Mass spectrometry analysis of protein post-translational modifications: phosphorylation and O-GlcNAc glycosylation

Ph.D. Thesis

Éva Klement

Dr. Botond Penke
Supervisor

Dr. Katalin F. Medzihradszky
Consultant

Department of Medical Chemistry
University of Szeged

Proteomics Research Group
Biological Research Center, Hungarian Academy of Sciences

Szeged

2009

Accepted publications related to this thesis

I. Hlavanda E*, **Klement E***, Kókai E, Kovács J, Vincze O, Tokési N, Orosz F, Medzihradszky KF, Dombrádi V, Ovádi J. Phosphorylation Blocks the Activity of Tubulin Polymerization-promoting Protein (TPPP): IDENTIFICATION OF SITES TARGETED BY DIFFERENT KINASES. *Journal of Biological Chemistry* 282 (40): 29531-9 (2007)
(*These authors equally contributed to the work.)

II. Dorjgotov D, Jurca ME, Fodor-Dunai C, Szucs A, Otvös K, **Klement E**, Bíró J, Fehér A. Plant Rho-type (Rop) GTPase-dependent activation of receptor-like cytoplasmic kinases in vitro. *FEBS Letters* 583 (7):1175-82 (2009)

III. Kovács L, Alexa A, **Klement E**, Kókai E, Tantos Á, Gógl G, Sperka T, Medzihradszky KF, Tözsér J, Dombrádi V, Friedrich P. Regulation of calpain B from *Drosophila melanogaster* by phosphorylation. *FEBS Journal* 276 (17): 4959-72. (2009)

Submitted publication related to this thesis

IV. **Klement E**, Lipinszki Z, Kupihár Z, Udvardy A, Medzihradszky KF. Enrichment of O-GlcNAc modified proteins by the periodate oxidation – hydrazide resin capture approach

Abbreviations

BEMAD	β-elimination/Michael addition
CID	collision-induced dissociation
CDK5	cyclin-dependent kinase 5
DHB	2,5-dihydroxy-benzoic acid
EDTA	ethylenediaminetetraacetic acid
EGF	epidermal growth factor
ERK	extracellular signal-regulated kinase
ESI	electrospray ionization
H-CPG	controlled-pore-glass-bound hydrazide
HPLC	high performance liquid chromatography
IDA	iminodiacetic acid
IMAC	immobilized metal affinity chromatography
LC-MS/MS	liquid chromatography - tandem mass spectrometry
MALDI-TOF	matrix assisted laser desorption ionization - time-of-flight
MS	mass spectrometry
MS/MS	tandem mass spectrometry
MSA	multistage activation
NTA	nitrilotriacetic acid
NCBI	National Center for Biotechnology Information
O-GlcNAc	O-linked β-N-acetylglucosamine
PKA	cAMP-dependent protein kinase
PSD	post source decay
PTM	post-translational modification
Q-TOF	quadrupole time-of-flight
R	resolution
ROP	Rho-type GTPase of plants
RRK1	ROP interacting receptor-like kinase 1
SDS	sodium dodecyl sulfate
TOF	time-of-flight

TPPP/p25	tubulin polymerization promoting protein
UPLC	ultra performance liquid chromatography

Table of contents

1. Introduction

A comprehensive analysis at the protein level is required for a deeper understanding of the various biological processes. This includes the identification of proteins with particular functions, mapping of interacting networks, follow-up of the changes upon external stimuli or in different developmental stages, but also the characterization of the post-translational modifications (PTM), as they affect structure, function and activity of the proteins. Over 200 modifications have been described in the literature[1] and the list is growing. The modifications display a high diversity in chemical structure, size, stoichiometry, complexity or cellular localization. Mass spectrometry (MS) has become an indispensable analytical tool in the characterization of the PTMs. One of its strengths is its high detection sensitivity that allows protein analysis at the low femtomole level. Moreover, with the continuous improvements in instrumentation the analysis of a wide variety of modifications can be accomplished by this technique. The MS analysis of the different PTMs often necessitates prior isolation of the modified molecules, especially those with lower stoichiometry. Therefore, efficient enrichment methodologies have to be developed for the analysis of these PTMs. In my thesis work I focused on the enrichment and mass spectrometry analysis of two regulatory modifications: phosphorylation and O-linked β-N-acetylglucosamine glycosylation.

1.1. Protein mass spectrometry

At early stages mass spectrometry was used for the analysis of small volatile compounds. The analysis of nonvolatile biopolymers could be achieved only by the introduction of new ionization techniques: plasma desorption (PDMS), fast atom bombardment (FAB) and liquid secondary ionization (LSIMS). However, matrix assisted laser desorption ionization (MALDI) and electrospray ionization (ESI) represented a real breakthrough. Both are soft ionization techniques yielding mostly protonated molecules without significant fragmentation.

MALDI was invented by Hillenkamp and Karas[2]. The analyte is mixed with the solution of a small organic molecule, called matrix. This mixture is then applied to a metal plate and let to dry leading to the co-crystallization of the analyte and matrix. The matrix serves a dual-purpose. First, it absorbs the laser irradiation that induces sublimation of the

matrix/analyte co-crystals to the gas phase; second, it protects the analyte from decomposition by the laser energy. Gas-phase proton-transfer reactions facilitate ionization. This source operates either under high vacuum or at atmospheric pressure (AP-MALDI). Mainly singly charged ions are formed this way.

Nanoelectrospray – the ionization method usually utilized in protein analysis instead of the regular ESI – is the 'miniaturized' version of the ionization technique developed by Fenn and co-workers[3;4]. High potential is applied to a liquid emerging from a capillary at low flow rates (~50 nL/min-1 μL/min). Due to the accumulation of charges the drop at the tip of the capillary adopts a conical shape called 'Taylor cone'[4] from which small charged droplets are ejected producing a fine spray. The droplets shrink as the solvent evaporates, and Coulomb repulsion forces split the droplets into smaller and smaller parts until ions are desorbed from the surface. The sample ionized at atmospheric pressure is introduced into the high vacuum region of the mass spectrometer through several pumping stages. A heated capillary or a heated curtain gas at the entrance compensates for the cooling caused by the adiabatic expansion of the sample and solvent. Characteristic of this ionization technique is the formation of multiply charged molecular ions. As ions are detected according to their mass-to-charge (m/z) ratio, this feature largely extends the mass range of the molecules that can be analyzed far above the usual mass range of the mass analyzers. Chromatographic systems with the appropriate flow rate and mass spectrometry-compatible mobile phase may be directly linked to mass spectrometers equipped with an ESI source, thus combining fractionation and detection of mixture components. The chromatography also facilitates the concentration of diluted samples.

Mass analyzers used in proteomics are time-of-flight (TOF), quadrupole (Q), 3D and linear ion traps, and most recently, Fourier transform ion cyclotron resonance (FT-ICR) and orbitrap analyzers[5;6]. The most important features that characterize the performance of a mass analyzer are data acquisition speed, detection range, mass resolution (R), mass accuracy and detection sensitivity.

In TOF instruments the accelerated ions are ejected into a flight tube. Their flight time in this field free region to the detector is proportional to their m/z ratio. The ions are transmitted in packages, so this analyzer is well suited with a pulsed ion source like MALDI. If TOF is coupled with a continuous ion source (e.g. ESI), the ions are first trapped, then

orthogonally accelerated into the flight tube by a pusher electrode. The TOF analyzer has a high detection sensitivity as all the ions formed are guided to the detector. The mass range is theoretically unlimited, practically proteins up to a few hundred kDa can be detected with MALDI-TOF MS in linear mode[7]. High resolution (R~10,000) is achieved by the introduction of an ion mirror, called reflectron[8]. This device consists of a series of electrodes with increasing potential. It compensates for the kinetic energy dispersion of the ions of the same mass, as faster ions penetrate deeper and spend a longer time in the reflectron than the slower ones. In addition, the length of the flight tube and accordingly the flight time is doubled. However, increased resolution in reflectron mode goes with decreased sensitivity. In MALDI-TOF instruments an additional feature increases resolution, called delayed extraction[9]. Ions produced in the source are extracted into the flight tube with a short delay to compensate for the space and energy dispersion of the ions during the ionization process as well as for the slight time differences when the ionization actually occurred.

Quadrupole mass analyzers consist of four parallel rods. An oscillating electric field is produced by applying radio frequency (rf) with superimposed direct current (dc) to the rods. Ions with particular m/z having stable trajectories within the quadrupole reach the detector, all other ions with unstable trajectories collide to the rods. Keeping rf and dc voltages constant allows selection of an ion with particular m/z, while changing the potentials the different masses are consecutively detected. In rf-only mode all ions are transmitted. This operation is set in quadrupoles used as a collision cell in tandem mass spectrometers as well as in Q-TOF tandem instruments during the MS survey. Quadrupole analyzers have low resolution, $R \leq 2000$. The detection limit typically ranges to 2000 m/z.

A 3D ion trap is made up of a ring electrode and two hyperbolic end cap electrodes. A more recent development, the linear trap consists of four parallel rods. As in quadrupole instruments, an alternating rf potential with overlapping direct potential is applied to the electrodes. However, the mode of operation is different: all the ions with different m/z are stored together in the trap, then ions of increasing m/z are successively expelled by resonance ejection applying additional rf to the end caps. Low pressure helium gas in the trap prevents expansion of the ion cloud. The number of ions trapped in the device is controlled to overcome space-charge effects. Ion trap instruments feature high sensitivity, especially linear

traps. The resolution depends on the scan speed and the width of the mass window ($R \leq 4000$). The detection range is generally 2000 m/z, though it can be extended to 4000 m/z.

Since soft ionization techniques mostly provide information only on the mass of the intact molecule, structural information of the compounds can be obtained by tandem mass spectrometry (MS/MS) which involves two or more stages of mass analysis with a dissociation process in between. MS/MS is realized either in space or in time. In tandem-in-space instruments two mass analyzers are coupled, like Q-TOF, triple qudrupole (Q_3), Q-trap or TOF-TOF. Tandem-in-time can be performed in ion storage devices, i.e. in ion trap or FT-ICR mass spectrometers. In a special way, fragmentation spectra can also be recorded in MALDI-TOF instruments as follows.

In a MALDI-TOF instrument the monomolecular dissociation of an ion selected by the appropriate gating device is followed by continuously adjusting the reflectron voltage. The process is called post source decay (PSD) analysis. Since only a small portion of the ions has sufficient vibrational energy deposited upon MALDI ionization, this technique produces lower quality MS/MS spectra and the information content of the PSD spectra is very sequence-dependent. Sensitivity is increased at the expense of resolution and accuracy, average masses are detected with ±1 Da tolerance.

In a Q-TOF mass spectrometer the ion of interest is selected by the quadrupole analyzer, then it is subjected to collision-induced dissociation (CID) in the quadrupole collision cell and the fragments are measured orthogonally by a TOF mass analyzer. Here, an effective fragmentation – that can be influenced by adjusting the pressure of the collision gas as well as by changing the collision energy – is combined with the high accuracy (50 ppm) measurement of the fragments.

During MS/MS analysis in ion trap instruments all the ions that were stored in the trap are ejected except the ion of interest, then collisional dissociation is facilitated by excitation at a given frequency. The fragments are analyzed by resonance ejection or alternatively, a fragment can be selected for further dissociation to produce MS(3) or with additional cycles higher order MS(n) spectra. A limitation of ion trap instruments is however, that fragment ions with m/z lower than 1/3 of the precursor ion mass are lost because of stability reasons.

MS analysis of the proteins is mostly done at the peptide level. Enzymatic or chemical cleavages are performed to cut the proteins into smaller parts. The most frequently used

enzyme is trypsin because it is highly specific, cleaving at the C-termini of arginine and lysine residues. The masses of the peptides produced by specific enzymatic cleavages are sequence dependent, therefore, characteristic for a particular protein.

In the simplest approach, the masses detected in the spectrum of the unfractionated protein digest are compared to the theoretical peptide masses calculated for the proteins present in the database considering the modifications during sample handling, the enzyme specificity and the accuracy of the mass measurement. This approach is restricted to protein identification in low complexity samples.

For the identification of proteins in complex mixtures, as well as for the characterization of PTMs the peptides are first fractionated, then selected one by one for fragmentation analysis. Low energy fragmentation of peptides triggered by collisional activation (CID) or by MALDI (PSD) generates sequence specific product ions. Upon collision with an inert gas, the kinetic energy is partially converted to vibrational energy. In instruments equipped with a collision cell, collision of a low energy ion (10-100 eV) with nitrogen as the collision gas results in fragmentation reactions at the protonated amide bonds. N-terminus related cyclic b-ions and C-terminus derived y-ions are mainly formed this way[10-12]. Moreover, multiple collisions with the collision gas induce further consecutive fragmentation of the b-ions leading to abundant b_2 ions in the MS/MS spectrum. Usually, no b_1 ion is observed because the cyclic structure cannot be formed. In ion trap instruments, the internal energy gathered from collisions with the helium gas is generally insufficient for fragmentation. Here, a *single* fragmentation step is triggered by resonance excitation of the selected precursor producing b- and y-type ions, but this ’peeling’ of the b-ions is not observed.

Fragmentation at distinct amide bonds occurs with different probability. The most preferred fragmentations are observed N-terminally to proline and glycine residues, and C-terminally to aspartic acids[13]. Additional ions detected in the CID spectra of peptides originate from the loss of small neutral molecules. The split of a water molecule - mainly from serine, threonine, aspartic acid or glutamic acid residues - results in an 18 Da loss. Ammonia loss (-17 Da) usually originates from the side chains of arginine, lysine, asparagine or glutamine residues. Immonium ions with the structure $NH_2{=}CHR^+$ in the low mass region of the MS/MS spectrum are diagnostic for the amino acids present in the molecule[14].

The signals characteristic for a particular PTM are deposited on the peptide fragmentation, facilitating the identification and localization of the modification. The assignment of the modifications that are stable under collisional activation is based on the mass shift of the fragments retaining the modification. Other modifications are easily eliminated without any indication on the site of the modification. Here, localization of the modification requires chemical derivatization or complementary fragmentation techniques that leave the modification unaffected. In between is a group of modifications that undergo partial fragmentation. In addition, the mass fragmented upon collisional activation is not necessarily the same as the mass of the modification, e.g. oxidation of methionine causes a mass shift of 16 Da whereas upon fragmentation a partial loss of 64 Da is observed.

1.1.1. Phosphorylation

One of the most studied PTMs is phosphorylation because of its extremely important biological function. Protein phosphorylation plays a key role in signal transduction pathways, but also in the regulation of other biological processes, often in co-operation with other types of PTMs[15-18]. Phosphorylation is a dynamic process, controlled by two groups of enzymes: kinases introduce while phosphatases remove the phosphate modification.

The analysis of phosphorylation is hampered by the fact that the modification generally occurs at low stoichiometry. In addition, phosphopeptides show lower ionization efficiency compared to unmodified peptides. Their signal is suppressed because of the negative charge on the phosphate group, impairing their detection. Thus, enrichment of the phosphopeptides from a protein digest is essential. The main approaches in phosphopeptide enrichment are affinity based isolation and chemical derivatization.

Affinity based isolation methods are immobilized metal affinity chromatography (IMAC) and enrichment on metal oxides (titanium dioxide or zirconium dioxide). In IMAC, the phosphate group is trapped by a metal ion (usually Fe^{3+} or Ga^{3+}) coordinated by an immobilized chelating agent (iminodiacetic acid, IDA or nitrilotriacetic acid, NTA)[19;20]. Binding of acidic nonphosphorylated peptides is eliminated by methyl esterification of the carboxylic groups[21]. In 2005, a very impressive procedure for phosphopeptide enrichment was introduced[22]. Under highly acidic conditions phosphopeptides are bound to titanium dioxide,

then released upon ammonium hydroxide treatment. The simplicity and high selectivity of this method made it a very attractive alternative to IMAC.

Chemical derivatization methods include phosphoramidate chemistry (PAC)[23;24] or β-elimination/Michael addition reactions (BEMAD) followed by solid-phase capture[25] or affinity chromatography[26]. In the PAC approach, the carboxylic acid side chains are first converted to methyl esters, then a new thiol moiety is introduced into the phosphopeptides by coupling cysteamine to the phosphate group via a phosphoramidate linkage, followed by covalently binding to a thiol reactive resin. The procedure involves multiple reactions with internal clean-up steps making it a less favorable isolation technique. In the BEMAD strategy the phosphate group is removed by β-elimination under basic conditions[27] followed by Michael addition of a nucleophile on the resulting dehydroamino acid. Additional thiol group in the nucleophile allows selective isolation of the phosphopeptides by coupling to a solid-phase whereas a biotin tag enables affinity enrichment. The shortcoming of this approach is, that it is restricted to serine and threonine modified peptides. Moreover, O-sulfated, O-glycosylated and unmodified serines and threonines might also undergo β-elimination resulting in false positive identification.

A further possibility for the enrichment of phosphopeptides is ion exchange on strong cation exchange columns[28] or electrostatic repulsion hydrophilic interaction chromatography (ERLIC)[29;30].

Different paradigms should be followed when studying phosphorylation in a single protein or performing large-scale phosphoproteome analysis. In the former one, in depth analysis of the particular protein is performed to reveal all possible phosphorylation sites. For example, in a hypothesis-driven multiple-stage MS strategy described by Chang et al[31], all the potential phosphorylation sites were considered being modified and MS(2) and MS(3) spectra were acquired for all of the ions at a particular m/z matching to possible phosphopeptides to confirm the modification. If phosphopeptide-enrichment is utilized, IMAC and titanium dioxide are preferred, the other methods involve too many steps each leading to significant sample losses.

In a large-scale analysis phosphopeptides are identified somewhat incidentally. As only a partial overlap is observed between the sets of phosphopeptides isolated by different enrichment techniques, combination of these methods increases the overall outcome[23].

Phosphopeptide identification and site assignment by a probability-based database search is not as straightforward as peptide identification, therefore, a reference-facilitated analysis was proposed by Imanishi et al[32]. After phosphopeptide enrichment the sample is split and one half is dephosphorylated to serve as a reference for validation purposes.

In mass spectrometry, phosphorylation is characterized by an 80 Da mass shift, a shared feature with sulfation. The two modifications differ only by 0.009 Da. Fortunately, they can be distinguished by different fragmentation behavior upon collisional activation. In sulfopeptides the modification falls off completely as SO_3 (-80 Da) resulting in the same fragmentation as the unmodified peptide[33], whereas peptides phosphorylated on serine or threonine residues undergo a partial β-elimination of phosphoric acid (-98 Da) from the precursor ion as well as the fragments[34]. Tyrosine phosphorylation has been described as stable not displaying any significant losses, though a weak 80 Da loss might be observed. Additionally, phosphotyrosines can be recognized by an immonium ion at m/z 216. A further diagnostic signal in the CID spectrum of the phosphopeptides is PO_3^- detected at m/z 79 in negative ion mode[35;36].

1.1.2. O-GlcNAc glycosylation

Another important post-translational modification involved in signal transduction is O-linked β-N-acetylglucosamine (O-GlcNAc)[37-39]. This carbohydrate modification is attached to serine and threonine residues of cytosolic and nuclear proteins[40;41]. O-GlcNAc has been found to regulate other cellular processes, e.g. protein degradation[42;43] or transcription[41;44-46] as well. Its special importance was proven in the brain[47;48]. Proteins in the synaptic trafficking were shown to bear this modification[49]. O-GlcNAc has also been linked to neurodegenerative diseases, e.g. Alzheimer's disease[50;51]. Due to its occurrence and function, this modification significantly differs from complex N- or O-linked carbohydrate modifications, but shows similarity to phosphorylation as both are dynamic, substoichiometric, regulatory modifications. Moreover, an interplay between O-GlcNAc and phosphorylation has been suggested[16-18].

Being a substoichiometric modification, the analysis of the O-GlcNAc glycosylation requires prior enrichment. Different strategies are described in the literature: BEMAD[52], chemoenzymatic labeling[17;48;53] and lectin weak affinity chromatography (LWAC)[49;54]. The

BEMAD technique is analogous to that of phosphorylation. In chemoenzymatic labeling a ketone or azido moiety is introduced enzymatically into O-GlcNAc modified proteins. Then a biotin tag is attached through this ketone or azido functionality that enables affinity purification of the modified proteins. The LWAC approach is based on the weak affinity of O-GlcNAc to the lectin wheat germ agglutinin (WGA). An adequately long WGA column ensures separation of nonglycosylated peptides from O-GlcNAc modified ones.

In MS analysis, O-glycosylated peptides feature a characteristic fragmentation behavior as the glycosidic bond is much more susceptible to fragmentation than the amide bond[55-57]. Therefore, the O-GlcNAc modification is easily recognized by the sugar loss from the precursor ion and the oxonium ion observed at m/z 204[56;57]. Peptide fragmentation is underrepresented in the CID spectrum, especially in ion trap instruments, rendering peptide identification difficult. Moreover, the sugar loss proceeds via a gas phase rearrangement leaving no telltale sign of the site of the modification. Conversion of the facile O-GlcNAc modification to a stabile label by BEMAD[58] might assist site assignment.

