1

Discussion – Part 1

Despite technological and methodical novelties for LC-MS/MS-based phosphoproteomics increased the number of known phosphorylation sites, multi-phosphorylated peptides are still underrepresented. I optimized a protocol to overcome difficulties of their analysis, caused by their special physico-chemical properties and fragmentation behaviour. For instance, poor HPLC separation of such peptides under standard conditions was often underestimated in many phosphoproteomic studies. Further, the low ionization efficiency often requires a special inclusion list or exclusion of tandem MS precursor signal in a second LC-MS/MS experiment to acquire spectra of multiphosphorylated peptides.

To deplete unmodified peptides from the sample, I established a phosphopeptide enrichment protocol on titanium dioxide, which is compatible with subsequent HPLC separation of phosphopeptides and either electro-spray ionization or matrix-assisted laser desorption/ionization. The method is highly robust and does not require further modification of unmodified peptides by esterification of the carboxylic acid moieties, thereby allowing a more general application. Further, this protocol can also be applied to samples with detergents, since the high percentage of organic modifier will remove these buffer additives in the loading and washing steps of the titanium enrichment.

Often the most critical step is separation of the peptides by RP-chromatography before mass spectrometric analysis. To improve this step, I implemented separation on monolithic columns resulting in shorter elution fractions of less than five seconds at high resolution and allows direct deposition of the eluting peptides on a MALDI sample plate. Peptides with different number of phospho-sites were reliably separated with the highest phosphorylated form eluting first.

Further we developed a robust MALDI detection protocol with DHB as optimal matrix for ionization of phosphopeptides. Using this protocol, multiply phosphorylated peptides could be detected down to 20 fMol on the spot. Improved identification of phosphorylated peptides was achieved by elimination of neutral loss peptide ions in the upper mass range that often dominate tandem MS spectra of multi-phosphorylated peptides. This step is not available for ESI MS/MS spectra, since the multiply charged precursor ions form a region of neutral loss peaks in the central mass range, thereby overlapping with many fragment ions. In comparison to standard protein phosphorylation analysis peptides with higher phosphorylation were identified and the phosphorylation sites were assigned with increased accuracy. ESI-CID

MS/MS was not able to identify the same phosphopeptides at low concentration, but the recently introduced ETD fragmentation technique was as efficient as the MALDI method.

Although direct analysis of phosphopeptides from complete cell lysates cannot be provided by the presented procedure, it is able to detect phosphopeptides reliably after further reduction of the sample complexity as for instance by purification of protein complexes with immuno-precipitation. Interestingly, only one third of the sample amount used for ESI-MS/MS analysis was required to identify the major phosphopeptides. The residual amount could be used to evaluate the findings in experimental repeats or to obtain MS/MS data for peptides that have not been identified in this step.

A further problem, which comes more into focus of phosphoproteomics, is comparison of the phosphorylation state of a selected protein from different samples or under different conditions. Direct comparison of two individually analyzed samples shows significant shortcomings such as differences of sample amounts, separation conditions, or peptide ionization and detection conditions. Next to these technical problems, data interpretation plays a crucial role for the comparison of two different samples. Therefore, not only the identification score, which is provided by the search algorithm, is important for identification of the respective peptide sequence in both samples. Moreover, reliable localization of the peptide modification within the sequence improves the comparison. MALDI-MS/MS analysis addressed both questions, due to the separation of neutral loss ions and sequence specific fragments by the analysis of singly charged peptides and increased mass accuracy of fragment ions. To address such quantitative question by mass spectrometry, diverse isotope dilution methods have become quite popular in recent years. The presented protocol allows implementation of any of the published isotope labeling strategies such as SILAC (Ong S.E. *et al.*, 2002), ICPL (Schmidt A. *et al.*, 2005), and iTRAQ (Ross P.L. *et al.*, 2004).

PART B ANALYSIS OF ACID LABILE PROTEIN PHOSPHORYLATIONS

Protein purification, mutation studies of CtsR, and x-ray structural analysis of the transcription factor CtsR-DNA complex were carried out by J. Fuhrmann in the Group of T. Clausen at the IMP Vienna. To demonstrate the importance of the identified protein arginine phosphorylation in its biological context, these data are shortly described in the results section. Initial protein full mass determination and protein fragmentation experiments were performed by C. Stingl a former member of the Protein Chemistry Facility at the IMP Vienna. These data clearly demonstrate the phosphorylation state and are highly consistent with the data from the analysis of protein digests. NMR measurements were carried out at the NMR facility by Dr. L. Brecker of the Department for Organic Chemistry at the University of Vienna.

General Introduction

Nitrogen-bound Protein Phosphorylation Processes in Cells

The methodology for mass spectrometry based identification of protein phosphorylations has improved significantly by the introduction of enrichment techniques. Nevertheless, only few studies report on phosphorylation events in bacteria, since many phosphorylations in these organisms are attached to other amino acid side chains. Unlike eucaryotic cells, phosphorylation on serine, threonine or tyrosine is mostly absent in bacteria (Miller M.L. *et al.*, 2009). This observation is mainly due to a different set of protein kinases and phosphatases, which are expressed in bacteria and couple phospho-moieties to nitrogen atoms of histidine or arginine side chains (figure B_01).

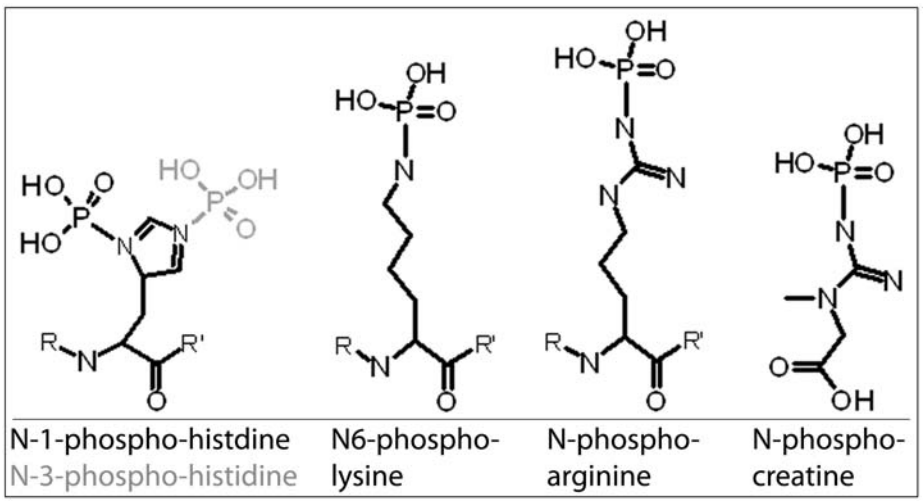


Figure B_01 **Nitrogen-Bound Phosphorylations in Living Cells**. Phosphohistidine is highly important for bacteria to sense molecules in the extra-cellular medium. Phospholysine is seldomly observed as reaction intermediate of phospho-transfer catalyzing enzymes. Phosphoarginine (invertebrates) and phosphocreatine (vertebrates) are important in energy homeostasis, since they buffer high ATP levels. Arginine phosphorylation in proteins was suggested in earlier studies.

A major part of known bacterial phospho-signalling is based on histidine and aspartate phosphorylation in two component signalling systems or phospho-relays (figure B_02A) (Mitrophanov A.Y. and Groisman E.A., 2008, Hoch J.A. and Silhavy T.J., 2005). Since bacteria inhabit precarious environments they need reliable detection systems for humidity, acidity, nutrients, temperature, or osmolarity, which are in majority based on two-component systems (Hoch J.A. and Silhavy T.J., 2005, p. 9). Further, these protein systems are also involved in chemotaxis by which bacteria detect different toxins in their environment and are able to move to the most fortunate place to grow (Parkinson J.S. and Kofoid E.C., 1992). Another very important task of two component signalling system is to report detected changes in the environment, to allow adaptation to altered growth conditions, for instance if density of bacteria becomes very high (Freeman J.A. *et al.*, 2000, Gov Y. *et al.*, 2004). Furthermore, two component signalling systems can also be found in plants where they developed a diverse protein class that has multiple functions (Mizuno T., 2005). Both systems, from bacteria and plant, consist of an often membrane-associated receptor that has the primary protein histidine kinase domain in its cytoplasmic domain (figureB_02B). Upon sensing of substrate molecules two receptor proteins dimerize and the structure of the intra-cellular kinase domain is altered, thereby enabling the kinase to autophosphorylate on a histidine, which is often conserved throughout different species. Similar to receptor tyrosine kinases in mammalian cells, the histidine residue of one receptor subunit is always phosphorylated by the active kinase domain of the second subunit and not by autophosphorylation within a single receptor molecule (Laub M.T. and Goulian M., 2007). This reaction product, phospho-histidine, is then the substrate for phosphorylation of a highly conserved aspartate residue in the response regulator protein. To demonstrate that the histidine phosphorylation is not the signalling molecule in this process, it was shown that the response regulator protein is phosphorylated by incubation with small phosphor-His analogues such as phosphoramidate or imidazol-N-phosphate (Hoch J.A. and Silhavy T.J., 2005, p18f.). Since the phospho-acceptor is responsible for the transfer step, the recognition of the dimerization domain of the receptor

histidine kinase by the response regulator must be highly specific to exclude activation of other phosphorylation pathways. Upon aspartate phosphorylation, the associated acceptor protein is able to interact with the flagellum system to change the direction of movement or it is a transcription factor and regulates expression of specific proteins.

The "so called" phospho-relay systems are a special form of the two-component response systems that involve multiple phospho-transfer reaction between histidine and aspartate in different proteins. The increased complexity of such systems allows a better regulation of the signal transduction process. This process is similar to the mammalian receptor tyrosine kinase systems, which are involved in sensing of growth factors and therefore are often associated with cancers (Blume-Jensen P. and Hunter T., 2001). Often, histidine kinases of two-component response systems are not only required to transfer the phosphate group onto the histidine, but many of them also have phosphatase activity that removes the modification to silence the signalling process (Laub M.T. and Goulian M., 2007).

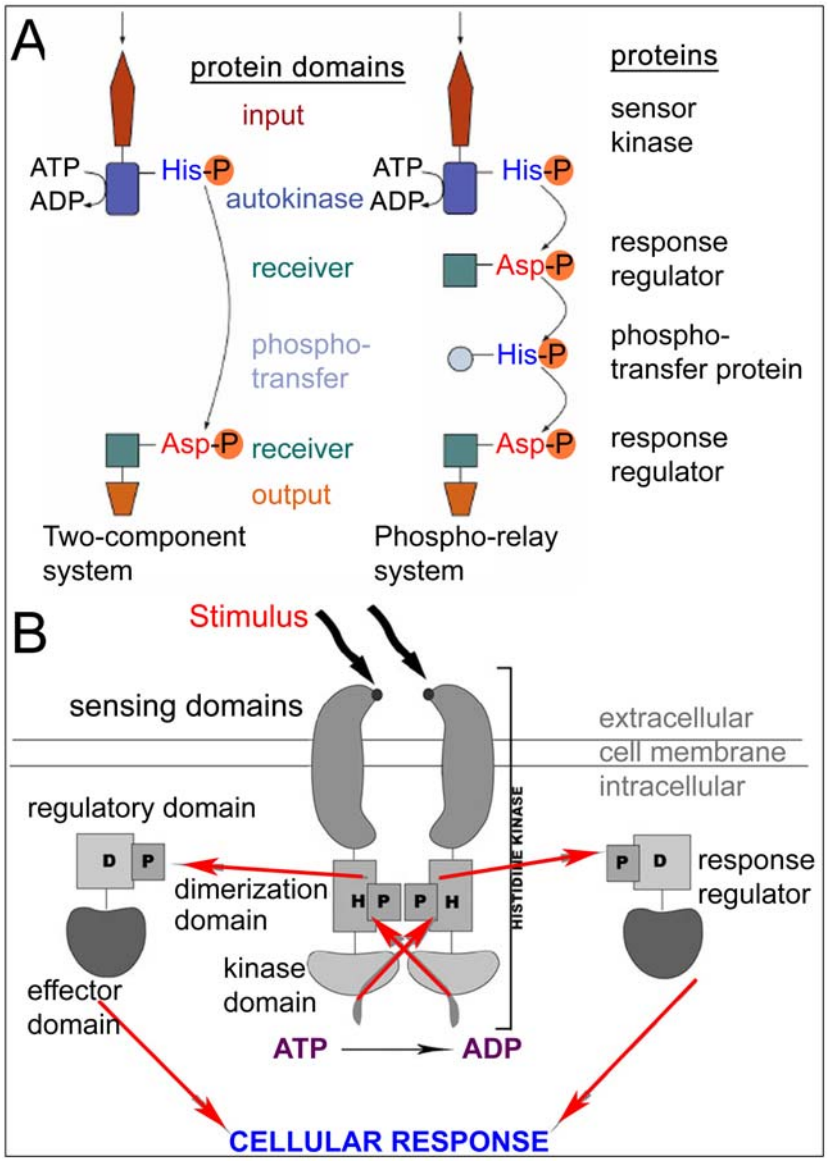


Figure B_02 **Receptor Histidine Kinases in Bacteria and Plants.** A. Two forms of receptor histidine kinases have evolved; two-component systems provide direct activation of the transcription factor after autophosphorylation of the receptor kinase. Phospho-relays include three phospho-transfer steps, first from His to Asp, than to a His of a transfer protein, and the third step is again from His to Asp of an effector protein. (modified from Mitrophanov A.Y. and Groisman E.A., 2008) B. Scheme of an active

receptor histidine kinase. Upon interaction with a target molecule the receptor dimerizes, autophosphorylates conserved His residues in the dimerization domain. The response regulator translocates the phosphorylation site onto an Asp residue and thereby activates it output domain (effector domain). (modified from Attwood P.V. *et al.*, 2007)

To a yet unkown percentage also phosphorylation of arginine is present in bacterial proteins. Phospho-arginine is a rarely described protein modification and only one mechanism, where phosphorylated arginine is involved in adaptation to stress conditions, has been elucidated recently (Fuhrmann J. *et al.*, 2009). Other studies in histones from bovine heart also assumed phosphorylation of arginine, but biological meaning and the responsible kinase were not identified (Wakim B.T. and Aswad G.D., 1994; Wakim B.T. *et al.*, 1995). Furthermore, there are only few studies on phosphoarginine phosphatases, which remove this modification to end signalling (Kumon A. *et al.*, 1996). So far only a general protein phosphatase has been described to have activity towards arginine phosphorylated peptides and the possibility of a dual activity of arginine modifying kinases, similar to histidine kinases in two-component systems, cannot be excluded. The limited knowledge makes it difficult to select the right phosphatase inhibitors during sample preparation, so that a certain percentage of modification is hydrolyzed by residual phosphatase activity.

Phosphorylation of guanidyl moieties also occurs in eukaryotic cells, where creatine and arginine are converted to phosphagens, thereby representing an alternative energy storage system (Attwood P.V. *et al.*, 2007, Besant P.G. *et al.*, 2009). While higher eukaryots such as mammals, fish, or birds are able to utilize creatine (Curtin N.A. and Woledge R.C., 1978), invertebrate organisms phosphorylate the free amino acid arginine on its guanidyl function (Padwardhan P. and Miller W.T., 2007). The activity of the corresponding creatine or arginine kinases depends on the intracellular ADP and ATP levels. If the ADP level is low and much ATP has been produced the cell, the phosphagen kinase transfers a phosphate moiety from ATP to the guanidyl function of creatine or arginine, forming one molecule of ADP, which then can further be used for ATP production. Under conditions, where ATP is present in low levels and high levels of phosphagens exist, the creatine/arginine kinases are able to catalyze the reversed reaction from phosphagen to creatine or arginine and form ATP.

N-phosphorylations on proteins have first been described in the 1960s, where biologists discovered phosphorylated proteins that lose the radioactive phosphorylation signal upon acidic treatment (Slater E.C. and Kemp A., 1964; Wålinder O., 1969). Localization of this modification by different methods pointed towards the amino acids histidine, lysine, and

arginine, which could be modified at nitrogen atoms in their side chains. Phosphorylation of small chemical molecules like butylamine was used as model system to study such protein modifications: the kinetics of pH dependent decay and the reactivity of N-phosphorylations. In contrast to arginine and lysine, where only one phosphorylated form exists, chemical phosphorylation of histidine leads to two different products, 1-phosphohistidine and 3-phosphohistidine. Up to now it has not been shown, if only one of the two forms has biological relevance or if histidine kinases are able to produce both isoforms.

All N-phosphorylations and the associated aspartic acid phosphorylation share high instability at lower pH values. On the other side, N-bound phospho-moieties are highly stable at alkaline conditions, where phospho-esters of serine or threonine can be removed.

I. Chemical Properties of Phospho-Amidates on Biomolecules

In comparison to the acid-stable Ser-, Thr-, Tyr-bound phospho-esters, phosphate moieties attached to His, Lys, Arg have a covalent nitrogen-phosphorous bond (figure B_01). Interestingly, the nitrogen atoms in the side chains of these amino acids retain their high proton affinity upon phosphorylation (Attwood P.V. *et al.*, 2007, Besant P.G. *et al.*, 2009). The pKa values for these protonation sites are only slightly lower than that of the unmodified side chain (see table B_1). This observation is unusual, because the phosphate moiety is an electron withdrawing group, which should be able to delocalize the free electron pair of the nitrogen atom into different structures with N=P double bonds (figure B_03). In phosphoarginine and phosphohistidine two different electron withdrawing groups compete for the lone pair electrons of the sp^2-hybridized nitrogen atom.

	phosphorylation		**NH of side chain**
compound	**pK_a1**	**pK_a2**	**pK_a**
histdine			6
N-1-phosphohistidine		< 3.0	7.04
N-3-phosphohistidine		< 3.0	6.18
lysine			10.5
N-ε-phospholysine		2.9	9.9
arginine			12
N-ω-phosphoarginine	1.96	4.5	11.2

Table B_1 **pKa Values for Nitrogen-Bound Phosphorylations**. The pKa of the phospho-moiety is similar to phospho-esters of Ser, Thr, Tyr (pKa1 = 1.2, pKa2 = 4.6). The pKa of the NH-group in the side chain is almost not altered in comparison to the unphosphorylated amino acid. Due to structural similarity, phospholysine has the same pKa values like phosphoramidate (NH_4PO_3).

Due to stronger withdrawing effect of the guanidyl group and the imidazole ring, they show the dominating effects, thereby making formation of N=P bonds unfavourable (Ruben E.A. *et al.*, 2005). In case of lysine phosphorylation the nitrogen atom is sp^3-hybridized, which precludes formation of double bonds. The pKa for the alkylic nitrogen atom is lowered from 11.2 to 9.2 in the phosphorylated form. Upon phosphorylation all phosphorous-bound nitrogen atoms encounter a reduction of the pKa value, which is predominantly due to inductive withdrawing effects. (Besant P.G. *et al.*, 2009).

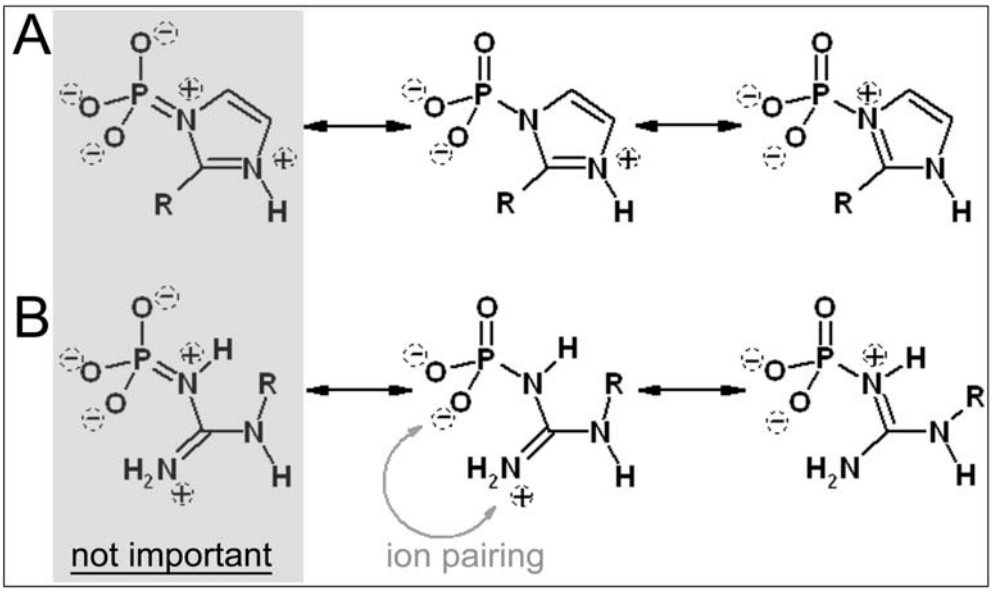


Figure B_03 **Charge Distribution and Mesomeric Structures of Phosphohistidine and Phosphoarginine at Neutral pH**. Mesomeric structures, having a single positive and two negative partial charges, are highly important, whereas the form with a triply negative charged phospho-moiety is almost not abundant.

Under acidic pH conditions N-bound phosphorylations rapidly hydrolyze, which can be described by an SN_2-like mechanism (figure B_04) (Sickmann A. and Meyer H.E., 2001; Besant P.G. *et al.*, 2009). In this reaction the phospho-moiety is transferred to a water molecule. The N-P bond is weakened in the transition state, leading to a larger distance between N and P atom. At this stage, the O-atom of a water molecule is able to interact with the P-atom from the other side, since the phospho-moiety adopts a planar structure in the transition state. Finally, the modification is eliminated as monohydrogenphosphate. The free reaction energy ΔG° for the hydrolysis of phosphoarginine is -7.3 kJ/mol, hydrolysis of phospholysine has a higher ΔG° of -42 kJ/mol (Besant P.G. *et al.*, 2009). Dissociation of the "low energy" N-P bond is the rate-determining step in this reaction (Besant P.G. *et al.*, 2009). In case of phospho-arginine and phospho-creatine, positively charged guanidyl groups interact

with the negatively charged phospho-moiety thereby forming an ion pair, which slightly increases their resistance towards hydrolysis. Upon hydrolysis this pair has to be disrupted, which increases the free energy of the transition state and makes phosphagen molecules ideal for energy storage, because the phosphorylated forms are quite stable under physiological conditions. While phosphorylated arginine can be stored for quite a long while at pH 7, phospho-histidine and phospho-creatine will decay within a few days.

At basic pH all phosphoamidates have a high stability because the amine anion is unsuitable as leaving group.

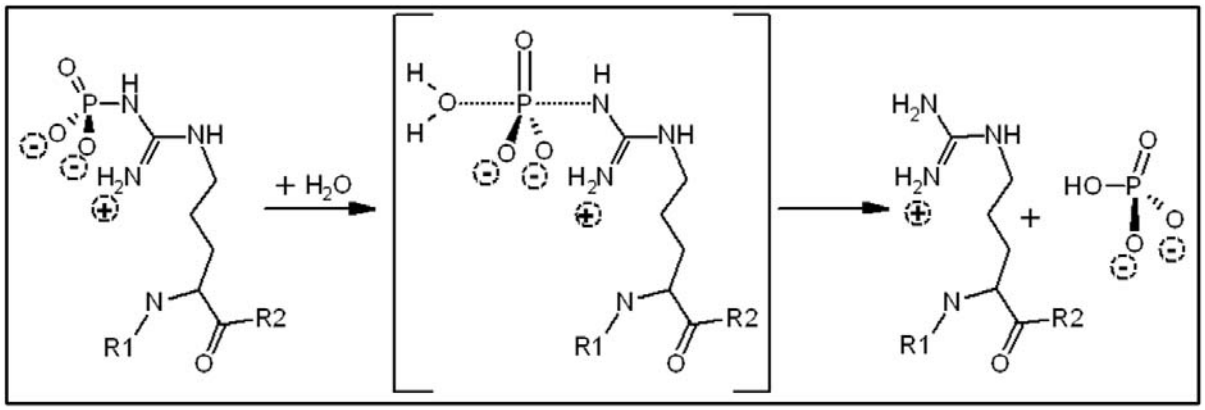


Figure B_04 **Acidic Hydrolysis of the Phospho-Amidate of Arginine Follows a S_N2-like Mechanism.**

II. Analysis of Nitrogen-bound Phosphorylations on Peptides and Proteins

One of the most sensitive techniques to study protein phosphorylation is the introduction of a radioactive phosphorous atom, which can be either ^{32}P or ^{33}P. Detection of the radioactive radiation facilitates the identification of phosphoproteins and mutation of putatively phosphorylated residues helps to identify a phosphorylation site. The first protein phosphorylations, which dissociated upon sample acidification, were discovered in the 1960s. Since as previously described serine, threonine and tyrosine phosphorylations are acid-stable, other amino acids have to be the targets for this type of kinase modification. Later studies revealed that the amino acids histidine, lysine, arginine, and aspartic acid are also phospho-acceptors (Weller M., 1978; Levy-Favatier F., 1987). Next to mutational studies in combination with acidic removal of N-phosphorylations, thin layer chromatography of amino acids was used to determine the type of phospho-amino acid (Hoch J.A. and Silhavy T.J., 2005). Thereby amino acids were released from the proteins after incorporation of a radioactive phospho-moiety by alkaline hydrolysis and mixed with unlabeled standard. After

one- or two-dimensional separation by thin layer chromatography the localization of the radioactive spot revealed the phosphorylated amino acid in the protein.

Later arginine-phosphorylation was detected by EDMAN-sequencing of the unmodified and the phosphorylated peptide. The comparison of both results shows a loss of the amino acid signal at the position of the modified residue for the phosphorylated peptide. Standard amino acids can be used to validate the finding of an N-phosphorylation and to exclude a different type of modification. Unfortunately, the commercial phospho-arginine standard is no longer available.

III. Mass spectrometry of protein N-phosphorylations

Due to the acid-lability, detection of peptides with N-phosphorylation is still challenging for mass spectrometry. Acidic sample preparation and dilution steps enhance hydrolysis of the phospho-amidate. Analysis time seems to be the critical point. Whereas normal peptides can be kept in acidic solvents for months or even years, peptides with labile modification should be analyzed within one day otherwise the chance to miss the modification is quite high (Zu X.L. *et al.*, 2007). Recent studies on phospho-Arg in proteins have not been conducted with mass spectrometry since this technique was not jet available (Wakim B.T. *et al.*, 1995). First MS-detection of His-phosphorylated peptides was achieved by MALDI MS and the acid lability was a good criterion to distinguish histidine phosphorylation from acid-stable phospho-esters in a hydrolysis step of a second sample (Attwood P.V. *et al.*, 2007, Wålinder O., 1969). If peptides with histidine phosphorylation were analyzed by MALDI mass spectrometry in positive ion mode, almost exclusively the unphosphorylated form was observed. Interestingly, the intact phosphopeptide could be detected in negative ionization mode. ESI-MS allowed detection of both peptide forms in positive ionization mode, because more protons are transferred to the peptide.

The first enrichment protocol for peptides with His phospho-modification is based on the affinity of the imidazole ring to copper and the negative charge of the phospho-moiety at low pH conditions (Ross A.R., 2007). Ross demonstrated that peptides without phosphorylation are effectively removed under mild acidic conditions and phosphorylation of other amino acids can be washed away by switching to conditions to formation a stable imidazole-copper complex of the histidine side chain.

Sample preparation and mass spectrometric detection of histidine phosphorylated protocols has not been intensely investigated so far. Especially the biochemical sample

processing requires significant improvement. Since these modifications are labile, the biggest difficulty for extensive studies is to accumulate enough peptide. Previously, it was demonstrated that low energy CID activation of peptides with phosphorylated histidines leads to two different neutral loss products (Medzihradszky K.F. *et al.*, 1997). While a signal at M-80 Da, which stems from elimination of HPO_3, could easily be explained from the chemical nature of the histidine phosphorylation, the observed loss of phosphoric acid (M-98 Da) was unexpected, because of the nitrogen-phosphorus-bond (Medzihradszky K.F. *et al.*, 1997, Zu X.L. *et al.*, 2007). Since the unmodified peptide did not show extensive water elimination, a concerted mechanism for the phosphoric acid loss is under discussion. To explain this neutral loss behaviour, a phospho-transfer from the histidine side chain to a free OH-moiety within the peptide has to take place in the first step, which is followed by the elimination of phosphoric acid. In their report, Attwood et al. described an adjacent serine residue as the phospho-acceptor (Zu X.L. *et al.*, 2007). Recently, Kleinijenhuis et al. investigated the fragmentation behaviour of histidine phosphorylation upon low energy electron activation (Kleinijenhuis A.J. et al., 2007). Despite the low stabilty in solution and high tendency to lose H_3PO_4 upon thermal activation no neutral loss is observed in ECD, ETD or EDD processes.

Part 2. Identification and Characterization of Arginine-Phosphorylated peptides and proteins

a the CtsR phospho-switch

Cells encounter a variety of stresses which affect the stability and folding of proteins. Bacteria developed therefore highly efficient protein quality systems, which are able to remove misfolded or aggregated proteins and often consist of chaperones and proteases. Since these degradation complexes can cause problems during normal cellular growths, they are only expressed under stress conditions (Hecker M. *et al.*, 1996). Most of these stress factors are encoded in a small number of genes and share the same promoters. Therefore, regulation is often achieved on the transcriptional level by inhibition of the DNA transcription process, for instance by the *Class three stress gene regulator* (CtsR) protein (Derre I. *et al.*, 1999). CtsR is expressed in Gram-positive bacteria and is encoded in the ClpC operon, which consists of the genes clpC, ctsR, mcsA and mcsB (Krüger E. *et al.*, 1998; Derre I. *et al.*, 1999). ClpC is a protein chaperon that is able to unfold misfolded proteins with the help of ATP. McsA and McsB (*modulator of class three stress gene regulator A and B*) are required

to phoshorylate CtsR. McsB is the kinase, applying the phosphorylation; McsA is believed to act as activator or stabilizer of McsB (Kirstein J. and Turgay K., 2005). This operon is not completely conserved, so that few species actually miss the mcsA and mcsB genes.

The transcription factor CtsR is bound to the promoter region of the gene under normal growth, thereby inhibiting the transcription of the upstream genes (Krüger E. *et al.*, 2001; Kirstein J. and Turgay K., 2005). When the cell encounters stress CtsR dissociates from the DNA and allows expression of the stress genes. To prevent re-association of the complex, the DNA-binding domain is phosphorylated by the protein kinase McsB (Kirstein J. and Turgay K., 2005; Kirstein J. *et al.*, 2005).

The protein structure consists mainly of helices (figure B_05), which form two protein domains (Fuhrmann J. *et al.*, 2009). The N-terminal domain represents a winged helix-turn-helix motif, which is a common DNA binding fold and involves many basic amino acids (R, K, H). The C-terminus is the protein dimerization domain and consists of a bundle of 4 helices, to interact with another CtsR molecule via a large hydrophobic surface. Upon interaction with DNA, the helix-turn-helix motif (amino acids 2-72) coordinates the phosphate backbone of the DNA by ion pair formation with positively charged amino acids and interacts with the major groove. The wing domain is a short peptide strand, which is stabilized by a ß-sheet structure with other residues in the N-terminal domain and comprises the amino acids 57-71. In complex with DNA the tip of the wing (aa 62-66, RGGGG) is fully inserted into the DNA minor groove and interacts with the organic bases. The major function of this domain is the recognition of the correct DNA sequence in the promoter of the regulated genes. A second, yet not fully described function arises from the flexibility of this region that is even increased at elevated temperatures. Under such conditions the wing could leave the minor groove and thereby initiate the dissociation of the complex.

Interestingly, both CtsR subunits recognize parallel DNA-sequence repeats, which is only possible by a bend of the N-terminal domain of the second subunit. Most dimeric transcription factors bind to palindromic DNA-repeats. The flexibility of CtsR lies in the linker region (aa 72-78) that is able to turn the second DNA-binding domain by 180°.

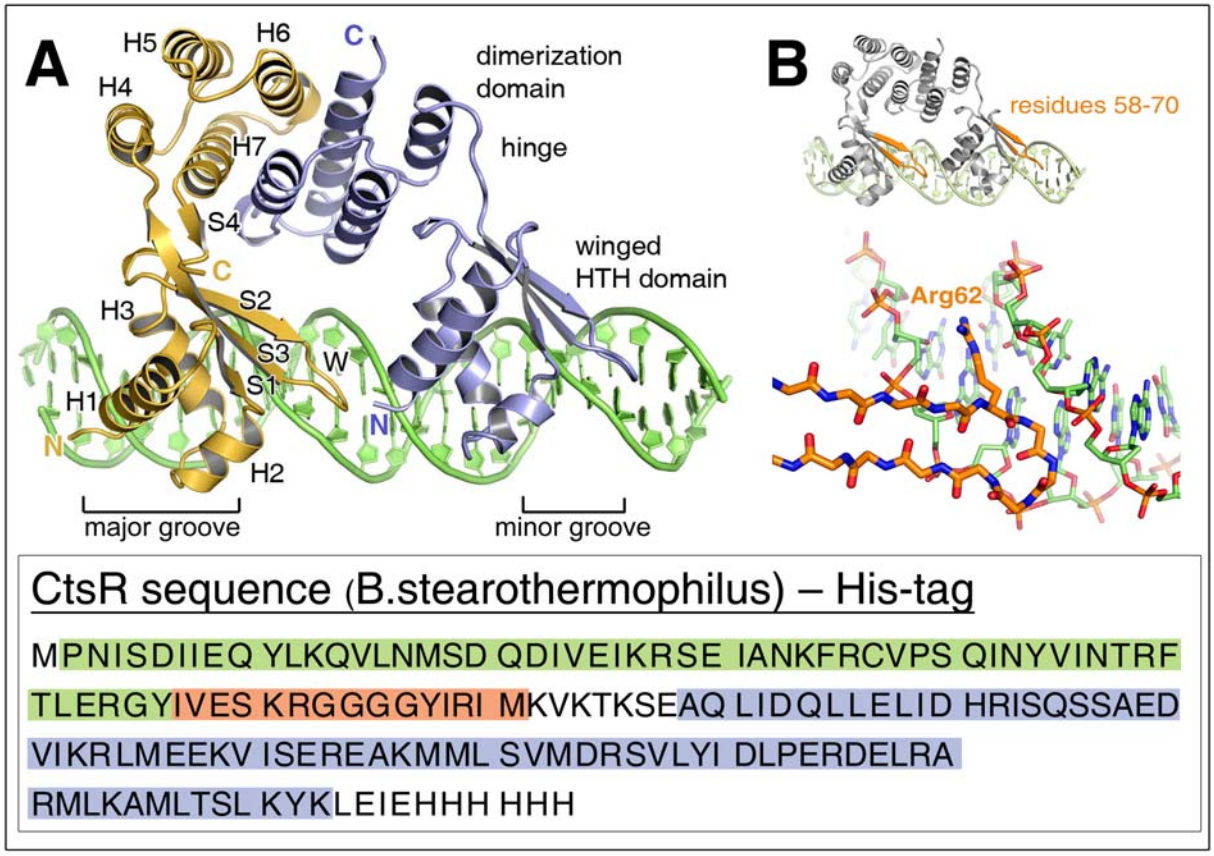


Figure B_05 **CtsR Sequence and Structure of the CtsR Dimer Bound to the DNA Recognition Sequence.** A. Each CtsR subunit consists of an N-terminal DNA-binding domain and a C-terminal dimerization motif. The DNA-binding domain binds to the major groove of the DNA and is a "so called" helix-turn-helix motif, comprising amino acids 2-56 (green). It is extended by a wing domain, which interacts with the minor groove (aa 57-71, red). Only the dimer, which it is formed by interaction of the C-terminal helix bundles of two CtsR molecules (aa 79-153, highlighted in blue), is stably attached to the DNA. B. The wing domain is completely inserted into the minor groove of the DNA where the arginine at the tip forms hydrogen bonds with the DNA bases.

To establish the kinase-substrate-protein system which could be studied *in vitro* with high expression rates and protein stability the sequences of CtsR and McsB from *bacillus stearothermophilus* were cloned into a vector. For determination of the phosphorylation state and localization of the phosphorylation sites the proteins were constructed with a C-terminal His-tag and individually expressed in the bacterium Escherichia coli. Cells were lysed and the proteins were purified by Ni2+/NTA chromatography and gel filtration.

Phosphorylation Assays and Intact Protein Mass Spectrometry

The phosphorylation on CtsR was applied by the kinase McsB in an *in vitro* phosphorylation assay and the functionality of the system was validated by autoradiography of incorporated radioactive ^{31}P, phosphorylation specific staining of the polyacrylamide electrophoresis gel and mass spectrometry of intact protein see figure B_06A+B. For phosphorylation, proteins were incubated in 30 mM HEPES pH 8, 50 mM KCl, 5 mM MgCl2, 2 mM ATP, 10 mM CtsR, and 5 mM McsB for 10 min at 35°C. After the reaction CtsR was separated on a heparin column to remove the reaction components and the kinase. Further, reversed phase chromatography was applied on a 200 µm monolithic column, using acidic buffer conditions (DVB-PS co-polymer, Dionex, Schley C. *et al.*, 2006), to prepare the protein for ESI MS analysis. The CtsR protein was strongly retained on the column and eluted at a high percentage of organic solvent, so that the collected elution fraction could directly be used for off-line ESI mass spectrometry.

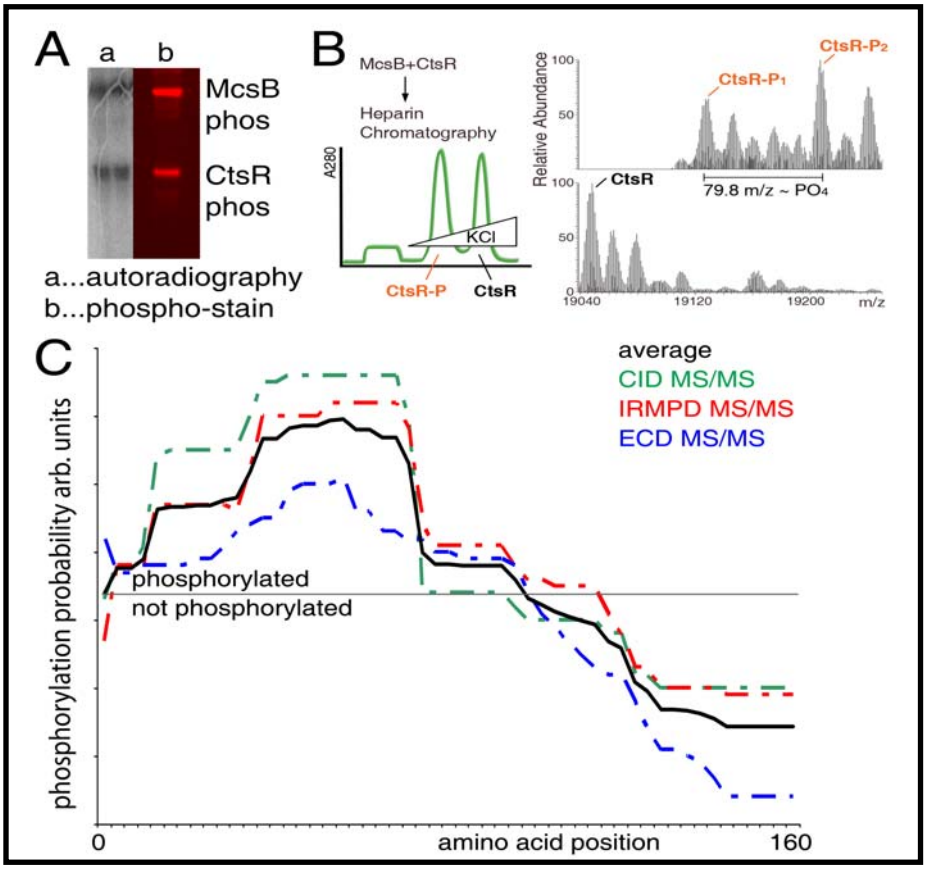


Figure B_06 **Detection of CtsR Protein Phosphorylation**. A. Autoradiography, which detects incorporation of radioactive phosphorous, and phospho-specific gel staining exhibited autophosphorylation of the kinase McsB and phosphorylation of the target protein CtsR upon incorporation with ATP. B. Unphosphorylated and phosphorylated CtsR can be separated by heparin-chromatography, whereof the unmodified form binds stronger to the chromatographic resin. MS of the phosphoprotein fraction reveals signals for monophosphorylation and

diphosphorylation of CtsR. C. *Top down* experiments of the monophosphorylated form suggest phosphorylation of the wing domain and the N-terminal DNA-binding motif with high probabilities, while modification of the C-terminus can be excluded based on the obtained results.

The protein was infused into a high resolution Fourier Transform Ion Cyclotron Resonance (FT-ICR) mass spectrometer (LTQ-FT, Thermo-Fisher) via static nano-ESI ionization. After optimization of the ionization conditions, mass spectra were acquired with a resolution of 100000 from 750-850 m/z, including charge states from $[M+23H]^{23+}$ to $[M+25H]^{25+}$. The final spectrum was averaged from 100 spectra. The spectrum was deconvoluted with the Xtract All software tool, which is part of the XCalibur QualBrowser software (Thermo-Fisher). The spectrum had evidence that up to two phosphorylations are incorporated into the CtsR protein by MscB (figure B_06). To identify the putative phosphorylation sites, the full protein was sequenced in the mass spectrometer using a *top down* MS/MS approach, since previous *bottom-up* LC-MS/MS studies did not lead to the detection of a phosphopeptide. For MS/MS analysis the 25fold charged protein ion ($[M+25H]^{25+}$) for the unmodified protein and the singly phosphorylated species were isolated in a 10 Da window and fragmented by CID, IRMPD and ECD activation. The fragment ion spectra (200-1400 m/z) were acquired with resolution of 100000 and 500 spectra were averaged for the final spectrum. Fragment ions were extracted from the mass spectrum and searched with the ProMass software Vs2.5 SR-1. Fragments, which matched to the unmodified protein sequence, were used as basal score. Since no phosphorylation was identified in bottom up approaches, all putative phosphorylation targets (Ser, Thr, Tyr, His, Arg, Lys, Cys, Asp) were included in the interpretation process. Identified protein fragments did not localize the phosphorylation to a specific amino acid. A high probability was found for the complete N-terminal domain, since, compared to the basal score, the identification score was increased for protein sequences with phosphorylated amino acids in this region. Highest scores were observed for the amino acids 55-82 which represent the wing domain. These results also excluded the C-terminus since all phosphorylated protein isoforms in this area scored lower than the unmodified protein.

New Bottom Up Strategy to Identify Arginine-Phosphorylation

Since previous attempts to localize the phosphorylation site in CtsR after proteolytic cleavage were not successful, the problematic sample preparation step had to be found. Top

down analysis clearly localized the phosphorylations to the N-terminal domain but not to a specific amino acid. Purified CtsR (20 µM) was mixed with 10 µM McsB, 5mM $MgCl_2$, 1 mM ATP, 30 mM Tris pH 8.0, and 10 mM β-mercaptoethanol and incubated at 40°C for 30 min in the presence of 2% glycerol. CtsR was purified by heparin chromatography to remove the reaction components. Cysteines were reduced with DTT and subsequently alkylated with iodoacetamide, as described elsewhere. In the first experiment, freshly prepared and phosphorylated CtsR protein was proteolytically cleaved by subtilisin, chymotrypsin and trypsin, individually. While trypsin produces a well defined cleavage pattern, chymotrypsin has a preference for hydrophobic amino acids and subtilisin is able to cleave at any amino acid. The peptide mixtures were not acidified to reduce the protease activity, all samples were analyzed with MALDI MS within 30 min after proteolytic cleavage and the residual was kept at -20°C .

Since phosphopeptides already eliminate H_3PO_4 from phospho-Ser, phospho-Thr, and phospho-His in MALDI-reTOF-MS experiments, each phosphopeptide gives rise to two signals. The major signal represents the intact phosphopeptide, which is accompanied by a minor neutral loss signal with a mass shift of -98 Da. The isotopic pattern of the neutral loss peptide ion is often not resolved, because the elimination occurs in the reflectron, at an analysis step, where high resolution cannot be achieved anymore. This characteristic signal, combined with the mass difference of 79.966 Da between unmodified and phosphorylated peptide allowed us the identification of phosphorylated peptides in the chymotrypsin and trypsin cleaved samples. Analysis of the subtilisin digest by MALDI MS was impaired, because of high salt concentration, interfering with crystallization and ionization of peptides. The chymotryptic digest had a phosphopeptide with the sequence IVESKRGGGGYIRIM (m/z 1715.84, aa 57-71) among the 10 most abundant peptide ions (figure B_07A). Phosphorylation of this peptide was further confirmed by MS/MS fragmentation, revealing a very intense loss of phosphoric acid. Phosphatase treatment of a fraction of the residual peptide digest also resulted in a loss of the phosphopeptide signal and showed an intense signal of the unmodified peptide (m/z 1635.89). The peptide sequence was identified from both MS/MS spectra of the unmodified and the phosphorylated peptide which clearly showed four adjacent glycines. Furthermore, one additional putative phosphopeptide was found in the chymotrypsin sample at 2434.22 m/z which also comprised the wing domain, see table TB_2. Phosphopeptides were not strongly represented in the trypsin cleaved CtsR sample, but all putative phosphopeptides pointed to phosphorylation of the central protein region, the wing domain with the block of four glycines (table TB_2).

Table B_2 **Phosphopeptides of CtsR detected by MALDI TOF MS**.
MALDI data identified several phosphopeptides in the region of the wing domain. Interestingly there was no hint for other phosphorylation sites from these data.

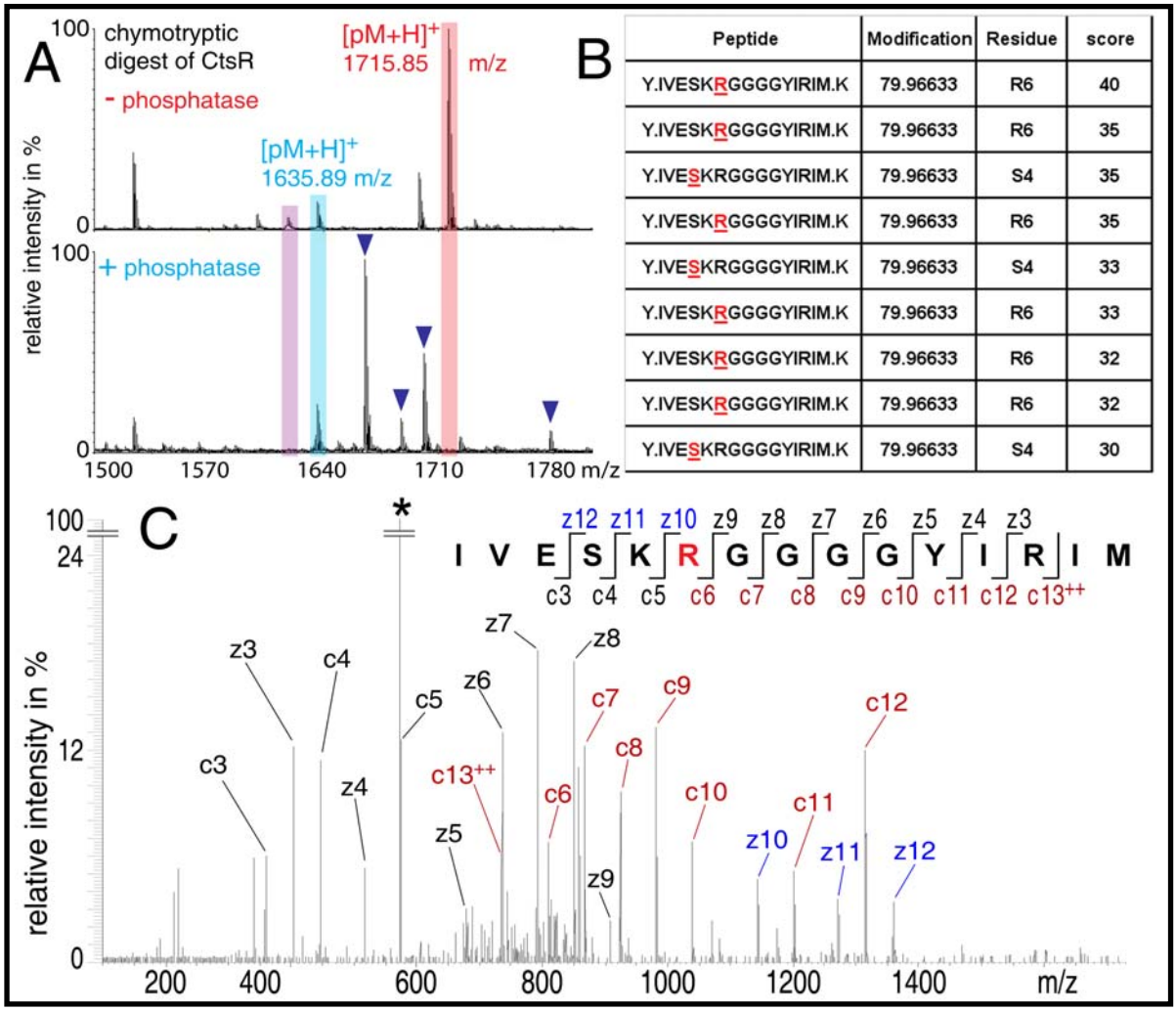
protease	sequence	aa	mass	
trypsin			no phos	phos
	RGGGGYIR	61-68	834.446	914.4123
	RGGGGYIRIMKVK	61-73	1433.829	1513.795
	RGGGGYIRIMKVKTK	61-75	1662.972	1742.938
	GYIVESK**R**GGGGYIR	54-68	1610.853	1690.819
	GYIVESK**R**GGGGYIRIMK	54-71	1983.072	2063.039
chymotrypsin				
	IVESK**R**GGGGYIRIM	56-70	1634.893	1714.859
	TLERGYIVESK**R**GGGGYIRIM	50-70	2354.253	2434.219

After identification of the phosphopeptide we tried to localize the phosphorylation within the peptide sequence, which includes one Ser and one Tyr as putative kinase targets. Extensive H_3PO_4 loss in MALDI-MS spectra and MS/MS analysis on the LTQ Orbitrap mass spectrometer localized the phosphorylation to Ser4 or Arg6, when arginine was also included as putative kinase target. Phosphorylation of Tyr11 was excluded from these data (figure B_07B). The sequence of y-ions in MALDI MS/MS pointed to an N-terminal localization of the modification.

Finally, unambiguous localization of the phosphorylation was only possible with a peptide fragmentation method, which keeps the modification attached to the amino acid. In combination with the high resolving power of FT-ICR MS, electron-capture dissociation (ECD) was the method of choice. The chymotryptic digest of freshly phosphorylated CtsR protein was separated by nanoRP-HPLC and detected on-line on an LTQ-FT mass spectrometer. All peptides with a charge state of 3+ or higher were automatically selected for fragmentation by ECD. Database search of the obtained mass spectra identified only the unphosphorylated and the phosphorylated form of the previously found peptide IVESKRGGGGYIRIM in charge state 3+. Manual interpretation of the phosphopeptide spectrum clearly localized the phosphorylation to Arg6 which represents Arg62 of the protein sequence. No evidence was found for phosphorylation at Ser4, Lys5, Tyr11, or Arg13 of the peptide sequence.

Next to the modification at Arg62 in the protein sequence, further experiments also revealed phosphorylation of Arg28, Arg36, Arg49, and Arg54. Arg28, Arg49, and Arg62 are highly conserved in the CtsR variants of different bacteria and are involved in DNA

recognition. In the complex with DNA, Arg28 and Arg62 are inserted into the major DNA groove. Arg36, Arg49 and Arg54 form salt bridges to the negatively charged phosphate backbone and thereby help to localize the DNA strand and stabilize the complex.



Peptide	Modification	Residue	score
Y.IVESKRGGGGYIRIM.K	79.96633	R6	40
Y.IVESKRGGGGYIRIM.K	79.96633	R6	35
Y.IVESKRGGGGYIRIM.K	79.96633	S4	35
Y.IVESKRGGGGYIRIM.K	79.96633	R6	35
Y.IVESKRGGGGYIRIM.K	79.96633	S4	33
Y.IVESKRGGGGYIRIM.K	79.96633	R6	33
Y.IVESKRGGGGYIRIM.K	79.96633	R6	32
Y.IVESKRGGGGYIRIM.K	79.96633	R6	32
Y.IVESKRGGGGYIRIM.K	79.96633	S4	30

Figure B_07 **Identification of the Phosphorylated Arginine in the Wing Domain at Position 62**. A. MALDI-MS analysis of the chymotryptic digest identifies a phosphopeptide at an m/z of 1715.85 (red box) with a neutral loss of H_3PO_4 at 1617.88 m/z (lilac). The unphosphorylated peptide (blue) was also observed at 1635.89 m/z. Phosphatase treatment of the peptide mixture with calf intestinal phosphatase removed the signal of the phosphopeptides, but residual chymotryptic activity generated fragments of the phosphatase (blue triangles). B. Further investigation by ESI-CID MS/MS resulted in ambiguous localization of the phosphorylation site within the peptide to S4 or R6. C. Only ECD fragmentation was able to reveal the arginine phosphorylation without doubts, since the modification is preserved. Fragments, carrying the intact phospho-moiety, are high-lighted in blue for z-ions and red for c-ions.

Biological Function of the phosphorylation

As described above, CtsR is a DNA-binding molecule that is removed from the DNA under stress conditions to allow stress gene expression. Within the complex the N-terminal domain establishes the contact to the DNA, while C-terminal domains of two CtsR molecules are fused together to form a stable dimer. MS analysis revealed phosphorylation of the N-terminal domain and all identified phospho-sites were arginines, which are crucial to maintain the protein-DNA interaction. As demonstrated by the blue colour in figure B_08A, many basic amino acids are located at the Protein-DNA interface and form ionic interaction with the negatively charged phosphate backbone. Structural data imply important interactions of Arg28, Arg36, Arg49, Arg54, and Arg62 with the DNA, mostly with the phosphate backbone. The location of these arginines in the protein-DNA complex is demonstrated in figure B_08A. Phosphorylation inverts the charge of arginine side chains. Therefore, interaction with DNA is impaired and the re-association of the complex is inhibited.

The tip of the wing domain, which is formed by the amino acids RGGGG (aa 62-66), is completely inserted into the minor groove of the DNA (figure B_05B). Together with the helix-turn-helix motive, these residues recognize the sequence of the DNA bases to find the correct binding side. Arg62, which lies in the wing-domain, interacts with the organic bases within the DNA-Helix and favours only two combinations of DNA sequences. Binding assays with phosphorylated CtsR or a mutant form of the protein with glutamate in position of Arg62, to mimic the phosphorylation, exhibited quantitatively inhibition of DNA binding, while mutating Arg62 to Lys results in increased stability of the protein-DNA complex.

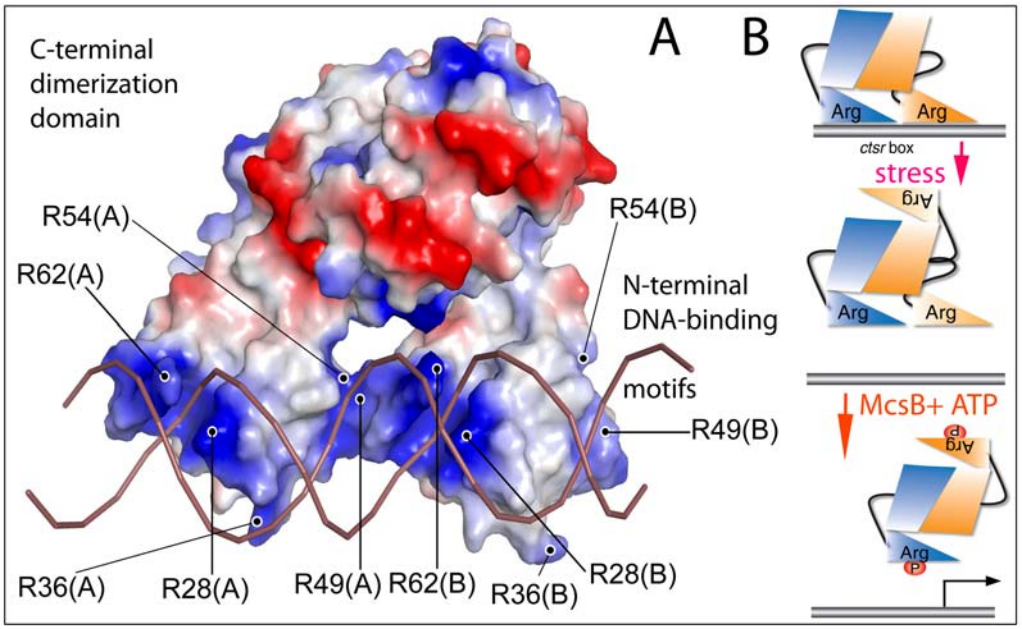


Figure B_08 **Localization of Phosphorylation Sites in the CtsR$_2$/DNA Structure**. Five different phosphorylated arginine residues were identified in CtsR, Arg28, Arg36, Arg49,

Arg54, and Arg62. All are located in the N-terminal domain of the CtsR subunits (A/B) and involved in protein-DNA interactions. Especially Arg62 in the wing domain (see figure B_05B) is essential for the stability of the complex. Arg28 and Arg62 interact with the bases in the major groove, the other phosphorylation sites coordinate the position of the DNA phosphate backbone. Phosphorylation causes electrostatic repulsion of the DNA and thereby impairs association of the complex. (red negative charge, blue positive charge, white uncharged). B Putative mechanism of CtsR regulation. Upon stress the $CtsR_2$/DNA-complex is destabilized and free CtsR is subsequently phosphorylated by McsB.

CtsR functions as suppressor of stress-gene transcription under normal growth conditions. Upon stress exposure, cells have to adapt quickly and therefore CtsR has to be removed from the DNA and re-association of the complex has to be inhibited. *In vitro*, the CtsR dimer requires only one DNA-recognition site to bind short DNA strands with one of the two helix-turn-helix domains. The location of the second DNA-binding domain is unknown. In a putative biological mechanism (figure B_08B), the CtsR dimer becomes more flexible under heat stress conditions and is able to release one helix-turn-helix domain from the DNA, which is then phosphorylated by McsB (Fuhrmann J. *et al.*, 2009). This leads to destabilization and dissociation of the complex. Free CtsR is further phosphorylated by McsB which guarantees inhibition of the re-association of the complex and allows transcription of the DNA region. A hint for such a mechanism is provided by NMR studies of a different transcription factor, which has a DNA-binding helix-turn-helix motif with a wing domain similar to that of CtsR (Derre *et al.*, 2000). This study shows that upon heat shock the highest gain of flexibility lies within the wing region. Addition of McsB alone to the $CtsR_2$/DNA complex in the absence of ATP is not sufficient to release the CtsR, demonstrating that the stress-response is not exclusively mediated by binding of McsB to CtsR.

b Characterization of the Protein Arginine Kinase McsB

Studying the regulation of CtsR activity by phosphorylation revealed the kinase activity of McsB towards arginine. Earlier, this protein kinase was reported to have a tyrosine specificity that was determined by thin layer chromatography of amino acids (Kirstein J. and Turgay K., 2005). For this experiment, amino acids of the purified McsB and CtsR proteins were released by acidic hydrolysis and phospho-amino acid analysis was performed by two-

dimensional thin-layer chromatography. The observed signals overlapped with the phospho-tyrosine standard (Kirstein J. and Turgay K., 2005). In a second experiment mutation studies showed a loss of kinase activity after exchange of specific Tyr residues to Ala in the protein sequence. Although both studies pointed towards tyrosine kinase activity, they are not sufficient to determine it unambigously. The hydrolysis of proteins and the second dimension of chromatographic separation were carried out under strong acidic conditions, which undoubtely resulted in the loss of the arginine phosphorylation according to our data. The mutation of several tyrosines in the second experiment could interfere with the protein structure and stability and therefore lead to loss of the kinase activity.

Peptide Phosphorylation by McsB reveals Arginine Specificity

To determine the kinase activity of McsB unambiguously, we conducted *in vitro* kinase assays with small synthetic peptides that had sequence similarity to the already identified phosphorylation site in CtsR. Peptides (40 µM) were individually mixed with 5 mM $MgCl_2$, 1mM ATP, 20 mM Tris pH 8, and 15 µM McsB to final volume of 100 µl. The reaction was performed at 40°C for 3 hours. Phosphorylated peptide products were analyzed by mass spectrometry and ^{31}P-NMR spectroscopy (figure B_09). Peptides, which revealed incorporation of a phospho-moiety, were conducted to MS/MS fragmentation to localize the site of modification.

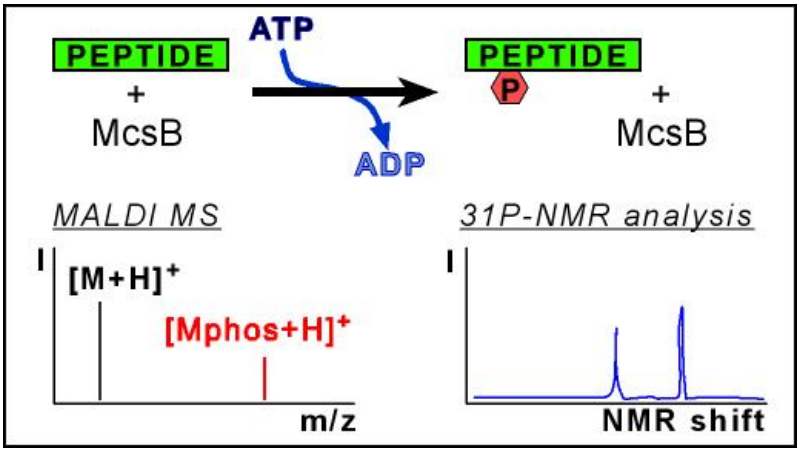


Figure B_09 **Schematic Illustration of the Peptide Phosphorylation Assay**. After incubation of the oligopeptide with McsB and ATP, the incorporation of phosphate is measured with MALDI-TOF mass spectrometry and the type of phosphorylation is evaluated by ^{31}P-NMR spectroscopy as independent analytical technique.

For phosphorylation, synthetic peptides with different amino acids in position of the putative kinase target were incubated with purified McsB kinase from B.stearothermophilus and ATP

(peptides see table TB_3). The reaction mixtures were diluted and directly analyzed by MALDI mass spectrometry. Synthetic peptide PepR_01, exactly constituting the CtsR amino acids 61-73 comprises two arginines that could serve as kinase target. MS/MS analysis showed that there was no evidence for a phosphorylation of R69 in the full length CtsR protein. Unfortunately, both arginines were phosphorylated by McsB on the synthetic peptide. To eliminate such side effects, I optimized the peptide sequence by exchanging the second arginine to lysine, resulting in the model peptide ac-KRGGGGYIKIIKV, PepR_02.

Table B_3 **Synthetic Peptides for the N-Phosphorylation Study**

Peptide	Sequence	Mass		Position	N-term	C-term
		no Phos	Phos	of Mod.		
PepR_01	KRGGGGYIRIMKV	1433.829	1513.795	R2, R9	free	free
PepR_02	KRGGGGYIKIIKV	1429.877	1509.843	R2	ac	free
PepR_03	KRGGGGFIKIIKV	1413.882	1493.848	R2	ac	free
PepR_04	KYGGGGYIKIIKV	1436.839	1516.806	Y2	ac	free
PepR_05	KR(Me)2GGGGYIKIIKV	1457.908		-	ac	free
PepR_06	KTGGGGYIKIIKV	1374.824	1454.790	T2	ac	free
PepR_07	KSGGGGYIKIIKV	1360.808	1440.774	S2	ac	free
PepR_08	KHGGGGYIKIIKV	1410.834	1491.801	H2	ac	free
PepR_09	KKGGGGYIKIIKV	1401.870		-	ac	free
PepR_10	KDGGGGYIKIIKV	1388.803		-	ac	free
PepR_11	KRGGGGYIEIIKV	1430.825	1510.791	R2	free	free
PepR_12	ERGGGGYIKIIKV	1430.825	1510.791	R2	free	free
PepR_13	ERGGGGYIEIIEV	1432.720		R2	free	free
PepR_14	YIVESKRGGGGY	1284.646	1364.613	R7	free	free
PepR_15	YIVESKRGGGGYIRIIK	1908.095	1988.061	R7	free	free
PepR_16	YIVESKRGGGGYIKIIK	1880.088	1960.055	R7	free	free
PepR_17	KRGGGGYIKIIKV	1443.893	1523.859	R2	ac	OCH3
PepR_18	KRGGGGYIKIIKV	1428.893	1508.859	R2	ac	NH2
PepR_19	KAQPLRAKFPAQPAK	1649.973	1729.939	R6	ac	free
PepR_20	KAQPLRAKFPAQPAK	1648.989	1728.955	R6	ac	NH2
PepR_21	KAEPLRAKFPAQPAK	1649.973	1729.939	R6	ac	NH2
PepR_22	KAQPLRAKFPAEPAK	1649.973	1729.939	R6	ac	NH2
PepR_23	KAQPLRAKFPASPAK	1607.962	1687.929	R6	ac	NH2

Upon phosphorylation with the purified kinase, I could show that only a single phosphorylation is incorporated into this amino acid sequence. Based on this sequence, new peptides were synthesized, where the arginine at position two was replaced by other amino acids (Ser,Thr,Tyr,Lys,His,Asp), which could be putatively phosphorylated to test, if McsB also modifies these residues. Mass spectrometry clearly demonstrates that only the peptide with arginine at position 2 (PepR_02) is phosphorylated (figure B_10A+B, spectral range: [M+H]-50 Da to [M+H]+150 Da). Furthermore, we studied, if modification of the arginine

side chain, for instance by methylation, impairs phosphorylation. Therefore I synthesized a peptide with a symmetrically dimethylated arginine in position two ac-KR$(Me)_2$GGGGYIK-IIKV, PepR_5. MALDI MS demonstrates that this substrate was not modified by McsB in the kinase assay (figure B_10A+B). Since tyrosine kinase activity was described for McsB previously, we also replaced the Tyr in position 7 of the model peptide with phenylalanine, which gave the sequence ac-KRGGGGFIKIIKV. After incubation with McsB and ATP, this peptide still revealed incorporation of a phospho-moiety, which would not have been observed with a tyrosine-modifying kinase (figure B_10A+B).

Exchange of other amino acids in the model sequence revealed a preference of McsB to phosphorylate basic peptides, while modification of acidic peptides is significantly reduced. Peptides ac-KRGGGGYIEIIKV and ac-ERGGGGYIKIIKV exhibited lower signals for the phosphorylated form in comparison to the model peptide PepR_2 and the version with three glutamic acids in position of the Lys residues could not be phosphorylated at all (figure B_10A+B). This observation is in good agreement with the observation that only the basic N-terminal DNA-binding domain of CtsR is phosphorylated. The identified phosphopeptides of CtsR and further studies with synthetic peptides revealed no specific recognition sequence for McsB but generally higher activity towards sequences with multiple basic residues.

Detection of Phospho-Arginine by ^{31}P-NMR Spectroscopy

Since so far only mass spectrometry based methods localized the phosphorylation to the side chain of arginine, a second analytical method should be applied to validate this finding. For this purpose, ^{31}P-NMR spectroscopy proofed to be the best tool to study this modification on small peptides. NMR spectroscopy measures the relaxation of the nuclear magnetic spin after perturbation by a radio impulse. Naturally, phosphorous occurs exclusively as ^{31}P-isotope, guaranteeing NMR detection of low amounts of phosphopeptide. The model peptide ac-K**R**GGGGYIKIIKV was prepared and purified by RP-HPLC at neutral pH with a $(NH_4)H_2PO_4$ buffer to yield approximately 50-100 µg pure phospho-peptide. The phosphate buffer was exchanged by solid phase extraction on C18 cartridges (Phenomenex, Germany) to heptafluorobutyric acid at pH 6.8. To exclude the presence of O-phosphorylations, I also measured the NMR resonance of the chemically synthesized phosphopeptides ac-K**S**GGGGYIKIIKV, ac-K**Y**GGGGYIKIIKV, and ac-K**T**GGGGYIKIIKV that were treated exactly like the arginine-phosphorylated peptide. All NMR analyses were carried out by Dr. L. Brecker at the NMR facility of the Organic Chemistry Department of the University of Vienna on a 600 MHz NMR instrument (DRX 600, Bruker). The exitation

frequency for the phosphorous nuclei was 242.94 MHz. The chemical shift for all four phosphorylation types was similar to that reported earlier and the arginine phosphorylation was again clearly confirmed and distinguished from phospho-esters (Kumon A. *et al.*, 1996). The resonance signals for pSer (+5.04 ppm) and pThr (+4.41 ppm) were shifted downfields in comparison to the reference at 0 ppm, the pY resonance (+0.86 ppm) was similar to the reference. Only the signal of the arginine phosphorylated peptide ac-K**R**GGGGGYIKIIKV had an upfield shift to -2.46 ppm, which is similar to free phospho-arginine (without the peptide context) at -2.52 ppm. (figure B_10C)

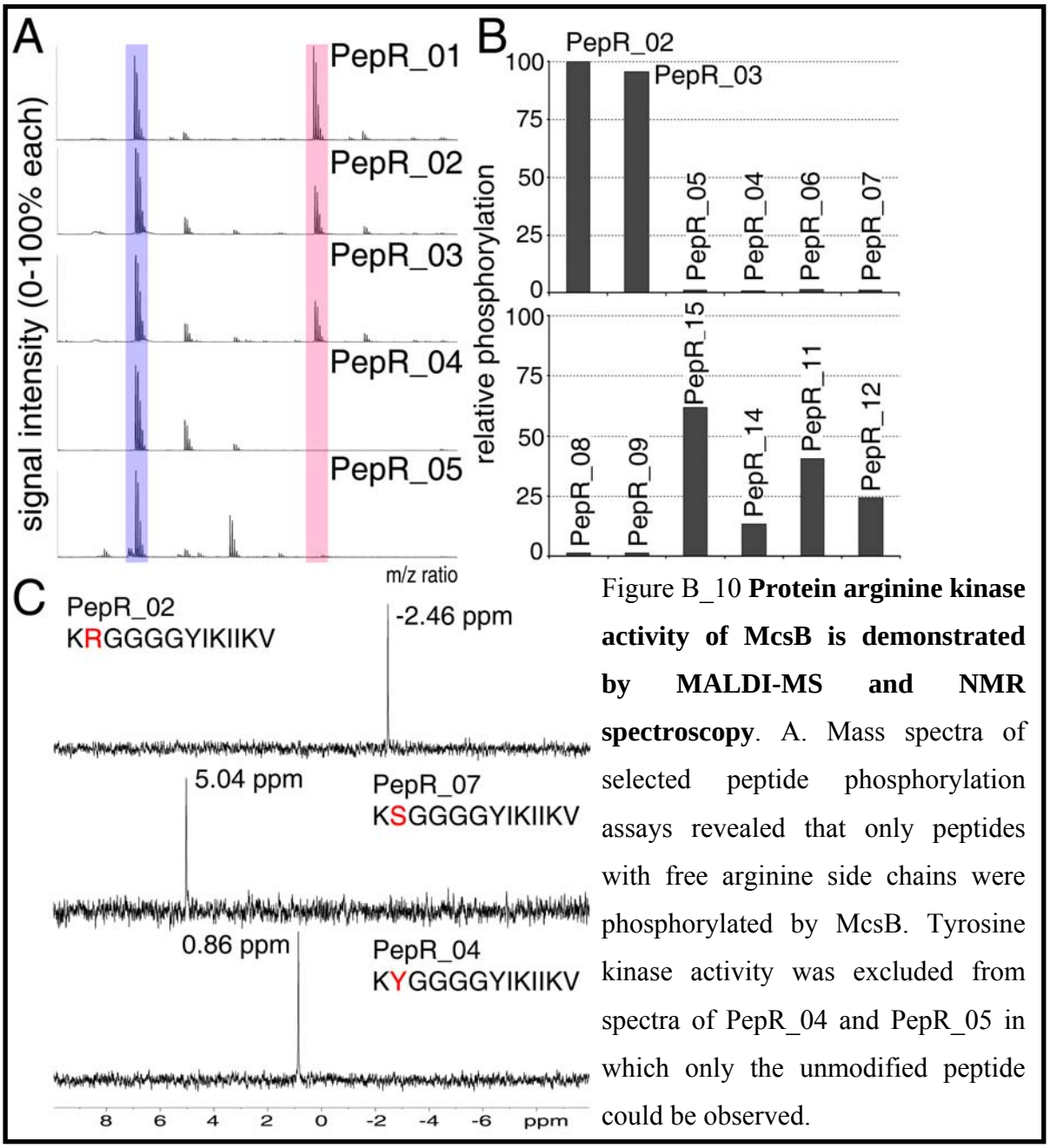


Figure B_10 **Protein arginine kinase activity of McsB is demonstrated by MALDI-MS and NMR spectroscopy**. A. Mass spectra of selected peptide phosphorylation assays revealed that only peptides with free arginine side chains were phosphorylated by McsB. Tyrosine kinase activity was excluded from spectra of PepR_04 and PepR_05 in which only the unmodified peptide could be observed.

B. Other putative kinase targets such as Ser, Thr, His, or Lys were also not modified by the arginine kinase McsB. The kinase preferred phosphorylation of peptide strands with basic

amino acids, while acidic residues have inhibitory function. C. ^{31}P-NMR spectroscopy of purified phosphopeptide PepR_02 (enzymatic phosphorylation), PepR_04 and PepR_07 (both chemical phosphorylation) reveals different chemical shifts of the phosphorous. The absolute values are close to the published chemical shifts of the free amino acids; pArg -2.5 ppm, pTyr 0.5 ppm, pSer 4.8 ppm.

c Stability of the Arginine-Phosphorylation and Enrichment by TiO_2-metal oxide affinity chromatography

Having the opportunity to selectively modify Arg residues with a phospho-moiety in a peptide background and reports about the presence of this modification in living cells, brought up the idea to establish an enrichment method that allows detection of these modifications by mass spectrometry. Enrichment of Ser, Thr, and Tyr phosphorylated peptides is almost a standard technique in proteomics and even His-phosphorylated peptides can be enriched (Thingholm T.E. *et al.*, 2009; Ross A.R., 2007). As previous experiments and publications revealed, the arginine-bound modification is sensitive to acid treatment and might be also unstable under alkaline pH conditions (Sickmann A. and Meyer H.E., 2001). Unfortunately, the majority of proteomic separation techniques (C18-RP-HPLC, SCX) and sample preparation steps (quenching of the protease activity, MALDI crystallization) are carried out at low pH.

The first step towards selective enrichment was to optimize purification of phosphorylated peptides from the kinase assays. These model peptides serve to measure the hydrolysis rate at different pH/temperature combinations, to determine the stability of the modification. Since Arg represents a putative target for tryptic cleavage, we also evaluated the ability of trypsin to cleave a peptide strand at the modified amino acid. Based on these data, we established a novel phosphopeptide enrichment protocol that allows purification of Arg-phosphopeptides and further detection by mass spectrometry. The solid phase for the enrichment protocol is TiO_2, but in contrast to most TiO_2 enrichment protocols the pH value is adjusted to approximately 4-5. First, we carefully optimized the method using the standard phosphopeptides mix A (table TA_1). Applying the novel protocol to a chymotryptic digest of an *in vitro* kinase assay of CtsR and McsB clearly depleted unmodified peptides and lead to the identification of 6 phosphorylation sites in CtsR and 15 phosphorylated arginines in McsB.

Purification of peptides with acid-labile modifications

To study arginine phosphorylation in proteins more intensely, we first aimed to optimize the purification procedure under non-hydrolyzing conditions for standard phosphopeptides. *In vitro*, the activity of the protein arginine kinase McsB is very high towards proteins, while small peptides are only poorly phosphorylated. Therefore we required high-resolution chromatographic separation of the phosphorylation mixture. Only RP-chromatography provides good separation, almost complete recovery of the analyte and has been tested under different pH conditions. Initial experiments showed that arginine phosphorylation is not hydrolyzed under alkaline pH, moreover, the stability is even increased. Therefore we evaluated RP separation at neutral and alkaline pH conditions on specific pH resistant silica-based C18 material GeminiNX C18 (Phenomenex, 15 cm x 4.2 mm).