2. Aims

The analysis of two post-translational modifications was addressed in this thesis: phosphorylation and O-GlcNAc glycosylation.

1. Phosphorylation
   - Studying the IMAC phosphopeptide enrichment with different multivalent cations and various binding and wash conditions (pH, solvent composition) we attempted to improve IMAC performance.
   - Applying IMAC and titanium dioxide enrichment strategies we intended to reveal the phosphorylation sites in three different biological events:
     - To determine the *in vitro* (human) and *in vivo* (bovine) phosphorylation sites of the tubulin polymerization promoting protein (TPPP/p25).
     - To map the autophosphorylation sites in the *Medicago* RRK1 kinase, and to identify the phosphorylation sites in its activator, the ROP6 GTPase.
     - To reveal the *in vitro* and *in vivo* phosphorylation sites of *Drosophila* calpain B.

2. O-GlcNAc glycosylation
   - Modifying the oxidation and cleavage steps in the periodate oxidation/hydrazide capture based enrichment technique originally developed for N-glycosylated proteins we aimed to extend the method for the analysis of O-GlcNAc modified proteins.

3. Materials and methods

Chemicals: Sequencing grade side-chain protected porcine trypsin (modified by reductive methylation) was ordered from Promega. C18 ZipTip was from Millipore. Ni-NTA Magnetic Agarose Beads were bought from Qiagen. Titanium dioxide was from SunCrom GmbH. Affi-gel was from BioRad. High purity solvents (HPLC grade) were purchased from Sigma and Merck. All other chemicals were obtained from Sigma.

TPPP/p25 isolated from bovine brain, as well as i*n vitro* phosphorylated human recombinant TPPP/p25 was provided by Dr. Judit Ovádi's group (Emma Hlavanda, Dr. Ferenc Orosz) at the Institute of Enzymology of the BRC HAS. *In vitro* phosphorylated recombinant *Medicago* RRK1 kinase and RO6 GTPase was from Dr. Attila Fehér's group (Dr. Dulguun Dorjgotov) at the Institute of Plant Biology of the BRC HAS. *In vitro* and *in vivo* phosphorylated *Drosophila* calpain B was provided by Dr. Viktor Dombrádi's group (László Kovács) at the Department of Medical Chemistry, University of Debrecen. Controlled pore glass hydrazide (H-CPG) was synthesized from LCA-CPG (CPG Inc., USA) by Dr. Zoltán Kupihár in the Nucleic Acids Laboratory of the Department of Medical Chemistry, University of Szeged.

In-gel digestion: Salts, SDS and Coomassie brilliant blue were removed by repeated wash with 25 mM ammonium bicarbonate in 50% acetonitrile. Disulfide bridges were reduced with 10 mM dithiothreitol dissolved in 25 mM ammonium bicarbonate, pH 8 at 56 °C for 30 min, then free sulfhydryls were derivatized with 55 mM iodoacetamide in 25 mM ammonium bicarbonate, pH 8 in the dark for 30 min. Excess reagents were removed by repeated washes with the above solution. The gel slices were dried, then rehydrated with trypsin solution (5 ng/μL in 25 mM ammonium bicarbonate, pH 8). Digestion proceeded at 37 °C for 4-6 h. The resulting peptides were extracted with 1% formic acid in 50% acetonitrile.

In-solution digestion: Protein was dissolved in 6 M guanidine hydrochloride in 25 mM ammonium bicarbonate, pH 8. Disulfide bridges were reduced with dithiothreitol (at a 500x

molar excess) at 56 °C for 1 h, and the free sulfhydryls were derivatized with iodoacetamide (at a 1100x molar excess) in the dark for 1 h. Then the sample was tenfold diluted with 25 mM ammonium bicarbonate, pH 8. Digestion with trypsin at an enzyme ratio of 1:10 (w/w) was performed at 37 °C for 4 h. The digest was acidified with trifluoroacetic acid and desalted on C18 ZipTip.

Methyl esterification: The protocol of Ficarro[21] was applied. The digest was dried down and redissolved in 50 µL of 2 M hydrochloric acid in methanol. Reaction proceeded at room temperature for 2 h. Then the sample was dried down again and redissolved in 10 µL of a mixture of equal volumes of water, methanol and acetonitrile.

Phosphopeptide enrichment by IMAC: In NTA Magnetic Agarose Beads Ni^{2+} was displaced by different metal ions as described by Thompson[20]. The beads were washed three times with water, 100 mM EDTA in 0.1% acetic acid and 0.1% acetic acid, then loaded with a 100 mM solution of the appropriate metal salt (twice, 5 min vortex), and washed three times with water. The sample dissolved in the appropriate buffer was loaded onto the beads and vortexed for 15 min. Then the supernatant was discarded, the beads washed three times and the phosphopeptides eluted by 15 min vortexing. Loading, wash and elution conditions are listed in Table 1. An aliquote of the eluate was either directly analyzed by MALDI-TOF MS or diluted with 0.1% formic acid in water and loaded on the trapping column for LC-MS/MS.

#	Load	Wash	Elute
1	20 mM ammonium acetate, pH 6.5	20 mM ammonium acetate, pH 6.5	0.85% phosphoric acid
2	20 mM ammonium acetate, pH 6.5	20 mM ammonium acetate, pH 6.5 in 30% acetonitrile	0.42% phosphoric acid in 50% acetonitrile
3	20 mM ammonium acetate, pH 4	20 mM ammonium acetate, pH 4 in 30% acetonitrile	0.42% phosphoric acid in 50% acetonitrile
4	20 mM ammonium acetate, pH 8	20 mM ammonium acetate, pH 8 in 30% acetonitrile	0.42% phosphoric acid in 50% acetonitrile
5	100 mM acetic acid, pH 2	70 mM acetic acid, pH 2 in 30% acetonitrile	0.42% phosphoric acid in 50% acetonitrile
6	100 mM ammonium acetate, pH 7	100 mM ammonium acetate, pH 7 in 30% acetonitrile	0.42% phosphoric acid in 50% acetonitrile
7	70 mM ammonium acetate, pH 7 in 30% acetonitrile	70 mM ammonium acetate, pH 7 in 30% acetonitrile	0.42% phosphoric acid in 50% acetonitrile
8	water/methanol/acetonitrile	water/methanol/acetonitrile	0.42% phosphoric acid in 50% acetonitrile

Table 1. Load, wash and elution conditions used in IMAC phosphopeptide enrichment.

Phosphopeptide enrichment by titanium dioxide: A modified protocol of Larsen[22] was applied. The digest was dried down, redissolved in 1% trifluoroacetic acid in 50% acetonitrile

and loaded on titanium dioxide suspended in the same solvent. After a few minutes vortexing supernatant was discarded and titanium dioxide washed three to five times with the above loading solvent, then twice with water. Phosphopeptides were eluted with 1% ammonium hydroxide. For MALDI-TOF analysis, an aliquote of the eluate was spotted on the target, let to dry, then redissolved in 0.1% formic acid in 50% acetonitrile, let to dry and redissolved again before adding the matrix. This sample preparation technique was used in order to eliminate ammonium salts and aid crystallization. For LC-MS/MS analysis, an aliquote of the eluate was acidified with 0.1% formic acid and loaded on the trapping column.

Sodium periodate oxidation: O-GlcNAc peptide standard (TAPTgSTIAPG) was dissolved in 100 mM sodium acetate. The oxidation was carried out by the addition of sodium periodate at a final concentration of 20 mM at pH 3, 6.5 or 9. The pH was adjusted with acetic acid to pH 3 and with sodium hydroxide to pH 9. The reaction proceeded for 0.5-24 h, then the reaction mixture was either desalted on C18 ZipTip or the oxidation was terminated by 1-5 equivalents of sodium sulfite prior to the reaction with the hydrazide.

Hydrazone formation: The pH of the reaction mixture was adjusted to pH 5-6 and benzoic acid hydrazide was added at a final concentration of 50 mM. The reaction proceeded at room temperature for 2-24 h, then the sample was desalted on C18 ZipTip.

Cleavage of the hydrazone product: Three different approaches were used for the hydrazone cleavage. 1) The hydrazone was cleaved by the addition of sodium periodate at a final concentration of 20 mM in 50 mM sodium acetate at pH 6. 2) Alternatively, 1-10% formic acid was used to cleave the hydrazone bond at room temperature. 3) Conversion of the hydrazone to oxime was performed by the addition of hydroxylamine hydrochloride at a final concentration of 200 mM in 50 mM sodium acetate at pH 5.

O-GlcNAc enrichment by hydrazide resin capture: The proteins were solubilized with 6 M guanidine hydrochloride or 0.5% SDS detergent in 50 mM sodium acetate or triethylamine phosphate buffer (pH 5-6). The oxidation was performed at 37 °C for 6-8 h in the dark by the addition of sodium periodate at a final concentration of 20 mM. The oxidation was terminated by 1-5 equivalents of freshly prepared sodium sulfite solution for 20 min. If needed, the pH of the solution was readjusted to pH 5-6. The hydrazide resin (stored in isopropanol) was washed three times with water and added to the reaction mixture. The coupling proceeded overnight (16-20 h) with gentle agitation by vertical rotation. The supernatant was discarded

and the resin washed five times with 0.5 M triethylamine phosphate, pH 8.5 in 30% acetonitrile or 0.5 M sodium-phthalate, pH 7 in 30% acetonitrile, then three times with 100 mM triethylamine phosphate, pH 8. The disulfide bridges were reduced with dithiothreitol (20 mM in 0.5% SDS, 100 mM triethylamine phosphate, pH 8) at 56 °C for 1 h and the free sulfhydryls derivatized with iodoacetamide at a final concentration of 50 mM at pH 8 for 1 h in the dark. The resin was washed three times with 50 mM triethylamine phosphate, pH 7.5. Tryptic digestion proceeded at 37 °C overnight. Nonspecifically bound tryptic peptides were removed by washing the resin five times with each 0.5 M triethylamine phosphate, pH 8.5 in 30% acetonitrile; 0.5 M sodium-phthalate, pH 7 in 30% acetonitrile and 0.1% trifluoroacetic acid in 50% acetonitrile; then three times with isopropanol and finally with 50 mM sodium acetate, pH 6.5 or water. O-GlcNAc modified peptides were eluted in four different ways. 1) Periodate elution was performed by 20 mM sodium periodate in 50 mM sodium acetate, pH 5-6 overnight. 2) Elution as an oxime was carried out by the addition of 200 mM hydroxylamine hydrochloride in 50 mM sodium acetate, pH 5 overnight with vertical rotation. 3) In the BEMAD elution, a solution of 250 mM sodium hydroxide and 50 mM cysteamine hydrochloride was added to the resin and incubated at 37 °C for 4 h. In all three cases, the eluate was collected, then the resin washed with 50 mM sodium acetate, pH 6.5 and added to the eluate. Additionally, the resin was washed twice with 100 mM acetic acid in 80% acetonitrile collected separately. The first eluate was concentrated in a speedvac, desalted on C18 ZipTip, dried, then redissolved in water. The second eluate was dried and redissolved in water. 4) The acidic cleavage was performed with 1-10% formic acid or 0.1-1% trifluoroacetic acid at room temperature, 37 °C or 60 °C for 0.5-24 h. The eluate was collected, then the resin washed with the same solution and added to the eluate. Finally the resin was washed twice with 100 mM acetic acid in 80% acetonitrile collected separately. The eluates were dried and redissolved in water.

MALDI-TOF MS analyses were performed on a Bruker Reflex III mass spectrometer.

MALDI-TOF MS in reflectron mode: An aliquot of the sample was spotted onto the target plate using the dried droplet protocol. 2,5-dihydroxy-benzoic acid served as the matrix. Delayed extraction was set to 200 or 400 ns. An acceleration voltage of 20 kV was used. Masses were detected in the 800-4000 Da range. The resolution was sufficient for monoisotopic mass determination. Either a three point external calibration on standard

peptides (angiotensin II, m/z 1046.54; α-melanocyte stimulating hormone, m/z 1664.80 and adrenocorticotropic hormone [18-39], m/z 2465.20) or a two point internal calibration with trypsin autolysis products (m/z 842.51 and m/z 2211.10) was applied.

MALDI-TOF PSD: For sequence information on selected peptides, PSD analysis was performed in 10-12 steps, lowering the reflectron voltage by 25% at each step, and eventually stitching the data together. Acceleration voltage was set at 25 kV. In PSD experiments average masses were determined.

LC-MS/MS experiments were performed on three different HPLC-MS systems with the following instrument settings:

1. Agilent 1100 nanoLC system on-line coupled to an XCT Plus ion trap. HPLC conditions: Samples were injected onto a Zorbax trap column (C18, 5 µm, 0.3 mm x 5 mm) at a flow rate of 10 µL/min for 5 min in solvent A and separated on a Zorbax column (C18, 3.5 µm, 75 µm x 150 mm) at a flow rate of 300 nL/min. The gradient was 5-45% B in 40 min. Solvent A was 0.1% formic acid in water, solvent B 0.1% formic acid in acetonitrile. MS conditions: Spray voltage was 2.1 kV. A drying gas at a temperature of 300 °C and a flow rate of 6 L/min was used. Ion charge control was set at $5x10^5$. Smart fragmentation amplitude was 30-200%. Samples were analyzed in information-dependent acquisition mode: MS acquisitions were followed by three collision-induced dissociation (CID) analyses on computer-selected multiply charged ions. In phosphopeptide analysis, occasionally a preferred mass list was included in data acquisition with m/z values calculated from MALDI-TOF data or derived from previous analyses. Data were processed by Data Analysis (v5.2).

2. Waters Micromass nanoAQUITY UPLC on-line coupled to a Premier Q-TOF mass spectrometer. HPLC conditions: Samples were injected onto a Symmetry trap column (C18, 5 µm, 180 µm x 20 mm) at a flow rate of 15 µL/min for 3 min in 3% of solvent B and separated on an AQUITY UPLC BEH C18 column (C18, 1.7 µm, 75 µm x 200 mm) at a flow rate of 300 nL/min with a gradient of 10-50% B in 40 min (short gradient) or 90 min (long gradient). Solvent A was 0.1% formic acid in water, solvent B 0.1% formic acid in acetonitrile. MS conditions: The spray voltage was set at 3-3.5 kV and the cone voltage at 26-28 V. Desolvation temperature was 180 °C. An information-dependent acquisition mode was used: 1 sec MS surveys were followed by 5 sec MS/MS scans on computer-selected multiply

charged ions. Collision energy was adjusted to m/z and charge of the precursor. A 30 sec dynamic exclusion was used. Data were processed by Mascot Distiller (v2.2.1.0).

3. Eldex MicroPro nanoLC system equipped with Spark Endurance autosampler and on-line coupled to a Thermo Scientific LCQ Fleet ion trap mass spectrometer. HPLC conditions: Samples were directly injected onto a Waters Atlantis dC18 column (C18, 3 µm, 75 µm x 100 mm) at a flow rate of 250 nL/min and separated by a gradient of 5-45% B in 40 min. Solvent A was 0.1% formic acid in water, solvent B 0.1% formic acid in acetonitrile. MS conditions: Spray voltage was 1-1.4 kV. Capillary temperature was 180 °C. Automatic gain control was set at $1x10^4$ ions for MS, $2x10^2$ for zoom scan and $5x10^3$ for MS(n). Each MS survey scan was followed by a 6 Da zoom-in scan on the most abundant ion in the survey. If it was found multiply charged then a CID acquisition followed. Dynamic exclusion was enabled for 30 sec. Collision energy was 35 V. Data were processed by Mascot Distiller (v2.2.1.0). In the analysis of O-GlcNAc modified peptides, MS(3) experiments and MS(2) experiments with multistage activation (MSA) were also performed with neutral loss masses of m/z 115.5, 77.0 and 58.0 for the 231 Da loss and m/z 53.0, 35.0 and 26.5 for the 105 Da loss. The 6 most abundant ions in the MS(2) spectrum were considered when selecting neutral loss products for the MS(3).

Database search: MS and MS/MS data were searched against the NCBI non-redundant protein database using on-line Mascot (v2.1 and v2.2, www.matrixscience.com) and Protein Prospector (v4.0.5-v5.1.8, prospector.ucsf.edu) search engines or an in-house Mascot server (v2.2.04). For MALDI-TOF data, monoisotopic masses within 200 ppm mass accuracy were considered in the MS-search, while average mass fragment ions within 1500 ppm for the PSD. For Agilent ion trap data, monoisotopic masses with precursor mass tolerance of ± 1 Da and fragment mass tolerance of ± 0.8 Da were submitted. For Waters Q-TOF data, monoisotopic masses with precursor mass tolerance of ± 50 ppm and fragment mass tolerance of ± 0.1 Da were used. For Thermo ion trap data, monoisotopic masses with precursor mass tolerance of ± 0.6 Da and fragment mass tolerance of ± 1 Da were submitted. No species restriction was used for protein identification. In the analysis of post-translational modifications, data were first searched against the full NCBI database with no species restriction, then the search was repeated with species restriction to the organism studied. In protein identification cysteine carbamidomethylation was considered as fixed modification,

and acetylation of protein N-termini, methionine oxidation and pyroglutamic acid formation from N-terminal glutamine residues as variable modifications. In phosphopeptide analysis, phosphorylation of serine, threonine and tyrosine residues were included as additional variable modifications.

4. Results and Discussion

4.1. Protein phosphorylation

4.1.1. Optimization of IMAC with different complexing metal ions

Phosphorylation is the most frequently studied post-translational modification. Since it is usually present at low stoichiometry, enrichment of phosphopeptides is highly recommended. Until recently, IMAC was the most popular method for phosphopeptide enrichment. Though there are a few other immobilized chelating agents described in the literature[59;60], the two most widely used are IDA and NTA in combination with Fe^{3+} and Ga^{3+}. In the present experiments, agarose-bound NTA was loaded with different multivalent metal ions in order to investigate whether an improved IMAC system could be developed. Transition metal ions are well known for their complexing properties. From this group Co^{2+}, Cu^{2+}, Hg^{2+}, Mn^{2+}, Ni^{2+}, Zn^{2+} and Fe^{3+} were selected. Alkaline earth metals - Ca^{2+}, Mg^{2+} and Ba^{2+} - were also tested because of the low solubility of their phosphate salts. Successful isolation of phosphoproteins by $Al(OH)_3$ [61] initiated experiments with Al^{3+}. Finally, Pb^{2+} was also included in this study.

A mixture of synthetic phosphopeptides modified at serine, threonine and tyrosine residues and fetuin tryptic digest (Appendix) served as standards. Bovine fetuin (P12763) is a glycosylated serum protein containing four phosphorylation sites at Ser-138, Ser-320, Ser-323 and Ser-325. The sequence, mass and pI (calculated by scansite.mit.edu/calc_mw_pi.html) of the five phosphopeptides observed in the fetuin digest (denoted as F1-F5) as well as of the four synthetic phosphopeptides (P1-P4) are summarized in Table 2. The synthetic phosphopeptides P1-P4 are singly phosphorylated. Fetuin phosphopeptide F2 corresponds to the same phosphorylation site as F1 with one missed cleavage site. Phosphopeptides F3-F5 represent identical sequences singly, doubly and triply phosphorylated. The masses of the

	Sequence	MH^+	pI
F1	CDSSPDpSAEDVR	1465.53	3.40
F2	CDSSPDpSAEDVRK	1593.63	3.85
F3	HTFSGVApSVESSSGEAFHVGK	2199.97	5.22
F4	HTFSGVApSVESSpSGEAFHVGK	2279.94	4.49
F5	HTFSGVApSVEpSSpSGEAFHVGK	2359.90	3.83
P1	RIEVALpTK	1009.54	5.76
P2	RKKRIpSALPG	1205.69	9.99
P3	ELFDDPSpYVNVQNL	1732.75	2.68
P4	NQLYNELNLGREEpYDVL	2161.98	3.53

Table 2. Sequence, protonated molecular mass and pI of the phosphopeptides included in IMAC phosphopeptide enrichment. F1-F5 are fetuin phosphopeptides, P1-P4 are synthetic phosphopeptides.

phosphopeptides are in the 1000-2400 m/z range. P1 and F3 are neutral, P3, P4, F1, F2, F4 and F5 are acidic phosphopeptides, while P2 is basic. Five picomoles of phosphopeptide standards or 1 pmole of fetuin digest were loaded on the IMAC beads. The binding conditions are listed in Table 1 in Materials and methods.