The separation step requires a volatile buffer in the HPLC solvents, since these peptides will be used in further experiments which should not be influenced by additives from the solvents. Several buffer substances were tested for their solubility at high percentage of organic modifier (ACN), their UV absorption at 214 nm, and volatility in the freeze-drying process, whereof ammonium bicarbonate performed best. This was mainly due to the low UV-absorption of the buffer that also allows detection of low levels of phosphopeptides. Furthermore, this substance allows 10 mM concentration at 90% organic modifier, where phosphate buffers already precipitated. Final HPLC solvents for isolation of standard peptides consisted of 10 mM ammonium bicarbonate instead of the acidic modifier. At neutral pH, we were able to separate 2 of 3 Arg-phosphoryated peptides sufficiently from their unphosphorylated counterparts, whereas the peptide PepR_16 and the histidine-phosphorylated PepR_8 could not be resolved sufficiently, so that significant amounts of the unphosphorypated peptide were found in the phosphopeptide fraction. Shifting the LC conditions to pH 9 resulted in baseline resolution for all peptides, with the phosphorylated peptide eluting earlier than the unmodified peptide. Peptides with histidine phosphorylations were also resolved under these conditions after chemical modification with phosphoric acid amidate. Interestingly, for most peptides the phosphorylated form could be detected in a single sharp peak, while the unmodified peptide eluted in a broad signal of several minutes or was not even distinguished from the background signal. Acidic washing after some separation steps, released tightly bound unmodified peptides and thereby reconstituted the capacity of the C18-HPLC column.

Stability of ARG – phosphorylation

As basis for an enrichment protocol, I determined the limitations of acidic treatment in a simple hydrolysis experiment. Therefore, the arginine phosphorylated model peptide ac-K**R**GGGGYIKIIKV was incubated in a phosphate buffered solution at different pH values and temperatures. The formation of the unphosphorylated product was detected by LC-MS for different time points. All datapoints were measured at least in triplicates to obtain solid statistics. Ammonium phosphate was chosen as reaction buffer, since all reactions should be carried out with the same chemical background. Furthermore, it was easily removed on the trapping column prior to HPLC-MS analysis and provided good buffering capacity over almost the complete pH range. On the other side, high phosphate concentration could influence the reaction, since phosphate is one of the reaction products so that the obtained hydrolysis rates might be lower in comparison to other buffer systems.

First, contribution of nano-HPLC separation to the hydrolysis was determined by 10 consecutive injections of freshly purified peptide and separation with the optimized HPLC method. The abundances of non-phosphorylated and phosphorylated peptide were determined from the mass spectrometric data by their respective peak areas for the doubly and triply charged peptide ions. In average, the hydrolyzed peptide was found at approximately 7 % intensity with little deviation between all runs. To avoid hydrolyzing HPLC conditions, I also studied the elution behaviour of both peptides at pH 4.5 and 7 in comparison to the standard separation conditions at pH 2.6. Under standard nanoHPLC conditions both peptides were observed as sharp peaks with the phosphopeptide eluting approximately 40 seconds after the unmodified form, whereas at pH 4.5 I saw a broadening of both elution profiles and a reduction of the elution time difference. This effect could be due to a change in charge states of carboxylic acid and phospho-amidate moieties, which then interact differently with the solid phase and form broad peaks. At pH 7 the elution order changed, so that the phosphorylated peptide eluted before the unmodified form. The unmodified peptide could not be detected in the pure phosphopeptide sample, but indvidual injection of this peptide revealed a broad elution over more than 5 min even at short gradient times. This behaviour agreed with the observation in the purification step, where only the phosphorylated form of the model peptide ac-K**R**GGGGYIKIIKV could be isolated in a distinct fraction. Despite small losses due to hydrolysis, we decided to apply the acidic standard nano-HPLC-MS conditions to analyze the stability data, since only then both peptide forms were reliably detected and the peak shapes allowed quantification on basis of the UV-chromatogram and the MS elution profiles.

Incubation under strong acidic conditions (pH 1 or 2) led to rapid hydrolysis of the phospho-amidate bond at arginine with a half life of approx. 60 min at 25°C (figure B_11A). At increased temperature (60 °C) this effect was even more severe, so that the phosphate moiety was completely lost within the first 15-25 min (figure B_11B). Furthermore, also acidic hydrolysis of the peptide was observed, which led to a slight increase of the curve at longer reaction times, while the background signal of the phosphopeptide elution window remained stable. Further increase of the temperature to 95°C led to a loss of the phospho-amidate within the first 5 min and significant peptide hydrolysis for long reaction times. Similar experiments with the phospho-Ser peptide ac-K**S**GGGGGYIKIIKV and the

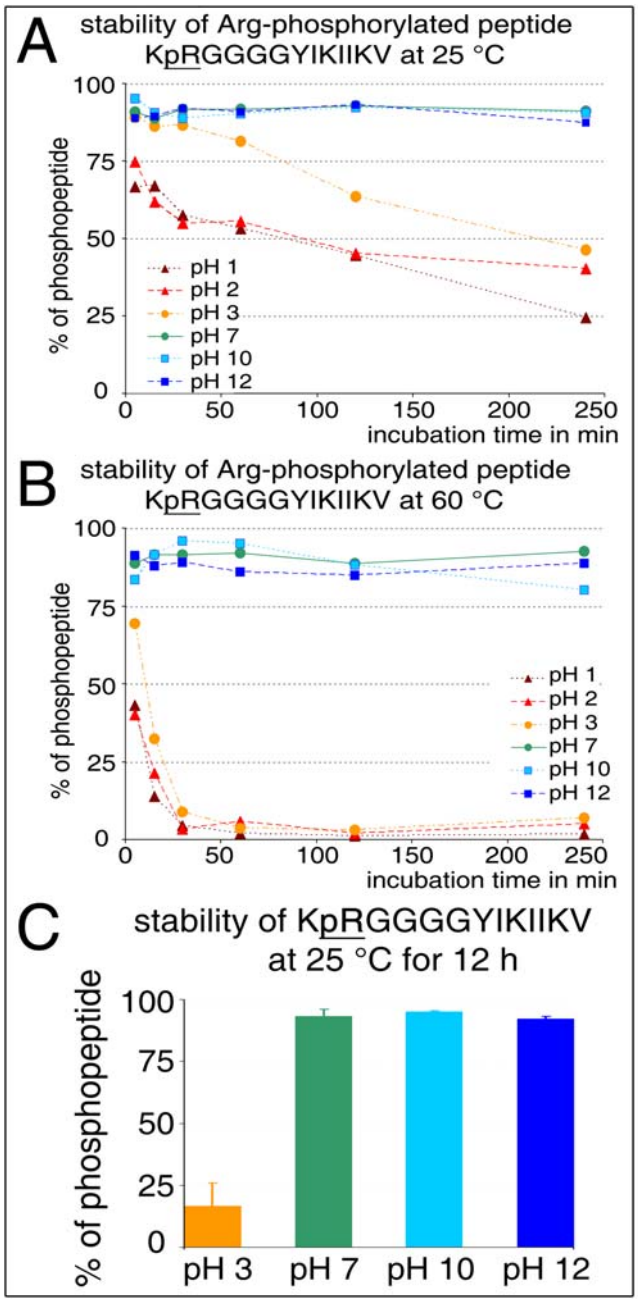


Figure B_11 **pH/Temperature Stability of the Arginine-Phosphorylation.** A. At room temperature, the phospho-amidate bond hydrolyses slowly under acidic conditions but is completely stable above pH 7. Moderate acidic pH (pH 3) exhibits significant slower hydrolysis rates. B. Elevated temperatures accelerate the hydrolysis under strong and moderate acidic conditions dramatically. High stability is again observed for neutral and alkaline pH conditions. C. Acidic hydrolysis over 14 h leads to loss of approx. 85% of the phosphopeptide at room temperature. In contrast, almost no hydrolysis is observed for pH conditions above pH 7.

phospho-Tyr peptide ac-K**Y**GGGGYIKIIKV showed no significant loss of the phosphorylation even after 4 h of incubation. Only little hydrolysis was observed at pH 3/25°C within the first 30 min (figure B_11A). At longer reaction times, the signal of the phosphopeptide rapidly decreased, thereby demonstrating also instability under moderate acidic conditions. At pH 3/60°C the phospho-amidate bond was hydrolyzed rapidly within the first 15 min, similar to pH 1 and pH 2 (figure B_11B).

Incubation at 25 and 60°C under neutral pH conditions exhibited only basal hydrolysis during the HPLC separation (figure B_11A+B), while raising the temperature to 95°C increased the hydrolysis rate and resulted in a half life of approximately 150 min (Fuhrmann J. *et al.*, 2009). As demonstrated in (figure B_11A+B), highest stability is observed for moderate alkaline conditions; at pH 10 or pH 12. The percentage of unmodified peptide did not change dramatically over the applied incubation period and was independent from the temperature. Incubation of samples at pH 14 impaired HPLC-MS/MS analysis of longer timepoints, but the modification was still intact after short reaction times.

To test the long term stability, as for instance in proteolytic cleavage, we performed hydrolysis of the phosphopeptide at 25 °C for 14 h and measured the decay of the phosphorylation at pH 3, pH 7, pH 10, and pH 12. In comparison to higher pH values, where no significant hydrolysis was detected, the amount of phosphopeptide was significantly decreased at pH 3 and the unmodified peptide was present to 85% (figure B_11C). This observation explains why samples with phosphorylated arginine should not be stored after acidification. Incubation of the sample at pH 7 for 5 days led to no significant formation of unmodified peptide, which is in agreement with the literature (Besant P.G. *et al.*, 2009).

Apart from the abundance, the acidic hydrolysis is one explanation why arginine phosphorylation was seldomly detected in recent phospho-proteomic studies, because almost all sample separation steps following the enzymatic cleavage were carried out at low pH. Strong cation exchange separation, for instance, is commonly performed at pH below 3, while for metal oxide-affinity chromatography on TiO_2 or ZrO_2 even lower pHs are applied in the loading and washing steps. As adaptation to this effect, we aborted the enzymatic digest by freezing the sample instead of adding formic acid as foreseen by many protocols. Further, we changed the HPLC loading buffer on the pre-column for LC-MS/MS analysis to a mixture of 0.1% TFA, 0.05% HFBA and adjusted the pH to 4 with ammonia, because especially the extended sample wash for 40 min after TiO_2 enrichment led to undesired hydrolysis. Extending the wash on the pre-column was necessary to remove the phosphate buffer, which would otherwise impair mass spectrometry by formation of large phosphate clusters.

Tryptic activity at phosphorylated arginine residues

Since phosphorylation completely inverts the charge of the guanidyl moiety in the amino acid side chain of arginine, tryptic cleavage at the modified residue might be limited. For this experiment, we acetylated lysine side chains and the N-terminal amino acid of the model peptide YIVESKRGGGGYIKIIK before we set up the kinase reaction, because we were not sure, if the arginine phosphorylation is maintained upon acetylation (figure B_12A). Phosphorylation with McsB was performed as described to give the peptide ac-YIVESK(ac)**R**GGGGYIK(ac)IIK(ac). The phosphorylated peptide was separated from the not phosphorylated peptide by RP-separation at pH 9 as described above. A fraction of the phosphopeptide and a fraction of the unphosphorylated peptide were incubated with trypsin in a 50:1 ratio (peptide/trypsin) at 37°C for 14 h. The efficiency of the tryptic digest was determined after LC-MS/MS analysis from the peak areas for the full length peptides and respective cleavage products. All peptide forms were identified manually based on their fragment ion spectra.

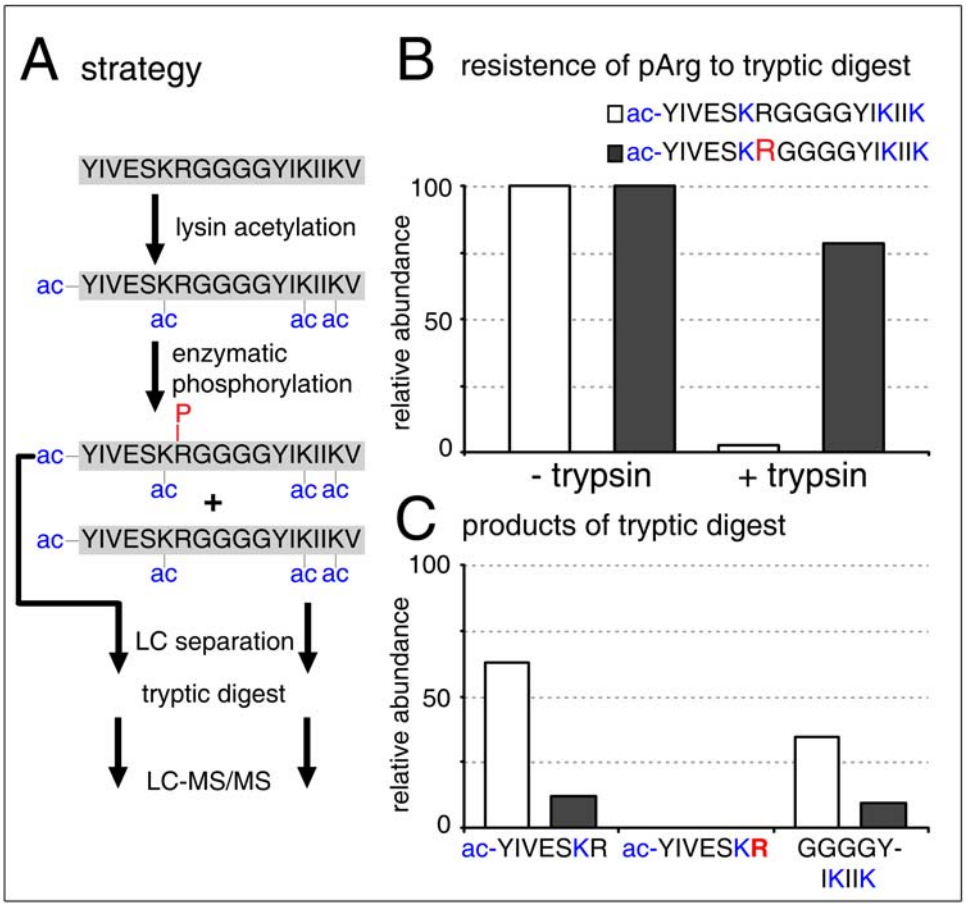


Figure B_12 **Trypsin is Not Able to Cleave at Phosphorylated Arginine**. A. Strategy to study tryptic activity at phosphorylated arginine, other putative targets of trypsin were blocked by acetylation prior to phosphorylation with McsB. Phosphorylated and unphosphorylated peptide forms were separated and individually incubated with trypsin over

night. B. Almost 80% of the phosphorylated precursor remains after tryptic digestion, while the unphosphorylated peptide is completely cleaved, demonstrating that trypsin is not able to recognize phosphorylated arginines. C. A phosphorylated cleavage product was not identified, but low amounts of unmodified cleavage products were observed in the phosphopeptide sample. This suggests that trypsin can only cleave after hydrolysis of the phospho-amidate.

Trypsin cleaves the unphosphorylated peptide completely, so that only the N-terminal and the C-terminal peptide fragments are detected. The phosphopeptide resists tryptic cleavage almost completely (figure B_12B+C). Nevertheless, unphosphorylated tryptic fragments were also identified after enzymatic cleavage of the phosphopeptide. These fragments resulted from the cleavage of the unmodified peptide, which is present at very low amounts or could be formed by phospho-amidate hydrolysis during digestion (figure B_12C). Since I did not detect a phosphorylated fragment in any trypsin treated phosphopeptide sample, I concluded that trypsin is not able to cleave at the modified arginines.

Enrichment of ARG-phosphorylated peptide using TiO_2-MOAC

As the stability experiment demonstrated (figure B_11A), strong acidic conditions led to a rapid hydrolysis of the phospho-amidate, thereby precluding the application of the previously described TiO_2 enrichment protocol. To utilize the features of TiO_2 enrichment, the pH of the loading and wash solvents has to be adjusted carefully, to minimize acidic hydrolysis. Moderate acidic conditions revealed higher stability for 30-60 min, which is long enough for phosphopeptide enrichment on TiO_2.

The loading buffer was made up from 40% acetic acid (AcOH) that contained 0.4% HFBA and was adjusted to pH 4.5 with ammonia. Per 1 ml of loading buffer 600 mg DHB were added to the AcOH solution and dissolved by carefully adjusting the pH in a range of 4-5. This mixture was further diluted by the same volume of ACN to give a 300 mg/ml DHB solution in 50% ACN, 20% AcOH and 0.2% HFBA. To remove unphosphorylated peptides in the first washing step, ACN was added to the loading buffer up to 65% ACN, 200 mg/ml DHB, 13% AcOH and 0.13% HFBA. The second wash solvent contained 75% ACN, 10% AcOH and 0.1% HFBA and is required to remove DHB. Phosphopeptide elution was achieved with 30 mM $(NH_4)_2HPO_4$ at a pH of 10.5. To further minimize hydrolysis, the sample was mixed with the loading buffer after column preparation and immediately before enrichment. First, 1 pmol the previously described standard phosphopeptide mixture A (table A_01) was spiked into 4 pmol of a tryptic digest of bovine serum albumin (BSA). This

sample was treated with the novel protocol to enrich phosphopeptides and deplete peptides from bovine serum albumin. LC-MS/MS analysis of the untreated sample and the eluted fractions was performed on an LCQ-XP ion trap mass spectrometer. MS/MS spectra and retention times were used to identify the phosphopeptides, since retention times were highly reproducible for all samples. Peak areas of the standard peptides in their most abundant charge states and 16 highly abundant tryptic BSA peptides were assigned manually after filtering for the precursor mass in a 1.6 Da mass window to determine the enrichment efficiency.

For almost all phosphopeptides, we observed higher signal intensities in comparison to the untreated standard mixture, which demonstrated strong enrichment by the method (figure B_13A). The average signal area of phosphorylated peptides was increased by a factor of 1.66. This increase is possibly due to reduced ionization suppression in the enriched sample, since coeluting unphosphorylated peptides from BSA are removed. Although two phosphopeptides had reduced signal areas, peptides with multiple phosphorylations were up to 2.5 fold more intense without the BSA background. Depletion of unphosphorylated peptides was monitored on peptides NoP1 and NoP2 of mix A and 16 highly abundant BSA peptides that were more than 3 times identified by MASCOT in the MS/MS dataset of the untreated sample (figure B_13A). Noise peaks significantly contributed to the signal area of small signals, as for instance for quantification of BSA peptides after enrichment. In average, unphosphorylated peptides were removed to 87% for all BSA peptides, but three peptides with more than four acidic amino acids in the sequence remained with relatively high intensities of 15-70%. Subtraction of these three acidic peptides reduced the average percentage of retained non phosphorylated peptides to 7%.

To test the protocol for enrichment of arginine phosphorylated peptides 1 pmol of the standard peptide ac-K**R**GGGGYIKIIKV (PepR_02) was added to 2 pmol tryptic BSA digest and the TiO_2 enrichment was performed. Untreated and TiO_2-enriched samples were analyzed in triplicates by HPLC-MS/MS on an Ultimate 3000 HPLC system, coupled to an LTQ-OrbitrapXL mass spectrometer. Peak areas of the eluting phosphopeptide were determined with a mass accuracy of 5 ppm and compared for both sample types. These data revealed again a slight increase of the phospho-peptide signal after enrichment by a factor of 1.3 (figure B_13A). Furthermore, these data clearly show that peptides with a very basic sequence and arginine-phosphorylation are efficiently trapped on TiO_2. The phospho-amidate moiety is not extensively hydrolyzed in the loading step, during sample washing, and HPLC separation.

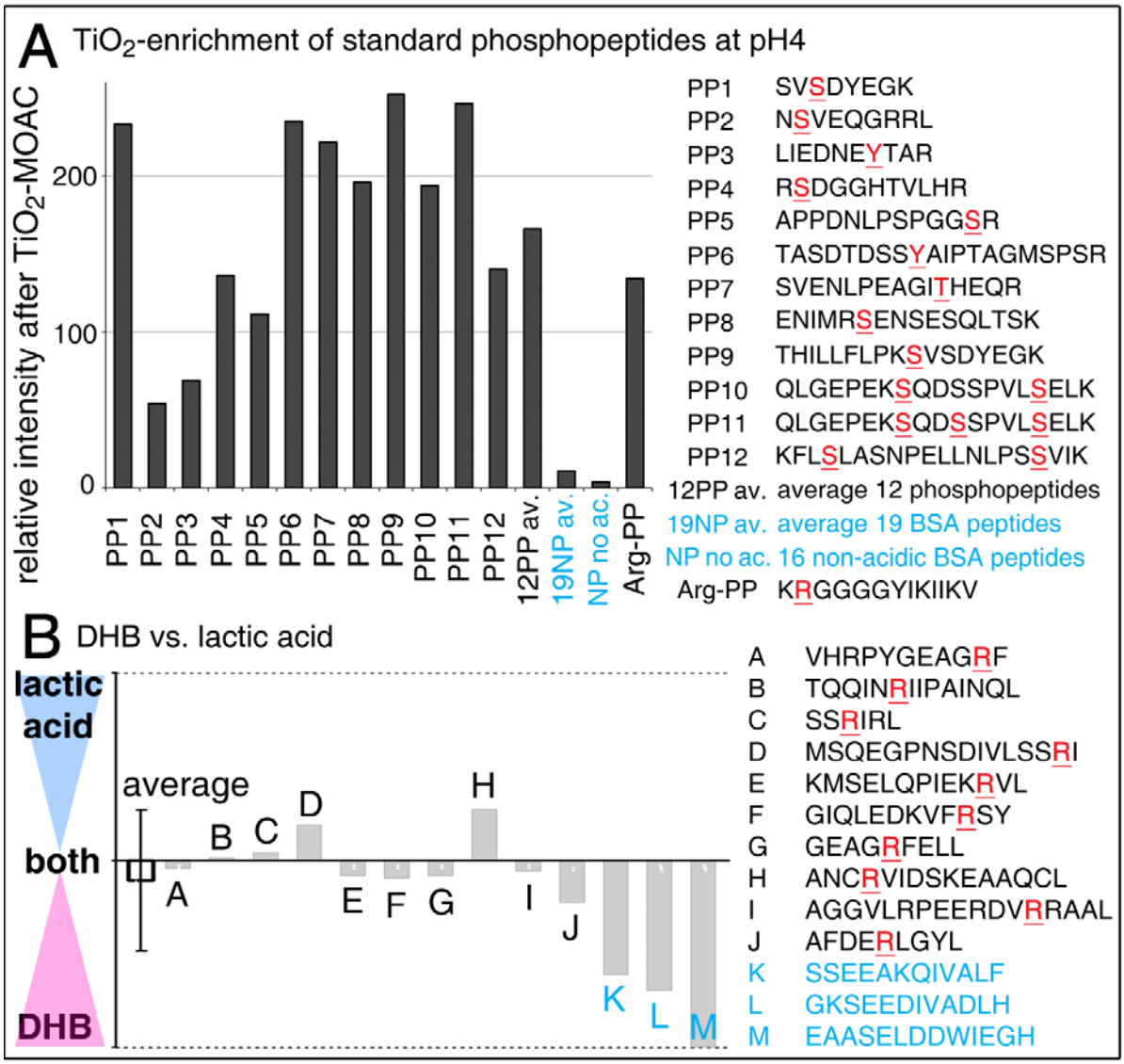


Figure B_13 **TiO_2-Enrichment of Peptides with Acid Labile, N-Bound Phosphorylations**. A. Feasibility of the novel enrichment protocol. Unphosphorylated peptides from 4 pmol BSA are reliably depleted in the loading and washing steps, but peptides with multiple acidic groups are still retained (light blue). Standard phosphopeptides (1 pmol) are retained and efficiently eluted under alkaline pH conditions. The increase in signal intensity after enrichment is mainly due to reduced ionization suppression effects. B. Down-facing bars represent peptides that have higher signal intensity after DHB enrichment; up-facing bars are better retained in the lactic acid method. Arginine-phosphorylated peptides are equally well enriched, when DHB is replaced by lactic acid as masking substance in the washing step, but unphosphorylated peptides are more efficiently depleted by the lactic acid protocol.

Comparison of DHB and lactic acid as replacement agents

Side effects of DHB in the loading buffer were huge background signals in the UV chromatogram and in mass spectra, which stem from impurities of DHB and interfere with

quantification of phosphopeptide signals. As described above, peptides without phosphorylation but many acidic amino acids in their sequence are only poorly depleted upon TiO_2 enrichment with the DHB method. To circumvent laborious washing of the sample or even DHB re-crystallization prior to enrichment, I tested, if lactic acid provides equal enrichment quality. This reagent was earlier described as masking reagent for MOAC and is also included in commercial phosphopeptide enrichment kits (Jensen S.S. and Larsen M.R., 2007). In comparison to DHB, lactic acid has a lower potential for crystallization and can be effectively removed by the sample washing on the HPLC trapping column. We prepared the loading buffer with 300 mg/ml lactic acid instead of DHB as was the washing buffer A and performed the TiO_2 enrichment with equal amounts (2 pMol) of chymotryptic digest of McsB kinase after 30 min autophosphorylation. After LC-MS/MS analysis and database search with Protein Discoverer (Vs 1.1.0.242), we filtered the total ion current for mass traces of identified phosphopeptides and their unphosphorylated counterparts. By comparing the sum of peak areas from different charge states and of individual peptides, I demonstrated that lactic acid has a slightly lower efficiency to trap phosphopeptides (FIG B_13B). Interestingly, many peptides with a methionine residue in their sequence were found in oxidized form in the DHB treated sample, while lactic acid enrichment showed lesser oxidation of methionines. This uneven oxidation of both samples lead to the effect that individual peptide species were stronger represented in the DHB sample, because the peptide signals were not split over different oxidation states. This observation suggests a general oxidation step before phosphopeptide enrichment, which completely converts methionine to methionine sulfoxide and therefore increases sensitivity for such peptides. Focussing on the detected phosphopeptides, the difference of DHB and lactic acid was marginal and both reagents guaranteed stability of the arginine-phosphorylation. Moreover, of all identified unphosphorylated peptides that were identified by both enrichment methods the signal intensity was significantly lower for the lactic acid protocol. Even the strong acidic peptides, which were not depleted by DHB, were effectively removed when lactic acid was applied in the enrichment step (FIG B_13B). Since lactic acid efficiently prevents binding of acidic unphosphorylated peptides and can be effectively removed by washing, it might be the better choice for complex samples such as proteolytic digests of cell lysates and tissues.

Phosphosite Assignment on McsB and CtsR

Earlier publications report autophosphorylation of McsB, next to phosphorylation of its target protein CtsR in vitro (Kirstein J. *et al.*, 2005). These studies have been conducted with

proteins from B.subtilis, whereas we detected the arginine-phosphorylation in proteins from B.stearothermophilus. The amino acid structure of both CtsR variants is highly conserved around the previously identified phosphorylation sites, pointing to evolutionary conservation of the protein activity. Previous experiments with small model peptides raised the question, if autophosphorylation of McsB might have an inhibitory effect on McsB activity. When we tried to phosphorylate the small model peptides, only low amounts of the desired product were formed and longer incubation times did not significantly increase the percentage.

Our collaborators established the protein systems from both bacteria, *B.subtilis* and *B.stearothermophilus, in vitro*. To study the phosphorylation of the proteins we incubated CtsR from both bacteria individually with their respective kinase McsB and ATP for 30 min. Furthermore, we also tested the phosphorylation of *B.subtilis* CtsR by McsB from the thermophilic organism. After incubation we performed chymotryptic digestion of the reaction mixture and enrichment of phosphopeptides from 2 pMol protein on TiO_2. Eluted peptide fractions were loaded onto a C18 trapping column at pH 4.5, washed for 40 min, and subsequently separated by nanoRP-HPLC prior to online detection and fragmentation in the mass spectrometer. Peptide fragmentation was performed with three different MS/MS methods: multi-stage activation CID (pseudo-MS^n), higher-energy collisional dissociation (HCD) and electron-transfer dissociation (ETD). For each method two individual phosphopeptide samples were prepared. From these three methods, ETD covered the lowest number of phospho-sites, which is due to the small peptide size, achieved by chymotryptic cleavage (FIG B_14). This leads mainly to peptides with 2+ charge state, of which ETD fragmentation often results in low sequence information and therefore such spectra are lost in the automated database search. On triply charged ions phospho-sites were assigned very accurately by this method, suggesting to repeat the analysis with a different protease. HCD-MS/MS analysis proved to identify the highest number of phosphopeptides but localized the modification site to other amino acid residues in 1/3 of all identified peptides. Wrong assignment of phosphorylated amino acids was even amplified upon fragmentation by the pseudo-MS^n method, where only 1/3 of the hits were assigned to arginine phosphorylation.

Interestingly, searching ETD spectra with MASCOT led to the lowest identification of Arg-phosphorylated peptides, while Sequest reported this modification on almost all phosphopeptides. This disagreement of both search engines was only observed with ETD data, where MASCOT might have a bias towards identification of phospho-esters. In general, Sequest reported slightly more phosphopeptides than MASCOT, but identification scores were often quite low, so that manual inspection of the spectra was necessary.

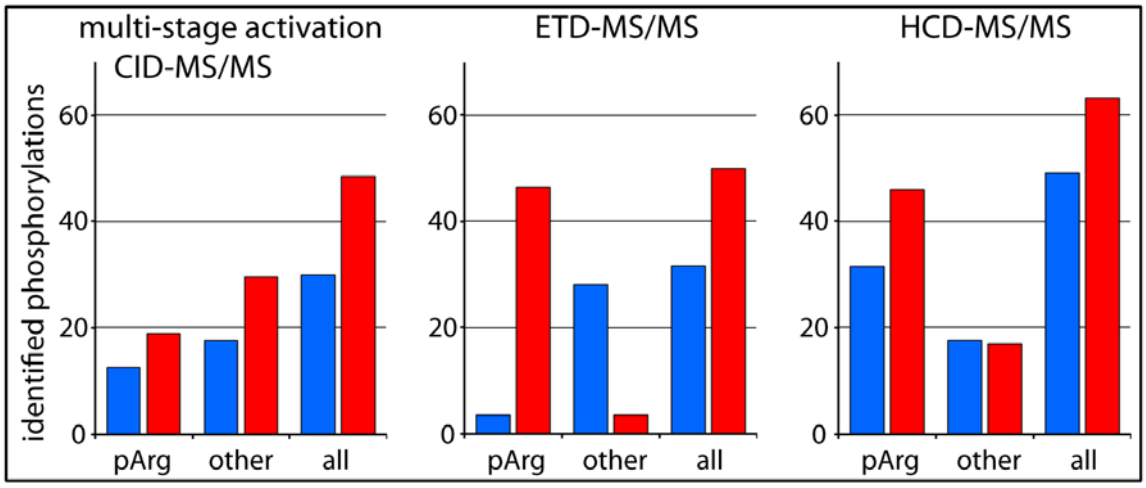


Figure B_14 **Optimization of the LC-MS/MS Identification Process for Arginine-Phosphorylated Peptides**. Three MS/MS principles were tested for their ability to identify arginine phosphorylated peptides from chymotryptic digested CtsR and McsB (*B. stearothermophilus*). The obtained data were searched against an "in house" database with MASCOT (blue) and Sequest (red) search algorithms. Multi-stage activation CID and ETD MS/MS methods identify less phosphopeptides than HCD. Only 1/3 of all phosphorylations were assigned to arginine from multi-stage activation CID-MS/MS data. Since other phosphorylations than phospho-Arg can be excluded from the sample preparation, identification of Ser, Thr, Tyr, or His modification resulted from ambiguous assignment by the search engine. Interestingly, only Sequest reliably identified the Arg-phosphorylation from ETD data, while MASCOT was always biased towards phospho-esters. HCD MS/MS results in approx. 30% wrong identification of the phosphorylated amino acid. Sequest delivers more identified phosphopeptide hits than MASCOT.

Combined results of all analyses revealed similar phosphorylation patterns of the N-terminal domain of both CtsR variants (Figure B-15), whereas one additional phosphorylation site was localized to the *B.subtilis* CtsR C-terminal domain. All arginines are involved in protein-DNA interactions and thus the complex is destabilized upon phosphorylation. Since chymotrypsin cleaves at hydrophobic amino acids, often resulting in a complex cleavage pattern, most phosphorylation sites were found on at least two overlapping peptide sequences (table TB_4). A benefit of the acidic HPLC separation was the formation of unphosphorylated peptides that were separately detected and identified. The high sequence coverage of the unmodified peptides allowed direct comparison of MS/MS spectra between unphosphorylated and phosphorylated peptides and confirmed their identification. We found 15 arginine-bound phosphorylations in McsB, representing the majority of arginines in this protein (Figure B-15

and table B-04). In comparison to CtsR, where the phosphorylated arginines impair complex formation, the phospho-sites of McsB are widespread over the complete protein, which points to random phosphorylation. Interestingly, also arginine residues that were predicted to mediate the phospho-transfer in the active site or directly bind the ATP were found in phosphorylated form, which might be an explanation for the reduced activity at long reaction times.

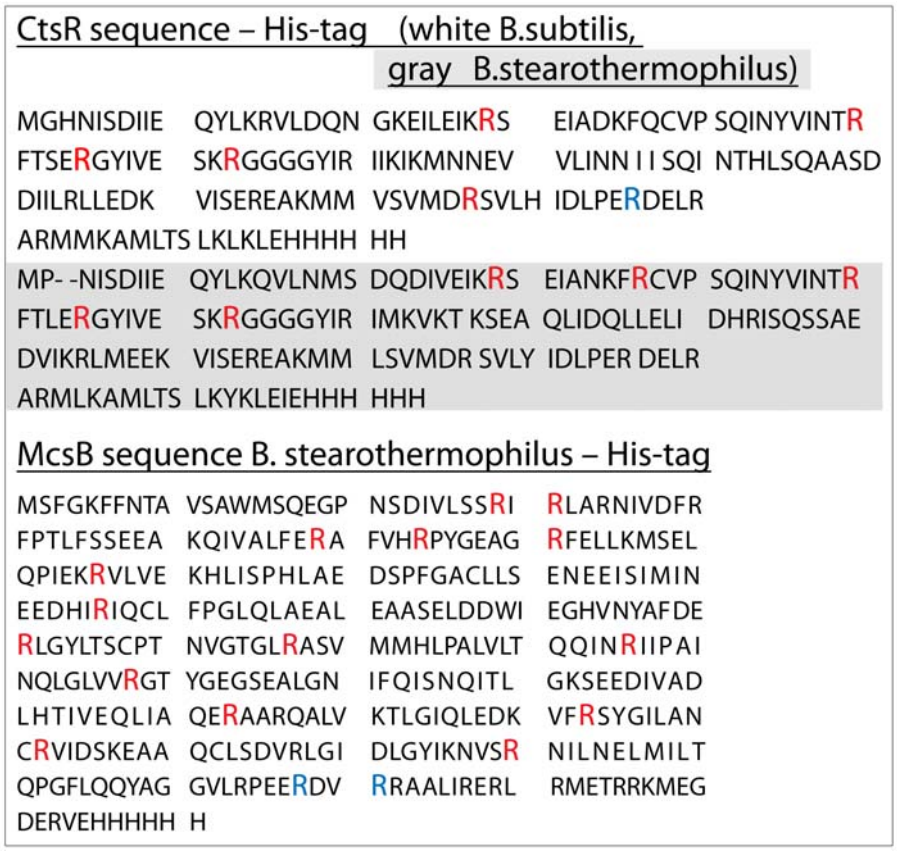


Figure B_15 **Identified Arginine Phosphorylations in CtsR and McsB**. Two variants of CtsR from *B.stearothermophilus* and *B.subtilis* were incubated with the respective McsB variants and phosphorylation sites were identified by LC-MS/MS. Most of the phosphorylations were found in identical positions of both proteins. Furthermore, cross reactivity of McsB (*B.stearothermophilus*) with CtsR (*B.subtilis*) revealed the same phosphorylation sites, thereby pointing to an evolutionary conserved mechanism of CtsR regulation. The kinase itself shows extensive autophosphorylation, which is depicted by 15 different phosphoarginines. Red arginines are confirmed phosphosites, blue arginines were only ambiguously assigned in case of McsB or the spectrum quality was questionable.

Table TB_4 **Phosphopeptides of McsB after Chymotryptic Cleavage**. Peptides in the region 329-345 show ambiguous assignment of phosphosites, isoforms of the other peptides with two different phosphorylation sites were chromatographically separated and could be assigned to one of the two forms.

	aa		Position	no phos	phos		phos	phos
Sequence	start	end	of Phos	m/z	m/z phos	MH+	ΔM ppm	RT min
AFDERLGY	157	164	R4	485.735	525.718	970.463	-0.29	30.54
AFDERLGYL	157	165	R4	542.277	582.260	1083.547	0.17	34.88
AFDERLGYLTSCPTNVGTGL	157	176	R4	1086.026	1126.010	2171.045	0.75	36.31
AGGVLRPEERDVRRAAL	329	345	R10 o R13	622.353	649.009	1865.045	-0.87	32.42
AGGVLRPEERDVRRAAL	329	345	R10 o R14	933.026	973.009	1865.045	0.41	35.13
ANCRVIDSKEAAQCL	279	293	R4	867.917	907.900	1734.827	0.26	26.28
ELLKMSELQPIEKRVL	73	88	R14	971.556	1011.539	1942.104	0.37	32.97
GEAGRFELL	67	75	R5	496.264	536.247	991.521	0.57	34.70
GILANCRVIDSKEAAQCL	276	293	R7	673.344	699.999	2018.016	0.14	31.69
GILANCRVIDSKEAAQCL	276	293	R7	1009.512	1049.495	2018.016	0.14	31.69
GIQLEDKVFRSY	264	275	R10	727.885	767.869	1454.763	-0.27	32.53
GLVVRGTYGEGSEALGNIF	204	222	R5	970.000	1009.983	1938.992	-0.47	38.87
IAQERAARQAL	249	259	R8	613.852	653.835	1226.696	-0.20	23.02
IKNVSRNILNELM	305	317	R6	780.432	820.415	1559.857	-0.01	33.79
INEEDHIRIQ	119	128	R8	633.825	673.808	1266.643	-0.69	25.45
KMSELQPIEKRVL	76	88	R11	793.950	833.934	1586.893	1.12	29.66
LTSCPTNVGTGLRASVM	165	181	R13	593.627	620.282	1778.865	2.47	28.55
LTSCPTNVGTGLRASVMM	165	182	R13	642.638	669.294	1925.900	0.34	28.34
MINEEDHIRIQ	118	128	R9	707.343	747.326	1413.678	-0.63	26.14
MSQEGPNSDIVLSSRI	15	30	R15	866.931	906.914	1732.855	0.63	31.91
MSQEGPNSDIVLSSRIRL	15	32	R15 o R17	1010.147	1050.130	2018.034	0.89	34.48
NVGTGLRASVM	171	181	R7	560.792	600.776	1120.578	-0.27	24.61
RIIPAINQL	195	203	R1	519.327	559.310	1037.646	-0.22	34.48
RIIPAINQLGL	195	205	R1	604.379	644.363	1207.752	-0.45	37.98
RPEERDVRRAAL	334	345	R5 o R8	734.411	774.394	1467.814	-0.93	23.65
SCPTNVGTGLRASVM	168	181	R11	783.375	823.358	1565.743	1.02	26.21
SSEEAKQIVALFERAF	46	61	R14	912.978	952.961	1824.949	0.44	39.00
SSRIRL	27	32	R3 o R5	366.230	406.213	731.452	0.32	27.40
TQQINRIIPAIN	190	201	R6	690.901	730.885	1380.796	-0.09	32.59
TQQINRIIPAINQL	190	203	R6	811.473	851.457	1621.940	0.63	36.33
TSCPTNVGTGLRASVM	166	181	R12	556.268	582.923	1666.788	-0.26	26.46
VHRPYGEAGRF	62	72	R3 o R10	644.830	684.813	1288.653	-1.42	25.88
VSRNILNELM	308	317	R3	594.824	634.807	1188.640	-0.16	32.84
VVRGTYGEGSEALGNIF	206	222	R3	884.948	924.931	1768.888	0.69	35.84

d Fragmentation of peptides with arginine-phosphorylation

In addition to the biochemical properties, also the fragmentation behaviour in MS of peptides with arginine phosphorylation differs significantly from that of other phosphorylated amino acids. Upon thermal activation, many phosphopeptides undergo a process in which a neutral phosphoric acid molecule is eliminated from the peptide (Palumbo A.M. and Reid G.E., 2009). Since this process is almost quantitative, MS/MS spectra exhibit only a single intense fragment ion of the precursor M-H_3PO_4. The H_3PO_4-elimination has been extensively studied and for the phospho-esters of Ser and Thr several reaction mechanisms have been described on the base of a β-elimination reaction similar to the formation of a dehydro-amino acid in solution at high pH (Boersema P.J. *et al.*, 2009). The influence of the position of the phosphorylation and charges in the vicinity are not yet fully understood and currently heavily discussed. Since this reaction is energetically preferred in comparison to the dissociation of the amide bond in the peptide backbone, only few fragment ions are observed upon low energy thermal activation such as CID-MS/MS in ion trapping mass spectrometers. In contrast, phospho-Tyr is not able to eliminate the H_3PO_4 molecule, since the aromatic side chain impairs β-elimination of hydrogen and formation of a triple bond (Boersema P.J. *et al.*, 2009). Therefore, peptides with this modification often exhibit fragment ions with the intact modification and losses of the phospho-moiety HPO_3 (-79.966 Da) are rarely observed. Furthermore, this modification can still be detected as immonium ion of tyrosine, which allows identification of phospho-Tyr in peptides by screening for a signal at 216.043 Da (Steen H. *et al.*, 2001, see figure A_6C).

Completely surprising was the finding that phospho-His also eliminates H_3PO_4 under low-energy CID conditions, which can not be explained by the chemical structure of a nitrogen-attached phospho-moiety (Medzihradszky K.F. *et al.*, 1997, Zu X.L. *et al.*, 2007). Detailed MS/MS studies demonstrated that the water molecule stems from a serine residue next to the phospho-His in the peptide sequence. The obtained data point to a concerted fragmentation mechanism: first, the phospho-moiety is transferred onto the OH-moiety of the serine from where it is subsequently eliminated as H_3PO_4 (Zu X.L. *et al.*, 2007).

In previously mentioned CID-MS/MS studies of the CtsR protein a similar fragmentation was observed, where peptides with arginine phosphorylation are able to undergo neutral loss of phosphoric acid. Similar to phospho-His, the chemical structure reveals an N-P bond and is therefore not able to explain elimination of H_3PO_4. Since the arginine side chain does not posses further free OH moieties that would support the H_3PO_4

loss, the required OH-groups have to be contributed by other amino acids such as Ser, Thr, Asp or Glu. We studied the ability for phosphoric acid loss and backbone fragmentation in comparison to other peptide phosphorylations on Ser, Thr, Tyr, and His and implemented specific amino acids into the peptide sequence that are able to serve as OH-donors.

Preparation of Peptides with Histidine-Phosphorylation

In order to compare the fragment ion spectra of peptides with phospho-Arg with spectra from other types of phosphorylation, I established chemical phosphorylation of histidines with potassium phosphoramidate following the method described by Wei and Matthews (Wei Y.-F. and Mattews H.R., 2001). Short, phosphorous oxychloride was converted to ammonium phosphoramidate by reaction with an aqueous solution of 15% ammonia at 0°C. The pH of the solution was lowered to 6 with acetic acid and the organic salt was extracted with acetone. Ammonium phosphoramidate was crystallized at 0°C and subsequently crystals were washed with ethanol and diethylether. To expel the ammonia, the crystalline product was dissolved in 50% KOH solution and stirred at 60° until ammonia formation stopped. The reaction mixture was again acidified and extracted with ethanol and crystallized at low temperature. The product was washed again with ethanol and diethylether and crystals were dried under vacuum for 24 h.

The peptide ac-KHGGGGYIKIIKV was diluted in a HEPES buffered solution at pH 8 and twenty fold excess w/w of phosphoramidate was added to the solution. Phosphorylation was carried out at 30°C for 2h and the reaction mixture was directly separated by semi-preparative HPLC at pH 9. Fractions, containing the phosphopeptide, were acidified and directly infused into the mass spectrometer by static nano-electrospray ionization. Since some experiments demonstrated significant loss of the modification upon freeze-drying, the exact amount of dry phosphopeptide could not be weighed. Storage for 24 h at low temperatures -80°C, -20°C, or 4°C - did not tremendously affect the stability of the modification, while hydrolysis was already observed, when the sample was kept at room temperature.

Fragmentation conditions

MS/MS fragmentation studies were carried out on an LTQ-Orbitrap XL ETD, an LTQ-Orbitrap Velos ETD mass spectrometer (both Thermo-Fisher, Bremen), and an 4800 MALDI-TOF/TOF Analyzer instrument (LifeTechnology, Darmstadt). Peptides were diluted to 10 mM concentration with H_2O and further to 5 µM with 70% ACN, 0.2% FA for offline-nanoESI experiments or 70% ACN, 0.1% TFA for MALDI ionization. The MALDI samples

were co-crystallized on the sample plate with 10mg/ml DHB in 70% ACN, 0.2% TFA, 0.1% H_3PO_4. To avoid extensive acidic hydrolysis, all peptide samples were analyzed by MS within 30 min after the acidic dilution and crystallization.

For ESI experiments, peptides were mounted into static nano-ESI-Emitters ES380 (Proxeon, Denmark) and directly infused into the Orbitrap mass spectrometer with a capillary voltage of 800 V. Fragmentation experiments were carried out on the doubly and triply charged peptide ions. Low energy CID MS/MS and MS^3 experiments were carried out at normalized collision energy (CE) of 35% and activation time of 35 ms. For HCD fragmentation, conditions were adjusted to leave a precursor ion intensity of 15 - 30 %. Peptides PEPR_19 to 23 were fragmented with an activation time of 10 ms and optimized collision energy for each peptide, approx. 25-28%. Prior to ETD experiments, the ETD source was tuned to optimal emission of fluoranthen anions. All fragment ion spectra were analyzed in the Orbitrap mass analyzer at a resolution of 60000 to allow unambiguous identification of the respective fragment ions. For MALDI MS 5000-10000 subspectra were recorded and summed up for the final mass spectrum and MS/MS spectra were acquired at elevated gas pressure in the collision tube.

Data Interpretation

All tandem MS spectra were manually evaluated. For ESI MS/MS a mass accuracy of 0.01 Da was applied, MALDI data were searched with a mass tolerance of 0.1 Da. Fragment ions were calculated with the GPMAW software Vs. 7.0 (www.gpmaw.com, Odense, Denmark). Fragment ion intensities were used to compare the fragmentation behaviour of differently phosphorylated peptides.

Stability of the phospho-amidate in ETD-MS/MS

The phospho-amidates of His are completely stable during ion-electron reactions in ETD processes and previous studies with arginine-phosphorylation also demonstrated the stability (Kleinijenhuis A.J. *et al.*, 2007; Fuhrmann J. *et al.*, 2009). This observation makes ETD-MS/MS a very important tool for the identification of such modifications. On the other side, fragmentation efficiency of this method varies dramatically for different charge states and also depends on the amino acid sequence, therefore limiting general application of this fragmentation technique. Another limiting factor associated with amino acid phosphorylation is the reduced peptide charge, since phospho-moieties add a negative charge to the peptide. This effect reduces the occupation of higher charge states, which are required for ETD MS/MS.

The ETD spectrum of the model peptide K**R**GGGGYIKIIKV in figure B_16A is obtained from the triply charged precursor at 504.2884 m/z. Apart from the C-terminal valine, all amino acids are covered by singly charged c- and z-type fragment ions and for the largest ions also doubly charged fragments were detected. None of the fragment or molecular ion species reveals a neutral loss of the phospho-moiety. Supplemental activation, an additional CID activation at very low energies adjacently to the ETD activation, does not induce neutral loss of H_3PO_4, but improved fragmentation was also not observed.

Similarly to arginine phosphorylation, the peptide with phospho-His ac-K**H**GGGGYIKIIKV formed fragment ions with intact phosphorylation upon ETD activation (figure B_16C). Nevertheless, some fragments with low intensity represent a different location of the phospho-moiety to the C-terminal lysines (data not shown). This reaction was not described in the original publication, but was recently studied by Kowalewska et al. and typically requires longer reaction times of approximately 24 h to be quantitative (Kowalewska K. *et al.*, 2010). A fraction of the phosphopeptide is modified at a lysine residue in the C-terminus, whereas it is not possible to distinguish between the two sites. Calculated from the signal intensities, less than 5% carry this modification. The contribution of the N-terminal lysine residue to the phosphopeptide signal is even more difficult to estimate, since only the c1 and the z12 ion are able to distinguish between the lysine and histidine modification. The c1 ion was not detected, but the signals of the z12 ion represented only 2% signal intensity of the unmodified form. These data show that side reactions have only limited contribution to the signal of the histidine phosphorylated peptide and further fractionation to remove the side products is not necessary. Shorter reaction times and lowering the pH to pH 7 might further suppress the formation of lysine phosphorylated peptides in phospho-His studies.

CID-fragmentation of N-bound phosphorylation on arginine

Previous reports demonstrated that next to Ser and Thr phosphorylation also phospho-His containing peptides undergo a nominal loss of H_3PO_4 upon thermal activation (Medzihradszky K.F. *et al.*, 1997). Interestingly, fragments ions exhibit an unmodified histidine and water loss from a different location in the peptide, which led to the assumption of translocation of the phospho-moiety during the activation process. Since the recently described protein arginine phosphorylation reveals similar neutral loss fragments, we tried to solve the mechanism of this phosphoric acid loss. Off-line fragmentation experiments with phospho-Arg model peptide ac-K**R**GGGGYIKIIKV demonstrate intense loss of H_3PO_4 (- 98

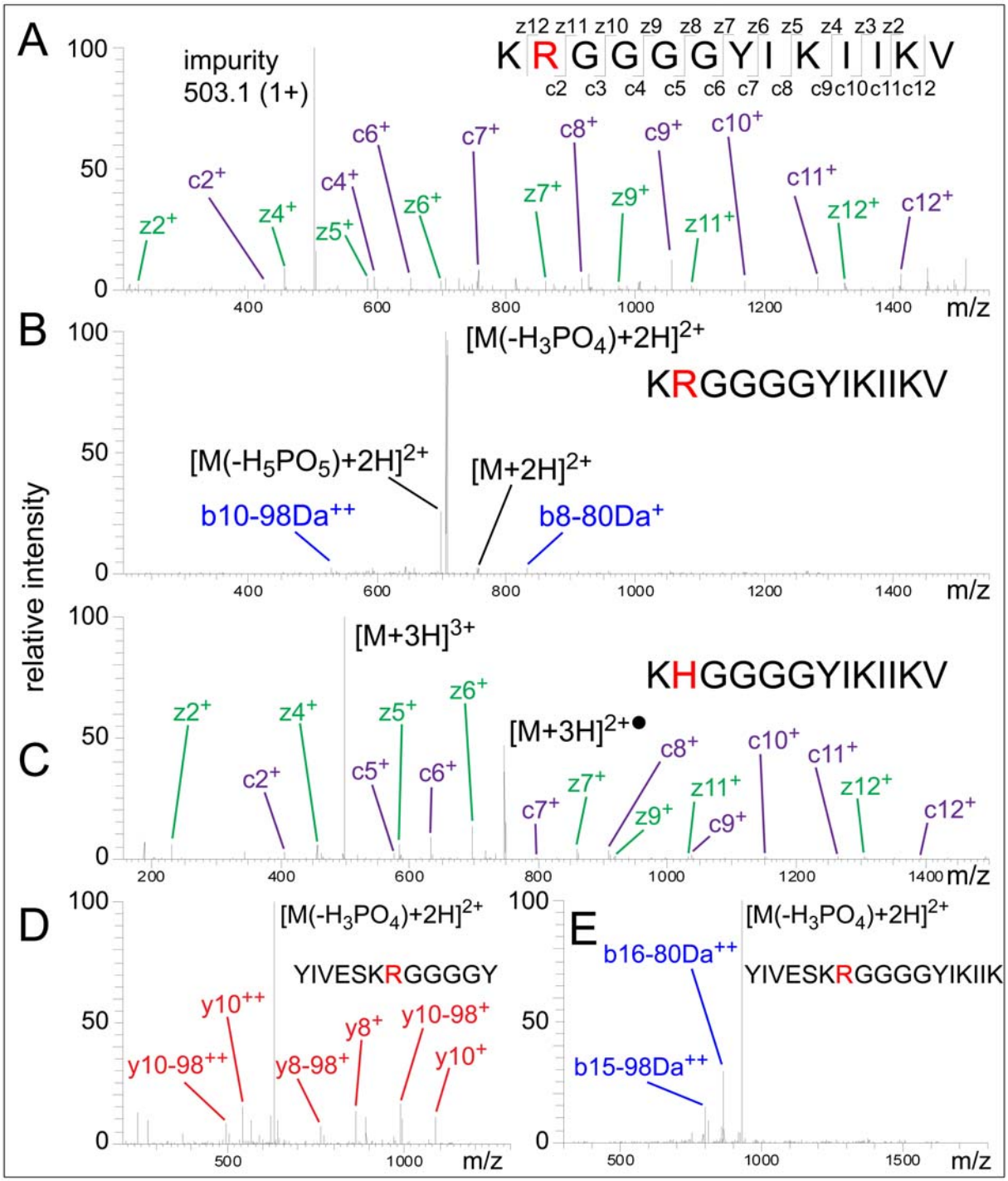


Figure B_16 **MS/MS Analysis of Arg- and His-Phophorylated Model Peptides**. A. The position of the phosphorylation in PepR_02 is only revealed by ETD MS/MS analysis with high sequence coverage. B. CID-MS/MS of the same peptides results in extensive elimination of phosphoric acid, whereas formation of sequence specific fragments is suppressed. Further, an additional loss of water next to the H_3PO_4 molecule can be observed ($[M(-H_5PO_5)+2H]^{2+}$). C. Similarly, ETD MS/MS is required to locate the phosphorylation site to His2 in the model peptide PepR_08, which was chemically phosphorylated with phosphoramidate. D+E. Also

PepR_14 and PepR_16 undergo extensive neutral loss of H_3PO_4 in MS/MS analysis upon thermal activation.

Da, $[M(-H_3PO_4)+2H]^{2+}$ at 706.9405 m/z) in ion trap CID MS/MS spectra (figure B_16B) and additional water loss (- 116 Da, $[M(-H_5PO_5)+2H]^{2+}$ at 697.9352 m/z) as most abundant signals. A similar fragmentation was also observed on other model peptides PepR_19 to PepR_23 (figure B_16D+E). Sequence specific fragment ions were only detected at low intensity. Since the phospho-moiety is attached to a nitrogen atom and only three oxygen atoms are covalently bound, two explanations for the loss of H_3PO_4 are possible: first, as recently described for pHis, the phospho-moiety is involved in a phospho-transfer onto an OH-function and subsequently eliminated as H_3PO_4. Second, the signal is due to extensive loss of two individual molecules (HPO_3 and H_2O) from the peptide that has the sum of 98 Da. The additional water loss M-H_5PO_5 is more likely explained by a concerted mechanism than by three independent neutral loss reactions.

ESI-MS^3 analysis of the neutral loss signal M-98Da at 706.9405 m/z demonstrates clearly that the loss of H_3PO_4 is split to two different locations in the peptide. From the dominating b-type ion series (b▲) only the previously attached phospho-moiety HPO_3 (79.966 Da) was eliminated, thereby causing a mass shift of this sequence by -79.966 Da compared to fragments with intact phosphorylation (figure B_17A+B). Another series of b-ions (b●), from which H_3PO_4 was eliminated, was present at very low intensity except for the b10 ion (figure B_17B); this fragment had a quite abundant loss of H_3PO_4. Curiously, the series of y-ions (y■) exhibited a water loss (-18.01 Da, figure B_17A+C) and intact y-ions were completely absent. Taken together, both losses, -79.966 Da from the b-ions and -18.0103 Da from the y-ions, exactly represent the elimination of H_3PO_4 from the whole peptide. Since water loss from the C-terminal amino acid is only a minor reaction of the unphosphorylated peptide, the observed fragmentation behaviour assumes a concerted neutral loss reaction that requires the C-terminal carboxylic acid group to eliminate H_3PO_4. Compared to other phosphorylations on S or H, this fragmentation pattern is more similar to that of phosphoHis, where a phospho-transfer mechanism is currently discussed (Zu X.L. *et al.*, 2007, figure B_17D). Both spectra exhibit dominant b▲-ion series, whereas pSer fragmentation resulted in the b● fragment ions that are associated with β-elimination of H_3PO_4 within the side chain (figure B_17E). Here, b-ions are highly abundant, from which the whole H_3PO_4 molecule was eliminated, and y-ions did not extensively lose water. Interestingly, the pArg-peptide shows only a single y-ion series with loss of H_2O, while intact

y-ions are also present in the pHis-peptide spectrum next to y-H_2O. The observation of a less intense b-ion series (b●), which had lost H_3PO_4, and the formation of a [M-(H_3PO_4+H_2O)+2H]$^{2+}$ ion points to the presence of a second neutral loss pathway. In this pathway different OH moieties are involved to eliminate the H_3PO_4 molecule.

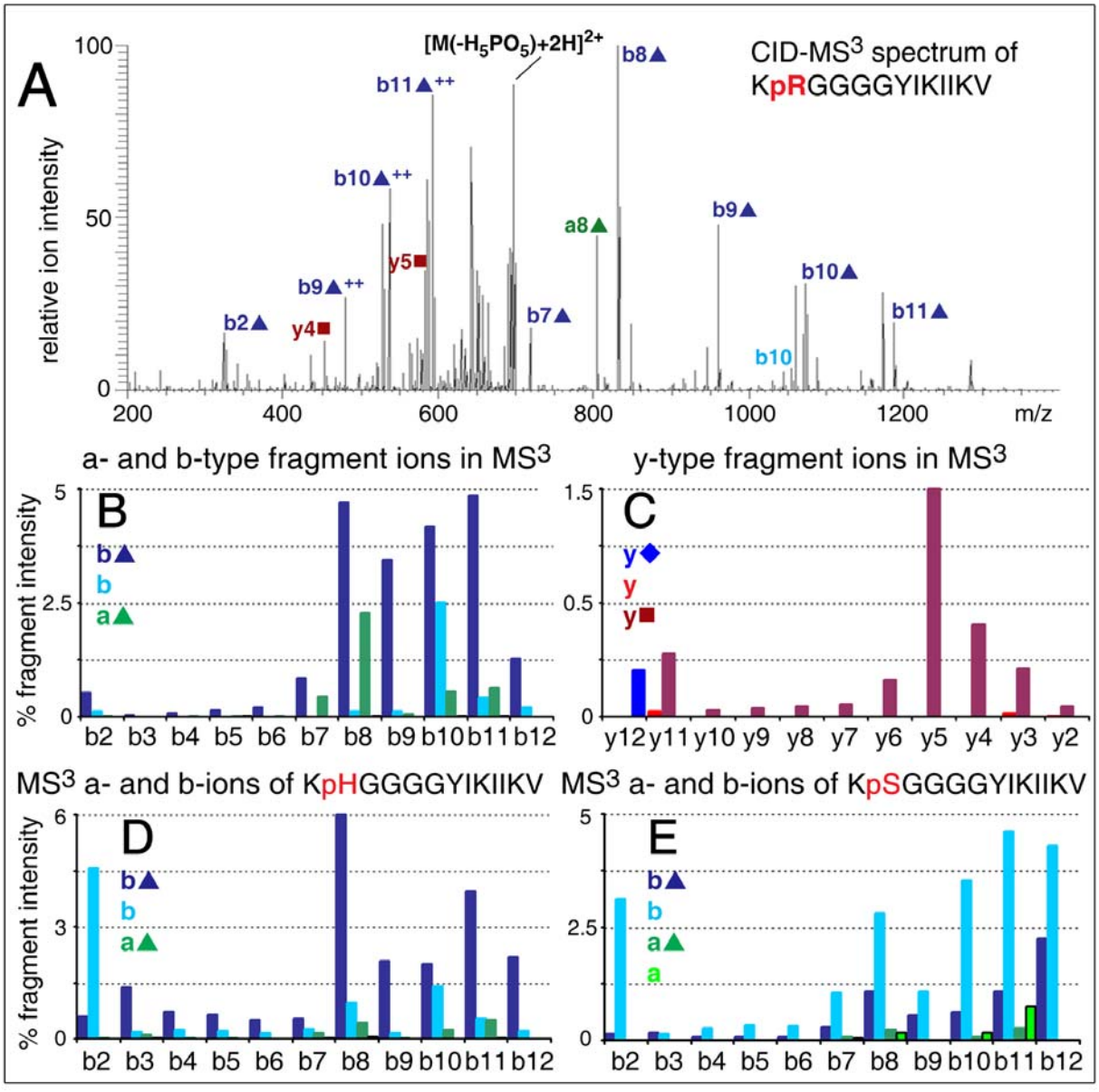


Figure B_17 **Fragmentation Analysis of the Neutral Loss Species M-98Da by MS^3**. A. MS^3 fragmentation of PepR_02 results predominantly in b-ions that had lost the phospho-moiety (-80 Da, b▲), which was not expected since the precursor eliminated H_3PO_4 (-98 Da). B. The same fragmentation behaviour was observed on all N-terminal fragment ions (a▲ and b▲). The intensity for b-H_3PO_4 ions (b) was significantly lower. C. The water molecule stems from the C-terminal COOH-moiety, as becomes obvious from the y-ion series that was only identified with a loss of a water molecule. D. Similar fragmentation behaviour is observed

upon fragmentation of phospho-histidine containing peptides, where a phospho-transfer reaction is under discussion. E. The β-elimination pathway in the Ser-phosphorylated peptide PepR_07 leads to b-98Da fragments, therefore resulting in a completely different fragment ion pattern. Intact y-ions are identified in this spectrum (data not shown.)

Formation of neutral loss species depends on activation parameters

The applied activation parameters (CE 35 and activation time 30 ms) were optimized for the fragmentation of unphosphorylated peptides and therefore delivered more energy than required for the elimination of H_3PO_4 alone. To exclude the possibility of a loss of two individual neutral molecules due to the high energy level, we screened different collision energies to see, if signals from single molecule loss species, either M-H_2O or M-HPO_3, have higher abundance at low collision energies. Interestingly, high resolution data clearly demonstrate the formation of the M-98 Da signal (red) even at low CID energies as demonstrated in figure B_18A, whereas formation of M-H_2O or M-HPO_3 (green) was not detected. Moreover, parallel to the M-H_3PO_4 signal, also the species with combined water/H_3PO_4 loss (blue) appeared already at low collision energies, while the precursor intensity (black) dropped immediately at CE 17. Since not only the applied energy but also the activation time might be of importance for this reaction, we conducted experiments at lower activation times with fixed CID energies in which we observed the same fragmentation behaviour (figure B_18B).

Data acquisition was also performed in the ion trap analyzer, since the ion trap requires significantly less time but has lower resolution. Acquisition of MS/MS spectra in the ion trap is performed in app.100 ms, which would therefore allow us to detect meta-stable ions. The orbitrap provides the exact mass and high resolution to determine the elimination products, but requires 1-2 seconds acquisition time, so that activated ions can undergo further fragmentation. Ion trap fragmentation of the model peptide with 5 ms activation time revealed identical results for the formation of neutral loss species (figure B_18C+D). These data demonstrate that the observed neutral loss of H_3PO_4 is not a product of the long detection process in the orbitrap analyzer. The M-H_3PO_4 neutral loss ion reaches a plateau at approx. 82%, and the M(-H_5PO_5) signal gains a maximal intensity of 16%. To study the formation of other neutral loss species, we applied different CE/activation time combinations. At higher collision energy and lower activation times (2 ms), both neutral loss pathways for M-HPO_3 and M-(H_3PO_4+H_2O) were significantly stronger represented (figure B_18D) and the intensity of the M-H_3PO_4 ion reduced. From this data I concluded that different fragmentation

pathways exist, which differ by slightly higher collision energy. Although ion trap MS provides shorter scan times, fragment ions according to a loss of H_2O from the intact phosphopeptide were never detected. The M-HPO_3 signal at 715.9457 m/z was observed at low intensities and required higher CE than elimination of phosphoric acid.

PSD fragmentation in MALDI-TOF/TOF-MS/MS is measured within very short times of usually less than 1 ms but at rather high collision energies. Even under these conditions we observed the loss of H_3PO_4 (M-98Da) as base peak. Loss of HPO_3 (M-80Da) had only 4% intensity. Interestingly, the combined loss of H_3PO_4 and H_2O (M-116Da), which was quite intense in ESI-MS/MS experiments, had less intensity (~2%) than the M-80Da signal in MALDI-PSD spectra. This observation assumes that the additional water loss might require mobile protons, to form intermediate OH moieties at the amide bonds in the peptide backbone.

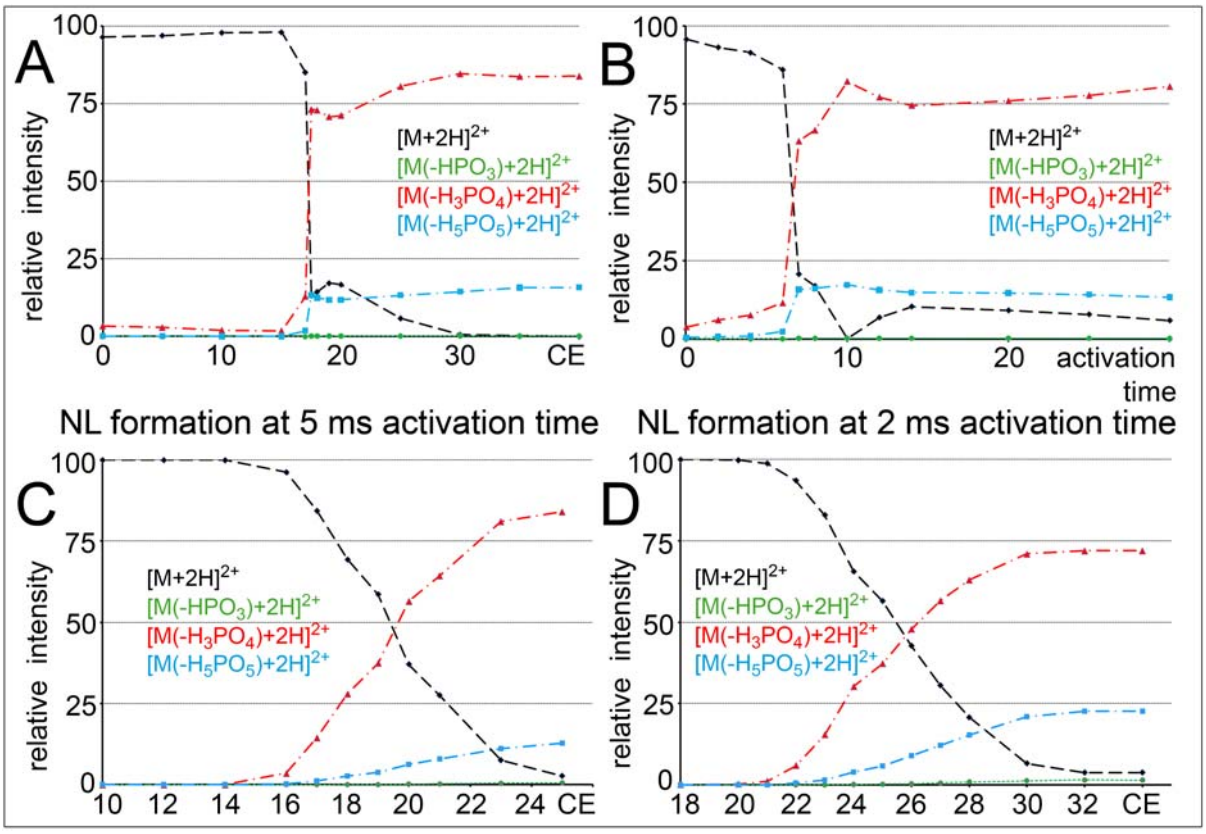


Figure B_18 **Formation of Neutral Loss Species Depends on Activation Parameters**. A. Screening the collision energy at 35 ms activation time and acquisition of fragment ions in the orbitrap analyzer leads to elimination of H_3PO_4 (red) at a minimum CE of 17 for the doubly charged precursor. The precursor ion (black) is completely fragmented at CE of 30. B. Activation times exceeding 6 ms are required for extensive neutral loss of H_3PO_4 and H_5PO_5 (blue) at collision energy of 35. The data were again acquired in the orbitrap analyzer. C+D.

Acquisition of MS/MS spectra in the ion trap provides better investigation of the neutral loss process. C. At 5 ms activation time, the loss of H_3PO_4 is first initiated at CE 16 and gains up to 82% relative intensity. Elimination of H_5PO_5 requires slightly more energy and reaches maximal ~16% relative intensity. D. Lower activation times (shown for 2 ms) require higher CE to initiate the neutral loss. At higher CE elimination of H_5PO_5 is significantly increased as well as the loss of the phospho-moiety (HPO_3, green) alone, which gains little intensity.

MS/MS studies of peptides with amidate- or methylester-blocked OH-moieties

To substantiate our first idea of the neutral loss mechanism which involves the C-terminal COOH-moiety of PepR_02, we modified the OH-groups by methyl esters or amidation. Thereby, we generated peptides without stable OH-moieties which could undergo neutral loss. The remaining OH moiety in the side chain of tyrosine is not able to eliminate H_3PO_4, therefore it is not considered as an acceptor in this neutral loss study.

ESI-CID-MS/MS of the esterified and arginine-phosphorylated peptide PepR_17 ac-KpRGGGGYIKIIKV-OCH_3 displays again a strong neutral loss of H_3PO_4 which represents the base peak of the spectrum, similar to PepR_02. This fragmentation behaviour was quite surprising, since the C-terminal COOH moiety was modified and could not serve as phospho-acceptor. Upon esterification of the C-terminus, the neutral loss species M-HPO_3 (722.9512 m/z) gained signal intensity by a factor of 27 relatively to the M-H_3PO_4 (713.9467 m/z) neutral loss ion in comparison to PepR_02 (figure B_19A). A similar, but lower increase upon esterification of the C-terminus was also observed for peptides with serine or threonine phosphorylation (PepR_06 and 07). Further, the combined neutral loss of H_3PO_4 and H_2O (M-116Da) was significantly reduced in the CID_MS/MS spectrum of PepR_17.

Fragment ions of the peptide backbone showed around 2-3 times higher signal intensities in CID MS/MS compared to the spectrum of PepR_02. To a high percentage the phospho-moiety was preserved on identified b-ions during the fragmentation, while the b▲ series was dramatically reduced (figure B_19B) in comparison to b and b● ions. Similar to PepR_02, strong b-H_3PO_4 ions were observed starting at b10-H_3PO_4. CID-MS^3 analysis of M-H_3PO_4 resulted in a strong loss of methanol from the methylated C-terminus (697.9349 m/z) in addition to the H_3PO_4 loss in the first tandem MS step. Despite loss of H_3PO_4 in the first fragmentation step, MS^3 analysis of this species revealed intact y-ions and a less intense y■ series, suggesting that a yet unknown functional group acts as phospho-acceptor in the neutral loss reaction (figure B_19C). A hint towards the involvement of amide bond carbonyl groups is provided by the formation of b●-ions. In MS/MS- and MS^3-spectra this b●-ion series that

results from elimination of H_3PO_4 from the N-terminal part, has significantly increased intensity in comparison to the peptide with the free COOH-group (figure B_19B). Amidation of the C-terminal COOH moiety leads to similar fragmentation behavior of the peptide, which further strengthens the hypothesis of different fragmentation pathways for the neutral loss of H_3PO_4 in CID-MS/MS experiments (figure B_19D).

Interestingly, also peptides with phosphoesters on Ser and Thr and methylesters on the C-terminus show increased intensity of sequence specific b- and y-ions but the overall intensity gain was lower than for pArg. From MALDI-TOF/TOF MS/MS fragmentation I obtained similar fragmentation patterns, but interestingly here the neutral loss species M-HPO_3 gained relative intensity to approximately 80% of the M-H_3PO_4 peptide ion.

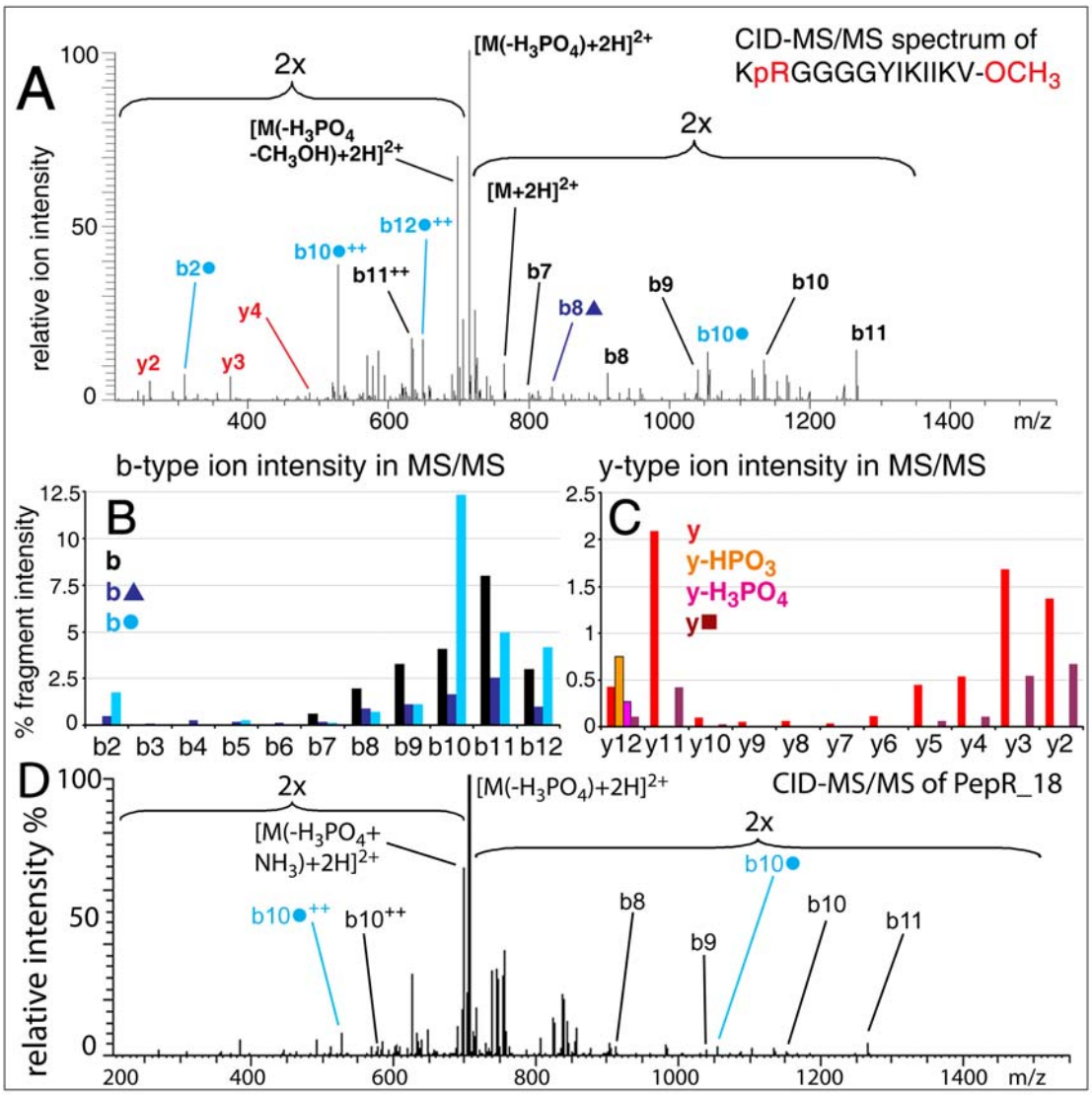


Figure B_19 **Neutral Loss of Peptides Without Acidic OH-Moieties**. A. CID-activation of PepR_17 results in extensive elimination of H_3PO_4 from the precursor, but the intensity of other fragments is significantly increased. B. N-terminal fragment ions are present with intact

phospho-moieties or exhibit an elimination of H_3PO_4, which is significantly different from the peptide with free C-terminus. Fragments with a loss of the phospho-moiety (-80 Da) have the lowest intensity. C. Intact y-ions can be observed with higher intensity than y-ions with water loss. D. The amidated peptide PepR_18 leads to a similar fragment ion spectrum like PepR_17.