The ability of the metal ions to capture the synthetic phosphopeptides P1-P4 as well as their selectivity towards the phosphopeptides F1-F5 in the fetuin tryptic digest was considered when evaluating the results (Figure 1). Fe^{3+} and Cu^{2+} behaved similarly as they both trapped all four synthetic peptides, but when analyzing the fetuin tryptic digest, they exhibited high nonspecific binding of acidic peptides. The difference between these two ions is their pH optimum, pH 4-8 vs. pH 6.5-8. Al^{3+} and Pb^{2+} could also bind all four synthetic phosphopeptides at pH 6.5-8, with markedly lower nonspecific binding in the fetuin digest. Al^{3+} worked well in parallel experiments, but results with Pb^{2+} did not prove to be reproducible. Mn^{2+}, Zn^{2+}, Hg^{2+}, Co^{2+} and Ni^{2+} could bind only three (P2, P3 and P4) out of the four synthetic standards (optimal pH 4), yet they exhibited higher selectivity for the phosphate group when analyzing the fetuin tryptic digest. Ca^{2+}, Mg^{2+} and Ba^{2+} showed the same preferences as the former group with a shifted pH optimum (pH 4-6.5).

The pH of the loading solution has high impact on the binding of phosphopeptides to the IMAC beads. One aspect of pH dependence in IMAC phosphopeptide enrichment is the complex stability between the metal ion and the chelating agent. The higher is the stability constant the lower pH can be applied. For Fe^{3+} and Ga^{3+} ions the generally used solvent is 0.1 M acetic acid (pH 2), as they have high stability constants ($\lg K_{\text{Fe-NTA}}$ 15.8), while Ca^{2+}, Mg^{2+} and Ba^{2+} – having lower stability constants ($\lg K_{\text{M-NTA}}$ 4.7-6.3) – cannot be used at such a low pH. Another aspect of pH is peptide protonation. The charge state of a peptide at a

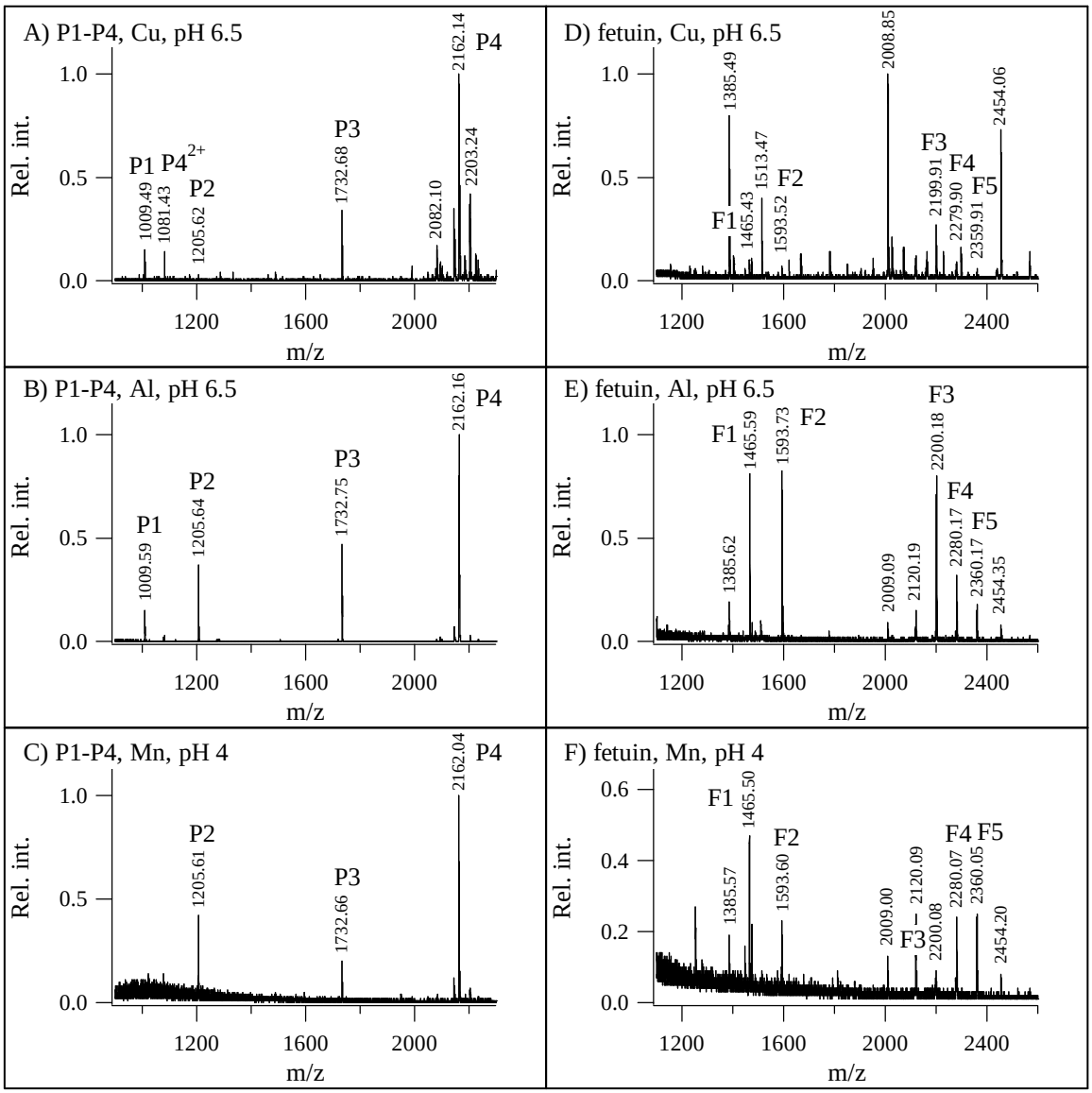


Figure 1. Performance of different metal ions in the capture of synthetic phosphopeptides P1-P4 (A-C) and in the enrichment of the phosphopeptides F1-F5 from fetuin tryptic digest (D-F). Spectra were recorded on a MALDI-TOF mass spectrometer.

given pH depends on the pK values of the acidic and basic groups at the peptide termini and the amino acid side chains. When binding was performed at low pH (binding condition #5), only the acidic phosphopeptides P3 and P4 were captured. Flow-through of the neutral and basic phosphopeptides P1 and P2 is most likely due to charge repulsion. Considering both aspects, an optimal pH for the particular metal ion has to be found. The effect of pH on the performance of Al-NTA in the capture of the phosphopeptide standards P1-P4 is presented in

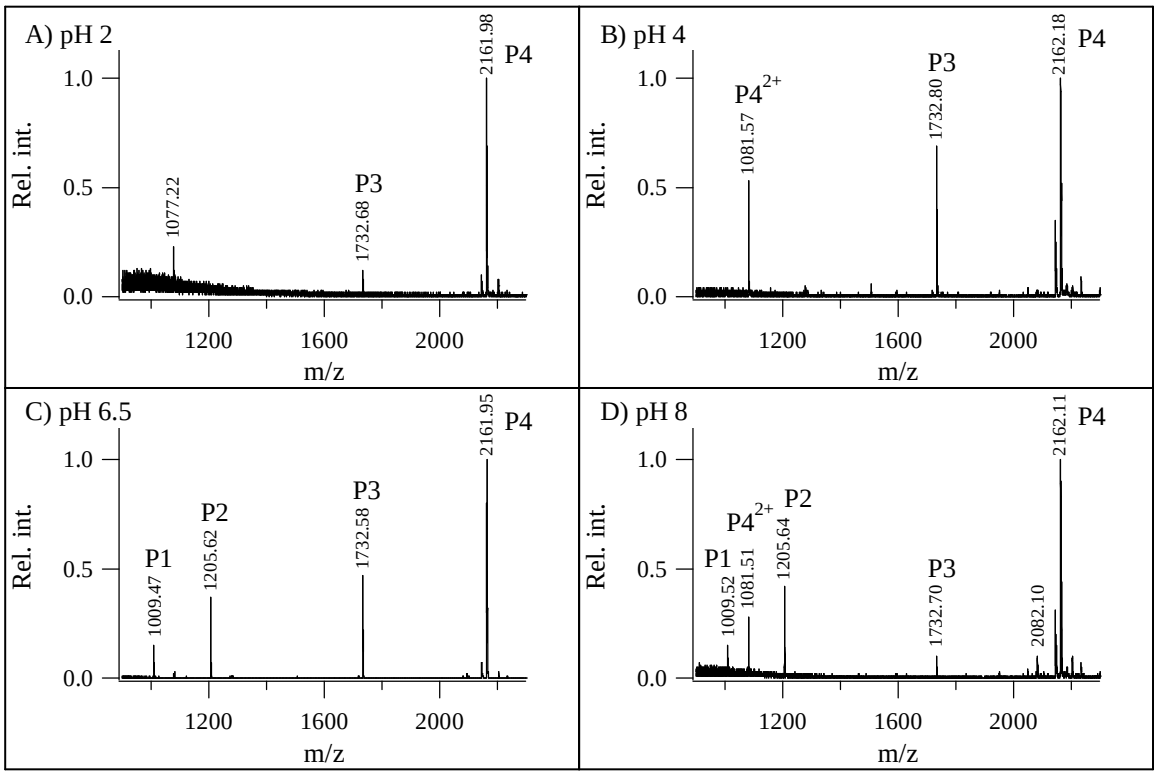


Figure 2. The effect of pH of the loading solution on phosphopeptide enrichment by Al-NTA. Spectra were acquired on a MALDI-TOF mass spectrometer.

Figure 2. At low pH (pH<4) only the acidic phosphopeptides P3 and P4 were retained. Optimal binding was observed at pH 6.5. At higher pH (pH 8) the performance dropped again, most likely due to the hydrolysis of Al^{3+}. Nevertheless, shifting the pH could be used for the fractionation of the phosphopeptides.

There are several protocols in the literature using organic modifier in the wash and/or load buffers to improve IMAC performance. Hexafluoroisopropanol as an additive was found to increase the recovery of the phosphopeptides, especially of the basic ones, most likely due to conformational factors[62]. In Fe-NTA or Fe-IDA enrichment procedures acetonitrile is often included in the load and wash buffer in order to reduce nonspecific binding to the beads and/or to wash off nonphosphorylated peptides[19;63;64]. Enrichment performance of Al-NTA without or with acetonitrile in the wash buffer (binding conditions #1 and #2) is compared in Figure 3. Indeed, organic modifier added to the wash buffer significantly reduced nonspecific binding to the beads. The nonphosphorylated peptides observed in binding condition #1

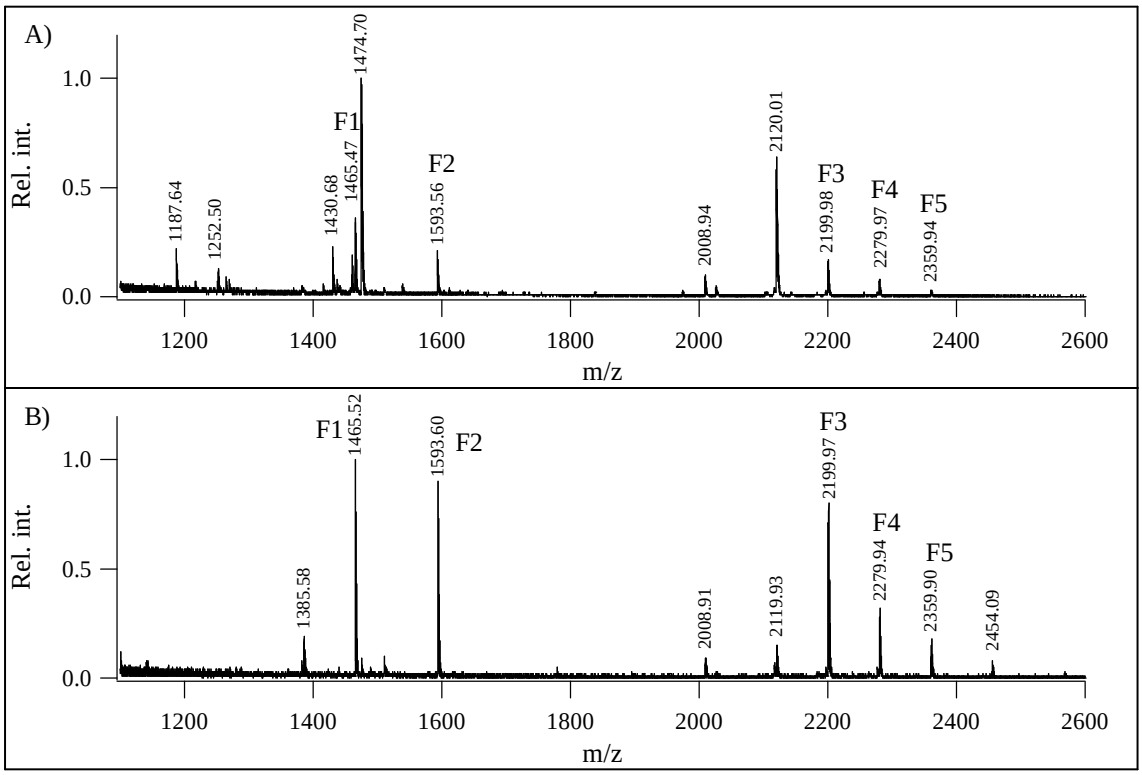


Figure 3. Al-NTA enrichment of phosphopeptides from fetuin tryptic digest without (A) and with 30% acetonitrile (B) in the wash buffer. Spectra were acquired by MALDI-TOF mass spectrometry.

(without acetonitrile) are the same that were captured by unloaded NTA beads, thus this nonspecific binding comes from the interaction with the solid support.

Another source of nonspecific binding is the capture of acidic peptides, as seen for Cu^{2+} in Figure 1D. The use of acetic acid or acetate salts in the load and wash buffers aims to reduce this type of nonspecific binding. To further reduce the aspecific binding of acidic nonphosphorylated peptides, methyl esterification of the carboxylic acid side chains and the C-terminus can be performed[21]. However, incomplete methylation of peptides with multiple aspartic acid or glutamic acid residues, and partial hydrolysis of asparagine and glutamine residues followed by esterification under the conditions used (2 M hydrochloric acid in methanol at room temperature for 2 h) may lead to signal splitting and therefore reduced sensitivity as demonstrated in Figure 4. In order to avoid the incorporation of additional carboxylic groups into the proteins, iodoacetamide is preferred over iodoacetic acid for the alkylation of the cysteine residues.

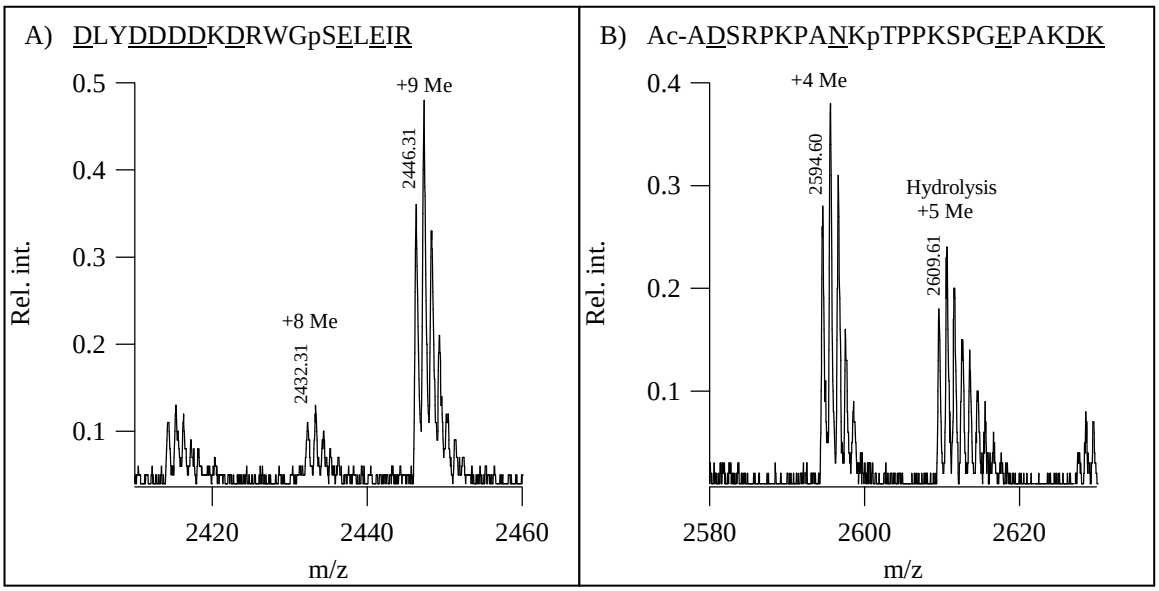


Figure 4. Signal splitting due to incomplete esterification of aspartic or glutamic acids (A) or hydrolysis and esterification of asparagine (B) during methyl esterification. (MALDI-TOF data)

As Al^{3+} performed best in preceding experiments, its efficiency was tested in a complex digest mixture as well. 5 picomoles of BSA digest were added to 500 femtomoles of fetuin digest and loaded on Al-NTA using binding condition #6. Though some aspecific binding of nonphosphorylated BSA peptides was observed in the eluate, fetuin

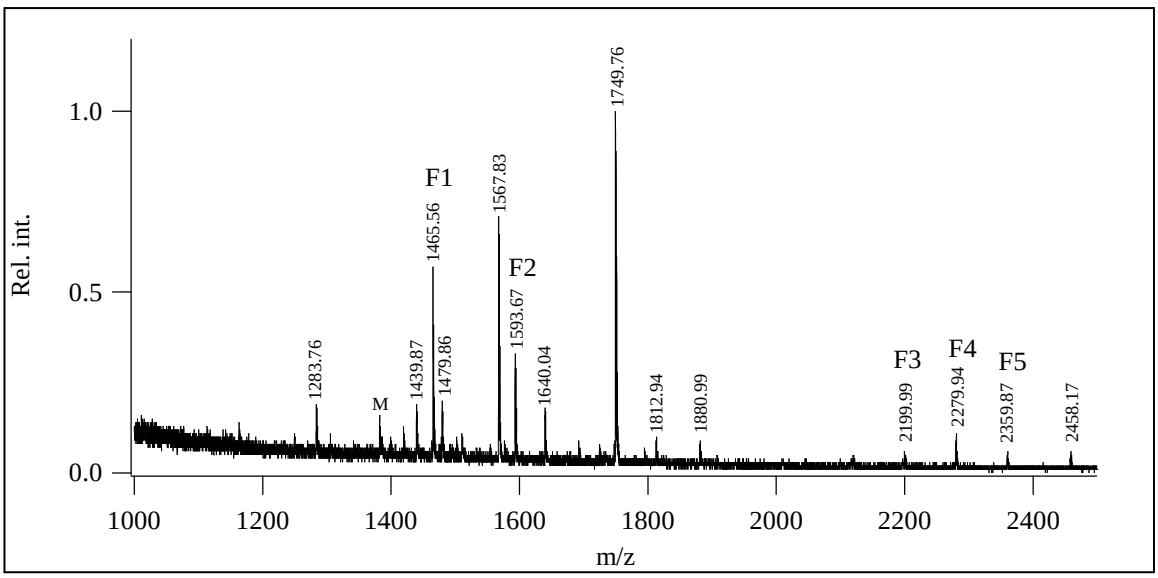


Figure 5. Enrichment of fetuin phosphopeptides from a 1:10 mixture of fetuin:BSA digest. Fetuin phosphopeptides are denoted as F1-F5. M stands for metastable ion. (MALDI-TOF data)

phosphopeptides were efficiently enriched (Figure 5). When analyzing complex mixtures, higher ammonium acetate concentration is recommended to suppress nonspecific binding to the beads.

Under optimized conditions, Al^{3+} performed superior compared to other metal ions investigated, in terms of efficiency as well as selectivity. Despite of possible sample losses, methyl esterification is recommended with Fe^{3+} or Cu^{2+} to eliminate aspecific binding of acidic nonphosphorylated peptides. Though other bivalent cations were less efficient in phosphopeptide capture, they might also find some applications due to their higher selectivity for the phosphate group. In conclusion, our experiments did not lead to significant improvement of the existing IMAC methods. However, Al^{3+}-IMAC offers an alternative with the attractive feature of phosphopeptide-fractionation, that we intend to further investigate.

4.1.2 In vitro and in vivo phosphorylation of TPPP/p25

TPPP/p25 is a brain specific protein involved in the assembly of tubulin/microtubules[65]. *In vitro* phosphorylation experiments showed a connection between phosphorylation of TPPP/p25 and its microtubule assembling activity. Therefore, *in vitro* phosphorylated TPPP/p25 was subjected to MS analysis to reveal the phosphorylation sites.