Suitability of different OH-moieties for phospho-transfer

A neutral loss fragmentation pathway, combining the elimination of two molecules from different locations in the peptide, raised the question, if other OH-moieties than the C-terminal COOH can also take part in this mechanism. Therefore we synthesized a new set of peptides (table TB_03, PepR_19 to PepR_23), to provide OH-groups at different locations within the peptide and of various reactivity: the C-terminus and the side chains of glutamate or serine. First, we tested the phosphorylated peptides for their ability to eliminate H_3PO_4 upon low energy CID activation and if other neutral loss species can be detected in the resulting spectra. A good indicator for a OH moiety to be involved in the neutral loss process is the intensity of M-HPO_3 ion, which is increased, if the OH group is less suitable to eliminate H_3PO_4 via this pathway, as demonstrated by the esterified and amidated peptides above. Furthermore, under such conditions backbone breaks and side reactions are more frequently observed.

Peptides PepR_19 and 20 exhibited lower and upper limits of the neutral loss reaction, while the HPO_3-loss was almost absent in the peptide with free C-terminus (Peptides PepR_19), the C-terminally amidated peptide had the highest percentage of this fragmentation type (up to 16%) under the applied conditions (figure B_20A). Replacement of glutamine residues by glutamic acid as phospho-acceptor either in position 3 or 12 significantly reduced the peptides ability to undergo the H_3PO_4 loss in comparison to PepR_19, which was demonstrated by the higher abundance of $[M(-HPO_3)+2H]^{2+}$. Insertion of Ser at position 12 resulted in the second strongest loss of HPO_3 from the peptide that was almost as intense as that of the fully amidated peptide. Therefore, serine had the lowest ability to undergo elimination of H_3PO_4 under the applied conditions. The contribution of Ser- or Thr-directed H_3PO_4-loss from N-phosphorylated peptides to the intensity of the M-98Da signal should be very small, if the peptide has a free C-terminal COOH group.

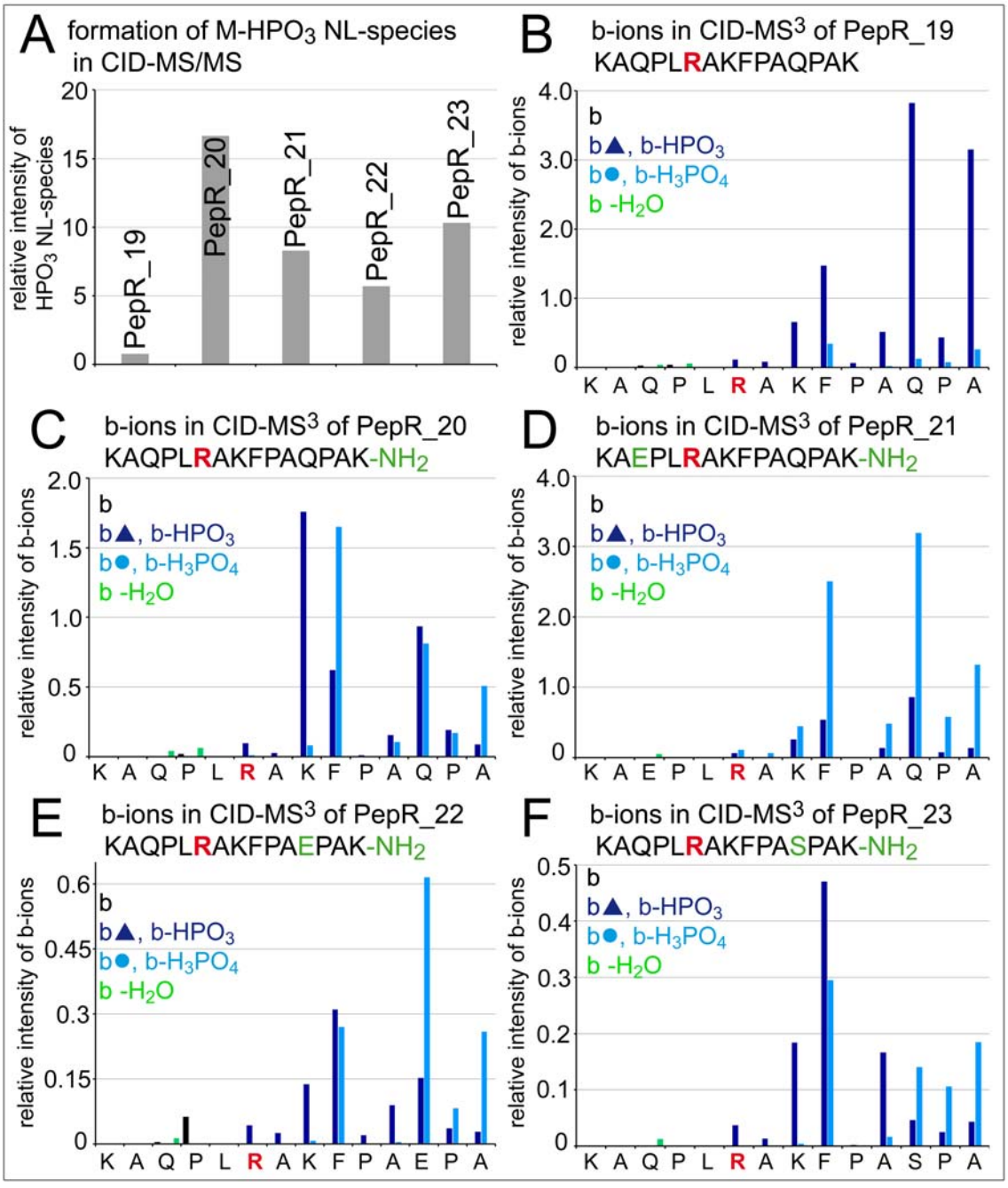
Fragment analysis of the M-H_3PO_4 neutral loss species in MS^3 spectra demonstrated clearly that the neutral loss required an OH-moiety (figure B_20B-K). While the peptide with

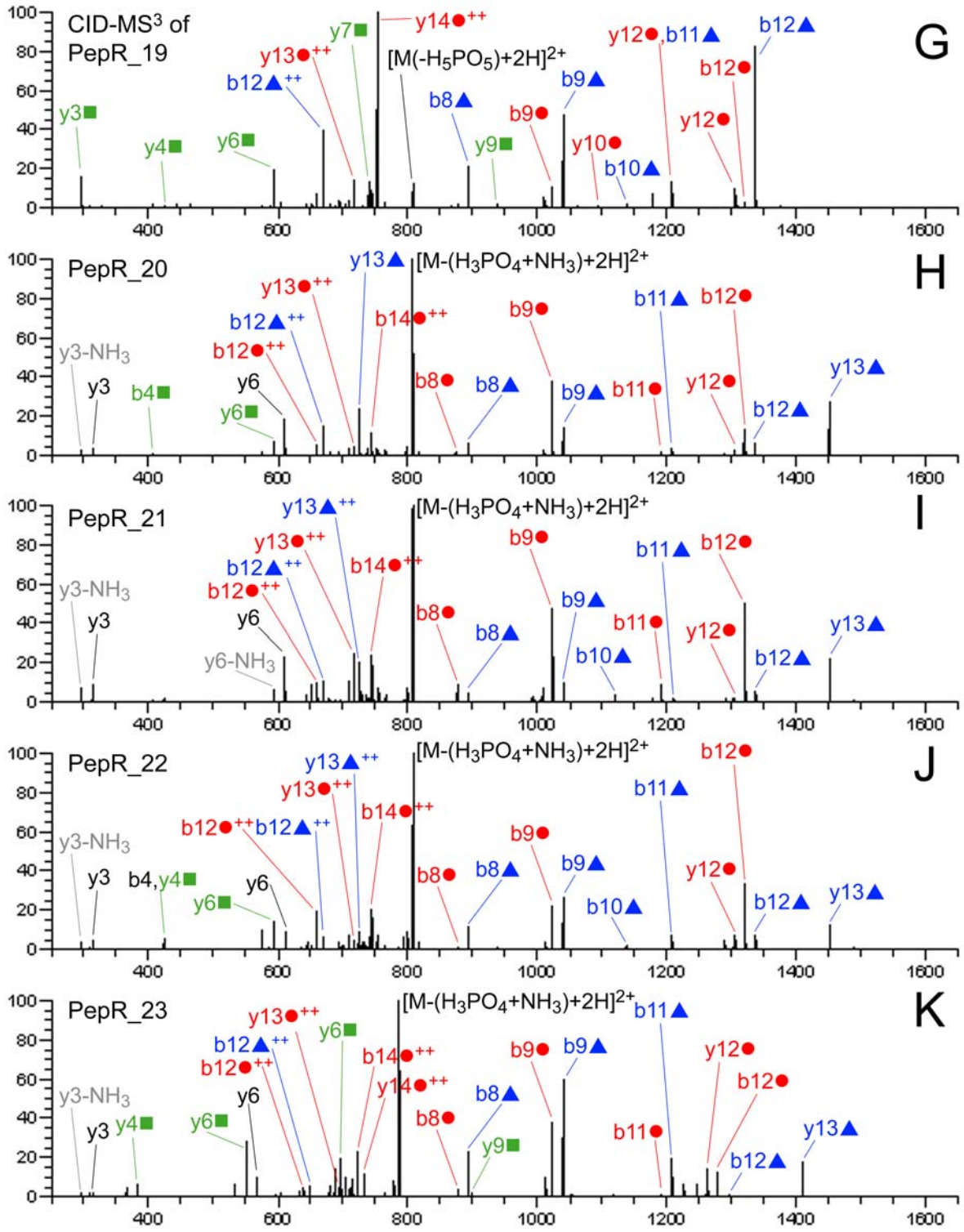
the free C-terminal COOH (PepR_19) produced exclusively y-H_2O ions (y■) and b-ions, which had lost HPO_3 (b▲), the amidated peptide PepR_20 exhibited also abundant series of intact y-ions and b-H_3PO_4 type ions b● (figure B_20B,C,G,H). This fragmentation pattern was similar to PepR_17 and 18 and might be explained by neutral loss from intermediate OH-groups that were formed by the mobile proton. If the glutamic acid residue was located N-terminal to the phosphorylation site (PepR_21), then the dominant b-ion series showed intensive H_3PO_4-loss as well as water loss for b-ions before the phosphorylated arginine residue, but included the glutamic acid (figure B_20D,I). For this peptide, predominantly intact y-ions were observed. PepR_22 had the glutamic acid residue located between the phosphosite and the amidated C-terminus at position 12. MS^3 analysis of this peptide after H_3PO_4 elimination clearly allowed localization of the reactive OH moiety to this residue. The y-ion series exhibited intense water loss starting from the y4 ionm, the first ion with the glutamic acid. Further, y-ions which comprise the pArg predominantely lost H_3PO_4 except for y13. The series of b-ions with a loss of HPO_3 was dominant from b6 to b11 (figure B_20E,J). Starting from b12, at the glutamate, the H_3PO_4 loss became more intense, which clearly revealed that this amino acid is mainly involved in the elimination of H_3PO_4. Similar observations were made for peptide PepR_23 in which the glutamic acid is replaced by serine (figure B_20F+H).

All MS^3 spectra of peptides PepR_20 to PepR23 contain NL species that can be explained by different NL mechanisms than the phospho-transfer onto the free OH-moiety alone. For instance, fragment ions with water losses of b3 or b4 without the N-terminal Glu or elimination of H_3PO_4 from all fragments comprising Arg6. Such fragments can be explained by the significantly lower ability of theses peptides, to undergo the phospho-transfer / H_3PO_4 elimination with COOH moieties on Glu or hydroxyl groups of Ser instead of the free C-terminus. Under these circumstances, the NL pathways that led to the elimination of H_3PO_4 from peptides without free OH groups, gain higher importance in comparison to PepR_19, which predominantly eliminated the water from the C-terminus.

Figure B_20 **Neutral Loss Fragmentation of Peptides With Different Cleavable OH-Groups**. Peptides PepR_19 to PepR_23 were subjected to different fragmentation methods to study the neutral loss products. A. The tendency to eliminate HPO_3 instead of H_3PO_4 is higher, if no or other OH groups than the C-terminus are involved. The peptide without OH-moieties exhibits the highest elimination of HPO_3, followed by the Ser-containing peptide. B+G. The MS^3 spectrum of PepR_19 shows b-HPO_3 (b▲) and y-H_2O (y■) signals, which

were expected after elimination of H_3PO_4. Y-ions, having the phosphorylated arginine in their sequence, exhibit H_3PO_4 loss. C+H. Fragment analysis of the H_3PO_4 neutral loss ion of PepR_20 exhibits diverse elimination of HPO_3 or H_3PO_4 on the fragment ion level. D+I. In PepR_21 the N-terminal Glu residue is involved in the elimination of H_3PO_4, as demonstrated by intense b-H_2O and b-H_3PO_4 signals. E+J. PepR_22 forms b-HPO_3 fragments up to A11; from E12 H_3PO_4 is eliminated, which clearly locates the second neutral loss position to this amino acid. F+K. Insertion of Ser (PepR_23) at Pos12 results in a similar fragment pattern in MS^3 spectra as observed for PepR_22 (E+J), with H_3PO_4 elimination from R6 and S12. Interestingly, the b● ions have lower intensity than similar fragments of PepP_22.





Revealing the Phospho-Transfer Process upon Thermal Activation of Phospho-Arg-Containing Peptides

The great differences for H_3PO_4-elimination from Arg-phosphorylated peptides PepR_19 to 23 with different OH-moieties gives rise to two possible questions: are the different OH-moieties less accessible for the phospho-transfer or is the up-following

elimination of H_3PO_4 inhibited. To test, if the second reason explains the fragmentation behaviour, we wanted to know, whether we can identify phosphorylated y-type fragment ions which do not carry the originally phosphorylated arginine, but an amino acid with free OH-moiety. Unfortunately, there is only a little chance to detect such fragment ions, which is due to the fact that the phospho-transfer does not alter mass or charge state. Keeping these two parameters during the transfer process, leads to additional thermal activation of the product species by the mass spectrometer, allowing elimination of H_3PO_4 since activation parameters are relatively independent from the type of modification.

First, exclusive phosphorylation of Arg6 of peptides PepR_19 to 23 was determined by intensive investigation of ETD spectra of the 3+ and 4+ charged precursor ions that did not exhibit signals for phosphorylation of a different amino acid. Therefore, phosphorylated fragment ions in CID-MS/MS spectra, pointing to other residues than Arg5 have to be formed upon thermal activation. Low-energy and high-energy CID-MS/MS spectra of peptides PepR_22 and PepR_23 revealed signals for putative phosphorylation on y4- and y6-ions, see figure B_21ABI+II. In all cases, the mass increase, caused by addition of the phosphorylation differed only in the fourth digit from the theoretical value. Y6-ions were quite abundant, since they result from an amide bond break N-terminal to praline, which occurs more frequently than on other amino acids. The high abundance is due to the rigid ring structure of this amino acid, which is not able to release thermal energy by motion or rotation. In case of the PepR_22, the signal of the phosphorylated fragment (691.3180 m/z) was only approx. 10% of that of the unphosphorylated y6 at 611.3517 m/z, which could be explained by the low stability of the formed phospho-anhydride structure in the glutamate side chain. For PepR_23, signals of phosphorylated y4 (481.2176 m/z) and y6 (649.3075 m/z) were even more intense than that of the unmodified fragment ions (y4 401.2512 m/z, y6 569.3411 m/z) (figure B_21ABI+II). All putatively phosphorylated y-fragment ions were absent in CID-MS^3 spectra of the phosphoric acid neutral loss ion, further pointing to involvement of amino acids in these fragments in the phospho-transfer step.

All respective fragment ion species could be isolated and fragmented in a further MS/MS step. CID-MS^3 spectra of unphosphorylated y6-fragments confirmed the amino acid sequence of the fragment ions. Fragmentation of the phosphorylated y6-fragment of peptide PepR_22 with the phospho-anhydride function resulted in excessive loss of phosphoric acid, however sequence related fragment ions could not be detected figure B_21CI+II. The y6-fragment of PepR_23, which contained Ser instead of glutamate allowed identification of the amino acids in both unmodified and phosphorylated form. Moreover, in these spectra we

found also fragments of y6 which carried the phospho-moiety. Furthermore, MS^4 analysis after H_3PO_4 elimination could further be used for sequence validation.

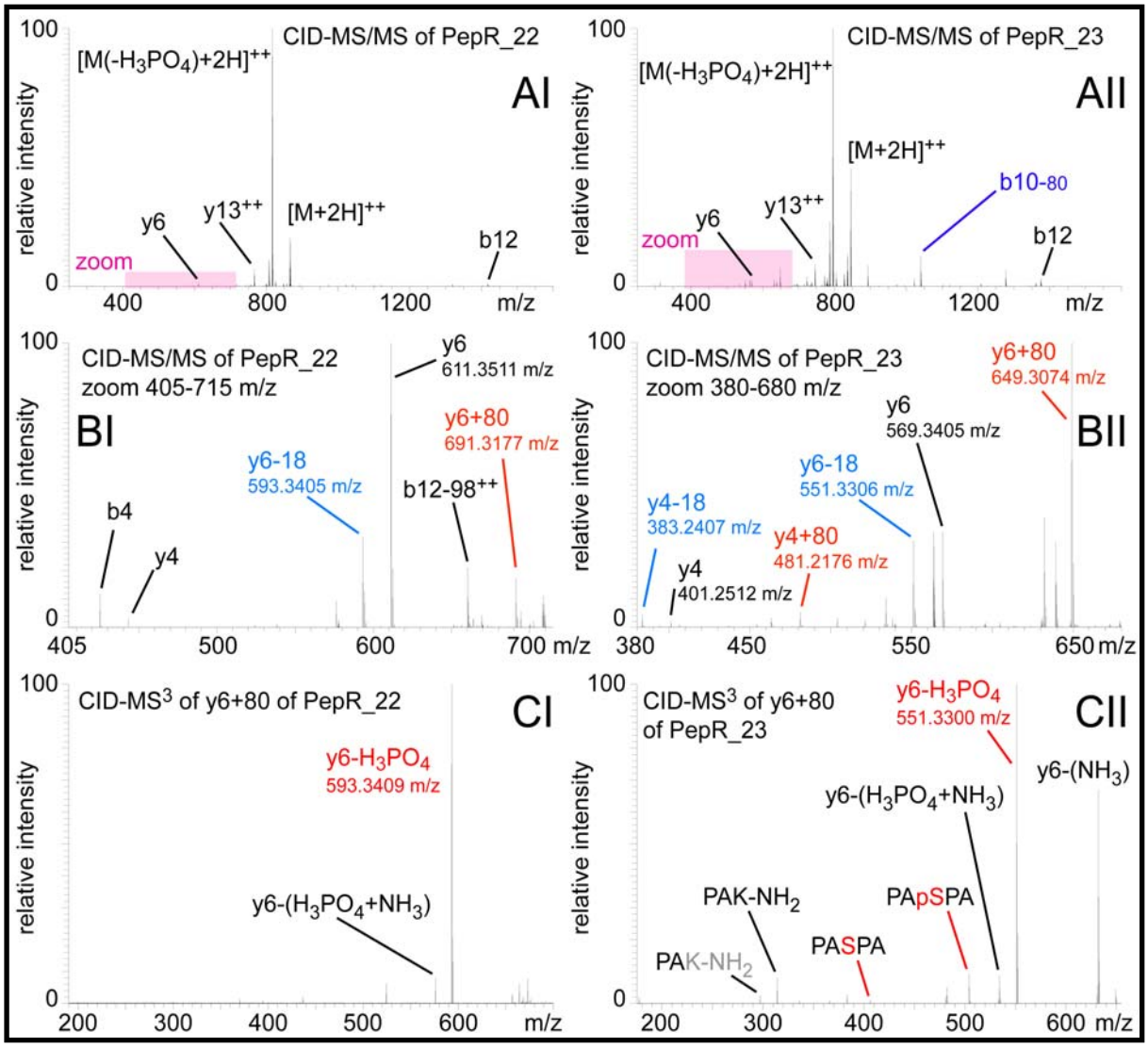


Figure B_21 **CID-Induced Translocation of the Phospho-Moiety Onto Glutamate and Serine** A. CID-MS/MS spectra of PepR_22 (I) and PepR_23 (II) showed mainly neutral loss of H_3PO_4. B Close inspection of the low mass area revealed presence of phosphorylated y-ions, which did not comprise the originally phosphorylated arginine. BI PepR_22 exhibited a phosphorylated y6 ion (y6+80) next to the unmodified form y6 and that with water loss (y6-18). BII Two phosphorylated y-ions were identified in the spectrum of PepR_23, y4+80 and y6+80. C Fragmentation of the phosphorylated y6-ions of both peptides exhibits strong neutral loss of phosphoric acid in both cases, further confirming phosphorylation of the respective OH-donors under CID conditions.

High energy CID fragmentation in ESI-HCD-MS/MS and MALDI-TOF/TOF-MS/MS

At elevated collision energy, bond dissociations in the peptide backbone of phosphorylated peptides occur more frequently. Although elimination of phosphoric acid is the energetically favoured fragmentation pathway, further bond breaks are induced by the high energy level. On average, only a single bond should be dissociated in the fragmentation reaction, but this also involves a high percentage of fragments with dissociation of two bonds. On the other side, a considerable amount of unfragmented precursor is also left in the HCD spectrum. In comparison, under low energy CID conditions, the single bond break is achieved by longer activation times at lower collision energy.

Application of higher collision energies to fragment Arg-phosphorylated peptides induces the formation of peptide backbone fragment ions with intact modification as well as fragments with neutral loss. Higher fragmentation energy and the relatively long detection time in the orbitrap analyzer induce subsequent fragmentation reactions and phospho-loss from the fragments, which leads to a plethora of fragment ions. Most importantly, fragments with intact N-phosphorylation are formed and can be used for identification of the original phosphorylation site. Neutral loss of water or phosphoric acid in MS^n spectra are often ambiguously interpreted for a selected fragment ion, especially if other putative phospho-sites such as Ser or Thr are present in the sequence. Due to the high number of observed fragments, some fragment ions overlap with their isotopic patterns, often reducing the chance to identify the original modification.

Similar to the low energy fragmentation, under these fragmentation conditions predominantly a-HPO_3 and b-HPO_3 fragments are observed next to that with intact modification. In comparison, the phospho-Ser peptide results predominantly in b-type fragment ions with a neutral loss of H_3PO_4. Interestingly, y-ions are observed with highest intensity at their expected mass, and the water loss is significantly reduced.

Discussion – Part II

Optimization of the protein phosphorylation analysis protocol led to the identification of arginine-phosphorylation in the bacterial transcription factor CtsR and first description of a protein arginine kinase: McsB (Fuhrmann J. *et al.*, 2009). Both proteins have been targets of recent studies, which revealed incorporation of phosphate, but failed at localization of the modified amino acids. So far, arginine phosphorylation was only indirectly detected by autoradiography of proteins, followed by acidic release of the phosphate, or loss of the signal in EDMAN-degradation at the position of putatively modified arginine. By applying electron capture dissociation, a special fragmentation method in mass spectrometry, which allows detection of labile modifications, we identified several phosphorylated arginines in the sequence of CtsR. This dimeric protein is in tight complex with the DNA and thereby inhibits transcription of stress-related genes. Upon exposure to stress, the protein is released from the DNA and arginines in the DNA-binding motif are phosphorylated by McsB to inhibit re-association of the $CtsR_2$/DNA-complex. The inhibition is predominantly based on the inversion of the charge of arginine side chains upon phosphorylation, leading to electrostatic repulsion of the CtsR from the DNA. This model was further strengthened by biochemical data on the affinity of several CtsR mutants towards DNA. Interestingly, only arginines in the DNA-binding motif were phosphorylated, while the arginine residues in the C-terminal dimerization domain were not targeted by McsB.

McsB was previously described as protein tyrosine kinase, based on phospho-amino acid analysis after acidic hydrolysis and tyrosine mutation. Our studies clearly revealed specificity towards arginine. To confirm the McsB kinase activity we conducted phosphorylation assays on small peptides and measured the incorporation of phosphate by MALDI mass spectrometry and ^{31}P-NMR spectroscopy. Peptide phosphorylation assays exhibited an additional peptide signal in MALDI mass spectra only when the peptide had an unmodified arginine. Other amino acids, which can be phosphorylated, were not modified by McsB. The NMR spectrum shows only a single signal, which has a chemical shift close to that of free phospho-arginine. Both analyses delivered clear evidence for the arginine phosphorylation and excluded modification of tyrosine by McsB. Furthermore, we demonstrated that McsB prefers arginine residues in positively charged domains, being often associated with DNA-binding proteins.

Detection of arginine phosphorylation in biological samples is still a challenging task, since acidic conditions should be avoided and the knowledge of phospho-arginine

phosphatases is limited. To establish an enrichment procedure that allows purification of arginine-phosphorylated peptides and subsequent detection by mass spectrometry, the hydrolysis rate was determined for several pH/temperature combinations. As previously described, the hydrolysis is significantly enhanced under acidic conditions but already at pH 3 the hydrolysis rate is reduced. The modification was completely stable under neutral and alkaline conditions. Based on these data, a purification protocol after enzymatic phosphorylation of model peptides was established, to prepare pure phosphopeptides for fragmentation experiments in the mass spectrometer and optimization of LC-MS/MS based identification.

Using the described protocol we detected 15 different autophosphorylation sites in the McsB sequence, which are applied upon incubation with ATP. Some of the identified phosphoarginines are required for enzymatic activity. This might explain reduced phosphorylation of small model peptides at long reaction times.

Mechanism of Neutral Loss of Phosphoric Acid Under CID Conditions

Screening of MS activation parameters demonstrated that loss of H_3PO_4 is the preferred fragmentation pathway, which already takes place at lowest fragmentation energies. This reaction involves bond breaks at two different positions in the peptide: at an acidic OH moiety and the phosphorylated guanidyl side chain. Individual fragmentations from both positions are excluded. Since each single molecule loss would require less energy than their combination, higher signal intensities of these single loss species should be observed at low CE. Nevertheless, elimination of H_3PO_4 is still preferred under these fragmentation conditions. The observed fragmentation mechanism is independent of mobile protons, since it also occurs in MALDI MS/MS analysis with high frequency. Elimination of a single water molecule from the phosphorylated peptide was never observed and the peptide signal for an N-P bond break (M-80Da) alone is not present at low CEs, but gains intensity at higher energy levels. This behaviour would account for a high energy fragmentation pathway, which is also observed for phospho-esters of Ser and Thr. Furthermore, water loss from the involved OH-moiety is not extensively present upon CID MS/MS analysis of the unphosphorylated peptide. These observations strengthen the theory of a concerted fragmentation pathway, which connects the elimination of HPO_3 from the phosphorylated side chain and the water loss. This mechanism could be similar to the phospho-transfer described for peptides with His-phosphorylation.

Interestingly, H_3PO_4 is also eliminated from peptides that do not have acidic OH moieties. This observation is in agreement with a second less intense neutral loss species of peptides with acidic OH-moieties. These are able to eliminate an additional water molecule next to the H_3PO_4 loss to form [M(-H_5PO_5)] ions. At elevated collision energies this type of fragmentation is significantly increased upon low and higher energy CID activation. Since this fragmentation is significantly suppressed upon fragmentation of singly charged peptide ions by MALDI TOF/TOF analysis, maybe the process requires mobile protons. These protons are located at the amide bonds and are able to intermediately form OH groups with the carbonyl O-atoms. Although the pathway, including the peptides C-terminal COOH moiety, seems to require the lowest activation energy under standard fragmentation conditions, the resulting tandem MS spectrum is always a mixture of different fragmentation pathways. Abundance of mixed fragmentation products increases the complexity to elucidate the fragmentation mechanism into detail. A problem of the phospho-transfer mechanism is caused by the equal masses of the precursor and intermediate product, a mixed anhydride of phosphoric acid and a carbonic acid. If the applied CE initiates the translocation of the phospho-moiety from the arginine to the OH group, then the newly formed species is further activated and almost immediately fragmented.

MS/MS spectra of peptides with Ser or Glu as OH donors instead of the C-terminal COOH moiety exhibit phosphorylated fragment ion signals, which do not comprise the originally phospho-Arg, but the OH-donor. Since phosphorylation and sequence of these fragments were confirmed by MS^n studies, these signal are the first evidence for the translocation of phospho-moieties of N-phosphorylated peptides under thermal activation conditions.

Outlook – Identification of N-bound Phosphorylations in Cells

The limited knowledge about abundance and biological function of nitrogen-bound protein phosphorylations in cells as well as enzymes that apply or remove this modification makes this research field highly important and interesting. Especially, since bacteria seem to use such modifications more often than eukaryotes, identification of the respective proteins might be the basis for the development of novel antibiotics, which will be directed against specific bacteria and cause less side effects to eukaryotic cells. The presence of these modifications in eukaryotic cells has been demonstrated for phospho-His in yeast and phospho-Arg in liver and heart, whereas the latter did not directly identify the modification (Wakim B.T. and Aswad G.W., 1994; Levy-Favatier F. *et al.*, 1987).

Phosphorylation of His, Arg, and Lys leads to charge inversion in the side chains of these amino acids which alters the interaction with other molecules. Due to the positive charge of the unmodified side chain, these amino acids have often crucial roles in DNA or RNA recognition. As demonstrated by the CtsR/McsB regulatory system, especially DNA-bound transcription factors are effectively removed from the DNA and binding is inhibited. Therefore, N-phosphorylations would be highly valuable to adjust the activity of transcription factors and could also be present in eukaryotes.

In bacteria and plants, sensing of extra-cellular conditions is often associated with receptor-His-kinases, which demonstrates the importance of this post-translational modification. Starting characterization of these modifications in bacteria is also supported by the low amount of phospho-esters, being present in these cells (Miller M.L. *et al.*, 2009). Since the presented enrichment strategy allows purification of all peptide phosphorylations, a large fraction of Ser, Thr, or Tyr phosphorylation would definitely impair identification of N-bound phosphorylations. So far, approximately 100 phospho-esters have been identified in bacteria from rather large sample amounts, demonstrating that this modification is underrepresented. Since also other signalling processes might use phosphorylation of arginine or histidine, phosphoproteomic studies in bacteria could lead to the identification of novel therapeutic targets. Furthermore, this could be the basis to develop strain-specific antibiotics, to deal with the rising problem of multi-drug resistant bacteria in hospitals or huge farms.

Many histidine kinases were also identified in plants that also utilize the two-component systems to screen environmental conditions. Since plants have developed both phospho-signalling forms, with stable phospho-esters and labile phospho-amidates, this

system is not yet optima for proteomic studies. Furthermore, many separation methods and sample preparation steps have not yet been adopted to plant samples.

Animals, in comparison, developed multiple cellular pathways which rely on Ser, Thr, and Tyr phosphorylation. O-phosphorylation is found on numerous proteins and should be removed before N-phosphorylations are analyzed. Nevertheless, recent publications about protein arginine phosphorylations in rat and bovine cells indicate the presence of this modification also in higher eukaryotes. Furthermore, it is believed that His-phosphorylation occurs approximately 100 times more often in eukaryotic cells than tyrosine phosphorylation (Steeg P.S. *et al.*, 2003). A model system for histidine phosphorylation in eukaryotic cells is the membrane-bound SLN1 receptor kinase of yeast cells, which is a bacterial-derived histidine kinase (Thomason P. and Kay R., 2000). This receptor is the initiating protein of a phospho-relay system and involved in osmolarity sensing and represents a second signalling mechanism to activate salt stress response.

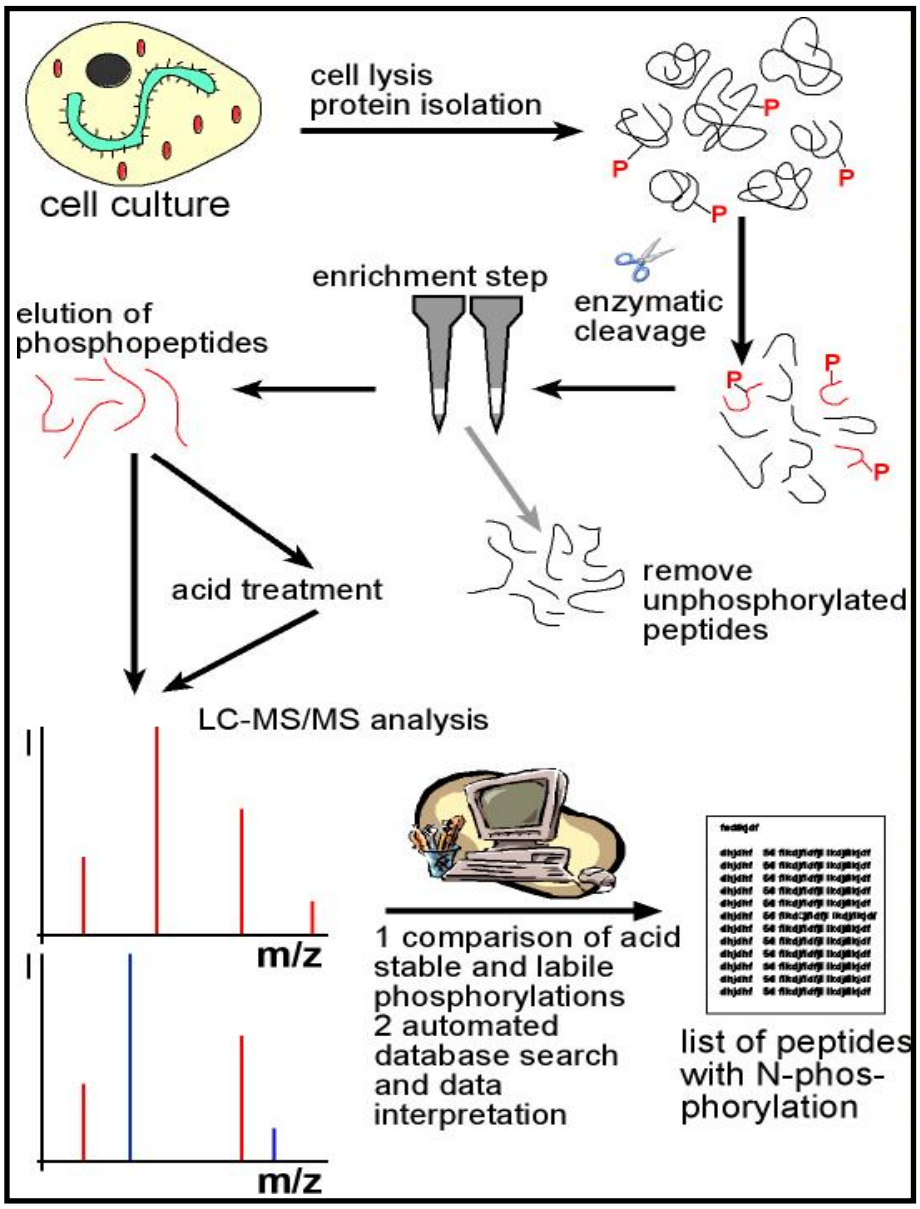


Figure B_22 **Strategy for Identification of Peptides With N-Bound Phosphorylations from Whole Cell Lysates**.

Appendix

Chemical Cross-Linking and Mass Spectrometry

Introduction

Protein Structure and Function

After synthesis in the ribosomes, most proteins undergo an extensive process, where they receive a specific folding that is often combined with the formation of intracellular disulfide bonds. This protein folding is required to guarantee the optimal function of the protein and often associated with chaperone function to diminish aggregation of misfolded or unfolded proteins. During this process, protein domains are established and contacts between different parts of the protein are correctly formed.

A loss of the specific structure normally leads to replacement and intracellular cleavage of the protein, but it can also have severe effects on the cellular health, if large aggregates are formed. Folding stresses can be associated with aggregation of the misfolded proteins to large particles, so called “inclusion bodies”, which are difficult to remove by disaggreases and proteases. Therefore, studying protein structure is a highly important field of biochemistry and many methods have been developed to reveal protein folds in different levels of resolution. The high-resolution techniques, NMR spectroscopy and X-ray crystallography, have resolved many different structures, but they have a significant bias towards stable and soluble proteins. Low-resolution methods are often not used to reveal new structures but they can deliver important details of structural changes and protein-protein interactions. Method such as cryo-electron microscopy or chemical cross-linking that result in low resolution structures are often less sensitive to buffer conditions, protein amount, purity, and stability. On the other side, data interpretation of distance constrains, which are obtained from these methods often require a high resolution structure of the protein of interest or at least a related protein or protein domains.

Chemical cross-linking of proteins

In most cases, protein complexes are not stabilized by covalent bonds, so that they associate and dissociate depending on intracellular conditions. Such alterations might be due to extra-cellular signals, drugs or a changed localization of the protein that could induce

structural changes, which favour or diminish complex formation. High-resolution techniques for structural investigations can often not be applied to visualize theses changes. This is due to the fact that their sample preparation conditions and analysis times are rather unsuitable to stabilized transient interactions.

Chemical cross-linking was introduced in the 1970s to identify specific protein complexes and identify putative new binding partners. Since the late 1990s this technique evolved from its nichè, when mass spectrometry was available to unambiguously identify binding partners as well as putative interaction sides as demonstrated in figure App_01. Since then, the combination of chemical cross-linking and mass spectrometry became a popular tool to study protein structures and protein-substrate complexes.

Upon chemical cross-linking, a small organic molecule with two defined reactive moieties is added to the protein solution, where it interconnects reactive groups of spatially near protein domains in a specific distance. The aim of the method is to maintain the structural integrity of the complex, which is achieved by incorporation of a low number of cross-links into the protein. Subsequently, this reaction mixture is often separated by gel electrophoresis to selectively enrich the cross-linking products that often have different electrophoretic mobility (Sinz A., 2003). These products are conducted to enzymatic cleavage and analysed by nanoRP-HPLC-MS/MS. Cleavage of such proteins generates up to four different types of cross-linking products: peptides, having a cross-linker molecule bound at the reactive amino acid side chain, interconnected amino acids within one peptide, and products with two connected peptide strands, stemming from the same protein or two different proteins (figure App_02A). Subsequent data interpretation is not as straight forward as simple peptide identification, since attached peptide sequences can have a high variability which is not considered by standard search algorithms.

Recently, specifically designed cross-linking reagents have been used to facilitate the identification process (Sinz A., 2003). For instance, introduction of isotopes in the spacer region of the reagent leads to a defined mass shift of ion signals of cross-linking products in the mass spectrum. Upon application of a mixture of two isotopically coded reagents for chemical cross-linking, peptide signals, which are modified by a cross-linker are easily identified by the presence of two peptide isotope patterns in the MS spectrum. Normal peptides have only a single isotopic envelope. Implementing a third functional group into the cross-linker can help to purify reaction products after enzymatic cleavage, for instance by affinity chromatography. This method improves detection of cross-linked peptides that are often observed at very low intensity, since all normal peptides are removed from the reaction

mixture. Finally, to facilitate identification of the connected peptide strands, cross-linking reagents that can be cleaved in a first fragmentation step in the mass spectrometer are available. The peptide sequences are than investigated by MS^3 fragmentation of the cleavage products

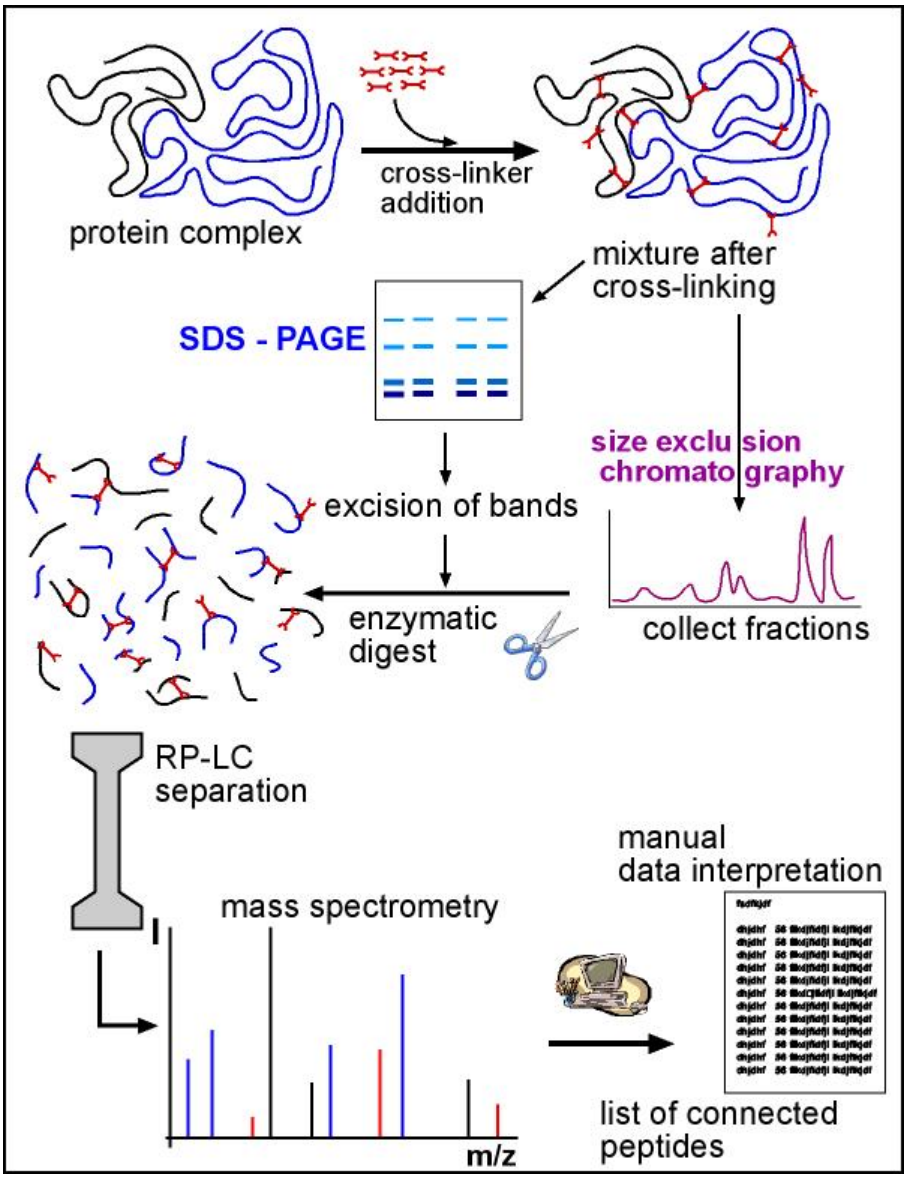


Figure App_01 **Strategy for Structural Analysis of Protein Complexes by Chemical Cross-Linking and Mass Spectrometry**. After fixation of the three-dimensional structure by incorporation of covalently bound cross-linkers, the reaction mixture is separated according to the size of the components. Enzymatic digestion produces small peptides fragments, which are analyzed by HPLC-MS/MS. Interpretation of MS data is carried out by a combination of automatic and manual interpretation steps.

Analysis and interpretation of cross-linking data in structural context requires at least a low resolution structure of the investigated protein or a structural homologue. First cross-

linked peptides are identified based on their mass. Additionally, a database consisting of all putative cross-linking products has to be generated. Therefore, the masses of both peptides and an additional mass shift of the cross-linker have to be added for all possible combinations of interconnected peptide. Due to the high number of theoretically possible cross-linking products, the number of protein entries is often resticted. In the second step candidates are evaluated on the basis of their fragment ion spectra. Despite some effort has been put into automated analysis of MS/MS spectra of cross-linked peptides in the recent years, data processing is still a laborious step and often only manual inspection allows unambiguous assignment. This is mainly due to fact that two different peptide sequences are represented in the fragment ion spectrum and the modification at the cross-linked amino acid is highly variable in its size and composition as the schematic demonstration in figure App_02B shows.

Next, the identified interconnected regions in the protein have to be matched into a structure model with regard to the spacer length of the cross-linker and the flexibility of the protein chains. The length of the spacer-region between the two reactive moieties of the cross-linker defines the maximal distance for this linkage and can be used as a distance constrain to map structural features.

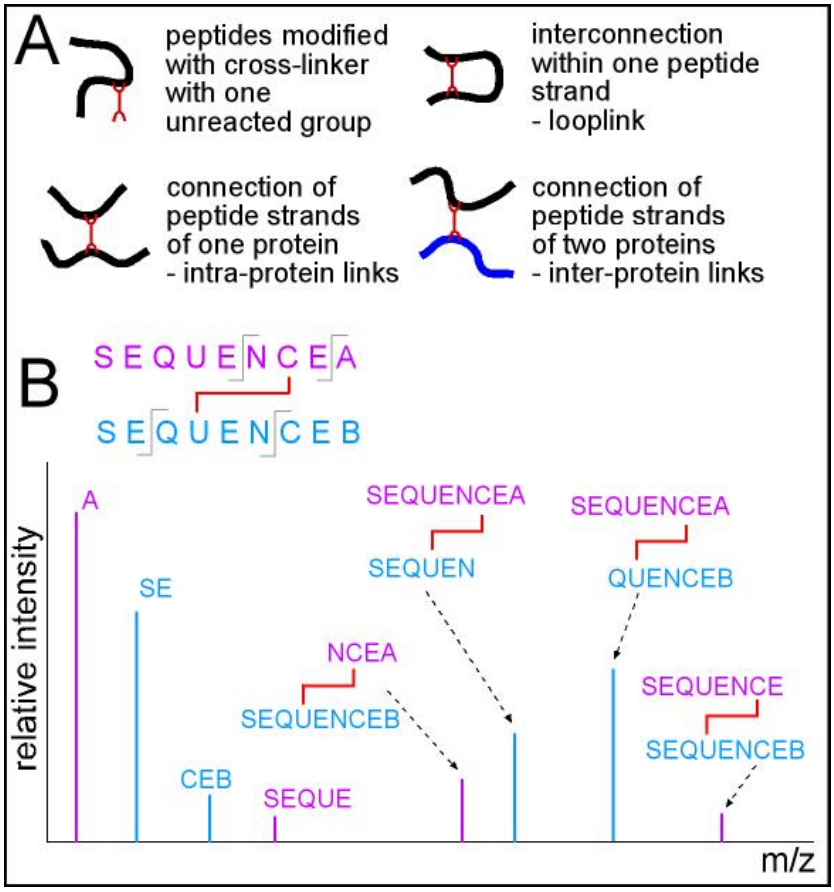


Figure App_02 **Cross-Linking Products and MS/MS Fragmentation**. A Cross-linking of two different proteins gives rise to four possible products. B Fragmentation analysis of cross-linking products leads to complex fragment ion spectra. This complexity is caused by backbone fragmentations of both interconnected peptide strands and the variability of the modification for each cross-linking product and fragment ion.

A related method uses mono-reactive reagents of a defined reactivity, for instance towards primary amines, to map the surface accessibility of protein domains. These data can further be used to refine the topology and to identify localisation of proteins and their domains within a complex. Data from protein labeling experiments are more difficult to interpret based on the protein or complex structure, but incubation of single proteins and comparison to the protein complex can easily reveal the interacting peptide stretches. This technique requires significant complex stability and is not applied to study transient protein interactions.

Projects

A In co-operation with Prof. Dr. A. Sinz and M. Müller of the University of Halle in Germany we investigated the structural changes of the PPAR2α receptor upon interaction with two different inhibitors.

Recently, chemical cross-linking in combination with mass spectrometry was used to determine interaction epitopes of proteins with peptides or larger biomolecules such as other proteins or DNA. Protein-small molecule interactions can also induce decent structural changes but these were believed to be indistinguishable from the natural flexibility of protein structures when analyzed with chemical cross-linking. To demonstrate different conformational changes in peroxisome proliferator-activated receptors (PPARs) caused by the binding of two different antagonists (GW6471, IC_{50}=0.24 μM and YS81, IC_{50}=7 μM), the cross-linking protocol and LC-MS/MS analysis were optimized. PPARs are nuclear receptors which sense fatty acids and eicosanoids as well as anti-diabetic drugs (Qi C. *et al.*, 2000; Plutzky J., 2003). To bind to specific DNA sequences and act as transcription factor, these proteins dimerize with retinoic X receptors (Issemann I. *et al.*, 1993). The ligand binding domain of PPARα was expressed, purified, and cross-linked in the presence of 10 μM of one of the two antagonists. To determine cross-links which are due to conformational changes, results were compared to a sample without incubation with inhibitor. The cross-linking reagents, bis-sulfosuccinimidyl-glutarate BS^2G and bis-sulfosuccinimidyl-suberate BS^3, were amine reactive and targeted lysine side chains. In individual experiments, cross-linkers were added in 50 and 100 molar excess over the purified receptor and incubated for 30 and 60 min. Samples were digested "*in gel*" by combination of trypsin and Asp-N to generate small peptide strands which could efficiently be eluted from the polyacrylamide gel. Sample analysis was performed by nanoC18RP-HPLC on an Ultimate Dual Gradient system which

was directly coupled to an LTQ Orbitrap hybrid mass spectrometer. The mass spectrometer was operated in positive ionization mode.

To improve detection of cross-linking products by mass spectrometry, two different fragmentation techniques were applied for fragmentation of eluting peptides. After detection of the peptide signals in the survey scan, the three most abundant precursor ions were selected for fragmentation by low-energy CID MS/MS in the ion trap part and a second fragmentation of the same precursors was carried out in the C-trap under higher collision energy. Fragmentation of cross-linked peptides under standard conditions produces often an insufficient number of fragments to identify both strands and it is associated with extensive neutral losses of small molecules such as water or ammonia. High-energy fragmentation induces more backbone fragmentations and provides high resolution and high mass accuracy since these scans are measured in the orbitrap analyzer.

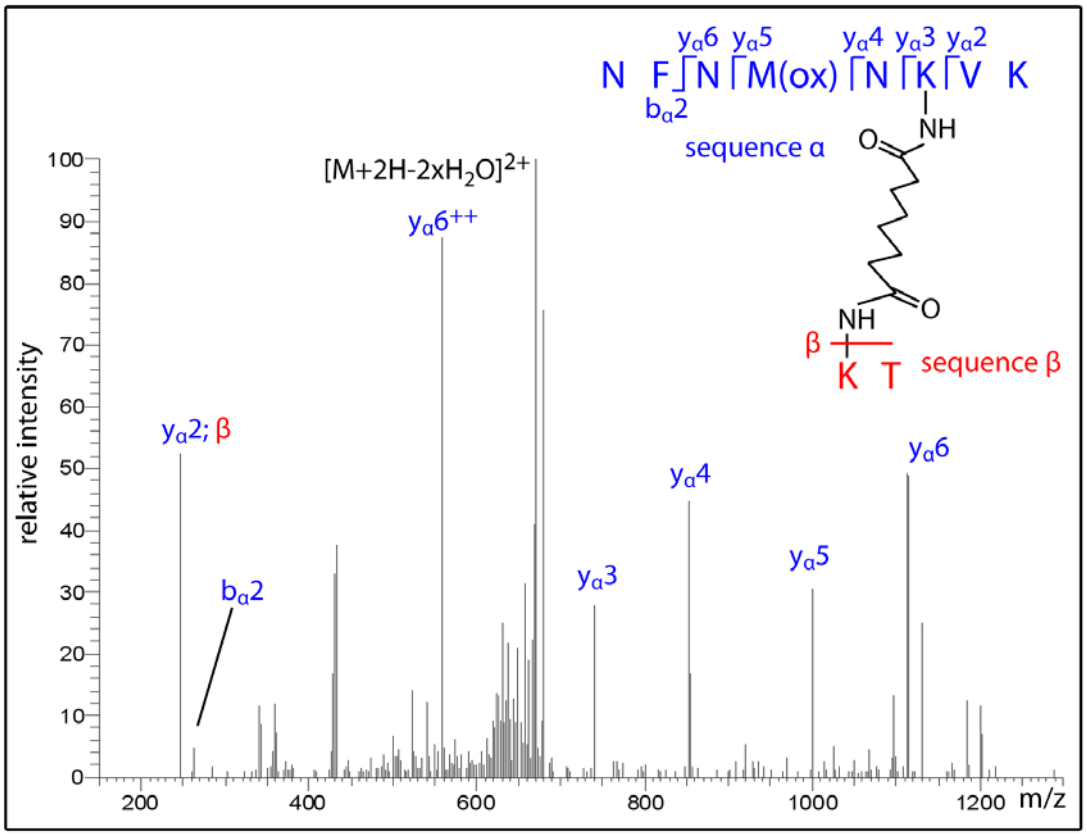


Figure App_03 **MS/MS spectrum of the identified cross-link NFNMNKVK-KT from PPAR2α**. After incubation with the inhibitor, cross-linker BS^3 was added to the reaction mixture. The resulting cross-link sample was enzymatically cleaved with a combination of the proteases AspN and trypsin and subjected to HPLC-MSMS analysis. The presented cross-link was identified from the fragment ion spectrum and the two isotopic envelopes of the precursor signal which resulted from the reaction with isotopic variants of the cross-linker with 0 and 4 deuterium atoms.

Obtained data for both inhibitors indicate different binding modes of the two substances to the PPAR ligand binding domain. While GW6471 induces significant conformational changes which are indicated by the newly formed cross-links between Lys222/Lys449 and Lys252/Lys449. These cross-links were not found in the free receptor form or the sample with YS81 as inhibitor. A possible explanation for the two binding modes comes from the size of the inhibitor molecules, while GW6471 is able to fill the binding pocket completely, YS81 is too small and will leave higher flexibility. The obtained distance constraints were compared to the crystal structure of the ligand binding domain of PPAR2α in complex with the antagonist GW6471.

B Localization and orientation of protein subunits of the core complex of salmonella's Type III secretion system were studied with acetyl labelling and chemical cross-linking of lysines in co-operation with Dr. T. Marlovits, O. Schraidt and M. Brunner (Institute for Molecular Biotechnology, IMBA, Vienna)

Structural studies of integral membrane proteins and protein complexes is a challenging task because these proteins tend to precipitate even at low concentrations which often impairs crystallization and NMR analysis. For the investigation of large integral membrane protein complexes EM is often the method of choice (Marlovits T.C. *et al.*, 2004). Unfortunately, the obtained low resolution structure often misses the protein orientation localization of individual proteins and protein domains. By combination of EM and chemical cross-linking/labelling we were able to refine the structure of the core complex of the Type III secretion system (needle complex) of salmonella thyphimurium.

The core complex of the type III secretion system is approx. 3 MDa large and consists of 5 different proteins which are present in multiple copies (figure App_04, Zhou D. and Galán J., 2001). These proteins, InvG, PrgH, PrgK, PrgJ, and PrgI interact with numerous other proteins in the cytoplasm such as chaperones which relocate the complex within the membrane or prepare the secreted effector proteins. Two proteins PrgH and PrgK form a ring-like structure which permeates the inner cell membrane of salmonella. These proteins interact with chaperones and disassembled proteins which enter the channel of the needle at this position. PrgH is thereby located at the outer rim, while PrgK forms a central ring and the entrance to the secretion channel (Marlovits T.C. *et al.*, 2004).

Attached to the outer membrane is a second ring which consists of the proteins InvG and PrgJ. It is widely accepted that InvG is the major structural component of this ring which crosses the outer bacterial membrane, whereas PrgJ is assumed to form the inner basis of the outstretching needle. InvG is also involved in protein-protein interactions with proteins in the outer membrane which stabilize and localize the needle. Finally, multiple PrgI subunits form the needle tip which then interacts with the host cell membrane to transfer the secreted proteins.

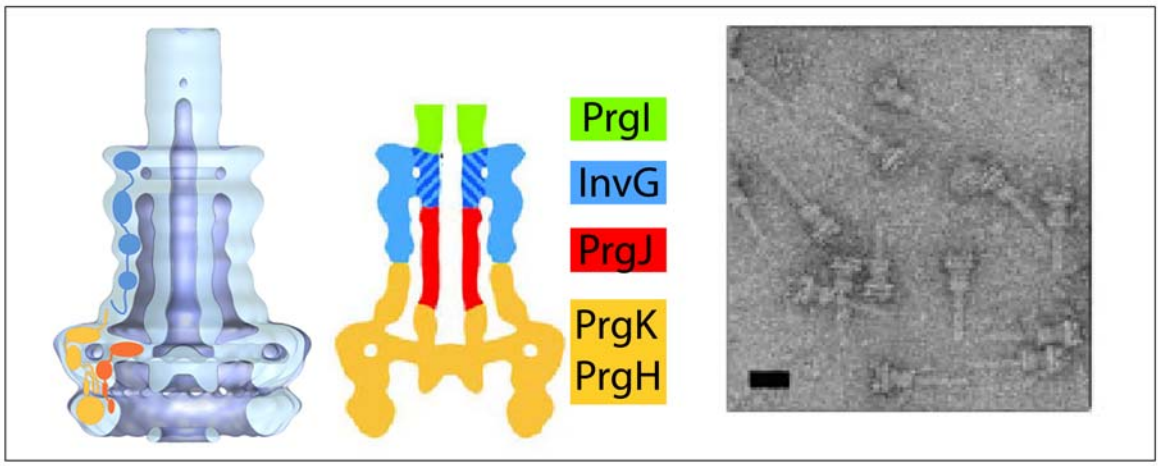


Figure App_04 **The Core Complex of the Samonella Type III Secretion System**. The so-called needle complex consists of multiple copies of 5 proteins, InvG, PrgH, PrgI, PrgJ, and PrgK. PrgK (orange in left structure) and PrgH (yellow) form the intracellular ring which crosses the inner membrane. The outer membrane ring consists of the protein InvG (blue) which stabilizes the needle rod which is constructed from PrgJ subunits (red). The needle tip which interacts with the host cell membrane is solely made up from PrgI (green) molecules. In EM, all structures of the complex can be identified, but localization and orientation of individual proteins is elusive.

Individual protein species were removed by mutation of the respective gene by our coworkers in the laboratory of T. Marlovits. To enlighten the location of the subunits and the altered surface accessibility, we set up an acetyl labelling experiment. The wild type (wt) complex and the mutant which lacks the InvG protein, and therefore the outer membrane ring (figure App_05A EM images, left wt complex, right InvG mutant), were labelled on lysine residues by N-sulfosuccinimidyl acetate. Equal amounts of both complexes were digested with chymotrypsin and analyzed by nanoRP-HPLC-MS/MS. To quantify the incorporation of

the acetyl label, signal areas for identified peptides which posses a lysine residue were extracted for the label-free and the acetylated form. By comparing the percentage of acetyl incorporation between the two complex forms, we found a significant increase of acetylation on all peptides of PrgJ (data not shown) and for Lys168 of PrgK (see figure App05A) in the InvG mutant complex. On average, PrgJ peptides showed six times higher acetylation while the acetylation of other lysines in the whole complex was not affected by the mutation. The first observation confirms the location of the PrgJ to the center of the complex, while it is not accessible for the reagent in the wt complex, removal of the surface protein InvG facilitates acetylation. Mapping of Lys168 of PrgK into a model of the complex revealed its location to the interface between inner membrane ring and outer membrane ring (figure App_05A). Removal of the InvG protein brings this residue to the newly formed surface and therefore enhances reaction with the acetylation reagent.

To study such large protein complexes by chemical cross-linking, we had to adapt the complete sample preparation procedure and LC-MS/MS method to increase the sensitivity. Cross-Linking was attained by addition of BS_2G in non-deuterated and fourtimes deuterated isotope–coded form to facilitate identification of cross-linking products. Thereby, two isotope envelopes are generated in the MS scan which are separated by 4 m/z. Separation of the cross-linked sample by SDS-Page revealed covalently connected dimers of all components of the complex and heterodimers of InvG with PrgI and PrgJ and PrgH with PrgK. SDS-PAGE delivers a good resolution of protein but the subsequent enzymatic cleavage has low efficiency and especially long polypeptide chains remain in the gel upon elution which excluded them from MS analysis. The in-gel processed samples delivered no interconnected peptides. Since cross-linking expects the formation of bigger peptide assemblies we decided to perform an in solution digest. Due to the nature of the transmembrane domains, the protein complex required the presence of a detergent to be stable in solution which reduces enzymatic activity of the cleaving protease dramatically. Therefore we chose to add enzyme in a sample/enzyme ratio of 15:1 in 4 steps. The second addition was made after 3 h of digestion and then every 10 h and the digest was carried out for 35 h (over two nights). Since no effective orthogonal separation technique such as strong cation exchange chromatography was available for the limited sample amount, we applied a very long gradient on the nano-C18RP-HPLC separation system to guarantee efficient separation of individual peptide species. Mass spectrometric analysis was performed with the double fragmentation method (see project 1) which produces two fragment ion spectra for an individual peptide species by low-energy CID and high-energy HCD collision. Extended gradient time and the specific

detection method increased the size of raw data files to more than 1.5 GB which required data reduction before automatic database search with the Xquest software (Rinner O. *et al.*, 2008).

A second problem of cross-linking is detection and identification of products by mass spectrometry. Due to the chemical nature of the deuterium, the heavy peptide form eluted first from the column and was therefore more often selected for fragmentation analysis than the light form. This underrepresentation of the light isotope is due to limitations of the mass spectrometry software to identify such isotope pairs as cross-linking products of the same peptide. Normally such signals are excluded, because fragmentation would result in a mixed MS/MS spectrum of two different peptide sequences which cannot be interpreted by actual MS interpretation algorithms.

Chemical cross-linking of the Salmonella needle complex revealed interactions of the N-terminal domain of InvG with the C-terminal domain of PrgH. Thereby, Lys 38 of InvG and Lys368 in the PrgH sequence were covalently connected. This protein-protein interaction is quite interesting because it is located in between the two transmembrane rings of the complex. Deletion of the PrgH C-terminus by mutation in later experiments was associated with higher instability between these domains which lead to detachment of the inner membrane ring upon shearing forces during EM sample preparation. Further, the cross-linker connected the N-terminal PrgH domain with the C-terminal domain of PrgK at the basis of the inner membrane ring. This region faces towards the cytoplasm and is therefore highly accessible for modification. Based on these data we were able to determine the orientation of three components of the needle complex. Both proteins of the inner membrane ring, PrgH and PrgK, face into opposite directions. The N-terminus of PrgH is located in the cytoplasm and the C-terminus in the periplasmatic region between inner and outer membrane. The C-terminus of PrgK is mainly cytoplasmatic and interacts with the PrgH N-terminus, while the N-terminus protrudes through the membrane and forms the inner basis of the needle structure. The InvG protein adopts the same orientation as PrgH. The N-terminus faces towards the inner membrane, while the C-terminal domain perforates the outer cell membrane.

Cross-linking products were not found for the other two proteins which might be a consequence of the location of PrgJ and the amino acid composition of PrgI. PrgJ is found in the center of the structure and has no accessible surface area apart from the channel which might be to small for the cross-linking reagent to penetrate. Furthermore, it is not known which amino acids form this channel.

Modification of PrgI lysines leads to very hydrophobic peptides after enzymatic digest, therefore these peptides might be lost in the RP-separation step.

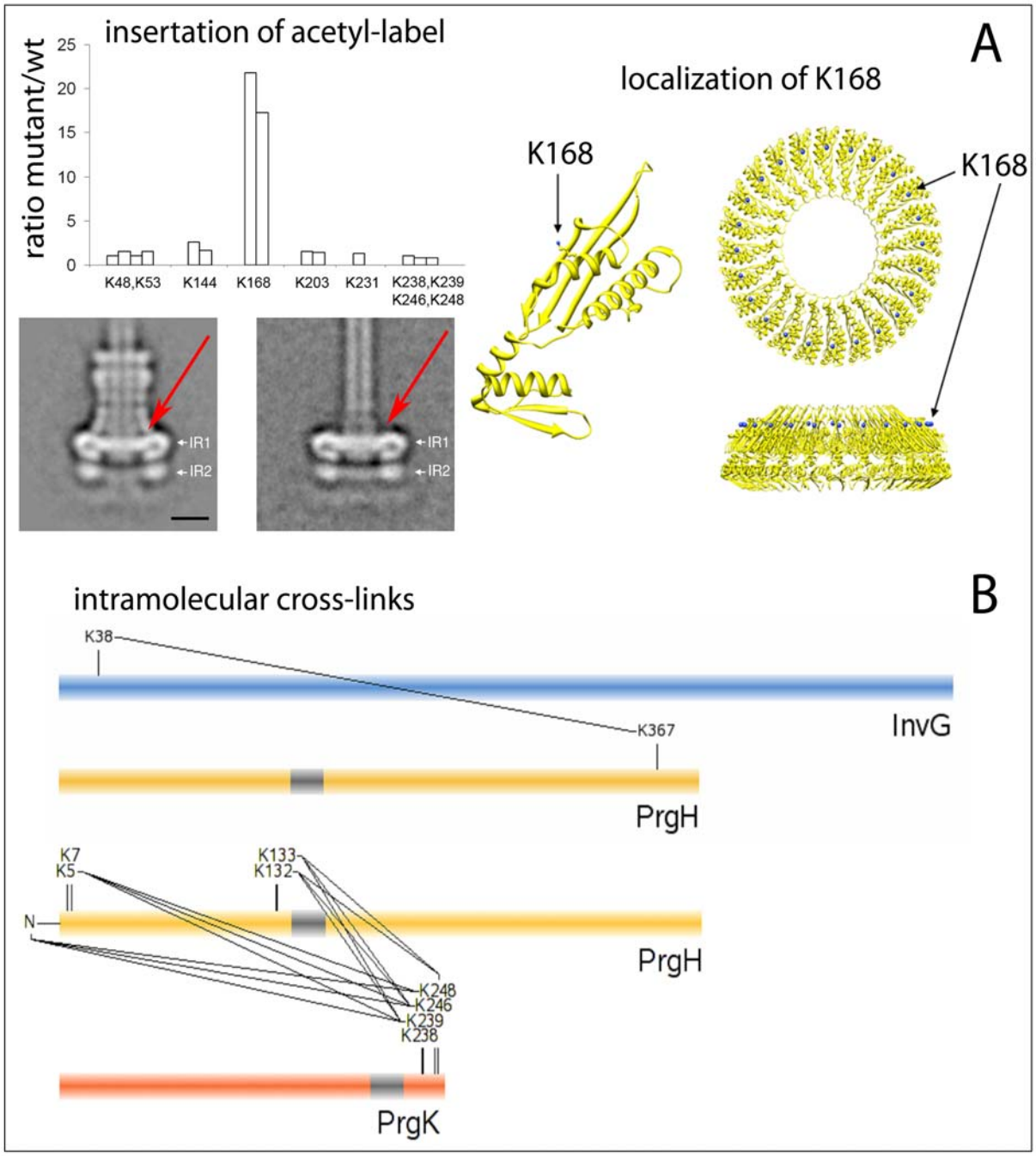


Figure App05 **Results for Acetyl-Labelling and Connected Domains Between Different Protein Species in the Needle Complex**. Acetyl labeling reveals enhanced incorporation of the label to Lys168 of PrgK when InvG is deleted from the complex. In the wt complex (left EM image) this residue is shielded by the outer membrane domain, whereas in the mutant complex (right EM) this residue is located at the newly formed surface. Modelling of the PrgK structure and the inner ring formed of PrgK subunits confirms localization of this residue to the interface between inner and outer membrane domains of the complex and explains higher accessibility in the InvG mutant. B Cross-linking analysis demonstrated interaction of the N-terminus of PrgH with the C-terminus of InvG at the interface of the two

membrane rings. The PrgH N-terminus and the PrgK C-terminus are located within close proximity in the intracellular domain of the needle complex. Gray bars indicate transmembrane domains.

C Interaction site determination and orientation of Myotilin-IG domains upon dimerization were studied in co-operation with Prof. K. Djinovic-Carugo (University of Vienna, Vienna)

Immunoglobin (IG)-domains are specific protein-protein interaction domains which are very common in proteins of the cytoskeleton. Thereby, two structurally similar domains interact with IG-domains of a different protein molecule and form tight complexes. To study the dimerization of protein IG-domains of myotilin, these domains were individually expressed with an N-terminal His-tag and purified. Dimer formation was monitored by small angle X-ray scattering (SAX) and interacting regions of the dimer should be identified by chemical cross-linking and mass spectrometry. The preformed complex was incubated with two cross-linking reagents in individual experiments. One cross-linker was ethylene glycol bis[succinimidylsuccinate] (EGS) which has a spacer length of 16.1 Å and interconnects two lysine-ε-amino groups. The second reagent system was EDC/NHS which represents a zero-length cross-linker that forms an amide bond between spacially close amine and carboxylic acid functions of Lys and Asp/Glu. Reaction products were separated by SDS-PAGE, protein bands which correspond to monomeric and dimeric forms were excised, tryptically digested, and analyzed by HPLC-MS/MS as described previously.

Both reagents were not isotopically labelled, so the easily recognizable isotopic envelopes in MS spectra were missing. Therefore, a different strategy for data interpretation and distinction of cross-linking products had to be established. After extraction of peptide signals from the survey scans, all signals in the range of 500-5000 m/z were compared to a set of theoretical cross-linking products using the GPMAW software (www.gpmaw.com). A sample which was not subjected to cross-linking was also matched against the database to reveal peptides which have similar mass to putative cross-linking products. Such false positives were discarded. Overlapping cross-links of dimeric and monomeric proteins were separated from the dimer-only matches since these cross-links define the three-dimensional structure within one subunit, but have no information on the interaction epitope. For further verification of a dimer specific cross-link, the abundance of the peptide signal was compared

between monomeric and dimeric sample. If the same signal was found only in the dimer sample, the link would be specific for the interaction of two IG domains.

Using this strategy we derived a single dimer specific link EHKR-EHK which interconnected the same amino acid regions of two IG domains see figure APP_05. The number of intramolecular links was significantly higher (table App_1).

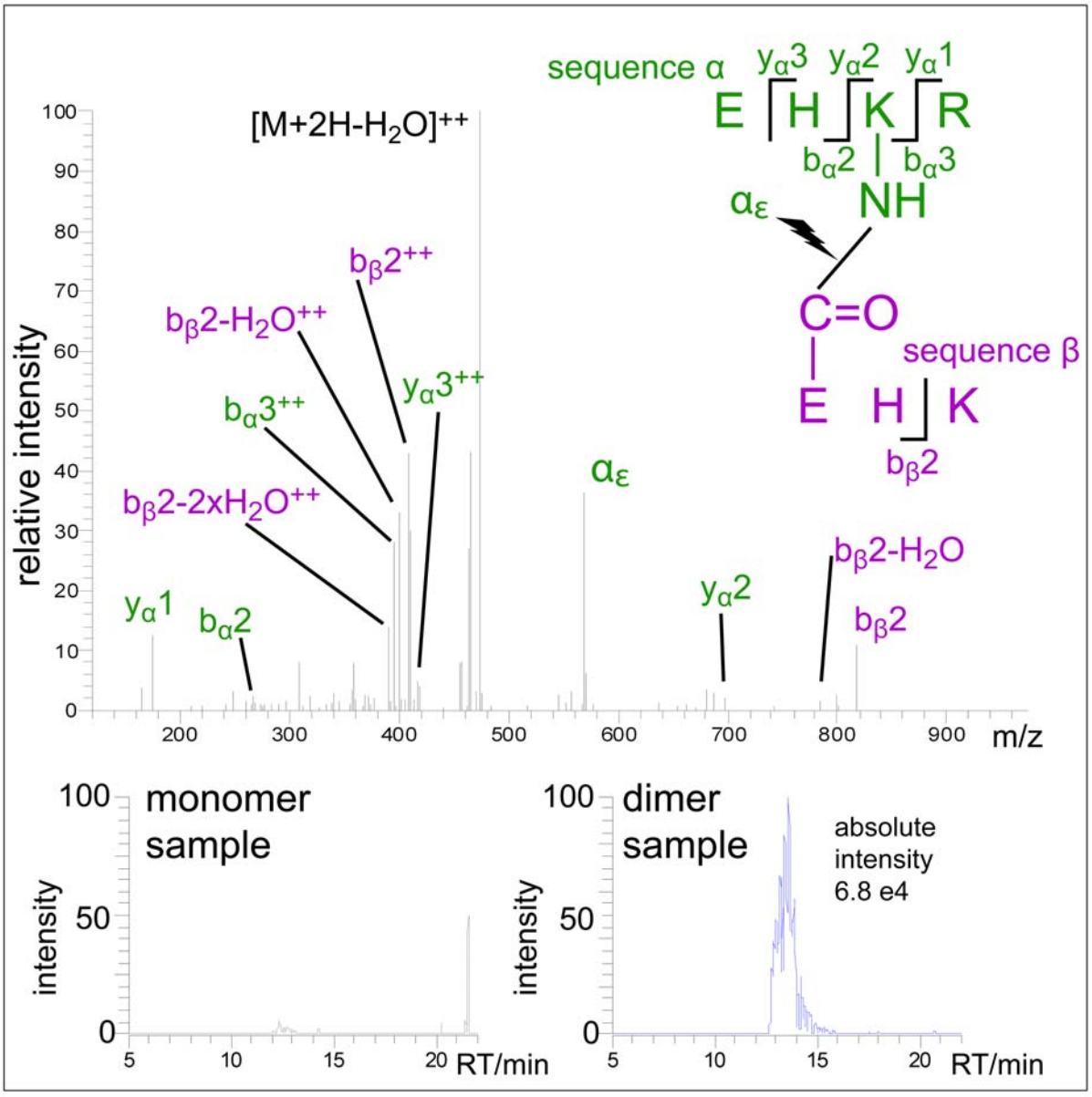


Figure App06 **Connection of the Same Peptide Strands Identifies the Dimerization Region of Myotilin-IG Domains.** The fragment ion spectrum clearly identified both peptide sequences of the cross-link EHKR-EHK (amino acids sequence A 153-156, sequence B 153-155) in the IG domain dimer. Sequence α is well identified by many fragment ions, while sequence β delivers only a single fragmentation. This ion is verified by two charge states and multiple water losses.

This cross-link product is exclusively found in the dimeric sample at a retention time of approx. 13 min; the same signal is completely absent in the monomeric sample.

Table App_1 **Identified Cross-Linking Products of Myotilin-IG Domains Cross-Linked with EDC/NHS**.

mass			peptide sequences			
observed	theoretical	deviation	A	aa	B	aa
963.512	963.511	0.001	EHK	153-155	EHKR	153-156
1290.709	1290.710	-0.001	DVNKK	221-225	MIVSEK	107-112
1396.754	1396.754	0.000	EHKR	153-156	VLEGDSVK	169-176
1430.785	1430.786	-0.001	DVNKK	221-224	VLEGDSVK	169-176
1536.911	1536.910	0.001	EHK	153-155	VTLLIKDVNK	215-224
1558.880	1558.881	-0.001	DVNKK	221-224	KVLEGDSVK	168-176
1576.885	1576.884	0.001	LFWKR	189-193	VLEGDSVK	169-176
1592.814	1592.814	0.000	EHKR	153-156	TVQSDDLHK	98-106
1626.845	1626.845	0.000	DVNKK	221-224	TVQSDDLHK	98-106
1772.944	1772.943	0.001	LFWKR	189-193	TVQSDDLHK	98-106
1917.898	1917.901	-0.003	NNEMVQFNTDR	194-204	EHKR	153-156
1970.153	1970.156	-0.003	VTLLIKDVNK	215-224	VLEGDSVK	169-176
2166.213	2166.210	0.003	VTLLIKDVNK	215-224	TVQSDDLHK	98-106
2323.135	2323.132	0.003	NNEMVQFNTDR	194-204	KVLEGDSVK	168-176

Appendix

Detection of Methyltransferase andDemethylase Activity on Small Model Peptides by Mass Spectrometry

The project was a co-operation with the laboratory of T. Jenuwein (now MPI for Immunology, Freiburg). Protein tagging and purification was performed by the members of the TJ-Lab.

Numerous reversible protein modification have been identified in the N-terminal domains of histones, the so-called “histone tails”. Since histones are tightly associated with DNA, these modifications are essential for the regulation of transcriptional processes and they comprise acetylation or methylation of lysines or arginines and phosphorylation of serines. Acetylation and phosphorylation are generally associated with transcriptional activation, while methylation often causes silencing of the respective DNA region. Furthermore, binding of specific transcription factors depends on the modification state of the DNA.

In most cases methylation and acetylation are antagonistic processes which are found on the same amino acids within the histone sequence in different cell states. The field of epigenetics is quite young, so that only a limited number of histone modifying enzymes has been described so far.

To test a higher number of putative enzymes with demethylase activity we developed a fast assay to detect removal of the methylation mark on small model peptides. These peptide sequences perfectly resemble the N-terminal strands of different histone variants (the so-called “histone tails”). The peptides already carry the respective lysine methylation which should be removed by the enzyme. In demethylase assays, candidate proteins are tagged for purification and the purified protein is incubated with individual model peptides. Assay conditions were optimized with regard to detergents, salt concentration, and buffer composition to allow direct MS analysis of the mixture. Therefore, samples were acidified and diluted with the matrix substance and directly placed on a MALDI sample plate. 4000 to 6000 sub spectra were automatically acquired for one MALDI MS spectrum which was then investigated for the presence of peptides in lower methylation states. The signals of these reaction products should exhibit a mass decrease of 14 Da in comparison to the precursor peptide.