Recombinant human TPPP/p25 (O94811) was *in vitro* phosphorylated by extracellular signal-regulated kinase 2 (ERK2), cyclin-dependent kinase 5 (CDK5) and cAMP-dependent protein kinase (PKA). The 1D SDS-PAGE fractionated protein bands were in-gel digested with trypsin. These experiments were simultaneously performed with our IMAC-optimization studies. Thus, two established methods were used to enrich the phosphopeptides: IMAC and titanium dioxide. IMAC was performed on Fe-NTA beads with prior methyl esterification. For titanium dioxide enrichment, a modified protocol of Larsen[22] was applied. After phosphopeptide enrichment, samples were analyzed by MALDI-TOF MS, MALDI-TOF PSD and LC-MS/MS on an ion trap mass spectrometer. Phosphopeptides were first identified by the 80 Da mass shift in comparison to the predicted molecular weights of tryptic peptides. Their identity was further confirmed by the β-elimination of phosphoric acid (-98 Da) from the precursor ion under CID or PSD conditions. The site of modification was assigned considering the 80 Da mass shift of the appropriate peptide fragments. Results are

m/z (z)	Sequence	Site	Kinase	Enrichment
1915.9 (1+)	AANRpTPPKSPGDPSKDR	Thr-14	CDK5	Fe-NTA
609.8 (2+)	TPPKpSPGDPSK	Ser-18	CDK5, ERK2	Fe-NTA
502.1 (3+)	TPPKpSPGDPSKDR	Ser-18	CDK5, ERK2	Fe-NTA
488.1 (3+)	TPPKpSPGDPSKDR	Ser-18	CDK5, ERK2	TiO_2
625.5 (3+)	AANRTPPKpSPGDPSKDR	Ser-18	CDK5, ERK2	TiO_2
1027.9 (3+)	RLpSLESEGAGEGAAASPELSALEEAFRR	Ser-32	PKA	Fe-NTA
780.0 (2+)	NVpTVTDVDIVFSK	Thr-92	PKA	Fe-NTA
506.3 (2+)	AIpSSPTVSR	Ser-159	PKA	Fe-NTA
506.3 (2+)	AISpSPTVSR	Ser-160	CDK5, ERK2	Fe-NTA
499.5 (2+)	AISpSPTVSR	Ser-160	CDK5, ERK2	TiO_2

Table 3. Identification of phosphorylation sites in human recombinant TPPP/p25 *in vitro* phosphorylated by CDK5, ERK2 and PKA. Phosphorylated residues are indicated as pS or pT.

summarized in Table 3. CDK5 and ERK2 phosphorylated Ser-18 and Ser-160, CDK5 also phosphorylated Thr-14, while PKA modified Ser-32, Thr-92 and Ser-159. These phosphorylation sites are in agreement with the consensus sequences of the kinases: -S/T-P- for ERK2 and CDK5, and -K/R-K/R-X-S/T-I/L/V- for PKA. The phosphorylation site assignment is illustrated by the comparison of the CID spectra of the nonphosphorylated ^{157}AISSPTVSR^{165} peptide vs. the peptides phosphorylated at adjacent serines, i.e. Ser-159 and Ser-160, according to the substrate specificity of the different kinases PKA and ERK2, respectively (Figure 6). The fragment ion at m/z 573.4 observed in all three CID spectra corresponds to the nonphosphorylated y_5 fragment. In the PKA modified peptide (middle panel), the nonphosphorylated fragment y_6 at m/z 660.5 excluded Ser-160 as the site of modification, and thus, unambiguously assigned the site of phosphorylation at Ser-159. In the ERK2 modified peptide (lower panel), the y_6 fragment is observed at m/z 740.4 which is 80 Da higher in mass than the nonphosphorylated one clearly identifying Ser-160 as the site of phosphorylation. A 98 Da loss from the precursor as well as fragment ions is characteristic for phosphopeptides modified at serine and threonine residues.

In order to confirm biological relevance of the *in vitro* phosphorylation analyses, experiments were also performed to reveal the *in vivo* phosphorylated sites of TPPP/p25. The *in vivo* phosphorylation of TPPP/p25 was investigated on the bovine homologue (Q27957). Sequence alignment of the human and bovine TPPP/p25 is shown in Appendix. The TPPP/p25 preparation isolated from bovine brain was subjected to 1D SDS-PAGE fractionation. Mass spectrometry analysis of the in-gel digested SDS-PAGE bands identified TPPP/p25 in the preparation with minor contamination of myelin basic protein (P02687). The

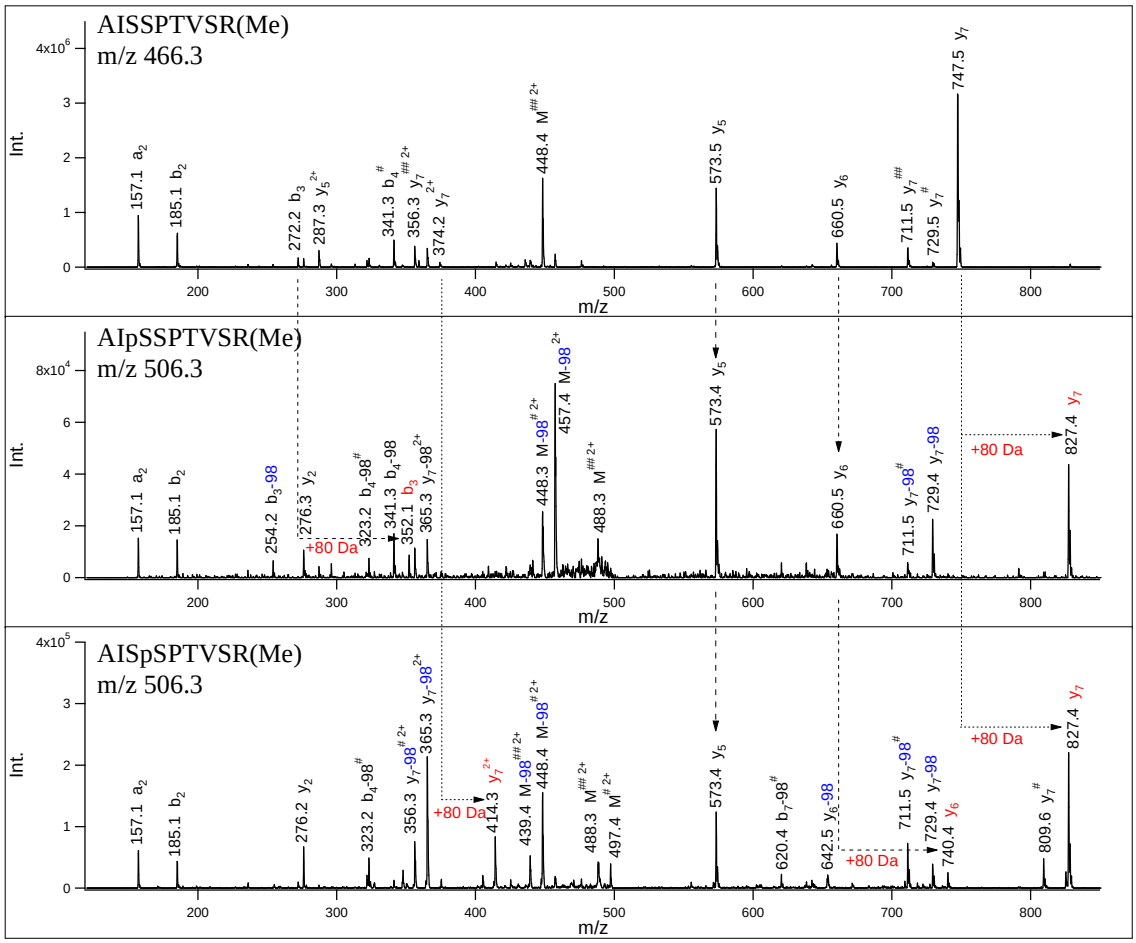


Figure 6. Phosphorylation site assignment in the AISSPTVSR(Me) peptide. Upper panel: nonphosphorylated, middle panel: phosphorylated by PKA, lower panel: phosphorylated by ERK2. Fragment ions observed with a mass shift of 80 Da indicate phosphorylation, 98 Da losses result from the β-elimination of phosphoric acid. # stands for water loss. Spectra were acquired on an ion trap mass spectrometer.

first set of phosphopeptide analysis was performed on the in-gel digested TPPP/p25. Even without phosphopeptide enrichment, low abundance peaks featuring the telltale 80 Da mass shift at m/z 2538.7 and m/z 2954.9 in the MALDI-TOF mass spectrum indicated phosphorylation in peptide sequences Ac-[2-24] and [28-56], respectively. After methyl esterification of the acidic side chains and the C-terminus followed by phosphopeptide enrichment by IMAC on Fe-NTA beads, additional singly and doubly phosphorylated peptides were observed in the MALDI-TOF mass spectrum (Table 4). PSD spectrum of the peptide at m/z 2954.9 produced fragments at m/z 1165.0 (y_{10}), 1635.2 (y_{16}), 1764.0 (y_{18}) and 2077.2 (y_{22}) revealing ^{28}RLpSLEAEGAGEGAAAAGAELSALEEAFRK^{56}, thus

MH^+	Sequence	Position	Phosphate	Enrichment
1656.9	Ac-ADSRPKPANKTPPK	Ac-[2-15]	1	✓
2296.0	Ac-ADSRPKPANKTPPKSPGEPAK	Ac-[2-22]	1	×
2337.2	Ac-ADSRPKPANKTPPKSPGEPAK	Ac-[2-22]	1	✓
2538.7	Ac-ADSRPKPANKTPPKSPGEPAKDK	Ac-[2-24]	1	×
2594.5	Ac-ADSRPKPANKTPPKSPGEPAKDK	Ac-[2-24]	1	✓
2674.5	Ac-ADSRPKPANKTPPKSPGEPAKDK	Ac-[2-24]	2	✓
1852.0	PKPANKTPPKSPGEPAK	[6-22]	1	✓
1626.8	PANKTPPKSPGEPAK	[8-22]	1	✓
1884.0	PANKTPPKSPGEPAKDK	[8-24]	1	✓
1473.7	TPPKSPGEPAKDK	[12-24]	1	✓
2924.3	RLSLEAEGAGEGAAAAGAELSALEEAFR	[28-55]	1	✓
2954.9	RLSLEAEGAGEGAAAAGAELSALEEAFRK	[28-56]	1	×
3052.3	RLSLEAEGAGEGAAAAGAELSALEEAFRK	[28-56]	1	✓

Table 4. Phosphopeptides observed by MALDI-TOF MS in the tryptic digest of TPPP/p25 isolated from bovine brain without or after IMAC enrichment.

phosphorylation at Ser-30. Further PSD spectra were acquired to determine the phosphorylation sites in the N-terminal Ac-^{2}ADSRPKPANKTPPKSPGEPAKDK^{24} peptide. Fragments indicated phosphorylation at Thr-12, but data were ambiguous. Therefore, a second set of phosphopeptide analysis was performed. In this experiment, TPPP/p25 isolated from bovine brain was in-solution digested with trypsin. The digest was subjected to phosphopeptide enrichment by titanium dioxide, then analyzed by LC-MS/MS on an ion trap mass spectrometer. A series of CID spectra unambiguously identified phosphorylation at Thr-12, as listed in Table 5. Phosphorylation site assignments were based on preferential peptide backbone cleavages between ^{7}LysPro^{8} and ^{12}ThrPro^{13}. Determination of the phosphorylation sites in the doubly phosphorylated Ac-[2-24] peptide was accomplished from the quadruply charged precursor at m/z 655.7 (Figure 7). The fragment ion observed at m/z 665.9 can only be assigned as y_{12}^{2+} with one phosphate in this fragment, i.e. phosphorylation

m/z (z)	Sequence	Position	Phosphate	Site(s)
575.0 (4+)	Ac-ADSRPKPANKpTPPKSPGEPAK	Ac-[2-22]	1	Thr-12
603.6 (4+)	Ac-ADSRPKPANKpTPPKSPGEPAKD	Ac-[2-23]	1	Thr-12
483.1 (5+)	Ac-ADSRPKPANKpTPPKSPGEPAKD	Ac-[2-23]	1	Thr-12
508.9 (5+)	Ac-ADSRPKPANKpTPPKSPGEPAKDK	Ac-[2-24]	1	Thr-12
655.7 (4+)	Ac-ADSRPKPANKpTPPKpSPGEPAKDK	Ac-[2-24]	2	Thr-12, Ser-16
537.1 (5+)	Ac-ADSRPKPANKpTPPKSPGEPAKDKAA	Ac-[2-26]	1	Thr-12
671.0 (4+)	Ac-ADSRPKPANKpTPPKSPGEPAKDKAA	Ac-[2-26]	1	Thr-12
691.3 (4+)	Ac-ADSRPKPANKTPPKSPGEPAKDKAA	Ac-[2-26]	2	
562.7 (5+)	Ac-ADSRPKPANKTPPKSPGEPAKDKAAK	Ac-[2-27]	1	

Table 5. Phosphopeptides observed by LC-MS/MS in the tryptic digest of TPPP/p25 isolated from bovine brain after TiO_2 enrichment. Phosphorylated residues are indicated as pS or pT.

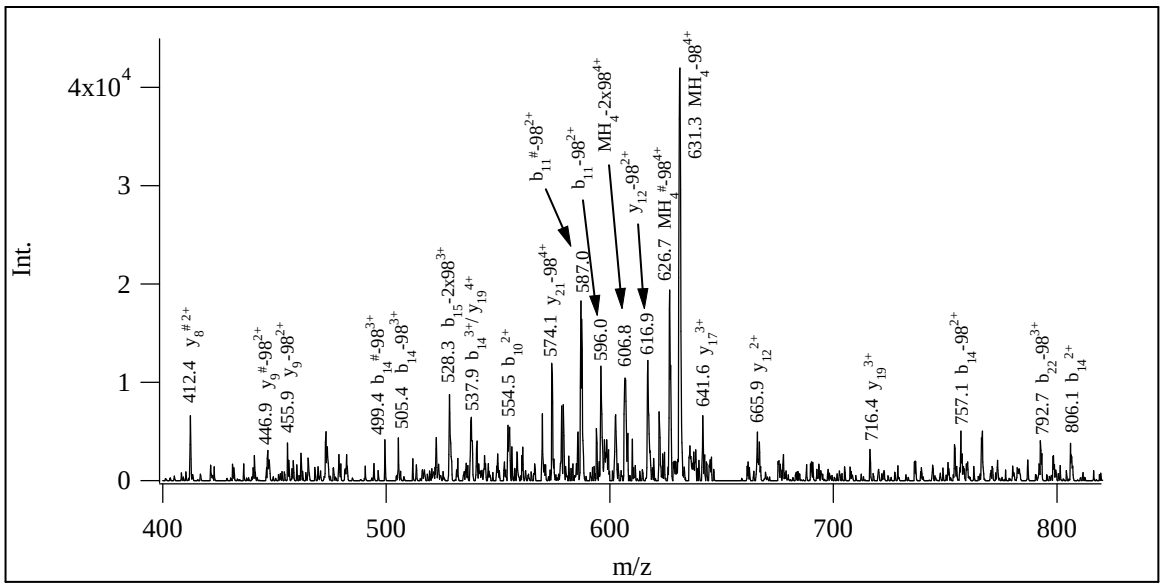


Figure 7. Phosphorylation site assignment in the Ac-ADSRPKPANKpTPPKpSPGEPAKDK [2-24] bovine TPPP/p25 peptide observed at m/z 655.7 (4+). Fragments assigned as y_{12}^{2+} and y_{17}^{3+} unambiguously identified Thr-12 and Ser-16 as the phosphorylation sites. The spectrum was acquired on an ion trap mass spectrometer.

at Ser-16. The fragment ion observed at m/z 641.6 corresponds to the doubly phosphorylated y_{17}^{3+} thus confirming an additional phosphate at position Thr-12. These phosphorylation site assignments are in agreement with our previous data found *in vitro* in the human protein, as well as with *in vivo* data found in the mouse and human TPPP/p25[66-68].

4.1.3. Phosphorylation of the Medicago RRK1 kinase and its activator the ROP6 GTPase

Rho-type GTPases of plants (ROP) are small GTP-binding proteins, key players in signal transduction. In *Medicago* interaction of the ROP6 GTPase with receptor like kinase 1 and 2 (RRK1 and RRK2) was demonstrated in yeast two-hybrid experiments. As *in vitro* kinase assays showed, the GTP-bound (active) ROP6 GTPase activated both kinases. Moreover, autophosphorylation of the kinases as well as phosphorylation of the ROP6 GTPase was also observed. MS analyses were performed to map the phosphorylation sites in the RRK1 kinase and the ROP6 GTPase. The proteins were in-gel digested with trypsin and the phosphopeptides enriched by Fe-NTA IMAC after methyl esterification or by titanium dioxide. The samples were analyzed by MALDI-TOF MS, MALDI-TOF PSD and by LC-MS/MS on ion trap and Q-TOF mass spectrometers.

Sequence	Position	Phosphate	Enrichment	MALDI-TOF MS	PSD or CID
GSSHHHHHHSSGLVPR	[h2-h17]	1	TiO_2	1849.0	1849.0 (1+)
GSHMASMTGGQQMGR	[h18-h32]	1	Fe-NTA	1629.7	1629.7 (1+)
GSHMASMoxTGGQQMGR	[h18-h32]	1	Fe-NTA	1645.7	-
GSEFGpTRR	[h33-r2]	1	Fe-NTA/TiO_2	1017.5/989.6	1017.5 (1+)/495.2 (2+)
YIRTGSFK	[r3-r10]	1	Fe-NTA/TiO_2	1065.5/1051.6	-/-
YIRpTGSFKR	[r3-r11]	1	Fe-NTA/TiO_2	1221.6/1207.7	1221.6 (1+)/-
YIRpTGpSFKR	[r3-r11]	2	TiO_2	1287.7	429.8 (3+)

Table 6. Autophosphorylation in the recombinant *Medicago* RRK1 kinase. Phosphorylated residues are indicated as pS or pT if the modification site could be located. In the position of the peptides the His-tag [h1-h39] and the native RRK1 [r2-r381] sequences were separately considered.

The autophosphorylation sites in the His-tagged RRK1 kinase (C1L342, Appendix) are summarized in Table 6. The N-terminal ^{h2}GSSHHHHHHSSGLVPR^{h17} phosphopeptide (h means within the His-tag sequence) was observed only in the titanium dioxide enrichment. The preference of titanium dioxide to hexaHis-peptides was noticed in other samples as well. In contrast, the ^{h18}GSHMASMTGGQQMGR^{h32} phosphopeptide was observed only after Fe-NTA enrichment. However, the site of the modification could not be located in either of these phosphopeptides. The ^{h33}GSEFGTRR^{r2} phosphopeptide - still part of the His-tag sequence - was observed in both enrichments. The PSD spectrum of the methyl esterified precursor observed at m/z 1017.5 after Fe-NTA enrichment as well as the CID spectrum of the peptide detected at m/z 495.2 (2+) after titanium dioxide enrichment revealed phosphorylation at Thr-h38. Within the native RRK1 sequence two phosphorylation sites were identified, both in the N-terminal ^{r3}YIRTGSFKR^{r11} peptide. Modification at Thr-r6 was determined from the PSD spectrum of the singly phosphorylated peptide observed at m/z 1221.6 after Fe-NTA enrichment. The doubly phosphorylated peptide observed after titanium dioxide enrichment revealed the second phosphorylation site at Ser-r8.

The determination of the phosphorylation sites in the constitutive active mutant form of the ROP6 GTPase (Q3ZVR9, mutation: G15V) was performed on differently His-tagged versions of the protein (Table 7). In the first construct (H_1, Appendix) two phosphopeptides were detected after enrichment: WGpSELEIR [$h_1$34-$h_1$41] and DLYDDDDKDRWGpSELEIR [$h_1$24-$h_1$41], both pointing to the same phosphorylation on Ser-$h_1$36 in the His-tag region. In the second construct (H_2) this serine was eliminated and the phosphorylation experiment repeated. This time phosphorylation was found on the peptides IPMSGSR [$h_2$38-r5] and YRIPMSGSR [$h_2$36-r5] at the junction of the His-tag and ROP6

Construct	Sequence	Position	Phosphate
H_1	WGpSELEIR	[$h_1$34-$h_1$41]	1
H_1	DLYDDDDKDRWGpSELEIR	[$h_1$24-$h_1$41]	1
H_2	IPMSGSR	[$h_2$38-r5]	1
H_2	YRIPMSGSR	[$h_2$36-r5]	1
H_3	GGSHHHHHHGMASMTGGQQMGR	[$h_3$2-$h_3$23]	1, 2
H_3	DLQMVPYGNpSMSGSR	[$h_3$39-r5]	1

Table 7. Phosphorylation of differently His-tagged recombinant *Medicago* ROP6 GTPase. Phosphorylated residues are indicated as pS or pT if the modification site could be located. In the position of the peptides the His-tag and the native ROP6 GTPase sequences were separately considered.

GTPase sequences, though the exact site of the modification could not be located. Since this sequence might have become target of the kinase due to the preceding RYRIP pattern, a third His-tag variant (H_3) was constructed. In this protein phosphorylation was detected again only on His-tag peptides: GGSHHHHHHGMASMTGGQQMGR [$h_3$2-$h_3$23] and DLQMVPYGNSMSGSR [$h_3$39-r5], both observed in multiple methionine oxidation forms. The hexaHis-containing peptide was found singly and doubly phosphorylated, but the sites of the modification could not be located. In the other peptide the phosphorylation site was identified at Ser-$h_3$48, just preceding the native ROP6 GTPase sequence. As all the prior modifications were linked to the N-terminal His-tag, another version was constructed with a C-terminal His-tag (H_4). In this protein, however, no phosphorylation could be confirmed. These results indicated that the phosphorylation of the ROP6 GTPase was artificially induced in the N-terminal His-tag sequence.