Using this protocol, we confirmed demethylase activity of the proteins Su(var)3-3 and JMJD2 as well as methyltransferase activity of SetD6.

References

Agrawal G.K., Thelen J.J., Large scale identification and quantitative profiling of phosphoproteins expressed during seed filling in oilseed rape. *Molecular and Cellular Proteomics* **2006**, 5, 2044–2059

Alpert A.J. Electrostatic Repulsion Hydrophilic Interaction Chromatography for Isocratic Separation of Charged Solutes and Selective Isolation of Phosphopeptides. *Analytical Chemistry* **2008**, 80 (1), 62-76

Annan R.S. and Carr S.A., Phosphopeptide analysis by matrix-assisted laser desorption time-of-flight mass spectrometry. *Analytical Chemistry* **1996**, 68, 3413–3421

Arnott D., Shabanowitz J., Hunt D.F. Mass Spectrometry of Proteins and Peptides: Sensitive and Accurate Mass Measurement and Sequence Analysis. *Clinical Chemistry* **1993**, 39 (9), 2005-2010

Attwood P.V., Piggott M.J., Zu X.L., Besant P.G. Focus on Phosphohistidine. *Amino Acids* **2007**, 32 (1), 145-156

Bailyes E.M., Bennett D.L., Hutton J.C., Proprotein-processing endopeptidases of the insulin secretory granule. *Enzyme* **1991**, 45, 301-313

Bantscheff M., Eberhard D., Abraham Y., Bastuck S., Bösche M., Hobson S., Mathieson T., Perrin J., Raida M., Rau C., Reader V., Sweetman G., Bauer A., Bouwmeester T., Hopf C., Kruse U., Neubauer G., Ramsden N., Rick J., Küster B., Drewes G. Quantitative chemical proteomics reveals mechanisms of action of clinical ABL kinase inhibitors. *Nature Biotechnology* **2007**, 25, 1035-1044

Beausoleil S.A., Jedrychowsky M., Schwartz D., Elias J.E., Villen J., Li J., Cohn M.A., Cantley L.C., Gygi S.P. Large-scale characterization of HeLa cell nuclear phosphoproteins. *Proceedings of the National Academy Sciences* **2004**, 101, 12130-12135

Besant P.G., Attwood P.V., Piggott M.J. Focus on Phospholysine and Phosphoarginine. *Current Protein and Peptide Science* **2009**, 10 (6), 536-550

Blume-Jensen P. and Hunter T. Oncogenic Kinase Signalling. *Nature* **2001**, 411, 355-365

Bodenmiller B., Müller L.N., Müller M., Domon B., Aebersold R. Reproducible isolation of distinct, overlapping segments of the phosphoproteome. *Nature Methods* **2007**, 4, 2331-237

Boersema P.J., Mohammed S., and Heck A.J.R. Phosphopeptide fragmentation and analysis by mass spectrometry. *Journal of Mass Spectrometry* **2009**, 44, 861-878

Cantin G.T., Shock T.R., Park S.K., Madhani H.D., Yates III J.R. Optimizing TiO2-based phosphopeptide enrichment for automated multidimensional liquid chromatography coupled to tandem mass spectrometry. *Analytical Chemistry* **2007**, 79, 4666–4673

Cavalli G. and Paro R. Chromo-domain proteins: linking chromatin structure to epigenetic regulation. *Current Opinion in Cell Biology* **1998**, 10, 354-360

Chang E.J., Archambault V., McLachlin D.T., Krutchinsky A.N., Chait B.T. Analysis of protein phosphorylation by hypothesis-driven multiple-stage mass spectrometry. *Analytical Chemistry* **2004**, 76, 4472–4483

Chang F., Steelmans L.S., Lee J.T., Shelton J.G., Navolanic P.M., Blalock W.L., Franklin R.L., McCubrey J.A. Signal transduction mediated by the Ras|Raf|MEK|ERK pathway from cytokine receptors to transcription factors: potential targeting for therapeutic intervention. *Leukemia* **2003**, 17, 1263-1293

Choi H., Lee H., Park Z.-Y., Detection of Multiphosphorylated Peptides in LC-MS/MS Analysis under Low pH Conditions. *Analytical Chemistry* **2008**, 80, 3007-3015

Claverol S., Burlet-Schiltz O., Gairin J.E., Monsarrat B. Characterization of Protein Variants and Post-Translational Modifications: ESI-MSn Analysis of Intact Proteins Eluted from Polyacrylamide Gels. *Molecular and Cellular Proteomics* **2003**, 2, 483-493

Cohen P. The role of protein phosphorylation in neural and hormonal control of cellular activity. *Nature* **1982**, 296 (5858), 613-620

Curtin N.A. and Woledge R.C. Energy Changes and Muscular Contraction. *Physiological Reviews* **1978**, 58, 690-761

Derre I., Rapoport G., Msadek T. CtsR a novel regulator of stress and heat-shock response controls *clp* and molecular chaperone gene expression in *Gram*-positive bacteria. *Molecular Microbiology* **1999**, 31 (1), 117-131

Fenn J.B., Mann M., Meng C.K., Wong S.F., Whitehouse C.F. Electrospray Ionization for Mass Spectrometry of Large Biomolecules. *Science* **1989**, 246, 64-71

Ferri N., Paoletti R., Corsini A. Lipid-modified proteins as biomarkers for cardiovascular disease: a review. *Biomarkers* **2005**, 10, 219-237

Ficarro S.B., McCleland M.I., Stukenberg P.T., Burke D.J., Ross M.M., Shabanowitz J., Hunt D.F., White F.M. Phosphoproteome analysis by mass spectrometry and its application to Saccharomyces cerevisiae. *Nature Biotechnology* **2002**, 20, 301–305

Flora J.W. and Muddiman D.C. Selective, Sensitive, and Rapid Phosphopeptide Identification in Enzymatic Digests Using ESI_FTICR_MS with Infrared Multiphoton Dissociation. *Analytical Chemistry* **2001**, 73, 3305-3311

Freeman J.A., Lilley B.N., Bassler B.L. A genetic analysis of the function of LuxN: a two-component hybrid sensor kinase that regulates quorum sensing in *Vibrio harveyi Molecular Microbiology* **2000**, 35, 139-149

Fuhrmann J., Schmidt A., Spiess, S., Lehner A., Turgay K., Mechtler K., Charpentier E., Clausen T., McsB Is a Protein Arginien Kinase That Phosphorylates and Inhibits the Heat-Shock Regulator CtsR., *Science* **2009**, 324, 1323-1327

Gao Y., Wang Y. A method to determine the ionization efficiency change of peptides caused by phosphorylation. *Journal of the American Society for Mass Spectrometry* **2007**, 18, 1973–1976

Gassner M., Stehlik D., Schrecker O., Hengstenberg W., Maurer W., Rüterjans H. The Phosphoenolpyruvate-Dependent Phosphotransferase System of *Staphylococcus aureus*. 2. 1H and 31P Nuclear-Magnetic-Resonance Studies on the Phosphocarrier Protein HPr, Phosphohistidines and Phosphorylated HPr. *European Journal of Biochemistry* **1977**, 75, 287-296

Ge Y., Lawhorn B.G., ElNaggar M., Strauss E., Park J.-H., Begley T.P., McLafferty F.W. Top Down Characterization of Larger Proteins (45 kDa) by Electron Capture Dissociation Mass Spectrometry. *Journal of the American Chemical Society* **2002**, 124 (4), 672-678

Ge Y., Lawhorn B.G., ElNaggar M., Sze S.K., Begley T.P., McLafferty F.W. Detection of four oxidation sites in viral prolyl-4-hydroxylase by top-down mass spectrometry. *Protein Science* **2003**, 12 (10), 2320-2326

Gov Y., Borovok I., Korem M., Singh V.K., Jayaswal R.K., Wilkinson B.J., Rich S.M., Balaban, N. Quorum Sensing in Staphylococci is Regulated via Phosphorylation of Three Conserved Histidine Residues. *The Journal of Biological Chemistry* **2004**, 279, 14665-14672

Gunawardena H.P., Emory J.F., McLuckey S.A. Phosphopeptide anion characterization via sequential charge inversion and electron-transfer dissociation. *Analytical Chemistry* **2006**, 78, 3788–3793

Hecker M., Schumann W., Völker U. Heat-shock and general stress response in *Bacillus subtilis*. *Molecular Microbiology* **1996**, 19, 417-426

Hicks G.R. and Raikhel N.V. Protein Import into the Nucleus: An Integrated View. *Annual Reviews of Cell and Developmental Biology* **1995**, 11, 155-188

Hoch J.A. and Silhavy T.J. Two-component Signal Transduction. *ASM Press Washington DC*, ISBN 1-55581-089-6, **1995**

Hunt D.F., Yates III. J.R., Shabanowitz J., Winston S., Hauer C.R. Protein sequencing by tandem mass spectrometry. *Proceedings of the National Academy of Sciences* **1986**, 83, 6233-6237

Hunter T. Signalling – 2000 and Beyond. *Cell* **2000**, 100, 113-127

Issemann I., Prince R.A., Tugwood J.D., Green S., The Peroxisome Proliferator-Activated Receptor:Retinoid X Receptor Heterodimer is Activated by Fatty acids and Fibrate Hypolipidemic Drugs. *Journal of Molecular Endocrinology* **1993**, 11, 37-47

Janek K., Wenschuh H., Bienert M., Krause E. Phosphopeptide analysis by positive and negative ion matrix-assisted laser desorption/ionization mass spectrometry. *Rapid Communications in Mass Spectrometry* **2001**, 15, 1593–1599

Jensen S.S., and Larsen M.R. Evaluation of the impact of some experimental procedures on different phosphopeptide enrichment techniques. *Rapid Communications in Mass Spectrometry* **2007**, 21, 3635–3645

Johnsson A.P. Mass spectrometry for protein and peptide characterization. *Cellular and Molecular Life Sciences* **2001**, 58, 868-884

Kalkum M., Lyon G.J., Chait B.T. Detection of secreted peptides by using hypothesis-driven multistage mass spectrometry. *Proceedings of the National Academy of Sciences* **2003**, 100, 2795–2800

Karas M. and Hillenkamp F. Laser Desorption Ionization of Proteins with Molecular Masses Exceeding 10,000 Da. *Analytical Chemistry* **1988**, 60 (20), 2299-2301

Keyse S.M. An emerging family of dual specificity MAP kinase phosphatases. *Biochimica et Biophysica Acta* **1995**, 1265, 152-160

King R.W., Jackson P.K., Kirschner M.W. Mitosis in Transition. *Cell* **1994**, 79, 563-571

Kirstein J. and Turgay K. A New Tyrosine Phosphorylation Mechanism Involved in Signal Transduction in *Bacillus subtilis*. *Journal of Molecular Microbiology and Biotechnology* **2005**, 9, 182-188

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*. *The EMBO Journal* **2005**, 24, 3435-3445

Kjellström S., and Jensen O.N. Phosphoric acid as a matrix additive for MALDI MS analysis of phosphopeptides and phosphoproteins. *Analytical Chemistry* **2004**, 76, 5109

Kleinijenhuis A.J., Kjeldsen F., Kallipolitis B., Haselmann K.F., Jensen O.N. Analysis of Histidine Phosphorylation Using Tandem MS and Ion-Electron Reactions. *Analytical Chemistry* **2007**, *79* (19), 7450-7456

Kochin V., Imanishi S.Y., Eriksson J.E. Fast track to a phosphoprotein sketch – MALDI TOF characterization of TLC-based tryptic phosphopeptide maps at femtomolar detection sensitivity. *Proteomics* **2006**, 6, 5676–5682

Kowalewska K., Stefanowicz P., Ruman T., Frąnczyk T., Rode W., Szewczuk Z. Electron Capture Dissociation Mass Spectrometric Analysis of Lysine-Phosphorylated Peptides. *Bioscience Reports* **2010**, published ahead of print 09 Feb 2010

Krüger E. and Hecker M. The first gene of the *Bacillus subtilis clpC* operon, ctsR, encodes a negative regulator of its own operon and other class III heat-shock genes. *Journal of Bacteriology* **1998**, 180, 6681-6688

Krüger E., Zühlke D., Witt E., Ludwig H., Hecker M. ClpC-mediated proteolysis in *Gram*-positive bacteria is autoregulated by the stability of a repressor. *The EMBO Journal* **2001**, 20, 852-863

Küster B. and Mann M. Identifying proteins and post-translational modifications by mass spectrometry. *Current Opinion in Structural Biology* **1998**, 8 (3), 393-400

Kumon A., Kodama H., Kondo M., Yokoi F., Hiraishi H. N^{ω}-Phophoarginine Phosphatase (17 kDa) and Alkaline Phosphatase as Protein Arginine Phosphatases. *Journal of Biochemistry* **1996**, 119, 719-724

Kweon H.K., Hakansson K. Selective zirconium dioxide-based enrichment of phosphorylated peptides for mass spectrometric analysis. *Analytical Chemistry* **2006**, 78, 1743–1749

Larsen M.R., Thingholm T.E., Jensen O.N., Roepstorff P., Jorgensen T.J. Highly selective enrichment of phosphorylated peptides from peptide mixtures using titanium dioxide microcolumns. *Molecular and Cellular Proteomics* **2005**, 4, 873-886

Laub M.T. and Goulian M. Specificity in Two-Component Signal Transduction Pathways. *Annual Reviews in Genetics* **2007**, 41, 121-145

Laugensen S., Roepsdorff P. Combination of two matrices results in improved performance of MALDI MS for peptide mass mapping and protein analysis. *Journal of the American Society for Mass Spectrometry* **2003**, 14, 992–1002

Levy-Favatier F., Delpech M., Kruh J. Characterization of an arginine-specific protein kinase tightly bound to rat liver. *European Journal of Biochemistry* **1987**, 166 (3), 617-621

Linding R., Jensen L.J., Pasculescu A., Olhovsky M., Colwil K., Bork P., Yaffe M.B., Pawson T. NetworKIN: a resource for exploring cellular phosphorylation networks. *Nucleic Acids Research*, **2008**, 36 (Database Issue),D 695-699

Lowery D.M., Mohammad D.H., Elia A.E.H., Yaffe M.B. The Polo-Box Domain – A Molecular Integrator of Mitotic Kinase Cascades and Polo-like Kinase Function. *Cell Cycle* **2004**, 3 (2), 128-131

Marlovits T.C., Kubori T., Sukhan A., Thomas D.R., Galán J.E., Unger V.M. Structural Insights Into the Assembly of the Type III Secretion Needle Complex. *Science* **2004**, 306, 1040-1042

Mazanek M., Mitulović G., Herzog F., Stingl C., Hutchins J.R., Peters J.M., Mechtler K. Titanium dioxide as a chemo-affinity solid phase in offline phosphopeptide chromatography prior to HPLC-MS/MS analysis. *Nature Protocols* **2007**, 2, 1059–1069

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. *Molecular and Cellular Proteomics* **2007**, 7, 299-307

Medzihradszky K.F., Phillipps N.J., Senderowicz L., Wang P., and Turck C.W. Synthesis and characterization of histidine-phosphorylated peptides. *Protein Science* **1997**, 6, 1405-1411

Miller M.L., Soufi B., Jers C., Blom N., Macek B., Mijakovic I. NetPhosBac – A predictor for Ser/Thr phosphorylation sites iin bacterial proteins. *Proteomics* **2009**, 9, 116-125

Mitrophanov A.Y. and Groisman E.A. Signal integration in bacterial two-component regulatory systems. *Genes & Development* **2008**, 22 (19), 2601-2611

Mizuno T. Two-Component Phosphorelay Signal Transduction Systems in Plants: from Hormone Response to Circardian Rythms. *Bioscience, Biotechnology, and Biochemistry* **2005**, 69, 2236-2276

Müller M.Q., de Koning L.J., Schmidt A., Ihling C., Syha Y., Rau. O., Mechtler K., Schubert-Zsilavecz M., Sinz A. An Innovative Method to Study Target Protein-Drug Interactions by Mass Spectrometry. *Journal of Medical Chemistry* **2009**, 52, 2875-2879

Nabetani T., Miyazaki K., Tabuse Y., Tsugita A. Analysis of acidic peptides with a matrix-assisted laser desorption/ionization mass spectrometry using positive and negative ion modes with additive monoammonium phosphate. *Proteomics* **2006**, 6, 4456–4465

Nakashima H., Toyoshima-Morimoto F., Taniguchi E., Nishida E. Identification of a Consensus Motif for Plk (Polo-like Kinase) Phosphorylation Reveals Myt1 as a PLK1 Substrate. *The Journal of Biological Chemistry* **2003**, 278, 25277-25280

Nigg E.A. Mitotic Kinases as Regulators of Cell Division and its Checkpoints. *Nature Reviews* **2001**, 2, 21-32

Nightingale K.P., O'Neill L.P., Turner B.M. Histone modifications: signalling receptors an potential elements of a heritable epigenetic code. *Current Opinions in Genetic and Development* **2006**, 16 (2), 125-136

Olsen J.V., Macek B., Lange O., Makarov A., Horning S., Mann M. Higher energy C-trap dissociation for peptide modification analysis. *Nature Methods* **2007**, 4, 709–712

Ong S.E., Blagoev B., Kratchmarova I., Kristensen D.B., Steen H., Pandey M., Mann M. Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Molecular and Cellular Proteomics* **2002**, 1, 376-386

Padwardhan P. and Miller W.T. Processive Phosphorylation: Mechanisms and Biological Importance. *Cell Signal* **2007**, 19 (11), 2218-2226

Palumbo A.M. and Reid G.E. Evaluation of Gas-Phase Rearrangement and Competing Fragmentation Reactions on Protein Phosphorylation Site Assignment Using Collision Induced Dissociation-MS/MS and MS3. *Analytical Chemistry* **2009**, 90, 9735-9747

Parkinson J.S. and Kofoid E.C. Communication Modules in Bacterial Signalling Proteins. *Annual Reviews Genetics* **1992**, 26, 71-112

Pinkse M.W., Uitto P.M., Hilhorst M.J., Ooms B., Heck A.J. Selective isolation at the femtomole level of phosphopeptides from proteolytic digests using 2D-NanoLC-ESI-MS/MS and titanium oxide precolumns. *Analytical Chemistry* **2004**, 76, 3933–3943

Pinske M.W., Mohammed S., Gouw J.W., van Breukelen B., Vos H.R., Heck A.J. Highly Robust, Automated, and Sensitive online TiO2-based Phosphoproteomics Applied to Study Endogenous Phosphorylation in *Drosophila melanogaster*. *Journal of Proteome Research* **2008**, 7, 687-97

Plutzky J. The Potential Role of Peroxisome Proliferator-Activated Receptors on Inflammation in Type 2 Diabetes Mellitus and Atherosclerosis. *American Journal of Cardiology* **2003**, 92, 34J-41J

Qi C., Zhu Y., Reddy J.K. Peroxisome Proliferator-Activated Receptors, Coactivators and Downstream Targets. *Cell Biochemistry and Biophysics* **2000**, 32, 187-204

Raggiaschi R., Lorenzetto C., Diodato E., Caricasole A., Gotta S., Terstappen G.C. Detection of phosphorylation patterns in rat cortical neurons by combining phosphatase treatment and DIGE technology. *Proteomics* **2006**, 6, 748–756

Reid G.E., Wu J., Chrisman P.A., Wells J.M., McLuckey S.A. Charge-State-Dependent Sequence Analysis of Protonated Ubiquitin Ions via Ion Trap Tandem Mass Spectrometry. *Analytical Chemistry* **2001**, 73 (14), 3274-3281

Rinner O., Seebacjer J., Walzthoeni T., Müller L.N., Beck M., Schmidt A., Müller M., Aebersold R. Identification of Cross-Linked Peptides from Large Sequence Databases. *Nature Methods* **2008**, 5 (4), 315-318

Roepstorff P. and Fohlman J. Proposal for a common nomenclature for sequence ions in mass spectra of peptides. *Biomedical Mass Spectrometry* **1984**, 11, 601

Ross A.R. Identification of Histidine Phosphorylation in Proteins Using Mass Spectrometry and Affinity-Based Techniques. *Methods in Enzymology* **2007**, 423, 549-72

Ross P.L., Huang Y.N., Marchese J.N., Williamson B., Parker K., Hattan S., Khainovski N., Pillai S., Dey S., Daniels S., Pukayastha S., Juhasz P., Martin S., Bartlet-Jones M., He F., Jacobson A., Pappin D.J. Multiplexed Protein Quantitation in *Saccharomyces cerevisiae* Using Amine-reactive Isobaric Tagging Reagents. *Molecular and Cellular Proteomics* **2004**, 3, 1154-1169

Rostom A.A. and Robinson C.V. Disassembly of intact multiprotein complexes in the gas phase. *Current Opinion in Structural Biology* **1999**, 9 (1), 135-141

Ruben E.A., Chapman M.S., Evanseck J.D., Generalized anomeric interpretation of the "high energy" N-P bond in *N*-methyl-*N'*-phosphorylguanidine: importance of reinforcing stereoelectronic effects in "high energy" phosphoester bonds. *Journal of the American Chemical Society* **2005**, 127, 17789-17798

Ruotolo B.T. and Robinson C.V. Aspects of native proteins are retained in vacuum. *Current Opinion in Chemical Biology* **2006**, 10 (5), 402-408

Rush J., Moritz A., Lee K.A., Guo A., Goss V.L., Spek E.J. Zhang H., Zha X.M., Polakiewicz R.D., Comb M.J. Immunoaffinity profiling of tyrosine phosphorylation in cancer cells. *Nature Biotechnology* **2005**, 23 (1), 94-101

Sanders S.L. and Schekman R. Polypeptide Translocation across the Endoplasmatic Reticulum Membrane. *The Journal of Biological Chemistry* **1992**, 267 (20), 13791-13794

Sarma K. and Reinberg D. Histone variants meet their match. *Nature Reviews Mollecular Cell Biology* **2005**, 6 (2), 139-149

Schley C., Altmeyer M.O., Swart R., Müller R., Huber C.G. Proteome analysis of *Myxococcus xanthus* by off-line two-dimensional chromatographic separation using monolithic poly-(styrene-divinylbenzene) columns combined with ion-trap tandem mass spectrometry. *Journal of Proteome Research* **2006**, 5, 2760–2768

Schlosser A., Pipkorn R., Bossemeyer D., Lehmann W.D. Analysis of Protein Phosphorylation of Elastase Digestion and Neutral Loss Tandem Mass Spectrometry. *Analytical Chemistry* **2001**, 73, 170-176

Schmidt A., Kellermann J., Lottspeich F. A novel strategy for quantitative proteomics using isotope-coded protein labels. *Proteomics* **2005**, 5 (1), 4-15

Shaw P.E. Peptidyl-prolyl isomerases: a new twist to transcription. *EMBO Reports* **2002**, 3, 521-526

Sickmann A. and Meyer H.E. Phosphoamino acid analysis. *Proteomics* **2001**, 1, 200-206

Sinz A. Chemical Cross-linking and Mass Spectrometry for Mapping Three-Dimensional Structures and Protein-Protein Interactions. *Journal of Mass Spectrometry* **2003**, 38, 1225-1237

Slater E.C. and Kemp A. Rate Labelling of Mitochondrial Phosphohistidine by Radioactive Inorganic Phosphate. *Nature* **1964**, 204, 1268-1271

Sleno L. and Vollmer D.A. Ion activation methods for tandem mass spectrometry. *Journal of Mass Spectrometry* **2004**, 39, 1091-1112

Sørensen E.S., Højrup P., Petersen T.E., Post-translational modifications of bovine osteopontin: Identification of twenty-eight phosphorylation and three O-glycosylation sites. *Protein Science* **1995**, 4, 2040–2049

Steeg P.S., Palmieri D., Ouatas T., Salerno M. Histidine Kinases and Histidine Phosphorylated Proteins in Mammalian Cell Biology, Signal Transduction, and Cancer. *Cancer Letters* **2003**, 190, 1-12

Steen H., Küster B., Fernandez M., Pandey A., and Mann M., Detection of Tyrosine Phosphorylated Peptides by Precursor Ion Scanning Quadrupole TOF Mass Spectrometry in Positive Ion Mode. *Analytical Chemistry* **2001**, 73, 1440-1448

Steen H., Jebanathirajah J. A., Rush J., Morrice N., Kirschner M. W., Phosphorylation analysis by mass spectrometry: Myths, facts, and the consequences for qualitative and quantitative measurements. *Molecular and Cellular Proteomics* **2006**, 5, 172–181

Stensballe A., Jensen O.N., Olsen J.V., Haselmann K.F., Zubarev R.A., Electron capture dissociation of singly and multiply phosphorylated peptides. *Rapid Communications in Mass Spectrometry* **2000**, 14, 1793–1800

Syka J.E., Coon J.J., Schroeder M.J., Shabanowitz J., Hunt D.F. Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proceedings of the National Academy of Sciences* **2004**, 101, 9528–9533

Teis D., Wunderlich W., Huber L.A. Localization of the MP1-MAPK Scaffold Complex to Endosomes Is Mediated by p14 and Required for Signal Transduction. *Developmental Cell* **2002**, 3, 803-814

Theodosiou A. and Ashworth A. MAP kinase phosphatases. *Genome Biology* **2002**, 3 (7), reviews3009.1-10

Thingholm T.E., Jensen O.N., Robinson P.J., Larsen M.R., SIMAC – A phosphoproteomic strategy for the rapid separation of mono-phosphorylated from multiply phosphorylated peptides. *Molecular and Cellular Proteomics* **2008**, 7, 661–671

Thingholm T.E., Jensen O.N., Larsen M.R. Analytical strategies for phosphoproteomics *Proteomics* **2009**, 9, 1451-1468

Thomas D., Guthridge M., Woodcock J., Lopez A. 14-3-3 Protein Signalling in Development and Growth Factor Responses. *Current Topics in Developmental Biology* **2005**, 67, 285-303

Thomason P. and Kay R. Eucaryotic Signal Transduction via Histidine Aspartate Phosphorelay. *Journal of Cell Science* **2000**, 113, 3141-3150

Trinidad J.C., Specht C.G., Thalhammer A., Schoepfer R., Burlingame A.L. Comprehansive Identification of Phosphorylation Sites in Postsynaptic Density Preparations. *Molecular and Cellular Proteomics* **2006**, 5, 914-922

Villen J., Beausoleil S. A., Gerber S. A., Gygi S. P. Large-scale phosphorylation analysis of mouse liver. *Proceedings of the National Academy of Sciences* **2007**, 104, 1488–1493

Villen J. and Gygi S.P. The SCX/IMAC enrichment approach for global phosphorylation analysis by mass spectrometry. *Nature Protocols* **2008**, 3, 1630-1638

Voet D., Voet J.G., Pratt C.W., **dt. Übersetzung Beck-Sickinger A.G., Hahn U.** Biochemie, ISBN 3-527-30519-X, *Wiley-VCH Verlag GmbH & Co KGaA*, Weinheim **2002**

Wakim B.T. and Aswad G.D. Ca2+-calmodulin-dependent Phosphorylation of Arginine i Histone 3 by a Nuclear Kinase from Mouse Leukemia Cells. *The Journal of Biological Chemistry* **1994**, 269 (4), 2722-2727

Wakim B.T., Grutoski P.S., Vaughan A.T.M., Engelmann G.L. Stimulation of a Ca2+-calmodulin-activated Histone 3 Arginine Kinase in Quiescent Rat Heart Endothelial Cells Compared to Actively Dividing Cells. *The Journal of Biological Chemistry* **1995**, 270 (39), 23155-23158

Wålinder O. Protein-bound acid labile phosphate. Isolation of 1-32P-phosphohistidine and 3-32P-phosphohistidine from some mammalian and microbial cell extracts incubated with adenosine triphosphate-32P. *The Journal of Biological Chemistry* **1969**, 244, 1065-69

Wålinder O., Zetterqvist O., Engström L. Intermediary phosphorylation of bovine liver nucleoside diphosphate kinase. Studies with rapid mixing techniques. *The Journal of Biological Chemistry* **1969**, 244, 1060-1064

Wei Y.-F. and Matthews H.R. Identification of Phosphohistidine in Proteins and Purification of Protein-Histidine-Kinases. *Methods in Enzymology* **1991**, 200, 388-414

Weller M. Protein-Bound Histidine, As Well As Protein-Bound Serine, Residues Are Sites Of Phosphorylation In The Synaptic Plasma Membrane. *Biochimica et Biophysica Acta* **1978**, 509, 491-498

Weston C.R. and Davis R.J. The JNK signal transduction pathway. *Current Opinion in Genetics and Development* **2002**, 12, 14-21

Xu C.F., Wang H., Li D., Kong X.P., Neubert T.A. Selective enrichment and fractionation of phosphopeptides from peptide mixtures by isoelectric focusing after methyl esterification. *Analytical Chemistry* **2007**, 79, 2007–2014

Yoo C., Patwa T.H., Kreunin P., Miller F.R., Huber C.G., Nesvizhskii A.I., Lubman D.M. Comprehensive analysis of proteins of pH fractionated samples using monolithic LC/MS/MS, intact MW measurement and MALDI-QIT-TOF MS. *Journal of Mass Spectrometry* **2007**, 42, 312–334

Zhang X., Ye J., Jensen O.N., Roepstorff P. Highly Efficient Phosphopeptide Enrichment by Calcium Phosphate Precipitation Combined with Subsequent IMAC Enrichment. *Molecular and Cellular Proteomics* **2007**, 6, 2032-2042

Zhou D. and Galán J. *Salmonella* entry into host cells: the work in concert of type III secreted effector proteins. *Microbes and Infection* **2001**, 3, 1293-1298

Zeng L. and Zhou M.-M. Bromodomain: an acetyl-lysine binding domain. *FEBS Letters* **2002**, 513, 124-128

Zu X.L., Besant P.G., Imhof A., and Attwood P.V. Mass spectrometric analysis of protein histidine phosphorylation. *Amino Acids* **2007**, 32, 347-357

Figures

Part I

A_01 Phosphoamino acids 002
A_02 Signalling cascades in cells 003
A_03 Kinase Activity and Phosphatase Activity 004
A_04 Proteomics workflows for Protein Identification 009
A_05 Enrichment Techniques for Phosphopeptides 012
A_06 Schematic Overview of MS/MS Methods 016
A_07 Properties of Peptides with Multiple Phosphorylations 021
A_08 Optimization of MALDI Detection of Phosphopeptides 026
A_09 MS/MS Spectrum of the Singly Phosphorylated Peptide (aa 33–48) from Tryptically Digested β-casein 029
A_10 Enrichment Strategy on TiO_2 030
A_11 Results TiO2 optimization 032
A_12 Improved Phosphopeptide Separation in RP-Chromatography 035
A_13 Improved Identification of Phosphopeptides from Osteopontin with MALDI-MS/MS 038
A_14 Comparison of MALDI-MS/MS analysis with ETD fragmentation of Phosphopeptides 042

Part II

B_01 Nitrogen-Bound Phosphorylations in Living Cells 045
B_02 Receptor Histidine Kinases in Bacteria and Plants 047
B_03 Delocalization of pArg and pHis 050
B_04 Hydrolysis Mechanism 051
B_05 Structure of the CtsR Homodimer in Complex with DNA 055
B_06 Intact Protein MS and Top Down Analysis 056
B_07 Identification of Phosphorylation at R62 of CtsR 060
B_08 Localization of Phosphorylated Arginine Residues 061
B_09 McsB Catalyzed Phosphorylation of Peptides 063
B_10 Arginine-Specificity of McsB demontrated MS and ^{31}P-NMR 066
B_11 Stability of Phospho-Arginine 070
B_12 Tryptic Activity at Arginine-Phosphorylation 072
B_13 TiO_2-based Enrichment of Peptides with Acid-Labile Phosphorylations 075
B_14 Comparison of MS/MS Methods for the Identification of Arginine-Phosphorylated Peptides 078
B_15 Identified Phosphorylation Sites of CtsR and McsB 079
B_16 MS/MS Analysis of Peptides with Phospho-Arg and Phospho-His 085
B_17 Neutral Loss from two Different Positions Revealed by ESI-MS^3 087
B_18 H_3PO_4-Eliminiation is the Lowest Energy Neutral Loss Pathway 089
B_19 Different Fragmentation of Peptides without free OH-Moieties 091
B_20 Acidic Amino Acids Partially Substitute the Function of the C-terminal COOH-Group 094-095
B_21 CID-Induced Translocation of the Phospho-Moiety Onto Glutamate and Serine 097
B_22 Analysis of N-Phosphorylations in Living Cells – Strategy 103

Appendix Cross-Linking

App_01 Strategy for Structural Analysis of Protein Complexes by Chemical Cross-Linking and Mass Spectrometry 106
App_02 Cross-Linking Products and MS/MS Fragmentation 107
App_03 MS/MS spectrum of the identified cross-link NFNMNKVK-KT from PPAR2α 109
App_04 The Core Complex of the Samonella Type III Secretion System 111
App_05 Results for Acetyl-Labelling and Connected Domains Between Different Protein Species in the Needle Complex 114
App_06 Connection of the Same Peptide Strands Identifies the Dimerization Region of Myotilin-IG Domains 116

Tables

Part I

TA_1 Synthetic phosphopeptides for method optimization. 027
TA_2 Removal of sequence-unspecific neutral loss ion signal improves automatic database identification of highly phosphorylated peptides. 037
TA_3 Mascot Phosphosite Determination 039

Part II

TB_1 pka values for pHis, pArg and pLys 049
TB_2 Phosphopeptides of CtsR from B.stearothermophilus 059
TB_3 Synthetic Peptides for Investigation of N-Bound Phosphorylations 064
TB_4 Autophosphorylation sites of McsB 080

Appendix – Cross-Linking

TApp_1 Identified Cross-Linking Products of Myotilin-IG Domains Cross-Linked with EDC/NHS 117

Abbreviations

ACN	acetonitrile
AcOH	acetic acid
ADP	adenosine diphosphate
ATP	adenosine triphosphate
BSA	bovine serum albumin
BS^2G	bis-sulfosuccinimidyl-glutarate
BS^3	bis-sulfosuccinimidyl-suberate
CE	collision energy
CHCA	α-cyano-4-hydroxycinnamic acid
CID	collision-induced dissociation
ClpC, clpC	class three stress response-related ATPase, small gene name
CtsR, ctsR	class three stress response regulator, small gene name
Da	Dalton
DHB	2,5-dihydroxybenzoic acid, gentisic acid
DNA	Desoxyribonucleic acid
DTT	dithiothreitol
DVB	divinylbenzole
ECD	electron-capture dissociation
EDC	1-ethyl-3-(*N*,*N*-dimethylaminopropyl) carbodiimide hydrochloride
EDD	electron-detachment dissociation
EGS	ethylene glycol bis[succinimidylsuccinate]
EM	electron microscopy
ERLIC	electrostatic repulsion, hydrophilic interaction chromatograpy
ESI	electrospray ionization
ETD	electron-transfer dissociation
FA	formic acid
FT	Fourier-transform
FT-ICR	Fourier-transform ion cyclotron resonance
GPMAW	general protein mass analysis for Windows
HCD	higher-energy collisional dissociation
HFBA	Heptafluorobutyric acid
HPLC	high performance liquid chromatography
IG	immunoglobulin
iTRAQ	isobaric tag for relative and absolute quantificantion
LC	liquid chromatography
ICPL	isotope coded protein labeling
IMAC	immobilized metal ion-affinity chromatogrpahy
IMP	institute for molecular pathology
IRMPD	infra-red multiphoton dissociation
LTQ	Linear trap-quadrupole
MALDI	matrix-assisted laser desorption/ionization
MAP	mitogen-activated protein
MAPK	mitogen-activated protein kinase
McsA, mcsA	modulator of class three stress response A, small gene name
McsB, mcsB	modulator of class three stress response B, small gene name
MOAC	metal oxide-affinity chromatography
MS	mass spectrometry
MS/MS	tandem mass spectrometry
MSn	multiplexed tandem mass spectrometry, for instance MS3

m/z	mass to charge
NHS	N-hydroxysuccinimide
NL	neutral loss
NMR	nuclear magnetic resonance
OcSNa	octylsulfonate sodium salt
PAGE	polyacrylamide gel electrophoresis
PPAR	peroxisome proliferator-activated receptors
phospho-, p-	phosphorylated
ppm	parts per million
PS	polystyrole
PSD	post-source decay
PTM	post-translational modification
RNA	Ribonucleic acid
RP	reversed phase
RPC	reversed phase chromatography
SAX	small angle X-ray scattering
SCX	strong cation exchange
SDS	sodium dodecylsulfate
SILAC	stable isotope labelling of amino acids in cell culture
TOF	time of flight
TFA	trifluoroacetic acid
UV	ultra-violet
WAX	weak anion exchange
wt	wild type

Amino acids

amino acid	3 letter code	1 letter code	amino acid	3 letter code	1 letter code	amino acid	3 letter code	1 letter code
alanine	Ala	A	cysteine	Cys	C	aspartate	Asp	D
glutamate	Glu	E	phenyl-alanine	Phe	F	glycine	Gly	G
histidine	His	H	isoleucine	Ile	I	lysine	Lys	K
leucine	Leu	L	methionine	Met	M	asparagine	Asp	N
proline	Pro	P	glutamine	Gln	Q	arginine	Arg	R
serine	Ser	S	threonine	Thr	T	valine	Val	V
tryptophane	Trp	W	tyrosine	Tyr	Y			

List of Publications

First author publications

1: Proteomics. 2008 Oct 29;8(21):4577-4592.
Enhanced detection and identification of multiply phosphorylated peptides using TiO(2) enrichment in combination with MALDI TOF/TOF MS.
Schmidt A., Csaszar E., Ammerer G., and Mechtler K.

2: Science. 2009 Jun 5;324(5932):1323-7.
McsB is a protein arginine kinase that phosphorylates and inhibits the heat-shock regulator CtsR.
Fuhrmann J., **Schmidt A.**, Spiess S., Lehner A., Turgay K., Mechtler K., Charpentier E., and Clausen T.

3: J Am Soc Mass Spectrom. 2006 Aug;17(8):1100-13. Epub 2006 Jun 5.
Isotope-Labeled Cross-Linkers and Fourier Transform Ion Cyclotron Resonance Mass Spectrometry for Structural Analysis of a Protein/Peptide Complex
Ihling C, **Schmidt A.**, Kalkhof S, Schulz D.M., Stingl C., Mechtler K., Haack M., Beck-Sickinger A.G., Cooper D.M.F., and Sinz A.

Contributions

1: BMC Bioinformatics. 2007 Jun 13;8:197.
MASPECTRAS: a platform for management and analysis of proteomics LC-MS/MS data
Hartler J., Thallinger G.G., Stocker G., Sturn A., Burkard T.R., Körner E., Rader R., **Schmidt A.**, Mechtler K., and Trajanoski Z.

2: Mol Cell. 2007 Apr 13;26(1):103-15.
Heterochromatin Formation in Drosophila Is Initiated through Active Removal of H3K4 Methylation by the LSD1 Homolog SU(VAR)3-3
Rudolph T., Yonezawa M., Lein S., Heidrich K., Kubicek S., Schäfer C., Phalke S., Walther M., **Schmidt A.**, Jenuwein T., and Reuter G.

3: J Biol Chem. 2009 Mar 27;284(13):8395-405. Epub 2009 Jan 13.
Dynamic histone H1 isotype 4 methylation and demethylation by histone lysine methyltransferase G9a/KMT1C and the Jumonji domain containing JMJD2/KDM4 proteins
Trojer P., Zhang J., Yonezawa M., Zheng H., **Schmidt A.**, Perez-Burgos L., Lobel P., Jenuwein T., Reinberg D.

4: PLoS Pathogens, accepted 2010 Feb 01, published April 2010
Topology and organization of the Salmonella typhimurium type III secretion needle complex components
Schraidt O., Lefebre M., Brunner M, Schmied W.H., **Schmidt A.**, Radics J., Mechler K, Galan J.E., and Marlovits T.

5: Proteins. 2007 Nov 1;69(2):254-69.
Annexin A2/P11 interaction: New insights into annexin A2 tetramer structure by chemical crosslinking, high-resolution mass spectrometry, and computational modeling
Schulz D.M., Kalkhof S., **Schmidt A.**, Ihling C., Stingl C., Mechtler K., Zschörnig O., and Sinz A.

6: J Med Chem. 2009 May 14;52(9):2875-9.
An innovative method to study target protein-drug interactions by mass spectrometry.
Müller M.Q., de Koning L.J., **Schmidt A.**, Ihling C., Syha Y., Rau O., Mechtler K., Schubert-Zsilavecz M., and Sinz A.

Further Publication

Mapping protein interfaces by chemical cross-linking and Fourier transform ion cyclotron resonance mass spectrometry: application to a calmodulin/adenylyl cylase 8 peptide complex
Schmidt A., Kalkhof S., Ihling C., Cooper D.M.F., and Sinz A.

Planned Publications 2010
Detection of N-bound phosphorylations in bacteria after TiO2 enrichment
Fragmentation of peptides with Arginine Phosphorylation

Presentations at Conferences
Posters
IMSC 2006, Prague
Multidimensional Chromatography and Quantitative Mass Spectrometry Based Approach for Identification of Differentially Expressed Proteins in Mice Livers
Schmidt A., Pichler P., Mitulovic G., Hasselblatt P., Stingl C., Ammerer G., Wagner E.F., and Mechtler K.
ASMS 2008, Denver
MALDI TOF/TOF Is a Convenient Tool for the Identification and Quantification of Multiply Phosphorylated Peptides from Low Complexity Samples
Schmidt A., Csaszar E., Mitulovic G., Ammerer G., and Mechtler K.
ASMS 2009
Arginine-Phosphorylation as Post-Translational Modification of Proteins
Schmidt A., Fuhrmann J., Stingl C., Brecker L., Clausen T., and Mechtler K.

Talks
Phospho-Proteomics Seminar – 13th and 14th October 2008, IMP Vienna
Analysis of multiphosphorylated peptides with Maldi TOF/TOF
(Andreas Schmidt)

Curriculum Vitae

Personal Details

Last Name:	Schmidt, birth name Schreiter
First Name:	Andreas
Current address:	Laxenburgerstrasse, 59/6, Vienna, A-1100, Austria
Date of Birth:	4th April 1981
Place of Birth:	Annaberg-Buchholz (Federal Republic of Germany) (former German Democratic Republic)
Nationality:	German
Parents:	Annelie Schmidt, birth name Schreiter (29th June 1957), Ralf Schmidt (18th October 1959),
Brother:	Norbert Schmidt (24th March 1984), married

Degrees and Education

1st January 2006 - PhD student at the Christian-Doppler.Laboratory for Proteome Analysis, Vienna
research projects:

A *phosphoproteome analysis of multiple phosphorylated peptides and N-phosphorylations - focus on method developement for reliable detection and quantification (N-phosphorylation analysis* in cooperation with Dr. Tim Clausen, Strucural Biology, IMP Vienna, Austria)

B *cross-linking studies for the identification of protein conformation changes upon small molecule binding and localization and relative structure of protein subunits in large assemblies* (with Prof. Dr. A. Sinz, University of Halle, Dr. Thomas Marlovits, IMBA Vienna, Austria, and Prof. Dr. Kristina Djinović-Carugo, University of Vienna, Vienna, Austria)

C *detection/localization and relative quantification of methylation and demethylation processes by protein histone methylases and demethylases* (in cooperation with Prof. Dr. Thomas Jenuwein, now MPI for Immunology, Freiburg, Germany)

1st August 2005 - 30th December 2005 research assistent at Prof. Dr. Dermot M.F. Cooper's Laboratory at the University of Cambridge (Cambridge, UK)
Introduction to microbiology to study effect of mutations in the adenylyl cyclase 8 calmodulin binding motif on its activation by Ca2+/calmodulin.

31st May 2005 Diploma thesis: *Studying the AC8/calmodulin interaction with chemical cross-linking and high resolution mass spectrometry*
Introduction to chemical cross-linking methods and application to a calmodulin/peptide complex.
Degree: Diploma Chemist (Univ.), equivalent to MSc (Chemistry)
Mark: very good

October 2000 - May 2005 Study of chemistry at the University of Leipzig
July 2002 Pre-Diploma
main study (2003-2005) focus on analytical chemistry, biochemistry and organic chemistry

September 1999 - July 2000 military service (Osterode/Harz)

10th July 1999 Senior High School Graduation

September 1992 - July 1999 school for higher education (Landkreisgymnasium Schlettau)
examination marks: Chemistry, 2.3
Mathematics, 2.9

September1990- August 1992 "polytechnical school" in Cranzahl

September 1987 - August 1990 elementary school in Cranzahl ("Wolfgang Goeftel Oberschule")

RESEARCH ARTICLE

Enhanced detection and identification of multiply phosphorylated peptides using TiO_2 enrichment in combination with MALDI TOF/TOF MS

Andreas Schmidt[1, 2], Edina Csaszar[2], Gustav Ammerer[1, 2] and Karl Mechtler[3, 4]

[1] Christian Doppler Laboratory for Proteome Analysis, University of Vienna, Vienna, Austria
[2] Institute for Biochemistry and Molecular Biology, University of Vienna, Vienna, Austria
[3] Institute for Molecular Biotechnology of the Austrian Academy of Sciences (IMBA), Vienna, Austria
[4] Research Institute of Molecular Pathology (IMP), Vienna, Austria

The analysis of PTMs such as phosphorylation has become an important field in MS because they can directly indicate protein states and interactions. Whereas the characterization of singly and doubly phosphorylated peptides has almost become routine, identifying phosphorylation events at multiple residues within a small region of a protein is still problematic. The identification of multiple modifications can be further hampered by low sequence information due to multiple neutral losses from phosphorylated side chains. Here we present a strategy for the analysis of complex phosphopeptides that combines peptide enrichment by titanium dioxide, separation by RP separation on monolithic columns and MS using high energy HE-CAD in a MALDI TOF/TOF analyser. Using synthetic phosphopeptides our approach is compared to multistage activation (MSA) MS/MS and the recently described electron transfer dissociation (ETD) method using an ESI-LTQ mass spectrometer.

Received: April 7, 2008
Revised: July 23, 2008
Accepted: July 25, 2008

Keywords:
Chromatographic enrichment / Mass spectrometry / Phosphopeptides / Phosphoproteomics / Phosphorylation

1 Introduction

PTMs such as phosphorylation, glycosylation and peptide-conjugation are vital for function of proteins and cellular pathways. Phosphorylation, a reversible modification that practically occurs in all cellular compartments is viewed as a key regulator of protein activity as well as a major mediator of protein–protein interactions. The extent of the modification depends on dynamic interactions of specific protein kinases, of which more than 500 different forms are supposed to be expressed in human cells [1] and a smaller number of protein phosphatases. Misregulation at the step of modification or dephosphorylation is often associated with severe physiological consequences for the organism. In recent years, MS gained a key role in the identification of phosphorylation sites and the cellular pathways and signalling networks behind them. For example, large-scale experiments designed towards identification of phosphopeptides led to a high number of newly described phosphorylation sites on many proteins [2, 3]. This situation triggered increased efforts to test the physiological relevance of these modification using combinations of molecular biology methods, of biochemistry and of bioinformatics [4]. Although there are now numerous cases for which the functional and regulatory role of phosphorylation on selected protein sites has been worked out,

Correspondence: Karl Mechtler, Dr.-Bohr-Gasse 3, Vienna A-1030, Austria
E-mail: mechtler@imp.univie.ac.at
Fax: +13-7987153

Abbreviations: DHB, 2,5-dihydroxybenzoic acid; **ECD,** electron capture dissociation; **ETD,** electron transfer dissociation; **FA,** formic acid; **IAA,** indolylacrylic acid; **MSA,** multistage activation; **PA,** phosphoric acid; **reTOF,** reflector TOF

there is still a vast area of events that have either escaped detection or whose relevance could not be resolved due to the complexity of the system. One of the remaining problems is also the question of how to attain complete sequence coverage by MS for proteins in question. Among the difficulties to resolve this problem area is the fact that proteins are sometimes phosphorylated by the same or even different kinases within small regions. Although considerable progress has been made in the identification of phosphopeptides in general, thereby allowing the confident assignment of single or double phosphorylation events, the analysis of multiple phosphorylated peptides is still a major challenge that has not yet attracted a large amount of attention [5, 6]. After proteolytic digestions of proteins a major problem is the traceability of a complex modification due to either low stoichiometry compared to the unphosphorylated or singly phosphorylated peptides. Yet other factors cause poor identification of these peptides by any mass spectrometric method whereof the ionization efficiency is the most important [7, 8]. Finally, the chromatographic properties of multiply phosphorylated peptides can also have a detrimental effect within routine purification protocols.

Development of enrichment methods has given a big boost to the characterization of phosphorylated peptides, because it solved the problem of signal suppression by unphosphorylated peptides. Enrichment can be achieved either by using the physico-chemical properties of the phosphate moiety, by chemical derivatization or by the affinity to antibodies or metal ions [9–23]. Today's affinity methods are predominantly based on metal ions with immobilized metal affinity chromatography (IMAC) as the most widespread approach [21]. IMAC allows affinity enrichment of phosphopeptides based on the affinity of acidic groups to trivalent metal atoms such as iron or gallium. The early problem that strong acidic peptides also show high affinity for the IMAC resin was solved by esterification of the free carboxy groups found at the C-terminus as well as the side chains of glutamate and aspartate [14]. A second method that has become quite popular throughout the last years is the enrichment of phosphopeptides on the surface of metal oxides such as titanium or zirconium. Theoretical arguments had suggested a specifically strong attraction of the phosphate moiety towards titanium dioxide (TiO_2) compared to other acidic functions like carboxyl groups or other anions such as halogens or acetate [17]. One advantage of this method is that selective enrichment of phosphorylated peptides does not depend on prior esterification procedures and is therefore not affected by side reactions or incomplete esterification. Earlier experiments revealed a significant loss of multiply phosphorylated peptides upon TiO_2 enrichment because of this effect, IMAC is currently favoured for the analysis of multiply phosphorylated peptides [19].

The final separation of phosphopeptide mixtures before MS analysis is routinely accomplished on C18 columns because the applied separation conditions allow easy coupling to ESI-MS. Although chromatographical behaviour of multiply phosphorylated peptides has not been investigated in depths, it has been speculated that their separation on C18 column can have major drawbacks such as a low recovery or broad retention width. Since ESI is the most commonly used ionization technique for phosphopeptides any bias of the C18 chromatographic method would also influence the generation of MS data. ESI mass spectrometers have an advantage as they offer high throughput and a wide range of fragmentation methods that can be successfully applied to phosphopeptides. However, early MS/MS experiments on phosphopeptides also exhibited the inadequacy of thermal or vibrational exitation to generate daughter ions with high amount of sequence information. This inadequacy is due to lower stability of the phosphate moiety in contrast to the bonds in the peptide backbone. This phenomenon of neutral phosphate loss is mainly connected to phosphoserine and phosphothreonine (in case of phosphotyrosine strong neutral loss ions are normally not observed).

One solution was fragmentation of the precursor in so-called MS^n or multistage activation (MSA) methods applied on IT instruments which gave significantly improved results for the identification [24–26]. During the MS/MS steps the precursor is dephosphorylated and in the final step fragmented, which often results in higher yields of backbone fragment ions. A recently described method fragments phosphorylated peptides at elevated energy levels in the so-called C-trap of linear IT-Orbitrap hybrid instruments which is believed to improve fragmentation of the peptide backbone in MS/MS analysis in comparison to low energy collisional fragmentation in normal ion trapping instruments [27].

A different fragmentation mechanism, electron capture dissociation (ECD), was introduced especially for the enhanced analysis of post-translationally modified peptides. This mechanism relies on the uptake of a single low energy electron followed by dissociation. Chemical bonds are cleaved prevailingly in the peptide backbone at the $N–C_\alpha$ bond as a result from a radical reaction cascade [28]. ECD provides MS/MS analysis of phosphopeptides without excessive loss of phosphoric acid (PA) from the precursor, so that the formed daughter ions still carry the intact phosphate moiety which benefits the assignment of the phosphosite within the amino acid sequence [29]. Since this technique is only available on Fourier transform ICR (FT-ICR) instruments excellent mass accuracy and resolution are achieved also for MS/MS data. While ECD works nicely for direct MS, the coupling to HPLC is still difficult.

Quite recently electron transfer dissociation (ETD) was described as a method where electron uptake and fragmentation takes place in the IT [30, 31]. Thereby electrons are delivered in the form of aromatic anions, while the positively charged peptides are trapped, during interaction of precursor ions with the anions electrons can be transferred leading to neutralization of the anion and a single charge of the peptide ion. Both methods have the disadvantage that fragmentation is only effective with multiply charged ions as precursors, typically three- or four-fold charged ions give the best results [28].

MALDI is rarely used for phosphopeptide analysis [32–35]. The major disadvantage is the immense signal suppression of phosphorylated peptides during ionization, when there are also unphosphorylated peptides present in the crystallized peptide mixture [35, 36]. A higher number of phosphate moieties in the peptide sequence significantly increases signal supression. Literature describes the invertion of this effect in negative ionization mode, which could be used to identify phosphopeptides from difficult mixtures, *e.g.* after in gel digest by comparing the signal intensities from both experiments [34].

MALDI analysis before and after phosphatase treatment provides another possibility to identify phosphorylated peptides, because they can be easily identified by the mass shift to the unmodified counterpart in the phosphatase treated sample. This method proved to be efficient with MALDI instruments, because only singly charged peptides have to be taken into account [37, 38]. Unfortunately both methods cannot answer the question of the localization of the phosphate moiety within the observed peptide unless only one single putatively phosphorylated amino acid is present.

In the case of serine or threonine phosphorylation, PSD fragments which belong to ions with a neutral loss of H_3PO_4 are observed in reflector TOF (reTOF) spectra as unresolved peak 98 Da below of the expected peptide mass. These signals are normally observed at low intensity levels but they can be easily detected because of their characteristic unresolved structure. In the case of multiphosphorylated peptides also a second peak may occur at $[M + H]^+$ minus 196 Da. The massive signal suppression of phosphopeptide signals in the presence of unphosphorylated peptides is the strongest argument against MALDI ionization for the direct analysis of protein phosphorylations. This is due to complex formation of the phosphate moiety on the metal surface of the MALDI target which is often stainless steel and its low proton affinity and therefore applied negative charge. Another handicap is the predominant formation of singly charged ions, which cover a quite broad mass range which is unsuitable for trapping instruments. The quality of MS/MS spectra upon CID fragmentation is also lower for singly charged peptides compared to multiply charged which are formed in ESI.

To obtain structural information of phosphorylated peptides, PSD fragmentation analysis is not often the method of choice, because of its low fragmentation efficiency. Although signals which are related to the neutral loss of the phosphate moiety dominate the resulting MS/MS spectra, many backbone fragments of low intensity are observed [32].

Today MALDI ionization is commonly used with a TOF/TOF mass analyser which provides MS/MS using either PSD alone or in combination with collisional activation of the ions in a short CID-tube. PSD is suitable to generate high mass daughter ions, while activation in the collision cell (up to 3×10^{-6} bar) mainly promotes the formation of low mass ions such as ammonium ions or small internal fragments. This approach is quite powerful, because it is applicable over a broad mass range in MS from 800 to over 4000 amu and a high mass accuracy can be achieved in tandem mass spectra.

Our study presents an alternative method for identification of multiply phosphorylated peptides which initially become purified on TiO_2 particles. Following the enrichment we introduce a fast monolithic chromatography which provides high recovery and effective separation and MALDI TOF/TOF MS/MS analysis. Using this approach we cannot only identify multiply phosphorylated proteins and their phosphosites but in combination with the iTRAQ chemistry we are also able to obtain quantitative data [39, 40].

2 Materials and methods

2.1 Materials

All organic solvents and ammonia were from VWR-Merck (Vienna, Austria) in chromatographic grade. Trypsin was purchased from Promega (Mannheim, Germany) and all chemicals for the tryptic digest, DTT, ammonium hydrogencarbonate (NH_4HCO_3), ammonium dihydrogenphosphate ($NH_4H_2PO_4$) and iodoacetamide, were from Sigma (Vienna, Austria) at highest available purity. MALDI matrix CHCA was purchased in 98% purity (Sigma) and recrystallized following the protocol from Bruker (Bremen, Germany). 2,5-Dihydroxybenzoic acid (DHB, 98% purity; Sigma–Aldrich, Vienna, Austria) for TiO_2 enrichment and octylsulphonate sodium salt (OcSNa, highest purity) were also from Sigma. MALDI matrix DHB was from Bruker. PA and heptafluorobutyric acid (HBFA) were from Fluka (Sigma–Aldrich), glacial acetic acid (AcOH) was from J.T. Baker (Griesheim, Germany), and TFA from Pierce (THP, Vienna, Austria). Osteopontin from bovine milk without BSA matrix was from R&D Systems (Minneapolis, USA) and β-casein from Sigma. Peptides were in-house synthesized using the FMOC strategy on a peptides synthesizer ABI 433A (Applied Biosystems, Darmstadt, Germany).

2.2 Enzymatic protein cleavage

A 100 μg/mL solution of β-casein was prepared with freshly prepared 50 mM NH_4HCO_3 solution. A 100 μL aliquot was used for tryptic cleavage. Osteopontin was dissolved in freshly prepared 50 mM NH_4HCO_3 in H_2O to a final protein concentration of 1 mg/mL and sonicated for 2 min. A pH of 7.7 was applied and aliquots of 10 μg were prepared and kept at −80°C. Prior to trypsination osteopontin was mixed with 50 mM NH_4HCO_3 solution to a final volume of 50 μL. Two microlitres of DTT (6.5 mM, in 50 mM NH_4HCO_3) were added to aliquots of β-casein and osteopontin and the mixtures were incubated at 56°C for 35 min. After cooling to room temperature 2 μL of iodoacetamide solution (27 mM, in 50 mM NH_4HCO_3) were applied and the sample was kept in the dark for 35 min to ensure a quantitative blocking of cysteines. Excess of iodoacetamide was removed afterwards

by adding 10 μL of the previously mentioned DTT solution. Trypsin was dissolved in 50 mM acetic acid and diluted to a final concentration of 500 ng/μL and kept at −80°C until use. An aliquot of 1 μg was diluted with 50 mM NH_4HCO_3 to a concentration of 50 ng/μL. Trypsin solution (4 μL) was added to each aliquot and after 4 h at 37°C additional 4 μL of the trypsin solution was added. The cleavage was carried out for 14 h at 37°C and the reaction was stopped after 14 h by addition of 10% TFA to a final concentration of 0.5%.

2.3 TiO_2 enrichment of phosphorylated peptides

TiO_2 packed into small tips (Glygen, Columbia, USA) was washed with 40 μL of ACN, 60 μL of 0.2% HFBA in 70% ACN and 100 μL of diluted loading buffer (200 mg/mL DHB, 33.3 mM OcSNa in 20% AcOH, 45% ACN, 0.1% HFBA). Tryptic digests from osteopontin (4.3 μM) and β-Casein (5.8 μM) were diluted to 2 μM solutions using 0.1% TFA and further dilution to 0.2 μM peptide concentration was accomplished by adding 50% ACN with 10% AcOH. Twenty microliters were mixed with 40 μL concentrated loading buffer (300 mg/mL DHB (Sigma), 50 mM OcSNa in 40% AcOH, 40% ACN, 0.2% HBFA) to a final DHB concentration of 200 mg/mL. Aliquots according to 1 and 3 pmol of each protein respectively were loaded onto four TiO_2 columns over a time span of 30 min. In order to remove unphosphorylated peptides, 60 μL of diluted loading buffer and 100 μL of 70% ACN (with 0.2 HFBA) were used as wash solvents. The complete liquid was removed from the TiO_2 tips before the elution buffer was applied. $NH_4H_2PO_4$ (50 mM) was prepared with H_2O and the pH was set to 11 with concentrated ammonia solution (25%). Over 20 min phosphopeptides were eluted with 25 μL of the $NH_4H_2PO_4$ solution in steps of 5 μL. The resulting sample was acidified with 5 μL of 5% H_3PO_4 to give a pH of 2–3. Enrichment efficiency was tested by MALDI reTOF MS spectra of treated and untreated sample.

2.4 HPLC separation coupled to MALDI detection

Separation of phosphorylated peptides was carried out using an Ultimate Plus HPLC system (Dionex, Germany) which was equipped with FAMOS autosampler, Switchos column switching device, UV detector (detection at 214 nm) and Probot fractionation device (all Dionex). Comparison of separation efficiencies of C18 and monolithic columns was carried out using PA in all HPLC solvents. In case of separation with C18 column, 0.1% H_3PO_4 was used as loading solvent and transport liquid, solvent A was 5% ACN with 0.1% H_3PO_4 and solvent B 80% ACN with 0.08% H_3PO_4 (flow rate 0.280 μL/min). The C18 column (PepMap C18, 30 cm × 75 μm × 3 μm, 100 Å, Dionex) was operated at elevated temperature of 40°C in combination with a trap column (PepMap C18, 5 mm × 300 μm × 5 μm, 100 Å, Dionex) to remove salts. Phosphopeptide separation on C18 material was performed after 10 min washing with a linear gradient from 0 to 60% B in 40 min and sample collection (15 s/fraction) started 5 min after precolumn switching. Separation on the monolithic column was performed with 0.05% H_3PO_4 as solvent A and transport liquid and 80% ACN with 0.04% H_3PO_4 in solvent B at a flow rate of 2.5 μL/min. Monolithic separation (5 cm × 200 μm, Dionex) was performed in direct injection mode with temperature of the column oven set to 60°C. A two step linear gradient from 3 to 15% B in 6 min and 15–65% B in 8 min followed by 0.5 min 65%B and 4 min 95% solvent B was applied 2 min after injection and samples were collected every 7 s, starting 1 min after injection.

2.5 MALDI MS

Acquisition parameters, resolution and laser energy were optimized prior to MS analysis on the 4800 MALDI TOF/TOF Analyser (Applied Biosystems) using the 14 peptide mixture from Table 1 (peptides 1–14) with 1 pmol/peptide. MS/MS optimization was performed on peptides 1 for determination of key acquisition parameters ($x2$ and $y2$ voltage) and 14 was used as a control for multiply phosphorylated peptides. The instrument was operated in positive ionization mode unless otherwise stated. Samples crystallized with CHCA were analysed in MS mode with 2000 shots/spot to ensure good S/N quality for precursor selection. DHB crystallized samples gave high spectral quality after 1500 laser shots. MS/MS was done with varying number of laser shots in the case of directly spotted samples, for LC-MALDI analysis 4000 laser shots were applied for each selected precursor. Enhanced fragmentation was achieved by applying an increased pressure (2×10^{-6} bar) in the collision cell.

Obtained MS/MS data were smoothed automatically using a Savitsky–Golay function of 4th order with 7 points/peak. Generation of MASCOT generic files (mgf) was carried out with the implemented PEAKStoMASCOT tool from the 4000 Explorer software version 3.5.1 (Applied Biosystems). Spectral images for manual validation were prepared with the Data Explorer Software version 4.9 (Applied Biosystems).

2.6 Online nano-RP-HPLC/nano-ESI MS

For online nano-RP-HPLC/nano-ESI MS/MS measurements, the HPLC system Ultimate Plus (Dionex) was directly coupled to the nano-ESI source (Proxeon, Odense, Denmark) instead of the Probot device. It was operated with 0.1% TFA as loading solvent, 5% ACN with 0.1% formic acid (FA) was solvent A and solvent B consisted of 80% ACN and 0.08% FA. After sample load, salts were removed on the trap column (PepMap C18, 5 mm × 300 μm × 3 μm, 100 Å, Dionex) for 40 min at a flow rate of 20 μL/min. For separation on a 15 cm C18 column (PepMap C18, 15 cm × 75 μm × 3 μ, 100 Å, Dionex), a gradient from 0 to 35% B in 25 min and to 60% B in 5 min (flow rate 0.275 μL/min) was applied. MS detection was carried out either on an LTQ-Orbitrap hybrid mass spectrometer (Thermo Fisher, Bremen) or LTQ mass spectrometer equipped with ETD (Thermo Fisher).

Table 1. Synthetic peptides

No.	Name	M_r	Sequence	$[M - H]^+$
1	NoP_1	1569.6696	EGVNDNEEGFFSAR	1570.6769
2	NoP_2	1263.6208	APPDNLPSPGGSR	1264.6281
3	OneP_1	963.3587	SV**S**DYEGK	964.3660
4	OneP_2	1137.5292	N**S**VEQGRRL	1138.5365
5	OneP_3	1302.5493	LIEDNE**Y**TAR	1303.5566
6	OneP_4	1313.5990	R**S**DGGHTVLHR	1314.6063
7	OneP_5	1343.5871	APPDNLPSPGG**S**R	1344.5944
8	OneP_6	1758.7938	SVENLPEAGI**T**HEQR	1759.8011
9	OneP_7	1931.8296	ENIMR**S**ENSESQLTSK	1932.8369
10	OneP_8	2026.0176	THILLFLPK**S**VSDYEGK	2027.0249
11	OneP_9	2093.8613	TASDTDSS**Y**AIPTAGMSPSR	2094.8686
12	TwoP_1	2229.9807	QLGEPEK**S**QDSSPVL**S**ELK	2230.9880
13	ThreeP_1	2309.9471	QLGEPEK**S**QD**S**SPVL**S**ELK	2310.9544
14	ThreeP_2	1792.6471	FG**SS**NTDSAGALG**T**LR	1793.6544
15	OneP_10	1655.6842	WWGSGPSGSGG**S**GGGKR	1656.6915
16	OneP_11	1792.7431	HWWGSGPSGSGG**S**GGGKR	1793.7504
17	TwoP_2	1735.6505	WWGSGPSG**S**GG**S**GGGKR	1736.6578
18	TwoP_3	1872.7094	HWWGSGPSG**S**GG**S**GGGKR	1873.7167
19	ThreeP_3	1815.6169	WWGSGP**S**G**S**GG**S**GGGKR	1816.6241
20	ThreeP_4	1952.6758	HWWGSGP**S**G**S**GG**S**GGGKR	1953.6830
21	FourP_1	1895.5832	WWG**S**GP**S**G**S**GG**S**GGGKR	1896.5905
22	FourP_2	2032.6421	HWWG**S**GP**S**G**S**GG**S**GGGKR	2033.6494
23	FourP_3	2863.1732	RELLNVPGE**S**IVE**S**L**SS**EESITR	2864.1805

Peptides that were used for optimization of MALDI MS/MS and chromatographic separation. Phosphorylated amino acids are in bold and underlined; peptides 1 and 2 are not phosphorylated.

2.7 Nano-ESI-MSA MS

The LTQ-Orbitrap mass spectrometer was operated in positive ionization mode and MS survey scans were carried out in the orbitrap. Survey scans were internally calibrated using a lock mass at 445.120024 m/z, and up to six signals were selected for MS/MS analysis in the IT part of the instrument in a single cycle. Neutral loss masses of 32.6, 49 and 98 m/z were set to perform MSA of phosphorylated peptides [29]. Signals with charge state one were completely rejected from MS/MS analysis whereas peptide masses which already have been fragmented were put onto an exclusion list for 180 s. Monoisotopic precursor selection (MPS) was activated because earlier experiments showed improved identification of peptides with this feature.

2.8 Nano-ESI-ETD MS

The LTQ instrument was operated in positive ionization mode and the nanospray source of Proxeon was used with off-line borosilicate emitters of the same company. The instrument was offline calibrated and optimization of ETD fragmentation resulted in an activation time of 90 ms. A 2.5 pmol/μL solution of peptide FourP_2 was directly injected with a flow rate of ~50 nL/min. The electrospray voltage was set to 850 V, fragment ion spectra were manually recorded in the mass range of 180–2000 Da for doubly and triply charged peptide ions. Ions were selected from a 1 Da mass window with automatic gain control. One microscan was acquired for each mass spectrum.

2.9 Data analysis

Analysis of MALDI MS data was accomplished using either MASCOT (version 2.2.1, Matrixscience, London, UK; www.matrixscience.com) or Protein Pilot (Applied Biosystems, Foster City, USA) [41]. Interpretation of all results of MALDI or ESI MS data *via* database searching was accomplished using an in-house database and all identified MS/MS spectra of phosphorylated peptides were manually confirmed. The in-house database consists of approximately 500 entries of several standard proteins used for LC-MS optimization, all MS relevant synthetic peptides of our laboratory and proteins which were analysed by LC-MS/MS in different other experiments. Using this database it was possible to perform MASCOT searches of MS optimization experiments with synthetic standard peptides. For identification of nonphosphopeptides, a general MASCOT score of 30 was applied. Phosphopeptides with scores above 15 were manually interpreted, if they match to the '*in silico*' digest of osteopontin. The significant similarity of ESI and MALDI data for CID activation was helpful to annotate low scoring ESI spectra of multiply phosphorylated peptides.

MASCOT search parameters were 0.1 Da mass error for precursor ions and 0.3 Da for fragment ion signals, in case of tryptic digests carbamidomethylation was selected as fixed modification. Phosphorylation of S, T or Y was set as variable modification, loss of water during sample preparation was also considered as well as methionine oxidation. Four missed cleavages were allowed with trypsin as enzyme. Precursor charge state was set to 1 and MALDI TOF/TOF was selected as instrument type.

Data analysis of MSA results of osteopontin was performed with the MASCOT Daemon software (Matrixscience, GB). Parameters for database search were equal to MALDI experiments except of charge state which was automatically assigned, MS tolerance of 3 ppm and MS/MS tolerance of 0.6 Da. In this search, the instrument type was set to LCQ.

Acquired ETD data were analysed with MASCOT with following parameter settings: MS tolerance 2 Da, fragment mass tolerance 0.8 Da, maximal missed cleavages 1, instrument type ETD-trap and phosphorylation of S or T as variable modification.

3 Results

3.1 Optimization of crystallization conditions for detection of phosphorylated peptides

A crucial first step for obtaining reliable identification of complex phosphopeptides with MALDI MS was to establish a suitable protocol for matrix crystallization. In previous studies, several different matrices have been advertized as suitable for phosphopeptide analysis by MALDI, however, so far there exists no commonly accepted protocol. One important aspect in matrix selection is the composition and concentration of the crystallization mixture as the eventual crystal structure and size play an important role in the ionization efficiency of peptides. Commonly, best results are obtained from sample spots with homogeneously dispersed small crystals. With regard to phosphopeptide analysis another important point is the use of PA in the matrix solvent to prevent phosphopeptide binding to the stainless steel surface of the MALDI sample plate [42]. We expected that the addition of PA would be particularly important for a low detection limit of multiply phosphorylated peptides.

Our initial optimization experiments were carried out using a mixture of peptides that were in a low phosphorylation state consisting of peptides 1–13 (13PPMix) of Table 1 with 10 pmol of each peptide in 0.1% TFA from peptide stock solutions (1 mg/mL in water). This mixture was further diluted with 0.1% TFA to obtain a series of peptide concentrations from 10 nM to 10 μM. Four different MALDI matrices were chosen to test the ionization efficiency of phosphopeptides: CHCA, DHB, indolylacrylic acid (IAA) and 2,5-dihydroxyacetophenone (DHAP) [9]. CHCA crystallization was performed with a solution of 2 mg/mL in 70% ACN containing 0.1% TFA, the other matrices were dissolved to 10 mg/mL in 50% ACN, 0.1% TFA. We found that IAA (hint from ABRF forum) or DHAP showed good results for high amounts of phosphopeptides (>2 pmol/spot, data not shown) but yielded unsatisfactory and irreproducible detection at low amounts (<200 fmol/spot). For IAA, we surmise that this deficiency is due to a polymerization product that limits the number of useful laser shots that can be applied to the sample.

In further optimization experiments, we therefore focussed on the use of CHCA and DHB. CHCA was used at a concentration of 2 mg/mL in 70% ACN, 0.1% TFA and crystallization with DHB was performed with a 10 mg/mL solution in 50% ACN 0.1% TFA. Both matrices turned out to be sufficient for detection and identification of pure phosphopeptides at low amounts (10 fmol/spot). Although we could not detect multiphosphorylated peptides spiked into a nonsphophorylated peptide sample at such low concentrations using CHCA matrix. We were able to identify them at least at the concentrations of 200 fmol/spot. Comparable ionization efficiencies were observed for not phosphorylated and monophosphorylated peptides in samples of 50 fmol/μL synthetic phosphopeptides that were mixed with 50 fmol/μL of a tryptic digest of cytochrome *c* (Fig. 1) to simulate the conditions of biological samples. Only samples that were crystallized with DHB exhibited signals of multiply phosphorylated peptides at this concentration. Although these signals were not intense they were reproducibly detected in repeat measurements on ten sample spots. In a mixture containing exclusively phosphorylated peptides (peptides 15–22), signal suppression of multiply phosphorylated peptides was reduced only with DHB as matrix (data not shown), when CHCA and DHB were compared from the analysis of equal samples. It was described earlier that mixtures of both matrices provide better results than the pure substances alone [43]. CHCA (10 mg/mL) and DHB (10 mg/mL) were combined in different ratios, and ionization efficiency of the matrix mixtures was compared to pure matrix substances. In contrast to published evidence, we observed no significant improvements from these experiments (data not shown). We therefore chose pure DHB as matrix substance in all further experiments.

3.2 Optimization of DHB crystallization

For further improvement of our procedure, the signal intensity of phosphopeptides was used as a parameter. The 13 PPMix was crystallized with different amounts of DHB (from 50 to 1 mg/mL) and each mixture was analysed in triplicates. As expected high peptide concentrations (10 pmol−800 fmol/spot) work very well with higher amount of DHB (50 mg/mL) but formation of multiple layers of big crystals was observed. In the case of low peptide amounts and high DHB concentration, huge background signals were present especially in the low mass area (900–1300 m/z) affecting the detection of phosphorylated peptides.

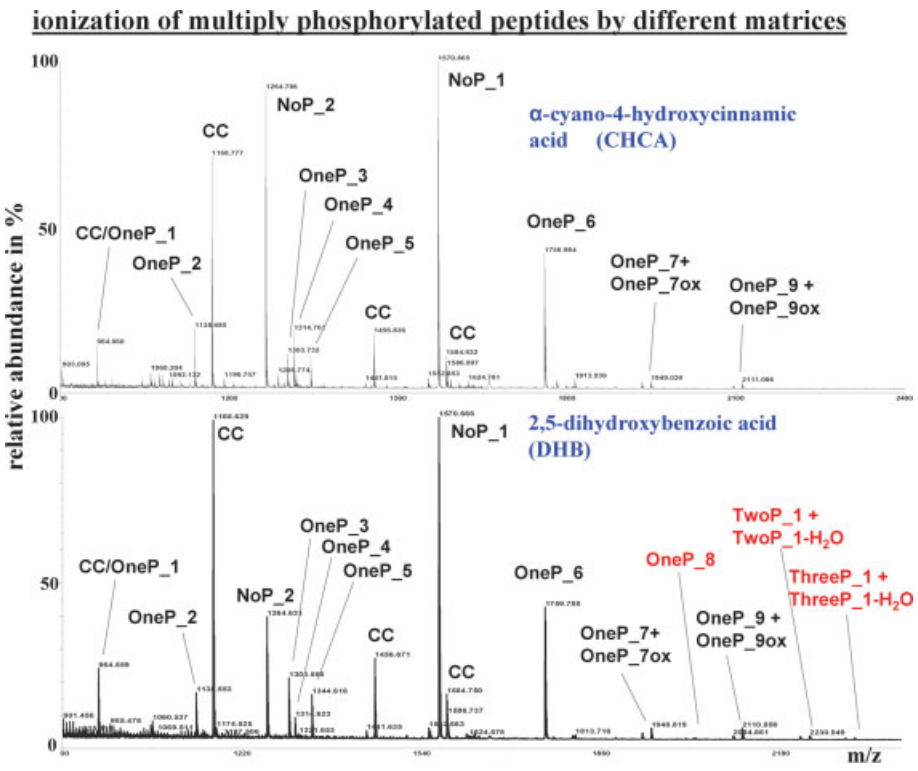


Figure 1. Ionization of multiply phosphorylated peptides with different matrices. Two reTOF mass spectra of a mixture of 13 phosphopeptides (Table 1, peptides 1–13) with a tryptic digest of horse heart cytochrome *c* (CC) (50 fmol/μL of each peptide) are shown (mass range 900–2400 Da). Peptides are indicated in reference to Table 1 in the case of peptides from CC they are named CC. The upper spectrum is taken from a mixture crystallized with HCCA (2 mg/mL) and the lower spectrum shows the sample with DHB (10 mg/mL). Multiply phosphorylated peptides TwoP_1 and ThreeP_1 are only detected in the DHB sample.