4.1.4. In vitro and in vivo phosphorylation of Drosophila calpain B

Calpains are Ca^{2+} dependent SH-proteases catalyzing irreversible limited proteolysis of their substrates[69]. Structural predictions of the *Drosophila* calpain B (Q9VT65, Appendix) pointed to possible phosphorylation sites in the protein. This initiated *in vitro* phosphorylation experiments, that showed elevated autoproteolytic activation of the enzyme, enhanced proteolysis of its substrates, as well as increased Ca^{2+} sensitivity upon phosphorylation. The phosphorylation sites involved in these processes were determined by MS analysis.

In vitro phosphorylation of the recombinant calpain B was carried out by three kinases: PKA, ERK1 and ERK2. The 1D SDS-PAGE fractionated protein bands were in-gel digested and the digests subjected to titanium dioxide phosphopeptide enrichment. MALDI-

m/z (z)	Sequence	Site	Kinase
591.7 (2+)	TGpSIDGFHLR	Ser-845	PKA
764.8 (2+)	qNpSVSKGDFQSLR	Ser-240	PKA
515.9 (3+)	QNpSVSKGDFQSLR	Ser-240	PKA
689.7 (3+)	IAPSLPPPpTPKEEDDPQR	Thr-747	ERK1, ERK2

Table 8. Phosphopeptides observed by LC-MS/MS in the *Drosophila* calpain B digests after TiO_2 enrichment. Phosphorylated residues are indicated as pS or pT.

TOF and LC-MS/MS analyses were performed on the enriched fractions. Two PKA and a single ERK phosphorylation site could be confirmed (Table 8). PKA phosphorylation at Ser-845 was identified from the precursor at m/z 591.7 (2+) corresponding to ^{842}TGpSIDGFHLR^{852}. The precursors at m/z 764.8 (2+) and 515.9 (3+) represent the same phosphopeptide ^{238}QNpSVSKGDFQSLR^{250}, in the former one the N-terminal glutamine is cyclized. ERK phosphorylation at Thr-747 was determined from the CID spectrum of the precursor at m/z 689.9 (3+) corresponding to ^{739}IAPSLPPPpTPKEEDDPQR^{756}.

It has already been mentioned that peptides phosphorylated on serine or threonine residues generally display abundant 98 Da losses from the precursor ion as well as the fragments upon collisional activation, as seen previously for TPPP/p25 phosphopeptides. This characteristic fragmentation of the modified peptides is often used for their targeted analysis: the 98 Da loss enables phosphoserine or phosphothreonine containing precursor ion selection by neutral loss scan in a Q-trap instrument or to perform MS^3 experiments in an ion trap. Interestingly, the calpain B ^{739}IAPSLPPPpTPKEEDDPQR^{756} phosphopeptide was different in this respect. No phosphoric acid loss from the precursor ion was detected that would have facilitated its recognition as a phosphopeptide. However, comparison to the CID spectrum of the nonphosphorylated peptide (m/z 663.0) confirmed its identity (Figure 8). The nonphosphorylated peptide and its phosphorylated counterpart show the same fragmentation pattern. The high abundance fragments observed correspond to peptide backbone cleavages N-terminally to proline residues. The site of modification was determined from the mass shift of the product ions. The b_5 fragment observed at m/z 482.3 in both CID spectra indicated that Ser-742 is unmodified, whereas the 80 Da mass shift in the y_{13}-y_{17} ion series – either in the +2 or +3 charge state – confirmed phosphorylation at Thr-747.

Autoradiography pointed to other ERK phosphorylation site(s) in the N-terminal 20 kDa segment of the protein that could not be identified by MS. The two potential ERK

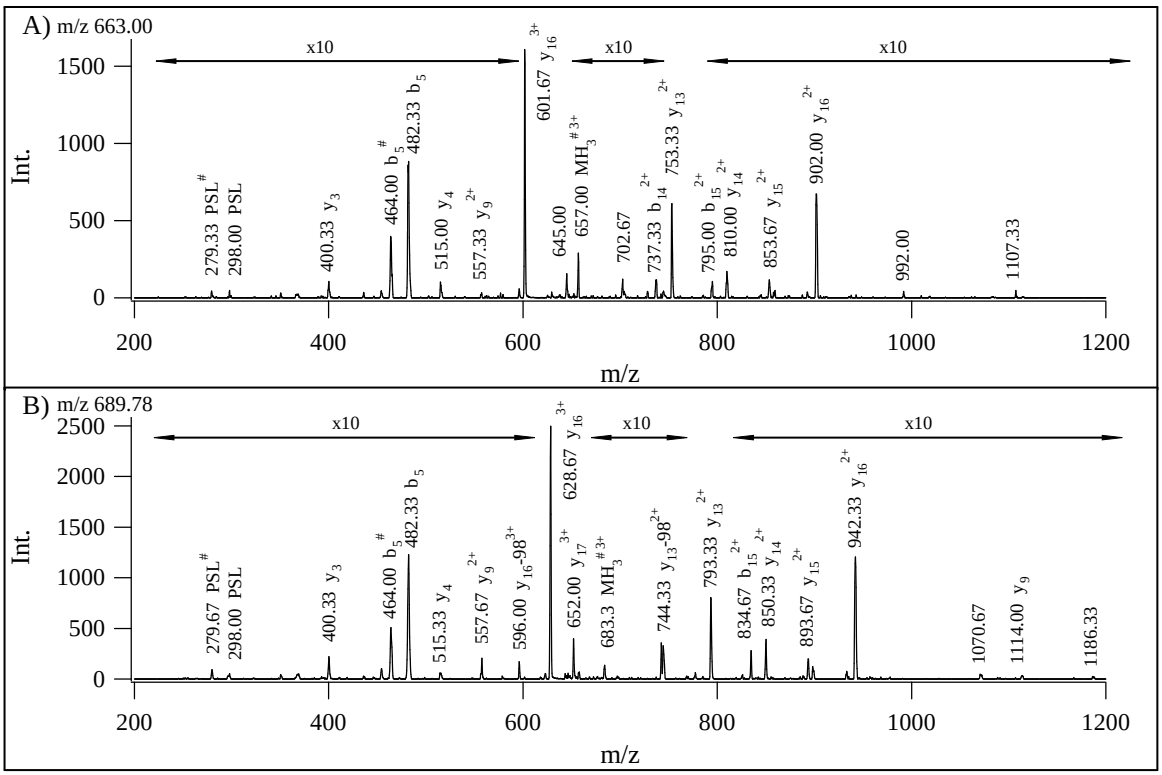


Figure 8. CID spectra of the nonphosphorylated (A) and ERK2 phosphorylated (B) calpain B IAPSLPPPTPKEEDDPQR peptide. Spectra were acquired on an ion trap mass spectrometer.

phosphorylation sites in this region of the protein are Thr-102 and Ser-176, but there are no arginines and lysines near and far to these sites, thus the predicted tryptic peptide (m/z 15626.4) is much bigger than the optimal size both for extraction from the gel and successful LC-MS/MS analysis or MALDI-TOF mass measurement in the presence of multiple much smaller peptides.

To test if the *in vitro* phosphorylation events occur *in vivo*, calpain B was immunoprecipitated from untreated S2 *Drosophila* cells, fractionated by SDS-PAGE and subjected to our general digestion and titanium dioxide enrichment protocol. None of the phosphopeptides observed *in vitro* were detected in this sample indicating that calpain B was not phosphorylated in resting cells. No calpain B phosphorylation could be confirmed in cells treated with the phosphatase inhibitor calyculin-A either. Then cells were treated with epidermal growth factor (EGF) to activate the ERK pathway and protein dephosphorylation was prevented by calyculin-A. In this experiment two residues, Thr-747 and Ser-240 were found phosphorylated. The extracted ion chromatograms at m/z 690 and 765 corresponding to

the calpain B phosphopeptides IAPSLPPPpTPKEEDDPQR and qNpSVSKGDFQSLR, respectively, are shown in Figure 9. The identity of the phosphopeptides was confirmed by CID experiments. The corresponding peaks are missing from the untreated sample.

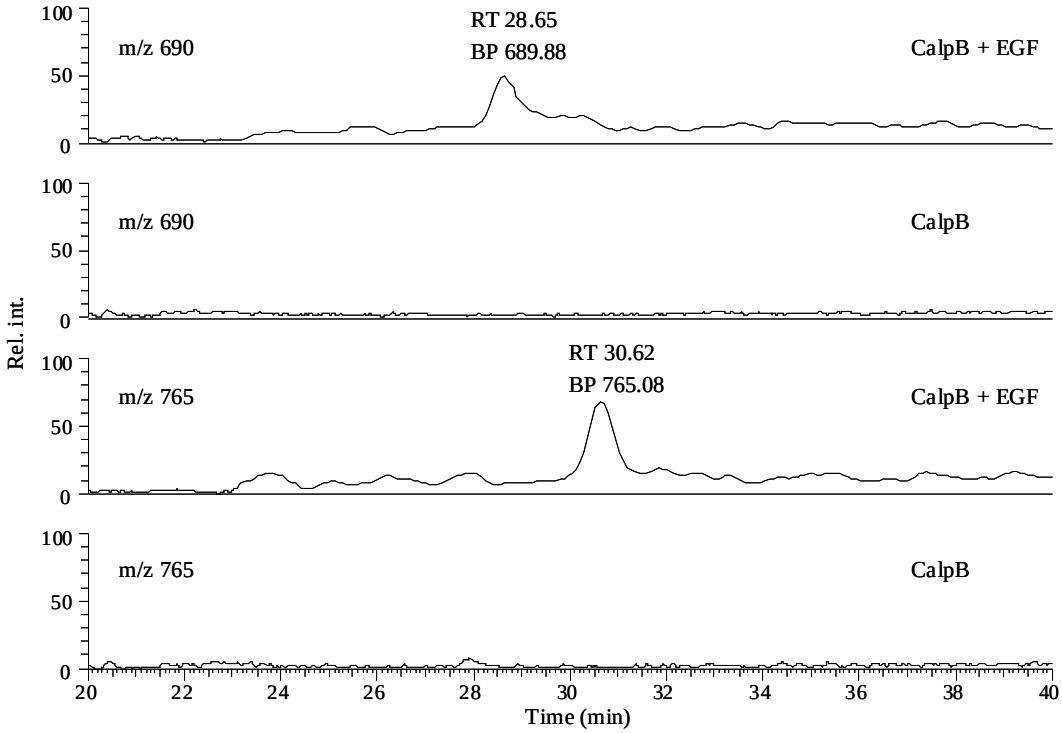


Figure 9. Extracted ion chromatograms at m/z values corresponding to calpain B phosphopeptides. Calpain B was isolated from S2 cells without (CalpB) or after EGF treatment (CalpB + EGF). The identity of the phophopeptides was confirmed by CID experiments. The retention time (RT, min) and the base peak m/z (BP) is given above the peaks.

Thus, *in vivo* modification of the ERK phosphorylation site at Thr-747 was proven upon EGF stimulation. No additional ERK phosphorylation sites were detected. Interestingly, Ser-240 was also phosphorylated in the same experiment. This site does not fit to the consensus sequence of the ERK kinases, but can be the target of PKA (as found *in vitro*) or the Ca^{2+}/calmodulin-dependent protein kinase II (by prediction). The EGF treatment possibly induced this phosphorylation event independently from the ERK activation.

4.2. Enrichment of O-GlcNAc modification

Analogous to phosphorylation the O-GlcNAc modification occurs at substoichiometric levels, therefore, enrichment is essential in the analysis of this PTM. None of the isolation methods developed for this modification display high sensitivity and selectivity, thus we sought to establish a more efficient enrichment strategy.

Periodate readily oxidizes exocyclic vicinal diols and cyclic *cis*-diol compounds to aldehydes. This reaction combined with hydrazide resin capture and release of the modified peptides by PNGaseF treatment is a widely used method for the analysis of proteins with N-linked complex carbohydrates due to the sialic acid and mannose content[70-72]. Cellulose[73] and cyclodextrins[74;75] – both containing *trans*-diol moiety – were also found to be oxidized by periodate, though at a much lower rate. This observation suggested that O-GlcNAc modified peptides might also be oxidized by periodate enabling the enrichment by the hydrazide resin capture.

4.2.1. Test reactions in solution

In order to test whether this approach was feasible, reactions for the oxidation, hydrazone formation and hydrazone cleavage were performed in solution using a commercially available O-GlcNAc modified peptide standard, TAPTgSTIAPG (m/z 1118.5).

As N-terminal serine and threonine residues – being vicinal amino alcohols – are also oxidized by periodate resulting in a glyoxylyl derivative with a mass shift of -31 Da and -45 Da, respectively[76], the initial oxidation product was α-N-glyoxylyl-APTgSTIAPG observed at m/z 1095.5 (Figure 10A). A second peak at m/z 1113.5 corresponds to the α-N-glyoxylyl hydrate. Unfortunately, the only basic residue in the O-GlcNAc peptide standard is destroyed by the periodate oxidation impairing ionization and detection. The oxidation product of the O-GlcNAc peptide standard and the respective derivatives were always observed as sodium/potassium adducts. At acidic and close neutral pH (pH 3-6.5) a partial oxidation of the carbohydrate ring (-2 Da) was observed within a few hours at room temperature. The ring opening was complete in an overnight reaction (16-20 h). At elevated temperatures the reaction was faster, at 37 °C the sugar ring was completely oxidized to the

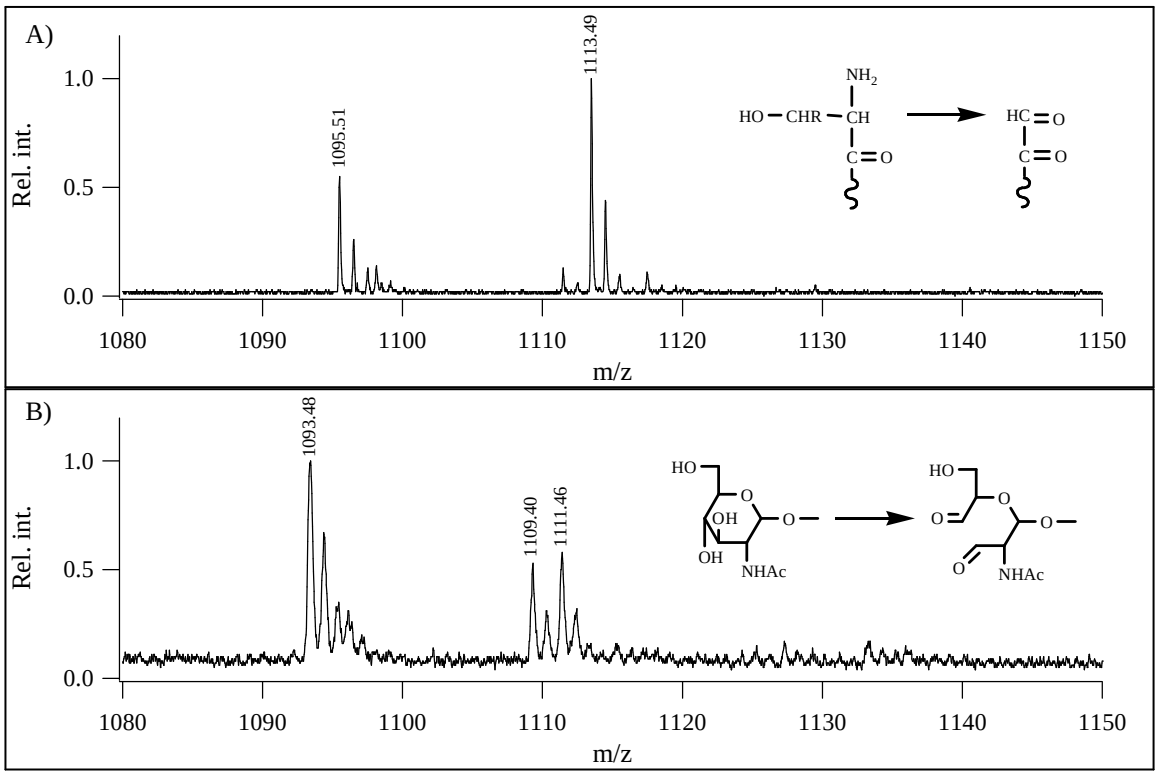


Figure 10. MALDI-TOF mass spectra of the periodate oxidation products of the O-GlcNAc modified peptide standard TAPTgSTIAPG. A) Initial oxidaton (-45 Da) occurs at the peptide N-terminus. B) The carbohydrate ring is split up (-2 Da) in a longer exposition at elevated temperatures.

dialdehyde derivative in 4-6 h (Figure 10B). At basic pH (pH~9) only the N-terminal oxidation was observed, ring opening did not occur even in a 24 h reaction.

Prior to the hydrazone formation, the oxidation reaction mixture can either be desalted[71] or sodium sulfite can be added to terminate the periodate oxidation[72]. Since in the latter approach the oxidation and the subsequent hydrazide capture are performed without a purification step in between, i.e. without sample losses, this route was preferred. Even if the sodium sulfite solution is prepared freshly, reaction with oxygen will consume a portion of the reductant. Therefore, 1 equivalent with an excess was required to terminate the periodate oxidation by reducing the periodate to iodate. However, in the following reaction step with benzoic acid hydrazide a side reaction of the N-terminal glyoxylyl group occurred. Instead of the N-terminal hydrazone formation, reduction of the glyoxylyl function to glycolyl (+2 Da) was observed. This side reaction can be attributed to the reducing property of hydrazides/hydrazines. The actual reducing agent is a diimide[77;78] which is formed by the

reaction of hydrazine/hydrazide with an oxidizing agent. In this case oxidizing iodo-compounds (IO_4^-, IO_3^-) act as co-oxidants to produce the diimide and thus facilitate the reduction of the N-terminal glyoxylyl group. Nevertheless, if periodate was reduced to iodide

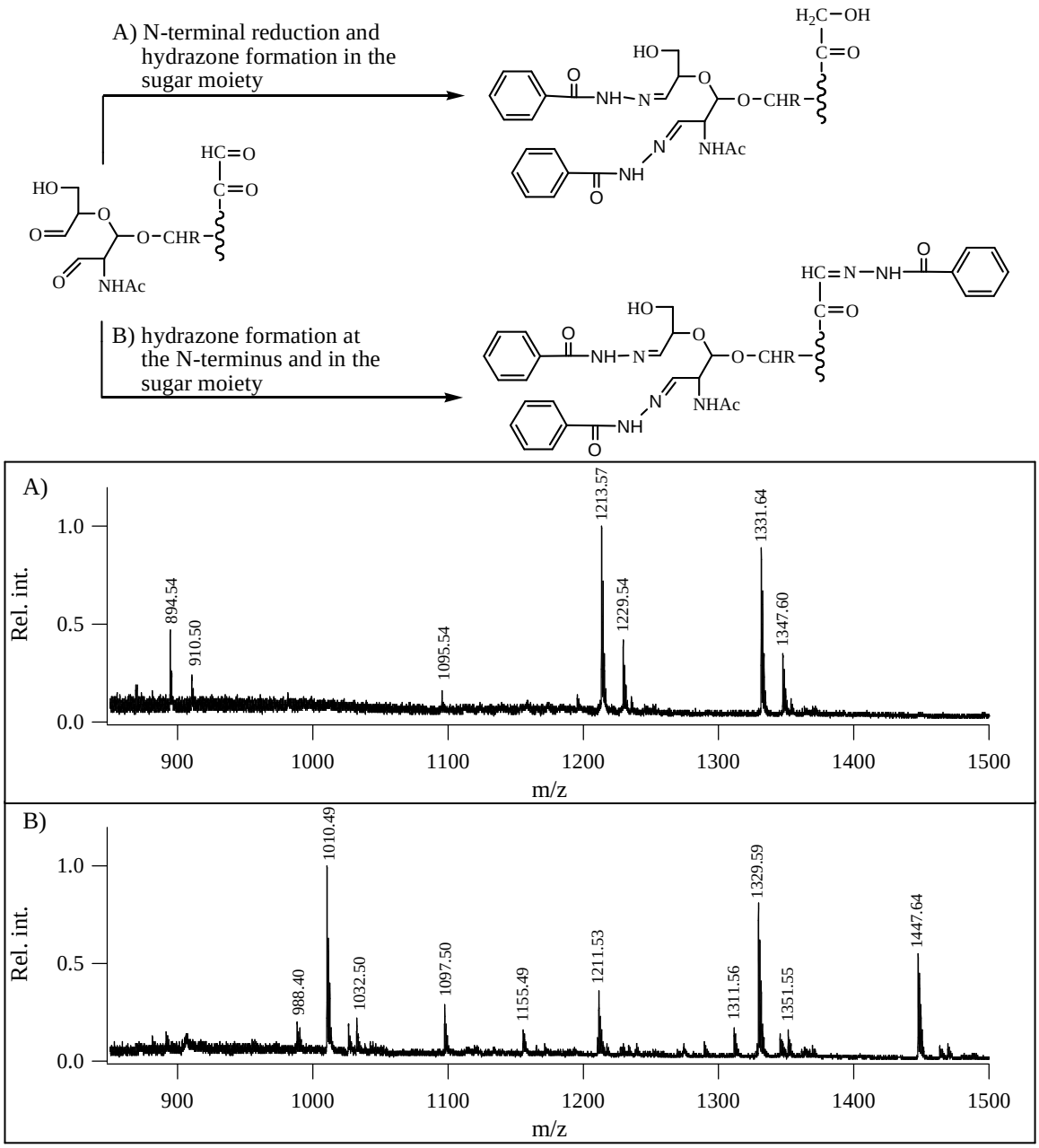


Figure 11. Hydrazone formation from oxidized O-GlcNAc modified peptide standard TAPTgSTIAPG with benzoic acid hydrazide. Prior oxidation with periodate was terminated by 2 equivalents of sodium sulfite (A) or 5 equivalents of sodium sulfite at pH 4-5 (B). Spectra were recorded on a MALDI-TOF mass spectrometer.

by 5 equivalents of sodium sulfite under acidic conditions (pH 4-5), no reduction occurred and the N-terminal glyoxylyl group was converted to hydrazone. This side reaction was not observed for the 'sugar-dialdehyde'.