If low matrix concentrations (2 and 1 mg/mL) were applied two effects were quite obvious. On one hand, low reproducibility of crystallization was achieved. The crystal spot varied in size and showed a ring of thin DHB needles which surrounded a surface of small crystals. On the other hand, only a low number of spectra could be acquired because of the high consumption of matrix during ionization. Matrix concentrations in the range of 5–20 mg/mL did not tremendously differ in their ionization behaviour for peptide amounts of 5 pmol−20 fmol/spot, while lower sample amounts performed better with 10 and 5 mg/mL DHB.

DHB dissolved in a mixture of 80% ACN v/v, 1% TFA v/v and 0.2% H_3PO_4 v/v gave the best ionization results and was used in a concentration of 10 mg/mL in all following experiments. Optimized matrix conditions for offline experiments were also tested in their LC-MALDI compatibility, where the conditions from offline experiments (10 mg/mL DHB in 80% ACN, 1%TFA, 0.1% H_3PO_4) also proved to be optimal for LC-MALDI experiments (comparison of ionization modes see Supporting Information).

3.3 Affinity enrichment of phosphorylated peptides

In order to effectively detect phosphopeptides in low complexity samples of purified proteins or protein complexes, an enrichment step using TiO_2 was introduced into the protocol. The first affinity purification demonstrated that the previously published method from Mazanek *et al.* [22] is not compatible with MALDI MS without further purification of the eluted phosphopeptide fraction. Dominating signals of octylsulphonate multimers were observed with their characteristic mass difference of 194 Da and suppressed all peptide signals completely. To prove enrichment efficiency, it was necessary to adapt this protocol for direct detection of enriched phosphopeptides by MALDI MS.

A major change was applied for sample loading were the previously described loading buffer was replaced by 200 mg/mL DHB (Sigma), 35 mM OcSNa, 23% AcOH, 40% ACN and 0.1 HFBA. Peptide samples were diluted with 50% ACN (containing 10% AcOH) to ensure a high amount of organic solvent which promotes phosphopeptide binding to TiO_2. Additionally volumes for washing were increased to 60 μL of diluted loading buffer and 100 μL of 70% ACN containing 0.2% HFBA. For effective elution of 50 mM $NH_4H_2PO_3$, pH 11 adjusted with NH_3, was applied to the titanium column.

Protein digests (300 fmol) of β-casein and osteopontin were treated according to the new protocol and the enrichment is displayed in Fig. 2 in comparison to the untreated sample. In case of β-casein, the signal of the monophosphorylated peptide FC**S**EEQQQTEDELQDK (amino acids 33–48, *m/z* 2061.83) which is nearly lost in the spectral background without enrichment represents the base peak after TiO_2 enrichment (black star). The four-fold phosphorylated peptide of beta casein at *m/z* 3122.26 (amino acids 1–25, RELEELNVPGEIVE**S**L**SSS**EESITR) which was not detected in the untreated digest is greatly enriched and two

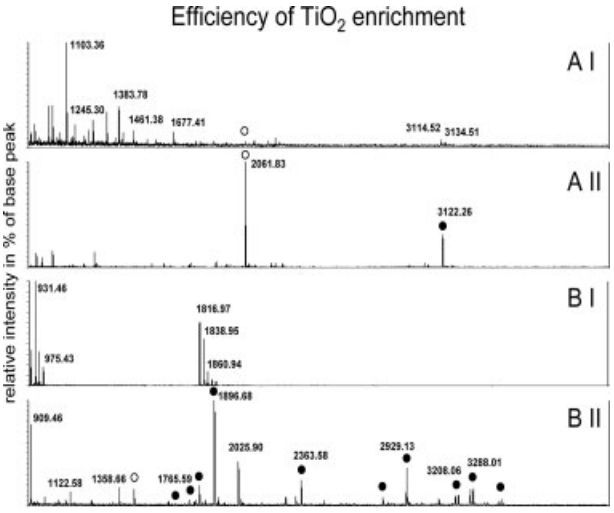


Figure 2. Efficiency of TiO_2 enrichment of tryptic digests of β-casein and osteopontin. MALDI reTOF MS spectra (mass range 900–4000 *m/z*) are shown for tryptic digest of β-casein (A) and osteopontin (B) before TiO_2 enrichment (A I and B I) and the eluted phosphopeptide fraction (A II and B II). White circles indicate single phosphorylated peptides, multiply phosphorylated peptides are indicated with black circles.

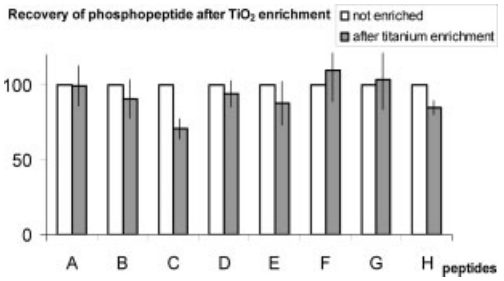


Figure 3. Recovery of phosphopeptides after TiO_2 enrichment. Eight synthetic peptides were mixed and split into two parts which were labelled with iTRAQ reagents 114 and 115, respectively. The 115 labelled fraction was treated according to the described TiO_2 enrichment protocol and sample 114 was added to the elution fractions in equal amount to loaded sample 115. The iTRAQ ratios were determined by MALDI TOF/TOF MS/MS MS and averaged over five individual enrichment steps.

adjacent signals of unphosphorylated peptides are completely removed. The seldom observed signal at *m/z* 2966.16 which belongs to the four-fold phosphorylated peptide without the N-terminal arginine residue is detected, but it has only low intensity because trypsin is not able to cleave this protein N-terminus effectively.

The MS spectrum of the osteopontin digest contains few major peaks with up to four sodium adduct signals, because the protein was purchased with a sodium containing buffer substance. Signals of phosphorylated peptides are not detected (Fig. 2) in this spectrum. After enrichment almost all major signals are related to phosphorylated peptides of osteopontin and most of them have multiple phosphosites (black circles). The base peak corresponds to the three-fold phosphorylated C-terminal peptide IRI**S**HELD**S**A**S**SEVN (*m/z* 1896.72).

To get more insight how the new protocol improves enrichment of multiply phosphorylated peptides, we set up an experiment to determine the recovery from TiO_2 in the elution step. Eight synthetic peptides of different phosphorylation state and hydrophobicity were mixed and two equal fractions were labelled with the iTRAQ reagents 114 and 115. 2 pmol/TiO_2 column of the 115 fraction were treated according to the described protocol in five individual enrichment experiments. Sample 114 was mixed to the eluted fractions in equal amount to the loaded sample 115. The resulting mixture was crystallized with DHB as matrix and iTRAQ ratios were determined in MS/MS experiments. The obtained ratios are shown in Fig. 3, the area of signal 114 was used as reference and is therefore set to 1. Almost all phosphopeptides are recovered to about 90% from the titanium column and there are no differences between singly (peptides A, B, C) and multiply phosphorylated peptides (two phosphates D, E, three phosphates F, G, four phosphates H) visible. Only peptide C which has strong hydrophobicity has a lower recovery from the column which can be due to a higher solubility in the washing solution compared to the other peptides.

3.4 HPLC fractionation of phosphorylated peptides

Experiments with mixtures of unphosphorylated, singly and multiply phosphorylated peptides showed a tremendous suppression of signals of multiply phosphorylated peptides in the MALDI ionization step (Fig. 1 and Supporting Information Fig. S1). Separation of phosphorylated peptides is necessary to reduce suppression and therefore enhance identification of multiply phosphorylated peptides. RP-HPLC is capable of separating complex peptide mixtures according to their hydrophobicity as well as phosphorylated peptides with different number and positions of phosphorylated amino acids, even if they share the same amino acid sequence. Unfortunately, low recovery rates are often observed for highly phosphorylated peptides upon separation on C18 columns under standard conditions. TFA as well as FA which are widely used for RP-HPLC separations lead to broad peaks which are also a reason why the detection limit of multiply phosphorylated peptides is quite high. Integrating H_3PO_4 as acidic component of the solvent system significantly improves peak shapes of multiply phosphorylated peptides on C18 columns because undesired interactions of the phosphate moiety with metal parts of the HPLC system, *e.g.* titanium valves are suppressed. Unfortunately, even low amounts of PA cannot be combined with online ESI MS detection.

Two separation systems were compared for their resolving power and recovery of multiply phosphorylated peptides

from simple peptide mixtures. Phosphopeptides from 1 and 3 pmol osteopontin tryptic digest were enriched on TiO_2 using the new protocol and the sample was split for separation on a monolithic column and an in-house packed C18 column. Samples representing 300 fmol and 1 pmol of the TiO_2 elution fraction were injected on each HPLC system and eluting fractions were spotted on a MALDI sample plate. The monolithic system was operated in direct injection mode to ensure high recovery of weakly binding peptides. A fractionation time of 7 s for the monolith was applied which resembled roughly the double peak width at half height and ensures small sample droplets. For C18 chromatography, the fraction time was set to 15 s which is about one and a half peak widths and was optimized in earlier experiments.

Interestingly, the order of the eluting peptides was different between monolithic and C18 column (Fig. 4). Highly phosphorylated peptides showed significantly increased retention times on C18 material compared to peptides in lower phosphorylation state. This becomes quite obvious from the selected ion chromatograms (SICs) of peptides OneP_8 and ThreeP_2 in Fig. 4, where the strongly hydrophobic monophosphorylated peptide OneP_8 elutes earlier from the C18 column than the triply phosphorylated peptide. Also peptides which share the same amino acid sequence but have a different number of phosphate moieties became separated while the peptide with less phosphoamino acids eluted earlier than the higher phosphorylated form (see TwoP_1 and ThreeP_1).

However this strong retention of multiply phosphorylated peptides could not be observed in the monolithic separation, but hydrophobicity of the peptide backbone is the driving force for retention of peptides which is demonstrated by comparing peptides OneP_8 and ThreeP_2. Peptides with high backbone hydrophobicity elute later from the monolithic column. Phosphopeptides with same amino acid sequence but different phosphorylation state eluted in reversed order to the C18 column from monolithic columns as shown for peptides TwoP_1 and ThreeP_1. This effect was also observed for multiply phosphorylated peptides from osteopontin. Some late eluting highly phosphorylated peptides did not show a sharp elution profile on C18 columns, for example peptide osteopontin B. In contrast, monolithic

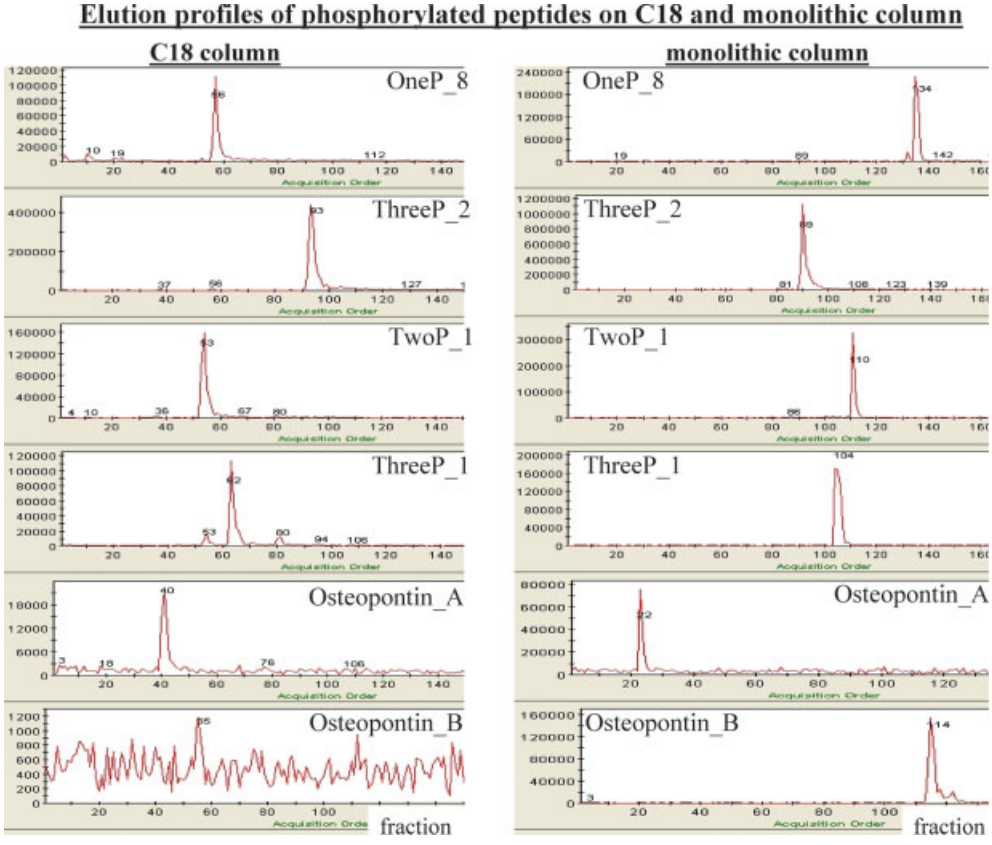


Figure 4. Elution profiles of phosphorylated peptides on C18 or monolithic column. SICs for several phosphopeptides, including two osteopontin peptides are displayed for separation on either C18 or monolithic column. Interestingly all highly phosphorylated peptides have significantly increased retention times on the C18 column compared to single phosphorylated peptides (OneP_8) or can even not be eluted in a sharp signal (osteopontin_B). Earlier elution of highly phosphorylated peptides (ThreeP_2) compared to singly phosphorylated ones is observed from the monolithic fractionation.

separation leads to a sharp elution profile and a high recovery rate. This peptide is three-fold phosphorylated SNVQSPDA-**T**EEDFT**S**HIE**S**EEMHDAPK ($[M + H]^+$ 3286.18 Da, amino acids 155–181, MSO at M22). Although spot time for the monolithic separation was only half of that of the C18 system, peptides were often detected only in a single spot or in two adjacent spots whereof one contained more than 80% of the peptide.

These observations lead to the impression that C18 chromatography contributes to higher detection limits for multiphosphorylated peptides because of a stronger retention compared to monophosphorylated peptides and undefined broad elution. Monolithic columns on the other hand do not have the peak capacity of C18 columns but allow fast and highly reproducible separations with sharp peak shapes. Based on our data we suggest the use of monolithic column for the analysis of enriched phosphopeptides from low complexity samples.

3.5 Phosphosite determination of osteopontin

Fractions of 500 fmol of TiO_2 enriched phosphopeptides from tryptic digest of osteopontin were analysed using either LC-ESI MSA MS/MS or LC MALDI TOF/TOF MS/MS. The enriched phosphopeptides were fractionated by nano-RP-HPLC using the standard set-up with C18 column and online MS was carried out with MSA which combines the quality of MS^n spectra with improved scan speed. A second phosphopeptide fraction was separated using the monolithic HPLC system and fractions were spotted every 7 s on a MALDI target sample plate, which was prespotted with 12 mg/mL DHB. Two hundred fractions were collected starting from 1 min after injection of the samples to ensure that also nonbinding peptides are found in the early fractions. Both approaches identified 15 phosphosites out of 28 predicted phosphosites of osteopontin with MASCOT scores above 20 (Fig. 5). To evaluate the obtained data, phosphopeptides were compared to the theoretical phosphorylation state according to the Swiss-Prot database entry for PTMs of osteopontin [5].

Interestingly, identified phosphopeptides from MSA showed a very high diversity of phosphorylation states for peptides with multiple phosphorylations and often monophosphorylated species scored highest. Peptides in expected phosphorylation states did not receive good identification results. One reason for poor identification of these peptides was that the final separation step was carried out on a C18 column with solvents containing FA as acidic component which resulted in peaks eluting over a broad time span thus decreasing the number of ions available for a single spectrum. This effect led to the acquisition of up to ten individual MS/MS spectra of a single peptide.

MALDI TOF/TOF MS/MS analysis showed a different picture herein almost all identified phosphopeptides were in there expected phosphorylation state or missed only one phosphate moiety. Signals for lower phosphorylation states

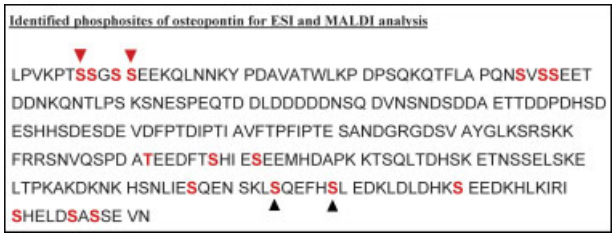


Figure 5. Identified phosphosites on osteopontin. Phosphorylated amino acids of bovine osteopontin which were identified by ESI-MSA and MALDI TOF/TOF MS/MS (bold red letters), black upfacing triangles mark phosphosites exclusively identified with ESI-MS/MS (MASCOT score >20), red downfacing triangles stand for phosphoamino acids only found in MALDI TOF/TOF data (score >30). All phosphopeptide spectra were manually interpreted to avoid false positive identification.

were less intense and often not selected for MS/MS analysis. Their abundance was elucidated by their mass traces which had up to 20 times lower intensity compared to the identified phosphopeptide species.

Significantly higher scores for multiphosphorylated peptides were achieved with the MALDI approach. This might be due to a lower number of background signals because maximal 200 masses *per* MS/MS spectrum were extracted for the mgf-file. Additionally the upper mass area, where signals of different neutral losses of phosphate moieties occur, was deleted manually for each peptide before database search. This significantly improved identification, because MASCOT is not able to recognize signals of neutral losses from the intact peptide (Supporting Information Fig. S2). Two major criteria were applied for manual deletion of this mass area, first the most intense signals of MS/MS data were related to neutral loss of H_3PO_4 or HPO_3 which was determined by manual inspection of the spectrum. As second criterion a database search in MASCOT with the original mgf-file was performed to decide if multiple phosphorylations were present which was used to determine the minimal range of masses that was deleted.

3.6 Phosphosite assignment

Although MSA leads to the observation of series of *b*- and *y*-ions, while in MALDI TOF/TOF spectra often only one type of fragment ions was predominantly present, exact pinpointing of the phosphorylation site was more successful with MALDI data (Table 2). This aspect is shown for three highly phosphorylated peptides of osteopontin which have been identified with ESI and MALDI.

MALDI data yield higher identification scores for multiply phosphorylated peptides compared to ESI so that the N-terminal peptide LPVKPT**S**G**S**EEK (MR 1764.61) as well as the C-terminus IRI**S**HELD**S**A**S**SEVN (MR 1895.71) of osteopontin do not even receive a score of 30 for the highest scoring spectrum from ESI data see Table 2. Manual interpretation was done on each phosphopeptide with score above

Table 2. Automatic annotation of phosphorylation sites by MASCOT

Phosphosite distribution		MASCOT score	
		ESI	MALDI
Sequence A			
A-1	LPVKPT**SS**G**SS**EEK	5.9	50.5
A-2	LPVKP**T**S**S**G**SS**EEK	17.8	45.0
A-3	LPVKP**TS**SG**SS**EEK	11.4	32.3
A-4	LPVKP**TSS**GS**S**EEK	8.8	29.2
A-5	LPVKP**TSS**G**S**SEEK	8.8	27.5
Sequence B			
B-1	IRI**S**HELD**S**A**S**SEVN	16.4	54.2
B-2	IRI**S**HELD**S**AS**S**EVN	10.2	44.1
B-3	IRI**S**HELDSA**SS**EVN	9.3	35.1
B-4	IRISHELD**S**A**SS**EVN	3.2	22.1
Sequence C			
C-1	SNVQSPDA**T**EEDFT**S**HIE**S**EEMHDAPK	34.4	70.8
C-2	SNVQ**S**PDATEEDFT**S**HIE**S**EEMHDAPK	15.7	50
C-3	**S**NVQSPDATEEDFT**S**HIE**S**EEMHDAPK	6.5	50
C-4	SNVQSPDATEEDF**TS**HIE**S**EEMHDAPK	20.8	46.4
C-5	SNVQSPDA**T**EEDF**T**SHIE**S**EEMHDAPK	34.4	45
C-6	SNVQ**S**PDATEEDF**T**SHIE**S**EEMHDAPK	10.3	39.6
C-7	**S**NVQSPDATEEDF**T**SHIE**S**EEMHDAPK	4.4	39.6
C-8	SNVQSPDA**T**EEDF**TS**HIESEEMHDAPK	19.2	22.5
C-9	SNVQ**S**PDA**T**EEDFTSHIE**S**EEMHDAPK	2.7	22.2
C-10	**S**NVQSPDA**T**EEDFTSHIE**S**EEMHDAPK	n.a.	22.2
C-11	SNVQ**S**PDATEEDF**TS**HIESEEMHDAPK	6.1	n.a.

MASCOT scores for different phosphosite distributions of three peptides of osteopontin (A, B, C) are listed in the table. Putatively identified phosphosites are underlined and in bold. From the Swiss-Prot database entry predicted distribution has a grey background. All peptides contain at least one additional T or S residue next to the phosphoamino acids, which is not phosphorylated. Only MALDI data lead to the identification of the expected phosphosite distribution with the highest score.

15. Interestingly, a high similarity on the fragment ion level was found between both activation mechanisms. Exact identification of the phosphorylated amino acid is often only achieved from MALDI data, because although fragment ions do not have high intensity relevant signals are only detected in MALDI MS/MS spectra. ESI spectra yield also a high number of neutral loss fragment ions which do not allow an unambiguous phosphosite alignment especially in the case of peptides containing other putative phosphoamino acids.

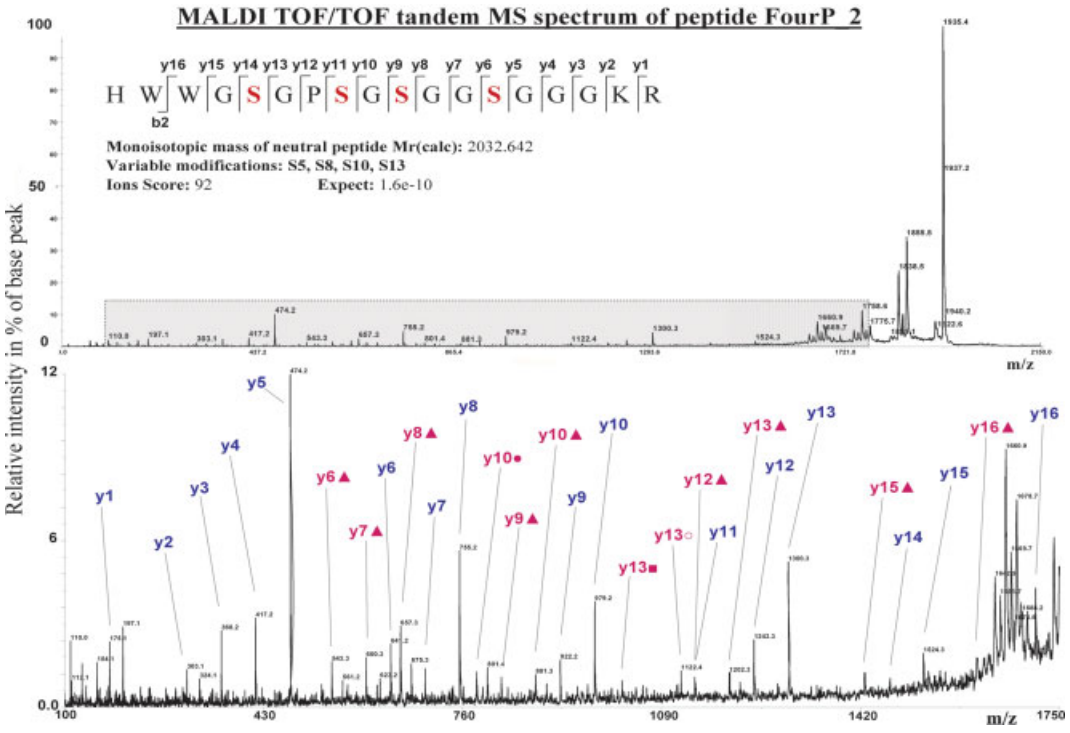
The score difference between first and second hit of the MALDI data is quite small for peptide LPVKPT**SS**G**SS**EEK. This result is due to the fact that almost all identified signals overlap between the two possible peptides because one phosphate is ambigously assigned to two adjacent amino acids. Only fragments at T6 or S7 are responsible for the identification. Assignment of the first phosphate moiety to S7 is difficult because of a neutral loss signal of water from the threonine side chain which is present next to the signal for the unmodified amino acid. The signal which includes the intact phosphoamino acid does not exist. Also two signals are observed for the first serine at position 7, one with intact phosphate moiety and one with neutral loss of H_3PO_4, no signal for the unmodified serine is present which determines the first phosphorylation to S7. The result of the corresponding ESI spectrum leads to the assumption of a different phosphate distribution. The phosphosites of the C-terminal peptide of osteopontin are correctly assigned for ESI and MALDI data but the ESI spectrum obtained a low score and was even not in the list of identified peptides for osteopontin. The third peptide SNVQSPDA**T**EEDFT-**S**HIE**S**EE**M**HDAPK (MSO at M22) presents another shortcoming of ESI data, where automatic interpretation does not always lead to a single distribution, but two or more forms receive the same score. The identification from MALDI data is quite unambiguous with a score difference of 20 between first and second hit (for annotated spectra see Supporting Information).

3.7 Relevance of the method in contrast to novel fragmentation techniques

To evaluate the viability of our method, MS/MS spectra of the four-fold phosphorylated peptide HWWG**S**G-P**S**G**S**GG**S**GGGKR (FourP_1) were compared for either MALDI TOF/TOF PSD fragmentation or ETD in an LTQ IT instrument (Figs. 6 and 7). For MALDI TOF/TOF experiments, 100 fmol peptide was crystallized with 0.8 μL of 12 mg/mL DHB matrix and 8000–12 000 shots were summed up for one MS/MS spectrum. Peaks were exported using the PEAKStoMASCOT tool from the 4000 Explorer Software (Applied Biosystems) and the upper mass region where neutral losses of H_3PO_4 and HPO_3 occur was manually removed from the mgf-file.

Base peak of the MALDI MS/MS spectrum (Fig. 6) was the neutral loss of a single H_3PO_4 molecule but also species with two or three neutral losses were identified with high intensity. Interestingly no signal for the loss of one molecule HPO_3 from the intact peptide was observed, but evidence of this loss was taken from backbone fragments carrying a phosphorylated amino acid. Nearly the complete *y*-ion series was identified in the spectrum, which suggests that the basic residues at the C-terminus are predominantly charged. Next to the original *y*-ion series, many *y*-ions with either loss of H_3PO_4 (triangles) or HPO_3 were detected with lower signal intensity. This series started at the first phosphorylated amino acid. Some fragment signals showed even more than a single neutral loss (circle or square). Interestingly all *y*-ions, resulting from the break N-terminally to phosphorylated amino acids, have low intensity, while fragments of the following unphosphorylated amino acids where significantly stronger.

ETD spectra were acquired on an LTQ-ETD instrument by direct infusion of a 2.5 pmol/μL solution of the peptide. Two- and three-fold charged precursors were selected for MS/MS experiments to reveal the influence of the charge



#	Immon.	b	Seq.	y	y-80	y-98	#
1	110.1	138.1	H				18
2	159.1	324.1	W	1896.6	1816.6	1798.6	17
3	159.1	510.2	W	1710.5	1630.5	1612.5	16
4	30.0	567.2	G	1524.4	1444.4	1426.4	15
5	140.0	734.2	S	1467.4	1387.4	1369.4	14
6	30.0	791.3	G	1300.4	1220.4	1202.4	13
7	70.1	888.3	P	1243.4	1163.3	1145.3	12
8	140.0	1055.3	S	1146.3	1066.3	1048.3	11
9	30.0	1112.3	G	979.3	899.3	881.3	10
10	140.0	1279.3	S	922.3	842.3	824.3	9
11	30.0	1336.4	G	755.3	675.3	657.3	8
12	30.0	1393.4	G	698.3	618.3	600.3	7
13	140.0	1560.4	S	641.3	561.3	543.3	6
14	30.0	1617.4	G	474.3			5
15	30.0	1674.4	G	417.3			4
16	30.0	1731.4	G	360.2			3
17	101.1	1859.5	K	303.2			2
18	129.1		R	175.1			1

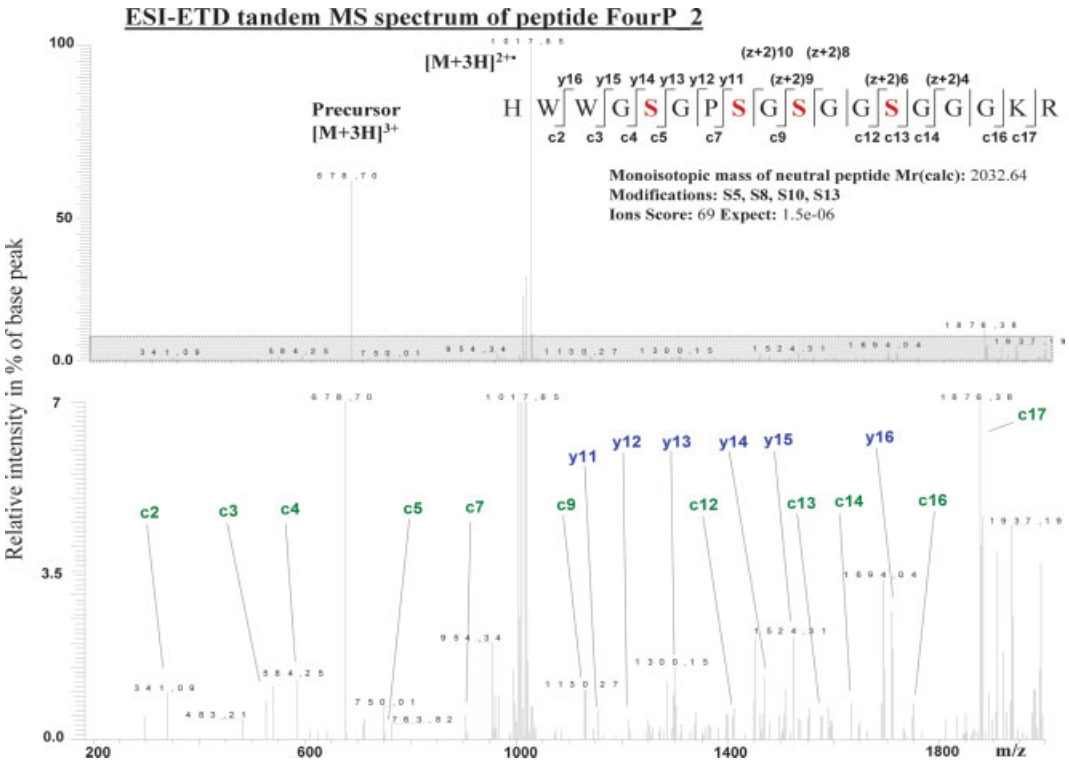
Figure 6. MALDI TOF/TOF tandem mass spectrum of peptide FourP_2. The MS/MS spectrum of the peptide was summed up from 12 000 shots. The upper spectrum shows the complete MS/MS spectrum, and the grey region is enlarged in the bottom. The spectrum is dominated by several neutral loss signals of either loss of H_3PO_4 or HPO_3 or combinations. A high number of *y*-type fragment ions could be identified in the spectrum (blue intact *y*-ions, red *y*-ions with loss of phosphate 1 × ▲, 2 × ●, 3 × ■). Spectral data were analysed by MASCOT and *y*-80 and *y*-98 ions were added to the fragment table manually. The mass accuracy of the MS/MS spectrum is 0.2 Da.

state on the applied fragmentation settings. The spectrum of the doubly charged precursor did not show formation of fragment ions, whereas the triply charged precursor was suitable for MS/MS analysis (Fig. 7). Base peak of the spectrum of FourP_1 is the radical ion $[M + 3H]^{2+\cdot}$ the second intense signal represents the precursor. Although fragment ions have only small intensity high sequence coverage is obtained for this peptide. Interestingly, a significant number of *y*-ions are present in the spectrum which was not expected from the underlying fragmentation mechanism.

As Figs. 6 and 7 show our method yields at least equal identification results as ETD fragmentation at low peptide amounts. Almost the complete peptide sequence with intact phosphate moieties was identified and *y*-ions that were not annotated by MASCOT can easily be assigned manually in the TOF/TOF MS/MS spectrum upon manual inspection. The set of *c*-ions, *y*-ions, $z + 1$-ions and $z + 2$-ions from the ETD spectrum cover also the whole peptide.

3.8 Isotopic labelling

Yet another benefit of our identification strategy is combination with multiple stable isotope labelling strategies for protein quantification, which is demonstrated in Fig. 8. The peptide FourP_1 was labelled with 4-plex iTRAQ reagents (isotopes 114 and 116) and mixed to give a 1:1 ratio. The spectrum is not severely altered by the addition of the iTRAQ moiety, all identified fragment ions carry a single intact iTRAQ label at the ε-amino function of K17. Signals of fragments with loss of the iTRAQ moiety are not detected. The



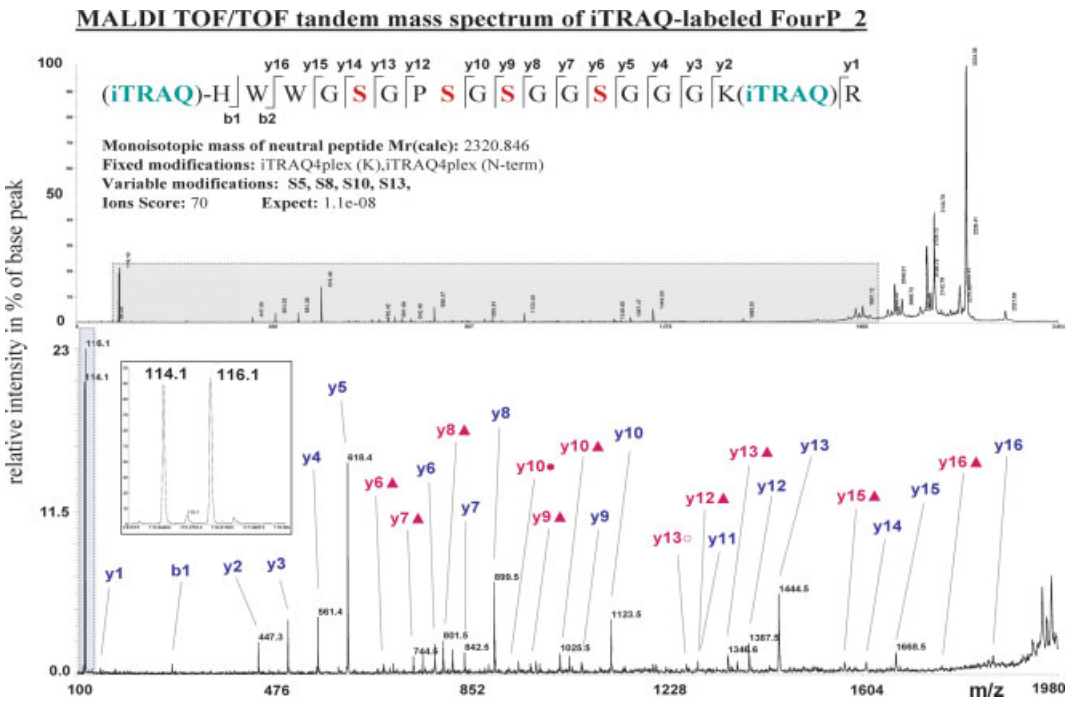
#	c	Seq.	y	z+1	z+2	z+2++	#
1	155.1	H					18
2	**341.2**	W	1896.6	1880.6	1881.6	941.3	17
3	**527.3**	W	**1710.5**	**1694.5**	**1695.5**	848.3	16
4	**584.3**	G	**1524.4**	1508.4	1509.4	755.2	15
5	**751.3**	S	**1467.4**	**1451.4**	1452.4	726.7	14
6	808.3	G	**1300.4**	**1284.4**	**1285.4**	643.2	13
7	**905.4**	P	**1243.4**	1227.4	1228.4	614.7	12
8	1072.3	S	**1146.3**	**1130.3**	**1131.3**	566.2	11
9	**1129.4**	G	979.3	963.3	**964.3**	**482.7**	10
10	1296.4	S	922.3	**906.3**	**907.3**	454.2	9
11	1353.4	G	755.3	**739.3**	**740.3**	370.7	8
12	**1410.4**	G	698.3	682.3	683.3	**342.2**	7
13	**1577.4**	S	641.3	**625.3**	**626.3**	313.6	6
14	**1634.4**	G	474.3	458.3	459.3	230.1	5
15	1691.5	G	417.3	**401.2**	**402.3**	201.6	4
16	**1748.5**	G	360.2	344.2	345.2	173.1	3
17	**1876.6**	K	303.2	287.2	288.2	144.6	2
18		R	175.1	159.1	160.1	80.6	1

Figure 7. ESI-IT ETD tandem mass spectrum of peptide FourP_2. 2.5 pmol/μL of FourP_2 peptide were injected using a static nanospray and ETD fragmentation was carried out on the triply charged precursor ion with optimized parameters. Enlarging the low intensity area below 7% of the base peak (grey box) reveals the fragment ions of *c*-, *z*- and *y*-type. Spectral data were analysed with MASCOT which led to an almost complete identification of all amino acids by combining the observed *c*-, *y*- and *z*-type ions. The mass accuracy of the MS/MS spectrum is 0.8 Da.

label is beneficial for fragment ions with a mass which is near to the area of H_3PO_4 losses from the complete peptide because this area is shifted further by two iTRAQ labels at the N-terminus and K17, while the corresponding fragment ions carry only a single iTRAQ modification. Identified *y*-ions are also almost complete for this spectrum, unfortunately two fragments at phosphorylated residues were not considered by MASCOT.

4 Discussion

Although a large number of tools exists that help in the identification of phosphorylation sites of peptides by MS, the analysis of multiply phosphorylated peptides has remained a difficult task. In this work, we have established a strategy that should help to overcome the limitations inherent in some of the more classical approaches. We developed a method consisting of three steps, enrichment, separation and detection/identification that allows analysis of highly phosphorylated peptides from digests of phosphoproteins and exact determination of the phosphorylated amino acids. Although our TiO_2 protocol is not designed and therefore not quite suitable for large-scale analysis of phosphorylation sites from whole cells, it yields excellent data for enriched phosphoproteins or protein complexes. As such it might be a valuable addition to our arsenal for determining complicated modification patterns.



#	b	Seq.	y	y-98	#
1	282.2	H			18
2	468.2	W	2040.7	1942.7	17
3	654.3	W	1854.6	1756.6	16
4	711.3	G	1668.5	1570.5	15
5	878.3	S	1611.5	1513.5	14
6	935.4	G	1444.5	1346.5	13
7	1032.4	P	1387.5	1289.5	12
8	1199.4	S	1290.4	1192.5	11
9	1256.4	G	1123.4	1025.4	10
10	1423.4	S	1066.4	967.4	9
11	1480.5	G	899.4	801.4	8
12	1537.5	G	842.4	744.4	7
13	1704.5	S	785.4	687.4	6
14	1761.5	G	618.4		5
15	1818.5	G	561.4		4
16	1875.5	G	504.3		3
17	2147.7	K	447.3		2
18		R	175.1		1

Figure 8. MALDI TOF/TOF MS/MS spectrum of iTRAQ-labelled FourP_2. The MALDI TOF/TOF MS/MS spectrum of 100 fmol iTRAQ-labelled peptide FourP_2 displayed in the figure. Again dominating peaks result from neutral losses of H_3PO_4 or HPO_3 and fragment ions as well as iTRAQ reporter ions have lower intensity. Interestingly *y*-type fragments with high intensity have an intact label and phosphorylation. In comparison to the unlabelled peptide not significant implication occurs upon labelling and the spectrum shows only small alterations such as the intact iTRAQ labels at H1 and K17.

Concerning the enrichment step, we opted for a purification protocol using TiO_2. In our opinion, TiO_2 has some advantages over the common IMAC method. First of all, acidic amino acid side chains must not necessarily be esterified which is a source of increased sample complexity if the reaction is not complete. Secondly, our protocol provides good recovery for multiple phosphorylated peptides also at variable sample amounts, thus allowing easy upscaling and downscaling of the enrichment procedure. In our experience, IMAC purification is sometimes sensitive to changes in the sample amount. Another important improvement in our procedure is the low phosphate concentration in the elution buffer, which on one hand ensures almost complete elution of multiple phosphorylated peptides from the TiO_2 surface and on the other hand allows direct and efficient analysis of the eluted fraction by MALDI MS.

With regard to the final separation step, monolithic HPLC proved to be an excellent alternative for separation of enriched phosphopeptide mixtures compared to HPLC on C18 columns [44, 45]. Advantages are reduced time requirement for chromatography combined with sharp elution profiles. Whereas separation on C18 material often results in broad peaks for multiply phosphorylated peptides. This detrimental effect is not only caused by the column material itself, but also by the solvent composition that might alter the retention behaviour of such peptides. We showed that PA as acidic component gave the best performance, but it cannot be used in combination with ESI ionization.

In contrast to ESI MS, MALDI is not affected by small amounts of phosphate in the HPLC solvent. It even proved to be necessary to prevent phosphopeptides from binding to the

surface of the steel target. Another benefit of the presence of phosphate in MALDI MS is the suppression of matrix cluster signals that can be formed during ionization. Interestingly, MALDI TOF/TOF MS/MS spectra of multiply phosphorylated peptides had a high number of fragment ions with intact phosphosites. These fragments were also observed in ESI MSA spectra which demonstrates high similarity between both activation methods for multiphosphorylated peptides. Concerning sensitivity problems which were discussed with MALDI MS, we found that our mass spectrometer had comparable sensitivity to common ESI approaches. Another benefit of the TOF/TOF analyser is high mass accuracy both in MS and MS/MS which allowed us unambiguous identification of phosphorylation sites. Excellent results could be obtained from off-line analysis low sample amounts with MALDI MS or ESI-ETD MS, demonstrating that our method is at least equally successful in the identification of complex phosphorylation patterns.

Finally, our approach is also compatible with relative quantification between two samples, if an isotope label is introduced. Almost all published labelling techniques can be applied because they are stable during the enrichment and separation process. We demonstrated the viability of the iTRAQ strategy to yield quantification data from MS/MS spectra of low sample concentrations. This method cannot be combined with other strategies for phosphopeptide identification like ETD or ECD because the isobaric label will not be fragmented to give the characteristic reporter ions which are necessary for quantification. Since identification and quantification is accomplished from a single LC run, slight differences which result from the separation conditions itself and from the crystallization step do not affect the quantification result.

This project was funded by the Christian-Doppler-Society, the Institute for Molecular Pathology (IMP), the University of Vienna. This work was further supported by the Austrian Proteomics Platform (APP) within the Austrian Genome Research Program (GEN-AU) and by the Mitocheck Project within the Sixth Framework Program of the European Commission. We thank all the members of the Protein Chemistry Facility and University MS unit for sharing HPLC and MS equipment, reagents and for fruitful discussions. We also thank the IMP Protein Chemistry Facility for the synthesis of phosphopeptides.

The authors have declared no conflict of interest.

5 References

[1] Linding, R., Jensen, L. J., Pasculescu, A., Olhovsky, M. *et al.*, NetworKIN: A resource for exploring cellular phosphorylation networks. *Nucleic Acids Res.* 2008, *36*, D695–D699.

[2] Agrawal, G. K., Thelen, J. J., Large scale identification and quantitative profiling of phosphoproteins expressed during seed filling in oilseed rape. *Mol. Cell. Proteomics* 2006, *5*, 2044–2059.

[3] Villen, J., Beausoleil, S. A., Gerber, S. A., Gygi, S. P., Large-scale phosphorylation analysis of mouse liver. *Proc. Natl. Acad. Sci.* 2007, *104*, 1488–1493.

[4] Nakajima, H., Tojoshima-Morimoto, F., Taniguchi, E., Nishida, E., Identification of a consensus motif for Plk (Polo-like kinase) phosphorylation reveals Myt1 as a Plk1 substrate. *J. Biol. Chem.* 2003, *278*, 25277–25280.

[5] Sørensen, E. S., Højrup, P., Petersen, T. E., Posttranslational modifications of bovine osteopontin: Identification of twenty-eight phosphorylation and three O-glycosylation sites. *Protein Sci.* 1995, *4*, 2040–2049.

[6] Lees, J. K., Buchkovich, K. J., Marshak, D. R., Anderson, C. W., Harlow, E., The retinoblastoma protein is phosphorylated on multiple sites by human cdc2. *EMBO J.* 1991, *10*, 4279–4290.

[7] Kang, J.-H., Toita, R., Oishi, J., Niidome, T., Katayama, Y., Effect of the addition of diammonium citrate to alpha-cyano-4-hydroxycinnamic acid (CHCA) matrix for the detection of phosphorylated peptide in phosphorylation reactions using cell and tissue lysates. *J. Am. Soc. Mass Spectrom.* 2007, *18*, 1925–1931.

[8] Steen, H., Jebanathirajah, J. A., Rush, J., Morrice, N., Kirschner, M. W., Phosphorylation analysis by mass spectrometry: Myths, facts, and the consequences for qualitative and quantitative measurements. *Mol. Cell. Proteomics* 2006, *5*, 172–181.

[9] Xu, C.-F., Wang, H., Li, D., Kong, X.-P., Neubert, T. A., Selective enrichment and fractionation of phosphopeptides from peptide mixtures by isoelectric focusing after methyl esterification. *Anal. Chem.* 2007, *79*, 2007–2014.

[10] Kochin, V., Imanishi, S. Y., Eriksson, J. E., Fast track to a phosphoprotein sketch – MALDI TOF characterization of TLC-based tryptic phosphopeptide maps at femtomolar detection sensitivity. *Proteomics* 2006, *6*, 5676–5682.

[11] Kane, S., Sano, H., Liu, S. C., Asara, J. M. *et al.*, A method to identify serine kinase substrates. Akt phosphorylates a novel adipocyte protein with a Rab GTPase-activating protein (GAP) domain. *J. Biol. Chem.* 2002, *277*, 22115–22118.

[12] Grønborg, M., Kristiansen, T. Z., Stensballe, A., Andersen, J. S. *et al.*, A mass spectrometry-based proteomic approach for identification of serine/threonine-phosphorylated proteins by enrichment with phospho-specific antibodies: Identification of a novel protein, Frigg, as a protein kinase A substrate. *Mol. Cell. Proteomics* 2002, *1*, 517–527.

[13] Ficarro, S. B., Salomon, A. R., Brill, L. M., Mason, D. E. *et al.*, Automated immobilized metal affinity chromatography/nano-liquid chromatography/electrospray ionization mass spectrometry platform for profiling protein phosphorylation sites. *Rapid Commun. Mass Spectrom.* 2005, *19*, 57–71.

[14] Ficarro, S. B., McCleland, M. l., Stukenberg, P. T., Burke, D. J. *et al.*, Phosphoproteome analysis by mass spectrometry and its application to Saccharomyces cerevisiae. *Nat. Biotechnol.* 2002, *20*, 301–305.

[15] Zhou, H., Tian, R., Ye, M., Xu, S. *et al.*, Highly specific enrichment of phosphopeptides by zirconium dioxide nanoparticles for phosphoproteome analysis. *Electrophoresis* 2007, *28*, 2201–2215.

[16] Yu, L.-R., Zhu, Z., Chan, K. C., Issaq, H. J. *et al.*, Improved TiO_2 enrichment of phosphopeptides from HeLa cells and high confident phosphopeptide identification by cross-validation of MS/MS and MS/MS/MS spectra. *J. Proteome Res.* 2007, *6*, 4150–4162.

[17] Kweon, H. K., Hakansson, K., Selective zirconium dioxide-based enrichment of phosphorylated peptides for mass spectrometric analysis. *Anal. Chem.* 2006, *78*, 1743–1749.

[18] Cantin, G. T., Shock, T. R., Park, S. K., Madhani, H. D., Yates, J. R., III, Optimizing TiO2-based phosphopeptide enrichment for automated multidimensional liquid chromatography coupled to tandem mass spectrometry. *Anal. Chem.* 2007, *79*, 4666–4673.

[19] Thingholm, T. E., Jensen, O. N., Robinson, P. J., Larsen, M. R., SIMAC – A phosphoproteomic strategy for the rapid separation of mono-phosphorylated from multiply phosphorylated peptides. *Mol. Cell. Proteomics* 2008, *7*, 661–671.

[20] Larsen, M. R., Thingholm, T. E., Jensen, O. N., Roepstorff, P., Jorgensen, T. J., Highly selective enrichment of phosphorylated peptides from peptide mixtures using titanium dioxide microcolumns. *Mol. Cell. Proteomics* 2005, *4*, 873–886.

[21] Jensen, S. S., Larsen, M. R., Evaluation of the impact of some experimental procedures on different phosphopeptide enrichment techniques. *Rapid Commun. Mass Spectrom.* 2007, *21*, 3635–3645.

[22] Mazanek, M., Mitulovic, G., Herzog, F., Stingl, C. *et al.*, Titanium dioxide as a chemo-affinity solid phase in offline phosphopeptide chromatography prior to HPLC-MS/MS analysis. *Nat. Protoc.* 2007, *2*, 1059–1069.

[23] Pinkse, M. W., Uitto, P. M., Hilhorst, M. J., Ooms, B., Heck, A. J., 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*, 3933–3943.

[24] Kalkum, M., Lyon, G. J., Chait, B. T., Detection of secreted peptides by using hypothesis-driven multistage mass spectrometry. *Proc. Natl. Acad. Sci.* 2003, *100*, 2795–2800.

[25] Chang, E. J., Archambault, V., McLachlin, D. T., Krutchinsky, A. N., Chait, B. T., Analysis of protein phosphorylation by hypothesis-driven multiple-stage mass spectrometry. *Anal. Chem.* 2004, *76*, 4472–4483.

[26] Beausoleil, S. A., Jedrychowsky, M., Schwartz, D., Elias, J. E. *et al.*, Large-scale characterization of HeLa cell nuclear phosphoproteins. *Proc. Natl. Acad. Sci.* 2004, *101*, 12130–12135.

[27] Olsen, J. V., Macek, B., Lange, O., Makarov, A. *et al.*, Higher-energy C-trap dissociation for peptide modification analysis. *Nat. Methods* 2007, *4*, 709–712.

[28] Stingl, C., Ihling, C., Ammerer, G., Sinz, A., Mechtler, K., Application of different fragmentation techniques for the analysis of phosphopeptides using a hybrid linear ion trap-FTICR mass spectrometer. *Biochim. Biophys. Acta* 2006, *1764*, 1842–1852.

[29] Stensballe, A., Jensen, O. N., Olsen, J. V., Haselmann, K. F., Zubarev, R. A., Electron capture dissociation of singly and multiply phosphorylated peptides. *Rapid Commun. Mass Spectrom.* 2000, *14*, 1793–1800.

[30] Syka, J. E., Coon, J. J., Schroeder, M. J., Shabanowitz, J., Hunt, D. F., Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proc. Natl. Acad. Sci.* 2004, *101*, 9528–9533.

[31] Gunawardena, H. P., Emory, J. F., McLuckey, S. A., Phosphopeptide anion characterization via sequential charge inversion and electron-transfer dissociation. *Anal. Chem.* 2006, *78*, 3788–3793.

[32] Bennett, K. L., Stensballe, A., Podtelejnikov, A. V., Moniatte, M., Jensen, O. N., Phosphopeptide detection and sequencing by matrix-assisted laser desorption/ionization quadrupole time-of-flight tandem mass spectrometry. *J. Mass Spectrom.* 2002, *37*, 179–190.

[33] Annan, R. S., Carr, S. A., Phosphopeptide analysis by matrix-assisted laser desorption time-of-flight mass spectrometry. *Anal. Chem.* 1996, *68*, 3413–3421.

[34] Janek, K., Wenschuh, H., Bienert, M., Krause, E., Phosphopeptide analysis by positive and negative ion matrix-assisted laser desorption/ionization mass spectrometry. *Rapid Commun. Mass Spectrom.* 2001, *15*, 1593–1599.

[35] Nabetani, T., Miyazaki, K., Tabuse, Y., Tsugita, A., Analysis of acidic peptides with a matrix-assisted laser desorption/ionization mass spectrometry using positive and negative ion modes with additive monoammonium phosphate. *Proteomics* 2006, *6*, 4456–4465.

[36] Gao, Y., Wang, Y., A method to determine the ionization efficiency change of peptides caused by phosphorylation. *J. Am. Soc. Mass Spectrom.* 2007, *18*, 1973–1976.

[37] Raggiaschi, R., Lorenzetto, C., Diodato, E., Caricasole, A. *et al.*, Detection of phosphorylation patterns in rat cortical neurons by combining phosphatase treatment and DIGE technology. *Proteomics* 2006, *6*, 748–756.

[38] Graham, M. E., Ruma-Haynes, P., Capes-Davis, A. G., Dunn, J. M. *et al.*, Multisite phosphorylation of doublecortin by cyclin-dependent kinase 5. *Biochem. J.* 2004, *381*, 471–481.

[39] Chong, P. K., Gan, C. S., Pham, K. T., Wright, P. C., Isobaric tags for relative and absolute quantitation (iTRAQ) reproducibility: Implication of multiple injections. *J. Proteome Res.* 2006, *7*, 424–435.

[40] Sachon, E., Mohammed, S., Bache, N., Jensen, O. N., Phosphopeptide quantitation using amine-reactive isobaric tagging reagents and tandem mass spectrometry: Application to proteins isolated by gel electrophoresis. *Rapid Commun. In Mass Spectrom.* 2006, *20*, 1127–1134.

[41] Shilov, I. V., Seymore, S. L., Patel, A. A., Loboda, A. *et al.*, The Paragon Algorithm, a next generation search engine that uses sequence temperature values and feature probabilities to identify peptides from tandem mass spectra. *Mol. Cell. Proteomics* 2007, *6*, 1638–1655.

[42] Kjellström, S., Jensen, O. N., Phosphoric acid as a matrix additive for MALDI MS analysis of phosphopeptides and phosphoproteins. *Anal. Chem.* 2004, *76*, 5109.

[43] Laugensen, S., Roepsdorff, P., Combination of two matrices results in improved performance of MALDI MS for peptide mass mapping and protein analysis. *J. Am. Soc. Mass Spectrom.* 2003, *14*, 992–1002.

[44] Yoo, C., Patwa, T. H., Kreunin, P., Miller, F. R. *et al.*, Comprehensive analysis of proteins of pH fractionated samples using monolithic LC/MS/MS, intact MW measurement and MALDI-QIT-TOF MS. *J. Mass Spectrom.* 2007, *42*, 312–334.

[45] Schley, C., Altmeyer, M. O., Swart, R., Müller, R., Huber, C. G., Proteome analysis of *Myxococcus xanthus* by off-line two-dimensional chromatographic separation using monolithic poly-(styrene-divinylbenzene) columns combined with ion-trap tandem mass spectrometry. *J. Proteome Res.* 2006, *5*, 2760–2768.

McsB Is a Protein Arginine Kinase That Phosphorylates and Inhibits the Heat-Shock Regulator CtsR

Jakob Fuhrmann, *et al.*

Science **324**, 1323 (2009);
DOI: 10.1126/science.1170088

Science (print ISSN 0036-8075; online ISSN 1095-9203) is published weekly, except the last week in December, by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005.

the profile from *Hoxd9* to *Hoxd1* was comparable to wild type. The robust gain in H3K4me3 marks over *Hoxd12* was not scored in older wild-type tail buds (fig. S2) and did not match any transcriptional activity, neither for *Hoxd12* nor for *Hoxd11* (Fig. 3B). In this case, both *Hoxd11* and *Hoxd12* were ready to be transcribed (*28*), yet they remained silent because they were moved away from the required enhancer sequence located telomeric to the breakpoint. In contrast, increased H3K4 trimethylation on the other side of the breakpoint (Fig. 3C) matched the premature activation of *Hoxd10*.

The DNA interval decorated by H3K27me3 marks in *inv* mutants was virtually identical to wild type (Fig. 3C), indicating that an integral cluster is not necessary to define the initial extent of the repressive domain; H3K27me3 marks were positioned over posterior genes even though these genes were disconnected from the rest of the cluster, thus ruling out the existence of a spreading mechanism sensu stricto for the implementation of this repression. In addition, the overall density of these marks on both sides of the break point was considerably below the wild-type situation (Fig. 3C). In the posterior half-cluster, H3K27me3 marks were distributed almost as in wild type over *Evx2* and *Hoxd13*, whereas a decrease was scored over the *Hoxd12* to *Hoxd11* intergenic region and 3′ to *Hoxd11* (Fig. 3C). In the anterior half-cluster, a similar reduction was detected at the *Hoxd10* locus, consistent with its premature activation and, to a lesser extent, over *Hoxd9* (Fig. 3C). This weakening in H3K27me3 signal over *Hoxd10* was not observed at the wild-type locus, even in older tail buds (fig. S2). The general decrease in H3K27 trimethylation around the break point suggests that a dense coverage of the *HoxD* cluster by this histone modification requires an intact clustered configuration. Whereas isolated parts of the gene cluster can be trimethylated at H3K27 independently of one another, these various parts may cooperate and synergize to mediate a dense pattern of methylation, potentially through local cis interactions.

These results shed light on the general regulatory strategy implemented by *Hox* gene loci during the earliest steps of mouse trunk development. Unlike in *Drosophila*, mammalian *Hox* gene loci appear refractory to transcription before transcription initiates, as indicated by high levels of H3K27me3 marks covering the *HoxD* locus early on. This likely reflects the necessity to prevent the premature activation of posterior genes at a time when anterior structures are being determined, which would be deleterious to the embryo. During gastrulation, this repression is counteracted by an activity progressing from the telomeric to the centromeric extremity of the cluster, illustrated by both an elevation of H3K4me3 level and the demethylation of H3K27me3. The region of transition between these two states of chromatin corresponds to the dynamic window wherein *Hoxd* genes become transcriptionally active. Alternatively, *Hox* genes could be activated from a persisting pool of nonexpressing stem cells. In this view, the chromatin modifications observed in our samples reflect the average of successive waves of transcriptional activation rather than a dynamic process occurring in the same cells. We do not favor this possibility because such a pool of *Hox*-negative cells would constitute a large fraction of the tissue sample, yet it has never been observed in gastrulating tail buds. Also, the nucleosomes of these stem cells would lack the repressive marks over the *HoxD* cluster, unlike in ESC. Finally, *Hox* genes are activated in cells already expressing more anterior combinations thereof.

We have shown that gene clustering is not necessary for the initial definition of the H3K27me3 landscape. However, clustering is required for a full repression to be consolidated and/or maintained over the cluster, which suggests a synergistic effect due to *Hox* genes' density. Likewise, whereas an integral cluster appears dispensable for selecting the sites of H3K4 trimethylation, gene clustering helps the coordination of this general transition in chromatin status because split clusters displayed premature H3K4me3 marks on either side of the breakpoint. Although the gain of H3K4me3 and the concurrent weakening of H3K27me3 at the mutant *Hoxd10* locus coincided with its early ectopic transcription, similar imbalances at the inverted *Hoxd11* and *Hoxd12* loci did not elicit the same transcriptional response. From this, we conclude that H3K4me3 chromatin modification is necessary but not sufficient for proper *Hox* gene transcriptional control and that remote enhancer sequences must have contributed to the maintenance of clustered organization during animal evolution.

References and Notes

1. D. Duboule, P. Dollé, *EMBO J.* **8**, 1497 (1989).
2. A. Graham, N. Papalopulu, R. Krumlauf, *Cell* **57**, 367 (1989).
3. R. Krumlauf, *Cell* **78**, 191 (1994).
4. J. C. Izpisúa-Belmonte, H. Falkenstein, P. Dollé, A. Renucci, D. Duboule, *EMBO J.* **10**, 2279 (1991).
5. D. Duboule, *Development* **134**, 2549 (2007).
6. M. Kmita, D. Duboule, *Science* **301**, 331 (2003).
7. P. Dollé, J. C. Izpisúa-Belmonte, H. Falkenstein, A. Renucci, D. Duboule, *Nature* **342**, 767 (1989).
8. T. Kondo, D. Duboule, *Cell* **97**, 407 (1999).
9. S. Chambeyron, W. A. Bickmore, *Genes Dev.* **18**, 1119 (2004).
10. Y. B. Schwartz, V. Pirrotta, *Nat. Rev. Genet.* **8**, 9 (2007).
11. J. Simon, A. Chiang, W. Bender, *Development* **114**, 493 (1992).
12. M. van Lohuizen, *Cell. Mol. Life Sci.* **54**, 71 (1998).
13. R. Cao *et al.*, *Science* **298**, 1039 (2002).
14. B. Czermin *et al.*, *Cell* **111**, 185 (2002).
15. A. Kuzmichev, K. Nishioka, H. Erdjument-Bromage, P. Tempst, D. Reinberg, *Genes Dev.* **16**, 2893 (2002).
16. J. Müller *et al.*, *Cell* **111**, 197 (2002).
17. T. Kouzarides, *Cell* **128**, 693 (2007).
18. B. E. Bernstein *et al.*, *Cell* **120**, 169 (2005).
19. A. P. Bracken, N. Dietrich, D. Pasini, K. H. Hansen, K. Helin, *Genes Dev.* **20**, 1123 (2006).
20. T. I. Lee *et al.*, *Cell* **125**, 301 (2006).
21. J. L. Rinn *et al.*, *Cell* **129**, 1311 (2007).
22. J. Deschamps, J. van Nes, *Development* **132**, 2931 (2005).
23. A. P. Lee, E. G. Koh, A. Tay, S. Brenner, B. Venkatesh, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 6994 (2006).
24. B. E. Bernstein *et al.*, *Cell* **125**, 315 (2006).
25. F. Spitz, C. Herkenne, M. A. Morris, D. Duboule, *Nat. Genet.* **37**, 889 (2005).
26. P. Tschopp, B. Tarchini, F. Spitz, J. Zakany, D. Duboule, *PLoS Genet.* **5**, e1000398 (2009).
27. T. Montavon, J. F. Le Garrec, M. Kerszberg, D. Duboule, *Genes Dev.* **22**, 346 (2008).
28. A. S. Chi, B. E. Bernstein, *Science* **323**, 220 (2009).

29. We thank T. Montavon, A. Puglisi, and D. Schübeler for advice; F. Chabaud for cell culture; and P. Descombes and members of the Genomics Platform for their help with tiling arrays. N.S. was supported by a European Molecular Biology Organization long-term fellowship. This work was funded by the University of Geneva, the Federal Institute of Technology in Lausanne, the Swiss National Research Fund, the National Research Center Frontiers in Genetics, and the European Union program Crescendo. Data and analysis are available for download from ArrayExpress (accession E-TABM-677).

Supporting Online Material

www.sciencemag.org/cgi/content/full/324/5932/1320/DC1
Materials and Methods
Figs. S1 to S3
References

27 January 2009; accepted 17 April 2009
10.1126/science.1171468

McsB Is a Protein Arginine Kinase That Phosphorylates and Inhibits the Heat-Shock Regulator CtsR

Jakob Fuhrmann,[1]* Andreas Schmidt,[2]* Silvia Spiess,[3] Anita Lehner,[1] Kürşad Turgay,[4] Karl Mechtler,[1,5] Emmanuelle Charpentier,[3,6] Tim Clausen[1]†

All living organisms face a variety of environmental stresses that cause the misfolding and aggregation of proteins. To eliminate damaged proteins, cells developed highly efficient stress response and protein quality control systems. We performed a biochemical and structural analysis of the bacterial CtsR/McsB stress response. The crystal structure of the CtsR repressor, in complex with DNA, pinpointed key residues important for high-affinity binding to the promoter regions of heat-shock genes. Moreover, biochemical characterization of McsB revealed that McsB specifically phosphorylates arginine residues in the DNA binding domain of CtsR, thereby impairing its function as a repressor of stress response genes. Identification of the CtsR/McsB arginine phospho-switch expands the repertoire of possible protein modifications involved in prokaryotic and eukaryotic transcriptional regulation.

One of the most intensely studied stress-response pathways is the bacterial heat-shock response. In the Gram-positive model organism *Bacillus subtilis*, the heat-shock response is mediated by a complex regulatory network (*1*, *2*) that is under control of at least four

major transcriptional regulators, including the alternative sigma factor σB (*3*), the two-component response regulator CssR (*4*), and the repressors HrcA (*5*) and CtsR (*6*, *7*). The latter factor, CtsR, controls the expression of genes encoding the HSP100/Clp chaperones and the protease ClpP (*6*, *8*) that constitute the core of the bacterial protein quality control system (*9*, *10*). CtsR is encoded by the first gene of the *clpC* operon that includes *ctsR*, *mcsA*, *mcsB*, and *clpC* (*6*). The dimeric repressor consists of an N-terminal domain with a helix-turn-helix (HTH) motif and a C-terminal domain of unknown function (*11*). In *B. subtilis*, CtsR represses transcription of the *clpC* heat shock operon and the *clpE* and *clpP* genes by binding specifically to a seven-nucleotide direct repeat sequence located upstream of the transcriptional start sites (*7*). Stress-induced transcription of the *clp* genes depends on the inactivation of CtsR by McsB (*12*). McsB shows pronounced homology to phosphagen kinases (PhKs) and has been reported to exhibit tyrosine kinase activity (*12*, *13*). Under normal growth conditions, McsB is captured and inhibited by ClpC. However, when bacteria are exposed to stress situations, the ClpC chaperone preferentially interacts with misfolded proteins. It is assumed that the released McsB can now form a complex with CtsR, thereby displacing it from DNA and inducing the expression of heat-shock genes (*14*). Alternatively, the phosphorylation of CtsR by McsB may be critical for the release of the repressor from DNA (*12*). To clarify and delineate the precise function of CtsR and McsB in the bacterial stress response, we screened the respective proteins from various Gram-positive bacteria for recombinant production and succeeded in reconstituting the *Bacillus stearothermophilus* CtsR/McsB system in vitro.

To uncover how McsB modulates the repressor activity of CtsR, we performed electrophoretic mobility shift assays (EMSAs) (Fig. 1A). Addition of CtsR to the 258–base pair (bp) *clpC* promoter containing three *ctsr* half sites led to a substantial band shift caused by the formation of a $CtsR_4$/DNA complex. Addition of McsB yielded two lower migrating bands that represent $CtsR_2$/DNA and free DNA. The McsB-dependent release of CtsR was observed only in the presence of Mg/adenosine triphosphate (ATP), whereas addition of EDTA or phosphatase counteracted the effect of McsB. Because no protein-protein interaction could be detected by native gel analysis or size exclusion chromatography, we speculated that McsB and CtsR interact transiently and that phosphorylation of CtsR by McsB abolishes its binding to DNA. To test this hypothesis, CtsR was incubated with McsB in the presence of ATP, and subsequently, phosphorylated CtsR (CtsR-P) was separated from nonphosphorylated CtsR by heparin affinity chromatography (Fig. 1B). Mass spectrometry (MS) analysis of CtsR-P revealed two protein species with either one or two phosphate moieties per protomer (Fig. 1B). In contrast to unmodified CtsR, the isolated CtsR-P cannot bind to its target DNA, as deduced from isothermal titration calorimetry (ITC) and gel-shift experiments (Fig. 1, C and D). Removal of the phosphate group by alkaline phosphatase fully restored the DNA binding capability of CtsR. Thus, phosphorylation of CtsR by McsB is sufficient to inhibit the repressor function of CtsR.

To understand how phosphorylation of CtsR affects DNA binding, we determined the crystal structure of CtsR bound to a 26-bp DNA derivative of the *clpC* promoter (table S1). The $CtsR_2$/DNA structure revealed that the CtsR protomer is composed of two distinct domains: (i) an N-terminal DNA binding domain that adopts the winged HTH fold (residues 2 to 72) and (ii) a C-terminal dimerization domain (residues 79 to 153) that consists of four α helices organized in a four-helix bundle (Fig. 2A). The DNA reading heads of the major and minor groove comprise the recognition helix of the HTH motif and the extended β-hairpin wing, respectively. Key residues for recognizing and binding the *ctsr* consensus sequence are indicated in Fig. 2A and fig. S3. After obtaining a molecular model of the CtsR/DNA complex, we used MS to pinpoint individual phosphorylation sites. Our initial analyses of "in-solution" and "in-gel" digested CtsR-P were not successful; thus, we attempted to sequence mono-phosphorylated CtsR in a "top-down" MS experiment (Fig. 2B). Purified CtsR-P was directly infused into the mass spectrometer and fragmented by different techniques including electron-capture dissociation (ECD), collisionally activated dissociation (CAD), and infrared multiphoton dissociation (IRMPD). Mapping of the resulting modified protein fragments to the CtsR amino acid sequence revealed that the phosphorylation sites reside in the winged HTH domain. Furthermore, the broad distribution of modified fragments pointed to the existence of product isoforms with different phosphorylation sites. The highest probability for a phosphorylation event was observed for the region

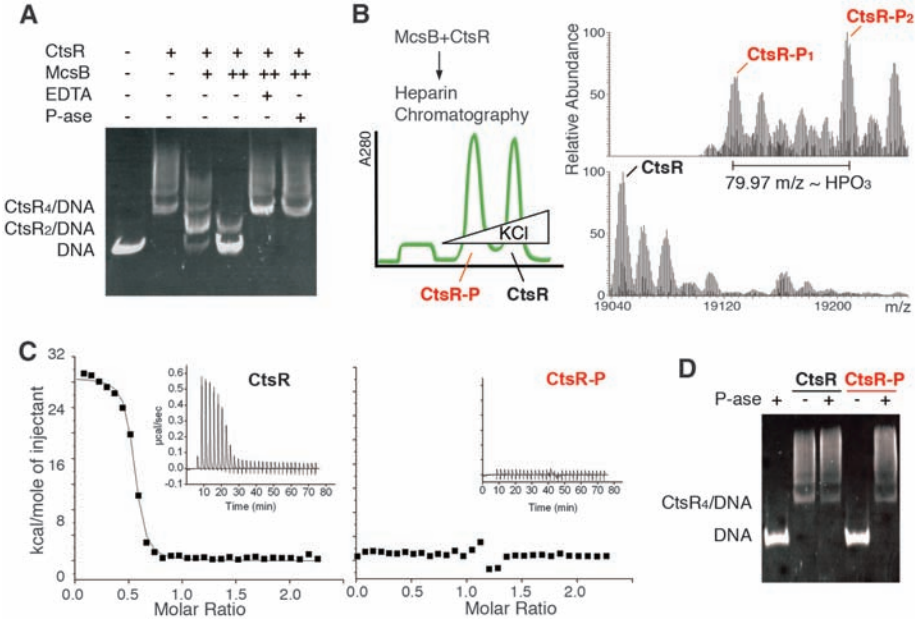


Fig. 1. Phosphorylation of CtsR impedes DNA binding. (**A**) EMSA analysis of the DNA binding capability of CtsR in the presence of McsB. CtsR was incubated with a *clpC* promoter fragment, McsB (+, 2 μM; ++, 8 μM), EDTA, and phosphatase (P-ase), as indicated. The promoter fragment, which was visualized by ethidium bromide staining of the native polyacrylamide gel, was either bound to one ($CtsR_2$/DNA) or two ($CtsR_4$/DNA) CtsR dimers. (**B**) Schematic presentation of the separation of CtsR-P from CtsR and McsB by heparin chromatography (left) and deconvoluted MS spectra of CtsR (average mass of 19047.2 daltons) and CtsR-P (19127.2 and 19207.1 daltons for mono- and di-phosphorylated isoforms, respectively) (right). (**C**) ITC analysis of $CtsR_2$/DNA complex formation. The 26-bp DNA duplex containing the *ctsr* box was injected into the sample cell containing either CtsR or CtsR-P (inset). The area under each peak was integrated and plotted against the molar ratio DNA/CtsR inside the sample cell. Thermodynamic values of CtsR/DNA complex formation are K_d = 22.2 ± 3.0 nM and n = 0.53 (reflecting the stoichiometry of bound DNA per CtsR protomer), whereas DNA binding of CtsR-P could not be detected by ITC. (**D**) EMSA analysis of the DNA binding capability of CtsR and CtsR-P, before and after phosphatase treatment.

[1]Research Institute of Molecular Pathology, Dr. Bohrgasse 7, A-1030 Vienna, Austria. [2]Christian Doppler Laboratory for Proteome Analysis, University of Vienna, Dr. Bohrgasse 3, A-1030 Vienna, Austria. [3]Max F. Perutz Laboratories, University of Vienna, Dr. Bohrgasse 9, A-1030 Vienna, Austria. [4]Institute for Biology–Microbiology, Freie Universität Berlin, Königin-Luisé-Str. 12-16, 14195 Berlin, Germany. [5]Institute for Molecular Biotechnology–IMBA, Dr. Bohrgasse 3, A-1030 Vienna, Austria. [6]The Laboratory for Molecular Infection Medicine Sweden, Umeå University, S-90187 Umeå, Sweden.

*These authors contributed equally to the work.
†To whom correspondence should be addressed. E-mail: clausen@imp.univie.ac.at

Tyr[55] to Asp[82], making up the β-hairpin wing (Fig. 2A). A lower, albeit still substantial, number of modified fragments matched the N-terminal segment from Ser[18] to Tyr[55].

To identify individual CtsR phosphorylation sites, we established a modified protocol for sample preparation and MS analysis (*15*). Most importantly, we implemented ECD and CAD fragmentation in two parallel MS/MS experiments. Only the ECD MS/MS spectrum of the phosphorylated CtsR peptide I_{57}VESKpRGGGGYIRIM$_{71}$ (*16*) allowed the unambiguous identification of Arg[62] as the site of modification (Fig. 2C). Both c- and z-fragment ion series unveiled a fragment of 236.067 daltons, reflecting the addition of a phosphate moiety (79.966 daltons) to an Arg residue (156.101 daltons). Moreover, CAD MS/MS of the I_{57}VESKpRGGGGYIRIM$_{71}$ phosphopeptide resulted in a discrete mass shift of 98 daltons, indicating the loss of phosphoric acid (fig. S1). This fragmentation behavior argues against a tyrosine kinase activity of McsB because phospho-tyrosine is stable upon CAD fragmentation (*17*). Further MS analysis led to the identification of two additional phosphorylation sites, Arg[28] and Arg[49] (fig. S2). Consistent with the results of the top-down approach, these amino acids are located within the winged HTH domain. Moreover, all Arg residues are strictly conserved in the CtsR protein family and play a crucial role in DNA binding, as predicted by our crystal structure. Arg[62] is a residue within the β wing and deeply invades the minor groove of the DNA duplex. In addition to undergoing extensive van der Waals contacts, the guanidinium group of Arg[62] forms hydrogen bonds with the DNA backbone and with one of the thymine pyrimidine carbonyls (Fig. 2A). Similarly, in the major groove of the CtsR consensus site, Arg[28] and Arg[49] bind to purine bases and coordinate the sugar-phosphate backbone, respectively (fig. S3).

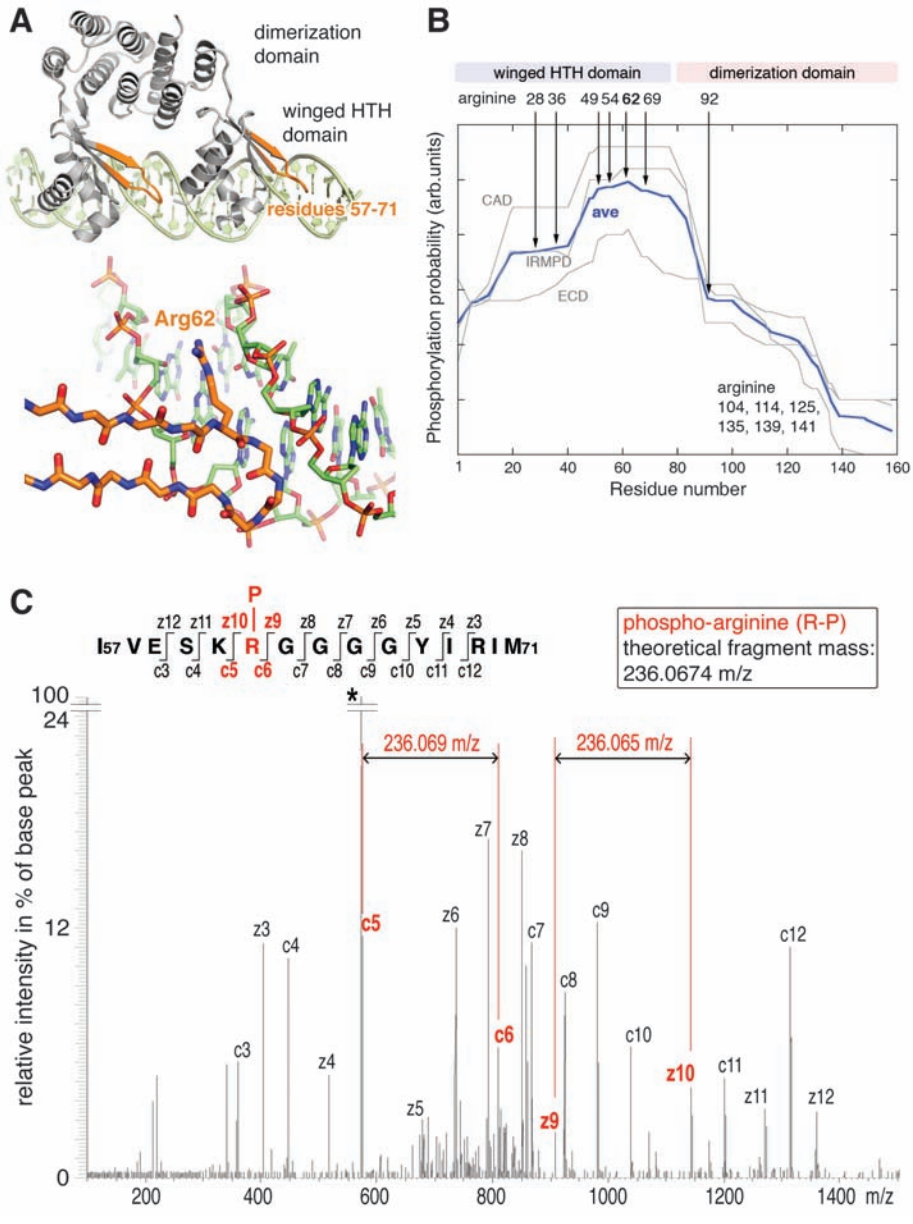


Fig. 2. Identification of arginine phosphorylation sites of CtsR. (**A**) Ribbon diagram showing the CtsR dimer (gray, with labeled domains) bound to the DNA direct repeat motif (green). The identified CtsR phosphopeptide I_{57}VESKpRGGGGYIRIM$_{71}$, which constitutes the β-hairpin of the winged HTH domain penetrating the DNA minor groove, is highlighted in orange. The lower panel illustrates the binding mode of Arg[62] (orange), the main phosphorylation site, at the floor of the DNA minor groove (green). (**B**) Phosphosite mapping with top-down MS. The mono-phosphorylated isoform of full-length CtsR was sequenced by three different fragmentation techniques. The blue line represents the average (ave) value of the three experimental setups and refers to the number of fragments additionally identified in CtsR-P, relative to unmodified CtsR. The residue with the highest phosphorylation score was Arg[62]. (**C**) ECD-MS/MS spectrum of the major phosphopeptide I_{57}VESKpRGGGGYIRIM$_{71}$ obtained after chymotryptic cleavage of phosphorylated CtsR. Individual fragments are labeled according to the c- or z-ion nomenclature. The characteristic mass difference of the phosphorylated Arg[62] is highlighted, and the threefold charged precursor ion is marked with an asterisk. m/z, mass/charge ratio.

To explore the functional relevance of the identified phosphosites, we conducted a mutational analysis of full-length CtsR by introducing various Arg-to-Lys mutations. Mutating the target sites in position 28, 49, and 62 (3RK) did not completely abolish, but did substantially reduce the phosphorylation of CtsR by McsB (Fig. 3A). Moreover, a mutant protein (8RK), in which the eight Arg residues located in the DNA binding region were replaced by Lys residues, was completely unsusceptible to McsB modification. Reintroduction of Arg[62] (7RK) markedly restored the phosphorylation potential. To study the direct effect of CtsR phosphorylation on DNA binding, we replaced Arg[62] by a phosphomimicking Glu residue. EMSA experiments clearly demonstrated that the Arg[62] → Glu[62] (R62E) mutant lost its capability to bind DNA (Fig. 3B), thus corroborating our finding that phosphorylation of CtsR alone is sufficient to inhibit its repressor activity. Conversely, replacing Arg[62] by Lys[62] did not alter the DNA binding ability of CtsR in band-shift assays. To test which state of CtsR is targeted by McsB, we incubated the kinase with DNA-bound and -unbound CtsR. Following the interaction with DNA over time revealed that McsB preferentially phosphorylates free CtsR, thereby preventing DNA complex formation (fig. S4). We conclude that the selective introduction of a negatively charged phosphate moiety functions as a molecular switch regulating DNA binding. Whereas the unphosphorylated CtsR binds with high affinity to its DNA consensus site and inhibits transcription of downstream genes, the McsB-phosphorylated CtsR repressor is not able to bind to DNA, thus allowing heat-shock gene expression.

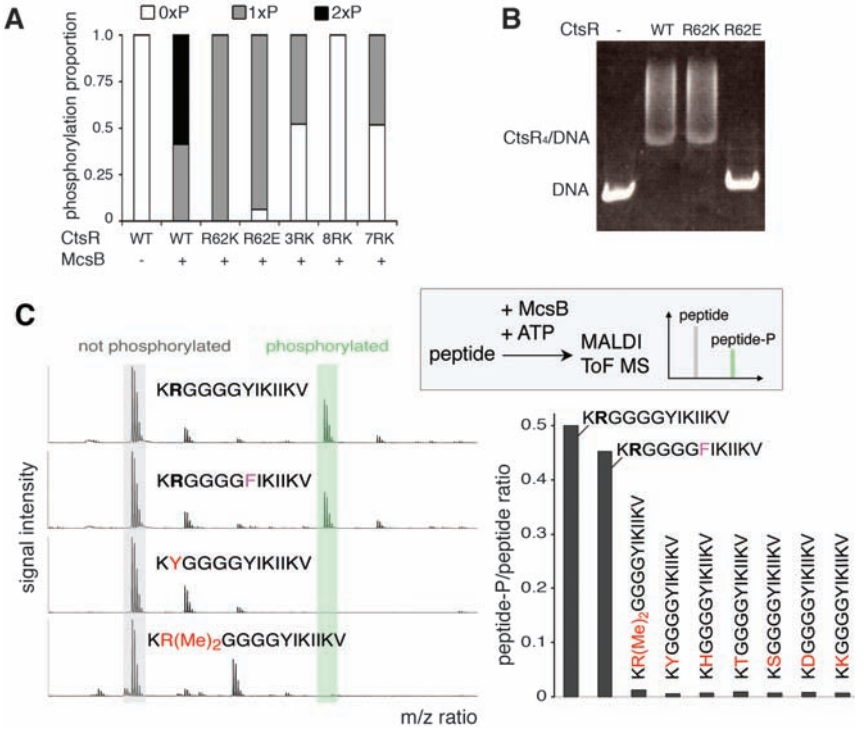


Fig. 3. Characterization of McsB-mediated arginine phosphorylation. (**A**) Phosphorylation level of CtsR Arg mutants analyzed by electrospray ionization–MS. The Arg-to-Lys mutants are 28/49/62 (3RK), 28/36/49/54/62/69/114/125 (8RK), and 28/36/49/54/69/114/125 (7RK). (**B**) DNA binding ability of different CtsR mutants in gel-shift assays. (**C**) Peptide phosphorylation assay (schematically shown in the inset). (Left) Matrix-assisted laser desorption/ionization–time-of-flight spectra of selected peptides after incubation with McsB. Non-phosphorylated and phosphorylated peptides are marked in gray and green, respectively. (Right) Effect of the exchange of Arg to other potential phospho-acceptor sites (shown in red) on the phosphorylation efficiency.

To verify our finding that McsB is a protein arginine kinase, we established an in vitro phosphorylation assay (Fig. 3C) using synthetic oligopeptides that resembled the CtsR sequence (residues 61 to 73). To avoid side effects during sample preparation that would preclude quantification of the phosphorylation reaction, we replaced one potential oxidation site (Met^{71}) and one arginine (Arg^{69}), yielding the 13-residue model substrate K_{61}RGGGGYIKIIKV$_{73}$. Systematic incorporation of potential phosphorylation sites (Tyr, Ser, Thr, His, Asp, and Lys) in position 62 revealed that only peptides with an Arg moiety are modified by McsB (Fig. 3C). Moreover, modification of the guanidinium group of Arg^{62} by asymmetric dimethylation prevented McsB-mediated modification. Additionally, we analyzed the purified phosphopeptide K_{61}RGGGGYIKIIKV$_{73}$ by ^{31}P nuclear magnetic resonance (NMR) spectroscopy (fig. S5). The chemical shift of about –2.4 parts per million (ppm) fits well to the measured NMR spectrum of free phospho-arginine (–3.0 ppm) (*18*), suggesting that the phosphate is attached via a phosphoramidate N-P linkage. Corresponding spectra of O-P linked phosphor compounds (as, for example, phospho-tyrosine, -serine, and -threonine) exhibit markedly higher chemical shifts of ~0.7 to 4.0 ppm (*18*). Thus, McsB is a protein kinase that acts exclusively on Arg residues, phosphorylating one of the amine nitrogens of the guanidinium group.

Phosphorylation of the free amino acid L-arginine by eukaryotic PhKs yields a chemically labile compound (*19*). We studied the CtsR/McsB system of a thermophilic organism living at ~55°C and thus explored the thermostability of a phosphorylated Arg residue present in a peptide context. For this purpose, we phosphorylated the K_{61}RGGGGYIKIIKV$_{73}$ peptide with McsB, incubated the purified phosphopeptide at different temperatures, and quantified the stability of the phosphorylation signal by high-performance liquid chromatography–MS analysis. The results clearly showed that peptide arginine phosphorylation is surprisingly stable up to 60°C. Dephosphorylation of the phosphopeptide occurred only at 95°C, with a half life $t_{1/2}$ of ~130 min (fig. S6). Therefore, phosphorylation of protein arginine residues should represent a relevant biological signal.

Sequence analysis indicated that the McsB protein arginine kinase exhibits no substantial homology to known Ser, Thr, Tyr, or His kinases. However, the catalytic domain of McsB is highly homologous to the catalytic domain of PhKs (*12*, *13*), which are involved in maintaining energy homeostasis but not in intermolecular signaling (*20*). Mutational analyses revealed that McsB and PhKs use a common mechanism to phosphorylate the terminal guanidinium group of substrates (fig. S7) (*14*). However, in contrast to PhKs where substrate specificity is primarily determined by the N-terminal domain, McsB harbors a distinct C-terminal domain that may redirect the substrate specificity from free Arg to protein-incorporated Arg residues.

McsB appears to be the founding member of a new class of protein kinases acting specifically on Arg residues. It should be noted that protein arginine phosphorylation has been reported previously (*21*). Remarkably, histone H3 was identified as a potential eukaryotic target (*22*), implying that Arg phosphorylation activity might be relevant for epigenetic regulation. However, these analyses failed to identify the corresponding kinase and obtained only indirect evidence for Arg modification. The thorough characterization of a protein arginine kinase presented in this work should provide the experimental tools to directly address the impact of Arg phosphorylation in prokaryotic and eukaryotic signaling pathways.

References and Notes

1. M. Hecker, W. Schumann, U. Volker, *Mol. Microbiol.* **19**, 417 (1996).
2. W. Schumann, M. Hecker, T. Msadek, in *Bacillus subtilis and Its Closest Relatives: From Genes to Cells*, A. L. Sonenshein, J. A. Hoch, R. Losick, Eds. (American Society for Microbiology Press, Washington, DC, 2002).
3. W. G. Haldenwang, R. Losick, *Nature* **282**, 256 (1979).
4. H. L. Hyyrylainen *et al.*, *Mol. Microbiol.* **41**, 1159 (2001).
5. A. Schulz, W. Schumann, *J. Bacteriol.* **178**, 1088 (1996).
6. E. Kruger, M. Hecker, *J. Bacteriol.* **180**, 6681 (1998).
7. I. Derre, G. Rapoport, T. Msadek, *Mol. Microbiol.* **31**, 117 (1999).
8. I. Derre, G. Rapoport, K. Devine, M. Rose, T. Msadek, *Mol. Microbiol.* **32**, 581 (1999).
9. S. Wickner, M. R. Maurizi, S. Gottesman, *Science* **286**, 1888 (1999).
10. R. T. Sauer *et al.*, *Cell* **119**, 9 (2004).
11. I. Derre, G. Rapoport, T. Msadek, *Mol. Microbiol.* **38**, 335 (2000).
12. E. Kruger, D. Zuhlke, E. Witt, H. Ludwig, M. Hecker, *EMBO J.* **20**, 852 (2001).
13. J. Kirstein, K. Turgay, *J. Mol. Microbiol. Biotechnol.* **9**, 182 (2005).
14. J. Kirstein, D. Zuhlke, U. Gerth, K. Turgay, M. Hecker, *EMBO J.* **24**, 3435 (2005).
15. Materials and methods are available as supporting material on *Science* Online. A detailed description of the MS approach is provided.
16. Single-letter abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.
17. A. Tholey, J. Reed, W. D. Lehmann, *J. Mass Spectrom.* **34**, 117 (1999).
18. T. L. James, *CRC Crit. Rev. Biochem.* **18**, 1 (1985).
19. A. Sickmann, H. E. Meyer, *Proteomics* **1**, 200 (2001).
20. W. R. Ellington, *Annu. Rev. Physiol.* **63**, 289 (2001).
21. H. R. Matthews, *Pharmacol. Ther.* **67**, 323 (1995).
22. B. T. Wakim, G. D. Aswad, *J. Biol. Chem.* **269**, 2722 (1994).

23. We thank D. Fiegen and D. Reinert (Boehringer Ingelheim) and the staff at Swiss Light Source for assistance with collecting synchrotron data, A. Carrieri for his contribution with the kinase assays, L. Becker for the NMR analysis of phosphopeptides, T. Krojer and D. Hellerschmied for their support in the structural analysis of CtsR, and C. Stingl and M. Mazanek for assisting MS analysis. The Research Institute of Molecular Pathology is funded by

Boehringer Ingelheim. J.F., S.S., E.C., and T.C. were supported by Wiener Wissenschafts, Forschungs und Technologiefonds; A.S. by the Christian-Doppler-Society; K.T. by the Deutsche Forschungsgemeinschaft; and T.C. by the European Molecular Biology Organization Young Investigator Program. This work was further supported by the Austrian Proteomics Platform (GEN-AU). The Protein Data Bank accession number for the $CtsR_2$/DNA complex is 3H0D.

Supporting Online Material
www.sciencemag.org/cgi/content/full/324/5932/1323/DC1
Materials and Methods
Figs. S1 to S8
Table S1
References

22 December 2008; accepted 13 April 2009
10.1126/science.1170088

Rhes, a Striatal Specific Protein, Mediates Mutant-Huntingtin Cytotoxicity

Srinivasa Subramaniam, Katherine M. Sixt, Roxanne Barrow, Solomon H. Snyder*

Huntington's disease (HD) is caused by a polyglutamine repeat in the protein huntingtin (Htt) with mutant Htt (mHtt) expressed throughout the body and similarly in all brain regions. Yet, HD neuropathology is largely restricted to the corpus striatum. We report that the small guanine nucleotide–binding protein Rhes, which is localized very selectively to the striatum, binds physiologically to mHtt. Using cultured cells, we found Rhes induces sumoylation of mHtt, which leads to cytotoxicity. Thus, Rhes-mHtt interactions can account for the localized neuropathology of HD.

Huntington's disease (HD), a genetically dominant neurodegenerative disorder, reflects expansion of a polyglutamine repeat in the protein huntingtin (Htt) (*1*). Mutant Htt (mHtt) occurs uniformly throughout the brain and peripheral tissues. Yet, HD is brain-specific with profound abnormal movements related to selective, gross degeneration of the corpus striatum and lesser damage to the cerebral cortex eliciting dementia (*2*, *3*). Molecular mechanisms causing mHtt cytotoxicity are unclear. mHtt forms protein aggregates, which may be neuroprotective with soluble mHtt linked to cytotoxicity (*4*–*7*). mHtt is sumoylated, which increases the soluble form of mHtt and elicits cytotoxicity and neurotoxicity in a *Drosophila* model of HD (*8*).

Rhes (Ras homolog enriched in striatum) is a small guanine nucleotide–binding protein (G protein) very selectively localized to the striatum (*9*). To determine whether Rhes binds to Htt, we overexpressed Rhes in HEK293 cells where it bound to both wild-type (wt) Htt and mHtt (Fig. 1A) (*10*). In conditionally immortalized Htt knock-in striatal neuronal cells (*11*), which lack endogenous Rhes (fig. S1C), overexpressed Rhes bound robustly to endogenous mHtt (Fig. 1B). In HD transgenic mice (*12*), endogenous striatal mHtt coprecipitated with Rhes (Fig. 1C). In the presence of purified Rhes and Htt, Rhes bound much more to mHtt than wtHtt protein (fig. S1A). Rhes did not bind to ataxin (fig. S1B), a polyglutamine-repeat protein involved in another neurodegenerative disorder, spinocerebellar ataxia.

To ascertain whether Rhes influences mHtt cytotoxicity, we used several cell lines. In

The Solomon H. Snyder Department of Neuroscience, Johns Hopkins University School of Medicine, 725 North Wolfe Street, Baltimore, MD 21205, USA.

*To whom correspondence should be addressed. E-mail: ssnyder@jhmi.edu

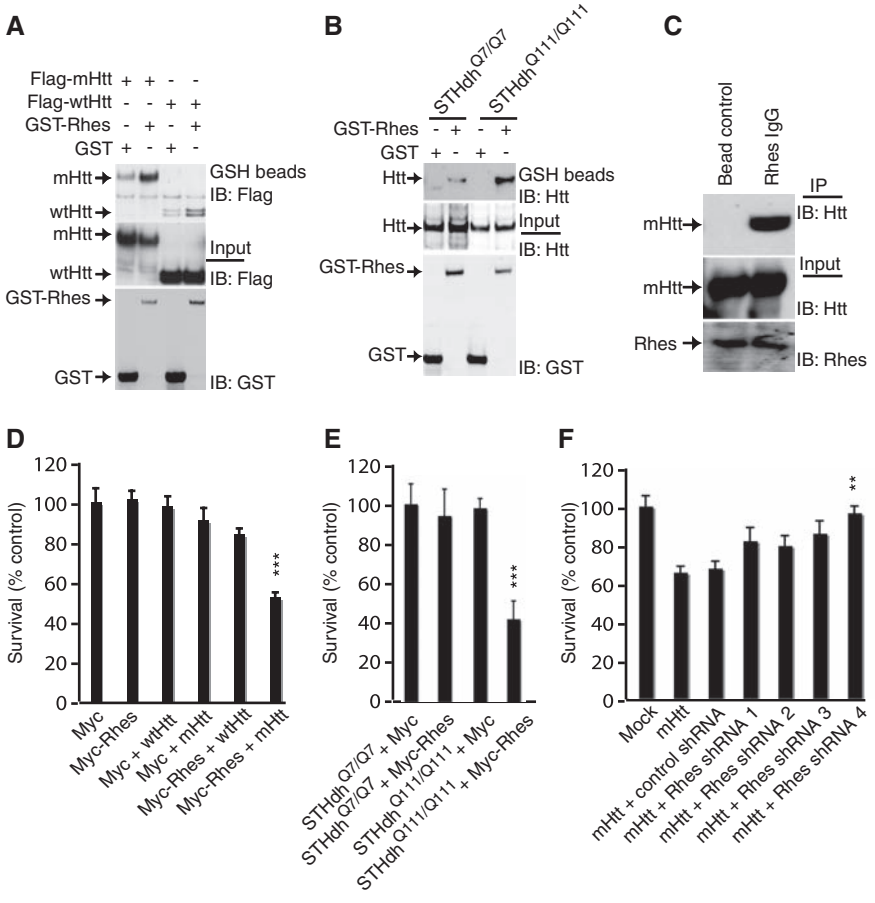


Fig. 1. Rhes binds Htt and affects cell survival. (**A**) Rhes interacts with N-terminal Htt. HEK293 cells were transfected with glutathione *S*-transferase (GST) or GST-Rhes together with Flag-tagged Htt or the N-terminal fragment containing 171 amino acids and 18 glutamines (wtHtt) or 82 glutamines (mHtt). After 48 hours, cell lysates were glutathione (GSH) precipitated and immunoblotted (IB) for Flag. (**B**) Rhes interacts with full-length Htt. Striatal cells expressing wild-type Htt (ST$Hdh^{Q7/Q7}$) or mutant Htt (ST$Hdh^{Q111/Q111}$) were transfected with GST or GST-Rhes. After 48 hours, cell lysates were GSH-precipitated and immunoblotted for Htt. Htt and GST inputs are shown. (**C**) Rhes interacts with mHtt in striatum. Striatum of transgenic mice expressing mHtt was lysed and immunoprecipitated with Rhes antibody or immunoglobulin IgG alone (bead control). Immunoprecipitates were probed with an N-terminal–specific Htt antibody (N-Htt). (**D**) Rhes reduces cell survival. HEK293 cells were transfected with Myc/Myc-Rhes and wtHtt–mHtt constructs. ***P < 0.005 versus mHtt alone. (**E**) Wild-type (ST$Hdh^{Q7/Q7}$) or mutant (ST$Hdh^{Q111/Q111}$) striatal cells were transfected with Myc/Myc-Rhes. ***P < 0.005 versus Myc. (**F**) Depletion of Rhes prevents PC12 cell death. Control short hairpin–mediated (shRNA) or Rhes shRNA 1 to 4 were cotransfected with mHtt. Only shRNA4 was significantly cytoprotective (**P < 0.01 versus control shRNA). After 48 hours, cell survival was measured by MTT.

Isotope-Labeled Cross-Linkers and Fourier Transform Ion Cyclotron Resonance Mass Spectrometry for Structural Analysis of a Protein/Peptide Complex

Christian Ihling,[†*] Andreas Schmidt,[†*] Stefan Kalkhof, and Daniela M. Schulz

Biotechnological-Biomedical Center, Faculty of Chemistry and Mineralogy, University of Leipzig, Leipzig, Germany

Christoph Stingl and Karl Mechtler

Institute of Molecular Pathology, Vienna, Austria

Michael Haack and Annette G. Beck-Sickinger

Department of Bioscience, Pharmacy, and Psychology, Institute of Biochemistry, University of Leipzig, Leipzig, Germany

Dermot M. F. Cooper

Department of Pharmacology, University of Cambridge, Cambridge, United Kingdom

Andrea Sinz

Biotechnological-Biomedical Center, Faculty of Chemistry and Mineralogy, University of Leipzig, Leipzig, Germany

For structural studies of proteins and their complexes, chemical cross-linking combined with mass spectrometry presents a promising strategy to obtain structural data of protein interfaces from low quantities of proteins within a short time. We explore the use of isotope-labeled cross-linkers in combination with Fourier transform ion cyclotron resonance (FTICR) mass spectrometry for a more efficient identification of cross-linker containing species. For our studies, we chose the calcium-independent complex between calmodulin and a 25-amino acid peptide from the C-terminal region of adenylyl cyclase 8 containing an "IQ-like motif." Cross-linking reactions between calmodulin and the peptide were performed in the absence of calcium using the amine-reactive, isotope-labeled (d_0 and d_4) cross-linkers BS^3 (*bis*[sulfosuccinimidyl]suberate) and BS^2G (*bis*[sulfosuccinimidyl]glutarate). Tryptic *in-gel* digestion of excised gel bands from covalently cross-linked complexes resulted in complicated peptide mixtures, which were analyzed by nano-HPLC/nano-ESI-FTICR mass spectrometry. In cases where more than one reactive functional group, e.g., amine groups of lysine residues, is present in a sequence stretch, MS/MS analysis is a prerequisite for unambiguously identifying the modified residues. MS/MS experiments revealed two lysine residues in the central α-helix of calmodulin as well as three lysine residues both in the C-terminal and N-terminal lobes of calmodulin to be cross-linked with one single lysine residue of the adenylyl cyclase 8 peptide. Further cross-linking studies will have to be conducted to propose a structural model for the calmodulin/peptide complex, which is formed in the absence of calcium. The combination of using isotope-labeled cross-linkers, determining the accurate mass of intact cross-linked products, and verifying the amino acid sequences of cross-linked species by MS/MS presents a convenient approach that offers the perspective to obtain structural data of protein assemblies within a few days. (J Am Soc Mass Spectrom 2006, 17, 1100–1113) © 2006 American Society for Mass Spectrometry

Tremendous progress has been made during the past few years in mapping interface regions between interacting proteins, which is crucial for understanding signal transduction pathways and regulation of proteins. Currently, X-ray crystallography and NMR spectroscopy are the primarily applied analytical methods to structurally characterize protein complexes. Yet, both techniques possess inherent insufficiencies: NMR spectroscopy requires rather large quantities of

Published online June 5, 2006

Address reprint requests to Dr. A. Sinz, Biotechnological-Biomedical Center, Faculty of Chemistry and Mineralogy, University of Leipzig, Linnéstrasse 3, D-04103 Leipzig, Germany. E-mail: sinz@chemie.uni-leipzig.de

* Both authors contributed equally to the present work.

† Also at the Institute of Molecular Pathology, Vienna, Austria.

1044-0305/06/$32.00
doi:10.1016/j.jasms.2006.04.020

Received September 22, 2005
Revised April 1, 2006
Accepted April 3, 2006

isotope-labeled (^{13}C and ^{15}N) pure protein in a specific solvent, and for X-ray studies the protein has to be crystallized. An alternative strategy consists in subjecting the complex constituents to chemical cross-linking reactions, followed by a mass spectrometric analysis of the created cross-linked products [1–3]. The procedure is fast, requires only low amounts of material, and offers the opportunity to gain insight into three-dimensional structures of protein complexes in solution. Matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) [4] and electrospray ionization (ESI) [5] mass spectrometry are the methods of choice for cross-linking studies because of their high sensitivity enabling rapid analysis of the complex mixtures obtained after enzymatic digestion of the cross-linking reaction mixtures. One major shortcoming of chemical cross-linking approaches, however, is the tremendous sample complexity attributable to the wide variety of cross-linked products that can be created during the cross-linking reaction. To circumvent this disadvantage, a number of different strategies have been developed that have in common to facilitate identification of cross-linker containing species, e.g., by employing isotope-labeled cross-linkers [6, 7], isotope-labeled proteins [8], cleavable cross-linkers [9], fluorogenic cross-linkers [10], or cross-linkers creating a characteristic marker ion during MS/MS analysis [11]. Another strategy to selectively enrich cross-linker containing species by using trifunctional cross-linkers containing a biotin moiety has been described recently [12–14].

In the present study, we utilize isotope-labeled cross-linkers, which are employed as 1:1 mixtures of nondeuterated and four-times deuterated derivatives. Thus, interpretation of mass spectra is greatly facilitated as cross-linker containing species are easily detected based on their characteristic isotope patterns. We are implementing FTICR mass spectrometry, which offers distinct advantages for the analysis of complex biological samples, such as excellent mass resolution, ultra-high mass measurement accuracy, and high sensitivity [15, 16] to evaluate its use in the area of cross-linking. Top-down approaches employing FTICR-MS for analyzing cross-linked products have been described by other groups [17–19] but, so far, that strategy has been exclusively employed to determine low-resolution three-dimensional structures of proteins from intramolecular cross-linking experiments. Our aim is to analyze the cross-linked species in a bottom-up approach, in which the cross-linked products are enzymatically or chemically digested before an FTICR-MS analysis is performed on the peptide mixtures [20, 21]. Therefore, the approach has the potential to be applied to investigate even large protein assemblies—given that the cross-linked species can easily be discriminated from underivatized peptides, e.g., by isotope-labeling. In the present study, we are making use of an FTICR mass spectrometer equipped with a linear ion trap in front of the ICR cell, which offers the possibility to conveniently acquire MS/MS data during LC/MS analysis.

As a model system to demonstrate the feasibility of our approach, we chose the complex between calmodulin and a peptide derived from the C-terminal region of adenylyl cyclase 8 (AC8) that is created in the absence of calcium. CaM is an acidic protein comprised of 148 amino acids, which belongs to the family of EF-hand proteins, and is found ubiquitously in animals, plants, fungi, and protozoa [22]. CaM/peptide complexes have been used frequently in our group for developing and improving chemical cross-linking strategies [12, 20, 21], as a wealth of structural data is available for the various CaM/target complexes (http://calcium.uhnres.utoronto.ca/ctdb/).

AC8 belongs to the family of adenylyl cyclases, which are important signalling proteins, possessing two distinct CaM binding sites [23]: one calcium-dependent binding site is located at its N-terminus, whereas a so-called "IQ-like motif" is located at its C-terminus. "IQ-like motifs" are well known to bind to CaM in a Ca^{2+}-dependent as well as a Ca^{2+}-independent manner [24]. Therefore, we chose the complex between CaM, and a peptide derived from the C-terminal region of AC8, to be well suited for demonstrating our strategy using isotope-labeled cross-linkers combined with high-resolution mass spectrometry.

A 25-amino acid peptide with amidated C- terminus comprising the "IQ-like motif" of AC8 was synthesized and subjected to chemical cross-linking with CaM using the two amine-reactive, isotope-labeled cross-linkers BS^3 (*bis*[sulfosuccinimidyl]suberate) and BS^2G (*bis*[sulfosuccinimidyl]glutarate) in the absence of calcium. In a previous publication, we had conducted structural studies on the CaM/AC8 peptide complex that is created in the presence of calcium using isotope-labeled cross-linkers and FTICR mass spectrometry [25]. However, in this earlier study, we did not perform MS/MS experiments, which are essential for pinpointing the cross-linked amino acids as well as for verification of the amino acid sequences of the created cross-linked products. In that respect, the present study extends our initial studies using isotope-labeled cross-linkers and high-resolution mass spectrometry for a structural analysis of protein complexes.

The combination of using isotope-labeled cross-linkers, determining the accurate mass of the intact cross-linked products, and verifying the amino acid sequences of cross-linked species by MS/MS presents a convenient approach that offers the perspective to obtain structural data of protein assemblies within a few days.

Experimental

Materials

Bovine brain CaM was obtained from Calbiochem (Schwalbach am Taunus, Germany) and used without further purification. The purities of CaM and the AC8 peptide were checked by HPLC, ESI-FTICR and

MALDI-TOF mass spectrometry, and SDS-PAGE. The isotope-labeled (d_0 and d_4) cross-linkers BS^3 and BS^2G were a generous gift from Pierce (Rockford, IL). Trypsin (sequencing grade) was obtained from Roche Diagnostics (Mannheim, Germany). LHRH (luteinizing hormone releasing hormone), 2,5-dihydroxybenzoic acid, and proteins for MALDI-TOF-MS calibration were purchased from Sigma (Taufkirchen, Germany). Chemicals were obtained from Sigma at the highest available purity. Nano-HPLC solvents were spectroscopic grade (Uvasol, VWR, Darmstadt, Germany). Water was purified with a Direct-Q5 water purification system (Millipore, Eschborn, Germany).

AC8 Peptide Synthesis

The AC8 peptide was synthesized by solid-phase peptide synthesis using the automated multiple peptide synthesizer Syro II (MultiSyntech, Witten, Germany). The peptide chain was synthesized by Fmoc/*tert*-butyl strategy applying a double coupling procedure with a ten-fold excess of amino acid, HOBt, and DIC in DMF for 40 min twice. N^{α}-Fmoc-protected amino acids and Rink amid resin (polymer) were purchased from Novabiochem (Läufelfingen, Switzerland), HOBt and DIC from Iris Biotech (Marktredwitz, Germany), and DMF from BioSolve (Valkenswaard, The Netherlands). Fmoc deprotection was accomplished by 40% piperidine in DMF for 3 min, followed by 20% piperidine in DMF for 7 min, and finally by 40% piperidine in DMF for 5 min. Removal of the amino acid side-chain protecting groups and cleavage of the peptide from the resin were accomplished simultaneously by using TFA/thioanisole/thiocresol (90/5/5, vol/vol) for 3 h. The deprotected peptides were allowed to precipitate for 15 min in ice-cold diethyl ether. The suspension was centrifuged at 5 °C, the diethyl ether was decanted, and the peptide was suspended in fresh diethyl ether. The peptide was washed by repeated centrifugation, decanting, and suspending the peptide in diethyl ether. Finally, the peptide was dissolved in *tert*-butyl alcohol/water (1:3, vol/vol) and lyophilized. Piperidine, *tert*-butyl alcohol as well as the chemicals for cleavage were purchased from Fluka (Taufkirchen, Germany).

Cross-Linking Reactions

For cross-linking experiments with the homobifunctional cross-linking reagents BS^3-d_0/d_4 and BS^2G-d_0/d_4, an aqueous CaM stock solution (1 mg/ml) was diluted with 20 mM Hepes buffer (pH 7.5) containing 1 mM EGTA to give solutions containing CaM at a concentration of 10.6 μM (volume 964 μl). After an incubation time of 10 min, 28 μl of an AC8 peptide solution (1 mg/ml) were added. The mixture was incubated at room temperature for 30 min. The cross-linking reactions were started by adding 8 μl of solutions containing either BS^3-d_0/d_4 or BS^2G-d_0/d_4 (each in a 1:1 mixture of d_0 and d_4 reagent at concentrations of 62.5 mM, 125 mM, or 250 mM in DMSO), thus yielding 50-, 100-, and 200-fold molar excesses of cross-linker over the protein/peptide concentration. One solution was prepared without adding cross-linker solution, but 8 μl of DMSO were added instead. The reaction mixtures were incubated at room temperature, and 200-μl aliquots were taken after 5 min, 15 min, 30 min, 60 min, and 120 min. The reactions were terminated by adding 20 μl of a 220 mM NH_4HCO_3 (final concentration 20 mM) solution to each aliquot. The solutions were desalted using Microcon-YM-3 filters (Millipore) and stored at −20 °C before MALDI-TOF-MS and SDS-PAGE were conducted.

SDS-PAGE and Enzymatic Proteolysis

Following separation of the reaction mixtures by one-dimensional SDS-PAGE (15%, Coomassie staining [26]), the bands of interest were excised and *in-gel* digested as described previously [27]. Trypsin (50 ng/μL) was used as digestion enzyme for all samples. Depending on the volume of the gel pieces, between 6 and 8 μl of enzyme solution were added. The gel pieces were covered with 50 mM NH_4HCO_3 solution and incubated at 37 °C for 16 h.

MALDI-TOF Mass Spectrometry

MALDI-TOF mass spectrometry of the cross-linking reaction mixtures was performed on a Voyager DE RP Biospectrometry Workstation (Applied Biosystems, Foster City, CA) equipped with a nitrogen laser (337 nm). The instrument was run in positive ionization mode and measurements were performed in the linear mode (detection range *m/z* 13,000 to 45,000) using 2,5-dihydroxybenzoic acid as matrix. A saturated matrix solution was prepared in 30% (vol/vol) acetonitrile, 69.9% (vol/vol) water, and 0.1% (vol/vol) TFA. Samples were prepared using the dried droplet method by spotting 0.5 μl of matrix solution and 0.5 μl of sample solution onto the target. Spectra from 100 to 300 laser shots were accumulated to one spectrum. The instrument was calibrated using cytochrome *c* ($[M + H]^+_{average}$ at *m/z* 12,361) and myoglobin ($[M + H]^+_{average}$ at *m/z* 16,952). Data acquisition and data processing were performed using the Voyager software version 5.1 and the Data Explorer software version 4.0 (Applied Biosystems).

Nano-HPLC/Nano-ESI-FTICR Mass Spectrometry

Peptide mixtures from enzymatic digests were separated by nano-HPLC on an Ultimate Nano-LC system (Dionex, Idstein, Germany) equipped with a Switchos II column switching module and a Famos Micro Autosampler with a 20-μl sample loop. Samples were injected by the autosampler and concentrated on a trapping column (PepMap, C18, 5 mm × 300 μm, 3 μm, 100 Å, Dionex, Idstein, Germany) with water containing

0.1% formic acid at flow rates of 20 μl/min. After 3 min for desalting, the peptides were eluted onto the separation column (PepMap, C18, 150 mm × 75 μm, 3 μm, 100 Å, Dionex), which had been equilibrated with 95% A (Solvent A: water containing 0.1% formic acid). Peptides were separated using the following gradient: 0–30 min: 5–50% B, 30–31 min: 50–95% B, 31–45 min: 95% B (Solvent B: acetonitrile containing 0.1% formic acid) at flow rates of 200 nl/min and detected based on their UV absorptions at 214 and 280 nm.

The nano-HPLC system was coupled *on-line* to an APEX II FTICR mass spectrometer equipped with a 7 tesla superconducting magnet (Bruker Daltonics, Billerica, MA) and a nano-electrospray ionization source (Agilent Technologies, Waldbronn, Germany). For nano-ESI-MS, coated fused-silica PicoTips (tip ID 8 μm, New Objective, Woburn, MA) were applied. The capillary voltage was set to -1400 V. Mass spectral data were acquired in the broadband mode over an *m/z* range of 400 to 2000 with 256 k data points, four scans were accumulated per spectrum, and 400 spectra were recorded for each LC/MS run. MS data acquisition was initialized with a trigger signal from the HPLC system 7 min after initiation of the LC gradient. Data were acquired over 34.5 min. Calibration of the instrument was performed with CID fragments (capillary exit voltage 200 V) of the LHRH peptide. Data acquisition and data processing were performed using the XMASS software, version 7.0.2 (Bruker Daltonics). Before Fourier transformation the time-domain signals were doubly zero filled, followed by apodization with a sine function. Four hundred single spectra were projected into one final mass spectrum using the "Projection" tool in the XMASS software [28]. The obtained spectra were internally recalibrated using autolytic peptide signals from trypsin. The XMASS software, version 7.0.2, was used for manual generation of peak lists containing the neutral monoisotopic peptide masses. An in-house software ("IsoFind") was employed to search for signals exhibiting a mass difference of 4 u and to calculate the ratios of the respective signal intensities.

MS/MS Analysis

Products of the cross-linking reactions conducted with BS^2G (gel bands D, 30, 60, 120 min reaction times, Figure 2), were additionally analyzed by MS/MS experiments. The peptide mixtures resulting from the tryptic *in-gel* digests were separated by C18-RP-chromatography on a nano-HPLC system (Ultimate 3000, Dionex; pre-column: PepMap, C18, 5 mm × 300 μm, 3 μm, 100 Å, Dionex; separation column: PepMap, C18, 150 mm × 75 μm, 3 μm, 100 Å, Dionex; solvents A: 5% acetonitrile, 0.1% formic acid in water, B: 80% acetonitrile, 0.08% formic acid in water) using a gradient from 0% B to 60% B in 90 min followed by isocratic elution with 90% B for 3 min. An LTQ-FT mass spectrometer (Thermo Electron, Bremen, Germany) with a 7 T magnet, equipped with a nano-ESI source (Proxeon Biosystems, Odense, Denmark; emitter: distal coated PicoTips, tip ID 15 μm, New Objective) was *on-line* coupled to the nano-HPLC system. MS data were acquired over 100 min in data-dependent MS2 mode: each high-resolution full scan (*m/z* 300 to 2000, resolution at *m/z* 400 was set to 100.000) in the ICR cell was followed by 10 product ion scans in the linear trap for the 10 most intense signals in the full-scan mass spectrum (isolation window 3 u). Dynamic exclusion (exclusion duration 20 s, exclusion window ±5 ppm) was enabled to allow detection of less abundant ions. Data analysis was performed using the SEQUEST search algorithm (BioWorks 3.2, Thermo Electron).

(a)

^{1}ADQLT	^{6}EEQIA	^{11}EFKEA	^{16}FSLFD	^{21}KDGDG	^{26}TITTK
^{31}ELGTV	^{36}MRSLG	^{41}QNPTE	^{46}AELQD	^{51}MINEV	^{56}DADGN
^{61}GTIDF	^{66}PEFLT	^{71}MMARK	^{76}MKDTD	^{81}SEEEI	^{86}REAFR
^{91}VFDKD	^{96}GNGYI	^{101}SAAEL	^{106}RHVMT	^{111}NLGEK	^{116}LTDEE
^{121}VDEMI	^{126}READI	^{131}DGDGQ	^{136}VNYEE	^{141}FVQMM	^{146}TAK

(b)

^{1}YSLAA ^{6}VVLGL ^{11}VQSLN ^{16}RQRQK ^{21}QLLNE-NH_2

Figure 1. Amino acid sequences of (**a**) CaM and (**b**) AC8 peptide. CaM is N-acetylated (Ac-Ala) and contains a trimethylated lysine in position 115 (the modified amino acids Ala-1 and Lysine 115 are underlined). The "IQ-like motif" in the AC8 peptide comprises amino acids 11–16 (VQSLNR underlined).

Identification of Cross-Linked Products

Cross-linked products were identified using the GPMAW (General Protein Mass Analysis for Windows) software, versions 6.01 and 6.2 (Lighthouse Data, Odense, Denmark) (available at: http://welcome.to/gpmaw) [29] and the ExPASy Proteomics tools in the Swiss-Prot Database (available at: www.expasy.ch). Proteolytic cleavages at modified amino acids, such as the trimethylated K115 in CaM as well as amino acids modified by cross-linking reagents were excluded. The N-terminus of CaM was excluded from possible cross-linking as it is acetylated (Figure 1).

Structure of Calcium-Free CaM

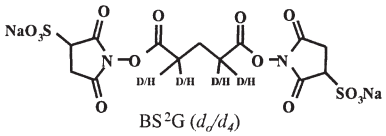
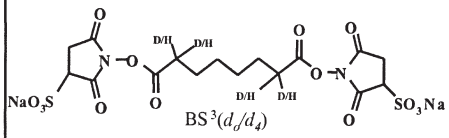
The NMR structure of apo Cam (Ca^{2+}-free CaM) is deposited in the RCSB Protein Data Bank (http://www.rcsb.org/pdb/) under the entry "1DMO". The structure was visualized using the VMD-XPLOR visualization package (available at: http://vmd-xplor.cit.nih.gov).

Results and Discussion

Characterization of CaM and AC8 Peptide

A prerequisite for successfully conducting cross-linking experiments consists in an in-depth characterization of

Table 1. Chemical structures and spacer lengths of the amine–reactive, homobifunctional, and isotope–labeled (d_0 and d_4) cross–linking reagents BS^3 and BS^2G

Cross-Linker	*Chemical formula*	*Spacer length*
BS^2G *Bis*(sulfosuccinimidyl)glutarate	 BS^2G (d_0/d_4)	7.7 Å
BS^3 *Bis*(sulfosuccinimidyl)suberate	 BS^3(d_0/d_4)	11.4 Å

the proteins regarding their post-translational modifications, splice variants, or amino acid exchanges. The characterization of CaM has been described in detail in previous publications of our group [20, 21]. Exact mass measurements, after spectral deconvolution, of the intact protein and peptide performed by ESI-FTICRMS yielded a most abundant mass of 16,790.921 u for CaM (simulated: 16,790.884 u, Δm = 0.037 u, 2.1 ppm) and a monoisotopic mass of 2838.650 u for the AC8 peptide (calculated: 2838.619 u, Δm = 0.031 u, 10.9 ppm). The amino acid sequences of CaM and AC8 peptide are presented in Figure 1.

Cross-Linking Reactions

Cross-linking experiments between CaM and the AC8 peptide were conducted as described in the Experimental section. The homobifunctional, amine-reactive cross-linking reagents BS^3-d_0/d_4 and BS^2G-d_0/d_4 (Table 1) were used for cross-linking experiments. Homobifunctional cross-linking reagents contain two identical functional groups on either side of the molecule that are separated by a spacer bridging a defined distance [30]. The homobifunctional sulfo-NHS esters BS^3 and BS^2G are highly reactive towards primary amine groups, but are also susceptible to hydrolysis. Upon cross-linking, BS^3 and BS^2G produce amide bond cross-linked molecules causing mass shifts of 138.077 u (BS^3-d_0) and 96.030 u (BS^2G-d_0), respectively, whereas peptides modified with partially hydrolyzed cross-linkers exhibit mass increases of 156.092 u (BS^3-d_0) and 114.045 u (BS^2G-d_0), respectively. 1:1 mixtures of nondeuterated (d_0) and four-times deuterated (d_4) reagents were employed to facilitate identification of cross-linker containing species based on their characteristic isotope doublet with a mass difference of 4.025 u in the deconvoluted ESI-FTICR mass spectra.

SDS-PAGE

After the cross-linking reactions, reaction mixtures were separated by one-dimensional SDS-PAGE and the gels were stained with Coomassie Brilliant Blue. CaM exhibited three bands on the gel (one intense and two faint bands, Figure 2). We assume that—despite using EGTA as calcium-complexing agent—a minor part of CaM molecules are present in different calcium-loaded states, exhibiting different mobilities on the gel. As shown for the cross-linking reaction mixture obtained with BS^2G-d_0/d_4 (Figure 2), an intense band at ~17 kDa probably represents intramolecularly cross-linked CaM (Band C, Figure 2). Bands of nonreacted CaM (Band B,

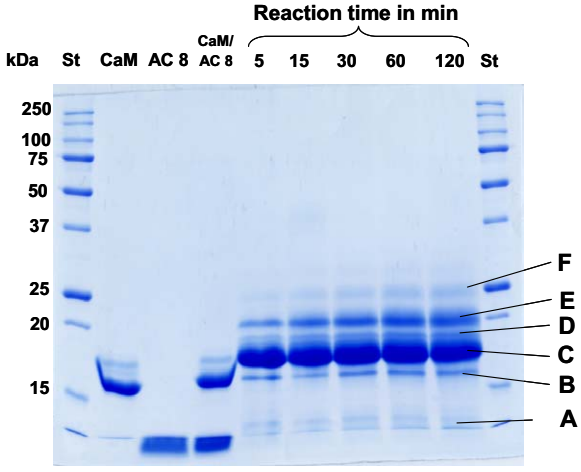


Figure 2. SDS-PAGE of CaM/AC8 peptide cross-linking reaction mixture with BS^2G-d_0/d_4 (100-fold excess, 1 mM EGTA), reaction times 5 to 120 min. St: molecular weight standard. Band A: intramolecularly cross-linked AC8 peptide; Band B: CaM; Band C: intramolecularly cross-linked CaM; Bands D and E: CaM/AC8 peptide complexes (1:1); Band F: CaM/AC8 peptide (1:2) complex.

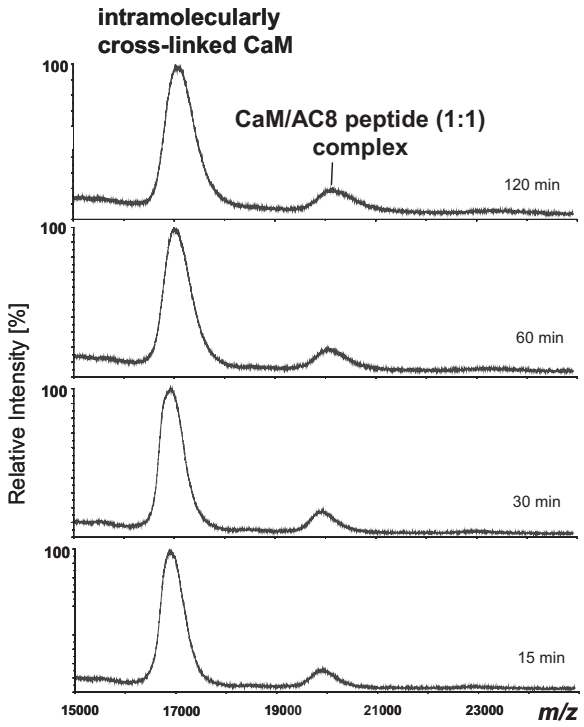


Figure 3. MALDI-TOF mass spectra of CaM/AC8 peptide cross-linking reaction mixtures obtained with BS^2G (100-fold excess over protein / peptide concentration), 1 mM EGTA.

Figure 2) and intramolecularly cross-linked AC 8 peptide (Band A, Figure 2) are also visible on the gel. Two bands migrating at ~19 and 20 kDa were assigned to CaM/AC8 peptide (1:1) complexes (Bands D and E) exhibiting different mobilities on the gel due to the presence of differently cross-linked species (Figure 2). Also, formation of a CaM/AC8 peptide (1:2) complex was observed (Band F, Figure 2) as was visualized by the presence of a faint gel band migrating at ~25 kDa. Aggregation of proteins caused by excessive cross-linking was not observed for any of the cross-linking reagents, as no gel bands were visible in the higher mass range.

MALDI-TOF Mass Spectrometry

MALDI-TOF-MS was employed to estimate the extent of chemical cross-linking over the course of the cross-linking reaction. For BS^2G, formation of a CaM/ AC8 peptide (1:1) complex at m/z ~19,900 was observed with the signal intensity slightly increasing with increasing reaction time (Figure 3). From the ~300 u-mass shift of the CaM signal (signal at m/z ~17,100 instead of m/z 16,790 for noncross-linked CaM), the incorporation of three BS^2G molecules per CaM molecule was estimated (Figure 3). The broadness of the signal for intramolecularly cross-linked CaM gives a first hint on the variety of intramolecular modifications that have been created in the course of the cross-linking reaction.

Analysis of Cross-Linked Products

Following SDS-PAGE separation of the cross-linking reaction mixtures, bands of the cross-linked CaM/AC8 peptide (1:1) complexes (Bands D and E, Figure 2) as well as bands originating from intramolecularly cross-linked CaM (Band C, Figure 2) were excised from the gel and subjected to enzymatic *in-gel* digestion with trypsin. CaM/AC8 peptide (1:2) complexes that were observed in SDS-PAGE (Band F, Figure 2) were assumed to be artifacts and were not investigated further. Now, we cannot rule out that CaM/AC8 peptide (1:2) complexes are also created in vivo. The band corresponding to intramolecularly cross-linked CaM could either originate from unbound CaM or, alternatively, it represents the CaM conformation within the complex, which dissociates during SDS-PAGE separation as it is not covalently attached to its binding partner by a cross-linker molecule. The intricate peptide mixtures resulting from *in-gel* digestion were analyzed by nano-HPLC/nano-ESI-FTICR mass spectrometry with respect to the presence of intra- and intermolecular cross-linked products as well as peptides, which are modified by a partially hydrolyzed cross-linker.

Cross-Linked Products Obtained with BS^2G-d_0/d_4

BS^2G was employed for conducting cross-linking reactions as a 1:1 mixture with its deuterated derivative (d_4) (Table 1). Thus, an additional criterion for the identification of cross-linked products is introduced, as every species containing one cross-linker molecule should exhibit a doublet with a distance of 4.025 u in the deconvoluted ESI-FTICR mass spectra. One intermolecular cross-linked product was identified during our first set of experiments from band E (Figure 2), in which amino acid sequence 75–86 (central α-helix) of CaM is connected with amino acids 17–25 of AC8 peptide (Table 2). Lys-20 is the exclusive reaction site in the respective AC8 sequence that has reacted with Lys-75 or Lys-77 of CaM. For clarification, which of the lysine residues at position 75 or 77 in CaM had reacted, MS/MS experiments were conducted (see below). The deconvoluted ESI-FTICR mass spectrum, from which this cross-linked product was identified, is presented in Figure 4 with the inset showing the magnified signal of the identified intermolecular cross-linked product. The doublet of isotope patterns with a mass difference of 4.025 u caused by using a 1:1 mixture of d_0- and d_4-cross-linker greatly facilitates identification of cross-linker-containing species.

Cross-Linked Products Obtained with BS^3-d_0/d_4

As in the case of BS^2G, BS^3 was employed as 1:1 mixture with its fully deuterated derivative (d_4). With a 100-fold excess of BS^3-d_0/d_4 over protein/peptide, one intermolecular cross-linked product was identified between CaM and AC8 pointing to amino acids 75–86 of CaM

Table 2. Intermolecular cross-linked products between CaM and AC8 peptide as well as peptides that are modified by partially hydrolyzed cross–linker using BS^2G–d_0/d_4. Reaction times and molar excess of cross–linking reagent are indicated, labeling of gel bands according to Figure 2

Intermolecular Cross-Linked Products					
CaM	AC8 Peptide	Reacted amino acids	$[M + H]^+$ calc.	$[M + H]^+$ exp.	Δm [ppm]	Detected in SDS-PAGE band
75–86	17–25	K75, K77 (CaM)-K20 (AC8)	2747.363 (BS^2G-d_0)	2747.370 (BS^2G-d_0)	2	Band E, 100-fold excess, 60 min
75–86	17–25	K75, K77 (CaM)-K20 (AC8)	2751.388 (BS^2G-d_4)	2751.393 (BS^2G-d_4)	2	Band E, 100-fold excess, 60 min
Peptides Modified by Partially Hydrolyzed Cross-Linker						
CaM	AC8 Peptide	Reacted amino acids	$[M + H]^+$ calc.	$[M + H]^+$ exp.	Δm [ppm]	Detected in SDS-PAGE band
—	1–16	Y1	1817.016 (BS^2G-d_0)	1817.016 (BS^2G-d_0)	0	Band E, 50-fold excess, 60 min
—	1–16	Y1	1821.040 (BS^2G-d_4)	1821.034 (BS^2G-d_4)	4	Band E, 50-fold excess, 60 min
—	1–16	Y1	1817.016 (BS^2G-d_0)	1817.021 (BS^2G-d_0)	2	Band E, 100-fold excess, 30 min
—	1–16	Y1	1821.040 (BS^2G-d_4)	1821.044 (BS^2G-d_4)	2	Band E, 100-fold excess, 30 min

Figure 4. Deconvoluted ESI-FTICR mass spectrum of a tryptic digest of a cross-linking reaction mixture (CaM/AC8 peptide (1:1) complex created with 100-fold excess BS^2G-d_0/d_4, 1 mM EGTA, reaction time 60 min). Signals of the identified cross-linked product between amino acids 75–86 of CaM and amino acids 17–25 of AC8 peptide (Table 3) are shown enlarged. Please note that neutral monoisotopic masses are given. The signal intensities of the characteristic isotope doublet are not equal, as the BS^2G-d_0/d_4 mixture was not exactly 1:1. Signals originating from CaM peptides (diamond), AC8 peptides (triangle), cross-linked products (open star), and autolytic peptides of trypsin (filled circle) are indicated; n.a.: signal not assigned.

and the complete AC8 peptide being connected (Table 3). In the AC8 peptide, either the N terminus or, more likely, the ε-amino group of the lysine in position 20 has reacted with one of the lysine residues in position 75 or 77 of the CaM central α-helix.

Peptides Modified by Hydrolyzed Cross-Linker

In addition to inter- and intramolecularly cross-linked products, we observed a large number of peptides that had been modified by partially hydrolyzed BS^3 or BS^2G (Tables 2, 3, and 4).The modified amino acids do not yield direct information on the interface between CaM and AC8; however, they provide valuable information on the surface topology of the complex. As is readily visible from Figure 6, these residues are solvent-exposed in calcium-free CaM and thus, are quite easily modified by the cross-linker.

Identification of Cross-Linked Amino Acids by MS/MS

To unambiguously identify the amino acids that had actually been cross-linked, additional MS/MS experiments were conducted on the tryptic peptides of Band D (Figure 2), derived from cross-linking reaction mixtures that had been created using a 100-fold molar excess of BS^2G-d_0/d_4. MS/MS experiments yielded a wealth of information on the various cross-linked products that had been created in one single cross-linking reaction (Table 4). As far as intermolecular cross-linking

Table 3. Intermolecular cross–linked products between CaM and AC8 peptide as well as peptides that are modified by partially hydrolyzed cross–linker using BS^3G–d_0/d_4. Reaction times and molar excess of cross–linking reagent are indicated, labeling of gel bands according to Figure 2. MSO: oxidized methionine

Intermolecular cross-linked products						
CaM	AC8 peptide	Reacted amino acids	$[M + H]^+$ calc.	$[M + H]^+$ exp.	Δm [ppm]	Detected in SDS-PAGE band
75–86	1–25	K75, K77 (CaM)-Y1, K20 (AC8)	4457.382 (BS^3-d_0)	4457.426 (BS^3-d_0)	10	Band E, 100-fold excess, 60 min
75–86	1–25	K75, K77 (CaM)-Y1, K20 (AC8)	4461.407 (BS^3-d_4)	4461.411 (BS^3-d_4)	1	Band E, 100-fold excess, 60 min
Peptides Modified by Partially Hydrolyzed Cross-Linker						
CaM	AC8 Peptide	Reacted amino acids	$[M + H]^+$ calc.	$[M + H]^+$ exp.	Δm [ppm]	Detected in SDS-PAGE band
1–37	—	K13, K21, K30	4348.106 (BS^3-d_0)	4348.143 (BS^3-d_0)	1	Band C, 100-fold excess, 60 min
1–37	—	K13, K21, K30	4352.131 (BS^3-d_4)	4352.143 (BS^3-d_4)	3	Band C, 100-fold excess, 60 min
91–126	—	K94	4293.105 (BS^3-d_0)	4293.108 (BS^3-d_0)	1	Band C, 100-fold excess, 60 min
91–126	—	K94	4297.130 (BS^3-d_4)	4297.149 (BS^3-d_4)	5	Band C, 100-fold excess, 60 min
91–126	—	K94	4293.105 (BS^3-d_0)	4293.076 (BS^3-d_0)	7	Band E, 50-fold excess, 60 min
91–126	—	K94	4297.130 (BS^3-d_4)	4297.105 (BS^3-d_4)	6	Band E, 50-fold excess, 60 min
91–126	—	K94	4309.100, (MSO, BS^3-d_0)	4309.070 (MSO, BS^3-d_0)	7	Band E, 50-fold excess, 60 min
91–126	—	K94	4313.125 MSO (BS^3-d_4)	4313.093 (MSO, BS^3-d_4)	7	Band E, 50-fold excess, 60 min
91–126	—	K94	4309.100 (MSO, BS^3-d_0)	4309.108 (MSO, BS^3-d_0)	2	Band C, 100-fold excess, 60 min
91–126	—	K94	4313.125 (MSO, BS^3-d_4)	4313.110 (MSO, BS^3-d_4)	3	Band C, 100-fold excess, 60 min

Table 4. MS/MS data of cross–linked products between CaM and AC8 peptide as well as peptides that are modified by partially hydrolyzed cross–linker BS^2G–d_0/d_4. Reaction time: 30 min, 100–molar excess of BS^2G–d_0/d_4, 1mM EGTA, band D (see Figure 2). Species containing two cross–linker molecules (one cross–link plus one partially hydrolyzed cross–linker) are printed in shaded areas

Cross-linked sequences		Cross-linked residue		Theoretical precursor mass	Experimental precursor mass	Cross-linker BS^2G	Identified fragment ions			
CaM	AC8 peptide	CaM	AC8 peptide	$[M + H]^+$	$[M + H]^+$		b ion series CaM	AC 8 peptide	y ion series CaM	AC8 peptide
14–30	19–25	K21	K20	2811.404	2811.412	d0	3–9,11,12,14,16	—	3–8,11–16	—
				2815.429	2815.434	d4	3–7,9–16	—	3–12,14–16	—
22–30	19–25	K30	K20	1873.950	1873.939	d0	3,4,6–8	—	1–3,5–8	—
				1877.975	1877.934	d4	3,4,6,8	—	1–3,5–8	—
75–86	19–25	K77	K20	2447.208	2447.204	d0	1–4,6–11	—	2–11	—
				2451.233	2451.228	d4	2–11	—	2–11	—
					Oxidation at M76					
75–86	19–25	K77	K20	2463.207	2463.212	d0	4,6,8–11	2,4–6	2–11	2–6
				2467.232	2467.239	d4	3–11	—	2–10	—
					Ammonia loss and oxidation at M76					
75–86	19–25	K77	K20	2446.180	2446.174	d0	2–11	—	2–11	—
				2450.205	2450.210	d4	2–11	—	2–10	—
					Two cross-linkers					
75–86	19–25	K75	K20	2561.240	2561.244	2x d0	1–8,10,11	—	2–11	—
75–86	19–25	K77	K20	2561.240	2561.244	2x d0	1–3,5,6,8–11	—	2–11	—
75–86	19–25	K75	K20	2565.265	2565.272	d0 K75,d4 K77	1–11	—	2–11	—
75–86	19–25	K75	K20	2565.256	2565.272	d4 K75,d0 K77	1–11	—	2–11	—
75–86	19–25	K77	K20	2565.256	2565.272	d4 K75,d0 K77	2–11	—	2–10	—
75–86	19–25	K77	K20	2565.256	2565.272	d0 K75,d4 K77	1–11	—	2–11	—
75–86	19–25	K75	K20	2569.281	2569.280	2x d4	1,3–10	—	2,4–9,11	—
75–86	19–25	K77	K20	2569.281	2569.280	2x d4	2–8,10,11	—	2–11	—
76–86	19–25	K77	K20	2319.113	2319.115	d0	2–8,10	—	2–10	—
				2323.238	2323.157	d4	2–4,6,7,9,10	—	2–10	—
					Oxidation of M76					
76–86	19–25	K77	K20	2335.112	2335.118	d0	2–8	—	3–10	—
				2339.137	2339.143	d4	2–8	—	2–10	—
91–106	19–25	K94	K20	2721.384	2721.397	d0	3–5,7–13,15	—	2–14	—
				2725.409	2725.411	d4	3–15		2–15	—
					Ammonia loss					
91–106	19–25	K94	K20	2704.357	2704.367	d0	4–15	—	2–12,15	—
				2708.382	2708.387	d4	4–15	—	2–12	—
					Peptides modified by hydrolyzed cross-linker					
CaM		modified residue		1958.923	1958.929	d0	2–16	—	2–16	—
14–30		K21		1962.948	1962.951	d4	2–16	—	2–16	—

(*Continued*)

Table 4. Continued

Modified sequence		Modified residue		Theoretical precursor mass	Experimental precursor mass	Cross-linker BS^2G	Identified fragment ions			
CaM	AC8 peptide	CaM	AC8 peptide	$[M + H]^+$	$[M + H]^+$		b ion series CaM	AC 8 peptide	y ion series CaM	AC8 peptide
14–37	—	K21	—	2745.329	2745.322	d0	3–10,12–14,16–23	—	16,18,19,21	—
				2749.354	2749.361	d4	3–23	—	2–7,9–15,17–21,23	—
—	—	K30	—	2745.329	2745.322	d0	3–7,9–11,13–23	—	2–11,13–16,18,19,21	—
				2749.354	2749.361	d4	3–7,9–23	—	2–8,10–21,23	—
					Ammonia loss					
14–37	—	K21	—	2728.302	2728.315	d0	4–10,11,13–23	—	2–23	—
				2732.327	2732.347	d4	6,6,7,9–23	—	2–10,12,14–23	—
—	—	K30	—	2728.302	2728.315	d0	4–12,14,15,17–23	—	2–23	—
				2732.327	2732.347	d4	3,6–10,12–23	—	2–7,9,11–15,16,18–23	—
					Oxidation of M36					
14–37	—	K21	—	2761.328	2761.334	d0	3,5–10,12–23	—	2–7,9–12,15–22	—
				2765.353	2765.351	d4	3–10,12,14–23	—	2–7,10,12,14,17–23	—
—	—	K30	—	2761.328	2761.334	d0	3,5–23	—	2–12,14,16–18,20,21	—
				2765.353	2765.351	d4	3–11,13–23	—	2–7,9–14,16–18,20–22	—
					Oxidation of M36 and ammonia loss					
14–37	—	K21	—	2744.301	2744.321	d0	3–6,8–10,12,14,16–23	—	2–7,10–21,23	—
				2748.326	2748.344	d4	3–6,9,10,12,15–23	—	2–7,9–22	—
—	—	K30	—	2744.301	2744.321	d0	3–23	—	2–13,15–23	—
				2748.326	2748.344	d4	3–6,8–14,17–23	—	2–8,10,11,13–22	—
					Oxidation of M36, two cross-linkers					
14–37	—	K21, K30	—	2875.360	2875.334	2x d0	5,7,9–16,2		9–17,19	—
				2879.385	2879.386	d0 K21,d4 K30	3–10,12–23		2–21	—
				2879.385	2879.386	d4 K30,d0 K21	3–23		2–13,15,17–21	—
				2883.410	2883.419	2x d4	3–9,11–23		2–15,17–23	—
22–37	—	K30	—	1807.874	1807.880	d0	3–15		2–11,13,14	—
				1811.899	1811.904	d4	3–15		2–15	—
					Oxidation of M36					
22–37	—	K30	—	1823.873	1823.874	d0	3–15	—	2–15	—
				1827.898	1827.896	d4	4–15	—	2–15	—
75–86	—	K75	—	1594.727	1594.730	d0	1–11	—	2–11	—
				1598.752	1598.755	d4	1–11	—	2–11	—
—	—	K77	—	1594.727	1594.730	d0	2–11	—	2–11	—
				1598.752	1598.755	d4	2–11	—	2–11	—

(*Continued*)

Table 4. Continued

Modified sequences		Modified residue		Theoretical precursor mass	Experimental precursor mass	Cross-linker BS^2G	Identified fragment ions			
CaM	AC8 peptide	CaM	AC8 peptide	$[M + H]^+$	$[M + H]^+$		b ion series CaM	AC 8 peptide	y ion series CaM	AC8 peptide
					Oxidation of M76					
75–86	—	K75	—	1610.725	1610.726	d0	1–3,6–11	—	1–11	—
				1614.751	1614.751	d4	1–11	—	1–11	—
	—	K77	—	1610.726	1610.726	d0	2–11	—	2–11	—
				1614.751	1614.751	d4	2–11	—	2–11	—
					Two cross-linkers					
75–86	—	K75, K77	—	1708.758	1708.763	2x d0	2–8	—	1–6,8–11	—
				1712.783	1712.787	d0 K75,d4 K77	2–11	—	2–11	—
				1712.783	1712.789	d4 K75,d0 K77	2–11	—	2–11	—
				1716.808	1716.813	2x d4	2–11	—	2–11	—
					Oxidation of M76					
75–86	—	K75, K77	—	1724.757	1724.759	2x d0	2–11	—	2–11	—
				1728.782	1728.785	d0 K75,d4 K77	2–11	—	2–11	—
				1728.782	1728.785	d4 K75,d0 K77	2–11	—	2–11	—
				1732.807	1732.811	2x d4	2–11	—	2–11	—
76–86	—	K77	—	1466.632	1466.632	d0	2–10	—	2–10	—
				1470.657	1470.660	d4	2–10	—	2–10	—
					Oxidation of M76					
76–86	—	K77	—	1482.631	1482.634	d0	2–10	—	2–10	—
				1486.656	1482.653	d4	2–10	—	2–10	—
					Oxidation of M76					
76–90	—	K77	—	1985.880	1985.883	d0	2–9,11,12,14	—	2–14	—
				1989.905	1989.901	d4	2,3,5,7–10,12–14	—	3,5–8,10–14	—
91–106	—	K94	—	1868.903	1868.910	d0	3–6,9–15	—	2–14	—
				1872.928	1872.932	d4	3–7,9–15		2–15	—
—	19–25	—	K20	985.531	985.533	d0	—	2–6	—	1–5
				989.556	989.558	d4	—	2–6	—	1–6
					Ammonia loss					
—	19–25	—	K20	968.502	968.506	d0	—	2,4–6	—	1–6
				972.527	972.532	d4	—	2–6	—	2–6

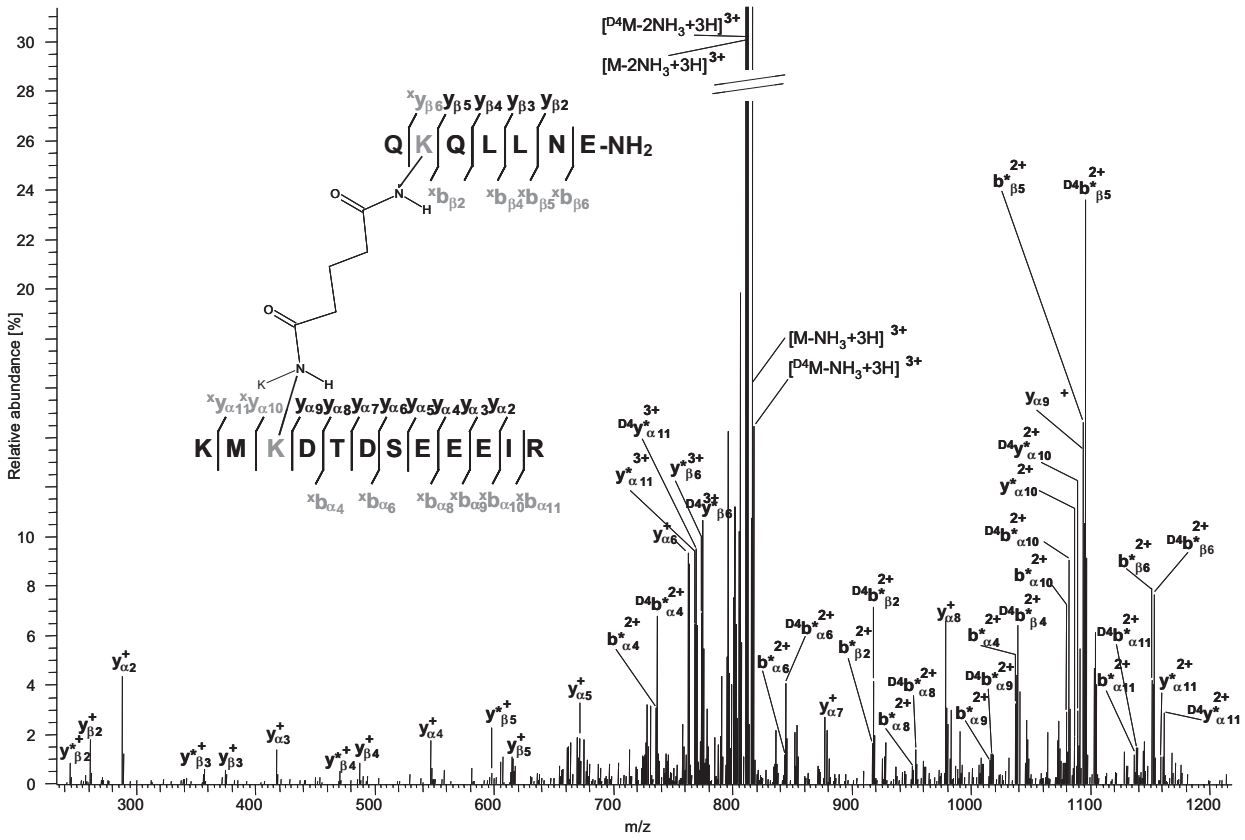


Figure 5. Product ion spectrum of the signal at *m/z* 823.08 ($[M + 3H]^{3+}$) corresponding to a cross-linked product between CaM (amino acids 75 to 86, M-76 oxidized, lower sequence) and AC8 peptide (amino acids 19 to 25, upper sequence) cross-linked with 100-fold excess BS^2G-d_4, 1 mM EGTA, reaction time 30 min. Isolation of the precursor ion with an isolation window of 3 u also includes the cross-linked product with BS^2G-d_0. Ions consisting of one peptide fragment plus cross-linker and the intact cross-linked peptide are printed in grey and are denoted in the amino acid sequence as ^{x}b or ^{x}y ions, respectively. Fragment ions that are created by an additional loss of ammonia are indicated as y* and b* ions.

between CaM and AC8 peptide is concerned, MS/MS data pointed to lysine residues at positions 75 and 77 in the central α-helix of CaM, Lys-21, and Lys-30 in the N-terminal lobe of CaM, and Lys-94 in the C-terminal lobe of CaM to be cross-linked with Lys-20 in the AC8 peptide (Table 4). In Figure 5, the product ion spectrum of the triply charged ion of the cross-linked product between CaM (amino acids 75 to 86, Met-76 oxidized) and AC8 peptide (amino acids 19 to 25), is presented. For both peptides comprising the cross-linked product, nearly complete y type ion series and a number of b type ions were observed, with the fragmentation patterns unambiguously identifying Lys-77 of CaM as the amino acid to be cross-linked with Lys-20 of the AC8 peptide. In total, MS/MS information was obtained for six intermolecular cross-linked products, with a number of cross-linked products being also present as oxidized species (Met-76 oxidized). Moreover, ammonia loss from the precursor ion was frequently observed in the product ion spectra. Three of the six intermolecular cross-linked products revealed cross-links between Lys-75 or Lys-77 of CaM with lysine 20 of the AC8 peptide (Table 4). One cross-linked product was found for each of the CaM lysine residues at positions 21, 30, or 94 with Lys-20 of the AC8 peptide. MS/MS analysis of peptides that had been modified by a partially hydrolyzed cross-linker revealed the same lysines that had already been found to be involved in intermolecular cross-linking (Lys-21, Lys-30, Lys-75, Lys-77, and Lys-94 of CaM and Lys-20 of AC8 peptide) (Table 4).

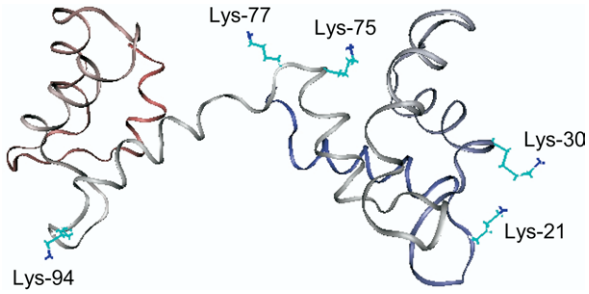


Figure 6. NMR structure of calcium-free CaM (pdb entry 1DMO) with the side chains of lysines indicated that were found to be cross-linked with the AC8 peptide

MS/MS experiments exposed that a number of cross-linked products had been created, in which two cross-linker molecules are present (Table 4, shaded areas)—thus making the complexity of created cross-linked species even more fascinating. For these species, the isotope pattern obviously is 1:2:1. Here, the first cross-linker molecule connects two amine groups of the binding partners, whereas the second cross-linker molecule is partially hydrolyzed due to the absence of a suitable amine group in the correct cross-linking distance. In the present case, either Lys-75 or Lys-77 reacted with Lys-20 in the AC8 peptide, whereas the second cross-linker molecule reacted at the remaining free lysine on one site and was hydrolyzed at its other reactive site (Table 4). All possible combinations of cross-linked species with incorporated nondeuterated and four-times deuterated cross-linkers were observed.

Structure of the Calcium-Free CaM/AC8 Peptide Complex

In Figure 6, an NMR structure of calcium-free CaM (apo-CaM) is presented, in which the side chains of these lysines are indicated that have been found to be involved in intermolecular cross-linking between CaM and the AC8 peptide based on MS/MS experiments. The intermolecular constraints found for the homobifunctional amine-reactive cross-linkers BS^2G and BS^3 (spacer lengths 7.7 and 11.4 Å, Table 1) point conclusively to Lys-75 or Lys-77 of the central α-helix of CaM as well as Lys-21 and Lys-30 in the N-terminal lobe, and Lys-94 in the C-terminal lobe of CaM to be cross-linked with Lys-20 of the AC8 peptide. One drawback of the present study consists in that only a few reactive sites (N-terminus and Lys-20) are present in the AC8 peptide, thus reducing the number of possibly created cross-linked products. For future studies, cross-linking reagents with different specificities, such as photoreactive cross-linkers, should be evaluated [13]. Without conducting further cross-linking studies, we are presently unable to propose a structural model of the CaM/AC8 peptide complex. From the calcium-free structure of CaM (Figure 6), one could envision that Lys-20 of the AC8 peptide is in the correct distance to CaM lysines at positions 21, 30, 75, and 77 to be cross-linked by BS^2G and BS^3. However, the observed cross-linked product between Lys-94 of CaM and Lys-20 of the AC8 peptide is not in agreement with the proposed arrangement. Further cross-linking studies will reveal if the AC8 peptide might even bind to CaM in different orientations.

Conclusions

The combination of employing isotope-labeled cross-linkers and a mass spectrometric technique yielding data with high mass accuracy has proven to be extremely powerful in rapidly identifying cross-linker containing species from the complex peptide mixtures created from enzymatic digestion of the cross-linking reaction mixtures. In case more than one reactive functional group, e.g., ε-amine groups of lysine residues, is present in a sequence stretch, an in-depth MS/MS analysis is essential for unambiguously identifying sites of modification. Further cross-linking studies will have to be conducted to propose a structural model for the CaM/AC8 peptide complex, which is formed in the absence of calcium.

Acknowledgments

The authors are indebted to Pierce Inc. for the generous gift of isotope-labeled cross-linkers BS^3 and BS^2G. The junior research group of AS is funded by the Saxon State Ministry of Higher Education, Research, and Culture and the Deutsche Forschungsgemeinschaft (DFG project Si 867/7-1). Financial support from the Thermo Electron Corporation (Mattauch-Herzog award of the German Society for Mass Spectrometry to AS) is also gratefully acknowledged.