The formation of the hydrazone was studied with oxidized O-GlcNAc peptide standard and benzoic acid hydrazide. As the addition of a nucleophile onto the carbonyl requires mildly acidic conditions, the reaction was carried out at pH 5-6. The products of the reaction depended on how the oxidation was terminated. If 1-4 equivalents of sodium sulfite were added to the periodate reaction prior to the hydrazone formation, the N-terminal glyoxylyl was reduced to glycolyl and the 'sugar-dialdehyde' converted to a semi-hydrazone (m/z 1213.6) or a double hydrazone (m/z 1331.6) as shown in Figure 11A. The peak at m/z 894.5 corresponds to the sugar loss. If the oxidation was terminated with 5 equivalents of sodium sulfite at pH 4-5, the N-terminal glyoxylyl as well as the aldehyde functions in the sugar moiety were converted to hydrazones (Figure 11B). The peak at m/z 1211.5 corresponds to the single, the peak at m/z 1329.6 to the double and the peak at m/z 1447.6 to the triple hydrazone. The peak at m/z 1010.5 is the deglycated N-terminal hydrazone derivative. The hydrazone formation was not complete even in an overnight reaction.

In the enrichment of N-glycoproteins, the release of the modified peptides is performed by PNGaseF treatment cleaving the peptide from the hydrazide captured sugar structure[70-72]. The site of the modification is recognized by the conversion of the respective asparagine to aspartic acid. O-glycosylated peptides could be released similarly by removing the sugar by β-elimination. It also would leave a telltale sign, a double bond on the previously modified amino acid, if the peptide survives the hydrolysis. That is especially problematic when threonine is modified, since this residue is much less prone to β-elimination than a glycosylated serine. Moreover, β-elimination of water from unmodified serine or threonine residues of nonspecifically bound peptides can lead to false positive results. Therefore, alternative reactions had to be considered for the release of the modified peptides (Figure 12).

Cleavage of the hydrazone bond releases the peptides with the sugar modification. As the hydrazones are acid labile[76;79], the hydrazone bond was split by 1-10% formic acid at room temperature. The reaction yielded the initial N-terminal glyoxylyl group and the 'sugar-dialdehyde'. However, cleavage of the hydrazone bond under acidic conditions can lead to the loss of the sugar modification resulting in decreased sensitivity. Another approach is to cleave

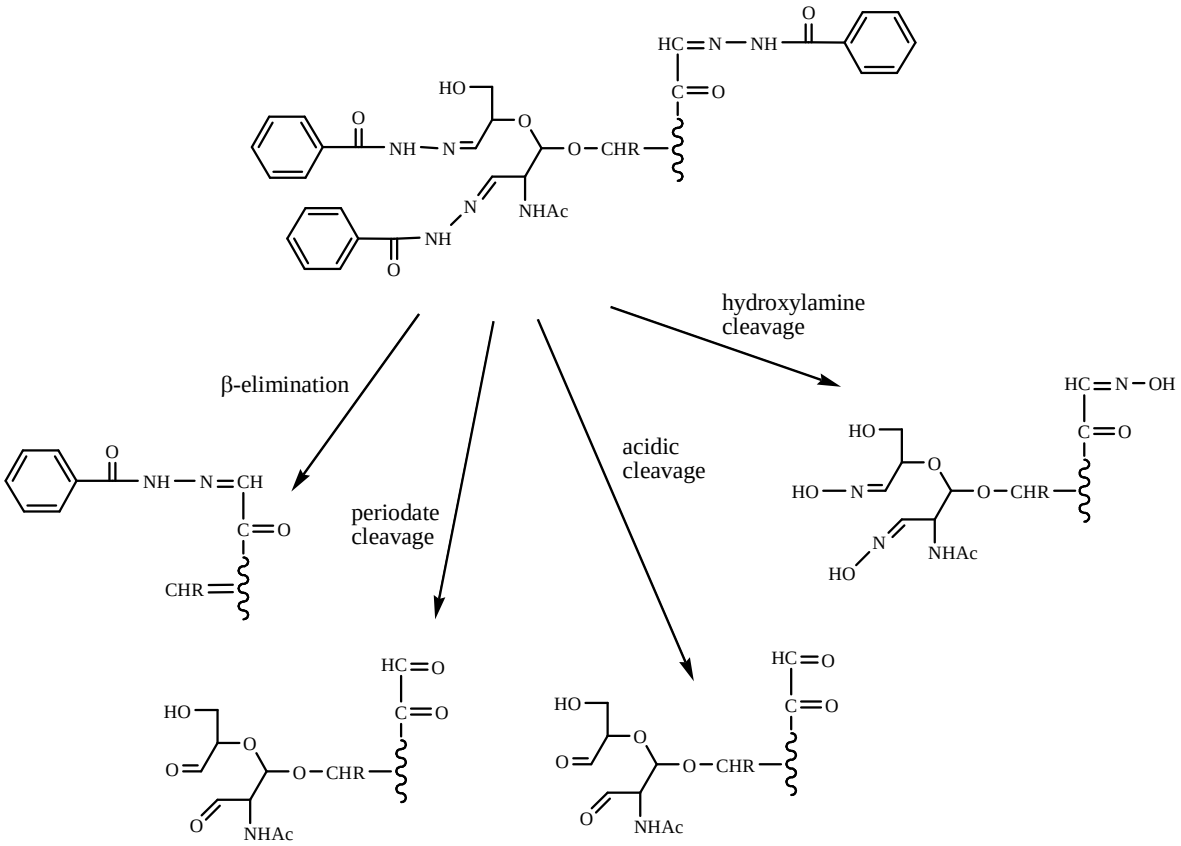


Figure 12. Cleavage of the hydrazone derivative by β-elimination, periodate, acidic or hydroxylamine treatment.

the hydrazone by periodate[80]. Again, the hydrazone was converted to the initial aldehyde, either the N-terminal glyoxylyl group or the 'sugar-dialdehyde'. Conversion of the hydrazone to oxime by hydroxylamine provides an additional option for the release of the modified peptides. Analogous to the hydrazone formation the reaction was performed under mildly acidic conditions (pH 5). An overnight reaction was required for a nearly complete conversion.

4.2.2. Solid-phase capture

As a next step, the in-solution reactions had to be adjusted to solid-phase capture. Since protein digestion most likely produces at least some peptides with N-terminal serine or threonine residues, and the oxidation and capture of these by the hydrazide resin prevents the selective isolation of the O-GlcNAc modification, enrichment at the protein level was

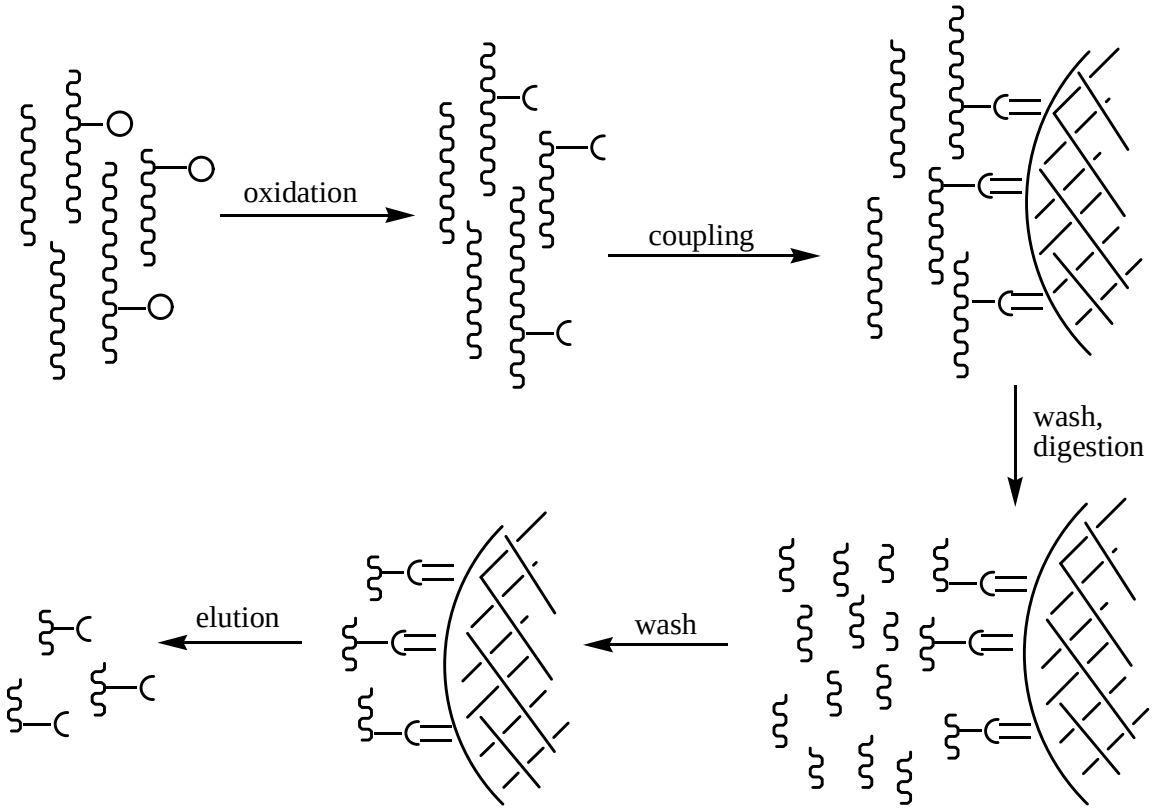


Figure 13. Reaction scheme for the solid-phase capture.

preferred. The reaction scheme is shown in Figure 13. In these experiments, α-crystallin was used as an O-GlcNAc modified protein standard. The commercially available α-crystallin (Appendix) contains both the A and B chains (P02470 and P02510, respectively). In chain A an O-GlcNAc modification was found at position Ser-162[81], whereas in chain B Thr-170 was described to be O-GlcNAc modified[82].

Ammonium salts and primary or secondary amines would interfere with the coupling reaction, therefore, these compounds were avoided in the reaction buffer. Otherwise, the type of the buffer – sodium acetate and triethylamine phosphate were used here – did not have influence on the enrichment. On the contrary, the type of the denaturing agent had a great impact on the performance. Moderate yields were achieved with 6 M guanidine hydrochloride while the use of SDS greatly enhanced the signal of the O-GlcNAc modified crystallin peptide in the eluate. This improvement might be attributed to electrostatic interactions with the solid support. The coupling reaction proceeds under mildly acidic conditions (pH 5-6). The hydrazide resin bears a net positive charge at this pH. Therefore, solubilization with the

negatively charged SDS brings the proteins in close proximity to the resin facilitating the reaction between the aldehyde groups of the oxidized sugar moiety and the hydrazide function of the solid support. Accordingly, nonspecific binding is also increased if SDS is used as the detergent, therefore, extensive wash procedure is necessary.

Two types of hydrazide resins were tested, agarose-bound hydrazide (Affi-Gel Hz) and hydrazide coupled to silica support (H-CPG). The Affi-Gel is widely used in the enrichment of N-linked glycopeptides[70;72]. The H-CPG resin has been applied to the selective isolation of oxidative stress related 4-hydroxynonenal modified proteins[79]. An advantage of the H-CPG resin is that it features a rigid support with swelling properties independent form the solvent used, whereas Affi-Gel displays different swelling in aqueous and organic solvents. Additionally, the acidic cleavage of the hydrazone bond worked only with the H-CPG resin, but not with Affi-Gel. Otherwise, no significant difference in the performance of the two resins in the O-GlcNAc enrichment was observed.

Similarly to the in-solution reactions, different strategies were followed for the release of the O-GlcNAc modified peptides. Though the hydrazone bond was found to be cleaved by periodate in solution, this approach did not prove to be feasible when applied to solid-phase bound crystallin. The β-elimination/Michael-addition strategy provided the O-GlcNAc modified α-crystallin A peptide AIPVgSREEKPSSAPSS - converted to the S-aminoethyl-cysteine analogue (m/z 1701.0) - only at moderate yields (Figure 14B). As threonine shows reduced reactivity in β-elimination, even lower efficiency is expected for threonine modified peptides. The acidic cleavage of the hydrazone bond was performed with formic acid (1-10%) or trifluoroacetic acid (0.1-1%) at increasing temperatures (25 °C, 37 °C and 60 °C). This method worked well with the H-CPG resin, but was not applicable to the Affi-Gel. The release of the oxidized O-GlcNAc modified peptide AIPVgSREEKPSSAPSS (m/z 1842.9) and its truncated forms AIPVgSREEKPS (m/z 1413.6) and AIPVgSREEKPSSAPS (m/z 1755.8) increased with lower pH and elevated temperatures, but unfortunately, the elimination of the sugar moiety (m/z 1641.8) also increased. In summary, 1% formic acid at 37 °C for 2 h or 10% formic acid at room temperature for 2 h gave highest yields of the oxidized O-GlcNAc modified peptides with relatively low sugar loss (Figure 14C). Among the different cleavage strategies, the overnight hydroxylamine treatment proved to be the most successful for the release of the oxidized O-GlcNAc modified peptide

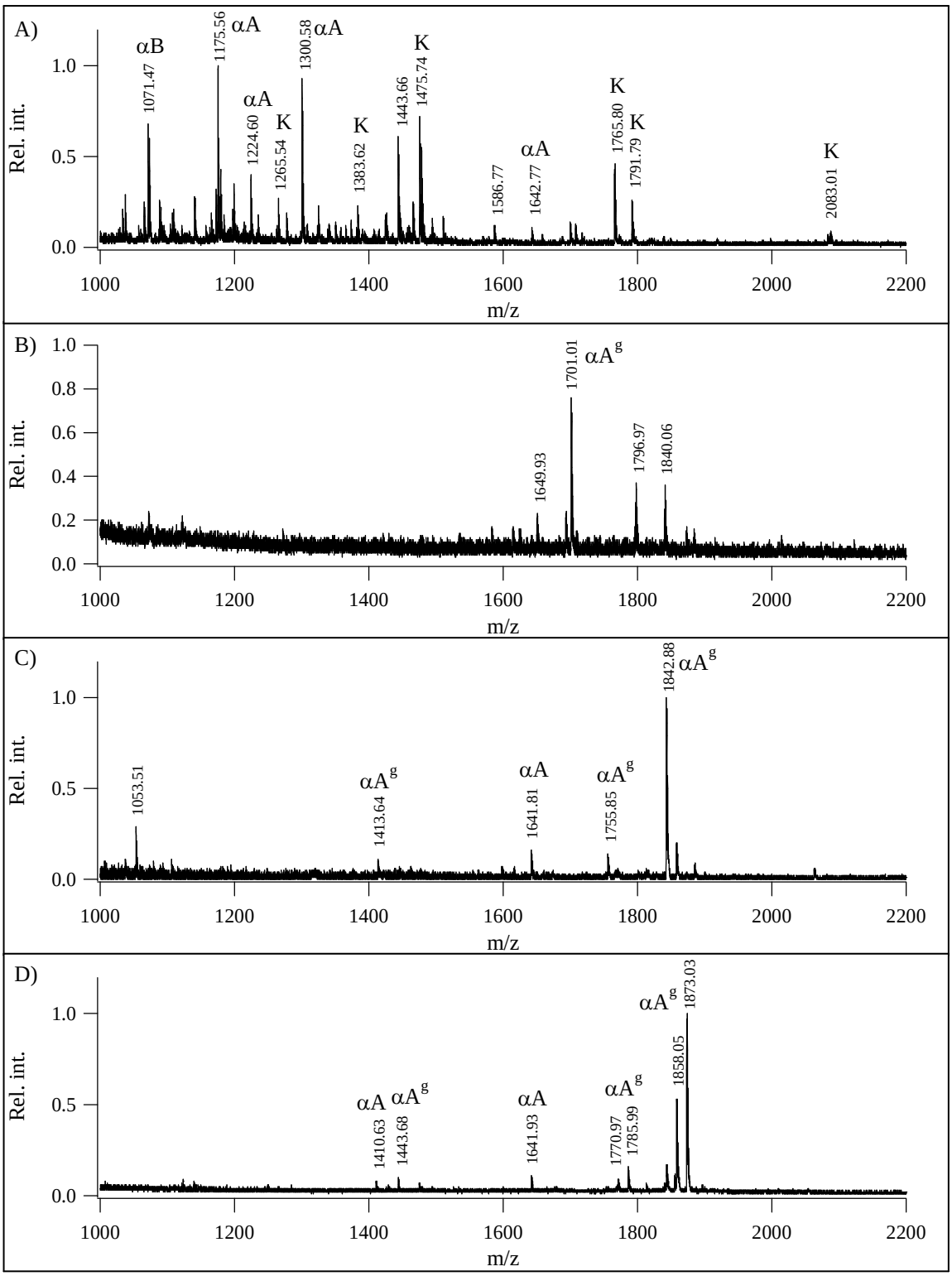


Figure 14. Enrichment of O-GlcNAc modified peptides from α-crystallin. Flow-through after the tryptic digestion (A) and elution by BEMAD (B), 1% formic acid (C) or hydroxylamine (D). K denotes keratin, αA and αB stands for α-crystallin A and B chain peptides, αA^g denotes the respective derivatives of the O-GlcNAc modified α-crystallin peptides. Spectra were recorded on a MALDI-TOF mass spectrometer.

AIPVgSREEKPSSAPSS as an oxime observed at m/z 1873.0 (Figure 14D). The signals at m/z 1443.7 and 1786.0 correspond to the truncated forms AIPVgSREEKPS and AIPVgSREEKPSSAPS, respectively. The α-crystallin B glycopeptide EEKPAVgTAAPK was observed only in the LC-MS/MS analysis of the hydroxylamine cleaved sample at m/z 686.38 (2+).

As mentioned previously, the use of SDS necessitates an extensive wash procedure. In order to optimize the wash steps, α-crystallin was mixed with BSA in 1:20 w/w ratio. Different buffers, pH, detergents and organic solvents were used at the protein level as well as after tryptic digestion at the peptide level. In all cases, BSA peptides dominated in the supernatant after the tryptic digestion, suggesting that neither of the different wash conditions could efficiently break the strong adhesion between the SDS wrapped proteins and the hydrazide resin. On the contrary, great differences were found in the wash efficiency at the peptide level, though none of the wash conditions alone was sufficient for the removal of the nonspecific binding. In the enrichment of N-glycopeptides, a high salt wash (1.5 M sodium chloride) is combined with an acidic (0.1% trifluoroacetic acid in 80% acetonitrile) and an organic wash (methanol)[70]. Improved wash efficiency was found when a mildly basic phosphate buffer (0.5 M triethylamine phosphate, pH 8.5 in 30% acetonitrile) was used instead of the high salt wash and methanol was replaced with isopropanol. Additionally,

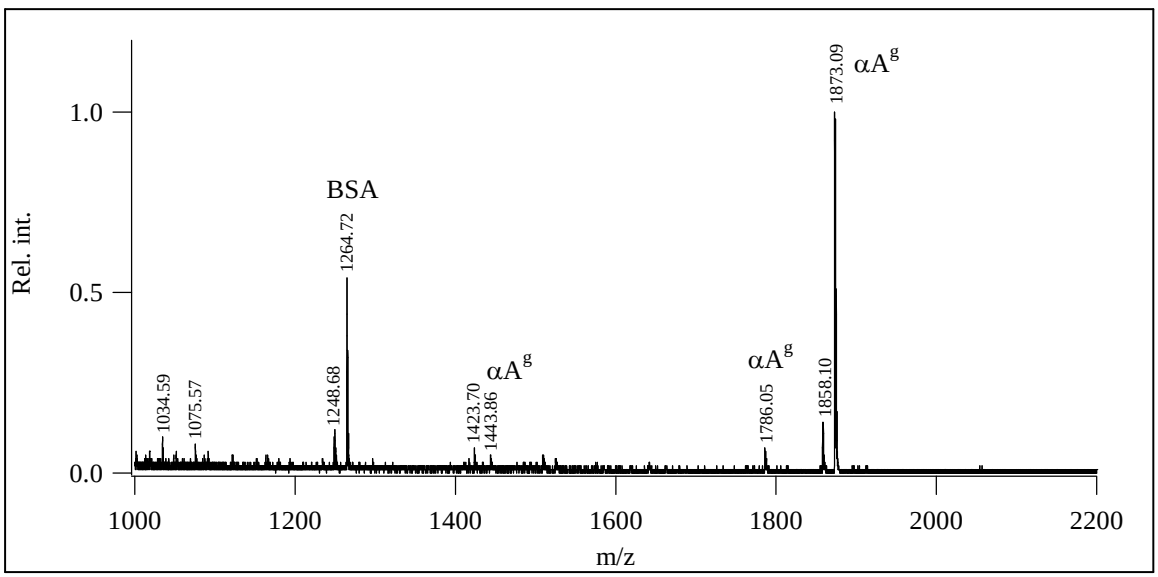


Figure 15. Enrichment of O-GlcNAc modified α-crystallin peptides from a mixture of crystallin:BSA 1:20 (w/w). αA^g denotes the oxime derivatives of the O-GlcNAc modified α-crystallin peptides. The spectrum was acquired on a MALDI-TOF mass spectrometer.

0.5 M sodium-phthalate, pH 7 in 30% acetonitrile further reduced the nonspecific binding (Figure 15). The oxime derivative of the O-GlcNAc modified α-crystallin A peptide is observed as the most abundant peak at m/z 1873.1. The signals detected at m/z 1248.7 and 1264.7 correspond to the tryptic BSA peptide EK*VLTSSAR with a yet unknown modification at the lysine residue that was converted to an aldehyde upon periodate treatment and also captured by the hydrazide resin (CID data, not shown).

4.2.3. Fragmentation of the oxime derivatives of O-GlcNAc modified peptides

O-GlcNAc modified peptides show characteristic fragmentation as compared to unmodified peptides. Since the glycosidic bond is more easily cleaved than the amide bond, extensive 203 Da loss is observed in the MS/MS spectra. Moreover, due to a gas-phase rearrangement the sugar is eliminated without “marking” the originally modified residue in the process. The oxime derivatives of the oxidized O-GlcNAc modified peptides show slightly different fragmentation. Since the carbohydrate ring is split open in the periodate oxidation, cleavage of the C-O bond on either side of the carbonyl C results in two characteristic losses as demonstrated in Figure 16.

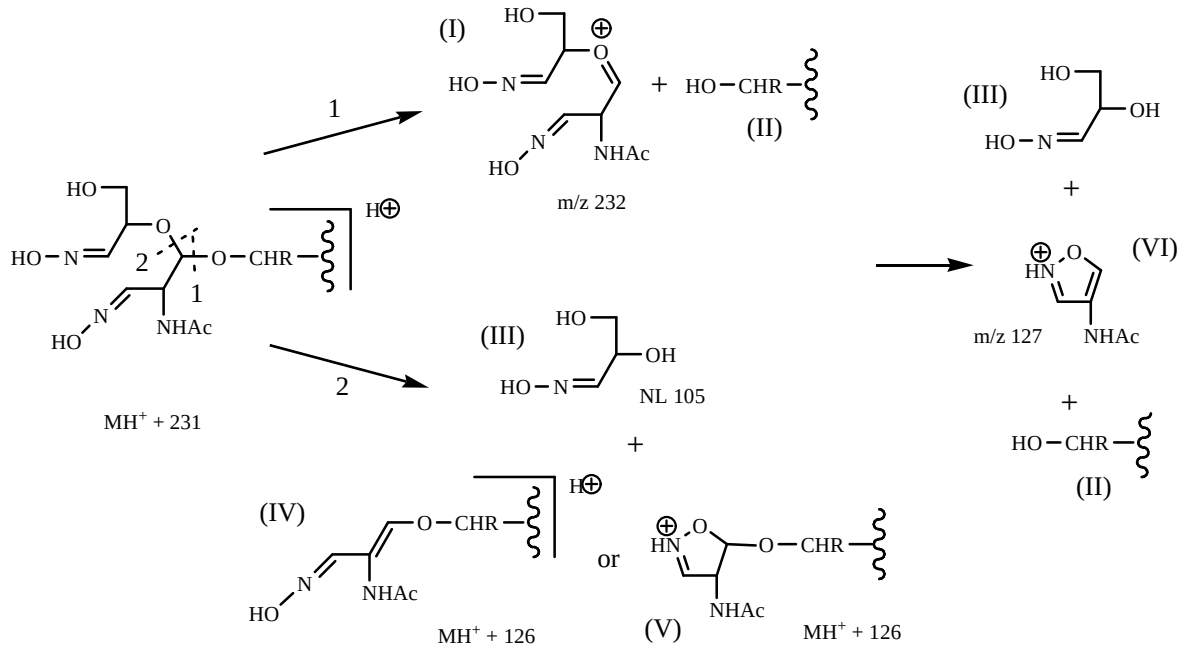


Figure 16. Fragmentation of the oxime derivatives of oxidized O-GlcNAc modified peptides.

In pathway 1, the sugar is eliminated in the same manner, as from O-glycosylated peptides in general, i.e. without any indication where the modification was originally located. Depending on where the charge is retained, an oxonium ion (I) is observed at m/z 232 and the 'unmodified' peptide is detected of the same charge as (constant neutral loss) or one charge less than (constant charged loss, II) the precursor. If the adjacent acetal bond is cleaved (pathway 2), only the half of the carbohydrate derivative is released as a neutral. The result is a neutral loss of 105 Da (III) and a mass signature of 126 Da on the modified amino acid. We assume a conjugated double bond (IV) or a cyclic structure (V) in this 126 Da mass modification. Further fragmentation of the oxonium ion (I) or the truncated derivative (IV or V) leads to a product ion of m/z 127, we hypothesize an isoxasolium structure (VI) for this product.

Distinct fragmentation pattern is observed in ion traps and Q-TOF instruments as illustrated by the CID spectra of the O-GlcNAc modified crystallin peptide AIPVgSREEKPSSAPSS (Figure 17). In ion traps, due to the single activation of the precursor

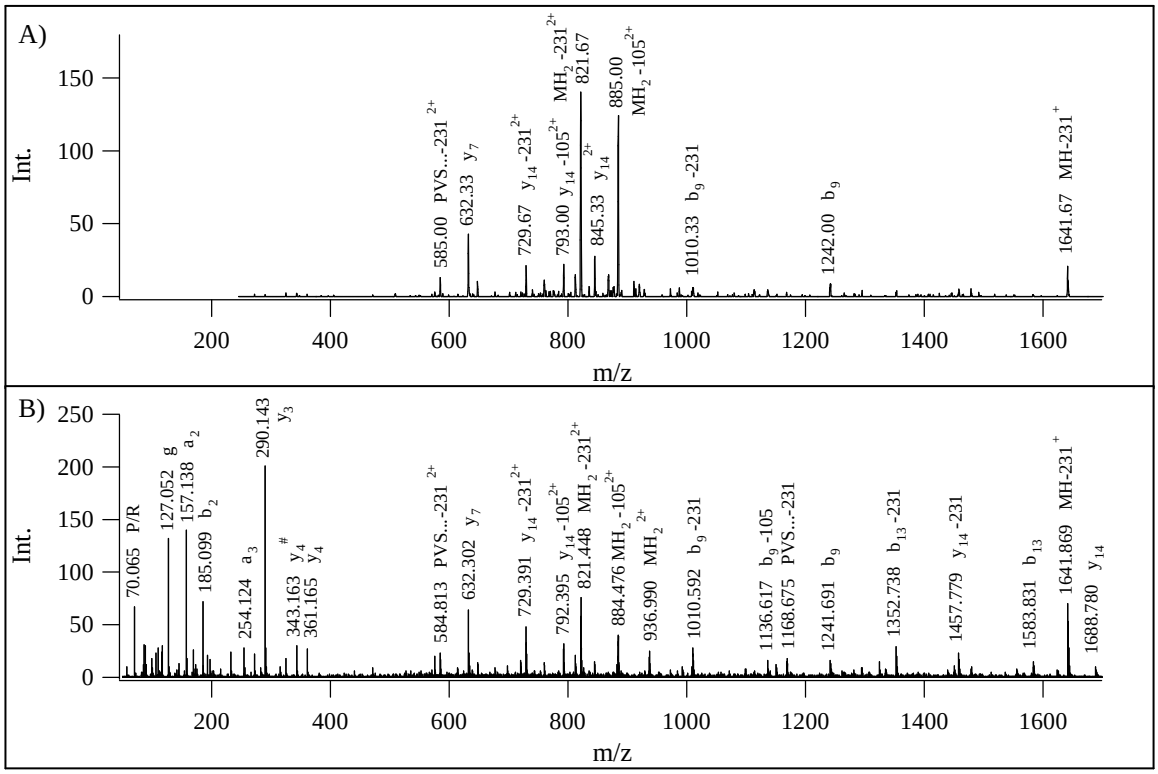


Figure 17. CID spectrum of the oxime derivative of the O-GlcNAc modified α-crystallin AIPVgSREEKPSSAPSṢ peptide (m/z 937.0) recorded in an ion trap (A) or a Q-TOF (B) mass spectrometer.

ion at its secular frequency, mainly the cleavage of the sugar derivative is observed as a neutral loss of 231 or 105 Da, or as a charged loss of 232^+ along with the oxonium ion. As the amide bond N-terminally to proline is also relatively facile, the few sequence ions detected (b_9, y_7 and y_{14}) are products of these cleavages. The poor fragmentation of glycopeptides in ion trap instruments might be compensated by MS(3) experiments or MS(2) with multistage activation (MSA). In MS(3) experiments the product ion corresponding to the neutral loss of 231 Da in the MS(2) spectrum is selected for further fragmentation. In MSA experiments the two fragmentation steps are merged into a single step, the precursor and its constant neutral loss product are consecutively activated in the same fragmentation cycle.

In Q-TOF instruments multiple collisions occur in the collision cell. As a consequence, peptide fragmentation is also observed. The fragments might retain the modification or lose it partially (-105 Da) or entirely (-231 Da). Moreover, instead of the oxonium ion at m/z 232, a fragment of the carbohydrate derivative is observed at m/z 127.

To sum up, a chemical derivatization method has been developed for the enrichment of O-GlcNAc modified proteins. The procedure involves periodate oxidation followed by hydrazide resin capture, on-resin digestion and cleavage of the O-GlcNAc modified peptides by hydroxylamine treatment. Further development is underway in order to reduce the nonglycosylated background and increase the efficiency of this isolation method.

5. Summary

Our results regarding two post-translational modifications - phosphorylation and O-GlcNAc glycosylation - are discussed in this thesis.

In the *phosphorylation* studies, the efficiency and selectivity of different complexing metal ions was investigated in IMAC-based phosphopeptide enrichment. Binding and wash conditions with respect to pH and solvent composition were optimized for the different metal ions. Fe^{3+} is well known to have high binding affinity for the phosphate group but also binds carboxylic groups impairing selective enrichment. Al^{3+} proved to be a real alternative possessing high binding affinity for phosphopeptides with significantly lower nonspecific binding. Other metal ions were less efficient in phosphopeptide enrichment.

IMAC as well as titanium dioxide enrichment was used to determine *in vitro* and *in vivo* phosphorylation in three different biological processes. 1) In the brain specific TPPP/p25, *in vitro* phosphorylation of the recombinant human protein by ERK2 and CDK5 proceeded at sites Ser-18 and Ser-160, by CDK5 also at Thr-14, whereas PKA phosphorylated Ser-32, Thr-92 and Ser-159. *In vivo* phosphorylation sites in the bovine protein were identified at Thr-12, Ser-16 and Ser-30, in accordance with the phosphorylation sites found in human and mouse TPPP/p25 proteins. 2) Activation of the recombinant *Medicago* RRK1 kinase by the ROP6 GTPase resulted in the autophosphorylation of the kinase at Thr-6 and Ser-8. Phosphorylation of the activator ROP6 GTPase turned out to be artificially induced at multiple sites in the N-terminal His-tag sequence. 3) *In vitro* phosphorylation of the *Drosophila* calpain B by PKA occurred at Ser-240 and Ser-845. The ERK1 and ERK2 kinases phosphorylated Thr-747 *in vitro*. This phosphorylation site as well as phosphorylation at Ser-240 was confirmed *in vivo* in EGF treated cells.

The second part of the thesis deals with *O-GlcNAc glycosylation*. A method has been developed for the enrichment of O-GlcNAc modified proteins. The strategy is analogous to the isolation of N-glycosylated proteins by the hydrazide resin capture approach, however, the oxidation reaction as well as the release of the glycopeptides had to be modified due to the differences in the two types of carbohydrate modification. The reaction conditions were first optimized in solution with an O-GlcNAc modified peptide standard, then the procedure was adjusted to solid-phase capture using α-crystallin as an O-GlcNAc modified protein standard. The efficiency of the method was demonstrated by the enrichment of the O-GlcNAc modified α-crystallin peptides from α-crystallin:BSA mixtures.

6. Acknowledgements

I wish to thank to Prof. Botond Penke and Dr. Katalin F. Medzihradszky for giving me the opportunity to join their groups. Prof. Botond Penke always supported me during and after my Ph.D. studies. Dr. Katalin F. Medzihradszky introduced me to protein mass spectrometry and supervised my scientific progress.

I am grateful to the collegues at the Biological Research Center, Dr. Zsuzsanna Darula, Dr. Éva Hunyadi-Gulyás, Emília Szájli and Gyula Orosz for their help and useful discussions.

Thanks are due to the collegues at the Department of Medical Chemistry, University of Szeged, Dr. Zoltán Szabó, Dr. Zoltán Kele, Dr. Györgyi Ferenc, Dr. Zoltán Kupihár, Dr. Lajos Kovács and Attila Bokros for technical help and ideas.

I should thank the collaborators, Dr. Judit Ovádi, Dr. Ferenc Orosz, Emma Hlavanda, Dr. Attila Fehér, Dr. Dulguun Dorjgotov, Dr. Viktor Dombrádi, László Kovács, Anita Alexa, Ágnes Tantos.

I would like to acknowledge the Hungarian Academy of Sciences for financial support.

Kromat Ltd. is acknowledged for providing the use of an Agilent 1100 nanoLC-XCT Plus IonTrap system.

References

1. Garavelli, J. S. *Proteomics* 2004, *4*, 1527-33.
2. Karas, M.; Hillenkamp, F. *Analytical Chemistry* 1988, *60*, 2299-301.
3. Fenn, J. B.; Mann, M.; Meng, C. K.; Wong, S. F.; Whitehouse, C. M. *Science* 1989, *246*, 64-71.
4. Wilm, M. S.; Mann, M. *International Journal of Mass Spectrometry and Ion Processes* 1994, *136*, 167-80.
5. de Hoffmann, E.; Stroobant, V. *Mass Spectrometry Principles and Applications*, 2nd edition ed.; Wiley: Chichester, 2002.
6. Scigelova, M.; Makarov, A. *Proteomics* 2006, 16-21.
7. Moniatte, M.; vanderGoot, F. G.; Buckley, J. T.; Pattus, F.; VanDorsselaer, A. *Febs Letters* 1996, *384*, 269-72.
8. Mamyrin, B. A.; Karataev, V. I.; Shmikk, D. V.; Zagulin, V. A *Soviet Physics - JETP* 1973, *37*, 45.
9. Brown, R. S.; Lennon, J. J. *Analytical Chemistry* 1995, *67*, 1998-2003.
10. Biemann, K. *Methods in Enzymology* 1990, *193*, 886-87.
11. Paizs, B.; Suhai, S. *Mass Spectrometry Reviews* 2005, *24*, 508-48.
12. Yalcin, T.; Khouw, C.; Csizmadia, I. G.; Peterson, M. R.; Harrison, A. G. *Journal of the American Society for Mass Spectrometry* 1995, *6*, 1165-74.
13. Yu, W.; Vath, J. E.; Huberty, M. C.; Martin, S. A. *Analytical Chemistry* 1993, *65*, 3015-23.
14. Falick, A. M.; Hines, W. M.; Medzihradszky, K. F.; Baldwin, M. A.; Gibson, B. W. *Journal of the American Society for Mass Spectrometry* 1993, *4*, 882-93.
15. Fischle, W.; Wang, Y. M.; Allis, C. D. *Nature* 2003, *425*, 475-79.
16. Wang, Z.; Gucek, M.; Hart, G. W. *Proceedings of the National Academy of Sciences of the United States of America* 2008, *105*, 13793-98.
17. Wang, Z.; Pandey, A.; Hart, G. W. *Molecular & Cellular Proteomics* 2007, *6*, 1365-79.
18. Tallent, M. K.; Varghis, N.; Skorobogatko, Y.; Hernandez-Cuebas, L.; Whelan, K.; Vocadlo, D. J.; Vosseller, K. *Journal of Biological Chemistry* 2009, *284*, 174-81.

19. Posewitz, M. C.; Tempst, P. *Analytical Chemistry* 1999, *71*, 2883-92.
20. Thompson, A. J.; Hart, S. R.; Franz, C.; Barnouin, K.; Ridley, A.; Cramer, R. *Analytical Chemistry* 2003, *75*, 3232-43.
21. Ficarro, S. B.; McCleland, M. L.; Stukenberg, P. T.; Burke, D. J.; Ross, M. M.; Shabanowitz, J.; Hunt, D. F.; White, F. M. *Nature Biotechnology* 2002, *20*, 301-05.
22. Larsen, M. R.; Thingholm, T. E.; Jensen, O. N.; Roepstorff, P.; Jorgensen, T. J. D. *Molecular & Cellular Proteomics* 2005, *4*, 873-86.
23. Bodenmiller, B.; Mueller, L. N.; Mueller, M.; Domon, B.; Aebersold, R. *Nature Methods* 2007, *4*, 231-37.
24. Zhou, H. L.; Watts, J. D.; Aebersold, R. *Nature Biotechnology* 2001, *19*, 375-78.
25. Thaler, F.; Valsasina, B.; Baldi, R.; Jin, X.; Stewart, A.; Isacchi, A.; Kalisz, H. M.; Rusconi, L. *Analytical and Bioanalytical Chemistry* 2003, *376*, 366-73.
26. Oda, Y.; Nagasu, T.; Chait, B. T. *Nature Biotechnology* 2001, *19*, 379-82.
27. Byford, M. F. *Biochemical Journal* 1991, *280*, 261-65.
28. Beausoleil, S. A.; Jedrychowski, M.; Schwartz, D.; Elias, J. E.; Villen, J.; Li, J. X.; Cohn, M. A.; Cantley, L. C.; Gygi, S. P. *Proceedings of the National Academy of Sciences of the United States of America* 2004, *101*, 12130-35.
29. Alpert, A. J. *Analytical Chemistry* 2008, *80*, 62-76.
30. Gan, C. S.; Guo, T. N.; Zhang, H. M.; Lim, S. K.; Sze, S. K. *Journal of Proteome Research* 2008, *7*, 4869-77.
31. Chang, E. J.; Archambault, V.; McLachlin, D. T.; Krutchinsky, A. N.; Chait, B. T. *Analytical Chemistry* 2004, *76*, 4472-83.
32. Imanishi, S. Y.; Kochin, V.; Ferraris, S. E.; de Thonel, A.; Pallari, H. M.; Corthals, G. L.; Eriksson, J. E. *Molecular & Cellular Proteomics* 2007, *6*, 1380-91.
33. Medzihradszky, K. F.; Darula, Z.; Perlson, E.; Fainzilber, M.; Chalkley, R. J.; Ball, H.; Greenbaum, D.; Bogyo, M.; Tyson, D. R.; Bradshaw, R. A.; Burlingame, A. L. *Molecular & Cellular Proteomics* 2004, *3*, 429-40.
34. Covey, T. R.; Sushan, B. I.; Bonner, R.; Shroder, W.; Hucho, F. *Methods in Protein Sequence Analysis*, Jörnvall, H.; Höög, J. O.Gustavsson, A. M., Eds.; Birkhauser Verlag: Basel, 1991.