References

1. Sinz, A. Chemical Cross-Linking and Mass Spectrometry for Mapping Three-Dimensional Structures of Proteins and Protein Complexes. *J. Mass Spectrom.* **2003,** *38,* 1225–1237.
2. Back J. W.; de Jong, L.; Muijsers, A. O.; de Koster, C. G. Chemical Cross-Linking and Mass Spectrometry for Protein Structural Modeling. *J. Mol. Biol.* **2003,** *331,* 303–313.
3. Sinz, A. Chemical Cross-Linking and FTICR Mass Spectrometry for Protein Structure Characterization. *Anal. Bioanal. Chem.* **2005,** *381,* 44–47.
4. Karas, M; Hillenkamp, F. Laser Desorption Ionization of Proteins with Molecular Masses Exceeding 10,000 Da. *Anal. Chem.* **1988,** *60,* 2299–2301.
5. Fenn, J. B.; Mann, M.; Meng, C. K.; Wong, S. F.; Whitehouse, C. M. Electrospray Ionization for Mass Spectrometry of Large Biomolecules. *Science* **1989,** *46,* 64–71.
6. Müller, D. R.; Schindler, P.; Towbin, H.; Wirth, U.; Voshol, H.; Hoving, S.; Steinmetz, M. O. Isotope-Tagged Cross-Linking Reagents. A New Tool in Mass Spectrometric Protein Interaction Analysis. *Anal. Chem.* **2001,** *73,* 1927–1934.
7. Collins, C. J.; Schilling, B.; Young, M. M.; Dollinger, G.; Guy, R. K. Isotopically Labeled Crosslinking Reagents: Resolution of Mass Degeneracy in the Identification of Cross-Linked Peptides. *Bioorg. Med. Chem. Lett.* **2003,** *13,* 4023–4026.
8. Taverner, T.; Hall, N. E.; O'Hair, R. A. J.; Simpson, R. J. Characterization of an Antagonist Interleukin-6 Dimer by Stable Isotope Labeling, Cross-Linking, and Mass Spectrometry. *J. Biol. Chem.* **2002,** *277,* 46487–46492.
9. Bennett, K. L.; Kussmann, M.; Bjork, P.; Godzwon, M.; Mikkelsen, M.; Sorensen, P.; Roepstorff, P. Chemical Cross-Linking with Thiol-cleavable Reagents Combined with Differential Mass Spectrometric Peptide Mapping—A Novel Approach to Assess Intermolecular Protein Contacts. *Protein Sci.* **2000,** *9,* 1503–1518.
10. Sinz, A.; Wang, K. Mapping Protein Interfaces with a Fluorogenic Cross-Linker and Mass Spectrometry: Application to Nebulin-Calmodulin Complexes. *Biochemistry* **2001,** *40,* 7903–7913.
11. Tang, X.; Munske, G. R.; Siems, W. F.; Bruce, J. E. Mass Spectrometry Identifiable Cross-Linking Strategy for Studying Protein–Protein Interactions. *Anal. Chem.* **2005,** *77,* 311–318.
12. Sinz, A.; Kalkhof, S.; Ihling, C. Mapping Protein Interfaces by a Trifunctional Cross-Linker Combined with MALDI-TOF and ESI-FTICR Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **2005,** *16,* 1921–1931.
13. Trester-Zedlitz, M.; Kamada, K.; Burley, S. K.; Fenyö, D.; Chait, B. T.; Muir, T. W. A Modular Cross-Linking Approach for Exploring Protein Interactions. *J. Am. Chem. Soc.* **2003,** *125,* 2416–2425.
14. Hurst, G. B.; Lankford, T. K.; Kennel, S. J. Mass Spectrometric Detection of Affinity Purified Crosslinked Peptides. *J. Am. Chem. Soc.* **2004,** *15,* 832–839.
15. Comisarow, M. B.; Marshall, A. G. Fourier Transform Ion Cyclotron Resonance Spectroscopy. *Chem. Phys. Lett.* **1974,** *25,* 282–283.
16. Marshall, A. G. Milestones in Fourier Transform Ion Cyclotron Resonance Spectrometry Technique Development. *Int. J. Mass Spectrom.* **2000,** *200,* 331–356.
17. Kruppa, G. H.; Schoeniger, J. S; Young, M. M. A Top Down Approach to Protein Structural Studies Using Chemical Cross-Linking and Fourier

Transform Mass Spectrometry. *Rapid Commun. Mass Spectrom*. **2003,** *17,* 155–162.
18. Novak, P.; Haskins, W. E.; Ayson, M. J.; Jacobsen, R. B.; Schoeniger, J. S.; Leavell, M. D.; Young, M. M.; Kruppa, G. H. Unambiguous Assignment of Intramolecular Chemical Cross-Links in Modified Mammalian Membrane Proteins by Fourier Transform-Tandem Mass Spectrometry. *Anal. Chem*. **2005,** *77,* 5101–5108.
19. Novak, P.; Young, M. M.; Schoeniger, J. S.; Kruppa, G. H. A Top-Down Approach to Protein Structure Studies Using Chemical Cross-Linking and Fourier Transform Mass Spectrometry. *Eur. J. Mass Spectrom*. **2003**, 9, 623–631.
20. Schulz, D. M.; Ihling, C.; Clore, G. M.; Sinz, A. Mapping the Topology and Determination of a Low Resolution Three-Dimensional Structure of the Calmodulin-Melittin Complex by Chemical Cross-Linking and High-Resolution FTICR Mass Spectrometry: Direct Demonstration of Multiple Binding Modes. *Biochemistry* **2004,** *43,* 4703–4715.
21. Kalkhof, S.; Ihling, C.; Mechtler, K.; Sinz, A. Chemical Cross-Linking and High-Performance Fourier Transform Ion Cyclotron Resonance Mass Spectrometry for Protein Interaction Analysis: Application to a Calmodulin/Target Peptide Complex. *Anal. Chem*. **2005,** *77,* 495–503.
22. Crivici, A.; Ikura, M. Molecular and Structural Basis of Target Recognition by Calmodulin. *Annu. Rev. Biophys. Biomol. Struct*. **1995,** *24,* 85–116.
23. Gu, C.; Cooper, D. M. F. Calmodulin Binding Sites on Adenylyl Cyclase VIII. *J. Biol. Chem*. **1999,** *274,* 8012–8021.
24. Bähler, M.; Rhoads, A. Calmodulin Signaling via the IQ Motif. *FEBS Lett*. **2002,** *513,* 107–113.
25. Schmidt, A.; Kalkhof, S.; Ihling, C.; Cooper, D. M. F.; Sinz, A. Mapping Protein Interfaces by Chemical Cross-Linking and FTICR Mass Spectrometry: Application to a Calmodulin/Adenylyl Cyclase 8 Peptide Complex. *Eur. J. Mass Spectrom*. **2005**, 11, 524–534.
26. Laemmli, U. K. Cleavage of Structural Proteins During Assembly of Head of Bacteriophage-T4. *Nature* **1970,** *227,* 680–685.
27. Jensen, O. N.; Shevchenko, A.; Mann, M. In: Protein Structure, A Practical Approach; Creighton, T. E., Ed.; Oxford University Press: Oxford, **1997**, p 48.
28. *BioAPEX User's Manual Vol. 1.1*; Bruker Daltonics: Billerica, MA, 1996.
29. Peri, S.; Steen, H.; Pandey, A. GPMAW—A Software Tool for Analyzing Proteins and Peptides. *Trends Biochem. Sci*. **2001,** *26,* 687–689.
30. Hermanson, G. T. Bioconjugate Techniques; Academic Press Inc.: San Diego, 1996, pp 187–227.

OPEN ACCESS Freely available online

PLoS PATHOGENS

Topology and Organization of the *Salmonella typhimurium* Type III Secretion Needle Complex Components

Oliver Schraidt[1,2], Matthew D. Lefebre[3], Matthias J. Brunner[1,2], Wolfgang H. Schmied[1,2], Andreas Schmidt[4], Julia Radics[1,2], Karl Mechtler[1], Jorge E. Galán[3], Thomas C. Marlovits[1,2]*

1 Research Institute of Molecular Pathology, Vienna, Austria, **2** Institute of Molecular Biotechnology GmbH, Austrian Academy of Sciences, Vienna, Austria, **3** Yale University School of Medicine, Section of Microbial Pathogenesis, Boyer Center for Molecular Medicine, New Haven, Connecticut, United States of America, **4** CD-Laboratory for Proteome Analysis, Vienna, Austria

Abstract

The correct organization of single subunits of multi-protein machines in a three dimensional context is critical for their functionality. Type III secretion systems (T3SS) are molecular machines with the capacity to deliver bacterial effector proteins into host cells and are fundamental for the biology of many pathogenic or symbiotic bacteria. A central component of T3SSs is the needle complex, a multiprotein structure that mediates the passage of effector proteins through the bacterial envelope. We have used cryo electron microscopy combined with bacterial genetics, site-specific labeling, mutational analysis, chemical derivatization and high-resolution mass spectrometry to generate an experimentally validated topographic map of a *Salmonella typhimurium* T3SS needle complex. This study provides insights into the organization of this evolutionary highly conserved nanomachinery and is the basis for further functional analysis.

Citation: Schraidt O, Lefebre MD, Brunner MJ, Schmied WH, Schmidt A, et al. (2010) Topology and Organization of the *Salmonella typhimurium* Type III Secretion Needle Complex Components. PLoS Pathog 6(4): e1000824. doi:10.1371/journal.ppat.1000824

Editor: C. Erec Stebbins, The Rockefeller University, United States of America

Received January 11, 2010; **Accepted** February 12, 2010; **Published** April 1, 2010

Funding: This project was supported in part by ZIT (Zentrum fuer Innovation und Technology Vienna) grant CMCN (Center of Molecular and Cellular Nanostructure Vienna) to T.C.M. and by NIH Grant AI30492 to J.E.G. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

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

* E-mail: marlovits@imp.ac.at

Introduction

One of the most exciting recent developments in the field of bacterial pathogenesis is the discovery that many bacterial pathogens utilize supramolecular nanomachines to deliver bacterial proteins into eukaryotic cells. These proteins, which are collectively referred to as effectors, have the capacity to modulate a variety of cellular functions including cytoskeleton dynamics, vesicle traffic, cell cycle progression and transcription. At least four types of machines capable of transporting effectors have been identified. They are known as type II, type III, type IV, and type VI protein secretion systems [1–5]. Arguably the best understood of these machines are the type III secretion systems (T3SS), which are essential for the virulence of several important bacterial pathogens including *Salmonella enterica*, *Shigella* spp., *Yersinia* spp., enteropathogenic strains of *E. coli*, and *Vibrio cholerae*. A central component of T3SS is the needle complex, a multiprotein structure that mediates the passage of the effector proteins through the bacterial envelope. Although this structure was initially identified in *Salmonella enterica* serovar *typhimurium* (*S. typhimurium*) [6], it has been shown to be conserved in other bacteria encoding T3SSs [7,8]. The cylindrically shaped needle complex is composed of a multi-ring base (~25nm in width and ~30nm in length), associated to the bacterial envelope, and a needle-like extension that protrudes several nanometers (~20–50nm) from the bacterial surface. The needle is anchored to the base through another substructure, the inner rod, which together with the needle filament forms a ~3nm wide channel that serves as conduit for the proteins that travel this secretion pathway [9]. Assembly of the needle complex occurs in discrete steps that first lead to the assembly of the base substructure [10]. Once assembled, the base begins to work as a secretion machine although exclusively devoted to the secretion of the proteins required for the assembly of the inner rod and the needle. Only upon complete assembly, the secretion machine changes substrate specificity and becomes competent for the secretion of effector proteins destined for delivery into eukaryotic target cells [5]. This functional reprogramming is believed to involve significant conformational changes in the needle complex itself [11].

S. typhimurium encodes two different T3SS within its pathogenicity island 1 (SPI-1) and 2 (SPI-2), which in a coordinated fashion mediate bacterial uptake into and replication within epithelial cells. Previous biochemical and genetic studies have established that the SPI-1-encoded *S. typhimurium* needle complex is composed of the bacterial proteins PrgH, PrgK, and InvG, which make up the base substructure, and PrgI and PrgJ, which constitute the needle and inner rod substructures, respectively [6,12]. Cryo electron microscopy and single particle analysis have provided a ~17Å resolution density map of the *S. typhimurium* SPI-1-encoded needle complex [9]. Recently, atomic structures of soluble domains of protein components from needle complexes from various bacterial species have become available [13–15].

Author Summary

Many Gram negative pathogens such as *Salmonella*, *Yersinia*, or *Shigella* use the type III secretion system (T3SS) to initiate infection in eukaryotic cells, resulting in well known clinical symptoms ranging from mild headaches and diarrhea to life-threatening diseases such as typhoid fever or bubonic plague. The T3SS is a highly developed macromolecular system that serves as a platform to make physical contact between host cells and pathogens and mediates the translocation of bacterial toxins (effector proteins) into eukaryotic cells. Central to the T3SS is the mega-dalton sized membrane associated needle complex, which is composed of several soluble and membrane proteins; however, their organization within the needle complex critical for proper assembly and function is unclear. Here, we use an integrated experimental approach that combines cryo electron microscopy with bacterial genetics, site-specific labeling, mutational analysis, chemical derivatization and high-resolution mass spectrometry in order to determine the topographic organization of individual components of the *Salmonella typhimurium* needle complex and define sites critical for its stability. Our study provides insights into the organization of this evolutionary highly conserved system and is the basis for further functional analysis.

These studies have revealed the presence of a conserved domain within the main components of the base, which in the *S. typhimurium* SPI-1 T3SS are InvG, PrgH and PrgK. Given the fact that these three proteins apparently organize themselves in a ring-like fashion, it has been proposed that this domain may mediate the formation of these rings. Attempts have been made to dock these protein domains into the needle complex structure. However, the relatively low resolution of the available electron microscopy density map does not allow the confident placement of the atomic structures of the different protein domains without additional experimental verification. Consequently, different, and in some case mutually incompatible, locations have been proposed for various protein domains [15,16].

In this study, we have used a combination of methods including bacterial genetics, biochemistry, mass spectrometry and cryo electron microscopy/single particle analysis to experimentally assign specific protein domains to different substructures of the needle complex. In addition, we have identified specific interaction sites among components of the needle complex, which are critical for its stable assembly. Combined, this analysis provides the first experimentally validated topographic map of different components of the needle complex of the *S. typhimurium* SPI-1 TTSS.

Results

InvG forms the outer rings and neck region and PrgH and PrgK form the inner rings of the needle complex

The *S. typhimurium* needle complex component InvG belongs to the secretin family, which is composed of outer membrane proteins that are associated with several secretion systems in Gram-negative bacteria. These include proteins associated with type II (e.g. *Klebsiella pneumoniae* PulD) [17] and type III (e.g. *Yersinia* spp. YscC) protein secretion systems [18], type IV pilus assembly (e.g. *Neiseria* PilQ) [19] and filamentous bacteriophage secretion (e.g. filamentous bacteriophage pIV) [20]. Secretins form higher-ordered ring-like structures, which in the case of PulD are organized in the form of two rings connected by a central disc. This basic architecture creates two chambers of different size, one of which extends with its N-terminal domain into the periplasmic space as visualized by cryo electron microscopy of the trypsin resistant core of the PulD complex [21]. Comparison of the PulD structure with the needle complex shows striking similarities between the PulD rings and the outer rings of the base substructure of the needle complex. The similarities also extend to regions of the neck of the needle complex, which connect the outer rings with the inner rings (**Fig. 1A**). These similarities strongly suggest that InvG forms the outer rings of the needle complex of the SPI-1 T3SS base. However, this has not been formally demonstrated and the extent to which InvG may form the neck region of the needle complex has not been experimentally determined. In order to ascertain what needle complex substructures are specifically formed by InvG, we purified needle complexes from a *ΔinvG* mutant strain and analyzed their structure by cryo electron microscopy and single particle analysis (**Fig. 1**). Western blot analysis of the structures isolated from the *ΔinvG* mutant showed the presence of PrgH and PrgK at equivalent stoichiometry to that found in needle complexes isolated from wild type (**Fig. 1B**). Structures isolated from this mutant strain showed the presence of the inner ring 1 (IR1) and 2 (IR2) and an extending needle-like structure but lacked the upper rings and the neck region (**Fig. 1C, E**). Subtraction of two-dimensional class averages of samples obtained from a *ΔinvG* mutant strain (**Fig. 1E**) from a wild-type needle complex structure (**Fig. 1D**) revealed that InvG is localized at the apical side of the needle complex forming the outer ring structures and neck region (**Fig. 1F**). In comparison to PulD, InvG extends further into the periplasmic space reaching the largest inner ring (IR1) of the needle complex, presumably anchoring the outer rings and thus stabilizing the base substructure.

The secondary structural features of PrgH and PrgK strongly suggest that they form the inner rings of the needle complex base although no direct evidence for this hypothesis has been presented. Subtraction of two-dimensional class averages of samples obtained from a *ΔinvG* mutant strain from the wild-type needle complex structure also revealed that the ring structures from this mutant strain are virtually identical to the inner rings of the wild type needle complex (**Fig. 1F**). Since, beside PrgH and PrgK, the needle and inner rod proteins are the only main structural components of the *ΔinvG* substructure, these results formally demonstrate that PrgH and PrgK make up the inner rings of the base substructure. Taken together these results demonstrate that InvG is the main component of the outer rings and connecting neck of the needle complex while PrgH and PrgK are the main components of the inner rings.

Localization and topology of PrgH within the base substructure of the needle complex

Secondary structure prediction analysis indicates that PrgH contains a transmembrane domain (from amino acid 142 to 162), which would separate the protein into two soluble domains of roughly equivalent size but distinct secondary structures (**Fig. S1, S2**). The N-terminal domain is predicted to be rich in beta-sheets, whereas the recently solved crystal structure of the major part of the C-terminal domain (amino acids 170–362) showed that it has a modular arrangement of similar α/β domains [15]. The localization of the two soluble domains of PrgH relative to the assembled needle complex, however, is still unclear. This information is essential to guide the potential placement of the recently solved atomic structure of a soluble domain of PrgH within the protein density map of the needle complex. In order to determine the topology of PrgH within fully assembled complexes we constructed *S. typhimurium* strains expressing N- or C-terminal

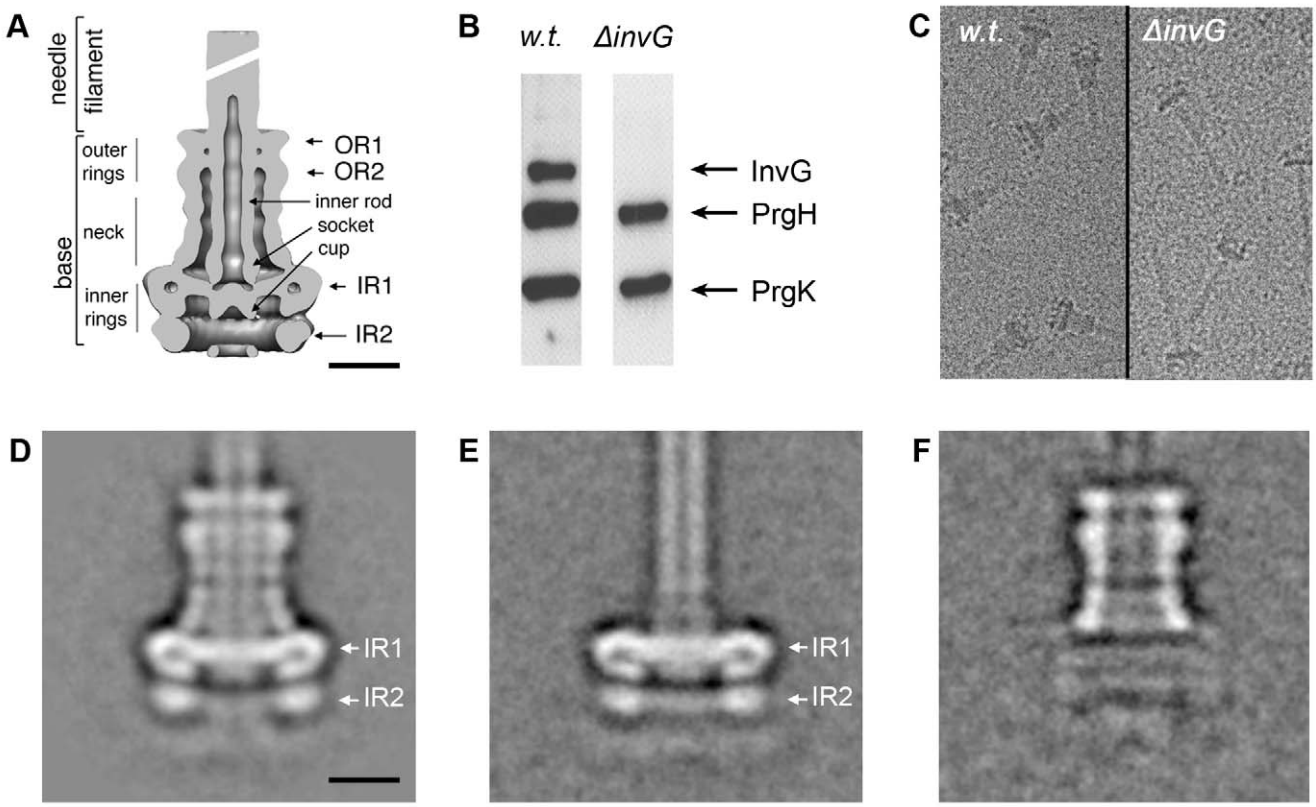


Figure 1. InvG forms the outer rings and neck region of the needle complex. (**A**) Cut-away view and description of individual substructures of the needle complex from *S. typhimurium*. (OR outer ring; IR inner ring) Bar = 10nm. (**B–F**) Analysis of complexes obtained from a *S. typhimuirum ΔinvG* strain: Western blot analysis (**B**), cryo electron microscopy images (**C**), and class averages of non-tilted complexes, (**D**, **E**) isolated from wild type and *ΔinvG* mutant *S. typhuimurium* strains, respectively. (**F**) Density difference between averaged images of wild type needle complexes and *ΔinvG* mutant complexes (**1F = 1D-1E**), indicates the position of the outer ring substructure.
doi:10.1371/journal.ppat.1000824.g001

poly-histidine tagged PrgH. Addition of the peptide tag did not affect the functionality of the needle complex as assayed by its ability to secrete the effector proteins SipB and SptP (**Fig. 2A**). Needle complexes isolated from these strains were labeled with Ni-NTA nanogold and analyzed by cryo electron microscopy and single particle analysis. Class-averages showed that the labeled N- and C-termini are located far away from each other, consistent with the bioinformatic prediction that the different termini of PrgH are topologically located on opposite sides of the inner membrane (**Fig. 2B and 2E**). Classification of gold-labeled needle complexes isolated from a strain expressing N-terminal poly-histidine tagged PrgH showed additional density exclusively on the cytoplasmic side of the base substructure suggesting that the N-terminus of PrgH faces the cytoplasm (**Fig. 2B**). Moreover, in side views of classified particles, the nanogold label can be seen at various vertical and horizontal positions below IR2, presumably due to a flexible tag and/or the existence of multiple copies of PrgH that are organized in a cylindrical fashion and that are not uniformly labeled by Ni-NTA nanogold. This observation is consistent with the diffuse appearance of the additional "density" observed below IR2 in a class average of *all* the labeled particles (**Fig. 2C and D**). In contrast, labeled needle complexes isolated from a strain expressing C-terminal poly-histidine tagged PrgH showed additional density above IR1 in close proximity to the neck region (**Fig. 2E–G**). In this case, the labeling was visualized as a distinct density, even in the class average of all particles suggesting a more rigid conformation of this domain of PrgH. Taken together, these results indicate that the amino terminus of PrgH faces the bacterial cytoplasm while its carboxy terminus is located within the periplasm of the bacterial envelope.

Localization of PrgK within the base substructure of the needle complex

PrgK is a lipoprotein with a canonical *sec*-dependent transport signal sequence that is processed upon secretion [6], and a predicted single transmembrane domain (AA 207–227) close to its carboxy terminus. Consequently, the large N-terminal domain of PrgK (starting from Cys-18) is predicted to be localized in the periplasm and anchored to the inner membrane via its transmembrane domain. The atomic structure of the PrgK homologue EscJ, which lacks a transmembrane domain, revealed that this protein is organized in two independent domains linked by a flexible linker [13,14] (**Fig. S1**). Although EscJ carries a 'ring forming' motif similar to that of PrgH and EscC (InvG) [15], EscJ rings have so far never been isolated and/or visualized. To investigate the topology of PrgK within the needle complex, we constructed *S. typhimurium* strains expressing N- or C-terminal poly-histidine tagged PrgK. Addition of the tag at the amino terminus resulted in a loss of type III dependent protein secretion (data not shown). In contrast, addition of the tag to the carboxy terminus did not negatively affect needle complex function as measured by SipB and SptP secretion (**Fig. 3A**). Needle complexes isolated from this strain, labeled with Ni-NTA and

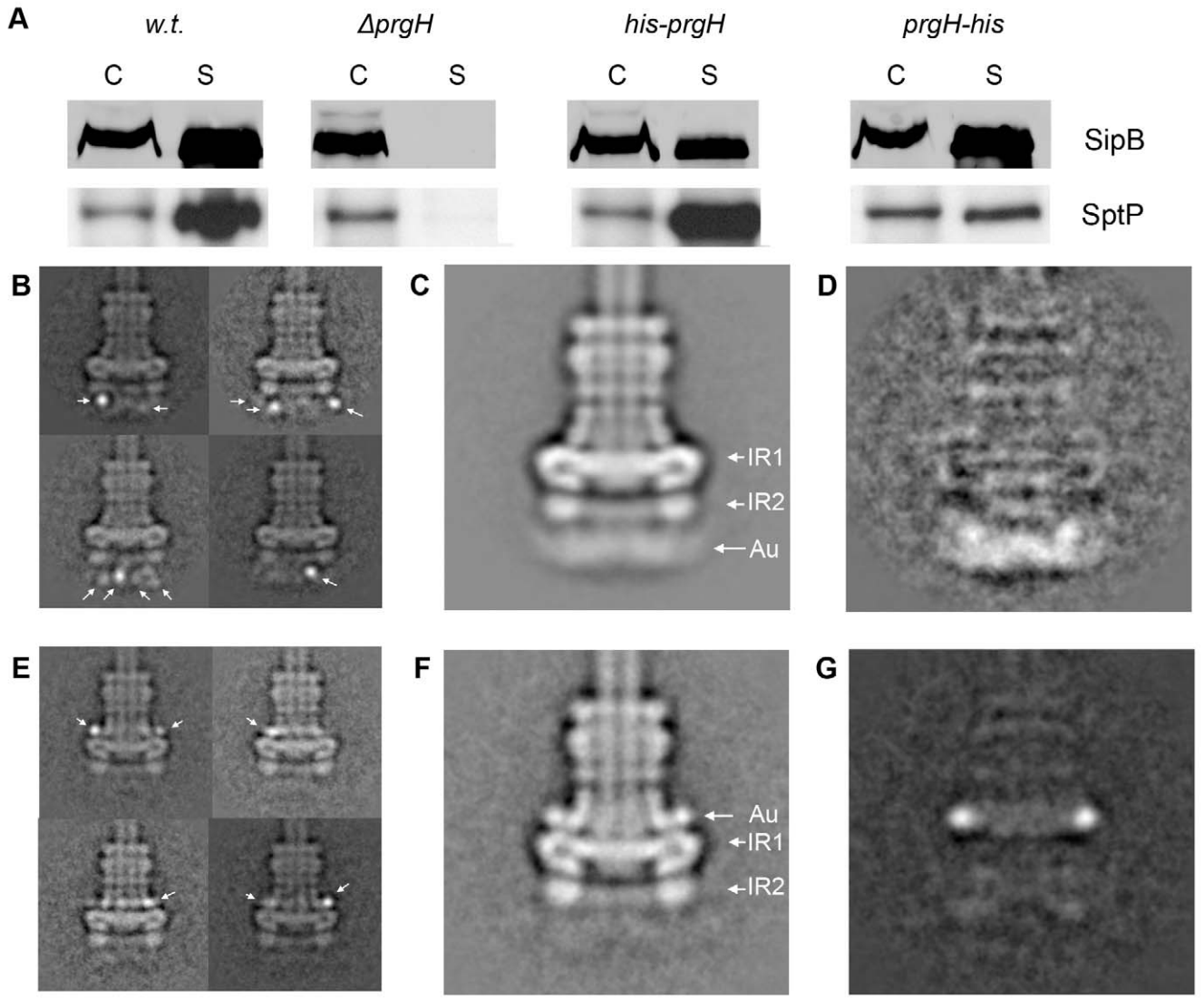


Figure 2. The N- and C-terminus of PrgH are located far away from each other within the needle complex. (**A**) N- and C-terminally poly-histidine tagged PrgHs are functional. Culture supernatants of a wild type *S. typhimurium* (w.t.), PrgH-deficient (*ΔprgH*), and mutant strains encoding either N- (his-PrgH) or C-terminally (PrgH-his) poly-histidine tagged PrgH were analyzed for the presence of the type III secreted proteins SipB and SptP by Western immunoblot. (C: whole cell lysates; S: culture supernatants). (**B**, **E**) Representative class averages obtained by single particle analysis of cryo electron microscopy images of Ni-NTA-labeled needle complexes derived from strains expressing either N-terminally (**B**) or C-terminally (**E**) tagged PrgH. (**C**, **F**) The total class averages (average of *all* particles) from the respective data set are shown in (**C**) (N-terminally labeled PrgH) and (**F**) (C-terminally labeled PrgH). The diffuse appearance of density at the basal side in (**C**) indicates that the Ni-NTA-nanogold (Au) label is present in various positions below IR2, but is more restricted above IR1 in C-terminally labeled complexes (**F**) (IR1 and IR2 = inner ring 1 and 2). (**D**, **G**) Density difference between the total averages of the labeled particles and unlabeled wild type complexes (panel D: w.t. needle complexes subtracted from labeled N-terminally tagged complexes (**2D = 2C-1D**); panel G: w.t. needle complexes subtracted from labeled C-terminally tagged complexes (**2G = 2F-1D**)).
doi:10.1371/journal.ppat.1000824.g002

examined by cryo electron microscopy and single particle analysis showed additional density exclusively on the cytoplasmic side of the base substructure (**Fig. 3B**). These results indicate that the carboxy terminal domain of PrgK faces the cytoplasm. Comparison of the position of the nanogold label on the C-terminus of PrgK (**Fig. 3C**) with that of the N-terminus of PrgH (**Fig. 2C**) by subtracting the total class averages from wild type (**Fig. 1D**) showed a more distinct localization of the nanogold label in PrgK (**Fig. 3D, 2D**). This observation suggests that the C-terminus of PrgK is less flexible than the N-terminus of PrgH and positioned closer towards the cup region of the needle complex (**Fig. 1A, 3E**).

Localization of PrgH and PrgK within IR1 and IR2

In order to refine the relative position of PrgH within the inner rings we constructed a strain of *S. typhimurium* that expresses a PrgH mutant in which we introduced a poly-histidine insertion linker after amino acid 267 (PrgH-267^{his}) (**Fig. S3**). The resulting strain expressed a functional T3SS system as shown by its ability to secrete the effector proteins SipB and SptP (**Fig. 4A**). Needle complexes isolated from this strain were labeled with Ni-NTA-nanogold and examined by cryo electron microscopy and single particle analysis. Additional density at the widest part of the IR1 (**Fig. 4B, 4C, Fig. S4**) was readily observed, indicating that PrgH is located at the periphery of IR1.

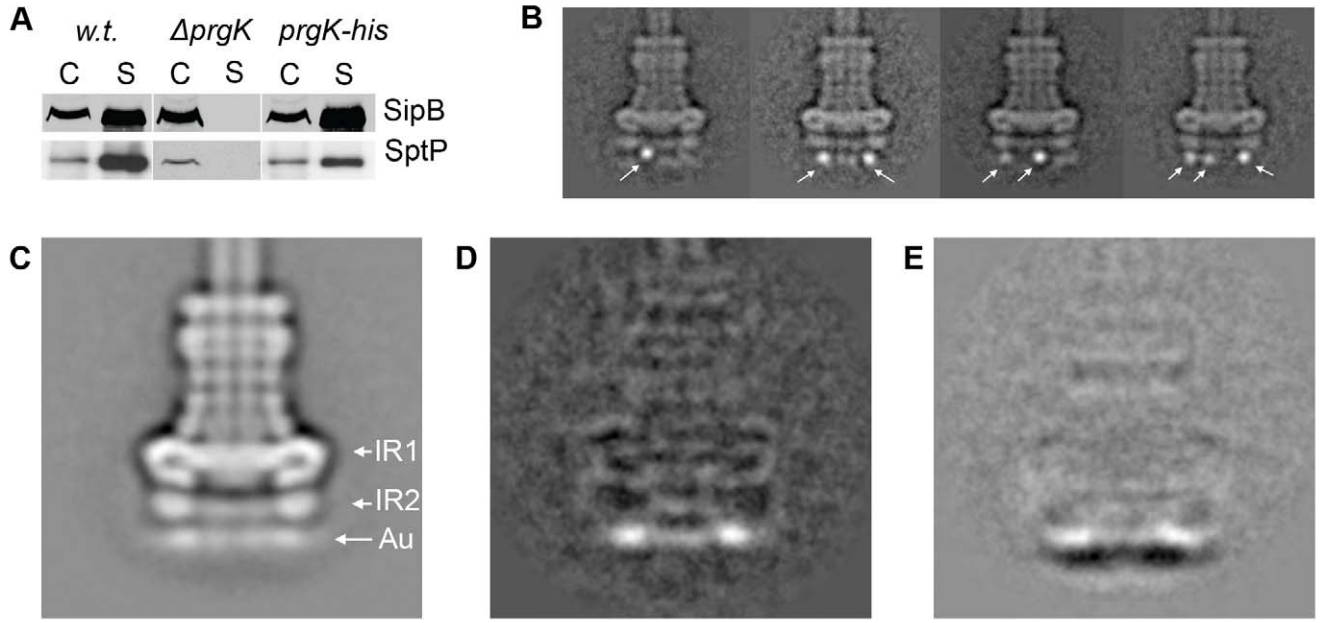


Figure 3. The C-terminus of PrgK is located at the basal (cytoplasmic) side of the needle complex. (**A**) Carboxy terminally poly-histidine tagged PrgK is functional. Culture supernatants of wild type *S. typhimurium* (w.t.), PrgK-deficient (*ΔprgK*), and a mutant strain encoding C-terminally (PrgK-his) poly-histidine tagged PrgK were analyzed for the presence of the type III secreted proteins SipB and SptP by Western immunoblot. (C: whole cell lysates; S: culture supernatants). (**B**) Representative class averages obtained by single particle analysis of cryo electron microscopy images of Ni-NTA-labeled needle complexes derived from strains expressing C-terminally poly-histidine tagged PrgK. (**C**) Total class average of Ni-NTA-labeled needle complexes obtained from a *S. typhimurium* strain expressing C-terminally poly-histidine tagged PrgK. The location of the density observed at the basal side (Au) is similar to the diffuse density observed in Ni-NTA-labeled needle complexes obtained from strains expressing N-terminally poly-histidine tagged PrgH (see **Fig. 2C**) (IR1 and IR2 = inner ring 1 and 2). (**D**) Resulting difference in density after subtraction of w.t. needle complexes from labeled C-terminally poly-histidine tagged PrgK complexes (**3D = 3C-1D**). (**E**) Subtraction of the total averages of labeled complexes obtained from a strain expressing N-terminally-tagged PrgH from a strain expressing C-terminally-tagged PrgK (**3E = 3D-2C**).
doi:10.1371/journal.ppat.1000824.g003

PrgK has been proposed to exist as an inner ring enclosed by an outer ring formed by PrgH [15]. This model was based on the crystallographic analysis of the PrgK homolog EscJ, which crystallized in a superhelical fashion. Subsequent modeling led to an approximately 180Å wide ring structure, which was proposed to be anchored to the inner membrane by the lipidated N-terminal cysteine of EscJ/PrgK. In this model, the EscJ/PrgK ring would be located on top of the outer leaflet of the inner membrane projecting into the periplasmic space [14]. An alternative model proposed for MxiJ, the Shigella spp. homolog of PrgK, positions this protein as radial spikes enclosed by an outer ring [16]. To gain insight into the organization of PrgK within the needle complex, we sought to obtain top views of the inner rings of the needle complex by electron microscopy for subsequent single particle analysis, which required the removal of the InvG rings. The proximity of the C-terminus of PrgH to InvG suggested the possibility that truncating the C-terminus of PrgH could weaken the interaction between these two proteins. We therefore constructed a S. *typhimurium* strain that expresses a PrgH mutant lacking its last 4 amino acids ($PrgH^{\Delta 4}$). The resulting strain expressed a functional T3SS as shown by its ability to secrete effector proteins (**Fig. 4A**) and the presence of intact needle complexes (**Fig. 4D**). However, when subjected to high pH treatment, the needle complexes not only disassembled the needle filament over time [9], but, in contrast to wild type, upon negative staining the complexes obtained from this mutant strain could be further disassembled to generate intact inner rings (IR1/IR2) separated from the smaller outer rings (**Fig. 4D**). We analyzed the inner rings (IR1/IR2) thus obtained by electron microscopy without imposing any symmetry or introducing any potential model bias. This analysis showed the presence of several concentric rings with different staining intensities (**Fig. 4E**). The two concentric rings with the larger diameter (~250Å and ~180Å) exhibited repeating subunits separated every 15 degrees thus resulting in a 24-fold symmetry. The accessibility of $PrgH\text{-}267^{his}$ to surface labeling and the appearance of additional density on the outermost surface of IR1 strongly suggested that the largest ring visualized in the inner ring particles must be made up by PrgH. On the other hand, the size, configuration, and density distribution of the second largest ring with an approximate diameter of 180Å are consistent with the hypothesis that this ring is formed by PrgK (**Fig. 4E, G, Fig. S6**). If this were the case, most of PrgK would be buried by the presence of InvG (top) and PrgH (outside). Consequently, this model would also predict that removal of InvG should expose otherwise buried residues. To test this hypothesis, we compared the ratio of accessibility of lysine residues of PrgK in complexes isolated from wild type or *ΔinvG S. typhimurium* strains (which exposes the apical side of IR1, see **Fig. 1**) by subjecting the isolated particles to acetylation using NHS-acetate and subsequent analysis by mass spectrometry. Lysine 168 from PrgK was found to be more frequently (>15 times) acetylated in tryptic peptides derived from PrgK obtained from complexes from a *ΔinvG* mutant strain than in PrgK peptides obtained from wild type needle complexes (**Fig. 4F, Table S1**). Consistent with this observation, modeling of PrgK using the EscJ structure as a template (see below) showed that the side chain amino group of K168 would be surface exposed in the absence of InvG (**Fig. 4G**).

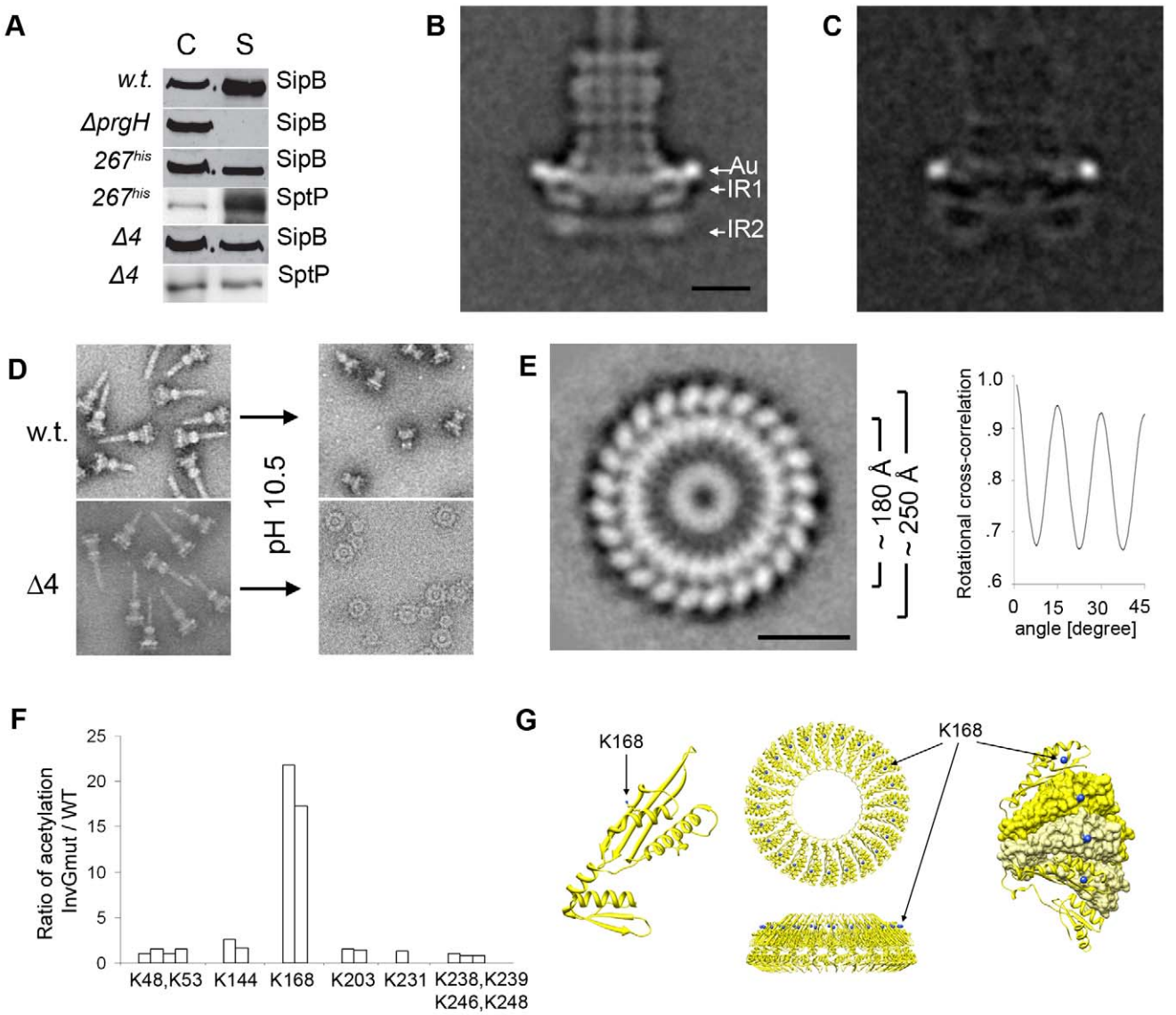


Figure 4. Organization of PrgH and PrgK within the lower ring of the needle complex. (**A**) Insertion of a poly-histidine linker at amino acid 267 of PrgH (267^{his}) or removal of four amino acids from its C-terminus (Δ4) does not alter its function. Culture supernatants of wild type *S. typhimurium* (w.t.), PrgH-deficient (*ΔprgH*), and mutant strains encoding a poly-histidine linker at amino acid 267 of PrgH (267^{his}) or a PrgH lacking the terminal four amino acids (Δ4) were analyzed for the presence of the type III secreted proteins SipB and SptP by Western immunoblot. (C: whole cell lysates; S: culture supernatants). (**B**) Representative class average obtained by single particle analysis of cryo electron microscopy image of Ni-NTA-labeled needle complexes obtained from a *S. typhimurium* strain expressing PrgH with a poly-histidine tag inserted at amino acid 267. A prominent gold label (Au) is seen at the widest side of the periplasmic face of the IR1 (IR1 and IR2 = inner ring 1 and 2). Bar = 10 nm. (Additional results of the analysis are shown in **Fig. S4**). (**C**) Resulting difference in density after subtraction of unlabeled w.t. needle complexes from labeled complexes isolated from strains expressing a poly-histidine insertion following amino acid 267 in PrgH. (**D**) Mutant needle complexes carrying truncated PrgH can be selectively disassembled into larger and smaller rings by shifting pH to 10.5 and subsequent negative staining, whereas wild type (w.t.) needle complexes maintain the integrity of the base. (**E**) *En face* class-average derived from single particle analysis from negatively stained electron microscopy images of inner rings substructures. The substructures were obtained by selective disassembly of needle complexes isolated from a mutant strain encoding for a C-terminally, four amino acid truncated PrgH. The ring substructure is organized in two larger concentric rings with different diameters (~180Å and ~250Å). Bar = 10nm. Rotational cross-correlation analysis revealed that the maximum of the cross-correlation peak is repeatedly obtained every 15°, demonstrating that the larger concentric rings of the inner ring structure from the *prgHΔ4* mutant strain exhibit 24 fold symmetry. (**F**) Surface accessibility of lysine K168 is increased in complexes obtained from a *S. typhimurium ΔinvG* strain. Primary amines of wild type and *ΔinvG* mutant complexes were acetylated with Sulfo-NHS-Acetate and the ratio of modified to non-modified peptides was determined by mass spectrometry. Each bar represents the ratio of acetylation of lysines between specific PrgK peptides obtained from the *ΔinvG* mutant and the same peptides obtained from a wild type complex (**Table S1**). The ratio of peptide acetylation is an average of two independent mass spectrometry measurements. (**G**) Lysine 168 is surface exposed in modeled PrgK rings. A monomeric PrgK structure was modeled using EscJ as a template and protein-protein contacts present in the EscJ crystal structure were used to form a PrgK ring (top and side view). Lysine 168 is located on top of the ring and its side group nitrogen is surface exposed as highlighted in blue in the surface view of the cut-out segments of individual monomers. The cut-out segment shows two PrgK molecules in surface representation (yellow and light-yellow) followed by one neighboring molecule on each side but displayed in ribbon style.

doi:10.1371/journal.ppat.1000824.g004

Taken together these results indicate that the inner rings are composed of two larger concentric rings: a peripheral ring formed by PrgH enveloping an inner ring formed by PrgK, which is shielded by PrgH (on the sides) and InvG (on the top).

Domain interactions among components of the type III secretion needle complex characterized by cross-linking and mass spectrometry

In order to refine the spatial relationship between the domains of the different proteins that make up the needle complex and to gain insight into the nature of potential domain-domain interactions, we used chemical cross-linking combined with high-resolution mass spectrometry. Purified needle complexes were incubated with the bi-functional cross-linking agent BS2G (d0/d4), which is able to covalently link primary amino groups at a distance up to ~8Å. Particles in which two or more complexes were cross-linked were separated from individual ones by re-purification on sucrose gradients, and tryptic fragments generated from these complexes were analyzed by mass spectrometry after protease digestion in solution. Several cross-links were identified between the different domains of InvG, PrgH, and PrgK (**Fig. 5A, Table 1**). A single cross-link between InvG (K38) and PrgH (K367) was found, establishing that the N-terminal domain of InvG and the C-terminal domain of PrgH are in close contact. This is also consistent with the observation that truncation of four amino acids from the carboxy terminal end of PrgH resulted in the destabilization of the needle complex (**Fig. 4D**) and supports the notion that the interaction of PrgH with InvG is very important for the linkage of these two substructures. From strains with longer truncations at the C-terminus of PrgH only complexes similar to those obtained from a *ΔinvG* mutant strain could be purified (data not shown), further demonstrating the importance of this domain for needle complex assembly.

We found two clusters of residues of the amino terminus of PrgH that cross-linked to the C-terminal end of PrgK (**Table 1** and **Fig. 5A**). In addition to the primary amino group at the N-terminal end of the protein, lysine residues at position 5, 132, and 133 of PrgH were found to cross-link to lysine residues at position 239, 246 and 248 of PrgK. The linkages found exclusively between the N-terminal domain of PrgH and the C-terminal tail of PrgK are in full agreement with the results obtained from the nanogold labeling experiments. The absence of any cross-links from the large N-terminal domain of PrgK is also consistent with the proposed concentric arrangement of PrgK shielded by PrgH and InvG.

Topology model of the *S. typhimurium* needle complex components

To construct a model for the topological arrangement of the components of the *S. typhimurium* needle complex, we used the available crystal structure of PrgH and modeled InvG and PrgK using the structures of their close homologues EscC and EscJ, respectively (Text S1). In addition, we incorporated in such model all of our experimental findings described above. The model shows that the N-terminal domain of PrgH and the short C-terminal domain of PrgK are in close contact at the cytoplasmic face of the needle complex (**Fig. 5B**). Furthermore, the N-terminal domain of InvG is localized at the lowest part of the neck-region oriented in such a way that allows its interaction with the C-terminal domain of PrgH (**Fig. 5B, Fig. S7**). This arrangement is supported by the finding of cross-linked peptides [IPVTGSGFAV$\mathbf{K_{38}}$DDSLR]-[DDWL$\mathbf{K_{367}}$G] encompassing these two domains (**Table 1**). The C-terminal PrgH peptide, however, is not visible in the atomic structure, hence this interaction can not be incorporated into a high-resolution model building (**Fig. 5D**).

Several models involving 12 or 14 fold symmetries have been proposed for the N-terminal region of the InvG homologue EscC. These ring models were placed at different positions within the needle complex depending on the proposed symmetries and correlating different volumes and diameters [15]. Our data indicate that the amino terminus of InvG should face the inner rings, which provides an unambiguous orientation to the InvG ring that is incompatible with some of the previously proposed models [15,16]. However, our data cannot clarify the issue of the differing symmetries in the proposed models, which would require experimental data at a sufficiently high resolution to show individual subunits within the outer ring and neck region.

PrgH and PrgK are the main constituents of the inner rings. Nanogold labeling of a needle complex isolated from strain expressing a PrgH mutant derivative with a his-tag after position 267 resulted in an additional density on the outer perimeter of the IR1. This observation led us to manually position the C-terminal domain of PrgH in such a way that amino acid 267 would be located at the periphery of IR1 (**Fig. 5C, E, Fig. S7**). The positioning of PrgK in our model takes into account the crystallographic contacts between the monomers observed in its homologue EscJ, resulting in a ring of roughly the same diameter to that proposed for EscJ. In our model we have also positioned the PrgK ring in such a way that PrgK K168 faces the periplasm, to account for our observation that K168 becomes accessible for derivatization in the absence of InvG (**Fig. 5C, E, Fig. S7**).

Discussion

The atomic structures of several soluble domains of needle complex components have become available. Attempts have been made to place those domains within the available structures of the entire needle complex to begin to generate an atomic model of the entire structure. However, the low resolution of the available needle complex structure has significantly hampered this objective and has prevented the unambiguous assignment of specific domains to specific protein densities within the needle complex. The cylindrical architecture of the needle complex further complicates the docking since, depending on the assumed stoichiometry or subunit number, modeled ring-like structures can be placed at different positions within the needle complex. Consequently, different studies have proposed incompatible locations for different proteins and/or protein domains within the needle complex [15,16]. We have used a multi-pronged approach to generate data allowing us to place domains of different protein components of the needle complex at specific sites of its structure. We have experimentally shown that InvG forms the outer rings. Although widely predicted from secondary structure analysis as well as the organization of homologues in other secretion systems (e.g. PulD family of proteins), this is the first experimental demonstration of this organization. Furthermore, our data demonstrated that InvG reaches deep into the periplasmic space making up the entire neck region and making direct contact with the inner rings, and in particular, the carboxy terminal domain of PrgH. Although the C-terminus of PrgH is not visible in the atomic structure therefore hindering high-resolution model building, the interaction between this domain and the amino terminus of InvG appears to be critical for the stability of the entire complex.

Our data also demonstrates that PrgH and PrgK are the main constituents of the inner rings. Our results indicate that the C-terminal domain of PrgH is localized at the periphery of IR1, as supported by the nanogold label of a tag introduced between amino acid 267 and 268, **Fig. 4B and 4C**). As previously

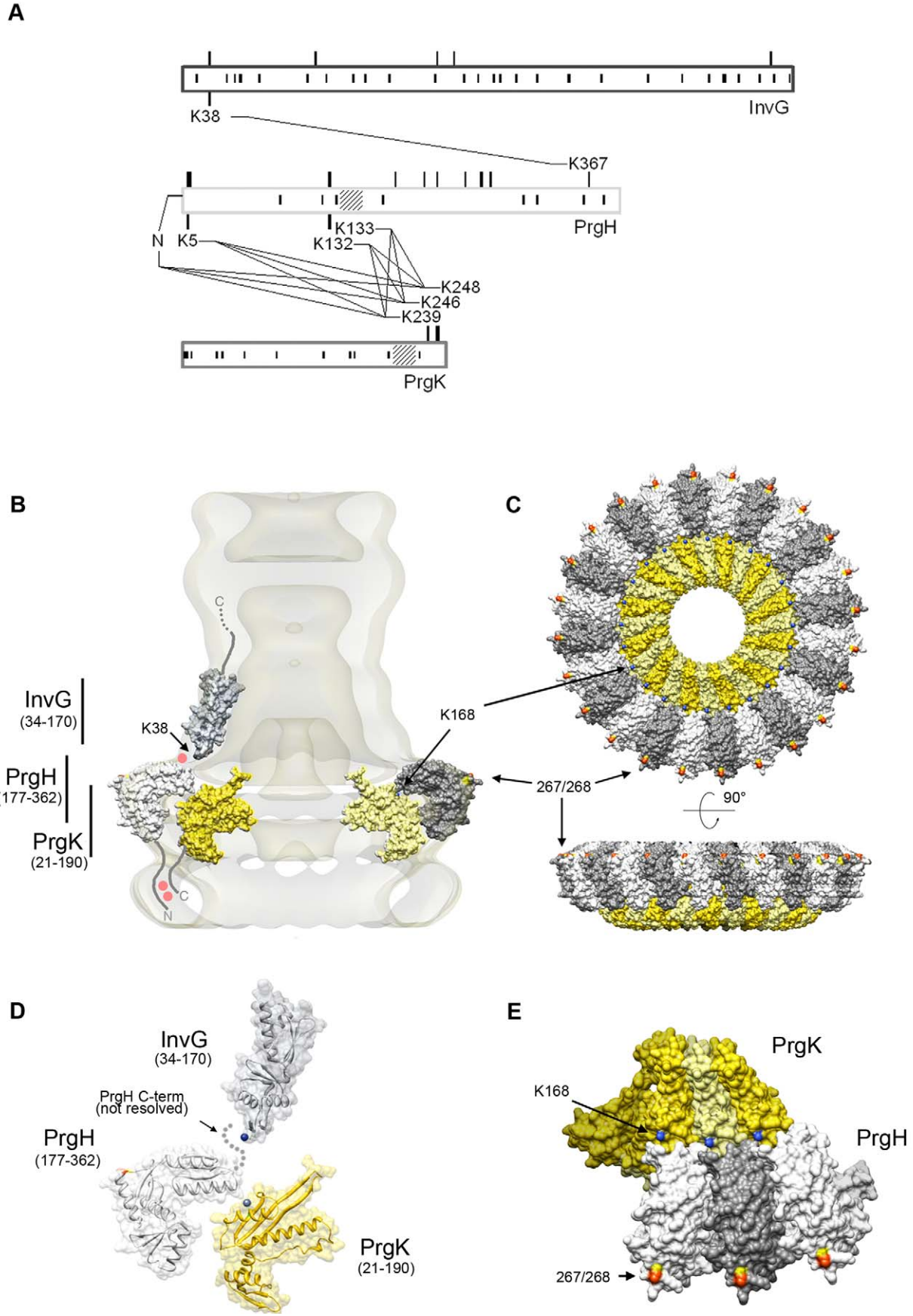

A
InvG
K38
K367
PrgH
N
K5
K133
K132
K248
K246
K239
PrgK
B
C
InvG
(34-170)
PrgH
(177-362)
PrgK
(21-190)
K38
K168
267/268
90°
D
E
InvG
(34-170)
PrgH C-term
(not resolved)
PrgH
(177-362)
PrgK
(21-190)
PrgK
K168
PrgH
267/268

Figure 5. Domain interactions and relative orientations of needle complex components. (**A**) Proximity of specific domains of the base proteins, InvG, PrgH, and PrgK within the needle complex. The block diagrams shows the three major base proteins, InvG, PrgH, and PrgK and covalent cross-links of peptides obtained from chemically derivatized needle complexes at primary amino groups. Amino acid position are indicated for the full length proteins prior signal peptide cleavage (processed InvG starts at Ser-25, and processed PrgK starts at Cys-18 [6]) protein-cross-links found are indicated with amino acid position and with crossing lines between proteins. While the position of non-derivatized lysines (presumably due to lack of surface exposure) is shown as vertical lines *within* the block diagram, the positions of derivatized lysines (surface accessible) is indicated as vertical lines extending from the block diagram (**Table S2**). Note that lysines within the N-terminal domain of PrgK are not derivatized, suggesting that the majority of PrgK within fully assembled needle complexes is not surface exposed. (**B–E**) Topographic model of the needle complex: Localization of InvG, PrgH, and PrgH within the base of the needle complex. The N-terminal domain of InvG (blue-grey) reaches far down into the neck region and is in close contact with the C-terminal domain of PrgH (white and grey), which resembles the larger of the two concentric rings. Insertion of a poly-histidine tag between amino acid 267 (yellow) and 268 (orange) and subsequent Ni-NTA nanogold labeling further determines the position of this domain within the complex. Sites of interaction found by cross-linking and mass spectrometry, for which in the case of PrgH and EscJ/PrgK no atomic structure is available, are labeled as red dots. The N-terminal domain of PrgH is pointing to the cytoplasmic side of the complex, and interacts with the C-terminal domain of PrgK. For both, no high resolution structure is available as of yet. The N-terminal domain of PrgK is located within the complex and is therefore packed into its position by PrgH from the side and InvG from the top. (**C**) Top and side view of the modeled PrgH (white/grey alternating) and PrgK (yellow/bright yellow alternating), as well as sites accessible for nanogold labeling (267/268) and chemical derivatization (K168). (**D**) Proposed relative position of protein domains from PrgH, InvG, and PrgK. The side group nitrogen of K38 (InvG) and K168 (PrgK) are highlighted in blue. (**E**) Top-view of three PrgK and PrgH monomers extracted from the modeled inner ring structure highlighting sites for chemical derivatization (PrgK (K168, blue)) and Ni-NTA nanogold labeling of a poly-histidine insertion at position 267/268 (yellow/orange) within PrgH.
doi:10.1371/journal.ppat.1000824.g005

proposed [15], our results suggest that PrgK is organized in a smaller diameter ring structure engulfed by the PrgH ring on the side and covered by the InvG neck region on top.

Multicomponent macromolecular complexes are central to many fundamental processes in biology. To gain insight into mechanistic details, knowledge not only of the atomic structure of the different subunits, but also their orientation relative to one another is essential. Therefore, our studies offer an essential view of the architecture of this remarkable bacterial nanomachine and will be the basis for further functional studies.

Materials and Methods

Bacterial strains and plasmids

All strains were derived from the non-flagellated *S. enterica* serovar *typhimurium* strain SJW2941. The *invG*, *prgH*, and *prgK* mutant alleles were introduced into this strain by P22HT*int*-mediated transduction as described elsewhere [22]. The needle complexes lacking InvG were purified from the strain SB1171 (all "SB" strains are described elsewhere) [10]. PrgH modified needle complexes were obtained by complementing the strain SB906, which harbors a chromosomal deletion of *prgH*, with plasmids expressing C-terminal truncated PrgH, an 18x-C-terminal, a 6x-N-terminal, or a 6x internal (after M267) poly-histidine tagged PrgH. A longer C-terminal poly-histidine tag was chosen because a 6xhistidine-tag yielded a poor Ni-NTA-NanoGold label. All plasmids were based on the low copy vector plasmid pWSK29 and expressed the tagged PrgH under its natural promoter (500bp upstream sequence). The w.t. strain SB905 was used as a negative control for the Ni-NTA-NanoGold labeling experiments and for the detection of adjacent epitopes by mass spectrometry. An allele of PrgK with a C-terminal 6x-poly-hisitidine tag was introduced into SB905 by homologous recombination as described previously [22]. Plasmids expressing the *hilA* positive transcriptional regulator

Table 1. List of hetero-cross-linked peptides determined by chemical derivatization and mass spectrometry.

Crosslinked proteins	Crosslinked lysines	Peptide1	Peptide2	Mass Error (ppm)	xQuest Score
InvG-PrgH	K38—K367	IPVTGSGFVA**K**DDSLR	DDWL**K**G	2.1	17.7
PrgH-PrgK	N-Term-K239	**M**ETSK	**K**GITADDK	1.3	25.8
PrgH-PrgK	N-Term-K246	**M**ETSK	GITADD**K**AK	1.3	26.7
PrgH-PrgK	N-K248	**M**ETSK	KGITADD**K**AK	0.7	19.1
PrgH-PrgK	N-K248	**M**ETSK	A**K**SSNE	1.1	20.5
PrgH-PrgK	K5-K239	METS**K**EK	**K**GITADDK	0.5	21.4
PrgH-PrgK	K5-K246	METS**K**EK	GITADD**K**AK	1.0	18.4
PrgH-PrgK	K5-K248	METS**K**EK	A**K**SSNE	0.0	15.4
PrgH-PrgK	K132-K239	LETSA**K**K	**K**GITADDK	1.4	23.4
PrgH-PrgK	K132-K246	LETSA**K**K	GITADD**K**AK	0.9	15.2
PrgH-PrgK	K132-K248	LETSA**K**K	A**K**SSNE	1.0	22.4
PrgH-PrgK	K133-K239	**K**NEPR	**K**GITADDK	1.1	21.2
PrgH-PrgK	K133-K246	**K**NEPR	GITADD**K**AK	1.3	24.7
PrgH-PrgK	K133-K248	**K**NEPR	A**K**SSNE	1.7	17.8

Only peptides with an xQuest score of higher than 15 are displayed as those cross-linking results are likely to be reliable [25] (Text S1). The respective modified amino acid is displayed in bold. Under "Crosslink" listed positions are the absolute positions of the respective protein (including signal peptides), "Error" the deviation of measured precursor ion mass from theoretical precursor ion mass in parts per million, and "Score" the xQuest score.
doi:10.1371/journal.ppat.1000824.t001

gene under the control of P_{BAD} promoter (pSB667 or pSB1418) were used to over-express the SPI-I TTSS regulon as described elsewhere [23].

Bacterial secretion assay

The preparation and analysis of cultured supernatant proteins was conducted as described elsewhere [22].

Needle complex expression and purification

Needle complex purification was based on the purification protocol previously published [9]. Needle complexes were purified from 2 L of bacterial culture. For the strain SB1171 the protocol was up-scaled to 18 L of bacterial culture to obtain a yield comparable to the other strains. Details are provided under Text S1. To further improve purity of the sample or to separate inter-particularly cross-linked particles from single particles, a sucrose gradient centrifugation was performed. A continuous 10–25% sucrose gradient in thin-wall tubes of the Sorvall TH-660 rotor was made with a 0.1% LDAO, 10 mM sodium phosphate (pH 7.4), 0.5 M NaCl buffer. The sample was applied to the top of the gradient and centrifuged for 3.5 hours at 50 krpm. Two 1.1 ml and four 0.45 ml fractions were collected from the top of the tube and were diluted in sucrose-free buffer and pelleted at 90 krpm for 30 minutes in a Sorvall S100-AT4 rotor. The pellets were re-suspended in 0.1 ml 0.1% LDAO, 10 mM sodium phosphate (pH 7.4), 0.5 M NaCl. Integrity and purity of the sample was verified by negative stain EM and SDS-PAGE.

Disassembly of needle complexes to inner and outer rings

The needle filament was removed by incubating purified needle complexes in 10 mM sodium phosphate (pH 10.4), 0.5 M NaCl, 0.1% LDAO for 15 min at 37°C. Subsequently the pH was re-adjusted by purifying the sample by sucrose gradient centrifugation as described above. The sample was observed by negative stain electron microscopy as described below.

Ni-NTA-NanoGold labeling of poly-histidine-tagged needle complexes

Samples tending to aggregate (His-PrgH, PrgK-His) were purified in presence of 10mM imidazol. Free thiols of sucrose gradient purified needle complexes were blocked by incubating the sample with 1 mM N-ethylmaleimide at 4°C over night. Subsequently, imidazole was added to a concentration of 20 mM and the sample was incubated with 5 µM nickel-nitrilotriacetic acid (Ni-NTA) NanoGold (Nanoprobes, Stony Brook, NY) at room temperature for 10 minutes. The sample was gel-filtrated with a Sephacryl 300 column to remove unbound Ni-NTA-NanoGold and analyzed by cryo-electron microscopy.

Chemical derivatization, mass spectrometry (MS) and MS data analysis

Chemical derivatization was performed by incubating sucrose gradient purified sample (about 1mg/ml) on ice with 600 µM Sulfo-NHS Acetate (Pierce, Rockford, USA) for 60 minutes or 200 µM BS2G-d0/d4 (Pierce, Rockford, USA) for either 30 or 60 minutes.. Details about chemical derivatization and subsequent MS analysis are provided in Text S1.

Electron microscopy and image processing

Samples were applied to glow-discharged carbon-coated 400 mesh hexagonal Cu/Pd-grids. For negative stain images 5 µl of sample was applied to the grid and subsequently stained with 2% PTA (phosphotungstate), pH 7.0 (**Fig. 4D**) or NanoVan (Nanoprobes, Stony Brook, NY) (**Fig. 4E, Fig. S5**). Overview images were acquired at 44,000-fold magnification in a Morgani TEM (FEI Company, Hillsboro, USA) at 80kV using an 11 megapixel CCD camera. High-resolution data was collected with a FEI Tecnai Polara at 300kV using a Gatan Ultrascan 4000 UHS CCD camera (16 mega-pixel, 4k×4k, 15 micron pixel size). Images were acquired at 112,968-fold magnification, which corresponds to 1.33 Å/pixel at the level of the specimen, with underfocus values ranging from 1.2–3.5 µm.

For cryo electron microscopy 5 µl sample was applied to glow-discharged grids before vitrification by plunge freezing in liquid ethane. Low-dose data was collected with a FEI Tecnai Polara at 300kV using a Gatan Ultrascan 4000 UHS CCD camera (16 mega-pixel, 4k×4k, 15 micron pixel size). Images were acquired at 71,949-fold magnification (2.08 Å/pixel) with under focus values ranging from 1.2–3.5 µm.

Individual particle projections were extracted, combined into a dataset and processed by IMAGIC-5 (Image Science Software GmbH, Germany). The contrast reversals imposed by the contrast transfer function (CTF) of the objective lens were corrected for each particle projection using the mean under focus value of the respective CCD-images as determined by the program CTFFIND3 [24].

Supporting Information

Text S1 Supplementary Information Protocols
Found at: doi:10.1371/journal.ppat.1000824.s001 (0.05 MB RTF)

Figure S1 Block diagram of the three major base proteins, InvG, PrgH, and PrgK of the needle complex and region of atomic structures of PrgH and the homologues EscC (InvG) and EscJ solved. Vertical lines extending from the block diagram are positions of lysines that can be chemically derivatized and are presumably surface exposed. Remaining non-derivatizable (and probably not surface exposed) lysines are indicated as vertical lines within the block diagram.. Amino acid position for InvG, PrgH, and PrgK are indicated for the full length proteins prior signal peptide cleavage (processed mature InvG starts at Ser-25, and PrgK starts at Cys-18, respectively).
Found at: doi:10.1371/journal.ppat.1000824.s002 (2.84 MB TIF)

Figure S2 Sequence alignment and secondary structure prediction of the major part of the N-terminal domain of PrgH (1–122) from various species (S. typhimurium (PrgH), E. coli (EprH), Y. enterocolitica (Ye3550), S. flexneri (MxiG)). The N-terminal domain of PrgH is predicted to be mostly composed of beta-strands. Highly conserved and similar residues are indicated in red.
Found at: doi:10.1371/journal.ppat.1000824.s003 (0.47 MB TIF)

Figure S3 Ribbon diagram of PrgH (177–362) and position for insertion of a poly-histidine tag following amino acid 267 for Ni-NTA-nanogold labeling.
Found at: doi:10.1371/journal.ppat.1000824.s004 (0.23 MB TIF)

Figure S4 Single particle analysis of Ni-NTA labeled needle complexes with a poly-histidine insertion following position 267 in PrgH. Representation of various views of class-averages, variance-images, and individual particles of PrgH-267^{his} labeled needle complexes show the presence of the nanogold label at the outermost perimeter of the inner rings. The highest variance was observed at the position of the nanogold label, which indicates that the individual particles are not uniformly labeled. This is also evident, when individual particles are visualized.

Found at: doi:10.1371/journal.ppat.1000824.s005 (4.35 MB TIF)

Figure S5 PrgK resembles the smaller concentric ring of the PrgHΔ4 inner ring substructure. *En face* view of the most prominent class-average derived after hierarchical clustering from images of negatively stained inner rings substructures from PrgHΔ4 complexes. The smaller concentric ring shares similarity in dimension (~180Å) and organization to the modeled PrgK ring. In order to allow a comparison between the class average of the inner ring substructure and the modeled PrgK ring, only a segment of the latter is shown.
Found at: doi:10.1371/journal.ppat.1000824.s006 (0.53 MB TIF)

Figure S6 Ribbon diagram of modeled InvG and PrgK domains. The structure of InvG (white) and PrgK (yellow) were obtained by structure homology-modeling based on EscC (blue) and EscJ (blue) templates, using the SWISS-MODEL server (http://swissmodel.expasy.org/). Note that in EscJ Asn134 to Gln139 is not resolved in the X-ray structure, however, the corresponding amino acids in PrgK (Asp133 to Lys144) have been modeled using the the SWISS-MODEL server, indicating a possible conformation of this amino acid stretch (marked with an *). This domain (also marked with an *) is shown with a 70% transparency setting in Fig. S7.
Found at: doi:10.1371/journal.ppat.1000824.s007 (0.43 MB TIF)

Figure S7 Overview of needle complex substructures and organization. Side view of half-sectioned base and location of individual protein domain. (* indicates to a possible confirmation of amino acids 133–144 in PrgK based on modeling using the SWISS-MODEL server (Fig. S6). Concentric ring-organization of IR1/2 revealed from top-viewed disassembled complexes and ring models of PrgH and PrgK.
Found at: doi:10.1371/journal.ppat.1000824.s008 (1.48 MB TIF)

Table S1 Comparison of surface accessibility in wild type and Δ*invG* mutant complexes. Peptides from NHS-acetate derivatized complexes were identified by MS/MS sequencing and semi-quantified by MS peak integration (Text S1). The table displays the modified lysine (Lysine), the identified corresponding peptide (Peptide) and the degree of NHS-acetate modification for the wild type (%w.t.) and the *ΔinvG* mutant (%ΔinvG). The ratio of peptide acetylation is an average of two independent mass spectrometry measurements. To highlight which lysines get stronger modified in the *ΔinvG* mutant (and are most likely more surface exposed) the ratio of *ΔinvG* mutant and wild type acetylation is determined (%ΔinvG/%w.t.). Due to the presence of several lysines within the peptides "K48, K53" and the C-terminally located peptides ("C-term" = K238, K239, K246, K248) the exact position of the derivatization could not be distinguished.
Found at: doi:10.1371/journal.ppat.1000824.s009 (0.07 MB RTF)

Table S2 List of mono-links and loop-links determined by chemical derivatization and mass spectrometry. In order to specify all lysines being accessible for derivatization, hetero-cross-links (cross-link between two different proteins, see Table 1), mono-links (modified peptides that reacted with only one functional group of the bivalent cross-linker (BS2G (0/d4))) and loop-links (cross-links within one polypeptide chain or between homo-proteins) were determined. Only peptides with an xQuest score of higher than 15 are displayed as those cross-linking results are likely to be reliable (see Text S1). The respective modified amino acid is displayed in bold. Under "Monolink/Looplink" listed positions are the absolute positions of the respective protein (including signal peptides), "Error" the deviation of measured precursor ion mass from theoretical precursor ion mass in parts per million, and "Score" the xQuest score.
Found at: doi:10.1371/journal.ppat.1000824.s010 (0.10 MB RTF)

Acknowledgments

We would like to thank Christoph Stingl and Peter Pichler for their support during initial MS analysis and Günter Resch (EM facility) and all the other IMP/IMBA service facilities for their assistance.

Author Contributions

Conceived and designed the experiments: OS MDL MJB WHS JEG TCM. Performed the experiments: OS MDL MJB WHS AS JR. Analyzed the data: OS MDL MJB WHS AS JR JEG TCM. Wrote the paper: OS MDL JEG TCM. Supervised mass spectrometry: KM. Supervised project: JEG TCM.

References

1. Cianciotto NP (2005) Type II secretion: a protein secretion system for all seasons. Trends Microbiol 13: 581–588.
2. Johnson TL, Abendroth J, Hol WG, Sandkvist M (2006) Type II secretion: from structure to function. FEMS Microbiol Lett 255: 175–186.
3. Llosa M, Roy C, Dehio C (2009) Bacterial type IV secretion systems in human disease. Mol Microbiol 73: 141–151.
4. Filloux A, Hachani A, Bleves S (2008) The bacterial type VI secretion machine: yet another player for protein transport across membranes. Microbiology 154: 1570–1583.
5. Galan JE, Wolf-Watz H (2006) Protein delivery into eukaryotic cells by type III secretion machines. Nature 444: 567–573.
6. Kubori T, et al. (1998) Supramolecular structure of the Salmonella typhimurium type III protein secretion system. Science 280: 602–665.
7. Sekiya K, et al. (2001) Supermolecular structure of the enteropathogenic Escherichia coli type III secretion system and its direct interaction with the EspA-sheath-like structure. Proc Natl Acad Sci U S A 98: 11638–11643.
8. Blocker A, et al. (2001) Structure and composition of the Shigella flexneri "needle complex", a part of its type III secreton. Mol Microbiol 39: 652–663.
9. Marlovits TC, et al. (2004) Structural insights into the assembly of the type III secretion needle complex. Science (New York, NY) 306: 1040–1042.
10. Sukhan A, Kubori T, Wilson J, Galan JE (2001) Genetic analysis of assembly of the Salmonella enterica serovar Typhimurium type III secretion-associated needle complex. J Bacteriol 183: 1159–1167.
11. Marlovits TC, et al. (2006) Assembly of the inner rod determines needle length in the type III secretion injectisome. Nature 441: 637–640.
12. Sukhan A, Kubori T, Galan JE (2003) Synthesis and localization of the Salmonella SPI-1 type III secretion needle complex proteins PrgI and PrgJ. J Bacteriol 185: 3480–343.
13. Crepin VF, et al. (2005) Structural and functional studies of the enteropathogenic Escherichia coli type III needle complex protein EscJ. Mol Microbiol 55: 1658–1670.
14. Yip CK, et al. (2005) Structural characterization of the molecular platform for type III secretion system assembly. Nature 435: 702–707.
15. Spreter T, et al. (2009) A conserved structural motif mediates formation of the periplasmic rings in the type III secretion system. Nat Struct Mol Biol 16: 468–476.
16. Hodgkinson JL, et al. (2009) Three-dimensional reconstruction of the Shigella T3SS transmembrane regions reveals 12-fold symmetry and novel features throughout. Nat Struct Mol Biol 16: 477–485.
17. Hardie KR, Lory S, Pugsley AP (1996) Insertion of an outer membrane protein in Escherichia coli requires a chaperone-like protein. EMBO J 15: 978–988.
18. Koster M, et al. (1997) The outer membrane component, YscC, of the Yop secretion machinery of Yersinia enterocolitica forms a ring-shaped multimeric complex. Mol Microbiol 26: 789–797.
19. Collins RF, Davidsen L, Derrick JP, Ford RC, Tonjum T (2001) Analysis of the PilQ secretin from Neisseria meningitidis by transmission electron microscopy reveals a dodecameric quaternary structure. Journal of Bacteriology 183: 3825–3832.
20. Linderoth NA, Simon MN, Russel M (1997) The filamentous phage pIV multimer visualized by scanning transmission electron microscopy. Science 278: 1635–1638.
21. Chami M, et al. (2005) Structural Insights into the Secretin PulD and Its Trypsin-resistant Core. J Biol Chem 280: 37732–37741.
22. Kaniga K, Bossio JC, Galan JE (1994) The Salmonella typhimurium invasion genes invF and invG encode homologues of the AraC and PulD family of proteins. Mol Microbiol 13: 555–568.

23. Kimbrough TG, Miller SI (2000) Contribution of Salmonella typhimurium type III secretion components to needle complex formation. Proc Natl Acad Sci U S A 97: 11008–1113.
24. Mindell JA, Grigorieff N (2003) Accurate determination of local defocus and specimen tilt in electron microscopy. J Struct Biol 142: 334–347.
25. Rinner O, et al. (2008) Identification of cross-linked peptides from large sequence databases. Nat Methods 5: 315–318.

Molecular Cell
Article

Heterochromatin Formation in *Drosophila* Is Initiated through Active Removal of H3K4 Methylation by the LSD1 Homolog SU(VAR)3-3

Thomas Rudolph,[1] Masato Yonezawa,[2] Sandro Lein,[1] Kathleen Heidrich,[1] Stefan Kubicek,[2] Christiane Schäfer,[1] Sameer Phalke,[1] Matthias Walther,[1] Andreas Schmidt,[3] Thomas Jenuwein,[2] and Gunter Reuter[1,*]
[1] Institute of Biology, Developmental Genetics, Martin Luther University Halle, D-06120 Halle, Germany
[2] Research Institute of Molecular Pathology
[3] Department of Biochemistry, Christian-Doppler Laboratory for Proteome Analysis, University of Vienna
The Vienna Biocenter, A-1030 Vienna, Austria
*Correspondence: reuter@genetik.uni-halle.de
DOI 10.1016/j.molcel.2007.02.025

SUMMARY

Epigenetic indexing of chromatin domains by histone lysine methylation requires the balanced coordination of methyltransferase and demethylase activities. Here, we show that SU(VAR)3-3, the *Drosophila* homolog of the human LSD1 amine oxidase, demethylates H3K4me2 and H3K4me1 and facilitates subsequent H3K9 methylation by SU(VAR)3-9. *Su(var)3-3* mutations suppress heterochromatic gene silencing, display elevated levels of H3K4me2, and prevent extension of H3K9me2 at pericentric heterochromatin. SU(VAR)3-3 colocalizes with H3K4me2 in interband regions and is abundant during embryogenesis and in syncytial blastoderm, where it appears concentrated at prospective heterochromatin during cycle 14. In embryos of *Su(var)3-3*/+ females, H3K4me2 accumulates in primordial germ cells, and the deregulated expansion of H3K4me2 antagonizes heterochromatic H3K9me2 in blastoderm cells. Our data indicate an early developmental function for the SU(VAR)3-3 demethylase in controlling euchromatic and heterochromatic domains and reveal a hierarchy in which SU(VAR)3-3-mediated removal of activating histone marks is a prerequisite for subsequent heterochromatin formation by H3K9 methylation.

INTRODUCTION

Covalent modifications of histone N termini control chromatin structure and define epigenetic marks for transcriptional competence during development. Methylation of histone H3 lysine 9 (H3K9) is a hallmark of constitutive heterochromatin (Rea et al., 2000; Nakayama et al., 2001; Peters et al., 2002; Schotta et al., 2002), whereas methylation of H3K4 is associated with actively transcribed chromatin (Zhang and Reinberg, 2001; Bernstein et al., 2002; Kouzarides, 2002). In eukaryotes, a large fraction of the genome is in an inaccessible heterochromatic state. Factors defining heterochromatin were first identified in *Drosophila* by genetic dissection of gene silencing in position-effect variegation (PEV) with the help of dominant modifier mutations (Schotta et al., 2003), so-called suppressors of variegation or *Su(var)* genes. The currently characterized *Su(var)* genes encode either heterochromatin components or factors that impair spreading of heterochromatin into euchromatin (Ebert et al., 2006). For example, the H3K9 methyltransferase SU(VAR)3-9, its associated H3K9 methyl binder HP1, and the Zn-finger protein SU(VAR)3-7 (Jaquet et al., 2002) work together in a complex to maintain stable gene silencing in heterochromatin (Ebert et al., 2006).

At the end of embryonic development, which is syncytial in *Drosophila*, the transition of cleavage chromatin into somatic and germline chromatin occurs during or shortly before blastoderm formation. Heterochromatin first becomes visible at the apical pole of early blastoderm nuclei as an intensely staining region when chromosomes are organized in a typical Rabl conformation (Foe et al., 1993). The molecular mechanisms that initiate heterochromatin formation during early development are currently unknown.

Based on its thermodynamic stability, histone lysine methylation has been regarded as a very robust histone modification that would be ideally suited to impart more long-term epigenetic information. With the discoveries of histone lysine demethylases (Shi et al., 2004; Tsukada et al., 2006), the epigenetic stability of histone lysine methylation has to be reconsidered by comparing the activities of methyltransferases (HMTase) and demethylases that either target activating (e.g., H3K4) or repressive (e.g., H3K9) histone modifications (Bannister and Kouzarides, 2005). For example, the human amine oxidase LSD1 catalyzes removal of methyl groups from H3K4me1 and H3K4me2 (Shi et al., 2005) and, when complexed with

other factors, also destabilizes H3K9me2 (Metzger et al., 2005). LSD1 function is present in unicellular organisms, and highly related genes are conserved throughout multicellular genomes. However, with the exception of knockdown studies in *C. elegans* (Shi et al., 2004) and a recent study in *S. pombe* (Nicolas et al., 2006), no in vivo analysis for H3K4 demethylation during metazoan development has been described.

Here, we show that the dominant PEV modifier *Su(var)3-3* encodes the *Drosophila* homolog of LSD1 and is an H3K4 demethylase that associates with prospective heterochromatin in early embryonic development. The SU(VAR)3-3 enzyme directly restricts the extent of H3K4me2 at heterochromatic sequences in somatic cells and safeguards transcriptional silence in primordial germ cells. Loss of the SU(VAR)3-3 demethylase indirectly impairs the full expansion of H3K9me2 marks at heterochromatin. SU(VAR)3-3 is an abundant nuclear protein at cleavage stage embryos. In syncytial blastoderm, it is restricted to heterochromatic regions and complexes with other silencing factors including the SU(VAR)3-9 HMTase, HP1, and the histone deacetylase RPD3. Analysis of *Su(var)3-3* mutant embryos and flies suggests a function for SU(VAR)3-3 in the protection of heterochromatic domains against antagonistic H3K4 methylation. Together, our data define an early developmental role for the SU(VAR)3-3 demethylase in initiating heterochromatin formation and gene silencing in *Drosophila*.

RESULTS

The *Drosophila* Homolog of LSD1 Is the Dominant PEV Modifier *Su(var)3-3*

Su(var)3-3 mutations belong to the strongest dominant suppressors for *white* gene silencing in the w^{m4} PEV rearrangement (Wustmann et al., 1989). *Su(var)3-3* mutations are characterized by a number of interesting phenotypic effects. Homozygous mutant females produce no oocytes (agametic), whereas in males only immobile sperms are found (Szabad et al., 1988). These mutations are also butyrate sensitive and display recessive lethal interactions with additional heterochromatin, such as an extra copy of the Y chromosome in XXY females or in XYY males (Reuter et al., 1982).

We mapped *Su(var)3-3* to polytene chromosome region 77A3 by male recombination and complementation analysis examining recessive female and male sterility with a series of deletions (Ryder et al., 2004; see Figure S1 in the Supplemental Data available with this article online). This region contains the two genes CG17147 and CG17149 (Crosby et al., 2007). Genetic analysis proves that CG17149 is identical with *Su(var)3-3*. The PiggyBac insert *PBac{WH}f00678* in CG17149 (Parks et al., 2004) is allelic to *Su(var)3-3*. *Su(var)3-3* encodes a gene product that shows high amino acid conservation to the human demethylase LSD1, including 71% identity in the SWIRM domain and 75.3% in the amine oxidase domain (Figure 1A).

All 22 *Su(var)3-3* mutations were isolated as dominant suppressors of *white* variegation in w^{m4}, and 15 alleles displayed mutational lesions in the protein-coding sequences (Figure 1A). In the remaining seven alleles, no transcript is detected by RT-PCR. Phenotypic rescue of the dominant suppressor effect of *Su(var)3-3* heterozygotes is restored by an extra gene copy of a wild-type allele of *Su(var)3-3* (Figure 1C). Genetic crosses demonstrate that the *Su(var)3-3* mutations also strongly suppress heterochromatic PEV in other rearrangements (Figure 1B), such as *brown*, *Stubble*, *yellow*, or a *lacZ* transgene (Sinclair et al., 1992; Le et al., 1995; Lu et al., 1998). The general suppressor effect of *Su(var)3-3* mutations in PEV illustrates its crucial requirement for heterochromatic gene silencing.

H3K9 methylation catalyzed by the evolutionarily conserved HMTase SU(VAR)3-9 is a hallmark for constitutive heterochromatin in all eukaryotic cells (Rea et al., 2000; Nakayama et al., 2001; Schotta et al., 2002). The strength and extent of gene silencing in PEV are directly correlated with the dosage of the heterochromatin-associated factors SU(VAR)3-9, HP1, and SU(VAR)3-7 (Ebert et al., 2004). Additional genomic copies of all three genes cause strong enhancement of *white* gene silencing in w^{m4} (Figure 1C). *Su(var)3-3* mutations dominate in their suppressor effect the PEV enhancement induced by additional *Su(var)3-9*, *Su(var)2-5* (HP1), and *Su(var)3-7* gene copies. Even in w^{m4};*Su(var)3-3*/+ flies carrying two additional wild-type copies of *Su(var)3-9*, enhancement of *white* gene silencing is strongly suppressed (Figure 1C). This epistatic effect of *Su(var)3-3* mutations on PEV enhancement suggests that SU(VAR)3-3 functions upstream of the SU(VAR)3-9 H3K9 methyltransferase in control of heterochromatin formation.

SU(VAR)3-3 Is an H3K4 Demethylase that Also Affects H3K9 Methylation

Histone demethylase activity of SU(VAR)3-3 and of the mouse LSD1 homolog was studied using baculovirus-expressed proteins and calf thymus histones by western blot with methyl-lysine-specific histone antibodies. Like LSD1 (Shi et al., 2004), recombinant SU(VAR)3-3 specifically reduces H3K4me1 and H3K4me2, but not H3K9 (Figure 2) or any of the other known methylation sites (data not shown). No enzymatic activity is detected for a point mutant protein (G316R) that impairs the amine oxidase domain. This mutation is identical to the $Su(var)3\text{-}3^{09}$ mutant, which is a strong silencing suppressor in vivo. We next confirmed the demethylase activity by mass spectrometry using recombinant SU(VAR)3-3 and histone H3 peptides (1–20) which were either mono-, di-, or trimethylated at H3K4. SU(VAR)3-3 demethylates H3K4me1 and H3K4me2 but, as expected, cannot convert H3K4me3 and H3K9me1 or me2 (Figure 2D). These enzymatic data demonstrate that *Drosophila* SU(VAR)3-3 and mammalian LSD1 are functionally conserved histone H3K4 demethylases.

To address the function of SU(VAR)3-3 in vivo, we first examined the chromosomal distribution of the protein

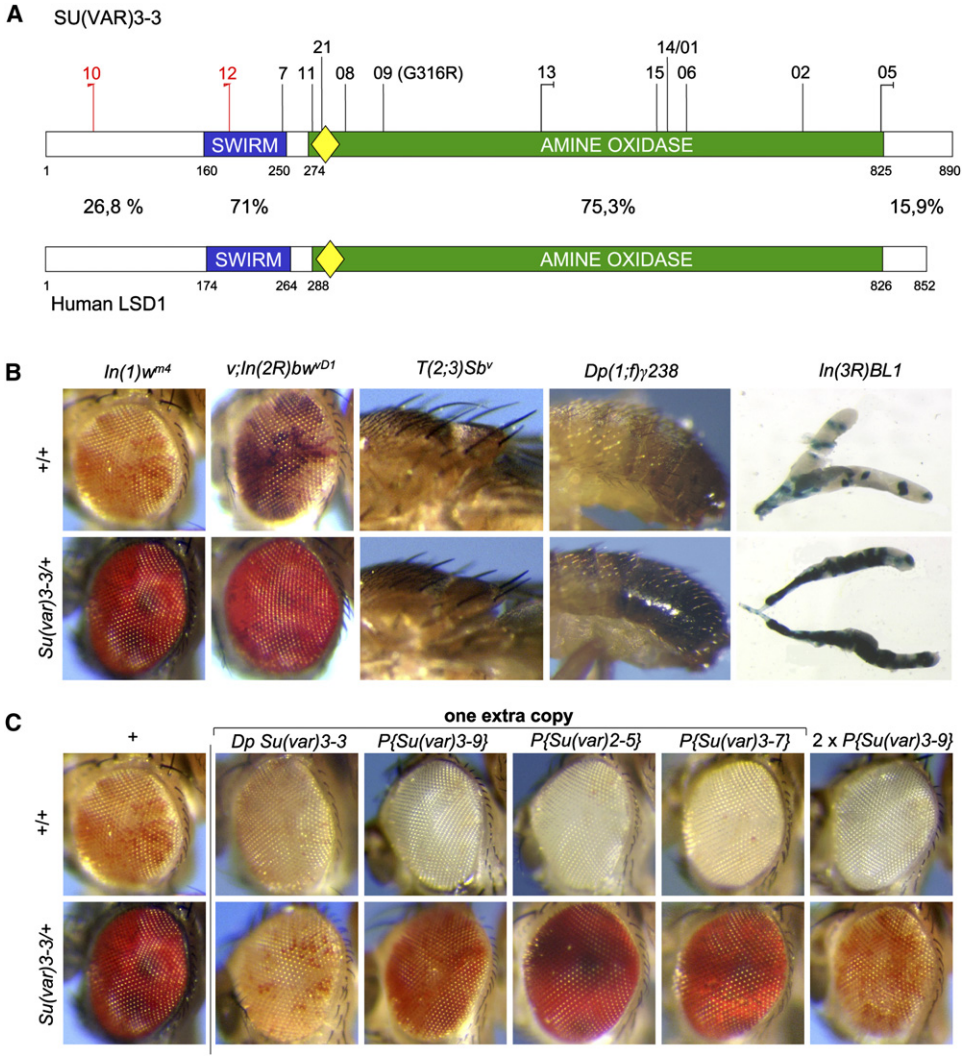


Figure 1. The *Drosophila* Homolog of Histone Demethylase LSD1 Is the PEV Suppressor SU(VAR)3-3

(A) Protein structure of SU(VAR)3-3 and positions of the molecularly characterized mutations. SU(VAR)3-3 shows high identity with the human LSD1 protein, with 71% in the SWIRM and 75.3% in the amine oxidase domain. Yellow lozenge indicates the FAD binding domain.

(B) SU(VAR)3-3 is required for heterochromatic gene silencing. All *Su(var)3-3* mutations were isolated as dominant suppressors of white variegation in w^{m4}. They also suppress gene silencing in other PEV rearrangements (*brown*, *Stubble*, *yellow*, *lacZ*).

(C) Genomic extra copies of *Su(var)3-9*, *Su(var)2-5*, and *Su(var)3-7* display enhanced silencing of the *white* gene in *In(1)*w^{m4} (upper row), which is impaired in the presence of the strong PEV suppressor mutant *Su(var)3-3* (lower row). Rescue of the dominant suppressor effect of *Su(var)3-3* mutations is found in w^{m4} flies heterozygous for a *Su(var)3-3* mutation and a duplication (second column).

along polytene chromosomes using a SU(VAR)3-3-specific antibody (Figure 3A). In salivary gland chromatin, SU(VAR)3-3 is exclusively found at euchromatin and preferably associates with interband regions that also accumulate H3K4me2 (Ebert et al., 2006). Our immunocytology indicates significant colocalization of SU(VAR)3-3 and H3K4me2 at interbands (Figure S2). Immunocytology in *Su(var)3-3* null larvae revealed elevated H3K4me2 levels and reduction in heterochromatic H3K9me2 (Figure 3B). Western blot analysis proved elevated H3K4me2 and reduced H3K9me2 levels (Figure S3). The apparent interdependent shift between increased H3K4me2 and decreased H3K9me2 in *Su(var)3-3* null larvae suggests that SU(VAR)3-3 not only directly regulates H3K4me2 but also indirectly affects dimethylation of H3K9. In addition, an interesting change in the distribution of H3K9me3 is detected in *Su(var)3-3* null larvae, which, compared to wild-type larvae, display many sites of H3K9me3 along all chromosomes (Figure 3B).

We also generated transgenic lines expressing SU(VAR)3-3 under control of the yeast UAS promotor. Forced expression of SU(VAR)3-3 in development was controlled by different GAL4 driver elements, including tubulin-GAL4 for global and SGS3-GAL4 for salivary

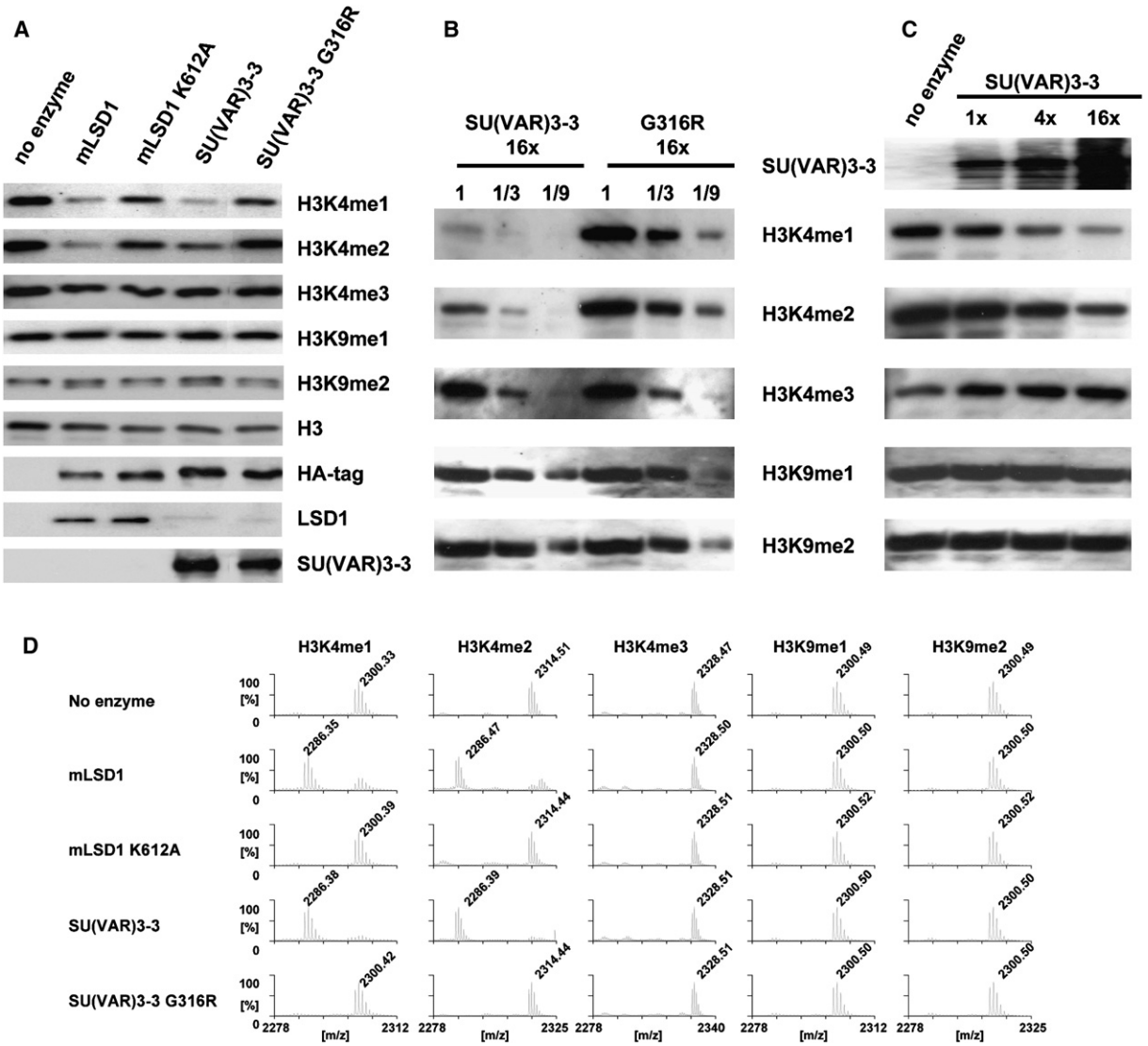


Figure 2. SU(VAR)3-3 Demethylates In Vitro Mono- and Dimethyl H3-K4

Western blot analysis with α-mono-, α-di-, and α-trimethyl H3K4; α-mono- and α-dimethyl H3K9; and α-H3, α-HA, α-LSD1, and α-SU(VAR)3-3.

(A) Histone demethylation assays on bulk histones without enzyme and with mLSD1, mLSD K612A, SU(VAR)3-3, and SU(VAR)3-3 G316R recombinant protein.

(B) Western blot analysis of serial dilutions of the reaction products from HDM assays on bulk histones with SU(VAR)3-3 wild-type and G316R mutant protein. About 80% of H3K4me1 and H3K4me2 is removed from bulk histones upon incubation with SU(VAR)3-3. 1 corresponds to 4.5 μl of reaction products.

(C) Concentration-dependent H3K4 demethylation activity of SU(VAR)3-3 on bulk histones. 1x corresponds to 2 μg of SU(VAR)3-3 protein.

(D) Mass spectrometric analyses of H3K4me1, -me2, and -me3 and H3K9me1 and -me2 levels in synthetic H3 peptides after incubation without enzyme and with mLSD1, mLSD K612A, SU(VAR)3-3, and SU(VAR)3-3 G316R.

gland-specific overexpression. Forced expression of SU(VAR)3-3 in salivary glands results in reduction of H3K4me2 at interbands in polytene chromosomes (Figure 3C), whereas the level of H3K4me3 is not changed (data not shown). Interestingly, we also detected reduced staining for H3K9me2 and H3K9me3 in chromocenter heterochromatin after overexpression (Figure 3C). Although SU(VAR)3-3 cannot be detected at the heterochromatic chromocenter of wild-type polytene chromosomes, forced overexpression results in weak association of SU(VAR)3-3 with this domain (Figure 3C).

We next examined these global changes in H3K4 and H3K9 methylation on gene silencing in PEV. Since all transgene constructs contain the *mini-white* marker, we studied the effect of ectopic SU(VAR)3-3 expression on PEV in the *T(2;3)Sb*^{v} rearrangement by measuring frequency of *Stubble* (*Sb*) mutant gene inactivation (Sinclair et al., 1992). Reduction of H3K9 methylation after ectopic

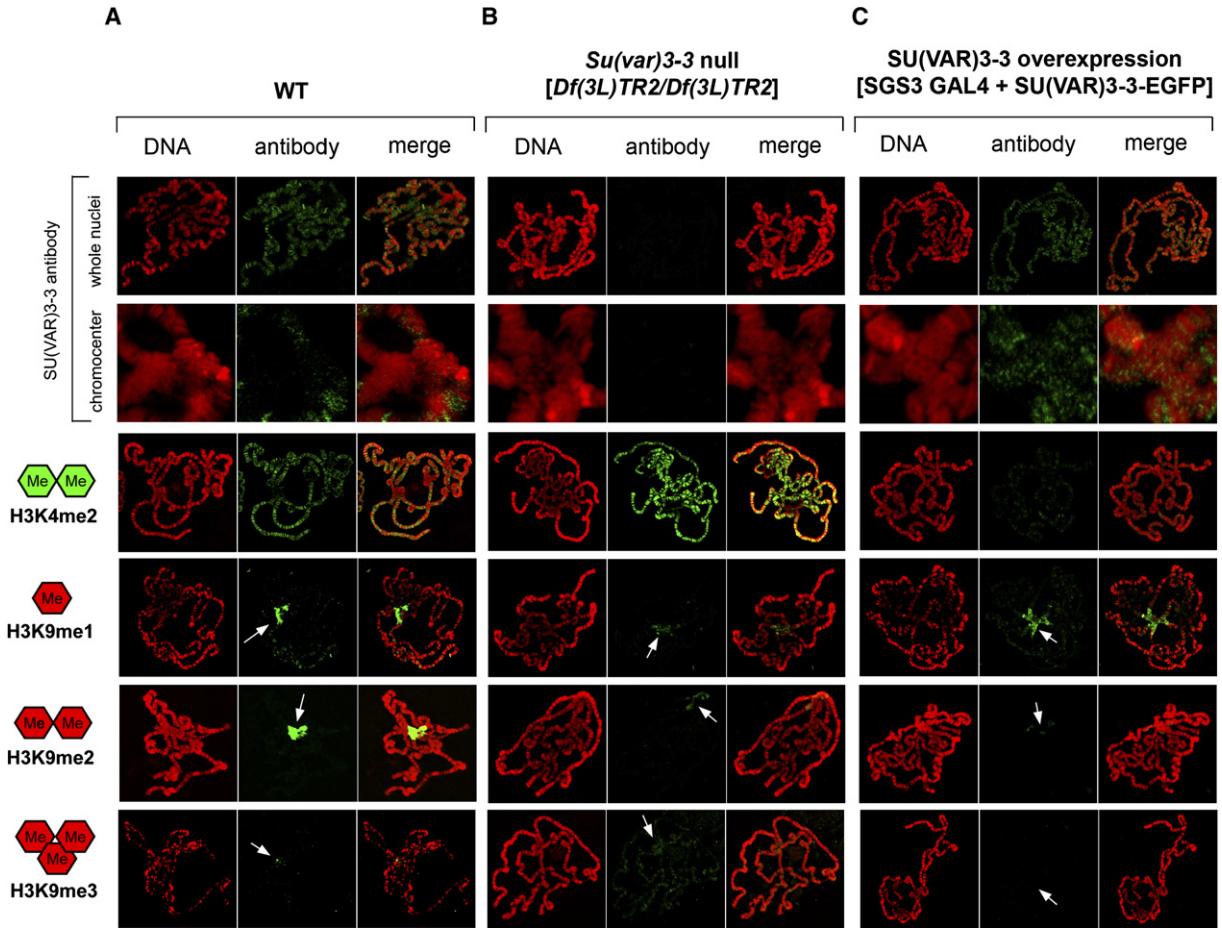


Figure 3. SU(VAR)3-3 Is a H3K4 Demethylase that Shows Ectopic Effects on In Vivo H3K9 Methylation

(A) Salivary gland polytene chromosomes from wild-type larvae stained with α-SU(VAR)3-3, α-dimethyl H3K4, and α-mono-, α-di-, or α-trimethyl H3K9.

(B) Histone methylation marks in *Su(var)3-3* null polytene chromosomes. A loss of SU(VAR)3-3 causes chromosomewide increase in H3K4me2 and decreases the amount of H3K9 di- and trimethylation in chromocenter heterochromatin. Immunostaining with α-SU(VAR)3-3, α-dimethyl H3K4, and α-mono-, α-di-, or α-trimethyl H3K9.

(C) Ectopic effect on histone methylation after SU(VAR)3-3EGFP overexpression using a GAL4-SGS3 driver element results in heterochromatin association of SU(VAR)3-3EGFP, strong reduction of H3K4me2, and H3K9me2, and -me3. H3K9 monomethylation is not affected. Immunolabeling with α-SU(VAR)3-3, α-dimethyl H3K4, and α-mono-, α-di-, or α-trimethyl H3K9. DNA was stained with DAPI (red). Arrows point to chromocenter region.

SU(VAR)3-3 overexpression coincides with significant suppression of PEV in *T(2;3)Sb^{v}*, which is reflected by an increase in the number of *Stubble* mutant bristle to 83.3% as compared to 51.7% in control flies. In contrast to forced SU(VAR)3-3 overexpression, controlled expression by additional genomic copies of *Su(var)3-3* results in weak enhancement of PEV. Strong suppression of PEV is the typical loss-of-function phenotype of *Su(var)3-3* (Figure 1C). Therefore, a dose-dependent activity of SU(VAR)3-3 is required for establishment of heterochromatic gene silencing.

Heterochromatin Formation during Embryonic Development Depends on SU(VAR)3-3

To investigate the function of SU(VAR)3-3 in heterochromatin formation during early embryonic development, we first studied nuclear distribution of the protein in wild-type embryos and in embryos produced from females carrying either the loss-of-function *Su(var)3-3^{12}* allele or a deletion of the gene. We analyzed distribution of the H3K4me2, H3K9me2, and H3K9me3 marks during embryonic development in somatic as well as in germline precursor cells (Figure 4). In early cleavage, SU(VAR)3-3 displays a uniform nuclear distribution (Figure 4A). During the cleavage divisions, the protein becomes redistributed (Figure 4B). At metaphase, a significant amount of SU(VAR)3-3 is detected at the chromosome periphery, whereas it localizes between the chromosomes during anaphase. A rather uniform interphase chromatin association of SU(VAR)3-3 is detected until syncytial blastoderm (Figure 4A). When cellular blastoderm is formed (cycle 14), SU(VAR)3-3 becomes preferably associated

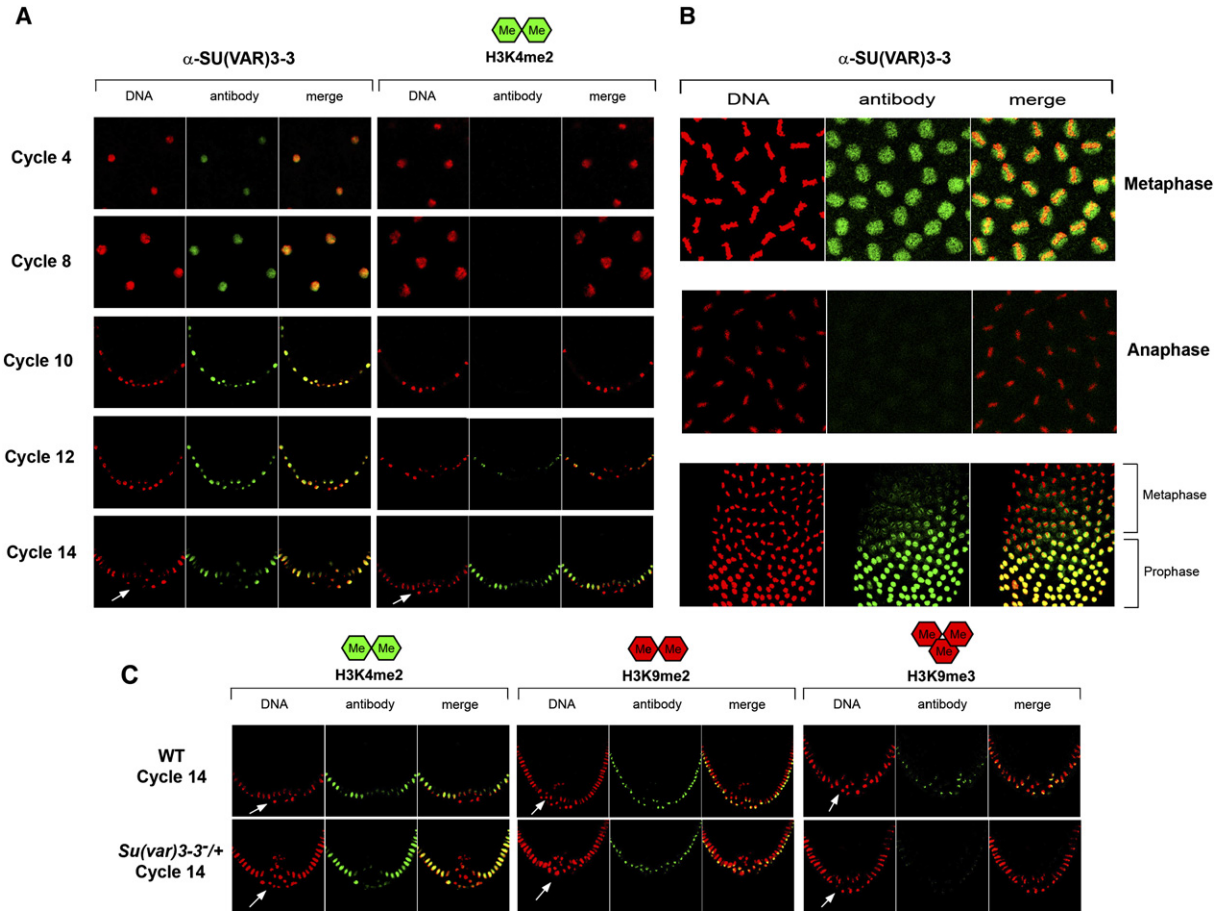


Figure 4. Chromatin Association of SU(VAR)3-3 and Distribution of Histone Modification Marks at Cleavage and during Development of Somatic and Germline Precursor Cells

(A) During cleavage, SU(VAR)3-3 shows a uniform chromatin association until cycle 12. H3K4 dimethylation first becomes visible at the end of cleavage (cycle 12) and is excluded from the heterochromatic compartment and pole cells, the primordial germ cells of *Drosophila*. At cycle 14, SU(VAR)3-3 becomes enriched at prospective heterochromatin and remains uniformly associated with pole cell chromatin. H3K4 dimethylation is excluded from heterochromatin and pole cell chromatin.

(B) Redistribution of SU(VAR)3-3 during cleavage mitosis. The protein becomes accumulated at the chromosomal periphery during metaphase, whereas only weak uniform staining is detected during anaphase.

(C) H3K4me2, H3K9me2, and H3K9me3 staining in wild-type embryos and embryos produced by *Su(var)3-3*/+ females. Increase of H3K4me2 in blastoderm cells and significant H3K4me2 staining in pole cells of mutant embryos. Concomitant reduction of heterochromatic H3K9 di- and trimethylation in somatic and pole cell nuclei is observed. Arrowheads point to pole cells.

with pericentric heterochromatin (Figures 4A and 5) that is located in a typical Rabl conformation as DAPI bright material situated at the apical pole of blastoderm nuclei (Foe et al., 1993). Later in development, SU(VAR)3-3 is only found over euchromatin. In the germline precursor pole cells, SU(VAR)3-3 remains uniformly distributed within the nucleus (Figure 4A).

Heterochromatin association of SU(VAR)3-3 in blastoderm nuclei coincides with the appearance of H3K4 methylation. H3K4 methylation is first visible at syncytial blastoderm (cycle 12) in a speckled distribution (Figure 4C). At cycle 14, H3K4 methylation becomes restricted to the euchromatic compartment of the nuclei. At this time, SU(VAR)3-3 is enriched at pericentric heterochromatin, and all H3K9 methylation states appear in this domain. In contrast to later stages of *Drosophila* development (Ebert et al., 2004), early heterochromatin shows more intense staining for H3K9me3 (Figure 5B). In pole cells, intense staining for H3K9me2 was found (Schaner et al., 2003). Indeed, all H3K9 methylation states can be detected in pole cells (Figure 4 and data not shown).

To study the role of SU(VAR)3-3 in heterochromatin formation during early embryonic development and in particular at cycle 14, we analyzed embryos produced after a cross of females and males heterozygous for a *Su(var)3-3* null mutation. In most of these embryos, the amount of SU(VAR)3-3 over heterochromatin is significantly reduced (Figure 5A). These embryos show a more extended

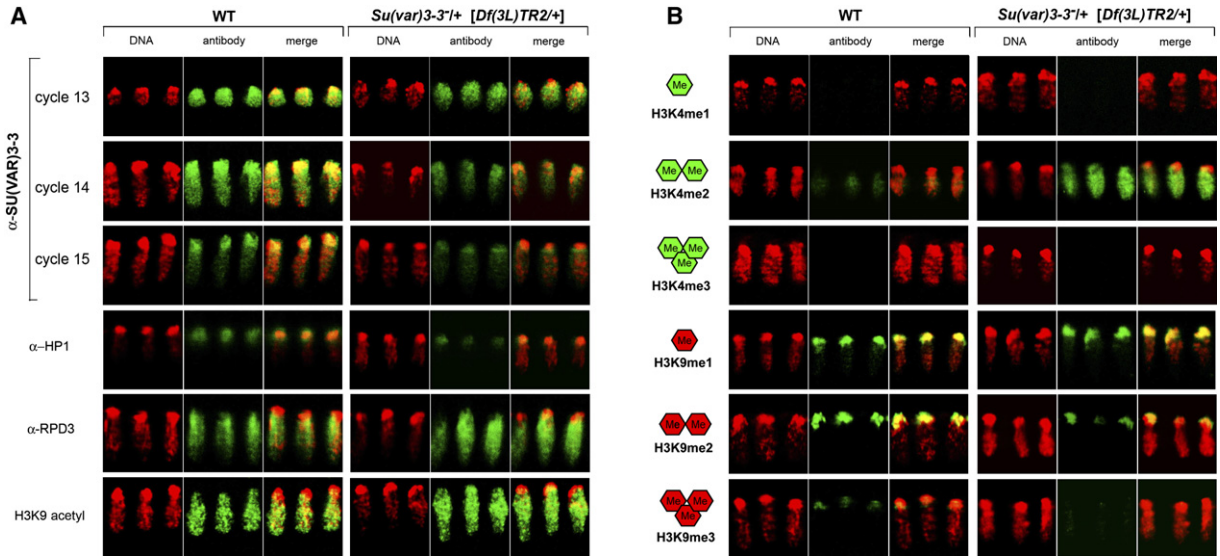


Figure 5. High-Resolution Analysis of SU(VAR)3-3, HP1, and RPD3 Binding and Histone Modifications in Early Blastoderm Cells

(A) Immunostaining of wild-type (left panel) and embryos of *Su(var)3-3*$^{-}$/+ females (right panel) with α-SU(VAR)3-3, α-HP1, α-RPD3, and α-acetyl H3K9. SU(VAR)3-3 distribution is shown for cycles 13–15 as indicated. All other photos represent cycle 14 embryos. In wild-type embryos, SU(VAR)3-3 is uniformly distributed during cycle 13, becomes enriched in heterochromatic regions during cycle 14, and afterwards is accumulated at the boundary region between eu- and heterochromatin (cycle 15). SU(VAR)3-3 and HP1 staining is significantly reduced in mutant embryos. RPD3 binding is almost uniform along the apicobasel axis of blastoderm nuclei in both wild-type and mutant embryos. In contrast, extended acetyl H3K9 staining toward heterochromatin is found in mutant embryos.

(B) Compared to wild-type (left panel) in *Su(var)3-3* mutant embryos (right panel), the euchromatic mark H3K4me2 is elevated and spreads into the heterochromatic compartment. In mutant embryos, the heterochromatic trimethyl H3K9 is significantly reduced. No effects are found on mono- and trimethyl H3K4 and monomethyl H3K9. H3K9me2 appears only weakly reduced in *Su(var)3-3* mutant embryos.

staining of H3K4me2 at the prospective heterochromatin compartment (Figure 5B). Conversely, H3K9me2 and staining for HP1 appear to be reduced. Distribution of H3K9me3 is markedly affected and now shows foci throughout the nuclei. The amount of H3K9me1 appears not to be affected. Western blot analysis of 0–4 hr old embryos also revealed increase in global H3K4me1 and H3K4me2 levels. In these collections, which contain only a fraction of cycle 13–14 embryos, reduction of H3K9me3 but no obvious global change in H3K9me2 is detected, suggesting that mainly local concentration of these marks at prospective heterochromatin in early embryos controls heterochromatin formation (Figure S3B). Furthermore, immunostaining for H3K9 acetylation in cycle 14 embryos shows expansion of H3K9ac distribution toward heterochromatin. These data indicate that both demethylation of H3K4 by SU(VAR)3-3 and deacetylation of H3K9 by RPD3 are coordinated during chromatin differentiation in early embryogenesis.

In *Drosophila*, the primordial germ cells become separated from the soma at the posterior end of the embryo as pole cells at the end of cleavage (cycle 10). Pole cell nuclei are devoid of any H3K4me2 (Schaner et al., 2003) and show intense staining for H3K9me2 and H3K9me3 (Figure 4C). In *Su(var)3-3* mutant embryos, significant H3K4 methylation is found. Appearance of H3K4 methylation coincides with significant reduction of H3K9me2 and H3K9me3 (Figure 4C). This demonstrates that SU(VAR)3-3 also plays an essential role in establishment of transcriptional silence in pole cells. Together, this mutant analysis suggests a developmentally regulated function for SU(VAR)3-3 in protecting heterochromatic regions against expansion of H3K4 methylation during early stages of *Drosophila* development.

SU(VAR)3-3 Associates with Other Silencing Factors and Controls Heterochromatin Spreading in PEV

Histone H3K4 and H3K9 methylation represent antagonistic indexing marks. Similar to mammalian SUV39H1 (Nishioka et al., 2002), *Drosophila* SU(VAR)3-9 HMTase activity is inhibited by pre-existing methylation of H3K4, as observed in in vitro methylation assays with H3K4me1 and H3K4me2 histone H3 peptides (1–20) (Figure 6A). Addition of recombinant SU(VAR)3-3 relieves this inhibition (Figure 6A). To examine if the SU(VAR)3-3 demethylase may associate with the SU(VAR)3-9 HMTase in vivo, we analyzed extracts from early embryos by coimmunoprecipitation using a protein A Sepharose-coupled SU(VAR)3-3-specific antibody. Together with SU(VAR)3-3, three proteins, SU(VAR)3-9, HP1, and RPD3, are coimmunoprecipitated (Figure 6B). These data indicate the existence of a silencing complex, in which the SU(VAR)3-3 demethylase and the RPD3 histone deacetylase can provide activities in resetting pre-existing chromatin modifications

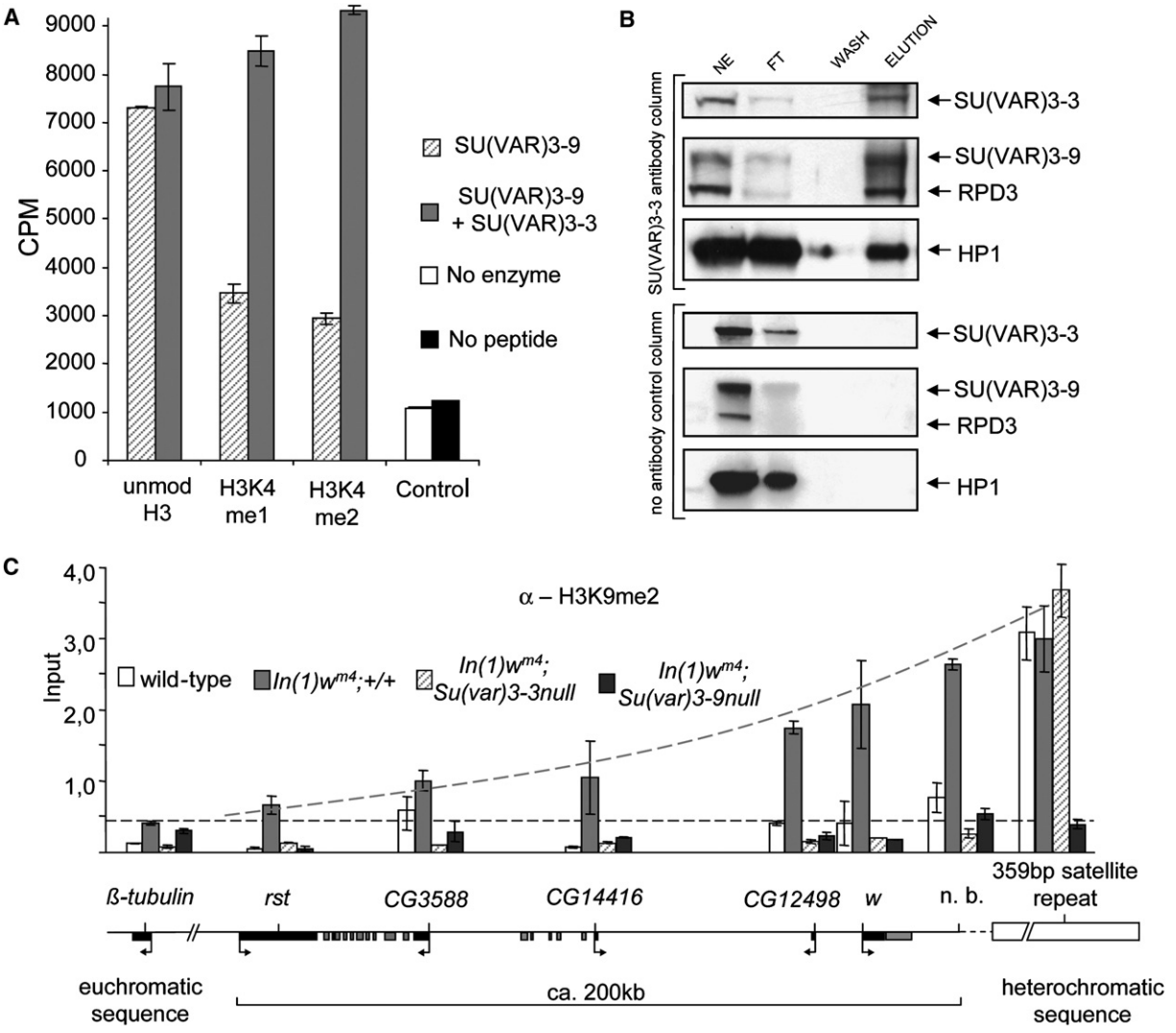


Figure 6. SU(VAR)3-3 Abolishes the Inhibitory Effect of K4-Methylated H3 Substrates on SU(VAR)3-9 Lysine 9 Methylation Activity, Associates with SU(VAR)3-9/HP1 Heterochromatin Complexes, and Impairs Heterochromatin Spreading in PEV

(A) HMTase assay with SU(VAR)3-9 and H3K4-methylated peptides (1–20) after incubation with SU(VAR)3-3. The amount of incorporated label was measured by scintillation counting after 20 min (error bars indicate minimum and maximum values).

(B) Affinity purification of SU(VAR)3-3 complexes from early (0–3 hr) embryos using a covalently Sepharose A-coupled SU(VAR)3-3 antibody. Immunoblotting of copurified proteins with α-SU(VAR)3-3, α-SU(VAR)3-9, α-HP1, and α-RPD3.

(C) *Su(var)3-3* mutations impair heterochromatin spreading. ChIP analyses along the *white-roughest* euchromatic region juxtaposed in w^{m4} to pericentric heterochromatin.

In w^{m4};+/+ flies, a gradient of H3K9me2 along the *white-roughest* region is found (gray bars), whereas no H3K9 dimethylation is detected over this region at its normal position in the X chromosome (white bars). In *Su(var)3-9* null mutant flies, H3K9me2 is not induced (black bars), whereas in *Su(var)3-3* null flies, spreading of H3K9me2 is blocked (gray shaded bars). Error bars indicate standard deviation.

for the subsequent methylation and binding of H3K9 by SU(VAR)3-9 and HP1. A requirement for H3K9 deacetylation by HDAC1 in a sequential H3K9 methylation by SU(VAR)3-9 to establish heterochromatic gene silencing was already shown (Czermin et al., 2001).

SU(VAR)3-3 association with SU(VAR)3-9 indicates that spreading of H3K9 methylation along euchromatic regions depends on the activity of SU(VAR)3-3. To analyze this in more detail, we performed chromatin immunoprecipitation (ChIP) for H3K9me2 across the euchromatic region flanking the w^{m4} rearrangement. In w^{m4}, the *white-roughest* region is located next to pericentric heterochromatin, and heterochromatization results in variable silencing of the genes located in this euchromatic region. The ChIP analysis was done with chromatin material prepared from adult fly heads and using the heterochromatic 395 bp satellite and euchromatic β-tubulin gene sequences as controls. In the w^{m4} rearrangement, a gradient of

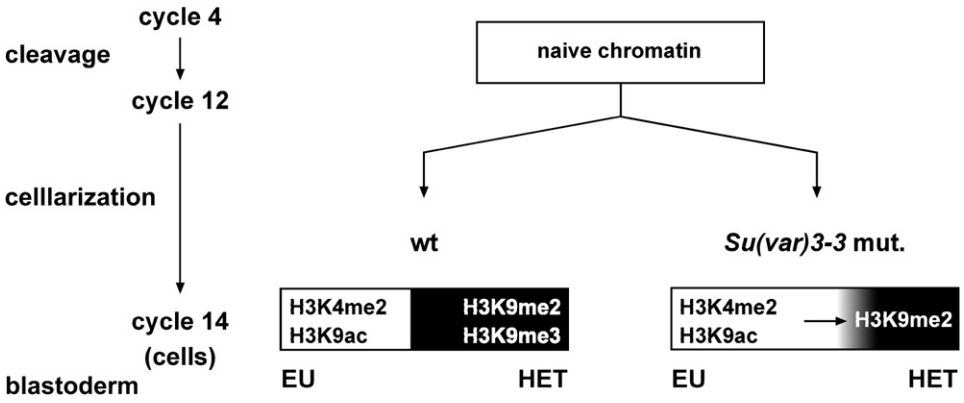


Figure 7. Coordinated Histone Demethylation and Methylation Define the Boundary between Eu- and Heterochromatin in Early Embryonic Development

SU(VAR)3-3, the *Drosophila* homolog of human LSD1, associates in early embryonic development with heterochromatin. Protection of heterochromatin against the active histone methylation marks H3K4me1 and -me2 by the H3K4 demethylase SU(VAR)3-3 is a prerequisite for heterochromatin formation and establishment of heterochromatic H3K9 methylation by SU(VAR)3-9. The H3K4 demethylase SU(VAR)3-3 associates with the SU(VAR)3-9 complex for coordinated function in determination of the boundary between heterochromatic and euchromatic chromatin domains.

H3K9me2 along the *white-roughest* region is detected, whereas no significant H3K9me2 is present across this region in its wild-type chromosomal configuration (Figure 6C). H3K4 methylation could not be detected within the *white-roughest* region in adult heads and salivary glands (data not shown), suggesting that the genes located in this region are not expressed in the majority of cells within the analyzed tissues. The amount of H3K9me2 in the *white-roughest* region within w^{m4} strongly depends on SU(VAR)3-9. In *Su(var)3-9* null mutant flies, no spreading of H3K9me2 is found (Figure 6C), whereas in flies with additional *Su(var)3-9* gene copies, regional expansion of H3K9me2 is significantly enhanced (data not shown). Importantly, in *Su(var)3-3* null flies, spreading of heterochromatin is impaired and no H3K9me2 is detected along the *white-roughest* region in w^{m4} (Figure 6C). This finding indicates that SU(VAR)3-3 is required to allow extended decoration of H3K9me2 into the flanking euchromatic region.

Early Initiation and Maintenance of Heterochromatic Gene Silencing in PEV

In early embryos, SU(VAR)3-3 associates with RPD3, SU(VAR)3-9, and HP1. We therefore investigated by clonal analysis whether initiation of heterochromatic gene silencing specifically occurs early in development. In w^{m4} *Su(var)3-9*/+ heterozygotes, Su^+/Su^+ cell clones were induced with the flipase/FRT mitotic recombination system (Xu and Rubin, 1993), and their phenotype was inspected in adult flies for restored *white* variegation. Flipase expression under the control of the *eyeless* promotor (Newsome et al., 2000) results in eye clones whose size depends on the timing of their induction (Figure S4). Although in all Su^+/Su^+ clones two wild-type copies of *Su(var)3-9* are present, *white* variegation is restored only in large clones, which were induced early in development (Figure S4A). All smaller clones, which are induced at later developmental time points, show a suppressor phenotype. Their indicative brownish color is due to homozygosity of the *red* eye marker mutation (Figure S4B). The size of the white mottled clones corresponds to the size of clones induced by X-ray treatment of w^+/w embryos or early first-instar larvae (Becker, 1966), suggesting that suppression of heterochromatic gene silencing by a *Su(var)3-9* mutation in w^{m4} is established early and then stably maintained.

Reversion of suppressed silencing could also depend on the number of cell divisions. We therefore induced Su^+/Su^+ clones in *Su(var)3-9*/+ heterozygotes carrying a dominant *Minute* mutation [$M(3)w^{124}$] on the *Su(var)3-9* mutant chromosome. The Su^+/Su^+ clones automatically convert to $Minute^+$ (Figure S4) and overgrow larger territories, since *M*/+ neighboring cells are significantly delayed in development (Garcia-Bellido et al., 1973). Late-induced Su^+/Su^+ clones frequently occupy the entire eye. Because these late clones still manifest a suppressor phenotype, additional cell divisions do not interfere with maintenance of the established suppressor effect (Figures S4C and S4D). The size of the Su^+/Su^+ clones reverting back to a white mottled phenotype suggests that heterochromatic gene silencing in w^{m4} is already defined at embryogenesis or first larval instar stage and than remains stably maintained.

DISCUSSION

Our data support a model in which heterochromatin formation and gene silencing in PEV are defined during early embryonic development of *Drosophila*. A dynamic balance between HMTases and demethylases controls establishment of the functionally antagonistic histone H3K4 and H3K9 methylation marks at the border region of euchromatin and heterochromatin (Figure 7). In transcriptionally silent cleavage nuclei, chromatin is in a naive

state with only little H3K9me2 and with H3K4 methylation completely missing. A dramatic transition of chromatin structure occurs during blastoderm formation and cellularization by establishing H3K4 and H3K9 methylation. In contrast to H3K9 acetylation, which is already found in cleavage chromatin, H3K4 methylation at prospective euchromatin appears first at the end of cleavage in cycle 12. In parallel, di- and trimethylation of H3K9 and HP1 binding establish heterochromatin. Pole cells, which are the primordial germ cells of *Drosophila*, are in a transcriptionally silent state and show extensive H3K9me2 and H3K9me3. During the definition of the euchromatin-heterochromatin boundaries in blastoderm cells and for the establishment of repressive H3K9 methylation marks in primordial germ cells, the SU(VAR)3-3 demethylase plays an early and inductive regulatory role. SU(VAR)3-3 might also be involved in control of early transcriptional activities within *Drosophila* pericentromeric sequences preceding heterochromatin formation, as suggested by a model of heterochromatin formation that depends on the RNAi pathway (Pal-Bhadra et al., 2004).

A Mechanistic Hierarchy for the Establishment of Heterochromatin in Somatic Cells

Genetic analysis revealed that SU(VAR)3-3 functions upstream of the H3K9 HMTase SU(VAR)3-9 and the heterochromatin-associated proteins HP1 and SU(VAR)3-7 in control of gene silencing in PEV. Combined with earlier studies of epigenetic interactions (Schotta et al., 2002), heterochromatic gene silencing is established by a sequential action of SU(VAR)3-3, SU(VAR)3-9, the amount of Y heterochromatin, HP1, and SU(VAR)3-7. RPD3 also acts upstream of SU(VAR)3-9, because *Rpd3* mutations dominate the dose-dependent PEV enhancer effect of SU(VAR)3-9 (Czermin et al., 2001). Additional genomic copies of *Su(var)3-3* are epistatic to a *Rpd3* mutation (Figure S5) placing the H3K4 demethylase SU(VAR)3-3 together with RPD3 at the top of a mechanistic hierarchy controlling heterochromatic gene silencing in *Drosophila*. Such a role is in agreement with the enriched association of SU(VAR)3-3 to prospective heterochromatin in early blastoderm nuclei. In *Su(var)3-3* null embryos, there is an extension of H3K4me2 and concomitant reduction of H3K9me3 at prospective heterochromatin, suggesting that SU(VAR)3-3 has a protective function at heterochromatic regions to restrict expansion of H3K4 methylation. Similarly, H3K9 acetylation becomes expanded toward heterochromatin. H3K4 methylation precedes H3K9 methylation in blastoderm nuclei, and both SU(VAR)3-3 and SU(VAR)3-9 are abundant proteins within cleavage chromatin. A developmentally regulated silencing complex between SU(VAR)3-3, RPD3, and SU(VAR)3-9 (Figure 5B) is therefore likely to dictate the distinction between euchromatic and heterochromatic domains during early embryogenesis. A comparable functional crosstalk between human LSD1 and HDAC1/2, which depends on nucleosomal substrates and the CoREST protein, was currently shown (Lee et al., 2006). The interaction between SU(VAR)3-3 and RPD3 could also explain butyrate sensitivity of *Su(var)3-3* mutations (Reuter et al., 1982). The effect of SU(VAR)3-3 on heterochromatin formation during blastoderm could involve both maternal and zygotic protein. Association of SU(VAR)3-3 with cleavage chromatin is dependent on maternal sources. In contrast, all other effects on gene silencing are zygotically determined, and no maternal effects on PEV were found in any of the *Su(var)3-3* mutations. This is also supported by clonal analysis showing early onset and stable maintenance of gene silencing in PEV (Figure S5).

SU(VAR)3-3 Controls Transcriptional Silence in Primordial Germ Cells

Transcriptional silence in primordial germline cells of *Drosophila* is regulated by *nanos* (*nos*), *pumillio* (*pum*), and *germ cell-less* (*gcl*), since mutations in these genes result in premature activation of transcription in germ cells (Asaoka-Taguchi et al., 1999; Deshpande et al., 1999; Leatherman et al., 2002). However, in *nos* null embryos, only about 50% display H3K4 methylation signals (Schaner et al., 2003), suggesting that several other factors contribute to establishment and maintenance of transcriptional silence. SU(VAR)3-3 is likely to be a main component in this control because every pole cell nucleus in *Su(var)3-3* mutants displays H3K4 methylation and significant reduction of H3K9me2 and H3K9me3. Homozygous *Su(var)3-3* mutant females are completely sterile and do not develop oocytes in their egg chambers, which show a spectrum of developmental abnormalities and become arrested already at stage three to four (Szabad et al., 1988). This suggests that SU(VAR)3-3 is also required during oogenesis in *Drosophila*. In mammals, two of the three *nos* homologs are required for female fertility. For *nanos-3*, a function in primordial germ cell specification was demonstrated in knockout mice (Tsuda et al., 2003). Based on our data, a likely function for LSD1 during germline specification in mammals is predicted.

Su(var)3-3 mutant males are sterile and produce immobile sperm, and in primary spermatocytes of *Su(var)3-3* males, crystals are frequently found (G.R., unpublished data), a phenotype typical for X0 males or XY males carrying a deletion for the *crystal* (*cry*) locus (Tritto et al., 2003). Whether or not SU(VAR)3-3 also interferes with control of activation of Y chromosomal genes in the male germline remains to be studied. Furthermore, alleles of *Su(var)3-3* increase crossover in pericentric regions (Westpahl and Reuter, 2002), suggesting that SU(VAR)3-3 is also involved in control of heterochromatin packaging in female meiosis. All these phenotypic effects indicate that the H3K4 demethylase SU(VAR)3-3 is an important regulator in protecting chromatin functions in germline cells.

EXPERIMENTAL PROCEDURES

Drosophila Cultures, Stocks, and Genetic Analysis

Flies were reared on standard medium at 25°C. Chromosomes and mutations not noted here are described in Crosby et al. (2007).

Su(var)3-3 mutations were isolated by their dominant suppressor effect on *white* variegation of *In(1)w^{m4}* in the background of *E(var)* mutations after EMS (2.5 mM) mutagenesis (Reuter et al., 1986). The general effect of *Su(var)3-3* on gene silencing was studied in *Dp(1;f)γ238* (Le et al., 1995) variegating for *yellow*, *T(2;3)BL1* showing P[(*w*$^{+}$), *HS-lacZ*] transgene silencing (Lu et al., 1998) and *T(2;3)Sb^{V}* (Sinclair et al., 1992). *Su(var)3-3^{12}*-dependent compensation of PEV enhancement was studied in genotypes with one or two additional genomic copies of *Su(var)3-9* and one additional copy of *HP1* or *Su(var)3-7*.

Immunostaining of Polytene Chromosomes and Embryos

Preparation and staining of polytene chromosomes were performed as described (Schotta et al., 2002). Chromosomes were incubated with rabbit polyclonal α-mono-, α-di-, and α-trimethyl H3K4 or H3K9, respectively (Upstate; 1:50/5% dry milk) or α-SU(VAR)3-3 (1:300/5% dry milk) overnight at 4°C, followed by incubation with Alexa Fluor 488 (Molecular Probes) conjugated secondary antibody for 2 hr at 37°C (1:100/5% dry milk). Preparations were examined with confocal laser scanning microscopy (LSM 510; Zeiss). *Drosophila* embryos were fixed with the boiling fix method as described (Rothwell and Sullivan, 2000). SU(VAR)3-3-specific rabbit antibodies were raised against purified bacterial 6× HIS–SU(VAR)3-3 (amino acids 1–150) antigen.

Molecular Cloning of Su(var)3-3 and Constructs

Su(var)3-3 was mapped to the genomic region 76D-77B by P-transposase-induced male recombination using a series of *P* element insertions. After a cross of +/Y; *CyO*, *H{w*$^{+mC}$ = P2-3}HoP2.1/+; *Ly Su(var)3-3^{03} Sb/P* element males, *w^{m4h}*/+/+, *Ly*+, and +*Sb* recombinant chromosomes were selected and tested for the presence of *Su(var)3-3^{03}* by a cross to *Su(var)3-3^{12}/TM3*, *Sb Ser* flies (sterility test). For molecular analysis of *Su(var)3-3* mutations, genomic DNA from *Su(var)3-3^{x}/Df(3L)TR2* heterozygotes was used.

Full-length ORFs of *Su(var)3-3* (GenBank accession number AY094837) and mLSD1 (GenBank accession number BC059885) were amplified by PCR and cloned into expression vectors. Point mutations were introduced by site-directed mutagenesis. All constructs were verified by sequencing.

Immunoaffinity Purification of SU(VAR)3-3 Complexes

Nuclear extract from *Drosophila* embryos (0–3 hr) was incubated with crosslinked SU(VAR)3-3 antibody for 4 hr at 4°C. The antibodies were coupled to protein A Sepharose (Amersham) using standard techniques. The beads were washed with buffer A (50 mM Tris [pH 7.5]/150 mM NaCl/0.1% NP-40) followed by a wash with buffer A plus 250 mM KCl. The proteins were eluted with 0.1 M glycine (pH 2.5) and neutralized with 1:10 volume of 1 M Tris (pH 8.0). α-RPD3 (ab1767; 1:1000), α-HP1 (ab24726; 1:3000), α-SU(VAR)3-9 (Schotta et al., 2002; 1:1000), and α-SU(VAR)3-3 (1:3000) were used for western blotting. The detection was done according to manufacturer's instructions (Amersham; ECL-Kit).

In Vitro Histone Demethylase Assays

N-terminally FLAG-HA-tagged SU(VAR)3-3 and mLSD1 were baculovirally expressed in Sf9 cells and purified with M2 agarose (Sigma). Typically, 1 μg of recombinant protein was incubated with 11.25 μM histone H3 peptides (1–20) or 2.5 μg of calf thymus bulk histone (Roche) for 4 hr at 25°C in HDM buffer (10 mM HEPES (pH 7.9), 60 mM KCl, 1% glycerol) in a final volume of 20 μl. Histones were analyzed by western blot using methyl-specific antibodies (Peters et al., 2003). α-HA (sc-805; 1:1000), α-LSD1 (ab17721; 1:1000), and α-SU(VAR)3-3 (1:2000) were used to evaluate enzyme concentration. For mass spectrometry, peptide samples were acidified by adding 1 μl of 10% (v/v) TFA followed by purification using ZipTip C-18 microcolumns (Millipore). Peptides were eluted with 50% (v/v) ACN, 0.1% (v/v) TFA, and 5 μl was mixed with an equal volume of 10 mg/ml 4-hydroxy-α-cyanocinnamic acid (HCCA, Sigma, in 70% [v/v] ACN, 0.1% [v/v] TFA), of which 1 μl was placed on a MALDI sample plate. Spectra were acquired on a MALDI-TOF/TOF mass spectrometer 4800 Analyzer (Applied Biosystems, Darmstadt) with 8000 shots per spectrum. In HMTase assays, 3 μg of recombinant SU(VAR)3-3 and 1 μg histone H3 peptides were incubated overnight at 25°C in 16.7 mM HEPES-KOH (pH 7.9), 100 mM KCl, and 1.7% glycerol in a final volume of 20 μl. Then, 2.5 μl 10× HIM buffer (Czermin et al., 2001), 1 μl of 1.25 M glycine, 0.5 μl of SAM (1 μCi/μl; Amersham), and 1 μg of recombinant SU(VAR)3-9 were added and incubated for 20 min at 25°C. Reactions were stopped and incorporated radioactivity was measured according to Czermin et al. (2001).

Chromatin Immunoprecipitation

Fly heads and salivary glands were fixed with 1.8% formaldehyde for 30 min at room temperature, resuspended in RIPA buffer (140 mM NaCl, 10 mM Tris-HCl [pH 8.0], 1 mM EDTA, 1% Triton X-100, 0.1% SDS, 0.1% DOC), and lysed by sonication. The lysate was cleared by centrifugation, preabsorbed by incubation with protein G Sepharose beads (Amersham), and incubated with 7 μg α-dimethyl H3K9 (Upstate) overnight at 4°C. Antibody complexes were bound to protein G Sepharose beads. DNA was recovered and dissolved in 150 μl water. Control mock immunoprecipitations were done in parallel without antibodies. Real-time PCR analysis was performed according to Dellino et al. (2004) and DNA from 5 μl of each sample amplified in 20 μl reactions with 2× SYBR Green Super Mix (Bio-Rad). Primer sequences are available upon request.

Supplemental Data

Supplemental Data include five figures and can be found with this article online with this article at http://www.molecule.org/cgi/content/full/26/1/103/DC1/.

ACKNOWLEDGMENTS

We thank Susanne Opravil, Gustav Ammerer, Kristin Wächter, Enrica Müller, Sandy Mietsch, Maria Kube, and Kathrin Kittlaus for useful reagents and experimental support. Research in the laboratory of G.R. is supported by grants from the Deutsche Forschungsgemeinschaft (DFG), Land Sachsen-Anhalt, and the European Union (EU-network LSHG-CT 2000-503433). Studies conducted in the lab of T.J. were supported by the Institute of Molecular Pathology (IMP) through Boehringer Ingelheim and by grants from the Vienna Economy Promotion Fund (WWFF), the European Union (EU-network HPRN-CT 2000-00078), and the Austrian GEN-AU Initiative, which is financed by the Austrian Ministry of Education, Science, and Culture. M.Y. is a Japan Society for the Promotion of Science (JSPS) research fellow.

Received: October 6, 2006
Revised: December 21, 2006
Accepted: February 28, 2007
Published: April 12, 2007

REFERENCES

Asaoka-Taguchi, M., Yamada, M., Nakamura, A., Hanyu, K., and Kobayashi, S. (1999). Maternal Pumillio acts together with Nanos in germ line development in *Drosophila* embryos. Nat. Cell Biol. *1*, 431–437.

Bannister, A.J., and Kouzarides, T. (2005). Reversing histone methylation. Nature *436*, 1103–1106.

Becker, H.J. (1966). Genetic and variegation mosaics in the eye of Drosophila. In Current Topics in Developmental Biology, A.A. Moscona and A. Monroy, eds. (New York: Academic Press), pp. 155–171.

Bernstein, B.E., Humphrey, E.L., Erlich, R.L., Schneider, R., Bouman, P., Liu, J.S., Kouzarides, T., and Schreiber, S.L. (2002). Methylation

of histone H3 Lys4 in coding regions of active genes. Proc. Natl. Acad. Sci. USA *99*, 8695–8700.

Crosby, M.A., Goodman, J.L., Strelets, V.B., Zhang, P., Gelbart, W.M., and the FlyBase Consortium (2007). FlyBase: genomes by the dozen. Nucleic Acids Res. *35*, 486–491.

Czermin, B., Schotta, G., Hülsmann, B.B., Brehm, A., Becker, P.B., Reuter, G., and Imhof, A. (2001). Physical and functional interaction of SU(VAR)3-9 and HDAC1 in *Drosophila*. EMBO Rep. *2*, 915–919.

Dellino, G.I., Schwartz, Y.B., Farkas, G., McCabe, D., Elgin, S.C., and Pirrotta, V. (2004). Polycomb silencing blocks transcription initiation. Mol. Cell *13*, 887–893.

Deshpande, G., Calhoun, G., Yanowitz, J.L., and Schedl, P.D. (1999). Novel functions of nanos in downregulating mitosis and transcription during the development of the *Drosophila* germline. Cell *99*, 271–281.

Ebert, A., Schotta, G., Lein, S., Kubicek, S., Krauss, V., Jenuwein, T., and Reuter, G. (2004). *Su(var)* genes regulate the balance between euchromatin and heterochromatin in *Drosophila*. Genes Dev. *18*, 2973–2983.

Ebert, A., Lein, S., Schotta, G., and Reuter, G. (2006). Histone modification and the control of heterochromatic gene silencing in *Drosophila*. Chromosome Res. *14*, 377–392.

Foe, V.E., Odell, G.M., and Edgar, B.A. (1993). Mitosis and morphogenesis in the Drosophila embryo: point and counterpoint. In The Development of Drosophila melanogaster, M. Bate and A. Martinez-Arias, eds. (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press), pp. 149–300.

Garcia-Bellido, A., Ripoll, P., and Morata, G. (1973). Developmental compartmentalisation of the wing disc of *Drosophila*. Nat. New Biol. *245*, 251–253.

Jaquet, Y., Delattre, M., Spierer, A., and Spierer, P. (2002). Functional dissection of the *Drosophila* modifier of variegation *Su(var)3-7*. Development *129*, 3975–3982.

Kouzarides, T. (2002). Histone methylation in transcriptional control. Curr. Opin. Genet. Dev. *12*, 198–209.

Le, M.H., Duricka, D., and Karpen, G.H. (1995). Islands of complex DNA are widespread in *Drosophila* centric heterochromatin. Genetics *141*, 282–303.

Leatherman, J.L., Levin, L., Boero, J., and Jongens, T.A. (2002). Germ cell-less acts to repress transcription during establishment of the *Drosophila* germ cell lineage. Curr. Biol. *12*, 1681–1685.

Lee, G.M., Wynder, C., Bochar, D.A., Hakimi, M.-A., Cooch, N., and Shiekhattar, R. (2006). Functional interplay between histone demethylase and deacetylase enzymes. Mol. Cell. Biol. *26*, 6395–6402.

Lu, B.Y., Ma, J., and Eissenberg, J.C. (1998). Developmental regulation of heterochromatin-mediated gene silencing in *Drosophila*. Development *125*, 2223–2234.

Metzger, E., Wissmann, M., Yin, N., Müller, J.M., Schneider, R., Peters, A.H.F.M., Günther, T., Buettner, R., and Schüle, R. (2005). LSD1 demethylates repressive histone marks to promote androgene-receptor-dependent transcription. Nature *437*, 436–439.

Nakayama, J., Rice, J.C., Strahl, B.D., Allis, C.D., and Grewal, S.I. (2001). Role of histone H3 lysine 9 methylation in epigenetic control of heterochromatin assembly. Science *292*, 110–113.

Newsome, T.P., Asling, B., and Dickson, B.J. (2000). Analysis of *Drosophila* photoreceptor axon guidance in eye-specific mosaics. Development *127*, 851–860.

Nicolas, E., Lee, M.G., Hakimi, M.-A., Cam, H., Grewal, S.S., and Shiekhattar, R. (2006). Fission yeast homologs of human histone H3 lysine 4 demethylase regulate a common set of genes with diverse functions. J. Biol. Chem. *281*, 35983–35988.

Nishioka, K., Chuikov, S., Sarma, K., Erdjument-Bromage, H., Allis, C.D., Tempst, P., and Reinberg, D. (2002). Set9, a novel histone H3 methyltransferase that facilitates transcription by precluding histone tail modifications required for heterochromatin formation. Genes Dev. *16*, 479–489.

Pal-Bhadra, M., Leibovitch, B.A., Gandhi, S.G., Rao, M., Bhadra, U., Birchler, J.A., and Elgin, S.C. (2004). Heterochromatic silencing and HP1 localization in Drosophila are dependent on the RNAi machinery. Science *303*, 669–672.

Parks, A.L., Cook, K.R., Belvin, M., Dompe, N.A., Fawcett, R., Huppert, K., Tan, L.R., Winter, C.G., Bogart, K.P., Deal, J.E., et al. (2004). Systematic generation of high-resolution deletion coverage of the *Drosophila melanogaster* genome. Nat. Genet. *36*, 288–292.

Peters, A.H., Mermoud, J.E., O'Carroll, D., Pagani, M., Schweizer, D., Brockdorff, N., and Jenuwein, T. (2002). Histone H3 lysine 9 methylation is an epigenetic imprint of facultative heterochromatin. Nat. Genet. *30*, 77–80.

Peters, A.H.F.M., Kubicek, S., Mechtler, K., O'Sullivan, J., Derijck, A.A.H.A., Perez-Burgos, L., Kohlmaier, A., Opravil, S., Tachibana, M., Shinkai, Y., et al. (2003). Partitioning and plasticity of repressive histone methylation states in mammalian chromatin. Mol. Cell *12*, 1577–1589.

Rea, S., Eisenhaber, F., O'Carroll, D., Strahl, B.D., Sun, Z.-W., Schmid, M., Opravil, S., Mechtler, K., Ponting, C.P., Allis, C.D., and Jenuwein, T. (2000). Regulation of chromatin structure by site-specific histone H3 methyltransferases. Nature *406*, 593–599.

Reuter, G., Dorn, R., and Hoffmann, H.-J. (1982). Butyrate sensitive suppressor of position-effect variegation mutations in *Drosophila melanogaster*. Mol. Gen. Genet. *188*, 480–485.

Reuter, G., Dorn, R., Wustmann, G., Friede, B., and Rauh, G. (1986). Third chromosome suppressor of position-effect variegation loci in Drosophila melanogaster. Mol. Gen. Genet. *202*, 481–487.

Rothwell, W.F., and Sullivan, W. (2000). Fluorescent analysis of Drosophila embryos. In Drosophila Protocols, W. Sullivan, M. Ashburner, and R.S. Hawley, eds. (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press), pp. 141–158.

Ryder, E., Blows, F., Ashburner, M., Bautista-Llacer, R., Coulson, D., Drummond, J., Webster, J., Gubb, D., Gunton, N., Johnson, G., et al. (2004). The DrosDel collection: a set of P-element insertions for generating custom chromosomal aberrations in *Drosophila melanogaster*. Genetics *167*, 797–813.

Schaner, C.E., Deshpande, G., Schedl, P.D., and Kelly, W.G. (2003). A conserved chromatin architecture marks and maintains the restricted germ cell lineage in worms and flies. Dev. Cell *5*, 747–757.

Schotta, G., Ebert, A., Krauss, V., Fischer, A., Hoffmann, J., Rea, S., Jenuwein, T., Dorn, R., and Reuter, G. (2002). Central role of *Drosophila* SU(VAR)3-9 in histone H3-K9 methylation and heterochromatic gene silencing. EMBO J. *21*, 1121–1131.

Schotta, G., Ebert, A., Dorn, R., and Reuter, G. (2003). Position-effect variegation and the genetic dissection of chromatin regulation in *Drosophila*. Semin. Cell Dev. Biol. *14*, 67–75.

Shi, Y., Lan, F., Matson, C., Mulligan, P., Whetstine, J.R., Cole, P.A., Casero, R.A., and Shi, Y. (2004). Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. Cell *119*, 941–953.

Shi, Y., Matson, C., Lan, F., Iwase, S., Baba, T., and Shi, Y. (2005). Regulation of LSD1 demethylase activity by its associated factors. Mol. Cell *19*, 857–864.

Sinclair, D.A., Ruddell, A.A., Brock, N.J., Lloyd, V.K., and Grigliatti, T.A. (1992). A cytogenetic and genetic characterization of a group of closely linked second chromosome mutations that suppress position-effect variegation in *Drosophila melanogaster*. Genetics *130*, 333–344.

Szabad, J., Reuter, G., and Schröder, M.B. (1988). The effect of two mutations connected with chromatin functions on female germ-line cells of *Drosophila*. Mol. Gen. Genet. *211*, 56–62.

Tritto, P., Specchia, V., Fanti, L., Spinelli, L., Berloco, M., Dálessandro, R., Pimpinelli, S., Palumbo, G., and Bozzetti, M.P. (2003). Structure, regulation and evolution of the crystal-Stellate system of *Drosophila*. Genetica *117*, 247–257.

Tsuda, M., Sasaoka, Y., Kiso, M., Abe, K., Haraguchi, S., Kobayashi, S., and Saga, Y. (2003). Conserved role of nanos proteins in germ cell development. Science *301*, 1239–1241.

Tsukada, Y., Fang, J., Erdjument-Bromage, H., Warren, M.E., Borchers, C.H., Tempst, P., and Zhang, Y. (2006). Histone demethylation by a family of JmjC domain-containing proteins. Nature *439*, 811–816.

Westpahl, T., and Reuter, G. (2002). Recombinogenetic effects of suppressor of position-effect variegation in *Drosophila*. Genetics *160*, 609–621.

Wustmann, G., Szidonya, J., Taubert, H., and Reuter, G. (1989). The genetics of position-effect modifying loci in *Drosophila melanogaster*. Mol. Gen. Genet. *217*, 520–527.

Xu, T., and Rubin, G. (1993). Analysis of genetic mosaics in developing and adult *Drosophila* tissues. Development *117*, 1223–1237.

Zhang, Y., and Reinberg, D. (2001). Transcription regulation by histone methylation: interplay between different covalent modifications of the core histone tails. Genes Dev. *15*, 2343–2368.