35. Annan, R. S.; Huddleston, M. J.; Verma, R.; Deshaies, R. J.; Carr, S. A. *Analytical Chemistry* 2001, *73*, 393-404.
36. Neubauer, G.; Mann, M. *Analytical Chemistry* 1999, *71*, 235-42.
37. Wells, L.; Vosseller, K.; Hart, G. W. *Science* 2001, *291*, 2376-78.
38. Comer, F. I.; Hart, G. W. *Journal of Biological Chemistry* 2000, *275*, 29179-82.
39. Vosseller, K.; Wells, L.; Lane, M. D.; Hart, G. W. *Proceedings of the National Academy of Sciences of the United States of America* 2002, *99*, 5313-18.
40. Roquemore, E. P.; Chou, T. Y.; Hart, G. W. *Guide to Techniques in Glycobiology* 1994, *230*, 443-60.
41. Hart, G. W. *Annual Review of Biochemistry* 1997, *66*, 315-35.
42. Zhang, F. X.; Su, K. H.; Yang, X. Y.; Bowe, D. B.; Paterson, A. J.; Kudlow, J. E. *Cell* 2003, *115*, 715-25.
43. Sumegi, M.; Hunyadi-Gulyas, E.; Medzihradszky, K. F.; Udvardy, A. *Biochemical and Biophysical Research Communications* 2003, *312*, 1284-89.
44. Jackson, S. P.; Tjian, R. *Cell* 1988, *55*, 125-33.
45. Lefebvre, T.; Planque, N.; Leleu, D.; Bailly, M.; Caillet-Boudin, M. L.; Saule, S.; Michalski, J. C. *Journal of Cellular Biochemistry* 2002, *85*, 208-18.
46. Vosseller, K.; Sakabe, K.; Wells, L.; Hart, G. W. *Current Opinion in Chemical Biology* 2002, *6*, 851-57.
47. Cole, R. N.; Hart, G. W. *Journal of Neurochemistry* 2001, *79*, 1080-89.
48. Khidekel, N.; Ficarro, S. B.; Peters, E. C.; Hsieh-Wilson, L. C. *Proceedings of the National Academy of Sciences of the United States of America* 2004, *101*, 13132-37.
49. Vosseller, K.; Trinidad, J. C.; Chalkley, R. J.; Specht, C. G.; Thalhammer, A.; Lynn, A. J.; Snedecor, J. O.; Guan, S. H.; Medzihradszky, K. F.; Maltby, D. A.; Schoepfer, R.; Burlingame, A. L. *Molecular & Cellular Proteomics* 2006, *5*, 923-34.
50. Griffith, L. S.; Mathes, M.; Schmitz, B. *Journal of Neuroscience Research* 1995, *41*, 270-78.
51. Li, X.; Lu, F.; Wang, J. Z.; Gong, C. X. *European Journal of Neuroscience* 2006, *23*, 2078-86.
52. Wells, L.; Vosseller, K.; Cole, R. N.; Cronshaw, J. M.; Matunis, M. J.; Hart, G. W. *Molecular & Cellular Proteomics* 2002, *1*, 791-804.

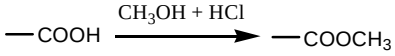
53. Khidekel, N.; Arndt, S.; Lamarre-Vincent, N.; Lippert, A.; Poulin-Kerstien, K. G.; Ramakrishnan, B.; Qasba, P. K.; Hsieh-Wilson, L. C. *Journal of the American Chemical Society* 2003, *125*, 16162-63.
54. Chalkley, R. J.; Thalhammer, A.; Schoepfer, R.; Burlingame, A. L. *Proc.Natl.Acad.Sci.U.S.A* 2009, *106*, 8894-99.
55. Medzihradszky, K. F.; Gillececastro, B. L.; Settineri, C. A.; Townsend, R. R.; Masiarz, F. R.; Burlingame, A. L. *Biomedical and Environmental Mass Spectrometry* 1990, *19*, 777-81.
56. Greis, K. D.; Hayes, B. K.; Comer, F. I.; Kirk, M.; Barnes, S.; Lowary, T. L.; Hart, G. W. *Analytical Biochemistry* 1996, *234*, 38-49.
57. Chalkley, R. J.; Burlingame, A. L. *Journal of the American Society for Mass Spectrometry* 2001, *12*, 1106-13.
58. Rademaker, G. J.; Pergantis, S. A.; Blok-Tip, L.; Langridge, J. I.; Kleen, A.; Thomas-Oates, J. E. *Analytical Biochemistry* 1998, *257*, 149-60.
59. Kinoshita, E.; Yamada, A.; Takeda, H.; Kinoshita-Kikuta, E.; Koike, T. *Journal of Separation Science* 2005, *28*, 155-62.
60. Feng, S.; Ye, M. L.; Zhou, H. J.; Jiang, X. G.; Jiang, X. N.; Zou, H. F.; Gong, B. L. *Molecular & Cellular Proteomics* 2007, *6*, 1656-65.
61. Wolschin, F.; Wienkoop, S.; Weckwerth, W. *Proteomics* 2005, *5*, 4389-97.
62. Barnouin, K. N.; Hart, S. R.; Thompson, A. J.; Okuyama, M.; Waterfield, M.; Cramer, R. *Proteomics* 2005, *5*, 4376-88.
63. Nuhse, T. S.; Stensballe, A.; Jensen, O. N.; Peck, S. C. *Molecular & Cellular Proteomics* 2003, *2*, 1234-43.
64. Liu, H. Z.; Stupak, J.; Zheng, J.; Keller, B. O.; Brix, B. J.; Fliegel, L.; Li, L. *Analytical Chemistry* 2004, *76*, 4223-32.
65. Hlavanda, E.; Kovacs, J.; Olah, J.; Orosz, F.; Medzihradszky, K. F.; Ovadi, J. *Biochemistry* 2002, *41*, 8657-64.
66. Trinidad, J. C.; Thalhammer, A.; Specht, C. G.; Lynn, A. J.; Baker, P. R.; Schoepfer, R.; Burlingame, A. L. *Molecular & Cellular Proteomics* 2008, *7*, 684-96.
67. DeGiorgis, J. A.; Jaffe, H.; Moreira, J. E.; Carlotti, C. G.; Leite, J. P.; Pant, H. C.; Dosemeci, A. *Journal of Proteome Research* 2005, *4*, 306-15.

68. Collins, M. O.; Yu, L.; Coba, M. P.; Husi, H.; Campuzano, L.; Blackstock, W. P.; Choudhary, J. S.; Grant, S. G. N. *Journal of Biological Chemistry* 2005, *280*, 5972-82.
69. Goll, D. E.; Thompson, V. F.; Li, H. Q.; Wei, W.; Cong, J. Y. *Physiological Reviews* 2003, *83*, 731-801.
70. Zhang, H.; Li, X. J.; Martin, D. B.; Aebersold, R. *Nature Biotechnology* 2003, *21*, 660-66.
71. Pan, S.; Wang, Y.; Quinn, J. F.; Peskind, E. R.; Waichunas, D.; Wimberger, J. T.; Jin, J. H.; Li, J. G.; Zhu, D.; Pan, C.; Zhang, J. *Journal of Proteome Research* 2006, *5*, 2769-79.
72. Sun, B. Y.; Ranish, J. A.; Utleg, A. G.; White, J. T.; Yan, X. W.; Lin, B. Y.; Hood, L. *Molecular & Cellular Proteomics* 2007, *6*, 141-49.
73. Hou, Q. X.; Liu, W.; Liu, Z. H.; Bai, L. L. *Industrial & Engineering Chemistry Research* 2007, *46*, 7830-37.
74. Kobayashi, M.; Urayama, T.; Suzawa, I.; Takagi, S.; Matsuda, K.; Ichishima, E. *Agricultural and Biological Chemistry* 1988, *52*, 2695-702.
75. Pumera, M.; Jelinek, I.; Jindrich, J.; Coufal, P.; Horsky, J. *Journal of Chromatography A* 2000, *891*, 201-06.
76. Geoghegan, K. F.; Stroh, J. G. *Bioconjugate Chemistry* 1992, *3*, 138-46.
77. Corey, E. J.; Mock, W. L.; Pasto, D. J. *Tetrahedron Letters* 1961, *2*, 347-52.
78. Smit, C.; Fraaije, M. W.; Minnaard, A. J. *Journal of Organic Chemistry* 2008, *73*, 9482-85.
79. Roe, M. R.; Xie, H. W.; Bandhakavi, S.; Griffin, T. J. *Analytical Chemistry* 2007, *79*, 3747-56.
80. Enders, D.; Wortmann, L.; Peters, R. *Accounts of Chemical Research* 2000, *33*, 157-69.
81. Roquemore, E. P.; Dell, A.; Morris, H. R.; Panico, M.; Reason, A. J.; Savoy, L. A.; Wistow, G. J.; Zigler, J. S.; Earles, B. J.; Hart, G. W. *Journal of Biological Chemistry* 1992, *267*, 555-63.
82. Roquemore, E. P.; Chevrier, M. R.; Cotter, R. J.; Hart, G. W. *Biochemistry* 1996, *35*, 3578-86.

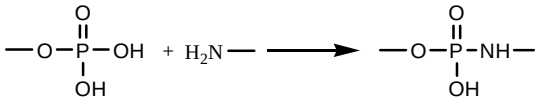
Appendix

Reaction schemes

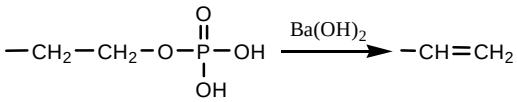
1. methyl esterification



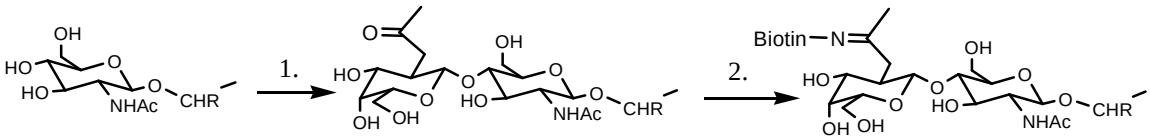
2. phosphoramidate formation



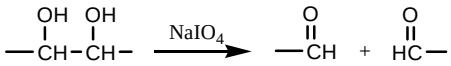
3. β-elimination



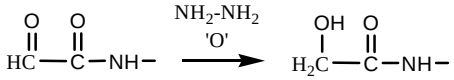
4. chemoenzymatic labeling



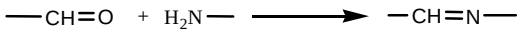
5. periodate oxidation



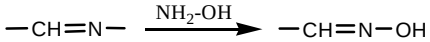
6. reduction of glyoxylyl groups



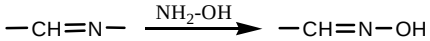
7. hydrazone formation



8. hydrazone cleavage by periodate or acids



9. hydrazone cleavage by hydroxylamine



Sequence and theoretical tryptic peptide masses of fetuin (P12763) considering cysteine pyridylethylation as fixed modification and pyroglutamic acid formation from N-terminal glutamine residues (denoted as q) as variable modification. One missed cleavage was allowed.

```
MKSFVLLFCL AQLWGCHSIP LDPVAGYKEP ACDDPDTEQA ALAAVDYINK HLPRGYKHTL NQIDSVKVWP
RRPTGEVYDI EIDTLETTCH VLDPTPLANC SVRQQTQHAV EGDCDIHVLK QDGQFSVLFT KCDSSPDSAE
DVRKLCPDCP LLAPLNDSRV VHAVEVALAT FNAESNGSYL QLVEISRAQF VPLPVSVSVE FAVAATDCIA
KEVVDPTKCN LLAEKQYGFC KGSVIQKALG GEDVRVTCTL FQTQPVIPQP QPDGAEAEAP SAVPDAAGPT
PSAAGPPVAS VVVGPSVVAV PLPLHRAHYD LRHTFSGVAS VESSSGEAFH VGKTPIVGQP SIPGGPVRLC
PGRIRYFKI
```

MH^+	Sequence	Position
816.4210	ALGGEDVR	[238-245]
833.3651	qYGFCK	[226-231]
850.3916	QYGFCK	[226-231]
870.4944	HLPRGYK	[51-57]
895.4706	CNLLAEK	[219-225]
919.5295	LCPGRIR	[349-355]
1154.6164	HTLNQIDSVK	[58-67]
1252.6208	qDGQFSVLFTK	[121-131]
1269.6474	QDGQFSVLFTK	[121-131]
1385.5638	CDSSPDSAEDVR	[132-143]
1428.7805	GSVIQKALGGEDVR	[232-245]
1445.7246	qYGFCKGSVIQK	[226-237]
1462.7511	QYGFCKGSVIQK	[226-237]
1474.8376	TPIVGQPSIPGGPVR	[334-348]
1502.7962	GYKHTLNQIDSVK	[55-67]
1513.6587	CDSSPDSAEDVRK	[132-144]
1663.8724	EVVDPTKCNLLAEK	[212-225]
1692.9180	HTLNQIDSVKVWPR	[58-71]
1726.8444	CNLLAEKQYGFCK	[219-231]
1836.9135	LCPDCPLLAPLNDSR	[145-159]
1965.0085	KLCPDCPLLAPLNDSR	[144-159]
2008.9545	qQTQHAVEGDCDIHVLK	[104-120]
2025.9811	QQTQHAVEGDCDIHVLK	[104-120]
2106.1641	TPIVGQPSIPGGPVRLCPGR	[334-353]
2120.0043	HTFSGVASVESSSGEAFHVGK	[313-333]
2454.1129	EPACDDPDTEQAALAAVDYINK	[29-50]

Sequence alignment of human (O94811) and bovine (Q27957) TPPP/p25

```
human     1   MADKAKPAKAANRTPPKSPGDPSKDRAAKRLSLESEGAGEGAAAS-PELS    49
              ||| ::| |.||:|||||||:|:||:|||||||:|||||||||: .|||
bovine    1   MAD-SRP-KPANKTPPKSPGEPAKDKAAKRLSLEAEGAGEGAAAAGAELS    48

human    50   ALEEAFRRFAVHGDARATGREMHGKNWSKLCKDCQVIDGRNVTVTDVDIV    99
              |||||||:|||||||||:|||||||||||||:|||||||:|||||||||
bovine   49   ALEEAFRKFAVHGDARASGREMHGKNWSKLCRDCQVIDGRSVTVTDVDIV    98

human   100   FSKIKGKSCRTITFEQFQEALEELAKKRFKDKSSEEAVREVHRLIEGKAP   149
              ||||||||||||||||:||||||||||||||||:|||||||:|||||||
bovine   99   FSKIKGKSCRTITFEQFKEALEELAKKRFKDKSAEEAVREVHKLIEGKAP   148

human   150   IISGVTKAISSPTVSRLTDTTKFTGSHKERFDPSGKGKGKAGRVDLVDES   199
              ||||||||||||||||||||:||||||||||||||:||||:||||||||||
bovine  149   IISGVTKAISSPTVSRLTDTSKFTGSHKERFDPSGRGKGRAGRVDLVDES   198

human   200   GYVSGYKHAGTYDQKVQGGK   219
              |||.||||||||||||||||
bovine  199   GYVPGYKHAGTYDQKVQGGK   218
```

Sequence of His-tagged *Medicago* RRK1 (C1L342)

MGSSHHHHHH SSGLVPRGSH MASMTGGQQM GRGSEFGTRR YIRTGSFKRL FSFTKRGLGE PVLSPKGEEN
EKSLKILPYE EETCQRPTWK CFSYEELFDA TNGFNSENMV GKGGYAEVYK GTLKNGEEIA VKRLTRASKD
ERKEKEFLTE IGTIGHVRHS NVLSLLGCCI DNGLYFVFEL STTGSVSSIL HDEKLAPLDW KTRHKIVVGT
ARGLHYLHKG CKRRIIHRDI KASNILLTKD FEPQISDFGL AKWLPSQWTH HSIAPIEGTF GHLAPEYYLH
GVVDEKTDVF AFGVFLLEVI SGRKPVDVSH QSLHSWAKPI LNKGDIEELV DARLEGEYDI TQLKRLAFAA
SLCIRASSTW RPSMTEVLEI MEEGEMDKEK WKMPEEEEEQ EEEFWGFEDL EYEYDSSFSM SLIDSIESN

(The His-tag sequence is underlined.)

Differently His-tagged *Medicago* ROP6 sequences

H_1

MGGSHHHHHH GMASMTGGQQ MGRDLYDDDD KDRWGSELEI RYRIPMSGSR FIKCVTVGDV AVGKTCLLIS
YTSNTFPTDY VPTVFDNFSA NVVVDGSTIN LGLWDTAGQE DYNRLRPLSY RGADVFLLAF SLISKASYEN
IAKKWIPELR HYAPGVPIIL VGTKLDLRDD SQFFQDHPGA APITTAQGEE LRKLIGAPVY IECSSKTQKN
VKAVFDSAIK VVLQPPKQKK TKRKGQKACS IL

H_2

MGGSHHHHHH GMASMTGGQQ MGRDLYDDDD KDWIRYRIPM SGSRFIKCVT VGDVAVGKTC LLISYTSNTF
PTDYVPTVFD NFSANVVVDG STINLGLWDT AGQEDYNRLR PLSYRGADVF LLAFSLISKA SYENIAKKWI
PELRHYAPGV PIILVGTKLD LRDDSQFFQD HPGAAPITTA QGEELRKLIG APVYIECSSK TQKNVKAVFD
SAIKVVLQPP KQKKTKRKGQ KACSIL

H_3

MGGSHHHHHH GMASMTGGQQ MGRDLYDDDD KDRWIRPRDL QMVPYGNSMS GSRFIKCVTV GDVAVGKTCL
LISYTSNTFP TDYVPTVFDN FSANVVVDGS TINLGLWDTA GQEDYNRLRP LSYRGADVFL LAFSLISKAS
YENIAKKWIP ELRHYAPGVP IILVGTKLDL RDDSQFFQDH PGAAPITTAQ GEELRKLIGA PVYIECSSKT
QKNVKAVFDS AIKVVLQPPK QKKTKRKGQK ACSIL

H_4

MSGSRFIKCV TVGDVAVGKT CLLISYTSNT FPTDYVPTVF DNFSANVVVD GSTINLGLWD TAGQEDYNRL
RPLSYRGADV FLLAFSLISK ASYENIAKKW IPELRHYAPG VPIILVGTKL DLRDDSQFFQ DHPGAAPITT
AQGEELRKLI GAPVYIECSS KTQKNVKAVF DSAIKVVLQP PKQKKTKRKG QKACSILLEH HHHHH

(The His-tag sequences are underlined.)

Sequence of *Drosophila* calpain B (Q9VT65)

```
MYGIDNYPKN YHASGIGLVN LAALGYSKNE VSGGNEGGGA PKPKAGLYPS LPYPSSESVG GMPYVVKQTS
HAQNASYAGP TMGMGMPVPE APSAPAPYPS ATPYPGSGLY PSLPSANVSS LPYPTAPMAP YPTGMPYPTG
MPQPNLPYPA APLAPYPSAM PGLPGMPMPY APMPTSPAPQ HNIGFPALPY PTAPPPESAP TQEEEPSVGV
AELSFTSVKV PENQNMFWMG RKATSARQNS VSKGDFQSLR DSCLANGTMF EDPDFPATNA SLMYSRRPDR
YYEWLRPGDI ADDPQFFVEG YSRFDVQQGE LGDCWLLAAA ANLTQDSTLF FRVIPPDQDF QENYAGIFHF
KFWQYGKWVE VVIDDRLPTY NGELIYMHST EKNEFWSALL EKAYAKLHGS YEALKGGTTC EAMEDFTGGV
TEWYDIKEAP PNLFSIMMKA AERGSMMGCS LEPDPHVLEA ETPQGLIRGH AYSITKVCLM DISTPNRQGK
LPMIRMRNPW GNDAEWSGPW SDSSPEWRFI PEHTKEEIGL NFDRDGEFWM SFQDFLNHFD RVEICNLSPD
SLTEDQQHSS RRKWEMSMFE GEWTSGVTAG GCRNFLETFW HNPQYIISLE DPDDEDDDGK CTAIVALMQK
NRRSKRNVGI DCLTIGFAIY HLTDRDMQVK PQGLNFFKYR ASVARSPHFI NTREVCARFK LPPGHYLIVP
STFDPNEEGE FIIRVFSETR NNMEENDDEV GFGETDDRIA PSLPPPTPKE EDDPQRIALR RLFDSVAGSD
EEVDWQELKR ILDHSMRDVM VGSDGFSKDA VRSMVAMLDK DRSGRLGFEE FEALLTDIAK WRAVFKLYDT
RRTGSIDGFH LRGALNSAGY HLNNRLLNAL AHRYGSREGQ IPFDDFLMCA IKVRTFIEMF RERDTDNSDT
AFFNLDDWLE RTIYS
```

Sequences of α-crystallin A (P02470) and B (P02510) chains

α-crystallin A:

```
MDIAIQHPWF KRTLGPFYPS RLFDQFFGEG LFEYDLLPFL SSTISPYYRQ SLFRTVLDSG ISEVRSDRDK
FVIFLDVKHF SPEDLTVKVQ EDFVEIHGKH NERQDDHGYI SREFHRRYRL PSNVDQSALS CSLSADGMLT
FSGPKIPSGV DAGHSERAIP VSREEKPSSA PSS
```

α-crystallin B:

```
MDIAIHHPWI RRPFFPFHSP SRLFDQFFGE HLLESDLFPA STSLSPFYLR PPSFLRAPSW IDTGLSEMRL
EKDRFSVNLD VKHFSPEELK VKVLGDVIEV HGKHEERQDE HGFISREFHR KYRIPADVDP LAITSSLSSD
GVLTVNGPRK QASGPERTIP ITREEKPAVT AAPKK
```